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

eMERGE: Laura Rasmussen-Torvik

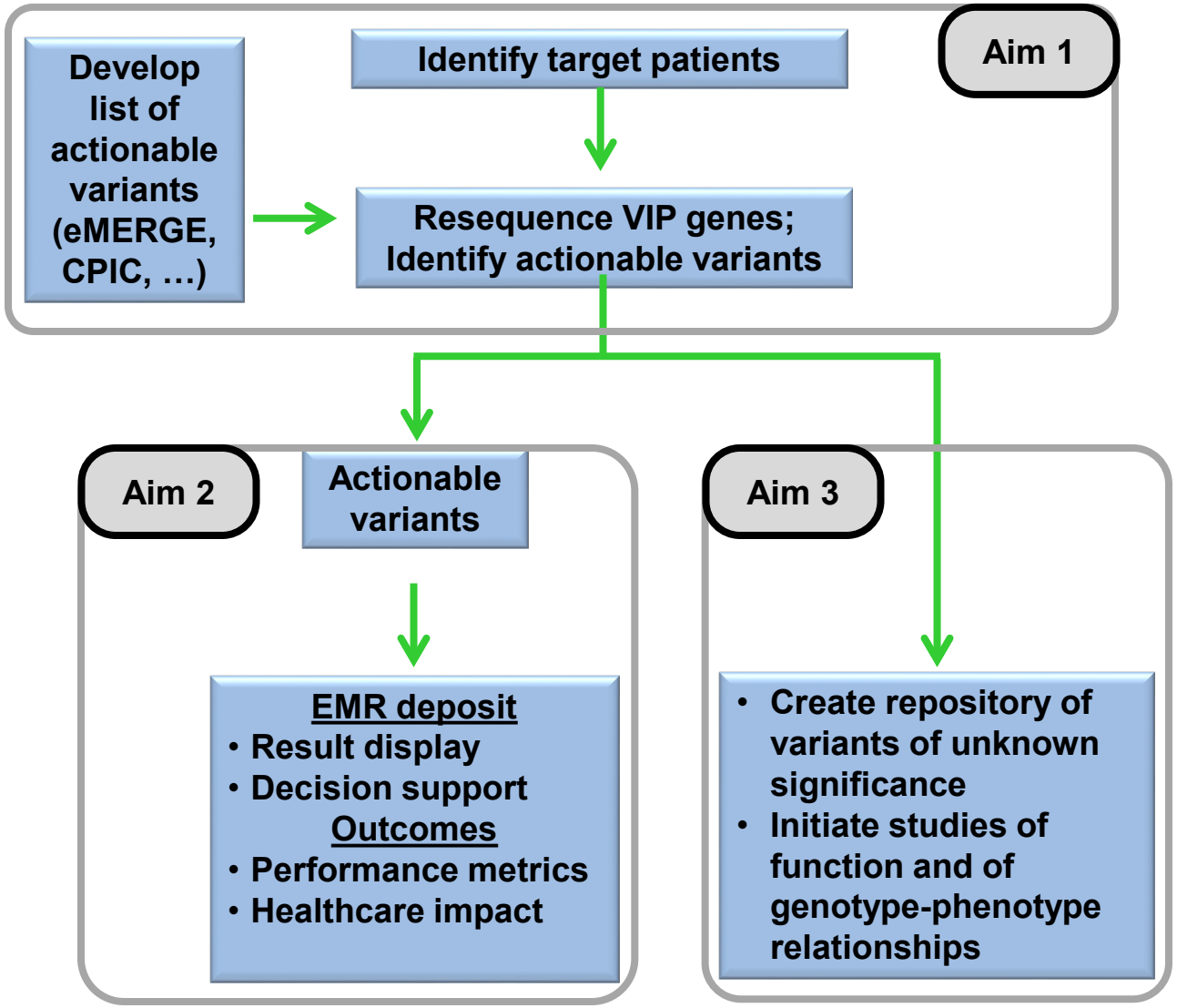
Reaction: Heidi Rehm

Summary: Dick Weinshilboum

Panel: Murray Brilliant, David Carey, John Carpten, Kim Doheny

PROGRESS TO DATE

eMERGE PGx – Overview by Aim



Progress Toward Target Enrollment (as of mid Jan 2014)

 # 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

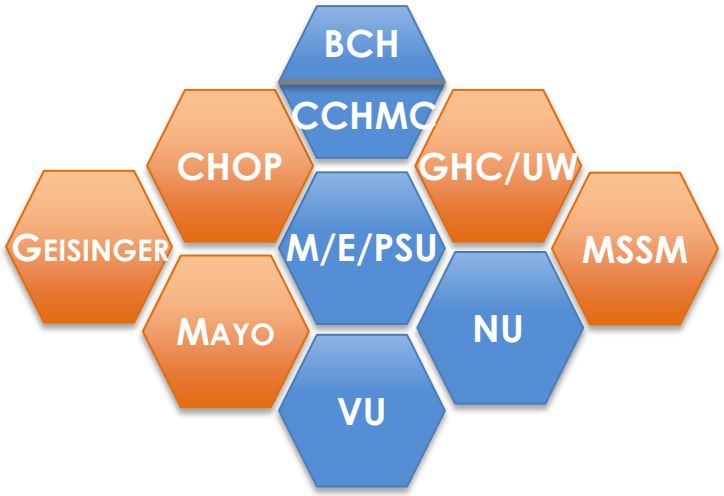
- Excellent results to date:
  - 32 diverse HapMap trios produced an average depth of coverage per sample of 496x

  - genotypes derived from this PGRNseq data were 99.9% concordant with existing SNV data on these samples from the 1000 Genomes project

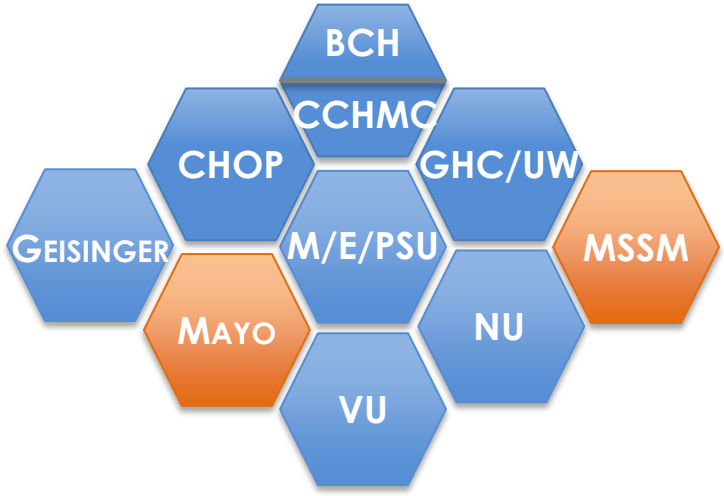
PGx platform

- 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*

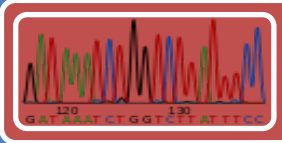
eMERGE PGx Project Summary

Specific Aim 1

Recruit / Collect Samples

PGRN-Seq Sequencing

Specific Aim 2

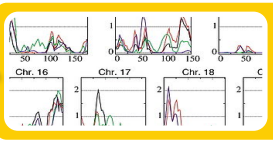


Clinical Variant Validation

Return Results: EHRintegration, CDS, Patient & Clinician Education

Outcomes Measures

Specific Aim 3



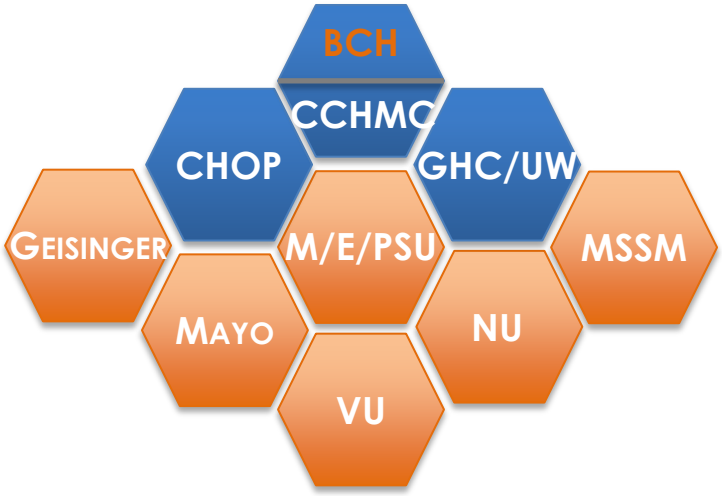
Populate Variant and Phenotype Data Repository (SPHINX)

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

Concordance	Raw (%)		Filtered (%)	
	Concordance	Discordance	Concordance	Discordance
CIDR vs. Mt. Sinai	98.125	1.436	99.421	0.329
CIDR vs. UW	97.859	1.223	99.127	0.393
UW vs. Mt. Sinai	98.001	1.215	99.130	0.468

Variants

	CIDR	Mt. Sinai	UW
CIDR	946361	10154	20312
Mt. Sinai	1540	937747	15152
UW	975	4429	927024

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.

PGRNSeq Concordance - vs. Seq Center

Site	UW	Mayo	Mt. Sinai	Northwestern	CHOP	Marshfield	Vanderbilt
Date Rec.	7/31	9/26	10/06	10/31	10/31	11/01	11/05
# Match Variants	10,861	8,625	12,582	7,354	11,814	5,028	6,893
# Filtered Var.	9,389	7,411	10,712	6,453	10,262	4,275	6,014
Discord (Het. / Hom)	0.211% / 0.023%	0.462% / 0.035%	0.323% / 0.024%	0.539% / 0.029%	0.380% / 0.029%	0.829% / 0.043%	0.638% / 0.038%
Raw Discrepant	0.003%	0.073%	0.048%	0.054%	0.061%	0.080%	0.060%
Raw Singleton Discord	0.015% / 0.007%	0.044% / 0.007%	0.050% / 0.007%	0.127% / 0.017%	0.062% / 0.008%	0.224% / 0.028%	0.124% / 0.016%
Filt. Discord	0 % / 0 %	0.002% / 0%	0.001% / 0%	0.003% / 0%	0.001% / 0%	0.006% / 0%	0.004% / 0%
Filt. Discrep.	4.826 %	6.766%	7.868%	5.661%	6.321%	6.960%	5.751%
Filt. Singleton Discord	0 % / 0 %	0.001% / 0%	0% / 0%	0.001% / 0%	0% / 0%	0.001% / 0%	0.003% / 0%

PGRNSeq Concordance - vs. SPHINX

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

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

EXTRA SLIDES

Cross-Site Comparison - eMERGE

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

eMERGE PGx QC Details

- Concordance checks
  - Concordance with VCF from sequencing center (typically single-called)
  - Concordance with group-called site vs. combined release
- Inconsistency checks
  - Duplicate study samples and controls called with different IDs
  - All samples renamed to eMERGE or Coriel IDs
  - VCF file checked for inconsistency (same ID, discordant calls)

eMERGE PGx QC results

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