One Precision Medicine

Please

Richard Gibbs AC Ph.D.,
Baylor College of Medicine –
Human Genome Sequencing Center

- BCM is a joint owner of Baylor-Miraca Genetics Laboratories (BMGL)
- BCM owns stock in Codified Genomics
Clinical Genetics Catch 22:

1: Need medical relevance to incentivize clinical application,

2: Need discovery to drive clinical relevance,

3: Need large numbers of well characterized research participants and freely accessible data to catalyze discovery,

4: The best source of data is the clinic – but not yet enough Medical relevance,
Somewhere between 2012-2018 coined the ‘virtuous cycle’

Overcoming the Catch-22

Complete reciprocity between the Clinical Interface and Research/Discovery.

- Returning Valuable Genetic Data
- Recovering Clinical Records
- Driving Discovery
- Improving Clinical Value
- etc
Personal genomes and whole exome sequencing:

- HGP, HapMap, etc
- Watson Individual variation
- Lupski Disease gene
- Beery Medically Actionable

Clinical Genomes (>35x)

‘Kringle repeats’ in LPA, All STRs (e.g. Fragile X, Huntingtons)
CYP2D6 Haplotypes
Complex rearrangements

VCRome 2.1
Exome 42 Mbp

BCM Capture Reagent: HGSC VCRome
Developed at BCM:HGSC and commercially available at Roche (SeqCap EZ HGSC VCRome)
High frequency
Few Actionable Alleles
Missing ‘Heritability’

Low frequency/High impact
Not related to sporadic cases in a simple way

BCM Capture Reagent: HGSC VCRome
Developed at BCM:HGSC and commercially available at Roche (SeqCap EZ HGSC VCRome)
~220,000 EXOMES, WGS, PANELS

RESEARCH (~160,000)

>25,000 Clinical Samples

DIAGNOSIS (~15,000)

>2% of lives in USA (DIGNITY + CHI)

Texas Medical Center (7M per year)

HGSC

BCM-JOHNS HOPKINS CENTER FOR MENDELIAN GENOMICS (~11,500)

James Lupski
David Valle

CommonSpirit

Baylor College of Medicine
JOHNS HOPKINS SCHOOL of MEDICINE

Genetic Discovery
# Tremendous Progress in Mendelian Discovery


<table>
<thead>
<tr>
<th></th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples</td>
<td>61,286</td>
</tr>
<tr>
<td>Families</td>
<td>22,742</td>
</tr>
<tr>
<td>Collaborators</td>
<td>3,928</td>
</tr>
<tr>
<td>Countries</td>
<td>80</td>
</tr>
</tbody>
</table>

~263 novel discoveries per year
Average 1 novel discovery per 28 WES

## Recruitment

*Unsolved Clinical Cases feed discovery*

*Discovery Feeds Clinical Diagnoses*

(GTR Genetics Testing Registry)
4 subjects with \textit{de novo} NS mutations in \textit{AHDC1} define new disease: Xia-Gibbs Syndrome (XGS)

De Novo Truncating Mutations in \textit{AHDC1} in Individuals with Syndromic Expressive Language Delay, Hypotonia, and Sleep Apnea


Clinical whole-exome sequencing (WES) for identification of mutations leading to Mendelian disease has been offered to the medical community since 2011. Clinically undiagnosed neurological disorders are the most frequent basis for test referral, and currently, approximately 25% of such cases are diagnosed at the molecular level. To date, there are approximately 4,000 "known" disease-associated loci, and many are associated with striking dysmorphic features, making genotype-phenotype correlations relatively straightforward. A significant fraction of cases, however, lack characteristic dysmorphism or clinical pathognomonic traits and are dependent upon molecular tests for definitive diagnoses. Further, many molecular diagnoses are guided by recent gene-disease association discoveries. Hence, there is a critical interplay between clinical testing and research leading to gene-disease association discovery. Here, we describe four probands, all of whom presented with hypotonia, intellectual disability, global developmental delay, and mildly dysmorphic facial features. Three of the four also had sleep apnea. Each was a simplex case without a remarkable family history. Using WES, we identified \textit{AHDC1} de novo truncating mutations that most likely cause this genetic syndrome.

Clinical presentation of 4 subjects.
Sept 10, 2019

“Dear Prof Gibbs: I have had a diagnosis of Xia-Gibbs syndrome for one of my patients through the 100,000 genomes project. She is nearly 13 yrs old with a de novo nonsense variant in AHDC1. I have been seeing her for 12 yrs so it’s a fantastic to finally have an explanation for the family………

Nora Shannon Consultant Clinical Geneticist
AHDC1 Diagnoses:

- *De novo* truncating mutations = relatively homogeneous phenotypes,
- Possible ‘gradient of severity’ along the coding region,
- *De novo* missense being reported as ‘pathogenic’, with milder phenotypes,
- *De novo* CNVs reported as pathogenic – mixed phenotypes,

**BUT – CLINICAL DIAGNOSTICIANS ‘RUNNING AHEAD’!!**

- *De novo* missense being reported as ‘pathogenic’, with milder phenotypes,
- *De novo* CNVs reported as pathogenic – mixed phenotypes,
7 AHDC1 Missense Cases:

- Reported variants cluster: ●
- SOME likely pathogenic (5 of 7),
- High-CADD common variants?
- Demands a ‘functional correlate’
AHDC1 Function:

1: Few clues from homologies:
   - DNA Repair?
   - Chromatin maintenance?

2: Transfection/ Visualization, reveals signs of ‘nucleolar stress’,
LESSONS FROM MENDELIAN DISCOVERY

I: Driving deep-dives into biology,

II: More De novo mutations than expected

III: Dual-Diagnoses  
~ 5% of diagnosed cases have >1 disorder

IV: Re-Analysis Works – More ‘Genetics’ than thought:

Jennifer Posey  
Pengfei Liu

July 2019 NEJM
Clan Genomics and the Complex Architecture of Human Disease

James R. Lupski,1,2,3,* John W. Belmont,1,2 Eric Boerwinkle,4,5 and Richard A. Gibbs1,6,*

an alternative to the CVCD hypothesis

Recent new mutations and novel combinations account for many medically actionable variants
‘Genetic Burden’: Real Examples:

I: Charcot-Marie Tooth
Statistical Evidence

Model interactions in zebrafish

37 unrelated families,

Sub-optimal dose morpholino A1 + Sub-optimal dose morpholino A2 = phenotype

II: Parkinson’s Disease
2016 – 150 probands
No burden signal

• Two novel loci implicated (TKN2, TNR)
• Genetically heterogeneous

2018 – 1,516 probands
Burden signal!

Laurie Robak  Iris Jansen
Joshua Shulman
International Parkinson’s Disease Genetics Consortium

Whole-Exome Sequencing in Familial Parkinson Disease

• 22% of PD cases have 2 or more putative damaging variants
**2010 – PRESENT: Unsuccessful ‘burden studies’**
- No Evidence Yet for Oligogenics: New loci, underpowered for interactions?

### IMMEDIATE COMMUNICATION

Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy

G McMichael, JN Bainbridge, EE Haan, M Corbett, A Gardner, S Thompson, BWM van Bon, CL van Eyk, J Broaders, C Reynolds, ME O’Callaghan, LS Nguyen, DL Adelson, R Russo, S Jhangiani, H Dodds-Stampanoni, DM Muzny, RA Gibbs, J Gecz, and AH MacLennan

- 183 Probands (+/- parents)
- WES
- 14% ‘genetic causes’
- Genetically heterogeneous

### Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease


- 128 Families
- WES
- Potential genetic cause in >85% of families
- 41 known genes, 41 candidate genes

### Whole-Exome Sequencing Identifies Novel Variants for Tooth Agenesis

N. Dinckan, R. Du, L. E. Petty, more...

First Published August 16, 2017 | Research Article

- ~150 Probands (+/- parents)
- WES
- Two novel loci implicated (TKN2, TNR)
- Genetically heterogeneous

### Clinical Medicine

Molecular etiology of arthrogryposis in multiple families of mostly Turkish origin

Molecular etiology of arthrogryposis in multiple families of mostly Turkish origin

Yavuz Bayram,1 Ender Karaca,1 Zeynep Coban Akdemir,1 Elf Oztan Yilmaz,2 Gulsen Alpay Tayfun,1 Hatip Aydin,4 Deniz Torun,4 Sevcan Tug Bozdagcan,4 Alper Gezdirici,7 Sedat Isikay,4 Mehmed M. Attik,1 Tomasz Gambijn,1 Tamar Harel,1 Ayman W. El-Hattab,4 Wu-Lin Chang,1 Davut Pehlivan,1 Shalini N. Jhangiani,10 Donna M. Muzny,9 Ali Karaman,11 Tamer Celik,12 Ozge Ozalp Yuregir,13 Timur Yildirim,11 Ilhan A. Bayhan,16 Eric Boerwinkle,20,21 Richard A. Gibbs,30 Nurset Elcioglu,4 Beyhan Tuyusuz,4 and James R. Lupski14,30,37

ARTICLE
The Genomics of Arthrogryposis, a Complex Trait: Candidate Genes and Further Evidence for Oligogenic Inheritance

Davut Pehlivan,1,2,8 Yavuz Bayram,1,3,28 Nilay Gunes,4 Zeynep Coban Akdemir,1 Anju Shukla,5 Tatjana Biehtals,6 Burcu Tabakci,7 Yavuz Sahin,8 Alper Gezdirici,7 Jawid M. Fatih,9 Elif Yilmaz Gulec,9 Gozde Yesil,10 Jaya Punetha,1 Zeynep Ocak,7 Christopher M. Grochowski,1 Ender Karaca,1 Hatice Mutlu Albayrak,11 Penyasaamy Radhakrishnan,12 Hakan Bagis Erdem,13 Ibrahim Sahin,13 Timur Yildirim,14 Ilhan A. Bayhan,14 Aysegul Bursali,14 Muhsin Elmas,15 Zafer Yuksel,16 Ozturk Ozdemir,17 Fatma Silan,17 Onur Yildiz,17 Osman Yesilbas,16 Sedat Isikay,19 Burhan Balta,20 Shen Gu,1 Shalini N. Jhangiani,10 Harsha Dodgapaneni,21 Jianhong Hu,25 Donna M. Muzny,24 Baylor-Hopkins Center for Mendelian Genomics, Eric Boerwinkle,20,21,22 Richard A. Gibbs,1,21 Konstantinos Tsiakas,23 Maja Hempel,6 Katta Mohan Girisha,5 Davut Gul,24 Jennifer E. Posey,1 Nurset H. Elcioglu,7,25 Beyhan Tuyusuz,4 and James R. Lupski14,21,26,27,4

Arythrogryposis: ~65% diagnosed (54/85)
Relative frequency of multiple molecular diagnoses is higher in a cohort with high identity-by-descent, and this is driven by high rate of AR + AR dx.

<table>
<thead>
<tr>
<th></th>
<th>N. Am cohort</th>
<th>Turkish cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of multiple molecular diagnoses</td>
<td>4.9% (101/2076)</td>
<td>22.0% (18/82)</td>
</tr>
<tr>
<td>Rate of AR+AR</td>
<td>10.9% (11/101)</td>
<td>88.9% (16/18) P&lt;0.0001</td>
</tr>
</tbody>
</table>
Arythrogryposis: AOH and Dual Diagnoses

- First link of (cryptic) consanguinity to more ‘dual diagnoses’
- Shows ‘aggregation’ of pathogenic alleles through different mechanisms of disequilibrium,
The Last Eight Years- *The Moon Mission*: Pediatrics

Mendelian Genetics – in ‘good shape’,

Foundation is driving ‘oligogenic genetic studies,

Major Lesson: WES of children demonstrated that merging research and clinic speeds accrual of large numbers of participants and promotes rapid progress!

**THE ‘MARS’ GENETICS MISSION:**

*Establishing routine genomics in the adult clinic in order to boost the number of available, well phenotyped, ACCESSIBLE genomes for research studies.*
Isn’t that what everyone wants to do...?

<table>
<thead>
<tr>
<th>PROGRAM</th>
<th>UK Biobank</th>
<th>Genome England</th>
<th>TOPMed</th>
<th>CCDG</th>
<th>All of Us</th>
<th>IDEAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIZE (Millions)</td>
<td>&lt;0.5</td>
<td>0.1-5.0</td>
<td>0.1-0.5</td>
<td>~0.1</td>
<td>1.0</td>
<td>&gt;1.0</td>
</tr>
<tr>
<td>CLINICAL REQUIREMENTS</td>
<td>None</td>
<td>Rare disease, Cancer</td>
<td>None</td>
<td>Common disease</td>
<td>None</td>
<td>Rare and common genetic disease</td>
</tr>
<tr>
<td>DATA ACCESS</td>
<td>Open</td>
<td>Controlled</td>
<td>Open</td>
<td>Open</td>
<td>Open</td>
<td>Open</td>
</tr>
<tr>
<td>PHENOTYPING</td>
<td>Questionnaire</td>
<td>Clinical</td>
<td>Varies by study</td>
<td>Varies by study</td>
<td>Questionnaire</td>
<td>Clinical, medical record</td>
</tr>
<tr>
<td>RETURN OF RESULTS</td>
<td>No</td>
<td>Yes- via clinic</td>
<td>No</td>
<td>No</td>
<td>Yes- direct to participant</td>
<td>Optional via clinic</td>
</tr>
<tr>
<td>GOVERNANCE</td>
<td>Program directed</td>
<td>Program directed</td>
<td>Program directed</td>
<td>Program directed</td>
<td>Program directed</td>
<td>‘bottom up’</td>
</tr>
</tbody>
</table>
DELIVERING GENETIC TESTS IN PRECISION MEDICINE
AN ONGOING EXPERIMENT:

I: DIRECT TO PARTICIPANT

**PROS:**
- Centralized
- Scalable

**CONS**
- Participants unprepared
- Physicians unengaged
- No cost models

II: PHYSICIAN/HEALTH CARE PROVIDER PATHWAY

**PROS:**
- Physician, System ordered
- Early embedding in EHR
- Flexible results return

**CONS**
- Scalability
- Physician readiness
- System economics
eMERGE III and All of Us

- 25,000 Participants,
- 11 collection sites
- Clinician reporting,
- ACMG +10 reporting

- 1,000,000 Participants
- Direct report to participants
- >250,000 accrued, ROR being decided

- Diversity at the scale of 1 million people: demographically, geographically, medically, and especially those underrepresented in biomedical research
- Focus on participants as partners: included in governance, invited to co-invent systems and give input into the science, choice to receive data and information back
- Diversity of data types collected longitudinally: clinical, environmental, genetic, behavioral, socioeconomic – with ability to recontact participants
- National/international, open resource for all: open to the public and all researchers, open source software & tools
Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network

The eMERGE Consortium**
# eMERGE III Haromization:

<table>
<thead>
<tr>
<th>Item</th>
<th>Challenge</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection sites</td>
<td>sample type</td>
<td>agreed to blood(^a),(^b)</td>
</tr>
<tr>
<td></td>
<td>sample quality</td>
<td>minimal quantity specified(^a)</td>
</tr>
<tr>
<td></td>
<td>intake formats</td>
<td>standard tables supplied to sites</td>
</tr>
<tr>
<td></td>
<td>phenotypes</td>
<td>not shared unless indication for testing</td>
</tr>
<tr>
<td></td>
<td>patient ID structure</td>
<td>naming conventions</td>
</tr>
<tr>
<td></td>
<td>indications for testing</td>
<td>selected 40 “hard coded”</td>
</tr>
<tr>
<td>Assay development</td>
<td>gene targets</td>
<td>selected by consensus</td>
</tr>
<tr>
<td></td>
<td>capture strategy</td>
<td>agreed exons (+/− 15 bases)/SNPs; capture probes spanned min 100 bases</td>
</tr>
<tr>
<td></td>
<td>capture reagents</td>
<td>two platforms supported (Nimblegen and Illumina Rapid Capture)</td>
</tr>
<tr>
<td></td>
<td>Sanger validation</td>
<td>rare variants always Sanger validated; for common SNVs, stopped validation after 5 confirmations</td>
</tr>
<tr>
<td></td>
<td>CNV validation</td>
<td>all CNVs by orthogonal technology</td>
</tr>
<tr>
<td>Validation/proficiency</td>
<td>technical performance/coverage</td>
<td>min standards (200×; 95% coverage, etc.)</td>
</tr>
<tr>
<td></td>
<td>ongoing proficiency</td>
<td>interlaboratory exchange or eMERGE samples and use of standard CAP NGS PT</td>
</tr>
<tr>
<td>Primary analysis</td>
<td>CNV calling parameters</td>
<td>3+ exons</td>
</tr>
<tr>
<td></td>
<td>pharmacogenomics</td>
<td>report variants and inferred diplotypes</td>
</tr>
<tr>
<td>Variant classification</td>
<td>initial harmonization</td>
<td>required harmonization of all medically significant differences observed 5 or more times in tested genes</td>
</tr>
<tr>
<td></td>
<td>ongoing classifications</td>
<td>required consensus between labs or elevation to Clinical Annotation WG for network consensus</td>
</tr>
<tr>
<td>Report content(^a)</td>
<td>consensus content</td>
<td>67 genes and 14 SNVs</td>
</tr>
<tr>
<td></td>
<td>site-specific genes and SNVs</td>
<td>see Figure 4 and Table S7</td>
</tr>
<tr>
<td></td>
<td>updates</td>
<td>variant reclassifications provided</td>
</tr>
<tr>
<td>Data delivery</td>
<td>physician clinical reports</td>
<td>PDBs, consumable xml structure; GenelInsight</td>
</tr>
<tr>
<td></td>
<td>network access to interpreted variants and de-identified reports</td>
<td>GenelInsight de-identified case repository, DNA nexus Commons</td>
</tr>
<tr>
<td></td>
<td>community data sharing</td>
<td>dbGaP and ClinVar submissions</td>
</tr>
</tbody>
</table>
eMERGE III Custom Reporting:

Figure 4. Site-Specific Reportable List of Genes/SNPs for Which Pathogenic or Likely Pathogenic Variants Will Be Reviewed:
(A) Consensus list of variable SNPs/gene. Inclusions are indicated with a green dot and exclusions are indicated by no dot.
(B) Site-specific list. Non-consensus genes/SNPs with site-specific exclusions indicated with a green dot.
(C) Site-specific list. Non-consensus genes/SNPs for inclusion in PIH genes. Inclusions are indicated with a green dot. UW/KP, University of Washington/Kaiser Permanente Washington; CHOP, Children's Hospital of Philadelphia; COHSC, Cincinnati Children’s Hospital Medical Center.
eMERGE III Contributions:

I: Calibrated ‘return rates’,

II: Supported development of automated reporting,

III: Demanded data structure standards: e.g. FHIR
Tools for integrated Clinical Research and Support:

I: DATA LAKE (gene discovery)
- Variant/Genotype Data
- Sample Data
- QC/Tracking Metrics
- Phenotype/EMR
- Annotation/Ontology

II: DATA COMMONS (dashboard, analytics)
- Variant, Annotation, and Phenotype Data

III: NEPTUNE (variant interpretation)
- VIP Curated Variants
- Sample, Local Identifier, PMI/Metadata

IV: PARLIAMENT (SV discovery by Consensus on the Cloud)

V: ARBOR (Report tracking)

Assessing structural variation in a personal genome—towards a human reference diploid genome


**RESEARCH ARTICLE**

Open Access
Driving Adult Genetic Discovery:

A CVD SCREEN PILOT – 5,000 Participants

First conceived by John Belmont >5 years ago

Clinical return, Research consented

BCM the Common SPirit

**HeartCare Panel**

- Blood or saliva collection in clinic
- 158 gene NGS panel + PGx + PRS + LPA
- Results returned to clinician and become part of medical record
- Only actionable genetic changes are reported (Pathogenic or Likely Pathogenic)
- Clinician will discuss test results and any recommended management changes
- Genetic counseling services and referral to adult genetics clinic are available

- Aortopathy
- Arrhythmia
- Cardiomyopathy
- Dyslipidemia
- Pharmacogenetics
- Polygenic Risk Score
- LPA Risk Variants
Baylor College of Medicine
1: Lipid Clinic –
2: Other Cardiology
3: Primary care
Texas Medical Center:
54 Institutions, 7M patient visits
BCM affiliate – Largest US HC System, 22 States
700 Care sites
154 Hospitals
25,000 physicians
~ 25% of lives in the USA
1: “CVD and related disease risk..”

2: “Research and discovery on CVD and other genetic diseases…”,

Consent Language:
**HeartCare Genes (158): 5,000 sample pilot**

- Pharmacogenetics
- Polygenic Risk Score
- LPA Risk Alleles
- Other monogenic risk alleles

<table>
<thead>
<tr>
<th>Arrhythmias</th>
<th>Cardiomyopathies</th>
<th>Aortopathies</th>
<th>Dyslipidemias</th>
<th>Other CV Related</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Long QT syndrome (LQTS)</td>
<td>• Hypertrophic Cardiomyopathy (HCM)</td>
<td>• Connective Tissue Disorders:</td>
<td>• Familial Hypercholesterolemia</td>
<td>• Amyloidosis</td>
</tr>
<tr>
<td>• Short QT syndrome</td>
<td>• Dilated Cardiomyopathy (DCM)</td>
<td>• Ehlers-Danlos syndrome (EDS)</td>
<td>• Familial Combined Hyperlipidemia</td>
<td>• Pompe Disease</td>
</tr>
<tr>
<td>• Brugada syndrome</td>
<td>• Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)</td>
<td>• Marfan syndrome (MFS)</td>
<td>• Tangier disease / Hypoalphalipoproteinemia</td>
<td>• Fabry Disease</td>
</tr>
<tr>
<td>• Wolff-Parkinson-White (WPW) syndrome</td>
<td>• Left ventricular non-compaction</td>
<td>• Loeys-Dietz syndrome (LDS)</td>
<td>• Sitosteroledema</td>
<td>• Pulmonary</td>
</tr>
<tr>
<td>• Catecholaminergic polymorphic ventricular tachycardia (CPVT) Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)</td>
<td></td>
<td>• Thoracic Aortic Aneurysm and Dissection (TAAD)</td>
<td>• Abetalipoproteinemia</td>
<td>• Arterial Hypertension</td>
</tr>
<tr>
<td>• Catecholaminergic polymorphic ventricular tachycardia (CPVT) Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)</td>
<td></td>
<td></td>
<td></td>
<td>• Muscular Dystrophy (DMD)</td>
</tr>
</tbody>
</table>
Familial Hypercholesterolemia (FH)
- 6 (3.8%) LDLR, ApoA mutations

Other High Penetrance Alleles
- 7 (4.4%) MYH7, MYBPC3..etc

LPA Risk Variants-
- 34 (14.7%),
- LP(a) measurement recommended

Polygenic Risk Score
- 15 (6.5%) in top 5% risk for CAD

Pharmacogenetics
- 87 (38%), simvastatin and warfarin
  - All consent for ‘general research’,
HeartCare: Priming the Pump for WGS and Full Deployment

RESEARCH

Clinical Tests

>20% of lives in USA (DIGNITY + CHI)

Texas Medical Center (7M per year)
CONCLUSIONS

• Mendelian discovery –
  • Illustrates the ‘Virtuous Cycle’,

• Adult CVD risk is our gateway to large scale adult genome testing in the clinic,

Venue for deployment begins in Academic Center then spreads to large HCS,
ACKNOWLEDGEMENTS

Kimberly Walker
Christie Kovar
Viktoriya Korchina
Ritika Raj

Jessica de la Cruz
Doreen Ng

Nara Sobreira
Ada Hamosh
Allan Scott
Kim Doheny

FUNDING:
NIH/NHGRI/NHLBI

NHGRI/NCI U01 HG006485
NHGRI U54 HG006542
NHGRI U54 HG003273
NINDS R01 NS058529
NIH/NIGMS T32 GM07526