Whole Genome Sequencing Analysis of an Idiopathic Intellectual Disability Syndrome.

Gholson J. Lyon[1,2,3], Jason O'Rawe[1,2], Yiyang Wu[1,2], Alan Rope[4], Reid J. Robison[3,5], Kai Wang[3,6], Jeffrey Swensen[4]

1) Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, NY, USA; 2) Stony Brook University, Stony Brook, NY, USA; 3) Utah Foundation for Biomedical Research, Salt Lake City, UT, USA; 4) Department of Pediatrics, University of Utah, Salt Lake City, UT, USA; 5) Intermountain Healthcare, Salt Lake City, UT, USA; 6) University of Southern California, Los Angeles, CA, USA.

Abstract

We present a new idiopathic intellectual disability syndrome accompanied with distinctive facial dysmorphology. X-chromosome inactivation assays reveal skewing in the mother, suggesting the possibility of an X-linked disorder. High-density genotyping arrays were performed on both children revealing no known causal or pathogenic SNVs and no known CNVs that might contribute to the phenotype. Whole genome sequencing was performed with Complete Genomics and subsequent sequence analysis led to the identification of several rare variants that may or may not contribute to the syndrome.

Presentation of the Phenotype

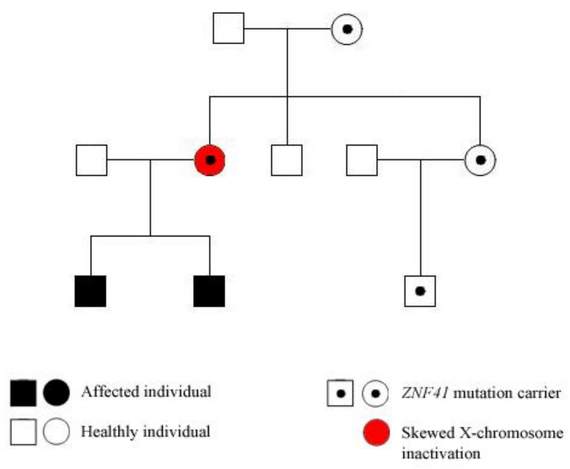
The propositi are two affected male brothers (Fig. 1), aged 10 and 12 respectively, with severe intellectual disability, autism-like behavior, attention deficit issues, and very distinctive facial features (Fig. 1), including broad, upturned nose, sagging cheeks, downward sloping palpebral fissures, relative hypertelorism, high-arched palate, and prominent ears. Their parents are nonconsanguineous and are both healthy, and the family history does not demonstrate any members with anything resembling this current syndrome.

Both boys were diagnosed with IUGR (intra-uterine growth retardation) very early and delivered through C-section. The younger boy had mild heart murmurs at birth but echocardiography confirmed the absence of any cardiovascular abnormalities. Like many infants, he was treated with light for neonatal jaundice.

Additional phenotypic symptoms shared by both brothers include microcephaly, ventriculomegaly, congenital hydrocephalus, cerebral atrophy, oculomotor dysfunction, frequent otitis media, hearing loss, sacral caudal abnormalities, dysplastic toenails, growth retardation, hypotonia and global developmental delays, especially in the areas of gross motor and verbal expression. The younger brother also suffers from frequent episodes of contact dermatitis and eczema as well as mild asthma, but no longer requiring medication. The elder brother, on the other hand, has partial agenesis of the corpus callosum and cutis aplasia congenita, which healed at 4 days old.

Conclusions

There are many challenges in showing how any one mutation can contribute toward a clear phenotype, particularly in the context of genetic background and possible environmental influences. We highlight the importance of comprehensive, high quality raw and processed NGS data sets, so as to reduce the false negative rate. Recent work demonstrates the existence of complex genic dependencies where some variants are deemed necessary, but not entirely sufficient for causing disease. Accordingly, we are in the process of sequencing the whole genome of the unaffected male cousin along with a large portion of the extended family to more comprehensively (and accurately) explore these polygenic disease models. Indeed, we have demonstrated that the ZNF41 mutation shows potential for being a disease contributing mutation as it has been identified by more than a single bioinformatics analysis as being relevant and has also been previously identified as mutated in other families with intellectual disability; however, the presence of the mutation in an unaffected cousin demonstrates that it is not altogether sufficient.



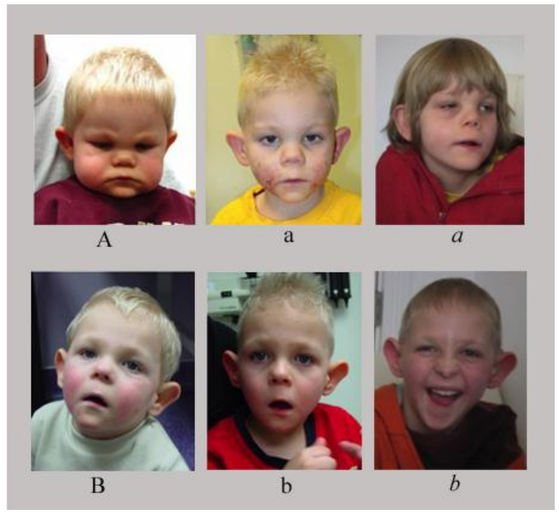


Fig. 1 Facial phenotype of the younger brother at age 19 months (A), 3.5 years (a) and 7 years (a) ; and the elder brother at age 3 years (B), 5 years (b) and 9 years (b) , with broad, upturned nose, sagging cheeks, downward sloping palpebral fissures, relative hypertelorism, high-arched palate, and prominent ears. A pedigree displaying intra familial relationships as well as the segregation pattern of a particularly interesting mutation in the ZNF41 gene is shown.

Germline mosaic

VAAST	SVS	Omicia Opal
PION c.10C>T	**PION** c.10C>T	**PION** c.10C>T
		OPLAH c.3499_3500del

Autosomal recessive

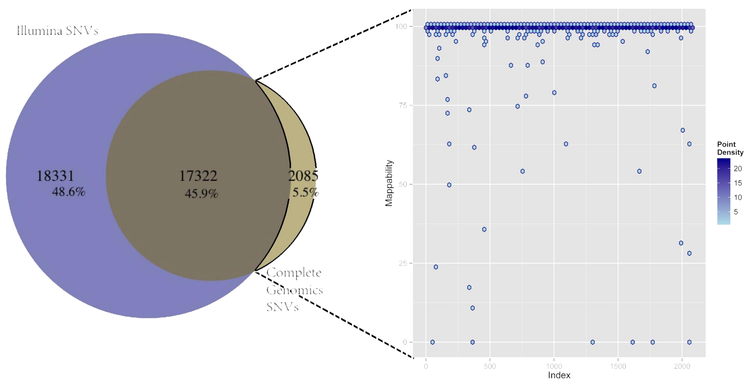
VAAST	SVS	Omicia Opal
TMEM92 c.269G>A	-	**TMEM92** c.269G>A

X-linked

VAAST	SVS
ZNF41 c.1191C>A	**ZNF41** c.1191C>A
ASB12 c.739G>T	**ASB12** c.739G>T
TAF1 c.4010T>C	**TAF1** c.4010T>C

Complete Genomics WGS

Whole genomes of the mother, father and both affected boys were sequenced and analyzed with the Complete Genomics WGS sequencing and bioinformatics pipeline. Although we find that the Complete Genomics (CG) WGS pipeline is known to produce sets of variants with high sensitivity, we find that the pipeline also potentially misses a relatively large number of true exonic variants that might indeed be functional. However, we also show that CG WGS identifies some variants missed by exon capture and sequencing with Agilent/Illumina, which is further supported by high mappability scores of these variants, indicating that they should have been found by the Agilent/Illumina platform but were likely missed due to poor capture.



Illumina WGS

Whole genomes of the entire pedigree are currently being sequenced using the Illumina HiSeq2000 platform and will be analyzed by, among others, the BWA-GATK alignment and variant discovery pipeline. Unlike the Complete Genomics WSG pipeline, we expect the Illumina WGS pipeline to reveal a number of variant sets with varying degrees of sensitivity and specificity depending on the particular bioinformatics chosen to analyze the raw data, and we expect the sensitivity and specificity to also depend on the number of pipelines used to analyze the data.

	Specificity		Sensitivity		Known SNPs			Novel SNPs		
	Mean*	SD	Mean*	SD	#Total	#cSNP	Ti/Tv	#Total	#cSNP	Ti/Tv
SOAPsnp	99.82	0.039	94.53	2.287	30,022	17,409	2.77	875	419	1.94
GATK	99.72	0.085	95.33	1.161	29,620	17,306	2.8	365	206	2.34
SNVer	99.78	0.044	92.32	4.339	28,242	17,111	2.85	490	253	2.52
GNUMAP	99.64	0.065	86.67	3.286	24,893	15,144	3.03	1,091	659	1.28
SAMTools	99.59	0.158	94.45	4.221	29,577	17,449	2.78	949	539	1.33
ANY pipeline	99.62	0.113	97.72	1.215	33,947	19,638	2.68	2,163	1,182	1.23
>=2 pipelines	99.69	0.074	96.68	2.298	31,099	18,108	2.77	639	323	2.17
>=3 pipelines	99.73	0.045	95.65	3.143	29,363	17,257	2.84	416	230	2.56
>=4 pipelines	99.82	0.041	92.63	3.412	26,772	16,097	2.91	318	193	2.67
5 pipelines	99.87	0.015	80.61	5.266	21,174	13,320	3.12	234	149	2.83

Questions? Email: GholsonJLyon@gmail.com