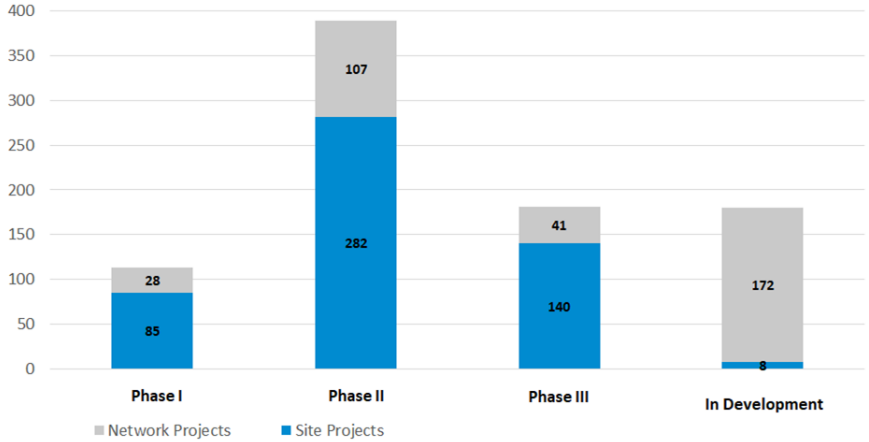
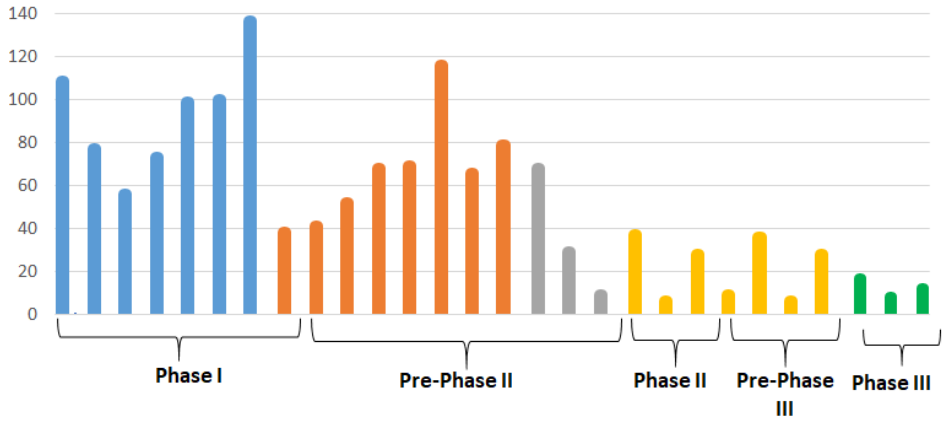
eMERGE Impact: Publications & Data





As of May 2020 there were 863 network and sites site-specific projects. 683 have been published.

eMERGE dbGaP Submissions
The have been 1473 external downloads as of May 2020

eMERGE Discovery & Implementation

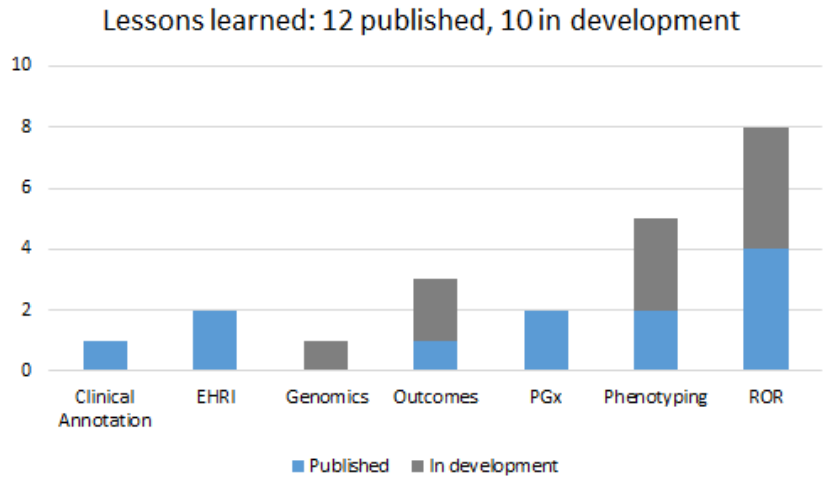
eMERGE Discovery Cohort	GWAS I-III	PGRNseq	Exome chip	Whole Exome Sequencing	Whole Genome Sequencing	eMERGEseq
157,480 total phenotype & genotype data available	105,108	9,010	12,865	3,745	1,796	24,956

- The eMERGE network launched with the goal of advancing discovery based EHR research, focused on establishing methods for developing and validating electronic phenotyping algorithms across multiple sites and EHRs.
- eMERGE evolved to focus on advancing translational efforts, returning clinically relevant findings as sites recruited, sequenced, and returned results from both the targeted Pharmacogenomics Research Network sequence platform (PGRNseq, Phase II) and a custom eMERGE sequencing panel centered around ACMG variants (eMERGEseq, Phase III).
- The eMERGE Network has balanced discovery while informing clinical research in the realm of return of genomic variants, adding to the field of genomic research and clinical implementation.

eMERGE: Future considerations for genomic research

Transitioning from genetic variant return to overall assessment of genomic risk & disease prevention

- **Enrollment & return for genomic testing of large cohorts**
  - Lessons from variable site experimental design ‘experiments of nature’, participant & provider surveys.
  - Benefits of uniform study design, single IRB.
- **Development of novel, complex algorithms for disease prediction**
  - Lessons from electronic phenotyping of clinical diseases and outcomes of interest.
  - Benefits of common data models and coding. Complexity of NLP and transferability of scripts across EMR



- **Integration of family history and clinical data to assess overall risk of a disease**
  - Lessons for streamlining data collection & compilation from RoR, outcomes, & genomic data.
  - Benefits of consistent and centralized variables and data definitions.
- **Return of results and integration for clinical implementation**
  - Lessons from EHRI, RoR, health care provider survey.
  - Necessity of provider engagement, participant comprehension, XML & FHIR standards for efficient transfer.

Clinical Annotation impact on the field

- Developed a better understanding of the rates of secondary findings in biobank populations which helps research programs estimate resources required for adding secondary finding return to genomic sequencing programs and informs consideration of screening.
- Network's additional non-ACMG IF to return were considered by other consortia and by ACMG.
- Developed pre-reporting interlaboratory harmonization approach to improve consistency in interpretation across clinical laboratories. This approach is now being applied in the All of Us Research program.
- Penetrance analysis ongoing on Arrhythmias (VUMC), FH (Mayo), Breast cancer (Columbia), and Aortopathy (Mayo)

Penetrance lessons learned

- Penetrance analysis is heavily influenced by ascertainment bias, this should be taken into consideration during study design.
- Penetrance should be examined across the lifespan as clinical manifestations of the disease occur at a variety of ages and not at a set point in time.
- Genetic test results can influence coding of diseases in the EHR, even if the participant does not currently present the condition, this can also influence penetrance analysis.
- Penetrance analysis can be effected by small sample sizes for specific conditions, for rare disorders and risk alleles, larger sample sizes will be required to obtain meaningful data.

Penetrance future considerations

- Future networks should be clear about their goals in terms of penetrance analysis during the startup phase of the project.
- Analysis of the data should begin early on, so any issues with data interpretation, collection, and missingness can be addressed.
- Care should be taken when enrolling a prospective cohort as unclear ascertainment can unnecessarily reduce sample size in penetrance analyses.

RoR impact on the field

- Guidance for investigators in navigating IRB for RoR projects
- Impact on providers
- Impact on participants: comprehension and personal value
- How do participants react to ‘neutral’ results
- Familial communication
- The RoR process across different settings in eMERGE
- RoR in the Pediatric setting
- Challenges encountered in RoR and how to be prepared for these: decedents, non-responders, previously tested, phenotype-genotype mismatch, reinterpretation of results
- RoR in low resource settings (intersection of SDH and Precision Med)

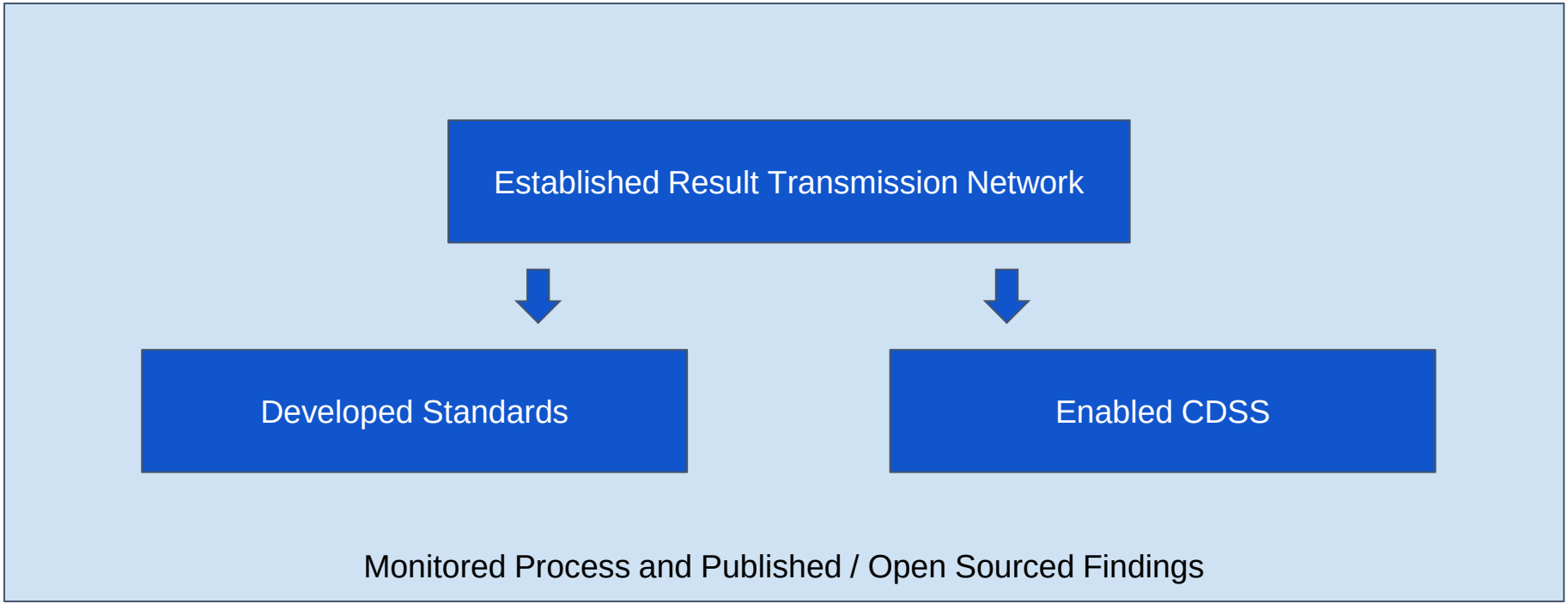
RoR lessons learned

- Coordinate the ROR processes across sites prior to the initiation of the study
- Coordinate participant survey or other cross-site studies before sites are sending surveys out.
- Importance of coordinating survey and interview studies across sites before sites are collecting these data
- Integrating pediatric data with adult data is a challenge and how to do this should be addressed at the outset of future studies.

RoR future considerations

- Disclosure of PRS has its own challenges: numeracy
- 25% of recruited cohorts with high PRSs will need FTF RoR (compared to 5% in eIII)
- Efficient and scalable means for RoR
- Innovations: chatbots, nurses with genetics focus, videos
- RoR protocols should account for non-responders, decedents.
- Provider education and engagement to use risk estimates that include PRS
- Familial communication of polygenic risk is uncharted territory

EHRI impact on the field



EHRI lessons learned

- Clinical-grade cross-institutional interface projects require significant effort
  - Clinical and laboratory sites are extremely heterogeneous (data production, format and structure of reports, interpretation standards, exchange of data between systems, how it is stored in the EHR, how it is consumed and used for CDSS)
  - IT processes and clinical processes are symbiotic (when you are designing IT processes you are also designing clinical processes)
  - Cross site differences need to be assessed early to get networks off the ground
- Some genomics functionality may not be practical to implement in real world clinical systems
- Forums to share lessons learned as implementations occur can be very helpful

EHRI future considerations

- **Start early** on designing network topology and inter-institutional interfaces
- Recognize cloud infrastructure and traditional infrastructure are different - **start early** on designing for this as well
- **Start early** working on FHIR standard-based germline variant communication
- **Start early** on engaging cross industry stakeholders who will influence rate of FHIR adoption
- **Start early** on thinking about how eMERGE internal network/app development could be used to encourage commercial players to leverage standard interfaces

Outcomes impact on the field

- Defined and harmonized measureable health services and clinical outcomes across a broad range of genomic medicine scenarios
- Reported consequences of returning targeted sequencing results at scale in non-research settings

Outcomes lessons learned

- Context is important to understand the changes (or lack thereof) in health services delivered.
- Context is difficult to uniformly capture across a large cohort; complementary qualitative assessments may be critical.
- Pediatric cohorts offer the potential for longitudinal studies in the future.

Outcomes future considerations

- eMERGE III natural experiment without a formal study design improves the breadth of lessons learned but represents a challenge to generalizing results
  - Narrow recruitment to those who are unselected for trait or prior test result
  - Standardize return of results when possible which will reduce variability in downstream participant and provider response.

Genomics impact on the field

- Generated one of the largest (~105K) genetic data sets with moderate diversity linked to EHR for deep phenotyping
- Understanding of genetic ancestry compared to self-reported (and observed-) ancestry
- Provided guidance with controlling for ascertainment bias in genome-wide association studies
- Development of common phenotypes to assist in covariate adjustment in regression models
  - Aggregate ICD-9 (10), CPT codes, labs, demographics assist in covariate development without straining individual site resources

Genomics lessons learned

- Adding and removing participants once a large dataset is compiled costs significant amounts of time and resources.
- Clearly defined data freezes taking into account diversity, phenotypic data, and discovery & implementation goals should be outlined at the beginning of the network to maximize data delivery and analysis time at the sites.
- Standard naming conventions are necessary when trying to combine files from multiple sequencing centers will maximize efficiency and turn around time.
- Cloud computing can be used to set up standard pipelines for analysis, saving time, resources, and improving consistency.

Genomics future considerations

- Understanding the framework of discovery and implementation and balancing at a network level is critical for progress
- Understanding the framework of discovery in a data set enriched for European-ancestry while striving for more diversity at a network level is critical for the future of genomic medicine

PGx impact on field

- Created relationship with CPIC to generate additional association information to be used in guideline statements

- Collaborations with IGNITE to investigate strategies for the clinical implementation of CYP2D6 genotyping to guide drug prescribing; describe the scope and use of CYP2D6 and CYP2C19 testing to inform opioid therapy for pain management and antidepressant therapy; and investigate required clinical implementation of CYP2D6 to inform opioid therapy

PGx lessons learned

- Some genotyping chips do not contain all data needed for full guideline implementation, particularly in non-white racial and ethnic groups.
- CYP2D6 is critical for many medications and associated guidelines.
- After deployment, sufficient time should be allowed for impact to be demonstrated.
- PGx phenotypes should be identified early on.
- Limitations of current PGx guidelines should be considered, particularly concerning non-white populations.

PGx future considerations

- Allow sufficient time and effort for implementation, even at sites with prior PGx implementation experience
- Not necessarily an easy add on to currently funded non-PGx projects
- Prioritize PGx phenotypes early, recognizing this can require complex genotyping, interpretation and EHR extraction of drugs, outcomes, timing
- Include with these phenotypes sub-phenotypes for examination of practitioner and patient interaction with CDS
- Consider limitations of current PGx guidelines, particularly with regard to non-white populations (and foster more discovery research specifically in these populations)

NLP Phenotyping impact on the field

- NLP improves phenotyping performance and enables high-fidelity phenotyping
- It is feasible to develop portable NLP algorithms with reproducible performance
- NLP requires dedicated
  - Notes server
  - Pre-processing pipelines
  - Skilled NLP Personnel

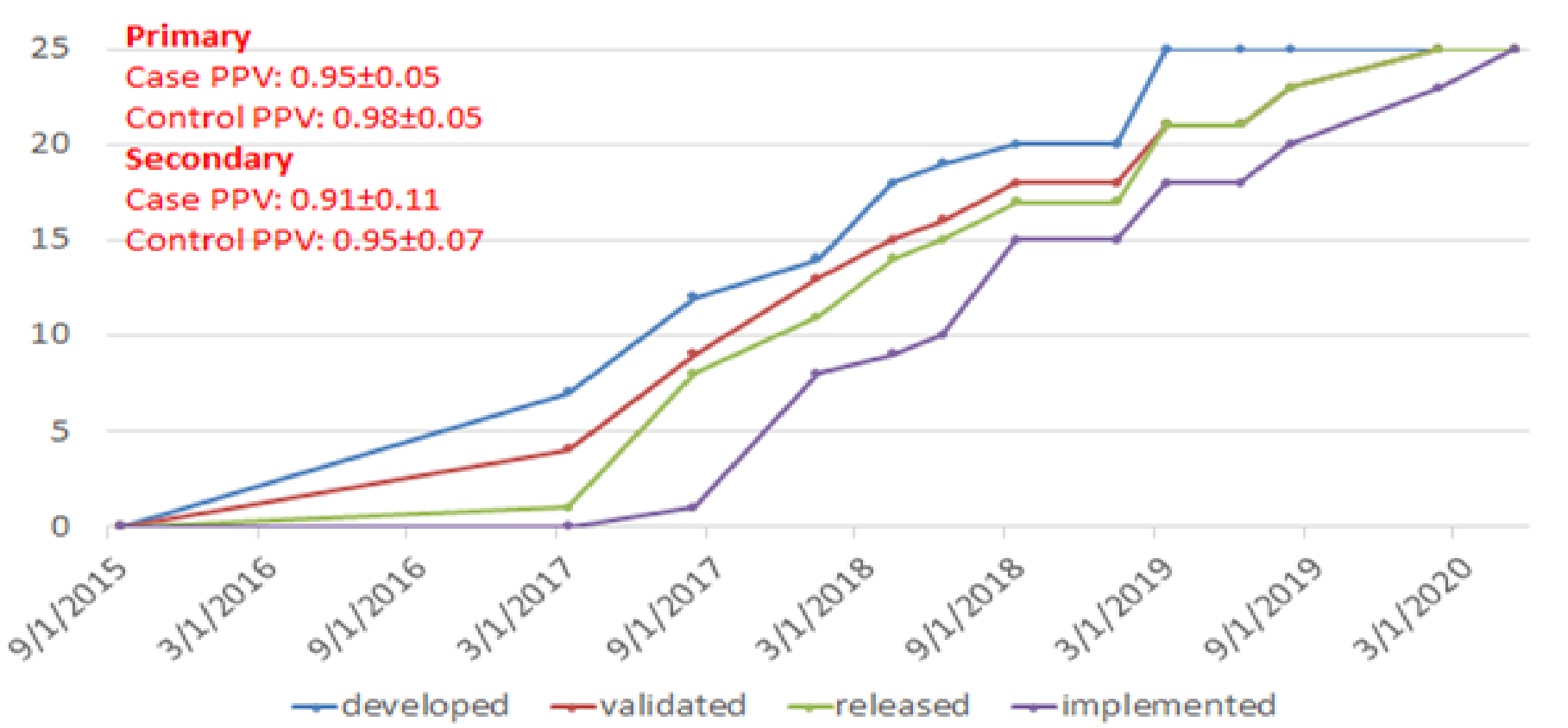
NLP lessons learned

- Unlike structured EHR data, NLP has extra requirements for privacy protection, technical infrastructure setup, and high-fidelity notes provision
- Unlike SQL scripts, NLP software have IP issues
- NLP performance may vary across sites due to heterogeneity in notes naming and structuring
- Negation remains an open NLP challenge
- NLP pipelines may not generalize to rare phenotypes

NLP future considerations

- Start with semi-structured notes
- Adopt a standard terminology for document types
- Enhance code modularization
- Standardize the protocol for NLP validation and implementation
- Standardize the documentation for NLP algorithms
- Improve communication across sites

Phenotyping timeline



Timeline of five NLP components