Whole genome analysis of an extended pedigree with Prader–Willi Syndrome, Hereditary Hemochromatosis, Familial Dysautonomia, Tourette Syndrome and other illnesses

Han Fang[1,2], Yiyang Wu[1,2], Jason A. O'Rawe[1,2], Laura T. Jimenez Barrón[1,3], Gareth Highnam[4], David Mittelman[4], Gholson J. Lyon[1,2]

[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] Centro de Ciencias Genomicas, Universidad Nacional Autonoma de Mexico, Cuernavaca, Morelos, MX. [4] Gene by Gene, Ltd., Houston, TX, USA

Background

We report here our analyses and discovery of an extended pedigree with Prader–Willi Syndrome (PWS), Hereditary Hemochromatosis (HH), Familial Dysautonomia (FD), Tourette Syndrome and other illnesses. Since genetic architectures between these diseases are heterogeneous, we chose to perform whole genome sequencing (WGS) on nine people of this pedigree, enabling a wide scope of variant calling from a single SNP to large structural events. To reduce false positive/negative variant calls, we used more than one pipeline to detect SNPs, INDELs, large structural variations, and copy number variations.

Results

First, we used ERDS to identify two deletions with size of 450Kb and 4.8Mbp, spanning the regions of 15q11.2, 15q12, 15q13.1 in the boy with PWS. These deletions were confirmed with Illumina 2.5M array data using PennCNV. Second, the mother with Hereditary Hemochromatosis is homozygous for the C282Y variant in *HFE*, with the variant being called by HaplotypeCaller and FreeBayes. These two variants were previously reported in the literature, suggesting they are likely large effect-size variants contributing to this phenotype. Third, none of the family members with FD carry any previously reported variants in *IKBKAP* that have been implicated in the autosomal recessive transmission of FD. The WGS data had good sequence coverage (> average coverage 40x) for this gene but we did not identify any novel rare variants.

It is likely that FD is dominantly inherited in the family, which has not been reported in any detailed manner. Thus, this is possibly a novel type of FD that might be relevant to variants in other genes. To investigate, we leveraged the power of the large pedigree and WGS with the use of pVAAST and CADD. Fourth, pharmacogenomic analyses were performed using TuteGenomics and PharmGKB platforms. We found pharmacogenetic variants influencing the metabolism of Coumadin and Simvastatin, which were being routinely prescribed to the daughter.

We highlight the importance of detailed phenotyping and sharing of both genomic and phenotyping data, due to extreme heterogeneity of illnesses across families and insufficient knowledge of the genetic architecture of most diseases. Ongoing effort will focus on identifying variants that might be relevant to FD and other illnesses segregating in the pedigree.

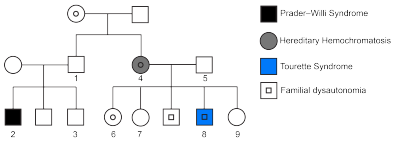

Figure 1. **A three-generation pedigree with multiple rare diseases in this study**. The pedigree members with a number underneath indicates that WGS has been performed on this person.

Table 1. **A list of variants (MAF < 1% in ESP and 1000G) found in the three people with Familial Dysautonomia**. These variants are heterozygous, called by at least one pipelines, within coding regions, ranked high by pVAAST, and a CADD c-score greater than 20. The alternative allele frequency (AF) is computed based on the European population in the ExAC database.

Gene	Genomic coordinates	Change	Effect	CADD	Zygosity & AAF
PLA2G4E	chr15: 42281727	C>T	missense	36.0	Het, 0.007%
SLC35C2	chr20: 44987076	A>G	missense	27.5	Het, 0.006%
HHATL	chr3: 42740594	C>T	missense	23.8	Het, 0.04%
GRSF1	chr4: 71701950	C>T	missense	23.0	Het, 0.1%
GABRA6	chr5: 161128514	C>T	missense	18.0	Het, 0.003%
ALG3	chr3: 183963549	T>C	missense	18.3	Het, 0.1%
NISCH	chr3: 52505834	A>T	missense	17.8	Het, 0.06%
RBM27	chr5: 145634638	T>C	missense	18.7	Het, 0.05%
MYO1H	chr12: 109862622	A>G	missense	19.5	Het, 0.002%

Table 2. **A list of variants with previous evidence in ClinVar were found in the pedigree members**. The mother with Hereditary Hemochromatosis is homozygous for the C282Y variant in *HFE*. The carriers are represented by the numbers shown in Figure 1. AR-autosomal recessive, CH- compound heterozygous

Gene	Genomic coordinates	Change & Effect	Zygosity & Carriers	AAF	Supporting Evidence
HFE	chr6: 26093141	G>A missense	Hom: 1, 4 Het: 2, 3, 6, 7, 8, 9	0.007%	Hemochromatosis (AR)
BRIP1	chr17: 59937223	G>C missense	Het: 1, 4, 6, 7, 8	0.04%	Breast cancer, early-onset (AR)
MKKS	chr20: 10393439	G>T missense	Het: 4, 9	0.9%	Mckusick-kaufman Syndrome (AR)
PRSS1	chr7: 142458451 chr7: 142458526	A>T missense A>G missense	Compound Het: 1, 2, 4, 6, 7, 8, 9	47% 3%	Hereditary pancreatitis (CH)

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/day (Consider genetic testing results)	*VKORC1*: A/G (rs9923231) *CYP2C9*: *1/*1 (rs1799853, rs1057910)
Simvastatin	20 mg/day Increased risk of myopathy with 40mg Simvastatin	20 mg/day	80 mg/day	*SLCO1B1*: T/C (rs4149056)

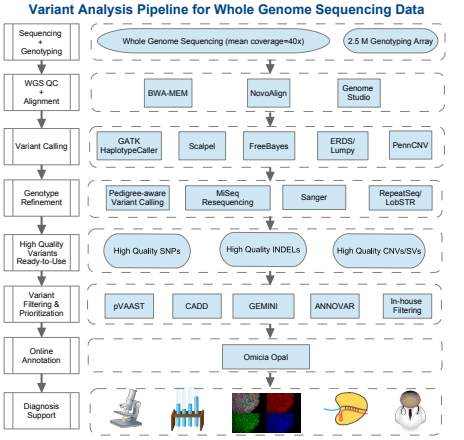

Figure 2. **Lyon lab variant analysis pipeline for whole genome sequencing data**. The left-hand side is the major analysis work flow while the right-hand side are the details of each procedure.

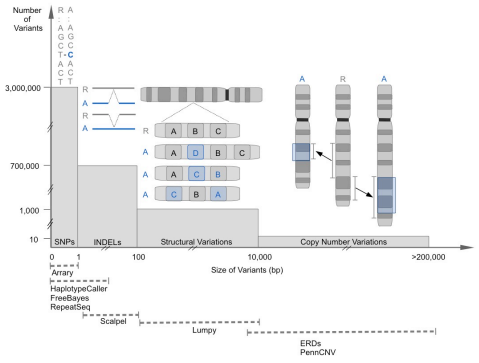

Figure 3. **Interrogate human genome from a single-codon resolution to large structural events with WGS**. Due to the fact that variations are of different sizes in the human genome, to accurately detect variants, one would need to leverage of the advantages of various bioinformatics algorithms.

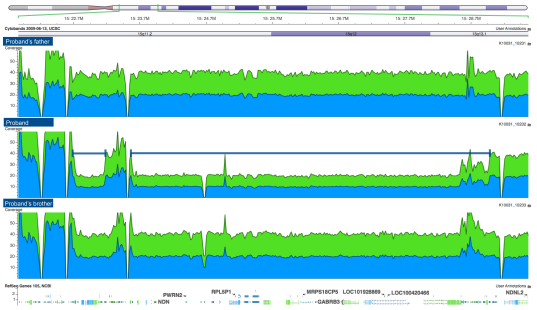

Figure 4. **Heterozygous deletion expanding the region 15q11.2 to 15q13 in the proband (breakpoints: chr15: 22,749,401- 28,566,000)**. This deletion is not detected from the proband's father and brother. The allele of the proband was inherited from the mother.

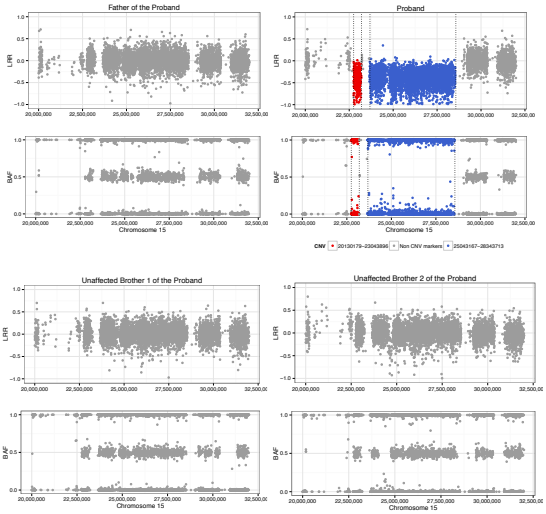

Figure 5. **Validation of the copy number variant using Illumina 2.5m microarray data**. We used PennCNV to call this deletion from the microarray data., which is also only detected from the proband, but not from the father and the two unaffected brothers. The dash lines in the figure of proband indicates the interval of the ERDS copy number variant call.