ClinGen: Sharing genomic variant data to increase knowledge and improve patient care

David H. Ledbetter, PhD, FACMG on behalf of the ClinGen Team
March 28, 2014
Conflict of Interest Statement

- No Conflicts to Disclose
International Standards for Cytogenomic Arrays (ISCA) Consortium

• Established in 2007 (ACMG-F/Luminex and NIH)
• Over 190 member institutions
• Over 2,800 individual members worldwide
Expanding the ISCA Efforts

• Realizing the goals of ISCA are not unique to structural variation, we evolved into ICCG (2012) to include sequence-level variation efforts.

Project PIs:
Christa Martin (Geisinger), Heidi Rehm (Harvard),
David Ledbetter (Geisinger), Joyce Mitchell (Utah), Robert Nussbaum (UCSF)
Goals of ICCG

- Work toward **standardizing** test designs and data interpretation related to structural and sequence-level variation
- Develop a **centralized database** of clinically relevant variant annotations to **share data** for clinical and research purposes
- Implement a **QC** and **expert consensus process** for curating data submitted across laboratories and developing evidence based classifications
- **Support** the needs of our stakeholders, including laboratories, clinicians, researchers, AND families
- Advocate for **responsible data sharing**
Towards a Unified Clinical Genomics Database...
The Clinical Genome Resource (ClinGen)

Collaboration between:

- **NHGRI U41 Grant**
  PIs: Rehm (Harvard), Martin (Geisinger), Nussbaum (UCSF), Ledbetter (Geisinger)

- **NHGRI U01 “Clinically Relevant Variant Resource” Grants**
  Grant 1 PIs: Plon (Baylor), Bustamante (Stanford)
  Grant 2 PIs: Berg (UNC), Watson (ACMG), Ledbetter/Williams (Geisinger)

- **NCBI**
  ClinVar
The Clinical Genome Resource (ClinGen)

• **Purpose:** Create a centralized repository and interconnected resources of clinically annotated genes and variants to improve our understanding of genomic variation and optimize its use in genomic medicine.

• **Main activities:**
  – Facilitate deposition of variants and phenotypes into ClinVar & develop methods for defining pathogenicity (Harvard/Geisinger/UCSF)
  – Organize clinical curation groups, lead consensus efforts for clinical actionability, & ensure interoperability with electronic health records (UNC/ACMG/Geisinger)
  – Build informatics support, facilitate data access by the community, and develop novel machine learning algorithms (Baylor/Stanford)
Other Genomic Data Standards/Data Collection Workgroups

Clinical Domain Workgroups

Bioinformatics & IT Workgroups

Data Collection into ClinVar; methods for defining pathogenicity
Harvard/Geisinger/UCSF

Clinical curation groups for clinical actionability;
EHR interoperability
UNC/Geisinger/ACMG

Informatics resources for data analyses and access;
machine learning
Stanford/Baylor

Scientific Advisory Board

Steering Committee
Harvard/Geisinger/UCSF/Utah, UNC/Geisinger/ACMG,
Stanford/Baylor, NHGRI, NCBI

External Stakeholder Committee

NHGRI

Data Analysis Subcommittee

ELSI and Genetic Counseling WG

Structural Variant WG

Sequence Variant WG

Phenotyping WG

Other Genomic Data

Sequence Variant WG

ELSI, Education, Engagement, Access WG
Data Sharing Benefits Everyone

• Benefits to users (laboratories, clinicians, researchers)
  – Allows others to easily identify similar cases
  – Informs result interpretation (clinical assertion comparison)
  – Informs patient counseling (phenotype comparison)

• Benefits to submitters
  – Quality assurance measure
    • Inter-laboratory comparison
    • Intra-laboratory comparison
## Data Submissions

<table>
<thead>
<tr>
<th>ClinVar Submitters</th>
<th>Variants</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMIM</td>
<td>23524</td>
<td>3077</td>
</tr>
<tr>
<td>Harvard Medical School and Partners Healthcare</td>
<td>6996</td>
<td>155</td>
</tr>
<tr>
<td>International Standards for Cytogenomic Arrays (≥40,000 dbGaP)</td>
<td>4194</td>
<td>46</td>
</tr>
<tr>
<td>GeneReviews</td>
<td>2913</td>
<td>287</td>
</tr>
<tr>
<td>ARUP Laboratories</td>
<td>1415</td>
<td>6</td>
</tr>
<tr>
<td>LabCorp</td>
<td>1391</td>
<td>140</td>
</tr>
<tr>
<td>Sharing Clinical Reports Project</td>
<td>902</td>
<td>2</td>
</tr>
<tr>
<td>Finland Institute for Molecular Medicine</td>
<td>840</td>
<td>39</td>
</tr>
<tr>
<td>Tuberous Sclerosis Database</td>
<td>431</td>
<td>1</td>
</tr>
<tr>
<td>ClinSeq Project</td>
<td>425</td>
<td>35</td>
</tr>
<tr>
<td>Leiden Muscular Dystrophy Database</td>
<td>220</td>
<td>10</td>
</tr>
<tr>
<td>GeneDx</td>
<td>205</td>
<td>3</td>
</tr>
<tr>
<td>Emory Genetics Laboratory</td>
<td>48</td>
<td>13</td>
</tr>
<tr>
<td>American College of Medical Genetics and Genomics</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>Osteogenesis Imperfecta Database; University of Leicester</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>Ambry Genetics</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Other laboratories (19)</td>
<td>52</td>
<td>25</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>43604</strong></td>
<td><strong>3844</strong></td>
</tr>
</tbody>
</table>
Submitted laboratory calls are not changed without discussion with the laboratory.
ISCA/ICCG Submitter Benefits: Quality Assurance

1) Quality Assurance

- Consistency of calls overtime
- Alerts to any changes in interpretations (e.g., 16p11.2 del)

2) Standardization of CNV interpretations across laboratories

Rate of Conflicts Identified through Curation

<table>
<thead>
<tr>
<th></th>
<th>Intra-lab</th>
<th>Inter-lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td></td>
<td>8%</td>
</tr>
<tr>
<td>1%</td>
<td></td>
<td>7%</td>
</tr>
<tr>
<td>2%</td>
<td></td>
<td>6%</td>
</tr>
<tr>
<td>3%</td>
<td></td>
<td>5%</td>
</tr>
<tr>
<td>4%</td>
<td></td>
<td>4%</td>
</tr>
<tr>
<td>5%</td>
<td></td>
<td>3%</td>
</tr>
<tr>
<td>6%</td>
<td></td>
<td>2%</td>
</tr>
<tr>
<td>7%</td>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>8%</td>
<td></td>
<td>0%</td>
</tr>
</tbody>
</table>
ICCG Sequence Variant Example: U41/ClinVar Pilot Project

Comparison of three laboratories classifications for variants in 12 RASopathy genes: BRAF, CBL, HRAS, KRAS, MAP2K1, MAP2K2, NRAS, PTPN11, RAF1, SHOC2, SOS1, SPRED1

<table>
<thead>
<tr>
<th>Scope</th>
<th>Number of alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total submitted to ClinVar</td>
<td>997</td>
</tr>
<tr>
<td>Multiple assertions</td>
<td>269</td>
</tr>
</tbody>
</table>

53 discrepancies:
60% differ based upon likelihood (Benign vs LB, P vs LP)
34% differed VUS vs Likely Pathogenic/Likely Benign
6% differed VUS vs Pathogenic
ICCG Experience: Inter-laboratory conflicts

Structural Variation

7% discrepant

Sequence Variation

20% discrepant
June 18, 2013

AMA Adopts New Policies on Second Day of Voting at Annual Meeting

For immediate release:
June 18, 2013

CHICAGO - The American Medical Association (AMA), the nation’s largest physician organization, voted today during its Annual Meeting to adopt the following new policies on emerging issues in public health and science:

Support of Public Access to Genetic Data
The AMA today adopted policy that encourages companies, laboratories, researchers and providers to publicly share data on genetic variants and the clinical significance of those variants through a system that assures patient and provider privacy.

“On the heels of the U.S. Supreme Court decision that genetic information is not patentable, this policy urges collaboration and data sharing with privacy protections to advance genomic medicine,” said Dr. Kobler. “Genetic analyses done collectively with the ability to compare genetic variants and analysis will allow researchers and health care professionals to more quickly identify and adopt advances in genomic medicine to benefit patients.”
Public Access

- ClinVar
- dbVar
- UCSC Genome Browser
- Array Vendor Software Analysis Tools
- ICCG Website Search Tool
Welcome to the ISCA Database Search. We recommend using Firefox, Chrome, or Safari web browsers. For optimum functionality, please set your browser to allow popups and javascript execution from our web site. Thanks.
ISCA/ICCG tracks
Evidence-based Process

- Systematic method for integrating research evidence with clinical expertise
- Evaluation of data based on strength, credibility and reproducibility
- Ultimate goal of translating information into clinical recommendations or practice guidelines
Application in Genomic Medicine

• Challenging due to the rarity of genomic disorders . . . why large, shared datasets are needed!

• Goals:
  1) Use evidence-based review to assess genes/regions targeted on ICCG/ISCA clinical testing designs
  2) Develop a genome-wide dosage sensitivity map to aid in clinical array interpretation and inform array design
     • Haploinsufficiency (loss)
     • Triplosensitivity (gain)
ICCG Curation Efforts to Create a Medical Genome

Goal: To define genes with medical relevance

Variant types:

- CNV – haploinsufficient
- CNV – triplosensitive
- Other structural changes (translocations, inversions, etc.)
- Sequence variants (substitutions, small indels)
  - Loss-of-function
  - Gain-of-function
- Imprinted loci
- Repeat expansions
The Medical Genome for CNVs

Rating System

Highest -- 3, 2, 1, 0, unlikely dosage sensitive -- Lowest

Review

Towards an evidence-based process for the clinical interpretation of copy number variation


The evidence-based review (EBR) process has been widely used to develop standards for medical decision-making and to explore complex medical questions. This paper describes the creation of an EBR process for the clinical interpretation of copy number variation.
Evidence for Dosage Sensitivity

Is this genomic region associated with a clinical phenotype?

Is this clinical phenotype associated with dosage sensitivity?

How many lines of evidence are there to support dosage sensitivity?

Are CNVs involving this genomic region enriched in disease populations?

Is there any compelling evidence to refute its dosage sensitivity?

Peer-reviewed literature
Case databases
...or other type of mutation mechanism
Correlates with dosage rating
Case/control studies

Evidence-Based Rating Can Help Standardize Clinical Interpretation

3 = Sufficient evidence = “pathogenic”

2 = Emerging evidence = “likely pathogenic” or “uncertain”

1 = Little evidence = “uncertain”

0 = No evidence = “uncertain” or “likely benign”

Unlikely dosage sensitive = “likely benign” or “benign”

To Date: >500 genes reviewed
A publicly available genomic database has many benefits, but it must be maintained in a way that:

- Respects patient privacy AND
- Promotes active use among the clinical and research communities

Current policies of similar databases initially developed to accommodate traditional research applications

However, many of the requirements may represent barriers to potential clinical users:

- IRB-approved protocols
- Annual usage reports
- Requirement to fully download entire datasets in order to search for variants
Possible Solution: Tiered Access System?

Actively working with NIH to determine most appropriate solutions to support the different stakeholder groups.
Community Outreach

Connect with us!
- On Twitter: @TheICCG
- On Facebook: InternationalCollaborationForClinicalGenomicsICCG

2014 ICCG/ClinGen Meeting
- June 10-12 Bethesda North Marriott
Acknowledgements

- **U41 PIs**
  - Heidi Rehm (Harvard Partners)
  - Christa Martin (Geisinger)
  - Robert Nussbaum (UCSF)

- **ICCG Staff**
  - Danielle Metterville (Harvard Partners)
  - Erin Rooney Riggs (Geisinger)

- **UNC/Geisinger/ACMG PIs**
  - Jonathan Berg (UNC)
  - David Ledbetter (Geisinger)
  - Mike Watson (ACMG)

- **Stanford/Baylor PIs**
  - Carlos Bustamante (Stanford)
  - Sharon Plon (Baylor)

- **NCBI**
  - Deanna Church
  - Mike Feolo
  - Melissa Landrum
  - Donna Maglott
  - Jose Mena
  - Jim Ostell
  - George Riley
  - Steve Sherry

- **Contribution Laboratorys**
  - ARUP
  - Emory University
  - GeneDx
  - Mayo Clinic
  - Partners Laboratory for Molecular Medicine
  - University of Chicago
  - All ICCG Laboratories

- **Grant Support**
  - NICHD
    - HD064525
  - NIMH
    - RO1 MH074090
  - NHGRI
    - U41 HG006834
    - U01 HG007436
    - U01 HG007437
  - ACMG/Luminex

[www.iccg.org](http://www.iccg.org)