Phenotype Discussion Summary

Josh Denny / Marylyn Ritchie

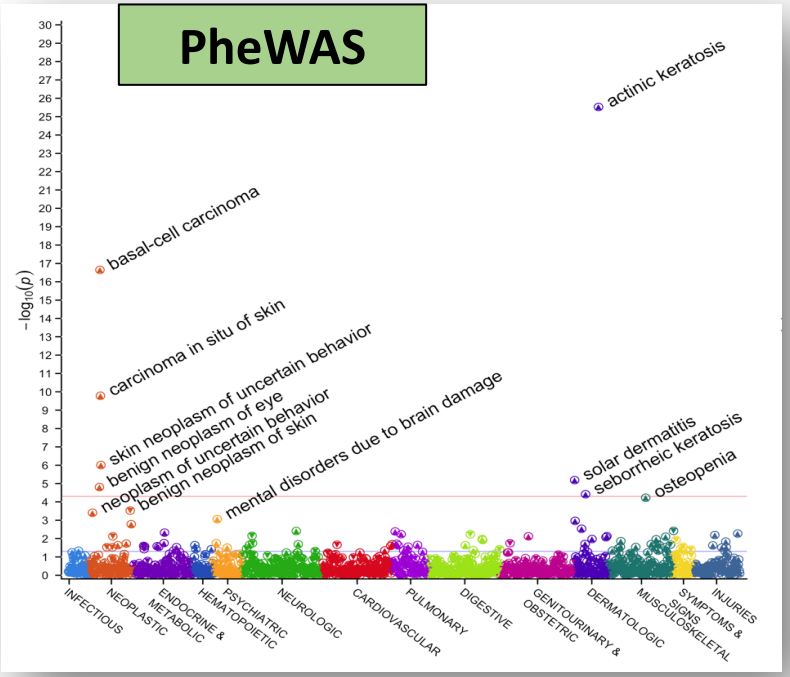
Key points

- eMERGE has pioneered use of phenotyping in the EHR, and is a model for other networks repurposing EHRs
- In general phenotype creation is still hard though has accelerated some (e1=14, e2=29, e3=27)
- PheKB has 154 mostly rule-based phenotypes, 75 have (already) been attributed to eMERGE
- Use of common data models and phenotype languages/models (OMOP, FHIR) should accelerate phenotype translation across sites
- Machine learning represents an opportunity to accelerated some but still require gold standard to train and portability has not been as robustly demonstrated as rule-based algorithms

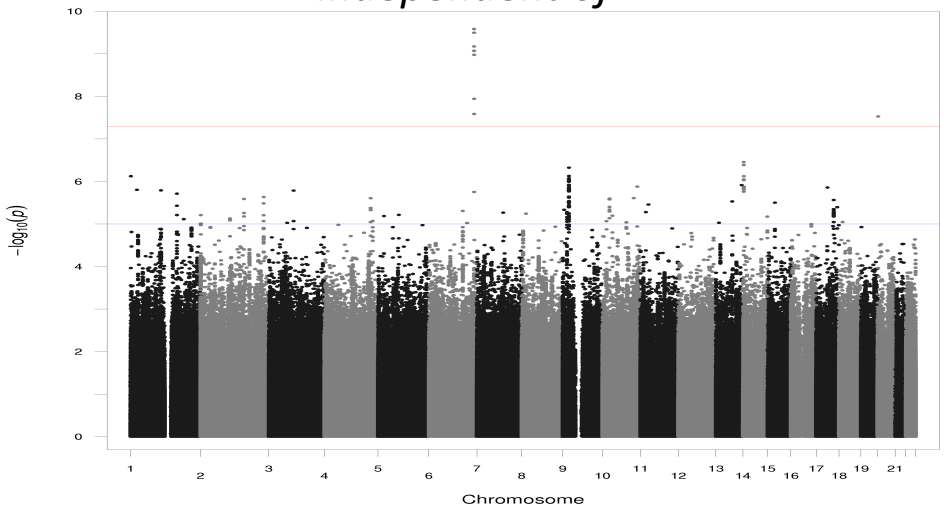
Phenotypes in PheKB now

	Public (n = 44)	Non-public (n = 110)	%
ICD-9 or -10 codes	39	73	73%
Medications	31	51	53%
CPT codes	23	44	44%
NLP	28	36	42%
Laboratory test results	21	37	38%
Vital signs	5	14	12%

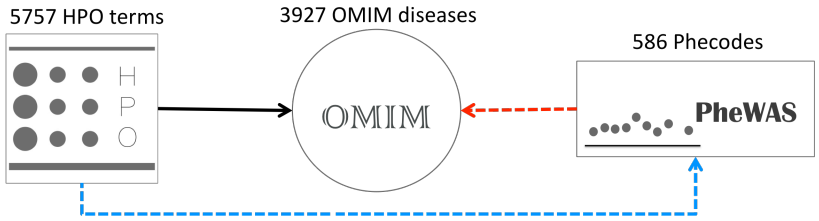
PheWAS



GWAS discovery in innovative phenotypes: MACE on Statins identified locus *independent of ΔLDL*



Clustering phenotypes: phenotype risk scores & mining for human phenotype ontology



$$PRS_i = \sum_{j=1}^{k} \left\{ \begin{matrix} 1 \\ 0 \end{matrix} \right\}_{ij} \log \frac{n_{total}}{n_j}, \text{ where j= Mendelian gene phenotypes}$$

Key points - 2

- There is a tradeoff: complicated phenotypes that take more time vs. simpler algorithms we can extend. Where is the greatest value for eMERGE?
- Multimodal phenotypes and use of text records/NLP are a hallmark of many eMERGE phenotypes
- Long history of phenotype innovation - pharmacogenomic, longitudinal phenotypes, OCR, portable NLP modules, KNIME, deeper phenotyping, PheWAS, phenotype risk scores
- Where does sequencing take us for EHR utility?