# Trans-ethnic Meta-analysis of Vitamin D Receptor Gene Variants in Context of Polygenic Risk, Associated with Difference in Incidence of Hypertension-Mediated Intracerebral Hemorrhage Across Ethnic Groups

# Trans-ethnic Meta-analysis of Vitamin D Receptor Gene Variants in Context of Polygenic Risk, Associated with Difference in Incidence of Hypertension-Mediated Intracerebral Hemorrhage Across Ethnic Groups

Jonathan Henry

A Thesis in the Field of Biology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

May 2021

## Abstract

Stroke is second in rates of mortality experienced globally (Katan & Luft, 2018). Defined by a bleed or blockage of a blood vessel in the brain, ischemic and hemorrhagic strokes, respectively, are the two primary types (Katan & Luft, 2018). Hemorrhagic stroke is responsible for the most mortality and disease burden. Post-Hemorrhagic stroke disability often includes declines in cognition, behavior and motor functions (Dajpratham et al., 2020). As with other diseases, stroke is twice as likely to occur in people of color, Black/African Americans especially. The same can be said for hypertension, which is also a prime risk factor for the incidence of hemorrhagic stroke (Girouard & Ladecola, 2006). Allocation of genomic profiles via genome-wide association studies (GWAS) have allowed for ethnicity-based assessment of hypertension-mediated intracerebral hemorrhage (ICH), as influenced by Vitamin D receptor (VDR) gene variants, a novel genetic constituent with believed connection to hypertension (Maciejewski et al., 2019). In this thesis, we investigated hemorrhagic stroke's genetic correlates was addressed with polygenic risk scores tailored to hypertension-mediated ICH as influenced by VDR gene variants. Though statistically insignificant, possibly a result of a lack of representation of black subjects, findings showed increased incidence of hypertension-mediated ICH in black subjects, non-lobar subtype of ICH, consistent with regions where systolic blood pressure polygenic risk scores would have more effect in. Future inquiry into VDR genes and other novel genetic considerations would be beneficial especially when aligned with addressing disparities in disease burden and mortality.

Acknowledgments

I would like to express my deepest gratitude in thanking Dr. Christopher D. Anderson, Dr. Evangelos Pavlos Myserlis, Christina Kourkoulis, Dr. Jonathan Rosand, and the rest of the RosAnderson Lab in the Center for Genomic Medicine at The Massachusetts General Hospital for their continued support and guidance throughout this process. I would also like to thank Dr. James Morris for his support, direction and assistance.

Table of Contents

List of Tables

# Chapter I.

## Introduction

Second only to heart disease, strokes are the second leading cause of death globally (The top 10 causes of death, 2020) (Katan & Luft, 2018). A stroke is defined by the manifestation of a bleed, defined as a hemorrhagic stroke, or blockage, defined as an ischemic stroke, in a cranial blood vessel (Katan & Luft, 2018). Responsible for supplying oxygen and nutrients which are both essential for proper neurological function, including breathing, behavior, muscular function and metabolic homeostasis, among many others, strokes have many negative resultant clinical manifestations. These manifestations are comprised of paralysis, speech and language difficulties, memory loss and changes to mood and behavior such as depression, forgetfulness, carelessness, irritability and confusion among others (Dajpratham et al., 2020). While ischemic strokes account for the majority of stroke occurrences, hemorrhagic strokes are responsible for the large majority of stroke-associated mortality and post-stroke disease burden (Johnson et al., 2019). There have been many research efforts and many more continue to investigate how to address these post-stroke effects, however, there remains much to be desired with respect to our understanding and approaches that are best suited to prevent, explain the pathophysiology, and remediate these outcomes.

Many of these research efforts are surrounded around the investigation of risk factors, defined as a variable that is associated with increasing the risk of the development of a disease (Burt, 2001), and mediators, such as comorbid factors,

medications, and genetics as conduits to possible predispositions for the incidence of a stroke and stroke recurrence, the innerworkings of the pathophysiology and influences of post-stroke recovery (Iadecola & Anrather, 2011). Risk factors are often separated into two categories; they are modifiable risk factors and the other is non-modifiable risk factors. Modifiable risk factors speak to those risk factors that can be altered or changed, while non-modifiable risk factors are those which traditionally cannot be changed. Diet, exercise, Body Mass Index, controlled blood pressure, smoke cessation and medication adherence are modifiable risk factors that have been widely used as points of interest in research studies looking to improve stroke prevention (Romero, Morris & Pikula, 2008). Age, sex, racial and ethnic background and genetics are focuses on the realm of non-modifiable risk factors for stroke research. With respect to stroke research, genetics offers insight into the architecture of patents' genomes as a perspective that gives possible explanations for increased incidence of stroke and prediction of outcomes (Markus, 2011). Putting aside the budding prospect of gene therapies and gene-altering technologies such as Clusters of Regularly Interspersed Short Palindromic Repeats (CRISPR), that are believed to eventually become regular approaches to address genetic components of disease as a means of therapy (Ledford, 2016), genetics is especially non-modifiable in the sense that there aren't any reliable approaches that address or account for genetics while considering stroke prevention and recover specifically at the genetic level. However, genetics have proven to offer insight into the truly heterogenous makeup that edifies the manifestation of stroke occurrence (Woo, 2004). Strokes are influenced by many risk factors, many of which are present in patients who have strokes. One of these factors is hypertension. Hypertension, defined as consistent elevated pressure level

(Girouard & Ladecola, 2006), is associated with many comorbid factors that increase the risk of the incidence of stroke (Johansson, 1999). With the underlying heterogeneity associated with stroke risk, along with hypertension as a risk factor that influences many other risk factors for stroke risk, such as heart disease and kidney disease, looking into how hypertension may play a role in the mediation of strokes, especially hemorrhagic stroke and more specifically intracerebral hemorrhage (ICH) given heightened mortality and disease burden associated with them, may offer understanding and discernment into the heterogenous manifestation of ICH.

An important consideration in research efforts of all disease, let alone the genetics of stroke, is that of social determinants of health. Defined as the socio-economic conditions that influence health-outcomes among different groups in different ways (Hill-Briggs et al., 2020), these determinants have many influences entrenched in their makeup. Connecting genetic investigation to these determinants is necessary due to the fact that a person's genetic makeup is largely outlined by their race and ethnicity. For example, genetic differences among populations with an ancestry coming predominantly from an African origin in comparison to someone with ancestry coming from European origin have different genetic makeups, possibly making for different pathophysiological reasonings for the predisposition to, or development of various diseases (Furr, 2002). Having a great influence in the genetic architecture due to evolutionary forces shaped by climate, diet and migration, among other factors, the racial and ethnic makeup of an individual or population also can also be a strong predictor of socio-economic conditions, especially in countries like The United States of America, due social issues that remain prevalent, such racism, that have strong effects on the living conditions of minority

populations (Smith et al., 2016). Health equity research focused on minority populations have shown that both biological and nonbiological risk factors are major influences in health disparities. These additional concerns also contribute to the aforementioned heterogeneity in the manifestation of stroke and have been thought to be a considerable contributor to the inequality seen in the mortality and disease burden associated with strokes (Stroke Facts | cdc.gov, 2020). Specific to the incidence of stroke, the incidence of stroke was seen to be nearly two-times as likely in black populations in comparison to white populations and although there have been improvements of stroke mortality among Hispanics, this population group has seen increases in the mortality as a result of stroke since 2013 (Stroke Facts | cdc.gov, 2020).

Genetic tools and our interpretations of how genetics influence disease have seen massive improvements in the last two decades, both in the ascertainment of genetic data and also in the analysis of said data. The same can be said about how genetics influence stroke exclusively. However, the access to and involvement in these developments and discoveries is not equitable, especially for minority populations (Nazroo, 1998) (Burnett-Hartman et al., 2020). Without continued efforts to address disparities in healthcare and clinical research, genetic technologies, understandings and the translation to clinical approaches like precision medicine, progressions in the application genetics in addressing disease will broaden the gap of the inequalities rooted in health disparities. Considering the excessive rates of mortality and disease burden that are associated with hemorrhagic strokes, the heterogeneity of the underlying genetic influence on the pathophysiology of this disease, the differences in genetic makeup among different races and ethnicities and also the need for a better understanding of genetic influences among minority

populations, this research study investigated the influence of Vitamin D Receptor (VDR) variants as a novel facilitator of hypertension-mediated Intracerebral Hemorrhage and stratified the findings based on race and ethnicity. The use VDR was chosen specifically as an intermediary to the prominent components of the heterogeneity of stroke occurrence; including previously considered pathophysiological influence on blood pressure, unique genetic architecture among different racial and ethnic backgrounds and how both of these components may lead to increased incidence of strokes, specifically Intracerebral Hemorrhage.

## Background of the Problem

Vitamin D Receptor genes are implicated in many biological pathways and associated with different diseases. Two of these diseases are cardiovascular disease and elevated blood pressure (Vélayoudom-Céphise et al., 2011) (Lin, Zhang, Li, Gai & Li, 2019). Vitamin D levels deficiencies have been found to be far more prevalent in individuals with darker skin tones and higher Body Mass Indexes (BMI), with Vitamin D deficiencies being linked to the increased need of additional synthesis of Vitamin D by these skin tones, with little well-known reasoning with the association to increased BMI, a possible bystander result (Weishaar, Rajan & Keller, 2016).

With the connection of VDR genes to possible pathways that are involved with hypertension, which is also connected to increased BMI, and a major risk factor heart and kidney disease, all of which are also strong risk factors for the incidence of stroke, this study used the interconnectedness of these risk factors as justification for the

investigation of the genetic predisposition associated with VDR genes in relation to ICH, as mediated by hypertension.

## Definition of Terms

- Case-control study – a study designed to determine the association of a risk factor or exposure to a specific disease.
- Case – group with a known outcome (disease)
- Control – group without a known outcome (disease)
- Exposure – the state of being exposed or subjected to something that impacts a medical outcome
- Gene – generationally transmissible unit of hereditary
- Genetic Contribution – the quantified amount or consideration of the genetic makeup of an individual that influences the incidence of disease
- Genetic Variant – alteration to a normally seen sequence of DNA
- Genome-Wide Association Study (GWAS) – scientific approach to identify the inheritance of a set or sets of genetic variants in relation to a particular phenotype.
- Genotype – the genetic makeup of a single organism
- Genotyping – the procedure of determining variances in the genetic make-up of individuals
- Hypertension (HTN) – irregularly high blood pressure
- Ligand-activated – a receptor and/or biological response that requires the binding of a ligand to initiate its action

- Linear Regression – an approach for modeling the relationship amongst a dependent variable and at least one other independent variable.
- Mendelian Randomization – method of using a measured rate of variation in genes of a known function to investigate their causal effect of a modifiable exposure on disease.
- Modifiable Risk Factor – a risk factor that can be changed by the individual of which possess it.
- Neuro-ICU – an intensive care unit devoted to life threatening neurological injuries or complications.
- Perfusion – the passage of blood and other fluids through the circulatory system or lymphatic system to an organ or tissue.
- Phenotype – a physical trait
- Polygenic Risk Score - quantification based on variation in multiple genetic loci and their associated weights
- Risk Factor – an exposure that increases the likelihood of a particular medical outcome
- Single Nucleotide Polymorphism (SNP) – most common variant of a gene amongst a population
- Socio-economic Status – social and economic standing of an individual or group of individuals in comparison to other populations.
- Transcription – the biological process where a gene's sequence is copied into a new code, known as RNA

- Transcription Factor – a biological influence that can alter the rate of transcription of a gene.

## Stroke: Biology, Types, Diagnosis, Mortality and Disease Burden

The brain is a very complex organ that relays its control and function via elaborate electrical circuitry outlined by what is known as the nervous system. The nervous system is composed of nerve cells, commonly referred to as neurons. These neurons are in constant communication with each other and are in need consistent, unimpeded supply of energy. This energy is furnished via the breakdown of energy substrates such as glucose via intricate chemical reactions. In the latter part of these essential reactions, oxygen is required. Oxygen in the brain is supplied via blood flow, sometimes referred to as perfusion. The brain is in receipt of both glucose and oxygen through the blood flow that is supplied via the heart, through the arteries found in the neck, including the carotid artery. These arteries diverge and are embedded deep into the brain. In the event that the supply of oxygen to the brain is insufficient, the breakdown of energy substrates is interrupted resulting in insufficient energy needed for normal brain function. Increased presence of toxic byproducts as a result of incomplete energy substrate breakdown is also a result and leads to further brain damage, which can be irreversible if the blood flow is not restored.

Strokes are a specific occasion where the interruption and damage to brain function mentioned prior occurs. Oddities and malfunctions in the blood vessels in the brain are a prominent reason for strokes. The first type of stroke is ischemic stroke.

Characterized a blockage of cerebral blood flow, ischemic strokes make up the roughly 80% of strokes that occur. Another form of stroke that is less problematic is a Transient Ischemic Attack (TIA), where signs and symptoms of an ischemic event dissipate within roughly 24 hours of onset. The second primary form of stroke is referred to as a hemorrhagic stroke. This is type is characterized by bleeding of a blood vessel in the brain.

Intracerebral hemorrhage is second in rank among the most common cause of stroke, totaling between 10-15% of all strokes (Smith & Eskey, 2011). Although this is much less than ischemic stroke, ICH is associated with much more mortality, ranging anywhere from 35-52%, a figure which is roughly 5 times larger than the mortality found to be associated with ischemic incidents (Smith & Eskey, 2011). Once the sudden onset of focal neurological symptoms is presented in clinic, they are assumed to be vascular in source until confirmed otherwise (Smith & Eskey, 2011). While patient presentations of vomiting, systolic blood pressure greater than 220 mmHg, severe headache, decreased or complete loss of consciousness are usually synonymous with a hemorrhagic stroke, this confirmation is found in neuroimaging. Rapid neuroimaging with both computerized tomography scan, CT-scan, and magnetic resonance imaging (MRI) scan are recommended to distinguish ischemic stroke from ICH. After an ICH is confirmed, immediate admission of the patient to a neuro-intensive care unit (Neuro-ICU) is recommended (Steiner & Jüttler, 2008).

A chief cause of ICH in adults is hypertension (HTN). Related to the fundamental relationship and mechanism associated with systemic blood pressure and circulation via arteries that come from major intracranial vessels, these vessels can develop hyperplasia,

enlargement of the vessels, and degeneration which predisposes these vessels to necrosis, and eventual rupture. The lenticulostriate arteries coming from the middle cerebral arteries, as well as the thalamoperforating and perforators coming from the basilar artery are involved in this process as well. Further alluding to the connectedness of hypertension and ICH, studies have found that the common location macroscopic hypertensive ICH reflect the locations supplied by the small vessels and perforators with 60% to 65% of bleeds in the putamen and internal capsule, 15% to 25% found in the thalamus and 5% to 10% in the pons (Smith & Eskey, 2011). Studies have also shown that while deep locations are most characteristic, hypertension remains a substantial risk factor for lobar hemorrhage in the absence of other causal risk factors (Broderick, Brott, Tomsick & Leach, 1993) (Jackson & Sudlow, 2006) (Smith & Eskey, 2011).

## Genetic Implications of Stroke Research

Given that stroke is a multifaceted disease, consequential of several risk factors and pre-existing diseases, looking into stroke at a genetic level has developed into focused approach for possible insight. Efforts to elucidate the underlying mechanism of stroke from a genetic level has a been a focus of stroke related predisposition, pathophysiology and recovery. This work has been furnished by two primary approaches, candidate gene approach and genome-wide and sequencing studies (Falcone & Woo, 2017). The candidate gene approach looks to assess the plausibility of an association between a specific genetic variant of a specific gene or region of the genome and the risk of disease. The gene candidates can be chosen based on prior knowledge of molecular function, environmental impact on the epigenetics of the genes of interest, or the

association of the selected genes and the associated phenotypes (Falcone & Woo, 2017). While this approach is relatively easy to conduct with smaller sample sizes, the effectiveness of this approach is largely based on what candidates have been chosen, the rate of variation amongst different population groups and also the probability of false-negative or false-positive results as a result of biases (Falcone & Woo, 2017).

As approaches have advanced in both the technology used and the interpretations being applied have made for the more efficient approach to be based around the assessment of large amounts, in the millions, of markers across the entire genome. This approach is defined as a Genome-wide association study (GWAS). Unlike candidate gene studies, GWAS requires large sample sizes to supply the power to validate the findings. Motivated by the GWAS' ability to help with the interrogation of complex traits, the large datasets from GWAS are able to be used in many different ways for the purpose of elucidating more findings and conclusions (Maier, Visscher, Robinson & Wray, 2017) (Duncan et al., 2019). One of these approaches that uses large data sets form GWAS is polygenic risk scores (PRS). Sometimes referred to as genetic risk scores or risk profiling, PRS have become widely used for the scoring and quantified prediction of the risk for the development of a disease. In line with precision medicine, the outlook for this type of genetic inquiry is believed to have the potential to improve health outcomes by accelerating diagnostics and coordinating patents with specific treatments that would address their needs (Duncan et al., 2019). Optimism continues to grow as PRS have shown reliable prediction for several complex genetic phenotypes such as diabetes, depression and blood pressure (Maier, Visscher, Robinson & Wray, 2017) (Duncan et al., 2019). PRS are scored by a calculation of the summation of risk alleles that are weighted

by their individual effect sizes as outlined by their derivation from the GWAS dataset being used. PRS are built specifically for the use of whatever complex genetic phenotype that is applicable based on the GWAS dataset being used (Maier, Visscher, Robinson & Wray, 2017).

## Implications of VDR Gene

One specific genetic element that has been deliberated upon with respect to its contribution to disease is that of the Vitamin D Receptor (VDR) gene. The VDR is in a family of steroid and thyroid hormone receptor transcription factors. VDR facilitate the effects of 1,25-dihydroyvitamin D (1,25(OH)2D) on gene expression (Uitterlinden, Fang, van Meurs, Pols & van Leeuwen, 2004). The prevalence and downstream functions attributed to the vitamin D hormone is modulated by the Vitamin D Receptor (Khan et al., 2016). This receptor is a ligand-activated transcription factor that mediates the expression of the gene (Khan et al., 2016) (Pike & Meyer, 2010) (Valdivielso & Fernandez, 2006). Vitamin D is an essential element in the regulation and homeostasis of the endocrine system. Happening prior to the activation via the ligand, the VDR specifically binds to sequences. Depending on the variants of these sequences, varying levels of transcription are conducted resulting in varying levels of serum Vitamin D levels that are available for biological function. Having involvement in a wide variety of biological functions associated with immune response, cell proliferation and differentiation and bone formation, vitamin D has also been implicated in common diseases such as diabetes, heart disease and cancer (Uitterlinden, Fang, van Meurs, Pols

& van Leeuwen, 2004). One biological manifestation that remains inconclusive is VDR gene's relation and potential causation of hypertension (Zhang et al., 2020).
Having previously been associated with control of calcium levels, keeping them in homeostasis, Vitamin D levels have regulatory processes on calcium absorption (Christakos, Dhawan, Porta, Mady & Seth, 2011). Roughly 40% of African Americans in the United States of America develop hypertension, much higher than the just 25% of white Americans. Studies investigated whether lower levels of vitamin D is associated with higher blood pressure and the risk for the development of hypertension, but more research is needed to confirm or deny this relationship (Vaidya & Forman, 2010). Other studies have shown VDR gene variants with purported association to the incidence of increased blood pressure (Iqbal, Maqbool & Khan, 2018) (Khan et al., 2016) (Maciejewski et al., 2019) (Muray, Parisi, Card's, Craver & Fernandez, 2003) (Pike & Meyer, 2010) (Valdivielso & Fernandez, 2006). The considerations of disease burden, mortality, underlying mechanisms of hypertension, calcium-regulation as a possible influence on the incidence of stroke (Anderson & Rosand, 2019) and Vitamin D and VDR gene variants make for an interesting perspective and potential assessment of hypertension as a mediator of ICH with possible genetic influences via VDR gene variants.

## Question and Hypothesis

With the consideration of the influences and underlying mechanism of hypertension and its connection to incidence of ICH, the increased incidence of hypertension in populations of non-European ancestry who also exhibit deficiencies in

Vitamin D levels, as well as the prospect of Vitamin D levels having influence on blood pressure, this study questions whether VDR gene variants and VDR variants previously considered for association with hypertension have an disproportionate influence on hypertension mediated ICH for non-European populations, both African American and Hispanic/Latin-American.

I pose the hypothesis that non-Europeans, both African American and Hispanic/Latin-Americans, will show an increased incidence of hypertension-mediated ICH based on the presence of VDR gene variants in comparison to that of White-Americans.

VDR gene variants previously considered to have an association with hypertension (Iqbal, Maqbool & Khan, 2018) (Khan et al., 2016) (Maciejewski et al., 2019) (Muray, Parisi, Card's, Craver & Fernandez, 2003) (Pike & Meyer, 2010) (Valdivielso & Fernandez, 2006).

- rs2228570 (*Fok*I)
- rs1544410 (*Bsm*I)
- rs7975232 (*Apa*I)
- rs731236 (*Taq*I)
- rs11568820 (Cdx2)
- rs4516035 (*Eco*RV)
- rs757343 (Tru9I)

## Implication of Research

Possible implications of this research include the use of PRS to uncover underlying genetic influence that could lead to future efforts directed by VDR gene variants. This study could potentially show the dexterity of PRS in context of risk factor mediated incidence of disease while considering novel genetic influences. This research will highlight the importance of diversity in clinical research participation, whether that be from significant findings with respect to stratification by race and ethnicity, or the need for more diverse participation in the form of a need of additional power to validate findings and elucidate any confounding or inconclusive findings.

# Chapter II.

# Methods

## Description of Datasets

This study used the individual and summary-level genetic data collected via Genome Wide Association Studies. The genetic data was allocated from GWAS across cohorts that include Genetic and Environmental Risk Factors for Hemorrhagic Stroke (GERFHS I, II & III) (Woo et al., 2002), "Genetics of Cerebral Hemorrhage on Anticoagulation" (GOCHA) (Chang et al., 2009), and European Samples from research participants from cohorts in the International Stroke Genetics Consortium (Traylor et al., 2012).. The Validation analysis, tables 12-16 was conducted with the MEGASTROKE trans-ancestral cohort, which included all stroke, ischemic stroke and subtypes) and the CARDIoGRAMplusC4D 1000 Genomes-based GWAS (60,801 CAD cases and 123,504 controls, mainly European) (Malik et al., 2018).

Below is a list of VDR gene variants that were previously considered for their association with Hypertension. (Iqbal, Maqbool & Khan, 2018) (Khan et al., 2016) (Maciejewski et al., 2019) (Muray, Parisi, Card's, Craver & Fernandez, 2003) (Pike & Meyer, 2010) (Valdivielso & Fernandez, 2006)

- rs2228570 (*Fok*I)
- rs1544410 (*Bsm*I)
- rs7975232 (*Apa*I)

- rs731236 (*Taq*I)
- rs11568820 (Cdx2)
- rs4516035 (*Eco*RV)
- rs757343 (Tru9I)

Individual-level ICH GWAS data contained GOCHA, GERFHS, and European participants of the International Stroke Genetics Consortium (ISGC) (ERU/ISGC): The GOCHA cohort has a total of 841 subjects, 436 of them are ICH cases (199 non-lobar / 237 lobar) and 405 of them are ICH-free controls. GERFHS has a total of 1642 subjects. 848 of them are ICH cases (508 non-lobar / 340 lobar), and 794 of them were ICH-free controls. The European cohort from the ISGC has a total of 1100 subjects, 577 of them are ICH cases (368 non-lobar / 209 lobar) and 523 of them are ICH-free controls (Woo et al., 2014). All three datasets, GOCHA, GERFHS and ERU/ISGC were imputed separately using the Haplotype Reference Consortium (HRC) reference panel on the Michigan Imputation Server. Imputation allows for the introduction of variability in the calculation of the PRs at an individual level without any alteration to the genetics that underly the model (Chen et al., 2020) Imputation bootstraps additional variants that can then be included in PRS.

The quality and control steps that were taken after imputation of each dataset included the retaining of only those SNPs with minor allele frequency (MAF) > 0.01, with imputation quality ($R^2$) > 0.4 and we excluded multi-allelic SNPs. Subjects with the following missing phenotypic data from each study were removed: ICH status, age, and sex

Summary-level GWAS stroke data (and related phenotypes) for validation were conducted The MEGASTROKE summary statistics (Malik et al., 2018) included all stroke (AS), ischemic stroke (IS), large artery stroke (LAS), cardioembolic stroke (CES), small vessel stroke (SVS) and coronary artery disease (CAD) (Won et al., 2015). This cohort includes 60, 801 cases and 123,504 controls, mainly of European ancestry.

## Construction of the Polygenic Risk Scores

The following steps were taken for the generation of the SBP PRS. QC of SBP summary statistics was conducted. This included retaining SNPs with MAF >0.01. This lowered the SNP total from 7,088,122 to 7,088,084. The ambiguous AT/GC SNPs were removed. This left 6,012,833 SNPs. The removal of duplicate SNPs left a remaining 6,012,833 SNPs. Construction of PRS included the pruning and thresholding method using PLINK software v1.90 (Chang et al., 2015) (Won et al., 2015), kept only SNPs with p-value ≤ 5x10-8, pruned at r2<0.4 using European 1000 Genomes phase 3 as reference LD panel, distance of SNPs 250kb

The following steps were taken to identify common SNPs lying within the VDR gene region. There were 119 original VDR gene SNPs identified via a ClinVar search but were too rare and were unidentifiable in the GWAS data included in this study. As a plausible alternative, coordinates of the VDR gene from the UCSC genome browser were found and used (Kent et al., 2002). SNPs from the respective region from the SBP GWAS data were extracted.

Finally, construction of the following PRSs took place. Constructs included, 1) SBP PRS, 2) SBP PRS + VDR SNPs (all), 3) SBP PRS + VDR SNPs (excluding those

previously associated with HTN), 4) VDR SNPs (all) and 5) VDR SNPs (excluding those previously associated with HTN). For the PRSs exclusion of any duplicate SNPs was ensured after the addition of the VDR SNPs

## Application of the PRS and Association Testing

Individual-level GWAS data (GOCHA, GERFHS, EUR/ISGC) was used to perform each of the following steps in each cohort separately: For each PRS as described above, the association estimates (betas) with SBP were used to generate five different PRSs using the PLINK software v1.90 (Chang et al., 2015). Standardization of each PRS [(PRS – mean of PRS) / (standard deviation of PRS)] was done in order to be able to express each association as per one standard deviation increase. The PRS was entered in as a linear predictor in a logistic regression, modeling risk for ICH. In each logistic regression we also included age, sex, and two principal components reflecting ancestry. The above steps were performed for ICH subtypes. The lobar subtype was modeled as lobar ICH versus all other. The non-lobar subtype was also modeled as non-lobar ICH versus all others. After performing the above analyses for each PRS within each study, meta-analysis was done under an inverse-variance, random effects model across GOCHA, GERFHS, and EUR/ISGC.

## Race and Ethnicity Analysis

For each race/ethnicity separately for the GOCHA study, for which there were non-white participants, the lobar subtype was modeled as lobar ICH versus all other. The non-

lobar subtype was also modeled as non-lobar ICH versus all others. After performing the above analyses for each PRS within each study, meta-analysis was done under an inverse-variance, random effects model.

Race and ethnicity were separated into three groups, white, black and Hispanic all of which were self-reported by the subjects. An adjustment for age, sex and principal components was not done due to limited sample size. Results from this portion of the analysis are unadjusted. All association analyses were performed using R software version 3.6.1. All association analyses and meta-analyses were performed using R software version 3.6.1 as well.

# Chapter III.

## Results

After duplicate removal, the SBP PRS contained 3,133 SNPs. The SBP PRS with all VDR gene SNPs derived from the USCS Genome Browser contained 3293 SNPs. The SBP PRS containing VDR gene SNPs, excluding those previously considered associated with HTN, had 3289. The PRS containing only VDR gene SNPs had 160 SNPs and the PRS containing VDR gene SNPs, excluding those previously considered associated, had 156 SNPs.

For the SBP PRS, all ICH, non-lobar and lobar cohorts were found to have Odds Ratios (OR), per 1 SD, of 1.08, 1.07 and 1.02 respectively. The all ICH, non-lobar and lobar cohorts were found to have Confidence intervals (CI) at 95%, of 1.01-1.15, 1.00-1.15 and 0.94-1.11, respectively. The p-values were found to be 0.03, 0.06 and 0.56 for all ICH, non-lobar ICH and lobar ICH respectively. These results were from inverse-variance, random effects meta-analysis across GOCHA, GERFHS and EUR/ISGC (Table 1).

For the SBP PRS containing all VDR gene SNPs, all ICH, non-lobar ICH and Lobar ICH, the ORs (per 1 SD) were found to be 1.08, 1.07 and 1.02 respectively. The 95% CI were found to be 1.01-1.15, 1.00-1.15 and 0.94-1.11 for all ICH, non-lobar ICH and lobar ICH respectively. The p-values were found to be 0.03, 0.06 and 0.57 for all ICH, non-lobar ICH and lobar ICH respectively. These results were from inverse-variance, random effects meta-analysis across GOCHA, GERFHS and EUR/ISGC (Table 2).

For the SBP PRS containing VDR gene SNPs, excluding those previously considered associated with HTN, all ICH, non-lobar ICH and lobar ICH, the ORs (per 1 SD) were found to be 1.08, 1.07 and 1.02 respectively. The all ICH, non-lobar and lobar cohorts were found to have 95% CI of 1.01-1.15, 1.00-1.15 and 0.94-1.11 respectively. The p-values were found to be 0.03, 0.06, and 0.57 for all ICH, non-lobar ICH and lobar ICH respectively. These results were from inverse-variance, random effects meta-analysis across GOCHA, GERFHS and EUR/ISGC (Table 3).

For the PRS containing only VDR gene SNPs, all ICH, non-lobar ICH and lobar ICH, the ORs (per 1 SD) were found to be 1.00 for all categories. The all ICH, non-lobar and lobar cohorts were found to have 95% CI of 0.94-1.07, 0.93-1.09, and 0.92-1.08 respectively. The p-values were found to be 1.00, 0.93, and 0.97 for all ICH, non-lobar ICH and lobar ICH respectively. These results were from inverse-variance, random effects meta-analysis across GOCHA, GERFHS and EUR/ISGC (Table 4).

For the PRS containing all VDR gene SNPs excluding those previously considered associated with HTN, all ICH, non-lobar ICH and lobar ICH, the ORs (per 1 SD) were found to be 1.00, 1.01, and 1.00 respectively. The all ICH, non-lobar and lobar cohorts were found to have 95% CI of 0.94-1.07, 0.93-1.09 and 0.92-1.08 respectively. The p-values were found to be 10 for all ICH, non-lobar ICH and lobar ICH respectively. These results were from inverse-variance, random effects meta-analysis across GOCHA, GERFHS and EUR/ISGC (Table 5).

For the race and ethnicity-based analysis, the GOCHA cohort was used. This cohort contained 841 total subjects with 778 white, 43 black and 18 Hispanic. The

percentages for the white, Black and Hispanic groups were 92.5%, 5.1%, 2.1% respectively (Table 6).

For the race and ethnicity-based analysis for the SBP PRS, ORs (per 1 SD) for all ICH, non-lobar ICH and lobar ICH were found to be 1.02, 1.06 and 0.98 for Whites, 1.45, 1.15 and 1.40 for Blacks and 0.65, 2.10 and 0.21 for Hispanic subjects respectively. The 95% CI for all ICH, non-lobar ICH and lobar ICH were found to be 0.88-1.18, 0.89-1.25, 0.83-1.14 for whites, 0.84-2.65, 0.66-2.03 and 0.73-2.83 for blacks and 0.24-1.48, 0.82-7.46 and 0.027-0.717 for Hispanics respectively. P-values for all ICH, non-lobar ICH and lobar ICH were found to be 0.79, 0.52 and 0.76 for whites, 0.20, 0.62 and 0.32 for blacks and 0.33, 0.17 and 0.0487 for Hispanics respectively. An adjustment for age, sex and principal components was not done due to limited sample size (Table 7).

For the race and ethnicity-based analysis for the SBP PRS containing all VDR gene SNPs, ORs (per 1 SD) for all ICH, non-lobar ICH and lobar ICH were found to be 1.02, 1.11 and 0.97 for Whites, 1.43, 1.11 and 1.45 for Blacks and 0.71, 2.16 and 0.256 for Hispanic subjects respectively. The 95% CI for all ICH, non-lobar ICH and lobar ICH were found to be 0.89-1.18, 0.90-1.26 and 0.83-1.14 for whites, 0.83-2.62, 0.64-1.97 and 0.75-3.05 for blacks and 0.27-1.60, 0.83-7.84 and 0.041-0.791 for Hispanics respectively. P-values for all ICH, non-lobar ICH and lobar ICH were found to be 0.75, 0.46 and 0.73 for whites, 0.22, 0.71 and 0.29 for blacks and 0.42, 0.16 and 0.056 for Hispanics respectively. An adjustment for age, sex and principal components was not done due to limited sample size (Table 8).

For the race and ethnicity-based analysis for the SBP PRS containing all VDR gene SNPs excluding those previously considered associated with HTN, ORs (per 1 SD)

for all ICH, non-lobar ICH and lobar ICH were found to be 1.03, 1.07 and 0.97 for Whites, 1.43, 1.12 and 1.45 for Blacks and 0.70, 2.15 and 0.253 for Hispanic subjects respectively. The 95% CI for all ICH, non-lobar ICH and lobar ICH were found to be 0.89-1.18, 0.90-1.26 and 0.83-1.14 for whites, 0.83-2.623, 0.64-1.98 and 0.75-3.03 for blacks and 0.27-1.59, 0.83-7.82 and 0.040-0.782 for Hispanics respectively. P-values for all ICH, non-lobar ICH and lobar ICH were found to be 0.76, 0.46 and 0.73 for whites, 0.21, 0.70 and 0.29 for blacks and 0.41, 0.16 and 0.056 for Hispanics respectively. An adjustment for age, sex and principal components was not done due to limited sample size (Table 9).

For the race and ethnicity-based analysis for the PRS containing all VDR gene SNPs, ORs (per 1 SD) for all ICH, non-lobar ICH and lobar ICH were found to be 1.04, 1.08 and 0.98 for Whites, 0.82, 0.65 and 1.35 for Blacks and 3.97, 1.17 and 4.28 for Hispanic subjects respectively. The 95% CI for all ICH, non-lobar ICH and lobar ICH were found to be 0.90-1.19, 0.91-1.28 and 0.84-1.14 for whites, 0.42-1.53, 0.32-1.24 and 0.63-3.15 for blacks and 1.12-27.33, 0.37-4.33 and 1.05-37.79 for Hispanics respectively. P-values for all ICH, non-lobar ICH and lobar ICH were found to be 0.62, 0.37 and 0.77 for whites, 0.53, 0.20 and 0.45 for blacks and 0.07, 0.79 and 0.098 for Hispanics respectively. An adjustment for age, sex and principal components was not done due to limited sample size (Table 10).

For the race and ethnicity-based analysis for the PRS containing all VDR gene SNPs excluding those previously considered associated with HTN, ORs (per 1 SD) for all ICH, non-lobar ICH and lobar ICH were found to be 1.04, 1.08 and 0.98 for Whites, 0.84, 0.67 and 1.36 for Blacks and 3.98, 1.15 and 4.52 for Hispanic subjects respectively. The

95% CI for all ICH, non-lobar ICH and lobar ICH were found to be 0.90-1.19, 0.91-1.28 and 0.84-1.14 for whites, 0.43-1.56, 0.33-1.26 and 0.64-3.18 for blacks and 1.11-27.40, 0.36-4.25 and 1.07-43.49 for Hispanics respectively. P-values for all ICH, non-lobar ICH and lobar ICH were found to be 0.62, 0.36 and 0.76 for whites, 0.57, 0.22 and 0.44 for blacks and 0.07, 0.81 and 0.096 for Hispanics respectively. An adjustment for age, sex and principal components was not done due to limited sample size (Table 11).

As described previously, summary statistics were derived from the MEGASTROKE and CAD cohort for the validation analyses. Using the five PRSs described above, the respective SNPs, SNP weights (betas) and standard errors from the SBP GWAS, the MEGASTROKE and the CAD summary statistics. For the SBP PRS validation analysis, AS had an OR, 95% CI and p-value of 1.029, 1.027-1.031 and 4.07x10-234 respectively. IS had an OR, 95% CI and p-value of 1.031, 1.029-1.033 and 1.14x10-233 respectively. LAS had an OR, 95% CI and p-value of 1.047, 1.043-1.051 and 3.73x10-126 respectively. CES had an OR, 95% CI and p-value of 1.016, 1.013-1.038 and 2.78x10-22. SVS had an OR, 95% CI and p-value of 1.034, 1.031-1.038 and 3.33x10-80 respectively. CAD had an OR, 95% CI and p-value of 1.033, 1.031-1.035 and 1.02x10-189 respectively (Table 12). AS had 2,995 SNPs, IS had 2,986 SNPs, LAS had 3,008 SNPs, CES had 3,014 SNPs, SVS had 2,599 SNPs and CAD had 3,024 SNPs available in their respective outcome datasets (Table 12).

For the SBP PRS with all VDR gene SNPs validation analysis, AS had an OR, 95% CI and p-value of 1.029, 1.027-1.031 and 1.24x10-240 respectively. IS had an OR, 95% CI and p-value of 1.031, 1.029-1.033 and 2.21x10-238 respectively. LAS had an OR, 95% CI and p-value of 1.047, 1.043-1.051 and 2.35x10-126 respectively. CES had

an OR, 95% CI and p-value of 1.017, 1.013-1.020 and 2.69x10-22. SVS had an OR, 95% CI and p-value of 1.033, 1.031-1.039 and 5.40x10-83 respectively. CAD had an OR, 95% CI and p-value of 1.033, 1.031-1.036 and 1.02x10-189 respectively (Table 13). AS had 3,147 SNPs, IS had 3,137 SNPs, LAS had 3,161 SNPs, CES had 3167 SNPs, SVS had 2,718 SNPs and CAD had 3,183 SNPs available in their respective outcome datasets (Table 13).

For validation analysis, the SBP PRS with VDR gene SNPs, excluding the SNPs previously considered associated with HTN, AS had an OR, 95% CI and p-value of 1.029, 1.027-1.031 and $2.22x10^{-240}$ respectively. IS had an OR, 95% CI and p-value of 1.031, 1.029-1.033 and $2.58x10^{-238}$ respectively. LAS had an OR, 95% CI and p-value of 1.047, 1.043-1.051 and $2.84x10^{-126}$ respectively. CES had an OR, 95% CI and p-value of 1.017, 1.013-1.020 and $2.92x10^{-22}$. SVS had an OR, 95% CI and p-value of 1.035, 1.031-1.039 and $8.04x10^{-83}$ respectively. CAD had an OR, 95% CI and p-value of 1.033, 1.031-1.036 and $5x10^{-196}$ respectively (Table 14). AS had 3,143 SNPs, IS had 3133 SNPs, LAS had 3,157 SNPs, CES had 3,163 SNPs, SVS had 2,714 SNPs and CAD had 3,179 SNPs available in their respective outcome datasets (Table 14).

The validation analysis of all of the VDR gene SNPs for the PRS, AS had an OR, 95% CI % and p-value of 1.01, 1.00-1.03 and 0.16 respectively. IS had an OR, 95% CI and p-value of 1.00, 0.98-1.02 and 0.85 respectively. LAS had an OR, 95% CI and p-value of 0.98, 0.94-1.02 and 0.38 respectively. CES had an OR, 95% CI and p-value of 0.97, 0.93-1.01 and 0.11. SVS had an OR, 95% CI and p-value of 1.14, 1.10-1.20 and $2.28x10^{-9}$ respectively. CAD had an OR, 95% CI and p-value of 1.08, 1.05-1.10 and $7.70x10^{-10}$ respectively (Table 15). AS had 152 SNPs, IS had 151 SNPs, LAS had 153

SNPs, CES had 153 SNPs, SVS had 119 SNPs and CAD had 159 SNPs available in their respective outcome datasets (Table 15).

The validation analysis for the PRS containing all of the VDR gene SNPs, excluding those previously considered associated with HTN, AS had an OR, 95% CI and p-value of 1.01, 0.99-1.03 and 0.19 respectively. IS had an OR, 95% CI and p-value of 1.00, 0.98-1.02 and 0.86 respectively. LAS had an OR, 95% CI and p-value of 0.98, 0.94-1.03 and 0.37 respectively. CES had an OR, 95% CI and p-value of 0.97, 0.93-1.01 and 0.11. SVS had an OR, 95% CI and p-value of 0.97, 1.10-1.21 and $3.44 \times 10^{-9}$ respectively. CAD had an OR, 95% CI and p-value of 1.07, 1.05-1.10 and $6.13 \times 10^{-9}$ respectively (Table 16). AS had 148 SNPs, IS had 147 SNPs, LAS had 149 SNPs, SVS had 115 SNPs and CAD had 155 SNPs available in their respective outcome datasets (Table 16).

# Chapter IV

## Discussion

The results found in this investigation allude to some interesting perspectives. While there were statistically significant findings in the SBP PRS with VDR gene SNPs in the three ICH cohorts (GOCHA, GERFHS, and EUR/ISGC) as stratified by race and ethnicity, there were some indications that there are differences in hypertension mediated ICH stratified by race and ethnicity. The ORs point to increased hypertension-mediated ICH in blacks and Hispanics in comparisons to whites while the CI and p-values are illustrations that more power is needed to confirm this association at a statistically significant level. (Tables 7-11). More specifically, when stratified by race/ethnicity in the GOCHA cohort, the differences that were observed between whites and blacks were most prominent in the non-lobar ICH subtype risk. This is in-line with the hypothesis because of the fact that hypertension-mediated ICH affects the deep regions of the brain, the SBP PRS would have more effect in the non-lobar ICH risk, especially when stratified by race and ethnicity when adding the VDR SNPs.

The analysis shows statistically significant relationships between VDR gene SNPs and small vessel disease and CAD, represented in PRS validation analysis (Tables 15-16). This, too, alludes to the plausibility of the VDR gene SNPs having an influence or association with some underlying mechanism associated with the incidence of stroke and possibly, more specifically, hypertension-mediated ICH. This can be assumed with the statistically significant finding of VDR gene SNPs' association with CAD and small vessel stroke's previous association with the incidence of stroke.

Disease of the small intracerebral vessels has been assumed causal for a large portion of small, lacunar and primary intracerebral brain bleeds (Allstair Lammie, 2000). Grounded in occurrence of ruptures of an intracerebral vessel, assumed consequential of a focal pathology at the vessel wall, bleeds in in these small vessels have also been associated with increased levels of blood pressure and previous diagnosis of hypertension (Allstair, 2006). This notion was also elucidated in studies showing the hypertensive vasculopathy and inflammatory processes and changes that occur in cerebral small vessel disease (CSVD) (Liu, Dong, Lyu, Chen & Li, 2018), with genetic studies showing plausible novel loci being connected to CSVD (Chung et al., 2019), alluding to more genetic implications in this relationship. With these previous considerations and findings from earlier studies, the results from the validation analysis pointing to VDR genes association with SVS could be indication of VDR genes being implicated in the underlying mechanism of hypertension-mediated ICH, having influence on the vasculopathy and perforation vessels that lead to this kind of brain bleed. The use of VDR genes could be an

The validation analysis' showing an association between VDR gene SNPs and CAD is an interesting finding. This is due to the fact that ischemic stroke and CAD have common risk factors, with CAD often being a predisposition to the incidence of ischemic stroke (Bhatia, et al., 2019). These findings have been corroborated by the frequency asymptomatic CAD being found in patients who had a stroke or TIA in autopsies, with studies showing up to 80% of patients showing the presence of coronary plaques and 40% showing stenosis of the coronary artery (Amarenco et al., 2013). While the focus of this study was ICH, studies have shown that patients with a history of CAD, history of

stroke, TIA is found to increase the incidence of both ischemic and hemorrhagic stroke (Ducrocq et al., 2013).

## Future Directions and Considerations

The findings in this study suggest the VDR gene may be an influential component of the underlying mechanism of stroke, ICH specifically. Future directions may include a more robust investigation of VDR gene variants across different cohorts. With the inequality of disease burden being shown in populations of color, increased efforts need to be made to include race and ethnicity as a component of research. A fruitful approach may include the juxtaposition of the use of PRS with another gene of a similar size to illustrate that the VDR gene's small size is rich its association as opposed to providing illegitimate findings, confounded by its small size.

From a public health perspective, stroke is the second leading cause of death world-wide, populations of color having twice the likelihood of experiencing a stroke than white populations (Stroke Facts | cdc.gov, 2020). This statistic alone speaks to the need to address disparities of disease burden, in-line with the overall need to address health disparities that include access to equitable healthcare and health literacy. It is essential that the advancements in genetic technologies, approaches and understandings are inclusive and provide an equitable opportunity for all populations. Not doing so has the propensity to make the clinical application of genetic understandings and approaches another metric of inequality, further widening the gap among diverse populations Martin et al., 2019).

Genetic research and treatment of diseases are seeing advancements that provide health professionals, scientist and researchers with great optimism about the future of medicine. With these advancements it is also important to consider the foundational elements that make up the infrastructure that furnishes the improvements and progressions of clinical research and its translation to healthcare. With respect to genetics, the subjects from which the genetic data is derived are the infrastructure from which conclusion are reached. If the infrastructure, subjects, form which the genetic data that is being used to draw conclusions does not reflect the eclectic and diverse populations that are being seen in clinics and hospitals across the world, there is a risk of continued disparities in the form of inapplicable and implausible to the populations that are underrepresented. Analogous to having diverse representation on a schoolboard, committee or representatives in a government body, diversity is needed to reflect the nuances and innerworkings of a community, group, race, ethnicity, among other categories. The same can be said for genetics. With many complex and heterogeneous mechanisms that connect genes to disease, without proper diversification of populations represented in the genetic data that is being interpreted, these nuances have a higher likelihood of being unaccounted for in the interpretations from research being conducted. Efforts are being made. It is essential that these efforts continue to be made and integrated in the routine considerations of genetic research.

## Limitations

Limitations to this study include sample sizes of the ICH GWAS cohorts overall. The sample size of non-white participants in the GOCHA cohort was limited which the

power of the analysis. Another limitation is the fact that the composition of the SBP GWAS cohort is primarily of European ancestry. Constructing a PRS from a multi-ethnic SBP GWAS data will more accurately reflect the genetic composition in non-Europeans and would allow for more unbiased inferences. There was no validation cohort to estimate the actual effect of the SBP PRS on blood pressure. Only one method of PRS construction, including pruning and thresholding was used. Other more robust methods such as those modeling linkage disequilibrium, such as LDpred, could potentially offer more in-depth explanation for more variance and thus yield better estimates. Also of note, the r2 value that was chosen to define independent variants is slightly higher than normal for the mendelian randomization validation analysis. A more stringent r2 value, such as 0.1, is warranted in future investigations.

## Conclusions

This study showed no statistically significant findings with respect to VDR gene-influenced hypertension-mediated ICH, however, findings did show increased incidence amongst non-white populations. This finding has the potential to be confirmed with more statistical power coming from datasets with more black and Hispanic subjects. The validation analysis showed statistically significant associations amongst VDR gene variants and SVS and CAD. SVS and CAD have been shown to play a role in the predisposition to, and underlying mechanism of stroke, ICH specifically.

List of Tables

Table 1. SBP PRS

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| SBP PRS | 1.08 | 1.01-1.15 | 0.03 |
| Non-lobar ICH | | | |
| SBP PRS | 1.07 | 1.00-1.15 | 0.06 |
| Lobar ICH | | | |
| SBP PRS | 1.02 | 0.94-1.11 | 0.56 |

*SBP PRS, analyzed in groups containing All ICH, non-lobar ICH and lobar ICH, detailing the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value.*

. Table 2. SBP PRS + VDR Gene SNPs (All)

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| SBP PRS | 1.08 | 1.01-1.15 | 0.03 |
| Non-lobar ICH | | | |
| SBP PRS | 1.07 | 1.00-1.15 | 0.06 |
| Lobar ICH | | | |
| SBP PRS | 1.02 | 0.94-1.11 | 0.57 |

*SBP PRS containing all of the VDR gene SNPs that were derived from the USCS Genome Browser, analyzed in groups containing All ICH, non-lobar ICH and lobar ICH, detailing the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value*

Table 3. SBP PRS + VDR Gene SNPs (Excluding those SNPs previously considered associated with HTN)

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| SBP PRS | 1.08 | 1.01-1.15 | 0.03 |
| Non-lobar ICH | | | |
| SBP PRS | 1.07 | 1.00-1.15 | 0.06 |
| Lobar ICH | | | |
| SBP PRS | 1.02 | 0.94-1.11 | 0.57 |

*SBP PRS containing all of the VDR gene SNPs that were derived from the USCS Genome Browser except for those VDR gene SNPs that were previously considered for association with HTN. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing All ICH, non-lobar ICH and lobar ICH.*

Table 4. Only VDR Gene SNPs (All)

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| SBP PRS | 1.00 | 0.94-1.07 | 1.00 |
| Non-lobar ICH | | | |
| SBP PRS | 1.00 | 0.93-1.09 | 0.93 |
| Lobar ICH | | | |
| SBP PRS | 1.00 | 0.92-1.08 | 0.97 |

*All of the VDR gene SNPs that were derived from the USCS Genome Browser. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing All ICH, non-lobar ICH and lobar ICH.*

Table 5. Only VDR Gene SNPs (Excluding those previously considered with association with HTN)

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| SBP PRS | 1.00 | 0.94-1.07 | 0.97 |
| Non-lobar ICH | | | |
| SBP PRS | 1.01 | 0.93-1.09 | 0.89 |
| Lobar ICH | | | |
| SBP PRS | 1.00 | 0.92-1.08 | 0.97 |

*All of the VDR gene SNPs that were derived from the USCS Genome Browser Excluding those previously considered associated with HTN. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value.*

Table 6. Racial and Ethnic Breakdown of GOCHA Cohort

| | All Cohort (n=841) |
|---|---|
| White | 778 (92.5%) |
| Black | 43 (5.1%) |
| Hispanic | 18 (2.1%) |
| | N missing = 2 |

*This is the racial and ethnic breakdown of the GOCHA cohort used for the race and ethnic based analysis. There were 2 subjects from this cohort that did not have corresponding race and ethnic phenotypic information. Of the total 841 subjects, 778 were White, 43 were Black and 18 were Hispanic, representing 92.5 %, 5.1% and 2.1% of the total cohort respectively.*

Table 7. Racial and Ethnic Breakdown of GOCHA Cohort

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| White | 1.02 | 0.88-1.18 | 0.79 |
| Black | 1.45 | 0.84-2.65 | 0.20 |
| Hispanic | 0.65 | 0.24-1.48 | 0.33 |
| Non-lobar ICH | | | |
| White | 1.06 | 0.89-1.25 | 0.52 |
| Black | 1.15 | 0.66-2.03 | 0.62 |
| Hispanic | 2.10 | 0.82-7.46 | 0.17 |
| Lobar ICH | | | |
| White | 0.98 | 0.83-1.14 | 0.76 |
| Black | 1.40 | 0.73-2.83 | 0.32 |
| Hispanic | 0.21 | 0.027-0.717 | 0.0487 |

*SBP PRS detailing the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing All ICH, non-lobar ICH and lobar ICH, categorized by race and ethnic identity. An adjustment for age, sex and principal components was not done due to limited sample size.*

Table 8. SBP PRS + VDR Gene SNPs (All) (Race and Ethnicity-Based Analysis)

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| White | 1.02 | 0.89-1.18 | 0.75 |
| Black | 1.43 | 0.83-2.62 | 0.22 |
| Hispanic | 0.71 | 0.27-1.60 | 0.42 |
| Non-lobar ICH | | | |
| White | 1.07 | 0.90-1.26 | 0.46 |
| Black | 1.11 | 0.64-1.97 | 0.71 |
| Hispanic | 2.16 | 0.83-7.84 | 0.16 |
| Lobar ICH | | | |
| White | 0.97 | 0.83-1.14 | 0.73 |
| Black | 1.45 | 0.75-3.05 | 0.29 |
| Hispanic | 0.256 | 0.041-0.791 | 0.056 |

*SBP PRS, containing all of the VDR gene SNPs that were derived from the USCS Genome Browser, detailing the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing All ICH, non-lobar ICH and lobar ICH, categorized by race and ethnic identity. An adjustment for age, sex and principal components was not done due to limited sample size.*

Table 9. SBP PRS + VDR Gene SNPs (Excluding Those VDR Gene SNPs Previously Considered Associated With HTN) (Race and Ethnicity-Based Analysis)

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| White | 1.03 | 0.89-1.18 | 0.76 |
| Black | 1.43 | 0.83-2.63 | 0.21 |
| Hispanic | 0.70 | 0.27-1.59 | 0.41 |
| Non-lobar ICH | | | |
| White | 1.07 | 0.90-1.26 | 0.46 |
| Black | 1.12 | 0.64-1.98 | 0.70 |
| Hispanic | 2.15 | 0.83-7.82 | 0.16 |
| Lobar ICH | | | |
| White | 0.97 | 0.83-1.14 | 0.73 |
| Black | 1.45 | 0.75-3.03 | 0.29 |
| Hispanic | 0.253 | 0.040-0.787 | 0.056 |

*SBP PRS, containing all of the VDR gene SNPs that were derived from the USCS Genome Browser excluding those previously considered associated with HTN, detailing the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing All ICH, non-lobar ICH and lobar ICH, categorized by race and ethnic identity. An adjustment for age, sex and principal components was not done due to limited sample size.*

Table 10. Only VDR Gene SNPs (All) (Race and Ethnicity-Based Analysis)

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| White | 1.04 | 0.90-1.19 | 0.62 |
| Black | 0.82 | 0.42-1.53 | 0.53 |
| Hispanic | 3.97 | 1.12-27.33 | 0.07 |
| Non-lobar ICH | | | |
| White | 1.08 | 0.91-1.28 | 0.37 |
| Black | 0.65 | 0.32-1.24 | 0.20 |
| Hispanic | 1.17 | 0.37-4.33 | 0.79 |
| Lobar ICH | | | |
| White | 0.98 | 0.84-1.14 | 0.77 |
| Black | 1.35 | 0.63-3.15 | 0.45 |
| Hispanic | 4.28 | 1.05-37.79 | 0.098 |

*All of the VDR gene SNPs, only, that were derived from the USCS Genome Browser excluding those previously considered associated with HTN, detailing the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing All ICH, non-lobar ICH and lobar ICH, categorized by race and ethnic identity. An adjustment for age, sex and principal components was not done due to limited sample size*

Table 11. Only VDR Gene SNPs Excluding Those Previously Considered Associated with HTN (Race and Ethnicity-Based Analysis)

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| All ICH | | | |
| White | 1.04 | 0.90-1.19 | 0.62 |
| Black | 0.84 | 0.43-1.56 | 0.57 |
| Hispanic | 3.98 | 1.11-27.40 | 0.07 |
| Non-lobar ICH | | | |
| White | 1.08 | 0.91-1.28 | 0.36 |
| Black | 0.67 | 0.33-1.26 | 0.22 |
| Hispanic | 1.15 | 0.36-4.25 | 0.81 |
| Lobar ICH | | | |
| White | 0.98 | 0.84-1.14 | 0.76 |
| Black | 1.36 | 0.64-3.18 | 0.44 |
| Hispanic | 4.52 | 1.05-43.49 | 0.096 |

*VDR gene SNPs, excluding those previously considered associated with HTN, that were derived from the USCS Genome Browser excluding those previously considered associated with HTN, detailing the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing All ICH, non-lobar ICH and lobar ICH, categorized by race and ethnic identity. An adjustment for age, sex and principal components was not done due to limited sample size.*

Table 12. SBP PRS - Validation Analysis

| | **OR (Per 1 SD)** | **95% CI** | **P-value** |
|---|---|---|---|
| **AS** | 1.029 | 1.027-1.031 | $4x10^{-234}$ |
| **IS** | 1.031 | 1.029-1.033 | $1.14x10^{-233}$ |
| **LAS** | 1.047 | 1.043-1.051 | $3.73x10^{-126}$ |
| **CES** | 1.016 | 1.013-1.020 | $2.78x10^{-22}$ |
| **SVS** | 1.034 | 1.031-1.038 | $3.33x10^{-80}$ |
| **CAD** | 1.033 | 1.031-1.035 | $1.02x10^{-189}$ |

*SBP PRS Validation Analysis detailing the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing AS, IS, LAS, CES, SVS and CAD. The MEGASTROKE cohort contained 2995 SNPs for AS, 2986 SNPs for IS, 3008 SNPs for LAS, 3014 SNPs for CES, 2599 SNPs for SVS and 3024 SNPs for CAD.*

Table 13. SBP PRS + VDR Gene SNPs (All) - Validation Analysis

| | **OR (Per 1 SD)** | **95% CI** | **P-value** |
|---|---|---|---|
| **AS** | 1.029 | 1.027-1.031 | $1.24x10^{-240}$ |
| **IS** | 1.031 | 1.029-1.033 | $2.01x10^{238}$ |
| **LAS** | 1.047 | 1.043-1.051 | $2.35x10^{-126}$ |
| **CES** | 1.017 | 1.013-1.020 | $2.69x10^{-22}$ |
| **SVS** | 1.035 | 1.032-1.039 | $5.40x10^{-83}$ |
| **CAD** | 1.033 | 1.031-1.036 | $2.35x10^{-196}$ |

*Validation Analysis of SBP PRS containing all VDR gene SNPs, derived from the USCS Genome Browser, detailing the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing AS, IS, LAS, CES, SVS and CAD. The MEGASTROKE cohort contained 2995*

*SNPs for AS, 2986 SNPs for IS, 3008 SNPs for LAS, 3014 SNPs for CES, 2599 SNPs for SVS and 3024 SNPs for CAD.*

Table 14. SBP PRS + VDR Gene SNPs (Excluding Those Previously Considered Associated with HTN) - Validation Analysis

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| **AS** | 1.029 | 1.027-1.031 | $2.22x10^{-240}$ |
| **IS** | 1.031 | 1.029-1.033 | $2.58x10^{-126}$ |
| **LAS** | 1.047 | 1.043-1.051 | $2.84x10^{-126}$ |
| **CES** | 1.017 | 1.013-1.020 | $2.92x10^{-22}$ |
| **SVS** | 1.035 | 1.031-1.039 | $8.04x10^{-83}$ |
| **CAD** | 1.033 | 1.031-1.036 | $5x10^{-196}$ |

*Validation Analysis of SBP PRS containing all VDR gene SNPs excluding those previously considered associated with HTN, derived from the USCS Genome Browser. Statistics of the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value shown. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing AS, IS, LAS, CES, SVS and CAD.*

Table 15. Only VDR Gene SNPs (ALL) - Validation Analysis

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| **AS** | 1.01 | 1.00-1.03 | 0.16 |
| **IS** | 1.00 | 0.98-1.02 | 0.85 |
| **LAS** | 0.98 | 0.94-1.02 | 0.38 |
| **CES** | 0.97 | 0.93-1.01 | 0.11 |
| **SVS** | 1.14 | 1.01-1.20 | 2.28x10-9 |
| **CAD** | 1.08 | 1.05-1.10 | $7.70x10^{-10}$ |

*Validation Analysis of all VDR gene SNPs derived from the USCS Genome Browser. Statistics of the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value shown. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing AS, IS, LAS, CES, SVS and CAD. SVS and CAD showed statistically significant results for SVS and CAD's OR, CI at 95% and p-value.*

Table 16. PRS - VDR Gene SNPs (Excluding Those Previously Considered Associated with HTN) - Validation Analysis

| | OR (Per 1 SD) | 95% CI | P-value |
|---|---|---|---|
| **AS** | 1.01 | 0.99-1.03 | 0.19 |
| **IS** | 1.00 | 0.98-1.02 | 0.86 |
| **LAS** | 0.98 | 0.94-1.03 | 0.37 |
| **CES** | 0.97 | 0.93-1.01 | 0.09 |
| **SVS** | 1.15 | 1.01-1.21 | $3.44 \times 10^{-9}$ |
| **CAD** | 1.07 | 1.05-1.10 | $6.13 \times 10^{-9}$ |

*Validation Analysis of VDR gene SNPs excluding those previously considered associated with HTN, derived from the USCS Genome Browser. Statistics of the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value shown. These statistics detail the odds ratios (OR) per 1 standard deviation (SD), 95% confidence interval (CI) and p-value, analyzed in groups containing AS, IS, LAS, CES, SVS and CAD. SVS and CAD showed statistically significant results for SVS and CAD's OR, CI at 95% and p-value.*

References


Alistair Lammie, G. (2000). Pathology of small vessel stroke. *British Medical Bulletin*, *56*(2), 296-306. doi: 10.1258/0007142001903229

Alistair, D. (2006). Hypertensive Cerebral Small Vessel Disease and Stroke. *Brain Pathology*, *12*(3), 358-370. doi: 10.1111/j.1750-3639.2002.tb00450.x

Amarenco, P., Bogousslavsky, J., Caplan, L., Donnan, G., & Hennerici, M. (2009). Classification of Stroke Subtypes. *Cerebrovascular Diseases*, *27*(5), 493-501. doi: 10.1159/000210432

Amarenco, P., Lavallée, P., Labreuche, J., Ducrocq, G., Juliard, J., & Feldman, L. et al. (2013). Coronary Artery Disease and Risk of Major Vascular Events After Cerebral Infarction. *Stroke*, *44*(6), 1505-1511. doi: 10.1161/strokeaha.111.000142

Anderson, C., & Rosand, J. (2019). The yin and yang of magnesium and calcium. *Neurology*, 10.1212/WNL.0000000000006991. doi: 10.1212/wnl.0000000000006991

Bentley, A., Callier, S., & Rotimi, C. (2017). Diversity and inclusion in genomic research: why the uneven progress?. *Journal Of Community Genetics*, *8*(4), 255-266. doi: 10.1007/s12687-017-0316-6

Bhatia, R., Sharma, G., Patel, C., Garg, A., Roy, A., & Bali, P. et al. (2019). Coronary Artery Disease in Patients with Ischemic Stroke and TIA. *Journal Of Stroke And Cerebrovascular Diseases*, *28*(12), 104400. doi: 10.1016/j.jstrokecerebrovasdis.2019.104400

Broderick, J., Brott, T., Tomsick, T., & Leach, A. (1993). Lobar hemorrhage in the elderly. The undiminishing importance of hypertension. *Stroke*, *24*(1), 49-51. doi: 10.1161/01.str.24.1.49

Broderick, J., Brott, T., Tomsick, T., Miller, R., & Huster, G. (1993). Intracerebral hemorrhage more than twice as common as subarachnoid hemorrhage. *Journal Of Neurosurgery*, *78*(2), 188-191. doi: 10.3171/jns.1993.78.2.0188

Burgess, S. (2014). Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *International Journal Of Epidemiology*, *43*(3), 922-929. doi: 10.1093/ije/dyu005

Burnett-Hartman, A., Blum-Barnett, E., Carroll, N., Madrid, S., Jonas, C., & Janes, K. et al. (2020). Return of Research-Related Genetic Test Results and Genetic Discrimination Concerns: Facilitators and Barriers of Genetic Research Participation in Diverse Groups. *Public Health Genomics*, *23*(1-2), 59-68. doi: 10.1159/000507056

Burt, B. (2001). Definitions of Risk. *Journal Of Dental Education*, *65*(10), 1007-1008. doi: 10.1002/j.0022-0337.2001.65.10.tb03442.x

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1).

Chang, Y., Cortellini, L., Rosand, J., Singer, D., Schwab, K., & Selim, M. et al. (2009). Exploiting Common Genetic Variation to Make Anticoagulation Safer. Stroke, 40(3, Supplement 1), S64-S66. doi: 10.1161/strokeaha.108.533190

Chen, S., Dias, R., Evans, D., Salfati, E., Liu, S., Wineinger, N., & Torkamani, A. (2020). Genotype imputation and variability in polygenic risk score estimation. *Genome Medicine*, *12*(1). doi: 10.1186/s13073-020-00801-x

Choi, S., Mak, T., & O'Reilly, P. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759-2772. doi: 10.1038/s41596-020-0353-1

Christakos, S., Dhawan, P., Porta, A., Mady, L., & Seth, T. (2011). Vitamin D and intestinal calcium absorption. *Molecular And Cellular Endocrinology*, *347*(1-2), 25-29. doi: 10.1016/j.mce.2011.05.038

Chung, J., Marini, S., Pera, J., Norrving, B., Jimenez-Conde, J., & Roquer, J. et al. (2019). Genome-wide association study of cerebral small vessel disease reveals established and novel loci. *Brain*, *142*(10), 3176-3189. doi: 10.1093/brain/awz233

Dajpratham, P., Pukrittayakamee, P., Atsariyasing, W., Wannarit, K., Boonhong, J., & Pongpirul, K. (2020). The validity and reliability of the PHQ-9 in screening for post-stroke depression. *BMC Psychiatry*, *20*(1). doi: 10.1186/s12888-020-02699-6

Dichgans, M., Pulit, S., & Rosand, J. (2019). Stroke genetics: discovery, biology, and clinical applications. *The Lancet Neurology*, *18*(6), 587-599. doi: 10.1016/s1474-4422(19)30043-2

Ducrocq, G., Amarenco, P., Labreuche, J., Alberts, M., Mas, J., & Ohman, E. et al. (2013). A History of Stroke/Transient Ischemic Attack Indicates High Risks of Cardiovascular Event and Hemorrhagic Stroke in Patients With Coronary Artery Disease. *Circulation*, *127*(6), 730-738. doi: 10.1161/circulationaha.112.141572

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., & Feldman, M. et al. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, *10*(1). doi: 10.1038/s41467-019-11112-0

Elliott, J., Bodinier, B., Bond, T., Chadeau-Hyam, M., Evangelou, E., & Moons, K. et al. (2020). Predictive Accuracy of a Polygenic Risk Score–Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *JAMA*, *323*(7), 636. doi: 10.1001/jama.2019.22241

Falcone, G., & Woo, D. (2017). Genetics of Spontaneous Intracerebral Hemorrhage. *Stroke*, *48*(12), 3420-3424. doi: 10.1161/strokeaha.117.017072

Furr, L. (2002). Perceptions of Genetics Research As Harmful to Society: Differences among Samples of African-Americans and European-Americans. *Genetic Testing*, *6*(1), 25-30. doi: 10.1089/109065702760093889

Georgakis, M., Gill, D., Rannikmäe, K., Traylor, M., Anderson, C., & Lee, J. et al. (2019). Genetically Determined Levels of Circulating Cytokines and Risk of Stroke. *Circulation*, *139*(2), 256-268. doi: 10.1161/circulationaha.118.035905

Girouard, H., & Iadecola, C. (2006). Neurovascular coupling in the normal brain and in hypertension, stroke, and Alzheimer disease. *Journal Of Applied Physiology*, *100*(1), 328-335. doi: 10.1152/japplphysiol.00966.2005

Greenberg, S., Eng, J., Ning, M., Smith, E., & Rosand, J. (2004). Hemorrhage Burden Predicts Recurrent Intracerebral Hemorrhage After Lobar Hemorrhage. *Stroke*, *35*(6), 1415-1420. doi: 10.1161/01.str.0000126807.69758.0e

Hata, J., & Kiyohara, Y. (2013). Epidemiology of Stroke and Coronary Artery Disease in Asia. *Circulation Journal*, *77*(8), 1923-1932. doi: 10.1253/circj.cj-13-0786

Hemani, G., Tilling, K., & Davey Smith, G. (2017). Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLOS Genetics, 13(11), e1007081. doi: 10.1371/journal.pgen.1007081

Hemani G, Zheng J, Elsworth B, Wade KH, Baird D, Haberland V, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC, The MR-Base Collaboration. The MR-Base platform supports systematic causal inference across the human phenome. eLife 2018;7:e34408. doi: 10.7554/eLife.34408

High blood pressure (hypertension) - Symptoms and causes. (2020). Retrieved 14 March 2021, from https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/symptoms-causes/syc-20373410

Hill-Briggs, F., Adler, N., Berkowitz, S., Chin, M., Gary-Webb, T., & Navas-Acien, A. et al. (2020). Social Determinants of Health and Diabetes: A Scientific Review. *Diabetes Care*, *44*(1), 258-279. doi: 10.2337/dci20-0053

Howell, A., Zheng, J., Haycock, P., McAleenan, A., Relton, C., Martin, R., & Kurian, K. (2018). Use of Mendelian Randomization for Identifying Risk Factors for Brain Tumors. *Frontiers In Genetics*, *9*. doi: 10.3389/fgene.2018.00525

Iadecola, C., & Anrather, J. (2011). Stroke research at a crossroad: asking the brain for directions. *Nature Neuroscience*, *14*(11), 1363-1368. doi: 10.1038/nn.2953

Iqbal, M., Maqbool, S., & Khan, T. (2018). Association of low penetrance vitamin D receptor Tru9I (rs757343) gene polymorphism with risk of premenopausal breast cancer. *Journal Of International Medical Research*, *46*(5), 1801-1814. doi: 10.1177/0300060518761304

Jackson, C., & Sudlow, C. (2006). Is hypertension a more frequent risk factor for deep than for lobar supratentorial intracerebral haemorrhage?. *Journal Of Neurology, Neurosurgery & Psychiatry*, *77*(11), 1244-1252. doi: 10.1136/jnnp.2006.089292

Johansson, B. (1999). HYPERTENSION MECHANISMS CAUSING STROKE. *Clinical And Experimental Pharmacology And Physiology*, *26*(7), 563-565. doi: 10.1046/j.1440-1681.1999.03081.x

Johnson, C., Nguyen, M., Roth, G., Nichols, E., Alam, T., & Abate, D. et al. (2019). Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, *18*(5), 439-458. doi: 10.1016/s1474-4422(19)30034-1

Juraschek, S., Ishak, A., Mukamal, K., Cohen, M., & Beach, J. (2019). Impact of Clinic-Based Blood Pressure Approaches on Blood Pressure Measurement. *American Journal Of Hypertension*, *33*(1), 26-30. doi: 10.1093/ajh/hpz118

Katan, M., & Luft, A. (2018). Global Burden of Stroke. *Seminars In Neurology*, *38*(02), 208-211. doi: 10.1055/s-0038-1649503

Kelly, P., Furie, K., Shafqat, S., Rallis, N., Chang, Y., & Stein, J. (2003). Functional recovery following rehabilitation after hemorrhagic and ischemic stroke11No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit on the authors or on any organization with which the authors are associated. *Archives Of Physical Medicine And Rehabilitation*, *84*(7), 968-972. doi: 10.1016/s0003-9993(03)00040-6

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996-1006.

Khan, R., Riestra, P., Gebreab, S., Wilson, J., Gaye, A., Xu, R., & Davis, S. (2016). Vitamin D Receptor Gene Polymorphisms Are Associated with Abdominal Visceral Adipose Tissue Volume and Serum Adipokine Concentrations but Not with Body Mass Index or Waist Circumference in African Americans: The Jackson Heart Study. *The Journal Of Nutrition*, *146*(8), 1476-1482. doi: 10.3945/jn.116.229963

Klatsky, A., Armstrong, M., & Friedman, G. (1990). Risk of cardiovascular mortality in alcohol drinkers, ex-drinkers and nondrinkers. *The American Journal Of Cardiology*, *66*(17), 1237-1242. doi: 10.1016/0002-9149(90)91107-h

Krishnamurthi, R., Barker-Collo, S., Parag, V., Parmar, P., Witt, E., & Jones, A. et al. (2018). Stroke Incidence by Major Pathological Type and Ischemic Subtypes in the Auckland Regional Community Stroke Studies. *Stroke*, *49*(1), 3-10. doi: 10.1161/strokeaha.117.019358

Ledford, H. (2016). CRISPR: gene editing is just the beginning. *Nature*, *531*(7593), 156-159. doi: 10.1038/531156a

Lewis, S. (2010). Mendelian Randomization as Applied to Coronary Heart Disease, Including Recent Advances Incorporating New Technology. *Circulation: Cardiovascular Genetics*, *3*(1), 109-117. doi: 10.1161/circgenetics.109.880955

Lin, L., Zhang, L., Li, C., Gai, Z., & Li, Y. (2019). Vitamin D and Vitamin D Receptor: New Insights in the Treatment of Hypertension. *Current Protein & Peptide Science*, *20*(10), 984-995. doi: 10.2174/1389203720666190807130504

Liu, Y., Dong, Y., Lyu, P., Chen, W., & Li, R. (2018). Hypertension-Induced Cerebral Small Vessel Disease Leading to Cognitive Impairment. *Chinese Medical Journal*, *131*(5), 615-619. doi: 10.4103/0366-6999.226069

Maciejewski, A., Kowalczyk, M., Herman, W., Czyżyk, A., Kowalska, M., Żaba, R., & Łącka, K. (2019). Vitamin D Receptor Gene Polymorphisms and Autoimmune Thyroiditis: Are They Associated with Disease Occurrence and Its Features?. *Biomed Research International*, *2019*, 1-9. doi: 10.1155/2019/8197580

Maier, R., Visscher, P., Robinson, M., & Wray, N. (2017). Embracing polygenicity: a review of methods and tools for psychiatric genetics research. *Psychological Medicine*, *48*(7), 1055-1067. doi: 10.1017/s0033291717002318

Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., & Mishra, A. et al. (2018). Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nature Genetics, 50(4), 524-537. doi: 10.1038/s41588-018-0058-3

Marees, A., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal Of Methods In Psychiatric Research*, *27*(2), e1608. doi: 10.1002/mpr.1608

Markus, H. (2011). Stroke genetics. *Human Molecular Genetics*, *20*(R2), R124-R131. doi: 10.1093/hmg/ddr345

Martin, A., Kanai, M., Kamatani, Y., Okada, Y., Neale, B., & Daly, M. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. Nature Genetics, 51(4), 584-591. doi: 10.1038/s41588-019-0379-x

Muray, S., Parisi, E., Card??s, A., Craver, L., & Fern??ndez, E. (2003). Influence of vitamin D receptor gene polymorphisms and 25-hydroxyvitamin D on blood pressure in apparently healthy subjects. *Journal Of Hypertension*, *21*(11), 2069-2075. doi: 10.1097/00004872-200311000-00016

Nazroo, J. (1998). *Sociology Of Health &Amp; Illness*, *20*(5), 710-730. doi: 10.1111/1467-9566.00126

Perkins, S., Bacchetti, P., Davey, C., Lindsell, C., Mazumdar, M., & Oster, R. et al. (2016). Best Practices for Biostatistical Consultation and Collaboration in Academic Health Centers. *The American Statistician*, *70*(2), 187-194. doi: 10.1080/00031305.2015.1077727

Pike, J., & Meyer, M. (2010). The Vitamin D Receptor: New Paradigms for the Regulation of Gene Expression by 1,25-Dihydroxyvitamin D3. *Endocrinology And Metabolism Clinics Of North America*, *39*(2), 255-269. doi: 10.1016/j.ecl.2010.02.007

Qureshi, A., Tuhrim, S., Broderick, J., Batjer, H., Hondo, H., & Hanley, D. (2001). Spontaneous Intracerebral Hemorrhage. *New England Journal Of Medicine*, *344*(19), 1450-1460. doi: 10.1056/nejm200105103441907

Relton, C., & Davey Smith, G. (2015). Mendelian randomization: applications and limitations in epigenetic studies. *Epigenomics*, *7*(8), 1239-1243. doi: 10.2217/epi.15.88

Romero, J., Morris, J., & Pikula, A. (2008). Review: Stroke prevention: modifying risk factors. *Therapeutic Advances In Cardiovascular Disease*, *2*(4), 287-303. doi: 10.1177/1753944708093847

Rostand, S. (2010). Vitamin D, Blood Pressure, and African Americans: Toward a Unifying Hypothesis: Figure 1. *Clinical Journal Of The American Society Of Nephrology*, *5*(9), 1697-1703. doi: 10.2215/cjn.02960410

Smith, C., Fullerton, S., Dookeran, K., Hampel, H., Tin, A., & Maruthur, N. et al. (2016). Using Genetic Technologies To Reduce, Rather Than Widen, Health Disparities. *Health Affairs*, *35*(8), 1367-1373. doi: 10.1377/hlthaff.2015.1476

Smith, S., & Eskey, C. (2011). Hemorrhagic Stroke. *Radiologic Clinics Of North America*, *49*(1), 27-45. doi: 10.1016/j.rcl.2010.07.011

Steiner, T., & Jüttler, E. (2008). American guidelines for the management of spontaneous intracerebral hemorrhage in adults: European perspective. *Polish Archives Of Internal Medicine*, *118*(4), 181-182. doi: 10.20452/pamw.354

Stroke Facts | cdc.gov. (2020). Retrieved 7 February 2021, from https://www.cdc.gov/stroke/facts.htm

Stroke Statistics | Internet Stroke Center. (2020). Retrieved 14 March 2021, from http://www.strokecenter.org/patients/about-stroke/stroke-statistics/

Suarez, J., Tarr, R., & Selman, W. (2006). Aneurysmal Subarachnoid Hemorrhage. *New England Journal Of Medicine*, *354*(4), 387-396. doi: 10.1056/nejmra052732

Teumer, A. (2018). Common Methods for Performing Mendelian Randomization. *Frontiers In Cardiovascular Medicine*, *5*. doi: 10.3389/fcvm.2018.00051

The top 10 causes of death. (2020). Retrieved 3 January 2021, from https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends In Ecology & Evolution*, *29*(12), 673-680. doi: 10.1016/j.tree.2014.10.004

Traylor, M., Farrall, M., Holliday, E., Sudlow, C., Hopewell, J., & Cheng, Y. et al. (2012). Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. The Lancet Neurology, 11(11), 951-962. doi: 10.1016/s1474-4422(12)70234-x

Uitterlinden, A., Fang, Y., van Meurs, J., Pols, H., & van Leeuwen, J. (2004). Genetics and biology of vitamin D receptor polymorphisms. *Gene*, *338*(2), 143-156. doi: 10.1016/j.gene.2004.05.014

Vaidya, A., & Forman, J. (2010). Vitamin D and Hypertension. *Hypertension*, *56*(5), 774-779. doi: 10.1161/hypertensionaha.109.140160

Valdivielso, J., & Fernandez, E. (2006). Vitamin D receptor polymorphisms and diseases. *Clinica Chimica Acta*, *371*(1-2), 1-12. doi: 10.1016/j.cca.2006.02.016

Vélayoudom-Céphise, F., Larifla, L., Donnet, J., Maimaitiming, S., Deloumeaux, J., & Blanchet, A. et al. (2011). Vitamin D deficiency, vitamin D receptor gene polymorphisms and cardiovascular risk factors in Caribbean patients with type 2 diabetes. *Diabetes & Metabolism*, *37*(6), 540-545. doi: 10.1016/j.diabet.2011.05.005

Warren, H., Evangelou, E., Mosen, D., Mifsud, B., Pazoki, R., & Gao, H. et al. (2018). GENETIC ANALYSIS OF OVER ONE MILLION PEOPLE IDENTIFIES 535 NOVEL LOCI ASSOCIATED WITH BLOOD PRESSURE AND RISK OF CARDIOVASCULAR DISEASE. *Journal Of Hypertension*, *36*(Supplement 1), e229. doi: 10.1097/01.hjh.0000539644.13726.3b

Weishaar, T., Rajan, S., & Keller, B. (2016). Probability of Vitamin D Deficiency by Body Weight and Race/Ethnicity. *The Journal Of The American Board Of Family Medicine*, *29*(2), 226-232. doi: 10.3122/jabfm.2016.02.150251

Won, Hong-Hee, Hall, Leanne M, Willenborg, Christina, Kanoni, Stavroula, Saleheen, Danish, Nelson, Christopher P, . . . McPherson, Ruth. (2015). A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. Nature Genetics, 47(10), 1121-1130.

Woo, D., Falcone, G., Devan, W., Brown, W., Biffi, A., & Howard, T. et al. (2014). Meta-analysis of Genome-wide Association Studies Identifies 1q22 as a Susceptibility Locus for Intracerebral Hemorrhage. The American Journal Of Human Genetics, 94(4), 511-521. doi: 10.1016/j.ajhg.2014.02.012

Woo, D., Haverbusch, M., Sekar, P., Kissela, B., Khoury, J., & Schneider, A. et al. (2004). Effect of Untreated Hypertension on Hemorrhagic Stroke. *Stroke*, *35*(7), 1703-1708. doi: 10.1161/01.str.0000130855.70683.c8

Woo, D., Sauerbeck, L., Kissela, B., Khoury, J., Szaflarski, J., & Gebel, J. et al. (2002). Genetic and Environmental Risk Factors for Intracerebral Hemorrhage. Stroke, 33(5), 1190-1196. doi: 10.1161/01.str.0000014774.88027.22

Zhang, D., Cheng, C., Wang, Y., Sun, H., Yu, S., & Xue, Y. et al. (2020). Effect of Vitamin D on Blood Pressure and Hypertension in the General Population: An Update Meta-Analysis of Cohort Studies and Randomized Controlled Trials. *Preventing Chronic Disease*, *17*. doi: 10.5888/pcd17.190307

Zilbermint, M., Hannah-Shmouni, F., & Stratakis, C. (2019). Genetics of Hypertension in African Americans and Others of African Descent. *International Journal Of Molecular Sciences*, *20*(5), 1081. doi: 10.3390/ijms20051081

Traylor, M., Farrall, M., Holliday, E., Sudlow, C., Hopewell, J., & Cheng, Y. et al. (2012). Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. The Lancet Neurology, 11(11), 951-962. doi: 10.1016/s1474-4422(12)70234-x