Title
Elucidating the Genetics of Vaccine-Induced Febrile Seizure Through a GWAS and Exome Sequencing

Author
Rojo, Carlos

Publication Date
2014

Peer reviewed|Thesis/dissertation

Elucidating the Genetics of Vaccine-Induced Febrile Seizure Through a GWAS and Exome Sequencing

by

Carlos Roberto Rojo

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

Acknowledgements and Dedication

I thank all the study's participants and their families for their willingness to participate and support this research. I thank my research supervisor and thesis committee chair, Dr. Mark Seielstad, for affording me the opportunity to work on this project and for serving as my committee's chair. My sincere thanks also to Dr. Noah Zaitlen and Dr. Elad Ziv, who reviewed this manuscript and provided helpful suggestions. I also thank Dr. Nicola P. Klein (and colleagues) whose group performed much of the work involved in the design of this study and acquisition of samples. I am also grateful to the University of Texas at Austin's Texas Advanced Computing Center (TACC) for providing me, via NSF's XSEDE, with the computational resources necessary to store and process the genotype and exome sequencing data.

Finalmente, este trabajo esta dedicado a la familia Rojo, cuyo apoyo, paciencia, y consejos me han guiado durante mi carrera y ayudado a superar cualquier obstáculo.

A mis padres, hermanos, hermana, y a todos mis sobrinos: gracias.

Abstract

Elucidating the Genetics of Vaccine-Induced Febrile Seizure Through a GWAS and Exome Sequencing

Author: Carlos Roberto Rojo

Despite the tremendous success of vaccination efforts across the globe, outbreaks of vaccine-preventable childhood diseases in communities around the United States are increasing. These localized outbreaks, due in part to decreased vaccination coverage, are likely the result of a public skepticism about the safety of childhood vaccinations. To help restore trust in the vaccination effort, it is important to understand a vaccine's full range of side effects. For example, an increased risk of febrile seizure has been associated with a variety of common childhood vaccines including measles-containing vaccines such as the measles, mumps, and rubella vaccines. Currently, however, the biological cause of vaccination induced febrile seizure (VIFS) remains unclear and little is known as to why only a subset of the vaccine's recipients is affected. Towards these goals, we conducted a genome-wide association study and exome sequencing analysis to survey the genetic architecture of VIFS following measles vaccination. 274 California residents, 133 of whom experienced febrile seizure post-measles vaccination and 141 matched controls, were genotyped and exome sequenced. Using the software PLINK v.1.07, 5,757,137 markers were tested for association with VIFS, which highlighted potential genes of interest in VIFS that either regulate neuronal activity or cytokine production (rs202194476 in *PTPRD,* $p = 1.6x10^{-6}$; rs11186481 near *SH2D4B,* $p = 7.7x10^{-6}$; and rs56682383 in *DPYD,* $p = 2.4X10^{-5}$). After using the Burrows Wheeler Aligner (BWA), Picard, and the Genome Analysis Toolkit (GATK) to process the exome sequencing data, optimized sequence kernel association testing (SKAT-O) was used to implement gene-based association testing. This analysis highlighted potential genes of interests with roles in immune system regulation and apoptosis (*BATF,* p =

$6.1x10^{-6}$; *DDI2*, p = $6.5x10^{-5}$; NDUFS3, p = $8.9x10^{-5}$; *DEFB126,* p = $2.1x10^{-5}$). As the first GWAS and exome sequencing analysis of VIFS, this study provides the first large-scale look at the underlying genetic architecture of VIFS. As such, it builds the foundation for our understanding of a serious but unexplained side effect of the common measles vaccine, which could not only help restore any lost trust in the safety of vaccination but also potentially work towards a strategy to stratify a patient's risk for developing VIFS.

Table of Contents

List of Tables

List of Figures

Elucidating the Genetics of Vaccine-Induced Febrile Seizure Through a GWAS and Exome Sequencing

Author: Carlos Roberto Rojo

INTRODUCTION

Despite the tremendous success of vaccination efforts across the globe, outbreaks of vaccine-preventable childhood diseases in communities around the United States are increasing, including measles [1, 2] and pertussis [3]. These localized outbreaks, due in part to decreased vaccination coverage [1-4], are likely the result of a public skepticism about the safety of childhood vaccinations. Although much of this skepticism may find its main root in the discredited claims that childhood vaccination is associated with neurological conditions such as autism [5, 6], experiencing any unexplained, vaccine-induced adverse side effect could also contribute to distrust in the safety of vaccines. An example of such an adverse effect is vaccine-induced febrile seizure (VIFS). An increased risk of febrile seizure (FS) has been associated with a variety of common childhood vaccines including the measles, mumps, and rubella vaccine [7]. Currently, however, the biological cause of VIFS remains unclear and little is known as to why only a subset of the vaccine's recipients is affected. Understanding the basis for FS post-vaccination would thus serve two goals: 1.) elucidate the biological basis of FS both within and without the context of vaccination, and 2.) work towards developing strategies to stratify a patient's risk for developing VIFS. Towards these goals, this study will perform the first extensive genetic study of VIFS, utilizing both a genome-wide association study and exome sequencing, to explore the genetic basis underlying an increased risk for febrile seizure after administration of a measles-containing vaccine.

BACKGROUND

Febrile Seizure and Vaccine-Induced Febrile Seizure

Febrile seizure is the most common neurological event in childhood with a worldwide incidence of approximately 1 in 20 [8]. FS is characterized by convulsions typically lasting <10-15 minutes that occur in the context of fever without a concurrent central nervous system infection (e.g. meningitis), typically in children aged 6 months to five years [8, 9]. A variety of studies have demonstrated that many common childhood vaccinations are associated with an increased risk of febrile seizure including the measles-containing vaccines (MMR and MMRV) [7]. Although vaccine-induced febrile seizure is relatively rare (~0.1% of vaccine recipients develop VIFS), is usually non-recurrent, and has not been shown to cause any long-term neurological or developmental problems, these seizure episodes often result in visits to intensive care units [10], are concerning to parents and patients alike, and could diminish public confidence in the safety of vaccination efforts.

Potential Causes of Vaccine-Induced Febrile Seizure

When attempting to understand the cause of vaccine-induced febrile seizure, two main questions arise: what is the link between fever and seizure, and what role does vaccination play in triggering febrile seizure? Two plausible explanations for how fever might induce seizure are: 1.) hyperthermic conditions affect neuronal activity, predisposing the brain to seizure (temperature-dependent mechanism), and 2.) other factors that cause fever, or are released during the course of fever, affect neuronal activity and alter seizure risk (temperature-independent mechanism). In support of the former, it has been shown that synaptic activity is demonstrably altered at elevated temperature. For example, Kang et al found that trafficking of mutant forms of the $GABA_A$ γ2 receptor to the cell surface (a process important in the inhibition of synaptic activity) is highly temperature-dependent and significantly decreased at elevated temperatures [11]. This work is particularly interesting because mutations in GABA receptor genes are often seen in patients with generalized seizures and/or recurring FS [11, 12]. In

support of a temperature-independent cause of febrile seizure, a variety of studies have linked chemokines that are normally produced during the course of fever with an increased rate of febrile seizure [13, 14]. Importantly, many of these immunological factors are known to directly affect brain activity (e.g. interleukin IL-1β increases hyperexcitability of neurons [8, 14]). Given that vaccines are designed to elicit a strong immune response, fever, and release of pyrogenic and pro-inflammatory cytokines, it is therefore possible that either: 1.) cytokines released during the course of fever post-vaccination directly modulate neuronal activity and seizure risk, and/or 2.) vaccine-induced fever, in the context of mutations that increase susceptibility to seizures in hyperthermic conditions, triggers febrile seizure post-vaccination. Thus, both the temperature-dependent and -independent causes of FS begin to clarify two potential biological bases for how vaccines might induce febrile seizure. However, very little is known as to why some patients develop FS after vaccinations while others do not. Past work studying FS suggests that genetics might play a significant role.

Genetics of Vaccine-Induced Febrile Seizure

Given the result of a variety of family, twin, and linkage studies, genetics is believed to play a significant role in predisposition towards febrile seizure risk. For example, a family history of FS was identified in approximately 30% of patients experiencing FS [15]. Twin and family studies have reported that febrile seizure concordance rates are 3 to 4 times [16, 17] higher in monozygotic twins than in dizygotic twins and have estimated the heritability to be 75% [16], implying a significant genetic etiology. Additionally, a variety of genes have been associated with febrile seizure, almost all of which are involved in affecting neuronal excitability/activity. For example, mutations have been identified in genes coding for voltage-gated sodium ion channels (*SCN1A*) [18], GABA receptors [11, 12], and a variety of interleukins such as IL-1β (an immunological actor known to affect neuronal excitability) [13, 14]. Although some subtypes of FS such as familial simple febrile seizure are monogenic [18], the genetic architecture of VIFS is likely complex and might include contributions from a number of variants

of more modest effect. Thus, despite the progress made via use of family, twin, and linkage studies to understand FS, a detailed investigation of VIFS requires analysis of variants on a genomic scale. By utilizing a genome-wide association study and next-generation exome sequencing of patients who experienced FS post-vaccination against a measles containing vaccine (such as MMR or MMRV), this study provides the first detailed look at the genetic architecture of FS in the context of vaccination. By doing so, this study aims to not only shed light on the specific problem of VIFS but also on the etiology and biological basis of FS in general.

RESULTS

Study Population

In total, 275 samples (133 cases and 142 controls) were genotyped on the UK Biobank Axiom Array (UKBA) and exome sequenced using Illumina technology (see Methods). The samples were collected from Kaiser Permanente's Northern California Division and represent a diverse subset of populations, races, and ethnicities (Table 1).

	Sex		Race						Ethnicity	
	M	**F**	**White**	**Black**	**Asian**	**Nhpi**	**Naan**	**Other****	**Hisp.**	**Non-Hisp**
Case	77	56	75	21	34	7	8	1	49	80
Control*	65	76	75	16	29	3	0	26	52	89
Total	**142**	**132**	**150**	**37**	**63**	**7**	**8**	**27**	**101**	**169**

Table 1: Demographics of the study's cases and controls. Race and ethnicity were self-reported by study participants. Participants could select multiple racial designations for themselves. Consequently, the sums of race/ethnicity counts may not add up to the number of study participants. In addition to race, participants self-identified ethnically as Hispanic, Non-Hispanic, Other or Refuse to Answer (latter two not shown as counts for both were 0). Each case was matched on age, race, ethnicity, and vaccine administered when possible. *20 controls were first-degree relatives of a study cases and were designated "family-controls". **There was no difference in how controls were selected. The difference in "Other" counts between cases and controls is thus an anomaly. *Nhpi: Native Hawaiians and Pacific Islanders; Naan: Native American/Alaska Native*

We used Multidimensional Scaling (MDS) Analysis via PLINK [19] to compare our study's genotype data to HapMap populations and thereby better gauge the diversity of our cohort. As can be seen in Figure 1, although the majority of the samples fall between the European (CEU) and Mexican (MEX) populations, the cohort's diversity is quite high, including

samples of Asian, African, Native American ancestry, and a handful of participants with a particularly heterogeneous genetic makeup. Many of these heterogeneous samples were difficult to match with a control, resulting in less than ideal case/control matching for some samples (see Supplementary Figure S1 for a distribution of case/control status on the basis of the first two MDS components). Given that matching was based on self-identified race/ethnicities, some mismatching was unavoidable. However, this highlights the importance of correcting for cryptic population structure in the subsequent association analyses.

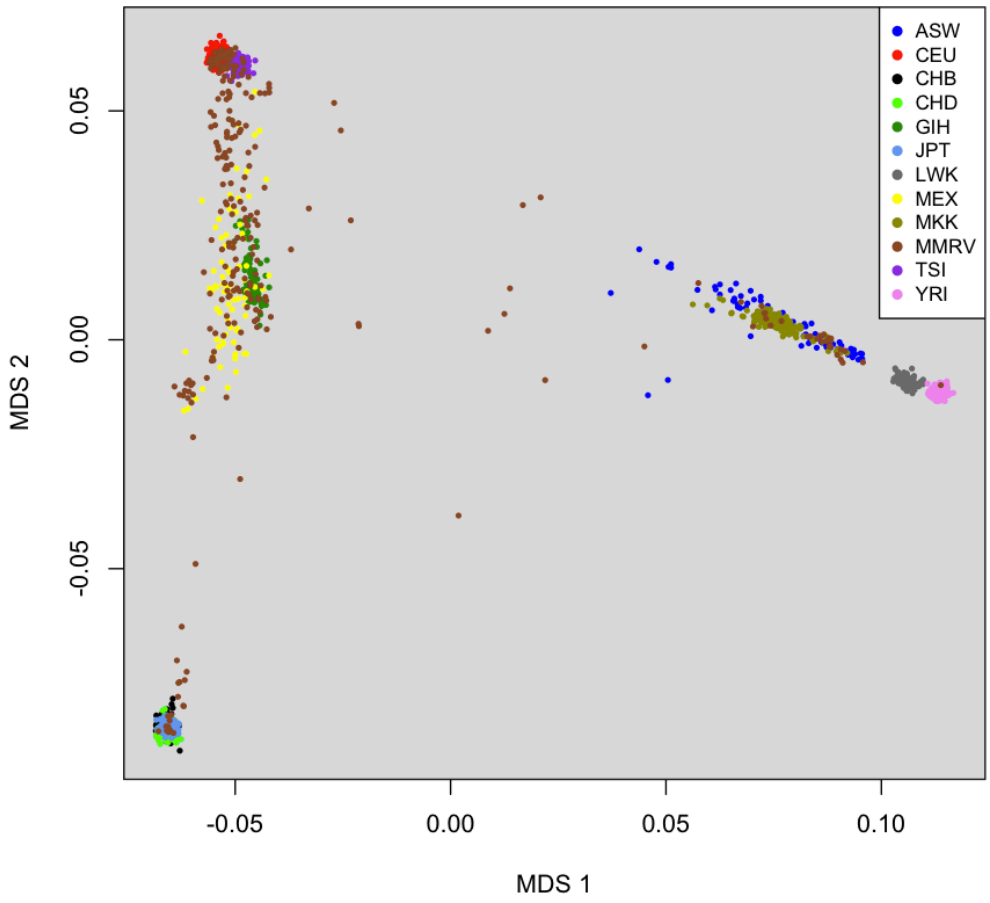


Figure 1: Plot of the first two components of the multi-dimensional scaling analysis, comparing this study's cohort genotype data with that of designated populations from the HapMap project. Although our cohort is diverse, the majority of the study participants fall between European (CEU, TSI) and Mexican (MEX) descent. *MMRV: Measles-Mumps-Rubella-Varicella (present study's cohort); ASW: African Ancestry, Southwest USA; CEU: Utah residents with Northern and Western European ancestry from the CEPH collection; CHB: Han Chinese in Beijing, China; CHD: Chinese in Denver, CO.; GIH: Gujarati Indians in Houston, TX; JPT: Japanese in Tokyo, Japan; LWK: Luhya in Webuye, Kenya; MEX: Mexican ancestry in Los Angeles, California; MKK: Maasai in Kinyawa, Kenya; TSI: Toscani in Italia; YRI: Yoruba in Ibadan, Nigeria.*

Genome-wide association study

813,270 markers from the UKBA array were genotyped and included in this study. To provide a more complete coverage of the genome, imputation was performed (via IMPUTE2 [20]) which produced 37,216,202 additional markers. After quality control (see Methods) a total of 5,757,137 were included in the analysis. These UKBA and imputed markers were tested for association with VIFS by performing logistic regression analysis, controlling for the first 10 MDS components and sex (genomic inflation factor, $\lambda = 1.04$). As can be seen in Figure 2, no marker reached genome-wide significance, defined as 5×10^{-8} (see Supplementary Figure S2 for the Manhattan and QQ-plots using only the original UKBA markers without imputation, along with the QQ-plot for the data in Figure 2). However, by looking at the Manhattan plot in Figure 2 and a list of the markers with the lowest p-values (Table 2), a few potential regions of interest emerge.

The marker with the lowest p-value is rs202194476 ($p = 1.6 \times 10^{-6}$), which lies in the *PTPRD* (protein tyrosine phosphatase receptor delta) locus on chromosome 9. Recent work associated a SNP in *PTPRD* with immune response strength after vaccination, specifically IFNγ production after administration of a rubella/measles-containing vaccine such as MMR [21] (the SNP identified in [21], rs16928280, was not included in our analysis since it did not pass quality control). Additionally, *PTPRD* and its gene product PTPδ have been implicated in the regulation of neuronal function and development in a variety of contexts including synapse development and activity [22], increased risk of seizure [22, 23], attention deficit disorder [24], and restless leg syndrome [25]. Thus, among other roles [26], *PTPRD* appears to play a role in regulating both the immune system response post-vaccination and neuronal activity and function. Although our data is inconclusive, given this previous work and the fact that another SNP in *PTPRD* is at least modestly suggestive in our GWAS (rs10125887, $p = 2.9 \times 10^{-5}$), this hints at the possibility that *PTPRD* is an important gene in stratifying VIFS risk, potentially via the regulation of cytokine production and/or synapse development and inhibition.

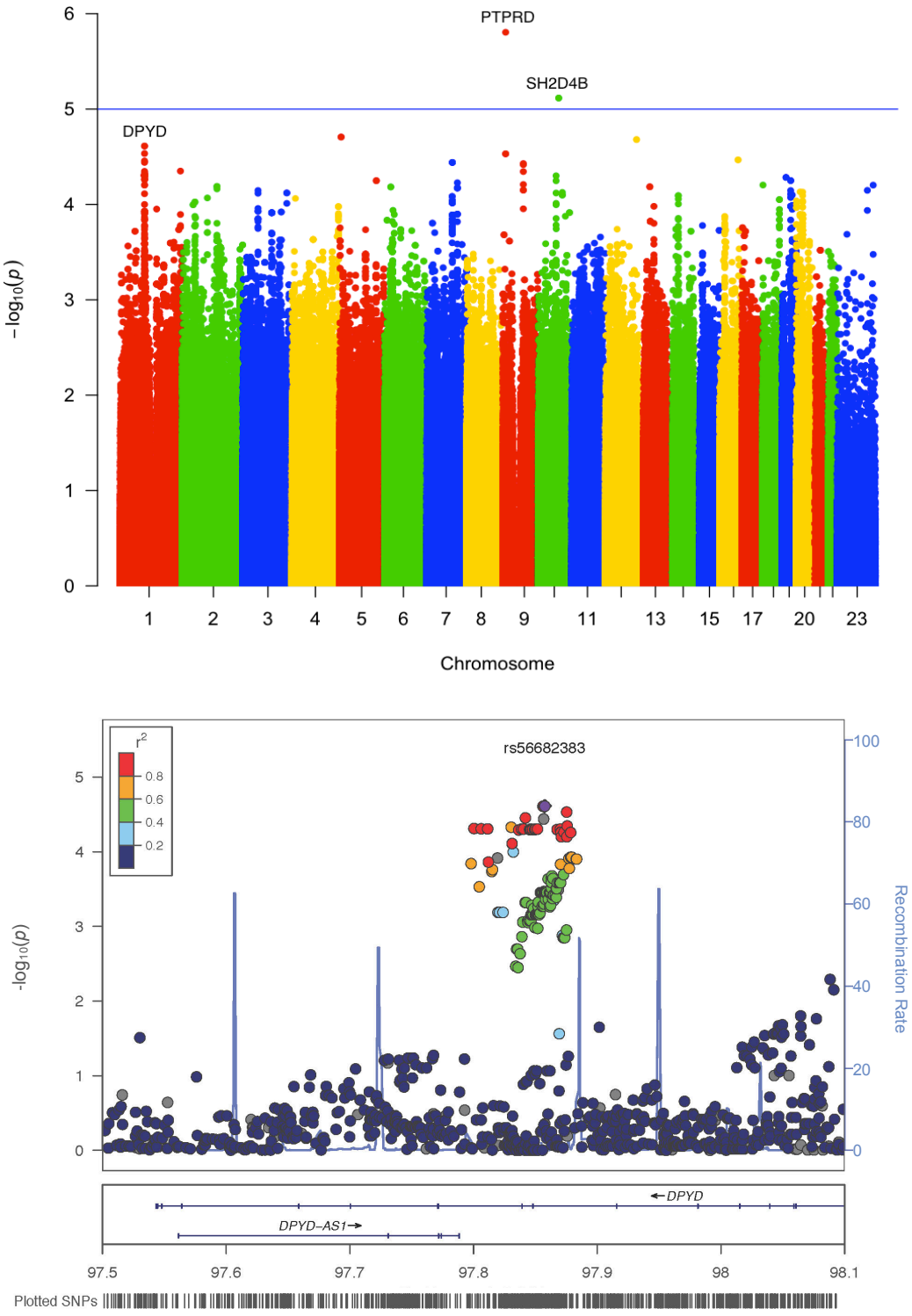


Figure 2: *Top:* Manhattan plot showing association p-values for the merged UKBA array and imputed marker genotype set, which contains 5,757,137 markers (note: blue line at $-log_{10}(p) = 5$ is for reference and does not denote significance level); *Bottom:* Regional association plot of the chromosome 1 region partially containing the *DPYD* locus. SNPs are colored by the extent of LD with the top SNP, rs56682383, shown in purple.

SNP	CHR	BP	MAF	Gene	p-value
rs202194476	9	10356124	0.24	PTPRD	$1.6x10^{-6}$
rs11186481	10	82419693	0.16	SH2D4B (near)	$7.7x10^{-6}$
rs10058390	5	6576320	0.31	LOC255167 (near)	$2.0x10^{-5}$
rs79201062	12	124719241	0.06	ZNF664-FAM101A	$2.1x10^{-5}$
rs56682383	1	97857539	0.15	DPYD	$2.4X10^{-5}$
rs12039093	1	97856036	0.16	DPYD	$2.4X10^{-5}$
rs74106717	1	97875200	0.15	DPYD	$2.9X10^{-5}$
rs10125887	9	10356509	0.26	PTPRD	$2.9X10^{-5}$
rs9673361	16	73930273	0.16	-	$3.4X10^{-5}$
rs12040045	1	97841850	0.17	DPYD	$3.5X10^{-5}$
rs199503697	1	97856601	0.15	DPYD	$3.6X10^{-5}$
rs13233554	7	101660862	0.48	CUX1	$3.6X10^{-5}$
rs13243790	7	101666868	0.48	CUX1	$3.6X10^{-5}$
rs7809609	7	101667468	0.48	CUX1	$3.6X10^{-5}$
rs144238891	9	81389257	0.16	-	$3.7X10^{-5}$
rs1958928	9	81386049	0.11	-	$3.8X10^{-5}$
rs12408999	1	242214828	0.33	-	$4.5X10^{-5}$
rs11165865	1	97875670	0.17	DPYD	$4.5X10^{-5}$
rs11137832	9	81388856	0.16	-	$4.5X10^{-5}$
rs74105110	1	97830355	0.20	DPYD	$4.7X10^{-5}$
rs74104367	1	97800233	0.18	DPYD	$4.9X10^{-5}$
rs78917795	1	97800419	0.18	DPYD	$4.9X10^{-5}$
rs12024103	1	97806124	0.18	DPYD	$4.9X10^{-5}$
rs74104372	1	97811356	0.18	DPYD	$4.9X10^{-5}$

Table 2: Table listing the 20 markers with the lowest p-values, along with the corresponding minor allele frequency (MAF) and gene. All markers (except for those marked "near") lie within the gene listed.

The second lowest p-value belonged to a SNP located near the *SH2D4B* (Src-homology 2-containing protein 4B) locus (rs11186481, $p = 7.7 \times 10^{-6}$). The gene product SH2D4B is in a family of T-cell specific adaptor proteins potentially involved in T-cell signal transduction although the role of SH2D4B in signal transduction has yet to be fully elucidated [27]. No other SNPs within or near this locus were found to have any sort of a suggestive p-value, possibly because rs11186481 is bound by regions with higher recombination rates (see Supplementary Figure S3).

Another potential region of interest is on chromosome 1. Figure 2 reveals a series of SNPs within a tight region on chromosome 1 that display a marked pattern of similar association

values (see Table 2 as well). These SNPs all fall within the *DPYD* (dihydropyrimidine dehydrogenase) locus (see Figure 2, Bottom). The DPYD enzyme is involved in the catabolism of pyrimidines (uracil and thymine) and mutations in the gene are most commonly implicated in increased toxicity during 5-fluorouracil chemotherapy treatment, due to increased breakdown of the drug [28]. Accumulation of DPYD breakdown products in cells has also been linked to fostering the epithelial to mesenchymal transition and extravasation of tumorigenic cells [29]. Interestingly, DPYD also has neuromodulatory effects, as it is an important regulator of the production of β-alanine, a modulator of inhibitory neurotransmission in the brain [30]. In fact, β-alanine analogs, like clorazepate, are used as adjunctive therapy for the treatment of partial seizures [30], providing a potential link between *DPYD* and VIFS. Furthermore, mutations in *DPYD* have been associated with autism [31], intellectual disability [32], and schizophrenia [33].

Although little is known about VIFS, FS without the context of vaccination has been studied in some detail and a variety of loci have been implicated. The FS-associated genes segregate into two general classes: cytokine genes and loci involved in regulating neuronal activity. Some commonly associated loci are *SCN1A* [18], GABA receptor loci such as *GABRG2* [12], and interleukin genes such as *IL-6* and *IL-1β* [13, 14]. Investigating how strongly some of these commonly cited genes/SNPs are associated in our GWAS could provide evidence as to whether the genetics of VIFS is similar to FS. After analyzing the SNP with the lowest p-value in these genes, there is only nominal significance (p-value < 0.05), if any, for association with VIFS (Table 3).

SNP	Gene	OR	Beta	SE	p-value
rs11890028	SCN1A	0.53	-2.54	0.25	0.011
rs79877185	GABRG2	3.54	2.03	0.62	0.043
rs3917356	IL-1β	0.66	-2.17	0.19	0.030
rs2066992	IL-6	1.28	0.93	0.27	0.355

Table 3: Association statistics for SNPs in this study within genes that have been previously implicated in FS. The SNP with the lowest p-value in each gene is shown.

Using the NHGRI GWAS catalog of published SNP-trait association studies [34], I searched for variants in these genes that had been previously associated with FS or a related neurological disorder like epilepsy. One SNP within the genes listed in Table 3 was shown to have a suggestive p value (rs11890028 in *SCN1A,* $p = 4 \times 10^{-6}$) when tested for association with genetic generalized epilepsy [35]. Although the direction of effect is the same in our study as in [35] (OR = 0.77 for rs11890028 in [35]), given that the p-value of this SNP is at best suggestive in [35] and only nominally significant in our study, more work needs to be done to determine whether this SNP (and the *SCN1A* locus) is truly associated with FS and/or VIFS. Thus, overall there appears to be evidence of only nominal significance to state that common FS genes like *SCN1A, GABARG2, AND IL-1β* are associated with VIFS. Taken together, these results suggest that either our study was underpowered to pick up significant signal from the common FS genes, or that the genetic architecture of VIFS might be distinct from that of FS.

Exome Sequencing Analysis: Gene-Based Association

This study sought to pair a GWAS with exome sequencing analysis to also highlight the potential contribution of rare variation to the development of VIFS. Towards this end, exome sequencing data on all 275 samples was processed, filtered, and analyzed (see Methods). Gene-based association tests were invoked through the SKAT package in R (http://cran.r-project.org/web/packages/SKAT/index.html). Specifically, since very little is known about the genetic architecture of VIFS and what sort of variants might be in our cohort, the optimized sequence kernel association test (SKAT-O) was implemented which boasts increased power over other tests such as SKAT or burden tests when causal, non-causal, and multi-directional variants are present [36]. Using RefSeq gene definitions, the variants within these genomic regions were pooled into gene sets and iteratively tested for association with VIFS, controlling again for sex and the top 10 components of the MDS analysis. In total, 18,964 genes were

tested, which when using Bonferroni correction results in a significance threshold of 2.6×10^{-6} (= 0.05/18,964). The genes with $p < 5 \times 10^{-4}$ are listed in Table 4 below.

Gene	CHR	Gene Annotation	p-value
BATF	14	Basic Leucine Zipper Transcription Factor, ATF-Like	6.1×10^{-6}
DDI2	1	DNA-Damage Inducible 1 Homolog 2	6.5×10^{-5}
NDUFS3	11	NADH-Ubiquinone Oxidoreductase Fe-S Protein 3	8.9×10^{-5}
SSH1	12	Slingshot, Drosophila, Homolog Of, 1	0.00011
TMEM55B	14	Transmembrane Protein 55B	0.00013
TTC12	11	Tetratricopeptide Repeat Domain-Containing Protein 12	0.00019
TBPL2	14	TATA Box-Binding Protein-Like Protein 2	0.00020
ZNF484	9	Zinc Finger Protein 484	0.00021
YY1AP1/ HCCA2	1	YY1-Associated Protein/Hepatocellular Carcinoma-Associated Protein 2	0.00023

Table 4: Gene-based association tests on exome sequencing data. Genes with p-value < 5×10^{-4} are shown. SKAT-O was used to perform association testing (see Methods).

The top result (nearly reaching exome-wide significance with $p = 6.1 \times 10^{-6}$) is the *BATF* locus (Basic Leucine Zipper Transcription Factor, ATF-Like) which has been shown to be a critical regulator of T-cell and B-cell activity. For example, loss of BATF results in reduced CD8(+) T-cell function, limited T-cell related immunopathology, and promotion of viral persistence after infection with the lymphocytic choriomeningitis virus [37]. Additionally, it has also been shown that BATF-induced regulation of T-cell activity and cytokine production is important in a variety of other contexts, from T-cell exhaustion in HIV [38] to the autoimmunity characteristic of asthma [39]. Recently, Tussiwand et al also implicated BATF in the development of dendritic cells operating during infection and mediated by the cytokines interleukin IL-12 and IFN-γ [40]. This result is particularly interesting since the response to measles is known to involve both IL-12 and IFN-γ, along with the fact that the measles virus is capable of immunosuppression and down-regulation of IL-12 production [41].

The next two most associated genes are *DDI2* (DNA-damage inducible 1, homolog 2) and *NDUFS3* (NADH-Ubiquinone Oxidoreductase Fe-S Protein 3), both of which are involved in cell death. Although little is known about *DDI2,* it is a Bcl-2 homolog and thus likely involved in

the induction of apoptosis. *NDUFS3* has been better characterized and is a member of the electron transport chain in mitochondria [42]. Mitochondrial disruption is known to be a means of cell death [43] and Martinvalet et al recently showed that granzymes produced by cytotoxic, activated T-cells are responsible for inducing the cleavage of *NDUFS3,* disrupting electron transport, producing reactive oxygen species, and thus catalyzing initiation of the apoptotic pathway [42].

As was done for the GWAS data (Table 3), the level of association for some genes commonly implicated in FS (e.g. *SCN1A, GABARG2, IL-1β,* and *IL-6*) was investigated (Table 5). Similar to the results for the GWAS analysis (Table 3), the lack of significance of any of these genes suggests that either our study was underpowered to pick up significant signal from the common FS genes, or that the genetic architecture of VIFS might be distinct from that of FS.

Gene	All Markers	Markers Tested	p-value
SCN1A	250	160	0.18
IL6	36	26	0.57
IL-1β	35	28	0.16
GABARG2	60	47	0.25

Table 5: Gene-based association test results for genes commonly implicated in FS. SKAT-O was used to perform association testing (see Methods). Along with the p-value, the results indicate the number of markers/variants within the gene that was tested ("All Markers") as well as the number of markers included in the association testing after removing those with a high missing rate or those that are non-polymorphic ("Markers Tested").

Exome-Sequencing Analysis: High Effect Variants

The analysis of the SKAT-O results (Table 4) suggests that both immune system activity (in terms of T-cell, B-cell, and dendritic cell activity) and apoptosis related genes could be associated with VIFS. Given the rarity of VIFS (~0.1% incidence), it seems likely that VIFS-associated variants are rare and are of high impact (e.g are non-synonymous, frameshifts, etc.). To explore this possibility and potentially gain better resolution by eliminating non-causal variants from the analysis, I performed SKAT-O again, this time after restricting the analysis to variants that are predicted to be of "High" impact (see Methods). This restriction limited the

analysis to 6,084 genes, resulting in a multiple-testing corrected significance threshold of 8.2 x 10^{-6} (= 0.05/6,084). The top results of this analysis are shown in Table 6 below.

Gene	CHR	Gene Annotation	p-value
DEFB126	20	Defensin, Beta 126	2.1×10^{-5}
BCL2L11	2	BCL2-Like 11	0.00040

Table 6: Gene-based association tests on exome sequencing data after restricting for variants with predicted "high" impact. Genes with p-value < 5 x 10^{-04} are shown. SKAT-O was used to perform association testing.

Similar to the results in Table 3, Table 6 highlights the potential importance of two classes of genes: 1.) genes critical for regulating the response to pathogens (i.e. *DEFB126)* and 2.) genes involved in the apoptotic pathway (i.e. *BCL2L11*). DEFB126 (Defensin-Beta 126) is a member of the β-defensin family of proteins which possess anti-microbial activities, primarily through induction of pathogen membrane permeabilization, and thus serves as a first-line of defense against infection [44]. In addition to this, recent work has also shown that DEFB126 can not only neutralize invading pathogens, but it also has potent anti-inflammatory effects, including regulation of key cytokines such as IL-α, IL-1β, IL-6 and TNF-α [45].

BCL2L11 (Bcl2-Like L11) is a member of a family of pro-survival genes involved in the apoptotic pathway. In addition to its well known role in tumorigenesis and lymphoma [46], tight regulation of the Bcl2/Bim apoptotic pathway is also a key component during the course of an immune response to pathogens. Temporal regulation of pro-survival (e.g. Bcl2) and pro-apoptotic (e.g. Bim) proteins is key to ensuring the correct timing and magnitude of immune cell proliferation and death during an infection [47]. Recent work has shown that Bcl2's binding partner, Bim, is critical to the downregulation of T cell activity following clearance of viral burden [47]. In this work, the absence of Bim in mice did not affect viral clearance, but did prevent normal T-cell death after viral clearance and cytokine withdrawal.

The common FS genes listed in Table 5 did not contain variants of predicted "High" effect in our study and thus were not analyzed after restricting for high-impact variants.

DISCUSSION

By utilizing advanced genotyping and sequencing technologies, researchers have the unique opportunity of elucidating variants, genes, or pathways that are involved in a variety of diseases with unknown or complex etiologies. Here I report the results of a genome-wide association and exome sequencing study aimed at highlighting the genetic determinants of vaccine-induced febrile seizure (VIFS). Despite the fact that our study was underpowered as a result of a small sample size, our results not only provide the first look at the landscape of genome-wide variation that is relevant in VIFS but they also help to generate hypotheses about the underlying genetic architecture of VIFS.

Genetics of VIFS

In this study, none of the classic FS or epilepsy related genes (such *SCN1A, IL-1β, or GABARB2*) showed highly significant signs of association with VIFS. Given that there is substantial genetic variability even between the different subtypes of FS (e.g. *SCN1A* is associated with epilepsy and some forms of generalized epilepsy with febrile seizure [48], but not associated to FS in other disease contexts [49]), it seems reasonable to assume that the genetics underlying VIFS could be distinct as well. Furthermore, the unique nature of VIFS (i.e. it occurs in the context of an acute immune response, is usually non-recurrent, and results in no future neurological problems), would further suggest that the genetic landscape is specific to VIFS and different from that of common forms of recurrent FS or epilepsy. After all, if genes that are involved in epilepsy or recurrent FS were implicated in VIFS, it would be a challenge to explain why few VIFS patients ever present with recurrent episodes of FS. Thus, although the results of this study did not reach genome- or exome-wide significance, they do support the hypothesis that VIFS is a distinct phenotype and the genes involved in predisposing to VIFS are different from those in other neurological disorders, such as FS or epilepsy.

Given that VIFS is a neurological effect that occurs in the presence of a strong immune response, an interesting question is whether the mutational burden of a VIFS patient is enriched

for neurological or immunological genes. While the GWAS does suggest genes involved in the regulation of neuronal activity (*PTPRD, DPYD),* the common theme between the GWAS and exome sequencing data is the presence of genes that are either regulators of immune cell activity and cytokine production (*PTPRD, DEFB126)* or that might play a role in programmed immune cell death during the course of an immune response (*NDUFS3, BCL2L11*). Thus, it seems possible that VIFS is a result of an abnormal immune response, resulting in fever exacerbation (via overproduction of pyrogenic cytokines [8]) and/or increased production of neuromodulatory cytokines [8]. With a background of mutations that modestly affects neurotransmission and/or neuronal activity (such as *DPYD, PTPRD*), this increase in immune activity could potentially trigger VIFS in a subset of susceptible patients. Taken together, our results suggest that the primary driving force behind VIFS might be a regulation (or deregulation) of the immune response after vaccination which enhances the risks of seizure. Of course, given the limitations of this study, these results need to be properly corroborated to determine if they are actually significant and reproducible.

Limitations and Future Directions

The main limitation of this study was the fact that our cohort was relatively small (reducing our power to detect association). In addition to including more case/control pairs, one way to quickly and cheaply increase power in the future is to recruit and sequence additional controls (or utilize existing databases), a strategy which is known to increase power [50]. Our association testing was also limited by the fact that it did not include a variety of covariates, like age at vaccination, that are known to be relevant in this study [51] (these data were not available at the time of analysis). In addition to these main limitations, our study contained other potentially limiting factors which could be addressed in future work. For example, the way in which controls were selected for this study could have led to misclassification bias. As outlined in the Methods, the controls were children who did not experience VIFS and who also did not experience fever (within 7-10 days after vaccination). Given that this was the first extensive

genetic study of VIFS, sample recruitment followed the reasonable rationale of excluding any control that might bear similar genetic traits (in the form of a shared propensity for fever) to the population of cases. However, given that this was a retrospective study that relied on hospital records to determine the existence of vaccine-induced fever, it is conceivable that some controls did in fact exhibit fever but did not seek any medical attention. Even if all controls were accurately classified, though, restricting controls to those who did not experience fever could confound the analysis and prevent the study from discriminating between vaccination-induced fever and vaccination-induced febrile seizure. For example, in a case where the immunomodulatory genes *PTPRD and BATF* are actually significantly associated with VIFS, it would be unclear if they are involved in vaccination-induced fever or direct modulation of seizure risk. By setting up a control population that experienced fever but not seizure, the study could then distinguish between immunological or other genetic factors that are relevant in fever-induction and factors that specifically and directly result in an increased risk of febrile seizure in the context of vaccination. Furthermore, this control selection strategy would reduce misclassification bias since it is easier to identify controls that experience fever but not seizure compared with choosing controls that did not experience either. Lastly, given that measles-containing vaccines commonly induce fever (~25% of the time [52]) while VIFS is relatively rare (~0.1%), the shared genetic traits between those who experience fever and those who experience fever + seizure seem unlikely to significantly overlap.

Another potential limitation of this study was that cases and controls were chosen without regard to their vaccination history/schedule. Given that other vaccines are associated with an increased risk of VIFS (e.g. pertussis) and that there are differences in risk of VIFS when administering different vaccines (such as MMR versus the MMRV vaccine) [7], vaccination history is potentially an important factor. A pre-existing or coincident modulation of the immune system by vaccination, regardless of whether an administered vaccine is associated with febrile seizure or not, could increase (or decrease) the risk of VIFS upon

vaccination with a measles-containing vaccine. If feasible, future work on VIFS should control for vaccination history and concurrent vaccinations.

CONCLUSION

Although this study has its limitations, as the first extensive investigation into the genetics of vaccine-induced febrile seizure, it has laid the groundwork for future studies. Furthermore, this work highlights potential genes of interest in VIFS, including those involved in modulating the immune system response (*BATF, PTPRD, DEFB126),* apoptosis (*BCL2L11, NDUFS3*), and neuronal activity (*PTPRD, DPYD).* By taking the first step towards understanding the full range of effects of the common measles-containing vaccines, this study will help instill greater trust in the vaccination effort and potentially guide development of a pharmacogenetic strategy to predict and prevent the onset of VIFS.

METHODS

Study Design

Ethics Statement and Informed Consent

The healthcare providers of potential study participants were notified by letter that their patients might be eligible for the study. In this letter, we requested permission to contact their patients, allowing them 3 weeks to respond. They were also informed that we would interpret a lack of response within these 3 weeks as permission to contact the patient. Once the provider gave permission, or if they did not respond within three weeks after notification, eligible cases and controls were mailed a letter that introduced the study, the scope of their participation, and included information on how to easily opt out of the study and any future contact (telephone numbers to call for information along with an opt-out mailer and postage-paid envelope). After waiting 2-3 weeks for a potential reply, a research assistant called all participating potential cases and controls to offer additional information, answer questions, conduct a short interview, and seek verbal consent to participate in the study. For cases whose parents consent to participate who meet the definition of extreme-trait phenotype, we also asked about first-degree

family members who might also wish to consent to participate. We enrolled these family members at that time if they were available at the time of the interview with the case or at a later date if unavailable. Separate telephone scripts/interviews were used depending on whether the adult respondent is answering as the parent of a case or control or for him/herself as a family member control. Scripts scenarios included one for child case, family member available at time of interview with case (adult or child), family member not available at time of interview with case (adult or child), and the control script for non-family members (adult or child). Following the round of telephone interviews, each study participant was then sent informed consent forms (and assent forms if participating child was over the age of 7) along with saliva kits. Forms were signed and returned via the included postage-paid envelope.

Identification of cases/controls, extreme cases, and inclusion criteria

In order to participate, a case had to meet the following criteria: 1.) Have experienced an acute episode of chart-confirmed, physician-diagnosed FS in the emergency room or inpatient setting; 2.) Have received any measles-containing vaccine 7 to 10 days prior to the FS event (i.e., MMRV or MMR) regardless of other vaccines given; 3.) Patient must have been ≤ 5 years at time of FS event. These criteria are the same as those in a previous study [53]. All cases were pooled from the CDC's Vaccine Safety Datalink (VSD) participating institution Kaiser Permanente-Northern California's (KPNC) records dating back to January 2000. Cases from January 2000 through October 2008 were identified as part of a previous VSD study [53]. These cases have already been chart confirmed as having a physician diagnosis of FS. Additional potential cases were identified via searching KPNC databases for ICD-9 seizure codes using the same parameters as used in the original VSD study [53]. Following identification, potential cases underwent a brief medical record review to assess whether the individual met inclusion criteria. Furthermore, we identified a subgroup of cases known to exhibit an "extreme-trait phenotype". Children in the extreme-trait phenotype sub-group are defined as cases who meet all inclusion criteria, but who also had a generalized FS lasting >10 minutes.

Controls were KPNC members who were unrelated to cases and who were identified within the electronic VSD database as having received a measles-containing vaccine without reported history of any seizure or record of a medically-attended visit for fever during the 7-10 days following receipt of a measles-containing vaccine. The rationale for excluding children who had medically-attended visit for fever 7-10 days after a measles-containing vaccine was out of concern that such children may have similar biologic/genetic traits as children who experienced a febrile seizure 7-10 days after receipt of a measles-containing vaccine. Controls were matched to cases based on age at vaccination, medical facility, race, and the specific vaccine which was administered to the matched case prior to the FS (i.e. MMR, MMRV). For those controls who were less than 6 years of age at the time of enrollment, KPNC determined at the time of genome analysis if any of these controls experienced a FS subsequent to study enrollment.

DNA Sample collection

All study participants were sent informed consent forms (and assent forms for children over 7 years of age) to be signed, an age-appropriate saliva specimen kit to be filled, and a pre-addressed, postage-paid envelope for return. All materials were marked with pre-labeled and matching study identification numbers. Specimens were banked at the KPNC's Division of Research Biobank Repository and completed consents/assents were stored at KPNC's Division of Research.

GWAS Data and Analysis

Chip Design and summary statistics

We used the UK Biobank Axiom Array (UKBA) to genotype participants. The UKBA contains 820,967 total markers, including SNPs, indels, and CNVs. 94,490 of these markers are associated with or have possible roles in phenotypic variation, 111,904 markers are coding variants (principally missense and protein truncating), and 629,368 markers provide genome-wide coverage. Although the UKBA was designed for Caucasian populations the coverage of

other ethnicities has been shown to be comparable to chips designed for other populations. In total, all 275 samples were genotyped at 813,270 markers. Summary statistics of the genotyping dataset are: average missingness per individual: 0.2%; average missingness per marker: 0.2%; number of individuals/variants with missingness greater than 5%: 0. There was no significant difference in the missing genotyping rate between cases (0.22%) and controls (0.25%).

Imputation

Genome-wide imputation was performed via IMPUTE2 [20], using the 1000 Genomes Phase I integrated reference haplotypes dataset (NCBI Build 37). This reference dataset contains genotype likelihoods for 36,820,992 SNPs, 1,384,273 short bi-allelic indels and 14,017 structural variations, and includes data from 14 different populations. The populations included in the reference dataset are (given as "population (# of samples)"): ASW (61), CEU (85), CHB (97), CHS (100), CLM (60), FIN (93), GBR (89), IBS (14), JPT (89), LWK (97), MXL (66), PUR (55), TSI (98), and YRI (88). Both the study and reference data were aligned to the + strand. Prior to imputation the study data was separated by chromosome and each chromosome was then segmented into 5mb regions (with a buffer of +/-250 kb) and imputed using an effective population size of 20,000. IMPUTE's original algorithm, which integrates over the unknown phase of the study data during the course of the imputation analysis and thus does not rely on pre-phasing of study haplotypes, was implemented. Although a more computationally intensive method, this method provides better imputation accuracy. IMPUTE2 calculates imputation accuracy by masking and genotyping study variants, and providing concordance rates between the masked/imputed genotypes and the actual study genotypes. Imputation accuracy was strong, with an overall concordance rate for each imputed region > 96%. In total, imputation generated 37,216,202 markers. This dataset was merged with the UKBA genotypes (813,270 markers) and the association study was performed as described below.

Association Study

PLINK (v.1.07) [19] was used to apply a logistic regression test to analyze the association between UKBA markers and case/control status (PLINK v1.9 was used for some additional data management such as removing duplicate SNPs). The results were adjusted for the following covariates: sex and the first ten components of a multidimensional scaling analysis (produced via PLINK v.1.07). Although inclusion of additional covariates is necessary for this particular study (e.g. age at vaccination, type of measles-containing vaccine administered, severity or length of seizure, etc.) this data was unavailable at the time of analysis. Of the 142 controls, 20 were family controls (first-degree relatives to a case), chosen for the study to help in potential cosegregation analysis. These, along with a duplicate sample, were removed from the analysis. Variants were included if their call rate > 95%, minor allele frequency > 0.01; samples were included if genotyping rate was > 90%. No samples were removed due to missing genotypes, but 4,606,974 and 28,486,225 markers were removed due to failing call rate and minor allele frequency tests, respectively. Ultimately, 5,757,137 markers were included in the analysis shown in Figure 2.

Manhattan, Regional Association, and QQ Plots

Manhattan and QQ plots were created via the R-package 'qqman' (http://biorxiv.org/content/early/2014/05/14/005165) using association statistics generated via PLINK (v.1.07). LocusZoom [54] was used to generate the regional association plots. To calculate LD, the hg19/1000 Genomes (Mar 2012) EUR dataset was used.

Exome Sequencing Data and Analysis

Variant Calling-Primary Analysis

We exome sequenced the 275 samples on 55 lanes using Illumina technology, achieving a mean target coverage of >70x and reads of ~100bp long. The exome was defined by the intervals in Roche's SeqCap EZ Human Exome Library v3.0. I then took the resulting reads and aligned them to the hg19 reference genome with BWA using the BWA-mem

algorithm (Burrows Wheeler Aligner) [55], removed duplicates and sorted reads with Picard (http://picard.sourceforge.net), and applied GATK (the Genome Analysis Toolkit) [56] for purposes of base quality score recalibration, indel realignment, SNP and INDEL discovery and genotyping (GATK's HaplotypeCaller was used to perform the variant calls) across all 275 samples simultaneously using variant quality score recalibration according to the GATK Best Practices recommendations [57] (http://www.broadinstitute.org/gatk/guide/bestpractices?bpm=DNAseq). Variants were called with a 1000 base pair buffer around the Roche library target intervals to potentially capture noncoding variation present in our dataset.

Variant Annotation

The VCF file produced by the variant calling pipeline above was functionally annotated with SnpEff [58] using the GRCh37.74 reference genome. Variants were assigned fields such as predicted effect (e.g. non-synonymous), impact (e.g. high, moderate, low), gene, and codon change, among others. The resultant annotated VCF file was processed by SnpSift [59] to calculate for each variant the frequency of homozygous, heterozygous, and total allele counts in cases and controls. SnpSift was also used to filter the VCF file based on SnpEff annotations.

Association Analysis

The R package 'SKAT' was used to perform the gene-based association analyses. SKAT-O is an optimized sequence kernel association test which performs association tests on user-defined gene sets, focusing on rare variant effects [36]. UCSC's Table Browser (http://genome.ucsc.edu/cgi-bin/hgTables) was used to generate a list of genes based on RefSeq gene definitions, with the GRCh37/hg19 serving as the reference genome. SKAT-O analysis was performed under two general conditions: 1.) including all variants that passed GATK filtering, and 2.) all variants that passed GATK filtering that also had a SnpEff predicted "High" impact on the gene. Annotated (and filtered) VCF files were first converted to PLINK compatible binary file sets via vcftools [60], read into SKAT, and used to calculate model parameters and residuals of the null model of no association between genetic variables and the

outcome phenotype via SKAT's Null Model function. Additionally, since when dealing with small sample sizes and binary traits large-sample based p-value calculations can produce overly conservative results [36], SKAT's small sample-size adjustment was invoked which increases the power of detection for small sample sizes (see http://cran.rproject.org/web/packages/SKAT/SKAT.pdf).

Supplementary Figures

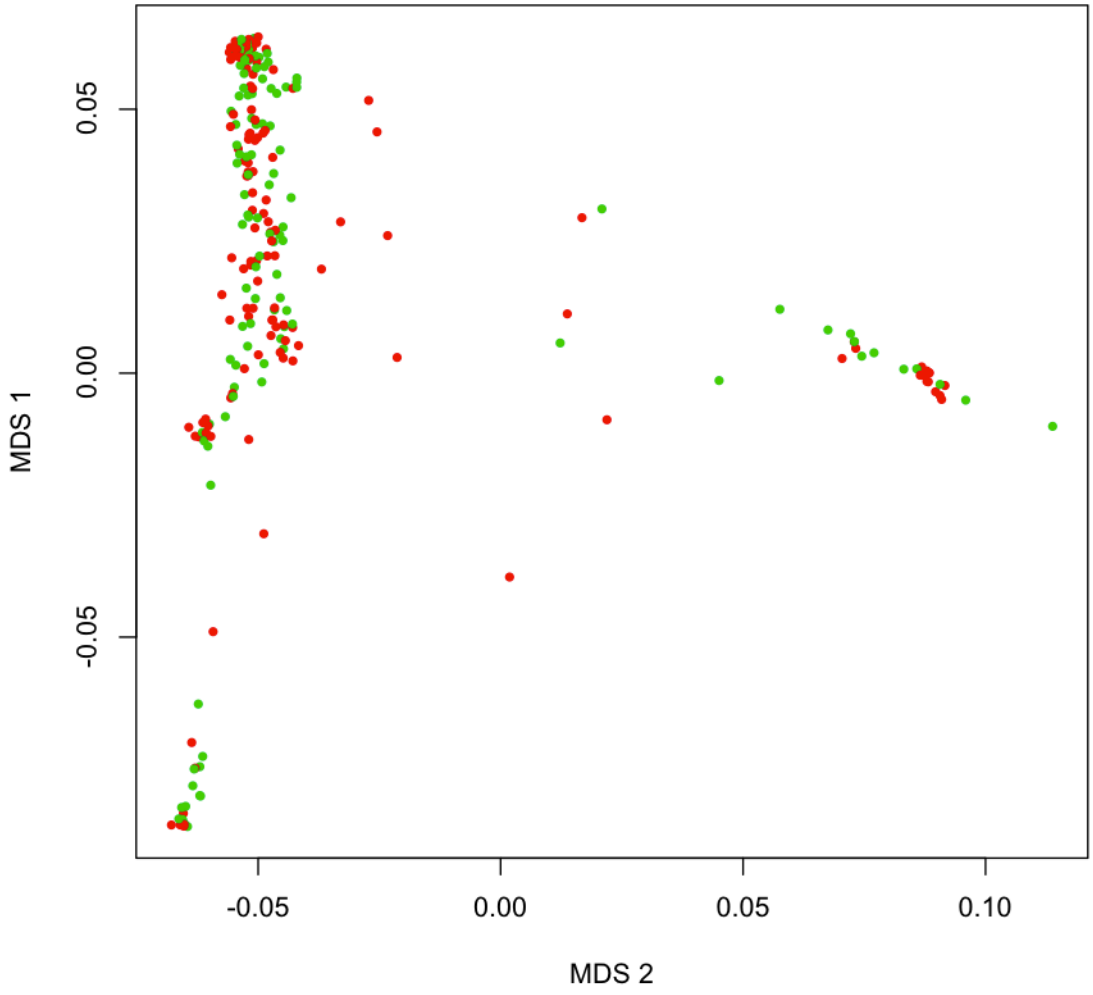


Figure S1: Distribution of cases (red) and controls (green) on the basis of the top two MDS components. Note the samples that appear to not have a closely matched case (for controls) or control (for cases).

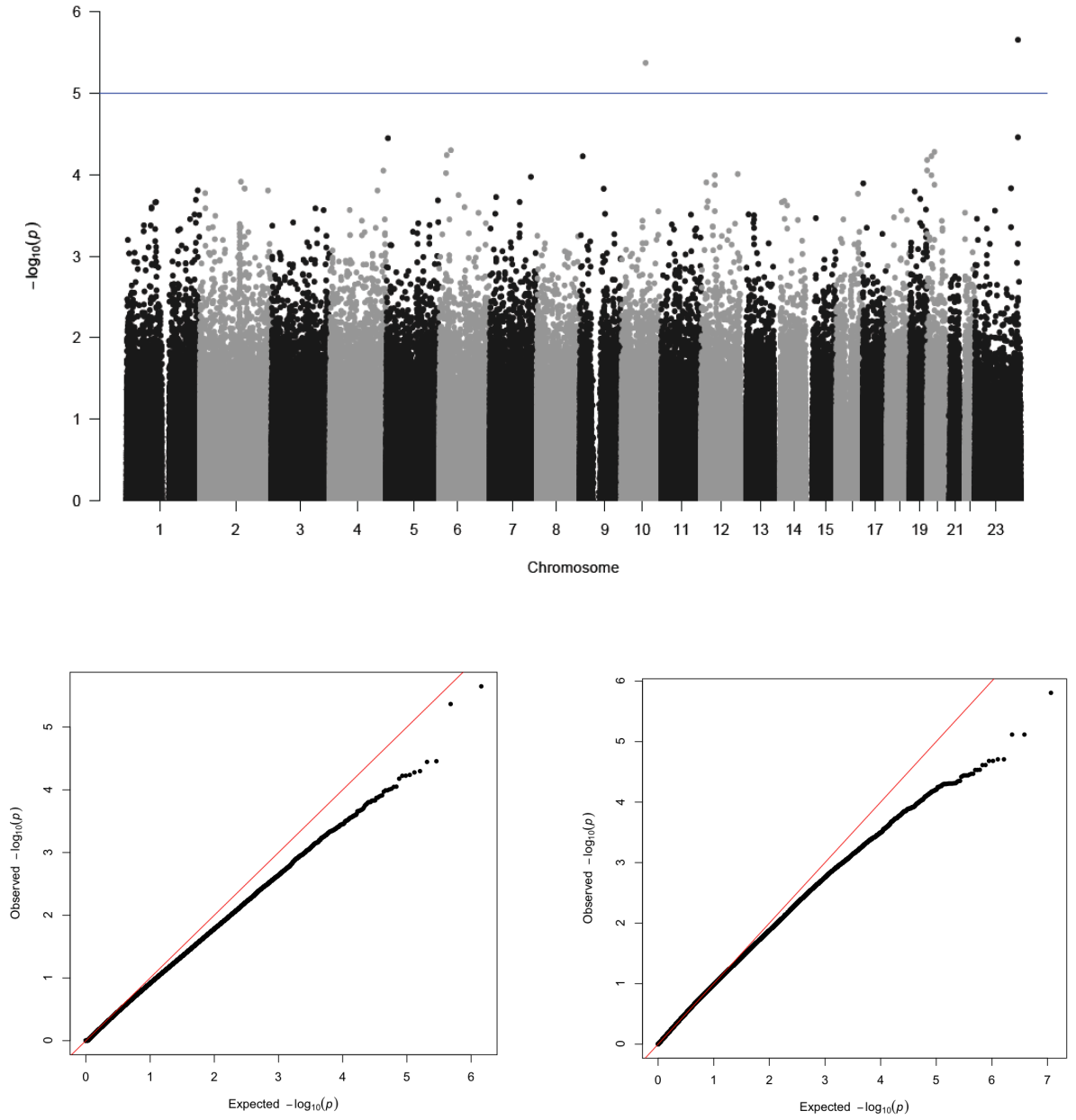
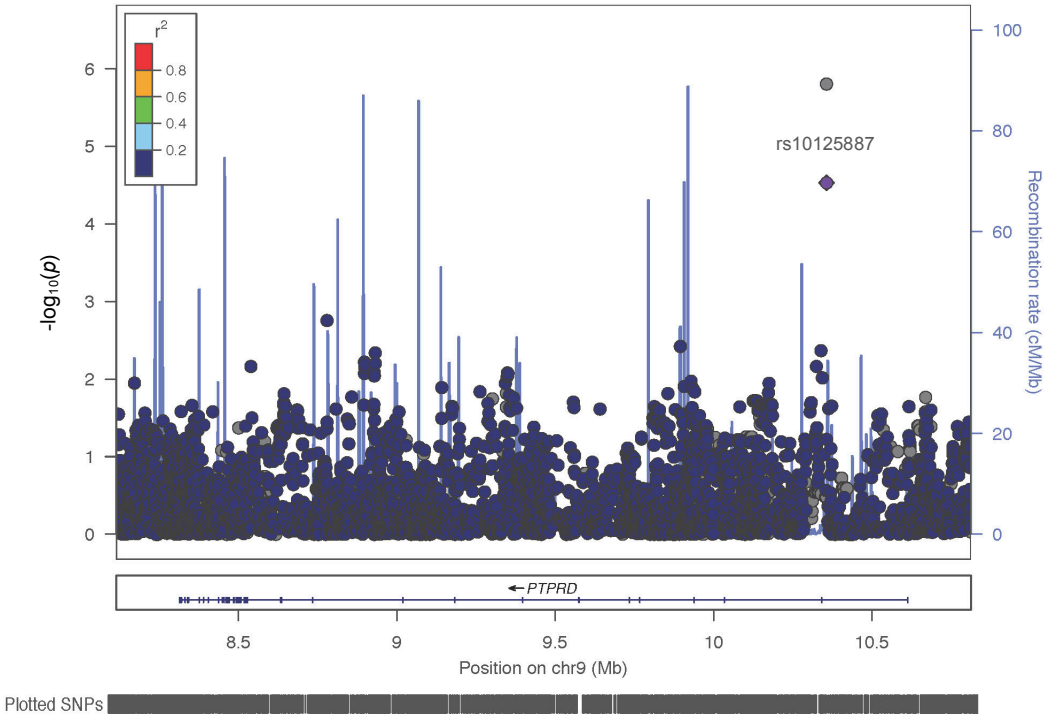


Figure S2: *Top:* Manhattan plot showing association p-values for the UKBA array prior to imputation and without filtering for minor allele frequency or missiningess. 813,270 markers are included. (note: blue line at $-log_{10}(p) = 5$ is for reference only and does not denote significance level); *Bottom:* Quantile-Quantile plots of p-values in Manhattan plot before (bottom-left) and after (bottom-right) imputation, showing a lack of enrichment for markers with low p-values.



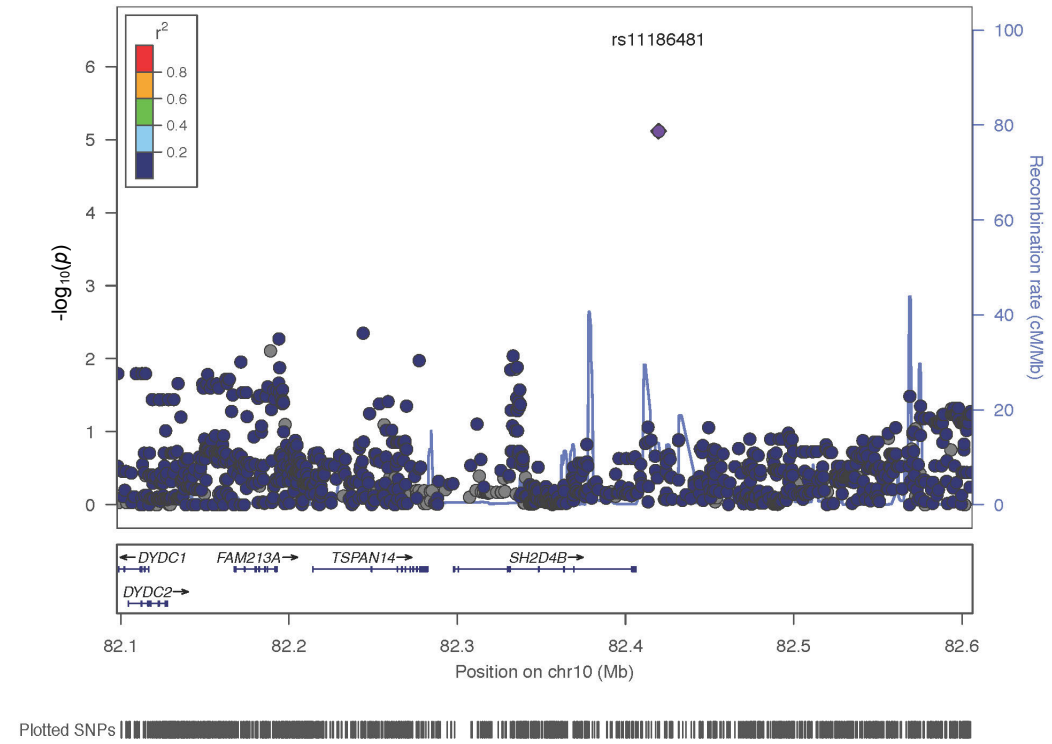


Figure S3: *Top:* Regional association plot of the *PTPRD* locus, containing the top SNP in the GWAS, rs202194476. SNPs are colored by the extent of LD with rs10125887 since LD data was unavailable for rs202194476. *Bottom:* Regional association plot of the chromosome 10 region containing rs11186481, near the *SH2D4B* locus. SNPs are colored by the extent of LD with the top SNP, rs11186481. For both plots recombination rates are overlaid.

References

1. Gahr, P., et al., *An outbreak of measles in an undervaccinated community.* Pediatrics, 2014. **134**(1): p. e220-8.

2. Sugerman, D.E., et al., *Measles outbreak in a highly vaccinated population, San Diego, 2008: role of the intentionally undervaccinated.* Pediatrics, 2010. **125**(4): p. 747-55.

3. Kline, J.M., et al., *Pertussis: a reemerging infection.* Am Fam Physician, 2013. **88**(8): p. 507-14.

4. Glanz, J.M., et al., *Association between undervaccination with diphtheria, tetanus toxoids, and acellular pertussis (DTaP) vaccine and risk of pertussis infection in children 3 to 36 months of age.* JAMA Pediatr, 2013. **167**(11): p. 1060-4.

5. Wakefield, A.J., et al., *Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children.* Lancet, 1998. **351**(9103): p. 637-41.

6. *Retraction--Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children.* Lancet, 2010. **375**(9713): p. 445.

7. Maglione, M.A., et al., *Safety of Vaccines Used for Routine Immunization of US Children: A Systematic Review.* Pediatrics, 2014. **134**(2): p. 325-337.

8. Choy, M., et al., *Inflammatory processes, febrile seizures, and subsequent epileptogenesis.* Epilepsy Curr, 2014. **14**(1 Suppl): p. 15-22.

9. Annegers, J.F., et al., *Factors prognostic of unprovoked seizures after febrile convulsions.* N Engl J Med, 1987. **316**(9): p. 493-8.

10. Wilson, K., et al., *Adverse events following 12 and 18 month vaccinations: a population-based, self-controlled case series analysis.* PLoS One, 2011. **6**(12): p. e27897.

11. Kang, J.Q., W. Shen, and R.L. Macdonald, *Why does fever trigger febrile seizures? GABAA receptor gamma2 subunit mutations associated with idiopathic generalized epilepsies have temperature-dependent trafficking deficiencies.* J Neurosci, 2006. **26**(9): p. 2590-7.

12. Todd, E., et al., *GABAA receptor biogenesis is impaired by the γ2 subunit febrile seizure-associated mutation, GABRG2(R177G).* Neurobiol Dis, 2014. **69**: p. 215-24.

13. Shahrokhi, A., et al., *Association of IL6 single nucleotide polymorphisms with febrile seizures.* J Neurol Sci, 2014. **342**(1-2): p. 25-8.

14. Dubé, C., et al., *Interleukin-1beta contributes to the generation of experimental febrile seizures.* Ann Neurol, 2005. **57**(1): p. 152-5.

15. Kaputu Kalala Malu, C., et al., *[Epidemiology and characteristics of febrile seizures in children].* Rev Med Liege, 2013. **68**(4): p. 180-5.

16. Tsuboi, T., *Genetic analysis of febrile convulsions: twin and family studies.* Hum Genet, 1987. **75**(1): p. 7-14.

17. Corey, L.A., et al., *The occurrence of epilepsy and febrile seizures in Virginian and Norwegian twins.* Neurology, 1991. **41**(9): p. 1433-6.

18. Mantegazza, M., et al., *Identification of an Nav1.1 sodium channel (SCN1A) loss-of-function mutation associated with familial simple febrile seizures.* Proc Natl Acad Sci U S A, 2005. **102**(50): p. 18177-82.

19. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.

20. Howie, B.N., P. Donnelly, and J. Marchini, *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.* PLoS Genet, 2009. **5**(6): p. e1000529.

21. Kennedy, R.B., et al., *Genome-wide SNP associations with rubella-specific cytokine responses in measles-mumps-rubella vaccine recipients.* Immunogenetics, 2014. **66**(7-8): p. 493-9.

22. Takahashi, H., et al., *Selective control of inhibitory synapse development by Slitrk3-PTPδ trans-synaptic interaction.* Nat Neurosci, 2012. **15**(3): p. 389-98, S1-2.

23. Speed, D., et al., *A genome-wide association study and biological pathway analysis of epilepsy prognosis in a prospective cohort of newly treated epilepsy.* Hum Mol Genet, 2014. **23**(1): p. 247-58.

24. Elia, J., et al., *Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes.* Mol Psychiatry, 2010. **15**(6): p. 637-46.

25. Schormair, B., et al., *PTPRD (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome.* Nat Genet, 2008. **40**(8): p. 946-8.

26. Ortiz, B., et al., *Loss of the tyrosine phosphatase PTPRD leads to aberrant STAT3 activation and promotes gliomagenesis.* Proc Natl Acad Sci U S A, 2014. **111**(22): p. 8149-54.

27. Lapinski, P.E., et al., *The T-cell-specific adapter protein family: TSAd, ALX, and SH2D4A/SH2D4B.* Immunol Rev, 2009. **232**(1): p. 240-54.

28. Offer, S.M., et al., *Comparative functional analysis of DPYD variants of potential clinical relevance to dihydropyrimidine dehydrogenase activity.* Cancer Res, 2014. **74**(9): p. 2545-54.

29. Shaul, Y.D., et al., *Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition.* Cell, 2014. **158**(5): p. 1094-109.

30. Tiedje, K.E., et al., *Beta-alanine as a small molecule neurotransmitter.* Neurochem Int, 2010. **57**(3): p. 177-88.

31. Carter, M.T., et al., *Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder.* Clin Genet, 2011. **80**(5): p. 435-43.

32. Willemsen, M.H., et al., *Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability.* J Med Genet, 2011. **48**(12): p. 810-8.

33. Xu, B., et al., *De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia.* Nat Genet, 2012. **44**(12): p. 1365-9.

34. Welter, D., et al., *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1001-6.

35. Steffens, M., et al., *Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32.* Hum Mol Genet, 2012. **21**(24): p. 5359-72.

36. Lee, S., M.C. Wu, and X. Lin, *Optimal tests for rare variant effects in sequencing association studies.* Biostatistics, 2012. **13**(4): p. 762-75.

37. Grusdat, M., et al., *IRF4 and BATF are critical for CD8$^+$ T-cell function following infection with LCMV.* Cell Death Differ, 2014. **21**(7): p. 1050-60.

38. Quigley, M., et al., *Transcriptional analysis of HIV-specific CD8+ T cells shows that PD-1 inhibits T cell function by upregulating BATF.* Nat Med, 2010. **16**(10): p. 1147-51.

39. Übel, C., et al., *The activating protein 1 transcription factor basic leucine zipper transcription factor, ATF-like (BATF), regulates lymphocyte- and mast cell-driven immune responses in the setting of allergic asthma.* J Allergy Clin Immunol, 2014. **133**(1): p. 198-206.e1-9.

40. Tussiwand, R., et al., *Compensatory dendritic cell development mediated by BATF-IRF interactions.* Nature, 2012. **490**(7421): p. 502-7.

41. Griffin, D.E. and M.M. Oldstone, *Measles. Pathogenesis and control. Introduction.* Curr Top Microbiol Immunol, 2009. **330**: p. 1.

42. Martinvalet, D., et al., *Granzyme A cleaves a mitochondrial complex I protein to initiate caspase-independent cell death.* Cell, 2008. **133**(4): p. 681-92.

43. Chowdhury, D. and J. Lieberman, *Death by a thousand cuts: granzyme pathways of programmed cell death.* Annu Rev Immunol, 2008. **26**: p. 389-420.

44. White, S.H., W.C. Wimley, and M.E. Selsted, *Structure, function, and membrane integration of defensins.* Curr Opin Struct Biol, 1995. **5**(4): p. 521-7.

45. Liu, H., et al., *Human beta-defensin DEFB126 is capable of inhibiting LPS-mediated inflammation.* Appl Microbiol Biotechnol, 2013. **97**(8): p. 3395-408.

46. Nieters, A., et al., *PRRC2A and BCL2L11 gene variants influence risk of non-Hodgkin lymphoma: results from the InterLymph consortium.* Blood, 2012. **120**(23): p. 4645-8.

47. Pellegrini, M., et al., *Shutdown of an acute T cell immune response to viral infection is mediated by the proapoptotic Bcl-2 homology 3-only protein Bim* 2003, PNAS. p. 14175–14180.

48. Baum, L., et al., *Case-control association study of polymorphisms in the voltage-gated sodium channel genes SCN1A, SCN2A, SCN3A, SCN1B, and SCN2B and epilepsy.* Hum Genet, 2014. **133**(5): p. 651-9.

49. Cordelli, D.M., et al., *Fever as a seizure precipitant factor in Panayiotopoulos syndrome: a clinical and genetic study.* Seizure, 2012. **21**(2): p. 141-3.

50. Hong, E.P. and J.W. Park, *Sample size and statistical power calculation in genetic association studies.* Genomics Inform, 2012. **10**(2): p. 117-22.

51. Rowhani-Rahbar, A., et al., *Effect of age on the risk of Fever and seizures following immunization with measles-containing vaccines in children.* JAMA Pediatr, 2013. **167**(12): p. 1111-7.

52. CDC: Possible Side-effects from Vaccines, M.; Available from: http://www.cdc.gov/vaccines/vac-gen/side-effects.htm - mmrv.

53. Klein, N.P., et al., *Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures.* Pediatrics, 2010. **126**(1): p. e1-8.

54. Pruim, R.J., et al., *LocusZoom: regional visualization of genome-wide association scan results.* Bioinformatics, 2010. **26**(18): p. 2336-7.

55. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

56. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.

57. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data.* Nat Genet, 2011. **43**(5): p. 491-8.

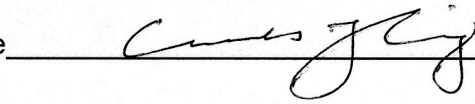
58. Cingolani, P., et al., *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.* Fly (Austin), 2012. **6**(2): p. 80-92.

59. Cingolani, P., et al., *Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift.* Front Genet, 2012. **3**: p. 35.

60. Danecek, P., et al., *The variant call format and VCFtools.* Bioinformatics, 2011. **27**(15): p. 2156-8.

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature  Date 09/16/2014