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

Rebecca Truty, PhD
Invitae
Financial Disclosure

Employee and stockholder of Invitae
Background

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

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

- We used a “genome-mixing” approach to investigate mosaicism detection 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’s 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 downsampling to simulate lower depth of coverage.

- For heterozygous variants expected to be at 0.5 allele balance in an un-mixed sample, determine the following:
  - Observed allele balance in mixed samples
  - Observed depth of coverage
  - Absence of the variant
Low allele balance sensitivity

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

- Computed for each
  - Titration level
  - Observed depth of coverage bin
  - Variant type / length
  - Genomic contexts

- High sensitivity
  - > 10% allele balance
  - > 200x

SNVs
“Easy” genomic regions

Coverage depth
- \geq 1000
- 900 – 1000
- 800 – 900
- 700 – 800
- 600 – 700
- 500 – 600
- 400 – 500
- 300 – 400
- 200 – 300
- 100 – 200
- 0 – 100

Nominal allele balance

Sensitivity vs Allele balance
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

Threshold

Observed allele balance distribution of heterozygous variants in unmixed genomes
Choosing a threshold to minimize FP and FN mosaics

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

Threshold

Observed allele balance distribution of heterozygous variants in unmixed genomes
Choosing a threshold to minimize FP and FN mosaics

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

Threshold

FP

FN

0.4
Choosing a threshold to minimize FP and FN mosaics

Positive Predictive Value (PPV) vs Allele

PPV depends on the relative abundance of mosaic and non-mosaic variants in the tested population.

We can detect mosaic variants and distinguish them from non-mosaic variants with high PPV and sensitivity.
Validation conclusions

- Sensitivity can be reduced at both high and low allele balance
- Sensitivity is reduced at lower depth of coverage
- The appropriate threshold for calling mosaic and non-mosaic variants needs to balance FP and FN mosaic calls
- Our bioinformatics approach provides high sensitivity for mosaic variants present at ~10% - 30% allele balance
- For CNVs, the sensitivity is limited to ~ 20% - 30% allele balance
Clinical data indicating mosaicism at Invitae

- Investigated genetic testing results in 472,991 individuals
  - 1606 different genes represented
  - Equivalent to 20 million single-gene tests
- 2503 mosaic variants found
  - 2108 SNVs
  - 283 small indels
  - 112 CNVs (del/dup)
- Observed allele balance 7% - 40%
  - Up to 90% allele balance in X-linked genes in males
Estimating the prevalence of mosaicism in hereditary disorders

- Mosaic variants found in 288 genes associated with disorders across clinical areas
- Mosaic variants contribute to ~1% of positive test results
- 70% of mosaic variants were in genes associated with autosomal dominant disorders, 93% if AD/AR included
- 40% of mosaic variants were classified as LP/P
- 41 LP/P mosaics found in X-linked genes, 23 in pediatric cases, 19 in females, 22 in males
Prevalence of mosaicism by gene and clinical area

- Only genes with >10 LP/P mosaic variants or a mosaic variant prevalence of >10% are shown.
- Gene counts are for (genes with LP/P mosaic variants) / (all genes in the clinical area).
Additional evidence of mosaicism

- Mosaic variants were typically not present in a second tissue when it was available to test.
Additional evidence of mosaicism

- 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
Mosaicism enriched in older individuals with cancer

Individuals with mosaic variants in hereditary cancer genes were considerably older than those with non-mosaic variants.
Mosaicism and symptom severity

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

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)

* In many cases, detailed clinical data is not available and thus the variant is not eligible - better clinical data can improve variant interpretation in these cases   See Poster 757!!
Limitations and Open Questions

● Limitations to our 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:
  ○ Mosaicism arising early in embryogenesis
  ○ Hematologic malignancy or clonal hematopoiesis
  ○ Complex structural variation
  ○ Maternal engraftment
  ○ Mosaicism due to reversion of a germline mutation
Conclusions

- We can detect mosaic variants and distinguish them from non-mosaic variants with high sensitivity and high positive predictive value (PPV)
  - 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 clinical hereditary disorders

- Occurrence of mosaicism can correlate with certain clinical observations
  - Individuals with mosaic variants in hereditary cancer genes were considerably older than those with constitutional variants
  - Mosaic variants were found to be transmitted to offspring 4% of the time
Acknowledgements

Curtis Kautzer, Michael Kennemer, Jennifer Rhees, Amanda Stafford, Robert L. Nussbaum, Swaroop Aradhya
Backup
Validation study

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

- We used a "genome-mixing" approach to investigate mosaicism detection 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's 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>
Small variant validation
Distinguishing mosaic and non-mosaic variants

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

When there is a low prior for mosaics:
- $6\sigma$ threshold reduces mosaic FP
- High PPV
Small variant validation

Distinguishing mosaic and non-mosaic variants

Genes with high prevalence of mosaicism (e.g. GATA1, ZIC2, TP53, WDR45, PITX3, ACTB)
→ Use 3\(\sigma\) threshold
→ Moderate PPV for high prior
→ High NPV

Genes with low prevalence of mosaicism
→ Use 6\(\sigma\) threshold
→ High PPV for all priors
→ Moderate sensitivity

Positive Predictive Value (PPV) vs Allele balance

Sensitivity vs Allele balance
CNV validation

Titration experiment using clinical samples with CNVs

Partial gene del ALG1

~25% allele balance - confident copy number = 2

~30% allele balance - low quality calls

~35% allele balance - confident copy number = 1
## CNV validation

<table>
<thead>
<tr>
<th></th>
<th>Not confidently called requires manual inspection</th>
<th>Confidently called as integer ploidy CNV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Confidently called as</strong></td>
<td><strong>copy number = 2</strong></td>
<td><strong>Confidently called as</strong></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></td>
</tr>
</tbody>
</table>

Confidently called as copy number = 2

Not confidently called requires manual inspection may be called as mosaic

Confidently called as integer ploidy CNV

Observed allele balance (from SNVs)
CNV validation: Clinical example

Mosaic deletion
Ryr1 Exons 48-49

NGS:

Array confirmation:
Mosaicism and symptom severity

Mosaic variants in early onset* gene are observed at higher allele balance compared to mosaic variants in late onset* genes.

* Onset based on manual gene curation

Individuals with mosaic variants in hereditary cancer genes were considerably older than those with constitutional variants.
Other layouts / junk
Mosaicism and symptom severity

When we assessed the clinical data provided* to aid in interpretation of variants in a gene linked to a disease with very specific clinical features (e.g., NF1, TSC1, NIPBL):

- Individuals with mosaic variants were less likely to meet the strict clinical criteria that allowed for a variant to be classified as pathogenic, compared to individuals with non-mosaic variants (Fisher’s exact test p<0.001)

* In many cases, detailed clinical data is not available and thus the variant is not eligible - better clinical data can improve variant interpretation in these cases
See Poster 757!!
Prevalence of mosaicism by gene and clinical area

Hereditary Cancer

Cardiology

Neurology

Peds and rare disease
Clinical areas

Comprehensive Neuromuscular Disorders
Comprehensive Myopathy
Cardiomyopathy and Skeletal Muscle Disease
Arrhythmia and Cardiomyopathy Comprehensive
Cardiomyopathy Comprehensive
Dilated Cardiomyopathy
Arrhythmia Comprehensive
Aortopathy Comprehensive
Hypertrophic Cardiomyopathy
Ehlers–Danlos Syndrome
Comprehensive Carrier Screen
Retinoblastoma Test
Peutz–Jeghers Syndrome Test
Basal Cell Nevus Syndrome
Myelodysplastic Syndrome/Leukemia
Multi-Caner
Pediatric Nervous System/Brain Tumors
Common Hereditary Cancers
Prostate Cancer
Li–Fraumeni Syndrome Test
Bone Marrow Failure Syndromes
Gastric Cancer
Hereditary Cancer Syndromes
Breast and Gyn Cancers
Breast Cancer
Breast and Gyn Cancers Guidelines-Based
Familial Adenomatous Polyposis Test
Pancreatic Cancer
Hereditary Cancer Syndromes
Breast Cancer Guidelines-Based
Sarcoma
Colorectal Cancer
Hereditary breast cancer extended panel
Colorectal Cancer Guidelines-Based
Breast Cancer High-Risk
Melanoma
Nervous System/Brain Cancer
Renal/Urinary Tract Cancers
Breast Cancer STAT
Thyroid Cancer
Comprehensive Neuromuscular Disorders
Comprehensive Neuropathies
Comprehensive Myopathy
Hereditary Spastic Paraplegia Comprehensive
Charcot-Marie-Tooth Disease Comprehensive
WAGR Syndrome Test
Tuberous Sclerosis Complex
Cornelia de Lange Syndrome
Neurodegeneration with Brain Iron Accumulation
Urea Cycle Disorders
Neuroblasticoma Type 1 Test
Primary Immunodeficiency
Epilepsy
Primary Gliary Dyskinesia
RASopathies Comprehensive
Clinical areas

Cardiology
- Comprehensive Neuromuscular Disorders
- Comprehensive Myopathy
- Cardiomyopathy and Skeletal Muscle Disease
- Arrhythmia and Cardiomyopathy Comprehensive
- Cardiomyopathy Comprehensive
- Dilated Cardiomyopathy
- Arrhythmia Comprehensive
- Aortopathy Comprehensive
- Hypertrophic Cardiomyopathy
- Ehlers–Danlos Syndrome
- Comprehensive Carrier Screen

Cancer
- Retinoblastoma Test
- Peutz–Jeghers Syndrome Test
- Basal Cell Nevus Syndrome
- Myelodysplastic Syndrome/Leukemia
- Multi–Cancer
- Pediatric Nervous System/Brain Tumors
- Common Hereditary Cancers
- Prostate Cancer
- Li–Fraumeni Syndrome Test
- Bone Marrow Failure Syndromes
- Gastric Cancer
- Hereditary Cancer Syndromes
- Breast and Gyn Cancers
- Breast Cancer
- East and Gyn Cancers Guidelines–Based
- Familial Adenomatous Polyposis Test
- Pancreatic Cancer
- Hereditary Cancer Syndromes
- Breast Cancer Guidelines–Based
- Sarcoma
- Colorectal Cancer
- Hereditary breast cancer, extended panel
- Colorectal Cancer Guidelines–Based
- Breast Cancer High–Risk
- Melanoma
- Nervous System/Brain Cancer
- Renal/Urinary Tract Cancers
- Breast Cancer STAT
- Thyroid Cancer

Ped/Rare
- Comprehensive Neuromuscular Disorders
- Comprehensive Neuropathies
- Comprehensive Myopathy
- Hereditary Spastic Paraplegia Comprehensive
- Charcot–Marie–Tooth Disease Comprehensive
- WAGR Syndrome Test
- Tuberous Sclerosis Complex
- Cornelia de Lange Syndrome
- Leukodystrophy with Brain Iron Accumulation
- Urea Cycle Disorders
- Neurofibromatosis Type 1 Test
- Primary Immunodeficiency
- Epilepsy
- Primary Ciliary Dyskinesia
- RASopathies Comprehensive

Neurology
Clinical areas
Prevalence by gene and panel

% of PLP variants that are mosaic

# of PLP variants observed

COL3A1
ACTA2
FLNA
CACNA1C
ABCC9
HBA1
GJB2
HBB
COL7A1
FMR1

Cardiology

Carrier

Comprehensive Neuromuscular Disorders
Comprehensive Myopathy
Cardiomyopathy and Skeletal Muscle Disease
Arrhythmia and Cardiomyopathy Comprehensive
Cardiomyopathy Comprehensive
Dilated Cardiomyopathy
Arrhythmia Comprehensive
Aortopathy Comprehensive
Hypertrophic Cardiomyopathy
Ehlers–Danlos Syndrome
Comprehensive Carrier Screen

% of order with a mosaic variant

0 1 2 3 4

Carrier

Cardiology
Prevalence by gene and panel

- CHEK2
- ATM
- APC
- TP53
- PTEN
- RB1
- SMARCA4
- SMARCB1
- CDKN1C
- CDKN1B
- GATA1
- PIK3CA

% of PLP variants that are mosaic

# of PLP variants observed

Cancer

- Retinoblastoma Test
- Peutz–Jeghers Syndrome Test
- Basal Cell Nevus Syndrome
- Myelodysplastic Syndrome/Leukemia
- Multi–Cancer
- Pediatric Nervous System/Brain Tumors
- Common Hereditary Cancers
- Prostate Cancer
- Li–Fraumeni Syndrome Test
- Bone Marrow Failure Syndromes
- Gastric Cancer
- Hereditary Cancer Syndromes
- Breast and Gyn Cancers
- Breast Cancer
- Breast and Gyn Cancers Guidelines–Based
- Familial Adenomatous Polyposis Test
- Pancreatic Cancer
- Hereditary Cancer Syndromes
- Breast Cancer Guidelines–Based
- Sarcoma
- Colorectal Cancer
- Hereditary breast cancer, extended panel
- Colorectal Cancer Guidelines–Based
- Breast Cancer High–Risk
- Melanoma
- Nervous System/Brain Cancer
- Renal/Urinary Tract Cancers
- Breast Cancer STAT
- Thyroid Cancer
Prevalence by gene and panel

% of PLP variants that are mosaic

# of PLP variants observed

MTM1
PNPLA6
WNK1
AGRIN
CHRN1D
NF1
TSC2
CDKL5
WDR45
SMC1A
BCOR
DOCK7
SLC25A15
MAF
SLC8A6
CORO1A
DPAGT1
EHMT1

Neurology
Comprehensive Neuromuscular Disorders
Comprehensive Neuropathies
Comprehensive Myopathy
Hereditary Spastic Paraplegia Comprehensive
Charcot–Marie–Tooth Disease Comprehensive
WAGR Syndrome Test
Tuberous Sclerosis Complex
Cornelia de Lange Syndrome
Neurodegeneration with Brain Iron Accumulation
Urea Cycle Disorders
Neurofibromatosis Type 1 Test
Primary Immunodeficiency
Epilepsy
Primary Ciliary Dyskinesia
RASopathies Comprehensive

Ped/Rare

% of order with a mosaic variant

0
1
2
3
4