Genomic Testing: Actionability, Validation, and Standard of Lab Reports

eMERGE: Laura Rasmussen-Torvik

Reaction: Heidi Rehm
Summary: Dick Weinshilboum
Panel: Murray Brilliant, David Carey, John Carpten, Kim Doheny
PROGRESS TO DATE
eMERGE PGx – Overview by Aim

Aim 1: Identify target patients
- Resequence VIP genes; Identify actionable variants
  - Create repository of variants of unknown significance
  - Initiate studies of function and of genotype-phenotype relationships

Aim 2: Actionable variants
- EMR deposit
  - Result display
  - Decision support
  - Outcomes
  - Performance metrics
  - Healthcare impact

Aim 3: Outcomes
- Create repository of variants of unknown significance
- Initiate studies of function and of genotype-phenotype relationships
# Subjects accrued with samples
  = 3841 / 8543 target
  (mixture of sites recruiting de novo, sites recruiting from biobank w/ and w/out clinical samples)

# Sequenced
  = 2450 / 8993 target

# CLIA genotyped (for return)
  = 1396 / 8543 target
PGx platform

• NGS capture reagent

• Genes selected by PGRN community (84 total)
  – Sequence capture = the complete coding regions plus sequence 2 kilobases (kb) up- and 1 kb down-stream to assess variation within nearby regulatory regions
  – also includes known variants present on other commercially available pharmacogenetic panel genotyping platforms, such as Affymetrix’s DMET+ platform and Illumina’s ADME platform
PGx platform

• Batches of 24 (or 48) processed through Illumina flowcell lane

• Excellent results to date:
  – 32 diverse HapMap trios produced an average depth of coverage per sample of 496x
  – genotypes derived from this PGRNseq data were 99.9% concordant with existing SNV data on these samples from the 1000 Genomes project
PGx platform

• Diverse implementation across eMERGE-PGx
  – 7 sites running samples at CIDR
  – 2 sites running samples only at CIDR, other 5 running at 2 locations
  – 1 site using Ion Torrent, others using Illumina HiSeq 2500/2000
Comparing Site Implementation - PGRNSeq

Running PGRNSeq on Site

Returning some results directly from PGRNSeq*
eMERGE PGx Project Summary

Specific Aim 1
- Recruit / Collect Samples
- PGRN-Seq Sequencing
- Clinical Variant Validation

Specific Aim 2
- Return Results: EHRintegration, CDS, Patient & Clinician Education
- Outcomes Measures

Specific Aim 3
- Populate Variant and Phenotype Data Repository (SPHINX)
Clinical validation

• PGRNSeq generally run on research samples

• In eMERGE, generally (but not always)
  – PGRNseq = sequencing = research results
  – CLIA (validation) = genotyping = clinical results
Comparing Site Implementation Details

Drug-Genome pairs study
CYP2C19-Clopidogrel
VKORC1/CYP2C9-Warfarin*
SLCO1B1-Simvastatin

* BCH DGI only VKORC1/CYP2C9-Warfarin
* Geisinger and M/E/PSU also have CYP4F2-Warfarin
Clinical Validation of PGRNSeq research results

- 6 sites validating some samples at JHU DDL (custom Sequenom panel)

- Other sites using Sanger, Illumina ADME, Sequenom ADME

- Many sites validating at more than 1 location, using more than 1 method
PGX STRATEGIES
PGRNSeq calling pipelines / QC—CC

• Cross-site comparison
  – Each site performing sequencing is running 32 HapMap trios along with eMERGE study samples
  – eMERGE-CC is calculating concordance to determine how similar the platform and variant calling is performing across sites

  – Two concordance checks being run
    1. Compare VCF across sites on HapMap trios
    2. Compare VCF on eMERGE study samples generated by sequencing facility and VCF generated by eMERGE-CC pipeline
Cross-Site Comparison - HapMap

<table>
<thead>
<tr>
<th></th>
<th>Raw (%)</th>
<th>Filtered (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Concordance</td>
<td>Discordance</td>
</tr>
<tr>
<td>CIDR vs. Mt. Sinai</td>
<td>98.125</td>
<td>1.436</td>
</tr>
<tr>
<td>CIDR vs. UW</td>
<td>97.859</td>
<td>1.223</td>
</tr>
<tr>
<td>UW vs. Mt. Sinai</td>
<td>98.001</td>
<td>1.215</td>
</tr>
</tbody>
</table>

**Variants**

<table>
<thead>
<tr>
<th></th>
<th>CIDR</th>
<th>Mt. Sinai</th>
<th>UW</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIDR</td>
<td>946361</td>
<td>10154</td>
<td>20312</td>
</tr>
<tr>
<td>Mt. Sinai</td>
<td>1540</td>
<td>937747</td>
<td>15152</td>
</tr>
<tr>
<td>UW</td>
<td>975</td>
<td>4429</td>
<td>927024</td>
</tr>
</tbody>
</table>

At the intersection of 2, it shows the number of filtered SNPs that are in the horizontal, but not in the vertical.
eMERGE Variant Calling Pipeline

- GATK
- All variants kept in VCF, annotated by FILTER status
- Variants filtered under the following:
  - QUAL <= 50 (QualFilter)
  - ABHet > 0.75 (ABFilter)*
  - QD < 5.0 (QDFilter)*

- Performing 2 variant calling runs at different time points
  - Multi-sample calling run on the batch sent from sequencing center for each site independently
  - Multi-sample calling run on the entire eMERGE set quarterly

* ABHet and QD fields not present in completely referent positions.
<table>
<thead>
<tr>
<th>Site</th>
<th>UW</th>
<th>Mayo</th>
<th>Mt. Sinai</th>
<th>Northwestern</th>
<th>CHOP</th>
<th>Marshfield</th>
<th>Vanderbilt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date Rec.</td>
<td>7/31</td>
<td>9/26</td>
<td>10/06</td>
<td>10/31</td>
<td>10/31</td>
<td>11/01</td>
<td>11/05</td>
</tr>
<tr>
<td># Match Variants</td>
<td>10,861</td>
<td>8,625</td>
<td>12,582</td>
<td>7,354</td>
<td>11,814</td>
<td>5,028</td>
<td>6,893</td>
</tr>
<tr>
<td># Filtered Var.</td>
<td>9,389</td>
<td>7,411</td>
<td>10,712</td>
<td>6,453</td>
<td>10,262</td>
<td>4,275</td>
<td>6,014</td>
</tr>
<tr>
<td>Discord (Het. / Hom)</td>
<td>0.211% / 0.023%</td>
<td>0.462% / 0.035%</td>
<td>0.323% / 0.024%</td>
<td>0.539% / 0.029%</td>
<td>0.380% / 0.029%</td>
<td>0.829% / 0.043%</td>
<td>0.638% / 0.038%</td>
</tr>
<tr>
<td>Raw Discrepant</td>
<td>0.003%</td>
<td>0.073%</td>
<td>0.048%</td>
<td>0.054%</td>
<td>0.061%</td>
<td>0.080%</td>
<td>0.060%</td>
</tr>
<tr>
<td>Raw Singleton Discord</td>
<td>0.015% / 0.007%</td>
<td>0.044% / 0.007%</td>
<td>0.050% / 0.007%</td>
<td>0.127% / 0.017%</td>
<td>0.062% / 0.008%</td>
<td>0.224% / 0.028%</td>
<td>0.124% / 0.016%</td>
</tr>
<tr>
<td>Filt. Discord</td>
<td>0 % / 0 %</td>
<td>0.002% / 0%</td>
<td>0.001% / 0%</td>
<td>0.003% / 0%</td>
<td>0.001% / 0%</td>
<td>0.006% / 0%</td>
<td>0.004% / 0%</td>
</tr>
<tr>
<td>Filt. Discrep.</td>
<td>4.826%</td>
<td>6.766%</td>
<td>7.868%</td>
<td>5.661%</td>
<td>6.321%</td>
<td>6.960%</td>
<td>5.751%</td>
</tr>
<tr>
<td>Filt. Singleton Discord</td>
<td>0 % / 0 %</td>
<td>0.001% / 0%</td>
<td>0% / 0%</td>
<td>0.001% / 0%</td>
<td>0% / 0%</td>
<td>0.001% / 0%</td>
<td>0.003% / 0%</td>
</tr>
</tbody>
</table>
## PGRNSeq Concordance - vs. SPHINX

<table>
<thead>
<tr>
<th>Site</th>
<th>UW</th>
<th>Mayo</th>
<th>Mt. Sinai</th>
<th>Northwestern</th>
<th>CHOP</th>
<th>Marshfield</th>
<th>UW</th>
<th>Vanderbilt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date Rec.</td>
<td>7/31</td>
<td>9/26</td>
<td>10/06</td>
<td>10/31</td>
<td>10/31</td>
<td>11/01</td>
<td>10/31</td>
<td>11/05</td>
</tr>
<tr>
<td># Match Variants</td>
<td>10,616</td>
<td>8,558</td>
<td>12,485</td>
<td>7,247</td>
<td>11,760</td>
<td>4,962</td>
<td>12,454</td>
<td>6,830</td>
</tr>
<tr>
<td># Filtered Var.</td>
<td>9,727</td>
<td>7,872</td>
<td>11,528</td>
<td>6,680</td>
<td>10,850</td>
<td>4,521</td>
<td>11,285</td>
<td>6,285</td>
</tr>
<tr>
<td>Discord (Het. / Hom)</td>
<td>0.132% / 0.003%</td>
<td>0.040% / 0.001%</td>
<td>0.042% / 0.001%</td>
<td>0.043% / 0.002%</td>
<td>0.041% / 0.001%</td>
<td>0.046% / 0.002%</td>
<td>0.040% / 0.001%</td>
<td>0.054% / 0.001%</td>
</tr>
<tr>
<td>Raw Discrepant</td>
<td>0.041%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0.001%</td>
</tr>
<tr>
<td>Raw Singleton Discord</td>
<td>0.037% / 0.001%</td>
<td>0.019% / 0.002%</td>
<td>0.011% / 0.002%</td>
<td>0.054% / 0.015%</td>
<td>0.013% / 0.002%</td>
<td>0.096% / 0.011%</td>
<td>0.007% / 0.001%</td>
<td>0.092% / 0.005%</td>
</tr>
<tr>
<td>Filt. Discord</td>
<td>0.008% / 0%</td>
<td>0.003% / 0%</td>
<td>0.003% / 0%</td>
<td>0.003% / 0%</td>
<td>0.003% / 0%</td>
<td>0.003% / 0%</td>
<td>0.003% / 0%</td>
<td>0.002% / 0%</td>
</tr>
<tr>
<td>Filt. Discrep.</td>
<td>1.449%</td>
<td>1.415%</td>
<td>1.526%</td>
<td>1.318%</td>
<td>1.261%</td>
<td>1.219%</td>
<td>2.355%</td>
<td>1.151%</td>
</tr>
<tr>
<td>Filt. Singleton Discord</td>
<td>0% / 0%</td>
<td>0% / 0%</td>
<td>0% / 0%</td>
<td>0% / 0%</td>
<td>0% / 0%</td>
<td>0.002% / 0%</td>
<td>0% / 0%</td>
<td>0% / 0%</td>
</tr>
</tbody>
</table>
Comparison of research and clinical pharmacogenetic results

• To evaluate PGRNSeq (research) platform

• Complicated by different report formats
  • Standardization of reports and comparison methods will benefit the wider community

• Forcing sites to develop policies about non-concordant (really good) research results with clinical genotyping
CLIA genotype results in EHR systems

• Development of systems to integrate genotypes as computed results (EHRI group)

  – How do we integrate and document *clinical interpretation* as part of these systems?
    • This is particularly complicated when receiving results from multiple outside laboratories

  – What do we do if interpretation (i.e. actionability) changes?
Summary

• Genomic testing
  – large scale use and comparison of NGS platform across sites

• Validation
  – comparison of clinical genotyping to research PGRNSeq samples

• Lab reports
  – How to create reports that can be
    • compared to sequencing easily
    • displayed as computed results, AND incorporate interpretation

• Actionability
  – What do we do if/ when interpretation changes
# Cross-Site Comparison - eMERGE

<table>
<thead>
<tr>
<th>eMERGE Site</th>
<th># Samples</th>
<th>First release</th>
<th>Variants called using eMERGE multi-sample calling pipeline</th>
<th>Variant comparison with VCF from sequencing center</th>
<th>Raw Discordance rate (multi-sample calling within site versus site VCF)</th>
<th>Filtered Discordance rate</th>
<th>Raw Discordance with combined release (multi-sample calling within site versus multi-sample calling combined all sites)</th>
<th>Filtered Discordance with Combined release</th>
</tr>
</thead>
<tbody>
<tr>
<td>20130731_uw</td>
<td>322</td>
<td>20131106</td>
<td>9/30/2013</td>
<td>11/5/2013</td>
<td>0.234%</td>
<td>0%</td>
<td>0.135%</td>
<td>0.008%</td>
</tr>
<tr>
<td>20130926_mayo</td>
<td>318</td>
<td>20131106</td>
<td>11/2/2013</td>
<td>11/5/2013</td>
<td>0.497%</td>
<td>0.002%</td>
<td>0.041%</td>
<td>0.003%</td>
</tr>
<tr>
<td>20131009_mtsinai</td>
<td>311</td>
<td>20131106</td>
<td>10/19/2013</td>
<td>11/5/2013</td>
<td>0.357%</td>
<td>0.001%</td>
<td>0.043%</td>
<td>0.003%</td>
</tr>
<tr>
<td>20131031_chop</td>
<td>300</td>
<td>20131106</td>
<td>11/4/2013</td>
<td>11/5/2013</td>
<td>0.409%</td>
<td>0.001%</td>
<td>0.042%</td>
<td>0.003%</td>
</tr>
<tr>
<td>20131031_nw</td>
<td>94</td>
<td>20131106</td>
<td>11/1/2013</td>
<td>11/5/2013</td>
<td>0.568%</td>
<td>0.003%</td>
<td>0.045%</td>
<td>0.003%</td>
</tr>
<tr>
<td>20131031_uw</td>
<td>594</td>
<td>20131106</td>
<td>11/2/2013</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.041%</td>
<td>0.003%</td>
</tr>
<tr>
<td>20131101_marshfield</td>
<td>96</td>
<td>20131106</td>
<td>11/2/2013</td>
<td>11/5/2013</td>
<td>0.872%</td>
<td>0.006%</td>
<td>0.048%</td>
<td>0.003%</td>
</tr>
<tr>
<td>20131105_vanderbilt</td>
<td>84</td>
<td>20131106</td>
<td>11/6/2013</td>
<td>11/6/2013</td>
<td>0.676%</td>
<td>0.004%</td>
<td>0.055%</td>
<td>0.002%</td>
</tr>
</tbody>
</table>
eMERGE PGx QC Details

• Concordance checks
  o Concordance with VCF from sequencing center (typically single-called)
  o Concordance with group-called site vs. combined release

• Inconsistency checks
  o Duplicate study samples and controls called with different IDs
  o All samples renamed to eMERGE or Coriel IDs
  o VCF file checked for inconsistency (same ID, discordant calls)
## eMERGE PGx QC results

<table>
<thead>
<tr>
<th></th>
<th>Raw</th>
<th>Filtered (combined release)</th>
</tr>
</thead>
<tbody>
<tr>
<td># base-pair positions</td>
<td>968,004</td>
<td>925,335</td>
</tr>
<tr>
<td># variants</td>
<td>27,396</td>
<td>29,491</td>
</tr>
<tr>
<td># SNPs</td>
<td>26,994</td>
<td>24,633</td>
</tr>
<tr>
<td># novel variants</td>
<td>12,569</td>
<td>12,189</td>
</tr>
<tr>
<td>Singletons</td>
<td>12,748</td>
<td>12,273</td>
</tr>
<tr>
<td>Doubletons</td>
<td>2,905</td>
<td>2,718</td>
</tr>
<tr>
<td># control inconsistencies</td>
<td>1,818</td>
<td>567</td>
</tr>
<tr>
<td># sample inconsistencies</td>
<td>502</td>
<td>104</td>
</tr>
</tbody>
</table>