Implicating Sequence Variants in Human Disease

Clinical Implications

Working Group Members

Euan Ashley

David Adams

David Dimmock

David Valle

Five Brief Sections

I. Defining “clinical” in context of DNA sequence variants.

II. Communicating variant information in a clinical setting: returning results to clinicians and patients.

III. A composite clinical example.

IV. Communicating clinical implication and the NGS report.

V. Some open questions.

What Does NGS Mean to a Clinician?

- A way to simultaneously test many (known) candidate genes quickly and inexpensively.
  - Currently offered as gene panels with well-worked-out test characteristics.
- Something used by researchers to find genetic causes of disease (rare or common).
  - Devolves into Sanger sequencing/traditional genetic testing.
- A clinical test to look for unsuspected or novel diagnoses in single individuals or families.
  - Most will be similar but not identical to known diseases and other unknown families.

The Spectrum of Clinical Uses for Genetic Testing

- Diagnostics
- Therapeutic decision making
  - Pharmacogenomics
  - Cancer therapy
- Decision-making surrounding reproduction
- Risk assessment

The Spectrum of Clinical Uses for Genetic Testing

- Diagnostic Certainty
  - Pre-implantation testing.
  - Recommending a prophylactic mastectomy.
- Clinical Judgment
  - Lower probability but life-saving therapies.
  - Diagnosis itself may be therapeutic, e.g. parental guilt about disease causation.
- Often a balance in practice
  - Requires enough information to assess certainty and significance of results

Variant Implication vs. Gene Implication

- Clinical Reporting of DNA Sequence Variants
  - Clinical issues well addressed by existing recommendations, e.g. ACMG 2007 standards for interpretation and reporting.
- Gene Implication
  - A particular feature of next generation sequencing.
  - Not well addressed by existing standards.
  - May benefit from a scoring rubric to categorize levels of evidence.
  - Biological implication is easy to hypothesize—care must be taken not to overstate evidence for causation.

Gene Implication Rubric Example

1. Known, well-characterized gene, likely implicated.
2. Related to known gene, some evidence for implication.
3. Biological understanding suggests possible implication.
4. Unknown relationship with phenotype.
5. Biological understanding suggests no implication.
6. Well characterized gene, unlikely to be implicated.

Returning Results: General Principles I

- NGS data generally arrives in the clinical setting as a report of results.
- Some results should be immediately apparent/highlighted.
  - Unambiguously clinically actionable.
  - Potentially severe health consequences.
  - Variant/gene data is clearly interpretable.
- ACMG recommendations pending.

Returning Results: General Principles II

- The final responsibility for interpreting and returning results resides with the ordering clinician.
  - “Variant of unknown significance.”
  - The use of consultants may be necessary.
- The level of expertise for any given variant/gene will vary among clinicians.
- The level of expertise in evaluating “raw” NGS data will vary among clinicians.

Returning Results: The Central Issue

- How is uncertainty conveyed to the clinician and to the patient?
- How to provide enough information to the clinician to allow clinical decision-making, while not:
  - Burying important information in a mass of data.
  - Assuming an excessively high level of analytic expertise of the clinician.

Real Example Sent to Pediatrician

Array Type SNP
Diagnosis Comment:
NORMAL; POSSIBLE UPD 21
Interpretation Comment:
arr (1,800,000 SNP/CN)x2

The whole genome chromosome SNP microarray (CSM) copy number analysis was normal. No significant DNA copy number changes in the 1,800,000 region specific SNP and copy number assay (Affymetrix, Inc.) were detected. There was, however, an extended contiguous region of allele homozygosity (~15 Mb observed in the chromosome 21 analysis which is suggestive of UPD (uniparental disomy) 21. There are NO confirmed imprinted genes on chromosome 21, although the homozygotic stretch of genes are subject to pairing of recessive gene disorders. If UPD confirmation is desired, DNA from the present sample should be sufficient for the additional analysis, although blood from at least one parent would be needed (test #470054). Approximate Mb linear position start-end): Chr.21:20.5-35.4 (15 Mb)

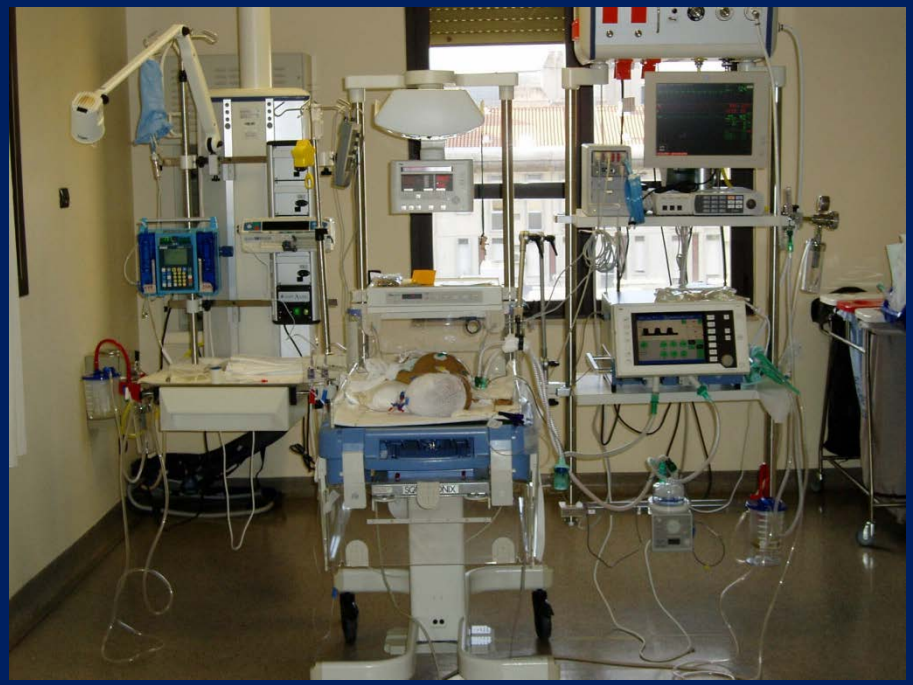
To find the genes present within the interval above, link to UCSC genome browser: (http://genome.ucsc.edu/cgi-bin hgGateway); enter start and end linear positions (chr21:20 500,000-35,400,000), press enter and go to NCBI on the menu bar for the inclusive gene list. This assay has the interprobe coverage sensitivity and specificity to reveal dosage changes across the genome at a median distance of less than 1 kb. Results of this test are for investigational purposes only. The performance

Returning Results: The Central Issue

- How is uncertainty conveyed to the clinician and to the patient?
- How to provide enough information to the clinician to allow clinical decision-making, while not:
  - Burying important information in a mass of data.
  - Assuming an excessively high level of analytic expertise of the clinician.
  - Reports should err on the side of providing more information rather than the side of oversimplification → requires a standard language.

(Composite) Example: The Family

- A family presents with a child with a complex, severe medical syndrome.
- Conventional testing does not provide a diagnosis and a clinical exome sequence is obtained.



Example: The Lab Director

- The lab director carefully analyzes the results looking for known genes that would explain the phenotype written on the test requisition form.
- No such known genes/variants are found.
- Several secondary variants of minor significance are returned in a two-page paper report.

Example I: The Consultant

- Six months pass.
- The patient is sent to a consultant who has just read about a gene reported to be mutated in five families with similar signs and symptoms.
- Knockdown of the gene in zebrafish produced a phenotype with similarities to the affected families.
- She (the consultant) wants to see if the exome sequence detected any variants in the new gene.
- The gene is not commented on in the report; it wasn't a known disease-associated gene when the report was assembled.

Example: The Consultant

- The testing lab is hesitant to return a full variant list because it includes variants that have not been CLIA/Sanger verified, vary in quality, etc..
- The consultant eventually manages to obtain a full list of detected variants.
- She notes that there are no coding variants in the gene she was interested in.
- However, there is also no indication of how well the gene was covered by the exome sequencing.

Example: The Researcher

- She asks a research colleague to sequence the new gene.
- He performs Sanger sequencing of the exons but does not find any likely variants.
- He asks to look at the exome sequence.

Example: The Researcher

- He searches the exome variant list and finds pathogenic-looking variants in a different gene in the same pathway.
- The new gene is not known to cause human disease.
- Based on well-known cell biology, however, the mutations in the pathway-associated gene he has found are predicted alter cell physiology in the same way as mutations in the known gene.

Example: The Primary Clinician

- The family calls their primary clinician to tell her that they are 10 weeks pregnant.
- They would like to know if the follow-up work on the exome sequence detected anything they could use to test the current pregnancy.

Issues to Consider

- Data revisiting
- Data reporting
- Assessment of clinical significance
- Data access
- Thresholds for clinical use

Approaches to Communicating Exome Results

- Return only results deemed important by testing lab.
  - This is the current standard.
- Return larger prioritized list.
- Return selected categories of results.
- Return all results in an annotated form with summary of important findings.
  - Requires a website or DVD.
  - “The Radiology Model.”
  - Requires standardized annotation.
  - Runs the risk of overwhelming an inexpert interpreter.

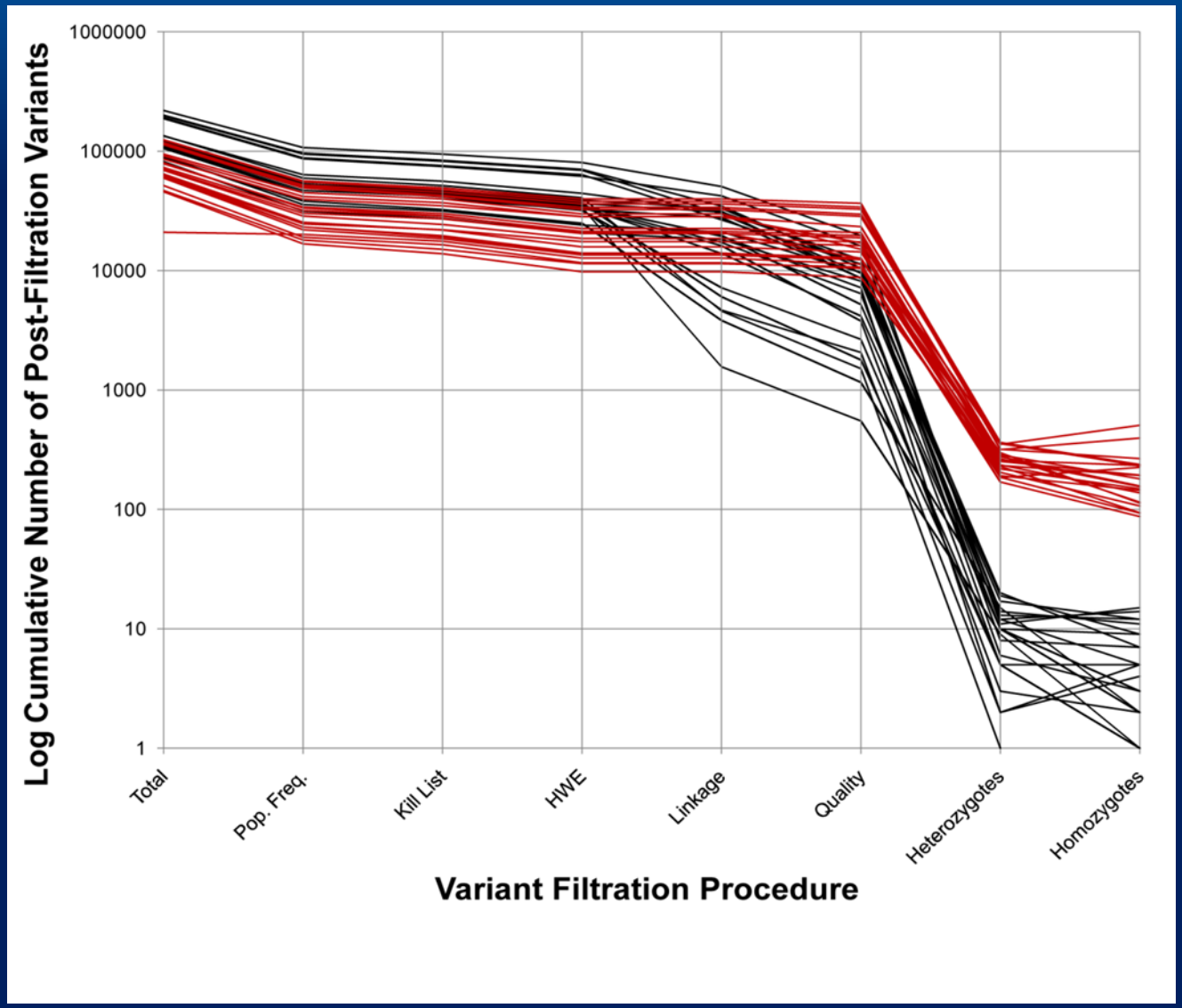
What is Optimal Report Content for Clinical Use?

- Quantitative information about likely pathogenicity (variant) and disease association (gene).
- Information about how analysis was done (assumptions, criteria for selecting/flagging results).
- Information about control populations used in analysis.
- Limitations of analytical methods.
  - General (ways method often fails).
  - Specific (what was and was not covered in specific exome sequencing instance).
- Most of these data are in need of standardization.

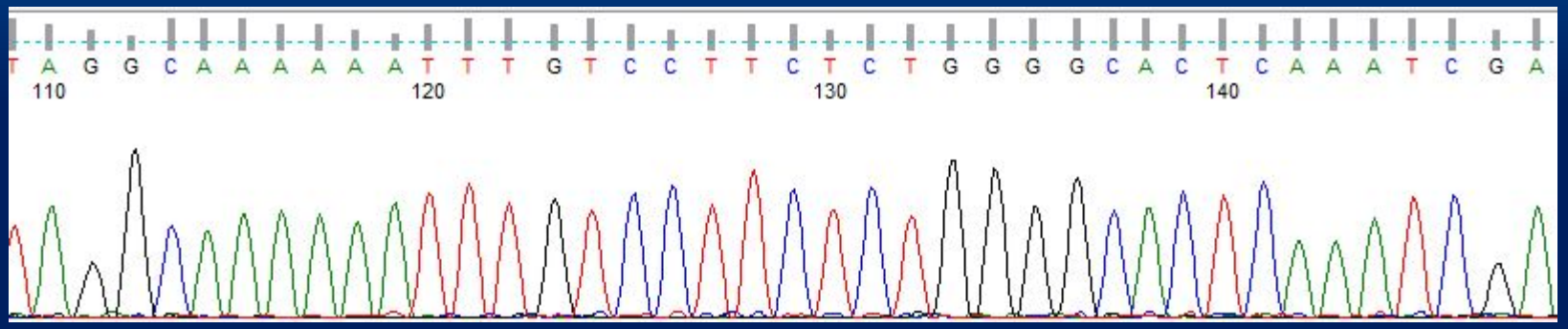
Testing Offered by Current Clinical Exome Labs

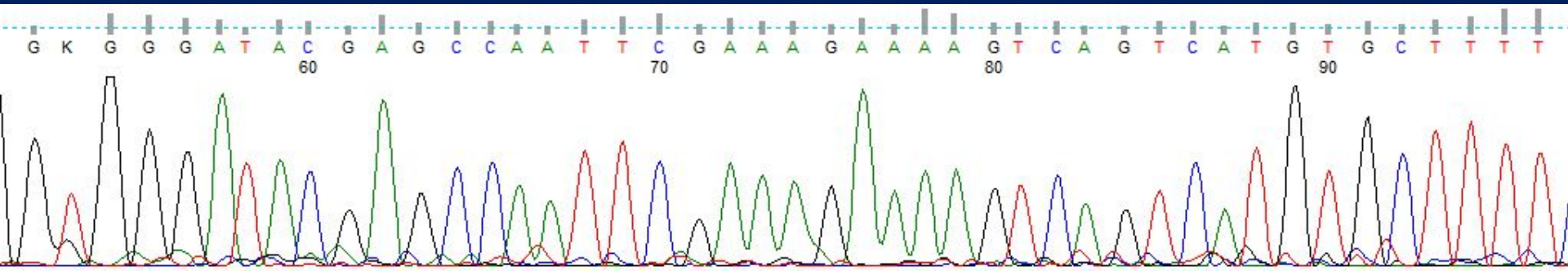
Laboratory	A	B	C	D	E	F
Tiered Testing	+	+	-	-	-	-
Family DNA Required	+	-	+	-	-	-
Parental Exome Performed	+	-	+	-	-	-
Return of Secondary Variants	+	+	-	+	-	-
Variant Return if <18 years	-	+	-	-	-	-
Focused Report Available	+	+	-	+	-	-
Re-analysis of exome	-	+	+	-	+	?

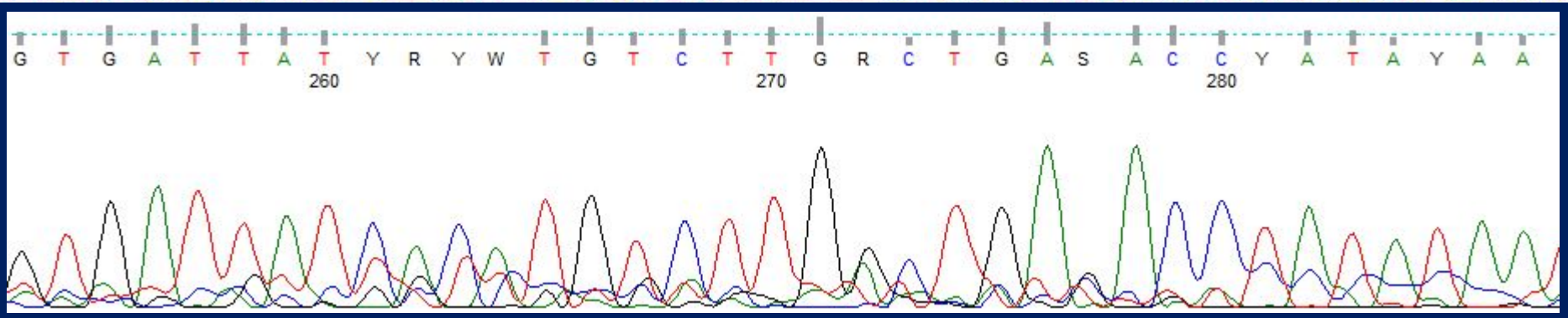
Assumptions Made During Data Analysis



Failure Modes: Sanger <> NGS







Limited Data Currently Being Returned

**Quality Metrics:*

Mean Depth of Coverage[1]	100X
Exome targeted region covered[2]	>99%
Quality threshold[3]	95%

The above values represent metrics from this [redacted] evaluation. [1]Mean depth of coverage refers to the sequence mean read depth across the [redacted] targeted region, defined as coding exons and splice junctions of [redacted] kit targeted protein coding RefSeq genes. [2]The total [redacted] target region covered at 1x. [3]The quality threshold refers to the percentage of the [redacted] defined target region where read depth was at least 10x coverage to permit high quality exome variant base calling, annotation and evaluation. Average quality thresholds may range from 90-95% of the [redacted] targeted region, indicating a small portion of the target region may not be covered with sufficient depth or quality to confidently call variant positions.

Selected Outstanding Questions

- Are there different thresholds for:
  - Publication
    - High evidence cutoff may inhibit dissemination of information.
    - Inadequate/excessive statement of evidence may lead to unintentional clinical use.
  - Clinical Use
    - High cutoff does not allow for clinical judgment.
      - Ultra-rare conditions – just a few families.
    - Low cutoff with inadequate/excessive statement of evidence may lead to medical error/patient harm.

Other Observations

- Common disease will be particularly challenging:
  - Multiple small effects.
  - Gene interactions.
  - Challenges in establishing gene causation.
  - Conflicting literature
    - Large number of potential stakeholders leads to large, internally-inconsistent body of journal literature.
    - Example *MTHFR* c.677C>T

Questions

- How much NGS analytic detail should be incuded in a clinical report?
- Can analytic results ad procedures be reported in a standarized manner?
- What is the best way to report clinical NGS studies?
- Are there different criteria for publication and clinical use? Different clinical examples?
- At the clinical level (clinician & patient) is an NGS study a one-time measurement or a reusable resource?
- Should a gene that is not well known to be associated with the presenting phenotype *ever* be reported/flagged?