Uncovering Novel Cytogenetic and Molecular Etiologies for Male Infertility

Uncovering novel cytogenetic and molecular etiologies for male infertility

A dissertation presented

by

Samantha Linn Price Schilit

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Genetics and Genomics

Harvard University

Cambridge, Massachusetts

March 2019

Dissertation Advisor: Professor Cynthia Morton Samantha Linn Price Schilit

Uncovering novel cytogenetic and molecular etiologies for male infertility

Abstract

Infertility affects 10-15% of couples, making it one of the most frequent health disorders for reproductive aged individuals. Over 20% of cases are idiopathic, and it is commonly thought that many individuals have an underlying genetic etiology. Identifying genes involved in unexplained infertility provides information to support diagnosis, genetic counseling and potential therapeutic intervention.

One approach to identifying genes involved in infertility is to examine the phenotype-genotype correlation in subjects with a clinical phenotype accompanied by a balanced chromosomal aberration (BCA), as is the foundation of gene discovery in the Developmental Genome Anatomy Project (DGAP). Participant DGAP230 has severe oligospermia and 46,XY,t(20;22)(q13.3;q11.2). While BCAs may reduce fertility by production of unbalanced gametes, a chromosomal rearrangement may also disrupt or dysregulate genes important in fertility.

Using large-insert genome sequencing, chromosomal breakpoints were determined with nucleotide-level precision. Investigation of genes in the topologically associated domains at the sites of the rearrangement and subsequent use of the novel 3C-PCR technique revealed exclusive dysregulation of synaptonemal complex protein 2 (*SYCP2*) from the derivative chromosome 20 (der(20)) allele, which resides 1.5 Mb centromeric to the der(20) breakpoint. 4C-seq from the *SYCP2* promoter revealed interactions eight Mb downstream of the der(20) breakpoint in chromosome 22, supporting a model of enhancer adoption.

SYCP2 encodes synaptonemal complex protein 2, a member of the synaptonemal complex involved in male meiosis I homologous chromosome synapsis. To assess its impact on

meiosis, we misexpressed *RED1*, the functional axial element homolog of *SYCP2* in *Saccharomyces cerevisiae*. By performing Red1 immunolocalization on surface-spread meiotic nuclei, we discovered that excess Red1 forms polycomplexes, disrupting the structural integrity of the synaptonemal complex by preventing incorporation of the transverse filament Zip1. This model highlights aberrant meiosis as the defect in spermatogenesis leading to severe oligospermia. Finally, we reveal two novel frameshift mutations in *SYCP2* identified in azoospermic men, which further support a mechanism of *SYCP2*-mediated infertility.

This thesis provides the first evidence of *SYCP2*-mediated infertility in humans and illuminates that rearrangement breakpoints should be considered as an alternative etiology to that of segregation of unbalanced gametes in infertile men harboring BCAs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

3C: chromosome conformation capture
4C-seq: circular chromatin conformation capture sequencing
aCGH: array-based comparative genomic hybridization
ADGRG2: adhesion G protein-coupled receptor G2
ART: assisted reproductive technologies
AZF: azoospermia factor
BCA: balanced chromosomal aberration
BLAST: Basic local alignment search tool
BLAT: BLAST-like alignment tool
CBAVD: congenital bilateral absence of the vas deferens
CeRA: Centre of Reproductive Medicine and Andrology
CFTR: cystic fibrosis transmembrane conductance regulator
chr20: chromosome 20
chr22: chromosome 22
ClinGen: Clinical Genome Resource
CNV: copy number variant
DAZ: deleted in azoospermia
dbSNP: Database of Single Nucleotide Polymorphisms
DDX3Y: DEAD box polypeptide 3
der(20): derivative chromosome 20
der(22): derivative chromosome 22
DGAP: Developmental Genome Anatomy Project
DIC: differential interference contrast
DSD = disorder of sex development
ES: exome sequencing
FBS: fetal bovine serum
FSH: follicle stimulating hormone
GEMINI: Genetics of Male Infertility Initiative
GFP: green fluorescent protein
gnomAD: Genome Aggregation Database
GnRH: gonadotropin-releasing hormone
GS: genome sequencing
hCG: human chorionic gonadotropin
HPG: hypothalamic-pituitary-gonadal
IACUC: institutional animal care and use committee
IBD: identity by descent
ICSI: intracytoplasmic sperm injection
IMIGC: International Male Infertility Genomics Consortium
IUI: intrauterine insemination
IVF: *in vitro* fertilization
KDM5D: lysine-specific demethylase 5D
LCL: lymphoblastoid cell line
LH: luteinizing hormone
MAF: minor allele frequency
MAR: mixed agglutination reaction
MMAF: multiple morphological abnormalities of the sperm flagella
MSCI: meiotic sex chromosome inactivation
NEB: New England Biolabs
NGS: next-generation sequencing

NIGMS: National Institute of General Medical Sciences
NOA: nonobstructive azoospermia
OA: obstructive azoospermia
OAT: oligoasthenoteratozoospermia
PAS: Periodic acid-Schiff
PBS: phosphate-buffered saline
PCR: polymerase chain reaction
PGD: preimplantation genetic diagnosis
PITT: pronuclear injection-based targeted transgenesis
pLI: probability of loss-of-function intolerance
RIPA: radioimmunoprecipitation assay buffer
RPMI: Roswell Park Memorial Institute
RT-PCR: real-time polymerase chain reaction
SCOS: Sertoli cell-only syndrome
SRY: sex-determining region on the Y chromosome
SYCP2: synaptonemal complex protein 2
SYCP3: synaptonemal complex protein 3
TAD: topologically associating domain
TESE: testicular sperm extraction
TLA: targeted locus amplification
TOPMed: Trans-Omics for Precision Medicine
UCSC: University of California Santa Cruz
UKM: University Hospital Münster
UNMC: University of Nebraska Medical Center
VUS: variants of uncertain significance
WHO: World Health Organization
YCMD: Y chromosome microdeletion

ACKNOWLEDGEMENTS

"*If you put your mind to it, you can accomplish anything.*"

Dr. Emmett Brown, *Back to the Future*

When I was accepted to the Harvard University Division of Medical Sciences Program in Biological and Biomedical Sciences in late February 2012, I could have never anticipated the difficult journey that awaited me. However, the last seven years have offered remarkable gifts. My time at Harvard has validated me as a scientist, given me confidence in my abilities as an independent investigator, helped me make discoveries that I truly believe can improve human health, and prepared me for my dream career. Below, I thank those who have helped me along the way.

Research is a collaborative effort, and I want to acknowledge the two laboratories where I trained during my time in graduate school. I want to thank Jesse Gray for being my principal investigator for the first few years of my graduate school journey. His lab taught me the bench and computational skills required to be an independent investigator. I am so grateful that Cynthia Morton gave me the chance to join her lab as a fourth-year student. She provided the freedom to drive my own project, a process that enabled me to engage passionately and creatively with my work. I'd also like to thank my previous and current labmates: Jameson Aubut, Suresh Boppana, Yvonne Chekaluk, Ozden Altiok Clark, Ben Currall, Scott Gallagher, Anne Giersch, Linda Johnson, Tammy Kammin, Netta Mäkinen, Shreya Menon, Zehra Ordulu, Raul Piña-Aguilar, Nancy Robertson, Weili Shi, Jun Shen, Tracy Song, Care Tangshewinsirikul, Ellen Wilch and Cinthya Zepeda Mendoza.

Several other researchers also helped nurture my project into the satisfying story it is today. My dissertation advisory committee, including chair Raju Kucherlapati and committee members Ray Anchan, David Page, and Jim Gusella, provided constructive criticism and helped

to refine next steps at every meeting. I appreciate the dissertation examination committee including chair Raju Kucherlapati and examiners Mitzi Kuroda, Matthew Meselson, and Kellee Siegfried for their generous time in reviewing the body of work I have completed for this thesis. Thank you to the Developmental Genome Anatomy Project (DGAP), which served as the infrastructure for my dissertation work. In addition to Cynthia Morton, I want to thank the other principal investigators on this project, Mike Talkowski, Jim Gusella and Eric Liao, for their collaborative spirit and interest in my work. Former DGAP coordinators Tammy Kammin and Ellen Wilch were instrumental in providing me with the necessary information to work on the DGAP230 case, and Carrie Hanscom is responsible for resolving DGAP230's breakpoints.

Beyond DGAP, I owe a huge thank you to Shreya Menon, an undergraduate researcher who has worked with me for the past two years and made the analysis of my 4C-seq data possible. It has been an honor to work with such a talented and motivated scientist. I also recognize Sizun Jiang, who created a protocol for CRISPR/Cas9-mediated genome editing in lymphoblastoid cells and assisted in planning and troubleshooting my enhancer causality experiments. The most remarkable collaborator on this dissertation work was Amy MacQueen, a professor from my alma mater Wesleyan University, who specializes in the yeast synaptonemal complex. Amy provided tremendous help modeling axial element overexpression in *S. cerevisiae*, making the strains as a didactic tool in her undergraduate class and inviting me to stay with her in Connecticut as I completed my experiments. Understanding the molecular mechanisms by which *SYCP2* overexpression leads to severe oligospermia would have been impossible without her enormous contribution, and I still can't believe how lucky I am that the DGAP230 project led me back to yeast research at Wesleyan.

I want to thank my collaborators who continue to push forward the future directions of this project. To test the hypothesis that *SYCP2* overexpression leads to severe oligospermia and male infertility in a mammalian system, Masato Ohtsuka and Channabasavaiah Gurumurthy have been using their well-established pronuclear injection-based targeted transgenesis (PITT) technique to

make a transgenic Sycp2 overexpression mouse with inducible, ubiquitous and stable expression of Sycp2. While I initiated the project by designing and testing the construct for this mouse and receiving IACUC approval for the protocol, the project will be taken over by Raul Piña-Aguilar. Thank you to Masato, Guru, and Raul for their hard work and time as they continue this important work. In an extension to my thesis project, we are exploring the prevalence of *SYCP2*-mediated infertility in the general population. To accomplish this goal, I have established relationships with several different male infertility cohorts to look for *SYCP2* pathogenic variants and interesting variants of uncertain significance. I want to acknowledge the leaders of these cohorts: Frank Tüttelmann (International Male Infertility Genomics Consortium [IMIGC]), Don Conrad (Genetics of Male Infertility Initiative [GEMINI], Washington University School of Medicine and Oregon Health & Science University, USA), Dan Zhang (Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China), Digumarthi V.S. Sudhakar and Kumarasamy Thangaraj (Centre for Cellular and Molecular Biology, Hyderabad, India) and Raul Piña-Aguilar (Instituto de Ciencias en Reproduccion Humana [Vida], León, Guanajuato, Mexico). I want to highlight the work of Frank Tüttelmann, Corinna Friedrich, and Sabine Kliesch, who discovered the two additional infertile males with *SYCP2* pathogenic variants through the IMIGC. Shreya Menon has helped to optimize a functional assay to determine pathogenicity of *SYCP2* variants of uncertain significance. I appreciate these collaborators for their commitment to making my project's future directions a reality.

My graduate career would have been incomplete without the Program in Genetics and Genomics (PGG) led by Fred Winston and the Leder Human Biology and Translational Medicine Program (LHBTM) led by Connie Cepko and Thomas Michel. These programs gave me the supplemental resources necessary to develop my own specific training in clinical genetics. I am so grateful for the additional mentorship from Fred and Connie, who became my "BBS parents" during challenging times in graduate school. Thank you for being trustworthy, ethical, and my advocates when I needed support. I received additional encouragement from Dean David

Cardozo, who treated me with respect, dignity and concern when I encountered my lowest moments in graduate school. I also thank the Personal Genetics Education Project (pgEd), especially director Marnie Gelbart and former curriculum director Lauren Tomaselli, who helped me connect with genetics on a broader level and gave me a platform to engage with the public directly on issues related to clinical genetics.

My graduate school trajectory would not have been possible without my funding sources, the National Science Foundation Graduate Research Fellowship (DGE1144152) and the Ruth L. Kirschstein National Research Service Award Individual Predoctoral Fellowship from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (F31HD090780-01). As always, any opinion, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of my funding agencies.

Thanks to the friends and communities who helped me maintain my sanity and happiness these past seven years. I'm grateful for my BBS classmates, especially Avery Davis Bell, Mary Gearing, Kelsey Tyssowski, and Rebecca Fine, who were the first responders when urgent need struck for an Aliquots café warm chocolate chip cookie and a companion to share it with. I owe my mental, physical and emotional health to the wonderful Harvard Dragon Boat, Dudley Rowing, and Camp Casco organizations. There's no better stress relief than watching sunrise or sunset on the Charles River with your closest friends and there's no better way to add perspective to your life than working with resilient and inspiring childhood cancer survivors.

This thesis is written in memory of my father Robert Schilit, who died just six weeks before I moved to Massachusetts to start my PhD. My dad was my greatest advocate and encouraged my academic pursuits. He knew of my plans to attend Harvard, and every success that followed in graduate school has been celebrated in memory of him. In the wake of my Dad's passing, I have developed enormous gratitude for the growth of my nuclear family, including the addition of my mom's husband Marty, brother-in-law Adam, niece Ava, nephew Matthew, fiancé Jeremy, and in-laws-to-be Lila, Marc and Aaron. These family members have been my lifeline during graduate

school. My mom has dropped everything to support me in times of crisis, Adam has commiserated with me as we both faced struggles in our respective research experiences, Ava and Matthew have brightened every day since their births, and Jeremy's parents have offered advice before interviews and presentations.

I dedicate this thesis to my twin sister, Arielle Nitenson, and my fiancé, Jeremy Rosenweig. Rel, you have always been my role model. After exiting from the womb two minutes before me, you continued to do everything first, including publishing research articles and getting your PhD. At 28 years old, you became a mother and an assistant professor! While at times it has been intimidating to compare our situations, I am so proud of you and the courage it has taken to be a trailblazer. You inspire me to work hard and find balance, just as you have done in your own life. Thanks for entertaining my crazy idea to publish *My Identical Twin Sequenced Our Genome*, which ended up being a highlight in my graduate career. In addition, thank you for editing this entire thesis at eight months pregnant. Finally, I can't thank you enough for bringing the most incredible children into this world. I am so proud to be Aunt Sammi!

Jeremy, I can't begin to express my gratitude for how much you have enriched my life. We have created a truly equal partnership, where we support each other personally and professionally. You use the perfect mix of humor and compassion to ground me when I'm stressed and have enabled me to prioritize completing this degree by taking on a disproportionate amount of joint responsibilities (including wedding planning!). Thanks for giving me the opportunity to dream big as I complete my PhD and begin an intense fellowship program. I love the life we have created with our two fuzzy sons, Sgt. Pepper and Herbie, and our third pet Oscar, the meyer lemon tree. I cannot wait to marry you just weeks after submitting this dissertation, so we can finally become a family. I love you so much!

CHAPTER 1

INTRODUCTION

1.1. Infertility: a common disorder with widespread consequences

According to the American Society for Reproductive Medicine, infertility is a disorder "defined by the failure to achieve a successful pregnancy after 12 months or more of appropriate, timed, unprotected intercourse or therapeutic donor insemination."[1] It affects 10-15% of couples, making it one of the most common disorders for individuals between the ages of 20-45 years[2].

Infertility can cause substantial emotional, social, and financial stress on couples. The majority of young Americans view parenthood as a future desired state[3]. The inability to meet this expectation leads to a variety of reactions including negative identity, a sense of inadequacy, a feeling of lack of personal control, grief and sense of loss, anger and resentment, anxiety and stress, low life satisfaction, depression, isolation and shame[4-7].

These negative consequences extend beyond the individual. Within a couple, infertility-related stress can lead to marital distress and dissatisfaction[8]. Infertility in heterosexual couples affects each partner differently: more women have concerns about a loss of emotional intimacy in their relationship, while men more frequently experience a lack of sexual satisfaction due to pressure to perform for conception over pleasure[9]. Infertility may also cause challenges to the couple within the context of their society. In some pronatalist cultures including those from Israel, Pakistan, and southern Africa, adult status is obtained by bearing children[6]. Childlessness leads to ostracism from the community, including grounds for divorce in some Bangladeshi settlements[6].

There is also a tremendous financial burden to accessing fertility care and assisted reproductive technologies (ART), such as intrauterine insemination (IUI) and *in vitro* fertilization (IVF). In the United States, most patients pay out-of-pocket for treatments, as only six states mandate insurance coverage for infertility treatments[10]. The median out-of-pocket price for infertility treatments including couples receiving either non-cycle-based treatment, medication only, IUI, or IVF is $5,338[11]. The most expensive intervention is IVF, which has a median out-of-

pocket expense of $19,234 for one cycle, representing about half of an individual's annual disposable income[11,12]. In addition to treatment costs, time spent on office visits adds up to an average of two work weeks a year[13]. The problem worsens when considering that ART interventions may take a long time to be successful; at least six cycles with timed intercourse are necessary to cover the period in which 90% of conceptions will occur[14]. IUI and IVF are also used for couples with unexplained infertility, where the likelihood of achieving pregnancy is lower[15].

The high cost for treatment serves as a barrier for access to care. In the United States, couples earning a cumulative income of less than $100,000 are more likely to be dissuaded from choosing IVF intervention[11]. This leads to the ethical concern that the high out-of-pocket costs for infertility treatment discriminate against groups of lower socioeconomic status[10]. Moreover, disparities exist between the outcomes of infertility treatment; women of African American, Asian, and Hispanic backgrounds experience longer time to conception, lower implantation and clinical pregnancy rates, and higher miscarriage from ART than their Caucasian counterparts[16-18].

Men comprise the largest understudied group in infertility research and treatment, despite the fact that 40-50% of infertile couples have male factor infertility[19,20]. While intracytoplasmic sperm injection (ICSI) and IVF now provide a more direct role for men in infertility treatment, women's bodies have been the focus of most medical interventions. The clinical file is sometimes linked with the woman's medical record and not her partner's, and men aren't always a part of an infertility consultation[20,21]. The disproportionate focus of the role of the woman in fertility often results from a male dominant culture which blames women for fertility challenges as a way of deflecting a threat to male masculinity, potency and virility[21-23]. Like most women, the majority of men desire parenthood and expect to be fathers[24,25]. Facing an infertility diagnosis results in profound grief, loss of control, a sense of inadequacy, and isolation[26]. The exclusion of men from most scientific and psychosocial literature in the context of infertility means that less is known about the mechanisms underlying and consequences of male infertility[20]. This highlights an unmet need to understand male infertility.

1.2. An overview of the adult male reproductive system

Male fertility relies on the successful production of healthy gametes called spermatozoa. The male reproductive system can be broken into three distinct modules responsible for sperm development: pre-testicular, testicular, and post-testicular[2]. Pre-testicular contributions to gametogenesis rely on a properly functioning hypothalamic-pituitary-gonadal (HPG) axis, which uses hormones to signal initiation of gametogenesis in the testes[27]. The HPG axis begins with the hypothalamic secretion of gonadotropin-releasing hormone (GnRH), which promotes production of luteinizing hormone (LH) and follicle stimulating hormone (FSH) from the anterior pituitary. LH stimulates Leydig cells in the testes to release testosterone, which together with FSH communicate with Sertoli cells in the testes to support sperm production[28]. Testosterone, along with inhibin B, facilitate negative feedback of the HPG axis by controlling activity from the hypothalamus and pituitary[27-29].

Spermatogenesis occurs in the seminiferous tubules of the testes (Figure 1.1). Germ cells develop in a spatially organized fashion from the basal membrane to the lumen[30]. Spermatogenesis involves the differentiation of primordial germs cells (spermatogonia) to mature sperm (spermatozoa) in a process that takes approximately 70 days[28]. Spermatogonia undergo several rounds of mitosis to renew the germ cell population while creating cells capable of differentiating. Some of these daughter cells develop into diploid primary spermatocytes, which become haploid secondary spermatocytes after the first meiotic division. Secondary spermatocytes differentiate into spermatids after a second meiotic division, which then mature into spermatozoa through a process called spermiogenesis[30]. Finally, mature sperm are released from the seminiferous epithelium through a process called spermiation[31].

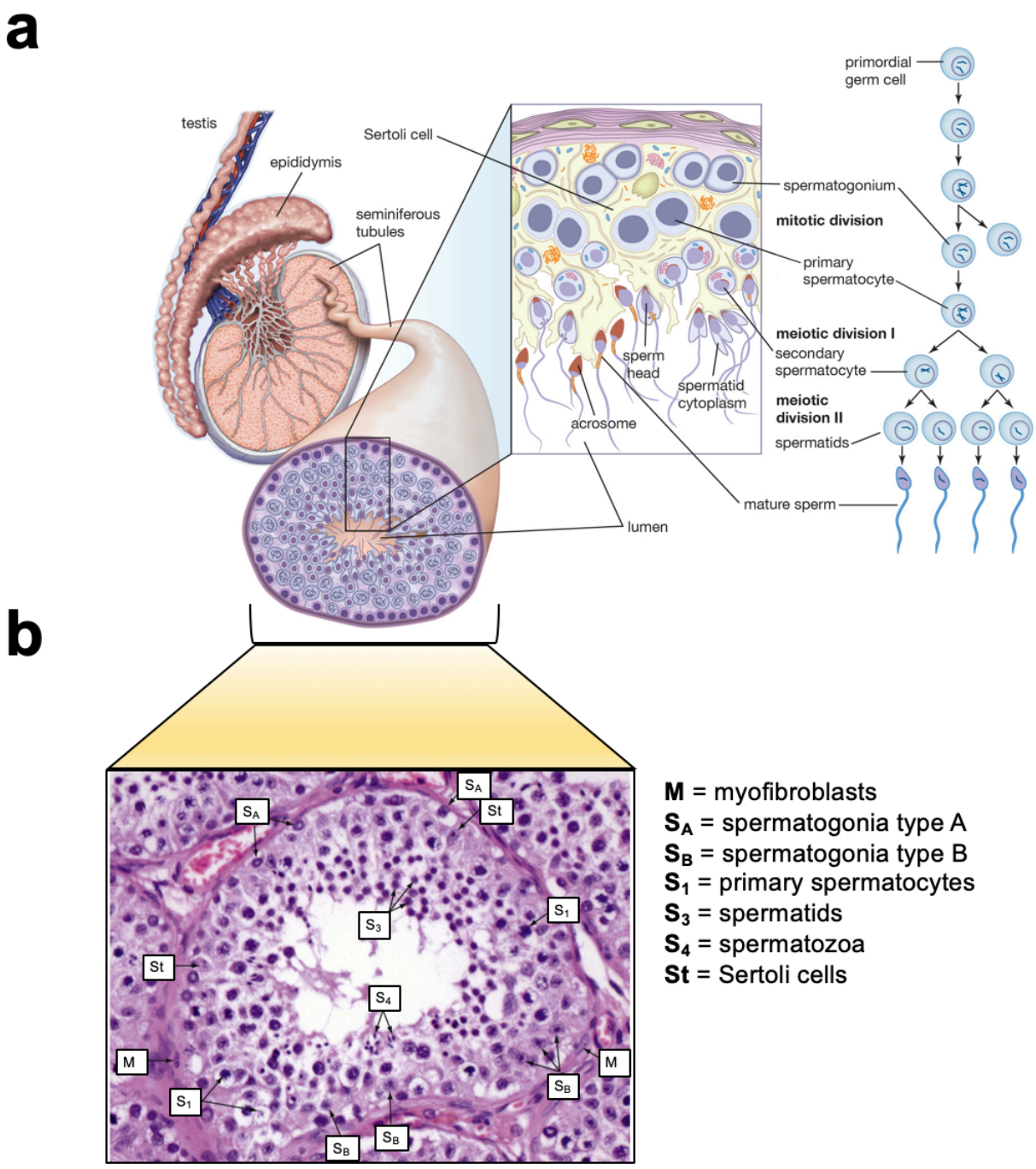


Figure 1.1. Anatomy of healthy adult seminiferous tubules
a) Spermatogenesis takes place in the seminiferous tubules of adult testes. The magnification of a seminiferous tubule wedge highlights various cell types created in the differentiation of primordial germ cells to mature sperm (adapted from [32]). **b)** Histology of a transverse section of a seminiferous tubule reveals spatial organization of sperm development (H & E X640, adapted from [30]).

The post-testicular phase involves final processing of semen for export out of the body through ejaculation. First, sperm in the seminiferous tubules are transported to the epididymis to undergo functional development through a process called epididymal maturation[31]. Next, semen enters the vas deferens and travels to the ejaculatory ducts, where it is combined with secretory products of male accessory glands including the seminal vesicle and the prostate[27]. The final ejaculate, which is composed of 10% sperm by volume, exits through the urethra[27,28]. Sperm undergo a final maturation step called capacitation, which occurs in the female reproductive tract and enables sperm to develop a hyperactivated form of forward progressive motility, acrosome reaction, and the ability to penetrate and fertilize the egg[31].

1.3. Diagnostic work-up for male infertility

A systematic diagnostic work-up is necessary to identify the best treatment options for infertile couples with male factor infertility. The World Health Organization's (WHO) current recommendation involves an initial evaluation including medical history, physical exam and semen analysis followed by referral to a urologist, an andrologist, or another male reproduction specialist in cases of abnormal findings (Figure 1.2)[33]. The medical history in the initial evaluation first collects information about the reproductive history such as coital frequency, previous fertility, information about the partner's fertility, and sexual history including sexually transmitted infections[33]. Patients are then asked about lifestyle factors, including BMI, smoking, and exposure to heat, which have been shown to influence semen parameters although diagnosing infertility by these factors is still controversial[33,34]. An understanding of medications taken can also inform a diagnosis as different drugs can reduce fertility by cytotoxicity from radiation with chemotherapy or inhibition of the HPG axis through exogenous testosterone[2]. Finally, questions are asked relating to anatomical dysfunction including proximal trauma, surgeries including vasectomies, torsion, cryptorchidism, erectile and ejaculatory dysfunction, and genitourinary infections[35].

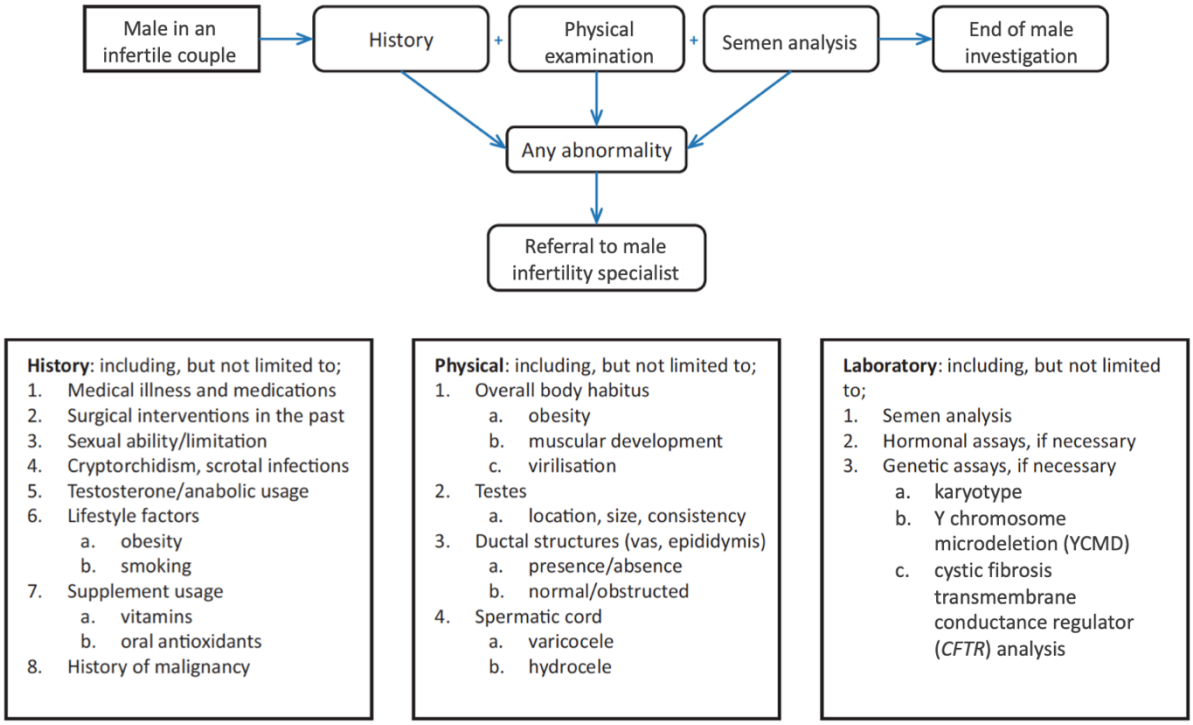


Figure 1.2. Flowchart for diagnosing male infertility
Current recommendations from the WHO suggest an initial evaluation before referral to a male infertility specialist. Initial diagnoses are made based upon information from medical history, physical examination, and semen analysis. Future hormonal tests, imaging and genetic assays may be warranted (adapted from [33]).

After obtaining a medical history, signs of endocrine-related disorders, scrotal pathologies and penile pathologies are investigated by physical exam[2]. A look at secondary sexual characteristics including musculature, hair distribution, and breast tissue can suggest a pre-testicular etiology including testosterone deficiency or hormonal imbalance[35]. A genital exam explores scrotal pathologies including varicoceles and small testicular size, which when atrophied below 12 ml may indicate primary testicular failure[2,35]. Palpation of other structures can indicate forms of post-testicular obstruction including absence of the vas deferens[2,35,36]. Finally, examination of the penis may reveal pathologies such as hypospadias that may indicate potential challenges with sperm placement into the vagina[35].

The final part of an initial evaluation is a semen analysis, as multiple semen parameters can be predictive of testicular production, function, and maturation[33,37,38]. The work-up includes a single collection of ejaculate after two to seven days of abstinence and assesses parameters including volume, pH, and sperm concentration, count, motility, morphology, and vitality[39]. A deviation from normal parameters can provide insight into a diagnosis (Table 1.1). Normozoospermia is described when the routine spermatozoa evaluation shows values above the lower reference value limits. Alternatively, results below the lower reference values for sperm volume, concentration, motility, and morphology describe hypospermia, oligozoospermia, asthenozoospermia, and teratozoospermia, respectively. Low sperm counts can further be delineated further based upon sperm concentrations less than five million/ml ejaculate (severe oligozoospermia), spermatozoa absent from fresh semen but visible in a pellet after centrifugation (cryptozoospermia), and the complete absence of sperm in the ejaculate (azoospermia)[40].

Parameters can also help explain the etiology of infertility. For example, after ruling out collection error, a low ejaculate volume may suggest potential retrograde ejaculation, ejaculatory duct obstruction, prostatitis, inflammation of the seminal vesicles, or androgen deficiency[36,41,42]. A high concentration of white blood cells and very basic pH measurements can indicate infections[2]. Alternatively, agglutination of motile spermatozoa and confirmation of anti-sperm antibodies by

mixed agglutination reaction (MAR) or immunobead testing support immunologic infertility[40]. Finally, low concentrations of seminal fluid markers including α-glucosidase, fructose, and zinc reflect potential obstruction in the epididymis, seminal vesicles, and prostate, respectively[2].

If an abnormality in sperm count or concentration is detected by initial evaluation and confirmed by a second semen analysis, then hormone analysis may be warranted to pinpoint further the etiology[33]. The measurement of total testosterone concentration and serum FSH can help distinguish pre-testicular from testicular and post-testicular etiologies[43,44]. Low levels of testosterone (below 8-12 nmol/l) and associated low serum FSH levels (<1 mIU/ml) suggest hypogonadotropic hypogonadism, which can be treated with human chorionic gonadotropin (hCG) or FSH[2,35,45]. Testosterone can also be abnormally high in pre-testicular etiologies if the patient uses exogenous testosterone or illicit anabolic androgenic steroids, in which case cessation of use can sometimes restore sperm production[36,45,46]. Alternatively, high FSH (>7-7.6 mIU/ml) often indicates primary testicular failure, with more severe defects including Sertoli cell-only syndrome (SCOS) correlating with more elevated FSH[2,36]. However, FSH may also be normal with testicular pathologies that occur later in spermatogenesis such as meiotic arrest[2,44]. This may be hard to distinguish from post-testicular etiologies, which often show normal hormone levels[2,36]. To distinguish further the etiologies as a way of identifying proper treatments for a patient, a genetic work-up is then recommended[44].

Table 1.1. Lower reference limits of semen parameters
Semen parameter reference limits have been adapted from [2,39,40].

Parameter	Lower reference limit values
Macroscopic appearance	
Semen volume	≥ 1.5 ml
pH	≥ 7.2
Routine spermatozoa evaluation	
Total sperm count	≥ 39 million/ejaculate
Sperm concentration	≥ 15 million/ml
Total motility	≥ 40% progressive and nonprogressive
Progressive motility	≥ 32% fast and slow
Sperm morphology	≥ 4% normal
Follow-up testing	
White blood cells (used if round cells are found on initial microscopic evaluation)	< 1 million/ml peroxidase-positive leukocytes
Mixed agglutination reaction (MAR) test or immunobead test (used if agglutination is found on initial microscopic evaluation)	< 50%
Vitality (used if total motility is < 40%)	≥ 58% live spermatozoa
Biochemical analysis of seminal fluids	
Zinc	≥ 2.4 µmol/ejaculate
Fructose	≥ 13 µmol/ejaculate
α-glucosidase	≥ 20 mU/ejaculate

1.4. Diagnostic genetic testing for male infertility

Currently, the most common genetic testing for male infertility involves sequencing of the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, assessing Y chromosome microdeletions (YCMD), and karyotype analysis[2,33,35,36,47-51].

Variants in *CFTR* are found in 80-90% of cases of congenital bilateral absence of the vas deferens (CBAVD), a condition that occurs in 1-2% of infertile males and 25% of men with obstructive azoospermia (OA)[47,52]. While men harboring two variant *CFTR* alleles with severe functional defects have symptomatic cystic fibrosis, compound heterozygotes with less severe functional defects in *CFTR* or men with a single pathogenic allele may have only the CBAVD phenotype[52]. Absence of the vas deferens may be detected by physical exam, but a genetic understanding of the etiology is critical for proper treatment of a male with *CFTR* variants. While males with CBAVD resulting from *CTFR* pathogenic variants can have children with the assistance of ART such as testicular sperm extraction (TESE) followed by ICSI and IVF, genetic testing of *CFTR* is recommended for the biological mother. Carrier screening resulting in positive findings for the couple offers the subsequent option of preimplantation genetic diagnosis (PGD) to eliminate the risk of transmitting cystic fibrosis to offspring[33,47,52]. While *ADGRG2* (encoding the epididymal- and efferent-ducts-specific adhesion G protein-coupled receptor G2) was recently identified as a second gene with variants causal for CBAVD when mutated, testing for this gene is not yet a routine part of diagnostic genetic testing for OA[53].

Males with nonobstructive azoospermia (NOA) or severe oligospermia should be tested for 0.8-7.7 Mb deletions in the azoospermia factor (AZF) region of the Y chromosome (or YCMD), which are found in 5-15% of these patients[33,50,54]. The AZF region resides in Yq11 and is subdivided into AZFa (0.8 Mb), AZFb (6.2 Mb) and AZFc (3.5 Mb)[48,50,55]. The AZF region is critical for fertility because several genes responsible for spermatogenesis map therein including DEAD box polypeptide 3 (*DDX3Y*) in AZFa, lysine-specific demethylase 5D (*KDM5D*) in AZFb, and deleted in azoospermia (*DAZ*) in AZFc[56,57]. These genes play a variety of roles in sperm

development including RNA metabolism in pre-meiotic germ cells, chromatin remodeling in meiosis, and translation regulation, respectively[56,57]. In order to identify which if any YCMDs are present, polymerase chain reaction (PCR) assays are performed for markers inside the region and flanking the borders of each AZF subregion[50,58]. The diagnosis is critical because treatment options vary based upon the deletion. No spermatozoa will be found from TESE performed on men with complete AZFa, AZFb, or AZFb/c microdeletions, so treatment alternatives including donor sperm and adoption might be considered[33,50,59]. Alternatively, males with complete AZFc deletions may have some residual spermatogenesis, with a 50% success rate from TESE[47,60]. Sperm cryopreservation is also warranted as sperm production decreases with age in men with AZFc deletions[60]. This knowledge is also important for genetic counseling, as any male offspring conceived from a male with an AZFc deletion will inherit the same Y chromosome. Thinking ahead, sperm cryopreservation in young adulthood for male offspring might be recommended in anticipation of decreasing spermatogenesis with age[33].

Finally, karyotyping, which assesses the number and structure of chromosomes, is also recommended for all men with NOA or severe oligospermia, as 15% of men with NOA and 4% of men with moderate to severe oligospermia have chromosomal abnormalities[33,43,50]. The most common finding for men with NOA is Klinefelter syndrome (47,XXY and variants such as 48,XXXY and 46,XY/47,XXY), which is found in 14% of their karyotypes[50,61,62]. This diagnosis is helpful for predicting the prognosis of TESE, as sperm retrieval has been successful in 40-50% of men with Klinefelter syndrome[63,64]. Another sex chromosome abnormality that may be found is 46,XX testicular disorder of sex development (DSD), or de la Chapelle syndrome, which has a rarer frequency of 1 in 20,000[50,65,66]. In most cases, the male phenotype results from a paternal translocation of the gene *SRY* (sex-determining region on the Y chromosome) from the short arm of the Y chromosome to the short arm of the X chromosome[65]. 46,XX testicular DSD males lack germinal cells, so TESE is not advisable[50,65,66].

Balanced chromosomal aberrations (BCAs), which have an abnormal order of the

chromosomes without any cytologically detectable gain or loss of genetic material, are found five to ten times more frequently in infertile men than in the general population[58,67]. BCAs can be categorized into Robertsonian translocations, reciprocal translocations involving a sex chromosome, reciprocal translocations with autosomes, insertions, and inversions[67]. Of these types of BCAs, reciprocal translocations involving sex chromosomes often result in more severe phenotypes with a higher incidence in azoospermic men than in oligospermic men[67]. Y;autosome (Y;A) translocations identified in azoospermic men often involve a breakpoint in Yq11, which disrupts the AZF region critical for spermatogenesis[67]. X;autosome (X;A) translocations are thought to impact fertility severely because they may lead to X-reactivation during meiotic prophase, disrupting critical meiotic sex chromosome inactivation (MSCI)[48,68,69].

The mechanism is less clear for other classes of BCAs, which have extremely variable outcomes ranging from azoospermia to normal semen parameters[47]. There is a general assumption that individuals with BCAs produce unbalanced gametes as a product of meiosis, which are selected against during spermatogenesis resulting in a lower sperm count and subsequent infertility[48,58]. However, this isn't the entire truth. While carriers of BCAs are more likely to have low sperm counts than karyotypically normal men, there is no significant relationship between fertility and sperm counts above 20 million/ml, so slight decreases in sperm count above that level don't imply a decrease in fertility[38]. In one cohort from the Czech Republic, the average sperm count of men with reciprocal translocations was 66.5 million/ml compared to 72.7 million/ml in controls, and less than 3% of reciprocal translocation carriers had a sperm count of <5 million/ml[70]. In a Japanese cohort that used a cutoff of <5 million/ml, there was no significant enrichment of BCA carriers in the NOA or severe oligospermia groups compared to controls[71]. It is also true that men with BCAs have more signs of spermatocyte apoptosis, such as externalized phosphatidylserine and DNA fragmentation[72], but this isn't necessarily due to selection against unbalanced gametes. In male BCA carriers, the distributions of meiotic segregation products at different spermatogenic stages show concordance, suggesting that there is no cellular selection

based on chromosomal imbalances from post-meiotic spermatocytes to mature spermatozoa[73].

Despite the misinterpretation that carriers of BCAs have reduced fertility due to unbalanced gametes decreasing sperm count, it is true that unbalanced gametes double the risk of miscarriages[62,74]. In 3-5% of couples with recurrent miscarriages, at least one partner is found by karyotype analysis to have a balanced reciprocal translocation[74]. In addition, unbalanced gametes can lead to congenital malformations in surviving offspring[62]. As a result, identification of a BCA can alter treatment options, as PGD with IVF provides identification of balanced or normal embryos prior to transfer[33,50].

1.5. Unexplained infertility: challenges and opportunities

While many factors contribute to infertility including anatomical defects, hormone dysregulation, environmental exposures, age and certain genetic syndromes, at least one in five cases of infertility are unexplained[48]. For male infertility, 40-72% of men lack a specific causal diagnosis beyond a descriptive category of male factor infertility[19,47,51]. Genetic defects may be responsible for many of these idiopathic cases. Indeed, mutations in over 200 genes have been shown to decrease fertility in animal models, and yet few genetic causes of infertility have been validated in humans[75]. This may result from the decreased reproductive fitness of infertile individuals that reduces the number of large families available for genetic analysis in humans as well as the genetic heterogeneity of the disorder[51,76].

Given that 30-50% of these cases are estimated to have genetic etiologies, identifying genes involved in unexplained infertility could be a rich area of study[47,51]. Uncovering these novel causes not only informs an understanding of mechanisms regulating fertility, but also provides clinical information to support diagnoses, genetic counseling, and therapeutic intervention.

While not currently a routine diagnostic for male infertility, the application of array-based comparative genomic hybridization (aCGH) to investigate copy number variants (CNVs) in subjects with male infertility has revealed novel variants on both sex chromosomes and autosomes that are risk factors for or causative for spermatogenic failure[77,78]. In addition, with the development of large scale sequencing approaches through next-generation sequencing (NGS) and subsequent genome-wide approaches in both small case studies and large consortia including the Genetics of Male Infertility Initiative (GEMINI) and the International Male Infertility Genomics Consortium (IMIGC), some success has been made in identifying monogenic forms of male infertility (Table 1.2)[50,51,58,76,79-83].

Table 1.2. Monogenic causes of non-syndromic male infertility in humans
This list of genes has been self-curated for at least moderate evidence of gene-disease association according to the Clinical Genome Resource (ClinGen) framework[84]. OMIM numbers are written in parentheses next to the gene symbol.

Nonobstructive azoospermia (NOA) or severe oligospermia	**Morphological and/or functional anomalies**
CCDC155 (618125)[76]	**Acephalic spermatozoa**
DBY (400010)[85]	*PMFBP1* (618085)[86]
DNAH6 (603336)[87]	*SUN5* (613942)[88]
FANCM (609644)[89]	
HIWI (605571)[90]	**Asthenozoospermia**
KLHL10 (608778)[91]	*CATSPER1* (606389)[92]
MCM8 (608187)[93]	*SLC26A8* (608480)[94]
MEIOB (617670)[87]	
NANOS2 (608228)[76]	**Globozoospermia**
PLK4 (605031)[95]	*DPY19L2* (613893)[96,97]
SPINK2 (605753)[98]	*SPATA16* (609856)[99]
SPO11 (605114)[76]	
SYCE1 (611486)[100]	**Macrozoospermia**
SYCP3 (604759)[101]	*AURKC* (603495)[102]
TAF4B (601689)[103]	
TDRD7 (611258)[104]	**Multiple morphological abnormalities of the sperm flagella (MMAF)**
TDRD9 (617963)[105]	
TEX11 (300311)[106,107]	*ARMC2*[108]
TEX14 (605792)[76,87]	*CFAP43* (617558)[109]
TEX15 (605795)[110]	*CFAP44* (617559)[109]
WNK3 (300358)[76]	*CFAP69* (617949)[111]
XRCC2 (600375)[112,113]	*DNAH1* (603332)[114,115]
ZMYND15 (614312)[103]	
	Oligoasthenoteratozoospermia (OAT)
	CDC14A (603504)[116]
	SEPT12 (611562)[117,118]
	Oocyte activation failure
	PLCZ1 (608075)[119]

While mostly confined to the research realm, these discoveries will hopefully be employed clinically as they may be informative for predicting therapeutic outcomes in patients[47,120]. For example, men with *AURKC* mutations have sperm that are often polyploid[102]. Due to the high risk of aneuploidies from even normal-appearing spermatozoa, ICSI is not recommended for these patients[120]. Alternatively, men with *DPY19L2* or *SPATA16* pathogenic variants have globozoospermia characterized by acrosome-deficient sperm[96,99]. Because of their inability to activate oocytes, artificial oocyte activation can improve outcomes for fertilization rate, embryo formation and clinical pregnancy with ICSI and IVF[121,122]. This is in contrast to men who have multiple morphological abnormalities of the sperm flagella (MMAF) with mutations from *CFAP43, CFAP44*, *CFAP69,* or *DNAH1*, where ICSI without any additional activation procedure is expected to have a high rate of success[47,109,111,115].

Given that only about 4% of causal genetic diagnoses have been established, there is great need to both identify additional genetic etiologies for male infertility and to improve recommendations for diagnostics in male infertility. Future use of aCGH, NGS, or gene panels may establish a definitive causal diagnosis, offer prognostic value for TESE and clinical pregnancy, and assess risks for potential offspring[51].

1.6. Questions addressed in this thesis

One approach to identifying genes involved in infertility is to examine the phenotype-genotype correlation in subjects with unexplained infertility accompanied with *de novo* BCAs, as is outlined in the Developmental Genome Anatomy Project (DGAP). One participant in this study, designated DGAP230, has severe oligospermia and an abnormal karyotype, 46,XY,t(20;22)(q13.3;q11.2). Based upon these observations, we hypothesize that the translocation breakpoints disrupt or dysregulate fertility genes near the breakpoints. The following questions are addressed in this thesis. First, can we identify a candidate gene disrupted or dysregulated by breakpoints of DGAP230's chromosome rearrangement? Next, what is the mechanism by which the candidate gene is misexpressed? Finally, what is the impact of this candidate gene on severe oligospermia?

Upon successful completion of this project, we will have identified a gene implicated in unexplained male infertility, which will both improve our understanding of mechanisms regulating fertility and demonstrate an additional mechanism for BCA-mediated infertility.

CHAPTER 2

A BALANCED TRANSLOCATION IN DGAP230 RESULTS IN *SYCP2* DYSREGULATION BY AN ENHANCER ADOPTION MECHANISM

2.1. Background

The Developmental Genome Anatomy Project (DGAP) is an NIH-funded research study that identifies genes disrupted or dysregulated by chromosomal rearrangements in subjects with a clinical finding presumed to have a genetic etiology. Genes disrupted or dysregulated by such rearrangements are a well-recognized paradigm in human genetics for underlying abnormal phenotypes (for example, Appendix A)[123-132]. By using the well-established DGAP infrastructure, this project aims to identify novel genes important for fertility and to explore another explanation for how BCAs reduce fertility in infertile patients (Figure 2.1). To identify infertile patients for this study who are most likely to have a disrupted or dysregulated fertility gene as a result of their BCA, eligibility criteria were developed to rule out patients with explained infertility or evidence that their infertility is solely caused by the production of unbalanced gametes (Table 2.1).

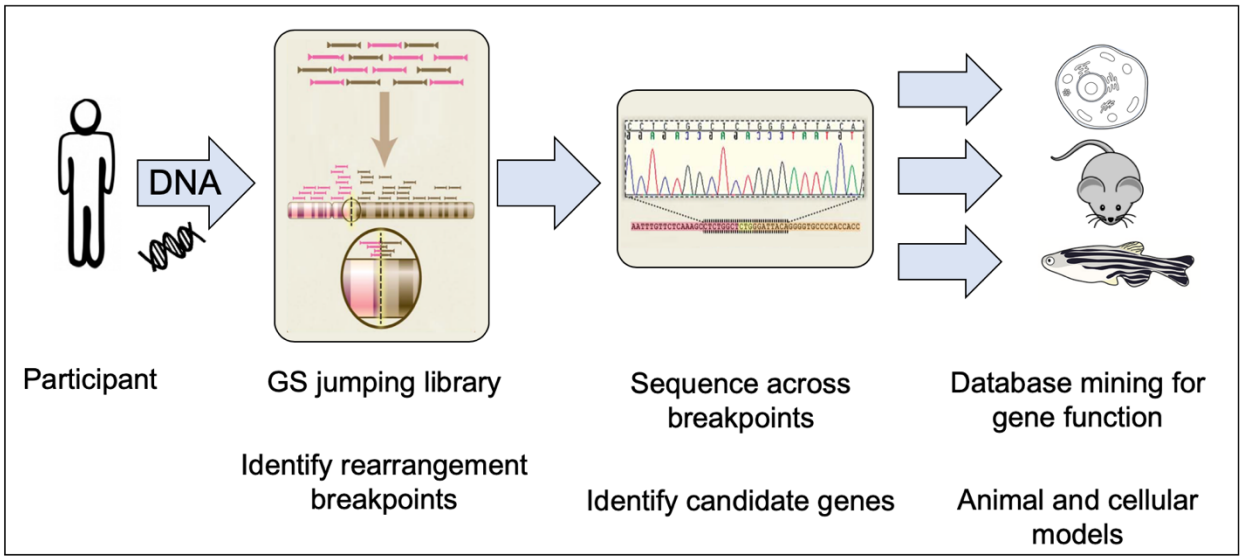


Figure 2.1. Developmental Genome Anatomy Project pipeline
The Developmental Genome Anatomy Project (DGAP) enrolls participants with BCAs and congenital or distinct phenotypes. Genome sequencing (GS) on jumping libraries is performed on DNA collected from the participant to identify rearrangement breakpoints to 2-3 kilobase resolution. Sanger sequencing across the breakpoints then pinpoints the genomic locations to nucleotide resolution, which directs an investigation of candidate genes that may be disrupted or dysregulated by the breakpoints. From there, candidate genes are assessed for causality in order to implicate the gene in the participant's phenotype.

Table 2.1. Eligibility criteria for recruiting infertile DGAP participants

Inclusions	Exclusions

Participants must have:
• Met the criteria for infertility according to the American Society for Reproductive Medicine.
• A karyotype confirming a BCA, such as a translocation or an inversion.
• In the case of an achieved pregnancy, no liveborn, as well as a confirmed euploid conceptus.
	Participants must not have:
	• Any live births.
	• Explained genetic or environmental etiology.
	• An inherited translocation, unless it is inherited from the participant's parent of the opposite sex.
	• A common heteromorphism as the sole BCA such as:
	  ○ Robertsonian translocations
	  ○ The following inversions: inv(9)(p12q13), inv(Y)(p11q11)
	  ○ Variants interpreted to be clinically insignificant:
	  ▪ Variants of constitutive heterochromatin length (1qh+, 9qh+, 16qh+, Yqh+)
	  ▪ Variants of short (p) arm length on chromosomes 13, 14, 15, 21 and 22 (13 ps+, 13pstk+, 14ps+, 14pstk+, 15ps+, 15pstk+, 21ps+, 21pstk+, 22ps+, 22pstk+)

2.2. Methods

2.2.1. DGAP participant recruitment

DGAP enrolled participants DGAP230, with 46,XY,t(20;22)(q13.3;q11.2) and DGAP278-02, a karyotypically normal age- and sex-matched control, by obtaining informed consent, medical records and blood samples under a protocol approved by the Partners HealthCare Systems Institutional Review Board.

2.2.2. Acquisition of lymphoblastoid cell lines (LCLs)

Epstein-Barr virus-transformed LCLs were generated at the Genomics and Technology Core in the Center for Human Genetic Research at Massachusetts General Hospital (Boston, MA, USA). Confirmatory GTG-banded karyotyping was performed on the DGAP230 LCL at the Brigham and Women's Hospital CytoGenomics Core Laboratory (Boston, MA, USA). Two additional karyotypically normal age- and sex-matched control LCLs, GM20184 and GM20188, were obtained from the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Institute for Medical Research (Camden, NJ, USA).

2.2.3. Large-insert jumping library sequencing and analysis

Large-insert ("jumping library") GS and subsequent Sanger sequencing identified the precise breakpoints of the DGAP230 chromosomal rearrangement (using primers SS_58F-SS_61R listed in Table 2.2)[128,133,134]. Chromatograms were analyzed in Geneious (Version 7.0, Biomatters) and described using next-generation cytogenetic nomenclature[135].

2.2.4. Delineation of topologically associating domains (TADs) disrupted by DGAP230's rearrangement breakpoints

TADs disrupted by the breakpoints in DGAP230 were identified according to human embryonic stem cell Hi-C domains from the Hi-C project[136]. The University of California Santa Cruz (UCSC) Genome Browser was used to delineate genes residing in these regions[137].

2.2.5. Culturing of LCLs

LCLs were grown at 37°C, 5% CO_2, in R10 medium (Roswell Park Memorial Institute [RPMI] 1640 Medium without L-Glutamine supplemented with 10% heat-inactivated fetal bovine serum (FBS), 100 U/ml Penicillin-Streptomycin-Glutamine, and 2 mM L-glutamine [all Gibco]). Cells were fed fresh R10 every two or three days as needed. For clonal expansion of single lymphoblastoid cells in the CRISPR/Cas9-mediated genome editing experiment, 30% 0.45 µm-filtered and irradiated conditioned R10 with 45% fresh R10 and 25% FBS was used and feeding only occurred to offset media evaporation.

2.2.6. RNA extraction, cDNA synthesis, and quantitative real-time polymerase chain reaction (RT-PCR) of lymphoblastoid cells

Forty to fifty million cells were collected for each RNA extraction. Cells were pelleted for 5 minutes at 200 x g, washed in phosphate-buffered saline (PBS; Gibco), and resuspended in 1 ml TRIzol (Invitrogen). Cell lysis and phase separation with chloroform (Sigma) were performed as described from the TRIzol manufacturer protocol. Total RNA was isolated using the RNeasy Midi Kit (Qiagen) with in-column RNase-free DNase I (Qiagen) digestion. RNA was converted to cDNA using 4 µg of RNA per reaction, random hexamer primers, and the SuperScript III First-Strand Synthesis System (Invitrogen) according to manufacturer instructions. For quantitative RT-PCR, cDNA was amplified with primers SS_36F-SS_55R and SS_62F-SS_147R (exon-spanning when possible; sequences reported in Table 2.2) and SsoFast EvaGreen Supermix (Bio-Rad) using the

CFX Connect Real-Time PCR Detection System (Bio-Rad). Standard curves were employed routinely to quantify amplicons from each primer pair and assess expression of each respective gene in lymphoblastoid cells.

2.2.7. Identification and evaluation of variable regions in *SYCP2* and *GAPDH*

SYCP2 and *GAPDH* exons were compared to the Database of Single Nucleotide Polymorphisms (dbSNP) to identify variable nucleotides[138]. To assess heterozygosity in DGAP230, genomic DNA of DGAP230 lymphoblastoid cells that had been pelleted and washed with PBS (Gibco) was extracted using the DNeasy Blood and Tissue Kit (Qiagen). PCR was performed with LongAmp Taq 2X Master Mix (New England Biolabs, NEB) using customized primers (Integrated DNA Technologies). After amplification confirmation with agarose gel electrophoresis, and purification with the QIAquick PCR Purification Kit (Qiagen), Sanger sequencing reactions of PCR products were carried out with an ABI3730xl DNA analyzer. Chromatograms were aligned and multiple single nucleotide variants were called using Geneious (Version 7.0, Biomatters). The target exonic regions were selected based upon presence of multiple single nucleotide variants (amplified and Sanger sequenced by primers SS_174F-SS_175R and SS_192F-SS_193R given in Table 2.2). DGAP230 lymphoblastoid cell cDNA was acquired as described above (Section 2.2.6) and assessed by the same method as the genomic DNA (amplified and Sanger sequenced with primers SS_175R-SS_193R as reported in Table 2.2).

2.2.8. 3C-PCR

3C-PCR was performed as described in Chapter 3[139]. In brief, standard chromosome conformation capture (3C) libraries were generated[140-144] and subjected to rearrangement-spanning nested PCR and subsequent Sanger sequencing using primers SS_216R-SS_227R (listed in Table 2.2).

2.2.9. Circular chromatin conformation capture sequencing (4C-seq) and analysis

4C-seq was adapted from previously described protocols with the following specifications[141-143]. Ten million cell aliquots of lymphoblastoid cells were crosslinked by adding formaldehyde (MilliporeSigma) to a final concentration of 2% for 10 minutes before the reaction was quenched with glycine (Promega) at a final concentration of 125 mM. After lysis, chromatin was further released by douncing with a 21.5-gauge needle and digested using 1500 U *Hin*dIII (NEB) per reaction at 37°C overnight. After enzyme inactivation, ligation, reverse crosslinking, DNA isolation, and purification, DNA was subjected to a second digestion by 50 U *Csp*6I (ThermoFisher Scientific) per reaction at 37°C overnight. Enzyme inactivation, ligation, DNA isolation and purification were next performed to create the 4C libraries. Each library was amplified by inverse PCR with primers containing multiplexed overhangs (SS_198F-SS_212R as reported in Table 2.2) using 50 ng of 4C template per reaction and purified by Qiaquick PCR purification kit (Qiagen). Samples were submitted to The Biopolymers Facility in the Department of Genetics at Harvard Medical School (Boston, MA, USA) for quality control on an Agilent 2200 TapeStation D1000 HS ScreenTape and by SYBR qPCR assay. Samples were spiked with 30% PhiX prior to sequencing on an Illumina HiSeq 2500 sequencer using the HiSeq Single-Read Rapid Cluster Kit v2 (Illumina) for 100 cycles. At least 14 million pass-filter reads were acquired per sample.

4C-seq datasets were imported into the Galaxy-BioTeam Appliance (https://gbsc-galaxy.stanford.edu) and demultiplexed using Barcode Splitter (Galaxy Version 1.0.0)[145]. Reads were aligned to GRCh37/hg19 and a custom-made derivative chromosome 20 (der(20)) hg19 reference genome using Bowtie2 (Galaxy Version 0.2). The der(20) was made using a Perl script that parsed through chromosomes 20 and 22 and concatenated parts of the chromosomes at the resolved breakpoints (Section 2.2.11). Bam files were then analyzed using FourCSeq (Version 1.16.0) with normal parameters established by the FourCSeq protocol with the exception of using

the der(20) chromosome for some analyses[146]. Interactions between fragments were then mapped to a circular plot using Circos (Version circos-0.69-6), by applying the relative coordinates of each fragment within the respective chromosome[147]. P-adjusted significance values were used to superimpose a colored heatmap onto the plot. These were then transformed into linear plots using Photoshop (Version CC 2015) by converting the image from Rectangular mode to Polar mode.

2.2.10. CRISPR/Cas9-mediated genome editing in lymphoblastoid cells

CRISPR/Cas9-mediated deletion of chromosome 22 (chr22) putative enhancers in the DGAP230 LCL was performed as previously described with the following parameters[148]. A Cas9-expressing stable DGAP230 cell line was developed by transduction with lentiCas9-Blast lentivirus, selection in blasticidin S HCl (Gibco), and validation by Western blot with primary antibodies anti-Cas9 (Abcam, catalog number 7A9-3A3) and anti-GAPDH (Cell Signaling Technology, catalog number 5174) and secondary antibodies IRDye 680RD Donkey anti-Mouse IgG and IRDye 800CW Donkey anti-Rabbit IgG (both LI-COR). This cell line was transduced with lentivirus derived from lentiviral vectors containing custom-made dual targeting sgRNAs that flank the putative enhancer region as well as a lentiGuide-Puro empty vector control. The dual targeting sgRNA vector was constructed and validated with primers SS_278F-SS_289F (given in Table 2.2) and the previously described H1 bridge dsDNA block[148]. Selection of infected cells in puromycin (Gibco) yielded heterogeneous cultures, which were then either grown for one to four months or sorted into single cells in 96-well cell culture plates (Corning) using a MoFlo *Astrios* EQ cell sorter with a 100 µm nozzle at the Harvard Medical School Division of Immunology Flow Cytometry Core (Boston, MA, USA). Single cell clones were PCR-validated for homozygous deletions after genomic DNA extraction with the DNeasy Blood and Tissue Kit (Qiagen) by using LongAmp Taq 2X Master Mix (NEB) and primers SS_304F-SS_309R (given in Table 2.2). To determine the relative amount of putative enhancer deletion in heterogeneous cultures derived

from the dual targeting sgRNA lentivirus transduction and grown for different lengths of time, quantitative RT-PCR of 12 ng genomic DNA was performed with primers SS_308F-SS_309R and SS_52F-SS_53R (sequences reported in Table 2.2) and SsoFast EvaGreen Supermix (Bio-Rad) using the CFX Connect Real-Time PCR Detection System (Bio-Rad). Standard curves were employed routinely to quantify amplicons from each primer pair.

2.2.11. Code availability

The GetDerivative.pl script used to create the der(20) and der(22) can be found at https://github.com/mnnshreya/Morton-DGAP230.

2.2.12. Statistical analyses

For quantitative RT-PCR experiments, cDNA for three DGAP230 LCL replicates and three distinct age- and sex-matched control LCLs were each assessed in technical duplicates, normalized to *GAPDH* expression, and evaluated using an unpaired two-tailed t-test (Excel). Statistical significance for RNA analyses were determined by a p-value of $p<0.05$ and figure error bars show standard error of the mean. For identifying and assessing variable regions in DGAP230 lymphoblastoid cells, Sanger sequencing was performed on at least three different preparations of gDNA and/or cDNA. 3C-PCR and 4C-seq were each performed in triplicate, creating both 3C and 4C libraries from three independent cultures for the DGAP230 LCL and three different control LCLs. For differential analysis of 4C-seq datasets using FourCSeq (Version 1.16.0)[146], read-counts were variance-stabilized and fit with a monotonic decay that was used to calculate z-scores. Z-scores were converted to p-values using a normal cumulative distribution curve and adjusted for multiple testing by the Benjamini-Hochberg method. A differential analysis was then conducted and significant interactions were detected by a p-adjusted value of $p_{adj}<0.01$. For quantitative RT-PCR of the putative enhancer deletion, genomic DNA for two biological replicates of DGAP230 LCLs transduced with the dual targeting sgRNA lentivirus and grown for one month,

15 biological replicates of DGAP230 LCLs transduced with the dual sgRNA lentivirus and each grown for four months, and two biological replicates of DGAP230 LCLs transduced with the lentiGuide-Puro empty vector control were each assessed in technical triplicates, normalized to an intronic region of genomic DNA, and evaluated using a heteroscedastic one-tailed t-test for one versus four months and homoscedastic one-tailed t-test for dual targeting sgRNA lentivirus versus empty vector control (Excel). Statistical significance for quantitative RT-PCR analyses on genomic DNA were determined by a p-value of $p<0.05$ and figure error bars show standard error of the mean.

Table 2.2. Primers used in chapter 2

Name	Sequence 5'-3'	Target region	Description
SS_36F	TTCACCACCATGGAGA AGGC	*GAPDH* cDNA, human	For qPCR control with SS_37R; amplicon = 112 bp
SS_37R	TCTCATGGTTCACACCC ATGAC	*GAPDH* cDNA, human	For qPCR control with SS_36F; amplicon = 112 bp
SS_40F	TTGGAAAAGGGACAGC CAAG	*SYCP2* cDNA, human	For qPCR with SS_41R; amplicon = 108 bp
SS_41R	GGTTGCTTTTCGTGGA AGTCTG	*SYCP2* cDNA, human	For qPCR with SS_40F; amplicon = 108 bp
SS_52F	CAGACTCCTGTGGTCA AGCA	*ATP2B1,* human	For qPCR with SS_53R; amplicon = 147 bp (control primer pair published in [149])
SS_53R	TTCGTCAGTCAACCCCT TTC	*ATP2B1,* human	For qPCR with SS_52F; amplicon = 147 bp (control primer pair published in [149])
SS_54F	TGACCACGTTTTCAGCT GTG	*CDH4* cDNA, human	For qPCR with SS_55R; amplicon = 103 bp
SS_55R	TTGGTGGCATTGATGT GCAG	*CDH4* cDNA, human	For qPCR with SS_54F; amplicon = 103 bp
SS_58F	GTTCGTGTCTCAGGTT CAGCCAG	upstream of DGAP230 rearrangement breakpoint on chr20	For amplification with SS_59R of der(20) breakpoint; amplicon = 976 bp (also used for subsequent Sanger sequencing)
SS_59R	GCACAGTTTTGATCCTG TCTTGTGG	downstream of DGAP230 rearrangement breakpoint on chr22	For amplification with SS_58F of der(20) breakpoint; amplicon = 976 bp (also used for subsequent Sanger sequencing)
SS_60F	GATAAGCCAATAACCA CGACCTGAG	upstream of DGAP230 rearrangement breakpoint on chr22	For amplification with SS_61R of der(22) breakpoint; amplicon = 832 bp (also used for subsequent Sanger sequencing)
SS_61R	GGATTAGGACAGGCAG GAGCAAG	downstream of DGAP230 rearrangement breakpoint on chr20	For amplification with SS_60F of der(22) breakpoint; amplicon = 832 bp (also used for subsequent Sanger sequencing)
SS_62F	TGCCACAATGCACAGA CAAC	*CDH26* cDNA, human	For qPCR with SS_63R; amplicon = 93 bp
SS_63R	TCTTCCAGCACATTGG CAAC	*CDH26* cDNA, human	For qPCR with SS_62F; amplicon = 93 bp
SS_70F	CCACCCAAACCTTAGA AAGCTG	*SNAP29* cDNA, human	For qPCR with SS_71R; amplicon = 100 bp

SS_71R	TCGAAGGTGTGGGTTC TTTG	*SNAP29* cDNA, human	For qPCR with SS_70F; amplicon = 100 bp
SS_76F	ACCAGCAACAGTATCG GGAAG	*SLC7A4* cDNA, human	For qPCR with SS_77R; amplicon = 85 bp
SS_77R	AGGCAGATGTTGAGGA CGATG	*SLC7A4* cDNA, human	For qPCR with SS_76F; amplicon = 85 bp
SS_92F	ACAGGTCAAGGTGTTC AACG	*PPP1R3D*, human	For qPCR with SS_93R; amplicon = 123 bp
SS_93R	AGGCAATGCAGGGTGA ACTC	*PPP1R3D*, human	For qPCR with SS_92F; amplicon = 123 bp
SS_94F	AGCACTACAGGAAAGA CAGTCC	*FAM217B* cDNA, human	For qPCR with SS_95R; amplicon = 120 bp
SS_95R	TGGGCCAGCATTCATA TTGC	*FAM217B* cDNA, human	For qPCR with SS_94F; amplicon = 120 bp
SS_98F	TTGTGAAATCGCGTGG ACTG	*C20orf197* cDNA, human	For qPCR with SS_99R; amplicon = 72 bp
SS_99R	TACCAACAGCTGCTCA GTGG	*C20orf197* cDNA, human	For qPCR with SS_98F; amplicon = 72 bp
SS_102F	AACATGATGCTGCCCA ACTG	*USP41* cDNA, human	For qPCR with SS_103R; amplicon = 138 bp
SS_103R	TGGCACAGTCAAGGCA AATC	*USP41* cDNA, human	For qPCR with SS_102F; amplicon = 138 bp
SS_106F	ATGGAGCATGCAGAGG GAAG	*ZNF74* cDNA, human	For qPCR with SS_107R; amplicon = 87 bp
SS_107R	TTGCAAATGCCCTGCT GTTG	*ZNF74* cDNA, human	For qPCR with SS_106F; amplicon = 87 bp
SS_110F	GCATGCCTGTAACCAT GTCAC	*SCARF2* cDNA, human	For qPCR with SS_111R; amplicon = 89 bp
SS_111R	TGCCATTGCTACACTTG GTC	*SCARF2* cDNA, human	For qPCR with SS_110F; amplicon = 89 bp
SS_114F	AACAACGATGCCGGAT ACAG	*KLHL22* cDNA, human	For qPCR with SS_115R; amplicon = 130 bp
SS_115R	TGTTGTCCAGCACAGC AATG	*KLHL22* cDNA, human	For qPCR with SS_114F; amplicon = 130 bp
SS_118F	AGCCATGTTTTCCTGAA GGC	*MED15* cDNA, human	For qPCR with SS_119R; amplicon = 137 bp
SS_119R	AGGCTCTGGAGTGCAT TCATAG	*MED15* cDNA, human	For qPCR with SS_118F; amplicon = 137 bp
SS_122F	AAGCGCCACTTCTCAG AAAC	*SERPIND1* cDNA, human	For qPCR with SS_123R; amplicon = 149 bp
SS_123R	TGAGCAGTTTCCCCTC CTTTC	*SERPIND1* cDNA, human	For qPCR with SS_122F; amplicon = 149 bp
SS_126F	CAGGTTTTGCATGTCCT CTTGG	*PI4KA* cDNA, human	For qPCR with SS_127R; amplicon = 122 bp
SS_127R	TTGCTGTAACGCCCAA ATGC	*PI4KA* cDNA, human	For qPCR with SS_126F; amplicon = 122 bp
SS_130F	AACCGCCGTTTTAAGAT CGG	*CRKL* cDNA, human	For qPCR with SS_131R; amplicon = 128 bp

SS_131R	ATTGGTGGGCTTGGATACCTG	*CRKL* cDNA, human	For qPCR with SS_130F; amplicon = 128 bp
SS_134F	TGAGGCCCAAGGAGTTTTTC	*AIFM3* cDNA, human	For qPCR with SS_135R; amplicon = 108 bp
SS_135R	TGAAGCCATCCTTGAACACG	*AIFM3* cDNA, human	For qPCR with SS_134F; amplicon = 108 bp
SS_138F	AAGTTTGCAACTGGCCAGTG	*LZTR1* cDNA, human	For qPCR with SS_139R; amplicon = 109 bp
SS_139R	CAGCAAAGATCCACAGCTTGTC	*LZTR1* cDNA, human	For qPCR with SS_138F; amplicon = 109 bp
SS_142F	AAAGGACACAGTTACCCACCTG	*THAP7* cDNA, human	For qPCR with SS_143R; amplicon = 104 bp
SS_143R	GGTGGAGAAAATGGAGTTGTGG	*THAP7* cDNA, human	For qPCR with SS_142F; amplicon = 104 bp
SS_146F	TGACCAACTTCCTTGTGACG	*P2RX6* cDNA, human	For qPCR with SS_147R; amplicon = 134 bp
SS_147R	TTTTACACCGTGGCTGTGTG	*P2RX6* cDNA, human	For qPCR with SS_146F; amplicon = 134 bp
SS_174F	TTGTGGACTCAACCCTTAGTCA	*SYCP2*, human	For PCR with SS_175R; amplicon = 2945 bp
SS_175R	GGACTTCCTCCCCCTTGTAA	*SYCP2*, human	For PCR with SS_174F; amplicon = 2945 bp (also used for subsequent Sanger sequencing)
SS_185F	GCTTGTGCGGATAGGTCAAT	*SYCP2* cDNA, human	For PCR with SS_186R; amplicon = 2404 bp
SS_186R	TGTTTCCAACAGTGTGCTGA	*SYCP2* cDNA, human	For PCR with SS_185F; amplicon = 2404 bp
SS_192F	TGAGCAGTCCGGTGTCACTA	*GAPDH*, human	For PCR with SS_193R; amplicon = 619 bp for cDNA and 859 bp for gDNA (also used for subsequent Sanger sequencing)
SS_193R	TGACTCCGACCTTCACCTTC	*GAPDH*, human	For PCR with SS_192F; amplicon = 619 bp for cDNA and 859 bp for gDNA
SS_198F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAGGCGAGGGGTGCTGAGCTGGGTAC	*SYCP2* promoter, human	For 4C-seq inverse PCR with SS_212R; multiplexed for library "DGAP230_A2" using the barcode TAAGGCGA
SS_199F	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACTAGGGGGTGCTGAGCTGGGTAC	*SYCP2* promoter, human	For 4C-seq inverse PCR with SS_212R; multiplexed for library "GM20184_B2" using the barcode CGTACTAG

SS_200F	AATGATACGGCGACCA CCGAACACTCTTTCCCT ACACGACGCTCTTCCG ATCTAGGCAGAAGGGG TGCTGAGCTGGGTAC	*SYCP2* promoter, human	For 4C-seq inverse PCR with SS_212R; multiplexed for library "DGAP230_B1" using the barcode AGGCAGAA
SS_201F	AATGATACGGCGACCA CCGAACACTCTTTCCCT ACACGACGCTCTTCCG ATCTTCCTGAGCGGGG TGCTGAGCTGGGTAC	*SYCP2* promoter, human	For 4C-seq inverse PCR with SS_212R; multiplexed for library "DGAP230_D3" using the barcode TCCTGAGC
SS_202F	AATGATACGGCGACCA CCGAACACTCTTTCCCT ACACGACGCTCTTCCG ATCTGGACTCCTGGGG TGCTGAGCTGGGTAC	*SYCP2* promoter, human	For 4C-seq, inverse PCR with SS_212R; multiplexed for library "GM20188_A1" using the barcode GGACTCCT
SS_203F	AATGATACGGCGACCA CCGAACACTCTTTCCCT ACACGACGCTCTTCCG ATCTTAGGCATGGGGG TGCTGAGCTGGGTAC	*SYCP2* promoter, human	For 4C-seq inverse PCR with SS_212R; multiplexed for library "DGAP278-02_A" using the barcode TAGGCATG
SS_212R	CAAGCAGAAGACGGCA TACGACTTGGCCTTTCA AGCCGGC	*SYCP2* promoter, human	For 4C-seq inverse PCR with SS_198F-SS_203F
SS_216R	GAAGAGGAGCTTCTTA ATGTACGC	*SYCP2*, human	For 3C-PCR second amplification with SS_220F; amplicon = 511 bp
SS_220F	ATTGCTTGAACCAGGA GGTG	chr22, proximal to rearrangement breakpoint	For 3C-PCR second amplification with SS_216R; amplicon = 511 bp (also used for subsequent Sanger sequencing)
SS_225F	TTTCTCTCCTGTTCCCA AGG	chr22, proximal to rearrangement breakpoint	For 3C-PCR first amplification with SS_227R; amplicon = 1843 bp
SS_227R	AGGTTGATCCTTGTTGA AATTGTT	*SYCP2*, human	For 3C-PCR first amplification with SS_225F; amplicon = 1843 bp
SS_278F	AAAGGACGAAACACCG CAGATACGAACAAAGA ATCGGTTTTAGAGCTAG AAATAGCAAG	Plasmid lentiGuide-Puro and H1 bridge dsDNA block (underlined sequences); chr22 putative enhancer target region	For cloning of dual targeting sgRNAs into plasmid lentiGuide-Puro (Addgene plasmid #52963), used with SS_287R
SS_287R	TTCTAGCTCTAAAACCA TTGTGGATCTGAGTGG	Plasmid lentiGuide-Puro and H1 bridge	For cloning of dual targeting sgRNAs into plasmid

	GAGGATCCAAGGTGTCTCATAC	dsDNA block (underlined sequences); chr22 putative enhancer target region	lentiGuide-Puro (Addgene plasmid #52963), used with SS_278F
SS_289F	GAGGGCCTATTTCCCATGATT	Plasmid lentiGuide-Puro	For Sanger sequencing of sgRNAs inserted into plasmid lentiGuide-Puro (Addgene plasmid #52963)
SS_304F	CCTTTATCCCTAGGCAGCGT	chr22, 5' flanking region for chr22 putative enhancer deletion	For PCR with SS_305R; amplicon = 342 bp in cells with at least one intact chr22 putative enhancer
SS_305R	CAATCGCTTCAACTCCACCC	chr22, 5' flanking region for chr22 putative enhancer deletion	For PCR with SS_304F; amplicon = 342 bp in cells with at least one intact chr22 putative enhancer
SS_306F	TCTTCTCCCTGGGCATGAAC	chr22, 3' flanking region for chr22 putative enhancer deletion	For PCR with SS_307R; amplicon = 236 bp in cells with at least one intact chr22 putative enhancer
SS_307R	AGAGCCAGGATAAGACTTGAGT	chr22, 3' flanking region for chr22 putative enhancer deletion	For PCR with SS_306F; amplicon = 236 bp in cells with at least one intact chr22 putative enhancer
SS_308F	TCCTTTATCCCTAGGCAGCG	chr22, across chr22 putative enhancer deletion	For PCR and qPCR with SS_309R; amplicon = 233 bp in cells with at least one deletion of the chr22 putative enhancer
SS_309R	TGACGATTCAGAGATTTGTGACT	chr22, across chr22 putative enhancer deletion	For PCR and qPCR with SS_308F; amplicon = 233 bp in cells with at least one deletion of the chr22 putative enhancer

2.3. Results

2.3.1. Clinical report for DGAP230

This study focuses on identifying the genetic etiology of infertility for one male research participant, designated DGAP230, who presented with a two-year history of infertility at age 28. His evaluation showed severe oligospermia (<2 million/ml) with normal semen volume. Karyotyping revealed the reciprocal translocation 46,XY,t(20;22)(q13.3;q11.2) (Figure 2.2a). This was considered balanced, as aCGH identified no cytologically detectable gains or losses of genomic material at either breakpoint. Genetic testing for YCMD and *CFTR* mutations were negative. DGAP230 has no dysmorphic features and has normal serum levels of FSH, LH and testosterone. Given DGAP230's normal hormonal levels, normal semen volume and absence of *CFTR* mutations, we predicted that the etiology of his severe oligospermia likely originates in the testes[35]. Based on these initial observations, we hypothesized that the t(20;22) breakpoints disrupt or dysregulate genes, which may be important in fertility, perhaps by influencing gametogenesis.

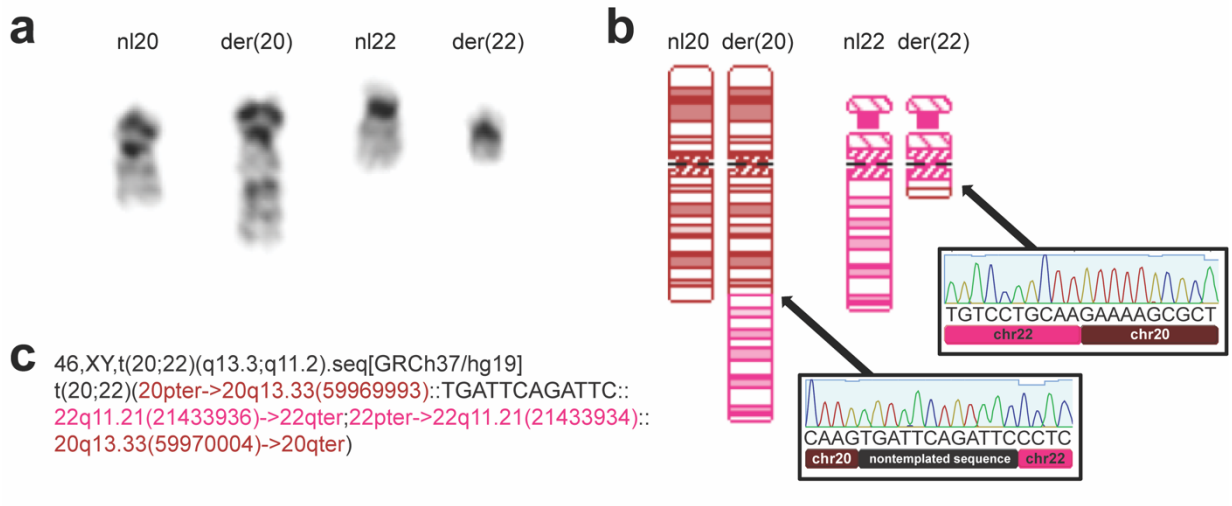


Figure 2.2. DGAP230 breakpoint characterization
a) DGAP230 composite partial karyotype of normal (left) and derivative (right) GTG-banded chromosomes 20 and 22. **b)** The DGAP230 ideogram indicates a simple long arm (q-q) translocation between chromosomes 20 (nl20) and 22 (nl22). Chromatograms from Sanger sequencing show rearrangement breakpoints at nucleotide level resolution. **c)** Next-Generation Cytogenetic Nomenclature is provided for DGAP230[135]. The following software was used to develop this figure: http://www.geneious.com, http://www.cydas.org/OnlineAnalysis/, http://boston.bwh.harvard.edu/.

2.3.2. Identification of candidate genes for DGAP230's severe oligospermia

To identify genes disrupted or potentially dysregulated by the t(20;22) breakpoints, DGAP230's translocation was elucidated, directing an investigation of candidate genes. We refined DGAP230's chromosomal breakpoints to nucleotide resolution using large-insert ("jumping library") sequencing and subsequent Sanger sequencing (Figure 2.2b,c).

DGAP230's breakpoint locations were then used to identify candidate genes that may cause his severe oligospermia. Genetic regions specifically disrupted by the t(20;22) could be impacted in two ways. If the breakpoint occurs within the open reading frame of any gene, the gene is considered disrupted. Alternatively, positional effects on genes near breakpoints could occur due to loss or gain of regulatory elements in the genetic neighborhood. We defined the extent of these effects by the breakpoint-residing topologically associating domains (TADs), compartments of chromatin with high interactions in three-dimensional space, because TAD disruption by structural rearrangements has been shown to rewire gene expression and induce pathogenicity[136,150,151] (Figure 2.3). As a result, our list of candidate genes includes all genes that are disrupted or potentially dysregulated by residing within the breakpoint-containing TADs.

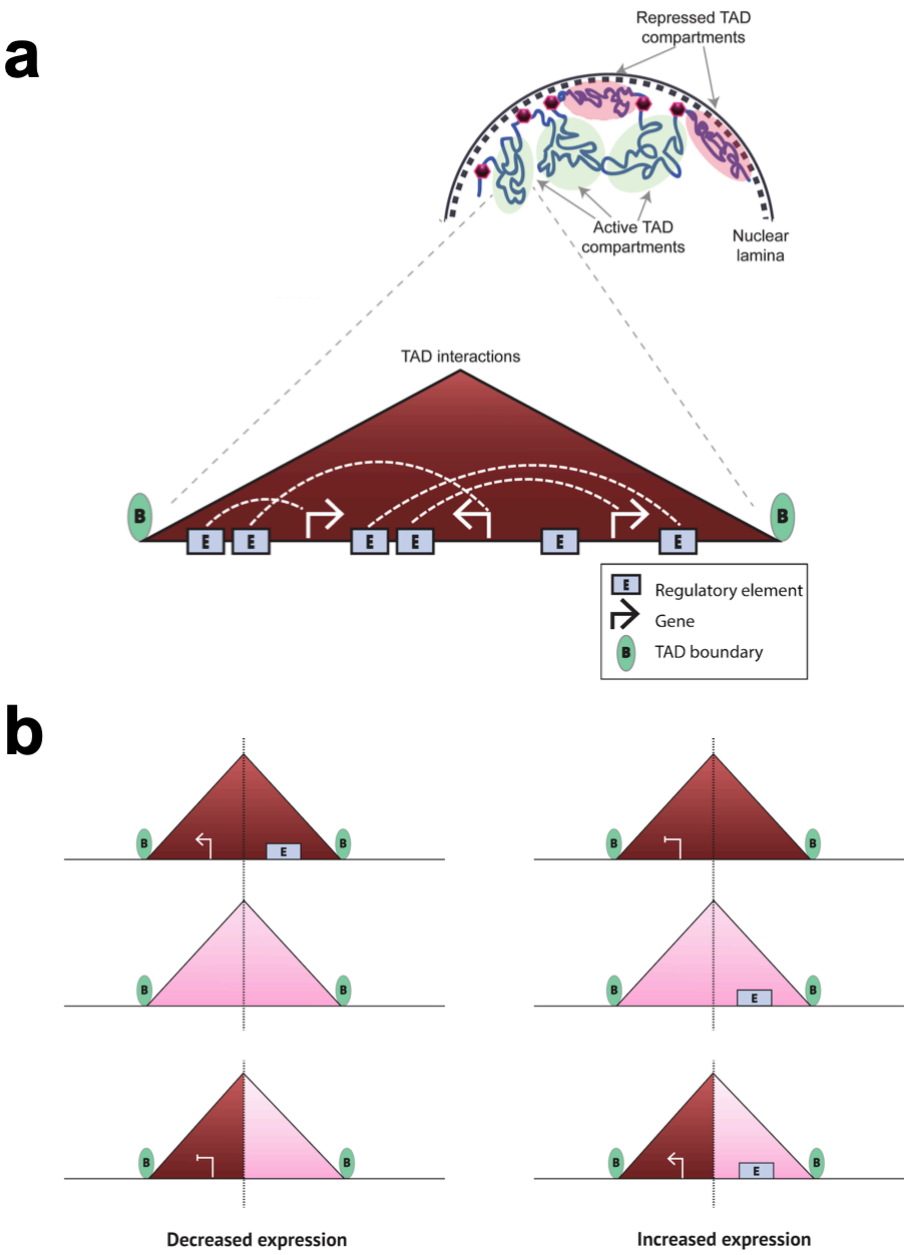


Figure 2.3. Topologically associating domains are biologically relevant delineations of regulatory regions

a) Topologically associating domains (TADs) are naturally occurring compartments of chromatin ranging in size from hundreds of kilobases to a few megabases[136]. These compartments (triangles) closely interact with themselves in three-dimensional space but not neighboring regions and are separated by boundaries (green ovals), which restrict regulatory elements from interacting with genes residing outside the boundary[152] (image adapted from [153]). **b)** Structural rearrangements can rewire gene expression by changing the position of regulatory elements relative to gene promoters[154]. For example, a translocation between two chromosomes (mahogany and pink) may either remove a regulatory region from a TAD on the derivative chromosome leading to decreased gene expression (left), or introduce a new regulatory element resulting in ectopic expression of the gene (right).

We computationally assessed predicted TADs by using stem cell Hi-C domains from the Hi-C project (chromosome.dsc.edu/) and converting them to hg19 for comparison to the breakpoints[136]. We identified six genes in the chromosome 20 (chr20) breakpoint-containing TAD and 14 genes in the chromosome 22 (chr22) breakpoint-containing TAD (Figure 2.4a). While one gene, *CDH4*, was directly disrupted by the breakpoint on chr20, it is not a strong candidate for severe oligospermia. *CDH4* encodes cadherin-4 or R-cadherin, which has predominant expression in the brain and plays a role in retinal axon outgrowth and visual system development[155-157]. The disruption is likely not pathogenic because it has a high haploinsufficiency score (%HI = 37)[158], suggesting that disruption of only one copy of *CDH4* is insufficient to induce pathogenicity. In addition, paternally-inherited *CDH4* deletions are reported in DECIPHER, which would be incompatible with a phenotype of male infertility[159].

To determine if DGAP230's translocation dysregulates candidate genes, we sought to assess their gene expression. Determining positional effects on genes near the breakpoints is not trivial, as we do not have reproductive tissues from DGAP230, such as testis, where the candidate genes are most likely to be expressed. However, Epstein-Barr virus-transformed LCLs are often useful in making important observations about the effects of structural rearrangements in DGAP subjects[160,161]. Although not all DGAP230's candidate genes are normally expressed in LCLs, we hypothesized that we would be able to detect genes that exhibit constitutive misexpression or overexpression compared to age- and sex-matched controls[157,158]. By assessing expression of every candidate gene by qPCR, we identified one gene on chr20, *SYCP2*, that is expressed about 20-fold in DGAP230 LCLs relative to age- and sex-matched controls, a result that was confirmed to be statistically significant by an unpaired two-tailed t-test ($p<0.02$). No other significant differences were noted in expression between DGAP230 and control LCLs (Figure 2.4b).

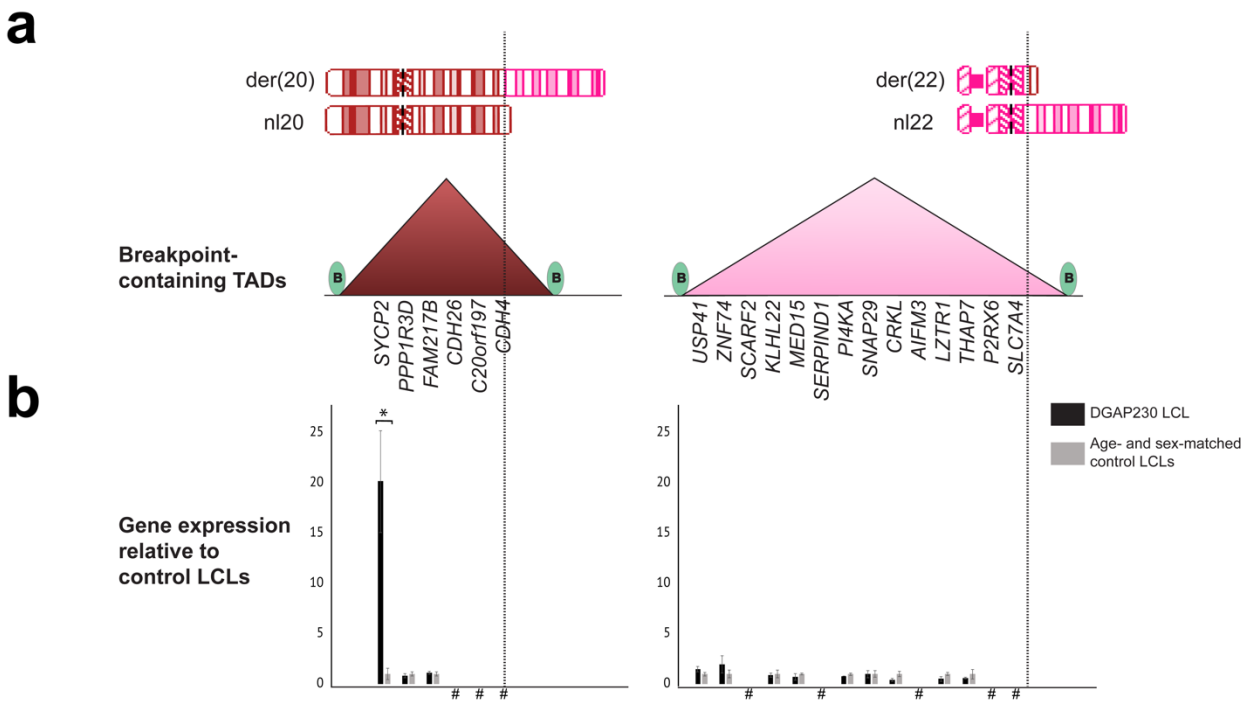


Figure 2.4. Topologically associating domains disrupted by DGAP230's t(20;22) define candidate genes

a) Vertical lines indicate positions of rearrangement breakpoints across ideograms of normal and derivative chromosomes 20 (mahogany) and 22 (pink). Genes residing in breakpoint-containing topologically associating domains (TADs) are listed below each TAD (triangle) defined by boundary regions (green ovals). **b)** qPCR of candidate genes in DGAP230 lymphoblastoid cells relative to age- and sex-matched control LCLs. Statistical significance was determined by an unpaired two-tailed t-test (N=3; $p<0.02$) and all results display mean ± standard error. # = genes deemed not expressed in LCLs by qPCR standard curve; * = $p<0.05$.

2.3.3. Determination of the cytogenetic etiology of *SYCP2* dysregulation

To determine if *SYCP2* misexpression is caused by DGAP230's t(20;22), we next characterized the mechanism by which *SYCP2* is dysregulated. Given that *SYCP2* is not expressed in karyotypically normal LCLs, we hypothesized that only the der(20)-residing allele would be expressed if the t(20;22) were responsible. To test the hypothesis that only one allele of *SYCP2* is expressed, we identified an exonic polymorphic region in *SYCP2* from DGAP230 genomic DNA and assessed expression of both allelic variants in DGAP230 LCL cDNA[162]. Sequence traces detected only one *SYCP2* haplotype in DGAP230 LCL cDNA, suggesting expression from a single allele in DGAP230 LCLs (Figure 2.5a).

To determine if the expressed *SYCP2* allele resides in *cis* with the der(20) translocation breakpoint, we sought to amplify selectively the der(20) allele using primers that span the translocation junction for subsequent Sanger sequencing. To overcome a technical challenge of the 1.5 Mb distance, we applied 3C-PCR (Chapter 3)[139], which capitalizes on principles underlying 3C to bring fragments containing the translocation junction and der(20)-residing *SYCP2* allele closer together, thus enabling PCR across the junction of a ligation product including the *cis* allele (Figure 2.5b)[144,163]. This resulted in production of amplicons in DGAP230 but not in karyotypically normal LCLs, demonstrating specificity for the t(20;22) (Figure 2.5b). Sanger sequencing of this amplicon revealed the expressed allele (Figure 2.5c), suggesting that *SYCP2* expression is detected exclusively from the der(20) allele in the DGAP230 LCL.

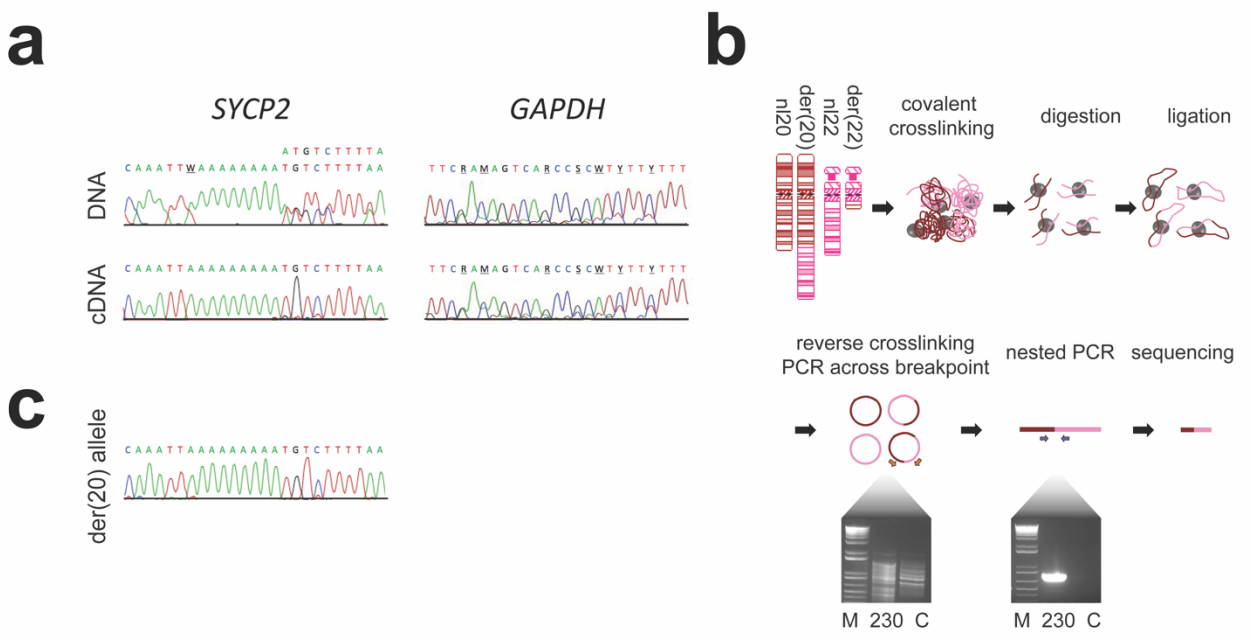


Figure 2.5. *SYCP2* is overexpressed exclusively from the der(20) allele
a) Comparison of DNA and RNA from polymorphic regions in *SYCP2* and *GAPDH* exons. Sequence traces are shown for genomic DNA (top) and cDNA (bottom) from DGAP230 lymphoblastoid cells. The *SYCP2* DNA trace shows heterozygosity at SNPs rs568645874 and rs199662252 while the *SYCP2* cDNA trace only detects one haplotype. In contrast, the *GAPDH* control shows heterogeneity in both DNA and cDNA (SNPs rs45568532, rs551180067 and rs11549332). W = A/T; R = A/G ; M = A/C ; S = C/G ; Y = C/T. **b)** Overview of 3C-PCR, a protocol that couples proximity ligation with breakpoint-spanning PCR to capture *cis* sequences distant from the chromosomal rearrangement[139]. Gel electrophoresis of PCR products from the first PCR across the breakpoint (left) and the second nested PCR using amplification products from the first PCR as a substrate (right) is shown. M = marker; C = control. **c)** Sanger sequencing traces for the der(20) allele, compared to the DNA and cDNA Sanger sequencing traces in Figure 2.5a (N=3).

With knowledge that the misexpression of *SYCP2* derives from the der(20), we hypothesized that this dysregulation is mediated by an "enhancer adoption" mechanism, a long-range *cis*-regulatory mutation that results in a gain of regulatory elements and subsequent promiscuous gene expression[154,164]. To explore this mechanism, we performed 4C-seq using the *SYCP2* promoter as bait in DGAP230 and age- and sex-matched control LCLs, and FourCSeq analysis for both nl20 and der(20) chromosomes[146]. A single statistically significant fragment was identified with 210-fold increased interaction in the DGAP230 LCL (N=3; p_{adj} control = 0.0053), which maps to a genomic region on chr22, 8 Mb downstream of the *SYCP2* promoter (Figure 2.6a). A closer look at this region revealed many signatures of enhancer activity including a high ratio of H3K4me1 to H3K4me3, H3K27ac, and two regions that demonstrate DNaseI hypersensitivity in fetal testis tissue (Figure 2.6b)[165,166]. Taken together, our findings support an enhancer adoption model, where an active enhancer residing in the segment of chr22 translocated to the der(20) may enter a newly formed chromatin contact encompassing the *cis*-residing *SYCP2* allele, resulting in illegitimate overexpression.

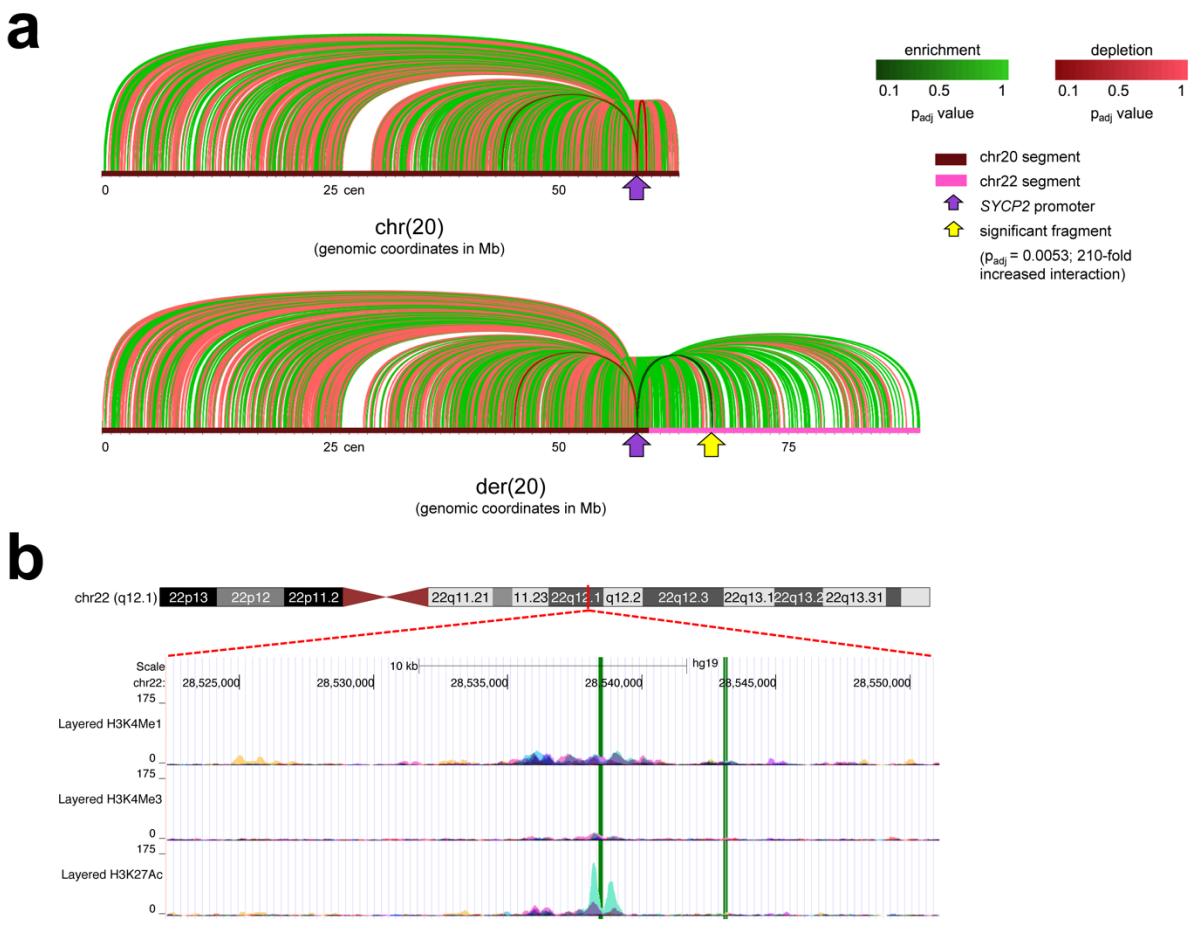


Figure 2.6. Investigation of chromatin contacts with the *SYCP2* promoter
4C-seq was performed on three biological replicates of the DGAP230 LCL and control LCLs from three different karyotypically normal age- and sex-matched controls. **a)** FourCSeq analysis for chr20 (top) and der(20) (bottom) reveals differential interactions of DNA fragments with the *SYCP2* promoter (purple arrow)[146]. Plots were made in Circos and linearized using Adobe Photoshop. Lines represent each DNA fragment found to interact with the *SYCP2* promoter. A single statistically significant fragment (yellow arrow) with DGAP230 enrichment appears in the der(20) analysis (210-fold enrichment, p_{adj} control = 0.0053). Cen = centromere. **b)** The significant DNA fragment and flanking enriched regions were consolidated to define a 29 kb putative interacting region. Tracks indicating enhancer activity from seven different cell lines are displayed in the UCSC genome browser and green bars highlight two regions of DNaseI hypersensitivity in fetal testis tissue according to ENCODE[165-167].

In order to implicate the genomic region identified by 4C-seq as the adopted enhancer for *SYCP2* dysregulation, we sought to determine causality by deleting the putative enhancer region using CRISPR/Cas9 and assessing the subsequent impact on *SYCP2* expression (Figure 2.7a)[148]. We established Cas9-expressing stable DGAP230 lymphoblastoid cells (Figure 2.7b) which were subsequently transduced with lentivirus derived from lentiviral vectors either containing dual targeting sgRNAs that flank the putative enhancer region or an empty vector control. Selection of dual targeting sgRNA-infected cells yielded a heterogeneous culture that demonstrates successful deletion of the target region (Figure 2.7c). Isogenic clones grown from single cell sorting revealed that only cells with biallelic intact enhancer regions survive (Figure 2.7d). To test whether the putative enhancer region is important for cell viability, dual targeting sgRNA-infected heterogeneous cultures were grown for variable lengths of time and quantitative RT-PCR of extracted genomic DNA was assessed for relative amounts of the targeted deletion. While the targeted deletion is enriched in dual targeting sgRNA-infected heterogeneous cells cultured for one month relative to the background signal from empty vector control-infected cultures, cells with the targeted deletion are selected out of the dual targeting sgRNA-infected heterogeneous population by four months (Figure 2.7e). These findings suggest that the putative enhancer region targeted by CRISPR/Cas9 is important for cell growth or viability. Due to the hypothesized biological significance of this region, we were unable to generate DGAP230 LCLs with a deleted putative enhancer region in order to study further its role in *SYCP2* overexpression.

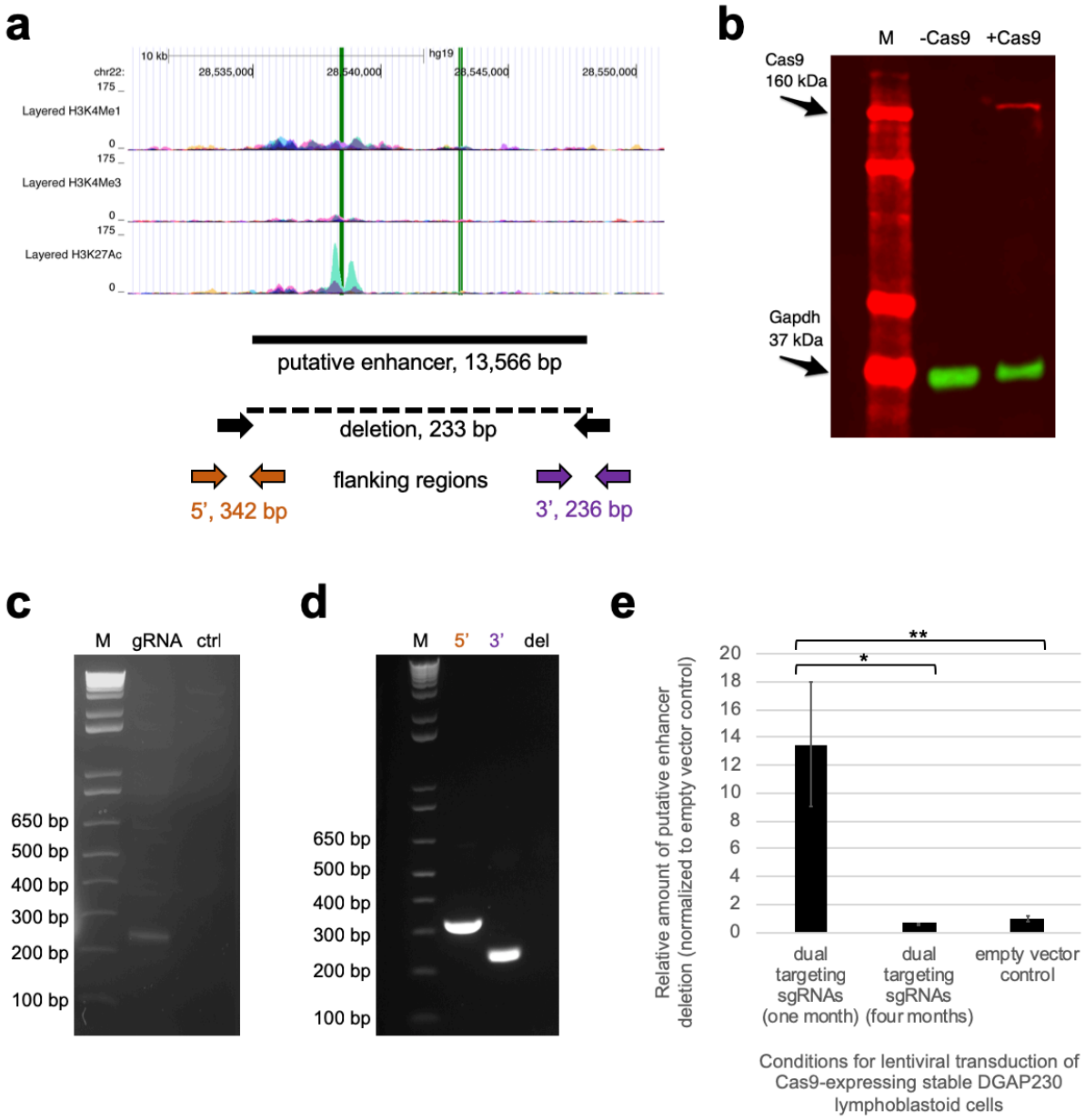


Figure 2.7. CRISPR/Cas9-mediated deletion of putative adopted enhancers
a) The targeted deletion (black solid line) is mapped to the significant DNA fragment identified by 4C-seq. Tracks indicating enhancer activity from seven different cell lines are displayed in the UCSC genome browser and green bars highlight two regions of DNaseI hypersensitivity in fetal testis tissue according to ENCODE[165-167]. The dotted line and black arrows below the targeted deletion represent the predicted amplicon size by PCR of a deleted substrate. Orange and purple arrows represent the predicted PCR amplicon size of the 5' and 3' flanking regions, respectively. **b)** A Cas9-expressing stable LCL was established for DGAP230, as indicated by Cas9 protein expression in a Western blot. M = marker; -Cas9 = uninfected DGAP230 LCL; +Cas9 = Cas9-expressing stable DGAP230 LCL.

Figure 2.7. (Continued)
c) PCR amplification across the deletion in genomic DNA from the dual targeting sgRNA-infected culture reveals some CRISPR/Cas9-deleted target regions in the heterogeneous lymphoblastoid cell population. M = 1 kb plus DNA ladder; gRNA = deletion assessment in the heterogeneous culture infected with the dual targeting sgRNA lentivirus; ctrl = deletion assessment in the culture infected by the empty vector control lentivirus. **d)** Example of PCR analysis for an isogenic LCL grown from a single cell after dual targeting sgRNA-infection of Cas9-expressing stable DGAP230 lymphoblastoid cells. The presence of amplicons for the 5' and 3' flanking regions and absence of an amplicon across the deletion are interpreted as no deletion of the targeted enhancer region. M = 1 kb plus DNA ladder; 5' = 5' flanking region; 3' = 3' flanking region; del = deletion targeted by the dual targeting sgRNA lentivirus. **e)** Quantitative RT-PCR of the putative enhancer deletion in DGAP230 LCLs transduced with the dual targeting sgRNA lentivirus and grown for one month compared to four months show a loss of modified cells over time. Two biological replicates for the one-month and control conditions and 15 biological replicates for the four-month condition were each tested in triplicate and grouped for graphing and statistical analyses. Results are normalized to the empty vector control value and graphed as mean ± standard error. Results were found to be statistically significant between the one-month and four-month groups by a heteroscedastic one-tailed t-test ($p<0.02$) and between the one-month and control conditions by a homoscedastic one-tailed t-test ($p<0.01$). * = $p<0.05$; ** = $p<0.01$.

2.4. Contributions

DGAP230 was enrolled through DGAP with principal investigators Drs. Cynthia Morton, Bradley Quade, James Gusella, and Richard Maas. Clinical information was obtained through the Coordinating and Administrative Core (CAC) from Dr. Cynthia Morton's lab, including assistance from coordinators Shahrin Ahsan, Tammy Kammin and Ellen Wilch. Large-insert jumping library sequencing analysis was performed in Dr. James Gusella's lab by Dr. Michael Talkowski and Carrie Hanscom. The GTG-banded karyotype was prepared by Dr. Shumei Wang and interpreted by Dr. Cynthia Morton. Samantha Schilit delineated TADs disrupted by DGAP230's breakpoints, performed RNA extraction, cDNA synthesis, and qRT-PCR of lymphoblastoid cells, identified and evaluated variable regions in *SYCP2* and *GAPDH*, and performed 3C-PCR to phase *SYCP2* alleles. Samantha Schilit also optimized the protocol for and made all 4C-seq libraries, which were then sequenced by the Biopolymers Facility in the Department of Genetics at Harvard Medical School and analyzed by Shreya Menon. The CRISPR/Cas9-mediated genome editing in lymphoblastoid cells protocol was developed by Sizun Jiang, who also provided guidance for designing guide RNAs and troubleshooting as well as materials including 293FT cells and the H1 bridge dsDNA block. Samantha Schilit performed the genome editing protocol with assistance from Shreya Menon, including establishment and validation of Cas9-expressing stable LCLs, cloning dual targeting sgRNAs into lentiviral vectors, transducing dual targeting sgRNA lentivirus into LCLs, and validating genomic deletions of single cell clones.

CHAPTER 3

3C-PCR IS A NOVEL PROXIMITY LIGATION-BASED APPROACH TO PHASE CHROMOSOMAL REARRANGEMENT BREAKPOINTS WITH DISTAL ALLELIC VARIANTS

3.1. Preface

3C-PCR is an inexpensive and efficient proximity ligation-based method for phasing chromosomal rearrangement breakpoints with distal allelic variants, which was developed to phase the chr20 and der(20) *SYCP2* alleles in DGAP230 lymphoblastoid cells (Chapter 2). In this chapter, we describe details of the technique including its potential application for clinical diagnostic laboratories. This work has been published in *Human Genetics*[139].

3.2. Background

In the past two decades, efforts to annotate the human genome have revealed a significant functional role for noncoding sequences. Genomic structural variation, such as copy-number variants and genomic rearrangements, have been shown to lead to genomic disorders[168]. Many of these variants result in an abnormal phenotype by altering long-range control of gene expression[169]. This is mediated by the disruption of TADs and subsequent promiscuous enhancer-promoter interactions that lead to pathogenic misexpression[128,150,164]. Given the clinical significance of long-range *cis* regulatory mutations, recent research has focused on predicting clinical outcomes for subjects with structural chromosomal rearrangements by considering dysregulation of genes that reside in the disrupted TADs[151,170].

If a dysregulated gene is associated with an autosomal recessive disease phenotype and subsequent sequencing of the gene reveals a second pathogenic variant, phasing is critical for clinical interpretation. While variants in *cis* may not manifest in the disease phenotype, variants that reside in *trans* result in a compound heterozygote[171]. The vast difference in clinical interpretation highlights a critical need for a method capable of deciphering large haplotypes across derivative chromosomes. There is great interest in applying this technology to *de novo* BCAs, because long-range position effects explain clinical phenotypes in a substantial proportion of subjects with BCAs[128].

While computational and experimental phasing has been used to identify haplotypes since the 1980s, current methods are insufficient to resolve a haplotype that spans megabase distances on derivative chromosomes, as requisite for a TAD-disrupting chromosomal rearrangement[172]. Computational haplotype phasing, which relies on genotype data from unrelated individuals using statistical approaches or from families using identity by descent (IBD), cannot be applied to nonrecurring genomic rearrangements because they are not common in the population or may not be inherited[172]. While experimental techniques such as long-range PCR, Drop-Phase, and targeted locus amplification (TLA) do not require population or family genotyping data, they are

limited by genomic distance, losing efficacy beyond 30, 200, and 400 kb, respectively[173-175]. Other technologies that physically separate chromosomes before genotyping, such as by microdissection using a computer-directed laser beam or by dispersion using a microfluidic device, may span large enough distances[176,177]; however, these techniques require specialized equipment and are labor intensive making them difficult to apply broadly. Even experimental techniques with straightforward protocols that can easily be translated to other labs, like HaploSeq, are still limiting in that they are expensive and require substantial computational expertise due to the cost and subsequent analysis of NGS[163].

In this chapter, we describe 3C-PCR, an inexpensive and efficient proximity ligation-based approach to phase chromosomal rearrangement breakpoints with distal allelic variants. Our method adapts the use of canonical 3C libraries by employing a novel nested PCR strategy with primers anchored across the rearrangement breakpoints and subsequent Sanger sequencing[144]. 3C has become a widely used method that can be performed in a matter of days using standard molecular biology equipment and the routine methods of PCR and Sanger sequencing in diagnostic laboratories[140]. By combining these simple and accessible methods, 3C-PCR makes possible phasing variants at a distance of over a megabase from a chromosomal rearrangement without the expense of specialized equipment, NGS or extensive computational analysis.

3.3. Methods

3.3.1. Acquisition of LCLs

Participants DGAP230, with 46,XY,t(20;22)(q13.3;q11.2), and DGAP278-02, a karyotypically normal age- and sex-matched control, were enrolled through DGAP. DGAP obtained informed consent, medical records and blood samples under a protocol approved by the Partners HealthCare Systems Institutional Review Board. Epstein-Barr virus transformed LCLs were generated at the Genomics and Technology Core in the Center for Human Genetic Research at Massachusetts General Hospital (Boston, MA, USA). Jumping library GS and subsequent Sanger sequencing identified the precise breakpoints of the DGAP230 chromosomal rearrangement as previously described and reported[128,133,134]. Two additional karyotypically normal age- and sex-matched control LCLs, GM20184 and GM20188, were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research (Camden, NJ, USA).

3.3.2. Identification of a variable region on chr20

TADs disrupted by the breakpoints in DGAP230 were identified according to human embryonic stem cell Hi-C domains from the Hi-C project[136]. The University of California Santa Cruz Genome Browser was used to delineate regions located over a megabase away from the t(20;22) breakpoints within the same TAD[137]. These sequences were compared against dbSNP to identify highly variable regions in the distal TAD-residing sequences[138].

To assess heterozygosity of these candidate regions in DGAP230 and control LCLs, genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen). PCR was performed using LongAmp Taq 2X Master Mix (NEB) and customized primers (Integrated DNA Technologies) designed to amplify potential variable regions. After amplification confirmation with agarose gel electrophoresis, Sanger sequencing reactions of PCR products were carried out with an ABI3730xl DNA analyzer. Chromatograms were aligned and multiple single nucleotide variants

were called using Geneious (Version 7.0, Biomatters). A target region was selected based upon the presence of several single nucleotide variants in the chromatograms for all experimental and control samples.

3.3.3. Generation of 3C libraries

3C libraries were generated as previously described[140-144]. In brief, 10 million cell aliquots of LCLs were crosslinked with 2% formaldehyde (Sigma-Aldrich) and lysed. Chromatin was digested with *Hin*dIII-HF (NEB), ligated with T4 DNA ligase (NEB), and reverse crosslinked by incubation with Proteinase K (NEB) and RNase A (EMD Millipore). DNA libraries were purified by phenol/chloroform/IAA extraction (Sigma-Aldrich), MaXtract High Density Tubes (Qiagen), and subsequent ammonium acetate precipitation (Sigma-Aldrich). 3C libraries were generated in triplicate, with three independent cultures for the DGAP230 LCL and three different control LCLs.

3.3.4. Design of primers for nested PCR approach

Primer design was adapted from 3C protocols, but with adjustments to accommodate target regions further than 80-150 bp from the restriction enzyme digestion site and PCR amplicons longer than 160-300 bp, as previously described[140]. Sequences were obtained for two predicted *Hin*dIII-digested fragments: one of the target region on chr20, and a second containing the sequence on chr22 most proximal to the der(20) breakpoint. A synthetic sequence of a potential ligation product from these two fragments was designed in SeqBuilder (Version 14.1.0.118, DNASTAR) by concatenating the two sequences at their respective *Hin*dIII restriction sites. Primers spanning both fragments and the target variable region were designed in Primer3Plus and assessed for sequence specificity using BLAST-like alignment tool (BLAT)[178,179]. Nested primer pairs were designed such that one primer pair flanked the entire substrate recognized by the second primer pair (Table 3.1).

Table 3.1. Primers used in chapter 3

Name	Sequence (5'-3')	Description
SS_232F	GAGGGCACACCTGGTTTTTA	For amplification of genomic target region with SS_233R; amplicon = 3044 bp
SS_233R	AATCCCCAAGAAGAGGAGGA	For amplification of genomic target region with SS_232F; amplicon = 3044 bp
SS_254F	AAGGCTGTCACCACAGGAAG	For first amplification with SS_258R; amplicon = 2843 bp
SS_258R	GTGAGACCCCATCCCAATTA	For first amplification with SS_254F; amplicon = 2843 bp
SS_256F	GACAGACTCAGCGGTTCAGA	For nested PCR with SS_257R; amplicon = 2655 bp
SS_257R	ATATGGCGAAACCCAGTGTC	For nested PCR with SS_256F; amplicon = 2655 bp
SS_230F	CTCTTTGCTAGGGTGCCAAG	For Sanger sequencing

3.3.5. Rearrangement-specific amplification and sequencing

Nested PCRs of breakpoint-spanning fragments were performed using LongAmp Taq 2X Master Mix (NEB). The first PCR reaction amplified ~300 ng 3C libraries for all experimental and control samples using the outer primer pair (SS_254F and SS_258R listed in Table 3.1) and generous thermocycling conditions including a long extension time and low annealing temperature (3 minutes at 94ºC, 35 cycles X [30 seconds at 94ºC, 30 seconds at 56ºC, 2.5 minutes at 65ºC], 10 minutes at 65ºC, hold 4ºC). Amplicons were purified using a QIAquick PCR purification kit (Qiagen). After quantification, ~100 ng purified amplicons were used as substrates for a second PCR reaction using the inner primer pair (SS_256F and SS_257R listed in Table 3.1) and more stringent conditions with a shorter extension time and higher annealing temperature (3 minutes at 94ºC, 45 cycles X [30 seconds at 94ºC, 2 minutes at 65ºC], 10 minutes at 65ºC, hold 4ºC). Nested PCR amplicon specificity was evaluated using agarose gel electrophoresis. Amplicons were purified using a QIAquick PCR purification kit (Qiagen) and Sanger sequenced with an ABI3730xl DNA analyzer using primer SS_230F (Table 3.1). 3C-PCR chromatograms were aligned to genomic DNA chromatograms for comparison and nucleotide variants were called using Geneious (Version 7.0, Biomatters).

3.4. Results

To develop an assay capable of phasing allelic variants over a megabase away from a breakpoint of a chromosomal rearrangement within the same TAD, we searched for an LCL that has a BCA with at least one breakpoint located over a megabase away from a TAD boundary. Through DGAP, we selected the DGAP230 LCL, with 46,XY,t(20;22)(q13.3;q11.2) and a distance of more than 1.4 Mb between the chr20 breakpoint and the upstream boundary of the TAD in which it resides (Figure 3.1a)[128]. To ensure assay specificity, we also selected three karyotypically normal age- and sex-matched control LCLs: DGAP278-02, GM20184 and GM20188. As a source for allelic variation, we identified a highly variable region 1.3 Mb upstream of the chr20 breakpoint. Sanger sequencing of this target region showed heterozygosity at several bases in DGAP230 as well as all control cell lines (Figure 3.1b).

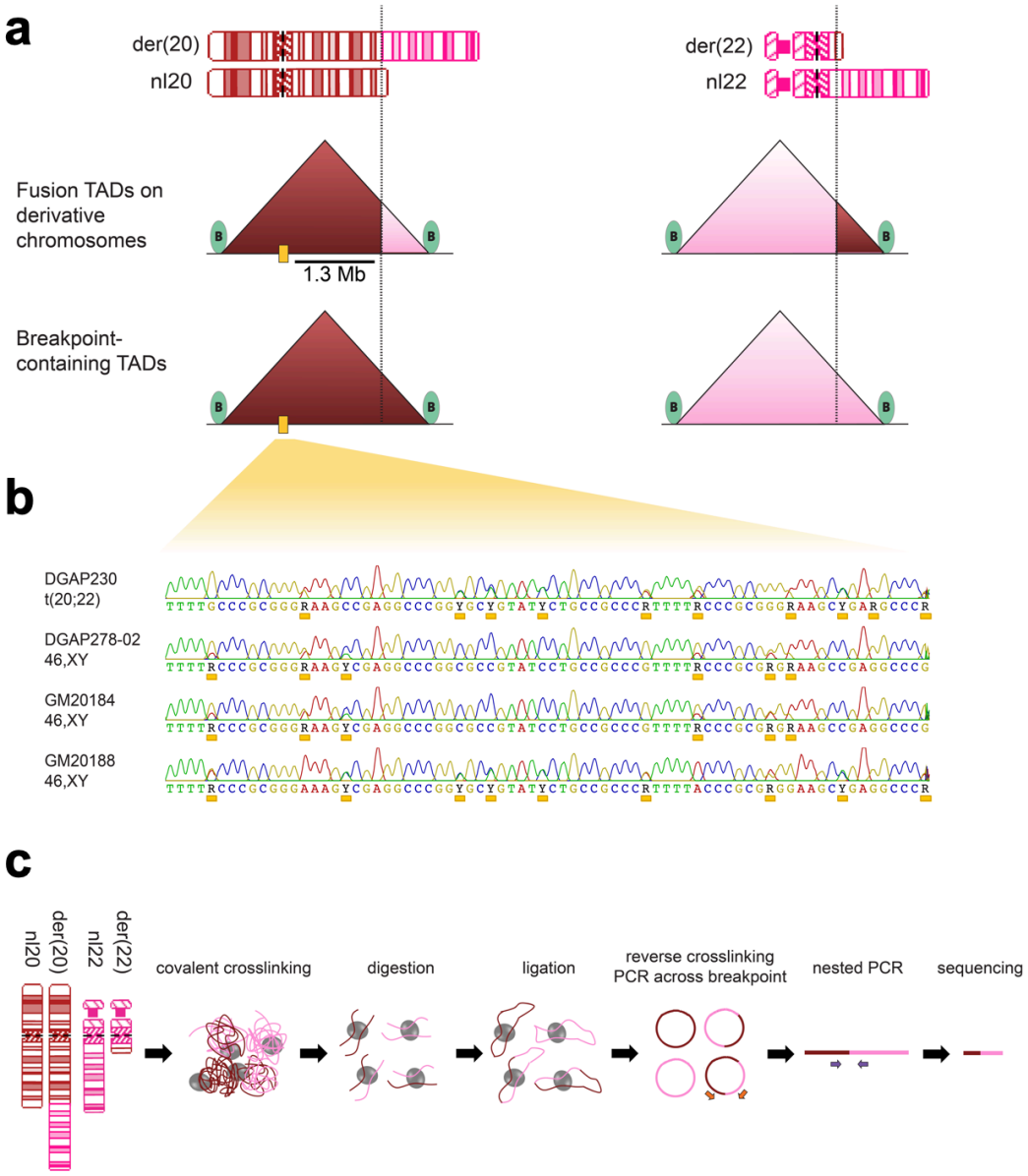


Figure 3.1. Experimental system for 3C-PCR

a) The LCL, designated DGAP230, has a balanced translocation between the long (q) arms of chromosomes 20 (mahogany color) and 22 (light pink color; top). Translocation breakpoints reside near the boundaries (green ovals) of predicted TADs (triangular shapes), enabling assessment of a distal region with multiple single nucleotide variants (yellow box) within the same chromatin loop (bottom). **b)** Chromatograms from Sanger sequencing of the target region reveal a highly variable region in DGAP230 and control cell lines. Single nucleotide variants are indicated by a small orange box below the corresponding nucleotide. R= A/G; Y= C/T. **c)** In 3C-PCR, coupling proximity ligation with breakpoint-spanning PCR can capture *cis* sequences distant from the chromosomal rearrangement. Chromatin conformation capture libraries are generated by covalent crosslinking of chromatin, enzymatic digestion, and ligation of proximal genomic fragments to bring high frequency three-dimensional interactions into two-dimensional linear space. Reverse-crosslinked ligation products are then subjected to two rounds of nested PCR to select for specific amplicons that cross the breakpoint junction and include the *cis* target region for subsequent Sanger sequencing.

We next set out to develop a method capable of determining the haplotype of the target variable region on the der(20). If the target region and chr20 breakpoint were located only a few kilobases apart, phasing could be accomplished by selectively amplifying the der(20) allele using primers that span the translocation junction to produce an amplicon containing the target region in *cis*, which could be resolved by Sanger sequencing. However, the 1.3 Mb distance between the breakpoint and the target region render this strategy unsuccessful because PCR performs at distances three orders of magnitude smaller.

To overcome this technical challenge, we developed a strategy called 3C-PCR. This method capitalizes on principles underlying 3C technologies developed by Dekker and Kleckner in 2002, which show that when crosslinked DNA is enzymatically digested into genomic fragments and then ligated to other fragments in close physical proximity, sequences in *cis* have a higher interaction frequency than those in *trans*[144,180]. We hypothesized that we could use 3C to bring fragments containing the translocation junction and der(20) target region closer together, thus enabling PCR across the junction of a ligation product including the *cis* target region. Given the strong possibility of amplifying nonspecific sequences from a complex 3C library with diverse ligation products, we pursued a nested PCR step to improve specificity (Figure 3.1c)[181].

Using the predicted ligation product as a substrate, we designed nested primers that would span the target region on chr20, the enzymatic digestion and ligation site, and the chr22 genomic fragment near the breakpoint (Figure 3.2a,b). As expected, the first amplification resulted in several nonspecific PCR products for all DGAP230 and control LCL 3C libraries (Figure 3.2c). However, after performing nested PCR on products purified from the first amplification, we produced DNA fragments of predicted size from all the DGAP230 samples but from none of the controls, suggesting that nested PCR recognized the predicted proximity ligation product from the *cis*-interacting der(20) chromosome only present in the DGAP230 samples (Figure 3.2c).

As evidence that the predicted proximity ligation product is the substrate for amplification, nested PCR on negative control genomic libraries without crosslinking, digestion or ligation

yielded no PCR-amplified products (Figure 3.3a). Additionally, *Hin*dIII digestion and subsequent agarose gel electrophoresis of the amplicon from the DGAP230 3C library-nested PCR confirmed derivation from the predicted ligation product (Figure 3.3b,c). Sequencing of all three amplicons revealed no heterogeneity and identical sequences, providing evidence that this is the haplotype of the target region on the der(20) (Figure 3.2d).

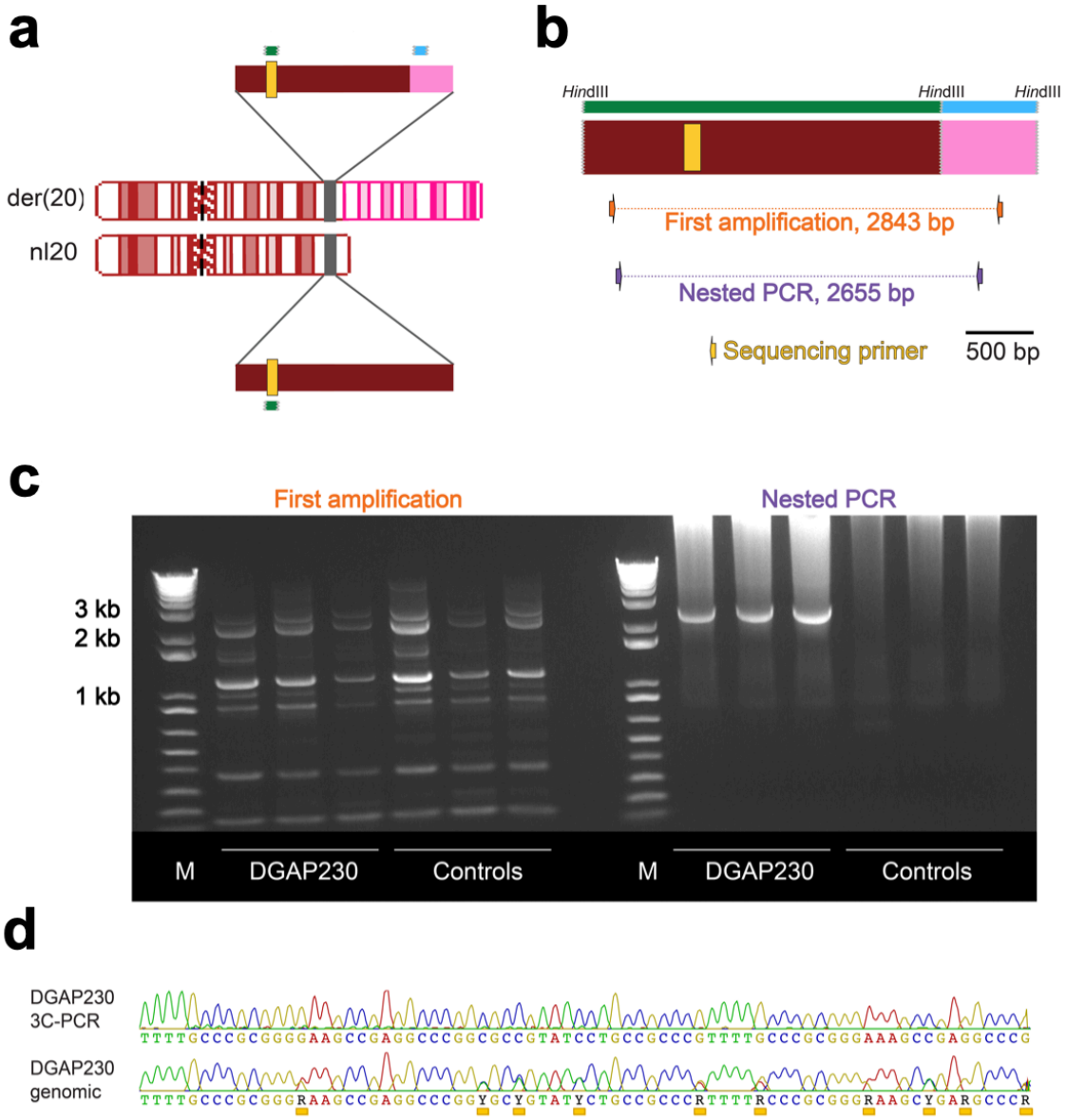


Figure 3.2. 3C-PCR assay validation
a) The goal of the assay in the DGAP230 experimental system is to differentiate the target region (yellow box) on the der(20) chromosome (top) from the target region on the normal chr20 (bottom). The small green bar represents the 3C genomic fragment that contains the target region and the small blue bar represents the digested genomic fragment containing a breakpoint-proximal region from the segment of chr22 translocated to the der(20). Rough gray edges reflect enzymatic digestion at flanking *Hin*dIII restriction sites. **b)** Schematic of nested PCR amplifications for the predicted ligation product with the target region (green bar above mahogany map) and the chr22 fragment (blue bar above light pink map). **c)** Gel electrophoresis displays products from the first PCR across the breakpoint for experimental and control 3C libraries (left), and the second nested PCR (right; N=3)**.** Key DNA fragment sizes of the markers (M) are indicated on the left. **d)** Sanger sequencing traces of the target variable region from the nested PCR amplicon (top) and genomic DNA from the same cell line (bottom; N=3).

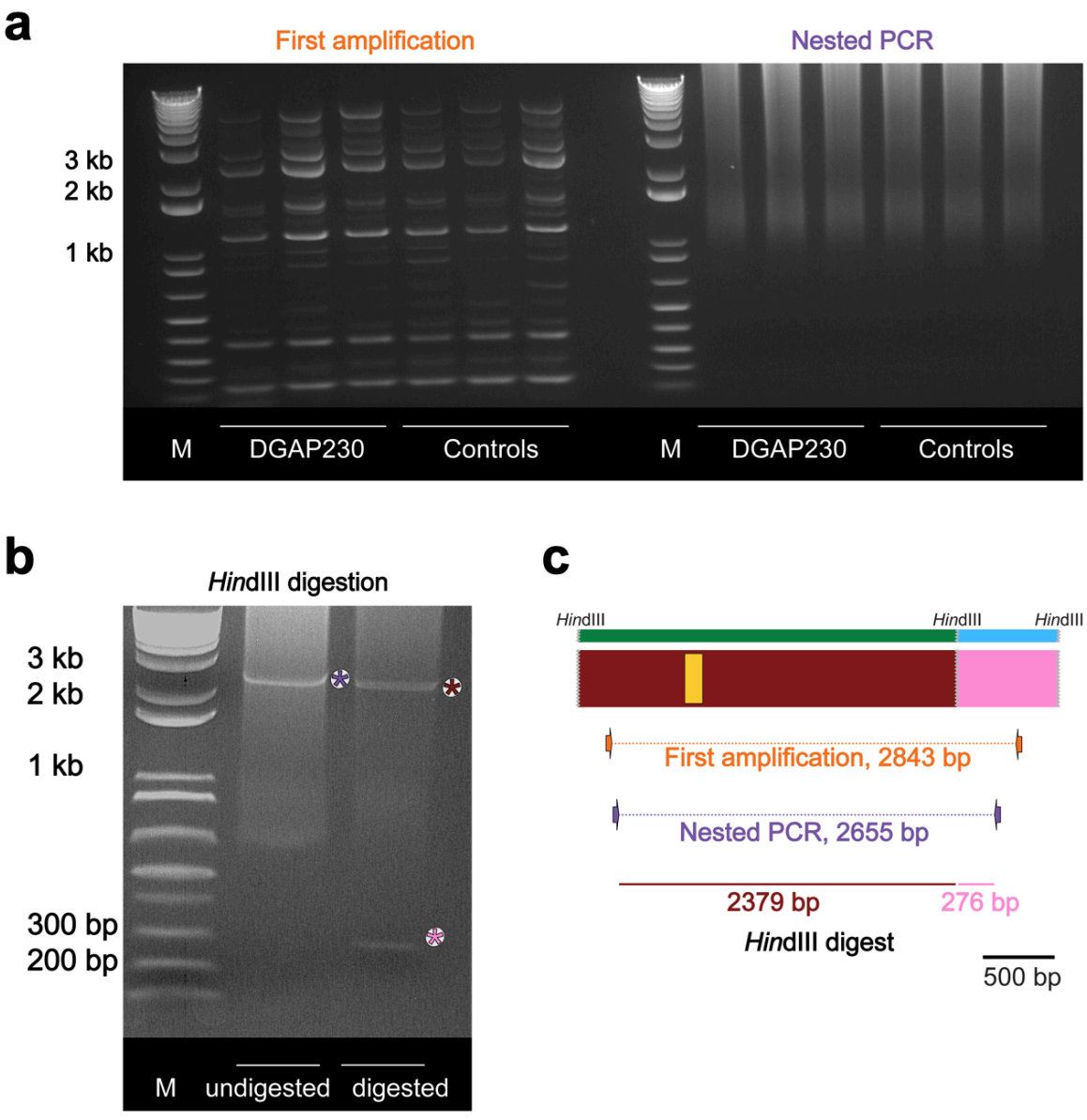


Figure 3.3. 3C-PCR control experiments
a) Gel electrophoresis displays products from the first PCR across the breakpoint for negative control genomic libraries without crosslinking, digestion or ligation (left), and the corresponding second nested PCR (right; N=3)**.** Key DNA fragment sizes of the markers (M) are indicated on the left. **b)** Gel electrophoresis of the amplicon from the DGAP230 3C nested PCR before (left) and after (right) *Hind*III digestion. Asterisks are colored to identify each band. Purple = entire amplicon after nested PCR; mahogany = fragment of ligation product derived from chr20; light pink = fragment of ligation product derived from chr22. Key DNA fragment sizes of the markers (M) are indicated to the left of the gel. **c)** Schematic of nested PCR amplifications for the predicted ligation product and predicted amplicon fragments after *Hind*III digestion. The thin green bar above the mahogany bar represents the 3C genomic fragment that contains the target region and the thin blue bar above the light pink bar represents the digested genomic fragment containing a breakpoint-proximal region from the segment of chr22 translocated to the der(20). Rough gray edges mark *Hind*III restriction sites.

3.5. Analysis

3C-PCR is an inexpensive and efficient proximity ligation-based approach to phase chromosomal rearrangement breakpoints with distal allelic variants. We anticipate that the simplicity of this approach will expedite its adoption in future clinical practice to determine compound heterozygosity in cases where a gene dysregulated by a disrupted TAD harbors a second pathogenic variant.

3C-PCR serves as a novel application to the widely used 3C method and differentiates itself from other adaptions of 3C in its ease, technical capabilities and versatility[144]. 3C-PCR targets the allele of a variable locus in *cis* with a chromosomal rearrangement on a derivative chromosome by a simple nested PCR strategy on 3C libraries, eliminating the need for costly and time consuming NGS and computational analysis used in other proximity ligation-based phasing methods[163,173]. In addition, these other phasing methods are also technically inferior to 3C-PCR in that HaploSeq has a sparse ascertainment density resulting in less than a 25% chance of detecting the distal allelic variant of interest as opposed to 100% for 3C-PCR, and TLA can only haplotype distances of up to 300 kb, less than a third of the capabilities of 3C-PCR[182].

In our system, nonspecific amplification of 3C libraries is ameliorated by a two-step nested PCR. This differs from standard PCR of 3C libraries to determine semi-quantitative interaction frequencies, because primers can be designed to flank closely the restriction enzyme digestion sites of the two genomic fragments in question, allowing for short PCR extension times that select for a small 160-300 bp amplicon[140]. In our assay, resulting amplicons must include the target region residing anywhere in the enzymatically digested genomic fragments (e.g., 2 kb when considering that restriction endonucleases with six base pair recognition sequences produce genomic fragments about 4 kb in size). Our optimized nested PCR strategy compensates for the nonspecific amplicons produced from longer extension times. The first PCR amplifies all possible products, with generous conditions including a long extension time and low annealing temperature. To prevent a biased overamplification of certain products, the number of cycles

allows for amplification within the linear range. The subsequent nested PCR applies stricter conditions with a shorter extension time and a much higher annealing temperature to select for the specific amplicon of interest. More cycles are used to compensate for the less efficient PCR.

Of note, this technique relies on the assumption that sequences in *cis* will have higher interaction frequencies than those in *trans*. While ligation products containing the *trans* target region and the breakpoint-proximal fragment would be much less common, they may still be present. To alleviate these concerns, PCR products detected in the DGAP230 cell line with the t(20;22) substrate are expected more frequently than in karyotypically normal cells. Indeed, our results identified an amplicon of the predicted size from the nested PCR in three independent 3C libraries performed on the experimental cell line and no products in three different 3C libraries derived from karyotypically normal LCLs (Figure 3.2c). Sanger sequencing of the same haplotype in all three replicates provides evidence of detection of the higher frequency *cis* interaction event (Figure 3.2d).

Our novel method does have some limitations. 3C-PCR targets a specific region, so customized primers must be designed and synthesized to probe the region of interest. The breakpoint of interest must also be resolved to near-nucleotide resolution (on the order of a couple kilobases), as is done by mate-pair or large-insert jumping libraries, to identify a genomic region known to reside on the derivative chromosome close to the breakpoint. If breakpoint information is only available at the resolution level of a karyotype, we predict that 3C-PCR will be successful if (1) there is a genomic region known with certainty to reside in *cis* with the breakpoint, and (2) if this region is less than 30 Mb away from the allelic variant, as a higher interaction frequency for *cis* sequences compared to *trans* sequences persists for genomic distances of up to 30 Mb in proximity ligation assays (only ~0.6% for *trans* interactions but increasingly to 2% at larger distances)[163]. This strong bias for *cis* interactions also provides versatility in 3C-PCR, as indels, that may alter genomic distances on the order of 1-10,000 bp between the breakpoint and the allelic variant, would not significantly influence interaction frequencies[183]. Similarly, due to this

long-spanning *cis*-interaction bias relative to the 880 kb median size of TADs, the variant of interest does not need to reside in the same TAD as the rearrangement breakpoint[136].

Due to dependence of this technology on discriminating *cis* versus *trans* by proximity ligation, 3C-PCR will inherently work better for balanced translocations than for balanced inversions, in which both sides of the breakpoint derive from the same chromosome. Efficacy will depend on the difference in interaction frequency of the breakpoint-proximal genomic region and the variant of interest on the inverted and normal chromosomes, which will be affected by many factors including linear distance and the presence of TADs, enhancer-promoter interactions and insulator elements[180].

Due to the requirement to make proximity ligation libraries, another limitation is that 3C-PCR requires intact chromatin from tissue or cultured cells. Finally, the assay is also dependent upon successful PCR, which may be impacted by the specific ligation product's GC or AT content, predicted secondary structure, or length. However, these limitations are less prohibitive than other technologies capable of phasing at distances over a megabase, including targeted haplotyping by dilution, single-chromosome sequencing, and HaploSeq, all of which require costly NGS and intensive labor[163,176,182,184]. 3C-PCR can phase distal variants with low cost and limited labor, using standard molecular biology reagents and equipment. As clinical diagnostic laboratories enter the era of "next-gen cytogenetics," determining allelic nucleotide variant(s) of the sequence of a gene dysregulated by a structural chromosomal rearrangement will become essential. In these cases, 3C-PCR will be integral to clinical interpretation and prediction of disease phenotypes.

3.6. Contributions

All experiments were designed, performed, and analyzed by Samantha Schilit. The manuscript was written by Samantha Schilit with guidance from Dr. Cynthia Morton, who served as the corresponding author on the paper.

CHAPTER 4

SYCP2 ABERRATIONS CAUSE MALE INFERTILITY BY INTERFERING WITH MEIOSIS

4.1. Background

In this chapter, we investigate how *SYCP2* dysregulation can explain DGAP230's phenotype of severe oligospermia. *SYCP2* encodes synaptonemal complex protein 2, an axial element in the proteinaceous synaptonemal complex[185]. SYCP2 nonspecifically interacts with the minor groove of DNA and serves as a scaffold for recruiting SYCP3 through its coiled-coil domain, thus facilitating formation of the synaptonemal complex[186-188]. Synaptonemal complex assembly is a meiosis-specific process that plays a role in pairing, recombination and segregation of homologous chromosomes during meiosis I (Figure 4.1)[188,189]. SYCP2 is important for spermatogenesis, as mice homozygous for coiled-coil domain-deficient *Sycp2* show male-specific infertility by compromising homologous chromosome synapsis, leading to spermatocyte apoptosis[186].

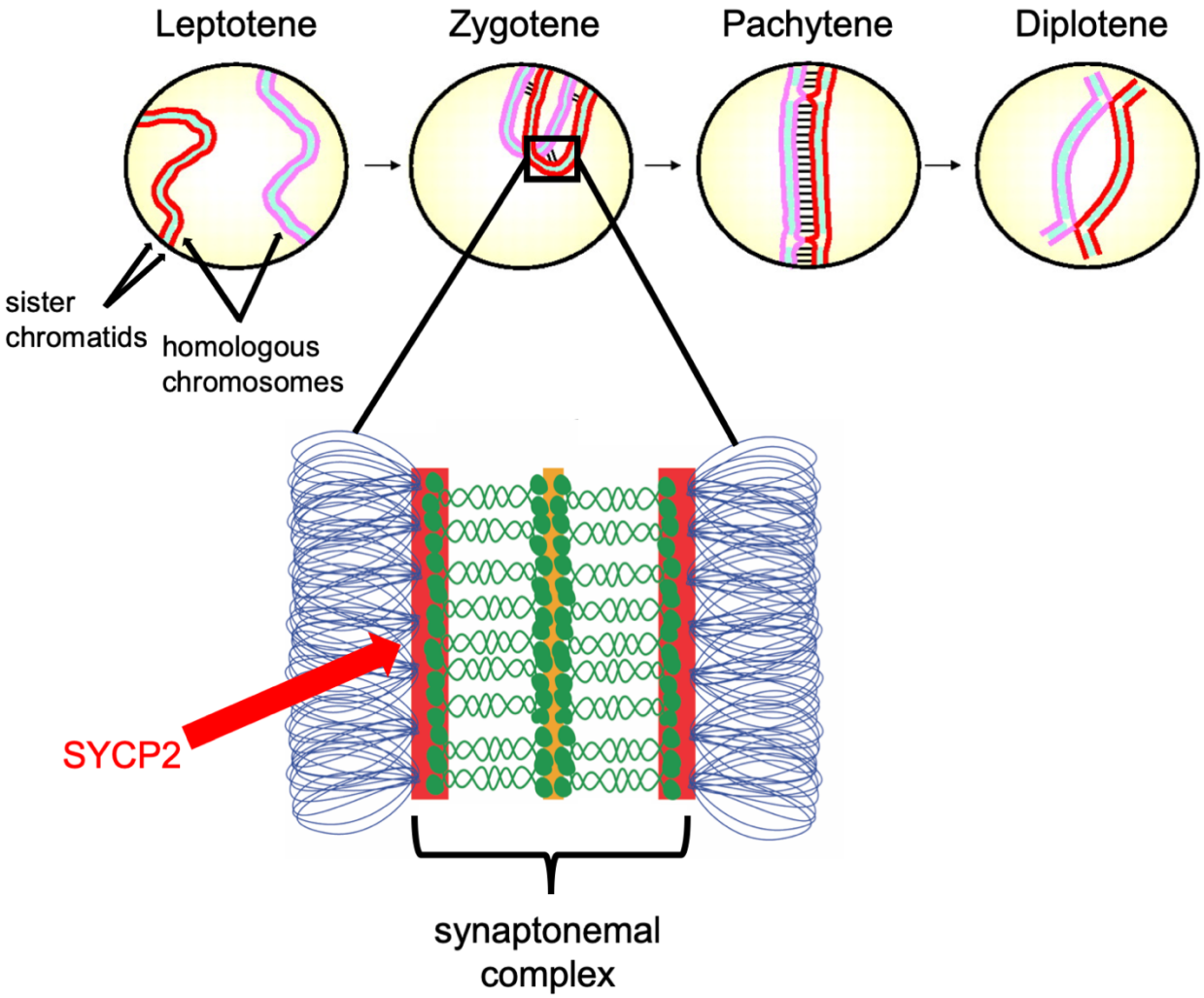


Figure 4.1. The synaptonemal complex in prophase I of meiosis
Meiosis is a type of cell division in gametogenesis that reduces the number of chromosomes to generate haploid cells. The accurate segregation of homologous chromosomes requires their pairing, which occurs during prophase I of meiosis through chromosome synapsis and recombination (top; adapted from [190]). This is facilitated by the synaptonemal complex (bottom), a zipper-like structure with axial elements (red), transverse filaments (green) and a central element (orange) that interacts with homologous chromosomes (blue).

4.2. Methods

4.2.1. Western blot analysis

Twenty-five million lymphoblastoid cells were collected for each protein extraction. Cells were pelleted for 5 minutes at 200 x g, washed in PBS (Gibco), and resuspended in 200 µl Radioimmunoprecipitation assay buffer (RIPA; Sigma-Aldrich) with cOmplete protease inhibitors (Roche). After trituration and agitation for 30 minutes at 4°C, cellular debris was pelleted for 20 minutes at 17,000 x g and 6X Laemmli buffer (375 mM Tris-HCl, 9% SDS, 50% glycerol, 9% ß-mercaptoethanol, and 0.03% bromophenol blue [all Sigma]) was added to the supernatant to a final concentration of 1X. Samples were passed through a 27-gauge syringe 10 times and boiled at 95°C for 10 minutes. Samples and a positive control of human testis tissue lysate (Abcam, catalog number ab30257) were loaded into NuPAGE 4-12% Bis-Tris Protein gels (Novex) and run with NuPAGE MOPS SDS Running Buffer (Novex) in an XCell SureLock Mini-Cell Electrophoresis System (Novex) according to manufacturer instructions. Gels were transferred to 0.2 µm nitrocellulose membranes (Bio-Rad) using NuPAGE Transfer Buffer and an XCell II Blot Module (Invitrogen) overnight at 100 mA. Membranes were probed with primary antibodies anti-SYCP2 (EMD-Millipore, catalog number ABE2622) and anti-GAPDH (Cell Signaling Technology, catalog number 5174) and secondary antibody IRDye 800CW Donkey anti-Rabbit IgG (LI-COR). Membranes were visualized using the Odyssey Fc Imaging System (LI-COR) and signal intensities were quantified in ImageJ (Version 1.46).

4.2.2. Construction of yeast strains

All strains used in this study (Table 4.1) are isogenic to BR1919-8B[191]. Strains were created by standard genetic crosses and transformation procedures. For the development of strains AM3762 and AM4282, a *TRP1:*P_{GAL1} promoter cassette was amplified from *pFA6a-TRP1-*P_{GAL1}[192] using Velocity polymerase (Bioline) and primers AJM1741-AJM1742 (Table 4.2).

4.2.3. Cytological analysis and imaging

Induction, sporulation, chromosome spreading, immunostaining, imaging and analysis were performed on diploid strains as previously described[193,194] using the following parameters. Overnight cultures of YAM2592 (wild type) and AM3762 (P_{GAL1}-*RED1*) were resuspended in 2% potassium acetate and split into two cultures per strain. Immediately after resuspension, one culture per strain was induced with 2 μm β-estradiol (Sigma E2257, prepared in DMSO) and the second received the corresponding volume of DMSO as an uninduced control. Cells were collected at 26 hours post sporulation and induction for chromosome spreading. The following primary antibodies were used for immunostaining at a 1:100 dilution: rabbit anti-Red1 (kind gift of G.S. Roeder[195]) and affinity-purified rabbit anti-Zip1 (raised at YenZym Antibodies, LLC, against a C-terminal fragment of Zip1[196]). The secondary antibody, donkey anti-rabbit conjugated to Alexa Fluor 488 (Abcam), was used at a 1:200 dilution. Imaging was carried out for eight different chromosome spreads derived from two distinct cultures per condition using a Deltavision RT imaging system (Applied Precision) adapted to an Olympus (IX71) microscope. Zip1 lengths were measured using the Softworx Measure Distance Tool.

4.2.4. RNA extraction, cDNA synthesis, and quantitative RT-PCR of *S. cerevisiae*

Induction and sporulation were performed on diploid strains as previously described[193,194]. Overnight cultures of YAM2592 (wild type), AM4063 (wild type), AM3762 (P_{GAL1}-*RED1*), AM4282 (P_{GAL1}-*RED1*), AM4283 (Δ*zip1*), AM4284 (Δ*red1*) and AM4286 (Δ*red1*) were resuspended in 2% potassium acetate and split into two cultures per strain. Immediately after resuspension, one culture per strain was induced with 2 μm β-estradiol (Sigma E2257, prepared in DMSO) and the second received the corresponding volume of DMSO as an uninduced control. Cells from 5 ml cultures were collected at 0, 6, and 26 hours post sporulation and induction and saved in 500 μl TRIzol (Invitrogen). RNA extraction, cDNA synthesis, and quantitative RT-PCR in *S. cerevisiae* were performed as described in the corresponding materials and methods section for LCLs

(Section 2.2.6) with the following differences. Cells were disrupted by vortex with 0.5 mm glass beads (BioSpec Products) for 30 minutes at 4°C before phase separation by MaXtract High Density Tubes (Qiagen). Total RNA was isolated using the RNeasy Mini Kit (Qiagen) and RNA was converted to cDNA using 60 ng of RNA per reaction, oligo(dT) primers, and the SuperScript III First-Strand Synthesis System (Invitrogen). For quantitative RT-PCR, cDNA was amplified with primers SS_326F-SS_341R (given in Table 4.2). Standard curves were routinely employed to quantify amplicons from each primer pair and assess expression of each respective gene in *S. cerevisiae* cells.

4.2.5. IMIGC participant recruitment

Patients attended the Centre of Reproductive Medicine and Andrology (CeRA) of the University Hospital Münster (UKM) for infertility treatment and were diagnosed using WHO reference ranges for semen parameters[39]. After routine clinical diagnostics, patients with explained infertility including malignant disease, exposure to chemotherapy or radiation, structural and numerical chromosomal aberrations, and YCMD were excluded from the study. All participants gave written informed consent for evaluation of their clinical data and genetic analysis of their DNA samples, according to protocols approved by the Ethics Committee of the State Medical Board and the Medical Faculty in Münster (Kennzeichen 2010-578-f-S).

4.2.6. Exome sequencing and analysis

Exome sequencing (ES) was performed in 520 patients with diverse infertility phenotypes to identify possible deleterious sequence variants which might be causal for male infertility. Genomic DNA was isolated using standard procedures as previously described[197]. Samples were prepared, enriched, and indexed for ES according to the manufacturer's protocol for SureSelect^{QXT} Target Enrichment for Illumina Multiplexed Sequencing Featuring Transposase-Based Library Prep Technology (Agilent). For multiplexed sequencing, libraries were index-

tagged using appropriate pairs of index primers. To capture libraries, SureSelect^{XT} Human All Exon Kits (v4, v5 and v6) were used. Quantity and quality of the libraries were determined with an Agilent TapeStation 2200 and the final concentration was adjusted to 1.6 pM. Sequencing was performed on the Illumina HiScan®SQ System, the Illumina NextSeq®500 System or the Illumina HiSeqX® System using the TruSeq SBS Kit v3 - HS (200 cycles), the NextSeq 500 V2 High-Output Kit (300 cycles) or the HiSeq Rapid SBS Kit V2 (300 cycles) respectively.

Reads were trimmed with Cutadapt v1.15[198] and aligned to GRCh37.p13 using BWA-MEM v0.7.17[199]. Base quality recalibration and variant calling were performed using the GATK toolkit v3.8[200] with haplotype caller according to their best practice recommendations. Resulting variants were annotated with Ensembl Variant Effect Predictor[201].

Before evaluation of variants in *SYCP2*, participants with likely pathogenic or pathogenic variants in *TEX11, NR5A1*, and *DMRT1* were excluded from the study[107,197,202]. Rare variants in *SYCP2* were selected by a minor allele frequency (MAF) < 1% in the Genome Aggregation Database (gnomAD) browser (http://gnomad.broadinstitute.org)[205] and assessed for functional consequences at the protein level.

4.2.7. Sanger sequencing

Identified variants in *SYCP2* were verified by Sanger sequencing according to standard procedures using primers SS_369F-SS_372R (given in Table 4.2) and a 3730 DNA Analyzer (Applied Biosystems)[197]. When parental DNA was available, segregation was analyzed within the family using Sanger sequencing. Sequence analysis and visualization of the chromatograms was performed with CodonCode Aligner software (Version 8.0.1).

4.2.8. Testicular biopsy histopathology

Testicular biopsies were fixed overnight in Bouin's solution, washed with 70% ethanol, and embedded in paraffin. Subsequently, 5 µm sections were stained with Periodic acid-Schiff (PAS)

and hematoxylin according to previously published protocols[203]. Slides were evaluated and documented using an Axioskop microscope (Zeiss, Oberkochen, Germany).

4.2.9. Statistical analyses

For Western blot analysis, signal intensities of SYCP2 relative to GAPDH for three DGAP230 LCL replicates and three distinct age- and sex-matched control LCLs were assessed and evaluated using an unpaired one-tailed t-test (Excel). Statistical significance for protein analyses were determined by a p-value of $p<0.05$ and figure error bars show standard error of the mean. For yeast cytology, at least 50 surface-spread meiotic nuclei were assessed per condition using eight different chromosome spreads derived from two distinct cultures. Graphpad Prism7 software was used for scatterplot generation of Zip1 filament length (with error bars indicating standard error of the mean) and statistical significance was determined by a Mann-Whitney U test, using VassarStats Concepts and Applications of Inferential Statistics (http://vassarstats.net/utest.html). Yeast qPCR experiments were performed in triplicate in two different strain backgrounds (N=6), normalized to *ACT1* expression, and evaluated using an unpaired one-tailed t-test for *RED1* and an unpaired two-tailed t-test for *ZIP1* (Excel). Statistical significance for RNA analyses were determined by a p-value of $p<0.05$ and figure error bars show standard error of the mean.

Table 4.1. Yeast strains used in chapter 4

Strain Name	Alias	Genotype	Source	Notes
YAM1252	wild type background	*lys2ΔNhe his4-260,519* *lys2ΔNhe his4-260,519* *leu2-3,112 MATα trp1-289* *leu2-3,112 MAT**a** trp1-289* *ura3-1 thr1-4 ade2-1* *ura3-1 thr1-4 ade2-1*	Amy MacQueen	BR1919 background from[191]
YAM2592 and AM4063	Ecm11-epitope tag heterozygote	YAM1252 *ECM11* *ECM11-13MYC::kanMX4* *ndt80Δ::LEU2* *ndt80Δ::LEU2*	Amy MacQueen	For cytology and qPCR; at least one parent is different between the two strains (although identical in strain background and genotype)
AM3762 and AM4282	P_{GAL1}-*RED1*	AM4063 *TRP1-P_{GAL1}-RED1* *RED1* *ura3::P_{GPD1}-GAL4(848).ER::URA3* *URA3*	This study	For cytology and qPCR; at least one parent is different between the two strains (although identical in strain background and genotype)
AM4283	Δ*zip1*	AM4063 *zip1Δ::URA3* *zip1Δ::URA3*	Amy MacQueen	For qPCR negative control
AM4284 and AM4286	Δ*red1*	AM4063 *red1Δ::HYG* *red1Δ::HYG* *ura3::PGPD1-GAL4(848).ER::URA3* *URA3*	Amy MacQueen	For qPCR negative control; at least one parent is different between the two strains (although identical in strain background and genotype)

Table 4.2. Primers used in chapter 4

Name	Sequence 5'-3'	Target region	Description
SS_326F	TTGAGACTGGCATCGC AATG	*RED1, S. cerevisiae*	For qPCR with SS_327R; amplicon = 89 bp
SS_327R	TTTGGATTGAGGGACA CTGC	*RED1, S. cerevisiae*	For qPCR with SS_326F; amplicon = 89 bp
SS_332F	AACAAAAAGGAGGCGG ATGC	*ZIP1, S. cerevisiae*	For qPCR with SS_333R; amplicon = 96 bp
SS_333R	TTCGCTCAACTGACCT GAAC	*ZIP1, S. cerevisiae*	For qPCR with SS_332F; amplicon = 96 bp
SS_340F	TGGATTCTGAGGTTGC TGCTTTG	*ACT1* cDNA, *S. cerevisiae*	For qPCR with SS_341R; amplicon = 103 bp
SS_341R	ACGATAGATGGGAAGA CAGCAC	*ACT1* cDNA, *S. cerevisiae*	For qPCR with SS_340F; amplicon = 103 bp
AJM1741	ATTTTTTAATCAGTGAG GACCACAAAGGGACAG CAAATACGGTGATAAG AGAATTCGAGCTCGTTT AAAC	*pFA6a-TRP1-*P_{GAL1} (underlined sequence) and *RED1* promoter, *S. cerevisiae*	Forward primer to amplify *TRP1-*P_{GAL1} and place it upstream of *RED1*. Used with AJM1742
AJM1742	AAGTCATTTTTCAGGCA AACACCAAAAATCTTTT TCTTCAAACCTTCCATT TTGAGATCCGGGTTTT	*pFA6a-TRP1-*P_{GAL1} (underlined sequence) and *RED1* promoter, *S. cerevisiae*	Reverse primer to amplify *TRP1-*P_{GAL1} and place it upstream of *RED1*. Used with AJM1741
AJM1743	ACGATTTCGCAGCAGG ATCAGATGG	*RED1* promoter, *S. cerevisiae*	For genotyping of *TRP1-*P_{GAL1}*-RED1*
SS_369F	TGGGCCATGAGTAGAC AAGTG	*SYCP2*, human	For PCR with SS_370R to amplify variant c.2793_2797del in exon 31; amplicon = 703 bp (also used for Sanger sequencing)
SS_370R	ACCTCTCTGGAAATAAG TTGTTTTGA	*SYCP2*, human	For PCR with SS_369F to amplify variant c.2793_2797del in exon 31; amplicon = 703 bp (also used for Sanger sequencing)
SS_371F	TCACTAGATTCAGACAT CTGTTTTG	*SYCP2*, human	For PCR with SS_372R to amplify variant c.3067_3071del in exon 33; amplicon = 564 bp (also used for Sanger sequencing)
SS_372R	GATGATAACTGGAATG GGAAGAT	*SYCP2*, human	For PCR with SS_371F to amplify variant c.3067_3071del in exon 33; amplicon = 564 bp (also used for Sanger sequencing)

4.3. Results

4.3.1. Analysis of the impact of *SYCP2* misexpression on severe oligospermia

In response to the intriguing role that SYCP2 plays in meiosis and spermatogenesis, we pursued identification of *SYCP2* overexpression at the protein level. Using Western blot, we found that DGAP230 lymphoblastoid cells express over five-times more SYCP2 than age- and sex-matched controls, which achieved statistical significance by an unpaired one-tailed t-test ($p<0.0032$) (Figure 4.2).

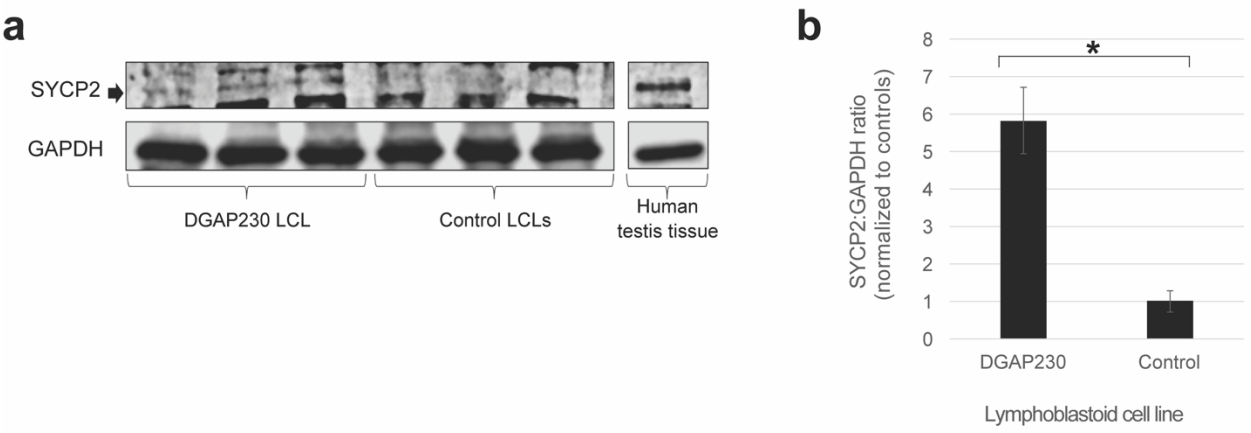


Figure 4.2. Western blot of SYCP2 in DGAP230 and age- and sex-matched control LCLs
a) The SYCP2 band, which has been confirmed by a human testis tissue lysate positive control, shows more abundant levels of expression in three biological replicates of DGAP230 lymphoblastoid cells compared to control LCLs from three different karyotypically normal fathers. **b)** Three technical replicates of each biological replicate were quantified in ImageJ (Version 1.46) and averaged. Biological replicates were then grouped into DGAP230 and control LCLs for graphing and statistical analysis. Results were found to be statistically significant by an unpaired one-tailed t-test ($N=3$; $p<0.0032$) and are graphed as mean ± standard error. * = $p<0.05$.

Based upon DGAP230's phenotype, overexpression of *SYCP2* at the RNA and protein levels in the LCL, and current literature on *SYCP2*, we hypothesized that *SYCP2* misexpression may lead to defects in meiosis, resulting in problems with spermatogenesis, thus leading to DGAP230's phenotype of severe oligospermia.

SYCP2 misexpression may inhibit proper meiosis in DGAP230 because of its role in facilitating homologous chromosome synapsis in meiosis I. It could influence meiosis in two opposing ways. First, it is possible that accumulation of SYCP2 leads to excess synaptonemal complex formation, gluing together synapsed chromosomes and preventing proper separation of homologs in anaphase I. Alternatively, excess SYCP2 may outcompete binding of other axial element proteins such as synaptonemal complex protein 3 (SYCP3), preventing proper stoichiometry and causing poor integrity of the synaptonemal complex with resulting asynapsis.

Because meiosis is evolutionarily conserved, the unicellular eukaryote *Saccharomyces cerevisiae* serves as an excellent model organism to study the molecular mechanisms of meiosis[204]. We positioned the yeast functional homolog of *SYCP2*, *RED1*, under control of the inducible promoter P_{GAL1}[187] in a strain background that contains *GAL4-ER*, which induces constitutive and overexpression from P_{GAL1} promoters in the presence of β-estradiol (Figure 4.3a,b).

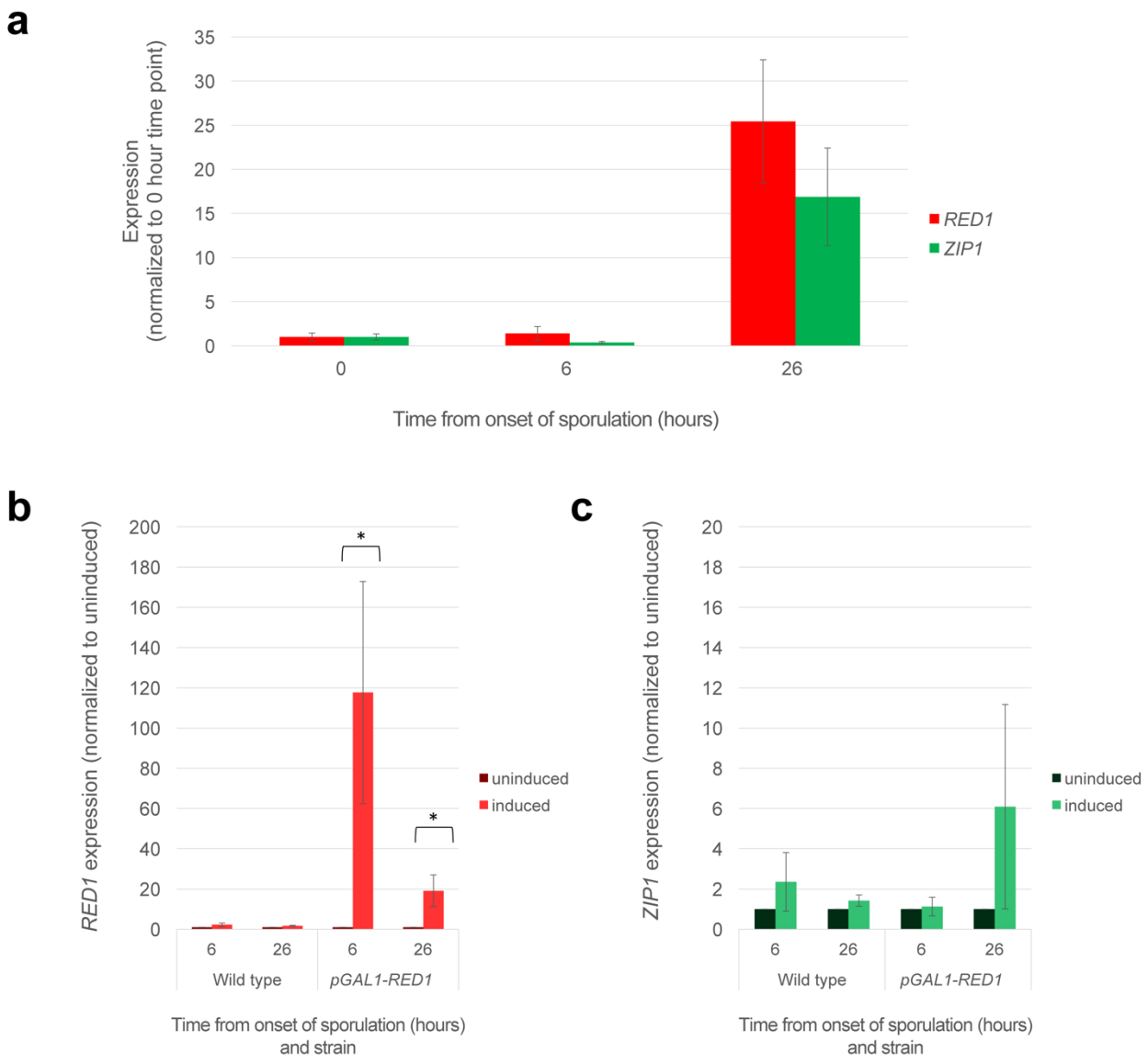


Figure 4.3. Control qPCR experiments for *RED1* and *ZIP1* transcription in sporulating *S. cerevisiae*

a) Measurement of *RED1* and *ZIP1* transcript levels in wild type sporulating cells at various time points reveals gene induction between 6 and 26 hours after placement into sporulation medium. **b)** *RED1* is both constitutively expressed and overexpressed in the induced *pGAL1-RED1* strain background compared to uninduced *pGAL1-RED1*. These are statistically significant according to an unpaired one-tailed t-test ($p<0.0201$ and $p<0.026$, respectively), when tested in triplicate in two different strain backgrounds (N=6). * = $p<0.05$. **c)** *ZIP1* levels are not altered significantly with the induction by β-estradiol in either wild type or *pGAL1-RED1* strains; differences were deemed not statistically significant according to an unpaired two-tailed t-test when tested in triplicate in two different strain backgrounds. All results display mean ± standard error.

To learn how induction of excess Red1 impacts its function as an axial element protein, we used immunolocalization to label Red1 on surface-spread meiotic nuclei. Excess Red1 shows an aggregation of signal, which we interpret to be a polycomplex structure (Figure 4.4a). We hypothesized that mislocalization of Red1 may influence its ability to serve as a scaffold for the synaptonemal complex. To determine how this impacts structural integrity of the synaptonemal complex, we used immunostaining of surface-spread meiotic nuclei to visualize Zip1, a transverse filament in the yeast synaptonemal complex that forms linear structures at the interface of each synapsed chromosome pair (Figure 4.4b,c). We discovered that excess Red1 inhibits Zip1 formation, a result that was found to be statistically significant after quantifying the cumulative length of Zip1 per nucleus in a population of 50 cells (Figure 4.4d). This inhibition was confirmed to be transcription-independent, ensuring that loss of Zip1 on meiotic chromosomes results from mislocalization of Red1 as opposed to a decrease in total levels of Zip1 (Figure 4.3c).

These results support the latter model, where excess axial element protein disrupts the integrity of the synaptonemal complex, leading to asynapsis, or the inability of homologous chromosomes to pair.

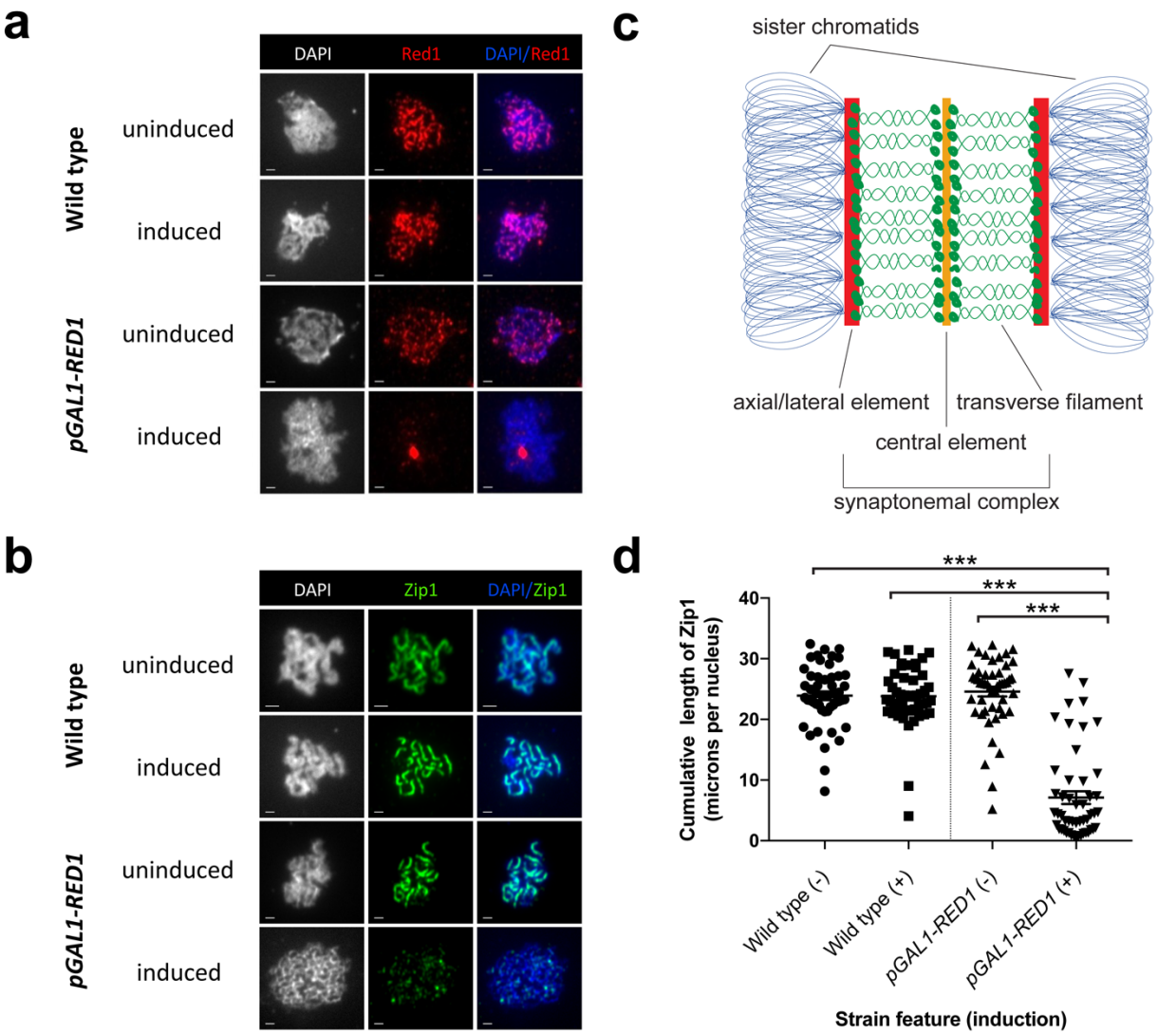


Figure 4.4. Consequences of synaptonemal complex axial element misexpression on meiosis in *S. cerevisiae*

a) Anti-Red1 immuno- and DAPI-staining of surface-spread pachytene chromosomes reveal constitutive *RED1* induction during sporulation leads to Red1 polycomplex formation (scale bar represents one micron in length). **b)** Anti-Zip1 immuno- and DAPI-staining of surface-spread pachytene chromosomes show constitutive *RED1* induction during sporulation obliterates Zip1 formation at the interface of homologous chromosomes (scale bar represents one micron in length). **c)** The synaptonemal complex is evolutionary conserved between mammals and budding yeast. SYCP2 and Red1 are axial element functional homologs and SYCP1 and Zip1 are transverse filament homologs in the mammalian and *S. cerevisiae* synaptonemal complex, respectively. **d)** Quantification of the cumulative length of synaptonemal complex Zip1 in surface-spread meiotic nuclei demonstrates a statistically significant decrease upon induction of *RED1*, as analyzed by Mann-Whitney U test (N = 50). Error bars represent mean ± standard error. *** = $p<0.0001$.

4.3.2. Identification of additional male infertility cases with *SYCP2* pathogenic variants

Current evidence from our work and the literature suggests that proper SYCP2 dosage is important for its function. In this project, we have found that misexpression of *SYCP2* leads to a loss of function by stoichiometric imbalance. On the other hand, *Sycp2* knockout mice also have an infertility phenotype[186]. Ultimately, it appears that either too much or too little SYCP2 results in a loss of function leading to infertility.

Under the assumption that genetic variants causing infertility would be eliminated from the general population by natural selection, we analyzed allele frequencies of *SYCP2* variants from over 140,000 genomes in the Genome Aggregation Database (gnomAD) browser[205]. We found that *SYCP2* is severely depleted for loss of function mutations including stop-gained and essential splice sites variants, as demonstrated by a pLI (probability of loss-of-function intolerance) of 1.00[205]. This extreme intolerance to loss of function could be explained by the inability to segregate these mutations due to a phenotype of infertility, which would further support *SYCP2* pathogenicity in humans. While we have found that *SYCP2* is significantly depleted for loss-of-function variants, these aggregation datasets lack phenotypic information that would be necessary to assess whether pathogenic variants lead to infertility. To test the model that pathogenic *SYCP2* variants are more prevalent in infertile men than in fertile men, we searched for *SYCP2* variants in the male infertility cohort from the International Male Infertility Genomics Consortium (IMIGC).

Patient 1 from the IMIGC presented with a 2.5-year history of infertility at age 27. He was diagnosed with cryptozoospermia (sperm concentration <0.1 million/ml) according to the current WHO reference ranges for semen parameters[39]. Patient 1 has no dysmorphic features and has normal serum levels of FSH, LH and testosterone. Genetic testing for structural and numerical chromosomal aberrations, YCMD, *TEX11, NR5A1*, and *DMRT1* were negative. ES and subsequent validation by Sanger sequencing revealed a heterozygous deletion in exon 31 of *SYCP2*: c.2793_2797del causing p.Lys932SerfsTer3 (Figure 4.5a). This variant is absent in both

gnomAD and Trans-Omics for Precision Medicine (TOPMed; https://bravo.sph.umich.edu) databases[205]. Segregation was assessed in all available family members, which revealed that the deletion is inherited from the mother while the father is wild type at this position (Figure 4.5a).

Patient 2 presented with a 17-year history of infertility at age 39. He was diagnosed with azoospermia and histopathological analysis from TESE revealed a phenotype of meiotic arrest at the pachytene spermatocyte stage (Figure 4.5b). Patient 2 has borderline testicular atrophy (right: 12 ml, left: 11 ml [reference range 12-15 ml]), elevated serum levels of FSH (44.9 IU/l [reference range 1-7 IU/l]) and LH (12.4 IU/l [reference range 2-10 IU/l]), and low serum levels of testosterone (8.2 nmol/l [reference range >12 nmol/l]). Genetic testing for structural and numerical chromosomal aberrations, YCMD, *TEX11, NR5A1*, and *DMRT1* were negative. ES and subsequent validation by Sanger sequencing revealed a heterozygous deletion in exon 33 of *SYCP2*: c.3067_3071del causing p.Lys1023LeufsTer2 (Figure 4.5c). This variant is absent in both gnomAD and TOPMed databases[205].

In order to predict the impact of the early termination of *SYCP2* resulting from these variants, we compared them to the coiled-coil domain region identified to be responsible for fertility in the *Sycp2* knockout mouse[186]. By searching for homology to the mouse *Sycp2* coiled-coil domain using a basic local alignment search tool (BLAST; https://blast.ncbi.nlm.nih.gov/), we identified the putative coiled-coil domain of human *SYCP2*[186,206]. The domain resides in exons 41-44, encoding residues 1408-1505. We discovered that both termination events reside upstream of the coiled-coil domain, suggesting that these alleles would encode nonfunctional truncated peptides.

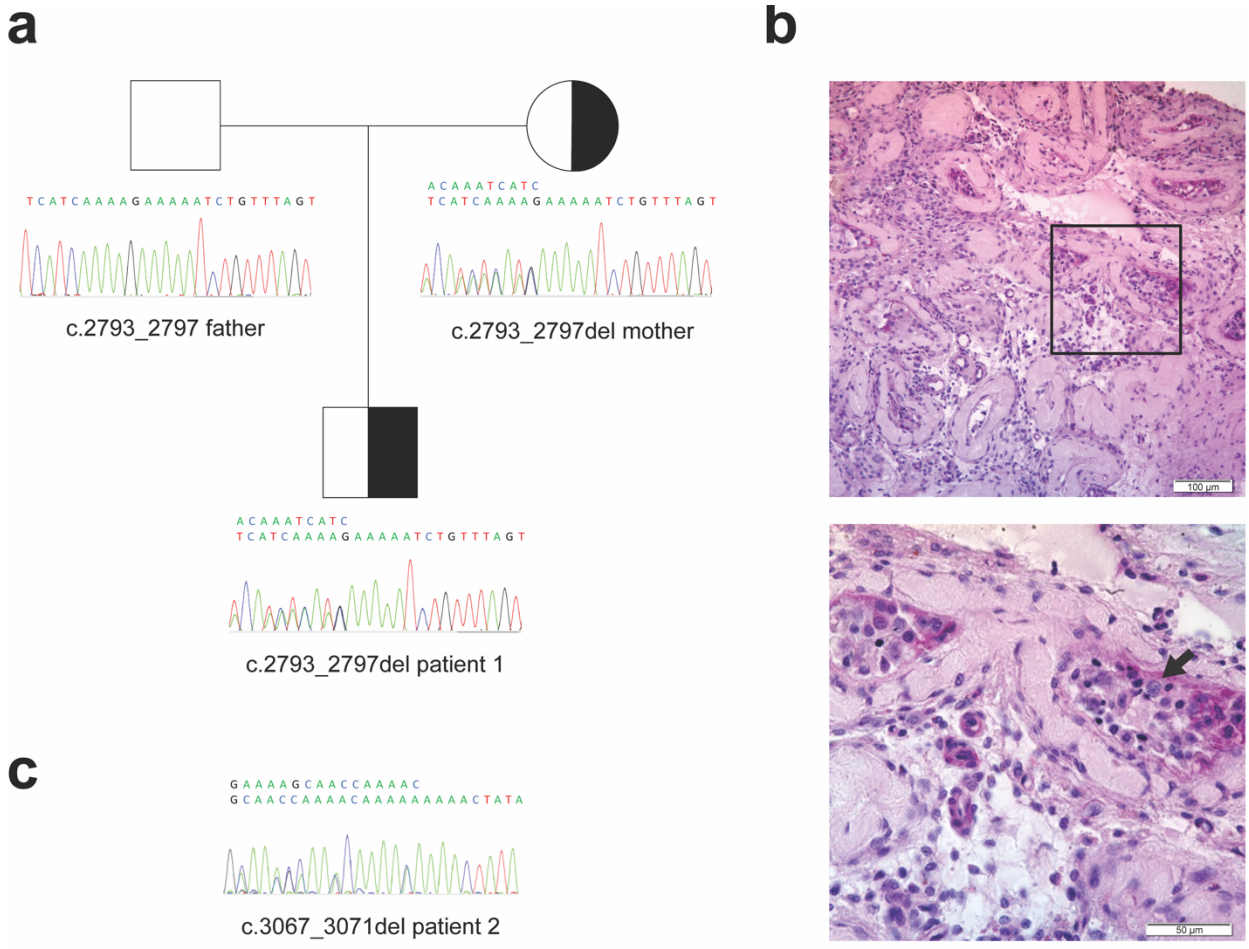


Figure 4.5. Identification of heterozygous five base pair deletion frameshift mutations in *SYCP2* from IMIGC participants

a) Patient 1 with cryptozoospermia carries a heterozygous deletion in *SYCP2* (c.2793_2797del, p.Lys932SerfsTer3) which was inherited from his mother while his father is wild type in this position. **b)** Histological PAS staining of a testis biopsy from Patient 2 shows a phenotype of meiotic arrest at the pachytene spermatocyte stage leading to nonobstructive azoospermia. The majority of the tubules in an overview section (top) are degenerated to the phenotype of tubular ghosts with a few tubular cross sections that present with spermatocytes. A closer look at the boxed section (magnified in the bottom image) shows disorganized seminiferous epithelium with a single pachytene spermatocyte (black arrow). **c)** Patient 2 carries a heterozygous deletion in *SYCP2* (c.3067_3071del, p.Lys1023LeufsTer2).

4.4. Contributions

Samantha Schilit performed the Western blot analysis. All yeast experiments were performed in the laboratory of Dr. Amy MacQueen at Wesleyan University. Dr. Amy MacQueen constructed the yeast strains and designed the cytological analysis and imaging experiments. Samantha Schilit performed the cytological analysis and imaging experiments, and designed, optimized and performed all additional yeast experiments including RNA extraction, cDNA synthesis, and qRT-PCR. Samantha Schilit initiated the collaboration with Drs. Frank Tüttelmann, Corinna Friedrich, and Sabine Kliesch from University Hospital Münster (UKM), who contributed the two patients from the IMIGC to this study.

CHAPTER 5

DISCUSSION

5.1. Conclusion

Male infertility is a common disorder among reproductive-aged couples and the majority of patients lack a specific etiologic diagnosis[51]. Understanding the precise causes of male infertility may directly inform therapies for infertile couples. For example, different recommendations will be offered to azoospermic males with the varied etiologies of AZFa deletions, Klinefelter syndrome, and mutations in *CFTR*[47,52,63].

Identifying genetic etiologies for human male infertility has been hindered by smaller pedigrees inherent to decreased reproductive fitness and genetic heterogeneity of the disorder. In addition, genetic evidence of a disorder may not be investigated thoroughly. For example, balanced reciprocal translocations identified by karyotype analysis in infertile men are rarely followed up beyond reporting a risk for segregation of unbalanced gametes. As a result, a deep investigation into single case studies can be illuminating for uncovering novel genetic etiologies for male infertility.

In this study, we identified a balanced reciprocal translocation in a severe oligospermic male designated DGAP230. While it is generally presumed that balanced reciprocal translocations reduce fertility due to production of unbalanced gametes[74] or meiotic silencing of unsynapsed chromatin[79], this does not account for the specific phenotype of severe oligospermia or azoospermia because the majority of men with balanced reciprocal translocations have normal sperm counts[70]. In addition, most men with low sperm counts and a balanced reciprocal translocation have rearrangement breakpoints that cluster in distinct genomic regions, suggesting that as opposed to a nonspecific mechanism of meiotic segregation, there may be something intrinsic to these genomic regions that is important for fertility[207,208].

In the case of DGAP230, a structural rearrangement leads to dysregulation of a synaptonemal complex gene *SYCP2*, which resides distal to one of the rearrangement

breakpoints. This impact of a chromosomal structural rearrangement on gene expression supports the finding that translocation breakpoints can influence gene expression by dysregulating genes residing within the same TAD[154].

While it is known that many different synaptonemal complex proteins have the intrinsic ability to self-assemble into polycomplexes when overexpressed, mutated, or expressed in mitotic cells[209-213], this study demonstrates the first observation of Red1 overexpression leading to polycomplex formation. We predict that aggregation may be mediated by misexpression before formation of meiotic chromosomes and expression of other axial element proteins as well as the presence of a coiled-coil domain, which facilitates protein complex interactions. The resulting asynapsis phenocopies *red1* mutants in *S. cerevisiae* as well as coiled-coil domain-deficient *Sycp2* mice[186,214]. We believe that our finding of asynapsis resulting from axial element misexpression is directly related to DGAP230's phenotype of severe oligospermia; asynapsis triggers checkpoint-mediated apoptosis of spermatocytes during spermatogenesis[68,215,216], which reduces sperm count and has been shown to lead to male-specific infertility[186]. Therefore, DGAP230's phenotype of severe oligospermia and infertility is likely due to asynapsis-triggered cell death in spermatocytes (Figure 5.1).

The identification of two novel frameshift mutations in *SYCP2* from men with cryptozoospermia and meiotic arrest through the IMIGC further supports the role of *SYCP2* in human male fertility. Both variants are extremely rare consistent with the inability to segregate these mutations in the general population due to a phenotype of infertility. This is also supported by the maternal inheritance of the variant in Patient 1, as *SYCP2*-mediated pathogenicity has been shown to cause male infertility but not female infertility in a mouse model[186].

Stop codons resulting from both frameshift mutations reside upstream of the coiled-coil domain, which is critical for functionality of the protein[186]. It is important to note, however, that both cases represent heterozygous mutations which would support an autosomal dominant disease model. This is discordant with the *SYCP2* knockout mouse model that only demonstrates

a male infertility phenotype in homozygotes[186]. However, haploinsufficiency discordance has been observed between human and mouse for the male infertility gene *SYCP3*, the other mammalian axial element in the synaptonemal complex as well as numerous other mouse models[101,217]. *SYCP2*'s high pLI also supports a haploinsufficient model, as extreme intolerance to loss-of-function according to constraint analysis is strongly correlated with haploinsufficiency[205]. While less common than autosomal recessive forms of male infertility, autosomal dominant forms have been identified for mutations in *HIWI*[90], *KLHL10*[91], *PLK4*[95], *SYCP3*[101], and *SPINK2*[98].

Another potential concern is that the phenotypes of cryptozoospermia and meiotic arrest are distinctive from each other. One explanation is that the subject with meiotic arrest may have an unidentified pathogenic variant on the other *SYCP2* allele, leading to a more severe phenotype. This phenotype gradation has been observed for mutations in *SPINK2*, which has a phenotype of azoospermia in homozygotes and oligospermia in heterozygotes[98]. It is also possible that the differences merely reflect variable expressivity, as has been observed in the male infertility genes *DBY* (Sertoli cell-only syndrome and severe hypospermatogenesis), *KLHL10* (severe oligospermia and oligospermia), *TAF4B* (azoospermia and oligosperma), *TDRD9* (azoospermia and cryptozoospermia), and *TEX11* (complete meiotic arrest and mixed testicular atrophy)[85,91,103,105-107].

It is well known that homologous chromosome synapsis is critical for spermatogenesis. Indeed, several genes implicated in human male infertility are members of the synaptonemal complex (*SYCP3* and *SYCE1*) or are otherwise required for synapsis (*SPO11, MEIOB, TEX11*, and *TEX15*)[76,87,100,101,106,107,110]. Before this study, *SYCP2* was considered a strong candidate gene for human male infertility because it encodes a protein that interacts directly with SYCP3 and TEX11, serves as an axial element in the synaptonemal complex, and is required for male fertility in the mouse[186,187,218]. DGAP230 and the two participants from the IMIGC represent the first human cases of putative *SYCP2*-mediated male infertility.

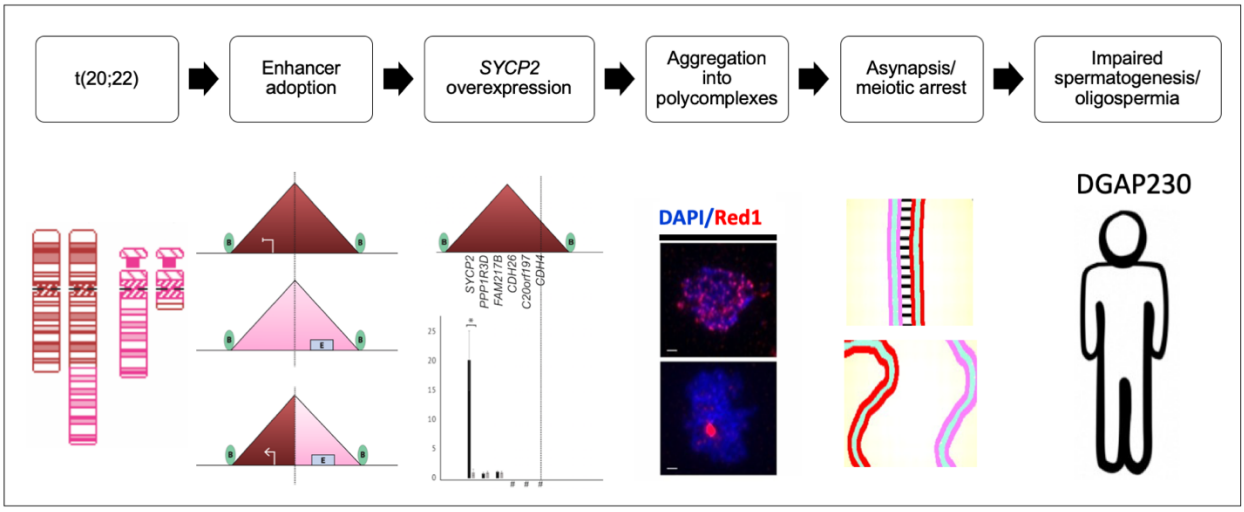


Figure 5.1. Overall summary of cytogenetic and molecular etiology for DGAP230's severe oligospermia
DGAP230 has a balanced translocation that, through a mechanism of enhancer adoption, leads to *SYCP2* overexpression. Excess axial elements aggregate into polycomplexes, preventing synaptonemal complex formation and leading to asynapsis and subsequent meiotic arrest that results in DGAP230's phenotype of severe oligospermia.

5.2. Future directions

The future directions for this project address limitations of the dissertation research and advance our understanding of the role of *SYCP2* in human male infertility.

First, the impact of axial element overexpression in the yeast model cannot directly demonstrate the downstream consequences of asynapsis. To test if *SYCP2* misexpression-mediated meiotic defects lead to impaired spermatogenesis, we are creating a transgenic mouse that models the dysregulation (Figure 5.2). In collaboration with Dr. Channabasavaiah Gurumurthy, Director of the University of Nebraska Medical Center (UNMC) Mouse Genome Engineering Core Facility, we are using his well-established pronuclear injection-based targeted transgenesis (PITT) technique (Appendix B) to make a transgenic *Sycp2* overexpression (*Sycp2 cTg*$^{+/+}$) mouse with inducible, ubiquitous and stable expression of *Sycp2*. While traditional transgenic techniques using pronuclear injection suffer from random integration, which leads to variation in copy number, varied expression state of the integration site, and potential endogenous gene disruption, PITT ameliorates these problems by performing pronuclear injection in embryos from a C57BL/6 seed mouse that has a "landing pad" at a predetermined genomic locus, facilitating targeted integration[219,220]. The landing pad includes an *attP* site, which enables efficient ϕC31 integrase-mediated recombination when a donor DNA template with the transgene construct and an *attB* site is introduced[221]. This landing pad resides in the *ROSA26* locus, providing stable and ubiquitous expression of the integrated transgene[219].

As of this writing, I have created a transgene construct that was used by Dr. Gurumurthy and his collaborator, Dr. Masato Ohtsuka from Tokai University in Japan, for microinjection and genotyping of offspring to identify transgenic founders.

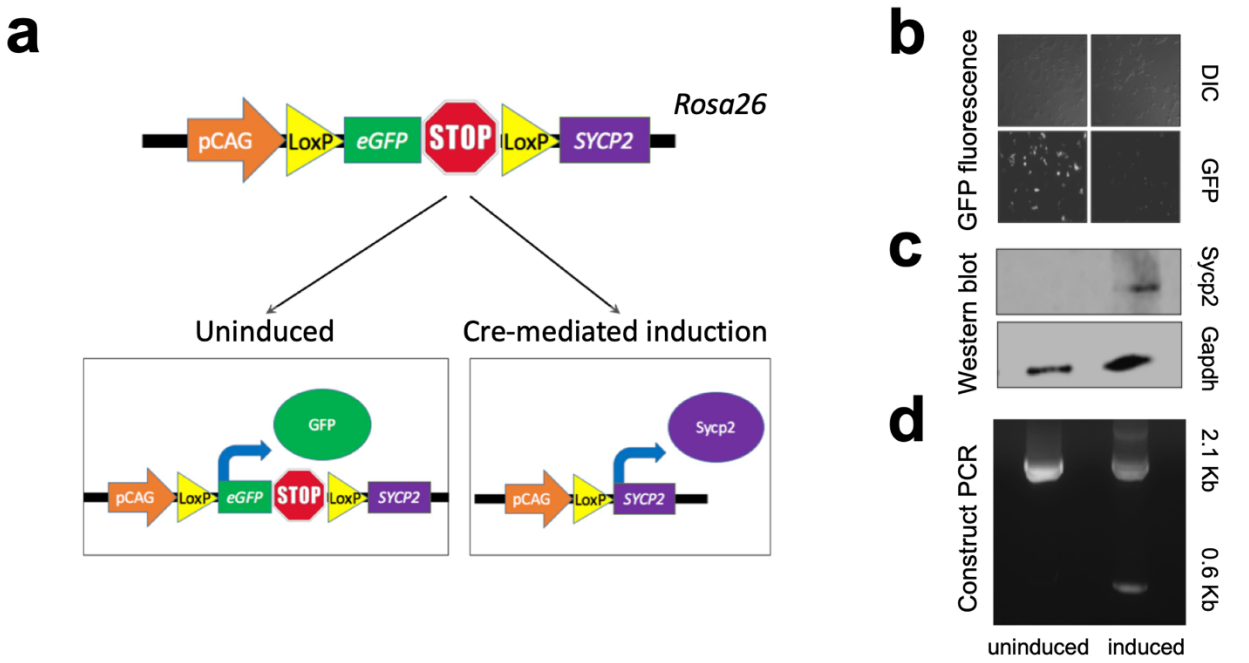


Figure 5.2. *Sycp2* overexpression transgene construct
a) Schematic of the transgene, which contains a genetically inducible system. *LoxP* sites flank a stop codon upstream of the *Sycp2* cDNA sequence. Cre recombinase is used to induce translation of *Sycp2* transcripts by excising the upstream stop codon. **b-d)** The *Sycp2* overexpression construct was transfected into 239T cells without (uninduced) or with (induced) a Cre-expressing construct. Assessments of GFP expression (**b**), Sycp2 protein levels (**c**), and excision of the DNA flanked by *LoxP* sites in the transgene (**d**) were used to validate the construct. GFP = green fluorescent protein; DIC = differential interference contrast.

Upon receiving transgenic founders from our collaborators, we will assay male sterility and functional and structural defects in the testes in house as has been outlined in our approved Institutional Animal Care and Use Committee (IACUC) protocol (BWH Protocol #2016N000430: Study of a Mouse Model for a Novel Human Infertility Gene, *SYCP2*) (Figure 5.3). To test for male sterility, we will mate induced male *Sycp2 Tg*$^{+/-}$; *CMV-Cre*$^{+/-}$ mice with wild type C57BL/6 females and determine the viability of offspring relative to the same cross with uninduced males (*Sycp2 cTg*$^{+/-}$). If we see reduced fertility from the *Sycp2 Tg*$^{+/-}$; *CMV-Cre*$^{+/-}$ mice, we will determine if this results from severe oligospermia by comparing sperm concentrations in induced and uninduced males. We will also look for testicular structural defects by harvesting testis tissue from induced (*Sycp2 Tg*$^{+/-}$; *CMV-Cre*$^{+/-}$) and uninduced (*Sycp2 cTg*$^{+/-}$) male mice and performing histology as well as staining of surface-spread spermatocyte nuclei. This work will more directly implicate *SYCP2* misexpression in DGAP230's phenotype of severe oligospermia.

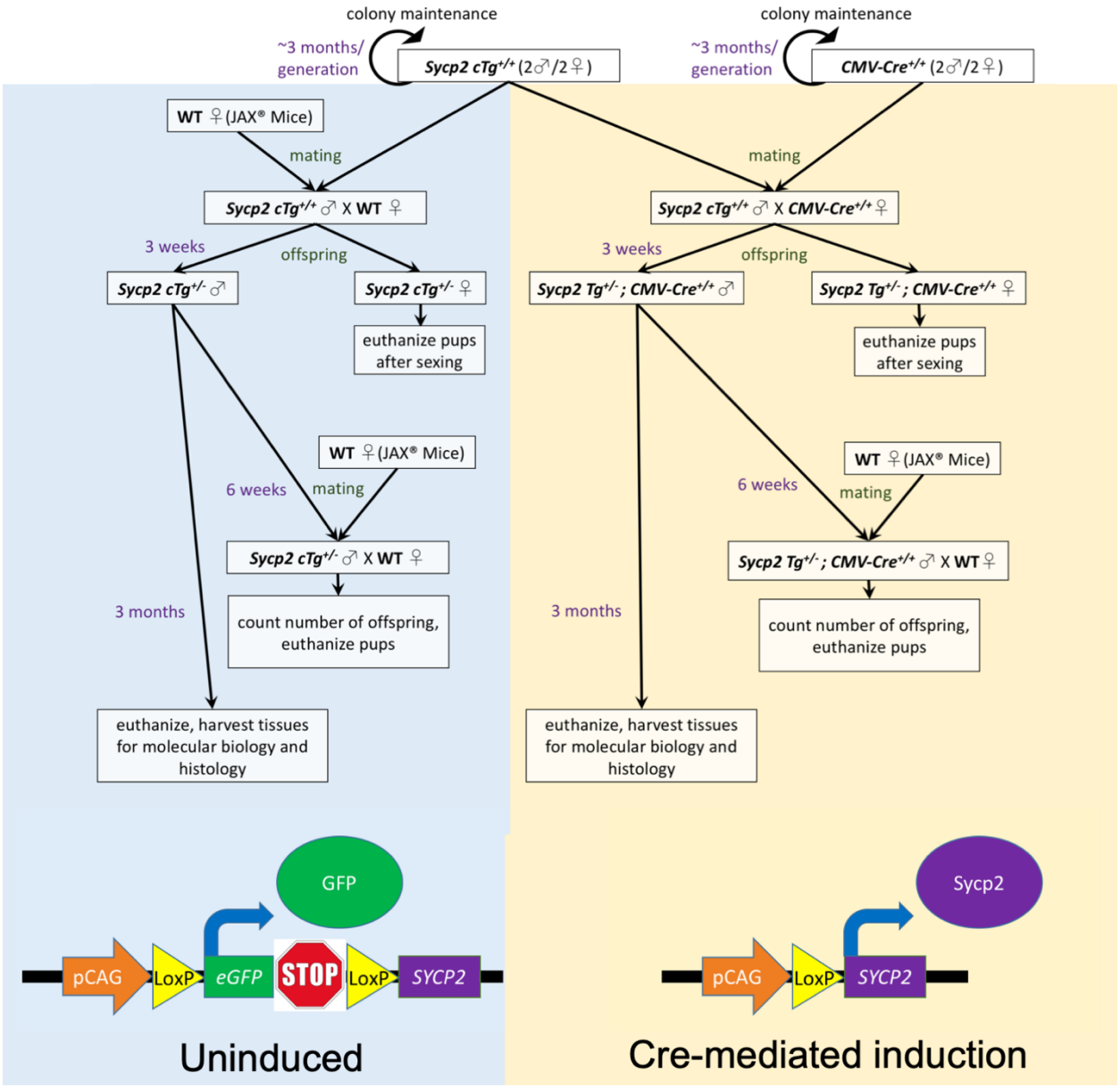


Figure 5.3. Flowchart for animal experimental protocol
This schematic demonstrates the proposed plan in the Harvard Medical Area Standing Committee on Animals protocol "Study of a Mouse Model for a Novel Human Infertility Gene, *SYCP2*" (BWH Protocol #2016N000430).

While we have identified *SYCP2* misexpression in DGAP230 and heterozygous *SYCP2* frameshift mutations in two infertile men from the IMIGC, the scope of this dissertation is still limited in understanding the prevalence of *SYCP2*-mediated infertility in the general population. To date, we have established collaborations with five male infertility cohorts to continue to test the model that pathogenic *SYCP2* variants are more prevalent in infertile men than in fertile men (Table 5.1).

In addition to identifying pathogenic *SYCP2* variants in men with unexplained infertility, we also expect to identify variants of uncertain significance (VUS). To determine the pathogenicity of interesting but uncertain *SYCP2* variants identified in infertile men, we are developing a functional assay that analyzes the impact of the VUS on synaptonemal complex initiation. This will focus on variants identified in the coiled-coil domain, which is critical for interaction of SYCP2 and SYCP3[186,205]. The coiled-coil domain encodes an α-helical structural motif that contains heptad repeats (designated **a, b, c, d, e, f,** and **g**), where the first and fourth residues (**a** and **d**) of the heptad are hydrophobic and the fifth and seventh residues (**e** and **g**) are hydrophilic. This creates a nonpolar face in three dimensional space that enables “knobs-into-holes” packing with other coiled-coil domains[222]. Variants will be prioritized for altering a residue at the **a, d, e** or **g** positions, especially if it changes the polarity of the residue, because these may be most damaging to coiled-coil domain interactions and could work in a dominant fashion. For example, one *SYCP2* VUS identified in an infertile man from GEMINI is T1481A (rs374889210). In addition to having a very low allele frequency in gnomAD of 4.0156E-5, this amino acid resides in the conserved **g** position with a residue change from polar to nonpolar[205]. To test the impact of this VUS and others on recruitment of SYCP3, the functional assay tests the binding affinity of SYCP3 to SYCP2 with the coiled-coil domain encompassing the variant in question (Figure 5.4). Our collaborations with male infertility cohorts and our development of functional assays to determine pathogenicity of prioritized variants of uncertain significance extended our findings beyond DGAP230 to suggest more generally that *SYCP2* pathogenicity may cause infertility in men.

Table 5.1. Male infertility cohorts

Cohort	Sequencing method	Research contact
International Male Infertility Genomics Consortium (IMIGC)	Exome sequencing	Dr. Frank Tüttelmann
Genetics of Male Infertility Initiative (GEMINI)	Exome sequencing	Dr. Don Conrad
Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China	Exome sequencing	Dr. Dan Zhang
Centre for Cellular and Molecular Biology, Hyderabad, India	Sanger sequencing of coiled-coil domain	Drs. Digumarthi V.S. Sudhakar and Kumarasamy Thangaraj
Instituto de Ciencias en Reproduccion Humana (Vida), León, Guanajuato, Mexico	Sanger sequencing of coiled-coil domain	Dr. Raul Piña-Aguilar

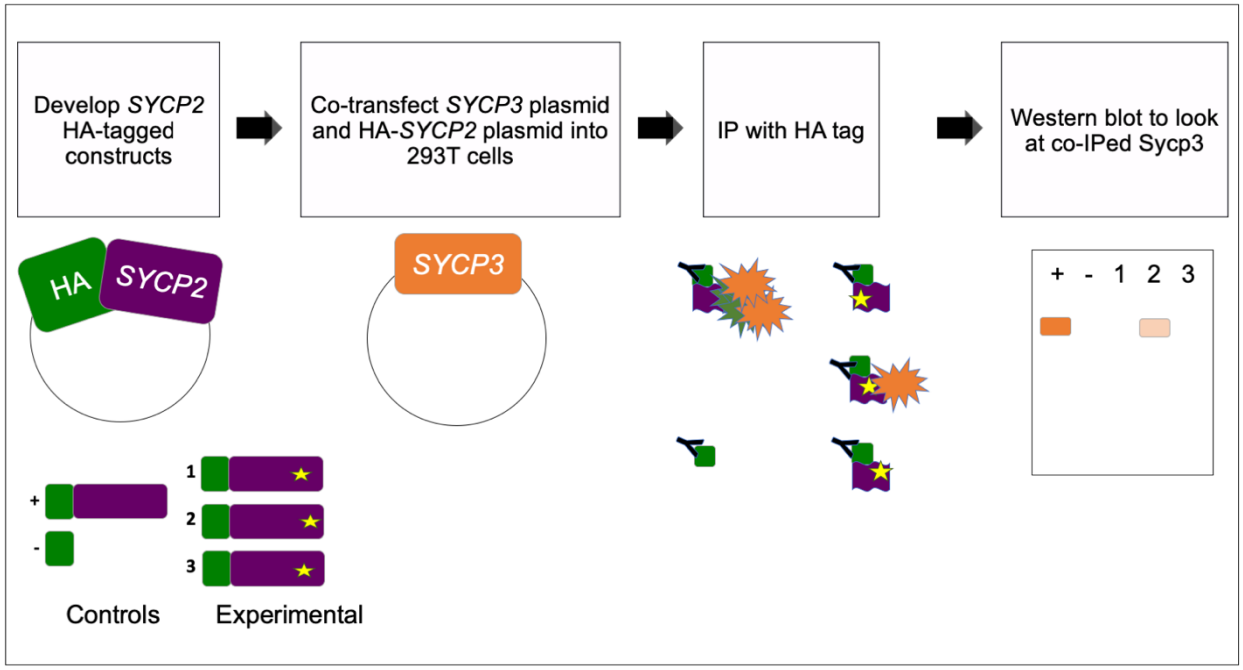


Figure 5.4. Functional assay for *SYCP2* variants of uncertain significance
This assay uses co-immunoprecipitation to determine how variants in the SYCP2 coiled-coil domain influence recruitment of SYCP3. Plasmids containing *SYCP3* and an HA-tagged *SYCP2* with the variant in question will be co-transfected into 293T cells. The positive control is the normal full length *SYCP2*, while the negative control is a construct with only the HA sequence. Each experimental construct includes a VUS in the coiled-coil domain (indicated by a yellow star). Protein lysates will be collected and immunoprecipitated for the HA tag. These lysates will then be assessed by Western blot for the presence of SYCP3 (orange box). Detection of SYCP3 indicates successful recruitment by SYCP2.

5.3. Significance of thesis findings

This dissertation offers many contributions to the field of clinical genetics. First, the impact of the specific der(20) breakpoint on male infertility in the DGAP230 case suggests that in men with balanced chromosomal rearrangements and a phenotype of infertility, pathogenesis by specific breakpoints should be considered as an alternative etiology to that of segregation of unbalanced gametes. Second, the discovery of *SYCP2* dysregulation resulting from the balanced translocation represents the first DGAP case to have an overexpressed candidate gene, which has inspired future considerations of a potential enhancer adoption mechanism in subsequent DGAP cases (Chapter 2). Our study of overexpressed *RED1* in *S. cerevisiae*, in conjunction with previous literature on the loss of *RED1* in budding yeast and of *Sycp2* in mouse[186,214], suggests that the synaptonemal complex formation is sensitive to dosage of axial elements (Chapter 4). Additionally, we developed a new technique called 3C-PCR, which differentiates alleles between normal and derivative chromosomes and has potential application for assessing compound heterozygosity in cytogenetic diagnostic labs (Chapter 3)[139]. Finally, the DGAP230 case and participants identified through the IMIGC provide the first evidence for *SYCP2*-mediated male infertility in humans.

BIBLIOGRAPHY

REFERENCES

1. Practice Committee of American Society for Reproductive, M. Definitions of infertility and recurrent pregnancy loss: a committee opinion. *Fertil. Steril.* **99**, 63 (2013).

2. Kliesch, S. Diagnosis of Male Infertility: Diagnostic Work-up of the Infertile Man. *European Urology Supplements* **13**, 73-82 (2014).

3. Thornton, A. & Young-DeMarco, L. Four Decades of Trends in Attitudes Toward Family Issues in the United States: The 1960s Through the 1990s. *Journal of Marriage and Family* **63**, 1009-1037 (2001).

4. Williams, M.E. Toward greater understanding of the psychological effects of infertility on women. *Psychotherapy in Private Practice* **16**, 7-26 (1997).

5. Hadley, R. & Hanley, T. Involuntarily childless men and the desire for fatherhood. *J Reprod Infant Psyc* **29**, 56-68 (2011).

6. Greil, A.L., Slauson-Blevins, K. & McQuillan, J. The experience of infertility: a review of recent literature. *Sociol. Health Illn.* **32**, 140-162 (2010).

7. Luk, B.H. & Loke, A.Y. The Impact of Infertility on the Psychological Well-Being, Marital Relationships, Sexual Relationships, and Quality of Life of Couples: A Systematic Review. *J. Sex Marital Ther.* **41**, 610-625 (2015).

8. Gana, K. & Jakubowska, S. Relationship between infertility-related stress and emotional distress and marital satisfaction. *J. Health Psychol.* **21**, 1043-1054 (2016).

9. Luk, B.H.K. & Loke, A.Y. Sexual satisfaction, intimacy and relationship of couples undergoing infertility treatment. *J. Reprod. Infant Psychol.*, 1-15 (2018).

10. Ethics Committee of the American Society for Reproductive Medicine. Disparities in access to effective treatment for infertility in the United States: an Ethics Committee opinion. *Fertil. Steril.* **104**, 1104-1110 (2015).

11. Wu, A.K., Odisho, A.Y., Washington, S.L., 3rd, Katz, P.P. & Smith, J.F. Out-of-pocket fertility patient expense: data from a multicenter prospective infertility cohort. *J. Urol.* **191**, 427-432 (2014).

12. Chambers, G.M., Sullivan, E.A., Ishihara, O., Chapman, M.G. & Adamson, G.D. The economic impact of assisted reproductive technology: a review of selected developed countries. *Fertil. Steril.* **91**, 2281-2294 (2009).

13. Wu, A.K., Elliott, P., Katz, P.P. & Smith, J.F. Time costs of fertility care: the hidden hardship of building a family. *Fertil. Steril.* **99**, 2025-2030 (2013).

14. Gnoth, C., Godehardt, D., Godehardt, E., Frank-Herrmann, P. & Freundl, G. Time to pregnancy: results of the German prospective study and impact on the management of infertility. *Hum. Reprod.* **18**, 1959-1966 (2003).

15. Mol, B.W., Tjon-Kon-Fat, R., Kamphuis, E. & van Wely, M. Unexplained infertility: Is it over-diagnosed and over-treated? *Best Pract. Res. Clin. Obstet. Gynaecol.* **53**, 20-29 (2018).

16. Armstrong, A. & Plowden, T.C. Ethnicity and assisted reproductive technologies. *Clin Pract (Lond)* **9**, 651-658 (2012).

17. Missmer, S.A., Seifer, D.B. & Jain, T. Cultural factors contributing to health care disparities among patients with infertility in Midwestern United States. *Fertil. Steril.* **95**, 1943-1949 (2011).

18. Seifer, D.B., Frazier, L.M. & Grainger, D.A. Disparity in assisted reproductive technologies outcomes in black women compared with white women. *Fertil. Steril.* **90**, 1701-1710 (2008).

19. Kumar, N. & Singh, A.K. Trends of male factor infertility, an important cause of infertility: A review of literature. *J. Hum. Reprod. Sci.* **8**, 191-196 (2015).

20. Culley, L., Hudson, N. & Lohan, M. Where are all the men? The marginalization of men in social scientific research on infertility. *Reprod Biomed Online* **27**, 225-235 (2013).

21. Meerabeau, L. Husbands' participation in fertility treatment: they also serve who only stand and wait. *Sociol. Health Illn.* **13**, 396-410 (1991).

22. Throsby, K. & Gill, R. “It’s Different for Men”: Masculinity and IVF. *Men and Masculinities* **6**, 330-348 (2004).

23. Nachtigall, R.D., Becker, G. & Wozny, M. The effects of gender-specific diagnosis on men's and women's response to infertility. *Fertil. Steril.* **57**, 113-121 (1992).

24. Tichenor, V., McQuillan, J., Greil, A.L., Contrepas, R. & Shreffler, K.M. The importance of fatherhood to U.S. married and cohabiting men. *Fathering* **9**, 232-251 (2011).

25. Marsiglio, W., Hutchinson, S. & Cohan, M. Young Men's Procreative Identity: Becoming Aware, Being Aware, and Being Responsible. *Journal of Marriage and Family* **63**, 123-135 (2001).

26. Webb, R.E. & Daniluk, J.C. The End of the Line:Infertile Men's Experiences of Being Unable to Produce a Child. *Men and Masculinities* **2**, 6-25 (1999).

27. O'Donnell, L., Stanton, P. & de Kretser, D.M. Endocrinology of the Male Reproductive System and Spermatogenesis. in *Endotext* (eds. De Groot, L.J.*, et al.*) (South Dartmouth (MA), 2000).

28. Molina, P.E. *Endocrine physiology*, (McGraw-Hill Education, New York, 2018).

29. Meachem, S.J., Nieschlag, E. & Simoni, M. Inhibin B in male reproduction: pathophysiology and clinical relevance. *Eur. J. Endocrinol.* **145**, 561-571 (2001).

30. Young, B. & Wheater, P.R. *Wheater's functional histology : a text and colour atlas*, (Churchill Livingstone Elsevier, Oxford, 2006).

31. Borg, C.L., Wolski, K.M., Gibbs, G.M. & O'Bryan, M.K. Phenotyping male infertility in the mouse: how to get the most out of a 'non-performer'. *Hum. Reprod. Update* **16**, 205-224 (2010).

32. The Editors of Encyclopaedia Britannica. Spermatogenesis. in *Encyclopædia Britannica* (Encyclopædia Britannica, inc., 2016).

33. Barratt, C.L.R.*, et al.* The diagnosis of male infertility: an analysis of the evidence to support the development of global WHO guidance-challenges and future research opportunities. *Hum. Reprod. Update* **23**, 660-680 (2017).

34. Bieniek, J.M. & Lo, K.C. Recent advances in understanding & managing male infertility. *F1000Res* **5**, 2756 (2016).

35. Gudeloglu, A. & Parekattil, S.J. Update in the evaluation of the azoospermic male. *Clinics (Sao Paulo)* **68 Suppl 1**, 27-34 (2013).

36. Practice Committee of the American Society for Reproductive Medicine in collaboration with the Society for Male, R. & Urology. Evaluation of the azoospermic male: a committee opinion. *Fertil. Steril.* **109**, 777-782 (2018).

37. Jedrzejczak, P., Taszarek-Hauke, G., Hauke, J., Pawelczyk, L. & Duleba, A.J. Prediction of spontaneous conception based on semen parameters. *Int. J. Androl.* **31**, 499-507 (2008).

38. Macleod, J. & Gold, R.Z. The male factor in fertility and infertility. II. Spermatozoon counts in 1000 men of known fertility and in 1000 cases of infertile marriage. *J. Urol.* **66**, 436-449 (1951).

39. World Health Organization. *WHO laboratory manual for the examination and processing of human semen*, (World Health Organization, Geneva, 2010).

40. Cooper, T.G. Semen analysis. in *Andrology: male reproductive health and dysfunction* (eds. Nieschlag, E., Behre, H.M. & Nieschlag, S.) xvii, 629 p. (Springer, Heidelberg, 2010).

41. Alshahrani, S., McGill, J. & Agarwal, A. Prostatitis and male infertility. *J. Reprod. Immunol.* **100**, 30-36 (2013).

42. Roberts, M. & Jarvi, K. Steps in the investigation and management of low semen volume in the infertile man. *Can Urol Assoc J* **3**, 479-485 (2009).

43. Jungwirth, A*., et al.* European Association of Urology guidelines on Male Infertility: the 2012 update. *Eur. Urol.* **62**, 324-332 (2012).

44. Simoni, M. & Nieschlag, E. Endocrine Laboratory Diagnosis. in *Andrology: male reproductive health and dysfunction* (eds. Nieschlag, E., Behre, H.M. & Nieschlag, S.) xvii, 629 p. (Springer, Heidelberg, 2010).

45. McLachlan, R.I. Approach to the patient with oligozoospermia. *J. Clin. Endocrinol. Metab.* **98**, 873-880 (2013).

46. Kolettis, P.N., Purcell, M.L., Parker, W., Poston, T. & Nangia, A.K. Medical testosterone: an iatrogenic cause of male infertility and a growing problem. *Urology* **85**, 1068-1073 (2015).

47. Okutman, O., Rhouma, M.B., Benkhalifa, M., Muller, J. & Viville, S. Genetic evaluation of patients with non-syndromic male infertility. *J. Assist. Reprod. Genet.* (2018).

48. Shah, K., Sivapalan, G., Gibbons, N., Tempest, H. & Griffin, D.K. The genetic basis of infertility. *Reproduction* **126**, 13-25 (2003).

49. Olesen, I.A.*, et al.* Clinical, genetic, biochemical, and testicular biopsy findings among 1,213 men evaluated for infertility. *Fertil. Steril.* **107**, 74-82 e77 (2017).

50. Krausz, C. & Riera-Escamilla, A. Genetics of male infertility. *Nat Rev Urol* **15**, 369-384 (2018).

51. Tuttelmann, F., Ruckert, C. & Ropke, A. Disorders of spermatogenesis: Perspectives for novel genetic diagnostics after 20 years of unchanged routine. *Med Genet* **30**, 12-20 (2018).

52. Yu, J., Chen, Z., Ni, Y. & Li, Z. CFTR mutations in men with congenital bilateral absence of the vas deferens (CBAVD): a systemic review and meta-analysis. *Hum. Reprod.* **27**, 25-35 (2012).

53. Patat, O.*, et al.* Truncating Mutations in the Adhesion G Protein-Coupled Receptor G2 Gene ADGRG2 Cause an X-Linked Congenital Bilateral Absence of Vas Deferens. *Am. J. Hum. Genet.* **99**, 437-442 (2016).

54. Ferlin, A.*, et al.* Male infertility: role of genetic background. *Reprod Biomed Online* **14**, 734-745 (2007).

55. Vogt, P.H.*, et al.* Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet.* **5**, 933-943 (1996).

56. Navarro-Costa, P., Plancha, C.E. & Goncalves, J. Genetic dissection of the AZF regions of the human Y chromosome: thriller or filler for male (in)fertility? *J Biomed Biotechnol* **2010**, 936569 (2010).

57. Krausz, C. & Casamonti, E. Spermatogenic failure and the Y chromosome. *Hum. Genet.* **136**, 637-655 (2017).

58. Massart, A., Lissens, W., Tournaye, H. & Stouffs, K. Genetic causes of spermatogenic failure. *Asian J Androl* **14**, 40-48 (2012).

59. Kamp, C.*, et al.* High deletion frequency of the complete AZFa sequence in men with Sertoli-cell-only syndrome. *Mol. Hum. Reprod.* **7**, 987-994 (2001).

60. Krausz, C., Quintana-Murci, L. & McElreavey, K. Prognostic value of Y deletion analysis: what is the clinical prognostic value of Y chromosome microdeletion analysis? *Hum. Reprod.* **15**, 1431-1434 (2000).

61. Rives, N.*, et al.* Assessment of sex chromosome aneuploidy in sperm nuclei from 47,XXY and 46,XY/47,XXY males: comparison with fertile and infertile males with normal karyotype. *Mol. Hum. Reprod.* **6**, 107-112 (2000).

62. Donker, R.B.*, et al.* Chromosomal abnormalities in 1663 infertile men with azoospermia: the clinical consequences. *Hum. Reprod.* **32**, 2574-2580 (2017).

63. Schiff, J.D.*, et al.* Success of testicular sperm extraction [corrected] and intracytoplasmic sperm injection in men with Klinefelter syndrome. *J. Clin. Endocrinol. Metab.* **90**, 6263-6267 (2005).

64. Fullerton, G., Hamilton, M. & Maheshwari, A. Should non-mosaic Klinefelter syndrome men be labelled as infertile in 2009? *Hum. Reprod.* **25**, 588-597 (2010).

65. Vorona, E., Zitzmann, M., Gromoll, J., Schuring, A.N. & Nieschlag, E. Clinical, endocrinological, and epigenetic features of the 46,XX male syndrome, compared with 47,XXY Klinefelter patients. *J. Clin. Endocrinol. Metab.* **92**, 3458-3465 (2007).

66. de la Chapelle, A. The etiology of maleness in XX men. *Hum. Genet.* **58**, 105-116 (1981).

67. Van Assche, E.*, et al.* Cytogenetics of infertile men. *Hum. Reprod.* **11 Suppl 4**, 1-24; discussion 25-26 (1996).

68. Burgoyne, P.S., Mahadevaiah, S.K. & Turner, J.M. The consequences of asynapsis for mammalian meiosis. *Nat Rev Genet* **10**, 207-216 (2009).

69. Homolka, D., Ivanek, R., Capkova, J., Jansa, P. & Forejt, J. Chromosomal rearrangement interferes with meiotic X chromosome inactivation. *Genome Res.* **17**, 1431-1437 (2007).

70. Vozdova, M.*, et al.* Balanced chromosomal translocations in men: relationships among semen parameters, chromatin integrity, sperm meiotic segregation and aneuploidy. *J. Assist. Reprod. Genet.* **30**, 391-405 (2013).

71. Xie, C., Chen, X., Liu, Y., Wu, Z. & Ping, P. Multicenter study of genetic abnormalities associated with severe oligospermia and non-obstructive azoospermia. *J. Int. Med. Res.* **46**, 107-114 (2018).

72. Brugnon, F.*, et al.* Study of two markers of apoptosis and meiotic segregation in ejaculated sperm of chromosomal translocation carrier patients. *Hum. Reprod.* **21**, 685-693 (2006).

73. Oliver-Bonet, M.*, et al.* From spermatocytes to sperm: meiotic behaviour of human male reciprocal translocations. *Hum. Reprod.* **19**, 2515-2522 (2004).

74. Kavalier, F. Investigation of recurrent miscarriages. *BMJ* **331**, 121-122 (2005).

75. Matzuk, M.M. & Lamb, D.J. Genetic dissection of mammalian fertility pathways. *Nat. Cell Biol.* **4 Suppl**, s41-49 (2002).

76. Fakhro, K.A.*, et al.* Point-of-care whole-exome sequencing of idiopathic male infertility. *Genet. Med.* (2018).

77. Tuttelmann, F.*, et al.* Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PLoS One* **6**, e19426 (2011).

78. Lo Giacco, D.*, et al.* Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility. *J. Med. Genet.* **51**, 340-344 (2014).

79. Mitchell, M.J.*, et al.* Single gene defects leading to sperm quantitative anomalies. *Clin. Genet.* **91**, 208-216 (2017).

80. Nakamura, R.*, et al.* Next-generation sequencing of 28 ALS-related genes in a Japanese ALS cohort. *Neurobiol. Aging* **39**, 219 e211-218 (2016).

81. Robay, A.*, et al.* A systematic review on the genetics of male infertility in the era of next-generation sequencing. *Arab J Urol* **16**, 53-64 (2018).

82. Patel, B.*, et al.* Comprehensive genetic testing for female and male infertility using next-generation sequencing. *J. Assist. Reprod. Genet.* **35**, 1489-1496 (2018).

83. Vendrell, X. New genetic point mutations in male infertility. in *Reproductomics : the -omics revolution and its impact on human reproductive medicine* (eds. Horcajadas, J.A.

& Gosálvez, J.) 47-62 (Elsevier/Academic Press, London, United Kingdom ; San Diego, CA, 2018).

84. Strande, N.T.*, et al.* Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. *Am. J. Hum. Genet.* **100**, 895-906 (2017).

85. Foresta, C., Ferlin, A. & Moro, E. Deletion and expression analysis of AZFa genes on the human Y chromosome revealed a major role for DBY in male infertility. *Hum. Mol. Genet.* **9**, 1161-1169 (2000).

86. Zhu, F.*, et al.* Mutations in PMFBP1 Cause Acephalic Spermatozoa Syndrome. *Am. J. Hum. Genet.* **103**, 188-199 (2018).

87. Gershoni, M.*, et al.* A familial study of azoospermic men identifies three novel causative mutations in three new human azoospermia genes. *Genet. Med.* (2017).

88. Zhu, F.*, et al.* Biallelic SUN5 Mutations Cause Autosomal-Recessive Acephalic Spermatozoa Syndrome. *Am. J. Hum. Genet.* **99**, 1405 (2016).

89. Kasak, L.*, et al.* Bi-allelic Recessive Loss-of-Function Variants in FANCM Cause Non-obstructive Azoospermia. *Am. J. Hum. Genet.* **103**, 200-212 (2018).

90. Gou, L.T.*, et al.* Ubiquitination-Deficient Mutations in Human Piwi Cause Male Infertility by Impairing Histone-to-Protamine Exchange during Spermiogenesis. *Cell* **169**, 1090-1104 e1013 (2017).

91. Yatsenko, A.N.*, et al.* Non-invasive genetic diagnosis of male infertility using spermatozoal RNA: KLHL10 mutations in oligozoospermic patients impair homodimerization. *Hum. Mol. Genet.* **15**, 3411-3419 (2006).

92. Avenarius, M.R.*, et al.* Human male infertility caused by mutations in the CATSPER1 channel protein. *Am. J. Hum. Genet.* **84**, 505-510 (2009).

93. Tenenbaum-Rakover, Y.*, et al.* Minichromosome maintenance complex component 8 (MCM8) gene mutations result in primary gonadal failure. *J. Med. Genet.* **52**, 391-399 (2015).

94. Dirami, T.*, et al.* Missense mutations in SLC26A8, encoding a sperm-specific activator of CFTR, are associated with human asthenozoospermia. *Am. J. Hum. Genet.* **92**, 760-766 (2013).

95. Miyamoto, T.*, et al.* A PLK4 mutation causing azoospermia in a man with Sertoli cell-only syndrome. *Andrology* **4**, 75-81 (2016).

96. Harbuz, R.*, et al.* A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. *Am. J. Hum. Genet.* **88**, 351-361 (2011).

97. Zhu, F., Gong, F., Lin, G. & Lu, G. DPY19L2 gene mutations are a major cause of globozoospermia: identification of three novel point mutations. *Mol. Hum. Reprod.* **19**, 395-404 (2013).

98. Kherraf, Z.E.*, et al.* SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes. *EMBO Mol. Med.* **9**, 1132-1149 (2017).

99. Dam, A.H.*, et al.* Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. *Am. J. Hum. Genet.* **81**, 813-820 (2007).

100. Maor-Sagie, E.*, et al.* Deleterious mutation in SYCE1 is associated with non-obstructive azoospermia. *J. Assist. Reprod. Genet.* **32**, 887-891 (2015).

101. Miyamoto, T.*, et al.* Azoospermia in patients heterozygous for a mutation in SYCP3. *Lancet* **362**, 1714-1719 (2003).

102. Dieterich, K.*, et al.* Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. *Nat. Genet.* **39**, 661-665 (2007).

103. Ayhan, O.*, et al.* Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *J. Med. Genet.* **51**, 239-244 (2014).

104. Tan, Y.Q.*, et al.* Loss-of-function mutations in TDRD7 lead to a rare novel syndrome combining congenital cataract and nonobstructive azoospermia in humans. *Genet. Med.* (2017).

105. Arafat, M.*, et al.* Mutation in TDRD9 causes non-obstructive azoospermia in infertile men. *J. Med. Genet.* **54**, 633-639 (2017).

106. Yang, F.*, et al.* TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. *EMBO Mol. Med.* **7**, 1198-1210 (2015).

107. Yatsenko, A.N.*, et al.* X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *N. Engl. J. Med.* **372**, 2097-2107 (2015).

108. Coutton, C.*, et al.* Bi-allelic Mutations in ARMC2 Lead to Severe Astheno-Teratozoospermia Due to Sperm Flagellum Malformations in Humans and Mice. *Am. J. Hum. Genet.* (2019).

109. Tang, S.*, et al.* Biallelic Mutations in CFAP43 and CFAP44 Cause Male Infertility with Multiple Morphological Abnormalities of the Sperm Flagella. *Am. J. Hum. Genet.* **100**, 854-864 (2017).

110. Colombo, R., Pontoglio, A. & Bini, M. Two Novel TEX15 Mutations in a Family with Nonobstructive Azoospermia. *Gynecol. Obstet. Invest.* (2017).

111. Dong, F.N.*, et al.* Absence of CFAP69 Causes Male Infertility due to Multiple Morphological Abnormalities of the Flagella in Human and Mouse. *Am. J. Hum. Genet.* **102**, 636-648 (2018).

112. Yang, Y.*, et al.* XRCC2 mutation causes meiotic arrest, azoospermia and infertility. *J. Med. Genet.* **55**, 628-636 (2018).

113. Zhang, Y.X.*, et al.* XRCC2 mutation causes premature ovarian insufficiency as well as non-obstructive azoospermia in humans. *Clin. Genet.* (2018).

114. Wang, X.*, et al.* Homozygous DNAH1 frameshift mutation causes multiple morphological anomalies of the sperm flagella in Chinese. *Clin. Genet.* **91**, 313-321 (2017).

115. Ben Khelifa, M.*, et al.* Mutations in DNAH1, which encodes an inner arm heavy chain dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella. *Am. J. Hum. Genet.* **94**, 95-104 (2014).

116. Imtiaz, A.*, et al.* CDC14A phosphatase is essential for hearing and male fertility in mouse and human. *Hum. Mol. Genet.* **27**, 780-798 (2018).

117. Kuo, Y.C.*, et al.* SEPT12 mutations cause male infertility with defective sperm annulus. *Hum. Mutat.* **33**, 710-719 (2012).

118. Lin, Y.H.*, et al.* SEPTIN12 genetic variants confer susceptibility to teratozoospermia. *PLoS One* **7**, e34011 (2012).

119. Escoffier, J.*, et al.* Homozygous mutation of PLCZ1 leads to defective human oocyte activation and infertility that is not rescued by the WW-binding protein PAWP. *Hum. Mol. Genet.* **25**, 878-891 (2016).

120. Ray, P.F.*, et al.* Genetic abnormalities leading to qualitative defects of sperm morphology or function. *Clin. Genet.* **91**, 217-232 (2017).

121. Rybouchkin, A.V., Van der Straeten, F., Quatacker, J., De Sutter, P. & Dhont, M. Fertilization and pregnancy after assisted oocyte activation and intracytoplasmic sperm injection in a case of round-headed sperm associated with deficient oocyte activation capacity. *Fertil. Steril.* **68**, 1144-1147 (1997).

122. Kochhar, P.K. & Ghosh, P. Intracytoplasmic Sperm Injection with Assisted Oocyte Activation Resulting in Successful Pregnancies and Live Birth in Couples with Globozoospermia: A Report of Two Cases. *J. Hum. Reprod. Sci.* **11**, 72-74 (2018).

123. Higgins, A.W.*, et al.* Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am. J. Hum. Genet.* **82**, 712-722 (2008).

124. Schilit, S.L.*, et al.* Estrogen-related receptor gamma implicated in a phenotype including hearing loss and mild developmental delay. *Eur. J. Hum. Genet.* **24**, 1622-1626 (2016).

125. Currall, B.B.*, et al.* Loss of LDAH associated with prostate cancer and hearing loss. *Hum. Mol. Genet.* **27**, 4194-4203 (2018).

126. Wilch, E.S. & Morton, C.C. Historical and Clinical Perspectives on Chromosomal Translocations. in *Chromosome Translocation* (ed. Zhang, Y.) 1-14 (Springer Singapore, Singapore, 2018).

127. Mukherjee, K.*, et al.* Actin capping protein CAPZB regulates cell morphology, differentiation, and neural crest migration in craniofacial morphogenesisdagger. *Hum. Mol. Genet.* **25**, 1255-1270 (2016).

128. Redin, C.*, et al.* The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet* **49**, 36-45 (2017).

129. Lindgren, A.M.*, et al.* Haploinsufficiency of KDM6A is associated with severe psychomotor retardation, global growth restriction, seizures and cleft palate. *Hum. Genet.* **132**, 537-552 (2013).

130. Kim, H.G., *et al.* Translocations disrupting PHF21A in the Potocki-Shaffer-syndrome region are associated with intellectual disability and craniofacial anomalies. *Am. J. Hum. Genet.* **91**, 56-72 (2012).

131. Lachke, S.A., *et al.* The cell adhesion gene PVRL3 is associated with congenital ocular defects. *Hum. Genet.* **131**, 235-250 (2012).

132. Brown, K.K., *et al.* Deletion of an enhancer near DLX5 and DLX6 in a family with hearing loss, craniofacial defects, and an inv(7)(q21.3q35). *Hum. Genet.* **127**, 19-31 (2010).

133. Talkowski, M.E., *et al.* Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am. J. Hum. Genet.* **88**, 469-481 (2011).

134. Hanscom, C. & Talkowski, M. Design of large-insert jumping libraries for structural variant detection using illumina sequencing. *Curr Protoc Hum Genet* **80**, 7 22 21-29 (2014).

135. Ordulu, Z., *et al.* Describing sequencing results of structural chromosome rearrangements with a suggested next-generation cytogenetic nomenclature. *Am J Hum Genet* **94**, 695-709 (2014).

136. Dixon, J.R., *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).

137. Rosenbloom, K.R., *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670-681 (2015).

138. Sherry, S.T., *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308-311 (2001).

139. Schilit, S.L.P. & Morton, C.C. 3C-PCR: a novel proximity ligation-based approach to phase chromosomal rearrangement breakpoints with distal allelic variants. *Hum. Genet.* **137**, 55-62 (2018).

140. Miele, A., Gheldof, N., Tabuchi, T.M., Dostie, J. & Dekker, J. Mapping chromatin interactions by chromosome conformation capture. *Curr. Protoc. Mol. Biol.* **Chapter 21**, Unit 21 11 (2006).

141. van de Werken, H.J., *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89-112 (2012).

142. Splinter, E., de Wit, E., van de Werken, H.J., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58**, 221-230 (2012).

143. Gheldof, N., Leleu, M., Noordermeer, D., Rougemont, J. & Reymond, A. Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method. *Methods Mol. Biol.* **786**, 211-225 (2012).

144. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-1311 (2002).

145. Afgan, E.*, et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537-W544 (2018).

146. Klein, F.A.*, et al.* FourCSeq: analysis of 4C sequencing data. *Bioinformatics* **31**, 3085-3091 (2015).

147. Krzywinski, M.*, et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639-1645 (2009).

148. Jiang, S.*, et al.* CRISPR/Cas9-Mediated Genome Editing in Epstein-Barr Virus-Transformed Lymphoblastoid B-Cell Lines. *Curr. Protoc. Mol. Biol.* **121**, 31 12 31-31 12 23 (2018).

149. Hu, V.W.*, et al.* Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. *Autism Res.* **2**, 78-97 (2009).

150. Lupiáñez, D.G.*, et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).

151. Ordulu, Z.*, et al.* Structural chromosomal rearrangements require nucleotide-level resolution: lessons from next-generation sequencing in prenatal diagnosis. *Am. J. Hum. Genet.* **99**, 1015-1033 (2016).

152. Spielmann, M., Lupianez, D.G. & Mundlos, S. Structural variation in the 3D genome. *Nat Rev Genet* **19**, 453-467 (2018).

153. Matharu, N. & Ahituv, N. Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. *PLoS Genet* **11**, e1005640 (2015).

154. Lupiáñez, D.G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet.* **32**, 225-237 (2016).

155. Andrews, G.L. & Mastick, G.S. R-cadherin is a Pax6-regulated, growth-promoting cue for pioneer axons. *J. Neurosci.* **23**, 9873-9880 (2003).

156. Babb, S.G.*, et al.* Zebrafish R-cadherin (Cdh4) controls visual system development and differentiation. *Dev. Dyn.* **233**, 930-945 (2005).

157. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580-585 (2013).

158. Huang, N., Lee, I., Marcotte, E.M. & Hurles, M.E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* **6**, e1001154 (2010).

159. Firth, H.V.*, et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524-533 (2009).

160. Kim, H.G.*, et al.* Hypogonadotropic hypogonadism and cleft lip and palate caused by a balanced translocation producing haploinsufficiency for FGFR1. *J Med Genet* **42**, 666-672 (2005).

161. Lu, W.*, et al.* Disruption of ROBO2 is associated with urinary tract anomalies and confers risk of vesicoureteral reflux. *Am J Hum Genet* **80**, 616-632 (2007).

162. Kim, H.G.*, et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 199-207 (2008).

163. Selvaraj, S., J, R.D., Bansal, V. & Ren, B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat. Biotechnol.* **31**, 1111-1118 (2013).

164. Lettice, L.A.*, et al.* Enhancer-adoption as a mechanism of human developmental disease. *Hum. Mutat.* **32**, 1492-1499 (2011).

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

166. Rosenbloom, K.R.*, et al.* ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* **41**, D56-63 (2013).

167. Kent, W.J., *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996-1006 (2002).

168. Stankiewicz, P. & Lupski, J.R. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61**, 437-455 (2010).

169. Kleinjan, D.A. & van Heyningen, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76**, 8-32 (2005).

170. Zepeda-Mendoza, C.J., *et al.* Computational prediction of position effects of apparently balanced human chromosomal rearrangements. *Am. J. Hum. Genet.* (2017).

171. Duzkale, H., *et al.* A systematic approach to assessing the clinical significance of genetic variants. *Clin. Genet.* **84**, 453-463 (2013).

172. Browning, S.R. & Browning, B.L. Haplotype phasing: existing methods and new developments. *Nat Rev Genet* **12**, 703-714 (2011).

173. de Vree, P.J., *et al.* Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat. Biotechnol.* **32**, 1019-1025 (2014).

174. Regan, J.F., *et al.* A rapid molecular approach for chromosomal phasing. *PLoS One* **10**, e0118270 (2015).

175. McDonald, O.G., Krynetski, E.Y. & Evans, W.E. Molecular haplotyping of genomic DNA for multiple single-nucleotide polymorphisms located kilobases apart using long-range polymerase chain reaction and intramolecular ligation. *Pharmacogenetics* **12**, 93-99 (2002).

176. Ma, L., *et al.* Direct determination of molecular haplotypes by chromosome microdissection. *Nat Methods* **7**, 299-301 (2010).

177. Fan, H.C., Wang, J., Potanina, A. & Quake, S.R. Whole-genome molecular haplotyping of single cells. *Nat. Biotechnol.* **29**, 51-57 (2011).

178. Untergasser, A., *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**, W71-74 (2007).

179. Kent, W.J. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656-664 (2002).

180. Denker, A. & de Laat, W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* **30**, 1357-1382 (2016).

181. Dekker, J. The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods* **3**, 17-21 (2006).

182. Snyder, M.W., Adey, A., Kitzman, J.O. & Shendure, J. Haplotype-resolved genome sequencing: experimental methods and applications. *Nat Rev Genet* **16**, 344-358 (2015).

183. Mills, R.E.*, et al.* Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res.* **21**, 830-839 (2011).

184. Kaper, F.*, et al.* Whole-genome haplotyping by dilution, amplification, and sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5552-5557 (2013).

185. Schalk, J.A.*, et al.* Isolation and characterization of the human SCP2 cDNA and chromosomal localization of the gene. *Mamm. Genome* **10**, 642-644 (1999).

186. Yang, F.*, et al.* Mouse SYCP2 is required for synaptonemal complex assembly and chromosomal synapsis during male meiosis. *J. Cell Biol.* **173**, 497-507 (2006).

187. Offenberg, H.H.*, et al.* SCP2: a major protein component of the axial elements of synaptonemal complexes of the rat. *Nucleic Acids Res.* **26**, 2572-2579 (1998).

188. Winkel, K., Alsheimer, M., Ollinger, R. & Benavente, R. Protein SYCP2 provides a link between transverse filaments and lateral elements of mammalian synaptonemal complexes. *Chromosoma* **118**, 259-267 (2009).

189. Kolas, N.K.*, et al.* Male mouse meiotic chromosome cores deficient in structural proteins SYCP3 and SYCP2 align by homology but fail to synapse and have possible impaired specificity of chromatin loop attachment. *Cytogenet Genome Res* **105**, 182-188 (2004).

190. Tsai, J.H. & McKee, B.D. Homologous pairing and the role of pairing centers in meiosis. *J. Cell Sci.* **124**, 1955-1963 (2011).

191. Rockmill, B. & Roeder, G.S. Telomere-mediated chromosome pairing during meiosis in budding yeast. *Genes Dev.* **12**, 2574-2586 (1998).

192. Longtine, M.S.*, et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**, 953-961 (1998).

193. MacQueen, A.J. & Rockmill, B. Analysis of Meiotic Chromosome-Associated Protein Dynamics Using Conditional Expression in Budding Yeast. *Methods Mol. Biol.* **1471**, 157-174 (2017).

194. Voelkel-Meiman, K., Moustafa, S.S., Lefrancois, P., Villeneuve, A.M. & MacQueen, A.J. Full-length synaptonemal complex grows continuously during meiotic prophase in budding yeast. *PLoS Genet* **8**, e1002993 (2012).

195. Smith, A.V. & Roeder, G.S. The yeast Red1 protein localizes to the cores of meiotic chromosomes. *J. Cell Biol.* **136**, 957-967 (1997).

196. Sym, M., Engebrecht, J.A. & Roeder, G.S. ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. *Cell* **72**, 365-378 (1993).

197. Ropke, A.*, et al.* Comprehensive sequence analysis of the NR5A1 gene encoding steroidogenic factor 1 in a large group of infertile males. *Eur. J. Hum. Genet.* **21**, 1012-1015 (2013).

198. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).

199. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

200. McKenna, A.*, et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303 (2010).

201. McLaren, W.*, et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

202. Tewes, A.C., Ledig, S., Tuttelmann, F., Kliesch, S. & Wieacker, P. DMRT1 mutations are rarely associated with male infertility. *Fertil. Steril.* **102**, 816-820 e813 (2014).

203. Brinkworth, M.H., Weinbauer, G.F., Schlatt, S. & Nieschlag, E. Identification of male germ cells undergoing apoptosis in adult rats. *J. Reprod. Fertil.* **105**, 25-33 (1995).

204. Kerr, G.W., Sarkar, S. & Arumugam, P. How to halve ploidy: lessons from budding yeast meiosis. *Cell. Mol. Life Sci.* **69**, 3037-3051 (2012).

205. Lek, M.*, et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).

206. Simm, D., Hatje, K. & Kollmar, M. Waggawagga: comparative visualization of coiled-coil predictions and detection of stable single alpha-helices (SAH domains). *Bioinformatics* **31**, 767-769 (2015).

207. Zhang, H.G.*, et al.* Male carriers of balanced reciprocal translocations in Northeast China: sperm count, reproductive performance, and genetic counseling. *Genet. Mol. Res.* **14**, 18792-18798 (2015).

208. Vogt, P. *Genetics of human infertility*, (Karger, Basel ; New York, 2017).

209. Zickler, D. & Kleckner, N. Meiotic chromosomes: integrating structure and function. *Annu. Rev. Genet.* **33**, 603-754 (1999).

210. Sym, M. & Roeder, G.S. Zip1-induced changes in synaptonemal complex structure and polycomplex assembly. *J. Cell Biol.* **128**, 455-466 (1995).

211. Ollinger, R., Alsheimer, M. & Benavente, R. Mammalian protein SCP1 forms synaptonemal complex-like structures in the absence of meiotic chromosomes. *Mol. Biol. Cell* **16**, 212-217 (2005).

212. Jeffress, J.K.*, et al.* The formation of the central element of the synaptonemal complex may occur by multiple mechanisms: the roles of the N- and C-terminal domains of the Drosophila C(3)G protein in mediating synapsis and recombination. *Genetics* **177**, 2445-2456 (2007).

213. Yuan, L.*, et al.* The synaptonemal complex protein SCP3 can form multistranded, cross-striated fibers in vivo. *J. Cell Biol.* **142**, 331-339 (1998).

214. Rockmill, B. & Roeder, G.S. Meiosis in asynaptic yeast. *Genetics* **126**, 563-574 (1990).

215. Champion, M.D. & Hawley, R.S. Playing for half the deck: the molecular biology of meiosis. *Nat. Cell Biol.* **4 Suppl**, s50-56 (2002).

216. Morelli, M.A. & Cohen, P.E. Not all germ cells are created equal: aspects of sexual dimorphism in mammalian meiosis. *Reproduction* **130**, 761-781 (2005).

217. Yuan, L.*, et al.* The murine SCP3 gene is required for synaptonemal complex assembly, chromosome synapsis, and male fertility. *Mol. Cell* **5**, 73-83 (2000).

218. Yang, F.*, et al.* Meiotic failure in male mice lacking an X-linked factor. *Genes Dev.* **22**, 682-691 (2008).

219. Ohtsuka, M.*, et al.* One-step generation of multiple transgenic mouse lines using an improved Pronuclear Injection-based Targeted Transgenesis (i-PITT). *BMC Genomics* **16**, 274 (2015).

220. Schilit, S.L., Ohtsuka, M., Quadros, R.M. & Gurumurthy, C.B. Pronuclear Injection-Based Targeted Transgenesis. *Curr Protoc Hum Genet* **91**, 15 10 11-15 10 28 (2016).

221. Tasic, B.*, et al.* Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proc Natl Acad Sci U S A* **108**, 7902-7907 (2011).

222. Burkhard, P., Stetefeld, J. & Strelkov, S.V. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* **11**, 82-88 (2001).

APPENDIX A

European Journal of Human Genetics (2016) 24, 1622–1626

www.nature.com/ejhg

SHORT REPORT

Estrogen-related receptor gamma implicated in a phenotype including hearing loss and mild developmental delay

Samantha LP Schilit[1,15], Benjamin B Currall[2,3,15], Ruen Yao[4], Carrie Hanscom[5], Ryan L Collins[5], Vamsee Pillalamarri[5], Dong-Young Lee[6], Tammy Kammin[2], Cinthya J Zepeda-Mendoza[2,3], Tarja Mononen[7], Lisa S Nolan[8], James F Gusella[1,3,5,9,10], Michael E Talkowski[3,5,9,10,11,12], Jun Shen[3,6,13] and Cynthia C Morton*[,2,3,6,9,14]

Analysis of chromosomal rearrangements has been highly successful in identifying genes involved in many congenital abnormalities including hearing loss. Herein, we report a subject, designated DGAP242, with congenital hearing loss (HL) and a *de novo* balanced translocation 46,XX,t(1;5)(q32;q15)dn. Using multiple next-generation sequencing techniques, we obtained high resolution of the breakpoints. This revealed disruption of the orphan receptor *ESRRG* on chromosome 1, which is differentially expressed in inner ear hair cells and has previously been implicated in HL, and disruption of *KIAA0825* on chromosome 5. Given the translocation breakpoints and supporting literature, disruption of *ESRRG* is the most likely cause for DGAP242's phenotype and implicates *ESRRG* in a monogenic form of congenital HL, although a putative contributory role for *KIAA0825* in the subject's disorder cannot be excluded.
European Journal of Human Genetics (2016) **24,** 1622–1626; doi:10.1038/ejhg.2016.64; published online 6 July 2016

INTRODUCTION

Hearing loss (HL) is one of the most common birth defects, affecting over one in a thousand newborns. Over half of these cases can be attributed to a genetic etiology.[1] Chromosomal abnormalities have been implicated in several genetic forms of congenital HL including Down, Smith-Magenis and Branchio-Oto-Renal syndromes.[2] In cases of congenital HL accompanied with balanced chromosomal abnormalities (BCAs), high resolution of chromosomal breakpoints can reveal genes that have been disrupted, implicating them in auditory pathology.[3] By employing this strategy, the Developmental Genome Anatomy Project (DGAP, dgap.harvard.edu) has successfully resolved the genetic etiology of HL in several cases of BCAs.[4–6] Herein, we report the high-resolution breakpoints from a *de novo* translocation discovered in a 6-year-old female with congenital bilateral sensorineural hearing loss (SNHL) accompanied with a developmental disorder of speech and language.

MATERIALS AND METHODS

The subject, referred to as DGAP242, was recruited into DGAP after identification of a BCA and congenital HL. Informed consent, medical records and blood samples were obtained through the DGAP protocol approved by the Partners HealthCare Systems Institutional Review Board.

Clinical report

DGAP242 was delivered at 40 weeks gestation with a birth weight of 3.36 kg (50th percentile) to a 40-year-old G3P3 mother after elective cesarean section. The pregnancy was remarkable for gestational diabetes, hypothyroidism and high blood pressure. There were no prenatal risk factors for HL, including no known family history of congenital or childhood HL. The paternal grandfather started wearing hearing aids after age 60 years with hearing impairment attributed to age-related hearing loss (ARHL) and occupational noise exposure (Figure 1a).

DGAP242 did not pass her newborn hearing screening and was referred for diagnostic audiological evaluation. Absent otoacoustic emissions, automated auditory brainstem response and auditory steady state response measurements indicated congenital, bilateral SNHL. *GJB2* testing was negative for pathogenic variants. DGAP242 demonstrated hypotonia and motor delay, standing at 16 months and walking at 20 months of age. At this time, chromosome analysis and array-based comparative genomic hybridization of the proband and her parents revealed a *de novo* apparently balanced translocation between chromosomes 1 and 5: 46,XX,t(1;5)(q32;q15)dn.

Upon enrollment in DGAP 6 months later, several features were noted. DGAP242 had delayed speech, slightly posteriorly rotated ears and mild epicanthal folds. Play audiometry indicated mild-to-moderate high-frequency SNHL (Figure 1b, top). She started receiving speech therapy at age three years, and started wearing hearing aids regularly at age four. In a follow-up at 6 years of age, DGAP242 had bilateral, symmetrical, moderate SNHL and receptive language disorder (Figure 1b, bottom) and significant problems in listening comprehension and in understanding complex sentence structures.

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA; [2]Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA, USA; [3]Harvard Medical School, Boston, MA, USA; [4]Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, Shanghai, China; [5]Molecular Neurogenetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; [6]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; [7]Department of Clinical Genetics, Kuopio University Hospital, Kuopio, Finland; [8]UCL Ear Institute, University College London, London, UK; [9]Medical and Population Genetics Program, Broad Institute, Cambridge, MA, USA; [10]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; [11]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; [12]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA; [13]Laboratory for Molecular Medicine, Partners Personalized Medicine, Partners HealthCare, Cambridge, MA, USA; [14]Manchester Academic Health Science Centre, University of Manchester, Manchester, UK
*Correspondence: Dr CC Morton, Department of Pathology, Brigham and Women's Hospital, New Research Building, Rm. 160D, 77 Avenue Louis Pasteur, Boston, MA 02115, USA. Tel: +1 617 525 4535; Fax: +1 617 525 4533; E-mail: cmorton@partners.org
[15]These authors contributed equally to this work.
Received 29 January 2016; revised 5 May 2016; accepted 21 May 2016; published online 6 July 2016

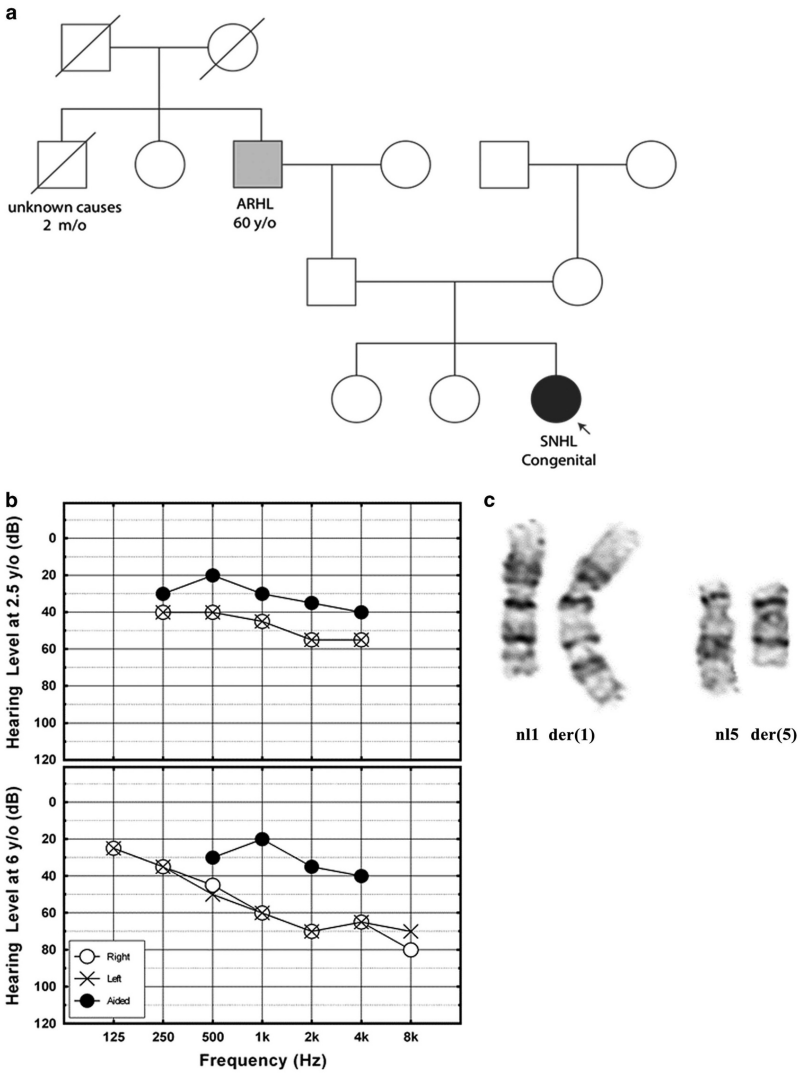


Figure 1 (**a**) DGAP242 pedigree (SNHL, sensorineural hearing loss; ARHL, age-related hearing loss). (**b**) Audiometric evaluation of both ears simultaneously, using play audiometry in sound field, shows downward sloping mild-to-moderate hearing loss in DGAP242 at age 2.5 years (top). Audiometric evaluation by pure tone audiometry shows bilateral, symmetrical and moderate SNHL in DGAP242 at age six years. (**c**) DGAP242 partial karyotype of normal (left) and derivative (right) GTG-banded chromosomes 1 and 5.

Additional observations included difficulty with abstract visual reasoning and balance tasks, but normal visual acuity and appropriate gross and fine motor skills (Table 1). DGAP242's clinical description and refined translocation breakpoints are deposited in DECIPHER (decipher.sanger.ac.uk/patient/317883).[7]

Lymphoblastoid cell line

Epstein–Barr virus transformation of lymphocytes, library generation, sequencing and bioinformatic analysis were performed at the Genomics and Technology Core in the Center for Human Genetic Research at Massachusetts General Hospital (Boston, MA, USA). Confirmatory GTG-banded karyotyping was performed on the Epstein–Barr virus-transformed lymphoblastoid cell line at the Brigham and Women's Hospital CytoGenomics Core Laboratory (Boston, MA, USA).

Whole-genome jumping library

Large-insert ('jumping library') whole-genome sequencing was performed as previously described.[8,9] In brief, genomic DNA was sheared to ~3 kbp, end-ligated with *EcoP*15I restriction adapters, size-selected and then circularized about an internal biotinylated adapter. Circular DNAs were *EcoP*15I restriction digested and purified using internal adapter pull-down, resulting in short DNA

fragments whose ends reflect ~3 kbp of physical separation in the human reference genome. Fragments were ligated to Illumina TruSeq multiplexing adapters and amplified by PCR. The jumping library was used for next-generation sequencing (NGS) on an Illumina HiSeq 2000 (Illumina, Inc., San Diego, CA, USA) using PE25 chemistry (2 × 25 bp reads). Data processing and analysis were performed as previously described (Figure 2a).[10,11]

Table 1 HPO-coded phenotypes for DGAP242

Phenotype	*Code*
Bilateral sensorineural hearing impairment	HP:0008619
Congenital sensorineural hearing impairment	HP:0008527
Infantile muscular hypotonia	HP:0008947
Motor delay	HP:0001270
Expressive language delay	HP:0002474
Poor speech	HP:0002465
Posteriorly rotated ears	HP:0000358
Epicanthus	HP:0000286
Receptive language delay	HP:0010863

Targeted locus amplification library

A targeted locus amplification library was prepared as previously described.[12] In brief, genomic DNA was crosslinked, enzymatically digested and then re-ligated to create circularized DNA fragments. After de-crosslinking, breakpoint regions were amplified using PCR with anchored primers designed to target regions upstream and downstream of the presumed breakpoints. PCR products were sheared by sonication, and then barcode adapters were added. After amplification, the library was used for NGS on an Ion Torrent PGM with a 316v2 chip (Figure 2b).

RESULTS

Fine-resolution mapping of DGAP242's translocation breakpoints was performed using sequential high-resolution methods. First, metaphase chromosome analysis confirmed the band-level resolution of the clinical karyotype indicating a balanced translocation, t(1;5)(q32;q15)dn (Figure 1c). Large-insert whole-genome sequencing was then performed substantiating the reciprocal translocation of chromosomes 1 and 5, and resolving the breakpoints to a 965 bp window on chr1 (216 952 149 to 216 953 114) and a 305 bp window on chr5 (93 556 359 to 93 556 664; Figure 2a).[8,9] Breakpoints of the translocation resulted in a revision of the karyotype to t(1;5)(q41;q15)dn. The chr1 breakpoint region includes

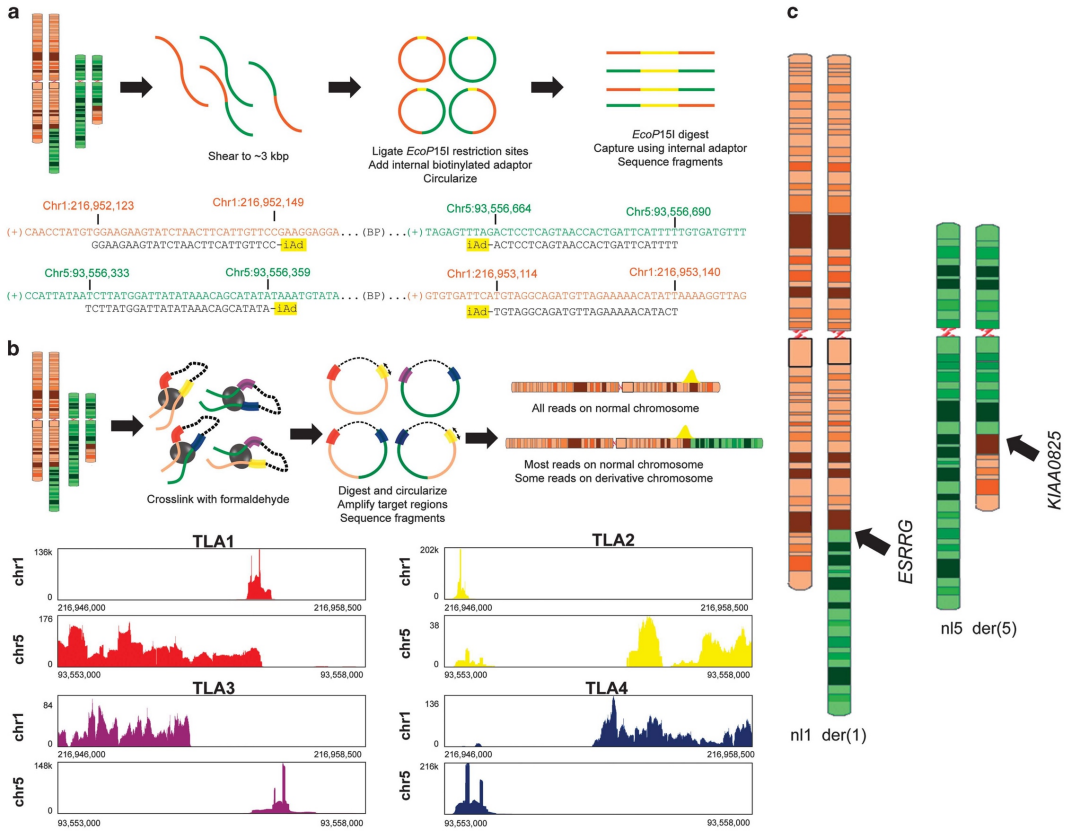


Figure 2 (**a**) Diagram of jumping library methods and nearest aligned sequences adjacent to the translocation breakpoints (iAd, internal adapter for circularization). (**b**) Diagram of targeted locus amplification (TLA) library and histogram of nearest aligned reads adjacent to breakpoints. (**c**) Ideograms illustrate a simple q-q translocation disrupting *ESRRG* on chr1 and *KIAA0825* on chr5.

a large ~500 bp AT-rich region that precluded identifying the precise junction to nucleotide resolution using standard polymerase-based amplification and Sanger sequencing.

In an effort to improve resolution of the breakpoints, a secondary NGS approach was pursued by constructing a targeted locus amplification library. Using four primer sets, regions were targeted either upstream or downstream of the putative breakpoints discovered in the jumping library (Supplementary Table 1). All primer sets had reads that aligned to both the targeted chromosome and to an untargeted chromosome (Figure 2b). These reads confirmed the t(1;5), refining the breakpoint region further to an 816 bp window on chr1 (216 952 298 to 216 953 114) and a 153 bp window on chr5 (93 556 359 to 93 556 512).

On the basis of these orthogonal NGS approaches, we mapped the breakpoints to sub-kilobase resolution. The karyotypic description of the chromosome rearrangement is given according to Next-Generation Cytogenetic Nomenclature[13] and newly proposed recommendations by the Human Genome Variation Society (www.hgvs.org/mutnomen/comments004.html):

Next Generation Cytogenetic Nomenclature (NGCN)
46,XX,t(1;5)(q32;q15)dn.seq[GRCh37/hg19]
t(1;5)(1pter->1q41(216,952,298~216,953,114)::
5q15(93,556,359~93,556,512)->5qter;
5pter->5q15(93,556,359~93,556,512)::
1q41(216,952,298~216,953,114)->1qter)dn

Human Genome Variation Society (HGVS)
46,XX,t(1;5)(q32;q15)dn.seq[GRCh37] t(1;5)(q41;q15)dn
g.[chr1:pter_cen_(216952298_216953114)::
chr5:(93556359_93556512)_qter]
g.[chr5:pter_cen_(93556359_93556512)::
chr1:(216952298_216953114)_qter]

The t(1;5) disrupts two genes, *ESRRG* and *KIAA0825*, suggesting a single copy loss for both genes (Figure 2c). Both genes have low haploinsufficency scores according to DECIPHER (*KIAA0825*, %HI=7.51; *ESRRG*, %HI=0.51), further arguing that these disruptions could be pathogenic.[14] Of the two genes, only *ESRRG* is significantly enriched in cochlear hair cells according to inner ear expression data from the Shared Harvard Inner Ear Laboratory Database (SHIELD, shield.hms.harvard.edu; fold enrichment=4.55, FDR=8.20E−03).[15] This indicates that loss of *ESRRG* is more likely to have a HL phenotype than *KIAA0825* (Supplementary Table 2). Sanger sequencing of coding regions and flanking exon–intron boundaries of *ESRRG* did not identify any additional pathogenic variant, further supporting a haploinsufficiency mechanism of pathogenicity. In addition to disrupted genes, potentially dysregulated genes within the topologically associating domains containing the breakpoints were assessed, due to possible changes in regional regulators of transcription (Supplementary Figure 1).[16]

DISCUSSION

In individuals with congenital anomalies and a BCA with normal array-based comparative genomic hybridization results, investigation of the chromosomal breakpoints frequently elucidates causal genes that have been disrupted.[3] Directly disrupted genes can be associated with a haploinsufficiency mechanism of pathogenicity, as one allele is lost due to interruption of the coding sequence. In this report, we analyzed genes disrupted by an apparently balanced *de novo* chromosomal translocation in a subject with congenital SNHL.

High-resolution analysis of the t(1;5) in DGAP242 illuminated the possible genetic etiology in two predicted haploinsufficient genes: *KIAA0825* and *ESRRG*. Although little is known about *KIAA0825*, which limits assessment of its contribution to DGAP242's phenotype, its lack of enrichment in inner ear cells does not support its involvement in DGAP242's HL. However, disruption of *ESRRG*, which is preferentially expressed in inner ear hair cells, could explain DGAP242's HL. *ESRRG* encodes estrogen-related receptor gamma (ERRγ), a dimeric orphan nuclear transcription factor that is a paralog to ERRβ, a known HL gene.[17–19] In addition, *ESRRG* has been shown by a candidate gene association study to be associated with age-related HL in females, a finding that has been supported by hair cell-specific expression of *Esrrg* in the adult mouse inner ear, and significantly elevated hearing thresholds in $Esrrg^{-/-}$ mice relative to their wild-type littermates.[20] These data suggest that disruption of *ESRRG* may reduce its expression in hair cells and result in congenital HL in humans.

ESRRG is hypothesized to have a role in DGAP242's congenital, bilateral SNHL: *ESRRG* is predicted to be haploinsufficient, is differentially expressed in the inner ear, and has been formerly associated with HL. To our knowledge, this is the first report of a heterozygous disruption of *ESRRG* implicated in human HL. This study supports the candidate gene association study and $Esrrg^{-/-}$ mouse studies that previously implicated *ESRRG* in age-related hearing loss.[20] Given these findings, evaluation of *ESRRG* in other cases of unexplained congenital bilateral SNHL may be warranted.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the patient and her family for their participation in our study. This research was supported by the National Institute of General Medical Sciences GM061354 (CCM, JFG), the National Institute of Mental Health MH095867 (MET), the National Institute on Deafness and Other Communication Disorders R03DC013866 (JS) and F32DC012466 (BBC) and the National Science Foundation DGE1144152 (SLPS). Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding institutions. This study makes use of data generated by the DECIPHER community. A full list of centers that contributed to the generation of the DECIPHER data is available from http://decipher.sanger.ac.uk and via E-mail from decipher@sanger.ac.uk. Funding for DECIPHER was provided by the Wellcome Trust.

1 Morton CC, Nance WE: Newborn hearing screening–a silent revolution. *N Engl J Med* 2006; **354**: 2151–2164.
2 Giersch ABS, Morton CC: Cytogenetics and cochlear expressed sequence tags (ESTs) for identification of genes involved in hearing and deafness. In: Keats BJB, Popper AN, Fay RR (eds): *Genetics and Auditory Disorders*. New York: Springer, 2002, ppxii, 322 p.
3 Higgins AW, Alkuraya FS, Bosco AF *et al*: Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am J Hum Genet* 2008; **82**: 712–722.
4 Anger GJ, Crocker S, McKenzie K *et al*: X-linked deafness-2 (DFNX2) phenotype associated with a paracentric inversion upstream of POU3F4. *Am J Audiol* 2014; **23**: 1–6.
5 Brown KK, Reiss JA, Crow K *et al*: Deletion of an enhancer near DLX5 and DLX6 in a family with hearing loss, craniofacial defects, and an inv(7)(q21.3q35). *Hum Genet* 2010; **127**: 19–31.
6 Williamson RE, Darrow KN, Michaud S *et al*: Methylthioadenosine phosphorylase (MTAP) in hearing: gene disruption by chromosomal rearrangement in a hearing impaired individual and model organism analysis. *Am J Med Genet A* 2007; **143 A**: 1630–1639.
7 Kohler S, Doelken SC, Mungall CJ *et al*: The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res* 2014; **42**: D966–D974.
8 Hanscom C, Talkowski M: Design of large-insert jumping libraries for structural variant detection using illumina sequencing. *Curr Protoc Hum Genet* 2014; **80**: 7 22 21–29.
9 Talkowski ME, Ernst C, Heilbut A *et al*: Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am J Human Genet* 2011; **88**: 469–481.

10 Chiang C, Jacobsen JC, Ernst C *et al*: Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat Genet* 2012; **44**: 390–397.
11 Brand H, Collins RL, Hanscom C *et al*: Paired-duplication signatures mark cryptic inversions and other complex structural variation. *Am J Hum Genetics* 2015; **97**: 170–176.
12 de Vree PJ, de Wit E, Yilmaz M *et al*: Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat Biotechnol* 2014; **32**: 1019–1025.
13 Ordulu Z, Wong KE, Currall BB *et al*: Describing sequencing results of structural chromosome rearrangements with a suggested next-generation cytogenetic nomenclature. *Am J Hum Genet* 2014; **94**: 695–709.
14 Huang N, Lee I, Marcotte EM, Hurles ME: Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 2010; **6**: e1001154.
15 Scheffer DI, Shen J, Corey DP, Chen ZY: Gene expression by mouse inner ear hair cells during development. *J Neurosci* 2015; **35**: 6366–6380.
16 Lupianez DG, Kraft K, Heinrich V *et al*: Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015; **161**: 1012–1025.
17 Collin RW, Kalay E, Tariq M *et al*: Mutations of ESRRB encoding estrogen-related receptor beta cause autosomal-recessive nonsyndromic hearing impairment DFNB35. *Am J Hum Genet* 2008; **82**: 125–138.
18 Hong H, Yang L, Stallcup MR: Hormone-independent transcriptional activation and coactivator binding by novel orphan nuclear receptor ERR3. *J Biol Chem* 1999; **274**: 22618–22626.
19 Horard B, Vanacker JM: Estrogen receptor-related receptors: orphan receptors desperately seeking a ligand. *J Mol Endocrinol* 2003; **31**: 349–357.
20 Nolan LS, Maier H, Hermans-Borgmeyer I *et al*: Estrogen-related receptor gamma and hearing function: evidence of a role in humans and mice. *Neurobiol Aging* 2013; **34**: e2071–e2079.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)

SUPPLEMENTARY INFORMATION

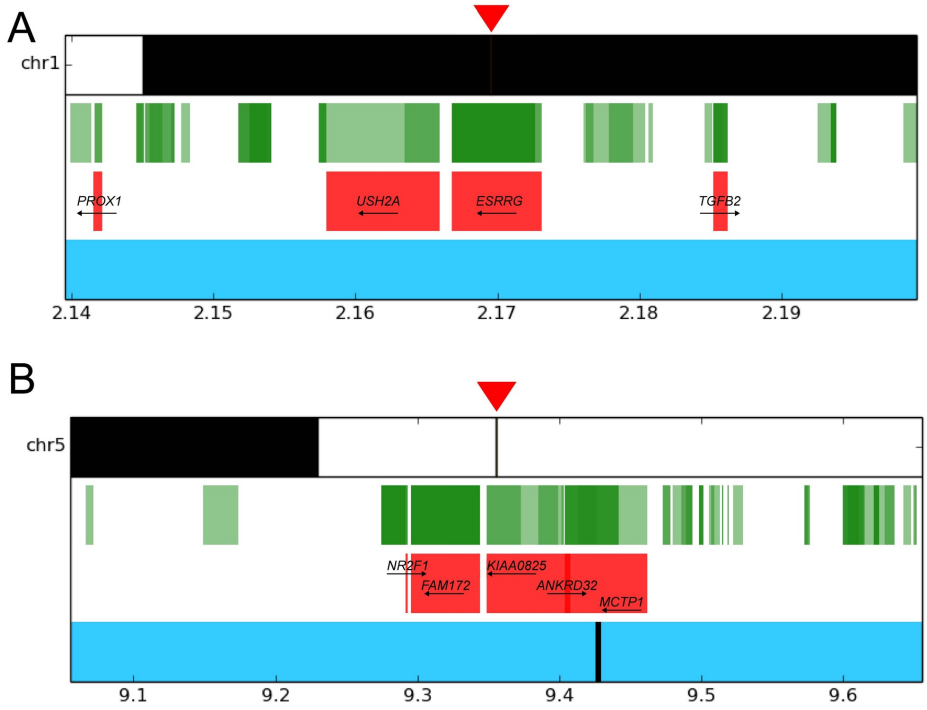
SUPPLEMENTARY METHODS

Topological Domains and Genes Predicted to Show Haploinsufficiency Bordering DGAP242 Translocation Breakpoints

Six Mb windows surrounding the chromosomal breakpoints were analyzed for topologically associating domains (TADs), gene annotation, predicted haploinsufficiency (HI), OMIM-associated phenotypes and inner ear expression. TADs were predicted using human embryonic stem cell Hi-C domains from the Hi-C project (chromosome.sdsc.edu) and converted to hg19 for comparison to the breakpoints.[1] Gene annotation was obtained from the University of California Santa Cruz Genome Browser (genome.ucsc.edu).[2] If the breakpoint occurred within the open reading frame of any gene, that gene was considered disrupted. If a gene resided within the TAD, it was evaluated for potential dysregulation. To assess pathogenicity, HI prediction was downloaded from DECIPHER (decipher.sanger.ac.uk).[3] Only disrupted or dysregulated genes with HI scores of $\leq$10% were considered as possibly pathogenic. Inner ear cell-type specific gene expression was determined using the Shared Harvard Inner Ear Laboratory Database (SHIELD) (shield.hms.harvard.edu).[4,5] Possible effects of disrupted or dysregulated pathogenic genes were determined by assessment of OMIM phenotypes from the Online Mendelian Inheritance in Man Catalog of Human Genes and Genetic Disorders (www.omim.org) and mapping of phenotypes to genes implicated within TAD regions.[6]

SUPPLEMENTARY FIGURES

Supplementary Figure 1. Topologically associating domains (TADs) disrupted by DGAP242's chromosomal translocation and genes predicted to show haploinsufficiency (HI). The top row is the chromosome section, containing the banding patterns of the 6 Mb chromosome region surrounding the breakpoint (red arrowhead). The second row includes genes present in that section of the chromosome (green rectangles). The third row corresponds to predicted HI genes (red rectangles). The fourth row includes the topological domain data, with blue rectangles representing Hi-C domains, and black rectangles representing domain boundary regions, or regions where no contacts were detected. Units are in bp x 10^7. Panel A: Chromosome 1 (chr1) TAD and HI presence. Predicted HI genes are *PROX1*, *USH2A* (OMIM #276901), *ESRRG* and *TGFB2* (OMIM #614816), which are located ~2.7, ~1.1, 0, and ~1.5 Mb from the translocation breakpoint, respectively. Note that *ESRRG* is directly disrupted by the translocation. Panel B: Chromosome 5 (chr5) TAD and HI presence. Predicted HI genes are *NR2F1* (OMIM #615722), *FAM172A*, *KIAA0825*, *ANKRD32* and *MCTP1*, which are located ~625, ~108, 0, ~519, and ~483 kb from the translocation breakpoint, respectively. Note that *KIAA0825* is directly disrupted by the translocation. Gene transcript orientation for haploinsufficient genes is denoted by a forward arrow (encoded on the positive reference strand) or reverse arrow (encoded on the negative reference strand).


A
chr1
PROX1
USH2A
ESRRG
TGFB2
2.14
2.15
2.16
2.17
2.18
2.19
B
chr5
NR2F1
FAM172
KIAA0825
ANKRD32
MCTP1
9.1
9.2
9.3
9.4
9.5
9.6

3

SUPPLEMENTARY TABLES

Supplementary Table 1. Anchored primers for targeted locus amplification

Name	Target region	Sequence (5'-3')
TLA1a	chr1_downstream_fw1	AACATTTACAAATCCCCAACAC
TLA1b	chr1_downstream_rev1	GGATCAGTGGTGCAGAAATA
TLA2a	chr1_upstream_fw1	GGACAGATTTGCACTTGTTAC
TLA2b	chr1_upstream_rev1	CTTGAAATCGCCAGCATTTG
TLA3a	chr5_downstream_fw1	TCAAGGTAATGAAATTCAGGGA
TLA3b	chr5_downstream_rev1	GTTCTCACTACTGACACCTTAA
TLA4a	chr5_upstream_fw1	TCAAGGAGTAAGACCCAGAG
TLA4b	chr5_upstream_rev1	AGGGTCTGGGTGCATATATT

Supplementary Table 2. Expression of candidate genes in mouse inner ear hair cells in ascending order of false discovery rate (FDR). Genes that both reside within a TAD disrupted by the breakpoints and have a HI score ≤10% were assessed for enriched expression in hair cells, using the SHIELD database (shield.hms.harvard.edu).[4,5] An asterisk (*) denotes significantly enriched expression levels in inner ear hair cells (FDR < 0.01).

Gene	Fold Enrichment in Hair Cells	False Discovery Rate (FDR)
*USH2A**	24.76	2.91E-05
*FAM172A**	2.93	2.18E-03
*ESRRG**	4.55	8.20E-03
MCTP1	1.16	5.97E-02
PROX1	5.48	7.11E-02
KIAA0825	1.32	3.14E-01
ANKRD32	2.39	3.33E-01
NR2F1	1.23	6.80E-01
TGFB2	1.85	7.43E-01

SUPPLEMENTARY DISCUSSION

Assessment of candidate genes in the breakpoint-containing TADs

When we extended our analysis to the identification of possible dysregulated genes by examining the TADs disrupted by the breakpoints and identifying genes with a HI score ≤10%, several additional predicted HI genes were identified within the chr1 and chr5 TAD breakpoint regions: *PROX1*, *USH2A*, *TGFB2*, *NR2F1*, *FAM172A*, *ANKRD32* and *MCTP1* (Supplementary Figure 1). Of these candidates, only *USH2A*, *TGFB2* and *NR2F1* are associated with an OMIM phenotype (#276901 Usher Syndrome, Type IIA, USH2A; #614816 Loeys-Dietz Syndrome 4, LDS4; and #615722 Bosch-Boonstra-Schaaf Optic Atrophy Syndrome, BBSOAS; respectively). Of all three diseases, only Usher Syndrome has a phenotype of SNHL.

When we analyzed the inner ear expression pattern of these genes using data from the SHIELD database, only *USH2A* and *FAM172A* have significantly enriched expression levels in inner ear hair cells, suggesting a potential role for these genes in HL (Supplementary Table 2).[4,5] However, of these genes, *FAM172A* is not predicted to be associated with pathogenicity and *USH2A* is associated with an autosomal recessive disorder, which is inconsistent with the predicted HI score (%HI = 4.17) and any dysregulation that may have occurred by the heterozygous chromosomal rearrangement in DGAP242.

Etiology of the t(1;5)

The cause of DGAP242's *de novo* translocation is unknown. One possibility is that the AT-rich nature of the sequences at the chromosome 1 and 5 breakpoints made them more susceptible to rearrangement. AT-rich sequences have been hypothesized to induce genome instability by forming vulnerable secondary structures. Lesions in these regions could then be repaired by the

non-homologous end-joining repair pathway, leading to improper ligation of non-homologous chromosomes.[7] Another possibility is that this translocation occurred by a microhomology-mediated repair mechanism, which has been reported to underpin many non-recurrent translocations.[8] It has not escaped our attention that a 16 bp region of microhomology (TGACCTTGGACAGCTG) resides 55 bp downstream of the chromosome 1 breakpoint and 643 bp downstream of the chromosome 5 breakpoint.

7

SUPPLEMENTARY REFERENCES

1. Dixon JR, Selvaraj S, Yue F *et al*: Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; **485:** 376-380.

2. Rosenbloom KR, Armstrong J, Barber GP *et al*: The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* 2015; **43:** D670-681.

3. Huang N, Lee I, Marcotte EM, Hurles ME: Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet* 2010; **6:** e1001154.

4. Scheffer DI, Shen J, Corey DP, Chen ZY: Gene Expression by Mouse Inner Ear Hair Cells during Development. *J Neurosci* 2015; **35:** 6366-6380.

5. Shen J, Scheffer DI, Kwan KY, Corey DP: SHIELD: an integrative gene expression database for inner ear research. *Database (Oxford)* 2015; **2015:** bav071.

6. Online Mendelian Inheritance in Man, OMIM®. Johns Hopkins University (Baltimore, MD): McKusick-Nathans Institute of Genetic Medicine, 2015.

7. Kato T, Kurahashi H, Emanuel BS: Chromosomal translocations and palindromic AT-rich repeats. *Curr Opin Genet Dev* 2012; **22:** 221-228.

8. Verdin H, D'Haene B, Beysen D *et al*: Microhomology-mediated mechanisms underlie non-recurrent disease-causing microdeletions of the FOXL2 gene or its regulatory domain. *PLoS Genet* 2013; **9:** e1003358.

APPENDIX B

Pronuclear Injection-Based Targeted Transgenesis

Samantha L.P. Schilit,[1,6,7] Masato Ohtsuka,[2,3,6] Rolen M. Quadros,[4] and Channabasavaiah B. Gurumurthy[4,5,7]

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts
[2]Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Isehara, Kanagawa, Japan
[3]The Institute of Medical Sciences, Tokai University, Isehara, Kanagawa, Japan
[4]Mouse Genome Engineering Core Facility, University of Nebraska Medical Center, Omaha, Nebraska
[5]Developmental Neuroscience, Munroe Meyer Institute, University of Nebraska Medical Center, Omaha, Nebraska
[6]These authors contributed equally to this work
[7]These are co-corresponding authors

Microinjection of DNA expression cassettes into fertilized zygotes has been a standard method for generating transgenic animal models. While efficient, the injected DNA integrates randomly into the genome, leading to potential disruption of endogenous genes or regulatory elements, variation in copy number, or integration into heterochromatic regions that inhibit transgene expression. A recently developed method addresses such pitfalls of traditional transgenesis by targeting the transgene to predetermined sites in the genome that can safely harbor exogenous DNA. This method, called Pronuclear Injection-based Targeted Transgenesis (PITT), employs an enzymatic transfer of exogenous DNA from a donor vector to a previously created landing-pad site in the mouse genome. DNA transfer is achieved using molecular tools such as the Cre-*LoxP* recombinase and PhiC31-*attB/P* integrase systems. Here, we provide protocols for performing PITT and an overview of the current PITT tools available to the research community. © 2016 by John Wiley & Sons, Inc.

Keywords: pronuclear injection • targeted transgenesis • PITT • Cre-*LoxP* recombination • PhiC31-*attB/P* integration

How to cite this article:
Schilit, S.L., Ohtsuka, M., Quadros, R.M., and Gurumurthy, C.B. 2016. Pronuclear injection-based targeted transgenesis. *Curr. Protoc. Hum. Genet.* 91:15.10.1-15.10.28.
doi: 10.1002/cphg.23

INTRODUCTION

Transgenic animals are invaluable for studying gene function and modeling human disease. The canonical method to create transgenic animals involves injecting an exogenous DNA of interest (DOI) into fertilized zygotes (also called pronuclei). These injected zygotes are then transferred to recipient females for gestation, leading to the birth of transgenic founder offspring.

While efficient, this approach results in random integration of the DOI into the genome. The inability to control copy number and integration site can lead to multiple problems including disruption of endogenous genes, multiple integration sites, and repressed expression of the transgene by epigenetic silencing or proximity to heterochromatic regions and local regulatory elements.

Current Protocols in Human Genetics 15.10.1-15.10.28, October 2016
Published online October 2016 in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cphg.23

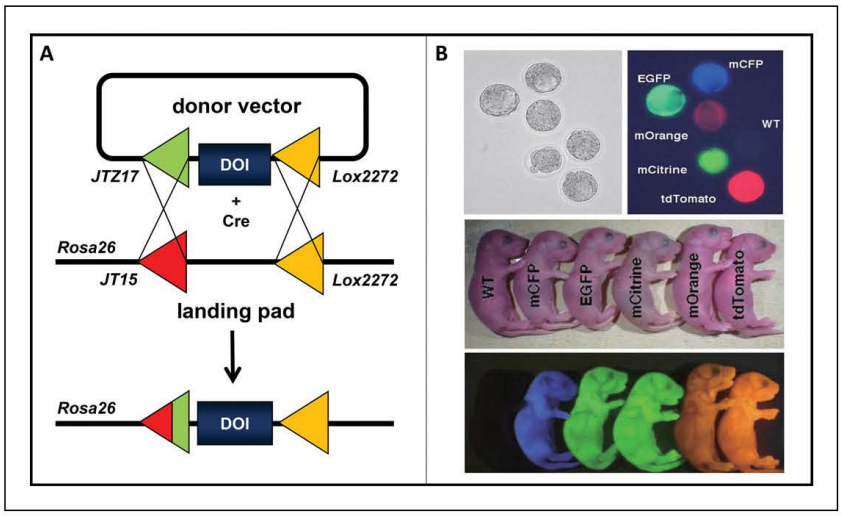


Figure 15.10.1 Schematic of the Cre-PITT system. (**A**) Donor vectors containing a project-specific DNA of interest (DOI) flanked by mutant *LoxP* sites are injected along with Cre (plasmid or Cre mRNA) into fertilized eggs collected from seed mice. (**B**) Blastocysts (top) and neonates (bottom) resulting from PITT of CAG-driven fluorescent reporters at the *Rosa26* locus exhibit ubiquitous and high expression that is consistent between transgenic littermates and from generation to generation (adapted from Ohtsuka et al., 2010). The PhiC31-PITT platform follows a similar methodology except that it uses PhiC31 (instead of Cre) and *attB/P* (instead of *LoxP* elements).

Pronuclear Injection-Based Targeted Transgenesis

To overcome the challenges of random DOI integration, the transgene can be targeted to a specific locus by using gene targeting to embryonic stem (ES) cells. Transgenic ES cells are then injected into blastocysts, which are transferred to surrogate mothers to create chimeric founder animals. However, this method still has pitfalls, as it is time consuming, labor intensive, and more expensive than pronuclear injection-based transgenesis.

Using the best features from both of these standard techniques, we have developed Pronuclear Injection-based Targeted Transgenesis (PITT), a method in which the transgene can be inserted at a predetermined locus through pronuclear injection (Ohtsuka et al., 2010). PITT includes two major steps. The first step involves generation of a "seed mouse" strain by inserting heterotypic recombination (*LoxP*) or integration (*attP*) sites at specific loci in the genome, which will ultimately serve as landing pads for DOI sequences. In the second step, a donor DNA cassette that contains compatible *LoxP* or *attB* sites is injected into pronuclei that have been isolated from the seed mouse strain. The donor cassette gets inserted at the landing pad sites through Cre-recombination or PhiC31-integration, respectively. An overall schematic of the PITT process is shown in Figure 15.10.1. Over the years, our group and other laboratories have improved this method and developed additional tools (Tasic et al., 2011; Ohtsuka et al., 2012b, 2013, 2015). This unit provides a detailed protocol for performing PITT.

The PITT method involves four major steps: (1) designing and building of PITT donor DNA constructs (Basic Protocol 1); (2) synthesis and purification of DNA and RNA components for microinjection (Basic Protocol 2); (3) isolation of embryos from seed mice, microinjection of PITT components, and embryo transfer (Basic Protocol 3); and (4) genotyping of offspring to identify transgenic founders (Basic Protocol 4). A general overview of these steps is presented in Figure 15.10.2.

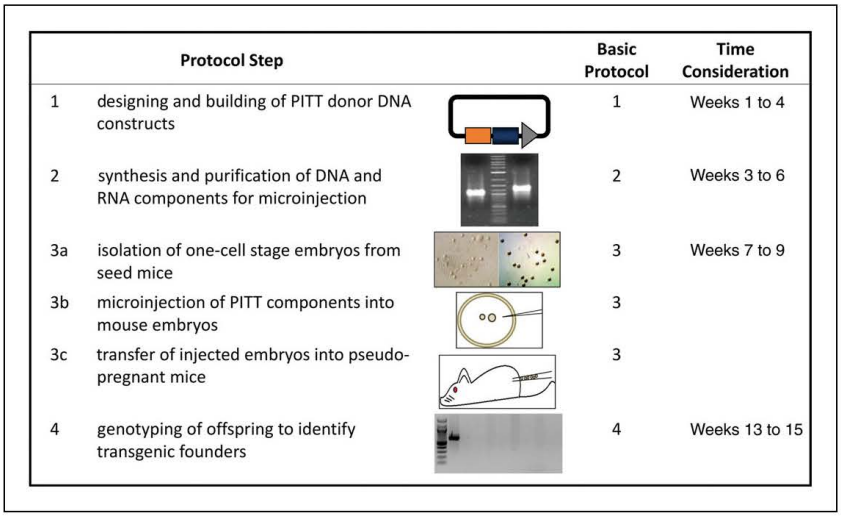


Figure 15.10.2 Overview of PITT steps.

DESIGNING AND BUILDING OF PITT DONOR DNA CONSTRUCTS

BASIC PROTOCOL 1

This protocol involves two major steps: (a) designing the PITT donor construct, and (b) building the PITT donor vector.

Designing the PITT Donor Construct

The PITT donor DNA construct contains the DOI flanked by elements needed for integration into the endogenous landing pad. Table 15.10.1 describes these required sequence elements. There are currently two PITT platforms available, based on the enzyme used for the efficient targeted integration: Cre-PITT and PhiC31-PITT. The Flp-*FRT*

Table 15.10.1 Sequence Elements Used in the PITT Donor Vectors and Seed Mice

Sequence element in the PITT donor vector			Corresponding element in the landing pad of the seed mouse		
PITT Platform	Element	Sequence	Element	Sequence	Notes
Cre-PITT	*Lox2272*	ATAACTTCGTATAGG ATACTTTATACGAAG TTAT	*Lox2272*	ATAACTTCGTATAGG ATACTTTATACGAAG TTAT	Donor DNA between the *LoxP* variants gets inserted through Cre-recombination-mediated cassette exchange.
	JTZ17	ATAACTTCGTATAGC ATACATTATAGCAA TTTAT	*JT15*	AATTATTCGTATAGC ATACATTATACGAAG TTAT	
PhiC31-PITT	*attB*	CCGCGGTGCGGGTGC CAGGGCGTGCCCTTG GGCTCCCCGGGCGCG TACTCCAC	*attP*	GTAGTGCCCCAACTG GGGTAACCTTTGAGT TCTCTCAGTTGGGGG CGTAG	The entire donor plasmid gets inserted through PhiC31-integrase mediated integration.

recombination system is inefficient, so it is not used as a major PITT platform. We instead currently use Flp-*FRT* recombination as an additional tool to remove excess sequences (that come from the vector backbone) after generation of founder transgenic mice.

Platform selection depends on the DOI. For example, Cre-PITT would not be suitable if the DOI consists of *LoxP* sites, because it would result in aberrant recombination in the PITT donor DNA construct. Such a donor construct can be inserted using PhiC31-PITT. Similarly, presence of *attB/P* sites within the DOI would interfere with PhiC31-PITT, in which case Cre-PITT should be used.

A PITT donor vector consists of two major classes of sequence elements: DOI elements and PITT elements. The DOI is the primary transgenic DNA cargo to be inserted into the genome, whereas PITT elements help achieve the targeted insertion of the DOI into a genomic landing pad. The DOI and PITT elements are typically assembled in a standard bacterial plasmid backbone that contains essential plasmid features, such as an origin of replication and an antibiotic selection marker. Some previously developed PITT donor vectors are listed in Table 15.10.2.

The composition of DOI elements depends on the transgenic project. A simple DOI may have an expression cassette with a promoter driving a cDNA or microRNA followed by a polyA signal sequence. A more complex DOI may have an inducible reporter cassette followed by an internal ribosome entry site (or a viral 2A peptidase) with another expression cassette encoding a second reporter and a polyA terminator sequence.

In a donor vector, PITT elements (e.g., *LoxP* or *attB* sites) flank the DOI. The choice and architecture of PITT elements in a donor vector depends on the chosen recombinase/integrase system and the seed mouse strain. For example, heterotypic *LoxP* sites or *attB* sites are included if Cre-PITT or PhiC31-PITT are used, respectively. Our recommendations for compatible combinations of PITT elements needed in a typical PITT donor vector and the corresponding seed mouse landing pads that are available to the scientific community are listed in Tables 15.10.2 and 15.10.3. Schematics of some donor vectors and landing pads in PITT seed mouse strains are shown in Figures 15.10.3 and 15.10.4, respectively.

Once the DOI and PITT elements have been selected and the theoretical sequences of the donor vector have been designed, the vector can be built by custom synthesis from commercial vendors (e.g., Bio Basic, Integrated DNA Technologies, GENEWIZ, GeneArt, GenScript, or other companies). Alternatively, DOI elements can be cloned into preexisting plasmid donor vectors. We have made available donor vectors with certain PITT recombination/integration sites and commonly used DOI promoters and polyA signal sequences (Table 15.10.2). These plasmids include multiple restriction enzyme sites that enable cloning of the desired cDNA or expression cassette.

Building the PITT Donor Vector

This step involves standard molecular biology and recombinant DNA techniques for donor vector cloning. The protocol steps described below provide a choice between two types of cloning methods: a conventional restriction endonuclease (RE)-based method and a more modern technique called Gibson assembly.

Materials

Donor vector backbone (selected from Table 15.10.2)
QIAquick PCR Purification Kit (Qiagen, cat. no. 28104)
Low-melt agarose, e.g., SeaPlaque GTG Agarose (Lonza, cat. no. 50111)
Modified TE: 10 mM Tris, 0.1 mM EDTA, pH 8.0 (e.g., Affymetrix, cat. no. 75793)
Alkaline phosphatase, Calf Intestinal (CIP; NEB, cat. no. M0290)

Phenol, TE-saturated (e.g., Sigma-Aldrich, cat. no. 77607 or Nacalai Tesque, cat. no. 26829-54)
3 M sodium acetate (NaOAc) buffer solution, pH 5.2 (e.g., Sigma Aldrich, cat. no. S7899 or Nacalai Tesque, cat. no. 31150-64)
100% and 70% (v/v) ethanol (Decon Laboratories, cat. no. 07-678-005)
High fidelity DNA polymerase, e.g., Phusion (NEB, cat. no. M0530) or KOD-Plus (Toyobo, cat. no. F0934K)
Primers with appropriate sequence extensions, to amplify the desired cDNA or expression cassette while adding flanking sequences for cloning
Quick Ligation Kit (NEB, cat. no. M2200)
Gibson Assembly Master Mix (NEB, cat. no. E2611)
Competent cells, e.g., NEB 5-alpha competent *E. coli*, high efficiency (NEB, cat. no. C2987)
LB broth with agar (Sigma-Aldrich, cat. no. L3147), containing appropriate antibiotics (e.g., 100 μg/ml ampicillin or 25 μg/ml kanamycin)
LB medium (MP Biomedicals, cat. no. 113002022) with appropriate antibiotics
Plasmid Mini Kit (Qiagen, cat. no. 12125)
Primers for sequencing donor vector

Additional reagents and equipment for the extraction and precipitation of DNA (*APPENDIX 3C*), polymerase chain reaction (Kramer and Coen, 2001), agarose gel electrophoresis (*UNIT 2.7*; Jarco et al., 2001), and bacterial transformation.

NOTE: *UNIT 2.7 does not include anything on PCR.*

Select and prepare the donor vector backbone

1. Decide on the optimal recombinase/integrase platform for your project.

   There are two major platforms for PITT: Cre and PhiC31. The Cre platform uses two heterotypic loxP sites (e.g., JTZ17 and Lox2272) flanking a DOI. The PhiC31 platform has an attB site on one end of the DOI. If your DOI contains LoxP or attB, do not choose the Cre or PhiC31 platforms, respectively. You may also choose to use both platforms in tandem (Ohtsuka et al., 2015).

2. Select the appropriate plasmid backbone for your project based upon your platform of choice, available restriction sites, and the seed mouse strain that you plan to use (refer to Table 15.10.2).

   If restriction sites are not compatible with your DOI, select Gibson assembly instead of RE-based cloning in steps 7 to 11.

3. Digest the selected plasmid (e.g., 0.5 μg) using the specified restriction enzyme(s) (Table 15.10.2) according to the manufacturer's instructions.

4. Purify the digested plasmid fragments using the QIAquick PCR Purification Kit (follow manufacturer's protocol) or by gel purification as follows:

   a. Fractionate the digested DNA by electrophoresis through a 1% low-melt agarose gel.
   b. Stain the gel with ethidium bromide, illuminate the DNA on a LED trans-illuminator, and excise the desired DNA fragment from the gel.
   c. Transfer the gel fragment to a microcentrifuge tube. Keep the volume of the gel fragment ~100 μl or less.
   d. Add 2 volumes (200 μl) of modified TE to the sample and freeze at −80°C for at least 20 min.
   e. Thaw the sample at room temperature, and centrifuge the sample for 10 sec at 1,800 × *g* to pellet insoluble material and collect liquid in the bottom of the microcentrifuge tube. Transfer the supernatant to a new microcentrifuge tube.

Table 15.10.2 List of Representative[a] PITT Donor Vectors and Plasmids for mRNA Synthesis

Vector name (available from)	DNA of interest (DOI)	Landing pad (PITT platform)	Vector features	Restriction sites	Compatible seed mouse strains	References
pAOM (request from authors)	tdTomato expression cassette, "*CAG-tdTomato-pA*"	*JTZ17, Lox2272* (Cre-PITT)	Contains "*IRES-lacZ-pA*" and "*CAG-hyg-pA*" cassettes in a pUC119-based vector. Has amp^R for ampicillin resistance and *FRT* sites for Flp-mediated extra sequence excision.	tdTomato can be replaced with a different DOI using *Age*I and *Fse*I.	TOKMO-1 TOKMO-2	Ohtsuka et al. (2010)
pAOT (request from authors)	eGFP expression cassette with a synthetic miRNA against the gene encoding tyrosinase, "*CAG-eGFP-miR(Tyr-1/2)-pA*"	*JTZ17, Lox2272* (Cre-PITT)	Contains "*IRES-lacZ-pA*" and "*CAG-hyg-pA*" cassettes in a pUC119-based vector. Also has amp^R for ampicillin resistance and *FRT* sites for Flp-mediated extra sequence excision.	eGFP can be replaced with a different reporter using *Age*I and *Bsr*GI. The miRNA can be replaced with a different miRNA using *Bam*HI and *Bgl*II	TOKMO-1 TOKMO-2	Ohtsuka et al. (2010)
pA748 (request from authors)	Sucrose counterselection cassettes with GFPuv, "*sacB-GFPuv-sacB*" cassette	*JTZ17, Lox2272* (Cre-PITT)	Contains "*IRES-eGFP-pA*" and "*CAG-hyg-pA*" cassettes in a pUC119-based vector. Has amp^R for ampicillin resistance and *FRT* sites for Flp-mediated extra sequence excision.	The DOI cassette can be replaced with a different DOI using *Sfi*I.	TOKMO-1 TOKMO-2	Ohtsuka et al. (2010)
pAWV (Addgene #62710)	tdTomato expression cassette with a synthetic miRNA against the GFP gene, "*CAG-tdTomato-miR(eGFP)-pA*"	*JTZ17, Lox2272* (Cre-PITT)	pBR322-based vector. Contains a "*CAG-FLPe-pA*" cassette to aid in self-removal of extra sequence by Flp-recombination. Has amp^R for ampicillin resistance.	The entire DOI cassette can be replaced with a different DOI using *Not*I. The miRNA region can be replaced with a different miRNA using *Sal*I and *Bgl*II.	TOKMO-1 TOKMO-2	Ohtsuka et al. (2013); Miura et al. (2015)
pAWK (Addgene #62713)	Expression cassette for synthetic miRNA against the GFP gene, "*CAG-miR(eGFP)-pA*"	*JTZ17, Lox2272* (Cre-PITT)	pBR322-based vector. Contains a "*CAG-FLPe-pA*" cassette to aid in self-removal of extra sequence by Flp-recombination. Has amp^R for ampicillin resistance.	The entire DOI cassette can be replaced with a different DOI using *Not*I. The miRNA region can be replaced with a different miRNA using *Sal*I and *Bgl*II.	TOKMO-1 TOKMO-2	Ohtsuka et al. (2013); Miura et al. (2015)

continued

Table 15.10.2 List of Representative[a] PITT Donor Vectors and Plasmids for mRNA Synthesis, *continued*

Vector name (available from)	DNA of interest (DOI)	Landing pad (PITT platform)	Vector features	Restriction sites	Compatible seed mouse strains	References
pBFD (request from authors)	Dre expression cassette from the Thy1 promoter, "*Thy1-Dre-pA*"	*JTZ17*, *Lox2272* (Cre-PITT)	pBR322-based vector. Has amp^R for ampicillin resistance and *FRT* sites for Flp-mediated extra sequence excision.	The entire DOI cassette can be replaced with a different DOI by *Eag*I.	TOKMO-1 TOKMO-2	Ohtsuka et al. (2013)
pBDR (Addgene #62663)	Promoterless tdTomato cassette, "*tdTomato-pA*"	*JTZ17*, *Lox2272*, *attB* (Cre-PITT and/or PhiC31-PITT)	pIDTSmart-based vector. Has kan^R for kanamycin resistance and mutant *FRT* sites (*F14* and *F15*) for Flp-mediated extra sequence excision.	The entire DOI cassette can be replaced with a different DOI using *Age*I and *Eco*RI. The tdTomato can be replaced using *Age*I and *Fse*I.	TOKMO-3	Ohtsuka et al. (2015)
pBHL (request from authors)	Promoterless tdTomato cassette, "*tdTomato-pA*"	*JTZ17*, *Lox2272*, *attB* (Cre-PITT and/or PhiC31-PITT)	This plasmid is derived from pBDR but the order of *attB* and *JTZ17* is reversed.	The entire DOI cassette can be replaced with a different DOI using *Age*I and *Eco*RI. The tdTomato can be replaced using *Age*I and *Fse*I.	TOKMO-3	Unpublished
pBBI (Addgene #65795)	iCre expression cassette, "*T7-iCre-AAAA...*"	-	To be used for iCre mRNA synthesis. It is derived from pcDNA3.1 and has amp^R for ampicillin resistance.	Linearize with *Xba*I before mRNA synthesis.	-	Ohtsuka et al. (2013)
pBBK (Addgene #62670)	PhiC31o expression cassette, "*T7-PhiC31o-AAAA...*"	-	To be used for PhiC31o mRNA synthesis. It is derived from pcDNA3.1 and has amp^R for ampicillin resistance.	Linearize with *Xba*I before mRNA synthesis.	-	Ohtsuka et al. (2015)

Abbreviations: *AAAA...* = polyA stretch; amp^R = ampicillin resistance cassette; *CAG* = synthetic cytomegalovirus (CMV) early enhancer/ chicken ß-actin promoter/ rabbit ß-globin splice acceptor site for strong expression in mammalian cells; DOI = DNA of interest; *Dre* = phage D6 recombinase; *eGFP* = enhanced green fluorescent protein; *FLPe* = enhanced FLP recombinase; *GFPuv* = ultraviolet light-excitable green fluorescent protein; *hyg* = hygromycin resistance cassette; *iCre* = codon-improved Cre recombinase; *IRES* = internal ribosome entry site; kan^R = kanamycin resistance cassette; *lacZ* = ß-galactosidase; *pA* = polyA signal sequence; *PhiC31o* = codon-optimized PhiC31 integrase; *puro* = puromycin resistance cassette; *sacB* = levansucrase, which confers sensitivity to sucrose as counterselection; *T7* = T7 prokaryotic promoter; *tdTomato* = red fluorescent protein; *Thy1* = promoter frequently used for expression in neurons.

[a]Please note that this is not a comprehensive list of PITT donor vectors or plasmids for mRNA synthesis. Other versions of available plasmids may be found in our previous publications (Ohtsuka et al., 2010, 2012a, 2013, 2015). In addition, the construction of donor vectors for the R26 Cr4 mouse lines (described in Table 15.10.3) is in progress and these vectors can be requested from the authors. Tasic et al. (2011) have reported other PhiC31-based vector systems, which are available through Charles River Laboratories, USA, under the trade name TARGATT system.

Table 15.10.3 List of Representative[a] PITT Seed Mouse Strains

Common name	Strain name (genetic background)	Landing pad (PITT platform)	Locus	Available from	References
TOKMO-1	*Gt(ROSA)26Sor <tm1Maoh>* (129/C57BL/6J mixed)	*JT15, Lox2272* (Cre-PITT)	*Rosa26*	Authors	Ohtsuka et al. (2010)
TOKMO-2	*H2-T3<tm1Maoh>* (129/C57BL/6J mixed)	*JT15, Lox2272* (Cre-PITT)	*H2-Tw3*	Authors	Ohtsuka et al. (2010)
R26 Cr4-Cre-PITT	*CRISPR(ROSA)(Cre-PITT)CBG>* (129/C57BL/6N mixed)	*JT15, Lox2272* (Cre-PITT)	*Rosa26*	Authors	Quadros et al. (2015)
R26 Cr4-PhiC-PITT	*CRISPR(ROSA)(PhiC31-PITT)CBG>* (C57BL/6N)	*attP* (PhiC31-PITT)	*Rosa26*	Authors	Unpublished
TOKMO-3	*Gt(ROSA)26Sor<tm10 (PITT)Maoh>* (C57BL/6N)	*JT15, Lox2272, attP* (Cre-PITT and/or PhiC31-PITT)	*Rosa26*	RIKEN BioResource Center, RBRC06517	Ohtsuka et al. (2015)
R26 Cr4-CrePhiC-PITT	*CRISPR(ROSA)(Cre/PhiC31-PITT)CBG>* (129/C57BL/6N mixed)	*JT15, Lox2272, attP* (Cre-PITT and/or PhiC31-PITT)	*Rosa26*	Authors	Unpublished

[a]Please note that this is not a comprehensive list of PITT seed mice. Specifically, Tasic et al. (2011) reported other PhiC31-based seed mice, which are available through Charles River Laboratories, USA, under the trade name TARGATT system.

5. Dephosphorylate the plasmid fragment by adding 20 units of CIP and incubating at 37°C for 40 min.

   This step is important to prevent self-ligation of the vector during cloning.

6. Extract the DNA with phenol and precipitate with ethanol as follows:

   a. Add an equal volume of TE-saturated phenol, vortex, and centrifuge for 5 min at 13,000 × *g* to separate the aqueous and organic phases.
   b. Transfer the aqueous phase (top) to a new microcentrifuge tube, and add 0.1 volume 3 M NaOAc (pH 5.2) and 2.5 volumes ethanol to precipitate the DNA.
   c. Pellet the precipitated DNA by centrifugation for 10 min at 13,000 × *g*, room temperature, and aspirate the supernatant.
   d. Wash the pellet with 70% ethanol, centrifuge for one minute at 13,000 × *g*, room temperature, and aspirate the supernatant.
   e. Let the pellet air dry for 10 min.
   f. Suspend the dried pellet in modified TE (e.g., 2 µl).

Prepare the DOI

7. Amplify the DOI by polymerase chain reaction (PCR) using cDNA or an expression cassette as template, a high fidelity polymerase, and primers with extensions suitable for RE-based cloning or Gibson assembly. Alternatively, order a custom synthesis of the DOI with appropriate sequence extensions.

   Refer to The NEBuilder Assembly Tool to design optimal primers (http://nebuilder.neb.com).

   Make sure to take the intended DOI orientation into account when designing primers.

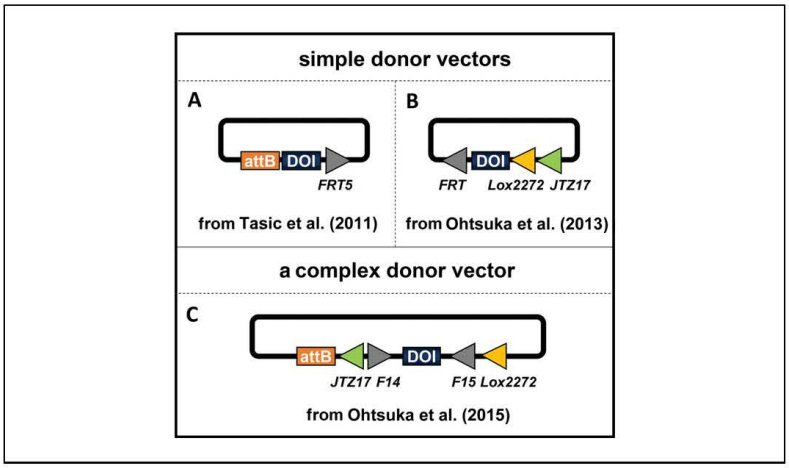


Figure 15.10.3 Examples of simple and complex PITT donor vectors. Simple PITT donor vectors contain either (**A**) *attB* elements for the PhiC31 platform or (**B**) *LoxP* elements for the Cre-PITT platform. (**C**) This complex PITT donor vector contains PITT elements for both platforms (Cre-*LoxP* and PhiC31-*attB*). While not a part of the PITT platform, *FRT* elements in donor vectors serve as a tool for removing extra sequences in the founder mice using Flp recombinase.

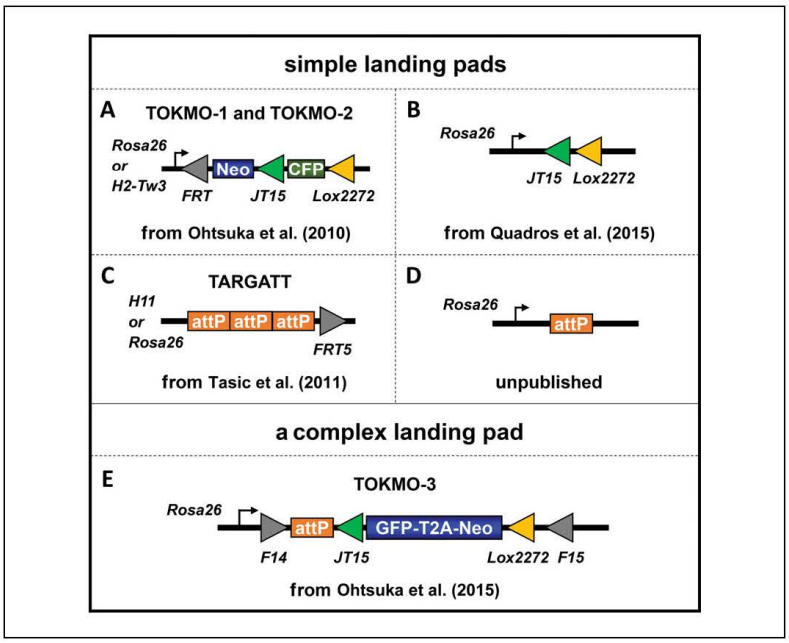


Figure 15.10.4 Examples of simple and complex landing pads in seed mouse strains. (**A-D**) Simple landing pads include the critical elements to facilitate targeted insertion for a single PITT-platform. Examples of landing pads include *LoxP* variant recombination sites (Cre-PITT; **A** and **B**) or *attP* integration sites (PhiC31-PITT; **C** and **D**). (**E**) Complex landing pads can include a combination of both *attP* and *LoxP* elements for using PhiC31-PITT and Cre-PITT platforms either independently or together. While not a part of the PITT platform, *FRT* elements in landing pads serve as a tool for removing extra sequences in the founder mice using Flp recombinase.

For RE-based cloning, the extensions should include RE sites that are compatible with the donor plasmid cloning site.

For Gibson assembly, use 15- to 80-bp extensions homologous to the ends of the digested plasmid backbone.

8. **For DOIs amplified by PCR**: Confirm the correct size amplicon by gel electrophoresis.
9. **For RE-based cloning method only**: Digest ~50 to 300 ng of the DOI fragment using the specified restriction enzyme(s), according to manufacturer's instructions.
10. **For all DOIs amplified by PCR or digested by restriction enzymes**: Purify the DNA sample using a QIAquick PCR Purification Kit, or by gel purification, phenol extraction, and ethanol precipitation as described in steps 4 and 6.

Assemble the donor vector and identify clones

11a. **RE-based cloning**: Combine appropriate amounts of the vector and DOI fragment, and ligate using the Quick Ligation Kit (follow manufacturer's instructions) for 5 min at room temperature.

11b. **Gibson assembly**: Combine appropriate amounts of vector, DOI fragment, and Gibson Assembly Master Mix (see manufacturer's instructions) in a thermocycler tube. Incubate the reaction in a thermocycler for 60 min at 50°C.

12. Store samples at 4°C or −20°C until transformation.
13. Transform sample into competent *E. coli* cells, and select transformed bacteria on LB agar medium containing the appropriate antibiotic. Incubate overnight at 37°C.

    In some cases a DOI may render the plasmid unstable. If so, try transforming into a different E. coli strain, such as SURE (stop unwanted rearrangement events), or try growing transformants at 30°C instead of 37°C.

14. Pick colonies to 5 ml LB medium with the appropriate antibiotic, incubate the liquid cultures overnight at 37°C, with shaking, and isolate plasmid DNA using the Qiagen Plasmid Mini Kit.

    Save the residual bacterial cultures at 4°C until after transformants are analyzed. The residual culture from positive colonies can be used as a starter culture to prepare larger amounts of DNA and for storage. This will save time by avoiding another transformation of positive clones.

15. Identify correct clones by restriction digestion analysis and confirm by DNA sequencing.

BASIC PROTOCOL 2

SYNTHESIS AND PURIFICATION OF DNA AND RNA COMPONENTS FOR MICROINJECTION

Microinjection components for PITT include the donor vector and an mRNA encoding Cre or PhiC31. While PITT donor vectors are unique for supporting a specific transgenic project, there are well-established plasmids available for generating mRNA encoding the enzymes (such as pBBI and pBBK, see Table 15.10.2 and Fig. 15.10.5). You can also directly inject plasmids with DNA encoding the recombinase or integrase enzymes, but we have found empirically that mRNA is ~5-fold more efficient for PITT (Ohtsuka et al., 2013). We have observed that plasmid injection is inferior to mRNA injection because the delay in translation of the recombinase/integrase results in a higher likelihood of

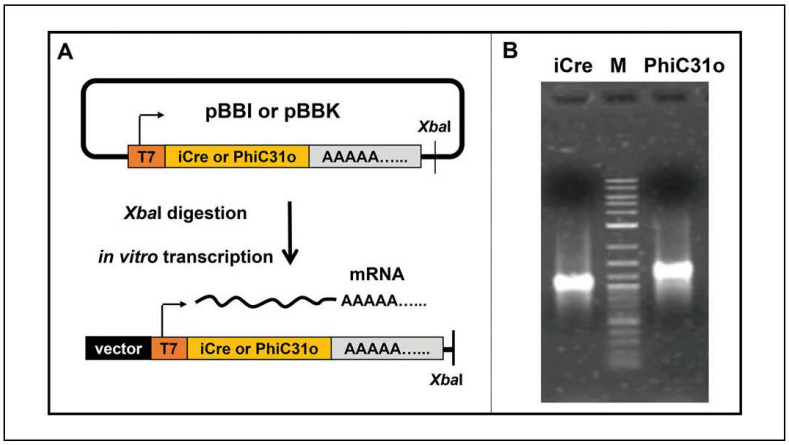


Figure 15.10.5 Generation of mRNA encoding Cre recombinase or PhiC31 integrase. (**A**) Plasmids (pBBI = iCre; pBBK = PhiC31o) are linearized with *Xba*I prior to in vitro transcription by T7 RNA polymerase. (**B**) An image of a non-denaturing agarose gel shows 1 µg of synthesized mRNAs fractionated alongside a 100-bp DNA ladder (M). The approximate sizes for iCre and PhiC31o mRNAs are 1.3 kb and 2.1 kb, respectively. Note that the mRNAs do not migrate perfectly with the DNA molecular weight marker. For accurate size analysis, RNA markers may be included in the gel.

mosaicism in founder mice, and therefore less efficient germline transmission (Ohtsuka et al., 2013). Furthermore, injection of the Cre and PhiC31-encoding plasmids may even result in their own undesired insertion into the genome. For these reasons, the protocol below describes the mRNA technique.

Materials *(see also Basic Protocol 1)*

Donor vector, transformed into *E. coli* (from Basic Protocol 1)
Recombinase/integrase in vitro transcription vector, transformed into *E. coli*
  Cre platform: pBBI, vector for iCre mRNA synthesis (Addgene, cat. no. 65795; *https://www.addgene.org/65795/*; Ohtsuka et al., 2013)
  PhiC31 platform: pBBK, vector for PhiC31o mRNA synthesis (Addgene, cat. no. 62670; *https://www.addgene.org/62670/*; Ohtsuka et al., 2015)
HiSpeed Plasmid Midi Kit (Qiagen, cat. no. 12643)
Ultrafree-MC filter, 0.45-µm pore size (Millipore, cat. no. UFC30HV00)
Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, cat. no. P11496)
*Xba*I (NEB, cat. no. R0145), to linearize mRNA synthesis vector
Nuclease-free water (Ambion, cat. no. AM9937), not DEPC-treated
25:24:1 (v/v/v) phenol/chloroform/isoamyl alcohol, pH 7.9 (e.g., Nacalai Tesque, cat. no. 25970-14)
Chloroform (e.g., Nacalai tesque, cat. no. 08402-55)
mMESSAGE mMACHINE T7 ULTRA Kit (Ambion, cat. no. AM1345)
MEGAclear Transcription Clean-Up Kit (Ambion, cat. no. AM1908)
Microinjection buffer (see recipe)

1.5-ml RNase-free microcentrifuge tubes
NanoDrop spectrophotometer or fluorescence microplate reader

Additional reagents and equipment for the spectroscopic quantification of nucleic acids (*APPENDIX 3D*; Gallagher and Desjardins, 2007)

Prepare plasmid DNA for donor and mRNA-encoding vector

1. Grow transformed bacterial cultures and isolate donor and mRNA-encoding plasmids using the Qiagen HiSpeed Plasmid Midi Kit, according to the manufacturer's protocol. Elute the DNA with TE.

   We recommend growing these transformed bacterial strains in 200 ml LB medium with appropriate drug selection prior to harvesting (to achieve ~50- to 100-μg yield).

Extract and precipitate the DNA

2. Add an equal volume of TE-saturated phenol, vortex to mix, and separate the phases by centrifugation for 5 min at 13,000 × *g*, room temperature.
3. Transfer the aqueous phase to a new microcentrifuge tube, add 0.1 volumes 3 M NaOAc (pH 5.2) and 2.5 volumes ethanol, and mix to precipitate the DNA.
4. Pellet the precipitated DNA by centrifugation for 10 min at 13,000 × *g*, room temperature.
5. Aspirate the supernatant, and wash the pellet with 70% ethanol.
6. Centrifuge 1 min at 13,000 × *g*, room temperature, aspirate the ethanol wash, and air dry the pellet for 10 min.
7. Suspend the DNA pellet in 50 to 100 μl modified TE, and store at 4°C overnight.
8. Filter the donor vector using a pre-equilibrated Ultrafree-MC filter for 2 min at 12,000 × *g*, 4°C.

   Filtering is critical to avoid clogging the injection needle during microinjection of one-cell stage embryos.

9. Determine plasmid concentration using the Quant-iT PicoGreen dsDNA Assay Kit.

   The final concentration may range from 200 to 2000 ng/μl.

10. Store plasmids at −20°C until ready to use.

Prepare mRNA

11. In a 1.5-ml microcentrifuge tube, digest 30 μg of the mRNA synthesis vector (e.g., 30 μl of 1 μg/μl vector) with 120 units *Xba*I in a total volume of 120 μl for 3 to 3.5 hr at 37°C. Analyze 0.5 μl by agarose gel electrophoresis to ensure complete digestion.
12. Add 280 μl nuclease-free water and 40 μl 3 M NaOAc (pH 5.2) to the digested sample.
13. Add 450 μl phenol/chloroform/isoamyl alcohol and vortex to extract the DNA.
14. Separate the phases by centrifugation for 5 min at 13,000 × *g*, room temperature, and transfer the ~440 μl aqueous phase to a new microcentrifuge tube.
15. Add 450 μl chloroform, and vortex to extract the DNA.
16. Separate the phases by centrifugation for 5 min at 13,000 × *g*, room temperature, and transfer the ~435 μl aqueous phase to a new microcentrifuge tube.
17. Add approximately 2.0 volumes (~870 μl) ethanol to precipitate the DNA.

    At this step, the DNA is visible in the solution.

18. Pellet the precipitated DNA by centrifugation for 10 min at 13,000 × *g*, room temperature. Remove the supernatant and wash the pellet with 70% ethanol.
19. Pellet the washed DNA by centrifugation for 1 min at 13,000 × *g*, room temperature, remove the supernatant, and let the pellet air dry for 10 min.
20. Suspend the dried pellet in 23 μl nuclease-free water.
21. Dilute 1 μl *Xba*I-linearized mRNA vector into 9 μl nuclease-free water and determine the concentration by NanoDrop or the Quant-iT PicoGreen dsDNA Assay Kit.
22. Dilute the sample to ~0.5 μg/μl with nuclease-free water.
23. Synthesize mRNA transcripts using the mMESSAGE mMACHINE T7 ULTRA Kit, according to the manufacturer's protocol but with the following adjustments to scale up the reaction for a large amount of stock solution:
    a. Prepare a 100 μl transcription reaction, using 5 μg of the linearized plasmid.
    b. Incubate the mixed sample for 3 hr at 37°C, add 5 μl TURBO DNase I, and incubate again for 15 min at 37°C.
    c. Divide the sample into two 52.5 μl aliquots.

    We omit the polyA tailing procedure following in vitro transcription because our plasmids already include a polyA sequence.
24. Recover mRNA using the MEGAclear Transcription Clean-Up Kit according to the manufacturer's protocol with the following adjustments:
    a. Treat each aliquot independently throughout the purification.
    b. When eluting RNA, select RNA elution option 2, and perform the second optional elution for a final volume of 100 μl for each eluate.
    c. Combine the eluates of both reaction aliquots (200 μl total) and perform the optional ethanol precipitation step with ammonium acetate (NH_4Ac).

    We recommend performing a 15-min centrifugation step at 4°C.

    d. Suspend the pellet in 50 to 100 μl nuclease-free water (included in the kit).
25. Filter the mRNA using a pre-equilibrated Ultrafree-MC filter for 2 min at 12,000 × *g*, 4°C.
26. Quantify the mRNA concentration using a NanoDrop spectrophotometer, and confirm that the RNA is good quality by 1% agarose gel electrophoresis.

    A representative gel image of iCre and PhiC31o mRNA preps is shown in Fig. 15.10.5.
27. Dilute the purified mRNA to 400 to 500 ng/μl, dispense 5-μl aliquots into RNase-free tubes, and store at −80°C.

Prepare the microinjection solution

28. Thaw frozen stocks of the donor vector (200 to 2000 ng/μl) and mRNA (400 to 500 ng/μl) on ice.
29. Gently tap the tubes to mix the solutions, and collect solutions in the bottom of the tubes by centrifugation for 3 min at 18,700 × *g*, 4°C.
30. Take 2 μl of donor vector from the top of the stock solution, and dilute it into microinjection buffer to a concentration of about 20 ng/μl (at least double the desired final concentration).

31. Take 2 to 4 µl of mRNA from the top of the stock solution, and dilute it into microinjection buffer to a concentration of about 1 to 2 ng/µl for iCre or 15 to 30 ng/µl for PhiC31o (at least double the desired final solution concentration).

    Optional: At this stage, the mRNA can be filtered again using the Ultrafree-MC filter.

32. Collect the diluted donor vector and mRNA solutions in the bottom of the tubes by centrifugation for 3 min at 18,700 × *g*, 4°C.

33. Carefully combine the appropriate volumes (typically 5 to 40 µl from the top of the working dilution tubes) of donor vector and mRNA solution in an RNase-free 1.5-ml microcentrifuge tube, mix gently by pipetting, and adjust final concentration with microinjection buffer, if needed, to make the final microinjection solution (typically ~100 µl).

    The final concentrations should be 10 ng/µl for donor vector, 0.5 to 1 ng/µl for iCre mRNA, and 7.5 to 15 ng/µl for PhiC31o mRNA.

34. Analyze 3 µl of injection solution by agarose gel electrophoresis to check the quality and concentration of nucleic acids in the injection mix.

35. Either use the injection solution immediately (Basic Protocol 3), or store at −80°C until use.

BASIC PROTOCOL 3

ISOLATION OF EMBRYOS FROM SEED MICE, MICROINJECTION OF PITT COMPONENTS, AND EMBRYO TRANSFER

Targeted transgenesis follows similar technical steps to traditional transgenesis, but is unique in that it utilizes zygotes isolated from seed mice as opposed to from wild type animals. These seed mouse strains have been previously developed by us (Ohtsuka et al., 2010, 2012a, 2015) and other groups (Tasic et al., 2011), so the generation of PITT seed mice is not described in this unit. If a given project requires development of a new seed mouse, it can be made using standard gene targeting in mouse embryonic stem cells (Pease and Saunders, 2011; Behringer et al., 2014) or more rapidly with CRISPR-based approaches (Quadros et al., 2015). The basic steps for this protocol include isolation of one-cell stage embryos from seed mice, microinjection of PITT components into mouse embryos, and transfer of injected embryos into pseudo-pregnant mice. These steps follow standard mouse transgenesis protocols that have been described previously (Pease and Saunders, 2011; Behringer et al., 2014). Notably, they are very similar to those described in *UNIT 15.7* (see Basic Protocol 3 of Harms et al., 2014), and are given below with slight modifications relevant to PITT. We also extend techniques from the previous unit by providing an alternative protocol for producing fertilized zygotes in place of steps 1 to 16 (see Support Protocol).

Materials

Mice: A list of seed mouse strains can be found in Table 15.10.3.

Egg donors: wild-type C57BL/6 female mice procured at 3-weeks old (Charles River Laboratories, Wilmington, MA).

Homozygous seed female mice (bred in-house), instead of wild type females, can also be used. However, we have found that hemizygous embryos survive better than homozygous embryos after microinjection.

Stud males: homozygous seed mice, 3- to 6-months old

Pseudo-pregnant recipients: Crl:CD1(ICR) female mice, 8- to 12-weeks old (purchased at 5- to 6-weeks old; Charles River Laboratories, Wilmington, MA)

Vasectomized males: CD1 male mice purchased at 5- to 6-weeks old (Charles River Laboratories, Wilmington, MA) and vasectomized as previously described (Behringer et al., 2014)

The vasectomized mice can be used from 6 months to 1 year of age.

2 IU/μl pregnant mare's serum gonadotropin (PMSG; National Hormone and Peptide Program, Harbor–UCLA Medical Center, Torrance, CA)

1 IU/μl human chorionic gonadotropin (hCG; National Hormone and Peptide Program, Harbor–UCLA Medical Center, Torrance, CA)

Lyophilized PMSG and hCG are supplied by National Hormone and Peptide Program in vials containing 2000 IU and 1000 IU, respectively. Upon first use, reconstitute in PBS (Millipore, cat. no. BSS-1006-B) to a final concentration of 100 IU/100 μl. Prepare 100-μl aliquots of this 20× stock solution and store at −80°C.

1× EmbryoMax M2 medium with phenol red (Millipore, cat. no. MR-015-D), for embryo handling and microinjection

1× EmbryoMax M2 medium with phenol red and hyaluronidase (Millipore, cat. no. MR-051-F), for dissociation of the cumulus oophorus complex

1× EmbryoMax KSOM medium with ½ amino acids (Millipore, cat. no. MR-106-D), for embryo incubation

Microinjection buffer (see recipe)

EmbryoMax filtered light mineral oil (Millipore, cat. no. ES-005-C)

Individually ventilated cages (Allentown, Lab Products, or Tecniplast)

35 × 10-mm Falcon tissue culture dish (Corning, cat. no. 353001)

Falcon IVF/organ culture dish (Corning, cat. no. 353653)

Flexipet pipette, 130-μm (Cook Medical, cat. no. K-FPIP-1130-10BS-5), for collecting embryos

Glass capillaries, 4-in × 1-mm (World Precision Instruments, cat. no. TW100F-4)

MicroFil 28-gauge × 97-mm long (World Precision Instruments, cat. no. MF28G)

1-ml subcutaneous syringe and needle (BD, cat. no. 309597)

150 × 15-mm tissue culture dish (Falcon, cat. no. 351058)

Nunc Lab-Tek chamber slide system (Lab-Tek, cat. no. 177372)

Holding micropipets (Origio, cat. no. MPH-SM-20)

Heraeus HERAcell 150i Tri–gas incubator, equipped with Coda Inline filters

Large slide warmer (Spectrum Scientifics, cat. no. 3875)

Dissecting scope: e.g., Leica MZ 9.5 with Plan 0.5× condenser lens (model 10 446 157), base (model 10 445 367), and tilt head

Heating glass (Live Cell Instrument, cat. no. HG-T-Z002) with temperature controller (Live Cell Instrument, model CU-301)

Mouth pipetting apparatus (assembled as described by Gurumurthy et al., 2016).

Glass micropipette puller (Sutter Instrument Co., model P97), outfitted with a 2.5 × 2.5-mm box filament (cat. no. FB255B)

Microinjection scope (example below) equipped with Narishige IM-300 microinjector, NT-88-V3 manipulator Nikon Eclipse TE 2000-E with DIC, equipped with Narishige IM 300 microinjector and NT-88-V3 micromanipulators, and Live Cell Instrument HG-T-Z002 heating glass with CU-301 temperature controller

Leica DM IRB with HC PLAN 10×/22 adjustable eyepiece (cat. no. 11 507 804), S70/0.30 condenser lens, C PLAN 4×/0.10 (cat. no. 11 506 074), N PLAN L 20×/0.40 CORR (cat. no. 11 506 057), N PLAN L 40×/0.55 CORR (cat. no. 11 506 059) objectives

Additional reagents and equipment for mouse husbandry, euthanasia, and surgery (*APPENDIX 3L*; Donovan et al., 2003) and transgenesis (Behringer et al., 2014; Pease and Saunders, 2011)

IMPORTANT: Be sure all animal methods are approved by your institutional animal care and use committee (IACUC).

Induce superovulation and isolate one-cell embryos from seed mice

1. House mice in individually ventilated cages on a 14 hr on/10 hr off light cycle (on at 06:00, off at 20:00).
2. On the day of the hormone injection (day 1), thaw a 100-µl vial of PMSG, and dilute with 1.9 ml PBS to obtain a final concentration of 5 IU/0.1 ml.
3. Intraperitoneally inject 10 to 20 donor female mice, each with 5 IU PMSG (0.1 ml) around 12:00 on day 1, and discard leftover hormone.
4. On day 3, thaw a 100-µl vial of hCG, and dilute with 1.9 ml PBS to obtain a final concentration of 5 IU/0.1 ml.
5. Approximately 48 hr after the PMSG injection, intraperitoneally inject each female mouse with 5 IU hCG (0.1 ml), and discard leftover hormone.
6. Breed with stud males overnight.
7. On the morning of day 4 (around hour 8:00), prepare the following dishes:
   a. Oviduct collection dish: 35-mm tissue culture dish with 2 ml M2 medium (one dish per up to 10 females).
   b. Hyaluronidase dish: 35-mm tissue culture dish with 1.5 ml M2 hyaluronidase medium (one dish per up to 10 females).
   c. Wash dish: 35-mm tissue culture dish with 1.5 ml M2 medium (at least two per injection batch).
   d. KSOM rinse dish: 35-mm tissue culture dish with 1.5 ml KSOM medium (pre-equilibrated).
   e. Incubation dish: IVF/organ culture dish with 1 ml KSOM medium (two per injection batch, both pre-equilibrated).
   f. Embryo Transfer dish: 35-mm tissue culture dish with 1.5 ml M2 medium.
8. Euthanize plugged donor females approximately 20 hr after hCG injection (about hour 8:00 on day 4).

   Be sure all methods are approved by your IACUC committee.
9. Surgically remove oviducts, place them in the oviduct collection dish, and maintain tissue samples at 37°C on a heated slide warmer.

   Make sure oviducts are removed within 10 min from euthanasia.
10. Once oviducts have been collected, clear the work space and begin dissociating the cumulus-oocyte complexes (COC), one at a time.

    The following steps are performed under a dissecting scope maintained at 37°C.

    a. Place an oviduct in the hyaluronidase dish.
    b. Dissect the COC by disrupting the ampulla with a pair of fine forceps.
    c. Continue to process the remaining oviducts while working efficiently.

    If all oviducts cannot be processed in under 10 min after they are dissected, it is better to plan a dissection of fewer animals in the future.
11. Once the last COC has been expelled from the ampulla, collect individual oocytes from the dish using the mouth pipetting apparatus attached to a 130-µm flexipet pipette.

    The first set of oocytes should begin to dissociate in the time it takes to harvest the rest of the COCs.

12. Transfer the oocytes to the wash dish (avoiding as much of the hyaluronidase medium and as many of the cumulus cells as possible).

    This will inactivate the residual hyaluronidase.

13. (Optional) Transfer oocytes to a second wash dish to remove residual cumulus cells and hyaluronidase.

14. One-by-one, collect the oocytes using the flexipet pipette into a fresh wash dish, and count the number of fertilized eggs (zygotes) and unfertilized oocytes. Record this information to calculate the fertilization efficiency.

    Zygotes can be distinguished from unfertilized oocytes by the presence of two pronuclei.

15. Transfer only zygotes to the KSOM rinse dish to wash residual M2 medium from the embryos.

16. Transfer the zygotes to the incubation dish until needed (typically 30 min to 1 hr), and incubate at 37°C, 5% CO_2 to maintain a pH range of 7.23 to 7.42.

    Zygotes should be injected within two hours, as zygotes that advance past the one-cell stage are no longer suitable for injection.

Pull microinjection needles

17. Pull injection capillaries on the morning of the microinjection using the pipette puller and the following program:

Glass:	TW100F-4
Heat:	Ramp +5
Pull:	70
Velocity:	120
Pressure:	200
Time:	100 delay

    Use sterile technique to keep the capillaries nuclease-free.

Microinject PITT components into mouse embryos

PITT requires both a cytoplasmic and nuclear injection into embryos. The mRNA must be translated in the cytoplasm, and the donor DNA needs to be delivered to the nucleus for targeted integration. This injection approach is very similar to the protocols followed for mouse genome editing using CRISPR/Cas (Harms et al., 2014).

18. Prewash a 28-gauge MicroFil three times with sterile microinjection buffer.

19. Connect the prewashed MicroFil to a 1-cc tuberculin syringe, and backfill 5 or 6 injection needles with 1 to 2 μl of microinjection solution each (from Basic Protocol 2).

20. Affix an injection needle to the microinjector, and store the remaining prefilled needles in a needle holder on ice as an additional precaution to prevent RNA degradation during this step.

    The needle holder is made from a 150-mm tissue culture dish outfitted with a 0.25-cm diameter rod-shaped modeling clay. The injection needles filled with the solution are pushed into the clay and the entire storage unit is placed directly in contact with the ice bath.

21. The following parameters are programmed into the Narishige IM-300 microinjector:

Injection pressure:	20 psi
Balance:	2.2 psi
Hold:	14 psi
Clear:	0.20 sec
Clear hold:	0.30 sec
Injection time:	0.08 sec

22. To prepare an injection slide, place two 150-µl drops of M2 medium side-by-side on a Lab-Tek chamber slide, flatten these drops into discs with a pipette tip to minimize their height, and overlay the flattened drops with ~1 ml of mineral oil. Maintain the temperature at 37°C with the heating glass.

23. Transfer zygotes to the injection slide. All zygotes must be injected within 10 min, so the number of zygotes taken per batch depends on efficiency of the injector.

    We usually inject about 25 to 30 zygotes per batch. A beginner might start with as few as 4 to 6 zygotes per batch, and the most experienced technician can inject as many as 50 zygotes in 10 min.

24. Check the general morphology of the zygotes under the microscope for the presence of zona pellucida, pronuclei, and both polar bodies. Discard embryos that contain more than two pronuclei.

25. Prior to injection, make sure that the needle is open by placing the tip of the injection needle next to an embryo.

    a. Press the "clear" button on the injector.
    b. If the embryo rotates freely, the needle flow is free of obstruction.
    c. If the embryo does not move, gently remove the tip of the injection needle using a scraping motion against the holding pipette. Check the needle again for flow rate. The needle should be discarded if the embryo moves too much, but can be used if the embryo rotates freely. If the needle opening is still obstructed, try breaking off more of the tip.

26. Using the holding micropipet, position the first zygote and fix by applying negative pressure.

27. Align the embryo and the microinjection needle so that both the opening of the needle and the pronucleus of the embryo are in the same focal plane.

28. Microinject the zygotes, observing the following:

    a. Maintain positive pressure on the injection needle at all times.
    b. Penetrate the zona pellucida and oolemma with the injection needle. Move forward into the closest pronucleus.
    c. A slight swelling of the pronucleus may be seen once the plasma membrane is penetrated. Otherwise, press the injection foot pedal to observe a slight swelling of the pronucleus.
    d. Retract the tip of the needle to the cytoplasm and inject another volume of microinjection solution. Carefully withdraw the capillary from the zygote if it hasn't already been removed by the force of the cytoplasm injection.

29. Repeat with the remaining zygotes.

30. Use the mouth pipetting apparatus to collect the injected zygotes and transfer them to the Embryo Transfer dish.

31. Remove lysed zygotes.

32. Incubate surviving zygotes at 37°C in KSOM until embryo transfer (usually within the next 1 to 2 hr).

Every new batch of microinjection solution should be assessed for toxicity by culturing about 30 injected zygotes overnight. In a successful injection session, 90% to 95% of zygotes should progress to the two-cell stage. If the solution batch is toxic, extensive lysis may be visible within an hour post injection.

Transfer injected embryos into pseudo-pregnant mice

33. Transfer injected embryos into pseudo-pregnant female mice.

    a. Obtain pseudo-pregnant mice by mating 8- to 12-week old CD1 females to vasectomized CD1 males on the day before microinjection between hours 12:00 and 16:00.
    b. On the morning of the injection, identify plug-positive (i.e., pseudo-pregnant) females for oviduct transfers.

    Typically, 10 to 20 CD1 females should be mated to obtain 4 to 8 plugged females.

    c. Transfer 15 to 25 viable manipulated embryos into the oviducts of each pseudo-pregnant foster mother following established surgical procedures (Behringer et al., 2014).

    The optimal number of embryos transferred is 18 total per female (9 per side).

PRODUCTION OF FERTILIZED EGGS THROUGH IN VITRO FERTILIZATION (IVF)

SUPPORT PROTOCOL

As an alternative to dissecting fertilized zygotes from mated mice (step 2 to 16, Basic Protocol 3), zygotes may be produced by in vitro fertilization using sperm from homozygous PITT seed mice as described here. IVF is advantageous because it reduces the number of stud mice needed and provides more scheduling flexibility for the researcher. We have also found that IVF produces a large number of synchronized and high quality embryos for microinjection.

The hormone treatment in this IVF protocol is slightly different from the natural mating protocol provided above (e.g., concentration, dosing, and timing of hormone administration), because the techniques are performed in different laboratories (natural mating: from CB Gurumurthy's laboratory in the United States; IVF: from Masato Ohtsuka's laboratory in Japan). Although slightly different, these protocols seem to work most optimally based on the hormones, animal sources, and housing conditions in these labs.

Materials *(also see Basic Protocol 3)*

Mice:

Stud males: homozygous seed mice, 3- to 6-months old (see Table 15.10.3)

Egg donors: wild type C57BL/6 female mice procured at 7-weeks old (CLEA Japan, Inc., Tokyo, Japan).

Homozygous seed female mice (bred in-house), instead of wild type females, can also be used. However, we have found that hemizygous embryos survive better than homozygous embryos after microinjection.

PMSG (ASKA Animal Health Co., Ltd., Tokyo, Japan), 1000 IU/ampule, 10 ampules

hCG (ASKA Animal Health Co., Ltd., Tokyo, Japan), 3000 IU/ampule, 10 ampules

NOTE: *The PMSG and hCG hormones from Basic Protocol 3 should be compatible with Support Protocol.*

Reconstitute lyophilized PMSG and hCG in saline (0.9% sodium chloride solution; Otsuka Normal Saline, Otsuka Pharmaceutical Factory, Inc.) to a final concentration of 7.5 IU/0.2 ml and store at −80°C until use.

HTF medium (Human Tubal Fluid; ARK Resource, Kumamoto, Japan; http://www.ark-resource.co.jp/english/products/pr_vitrification)
1× EmbryoMax M2 medium with phenol red (Millipore, cat. no. MR-015-D), for embryo handling and microinjection
Liquid paraffin (Nacalai Tesque, cat. no. 26137-85)

35 × 10-mm Falcon tissue culture dish (Corning, cat. no. 353001)
1-ml tuberculin syringe with 26G × 0.5-in needle (TERUMO, cat. no. SS-01T2613S)
Dissecting microscope (e.g., Olympus SZ11) equipped with a hot plate (e.g., KM-1, Kitazato)
Heraeus HERAcell 150i Tri–gas incubator, equipped with Coda Inline filters

IMPORTANT: Be sure all animal methods are approved by your institutional animal care and use committee (IACUC).

Stimulate ovulation

1. Intraperitoneally inject 7.5 IU of PMSG (in a 0.2 ml volume) into about 15 female mice at hour 18:00 on day 1.

   The later injection time for IVF (compared to the earlier hormone treatment for natural mating) allows for the microinjection and surgery procedures to fall at a reasonable time (e.g., between 16:30 and 19:30) on the day of IVF.

2. After 48 hr (at hour 18:00 on day 3), intraperitoneally inject 7.5 IU of hCG (in a 0.2 ml volume) into the female mice.

Isolate and capacitate sperm in vitro

3. On the morning of day 4 (between hours 8:00 and 8:30), about 30 min before egg collection, dissect the cauda epididymides from a stud homozygous male seed mouse, as previously described (Takahashi and Liu, 2010)

   Instead of using the freshly-isolated epididymides as a source of sperm, cryopreserved sperm samples may be used. We have observed, however, that using cryopreserved sperm leads to poor fertilization rates and poor quality embryos.

4. While grasping each cauda epididymis with forceps, make a small incision using a 26-gauge needle. Gently press the forceps against the cauda epididymis to release the spermatozoa-containing tissue from the incision. Transfer to a 35-mm dish containing 300 μl warm (37°C) HTF medium.

5. Incubate the sperm in a 5% CO_2 incubator for 1 to 1.5 hr to allow sperm capacitation.

Isolate oocytes

6. During sperm capacitation, euthanize the super-ovulated females, surgically remove their oviducts, and place the oviducts in 35-mm dishes containing M2 medium.

7. Tease the ampulla of the oviduct with a 26-gauge needle to release the egg–cumulus cell complex into the M2 medium.

8. Transfer the egg–cumulus cell complex into 250 μl HTF in a 35-mm dish, and cover with liquid paraffin.

Perform in vitro fertilization

9. At hour 9:30 on day 4, add 10 μl cultured sperm to the HTF drop containing oocytes, confirm sperm motility using a dissecting microscope, and incubate in a 5% CO_2 incubator to allow IVF.

10. After 5 to 6 hr, determine the success of IVF by identifying fertilized oocytes—i.e., oocytes with two pronuclei.

With freshly isolated sperm, the fertilization rate should be exceed 80%.

11. Proceed with microinjection at hour 16:30 (Basic Protocol 3, step 17).

GENOTYPING OF OFFSPRING TO IDENTIFY TRANSGENIC FOUNDERS

BASIC PROTOCOL 4

Although performing a Southern blot is traditionally considered the gold standard for confirming successful targeted cassette insertions, PCR is reliable for identifying PITT founder mice. While the exact genotyping strategy depends on the PITT project, primer sets must be designed to amplify both landing pad-DOI junctions in order to identify the precise insertion site. Additional primer sets should be used to probe the DOI. A schematic of a standard PCR-based strategy is shown in Figure 15.10.6.

With the exception of the chosen primer set, all genotyping PCRs follow the same standard protocol steps. There are two major steps in this protocol: (a) mouse tail DNA extraction and (b) PCR amplification followed by agarose gel electrophoresis.

Materials *(also see Basic Protocol 1)*

Mice: potential transgenic founders (from Basic Protocol 3) and wild-type controls
Cell lysis solution (Qiagen, cat. no. 158908)
Proteinase K, 20 mg/ml (5 PRIME, cat. no. 2500150)
Protein precipitation solution (Qiagen, cat. no. 158912)
100% and 70% (v/v) ethanol (e.g., Decon Laboratories, cat. no. 07-678-005)
DNA hydration solution (Qiagen, cat. no. 158914)
PCR 2× master mix: e.g., GoTaq Hot Start Green Master Mix (Promega, cat. no. M5122)
Nuclease-free water (such as from Thermo Fisher Scientific, cat. no. BP561-1)

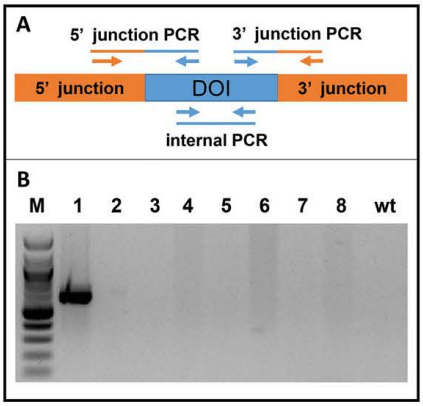


Figure 15.10.6 PCR-based genotyping of PITT transgenic founders. (**A**) Schematic of PCR primer sets typically used for genotyping: 5′ junction PCR, 3′ junction PCR, and internal PCR. Primer sets are designed to amplify 200 to 800 bp. (**B**) A sample agarose gel run with a 100-bp ladder (M) that demonstrates a transgenic founder identified by a 5′ junction PCR (lane 1) compared to other tested mice (2-8) and a wild-type control (wt).

Primer mix: combine equal volumes of forward and reverse primers from 100 pmol/μl stocks, resulting in a final mix containing 50 pmol/μl each primer. Use 1 μl primer mix in 100 μl of PCR master mix.

Heat block
Vortex mixer
Microcentrifuge
Thermocycler (BioRad T100 or equivalent)

Extract mouse tail DNA

1. Collect ~2- to 3-mm tail pieces from potential founder mice and wild-type controls in 1.5-ml microcentrifuge tubes.
2. Add 300 μl cell lysis solution containing 3 μl proteinase K, and incubate at 65°C overnight.

   To save time and reagent loss, make a master mix of lysis solution and proteinase K. This can then be distributed to 300-μl aliquots in separate microcentrifuge tubes.
3. Cool samples to room temperature, add 100 μl protein precipitation solution, and vortex thoroughly for ~20 sec.
4. Place the tubes on ice for 2 to 3 min, and then pellet insoluble material by centrifugation for 2 to 4 min at 17,000 × *g*, room temperature.
5. Transfer supernatants to new tubes containing 800 μl of 100% ethanol, and mix by inverting the tubes 8 to 10 times.
6. Pellet precipitated DNA by centrifugation for 2 to 4 min at 17,000 × *g*, room temperature.
7. Discard the supernatant, add 800 μl of 70% ethanol, and mix by inverting the tubes 8 to 10 times.
8. Pellet washed DNA by centrifugation for 2 to 4 min at 17,000 × *g*, room temperature.
9. Discard the supernatant, and collect the pellet and residual ethanol in the bottom of the tube by centrifugation for 1 min at 17,000 × *g*, room temperature.
10. Manually aspirate the remaining 70% ethanol using a 200-μl pipette tip, and air dry the DNA pellet for ~5 min (do not exceed 8 min).

    It is important to change tips between samples to avoid cross contamination.
11. Add 100 μl DNA hydration solution to the pellet, mix by flicking the side of the tube, and incubate the tubes for 15 to 30 min at 65°C to solubilize the DNA.

Genotype mice by PCR amplification of PITT landing pads and DOIs

Perform PCR for all primer sets on potential transgenic founder mice and wild type control DNA samples (see Figure 15.10.6). Use the PCR 2× master mix manufacturer's protocol to determine the exact parameters for each amplicon.

12. Prepare the PCR reactions by combining the following in the appropriate PCR tubes:

    1 μl mouse tail DNA
    7.5 μl 2× GoTaq Hot Start Green Master Mix (1× final)
    0.15 μl primer mix (7.5 pmol each primer)
    6.35 μl nuclease-free water

    For each primer set, prepare enough reaction mix (without mouse tail DNA) to analyze each sample.

13. Run the PCR reactions in a thermocycler.

1 cycle:	95°C	2 min
35 cycles:	95°C	30 sec
	55°C	30 sec
	72°C	1 min
Final step:	4°C	(hold)

The annealing temperature may vary slightly depending on the primer set.

14. Analyze PCR products by agarose gel electrophoresis to identify transgenic founders.

Example genotyping results are provided in Figure 15.10.6.

Transgenic founder mice should yield the correct size amplicons for all assayed primer sets. Wild-type control DNA samples should not yield any amplicons.

REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes and protocol steps. For common stock solutions, see APPENDIX 2D.

Microinjection buffer

5 mM Tris·Cl, pH 7.4 (Sigma Aldrich, cat. no. T2663)
0.1 mM EDTA (Sigma Aldrich, cat. no. E7889)
Sterile-filtered water (Sigma Aldrich, cat. no. W1503)
Sterilize using an Ultrafree-MC filter, 0.22-μm pore size (Millipore, cat. no. UFC30GV00)

COMMENTARY

Background Information

Historical perspectives on random and targeted transgenic technologies

Transgenesis has revolutionized many fields of biology—including genetics, gene regulation, medicine, and bioengineering—since its inception in the early 1980s (Jones, 2011). The initial successes in developing this technology were reported by Drs. Jon Gordon and Frank Ruddle, who performed the first genetic transformation by injecting DNA into mouse pronuclei, resulting in the integration and stable germ line transmission of the exogenous DNA (Gordon et al., 1980; Gordon and Ruddle, 1981). However, this technology suffered from the random nature of exogenous DNA integration, including variation in copy number and chromosome position effects at the site of integration, which led to inconsistent gene expression (Hogan, 1983). To overcome this challenge, the next generation of transgenesis techniques involved insertion of transgenes into embryonic stem cells at a specific site in the genome and subsequent injection of targeted ES cells into blastocyst-stage embryos (Gossler et al., 1986). The steps required to generate transgenic embryonic stem cells—electroporation/transfection, selection, and screening—and an additional generation of breeding to ensure germline transmission in the chimeric founders made the ES cell-based approach laborious and expensive.

We have adopted and combined the best features of traditional transgenesis and the ES cell-based approaches: injection of transgenic DNA into a pronucleus of the one-cell zygote and targeted integration to a specific genomic site. Our laboratory first developed a seed mouse containing a special landing pad in its genome that would enable efficient and targeted insertion of an injected DOI sequence in the presence of a recombination enzyme (Ohtsuka et al., 2010). We systematically tested and characterized a set of heterotypic *LoxP* sites to direct a DOI to a specific genomic locus using a non-reversible Cre recombination-mediated cassette exchange mechanism. We described this technique, including the ideal landing pad and compatible DNA elements that would be suitable for such targeted transgenesis, in 2010 (Ohtsuka et al., 2010). Soon after, another group adapted pronuclear injection-based targeted transgenesis to be used with *attB/P* sites and PhiC31 integrase (Tasic et al., 2011). Our lab subsequently improved the

PITT tools, including the development of a multiplexed PITT system that constitutes Cre-*LoxP*, PhiC31-*attB/P*, and Flp-*FRT* systems (Ohtsuka et al., 2015).

Critical Parameters

Here we address critical parameters of the microinjection solution and the seed mouse to maximize PITT efficiency.

Microinjection solution

PITT insertion efficiency and the viability of injected embryos are highly sensitive to the quality and concentration of nucleic acids in the microinjection solution.

Nucleic acid quality

We recommend filtering the nucleic acid solution with the Ultrafree-MC filter prior to preparing the injection mixture to prevent potential clogging of the injection needle. The concentration should be quantified after this step to account for decreased yield post filtration. It is also important to ensure high integrity of the synthesized mRNA. We recommend evaluating mRNA quality by gel electrophoresis both after mRNA synthesis and after injection mixture preparation. It is important to check mRNA quality after mRNA synthesis, because the synthesized mRNA solution contains variable amounts of incompletely synthesized RNAs. As a result, the concentration of the mRNA solution does not fully reflect the number of intact mRNA molecules within a given sample. Incomplete mRNA synthesis can be detected as a smear in the gel. Given that there may be trace RNase contamination in the donor vector solution, the microinjection solution should also be evaluated for mRNA integrity after it has been stored for 24 hr at room temperature or 4°C.

Nucleic acid concentration

Success of PITT is highly dependent on the concentration of nucleic acids. Although we have described the optimal concentrations of mRNA and donor vector for our own system, the ideal concentration should be optimized for each unique system. This is especially important because there is only a narrow range of recombinase/integrase concentrations that will maximize insertion efficiency while preserving embryo survival. Concentrations may vary because the methods of calculating nucleic acid concentration differ depending on the lab preference (NanoDrop, PicoGreen, UV absorbance, etc). In addition, regardless of the concentration of mRNA, the amount of recombinase/integrase translated in injected zygotes depends on the vector used for mRNA synthesis (e.g., Cre vs. iCre, PhiC31 vs. PhiC31o, or vector containing a polyA sequence vs. no polyA sequence and subsequent polyadenylation by the mMESSAGE mMACHINE T7 ULTRA Kit).

To determine the optimal concentration of mRNA in your system, inject pronuclei with microinjection solutions that test a range of mRNA concentrations. Develop these pronuclei in vitro and check the survival rate at the blastocyst stage (see Tables S3 and S4 in Ohtsuka et al., 2010, and Table S1 in Ohtsuka et al., 2015). Select the maximum concentration before blastocyst viability is reduced by more than 40% to 60%. This should be performed for each new batch of mRNA as well as for each unique microinjectionist to account for person-to-person technical variability. Given the investment in time and reagents to perform this optimization, we recommend preparing large batches of mRNA so that one optimization can be performed for many experiments.

While we have observed that increasing sizes of the DOI decreases insertion efficiency, we have not empirically defined the acceptable range for donor vector size. To date, we have performed PITT with donor vectors up to 14.4 kb in size (Ohtsuka et al., 2013).

Seed mouse

While we have generated many seed mice that have been used successfully for PITT, we have observed that some of these strains are less efficient at integrating exogenous DOI than others. Seed mouse parameters that may influence efficiency include 1) landing pad locus, 2) recombinase/integrase system, and 3) genetic background.

Landing pad locus

Rosa26 is the most frequently used genomic locus for targeted donor vector insertion because it is well established as an open chromatin region conducive to both efficient DOI integration and expression of inserted genes. However, several other loci can serve as landing pads including *H11*, *H2-Tw3*, *Hprt*, *TIGRE*, *Actb* and *AAVS1*.

Recombinase/integrase system

Although there are three available PITT platforms (Cre, PhiC31, and Flp), only Cre and PhiC31 systems have been proven to work in vivo. The Flp recombination system has been successfully used in vitro, but not yet in mouse embryos. In our experience,

Cre and PhiC31 are equally effective at mediating targeted transgenesis. Both systems show comparable insertion efficiencies from embryonic stem cell studies (Ohtsuka et al., 2015). Furthermore, we have successfully used the Cre and PhiC31 platforms together in our TOKMO-3 seed mouse, which increases the targeted insertion efficiency (Ohtsuka et al., 2015). To use this *i*mproved PITT (*i*-PITT) strategy, mRNAs from both platforms should be added to the microinjection solution for simultaneous translation of Cre and PhiC31 in injected fertilized eggs.

Seed mouse genetic background

It has been reported that mouse genetic background influences traditional transgenic rates. For example, FVB/N mice are known to be highly susceptible to transgenesis whereas C57BL/6 strains have lower integration efficiencies (Auerbach et al., 2003). While we have not empirically tested the influence of genetic background on PITT, we predict that strain differences may influence insertion efficiency, rate of successful transplantation, and embryo survival. The genetic backgrounds of our seed mice are reported in Table 15.10.3.

Troubleshooting

If PITT mice are not obtained for a specific project, the following factors should be checked.

1. Confirm that the donor vector, recombinase/integrase mRNA, and seed mouse are compatible. Ensure that the PITT elements in the DOI, recombinase/integrase of choice, and docking sites in the landing pad are all a part of the same Cre-PITT or PhiC31-PITT platform. To determine reagent compatibility for a given platform, please refer to our suggestions in Tables 15.10.2 and 15.10.3.

2. Ensure that you have evaluated the quality and concentration of nucleic acids in your microinjection solution. Our suggestions for assessment and optimization may be found in the critical parameters section of this protocol.

3. Consider the possibility that expression of the transgene of interest may affect embryo viability. If embryo health is influenced by expression of a specific transgene, engineer an inducible construct for spatial or temporal control of expression.

If the above factors are satisfactory, then we recommend repeating the microinjection with more (150 to 200) zygotes. Alternatively, instead of transferring injected fertilized eggs to pseudo-pregnant surrogate females at the two-cell stage, eggs may be cultured in vitro until they develop into blastocysts, which can then be genotyped to screen for targeted insertion prior to transplantation. It is also possible to increase the concentration of nucleic acids in the injection solution or inject a greater volume of the solution if embryo viability has not been compromised at the current concentration (determined by if more than 60% of injected zygotes develop into blastocysts and/or more than 15% of transplanted eggs survive to birth).

Anticipated Results

PITT is used to generate targeted transgenic founder lines for a DOI of interest. While the outcome for every PITT project is unique depending on the selected DOI, seed mouse, and recombinase/integrase platform, here we provide the general range of anticipated results for targeted integration and gene expression of the donor cassette.

Targeted integration of donor cassette

Targeted integration efficiency depends on the PITT system used. We have observed efficiencies ranging from 1.9% to 62.0% (Ohtsuka et al., 2010, 2012b, 2015). Efficiency will likely be greater than 10% in PITT projects that use recombinase/integrase mRNA injection instead of plasmid DNA encoding the recombinase/integrase and in projects that simultaneously employ both Cre and PhiC31 systems (Ohtsuka et al., 2013, 2015).

While random insertions could presumably occur using this technology, this only happens rarely compared to traditional transgenesis that uses a linear DNA fragment (~2.4% vs. 10% to 20%; Fielder et al., 2010; Ohtsuka et al., 2010). Excluding a minor fraction of cases, almost all PITT mice will have the integrated allele at the expected locus as a result of the proper recombination/integration events. Insertions are usually clean, without unintended insertions or deletions of genomic DNA. However, it is important to note that several insertion alleles can be anticipated when simultaneously using several platforms, such as Cre and PhiC31 in *i*-PITT (Ohtsuka et al., 2015).

Insertion of vector sequences at the landing pad

In some currently available PITT designs, insertion of the DOI leaves trace vector backbone sequences near PITT landing pads. These extra vector sequences include prokaryote-derived elements that can inhibit stable transgene expression. To ameliorate this problem,

many PITT platforms include *FRT* elements in the donor vectors that allow for genetic removal of these excess sequences by crossing PITT Tg founder mice with transgenic mice carrying a ubiquitously expressed Flp recombinase transgene (Ohtsuka et al., 2010, 2015). An example of a Flp transgenic mouse from JAX Mice that can be used for such purposes is *B6.Cg-Tg(Pgk1-FLPo)10Sykr/J.*

Gene expression of donor cassette

Unlike traditional transgenesis, each transgenic founder generated from a given PITT project will express the inserted transgene to similar levels. This is because the transgene resides in a predetermined locus with a single-copy configuration, which allows for consistent expression between transgenic siblings and from generation to generation.

Although a reliable transgene expression pattern can be obtained in PITT mice, the magnitude of cassette gene expression will vary between PITT projects based upon choice of locus, promoter, cDNA, and inclusion of additional sequence elements (e.g., multiple expression cassettes, microRNA sequences, internal ribosome entry sites, and polyadenylation signals). PITT can be used for ubiquitous expression, as well as tissue-specific expression when the DOI contains a tissue-specific promoter and is integrated into the *Rosa26* locus (Tsuchida et al., 2016).

The one factor that may hinder reliable gene expression is the presence of proximal prokaryote-derived sequences from the donor vector backbone, which have been integrated as excess cargo in the landing pad. We have shown that the removal of this excess donor vector sequence ensures stable inheritance of cassette gene expression (Ohtsuka et al., 2010).

In the unforeseen case that a transgene does not exhibit the expected expression pattern, it is possible that unannotated regulatory sequences reside in the DOI, such as enhancers or silencers. Such unknown DNA interactions can be overcome by incorporating insulator sequences at both sides of the DOI (Madisen et al., 2015).

Time Considerations

Targeted transgenesis experiments performed through ES cell-based approaches require at least one year to generate chimeric mice, which must be bred to ensure germline transmission of the transgene. On the other hand, PITT takes about 3 to 4 months to generate a founder mouse line, which invariably transmits the transgene to offspring. A typical time frame for the PITT experimental procedures is outlined below and depicted in Figure 15.10.2:

Weeks 1 to 4: Design and build PITT donor DNA constructs

Weeks 3 to 6: Synthesize and purify DNA and RNA components for microinjection

Weeks 4 to 6: Prepare injection components and initiate superovulation

Weeks 7 to 9: Isolate embryos from seed mice, microinject PITT components, and transfer embryos into pseudo-pregnant mice

Weeks 13 to 15: Genotype offspring to identify transgenic founders

Acknowledgments

S.L.P.S. was supported by the NSF Graduate Research Fellowship DGE1144152. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. M.O. acknowledges the staff from the Support Center for Medical Research and Education and Tokai University for microinjection and mRNA synthesis. We thank Hiromi Miura for technical assistance in preparation of mRNA, donor vector construction, and seed mouse development. M.O.'s lab was supported by Grant-in-Aid for Young Scientists (B) (23700514), Grant-in-Aid for Scientific Research (25290035 and 16H04685) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), and by MEXT-Supported Program for the Strategic Research Foundation at Private Universities, 2009-2013. This work was partially supported by an Institutional Development Award (IDeA) to C.B.G. (PI: Shelley Smith) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103471.

Literature Cited

Auerbach, A.B., Norinsky, R., Ho, W., Losos, K., Guo, Q., Chatterjee, S., and Joyner, A.L. 2003. Strain-dependent differences in the efficiency of transgenic mouse production. *Transgenic Res.* 12:59-69. doi: 10.1023/A:1022166921766.

Behringer, R., Gertsenstein, M., Nagy, K.V., and Nagy, A. 2014. Manipulating the mouse embryo: A laboratory manual. Fourth edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Donovan, J., Brown, P., Reeves, J. and Reeves, P. 2003. Introduction to basic mouse handling techniques. *Curr. Protoc. Hum. Genet.* 36:A.3L.1–A.3L.18. doi: 10.1002/0471142905.hga03ls36.

Fielder, T.J., Barrios, L., and Montoliu, L. 2010. A survey to establish performance standards for the production of transgenic mice. *Transgenic Res.* 19:675-681. doi: 10.1007/s11248-009-9335-3.

Gallagher, S. R. and Desjardins, P. R. 2007. Quantitation of DNA and RNA with absorption and fluorescence spectroscopy. *Curr. Protoc. Hum. Genet.* 53:A.3D.1–A.3D.21. doi: 10.1002/0471142905.hga03ds53.

Gordon, J.W. and Ruddle, F.H. 1981. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* 214:1244-1246. doi: 10.1126/science.6272397.

Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A., and Ruddle, F.H. 1980. Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc. Natl. Acad. Sci. U.S.A.* 77:7380-7384. doi: 10.1073/pnas.77.12.7380.

Gossler, A., Doetschman, T., Korn, R., Serfling, E., and Kemler, R. 1986. Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proc. Natl. Acad. Sci. U.S.A.* 83:9065-9069. doi: 10.1073/pnas.83.23.9065.

Gurumurthy, C.B., Takahashi, G., Wada, K., Miura, H., Sato, M., and Ohtsuka, M. 2016. GONAD: A novel CRISPR/Cas9 genome editing method that does not require ex vivo handling of embryos. *Curr. Protoc. Hum. Genet.* 88:15.8.1-15.8.12. doi.wiley.com/10.1002/0471142905.hg1508s88.

Harms, D.W., Quadros, R.M., Seruggia, D., Ohtsuka, M., Takahashi, G., Montoliu, L., and Gurumurthy, C.B. 2014. Mouse genome editing using the CRISPR/Cas system. *Curr. Protoc. Hum. Genet.* 83:15.7.1-15.7.27. doi: 10.1002/0471142905.hg1507s83.

Hogan, B. 1983. Molecular biology. Enhancers, chromosome position effects, and transgenic mice. *Nature* 306:313-314. doi: 10.1038/306313a0.

Jarcho, J. 2001. Restriction fragment length polymorphism analysis. *Curr. Protoc. Hum. Genet.* 1:2.7.1-2.7.15. doi: 10.1002/0471142905.hg0207s01.

Jones, D. 2011. Genetic engineering of a mouse: Dr. Frank Ruddle and somatic cell genetics. *Yale J. Biol. Med.* 84:117-124.

Kramer, M.F. and Coen, D.M. 2001. Enzymatic amplification of DNA by PCR: Standard procedures and optimization. *Curr. Protoc. Mol. Biol.* 56:15.1.1-15.1.14.

Madisen, L., Garner, A.R., Shimaoka, D., Chuong, A.S., Klapoetke, N.C., Li, L., van der Bourg, A., Niino, Y., Egolf, L., Monetti, C., Gu, H., Mills, M., Cheng, A., Tasic, B., Nguyen, T.N., Sunkin, S.M., Benucci, A., Nagy, A., Miyawaki, A., Helmchen, F., Empson, R.M., Knöpfel, T., Boyden, E.S., Reid, R.C., Carandini, M., and Zeng, H. 2015. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* 85:942-958. doi: 10.1016/j.neuron.2015.02.022.

Miura, H., Inoko, H., Tanaka, M., Nakaoka, H., Kimura, M., Gurumurthy, C.B., Sato, M., and Ohtsuka, M. 2015. Assessment of artificial miRNA architectures for higher knockdown efficiencies without the undesired effects in mice. *PloS ONE* 10:e0135919. doi: 10.1371/journal.pone.0135919.

Ohtsuka, M. 2014. Development of pronuclear injection-based targeted transgenesis in mice through Cre-loxP site-specific recombination. *Methods Mol. Biol.* 1194:3-19. doi: 10.1007/978-1-4939-1215-5_1.

Ohtsuka, M., Miura, H., Nakaoka, H., Kimura, M., Sato, M., and Inoko, H. 2012a. Targeted transgenesis through pronuclear injection of improved vectors into in vitro fertilized eggs. *Transgenic Res* 21:225-226. doi: 10.1007/s11248-011-9505-y.

Ohtsuka, M., Miura, H., Sato, M., Kimura, M., Inoko, H., and Gurumurthy, C.B. 2012b. PITT: Pronuclear injection-based targeted transgenesis, a reliable transgene expression method in mice. *Exp. Anim.* 61:489-502. doi: 10.1538/expanim.61.489.

Ohtsuka, M., Miura, H., Hayashi, H., Nakaoka, H., Kimura, M., Sato, M., Gurumurthy, C.B., and Inoko, H. 2013. Improvement of pronuclear injection-based targeted transgenesis (PITT) by iCre mRNA-mediated site-specific recombination. *Transgenic Res.* 22:873-875. doi: 10.1007/s11248-013-9703-x.

Ohtsuka, M., Ogiwara, S., Miura, H., Mizutani, A., Warita, T., Sato, M., Imai, K., Hozumi, K., Sato, T., Tanaka, M., Kimura, M., and Inoko, H. 2010. Pronuclear injection-based mouse targeted transgenesis for reproducible and highly efficient transgene expression. *Nucleic Acids Res.* 38:e198. doi: 10.1093/nar/gkq860.

Ohtsuka, M., Miura, H., Mochida, K., Hirose, M., Hasegawa, A., Ogura, A., Mizutani, R., Kimura, M., Isotani, A., Ikawa, M., Sato, M., and Gurumurthy, C.B. 2015. One-step generation of multiple transgenic mouse lines using an improved Pronuclear Injection-based Targeted Transgenesis (i-PITT). *BMC Genomics* 16:274. doi: 10.1186/s12864-015-1432-5.

Pease, S. and Saunders, T.L. (eds.). 2011. Advanced protocols for animal transgenesis. Springer-Verlag, Berlin.

Quadros, R.M., Harms, D.W., Ohtsuka, M., and Gurumurthy, C.B. 2015. Insertion of sequences at the original provirus integration site of mouse ROSA26 locus using the CRISPR/Cas9 system. *FEBS Open Bio.* 5:191-197. doi: 10.1016/j.fob.2015.03.003.

Takahashi, H. and Liu, C. 2010. Archiving and distributing mouse lines by sperm cryopreservation, ivf, and embryo transfer. *Methods Enzymol.* 476:53-69. doi: 10.1016/S0076-6879(10)76004-3.

Tasic, B., Hippenmeyer, S., Wang, C., Gamboa, M., Zong, H., Chen-Tsai, Y., and Luo, L. 2011. Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proc.*

Natl. Acad. Sci. U.S.A. 108:7902-7907. doi: 10.1073/pnas.1019507108.

Tsuchida, J., Matsusaka, T., Ohtsuka, M., Miura, H., Okuno, Y., Asanuma, K., Nakagawa, T., Yanagita, M., and Mori, K. 2016. Establishment of nephrin reporter mice and its use for chemical screening. *PLoS ONE* 11:e0157497. doi: 10.1371/journal.pone.0157497.

Key Reference

Ohtsuka, M. 2014. Development of pronuclear injection-based targeted transgenesis in mice through Cre-loxP site-specific recombination. *Methods Mol. Biol.* 1194:3-19. doi: 10.1007/978-1-4939-1215-5_1.

Some aspects of this protocol were adapted from this reference.

APPENDIX C

J Genet Counsel (2017) 26:276–278
DOI 10.1007/s10897-016-0046-7

ORIGINAL RESEARCH

My Identical Twin Sequenced our Genome

Samantha L.P. Schilit[1] · **Arielle Schilit Nitenson**[2]

Received: 19 February 2016 / Accepted: 8 November 2016 / Published online: 16 November 2016

Abstract With rapidly declining costs, whole genome sequencing is becoming feasible for widespread use. Although cost-effectiveness is driving increased use of the technology, comprehensive recommendations on how to handle ethical dilemmas have yet to reach a consensus. In this article, Sam shares her experience of undergoing whole genome sequencing. Despite the deeply private nature of the test, the results do not solely belong to Sam; her identical twin sister, Arielle, shares virtually the same genome and received results without a formal consent process. This article explores their parallel experiences as a way of highlighting the controversial ethics of a private test with familial implications.

Keywords Whole genome sequencing · Identical twins · Genetic testing · Privacy · Informed consent · Bioethics

"Who in this room would like to have their whole genome sequenced?" The excitement in the genetics seminar was palpable. "Now you can get sequenced and a clinical interpretation for only $3100," the medical geneticist continued. With increasing affordability, clinical whole genome sequencing is now accessible to ostensibly healthy individuals (Christensen et al. 2015; Jackson 2015). For the researchers in the audience, getting the opportunity to discover the secrets hiding within their cells would be as fortuitous as Charlie winning the golden ticket to Willy Wonka's chocolate factory.

✉ Samantha L.P. Schilit
Samanthaschilit@fas.harvard.edu

[1] Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

[2] Department of Neuroscience, Brown University, 185 Meeting Street, Providence, RI 02912, USA

Whole genome sequencing can be desired for many reasons including altruism towards scientific research, curiosity, and financial or emotional preparation in the case of suspected disease alleles (Sanderson et al. 2016). For Sam, a genetics PhD candidate, the idea of interacting with her own genome in the same way she had studied the genomes of research subjects would be a dream come true. Plus, with a strong family history of early onset cancer, she wanted to know if she carried a mutation in a cancer predisposition gene.

While genetic tests always influence biological family members to some extent, one family member in particular would be impacted dramatically by these results. Arielle, her identical twin sister, shares virtually all three billion base pairs that make up their genome.

Arielle and Sam were not novices to the world of genetic testing. In fact, it was a genetic test that revealed their zygosity at age 14. The test was a part of a research study at the annual Twins Days Festival in Twinsburg, Ohio. As the largest gathering of twins in the world, the Twins Days Festival is a goldmine for researchers who study heredity; they flock from around the world, exchanging gift cards and candy for a cheek swab and a completed questionnaire. A few months after participating in one such study, Arielle and Sam received a letter revealing their monozygosity (Morell et al. 2007). The results were confirmed by two subsequent tests over the next five years. The discovery changed Sam's life, igniting a passion for epigenetics and clinical genetics that inspired her to pursue a career in genetics research.

The first time Sam told Arielle about her desire to get sequenced was as a campaign, requesting that Arielle split the sequencing costs. After all, Arielle was also a scientist, and Sam believed that both sisters would be interested in learning about their shared genome. Instead, Arielle apprehensively wondered if Sam had fully considered the consequences of the test. What if the results found pathogenic variants that were not

medically actionable? She worried about the psychological impact this news may bring to them. As a former researcher at the National Human Genome Research Institute (NHGRI), Arielle was also aware of the risks associated with securing genomic information. Despite coming from a clinical test, would the results be private or used for research purposes? Finally, Arielle felt the test was too novel and would not provide Sam with the information she so greatly desired. She thought Sam should wait until technology improved, such as the optimization of long-range sequencing, to make the most out of this opportunity.

Arielle's trepidation was justified. As a clinical test, Sam's sequencing results were required to go into her medical record. Congress passed the Genetic Information Nondiscrimination Act (GINA) in 2008, but its scope is limited (Genetic Information Nondiscrimination Act of 2008). While GINA prevents genetic discrimination for employment and health insurance, it does not cover life, disability or long-term care insurance (Green et al. 2015). In addition, there are indeed significant limitations to today's predominant genome sequencing technology. In current clinically approved next-generation sequencing methods, repetitive regions of the genome and structural variants such as deletions, duplications, and rearrangements are not detectable (Cordero and Ashley 2012; Green et al. 2013; Kircher and Kelso 2010).

Although Sam appreciated her sister's concerns about these limitations, she chose to continue her fundraising quest. After obtaining sponsorship, she booked an appointment with her primary care physician, as a physician's signature was required for ordering the test.

Sam's doctor highlighted the risks associated with the test as he conducted a rigorous informed consent discussion. "Are you aware that the test results may reveal information about yourself that you would rather not know, such as predispositions for diseases that might not be curable? Are you aware that you could be subject to genetic discrimination for life insurance or long-term disability insurance? Are you aware that if the security system is breached that the electronic delivery of your results could be accessed by someone else?" Although Sam believed the benefits of the results outweighed the risks, receiving the test was only contingent on her consent. As the physician was not a trained clinical geneticist or genetic counselor, the conversation never exceeded the informed consent form provided by the testing company. Moreover, because this was the physician's first time ordering whole genome sequencing for a patient, he had limited experience in referring patients to genetic counselors. Arielle was neither offered counseling nor asked to sign consent forms, even though most risks applied to her as well. Arielle's skepticism remained, as she felt divorced from the process; by not meeting with the physician, she was never debriefed on possible implications or invited to ask questions that would assuage her concerns.

While scientific societies and ethical boards have proposed informed consent guidelines for whole genome sequencing since its first application to clinical diagnosis in 2009, there is currently no consensus on what constitutes the requirements for informed consent of a genetic test (Jamal et al. 2013; Platt et al. 2014; Roche and Berg 2015). For example, one study that performed a rigorous examination of these recommendations revealed that only two-thirds discuss "possible disadvantages, risks, or complications" as part of the informed consent process, and only a single set of recommendations explicitly included a discussion of relatives as a part of the informed consent conversation (Ayuso et al. 2013; Sijmons et al. 2011). These policies are insufficient given the prevalence of this controversy; it has been estimated that almost a quarter of genetics health professionals have faced the ethical dilemma of whether to provide a genetic test to one identical twin who wishes to know his or her genetic status when the other twin disapproves (McLean et al. 2013). While monozygotic twinning occurs spontaneously in approximately 1 in 250 live births, this rate is anticipated to double in the wake of assisted reproductive technologies (Hankins and Saade 2005). The growing number of identical twins highlights the urgency with which policies must be established to protect this population.

Weeks passed by and Sam started to worry about how she would handle the discovery of a pathogenic variant in her genome. She feared that she would learn something so terrifying that she wouldn't want to bear the responsibility of telling her sister. As her anticipation heightened, she started to feel protective about the privacy of her results. Why did she feel obligated to share her private medical test results with Arielle? Most recommendations regarding returning genomic results to family members prioritize privacy and confidentiality over the desires of relatives (Wolf et al. 2015). However, the one case where there is currently no consensus is how to handle the discovery of highly pathogenic but medically actionable variants that could have serious implications for closely related family members likely to harbor the same mutation (Godard et al. 2006; Sijmons et al. 2011; Wolf et al. 2015). Arielle refused to accept Sam's resistance to sharing the results. She felt violated that Sam could withhold information that equally belonged to both of them, despite the original source of the tested blood. After many conversations, the sisters realized they needed to devise a compromise on their own, given that neither of their interests were fully protected by current ethical recommendations. Despite Sam's legal right to submit her sample for whole genome sequencing without Arielle's consent, both sisters agreed that Sam had the responsibility to share her sequencing results with Arielle, as any important findings could inform Arielle's future healthcare decisions.

A few months later, the results finally arrived. After Sam met with her physician to discuss the clinical and pharmacogenomics reports, she prepared to share the results with Arielle. "The things I'm going to tell you today are only scratching the surface of the information that's here," Sam explained. "There are limitations to the test, and there are even more limitations to

our ability to analyze the results. We have clinical interpretation for almost 2,000 disease-causing genes out of over 25,000 genes in the human genome. Everything else is up to us to explore, and future scientific research to discover."

Despite the wealth of information provided, Sam admitted to Arielle that she was disappointed by the results. She was overwhelmed by the amount of data given and yet frustrated that the majority of it was currently uninterpretable. Arielle comforted her, reminding Sam that while the clinical interpretation was not comprehensive, they could at least feel relief in knowing that no severe clinically significant findings were identified to date.

Whole genome sequencing provided the sisters with more questions than answers. Arielle and Sam don't carry any known cancer predisposition mutations out of the clinically interpreted genes, but would any mutations in the 23,000 genes that were not assessed one day be shown to cause a cancer predisposition? What is hiding in the unexplored depths of their genome? While they face many uncertainties in the aftermath of receiving whole genome sequencing results, one thing is certain. There is a pressing need to educate patients on how their genetic testing will affect family members and to provide resources for vulnerable relatives. Sam and Arielle were fortunate to have devised a compromise that worked for them, but not all identical twins will have these important conversations.

Acknowledgments The authors thank Medullan for their generous trainee sponsorship and Ms. Tammy Kammin, Ms. Susan Price, Mr. Adam Nitenson, Ms. Erica Ramos, Dr. Robert Green, and Dr. Robert Morell for their input. This work was supported by the NSF Graduate Research Fellowships DGE1144152 (SLPS) and DGE0228243 (ASN). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Compliance with Ethical Standards

Conflict of Interest Samantha L.P. Schilit and Arielle Schilit Nitenson declare that they have no conflict of interest.

Human Studies and Informed Consent This article does not contain any studies with human participants performed by any of the authors.

Animal Studies No animal studies were carried out by the authors for this article.

References

Ayuso, C., Millan, J. M., Mancheno, M., & Dal-Re, R. (2013). Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. *European Journal of Human Genetics, 21*(10), 1054–1059. doi:10.1038/ejhg.2012.297.

Christensen, K. D., Dukhovny, D., Siebert, U., & Green, R. C. (2015). Assessing the costs and cost-effectiveness of genomic sequencing. *Journal of Personalized Medicine, 5*(4), 470–486. doi:10.3390/jpm5040470.

Cordero, P., & Ashley, E. A. (2012). Whole-genome sequencing in personalized therapeutics. *Clinical Pharmacology and Therapeutics, 91*(6), 1001–1009. doi:10.1038/clpt.2012.51.

Genetic Information Nondiscrimination Act of 2008, Pub. L. 110-223, 122 Stat. 881, codified as amended in scattered sections of 26, 29, and 42 U.S.C. https://www.gpo.gov/fdsys/pkg/PLAW-110publ233/html/PLAW-110publ233.htm. Accessed date May 6, 2016.

Godard, B., Hurlimann, T., Letendre, M., Egalite, N., & INHERIT BRCAs (2006). Guidelines for disclosing genetic information to family members: from development to use. *Familial Cancer, 5*(1), 103–116. doi:10.1007/s10689-005-2581-5.

Green, R. C., Rehm, H. L., & Kohane, I. (2013). Clinical Genome Sequencing. In G. S. Ginsburg & H. F. Willard (Eds.), *Genomic and personalized medicine* (2nd ed., pp. 102–122). London and Waltham: Elsevier/Academic Press.

Green, R. C., Lautenbach, D., & McGuire, A. L. (2015). GINA, genetic discrimination, and genomic medicine. *The New England Journal of Medicine, 372*(5), 397–399. doi:10.1056/NEJMp1404776.

Hankins, G. V., & Saade, G. R. (2005). Factors influencing twins and zygosity. *Paediatric and Perinatal Epidemiology, 19*(Suppl 1), 8–9. doi:10.1111/j.1365-3016.2005.00609.x.

Jackson, M. (2015). Veritas genetics breaks $1,000 whole genome barrier [press release]

Jamal, S. M., Yu, J. H., Chong, J. X., Dent, K. M., Conta, J. H., Tabor, H. K., & Bamshad, M. J. (2013). Practices and policies of clinical exome sequencing providers: analysis and implications. *American Journal of Medical Genetics. Part A, 161A*(5), 935–950. doi:10.1002/ajmg.a.35942.

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing–concepts and limitations. *BioEssays, 32*(6), 524–536. doi:10.1002/bies.200900181.

McLean, N., Delatycki, M. B., Macciocca, I., & Duncan, R. E. (2013). Ethical dilemmas associated with genetic testing: which are most commonly seen and how are they managed? *Genetics in Medicine, 15*(5), 345–353. doi:10.1038/gim.2012.138.

Morell, R. J., Brewer, C. C., Ge, D., Snieder, H., Zalewski, C. K., King, K. A., Drayna, D., et al. (2007). A twin study of auditory processing indicates that dichotic listening ability is a strongly heritable trait. *Human Genetics, 122*(1), 103–111. doi:10.1007/s00439-007-0384-5.

Platt, J., Cox, R., & Enns, G. M. (2014). Points to consider in the clinical use of NGS panels for mitochondrial disease: an analysis of gene inclusion and consent forms. *Journal of Genetic Counseling, 23*(4), 594–603. doi:10.1007/s10897-013-9683-2.

Roche, M. I., & Berg, J. S. (2015). Incidental findings with genomic testing: implications for genetic counseling practice. *Current Genetic Medicine Reports, 3*(4), 166–176. doi:10.1007/s40142-015-0075-9.

Sanderson, S. C., Linderman, M. D., Suckiel, S. A., Diaz, G. A., Zinberg, R. E., Ferryman, K., Wasserstein, M., et al. (2016). Motivations, concerns and preferences of personal genome sequencing research participants: baseline findings from the HealthSeq project. *European Journal of Human Genetics, 24*(1), 153. doi:10.1038/ejhg.2015.179.

Sijmons, R. H., Van Langen, I. M., & Sijmons, J. G. (2011). A clinical perspective on ethical issues in genetic testing. *Accountability in Research, 18*(3), 148–162. doi:10.1080/08989621.2011.575033.

Wolf, S. M., Branum, R., Koenig, B. A., Petersen, G. M., Berry, S. A., Beskow, L. M., Daly, M. B., et al. (2015). Returning a research Participant's genomic results to relatives: analysis and recommendations. *The Journal of Law, Medicine & Ethics, 43*(3), 440–463. doi:10.1111/jlme.12288.