ASHG 2015 Meeting Report

Yonglan Zheng

Olopade Lab Meeting

Oct 27th, 2015
<table>
<thead>
<tr>
<th>Session Topic/Title</th>
<th>Abstract/Poster Board Numbers</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start #</td>
<td>End #</td>
</tr>
<tr>
<td>Bioinformatics and Genomic Technology</td>
<td>1588</td>
<td>1917</td>
</tr>
<tr>
<td>Cancer Genetics</td>
<td>2567</td>
<td>2794</td>
</tr>
<tr>
<td>Cardiovascular Genetics</td>
<td>508</td>
<td>647</td>
</tr>
<tr>
<td>Clinical Genetics and Dysmorphology</td>
<td>2344</td>
<td>2489</td>
</tr>
<tr>
<td>Clinical Genetic Testing</td>
<td>1918</td>
<td>2099</td>
</tr>
<tr>
<td>Complex Traits and Polygenic Disorders</td>
<td>704</td>
<td>1021</td>
</tr>
<tr>
<td>Cytogenetics</td>
<td>2490</td>
<td>2538</td>
</tr>
<tr>
<td>Development</td>
<td>2539</td>
<td>2566</td>
</tr>
<tr>
<td>Epigenetics</td>
<td>391</td>
<td>467</td>
</tr>
<tr>
<td>Ethical, Legal, Social, and Policy Issues in Genetics</td>
<td>2155</td>
<td>2195</td>
</tr>
<tr>
<td>Evolutionary and Population Genetics</td>
<td>1447</td>
<td>1587</td>
</tr>
<tr>
<td>Genetic Counseling</td>
<td>2196</td>
<td>2215</td>
</tr>
<tr>
<td>Genetics/Genomics Education</td>
<td>2216</td>
<td>2241</td>
</tr>
<tr>
<td>Genome Structure, Variation, and Function</td>
<td>3026</td>
<td>3193</td>
</tr>
<tr>
<td>Health Services Research</td>
<td>2242</td>
<td>2262</td>
</tr>
<tr>
<td>Metabolic Disorders</td>
<td>2263</td>
<td>2343</td>
</tr>
<tr>
<td>Molecular Basis of Mendelian Disorders</td>
<td>2795</td>
<td>3025</td>
</tr>
<tr>
<td>Pharmacogenetics</td>
<td>648</td>
<td>703</td>
</tr>
<tr>
<td>Prenatal, Perinatal, and Reproductive Genetics</td>
<td>2100</td>
<td>2154</td>
</tr>
<tr>
<td>Psychiatric Genetics, Neurogenetics, and Neurodegeneration</td>
<td>1022</td>
<td>1253</td>
</tr>
<tr>
<td>Statistical Genetics and Genetic Epidemiology</td>
<td>1254</td>
<td>1446</td>
</tr>
<tr>
<td>Therapy for Genetic Disorders</td>
<td>468</td>
<td>507</td>
</tr>
</tbody>
</table>
http://www.ashg.org/2015meeting/

- The U. S. Precision Medicine Initiative. **Francis Collins.** NIH, Bethesda.
- Personalized medicine, an epidemiological perspective. **David Hunter.** Harvard T.H. Chan Sch Publ Hlth, Boston.
- Studying genetic variation in large cohorts. **Naomi Wray.** Univ Queensland, Australia.
Cancer Genetics in the Genomics Era
• Computational approaches for large-scale cancer genetics

Statistical Genetics: Complex Phenotypes, Complex Solutions
• Pitfalls in development of statistical methods for rare variant association studies

Hereditary Cancer Genes: Old and New
• Germline mutations in cancer predisposition genes in 1,120 children with cancer: A report from the Pediatric Cancer Genome Project
• Somatic TP53 mutations detected in germline testing: The importance of phenotypic correlation in cancer predisposition testing

Human-wide Association Studies: More Genotypes, More Phenotypes, More Diverse Populations
• Whole genome sequencing increase the power to detect trait-associated rare variants shifted towards high frequencies in the Sardinian

Decoding Variants in Coding Regions
• Phased annotation of protein-coding variants across 60,706 human exomes
• Assessing the pathogenicity of insertion and deletion variants with the Variant Effect

ASHG Awards -- Friday
Opening Up Big Data -- Saturday
Gruber Genetics Prize Award
Presentation and Rosalind Franklin Award -- CRISPR-Cas9 Genome Editing: Origins and Development of a Revolutionary Technology

Emmanuelle Charpentier, PhD, Helmholtz Centre for Infection Research in Braunschweig, Germany

ASHG William Allan Award -- Duchenne Muscular Dystrophy (DMD)

Kay E. Davies, DPhil
Dr. Lee's Professor of Anatomy, Department of Physiology, Anatomy and Genetics, and Associate Head of the Medical Sciences Division Director of the Medical Research Council Functional Genomics Unit, University of Oxford, UK

ASHG Curt Stern Award

Jennifer Doudna, PhD, University of California, Berkeley, CA

Leonid Kruglyak, PhD
Professor of Human Genetics and Professor of Biological Chemistry, University of California, Los Angeles Investigator, Howard Hughes Medical Institute
Computational approaches for large-scale cancer genetics
Computational approaches for large-scale cancer genetics

(B) Model of clonal architecture and tumor evolution, inferred from the original ~30x sequencing data.

(C) Ultra-deep sequencing and validation, revealing additional subclonal complexity.

(D) Incorporating the results of single-cell sequencing and intermediate time points to allow refinements to the model, including establishing an independent origin for the TP53 mutant clonal population.

Optimizing Cancer Genome Sequencing and Analysis. Cell Systems, 2015
Computational approaches for large-scale cancer genetics

By WGS of this tumor to a depth of greater than 300x and applying comprehensive analysis strategies:

(1) the commonly used sequencing depths of ~100x for exomes and 30x–50x for WGS are inadequate when there exists even moderate heterogeneity, impurity, contamination, aneuploidy, or combinations thereof;

(2) current analysis strategies relying on a single alignment algorithm and variant caller, using data from a single DNA library, suffer from poor sensitivity and specificity;

(3) deep sequencing substantially improves the discovery of variants across a range of VAFs and allows for a more definitive model of tumor clonal architecture; and

(4) this dataset, generated from the most deeply sequenced and cross-validated tumor described to date, will serve as a valuable community resource for improving tools and algorithms by providing an extremely high-confidence set of low-frequency mutations.
Pitfalls in Development of Statistical Methods for Rare Variant Association Studies

Methods Developed vs. Being Used

June 2008 - June 2015: a survey of 91 RVA methods and 24 RVA studies

- SKAT Series (36.7%)
- C-alpha (10%)
- Single Variant Tests (23.3%)
- Variable Threshold (10%)
- CMC / BRV / GRANVIL (20%)

Methods applied in more than two studies
# Pitfalls in Development of Statistical Methods for Rare Variant Association Studies

## Table 2. Summary of Statistical Methods for Rare-Variant Association Testing

<table>
<thead>
<tr>
<th>Description</th>
<th>Methods</th>
<th>Advantage</th>
<th>Disadvantage</th>
<th>Software Packages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burden tests</td>
<td>ARIEL test, CAST, CMC method, MZ test, WSS</td>
<td>are powerful when a large proportion of variants are causal and effects are in the same direction</td>
<td>lose power in the presence of both trait-increasing and trait-decreasing variants</td>
<td>EPACTS, GRANVIL, PLINK/SEQ, Rvttests, SCORE-Seq, SKAT, VAT</td>
</tr>
<tr>
<td>Adaptive burden tests</td>
<td>aSum, Step-up, EREC test, VT, KBAC method, RBT</td>
<td>are more robust than burden tests using fixed weights or thresholds; some tests can improve result interpretation</td>
<td>are often computationally intensive; VT requires the same assumptions as burden tests</td>
<td>EPACTS, KBAC, PLINK/SEQ, Rvttests, SCORE-Seq, VAT</td>
</tr>
<tr>
<td>Variance-component tests</td>
<td>SKAT, SSU test, C-alpha test</td>
<td>are powerful in the presence of both trait-increasing and trait-decreasing variants or a small fraction of causal variants</td>
<td>are less powerful than burden tests when most variants are causal and effects are in the same direction</td>
<td>EPACTS, PLINK/SEQ, SCORE-Seq, SKAT, VAT</td>
</tr>
<tr>
<td>Combined tests</td>
<td>SKAT-O, Fisher method, MIST</td>
<td>are more robust with respect to the percentage of causal variants and the presence of both trait-increasing and trait-decreasing variants</td>
<td>can be slightly less powerful than burden or variance-component tests if their assumptions are largely held; some methods (e.g., the Fisher method) are computationally intensive</td>
<td>EPACTS, PLINK/SEQ, MIST, SKAT</td>
</tr>
<tr>
<td>EC test</td>
<td>exponentially combines score statistics</td>
<td>is powerful when a very small proportion of variants are causal</td>
<td>is computationally intensive; is less powerful when a moderate or large proportion of variants are causal</td>
<td>no software is available yet</td>
</tr>
</tbody>
</table>

Abbreviations are as follows: ARIEL, accumulation of rare variants integrated and extended locus-specific; aSum, data-adaptive sum test; CAST, cohort allelic sums test; CMC, combined multivariate and collapsing; EC, exponential combination; EPACTS, efficient and parallelizable association container toolbox; EREC, estimated regression coefficient; GRANVIL, gene- or region-based analysis of variants of intermediate and low frequency; KBAC, kernel-based adaptive cluster; MIST, mixed-effects score test for continuous outcomes; MZ, Mom’s and Zeggini; RBT, replication-based test; Rvttests, rare-variant tests; SKAT, sequence kernel association test; SSU, sum of squared score; VAT, variant association tools; VT, variable threshold; and WSS, weighted sum statistic.

*More information is given in Table 3.

PMID: 24995866
Pitfalls in Development of Statistical Methods for Rare Variant Association Studies

Pitfall 1: Misuse of genetic data in simulation studies
   Cautions with site frequency spectra
   Cautions in sample size estimation
Pitfall 2: Improper “head start” in power comparisons
Pitfall 3: Overreliance on omnibus tests

“The well-worn dictum “all models are wrong, some models are useful” applies with a vengeance in this context” “the more appropriate uses of cell-line pharmacological data are for hypothesis generation and for elaborating on existing hypotheses, rather than for formal statistical prediction. Clues obtained by correlating drug-sensitivity patterns with molecular profiles will sometimes illuminate cellular mechanisms and pathways that advance our basic understanding, even if they are not directly predictive for the clinic.” -- Inconsistency in large pharmacogenomic studies. Nature 2013
RVA methods overemphasize association but underrate architecture
• Phenotypic assumptions motivate methods development
• Heterogeneity in genetic architecture is little quantified

“Our results imply that tens of thousands of individuals, extensive functional annotation, or highly targeted hypothesis testing will be required to confidently detect or exclude rare variant signals at complex disease loci.” -- The Power of Gene-Based Rare Variant Methods to Detect Disease-Associated Variation and Test Hypotheses About Complex Disease. PLoS Genetics 2015.

**Practical Recommendations**

**Method development**
• Use benchmark data: save your effort
• Publish computational environment: save other’s effort
• Release usable software: or the method will go unused!

**Association studies**
• Deploy all possible tests and scrutinize the architecture
Germline mutations in cancer predisposition genes in 1,120 children with cancer: A report from the Pediatric Cancer Genome Project

Pediatric Cancer Genome Project, is a $65 million project that began in 2010 with the aim of sequencing tumor/normal genome pairs of pediatric cancer patients. St. Jude collaborated with the Genome Institute at Washington University.

The clinical pipeline will include WGS, WES, and RNA-seq of the patient's tumor genome to look for somatic alterations, as well as WGS and WES of normal tissue to look for germline mutations. The researchers will return data on somatic alterations from 567 cancer-related genes as well as on germline mutations from 31 cancer predisposition genes and 22 non-cancer disease genes that will be put into patients' electronic medical records.
Germline mutations in cancer predisposition genes in 1,120 children with cancer: A report from the Pediatric Cancer Genome Project

The development of the clinical sequencing pipeline was completed in February 2013. It included whole-genome sequencing of 700 matched tumor/normal pairs as well as 2,000 exomes from 21 different types of pediatric cancers.

The second phase of the Pediatric Cancer Genome Project is a $30 million, two-year initiative. It has completed an 80-patient pilot study to validate the tumor sequencing protocol as well as a 1,120-patient (700 WGS and 420 WES) pilot study to validate the germline analysis.
Somatic TP53 mutations detected in germline testing: The importance of phenotypic correlation in cancer predisposition testing

Li-Fraumeni syndrome, named after both Dr. Li and his colleague Dr. Joseph Fraumeni.

Inherited (germline) p53 gene mutations led to this syndrome.

Frederick Pei Li (May 7, 1940 – June 12, 2015)
Somatic TP53 mutations detected in germline testing: The importance of phenotypic correlation in cancer predisposition testing

Analysis of DNA isolated from peripheral blood/saliva is typically used for diagnosis of hereditary cancer predisposition. Results from testing are accepted as representing a patient’s germline, as acquired somatic mutations are rare. While somatic TP53 mutations are detected in multiple cancer types, germline mutations are exceedingly rare and result in Li-Fraumeni Syndrome (LFS).

This study evaluated whether somatic interference / somatic mosaicism may be more common in genetic testing than previously anticipated. Among patients with pathogenic TP53 results from multi-gene panel testing, cases were selected with potentially abnormal NGS metrics, including decreased ratio of mutant to wild-type allele, >2 detected alleles or haplotypes, or large copy-number alterations. Clinical data was obtained from test requisition forms and compared to LFS testing criteria.
Somatic TP53 mutations detected in germline testing: The importance of phenotypic correlation in cancer predisposition testing

Among 166 TP53 positive cases, 25 were identified as higher risk for somatic interference. None of these families met LFS criteria. To date testing additional tissues confirmed somatic origin for 4/25 cases; two were subsequently diagnosed with a hematologic disorder.

We suggest that somatic TP53 mutations in blood/saliva may be more common than previously thought. Beyond using NGS quality control measures, clinician recognition of test results inconsistent with a LFS phenotype should create an index of suspicion, and caution is urged in the medical management of patients in whom the only criterion for LFS is a TP53 mutation.
Whole genome sequencing increase the power to detect trait-associated rare variants shifted towards high frequencies in the Sardinian

SardiNIA project
Sardinia is the second largest island in the Mediterranean. Its modern population numbers approximately 1.65 million and constitutes a genetically isolated founder population, which has already aided in the identification of genes involved in several Mendelian disorders

Project started in 2001; 151 papers as of Oct 27th, 2015
1st publication @PLoS Genetics, Aug 2006
3 papers @Nature Genetics, Sept 2015; 3 traits: lipid and blood inflammatory markers, hemoglobin levels, short stature
2,120 Sardinians; WGS + genotyping
“The entire computational process required approximately 20 years of computing time (6 CPU years for quality control and alignment and 14 CPU years for variant discovery and genotyping).”
Whole genome sequencing increase the power to detect trait-associated rare variants shifted towards high frequencies in the Sardinian

Other large-scale population genetics (germline)

**The UK10K project**  
Main paper, Nature, Sept 2015  
Imputation, Nature Communications, Sept 2015  
WGS (low-pass, 7×) or WES (80×)

**1000 Genomes Project Phase3**  
2,504 individuals from 26 populations; low-coverage WGS, deep WES, and dense microarray genotyping"
Phased annotation of protein-coding variants across 60,706 human exomes

Daniel MacArthur @Broad Institute, Harvard Medical School, Massachusetts General Hospital
ExAC, aggregate and harmonize exome sequencing data from a variety of large-scale sequencing projects
Picard -> BWA -> GATK
5 talks and 2 posters @ASHG2015

Assessing the pathogenicity of insertion and deletion variants with the Variant Effect

Rachel Karchin @Johns Hopkins
VEST-indel, Human Mutation, Oct 2015
What I feel

• Attractive good tools and platforms – academic groups from the same place put efforts together and make brands
• Well-known labs keep doing great; Companies are doing great too
• Think globally and act locally; Target big but focus and move first
• Numerous consortia, small or large, disease- or locus- specific, or reference; Phenotomes
• “Decreasing market value of disease oriented expertise; Increasing need to collaborate across disciplines; Increasing focus on quality” -- Hem/Onc: State of the Section FY16
• Big Data: the challenge for small groups in the era of cancer genomics; Too much data, too little time & Cultural barriers & Sharing
• “NCI Earmarks $8.5M for Creation of Genomic Data Analysis Centers”
• New NGS techniques; NGS in clinical practice
• Policy, ethic, regulation, education all matter