2018

Uncovering novel genes causing isolated gonadotropin releasing hormone deficiency using runs of homozygosity in outbred families

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

UNCOVERING NOVEL GENES CAUSING ISOLATED GONADOTROPIN RELEASING HORMONE DEFICIENCY USING RUNS OF HOMOZYGOSITY IN OUTBRED FAMILIES

by

ANNA KUTATELADZE

B.S., University of Virginia, 2016

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2018

Approved by

First Reader

James C. McKnight, Ph.D.
Associate Professor of Physiology and Biophysics

Second Reader

William F. Crowley Jr, M.D.
David K. Podolsky Professor of Medicine
Massachusetts General Hospital

ACKNOWLEDGMENTS

I would like to thank everyone in Dr. William F. Crowley Jr's lab for all of their support and feedback throughout the project. I would especially like to thank Dr. Margaret Lippincott M.D. for her guidance and mentorship throughout the project that was instrumental. I would also like to thank Dr. Rany Salem PhD for designing the QC protocol. Also, I would like to thank the research study participants who have made this study possible.

UNCOVERING NOVEL GENES CAUSING ISOLATED GONADOTROPIN RELEASING HORMONE DEFICIENCY USING RUNS OF HOMOZYGOSITY IN OUTBRED FAMILIES

ANNA KUTATELADZE

ABSTRACT

Background: Isolated GnRH Deficiency (IGD) is a rare Mendelian disorder, characterized by absent puberty and infertility. Over 30 causal IGD genes have been discovered, yet only 30-35% of IGD cases have a proven genetic etiology, highlighting the importance of new discovery methods. Homozygosity mapping, traditionally used to detect autozygous segments in inbred populations, has recently proven useful for gene discovery in outbred populations.

Methods: Strict quality control (QC) metrics for the SNP genotyped cohort and controls were used. Overall, 653 European IGD probands and 1,911 European controls remained. After pruning, 206,012 SNPs remained. Runs of Homozygosity (ROH) were found and a burden analysis was performed. Areas of known IGD genes, Expression Quantitative Trait Loci (eQTLs) of known genes and novel regions predicted to be associated with IGD were explored.

Results: While there was no statistical difference between the ROH burden in cases vs. controls, there was an unexpected trend towards a greater burden in the controls. Of the known gene locations and their eQTL regions, there was a trend in cases having a higher representation of ROHs within these regions. Four ROH regions unique to cases only

were also discovered. None of the individuals had a variant in the coding genes found in these regions.

Conclusions: SNP markers are found in coding regions and non-coding regions. This allows for a preview of potential disease causing variation in the non-coding region. Even though no coding gene of interest was found, the non-coding parts of the identified regions are of significant interest.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BU.....Boston University

CDP .....Constitutional Delay of Puberty

E.....Estrogen

eQTL..... Expression Quantitative Trait Loci

Froh..... Frequency of Run of Homozgyosity

FSH.....Follicle Stimulating Hormone

GnRH.....Gonadotropin Releasing Hormone 1

GnRHR .....GnRH Receptor

GWAS ..... Genome Wide Association Study

HA ..... Hypothalamic Amenorrhea

HPG Axis..... Hypothalamic Pituitary Gonadal Axis

HWE ..... Hardy-Weinberg Equilibrium

IBD .....Identical by Decent

IGD ..... Isolated GnRH Deficiency

KS .....Kallmann Syndrome

LD.....Linkage Disequilibrium

LH..... Luteinizing Hormone

MAF ..... Minor Allele Frequency

MGH.....Massachusetts General Hospital

nIHH .....Normosmic Isolated Hypogonadotropic Hypogonadism

PCA ..... Principle Components Analysis

PP.....Precocious Puberty

PsychChip..... Psychiatric Predisposition Microarray

QC.....Quality Control

REU .....Reproductive Endocrine Unit

ROH..... Run of Homozygosity

SNP.....Single Nucleotide Polymorphism

T..... Testosterone

TDT ..... Transmission disequilibrium test

VIF..... Variance Inflation Factor

WES..... Whole Exome Sequencing

WGS .....Whole Genome Sequencing

INTRODUCTION

Isolated Gonadotropin Releasing Hormone Deficiency

Isolated Gonadotropin Releasing Hormone (GnRH) Deficiency (IGD) is a mendelian disorder that results from the failure of the secretion or action of GnRH neurons. This failure disrupts the normal functioning of the hypothalamic-pituitary-gonadal (HPG) axis.

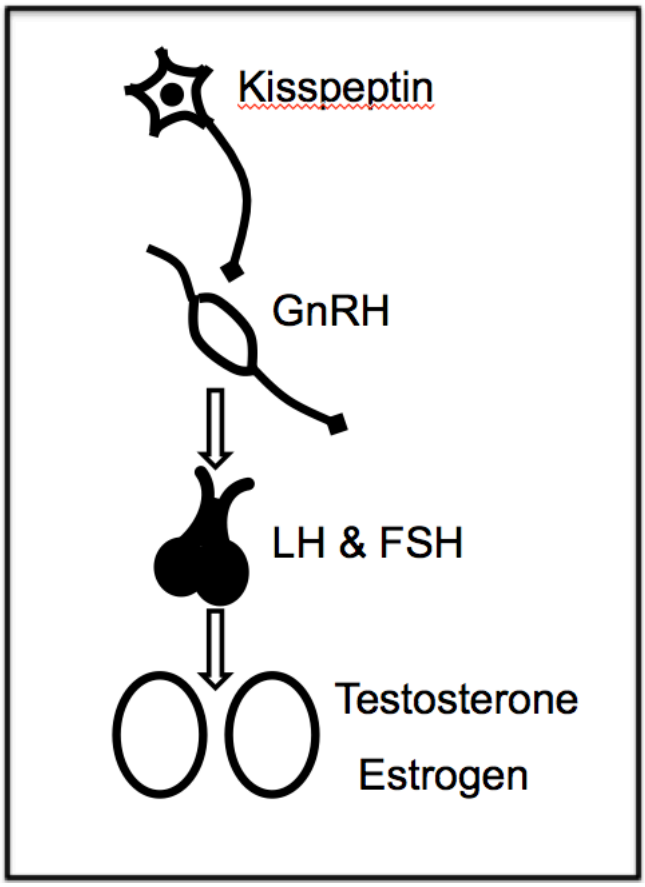


Figure 1: Kisspeptin + Hypothalamic-Pituitary-Gonadal Axis - Classic HPG axis plus kisspeptin. Although not discussed in this paper, kisspeptin released by kisspeptin neurons and their receptor GPR54 (KISS1R) were found by our group (Seminara 2003) to be essential for normal GnRH neuronal function.

The HPG axis (Figure 1), known as the reproductive axis, is an integrated endocrine signaling system that hormonally connects the hypothalamus, the anterior pituitary, and the gonads in both a 'feed forward' stimulatory system and a 'feedback' inhibitory system that act together to provide the delicate balance required for reproduction to proceed in a timely and physiological fashion at the appropriate developmental time. Starting cranial to caudal on this axis, GnRH neurons in the hypothalamus secrete pulsatile GnRH1, a decapeptide hormone, into the hypophyseal

portal capillaries that flow through the anterior pituitary (Amoss 1971; Matsuo 1971). From the hypophyseal portal capillaries, GnRH1 is able to bind to GnRH1 receptors (GnRHR1) on the anterior pituitary's gonadotrope cells. GnRHR1 is a 7-membrane-spanning, G-coupled protein receptor (GPCR) which activates the phosphatidylinositol-calcium secondary messenger and leads to the activation of protein kinase C (PKC) and eventually the physiological release of luteinizing hormone (LH) and follicle stimulating hormone (FSH) (Stojilkovic 1995; Naor 2000). Two genes for GnRH exist in mammals, GnRH1 and GnRH2, which bind to their respective receptors, GnRHR1 and GnRHR2. Not much, however, is known about the function of GnRH2 or its receptor other than that it appears to be a 'pseudogene' that is not transcribed in humans (Desaulniers 2017; Neill 2004). Therefore, this paper will refer to GnRH1 simply as GnRH and GnRHR1 as GnRHR.

LH and FSH are dimeric glycoprotein hormones that then travel through the vasculature and bind to their receptors on the ovaries in females and testes in males. FSH stimulates gametogenesis, i.e. follicular development in females and spermatogenesis in males, as well as the secretion of the dimeric gonadal protein

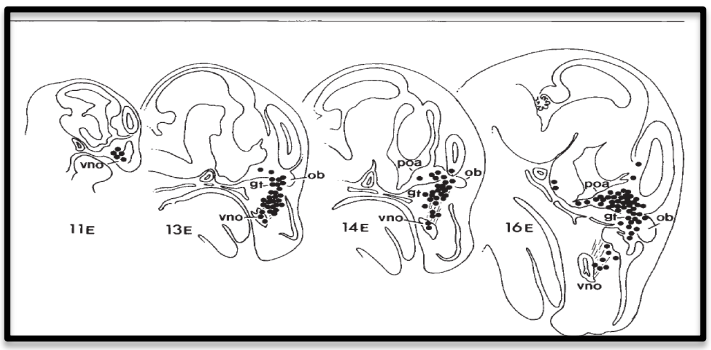


Figure 2: **GnRH neuronal migration in a mouse model**. The black dots represent GnRH neurons and the picture is divided into their location at embryological day 11-16. The GnRH neuron originates with the vomeralnasal organ (vno) and travels along the vomeralnasal and olfactory nerve to the olfactory bulb (ob) and eventually, by day 16, most GnRH neurons are located in the preoptic area (poa) of the hypothalamus. (Schwanzel-Fukuda 1989)

hormone, inhibin, that restrains FSH secretion in both females and males. LH stimulates the release of androgens from the gonads of both sexes: androgens that will eventually be converted to the sex steroid hormones estrogen (E) in females and testosterone (T) in males. These sex steroid hormones and inhibin together provide a negative feedback loop to restrain the output of the anterior pituitary gonadotropes of the HPG axis. As their levels increase, they inhibit gonadotrope secretion of LH and FSH and the secretion of GnRH from the GnRH neuron (Vadakkadath 2005).

The failure of the HPG axis in IGD most frequently occurs due to genetic defects within the hypothalamic GnRH neurons. Consequently, these patients typically present with hypogonadotropic hypogonadism characterized by low levels of gonadotropins (LH and FSH) and low (i.e. pre-pubertal) levels of sex steroids (E and T). In 1989, Schwanzel-Fukuda et al. found that GnRH neurons originate in the medial olfactory placode, i.e. extramural to the CNS, and subsequently migrate along the vomeronasal and olfactory nerves into the brain at the cribriform plate. From there, they proceed into the pre-optic area and eventually to the mediobasal area of the hypothalamus (Figure 2). There, the mature GnRH neuronal processes project into the hypophysial portal capillaries of the median eminence and secrete GnRH, which is eventually delivered by these capillaries to the anterior pituitary gland where they regulate pituitary hormone release from the gonadotropes (Schwanzel-Fukuda 1989, Wierman 2012).

IGD is heterogeneous in its phenotypic presentation. IGD patients can present with either a normal sense of smell (normosmia), known as normosmic isolated hypogonadotropic hypogonadism (nIHH), or they can present with the loss of the sense

of smell (anosmia), known as Kallmann Syndrome (KS). nIHH is prevalent in about 40% of IGD cases while KS accounts for about 60% of IGD cases (Balasubramanian 2011). IGD is a rare disease, with Kallmann Syndrome affecting 1:125,000 females and 1:30,000 males (Laitinen 2011). In addition to this difference in ability to smell, IGD patients also vary in severity and timing of reproductive defects and may or may not present with a variety of other non-reproductive phenotypes (Balasubramanian 2011).

Kallmann Syndrome (KS) was first documented in 1944 by Franz Josef Kallmann, an American geneticist who reported three different families affected with KS and, additionally, first noted normosmic IGD patients (Kallmann 1944). His observations thus first suggested a genetic etiology for IGD. In 1986, Ballabio et al. reported a boy with 3 different genetic diseases: KS, chondrodysplasia punctate and X-linked ichthyosis. His karyotype analysis and positional cloning revealed a large deletion on the X chromosome (Xp.22.3), inherited from the proband's mother. This study, along with others, defined this deletion that encompassed three genes—*KAL1* (now known as *ANOS1*), *ARSE*, and *STS*—that respectively caused the boy's KS, chondrodysplasia punctata, and ichthyosis, thus making *ANOS1* the first Kallmann Syndrome gene (Ballabio 1986; Bick 1989).

In 1989, Schwanzel-Fukuda et al, discovered that the GnRH neurons did not migrate normally in individuals with KS (Schwanzel-Fukuda 1989). In patients with the KS mutation, this migration is halted at the level of the cribiform plate, thereby affecting olfactory sensing and GnRH neuronal function. In 1991, *ANOS1* (*KAL1*) was shown to be

involved in neuronal cell adhesion and axonal path-finding thus officially establishing it as a KS-causing gene due to the inhibition of GnRH neuronal migration (Franco 1991).

As insightfully written by Franz Josef Kallmann in his 1944 article, "[t]here is no reason to assume that it must always be the same mechanism… or the same gene that causes the disease…." (Kallmann 1944). This quote emphasizes the genetic complexities of IGD that underlie our current understanding. Since the discovery of *ANOS1* (*KAL1*) in 1991, over 30 new genes have been shown, when mutated, to be sufficient to cause IGD. (Stamou 2015). Even with these advances in gene discovery, as shown in Figure 3, only about 30-35% of all IGD cases have a proven genetic etiology (Stamou 2015). This observation, along with the complexities and puzzles of puberty initiation itself, as shown in a special edition of Science Magazine that lists pubertal initiation as a top 125th unanswered scientific question (Science, July, 2005), implies that there are many more genes left to discover to have a clearer understanding of the pathophysiology behind IGD and ultimately the physiology behind puberty initiation.

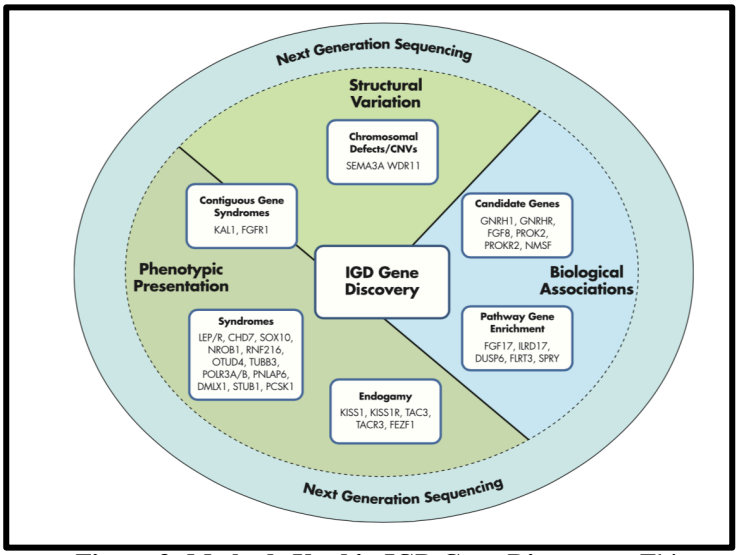


Figure 3: Methods Used in IGD Gene Discovery. This includes newer, faster strategies through new generation sequencing including whole exome sequencing and whole genome sequencing that are revolutionizing genetic discovery. (Stamou 2015)

Among the 30+ genes we now associate with IGD, there are two main aspects of the GnRH neuronal function that can be impaired and cause the disease (Figure 4). First, a genetic mutation in a particular gene, such as *ANOS1* (*KAL1*), can cause a neurodevelopmental failure that prevents normal GnRH neuronal migration.

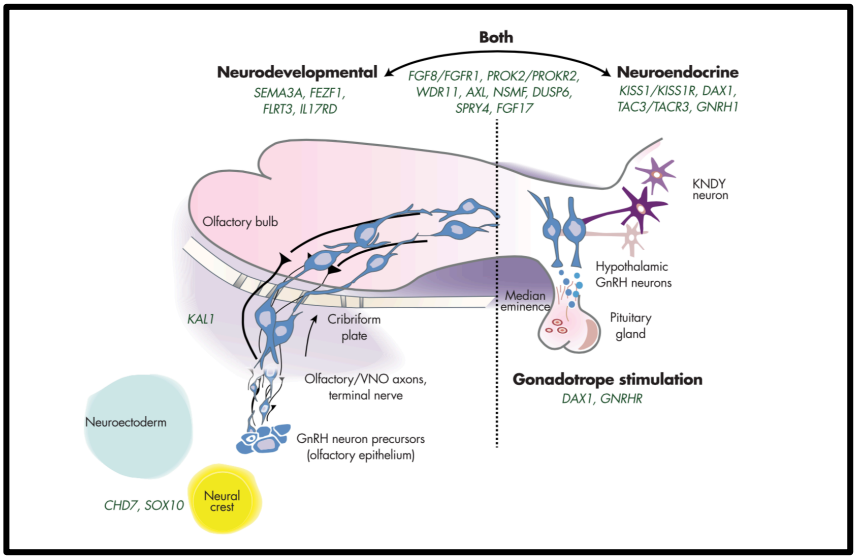


Figure 4: The Complexity of GnRH Migration and Signaling. GnRH neurons form from neuroectoderm and evidence also points to a subset forming from neural crest (Whitlock 2003). This picture shows on the left genes involved in neurodevelopment that when mutated, cause a failure in GnRH migration. This picture shows on the right genes involved in neuroendocrine function and when mutated, impair GnRH signaling. In the center of the picture are genes involved in both processes (Stamou 2015).

Second, a genetic mutation in a particular gene can cause a neuroendocrine failure, thereby preventing proper signaling of GnRH, such as Kisspeptin/GPR54 (*KISS1/KISS1R*). Some IGD-causing genes have also been shown to participate in both phenotypes of IGD (anosmic/normosmic) including Prokineticin 2/Prokineticin 2 Receptor (*PROK2/PROKR2*). Along with their complex regulatory functions, these IGD genes also exhibit variable penetrance/expressivity within families and among different families with the same mutations (Pitteloud 2007). Other phenotypes that share at least some genetic links to IGD are shown in Table 1 below.

Phenotype	Definition	IGD genetic link
Constitutional Delay of Puberty (CDP)	No signs of puberty in girls by age 13 or boys by age 14 but spontaneous puberty does occur before the age of 18 (Klein 2017)	In IHH pedigrees, 53% of CDP family members share the proband's mutation (Zhu, et al. 2015)
Precocious Puberty (PP)	Onset of puberty before age 8 in girls and before age 9 in boys (Klein 2017)	KISS1 misssense mutations were reported (Silveira, et al. 2010)
Hypothalamic Amenorrhea (HA)	Disruption of the HPG axis due to over-exercise, increased stress, or malnutrition (Liu 2008)	Reported misssense mutations found in HA patients include FGFR1, PROKR2, GNRHR, and KAL1. (Caronia, et al. 2015)
Isolated Congenital Anosmia	Inability to smell (Karstensen 2012)	The same PROKR2 mutations associated with KS in families with isolated congenital anosmia. (Mova-Plana et al, 2012)

Table 1: Phenotypes Genetically Related to IGD

Homozygosity mapping

Homozygosity mapping is a genetic discovery tool (classified under endogamy in Figure 3) that traditionally looks at haplotype blocks (a group of genes inherited from a single parent) passed down in consanguineous families such that an affected individual's consanguineous parents will each be a carrier of the same recessive gene mutation(s) that is/are identical by descent (IBD). This mutation will then be passed down to their affected child so that this child is homozygous for the recessive mutated gene (Ceballos 2018). As seen in Figure 5, the grandmother of the affected child's parents is a carrier for the recessive mutation. Despite crossing-over events, the majority of her original haplotype block, that includes the mutated allele, is passed down to both of the affected child's parents, and eventually, passed down in a homozygous manner to their affected

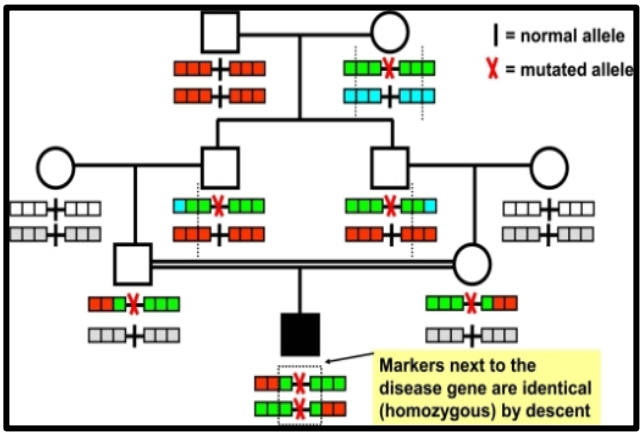


Figure 5: Traditional Homozygosity Mapping (Hildebrandt 2009)

child (Hildebrandt 2009).

Homozygosity mapping, therefore, allows for the discovery of homozygous mutations in recessive genes that are passed down by both of the affected individual's parents and the mutations are said to be IBD. These regions are said to be autozygous. The power in this method lies in the fact that detection of these haplotype blocks relies on identified linkage markers that are scattered throughout the genome with a majority (~80%) spanning the non-coding region. Botstein first described the use of these polymorphic marker loci in 1980 (Botstein 1980) that could serve as 'genome addresses' that could be used to identify physical positions on chromosomes. These identifiers include microsatellites and single nucleotide polymorphisms (SNPs) (Ball 2010). Bolino et al., in 1996, were among the first people to use long consecutive stretches of homozygous markers in a consanguineous family to predict regions of autozygosity (Bolino 1996). This phenomenon is known as a run of homozygosity (ROH). Between 1995 and 2003, almost 200 studies were published that used this 'homozygosity mapping' in consanguineous families to discover rare recessive mutated genes (Botstein 2003). In our center, an IGD gene mutation in GPR54 (KISS1R) was found in a

consanguineous family in a haplotype block that was 1.06 Mb long and located on chromosome 19p13.3 (Seminara, Crowley 2010).

Empirical studies have shown that there is a burden of homozygous deleterious variants in regions of homozygosity (ROH) compared to non-deleterious variants (Szpiech 2013). This observation increased the importance in detecting these ROHs, not just in consanguineous families, but also within non-consangineous families. Given the size of the human population, we are all inbred to some degree and therefore ROHs should be seen in all individuals. The typical size of a ROH in a consanguineous family, however, is relatively large (Figure 5) when compared to an outbred population (Figure 6) since the common ancestor of the consanguineous parents is comparatively recent (Ceballos 2018). ROHs, however, are seen in the general population and, on average, an individual's genome will encompass around 10 ROHs (McQuillan 2008). The amount of ROHs in a population is dependent on the effective population size (size of the reproductively capable population), where smaller populations tend to have more ROHs (Ceballos 2018). As next generation sequencing techniques have improved over time, a clearer and denser picture of the genome is emerging that allows for improved detection

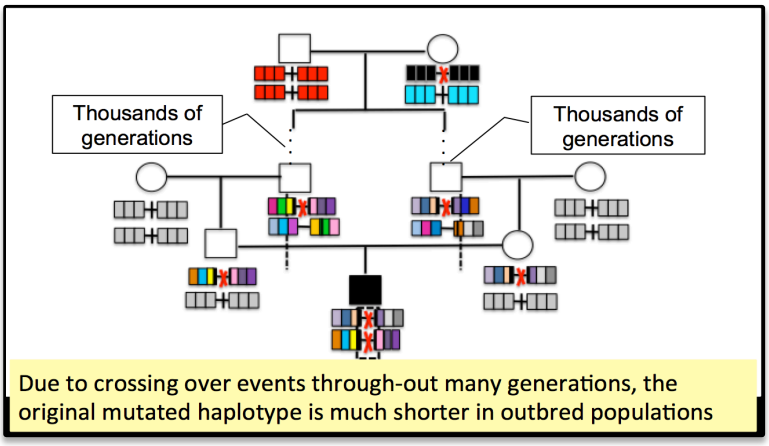


Figure 6: Homosygosity mapping in an outbred family.
Modeled after pedigree from Hildebrandt et al. 2009

of these ROHs in non-consanguineous families. In 2002, the International HapMap Project identified over 3 million common SNP variations in 269 individuals from 4 populations in its phase 1 and 2 studies. In phase 3, 1,301 individuals were analyzed from 11 different populations, which allowed high density single nucleotide polymorphism (SNP) arrays to become more widely available (International HapMap Constortium 2007).

These advances have, in turn, enabled researchers to make greater use of homozygosity mapping in outbred populations for gene discovery, including Keller et al., who showed an increased burden of ROHs in schizophrenia when compared to a control population (Keller 2012). Since IGD is a rare reproductive disorder, it would therefore be predicted to be under high negative selection (or purifying selection). Therefore, it might be expected that the disease risk alleles leading to IGD would be enriched for recessive mutations. Hence, homozygosity mapping should be an ideal tool to help uncover some of these recessive disease risk alleles (Keller 2012; Charlesworth 2009). Furthermore, SNP arrays also have the advantage of covering a large majority of the non-coding regions of the genome allowing this technique to theoretically detect potential modifier regions within the non-coding transcript that have previously been unassociated with IGD.

Specific Aims

Specific aims of the following thesis are:

1. Determine whether there is an increased burden of areas of autozygosity in unrelated probands of European descent when compared to control individuals of European descent.
    a. If validated, it would imply that the more ROHs an individual has, the higher risk they are of inheriting IGD. Inbred individuals, therefore, would have an increased risk for IGD when compared to outbred individuals since inbred individuals have larger runs and numbers of ROHs.
2. Investigate whether there are regions of shared SNPs (ROH) that harbor novel homozygous mutations in known genes or regulatory regions of these known genes that contribute to IGD.
    a. By using homozygosity mapping and pinpointing known gene locations and regulatory regions of these known genes, there is a potential for uncovering novel mutations within these areas.
3. Investigate whether there are regions of shared SNPs (ROH) that harbor novel homozygous mutations in novel genes or novel regulatory regions that contribute to IGD.
    a. By using homozygosity mapping, there is a potential for uncovering novel genes, novel pseudo genes, novel enhancer/repressor regions, novel RNAs or other novel transcripts within the coding or non-coding

region that are currently unknown to be associated with IGD but are found to be causal of IGD.

METHODS

Case Cohort

All participants had been recruited with approval from the Institutional Review Board of Partners Healthcare over a 40-year period and had given their informed consent to participate in a genetic study to understand the causes of hypogonadotropic hypogonadism and the genetics of human reproduction. 2,409 individuals were selected for genotyping analysis as they had a reproductive phenotype or were family members of an individual who exhibited a reproductive phenotype. SNP genotyping was performed at the Broad Institute from genomic DNA extracted from peripheral blood samples. For this analysis, the only reproductive phenotype considered was hypogonadotropic hypogonadism (nIHH or KS).

Control Cohort

2,199 controls from The Schizophrenia Psychiatric Genome-Wide Association Study were provided by Benjamin Neale, PhD. They also had undergone standard QC metrics and the Broad Institute as seen in the Psychiatric GWAS Consortium 2011.

SNP Genotyping: Psychiatric predisposition microarray

All SNP genotyping was performed at the Broad Institute from genomic DNA extracted from peripheral blood samples. Since IGD is a brain-based disease, the PsychArray-24 SNP platform (V1.1 and B) from Illumina was used for genotyping. This PsychChip includes ~271,000 proven tag SNPs from the Infinium HumanCore-24 Beadchip, ~277,000 markers from the infinium Exome-24 BeadChip and ~50,000 markers

associated with common psychiatric disorders (i.e. brain-based disease). The V1.1 version of the SNP platform includes 22,196 new SNPs not seen in the B version and 4,832 SNPs that were present in the B version that are missing from the V1.1 version. 1,044 individuals were genotyped using the V1.1 version while 1,365 individuals were genotyped using the B version. For this analysis, only shared SNPs between the two platforms were used. Three genotype calling algorithms were used to call variants and include Birdseed from the Birdsuite software, Autocall from the GenCall software, and Zcall. Birdseed and Autocall both excel in calling common variation (Minor allele frequency (MAF)>0.01) while Zcall was designed specifically for calling rare variants (MAF<0.01).

Audit of Cases

Of the 1,403 probands with either nIHH or KS (732/1403 with nIHH and 671/1,403 with KS) whose DNA samples had all undergone SNP genotyping using the PsychChip at the Broad Institute, every 20th individual was selected (72/1,403 individuals with 39/72 individuals having an nIHH phenotype and 33/72 individuals having a KS phenotype) to undergo an extensive records review which included evaluating and confirming each individual's documented sex, diagnosis, age, clinical characteristics that led to their diagnosis, including their hormone levels, pubertal development, imaging and smell phenotype, and any other recorded phenotype that had been given. The quality of the recorded phenotyping information was then assessed. If needed, a cases status was updated prior to case vs. control analysis. Only one individual's phenotype was updated.

Sample Processing and Initial Quality Control

Once genotyping data was returned from the Broad Institute's Core, every individual was validated for his or her correct identification number, correct sex, correct parent identification numbers if an individual's parents are in the study, and correct phenotype status (KS/nIHH phenotypes being the only individuals listed as affected). Everyone else, including individuals with other reproductive phenotypes, were labeled as unaffected for the purposes of this study. Once complete, individuals were removed who did not have a clear SNP call in each of the three calling algorithms. Since two versions of the PsychChip were used, pre-quality control measures, as well as some QC measures, were performed on each of the versions separately before merging them (Figure 7). PLINK (Purcell 2007) was used to perform all QC analysis, aside from principal component analysis (PCA), where Population Structure Inference in KING (Manichaikul 2010) was used. Of the 1,044 individuals genotyped on the V1.1 version of the PsychChip, 39 individuals did not pass all three calls and were removed, leaving 1,005 individuals. Of the 1,365 individuals genotyped on the B version of the PsychChip, 216 did not pass all three calls and were

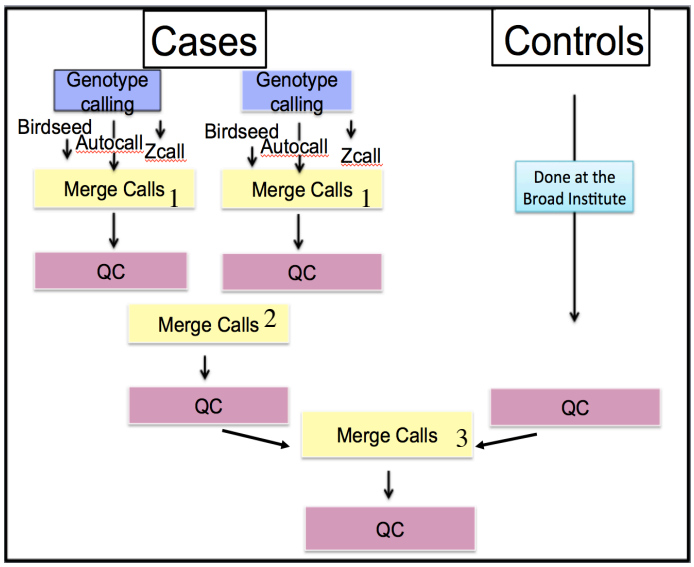


Figure 7: QC outline. The two columns under cases represent the two versions of the Illumina SNP chip used.

also removed, leaving 1,137 individuals. SNPs with a call rate of less than 97% were then removed from birdseed and autocall algorithms only (best suited for common variant calling) with an average of around 35,000 SNPs lost. Since Z-call performs ideally on rare SNPs (MAF<0.01), only the rare SNPs were kept on the Z-call algorithm before QC could be done. Individuals (n=4) with a call rate of less than 95% were also removed. Since birdcall and autocall perform more reliably on SNPs with MAF>0.01, they were subsequently merged together. SNPs from this merged data set with MAFs<0.01 (rare SNPs) were removed but kept as the only SNPs in the zcall calling algorithm. SNPs from the merged birdseed/autocall data set with a call rate of less than 97% were then removed from the merged birdseed/autocall data set and SNPs with a call rate of less than 97% in the zcall data set were removed from the zcall data. An average of 2,829 SNPs were removed between the two sets.

Quality Control on merged calls 1 (Figure 7: Merge Calls 1)

Once all three calls were merged on each of the two PsychChip versions, QC was as follows: a) SNPs with a call rate < 95% were removed using PLINK --missing argument to provide a 'missingness rate' for each SNP through generation of an lmiss dataset. b) Individuals with a call rate of < 98% were removed again using PLINK --missing argument to provide a missingness rate for each individual through generation of an imiss dataset. c) Inbreeding coefficient was determined using PLINK's --het argument to generate a .het text file. Individuals who fell out of the -0.2-0.2 frequency of heterozygosity (FHET) inbreeding metric range were then removed. d) The sex of each individual was confirmed using PLINK --check-sex argument. No SNPs remaining had a

call rate of less than 95%. Three individuals in the V1.1 version of the PsychChip and two individuals in the B version of the PsychChip did not pass the call rate greater than 98% and were subsequently removed. Two individuals from the V1.1 version of the PsychChip and three individuals from the B version of the PsychChip fell outside the FHET for inbreeding coefficient range and were removed. Finally, 15 individuals were removed from the V1.1 version of the PsychChip and 101 individuals were removed from the B version of the PsychChip for individuals that were either male but had an FHET Sex Determination Metric >0.8 or were female but had an FHET Sex Determination Metric <0.4.

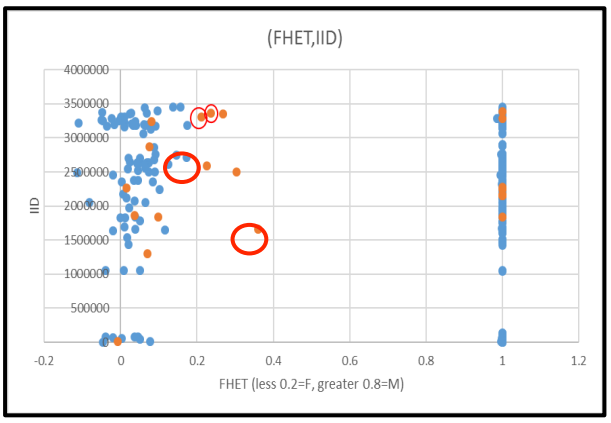


Figure 8: Example of FHET for SEX Determination Distribution After A Sex Check was Performed on Newer PsychChip. Distribution of FHET per individual. Traditionally, females with FHET less than 0.2 are considered to have passed the sex check but for our analysis, females with an FHET less than 0.4 were considered to have passed the sex check (circled in red) and males with an FHET greater than 0.8 were considered to have passed the sex check. The orange dots less than 0.4 that are not circled in red were males and did not pass the sex check and the orange dots greater than 0.8 were female and did not pass the sex check.

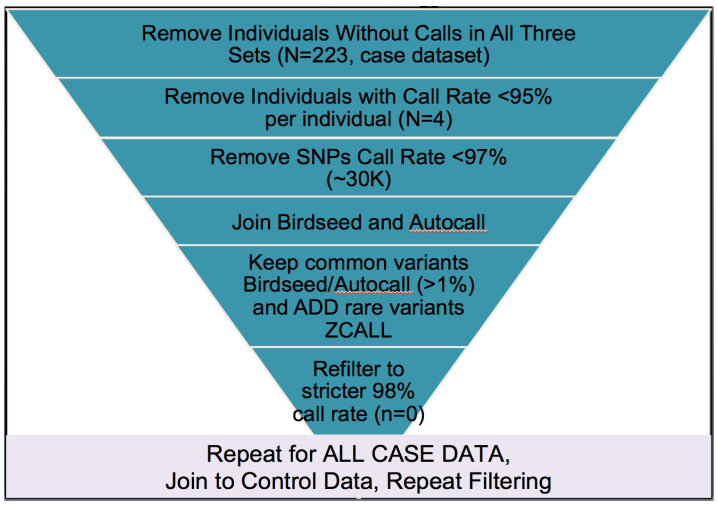


Figure 9: Summary of QC.

Traditionally, a female should have an FHET sex determination metric <0.2 but for this analysis, due to the distribution of the plot in Figure 8, individuals who were female with an FHET sex determination metric<0.4 passed the sex check (circled in red in Figure 8). A general summary of the QC process is seen in Figure 9.

Quality Control on merged PsychChips and Controls (Figure 7: Merge Calls 2 and 3)

The two different PsychChip versions had some SNPs that were located in the same location but given a slightly different name. Therefore, a list of SNP locations was generated with just one SNP name to represent that location. Thus, when these two SNPChip versions were merged, there was only one name for each SNP per location. Once the two SNPChip versions were merged using PLINK's --bmerge argument, two multiallelic SNPs were removed and 553,926 SNPs remained with 1,996 total individuals that were either probands, family members of probands, or other non-IGD phenotype individuals from the IGD cohort. The Schizophrenia Psychiatric Genome-Wide Association Study Consortium controls were then merged with the IGD cohort using PLINK's --bmerge argument. There were a total of 554,631 SNPs with 4,195 individuals. The QC at this step was as follows: a) SNPs with a call rate of < 95% were removed. b) Individuals with a call rate < 98% were removed. 120,083 SNPs had a call rate <95% and were removed, leaving 430,288 SNPs. All individuals passed with a call rate >98%.

PCA: European Population

Principle component analysis was done using --pca 5 argument using the Population Structure Inference in KING (Manichaikul 2010). The merged IGD cohort and Schizophrenia Psychiatric Genome-Wide Association Study Consortium controls were merged with HapMap3 (International HapMap Constortium 2007) and coordinates that correlate with ethic SNPs that define African populations (ASW, LWK, MKK, YRI), an

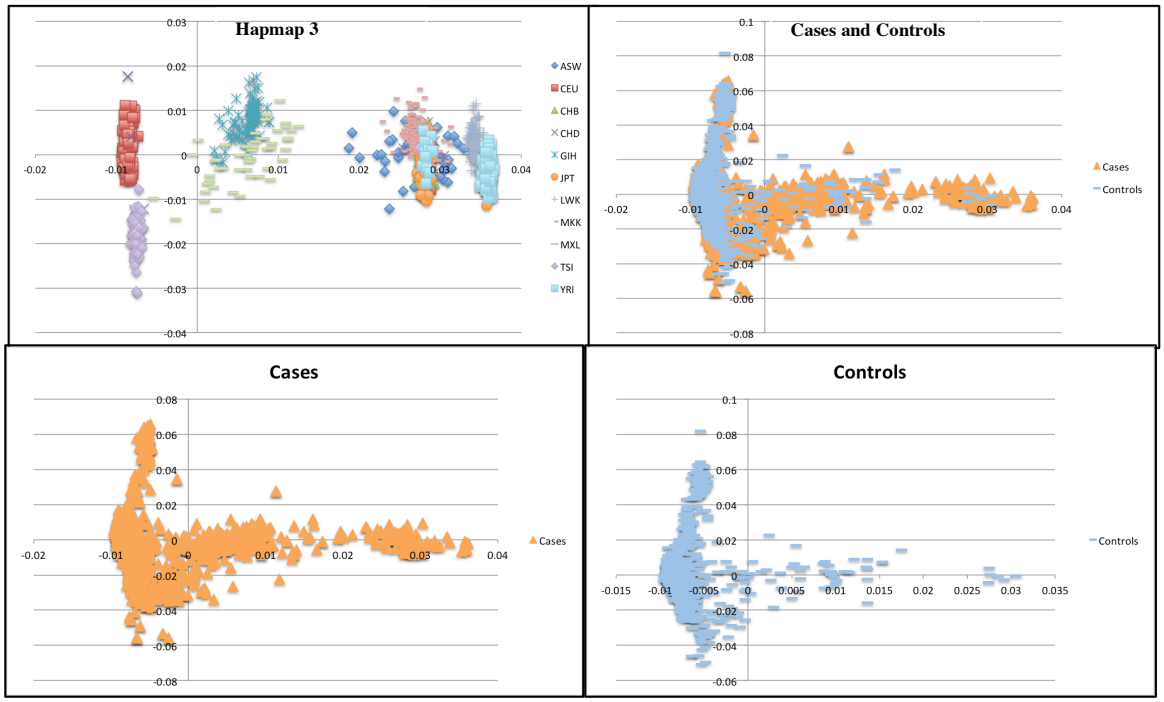


Figure 10: Representative Example of a Graph used to Define Different Ethnic Populations. The same coordinates are depicted with only the hapmap 3 individuals in the top left, the cases and controls together in the top right, the cases only in the bottom left, and the controls only in the bottom right.

Indian population (GIH), a Mexican population (MXL), East Asian populations (CHB, CHD, JPT) and European populations (CEU, TSI) were determined. Figure 10 shows an example of the graphs used to define the European populations. Two different HapMap3 populations define the Europeans, the Utah residence from Northern and Western European Descent (CEU) and the individuals from Tuscany, Italy that represent a

Southern European Descent. As shown in Figure 10, the Hapmap 3 graph depicts the CEU population with red squares in the upper left corner and the TSI population below it with purple diamonds on the lower left side. As shown in the three other graphs, the majority of the cases as well as the controls are seen roughly equally between the Northern/Western European Population and the Southern European Population. European populations from the merged IGD cohort and Schizophrenia Psychiatric Genome-Wide Association Study Consortium controls were then extracted. Out of the 4,195 total individuals, 3,130 individuals were isolated as European with 1,219 individuals from the IGD cohort and 1,911 individuals from the Schizophrenia Psychiatric Genome-Wide Association Study Consortium controls. A new data set with only European IGD probands was created for this analysis (to minimize the influence of any population stratification) plus all the European Schizophrenia Psychiatric Genome-Wide Association Study Consortium controls. There were 653 European probands and 1,911 European Schizophrenia Psychiatric Genome-Wide Association Study Consortium controls. Hardy-Weinberg Equilibrium (HWE) was then performed on the European dataset using PLINK's --hardy argument and removing any SNPs that had a $P<1e^{-7}$ and therefore would deviate from what would be expected in a normal population. 291 SNPs (115 of these SNPs had a MAF<1%) were removed that failed HWE, leaving 429,937 SNPs. Finally, a missingness by phenotype analysis using PLINK's --test-missing argument was performed to remove any SNP that had a $P<1e^{-7}$ and would therefore be under-represented in either the controls or the cases and could result in false ROHs. 452 SNPs

failed the missingness by phenotype of P<1e-7 and were removed, leaving 429,485 SNPs.

ROH Calling Using European Only Cohort

To remove SNPs in linkage disequilibrium (LD) with one another that would therefore be on the same haplotype block and cause false positives, PLINK's pruning function was used. The general pruning method uses a window size of 50 SNPs that moves every 5th SNP. Two different pruning methods were used: i) pairwise-pruning which considers the given correlation between two SNPs. If there is a significant pairwise genotypic correlation between the two, the SNP with the lower MAF will be removed in order to preserve the SNP that would be present in the population at a higher frequency and will be more informative about the haplotype block it represents; ii) pruning using the same method as pairwise with the additional use of the variance inflation factor (VIF) which accounts for non-linear combinations of SNPs and is therefore a more stringent pruning method. VIF is $1/(1-R^2)$ where $r^2>0.5$. Both pruning methods were performed using PLINK. The pairwise pruning was done using PLINK's --indep-pairwise argument, removing 164,963 SNPs and leaving 264,522 SNPs. The pruning taking into account the additional VIF was done using PLINK's --indep argument, removing 223,473 SNPs and leaving 206,012 SNPs. Runs of homozygosity (ROH) were found using PLINK's --homozyg argument and specifying for no heterozygous SNPs in a run, a minimum amount of 65 SNPs/run (Howrigan 2011), a minimum density of 50 kb SNPs (1 SNP every 50 kb) per run, overlapping ROHs matching by at least 95%, and finally, the maximum homozygosity gap between two SNPs in a run being 500kb, since larger

distances between two SNPs in a run could include areas such as the centromere in the chromosome. If two SNPs in a ROH have a distance larger than 500 kb between them, the ROH was broken into two different ROHs. Initial ROH analysis was performed on both the pairwise LD pruned dataset (33,054 ROH) and the VIF LD pruned dataset (34,893) to compare the two but due to the overall outperformance of the VIF LD pruning when compared to pairwise LD pruning, VIF LD pruning was eventually the only dataset used (Howrigan 2011).

ROH Burden Analysis in the European Only Cohort

A frequency histogram of the number of individuals with a specific number of ROHs was plotted in the European cases versus the European controls. An unpaired two-tailed T-test using GraphPad software was performed to determine if a significant result was achieved. The 'Froh,' or the proportion of the genome of an individual that is part of a ROH found by addition of the total length of all an individual's ROH in kb divided by the total SNP-mappable autosomal distance in our dataset of 2.97×10^6 kb (Keller 2012), of each individual was also determined and the average Froh in cases vs the average Froh in controls was calculated. An unpaired t test using GraphPad software was then performed to compare the average Froh in the cases vs. the average Froh in the controls. A histogram to visually assess the number of individuals with a certain length of ROH in cases vs. controls was also graphed using the VIF dataset only.

Homozygosity Mapping Analysis on Known Genes and their Predicted Regulatory Sites in the European Only Cohort

All individuals' ROHs were examined in the VIF LD pruned data to see if any ROHs overlapped with the locations of the 27 genes known to cause IGD when mutated (Balasubramanian 2003). The number of cases and controls with a run that spans a known gene were then recorded. Even though whole exome sequencing (WES) had been done on all of the SNP genotyped individuals in our cohort, there is a chance that a variant in a gene was missed. Therefore, if the location of a known IGD gene was found in a ROH of a case subject, that individual was queried for a variant in that gene. This step is also a good way to confirm the method used if we are able to uncover known homozygous mutations in individuals. The locations of significant cis-expression quantitative trait L\loci (cis-eQTLs) for the same 27 known IGD genes were found through GTEx's Gene eQTL Visualizer (GEV) summarized using a heat map (GTEx Consortium 2017) that signify a potential regulatory region. A eQTL is marked by a SNP and is associated with the expression level of the target gene. A cis-eQTL is found less than 1 Mb away, either downstream or upstream, from the target gene's transcription start site (GTEx Consortium 2017) while a trans-eQTL would be defined as a variant 5Mb away, either downstream or upstream, from the target gene (Nica 2013). A significant cis-eQTL in this study is defined as having a false discovery rate (FDR) of <5% and an absolute effect size, determined by effect of the alternate allele relative to the effect of the reference allele, greater than 0. The significance metric was relaxed for this study since, again, these locations were used to identify potential areas of interest and false positives

in this case are less concerning than false negatives. Locations of cis-eQTLs for this study were only analyzed if they were expressed in brain tissue or pituitary tissue since IGD is a neural disease and also related to the pituitary due to the neuroendocrine connection between the hypothalamus and the pituitary.

Homozygosity Mapping Analysis in Novel Regions of the Genome in the European Only Cohort

ROHs most likely to contain novel genes or regulatory regions were deduced by analyzing homozygous SNPs seen only in the cases and not in the controls in the VIF pruned data set. Statistically, the likelihood of 3 probands sharing a ROH not seen in controls is relatively high and as calculated by a two-tailed Fisher's exact test. The P value for 3 probands sharing a unique region not seen in any controls is 0.017. On the other hand, the likelihood of 4 probands sharing a ROH not seen in controls becomes more significant. Using the same two-tailed Fisher's exact test, the P value for this becomes 0.004. Therefore, only regions shared by 4 or more probands and no controls were analyzed as unique and potentially informative areas. These regions were then queried through the University of California Santa Cruz Genome Browser (Kent 2002) for coding genes and non-coding regulatory areas in these regions. Each proband that had a ROH that encompassed one of these unique regions was queried for a misssense or loss-of-function variants (frameshift mutations, insertions/deletions, or splice site altering mutations) with the variant having a MAF<10% in the gnomAD genome browser (Lek 2016) to ensure low frequency in the unaffected population and with a depth score of >5

to ensure accurate readings in the coding genes within that region using Gemini: a flexible framework for exploring genome variation (Paila 2013).

RESULTS

ROH Burden Analysis in the European Only Cohort

The number of ROHs seen in the European probands (cases) only was 8,232 when pairwise LD pruning was applied and 8,713 when VIF LD pruning was applied. This means that on average, each proband had ~12.61 ROHs when pairwise LD pruning was applied and ~13.34 ROHs when VIF pruning LD pruning was applied. The total number of ROHs seen in the European controls was 24,822 when pairwise LD pruning was applied and 226,180 when VIF LD pruning was applied. This means that on average, each control had ~13.00 ROHs when pairwise LD pruning was applied and ~13.70 ROHs when VIF LD pruning was applied. Thus, there is no significant difference between the average ROHs per individual in cases vs controls.

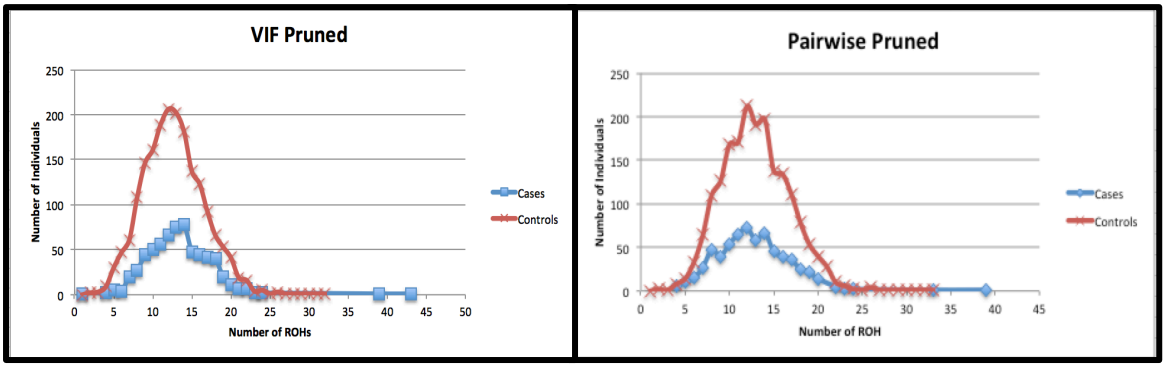


Figure 11: Frequency graph in European cases vs. European controls in both the pairwise pruned and VIF pruned data sets. P-value of the pairwise pruned cases vs. controls was 0.054 and p-value for the VIF pruned cases vs. controls was 0.024.

The frequency histogram of the number of individuals with a specific number of ROHs in the European probands (cases) vs. the European controls was then plotted in both the pairwise pruned data set and the VIF pruned data set, as shown in Figure 11, to visualize the distribution of each population. An unpaired two-tailed T-test was then performed for

the cases vs. the controls using GraphPad software in both the pairwise LD pruned data set and the VIF LD pruned data set. The p-values for both were not significant at $p<0.01$ with the pairwise LD pruned data set having a p-value of 0.054 and the VIF pruned data set having a p-value of 0.024. The average Froh of the cases was compared to the Froh of the controls, which takes into account not only the number of ROHs of the individuals, but also the length of these ROHs. The length corresponds to the amount of inbreeding an individual has in their family where inbred individuals have longer haplotype blocks and therefore a larger Froh. The average Froh of the cases in the pairwise LD pruned data set was 0.0064 and the average Froh of the controls in the pairwise LD pruned data set was 0.0068. An unpaired T test using GraphPad software was then used to compare the two values and the two-tailed p value equaled 0.072 and was not significant at a $p<0.01$. The same calculation was performed on the VIF pruned data set. The average Froh of the VIF pruned cases was 0.0068 and the average Froh of the VIF pruned controls was 0.0072. The two-tailed p-value for the unpaired T-test was 0.089 and therefore also not significant at a $p<0.01$. A histogram was used to visually see the distribution of the sum of all ROHs per individual in cases versus controls in the VIF pruned data set depicted in

figure 12.

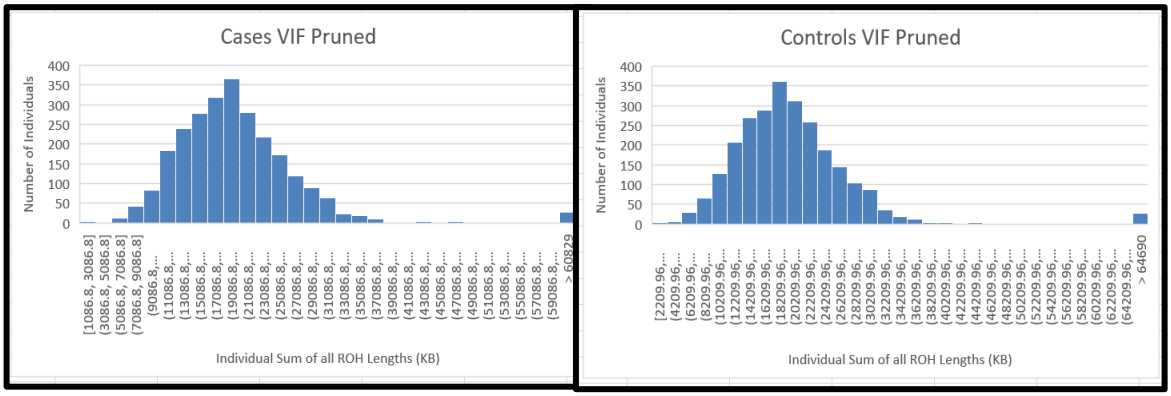


Figure 12: Histogram of Sum of All ROH Lengths

Homozygosity Mapping Analysis on Known Genes and their Predicted Regulatory Sites in the European Only Cohort

The locations of 26 autosomal genes known to cause IGD (7 of them being more common causes) as per Balasubramanian 2003 (Balasubramanian actually lists 27 genes but *ANOS1*, previously known as *KAL1*, is located on the X chromosome, which was not analyzed in this study) were recorded. ROHs in the European cases and controls that overlapped with one of these known 26 genes were recorded as shown in Table 2.

Most common					
Gene	Chr	Location beginning	Location end	Cases with ROH spanning region standardized to 500 cases	Controls with ROH spanning region standardized to 500 controls
CHD7	8	61,591,324	61,780,586	0.8	0.5
FGFR1	8	38,268,656	38,325,363	2.3	0.8
GNRHR	4	68,603,099	68,621,804	3.1	2.4
IL17RD	3	57,124,010	57,199,403	3.8	7.3
PROKR2	20	5,282,686	5,295,023	0.0	0.0
SOX10	22	38,368,319	38,380,539	0.0	1.6
TACR3	4	104,510,625	104,640,973	6.1	2.6
Less common					
AXL	19	41,725,108	41,767,671	0.8	0.8
CCDC141	2	179,694,484	179,914,786	1.5	0.8
DUSP6	12	89,741,837	89,746,296	0.0	0.3
FEZF1	7	121,943,712	121,950,131	3.8	3.4
FGF8	10	103,529,887	103,535,759	23.7	28.3
FGF17	8	21,900,428	21,906,319	0.0	0.0
FLRT3	20	14,304,639	14,318,313	0.8	0.8
GHRH1	20	35,879,490	35,885,311	4.6	3.9
HS6ST1	2	129,023,054	129,076,171	0.8	2.1
KISS1	1	204,159,469	204,165,619	0.8	1.0
KISS1R	19	917,342	921,015	0.0	0.5
POLR3B	12	106,751,436	106,903,976	0.0	0.8
PROK2	3	71,820,806	71,834,357	0.0	0.0
SEMA3A	7	83,587,659	83,824,217	0.0	0.5
SEMA3E	7	82,993,222	83,278,479	0.0	0.8
SPRY4	5	141,689,992	141,704,620	0.0	0.3
SRA1	5	139,929,652	139,937,041	9.2	8.9
TAC3	12	57,403,781	57,410,344	6.9	5.0
WDR11	10	122,610,687	122,669,038	2.3	0.0

Table 2: Number of Cases and Controls with ROHs that Overlap Known IGD Genes. (Balasubramanian 2003). Numbers of cases and controls are standardized to per 500 individuals for an accurate comparison taking into account population size.

Since all individuals from the cases were whole exome sequenced, their genomes were queried for the known gene that was contained within one of their ROH. Two of the cases, (IDs: 47601 and 1661001), had a ROH encompassing a more common causal IGD gene, *GNRHR*. When these individuals were queried in GEMINI, each was found to carry a homozygous mutation in GNRHR (p.Q106R hom and p.R262Q hom respectively). All other cases, when queried for their respective gene, did not have a variant in that gene. One of the advantages to using SNP genotyping is it allows us to preview information in

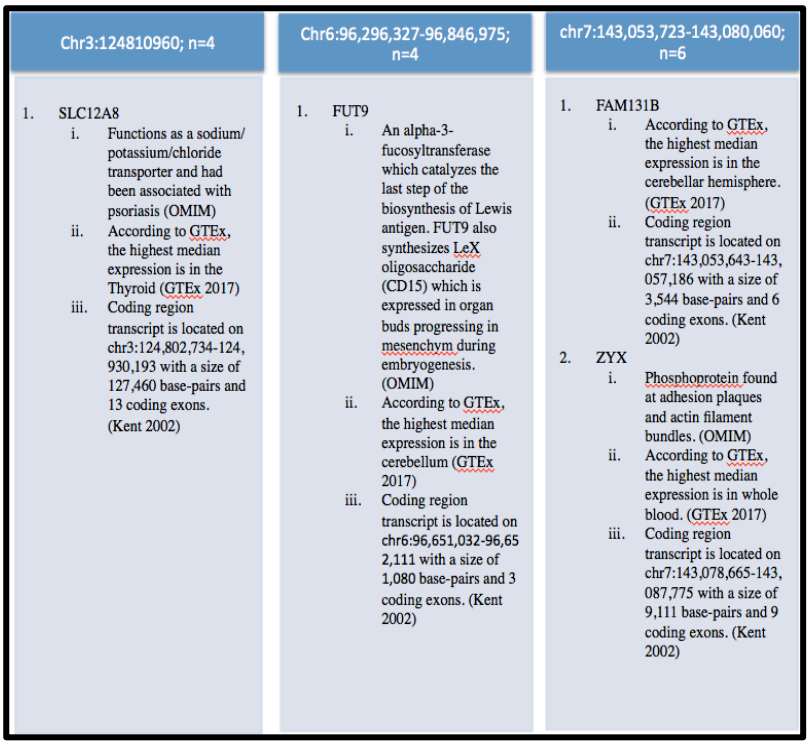
the non-coding regions of the genome. Our current understanding about variants found in the non-coding regions is still in its infancy; however, one useful place to start to understand these potential defects is looking at eQTLs or expression quantitative trait loci that can theoretically cause expression changes in their target gene. The Gene eQTL Visualizer in GTEx was therefore used to determine the locations of cis-eQTLs for the 26 known genes that affected expression levels in any brain tissue or in the pituitary. 19/26 genes had eQTLs that affected expression in brain tissue or in the pituitary. The total genomic location that included all eQTLs for a specific gene in the tissue specified, the largest positive effect size or largest negative effect size or both, if that applied, of an eQTL within that area, the total number of eQTLs within the region, and the number of cases/controls with ROHs that span all of the region were recorded as shown in Table 3.

Most Common								
eQTL Target Gene	**Chr**	**Tissue with eQTL**	**Number eQTLs**	**Most pos/most neg eQTL Effect Size**	**Location First eQTL**	**Location Last eQTL**	**Cases with ROH spanning region standardized to 500 cases**	**Controls with ROH spanning region standardized to 500 controls**
CHD7	8	Cortex	30	2.46	61,249,482	61,286,214	1.5	0.5
FGFR1	8	Hypothalamus	5	-3.56	37,886,171	37,994,786	2.3	0.8
		Pituitary	2	0.49	38,484,726	38,485,090	1.5	0.5
GNRHR	4	Anterior cingulate cortex	1	-0.43	68,539,110	68,539,110	3.1	2.4
		Cerebellar Hemisphere	89	0.8/-0.44	68,479,934	68,594,552	3.1	2.1
		Cerebellum	86	0.9	68,479,934	68,596,604	3.1	2.1
IL17RD	3	Caudate basal ganglia	6	-2.38	56,617,585	57,988,718	0.0	0.0
PROKR2	20	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SOX10	22	Amygdala	3	-2.02	38,113,919	38,856,718	0.0	1.0
TACR3	4	Hippocampus	2	0.9	103,903,209	103,903,209	23.7	12.8
		Pituitary	10	1.18	104,408,993	104,551,098	12.3	8.4
Less common								
eQTL Target Gene	**Chr**	**Tissue with eQTL**	**Number eQTLs**	**Most pos/most neg eQTL Effect Size**	**Location First eQTL**	**Location Last eQTL**	**Cases with ROH spanning region standardized to 500 cases**	**Controls with ROH spanning region standardized to 500 controls**
AXL	19	Cerebellum	1	0.21	41,833,167	41,833,167	0.8	0.8
		Nucleus Accumbens basal ganglia	2	-2.43	41,739,611	42,435,481	0.8	0.8
CCDC141	2	Nucleus Accumbens basal ganglia	2	0.47	179,363,685	179,363,807	3.1	2.4
		Putamen basal ganglia	9	0.5, -0.48	179,926,165	180,096,751	0.8	0.5
DUSP6	12	N/A		N/A	N/A	N/A	N/A	N/A
FEZF1	7	Cerebellum	10	0.57, -0.56	121,100,433	121,106,729	1.5	0.8
		Frontal Cortex	2	0.77	122,651,200	122,658,373	10.0	5.0
		Nucleus Accumbens basal ganglia	2	-1.77	121,268,017	121,411,003	1.5	0.8
		Putamen basal ganglia	17	2.46, -2.46	121,084,301	121,128,467	1.5	0.5
FGF8	10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
FGF17	8	Caudate Basal ganglia	8	0.2	21,895,773	21,895,763	0.0	0.0
		Cerebellum	11	0.24, -0.22	21,895,146	21,901,083	0.0	0.0
		Pituitary	18	0.36, -0.33	21,878,607	21,898,763	0.0	0.0
FLRT3	20	N/A	N/A	N/A	N/A	N/A	N/A	N/A
GHRH1	20	N/A	N/A	N/A	N/A	N/A	N/A	N/A
HS6ST1	2	Amygdala	2	1.01	129,026,457	129,045,871	3.1	3.7
		Pituitary	23	-0.4	129,045,584	129,084,062	0.8	2.1
KISS1	1	Nucleus Accumbens basal ganglia	7	-0.37	204,206,270	204,213,058	0.8	1.0
		Putamen basal ganglia	6	-0.55	204,166,920	204,175,579	0.8	1.0
KISS1R	19	Caudate Basal ganglia	1	0.97	919,009	919,009	0.0	0.5
		Cortex	1	0.88	919,009	919,009	0.0	0.5
		Nucleus Accumbens basal ganglia	1	0.96	919,009	919,009	0.0	0.5
POLR3B	12	Pituitary	1	1.07	106,781,264	106,781,264	0.0	0.8
PROK2	3	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SEMA3A	7	Cortex	85	-2.53	83,917,699	84,151,726	0.0	0.5
SEMA3E	7	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SPRY4	5	Amygdala	1	-1.94	141,141,967	141,141,967	1.5	0.3
		Anterior Cingulate Cortex	1	0.66	142,642,958	142,642,958	0.0	0.8
		Cerebellum	12	0.6, -0.31	141,804,052	141,822,539	0.8	0.3
SRA1	5	Amygdala	69	0.45	139,730,444	139,941,318	9.2	8.6
		Anterior Cingulate Cortex	110	0.4, -0.29	139,719,991	139,941,318	8.4	8.4
		Caudate Basal ganglia	100	0.36, -0.35	139,719,991	139,941,318	8.4	8.4
		Cerebellum	121	0.26	139,729,152	139,917,146	9.2	8.6
		Cortex	133	0.66, -0.58	139,719,991	139,941,318	8.4	8.4
		Frontal Cortex	109	0.45, -0.39	139,719,991	139,941,318	8.4	8.4
		Hypothalamus	86	0.48, -0.43	139,719,991	139,941,318	8.4	8.4
		Putamen basal ganglia	45	0.37	139,730,444	139,941,318	9.2	8.6
		Substantia nigra	103	0.48	139,719,991	139,941,318	8.4	8.4
		Pituitary	2	1.66, -1.24	139,205,737	139,205,737	8.4	5.0
TAC3	12	Cortex	1	0.8	56,579,826	56,579,826	4.6	2.6
		Pituitary	18	-0.62	57,317,393	57,395,678	5.4	2.9
WDR11	10	Amygdala	165	0.46, -0.6	122,557,474	122,644,992	2.3	0.0
		Anterior Cingulate Cortex	157	-1.17	121,732,133	122,646,126	2.3	0.0
		Caudate Basal ganglia	212	0.35, -0.53	122,511,995	122,634,406	2.3	0.0
		Cerebellar Hemisphere	179	0.35, -0.53	122,343,181	122,646,113	2.3	0.0
		Cerebellum	250	0.48, -0.74	122,299,376	122,634,099	2.3	0.0
		Cortex	251	0.59, -0.57	122,489,283	122,638,836	2.3	0.0
		Frontal Cortex	178	0.34, -0.41	122,593,020	122,646,113	2.3	0.0
		Hippocampus	83	-0.36	122,611,526	122,677,278	2.3	0.0
		Hypothalamus	150	0.44, -0.52	122,585,852	122,646,738	2.3	0.0
		Nucleus Accumbens basal ganglia	259	0.53, -0.72	122,534,587	122,634,955	2.3	0.0
		Putamen basal ganglia	250	0.70, -0.69	122,489,283	122,634,099	2.3	0.0
		Pituitary	262	0.44, -0.56	122,534,134	122,634,099	2.3	0.0

Table 3: Locations of eQTLs in Known IGD Genes seen in Neuronal or Pituitary Tissue Identified through GTEx and the Number of Cases and Controls with an ROH that Spans the Area eQTL regions are provided by GTEx. (GETx 2017) Numbers of cases and controls are standardized to per 500 individuals for an accurate comparison taking into account population size.

Homozygosity Mapping Analysis in Novel Regions of the Genome in the European Only Cohort

Four different chromosomal regions in the genome with a unique region of homozygous SNPs seen only in cases (probands) and not in controls were found: one on chromosome 3, one on chromosome 4, one on chromosome 6 and on chromosome 7. Individual probands who possessed an ROH that covered one of these unique areas were assessed as shown in Table 4.



Chr3:124810960; n=4	Chr6:96,296,327-96,846,975; n=4	chr7:143,053,723-143,080,060; n=6
1. SLC12A8 i. Functions as a sodium/ potassium/chloride transporter and had been associated with psoriasis (OMIM) ii. According to GTEx, the highest median expression is in the Thyroid (GTEx 2017) iii. Coding region transcript is located on chr3:124,802,734-124, 930,193 with a size of 127,460 base-pairs and 13 coding exons. (Kent 2002)	1. FUT9 i. An alpha-3-fucosyltransferase which catalyzes the last step of the biosynthesis of Lewis antigen. FUT9 also synthesizes LeX oligosaccharide (CD15) which is expressed in organ buds progressing in mesenchym during embryogenesis. (OMIM) ii. According to GTEx, the highest median expression is in the cerebellum (GTEx 2017) iii. Coding region transcript is located on chr6:96,651,032-96,65 2,111 with a size of 1,080 base-pairs and 3 coding exons. (Kent 2002)	1. FAM131B i. According to GTEx, the highest median expression is in the cerebellar hemisphere. (GTEx 2017) ii. Coding region transcript is located on chr7:143,053,643-143, 057,186 with a size of 3,544 base-pairs and 6 coding exons. (Kent 2002) 2. ZYX i. Phosphoprotein found at adhesion plaques and actin filament bundles. (OMIM) ii. According to GTEx, the highest median expression is in whole blood. (GTEx 2017) iii. Coding region transcript is located on chr7:143,078,665-143, 087,775 with a size of 9,111 base-pairs and 9 coding exons. (Kent 2002)

Figure 13: Coding Genes within the Unique Regions shared by Cases Only

The UCSC Genome Browser found 1 coding gene in the unique region 124,810,960 on chromosome 3 called Sodium Channel and Clathrin Linker 1 (*SLC12A8*). All four probands in table 4 that have a ROH spanning this area were queried in GEMINI for potential mutations in the coding areas of these genes but none of the four individuals had a homozygous variant in this gene. The UCSC Genome Browser also failed to find any coding genes within the unique region 82,547,364-82,587,050 on chromosome 4;

however, they did find a long intergenic non-coding RNA, *RP11-689K5.3* that has its highest median expression in the pituitary (GTEx 2017). The UCSC Genome Browser found one coding gene within the unique region 96,296,327-96,846,975 on chromosome 6 called fucosyltransferase 9 (*FUT9*). All four IGD probands in Table 4 with an ROH spanning the ROH were queried in Gemini for homozygous variants in this gene but none were found. Three non-coding transcripts were also found in this region: A pseudogene, *RN752797P,* with highest median expression in subcutaneous adipose (GTEx 2017), a pseudogene, *KRT18P50* with its highest median expression in subcutaneous adipose (GTEx 2017), and an antisense RNA, *UFL1-AS1*, with its highest median expression in the testis (GTEx 2017). Finally, the UCSC Genome Browser found 2 coding genes within the unique region of 143,053,713-143,080,060 on chromosome 7: a family with sequence similarity in 131 member B (*FAM131B*) and zyxin (*ZXY*) . None of the six probands in Table 4 with a ROH overlapping this region had a homozygous variant in either coding gene. Two non-coding transcripts within this unique region were also found: A long intergenic non-coding RNA, *RP11-563K23.1,* whose highest median expression occurs in the anterior cingulate cortex of the brain (GETx 2017) and an antisense RNA, *AC093673.5*, with the highest median expression in the uterus (GETx 2017). More information about the coding genes found in any of the four unique regions is found in Figure 13.

Unique region on Chr3:124810960; n=4											
Proband ID	**Diagnosis**	**Sex**	**Known Genetic Variation**	**Chr**	**First SNP**	**Last SNP**	**Location Start**	**Location End**	**Kb**	**# of SNPs**	**Density**
1859001	nIHH	M	none known	3	exm345486	exm347329	124810295	126751579	1941.29	172	11.287
2608001	Adult onset nIHH	M	none known	3	chr3_123166261	exm345522	123166261	124826536	1660.28	149	11.143
2555001	KS	M	none known	3	exm345486	rs9880406	124810295	126049305	1239.01	90	13.767
2245001	nIHH, cryptorchidism	M	none known	3	variant.77141	rs9865702	124810960	126092351	1281.39	93	13.778
Unique region on Chr4:82,547,364-82,587,050; n=4											
Proband ID	**Diagnosis**	**Sex**	**Known Genetic Variation**	**Chr**	**First SNP**	**Last SNP**	**Location Start**	**Location End**	**Kb**	**# of SNPs**	**Density**
5513	nIHH	M	none known	4	rs11098907	rs11099493	80745103	82587050	1841.95	90	20.466
68501	KS	M	DUSP6 het, FGFR1 het	4	exm409252	rs11099493	81188123	82587050	1398.93	68	20.572
47601	nIHH	M	GnRHR hom	4	exm405598	rs10022462	74270123	89243818	14973.7	1014	14.767
1661001	nIHH	F	GnRHR hom	4	exm402033	rs17031254	68380070	102325626	33945.6	1989	17.067
Unique region on Chr6:96,296,327-96,846,975; n=4											
Proband ID	**Diagnosis**	**Sex**	**Known Genetic Variation**	**Chr**	**First SNP**	**Last SNP**	**Location Start**	**Location End**	**Kb**	**# of SNPs**	**Density**
83301	nIHH, hyposmia	F	none known	6	rs9354118	rs9398152	95147902	97063724	1915.82	67	28.594
2010001	Adult onset nIHH, hyposmia	M	DUSP6 het, FGFR1 het	6	psy_rs12524279	rs10499010	95513192	97764445	2251.25	101	22.29
3001	nIHH, Fertile Eunuch	M	none known	6	rs2785917	rs6939572	96296327	99139350	2843.02	105	27.076
45902	nIHH	M	none known	6	rs16881970	rs10499010	72666374	97764445	25098.1	1284	19.547
Unique region on Chr7:143,053,723-143,080,060; n=6											
Proband ID	**Diagnosis**	**Sex**	**Known Genetic Variation**	**Chr**	**First SNP**	**Last SNP**	**Location Start**	**Location End**	**Kb**	**# of SNPs**	**Density**
2071001	KS	M	FGF8fx	7	exm663816	exm666209	141952371	143080060	1127.69	124	9.094
1568001	nIHH	M	IL17RD het, POLR3B het	7	exm663784	exm666209	141803090	143080060	1276.97	137	9.321
2229001	nIHH, cryptorchidism	M	none known	7	exm665388	exm666989	142630489	143771934	1141.45	145	7.872
2488001	KS	M	none known	7	exm665577	exm667470	142658922	144101728	1442.81	144	10.019
2562001	nIHH	M	none known	7	exm666127	exm667455	143053713	144098529	1044.82	88	11.873
1521001	KS	F	none known	7	exm666127	rs10255070	143053713	144114122	1060.41	92	11.526

Table 4: Four Unique Homozygous SNP Regions Shared by Four or more Cases and no Controls.

DISCUSSION

ROH Burden Analysis in the European Only Cohort

In this study, there were no significant findings to conclude that autozygosity is a risk factor for IGD. In fact, the data appears to show the opposite trend where, in general, the controls seem to have more ROHs, both when looking at their number and their length, than the cases. These results, however, do not rule out the hypothesis that autozygosity is a risk factor for IGD. Several limitations to this study include the small size of our patient population, the density of the SNP chips available, the limitation of the control data set, and the lack of ROH coverage of the X chromosome.

Due to the rarity of our disease model and its reproductive phenotype's negative selection on a population basis, the small effect size in an outbred population has proven to be challenging. To gain significant results from these populations, it has been shown that sample sizes need to be around 12,000-65,000 individuals (Keller 2011). Our population size was 653 probands. Second, the density of the PsychChip used for this analysis was 264,522 SNPs whereas the optimal standard density of a currently available microarray chip is now closer to 1 million SNPs (Ceballos 2018). With increasing densities of the SNP chip, higher quality call rates for ROHs can be achieved. For this study, the QC was stringent since two different versions of the PsychArray were used and eventually merged together. Also, different batches within a single version where genotyped at different times and then merged together, which could lead to differences in call rates, therefore, a considerable number of SNPs in this analysis were lost. Third, the control data set had its initial QC done at The Broad Institute whereas the cases initial QC

was performed at the REU (see methods), which could have caused variation in the QC of the cases versus the controls. Also, these controls come from The Schizophrenia Psychiatric Genome-Wide Association Study; thus there might be a chance that unaffected schizophrenic family members are in the control population data set. While there have been mixed results whether schizophrenics have an increased burden of ROH (Keller 2012), several studies, including Keller et al 2012, have shown that there is an increased burden of ROH in this population. If this is true and schizophrenic family members are in the control group, there could be an over inflation of the length and number of ROHs in the control population that might obscure differences in the IGD population. Finally, the X chromosome contains a higher rate of ROHs due to its low recombination rate (Curtis 2008) and would therefore be an ideal location where ROH would uncover novel causal variation, especially since one of our known IGD genes, *ANOS1* (*KAL1*) is located on the X chromosome. There is, however, a low SNP density on the X chromosome leaving it much harder to interpret (Ceballos 2018). Along with the stated limitations to this study, as already mentioned above but in a slightly different context, another reason the hypothesis of autozygosity being a risk factor for IGD cannot be ruled out is that in other disease models, such as schizophrenia, the results on whether there is a significant effect of autozygosity in their disease have been mixed with several studies finding a significance and others not (Keller 2012). These results shows that our understanding of ROHs in outbred populations, as well as our detection methods, can still be improved.

Homozygosity Mapping Analysis on Known Genes and their Predicted Regulatory Sites in the European Only Cohort

It was unlikely that novel variants in the coding regions of known genes would be uncovered using homozygosity mapping since whole exome sequencing has been performed on all the individuals that had undergone SNP genotyping in this analysis in the REU cohort. However, uncovering two IGD cohort probands with ROHs spanning the genomic location of the known IGD causing gene, *GNRHR,* both of whom turned out to also be homozygous for a variant in *GNRHR,* validates the approaches we employed to detect the ROHs. This validation still holds despite the fact that these homozygous mutations were already known in these individuals, since the individuals were uncovered using an unbiased method, where no indications of known gene variation were known when searching for individuals with ROHs that overlapped the region of interest. It is only after the individual IDs were found that their genetic variation was confirmed.

The advantage to using SNP genotyping analysis vs. whole exome sequencing (WES), however, lies not in finding defects in the coding region where WES is clearly superior. Rather, surfacing defects in the non-coding region of the genome is where ROH has its greatest value. The majority of the SNPs genotyped (80+ %) lie within the non-coding regions of the genome. These regions in the non-coding domains of the human genome are predicted to influence the expression of our known genes and currently are not accessible via WES. Therefore, this method provides a way to examine potential disease causing variants within both the non-coding regions and the eQTL domains of those genes. Typical of Mendelian conditions, many known IGD genes exhibit variable

penetrance/expressivity defined as whether the gene mutation causes the disease or not (penetrance) or to what extent the phenotype is expressed (expressivity) either within their family or across other IGD families. Expression quantitative trait loci (eQTL's) have been shown to account for at least some of this variability (Castel 2017). While it is still problematic to fully interpret what it means for a certain number of cases and controls to overlap a region of interest, it is interesting to note that there is an overall trend, when looking at Tables 2 and 3 where, when the numbers of cases and controls with ROHs spanning a region of interest is standardized to per 500 individuals, the cases have an overall higher representation of ROHs in these regions compared to the controls. With the decreasing price of next generation sequencing, it is increasingly feasible to perform whole genome sequencing (WGS) on individuals within disease cohorts. New pipelines are being rapidly developed to both manage and interpret these results and our understanding keeps improving as well. It is crucial for the complete understanding of human genetics to begin examining the non-coding region since 98% of the human genome is non-coding (Iulio 2018). Therefore, prioritizing the sequencing of certain individual's genomes that already have areas of interest with potential disease-causing variations within their non-coding region is a good place to start. The cases that overlap eQTL regions of known genes (roughly 194/653) would be high on this list.

Homozygosity Mapping Analysis in Novel Regions of the Genome in the European Only Cohort

Four regions of homozygous SNPs were found in which 4 or more cases had regions of ROHs that overlapped while no controls overlapped the region. This discovery analysis

eliminated the uncertainty in interpreting what it means for a control sample to overlap an area of interest thus making these 4 areas particularly intriguing. Some of these 4 individuals (Table 4) have genetic variation in a known IGD gene. However, due to the high levels of incomplete penetrance and incomplete expressivity that are frequently observed in our known genes, these known genetic variants certainty do not have to be the cause or the only cause of the disease. As our group previously documented, oligogenicity is prevalent in IGD (Sykiotis 2010; Balasubramanian 2010) and it is likely that another known gene variant could be working in synergy with another variant, either in a coding gene or non-coding regulatory region. While no case had a homozygous variant in a coding gene within the unique area of interest in the genome, this lack does not preclude contributory variation within the non-coding region of these unique areas, again underscoring the importance of understanding how to interpret these regions and also the contributing knowledge WGS will surely provide us. Cases overlapping these areas of interest would also be considered higher priority cases for whole genome sequencing.

Alternative ROH Detection Methods

Two main methods have been used to detect ROH: observational genotype-counting and model-based approaches (Ceballos 2018). The approach used in this study was an observational approach using PLINK (Purcell 2007). PLINK employs a sliding window that checks for the amount of homozygous SNPs within a given window and will call a ROH based on the criteria specified, whether it be no heterozygous SNPs allowed in a ROH, the minimum and maximum amount of SNPs that define a ROH, etc. A model-

based approach can be done through programs such as Beagle (Ceballos 2018, Browning 2013) that uses a hidden Markov model (HMM) to detect ROHs. PLINK, however, has been shown to outperform other models such as Beagle (Howrigan 2011), and was therefore the best program to use for this study.

Next Steps

Homozygosity mapping is a useful tool for detecting Mendelian disease variants, where a variation in a single gene gives rise to a phenotype that can be passed from parent to offspring. Complex Trait Genetics involves variations in multiple genes that can interact with the environment. While IGD is historically thought of as a Mendelian disease (Balasubramanian 2012), it is becoming clearer that there should be less of a divide in thinking between the two genetic disease approaches; rather a more holistic approach that combines both Mendelian disease and complex trait disease methods should be considered and might be more profitable to employ. Genome wide association studies (GWAS) uses SNP markers to identify variants associated with a given trait in an unbiased manner and have been extensively used and quite successful in identifying variants in complex common traits (Verma 2018). The QC'ed SNPChip dataset created in this study will be used as the platform in an IGD GWAS study.

Another method that uses this SNPChip platform as a marker is a transmission disequilibrium test (TDT). A TDT is a family-based association test based on linkage that tests for seemingly overtransmission of a disease-causing allele to an affected individual from his/her parents. The REU IGD cohort, being one of the largest cohorts for this rare

disease in the world, has a fair amount of genotyped affected individuals whose parents are also genotyped (termed trios) that will be used for this analysis.

REFERENCES

Acierno, JS., Shagoury, JK., Bo-Abbas, Y., Crowley, WF., Seminara, SB. (2003). A locus for autosomal recessive idiopathic hypogonadotropic hypogonadism on chromosome 19p13.3. *The Journal of Clinical Endocrinology and Metabolism,* 88(6), 2947-2950.

Amoss, M., Burgus, R., Blackwell, R., Vale, W., Fellows, R., Guillemin, R. (1971). Purification, animo acid composition and N-terminus of the hypothalamic luteinizing hormone releasing factor (LRF) of ovine origin. *Biolchemical and Biophysical Research Communications,* 44(1), 205-210.

Balasubramanian R, Crowley WF Jr. Isolated Gonadotropin-Releasing Hormone (GnRH) Deficiency. (2007) [Updated 2017 Mar 2]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1334/

Balasubramanian, R., Crowley, WF. (2011). Isolated GnRH Deficiency: A Disease Model Serving as a Unique Prism into the Systems Biology of the GnRH Neuronal Network. *Molecular Cell Endocrinology,* 346(1-2), 4-12

Balasubramanian, R., Dwyer, A., Seminara SB., Pitteloud, N., Kaiser, UB., Crowley, WF. (2010). Human GnRH Deficiency: A Unique Disease Model to Unravel the Ontogeny of GnRH Neurons. *Neuroendocrinology,* 92(2), 81-99

Ball, A. D., Stapley, J., Dawson, D. A., Birkhead, T. R., Burke, T., & Slate, J. (2010). A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (*Taeniopygia guttata*). *BMC Genomics*, *11*, 218.

Ballabio A, Parenti G, Tippett P, et al. X-linked ichthyosis, due to steroid sulphatase deficiency, associated with Kallmann syndrome (hypogonadotropic hypogonadism and anosmia): linkage relationships with Xg and cloned DNA sequences from the distal short arm of the X chromosome. Hum Genet. 1986;72:237–240

Bick, D., Curry, CJ., McGill, JR., Schorderet, DF., Moore, CM. (1989). Male infant with ichthyosis, Kallmann syndrome, chondrodysplasia punctata, and an Xp chromosome deletion. *American Journal of Medical Genetics,* 33(1), 100-107.

Bittles, AH., Neel, JV. (1994). The costs of human inbreeding and their implications for variations at the DNA level. *Nature Genetics,* 8, 117-121.

Bolino, A., Brancolini, V., Bono, Francesco., Bruni, A., Gambardella, A., Romeo, Giovani., Quattrone, A., Devoto, M. (1996). Localization of a gene responsible for

autosomal recessive demyelinating neuropathy with focally folded myelin sheaths to chromosome 11q23 by homozygosity mapping and haplotype sharing. *Human Molecular Genetics.* 5(7), 1051-1054.

Botstein, D., and Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33(Suppl), 228–237

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3), 314–331.

Browning, BL., Browning, SR. (2013). Improving the accuracy and efficiency of identity-by-decent detection in population data. *Genetics*, 194(2), 459-471.

Caronia, LM., Martin, C., Welt, C.K., Sykiotis, GP., Quinton, R., Thambundit, A., Avbelj, M., Dhruvakumar,S., Plummer, L., Hughes, VA., Seminara, SB., Boepple,PA., Sidis, Y., Crowley, WF., Martin, KA., Hall, JE., Pitteloud, N. A Genetic Basis for Functional Hypothalamic Amenorrhea.(2011) *The New England Journal of Medicine.* 364:215-225

Ceballos, FC., Joshi, PK., Clark, DW., Ramsay, M., Wilson, JF. (2018) Runs of Homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics.* DOI: 10.1038/nrg.2017.109 Epub.

Curtis, D., Vine, AE., Knight, J. (2008) Study of regions of extended homozygosity provides a powerful method to explore haplotype structure of human populations. *Annals of Human Genetics.*72, 261-278.

Desaulniers, AT, Cederberg, RA, Lents, CA, White,B. (2017). Expression and Role of Gonadotropin-Releasing Hormone 2 and Its Receptor in Mammals. *Frontiers in Endocrinology,* 8(269), 1-25

Franco, B., et al. (1991). A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. *Nature,* 353, 529-536.

GTEx Consortium (2017). Genetic effects on gene expression across human tissues. *Nature,* 550, 204-213.

Hildebrandt, F., Heeringa, S. F., Rüschendorf, F., Attanasio, M., Nürnberg, G., Becker, C., … Otto, E. A. (2009). A Systematic Approach to Mapping Recessive Disease Genes in Individuals from Outbred Populations. *PLoS Genetics*, *5*(1), e1000353.

Howrigan, DP., Simonson, MA., Keller, MC. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genetics*, 12:460

International HapMap Constortium. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449:851-862

Iulio, Jd., et al. (2018) The human noncoding genome defined by genetic diversity. *Nature Genetics*, doi:10.1038/s41588-018-0062-7

Kallmann F., Schoenfeld W., et al. (1994). The genetic aspects of primary eunuchoidism. *American Journal of Mental Deficiency,* 48, 203-236

Karstensen, HG., Tommerup, N. (2012) Isolated and syndromic forms of congenital anosmia. *Clinical Genetics.* 81(3), 210-215.

Keller, MC., Simonson, MA., Ripke, S., Neale, BM., Gejman, PV., Howrigan, DP., Lee, SH., Lencz, T., Levinson, DF., Sullivan, PF. (2012) Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor. *PLOS Genetics.* 8(4), e1002656.

Keller, MC., Visscher, PM., Goddard, ME. (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism dats. *Genetics.* 189, 237-249.

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

Klein, DA., Emerick, JE., Sylvester, JE., Vogt, KS. (2017) Disorders of Puberty : An Approach to Diagnosis and Management. *American Family Physician.* 96(9), 590-599

Laitinen, E-M., Vaaralahti, K., et al. (2011) Incidence, Phenotypic Features and Molecular Genetics of Kallmann Syndrome in Finland. *Orphanet Journal of Rare Diseases*. 6:41

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

Liu, JH., Bill, AH. (2008) Stress-associated or functional hypothalamic amenorrhea in the adolescent. *Annals of the New York Academy of Science.* 1135, 179-184

Manichaikul, A., Mychaleckyj, JC., Rich, SS., Daly, K., Sale, M., Chen, WM. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics.* 26(22), 2867-2873.

Matsuo, H., Baba, Y., Nair, RM., Arimura, A., Schally, AV. (1971) Structure of the procine LH- and FSH-releasing hormone. I. The proposed amino acid sequence. *Biochemical and Biophysical Research Communications.* 43(6), 1334-1339.

McQuillan, R., et al. (2008) Runs of Homozygosity in European Populations. *The American Journal of Human Genetics.* 83, 359-372

Moya-Plana,A., Villanueva, C., Laccourreye, O., Bonfils, P., de Roux, N. (2012) PROKR2 and PROK2 mutations cause isolated congenital anosmia without gonadotropic deficiency. *European Journal of Endocrinology.* 168(1), 31-37

Naor, Z., Benard, O., Seger, R. (2000) Activation of MAPK cascades by G-protein-coupled receptors: the case of gonadotropin-releasing hormone receptor. *Trends in Endocrinology and Metabolism:TEM.* 11(3), 91-99

Neill, JD., Musgrove, LC., Duck, LW. (2004) Newly recognized GnRH receptors: function and relative role. *Trends in Endocrinology and Metabolism:TEM.* 18(8), 383-392

Nica, A. C., & Dermitzakis, E. T. (2013). Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1620), 20120362. http://doi.org/10.1098/rstb.2012.0362

Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), {02/25/2018}. World Wide Web URL: https://omim.org/

Paila, U., Chapman, BA., Kirchner, R., Quinlan, AR. (2013). GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations. *PLoS Computational Biology.* 9(7), e1003153

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, MAR., Bender, D., Maller, Julian., Sklar, P., Bakker, PIW., Daly, MJ., Sham, PC. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analysis. *American Journal of Human Genetics.* 81(3), 559-575

Pitteloud, N., Quinton, R., Pearce, S., Ravivio,T., Acierno,J., Dwyer,A., Plummer, L., Hughes, V., Seminara, S., Cheng, Y-Z., Li, W-P., Maccoll, G., Eliseenkova, AV., Olsen, SK., Ibrahimi, OA., Hayes, FJ., Boepple, P., Hall, JE., Bouloux, P., Mohammadi, M., Crowley, WC. (2007) Digenic mutations account for variable phenotypes in idiopathic hypogonadotropic hypogonadism, *The Journal of Clinical Investigation*, 117(2), 457-463

Schwanzel-Fukuda, M., Bick, D., Pfaff, DW. (1989). Luteinizing hormone-releasing hormone (LHRH)-expressing cells do not migrate normally in an inherited hypogonadal (Kallmann) syndrome. *Brain Research. Molecular Brain Research,* 6(4), 311-326

Schwanzel-Fukuda, Pfaff, DW. (1989). Origin of luteinizing hormone-releasing hormone neurons. *Nature,* 338(6211), 161-164

Science. (2005). So Much More to Know... *Science,* 309(5731), 78-102

Seminara, S. B., & Crowley, W. F. (2008). Kisspeptin and GPR54: Discovery of a Novel Pathway in Reproduction. *Journal of Neuroendocrinology*, *20*(6), 727–731

Seminara, SB., Messager, S., Chatzidaki, EE., Thresher, RR., et al. (2003). The GPR54 Gene as a Regulator of Puberty. *New England Journal of Medicine,* 349:1614-1627

Silveira, LG., et al. (2010) Mutations of the KISS1 Gene in Disorders of Puberty. *Journal of Clinical Endocrinology and Metabolism.* 95(5): 2276-2280

Stamou, MI., Cox, KH., Crowley, WF. (2015) Discovering Genes Essential to the Hypothalamic Regulation of Human Reproduction Using a Human Disease Model: Adjusting to Life in the "-Omics" Era. *Endocrine Reviews*. 36(6):603-621

Stojikovic, SS., Catt, KJ. (1995) Expression and signal transduction pathways of gonadotropin-releasing hormone receptors, *Recent Progress in Hormone Research*, 50, 161-205

Sykiotis, GP., Plummer, L., Hughes, VA., Au, M., Durrani, S., Nayak-Young, S., Dwyer, AA., Quinton, R., Hall, JE., Gusella, JF., Seminara, SB., Crowley, WF Jr., Pitteloud, N. (2010) Oligogenic basis of isolated gonadotropin-releasing hormone deficiency, *Proceedings of the National Academy of Science of the United States of America*, 107(34), 15140-115144

Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zöllner, S., Rosenberg, N. A., & Li, J. Z. (2013). Long Runs of Homozygosity Are Enriched for Deleterious Variation. *American Journal of Human Genetics*, *93*(1), 90–102.

Vadakkadath,M., Atwood,CS. (2005). The role of hypothalamic-pituitary-gonadal hormones in the normal structure and functioning of the brain, *Cellular and Molecular Life Sciences*, 62(3), 257-270

Verma, SS., Ritchie, MD. (2018). Another Round of "Clue" to Uncover the Mystery of Complex Traits. *Genes*, 9(2), 61

Wierman, M., Kiselijak-Vassiliades,K., Tobet, S. (2012). Gonadotropin Releasing Hormone (GnRH) Neuron Migration: Initiation, Maintenance and Cessation as Critical Steps to Ensure Normal Reproductive Function, *Front Neuroendocrinology*, 32 (1), 43-52.

Zhu, J., Choa, R. E.-Y., Guo, M. H., Plummer, L., Buck, C., Palmert, M. R., Hirschhorn, J.N., Seminara, S.B., Chan, Y.-M. (2015). A Shared Genetic Basis for Self-Limited Delayed Puberty and Idiopathic Hypogonadotropic Hypogonadism. *The Journal of Clinical Endocrinology and Metabolism*, *100*(4), E646–E654

CURRICULUM VITAE