Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Genes Immun. Author manuscript; available in PMC 2017 January 06.

Published in final edited form as:

Genes Immun. 2016 July ; 17(5): 298–304. doi:10.1038/gene.2016.21.

Identification of genetic variants associated with susceptibility to West Nile virus neuroinvasive disease

Dustin Long[1], **Xutao Deng**[2,3], **Pardeep Singh**[4], **Mark Loeb**[4], **Adam S. Lauring**[5,*], and **M Seielstad**[2,3,*]

[1]School of Medicine, University of California, San Francisco, CA, USA

[2]Department of Laboratory Medicine, University of California, San Francisco, CA, USA

[3]Blood Systems Research Institute, San Francisco, CA, USA

[4]McMaster University, Hamilton, Ontario, Canada

[5]Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

Abstract

West Nile virus (WNV) infection results in a diverse spectrum of outcomes, and host genetics are likely to influence susceptibility to neuroinvasive disease (WNND). We performed whole exome sequencing of 44 individuals with WNND and identified alleles associated with severe disease by variant filtration in cases, kernel association testing in cases and controls, and SNP imputation into a larger cohort of WNND cases and seropositive controls followed by genome-wide association analysis. Variant filtration prioritized genes based on the enrichment of otherwise rare variants, but did not unambiguously implicate variants shared by a majority of cases. Kernel association demonstrated enrichment for risk and protective alleles in the HLA-A and HLA-DQB1 loci, which have well understood roles in antiviral immunity. Two loci, *HERC5* and an intergenic region between *CD83* and *JARID2*, were implicated by multiple imputed SNPs and exceeded genome-wide significance in a discovery cohort (n=862). SNPs at two additional loci, *TFCP2L1* and *CACNA1H*, achieved genome-wide significance after association testing of directly genotyped and imputed SNPs in a discovery cohort (n=862) and a separate replication cohort (n=1387). The context of these loci suggests that immunoregulatory, ion channel, and endothelial barrier functions may be important elements of the host response to WNV.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence: Adam Lauring, MD, PhD, 1150 W. Medical Center Dr. MSRB1 Room 5510B, Ann Arbor, MI 48109-5680, alauring@med.umich.edu.

*These authors contributed equally to this work

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

INTRODUCTION

West Nile Virus (WNV) is an arthropod borne RNA virus that causes West Nile fever and West Nile neuroinvasive disease [1]. Annual epidemics in the United States result in several thousand cases of WNV-associated disease [2]. The spectrum of WNV-associated disease is highly variable. Eighty percent of infections are asymptomatic, approximately 20% experience a nonspecific "flu-like" illness, and <1% progress to neuroinvasive forms of disease [3]. Host factors that influence the clinical spectrum of WNV disease remain poorly defined. Recognized epidemiologic risk factors for WNND include advanced age, diabetes, immunosuppression, alcohol abuse, cancer, and chemotherapy [4–9]. However, a significant number of patients who succumb to extreme forms of infection have no identifiable risk factor.

The existence of rare, extreme outcomes to WNV infection has stimulated interest in host genetic risk factors [10,11]. Candidate gene studies have associated single nucleotide polymorphisms (SNPs) in the 2′–5′ oligoadenylate synthetase (*OAS*) gene family [12,13] and the *CCRdel32* deletion [14,15] with risk for seroconversion and symptomatic infection, respectively. An association study of functional SNPs in genes with immune function identified a single SNP in *OAS1* that was associated ($P < 0.01$) with encephalitis and acute flaccid paralysis, and SNPs in *MX1* ($P < 0.05$ and *IRF3* ($P < 0.01$) that were associated with symptomatic infection [16]. A genome-wide scan of over 13,000 mostly nonsynonymous, coding SNPs in 560 neuroinvasive cases and 950 seropositive controls found tentative associations between WNND and SNPs in three genes, but showed no joint significance when examined in the study's predefined replication cohort [17].

These results are consistent with studies of other infection-related phenotypes, which have generally identified genetic associations of modest effect size, with varying success at replication. This limited success could result in part from a reliance on common and functional SNPs in immune genes that do not adequately capture and represent the full spectrum of rare and common genetic variation. The rarity and severity of WNND may suggest the contribution of relatively few, rare human risk alleles of large effect, as opposed to the combined effects of many common variants with smaller individual effects. Rare variants collectively form the largest class of human genetic sequence diversity and are believed to contribute significantly to health and disease [18–20].

We sought to identify rare, potentially deleterious variants enriched in subjects with WNND using whole exome sequencing, sequence kernel association testing (SKAT), and imputation. Subsequent genotyping and association testing of candidate risk alleles in a larger set of cases and controls identified multiple relatively common variants in *HERC5* and an intergenic region between *CD83* and *JARID2*, while rare variants in *TFCP2L1* and *CACNA1H* were associated with various forms of WNV disease at genome-wide significance thresholds. Our data suggest that genetic susceptibility to WNND is a complex trait and that rare and common variants contribute to the risk of severe outcomes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RESULTS

The overall design of this study is illustrated in Figure 1. From a large cohort of WNND cases and seropositive controls, we identified a subset of 44 young, otherwise healthy individuals with WNV encephalitis and performed exome sequencing. We then generated a list of rare variants from the exome sequence dataset that were enriched in this highly selected population of patients with an extreme outcome to WNV infection. We also used gene-level burden testing as implemented in SKAT-O to implicate genes in which rare variants were enriched in the encephalitis group, relative to individuals of similar ethnicity who were sequenced as part of the 1000 genomes project. Such tests are sensitive to allelic heterogeneity in genes. We then imputed additional genotypes into a larger collection of cases and controls using our own exome sequence data and the 1000 genomes reference haplotypes. These analyses identified a number of candidate risk variants that we included in a custom array to genotype in the original cohort and a second replication cohort. The characteristics of these cohorts are shown in Table 1.

Exome Sequencing

On average, 85% of targeted bases were covered at a read-depth of 8× or greater (Supplementary Figure 1). Average concordance between sequence variant calls and Illumina HumanOmni1-Quad v1.0 SNP calls was 99.4% for overlapping SNPs. The final variant call-set closely matched published standards for the number of expected coding variants and ratios of transitions to transversions, frameshift to non-frameshift InDels, and synonymous to non-synonymous SNPs (Supplementary Tables 1 and 2) [21,22].

In exome datasets, a rare or novel coding variant that is enriched to a significant degree in a cohort of affected, unrelated individuals is a plausible risk allele [23]. To identify rare variants potentially associated with WNND, we set allele frequency cut-offs for significant enrichment under autosomal recessive and autosomal dominant models based on an estimated trait prevalence of 1%, high penetrance, Hardy-Weinberg equilibrium, the existence of multiple risk variants per gene, and the predicted impact of the coding variant on protein function (Supplementary Tables 3 and 4). We used Sanger re-sequencing of selected variants and comparisons with available SNP genotyping to refine variant calling in our next generation sequencing data (for example, Supplementary Table 5). While this approach prioritized genes and identified putatively deleterious variants shared by up to 53% of subjects (Supplementary Table 4), the lack of a formal statistical framework with which to assess the results in a limited number of cases hinders more definitive statements about the involvement of these variants in WNV disease risk. Ultimately, upon replication genotyping in much larger numbers of cases and controls, variants in the *CACNA1H* locus identified in this manner were replicated at a threshold exceeding genome-wide significance (see below).

Kernel Association Testing (SKAT-O)

Even in our cohort of limited ethnic diversity, we considered the possibility that individuals harbored unique risk alleles within a given gene. Like burden testing, kernel association testing with SKAT-O can demonstrate statistical association for an aggregation of rare variants within a gene and is much less sensitive to errors in genomic annotation [24,25]. We

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

used this approach to compare variants identified in our WNND exome sequence dataset to those found in 379 subjects of European ancestry included in the 1000 Genomes Project phase 1 release.

A quantile-quantile (q-q) plot of p values for this analysis (Supplementary Figure 2) may indicate some residual stratification despite close matching of cases and controls (see MDS plot, Supplementary Figure 3). Multiple variants in HLA-A and HLA-DQB1 emerged as highly significant in this analysis ($p<10^{-10}$), and represent highly plausible candidate loci for the viral immune response (Supplementary Table 6). The high levels of polymorphism and the aggregation of rare risk and protective alleles at these loci represent the exact circumstances for which SKAT-O is designed to be maximally sensitive, and this constitutes one of the few successful applications of SKAT-O for any trait. After excluding highly polymorphic genes and signals driven entirely by numerous alleles of small effect in the control data (i.e., *AHNAK2* and *SPEN*), additional genes, including *NTF3, FMN2 and PPYR1*, maintained levels of significance considerably below the threshold for exome-wide significance ($p<2.5\times10^{-6}$, based on Bonferroni correction for 20,000 independent human genes). (Supplementary Table 6)

Imputation and Association Testing

To further explore the variants identified in the case-only exome data, we imputed all novel variants identified by sequencing, together with the latest 1000 Genomes reference haplotypes, into an additional 406 WNND cases and 456 seropositive controls, for which Illumina 1M SNP genotypes were available [17]. Rare alleles are more difficult to impute, and we observed a predictable degradation in performance below minor allele frequencies (MAFs) of 0.05 (Supplementary Table 7 and Supplementary Figure 4) [26–28]. Nevertheless, the inclusion of a custom reference panel allowed us to impute many variants that are rare in the general population, but enriched in WNND cases (Supplementary Figure 5).

A q-q plot of imputed genotypes (Supplementary Figure 6) shows a considerable excess of observed over expected variants associated with WNND in this imputed dataset, without signs of an overall inflation of p-values due to population stratification or other sources of experimental error. A baseline test of association on the directly genotyped SNPs also showed no evidence of stratification (Supplementary Figure 7). Multiple variants above the inflection point ($p<10^{-9}$) are located within the *HERC5* gene on chromosome 4 (lead SNP rs148556308; $P = 6.5\times10^{-10}$), and an intergenic locus on chromosome 6 between *CD83* and *JARID2* (lead SNP 6:14571587; $P = 4.0\times10^{-10}$) that includes a conserved *STAT5a* transcription factor-binding site (Figure 2 and Supplementary Table 8). Two additional single SNPs on chromosome 8 and chromosome 21 showed associations with $P < 5 \times 10^{-9}$.

Validation and replication of candidate variants

We selected the top candidates from each of the above analyses for direct genotyping, in an attempt to validate the associations observed in the imputed cohort (stage 1; n=862) and to replicate these associations in additional cases and controls (stage 2; n=1387). The replication cohort consisted of individuals recruited from the same population who were not included in our exome sequencing and imputation studies, and for whom genome-wide

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

genotyping data were not available (see Supplementary Methods). We used a custom Illumina iSelect BeadChip to genotype variants from 122 candidate risk loci in these two cohorts. Three hundred twenty-five of 373 variants selected for replication were successfully genotyped in 1,563 subjects (Supplementary Table 9). The failure of 13% of attempted SNPs is typical for the Illumina iSelect genotyping technology used (see Illumina Technical Note on Genotyping; San Diego, CA). Association statistics were computed separately for stage 1 and stage 2, and were also combined into a joint analysis of significance. Separate analyses were conducted for: 1) all neuroinvasive cases vs. mild/asymptomatic controls; 2) encephalitis cases vs. asymptomatic controls; and 3) acute flaccid paralysis cases vs. mild/asymptomatic controls.

The minimal p-value observed in the most inclusive and heterogeneous group of all WNND cases was $P = 2.3\times10^{-6}$ for rs11122852, within an intron of the *TFCP2L1* gene on chromosome 2 (Table 2). Two adjacent SNPs were also marginally associated in the joint analysis of all WNND cases (*rs6756142 and rs7563166* $p=9.6\times10^{-5}$-4.5×10^{-5}, OR = 3.6–3.9, Table 2). These SNPs did achieve genome-wide significance ($p=2.4\times10^{-8}$–5.6×10^{-6}, ORs 4.9) in the analysis of AFP-only cases (n=267) to controls (n=954). The use of this case definition noticeably elevated the significance of the *TFCP2L1*, but revealed no new risk loci, and may suggest this association is with AFP as an outcome rather than other neuroinvasive forms of disease.

An additional SNP in the *CACNA1H* gene (*rs78879053* $p=4.4\times10^{-7}$) achieved the pre-defined threshold for joint significance and approached genome-wide significance in the analysis of encephalitis cases (either encephalitis or meningoencephalitis, n=225). This variant was observed at a minor allele frequency of 1.3% in encephalitis cases and was not observed in controls (3).

While the SNPs that we evaluated in *TFCP2L1* and *CACNA1H* are rare in the control groups, the minor allele frequencies closely match those of individuals of CEPH and British descent in the 1000 Genomes Phase 3 release (Table 3, see also Supplementary Figure 3). A power analysis [29] demonstrates that, despite their low population frequencies, we were well powered to identify the reported associations for these SNPs (Table 3). For the SNPs in *CACNA1H* and *TFCP2L1*, there was no excess of minor allele homozygotes in the cases compared to the controls. The signal was driven almost entirely by individuals who were heterozygous for these rare alleles. For example, among WNND cases, there were 23 heterozygotes and 2 homozygotes for *rs6756142 and rs7563166,* and among encephalitis cases, there were 4 heterozygotes and 1 homozygote for *rs78879053.*

SNPs in the *HERC5* and *CD83-JARID2* loci identified as surpassing genome-wide significance in stage 1 (see above) failed in the replication genotyping, which is not uncommon with the Illumina iSelect custom genotyping platform.

Replication of previously described risk alleles

We attempted to replicate associations previously described for WNV disease. These included variants in *OAS1* (*rs34137742* and *rs10774671*), *MX1* (*rs7280422*), and *IRF3* (*rs2304207*) [16]. We tested these associations using similar case definitions and genetic

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

models. In our joint analysis, none of the above SNPs approached statistical significance (Supplementary Table 10), although we were unable to achieve accurate genotyping of *rs3213545* in *OASL*.

DISCUSSION

In this comprehensive study of genetic risk factors for WNND, we used three methods to identify variants associated with this severe outcome of WNV infection. In contrast to previous studies of host genetic risk factors for symptomatic WNV infection [12–17], this study was designed initially to detect rarer, high impact variants in coding regions of the genome. In the primary model of WNND, one locus encompassing *TFCP2L1* surpassed genome-wide thresholds of significance and was validated in the replication sample of cases and controls. Secondary analyses of WNND subtypes identified a risk locus in *CACNA1H* significantly associated with the diagnosis of encephalitis.

The most robust association produced from these analyses localized to an intronic region in *TFCP2L1* (transcription factor CP2-like 1). A variant (*rs17006292*) in *TFCP2L1* has been strongly associated with Behcet's disease among Han Chinese, and a related gene, *TFCP2*, has been shown to play a central regulatory role in responses to therapeutic interferon in patients with multiple sclerosis [30,31]. The associations were also found in the analysis of patients experiencing acute flaccid paralysis, but not encephalitis only.

We also identified a rare, but high impact, variant in *CACNA1H* observed exclusively in patients with WNV encephalitis. Mutations in this gene have been reported as rare causes of epilepsy [32]. Considered in this context, it is possible that the influence of this variant in the setting of WNV infection may primarily be upon symptomatic manifestations of disease, rather than the course and extent of infection itself.

Among the most promising associations, and surpassing thresholds of genome-wide significance, were multiple SNPs located within *HERC5* (Figure 2), an interferon stimulated gene, as well as an intergenic locus between *CD83* and *JARID2* that includes a conserved *STAT5a* transcription factor-binding site. Unfortunately, genotyping assays developed for these variants failed in the replication sample, and these results are currently based on imputation in about half of our study sample.

Unlike past studies, which focused primarily on well-characterized innate immune and interferon-associated genes, we observed no association with genes in these categories. We were unable to replicate previous associations reported for variants in *OAS1*, *MX1*, and *IRF3*[16]. Our findings suggest that non-coding regulatory, immunomodulatory, ion channel, and endothelial barrier functions may play important roles in the pathogenesis of WNND. Because none of the significant risk alleles identified is a functional mutation, they are unlikely to be causative variants themselves.

Only a small fraction of the candidate risk alleles identified by exome sequencing, kernel association testing, and imputation were found to be significantly associated with disease in our follow-up genotyping. This finding may not be surprising, given the limited overlap among loci identified by each of these methods. Some of the loci achieving significance in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

follow-up association testing were identified via imputation. This observation supports the findings of published simulations, which suggest that low-coverage sequencing and large-scale imputation is a highly efficient approach to GWAS of rare variants in complex human diseases [33].

Several factors may have contributed to our low overall validation rate, including a very limited sample size for the initial exome sequencing. First, a large number of platform-specific sequencing, alignment, variant calling, and annotation errors are known to exist in data generated by current technologies [34]. These challenges underscore the value of sequencing control subjects from the same study population under the same experimental conditions as cases. A second factor that may have limited our candidate validation rate relates to the accuracy of imputation for rare variants, although the approach has been validated in numerous previous GWAS studies. We attempted to counter this problem through the additional use of a custom imputation reference panel based on paired exome and array genotypes.

Beyond these technical challenges, our study has several limitations. First, we considered only the effects of autosomal sequence variation and did not evaluate copy-number variants, structural variants, or epigenetic modifications. Second, because multiple analytic methods, data types, and case-control definitions were used, determining the most appropriate significance thresholds for our final tests of association was more difficult than it would have been using a standard study design.

Our findings are similar to those reported from similar studies using NGS techniques to study the effects of rare variants on complex traits [35]. While we were able to identify rare risk alleles of moderate to large effect, we conclude that susceptibility to neuroinvasive WNV infection remains a highly complex trait. Noncoding and population-specific variants are likely to contribute significantly to the host-genetics of this disease. However, despite their limited ability to predict WNND risk in the majority of cases in this cohort, the associations identified may provide novel insights into the pathogenesis of severe WNV infection.

SUBJECTS AND METHODS

Description of cohorts

This study received institutional review board approval from McMaster University, McGill University, the University of California San Francisco, and the University of Michigan. Subjects were drawn from a previously described cohort [17]. Cases were defined as individuals meeting criteria for WNV infection as well as clinical criteria for meningitis, acute flaccid paralysis (AFP), or encephalitis (see Supplementary Methods). Controls were defined as individuals meeting the same criteria for infection, but not those for meningitis, AFP, or encephalitis. Informed consent was obtained from all patients or their surrogates.

Sequencing, variant calling, and annotation

We selected 44 subjects of European descent (42% female) with encephalitis who were young (mean 39, range 19–45) and otherwise healthy. Exon capture was performed with the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Agilent SureSelect Target Enrichment System (Santa Clara, CA), and sequencing was performed on the Illumina (San Diego, CA) platform using standard manufacturer protocols. Reads were aligned to human genome build 37 with BWA [36] and processed with SAMtools [37] and the Genome Analysis Toolkit [38]. Variant calling was performed using the UnifiedGenotyper and the VCFs were annotated with snpEff and the ENSEMBL annotation database [39]. Filtering models for SNPs and InDels were independently trained using the Variant Quality Score Recalibration module. We set passing thresholds at values corresponding to filtration of less than 0.5% of known, high-confidence SNPs and InDels.

Case-Only Variant Filtering

We parsed variant call-sets into a database populated with genomic data from several public sources (e.g 1000 Genomes Project [21], NHLBI [40], and UW NIEHS [41] genotypes, dbSNP data). We joined variants to each table by chromosome, position, reference and alternate alleles, and ran a series of queries (see Supplementary Methods) to filter based on biologic effect, sequence context, and estimates of allele frequency. We collapsed variants passing all filters into the genes within which they occurred, and prioritized based on dominant or recessive models accounting for the presence of compound heterozygotes.

Kernel association testing (SKAT-O)

We prepared a panel of case genotypes from the exome sequencing calls of all 44 WNND subjects and a panel of *in silico* control genotypes from 379 subjects of European ancestry included in the 1000 Genomes phase 1 release. We limited this analysis to SNPs in consensus coding sequences (common to both our exon capture kit and those used in the 1000 Genomes Project). Standard quality control procedures were independently applied to cases and controls. Kernel association testing was completed with SKAT-O [24] using the unified optimal test with small sample size adjustment, using RefSeq exons to define kernels and four multidimensional scaling (MDS) dimensions calculated in PLINK as PCA covariates [42].

Imputation and association testing

We prepared a custom imputation reference panel by merging exome and array data for the 36 sequenced cases that had previously been genotyped with the Illumina 1M Bead Chip (available through the Immport databse, https://immport/niaid.nih.gov). Because our custom reference panel was small and composed of cases only, we leveraged data from the 1000 Genomes Integrated Phase I release and IMPUTE2's dual reference panel option to improve phasing and to provide an adequate panel of control haplotypes [26]. Alleles from both reference panels were simultaneously imputed for 406 WNND cases and 456 controls from the same population that had been previously genotyped on an Illumina 1M SNP array [17]. We applied standard quality control filters and used PLINK identity-by-state analysis to exclude duplicates and cryptically related subjects. We performed principle component analysis with EIGENSOFT to characterize population stratification [43]. Case-control association testing was completed using the SNPTEST frequentist method with the first ten eigenvectors as covariates to adjust for any residual stratification [44].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Validation and replication of candidate variants

Candidate polymorphisms were systematically prioritized for inclusion on a custom Illumina GoldenGate array using an automated selection algorithm (see Supplementary Methods). We additionally included tag-SNPs for high-ranking variants identified with Haploview [45]. In total, 373 candidate and tag-SNPs were selected. These SNPs were genotyped in the original cohort of 406 cases and 456 controls as well as a pre-specified replication cohort of 513 cases and 874 controls. The quality control filters applied to this custom array were more stringent than those for generic, manufacturer-designed chips (see Supplementary Methods). Association testing was performed in PLINK.

Statistical analysis of association

The candidate variants selected for final association testing were drawn from multiple high-density data sources, and identified using multiple methods. We therefore adopted distinct significance thresholds for tests in different subject groups and for follow-up analyses. For tests of validation for associations observed in imputed genotypes in cases and controls, we selected the nominal threshold of $p<2.5\times10^{-6}$ for exome-wide significance [46]. This threshold is based on Bonferroni correction for testing at 20,000 independent genes. Although some non-coding variants from the 1M chip genotypes used for imputation and from nonexonic sequencing reads were included, exome data formed the backbone of the dataset.

For tests of replication in the case and control groups not used for imputation and hypothesis generation, we calculated a Bonferroni-corrected equivalent of $\alpha=0.05$ for 324 tests to arrive at a value of $p<1.5\times10^{-4}$. This is conservative, as the majority of the 324 probes were tag SNPs included for redundancy in the characterization of 122 distinct loci. A joint significance threshold of $p < 3.7\times10^{-5}$ was approximated using a sample size-weighted logarithmic mean of the two values above.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Paul Renauer for technical assistance and Amr Sawalha for a critical reading of the manuscript. This work was supported by a grant from Novartis and institutional funds from the Blood Systems Research Institute. DL was supported by the National Center for Research Resources, the National Center for Advancing Translational Sciences, and the Office of the Director, National Institutes of Health, through UCSF-CTSI Grant Numbers TL1 RR024129 and TL1 TR000144. ASL was supported by National Institutes of Health K08 AI081754 and a Clinician Scientist Development Award from the Doris Duke Charitable Foundation. ML was supported in part by the NIH Population Genetics Analysis Program.

References

1. Petersen LR, Brault AC, Nasci RS. West Nile virus: review of the literature. JAMA. 2013; 310:308–315. [PubMed: 23860989]
2. [accessed 3 Jun, 2015] West Nile. cdc.gov. http://www.cdc.gov/westnile/

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3. Mostashari F, Bunning ML, Kitsutani PT, Singer DA, Nash D, Cooper MJ, et al. Epidemic West Nile encephalitis, New York, 1999: results of a household-based seroepidemiological survey. Lancet. 2001; 358:261–264. [PubMed: 11498211]

4. Chowers MY, Lang R, Nassar F, Ben-David D, Giladi M, Rubinshtein E, et al. Clinical characteristics of the West Nile fever outbreak, Israel, 2000. Emerging Infect Dis. 2001; 7:675–678. [PubMed: 11585531]

5. Patnaik JL, Harmon H, Vogt RL. Follow-up of 2003 human West Nile virus infections, Denver, Colorado. Emerging Infect Dis. 2006; 12:1129–1131. [PubMed: 16836833]

6. Bode AV, Sejvar JJ, Pape WJ, Campbell GL, Marfin AA. West Nile virus disease: a descriptive study of 228 patients hospitalized in a 4-county region of Colorado in 2003. Clin Infect Dis. 2006; 42:1234–1240. [PubMed: 16586381]

7. Weiss D, Carr D, Kellachan J, Tan C, Phillips M, Bresnitz E, et al. Clinical findings of West Nile virus infection in hospitalized patients, New York and New Jersey, 2000. Emerging Infect Dis. 2001; 7:654–658. [PubMed: 11589170]

8. Nash D, Mostashari F, Fine A, Miller J, O'Leary D, Murray K, et al. The outbreak of West Nile virus infection in the New York City area in 1999. N Engl J Med. 2001; 344:1807–1814. [PubMed: 11407341]

9. Jean CM, Honarmand S, Louie JK, Glaser CA. Risk factors for West Nile virus neuroinvasive disease, California, 2005. Emerging Infect Dis. 2007; 13:1918–1920. [PubMed: 18258047]

10. Samuel CE. Host genetic variability and West Nile virus susceptibility. Proc Natl Acad Sci USA. 2002; 99:11555–11557. [PubMed: 12192094]

11. Suthar MS, Diamond MS, Gale M. West Nile virus infection and immunity. Nat Rev Micro. 2013; 11:115–128.

12. Lim JK, Lisco A, McDermott DH, Huynh L, Ward JM, Johnson B, et al. Genetic variation in OAS1 is a risk factor for initial infection with West Nile virus in man. PLoS Pathog. 2009; 5:e1000321. [PubMed: 19247438]

13. Yakub I, Lillibridge KM, Moran A, Gonzalez OY, Belmont J, Gibbs RA, et al. Single nucleotide polymorphisms in genes for 2′-5′-oligoadenylate synthetase and RNase L inpatients hospitalized with West Nile virus infection. J INFECT DIS. 2005; 192:1741–1748. [PubMed: 16235172]

14. Lim JK, McDermott DH, Lisco A, Foster GA, Krysztof D, Follmann D, et al. CCR5 deficiency is a risk factor for early clinical manifestations of West Nile virus infection but not for viral transmission. J INFECT DIS. 2010; 201:178–185. [PubMed: 20025530]

15. Glass WG, McDermott DH, Lim JK, Lekhong S, Yu SF, Frank WA, et al. CCR5 deficiency increases risk of symptomatic West Nile virus infection. J Exp Med. 2006; 203:35–40. [PubMed: 16418398]

16. Bigham AW, Buckingham KJ, Husain S, Emond MJ, Bofferding KM, Gildersleeve H, et al. Host genetic risk factors for West Nile virus infection and disease progression. PLoS ONE. 2011; 6:e24745. [PubMed: 21935451]

17. Loeb M, Eskandarian S, Rupp M, Fishman N, Gasink L, Patterson J, et al. Genetic variants and susceptibility to neurological complications following West Nile virus infection. J INFECT DIS. 2011; 204:1031–1037. [PubMed: 21881118]

18. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012; 337:64–69. [PubMed: 22604720]

19. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science. 2012; 337:100–104. [PubMed: 22604722]

20. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–753. [PubMed: 19812666]

21. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]

22. Pelak K, Shianna KV, Ge D, Maia JM, Zhu M, Smith JP, et al. The characterization of twenty sequenced human genomes. PLoS Genet. 2010; 6:e1001111. [PubMed: 20838461]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

23. Biesecker LG. Exome sequencing makes medical genomics a reality. Nat Genet. 2010; 42:13–14. [PubMed: 20037612]
24. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012; 91:224–237. [PubMed: 22863193]
25. Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat Genet. 2012; 44:886–889. [PubMed: 22772370]
26. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]
27. Li L, Li Y, Browning SR, Browning BL, Slater AJ, Kong X, et al. Performance of genotype imputation for rare variants identified in exons and flanking regions of genes. PLoS ONE. 2011; 6:e24945. [PubMed: 21949800]
28. Wang Z, Jacobs KB, Yeager M, Hutchinson A, Sampson J, Chatterjee N, et al. Improved imputation of common and uncommon SNPs with a new reference set. Nat Genet. 2012; 44:6–7.
29. Menashe I, Rosenberg PS, Chen BE. PGA: power calculator for case-control genetic association analyses. BMC Genet. 2008; 9:36. [PubMed: 18477402]
30. Hecker M, Goertsches RH, Fatum C, Koczan D, Thiesen H-J, Guthke R, et al. Network analysis of transcriptional regulation in response to intramuscular interferon-β-1a multiple sclerosis treatment. Pharmacogenomics J. 2012; 12:360–360. [PubMed: 22825046]
31. Hou S, Yang Z, Du L, Jiang Z, Shu Q, Chen Y, et al. Identification of a susceptibility locus in STAT4 for Behçet's disease in Han Chinese in a genome-wide association study. Arthritis Rheum. 2012; 64:4104–4113. [PubMed: 23001997]
32. Cain SM, Snutch TP. T-type calcium channels in burst-firing, network synchrony, and epilepsy. Biochim Biophys Acta. 2013; 1828:1572–1578. [PubMed: 22885138]
33. Pasaniuc B, Rohland N, McLaren PJ, Garimella K, Zaitlen N, Li H, et al. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. Nat Genet. 2012; 44:631–635. [PubMed: 22610117]
34. Lam HYK, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, et al. Performance comparison of whole-genome sequencing platforms. Nature Biotechnology. 2012; 30:78–82.
35. Heinzen EL, Depondt C, Cavalleri GL, Ruzzo EK, Walley NM, Need AC, et al. Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. Am J Hum Genet. 2012; 91:293–302. [PubMed: 22863189]
36. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. 2009; 25:1754–1760.
37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
38. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–498. [PubMed: 21478889]
39. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, 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; 6:80–92. [PubMed: 22728672]
40. [accessed 30 May, 2015] Exome Variant Server. evs.gs.washington.edu. http://evs.gs.washington.edu/EVS/
41. [accessed 30 May, 2015] NIEHS EGP - Exome Project. evs.gs.washington.edu. http://evs.gs.washington.edu/niehsExome/
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

43. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]
44. Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]
45. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. [PubMed: 15297300]
46. Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, et al. Exome sequencing and the genetic basis of complex traits. Nat Genet. 2012; 44:623–630. [PubMed: 22641211]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

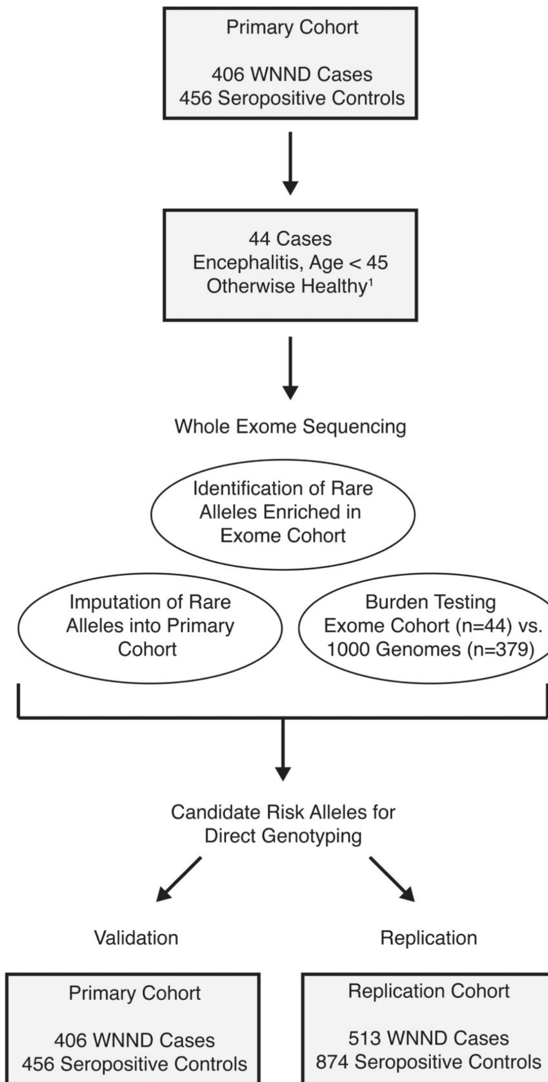

Figure 1.

Study Overview. Forty-four young, otherwise healthy individuals with WNV encephalitis were selected for exome sequencing. The past medical history in the entire cohort included only single reported instances of pyelonephritis, food poisoning, minor orthopedic surgery, appendectomy, and tonsillectomy. Candidate risk alleles were identified from the exome sequence data, burden testing, and genome-wide association testing after imputation into a larger collection of genotyped case and control individuals. These candidates were genotyped in the primary cohort (validation) and a replication cohort.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

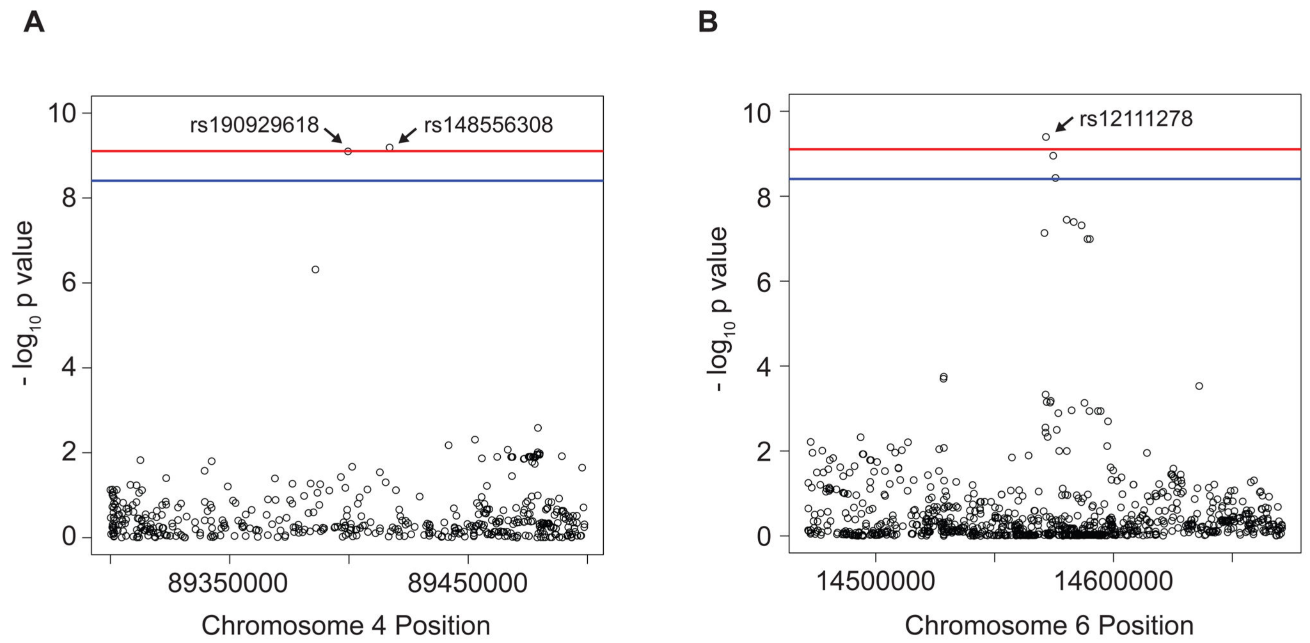

Figure 2.

Zoomed Manhattan plots of imputed variants associated with WNND in HERC5 (A) and an intergenic locus between *CD83* and *JARID2* (B). Red and blue lines indicate Bonferroni corrections for p-values of 0.01 and 0.05, respectively.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Phenotypic characteristics of cohorts

	Sequenced Cohort	Imputed Cohort Cases	Imputed Cohort Controls	Replication Cohort Cases	Replication Cohort Controls
Number	44	406	456	513	874
N (%) Female	23 (52.3)	187 (46.1)	244 (53.5)	251 (48.9)	533 (61.0)
Age at enrollment mean, [SD]	39.5 [5.4]	58.0 [14.8]	52.7 [12.8]	61.8 [14.6]	53.7 [13.3]
European-American	44	405	456	456	861
Hispanic or Latino				37	10
Asian				3	2
African American				12	
Other/Unknown		1		5	1
Acute flaccid paralysis	-	140	-	247	-
Encephalitis	8	55	-	30	-
Meningitis	-	55	-	121	-
Meningoencephalitis	36	156	-	115	-
Control	-	-	456	-	874

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Genetic loci associated with WNND

			Primary Cohort				Replication Cohort				Combined Cohort			
Locus	**rsID**	**Chr:Position**	**Case MAF**	**Control MAF**	***P***	**OR**	**Case MAF**	**Control MAF**	***P***	**OR**	**Case MAF**	**Control MAF**	***P***	**OR**
WNND vs. Control														
TFCP2L1	rs11122852	2:122029914	0.0125	0.0000	0.0061	NA	0.0462	0.0138	8.04E-06	3.47	0.0328	0.0094	2.25E-06	3.57
TFCP2L1	rs6756142	2:122033495	0.0062	0.0000	0.0526	NA	0.0326	0.0084	5.28E-05	3.97	0.0222	0.0058	4.48E-05	3.91
TFCP2L1	rs7563166	2:122031368	0.0062	0.0000	0.0526	NA	0.0326	0.0092	1.13E-04	3.64	0.0222	0.0063	9.60E-05	3.58
Encephalitis vs. Control														
CACNA1H	rs78879053	16:1236758	0.0000	0.0000	NA	NA	0.0278	0.0000	1.59E-09	NA	0.0133	0.0000	4.41E-07	NA
CACNA1H	rs113802594	16:1220480	0.0043	0.0012	0.1085	NA	0.0232	0.0023	8.65E-05	10.29	0.0133	0.0016	2.73E-04	8.58
AFP vs. Control														
TFCP2L1	rs11122852	2:122029914	0.0319	0.0000	1E-05	NA	0.0520	0.0138	1.48E-05	3.93	0.0449	0.0094	2.42E-08	4.94
TFCP2L1	rs6756142	2:122033495	0.0160	0.0000	0.0019	NA	0.0289	0.0084	2.50E-03	3.50	0.0243	0.0058	1.20E-04	4.30
TFCP2L1	rs7563166	2:122031368	0.0160	0.0000	0.0019	NA	0.0376	0.0092	1.20E-04	4.21	0.0300	0.0063	5.57E-06	4.88

NA - Odds ratios could not be calculated as allele not detected in controls

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Population allele frequency and statistical power

Phenotype	Locus	SNP	OR	Case MAF	Cases (n)	Control MAF	Controls (n)	Population MAF[a]	Power (%)[c]
WNND	TFCP2L1	rs11122852	3.57	0.0328	609	0.0094	954	0.0079	99
		rs7563166	3.58	0.0222	609	0.0063	954	0.0079	96
		rs6756142	3.91	0.0222	609	0.0058	954	0.0079	97
Encephalitis	CACNA1H	rs78879053	37.88	0.0133	225	0.0004	954	0.003[b]	96
	CACNA1H	rs113802594	5.40	0.0133	225	0.0025	954	0.005[b]	84
	OFCC1	rs72653717	NA	0.0067	225	0.0000	954	0[b]	NA
AFP	TFCP2L1	rs11122852	4.94	0.0449	267	0.0094	954	0.0079	100
		rs7563166	4.88	0.0300	267	0.0063	954	0.0079	99
		rs6756142	4.30	0.0243	267	0.0058	954	0.0079	95

[a] CEU and GBI populations, 1000 Genomes Phase 3 release

[b] Shown are MAF for European populations as MAF was 0 in CEU and GBI

[c] Based on alpha of 0.05, case prevalence of 0.00714 (WNND in 1/140 of exposed), the respective odds ratios, risk allele frequencies, and sample population numbers for each variant.