Whole genome sequencing analysis of a family with familial dysautonomia and neuropsychiatric symptoms

Han Fang[1,2], Yiyang Wu[1,2], Jason A. O'Rawe[1,2], David Mittelman[4,5], Gholson J. Lyon[1,2,3*]

[1]Stanley Institute for Cognitive Genomics, One Bungtown Road, Cold Spring Harbor Laboratory, NY, USA; [2]Stony Brook University, 100 Nicolls Rd, Stony Brook, NY, USA; [3]Utah Foundation for Biomedical Research, E 3300 S, Salt Lake City, Salt Lake City, UT, USA; [4]Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, USA, [5]Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA.

Background

We report here our discovery and whole genome sequencing (WGS) analysis of a family with dominantly inherited familial dysautonomia (Figure 1). The mother is affected with dysautonomia, hereditary hemochromatosis, and obsessive compulsive traits. The oldest daughter reports severe dysautonomic syncopal episodes, gastroparesis, glaucoma, visual and auditory hallucinations, urinary retention, and one prior stroke. One son is affected with dysautonomia, Tourette syndrome (TS), attention deficit disorder (ADD), and obsessive-compulsive disorder (OCD). Another son reports dysautonomia, asthma, seizure in response to pertussis, dyslexia, migraine, dysgraphia, ADD, sensory integration disorder, and arthritis.

Methods

WGS was performed on two affected siblings in the Illumina CLIA-certified lab. Three other DNA samples from the family also underwent WGS in the research setting at CSHL. DNA samples from the same five people were genotyped using the Illumina HumanOmni2.5-8 BeadChip. Genome-wide parametric linkage analysis was performed using Merlin (adjusted for linkage disequilibrium). Our group previously reported the low concordance rates across different variant calling pipelines. Thus, to reduce algorithm-induced biases, we used multiple pipelines for quality control, alignment, assembly, variant calling, genotype refinement, variant filtering, and variant annotation (Figure 2). The Empowered Genome Cohort data, implemented in the Ingenuity variant analysis system, was also used to filter out WGS-specific sequencing errors, which might have been identified as rare variants.

Results

Linkage analysis showed not enough power for identifying any disease relevant variant loci in a single family. None of the family members carry any previously reported variants in *IKBKAP* that have been implicated in the autosomal recessive transmission of familial dysautonomia. The WGS data had good sequence coverage for *IKBKAP* in all five people (mean coverage ~40x) but we did not identify any novel rare variants in this gene. ERDS and PennCNV did not reveal any CNVs that of relevance to the phenotype of familial dysautonomia. We have generated a list of rare variants of unknown significance (Table 1 & 2). Sanger sequencing results confirmed an insertion in *POU4F1* (Figure 4). We find relatively low concordance across three variant detection pipelines (Figure 5). Pharmacogenomic analyses reveal the recommended dosages of Coumadin and Simvastatin (Table 3). Five variants in three people with prior clinical evidence are also shown (Table 4).

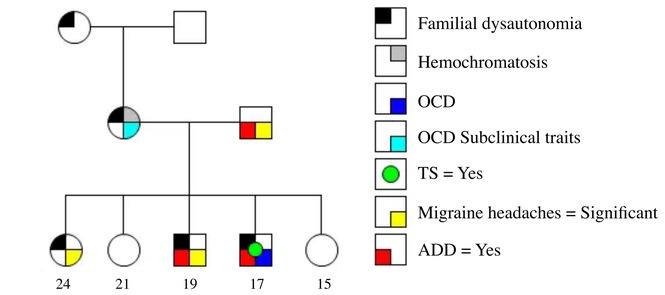

Figure 1. The pedigree in this study. The numbers below the children are their corresponding ages. Five people in the family have thus far been included in our WGS study: the parents, the 24-year-old daughter, the 17-year-old son, and the 15-year-old daughter.

Table 1. A subset list of rare variants (MAF < 1%) found in at least two affected people. These variants are called by at least two pipelines, within coding regions, with a high VAAST score, and a CADD c-score greater than 20.

Gene name	Genomic coordinates	Amino Acid change	Zygosity Effect	Quality Coverage	Population frequency
EPS15	chr1: 51829571	Gly>Arg	Heterozygous non-synon	197 47:24:23	Unknown
MYL7	chr7: 44179406	Gly>Ser	Heterozygous non-synon	71 23:16:7	0.69%
ERI1	chr8: 8865561	Pro>Ser	Heterozygous non-synon	156 49:30:19	Unknown
CCIN	chr9: 36170317	Arg>Gln	Heterozygous non-synon	223 36:14:22	0.14%
PCTP	chr17: 53844742	Cys>Tyr	Heterozygous non-synon	223 47:22:25	0.98%

Table 2. The insertion in *POU4F1* is found in two children with familial dysautonomia. There is no coverage in the mother's WGS data. Follow-up Sanger sequencing confirmed the existence of the homozygous insertion in all three people : affected mother, affected daughter, and affected son. *POU4F1* encodes a member of the POU-IV class of neural transcription factors. This protein is expressed in a subset of retinal ganglion cells and may be involved in the developing sensory nervous system.

Gene	Chr	Position	Region	Transcript Variant	Protein Variant	Inferred Activity
POU4F1	13	79176323	Exonic	c.484_486dipGGC	p.162_163_insG	Loss

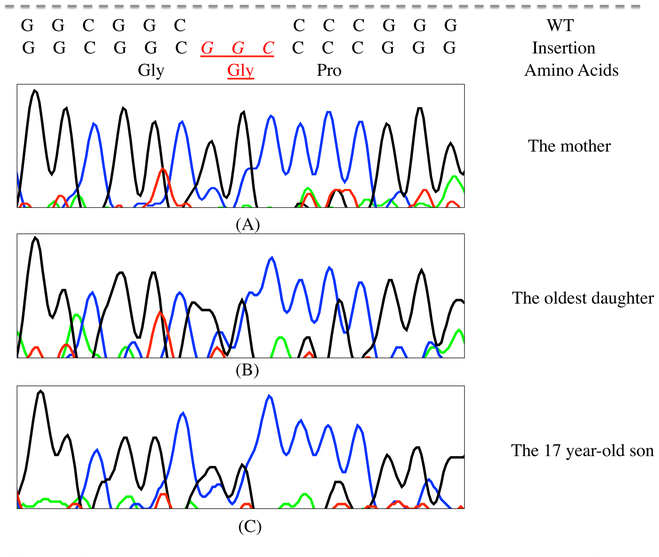

Figure 4. Sanger sequencing results confirmed the homozygous insertion in exon 2 of *POU4F1*. Sanger sequencing results of (A) the mother, (B) the oldest daughter, and (C) the 17-year-old son.

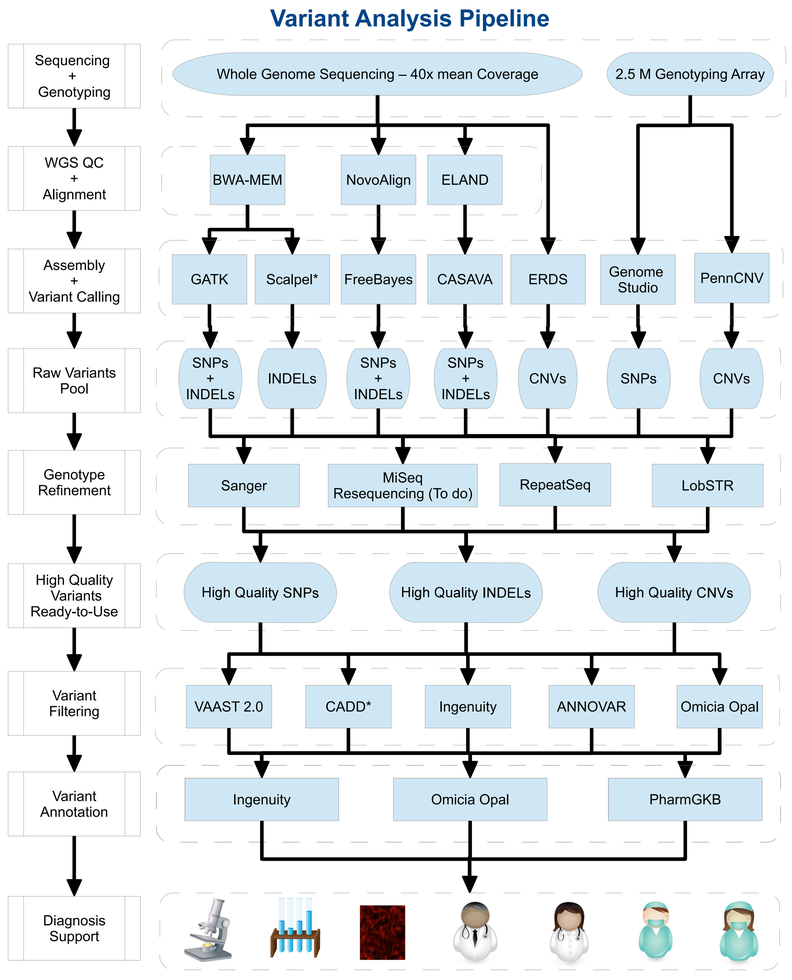

Figure 2. The left-hand side is the major analysis work flow while the right-hand side are the details of each procedure.* CADD (Combined Annotation Dependent Depletion, in press) : http://cadd.gs.washington.edu/
* Scalpel (in preparation) : Haplotype Micro-assembly for accurate INDEL detection: http://schatzlab.cshl.edu/

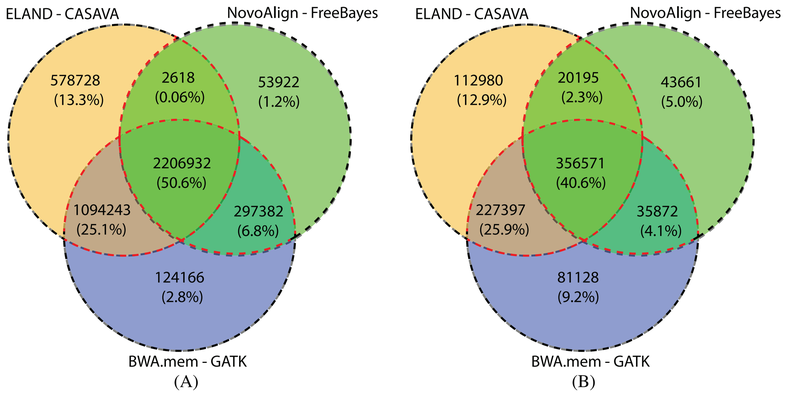

Figure 5. Venn diagrams of the variant calling results across three independent alignment and assembly combinations. 2x100 paired end reads for the WGS data were generated on Illumina HiSeq2000 platform. Reads were mapped to the hg19 reference genome. BWA-0.7.5a, GATK v.2.6-5, ELANDv2e, and CASAVA 1.8 were used here. (A) The Venn diagram is for SNP detection comparison. We noticed an even lower concordance rate compared to what we previously reported (57.4% across 5 pipelines). This is because we here use three different alignment steps. (B) The Venn diagram is for INDEL detection, the concordance rate is even lower: 40.6%. We focus on variants called by at least two pipelines, i.e. the central region surrounded by the red dash lines. This allows us to analyze 82.6% of the SNPs and 72.9% of the INDELs in our downstream analyses.

Table 3. Recommended dosages for Coumadin and Simvastatin dosages based on the oldest daughter's WGS results, in comparison to what she was actually prescribed in the absence of any genetic testing. Pharmacogenomics analyses were performed based on guidelines and algorithms from the International Warfarin Pharmacogenomics Consortium (IWPC) and the Clinical Pharmacogenomics Implementation Consortium (CPIC) in the PharmGKB database. People who are homozygous for major alleles at both sites in *CYP2C9* are designated as *1/*1.

Drug	Recommend dosages based on genotypes	Previous prescriptions	FDA recommendations	Genotypes
Coumadin (Warfarin Sodium)	5.85 mg/day	5 mg/day	2 to 10 mg daily Consider genetic testing results	*VKORC1*: A/G (rs9923231) *CYP2C9*: *1/*1 (rs1799853, rs1057910)
Simvastatin	20 mg/day Increased risk of myopathy with 40 mg Simvastatin	20 mg/day	80 mg/day	*SLCO1B1*: T/C (rs4149056)

Table 4. At least seven mutations with substantial clinical relevance were found in our study. The "Daughter" means the oldest daughter, and the "Son" means the 17-year-old son. "hom" = homozygous and "het" = heterozygous. "WT" = wild type. The mother is homozygous for the C282Y variant in *HFE*, and, consistent with the lack of symptoms for hemochromatosis in the children, we found that all three children are carriers for the C282Y variant, but not carrying any other reported hemochromatosis related variants.

Gene name	Genomic coordinates	Amino Acid change	Zygosity	Variation type	Population frequency	Clinical significance
HFE	chr6: 26093141	Cys>Tyr	Mother (hom) Daughter (het) Son (het)	non-synon	G:98% A:2%	ClinVar: Condition: Hereditary hemochromatosis
ELAC2	chr17: 12899902	Ala>Thr	Mother (WT) Daughter (het) Son (het)	non-synon	C:98% T:2%	ClinVar: Condition: Prostate cancer, hereditary
BRIP1	chr17: 59937223	Pro>Ala	Mother (het) Daughter (het) Son (het)	non-synon	G:97% C:3%	ClinVar: Condition: Breast cancer, early-onset
SLCO1B1	chr12: 21331549	Val>Ala	Mother (WT) Daughter (het) Son (WT)	non-synon	T:88% C:12%	PharmGKB: Strongly associated with an increased risk of statin-induced myopathy
ACADS	chr12: 121176083	Gly>Ser	Mother (het) Daughter (het) Son (het)	non-synon	G:82% A:18%	OMIM: Deficiency of butyryl-CoA dehydrogenase
IL23R	chr1: 67705958	Ard>Gln	Mother (WT) Daughter (het) Son (WT)	non-synon	G:97% A:3%	OMIM: Inflammatory Bowel Disease 17, Protection Against Psoriasis, Protection Against, Included
F5	chr1: 169519049	His>Tyr	Mother (het) Daughter (WT) Son (WT)	non-synon	C:98% T:2%	PharmGKB: Associated with risk of venous thromboembolism (VTE).

Conclusions

Here we show one example of individualized medicine in a family with neuropsychiatric symptoms and at least two rare diseases. WGS data provides much more clinically relevant and potentially actionable information when compared to genotyping arrays. Despite limited overall agreement between CLIA certified and non-CLIA certified analysis pipelines, utilizing variants identified by two or more pipelines enables WGS to act as a useful technique for clinical diagnostics. For rare diseases, we argue that family history and other family members' sequencing data should be incorporated into downstream analyses so as to eliminate false positive findings of disease-associated variants. We highlight the importance of detailed phenotyping and sharing of both genomic and phenotyping data because of extreme heterogeneity across families and insufficient knowledge of rare diseases. Ongoing effort will focus on identifying and proving familial dysautonomia relevant variants, extending the pharmacogenomic analyses to the other people in the pedigree, and including more family members in our study.

Questions? Email: GholsonJLyon@gmail.com