Frequency and Diagnostic Yield of Mosaic Variation Identified by Whole Exome Sequencing

Cecelia R. Miller, Erik Zmuda, Don J. Corsmeier, Ashita Dave-Wala, Michael Fisher, Vincent Magrini, Shalini C. Reshmi, Julie M. Gastier-Foster, Catherine E. Cottrell

Institute for Genomic Medicine at Nationwide Children's Hospital

background

Mosaicism occurs when a single fertilized egg develops with two or more genetically distinct cell populations.

Massively parallel sequencing (MPS), including whole exome sequencing (WES) and targeted panel analysis, has expanded our understanding of the contribution of somatic mosaicism in genetic disorders by enhancing our ability to identify disease-associated alterations occurring at low variant allele frequencies (VAF).

Among unselected clinical exome cohort studies, mosaic, pathogenic alterations have been reported at a frequency of ~1% (Yang, *et al*. 2013 *NEJM*; Retterer, *et al*. 2016 *Genet. Med.*). Additionally, it is known that certain genes associated with phenotypes such as epilepsy, autism, and overgrowth syndromes, demonstrate increased association with somatic mosaicism (Lim, *et al*. 2017 *Nat. Neurosci.*; Krupp, *et al.* 2017 *Am. J. Hum. Genet*.; Siegel, *et al*. 2018 *J. Investig. Dermatol.*).

With the aim to determine the frequency of mosaic alterations within our WES patient population, we reviewed 248 clinical WES cases submitted to our institution to establish mosaicism frequency in variants assessed as likely causal for the patient phenotype.

methods

DNA derived from peripheral blood was subject to target capture using Agilent SureSelect Human all Exon V6 followed by sequencing to 130x mean depth using Illumina instrumentation.

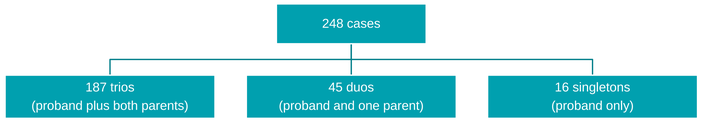
Sequencing data were demultiplexed and secondary analysis of the FASTQ performed by GenomeNext v1.1, which does the alignment, deduplication, and single sample variant calling with GATK Unified Genotyper 1.6-13.

GeneInsight (Sunquest) was used for annotation and tertiary analysis filtering based on clinician-provided phenotypes. The Integrative Genomics Viewer (IGV) v2.3 was used for sequencing visualization.

Variants were assessed according to ACMG/AMP consensus recommendations (Richards, *et al.* 2015 *Genet Med*.).

Mosaic variants were confirmed by Sanger sequencing or targeted deep sequencing of the region of interest.

results



71 cases (28.6%) had pathogenic or likely pathogenic variant(s) which likely contributed to the patient's phenotype.

4 of the 71 (5.6%) causative variants were confirmed to be mosaic.

3 mosaic variants were identified in parental samples due to the presence of a heterozygous variant in the child.

Table 1: Summary of mosaic variants assessed as likely causal for the patient phenotype

Case	Gene	Genomic change (hg19)	Nucleotide change	Predicted protein change	Variant allele frequency	Variant reads/total read depth	Relevant disease association
1	*ARIDIA*	chr1:27092947G>A	c.2879-1G>A	N/A	19%	18 of 96	(AD) Coffin-Siris syndrome 2
2	*ARX*	chrX:25025232C>T	c.1444G>A	p.Gly482Ser	12%	11 of 92	(XLR) Early infantile epileptic encephalopathy
3	*TRIP12*	chr2:230679862G>A	c.1684C>T	p.Arg562Ter	12%	21 of 173	(AD) Mental retardation 49
4	*PIK3R2*	chr19:18273784G>A	c.1117G>A	p.Gly373Arg	18%	18 of 101	(AD) Megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome 1

Abbreviations: AD, autosomal dominant; XLR, X-linked recessive

Case 1

- 6 year old female with global developmental delay, dysmorphic features, agenesis of the corpus callosum, hypotonia, club foot, and exotropia.
- WES identified a pathogenic mosaic *ARID1A* variant: it was *de novo* (PS2), a canonical splice alteration in a gene where loss of function alterations are a known mechanism of disease (PVS1), and absent from controls (PM2).

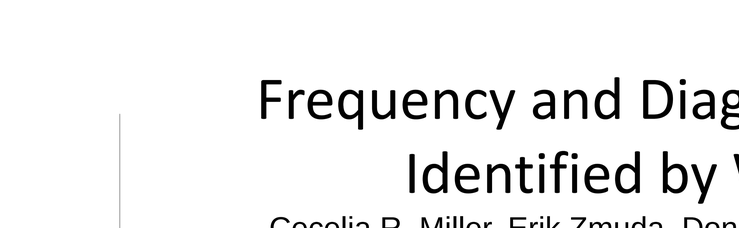


Figure 1: IGV visualization and Sanger sequencing confirmation of mosaic *ARID1A* variant

Case 2

- 6 month old male with infantile spasms, failure to thrive, severe hypotonia, seizures, microcephaly, esotropia, short palpebral fissures, protuberant tongue, and sparse scalp hair.
- WES identified a likely pathogenic mosaic *ARX* variant: it was *de novo* (PS2), absent from controls (PM2), predicted deleterious computationally (PP3) and occurred in a gene with low benign missense variation and where missense variation is associated with disease (PP2).

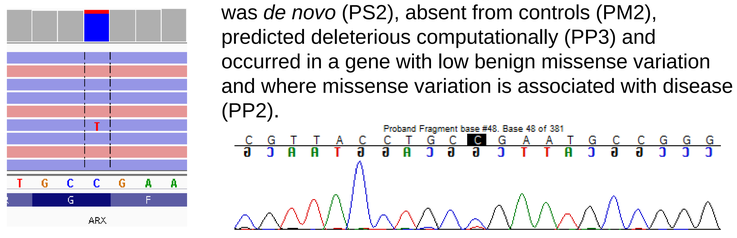


Figure 2: IGV visualization and Sanger sequencing confirmation of mosaic *ARX* variant

Table 2: Mosaic parental variants ascertained due to presence of a heterozygous variant in the child

Case	Gene	Genomic change (hg19)	Nucleotide change	Predicted protein change	Variant allele frequency	Variant reads/total read depth	Relevant disease association	Variant interpretation in heterozygous proband
5	*RYR1*	chr19:38948884 G>A	c.2119G>A	p.Gly707Ser	8%	45 of 596	(AD) Malignant hyperthermia susceptibility 1	Uncertain significance (PM1, PP3, PP5)
6	*ARID1A*	chr1:27100899 G>A	c.4181G>A	p.Ser1394Asn	6%	7 of 118	(AD) Coffin-Siris syndrome 2	Uncertain significance (PM2)
7	*ARID2*	chr12:46245833_ 46245834delAG	c.3927_3928 delAG	p.Gly1310Glufs Ter5	4%	2 of 45	(AD) Coffin-Siris syndrome 6	Likely pathogenic (PVS1, PM2)

Abbreviation: AD, autosomal dominant

Case 3

- 1 year old male with global developmental delay, seizures, chorea, hypotonia, short stature, ptosis, frontal bossing, micrognathia.
- WES identified a pathogenic mosaic *TRIP12* variant: it was *de novo* (PS2), a nonsense variant (PVS1) and absent from controls (PM2).

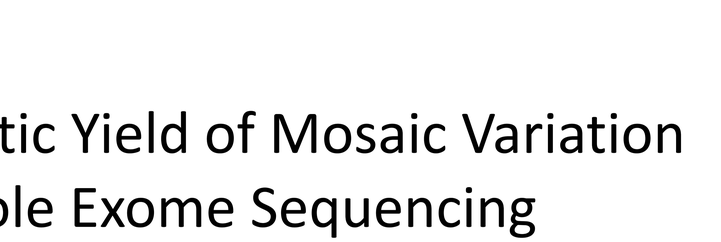


Figure 3: IGV visualization and Sanger sequencing confirmation of mosaic *TRIP12* variant

Case 4

- 2 year old female with macrocephaly, hirsutism, global developmental delay, and perisylvian polymicrogyria on MRI.
- WES identified a pathogenic mosaic *PIK3R2* variant: it was *de novo* (PS2), absent from controls (PM2), previous reported cosegregating with disease in multiple family members (PP1, Mirzaa, *et al*. 2015 *Lancet Neurol.*), demonstrated altered function *in vitro* (PS3, Riviere, *et al.* 2012 *Nature Genetics*), predicted deleterious computationally (PP3), and a recurrent hotspot alteration (PM1) with an increased prevalence in affected individuals compared to controls (PM_PS4).

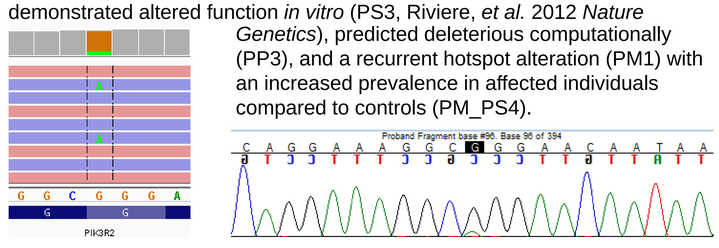


Figure 4: IGV visualization and Sanger sequencing confirmation of mosaic *PIK3R2* variant

discussion

Given adequate read depth, and appropriate bioinformatic processing parameters, MPS can effectively detect mosaic alterations at low VAFs.

The enhanced capability of MPS to detect mosaic alterations has improved our ability to identify disease-causing variants and establish recurrence risk.

Several challenges remain in the detection of mosaic variants by MPS including false positive sequencing/alignment artifacts, false negatives due to limitations in sensitivity in variant-calling algorithms, limitations in read depths, and the need to analyze disease-involved tissue.