Clinical Use of Germline Genome Analysis

Jill Hagenkord, MD
Chief Medical Officer
23andMe

Overview
- Sequencing terminology and workflow
- Gene classification
- Variant classification
- Ethical, legal, and social implications
  - GINA, incidental findings
- WGS Case
- Looking forward

Sequencing Terminology
- Whole Genome Sequencing (WGS)
  - Obtaining the "complete" sequence of all 2 x 3.2 billion bp of DNA
  - Not all regions of the genome are sequence-able by NGS
- Whole Exome Sequencing (WES)
  - 1-2% of the genome containing the exons
  - ~150,000 genes, 20 – 65 kb
  - The most complete exome is obtained by using WGS
- Medical exome or Mendeliome
  - 5,000 – 7,000 genes that are associated with human disease
  - Watch out for other definition of medical exome: WES + enriched for coverage of Mendeliome
- Disorder-specific multi-gene panels
  - Cardiomyopathies
  - X-linked intellectual disability

NGS Data Workflow
- Primary Analysis
  - De-multiplexing
  - Base calling
- Secondary Analysis
  - Read mapping
  - Variant calling
  - Variant files can be annotated with additional information about the variant from ESP, dbSNP, 1000 Genomes, conservation, or in silico functional predictions, etc.
  - Annotations can be added to filter variants based on variant type, quality, allele frequency, inheritance pattern, etc.
  - ACMG variant classification. Requires gene classification.
- Tertiary Analysis
  - Variant filtering
  - Variant classification
- Clinical Interpretation
  - Integrate findings with other laboratory and clinical data

Disclosure
- 23andMe, Chief Medical Officer, shareholder

Sequencing Terminology
- Whole Genome Sequencing (WGS)
- Whole Exome Sequencing (WES)
- Medical exome or Mendeliome
- Disorder-specific multi-gene panels

Filtering, WGS

3 Main Question Remain
1. Does the variant disrupt gene (protein) function?
2. Can disrupted function lead to disease?
3. Is this disrupted gene function causative for my patient’s presentation?

Gene classification

What constitutes an “established” disease gene?
- Identification through linkage analysis
- For candidate gene analyses w/o linkage: significant segregation with disease
- In vivo functional data (in vitro effect may not translate to disease)
- Large number of race-matched healthy controls (need thousands!)
- See ClinGen https://www.clinicalgenome.org

Moving to gene classification analogous to variant classifications
1. Definitively established disease gene
2. Gene with good evidence but needs additional data
3. Candidate gene (“GUS - gene of unknown significance”)

You can only have VUS’s in GUS’s!

Gene Disease Evidence Levels

<table>
<thead>
<tr>
<th>Evidence Level</th>
<th>Evidence Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH EVIDENCE</td>
<td>Strong evidence: ( \frac{\text{gene}}{\text{protein}} ) disrupted in disease, functional data, and/or other supporting evidence</td>
</tr>
<tr>
<td>STRONG</td>
<td>Strong evidence: ( \frac{\text{gene}}{\text{protein}} ) disrupted in disease, functional data, and/or other supporting evidence</td>
</tr>
<tr>
<td>MODERATE</td>
<td>Moderate evidence: ( \frac{\text{gene}}{\text{protein}} ) disrupted in disease, some supporting evidence</td>
</tr>
<tr>
<td>LIMITED</td>
<td>Limited evidence: ( \frac{\text{gene}}{\text{protein}} ) disrupted in disease, less supporting evidence</td>
</tr>
<tr>
<td>NO EVIDENCE</td>
<td>No evidence or evidence not available for the gene in the disease</td>
</tr>
</tbody>
</table>

Exome/Genome

Proposal for Evidence Required to Include a Gene In a Clinical Test?

<table>
<thead>
<tr>
<th>Evidence Level</th>
<th>Diagnostic Panels</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROACTIVE TESTS &amp; IFs*</td>
<td>Definitive evidence</td>
</tr>
<tr>
<td>Strong evidence</td>
<td>Moderate evidence</td>
</tr>
<tr>
<td>Limited evidence</td>
<td>Disputed evidence</td>
</tr>
</tbody>
</table>

*IF = Incidental findings

Reference: VUS in GUS's
### Variant Classification

#### Types of mutations

- **Silent mutation (Synonymous)** – does not change an amino acid
  - Silent mutations can cause disease (e.g., by affecting splicing)
  - Hutchinson-Gilford Progeria: LMNA Gly608Gly

- **Missense mutations (Nonsynonymous)** – change an amino acid
  - Conservative (variant amino acid has similar properties: size, charge)
  - Nonconservative (variant amino acid has very different properties)

- **Nonsense mutations**
  - TAC (Tyr)  TAA (STOP)
  - Can cause truncated protein or absent protein (via nonsense-mediated decay, NMD)

- **Frameshift mutations**
  - GCA TGT (Ala, Cys, …)
  - TTG T (Ala, Leu, …)
  - Insertions or deletions of a number of bases that is not a multiple of 3
  - Often lead to a nonsense mutation downstream → premature termination

- **Splice site mutations**
  - Intron inclusion
  - Exon skipping with or without a frameshift
  - Message instability due to frameshift → premature stop

- **Regulatory mutations in promoter, enhancer or mRNA UTR**
  - Decreased or absent protein expression

#### Functional Consequences

- **Loss-of-function (usually recessive – except in haploinsufficiency)**
  - **Null** – complete absence of a gene product or its function
  - **Leaky** – partial absence of a gene or its function

- **Gain-of-function (usually dominant)**
  - A mutation that confers a new function to the gene product (often acts as a “dominant negative” due to disruption of the wildtype gene product)

### Possible Splicing Mutations

- Variants that **can be assumed** to affect splicing:
  - Splice donor/acceptor +/− 1, 2

- Variants that **may affect splicing**:
  - Splice donor region: +3 → +6
  - Splice acceptor region: −3, −5 → −10
  - Silent variants affecting 1st and last 3 bases of exon
  - Branch site A: 20 - 50 bases upstream of the acceptor (mutations of the branch site are highly likely to affect splicing but identification of the branch site is more challenging)

- Variants **less likely** to affect splicing:
  - Splice donor region: +7 →
  - Splice acceptor region: −4, −11 →

### Loss of Function (LOF) variants

- **Nonsense, canonical splice site, frameshift leading to premature stop**

- **All of these commonly lead to nonsense mediated decay → LOF**
  - NMD may not be 100%
  - NMD typically does NOT happen if stop is 3' of the last 50 bases of the penultimate exon

- **Most truncating mutations will be deleterious to a protein, but:**
  - Have loss of function mutations in your gene been previously shown to be pathogenic?
  - Does your mutation occur 5' of the most 3' reported pathogenic truncating mutation?
  - Is the mutation predicted to cause NMD?
  - If not, how much of the normal protein is removed and what is added (frameshift mutations)?
  - Do you expect this to disrupt your protein based upon its function?

### Missense Variants

- **How conserved is the amino acid?**
  - Species alignments for your gene
  - PolyPhen-2
  - SIFT
  - Align GVGD

- **How significant is the biochemical change to the amino acid?**
  - BLOSUM scores, Grantham differences, etc

- **Are benign polymorphisms common in your gene or never seen?**
  - TTN – Common
  - PTPN11 – Rare

- **Do these affect the structure of the protein?**
  - PolyPhen sometimes uses structural data if available
  - Is the variant located in a critical domain (e.g., ion channel pore, DNA-binding domain of transcription factor, etc) and frequently mutated in disease?
Variant Classification

- The problem of dirty databases
  - >20% variants incorrectly/inconsistently classified
  - Lack of supporting evidence for classification

- 2014 ACMG/AMP/CAP Interpretation of Sequence Variants Workgroup
  - Evidence-based classification system for variants in clinically valid genes
  - 5 Tier system

<table>
<thead>
<tr>
<th>Tier</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likely Benign</td>
<td>Allele frequency in cases/controls</td>
</tr>
<tr>
<td>Likely Pathogenic</td>
<td>Segregation with disease</td>
</tr>
<tr>
<td>Pathogenic</td>
<td>Functional (in vivo) studies</td>
</tr>
<tr>
<td></td>
<td>To some extent: computational predictions (clinically validated)</td>
</tr>
</tbody>
</table>

ACMG 2014 Variant Classification

- **Benign**
  - Allele frequency in cases/controls

- **Likely Benign**
  - Segregation with disease

- **Unknown Significance**
  - Functional (in vivo) studies

- **Likely Pathogenic**
  - To some extent: computational predictions (clinically validated)

- **Pathogenic**
  - Allele frequency in cases/controls

New ACMG classification, PATHOGENIC

- **Very strong**
  - Same as change as previously established pathogenic variant
  - Segregation with disease in a patient
  - Functional studies support deleterious effect
  - Lack of segregation in unaffected members of a family

- **Strong**
  - Same as change as previously established pathogenic variant
  - Segregation with disease in a patient
  - Functional studies support deleterious effect
  - Lack of segregation in unaffected members of a family

- **Moderate**
  - Located in a mutational hot spot or critical functional domain
  - Absent in controls or very low frequency in ESP or 1000 Genomes
  - For AR disorders, detected de novo or in cis with a pathogenic variant
  - In-frame indels in a non-repeat region or stop loss variant
  - Located in a mutational hot spot or critical functional domain
  - Absent in controls or very low frequency in ESP or 1000 Genomes
  - For AR disorders, detected de novo or in cis with a pathogenic variant
  - In-frame indels in a non-repeat region or stop loss variant
  - Location of variant supports a deleterious effect
  - Lack of segregation in unaffected members of a family

New ACMG Variant Classification, Benign

- **Very strong**
  - Allele frequency (AF) > 5% in ESP or 1000 Genomes

- **Strong**
  - AF > expected for the disorder
  - Located in a mutational hot spot or critical functional domain
  - Absent in controls or very low frequency in ESP or 1000 Genomes
  - For AR disorders, detected de novo or in cis with a pathogenic variant
  - In-frame indels in a non-repeat region or stop loss variant
  - Location of variant supports a deleterious effect
  - Lack of segregation in unaffected members of a family

- **Supporting**
  - Type of variant does not fit known mechanisms of disease
  - Observed in trans with a pathogenic variant for a fully penetrant AD disorder or observed in cis with a pathogenic variant in any inheritance pattern
  - In-frame indels in a repetitive region without a known function
  - Multiple lines of computational evidence suggest no impact on gene or protein product
  - Variant found in a case with an alternate molecular basis for disease
  - Reputable database reports variant as benign but without supporting evidence

Scoring Rules for Classification

- **Pathogenic**
  - 1 Very strong AND
  - 1 Strong OR
  - ≥ 2 Moderate AND 1 Supporting
  - 2 Supporting
  - 1 Strong AND
  - ≥ 3 Moderate OR
  - ≥ 2 Moderate AND 2 Supporting
  - ≥ 1 Moderate AND 4 Supporting

- **Likely Pathogenic**
  - 1 strong AND
  - ≥ 1 moderate OR
  - ≥ 2 supporting
  - 1 strong AND
  - ≥ 3 moderate OR
  - ≥ 2 moderate AND 2 supporting
  - ≥ 1 moderate AND 4 supporting

- **Benign**
  - 1 Strong AND
  - ≥ 2 Moderate OR
  - ≥ 3 Moderate and 2 Supporting
  - ≥ 1 Moderate AND 4 Supporting

- **Likely Benign**
  - 1 Strong AND
  - 2 Supporting

- **Uncertain Significance**
  - Everything not pathogenic or benign

2 Main Work Up Types

- **“Known” Pathogenic**
  - Exists in public database as pathogenic (HGMD, ClinVar, LSDBs)
  - Requires review to ensure that it meets your laboratory’s standard for pathogenic
  - Exists in your own database, but need to have policy to review classification periodically

- **Novel Variant**
  - Has not been previously seen or reported to be pathogenic, but is the type of variant expected to cause disease
  - Work up based on variant type and mechanism of disease

Either way, the variant must be classified according to criteria established in your laboratory.
Recap: Variant assessment and results reporting

Lab Result

- Published + in house data
- Segregations studies
- Population frequency
- Amino acid conservation
- Predictions: PolyPhen, SIFT
- Splicing predictions

Variant Annotation

- Benign
- Likely benign
- Likely pathogenic
- Pathogenic

Variant classification

- Hereditary
- Li-Fraumeni syndrome
- Peutz-Jeghers syndrome
- FAP
- Multiple endocrine neoplasia
- Hereditary retinoblastoma
- PTEN
- Tuberous sclerosis
- Retinoblastoma
- Ewing sarcoma
- Hypersensitivity
- Malignant heart disease
- Marfan syndrome
- ApoB
- Loeys-Dietz syndrome
- Tachycardia hypertrophic
- Thyroid cancer
- Myotonic dystrophy
- Shwachman-Diamond syndrome
- Familial Mediterranean fever
- SCD
- Aortic aneurysm
- Connective tissue disease

Patient Data

- Birgit Funke

Now for the tricky stuff

Incidental Findings

 Gina
What GINA does (Genetic Information Nondiscrimination Act)

- Prohibits group and health insurers from using a person’s genetic information in determining eligibility or premiums
- Prohibits the insurer from requesting or requiring that a person undergo genetic testing
- Prohibits employers from using a person’s genetic information in making decisions such as hiring, firing, job assignments, or any other terms of employment
- Prohibits employers from requesting, requiring, or purchasing genetic information about persons or their families

What GINA does not do

- Does not prevent health care providers from recommending genetic tests to their patients
- Does not mandate coverage for any particular test or treatment
- Does not prohibit medical underwriting on the basis of current health status
- Does not cover life, disability, or long-term care insurance
- Does not apply to members of the US military or veterans obtaining care via the VA system
- Does not apply to health benefit plans for federal employees or the Indian Health Service

Case: WGS in a Silicon Valley Executive

- 42 y.o. C-level executive of a Silicon Valley biotech company who had WGS as part of a promotion from a sequencing company.
- Sent the hard drive of her genome to the medical geneticist ahead of visit, and the medical geneticist contacted you for help.
- PMH: mononucleosis at age 18, fracture of the tibia in a skiing accident 15 years ago, sees OB/GYN regularly and has been told that BP, glucose, and serum lipid are all WNL.
- No medications.
- Social Hx: Married with 3 children. No smoking or drugs. Drinks 1-2 glasses of wine per day.

H&P

- Results of her WGS

  - Heterozygous missense variants: \(3 \times 10^6\)
    - 9000 in exons (or canonical splice sites)
  - 48 CNVs (>1kb)
  - Filtering, sorting, and classification performed.
Subset of Results: What do they mean for the patient?

- Heterozygous for a CNV deletion that removes 20,103 nucleotides beginning 2700 bp upstream from the transcriptional start site of the IRGM gene. Decreased expression of IRGM → increased susceptibility to Crohn's and UC. Allele freq in European’s is 10%. Having one deleted allele increases risk by 1.5 fold.
- Heterozygous for a single bp deletion in exon 13 of the MSH2 that results in a frameshift and premature termination a few bases downstream. Known pathogenic variant. Lynch Syndrome. Duty to warn.
- Heterozygous for a missense variant in the KCNQ1 subunit of the voltage-gated K+ channel. Pathogenic variants in this gene are associated with familial arrhythmias such as Long QT Syndrome, familial atrial fibrillation, and short QT syndrome. This variant has never been seen before but the substituted glutamine has been seen in the normal gene sequence of other species.
- Homozygous for an ε4 allele at the ApoE locus associated with Alzheimer’s Disease.

Looking Ahead…

The implications of genetic abundance

HYPOTHESIS: Prior knowledge of DNA will…

- Reduce the complexity and residual uncertainty of current diagnostic algorithms
- Result in faster turn around time and treatment
- Lower overall cost of all clinical laboratory testing (!!!)
- Provide preventative information that would have not otherwise been ascertained

“Rule out genetic cause” testing

- In many diagnostic algorithms, we need to rule out genetic causes of disease (EVEN THOUGH THEY ARE RARELY CAUSATIVE), but we have devised assays that let us guess at what the DNA is doing, since we lived in the Age of Genetic Scarcity
- In the Age of Genetic Abundance, we can begin with knowledge of the DNA and refine our diagnostic algorithms accordingly.
- How do diagnostic algorithms change when we have prior knowledge of genetics?

Familial Hypercholesterolemia

- Hypercholesterolemia is common in Western societies
  - 1 in 3 or 71 million people in USA have high cholesterol
  - New data suggests 1.35% are FH
- Diagnosis of FH is done by clinical criteria and usually missed
  - 40% of FH goes unrecognized in routine practice in the US
  - FH heterozygotes have 10x greater mortality from heart disease in early adulthood
  - Heart disease risk estimation tools underestimate their risk and shouldn’t be used
  - More aggressive therapy needed
- FH meets WHO criteria for screening programs
- What if all current and newly diagnosed patients with high cholesterol were screened for FH?
Abnormalities of hemostasis

- The core lab of a mid-sized tertiary care center does ~300,000 coagulation evaluations per year

- Basic tests of hemostasis: PT, aPTT, platelet count

- Many pre-analytical confounders, eg incorrect sample handling (sensitive to time, temperature, and ratio of blood to anti-coagulant in the draw tube), etc

- Many analytic confounders, eg common interfering substances (aspirin, heparin, warfarin), suboptimal intra-laboratory standardization, platelet clumping, pregnancy/hormone status

- Abnormal PT, aPTT, and platelet counts are very common and must be further evaluated to determine the cause, including “rule out genetic causes”

Hemostasis: Hereditary factors can affect almost every step

What if we had knowledge of the DNA prior to starting down the algorithms for abnormal PT and PTT?

Diagnostic algorithm for von Willebrand Disease

I may have a prolonged PTT but no bleeding in vivo

ASCP PRISE Question.

A previously healthy 34-year-old man presents to the emergency room with a red, swollen right lower leg that is tender to touch. He also reports several episodes of shortness of breath. A Doppler ultrasound of the legs rules out deep vein thrombosis (DVT) and a spiral computed tomography (CT) shows bilateral pulmonary emboli. He is started on heparin and warfarin. Two days later, you are asked which laboratory tests should be drawn to evaluate this patient for hypercoagulability. The tests that should be ordered are:

A) antithrombin, proteins C and S activity, activated protein C resistance (APCR), and Prothrombin 20210 mutation.
B) proteins C and S activity, activated protein C resistance (APCR), lupus anticoagulant, and anticardiolipin antibodies.
C) activated protein C resistance (APCR), Factor V Leiden, lupus anticoagulant, homocysteine, and antithrombin activity.
D) factor VIII activity, homocysteine, proteins C and S activity, and Factor V Leiden.
E) activated protein C resistance (APCR), Prothrombin 20210, lupus anticoagulant, and anticardiolipin antibodies.
F) Just click a button and look at the DNA to rule out hereditary coagulopathy.
Questions?
jhagenkord@23andme.com