Analysis of mosaicism for sequence and copy number variants in a diverse set of hereditary disorders in a large clinical cohort

SWAROOP ARADHYA, PHD, FACMG
Disclosures & Acknowledgements

- Dr. Aradhya is medical director and full-time employee of Invitae, a clinical genetic information company in San Francisco, California.

Contributors to this study

- Rebecca Truty, PhD (Main contributor, bioinformatics analysis)
- Curtis Kautzer
- Michael Kennemer
- Jennifer Rhees
- Amanda Stafford
- Robert L. Nussbaum, MD, FACMG
- Swaroop Aradhya, PhD, FACMG (Principal investigator)
Mosaicism in Mendelian disease

Early Postzygotic Mutations

Late Postzygotic Mutations

Science. 2013;341:1237758
Mosaicism in Mendelian disease

Somatic Mosaicism

Gonosomal Mosaicism

Germline Mosaicism

Mosaicism in Mendelian disease

- Rare variation arising post-zygotically within Mendelian disease genes can lead to mosaicism and contribute to the pathogenesis of hereditary disorders.

- Because we routinely use high depth of coverage sequencing, we had the opportunity to study a clinical cohort of nearly half a million people to understand the prevalence of mosaicism in hereditary disease.

- Validation of mosaicism detection by NGS has to consider a variety of parameters and can be done through “genome-mixing” experiments.
Validation of NGS to detect mosaicism

- Establishing the sensitivity and specificity of mosaicism detection by NGS requires attention to both the chemistry and the bioinformatics pipeline.

- We performed “genome-mixing” to investigate mosaicism detection of >7000 variants by NGS.

<table>
<thead>
<tr>
<th>Percent of mixture comprised of Genome A</th>
<th>80</th>
<th>70</th>
<th>60</th>
<th>50</th>
<th>40</th>
<th>30</th>
<th>20</th>
<th>10</th>
<th>5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected allele balance for het calls in Genome A</td>
<td>40</td>
<td>35</td>
<td>30</td>
<td>25</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>2.5</td>
<td>1</td>
</tr>
</tbody>
</table>

- In-silico down-sampling to simulate lower depth of coverage.

- For heterozygous variants expected to be at 0.5 allele balance in an un-mixed sample, we determined the following when they were in mixed samples:
  - Observed allele balance in mixed samples
  - Observed depth of coverage
  - Absence of the variant
Sensitivity to low allele balance

Sensitivity = \frac{\text{variants observed in titration}}{\text{variants in unmixed genome}}

- Depth 350x average, >1500 genes
- Computed for each
  - Titration level
  - Observed depth of coverage
  - Variant type / length
  - Genomic contexts
- High sensitivity was observed for variants present at
  - > 10% allele balance
  - > 200x
Distinguishing mosaic and non-mosaic variants

Observed allele balance distribution of heterozygous variants in unmixed genomes
Distinguishing mosaic and non-mosaic variants

Observed allele balance distribution of variants in 20% allele balance titration

Observed allele balance distribution of heterozygous variants in unmixed genomes
Distinguishing mosaic and non-mosaic variants

Observed allele balance distribution of variants in 40% allele balance titration

Observed allele balance distribution of heterozygous variants in unmixed genomes

Possible to detect mosaic variants and distinguish them from non-mosaic variants with high positive predictive value and sensitivity
Mosaic intragenic copy number variants

- Single-exon, multi-exon, and whole-gene mosaic deletions were each reliably detected at ~30% or higher allele balance
- Detection of mosaic duplications showed no correlation to allele balance or depth of coverage

<table>
<thead>
<tr>
<th></th>
<th>Observed allele balance (from SNVs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Confidently called as copy number = 2</td>
</tr>
<tr>
<td>Whole gene dup NIPA1</td>
<td>9.3% 10.2% 10.3% 13.1% 17.2% 22.0% 28.6%</td>
</tr>
<tr>
<td>Partial gene dup AARS</td>
<td>14.5% 19.8% 20.4% 25.1% 30.2% 34.5% 39.1%</td>
</tr>
<tr>
<td>Whole gene del NPHP1</td>
<td>13.2% 18.7% 19.1% 24.1% 28.9% 33.8% 38.7%</td>
</tr>
<tr>
<td>Partial gene del ALG1</td>
<td>13.2% 18.7% 19.1% 24.1% 28.9% 33.8% 38.7%</td>
</tr>
<tr>
<td>Single exon del CTNNA3</td>
<td>14.5% 19.8% 20.4% 25.1% 30.2% 34.5% 39.1%</td>
</tr>
</tbody>
</table>
Outcomes from the validation study

- Sensitivity can drop at both high and low allele balance
- Sensitivity drops as depth of coverage decreases
- The appropriate threshold for calling mosaic and non-mosaic sequence variants needs to balance FP and FN mosaic calls
- Our bioinformatics approach provides high sensitivity for mosaic sequence variants present at ≥10% allele balance
- For CNVs, the sensitivity is limited to ~30% allele balance as lower limit for deletions and higher for duplications
Mosaic variants in a clinical cohort

- Investigated genetic testing results in 472,991 individuals
  - 1606 different genes represented
  - Equivalent to 20 million single-gene tests

- 2459 mosaic variants found
  - 2107 SNVs
  - 282 small indels
  - 70 CNVs (intragenic deletions/duplications)

- Observed allele balance 7% - 40%
  - Up to 90% allele balance in X-linked genes in males
Mosaic variants in a clinical cohort

- 286 genes. \(~1\%\) of positive test results
- 70\% in genes associated with autosomal dominant disorders
  - 93\% if AD/AR included
- **40\% classified as LP/P**
- 41 LP/P in X-linked genes
  - 19 in females
  - 22 in males
- Most mosaic variants in cancer genes
Mosaic variants in a clinical cohort

- Genes with >10 LP/P mosaic variants or prevalence of >10% are shown
- Gene counts = [genes with LP/P mosaic variants / all genes]
Mosaic variants in a clinical cohort

- Nearly **twice as many cancer-related genes** with mosaic variants relative to those associated with other types of disorders
- Patients tested for **neurological and cardiovascular disorders and pediatric rare disease** had only half the rate of mosaic variants compared with patients with cancer syndromes
Mosaic variants in a clinical cohort

- Mosaic variants were typically not present in a second tissue when it was available to test.
- Among cases in which both parents were tested, 35/36 mosaic variants were found to be de novo.
- Among individuals with a mosaic variant, 4% of their children carried the variant.

Result from testing a second specimen

- Negative (29)
- Non-mosaic (2)
- Mosaic (8)
Mosaic variants in a clinical cohort

- Individuals with mosaic variants in hereditary cancer genes were considerably older than those with non-mosaic variants ($p < 0.001$ Student’s t-test)

- Mosaic variants may translate to a later age of disease onset

- Caveat: impact of age-related clonal hematopoiesis
Mosaic variation and phenotypic correlation

- Do mosaic variants cause milder phenotypes?

- Cohort: Individuals suspected to have diseases that have highly specific diagnostic criteria (e.g., associated with \(NF1\), \(TSC1\), \(NIPBL\) genes)

- Findings: Individuals with a non-mosaic variant were more likely to meet diagnostic criteria than were individuals with a mosaic variant (Fisher’s exact test \(p<0.001\))

*Detailed clinical data was not available for several individuals and therefore mosaic variants in those individuals were not included in this study - better clinical data can improve variant interpretation in these cases
Limitations and future directions

- Limitations of this analysis
  - Technical artifacts due to difficult to sequence / map regions of the genome
  - Statistical fluctuations in read depth at a given allele

- Additional biological evidence is needed to elucidate the mechanism for low allele balance variants:
  - Hematologic malignancy or clonal hematopoiesis
  - Complex structural variation
  - Maternal engraftment
  - Mosaicism due to reversion of a germline mutation
Conclusions

- **Constitutional mosaic variants can be detected** by next-generation sequencing and distinguished from non-mosaic variants **with high sensitivity and high positive predictive value**
  - The appropriate allele balance thresholds to maximize positive and negative predictive values should be adjusted for prior probability of mosaicism in the gene in question

- **Mosaic variants contribute to ~1% of positive test results** in a cohort of nearly half a million people with a diverse set of hereditary disorders

- **Occurrence of mosaicism can correlate with clinical observations**
  - Mosaic variants virtually always de novo, as expected
  - Individuals with mosaic variants in hereditary cancer genes were considerably older than those with constitutional variants
  - Those with mosaic variants are less likely to show characteristic phenotypic features
  - Mosaic variants were found to be transmitted to offspring in 4% of cases
Thanks for your attention!
Balancing false positive and false negative calls

When there is a high prior for mosaics:
- 3σ threshold reduces mosaic FN
- High NPV

When there is a low prior for mosaics:
- 6σ threshold reduces mosaic FP
- High PPV
Choosing a threshold to **minimize** FP and FN mosaics

Possible to detect mosaic variants and distinguish them from non-mosaic variants with **high positive predictive value and sensitivity**