Sharing Data within ATHENA Breast Health Network WISDOM study
Olivier Harismendy, PhD
The ATHENA Breast Health Network

• Established network with a large community referral base
  • 5 University of California Medical Campuses
  • 13 Midwest hospitals (Sanford Health)
• >100 providers and researchers in the fields of primary care, radiology, pathology, surgery, medical oncology and radiation oncology
• >95,000 women enrolled to date

UC Davis
UC Irvine
UC Los Angeles
UC San Diego
UC San Francisco
ATHENA Strategic Initiatives

Prevention

Risk assessment as part of screening

Screening and Diagnosis

Risk based screening and prevention

Fewer false positives

Treatment

Tailored interventions based on molecular profiling

Prevention counseling for high risk women

Survivorship

Better models of care

Tailored frequency

Treatment Screening and Diagnosis

Adaptive learning in practice
The ATHENA Breast Health Network

PATIENTS AND PARTICIPATING SITES

SCREENING 150,000
BIOPSIES 40,000
DIAGNOSIS 20,000
CANCERS 10,000-12,000

DATA IN
- Electronic Patient Questionnaires
- Automated Risk Assessments
- Risk Models/Web Services
- Molecular Tests
- Disease Staging
- Treatment Decisions
- Outcomes

KNOWLEDGE OUT
- Personalized Risk Profile
- Personalized Biopsy Benefit
- Personalized Breast Cancer Treatment Options
- Options for Risk-based Trials
- Connection to BreastCancerTrials.org

RESEARCH/QUALITY IMPROVEMENT
- Biomarker Discovery
- Biomarker Validation
- Biospecimen Repository
- Comparative Effectiveness Research
- Strategies for Personalized Medicine
- Evidence-based Management
ATHENA Goals

- Personalize breast cancer risk
  - by assessing and communicating individuals’ breast cancer risk to their clinicians

- Build better predictive risk models
  - to drive the development of tailored screening and prevention strategies

- Identify women and families at high risk for hereditary cancer
  - offer them consultations with an Athena Breast Health Specialist, genetic counseling, gene tests, and breast cancer risk reduction and prevention interventions

- Identify low-risk patients at the time of diagnosis
  - offer more tailored treatment options, thereby reducing unnecessary morbidity from overtreatment

- Identify patients at high-risk for recurrence and death
  - based on their tumor biology, response to therapy, comorbidities, and lifestyle factors as well as developing and offering tailored interventions to those high-risk patients to reduce mortality

- Create a repository of clinical and biological data and specimens and a set of analytic tools
  - to conduct comparative effectiveness studies that will provide feedback and opportunities to accelerate learning and continuously improve treatment and intervention options
Athena Key Accomplishments

- Patient Reported Data Standardization at point of clinical care (Breast Imaging and Breast Health Center)
  - Collaboration with Salesforce (Health Questionnaire System)
  - Electronic medical record integration
    - Bi-directional data exchange with the Athena platform and several UC/Sanford Health site’s electronic medical record system
- Implementation of personalized breast cancer risk assessment as a part of clinical care screening mammography
  - Women who are found to be at high risk of developing breast cancer are contacted by an Athena Breast Health Specialist/Genetic Counselor to discuss risk reduction strategies and referrals
Athena Key Accomplishments

• Awarded Patient Centered Outcomes Research Institute (PCORI) grant for a personalized breast screening trial (Wisdom Study)

• Awarded NIH Molecular and Cellular Characterization (MCL) of Screen-Detected Lesions grant to study over-diagnosis and over treatment in breast cancer.
Screening Controversy

Some women screened too much, others too little

Women are caught in the middle...

and some are choosing not to screen at all
# Screening Controversy

## Comparison of Breast Cancer Screening Guidelines (January 2016)

<table>
<thead>
<tr>
<th>Recommended</th>
<th>ACOG</th>
<th>ACR/SBI</th>
<th>ACS</th>
<th>AMA</th>
<th>NCCN</th>
<th>USPSTF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age to Start Mammograms</td>
<td>40</td>
<td>40</td>
<td>45 Individual choice 40-44</td>
<td>40</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Age to Stop Mammograms</td>
<td>Annual as long as woman is in good health</td>
<td>When life expectancy is &lt;5-7 years</td>
<td>When life expectancy &lt;10 years</td>
<td>When life expectancy &lt;10 years</td>
<td>Upper age limit not established</td>
<td>74</td>
</tr>
<tr>
<td>Interval</td>
<td>Annual</td>
<td>Annual</td>
<td>Annual 45-54; 1-2 years 55+</td>
<td>Annual</td>
<td>Annual</td>
<td>2 years</td>
</tr>
<tr>
<td>Tomo-synthesis (3-D Mammography)</td>
<td>Further study to confirm whether cost-effective replacement for digital mammography alone as first-line screening</td>
<td>No longer investigational; represents an advance in breast imaging</td>
<td>Improvement in detection, lower chance of recall</td>
<td>Silent</td>
<td>Promising; definitive studies pending</td>
<td>Insufficient evidence to support routine use; grade “I”</td>
</tr>
<tr>
<td>Notes</td>
<td>Tomosynthesis shown to improve key screening parameters compared to digital mammography</td>
<td>40-44 Opportunity to begin screening; 45-54 Annual exam; 55+ 1-2 years Transition to biennial or opportunity for annual exam</td>
<td>Eligible at age 40, if they choose and their doctors agree; annual at 50</td>
<td></td>
<td>40-49 Grade “C” Individual decision; 50-74 Grade “B” biennial screening; 75+ Grade “I” Insufficient Evidence</td>
<td></td>
</tr>
</tbody>
</table>
Towards Tailored Screening

• Screening: one size does not fit all
  - Recognition that screening needs to reflect susceptibility

• Risk factors
  - family pedigree with breast cancer
  - hormonal factors (age of first childbirth, age of menarche, age of menopause etc)
  - breast density
  - atypical biopsy
  - gene carriers
    - high risk: BRCA1/BRCA2 to low risk variants: SNPs
Rationale for the WISDOM Study

• The only way to solve the screening debate is to conduct a randomized clinical trial to test the two approaches to screening (annual vs. risk-based)

• WISDOM ➔ Women Informed to Screen Depending On Measures of risk

\[\text{Wisdom} \text{hena Breast Health Network has received a $14M award from the Patient Centered Outcomes Research Institute (PCORI) to study this question to determine the best way forward for breast cancer screening.}\]
WISDOM Study Aims

Determine if personalized screening (as compared to annual screening):

1. **Is as safe**
   - Minimal or no increase in > stage 2B (node positive)
   - No increase in the rate of systemic therapy

1. **Is readily accepted**
   - Greater choice of personalized over annual screening in the self-assigned cohort;
   - Willingness to be randomized, greater adherence to recommended screening;
   - No overall increase in anxiety in the personalized screening arm;
   - No decisional regret

1. **Is less morbid**
   - Fewer recalls and biopsies;
   - Less low grade DCIS (less over-diagnosis)

1. **Enables prevention** as measured by
   - Greater uptake of risk reducing interventions
Personalized Screening Arm

Risk Calculator

- Mammography
  - Breast density
- Athena Health Questionnaire
  - Family history, comorbidities, previous biopsies, age, race/ethnicity
- Genomic profiling
  - BRCA, comprehensive hereditary breast cancer risk gene panel, SNPs
  - Saliva collection

Personalized Risk Profile:
Risk Assignment notification, assigned screening frequency

- Highest risk
- Elevated risk
- Average risk
- Lowest risk

Breast Health Specialist Counseling

Follow-Up:
- Mammography Frequency Assigned by Risk Profile
- Annual Athena Questionnaire to re-assess risk

No cancer: repeat
Cancer detected: Molecular profiling

risk calculator: BCSC model, Jeff Tice et al; USPSTF
Breast Cancer Genetics

- **Very rare variants**
  - Large effect size

- **Rare/uncommon variants**
  - Moderate effect size
  - (for example, BRCA1/2, HOXB13*, NBS1, CHEK2, PALB2)

- **Common variants**
  - Small effect size
  - (GWAS hits)

- **Allele frequency**
  - Rare
  - Uncommon
  - Common
BRCA1 & BRCA2

- ~One in 400-800 women have a mutation
- ~One in 40 women with Ashkenazi Jewish heritage
- Many have family history of breast, ovarian, or other cancers... but not all!
  - Adopted women
  - Family history not known
  - Passed through males
  - Non-penetrant families
# Wisdom Estimated Carrier Frequency

<table>
<thead>
<tr>
<th>Gene</th>
<th>Frequency Low end</th>
<th>Frequency High end</th>
<th>Women tested</th>
<th>Estimated Number (low end)</th>
<th>Estimated Number (high end)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>BRCA1</em></td>
<td>0.0005</td>
<td>0.002</td>
<td>50,000</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td><em>BRCA2</em></td>
<td>0.0005</td>
<td>0.002</td>
<td>50,000</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td><em>TP53</em></td>
<td>0.0001</td>
<td>0.0001</td>
<td>50,000</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><em>STK11</em></td>
<td>0.0001</td>
<td>0.0001</td>
<td>50,000</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><em>CDH1</em></td>
<td>0.0001</td>
<td>0.0001</td>
<td>50,000</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><em>PTEN</em></td>
<td>0.0001</td>
<td>0.0001</td>
<td>50,000</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><em>ATM</em></td>
<td>0.001</td>
<td>0.001</td>
<td>50,000</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td><em>CHEK2</em></td>
<td>0.01</td>
<td>0.01</td>
<td>50,000</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td><em>PALB2</em></td>
<td>0.01</td>
<td>0.01</td>
<td>50,000</td>
<td>500</td>
<td>500</td>
</tr>
</tbody>
</table>

Total Women Tested: 1,120
Total Estimated Numbers: 1,270
Polygenic risk score

- 77 GWAS SNPs
- 33k cases, 33k controls
  - Women with the highest 1% of scores were 3 times more likely to develop breast cancer
  - Women with the lowest 1% of scores had a 70% lower risk of developing breast cancer

*Mavaddat et al (2015)*
The WISDOM Study

- Eligible Patients
  - Consent
    - Randomized Cohort
      - Randomize
        - Annual Screening
        - Personalized Screening
    - Observational Cohort
      - Annual Screening
      - Personalized Screening

Adapts over time
Current general population guidelines

Are you 40? Mammogram q.1yr

Are you 40-49? You & your doctor discuss

Are you 50-74? Mammogram q.2yr
Current NCCN guidelines for BRCA1/2

Breast cancer risk management (Wisdom)

- Refer to high-risk specialty clinics with expertise
- Clinical breast exam q 6-12 mos.
- Mammogram + MRI (3 studies with consistent evidence of lower stage at diagnosis)
- Discuss option of risk-reducing mastectomies
- Consider risk-reducing medications

3e 75, “personalize”
The Personalized Arm: Better Buckets

Age 40-49 with a <1.3% 5-year risk
Repeat mammo at age 50

Age ≥50 *OR* >1.3% 5-year risk
Biennial mammo

Age 40-49 with ext. dense breasts
Annual mammo

Genetic test + OR >6% 5-year risk
Annual mammo + MRI
What testing is being done?

9 gene panel:
- BRCA1, BRCA2
- CDH1, PTEN, STK11, TP53
- ATM, CHEK2, PALB2

SNP score:
- 76 SNPs
Color Genomics Data

- ‘Result Report’ w pathogenic/likely pathogenic mutations in 9 genes
- Table w all variants per gene pathogenic/likely pathogenic/VUS/likely benign/benign
- Table with all SNP allele calls (major/minor alleles)
- VCF files of reportable range of sequenced genes and SNP regions (‘Result Data’)
- BAM file size: 42MB (21-152MB)
- VCF file size: 28KB (22-33KB)
Why trust a private cloud?

Customer Data
- Platform, Applications, Identity & Access Management
- Operating System, Network & Firewall Configuration
  - Client-side Data Encryption & Data Integrity Authentication
  - Server-side Encryption (File System and/or Data)
  - Network Traffic Protection (Encryption / Integrity / Identity)

Customer
- Responsible for security ‘in’ the Cloud

AWS
- Responsible for security ‘of’ the Cloud

AWS Global Infrastructure
- Regions
- Availability Zones
- Edge Locations

Compute
- Storage
- Database
- Networking
3 computation tiers
3 storage tiers
10GbE throughout
Full redundancy
RSA Two Factor Auth.
Remote data replication

1000+ cores
9TB+ RAM
1PB+ storage
Support for containers

Running Docker within Linux VM:
- Flexibility of the applications (bundle necessary libraries, retain provenance)
- Improve efficiency/ scalability/ economics / security of the cloud
Genomic Virtual Machines

Images

Best Practice Pipelines
• Germline variant calling
• Cancer variant calling
• Structural variant calling
• RNA-seq
• smallRNA-seq
• ChIP-seq

Li Ding, Jay Mashl - WashU
Brad Chapman- Harvard
Blue Collar Bioinformatics

B. Gruening - Galaxy
GA4GH API

Interoperability: One API, Many Apps

- Genome Browser
- Command-line Interface
- MapReduce Wrapper
- Repository (NCBI)
- Local Repository
- Repository (Google)
- Repository (EBI)
## Mining DNA Variants

<table>
<thead>
<tr>
<th></th>
<th>CS₁</th>
<th>CS₂</th>
<th>CS₃</th>
<th>CS₄</th>
<th>CS₅</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>V₁</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>...</td>
</tr>
<tr>
<td>V₂</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>V₃</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>V₄</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>V₅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **GAVariantSet**: Represents the set of variants.
- **GAVariant**: Represents a single variant.
- **GACallSet**: Represents the set of calls.
- **GACall**: Represents a single call.
<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>getReference</td>
<td>Gets a GAReference by ID. GET /references/{id} will return a JSON version of GAReference.</td>
</tr>
<tr>
<td>getReferenceBases</td>
<td>Lists GAReference bases by ID and optional range. GET /references/{id}/bases will return a JSON version of GAListReferenceBasesResponse.</td>
</tr>
<tr>
<td>getReferenceSet</td>
<td>Gets a GAREferenceSet by ID. GET /referencesets/{id} will return a JSON version of GAREferenceSet.</td>
</tr>
<tr>
<td>searchCallSets</td>
<td>Gets a list of GACallSet matching the search criteria. &lt;br&gt;POST /callsets/search must accept a JSON version of GASearchCallSetsRequest as the post body and will return a JSON version of GASearchCallSetsResponse.</td>
</tr>
<tr>
<td>searchReadGroupSets</td>
<td>Gets a list of GAReadGroupSet matching the search criteria. &lt;br&gt;POST /readgroupsets/search must accept a JSON version of GASearchReadGroupSetsRequest as the post body and will return a JSON version of GASearchReadGroupSetsResponse.</td>
</tr>
<tr>
<td>searchReads</td>
<td>Gets a list of GAReadAlignment matching the search criteria. &lt;br&gt;POST /reads/search must accept a JSON version of GASearchReadsRequest as the post body and will return a JSON version of GASearchReadsResponse.</td>
</tr>
<tr>
<td>searchReferenceSets</td>
<td>Gets a list of GAREferenceSet matching the search criteria. &lt;br&gt;POST /referencesets/search must accept a JSON version of GASearchReferenceSetsRequest as the post body and will return a JSON version of GASearchReferenceSetsResponse.</td>
</tr>
<tr>
<td>searchReferences</td>
<td>Gets a list of GAREference matching the search criteria. &lt;br&gt;POST /references/search must accept a JSON version of GASearchReferencesRequest as the post body and will return a JSON version of GASearchReferencesResponse.</td>
</tr>
<tr>
<td>searchVariantSets</td>
<td>Gets a list of GAVariantSet matching the search criteria. &lt;br&gt;POST /variantsets/search must accept a JSON version of GASearchVariantSetsRequest as the post body and will return a JSON version of GASearchVariantSetsResponse.</td>
</tr>
<tr>
<td>searchVariants</td>
<td>Gets a list of GAVariant matching the search criteria. &lt;br&gt;POST /variants/search must accept a JSON version of GASearchVariantsRequest as the post body and will return a JSON version of GASearchVariantsResponse.</td>
</tr>
</tbody>
</table>
API Reference

This API reference is organized by resource type as described in Google Genomics fundamentals. Each resource type has one or more data representations and one or more methods. The API Discovery file also lists all methods and resources.

This API provides an implementation of the Global Alliance for Genomics and Health v0.5.1 API as well as many other extensions.

Resource types

- AnnotationSets
- Annotations
- Callsets
- Datasets
- Jobs
- Readgroupsets
- Readgroupsets.coveragebuckets
- Reads
- References
- References.bases
- ReferenceSets
- Variants
- VariantSets
Interactions Overview
Security

Data User Agreement

Provides credentials (VPN, RSASecureID)

Data User & Contributor Agreement

UC San Diego School of Medicine

UCSF

Amends IRB

Global Alliance for Genomics & Health

UNIVERSITY OF CALIFORNIA SANTA CRUZ

Wisdom

color
Server Installation

GA4GH API development

GA4GH server Installation / maintenance

Secure queries
In progress

- Adding annotation directly to VCF files.
  - Current color VCF only contain DNA level info
  - GA4GH supports VEP and SNPEff format

- Building a web-query system using the API
  - “Which patients have loss of function variants (nonsense, frameshift, splice site)?”
  - “Which variants are considered pathogenic or likely pathogenic in the ClinVar database?”
  - “Which variants are new, absent from public databases (ExAC, ClinVar, dbSNP)?”
  - Scaling up to allow >60k patients
The vision
Acknowledgments

UC San Diego

- Lucila Ohno-Machado
- Barbara Parker
- Claudiu Fracas
- Antonios Koures
- Jihoon Kim
- Lisa Madlensky
- Tracy Layton

UC San Francisco

- Laura Esserman