The evolution of longevity in the context of epigenetic regulation and genetic sequence evolution

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Human Evolutionary Biology

have examined a dissertation entitled

The evolution of longevity in the context of epigenetic regulation and genetic sequence evolution

presented by Daniel Joseph Roger Richard

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature Terence D. Capellini

Typed name: Prof. Terence D. Capellini

Signature _____

Typed name: Prof. Daniel Lieberman

Signature David Reich

Typed name: Prof. David Reich

Signature _____

Typed name: Prof. Arcadi Navarro

Signature _____

Typed name: Prof.

Date: 8/31/2022

The evolution of longevity in the context of epigenetic regulation and genetic sequence evolution

A dissertation presented

By

Daniel Joseph Roger Richard

To

The Department of Human Evolutionary Biology

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

In the subject of

Human Evolutionary Biology

Harvard University

Cambridge, Massachusetts

August 2022

Dissertation Advisor: Terence D. Capellini Daniel Joseph Roger Richard

The evolution of longevity in the context of epigenetic regulation and genetic sequence evolution

Abstract

Modern-day humans enjoy an elongated lifespan relative to chimpanzees, even when accounting for environmental factors such as healthcare. This suggests that genetic changes to the biological processes underlying longevity have taken place over the course of human evolution, changes resultant from the forces of natural selection. The goal of this dissertation is to explore the potential means by which these forces of natural selection may operate to shape the human genome, and the implications of genetic changes for the manifestation of ageing phenotypes – in particular, focusing on the incidence of late-onset diseases such as osteoarthritis. Over three data chapters, I will approach this goal in several ways: (1) considering the role for selection acting on development of a derived human trait in influencing the genetic risk for development of late-onset disease, (2) how shifts in epigenetic regulation over the course of development and aging influence the sequence properties and disease associations of genetic variants, and (3) how selection operating across long-lived species may act at the protein-coding level to alter the activity and function of aging-associated proteins.

In the second chapter, I characterize the regulatory landscape of early knee development in mouse and human fetal tissues and consider the role that modifications to regulatory elements may have not only in the development of the knee, but in the potential for far-reaching consequences on knee disease later in life. Through my analyses, I develop an evolutionary model to describe the relationship between development, natural selection, and heritable risk for osteoarthritis incidence later in life. Ancient directional, and subsequent purifying, selection acts on developing knee regulatory regions, establishing and maintaining the derived human knee configuration; this imposes strong functional sequence constraint. More recently, genetic variants arising in these constrained regions via the effects of antagonistic pleiotropy

or random genetic drift contribute to a ‘violation of constraint’. This violation can lead to slight changes to knee morphology which may eventually contribute to knee dysfunction in late-life. Evidence in support of this model is established through the demonstration of changes to knee morphology, and incidence of spontaneous osteoarthritis, via functional experiments within the *GDF5* locus, and the finding that human osteoarthritis patients have elevated mutational load within these constrained knee regulatory regions relative to background populations.

In the third chapter I consider the interactions between developmental and aging processes more broadly, focusing on shifts in epigenetic regulation occurring across tissues as they transition from fetal to adult forms, and subsequently mature from a young-adult to old-adult state. I find that epigenetic trends occurring over development are continued across ageing, and that the genetic sequences subject to these trends have markedly divergent sequence properties depending on their directionality – suggesting that evolutionary pressures and regulatory activities act in a context-dependent manner. Furthermore, these sequences also diverge in their associations with heritable risk for late-onset diseases. I develop an evolutionary model in which sequences most epigenetically active in early adulthood have the strongest association with heritable risk for late-onset disease, while the behaviour of sequences most active during early development or far later in life suggest limitations to previously established evolutionary models of aging.

In the fourth chapter I consider the potential for natural selection favouring longevity to act at the level of protein-sequence changes. Building upon previous studies, which take a cross-species approach to identifying amino acid substitutions which stratify across long/short-lived organisms, I explore the nature of these sequence changes in terms of their potential effects on protein structures and functions, as well as their distribution across the proteome. I find that these amino acid substitutions are over-represented in protein complexes, and in particular the interfaces between interacting proteins. I go on to develop an evolutionary model in which natural selection favouring longevity may act at the level of complex formation, rather than modifying the catalytic activities of proteins themselves, so as to avoid the prohibitive sequence constraints imposed by pleiotropic effects.

Overall, this thesis establishes three evolutionary models to describe the results of findings made throughout the three data chapters, and in the discussion of the thesis we integrate these three models into a more general framework. These models are not mutually exclusive and instead may be complementary in explaining different aspects of the broader evolutionary program underlying the extension of lifespan along the human lineage.

Table of Contents

List of Figures

List of Tables

All supplementary tables are included as excel supplemental tables along with the submission of this dissertation.

Acknowledgments

I would like to first and foremost thank my advisor, Dr. Terence D. Capellini, for all the time, effort and thought he has shared with me over the years. It has been a privilege to get to work alongside him as a colleague and collaborator, and our early-morning coffee walks will be fondly remembered. He never fails to go above and beyond for his students and he is truly a remarkable mentor. To me, having worked under Dr. Capellini is one of my greatest good fortunes.

I would also like to thank Dr. Daniel Lieberman, for all of his advice and mentorship over the years, and for being a most excellent and compelling running companion. I would also like to thank the other members of my committee: Dr. Arcadi Navarro for being a steadfast collaborator and a researcher whose work I admire, and Dr. David Reich, whose penchant for asking insightful questions I have much appreciated. I'd also like to thank Drs. Maryellen Ruvolo and David Pilbeam for taking me under their wings so many years ago, and showing me how to be an evolutionary researcher as I was just starting out. To the past and present members of the Capellini Lab, I'd like to thank you for joining me in this experience, and for all the help, guidance and support you've given me. In particular I'd like to thank Dr. Pushpanathan Muthuirulan, for always being patient, kind, and devoted in helping others both in and out of the lab.

I'd also like to express my sincerest gratitude to Dr. Andrew Doxey, for taking a chance on an unpolished undergraduate so many years ago and starting me down the path to where I am today.

I want to express my deepest thanks to my support network for all that they've done for me over the course of my graduate career. To my friends, the McBrides, for being my second family, and Peter McBride for always being there for me, through it all. To my grandparents, Roger and Marilyn Richard, for always having an open home and fresh cookies at the ready. To Donald Gavas, who has always been a phone call away, and always wants to know what I'm up to. To my Tita, who is my biggest fan and never fails to bring a smile to my face. To my brother, for being a strong support and a stronger friend. To my

father who, through it all, has shown me nothing but pride and support. To my sister, for being my rock, and being so present and so supportive.

And finally, to my mother, for always being in my corner, always having my back, and for being the person I turn to. She has been a role model for me in how one can succeed regardless of the odds, and she has always taught me that hard work and learning were the only things that one can ever lose and always improve. She has been there for me every step of the way, and I cannot count the number of late-night phone calls and conversations where she's been my guiding beacon, my voice of reason, and my champion.

To all of these people I cannot express how much your support has meant to me, and I would not have made it here without all of you. Thank you.

Chapter I: Introduction

1.1: Life history as an evolutionary trait

One universal truth of nearly all organisms on Earth is that of expiration – a combination of extrinsic and intrinsic factors which contributes to gradually increasing mortality risk and, consequently, decreased individual survivorship over time. Mortality curves compared across a diverse range of species and taxonomic groups(Croft et al., 2015; Jones et al., 2014; Tejada-Martinez et al., 2022) demonstrates considerable variation in survivorship trajectories which are likely the result of evolutionary interactions between a species' physiology (i.e., heritable biological traits) and its environment. Mammals, broadly share similar patterns of age-dependent mortality and aging phenotypes – including an age-dependent mortality increase (Jones et al., 2014; Magalhães and Toussaint, 2002), suggesting that selection occurring quite anciently in the evolution of mammals acted on biological processes mediating functional decline. Within this general structure, however, exists substantial variation in aging rates(Austad, 1997) and overall lifespan, suggesting that, at a more rapid time-scale (e.g., within a given group, such as rodents and the naked mole rat(Lewis et al., 2016)) evolutionary innovations that accelerate or delay the aging process may have occurred within this background of mammalian aging physiology.

A key example of recent evolutionary processes shaping functional decline is in comparing life-histories and maximal longevities of humans and chimpanzees. In particular, modern humans exhibit an extended period of reduced early- and mid-life mortality, followed by a marked acceleration in mortality risk, a pattern far reduced in chimps(Jones et al., 2014). Humans also feature elevated estimates for both average life-expectancy and maximal lifespan compared to their wild or captive chimp counterparts(Gurven, M., Gomes, 2017). While in the past it has been suggested that the remarkable longevity of humans is largely an artifact of modern-day civilization(Williams, 1957), demographic observations of hunter-gatherers challenge this notion. These groups enjoy increased life-expectancy throughout life relative to both captive and wild chimps(Croft et al., 2015; Gurven and Kaplan, 2007), and in particular experience a depressed age-specific mortality for their first forty years(Gurven, M., Gomes,

2017), mirroring that observed in industrial populations(Croft et al., 2015; Jones et al., 2014). Thus, there appears to be convergence of these species-level shifts across traditional and industrial humans (summarized in Table 1.0).

Given the presence of similar extrinsic risk factors and causes of death between traditional humans and wild chimpanzees(Gurven, M., Gomes, 2017) the persistent extension of human longevity may result more from modifications to intrinsic physiological mechanisms, such as a decline in tissue maintenance and related function (i.e., senescence), than to external environmental interactions. This point is further suggested by observations that captive chimps, when benefiting from the same improvements in healthcare and diet (i.e., those that likely contribute to pre- and post-industrial human differences in lifespan and declines in mortality)(Finch, 2010; Riley, 2001), continue to lag behind humans in terms of both life expectancy and maximal lifespan(Gurven, M., Gomes, 2017). The increased importance of species-level differences, rather than environmental context, suggests that the modifications to these intrinsic physiological mechanisms may be encoded genetically, and that natural selection may have driven changes to life-history and the aging trajectory over the course of human evolution. It is worth noting, however, that other environmental influences, such as increased physical activity in humans, may also contribute to differences in maximal lifespan, and selection for increased physical activity may have indirectly lead to slowed senescence (Lieberman et al., 2021).

In this dissertation, we seek to understand the manner by which natural selection favouring longevity may have operated to shape the human aging process. Through our analyses leveraging a combination of novel experimental and public datasets, we develop models describing the possible nature of these evolutionary programs acting to genetically modify the regulation of genes involved in both developmental and adult-homeostatic biological processes, as well as modifying the sequences of these genes themselves. These models highlight the importance of antagonistic pleiotropic effects and extend this concept further to suggest a more nuanced confluence of evolutionary forces at play. Our work points to several avenues of future research to move towards a more complete understanding of the selective mechanisms and proximate biological mechanisms underlying the evolution of longevity.

Organism	Trait	Trait Value
Homo sapiens (traditional hunter-gatherers)	Average life-expectancy at fifteen years of age	37.7 (Gurven and Kaplan, 2007)
Homo sapiens	Maximum Observed Lifespan	122 (Robine and Allard, 1998)
Pan troglodytes (wild)	Average life-expectancy at fifteen years of age	14 (Gurven and Kaplan, 2007)
Pan troglodytes	Maximum Observed/Inferred Lifespan	54-66 (Bronikowski et al., 2011; Gurven, M., Gomes, 2017)
Homo sapiens	Maximum Estimated Lifespan[a]	126.1 (Weon and Je, n.d.)
Homo sapiens	Maximum Estimated Lifespan[b]	66-72 (Bogin and Smith, 1996; Hammer and Foley, 1996; Judge and Carey, 2000) (91 when *H.sapiens* excluded from predicting model)(Judge and Carey, 2000)
Australopithecus	Maximum Estimated Lifespan[b]	42-44 (Judge and Carey, 2000)
Homo habilis	Maximum Estimated Lifespan[b]	52-56 (Hammer and Foley, 1996; Judge and Carey, 2000)
Homo erectus	Maximum Estimated Lifespan[b]	60-63 (Hammer and Foley, 1996; Judge and Carey, 2000)

Table 1.0: Lifespan/expectancies for extinct/extant primates.

A: Extended Weibull model based on survival probabilities using life tables from an industrial population (Sweden)

B: Regression analyses on brain/body mass and lifespan using observations from a mix of extant primates and traditional/industrialized humans.

1.2: Evolution and life-history theory

Senescence may be broadly defined as the set of adverse biological changes in an organism which manifest as increased mortality risk in advancing chronological age(Comfort, 1954). Life history theory predicts a synchronization of somatic (i.e., adult tissue) and reproductive senescence, and thus a lack of a post-reproductive lifespan(Williams, 1957). This synchronization – wherein organisms would tend to be capable of reproduction up until death, has been suggested to be evolutionarily favoured under several different scenarios; below we describe three leading evolutionary models. It is important to note that all three of these models are not competing, but instead are complementary in explaining the presence of genetic factors that either directly promote early-life fitness benefits or permit late-life fitness costs.

The concept of diminished selective pressures acting on older and post-reproductive individuals is a central tenant to a third model for the evolution of age-related decline - the Mutation Accumulation Hypothesis(Charlesworth, 2001; Turan et al., 2019). This model proposes that age-related decline is driven, in part, through the accumulation of deleterious mutations impacting adult biological processes. This accumulation is permitted given the relative weakness of negative selective pressures acting later in life, a phenomenon often termed the ‘selective shadow’ of selective pressures. There is less experimental evidence in model organisms favouring this hypothesis(Elliott et al., 2002; Kaya et al., 2015), though the aforementioned GWAS analysis did find suggestive evidence in support of this model using human genetic data(Rodríguez et al., 2017).

The theory of antagonistic pleiotropy (AP) (Williams, 1957) broadly postulates that, given the pleiotropic roles of protein-coding genes within cells, genetic mutations to genes that confer some selective benefit to early-life development may have additional effects that manifest as negative consequences at a later stage of development (e.g., late adulthood). For example, mutations that confer a beneficial developmental growth phenotype, such as increased body size or height, could have pathogenic consequences far later in life, such as increasing the genetic risk of individuals to developing adult-onset diseases (e.g., osteoarthritis)(Capellini et al., 2017). Positive selection favouring mutations with early-life

benefits is substantially stronger than the forces of negative selection that would discourage mutations with late-life consequences; this is largely driven by the fact that fitness consequences decline with age, such that pressures favouring/discouraging mutations returns diminishing gains to fitness (Carnes and Olshansky, 1993; Ricklefs, 2008). There is now substantial evidence from model organisms to support antagonistic pleiotropy as being a very common, if not ubiquitous, phenomenon(Austad and Hoffman, 2018; Byars and Voskarides, 2020), and there is also some evidence to suggest that antagonistic pleiotropy may contribute to human late-life phenotypes as well(Byars and Voskarides, 2020). Furthermore, analysis of Genome-Wide Association Study (GWAS) datasets from early- and late-onset diseases yields results consistent with the effects of antagonistic pleitropy, wherein for example variants associated with old-onset disease tend to exist at higher frequencies. Moreover, there is an excess of variants which are pleiotropic across early- and late-onset diseases(Rodríguez et al., 2017).

The 'Disposable Soma' theory, which represents a special form of antagonistic pleiotropy (Rodriguez et al., 2017), suggests the existence of a functional trade-off between prioritizing physiological mechanisms that favour somatic tissue maintenance, and those favouring reproductive tissue activity and maintenance given a finite resource and the functional constraint of gene regulatory networks(Kirkwood and Holliday, 1979; Kirkwood and Rose, 1991). There is evidence to support the existence of such trade-off scenarios in wild populations of many different vertebrate species(Lemaître et al., 2015).

Given these models, the presence of a consistent post-reproductive lifespan is a unique, unexpected feature of humans(Alberts et al., 2013), with equivalents confined to similarly long-lived cetaceans(Bogin and Smith, 1996; Croft et al., 2015; Ellison, P.T., Ottinger, 2014). Theories such as the mother and (active) grandmother/parent hypotheses (Hamilton, 1966; Hawkes et al., 1998; Lieberman et al., 2021; Williams, 1957), and the 'embodied capital hypothesis'(Gurven, M., Gomes, 2017; Gurven, M., Kaplan, 2008; Kaplan, Hillard, Lancaster, Jane, Robson, 2003) have been proposed to explain how, despite the opposing selective forces (as described above) that would contribute to age-related decline, such a life-history trait might emerge. These hypotheses stem from observations that human children require extensive provisioning and develop more slowly(Bogin and Smith, 1996; Schuppli et al., 2012). The

grandparent hypothesis, for example, suggests that indirect fitness benefits accrued by longer-living individuals, within the context of substantial food-sharing thought to have emerged by early *Homo*(Pontzer, 2012), may have been sufficient to select for increased longevity. Similarly, the embodied capital hypothesis suggests that the extensive skill set required for successful exploitation of the human dietary niche takes a substantial time to develop, and thus can only become beneficial in terms of net caloric returns ('payoff') when the period of productive activity extends into later age, thus selecting for extension of lifespan to maximize provisioning of relatives(Gurven, M., Gomes, 2017; Gurven, M., Kaplan, 2008; Kaplan, Hillard, Lancaster, Jane, Robson, 2003). These two hypotheses need not be mutually exclusive; if anything, one might consider increasing caloric transfer to offspring through lifespan extension to be a general selective mechanism favoring longevity of both sexually active and post-reproductive individuals, and that the gradual increase in complexity of the human dietary niche would have acted synergistically with pre-existing selection for longevity through an 'additional payoff' scheme.

1.3: Lifespan shifts in the evolution of *Homo*

The above discussion on evolutionary life-history theories for delayed late-life mortality and overall longevity suggest the importance of key transitions during human evolution that likely altered factors such as extrinsic mortality risks, energy availability, social dynamics, etc. These transitions likely coincide with, and in fact favour, modifications to the intrinsic physiological mechanisms governing age-related decline, modifications which may have driven changes to human life history. Examining these transitions provides a context for how, when, and why genetic changes to these mechanisms were selected upon in humans. Regression analyses using brain and body masses from a wide range of extant primates and fossil remains (see Table 1.0) suggest that expanded maximal lifespans relative to *Pan* may have post-dated *Australopithecus* and arose with the emergence of *Homo*(Bogin and Smith, 1996; Hammer and Foley, 1996; Judge and Carey, 2000). These analyses also suggest further increases in lifespan between *H.habilis* and *H.erectus*, with additional increases in *H.sapiens*, albeit the latter are not expected to differ greatly from patterns in Neandertals (see(Trinkaus, 2011)). We note that these differences in lifespan are based on extrapolations of brain/body-mass, and do not take into account differences in extrinsic mortality factors between species of *Homo*, which could influence life expectancy rather than maximal theoretical

longevity. If one assumes an age at menopause of 50 years (Bogin and Smith, 1996; Isler and van Schaik, 2012; Judge and Carey, 2000), noting there is evidence for variation around this average (see (Emery Thompson M., 2017; He et al., 2009)), this would imply that early *Homo* was at least capable of living into a post-reproductive state, which coincides with the earliest predicted selection scenarios for grandmothering (Hawkes et al., 1998). Further, mathematical models attempting to consolidate increased brain size with energetics-imposed life-history constraints (e.g., delayed growth, suppressed reproductive effort) suggest that substantial alloparenting would have been necessary to support the endocranial volumes of ~600cm^3 as seen in early *Homo* (Hublin et al., 2015) while allowing for realistic population growth (Isler and van Schaik, 2012). Thus, the biological possibility for survival into post-reproductive life, paired with a requisite selection pressure for increased offspring provisioning would strongly imply selection for extended lifespan to have played a role in the evolution of *Homo*. Note, that this selection would not be limited to reduction of late-life degeneration; delayed tissue senescence at all ages (i.e., 'healthspan' (Kirkland and Peterson, 2009)) would cumulatively contribute to increases in post-reproductive lifespan, with the above kin-selection models predicting greater fitness returns the longer functionality (e.g., foraging ability) is extended later into an elongated life(Lieberman et al., 2021).

1.4: Genetic changes and longevity

As mentioned above, the consistency in lifespan differences observed between humans and chimps across different environmental contexts suggests modifications to intrinsic physiological mechanisms involved in tissue maintenance and decline (i.e., aging). These derived changes would need to have been genetically-encoded (i.e., heritable) in order for natural selection to operate upon them. However, genetic changes are not generated in a vacuum; they occur in the context of dense, complex gene regulatory networks that govern cellular behaviours. These gene regulatory networks are controlled at multiple levels; these include (1) the epigenetic level, which regulates the transcriptional potential of certain gene loci based on the presence and nature of chemical modifications to packaged genomic DNA, (2) the transcriptional level, the process which actually generates RNA transcripts from epigenetically-active loci, and (3) the post-transcriptional level, including both the translation process as well as post-translational modifications (as well as interactions with other proteins, e.g., to form complexes). Upstream of these

regulatory networks are the gene sequences themselves – that is, changes to the proteins encoded by gene loci also stand to impact cellular functions by modifying the activity and effects of protein ligands, enzymes, structural components, etc. Thus, genetic changes that are selected upon in the human lineage must impart a fitness-altering phenotype by means of modifying either gene-regulatory activity governing expression changes, or by directly modifying the encoded protein sequences themselves.

Many genetic studies have suggested that modifications to the regulatory sequences active in gene regulatory networks likely drive some of the key phenotypic differences observed between primates(King and Wilson, 1975; Rogers and Gibbs, 2014). The logic being that regulatory sequences (e.g., enhancers) often exhibit modularity – acting in a tissue or context-specific manner to facilitate fine-grained control of expression(Prud'homme et al., 2007; Rubinstein and de Souza, 2013), such that their modification has a lower potential for unintended deleterious consequences as a result of pleiotropic effects, relative to the coding sequences of the genes being regulated. It should also be noted that while there is variation in lifespan in modern-day humans, a component of which can be explained by genetic traits(Timmers et al., 2019), the averages in human and chimpanzee demographic patterns are markedly shifted, pointing to the importance of fixed genetic differences in mediating lifespan differences. We therefore consider the possible effects of both fixed and variable human sequence changes in our studies.

Additionally, while most gene sequences are much more highly conserved (given the heightened potential for pleiotropic effects) several studies have pointed to the role that protein-coding substitutions may have in contributing to the evolution of altered longevity(Farré et al., 2021a; Jobson et al., 2010; Kowalczyk et al., 2020; Li and de Magalhães, 2011; Muntané et al., 2018; Treaster et al., 2021). Thus, when studying the genetic evolution of lifespan in humans, the works described in this thesis include considerations for changes at both the *cis*- and *trans*-regulatory and protein sequence levels.

1.5: Overview of thesis

In the studies described in this thesis, I will consider three possible models describing the manner by which genetic sequence changes may have been shaped by evolutionary forces in order to modify the

human aging process, including age-related diseases. Chapter Two outlines a model wherein strong directional selection for a crucial part of human anatomy, the structure of the bipedal hindlimb knee joint, imparts a functional constraint on developmental gene regulatory networks. Subsequent violations of these constraint may take two forms: (1) direct antagonistic pleiotropy selecting genetic mutations favurable for another trait, and (2) genetic drift occurring in modern humans. This violation manifests as increased genetic risk for the development of a highly prevelant age-related knee pathology, osteoarthritis, possibly by means of slightly altering the bony structure of the knee which elicits improper joint loading – the effects of which are moderate early in life but compound over decades of joint use.

Chapter Three seeks to continue the concepts addressed in Chapter Two – i.e., the contribution of developmental gene-regulatory changes to the risk of late-life disease - and considers development and aging in a broader sense by taking a cross-tissue and cross-disease approach. By defining the epigenetic context of genomic regions active or inactive across tissues, and further how this context changes as tissues develop from their fetal to adult forms, we define epigenetic patterns suggesting a shared cross-tissue trajectory in how gene networks are regulated. By further comparing epigenetic shifts across young- and old-adult tissue samples, we study the interactions between development- and aging-associated epigenetic changes and focus on the nature of the sequences subject to modifications by these two processes. Through our analyses, we develop a novel model in which the shifting epigenetic context of mutations impacts their apparent contribution to late-life disease phenotypes, and thus influences the behaviour and strength of natural selection acting on these genetic variants.

In Chapter Four, we continue the emphasis placed on epigenetic processes as key players in the aging process by considering changes to the sequences of epigenetic proteins themselves, which in turn may significantly impact their roles in shaping the epigenome. We build on previous works(Farré et al., 2021a) and take a phylogenetic approach to protein evolution, defining amino acid substitutions that stratify across long- and short-lived species, and consider the potential consequences of these substitutions on protein structure and function. Through our analyses we propose a model in which natural selection acting to shape the aging process in different species has modified proteins not just at the single-structure

level, but also at the biochemical complex level, modifying the interaction domains of protein pairs as an additional means of modifying cellular mechanisms.

Finally, in the Discussion and Further Directions sections, I consider the merits and limitations of each of the three models described, as well as the potential for overlap and synergisms between them. I further discuss how using an evolutionary lens to understand the derived human lifespan, as well as the mechanisms by which natural selection may have acted, can help us to uncover novel aspects of aging biology. Future directions for this evolutionary line of research are discussed, with an emphasis for how selective models can inform the design of future studies which extend our understanding of longevity selection and aging-associated biology.

Chapter II

Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk

This chapter has been published as an article in *Cell*:

Richard, D., Liu, Z., Cao, J., Kiapour, A.M., Willen, J., Yarlagadda, S., Jagoda, E., Kolachalama, V.B., Sieker, J.T., Chang, G.H. and Muthuirulan, P., 2020. Evolutionary selection and constraint on human knee chondrocyte regulation impacts osteoarthritis risk. *Cell*, 181(2), pp.362-381.

2.1 Introduction

A derived feature of humans and our hominin ancestors is a unique bipedal gait (Darwin, 1888). From the last common ancestor with chimpanzees, natural selection shaped the hominin hindlimb to accommodate the biomechanical demands of bipedality, profoundly influencing human knee anatomy (Frelat et al., 2017; Jungers, 1988; Morrison, 1970; Rose, 1991). For example, the condyles of the distal femur expanded, dissipating the higher forces generated during bipedal walking and running. The proximal plateau and condyles of the tibia became more symmetrical, its epicondyles buttressed with bone mass, permitting more equal weight distribution and greater knee stabilization (Bramble and Lieberman, 2004; Lovejoy, 2007). These and other derived features were approximately fixed by 1.6 million years ago in *Homo erectus* and today variation in human knees is reduced compared to earlier *Homo* and differs from patterns found in African apes (Frelat et al., 2017; Lovejoy, 2007; Tardieu, 1999).

Derived human traits have been shaped by selection and its targeting of early development (Capra et al., 2013; Gokhman et al., 2017; Kanton et al., 2019; Tardieu, 1999). The development of knee features,

many present by birth, is tied to the regulation of chondrocytes that prefigure joints (Andersen, 1961; Gardner and O'Rahilly, 1968; Gray and Gardner, 1950; Mérida-Velasco et al., 1997; O'Rahilly, 1951). During knee development chondrocytes differentiate within the early joint to form femoral and tibial articular cartilage, while directly adjacent to the joint they arise to form the epiphyseal cartilages of the condyles and plateau (Decker et al., 2014). These cartilages remain protected from ossification until postnatal life to respond to mechanical loading and hormonal signals during growth (Tardieu, 2010). Adaptive changes to these developmental processes underlie key anatomical features of the unique human knee. To this end, modifications to chondrocyte networks, in the form of regulatory sequence changes, are the likely candidates as they typically are modular in effect (King and Wilson, 1975; McLean et al., 2011; Varki and Altheide, 2005). Here, we first investigate the evolutionary basis of human knee development by examining accessible/open chromatin sequences, regions often involved in gene regulation (e.g., enhancers) (Shlyueva et al., 2014), via epigenomic profiling of long bone chondrocytes in human and mouse. We find evidence of the effects of ancient selection on knee-specific regulatory elements that likely underlie derived morphology in humans.

Natural selection undoubtedly shaped human knee morphology but not without impacts to health. In the elderly today there is a high prevalence of knee osteoarthritis (OA), a condition in which joint- and epiphyseal-derived tissues deteriorate. Knee OA risk includes non-genetic factors, such as abnormal joint mechanics (Barr et al., 2016; Neogi and Felson, 2016; Neogi et al., 2013), obesity (Felson et al., 1988; Reyes et al., 2016), inflammation (Houard et al., 2013; Richette et al., 2011), and increased longevity (Berenbaum et al., 2018). However, knee OA risk is also ~40% heritable (Loughlin, 2015), a component that likely interacts with these modern environmental conditions. GWAS have revealed over 80 OA loci with ~95% of risk variants present in non-coding sequences, enriched near genes involved in chondrocyte and bone development (Klein et al., 2019; Miyamoto et al., 2007; Styrkarsdottir et al., 2018; Tachmazidou et al., 2019; Zengini et al., 2018). These findings suggest a key role for knee chondrocyte regulatory elements in mediating OA risk. Here, we build on our evolutionary analyses, and use chondrocyte open chromatin datasets to study how regulatory variation, shaped by ancient selection but also more recent evolutionary forces (genetic drift and antagonistic pleiotropy), influenced OA risk. We develop a model

that posits how nucleotide changes, occurring within evolutionarily constrained developmental regulatory sequences, violate this constraint and underlie genetic OA risk, and we find evidence using at-risk patient data and functional validation studies in human chondrocytes and the mouse model. Overall, our methods link an evolutionarily novel aspect of human anatomy to its pathogenesis, an approach that can be applied to other derived features to explore disease links.

2.2: Results

2.2.1: Epigenetic Profiling of Mouse and Human Long Bone Chondrocytes

To investigate the human knee from an evolutionary developmental genetics perspective, we used the assay for transposase-accessible chromatin with sequencing (ATAC-seq) (Buenrostro et al., 2013) to profile open chromatin regions (hereafter, “elements”) in chondrocytes from E15.5 mouse and stage-matched E59 human hindlimb proximal/distal femur and tibia samples (Methods, Fig. 2.1a, Supplementary Table S2.1). For the mouse, we catalogued a “general knee set” comprising elements from the distal femur and proximal tibia. This set was filtered for shared proximal femur, distal tibia, and embryonic brain elements to make a “knee-specific set” (Supplementary Table S2.2), consisting of elements unique to the distal femur (“DF-specific” set), the proximal tibia (“PT-specific” set) and both knee components (“knee-common (KC)-specific” set). An identical approach was used on homologous forelimb humeri and radii, and generated “elbow-specific” sets show minimal element overlap with knee-specific sets (Supplementary Table S2.2), highlighting each sets’ anatomical specificity. Comparing these datasets to those using FACS on *Col2a1*-eCFP-reporter mice, we found strong and significant overlap (Supplementary Table S2.2), indicating that dissection methods identify general chondrocyte cell populations typically marked by COL2A1. The lack of reliable FACs markers to isolate chondrocytes from procured human tissues makes this quality-control step important to ensure the reliability of dissection on rare human samples. We thus generated ATAC-seq element sets on E59 human skeletal structures using the same experimental/computational pipelines, and intersected human and mouse elements from equivalent tissues (e.g., distal-femur), finding a significant degree of overlap (Supplementary Tables S2.1, 2.2). Overlapping mouse and human element sets show substantial one-to-one orthology (Supplementary Table S2.1). Though there is regulatory divergence between species (Yue et al., 2014), this overlap is

sizable despite disparate collection sites/protocols and uncontrollable differences in time-points. Mouse and overlapping E59 human elements intersect with biological signals related to chondrocytes and skeletal morphogenesis as determined via GREAT (McLean et al., 2010) (Supplementary Table S2.1). We also found significant overlap with regulatory activity in other mouse (Guo et al., 2017) and human E47 limb samples (Cotney et al., 2013) and bone marrow-derived chondrocytes (BMDCs, ENCODE Project Consortium, 2012) (Supplementary Table S2.2). We further observed enrichments for motifs similar to chondrogenesis-associated transcription factors (TFs), which show significant overlap with ChIP-seq datasets (Supplementary Table S2.2).

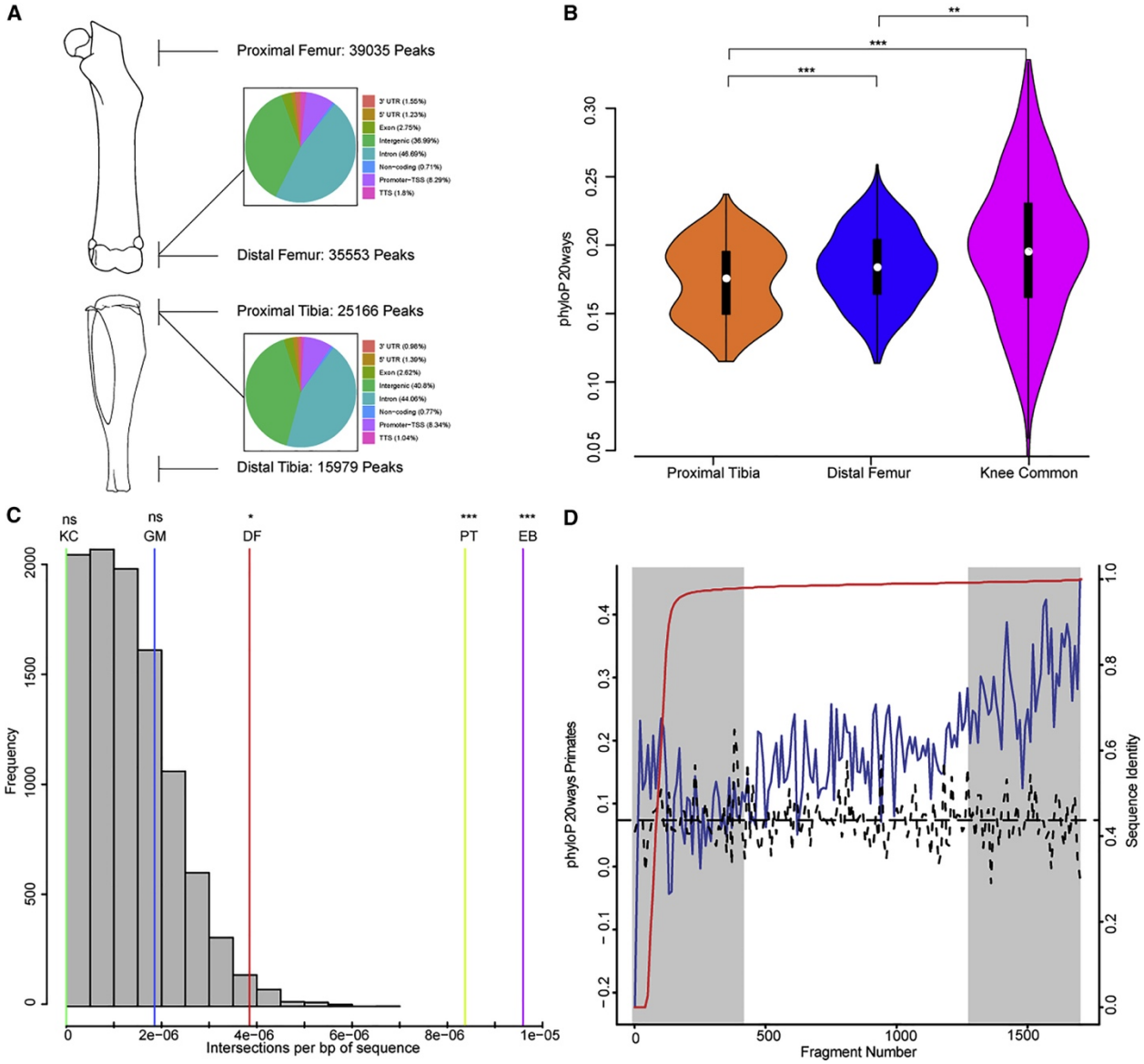


Figure 2.1: Sequence Features of Chondrocyte Epigenetic Profiles

(A) Diagram of ATAC-seq tissues. Numbers indicate replicate-consolidated peak calls before filtering with brain data. Inset: genomic distribution of lifted-over DF (top) and PT (bottom) elements.
(B) Per-bp conservation scores for 20 primates (phyloP20ways) aggregated for ATAC-seq sets and compared via Wilcoxon rank-sum test.
(C) Overlaps of human acceleration regions and ATAC-seq sets. KC, knee-common; GM, GM12878; DF, distal femur-specific; PT, proximal tibia-specific; EB, embryonic brain. Overlaps shown relative to set size (per bp of sequence) for background (gray) and target (colored) sets.
(D) Relationship between human-chimp sequence identity (red line, right y axis) and phyloP20ways score averaged over an ATAC-seq peak (blue line, left y axis) for DF-specific elements. Dashed curve and horizontal lines indicate average phyloP20ways scores for random background region set. Shaded regions reflect low/high %ID regions used in GREAT. Significance codes: not significant (ns), < 0.05 (*), < 0.01 (**), < 1e−5 (***).

2.2.2: Inter-species Sequence Variation in Element Sets

To understand how accessible elements involved in human knee development have been modified evolutionarily, we broadly examined how primates with diverse locomotor repertoires (Polk et al., 2009) modified these elements. We examined mouse and overlapping E59 human knee element sets for evidence of sequence evolution using phyloP (Pollard et al., 2010), a measure of nucleotide conservation/acceleration. Across primates, knee elements had elevated conservation, with strongest conservation towards each elements' center (Supplemental Figure S2.1), likely due to conserved TF sites (Buenrostro et al., 2013). When comparing DF-, PT-, and KC-specific sets, conservation tended to be highest in KC-specific elements - i.e., those that are likely pleiotropic across the knee (Fig. S2.1b, Supplementary Table S2.2). Conservation was reduced in elements unique to each bone's end, especially for the proximal tibia. These patterns were also observed for overlapping E59 human elements (Supplementary Table S2.2).

We next examined signatures of evolution shapccelerated regions) relative to chimps and other apes, often indicative of positive selection (Bird et al., 2007; Bush and Lahn, 2008; Gittelman et al., 2015;

Pollard et al., 2006a; Prabhakar et al., 2006). Knee-specific elements significantly overlapped these regions (Fig. 2.1C, Supplemental Table S2.4), largely recapitulated using overlapping E59 human elements, despite smaller set sizes. This suggests that ancient positive selection acted on these elements. Mouse brain elements, mapped to their human orthologs, also enrich for overlaps, as reported for human brain elements (Pollard et al., 2006a-b), while enrichment was not seen for B-lymphocyte elements (Buenrostro et al., 2013). When examining *en masse* accelerated sequences in knee-specific elements using GREAT, we identified enrichments for “ossification”, “skeletal phenotypes”, and “limb developmental expression” (Supplemental Table S2.4). Two example loci are shown in Supplemental Figure S2.1.

We next considered sequence change between humans and chimps (Methods). Nucleotide similarity (%ID) positively tracked with cross-primate sequence conservation for each element set (Fig. 2.1D, Supplemental Table S2.2). To understand the biological significance of different levels of human-chimp %ID, we ran GREAT on conserved sequences (i.e., third quartile of %ID) from DF- and PT-specific sets (Supplemental Table S2), and found enrichments for “regulation of ossification”, “skeletal system development” and “chondrocyte differentiation”. Since divergence in morphology can result from sequence changes in existing regulatory switches (Cotney et al., 2013), we ran GREAT on first quartile %ID sequences, which tended to retain conservation across primates (Fig. 2.1D, left shaded area), and found enrichments for “mild short stature”, “negative regulation of ossification”, etc. Weaker/insignificant enrichments were seen in the top/bottom 10% of sequences, with reduced strength due to smaller set sizes and the decreased power of GREAT to identify enrichments (Supplemental Table S2.2).

Given that human-specific enhancer mutations can impact development (Boyd et al., 2015), we considered the potential for base-pair alterations (chimp-human substitutions) in knee elements to alter sequence motifs, as shown for brain enhancers (Zehra and Abbasi, 2018). Identified motifs in knee sets (Supplemental Table S2.3) were combined with those of known chondrogenesis-related TFs (Liu et al., 2017) and used to identify the biased occurrence of alterations predicted to modify certain motifs (Fig. 2.2A, Supplemental Table S2.3, Methods).

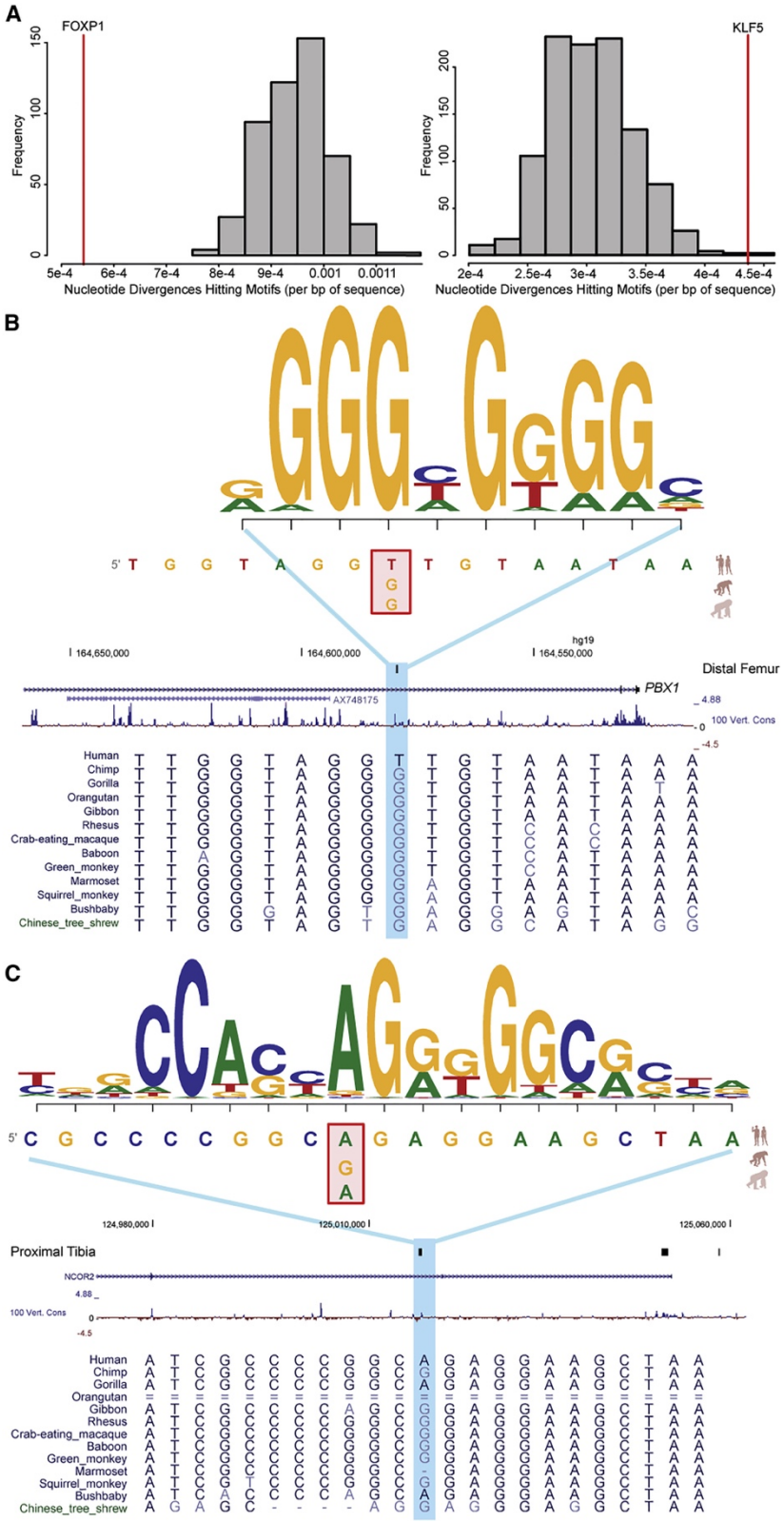

A
FOXP1
Frequency
Nucleotide Divergences Hitting Motifs (per bp of sequence)
KLF5
Frequency
Nucleotide Divergences Hitting Motifs (per bp of sequence)
B
5' T G G T A G G T T G T A A T A A
hg19
164,650,000
164,600,000
164,550,000
Distal Femur
PBX1
AX748175
4.88
0
100 Vert. Cons
-4.5
Human
Chimp
Gorilla
Orangutan
Gibbon
Rhesus
Crab-eating_macaque
Baboon
Green_monkey
Marmoset
Squirrel_monkey
Bushbaby
Chinese_tree_shrew
C
5' C G C C C C G G C A G A G G A A G C T A A
124,980,000
125,010,000
125,060,000
Proximal Tibia
NCOR2
4.88
100 Vert. Cons
0
-4.5
Human
Chimp
Gorilla
Orangutan
Gibbon
Rhesus
Crab-eating_macaque
Baboon
Green_monkey
Marmoset
Squirrel_monkey
Bushbaby
Chinese_tree_shrew

Figure 2.2: Sequence Modification of Putative Regulatory Elements

(A) Histograms showing depletion and enrichment of human-chimp base pair alterations intersecting predicted FOXP1 and KLF5 TF motifs, respectively, in DF-specific elements relative to randomized sets ($p < 0.01$). Counts of altered bases in background (gray) and element (red) sets shown as a fraction of total set size. See Table S3.

(B) KLF5 motif example in a DF-specific element intronic to PBX1, a TF regulating HOX expression and endochondral ossification (Capellini et al., 2006, Selleri et al., 2001). Predicted disruption by a human-derived T/G nucleotide change shown. A region of acceleration in a knee element identified upstream of PBX1 (Table S3). Red shading indicates altered base relative to motif logo, blue indicates genomic position of sequence. ATAC-seq regions and phyloPways conservation tracks, UCSC sequence alignment shown.

(C) Example of a human-chimp alteration predicted to improve a CTCF motif within a PT-specific element intronic to NCOR2, involved in skeletal biology (Blake et al., 2017). Format similar to (B).

Across knee-specific sets for several chondrocyte TF motifs there is a strong tendency for protection from modification (biased against possessing substitutions). For example, putative sequence motifs (i.e. sequences similar to a known TF motif) for FOXP1/FOXP2, factors important for skeletogenesis and knee morphology (Xu et al., 2018; Zhao et al., 2015), were biased against disruptive changes (Fig. 2.2A, left). Human FOXP1/FOXP2 ChIP-seq data also shows strong overlap with knee-specific elements (Supplemental Table S2.2). Conversely, we observed alterations biasing towards TF motifs (Supplemental Table S2.3). For example, predicted motifs for KLF5, involved in chondrogenesis (Shinoda et al., 2008), displayed such biases in the DF-specific set (Fig. 2.2A, right), which also significantly overlap ($p < 0.05$) KLF5 binding data (Supplementary Table S2.2, S2.3) (Fig. 2.2B for an example). We also observed in knee-specific sets a bias for alterations at predicted CTCF motifs, whose loss-of-function results in skeletal defects and spontaneous OA (Hijazi, 2018). Knee-specific sets also significantly overlap with CTCF binding data (Supplemental Table S2.2) (Fig. 2.2C for an example).

2.2.3: Examination of Intra-Species Sequence Diversity in Element Sets

As selection also shapes within-species sequence variation (Vitti et al., 2013), we acquired variant data for humans (Gibbs et al., 2015), chimps and gorillas (Prado-Martinez et al., 2013), examining whether mouse and overlapping E59 human element sets differ in patterns of sequence diversity within and between species. Considering human variation we found that sequence diversity in PT- and DF-specific elements is constrained compared to genomic backgrounds, as well as promoter-TSS and intronic elements (Fig. 2.3A, Supplemental Table S2.5, Methods). Comparing knee-sets directly we observed that DF- and PT-specific elements have significantly reduced diversity compared to KC-specific elements, with more marked constraint in PT relative to DF (Fig. 2.3B). These diversity patterns between specific knee element sets were not observed for orthologous forelimb sets, nor unfiltered 'general' knee sets (Fig. 2.3B, Supplemental Tables S2.5, S2.6, Supplemental Fig. S2.2).

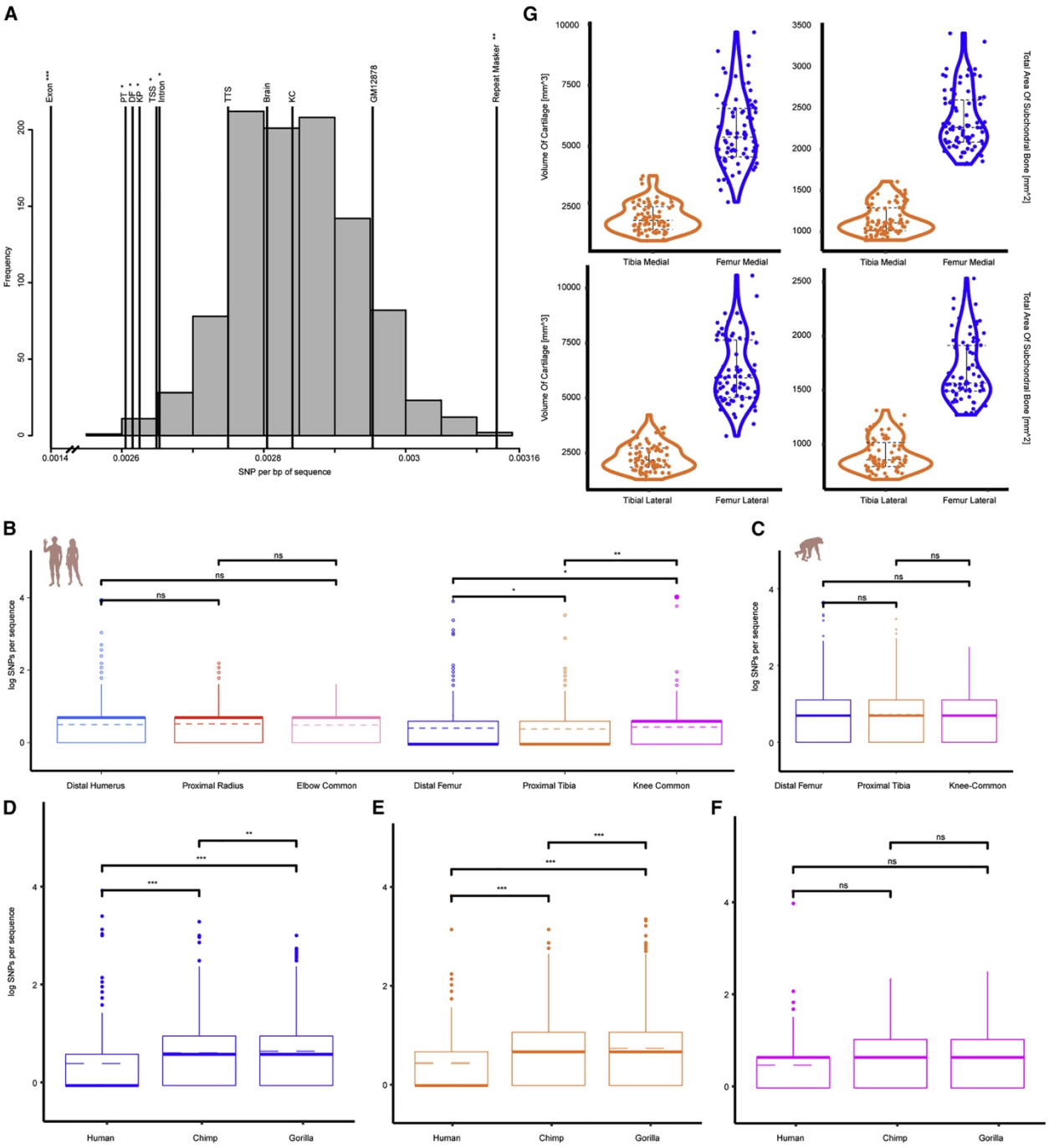


Figure 2.3. Intra-species Variation in Regulatory Elements and Human Variation in Knee Morphology

(A) Counts of common human variants per bp of sequence for element sets compared to random region sets along with other genomic features; labels correspond to Supplemental Table S2.5. (B) Common human variants intersecting elements in the elbow (left) and knee (right) specific sets were counted and compared across sets. (C) Comparisons of chimp sequence diversity across knee-specific sets. (D–F)

Common variants in human, chimp, and gorilla intersecting a given ATAC-seq peak counted for all variable sequences in knee-specific sets, expressed as “SNPs per sequence”—mean/median values shown in dashed/bold lines, respectively. (D) Distal femur-specific (DF). (E) Proximal tibia-specific (PT). (F) Knee-common-specific (KC). (G) Measurements of human medial/lateral tibia and femur via MRI dataset. Volume of cartilage (left) and total area of subchondral bone (right) for medial/lateral portions of both bones were measured across all subjects. Individual points are plotted alongside an outlined density curve, quartiles indicated in dashed lines. Significance codes: not significant (ns), < 0.05 (∗), < 0.01 (∗∗), < 1e−5 (∗∗∗).

Corresponding sets in chimps and gorillas were examined for similar patterns of diversity between DF-, PT- and KC-specific elements (Fig. 2.3C). No significant differences were observed, similarly true for the forelimb and ‘general knee’ sets (Supplemental Table S2.5). Comparing patterns of sequence diversity between humans and African apes, we observed consistently reduced variation within DF- and PT-specific elements in humans compared to corresponding chimp and gorilla elements (Fig. 2.3D-E). We note that no significant differences in diversity levels in KC-specific elements between species were observed (Fig. 2.3F, Supplemental Table S2.5). These patterns were recapitulated in overlapping E59 human elements (Supplemental Fig. S2.2A-C, J-L, Supplemental Tables S5-S6).

However, the human pattern we observed (Fig 2.3B-F) could reflect a more general increase in genetic diversity in apes, notably chimps (Prado-Martinez et al., 2013). We therefore performed an analysis of sequence set constraint by generating chimp genomic backgrounds and functional annotations and comparing diversity with knee-element sets. We found elevated background diversity across the chimp genome, indicated by the left skew in the background distribution. Regardless of skew, knee-specific sets in chimp behave markedly different; specifically, they each exhibit less constraint than promoter-TSS and intronic elements (Fig. 2.3A, Supplemental Fig. S2.2P, Supplemental Table S2.5). Overall, these data reveal a unique human pattern of constraint among DF- and PT-specific elements (i.e., reduction of genetic diversity relative to KC elements and functional annotations, not in African apes), suggesting earlier results were capturing, at least in part, these signals.

2.2.4: Examination of Human Knee Morphological Diversity

Given these patterns in sequence constraint, we sought a corresponding phenotypic signal in human knee morphology (Methods). Using morphometric data from OABCP MRI images we confirmed reduced variation in all proximal tibia measurements compared to matched distal femur features (Fig. 2.3G, Supplemental Table S2.5). Intersecting OABCP sequence data with knee elements we found significantly less variation in PT-specific compared to DT-specific elements (Supplemental Table S2.5), supporting a genotype-phenotype link between reduced knee regulatory sequence variation genome-wide and reductions in morphological diversity.

2.2.5: Examination of the Biological Impacts of Human Sequence Variation in Element Sets

We next examined how genetic diversity in knee elements generally impacts phenotypic and regulatory variation within species. Human and chimp sequences in knee sets partitioned by variation were examined using GREAT (Methods). For the least variable (first quartile) human and chimp elements, similar enrichments were observed for terms such as “cartilage development” and “chondrocyte differentiation.” Conversely, the most variable (third quartile) sequences showed different enrichments for humans and chimps. The most variable human elements yielded enrichments for “collagen catabolism/metabolism”, “anchoring collagen”, and “osteoarthritis”, functional terms not identified with variable chimp elements (Supplemental Table S2.5).

The initial findings in variable human sequences lead us to examine the relationship between common variation in knee sets and OA loci identified via GWAS. We aggregated 95 lead variants across 83 OA GWAS loci (Supplemental Table S2.7) and found that such variants were enriched in knee-specific element sets, including overlapping E59 human elements (Fig. 2.4A, Supplemental Table S2.7). We saw no enrichment of variants in less-specific human E47 limb datasets, BMDCs, nor for the brain. These findings suggest a specific link between human variation in knee chondrocyte regulatory sequences and OA risk.

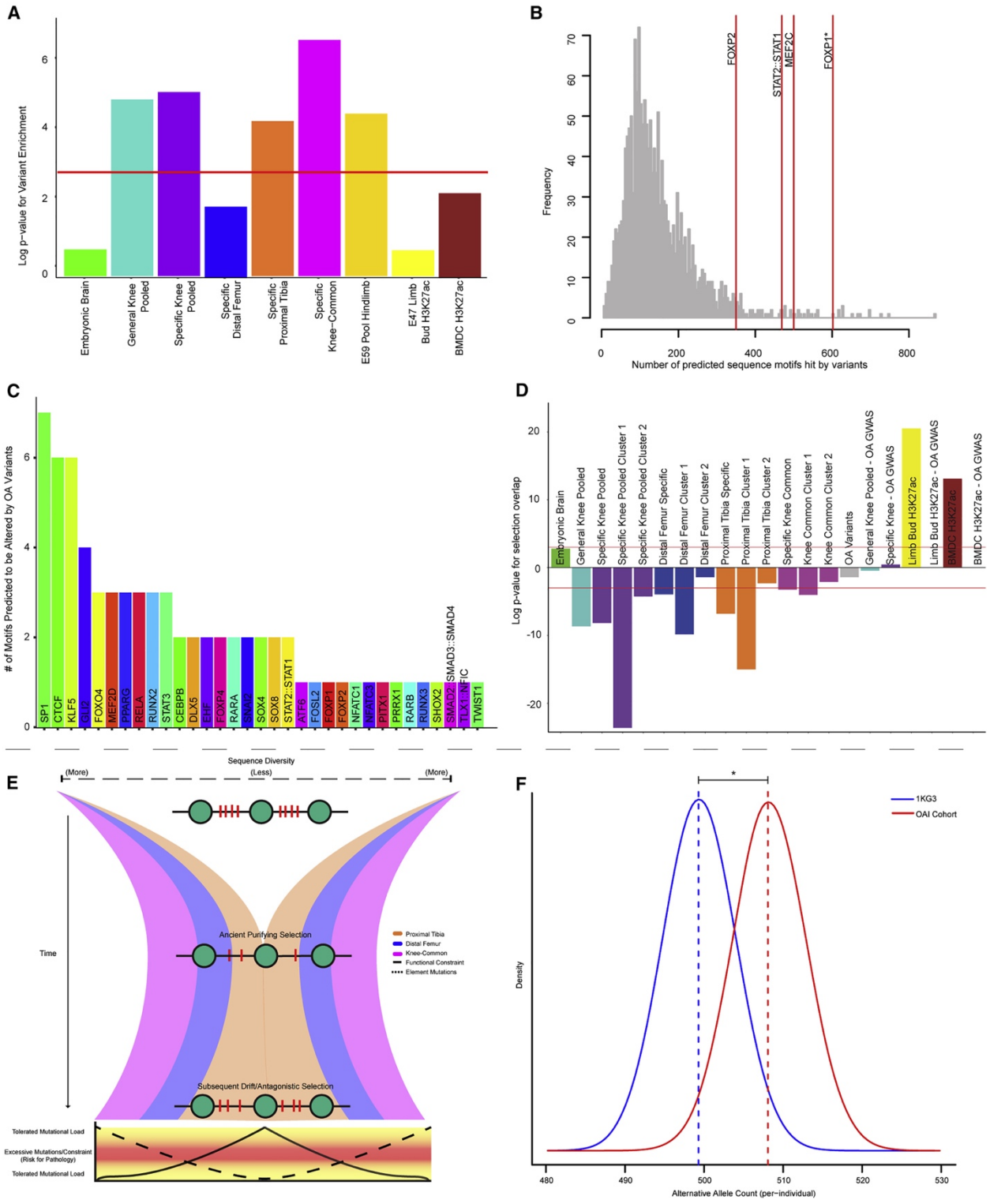


Figure 2.4: Human Variation in Knee Elements and Its Impacts on OA Risk

(A) Enrichments for OA risk variants in general knee, knee-specific, human E59 hind limb, and brain region sets, along with human E47 and BMDC H3K27ac ChIP-seq regions. Calculated Z score

enrichments over randomized sets shown as -log (p value); red line indicates significance threshold ($p < 0.05$). (B) Distribution of predicted motifs intersected by OA risk variants, counted per TF. Significantly enriched factors are indicated—FOXP1 significant following p value correction ($p < 0.05$). (C) Distribution of predicted motifs intersecting OA risk variants in knee elements for a set of chondrogenesis-related TFs. (D) Overlaps of region sets and signals of recent selection calculated for general knee, knee-specific, brain, and human H3K27ac ChIP-seq region sets, along with OA risk-variants and the regions intersecting them. Clustered set overlaps also shown. Hypergeometric tests represented as -log (p value), with sign denoting enrichment/depletion; red lines indicate respective significance cut-offs ($p < 0.05$). (E) Top: formalized model for the role of evolutionary history in modern heritable OA risk. Ancient selection acting on ancestral sequence diversity in regulatory elements establishes a derived knee configuration, which is subsequently maintained through ancient purifying selection (i.e., functional constraint). More recently, genetic drift, in combination with antagonistic selection for other traits, increases the frequency of alternative alleles in functionally constrained sites. Bottom: the presence of moderate mutational load in highly constrained elements (e.g., prox. tibia elements), or high mutational load in less-constrained elements (e.g., dist. femur elements), stand to disrupt knee homeostasis and promote pathology risk, while low mutational loads (or low sequence constraint) are more tolerated (i.e., harbor lesser risk of pathology). (F) The number of alternative alleles falling in chondrocyte regulatory elements counted per-individual for the 1KG3 population (blue) and the OAI patient cohort (red), shown as density distributions with mean values (dashed lines); significance bar indicates Student's t test result ($p < 0.05$).

As GWAS variants are thought to alter gene expression via TF modifications (Zhang and Lupski, 2015), we examined whether OA variants within knee elements modify sequence motifs (Supplemental Table S2.7). Although we had observed that human-chimp substitutions within knee-specific elements bias against altering FOXP1/FOXP2 motifs, OA variants bias towards altering them (adj. p-value < 0.05, Fig. 2.4B, Supplemental Table S2.7). We also observed these OA variants intersecting motifs for chondrogenesis-associated TFs (Figure 2.4C, Supplemental Table S2.7); in particular, pooled knee-specific and PT-specific elements have variants tending to overlap predicted CTCF and KLF5 motifs (adj.

p-value < 0.05). Furthermore, knee-specific elements capture a significant proportion of all OA GWAS variants intersecting these motifs (> 3-fold enrichment, adj. p-value < 0.05, Supplemental Table S2.7). We examined whether common variants in knee element sets (i.e., all SNPs with MAF >= 0.05) frequently modify chondrogenesis-related TF motifs but found a lack of significant bias (Supplemental Table S2.3), pointing to the specific enrichment of signals for OA GWAS variants. Nevertheless, at some functionally determined OA-annotated loci (i.e., found via GREAT) common variants were predicted to alter relevant motifs within knee-specific elements. Two examples are shown in Supplemental Fig. S2.3A-B.

Given the highly polygenic nature of OA (Hunter and Bierma-Zeinstra, 2019), smaller effect-size variants in knee elements may also cumulatively contribute to alter chondrocyte regulation, and ultimately the heritability of risk. We assessed the contributions that variants in knee elements, chondrocyte elements generally, and other genomic features have to OA risk heritability. Using LDSC (Finucane et al., 2015), we found that orthologous chondrocyte elements pooled from all long-bones, along with pooled human E59 hindlimb elements, captured variants explaining a significant proportion of OA heritability (adj. p < 0.05, Supplemental Table S2.7), a feature not seen using embryonic brain or B-lymphocyte elements. Variants captured by more refined knee-specific element sets did not reach LDSC significance, owing to limits in ascertaining significant partitioned heritability from very small region sets (Finucane et al., 2015). Interestingly, annotations pertaining to sequence conservation, i.e., GERP score (Davydov et al., 2010), primate phastCons (Hubisz et al., 2011) and predicted allele age (Rasmussen et al., 2014) also associated with OA heritability.

2.2.6: Examination of Recent Evolutionary Forces

The enrichments of OA GWAS variants in knee elements, and the heritability patterns seen in chondrocyte elements and conserved sequences reflect in part the effects of constraint on knee regulatory element function. But, how have more recent genetic drift and positive selection shaped human variant frequencies within knee-specific elements, and influenced the observed disease associations? We considered the behavior of variants within knee elements, investigating pairwise Fst (Weir and Cockerham, 1984) across 18 populations (Methods). PCA was used to visualize the separation of knee

variants based on shared behaviors in population stratification (Supplemental Fig. S2.3C, Supplemental Table S2.8). The first two components represented population differences between Eurasia/Africa (PC1) and Europe/Asia (PC2) populations (Supplemental Fig. S2.3D), capturing the majority (> 60%) of observed variation. K-means clustering of variants based on Fst resulted in two groups per knee set (Supplemental Table S2.8, Methods): Cluster 1 variants have a narrower Fst range across populations (Fig. S3E-F), while cluster 2 trends towards increased Eurasia/Africa divergence. Using GREAT, we determined whether clusters have divergent functional enrichments (Supplemental Table S2.8). Cluster 2 variants were functionally associated with “skeletal development” and related chondrocyte terms, while cluster 1 variants tended to occur near genes with annotations such as ‘arthritis’, ‘arthralgia’, ‘osteoarthritis’ and ‘knee joint’.

We next assessed whether recent positive selection shaped the patterns of clustered variants, and OA variants in general (Methods). We found that only 18% of regions containing clustered knee elements (cluster 1) are known to have undergone positive selection in the last 30,000 years (Jagoda et al., 2017; Pagani et al., 2016), a significant depletion (1.25 fold-decrease, adjusted p-value: $5.26e^{-11}$; Supplemental Table S2.8; Fig. 2.4D). No cluster 2 set was enriched in selection windows; rather, they trended towards depletion (Fig. 2.4D). We then examined loci for which OA GWAS variants fell in both knee elements and selection windows, identifying four (Supplemental Tables S2.7, S2.8). For three loci (*UNC5C-BMPR1B*, *ENPP1/3*, *LSM5*) the putative OA risk variant occurred on the non-selected, recombined haplotype (Supplemental Fig. S2.3G-J). However, at *GDF5,* OA risk variants were found to reside on the positively selected haplotype (Capellini et al., 2017; Miyamoto et al., 2007). OA GWAS variants, in general, were not enriched in selection regions (Fig. 2.4D), and by one test (Grossman et al., 2013) were strongly depleted (Supplemental Table S2.8). Lastly, TF motif analyses on clustered low- and high-Fst variants did not see show strong biases for intersections, indicating a consistent lack of directionality in regulatory modification more in line with genetic drift effects (Supplemental Table S2.3).

2.2.7: Formal Model of Chondrocyte Knee Developmental Regulation, Evolution, and OA Risk

Our analyses revealed that DF- and PT-specific regulatory elements exhibit evidence in humans of ancient positive selection (Fig. 2.1C), followed by more recent sequence constraint not observed in African-apes (Fig. 2.3A-E). Furthermore, the allele frequencies of human common variants within these elements, and OA risk variants in general, appear to have been shaped by genetic drift on the background of constraint (Supplemental Tables S2.7, S2.8), rather than recent positive selection (Fig. 2.4D). These variants have consistent links to OA (Supplemental Table S2.8), likely by impacting TF motifs under constraint (Fig. 2.4B-C), and contribute to overall heritability of risk (Supplemental Table S2.7). We propose a model in which violations to constraint in functional conserved sequences, tolerated during knee development, have pathological consequences later in life (Fig. 2.4E; Discussion). Two evolutionary mechanisms may cause violation to constraint thereby increasing genetic risk for OA: genetic drift and antagonistic selection. We now provide additional genetic and functional evidence addressing how both processes may drive heightened OA risk genome-wide as well as on a locus-specific level.

2.2.8: Functional Sequence Variation Within At-Risk Individuals

Since it is anticipated that genetic drift should increase the frequency of violations underlying OA risk, patients should have higher loads of violating mutations than the general population. To test this, we took the OAI patient cohort, consisting of individuals suffering from, or identified as being at high risk of developing OA, in comparison to the 1000Genomes population (Methods). We considered sequence variants falling in knee-specific elements and asked whether, across the set of variable sites, patients tended to possess more alternative alleles. Indeed, we observed a significant increase in the average number of alternative alleles possessed by subsets of the OAI cohort (Fig. 2.4F), results replicated using an additional population control (Pagani et al., 2016) (Supplemental Table S2.7). To confirm that it is specifically sequence violation within constrained knee regulatory elements that is related to the heightened OA risk in this cohort, we compared the number of alternative alleles falling within B-lymphocyte elements as a control, observing different variant behaviors (Supplemental Table S2.7).

2.2.9: Functional Interrogation of an OA Risk Locus in Humans and Mice

Another mechanism that may lead to violations of constraint is antagonistic selection or pleiotropy, in which positive selection increases a beneficial allele along with linked deleterious alleles. We refer to pleiotropy here in the context of linked genetic variants, rather than the multiple possible different effects of single variants (i.e. horizontal pleiotropy) Earlier, we uncovered one such locus, *Growth Differentiation Factor Five* (*GDF5)*. GDF5 is a BMP with quintessential roles in knee development across mammals (Basit et al., 2008; Rountree et al., 2004; Settle et al., 2003), and yet is the most reproducibly associated OA locus to date (Miyamoto et al., 2007; Zengini et al., 2018). At *GDF5*, selection on reduced height via a regulatory variant (rs4911178) in *GROW1*, an enhancer active in femoral and tibial growth plates, likely increased frequencies of linked OA risk variants (Capellini et al., 2017). As our model predicts high-frequency sequence violations contribute to OA risk, we explored this exemplar locus in-depth.

We first intersected *GDF5* GWAS OA risk variants (Zengini et al., 2018) with general E15.5 mouse/E59 human knee elements, and found putatively-causal variants, rs4911178 (G/A) in *GROW1* and rs6060369 (C/T) (Fig. 2.5A). Rs4911178 and *GROW1* do not impact knee morphology when deleted in mice (Capellini et al., 2017), implicating rs6060369 as causative. Rs6060369 overlaps a knee element, *R4*, located in sequence downstream of *GDF5* shown in rescue experiments to regulate expression and mediate knee morphology of *Gdf5* null (*bp*) mice (Chen et al., 2016; Pregizer et al., 2018). To characterize *R4* activity, we made stably-expressing *R4 lacZ* reporter mice and found at E14.5/15 expression in the early knee, including the condylar regions and notch but not the trochlea, and postnatally, expression restricted to femoral condyle and tibia plateau articular chondrocytes and cruciate ligaments (Fig. 2.5B). Compiling human functional genomics data on E47 limbs, E59 distal femur chondrocytes, and BMDCs, we found *R4* activity at each timepoint (Fig. 2.5A), indicating a similar regulatory time-course as in mouse.

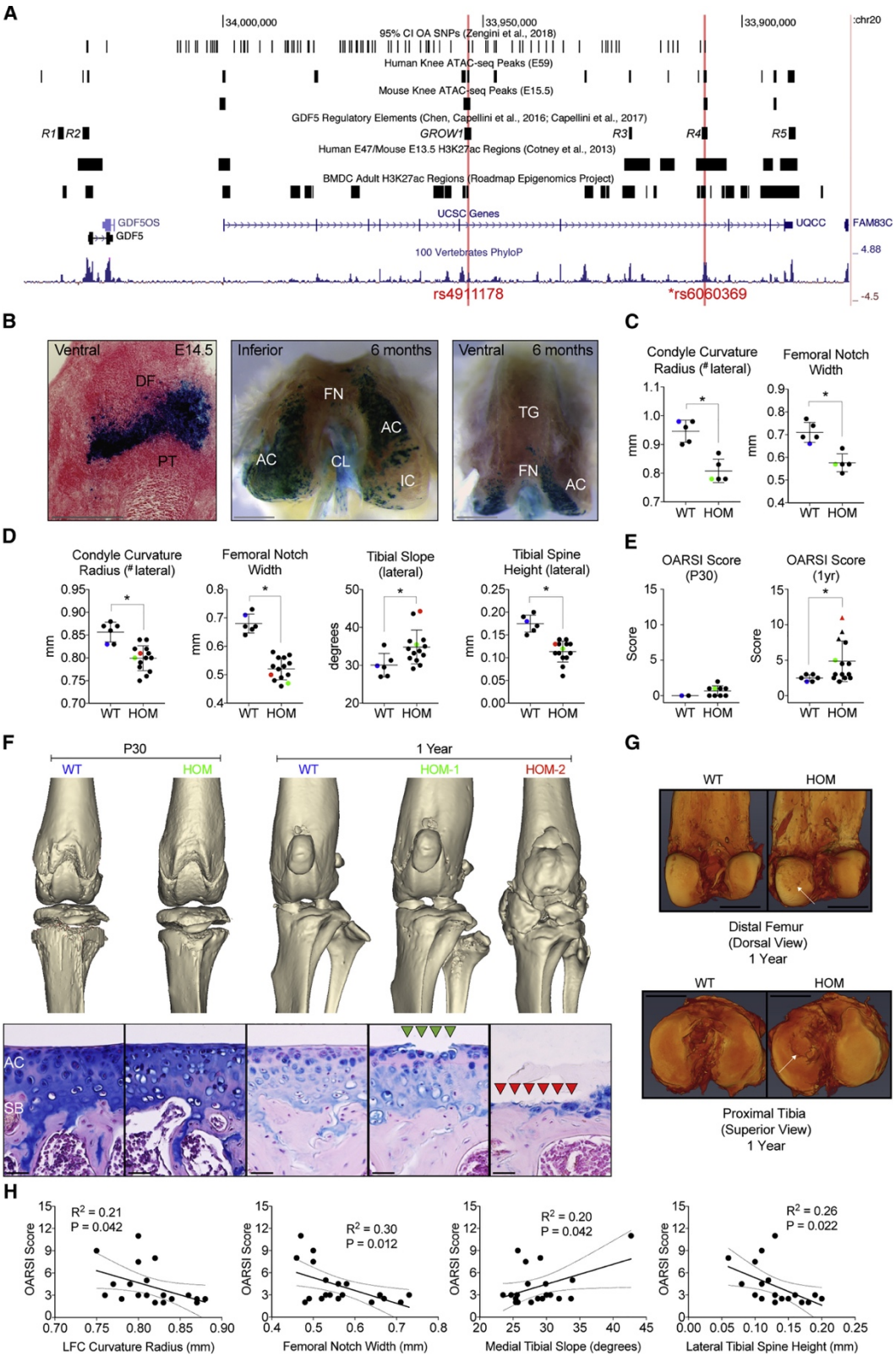


Figure 2.5: Functional Characterization of the *GDF5* Locus and *R4* Enhancer in the Mouse

(A) UCSC Genome Browser view of human GDF5 locus with intersections of OA variants, knee ATAC-seq regions (for human and mouse tissue), GDF5 enhancers, and H3K27ac ChIP-seq signals from human embryonic limbs (E47) and BMDCs. Rs4911178 (left red line) and rs6060369 (right red line) overlap functional knee sequences. (B) R4-driven lacZ expression in inferior-most distal femur (DF) and superior-most proximal tibia (PT) (left) tissues and in adult distal femora (right two images). AC, articular cartilage; FN, femoral notch; IC, inferior condyles; CL, cruciate ligaments; TG, trochlear groove. Scale bars, E14.5 = 250 μm; adult = 1 mm. (C) P30 knee anatomy in C57BL/6J R4 null mice. #, same trends for medial condyle. R4+/+ (wild-type [WT]), R4−/− homozygous (HOM). (D) 1-year knee anatomy. (E) OARSI scores on WT and HOM knees at P30 and 1 year. Triangles, heterotopic ossification. (F) 3D renditions (top) and histology (bottom) of WT/HOM knees with minimum (blue), mean (green), and maximum (red) OARSI scores. Heterotopic ossification observed in HOM knees with highest OARSI scores and most cartilage damage. Scale bars, 50 μm. (G) X-ray 3D cartilage scanning of WT/HOM distal femur condyles (top) and proximal tibia platforms (bottom) at 1 year, showing OA lesions (white arrows). Scale bars, 1 mm. (H) Linear regressions between knee morphology and OARSI score at 1 year. In (C)–(E) colored dots correspond to specimens shown in (F). Significance code: <0.05 (∗).

ASE on *R4*-deletion mice found that $R4^{+/-}$ significantly down-regulated *Gdf5* expression in E15.5 distal femur chondrocytes (deletion-allele *Gdf5* expression 69.6 ± 0.05% of control expression; N = 4/genotype; p = 0.001). At P30, *R4* loss ($R4^{-/-}$) led to significant distal femur changes, specifically, smaller femoral condyle curvature radius and notch sizes, among other changes (Fig. 2.5C; Fig. 2.7; Supplemental Fig. S2.4A-C, Supplemental Table S2.9). $R4^{+/-}$ mice displayed a non-significant trend in the same direction (Supplemental Table S2.9). Despite changes, a normal articular cartilage and joint cavity formed in $R4^{-/-}$ mice (Fig. 2.5F, Left). Given that $R4^{+/-}$ mice lacked overt phenotypes at P30, our 1 year comparisons focused on $R4^{+/+}$ and $R4^{-/-}$ mice. At 1 year, $R4^{-/-}$ mice exhibited exacerbated defects in the same distal femur features (Fig. 2.5D, $p < 0.05$) and new alterations to the proximal tibia (Supplemental Table S2.9). The observed changes are related to genetic variants and not aging as morphological comparisons occurred between genotypes at each time point separately. Strikingly, OA developed in $R4^{-/-}$ mice as observed via histology, OARSI score (Mann-Whitney test, p = 0.041), and cartilage imaging (Fig. 2.5E-G,

Supplemental Fig. S2.5). This phenotype included a loss of glycosaminoglycans throughout articular cartilage surfaces with minor fibrillations of isolated surfaces in most individuals, and severe damage up to complete denudation of isolated joint surfaces in some individuals (Fig. 2.5F). Using X-ray 3-D scanning we observed OA lesions on the femoral condyle and proximal tibia (Fig. 2.5G). In the most effected mice, heterotopic ossification was observed, a phenotype that occurs in late stage human OA (Fig. 2.5F). Finally, several morphologic features of *R4*$^{-/-}$ knees (e.g., condyle curvature, notch size, etc.) were correlated with OA severity; features which have been shown to change during human OA progression (Barr et al., 2016; Hunter et al., 2016; Neogi et al., 2013)(Fig. 2.5H, Supplemental Fig. S2.4D). However, as these measurement changes were also detected in *R4*$^{-/-}$ mice at P30, such shape alterations preceded and may have caused OA.

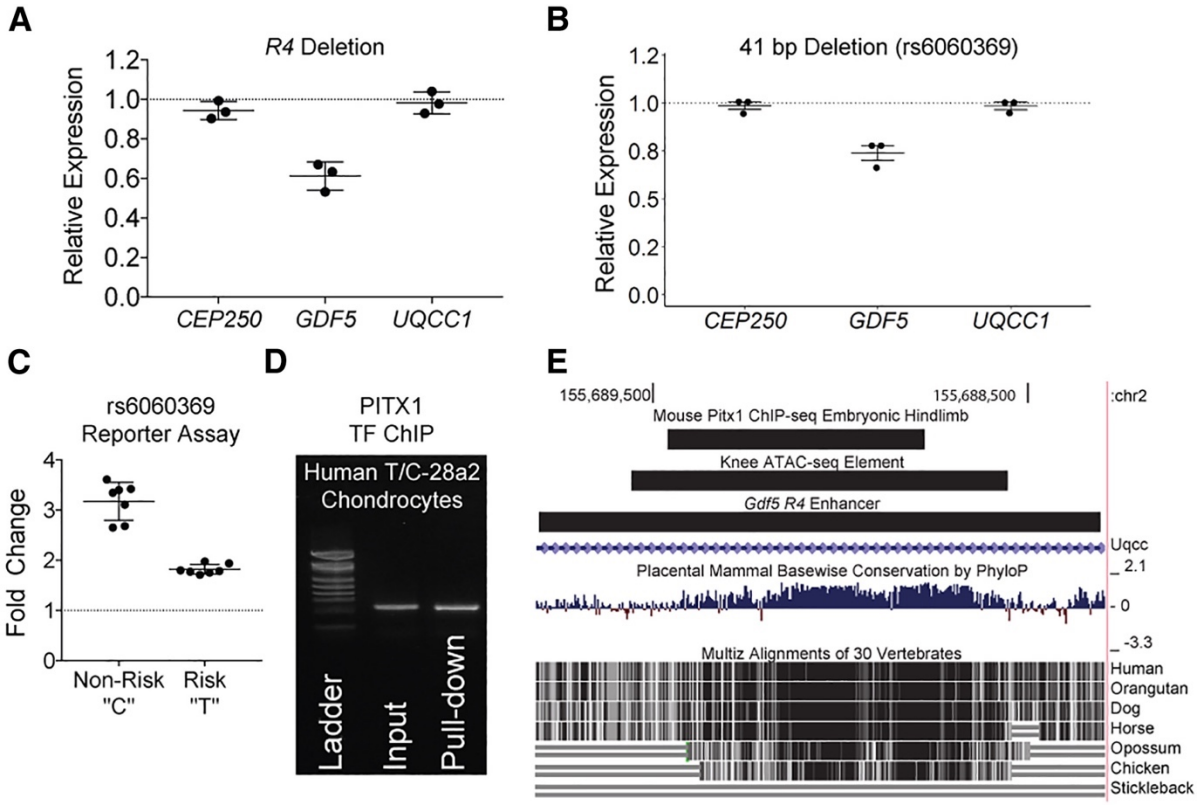


Figure 2.6. Functional Characterization of the R4 Enhancer in Human and Mouse Chondrocytes

(A) Expression by qRT-PCR of CEP250 (not significant [n.s.]), GDF5 ($p < 0.005$), and UQCC1 (n.s.) in human T/C-28a2 chondrocytes lacking R4. (B) Expression of CEP250 (n.s.), GDF5 ($p < 0.05$), and UQCC1 (n.s.) in cells lacking 41 bp containing rs6060369 in R4. (C) In vitro reporter analyses of R4-

driven luciferase activity comparing constructs with the OA risk “T” or non-risk “C” variant in T/C-28a2 cells. (D) PITX1 ChIP on a R4 sub-region containing rs6060369 in T/C-28a2 cells showing input (left) and pull-down (right) in image. (E) UCSC Genome Browser view of mouse R4 corresponding to knee ATAC-seq region and PITX1 binding via ChIP-seq (Infante et al., 2013).

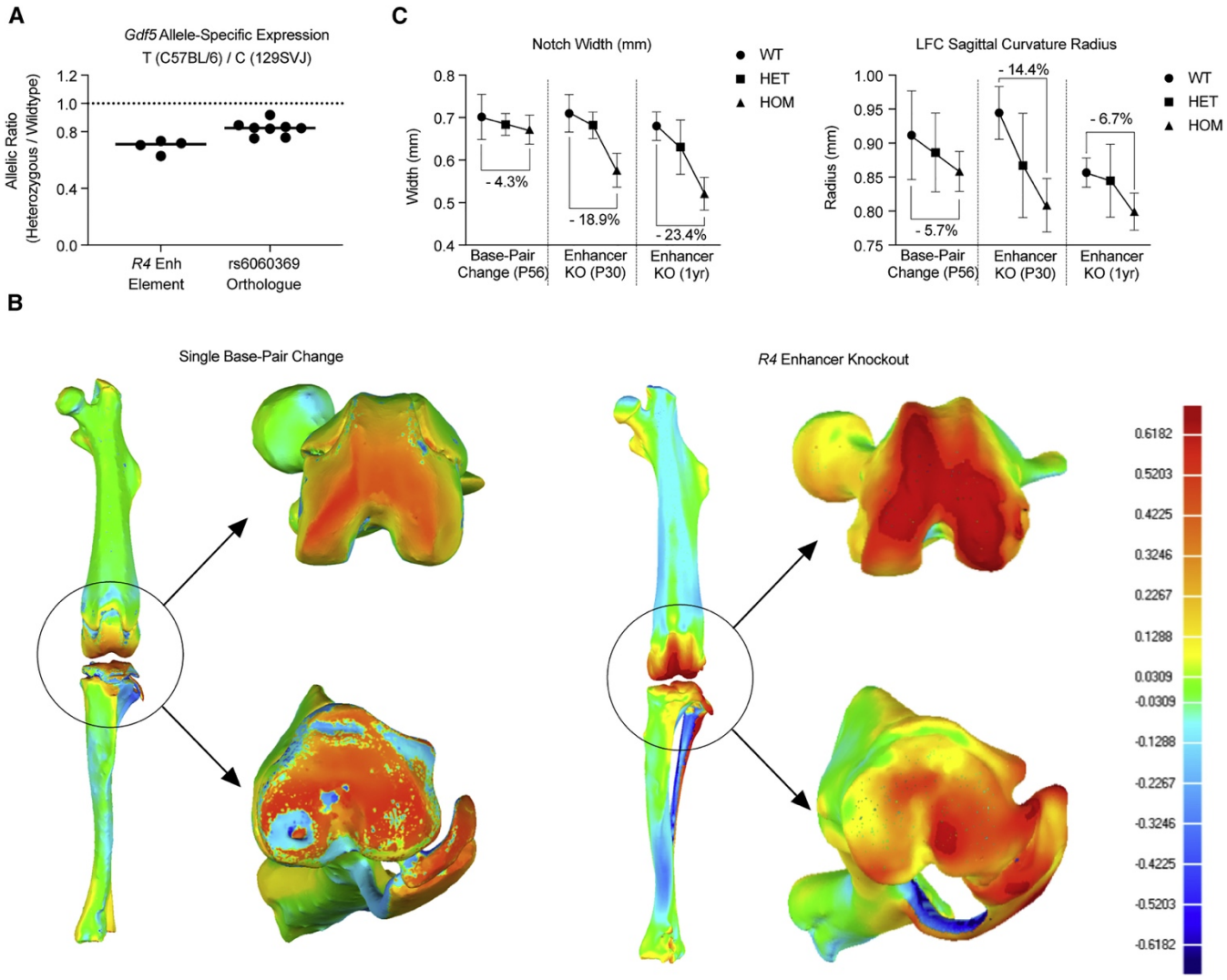


Figure 2.7. Functional Characterization of the rs6060369 OA Risk Allele in Humanized Replacement Mice (A) ASE assays on E14.5 distal femur chondrocytes from C57BL/6J/129SVJ R4rs6060369-T, rs6060369-A replacement mice (right, p = 0.0005) and C57BL/6J/129SVJ R4 heterozygous (R4+/−) mice (left, p = 0.001). (B) 3D comparative analysis indicating the locations of largest morphological differences between (left) WT R4rs6060369-A, rs6060369-A and HOM R4rs6060369-T, rs6060369-T hind limbs, as well as between (right) R4+/+ and R4−/− hind limbs (zoom-in images focus on inferior distal femur [top], superior

proximal tibia [bottom]). Areas with largest variations are highlighted in red (WT > HOM) and dark blue (WT < HOM), with minimal variation in green/yellow. (C) μCT measurements of indicated features in base pair replacement and R4 mice at postnatal days P30, P56, and 1 year.

We next explored *R4* function and its OA variant (rs6060369) in human and mouse chondrocytes *in vitro*. *R4* deletion in human T/C-28a2 articular chondrocytes caused significant down-regulation of *GDF5* expression ($p<0.005$), but not nearby genes (Fig. 2.6A, Methods). Deletion of a 41 bp subregion containing rs6060369 had a similar effect ($p < 0.05$, Fig. 2.6B). Luciferase reporter assays in these cells for rs6060369 found that the risk allele "T" drove significantly reduced expression (Fig. 2.6C, $p=0.000044$; Fisher-combined p=8.79E-09). We computationally predicted that Pituitary homeobox 1 (PITX1), a major TF in knee formation (Nemec et al., 2017) and OA risk factor (Butterfield et al., 2019; Picard et al., 2007), binds to this variant position. Using ChIP on T/C-28a2 cells (Fig. 2.6D), E15.5 distal femur and proximal tibia cartilage, and PITX1 ChIP-seq data (Fig. 2.6E), we validated PITX1 binding at this position, indicating that it likely mediates the *cis*-acting effects of rs6060369 in the *R4* element on *GDF5* expression.

We finally edited the rs6060369 orthologous position in C57BL/6J mice to contain the human risk "T" allele. ASE on mice with one humanized *R4* "T" allele significantly down-regulated *Gdf5* expression in E15.5 distal femur chondrocytes (replacement "T" allele *Gdf5* expression 82.3% +/-0.05% of control expression; N = 8/genotype; p=0.0005, Fig. 2.7A), revealing an endophenotype within an anatomically relevant context. Phenotyping P56 humanized mice found that several distal femur features, also dysmorphic in $R4^{-/-}$ mice (including notch width and lateral condyle sagittal curvature radius), display alterations in $R4^{\text{rs6060369-T, rs6060369-T}}$ compared to controls, and in the same direction of effect (Fig. 2.7B-C; Supplemental Table S2.9). While adjusted comparisons did not yield many significant differences, matched unadjusted analyses demonstrate marked differences in condyle curvature and width, and tibial spine size (Supplemental Table S2.9). These precise enhancer-mediated alterations are evident in heatmaps of morphological change between wildtype and $R4^{\text{rs6060369-T, rs6060369-T}}$ hindlimbs (Fig. 2.7B). As a subset of effected measures are correlated with OARSI score in $R4^{-/-}$ mice (Fig. 2.5H), and as knee

morphology and OA risk are complex polygenic phenotypes, these findings constitute evidence that the derived “T” variant at rs6060369 is a causal variant. Moreover, its location in a conserved TF binding site within a functional element that modifies knee development supports our model’s expectations that violations to constraint can confer elevated OA risk.

2.3: Discussion

2.3.1: Insights on the Evolutionary Developmental Regulation of the Human Knee

To examine the developmental genetic changes involved in the evolutionary transformation of the human knee, we profiled open chromatin regions from both human and mouse developmental samples, comparing patterns of inter- and intra-species sequence evolution across a variety of hindlimb and forelimb elements. Open chromatin regions specific to distal femur and proximal tibia, which reflect site-specific regulatory elements involved in knee chondrogenesis, exhibited several features reflecting the effects of ancient selection during primate and hominin evolution. These include reduced sequence conservation among primates relative to pleiotropic elements, enrichments for human accelerated regions, and significantly reduced diversity relative to pleiotropic elements and certain genome features (e.g. TSS and promoter elements) within humans but not chimps. These findings augment our limited understanding of prenatal human knee development, pointing to a series of adaptive regulatory modifications to human knee formation.

Furthermore, we observed that PT elements display a stronger reduction in genetic diversity than DF suggestive of greater purifying selection acting on this part of the knee to maintain its anatomical integrity. Importantly, we observed corresponding reductions in morphological variation in features of the proximal tibia relative to distal femur in the OAI dataset. As we were limited to studying patients with KL = 1 (i.e., the lowest score depicting initial OA onset) we cannot rule out confounding issues regarding the time-course of OA on different anatomical knee subdomains, though there is currently no consensus as to whether one particular bone end is the primary site of OA and morphological change. These findings on PT constraint in human genetic and morphological data, together with reduced PT sequence conservation across primates, suggest a more evolutionarily-labile feature for locomotor adaptations which, when a

more optimal configuration is achieved, becomes ‘locked-in’ and is subject to negative selection pressures to maintain its morphology.

We were unable to examine gained human-specific open chromatin regions, which could show additional evidence of regulatory evolution in the human lineage. Similarly, the assumption of conserved chimp regulatory activity for overlapping elements present in mouse and human may not hold in all cases. Resolving these questions, however, would require chimp developmental samples, which is infeasible and unethical. As the majority of regions identified here were shared across chondrocyte growth zones and hence were uninformative on specific anatomical evolution and patterns, use of chondrocytes derived from chimp iPS cells, while providing a general “chondrocyte” signal, would not inform on anatomically-specific open chromatin regions.

Nevertheless, given the expectation that changes to human knee development are mediated at the regulatory level, accelerated regions and more targeted nucleotide changes within existing knee-specific elements (i.e. isolated substitutions modifying TF binding) likely participated in altering conserved regulatory networks. For example, biased occurrence of human-chimp changes intersecting motifs of chondrogenesis-associated factors, such as KLF5, CTCF, and FOXP1/FOXP2, suggest that some network components were modified (KLF5 and CTCF) whilst ancient functional constraints precluded the exploitation of others (FOXP1/FOXP2). Once modified, these regulatory elements fell under new functional constraints to maintain the derived human knee, manifesting in reduced genetic diversity within modern humans at these sites relative to orthologous ape sequences, as well as to pleiotropically-acting knee elements.

2.3.2: Evolutionary Insights on the Genetic Risk of OA and Causal Variant Discovery

Despite evidence of ancient selection, human variants falling within knee elements show little evidence of recent positive selection (i.e., within the past 30,000 years), display weak population-level differentiation, exhibit no biases in their intersection of predicted TF motifs, and are generally enriched near genes that impact OA. These point to the predominant effects of genetic drift on functionally-constrained regulatory

elements in joint disease. To this end, we found that when regulatory constraint is violated, e.g., through the emergence of a variant disrupting a TF motif, there is greater likelihood for pathogenic consequences. Knee elements more variable in humans, but not chimps, are linked with OA gene annotations, knee-specific elements are enriched for OA GWAS variants, and notably, while predicted FOXP1/FOXP2 motifs are protected against human-chimp substitutions, they were also those likely to be disrupted by OA risk variants. We find it interesting that conditional excision of *FoxP2* using *Prx1-cre* mice leads to knee defects, changes in articular cartilage thickness and OA (Xu et al., 2018). These observations suggest that numerous genetic perturbations to constrained FOXP binding sites might also underlie OA risk in humans. Importantly, we also observed impacts at TGF-□ and BMP loci, which not only exhibit signals of adaptive evolution but are involved in knee development (Lyons and Rosen, 2019), OA (Wang et al., 2014; Wu et al., 2016) and possess GWAS variants overlapping knee-specific elements (e.g., *BMPR1B* and *GDF5*).

We formulate a model of human knee evolution that impacts OA risk at the regulatory level (Figure 2.4E). Normally, it would be expected that variants violating constraint in particular regulatory sequences - i.e. those integral to knee development, structure and/or maintenance - would be selected against due to the fitness consequences of an inefficient bipedal gait. However, recently these violations may have become tolerated due to buffers (e.g., improved health, medical care, footwear, etc.). If so, their selective removal on a genome-wide level would become less efficient (Carnes and Olshansky, 1993), permitting retention and accumulation. Additionally, prehistorically-tolerated variants in these elements – i.e., those that only mildly influenced OA risk in the past - may have become more potent risk effectors given complex interactions with modern factors, including obesity, inflammation, and changes in activity (Hunter and Bierma-Zeinstra, 2019). In both contexts, given typical late-life onset of OA the deleterious consequences of variants to aspects of morphology and joint homeostasis may emerge later, as joint stress compounds over decades, contributing to elevated disease risk. We observed that chondrocyte open-chromatin regions collectively capture a significant portion of OA heritability, and that the OAI patient cohort exhibited a greater proportion of alternative alleles in knee elements specifically. These indicate an increased common variant load in constrained elements, consistent with our model (Fig. 2.4E).

However, it remains unclear how OA prevalence differs between primates. Studies on captive baboons reveal similar prevalence to humans (Macrini et al., 2013), whereas in wild and captive gorillas and chimps showing OA at joint sites prevalence is lower than humans and captive baboons (Lowenstine et al., 2016). This suggests that the nature of OA in humans is quite different compared to African-apes. Understanding prevalence is also complicated as wild catarrhine primates typically show high levels of past trauma, which can be associated with OA and obscure connections to genetic effects (Lowenstine et al., 2016). We did find that both sequence conservation within primates, as well as the predicted age of variants, significantly associate with OA heritability. The latter has been suggested to represent negative selection acting on complex diseases, wherein recent risk variants have had less time to be removed (Gazal et al., 2017). These findings support our model and suggest that part of human OA heritability may result from how genetic variants violate a phylogenetic sequence constraint retained from primates.

Our evolutionary model suggests that violations to constraint may arise not only via genetic drift but also through antagonistic pleiotropy. Recent selection acting on other traits may oppose the functional constraints of a derived knee. Furthermore, recent selection may target the same loci used to sculpt the hominin knee because of their recurring importance to development, growth and homeostasis of the musculoskeletal system (Salazar et al., 2016; Chan et al., 2010; Capellini et al., 2017), and underlying modular *cis*-regulatory systems (e.g., Chen et al., 2016; DiLeone et al., 1998; Guenther et al., 2008; Indjeian et al., 2016). While we did not find enrichment of OA variants nor knee-specific regulatory elements in regions of recent positive selection, several loci were identified which fell within selection windows, and may represent examples of pleiotropy-driven sequence violation (here, again, we refer to pleiotropy at the level of linked genetic variants). Here, we show that one common OA risk variant (rs6060369), present on a haplotype under recent positive selection for height (Capellini et al., 2017), is located within a classic BMP locus involved in knee development, *GDF5*, and within a knee enhancer (*R4*) located using our ATAC-seq strategy. This enhancer and its variant (rs6060369) alter knee regulation *in vivo* and *in vitro*, and contribute to OA susceptibility.

The *R4* enhancer is found in humans, mice, and anole lizards (Wang et al., 2018), and is therefore part of a conserved BMP network involved in knee development. Loss of the conserved enhancer in human chondrocytes down-regulates *GDF5* expression, whereas its loss in mice (*R4*$^{-/-}$) down-regulates *Gdf5* expression locally and alters knee shape at the specific anatomical locations of activity in prenatal and early postnatal development. Note, the localized knee defects seen in *R4*$^{-/-}$ mice represent only a subset of those plaguing *Gdf5* null knees (Pregizer et al., 2018), revealing the modular effects that regulatory elements can have on morphology and disease risk. While the *R4*$^{-/-}$ knee remains functional, these mice do not require any experimental joint insult (e.g., surgical destabilization of the medial meniscus) to recapitulate several characteristic features of human OA progression.

The human OA risk "T" allele at rs6060369 on the selected haplotype now exhibits frequencies between 40-70% across Eurasia, making it present in billions of people, and confers a 1.3 to 1.8-fold increase in OA risk (Miyamoto et al., 2007; Zengini et al., 2018). As it resides at a PITX1 binding site, whose binding we demonstrate in human and mouse, this risk variant alters an evolutionary-constrained transcriptional complex governing normal *R4* function, which our model would predict to be causally linked to OA risk. Mice hetero/homozygous for the single base-pair human risk "T" allele display significant reductions to *Gdf5* expression *in vivo* and present with alterations in some of the same structures and directions of effect as found in *R4* null mice. Yet, this variant modifies knee shape not so dramatically as to markedly disrupt development or early postnatal locomotion. We suggest that by slightly altering knee shape in mice and humans, abnormal joint biomechanics and excessive cartilage wear and degeneration over time may lead to OA, especially when compounded with other risk factors. However, we cannot rule out other non-biomechanical influences for this variant on knee phenotypes. We believe this novel-engineered mouse model (and that of the *R4*$^{-/-}$ model) will provide an effective and realistic tool for testing OA treatments in the context of genetic and non-genetic risk factors. Moreover, that this single base-pair change has such an observable effect on phenotype lends support for GWAS variant testing in the mouse model, especially when it permits the proper three-dimensional and physiological context to observe an effect. Overall, the model proposed here represents a generalized means by which hypotheses linking functional sequences, evolutionary history, and modern-day disease may be made. We present genetic

and functional analyses, which seek to test such hypotheses in the context of heritable OA risk, and suggest an evolutionary framework in considering the molecular underpinnings of complex diseases as a useful avenue for scientific research.

2.4: Acknowledgments

The authors would like to thank: Drs Bruce, Brooks, and Bouxsein for microCT from the Center for Skeletal Research (NIH P30 AR066261) (Massachusetts General Hospital); EpigenDx for ASE; Applied StemCell for allelic mice; Drs. Sabeti (Harvard University (HU)), Zeng (Tufts University), Goldring (The Hospital for Special Surgery) for cell lines; the HU Genome Modification Facility for gene targeting; the HU Bauer Core Facility for sequencing; Tufts Medical Center and Boston Children's Hospital for histology; Dr. Worthington (HU) for statistics guidance; Dr. Price (Harvard T.H. Chan School of Public Health) for heritability discussion; Drs. Tabin (Harvard Medical School), Doxey (University of Waterloo), Lowe (Duke University), Bowes (iMorphics), Yau (Marcus Institute for Aging Research), Felson (Boston University), Pregizer (Boston Children's Hospital), Pilbeam (HU) and Capellini lab members for insight and manuscript review; and three reviewers for their insight into this work. AMK was supported by NIH/NIAMS (R01AR065462) and Orthopaedic Foundation at Boston Children's Hospital; VBK/GHC by NIH/NCATS BU-CTSI (1UL1TR001430), an American Heart Association Scientist Development Grant (17SDG33670323), a Hariri Research Award from the Hariri Institute for Computing and Computational Science & Engineering at Boston University, and NIH (5U01AG-018820 & 1R01AR070139); SY by HU PRISE; NR by NSF (BCS1518596); and TDC by NIH/NIAMS (1R01AR070139) and NSF (BCS1518596) for mouse research and the HU Dean's Competitive Fund and Milton Fund for the human research.

2.5: Author Contributions

T.D.C. supervised project. T.D.C., D.R., and N.R. conceived and designed the project. D.R. performed all computational analyses. E.J. performed selection analyses. J.W., M.Y., and P.M. performed ATAC-seq.

T.D.C. and J.C. designed mice. Z.L. performed mouse work, μCT, histology, and ASE. A.M.K. and S.H. performed morphometrics. J.T.S. and J.K. performed histology/OA analyses. A.M. and R.K. performed microscopy. Z.L. and S.Y. performed *in vitro* experiments. P.M. performed ChIP. P.M., D.E.M., and V.R. performed μCT. V.B.K. and G.H.C. performed morphometrics. T.D.C. and D.R. wrote/edited the manuscript with input from all authors.

2.6 Materials and Methods

Complete Materials and Methods text relating to this study are available in Supplemental File 1 of this Thesis.

Chapter III:

Shifting epigenetic contexts influence regulatory variation and disease risk

This chapter has been published as an article in Aging (Albany NY):

Richard, D. and Capellini, T.D., 2021. Shifting epigenetic contexts influence regulatory variation and disease risk. *Aging (Albany NY)*, *13*(12), p.15699.

3.1: Introduction

It has been suggested that the process of aging, and the concomitant manifestation of aging-related disease, is subject to both genetic and non-genetic factors impacting the regulatory networks (and subsequent behaviours) of aging cells(Brunet and Berger, 2014; Rodríguez-Rodero et al., 2011). Nongenetic regulation of aging refers to epigenetics; chemical changes to the genome (e.g., at the chromatin level) that impact transcriptional programs(Brunet and Berger, 2014), and which have been shown to accumulate with age(Booth and Brunet, 2016; Criscione et al., 2016; López-Otín et al., 2013). The epigenetic state of chromatin can be broadly classified into activating or repressing modifications(Dunham et al., 2012), referring, in part, to the increased/decreased accessibility of DNA to gene-regulatory machinery (e.g. transcription factors), and is established, maintained, and reset to switch between states(Dunham et al., 2012). Ample evidence suggests a causal relationship between changes in epigenetic state with age and hallmarks of aging in cells(Booth and Brunet, 2016; López-Otín et al., 2013). Much recent work has focused on elucidating this relationship and how, ultimately, this contributes to age-related tissue decline and adult diseases(Zhang et al., 2020a).

As aging can also be considered a continuation of development(Blagosklonny and Hall, 2009), the epigenetic changes that are retained from early-life development may have important consequences for the adult epigenome – establishing the context within which epigenetic aging occurs(Hanson et al., 2011). A ‘fetal programming’ model has been suggested whereby early epigenetic plasticity in response to

environmental and nutritional stimuli, while being adaptive and beneficial to fetal and early post-natal growth, has deleterious consequences later in life by contributing to adult disease risk(Gicquel et al., 2008; Godfrey et al., 2013; Hanson et al., 2011; Rinaudo and Wang, 2012). This serves as one possible mechanism for the theorized consequences of selection favouring early-life development at the cost of late-life function(Hanson et al., 2011; Kirkwood, 2008, 2002). Evidence supporting this model has been largely limited to DNA methylation(Conradt et al., 2018; Fraga et al., 2005; Hanson et al., 2011; Ollikainen et al., 2010), though replication of important loci findings has been difficult(Saffery and Novakovic, 2014).

Epigenetic marks established during development can persist into adulthood(Hanson et al., 2011), but they do so in the context of shifts in epigenetic states (see below) as tissues transition into their adult forms and functions. This transition process has been characterized with respect to DNA methylation, chromatin state, and gene expression across multiple tissues(Aagaard-Tillery et al., 2008; Horvath and Raj, 2018a; Yan et al., 2016). Furthermore, these fetal to adult epigenetic shifts can be compounded by additional modifications through aging-associated epigenetic changes. Such epigenetically-regulated biological pathways involved in development, such as Wnt signalling, subsequently take on a role in tissue homeostasis in adults and are implicated in age-related tissue decline(Maiese et al., 2008; Salminen and Kaarniranta, 2010) – suggesting a molecular link between processes mediating growth and aging(Blagosklonny and Hall, 2009; Magalhães, 2012). Thus, an important component of understanding the contributions of fetal programming as well as epigenetic aging to disease biology and risk is characterizing the epigenetic changes between fetal and adult tissues and how these might interact with subsequent aging-associated modifications.

While epigenomes vary between cell types(Rivera and Ren, 2013; Roadmap Epigenomics Consortium et al., 2015) and changes to epigenetic state with age may be expected to manifest differently, the aging epigenetic shifts outlined above have been repeatedly observed across tissues(Brunet and Berger, 2014; Day et al., 2013; Horvath and Raj, 2018a; Kubben and Misteli, 2017; López-Otín et al., 2013). Similarly, while age-related expression changes do exhibit tissue-specificity, there is evidence of potential synchronized changes across different sets of tissues(J. Yang et al., 2015), particularly for certain sets of

genes and pathways(Frenk and Houseley, 2018; J. Yang et al., 2015), and these changes may integrate at multiple different epigenetic levels(Bou Sleiman et al., 2020). Together, these findings suggest that a central trajectory for epigenetic state that reflects innate aging processes may exist(Horvath and Raj, 2018a), upon which extrinsic and cell-type effects are layered. Similarly, studies between fetal and adult tissues have found that, while epigenetic change is observed within individual tissues, there are also common trends of development (e.g., chromatin restriction, particularly at loci involved in early development)(Yan et al., 2016; Yuen et al., 2011; Zhu et al., 2013).

Importantly, the epigenetic state of genetic variants (e.g., single nucleotide polymorphisms) influences their regulatory effects, and subsequent association with heritable disease risk(Boyle et al., 2012). Thus, general epigenetic trends across early development and later aging may influence the phenotypic effects of regulatory mutations, albeit the extent to which this occurs is unknown. These phenotypes, if impacting an individual's fitness, may be acted upon by natural selection. Evolutionary theories have been proposed which suggest that mutations contributing to aging pathologies are 'allowed' to accumulate due to the reduced fitness consequences of disease in older, post-reproductive individuals(Charlesworth, 2001), or that beneficial mutations selected for early development become deleterious with age(Kirkwood and Rose, 1991; Richard et al., 2020; Rodríguez et al., 2017). Studying the added dimension of epigenetic context may provide a fresh perspective on theories of aging and selection. For example, deleterious mutations that change epigenetic context later in life may have different regulatory effects, and thus different fitness consequences, which alter the selective pressures acting on them.

In the present study, we seek to characterize common epigenetic trends between fetal and adult tissues, and examine the potential interaction of these developmental changes with later changes associated with epigenetic aging in adult tissues. We utilize our findings to propose a model for how evolutionary forces may have acted at these loci in humans, and how these forces in turn influence the distribution of mutations conferring heritable disease risk across a number of age-associated pathologies.

3.2: Results

3.2.1: Defining chromatin accessibility change, its genomic context, and loci subject to change

To investigate epigenetic changes occurring over the course of post-natal development and aging, we focused on chromatin accessibility, as it reflects the regulatory potential of a genetic locus and can be considered a property of the epigenome which integrates a number of possible epigenetic phenomena (e.g., regulatory factor binding, chromatin remodelling, etc.)(Klemm et al., 2019). We thus consider regions with altered chromatin accessibility as being indicative of epigenetic modifications or 'shifts' in context. As a read-out of accessibility we analyzed DNase-I hypersensitivity datasets acquired from primary human tissue, and obtained fetal/adult sample pairs for eight distinct primary tissues (spleen, lung, muscle, stomach, kidney, brain, heart, and skin, see Supplemental Table S3.1)(Davis et al., 2018; Dunham et al., 2012). For each tissue at each time-point, called accessible or "open" chromatin regions were consolidated across biological replicates, then further aggregated by tissue and stage (see Appendix C).

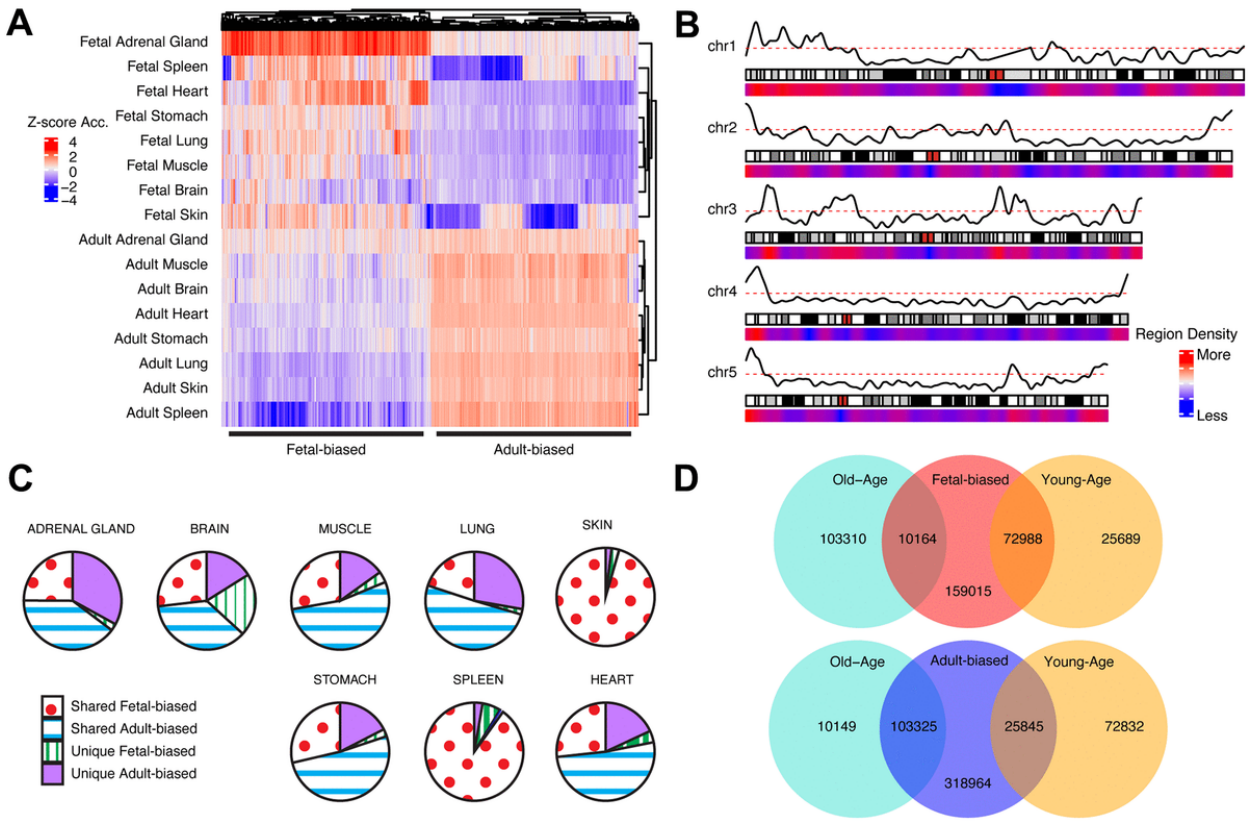


Figure 3.1. Cross-tissue accessibility. (**A**) Representative heatmap of Dnase-I accessibility for regions significantly different between fetal/adult tissues. Color scale indicates magnitude of chromatin accessibility signal (see Appendix C). Horizontal lines denote defined fetal-biased (left) and adult-biased regions. (**B**) Genomic distribution of regions changing accessibility in fetal and adult comparison. Red/blue: density of defined differentially-accessible regions. Solid black line: relative proportion of regions more accessible in adult (top) or fetal (bottom) tissues. First five autosomes shown (see Supplementary Figure 3.2). (**C**) The proportion of defined altered-accessibility regions between adult and fetal samples for indicated tissues which are unique to that tissue, or captured in the pan-tissue set. (**D**) Overlaps between regions defined as differentially-accessible in fetal/adult comparison and those defined in the young/old-age comparison. Directionality in accessibility change is significantly shared (see Supplementary Table 3.1).

We first identified chromatin regions exhibiting recurrent accessibility changes between fetal and adult samples across tissue types (see Appendix C, Figure 3.1A-B, Supplemental Figure S3.1-2). We define regions as ‘adult-biased’ if they exhibit increased differential accessibility in adults compared to in fetuses.

Conversely, we define regions as ‘fetal-biased’ if they exhibited decreased differential accessibility in adults compared to in fetuses. These ‘pan-tissue’ altered regions were compared to those defined in individual tissue comparisons, showing substantial but not complete overlap (Figure 3.1C) – suggesting that our approach captures cross-tissue signals of broader developmental changes and not tissue-specific effects. We next explored possible signals of epigenetic aging occurring in the context of fetal to adult changes, by further dividing our adult tissue samples into ‘younger’ and ‘older’ age categories (Methods, Supplemental Figure S3.3). We then assessed accessibility change between young and old occurring within the ‘adult-biased’ and ‘fetal-biased’ regions defined above (Figure 3.1D). This approach identified regions for which young-old differences mirrored fetal-adult differences, as well as regions where aging changes appear to counter developmental patterns. We observed a tendency for shared directionality in gains or losses of accessibility; i.e., adult-biased regions tended to also have increased accessibility in older adult samples, while regions losing accessibility in adult samples (i.e. are fetal-biased) continued this trend in older samples (chi-sq test, $p < 0.05$, Supplemental Table S3.1). In this text, we therefore refer to regions with greater accessibility in older samples as ‘old-biased’ and regions with lower accessibility in older samples as ‘young-biased’. As described in Appendix C, we considered histone mark and DNA methylation changes, key features of developmental(Roadmap Epigenomics Consortium et al., 2015; Yan et al., 2016; Zhu et al., 2013) and aging epigenetic changes(Booth and Brunet, 2016; Horvath and Raj, 2018a) as additional means to validate the behaviour of these region sets (see also Supplemental Table S3.1).

To gain insights into the roles these region sets have in transcriptional regulation, we next characterized the genomic distribution of our adult-biased and fetal-biased region sets using adult tissue epigenetic states(Roadmap Epigenomics Consortium et al., 2015) (Appendix C). We found that our region sets preferentially fell within different epigenetic states (e.g., enhancers, heterochromatin) depending on the nature of their accessibility shift (e.g., adult-biased, old-biased), suggesting that these shifts may be associated with altered regulatory biology at different loci, and that the interaction between fetal and adult shifts as well as young and old-age shifts heavily favours developmental changes to accessibility (see Supplemental Figures S3.4-S3.5, Appendix C).

As accessible chromatin regions often serve to regulate gene expression(Klemm et al., 2019) by acting as *cis*-regulatory sequences, we next sought to identify the potential role of our regions in regulatory changes occurring during development and aging. We did this by considering promoter-level accessibility (see Methods, Figure S6), promoter-capture (Hi-C) interactions(Jung et al., 2019), and regulatory-domain annotations(McLean et al., 2010) for genes which may be subject to control by these regulatory regions. We found a general pattern for enrichment of immune-related gene sets with the adult-biased set, while development-related (e.g. cellular proliferation) terms were enriched with fetal-biased regions, patterns echoed when considering old-biased and young-biased region sets, respectively (see Appendix C, Supplemental Table S3.2).

We next incorporated tissue-expression datasets looking for general gene expression trends between fetal and adult tissues (see Methods). We observe similar enrichment terms as well as significant overlaps with gene sets defined on the regulatory level (see Appendix C, Supplemental Table S3.2). Similarly, we utilized GTEx (gene tissue expression) datasets(GTEx Consortium, 2013) to look for corresponding shifts in gene expression with age, similar to previous work(Benayoun et al., 2019) (see Methods)(Supplemental Table S3.2). While we did not observe significant overlaps between these aging-expression gene sets and those defined using aging-accessibility changes, we did see significant overlaps with the fetal/adult expression comparisons, along with enriched gene sets with relevance to aging biology (see Appendix C, Supplemental Table S3.2, Discussion).

As development and aging are phenomena subjected to the actions of random and directed evolutionary forces(Carroll, 2008; Kirkwood, 2008; Kirkwood and Holliday, 1979), we next develop expectations for how these epigenetically-altered regions may have evolved over time.

3.2.2: Sequence evolution of epigenetically-altered regions

Development and aging are simultaneously very ancient and variable(Carroll, 2008; Jones et al., 2014) biological processes and are particularly divergent in key species(Ferris et al., 2018). Thus, it may be the

case that both development and age-associated regions have been shaped by a mix of evolutionary forces acting to either maintain or modify genetic sequences (e.g., regulatory enhancers). To address this possibility, we examined patterns of sequence conservation in our epigenetic datasets using phyloP(Pollard et al., 2010), a measure of nucleotide conservation and/or acceleration across species (Figure 3.2A). Across primates, we observed that fetal-biased regions tended to have greater sequence conservation than adult-biased regions, and furthermore that both sets differed significantly from those regions not defined as developmentally-altered (Supplemental Table S3.3). These patterns were similarly observed when comparing age-associated regions (Supplemental Table S3.3). These findings of conservation differences between sets suggests that the greater regulatory and developmental role associated with fetal-biased and young-biased regions (e.g., enriched for enhancer elements) exerts functional sequence constraint while adult-biased and old-biased regions (e.g., enriched for repressed segments) are less conserved across species.

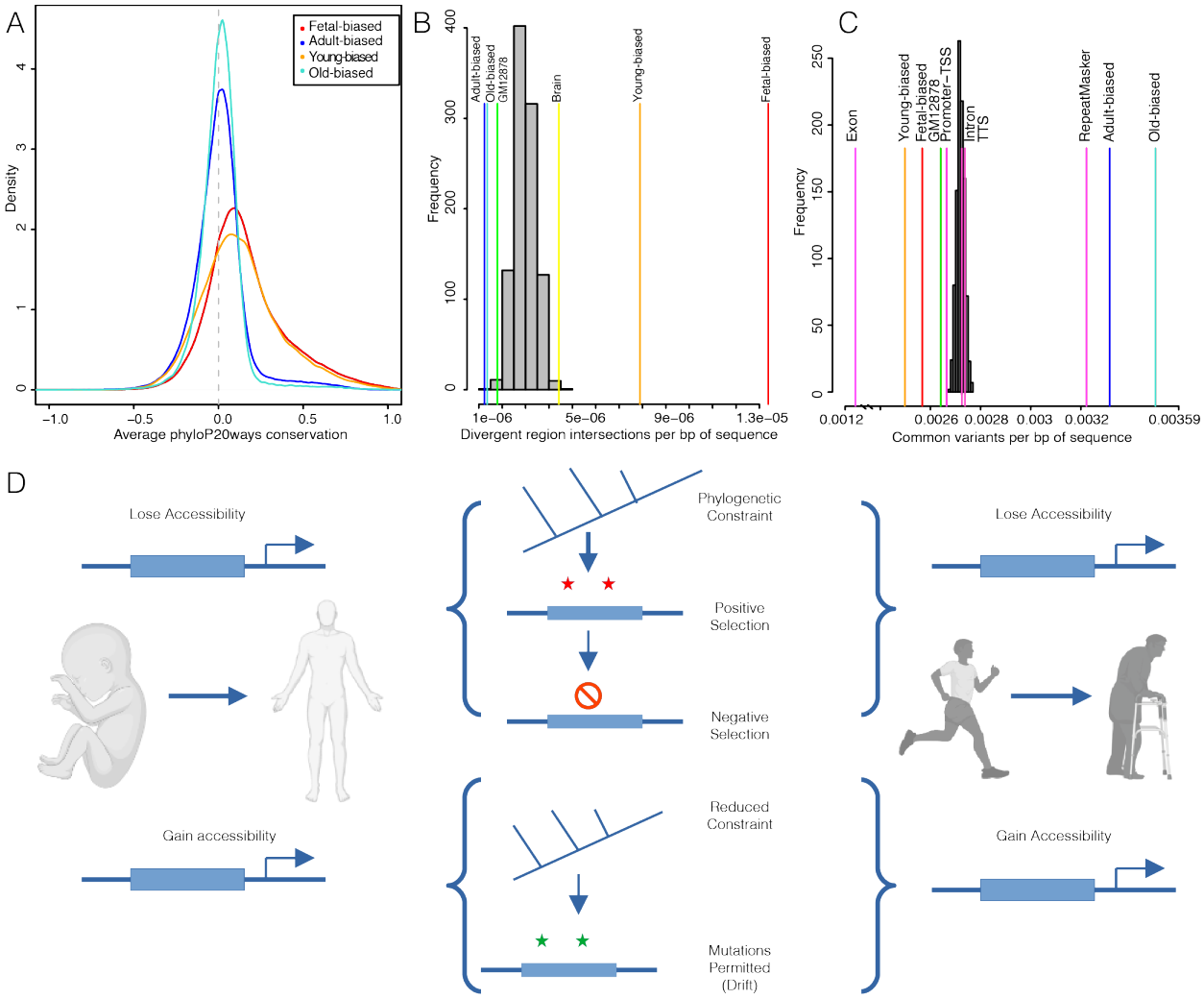


Figure 3.2: Sequence evolution of age-altered regions. (A) Distribution of average per-region sequence conservation (phyloP20ways) in differentially-accessible regions (see color legend). (B) Overlaps of developmental and age-related region sets and human acceleration regions. Overlaps shown relative to set size (per bp of sequence) for background (gray) and target (colored) sets; labels correspond to Supplemental Table S3.3. (C) Intersections of common human variants (per bp of sequence) for target (colored) and randomized (grey) region sets; labels correspond to results in Supplemental Table S3.3. (D) Diagram summarizing results of evolutionary sequence analyses. Accessible regions, here diagrammed as an upstream enhancer element (thick blue box), which either gain or lose accessibility over development (left) or ageing (right) exhibit different patterns of evolutionary sequence behavior. Created with BioRender.com.

Within this broader context of species diversity and evolution, humans and chimpanzees display marked and obvious differences in development and longevity(Carroll, 2008; Kirkwood, 1997). This relatively recent divergence is thought to be driven largely by non-coding changes to *cis*-regulation(Varki and Altheide, 2005). We therefore next looked for evidence of regulatory modifications to biological processes that may contribute to these human/chimp differences. To do this, we intersected our regions sets with sequences demonstrating significant divergence along the human lineage (e.g., 'human accelerated regions' (Hubisz and Pollard, 2014)). We found that fetal-biased regions were enriched for signals of acceleration while adult-biased regions were depleted (Figure 3.2B). Similar patterns were seen for young versus old-biased regions (Figure 3.2B, Supplemental Table S3.3). We found a number of genes involved in development and aging processes with putative nearby regulatory elements intersecting accelerated regions, two examples of which are shown in Supplemental Figure S3.7 (see also Appendix C).

To gauge the evolutionary interaction between sequence constraint across species and within-species variation, we next assessed modern-day human diversity within region sets (Methods). We found that genetic diversity in fetal-biased regions was markedly reduced (i.e., constrained) compared to genomic backgrounds, as well as to intronic and promoter-TSS elements (Figure 3.2C, Supplemental Table S3.3). Conversely, adult-biased regions were enriched for sequence diversity, at the level of annotated repeat elements. These patterns were accentuated when examining young- and old-biased region datasets, and comparing region sets directly (Supplemental Table S3.3, Appendix C). Importantly, when we considered sequence diversity within other ape species, we also observed a decrease in fetal-biased, and young-biased sequence diversity (relative to adult-biased and old-biased, respectively). This latter finding further suggests that fetal-biased regions are associated with conserved regulatory function that discourages mutation and/or drift (Supplemental Figure S8, Supplemental Table S3.3).

Overall, natural selection appears to have acted upon regions subject to accessibility shifts in development and aging, modifying some loci (i.e., accelerated divergence indicative of ancient positive selection) while protecting others (i.e., reduced variation indicative of more recent negative selection) (Figure 3.2D).

Importantly, selective forces, both positive and negative, manifest phenotypically through the effects of random genetic mutations, which act to modify gene regulatory networks to varying degrees. We next examine this relationship.

3.2.3: Epigenetic shifts in age-associated trait associations

In our above analyses, the observed signals of consistent inter- and intra-species conservation in regions most associated with early development (i.e., the fetal-biased set) follows with the expectation that variants negatively impacting early-life would be subject to stronger purifying selection(Carnes and Olshansky, 1993; Hanson et al., 2011; Ricklefs, 2008). Conversely, variants with later-manifesting effects, i.e., those within regions increasing in local accessibility with age (i.e., the adult-biased set), would be subjected to substantially weaker selection and may therefore be 'tolerated'(Williams, 1957; Wright et al., 2003a). To test expectations of the possible deleterious effects of variants subject to accessibility change over development and aging, we utilized GWAS datasets available from the UK Biobank(Sudlow et al., 2015). We extracted summary-statistics for a collection of 127 complex diseases/pathologies falling into aging-related categories(Chang et al., 2019), including metabolic disorders, cancers, cardiovascular disease, and musculoskeletal impairment (Supplemental Table S2.4, Methods). We similarly analyzed a set of developmental trait GWAS to act as a control for our fetal/adult accessibility comparisons, and finally considered longevity GWAS data (Supplemental Table S2.4, see Appendix C).

It has been suggested that the highly polygenic nature of complex traits and diseases reflects cumulative regulatory modification to a 'core' set of genes which functions most proximately in relevant biology(Boyle et al., 2017). Across age-associated diseases, this may reflect general aging processes, and regulatory variants impacting these would be expected to contribute to heritable aging-disease risk broadly. Given this rationale, we first considered the behaviour of individual SNPs nearby accessibility-altered regions across diseases, and subsequently these behaviours at the gene-locus level (below). We aggregated per-SNP associations across diseases as a singular cross-set metric of risk association (Appendix C). We confirmed that ClinVar variants, variants for which possible clinical significance have been described(Landrum et al., 2018), tended to be more risk-associated by this metric, as we would expect

(Supplemental Table S2.4). Additionally, across all diseases we individually performed enrichment tests for strongly-associated variants nearby our region sets, which corroborated the cross-disease results described below (Figure 3.3A, see Appendix C).

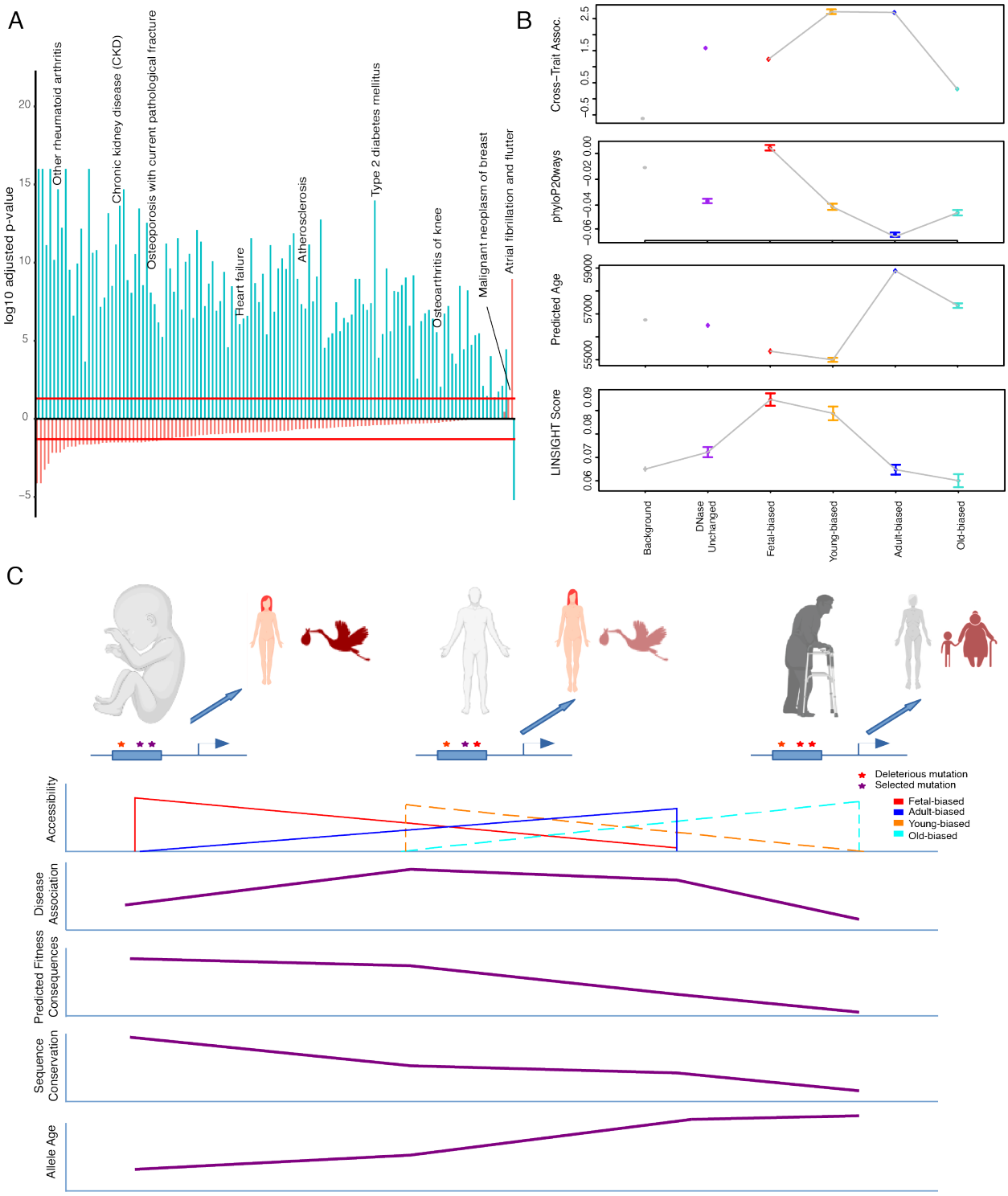


Figure 3.3: Epigenetic context and heritable disease associations. (**A**) Adjusted p-values for hypergeometric tests showing enrichment/depletion (positive/negative) for GWAS variants nearby regions increasing (blue) or decreasing (red) accessibility across adult tissues for a number of age-associated diseases (see Supplemental Table S3.4). (**B**) Cross-set disease associations, and additional per-SNP

metrics, for variants nearby developmental and age-altered region sets along with unaltered DNase sites and variants not nearby accessible regions. See Supplemental Table S3.4 and Methods. (**C**) Model for the effects of epigenetic context on disease association and sequence evolution. (Top): Example enhancer elements more accessible in fetal, adult, and old-adult tissues (left-right) which have been modified by mutations. (Left): Deleterious mutations disrupting regulation in development stand to have the biggest impact on fitness, while having a moderate effect on tissue homeostasis. (Middle): Mutations disrupting regulation in young-adult tissues have a moderate impact on fitness, but a larger effect on tissue homeostasis (particularly over adulthood). (Right): Mutations disrupting regulation in old-adult tissues have weak impacts on fitness, and a weak effect on tissue homeostasis (which has already deteriorated with age). (Bottom): Illustrating patterns of accessibility, disease association, sequence constraint and variant allele age for these sets of regions changing accessibility over time.

As variants in accessible non-coding regions likely have regulatory impacts generally(Klemm et al., 2019), we confirmed that variants within or nearby regions not classified as strictly developmental nor age-altered tended to have greater association than non-accessible variants. However, this control set had significantly less association than variants nearby sets of both developmentally- and age-altered regions (i.e. fetal/adult-biased, and young/old-biased regions) (Supplemental Table S3.4).

Considering first developmental change, we found that variants in regions gaining nearby accessibility in adults (i.e., adult-biased) have greater association with disease than in regions those losing nearby accessibility (i.e., fetal-biased) (Figure 3.3B). Unexpectedly, when looking at aging accessibility changes, we observed that variants in regions gaining nearby accessibility in older-samples actually have lower cross-disease associations than those in regions becoming more accessible in younger samples (Figure 3.3B, Supplemental Table S3.4). Furthermore, we found that for intersections of development and age-altered regions that the increased disease association with adult-biased regions was abrogated when intersected with old-biased regions. The magnitude of region-set differences in disease associations was also greater in the young/old-biased comparisons (see Figure 3.3B, Appendix C).

Taken together, these results would suggest that those variants most accessible in younger adults stand to have the greatest impact (in terms of association p-value) on late-life disease risk – a finding that may have important implications for understanding the development of disease over adulthood (see Figure 3.3C, Discussion).

We next considered these disease-association patterns at the gene-locus level. Briefly, for a given disease we assign the most significant nearby SNP to all genes, and subsequently rank genes based on their assigned GWAS signal. Gene ranks are then aggregated across diseases, looking for genes consistently ranked higher across sets (Appendix C). To confirm the behavior of this gene-ranking method, we compared the cross-set ranking of genes associated with homeostatic processes (based on GO annotations) to randomized gene sets, finding that these gene loci tend to harbor stronger genetic variants across a larger number of diseases than expected (compared to randomized sets), but not so for genes involved in reproductive organ development (see Appendix C, Supplemental Table S3.4).

We applied this method to the sets of development and age-associated genes we defined above and asked whether they tended to have more or less cross-disease GWAS signals than expected. Our sets defined by accessibility-region contacts supported our earlier findings on strong GWAS signals nearby development and age-altered regions – namely, loci of both adult-biased and young-biased gene sets were enriched for strong GWAS signals across diseases, while fetal-biased and old-biased gene sets were associated with relatively weaker GWAS signals (Supplemental Table S3.4). Sets defined by RNA-seq data showed more of a mix of enriched/depleted GWAS signals across developmental and age comparisons, reflecting the possibility that a mixture of genes increasing and decreasing expression over time may additively contribute to aging disease biology (see Appendix C).

Given our results, found at both genome-wide and gene-locus set levels, we finally sought to take an unbiased approach to identify relevant ‘core’ aging-related genes solely on the basis of aggregate GWAS signals (see Appendix C). Overall, we had limited success in defining a set of genes with clear, pan-tissue biological relevance, suggesting that, if such a core does exist, that it may be too broad, or the per-locus signals too moderate, for our method to robustly detect. However, since our results suggest the importance of altered epigenetics in modifying GWAS associations, we performed a similar gene-

prioritization analysis using variants occurring nearby altered-accessibility regions (Appendix C). This yielded markedly different enrichments for terms relating to immune processes and gene regulation (see Appendix C, Supplemental Table S2.4). One particular set of genes, involved in histone deacetylation, has repeatedly been linked to aging and epigenetics(Hall et al., 2013; Sen et al., 2016a) and was identified using our set of young-age regions (see Figure 3.4). We explore this set in more detail below.

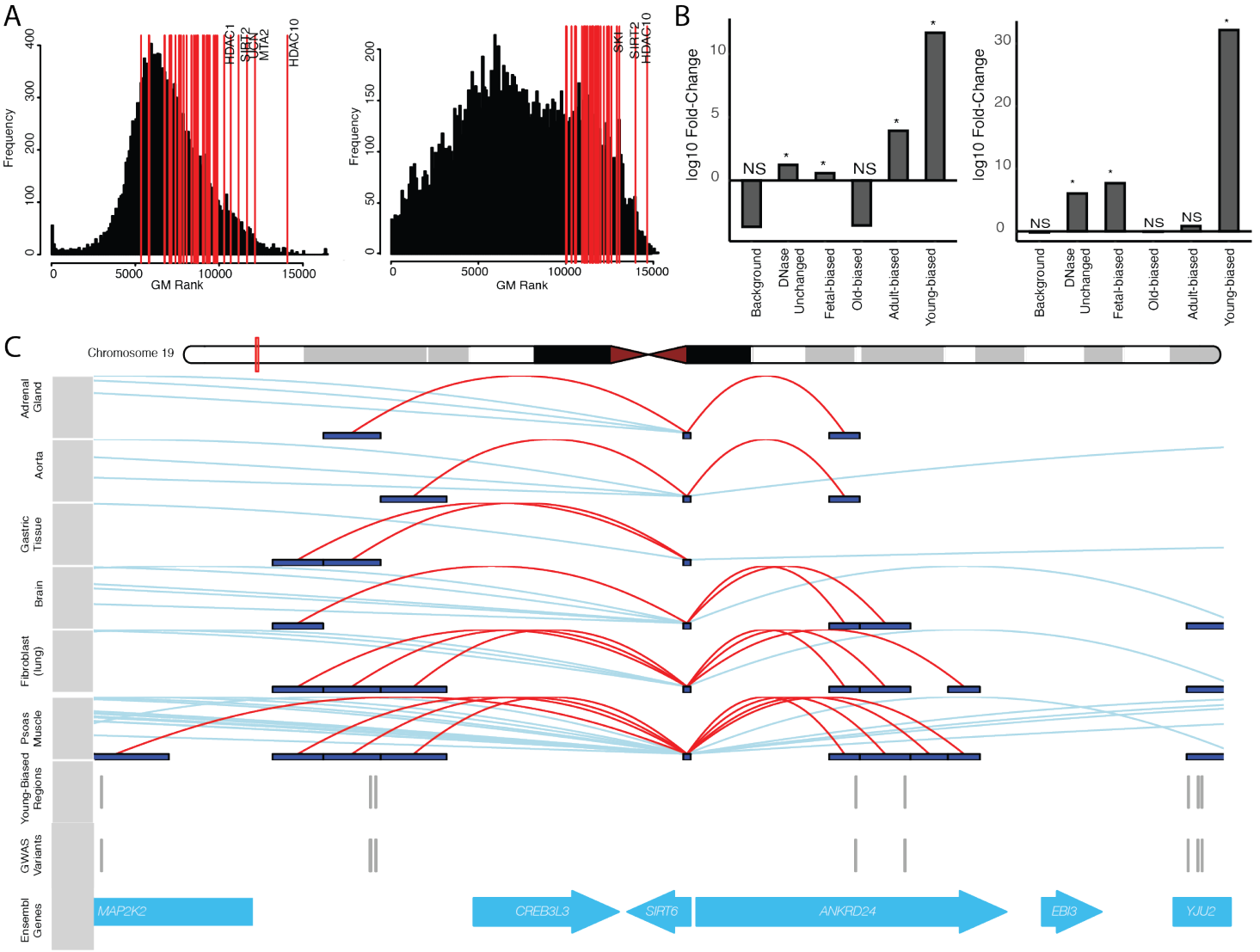


Figure 3.4: Altered-accessibility regions identify relevant aging biology. (**A**) (Left) Distribution of cross-disease ranks for all protein-coding genes, when ranking by local variants independent of accessibility

data (see Appendix C). Red lines indicate genes within the 'histone deacetylation' (HDAC) GO term; top ranked genes (by geometric mean) are indicated. (Right) Similar distribution of cross-disease ranks, ranking genes with variants nearby young-biased regions. Red lines indicate top HDAC genes by rank. (**B**) (Left) Fold-change of normal cumulative distribution function (CDF) p-values of variants within HDAC gene loci associated with different region sets, relative to CDF test performed using all variants, for cross-disease Z-score metric (see Appendix C). (Right) Similar plot for per-variant LINSIGHT scores. See Supplemental Table S3.4. (**C**) Variants directly intersecting young-biased regions which interact with the *SIRT6* promoter. (Top) Visualized promoter-capture data across multiple cell-types. (Bottom) Tracks indicating variants which overlap young-biased regions within the *SIRT6* gene locus.

3.2.4: Sequence evolution and disease association

Our previous analyses found that patterns of inter- and intra-species sequence conservation depended on epigenetic status (i.e. degree of accessibility) of regulatory elements. Subsequently, we found that the risk association of variants across a number of age-associated diseases also varied based on accessibility change in the vicinity of the variant. Much work has been done on understanding the relationship between sequence conservation and disease risk(Cooper et al., 2010; Cooper and Shendure, 2011; Hujoel et al., 2019; MacArthur et al., 2014). For example, a transcription factor (TF) binding site may be subjected to negative selection to conserve its sequence and hence function. Mutations that occur within this site would more likely impact *cis*-regulatory biology, and therefore manifest an association with disease. If this disease impacts fitness, then over time, such mutations will be eliminated, so that genetically 'older' mutations are less prevalent(Gazal et al., 2017; Hujoel et al., 2019). Given our interest in the evolution of development and aging processes, we wanted to investigate the role that epigenetics has on this disease-evolution relationship - and whether this holds with our data. By comparing the cross-trait associations of variants falling within and outside primate-conserved sequences (phastCons)(Siepel et al., 2005) (Appendix C), we found that variants within conserved sequences tend to have greater disease associations, along with younger estimated allele age (Supplemental Table S3.4). These patterns also hold true for phastCons sequences within age- and developmentally-altered regions

(Supplemental Table S3.4), and follow with previously observed enrichments for GWAS associations of conserved, younger (allele age) variants(Finucane et al., 2015; Gazal et al., 2017; Hujoel et al., 2019).

We next considered primate conservation, estimated allele age, and cross-set association of variants, looking for the effects of nearby accessibility change on these metrics (Appendix C). As an additional metric for predicted fitness consequences, particularly of non-coding variants, we also included per-bp LINSIGHT scores(Huang et al., 2017), which integrates data on chromatin accessibility, TF binding motifs, and comparative genomics.

First, we found that variants falling near fetal-biased regions were more conserved, younger, and had lower cross-set association, while variants near adult-biased regions behaved oppositely (Figure 3.3B, Supplemental Table S3.4). We also found that the predicted functional consequences associated with fetal-biased regions were greater than with adult-biased regions, despite the lower cross-set association with aging-associated diseases (Figure 3.3C, see Discussion). Variants falling near old-biased regions were less conserved, had older allele age, and had lower cross-set association than their young-biased counterparts (Supplemental Table S3.4). These old-biased regions were also associated with the lowest predicted functional consequences (in aggregate) of any set, while the set of young-biased regions had the second-highest average. To compare these behaviours with variants of annotated clinical significance we independently examined ClinVar variants, which while demonstrating increased cross-set association, tended to also be more conserved, younger, and have stronger predicted fitness consequences (Supplemental Table S3.4).

Collectively, our results indicate that variants stratified by nearby accessibility change violate the expected relationship between sequence conservation and disease association (behaviours instead observed for ClinVar variants). Namely, those regions exhibiting the highest sequence constraint (Figure 3.3C, left) do not also exhibit the strongest aging-disease associations, nor do those regions exhibiting the weakest constraint, as might be expected in a ‘mutation accumulation’ theory of aging(Charlesworth, 2001). However, when considering predicted functional consequences (LINSIGHT), which are not defined based on aging demographic data, this pattern is reversed (i.e. the most-constrained set, fetal-biased regions, had the strongest predicted consequences despite weaker aging-disease associations). This

unexpected behaviour may have important implications for evolutionary models of late-onset complex disease genetics. Based on our results, we propose such a model suggesting the outsized impact of regulatory sequences active in early adulthood on genetic contributions to aging-associated disease risk (see Figure 3.3C, Discussion).

Our proposed model suggests that focusing on disease risk loci containing such putative regulatory sequences (i.e. young-biased regions), should implicate sets of genes involved in aging biological processes. Our gene-level GWAS analyses using young-biased regions identified genes involved with histone deacetylation as being more consistently associated with aging-disease GWAS signal, a pattern which was diminished when considering gene-level associations in the absence of this epigenetic information (Figure 3.4A), and when using other region sets (e.g., old-biased regions) (Figure 3.4B). Histone deacetylation enzymes have known impacts on epigenetic aging biology(Hall et al., 2013; Sen et al., 2016a; Zhang et al., 2020a) and aging diseases(Tang et al., 2013). Within our young-biased enriched gene set we identified *SIRT6* and *SIRT7* as having multiple variants falling nearby young-biased regions which contacted their respective gene promoters (Figure 3.4C). Both these enzymes have been associated with maintaining heterochromatin during aging(Benayoun et al., 2015; Bi et al., 2020; Tasselli et al., 2017); *SIRT7* decreases expression with age, and antagonizes hMSC epigenetic aging(Bi et al., 2020), while *SIRT6* loss manifests an aging-like state(Mostoslavsky et al., 2006). It may be possible that decreased accessibility of regulatory regions controlling the expression of these genes are involved in decreases in sirtuin expression and heterochromatin(Benayoun et al., 2015).

3.3: Discussion

In this study we sought to describe how changing epigenetic context, defined here as changes to chromatin accessibility over both development and subsequent aging, influences the behaviour of evolutionary forces and genetic disease risk at the sequence level. To address this question, we defined genomic regions whose chromatin accessibility consistently shift over the course of development and aging. Our approach to identify epigenetic shifts relies on the observation that chromatin accessibility broadly reflects the regulatory capacity at a given locus (Klemm et al., 2019), though we acknowledge

that more subtle epigenetic changes (e.g. post-translational modifications, CpG methylation) may not be well captured by this accessibility-based definition of epigenetic context.

We performed several analyses suggesting that these regions reflect developmental and aging signatures from previous literature, including genomic features (e.g., repeat elements, CpG sites), epigenetic states (e.g., euchromatin/heterochromatin) and histone mark data. Gene sets associated with developmentally-altered regions were enriched for immune system function and cellular proliferation terms, echoing an earlier study of fetal to adult epigenetic changes(Yan et al., 2016). Furthermore, we found correspondence between these gene sets and genes whose RNA-seq expression generally shifted between fetal and adult tissues. Incorporating an independent RNA-seq dataset of adult age-stratified tissues we did not observe the same level of correspondence with age-altered regions – it is possible that some aspects of epigenetic aging (e.g., global de-repression(Booth and Brunet, 2016; López-Otín et al., 2013; Tsurumi and Li, 2012)) may account for this disconnect, whereby local accessibility changes are less-directly linked to local expression changes. Interestingly, comparing patterns of expression change in our fetal-adult and young-old comparisons, we saw similar gene-set enrichments (i.e., cell-cycling biased towards fetal and younger-age samples, immune responses biasing towards adult and older-age samples), suggesting that the continuation of epigenetic shifts we observed across development and aging (Figure 3.1D) may be mirrored at the expression level.

Given that epigenetic state impacts the potential regulatory effects of deleterious variants(Boyle et al., 2012), we looked to see if local development and/or ageing changes to epigenetic context impacts the strength of association between variants and aging-associated diseases. While it is possible that a number of these aging diseases share genetic correlations(Bulik-Sullivan et al., 2015), that these variants are associated with multiple age-associated diseases is also a key expectation for the functional relevance of age-altered regions. In other words, it is the change in epigenetic context that modifies the regulatory potential of these variants, and this has direct impacts on individual associations with multiple diseases.

According to the fetal programming model, we would expect that regulatory regions most active during early development, both dictating developmental processes as well as responding to environmental

perturbations(Hanson et al., 2011), would have an out-sized impact on the manifestation of adult-onset diseases. This would be evident in the increased associations of nearby variants with heritable risk for these diseases. However, we found that such fetal-biased regions were not those having the greatest impact with regards to aging disease associations, despite having greater predicted fitness consequences – finding instead that adult-biased regions are depleted for aging disease GWAS signals, and associated more with developmental diseases/traits (Table S3.4). A recent study of fetal chromatin accessibility at the single-cell level similarly found genetic associations with developmental traits (e.g., height) using regions accessible in different cell-types(Domcke et al., 2020). We suggest that the 'fetal programming' of epigenetic status during early development, genome-wide, has a more moderate impact on aging disease biology than has been previously suggested – though we note that certain developmental loci (e.g.. Wnt genes) can and do play a role in aging(Blagosklonny and Hall, 2009; Magalhães, 2012).

According to a model wherein epigenetic aging influences the phenotypic effects of regulatory mutations, we would expect that mutations with increased local accessibility in adult tissues, particularly aged adult tissues, would have stronger impacts on aging disease biology in these tissues (reflected in increased association with heritable disease risk). Here, we found that variants gaining nearby accessibility (i.e., adult-biased regions) have stronger associations across a number of aging-related diseases including several kinds of neoplasms, arthritis, and atherosclerosis. This finding suggests that the regulatory effects of deleterious variants may become 'uncovered' as tissues mature and follows with proposed links between development and ageing processes(Blagosklonny and Hall, 2009; Horvath and Raj, 2018a; Magalhães, 2012). However, we also found that regions most accessible later in life, when these diseases manifest, are actually associated with weaker GWAS variants. This young/old bias in aging-disease GWAS signal was far stronger than the fetal/adult bias (i.e., the young-biased set was more strongly enriched than adult-biased, and vice-versa). Taken together, these results suggest that (1) accessibility changes in aging tissues have a greater effect on aging tissue diseases, but (2) that variants more accessible earlier in adult life play a bigger regulatory role in contributing to disease risk than do those which gain accessibility later on. Disruptions to regulation in younger tissues may act to set tissues

down a path of increasing dysfunction and decline, especially if deleterious variants are able to (cumulatively) contribute to dysfunction as they gradually lose activity with age. In other words, by the time an individual reaches old-age their tissues have had sufficient time to accumulate these dysfunctional effects, 'setting the stage' for disease manifestation. Variants more active in old-age, by contrast, have less of an impact on disease manifestation, as their regulatory effects have had less time to integrate. It may be that the time at which disease prevention and/or intervention would be most effective is, perhaps non-intuitively, early in adult life rather than once phenotypes manifest.

We cannot rule out the effects of cell-type specific epigenetic (accessibility) shifts influencing the phenotypic impacts of regulatory sequence modifications on aging-associated disease risk. Similarly, it has been suggested that a facet of aging is 'epigenetic drift' – the accumulation of epigenomic aberrations that contribute to mis-regulation of gene regulatory networks, a component of which is tissue-specific(Teschendorff et al., 2013; Zampieri et al., 2015). However, the pan-signals which we do observe with respect to evolutionary forces, disease associations, and sets of implicated gene loci indicates the relevance of our approach in understanding the broader components of development and aging-accessibility changes, which may be complemented with future research focusing on those more tissue-specific components.

Regulation of general aging-related mechanisms, as well as increases in heritable disease risk, represent phenotypes upon which evolutionary forces may act to modify aging and mortality rates. We found that young-biased regions were enriched for signals of positive selection, a number of which implicated relevant aging-associated genes, and exhibited increased phylogenetic and within-human sequence constraint. Given that these behaviours are intermediate between those observed with regions more accessible in fetal and older-adult tissues, we suggest the following model (Fig. 3.3D):

Regulatory sequences most active during development are subjected to strong negative selection, both to maintain human-derived functional sequences and discourage subsequent modifications, as dysregulation of development would have the largest fitness consequences. Similarly, sequences most active during early adulthood are subjected to negative selection to maintain proper tissue maintenance and discourage disease. However, the strength of this selection is reduced, as we expect fitness

benefits/costs to diminish with age as individuals reproduce less frequently(Carnes and Olshansky, 1993; Ricklefs, 2008; Williams, 1957). Thus, despite the fact that mutations within or nearby these functional sequences stand to have the greatest impact on disease risk (as noted above) they are less-efficiently purged, and are allowed to accumulate over generations(Charlesworth, 2001). Finally, sequences most active in older adults are under relaxed selective pressures and allowed to drift – mutations are permitted and retained, particularly due to the reduced associations that these mutations have with heritable disease risk. Overall, this model suggests that considering the changing epigenetic context of disease-associated variants may help in prioritizing GWAS signals to loci involved in disease biology (e.g. as we saw for histone deacetlyases) and, ultimately, the aging processes driving tissue decline and eventual manifestation of aging-associated disease.

3.4: Materials and Methods

Accessibility Datasets:

DNase-I hypersensitivity datasets were obtained from ENCODE(Davis et al., 2018) for eight different fetal and adult tissues (adrenal gland, brain, heart, lung, muscle, skin, spleen and stomach) (see Supplemental Table S3.1 for accessions and metadata). Raw data was processed as described in the Supplemental Methods, with called open-chromatin regions consolidated across replicates and tissues in order to define a final set of reproducible regions. This aggregated set of peaks was then used to assess both pan-tissue, as well as per-tissue, accessibility changes between fetal and adult datasets using the limma package (version 3.46) in R(R Development Core Team, 2008; Ritchie et al., 2015). Differentially-accessible regions were defined using a Benjamini-Hochberg FDR(Benjamini and Hochberg, 1995) cutoff of < 0.05.

Adult DNase samples were further stratified in order to define age-altered chromatin regions, splitting samples used in the above analysis into those individuals younger than 50 (‘young-adult’) and those older than 50 (‘old-adult’), this age representing a roughly equal split of sample numbers. Not all tissues used in

the initial fetal/adult comparison were represented in these age-stratified sets – thus we restricted the tissue comparisons to brain, heart, lung, muscle and stomach tissues. A similar computational method as that used in defining fetal- and adult-biased regions was applied here (see Appendix C). We compared accessibility changes between young and old-adult samples within those regions exhibiting fetal/adult biases, defining young-biased and old-biased regions (again, using an FDR cutoff of < 0.05).

Promoter accessibility change:

All hg19 Refseq gene TSS were obtained from the UCSC Genome Browser(Karolchik et al., 2004) and padded 1kb up/downstream to define promoter regions. For each promoter region, DNase read coverage was compared between adult and fetal samples, with resulting data processed using a similar differential-accessibility method as that used above (see Appendix C). Significantly differentially-accessible promoters were defined using an FDR cut-off of 0.05. As an additional, more stringent analysis, we also defined differentially-accessible promoters based on intersections with the above defined region sets (see Appendix C).

Promoter capture datasets: Promoter-capture data was obtained from Jung et al., 2019(Jung et al., 2019); this dataset was generated from promoter-capture assays across a number of different tissues and cell-types. Given our pan-tissue approach, we considered all data (with the exception of OV2, as we excluded sex-specific tissues from all previous obtained datasets). To generate a set of genomic regions which show evidence of contacting gene promoters, we filtered interacting regions to those which contacted their respective promoters in at least two different tissues/cell-types. This moderate filter was used to exclude those regions for which interactions appear to be exclusive to one dataset, while allowing for regions that do not show such exclusivity.

Gene-set enrichment analyses: Gene sets generated in our analyses were tested for enrichment in different GO Biological Process terms using the ‘enrichGO’ function from the clusterProfiler(Yu et al., 2012a) package version 3.16.1, with semantically-similar GO terms collapsed and significantly-enriched terms defined as adjusted p-value < 0.05.

ENCODE RNA-seq datasets: Processed per-gene quantification files, as generated by the ENCODE pipeline were obtained from the ENCODE web portal(Davis et al., 2018) (see Table S3.2 for file accessions and metadata). Given the limited availability of adult tissue samples for use in differential-expression analysis, we instead defined a less-stringent method to identify broad changes in gene expression which demonstrate consistency across tissues (see Appendix C).

GTEx RNA-seq processing: Processed RNA-seq quantification files were obtained from the GTEx web portal(GTEx Consortium, 2013) for the following tissues (matching the above young/old-age accessibility comparison): brain (Brain - Cerebellum), heart (Heart – Left Ventricle), lung (Lung), muscle (Muscle - Skeletal) and stomach (Stomach). Similar to the processing performed in Benayoun et al.(Benayoun et al., 2019), we applied quality filters to remove lowly-expressed and non-coding genes, and subsequently used the same definitions of 'young-age' and 'old-age' (as in the above analyses) to calculate differential expression using limma-voom (see Appendix C).

Human sequence variation datasets: Variation data from the 1000 Genomes Project phase 3 (1KGP)(Gibbs et al., 2015) (n = 2504 individuals) in .vcf.gz format was obtained and intersected with our region sets using tabix(Li, 2011) (version 1.9) to obtain variants occurring within these altered-accessibility regions. Common variants were defined using a minor allele frequency (MAF) threshold of >= 0.05. These sets of intersected variants were subsequently used to compare sequence variation across region sets, as well as comparing region-intersected variation to genomic backgrounds and feature sets (see Appendix C).

GWAS summary statistics data: To define a set of aging-associated diseases for use in our analyses, we first used broadly-defined categories as described in Chang et al., 2019(Chang et al., 2019). This study described 92 age-related diseases grouped into broader disease categories based on analyses of large-scale demographic datasets. We took these classifications and manually extracted relevant GWAS phenotypes assessed by the UK Biobanks study(Sudlow et al., 2015), obtaining pre-processed summary statistics for these phenotypes provided by the Neale lab(Bulik-Sullivan et al., 2015) (https://nealelab.github.io/UKBB_ldsc/downloads.html). These data were subsequently utilized across several bioinformatic analyses (see Appendix C).

Additional computational methods, including implementations of statistical tests described in the Results, are described in detail in Supplemental File 2. The datasets supporting the findings of this study are publicly-available – accession codes and URLs are provided in Supplemental File 2 and Supplemental Tables.

3.5: Acknowledgments

We thank Dr. Anne Brunet for meaningful discussions relating to concepts of aging, development and epigenetics.

3.6: Author Contributions

DR: Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Methodology, Manuscript Writing; TDC: Conceptualization, Supervision, Funding acquisition, Investigation, Manuscript Writing, Project administration.

Chapter IV: The importance of protein structure, function, and interactions in convergent sequence evolution of long-lived species

4.1: Introduction

One universal feature of organisms on Earth is that of expiration – a combination of extrinsic and intrinsic factors which contributes to the progressive increase in mortality risk and, consequently, decreased survivorship over time. Comparing mortality curves for species across a diverse range of groups(Croft et al., 2015; Jones et al., 2014) indicates the presence of considerable variation in survivorship trajectories, and overall longevity. These result from evolutionary interactions between a species' physiology and its environment. Mammals, broadly, share similar patterns of age-dependent mortality and senescence phenotypes(Magalhães and Toussaint, 2002), suggesting that ancient selection acted on biological processes mediating functional decline. However, within mammals there is also substantial variation in senescence rates(Austad, 1997), suggesting that more recent evolutionary modifications may have occurred on this background of ancient selection. This is particularly true in instances where species exhibit substantially increased longevity relative to closely related taxa. For example, genomic studies of long-lived bats(Seim et al., 2013), whales(Keane et al., 2015), and naked mole-rats(Lewis et al., 2016) have revealed key genes/pathways which could be causally-linked with these lifespan extensions, corresponding with several 'hallmarks of aging'(López-Otín et al., 2013), including insulin/IGF signaling(Junnila et al., 2013), DNA repair(MacRae et al., 2015) and proteostasis(Draceni and Pechmann, 2019; Pan and Finkel, 2017), among others.

Of note, longevity studies have also highlighted the possibility that changes to epigenetic processes may act to alter aging programs in long-lived species(Horvath et al., 2021a; Keane et al., 2015; Wilkinson et al., 2021). Perhaps the best-characterized link between epigenetics and aging is DNA methylation patterns, which exhibit an extraordinarily consistent age-dependence(Horvath, 2013). This has been suggested to represent a biological 'clock' that is reflective of development and maintenance processes

throughout life whose pacing can reflect disease states and risk(Horvath and Raj, 2018a). Notably, this clock behaviour has been demonstrated in other species(Horvath, 2013; Horvath et al., 2021b; Prado et al., 2021), suggesting that the underlying biological basis for this phenomenon is conserved within mammals(Horvath et al., 2021b). DNA methylation patterns have been complemented with substantial research establishing the links between additional epigenetic features (e.g. chromatin states) and aging processes(Booth and Brunet, 2016; López-Otín et al., 2013; Moskalev et al., 2014).These studies have shown that epigenetic states shift with age in both a locus-specific and genome-wide manner, that these shifts can be causally linked with altered gene regulation (e.g., the deleterious expression of retrotransposons(Villeponteau, 1997; Wood et al., 2016)), and that mitigation and/or reversal of these shifts can ameliorate aging phenotypes and prolong lifespan(Benayoun et al., 2015; Booth and Brunet, 2016; Sen et al., 2016b; Zhang et al., 2020b). These epigenetic shifts are not limited to gene promoters/bodies to regulate expression, but also extend to more distal *cis*-regulatory elements (Booth and Brunet, 2016) containing disease-associated variants whose effects may only manifest upon age-associated remodeling(Richard and Capellini, 2021). Given the myriad of mechanistic ties between epigenetics and aging, it may be the case that proteins involved in epigenetic maintenance and remodeling are subject to evolutionary pressures for longevity, similar to how their target loci may have evolved (Lowdon et al., 2016; Prendergast et al., 2014).

The recurrent patterns observed in studies of notably long-lived species suggests that a comparative, evolutionary framework may help to illuminate additional genes/pathways involved in the aging process. Given a phylogenetic tree of species, patterns of stratified genetic changes can be used to identify genes with an impact on aging biology, and thus could be subject to selection for these effects. These findings are then often validated based on existing knowledge of aging/longevity-related processes(Farré et al., 2021b; Jobson et al., 2010; Kowalczyk et al., 2020; Li and de Magalhães, 2011; Muntané et al., 2018; Treaster et al., 2021). More recently, studies have moved from the gene-level down to individual non-synonymous protein-coding changes, looking for mutations which may modify longevity within primates(Muntané et al., 2018), and across mammals(Farré et al., 2021a). In particular, Farre et al., (Farré et al., 2021a) identified thousands of amino-acid (AA) substitutions associated with lifespan which are fixed or approach fixation in modern-day humans and are enriched in longevity-associated pathways

such as immune signaling(Ferrucci and Fabbri, 2018; Fulop et al., 2014) and coagulation(Kanapuru and Ershler, 2009). This study also showed that these changes could serve to improve the stability of these modified proteins in long-lived species. Given this resolution at the level of AA, a more detailed analysis of these modifications may help to elucidate the molecular targets (e.g. protein activities, complexes, etc.) upon which selection for longevity may have converged across long-lived species.

Much research effort has gone towards the study of non-synonymous mutations from the perspective of human variation and disease(Cargill et al., 1999; Cooper and Shendure, 2011; Lek et al., 2016; Ng and Henikoff, 2006). Computational approaches have been developed to predict the functional consequences of AA changes using a combination of evolutionary conservation (i.e., conserved sites are more likely to be functionally relevant)(Ng and Henikoff, 2006) and protein structure(Bromberg and Rost, 2007; Chasman and Adams, 2001; Sunyaev et al., 2001) information. In addition to computational predictions, functional assays have been used to test the effects of individual mutations *in-vitro* or *in-vivo*, while for key proteins (e.g., p53(Kato et al., 2003)) exhaustive scans of many possible mutational effects have been tested, an approach which has benefitted from modern high-throughput advancements(Fowler and Fields, 2014; Starita et al., 2017). Mutations can impact proteins in a number of ways, including changes to protein stability(Yue et al., 2005), modifying functional domains(Miller et al., 2015; F. Yang et al., 2015), and impacting interaction interfaces between pairs of proteins(Jubb et al., 2017a; Kumar et al., 2015; Yates and Sternberg, 2013). These effects can be modelled and tested on an individual protein basis, for example, considering human polymorphisms in genes acting as known disease risk-factors(Dakal et al., 2017), or recurrent mutations in key oncogenic genes(H. Liu et al., 2020). Furthermore, the field of cancer genomics has investigated patterns in mutation occurrence, highlighting key genes and pathways in cancer development(Campbell et al., 2020; Martínez-Jiménez et al., 2020), as well as identifying broader trends in mutations at the level of functional domains(Miller et al., 2015).

In this study, we consider the properties of these convergent amino acid substitutions in the context of protein structure and function, looking both for general trends which may suggest broader-scale evolutionary pressures on particular biological mechanisms (e.g. protein activities) as well as specific

examples wherein the functional consequences of these changes may be intimated. We make several key findings; (1) proteins subject to modification are over-represented in certain activities, (2) modified residues can cluster with, and in some cases directly overlap, mutations observed in cancer samples, and (3) modified residues are over-represented in protein-protein interaction interfaces, possibly impacting complex formation.

4.2: Results

4.2.1: CAAS description and gene-set integration:

We considered a set of convergent amino acid substitutions (CAAS) between long-lived and short-lived species, which we defined on the basis of their respective size-adjusted longevity (see Methods, including information on species used). Each CAAS can be defined as falling under one of four possible scenarios (Figure 4.1A); invariant across long-lived and short-lived species (#1), invariant across long-lived, variable across short-lived species (#2), variable across long-lived, invariant across short-lived (#3) and variable across both sets of species (#4). We first asked whether CAAS-containing genes are enriched in particular biological processes or pathways. We combined CAAS from Scenarios #1-3 for statistical purposes, considering them to be broadly different between long/short-lived species. Using this combined set, we found enrichments for terms relating to immune signaling, cell-cycle regulation, and DNA replication (Figure 4.1B), terms which were also enriched (at nominal significance) in individual Scenario subsets (Supplementary Table S4.1). These enrichments are consistent with those previously observed using similar CAAS definitions on lifespan-stratified species(Farré et al., 2021b). Conversely, Scenario #4 CAAS enrichments were dominated by immune-related processes, particularly the adaptive immune system, which may reflect the generally increased rate of evolution occurring in these proteins between species(Shultz and Sackton, 2019) (Supplementary Table S4.1). For all subsequent analyses we considered the combined Scenario #1-#3 CAAS only.

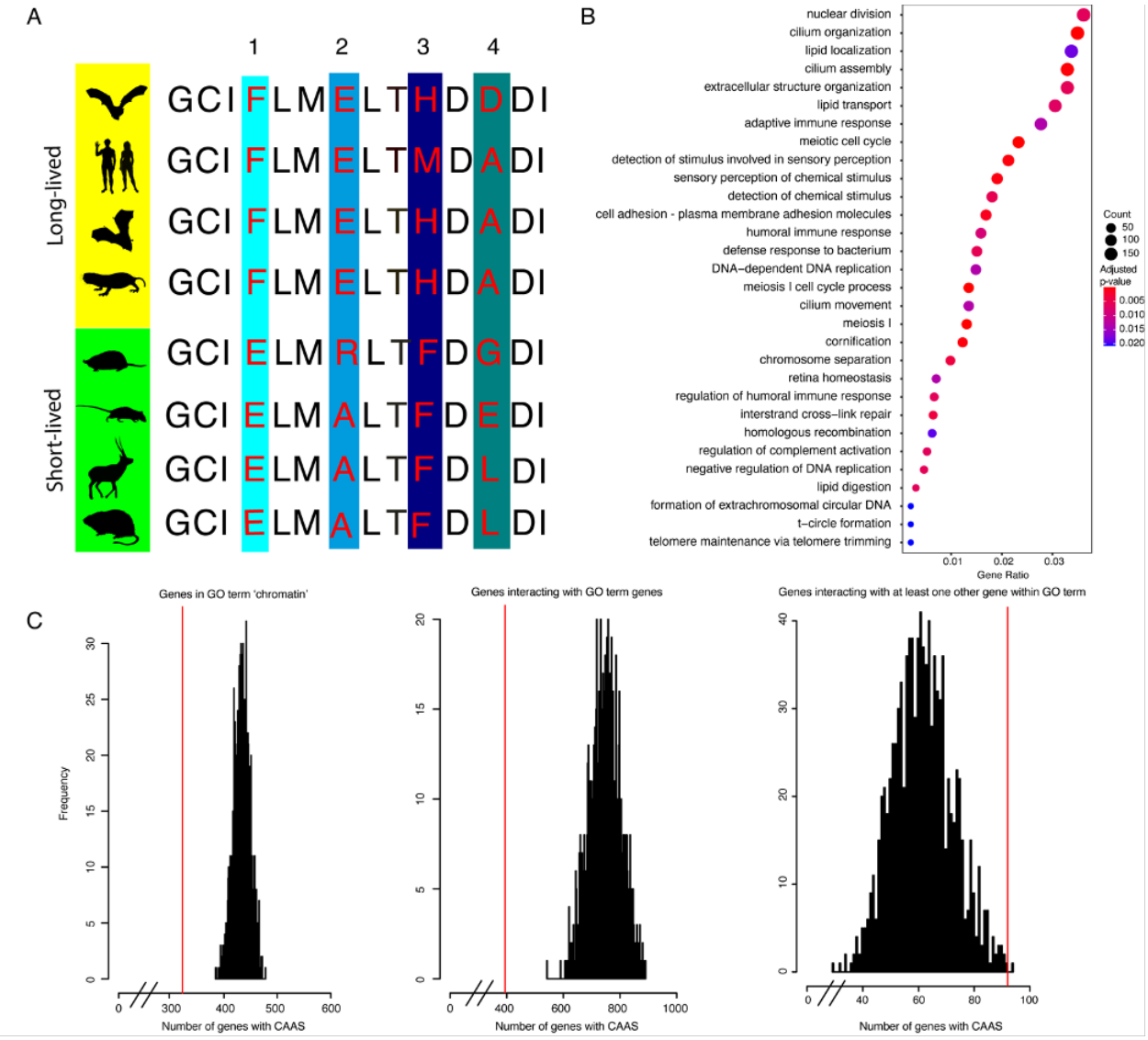


Figure 4.1: Convergent amino-acid substitutions across species. (**A**) Illustrative protein alignment, with CAAS (highlighted) defined under four different Scenarios. (**B**) GO term enrichment for CAAS defined under Scenarios 1-3. (**C**) Biases in CAAS occurrence within the GO term 'chromatin'. (Left): Lower occurrence of CAAS hits in genes of this set. (Middle): Lower occurrence of CAAS hits in genes which interact with genes in this set. (Right) within the GO term, CAAS-hit genes tend to interact with other CAAS-hit genes (Methods). See also Supplemental Table S4.1.

We next evaluated the occurrence of CAAS within genes involved in epigenetic regulation, as modifications to these genes can significantly impact aging(Benayoun et al., 2015; Booth and Brunet,

2016; Sen et al., 2016b; Zhang et al., 2020b). First, we aggregated a number of GO biological process terms related to epigenetic regulation (see Supplementary Table S4.1 for full gene sets) and tested whether these gene sets tended to be biased (in terms of associations with heritable disease risk) for a number of age-associated diseases curated from the UKBiobank datasets(Chang et al., 2019; Richard and Capellini, 2021; Sudlow et al., 2015) (Methods). Indeed, we observed that several epigenetic gene sets tended to have greater-than-expected per-gene rankings across a set of aging-related diseases (Supplementary Table S4.1). As we may similarly expect such associations of genes containing longevity-defined CAAS, we also observed that cross-set rankings of CAAS genes were also elevated (P=1.5e-21, Supplementary Table S4.1). Finally, we looked for the occurrence of CAAS genes within these epigenetic sets (Supplementary Table S4.1) and asked whether these genes were disproportionately associated with disease risk (i.e., when conditioning on the associations of their respective sets). We observed significantly-greater associations for CAAS genes within several terms, including 'heterochromatin' (GO0000792), and 'chromatin binding' (GO0003682) (adjusted P<0.05) (Supplementary Table S4.1). Given these conditioned associations with heritable disease risk, we asked whether our CAAS were enriched in these gene sets. We found that these sets of genes were less likely to contain CAAS than expected by chance (Figure 4.1C, left) (Supplementary Table S4.1).

Epigenetic regulation often occurs through the activity of protein complexes(Medvedeva et al., 2015). We thus considered the possibility that proteins which interact with proteins in a given GO term may also be the targets of CAAS, in that modifying these interacting proteins may impact epigenetic regulation in a similar manner to directly modifying epigenetic genes. Utilizing previously-annotated protein-protein interactions (PPI) from the STRING database(Szklarczyk et al., 2021), we found that interacting proteins were also depleted for CAAS hits (Figure 4.1C, middle) (Supplementary Table S4.1). We next considered the occurrence of CAAS within groups of interacting proteins. Considering all CAAS-hit proteins within a GO term, and all proteins interacting with these hits, we asked whether (a) the interacting proteins also had CAAS, and (b) whether those interacting proteins also fell within the same GO term. We found that these internal pairs of CAAS-hit genes occur at a higher frequency than expected across the majority of terms (with the exception of chromatin components) (Figure 4.1C, right) (see Supplementary Table S4.1).

However, when we performed a similar set of tests against randomly-selected GO terms (see Methods) we observed a similar pattern. Specifically, we observed a bias against CAAS occurrence within a term and their interactors, but significant over-representation of CAAS genes interacting within sets (Supplementary Table S4.1). That these patterns are not exclusive to epigenetic sets may suggest that, generally, when such cross-species protein-coding changes occur, they tend to co-occur between sets of functionally-related interacting proteins, a finding that has important implications for our understanding of how selection may act on longevity (see Discussion).

This could be driven by several possible scenarios; for example, CAAS-hit genes could be biased in their protein domain composition, such that PPI are more common between hits involved in coordinated biological processes (e.g., DNA damage sensing and repair). Another possibility is that CAAS directly modify the protein-protein interface of interacting pairs. We explore each scenario in more detail below (see Supplementary Figure S4.1 for results overview).

4.2.2: Domain-level integration

If the occurrence of multiple CAAS hits within a set is indicative of effects on longevity, and given that many of these proteins exert their biological effects via protein domains, we then considered (a) the occurrence of protein domains within CAAS-hit genes, and (b) where CAAS occur on individual proteins. Focusing first on epigenetic GO terms, we found several domains for which CAAS-hit genes were enriched, even after conditioning on inherent enrichment of terms for certain domains. Figure 4.2A illustrates these results for the GO cellular-component term ‘chromatin’, for which CAAS-hit genes are over-represented across several domains, including the PAS and SRC-1 domains (see Discussion). More generally, we found several protein domains which were over-represented in our global set of CAAS genes (Supplemental Table S4.2); these also significantly overlapped with domains enriched in aging-associated genes from the GenAge database(Tacutu et al., 2018), including epigenetic domains such as histone deacetylase domains, bromodomains, and helicases, along with domains involved in genome maintenance and repair (Figure 4.2B, Supplemental Table S4.2).

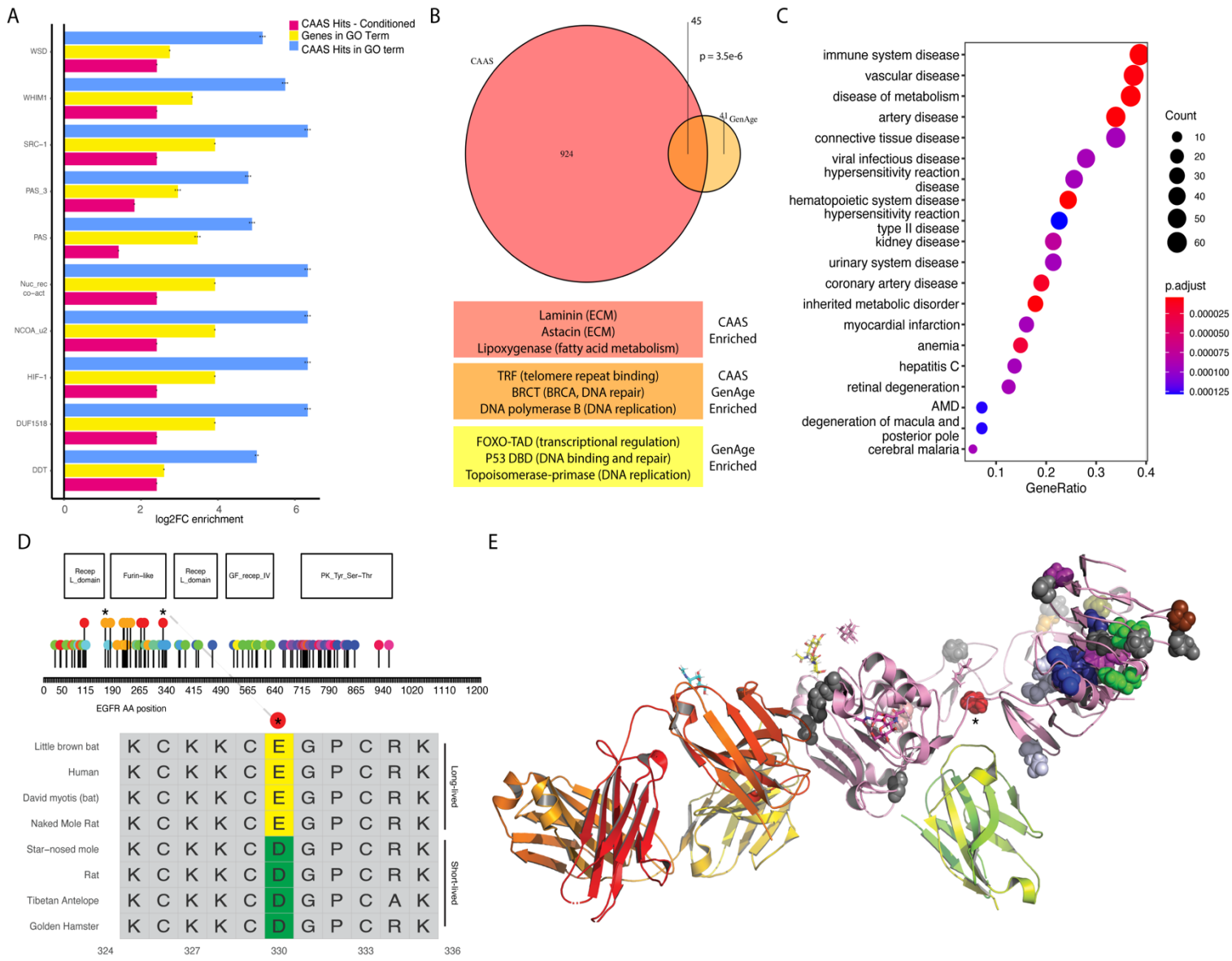


Figure 4.2: CAAS occurrence in functional domains and spatial clustering. (**A**) GO chromatin components have a greater occurrence of certain protein domains relative to the global occurrence of these domains (green bars). CAAS-hit chromatin components similarly have increased occurrence of these domains (blue), and when conditioned on the initial domain enrichments (orange). See Supplemental Table S4.2. (**B**) (Top) Venn diagram showing overlap between protein domains enriched in CAAS-hit genes (red) and GenAge genes (yellow). Relative to all unique PFAM domains occurring in CAAS genes (n = 3342) there is significant overlap between these sets. (Bottom) Enriched domains occurring in indicated sets, with parentheses indicating their broad functionality. (**C**) DOSE enrichments for genes with CAAS hits within mutational clusters. (**D**) (Top): Distribution of clustered mutations across *EGFR*. Different colored dots

represent mutations belonging to different clusters. Taller dots denote clusters containing CAAS, which are indicated with asterisk. Red dots denote the cluster highlighted in (E). (Bottom): Protein alignment indicating the D (long-lived) to E (short-lived) CAAS occurring within the indicated EGFR mutational cluster. (**E**) Protein crystal structure for EGFR, red atoms and asterisk indicating the position of the CAAS falling within the Furin-like domain. Differently-colored atoms indicate positions of other cancer-related mutations clustering within the EGFR protein.

Next, we looked for instances in which the potential targeting of longevity-relevant genes could be occurring through direct AA changes within specific functional protein domains (e.g., those we observed to be enriched in CAAS-hit genes). We found that the majority of CAAS-hit protein domains tended to occur at lower rates than the background occurrence of these domains across the proteome (Supplementary Table S4.2), the same evident for broader domain-clan groups. However, we found two exceptions to these trends: Serpin domains and the 'Aha1_BPI' clan occur more frequently in CAAS genes than the background occurrence of these domains (see Discussion).

We next examined those domains which showed biased occurrence within CAAS-hit genes of our assembled GO terms, and asked whether any CAAS directly hit these domains (Supplementary Table S4.2). We found the proportion of CAAS-hit genes with CAAS directly overlapping domains of interest varied across sets and domains. For example, of nine 'chromatin binding' CAAS-hit genes containing Helicase C domains, only one (WRN) had a CAAS directly within the domain. This domain has been implicated in surveys of WRN patient mutations(Huang et al., 2006) and is required for WRN accumulation at double strand breaks (DSBs)(Lan et al., 2005). Conversely, two of the three *NCOA* genes containing CAAS had these hits fall within their SRC-1 domains (involved in steroid receptor binding(Nolte et al., 1998)).

We also looked for instances in which domain-hitting CAAS are predicted to impact protein structure and/or function. Using the dbNSFP database of annotated coding mutations(X. Liu et al., 2020) we identified 279 domain-hit CAAS with predicted deleterious effects based on protein structure

(Supplementary Table S4.2) (Methods). The most consistently-predicted effect CAAS was within the *CFTR* gene, where an A1006V mutation within the ABC transporter domain may be associated with cystic fibrosis(George Priya Doss et al., 2008; Mercier et al., 1994). We also found an R1217K mutation in the BRCA2 domain of *BRCA2* of unknown clinical significance (see Supplementary Table S4.2).

4.2.3 Integration at the structural level - 3D clustering:

Another key factor in a protein's function is its three-dimensional configuration. As a complementary analysis we employed mutational clustering, an approach often used in cancer genomics to identify spatial 'hotspots' where tumor mutations tend to cluster(Niu et al., 2016b). By treating CAAS in a similar manner to somatic cancer mutations, we may identify important functionalities, especially since there is substantial evidence for overlap between biological mechanisms underlying both aging and cancer(Aunan et al., 2017; De Magalhães, 2013). Thus, we combined our set of CAAS with a previously published set of somatic non-truncating cancer mutations based on the TCGA(Campbell et al., 2020; Niu et al., 2016b) and applied the HotSpot3D(Niu et al., 2016a) algorithm to identify clusters of residue modifications (Methods). We identified 1876 clusters across 735 individual genes - and also yielded clusters in 40 additional genes not identified with cancer data alone (Supplementary Table S4.3). Our approach did not lead to substantial loss of existing gene hits (e.g., due to the addition of noise). This suggests that combining these data improves our power to detect spatial clusters, rather than obfuscating spatial patterns (Supplementary Table S4.3).

Of these 1876 clusters, 222 included CAAS across 218 unique genes. Specifically, this set of genes was largely enriched for biological processes related to immunity, such as cytokine regulation and innate/adaptive responses (Supplementary Table S4.3). Similar enrichment results were observed when we separated genes based on the CAAS scenario (Supplementary Table S4.3). When we tested these cluster-containing genes for enriched disease annotations (see Methods) we saw enrichments for several different disease categories. These included diseases related to immunity, along with several disease categories for which age is a known risk factor - including cardiovascular, metabolic, and arthritic disease (Figure 4.2C, Supplementary Table S4.3). Several CAAS-containing genes were associated with multiple

different diseases; e.g., *VEGFA* (Supplementary Figure S4.2), a growth factor known to be beneficial in aging tissues(Lähteenvuo and Rosenzweig, 2012) while also being associated with tumor angiogenesis(Carmeliet and P, 2005). Such multi-disease associations are expected for certain genes, particularly those with important biological roles and for which mutations may impact their functional domains.

As an illustrative example, we observed a cluster of four mutations within EGFR of which a CAAS formed the centroid (which was not detected using TCGA data alone). Three of these four mutations fell within a Furin-like domain while the fourth fell within a Receptor L-domain on the protein (Figure 4.2D). EGFR mutations are more common within this Furin-like domain, with several known or thought to be oncogenic(H. Liu et al., 2020). It may be that these clustered mutations similarly have oncogenic effects including, by association, our CAAS hit. This hit is one in which long-lived species have a fixed 'E' at this position, relative to a short-lived 'D' (Figure 4.2D-E) (see also below).

We next asked whether CAAS-containing mutational clusters hit certain domains at increased frequency (Methods). We observed 14 domains (FDR < 0.05) for which this was true (Supplementary Table S4.3). One of the strongest enrichments was observed for the "Matrixin" domain which occurred in clusters of five matrix-metalloprotease genes (*MMP7,8,9,10,14*). Another notable enriched hit was the p450 domain, for which we observed clustered hits in five different CAAS genes (*CYP19A1, CYP27A1, CYP2A13, CYP7A1, PTGIS*). We also observed a three-fold enrichment for the Serpin domain - which was expected given the general enrichment of CAAS directly hitting this domain.

We next looked for the occurrence of CAAS-containing mutational clusters within our epigenetic gene sets. We identified 13 genes (Supplementary Table S4.3), including the aforementioned *EGFR*. For example, we observed a cluster of mutations within the POLA1 protein which implicates two protein domains with distinct functions (polymerase elongation and exonuclease activity), both containing CAAS. Notably, as these domains were also identified in our analyses of protein-protein interactions (PPI), we thus considered the potential for mutational clusters to occur within PPI interfaces (see below).

Our ‘function-by-association’ analysis led us to also consider that CAAS may not only cluster with cancer mutations, but directly overlap at the same residues. Globally, we identified 1101 CAAS across 896 unique genes which directly overlapped with cancer mutations, though this overlap was not significantly more than expected from randomly-shuffled sets of CAAS (p = 0.31, Supplementary Table S4.3). A similar degree of overlap was observed when using an expanded set of mutations from the TCGA (Supplementary Table S4.3). Of these genes containing overlaps, 36 had clustered mutations including overlapping residues. Several of these genes have known roles in cancer and aging, including DNA repair genes like *PARP14*(CM et al., 2015) and *BRCA1*(Cao et al., 2003), and pro-proliferation proteins like *VEGFA*, *PRLR*(Lopez-Pulido et al., 2013), and *EGFR*(Sigismund et al., 2018). A number of these overlapping CAAS also hit protein domains, such as a CAAS within a cluster of five mutations in VEGFA which hits the PDGF/VEGF domain - such instances are presented in Supplemental Table 4.3.

4.2.4: CAAS occurring within protein-protein interaction domains:

The major finding that within a given GO term CAAS hits tended to occur between interacting pairs of genes suggests that CAAS are not randomly distributed across genes - but instead may reflect physical relationships between proteins, i.e., protein-protein interactions (PPI). Given the potential evolutionary targets presented by PPI(Choi et al., 2009), we made use of the ProtCID(Xu and Dunbrack, 2020) database which aggregates PPI at the protein-domain level. We looked for any instances of PPI occurring between our set of CAAS-hit proteins and observed 39 interaction pairs (across 32 unique gene pairs) for which CAAS fall within both protein domains, significantly more than expected by chance (p = 6.4e-12, Supplemental Table S4.4, Methods). These consisted of several examples with relevance to aging and cancer biology, including PLAT- and PLAU-SERPINE1 (involved in coagulation), several interactions between complement proteins, CD47-SIRPa(Willingham et al., 2012), and DNA damage response genes (e.g., FANCM-FAAP24)(Huang et al., 2010). We found seven instances where shared CAAS hits implicated genes from epigenetic gene sets (Supplemental Table S4.4). For example, multiple CAAS occur within similar interaction domains of POLA1, POLA2, and PRIM1, all of which are key components of the DNA polymerase alpha – primase complex(Weiner et al., 2007) (Figure 4.3A). We also reasoned

that a CAAS hit could occur in one domain of an interacting pair - potentially impacting how proteins interact in a unilateral fashion. Thus, we scanned our CAAS set, and again found a greater occurrence of these instances than expected by chance ($p = 1e-3$, Supplemental Table S4.4). Overall, these results suggest that a significant portion of CAAS may preferentially occur within interaction domains, possibly acting to improve, maintain, or disrupt the interactions between binding partners.

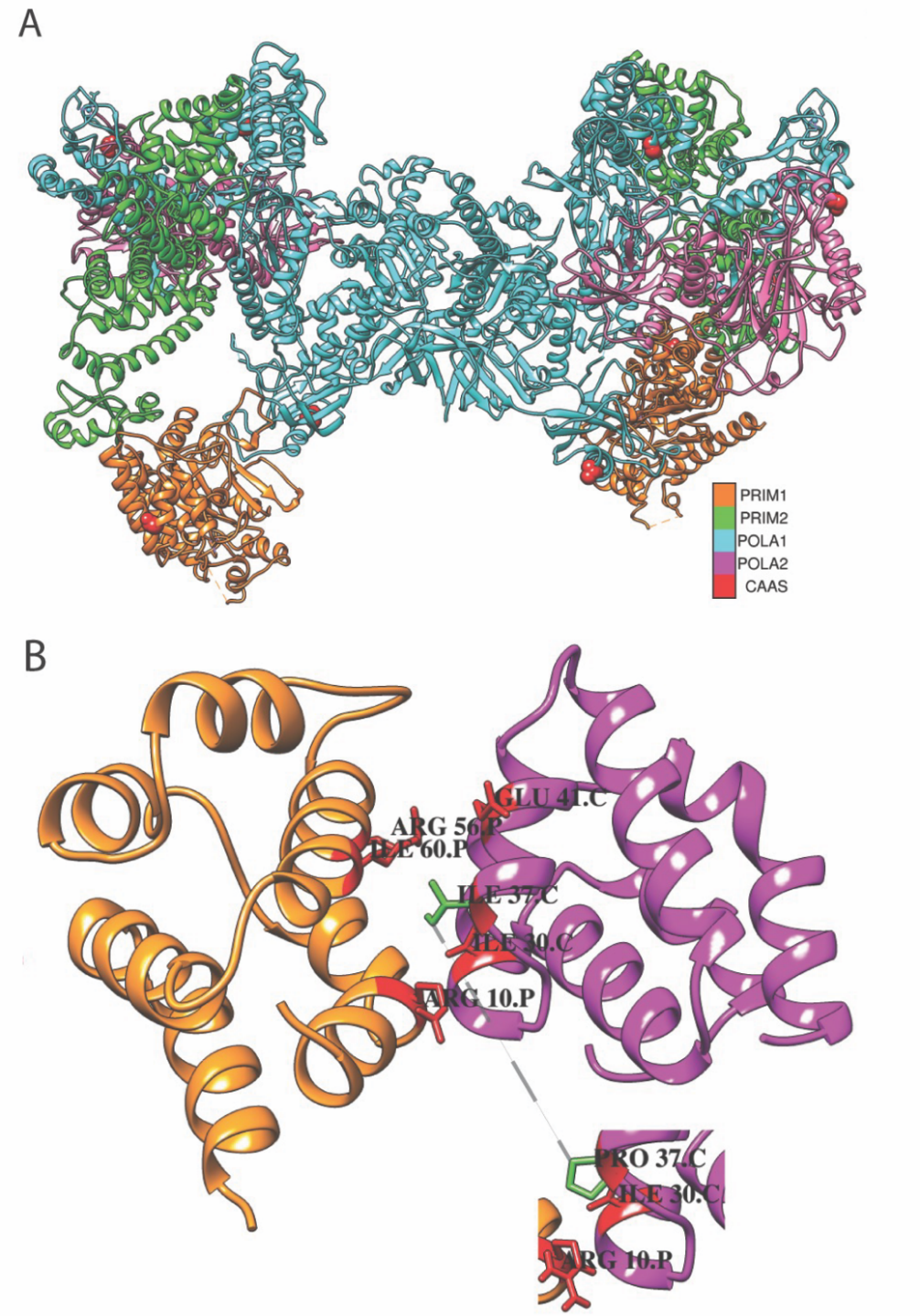


Figure 4.3: CAAS occurring at PPI interfaces. (**A**) Multiple CAAS occurring within the human DNA polymerase-primase complex. Ribbon colors as indicated, CAAS occurring in each chain are highlighted as red balls. PDBID: 5exr. (**B**) CASP9 - APAF1 interaction. Key residues involved in CASP9 (orange)

recruitment by APAF1 (purple) indicated(Qin et al., 1999a). (Left): long-lived residue at position 37 of APAF1 CARD domain. (Inset): short-lived residue at this position, which is a predicted unfavorable mutation ($\Delta\Delta G$ = 2.4 kcal mol^{-1} (Huang et al., 2020)). I37P residue side-chain shown in green.

To prioritize CAAS hits for potential functional follow-up studies, we applied a more stringent criteria by first defining residues as being directly involved in the PPI interface (Methods), and second looking for CAAS residues that fulfill these criteria. We found seven protein pairs in which CAAS residues were directly implicated at the PPI interface, observing significant enrichments despite this increased stringency (Supplemental Table S4.4). For example, APAF1 is a key component of the 'apoptosome' complex, which regulates apoptosis through the activation of the initiator caspase CASP9(Bao and Shi, 2006; Yoshida et al., 1998); we observed a CAAS interacting residue within the caspase recruitment domain (see Figure 4.3B, Discussion). Interestingly, we also observed multiple instances of ligand-receptor interactions for which CAAS are potential interface residues, including IL17/ IL17RA (Supplemental Figure S4.3), IL20/IL20RA, and transferrin/transferrin receptor (Supplemental Table S4.4). Ligand-receptor co-evolution is a well-studied phenomenon(Moyle et al., 1994), so instances of such concomitant modification of pairs identified with our CAAS sets may be expected.

Given our analyses on CAAS and interaction domains between pairs of different proteins, we also considered the possibility that CAAS may impact homodimer formation, or domain interactions within a single protein, again noting an enriched occurrence of such instances both at the levels of interaction domains and interface residues (see Supplemental Table S4.4).

We next looked for previous experimental data (del-Toro et al., 2019; Jankauskaite et al., 2019) which describes CAAS effects on PPI in the IntAct database(del-Toro et al., 2019); we found 10 interaction pairs for which mutations at CAAS residues were described; of these, three had possible mutational effects (see Supplemental Table S4.4). Most notable was a Q67K mutation at a critical binding residue in SIRPA, whose mutation disrupts CD47 binding(Liu et al., 2007). Using the SSIPE(Huang et al., 2020) server (see Methods) to predict the disruptive effects of mutations on dimeric complexes, we found that a CAAS

hitting an interaction domain of the aforementioned APAF1-CASP9 complex was a predicted strongly-unfavorable mutation ($\Delta\Delta G \geq 1.5$ kcal mol^{-1}(Huang et al., 2020)), leading to a substantial decrease in binding affinity (the short-lived residue yielding a predicted $\Delta\Delta G = 2.4$ kcal mol^{-1}) (Figure 4.3B). This further supports the possibility that a CAAS in this complex could alter apoptotic signaling cascades (see Discussion).

4.2.5: Spatial clustering and protein interactions

In our characterization of CAAS occurrence within both mutational clusters and protein-protein domain interactions, we observed several instances where genes were identified in both analyses. Specifically, we observed 28 interfaces for which one interaction domain intersected a mutational cluster (Supplemental Table S4.4). Of these, nine had a mutational cluster intersecting both domains. We found that a majority (7 of 9) of these instances were shared with our above CAAS-level analysis. A notable example is that we observed two mutational clusters occurring within the interaction domains of POLA1 and POLA2 (Supplemental Table S4.4). This congruence may be expected given that interaction domains may also have catalytic activity (e.g., exonuclease activity in POLA1), and that mutational clusters can preferentially occur within some functional protein domains (see above). When considering instances in which one interaction domain contains a mutational cluster, we identified 32 additional interfaces which were not defined as having CAAS falling directly within either domain (Supplemental Table S4.4). These additional hits expand our set of interfaces by including CAAS that may be spatially adjacent to an interface domain, without requiring that they fall within domains.

4.2.6: Functional impacts of CAAS on protein function *in-vitro*:

The advent of ‘deep mutational scan’ techniques has allowed for high-throughput testing of AA substitution effects. These studies generate large numbers of substitutions *in-vitro*, apply some selective factor (e.g. drug treatment) and assay biochemical metrics, most often DNA sequencing to test effects on cell survival/proliferation(Fowler and Fields, 2014; Starita et al., 2017). We looked for previous protein studies for which our particular CAAS modifications (i.e., long-lived to short-lived AA substitutions) had been made and assayed for some biochemical trait, taking advantage of a database for multiplex assays

of variant effects (MAVEdb(Esposito et al., 2019)). Despite the practical limitations of these studies(Starita et al., 2017) we found five studies (on p53, CCR5, MSH2, LDLRAP1, HMGCR) for which the observed long/short substitution had been assayed. Giacomelli et al.(Giacomelli et al., 2018) performed a mutational scan containing all p53 missense and nonsense mutations, exposing cells to etoposide and subsequently sequencing genomic DNA as a metric of a particular mutation's effects on survival. Our CAAS, a glutamine-to-alanine substitution at residue 51, was associated with positive phenotypic selection in cells following treatment (Figure 4.4A).

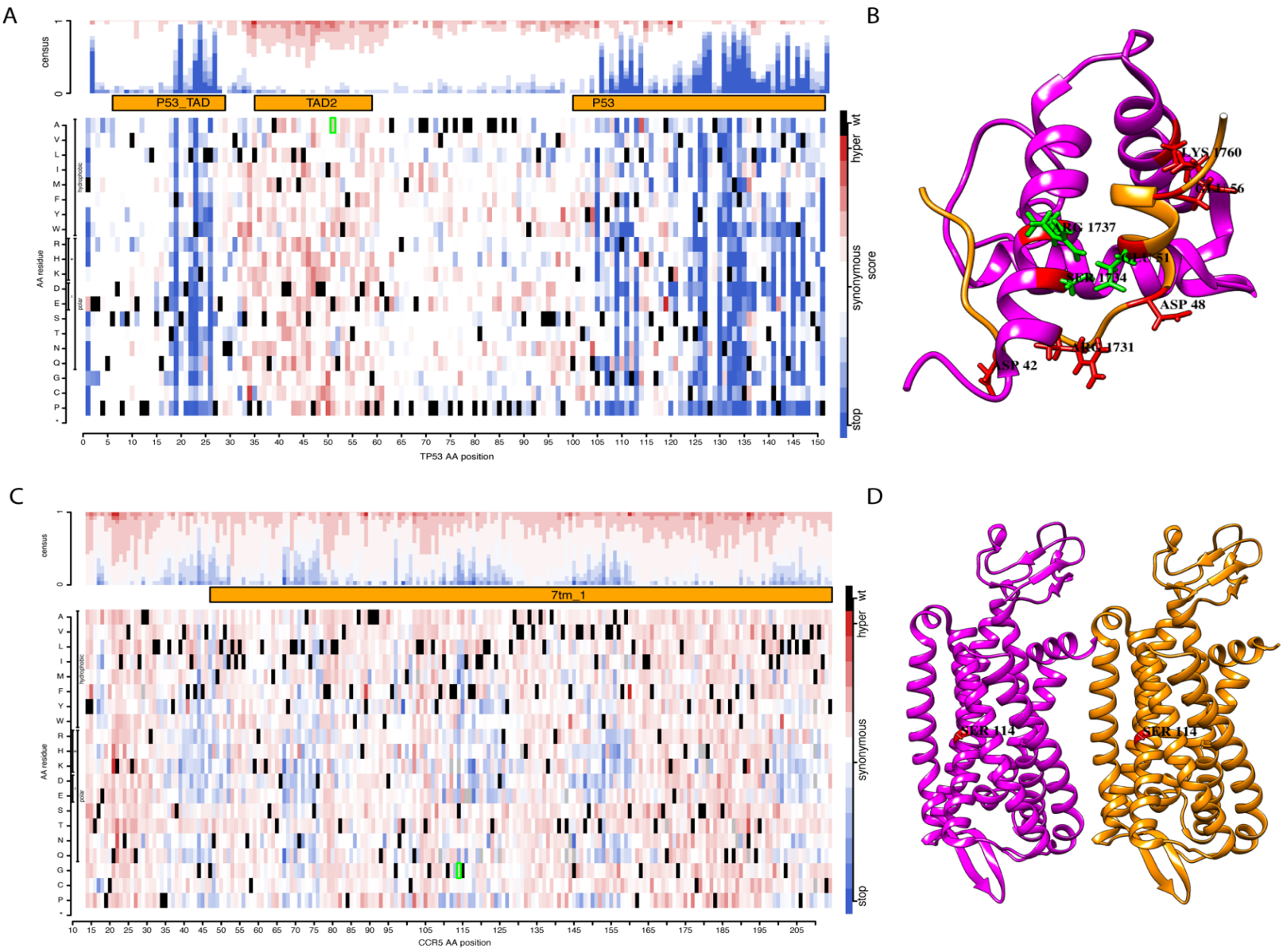


Figure 4.4: CAAS characterized in mutational scan studies. (**A**) Effects of substitutions engineered in TP53, as measured by variant enrichment following etoposide treatment in TP53-null A549 cells. Red/white/blue color scale refers to effects of termination codon (blue), synonymous mutations (white) and hyper-active mutations (red). The CAAS of interest, E51A, is indicated in green box within the Tad2 domain. Black boxes refer to the reference residue at a given position. Top graph indicates consensus effects of mutations (i.e., the predominate effects of mutations at this residue). (**B**) Protein structure of p53 (orange) in complex with Taz2 domain of p300 (pink), with key residues involved in stabilizing the complex(Miller Jenkins et al., 2015) labelled along with side-chains (red). Glu51 on p53 (CAAS-modified

residue), and the p300 residues with which it interacts, are highlighted in green. (**C**) Effects of substitutions engineered into CCR5, as measured by 2D7 antibody binding affinity in Expi293F cells. Red/white/blue color scale refers to log2 (mutant/wt) enrichment values centered at zero. Negative/positive values indicate decreased/increased binding affinity, respectively. (**D**) Protein structure of CCR5, showing the CAAS residue (S114G) within the 7tm_1 transmembrane domain. CCR5 homodimer complex shown (pink/orange).

Another example is CCR5, a key chemokine receptor involved in regulating effector cells of both the innate and adaptive immune system, with evidence to suggest a role in 'inflammaging' and age-related diseases such as atherosclerosis(Balistreri et al., 2007). Heredia et al.(Heredia et al., 2018) performed a mutational scan of CCR5, assaying the effects of missense mutations on CCR5 binding affinity to the monoclonal antibody 2D7, a blocker of HIV-I entry which competitively binds CCR5(Lee et al., 1999). A CAAS falling within one of the transmembrane domains of CCR5, a serine-to-glycine substitution at residue 114, leads to a slightly increased binding affinity for the 2D7 antibody (Figure 4.4B). Additional examples are shown in Supplemental Figure S4.4.

4.3: Discussion

Similar to how recurrent cancer mutations can illuminate key pathways involved in tumorigenesis and growth(Reyna et al., 2020), studying trends in convergent genomic changes in long-lived species has illuminated key aging pathways(Kowalczyk et al., 2020; Lambert and Portfors, 2017; Li and de Magalhães, 2011; Muntané et al., 2018). In this study, we sought to extend these analyses by considering convergent amino-acid substitutions (CAAS), as defined using species stratified based on longevity(Farré et al., 2021b) at the level of protein structure and function. As expected based on previous studies, we observed enrichments for CAAS in GO terms related to cell-cycling, DNA replication, and immune system biology. Given the known associations between epigenetics and aging(Horvath et al., 2021a; Keane et al., 2015; Wilkinson et al., 2021), we aggregated genes associated with epigenetic processes and identified several CAAS occurring within these sets (Supplemental Table S4.1). We also observed that CAAS-containing genes tended to be more associated with heritable disease risk across several aging-

associated diseases (Supplemental Table S4.1). These findings follow with the observation that genes with convergent longevity signals capture a significant portion of heritable longevity variance(Farré et al., 2021b), and that genomic lifespan associations also implicate aging-associated diseases such as cardiovascular decline and Alzheimer's disease(Timmers et al., 2019).

Interestingly, we found that epigenetic proteins and their interactors were depleted for CAAS occurrence. It may be the case that a number of these proteins are under strong purifying selection (e.g., due to functional constraint), and thus biased against CAAS occurrence. Indeed, genes under loss-of-function constraint(Karczewski et al., 2020) made up ~40% and ~30% of epigenetic proteins and their interactors, respectively (Methods). We instead observed that internal interacting pairs (within an epigenetic gene set) were enriched for CAAS hits, suggesting that part of the non-random distribution of CAAS could reflect protein partner co-occurrence. Unexpectedly, this finding was not unique to epigenetic genes, but instead was found consistently across GO terms. Thus, the effect of protein interactions on convergent substitutions is a general phenomenon across sets of functionally-related genes. This is consistent with a role for epistasis in protein evolution, wherein a mutation on one protein can impact the phenotypic effects of other mutations through direct or indirect physical interactions(Starr and Thornton, 2016), potentially altering evolutionary rates and trajectories(Dasmeh et al., 2017; Salverda et al., 2011).

Looking at the level of protein functional domains we found that those proteins containing CAAS possessed certain domains at a greater frequency. For example, CAAS genes in the GO term 'chromatin' had a disproportionate number of PAS and SRC-1 domains (Figure 4.2A). PAS domains are associated with redox signaling sensors(BL and IB, 1999) such as hypoxia-inducible factors (CAAS hits in this set included HIF1-3A), as well as the hormone-responsive NCoA co-activator family (NCoA1-3, which were also CAAS hits), the latter of which also contain SRC-1 domains. HIF factors and oxidative stress have been associated with longevity(Belenguer-Varea et al., 2020; Salmon et al., 2010; Zhang et al., 2009). It may be that modification to the nuclear regulators of the stress response modulate the sensitivity of this system to insult, particularly under chronic stress(Salmon et al., 2010). Additionally, we observed increased occurrence of some domain families, several of which overlap with domains over-represented

in aging-related genes(Tacutu et al., 2018) and related to aging processes. These included helicase domains (e.g., with CAAS hits in the RecQ family(Brosh and Bohr, 2007)), forkhead domains (with CAAS hits in the FOXO family(Greer and Brunet, 2005)), and PI3/PI4 kinase domains(Weichhart, 2018). These instances of over-occurrence of aging-relevant domains could imply that the evolutionary forces behind CAAS may target genes with particular protein activities, even if CAAS themselves tend to not fall within the majority of protein domains (Supplementary Table S4.2).

This tendency to fall outside protein domains may be expected if the majority of CAAS are predicted to be benign(Farré et al., 2021a) and domain modifications more likely to be damaging(Adzhubei et al., 2013). That we see biased occurrence of certain domains within CAAS-hit genes, but not directly hit by CAAS themselves, may be explained if CAAS are only tolerated when they subtly modify protein activity – e.g., by altering complex formation (see below) or indirectly impacting the active site via conformational shifts. Distal mutations can have profound impacts on enzymatic activity at active sites(Gagné et al., 2015; Osuna, 2021; Wang et al., 2021); this may be true for CAAS hits outside domains, though given the difficulty in predicting such distal effects(Osuna, 2021) this possibility remains an open question.

The only domains/clans for which direct CAAS-hits were enriched were the Serpin domain and the 'Aha/BPI' clan; Serpins are serine protease inhibitors with known roles in several important pathways, including mammalian coagulation(Law et al., 2006; Rau et al., 2007). A previous study of lifespan and AA substitutions within primates highlighted genes involved in coagulation (Muntané et al., 2018); our results here suggest that this pattern holds when considering lifespan diversity more broadly. Interestingly, Serpin domains were also detected in a study of adaptive selection in mammals(Slodkowicz and Goldman, 2020), establishing these as molecular targets upon which selection can act. CAAS genes containing domains of the 'Aha1_BPI' clan included several innate immunity genes (e.g. LBP); immunity has been previously linked with cross-species longevity differences(Farré et al., 2021b; Kowalczyk et al., 2020) and aging more generally(Franceschi et al., 2007). Notably, a longevity-associated variant in *BPIFB4* (which we also identify as a CAAS gene), has been described(Villa et al., 2015) and protects against age-linked atherosclerosis when genetically engineered in mice(Puca et al., 2020).

We next considered the spatial distribution of CAAS mutations, using a similar logic of applying spatial information in understanding cancer mutation. Such a spatial-clustering approach has been previously applied to associate positively-selected residues in mammals with putative functional effects in immune proteins(Slodkowicz and Goldman, 2020), and derived residues more broadly(Adams et al., 2017), supporting our use of spatial information as an important aspect of characterizing CAAS. Through this analysis we found numerous examples of CAAS spatially clustering with cancer mutations (Niu et al., 2016b). These included several proteins associated with both aging and cancer, such as BRCA1(Cao et al., 2003), VEGFA(Carmeliet and P, 2005; Lähteenvuo and Rosenzweig, 2012) and EGFR(H. Liu et al., 2020; Rongo, 2011). We also found several domains enriched for clustered mutations, including the Matrixin and p450 domains. Extracellular matrix (ECM) maintenance and turnover are important processes in longevity(75, 76) and cancer, while p450 enzymes are involved in processing environmental toxins, decrease expression with age, and can contribute to cancer risk(Agundez, 2005; Wauthier et al., 2007). Given the substantial links between aging and cancer(De Magalhães, 2013), that we see such instances of selection for tumorigenesis, as well as putative longevity selection, modifying similar spatial regions on proteins suggests that the molecular targets of selection, i.e., certain proteins or protein-activities, may be shared. This sharing is further supported by the observation that CAAS residues can directly overlap cancer-mutated residues, though they do not do so at a significantly-elevated rate (Supplementary Table S4.3). We also found that these mutational clusters can occur within protein-protein interaction domains, such that a molecular target for selection may be the interactions between protein pairs, in addition to the functional domains of individual proteins themselves.

Our finding of CAAS occurrence between protein pairs suggested a possible interplay between evolutionary forces and protein interactions. The unique evolutionary constraints experienced by protein pairs to preserve interaction kinetics can result in co-evolution, particularly at domains acting at the interface(Choi et al., 2009; Duarte et al., 2012; Teppa et al., 2017), and conversely modifications to partners can lead to large phenotypic changes(Jubb et al., 2017b). Furthermore, PPI domains can be modified evolutionarily to either increase or decrease the specificity of a protein for different possible

partners as gain/loss-of-function events(Plach et al., 2017; Sikosek and Chan, 2014), with simulations suggesting that as few as two substitutions may be sufficient to produce a novel interaction interface(Levy, 2010).

Interestingly, we found that CAAS were over-represented in annotated PPI domains, lending credence to the notion that PPI may be directly targeted for modification. This follows with the idea that the evolution of longevity may be influenced by complex relationships between proteins, with PPI acting as a more explicit form of epistasis(YV and SL, 2010). Indeed, PPI interfaces tend to be more evolutionarily conserved(Choi et al., 2009), as would be expected if these interactions are functionally important and members can exhibit evolutionary histories and modification rates that are influenced by partners(Goh and Cohen, 2002; Makino and Gojobori, 2007). We observed an enrichment of CAAS hitting both interaction domains of protein pairs as well as instances in which both proteins contained CAAS, one of which occurs within an interaction domain (Supplemental Table S4.4). In the case of the latter, it may be that CAAS occurring outside the interaction domain indirectly impact binding kinetics at the interface, such that the combination of a direct interface mutation on one protein and an indirect conformation-altering mutation in the partner protein are sufficient to achieve some phenotypic effect. Further *in-silico* modelling and functional testing may reveal the extent to which such a scenario holds across PPI pairs. It is also possible that members of a PPI are modified unilaterally (i.e., only one protein of a pair is CAAS hit), either to alter binding kinetics, or to allow for more or less promiscuous interactions with other potential partners(Aakre et al., 2015). As well, while we did not focus on homo-oligomers in this study, we did observe an enrichment of CAAS falling within these interaction domains (Supplemental Table S4.4). We describe these instances for use by others (Supplemental Table S4.4).

In our analysis of PPI modifications, we highlighted two examples of particular interest. One of which was between APAF1 and CASP9, wherein an I37P CAAS falls within the CARD domain of APAF1. The caspase-recruitment (CARD) domain of APAF1 is essential to recruiting and activating CASP9 to the assembling apoptosome complex(Li et al., 1997; Qin et al., 1999a; Yoshida et al., 1998). The short-lived residue within this domain was predicted to be strongly unfavorable in terms of binding affinity, suggesting

that the long-lived residue may act to improve the interactions between these two proteins. Indeed, a structural analysis of CASP9 recruitment by APAF1(Qin et al., 1999b) identified core residues making up the hydrophobic interface that confers specificity in CAPS9 recognition, which included Ile37 (Figure 4.3B). Several mutations in the CARD domain were also shown to abolish interactions with CASP9(Qin et al., 1999b), suggesting the possibility that mutation to the short-lived residue at Ile37 could have similar effects. Programmed cell death (PCD) can become dysregulated with age(Kavathia et al., 2009), increasing in some cell types but decreasing in others(Tower, 2015) - especially in the context of mitochondrial dysfunction(Kujoth et al., 2005) - and also plays an important role in cancer suppression(Lowe and Lin, 2000; Tower, 2015). It may be that modification to APAF1-CASP9 recruitment, a critical step in initiating apoptosis(Li et al., 1997; Yoshida et al., 1998), may impact aging/longevity by altering the sensitivity of PCD, either under homeostatic conditions or following severe stress/damage.

An additional CAAS-PPI occurrence was in the CD47/SIRPa interaction, where a Q67K mutation in SIRPA falls within the interaction domain. Point mutation experiments have demonstrated that Q67 plays a key role in mediating SIRPa binding to CD47(Liu et al., 2007). This study generated a mutant version of SIRPA with the short-lived residue (K) and found this to substantially inhibit CD47 binding, providing experimental evidence to suggest a functional effect of this CAAS. Furthermore, mutation of Q67 also reduces *in-vitro* cell adhesion and suppression of leukocyte migration, pointing at a potential for *in-vivo* effects as well. The CD47-SIRPA interaction acts in an immune inhibitory circuit(Logtenberg et al., 2020), particularly in controlling myeloid cells and phagocytosis(Veillette and Chen, 2018), a pathway often exploited by tumor cells seeking to avoid elimination(Logtenberg et al., 2020; Veillette and Chen, 2018; Willingham et al., 2012). Interestingly, CD47-SIRPA also plays a role in removal of aging cells and HSCs during normal homeostasis(Logtenberg et al., 2020); it is tempting to speculate that the increased affinity of this interaction conferred by the long-lived residue may represent a trade-off between improved maintenance of homeostasis at the cost of increased risk for tumor escape.

Finally, we looked for instances in which CAAS long/short-lived residue mutations have been phenotyped as part of deep mutational scans (Esposito et al., 2019). p53 is a well-known factor involved in a myriad of cellular processes(Kastenhuber and Lowe, 2017) as well as in aging(Rufini et al., 2013), for which we identified a CAAS falling within the transactivating domain (TAD2). This domain is important for binding of p53 to other proteins; a detailed NMR study(Miller Jenkins et al., 2015) of this domain in complex with the Taz2 domain of p300 found that the positively-charged GLU51 consistently forms salt bridges and/or hydrogen bonds with two residues of Taz2. It may be expected that, were this residue replaced with a neutrally-charged substitution that this may impact TAD2-Taz2 binding. Indeed, this study replaced GLU51 with an uncharged amino-acid (E51Q) and found that this change led to a two-fold decrease in binding affinity. It may therefore also be the case that the substitution of a similar neutral residue, the short-lived alanine at this position, also decreases binding affinity of p53 for protein partners. This could impact p53 function in certain instances but not others; E51A cells exhibited positive selection following etoposide treatment, indicating that the mutated p53 retained some WT-like activity(Giacomelli et al., 2018). That mutations in the transactivating domain generally appeared to retain functionality in the context of DSB-inducing agents suggests the effects of modifying p53 interactions at the TAD2 domain may only be evident under certain cellular conditions, a topic ripe for future research.

A CAAS falling within a transmembrane domain of CCR5 was associated with slightly altered binding of the 2D7 antibody(Heredia et al., 2018), though given 2D7 recognizes an epitope on the extracellular loop of CCR5(Khurana et al., 2005), this effect could be due to subtle conformational changes. Given its position within the protein, it is also possible that this CAAS impacts chemokine binding, particularly receptor activation through interactions between the N-terminus of cytokines and the transmembrane helix bundle(Blanpain et al., 2003; Howard et al., 1999). While characterizing CCR5-chemokine interactions with deep mutational scanning was infeasible in the initial study(Heredia et al., 2018), targeted functional assays specific to this particular S114G mutation could suggest that the long-lived residue (S114) facilitates increased sensitivity of CCR5 to chemokine signaling, with ramifications for immune regulation and possibly ‘inflammaging’(Franceschi et al., 2007).

Overall, our study extends our understanding of how evolutionary routes towards increased longevity can converge by considering convergence not only at the gene-level, but within the context of protein structure, function, and interactions. Given that these properties play an obvious role in the phenotypic effects of mutations, we believe these findings offer compelling concepts for studies characterizing the longevity-altering effects of protein changes. Importantly, these concepts need not be limited to changes generated by natural selection but can serve as a basis for guiding protein design of novel longevity-increasing mutations.

4.4: Materials and Methods

Convergent amino-acid substitution (CAAS) definition

The global set of genes considered in the CAAS analysis (see Supplemental Table S4.1) was obtained using the criteria previously described in Farre et al., 2021. Mammal species from Multiz protein alignments were selected based on the distribution of longevity quotients (LQ), selecting the top (longest-lived) and bottom (shortest-lived) four species. Longevity information was aggregated from the AnAge database, which is a record of largely zoo-based longevity data for different species. The number of species used at these extremes represented a trade-off between the total number of detected CAAS hits and the percent of hits which could be validated using additional statistical tests (see Figure S3 of Farre et al., 2021). The following species were selected as top: Myotis lucifugus (myoLuc2), Homo sapiens (hg19), Myotis davidii (myoDav1), and Heterocephalus glaber (hetGla2); and as bottom: Condylura cristata (conCri1), Rattus norvegicus (rn5), Pantholops hodgsonii (panHod1), and Mesocricetus auratus (mesAur1).

CAAS were annotated according to four different scenarios: (1) invariant across long-lived and short-lived species, (2) invariant across long-lived, variable across short-lived species, (3) variable across long-lived, invariant across short-lived, and (4) variable across both sets of species. For all CAAS, a phylogenetic ANOVA test that accounts for phylogenetic relationships (using the RRPP package in R(Collyer and

Adams, 2018; R Development Core Team, 2008)) was performed using 10,000 iterations for significance testing (Table S4.1).

Gene-set enrichment analyses

Gene set enrichments, whenever indicated in text, were performed using GO Biological Process (BP) terms via the 'enrichGO' function from the clusterProfiler(Yu et al., 2012b) package version 3.13.1. For the initial enrichment of CAAS-containing genes at the scenario-level (Table S4.1), the background set of all genes considered in the CAAS analysis (as described above and in Farre et al.(Farré et al., 2021a), and provided in Table S4.1) was used as the background set for enrichment testing. For other instances in which GO term enrichments were performed, subsequent methods text details the background sets used. Benjamini-Hochberg(Benjamini and Hochberg, 1995) p-value adjustments are applied to correct for the number of tests performed, with a minimum adjusted p-value threshold of 0.05 used to define enriched terms. Semantically-similar enriched GO terms were subsequently collapsed using the 'simplify' function from clusterProfiler, using default settings. Enrichment visualizations were performed using the 'dotplot' function from ggplot2, showing the top 30 enriched terms (see Table S4.1 for full sets). For the initial gene-set enrichments shown in Figure 1B, we combined CAAS falling under Scenarios #1-#3.

Collecting epigenetic gene sets and interactors

GO gene sets related to epigenetics were manually collected from MSigDB(Subramanian et al., 2005) using search criteria including 'chromatin', 'histone', and 'DNA methylation' (see Table S2 for full set of processes). To consider proteins interacting with those contained in these GO terms we downloaded the STRING database (9606.protein.actions.v11.0.txt.gz) and defined interactions as those having a 'combined' score of 700 or greater (as implemented in Ratnakumar et al. 2020(Ratnakumar et al., 2020) and suggested on the STRING version 11 FAQ). As a more stringent definition, we also defined interactions as those for which physical binding data (9606.protein.physical.links.detailed.v11.0.txt.gz) scores were >= 700. Results of analyses using both definitions of interactions are shown in Table S4.1.

Enriched gene-ranks for cross-disease heritability

A set of aging-related diseases was assembled based on epidemiological datasets(Chang et al., 2019) and used to extract GWAS summary statistics from the UK Biobanks dataset(Sudlow et al., 2015), as described more thoroughly in Richard and Capellini 2021(Richard and Capellini, 2021). Briefly, for each disease, SNPs were tied to genes using a 200kb window around protein-coding gene TSS, with the most significant association p-value assigned to that gene. All genes were subsequently ranked according to these assigned p-values. This was repeated across all 127 GWAS sets used in that study. For each gene, the number of times in which it ranked in the top 75^{th} percentile of ranks (i.e. was ranked highly in its association with the given disease) was counted. This establishes a genome-wide distribution of counts (from 1 to 127). For a given gene set (e.g. CAAS-hit genes), the distribution of counts for the target genes was compared to that of the global distribution of protein-coding genes (exclusive of the target set) using a one-sided (alternative = “greater”) Student’s t-test in base R. Resulting p-values were adjusted for the number of GO terms tested using a BH correction, with significance defined as adjusted $p < 0.05$. Given our interest in proteins interacting with epigenetic gene sets (see below), we also ran the above cross-rank enrichment testing on gene sets expanded to including interacting partners (see Table S4.1).

To test whether CAAS-hit genes had greater cross-disease association than expected, a similar method was applied, with the background distribution instead consisting of protein-coding genes (exclusive of CAAS-hit genes) subset to those considered in the initial CAAS definition (see above and Table S4.1).

In order to test whether CAAS-hit genes within a given GO term had proportionally greater cross-disease association, we tested the bias in gene-wise ranks conditioned on the GO term. The global distribution of protein-coding genes was subset to those genes included in a given GO term, with this used as a background distribution and compared to the distribution of rank-counts for the target set – CAAS-hit genes within this GO term, again using the Student’s t-test in base R. A BH adjustment was again

applied. As above, we also performed these conditional-enrichment tests using gene sets expanded to include interacting partners, for which we saw greater significance – possibly due to the larger size of the background gene sets for t-test comparison (see Table S4.1).

CAAS occurrence across proteins and interactors

For a given gene set (i.e. GO term), the frequency of CAAS occurrence was compared to random expectations by generating a set of 1000 randomly-generated gene sets of equal size (sampling from the global set of genes used in the initial CAAS definition). The number of CAAS-hit genes (ignoring multiple hits per-gene) in the target set was tested for statistical significance by comparing to the set of standardized background values using a CDF of the standard normal distribution with the 'pnorm' function in base R. P-values for significant deviations from the background distribution were corrected for the number of GO terms (n = 9) using a Benjamini-Hochberg FDR correction. Significance was defined as adjusted $p < 0.05$ (Table S4.1).

Next, for the given gene set, all interactors (as defined above) were defined via the STRING database, and the number of interacting proteins which were CAAS hit was counted. The same procedure was applied to all randomized background gene sets. The number of CAAS-hit interacting genes was tested for statistical significance against the background distribution using a normal distribution CDF as above. Testing was performed using both the 'combined score' based interaction definition, as well as those defined solely using physical binding scores (see Table S4.1).

Finally, for each gene set, the set of interacting genes containing CAAS was intersected with the set of CAAS-hit genes in the gene set itself to define 'internal interactors with shared CAAS'. This was also done for all randomly-generated gene sets, with statistical significance again calculated comparing the target value to a standardized background distribution using a normal distribution CDF, with BH correction applied for the number of GO terms tested. This was again done for both definitions of interactors.

To test whether our findings were unique to epigenetic gene sets, we randomly collected GO terms to run the above analyses. For each epigenetic GO term, all GO terms in the GO BP database (obtained from MsigDB – file c5.go.v7.4.symbols.gmt) were ranked by absolute (# of genes in epigenetic term - # of genes in term), with the 100 closest size-matched terms selected. Of these, five terms were randomly selected to include in the final set. The final sets of randomized GO terms were then combined to give a background set of unique terms. This was done to account for possible confounding effects of selecting random GO terms substantially smaller/larger than the epigenetic GO terms used. For each randomly-selected GO term, the above statistical tests for the occurrence of CAAS were performed, with BH correction applied for the total number of randomly-selected GO terms (n = 44) for each statistical test (see Table S4.1).

Protein Domain Occurrence

The PFAM human protein domain database was obtained (9606.tsv – Release 34.0) from (http://ftp.ebi.ac.uk/pub/databases/Pfam/current_release/proteomes/). Gene symbols were subsequently mapped to Uniprot accessions and intersected with this database, yielding lists of domain occurrence for genes in individual epigenetic GO terms, as well as a global list of genes considered in the CAAS background set. We tested the occurrence of certain PFAM domains within epigenetic GO terms as follows. For each PFAM domain occurring at least more than once within a given GO term, the total number of instances of this domain across all genes in the PFAM database (counting multiple hits in a given gene as a single count) was considered the total number of 'true' cases for that domain. We compared the occurrence counts of this domain in the GO set vs. the background using the 'phyper' function from base R. P-values from these hypergeometric tests were adjusted for the number of domains tested using BH correction for each GO set considered. Enrichments were plotted as log-FC values using ggplot2 (see Figure 4.2A). Within each GO term, we also extracted those genes containing at least one CAAS hit as the 'CAAS Target' set. Similar to above, the occurrence of a given PFAM domain within this target set was compared to the global occurrence of said domain using the 'phyper' function followed by BH correction. Finally, within a given GO term, we compared the number of occurrences for a given

domain in CAAS-hit genes to the occurrence of said domain in the GO term as a whole (counting multiple hits in a given gene as a single count). The 'phyper' function was again used along with BH correction for the number of domains tested for a given GO term.

Separately, to test the occurrence of PFAM domains within CAAS-hit genes as an entire set (not stratified by membership within GO terms), we performed similar per-domain hypergeometric tests, with the background defined as the occurrence of domains within the CAAS background set of genes (as presented in Table S4.1) which had entries in the PFAM database. To test the occurrence of PFAM domains within GenAge genes as a separate set, we performed similar per-domain hypergeometric tests, using the global occurrence of domains across PFAM as the background set. The overlap between enriched PFAM domains in CAAS-hit genes and GenAge genes was tested for significance using a hypergeometric test comparing the number of overlapped domains to the total number of enriched domains in the CAAS-hit set – relative to the total number of unique domains occurring within CAAS-hit proteins (see Table S4.2).

Within each gene, for a given PFAM domain hit, we took the annotated start-end positions of the HMM match and looked for CAAS hit residues falling in this range. Per-CAAS domain hits are presented in Table S4.2. For all protein domains directly hit by at least one CAAS residue, we counted the number of genes containing these CAAS-hit domains (counting multiple hits in a given gene as a single count) and compared to the number of genes in the background CAAS set containing these domains using the 'phyper' function (see Table S4.2), again applying BH correction for the number of domains tested.

dbNSFP Integration

The dbNSFP database(X. Liu et al., 2020) was obtained from http://database.liulab.science/dbNSFP. This database contains nucleotide modifications which are subsequently translated into non-synonymous mutations and fed into various predictive model software in order to assign the potential for deleterious effects of these mutant residues. In order to align our CAAS dataset to this database, we used the

backlocate program(Lindenbaum, 2015) with the hg19 reference genome sequence and gene annotation file (GRCh37.87.gtf) obtained from Ensembl. We intersected the chromosomal positions of mutations leading to CAAS residue modifications (via backlocate) with the chromosomal positions from dbNSFP, further requiring that the specific CAAS change (i.e. long-lived to short-lived amino acid) was shared with a matched dbNSFP entry. Given our interest in the predicted impacts of CAAS on protein structure/function, we considered the columns in dbNSFP describing the predicted mutation effects generated from different computational software, using the codes 'H', 'M' and 'D' provided by researchers. The number of software predictions of non-benign/tolerated effects for a given residue change were summed, with all residues having at least one predicted effect retained. To increase the stringency of these results, we further required that all residues in this set also directly intersected a PFAM protein domain, using our previously-generated per-CAAS PFAM mappings. See Table S4.2 for this complete set.

Spatial clustering of residues

We obtained the HotSpot3D(Niu et al., 2016b) software version 0.6.0 from Github (https://github.com/ding-lab/hotspot3d). Pre-processed proximity files were obtained from https://www.synapse.org/#!Synapse:syn8699796/wiki/421392 for the GRCh37 genome build, with gene names converted to PDB accessions using the 'hugo.uniprot.pdb.transcript.csv' file also made available. To generate a MAF format suitable for use with this software we implemented a custom-made script in base R. Briefly, for each CAAS we used the associated Refseq accession number to retrieve information via biomaRt version 2.44.4 (using the ENSEMBL 'hsapiens' database). For each CAAS we formatted the mutation code as 'p.[Long-lived AA][position][Short-lived AA]'. For CAAS falling under Scenarios 2 or 3 (multiple long-lived or short-lived AAs) we added one entry for each AA version (this not impacting the distribution of unique residue modifications along proteins for clustering purposes). We also obtained Supplemental Table 24 (41588_2016_BFng3586_MOESM25_ESM.xls) from the original HotSpot3D publication, which provides somatic mutation data for 1,818,754 mutations across 19 TCGA cancer types in a pre-prepared format for use with HotSpot3D. This dataset was subset to those genes included in the

CAAS background gene set, and for which pre-processed HotSpot3D files were available. The subset TCGA dataset was then pooled with the reformatted CAAS dataset to generate a combined input file for use with HotSpot3D. Residue clustering was performed using the following flags: '--3d-distance-cutoff 10' and '--meric-type intra'. The resulting cluster information was then filtered to retain only those cluster calls containing at least three residues; this set of calls was then further filtered to retain clusters containing CAAS residues. Additionally, the subset TCGA dataset was separately processed with HotSpot3D in order to determine the amount of additional information (i.e. additional cluster calls) provided by the inclusion of the CAAS dataset. The CAAS dataset by itself was too small to generate cluster calls using this algorithm (data not shown). Gene set enrichments were performed as described above, using as background the set of CAAS background genes for which pre-processed HotSpot3D files were available (and for which TCGA mutations were described). These were performed using all genes for which at least one cluster call contained a CAAS residue, as well as genes subset by the scenario of CAAS residues (1-3).

Clustered residue disease annotations

Enrichments for DOSE(Yu et al., 2015) terms were performed using the clusterProfiler(Yu et al., 2012b) package version 3.13.1 along with the DOSE package version 3.18.3, using as background the set of CAAS background genes for which pre-processed HotSpot3D files were available and applying a BH p-adjustment. Significant enrichments were visualized with the ggplot2 package version 3.3.5 using the 'dotplot' function for the top 20 enriched DOSE terms. In order to look at the distribution of gene associations within these enriched terms, we downloaded the set of human disease ontology terms (*https://github.com/DiseaseOntology/HumanDiseaseOntology/blob/main/src/ontology/HumanDO.obo*) and for the top 30 enriched DOSE terms (sorted by gene ratio) extracted all associated ontology hits. The set of genes possessing at least one CAAS-containing cluster was analyzed using the disgenet2r library (version 0.99.2)(Piñero et al., 2020) function 'gene2disease' to map genes to disease annotations. Finally, we took the gene-to-disease mappings and intersected them with our subset ontology hits. The

resulting object was visualized using the network plotting function built into the disgenetR using the following flags: prop = 5.

Visualization of mutant residues on EGFR

To visualize the mutated residues occurring in the EGFR protein, we first obtained the relevant protein sequence from Uniprot (P00533), and also retrieved the PFAM domain hits for this protein from our previous datasets. All called clusters in the EGFR protein were plotted as dots using base graphics in R, using a rainbow colour scheme to differentiate residues belonging to different clusters. All clusters for which CAAS residues were members were plotted at a taller height to help in visually distinguishing them, with asterisk used to denote CAAS residues themselves. Visualization of cluster distribution on the EGFR protein was performed using the 'visual' program from Hotspot3D with a PDB crystal structure obtained from the PDB database (accession 4KRP); CAAS residues were manually coloured red using the PyMol interface.

Domain occurrence of CAAS-containing clusters

We next considered those instances in which residues of CAAS-containing clusters fall within PFAM domains within their respective proteins. For a given cluster, we took all residue positions and intersected them with PFAM domain hits for that protein, assigning domains to residues when they intersected. This is different from the above analysis where we test for the occurrence of CAAS within PFAM domains – here we consider either the CAAS directly, or a residue which lies spatially adjacent to a CAAS within the 3D protein structure, and test domain-hitting biases in this expanded residue set. For a given domain, we compared the number of genes containing instances of residue-hit domains to the total number of genes containing this domain. We performed these tests for all domains which had at least two genes with domain-hitting residues (n = 30), and subsequently applied BH correction.

Overlap of CAAS and TCGA residues

Using our pooled MAF-formatted dataset of CAAS residues along with pre-processed TCGA residues (from(Niu et al., 2016b)), we counted the number of instances in which a TCGA-mutated residue on a given protein was also a CAAS-modified residue at this position. We next considered all proteins shared between the TCGA and CAAS sets. For each given protein we obtained the amino-acid sequence from Ensembl via biomaRt version 2.44.4, then counted the number of CAAS occurring within said protein. These CAAS were then randomly shuffled across the length of the protein to generate a randomized set of CAAS residues for that protein. This CAAS-shuffling was performed across all proteins containing at least one CAAS for 1000 iterations, yielding a background of 1000 sets of randomly-shuffled CAAS datasets to match the true target set. For each shuffled set we then counted the number of TCGA residue positions which overlapped the simulated CAAS residues, establishing a background for the number of TCGA-CAAS overlaps which would be expected if CAAS were randomly distributed across protein sequences. We then compared standardized values from this background distribution to the number of CAAS-TCGA overlaps of true CAAS residues via the 'pnorm' function in base R (Table S4.3).

As an additional test for the significance of overlap between CAAS residues and cancer mutations, we downloaded an expanded set of somatic mutations based on the TCGA(Campbell et al., 2020) (https://gdc.cancer.gov/about-data/publications/mc3-2017, using the file mc3.v0.2.8.PUBLIC.maf.gz). We similarly pooled this dataset with our MAF-formatted CAAS data, and again looked for overlap at the level of residue positions within proteins shared between these two sets. The same background of 1000 randomized CAAS sets was used to compare the number of true residue overlaps to that expected from randomly-shuffled residue positions using the same statistical approach as described above (Table S4.3).

ProtCID Integration

ProtCID(Xu and Dunbrack, 2020) is a database of protein-protein interactions which extends the resolution of PPI captured from crystal structures to the level of interactions between PFAM domains of protein binding partners, and provides additional information as to patterns of domain-domain interactions

across large sets of protein complexes. We subset our set of CAAS-containing proteins to those for which a CAAS intersected at least one annotated PFAM domain. Given this set of CAAS-hit PFAM proteins, the developers of this database kindly provided us with a dataset of all domain-domain interactions involving at least one of these proteins, along with the corresponding crystal structures. This domain-domain interaction dataset was separated into homo-dimeric and hetero-dimeric interactions for processing. For homo-dimeric interactions, we identified all instances in which CAAS residues fell within either one or both of the annotated PFAM domains involved in the interaction, as well as cases where a CAAS fell within a self-interacting domain (e.g. the PDGF domain of VEGFA) (see Table S4.3 for full set of homo-dimeric interfaces). For hetero-dimeric interactions, we identified all instances in which CAAS residues fell within either one or both of the annotated PFAM domains involved in an interaction.

To more stringently identify CAAS residues potentially involved in the protein-protein interface, we took the crystal structures obtained from ProtCID for all domain-domain interactions and imported them into R via the Rpdb library version 2.3. These structure files detail the chains in the protein pair which are involved in the protein-protein interaction, as well as the domain regions in these chains that are involved in the interface. The distances between all atoms of all residues in the interaction domain of one protein and all atoms of all residues in the interaction domain of the partner protein were calculated using the 'distances' function from Rpdb, and the norm of the xyz distance vectors was calculated using the 'norm' function. The minimum distance between all atoms of one residue and all atoms of a residue on the partner domain was assigned to each residue-residue pair. Interface residues on one domain were defined as those within 5Å of any other residue on the partner domain. These interface residue definitions were generated across all hetero-dimeric protein interfaces in our set, and subsequently intersected with our CAAS set to identify CAAS residues which may also function as potential interface residues (Table S4.3). This interface residue definition was also applied to all homo-dimeric protein interfaces in our set.

To test the significance of the observed number of CAAS residues falling within interaction domains (either hetero- or homo-dimer interactions), we used our sets of randomly-shuffled CAAS residues to establish the expected rate of CAAS-interaction domain overlaps if CAAS were randomly distributed

across proteins. For each randomly-shuffled set, we first assigned shuffled residues to PFAM domain annotations using the PFAM database. We then looked across all interactions in our set and looked for overlap between fake-CAAS PFAM hits and PFAM domains involved in these interactions, using the same algorithm as applied above for the true set of CAAS. This was done across all 1000 randomly-shuffled sets, defining background expected rates of occurrence of different scenarios of overlap. These scenarios included:

(1) Overlapping at least one homo-dimer interaction domain

(2) Overlapping both interaction domains of a homo-dimer, where both domains are the same PFAM type

(3) Overlapping both interaction domains of a homo-dimer, where domains are different PFAM types

(4) Overlapping at least one hetero-dimer interaction domain

(5) Overlapping both interaction domains of a hetero-dimer

(6) Overlapping at least one homo-dimer interacting residue (defined as above)

(7) Overlapping interacting residues within both interaction domains of a homo-dimer, where both domains are the same PFAM type

(8) Overlapping interacting residues within both interaction domains of a homo-dimer, where both domains are different PFAM types

(9) Overlapping at least one hetero-dimer interacting residue

(10) Overlapping interacting residues within both interaction domains of a hetero-dimer

To visualize the position of CAAS within protein complexes, we used SIFTS residue mappings (obtained from ftp://ftp.ebi.ac.uk/pub/databases/msd/sifts/flatfiles/tsv/pdb_chain_uniprot.tsv.gz) to map mutant residue numbers to their equivalent positions in PDB structures. For a given PDB structure, we mapped our set of CAAS residues, then subsequently used the Rpdb package to annotate CAAS residues as metadata in the PDB format. The resulting file was then imported to Chimera version 1.15 for visualization (see Figure 4.3A-B, Figure S4.3).

IntAct database and SSIPE integration

The IntAct database contains annotated experimental evidence of mutations shown to impact protein-protein interactions – this was obtained at the following link: ftp://ftp.ebi.ac.uk/pub/databases/intact/current/various/mutations.tsv. We intersected this set with our set of CAAS residues, requiring that the exact long-lived (human) to short-lived amino acid substitution was experimentally described. This yielded a set of three protein-protein interactions for which the CAAS mutation was described (Table S4.4).

We took our set of CAAS residues falling within protein interaction interface domains and formatted mutation strings (of the form long-lived vs. short-lived) for use with the SSIPE webserver(Huang et al., 2020), though, given that the webserver only handles protein dimers (and not larger complexes), we limited our searches to interactions involving only two proteins. Within this limited search space we identified the CASP9-APAF1 interaction as being significantly-deleteriously disrupted by the long-to-short residue change. Wild-type and mutant protein structures (generated by the SSIPE webserver) were visualized using Chimera version 1.15, as shown in Figure 4.4C.

Clusters and protein-protein interfaces

All cluster calls (generated from our above analyses in section '*Spatial clustering of residues*') for which CAAS residues were at least one member were considered. All residues within these clusters falling within annotated PFAM hits (in their respective proteins) were assigned to said domains. We then subsequently intersected these residue-domain hits with the sets of interaction domains from the ProtCID dataset, looking for instances in which either (a) a CAAS residue (belonging to a cluster) falls within an interaction domain (which can be considered a subset of the above analysis) or (b) a residue occurring adjacent to a CAAS residue (i.e. belonging to the same cluster) falls within an interaction domain. Finally, we intersected the set of interfaces identified in this analysis with the set generated using CAAS residues overlapping interaction domains, yielding sets of interfaces for which CAAS residues fall within both interaction domains while also being members of mutational clusters. We also identified interfaces for which CAAS residues do not fall within both interaction domains, but are members of mutational clusters

with adjacent residues falling within interaction domains. Both sets of interfaces are indicated in Table S4.4.

Multiplex assay of variant effect database

The rapimave package in R (version 0.2.0.9) was used to access the MaveDB database(Esposito et al., 2019), retrieving details on all entries. These data were intersected with our set of CAAS-hit proteins to identify proteins characterized in at least one study in the database, retrieving the results for all studies describing proteins of interest. Within each study, we looked for instances in which the CAAS residue(s) of interest were tested, at that position, and further that the long-to-short amino acid substitution was directly tested in the assay. This yielded five studies, the results of which are shown in Figure 4.4A,C and Supplemental Figure 4.1. To visualize the assay results for the experiments describing our CAAS hits, we used the mavevis library in R version (https://github.com/VariantEffect/mavevis). A modified version of the 'genophenogram' function from mavevis was implemented to highlight our particular CAAS residue changes in the genophenogram output, as well as integrating information on the distribution of PFAM domains across proteins of interest (as shown in Figure 4.4).

Intersection with loss-of-function gene sets

Data from gnomAD(Karczewski et al., 2020) was retrieved from https://gnomad.broadinstitute.org/downloads#v3variants (file: forweb_cleaned_exac_r03_march16_z_data_pLI_CNV-final.txt), with genes subsequently filtered for those with a pLi score of > 0.9. This yielded a set of 3230 unique loss-of-function-intolerant genes. This set was subsequently intersected with the aggregated set of unique epigenetic genes, along with their interactors, and expressed as a percentage overlap.

4.5: Acknowledgments

We thank Dr. Qifang Xu for graciously providing us with comprehensive ProtCID datasets for our genes of interest, and Dr. Andrew Doxey for insightful comments and feedback during the writing of this manuscript.

4.6: Author Contributions

DR: Conceptualization, Data curation, Formal Analysis, Investigation, Visualization, Methodology, Manuscript Writing, Manuscript Editing. GM: Generation of CAAS datasets, phylogenetic statistical validation tests, Conceptualization, Manuscript Editing; AN: Conceptualization, Manuscript Editing; TDC: Conceptualization, Supervision, Funding acquisition, Manuscript Editing, Project Administration.

Chapter V: Discussion and Conclusions

Throughout this dissertation, the analyses presented have contributed to our understanding of how natural selection, acting either in favour of derived developmental traits (e.g., hindlimb morphology) or successively elongated lifespans along the hominid lineage, may have modified genetic sequences controlling aging-related processes. While much work has been done here using evolutionary tenets to explain the history of these sequence changes, their patterns on the genome-wide level, and how they reflect function (e.g., in modifying certain transcription factor binding sites in enhancers), the actual causal mechanisms by which most of these sequence changes mediate alterations to organismal aging and overall lifespan remain to be determined. In this chapter, I will summarize the main findings and conclusions of this dissertation and discuss future avenues of research.

5.1: Key findings and discussion

There were several key conceptual advancements in the realm of evolutionary theory put forward throughout this dissertation. In Chapter 2, we considered a key human derived trait, the bipedal knee, and how its unique configuration has been achieved in part via natural selection acting on gene regulatory regions controlling chondrogenesis during early fetal joint development. By analyzing chromatin accessibility datasets generated from developing mouse and human skeletal tissues we identified signals of ancient positive selection, followed by more recent and potentially ongoing purifying selection, in knee regulatory regions. Furthermore, we found that these regulatory regions were enriched for heritability for osteoarthritis, while osteoarthritis risk loci (as defined through our mining of GWAS literature) were depleted for selection signals, and instead behaved as if subject to genetic drift (Figure S2.3). Since the onset of osteoarthritis is typically later in life, with age as a substantial risk factor(Loeser et al., 2016), this presented a potential example of antagonistic pleiotropy. This led us to formulate an evolutionary model (diagrammed in Figure 5.1) in which strong selection on developmental processes causes genetic variants in regulatory sequences to contribute to heritable risk for late-onset disease via a ‘violation of constraint’ – possibly via alterations to knee structure away from an optimized configuration.

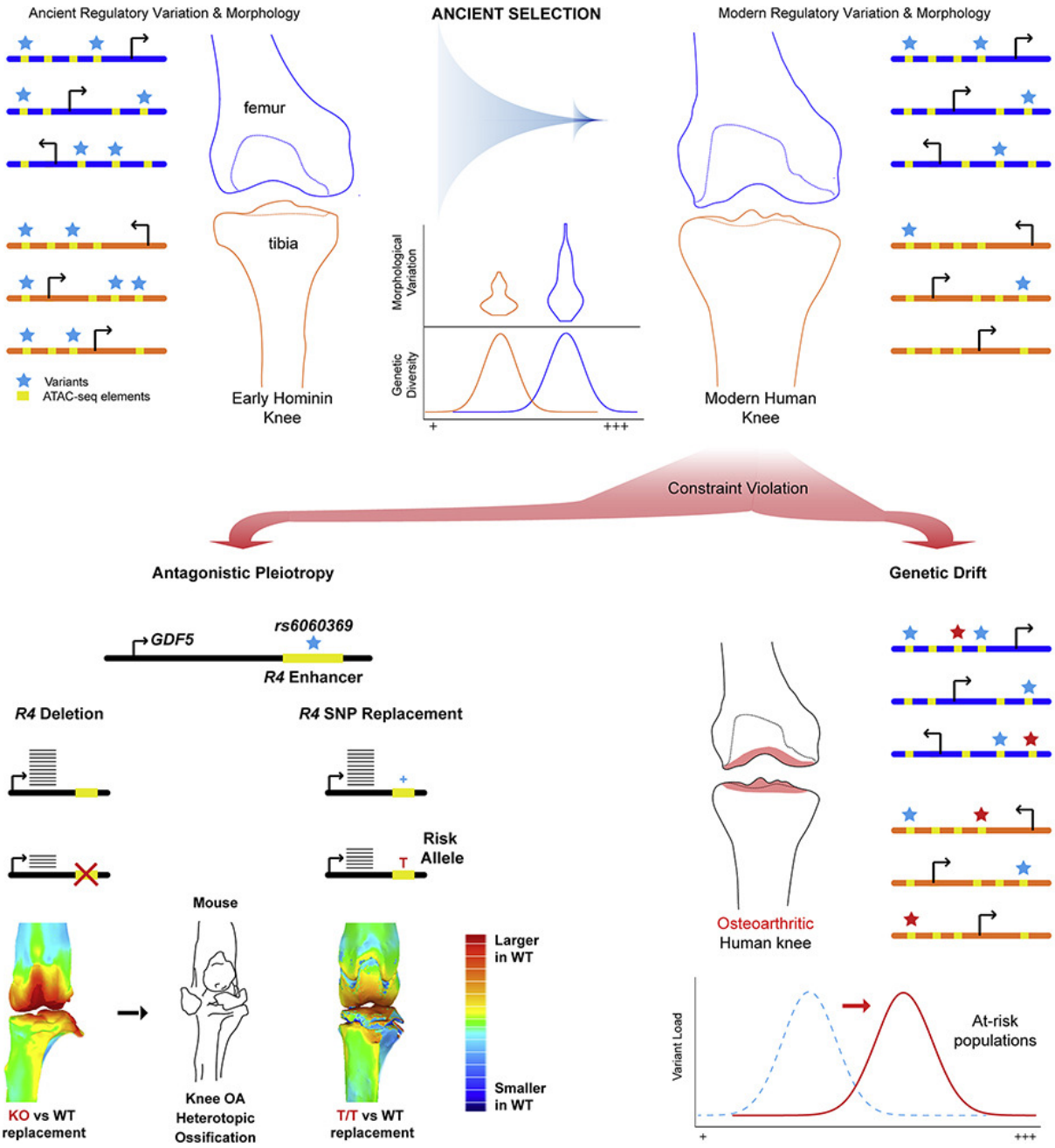


Figure 5.1: Graphical summary of model presented in Chapter 2. (Top): Ancient directional selection for a derived knee configuration is followed by a wave of purifying selection, imposing a functional constraint on knee regulatory regions. (Bottom): This constraint is violated by two processes; antagonistic pleiotropy favouring other traits (e.g. height), and the accumulation of random mutations which are inefficiently purged (due to moderate early-life effects).

Antagonistic pleiotropy has been used as a framework for the understanding of several age-related diseases, including Alzheimer's, coronary heart disease, etc.(Byars and Voskarides, 2020) – our study adds to this literature by suggesting that heritable osteoarthritis risk may be interpreted in a similar framework. Though we would note that the model presented here is a special case of antagonistic pleiotropy – in addition to direct selection for certain alleles beneficial for other traits (e.g., as was observed at the *GDF5* locus), our model suggests that ancient positive selection on the bipedal knee imposed unique biomechanical constraints on knee development. This established constraint goes on to cause emerging variants, which might otherwise be phenotypically neutral, to impose a risk-increasing effect for osteoarthritis; effects which are only apparent later in life. The effects of these variants are minimally deleterious to knee functionality earlier in life, or like other developmental traits the alleles would have been purged by negative selection. Thus, rather than selection for an early-life trait directly, it is instead the tolerable violation of previous selection on the early-life (i.e., developmentally-based) trait that contributes to late-life disease risk via antagonistic pleiotropy. It is worth noting that this unique conception of antagonistic pleiotropy is consistent with the contributions of mutation accumulation effects – variants which become deleterious later in life due to functional constraints (imposed during early development) are permitted to accumulate and contribute to increased late-life decline. We suggest that this general framework, that of considering the links between derived human traits and modern-human disease risk, may be applied to other traits and diseases. Indeed, studies in Alzheimer's genetics have suggested a role for evolution in shaping Alzheimer's-associated loci(Nitsche et al., 2020), as well as possible antagonistic pleiotropy acting on certain alleles (e.g. the apoE system(Byars and Voskarides, 2020; Finch and Sapolsky, 1999). Moreoever, it has been suggested that the enhanced neuroplastic potential that emerged over the course of human evolution contributes to the deleteriousness of brain aging(Neill, 2012). Interestingly, it has been observed that the brain regions most impacted by tau fiber aggregation are those which have been most expanded in the hominid lineage(Arendt et al., 2017; Rapoport, 1989), which has also been hypothesized to reflect possible developmental constraints contributing to disease risk(Arendt et al., 2017). It may be that, upon more detailed studies (e.g., considering the regulatory networks underlying early brain development), Alzheimer's disease presents a similar scenario of violated constraint leading to increased disease risk later in life. Though we note that

these developmental contributions could be balanced by the activity of many variants of small effect active in adult brain tissue(Keller and Miller, 2006)(see discussion below).

Given the links we observed between development and an age-related disease in Chapter 2, in Chapter 3 we considered the interactions between developmental and aging-related processes at the level of epigenetic regulation, and the importance of regulation for many diseases across the lifespan. The focus on epigenetic interactions is based on the observation that the epigenome plays a key role in orchestrating developmental processes(Kiefer, 2007), and is considered a key hallmark of aging(Brunet and Berger, 2014; López-Otín et al., 2013), while epigenetic studies (e.g., those focusing on epigenetic clocks) suggests that development and aging represent a continuous-progressing process that is recorded in the epigenetic 'record'(Horvath and Raj, 2018b; Raj and Horvath, 2020). By comparing consistent biases in chromatin accessibility across tissues in fetal and adult samples, and comparing these directional biases to those observed when comparing accessibility across young-adult and old-adult samples, we posit that the directionality of epigenetic trends through development are continued as adult individuals age. In support of this, we find that regions biased towards accessibility in fetal and young-adult tissues were enriched for active regulatory annotations (e.g., enhancers, promoters), active chromatin marks (e.g., H3K27ac) as well as bivalent domains(Bernstein et al., 2006), and associated with genes involved in growth processes, all of which are expected for epigenetic patterning to encourage growth and development. In contrast, regions biased towards accessibility in adult and old-adult tissues were enriched for repressed genomic annotations (e.g., heterochromatin, repeat elements), repressive chromatin marks (e.g., H3K9me3), and associated with genes involved in immune processes. These enrichments were also expected given epigenetic and expression observations in different aging tissues, as described in the heterochromatin loss(Lee et al., 2020; Villeponteau, 1997), retrotransposon(Gorbunova et al., 2021), and inflammaging(Nardini et al., 2018; Xia et al., 2016) theories of aging.

The sequence properties of these region sets also pointed to the increased effects of phylogenetic constraint, as well as ancient positive selection, and subsequent purifying selection acting on fetal- and

young-adult-biased regions, whereas adult- and old-adult-biased regions showed patterns more consistent with relaxed selection, and possibly increased permissivity to accumulated mutations. When we considered genome-wide heritability across many aging-related disease GWAS, we found that adult- and old-adult biased regions were enriched for nearby associated variants across the vast majority of diseases, while fetal- and young-adult-biased regions were depleted for variants. Under the model of antagonistic pleiotropy favouring developmental processes we discussed in Chapter 2, it would be expected that variants most accessible during development would be those most associated with heritable disease risk, as would be the expectation under a 'fetal programming' model of aging(Godfrey et al., 2013). Under this form of antagonistic pleiotropy, we would also expect that variants in regulatory regions most active in early development (e.g., the fetal knee joint) would be the most strongly associated with late-onset disease when these variants tolerably violate ancient sequence constraint.

However, our findings in Chapter 3 on a broader set of regulatory regions, that are shared across many tissues suggest that developmental regulatory activity does not account for a disproportional proportion of the genome-wide distribution of risk heritabilities for these diseases. It may be that the model we proposed for osteoarthritis is more unique to skeletal structures or structures acquiring functionality early in life and their compartmentally specific (i.e., modularized) regulatory architecture. Once a structure is patterned using its site-specific regulatory architecture for its final adult configuration in early development, regulatory regions active in the adult tissue have a smaller impact on disease risk – that is, there is an ontogenic constraint on the effects of variants active later in life. While less site-specific regulatory regions controlling general homeostatic processes may still impact tissue decline in the adult structures, as suggested by our findings on osteoarthritis in Chapter 3, the effects of ontogenic constraint dominate. In contrast, those tissues/structures which are more dynamic (e.g., the kidney), have more cross tissue regulatory useage, and therefore have less ontogenic rigidity patterned in early development may stand to be more greatly impacted by variants active in adult tissues. For example, it has been suggested that, given the large number of genes and loci involved in maintaining adult human brain function, that the brain regulatory landscape presents an enormous mutational target wherein many

variants (i.e., those epigenetically active in adult brain tissue) of small effect size spread across the genome can cumulatively contribute to risk for late-onset neurological disorders(Keller and Miller, 2006).

In contrast to models of antagonistic pleiotropy contributions to disease, the mutation accumulation theory would predict that regions most active in late-life should (1) be subjected to the weakest selection pressures due to reduced fitness consequence later in life, (2) accumulate more variants as a result of this reduced sequence constraint, and (3) that the accumulated variants should be associated with late-onset disease risk. Our set of old-adult-biased regions fulfill the first two expectations; however, we found that these regions were not those for which variants had the strongest cross-disease associations, arguing against this model as explaining a disproportionate amount of late-onset disease risk heritability. We first note that our GWAS aggregation method does not take into account effect sizes, such that variants of larger effect for any particular disease may not be proportionally represented – that is, we consider aggregate association signal across diseases. Secondly, we note that our aggregated GWAS data was based on late-onset disease incidence – it may be the case that regions most active in late-life (i.e., contemporaneously with disease onset) may instead be enriched for heritability for disease progression (i.e., individuals with a plethora of disease progressing variants may progress more quickly down disease pathophysiology), though we currently lack the genetic data to fully understand this possibility.

In between models favouring developmental processes and those favouring the effects of mutations active in late-life, we therefore propose a model where the time in which variants have the greatest association with late-onset disease risk is during early adulthood. During this period regulatory variants stand to impact homeostatic processes and contribute to tissue dysfunction – if these variants are most active in early adulthood and gradually lose their activity later into life, then their regulatory effects could accumulate over time, additively contributing to disrupted homeostasis. This would be in contrast to variants most active far later in life, i.e., contempareously with disease onset, where their potential deleterious effects may not have had time to integrate. For example, a regulatory variant may cause a slight down-regulation of a tissue-protective gene. If that variant is most active during development, but

diminishes in activity within adult tissue, this slight down-regulation will have less impact on adult tissue function; similarly, if this variant is most active far later in life, then the effects of this down-regulation have less time to accumulate. Down-regulation of that protective gene as one transitions from early- to late-adulthood would therefore stand to have the greatest impact on tissue function, 'setting the stage' for eventual disease manifestation.

There is substantial literature in the medical field to support the importance of early adulthood processes and functional decline later in life. Numerous studies have linked cardiovascular risk factors measured in mid-life and cognitive decline in late-life(Knopman et al., 2001; Qiu and Fratiglioni, 2015; Whitmer et al., 2005; Yaffe et al., 2021); importantly, a recent study demonstrated a much stronger association with cardiovascular risk factors measured in early adults, and suggested that young adulthood may be the time at which interventions to reduce cognitive decline would be most effective(Yaffe et al., 2021). Additionally, an observational cohort study of UK individuals found that individuals presenting with depressive symptoms in early adulthood had significantly elevated risk for developing physical multimorbidities later in life(Arias-de la Torre et al., 2021), while variability in blood pressure measurements in early adulthood (possibly indicative of cardiac remodelling) has been associated with deleterious changes to myocardial function later in life(Nwabuo et al., 2020). Early adults with diminished lung function are also at elevated risk of developing later comorbidities and suffering from premature death, though this may be driven in part by sub-optimal development of the lung earlier in utero(Agustí et al., 2017).

We would note that the late-onset diseases we focus on are complex traits – that is, the effects of many individual loci spread across the genome likely contribute to disease risk(Pasaniuc and Price, 2016). This organization of genetic contributions is important in the context of late-onset diseases, where it has been suggested that variants of small effect size are inefficiently purged by natural selection, and can go on to cumulatively contribute to disease risk(Keller and Miller, 2006; Wright et al., 2003). Furthermore, this facilitates an increased potential for antagonistic pleiotropy – individual alleles can be selected in favour of different traits at different loci, such that a sizable component of disease risk may be contributed by

different selective forces(Wright et al., 2003). For example, Blekhman et al.,(Blekhman et al., 2008) found that, while Mendelian-disease genes were under purifying selection across species, genes associated with complex diseases were not.

This complex organization of variants is a key component of our findings from Chapters 2 and 3. In Chapter 2, we found evidence to suggest that osteoarthritis risk is made up of a combination of general 'violation of constraint' across the genome, wherein variants arising in constrained regulatory regions are permitted to accumulate, and antagonistic pleiotropy acting in favour of certain traits with the example of selection for height in the *GDF5* locus. Both sets of phenomenon contribute to disease risk given the complex architecture of genomic heritability. In Chapter 3, we considered variants falling widely across the genome, implicitly relying on the complex nature of the late-onset diseases we studied.

In considering the evolution of lifespans across animals, not only can the expression level of a given gene be subject to change, as we focused on in our first two chapters with regards to changes in regulatory networks, but the coding sequence of the protein itself may also be subject to modification(Tejada-Martinez et al., 2022). Given our findings in Chaper 3 focusing on the role of epigenetic processes in aging, and how epigenetic context influences the selective forces acting on genomic sequences, we considered the potential for the actual proteins that mediate these processes to themselves be subject to natural selection. We built upon a previous study which sought to define amino acid substitutions in orthologous proteins that stratify based on the longevity of mammalian species (Farré et al., 2021a), considering the distribution of these 'convergent amino acid substitutions' (CAAS) across protein groups, as well as the individual protein structures themselves. We found that, while epigenetic proteins were depleted of CAAS as a set, CAAS-containing chromatin genes were disproportionately represented in certain activities, such as redox-sensing, while CAAS-containing genes in general were enriched for protein domains involved in fatty acid metabolism, DNA repair and replication, and telomere biology – domains which were similarly enriched for previously-established longevity-associated genes(Tacutu et al., 2018). Additionally, while CAAS positions themselves were depleted for occurrence in the majority of protein domain classes, we observed a significant enrichment within serpin domains. Serpin proteins are

involved in coagulation, which has been previously linked to evolution and lifespan in primates(Muntané et al., 2018), and serpins more generally have been linked with the immune system in modulating inflammatory responses(Bao et al., 2018). This enrichment for evolutionary signal in proteins with this immune-related domain follows with our findings in Chapter 3 highlighting the increased epigenetic (and transcriptomic) activity around immune-associated loci – it may be the case that late-expressing genes involved in immunity are subject to protein sequence evolution. Given that we know that immunity is linked with aging and lifespan(Singh et al., 2019), that immune-related proteins have undergone positive selection in great apes(Tejada-Martinez et al., 2022), and that there exists a relationship between time of expression and evolutionary rate(Cheng and Kirkpatrick, 2021), this scenario is plausible.

Through our analyses, we found that CAAS-containing proteins were over-represented in biochemically interacting pairs, suggesting the possibility that selection for longevity may be acting not only at the level of individual proteins, but at the protein-protein complex level. This also follows with our findings on over-representation of CAAS proteins for certain protein activities – proteins which share complementary protein activities (e.g., DNA binding domains, DNA polymerase domains) might be expected to interact as part of larger complexes (as in the example of the human primase complex we highlighted in Figure 4.3). A more in-depth analysis found that CAAS were over-represented in the annotated interaction domains between protein pairs, and further that these residues were also enriched directly at the interface (i.e., within 5A of the partner protein residues). A recent study of *D. melanogaster* proteins found that protein-protein interaction sites are hotspots for adaptive evolution(Peng et al., 2022), providing suggestive evidence supporting our findings on CAAS enrichments at interfaces. We posit that our results suggest direct modification of interaction dynamics, which could fall under three scenarios: (1) compensatory mutations in both members of a protein pair to preserve function, (2) modification to one member to strengthen or weaken the affinity of a known interaction, or (3) gain or loss of binding partners. We observed an enrichment of instances in which both proteins had CAAS in their respective interaction domains, suggesting the possibility of epistasis acting to preserve the interface (scenario 1 above), as has been suggested to occur for functionally-constrained complexes (Choi et al., 2009; Duarte et al., 2012; Teppa et al., 2017). We also found an enriched occurrence of CAAS falling within one interaction

domain of a pair, which would be congruent with the latter two scenarios. There is evidence in favour of interface mutations changing specificity of binding partner interactions (Lite et al., 2020; Plach et al., 2017; Sikosek and Chan, 2014), with simulations indicating that novel interaction interfaces may be achieved with small numbers of substitutions (Levy, 2010), so it is possible that the CAAS we observe at interfaces are mediating changes to complex formation.

This could represent a means for natural selection favouring longevity to avoid the potential for high fitness costs of protein mutation, as described in the following example based on the CASP9-APAF1 interface (Figure 4.3). If natural selection favoured changes to the regulation of caspase activation (e.g., to enhance cancer suppression(Lowe and Lin, 2000; Tower, 2015)), it could do so either by increasing the activity of CASP9 (e.g., by altering the active site), or by increasing the affinity of the CASP9-APAF1 interaction. It might be that altering the activity of CASP9 could have deleterious pleiotropic consequences (e.g., impacting apoptosis during development), whereas tuning the interaction with the activator APAF1 might reduce the potential for deleterious effects. We suggest that modifying the post-translational regulation of protein activity at the level of protein-protein interactions may represent an additional means of 'fine-tuning' biological processes that natural selection can use to modify phenotypes. This may potentially preclude the need for natural selection to modify protein functional domains (e.g., catalytic sites), for which the fitness costs would vastly limit the mutational landscape available for innovation. In support of this evolutionary concept, it has been found that catalytic residues exhibit far greater evolutionary constraints than protein-protein interaction residues across a large number of enzymes(Jack et al., 2016), results replicated in another study considering evolutionary rate as a function of distance from catalytic and protein-interaction sites(Sharir-Ivry and Xia, 2019). Interestingly, a study of the HoxA11-Foxo1a developmental complex found that changes outside the direct interaction interface can also mediate changes in protein-protein interactions, which the authors suggest may be a means to reduce the pleiotropic effects of protein evolution yet further(Brayer et al., 2011); this could serve as a further generalization of the model we propose.

I summarize the models developed throughout this dissertation in a graphic illustration (see Figure 5.2), using development/maintenance of the human knee and osteoarthritis risk as an exemplar that integrates the effects of both developmental and adult tissue processes. We highlight that these concepts are all complementary with one another; it is possible that each of these scenarios reflects some part of the broader evolutionary program shaping human longevity.

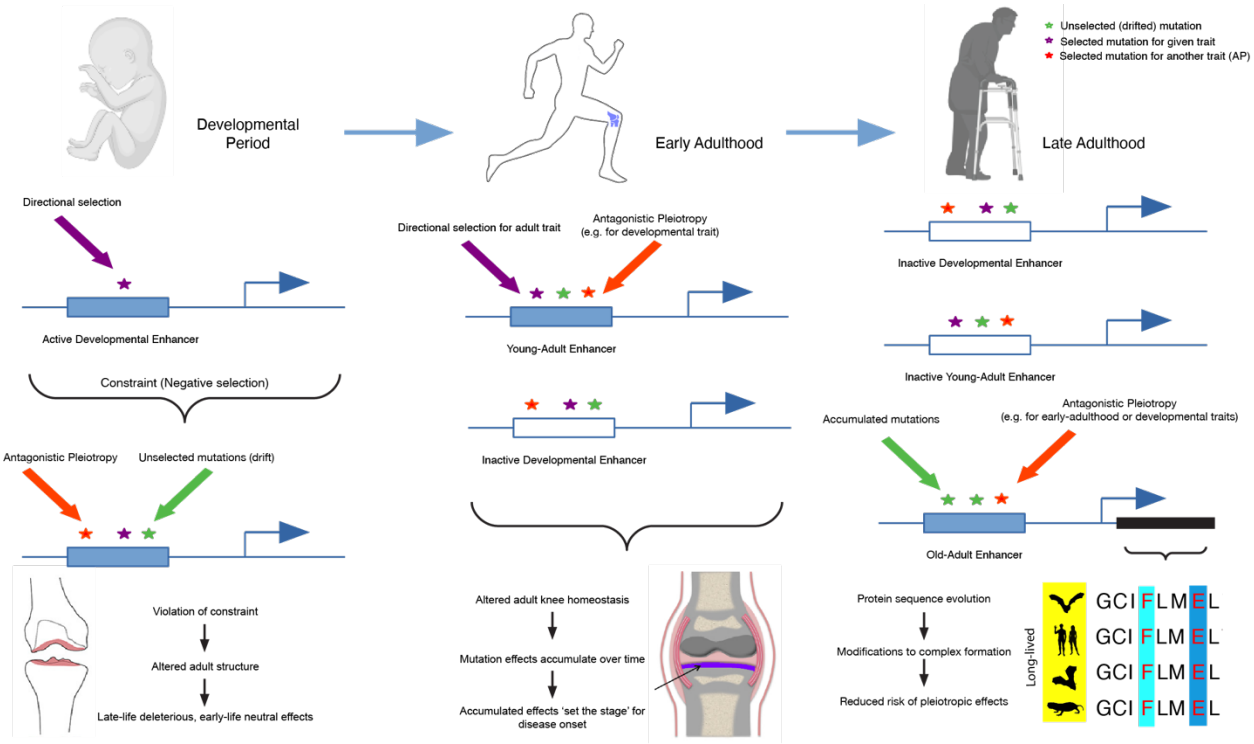


Figure 5.2: Integrative model emergent from studies described in thesis. (Left): Regulatory sequences active in early development are subjected to strong directional, and subsequent purifying, selection in favour of a given trait (here, derived knee morphology). Violation of this constraint alters adult knee structures which increases risk for late-onset disease but has moderate effects on early-life fitness. (Middle) Regulatory sequences active in young adulthood are subjected to a mixture of moderate directional selection (e.g., for some facet of adult knee biology), antagonistic pleiotropy favouring developmental traits, and moderate genetic drift. The effects of variants present in developmental enhancers on adult tissue processes is limited due to diminished epigenetic activity. Variants active in young adulthood stand to integrate their effects over time, cumulatively contributing to increased risk for late-onset disease (OA). (Right) Regulatory sequences active in old adulthood are under the weakest directional/purifying selection, and thus largely impacted by antagonistic pleiotropy favouring other traits

and the permitted accumulation of deleterious variants. However, these regions have moderate associations with disease incidence, possibly given that their epigenetic activity coincides with the timing of disease onset. Protein sequences of late-expressing genes (e.g., immune proteins) are subjected to evolutionary pressures favouring longevity. Selection acting at the level of complex formation is favoured given the decreased potential for pleiotropic consequences.

5.2: Study Caveats:

We acknowledge several limitations in the studies comprising this thesis.

5.2.1: Chapter 2:

Firstly, it is known that gains-of-enhancers in great apes (relative to primates)(Tejada-Martinez et al., 2022; Trizzino et al., 2017), as well as more generally gains/losses of enhancer activity between humans and chimpanzees(Long et al., 2016; McLean et al., 2011; Prabhakar et al., 2008) are associated with both developmental and aging-related gene loci and pathways. Given our use of mouse and human fetal tissues, and our inability for ethical and practical reasons to assess regulation on non-human primate tissues, we are unable to comment on the gain/loss of enhancers either within primates or between humans and chimpanzees. It may be that some of these evolved *cis*-regulatory sequences also contribute to knee development and are subject to unique evolutionary pressures – however, given the consistent patterns we observed when using either mouse-orthologous regions or human chromatin data independently, we maintain that our model is generally applicable. While we suggest that the early-life fitness effects of mutations that 'violate constraint' are moderate, and that late-life negative selection on these mutations is weak, we lack data on the reproductive consequences of such mutations that would allow the sort of inferences between osteoarthritis risk loci and reproductive success that have been made for other late-onset diseases such as coronary heart disease(Byars et al., 2017). Finally, osteoarthritis is a complex disease whose aetiology includes surrounding joint tissue (e.g., the synovium) as well as systemic factors (e.g. obesity, chronic inflammation)(Lee et al., 2013); we do not discount the effects of homeostatic processes active in adult knee tissues in contributing to osteoarthritis risk, but rather suggest it is the additive effects of deleterious alterations to developmental regulation, as well as

adult tissue homeostasis, that drive disease risk, with important implications to the rate of disease progression.

5.2.2 Chapter 3:

We considered DNase-I accessibility as a proxy for chromatin activity in our cross-tissue, cross-timepoint analyses. While it has been suggested that chromatin accessibility reflects the regulatory potential of a locus, integrating several different epigenetic phenomena(Klemm et al., 2019), it is possible that we lose finer-grained regulatory information through this use of a broad epigenetic read-out. We did seek to utilize more specific histone modification (ChIP-seq) information (see Appendix C) to compare fetal and adult tissue samples, finding concordance between the epigenetic annotation enrichments observed using our chromatin accessibility analysis and changes in chromatin state between fetal/adult tissues. However, we lacked sufficient resolution of our chromatin mark datasets to perform a similar young/old-adult comparison. As these datasets continue to be developed, these analyses may be done to confirm concordance with our accessibility results.

While we sought to identify broad trends in epigenetic state between fetal/young, and young/old-adult, and did so by explicitly requiring shared behaviour across our tissue sets, there are important differences in epigenetic state across diverse tissue types(Roadmap Epigenomics Consortium et al., 2015), epigenetic reprogramming in the transition between fetal and adult tissues has a tissue-specific component(Yuen et al., 2011), and certain facets of aging, e.g., ‘epigenetic drift’ also exhibit tissue-specific effects(Teschendorff et al., 2013; Zampieri et al., 2015). As the availability of larger per-tissue sample sizes increases with time, and further single-cell epigenetic datasets (e.g., scATAC-seq) become available, questions as to tissue-specific epigenetic changes that occur over the course of development and aging may be addressed in more detail.

5.2.3 Chapter 4:

Our initial definition of CAAS was based on the comparison of aligned proteins across four long-lived and four short-lived mammals. While additional computational analyses were performed to validate these

residue changes using additional independent sets of mammals (for details see Farre et al 2021), it may still be the case that the set of CAAS we used to perform our analyses may be biased by our species choice. Performing our analyses using alternate sets of CAAS based on additional species would allow us to confirm the robusticity of our results. We found enrichments of CAAS within protein-protein interaction domains, and further refined this by considering residues at the interaction interface. However, we note that we used a simple definition of interaction residues (using a molecular distance cutoff), and thus it may be that CAAS we considered to be interacting residues are not biochemically relevant for binding, and similarly we may be missing residues not directly adjacent to the partner protein that play an important role in mediating binding. More refined predictions of interaction residues using *in-silico* methods, such as BPred(Mukherjee and Zhang, 2011) or JET(Ripoche et al., 2017) may help to refine our interface analysis further. Additionally, while we highlighted two examples of CAAS occurring between interacting proteins, describing the potential effects of these residue changes on binding affinity via both computational predictions and experimental data, a comprehensive picture of the effects of CAAS mutations on protein interactions was not explored. While it is likely the case that the direction and magnitude of predicted effects are specific to the protein pairs and biological processes involved, any global trends may suggest general principles behind the evolutionary forces responsible for establishing these CAAS.

5.3 Future Directions:

In Chapter 2 we focused on the role of derived changes to developmental processes in contributing to osteoarthritis risk; as discussed, it may be the case that the uniqueness of the human knee as a derived feature renders it particularly sensitive to the effects of development. Nevertheless, our findings from Chapter 3, along with knowledge of the complex aetiology of osteoarthritis(Lee et al., 2013), suggest that a focus on the contributions of homeostatic processes in the adult knee tissue to osteoarthritis risk is also warranted. Epigenetic and transcriptomic datasets derived from adult knee samples have started to become available(Liu et al., 2018), although these are opportunistically sourced and thus skewed towards much older individuals. Using these datasets, as well as new datasets in the future from younger-adult

samples, would allow a comprehensive analysis comparing fetal- and adult-tissue associations with osteoarthritis, incorporating the methodology and concepts developed over the course of this dissertation.

As mentioned briefly above, expanding the cross-tissue, cross-time analyses performed in Chapter 3 to focus on tissue-specific epigenetic processes is an additional avenue of future research. With the increased availability of tissue-specific datasets from the ENCODE project(Abascal et al., 2020), as well as the potential for increased resolution at the level of single-cell epigenomic datasets across tissues (similar to that which has already been generated at the transcriptomic level(Jones et al., 2022)), future studies could elucidate development- and ageing-associated epigenetic changes – particularly in exploring whether the evolutionary model we developed in Chapter 3 holds when considering shifts in epigenetic context specific to individual tissues, as well as the potential interaction between tissue-specificity and diseases with obvious tissue biases (e.g., neurological disorders, diabetes, etc.). We would also note the potential in this model for the role of antagonistic pleitropy acting across adulthood – that is, variants that confer some benefit to tissue function in early adulthood could have deleterious consequences in late-life, either directly (e.g., altering regulatory networks in old-adult tissues) or indirectly (e.g., mediating tissue changes that impact late-life function). An increased resolution in the number and tissue representation of young/old-adult samples will help to elucidate the possible effects of adult-tissue antagonistic pleiotropy.

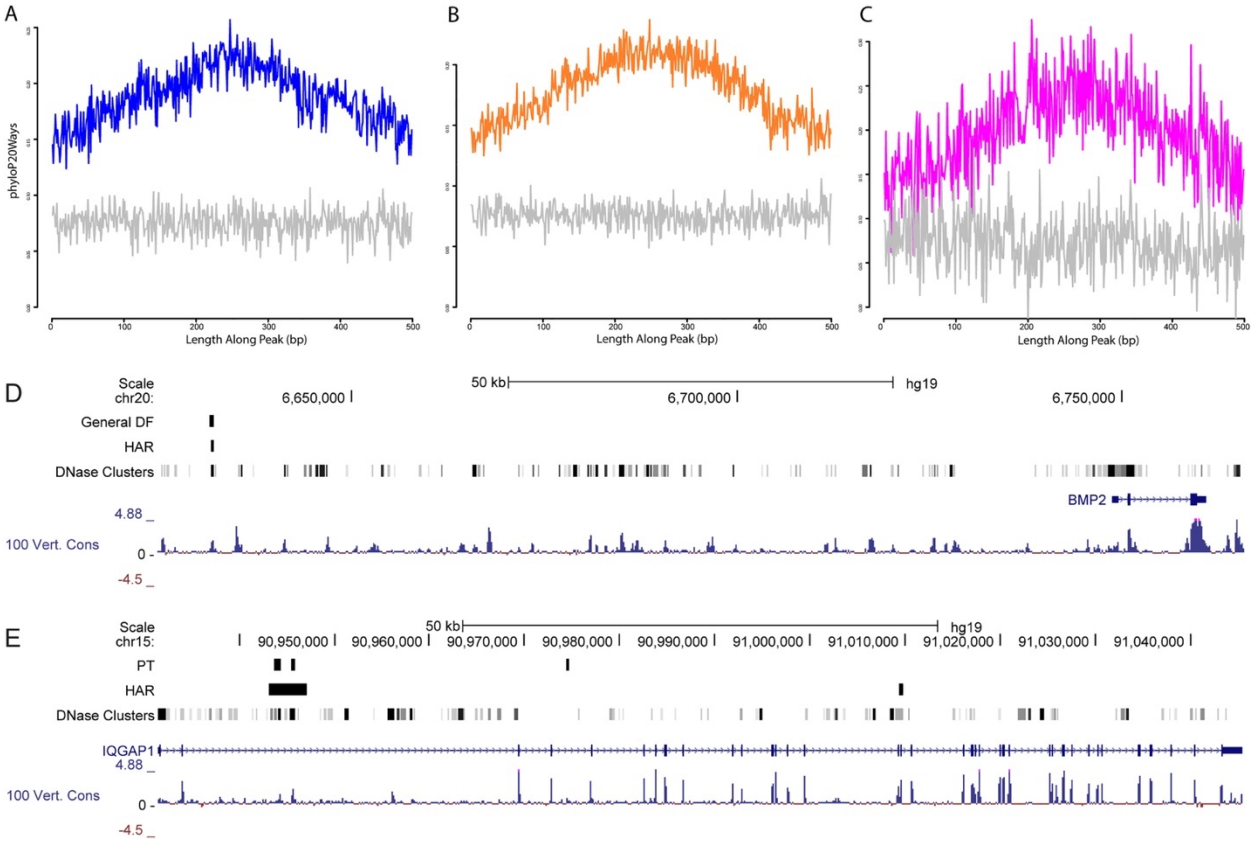
Our results presented in Chapter 4, which lead us to suggest a possible mode of natural selection action which favours fine-tuning of protein-protein interactions rather than protein functional domains, present a future direction of research focused on the role of antagonistic pleiotropy in shifting the preferred evolutionary landscape for modifications. It may be that preserving the functional domains of a certain protein (e.g., CASP9) is more important for developmental processes (e.g., developmental apoptosis), whereas modifying the interaction domain of that protein for a given partner is more beneficial for aging-related processes (e.g., cancer suppression). Given the biophysical constraints of protein structure and function, this could establish instances where protein changes favoring development come at the expense of aging processes, and vice-versa. Exploring the distribution of such evolutionary trade-off

scenarios across a broad set of proteins could lead us to a broader understanding of the role of antagonistic pleiotropy in shaping protein evolution in long-lived species.

5.4 Conclusion:

This dissertation sought to explore the evolutionary basis for the emergence of extended lifespan in humans, and through our analyses developed three main models for how natural selection favouring longevity can shape (1) regulatory sequences active during development, (2) regulatory sequences whose activity changes as a function of developmental/aging processes, and (3) the sequences of the aging-related proteins themselves (Figure 5.2). This work improves our conceptual understanding for the means by which natural selection may operate, and points towards biological processes and their associated loci which may be prime follow-up candidates for further elucidation – particularly with regards to functional validation. As we progress to a more complete understanding of the evolutionary pathways taken to achieve longevity, the integration of additional experimental data to bridge the divide between models, predictions and biological phenomenon will be a key progression in the field.

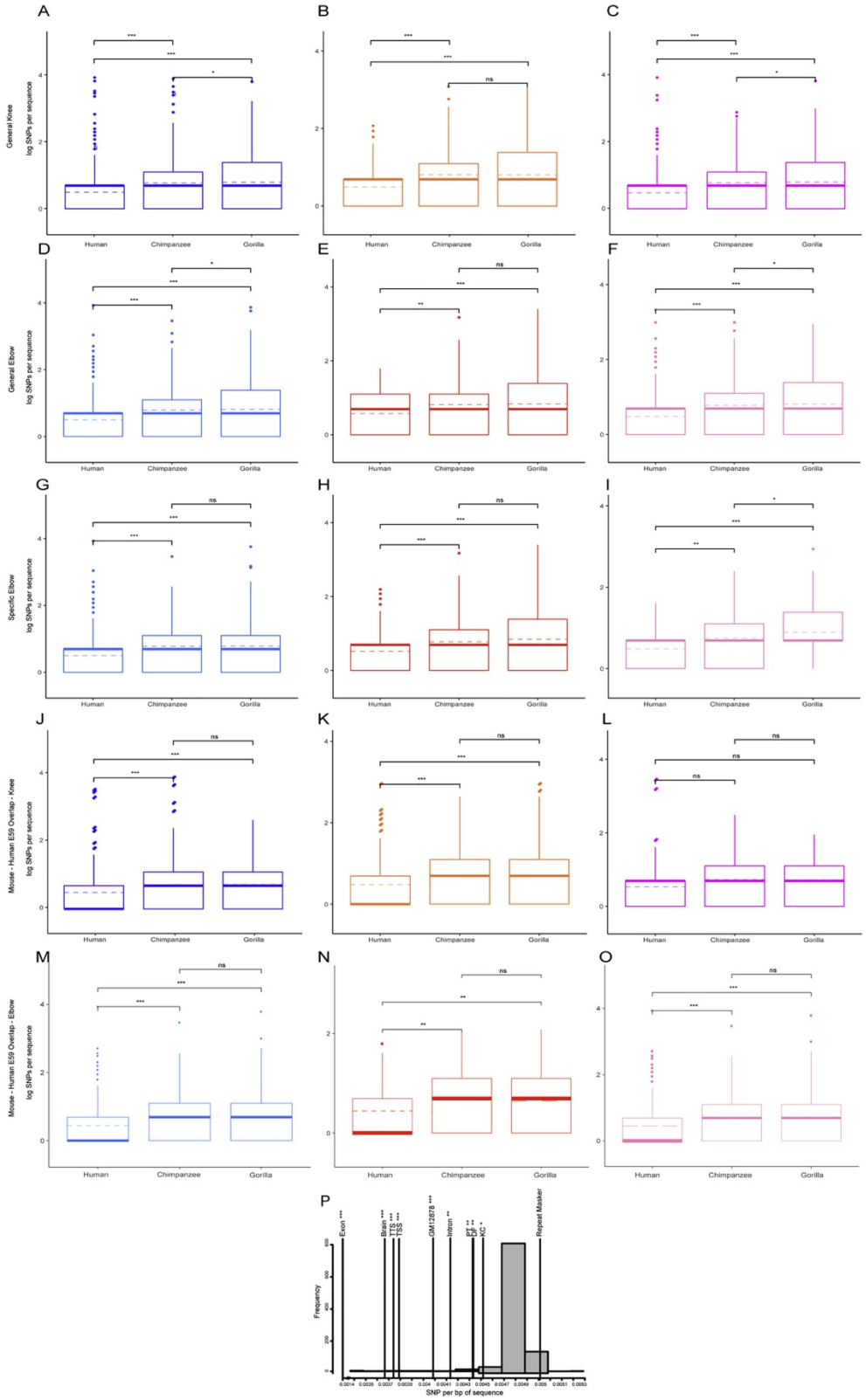
Appendix A: Supplemental Figures



Supplemental Figure 2.1: Related to Figure 2.1

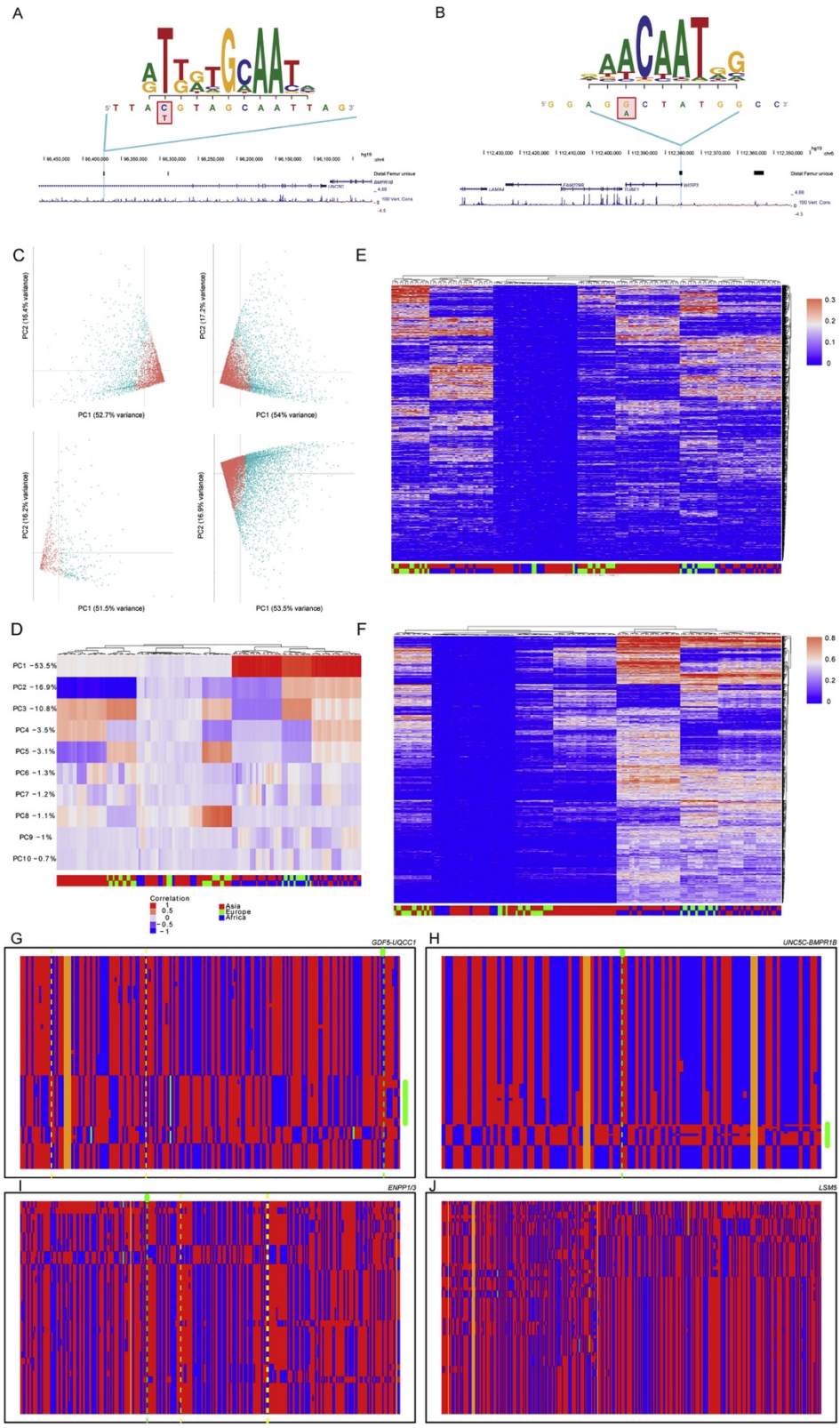
(A-C): phyloP scores (phyloP20ways) were averaged per-bp across all specific-knee ATAC-seq regions in a peak set and plotted in respective colors as a function of distance along a 1kb window centered on the peak middle; a random background distribution of conservation scores is indicated in gray (see Methods). (A) Specific Distal Femur (B) Specific Proximal Tibia (C) Specific Knee-common. (D-E): HAR enrichment gene loci screenshots. (D) A general distal-femur element overlaps a human-accelerated region (HAR) (Prabhakar et al., 2006) upstream of BMP2, a TGF-β family member with a well-established role in chondrogenic differentiation, that is expressed in the knee (Rosen, 2009) and which been linked to

osteoarthritis (Gamer et al., 2018, Valdes et al., 2006). ATAC-seq regions (“General DF”), ENCODE DNase hypersensitivity sites and phyloP100ways conservation tracks are shown on bottom. (E) A proximal-tibia-specific element overlaps an accelerated region (Bird et al., 2007) intronic to IQGAP1, a protein complex scaffold (Hedman et al., 2015) expressed in embryonic mouse cartilage (Cupit et al., 2004) which impacts skeletal development, particularly in the tibia (Dickinson et al., 2016). Format as in (D), with ATAC-seq track “PT” shown.



Supplemental Figure S2.2: Related to Figure 3.

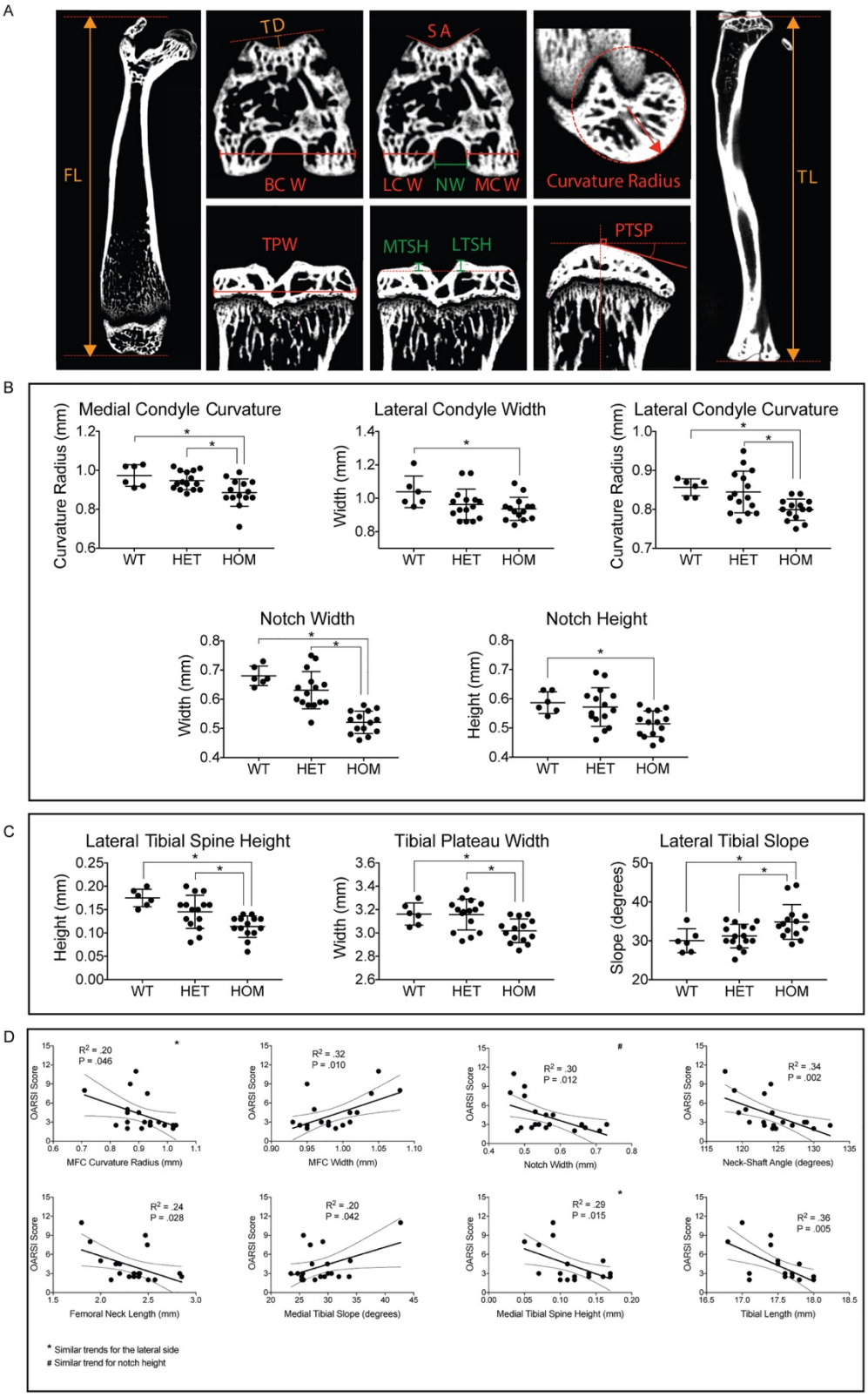
Intra-species sequence diversity in humans and apes. The number of intra-species variants intersecting a given ATAC-seq peak were counted for all variable sequences in a set within a given species and expressed as ‘SNPs per sequence’. Mean and median values of distributions are indicated in dashed and bold lines, respectively. Supplemental Table S2.5 lists numerical values for each comparison. (A) General Knee Distal Femur, (B) General Knee Proximal Tibia, (C) General Knee Common, (D) General Elbow Distal Humerus, (E) General Elbow Proximal Radius, (F) General Elbow Common, (G) Elbow Distal Humerus-Specific, (H) Elbow Proximal Radius-Specific, (I) Elbow-Common-Specific. (J) Knee Distal Femur-Specific, overlapping human E59 Distal Femur data. (K) Knee Proximal Tibia-Specific, overlapping human E59 Proximal Tibia data. (L) Knee-Common-Specific, overlapping pooled human E59 Distal Femur and Proximal Tibia data. (M) Elbow Distal Humerus-Specific, overlapping human E59 Distal Humerus data. (N) Elbow Proximal Radius-Specific, overlapping human E59 Proximal Radius data. (O) Elbow-Common-Specific, overlapping pooled human E59 Distal Humerus and Proximal Radius data. (P) Counts of common chimpanzee variants (via the GADP dataset) per bp of sequence for element sets were compared to random region sets along with other genomic features for enrichment/depletion analysis; labels correspond to results in Supplemental Table S2.5. Significance codes: not significant (ns), < 0.05 (∗), < 0.01 (∗∗), < 1e-5 (∗∗∗).



Supplemental Figure 2.3: Related to Figure 2.4

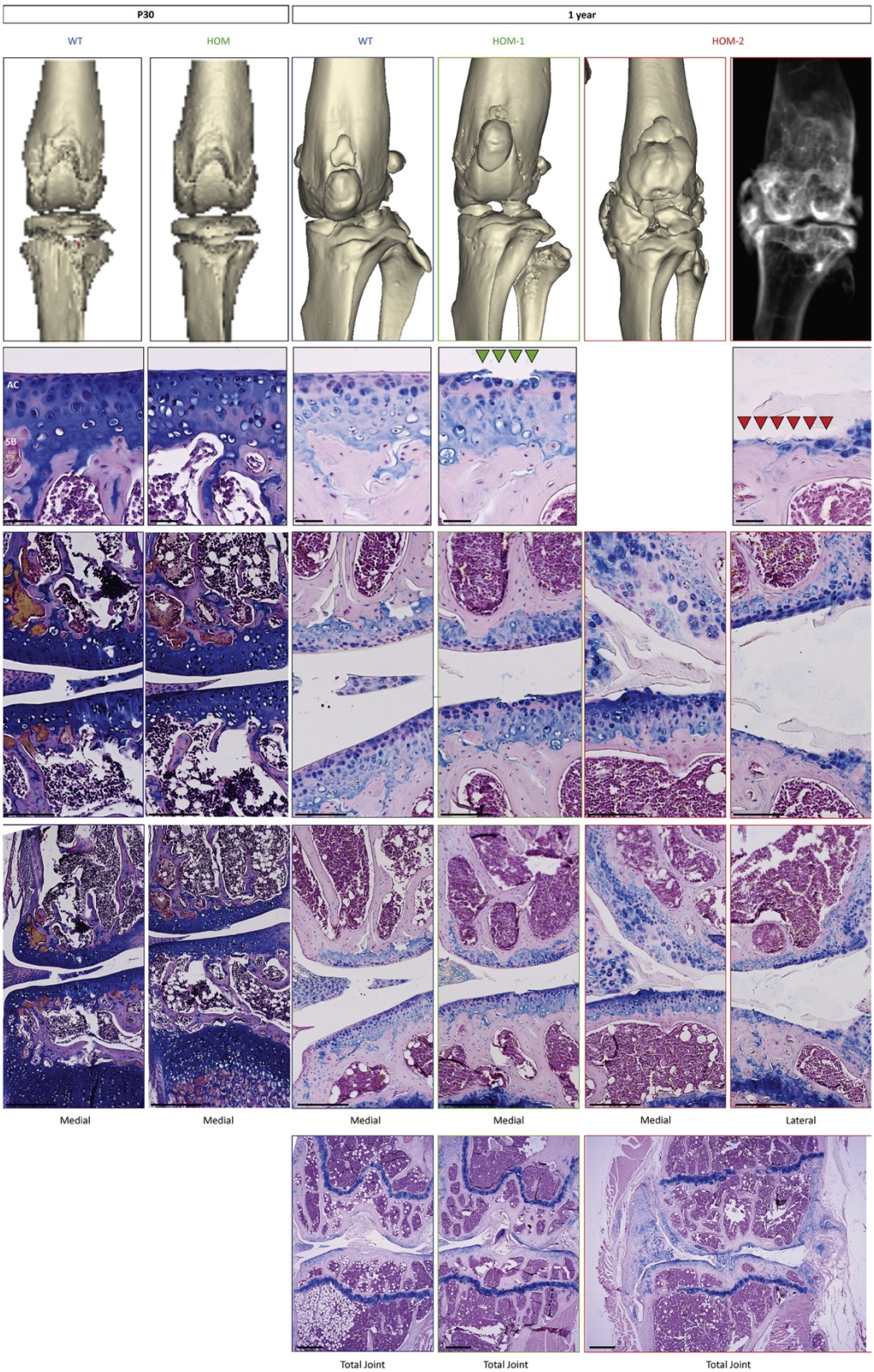
Examining modern human sequence variation in knee sets. (A-B): Motif alteration in putative regulatory elements near osteoarthritis-related genes. Red shading indicates altered base relative to motif logo, blue shading below indicates genomic position of sequence. ATAC-seq regions and phyloP100ways conservation tracks are shown on bottom. (A) Common human variant (rs265053: C/T) within an intron of UNC5C, improves a predicted CEBPB sequence motif downstream of BMPR1B (Baugé et al., 2013, Zhai et al., 2015). (B) Common human variant (rs2280153: G/A) improves a predicted SOX9 sequence motif within the WISP3 promoter (Kannu et al., 2009, Sen et al., 2004). (C) Common human variants within the indicated peak sets distributed across the first two PCs of an Fst-based PCA analysis. Two sub-groups based on k-means clustering are indicated in red (Cluster 1) and blue (Cluster 2). Variance explained by each PC is indicated on axes. (D) Representative correlation heatmap between PCs and population comparisons across all variants in the knee-specific pooled set. Pearson's correlation between PCs and population pairs are clustered, with populations colored by continent: red - Asia, green - Europe, blue - Africa. (E, F) Calculated Fst for common variants across populations. Fst values for a single variant across multiple pairwise population comparisons are shown, with each row representing a single variant; the particular pairwise comparison is indicated below, color scheme as for (D). Red/blue scale for Fst values is indicated on far right. Both variants and pairwise comparisons are grouped by hierarchical clustering. (E) Cluster 1 variants within the knee-specific pooled set. (F) Cluster 2 variants within the knee-specific pooled set. (G-J) Visual genotype plots for selected populations in a given locus. SNP sequences for all individuals are arranged according to a maximum-likelihood tree. Red/Blue indicates ancestral-derived assignments, respectively. Solid orange lines indicate SNPs which are gapped in chimpanzee, solid green lines indicates individuals carrying alleles besides the ancestral/derived annotations. Vertical yellow dashed lines indicate proxy variants, vertical green dashed lines indicate lead variants, when present in indicated region. For plots with a distinguishable haplotype (G, H) the alternate haplotype is indicated with a thick green line. (G) Visual genotypes for selected population (Volga-Ural region of Russia) (Capellini et al., 2017) within a ~150kb haploblock in the GDF5-UQCC1 locus. Distribution of genotypes indicates the presence of a high-frequency haploblock carrying the lead GWAS variant "A" at rs143383 (Miyamoto et al., 2007, Valdes et al., 2011) and linked variants (rs4911178, rs6060369). (H) Visual genotypes for selected population (South-East Asia) within a ~64kb haploblock in

the UNC5C-BMPR1B locus; variant rs2626053 (arcOGEN Consortium et al., 2012) appears on a low-frequency haploblock. (I) Visual genotypes for selected population (South-Western Europe) in the ENPP1/3 locus; given the lack of clearly-defined haploblock structure in the locus, a 200kb region centered on rs3850251 (arcOGEN Consortium et al., 2012, Klein et al., 2019) was used to generate this plot, with linked rs7744039, rs7773292, rs9493095 indicated. (J) Visual genotypes for selected population (Central Siberia) in the LSM5 locus; given the lack of clearly-defined haploblock structure in the locus, a 200kb region centered on rs4141788, linked to rs7785659 (arcOGEN Consortium et al., 2012), was used to generate this plot.



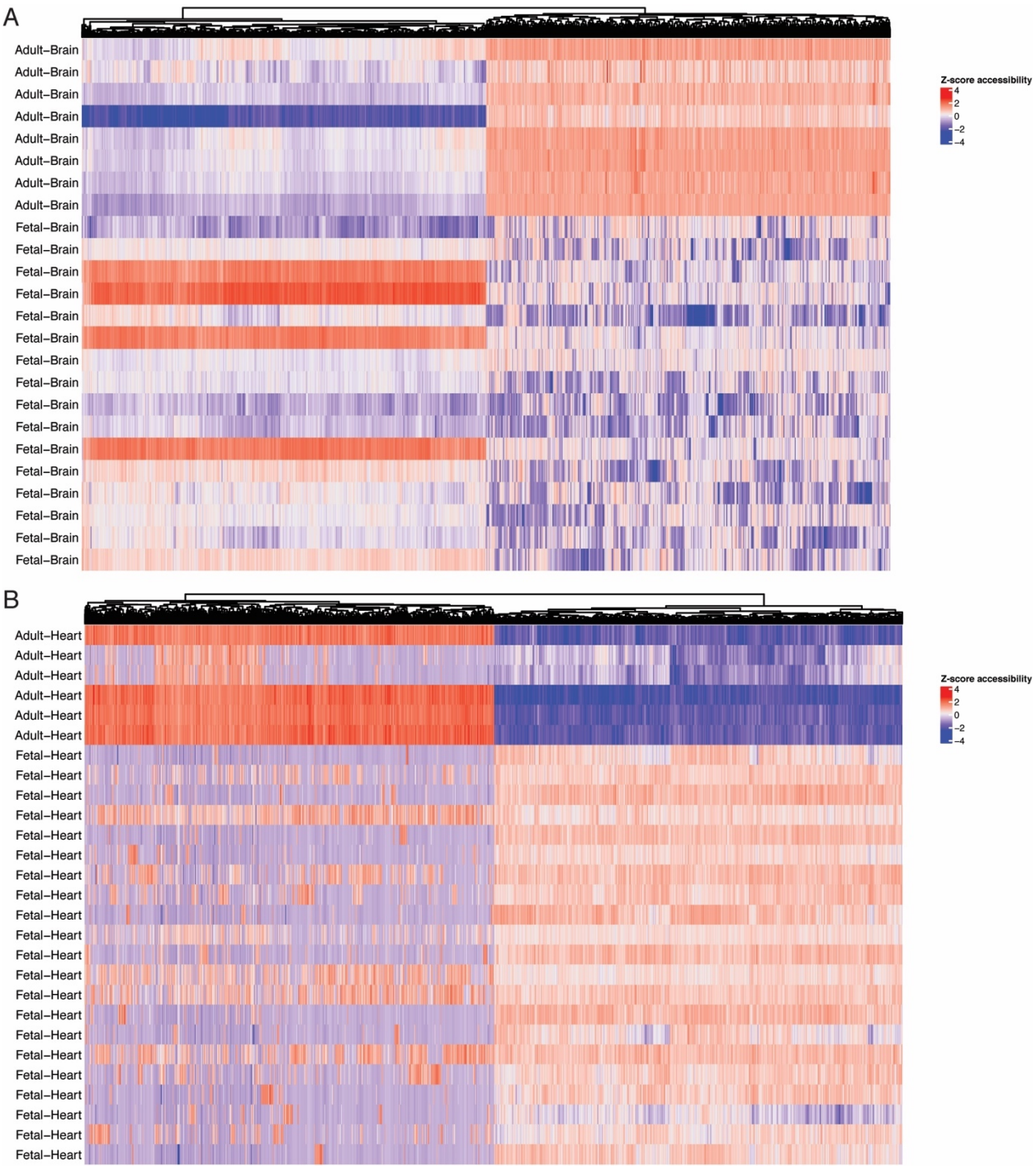
Supplemental Figure S2.4: Related to Figures 2.5 and 2.7.

Additional morphometric analyses on R4 enhancer null mice. (A) Quantified anatomical indices of the femur and tibia used in this study. Femoral length (FL), bicondylar width of the femur (BCW), trochlear groove depth (TD), width of the lateral femoral condyle (LCW), width of the medial femoral condyle (MCW), intercondylar notch width (NW), trochlear groove sulcus angle (SA), width of the tibial plateau (TPW), height of the medial tibial spine (MTSH), height of the lateral tibial spine (LTSH), posterior slope of the tibial plateau (PSTP), and tibial length (TL). Curvature radius of the femoral condyle and the posterior slope of the tibial plateau were measured across both medial and lateral compartments. (B,C) Morphological defects in 1 year old R4 enhancer null mice (HOM), compared to heterozygous mice (HET) and wild-type mice (WT). A number of measurements were carried out at 1 year of age (see (A)). This figure displays only those which revealed significant differences between control and R4 enhancer null mice in distal femur (B) and proximal tibia (C) structures. See Supplemental Table S2.9 for complete results of statistical comparisons for all measurements as well as significance values for each comparison. (D) Correlation between OARSI scores and morphometric measures of the distal femur and proximal tibia on R4 enhancer mice. Correlation (R2) and p value results for Pearson's correlation tests are indicated on each graph.



Supplemental Figure S2.5: Related to Figure 5.

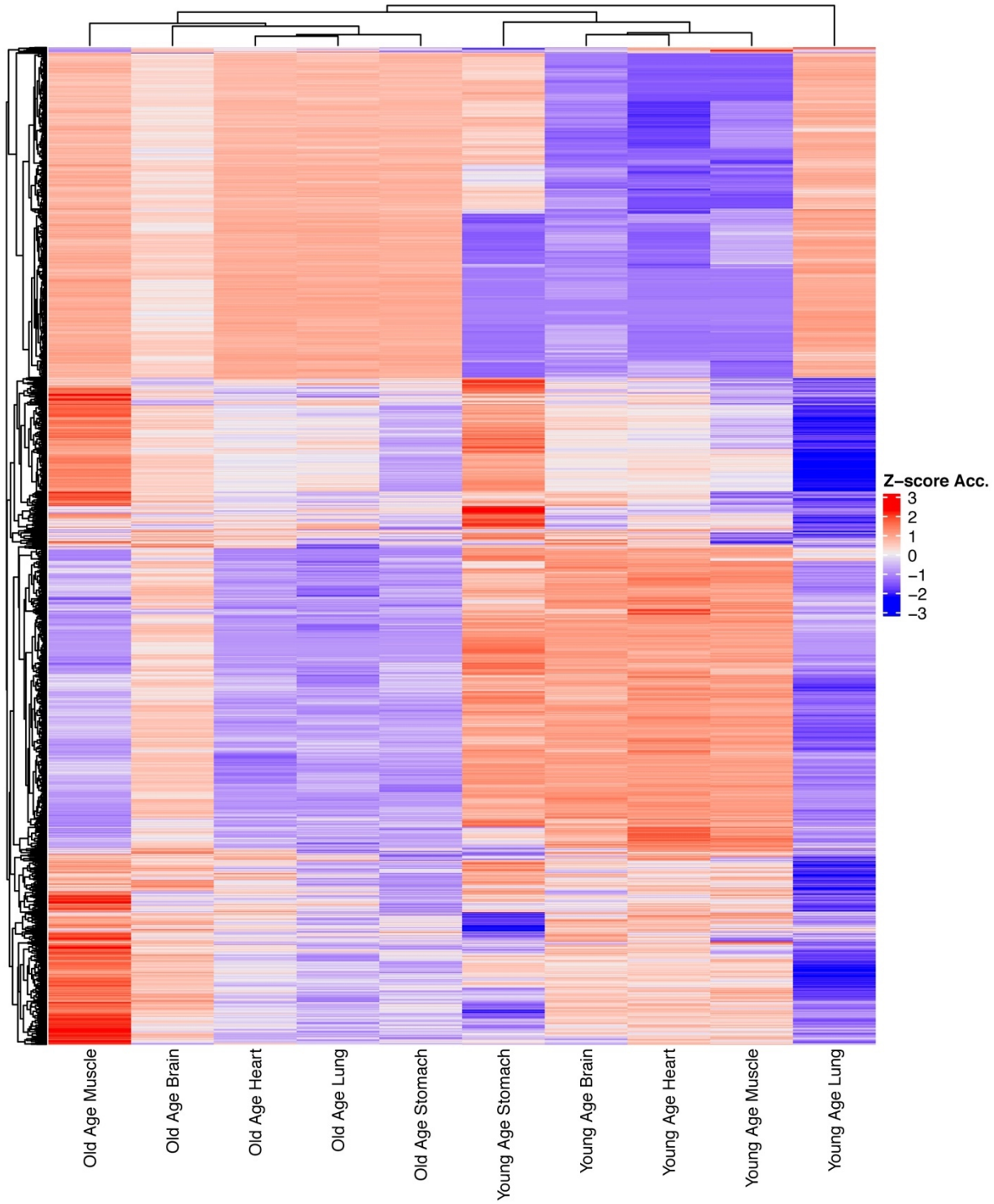
Additional histological sections on R4 enhancer null mice at P30 and 1 year. (Left) Comparison of WT and HOM knee joints at P30, showing microCT renditions (top row), high-magnification images of sections of the unaffected knee joint in both genotypes as reported in Figure 5 (second row), and low-magnification images of the of same medial sections but at different planes (third and fourth rows). (Right) Comparison of WT and two HOM knee joints of differing phenotypic severity (HOM-1 mild osteoarthritis (green); HOM-2 severe osteoarthritis (red)) at 1 year, showing microCT renditions (top row), including an X-ray image of HOM-2 specimen with heterotopic ossification (top row, far right), high-magnification images showing effected medial and lateral sections (when effected) as reported in Figure 2.5 (second row), low-magnification images of the same medial and lateral sections of the joint but in two different planes (third and fourth rows), and total joint images for all three specimens (fifth row). Note, the loss of articular cartilage matrix in the lateral compartment and clustering and loss of cells in the cartilage of the medial compartment in HOM knees. Scale bars, 50 μm (second row); 500 μm (third/fourth rows); 250 μm (fifth row).



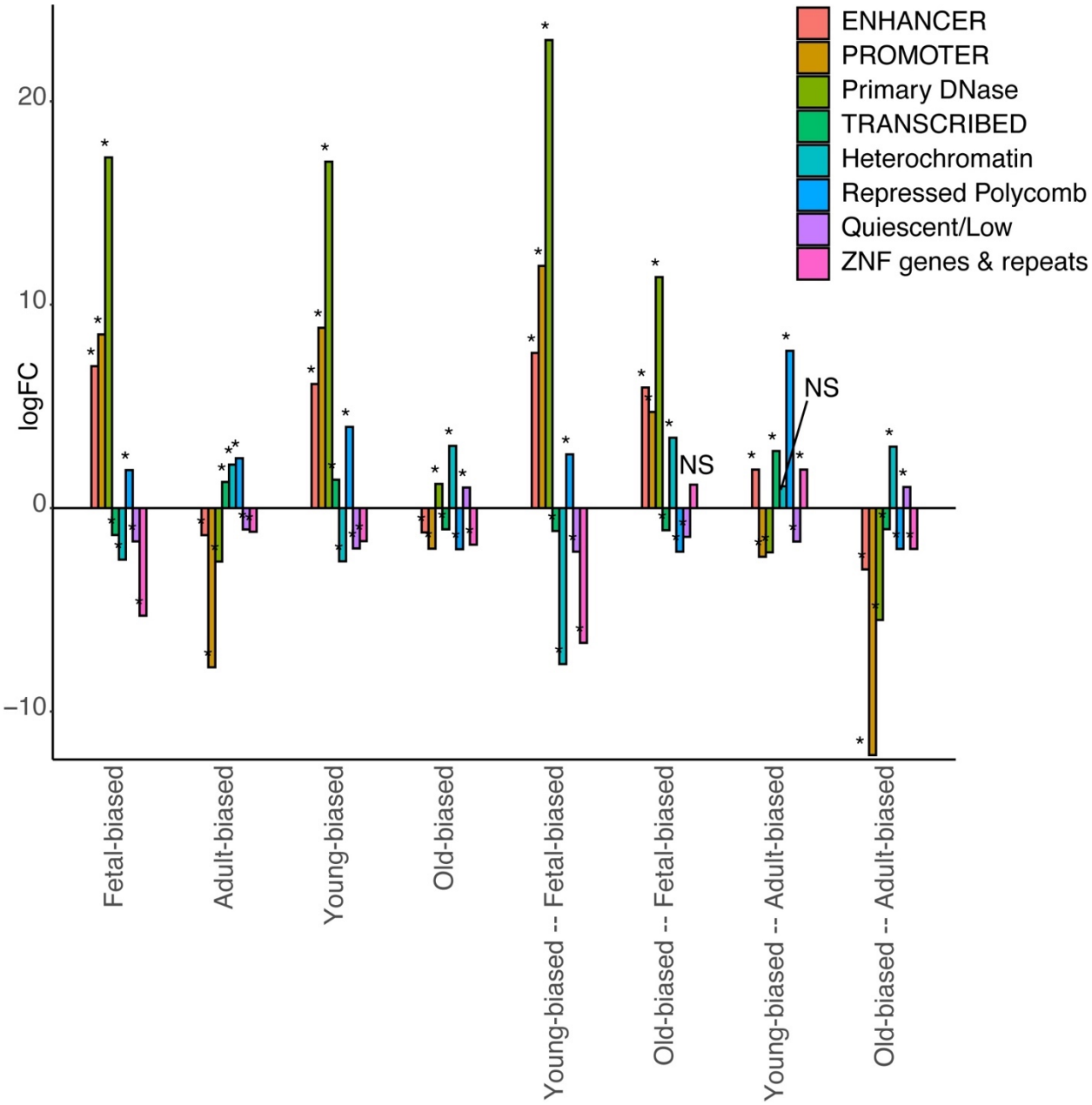
Supplemental Figure 3.1: Per-tissue heatmaps for brain and heart tissue. (A) Z-score accessibility values for regions defined as differentially-accessible comparing fetal and adult brain tissue samples. (B) Z-score accessibility values for regions defined as differentially-accessible comparing fetal and adult heart tissue samples. Red-blue colour scale indicates increased/decreased accessibility, z-score normalized per-column.

Supplemental Figure 3.2: Autosome distribution of accessibility-altered regions. Genomic distribution of regions changing accessibility in fetal/adult comparison. Red/blue: density of defined differentially-

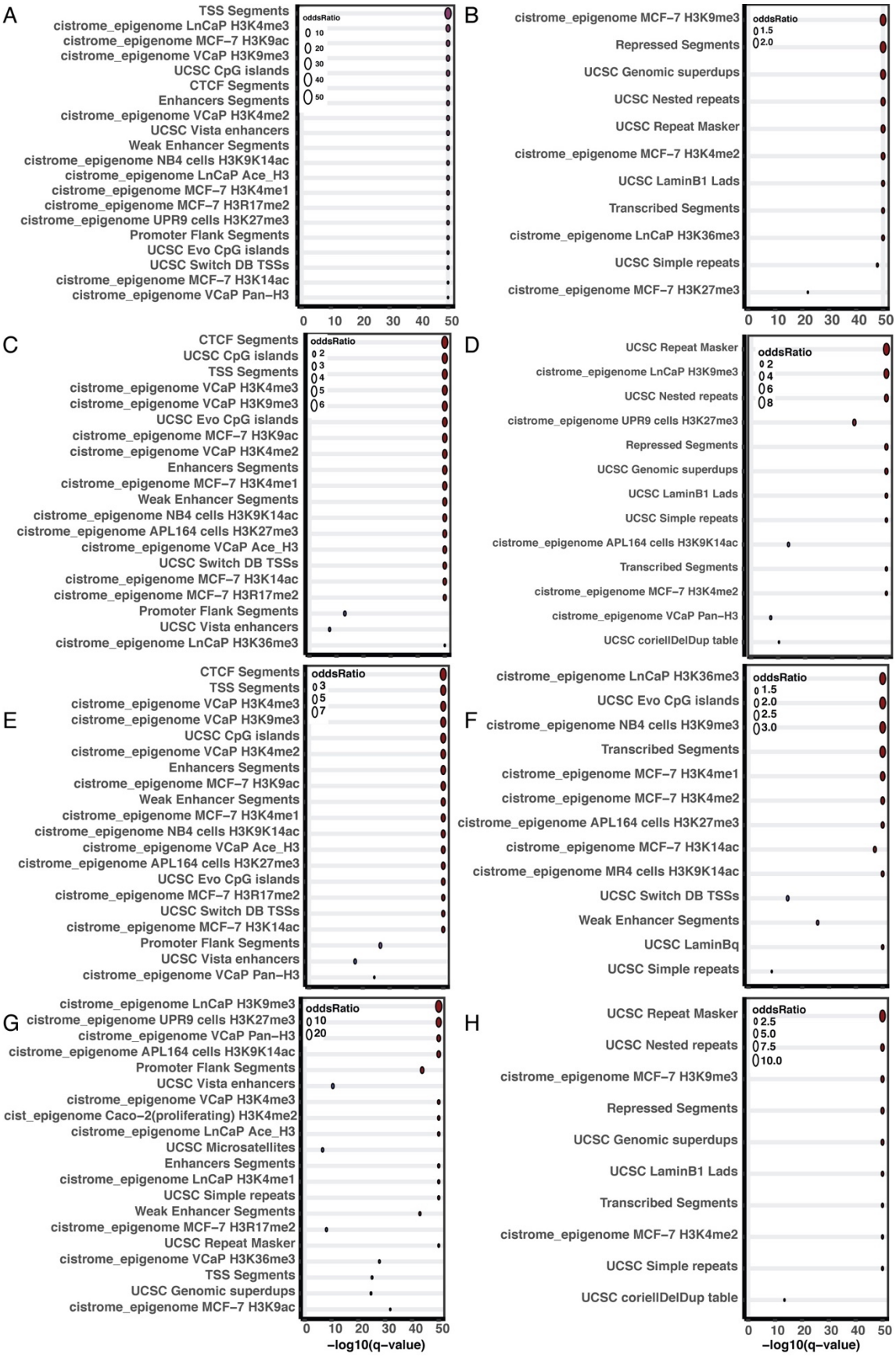
accessible regions. Line: relative proportion of regions more accessible in adult (top) or fetal (bottom) tissues.



Supplemental Figure 3.3: Changes in regional accessibility across young-age and old-age adult tissue samples. Equivalent to Figure 3.1A. Red-blue colour scale indicates increased/decreased accessibility, z-score normalized per-column.

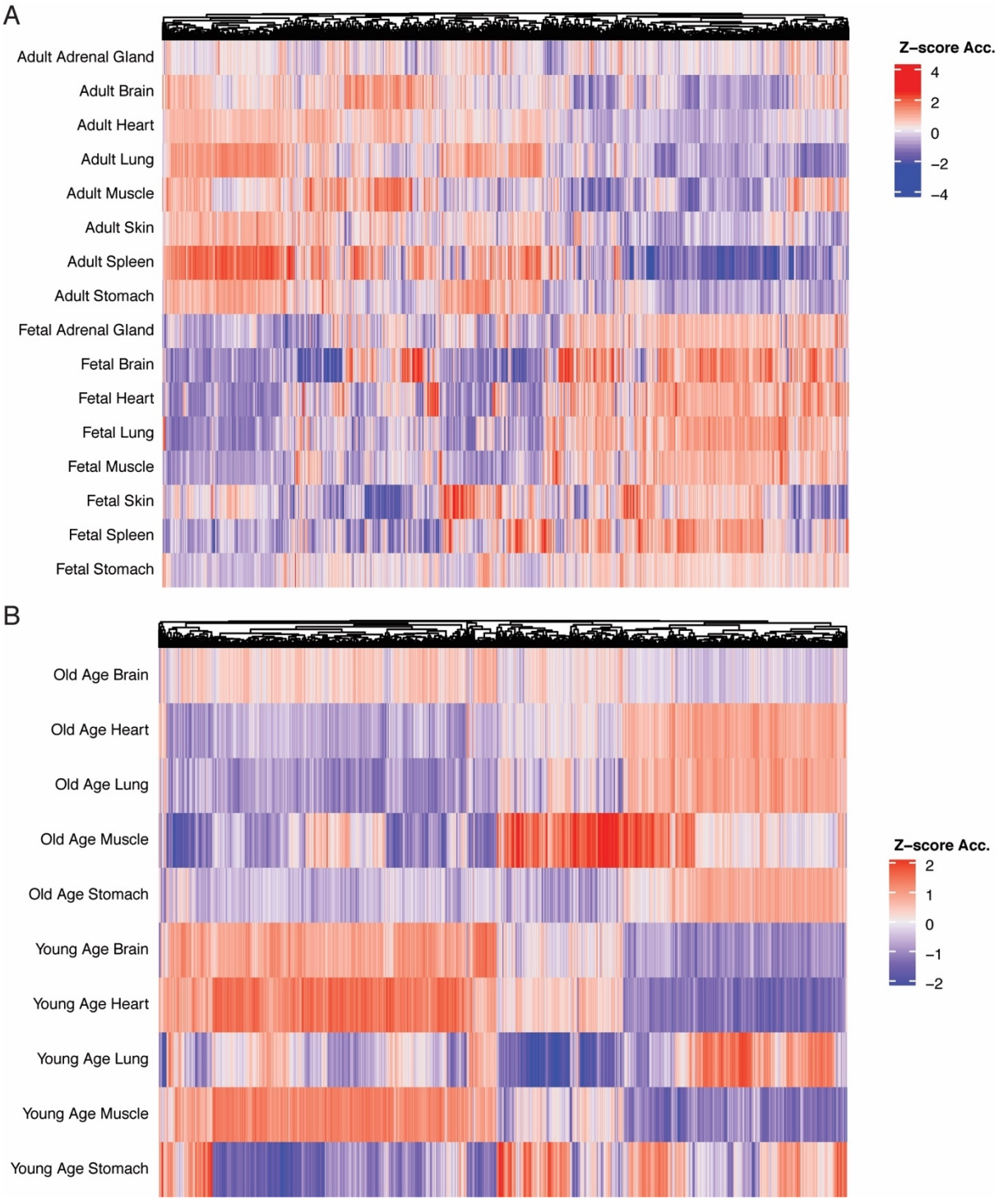


Supplemental Figure 3.4: Enrichment for epigenetic states. Bar plots indicate log10 logFC enrichment/depletion values for different region sets (e.g., young-biased regions) falling within different Roadmap HMM-annotated epigenetic states. Asterik (*) indicates significant hypergeometric test for enrichment/depletion of an indicated region set for indicated epigenetic state ($p < 0.05$). Intersection sets (e.g., young-biased, fetal-biased intersected regions) are indicated with "—".

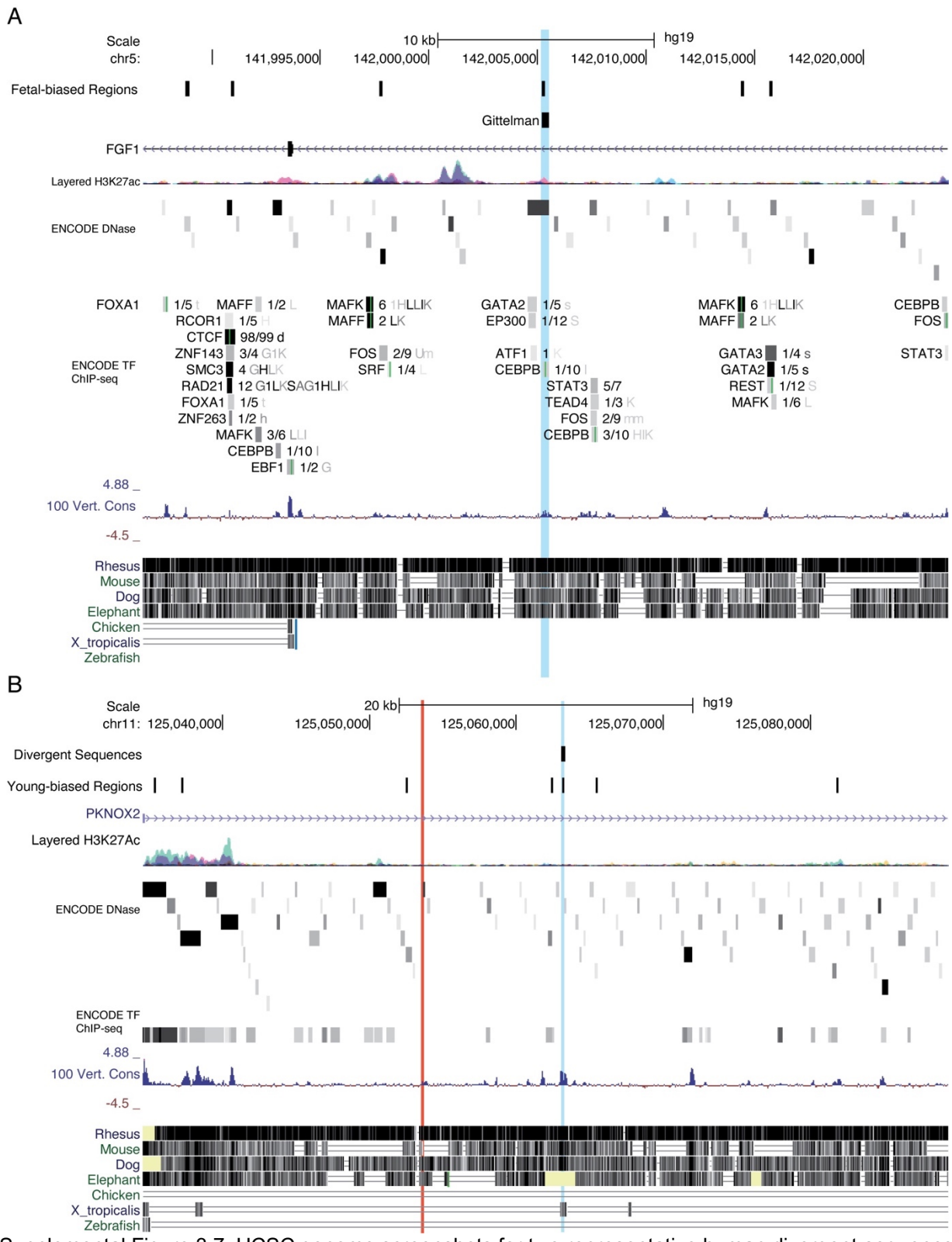


Supplemental Figure 3.5: LOLA enrichment plots. Enrichment q-values for top terms in the LOLA

regional-enrichment analyses. (A) Fetal-biased Regions (B) Adult-biased Regions (C) Young-biased Regions (D) Old-biased Regions (E) Young-biased – Fetal-biased Regions (F) Young-biased– Adult-biased Regions (G) Old-biased – Fetal-biased Regions (H) Old-biased – Adult-biased Regions. All listed terms are significant, q-value < 0.05. See also Supplementary Table 3.2.

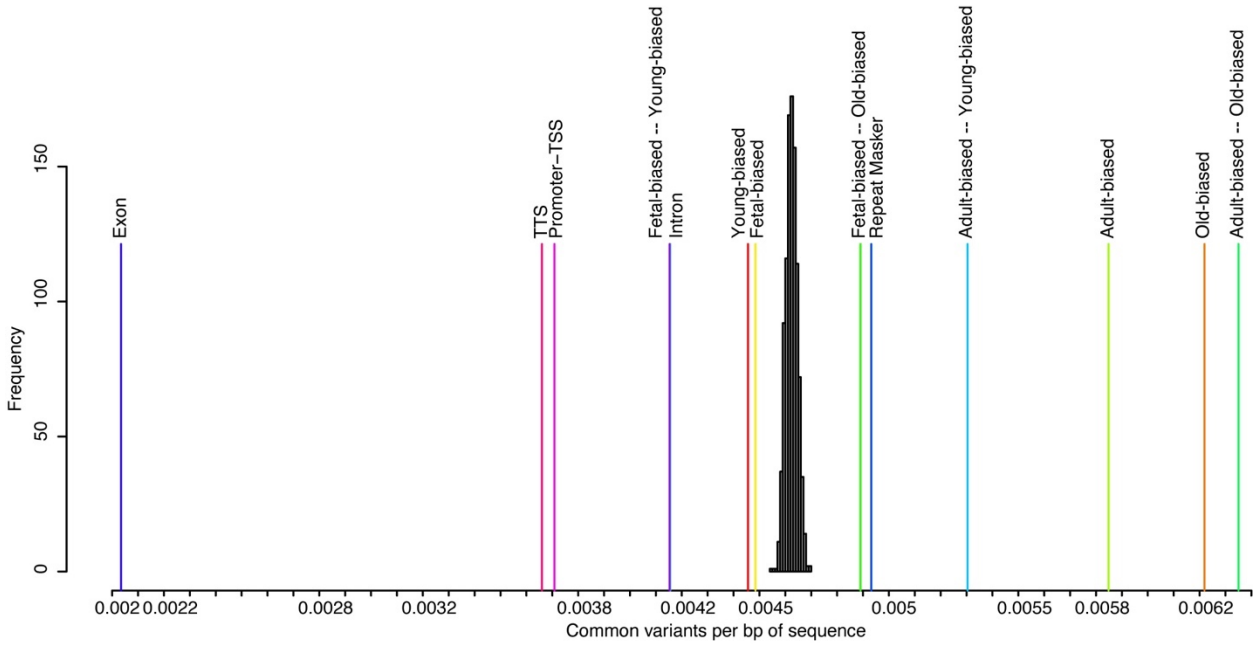

A
Adult Adrenal Gland
Adult Brain
Adult Heart
Adult Lung
Adult Muscle
Adult Skin
Adult Spleen
Adult Stomach
Fetal Adrenal Gland
Fetal Brain
Fetal Heart
Fetal Lung
Fetal Muscle
Fetal Skin
Fetal Spleen
Fetal Stomach
Z−score Acc.
4
2
0
−2
−4
B
Old Age Brain
Old Age Heart
Old Age Lung
Old Age Muscle
Old Age Stomach
Young Age Brain
Young Age Heart
Young Age Lung
Young Age Muscle
Young Age Stomach
Z−score Acc.
2
1
0
−1
−2

Supplemental Figure 3.6: Promoter accessibility heatmaps. Red-blue colour scale indicates increased/decreased accessibility, z-score normalized per-column. (A) Promoter accessibility differences between fetal and adult tissue samples, for significantly-altered promoters (adj. p-val < 0.05). (B) Promoter accessibility differences between young-age and old-age adult tissue samples, for significantly-altered promoters (adj. p-val < 0.05).



Supplemental Figure 3.7: UCSC genome screenshots for two representative human divergent-sequence loci. Additional tracks (top to bottom): Layered H3K27ac signal from ENCODE datasets, layered DNase-I

hypersensitivity sites from ENCODE datasets, aggregated ENCODE transcription-factor ChIP-seq data, phyloP100ways conservation track (per-bp), multiple-sequence alignment to human reference sequence. (A) A human-accelerated region (top track – highlighted in light blue) intersects a region losing accessibility in adult tissue (bottom track) intronic to the FGF1 gene (and which also has promoter-capture data to suggest promoter contact). Also intersects a possible CEBPB binding site (ENCODE TF-ChIP-seq track). (B) A human-accelerated region (top track - highlighted in light blue) intersects a region losing accessibility in old-age adult tissue (bottom track) intronic to the PKNOX2 gene (and which also has promoter-capture data to suggest promoter contact). Upstream of this region lies the variant rs590211(highlighted in red), which has been associated with human-longevity via GWAS studies.



Supplemental Figure S3.8: Chimpanzee genomic distribution plot. Counts of chimpanzee common variants per bp of sequence for region sets were compared to random region sets along with other genomic features; labels correspond to results in Supplementary Table S3.3.

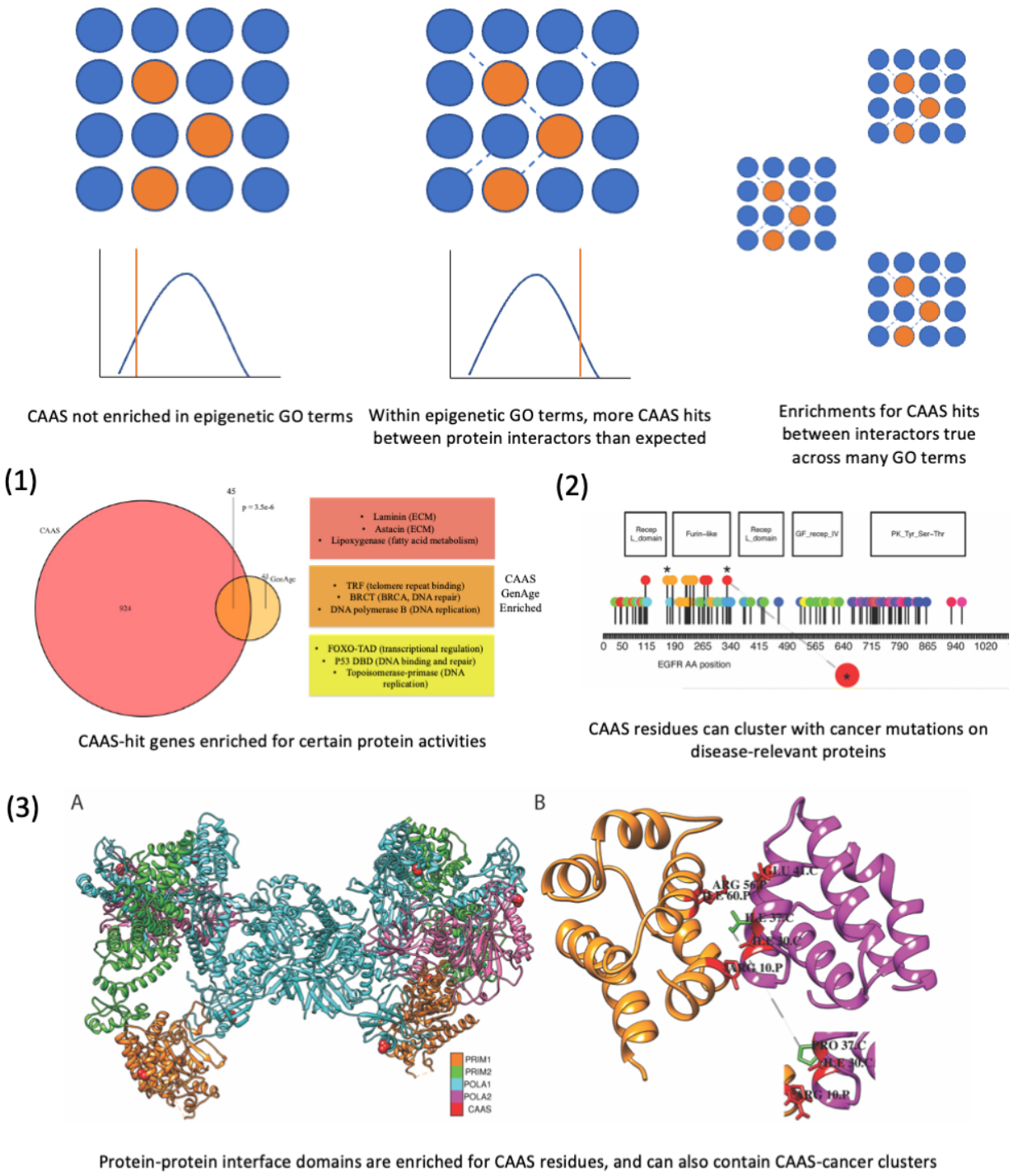


Figure S4.1: Graphical overview of main findings. (Top) CAAS gene hits are not enriched in assembled epigenetic GO terms (left) – however, when considering protein-protein interactions, CAAS hits interact with other CAAS hits within a GO term more than CAAS within randomized gene sets (middle). This pattern was not unique to epigenetic GO terms, but instead true across many GO terms (right). This lead

us to consider whether CAAS-hit genes were over-represented in certain protein activities (1), occurred in certain spatial patterns within proteins (2), or were themselves over-represented at protein-protein interfaces (3).

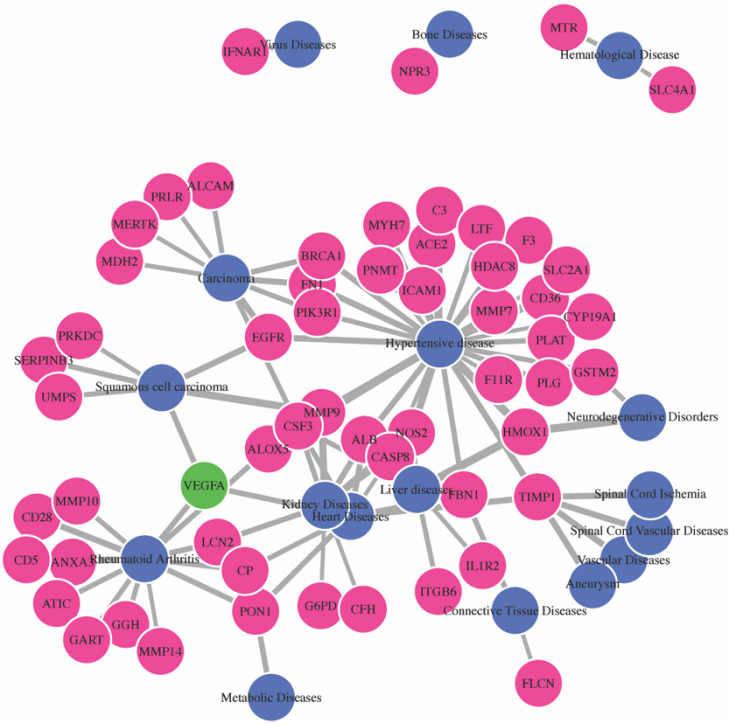


Figure S4.2: Gene-disease associations of clustered residue mutations. (**A**) Disgenet(Piñero et al., 2020) gene-disease association network for the enriched diseases highlighted in Figure 2C. Example of VEGFA highlighted in green.

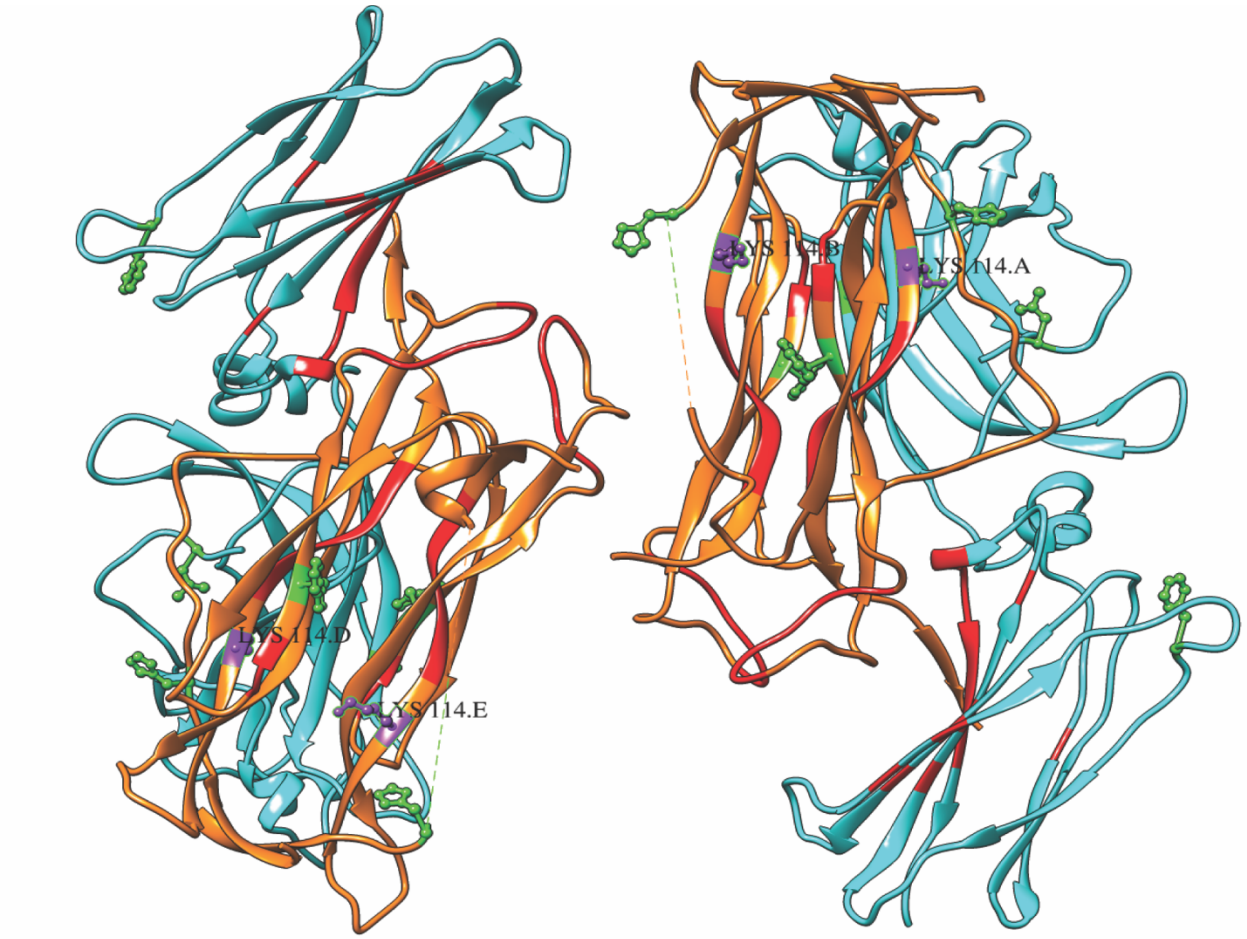


Figure S4.3: IL17 in complex with IL17RA. Cyan: IL17RA chains. Orange: IL17 ligand chains. Red: Putative defined interface residues (see Methods). Green: CAAS occurring within either protein sequence. Purple (labelled): One CAAS, at K114 of the IL17 ligand, was also considered a potential interface residue. PDBID: 4hsa.

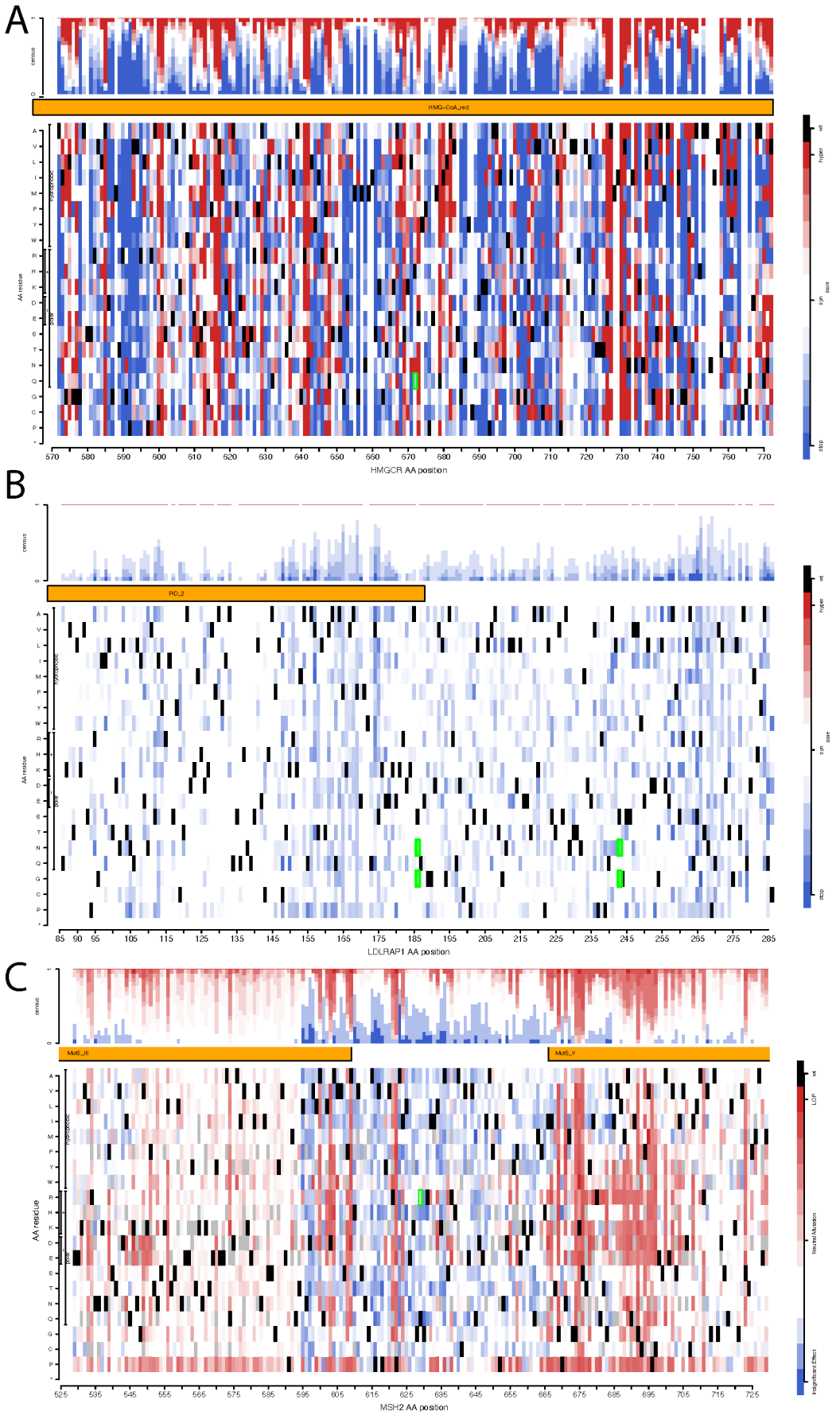

A
census
HMG-CoA_red
hydrophobic
polar
AA residue
hyper
syn
stop
HMGCR AA position
B
census
PID_2
hydrophobic
polar
AA residue
hyper
syn
stop
LDLRAP1 AA position
C
census
MutS_III
MutS_V
hydrophobic
polar
AA residue
LOF
Neutral Mutation
Insignificant Effect
MSH2 AA position

Figure S4.4: CAAS characterized in additional mutational scan studies.

(**A**) Effects of substitutions engineered in HMGCR, as measured by a functional complementation assay for the ability of mutant human HMGCR to recue growth of an hmg1 hmg2 double deletion yeast strain in rosuvastatin-supplemented media (MAVEDB accession: urn:mavedb:00000035-a-1). Red/white/blue color scale refers to effects of termination codon (blue), synonymous mutations (white) and hyper-active mutations (red). The CAAS phenotyped in this assay is indicated in green box. Black boxes refer to the reference residue at a given position. (**B**) Effects of substitutions engineered in LDLRAP1, as measured by a yeast two-hybrid assay to determine effects of mutations on LDLRAP1 binding affinity with partner proteins, wherein variant enrichment is assessed by ability for strains to grow via LDLRAP1-controlled expression of histidine production (MAVEDB accession: urn:mavedb:00000036-a-1). Red/white/blue color scale refers to effects of termination codon (blue), synonymous mutations (white) and hyper-active mutations (red). The CAAS phenotyped in this assay is indicated in green box. Black boxes refer to the reference residue at a given position. (**C**) Effects of substitutions engineered in MSH2, as measured by a functional complementation assay using the compound 6-TG, which is selectively toxic to MMR-deficient cells (MAVEDB accession: urn:mavedb:00000050-a-1). Positive scores in this assay correspond to loss of function (coloured red), while negative scores correspond to functionally-neutral variants (coloured white-blue). The CAAS phenotyped in this assay is indicated in green box. Black boxes refer to the reference residue at a given position.

Appendix B: Supplemental Table Legends

Supplementary Table 2.1. ATAC-seq datasets from mouse and human long bone samples, with downstream analyses. Related to Figure 2.1.

(Sheets 1-8, "IDR-Filt. Prox. Femur" – "IDR-Filt. Dist. Radius"): narrowPeak-formatted peak (mm10) information for called ATAC-seq peaks from indicated tissue following IDR replicate filtering. **(Sheet 9, "IDR Filtering Results")**: Summarized numbers of called peaks on individual biological replicates and final peak sets (mm10) following replicate consolidation. **(Sheet 10, "IDR Consistency Check"):** Results of analyses for within- and between-sample concordance. Columns A-C, E-G similar to those in Sheet 9. Self-consistency and rescue-ratios for the distal femur and proximal tibia replicate pairs used are indicated. **(Sheet 11, "Human E59 ATAC-seq")**: Summarized numbers of called peaks for ATAC-seq performed on human E59 embryonic tissue. **(Sheets 12-25, "General Knee Dist. Femur (hg19)" – "Embryonic Brain (hg19)")**: Lifted-over peak regions (hg19) in BED format from indicated tissue. "General" sets refer to original IDR-filtered, brain-filtered peak sets intersected as shown in Sheets 26-27, "Peak Set Intersections". "Specific" sets refer to knee/elbow peak set intersections, filtered for opposing bone-ends (e.g. distal tibia, proximal femur). **(Sheets 26-27, Peak Set Intersections):** Intersections between peak sets in mm10/hg19 (indicated) in generating the indicated general/specific joint sets. **(Sheet 28, "Fore-Hindlimb Intersections")**: Intersections of homologous forelimb/hindlimb peak sets. **(Sheet 29, "Human E59 – Mouse Intersections")**: Human E59 peak sets intersecting equivalent lifted-over mouse sets. Percent overlaps for mouse ('A') and human ('B') sets (respectively) are indicated. **(Sheet 30, "Human-Mouse Orthology")**: Mouse knee-specific sets were lifted-over to hg19, then lifted back to mm10 in order to confirm one-to-one orthology. The same analysis was performed for human E59 region sets. **(Sheet 31, "Genomic Annotations")**: Elements in each lifted-over set assigned genomic annotations based on distance to nearest TSS. **(Sheets 32 – 47, "Distal Femur Brain-filtered" – "Elbow-Common-Specific")**: GREAT output using ATAC-seq region sets. GREAT analysis was performed on sets of called ATAC-seq peaks, lifted over to hg19. Sheet titles indicate the set used - 'brain-filtered' sets are those in which ATAC-seq peaks were filtered for brain tissue, but not separated

from the opposing end of the long bone, nor separated at the knee/elbow. Table shows top hits per each category for which FDR (both binomial and hypergeometric tests) <= 0.05. See great.stanford.edu for details on column headings. Sheets reporting no GREAT data indicate that there were no significantly enriched terms using that dataset). **(Sheets 48 – 50, "DF - E59 Overlap" – "KC - E59 Overlap")**: GREAT results for lifted-over knee-specific sets intersecting human E59 data. **(Sheets 51-52, "Mouse_ATAC" – "Human_E59_ATAC")**: ATAC-seq quality/control data and primers. Column headings are self-explanatory. "Mouse_ATAC": ATAC-seq performed on E15.5 mouse embryonic tissue. "Human_E59_ATAC": ATAC-seq performed on E59 human embryonic tissue.

Supplementary Table 2.2. ATAC-seq dataset overlaps with evolutionary and functional genomics datasets. Related to Figure 2.1.

(Sheet 1, "Col2a-eCFP"): Intersections between gross-dissected proximal and distal femur ATAC-seq region sets and Col2a-eCFP-sorted equivalent region sets were performed using regioneR to test for significance of overlap compared to a randomized background. Columns: "Enrichment p-value": p-value for overlap enrichment test compared to randomized background. "Observed Intersections": Observed number of overlaps between indicated peak set and chromatin mark regions. "Enrichment Z-score": Z-score calculated as comparison of observed overlaps compared to randomized distribution. **(Sheets 2-4, "Overlaps - E47 Limb Bud - H3K27" – "Overlaps - Human E59 - ATAC")**: Intersections between ATAC-seq region sets and chromatin mark data were performed using regioneR to test for significance of overlap compared to a randomized background. Columns as above. Sheet "Overlaps - E47 Limb Bud - H3K27ac" provides overlap enrichment tests for specific and general knee sets, along with knee-specific sets separated by human sequence diversity, for H3K27ac ChIP-seq called regions obtained from human embryonic (E47) forelimb/hind limb bud. Sheet "Overlaps - BMDC - H3K27ac" provides overlap enrichment tests for specific and general knee sets, along with knee-specific sets separated by human sequence diversity, for H3K27ac ChIP-seq called regions obtained from bone marrow derived chondrocytes. Sheet "Overlaps - Human E59 – ATAC-seq" provides overlap enrichment tests for indicated knee-specific sets, for Human E59 ATAC-seq region sets from matched tissues (as indicated by

'Mouse Set' and 'Human Match' columns). **(Sheets 5-7, "ChIP-seq Overlaps - Dist. Femur" – "ChIP-seq Overlaps - KC")**: Overlaps between transcription factor ChIP-seq and ATAC-seq peak sets with regioneR, columns as above with additional: "Factor": TF assayed (mouse experiments indicated), "Source Reference": ENCODE or GEO Datasets experiment accession code for a given ChIP-seq experiment. **(Sheets 8, "Conservation - PhyloP20ways")**: phyloP20ways per-bp values for ATAC-seq region sets were compared across sets either through averaging per-bp values across 500bp for all regions ("Per-BP") or averaging values per-region ("Region-Averaged"). One-sided p-values for Wilcoxon rank-sum tests are shown, along with adjusted values. Yellow bars indicate region sets being compared. **(Sheets 9-14, "Dist. Femur Bottom 25%" – "Knee-Common Top 25%")**: Human-chimp sequence identity for ATAC-seq peaks was calculated for all sequences and used to generate top/bottom 25% slices. GREAT results for hg19 peaks in each of these slices are shown; tables show top hits per each category for which FDR (both binomial and hypergeometric tests) <= 0.05. See great.stanford.edu for details on column headings. Sheets reporting no GREAT data indicate that there were no significantly enriched terms using that dataset. **(Sheets 15 – 20, "Dist. Femur Bottom 10%" - "Knee-Common Top 10%"):** Equivalent GREAT results to above for top/bottom 10% slices. **(Sheets 21 – 23, "Dist. Femur - Peak Data" – "Knee-Common - Peak Data")**: Region information for human-chimp phyloP20ways analysis, for indicated set. "chromosome", "start" and "end" columns indicate the location of a given ATAC-seq peak in hg38 coordinates. "hg38.sequence" and "panTro4.sequence" columns contain the extracted sequences from multiz20ways MAF, with */- indicating gaps. "Raw.Match.Score", "Total.length" and "Sequence.Identity" refer to the number of matched nucleotides, length of sequence (ignoring shared gaps) and calculated sequence identity, respectively. "phyloP20ways.score" refers to phyloP20ways score averaged over the length of the indicated region. "random.region.phyloP20ways.score" refers to the averaged phyloP20ways score for a randomly-generated, size-matched region, and is used to generate the random distribution as seen in **Fig.1D**.

Supplementary Table 2.3. Transcription factor binding analyses using ATAC-seq sets. Related to Figure 2.2.

(Sheets 1-6, "Distal Femur (mm10)" – "Knee Common (hg19)"): HOMER *de-novo* motif analysis results for the indicated peak set, using original mm10 and lifted-over hg19 sequence sets. Columns: "Motif Logo": Logo of generated de novo motif. "Rank": Motifs ranked by calculated p-value. "P-value" and "log p-value": p-values for motif enrichment calculated by HOMER. "% of Targets/Background": Percent of target and background sequences predicted to contain a given motif. "STD(Bg STD)": Standard deviation of motif occurrence (bp) away from the center of the 500 bp sequence for target and background sets, respectively. "Similar Motif": Similar motif as identified through search of HOMER vertebrate motif library; similarity score shown in parentheses. **(Sheets 7 – 9, "Motif Alteration - Distal Femur" – "Motif Alteration – Knee-Common")**: Lifted-over knee region sets were tested for enrichment/depletion of human-chimp divergences predicted to alter potential binding motifs for a set of relevant transcription factors. Columns: "Transcription Factor Motif": transcription factor PWM for which predicted binding sites are being tested. "Distribution p-value": p-value for enrichment/depletion test against randomized background. "Adjusted p-value": p-value corrected for multiple tests of different TFs. "Direction of Bias": Whether a given region set had more (enrichment) or less (depletion) divergences predicted to alter a given TF motif than expected for randomized region sets. ChIP-seq signal enrichments for available TFs are shown in **Supplementary Table 2**, Sheets 5-7. **(Sheets 10-17, "Knee-Specific Pooled Cluster 1" – "SK-Common Cluster 2")**: Clusters of common human variants were tested for biases to modify potential binding motifs for a set of relevant transcription factors. Columns as above. "SK" sheets refer to specific-knee sets. **(Sheets 18-20, "Distal Femur Intra-species" – "Knee-Common Intra-species")**: Common human variants falling within ATAC-seq regions (of a fixed 500bp size) from the three knee-specific sets (unclustered) were tested for biases to modify potential binding motifs. Columns as above.

Supplementary Table 2.4. Putative knee regulatory regions and acceleration signals. Related to Figure 2.1.

(Sheet 1, "HAR Region Set"): Genomic coordinates (hg19) of regions showing evidence of acceleration in humans aggregated from multiple studies; citations for particular regions are indicated in the "source" column. **(Sheets 2-10, "Embryonic Brain" – "Knee-Common-Specific")**: Intersections between regions detailed in Sheet 1 and the indicated ATAC-seq region set. Columns: "chr", "start", "end": hg19

coordinates for a given ATAC-seq region intersecting a region of acceleration. “HAR.chr”, “HAR.start”, “HAR.end”: hg19 coordinates for region of acceleration intersected. “Nearest.Refseq.Gene”,"Entrez.ID", "Gene.Symbol”, : Nearest annotated TSS to the indicated ATAC-seq region, as RefSeq accession, Entrez ID or human gene symbol, respectively. “Nearest.TSS”: Distance (bp) from intersected region to nearest TSS, negative values denoting elements upstream. “Gene.Description”, “Gene.Type”: Short description and type of nearest annotated gene. **(Sheets 11-13, “Dist. Femur - E59 Overlap” – “Knee Common - E59 Overlap”)**: Intersections between mouse embryonic ATAC-seq regions with equivalent human E59 embryonic data, overlapping acceleration signals. Columns as in sheets 2-10. **(Sheet 14, "Intersection Enrichment Testing”)**: Accelerated region intersections for each set were compared to a 10K random background to test for enrichment of overlap. “Set”: ATAC-seq region set tested. “Set size (bp)”: size of indicated region set in base-pairs. “Number of Accelerated Region Intersections”: Number of accelerated region intersections in this set. “Number of Intersections per bp”: Accelerated region intersections expressed as a percentage of total set size. “Beta Distribution P-value”: Significance value for enrichment test against background using a fitted beta distribution. “Adjusted p-value”: Correction for multiple hypothesis testing using BH correction. **(Sheet 15, “Great Annotations”)**: GREAT output using an aggregate of all lifted-over mouse hindlimb ATAC-seq regions intersecting regions of acceleration. Table shows top 50 hits per each category for which FDR (both binomial and hypergeometric tests) <= 0.05. See great.stanford.edu for details on column headings. Sheets reporting no GREAT data indicate that there were no significantly enriched terms using that dataset.

Supplementary Table 2.5. Analyses of inter- and intra-species diversity in ATAC-seq sets. Related to Figure 2.3.

(Sheets 1 – 24, “Inter-species - DF General” – “Intra-Spec. (Gor) Elbow-Spec.”): Intra-species diversity across and within taxa. Columns: “Model”: Statistical method for which results are indicated. “Comparison”: Species/Sets being compared. “p-value”: Result of Tukey post-hoc testing for the indicated pair. “Direction (wrt first in pair)”: direction of variance between the pair, with respect to the first in the ‘Comparison’ pair. “t.ratio”: Test statistic for Tukey post-hoc test. Note for certain sheets resampling tests are indicated by “BOTTOM 25%” and “TOP 25%” headings, which refer to statistical results based on

taking the 25th and 75th percentile (respectively), rather than the mean, for human subsampling. The "N = 100 subsampling" heading refers to statistical results based on increasing the human sub-sample size for testing. Sheets 1-3, 7-9, 13-15, 19-21: Comparisons of intra-species diversity in humans, chimpanzees and gorillas for sequences in the indicated joint region sets, either "general" or "specific". Sheets 4-6, 10-12, 16-18, 22-24: Comparisons of intra-species diversity in sequences between forelimb/hindlimb joint region sets, either "general" or "specific", for human, chimpanzee and gorilla. **(Sheet 25, "Inter-Spec. E59 Knee Overlap")**: Comparisons of intra-species diversity in humans, chimpanzees and gorillas for knee-specific and elbow-specific region sets overlapping with matched human embryonic tissue (E59). **(Sheet 26, "Intra-Spec. E59 Knee Overlap")**: Comparisons of intra-species diversity in humans, chimpanzees and gorillas (comparing sets within species) for knee-specific and elbow-specific region sets overlapping with matched human embryonic tissue (E59). **(Sheet 27, "Human Seq. Constraint")**: Elements/regions in each set were intersected with common human variants, with intersections compared to a set of 1000 randomized region sets to test for depletion/enrichment. Columns: "Set": Genomic feature set tested. Order is equivalent to left-right in **Fig. 3A**. "Set size (bp)": Size of indicated region set in base-pairs. "Number of Common Variants Intersected": Number of common human variants intersecting this set. "Number of Intersections per bp": Common variant intersections expressed as a percentage of total set size. "Background Average": Average SNP/bp value from background region sets. "CDF P-value": Significance value for testing significant deviation from background using normal CDF. "Corrected CDF p-value": Correction for multiple hypothesis testing using BH correction. "Direction": Whether set value falls to the left/right of background distribution. **(Sheet 28, "Chimpanzee Seq. Constraint")**: Elements/regions in each set were intersected with common chimpanzee variants, with intersections compared to a set of 1000 randomized region sets to test for depletion/enrichment. Columns as indicated in Sheet 27, with the addition of 'Gamma P-value' and 'Corrected Gamma P-value', which refer to distribution p-values obtained when using a fitted gamma distribution in addition to those obtained using a normal CDF (Methods). **(Sheet 29, "Morphological Constraint")**: MRI measurement data for femur and tibia from the OAI dataset were tested for significant differences in variation across individuals; average and standard deviations of measurements for tibia/femur are indicated in columns B-E, with test statistics and resulting p-values for performed Fligner-

Killeen and Levene's tests shown. **(Sheet 30, "OAI Diversity"):** Genotype information from the OAI dataset was intersected with knee-specific elements to compare sequence diversity across sets. Column titles as for Sheets 1-24. "BOTTOM 25%" and "TOP 25%" headings refer to statistical results based on taking the 25th and 75th percentile (respectively), rather than the mean, for cohort subsampling. **(Sheets 31 – 42, "Dist. Femur Most Cons 25%" – "Elbow-Common Least Cons 25%")**: GREAT output using lifted-over ATAC-seq peaks from specific joint region sets, further separated based on the intersection of common human variants (averaged over sub-samples) using most/least constrained (i.e. least/most variable) 25% slices. **(Sheets 43-54, "Chimp Dist. Femur Most Cons 25%" – "Chimp Elbow-Comm. Least Cons 25")**: Equivalent GREAT output for lifted-over ATAC-seq peak sets separated on basis of chimpanzee variant intersections. Sheets reporting no GREAT data indicate that there were no significantly enriched terms using that dataset).

Supplementary Table 2.6. Analyses of intra-species diversity in ATAC-seq sets. Related to Figure 2.3. **(Sheets 1 – 36, "Human General Knee Distal Femur" – "Gorilla Elbow-Common-Specific")**: Genetic variants from human, chimpanzee and gorilla intersecting each element set. Columns: "chr", "start", "end": ATAC-seq region coordinates (padded to fixed length of 500 bp), sorted by sequence diversity. "Mean Values": Mean number of common variants intersecting the indicated region across sub-samples. "Standard Deviation": Standard deviation of common variant intersections across sub-samples. "25th Percentile": First quartile value for subsampled common variant intersections. "75th Percentile": Third quartile value for subsampled common variants intersections. "Subsample X": Number of common variants intersecting the indicated region within a particular subsample (X being a particular subsample number). Note that human subsamples refer to 200 subsamples of n = 25 from 1KG3 individuals. Gorilla subsamples consist of 5 subsamples of n=25 for 31 individuals. For chimpanzee diversity the entire set was used as a single sample for 25 individuals. Chimpanzee/Gorilla region coordinates are in hg18, human sheets are in hg19.

Supplementary Table 2.7. Heritable osteoarthritis risk in ATAC-seq peaks. Related to Figure 2.4.

(Sheet 1, "OA Lead Set"): Information on the lead OA risk-variants aggregated for analysis. Columns: "rsid": DBSNP160 accession. "chromosome", "Chromosome Start (hg38)", "Chromosome End (hg38)": Coordinates for the indicated SNP in hg38. **(Sheet 2, "OA Lead-Proxy Intersections")**: Intersections of independent OA risk loci and ATAC-seq region sets. Columns: "lead_RSID": lead variant identified in GWAS studies. "chrom", "bp", "proxy_chr", "proxy_bp": position (hg19) of lead variant and its proxies, respectively. "proxy_rsid": DBSNP accession for proxy variant. Columns G-P: Intersection between indicated ATAC-seq region set (e.g. "Distal Femur-Specific") and a given proxy variant, '1' indicating overlap. **(Sheet 3, "OA - Intersections – Selection")**: Those proxy variants intersecting ATAC-seq regions, subsequently intersected with EGDP selection windows. Columns: "Chr", "Start", "End": coordinates for a given selection window (hg19). "pclr": Results of test for ancient selection via 3P-CLR. Columns E-P: Results of test for recent selection in each of the indicated populations. '1' indicates a significant test result. "Genes": Genes within the indicated selection window. Columns R-AB are equivalent to those described for Sheet 2. **(Sheet 4, "rs143383 Pop. Allele Freq.")**: Frequency of the OA risk-variant rs143383 in EGDP populations. Columns: "ref", "alt": allele frequency of reference and alternate allele, respectively. "reference, alternate": Reference and alternate allele, respectively. "# of individuals": Number of individuals in a given EGDP population. "Final_Group": EGDP population for which allele frequencies are reported. "RSID": accession of given variant in chr:pos format. Columns H-J: Results (p-values) of recent selection tests for the indicated population within the selection window overlapping the given OA risk locus. "proxy": Proxy variant intersected by ATAC-seq data. "lead": Lead OA-risk variant in LD with proxy. "selected": Whether or not the current selection window gives a significant ($p < 0.05$) selection test for an EGDP population, as demarcated by "*". **(Sheets 5-7, "rs2626053 Pop. Allele Freq." – "rs7785659 Pop. Allele Freq.")**: Population allele frequencies for additional ATAC-seq intersected OA risk loci, with columns as described for "rs143383 Pop. Allele Freq.". **(Sheet 8, "OA Variant Enrichments")**: Results of enrichment testing for OA risk loci. Columns: "Set": the ATAC-seq peak set tested. "OA Variants Intersected": Number of unique OA risk loci intersected by peaks. "Average/SD Intersections in Background": Average/standard-deviation (of) number of unique OA risk loci intersected in randomized backgrounds (Methods). "Z-score": Resulting Z-score of target intersections. "Enrichment p-value": one-sided p-value for enrichment (Methods). **(Sheet 9, "OA Variant**

Motif Intersections"): OA variants captured by element sets were assessed for overlap with predicted chondrogenesis-associated TF binding sequences. Columns: "SNP_ID": String-formatted information for given variant, corresponding to columns C-F. "RSID": DBSNP identifier. "TF Motif": Transcription factor PWM matching a given sequence. Columns F-J: Information on PWM sequence match. "Predicted Effect": Whether the alternate allele results in a higher/lower PWM sequence match. "Element Set Capturing Variant": The element set intersecting the given variant. Note that some of these sets are overlapping, so duplicate rows are possible. **(Sheet 10, "Motif Biases - All OA Vars")**: Predicted TF binding sites overlapping any OA variant were counted to look for biases in the incidence of variants within certain predicted sites. Columns: "Factor": Specific TF PWMs compared to the superset distribution (Methods). "# of Variants": Number of OA variants which intersect predicted sites for this TF. "Average Intersections for all TFs": Average number of OA variants intersecting any predicted TF site. "Gamma Distribution p-value": Significance value for testing significant deviation from superset using a fitted gamma distribution. "Adjusted p-value": Correction for multiple hypothesis testing using BH correction. "Direction": Whether set value falls to the left/right of superset distribution. **(Sheet 11, "Motif Biases - Element OA Variants")**: Predicted TF binding sites overlapping element-captured OA variants were counted for particular TF PWMs, similar to Sheet 10. Only tests for which un-adjusted $p < 0.05$ are shown. Columns are as in Sheet 10, with "Set" indicating the element set considered. **(Sheet 12, "Element – OA Motif Proportions")**: The predicted TF binding sites intersected by element-captured OA variants were compared to all OA variants to look for biased occurrence in element sets. Columns: "Set": OA variants captured by a particular element set. "Factor": particular TF PWM being tested. "Number of Target TF Predicted Binding Sites": Number of predicted TF binding sites (of those tested) intersected by variants from this set. "Number of Predicted TF Binding Sites": Number of predicted sites for the target TF intersected by element variants. "Total Number of Target TF Predicted Binding Sites": Number of predicted sites for the target TF intersected by any OA variant. "Total Number of Predicted TF Binding Sites": Number of predicted TF binding sites (of those tested) intersected by any OA variant. "% of Target TF sites of Set": Percent of target TF sites intersected by element-variants, relative to all TFs tested. "% of Target TF Sites of Total": Percent of target TF sites intersected by any OA variant, relative to all TFs tested. "% of Target TF Sites Captured in Set": Percent of OA variant-intersected target TF sites captured

by element set. “Hypergeometric Test”: p-value for hypergeometric test. “Adjusted P-value”: Adjusting for multiple hypotheses with BH correction. “Direction”: whether target set captures more/less target sites than expected. “Fold-Change”: Enrichment/depletion of sites in target set. **(Sheet 13, “OA - Partitioned Heritability”)**: LDSC analysis was performed on OA GWAS summary statistics, utilizing a custom set of genomic features in addition to a modified baseline LD model. Columns: “Category”: Genomic annotation used to partition genome-wide heritability. Custom element set annotations are highlighted in yellow; other annotations are taken from the LDSC baseline model (http://data.broadinstitute.org/alkesgroup/LDSCORE), which are described in their original publications (Finucane et al., 2015; Gazal et al., 2017). “Prop._SNPs”: Proportion of SNPs captured by features; note that feature sets are overlapping, so column sums to > 1. “Prop._h2”: Proportion of genome-wide heritability, similarly sums to > 1.“Prop._h2_std_error”: Standard error of partitioned heritability estimate. “Enrichment”: Proportion of heritability / proportion of SNPs. “Enrichment_p”: Significance of heritability proportion captured by feature. “P_ADJ”: Adjusted p-values for multiple hypothesis correction. Columns I – K: Coefficients relating to LD score regression, see Finucane et al., 2015. **(Sheet 14, “OAI Cohort Variation”)**: The number of alternate alleles falling within indicated element sets for subsampled individuals from the OAI cohort as well as additional control populations (1KG3, EGDP) was compared (see Methods). T-test statistic, group averages and p-values shown were averaged over 1000 permutations (Student’s T-test, alternative of greater variation in OAI), with standard-errors similarly reported (‘sd’); headers indicate results when using n = 20 and n = 200 subsets.

Supplementary Table 2.8. Common human variant clustering and region selection tests. Related to Figure 2.4.

(Sheet 1, “Cluster Properties”): Columns: “Set”: Source of the indicated set of clustered variants from knee-specific ATAC-seq regions. “# of Variants”: Number of common variants grouped in the indicated cluster. “Cluster Mean Fst”: Average Fst value, across all pairwise population comparisons, for all common variants within the cluster. “Cluster SD Fst”: Standard deviation of Fst values across all pairwise population comparisons for all common variants in the cluster. “Cluster Max Fst”: Maximum Fst value, across all pairwise population comparisons, for all common variants within the cluster. **(Sheets 2-9,**

“Knee-Specific Pool Cluster 1” – “Knee-Common-Specific Cluster 2”): Position information for clusters described in “Cluster Properties” (hg19), with variant accessions indicated. **(Sheets 10 – 17, “GREAT KS Pool Cluster 1” – “GREAT Knee Common-Sp. Cluster 2”)**: GREAT output using clustered variants, sheet titles indicate set/cluster used, KS indicating ‘Knee-Specific’, Pool indicating “Pooled”, Sp. Indicating “Specific”. **(Sheet 18, “Selection Window Intersects”)**: Intersects of region sets and EGDP selection windows. Columns: “Chr_Mb”, “Chr”, “Start”, “End”: coordinates for a given selection window (hg19). “pclr”: Results of test for ancient selection via 3P-CLR. Columns F-Q: Results of test for recent selection in each of the indicated populations. ’1’ indicates a significant test result. “Genes”: Genes within the indicated selection window. Columns S-AE: Intersection between indicated ATAC-seq region set (e.g. “Distal Femur-Specific”) and a given selection window, ’0’ indicating no overlap. **(Sheet 19, “Grossman Region Intersects”)**: Intersects of region sets and CMS selection regions. Columns: “chr hg19”, “start”, “end”: Genomic coordinates for a given selection region. Columns E-Q: Intersection between indicated ATAC-seq region set and a given selection region. **(Sheet 20, “Ancient Sel. Window Testing”)**: Intersections between region sets and signals of ancient selection. Columns: “Region Set”: ATAC-seq region set used for testing. “Number.of.Windows.Intersected”: Number of 200kb windows intersected by a given set. “Number.of.Windows.Under.Selection”: Of the windows intersected, the number of windows for which at least one population had a significant selection signal. “Total.Number.of.Windows.Under.Selection”: Of all the windows tested across the genome, how many had at least one population with a significant selection signal. "Total.Number.of.Windows”: Number of 200kb windows across the genome used for testing. “Hypergeometric.Test”: p-value for a hyper-geometric test of over/under-representation. "Direction (Over/Under-represented)”: Direction of indicated hyper-geometric test (i.e. more/less selection signal intersections than expected). “Adjusted.P.value”: p-value for indicated hyper-geometric test after BH multiple-hypothesis correction. “Fold-Change”: Enrichment/depletion of selection window overlaps in target set. **(Sheet 21, “Recent Sel. Window Testing”)**: Intersections between region sets and signals of recent selection. Columns are as described for Sheet 3. **(Sheet 22, “Grossman Selection Enrichments”)**: Intersections between region sets and CMS selection regions. Columns: “Region Set”: Region sets typically used for overlap testing in this study, but also consisting of the following sets: ‘OA Variants’: Lead OA-risk variants and their proxies;

‘General Knee Pooled - OA Intersections’: Those regions from the general knee pooled set intersecting OA-risk variants and/or their proxies; ‘Limb Bud - OA Intersections’: Those regions from the human E47 limb bud H3K27ac set intersecting OA-risk variants and/or their proxies; ‘BMDC - OA Intersections’: Those regions from the human BMDC H3K27ac set intersecting OA-risk variants and/or their proxies. “Enrichment p-value”: p-value for overlap enrichment/depletion test compared to randomized background. "Observed Intersections”: Observed number of overlaps between indicated peak set and ChIP-seq mark regions. “Enrichment/Depletion Z-score”: Z-score calculated as comparison of observed overlaps compared to randomized distribution.

Supplementary Table 2.9. Experimental and bioinformatics results on the *GDF5 R4* enhancer. Related to Figures 2.5-7.

(Sheets 1 – 4, “*R4* null P30” – “*R4* rs6060369 P56 t-test”): P30, P56, and 1 year microCT statistics of *R4* Engineered Mice. **(Sheet 1, “*R4* null P30”)**: Differences in key morphological features of the *R4* null knee joint at P30. All outcome measures were defined as continuous variables (N=5 per genotype) and compared between the genotypes using multivariate linear regression analysis with anatomy used as dependent variable, and genotype and line as independent variables. Sidak post-hoc was used to correct for multiple comparisons between genotypes. P values are two-sided and the statistical significance was assessed at alpha = 0.05 for all the comparisons. Significant differences are highlighted in bold. WT: Wild Type; HET: Heterozygotes; HOM: Homozygotes. **(Sheet 2, “*R4* null 1 year”)**: Differences in key morphological features of the *R4* null knee joint at 1 year. Statistical treatment similar to ‘*R4* null p30’ results, with similarly-defined continuous variables (6 wildtype, 15 heterozygous and 14 homozygous). Formatting as above. **(Sheet 3, “*R4 rs6060369* P56”)**: Differences in key morphological features of the *R4 rs6060369* replacement knee joint at P56. All outcome measures were defined as continuous variables (6 wildtype, 8 heterozygous and 6 homozygous) and compared between the genotypes using oneway ANOVA analysis with anatomy used as dependent variable and genotype as independent variable. Sidak post-hoc was used to correct for multiple comparisons between genotypes. P values are two-sided and the statistical significance was assessed at alpha = 0.05 for all the comparisons. Formatting as above. **(Sheet 4, “*R4 rs6060369* P56 t-test”)**: Differences in key morphological features of

the *R4 rs6060369* knee joint at P56 body size normalized. All outcome measures were determined using paired T-test on littermate WT, Homozygotes, and Heterozygotes (N=6 mixed male and female per genotype). P values are two-sided. Significant differences, which are uncorrected, are highlighted in bold. **(Sheet 5, "*R4* null P30 histology data")**: Analysis of osteoarthritis on *R4* WT and HOM outcomes based on total joint OARSI Score at P30 indicating that significant joint osteoarthritis is absent at P30 in WT and HOM knees. A Mann-Whitney test was used to compare the OARSI Score between the wild type (N=5) and homozygous (N=5) knees. P-values are two-sided and the statistical significance was assessed at alpha = 0.05. **(Sheet 6, "*R4* null 1 Year histology data")**: Analysis of osteoarthritis on *R4* WT and HOM outcomes based on total joint OARSI Score and Osteophyte Score at 1 year. Significantly more osteoarthritis as measured using OARSI Score was observed in HOM knees. No significant difference was observed in Osteophyte Score between genotypes. A Mann-Whitney test was used to compare the OARSI Scores as well as Osteophyte Scores between the wild type (N=6) and homozygous (N=14) knees. P-values are two-sided and the statistical significance was assessed at alpha = 0.05. **(Sheets 7-9, "UniProbe - Non-Risk TTTAGCCGAGC" – "UniProbe - Risk vs Non-Risk")**: UniProbe Analysis of rs6060369 variant position. Two 15 bp sequences centered on the non-risk "C" (Sheet 5 - black font) and risk "T" (Sheet 6 - red font) alleles at rs6060369 were investigated for evidence of *in vitro* experimental TF binding affinities. Comparisons of the output for each allele (Sheet 7) reveals a number of reduced or gained TFs, for the risk allele with the PITX1 TF being highlighted (yellow) as a binding gain (i.e., >0.10 enrichment change) and factor expressed in and required for knee development. **(Sheet 10, "Construct Primers")**: Sequences of primers and sgRNAs used in this study to clone and target the *GDF5 R4* regulatory element in mouse and human cells.

Table S3.1 Region Set Characterization.

Contains information on region sets defined as differentially-accessible between either fetal/adult, or young-adult/older-adult tissue samples. Sheet 1 – "Sample Metadata": Sample Metadata for ENCODE

Dnase-I datasets used. Sheet 2 – “Developmental Regions”: Sets of regions differentially-accessible between fetal (FETAL-BIASED) and adult (ADULT-BIASED) samples. Sheet 3 - “Aging Regions”: Sets of regions differentially-accessible between young-age (YOUNG-BIASED) and old-age (OLD-BIASED) adult samples. Sheet 4 – “Development – Age Regions”: Chi-sq testing of overlaps between fetal/adult and young-age/old-age accessibility analyses. Sheet 5 - “Methylation Sites Analyses”: Intersections and enrichment tests for altered-accessibility region sets and methylation clock sites. Sheet 6 – “Histone Mark Accessions”: Sample Metadata for ENCODE Histone ChIP-seq datasets used. Sheets 7 – 14 – “FETALBIASED-GREAT” – “OLD-ADULT-BIASED -GREAT”: GREAT enrichment tool results for the indicated sets of regions, including intersection sets. See great.stanford.edu for details on column headings.

Table S3.2. Regulatory Gene Associations and Expression Data.

Describes results of analyses integrating additional chromatin-accessibility, as well as gene-expression, data with the differential-accessibility regions outlined in Results and Table S3.1.

Sheet 1 - “Development Promoter Access.”: Differential-accessibility results for gene promoters differentially-accessible between fetal (FETAL-BIASED) and adult (ADULT-BIASED) samples. Sheet 2 - “Aging Promoter Accessibility”: Differential-accessibility results for gene promoters differentially-accessible between young-adult (YOUNG-BIASED) and old-adult (OLD-BIASED) samples. Sheet 3 - “Region-Associated Genesets”: Sets of genes associated with the indicated region sets (see Supplemental Methods). Sheet 4 – “Genesets – GO Enrich”: Gene-set enrichment results for gene sets shown in Sheet 3. ‘Intersection’ refers to promoters with differentially-accessible regions intersecting, ‘Accessibility’ refers to promoters which themselves are differentially-accessible. Sheet 5 - “Developmental RNA-seq METADATA”: Sample Metadata for ENCODE RNA-seq pre-processed datasets used. Sheet 6 - “Aging RNA-seq Data METADATA”: Sample Metadata for GTEx RNA-seq pre-processed datasets used. Sheet 7 - “Region and Expression Overlaps”: Intersections between gene sets defined using altered-accessibility regions and those defined using gene expression datasets. ‘Intersection’ refers to promoters with differentially-accessible regions intersecting, ‘Accessibility’ refers to promoters which

themselves are differentially-accessible. Sheet 8 - “RNA-seq Genes - GO Enrich”: Gene-set enrichment results for gene sets defined using gene expression datasets.

Table S3.3: Evolutionary Sequence Analyses.

Describes the results of evolutionary sequence analyses, both between- and within-species, performed on sets of differentially-accessible regions outlined in Results and Table S3.1.
Sheet 1 – “phyloP20ways”: Comparing distributions of per-region-averaged phyloP20ways scores across different region sets. (Top) – Wilcoxon test results. (Bottom) – Student’s t-test results. See Supplemental Methods. Sheet 2 – “Divergent Sequence Set”: Aggregated set of human-divergent sequences used (hg19 coordinates). Sheet 3 - “Intersection Enrichment Testing”: Intersection-enrichment test results for divergent sequences intersecting the indicated region set. P-values shown using both a normal CDF as well as a fitted beta-distribution. Sheets 4 – 9, “BOCA Brain” – “GM12878”: Intersected divergent sequences broken down by region set. Genes nearest to each intersected divergent sequence are indicated. Final column indicates whether or not the particular region has a putative promoter contact with the indicated gene using processed promoter-capture datasets. Sheet 10 – “Zero-Hurdle Modelling”: Comparisons of sequence diversity (counts of common variants falling within the indicated region sets) between region sets, using variation data from humans, chimpanzees and gorillas. Sheet 11 - “Intra-species (Human) Diversity”: Comparison of human sequence diversity (counts of common variants falling within the indicated region sets) for indicated region sets when compared with randomized genomic region sets (see Supplemental File 2), enrichment/depletion evaluated using a normal CDF. Sheet 12 – “Intra-species (Chimp) Diversity”: Comparison of chimpanzee sequence diversity (counts of common variants falling within the indicated region sets) for indicated region sets when compared with randomized genomic region sets (see Supplemental File 2), enrichment/depletion evaluated using a normal CDF.

Table S3.4: Disease Associations.

Describes the results of analyses exploring the effects of changes to nearby epigenetic context (i.e. local accessibility) on the heritable aging-associated disease risk of human genetic variants, at the genome-wide and per-locus level. Sheet 1 – “UKB Summary Stat Files”: Metadata of UK Biobanks summary-

statistic datasets used. Sheet 2 - “SNP Enrichments - DNase Regions”: Enrichment results comparing the occurrence of significance-thresholded (1e-6) SNPs nearby DNase regions generally, when compared to randomly-generated region sets (see Supplemental Methods). Sheet 3 - “SNP Enrich - Hypergeo – ADULTBIASED”: Hypergeometric testing results for nearby significance-thresholded variants using the adult-biased region set. Additional developmental trait GWAS and longevity GWAS testing at bottom of sheet. Sheet 4 - “SNP Enrich - Hypergeo – FETALBIAS”: Hypergeometric testing results for nearby significance-thresholded variants using the fetal-biased region set. Additional developmental trait GWAS and longevity GWAS testing at bottom of sheet. Sheet 5 - “SNP Enrich - Hypergeo – YOUNGBIAS”: Hypergeometric testing results for nearby significance-thresholded variants using the young-biased region set. Sheet 6 - “SNP Enrich - Hypergeo – OLDBIAS”: Hypergeometric testing results for nearby significance-thresholded variants using the old-biased region set. Sheet 7 - “SNP Enrich - Intersects”: Hypergeometric testing results for nearby significance-thresholded variants using the indicated developmental/age-altered region intersect sets. Sheet 8 – “Effect-Size Dists”: Wilcoxon testing results for differences in absolute effect size values of variants nearby fetal-biased and adult-biased region sets, per-disease (see Supplemental File 2). Sheet 9 - “Per-SNP Comparisons”: Comparisons of indicated metric (e.g. Cross-Disease association metric) across variants nearby the indicated region sets using Tukey post-hoc corrections (see Supplemental File 2). Sheet 10 - “PhastCons Tests + CLINVAR”: (Top) Comparing the cross-disease association metric of variants falling within or outside phastCons elements globally, or when constraining by nearby altered-accessibility region sets. (Bottom) Comparing multiple per-variant metrics of ClinVar variants to genome-wide randomly sampled variants (see Supplemental File 2). Sheet 11 – “Genewise Rank Testing”: Comparing the cross-disease rankings (based on disease-associations) of the indicated gene sets compared to randomly-sampled sets of protein-coding genes (see Supplemental File 2). Sheet 12 - “Genewise Rank Agg”: Gene-set enrichments for sets of genes defined when aggregating within-disease gene rankings across diseases, when using an RRA-based approach (top), or when using an FGSEA rank-value based approach (bottom) (see Supplemental File 2). Top significant (q-value < 0.05) GO BP enrichments for each analysis are shown. Sheet 13 – “HDAC Gene Set”: Comparing the per-gene behaviour of top genes in the FGSEA-enriched ‘HDAC’ term in terms of different metrics, when compared to randomized gene sets. ‘ANN’ refers to the use of different region

sets (e.g. YOUNG_BIASED) in binning variants prior to assigning per-gene metrics (see Supplemental File 2). 'ALL' refers to enrichment/depletion testing independent of region set binning; log-odds values are calculated based on 'ALL' results p-values.

Table S4.1: Description of CAAS datasets.

Sheet 1: Description of convergent amino-acid substitutions used in this study. For each CAAS, the residue position, long and short-lived amino acid, and significance testing (see Methods) are shown, along with the Scenario in which each CAAS falls (see Methods).

Sheet 2: (Left) Global set of protein-coding genes considered in the definition of CAAS. (Right) set of age-associated genes obtained from GenAge.

Sheet 3: GO BP enrichments for genes hit by CAAS pooled from Scenarios 1-3.

Sheet 4: GO BP enrichments for genes hit by CAAS separated by Scenario.

Sheet 5: GO terms related to epigenetics used in this study, describing the number of genes in each set, as well as the number of genes containing at least one CAAS, and the number of interacting genes containing at least one CAAS.

Sheet 6: Statistical results of t-tests comparing the cross-disease ranking of genes in different gene sets relative to background ranks (see Methods).

Sheet 7: Statistical results of comparing the instances of CAAS residues in different contexts (as denoted in the 'TYPE' column) to randomly-sampled genesets.

Table S4.2: Protein functionality and CAAS

Sheet 1: Hypergeometric tests of all PFAM domains for which at least one CAAS in one protein overlapped. Columns refer to individual test parameters, as well as the final adjusted p-values for enrichment/depletion. 'Genes' column refers to CAAS-hit genes containing the indicated domain.

Sheet 2: Overlap of CAAS domain enrichments with domains enriched in GenAge genes (listed in Table S1). (Top): Hypergeometric test comparing the number of enriched domains in GenAge genes to the

number of overlapping enriched domains between GenAge and CAAS sets. (Bottom) Per-domain hypergeometric tests of all domains occurring at least once in GenAge genes.

Sheet 3: Subset of domains listed in Sheet 1 – enriched domains which overlapped domains enriched in GenAge genes.

Sheet 4: Mapping of all individual CAAS residues overlapping PFAM domains.

Sheet 5: Hypergeometric tests of all PFAM domains for which at least one CAAS directly overlaps at least one domain instance in a protein. Serpin domain (enriched) is highlighted in green.

Sheet 6: Hypergeometric tests of all PFAM clans for which at least one CAAS directly overlaps at least one instance of a domain in this clan. The 'Aha1_BPI' clan (enriched) is highlighted in green.

Sheets 7 – 14: Hypergeometric tests of individual PFAM domains occurring within genes of the indicated GO term. Columns largely as in previous sheets. 'Condition' column refers to the specific test performed. TARGET: GO term relative to global background, TARGET_CAAS: CAAS-hit genes within the GO term relative to the global background, COND_TARGET_CAAS: CAAS-hit genes within the GO term relative to all genes in the GO term. "HIT_GENES" column refers to those genes in the given set (GO term as a whole, CAAS-hit genes with GO term) containing at least one instance of the indicated domain.

Sheet 15: For all domains enriched in CAAS-hit genes within different GO terms (above Sheets 7-14), the hypergeometric tests have been aggregated. Additionally, the percentage of genes in the indicated GO term which have the indicated domain, and also are CAAS-hit is indicated in the far right column. Sheet 16: dbNSFP data on CAAS residue changes. DbSNFP dataset was subset to include those columns pertaining to computational predictions of functional effects of the indicated non-synonymous mutations. The 'PREDICTED_EFFECTS' column refers to the total number of different software predictions which describe a given mutation as possibly non-benign. 'PREDICTED_EFFECTS_STRING' refers to the individual prediction softwares making these non-benign predictions (separated by '@' symbol). The far-right columns indicate the PFAM domains which are overlapped by a given mutated residue.

Table S4.3: Spatial distribution of CAAS and cancer-related residues.

Sheet 1: HotSpot3D cluster calls for pooled set of residues from CAAS dataset and pre-processed TCGA mutation calls. Columns “CAAS_AA” and “CAAS_NUM” refer to the residue(s) within the given cluster call which are CAAS hits.

Sheet 2: Subset of Sheet 1 for all cluster calls occurring within genes in collected epigenetic GO terms.

Sheet 3: GO term enrichments for genes containing at least one cluster call containing a CAAS residue. (Left) Combining all calls across all CAAS residues. (Right) Gene sets generated by separating cluster calls by the scenario annotation of individual CAAS hits.

Sheet 4: Domain enrichments for mutational clusters directly overlapping PFAM domains within proteins.

Sheet 5: (Left) Residue position overlaps between TCGA and CAAS datasets. Dashes under ‘amino_acid_change’ refer to instances of multiple long-lived/short-lived amino acid substitutions in the defined CAAS. (Right) Statistical results comparing the number of overlapping residue positions between TCGA and CAAS datasets to that expected when randomly-shuffling CAAS data (see Methods).

Sheet 6: (Left) Subset of overlapping residues in Sheet 5 which also fell within cluster calls. (Right) Overlapping residues which both fall within cluster calls and overlap domain hits.

Table S4.4: CAAS occurring at protein-protein interfaces.

Sheet 1: (Top) ProtCID entries in which CAAS occur in at least one of the two indicated interaction domains of the interacting protein pair (as labelled in columns ‘Pfam1’ and ‘Pfam2’). See http://dunbrack2.fccc.edu/ProtCiD for details on column information. (Bottom) The CAAS residues hitting the indicated domains within respective proteins.

Sheet 2: (Top) ProtCID entries in which CAAS occur in both indicated interaction domains of the interacting protein pair. (Bottom) CAAS residues hitting the indicated domains within respective protein pairs.

Sheet 3: (Top) ProtCID entries in which CAAS occur in interaction domains, for entries involving homo-dimeric complexes that interact via the same domain type. (Bottom) CAAS residues hitting the indicated domains.

Sheet 4: (Top) ProtCID entries in which CAAS occur in interaction domains, for entries involving homo-dimeric complexes that interact via different domain types. (Bottom) CAAS residues hitting the indicated domains.

Sheet 5: More stringent definition of interactions based on intra-molecular distances in crystal structures. For each indicated scenario, columns are as follows. Columns A-F: Information on the interacting residues between the protein pair. G: Minimum distance (Angstroms) between the two residues. K: Information on the CAAS residue which is an interacting residue in the protein pair. Blanks indicate that the CAAS is on the other protein (geneB). L: Information on the CAAS residue which is an interacting residue in the protein pair. Blanks indicate that the CAAS is on the other protein (geneA). M: PFAM domain hit of residue in column K, if occurring. N: PFAM domain hit of residue in column L, if occurring.

Sheet 6: Enrichment testing results for the occurrence of CAAS residues in protein-protein interaction interfaces. See also Methods.

Sheet 7: (Top) ProtCID entries in which at least one residue in a cluster call overlaps an interaction domain in the indicated protein-protein interaction. Interactions here are shared with the previous CAAS-residue-hit interactions. Highlighted in yellow are interfaces for which cluster calls overlap interaction domains on both proteins in a pair.

(Bottom) Cluster call information for the clusters implicated in the highlighted protein-protein interactions. Highlighted in green is the interaction between POLA1 and POLA2 mentioned in the main text. Columns as in Table S4.3.

Sheet 8: (Left) Protein-protein interaction domain pairs which were overlapped by cluster calls, but are not hit by CAAS residues directly (expanding on the set of domains detailed in Sheet 7). (Right) Cluster call information for the clusters calls overlapping interaction domain pairs.

Appendix C: Supplemental Notes and Methods for Chapter 3

Comparing altered accessibility regions to genomic annotations, epigenetic states, and additional epigenetic datasets:

As past studies have found that altered distribution of certain histone marks (e.g., H3K27ac) are a key feature of fetal to adult epigenetic changes (1–3) as well as epigenetic aging (4), the changes in chromatin accessibility we observe likely also reflects, in part, histone mark modification.

To define the epigenetic context within which our development- and age-altered regions fall, we utilized genome-wide assignments of epigenetic state as defined by the Roadmap Epigenomics Project Consortium (3), which employs a Hidden Markov Model to assign one of several possible epigenetic annotations to 200bp segments of the genome, integrating both chromatin-modification and accessibility datasets to define state probabilities, for different epigenomes (e.g. skin, brain tissues, etc.). Given that our altered regions were defined using a pan-tissue approach, for each 200bp segment we subset those epigenetic states defined for adult tissue samples, and took the state definition recurrent in the majority of samples as an 'adult-majority' assignment (see Supplemental Methods). We next intersected our region sets with these assigned segments, comparing the distribution of regions falling within different epigenetic states to the genome-wide distribution of these states to look for biases (Figure S3.4). Adult-biased regions were enriched for epigenetic states associated with transcription, heterochromatin, and repressed Polycomb regions (Table S3.1). Conversely, fetal-biased regions were enriched for states associated with enhancers, promoters, and 'primary DNase', while also showing a more moderate enrichment for repressed Polycomb regions. Likewise, old-biased regions were enriched for heterochromatin and quiescent states, while young-biased regions were enriched for all other states (Table S3.1). By intersecting the fetal and adult as well as young and old-biased regions, we saw that the enrichments for different fetal and adult sets - i.e. adult-biased with heterochromatic states, fetal-biased with euchromatic states - overrode the young-biased and old-biased enrichment patterns (Figure S3.4). Utilizing publicly-

available epigenetic datasets and annotations through the LOLA (5) software (see Supplemental Methods), we again saw overlaps of the adult-biased region set for genomic annotations of 'repressed segments' and repeat sequences in this set, similar to the Roadmap epigenetic state results above (Figure S3.4). Considering fetal-biased regions, we observed enrichments for TSS segments, Promoter/enhancer segments, and Vista enhancers, along with annotated CpG islands. We also saw similar enrichments for young- and old-biased sets (relative to their Roadmap enrichment results), and again saw the overriding fetal and adult patterns of enrichments in intersection sets (Figure S3.5).

We next sought to validate the expected correspondence between development-associated chromatin accessibility and histone modifications, first using an independent dataset of fetal ChIP-seq experiments(1). This study defined fetal bivalent promoter regions, which are thought to poise expression of developmental genes for rapid induction upon appropriate signaling(6). Bivalent promoters tended to not be intersected by adult-biased regions, while fetal-biased regions were enriched in these sets ($p < 1e-16$, hypergeometric test, see Supplemental Methods). That these marked promoters responding to developmental signals lose accessibility in adult tissues would be expected(6), suggesting that our approach is capturing signals of epigenetic change in development. As additional validation of correspondence between development-, and potentially age-, associated chromatin accessibility and regions subject to histone modification, we again used LOLA enrichments, along with histone-mark ChIP-seq datasets acquired from primary tissues samples processed by ENCODE(7, 8).

ChIP-seq Analyses:

Given our use of DNA accessibility datasets, which should reflect the state of local chromatin with respect to chemical modifications increasing/decreasing accessibility, there is an expected concordance between open-chromatin regions defined by DNase-I hypersensitivity and the presence of nearby marks for histone post-translational modifications (i.e. histone ChIP-seq data). To first confirm this expected behaviour in our accessibility data obtained from ENCODE, we further obtained ChIP-seq datasets from fetal and adult tissues matching those used in our accessibility analyses (see Table S3.1 for accessions

and metadata). Datasets included H3K27ac (an active, euchromatin mark), H3K27me3 and H3K9me3 (facultative and constitutive heterochromatin marks, respectively). Replicable open-chromatin regions in fetal and adult tissues were compared to their respective called ChIP-seq peak datasets looking for adjacency between accessibility and chromatin marks (within 1kb, see Supplemental Methods). For H3K27ac marks in adult tissues, between 33-82% of replicable DNase peaks in a given tissue had adjacent ChIP-seq peak calls. For H3K27ac in fetal tissues, between 37-71% of replicable DNase peaks had adjacent ChIP-seq peak calls. For H3K27me3 in adult tissues, between 0.6-10% of replicable DNase peaks in a given tissue had adjacent ChIP-seq peak calls. For H3K27me3 in fetal tissues, between 4-27% of replicable DNase peaks had adjacent ChIP-seq peak calls. For H3K9me3 in adult tissues, between 0.05-4% of replicable DNase peaks in a given tissue had adjacent ChIP-seq peak calls. For H3K9me3 in fetal tissues, between 0.19-22% of replicable DNase peaks had adjacent ChIP-seq peak calls. The increased adjacency of DNase regions with H3K27ac (an active mark) compared to H3K27me3 and H3K9me3 (repressive marks) may be expected, given that DNase hypersensitivity should denote more accessible, active regions of chromatin.

We next asked whether the patterns of accessibility change we observed between fetal and adult tissue samples were also evident at the level of histone modifications. We thus applied a similar pipeline to that used in defining altered accessibility to define altered signals for histone marks (using ChIP-seq read coverage as an approximate, continuous metric) (see Supplemental Methods). This resulted in sets of H3K27ac, H3K27me3, and H3K9me3 peaks whose ChIP-seq signal significantly changed across tissues in comparing fetal and adult samples. Conditioning on the above DNase/ChIP-seq adjacency, we first asked whether significantly-DA DNase peaks tended to be adjacent to altered H3K27ac ChIP-seq peaks, above the general expectation for DNase peaks nearby H3K27ac peaks. We observed a 1.21 fold-change (FC) increase in the adjacency of altered DNase and ChIP-seq peaks (hypergeometric test p-value < 1e-16). Given this, we next asked whether, for these adjacent pairs, directionality was shared (i.e. DNase peaks gaining accessibility are adjacent to H3K27ac peaks gaining signal). We found that, of these adjacent pairs, those sharing direction (i.e. adult-biased DNase, adult-biased H3K27ac ChIP-seq) pairs were significantly over-represented (1.72 FC and 1.19 FC for adult/adult-biased and fetal/fetal-biased, respectively, hypergeometric tests comparing overlaps of sets, adjusted p-values < 1e-16).

We similarly checked this adjacency with H3K9me3 peaks changing signal across fetal/adult tissues. We did see a significantly-greater adjacency between significantly-DA DNase peaks and these altered H3K9me3 peaks, above general DNase/H3K9me3 adjacency (1.13 FC increase, hypergeometric test p-value < 1e-16). Of these adjacent pairs, those sharing direction (i.e. adult-biased DNase, adult-biased H3K9me3 ChIP-seq) were significantly under-represented, while those opposing direction were over-represented (1.152 FC and 1.212 FC for adult-biased DNase/fetal-biased H3K9me3 and fetal-biased DNase/adult-biased H3K9me3, respectively, hypergeometric tests comparing overlaps of sets, adjusted p-value < 1e-16). This follows with an expectation that regions gaining constitutive heterochromatic marks should lose local DNA accessibility, and vice-versa.

Next, we considered the adjacency of H3K27me3 changing signal across fetal/adult tissues. We did observe a slight, but significant, increased adjacency between significantly-DA DNase peaks and altered H3K27me3 peaks, above general DNase/H3K27me3 adjacency (1.03 FC increase, hypergeometric test p-value < 1e-16). Of these, those sharing direction (i.e. adult-biased DNase, adult-biased H3K27me3 ChIP-seq) were significantly over-represented (1.20 FC and 1.38 FC for adult/adult-biased and fetal/fetal-biased, respectively, hypergeometric tests comparing overlaps of sets, adjusted p-values < 1e-16).

Finally, we compared adjacent/overlapping (i.e., within 1 kb) developmentally-altered histone signals across different marks. For a given developmentally-altered H3K27ac peak, adjacent H3K27me3 peaks tended to also change (1.24 FC enrichment, hypergeometric test p-value < 1e-16), with regions gaining H3K27ac signal tending to lose adjacent H3K27me3 signal over development and vice-versa (1.59 FC and 1.09 FC for adult-biased H3K27ac/fetal-biased H3K27me3 and fetal-biased H3K27ac/adult-biased H3K27me3, respectively, adjusted p-values < 1e-16 and 1.9e-6, respectively). Comparing adjacent H3K27me3 and H3K9me3 developmentally-altered peaks, we observed opposing patterns, which may reflect their associations with predominantly facultative and constitutive heterochromatin, respectively. Altered H3K27ac and H3K9me3 peaks showed a small but significant degree of adjacency (~3%, 1.19 FC enrichment, hypergeometric test p-value < 1e-16), though the direction change of adjacent peaks were

not consistently biased between adult/adult-biased, fetal/fetal-biased, etc., which may reflect the limited number of adjacent pairs (data not shown).

We also considered the LOLA enrichments for external histone-mark datasets, observing that adult-biased regions showed strong enrichments with ChIP-seq datasets for repressive histone modifications H3K36me3, H3K9me3, and H3K27me3 (see Figure S3.5). Conversely, fetal-biased regions showed enrichments for both active (including H3K4me2/3, H3K9ac) and repressive (including H3K9me3 and H3K27me3) histone modifications.

Clock Sites Analysis:

Given the substantial literature on changes in DNA-level methylation across both development and aging, and the observed enrichments for annotated CpG sites in the above LOLA analyses, we next looked for correspondence between our development- and age-altered region sets and CpG sites. In particular, we considered so-called 'clock sites' capable of predicting age across the entire lifespan(9–11). Firstly, we re-confirmed the enrichment of CpG sites within developmental and age-altered DNase regions using UCSC annotated CpG sites (see Supplemental Methods), then confirmed that this enrichment held for clock sites, observing a small but significant capturing of these sites by developmentally-altered regions (40 of 353 clock sites, p-value < 1e-3 against 1000 randomized region sets). Of these regions, we saw that the fetal-biased set were enriched for overlaps with both clock sites losing methylation with age (hypo-methylated sites) and those gaining methylation with age (hyper-methylated sites), while the adult-biased set was not enriched for either set. We also saw a significant enrichment for clock sites by age-altered regions (16 of 353 clock sites, p-value < 1e-3 against 1000 randomized region sets). Of these, young-biased regions were enriched for overlaps of both hyper- and hypo-methylated clock sites, while we found no overlaps for clock sites with old-biased regions. Finally, we looked for overlaps between clock sites and our region sets at the gene-locus level – clock sites tied with particular genes (e.g. due to falling within promoter or gene-body regions) which overlap gene loci we associated with our region sets (see Supplemental Methods). This yielded significant overlaps for genes associated with developmentally-altered regions (58 genes, hypergeometric p-value = 0.005), though not those

associated with age-altered regions (12 genes, hypergeometric p-value = 0.19), and we observed no significant biases in direction sharing (e.g. old-age-associated genes and hypo-methylated regions – chi-sq test p-value > 0.05) (see Table S3.1).

Promoter capture datasets:

To better identify biological process whose cis-regulatory activity are subject to change we made use of a compendium of promoter-capture Hi-C interactions(12) (see Supplemental Methods) to identify possible promoter contacts made by our region sets. We also sought to incorporate accessibility information for gene promoters (in addition to the regions contacting them), and did this by (1) intersecting gene promoters with adult- or fetal-biased regions, or (2) similar to our treatment of region accessibility changes we also assessed promoter accessibility using DNase-seq read coverage across tissue samples (Figure S3.6, Supplemental Methods). Genome-wide, adult-biased regions tended to have more putative promoter contacts than fetal-biased regions, while old-biased regions tended to have less putative contacts than young-biased regions (zero-hurdle modeling, p-value << 1e-16). Gene promoters gaining accessibility are preferentially contacted by adult-biased regions, with those losing accessibility contacted by more fetal-biased regions than expected (chi-sq test, p < 1e-16), patterns which held when considering young- and old-age accessibility (chi-sq test, p < 1e-16). This bias was also true when considering gene promoter accessibility defined by intersection with our development- and age-altered region sets (see Supplemental Methods). In the context of enhancer-promoter interaction, we observed enrichments in the adult-biased set for gene-ontology terms associated with immune response, sensory perception, and keratinization (Table S3.2). Conversely, fetal-biased sets were enriched for many developmental terms, as well as terms relating to cellular proliferation and TGF-B signaling (Table S3.2). Echoing the fetal-biased enrichments, we found that old-biased regions were weakly enriched (adjusted p-value = 0.037) for chemokine-response terms, as well as sensory perception. However, no significant term enrichments were observed for young-biased regions and promoters.

As an additional means to consider the sets of genomic loci in which our development- and age-altered sets are distributed, we used the GREAT genome-ontology tool (see description of GREAT in

Supplemental Methods). Fetal-biased regions were located near genes associated with several developmentally-related terms, such as ‘animal organ morphogenesis’ and ‘embryo development’ (Table S3.2). The adult-biased region set yielded enrichments relating to immune processes, such as ‘innate immune response’ and ‘immune effector process’, as well as terms related to keratinization (Table S3.2). Young-biased regions were enriched for terms relating to cell-cycling, such as ‘mitotic cell cycle process’. Enrichments for old-biased regions were associated with immune processes such as ‘regulation of defense response’, while also hitting terms related to DNA break repair and ‘negative regulation of telomere maintenance’ (Table S3.2). Interestingly, when intersecting the fetal/adult and young/old-biased regions we saw a number of additional GREAT terms, while many signals persisted in intersect sets (Table S3.2). For example, adult-biased regions which were also more accessible in older-adult samples were enriched for the ‘positive regulation of immune response’ term; a signal of post-natal development of immune function would be expected(13) and that this signal persists into old-age might suggest that we also capture signals of inappropriate immune system behaviour (so-called ‘inflammaging’(14)).

RNA-seq expression datasets:

Given the biological signals we observed by associating our region sets with gene loci, we next looked to see if similar signals are evident with tissue expression datasets. We utilized ENCODE RNA-seq datasets(8) for fetal and adult tissues – however, given the limited availability of adult tissue samples we performed a less-stringent method for identifying genes whose expression changes over development (see Supplemental Methods). These broad sets of genes yielded similar enrichments to those seen previously on the regulatory level, with genes generally less-expressed in adult tissues enriched for terms involved in growth (e.g. cell-cycling) and chromatin regulation, while those generally more-expressed in adult tissues enriched for terms relating to immune response (e.g. ‘humoral immune response’), sensory perception and keratinization (Table S3.2). These gene sets significantly overlapped those genes associated with adult-biased and fetal-biased regions (all genes – 1.11 FC enrichment, hypergeometric p-value = 6.73e-10) and tended to share directionality (chi-sq test, p-value < 1e-16).

We performed a similar expression analysis using adult tissue samples, stratified by the same age categories used in our accessibility analyses, for those adult tissues available from the GTEx

dataset(15) which overlapped our adult-tissue accessibility datasets (brain, heart, lung, muscle and stomach) (see Supplemental Methods). Genes generally less-expressed in older samples were enriched for terms relating to growth, including cell-cycling, mitochondrial function, and protein synthesis/turnover (Table S3.2).

Conversely, genes generally more-expressed in older samples were enriched for terms relating to development, including terms such as 'ECM organization', 'ossification' and 'angiogenesis'. Whether or not this follows with the suggested role for aberrant mysregulation of developmental pathways in aging biology(16, 17) signalling pathways, is unclear however. Comparing these aging accessibility and expression-defined gene sets we did not observe significant overlaps (hypergeometric test, 1.04 FC, p-value = 0.19); this may be the result of a disconnect between epigenetic dysregulation and expression changes with aging at particular loci.

Finally, we looked for overlaps between gene expression in our fetal/adult and young/old-adult comparisons, finding that genes broadly less-expressed in adult tissues (relative to fetal) are also less expressed in older adult tissues (hypergeometric, p = < 1e-16). While we did not see significant overlap in the adult-biased/old-age-biased expression sets, those genes which did overlap were enriched for immune response terms similar to those seen in the adult-biased set (data not shown).

Divergent sequence intersection enrichments:

We took an aggregated set of sequences showing increased divergence along the human lineage (18–23) and intersected these with our region sets. Subsequently, we assigned each intersection to the nearest annotated gene, and asked whether these elements are actually contacted by these nearby genes via the promoter-capture datasets we had previously integrated with our region sets. These intersections, as well as whether the nearest annotated gene shows some contact data for the indicated region, are presented in Table S3.3. We highlight two example loci, one associated with the fetal-biased region set, the other with the young-biased region set (both of these sets showing general enrichments for overlaps with our aggregated sequence-divergence set, see Figure 3.2B, Figure S3.7, Table S3.3).

A region losing accessibility in adult tissues (i.e. a 'fetal-biased' region) intersects a human-accelerated region(20) intronic to FGF1, a fibroblast growth factor associated with numerous developmental

processes as well as tissue repair(24); this region also has promoter-capture data to suggest contact with the FGF1 promoter. A region losing accessibility in old-adult tissue intersects a human-accelerated region(20) intronic to the PKNOX2 gene, and which also has promoter-capture data to suggest contact with the PKNOX2 promoter. This region lies downstream of the variant rs590211, which has previously been identified in a GWAS of extreme longevity(25, 26).

Comparing sequence diversity between region sets:

Given the patterns of our different region sets in terms of the presence of common human sequence variation (relative to genomic backgrounds and other features, see Figure 3.2C), we directly compared the occurrence of common variants in different sets to one another in humans, chimps and gorillas (Table S3.3). Within humans, fetal-biased regions tended to have far lower variation when compared to every other set, with the exception of young-biased regions (for which the difference was insignificant). Conversely, adult-biased regions had greater variation when compared to every other set, with the exception of old-biased regions (which had higher variation). Accordingly, old-biased regions tended to have greater variation when compared to young-biased regions. Within both chimpanzees and gorillas these differences between accessibility-altered region sets were similarly observed (Table S3.3).

Developmental trait GWAS:

Considering our region sets comparing fetal/adult accessibility changes, we would expect that regions (which may potentially act as regulatory elements) more accessible in fetal tissues may have more of an impact on developmental processes than those regions less accessible in fetal tissues, and vice-versa when considering processes such as tissue homeostasis (e.g. in adult tissues). Therefore, in addition to our focus on aging-associated diseases/traits, we similarly collected a set of developmental traits/disease GWAS to confirm this expected behaviour with regards to developmental processes.
We observed that fetal-biased regions trended towards having greater numbers of nearby significance-thresholded SNPs (reported association p-value < 1e-6) compared to a general DNase background set

across almost all traits used (with the exception of childhood epilepsy). Significant enrichments (hypergeometric test, adjusted p-value < 0.05) were limited to birthweight(27) and height (28), though this may be due to the larger number of SNPs nearby target/background sets observed with these traits (see Table S3.4).

Conversely, adult-biased regions trended towards having decreased numbers of nearby significance-thresholded SNPs across almost all traits used (with the exception of childhood epilepsy). Significant (hypergeometric test, adjusted p-value < 0.05) depletions were observed for birth length, maternal-effect birth weight, childhood BMI, fetal-effect birth weight, gestational-duration and height (Table S3.4).

Longevity GWAS:

Given the patterns of association with our altered-accessibility region sets and aging-associated diseases, we also considered four different GWAS summary-statistics datasets for parental lifespan (29, 30). Compared to DNase regions generally, we observed that fetal-biased regions were not enriched for nearby significance-thresholded longevity SNPs (and trended slightly towards depletion). By contrast, adult-biased regions were significantly enriched for the nearby presence of such variants (hypergeometric test, adjusted p-value < 0.05). Similar to adult-biased regions, young-biased regions were significantly-enriched for two of the four longevity datasets, trending slightly with a third. Old-biased regions were neither significantly enriched nor depleted for longevity GWAS signals, unlike what was seen for aging-associated diseases in general.

Effect-size Distributions:

In addition to determining whether or not a given variant can act to significantly impact disease heritability, the epigenetic state of a region may also determine the magnitude of this impact. For those variants falling nearby developmentally-altered regions, we also considered the reported effect size for their respective diseases. We observed 40 diseases for which variants nearby adult-biased regions had significantly greater absolute effect sizes, compared to only 3 diseases for which nearby variants had significantly reduced effect sizes (Table S3.4). Given that lowering significance thresholds can increase

the amount of heritable variation explained for a given trait, we also considered the effect size distribution of all variants falling near our region sets. Nearly all diseases had biased distributions, with the majority (106 of 127) having larger absolute effect sizes for adult-biased regions (Table S3.4).

Per-disease enrichment testing:

For each GWAS set, we defined single nucleotide polymorphisms (SNPs) with strong association signals (p-value < 1e-6) and looked for the presence of nearby epigenetically-altered regions (Supplemental Methods). We observed that, generally, our accessibility data were enriched for nearby variants (Table S3.4), which is expected given that these data will capture non-coding regulatory elements which are concentrated for GWAS signal(31). First considering accessibility change between fetal and adult tissues, we found that of this general enrichment adult-biased regions associate with a significant proportion of variants across a majority of diseases, while fetal-biased regions associated with significantly less variants than expected (Table S3.4).

We next considered the effects of age-associated accessibility changes on age-related disease GWAS signals. Unexpectedly, we observed that old-biased regions, unlike adult-biased regions, are actually depleted of nearby strong variants across the majority of age-related diseases, while young-biased regions are enriched for such signals (Table S3.4). Furthermore, we found that for intersections of development and age-altered regions that this age-associated behaviour outweighs the earlier development behaviour. Of the general enrichment in adult-biased regions, a significant portion of this can be attested to adult-biased regions which lose accessibility in old-age (i.e. young-biased regions), while adult-biased regions which gain accessibility in old-age are actually depleted for such signals (Table S3.4). Conversely, of the general depletion in fetal-biased regions, an insignificant portion of this can be attested to fetal-biased, old-biased region intersects (hypergeometric test adjusted p-value > 0.05), while those strong variants which do fall nearby fetal-biased regions tend to be concentrated near those regions also considered young-biased (Table S3.4).

Gene set ranking tests:

To confirm the behaviour of our within-disease gene ranking strategy (see Supplemental Methods), we defined a positive-control gene set which would be expected to be strongly-associated with aging diseases using the GO term 'homeostatic process' (GO:0042592). When compared to randomly-sampled gene sets this set had significantly-increased cross-disease gene rankings (Table S3.4). As a negative control, we took a gene set which would not be expected to be strongly associated with aging diseases, those involved in the development of reproductive structures (GO:0003006). This set did not have significantly-increased cross-disease gene rankings.

When looking at gene sets defined by RNA-seq data, we found that genes generally less expressed in adult tissues (fetal-biased) were enriched for cross-disease GWAS signals, while genes more expressed in adults were actually significantly depleted for such signals (Table S3.4). Gene loci with increased expression in older adult tissues were enriched for GWAS signals, as were loci with decreased older-adult expression – suggesting the possibility that a mixture of genes increasing and decreasing expression over time may additively contribute to aging disease biology. It is worth noting that the fetal-biased (expression) genes significantly overlap with young-biased genes (defined by expression), possibly explaining the shared enrichment for GWAS signals, while adult-biased and old-biased genes (by expression) did not significantly overlap - though this overlap set itself, containing a number of immune-related genes, was enriched for GWAS signals (data not shown).

Cross-disease gene ranking genome-wide:

It has been suggested that the highly polygenic nature of complex traits and diseases reflects cumulative regulatory modification to a 'core' set of genes who functions most proximately in relevant biology (i.e., the 'Omnigenic model')(32). If this is indeed the case, we would expect that, for age-associated diseases across multiple tissues, those genes most involved with general pan-tissue aging processes would represent a 'core' set of genes whose dysregulation contribute to heritable risk across aging-associated diseases. We took an unbiased approach to relevant gene discovery, identifying a putative set of 'core'

aging-related genes solely on the basis of aggregate GWAS signals genome-wide (without considering accessibility change) (Supplemental Methods). The resulting set of genes was enriched for terms relating to keratinization, sensory perception of smell, and neuron-related terms (e.g. glutamate receptor signaling) (Table S3.4). We previously observed the former two terms in our region-association analyses, which may suggest that the effects of gene clustering (e.g. clustering of keratin genes, olfactory receptors) may bias our locus ranking method. We note that similar enrichments for these terms in our fetal/adult RNA-seq analyses were observed (Table S3.2), though whether this GWAS signal – expression - accessibility concordance is due to broad changes in accessibility and subsequent transcription in gene clusters is unclear. The fact that we observe consistent enrichments for keratinization and smell perception using the RRA-based method may indicate that this method is particularly sensitive to gene-clustering effects.

Our per-disease GWAS analyses suggested the importance of altered epigenetic state, particularly that which occurs between young/old adult tissues, in considering the risk association of variants with aging-associated diseases. Therefore, we looked for consistent cross-set ranking using variants occurring nearby age-association regions (Supplemental Methods). Again, applying an RRA-based method to different accessibility region sets yielded broadly similar terms relating to keratinization and smell perception. However, when applying a functional gene-set enrichment analysis (FGSEA)-based method, we saw greater differentiation in enrichment results. Ranking genes based on variants nearby fetal-biased regions yielded terms relating to developmental processes (e.g. embryonic development, skeletal system morphogenesis), while considering adult-biased regions again yielded enrichments for keratinization. Young-biased regions yielded enrichments for ‘histone deacetylation’ (discussed in more detail in main text), as well as terms relating to viral infection (e.g. ‘viral gene expression’). Finally, old-biased regions yielded the previously-seen enrichments for smell perception and keratinization, though also including enrichments for immune processes (e.g. ‘antibacterial humoral response’) and DNA methylation.

Intersection set comparisons:

We compared our developmentally-associated and age-associated regions directly, here explicitly comparing age-associated regions with developmental regions not changing with age as a more stringent

contrast (Supplemental Methods). Here we also saw the much-stronger biasing of young/old-biased regions; old-biased regions associating with significantly less cross-trait heritability than fetal-biased, while young-biased regions associated with significantly more heritability than all other sets (Table S3.4). Comparing development and age-altered intersection sets, we found that the strong disparity in GWAS associations between the young -and old-biased region sets outweighed the differences between the fetal- and adult-biased region sets. For example, the young-biased /fetal-biased set had the second-highest average cross-trait association, despite fetal-biased regions generally being associated with weaker GWAS signals in the previous fetal/adult comparison. Conversely, the weaker GWAS signals associated with the old-biased region set outweighed the generally-higher signals of the adult-biased region set, actually having a lower average cross-trait association than fetal-biased regions not significantly changing accessibility in the young/old accessibility analysis (see Table S3.4).

Supplementary Materials and Methods:

Please see the attached supplemental document.

Supplemental Text and Methods Citations:

1. I. Yan, H. Guo, B. Hu, R. Li, J. Yong, Y. Zhao, X. Zhi, X. Fan, F. Guo, X. Wang, W. Wang, Y. Wei, Y. Wang, L. Wen, J. Qiao, F. Tang, Epigenomic landscape of human fetal brain, heart, and liver. J. Biol. Chem. 291, 4386–4398 (2016).

2. J. Zhu, M. Adli, J. Y. Zou, G. Verstappen, M. Coyne, X. Zhang, T. Durham, M. Miri, V. Deshpande, P. L. De Jager, D. A. Bennett, J. A. Houmard, D. M. Muoio, T. T. Onder, R. Camahort, C. A. Cowan, A. Meissner, C. B. Epstein, N. Shoresh, B. E. Bernstein, Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell. 152, 642–654 (2013).

3. Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R.

S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. Nature. 518, 317–329 (2015).

4. L. N. Booth, A. Brunet, The Aging Epigenome. Mol. Cell. 62, 728–744 (2016).

5. N. C. Sheffield, C. Bock, LOLA: Enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics. 32, 587–589 (2016).

6. P. Voigt, W. W. Tee, D. Reinberg, A double take on bivalent promoters. Genes Dev. 27 (2013), pp. 1318–1338.

7. F. Yue, Y. Cheng, A. Breschi, J. Vierstra, W. Wu, T. Ryba, R. Sandstrom, Z. Ma, C. Davis, B. D. Pope, Y. Shen, D. D. Pervouchine, S. Djebali, R. E. Thurman, R. Kaul, E. Rynes, A. Kirilusha, G. K. Marinov, B. A. Williams, D. Trout, H. Amrhein, K. Fisher-Aylor, I. Antoshechkin, G. DeSalvo, L. H. See, M. Fastuca, J. Drenkow, C. Zaleski, A. Dobin, P. Prieto, J. Lagarde, G. Bussotti, A. Tanzer, O. Denas, K. Li, M. A. Bender, M. Zhang, R. Byron, M. T. Groudine, D. McCleary, L. Pham, Z. Ye, S. Kuan, L. Edsall, Y. C. Wu, M. D. Rasmussen, M. S. Bansal, M. Kellis, C. A. Keller, C. S. Morrissey, T. Mishra, D. Jain, N. Dogan, R. S. Harris, P. Cayting, T. Kawli, A. P. Boyle, G. Euskirchen, A. Kundaje, S. Lin, Y. Lin, C. Jansen, V. S. Malladi, M. S. Cline, D. T. Erickson, V. M. Kirkup, K. Learned, C. A. Sloan, K. R. Rosenbloom, B. L. De Sousa, K. Beal, M. Pignatelli, P. Flicek, J. Lian, T. Kahveci, D. Lee, W. J. Kent, M. R. Santos, J. Herrero, C. Notredame, A. Johnson, S. Vong, K. Lee, D. Bates, F. Neri, M. Diegel, T. Canfield, P. J. Sabo, M. S. Wilken, T. A. Reh, E. Giste, A. Shafer, T. Kutyavin, E. Haugen, D. Dunn, A. P. Reynolds, S. Neph, R. Humbert, R. S. Hansen, M. De Bruijn, L. Selleri, A. Rudensky, S. Josefowicz, R. Samstein, E. E. Eichler, S. H. Orkin, D. Levasseur, T. Papayannopoulou, K. H. Chang, A. Skoultchi, S. Gosh, C. Disteche, P. Treuting, Y. Wang, M. J. Weiss, G. A. Blobel, X. Cao, S. Zhong, T. Wang, P. J. Good, R. F. Lowdon, L. B. Adams, X. Q. Zhou, M. J. Pazin, E. A. Feingold, B. Wold, J. Taylor, A.

Mortazavi, S. M. Weissman, J. A. Stamatoyannopoulos, M. P. Snyder, R. Guigo, T. R. Gingeras, D. M. Gilbert, R. C. Hardison, M. A. Beer, B. Ren, A comparative encyclopedia of DNA elements in the mouse genome. Nature. 515, 355–364 (2014).

8. C. A. Davis, B. C. Hitz, C. A. Sloan, E. T. Chan, J. M. Davidson, I. Gabdank, J. A. Hilton, K. Jain, U. K. Baymuradov, A. K. Narayanan, K. C. Onate, K. Graham, S. R. Miyasato, T. R. Dreszer, J. S. Strattan, O. Jolanki, F. Y. Tanaka, J. M. Cherry, The Encyclopedia of DNA elements (ENCODE): Data portal update. Nucleic Acids Res. 46, D794–D801 (2018).

9. S. Horvath, K. Raj, DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. 19 (2018), pp. 371–384.

10. S. Horvath, V. Mah, A. T. Lu, J. S. Woo, O. W. Choi, A. J. Jasinska, J. A. Riancho, S. Tung, N. S. Coles, J. Braun, H. V. Vinters, L. S. Coles, The cerebellum ages slowly according to the epigenetic clock. Aging (Albany. NY). 7, 294–306 (2015).

11. G. Hannum, J. Guinney, L. Zhao, L. Zhang, G. Hughes, S. V. Sadda, B. Klotzle, M. Bibikova, J. B. Fan, Y. Gao, R. Deconde, M. Chen, I. Rajapakse, S. Friend, T. Ideker, K. Zhang, Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Mol. Cell. 49, 359–367 (2013).

12. I. Jung, A. Schmitt, Y. Diao, A. J. Lee, T. Liu, D. Yang, C. Tan, J. Eom, M. Chan, S. Chee, Z. Chiang, C. Kim, E. Masliah, C. L. Barr, B. Li, S. Kuan, D. Kim, B. Ren, A compendium of promoter-centered long-range chromatin interactions in the human genome. Nat. Genet. 51, 1442–1449 (2019).

13. A. Georgountzou, N. G. Papadopoulos, Postnatal innate immune development: From birth to adulthood. Front. Immunol. 8 (2017), p. 1.

14. C. Franceschi, P. Garagnani, G. Vitale, M. Capri, S. Salvioli, Inflammaging and ‘Garb-aging.’ Trends Endocrinol. Metab. 28, 199–212 (2017).

15. T. Gte. GTEx Consortium, The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–5 (2013).

16. M. V. Blagosklonny, M. N. Hall, Growth and aging: a common molecular mechanism. Aging (Albany. NY). 1 (2009), pp. 357–362.

17. J. P. Magalhães, Programmatic features of aging originating in development: aging mechanisms beyond molecular damage? FASEB J. 26, 4821–4826 (2012).

18. C. P. Bird, B. E. Stranger, M. Liu, D. J. Thomas, C. E. Ingle, C. Beazley, W. Miller, M. E. Hurles, E. T. Dermitzakis, Fast-evolving noncoding sequences in the human genome. Genome Biol. 8, R118 (2007).

19. E. C. Bush, B. T. Lahn, A genome-wide screen for noncoding elements important in primate evolution. BMC Evol. Biol. 8, 17 (2008).

20. R. M. Gittelman, E. Hun, F. Ay, J. Madeoy, L. Pennacchio, W. S. Noble, R. D. Hawkins, J. M. Akey, Comprehensive identification and analysis of human accelerated regulatory DNA. Genome Res. (2015).

21. K. S. Pollard, S. R. Salama, B. King, A. D. Kern, T. Dreszer, S. Katzman, A. Siepel, J. S. Pedersen, G. Bejerano, R. Baertsch, others, Forces shaping the fastest evolving regions in the human genome. PLoS Genet. 2, e168 (2006).

22. S. Prabhakar, J. P. Noonan, S. Pääbo, E. M. Rubin, Accelerated evolution of conserved noncoding sequences in humans. Science (80-.). 314, 786 (2006).

23. D. Kostka, A. K. Holloway, K. S. Pollard, Developmental Loci Harbor Clusters of Accelerated Regions That Evolved Independently in Ape Lineages. Mol. Biol. Evol. 35, 2034–2045 (2018).

24. D. M. Ornitz, N. Itoh, The Fibroblast Growth Factor signaling pathway. Wiley Interdiscip. Rev. Dev. Biol. 4, 215–266 (2015).

25. P. Sebastiani, N. Solovieff, A. T. DeWan, K. M. Walsh, A. Puca, S. W. Hartley, E. Melista, S. Andersen, D. A. Dworkis, J. B. Wilk, R. H. Myers, M. H. Steinberg, M. Montano, C. T. Baldwin, J. Hoh, T. T. Perls, Genetic signatures of exceptional longevity in humans. PLoS One. 7 (2012), doi:10.1371/journal.pone.0029848.

26. J. P. de Magalhães, A. Budovsky, G. Lehmann, J. Costa, Y. Li, V. Fraifeld, G. M. Church, The Human Ageing Genomic Resources: Online databases and tools for biogerontologists. Aging Cell. 8 (2009), pp. 65–72.

27. M. Horikoshi, R. N. Beaumont, F. R. Day, N. M. Warrington, M. N. Kooijman, J. Fernandez-Tajes, B. Feenstra, N. R. van Zuydam, K. J. Gaulton, N. Grarup, J. P. Bradfield, D. P. Strachan, R. Li-Gao, T. S. Ahluwalia, E. Kreiner, R. Rueedi, L.-P. Lyytikäinen, D. L. Cousminer, Y. Wu, E. Thiering, C. A. Wang, C. T. Have, J.-J. Hottenga, N. Vilor-Tejedor, P. K. Joshi, E. T. H. Boh, I. Ntalla, N. Pitkänen, A. Mahajan, E.

M. van Leeuwen, R. Joro, V. Lagou, M. Nodzenski, L. A. Diver, K. T. Zondervan, M. Bustamante, P. Marques-Vidal, J. M. Mercader, A. J. Bennett, N. Rahmioglu, D. R. Nyholt, R. C. W. Ma, C. H. T. Tam, W. H. Tam, S. K. Ganesh, F. J. A. van Rooij, S. E. Jones, P.-R. Loh, K. S. Ruth, M. A. Tuke, J. Tyrrell, A. R. Wood, H. Yaghootkar, D. M. Scholtens, L. Paternoster, I. Prokopenko, P. Kovacs, M. Atalay, S. M. Willems, K. Panoutsopoulou, X. Wang, L. Carstensen, F. Geller, K. E. Schraut, M. Murcia, C. E. M. van Beijsterveldt, G. Willemsen, E. V. R. Appel, C. E. Fonvig, C. Trier, C. M. T. Tiesler, M. Standl, Z. Kutalik, S. Bonàs-Guarch, D. M. Hougaard, F. Sánchez, D. Torrents, J. Waage, M. V. Hollegaard, H. G. de Haan, F. R. Rosendaal, C. Medina-Gomez, S. M. Ring, G. Hemani, G. McMahon, N. R. Robertson, C. J. Groves, C. Langenberg, J. Luan, R. A. Scott, J. H. Zhao, F. D. Mentch, S. M. MacKenzie, R. M. Reynolds, W. L. Lowe, A. Tönjes, M. Stumvoll, V. Lindi, T. A. Lakka, C. M. van Duijn, W. Kiess, A. Körner, T. I. A. Sørensen, H. Niinikoski, K. Pahkala, O. T. Raitakari, E. Zeggini, G. V. Dedoussis, Y.-Y. Teo, S.-M. Saw, M. Melbye, H. Campbell, J. F. Wilson, M. Vrijheid, E. J. C. N. de Geus, D. I. Boomsma, H. N. Kadarmideen, J.-C. Holm, T. Hansen, S. Sebert, A. T. Hattersley, L. J. Beilin, J. P. Newnham, C. E. Pennell, J. Heinrich, L. S. Adair, J. B. Borja, K. L. Mohlke, J. G. Eriksson, E. Widén, M. Kähönen, J. S. Viikari, T. Lehtimäki, P. Vollenweider, K. Bønnelykke, H. Bisgaard, D. O. Mook-Kanamori, A. Hofman, F. Rivadeneira, A. G. Uitterlinden, C. Pisinger, O. Pedersen, C. Power, E. Hyppönen, N. J. Wareham, H. Hakonarson, E. Davies, B. R. Walker, V. W. V. Jaddoe, M.-R. Järvelin, S. F. A. Grant, A. A. Vaag, D. A. Lawlor, T. M. Frayling, G. D. Smith, A. P. Morris, K. K. Ong, J. F. Felix, N. J. Timpson, J. R. B. Perry, D. M. Evans, M. I. McCarthy, R. M. Freathy, Genome-wide associations for birth weight and correlations with adult disease. Nature. 538, 248–252 (2016).

28. L. Yengo, J. Sidorenko, K. E. Kemper, Z. Zheng, A. R. Wood, M. N. Weedon, T. M. Frayling, J. Hirschhorn, J. Yang, P. M. Visscher, Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. Hum. Mol. Genet. 27, 3641–3649 (2018).

29. P. R. H. J. Timmers, N. Mounier, K. Lall, K. Fischer, Z. Ning, X. Feng, A. D. Bretherick, D. W. Clark, X. Shen, T. Esko, Z. Kutalik, J. F. Wilson, P. K. Joshi, Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. Elife. 8, 1–40 (2019).

30. L. C. Pilling, C. L. Kuo, K. Sicinski, J. Tamosauskaite, G. A. Kuchel, L. W. Harries, P. Herd, R. Wallace, L. Ferrucci, D. Melzer, Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. Aging (Albany. NY). 9, 2504–2520 (2017).

31. S. L. Klemm, Z. Shipony, W. J. Greenleaf, Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. 20 (2019), pp. 207–220.

32. E. A. Boyle, Y. I. Li, J. K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 169 (2017), pp. 1177–1186.

33. I. Dunham, A. Kundaje, S. F. Aldred, P. J. Collins, C. A. Davis, F. Doyle, C. B. Epstein, S. Frietze, J. Harrow, R. Kaul, J. Khatun, B. R. Lajoie, S. G. Landt, B. K. Lee, F. Pauli, K. R. Rosenbloom, P. Sabo, A. Safi, A. Sanyal, N. Shoresh, J. M. Simon, L. Song, N. D. Trinklein, R. C. Altshuler, E. Birney, J. B. Brown, C. Cheng, S. Djebali, X. Dong, J. Ernst, T. S. Furey, M. Gerstein, B. Giardine, M. Greven, R. C. Hardison, R. S. Harris, J. Herrero, M. M. Hoffman, S. Iyer, M. Kellis, P. Kheradpour, T. Lassmann, Q. Li, X. Lin, G. K. Marinov, A. Merkel, A. Mortazavi, S. C. J. Parker, T. E. Reddy, J. Rozowsky, F. Schlesinger, R. E. Thurman, J. Wang, L. D. Ward, T. W. Whitfield, S. P. Wilder, W. Wu, H. S. Xi, K. Y. Yip, J. Zhuang, B. E. Bernstein, E. D. Green, C. Gunter, M. Snyder, M. J. Pazin, R. F. Lowdon, L. A. L. Dillon, L. B. Adams, C. J. Kelly, J. Zhang, J. R. Wexler, P. J. Good, E. A. Feingold, G. E. Crawford, J. Dekker, L. Elnitski, P. J. Farnham, M. C. Giddings, T. R. Gingeras, R. Guigó, T. J. Hubbard, W. J. Kent, J. D. Lieb, E. H. Margulies, R. M. Myers, J. A. Stamatoyannopoulos, S. A. Tenenbaum, Z. Weng, K. P. White, B. Wold, Y. Yu, J. Wrobel, B. A. Risk, H. P. Gunawardena, H. C. Kuiper, C. W. Maier, L. Xie, X. Chen, T. S. Mikkelsen, S. Gillespie, A. Goren, O. Ram, X. Zhang, L. Wang, R. Issner, M. J. Coyne, T. Durham, M. Ku, T. Truong, M. L. Eaton, A. Dobin, A. Tanzer, J. Lagarde, W. Lin, C. Xue, B. A. Williams, C. Zaleski, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, P. Batut, I. Bell, K. Bell, S. Chakrabortty, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, G. Li, O. J. Luo, E. Park, J. B. Preall, K. Presaud, P. Ribeca, D. Robyr, X. Ruan, M. Sammeth, K. S. Sandhu, L. Schaeffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, Y. Hayashizaki, A. Reymond, S. E. Antonarakis, G. J. Hannon, Y. Ruan, P. Carninci, C. A. Sloan, K.

Learned, V. S. Malladi, M. C. Wong, G. P. Barber, M. S. Cline, T. R. Dreszer, S. G. Heitner, D. Karolchik, V. M. Kirkup, L. R. Meyer, J. C. Long, M. Maddren, B. J. Raney, L. L. Grasfeder, P. G. Giresi, A. Battenhouse, N. C. Sheffield, K. A. Showers, D. London, A. A. Bhinge, C. Shestak, M. R. Schaner, S. K. Kim, Z. Z. Zhang, P. A. Mieczkowski, J. O. Mieczkowska, Z. Liu, R. M. McDaniell, Y. Ni, N. U. Rashid, M. J. Kim, S. Adar, Z. Zhang, T. Wang, D. Winter, D. Keefe, V. R. Iyer, M. Zheng, P. Wang, J. Gertz, J. Vielmetter, E. C. Partridge, K. E. Varley, C. Gasper, A. Bansal, S. Pepke, P. Jain, H. Amrhein, K. M. Bowling, M. Anaya, M. K. Cross, M. A. Muratet, K. M. Newberry, K. McCue, A. S. Nesmith, K. I. Fisher-Aylor, B. Pusey, G. DeSalvo, S. L. Parker, S. Balasubramanian, N. S. Davis, S. K. Meadows, T. Eggleston, J. S. Newberry, S. E. Levy, D. M. Absher, W. H. Wong, M. J. Blow, A. Visel, L. A. Pennachio, H. M. Petrykowska, A. Abyzov, B. Aken, D. Barrell, G. Barson, A. Berry, A. Bignell, V. Boychenko, G. Bussotti, C. Davidson, G. Despacio-Reyes, M. Diekhans, I. Ezkurdia, A. Frankish, J. Gilbert, J. M. Gonzalez, E. Griffiths, R. Harte, D. A. Hendrix, T. Hunt, I. Jungreis, M. Kay, E. Khurana, J. Leng, M. F. Lin, J. Loveland, Z. Lu, D. Manthravadi, M. Mariotti, J. Mudge, G. Mukherjee, C. Notredame, B. Pei, J. M. Rodriguez, G. Saunders, A. Sboner, S. Searle, C. Sisu, C. Snow, C. Steward, E. Tapanari, M. L. Tress, M. J. Van Baren, S. Washietl, L. Wilming, A. Zadissa, Z. Zhang, M. Brent, D. Haussler, A. Valencia, N. Addleman, R. P. Alexander, R. K. Auerbach, S. Balasubramanian, K. Bettinger, N. Bhardwaj, A. P. Boyle, A. R. Cao, P. Cayting, A. Charos, Y. Cheng, C. Eastman, G. Euskirchen, J. D. Fleming, F. Grubert, L. Habegger, M. Hariharan, A. Harmanci, S. Iyengar, V. X. Jin, K. J. Karczewski, M. Kasowski, P. Lacroute, H. Lam, N. Lamarre-Vincent, J. Lian, M. Lindahl-Allen, R. Min, B. Miotto, H. Monahan, Z. Moqtaderi, X. J. Mu, H. O'Geen, Z. Ouyang, D. Patacsil, D. Raha, L. Ramirez, B. Reed, M. Shi, T. Slifer, H. Witt, L. Wu, X. Xu, K. K. Yan, X. Yang, K. Struhl, S. M. Weissman, L. O. Penalva, S. Karmakar, R. R. Bhanvadia, A. Choudhury, M. Domanus, L. Ma, J. Moran, A. Victorsen, T. Auer, L. Centanin, M. Eichenlaub, F. Gruhl, S. Heermann, B. Hoeckendorf, D. Inoue, T. Kellner, S. Kirchmaier, C. Mueller, R. Reinhardt, L. Schertel, S. Schneider, R. Sinn, B. Wittbrodt, J. Wittbrodt, G. Jain, G. Balasundaram, D. L. Bates, R. Byron, T. K. Canfield, M. J. Diegel, D. Dunn, A. K. Ebersol, T. Frum, K. Garg, E. Gist, R. S. Hansen, L. Boatman, E. Haugen, R. Humbert, A. K. Johnson, E. M. Johnson, T. V. Kutyavin, K. Lee, D. Lotakis, M. T. Maurano, S. J. Neph, F. V. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, E. Rynes, M. E. Sanchez, R. S. Sandstrom, A. O. Shafer, A. B. Stergachis, S. Thomas, B. Vernot, J. Vierstra, S. Vong, H. Wang, M. A.

Weaver, Y. Yan, M. Zhang, J. M. Akey, M. Bender, M. O. Dorschner, M. Groudine, M. J. MacCoss, P. Navas, G. Stamatoyannopoulos, K. Beal, A. Brazma, P. Flicek, N. Johnson, M. Lukk, N. M. Luscombe, D. Sobral, J. M. Vaquerizas, S. Batzoglou, A. Sidow, N. Hussami, S. Kyriazopoulou-Panagiotopoulou, M. W. Libbrecht, M. A. Schaub, W. Miller, P. J. Bickel, B. Banfai, N. P. Boley, H. Huang, J. J. Li, W. S. Noble, J. A. Bilmes, O. J. Buske, A. D. Sahu, P. V. Kharchenko, P. J. Park, D. Baker, J. Taylor, L. Lochovsky, An integrated encyclopedia of DNA elements in the human genome. Nature. 489, 57–74 (2012).

34. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5, 1752–1779 (2011).

35. Y. Yang, J. Fear, J. Hu, I. Haecker, L. Zhou, R. Renne, D. Bloom, L. M. McIntyre, LEVERAGING BIOLOGICAL REPLICATES TO IMPROVE ANALYSIS IN CHIP-SEQ EXPERIMENTS. Comput. Struct. Biotechnol. J. 9, e201401002 (2014).

36. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 841–842 (2010).

37. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25, 2078–9 (2009).

38. R Development Core Team, R: A Language and Environment for Statistical Computing (2008), (available at http://www.r-project.org).

39. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47 (2015).

40. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26, 139–140 (2009).

41. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B, 289–300 (1995).

42. M. Lawrence, R. Gentleman, V. Carey, rtracklayer: An R package for interfacing with genome browsers. Bioinformatics. 25, 1841–1842 (2009).

43. B. Gel, E. Serra, karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics. 33, 3088–3090 (2017).

44. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 32, 2847–9 (2016).

45. H. Chen, P. C. Boutros, VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 12, 35 (2011).

46. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. Genome Res. 12, 996–1006 (2002).

47. S. Horvath, DNA methylation age of human tissues and cell types. Genome Biol. 14 (2013), doi:10.1186/gb-2013-14-10-r115.

48. D. Karolchik, G. P. Barber, J. Casper, H. Clawson, M. S. Cline, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, others, The UCSC genome browser database: 2014 update. Nucleic Acids Res. 42, D764--D770 (2013).

49. B. Gel, A. Díez-Villanueva, E. Serra, M. Buschbeck, M. A. Peinado, R. Malinverni, regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics. 32, btv562 (2015).

50. D. Karolchik, A. S. Hinrichs, T. S. Furey, K. M. Roskin, C. W. Sugnet, D. Haussler, W. J. Kent, The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 32, 493D – 496 (2004).

51. S. Jackman, {pscl}: Classes and Methods for {R} Developed in the Political Science Computational Laboratory (2017), (available at https://github.com/atahk/pscl/).

52. A. Zeileis, C. Kleiber, S. Jackman, Regression Models for Count Data in {R}. J. Stat. Softw. 27 (2008) (available at http://www.jstatsoft.org/v27/i08/).

53. C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495 (2010).

54. G. Yu, L. G. Wang, Y. Han, Q. Y. He, ClusterProfiler: An R package for comparing biological themes among gene clusters. Omi. A J. Integr. Biol. 16, 284–287 (2012).

55. R. Kolde, S. Laur, P. Adler, J. Vilo, Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 28, 573–580 (2012).

56. B. A. Benayoun, E. A. Pollina, P. P. Singh, S. Mahmoudi, I. Harel, K. M. Casey, B. W. Dulken, A. Kundaje, A. Brunet, Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. Genome Res. 29, 697–709 (2019).

57. J. F. Fullard, M. E. Hauberg, J. Bendl, G. Egervari, M.-D. Cirnaru, S. M. Reach, J. Motl, M. E. Ehrlich, Y. L. Hurd, P. Roussos, An atlas of chromatin accessibility in the adult human brain. Genome Res. 28, 1243–1252 (2018).

58. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods. 10, 1213 (2013).

59. D. Richard, Z. Liu, J. Cao, A. M. Kiapour, J. Willen, S. Yarlagadda, E. Jagoda, V. B. Kolachalama, J. T. Sieker, G. H. Chang, P. Muthuirulan, M. Young, A. Masson, J. Konrad, S. Hosseinzadeh, D. E. Maridas, V. Rosen, R. Krawetz, N. Roach, T. D. Capellini, Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. Cell. 0 (2020), doi:10.1016/j.cell.2020.02.057.

60. J. Fox, S. Weisberg, An R Companion to Applied Regression, Third Edition (2019; https://socialsciences.mcmaster.ca/jfox/Books/Companion/).

61. M. L. Delignette-Muller, C. Dutang, fitdistrplus : An R Package for Fitting Distributions. J. Stat. Softw. 64, 1–34 (2015).

62. J. Mun, Advanced analytical models : over 800 models and 300 applications from the Basel II Accord to Wall Street and beyond (Wiley, 2008; https://dl.acm.org/citation.cfm?id=1386324).

63. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell. 38, 576–589 (2010).

64. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 20, 110–121 (2010).

65. R. A. Gibbs, E. Boerwinkle, H. Doddapaneni, Y. Han, V. Korchina, C. Kovar, S. Lee, D. Muzny, J. G. Reid, Y. Zhu, J. Wang, Y. Chang, Q. Feng, X. Fang, X. Guo, M. Jian, H. Jiang, X. Jin, T. Lan, G. Li, J. Li, Y. Li, S. Liu, X. Liu, Y. Lu, X. Ma, M. Tang, B. Wang, G. Wang, H. Wu, R. Wu, X. Xu, Y. Yin, D. Zhang, W. Zhang, J. Zhao, M. Zhao, X. Zheng, E. S. Lander, D. M. Altshuler, S. B. Gabriel, N. Gupta, N. Gharani, L. H. Toji, N. P. Gerry, A. M. Resch, P. Flicek, J. Barker, L. Clarke, L. Gil, S. E. Hunt, G. Kelman, E. Kulesha, R. Leinonen, W. M. McLaren, R. Radhakrishnan, A. Roa, D. Smirnov, R. E. Smith, I. Streeter, A. Thormann, I. Toneva, B. Vaughan, X. Zheng-Bradley, D. R. Bentley, R. Grocock, S. Humphray, T. James, Z. Kingsbury, H. Lehrach, R. Sudbrak, M. W. Albrecht, V. S. Amstislavskiy, T. A. Borodina, M. Lienhard, F. Mertes, M. Sultan, B. Timmermann, M.-L. Yaspo, E. R. Mardis, R. K. Wilson, L. Fulton, R. Fulton, S. T. Sherry, V. Ananiev, Z. Belaia, D. Beloslyudtsev, N. Bouk, C. Chen, D. Church, R. Cohen, C. Cook, J. Garner, T. Hefferon, M. Kimelman, C. Liu, J. Lopez, P. Meric, C. O'Sullivan, Y. Ostapchuk, L. Phan, S. Ponomarov, V. Schneider, E. Shekhtman, K. Sirotkin, D. Slotta, H. Zhang, G. A. McVean, R. M. Durbin, S. Balasubramaniam, J. Burton, P. Danecek, T. M. Keane, A. Kolb-Kokocinski, S. McCarthy, J. Stalker, M. Quail, J. P. Schmidt, C. J. Davies, J. Gollub, T. Webster, B. Wong, Y. Zhan, A. Auton, C. L. Campbell, Y. Kong, A. Marcketta, R. A. Gibbs, F. Yu, L. Antunes, M. Bainbridge, D. Muzny, A. Sabo, Z. Huang, J. Wang, L. J. M. Coin, L. Fang, X. Guo, X. Jin, G. Li, Q. Li, Y. Li, Z. Li, H. Lin, B. Liu, R. Luo, H. Shao, Y. Xie, C. Ye, C. Yu, F. Zhang, H. Zheng, H. Zhu, C. Alkan, E. Dal, F. Kahveci, G. T. Marth, E. P. Garrison, D. Kural, W.-P. Lee, W. Fung Leong, M. Stromberg, A. N. Ward, J. Wu, M. Zhang, M. J. Daly, M. A. DePristo, R. E. Handsaker, D. M. Altshuler, E. Banks, G. Bhatia, G. del Angel, S. B. Gabriel, G. Genovese, N. Gupta, H. Li, S. Kashin, E. S. Lander, S. A. McCarroll, J. C. Nemesh, R. E. Poplin, S. C. Yoon, J. Lihm, V. Makarov, A. G. Clark, S. Gottipati, A. Keinan, J. L. Rodriguez-Flores, J. O. Korbel, T. Rausch, M. H. Fritz, A. M. Stütz, P. Flicek, K. Beal, L. Clarke, A. Datta, J. Herrero, W. M. McLaren, G. R. S. Ritchie, R. E. Smith, D. Zerbino, X. Zheng-Bradley, P. C. Sabeti, I. Shlyakhter, S. F. Schaffner, J. Vitti, D. N. Cooper, E. V. Ball, P. D. Stenson, D. R. Bentley, B. Barnes, M. Bauer, R. Keira Cheetham, A. Cox, M. Eberle, S. Humphray, S. Kahn, L. Murray, J. Peden, R. Shaw, E. E. Kenny, M. A. Batzer, M. K. Konkel, J. A. Walker, D. G. MacArthur, M. Lek, R. Sudbrak, V. S. Amstislavskiy, R. Herwig, E. R. Mardis, L. Ding, D. C. Koboldt, D. Larson, K. Ye, S. Gravel, A. Swaroop, E. Chew, T. Lappalainen, Y. Erlich, M. Gymrek, T. Frederick Willems, J. T. Simpson, M. D. Shriver, J. A. Rosenfeld, C. D. Bustamante, S. B. Montgomery,

F. M. De La Vega, J. K. Byrnes, A. W. Carroll, M. K. DeGorter, P. Lacroute, B. K. Maples, A. R. Martin, A. Moreno-Estrada, S. S. Shringarpure, F. Zakharia, E. Halperin, Y. Baran, C. Lee, E. Cerveira, J. Hwang, A. Malhotra, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, F. C. L. Hyland, D. W. Craig, A. Christoforides, N. Homer, T. Izatt, A. A. Kurdoglu, S. A. Sinari, K. Squire, S. T. Sherry, C. Xiao, J. Sebat, D. Antaki, M. Gujral, A. Noor, K. Ye, E. G. Burchard, R. D. Hernandez, C. R. Gignoux, D. Haussler, S. J. Katzman, W. James Kent, B. Howie, A. Ruiz-Linares, E. T. Dermitzakis, S. E. Devine, G. R. Abecasis, H. Min Kang, J. M. Kidd, T. Blackwell, S. Caron, W. Chen, S. Emery, L. Fritsche, C. Fuchsberger, G. Jun, B. Li, R. Lyons, C. Scheller, C. Sidore, S. Song, E. Sliwerska, D. Taliun, A. Tan, R. Welch, M. Kate Wing, X. Zhan, P. Awadalla, A. Hodgkinson, Y. Li, X. Shi, A. Quitadamo, G. Lunter, G. A. McVean, J. L. Marchini, S. Myers, C. Churchhouse, O. Delaneau, A. Gupta-Hinch, W. Kretzschmar, Z. Iqbal, I. Mathieson, A. Menelaou, A. Rimmer, D. K. Xifara, T. K. Oleksyk, Y. Fu, X. Liu, M. Xiong, L. Jorde, D. Witherspoon, J. Xing, E. E. Eichler, B. L. Browning, S. R. Browning, F. Hormozdiari, P. H. Sudmant, E. Khurana, R. M. Durbin, M. E. Hurles, C. Tyler-Smith, C. A. Albers, Q. Ayub, S. Balasubramaniam, Y. Chen, V. Colonna, P. Danecek, L. Jostins, T. M. Keane, S. McCarthy, K. Walter, Y. Xue, M. B. Gerstein, A. Abyzov, S. Balasubramanian, J. Chen, D. Clarke, Y. Fu, A. O. Harmanci, M. Jin, D. Lee, J. Liu, X. Jasmine Mu, J. Zhang, Y. Zhang, Y. Li, R. Luo, H. Zhu, C. Alkan, E. Dal, F. Kahveci, G. T. Marth, E. P. Garrison, D. Kural, W.-P. Lee, A. N. Ward, J. Wu, M. Zhang, S. A. McCarroll, R. E. Handsaker, D. M. Altshuler, E. Banks, G. del Angel, G. Genovese, C. Hartl, H. Li, S. Kashin, J. C. Nemesh, K. Shakir, S. C. Yoon, J. Lihm, V. Makarov, J. Degenhardt, J. O. Korbel, M. H. Fritz, S. Meiers, B. Raeder, T. Rausch, A. M. Stütz, P. Flicek, F. Paolo Casale, L. Clarke, R. E. Smith, O. Stegle, X. Zheng-Bradley, D. R. Bentley, B. Barnes, R. Keira Cheetham, M. Eberle, S. Humphray, S. Kahn, L. Murray, R. Shaw, E.-W. Lameijer, M. A. Batzer, M. K. Konkel, J. A. Walker, L. Ding, I. Hall, K. Ye, P. Lacroute, C. Lee, E. Cerveira, A. Malhotra, J. Hwang, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, D. W. Craig, N. Homer, D. Church, C. Xiao, J. Sebat, D. Antaki, V. Bafna, J. Michaelson, K. Ye, S. E. Devine, E. J. Gardner, G. R. Abecasis, J. M. Kidd, R. E. Mills, G. Dayama, S. Emery, G. Jun, X. Shi, A. Quitadamo, G. Lunter, G. A. McVean, K. Chen, X. Fan, Z. Chong, T. Chen, D. Witherspoon, J. Xing, E. E. Eichler, M. J. Chaisson, F. Hormozdiari, J. Huddleston, M. Malig, B. J. Nelson, P. H. Sudmant, N. F. Parrish, E. Khurana, M. E. Hurles, B. Blackburne, S. J. Lindsay, Z. Ning, K. Walter, Y. Zhang, M. B. Gerstein, A. Abyzov, J. Chen, D. Clarke, H.

Lam, X. Jasmine Mu, C. Sisu, J. Zhang, Y. Zhang, R. A. Gibbs, F. Yu, M. Bainbridge, D. Challis, U. S. Evani, C. Kovar, J. Lu, D. Muzny, U. Nagaswamy, J. G. Reid, A. Sabo, J. Yu, X. Guo, W. Li, Y. Li, R. Wu, G. T. Marth, E. P. Garrison, W. Fung Leong, A. N. Ward, G. del Angel, M. A. DePristo, S. B. Gabriel, N. Gupta, C. Hartl, R. E. Poplin, A. G. Clark, J. L. Rodriguez-Flores, P. Flicek, L. Clarke, R. E. Smith, X. Zheng-Bradley, D. G. MacArthur, E. R. Mardis, R. Fulton, D. C. Koboldt, S. Gravel, C. D. Bustamante, D. W. Craig, A. Christoforides, N. Homer, T. Izatt, S. T. Sherry, C. Xiao, E. T. Dermitzakis, G. R. Abecasis, H. Min Kang, G. A. McVean, M. B. Gerstein, S. Balasubramanian, L. Habegger, H. Yu, P. Flicek, L. Clarke, F. Cunningham, I. Dunham, D. Zerbino, X. Zheng-Bradley, K. Lage, J. Berg Jespersen, H. Horn, S. B. Montgomery, M. K. DeGorter, E. Khurana, C. Tyler-Smith, Y. Chen, V. Colonna, Y. Xue, M. B. Gerstein, S. Balasubramanian, Y. Fu, D. Kim, A. Auton, A. Marcketta, R. Desalle, A. Narechania, M. A. Wilson Sayres, E. P. Garrison, R. E. Handsaker, S. Kashin, S. A. McCarroll, J. L. Rodriguez-Flores, P. Flicek, L. Clarke, X. Zheng-Bradley, Y. Erlich, M. Gymrek, T. Frederick Willems, C. D. Bustamante, F. L. Mendez, G. David Poznik, P. A. Underhill, C. Lee, E. Cerveira, A. Malhotra, M. Romanovitch, C. Zhang, G. R. Abecasis, L. Coin, H. Shao, D. Mittelman, C. Tyler-Smith, Q. Ayub, R. Banerjee, M. Cerezo, Y. Chen, T. W. Fitzgerald, S. Louzada, A. Massaia, S. McCarthy, G. R. Ritchie, Y. Xue, F. Yang, R. A. Gibbs, C. Kovar, D. Kalra, W. Hale, D. Muzny, J. G. Reid, J. Wang, X. Dan, X. Guo, G. Li, Y. Li, C. Ye, X. Zheng, D. M. Altshuler, P. Flicek, L. Clarke, X. Zheng-Bradley, D. R. Bentley, A. Cox, S. Humphray, S. Kahn, R. Sudbrak, M. W. Albrecht, M. Lienhard, D. Larson, D. W. Craig, T. Izatt, A. A. Kurdoglu, S. T. Sherry, C. Xiao, D. Haussler, G. R. Abecasis, G. A. McVean, R. M. Durbin, S. Balasubramaniam, T. M. Keane, S. McCarthy, J. Stalker, W. Bodmer, G. Bedoya, A. Ruiz-Linares, Z. Cai, Y. Gao, J. Chu, L. Peltonen, A. Garcia-Montero, A. Orfao, J. Dutil, J. C. Martinez-Cruzado, T. K. Oleksyk, K. C. Barnes, R. A. Mathias, A. Hennis, H. Watson, C. McKenzie, F. Qadri, R. LaRocque, P. C. Sabeti, J. Zhu, X. Deng, P. C. Sabeti, D. Asogun, O. Folarin, C. Happi, O. Omoniwa, M. Stremlau, R. Tariyal, M. Jallow, F. Sisay Joof, T. Corrah, K. Rockett, D. Kwiatkowski, J. Kooner, T. Tịnh Hiê`n, S. J. Dunstan, N. Thuy Hang, R. Fonnie, R. Garry, L. Kanneh, L. Moses, P. C. Sabeti, J. Schieffelin, D. S. Grant, C. Gallo, G. Poletti, D. Saleheen, A. Rasheed, A global reference for human genetic variation. Nature. 526, 68–74 (2015).

66. H. Li, Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics. 27, 718–9 (2011).

67. J. Prado-Martinez, P. H. Sudmant, J. M. Kidd, H. Li, J. L. Kelley, B. Lorente-Galdos, K. R. Veeramah, A. E. Woerner, T. D. O'connor, G. Santpere, others, Great ape genetic diversity and population history. Nature. 499, 471 (2013).

68. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27, 2987–93 (2011).

69. A. Y. Chang, V. F. Skirbekk, S. Tyrovolas, N. J. Kassebaum, J. L. Dieleman, Measuring population ageing: an analysis of the Global Burden of Disease Study 2017. Lancet Public Heal. 4, e159–e167 (2019).

70. C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, R. Collins, UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Med. 12, e1001779 (2015).

71. B. Bulik-Sullivan, H. K. Finucane, V. Anttila, A. Gusev, F. R. Day, P.-R. Loh, L. Duncan, J. R. B. Perry, N. Patterson, E. B. Robinson, M. J. Daly, A. L. Price, B. M. Neale, M. J. Daly, A. L. Price, B. M. Neale, An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015).

72. T. Beck, T. Shorter, A. J. Brookes, GWAS Central: A comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies. Nucleic Acids Res. 48, D933–D940 (2020).

73. A. Buniello, J. A. L. Macarthur, M. Cerezo, L. W. Harris, J. Hayhurst, C. Malangone, A. McMahon, J. Morales, E. Mountjoy, E. Sollis, D. Suveges, O. Vrousgou, P. L. Whetzel, R. Amode, J. A. Guillen, H. S. Riat, S. J. Trevanion, P. Hall, H. Junkins, P. Flicek, T. Burdett, L. A. Hindorff, F. Cunningham, H. Parkinson, The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 47, D1005–D1012 (2019).

74. K. Watanabe, S. Stringer, O. Frei, M. Umićević Mirkov, C. de Leeuw, T. J. C. Polderman, S. van der Sluis, O. A. Andreassen, B. M. Neale, D. Posthuma, A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. 51, 1339–1348 (2019).

75. N. M. Warrington, R. N. Beaumont, M. Horikoshi, F. R. Day, Ø. Helgeland, C. Laurin, J. Bacelis, S. Peng, K. Hao, B. Feenstra, A. R. Wood, A. Mahajan, J. Tyrrell, N. R. Robertson, N. W. Rayner, Z. Qiao, G. H. Moen, M. Vaudel, C. J. Marsit, J. Chen, M. Nodzenski, T. M. Schnurr, M. H. Zafarmand, J. P. Bradfield, N. Grarup, M. N. Kooijman, R. Li-Gao, F. Geller, T. S. Ahluwalia, L. Paternoster, R. Rueedi, V. Huikari, J. J. Hottenga, L. P. Lyytikäinen, A. Cavadino, S. Metrustry, D. L. Cousminer, Y. Wu, E. Thiering, C. A. Wang, C. T. Have, N. Vilor-Tejedor, P. K. Joshi, J. N. Painter, I. Ntalla, R. Myhre, N. Pitkänen, E. M. van Leeuwen, R. Joro, V. Lagou, R. C. Richmond, A. Espinosa, S. J. Barton, H. M. Inskip, J. W. Holloway, L. Santa-Marina, X. Estivill, W. Ang, J. A. Marsh, C. Reichetzeder, L. Marullo, B. Hocher, K. L. Lunetta, J. M. Murabito, C. L. Relton, M. Kogevinas, L. Chatzi, C. Allard, L. Bouchard, M. F. Hivert, G. Zhang, L. J. Muglia, J. Heikkinen, C. S. Morgen, A. H. C. van Kampen, B. D. C. van Schaik, F. D. Mentch, C. Langenberg, J. Luan, R. A. Scott, J. H. Zhao, G. Hemani, S. M. Ring, A. J. Bennett, K. J. Gaulton, J. Fernandez-Tajes, N. R. van Zuydam, C. Medina-Gomez, H. G. de Haan, F. R. Rosendaal, Z. Kutalik, P. Marques-Vidal, S. Das, G. Willemsen, H. Mbarek, M. Müller-Nurasyid, M. Standl, E. V. R. Appel, C. E. Fonvig, C. Trier, C. E. M. van Beijsterveldt, M. Murcia, M. Bustamante, S. Bonas-Guarch, D. M. Hougaard, J. M. Mercader, A. Linneberg, K. E. Schraut, P. A. Lind, S. E. Medland, B. M. Shields, B. A. Knight, J. F. Chai, K. Panoutsopoulou, M. Bartels, F. Sánchez, J. Stokholm, D. Torrents, R. K. Vinding, S. M. Willems, M. Atalay, B. L. Chawes, P. Kovacs, I. Prokopenko, M. A. Tuke, H. Yaghootkar, K. S. Ruth, S. E. Jones, P. R. Loh, A. Murray, M. N. Weedon, A. Tönjes, M. Stumvoll, K. F. Michaelsen, A. M. Eloranta, T. A. Lakka, C. M. van Duijn, W. Kiess, A. Körner, H. Niinikoski, K. Pahkala, O. T. Raitakari, B. Jacobsson, E. Zeggini, G. V. Dedoussis, Y. Y. Teo, S. M. Saw, G. W. Montgomery, H. Campbell, J. F. Wilson, T. G. M. Vrijkotte, M. Vrijheid, E. J. C. N. de Geus, M. G. Hayes, H. N. Kadarmideen, J. C. Holm, L. J. Beilin, C. E. Pennell, J. Heinrich, L. S. Adair, J. B. Borja, K. L. Mohlke, J. G. Eriksson, E. E. Widén, A. T. Hattersley, T. D. Spector, M. Kähönen, J. S. Viikari, T. Lehtimäki, D. I. Boomsma, S. Sebert, P. Vollenweider, T. I. A. Sørensen, H. Bisgaard, K. Bønnelykke, J. C. Murray, M. Melbye, E. A. Nohr, D. O. Mook-Kanamori, F. Rivadeneira, A. Hofman, J. F. Felix, V. W. V. Jaddoe, T. Hansen, C. Pisinger, A. A. Vaag, O. Pedersen, A. G. Uitterlinden, M. R. Järvelin, C. Power, E. Hyppönen, D. M. Scholtens, W. L. Lowe, G. Davey Smith, N. J. Timpson, A. P. Morris, N. J. Wareham, H. Hakonarson, S. F. A. Grant, T. M. Frayling, D. A. Lawlor, P. R. Njølstad, S. Johansson, K. K. Ong, M. I. McCarthy, J. R. B. Perry, D. M.

Evans, R. M. Freathy, Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. Nat. Genet. 51, 804–814 (2019).

76. J. P. Bradfield, S. Vogelezang, J. F. Felix, A. Chesi, Ø. Helgeland, M. Horikoshi, V. Karhunen, E. Lowry, D. L. Cousminer, T. S. Ahluwalia, E. Thiering, E. T. H. Boh, M. H. Zafarmand, N. Vilor-Tejedor, C. A. Wang, R. Joro, Z. Chen, W. J. Gauderman, N. Pitkänen, E. J. Parra, L. Fernandez-Rhodes, A. Alyass, C. Monnereau, J. A. Curtin, C. T. Have, S. E. McCormack, M. Hollensted, C. Frithioff-Bøjsøe, A. Valladares-Salgado, J. Peralta-Romero, Y. Y. Teo, M. Standl, J. T. Leinonen, J. C. Holm, T. Peters, J. Vioque, M. Vrijheid, A. Simpson, A. Custovic, M. Vaudel, M. Canouil, V. Lindi, M. Atalay, M. Kähönen, O. T. Raitakari, B. D. C. Van Schaik, R. I. Berkowitz, S. A. Cole, V. S. Voruganti, Y. Wang, H. M. Highland, A. G. Comuzzie, N. F. Butte, A. E. Justice, S. Gahagan, E. Blanco, T. Lehtimäki, T. A. Lakka, J. Hebebrand, A. Bonnefond, N. Grarup, P. Froguel, L. P. Lyytikäinen, M. Cruz, S. Kobes, R. L. Hanson, B. S. Zemel, A. Hinney, K. K. Teo, D. Meyre, K. E. North, F. D. Gilliland, H. Bisgaard, M. Bustamante, K. Bonnelykke, C. E. Pennell, F. Rivadeneira, A. G. Uitterlinden, L. J. Baier, T. G. M. Vrijkotte, J. Heinrich, T. I. A. Sørensen, S. M. Saw, O. Pedersen, T. Hansen, J. Eriksson, E. Widén, M. I. McCarthy, P. R. Njølstad, C. Power, E. Hyppönen, S. Sebert, C. D. Brown, M. R. Järvelin, N. J. Timpson, S. Johansson, H. Hakonarson, V. W. V. Jaddoe, F. A. Struan, A trans-ancestral meta-analysis of genome-wide association studies reveals loci associated with childhood obesity. Hum. Mol. Genet. 28, 3327–3338 (2019).

77. D. L. Cousminer, D. J. Berry, N. J. Timpson, W. Ang, E. Thiering, E. M. Byrne, H. Rob Taal, V. Huikari, J. P. Bradfield, M. Kerkhof, M. M. Groen-Blokhuis, E. Kreiner-Møller, M. Marinelli, C. Holst, J. T. Leinonen, J. R. B. Perry, I. Surakka, O. Pietiläinen, J. Kettunen, V. Anttila, M. Kaakinen, U. Sovio, A. Pouta, S. Das, V. Lagou, C. Power, I. Prokopenko, D. M. Evans, J. P. Kemp, B. St Pourcain, S. Ring, A. Palotie, E. Kajantie, C. Osmond, T. Lehtimäki, J. S. Viikari, M. Kähönen, N. M. Warrington, S. J. Lye, L. J. Palmer, C. M. T. Tiesler, C. Flexeder, G. W. Montgomery, S. E. Medland, A. Hofman, H. Hakonarson, M. Guxens, M. Bartels, V. Salomaa, J. M. Murabito, J. Kaprio, T. I. A. Sørensen, F. Ballester, H. Bisgaard, D. I. Boomsma, G. H. Koppelman, S. F. A. Grant, V. W. V. Jaddoe, N. G. Martin, J. Heinrich, C. E. Pennell, O. T. Raitakari, J. G. Eriksson, G. D. Smith, E. Hyppönen, M. R. Järvelin, M. I. McCarthy, S. Ripatti, E.

Widén, Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. Hum. Mol. Genet. 22, 2735–2747 (2013).

78. X. Liu, D. Helenius, L. Skotte, R. N. Beaumont, M. Wielscher, F. Geller, J. Juodakis, A. Mahajan, J. P. Bradfield, F. T. J. Lin, S. Vogelezang, M. Bustamante, T. S. Ahluwalia, N. Pitkänen, C. A. Wang, J. Bacelis, M. C. Borges, G. Zhang, B. A. Bedell, R. M. Rossi, K. Skogstrand, S. Peng, W. K. Thompson, V. Appadurai, D. A. Lawlor, I. Kalliala, C. Power, M. I. McCarthy, H. A. Boyd, M. L. Marazita, H. Hakonarson, M. G. Hayes, D. M. Scholtens, F. Rivadeneira, V. W. V. Jaddoe, R. K. Vinding, H. Bisgaard, B. A. Knight, K. Pahkala, O. Raitakari, Ø. Helgeland, S. Johansson, P. R. Njølstad, J. Fadista, A. J. Schork, R. Nudel, D. E. Miller, X. Chen, M. T. Weirauch, P. B. Mortensen, A. D. Børglum, M. Nordentoft, O. Mors, K. Hao, K. K. Ryckman, D. M. Hougaard, L. C. Kottyan, C. E. Pennell, L. P. Lyytikainen, K. Bønnelykke, M. Vrijheid, J. F. Felix, W. L. Lowe, S. F. A. Grant, E. Hyppönen, B. Jacobsson, M. R. Jarvelin, L. J. Muglia, J. C. Murray, R. M. Freathy, T. M. Werge, M. Melbye, A. Buil, B. Feenstra, Variants in the fetal genome near pro-inflammatory cytokine genes on 2q13 associate with gestational duration. Nat. Commun. 10, 1–13 (2019).

79. R. J. P. van der Valk, E. Kreiner-Møller, M. N. Kooijman, M. Guxens, E. Stergiakouli, A. Sääf, J. P. Bradfield, F. Geller, M. Geoffrey Hayes, D. L. Cousminer, A. Körner, E. Thiering, J. A. Curtin, R. Myhre, V. Huikari, R. Joro, M. Kerkhof, N. M. Warrington, N. Pitkänen, I. Ntalla, M. Horikoshi, R. Veijola, R. M. Freathy, Y. Y. Teo, S. J. Barton, D. M. Evans, J. P. Kemp, B. S. Pourcain, S. M. Ring, G. D. Smith, A. Bergström, I. Kull, H. Hakonarson, F. D. Mentch, H. Bisgaard, B. Chawes, J. Stokholm, J. Waage, P. Eriksen, A. Sevelsted, M. Melbye, C. M. van Duijn, C. Medina-Gomez, A. Hofman, J. C. de Jongste, H. Rob Taal, A. G. Uitterlinden, L. L. Armstrong, J. Eriksson, A. Palotie, M. Bustamante, X. Estivill, J. R. Gonzalez, S. Llop, W. Kiess, A. Mahajan, C. Flexeder, C. M. T. Tiesler, C. S. Murray, A. Simpson, P. Magnus, V. Sengpiel, A. L. Hartikainen, S. Keinanen-Kiukaanniemi, A. Lewin, A. Da Silva Couto Alves, A. I. Blakemore, J. L. Buxton, M. Kaakinen, A. Rodriguez, S. Sebert, M. Vaarasmaki, T. Lakka, V. Lindi, U. Gehring, D. S. Postma, W. Ang, J. P. Newnham, L. P. Lyytikäinen, K. Pahkala, O. T. Raitakari, K. Panoutsopoulou, E. Zeggini, D. I. Boomsma, M. Groen-Blokhuis, J. Ilonen, L. Franke, J. N. Hirschhorn, T. H. Pers, L. Liang, J. Huang, B. Hocher, M. Knip, S. M. Saw, J. W. Holloway, E. Melén, S. F. A. Grant, B. Feenstra, W. L. Lowe, E. Widén, E. Sergeyev, H. Grallert, A. Custovic, B. Jacobsson, M. R. Jarvelin, M.

Atalay, G. H. Koppelman, C. E. Pennell, H. Niinikoski, G. V. Dedoussis, M. I. Mccarthy, T. M. Frayling, J. Sunyer, N. J. Timpson, F. Rivadeneira, K. Bønnelykke, V. W. V. Jaddoe, A novel common variant in DCST2 is associated with length in early life and height in adulthood. Hum. Mol. Genet. 24, 1155–1168 (2015).

80. J. Fadista, L. Skotte, F. Geller, J. Bybjerg-Grauholm, S. Gørtz, P. A. Romitti, M. Caggana, D. M. Kay, H. Matsson, H. A. Boyd, D. M. Hougaard, A. Nordenskjöld, J. L. Mills, M. Melbye, B. Feenstra, Genome-wide meta-Analysis identifies BARX1 and EML4-MTA3 as new loci associated with infantile hypertrophic pyloric stenosis. Hum. Mol. Genet. 28, 332–340 (2019).

81. T.I.L.A.E., Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat. Commun. 9, 5269 (2018).

82. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G. L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods. 17, 261–272 (2020).

83. A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. D. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, D.

Haussler, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15, 1034–1050 (2005).

84. M. D. Rasmussen, M. J. Hubisz, I. Gronau, A. Siepel, Genome-Wide Inference of Ancestral Recombination Graphs. PLoS Genet. 10, e1004342 (2014).

85. Y. F. Huang, B. Gulko, A. Siepel, Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. Nat. Genet. 49, 618–624 (2017).

86. N. Harmston, E. Ing-Simmons, M. Perry, A. Barešić, B. Lenhard, GenomicInteractions: An R/Bioconductor package for manipulating and investigating chromatin interaction data. BMC Genomics. 16, 963 (2015).

References:

Aagaard-Tillery, K.M., Grove, K., Bishop, J., Ke, X., Fu, Q., McKnight, R., Lane, R.H., 2008. Developmental origins of disease and determinants of chromatin structure: Maternal diet modifies the primate fetal epigenome. Journal of Molecular Endocrinology 41, 91–102. https://doi.org/10.1677/JME-08-0025

Aakre, C.D., Herrou, J., Phung, T.N., Perchuk, B.S., Crosson, S., Laub, M.T., 2015. Evolving new protein-protein interaction specificity through promiscuous intermediates. Cell 163, 594. https://doi.org/10.1016/J.CELL.2015.09.055

Abascal, F., Acosta, R., Addleman, N.J., Adrian, J., Afzal, V., Aken, B., Akiyama, J.A., Jammal, O. al, Amrhein, H., Anderson, S.M., Andrews, G.R., Antoshechkin, I., Ardlie, K.G., Armstrong, J., Astley, M., Banerjee, B., Barkal, A.A., Barnes, I.H.A., Barozzi, I., Barrell, D., Barson, G., Bates, D., Baymuradov, U.K., Bazile, C., Beer, M.A., Beik, S., Bender, M.A., Bennett, R., Bouvrette, L.P.B., Bernstein, B.E., Berry, A., Bhaskar, A., Bignell, A., Blue, S.M., Bodine, D.M., Boix, C., Boley, N., Borrman, T., Borsari,

B., Boyle, A.P., Brandsmeier, L.A., Breschi, A., Bresnick, E.H., Brooks, J.A., Buckley, M., Burge, C.B., Byron, R., Cahill, E., Cai, L., Cao, L., Carty, M., Castanon, R.G., Castillo, A., Chaib, H., Chan, E.T., Chee, D.R., Chee, S., Chen, Hao, Chen, Huaming, Chen, J.Y., Chen, S., Cherry, J.M., Chhetri, S.B., Choudhary, J.S., Chrast, J., Chung, D., Clarke, D., Cody, N.A.L., Coppola, C.J., Coursen, J., D'Ippolito, A.M., Dalton, S., Danyko, C., Davidson, C., Davila-Velderrain, J., Davis, C.A., Dekker, J., Deran, A., DeSalvo, G., Despacio-Reyes, G., Dewey, C.N., Dickel, D.E., Diegel, M., Diekhans, M., Dileep, V., Ding, B., Djebali, S., Dobin, A., Dominguez, D., Donaldson, S., Drenkow, J., Dreszer, T.R., Drier, Y., Duff, M.O., Dunn, D., Eastman, C., Ecker, J.R., Edwards, M.D., El-Ali, N., Elhajjajy, S.I., Elkins, K., Emili, A., Epstein, C.B., Evans, R.C., Ezkurdia, I., Fan, K., Farnham, P.J., Farrell, N.P., Feingold, E.A., Ferreira, A.M., Fisher-Aylor, K., Fitzgerald, S., Flicek, P., Foo, C.S., Fortier, K., Frankish, A., Freese, P., Fu, S., Fu, X.D., Fu, Y., Fukuda-Yuzawa, Y., Fulciniti, M., Funnell, A.P.W., Gabdank, I., Galeev, T., Gao, M., Giron, C.G., Garvin, T.H., Gelboin-Burkhart, C.A., Georgolopoulos, G., Gerstein, M.B., Giardine, B.M., Gifford, D.K., Gilbert, D.M., Gilchrist, D.A., Gillespie, S., Gingeras, T.R., Gong, P., Gonzalez, A., Gonzalez, J.M., Good, P., Goren, A., Gorkin, D.U., Graveley, B.R., Gray, M., Greenblatt, J.F., Griffiths, E., Groudine, M.T., Grubert, F., Gu, M., Guigó, R., Guo, H., Guo, Yu, Guo, Yuchun, Gursoy, G., Gutierrez-Arcelus, M., Halow, J., Hardison, R.C., Hardy, M., Hariharan, M., Harmanci, A., Harrington, A., Harrow, J.L., Hashimoto, T.B., Hasz, R.D., Hatan, M., Haugen, E., Hayes, J.E., He, P., He, Y., Heidari, N., Hendrickson, D., Heuston, E.F., Hilton, J.A., Hitz, B.C., Hochman, A., Holgren, C., Hou, L., Hou, S., Hsiao, Y.H.E., Hsu, S., Huang, H., Hubbard, T.J., Huey, J., Hughes, T.R., Hunt, T., Ibarrientos, S., Issner, R., Iwata, M., Izuogu, O., Jaakkola, T., Jameel, N., Jansen, C., Jiang, L., Jiang, P., Johnson, A., Johnson, R., Jungreis, I., Kadaba, M., Kasowski, M., Kasparian, M., Kato, M., Kaul, R., Kawli, T., Kay, M., Keen, J.C., Keles, S., Keller, C.A., Kelley, D., Kellis, M., Kheradpour, P., Kim, D.S., Kirilusha, A., Klein, R.J., Knoechel, B., Kuan, S., Kulik, M.J., Kumar, S., Kundaje, A., Kutyavin, T., Lagarde, J., Lajoie, B.R., Lambert, N.J., Lazar, J., Lee, A.Y., Lee, D., Lee, E., Lee, J.W., Lee, K., Leslie, C.S., Levy, S., Li, B., Li, H., Li, N., Li, X., Li, Y.I., Li, Ying, Li, Yining, Li, Yue, Lian, J., Libbrecht, M.W., Lin, S., Lin, Y., Liu, D., Liu, J., Liu, P., Liu, T., Liu, X.S., Liu, Yan, Liu, Yaping, Long, M., Lou, S., Loveland, J., Lu, A., Lu, Y., Lécuyer, E., Ma, L., Mackiewicz, M., Mannion, B.J., Mannstadt, M., Manthravadi, D., Marinov, G.K., Martin, F.J., Mattei, E., McCue, K., McEown, M., McVicker, G.,

Meadows, S.K., Meissner, A., Mendenhall, E.M., Messer, C.L., Meuleman, W., Meyer, C., Miller, S., Milton, M.G., Mishra, T., Moore, D.E., Moore, H.M., Moore, J.E., Moore, S.H., Moran, J., Mortazavi, A., Mudge, J.M., Munshi, N., Murad, R., Myers, R.M., Nandakumar, V., Nandi, P., Narasimha, A.M., Narayanan, A.K., Naughton, H., Navarro, F.C.P., Navas, P., Nazarovs, J., Nelson, J., Neph, S., Neri, F.J., Nery, J.R., Nesmith, A.R., Newberry, J.S., Newberry, K.M., Ngo, V., Nguyen, R., Nguyen, T.B., Nguyen, T., Nishida, A., Noble, W.S., Novak, C.S., Novoa, E.M., Nuñez, B., O'Donnell, C.W., Olson, S., Onate, K.C., Otterman, E., Ozadam, H., Pagan, M., Palden, T., Pan, X., Park, Y., Partridge, E.C., Paten, B., Pauli-Behn, F., Pazin, M.J., Pei, B., Pennacchio, L.A., Perez, A.R., Perry, E.H., Pervouchine, D.D., Phalke, N.N., Pham, Q., Phanstiel, D.H., Plajzer-Frick, I., Pratt, G.A., Pratt, H.E., Preissl, S., Pritchard, J.K., Pritykin, Y., Purcaro, M.J., Qin, Q., Quinones-Valdez, G., Rabano, I., Radovani, E., Raj, A., Rajagopal, N., Ram, O., Ramirez, L., Ramirez, R.N., Rausch, D., Raychaudhuri, S., Raymond, J., Razavi, R., Reddy, T.E., Reimonn, T.M., Ren, B., Reymond, A., Reynolds, A., Rhie, S.K., Rinn, J., Rivera, M., Rivera-Mulia, J.C., Roberts, B.S., Rodriguez, J.M., Rozowsky, J., Ryan, R., Rynes, E., Salins, D.N., Sandstrom, R., Sasaki, T., Sathe, S., Savic, D., Scavelli, A., Scheiman, J., Schlaffner, C., Schloss, J.A., Schmitges, F.W., See, L.H., Sethi, A., Setty, M., Shafer, A., Shan, S., Sharon, E., Shen, Q., Shen, Y., Sherwood, R.I., Shi, M., Shin, S., Shoresh, N., Siebenthall, K., Sisu, C., Slifer, T., Sloan, C.A., Smith, A., Snetkova, V., Snyder, M.P., Spacek, D. v., Srinivasan, S., Srivas, R., Stamatoyannopoulos, G., Stamatoyannopoulos, J.A., Stanton, R., Steffan, D., Stehling-Sun, S., Strattan, J.S., Su, A., Sundararaman, B., Suner, M.M., Syed, T., Szynkarek, M., Tanaka, F.Y., Tenen, D., Teng, M., Thomas, J.A., Toffey, D., Tress, M.L., Trout, D.E., Trynka, G., Tsuji, J., Upchurch, S.A., Ursu, O., Uszczynska-Ratajczak, B., Uziel, M.C., Valencia, A., Biber, B. van, van der Velde, A.G., van Nostrand, E.L., Vaydylevich, Y., Vazquez, J., Victorsen, A., Vielmetter, J., Vierstra, J., Visel, A., Vlasova, A., Vockley, C.M., Volpi, S., Vong, S., Wang, H., Wang, M., Wang, Q., Wang, R., Wang, T., Wang, W., Wang, X., Wang, Y., Watson, N.K., Wei, X., Wei, Z., Weisser, H., Weissman, S.M., Welch, R., Welikson, R.E., Weng, Z., Westra, H.J., Whitaker, J.W., White, C., White, K.P., Wildberg, A., Williams, B.A., Wine, D., Witt, H.N., Wold, B., Wolf, M., Wright, J., Xiao, R., Xiao, X., Xu, J., Yan, K.K., Yan, Y., Yang, H., Yang, X., Yang, Y.W., Yardimci, G.G., Yee, B.A., Yeo, G.W., Young, T., Yu, T., Yue, F., Zaleski, C., Zang, C., Zeng, H., Zeng, W., Zerbino, D.R., Zhai, J., Zhan, L., Zhan, Y., Zhang, B.,

Zhang, Jialing, Zhang, Jing, Zhang, K., Zhang, L., Zhang, P., Zhang, Q., Zhang, X.O., Zhang, Y., Zhang, Z., Zhao, Y., Zheng, Y., Zhong, G., Zhou, X.Q., Zhu, Y., Zimmerman, J., 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 2020 583:7818 583, 699–710. https://doi.org/10.1038/S41586-020-2493-4

Adams, J., Mansfield, M.J., Richard, D.J., Doxey, A.C., 2017. Lineage-specific mutational clustering in protein structures predicts evolutionary shifts in function. Bioinformatics 33, 1338–1345. https://doi.org/10.1093/BIOINFORMATICS/BTW815

Adzhubei, I., Jordan, D.M., Sunyaev, S.R., 2013. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.] 0 7, Unit7.20. https://doi.org/10.1002/0471142905.HG0720S76

Agundez, J.A.G., 2005. Cytochrome P450 Gene Polymorphism and Cancer. Current Drug Metabolism 5, 211–224. https://doi.org/10.2174/1389200043335621

Agustí, A., Noell, G., Brugada, J., Faner, R., 2017. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. The Lancet Respiratory Medicine 5, 935–945. https://doi.org/10.1016/S2213-2600(17)30434-4

Alberts, S.C., Altmann, J., Brockman, D.K., Cords, M., Fedigan, L.M., Pusey, A., Stoinski, T.S., Strier, K.B., Morris, W.F., Bronikowski, A.M., 2013. Reproductive aging patterns in primates reveal that humans are distinct. Proceedings of the National Academy of Sciences 110, 13440–13445. https://doi.org/10.1073/pnas.1311857110

Andersen, H. (1961). Histochemical studies on the histogenesis of the knee joint and superior tibio-fibular joint in human foetuses. *Acta Anat* 64, 279–303.

arcOGEN Consortium, arcOGEN Collaborators, Zeggini, E., Panoutsopoulou, K., Southam, L., Rayner, N.W., Day-Williams, A.G., Lopes, M.C., Boraska, V., Esko, T., et al. (2012). Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* (London, England) 380, 815–823.

Arendt, T., Stieler, J., Ueberham, U., 2017. Is sporadic Alzheimer's disease a developmental disorder? Journal of Neurochemistry 143, 396–408. https://doi.org/10.1111/JNC.14036

Arias-de la Torre, J., Ronaldson, A., Prina, M., Matcham, F., Pinto Pereira, S.M., Hatch, S.L., Armstrong, D., Pickles, A., Hotopf, M., Dregan, A., 2021. Depressive symptoms during early adulthood and the development of physical multimorbidity in the UK: an observational cohort study. The Lancet Healthy Longevity 2, e801–e810. https://doi.org/10.1016/S2666-7568(21)00259-2/ATTACHMENT/A96E6B23-B715-496F-B742-86C4D2F9F20A/MMC1.PDF

Aunan, J.R., Cho, W.C., Søreide, K., 2017. The Biology of Aging and Cancer: A Brief Overview of Shared and Divergent Molecular Hallmarks. Aging and Disease 8, 628. https://doi.org/10.14336/AD.2017.0103

Austad, S.N., 1997. Comparative aging and life histories in mammals. Experimental Gerontology 32, 23–38. https://doi.org/10.1016/S0531-5565(96)00059-9

Austad, S.N., Hoffman, J.M., 2018. Is antagonistic pleiotropy ubiquitous in aging biology? Evolution, Medicine, and Public Health 2018, 287–294. https://doi.org/10.1093/EMPH/EOY033

Balistreri, C.R., Caruso, C., Grimaldi, M.P., Listì, F., Vasto, S., Orlando, V., Campagna, A.M., Lio, D., Candore, G., 2007. CCR5 receptor: biologic and genetic implications in age-related diseases. Ann N Y Acad Sci 1100, 162–172. https://doi.org/10.1196/ANNALS.1395.014

Bao, J., Pan, G., Poncz, M., Wei, J., Ran, M., Zhou, Z., 2018. Serpin functions in host-pathogen interactions. PeerJ 2018, e4557. https://doi.org/10.7717/PEERJ.4557/FIG-3

Bao, Q., Shi, Y., 2006. Apoptosome: a platform for the activation of initiator caspases. Cell Death & Differentiation 2007 14:1 14, 56–65. https://doi.org/10.1038/sj.cdd.4402028

Barr, A.J., Dube, B., Hensor, E.M.A., Kingsbury, S.R., Peat, G., Bowes, M.A., Sharples, L.D., and Conaghan, P.G. (2016). The relationship between three-dimensional knee MRI bone shape and total knee replacement-a case control study: data from the Osteoarthritis Initiative. *Rheumatology* (Oxford). 55, 1585–1593.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265.

Basit, S., Naqvi, S.K.-H., Wasif, N., Ali, G., Ansar, M., and Ahmad, W. (2008). A novel insertion mutation in the cartilage-derived morphogenetic protein-1 (CDMP1) gene underlies Grebe-type chondrodysplasia in a consanguineous Pakistani family. *BMC Med. Genet.* 9, 102.

Baugé, C., Girard, N., Lhuissier, E., Bazille, C., and Boumediene, K. (2014). Regulation and Role of TGFβ Signaling Pathway in Aging and Osteoarthritis Joints. *Aging Dis.* 5, 394–405.

Belenguer-Varea, Á., Tarazona-Santabalbina, F.J., Avellana-Zaragoza, J.A., Martínez-Reig, M., Mas-Bargues, C., Inglés, M., 2020. Oxidative stress and exceptional human longevity: Systematic review. Free Radic Biol Med 149, 51–63. https://doi.org/10.1016/J.FREERADBIOMED.2019.09.019

Benayoun, B.A., Pollina, E.A., Brunet, A., 2015. Epigenetic regulation of ageing: Linking environmental inputs to genomic stability. Nature Reviews Molecular Cell Biology. https://doi.org/10.1038/nrm4048

Benayoun, B.A., Pollina, E.A., Singh, P.P., Mahmoudi, S., Harel, I., Casey, K.M., Dulken, B.W., Kundaje, A., Brunet, A., 2019. Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. Genome Research 29, 697–709. https://doi.org/10.1101/gr.240093.118

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 289–300.

Berenbaum, F., Wallace, I.J., Lieberman, D.E., and Felson, D.T. (2018). Modern-day environmental factors in the pathogenesis of osteoarthritis. *Nat. Rev. Rheumatol.* 14, 1.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L., Lander, E.S., 2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell 125, 315–326. https://doi.org/10.1016/J.CELL.2006.02.041/ATTACHMENT/FF7DE1A6-7652-4AB5-BFD8-5892EAEB6060/MMC6.XLS

Bi, S., Liu, Z., Wu, Z., Wang, Z., Liu, X., Wang, S., Ren, J., Yao, Y., Zhang, W., Song, M., Liu, G.H., Qu, J., 2020. SIRT7 antagonizes human stem cell aging as a heterochromatin stabilizer. Protein and Cell 11, 483–504. https://doi.org/10.1007/s13238-020-00728-4

Bird, C.P., Stranger, B.E., Liu, M., Thomas, D.J., Ingle, C.E., Beazley, C., Miller, W., Hurles, M.E., and Dermitzakis, E.T. (2007). Fast-evolving noncoding sequences in the human genome. *Genome Biol.* 8, R118.

BL, T., IB, Z., 1999. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev 63, 479–506. https://doi.org/10.1128/MMBR.63.2.479-506.1999

Blagosklonny, M. V., Hall, M.N., 2009. Growth and aging: a common molecular mechanism. Aging. https://doi.org/10.18632/aging.100040

Blake, J.A., Eppig, J.T., Kadin, J.A., Richardson, J.E., Smith, C.L., Bult, C.J., and Group, M.G.D. (2016). Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. *Nucleic Acids Res*. 45, D723--D729.

Blanpain, C., Doranz, B.J., Bondue, A., Govaerts, C., De Leener, A., Vassart, G., Doms, R.W., Proudfoot, A., Parmentier, M., 2003. The Core Domain of Chemokines Binds CCR5 Extracellular Domains while Their Amino Terminus Interacts with the Transmembrane Helix Bundle *. Journal of Biological Chemistry 278, 5179–5187. https://doi.org/10.1074/JBC.M205684200

Blekhman, R., Man, O., Herrmann, L., Boyko, A.R., Indap, A., Kosiol, C., Bustamante, C.D., Teshima, K.M., Przeworski, M., 2008. Natural Selection on Genes that Underlie Human Disease Susceptibility. Current Biology 18, 883–889. https://doi.org/10.1016/J.CUB.2008.04.074

Bogin, B., Smith, B.H., 1996. Evolution of the human life cycle. American Journal of Human Biology 8, 703–716. https://doi.org/10.1002/(SICI)1520-6300(1996)8:6<703::AID-AJHB2>3.0.CO;2-U

Booth, L.N., Brunet, A., 2016. The Aging Epigenome. Molecular Cell 62, 728–744. https://doi.org/10.1016/j.molcel.2016.05.013

Bou Sleiman, M., Jha, P., Houtkooper, R., Williams, R.W., Wang, X., Auwerx, J., 2020. The Gene-Regulatory Footprint of Aging Highlights Conserved Central Regulators. Cell Reports 32, 108203. https://doi.org/10.1016/j.celrep.2020.108203

Boyd, J.L., Skove, S.L., Rouanet, J.P., Pilaz, L.-J., Bepler, T., Gordân, R., Wray, G.A., and Silver, D.L. (2015). Human-Chimpanzee Differences in a FZD8 Enhancer Alter Cell-Cycle Dynamics in the Developing Neocortex. *Curr. Biol.* 25, 772–779.

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., Cherry, J.M., Snyder, M., 2012. Annotation of functional variation in personal genomes using RegulomeDB. Genome Research 22, 1790–1797. https://doi.org/10.1101/gr.137323.112

Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. https://doi.org/10.1016/j.cell.2017.05.038

Bramble, D.M., and Lieberman, D.E. (2004). Endurance running and the evolution of Homo. Nature 432, 345.

Brayer, K.J., Lynch, V.J., Wagner, G.P., 2011. Evolution of a derived protein-protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. Proc Natl Acad Sci U S A 108. https://doi.org/10.1073/PNAS.1100990108/ASSET/39A013F6-14CE-4DCE-84CF-63B8E6E253C1/ASSETS/GRAPHIC/PNAS.1100990S108FIG01.JPEG

Bromberg, Y., Rost, B., 2007. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Research 35, 3823–3835. https://doi.org/10.1093/NAR/GKM238

Bronikowski, A.M., Altmann, J., Brockman, D.K., Cords, M., Fedigan, L.M., Pusey, A., Stoinski, T., Morris, W.F., Strier, K.B., Alberts, S.C., 2011. Aging in the natural world: comparative data reveal similar mortality patterns across primates. Science 331, 1325–8. https://doi.org/10.1126/science.1201571

Brosh, R.M., Bohr, V.A., 2007. Human premature aging, DNA repair and RecQ helicases. Nucleic Acids Research 35, 7527. https://doi.org/10.1093/NAR/GKM1008

Browning, B.L., Zhou, Y., and Browning, S.R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet*. 103, 338–348.

Brunet, A., Berger, S.L., 2014. Epigenetics of Aging and Aging-related Disease. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 69, S17–S20. https://doi.org/10.1093/gerona/glu042

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol*. 109, 21–29.

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., Daly, M.J., Price, A.L., Neale, B.M., Daly, M.J., Price, A.L., Neale, B.M., 2015. An atlas of genetic correlations across human diseases and traits. Nature Genetics 47, 1236–1241. https://doi.org/10.1038/ng.3406

Bush, E.C., and Lahn, B.T. (2008). A genome-wide screen for noncoding elements important in primate evolution. *BMC Evol. Biol.* 8, 17.

Butterfield, N.C., Curry, K.F., Steinberg, J., Dewhurst, H., Komla-Ebri, D., Mannan, N.S., Adoum, A.-T., Leitch, V.D., Logan, J.G., Waung, J.A., et al. (2019). Accelerating functional gene discovery in osteoarthritis. *BioRxiv* 836221.

Byars, S.G., Huang, Q.Q., Gray, L.A., Bakshi, A., Ripatti, S., Abraham, G., Stearns, S.C., Inouye, M., 2017. Genetic loci associated with coronary artery disease harbor evidence of selection and antagonistic pleiotropy. PLOS Genetics 13, e1006328. https://doi.org/10.1371/JOURNAL.PGEN.1006328

Byars, S.G., Voskarides, K., 2020. Antagonistic Pleiotropy in Human Disease. Journal of Molecular Evolution 88, 12–25. https://doi.org/10.1007/S00239-019-09923-2/TABLES/2

Campbell, P.J., Getz, G., Korbel, J.O., Stuart, J.M., Jennings, J.L., Stein, L.D., Perry, M.D., Nahal-Bose, H.K., Ouellette, B.F.F., Li, C.H., Rheinbay, E., Nielsen, G.P., Sgroi, D.C., Wu, C.L., Faquin, W.C., Deshpande, V., Boutros, P.C., Lazar, A.J., Hoadley, K.A., Louis, D.N., Dursi, L. Jonathan, Yung, C.K., Bailey, M.H., Saksena, G., Raine, K.M., Buchhalter, I., Kleinheinz, K., Schlesner, M., Zhang, Junjun, Wang, W., Wheeler, D.A., Ding, L., Simpson, J.T., O'Connor, B.D., Yakneen, S., Ellrott, K., Miyoshi, N., Butler, A.P., Royo, R., Shorser, S.I., Vazquez, M., Rausch, T., Tiao, G., Waszak, S.M., Rodriguez-Martin, B., Shringarpure, S., Wu, D.Y., Demidov, G.M., Delaneau, O., Hayashi, S., Imoto, S., Habermann, N., Segre, A. V., Garrison, E., Cafferkey, A., Alvarez, E.G., Heredia-Genestar, J.M., Muyas, F., Drechsel, O., Bruzos, A.L., Temes, J., Zamora, J., Baez-Ortega, A., Kim, H.L., Mashl, R.J., Ye, K., DiBiase, A., Huang, K. lin, Letunic, I., McLellan, M.D., Newhouse, S.J., Shmaya, T., Kumar, S., Wedge, D.C., Wright, M.H., Yellapantula, V.D., Gerstein, M., Khurana, E., Marques-Bonet, T., Navarro, A., Bustamante, C.D., Siebert, R., Nakagawa, H., Easton, D.F., Ossowski, S., Tubio, J.M.C., De La Vega, F.M., Estivill, X., Yuen, D., Mihaiescu, G.L., Omberg, L., Ferretti, V., Sabarinathan, R., Pich, O., Gonzalez-Perez, A., Taylor-Weiner, A., Fittall, M.W., Demeulemeester, J., Tarabichi, M., Roberts, N.D., Van Loo, P., Cortés-Ciriano, I., Urban, L., Park, P., Zhu, B., Pitkänen, E., Li, Y., Saini, N., Klimczak, L.J., Weischenfeldt, J., Sidiropoulos, N., Alexandrov, L.B., Rabionet, R., Escaramis, G., Bosio, M., Holik, A.Z., Susak, H., Prasad, A., Erkek, S., Calabrese, C., Raeder, B., Harrington, E., Mayes, S.,

Turner, D., Juul, S., Roberts, S.A., Song, L., Koster, R., Mirabello, L., Hua, X., Tanskanen, T.J., Tojo, M., Chen, J., Aaltonen, L.A., Rätsch, G., Schwarz, R.F., Butte, A.J., Brazma, A., Chanock, S.J., Chatterjee, N., Stegle, O., Harismendy, O., Bova, G.S., Gordenin, D.A., Haan, D., Sieverling, L., Feuerbach, L., Chalmers, D., Joly, Y., Knoppers, B.M., Molnár-Gábor, F., Phillips, M., Thorogood, A., Townend, D., Goldman, M., Fonseca, N.A., Xiang, Q., Craft, B., Piñeiro-Yáñez, E., Muñoz, A., Petryszak, R., Füllgrabe, A., Al-Shahrour, F., Keays, M., Haussler, D., Weinstein, J., Huber, W., Valencia, A., Papatheodorou, I., Zhu, J., Fan, Y., Torrents, D., Bieg, M., Chen, K., Chong, Z., Cibulskis, K., Eils, R., Fulton, R.S., Gelpi, J.L., Gonzalez, S., Gut, I.G., Hach, F., Heinold, M., Hu, T., Huang, V., Hutter, B., Jäger, N., Jung, J., Kumar, Y., Lalansingh, C., Leshchiner, I., Livitz, D., Ma, E.Z., Maruvka, Y.E., Milovanovic, A., Nielsen, M.M., Paramasivam, N., Pedersen, J.S., Puiggròs, M., Sahinalp, S.C., Sarrafi, I., Stewart, C., Stobbe, M.D., Wala, J.A., Wang, Jiayin, Wendl, M., Werner, J., Wu, Z., Xue, H., Yamaguchi, T.N., Yellapantula, V., Davis-Dusenbery, B.N., Grossman, R.L., Kim, Y., Heinold, M.C., Hinton, J., Jones, D.R., Menzies, A., Stebbings, L., Hess, J.M., Rosenberg, M., Dunford, A.J., Gupta, M., Imielinski, M., Meyerson, M., Beroukhim, R., Reimand, J., Dhingra, P., Favero, F., Dentro, S., Wintersinger, J., Rudneva, V., Park, Ji Wan, Hong, E.P., Heo, S.G., Kahles, André, Lehmann, K. Van, Soulette, C.M., Shiraishi, Y., Liu, F., He, Y., Demircioğlu, D., Davidson, N.R., Greger, L., Li, Siliang, Liu, D., Stark, S.G., Zhang, F., Amin, S.B., Bailey, P., Chateigner, A., Frenkel-Morgenstern, M., Hou, Y., Huska, M.R., Kilpinen, H., Lamaze, F.C., Li, C., Li, Xiaobo, Li, Xinyue, Liu, X., Marin, M.G., Markowski, J., Nandi, T., Ojesina, A.I., Pan-Hammarström, Q., Park, P.J., Pedamallu, C.S., Su, H., Tan, P., Teh, B.T., Wang, Jian, Xiong, H., Ye, C., Yung, C., Zhang, Xiuqing, Zheng, L., Zhu, S., Awadalla, P., Creighton, C.J., Wu, K., Yang, H., Göke, J., Zhang, Z., Brooks, A.N., Fittall, M.W., Martincorena, I., Rubio-Perez, C., Juul, M., Schumacher, S., Shapira, O., Tamborero, D., Mularoni, L., Hornshøj, H., Deu-Pons, J., Muiños, F., Bertl, J., Guo, Q., Gonzalez-Perez, A., Xiang, Q., Bazant, W., Barrera, E., Al-Sedairy, S.T., Aretz, A., Bell, C., Betancourt, M., Buchholz, C., Calvo, F., Chomienne, C., Dunn, M., Edmonds, S., Green, E., Gupta, S., Hutter, C.M., Jegalian, K., Jones, N., Lu, Youyong, Nakagama, H., Nettekoven, G., Planko, L., Scott, D., Shibata, T., Shimizu, K., Stratton, M.R., Yugawa, T., Tortora, G., VijayRaghavan, K., Zenklusen, J.C., Townend, D., Aminou, B., Bartolome, J., Boroevich, K.A., Boyce, R., Buchanan, A., Byrne, N.J., Chen, Z., Cho, S., Choi, W., Clapham, P., Dow, M.T., Dursi, Lewis

Jonathan, Eils, J., Farcas, C., Fayzullaev, N., Flicek, P., Heath, A.P., Hofmann, O., Hong, J.H., Hudson, T.J., Hübschmann, D., Ivkovic, S., Jeon, S.H., Jiao, W., Kabbe, R., Kahles, Andre, Kerssemakers, J.N.A., Kim, H., Kim, Jihoon, Koscher, M., Koures, A., Kovacevic, M., Lawerenz, C., Liu, J., Mijalkovic, S., Mijalkovic-Lazic, A.M., Miyano, S., Nastic, M., Nicholson, J., Ocana, D., Ohi, K., Ohno-Machado, L., Pihl, T.D., Prinz, M., Radovic, P., Short, C., Sofia, H.J., Spring, J., Struck, A.J., Tijanic, N., Vicente, D., Wang, Z., Williams, A., Woo, Y., Wright, A.J., Yang, Liming, Hamilton, M.P., Johnson, T.A., Kahraman, A., Kellis, M., Polak, P., Sallari, R., Sinnott-Armstrong, N., von Mering, C., Beltran, S., Gerhard, D.S., Gut, M., Trotta, J.R., Whalley, J.P., Niu, B., Espiritu, S.M.G., Gao, S., Huang, Y., Lalansingh, C.M., Teague, J.W., Wendl, M.C., Abascal, F., Bader, G.D., Bandopadhayay, P., Barenboim, J., Brunak, S., Carlevaro-Fita, J., Chakravarty, D., Chan, C.W.Y., Choi, J.K., Diamanti, K., Fink, J.L., Frigola, J., Gambacorti-Passerini, C., Garsed, D.W., Haradhvala, N.J., Harmanci, A.O., Helmy, M., Herrmann, C., Hobolth, A., Hodzic, E., Hong, C., Isaev, K., Izarzugaza, J.M.G., Johnson, R., Juul, R.I., Kim, Jaegil, Kim, J.K., Jan Komorowski, Lanzós, A., Larsson, E., Lee, Donghoon, Li, Shantao, Li, Xiaotong, Lin, Z., Liu, E.M., Lochovsky, L., Lou, S., Madsen, T., Marchal, K., Martinez-Fundichely, A., McGillivray, P.D., Meyerson, W., Paczkowska, M., Park, Keunchil, Park, Kiejung, Pons, T., Pulido-Tamayo, S., Reyes-Salazar, I., Reyna, M.A., Rubin, M.A., Salichos, L., Sander, C., Schumacher, S.E., Shackleton, M., Shen, C., Shrestha, R., Shuai, S., Tsunoda, T., Umer, H.M., Uusküla-Reimand, L., Verbeke, L.P.C., Wadelius, C., Wadi, L., Warrell, J., Wu, G., Yu, J., Zhang, Jing, Zhang, Xuanping, Zhang, Y., Zhao, Z., Zou, L., Lawrence, M.S., Raphael, B.J., Bailey, P.J., Craft, D., Goldman, M.J., Aburatani, H., Binder, H., Dinh, H.Q., Heath, S.C., Hoffmann, S., Imbusch, C.D., Kretzmer, H., Laird, P.W., Martin-Subero, J.I., Nagae, G., Shen, H., Wang, Q., Weichenhan, D., Zhou, W., Berman, B.P., Brors, B., Plass, C., Akdemir, K.C., Bowtell, D.D.L., Burns, K.H., Busanovich, J., Chan, K., Dueso-Barroso, A., Edwards, P.A., Etemadmoghadam, D., Haber, J.E., Jones, D.T.W., Ju, Y.S., Kazanov, M.D., Koh, Y., Kumar, K., Lee, E.A., Lee, J.J.K., Lynch, A.G., Macintyre, G., Markowetz, F., Navarro, F.C.P., Pearson, J. V., Rippe, K., Scully, R., Villasante, I., Waddell, N., Yang, Lixing, Yao, X., Yoon, S.S., Zhang, C.Z., Bergstrom, E.N., Boot, A., Covington, K., Fujimoto, A., Huang, M.N., Islam, S.M.A., McPherson, J.R., Morganella, S., Mustonen, V., Ng, A.W.T., Prokopec, S.D., Vázquez-García, I., Wu, Y., Yousif, F., Yu, W., Rozen, S.G., Rudneva, V.A., Shringarpure, S.S.,

Turner, D.J., Xia, T., Atwal, G., Chang, D.K., Cooke, S.L., Faltas, B.M., Haider, S., Kaiser, V.B., Karlić, R., Kato, M., Kübler, K., Margolin, A., Martin, S., Nik-Zainal, S., P'ng, C., Semple, C.A., Smith, J., Sun, R.X., Thai, K., Wright, D.W., Yuan, K., Biankin, A. V., Garraway, L., Grimmond, S.M., Adams, D.J., Anur, P., Cao, S., Christie, E.L., Cmero, M., Cun, Y., Dawson, K.J., Dentro, S.C., Deshwar, A.G., Donmez, N., Drews, R.M., Gerstung, M., Ha, G., Haase, K., Jerman, L., Ji, Y., Jolly, C., Lee, J., Lee-Six, H., Malikic, S., Mitchell, T.J., Morris, Q.D., Oesper, L., Peifer, M., Peto, M., Rosebrock, D., Rubanova, Y., Salcedo, A., Sengupta, S., Shi, R., Shin, S.J., Spiro, O., Vembu, S., Wintersinger, J.A., Yang, T.P., Yu, K., Zhu, H., Spellman, P.T., Weinstein, J.N., Chen, Y., Fujita, M., Han, L., Hasegawa, T., Komura, M., Li, J., Mizuno, S., Shimizu, E., Wang, Y., Xu, Y., Yamaguchi, R., Yang, F., Yang, Y., Yoon, C.J., Yuan, Y., Liang, H., Alawi, M., Borozan, I., Brewer, D.S., Cooper, C.S., Desai, N., Grundhoff, A., Iskar, M., Su, X., Zapatka, M., Lichter, P., Alsop, K., Bruxner, T.J.C., Christ, A.N., Cordner, S.M., Cowin, P.A., Drapkin, R., Fereday, S., George, J., Hamilton, A., Holmes, O., Hung, J.A., Kassahn, K.S., Kazakoff, S.H., Kennedy, C.J., Leonard, C.R., Mileshkin, L., Miller, D.K., Arnau, G.M., Mitchell, C., Newell, F., Nones, K., Patch, A.M., Quinn, M.C., Taylor, D.F., Thorne, H., Traficante, N., Vedururu, R., Waddell, N.M., Waring, P.M., Wood, S., Xu, Q., deFazio, A., Anderson, M.J., Antonello, D., Barbour, A.P., Bassi, C., Bersani, S., Cataldo, I., Chantrill, L.A., Chiew, Y.E., Chou, A., Cingarlini, S., Cloonan, N., Corbo, V., Davi, M.V., Duthie, F.R., Gill, A.J., Graham, J.S., Harliwong, I., Jamieson, N.B., Johns, A.L., Kench, J.G., Landoni, L., Lawlor, R.T., Mafficini, A., Merrett, N.D., Miotto, M., Musgrove, E.A., Nagrial, A.M., Oien, K.A., Pajic, M., Pinese, M., Robertson, A.J., Rooman, I., Rusev, B.C., Samra, J.S., Scardoni, M., Scarlett, C.J., Scarpa, A., Sereni, E., Sikora, K.O., Simbolo, M., Taschuk, M.L., Toon, C.W., Vicentini, C., Wu, J., Zeps, N., Behren, A., Burke, H., Cebon, J., Dagg, R.A., De Paoli-Iseppi, R., Dutton-Regester, K., Field, M.A., Fitzgerald, A., Hersey, P., Jakrot, V., Johansson, P.A., Kakavand, H., Kefford, R.F., Lau, L.M.S., Long, G. V., Pickett, H.A., Pritchard, A.L., Pupo, G.M., Saw, R.P.M., Schramm, S.J., Shang, C.A., Shang, P., Spillane, A.J., Stretch, J.R., Tembe, V., Thompson, J.F., Vilain, R.E., Wilmott, J.S., Yang, J.Y., Hayward, N.K., Mann, G.J., Scolyer, R.A., Bartlett, J., Bavi, P., Chadwick, D.E., Chan-Seng-Yue, M., Cleary, S., Connor, A.A., Czajka, K., Denroche, R.E., Dhani, N.C., Eagles, J., Gallinger, S., Grant, R.C., Hedley, D., Hollingsworth, M.A., Jang, G.H., Johns, J., Kalimuthu, S., Liang, S. Ben, Lungu, I., Luo, X., Mbabaali, F., McPherson, T.A.,

Miller, J.K., Moore, M.J., Notta, F., Pasternack, D., Petersen, G.M., Roehrl, M.H.A., Sam, M., Selander, I., Serra, S., Shahabi, S., Thayer, S.P., Timms, L.E., Wilson, G.W., Wilson, J.M., Wouters, B.G., McPherson, J.D., Beck, T.A., Bhandari, V., Collins, C.C., Fleshner, N.E., Fox, N.S., Fraser, M., Heisler, L.E., Lalonde, E., Livingstone, J., Meng, A., Sabelnykova, V.Y., Shiah, Y.J., Van der Kwast, T., Bristow, R.G., Ding, S., Fan, D., Li, L., Nie, Y., Xiao, X., Xing, R., Yang, S., Yu, Y., Zhou, Y., Banks, R.E., Bourque, G., Brennan, P., Letourneau, L., Riazalhosseini, Y., Scelo, G., Vasudev, N., Viksna, J., Lathrop, M., Tost, J., Ahn, S.M., Aparicio, S., Arnould, L., Aure, M.R., Bhosle, S.G., Birney, E., Borg, A., Boyault, S., Brinkman, A.B., Brock, J.E., Broeks, A., Børresen-Dale, A.L., Caldas, C., Chin, S.F., Davies, H., Desmedt, C., Dirix, L., Dronov, S., Ehinger, A., Eyfjord, J.E., Fatima, A., Foekens, J.A., Futreal, P. Andrew, Garred, Ø., Giri, D.D., Glodzik, D., Grabau, D., Hilmarsdottir, H., Hooijer, G.K., Jacquemier, J., Jang, S.J., Jonasson, J.G., Jonkers, J., Kim, H.Y., King, T.A., Knappskog, S., Kong, G., Krishnamurthy, S., Lakhani, S.R., Langerød, A., Larsimont, D., Lee, H.J., Lee, J.Y., Lee, M.T.M., Lingjærde, O.C., MacGrogan, G., Martens, J.W.M., O'Meara, S., Pauporté, I., Pinder, S., Pivot, X., Provenzano, E., Purdie, C.A., Ramakrishna, M., Ramakrishnan, K., Reis-Filho, J., Richardson, A.L., Ringnér, M., Rodriguez, J.B., Rodríguez-González, F.G., Romieu, G., Salgado, R., Sauer, T., Shepherd, R., Sieuwerts, A.M., Simpson, P.T., Smid, M., Sotiriou, C., Span, P.N., Stefánsson, Ó.A., Stenhouse, A., Stunnenberg, H.G., Sweep, F., Tan, B.K.T., Thomas, G., Thompson, A.M., Tommasi, S., Treilleux, I., Tutt, A., Ueno, N.T., Van Laere, S., Van den Eynden, G.G., Vermeulen, P., Viari, A., Vincent-Salomon, A., Wong, B.H., Yates, L., Zou, X., van Deurzen, C.H.M., van de Vijver, M.J., van't Veer, L., Ammerpohl, O., Aukema, S., Bergmann, A.K., Bernhart, S.H., Borkhardt, A., Borst, C., Burkhardt, B., Claviez, A., Goebler, M.E., Haake, A., Haas, S., Hansmann, M., Hoell, J.I., Hummel, M., Karsch, D., Klapper, W., Kneba, M., Kreuz, M., Kube, D., Küppers, R., Lenze, D., Loeffler, M., López, C., Mantovani-Löffler, L., Möller, P., Ott, G., Radlwimmer, B., Richter, J., Rohde, M., Rosenstiel, P.C., Rosenwald, A., Schilhabel, M.B., Schreiber, S., Stadler, P.F., Staib, P., Stilgenbauer, S., Sungalee, S., Szczepanowski, M., Toprak, U.H., Trümper, L.H.P., Wagener, R., Zenz, T., Hovestadt, V., von Kalle, C., Kool, M., Korshunov, A., Landgraf, P., Lehrach, H., Northcott, P.A., Pfister, S.M., Reifenberger, G., Warnatz, H.J., Wolf, S., Yaspo, M.L., Assenov, Y., Gerhauser, C., Minner, S., Schlomm, T., Simon, R., Sauter, G., Sültmann, H., Biswas, N.K., Maitra, A., Majumder, P.P., Sarin, R., Barbi, S., Bonizzato, G.,

Cantù, C., Dei Tos, A.P., Fassan, M., Grimaldi, S., Luchini, C., Malleo, G., Marchegiani, G., Milella, M., Paiella, S., Pea, A., Pederzoli, P., Ruzzenente, A., Salvia, R., Sperandio, N., Arai, Y., Hama, N., Hiraoka, N., Hosoda, F., Nakamura, H., Ojima, H., Okusaka, T., Totoki, Y., Urushidate, T., Fukayama, M., Ishikawa, S., Katai, H., Katoh, H., Komura, D., Rokutan, H., Saito-Adachi, M., Suzuki, A., Taniguchi, H., Tatsuno, K., Ushiku, T., Yachida, S., Yamamoto, S., Aikata, H., Arihiro, K., Ariizumi, S. ichi, Chayama, K., Furuta, M., Gotoh, K., Hayami, S., Hirano, S., Kawakami, Y., Maejima, K., Nakamura, T., Nakano, K., Ohdan, H., Sasaki-Oku, A., Tanaka, H., Ueno, M., Yamamoto, M., Yamaue, H., Choo, S.P., Cutcutache, I., Khuntikeo, N., Ong, C.K., Pairojkul, C., Popescu, I., Ahn, K.S., Aymerich, M., Lopez-Guillermo, A., López-Otín, C., Puente, X.S., Campo, E., Amary, F., Baumhoer, D., Behjati, S., Bjerkehagen, B., Futreal, P. A., Myklebost, O., Pillay, N., Tarpey, P., Tirabosco, R., Zaikova, O., Flanagan, A.M., Boultwood, J., Bowen, D.T., Cazzola, M., Green, A.R., Hellstrom-Lindberg, E., Malcovati, L., Nangalia, J., Papaemmanuil, E., Vyas, P., Ang, Y., Barr, H., Beardsmore, D., Eldridge, M., Gossage, J., Grehan, N., Hanna, G.B., Hayes, S.J., Hupp, T.R., Khoo, D., Lagergren, J., Lovat, L.B., MacRae, S., O'Donovan, M., O'Neill, J.R., Parsons, S.L., Preston, S.R., Puig, S., Roques, T., Sanders, G., Sothi, S., Tavaré, S., Tucker, O., Turkington, R., Underwood, T.J., Welch, I., Fitzgerald, R.C., Berney, D.M., De Bono, J.S., Cahill, D., Camacho, N., Dennis, N.M., Dudderidge, T., Edwards, S.E., Fisher, C., Foster, C.S., Ghori, M., Gill, P., Gnanapragasam, V.J., Gundem, G., Hamdy, F.C., Hawkins, S., Hazell, S., Howat, W., Isaacs, W.B., Karaszi, K., Kay, J.D., Khoo, V., Kote-Jarai, Z., Kremeyer, B., Kumar, P., Lambert, A., Leongamornlert, D.A., Livni, N., Lu, Y.J., Luxton, H.J., Marsden, L., Massie, C.E., Matthews, L., Mayer, E., McDermott, U., Merson, S., Neal, D.E., Ng, A., Nicol, D., Ogden, C., Rowe, E.W., Shah, N.C., Thomas, S., Thompson, A., Verrill, C., Visakorpi, T., Warren, A.Y., Whitaker, H.C., Zhang, Hongwei, van As, N., Eeles, R.A., Abeshouse, A., Agrawal, N., Akbani, R., Al-Ahmadie, H., Albert, M., Aldape, K., Ally, A., Appelbaum, E.L., Armenia, J., Asa, S., Auman, J.T., Balasundaram, M., Balu, S., Barnholtz-Sloan, J., Bathe, O.F., Baylin, S.B., Benz, C., Berchuck, A., Berrios, M., Bigner, D., Birrer, M., Bodenheimer, T., Boice, L., Bootwalla, M.S., Bosenberg, M., Bowlby, R., Boyd, J., Broaddus, R.R., Brock, M., Brooks, D., Bullman, S., Caesar-Johnson, S.J., Carey, T.E., Carlsen, R., Cerfolio, R., Chandan, V.S., Chen, H.W., Cherniack, A.D., Chien, J., Cho, J., Chuah, E., Cibulskis, C., Cope, L., Cordes, M.G., Curley, E., Czerniak, B., Danilova, L., Davis, I.J., Defreitas, T.,

Demchok, J.A., Dhalla, N., Dhir, R., Doddapaneni, H.V., El-Naggar, A., Felau, I., Ferguson, M.L., Finocchiaro, G., Fong, K.M., Frazer, S., Friedman, W., Fronick, C.C., Fulton, L.A., Gabriel, S.B., Gao, J., Gehlenborg, N., Gershenwald, J.E., Ghossein, R., Giama, N.H., Gibbs, R.A., Gomez, C., Govindan, R., Hayes, D.N., Hegde, A.M., Heiman, D.I., Heins, Z., Hepperla, A.J., Holbrook, A., Holt, R.A., Hoyle, A.P., Hruban, R.H., Hu, J., Huang, M., Huntsman, D., Huse, J., Iacobuzio-Donahue, C.A., Ittmann, M., Jayaseelan, J.C., Jefferys, S.R., Jones, C.D., Jones, S.J.M., Juhl, H., Kang, K.J., Karlan, B., Kasaian, K., Kebebew, E., Kim, H.K., Korchina, V., Kundra, R., Lai, P.H., Lander, E., Le, X., Lee, Darlene, Levine, D.A., Lewis, L., Ley, T., Li, H.I., Lin, P., Linehan, W.M., Liu, F.F., Lu, Yiling, Lype, L., Ma, Y., Maglinte, D.T., Mardis, E.R., Marks, J., Marra, M.A., Matthew, T.J., Mayo, M., McCune, K., Meier, S.R., Meng, S., Mieczkowski, P.A., Mikkelsen, T., Miller, C.A., Mills, G.B., Moore, R.A., Morrison, C., Mose, L.E., Moser, C.D., Mungall, A.J., Mungall, K., Mutch, D., Muzny, D.M., Myers, J., Newton, Y., Noble, M.S., O'Donnell, P., O'Neill, B.P., Ochoa, A., Park, Joong Won, Parker, J.S., Pass, H., Pastore, A., Pennell, N.A., Perou, C.M., Petrelli, N., Potapova, O., Rader, J.S., Ramalingam, S., Rathmell, W.K., Reuter, V., Reynolds, S.M., Ringel, M., Roach, J., Roberts, L.R., Robertson, A.G., Sadeghi, S., Saller, C., Sanchez-Vega, F., Schadendorf, D., Schein, J.E., Schmidt, H.K., Schultz, N., Seethala, R., Senbabaoglu, Y., Shelton, T., Shi, Y., Shih, J., Shmulevich, I., Shriver, C., Signoretti, S., Simons, J. V., Singer, S., Sipahimalani, P., Skelly, T.J., Smith-McCune, K., Socci, N.D., Soloway, M.G., Sood, A.K., Tam, A., Tan, D., Tarnuzzer, R., Thiessen, N., Thompson, R.H., Thorne, L.B., Tsao, M., Umbricht, C., Van Den Berg, D.J., Van Meir, E.G., Veluvolu, U., Voet, D., Wang, L., Weinberger, P., Weisenberger, D.J., Wigle, D., Wilkerson, M.D., Wilson, R.K., Winterhoff, B., Wiznerowicz, M., Wong, T., Wong, W., Xi, L., Yau, C., Zhang, Hailei, Zhang, Hongxin, Zhang, Jiashan, 2020. Pan-cancer analysis of whole genomes. Nature 2020 578:7793 578, 82–93. https://doi.org/10.1038/s41586-020-1969-6

Cao, L., Li, W., Kim, S., Brodie, S.G., Deng, C.X., 2003. Senescence, aging, and malignant transformation mediated by p53 in mice lacking the Brca1 full-length isoform. Genes & Development 17, 201. https://doi.org/10.1101/GAD.1050003

Capellini, T.D., Chen, H., Cao, J., Doxey, A.C., Kiapour, A.M., Schoor, M., and Kingsley, D.M. (2017). Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. *Nat. Genet.* 49, 1202.

Capellini, T.D., Chen, H., Cao, J., Doxey, A.C., Kiapour, A.M., Schoor, M., Kingsley, D.M., 2017. Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. Nat Genet 49, 1202.

Capellini, T.D., Di Giacomo, G., Salsi, V., Brendolan, A., Ferretti, E., Srivastava, D., Zappavigna, V., and Selleri, L. (2006). Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. *Development* 133, 2263–2273.

Capra, J.A., Erwin, G.D., McKinsey, G., Rubenstein, J.L.R., and Pollard, K.S. (2013). Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. 368, 20130025.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q., Lander, E.S., 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nature Genetics 1999 22:3 22, 231–238. https://doi.org/10.1038/10290

Carmeliet, P., P, C., 2005. VEGF as a key mediator of angiogenesis in cancer. Oncology 69 Suppl 3, 4–10. https://doi.org/10.1159/000088478

Carnes, B.A., and Olshansky, S.J. (1993). Evolutionary Perspectives on Human Senescence. *Popul. Dev. Rev.* 19, 793.

Carnes, B.A., Olshansky, S.J., 1993. Evolutionary Perspectives on Human Senescence. Population and Development Review 19, 793. https://doi.org/10.2307/2938414

Carrier, D.R., Kapoor, A.K., Kimura, T., Nickels, M.K., Satwanti, Scott, E.C., So, J.K., and Trinkaus, E. (1984). The energetic paradox of human running and hominid evolution [and comments and reply*]. Curr. Anthropol*. 25, 483–495.

Carroll, S.B., 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell 134, 25–36. https://doi.org/10.1016/J.CELL.2008.06.030

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J. and Myers, R.M., 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science*, 327(5963), pp.302-305.

Chang, A.Y., Skirbekk, V.F., Tyrovolas, S., Kassebaum, N.J., Dieleman, J.L., 2019. Measuring population ageing: an analysis of the Global Burden of Disease Study 2017. The Lancet Public Health 4, e159–e167. https://doi.org/10.1016/S2468-2667(19)30019-2

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.

Charles, M.D., Haloman, S., Chen, L., Ward, S.R., Fithian, D., and Afra, R. (2013). Magnetic resonance imaging-based topographical differences between control and recurrent patellofemoral instability patients. *Am. J. Sports Med.* 41, 374–384.

Charlesworth, B., 2001. Patterns of age-specific means and genetic variances of mortality rates predicted by the mutation-accumulation theory of ageing. Journal of Theoretical Biology 210, 47–65. https://doi.org/10.1006/jtbi.2001.2296

Chasman, D., Adams, R.M., 2001. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. Journal of Molecular Biology 307, 683–706. https://doi.org/10.1006/JMBI.2001.4510

Chen, H., Capellini, T.D., Schoor, M., Mortlock, D.P., Reddi, A.H., and Kingsley, D.M. (2016). Heads, shoulders, elbows, knees, and toes: modular Gdf5 enhancers control different joints in the vertebrate skeleton. *PLoS Genet.* 12, e1006454.

Cheng, C., Kirkpatrick, M., 2021. Molecular evolution and the decline of purifying selection with age. Nat Commun 12. https://doi.org/10.1038/S41467-021-22981-9

Choi, Y.S.Y., Yang, J.-S.S., Choi, Y.S.Y., Ryu, S.H., Kim, S., 2009. Evolutionary conservation in multiple faces of protein interaction. Proteins: Structure, Function, and Bioinformatics 77, 14–25. https://doi.org/10.1002/PROT.22410

Chokalingam, K., Hunter, S., Gooch, C., Frede, C., Florer, J., Wenstrup, R., and Butler, D. (2009). Three-dimensional in vitro effects of compression and time in culture on aggregate modulus and on gene expression and protein content of collagen type II in murine chondrocytes. *Tissue Eng. Part A* 15, 2807–2816.

CM, N., ER, A., KN, C., D, C., HJ, H., D, L., K, M., GL, M., Nicolae, C.M., Aho, E.R., Choe, K.N., Constantin, D., Hu, H.J., Lee, D., Myung, K., Moldovan, G.L., 2015. A novel role for the mono-ADP-

ribosyltransferase PARP14/ARTD8 in promoting homologous recombination and protecting against replication stress. Nucleic Acids Res 43, 3143–3153. https://doi.org/10.1093/NAR/GKV147

Coetzee, S.G., Coetzee, G.A., and Hazelett, D.J. (2015). motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* 31, 3847–3849.

Collyer, M.L., Adams, D.C., 2018. RRPP: An r package for fitting linear models to high-dimensional data using residual randomization. Methods in Ecology and Evolution 9, 1772–1779. https://doi.org/10.1111/2041-210X.13029

Comfort, A., 1954. BIOLOGICAL ASPECTS OF SENESCENCE. Biological Reviews 29, 284–329. https://doi.org/10.1111/j.1469-185X.1954.tb00598.x

Conradt, E., Adkins, D.E., Crowell, S.E., Raby, K.L., Diamond, L.M., Ellis, B., 2018. Incorporating epigenetic mechanisms to advance fetal programming theories. Development and Psychopathology 30, 807–824. https://doi.org/10.1017/S0954579418000469

Cooper, G.M., Goode, D.L., Ng, S.B., Sidow, A., Bamshad, M.J., Shendure, J., Nickerson, D.A., 2010. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. Nature Methods. https://doi.org/10.1038/nmeth0410-250

Cooper, G.M., Shendure, J., 2011. Needles in stacks of needles: Finding disease-causal variants in a wealth of genomic data. Nature Reviews Genetics. https://doi.org/10.1038/nrg3046

Cotney, J., Leng, J., Yin, J., Reilly, S.K., DeMare, L.E., Emera, D., Ayoub, A.E., Rakic, P., and Noonan, J.P. (2013). The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* 154, 185–196.

Criscione, S.W., Teo, Y.V., Neretti, N., 2016. The Chromatin Landscape of Cellular Senescence. Trends in Genetics 32, 751–761. https://doi.org/10.1016/J.TIG.2016.09.005

Croft, D.P., Brent, L.J.N., Franks, D.W., Cant, M.A., 2015. The evolution of prolonged life after reproduction. Trends in Ecology & Evolution 30, 407–416. https://doi.org/10.1016/j.tree.2015.04.011

Cupit, L.D., Schmidt, V.A., Miller, F., and Bahou, W.F. (2004). Distinct PAR/IQGAP expression patterns during murine development: implications for thrombin-associated cytoskeletal reorganization. *Mamm. Genome* 15, 618–629.

Dakal, T.C., Kala, D., Dhiman, G., Yadav, V., Krokhotin, A., Dokholyan, N. V., 2017. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene. Scientific Reports 7. https://doi.org/10.1038/S41598-017-06575-4

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.

Darwin, C. (1888). The descent of man and selection in relation to sex (Murray).

Das Neves Borges, P., Forte, A.E., Vincent, T.L., Dini, D., and Marenzana, M. (2014). Rapid, automated imaging of mouse articular cartilage by microCT for early detection of osteoarthritis and finite element modelling of joint mechanics. *Osteoarthr. Cartil.* 22, 1419–1428.

Dasmeh, P., Girard, É., Serohijos, A.W.R., 2017. Highly expressed genes evolve under strong epistasis from a proteome-wide scan in E. coli. Scientific Reports 2017 7:1 7, 1–6. https://doi.org/10.1038/s41598-017-16030-z

Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., and Onate, K.C., 2018 (2018). The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* 46, D794–D801.

Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., Onate, K.C., Graham, K., Miyasato, S.R., Dreszer, T.R., Strattan, J.S., Jolanki, O., Tanaka, F.Y., Cherry, J.M., 2018. The Encyclopedia of DNA elements (ENCODE): Data portal update. Nucleic Acids Research 46, D794–D801. https://doi.org/10.1093/nar/gkx1081

Davydov, E. V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLoS Comput. Biol.* 6, e1001025.

Day, K., Waite, L.L., Thalacker-Mercer, A., West, A., Bamman, M.M., Brooks, J.D., Myers, R.M., Absher, D., 2013. Differential DNA methylation with age displays both common and dynamic features across human tissues that are influenced by CpG landscape. Genome Biology 14, 1–19. https://doi.org/10.1186/gb-2013-14-9-r102

De Magalhães, J.P., 2013. How ageing processes influence cancer. Nature Reviews Cancer 2013 13:5 13, 357–365. https://doi.org/10.1038/nrc3497

Decker, R.S. (2017). Articular cartilage and joint development from embryogenesis to adulthood. *Semin. Cell Dev. Biol.* 62, 50–56.

Decker, R.S., Koyama, E., and Pacifici, M. (2014). Genesis and morphogenesis of limb synovial joints and articular cartilage. *Matrix Biol.* 39, 5–10.

del-Toro, N., Duesbury, M., Koch, M., Perfetto, L., Shrivastava, A., Ochoa, D., Wagih, O., Piñero, J., Kotlyar, M., Pastrello, C., Beltrao, P., Furlong, L.I., Jurisica, I., Hermjakob, H., Orchard, S., Porras, P., 2019. Capturing variation impact on molecular interactions in the IMEx Consortium mutations data set. Nature Communications 2019 10:1 10, 1–14. https://doi.org/10.1038/s41467-018-07709-6

Delignette-Muller, M.L., and Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. *J. Stat. Softw.* 64, 1–34.

Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al. (2016). High-throughput discovery of novel developmental phenotypes. *Nature* 537, 508–514.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., et al. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* 9, e1000582.

DiLeone, R.J., Russell, L.B., and Kingsley, D.M. (1998). An extensive 3' regulatory region controls expression of Bmp5 in specific anatomical structures of the mouse embryo. *Genetics* 148, 401–408.

Domcke, S., Hill, A.J., Daza, R.M., Cao, J., O'Day, D.R., Pliner, H.A., Aldinger, K.A., Pokholok, D., Zhang, F., Milbank, J.H., Zager, M.A., Glass, I.A., Steemers, F.J., Doherty, D., Trapnel, C., Cusanovich, D.A., Shendure, J., 2020. A human cell atlas of fetal chromatin accessibility. Science (1979) 370. https://doi.org/10.1126/science.aba7612

Draceni, Y., Pechmann, S., 2019. Pervasive convergent evolution and extreme phenotypes define chaperone requirements of protein homeostasis. Proc Natl Acad Sci U S A 116, 20009–20014. https://doi.org/10.1073/PNAS.1904611116/-/DCSUPPLEMENTAL

Duarte, J.M., Srebniak, A., Schärer, M.A., Capitani, G., 2012. Protein interface classification by evolutionary analysis. BMC Bioinformatics 2012 13:1 13, 1–16. https://doi.org/10.1186/1471-2105-13-334

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B.R., Landt, S.G., Lee, B.K., Pauli, F., Rosenbloom, K.R., Sabo, P., Safi, A., Sanyal, A., Shoresh, N., Simon, J.M., Song, L., Trinklein, N.D., Altshuler, R.C., Birney, E., Brown, J.B., Cheng, C., Djebali, S., Dong, X., Ernst, J., Furey, T.S., Gerstein, M., Giardine, B., Greven, M., Hardison, R.C., Harris, R.S., Herrero, J., Hoffman, M.M., Iyer, S., Kellis, M., Kheradpour, P., Lassmann, T., Li, Q., Lin, X., Marinov, G.K., Merkel, A., Mortazavi, A., Parker, S.C.J., Reddy, T.E., Rozowsky, J., Schlesinger, F., Thurman, R.E., Wang, J., Ward, L.D., Whitfield, T.W., Wilder, S.P., Wu, W., Xi, H.S., Yip, K.Y., Zhuang, J., Bernstein, B.E., Green, E.D., Gunter, C., Snyder, M., Pazin, M.J., Lowdon, R.F., Dillon, L.A.L., Adams, L.B., Kelly, C.J., Zhang, J., Wexler, J.R., Good, P.J., Feingold, E.A., Crawford, G.E., Dekker, J., Elnitski, L., Farnham, P.J., Giddings, M.C., Gingeras, T.R., Guigó, R., Hubbard, T.J., Kent, W.J., Lieb, J.D., Margulies, E.H., Myers, R.M., Stamatoyannopoulos, J.A., Tenenbaum, S.A., Weng, Z., White, K.P., Wold, B., Yu, Y., Wrobel, J., Risk, B.A., Gunawardena, H.P., Kuiper, H.C., Maier, C.W., Xie, L., Chen, X., Mikkelsen, T.S., Gillespie, S., Goren, A., Ram, O., Zhang, X., Wang, L., Issner, R., Coyne, M.J., Durham, T., Ku, M., Truong, T., Eaton, M.L., Dobin, A., Tanzer, A., Lagarde, J., Lin, W., Xue, C., Williams, B.A., Zaleski, C., Röder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Batut, P., Bell, I., Bell, K., Chakrabortty, S., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Li, G., Luo, O.J., Park, E., Preall, J.B., Presaud, K., Ribeca, P., Robyr, D., Ruan, X., Sammeth, M., Sandhu, K.S., Schaeffer, L., See, L.H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, Huaien, Hayashizaki, Y., Reymond, A., Antonarakis, S.E., Hannon, G.J., Ruan, Y., Carninci, P., Sloan, C.A., Learned, K., Malladi, V.S., Wong, M.C., Barber, G.P., Cline, M.S., Dreszer, T.R., Heitner, S.G., Karolchik, D., Kirkup, V.M., Meyer, L.R., Long, J.C., Maddren, M., Raney, B.J., Grasfeder, L.L., Giresi, P.G., Battenhouse, A., Sheffield, N.C., Showers, K.A., London, D., Bhinge,

A.A., Shestak, C., Schaner, M.R., Kim, S.K., Zhang, Z.Z., Mieczkowski, P.A., Mieczkowska, J.O., Liu, Z., McDaniell, R.M., Ni, Y., Rashid, N.U., Kim, M.J., Adar, S., Zhang, Zhancheng, Wang, T., Winter, D., Keefe, D., Iyer, V.R., Zheng, M., Wang, P., Gertz, J., Vielmetter, J., Partridge, E.C., Varley, K.E., Gasper, C., Bansal, A., Pepke, S., Jain, P., Amrhein, H., Bowling, K.M., Anaya, M., Cross, M.K., Muratet, M.A., Newberry, K.M., McCue, K., Nesmith, A.S., Fisher-Aylor, K.I., Pusey, B., DeSalvo, G., Parker, S.L., Balasubramanian, Sreeram, Davis, N.S., Meadows, S.K., Eggleston, T., Newberry, J.S., Levy, S.E., Absher, D.M., Wong, W.H., Blow, M.J., Visel, A., Pennachio, L.A., Petrykowska, H.M., Abyzov, A., Aken, B., Barrell, D., Barson, G., Berry, A., Bignell, A., Boychenko, V., Bussotti, G., Davidson, C., Despacio-Reyes, G., Diekhans, M., Ezkurdia, I., Frankish, A., Gilbert, J., Gonzalez, J.M., Griffiths, E., Harte, R., Hendrix, D.A., Hunt, T., Jungreis, I., Kay, M., Khurana, E., Leng, J., Lin, M.F., Loveland, J., Lu, Z., Manthravadi, D., Mariotti, M., Mudge, J., Mukherjee, G., Notredame, C., Pei, B., Rodriguez, J.M., Saunders, G., Sboner, A., Searle, S., Sisu, C., Snow, C., Steward, C., Tapanari, E., Tress, M.L., Van Baren, M.J., Washietl, S., Wilming, L., Zadissa, A., Zhang, Zhengdong, Brent, M., Haussler, D., Valencia, A., Addleman, N., Alexander, R.P., Auerbach, R.K., Balasubramanian, Suganthi, Bettinger, K., Bhardwaj, N., Boyle, A.P., Cao, A.R., Cayting, P., Charos, A., Cheng, Y., Eastman, C., Euskirchen, G., Fleming, J.D., Grubert, F., Habegger, L., Hariharan, M., Harmanci, A., Iyengar, S., Jin, V.X., Karczewski, K.J., Kasowski, M., Lacroute, P., Lam, H., Lamarre-Vincent, N., Lian, J., Lindahl-Allen, M., Min, R., Miotto, B., Monahan, H., Moqtaderi, Z., Mu, X.J., O'Geen, H., Ouyang, Z., Patacsil, D., Raha, D., Ramirez, L., Reed, B., Shi, M., Slifer, T., Witt, H., Wu, L., Xu, X., Yan, K.K., Yang, X., Struhl, K., Weissman, S.M., Penalva, L.O., Karmakar, S., Bhanvadia, R.R., Choudhury, A., Domanus, M., Ma, L., Moran, J., Victorsen, A., Auer, T., Centanin, L., Eichenlaub, M., Gruhl, F., Heermann, S., Hoeckendorf, B., Inoue, D., Kellner, T., Kirchmaier, S., Mueller, C., Reinhardt, R., Schertel, L., Schneider, S., Sinn, R., Wittbrodt, B., Wittbrodt, J., Jain, G., Balasundaram, G., Bates, D.L., Byron, R., Canfield, T.K., Diegel, M.J., Dunn, D., Ebersol, A.K., Frum, T., Garg, K., Gist, E., Hansen, R.S., Boatman, L., Haugen, E., Humbert, R., Johnson, A.K., Johnson, E.M., Kutyavin, T. V., Lee, K., Lotakis, D., Maurano, M.T., Neph, S.J., Neri, F. V., Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Rynes, E., Sanchez, M.E., Sandstrom, R.S., Shafer, A.O., Stergachis, A.B., Thomas, S., Vernot, B., Vierstra, J., Vong, S., Wang, Hao, Weaver, M.A., Yan, Y., Zhang, M., Akey, J.M., Bender,

M., Dorschner, M.O., Groudine, M., MacCoss, M.J., Navas, P., Stamatoyannopoulos, G., Beal, K., Brazma, A., Flicek, P., Johnson, N., Lukk, M., Luscombe, N.M., Sobral, D., Vaquerizas, J.M., Batzoglou, S., Sidow, A., Hussami, N., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M.W., Schaub, M.A., Miller, W., Bickel, P.J., Banfai, B., Boley, N.P., Huang, H., Li, J.J., Noble, W.S., Bilmes, J.A., Buske, O.J., Sahu, A.D., Kharchenko, P. V., Park, P.J., Baker, D., Taylor, J., Lochovsky, L., 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. https://doi.org/10.1038/nature11247

Eckstein, F., Ateshian, G., Burgkart, R., Burstein, D., Cicuttini, F., Dardzinski, B., Gray, M., Link, T.M., Majumdar, S., Mosher, T., et al. (2006). Proposal for a nomenclature for magnetic resonance imaging based measures of articular cartilage in osteoarthritis. *Osteoarthr. Cartil*. 14, 974–983.

Elliott, J.A., Stetson, M.H., Menaker, M., Grocock, C.A., Clarke, J.R., Farmer, D.S., King, J.R., Beach, R.F., Craig, G.B., 2002. A Test of Evolutionary Theories of Senescence. Science of Aging Knowledge Environment 2002, 487–497. https://doi.org/10.1126/SAGEKE.2002.37.CP17

Ellison, P.T., Ottinger, M.A., 2014. A comparative perspective on reproductive aging, reproductive cessation, post-reproductive life, and social behavior. Sociality, Hierarchy, Health: Comparative Biodemography: A Collection of Papers.

Emery Thompson M., E.P., 2017. Fertility and Fecundity, in: Chimpanzees AndHuman Evolution. pp. 217–258.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.

Esposito, D., Weile, J., Shendure, J., Starita, L.M., Papenfuss, A.T., Roth, F.P., Fowler, D.M., Rubin, A.F., 2019. MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect. Genome Biology 2019 20:1 20, 1–11. https://doi.org/10.1186/S13059-019-1845-6

Farré, X., Molina, R., Barteri, F., Timmers, P.R.H.J., Joshi, P.K., Oliva, B., Acosta, S., Esteve-Altava, B., Navarro, A., Muntané, G., 2021a. Comparative Analysis of Mammal Genomes Unveils Key Genomic Variability for Human Life Span. Molecular Biology and Evolution 38, 4948–4961. https://doi.org/10.1093/MOLBEV/MSAB219

Farré, X., Molina, R., Barteri, F., Timmers, P.R.H.J., Joshi, P.K., Oliva, B., Acosta, S., Esteve-Altava, B., Navarro, A., Muntané, G., 2021b. Comparative Analysis of Mammal Genomes Unveils Key Genomic Variability for Human Life Span. Molecular Biology and Evolution 38, 4948–4961. https://doi.org/10.1093/MOLBEV/MSAB219

Felson, D.T., Anderson, J.J., Naimark, A., Walker, A.M., and Meenan, R.F. (1988). Obesity and knee osteoarthritis: the Framingham Study. *Ann. Intern. Med*. 109, 18–24.

Ferrer-Admetlla, A., Liang, M., Korneliussen, T., and Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol*. 31, 1275–1291.

Ferris, E., Abegglen, L.M., Schiffman, J.D., Gregg, C., 2018. Accelerated Evolution in Distinctive Species Reveals Candidate Elements for Clinically Relevant Traits, Including Mutation and Cancer Resistance. Cell Reports 22, 2742–2755. https://doi.org/10.1016/j.celrep.2018.02.008

Ferrucci, L., Fabbri, E., 2018. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. Nature Reviews Cardiology 2018 15:9 15, 505–522. https://doi.org/10.1038/s41569-018-0064-2

Finch, C.E., 2010. Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition. Proceedings of the National Academy of Sciences 107, 1718–1724. https://doi.org/10.1073/pnas.0909606106

Finch, C.E., Sapolsky, R.M., 1999. The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms☆. Neurobiology of Aging 20, 407–428. https://doi.org/10.1016/S0197-4580(99)00053-6

Finger, F., Schörle, C., Zien, A., Gebhard, P., Goldring, M.B., and Aigner, T. (2003). Molecular phenotyping of human chondrocyte cell lines T/C-28a2, T/C-28a4, and C-28/I2. *Arthritis Rheum.* 48, 3395–3403.

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235.

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F.R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J.R.B., Okada, Y., Raychaudhuri, S., Daly, M.J., Patterson, N., Neale, B.M., Price, A.L., Patterson, N., Neale, B.M., Price,

A.L., 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature Genetics 47, 1228–1235. https://doi.org/10.1038/ng.3404

Folkesson, J., Dam, E.B., Olsen, O.F., Pettersen, P.C., and Christiansen, C. (2007). Segmenting articular cartilage automatically using a voxel classification approach. IEEE Trans. *Med. Imaging* 26, 106–115.

Fowler, D.M., Fields, S., 2014. Deep mutational scanning: a new style of protein science. Nature Methods 2014 11:8 11, 801–807. https://doi.org/10.1038/nmeth.3027

Fox, J., and Weisberg, S. (2019). An R Companion to Applied Regression, Third Edition.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T.D., Wu, Y.Z., Plass, C., Esteller, M., 2005. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102, 10604–10609. https://doi.org/10.1073/pnas.0500398102

Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., Panourgia, M.P., Invidia, L., Celani, L., Scurti, M., Cevenini, E., Castellani, G.C., Salvioli, S., 2007. Inflammaging and anti-inflammaging: A systemic perspective on aging and longevity emerged from studies in humans. Mechanisms of Ageing and Development 128, 92–105. https://doi.org/10.1016/J.MAD.2006.11.016

Frelat, M.A., Shaw, C.N., Sukhdeo, S., Hublin, J.-J., Benazzi, S., and Ryan, T.M. (2017). Evolution of the hominin knee and ankle. *J. Hum. Evol.* 108, 147–160.

Frenk, S., Houseley, J., 2018. Gene expression hallmarks of cellular ageing. Biogerontology 19, 547–566. https://doi.org/10.1007/s10522-018-9750-z

Fulop, T., Le Page, A., Fortin, C., Witkowski, J.M., Dupuis, G., Larbi, A., 2014. Cellular signaling in the aging immune system. Current Opinion in Immunology 29, 105–111. https://doi.org/10.1016/J.COI.2014.05.007

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. (2002). The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229.

Gagné, D., French, R.L., Narayanan, C., Simonović, M., Agarwal, P.K., Doucet, N., 2015. Perturbation of the Conformational Dynamics of an Active-Site Loop Alters Enzyme Activity. Structure 23, 2256–2266. https://doi.org/10.1016/J.STR.2015.10.011

Gamer, L.W., Pregizer, S., Gamer, J., Feigenson, M., Ionescu, A., Li, Q., Han, L., and Rosen, V. (2018). The Role of Bmp2 in the Maturation and Maintenance of the Murine Knee Joint. *J. Bone Miner. Res.* 33, 1708–1717.

Gardner, E., and O'Rahilly, R. (1968). The early development of the knee joint in staged human embryos. *J. Anat.* 102, 289–299.

Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.-R., Palamara, P.F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B.M., Gusev, A., et al. (2017). Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* 49, 1421–1427.

Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.-R., Palamara, P.F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B.M., Gusev, A., Price, A.L., 2017. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. Nature Genetics 49, 1421–1427. https://doi.org/10.1038/ng.3954

Gehrke, A.R., Schneider, I., de la Calle-Mustienes, E., Tena, J.J., Gomez-Marin, C., Chandran, M., Nakamura, T., Braasch, I., Postlethwait, J.H., Gómez-Skarmeta, J.L., et al. (2015). Deep conservation of wrist and digit enhancers in fish. *Proc. Natl. Acad. Sci*. 112, 803–808.

Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2015). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* 32, btv562.

George Priya Doss, C., Rajasekaran, R., Sudandiradoss, C., Ramanathan, K., Purohit, R., Sethumadhavan, R., 2008. A novel computational and structural analysis of nsSNPs in CFTR gene. Genomic Med 2, 23–32. https://doi.org/10.1007/S11568-008-9019-8

Giacomelli, A.O., Yang, X., Lintner, R.E., McFarland, J.M., Duby, M., Kim, J., Howard, T.P., Takeda, D.Y., Ly, S.H., Kim, E., Gannon, H.S., Hurhula, B., Sharpe, T., Goodale, A., Fritchman, B., Steelman, S., Vazquez, F., Tsherniak, A., Aguirre, A.J., Doench, J.G., Piccioni, F., Roberts, C.W.M., Meyerson, M., Getz, G., Johannessen, C.M., Root, D.E., Hahn, W.C., 2018. Mutational processes shape the

landscape of TP53 mutations in human cancer. Nature Genetics 2018 50:10 50, 1381–1387. https://doi.org/10.1038/s41588-018-0204-y

Gibbs, R.A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J.G., Zhu, Y., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.

Gibbs, R.A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J.G., Zhu, Y., Wang, J., Chang, Y., Feng, Q., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Lan, T., Li, G., Li, J., Li, Yingrui, Liu, S., Liu, Xiao, Lu, Y., Ma, X., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Xu, X., Yin, Y., Zhang, D., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Lander, E.S., Altshuler, D.M., Gabriel, S.B., Gupta, N., Gharani, N., Toji, L.H., Gerry, N.P., Resch, A.M., Flicek, P., Barker, J., Clarke, L., Gil, L., Hunt, S.E., Kelman, G., Kulesha, E., Leinonen, R., McLaren, W.M., Radhakrishnan, R., Roa, A., Smirnov, D., Smith, R.E., Streeter, I., Thormann, A., Toneva, I., Vaughan, B., Zheng-Bradley, X., Bentley, D.R., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Lehrach, H., Sudbrak, R., Albrecht, M.W., Amstislavskiy, V.S., Borodina, T.A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M.-L., Mardis, E.R., Wilson, R.K., Fulton, L., Fulton, R., Sherry, S.T., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Church, D., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., O'Sullivan, C., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H., McVean, G.A., Durbin, R.M., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T.M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Schmidt, J.P., Davies, C.J., Gollub, J., Webster, T., Wong, B., Zhan, Y., Auton, A., Campbell, C.L., Kong, Y., Marcketta, A., Gibbs, R.A., Yu, F., Antunes, L., Bainbridge, M., Muzny, D., Sabo, A., Huang, Z., Wang, J., Coin, L.J.M., Fang, L., Guo, X., Jin, X., Li, G., Li, Q., Li, Yingrui, Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Xie, Y., Ye, C., Yu, C., Zhang, F., Zheng, H., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Marth, G.T., Garrison, E.P., Kural, D., Lee, W.-P., Fung Leong, W., Stromberg, M., Ward, A.N., Wu, J., Zhang, M., Daly, M.J., DePristo, M.A., Handsaker, R.E., Altshuler, D.M., Banks, E., Bhatia, G., del Angel, G., Gabriel, S.B., Genovese, G., Gupta, N., Li, H., Kashin, S., Lander, E.S., McCarroll, S.A., Nemesh, J.C., Poplin, R.E., Yoon, S.C., Lihm, J., Makarov, V., Clark, A.G., Gottipati, S., Keinan, A., Rodriguez-Flores, J.L., Korbel, J.O., Rausch, T., Fritz, M.H., Stütz, A.M., Flicek, P., Beal, K., Clarke, L., Datta, A., Herrero, J., McLaren, W.M., Ritchie, G.R.S.,

Smith, R.E., Zerbino, D., Zheng-Bradley, X., Sabeti, P.C., Shlyakhter, I., Schaffner, S.F., Vitti, J., Cooper, D.N., Ball, E. V., Stenson, P.D., Bentley, D.R., Barnes, B., Bauer, M., Keira Cheetham, R., Cox, A., Eberle, M., Humphray, S., Kahn, S., Murray, L., Peden, J., Shaw, R., Kenny, E.E., Batzer, M.A., Konkel, M.K., Walker, J.A., MacArthur, D.G., Lek, M., Sudbrak, R., Amstislavskiy, V.S., Herwig, R., Mardis, E.R., Ding, L., Koboldt, D.C., Larson, D., Ye, Kai, Gravel, S., Swaroop, A., Chew, E., Lappalainen, T., Erlich, Y., Gymrek, M., Frederick Willems, T., Simpson, J.T., Shriver, M.D., Rosenfeld, J.A., Bustamante, C.D., Montgomery, S.B., De La Vega, F.M., Byrnes, J.K., Carroll, A.W., DeGorter, M.K., Lacroute, P., Maples, B.K., Martin, A.R., Moreno-Estrada, A., Shringarpure, S.S., Zakharia, F., Halperin, E., Baran, Y., Lee, C., Cerveira, E., Hwang, J., Malhotra, A., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Hyland, F.C.L., Craig, D.W., Christoforides, A., Homer, N., Izatt, T., Kurdoglu, A.A., Sinari, S.A., Squire, K., Sherry, S.T., Xiao, C., Sebat, J., Antaki, D., Gujral, M., Noor, A., Ye, Kenny, Burchard, E.G., Hernandez, R.D., Gignoux, C.R., Haussler, D., Katzman, S.J., James Kent, W., Howie, B., Ruiz-Linares, A., Dermitzakis, E.T., Devine, S.E., Abecasis, G.R., Min Kang, H., Kidd, J.M., Blackwell, T., Caron, S., Chen, W., Emery, S., Fritsche, L., Fuchsberger, C., Jun, G., Li, B., Lyons, R., Scheller, C., Sidore, C., Song, S., Sliwerska, E., Taliun, D., Tan, A., Welch, R., Kate Wing, M., Zhan, X., Awadalla, P., Hodgkinson, A., Li, Yun, Shi, X., Quitadamo, A., Lunter, G., McVean, G.A., Marchini, J.L., Myers, S., Churchhouse, C., Delaneau, O., Gupta-Hinch, A., Kretzschmar, W., Iqbal, Z., Mathieson, I., Menelaou, A., Rimmer, A., Xifara, D.K., Oleksyk, T.K., Fu, Yunxin, Liu, Xiaoming, Xiong, M., Jorde, L., Witherspoon, D., Xing, J., Eichler, E.E., Browning, B.L., Browning, S.R., Hormozdiari, F., Sudmant, P.H., Khurana, E., Durbin, R.M., Hurles, M.E., Tyler-Smith, C., Albers, C.A., Ayub, Q., Balasubramaniam, S., Chen, Y., Colonna, V., Danecek, P., Jostins, L., Keane, T.M., McCarthy, S., Walter, K., Xue, Y., Gerstein, M.B., Abyzov, A., Balasubramanian, S., Chen, J., Clarke, D., Fu, Yao, Harmanci, A.O., Jin, M., Lee, D., Liu, J., Jasmine Mu, X., Zhang, J., Zhang, Yan, Li, Yingrui, Luo, R., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Marth, G.T., Garrison, E.P., Kural, D., Lee, W.-P., Ward, A.N., Wu, J., Zhang, M., McCarroll, S.A., Handsaker, R.E., Altshuler, D.M., Banks, E., del Angel, G., Genovese, G., Hartl, C., Li, H., Kashin, S., Nemesh, J.C., Shakir, K., Yoon, S.C., Lihm, J., Makarov, V., Degenhardt, J., Korbel, J.O., Fritz, M.H., Meiers, S., Raeder, B., Rausch, T., Stütz, A.M., Flicek, P., Paolo Casale, F., Clarke, L., Smith, R.E., Stegle, O., Zheng-Bradley, X., Bentley, D.R., Barnes, B.,

Keira Cheetham, R., Eberle, M., Humphray, S., Kahn, S., Murray, L., Shaw, R., Lameijer, E.-W., Batzer, M.A., Konkel, M.K., Walker, J.A., Ding, L., Hall, I., Ye, Kai, Lacroute, P., Lee, C., Cerveira, E., Malhotra, A., Hwang, J., Plewczynski, D., Radew, K., Romanovitch, M., Zhang, C., Craig, D.W., Homer, N., Church, D., Xiao, C., Sebat, J., Antaki, D., Bafna, V., Michaelson, J., Ye, Kenny, Devine, S.E., Gardner, E.J., Abecasis, G.R., Kidd, J.M., Mills, R.E., Dayama, G., Emery, S., Jun, G., Shi, X., Quitadamo, A., Lunter, G., McVean, G.A., Chen, K., Fan, X., Chong, Z., Chen, T., Witherspoon, D., Xing, J., Eichler, E.E., Chaisson, M.J., Hormozdiari, F., Huddleston, J., Malig, M., Nelson, B.J., Sudmant, P.H., Parrish, N.F., Khurana, E., Hurles, M.E., Blackburne, B., Lindsay, S.J., Ning, Z., Walter, K., Zhang, Yujun, Gerstein, M.B., Abyzov, A., Chen, J., Clarke, D., Lam, H., Jasmine Mu, X., Sisu, C., Zhang, J., Zhang, Yan, Gibbs, R.A., Yu, F., Bainbridge, M., Challis, D., Evani, U.S., Kovar, C., Lu, J., Muzny, D., Nagaswamy, U., Reid, J.G., Sabo, A., Yu, J., Guo, X., Li, W., Li, Yingrui, Wu, R., Marth, G.T., Garrison, E.P., Fung Leong, W., Ward, A.N., del Angel, G., DePristo, M.A., Gabriel, S.B., Gupta, N., Hartl, C., Poplin, R.E., Clark, A.G., Rodriguez-Flores, J.L., Flicek, P., Clarke, L., Smith, R.E., Zheng-Bradley, X., MacArthur, D.G., Mardis, E.R., Fulton, R., Koboldt, D.C., Gravel, S., Bustamante, C.D., Craig, D.W., Christoforides, A., Homer, N., Izatt, T., Sherry, S.T., Xiao, C., Dermitzakis, E.T., Abecasis, G.R., Min Kang, H., McVean, G.A., Gerstein, M.B., Balasubramanian, S., Habegger, L., Yu, H., Flicek, P., Clarke, L., Cunningham, F., Dunham, I., Zerbino, D., Zheng-Bradley, X., Lage, K., Berg Jespersen, J., Horn, H., Montgomery, S.B., DeGorter, M.K., Khurana, E., Tyler-Smith, C., Chen, Y., Colonna, V., Xue, Y., Gerstein, M.B., Balasubramanian, S., Fu, Yao, Kim, D., Auton, A., Marcketta, A., Desalle, R., Narechania, A., Wilson Sayres, M.A., Garrison, E.P., Handsaker, R.E., Kashin, S., McCarroll, S.A., Rodriguez-Flores, J.L., Flicek, P., Clarke, L., Zheng-Bradley, X., Erlich, Y., Gymrek, M., Frederick Willems, T., Bustamante, C.D., Mendez, F.L., David Poznik, G., Underhill, P.A., Lee, C., Cerveira, E., Malhotra, A., Romanovitch, M., Zhang, C., Abecasis, G.R., Coin, L., Shao, H., Mittelman, D., Tyler-Smith, C., Ayub, Q., Banerjee, R., Cerezo, M., Chen, Y., Fitzgerald, T.W., Louzada, S., Massaia, A., McCarthy, S., Ritchie, G.R., Xue, Y., Yang, F., Gibbs, R.A., Kovar, C., Kalra, D., Hale, W., Muzny, D., Reid, J.G., Wang, J., Dan, X., Guo, X., Li, G., Li, Yingrui, Ye, C., Zheng, X., Altshuler, D.M., Flicek, P., Clarke, L., Zheng-Bradley, X., Bentley, D.R., Cox, A., Humphray, S., Kahn, S., Sudbrak, R., Albrecht, M.W., Lienhard, M., Larson, D., Craig, D.W., Izatt, T., Kurdoglu, A.A., Sherry, S.T., Xiao, C.,

Haussler, D., Abecasis, G.R., McVean, G.A., Durbin, R.M., Balasubramaniam, S., Keane, T.M., McCarthy, S., Stalker, J., Bodmer, W., Bedoya, G., Ruiz-Linares, A., Cai, Z., Gao, Y., Chu, J., Peltonen, L., Garcia-Montero, A., Orfao, A., Dutil, J., Martinez-Cruzado, J.C., Oleksyk, T.K., Barnes, K.C., Mathias, R.A., Hennis, A., Watson, H., McKenzie, C., Qadri, F., LaRocque, R., Sabeti, P.C., Zhu, J., Deng, X., Sabeti, P.C., Asogun, D., Folarin, O., Happi, C., Omoniwa, O., Stremlau, M., Tariyal, R., Jallow, M., Sisay Joof, F., Corrah, T., Rockett, K., Kwiatkowski, D., Kooner, J., Tịnh Hiê`n, T., Dunstan, S.J., Thuy Hang, N., Fonnie, R., Garry, R., Kanneh, L., Moses, L., Sabeti, P.C., Schieffelin, J., Grant, D.S., Gallo, C., Poletti, G., Saleheen, D., Rasheed, A., 2015. A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393

Gicquel, C., El-Osta, A., Le Bouc, Y., 2008. Epigenetic regulation and fetal programming. Best Practice and Research in Clinical Endocrinology and Metabolism. https://doi.org/10.1016/j.beem.2007.07.009

Gittelman, R.M., Hun, E., Ay, F., Madeoy, J., Pennacchio, L., Noble, W.S., Hawkins, R.D., and Akey, J.M. (2015). Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Res*.

Glasson, S.S., Chambers, M.G., Van Den Berg, W.B., and Little, C.B. (2010). The OARSI histopathology initiative - recommendations for histological assessments of osteoarthritis in the mouse. *Osteoarthr. Cartil.* 18 Suppl 3, S17-23.

Godfrey, K.M., Lillycrop, K.A., Burdge, G.C., Gluckman, P.D., Hanson, M.A., 2013. Non-imprinted epigenetics in fetal and postnatal development and growth, in: Nestle Nutrition Institute Workshop Series. S. Karger AG, pp. 57–63. https://doi.org/10.1159/000342552

Goh, C.S., Cohen, F.E., 2002. Co-evolutionary Analysis Reveals Insights into Protein–Protein Interactions. Journal of Molecular Biology 324, 177–192. https://doi.org/10.1016/S0022-2836(02)01038-0

Gokhman, D., Agranat-Tamir, L., Housman, G., Nissim-Rafinia, M., Nieves-Colón, M., Gu, H., Ferrando-Bernal, M., Gelabert, P., Lipende, I., Bontrop, R., et al. (2017). Recent Regulatory Changes Shaped Human Facial and Vocal Anatomy. *BioRxiv* 106955.

Gorbunova, V., Seluanov, A., Mita, P., McKerrow, W., Fenyö, D., Boeke, J.D., Linker, S.B., Gage, F.H., Kreiling, J.A., Petrashen, A.P., Woodham, T.A., Taylor, J.R., Helfand, S.L., Sedivy, J.M., 2021. The role

of retrotransposable elements in aging and age-associated diseases. Nature 596, 43. https://doi.org/10.1038/S41586-021-03542-Y

Gray, D.J., and Gardner, E. (1950). Prenatal development of the human knee and superior tibiofibular joints. *Am. J. Anat.* 86, 235–287.

Greer, E.L., Brunet, A., 2005. FOXO transcription factors at the interface between longevity and tumor suppression. Oncogene 2005 24:50 24, 7410–7425. https://doi.org/10.1038/sj.onc.1209086

Grossman, S.R., Andersen, K.G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D.J., Griesemer, D., Karlsson, E.K., Wong, S.H., et al. (2013). Identifying recent adaptations in large-scale genomic data. *Cell* 152, 703–713.

GTEx Consortium, T.Gte., 2013. The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580–5. https://doi.org/10.1038/ng.2653

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849.

Guenther, C., Pantalena-Filho, L., and Kingsley, D.M. (2008). Shaping skeletal growth by modular regulatory elements in the Bmp5 gene. *PLoS Genet.* 4, e1000308.

Guo, M., Liu, Z., Willen, J., Shaw, C.P., Richard, D., Jagoda, E., Doxey, A.C., Hirschhorn, J., and Capellini, T.D. (2017). Epigenetic profiling of growth plate chondrocytes sheds insight into regulatory genetic variation influencing height. *Elife* 6, e29329.

Gurven, M., Gomes, C., 2017. Mortality, Senescence, and Lifespan, in: Chimpanzees and Human Evolution. pp. 181–216.

Gurven, M., Kaplan, H., 2007. Longevity Among Hunter- Gatherers: A Cross-Cultural Examination. Population and Development Review 33, 321–365. https://doi.org/10.1111/j.1728-4457.2007.00171.x

Gurven, M., Kaplan, H., 2008. Beyond the grandmother hypothesis: evolutionary models of human longevity., in: The Cultural Context of Aging: Worldwide Perspectives. pp. 53–66.

Hall, J.A., Dominy, J.E., Lee, Y., Puigserver, P., 2013. The sirtuin family's role in aging and age-associated pathologies. Journal of Clinical Investigation. https://doi.org/10.1172/JCI64094

Hall, M.D., Murray, C.A., Valdez, M.J., and Perantoni, A.O. (2017). Mesoderm-specific Stat3 deletion affects expression of Sox9 yielding Sox9-dependent phenotypes. *PLoS Genet.* 13, e1006610.

Hamilton, W.D., 1966. The moulding of senescence by natural selection. J Theor Biol 12, 12–45.

Hammer, M.L.A., Foley, R.A., 1996. Longevity and life history in hominid evolution. Human Evolution 11, 61–66. https://doi.org/10.1007/BF02456989

Hanson, M., Godfrey, K.M., Lillycrop, K.A., Burdge, G.C., Gluckman, P.D., 2011. Developmental plasticity and developmental origins of non-communicable disease: Theoretical considerations and epigenetic mechanisms. Progress in Biophysics and Molecular Biology. https://doi.org/10.1016/j.pbiomolbio.2010.12.008

Hartfield, M., and Otto, S.P. (2011). Recombination and hitchhiking of deleterious alleles. *Evolution* (N. Y). 65, 2421–2434.

Hashemi, J., Chandrashekar, N., Gill, B., Beynnon, B.D., Slauterbeck, J.R., Schutt, R.C., Mansouri, H., and Dabezies, E. (2008). The geometry of the tibial plateau and its influence on the biomechanics of the tibiofemoral joint. *J. Bone Joint Surg. Am*. 90, 2724–2734.

Hawkes, K., O'Connell, J.F., Jones, N.G., Alvarez, H., Charnov, E.L., Pusey, A., Stoinski, T.S., Strier, K.B., Morris, W.F., Bronikowski, A.M., 1998. Grandmothering, menopause, and the evolution of human life histories. Proc Natl Acad Sci U S A 95, 1336–9. https://doi.org/10.1073/pnas.95.3.1336

He, C., Kraft, P., Chen, C., Buring, J.E., Paré, G., Hankinson, S.E., Chanock, S.J., Ridker, P.M., Hunter, D.J., Chasman, D.I., 2009. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. Nature Genetics 41, 724–728. https://doi.org/10.1038/ng.385

He, Y., Tsuei, J., and Wan, Y.-J.Y. (2014). Biological functional annotation of retinoic acid alpha and beta in mouse liver based on genome-wide binding. *Am. J. Physiol. Liver Physiol*. 307, G205–G218.

Hedman, A.C., Smith, J.M., and Sacks, D.B. (2015). The biology of IQGAP proteins: beyond the cytoskeleton. *EMBO Rep.* 16, 427–446.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.

Heredia, J.D., Park, J., Brubaker, R.J., Szymanski, S.K., Gill, K.S., Procko, E., 2018. Mapping Interaction Sites on Human Chemokine Receptors by Deep Mutational Scanning. J Immunol 200, 3825–3839. https://doi.org/10.4049/JIMMUNOL.1800343

Herlofsen, S.R., Bryne, J.C., Høiby, T., Wang, L., Issner, R., Zhang, X., Coyne, M.J., Boyle, P., Gu, H., Meza-Zepeda, L.A., et al. (2013). Genome-wide map of quantified epigenetic changes during in vitro chondrogenic differentiation of primary human mesenchymal stem cells. *BMC Genomics*, 14(1), 105.

Hijazi, A. (2018). Investigating the role of CCCTC-binding factor in osteoarthritis pathogenesis. *Osteoarthr. Cartil*. 26, S154.

Hormozdiari, F., Gazal, S., van de Geijn, B., Finucane, H.K., Ju, C.J.-T., Loh, P.-R., Schoech, A., Reshef, Y., Liu, X., O'Connor, L., et al. (2018). Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat. Genet*. 50, 1041–1047.

Horvath, S., 2013. DNA methylation age of human tissues and cell types. Genome Biology 14. https://doi.org/10.1186/gb-2013-14-10-r115

Horvath, S., Haghani, A., Macoretta, N., Ablaeva, J., Zoller, J.A., Li, C.Z., Zhang, J., Takasugi, M., Zhao, Y., Rydkina, E., Zhang, Z., Emmrich, S., Raj, K., Seluanov, A., Faulkes, C.G., Gorbunova, V., 2021a. DNA methylation clocks show slower progression of aging in naked mole-rat queens. bioRxiv 2021.03.15.435536. https://doi.org/10.1101/2021.03.15.435536

Horvath, S., Raj, K., 2018a. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nature Reviews Genetics. https://doi.org/10.1038/s41576-018-0004-3

Horvath, S., Raj, K., 2018b. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nature Reviews Genetics 2018 19:6 19, 371–384. https://doi.org/10.1038/s41576-018-0004-3

Horvath, S., Zoller, J.A., Haghani, A., Jasinska, A.J., Raj, K., Breeze, C.E., Ernst, J., Vaughan, K.L., Mattison, J.A., 2021b. Epigenetic clock and methylation studies in the rhesus macaque. Geroscience 43, 2441–2453. https://doi.org/10.1007/S11357-021-00429-8/FIGURES/4

Houard, X., Goldring, M.B., and Berenbaum, F. (2013). Homeostatic mechanisms in articular cartilage and role of inflammation in osteoarthritis. *Curr. Rheumatol. Rep.* 15, 375.

Howard, O.M.Z., Shirakawa, A.K., Turpin, J.A., Maynard, A., Tobin, G.J., Carrington, M., Oppenheim, J.J., Dean, M., 1999. Naturally Occurring CCR5 Extracellular and Transmembrane Domain Variants

Affect HIV-1 Co-receptor and Ligand Binding Function *. Journal of Biological Chemistry 274, 16228–16234. https://doi.org/10.1074/JBC.274.23.16228

Howell, S.M., Howell, S.J., and Hull, M.L. (2010). Assessment of the radii of the medial and lateral femoral condyles in varus and valgus knees with osteoarthritis. *J. Bone Joint Surg. Am*. 92, 98–104.

Huang, M., Kim, J.M., Shiotani, B., Yang, K., Zou, L., D'Andrea, A.D., 2010. The FANCM/FAAP24 Complex is Required for the DNA Inter-strand Crosslink-Induced Checkpoint Response. Mol Cell 39, 259. https://doi.org/10.1016/J.MOLCEL.2010.07.005

Huang, S., Lee, L., Hanson, N.B., Lenaerts, C., Hoehn, H., Poot, M., Rubin, C.D., Chen, D.F., Yang, C.C., Juch, H., Dorn, T., Spiegel, R., Oral, E.A., Abid, M., Battisti, C., Lucci-Cordisco, E., Neri, G., Steed, E.H., Kidd, A., Isley, W., Showalter, D., Vittone, J.L., Konstantinow, A., Ring, J., Meyer, P., Wenger, S.L., Von Herbay, A., Wollina, U., Schuelke, M., Huizenga, C.R., Leistritz, D.F., Martin, G.M., Mian, I.S., Oshima, J., 2006. The spectrum of WRN mutations in Werner syndrome patients. Human Mutation 27, 558–567. https://doi.org/10.1002/HUMU.20337

Huang, X., Zheng, W., Pearce, R., Zhang, Y., Zhang, Y., 2020. SSIPe: accurately estimating protein-protein binding affinity change upon mutations using evolutionary profiles in combination with an optimized physical energy function. Bioinformatics 36, 2429–2437. https://doi.org/10.1093/BIOINFORMATICS/BTZ926

Huang, Y.F., Gulko, B., Siepel, A., 2017. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. Nature Genetics 49, 618–624. https://doi.org/10.1038/ng.3810

Hubisz, M.J., Pollard, K.S., 2014. Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. Current Opinion in Genetics and Development. https://doi.org/10.1016/j.gde.2014.07.005

Hubisz, M.J., Pollard, K.S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform*. 12, 41–51.

Hublin, J.-J., Neubauer, S., Gunz, P., 2015. Brain ontogeny and life history in Pleistocene hominins. Philosophical Transactions of the Royal Society B: Biological Sciences 370, 20140062–20140062. https://doi.org/10.1098/rstb.2014.0062

Hujoel, M.L.A., Gazal, S., Hormozdiari, F., van de Geijn, B., Price, A.L., 2019. Disease Heritability Enrichment of Regulatory Elements Is Concentrated in Elements with Ancient Sequence Age and Conserved Function across Species. American Journal of Human Genetics 104, 611–624. https://doi.org/10.1016/j.ajhg.2019.02.008

Hume, M.A., Barrera, L.A., Gisselbrecht, S.S., and Bulyk, M.L. (2015). UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein–DNA interactions. *Nucleic Acids Res*. 43, D117–D122.

Hunter, D., Nevitt, M., Lynch, J., Kraus, V.B., Katz, J.N., Collins, J.E., Bowes, M., Guermazi, A., Roemer, F.W., Losina, E., et al. (2016). Longitudinal validation of periarticular bone area and 3D shape as biomarkers for knee OA progression? Data from the FNIH OA Biomarkers Consortium. *Ann. Rheum. Dis*. 75, 1607–1614.

Hunter, D.J., and Bierma-Zeinstra, S. (2019). Osteoarthritis. *Lancet* (London, England) 393, 1745–1759.

Indjeian, V.B., Kingman, G.A., Jones, F.C., Guenther, C.A., Grimwood, J., Schmutz, J., Myers, R.M., and Kingsley, D.M. (2016). Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. *Cell* 164, 45–56.

Infante, C.R., Park, S., Mihala, A.G., Kingsley, D.M., and Menke, D.B. (2013). Pitx1 broadly associates with limb enhancers and is enriched on hindlimb cis-regulatory elements. *Dev. Biol*. 374, 234–244.

Isler, K., van Schaik, C.P., 2012. How Our Ancestors Broke through the Gray Ceiling. Current Anthropology 53, S453–S465. https://doi.org/10.1086/667623

Jack, B.R., Meyer, A.G., Echave, J., Wilke, C.O., 2016. Functional Sites Induce Long-Range Evolutionary Constraints in Enzymes. PLOS Biology 14, e1002452. https://doi.org/10.1371/JOURNAL.PBIO.1002452

Jackman, S. (2017). {pscl}: Classes and Methods for {R} Developed in the Political Science Computational Laboratory.

Jagoda, E., Lawson, D.J., Wall, J.D., Lambert, D., Muller, C., Westaway, M., Leavesley, M., Capellini, T.D., Mirazón Lahr, M., Gerbault, P., et al. (2017). Disentangling Immediate Adaptive Introgression from Selection on Standing Introgressed Variation in Humans. *Mol. Biol. Evol.* 35, 623–630.

Jankauskaite, J., Jiménez-García, B., Dapkunas, J., Fernández-Recio, J., Moal, I.H., 2019. SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. Bioinformatics 35, 462–469. https://doi.org/10.1093/BIOINFORMATICS/BTY635

Jobson, R.W., Nabholz, B., Galtier, N., 2010. An Evolutionary Genome Scan for Longevity-Related Natural Selection in Mammals. Molecular Biology and Evolution 27, 840–847. https://doi.org/10.1093/MOLBEV/MSP293

Jones, O.R., Scheuerlein, A., Salguero-Gómez, R., Camarda, C.G., Schaible, R., Casper, B.B., Dahlgren, J.P., Ehrlén, J., García, M.B., Menges, E.S., Quintana-Ascencio, P.F., Caswell, H., Baudisch, A., Vaupel, J.W., 2014. Diversity of ageing across the tree of life. Nature 505, 169–173. https://doi.org/10.1038/nature12789

Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., Harper, W., Hemenez, M., Ponnusamy, R., Salehi, A., Sanagavarapu, B.A., Spallino, E., Aaron, K.A., Concepcion, W., Gardner, J.M., Kelly, B., Neidlinger, N., Wang, Z., Crasta, S., Kolluru, S., Morri, M., Tan, S.Y., Travaglini, K.J., Xu, C., Alcántara-Hernández, M., Almanzar, N., Antony, J., Beyersdorf, B., Burhan, D., Calcuttawala, K., Carter, M.M., Chan, C.K.F., Chang, C.A., Chang, S., Colville, A., Culver, R.N., Cvijović, I., D'Amato, G., Ezran, C., Galdos, F.X., Gillich, A., Goodyer, W.R., Hang, Y., Hayashi, A., Houshdaran, S., Huang, X., Irwin, J.C., Jang, S.R., Juanico, J.V., Kershner, A.M., Kim, S., Kiss, B., Kong, W., Kumar, M.E., Kuo, A.H., Leylek, R., Li, B., Loeb, G.B., Lu, W.J., Mantri, S., Markovic, M., McAlpine, P.L., de Morree, A., Mrouj, K., Mukherjee, S., Muser, T., Neuhöfer, P., Nguyen, T.D., Perez, K., Phansalkar, R., Puluca, N., Qi, Z., Rao, P., Raquer-McKay, H., Schaum, N., Scott, B., Seddighzadeh, B., Segal, J., Sen, S., Sikandar, S., Spencer, S.P., Steffes, L.C., Subramaniam, V.R., Swarup, A., Swift, M., van Treuren, W., Trimm, E., Veizades, S., Vijayakumar, S., Vo, K.C., Vorperian, S.K., Wang, W., Weinstein, H.N.W., Winkler, J., Wu, T.T.H., Xie, J., Yung, A.R., Zhang, Y., Detweiler, A.M., Mekonen, H., Neff, N.F., Sit, R. v., Tan, M., Yan, J., Bean, G.R., Charu, V., Forgó, E., Martin, B.A., Ozawa, M.G., Silva, O., Toland, A., Vemuri, V.N.P., Afik, S., Awayan, K., Botvinnik, O.B., Byrne, A., Chen, M., Dehghannasiri, R., Detweiler, A.M., Gayoso, A., Granados, A.A., Li, Q., Mahmoudabadi, G., McGeever, A., Olivieri, J.E., Park, M., Ravikumar, N., Stanley, G., Tan, M., Tan, W., Tarashansky, A.J., Vanheusden, R., Vorperian, S.K., Wang, P., Wang, S., Xing, G.,

Dethlefsen, L., Ho, P.Y., Liu, S., Maltzman, J.S., Metzger, R.J., Perez, K., Raquer-McKay, H., Sasagawa, K., Sinha, R., Song, H., Swarup, A., Vijayakumar, S., Wang, B., Yung, A.R., Artandi, S.E., Beachy, P.A., Clarke, M.F., Giudice, L.C., Huang, F.W., Huang, K.C., Idoyaga, J., Kim, S.K., Kuo, C.S., Nguyen, P., Rando, T.A., Red-Horse, K., Reiter, J., Relman, D.A., Sonnenburg, J.L., Wu, A., Wu, S.M., Wyss-Coray, T., 2022. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science (1979) 376. https://doi.org/10.1126/SCIENCE.ABL4896/SUPPL_FILE/SCIENCE.ABL4896_MDAR_REPRODUCIBILITY_CHECKLIST.PDF

Jubb, H.C., Pandurangan, A.P., Turner, M.A., Ochoa-Montaño, B., Blundell, T.L., Ascher, D.B., 2017a. Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. Progress in Biophysics and Molecular Biology 128, 3–13. https://doi.org/10.1016/J.PBIOMOLBIO.2016.10.002

Jubb, H.C., Pandurangan, A.P., Turner, M.A., Ochoa-Montaño, B., Blundell, T.L., Ascher, D.B., 2017b. Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. Progress in Biophysics and Molecular Biology 128, 3–13. https://doi.org/10.1016/J.PBIOMOLBIO.2016.10.002

Judge, D.S., Carey, J.R., 2000. Postreproductive Life Predicted by Primate Patterns. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 55, B201–B209. https://doi.org/10.1093/gerona/55.4.B201

Jung, I., Schmitt, A., Diao, Y., Lee, A.J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., Chiang, Z., Kim, C., Masliah, E., Barr, C.L., Li, B., Kuan, S., Kim, D., Ren, B., 2019. A compendium of promoter-centered long-range chromatin interactions in the human genome. Nature Genetics 51, 1442–1449. https://doi.org/10.1038/s41588-019-0494-8

Jungers, W.L. (1988). Relative joint size and hominoid locomotor adaptations with implications for the evolution of hominid bipedalism. *J. Hum. Evol*. 17, 247–265.

Junnila, R.K., List, E.O., Berryman, D.E., Murrey, J.W., Kopchick, J.J., 2013. The GH/IGF-1 axis in ageing and longevity. Nat Rev Endocrinol 9, 366–376. https://doi.org/10.1038/NRENDO.2013.67

Kanapuru, B., Ershler, W.B., 2009. Inflammation, Coagulation, and the Pathway to Frailty. Am J Med 122, 605. https://doi.org/10.1016/J.AMJMED.2009.01.030

Kannu, P., Bateman, J.F., Belluoccio, D., Fosang, A.J., and Savarirayan, R. (2009). Employing molecular genetics of chondrodysplasias to inform the study of osteoarthritis. *Arthritis Rheum*. 60, 325–334.

Kanton, S., Boyle, M.J., He, Z., Santel, M., Weigert, A., Sanchís-Calleja, F., Guijarro, P., Sidow, L., Fleck, J.S., Han, D., et al. (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* 574, 418–422.

Kaplan, Hillard, Lancaster, Jane, Robson, A., 2003. Embodied capital and the evolutionary economics of the human life span. Population and Development Review 152–182.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Aguilar Salinas, C.A., Ahmad, T., Albert, C.M., Ardissino, D., Atzmon, G., Barnard, J., Beaugerie, L., Benjamin, E.J., Boehnke, M., Bonnycastle, L.L., Bottinger, E.P., Bowden, D.W., Bown, M.J., Chambers, J.C., Chan, J.C., Chasman, D., Cho, J., Chung, M.K., Cohen, B., Correa, A., Dabelea, D., Daly, M.J., Darbar, D., Duggirala, R., Dupuis, J., Ellinor, P.T., Elosua, R., Erdmann, J., Esko, T., Färkkilä, M., Florez, J., Franke, A., Getz, G., Glaser, B., Glatt, S.J., Goldstein, D., Gonzalez, C., Groop, L., Haiman, C., Hanis, C., Harms, M., Hiltunen, M., Holi, M.M., Hultman, C.M., Kallela, M., Kaprio, J., Kathiresan, S., Kim, B.J., Kim, Y.J., Kirov, G., Kooner, J., Koskinen, S., Krumholz, H.M., Kugathasan, S., Kwak, S.H., Laakso, M., Lehtimäki, T., Loos, R.J.F., Lubitz, S.A., Ma, R.C.W., MacArthur, D.G., Marrugat, J., Mattila, K.M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Meigs, J.B., Melander, O., Metspalu, A., Neale, B.M., Nilsson, P.M.,

O'Donovan, M.C., Ongur, D., Orozco, L., Owen, M.J., Palmer, C.N.A., Palotie, A., Park, K.S., Pato, C., Pulver, A.E., Rahman, N., Remes, A.M., Rioux, J.D., Ripatti, S., Roden, D.M., Saleheen, D., Salomaa, V., Samani, N.J., Scharf, J., Schunkert, H., Shoemaker, M.B., Sklar, P., Soininen, H., Sokol, H., Spector, T., Sullivan, P.F., Suvisaari, J., Tai, E.S., Teo, Y.Y., Tiinamaija, T., Tsuang, M., Turner, D., Tusie-Luna, T., Vartiainen, E., Watkins, H., Weersma, R.K., Wessman, M., Wilson, J.G., Xavier, R.J., Neale, B.M., Daly, M.J., MacArthur, D.G., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020 581:7809 581, 434–443. https://doi.org/10.1038/s41586-020-2308-7

Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2013). The UCSC genome browser database: 2014 update. *Nucleic Acids Res*. 42, D764--D770.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*. 32, 493D – 496.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., Kent, W.J., 2004. The UCSC Table Browser data retrieval tool. Nucleic Acids Research 32, 493D – 496. https://doi.org/10.1093/nar/gkh103

Kassambara, A., and Mundt, F. (2017). Factoextra: extract and visualize the results of multivariate data analyses. R package version 1.0. 4. 2017.

Kastenhuber, E.R., Lowe, S.W., 2017. Putting p53 in Context. Cell 170, 1062–1078. https://doi.org/10.1016/J.CELL.2017.08.028

Kato, S., Han, S.Y., Liu, W., Otsuka, K., Shibata, H., Kanamaru, R., Ishioka, C., 2003. Understanding the function–structure and function–mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proceedings of the National Academy of Sciences 100, 8424–8429. https://doi.org/10.1073/PNAS.1431692100

Kavathia, N., Jain, A., Walston, J., Beamer, B.A., Fedarko, N.S., 2009. Serum markers of apoptosis decrease with age and cancer stage. Aging 1, 652–663. https://doi.org/10.18632/AGING.100069

Kaya, A., Lobanov, A. v., Gladyshev, V.N., 2015. Evidence that mutation accumulation does not cause aging in Saccharomyces cerevisiae. Aging Cell 14, 366–371. https://doi.org/10.1111/ACEL.12290

Keane, M., Semeiks, J., Webb, A.E., Li, Y.I., Quesada, V., Craig, T., Madsen, L.B., van Dam, S., Brawand, D., Marques, P.I., Michalak, P., Kang, L., Bhak, J., Yim, H.S., Grishin, N. V., Nielsen, N.H., Heide-Jørgensen, M.P., Oziolor, E.M., Matson, C.W., Church, G.M., Stuart, G.W., Patton, J.C., George, J.C., Suydam, R., Larsen, K., López-Otín, C., O'Connell, M.J., Bickham, J.W., Thomsen, B., deMagalhães, J.P., 2015. Insights into the Evolution of Longevity from the Bowhead Whale Genome. Cell Reports 10, 112. https://doi.org/10.1016/J.CELREP.2014.12.008

Keller, M.C., Miller, G., 2006. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? Behav Brain Sci 29, 385–404. https://doi.org/10.1017/S0140525X06009095

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006.

Khurana, S., Kennedy, M., King, L.R., Golding, H., 2005. Identification of a Linear Peptide Recognized by Monoclonal Antibody 2D7 Capable of Generating CCR5-Specific Antibodies with Human Immunodeficiency Virus-Neutralizing Activity. Journal of Virology 79, 6791–6800. https://doi.org/10.1128/JVI.79.11.6791-6800.2005/ASSET/1BF567FF-4319-4574-853B-C979E38FBB76/ASSETS/GRAPHIC/ZJV0110562930011.JPEG

Kiefer, J.C., 2007. Epigenetics in development. Developmental Dynamics 236, 1144–1156. https://doi.org/10.1002/DVDY.21094

King, M.-C., and Wilson, A.C. (1975). Evolution at Two Levels in Humans and Chimpanzees. *Science* (80). 188, 107–116.

Kirkland, J.L., Peterson, C., 2009. Healthspan, Translation, and New Outcomes for Animal Studies of Aging. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 64A, 209–212. https://doi.org/10.1093/gerona/gln063

Kirkwood, T.B.L., 1997. The origins of human ageing. Philosophical Transactions of the Royal Society B: Biological Sciences 352, 1765–1772. https://doi.org/10.1098/rstb.1997.0160

Kirkwood, T.B.L., 2002. Evolution of ageing. Mechanisms of Ageing and Development 123, 737–745. https://doi.org/10.1016/S0047-6374(01)00419-5

Kirkwood, T.B.L., 2008. Understanding ageing from an evolutionary perspective, in: Journal of Internal Medicine. J Intern Med, pp. 117–127. https://doi.org/10.1111/j.1365-2796.2007.01901.x

Kirkwood, T.B.L., Holliday, R., 1979. The evolution of ageing and longevity. Proceedings of the Royal Society of London - Biological Sciences 205, 531–546. https://doi.org/10.1098/rspb.1979.0083

Kirkwood, T.B.L., Rose, M.R., 1991. Evolution of senescence: late survival sacrificed for reproduction. Philosophical Transactions - Royal Society of London, B 332, 15–24. https://doi.org/10.1098/rstb.1991.0028

Klein, J.C., Keith, A., Rice, S.J., Shepherd, C., Agarwal, V., Loughlin, J., and Shendure, J. (2019). Functional testing of thousands of osteoarthritis-associated variants for regulatory activity. *Nat. Commun*. 10, 2434.

Klemm, S.L., Shipony, Z., Greenleaf, W.J., 2019. Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics. https://doi.org/10.1038/s41576-018-0089-8

Knopman, D., Boland, L.L., Mosley, T., Howard, G., Liao, D., Szklo, M., McGovern, P., Folsom, A.R., 2001. Cardiovascular risk factors and cognitive decline in middle-aged adults. Neurology 56, 42–48. https://doi.org/10.1212/WNL.56.1.42

Kokenyesi, R., Tan, L., Robbins, J.R., and Goldring, M.B. (2000). Proteoglycan production by immortalized human chondrocyte cell lines cultured under conditions that promote expression of the differentiated phenotype. *Arch. Biochem. Biophys*. 383, 79–90.

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 36, 765–771.

Kowalczyk, A., Partha, R., Clark, N.L., Chikina, M., 2020. Pan-mammalian analysis of molecular constraints underlying extended lifespan. Elife 9. https://doi.org/10.7554/ELIFE.51089

Kubben, N., Misteli, T., 2017. Shared molecular and cellular mechanisms of premature ageing and ageing-associated diseases. Nature Reviews Molecular Cell Biology. https://doi.org/10.1038/nrm.2017.68

Kujoth, C.C., Hiona, A., Pugh, T.D., Someya, S., Panzer, K., Wohlgemuth, S.E., Hofer, T., Seo, A.Y., Sullivan, R., Jobling, W.A., Morrow, J.D., Van Remmen, H., Sedivy, J.M., Yamasoba, T., Tanokura, M., Weindruch, R., Leeuwenburgh, C., Prolla, T.A., 2005. Medicine: Mitochondrial DNA mutations,

oxidative stress, and apoptosis in mammalian aging. Science (1979) 309, 481–484. https://doi.org/10.1126/SCIENCE.1112125/SUPPL_FILE/KUJOTH.SOM.PDF

Kumar, A., Butler, B.M., Kumar, S., Ozkan, S.B., 2015. Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine. Current Opinion in Structural Biology 35, 135–142. https://doi.org/10.1016/J.SBI.2015.11.002

Lähteenvuo, J., Rosenzweig, A., 2012. Effects of aging on angiogenesis. Circulation Research 110, 1252–1263. https://doi.org/10.1161/CIRCRESAHA.111.246116

Lambert, M.J., Portfors, C. V., 2017. Adaptive sequence convergence of the tumor suppressor ADAMTS9 between small-bodied mammals displaying exceptional longevity. Aging 9, 573–582. https://doi.org/10.18632/AGING.101180

Lan, L., Nakajima, S., Komatsu, K., Nussenzweig, A., Shimamoto, A., Oshima, J., Yasui, A., 2005. Accumulation of Werner protein at DNA double-strand breaks in human cells. J Cell Sci 118, 4153–4162. https://doi.org/10.1242/JCS.02544

Lanctôt, C., Moreau, A., Chamberland, M., Tremblay, M.L., and Drouin, J. (1999). Hindlimb patterning and mandible development require the Ptx1 gene. *Development* 126, 1805–1810.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J.B., Kattman, B.L., Maglott, D.R., 2018. ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Research 46, D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Law, R.H., Zhang, Q., McGowan, S., Buckle, A.M., Silverman, G.A., Wong, W., Rosado, C.J., Langendorf, C.G., Pike, R.N., Bird, P.I., Whisstock, J.C., 2006. An overview of the serpin superfamily. Genome Biology 2006 7:5 7, 1–11. https://doi.org/10.1186/GB-2006-7-5-216

Lee, A.S., Ellman, M.B., Yan, D., Kroin, J.S., Cole, B.J., van Wijnen, A.J., Im, H.J., 2013. A current review of molecular mechanisms regarding osteoarthritis and pain. Gene 527, 440–447. https://doi.org/10.1016/J.GENE.2013.05.069

Lee, B., Sharron, M., Blanpain, C., Doranz, B.J., Vakili, J., Setoh, P., Berg, E., Liu, G., Guy, H.R., Durell, S.R., Parmentier, M., Chang, C.N., Price, K., Tsang, M., Doms, R.W., 1999. Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. J Biol Chem 274, 9617–9626. https://doi.org/10.1074/JBC.274.14.9617

Lee, J.H., Kim, E.W., Croteau, D.L., Bohr, V.A., 2020. Heterochromatin: an epigenetic point of view in aging. Experimental & Molecular Medicine 2020 52:9 52, 1466–1474. https://doi.org/10.1038/S12276-020-00497-4

Lek, M., Karczewski, K.J., Minikel, E. V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., Abboud, H.E., Abecasis, G., Aguilar-Salinas, C.A., Arellano-Campos, O., Atzmon, G., Aukrust, I., Barr, C.L., Bell, G.I., Bergen, S., Bjørkhaug, L., Blangero, J., Bowden, D.W., Budman, C.L., Burtt, N.P., Centeno-Cruz, F., Chambers, J.C., Chambert, K., Clarke, R., Collins, R., Coppola, G., Córdova, E.J., Cortes, M.L., Cox, N.J., Duggirala, R., Farrall, M., Fernandez-Lopez, J.C., Fontanillas, P., Frayling, T.M., Freimer, N.B., Fuchsberger, C., García-Ortiz, H., Goel, A., Gómez-Vázquez, M.J., González-Villalpando, M.E., González-Villalpando, C., Grados, M.A., Groop, L., Haiman, C.A., Hanis, C.L., Hattersley, A.T., Henderson, B.E., Hopewell, J.C., Huerta-Chagoya, A., Islas-Andrade, S., Jacobs, S.B., Jalilzadeh, S., Jenkinson, C.P., Moran, J., Jiménez-Morale, S., Kähler, A., King, R.A., Kirov, G., Kooner, J.S., Kyriakou, T., Lee, J.Y., Lehman, D.M., Lyon, G., MacMahon, W., Magnusson, P.K., Mahajan, A.,

Marrugat, J., Martínez-Hernández, A., Mathews, C.A., McVean, G., Meigs, J.B., Meitinger, T., Mendoza-Caamal, E., Mercader, J.M., Mohlke, K.L., Moreno-Macías, H., Morris, A.P., Najmi, L.A., Njølstad, P.R., O'Donovan, M.C., Ordóñez-Sánchez, M.L., Owen, M.J., Park, T., Pauls, D.L., Posthuma, D., Revilla-Monsalve, C., Riba, L., Ripke, S., Rodríguez-Guillén, R., Rodríguez-Torres, M., Sandor, P., Seielstad, M., Sladek, R., Soberón, X., Spector, T.D., Tai, S.E., Teslovich, T.M., Walford, G., Wilkens, L.R., Williams, A.L., 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016 536:7616 536, 285–291. https://doi.org/10.1038/nature19057

Lemaître, J.F., Berger, V., Bonenfant, C., Douhard, M., Gamelon, M., Plard, F., Gaillard, J.M., 2015. Early-late life trade-offs and the evolution of ageing in the wild. Proceedings of the Royal Society B: Biological Sciences 282. https://doi.org/10.1098/RSPB.2015.0209

Levy, E.D., 2010. A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution. Journal of Molecular Biology 403, 660–670. https://doi.org/10.1016/J.JMB.2010.09.028

Lewis, K.N., Soifer, I., Melamud, E., Roy, M., McIsaac, R.S., Hibbs, M., Buffenstein, R., 2016. Unraveling the message: insights into comparative genomics of the naked mole-rat. Mammalian Genome 27, 259. https://doi.org/10.1007/S00335-016-9648-5

Li, H. (2011a). Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* 27, 718–719.

Li, H. (2011b). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993.

Li, H., 2011. Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics 27, 718–9. https://doi.org/10.1093/bioinformatics/btq671

Li, P., Nijhawan, D., Budihardjo, I., Srinivasula, S.M., Ahmad, M., Alnemri, E.S., Wang, X., 1997. Cytochrome c and dATP-Dependent Formation of Apaf-1/Caspase-9 Complex Initiates an Apoptotic Protease Cascade. Cell 91, 479–489. https://doi.org/10.1016/S0092-8674(00)80434-1

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat*. 5, 1752–1779.

Li, Y., de Magalhães, J.P., 2011. Accelerated protein evolution analysis reveals genes and pathways associated with the evolution of mammalian longevity. AGE 2011 35:2 35, 301–314. https://doi.org/10.1007/S11357-011-9361-Y

Lieberman, D.E., Kistner, T.M., Richard, D., Lee, I.M., Baggish, A.L., 2021. The active grandparent hypothesis: Physical activity and the evolution of extended human healthspans and lifespans. Proc Natl Acad Sci U S A 118. https://doi.org/10.1073/PNAS.2107621118/SUPPL_FILE/PNAS.2107621118.SAPP.PDF

Lindenbaum, P., 2015. JVarkit: java-based utilities for Bioinformatics. Figshare. https://doi.org/10.6084/M9.FIGSHARE.1425030.V1

Liu, C.-F., Samsa, W.E., Zhou, G., and Lefebvre, V. (2017). Transcriptional control of chondrocyte specification and differentiation. *Semin. Cell Dev. Biol*. 62, 34–49.

Liu, H., Zhang, B., Sun, Z., 2020. Spectrum of EGFR aberrations and potential clinical implications: insights from integrative pan-cancer analysis. Cancer Communications 40, 43–59. https://doi.org/10.1002/CAC2.12005

Liu, X., Li, C., Mou, C., Dong, Y., Tu, Y., 2020. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. Genome Medicine 12, 1–8. https://doi.org/10.1186/S13073-020-00803-9/FIGURES/4

Liu, Y., Chang, J.C., Hon, C.C., Fukui, N., Tanaka, N., Zhang, Z., Lee, M.T.M., Minoda, A., 2018. Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. Scientific Reports 8, 1–13. https://doi.org/10.1038/s41598-018-33779-z

Liu, Y., Tong, Q., Zhou, Y., Lee, H.W., Yang, J.J., Bühring, H.J., Chen, Y.T., Ha, B., Chen, C.X.J., Yang, Y., Zen, K., 2007. Functional Elements on SIRPα IgV domain Mediate Cell Surface Binding to CD47. J Mol Biol 365, 680. https://doi.org/10.1016/J.JMB.2006.09.079

Loeser, R.F., Collins, J.A., Diekman, B.O., 2016. Ageing and the pathogenesis of osteoarthritis. Nat Rev Rheumatol 12, 412–20. https://doi.org/10.1038/nrrheum.2016.65

Logtenberg, M.E.W., Scheeren, F.A., Schumacher, T.N., 2020. The CD47-SIRPα Immune Checkpoint. Immunity 52, 742–752. https://doi.org/10.1016/J.IMMUNI.2020.04.011

Long, H.K., Prescott, S.L., Wysocka, J., 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell. https://doi.org/10.1016/j.cell.2016.09.018

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., Kroemer, G., 2013. The Hallmarks of Aging. Cell 153, 1194–1217. https://doi.org/10.1016/J.CELL.2013.05.039

Lopez-Pulido, E.I., Muñoz-Valle, J.F., Del Toro-Arreola, S., Jave-Suárez, L.F., Bueno-Topete, M.R., Estrada-Chávez, C., Pereira-Suárez, A.L., 2013. High expression of prolactin receptor is associated with cell survival in cervical cancer cells. Cancer Cell International 13, 1–9. https://doi.org/10.1186/1475-2867-13-103/FIGURES/4

Loughlin, J. (2015). Genetic contribution to osteoarthritis development: current state of evidence. *Curr. Opin. Rheumatol*. 27, 284–288.

Lowdon, R.F., Jang, H.S., Wang, T., 2016. Evolution of Epigenetic Regulation in Vertebrate Genomes. Trends in Genetics 32, 269–283. https://doi.org/10.1016/J.TIG.2016.03.001

Lowe, S.W., Lin, A.W., 2000. Apoptosis in cancer. Carcinogenesis 21, 485–495. https://doi.org/10.1093/CARCIN/21.3.485

Lowenstine, L.J., McManamon, R., and Terio, K.A. (2016). Comparative Pathology of Aging Great Apes: Bonobos, Chimpanzees, Gorillas, and Orangutans. *Vet. Pathol*. 53, 250–276.

Lyons, K.M., and Rosen, V. (2019). BMPs, TGFβ, and border security at the interzone. *Curr. Top. Dev. Biol*. 133, 153–170.

Maas, S.A., and Fallon, J.F. (2005). Single base pair change in the long-rangeSonic hedgehog limb-specific enhancer is a genetic basis for preaxial polydactyly. *Dev. Dyn*. 232, 345–348.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., Barrett, J.C., Biesecker, L.G., Conrad, D.F., Cooper, G.M., Cox, N.J., Daly, M.J., Gerstein, M.B., Goldstein, D.B., Hirschhorn, J.N., Leal, S.M., Pennacchio, L.A., Stamatoyannopoulos, J.A., Sunyaev, S.R., Valle, D., Voight, B.F., Winckler, W., Gunter, C., 2014. Guidelines for investigating causality of sequence variants in human disease. Nature. https://doi.org/10.1038/nature13127

MacRae, S.L., Croken, M.M.K., Calder, R.B., Aliper, A., Milholland, B., White, R.R., Zhavoronkov, A., Gladyshev, V.N., Seluanov, A., Gorbunova, V., Zhang, Z.D., Vijg, J., 2015. DNA repair in species with extreme lifespan differences. Aging 7, 1171–1184. https://doi.org/10.18632/aging.100866

Macrini, T.E., Coan, H.B., Levine, S.M., Lerma, T., Saks, C.D., Araujo, D.J., Bredbenner, T.L., Coutts, R.D., Nicolella, D.P., and Havill, L.M. (2013). Reproductive status and sex show strong effects on knee OA in a baboon model. *Osteoarthr. Cartil*. 21, 839–848.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2018). cluster: Cluster Analysis Basics and Extensions.

Magalhães, J.P. de, Toussaint, O., 2002. The evolution of mammalian aging. Experimental Gerontology 37, 769–775. https://doi.org/10.1016/S0531-5565(02)00008-6

Magalhães, J.P., 2012. Programmatic features of aging originating in development: aging mechanisms beyond molecular damage? The FASEB Journal 26, 4821–4826. https://doi.org/10.1096/fj.12-210872

Maiese, K., Li, F., Chong, Z.Z., Shang, Y.C., 2008. The Wnt signaling pathway: Aging gracefully as a protectionist? Pharmacology and Therapeutics. https://doi.org/10.1016/j.pharmthera.2008.01.004

Makino, T., Gojobori, T., 2007. Evolution of Protein-Protein Interaction Network. Genome Dynamics 3, 13–29. https://doi.org/10.1159/000107601

Martínez-Jiménez, F., Muiños, F., Sentís, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., Mularoni, L., Pich, O., Bonet, J., Kranas, H., Gonzalez-Perez, A., Lopez-Bigas, N., 2020. A compendium of mutational cancer driver genes. Nature Reviews Cancer 2020 20:10 20, 555–572. https://doi.org/10.1038/s41568-020-0290-x

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2015). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 44, D110--D115.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol*. 28, 495.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., Bejerano, G., 2010. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28, 495. https://doi.org/10.1038/nbt.1630

McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T., et al. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* 471, 216–219.

McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T., Wenger, A.M., Bejerano, G., Kingsley, D.M., 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature 471, 216–9. https://doi.org/10.1038/nature09774

Medvedeva, Y.A., Lennartsson, A., Ehsani, R., Kulakovskiy, I. V., Vorontsov, I.E., Panahandeh, P., Khimulya, G., Kasukawa, T., Drabløs, F., 2015. EpiFactors: a comprehensive database of human epigenetic factors and complexes. Database 2015. https://doi.org/10.1093/DATABASE/BAV067

Mercier, B., Lissens, W., Novelli, G., Kalaydjieva, L., De Arce, M., Kapranov, N., Canki Klain, N., Estivill, X., Palacio, A., Cashman, S., Savov, A., Audrézet, M.P., Dallapicolla, B., Liebaers, I., Quéré, I., Raguénès, O., Verlingue, C., Férec, C., 1994. A cluster of cystic fibrosis mutations in exon 17b of the CFTR gene: a site for rare mutations. J Med Genet 31, 731–734. https://doi.org/10.1136/JMG.31.9.731

Mérida-Velasco, J.A., Sánchez-Montesinos, I., Espín-Ferra, J., Rodríguez-Vázquez, J.F., Mérida-Velasco, J.R., and Jiménez-Collado, J. (1997). Development of the human knee joint. *Anat. Rec*. 248, 269–278.

Miller Jenkins, L.M., Feng, H., Durell, S.R., Tagad, H.D., Mazur, S.J., Tropea, J.E., Bai, Y., Appella, E., 2015. Characterization of the p300 Taz2-p53 TAD2 Complex and Comparison with the p300 Taz2-p53 TAD1 Complex. Biochemistry 54, 2001–2010. https://doi.org/10.1021/ACS.BIOCHEM.5B00044/SUPPL_FILE/BI5B00044_SI_001.PDF

Miller, M.L., Reznik, E., Gauthier, N.P., Aksoy, B.A., Korkut, A., Gao, J., Ciriello, G., Schultz, N., Sander, C., 2015. Pan-Cancer Analysis of Mutation Hotspots in Protein Domains. Cell Systems 1, 197–209. https://doi.org/10.1016/J.CELS.2015.08.014/ATTACHMENT/5AEDF9BB-E370-4CC5-82A0-00D795AAB935/MMC2.XLS

Miyamoto, Y., Mabuchi, A., Shi, D., Kubo, T., Takatori, Y., Saito, S., Fujioka, M., Sudo, A., Uchida, A., Yamamoto, S., et al. (2007). A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nat. Genet*. 39, 529–533.

Morrison, J.B. (1970). The mechanics of the knee joint in relation to normal walking. *J. Biomech*. 3, 51–61.

Moskalev, A.A., Aliper, A.M., Smit-McBride, Z., Buzdin, A., Zhavoronkov, A., 2014. Genetics and epigenetics of aging and longevity. Cell Cycle 13, 1063. https://doi.org/10.4161/CC.28433

Mostoslavsky, R., Chua, K.F., Lombard, D.B., Pang, W.W., Fischer, M.R., Gellon, L., Liu, P., Mostoslavsky, G., Franco, S., Murphy, M.M., Mills, K.D., Patel, P., Hsu, J.T., Hong, A.L., Ford, E., Cheng, H.L., Kennedy, C., Nunez, N., Bronson, R., Frendewey, D., Auerbach, W., Valenzuela, D., Karow, M., Hottiger, M.O., Hursting, S., Barrett, J.C., Guarente, L., Mulligan, R., Demple, B., Yancopoulos, G.D., Alt, F.W., 2006. Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. Cell 124, 315–329. https://doi.org/10.1016/j.cell.2005.11.044

Moyle, W.R., Campbell, R.K., Myers, R. V., Bernard, M.P., Han, Y., Wang, X., 1994. Co-evolution of ligand-receptor pairs. Nature 1994 368:6468 368, 251–255. https://doi.org/10.1038/368251a0

Mukherjee, S., Zhang, Y., 2011. Protein-protein complex structure predictions by multimeric threading and template recombination. Structure 19, 955–966. https://doi.org/10.1016/J.STR.2011.04.006

Mun, J. (2008). Advanced analytical models : over 800 models and 300 applications from the Basel II Accord to Wall Street and beyond (Wiley).

Muntané, G., Farré, X., Rodríguez, J.A., Pegueroles, C., Hughes, D.A., de Magalhães, J.P., Gabaldón, T., Navarro, A., 2018. Biological Processes Modulating Longevity across Primates: A Phylogenetic Genome-Phenome Analysis. Molecular Biology and Evolution 35, 1990–2004. https://doi.org/10.1093/MOLBEV/MSY105

Nardini, C., Moreau, J.F., Gensous, N., Ravaioli, F., Garagnani, P., Bacalini, M.G., 2018. The epigenetics of inflammaging: The contribution of age-related heterochromatin loss and locus-specific remodelling and the modulation by environmental stimuli. Seminars in Immunology 40, 49–60. https://doi.org/10.1016/J.SMIM.2018.10.009

Neill, D., 2012. Should Alzheimer's disease be equated with human brain ageing?: A maladaptive interaction between brain evolution and senescence. Ageing Research Reviews 11, 104–122. https://doi.org/10.1016/J.ARR.2011.06.004

Nemec, S., Luxey, M., Jain, D., Huang Sung, A., Pastinen, T., and Drouin, J. (2017). Pitx1 directly modulates the core limb development program to implement hindlimb identity . *Development* 144, 3325–3335.

Neogi, T., and Felson, D.T. (2016). Osteoarthritis: Bone as an imaging biomarker and treatment target in OA. *Nat. Rev. Rheumatol*. 12, 503–504.

Neogi, T., Bowes, M.A., Niu, J., De Souza, K.M., Vincent, G.R., Goggins, J., Zhang, Y., and Felson, D.T. (2013). Magnetic resonance imaging-based three-dimensional bone shape of the knee predicts onset of knee osteoarthritis: data from the osteoarthritis initiative. *Arthritis Rheum*. 65, 2048–2058.

Ng, P.C., Henikoff, S., 2006. Predicting the Effects of Amino Acid Substitutions on Protein Function. http://dx.doi.org/10.1146/annurev.genom.7.080505.115630 7, 61–80. https://doi.org/10.1146/ANNUREV.GENOM.7.080505.115630

Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A.G. (2007). Recent and ongoing selection in the human genome. *Nat. Rev. Genet.* 8, 857–868.

Nitsche, A., Arnold, C., Ueberham, U., Reiche, K., Fallmann, J., Hackermüller, J., Horn, F., Stadler, P.F., Arendt, T., 2020. Alzheimer-related genes show accelerated evolution. Molecular Psychiatry 2020 26:10 26, 5790–5796. https://doi.org/10.1038/S41380-020-0680-1

Niu, B., Scott, A.D., Sengupta, S., Bailey, M.H., Batra, P., Ning, J., Wyczalkowski, M.A., Liang, W.-W.W., Zhang, Q., McLellan, M.D., Sun, S.Q., Tripathi, P., Lou, C., Ye, K., Mashl, R.J., Wallis, J., Wendl, M.C., Chen, F., Ding, L., Jay Mashl, R., Wallis, J., Wendl, M.C., Chen, F., Ding, L., 2016b. Protein-structure-guided discovery of functional mutations across 19 cancer types. Nature Genetics 2016 48:8 48, 827–837. https://doi.org/10.1038/ng.3586

Niu, B., Scott, A.D., Sengupta, S., Bailey, M.H., Batra, P., Ning, J., Wyczalkowski, M.A., Liang, W.W., Zhang, Q., McLellan, M.D., Sun, S.Q., Tripathi, P., Lou, C., Ye, K., Jay Mashl, R., Wallis, J., Wendl, M.C., Chen, F., Ding, L., 2016a. Protein-structure-guided discovery of functional mutations across 19 cancer types. Nature Genetics 2016 48:8 48, 827–837. https://doi.org/10.1038/ng.3586

Nolte, R.T., Wisely, G.B., Westin, S., Cobb, J.E., Lambert, M.H., Kurokawa, R., Rosenfeld, M.G., Willson, T.M., Glass, C.K., Milburn, M. V., 1998. Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. Nature 395, 137–143. https://doi.org/10.1038/25931

Nwabuo, C.C., Yano, Y., Moreira, H.T., Appiah, D., Vasconcellos, H.D., Aghaji, Q.N., Viera, A., Rana, J.S., Shah, R. v., Murthy, V.L., Allen, N.B., Schreiner, P.J., Lloyd-Jones, D.M., Lima, J.A.C., 2020. Association Between Visit-to-Visit Blood Pressure Variability in Early Adulthood and Myocardial Structure and Function in Later Life. JAMA Cardiology 5, 795–801. https://doi.org/10.1001/JAMACARDIO.2020.0799

O'Rahilly, R. (1951). The early prenatal development of the human knee joint. *J. Anat*. 85, 166–170.

O'Rahilly, R. (1979). Early human development and the chief sources of information on staged human embryos. *Eur. J. Obstet. Gynecol*. Reprod. Biol. 9, 273–280.

Ollikainen, M., Smith, K.R., Joo, E.J.H., Ng, H.K., Andronikos, R., Novakovic, B., Aziz, N.K.A., Carlin, J.B., Morley, R., Saffery, R., Craig, J.M., 2010. DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. Human Molecular Genetics 19, 4176–4188. https://doi.org/10.1093/hmg/ddq336

Osuna, S., 2021. The challenge of predicting distal active site mutations in computational enzyme design. Wiley Interdisciplinary Reviews: Computational Molecular Science 11, e1502. https://doi.org/10.1002/WCMS.1502

Owen Lovejoy, C. (2007). The natural history of human gait and posture. Part 3. The knee. *Gait Posture* 25, 325–341.

Pagani, L., Lawson, D.J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., DeGiorgio, M., Saag, L., et al. (2016). Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* 538, 238.

Pan, H., Finkel, T., 2017. Key proteins and pathways that regulate lifespan. The Journal of Biological Chemistry 292, 6452. https://doi.org/10.1074/JBC.R116.771915

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26, 419–420.

Pasaniuc, B., Price, A.L., 2016. Dissecting the genetics of complex traits using summary association statistics. Nature Reviews Genetics 2016 18:2 18, 117–127. https://doi.org/10.1038/NRG.2016.142

Peng, J., Svetec, N., Zhao, L., 2022. Intermolecular Interactions Drive Protein Adaptive and Coadaptive Evolution at Both Species and Population Levels. Molecular Biology and Evolution 39. https://doi.org/10.1093/MOLBEV/MSAB350

Picard, C., Azeddine, B., Moldovan, F., Martel-Pelletier, J., and Moreau, A. (2007). New emerging role of Pitx1 transcription factor in osteoarthritis pathogenesis. *Clin. Orthop. Relat. Res*. 462, 59–66.

Piñero, J., Ramírez-Anguita, J.M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., Furlong, L.I., 2020. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res 48, D845–D855. https://doi.org/10.1093/NAR/GKZ1021

Plach, M.G., Semmelmann, F., Busch, F., Busch, M., Heizinger, L., Wysocki, V.H., Merkl, R., Sterner, R., 2017. Evolutionary diversification of protein–protein interactions by interface add-ons. Proceedings of the National Academy of Sciences 114, E8333–E8342. https://doi.org/10.1073/PNAS.1707335114

Polk, J.D., Williams, S.A., and Peterson, J. V. (2009). Body size and joint posture in primates. *Am. J. Phys. Anthropol.* 140, 359–367.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 20, 110–121.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., Siepel, A., 2010. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 20, 110–121.

Pollard, K.S., Salama, S.R., King, B., Kern, A.D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J.S., Bejerano, G., Baertsch, R., et al. (2006a). Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* 2, e168.

Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006b). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443, 167.

Pontzer, H., 2012. Ecological Energetics in Early *Homo*. Current Anthropology 53, S346–S358. https://doi.org/10.1086/667402

Prabhakar, S., Noonan, J.P., Pääbo, S., and Rubin, E.M. (2006). Accelerated evolution of conserved noncoding sequences in humans. *Science* (80). 314, 786.

Prabhakar, S., Visel, A., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Morrison, H., Fitzpatrick, D.R., Afzal, V., et al. (2008). Human-specific gain of function in a developmental enhancer. *Science* 321, 1346–1350.

Prabhakar, S., Visel, A., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Morrison, H., Fitzpatrick, D.R., Afzal, V., Pennacchio, L.A., Rubin, E.M., Noonan, J.P., 2008. Human-specific gain of function in a developmental enhancer. Science 321, 1346–50. https://doi.org/10.1126/science.1159974

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. (2013). Great ape genetic diversity and population history. *Nature* 499, 471.

Prado, N.A., Brown, J.L., Zoller, J.A., Haghani, A., Yao, M., Bagryanova, L.R., Campana, M.G., E. Maldonado, J., Raj, K., Schmitt, D., Robeck, T.R., Horvath, S., 2021. Epigenetic clock and methylation studies in elephants. Aging Cell 20, e13414. https://doi.org/10.1111/ACEL.13414

Pregizer, S.K., Kiapour, A.M., Young, M., Chen, H., Schoor, M., Liu, Z., Cao, J., Rosen, V., and Capellini, T.D. (2018). Impact of broad regulatory regions on Gdf5 expression and function in knee development and susceptibility to osteoarthritis. *Ann. Rheum. Dis.* 77, 450.

Prendergast, J.G.D., Chambers, E. V., Semple, C.A.M., 2014. Sequence-Level Mechanisms of Human Epigenome Evolution. Genome Biology and Evolution 6, 1758–1771. https://doi.org/10.1093/GBE/EVU142

Prud'homme, B., Gompel, N., Carroll, S.B., 2007. Emerging principles of regulatory evolution. Proc Natl Acad Sci U S A 104, 8605–8612. https://doi.org/10.1073/pnas.0700488104

Puca, A.A., Carrizzo, A., Spinelli, C., Damato, A., Ambrosio, M., Villa, F., Ferrario, A., Maciag, A., Fornai, F., Lenzi, P., Valenti, V., di Nonno, F., Accarino, G., Madonna, M., Forte, M., Calì, G., Baragetti, A., Norata, G.D., Catapano, A.L., Cattaneo, M., Izzo, R., Trimarco, V., Montella, F., Versaci, F., Auricchio, A., Frati, G., Sciarretta, S., Madeddu, P., Ciaglia, E., Vecchione, C., 2020. Single systemic transfer of a human gene associated with exceptional longevity halts the progression of atherosclerosis

and inflammation in ApoE knockout mice through a CXCR4-mediated mechanism. European Heart Journal 41, 2487–2497. https://doi.org/10.1093/EURHEARTJ/EHZ459

Qin, H., Srinivasula, S.M., Wu, G., Fernandes-Alnemri, T., Alnemri, E.S., Shi, Y., 1999a. Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. Nature 1999 399:6736 399, 549–557. https://doi.org/10.1038/21124

Qin, H., Srinivasula, S.M., Wu, G., Fernandes-Alnemri, T., Alnemri, E.S., Shi, Y., 1999b. Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. Nature 1999 399:6736 399, 549–557. https://doi.org/10.1038/21124

Qiu, C., Fratiglioni, L., 2015. A major role for cardiovascular burden in age-related cognitive decline. Nature Reviews Cardiology 2015 12:5 12, 267–277. https://doi.org/10.1038/nrcardio.2014.223

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.

R Development Core Team (2008). R: A Language and Environment for Statistical Computing.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing.

Racimo, F. (2016). Testing for ancient selection using cross-population allele frequency differentiation. *Genetics* 202, 733–750.

Raj, K., Horvath, S., 2020. Current perspectives on the cellular and molecular features of epigenetic ageing. Experimental Biology and Medicine 245, 1532–1542. https://doi.org/10.1177/1535370220918329

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc*. 8, 2281.

Rapoport, S.I., 1989. Hypothesis: Alzheimer's disease is a phylogenetic disease. Medical Hypotheses 29, 147–150. https://doi.org/10.1016/0306-9877(89)90185-0

Rasmussen, M.D., Hubisz, M.J., Gronau, I., and Siepel, A. (2014). Genome-Wide Inference of Ancestral Recombination Graphs. *PLoS Genet.* 10, e1004342.

Ratnakumar, A., Weinhold, N., Mar, J.C., Riaz, N., 2020. Protein-Protein interactions uncover candidate 'core genes' within omnigenic disease networks. PLOS Genetics 16, e1008903. https://doi.org/10.1371/JOURNAL.PGEN.1008903

Rau, J.C., Beaulieu, L.M., Huntington, J.A., Church, F.C., 2007. Serpins in thrombosis, hemostasis and fibrinolysis. J Thromb Haemost 5, 102. https://doi.org/10.1111/J.1538-7836.2007.02516.X

Reyes, C., Leyland, K.M., Peat, G., Cooper, C., Arden, N.K., and Prieto-Alhambra, D. (2016). Association between overweight and obesity and risk of clinically diagnosed knee, hip, and hand osteoarthritis: a population-based cohort study. *Arthritis Rheumatol*. 68, 1869–1875.

Reyna, M.A., Haan, D., Paczkowska, M., Verbeke, L.P.C., Vazquez, M., Kahraman, A., Pulido-Tamayo, S., Barenboim, J., Wadi, L., Dhingra, P., Shrestha, R., Getz, G., Lawrence, M.S., Pedersen, J.S., Rubin, M.A., Wheeler, D.A., Brunak, S., Izarzugaza, J.M.G., Khurana, E., Marchal, K., von Mering, C., Sahinalp, S.C., Valencia, A., Abascal, F., Amin, S.B., Bader, G.D., Bandopadhayay, P., Beroukhim, R., Bertl, J., Boroevich, K.A., Busanovich, J., Campbell, P.J., Carlevaro-Fita, J., Chakravarty, D., Chan, C.W.Y., Chen, K., Choi, J.K., Deu-Pons, J., Diamanti, K., Feuerbach, L., Fink, J.L., Fonseca, N.A., Frigola, J., Gambacorti-Passerini, C., Garsed, D.W., Gerstein, M., Guo, Q., Gut, I.G., Hamilton, M.P., Haradhvala, N.J., Harmanci, A.O., Helmy, M., Herrmann, C., Hess, J.M., Hobolth, A., Hodzic, E., Hong, C., Hornshøj, H., Isaev, K., Johnson, R., Johnson, T.A., Juul, M., Juul, R.I., Kahles, A., Kellis, M., Kim, J., Kim, J.K., Kim, Y., Komorowski, J., Korbel, J.O., Kumar, S., Lanzós, A., Larsson, E., Lee, D., Lehmann, K. Van, Li, S., Li, X., Lin, Z., Liu, E.M., Lochovsky, L., Lou, S., Madsen, T., Martincorena, I., Martinez-Fundichely, A., Maruvka, Y.E., McGillivray, P.D., Meyerson, W., Muiños, F., Mularoni, L., Nakagawa, H., Nielsen, M.M., Park, Keunchil, Park, Kiejung, Pons, T., Reyes-Salazar, I., Rheinbay, E., Rubio-Perez, C., Saksena, G., Salichos, L., Sander, C., Schumacher, S.E., Shackleton, M., Shapira, O., Shen, C., Shuai, S., Sidiropoulos, N., Sieverling, L., Sinnott-Armstrong, N., Stein, L.D., Tamborero, D., Tiao, G., Tsunoda, T., Umer, H.M., Uusküla-Reimand, L., Wadelius, C., Wang, J., Warrell, J., Waszak, S.M., Weischenfeldt, J., Wu, G., Yu, J., Zhang, J., Zhang, X., Zhang, Y., Zhao, Z., Zou, L., Reimand, J., Stuart, J.M., Raphael, B.J., 2020. Pathway and network analysis of more than 2500 whole cancer genomes. Nature Communications 2020 11:1 11, 1–17. https://doi.org/10.1038/s41467-020-14367-0

Richard, D., Capellini, T.D., 2021. Shifting epigenetic contexts influence regulatory variation and disease risk. Aging 13, 15699–15749. https://doi.org/10.18632/AGING.203194

Richard, D., Liu, Z., Cao, J., Kiapour, A.M., Willen, J., Yarlagadda, S., Jagoda, E., Kolachalama, V.B., Sieker, J.T., Chang, G.H., Muthuirulan, P., Young, M., Masson, A., Konrad, J., Hosseinzadeh, S., Maridas, D.E., Rosen, V., Krawetz, R., Roach, N., Capellini, T.D., 2020. Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. Cell 0. https://doi.org/10.1016/j.cell.2020.02.057

Richette, P., Poitou, C., Garnero, P., Vicaut, E., Bouillot, J.-L., Lacorte, J.-M., Basdevant, A., Clément, K., Bardin, T., and Chevalier, X. (2011). Benefits of massive weight loss on symptoms, systemic inflammation and cartilage turnover in obese patients with knee osteoarthritis. *Ann. Rheum. Dis.* 70, 139–144.

Ricklefs, R.E., 2008. The evolution of senescence from a comparative perspective. Functional Ecology 22, 379–392. https://doi.org/10.1111/j.1365-2435.2008.01420.x

Riley, J.C., 2001. Rising life expectancy: a global history.

Rinaudo, P., Wang, E., 2012. Fetal programming and metabolic syndrome. Annual Review of Physiology 74, 107–130. https://doi.org/10.1146/annurev-physiol-020911-153245

Ripoche, H., Laine, E., Ceres, N., Carbone, A., 2017. JET2 Viewer: a database of predicted multiple, possibly overlapping, protein–protein interaction sites for PDB structures. Nucleic Acids Research 45, D236. https://doi.org/10.1093/NAR/GKW1053

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43, e47. https://doi.org/10.1093/nar/gkv007

Rivera, C.M., Ren, B., 2013. Mapping human epigenomes. Cell. https://doi.org/10.1016/j.cell.2013.09.011

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S.,

Carles, A., Dixon, J.R., Farh, K.H., Feizi, S., Karlic, R., Kim, A.R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., de Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–329. https://doi.org/10.1038/nature14248

Robine, J., Allard, M., 1998. The Oldest Human. Science (1979) 279, 1831h–11831. https://doi.org/10.1126/science.279.5358.1831h

Rodríguez-Rodero, S., Fernández-Morera, J.L., Menéndez-Torre, E., Calvanese, V., Fernández, A.F., Fraga, M.F., 2011. Aging genetics and aging. Aging and Disease.

Rodríguez, J.A., Marigorta, U.M., Hughes, D.A., Spataro, N., Bosch, E., Navarro, A., 2017. Antagonistic pleiotropy and mutation accumulation influence human senescence and disease. Nature Ecology and Evolution 1, 55. https://doi.org/10.1038/s41559-016-0055

Rogers, J., Gibbs, R.A., 2014. Comparative primate genomics: Emerging patterns of genome content and dynamics. Nature Reviews Genetics 15, 347–359. https://doi.org/10.1038/nrg3707

Rongo, C., 2011. Epidermal growth factor and aging: A signaling molecule reveals a new eye opening function. Aging (Albany NY) 3, 896. https://doi.org/10.18632/AGING.100384

Rosen, V. (2009). BMP2 signaling in bone development and repair. *Cytokine Growth Factor Rev*. 20, 475–480.

Rountree, R.B., Schoor, M., Chen, H., Marks, M.E., Harley, V., Mishina, Y., and Kingsley, D.M. (2004). BMP receptor signaling is required for postnatal maintenance of articular cartilage. *PLoS Biol*. 2, e355.

Rubinstein, M., de Souza, F.S.J., 2013. Evolution of transcriptional enhancers and animal diversity. Philosophical Transactions of the Royal Society B: Biological Sciences 368. https://doi.org/10.1098/rstb.2013.0017

Rufini, A., Tucci, P., Celardo, I., Melino, G., 2013. Senescence and aging: the critical roles of p53. Oncogene 2013 32:43 32, 5129–5143. https://doi.org/10.1038/onc.2012.640

Rux, D., Decker, R.S., Koyama, E., and Pacifici, M. (2019). Joints in the appendicular skeleton: Developmental mechanisms and evolutionary influences. *Curr. Top. Dev. Biol*. 133, 119–151.

Saffery, R., Novakovic, B., 2014. Epigenetics as the mediator of fetal programming of adult onset disease: what is the evidence? Acta Obstetricia et Gynecologica Scandinavica 93, 1090–1098. https://doi.org/10.1111/aogs.12431

Salazar, V.S., Gamer, L.W., and Rosen, V. (2016). BMP signalling in skeletal development, disease and repair*. Nat. Rev. Endocrinol*. 12, 203–221.

Salminen, A., Kaarniranta, K., 2010. Insulin/IGF-1 paradox of aging: Regulation via AKT/IKK/NF-κB signaling. Cellular Signalling. https://doi.org/10.1016/j.cellsig.2009.10.006

Salmon, A.B., Richardson, A., Pérez, V.I., 2010. Update on the oxidative stress theory of aging: Does oxidative stress play a role in aging or healthy aging? Free Radical Biology and Medicine 48, 642–655. https://doi.org/10.1016/J.FREERADBIOMED.2009.12.015

Salverda, M.L.M., Dellus, E., Gorter, F.A., Debets, A.J.M., van der Oost, J., Hoekstra, R.F., Tawfik, D.S., de Visser, J.A.G.M., 2011. Initial mutations direct alternative pathways of protein evolution. PLoS Genet 7. https://doi.org/10.1371/JOURNAL.PGEN.1001321

Schuppli, C., Isler, K., van Schaik, C.P., 2012. How to explain the unusually late age at skill competence among humans. Journal of Human Evolution 63, 843–850. https://doi.org/10.1016/J.JHEVOL.2012.08.009

Schwartz, H.J. (1989). Effect of chronic chromolyn sodium therapy in a beautician with occupational asthma. *J. Occup. Med*. 31, 112–114.

Seim, I., Fang, X., Xiong, Z., Lobanov, A. V., Huang, Z., Ma, S., Feng, Y., Turanov, A.A., Zhu, Y., Lenz, T.L., Gerashchenko, M. V., Fan, D., Hee Yim, S., Yao, X., Jordan, D., Xiong, Y., Ma, Y., Lyapunov, A.N., Chen, G., Kulakova, O.I., Sun, Y., Lee, S.G., Bronson, R.T., Moskalev, A.A., Sunyaev, S.R., Zhang, G., Krogh, A., Wang, J., Gladyshev, V.N., 2013. Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. Nature Communications 2013 4:1 4, 1–8. https://doi.org/10.1038/ncomms3212

Selleri, L., Depew, M.J., Jacobs, Y., Chanda, S.K., Tsang, K.Y., Cheah, K.S.E., Rubenstein, J.L.R., O'Gorman, S., and Cleary, M.L. (2001). Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* 128, 3543–3557.

Sen, M., Cheng, Y.-H., Goldring, M.B., Lotz, M.K., and Carson, D.A. (2004). WISP3-dependent regulation of type II collagen and aggrecan production in chondrocytes. *Arthritis Rheum.* 50, 488–497.

Sen, P., Shah, P.P., Nativio, R., Berger, S.L., 2016a. Epigenetic Mechanisms of Longevity and Aging. Cell. https://doi.org/10.1016/j.cell.2016.07.050

Sen, P., Shah, P.P., Nativio, R., Berger, S.L., 2016b. Epigenetic mechanisms regulating longevity and aging. Cell 166, 822. https://doi.org/10.1016/J.CELL.2016.07.050

Settle, S.H., Rountree, R.B., Sinha, A., Thacker, A., Higgins, K., and Kingsley, D.M. (2003). Multiple joint and skeletal patterning defects caused by single and double mutations in the mouse Gdf6 and Gdf5 genes. *Dev. Biol.* 254, 116–130.

Shannon, P., and Richards, M. (2014). MotifDb: An annotated collection of protein-DNA binding sequence motifs. R Packag. Version 1.

Sharir-Ivry, A., Xia, Y., 2019. Non-catalytic Binding Sites Induce Weaker Long-Range Evolutionary Rate Gradients than Catalytic Sites in Enzymes. Journal of Molecular Biology 431, 3860–3870. https://doi.org/10.1016/J.JMB.2019.07.019

Shinoda, Y., Ogata, N., Higashikawa, A., Manabe, I., Shindo, T., Yamada, T., Kugimiya, F., Ikeda, T., Kawamura, N., Kawasaki, Y., et al. (2008). Krüppel-like factor 5 causes cartilage degradation through transactivation of matrix metalloproteinase 9. *J. Biol. Chem*. 283, 24682–24689.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet*. 2014 154 15, 272.

Shultz, A.J., Sackton, T.B., 2019. Immune genes are hotspots of shared positive selection across birds and mammals. Elife 8. https://doi.org/10.7554/ELIFE.41815

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.D.W., Richards, S., Weinstock, G.M., Wilson, R.K., Gibbs, R.A., Kent, W.J., Miller, W., Haussler, D., 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Research 15, 1034–1050. https://doi.org/10.1101/gr.3715005

Sigismund, S., Avanzato, D., Lanzetti, L., 2018. Emerging functions of the EGFR in cancer. Mol Oncol 12, 3–20. https://doi.org/10.1002/1878-0261.12155

Sikosek, T., Chan, H.S., 2014. Biophysics of protein evolution and evolutionary protein biophysics. Journal of The Royal Society Interface 11. https://doi.org/10.1098/RSIF.2014.0419

Singh, P.P., Demmitt, B.A., Nath, R.D., Brunet, A., 2019. The Genetics of Aging: A Vertebrate Perspective. Cell 177, 200–220. https://doi.org/10.1016/J.CELL.2019.02.038

Slodkowicz, G., Goldman, N., 2020. Integrated structural and evolutionary analysis reveals common mechanisms underlying adaptive evolution in mammals. Proceedings of the National Academy of Sciences 117, 5977–5986. https://doi.org/10.1073/PNAS.1916786117

Sonnery-Cottet, B., Archbold, P., Cucurulo, T., Fayard, J.-M., Bortolletto, J., Thaunat, M., Prost, T., and Chambat, P. (2011). The influence of the tibial slope and the size of the intercondylar notch on rupture of the anterior cruciate ligament*. J. Bone Joint Surg. Br.* 93, 1475–1478.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.

Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., Fowler, D.M., 2017. Variant Interpretation: Functional Assays to the Rescue. American Journal of Human Genetics 101, 315. https://doi.org/10.1016/J.AJHG.2017.07.014

Starr, T.N., Thornton, J.W., 2016. Epistasis in protein evolution. Protein Science : A Publication of the Protein Society 25, 1204. https://doi.org/10.1002/PRO.2897

Stricker, S., Fundele, R., Vortkamp, A., and Mundlos, S. (2002). Role of Runx genes in chondrocyte differentiation. *Dev. Biol*. 245, 95–108.

Styrkarsdottir, U., Lund, S.H., Thorleifsson, G., Zink, F., Stefansson, O.A., Sigurdsson, J.K., Juliusson, K., Bjarnadottir, K., Sigurbjornsdottir, S., Jonsson, S., et al. (2018). Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis. *Nat. Genet.* 50, 1681–1687.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A

knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences 102, 15545–15550. https://doi.org/10.1073/PNAS.0506580102

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., Collins, R., 2015. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Medicine 12, e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., Kondrashov, A.S., Bork, P., 2001. Prediction of deleterious human alleles. Human Molecular Genetics 10, 591–597. https://doi.org/10.1093/HMG/10.6.591

Szeto, D.P., Rodriguez-Esteban, C., Ryan, A.K., O'Connell, S.M., Liu, F., Kioussi, C., Gleiberman, A.S., Izpisúa-Belmonte, J.C., and Rosenfeld, M.G. (1999). Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. *Genes Dev.* 13, 484–494.

Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., Jensen, L.J., von Mering, C., 2021. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res 49, D605–D612. https://doi.org/10.1093/NAR/GKAA1074

Tachmazidou, I., Hatzikotoulas, K., Southam, L., Esparza-Gordillo, J., Haberland, V., Zheng, J., Johnson, T., Koprulu, M., Zengini, E., Steinberg, J., et al. (2019). Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat. Genet*. 51, 230–236.

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, Di., Craig, T., DIana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V.E., de Magalhães, J.P., 2018. Human Ageing Genomic Resources: new and updated databases. Nucleic Acids Res 46, D1083–D1090. https://doi.org/10.1093/NAR/GKX1042

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.

Tang, J., Yan, H., Zhuang, S., 2013. Histone deacetylases as targets for treatment of multiple diseases. Clinical Science. https://doi.org/10.1042/CS20120504

Tardieu, C. (1999). Ontogeny and phylogeny of femoro-tibial characters in humans and hominid fossils: Functional influence and genetic determinism. Am. J. Phys. Anthropol. Off. Publ. *Am. Assoc. Phys. Anthropol.* 110, 365–377.

Tardieu, C. (2010). Development of the human hind limb and its importance for the evolution of bipedalism. *Evol. Anthropol. Issues, News, Rev.* 19, 174–186.

Tasselli, L., Zheng, W., Chua, K.F., 2017. SIRT6: Novel Mechanisms and Links to Aging and Disease. Trends in Endocrinology and Metabolism. https://doi.org/10.1016/j.tem.2016.10.002

Tejada-Martinez, D., Avelar, R.A., Lopes, I., Zhang, B., Novoa, G., de Magalhães, J.P., Trizzino, M., 2022. Positive Selection and Enhancer Evolution Shaped Lifespan and Body Mass in Great Apes. Molecular Biology and Evolution 39. https://doi.org/10.1093/MOLBEV/MSAB369

Teppa, E., Zea, D.J., Marino-Buslje, C., 2017. Protein–protein interactions leave evolutionary footprints: High molecular coevolution at the core of interfaces. Protein Science 26, 2438–2444. https://doi.org/10.1002/pro.3318

Teschendorff, A.E., West, J., Beck, S., 2013. Age-associated epigenetic drift: Implications, and a case of epigenetic thrift? Human Molecular Genetics 22, R7–R15. https://doi.org/10.1093/hmg/ddt375

Timmers, P.R.H.J., Mounier, N., Lall, K., Fischer, K., Ning, Z., Feng, X., Bretherick, A.D., Clark, D.W., Shen, X., Esko, T., Kutalik, Z., Wilson, J.F., Joshi, P.K., 2019. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. Elife 8, 1–40. https://doi.org/10.7554/elife.39856

Tower, J., 2015. Programmed cell death in aging. Ageing Res Rev 23, 90. https://doi.org/10.1016/J.ARR.2015.04.002

Treaster, S., Daane, J.M., Harris, M.P., 2021. Refining Convergent Rate Analysis with Topology in Mammalian Longevity and Marine Transitions. Molecular Biology and Evolution 38, 5190–5203. https://doi.org/10.1093/MOLBEV/MSAB226

Trinkaus, E., 2011. Late Pleistocene adult mortality patterns and modern human establishment. Proc Natl Acad Sci U S A 108, 1267–71. https://doi.org/10.1073/pnas.1018700108

Trizzino, M., Park, Y.S., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., Perry, G.H., Lynch, V.J., Brown, C.D., 2017. Transposable elements are the primary source of novelty in primate gene regulation. Genome Research 27, 1623–1633. https://doi.org/10.1101/GR.218149.116

Tsurumi, A., Li, W.X., 2012. Global heterochromatin loss: A unifying theory of aging? Epigenetics 7, 680–688. https://doi.org/10.4161/epi.20540

Turan, Z.G., Parvizi, P., Dönertaş, H.M., Tung, J., Khaitovich, P., Somel, M., 2019. Molecular footprint of Medawar's mutation accumulation process in mammalian aging. Aging Cell 18. https://doi.org/10.1111/ACEL.12965

Valdes, A.M., Evangelou, E., Kerkhof, H.J.M., Tamm, A., Doherty, S.A., Kisand, K., Tamm, A., Kerna, I., Uitterlinden, A., Hofman, A., et al. (2011). The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. *Ann. Rheum. Dis*. 70, 873–875.

Valdes, A.M., Van Oene, M., Hart, D.J., Surdulescu, G.L., Loughlin, J., Doherty, M., and Spector, T.D. (2006). Reproducible genetic associations between candidate genes and clinical knee osteoarthritis in men and women. *Arthritis Rheum.* 54, 533–539.

van der Kraan, P.M., Blaney Davidson, E.N., Blom, A., and van den Berg, W.B. (2009). TGF-beta signaling in chondrocyte terminal differentiation and osteoarthritis: Modulation and integration of signaling pathways through receptor-Smads. *Osteoarthr. Cartil*. 17, 1539–1545.

Varki, A., Altheide, T.K., 2005. Comparing the human and chimpanzee genomes: Searching for needles in a haystack. Genome Research 15, 1746–1758. https://doi.org/10.1101/GR.3737405

Varki, A., and Altheide, T.K. (2005). Comparing the human and chimpanzee genomes: Searching for needles in a haystack. *Genome Res*. 15, 1746–1758.

Veillette, A., Chen, J., 2018. SIRPα–CD47 Immune Checkpoint Blockade in Anticancer Therapy. Trends in Immunology 39, 173–184. https://doi.org/10.1016/J.IT.2017.12.005

Villa, F., Carrizzo, A., Spinelli, C.C., Ferrario, A., Malovini, A., Maciąg, A., Damato, A., Auricchio, A., Spinetti, G., Sangalli, E., Dang, Z., Madonna, M., Ambrosio, M., Sitia, L., Bigini, P., Calì, G., Schreiber, S., Perls, T., Fucile, S., Mulas, F., Nebel, A., Bellazzi, R., Madeddu, P., Vecchione, C., Puca, A.A., 2015. Genetic analysis reveals a longevity-associated protein modulating endothelial function and angiogenesis. Circulation Research 117, 333–345. https://doi.org/10.1161/CIRCRESAHA.117.305875

Villeponteau, B., 1997. The heterochromatin loss model of aging. Experimental Gerontology 32, 383–394. https://doi.org/10.1016/S0531-5565(96)00155-6

Visel, A., Thaller, C., and Eichele, G. (2004). GenePaint. org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res*. 32, D552--D556.

Vitti, J.J., Grossman, S.R., and Sabeti, P.C. (2013). Detecting natural selection in genomic data. *Annu. Rev. Genet*. 47, 97–120.

Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *PLoS Biol*. 4, e72.

Wang, J.S., Infante, C.R., Park, S., and Menke, D.B. (2018). PITX1 promotes chondrogenesis and myogenesis in mouse hindlimbs through conserved regulatory targets. *Dev. Biol*. 434, 186–195.

Wang, R.N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S., et al. (2014). Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis*. 1, 87–105.

Wang, X., Zhang, X., Peng, C., Shi, Y., Li, H., Xu, Z., Zhu, W., 2021. D3DistalMutation: A Database to Explore the Effect of Distal Mutations on Enzyme Activity. Journal of Chemical Information and Modeling 61, 2499–2508. https://doi.org/10.1021/ACS.JCIM.1C00318/SUPPL_FILE/CI1C00318_SI_002.XLSX

Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 40, D930–D934.

Wauthier, V., Verbeeck, R., Buc Calderon, P., 2007. The effect of ageing on cytochrome p450 enzymes: consequences for drug biotransformation in the elderly. Curr Med Chem 14, 745–757. https://doi.org/10.2174/092986707780090981

Weichhart, T., 2018. mTOR as Regulator of Lifespan, Aging, and Cellular Senescence: A Mini-Review. Gerontology 64, 127–134. https://doi.org/10.1159/000484629

Weiner, B.E., Huang, H., Dattilo, B.M., Nilges, M.J., Fanning, E., Chazin, W.J., 2007. An iron-sulfur cluster in the C-terminal domain of the p58 subunit of human DNA primase. J Biol Chem 282, 33444–33451. https://doi.org/10.1074/JBC.M705826200

Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* (N. Y). 38, 1358–1370.

Weon, B.M., Je, J.H., n.d. Theoretical estimation of maximum human lifespan. Biogerontology 10, 65–71. https://doi.org/10.1007/s10522-008-9156-4

Weston, A.D., Rosen, V., Chandraratna, R.A., and Underhill, T.M. (2000). Regulation of skeletal progenitor differentiation by the BMP and retinoid signaling pathways. *J. Cell Biol.* 148, 679–690.

Whitmer, R.A., Sidney, S., Selby, J., Johnston, S.C., Yaffe, K., 2005. Midlife cardiovascular risk factors and risk of dementia in late life. Neurology 64, 277. https://doi.org/10.1212/01.WNL.0000149519.47454.F2

Wilkinson, G.S., Adams, D.M., Haghani, A., Lu, A.T., Zoller, J., Breeze, C.E., Arnold, B.D., Ball, H.C., Carter, G.G., Cooper, L.N., Dechmann, D.K.N., Devanna, P., Fasel, N.J., Galazyuk, A. V., Günther, L., Hurme, E., Jones, G., Knörnschild, M., Lattenkamp, E.Z., Li, C.Z., Mayer, F., Reinhardt, J.A., Medellin, R.A., Nagy, M., Pope, B., Power, M.L., Ransome, R.D., Teeling, E.C., Vernes, S.C., Zamora-Mejías, D., Zhang, J., Faure, P.A., Greville, L.J., Herrera M., L.G., Flores-Martínez, J.J., Horvath, S., 2021. DNA methylation predicts age and provides insight into exceptional longevity of bats. Nature Communications 2021 12:1 12, 1–13. https://doi.org/10.1038/s41467-021-21900-2

Williams, G.C., 1957. PLEIOTROPY, NATURAL SELECTION, AND THE EVOLUTION OF SENESCENCE. Evolution (N Y) 11, 398–411. https://doi.org/10.1111/j.1558-5646.1957.tb02911.x

Willingham, S.B., Volkmer, J.-P.P., Gentles, A.J., Sahoo, D., Dalerba, P., Mitra, S.S., Wang, J., Contreras-Trujillo, H., Martin, R., Cohen, J.D., Lovelace, P., Scheeren, F.A., Chao, M.P., Weiskopf, K., Tang, C., Volkmer, A.K., Naik, T.J., Storm, T.A., Mosley, A.R., Edris, B., Schmid, S.M., Sun, C.K., Chua, M.-S.S., Murillo, O., Rajendran, P., Cha, A.C., Chin, R.K., Kim, D., Adorno, M., Raveh, T., Tseng, D., Jaiswal, S., Enger, P.Ø., Steinberg, G.K., Li, G., So, S.K., Majeti, R., Harsh, G.R., Rijn, M. van de, Teng, N.N.H.H., Sunwoo, J.B., Alizadeh, A.A., Clarke, M.F., Weissman, I.L., Van Rijn, M. De, Teng, N.N.H.H., Sunwoo, J.B., Alizadeh, A.A., Clarke, M.F., Weissman, I.L., 2012. The CD47-signal regulatory protein alpha (SIRPa) interaction is a therapeutic target for human solid tumors 109, 6662–6667.

Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet*. 13, 59–69.

Wood, J.G., Jones, B.C., Jiang, N., Chang, C., Hosier, S., Wickremesinghe, P., Garcia, M., Hartnett, D.A., Burhenn, L., Neretti, N., Helfand, S.L., 2016. Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in Drosophila. Proceedings of the National Academy of Sciences 113, 11277–11282. https://doi.org/10.1073/PNAS.1604621113

Wright, A., Charlesworth, B., Rudan, I., Carothers, A., Campbell, H., 2003. A polygenic basis for late-onset disease. Trends in Genetics. https://doi.org/10.1016/S0168-9525(02)00033-1

Wu, M., Chen, G., and Li, Y.-P. (2016). TGF-β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res.* 4, 16009.

Xia, S., Zhang, X., Zheng, S., Khanabdali, R., Kalionis, B., Wu, J., Wan, W., Tai, X., 2016. An Update on Inflamm-Aging: Mechanisms, Prevention, and Treatment. Journal of Immunology Research 2016. https://doi.org/10.1155/2016/8426874

Xu, Q., Dunbrack, R.L., 2020. ProtCID: a data resource for structural information on protein interactions. Nature Communications 2020 11:1 11, 1–16. https://doi.org/10.1038/s41467-020-14301-4

Xu, S., Liu, P., Chen, Y., Chen, Y., Zhang, W., Zhao, H., Cao, Y., Wang, F., Jiang, N., Lin, S., et al. (2018). Foxp2 regulates anatomical features that may be relevant for vocal behaviors and bipedal locomotion. *Proc. Natl. Acad. Sci*. 115, 8799–8804.

Yaffe, K., Vittinghoff, E., Hoang, T., Matthews, K., Golden, S.H., Zeki Al Hazzouri, A., 2021. Cardiovascular Risk Factors Across the Life Course and Cognitive Decline. Neurology 96, e2212. https://doi.org/10.1212/WNL.0000000000011747

Yan, I., Guo, H., Hu, B., Li, R., Yong, J., Zhao, Y., Zhi, X., Fan, X., Guo, F., Wang, X., Wang, W., Wei, Y., Wang, Y., Wen, L., Qiao, J., Tang, F., 2016. Epigenomic landscape of human fetal brain, heart, and liver. Journal of Biological Chemistry 291, 4386–4398. https://doi.org/10.1074/jbc.M115.672931

Yang, F., Petsalaki, E., Rolland, T., Hill, D.E., Vidal, M., Roth, F.P., 2015. Protein Domain-Level Landscape of Cancer-Type-Specific Somatic Mutations. PLoS Computational Biology 11. https://doi.org/10.1371/JOURNAL.PCBI.1004147

Yang, J., Huang, T., Petralia, F., Long, Q., Zhang, B., Argmann, C., Zhao, Y., Mobbs, C. V., Schadt, E.E., Zhu, J., Tu, Z., Ardlie, K.G., Deluca, D.S., Segrè, A. V., Sullivan, T.J., Young, T.R., Gelfand, E.T., Trowbridge, C.A., Maller, J.B., Tukiainen, T., Lek, M., Ward, L.D., Kheradpour, P., Iriarte, B., Meng, Y., Palmer, C.D., Winckler, W., Hirschhorn, J., Kellis, M., MacArthur, D.G., Getz, G., Shablin, A.A., Li, G., Zhou, Y.H., Nobel, A.B., Rusyn, I., Wright, F.A., Lappalainen, T., Ferreira, P.G., Ongen, H., Rivas, M.A., Battle, A., Mostafavi, S., Monlong, J., Sammeth, M., Mele, M., Reverter, F., Goldman, J., Koller, D., Guigo, R., McCarthy, M.I., Dermitzakis, E.T., Gamazon, E.R., Konkashbaev, A., Nicolae, D.L., Cox, N.J., Flutre, T., Wen, X., Stephens, M., Pritchard, J.K., Lin, L., Liu, J., Brown, A., Mestichelli, B., Tidwell, D., Lo, E., Salvatore, M., Shad, S., Thomas, J.A., Lonsdale, J.T., Choi, C., Karasik, E., Ramsey, K., Moser, M.T., Foster, B.A., Gillard, B.M., Syron, J., Fleming, J., Magazine, H., Hasz, R., Walters, G.D., Bridge, J.P., Miklos, M., Sullivan, S., Barker, L.K., Traino, H., Mosavel, M., Siminoff, L.A., Valley, D.R., Rohrer, D.C., Jewel, S., Branton, P., Sobin, L.H., Qi, L., Hariharan, P., Wu, S., Tabor, D., Shive, C., Smith, A.M., Buia, S.A., Undale, A.H., Robinson, K.L., Roche, N., Valentino, K.M., Britton, A., Burges, R., Bradbury, D., Hambright, K.W., Seleski, J., Korzeniewski, G.E., Erickson, K., Marcus, Y., Tejada, J., Taherian, M., Lu, C., Robles, B.E., Basile, M., Mash, D.C., Volpi, S., Struewing, J., Temple, G.F., Boyer, J., Colantuoni, D., Little, R., Koester, S., Carithers, L.J., Moore, H.M., Guan, P., Compton, C., Sawyer, S.J., Demchok, J.P., Vaught, J.B., Rabiner, C.A., Lockhart, N.C., 2015. Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. Scientific Reports 5. https://doi.org/10.1038/srep15145

Yang, X., Chen, L., Xu, X., Li, C., Huang, C., and Deng, C.-X. (2012). TGF-b/Smad3 Signals Repress Chondrocyte Hypertrophic Differentiation and Are Required for Maintaining Articular Cartilage*. J. Cell Biol*. 153, 1–12.

Yates, C.M., Sternberg, M.J.E., 2013. The Effects of Non-Synonymous Single Nucleotide Polymorphisms (nsSNPs) on Protein–Protein Interactions. Journal of Molecular Biology 425, 3949–3963. https://doi.org/10.1016/J.JMB.2013.07.012

Yoshida, H., Kong, Y.Y., Yoshida, R., Elia, A.J., Hakem, A., Hakem, R., Penninger, J.M., Mak, T.W., 1998. Apaf1 Is Required for Mitochondrial Pathways of Apoptosis and Brain Development. Cell 94, 739–750. https://doi.org/10.1016/S0092-8674(00)81733-X

Yu, G., Wang, L.G., Han, Y., He, Q.Y., 2012a. ClusterProfiler: An R package for comparing biological themes among gene clusters. OMICS A Journal of Integrative Biology 16, 284–287. https://doi.org/10.1089/omi.2011.0118

Yu, G., Wang, L.G., Han, Y., He, Q.Y., 2012b. ClusterProfiler: An R package for comparing biological themes among gene clusters. OMICS A Journal of Integrative Biology 16, 284–287. https://doi.org/10.1089/omi.2011.0118

Yu, G., Wang, L.G., Yan, G.R., He, Q.Y., 2015. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics 31, 608–609. https://doi.org/10.1093/BIOINFORMATICS/BTU684

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355–364.

Yue, P., Li, Z., Moult, J., 2005. Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. Journal of Molecular Biology 353, 459–473. https://doi.org/10.1016/J.JMB.2005.08.020

Yuen, R.K., Neumann, S.M., Fok, A.K., Pëaherrera, M.S., McFadden, D.E., Robinson, W.P., Kobor, M.S., 2011. Extensive epigenetic reprogramming in human somatic tissues between fetus and adult. Epigenetics and Chromatin 4, 7. https://doi.org/10.1186/1756-8935-4-7

YV, S., SL, K., 2010. Identification of epistatic effects using a protein-protein interaction database. Hum Mol Genet 19, 4345–4352. https://doi.org/10.1093/HMG/DDQ356

Zampieri, M., Ciccarone, F., Calabrese, R., Franceschi, C., Bürkle, A., Caiafa, P., 2015. Reconfiguration of DNA methylation in aging. Mechanisms of Ageing and Development. https://doi.org/10.1016/j.mad.2015.02.002

Zehra, R., and Abbasi, A.A. (2018). Homo sapiens-Specific Binding Site Variants within Brain Exclusive Enhancers Are Subject to Accelerated Divergence across Human Population. *Genome Biol. Evol.* 10, 956–966.

Zeileis, A., Kleiber, C., and Jackman, S. (2008). Regression Models for Count Data in {R}. J. Stat. Softw. 27.

Zengini, E., Hatzikotoulas, K., Tachmazidou, I., Steinberg, J., Hartwig, F.P., Southam, L., Hackinger, S., Boer, C.G., Styrkarsdottir, U., Gilly, A., et al. (2018). Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nat. Genet*. 50, 549–558.

Zhai, G., Doré, J., and Rahman, P. (2015). TGF-β signal transduction pathways and osteoarthritis. *Rheumatol. Int.* 35, 1283–1292.

Zhang, F., and Lupski, J.R. (2015). Non-coding genetic variants in human disease. *Hum. Mol. Genet.* 24, R102–R110.

Zhang, W., Qu, J., Liu, G.H., Belmonte, J.C.I., 2020a. The ageing epigenome and its rejuvenation. Nature Reviews Molecular Cell Biology. https://doi.org/10.1038/s41580-019-0204-5

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.

Zhang, Y., Shao, Z., Zhai, Z., Shen, C., Powell-Coffman, J.A., 2009. The HIF-1 Hypoxia-Inducible Factor Modulates Lifespan in C. elegans. PLOS ONE 4, e6348. https://doi.org/10.1371/JOURNAL.PONE.0006348

Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* 30, 1006–1007.

Zhao, H., Zhou, W., Yao, Z., Wan, Y., Cao, J., Zhang, L., Zhao, J., Li, H., Zhou, R., Li, B., et al. (2015). Foxp1/2/4 regulate endochondral ossification as a suppresser complex. *Dev. Biol*. 398, 242–254.

Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B. St., et al. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 33, 272–279.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., Bennett, D.A., Houmard, J.A., Muoio, D.M., Onder, T.T., Camahort, R., Cowan, C.A., Meissner, A., Epstein, C.B., Shoresh, N., Bernstein, B.E., 2013. Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell 152, 642–654. https://doi.org/10.1016/j.cell.2012.12.033