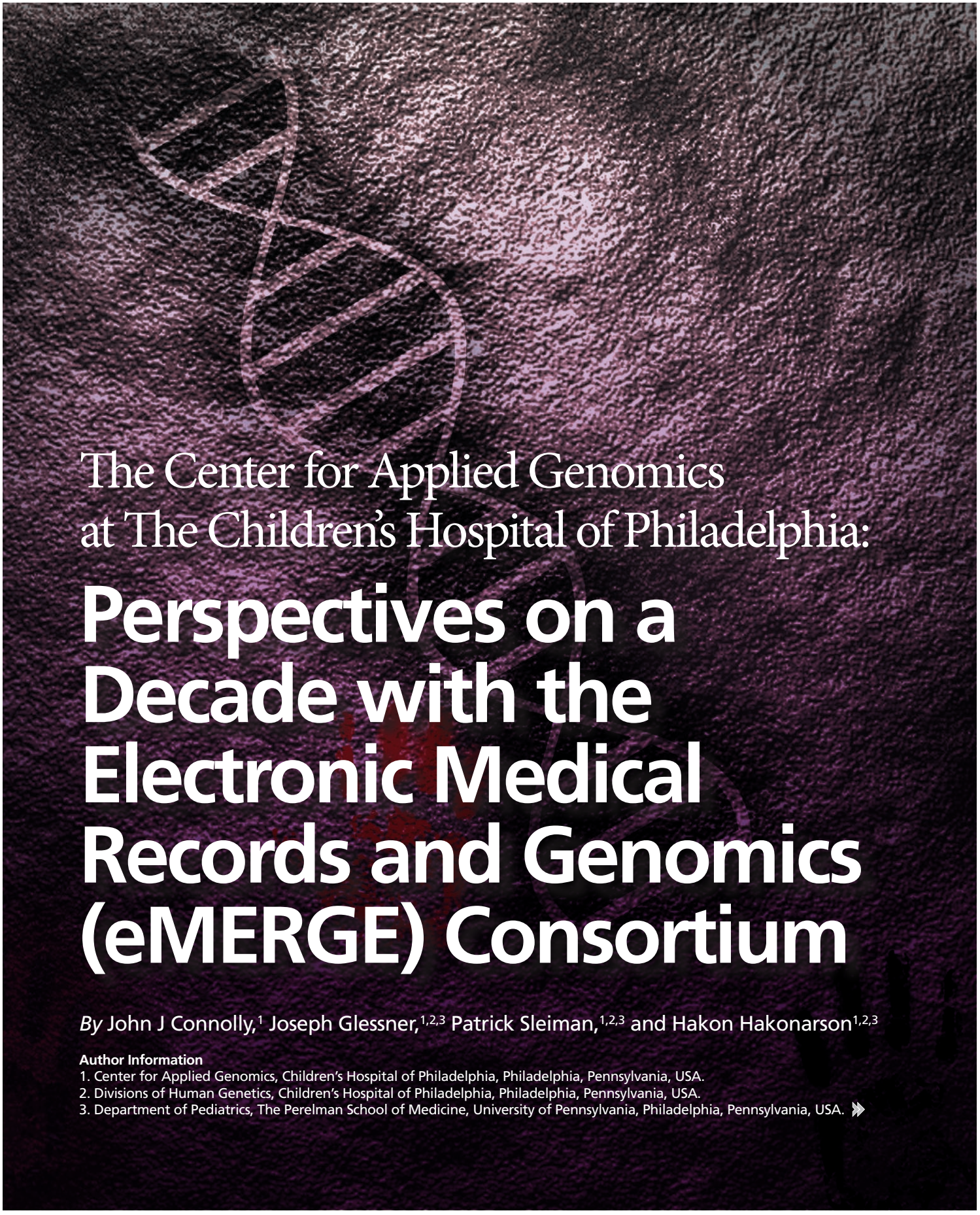


The Center for Applied Genomics at The Children's Hospital of Philadelphia:

Perspectives on a Decade with the Electronic Medical Records and Genomics (eMERGE) Consortium

By John J Connolly,[1] Joseph Glessner,[1,2,3] Patrick Sleiman,[1,2,3] and Hakon Hakonarson[1,2,3]

Author Information
1. Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.
2. Divisions of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.
3. Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

Don't let patients with TARGETABLE MUTATIONS get lost in the crowd

There are ~4,000 to 5,000 patients with *MET*ex14 in mNSCLC per year in the United States.[1-2]

Nearly 1 in 2 patients with mNSCLC may have a targetable oncogenic mutation,[3-10] but many patients are not tested for all potential targets (prevalence of *MET*ex14 ~3%).[4,9,11-15]

The National Comprehensive Cancer Network® (NCCN®) recommends testing for *ALK*, *KRAS*, *BRAF*, *EGFR*, ***MET*ex14**, *NTRK1/2/3*, *RET*, *ROS1* and PD-L1 in eligible newly diagnosed mNSCLC patients.[16*]

Up-front broad molecular profiling may help optimize first-line treatment for mNSCLC.

MET, mesenchymal-epithelial transition; *MET*x14, *MET* exon skipping; mNSCLC, metastatic non-small cell lung cancer.

*The NCCN Guidelines® for NSCLC provide recommendations for certain individual biomarkers that should be tested and recommend testing techniques but do not endorse any specific commercially available biomarker assays or commercial laboratories.

NCCN makes no warranties of any kind whatsoever regarding their content, use or application and disclaims any responsibility for their application or use in any way.

References: **1.** Vuong HG, et al. Clinicopathological implications of MET exon 14 mutations in non-small cell lung cancer — A systematic review and meta-analysis. *Lung Cancer.* 2018;123:76-82. **2.** Data on file. Novartis Calculation. Kantar Health. CancerMPact: lung (non-small cell) metastatic stage IV incidence and newly recurrent. Updated December 15, 2018. my.khapps.com **3.** Arcila ME, et al. EGFR exon 20 insertion mutations in lung adenocarcinomas: prevalence, molecular heterogeneity, and clinicopathologic characteristics. *Mol Cancer Ther.* 2013;12(2):220-229. **4.** Awad M, et al. MET Exon 14 mutations in non–small-cell lung cancer are associated with advanced age and stage-Dependent MET Genomic Amplification and c-Met overexpression. *J Clin Oncol.* 2016;34(7):721-730. **5.** Brustugun OT, et al. BRAF-mutations in non-small cell lung cancer. *Lung Cancer.* 2014;84(1):36-38. **6.** Nadal E, et al. KRAS-G12C mutation is associated with poor outcome in surgically resected lung adenocarcinoma. *J Thorac Oncol.* 2014;9(10):1513-1522. **7.** Oxnard GR, et al. Natural history and molecular characteristics of lung cancers harboring EGFR exon 20 insertions. *J Thorac Oncol.* 2013;8(2):179-184. **8.** Scheffler M, et al. K-ras mutation subtypes in NSCLC and associated co-occuring mutations in other oncogenic pathways. *J Thorac Oncol.* 2019;14(4):606-616. **9.** Shea M, et al. Management of advanced non-small cell lung cancers with known mutations or rearrangements: latest evidence and treatment approaches. *Ther Adv Respir Dis.* 2016;10(2):113-129. **10.** Vaishnavi A, et al. Oncogenic and drug sensitive NTRK1 rearrangements in lung cancer. *Nat Med.* 2013;19(11):1469-1472. **11.** Gierman HJ, et al. Genomic testing and treatment landscape in patients with advanced non-small cell lung cancer (aNSCLC) using real-world data from community oncology practices. *J Clin Oncol.* 2019;37(15 suppl):1585-1585. **12.** Griesinger F, et al. Biomarker testing in non-small cell lung cancer in routine care: Analysis of the first 3,717 patients in the German prospective, observational, nation-wide CRISP Registry. *Lung Cancer.* 2020;(152):174-184. **13.** Gutierrez ME, et al. Genomic profiling of advanced non-small cell lung cancer in community settings: gaps and opportunities. *Clin Lung Cancer.* 2017;18(6):651-659. **14.** Wempe MM, et al. A national assessment of diagnostic test use for patients with advanced NSCLC and factors influencing physician decision-making. *Am Health Drug Benefits.* 2020;13(3):110-119. **15.** Yu TM, et al. Multiple biomarker testing tissue consumption and completion rates with single-gene tests and investigational use of Oncomine Dx Target Test for advanced non-small-cell lung cancer: A single-center analysis. *Clin Lung Cancer.* 2019;20(1):20-29.e8. **16.** Referenced with permission from the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for Non-Small Cell Lung Cancer V.5.2021.

NOVARTIS

Novartis Pharmaceuticals Corporation
East Hanover, New Jersey 07936-1080

7/21

133636

THE MISSION of the Center for Applied Genomics (CAG) is to develop new and better ways to diagnose and treat children affected by rare and complex medical disorders,[1] including, but not limited to, ADHD, asthma, autism, diabetes, epilepsy, obesity, schizophrenia, pediatric cancer, and a wide range of rare diseases. Ultimately, the Center's goal is to generate new diagnostic tests and to guide physicians to the most appropriate therapies.

To help accomplish this mission, CAG is engaged in over 100 active collaborations with aligned research teams across the globe. Among CAG's most productive partnerships is its ongoing work with the electronic medical records and genomics (eMERGE) Network,[2] spanning more than 100 projects in the past decade. In this article, we review the history of collaboration between CAG and eMERGE and highlight a range of novel strategies aimed at delivering genomics-based individualized healthcare.

eMERGE Phases 2007–2026

The eMERGE Network was established in 2006 and has since funded a consortium of U.S. medical research institutions. The stated goal of the Network is *to develop, disseminate, and apply approaches to research that combine biorepositories with electronic medical record (EMR) systems for genomic discovery and genomic medicine implementation research.* This goal necessitates shared expertise in genomics, statistics, ethics, informatics, and clinical medicine, all of which are integrated in a comprehensive infrastructure across multiple sites. The Network has grown steadily over the years, evolving in tandem with developments in healthcare technologies (e.g., genomics, phenomics, and healthcare research and practice). The current membership of the funded consortium includes ten clinical centers,[2] a genotyping center (Broad), and a coordinating center (Vanderbilt). This overview presents a brief history of the consortium and its outputs through the years; the challenges faced and overcome; and an analysis of what we expect in the coming years.

"Collectively, disruptions of genes in this network (as determined by the presence of CNVs), account for ~12% of total ADHD cases when corrected for control occurrence. These results suggest that disrupted mGluR signaling/activity is involved in ADHD in a subset of patients that can be identified based on genetic profiling."

Phase 1 (P1: 2007-2011):

eMERGE Phase 1 was established in September 2007. Supported by the National Human Genome Research Institute (NHGRI), the network built a strong foundation for the more expansive iteration that exists today. The initial years of eMERGE were critical to establishing the workgroup structure and approaches to data-sharing, which remain a cornerstone of the eMERGE enterprise. In particular, eMERGE invested significant resources in creating a robust platform for developing shareable phenotype algorithms compatible with EMRs. These algorithms were critical to driving genome-wide association studies (GWAS) for a range of diseases.[3-6] Perhaps more importantly, they established a proof-of-principle that (often disparate) large medical centers could share rules, approaches, and, ultimately, case-control datasets to drive large-scale genomic studies. eMERGE continues to be a world-leader in innovating approaches to electronic algorithm development. We recount next the legacy programs developed during Phase 1.

P1 Legacy

Phenotyping:

Collectively, eMERGE has established a strong track-record in creating validated phenotyping algorithms in support of high-throughput genomics research, and sharing experiences and know-how in genomics discovery, methodology, and clinical decision support. Respective projects and resources have become increasingly mature through four phases of research and development. eMERGE published – and continues to maintain – a searchable catalog of EMR-based phenotyping algorithms called the Phenotype KnowledgeBase (PheKB), which is available to investigators worldwide.[7] It continues to develop and publish best practices for sharing genomic data and EMR-based phenotypes while protecting participants' privacy.[8-13]

Ethical, legal, and social implications (ELSI):

eMERGE published model consent language for EMR-linked biorepositories, intended to harmonize the consent process for the collection and storage of human biospecimens and data for future research, particularly those collections that have an EMR component. eMERGE continues to research ELSI in the light of EMR-based genomics programs, implementation, and data-sharing.

References

1. JJ Connolly JG, D Li, PMA Sleiman, H Hakonarson. The Center for Applied Genomics at The Children's Hospital of Philadelphia – Pediatric Perspectives on Genomic Medicine2019. Available from: https://www.thejournalofprecisionmedicine.com/the-journal-of-precision-medicine/the-center-for-applied-genomics-at-the-childrens-hospital-of-philadelphia-pediatric-perspectives-on-genomic-medicine/.
2. The Electronic Medical Records and Genomics Consortium 2021. Available from: https://emerge-network.org/.
3. Kullo IJ, Ding K, Shameer K, McCarty CA, Jarvik GP, Denny JC, et al. Complement receptor 1 gene variants are associated with erythrocyte sedimentation rate. Am J Hum Genet. 2011;89(1):131-8. Epub 20110623. doi: 10.1016/j.ajhg.2011.05.019. PubMed PMID: 21700265; PubMed Central PMCID: PMC3135803.
4. Crosslin DR, McDavid A, Weston N, Nelson SC, Zheng X, Hart E, et al. Genetic variants associated with the white blood cell count in 13,923 subjects in the eMERGE Network. Hum Genet. 2012;131(4):639-52. Epub 20111030. doi: 10.1007/s00439-011-1103-9. PubMed PMID: 22037903; PubMed Central PMCID: PMC3640990.
5. Denny JC, Ritchie MD, Crawford DC, Schildcrout JS, Ramirez AH, Pulley JM, et al. Identification of genomic predictors of atrioventricular conduction: using electronic medical records as a tool for genome science. Circulation. 2010;122(20):2016-21. Epub 20101101. doi: 10.1161/circulationaha.110.948828. PubMed PMID: 21041692; PubMed Central PMCID: PMC2991609.
6. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. Am J Hum Genet. 2011;89(4):529-42. doi: 10.1016/j.ajhg.2011.09.008. PubMed PMID: 21981779; PubMed Central PMCID: PMC3188836.
7. Kirby JC, Speltz P, Rasmussen LV, Basford M, Gottesman O, Peissig PL, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. J Am Med Inform Assoc. 2016;23(6):1046-52. Epub 20160328. doi: 10.1093/jamia/ocv202. PubMed PMID: 27026615; PubMed Central PMCID: PMC5070514.
8. Fullerton SM, Wolf WA, Brothers KB, Clayton EW, Crawford DC, Denny JC, et al. Return of individual research results from genome-wide association studies: experience of the Electronic Medical Records and Genomics (eMERGE) Network. Genet Med. 2012;14(4):424-31. Epub 20120223. doi: 10.1038/gim.2012.15. PubMed PMID: 22361898; PubMed Central PMCID: PMC3723451.
9. McGuire AL, Basford M, Dressler LG, Fullerton SM, Koenig BA, Li R, et al. Ethical and practical challenges of sharing data from genome-wide association studies: the eMERGE Consortium experience. Genome Res. 2011;21(7):1001-7. Epub 20110601. doi: 10.1101/gr.120329.111. PubMed PMID: 21632745; PubMed Central PMCID: PMC3129243.
10. Loukides G, Gkoulalas-Divanis A, Malin B. Anonymization of electronic medical records for validating genome-wide association studies. Proc Natl Acad Sci U S A. 2010;107(17):7898-903. Epub 20100412. doi: 10.1073/pnas.0911686107. PubMed PMID: 20385806; PubMed Central PMCID: PMC2867915.
11. Loukides G, Denny JC, Malin B. The disclosure of diagnosis codes can breach research participants' privacy. J Am Med Inform Assoc. 2010;17(3):322-7. doi: 10.1136/jamia.2009.002725. PubMed PMID: 20442151; PubMed Central PMCID: PMC2995712.
12. Malin B. Secure construction of k-unlinkable patient records from distributed providers. Artif Intell Med. 2010;48(1):29-41. Epub 20091028. doi: 10.1016/j.artmed.2009.09.002. PubMed PMID: 19875273.
13. Malin B, Benitez K, Masys D. Never too old for anonymity: a statistical standard for demographic data sharing via the HIPAA Privacy Rule. J Am Med Inform Assoc. 2011;18(1):3-10. doi: 10.1136/jamia.2010.004622. PubMed PMID: 21169618; PubMed Central PMCID: PMC3005867.

Powering Better Decisions for Your Patients

OmniSeq INSIGHT℠ delivers a comprehensive, actionable report for all solid tumors.

Learn more at oncology.labcorp.com/omniseq

labcorp | Oncology

Case Study: TPMT

Principle under study

Thiopurine S-methyltransferase (TPMT) is an enzyme that drives metabolism of purine analog drugs – namely azathioprine, 6-mercaptopurine and thioguanine – which are used as chemotherapeutic and immunosuppressant agents for treating lymphoid malignancies, leukemias, inflammatory bowel disease, and other immune conditions.[19,20] TPMT maps to chromosome 6p22.3 and has least 34 alleles, mostly associated with reduced activity *in vitro*.[20] Alleles *2 (rs1800462), *3A (rs1800460 and rs1142345), *3B (rs1800460), and *3C (rs1142345) account for 95% of all non-wild type alleles, and all four involve missense mutations (a mistake in DNA resulting in a wrong amino acid in a protein). Frequencies of these alleles vary across populations[19] – *3A is most commonly found in European cohorts (4.5%),[21] *3C is more prevalent in Africans or Asians, with 5.4–7.6% (reviewed in[22]). Across populations, approximately 0.3% of individuals carry two defective alleles (associated with negligible activity), about 10% are heterozygous (intermediate activity), and 89% have normal activity.[21,23] Individuals who do not have normal activity are at higher risk of developing myelosuppression within weeks of starting treatment with conventional doses of thiopurine. These can be fatal if unrecognized, particularly in homozygous individuals.[24] Appropriately, the US Food and Drug Administration (FDA) recommends TPMT testing prior to starting treatment with thiopurine drugs, and TPMT genotype-guided dosing recommendations are currently in use.[20,25]

Proof-of-principle

CAG[26] examined whether genome-array data captured on commercial platforms could be imputed to identify common defective TPMT alleles, which in turn could be used to identify individuals with high-risk genotypes. Results were intended to assess whether commercial arrays can be leveraged to pre-emptively capture high-risk individuals, whose data can be entered into the EMR in order to mitigate risk of thiopurine-associated adverse events. The research team imputed TPMT variants in 87,979 samples from the biobank at CAG using the 1000 Genomes Project as reference. In a sample of 630 subjects, Sanger sequencing (N = 59) and direct genotyping (N = 583) (12 samples overlapping in the two groups) were used to confirm concordance between the imputed and observed genotypes, as well as the sensitivity, specificity and positive and negative predictive values of the imputation. The SNPs that represent TPMT alleles *3A, *3B, and *3C ((rs1800460 and rs1142345) were imputed. Overall, 98.88% of individuals (623/630) were correctly imputed into carrying no risk alleles (553/553), heterozygous (45/46) and homozygous (25/31). The standout result here is that no individual was incorrectly identified as a normal metabolizer, which aligns with- normal dosing. Based on this finding, we can conclude that imputation of TPMT alleles from existing genomic data can be used as a first step in the screening of individuals at risk of developing serious adverse events secondary to thiopurine drugs. We can further conclude that, for appropriately-consented individuals and using appropriately-handled samples (e.g., CLIA-compliant), genomic data can be generated at low cost and used to identify pharmacogenetic actionable variants such as those relevant to TPMT. Moreover, such data can be integrated to the existing EMR, and remain "inactive" unless a provider recommends a PGx-relevant treatment. "Activation" of a relevant drug order would then helping trigger clinical decision support and, for example, recommend additional testing or more careful monitoring etc.

Phase 2 (P2: 2011–2015):

eMERGE Phase 2 expanded from 5 to 9 sites and was opened to pediatric institutions for the first time. The Network continued to innovate in electronic phenotyping, data-sharing, and genomics discovery.[14-17] Importantly, the Network undertook a large supplemental project early in Phase 2, which represented a major change in approach from discovery-based science toward implementation, specifically in returning reports based on individuals' genomic risk. Collectively, the Network's clinical centers adopted the Pharmacogenomics Global Research (PGRN) Network pharmacogenomics (PGx) panel and sequenced 84 PGx candidate genes in over 9,000 participants. A PGx-based risk report was created for all participants (N>9,000), and reports were returned to participants' EMRs, with clinical decision support and relevant outcomes capture downstream of this return. To fulfill this commitment, relevant workgroups pivoted to capture enhanced workflows for EMR integration and outcomes measurement and to anticipate and observe the ethical, legal, and social implications of returning reports. A number of relevant legacy programs are eviewed briefly in P2 Legacy, below.

P2 Legacy

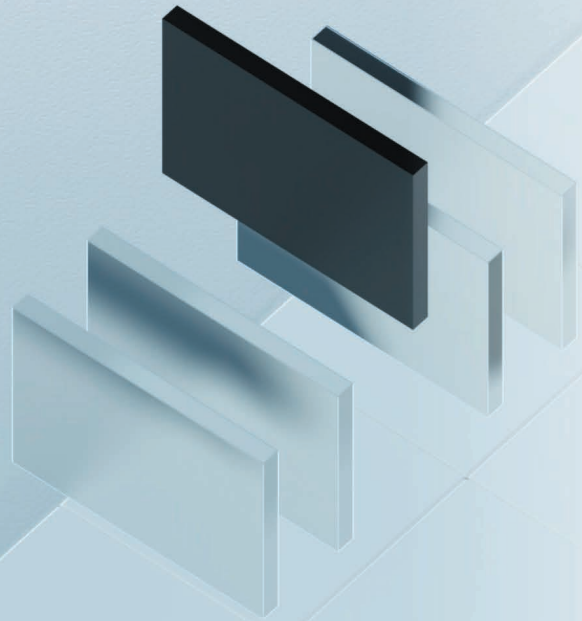
Pediatrics in eMERGE

As discussed previously,[1] translational research approaches in adults and children are often disconnected. Because children constitute a "special" population, they are often excluded from clinical research and trials. While the intention of such exclusions is to promote safety and protect vulnerable individuals, an unwelcome side-effect is the comparative under-investment in pediatric clinical trials.[18] Symptomatic of this issue was the previous ineligibility of pediatric sites under eMERGE Phase 1. Phase 2, with the admission of two pediatric sites – CHOP and a »

14. Van Driest SL, Wells QS, Stallings S, Bush WS, Gordon A, Nickerson DA, et al. Association of Arrhythmia-Related Genetic Variants With Phenotypes Documented in Electronic Medical Records. Jama. 2016;315(1):47-57. doi: 10.1001/jama.2015.17701. PubMed PMID: 26746457; PubMed Central PMCID: PMC4758131.
15. Herr TM, Bielinski SJ, Bottinger E, Brautbar A, Brilliant M, Chute CG, et al. Practical considerations in genomic decision support: The eMERGE experience. J Pathol Inform. 2015;6:50. Epub 20150928. doi: 10.4103/2153-3539.165999. PubMed PMID: 26605115; PubMed Central PMCID: PMC4629307.
16. Bush WS, Crosslin DR, Owusu-Obeng A, Wallace J, Almoguera B, Basford MA, et al. Genetic variation among 82 pharmacogenes: The PGRNseq data from the eMERGE network. Clin Pharmacol Ther. 2016;100(2):160-9. Epub 20160601. doi: 10.1002/cpt.350. PubMed PMID: 26857349; PubMed Central PMCID: PMC5010878.
17. Lingren T, Chen P, Bochenek J, Doshi-Velez F, Manning-Courtney P, Bickel J, et al. Electronic Health Record Based Algorithm to Identify Patients with Autism Spectrum Disorder. PLoS One. 2016;11(7):e0159621. Epub 20160729. doi: 10.1371/journal.pone.0159621. PubMed PMID: 27472449; PubMed Central PMCID: PMC4966969.
18. Chaudhari BP, Manickam K, McBride KL. A pediatric perspective on genomics and prevention in the twenty-first century. Pediatr Res. 2019. Epub 2019/10/03. doi: 10.1038/s41390-019-0597-z. PubMed PMID: 31578042.
19. Appell ML, Berg J, Duley J, Evans WE, Kennedy MA, Lennard L, et al. Nomenclature for alleles of the thiopurine methyltransferase gene. Pharmacogenet Genomics. 2013;23(4):242-8. doi: 10.1097/FPC.0b013e32835f1cc0. PubMed PMID: 23407052; PubMed Central PMCID: PMC3727893.
20. Relling MV, Gardner EE, Sandborn WJ, Schmiegelow K, Pui CH, Yee SW, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. Clin Pharmacol Ther. 2011;89(3):387-91. Epub 20110126. doi: 10.1038/clpt.2010.320. PubMed PMID: 21270794; PubMed Central PMCID: PMC3098761.
21. Schaeffeler E, Fischer C, Brockmeier D, Wernet D, Moerike K, Eichelbaum M, et al. Comprehensive analysis of thiopurine S-methyltransferase phenotype-genotype correlation in a large population of German-Caucasians and identification of novel TPMT variants. Pharmacogenetics. 2004;14(7):407-17. doi: 10.1097/01.fpc.0000114745.08559.db. PubMed PMID: 15226673.
22. Teml A, Schaeffeler E, Herrlinger KR, Klotz U, Schwab M. Thiopurine treatment in inflammatory bowel disease: clinical pharmacology and implication of pharmacogenetically guided dosing. Clin Pharmacokinet. 2007;46(3):187-208. doi: 10.2165/00003088-200746030-00001. PubMed PMID: 17328579.
23. Weinshilboum RM, Sladek SL. Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity. Am J Hum Genet. 1980;32(5):651-62. PubMed PMID: 7191632; PubMed Central PMCID: PMC1686086.
24. Sim SC, Kacevska M, Ingelman-Sundberg M. Pharmacogenomics of drug-metabolizing enzymes: a recent update on clinical implications and endogenous effects. Pharmacogenomics J. 2013;13(1):1-11. Epub 20121023. doi: 10.1038/tpj.2012.45. PubMed PMID: 23089672.
25. Relling MV, Gardner EE, Sandborn WJ, Schmiegelow K, Pui CH, Yee SW, et al. Clinical pharmacogenetics implementation consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing: 2013 update. Clin Pharmacol Ther. 2013;93(4):324-5. Epub 20130117. doi: 10.1038/clpt.2013.4. PubMed PMID: 23422873; PubMed Central PMCID: PMC3604643.
26. Almoguera B, Vazquez L, Connolly JJ, Bradfield J, Sleiman P, Keating B, et al. Imputation of TPMT defective alleles for the identification of patients with high-risk phenotypes. Front Genet. 2014;5:96. Epub 20140512. doi: 10.3389/fgene.2014.00096. PubMed PMID: 24860591; PubMed Central PMCID: PMC4026736.

TEMPUS

Discover, design, and deliver the right drug for the right patient



Integrated solutions for early discovery through commercialization

SEQUENCING

COMPANION DIAGNOSTICS

CLINICAL TRIAL MATCHING & SITE ACTIVATION

MULTI-MODAL DATA FOR R&D

We've partnered with **2/3 of the top 20 pharma oncology companies*** as well as **50+ biopharma companies.**

*Based on publically available 2020 oncology segment revenue

LEARN MORE AT
tempus.com/life-sciences

joint membership from Cincinnati Children's Hospital Medical Center (CCHMC) and Boston Children's Hospital (BCH) – represented both a symbolic and practical marker in redressing this imbalance. This change empowered the Network to pursue not only phenotypes more common to children (e.g., ADHD, asthma), but also to assess the etiological/age spectrum of diseases more traditionally focused on adult populations (e.g., clostridioides difficile (C-Diff) and venous thromboembolism (VTE)), illustrated in **Table 1**.

Table 1: eMERGE Phase 2 Phenotypes involving Pediatric Sites

	CAG Controls	CAG Cases	Total Controls	Total Cases
ADHD**	11,438	1,013	15,107	1,264
Asthma**	9,740	4,498	20,134	8,178
Atopic Dermatitis	8,072	1,695	16,643	2,078
GERD**	15,216	1,573	20,102	4,343
Autism	4,610	155	7694	155
Childhood Obesity	101	99	223	527
C. Diff	165	178	10,437	1,909
Extreme Obesity	42	2	1,293	7,239
VTE	469	140	22,984	17,960
All Algorithms involving Pediatrics	49,853	9,269	106,647	40,489

Note: Table only includes algorithms with pediatric data.
Cells are not mutually exclusive (e.g., the same person may be a control for ADHD and asthma).
**CAG-led algorithms.

Pharmacogenomics (PGx) and Clinical Decision Support (CDS)

The eMERGE PGx project genotyped drug-naive patients/participants, with relevant data returned to participants' EMRs as part of a nascent preemptive strategy. In a real-world demonstration of the potential of pro-active individualized medicine study, those participants who were subsequently prescribed one of 11 medications were consequently able to benefit from custom clinical-decision support genomic data, including information that could alter dosing or medication choice. A 2014 study from the CAG team provides an interesting example of the immediate readiness and applicability of the potential power of a preemptive approach to current clinical care (see **Case Study: TPMT**)

Pharmacogenomics and mitigating severe adverse events due to risk variants

Genome-wide data is increasingly accessible to healthcare providers, and methods that demonstrate accurate imputation of genotypes not directly probed by given arrays can further enhance healthcare decisions.[27] This individualized approach can be used as a model for the PGx in clinical practice,[28] and represents a potentially powerful too in mitigating adverse events. This is all the more pertinent given that severe adverse events (SAEs) following medication remain a major cause of morbidity and mortality in children. Increasingly genetic variants are being included in FDA labels as mediators of both efficacy and toxicity. Nevertheless, testing can still be expensive, time-consuming, and often requires an *a-priori* understanding of relevant genomic factors. An important lesson from CAG's TPMT study is that the widespread application of commercial panels makes it possible to prophylactically identify risk variants accurately, representing, at the very least, cost-effective supplement to existing PGx approaches to healthcare.

Gene Discovery related to Attention Deficit/Hyperactivity Disorder (ADHD)

CAG developed an ADHD electronic algorithm that was implemented and validated by Boston Children's and Cincinnati Children's hospitals. The positive predictive values, (PPV) 89% and 95% for cases and controls, respectively, and inclusion/exclusion modifications following external validation have been re-incorporated into the algorithm, increasing PPVs to 95%/96% for cases/controls in the CAG repository.

In a large-scale, genome-wide study comparing copy number variants (CNVs) in ADHD cases vs. controls, CAG investigators identified rare, recurring CNVs impacting specific GRM genes (i.e., *GRM1, GRM5, GRM7*, and *GRM8*) encoding for metabotropic glutamate receptors (mGluRs) in ADHD patients at significantly higher frequencies compared to healthy controls.[29] The large effect sizes (with odds ratio (OR) of >15) suggest that these mutations are likely highly penetrant for their effects on ADHD. Single cases with *GRM2* and *GRM6* deletions were also observed that were not found in controls.

When genes in the signaling pathway of GRM genes (i.e., a GRM/mGluR-network) were assessed, significant enrichment of CNVs was found to reside within this network in ADHD cases compared to controls. Collectively, disruptions of genes in this network (as determined by the presence of CNVs), account for ~12% of total ADHD cases when corrected for control occurrence. These results suggest that disrupted mGluR signaling/activity is involved in ADHD in a subset of patients that can be identified based on genetic profiling. Positive findings were further evaluated/replicated in multiple independent cohorts, totaling 3,500 ADHD cases and ~13,000 controls of European ancestry, with respective case-control cohorts genotyped on matched platforms. CNVs impacting seven out of eight metabotropic glutamate receptor (mGluR) genes were significantly enriched across all independent cohorts ($P \leq 2.1 \times 10^{-9}$).

Expanding the Bedside to Bench Potential

Leveraging the asthma phenotype algorithm developed in Phase 2, CAG researchers applied the rule set to all EMRs in the CAG biobank. Previous genome-wide associate studies (GWAS) of circulating vitamin D levels had identified a highly significant ($p < 1 \times 10^{-49}$) association with variants with the group-specific component (Gc, vitamin D binding) protein;[30] these results were used in a

27. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11(7):499-511. doi: 10.1038/nrg2796. PubMed PMID: 20517342.
28. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med. 2013. Epub 2013/06/08. doi: gim201372 [pii] 10.1038/gim.2013.72 [doi]. PubMed PMID: 23743551.
29. Elia J, Glessner JT, Wang K, Takahashi N, Shtir CJ, Hadley D, et al. Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. Nat Genet. 2012;44(1):78-84. Epub 2011/12/06. doi: ng.1013 [pii] 10.1038/ng.1013 [doi]. PubMed PMID: 22138692.
30. Ahn J, Yu K, Stolzenberg-Solomon R, Simon KC, McCullough ML, Gallicchio L, et al. Genome-wide association study of circulating vitamin D levels. Hum Mol Genet. 2010;19(13):2739-45. Epub 2010/04/27. doi: 10.1093/hmg/ddq155. PubMed PMID: 20418485; PubMed Central PMCID: 2883344.


Discover.
Create.
Accelerate.
Together Transforming Companion Diagnostics
1
Discovery
• Biomarker discovery expertise
• Platform independent multi-omics
• Bioinformatics support
2
Clinical
• Flexible assay design with quality control
• History in clinical trials and testing
• cGMP production capabilities
3
Regulatory
• Experienced CLIA laboratories
• 510(k), US & inter-national registration
• IVDR-ready
4
Commercialization
• Global capabilities and channel partnership opportunities
• Easily kit products for commercial use
• Portfolios with well-established reimbursement and CPT codes
bio-techne®
Learn more about Bio-Techne's end-to-end quality CDx solutions.

Mendelian randomization (MR) experiment to determine the true effect of vitamin D levels on asthma predisposition and severity.

Data from the Genotype-Tissue Expression Project (GTEx) consortium indicated that GC is expressed exclusively in liver. Using the GWAS data and liver expression quantitative trait loci (eQTL) data(31), the research group identified GC variants that were significantly associated with GC expression and circulating vitamin D levels to derive a polygenic risk score (PRS) for each individual in the database. The PRSs were subsequently regressed against asthma status and severity scores to determine the causal effect of circulating vitamin D levels on asthma. Critically, any number of variations in input data, such as gene expression, gene regulation, differential expression in a tissue between disease and control, could be used to determine observed functional effect on a phenotype.

Phase 3 (P3: 2015–2020):

eMERGE Phase 3 amplified the theme of returning genomics-informed results to consortium participants with more direct application to clinical care. Nine clinical sites collectively recruited 25,000 individuals who were assessed for rare variants, which were integrated into participants' EMRs and supplemented with a range of community resources.

This large-scale project represents an important real-world application of genomics-informed individualized medicine, which includes design, reporting, and outcomes research with coordination and harmonization across sites with different populations, health systems, and expertise. Technical requirements for this project include standardization of DNA capture, data harmonization and management, interpretation of results, and consensus on reporting metrics and methodology (review at[32]).

All eMERGE Phase 3 sites collectively developed a gene panel – eMERGEseq (https://emerge-network.org/the-emergeseq-platform/) – consisting of 109 genes and 1,551 single-nucleotide variant (SNV) sites, with priorities delineated by potential actionability of findings and local research interests.[32] The final total of 109 genes reflects the 56 original genes listed as actionable by original American College of Medical Genetics and Genomics (ACMG)[33] plus 6 nominated by each site. Relevant single-nucleotide variants (SNVs) included were: (1) ancestry-informative markers and QC/fingerprinting loci (n = 425), (2) HLA-informative (n = 272), (3) pathogenic return-of-results (RoR) SNVs in genes not included on the 109 gene panel (n = 14), (4) pathogenic or likely pathogenic SNVs in genes included for research only (not RoR) (n = 55), (5) SNVs related to site-specific discovery efforts (n = 718), and (6) pharmacogenomic variants (n = 125).

"A particularly strong component of CAG's approach is its highly-developed system and process for delivering electronic CDS to clinicians. This allows CAG to customize content in creative ways that make it instructive to clinicians who may not be experts in interpreting genomic information."

Variant classification was based on ACMG/Association of Medical Pathology criteria[34] with modifications established by an expert panel. -ACMG-56 genes required extensive clinical curation using the ClinGen framework for gene-disease validity assessment,[35] as well as assessment by the Clinical Annotation workgroup. The PGx workgroup selected variants to be included in the clinical reports provided to participants guided by Clinical Pharmacogenomics Implementation Consortium (CPIC) guidelines.[36] Twenty PGx variants across seven genes were included.

A total of 8,437,788 variants were detected in the 25,015 participants analyzed with the eMERGEseq panel. Extensive exclusion and filtration processed yielded 9,653 unique variants requiring further assessment. After expert review, these were further categorized as benign (1%), likely benign (8%), variant of uncertain significance (VUS, 69%), likely pathogenic (LP) (7%), pathogenic (P) (12%), or penetrance risk alleles (0.5%). A further 205 unique copy number variants were detected, of which 30% were deemed reportable. These data led to a total of 1,497 participants that have an LP/P variant that would require a positive report to be issued. Furthermore, 23,232 participants – or 93% of the cohort – had results indicative of at least one non-standard drug dosing in one of the 7 genes tested. These seven genes collectively encompass only 11 drug types, and again, the effort of finding these genes is justified by the massive potential for large-scale PGx testing to facilitate individualized guidelines for individuals across the healthcare spectrum.

P3 Legacy

Template for Genomic Medicine:

The implementation of clinical sequencing into Phase 3 provides a template for the large-scale clinical translation of genomic data in healthcare promised by the human genome project (see, e.g.,[37-39]). Further, the consortium's focus on integrating results with EMRs provides an important model for clinical data management, while implementation across sites with diverse populations and EMR systems resolves many questions on the "deliver-ability" of genomics results to highly heterogenous healthcare systems. Relevant outcomes continue to be tracked at respective sites and span the gamut of healthcare utilization including additional test order, new medications, and new procedures.

Clinical Decision Support (CDS):

A particularly strong component of CAG's approach is its highly-developed system and process for delivering electronic CDS to clinicians. This allows CAG to customize content in creative ways that make it instructive to clinicians who may not be experts in interpreting genomic information. The process also serves as a decision

31. Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 2011;7(5):e1002078. Epub 2011/06/04. doi: 10.1371/journal.pgen.1002078. PubMed PMID: 21637794; PubMed Central PMCID: 3102751.
32. Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network. Am J Hum Genet. 2019;105(3): 588-605. Epub 20190822. doi: 10.1016/j.ajhg.2019.07.018. PubMed PMID: 31447099; PubMed Central PMCID: PMC6731372.
33. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15(7):565-74. Epub 20130620. doi: 10.1038/gim.2013.73. PubMed PMID: 23788249; PubMed Central PMCID: PMC3727274.
34. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24. Epub 20150305. doi: 10.1038/gim.2015.30. PubMed PMID: 25741868; PubMed Central PMCID: PMC4544753.
35. Strande NT, Riggs ER, Buchanan AH, Ceyhan-Birsoy O, DiStefano M, Dwight SS, et al. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. Am J Hum Genet. 2017;100(6):895-906. Epub 20170525. doi: 10.1016/j.ajhg.2017.04.015. PubMed PMID: 28552198; PubMed Central PMCID: PMC5473734.
36. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clin Pharmacol Ther. 2011;89(3):464-7. Epub 20110126. doi: 10.1038/clpt.2010.279. PubMed PMID: 21270786; PubMed Central PMCID: PMC3098762.
37. van Ommen GJ, Bakker E, den Dunnen JT. The human genome project and the future of diagnostics, treatment, and prevention. Lancet. 1999;354 Suppl 1:Si5-10. doi: 10.1016/s0140-6736(99)90241-6. PubMed PMID: 10437848.
38. Chakravarti A. To a future of genetic medicine. Nature. 2001;409(6822):822-3. doi: 10.1038/35057281. PubMed PMID: 11236997.
39. Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011;470(7333):198-203. doi: 10.1038/nature09796. PubMed PMID: 21307932.

support conduit for clinical staff to communicate materials and guidance to patients and families. CAG carefully designs the process to account for the clinical and communication requirements of physicians and parents.

The system is capable of delivering summarized, highly-visible results to highlight clinically actionable variants, as well as one-click access to full clinical lab report that provides all variant findings. This system (architecture, interface, etc.) provides several options for messaging clinicians and clinical staff (messaging, paging etc.). Importantly, it provides effective decision support to clinicians that accounts not only for content but also context and manner in which information is presented,[40,41] particularly for CDS at the point of care. The technical and governance infrastructure developed in this work will be reusable for other genomic variants that impact medication dosing and provide a set of components for integrating actionable variants elsewhere in clinical workflows as well.

PGx Innovation:

Building on the groundbreaking work from Phase 2, Sood *et al.*[42] led a prospective study to identify patients carrying important variants pharmacogenomically relevant to asthma intervention. Asthma is the leading chronic disease in children. Several studies have identified genetic biomarkers associated with susceptibility and severity in both adult and pediatric cases.

As part of Phase 3, CAG evaluated health outcomes in 400 African American and European American pediatric cases, all of whom were regular users of inhaled corticosteroids. Patients were stratified by genotype using two single nucleotide polymorphisms (SNPs) in the beta-2 adrenergic receptor (ADRB2) gene – rs1042713 and rs1042714 – previously associated with asthma outcome. These correspond to nonsynonymous SNPs at positions 16 [arginine to glycine (Arg16Gly); rs1042713] and 27 [glutamic acid to glutamine (Glu27Gln); rs1042714] that are relatively common (minor allele frequencies ~40–50%) and have been well characterized in asthma PGx. The study controlled for adherence to National Heart, Lung and Blood Institute (NHLBI) guidelines using deep mining of EMRs to assess previous treatments. While no significant effect for rs1042713 (Arg16Gly) was identified, the study showed that participants homozygous for Gln27 had increased exacerbations while taking inhaled corticosteroids. This demonstrated for the first time that the Glu27 variant in ADRB2 is associated with increased frequencies of asthma exacerbations.

> *"The technical and governance infrastructure developed in this work will be reusable for other genomic variants that impact medication dosing and provide a set of components for integrating actionable variants elsewhere in clinical workflows as well."*

Importantly, this study follows previous eMERGE-funded efforts at CAG (e.g., TPMT-thiopurines, Case Study Inset) to illustrate how EMRs linked to genotypes can systematically be used to delineate health outcomes. By showing associations between genotype and response to inhaled corticosteroid treatment as reflected in frequency of asthma exacerbations for asthma, this study presents a model whereby systematic data-mining of biobanked samples coupled to EMRs can be used to identify genotype-health outcome associations and inform treatment. This example illustrates how data-mining of de-identified EMR data from unbiased large-scale recruitment and biobanking can successfully inform PGx effects and clinical decisions.

Geocoding

Although environmental exposures from fetal life through adulthood have persistent effects on both overall health and socioeconomic status,[43,44] a key knowledge gap has been the interaction of life exposures in terms of long-term health outcomes. For example, respiratory diseases such as upper respiratory tract infections, otitis media, tuberculosis, pneumonia, bronchiectasis, and allergies have all been linked to specific exposures in the environment,[45-55] all of which is available in existing eMERGE datasets as discrete EMR-derived outcomes.

As opposed to *bottom-up* approaches to deep-phenotyping, 'big-data' bioinformatics offers a complimentary, *top-down* mechanism for probing the environmental variables through geocoding. Geocoding simply refers to the process of identifying a geographic location from an address. A wide range of data can then be derived from those geolocations, such as directions or proximity associations. Examples include street distances to major roadways, toxic release inventory (TRI) sites, waste operation facilities, and nearby air emission plants, as well as, neighborhood traffic volume and road density, proximity to food deserts, clinics, recreational facilities, and healthcare facilities. Most commonly, studies map the locations to census tracts to derive demographic and socioeconomic variables from the census, community and crime surveys, and true environmental variables, such as air quality.

Using supplementary Phase 3 funding, CAG led a consortium-wide effort to perform geocoding across all samples in the eMERGE consortium.

40. Campbell R, CPHIMS C. The five" rights" of clinical decision support. Journal of AHIMA/American Health Information Management Association. 2013;84(10):42-7; quiz 8.
41. Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ: British Medical Journal. 2013;346.
42. Sood N, Connolly JJ, Mentch FD, Vazquez L, Sleiman PMA, Hysinger EB, et al. Leveraging electronic health records to assess the role of ADRB2 single nucleotide polymorphisms in predicting exacerbation frequency in asthma patients. Pharmacogenet Genomics. 2018;28(11):256-9. Epub 2018/10/20. doi: 10.1097/fpc.0000000000000355. PubMed PMID: 30334910.
43. Case A, Fertig A, Paxson C. The lasting impact of childhood health and circumstance. J Health Econ. 2005;24(2):365-89. doi: 10.1016/j.jhealeco.2004.09.008. PubMed PMID: 15721050.
44. Smith JP. The Impact of Childhood Health on Adult Labor Market Outcomes. Rev Econ Stat. 2009;91(3):478-89. doi: 10.1162/rest.91.3.478. PubMed PMID: 23585697; PubMed Central PMCID: PMC3625038.
45. Hodge L, Salome CM, Peat JK, Haby MM, Xuan W, Woolcock AJ. Consumption of oily fish and childhood asthma risk. Med J Aust. 1996;164(3):137-40. PubMed PMID: 8628130.
46. Weiland SK, Husing A, Strachan DP, Rzehak P, Pearce N, Group IPOS. Climate and the prevalence of symptoms of asthma, allergic rhinitis, and atopic eczema in children. Occup Environ Med. 2004;61(7):609-15. PubMed PMID: 15208377; PubMed Central PMCID: PMC1740799.
47. Gent JF, Triche EW, Holford TR, Belanger K, Bracken MB, Beckett WS, et al. Association of low-level ozone and fine particles with respiratory symptoms in children with asthma. JAMA. 2003;290(14):1859-67. doi: 10.1001/jama.290.14.1859. PubMed PMID: 14532314.
48. Aekplakorn W, Loomis D, Vichit-Vadakan N, Shy C, Wongtim S, Vitayanon P. Acute effect of sulphur dioxide from a power plant on pulmonary function of children, Thailand. Int J Epidemiol. 2003;32(5):854-61. PubMed PMID: 14559764.
49. Etzel RA, Montana E, Sorenson WG, Kullman GJ, Allan TM, Dearborn DG, et al. Acute pulmonary hemorrhage in infants associated with exposure to Stachybotrys atra and other fungi. Arch Pediatr Adolesc Med. 1998;152(8):757-62. PubMed PMID: 9701134.
50. Brooks WA, Yunus M, Santosham M, Wahed MA, Nahar K, Yeasmin S, et al. Zinc for severe pneumonia in very young children: double-blind placebo-controlled trial. Lancet. 2004;363(9422):1683-8. doi: 10.1016/S0140-6736(04)16252-1. PubMed PMID: 15158629.
51. Smith AH, Marshall G, Yuan Y, Ferreccio C, Liaw J, von Ehrenstein O, et al. Increased mortality from lung cancer and bronchiectasis in young adults after exposure to arsenic in utero and in early childhood. Environ Health Perspect. 2006;114(8):1293-6. PubMed PMID: 16882542; PubMed Central PMCID: PMC1551995.
52. Klonoff-Cohen HS, Edelstein SL, Lefkowitz ES, Srinivasan IP, Kaegi D, Chang JC, et al. The effect of passive smoking and tobacco exposure through breast milk on sudden infant death syndrome. JAMA. 1995;273(10):795-8. PubMed PMID: 7861574.
53. Mitchell EA, Ford RP, Stewart AW, Taylor BJ, Becroft DM, Thompson JM, et al. Smoking and the sudden infant death syndrome. Pediatrics. 1993;91(5):893-6. PubMed PMID: 8474808.
54. Schoendorf KC, Kiely JL. Relationship of sudden infant death syndrome to maternal smoking during and after pregnancy. Pediatrics. 1992;90(6):905-8. PubMed PMID: 1437432.
55. Landrigan PJ, Schechter CB, Lipton JM, Fahs MC, Schwartz J. Environmental pollutants and disease in American children: estimates of morbidity, mortality, and costs for lead poisoning, asthma, cancer, and developmental disabilities. Environ Health Perspect. 2002;110(7):721-8. PubMed PMID: 12117650; PubMed Central PMCID: PMC1240919.

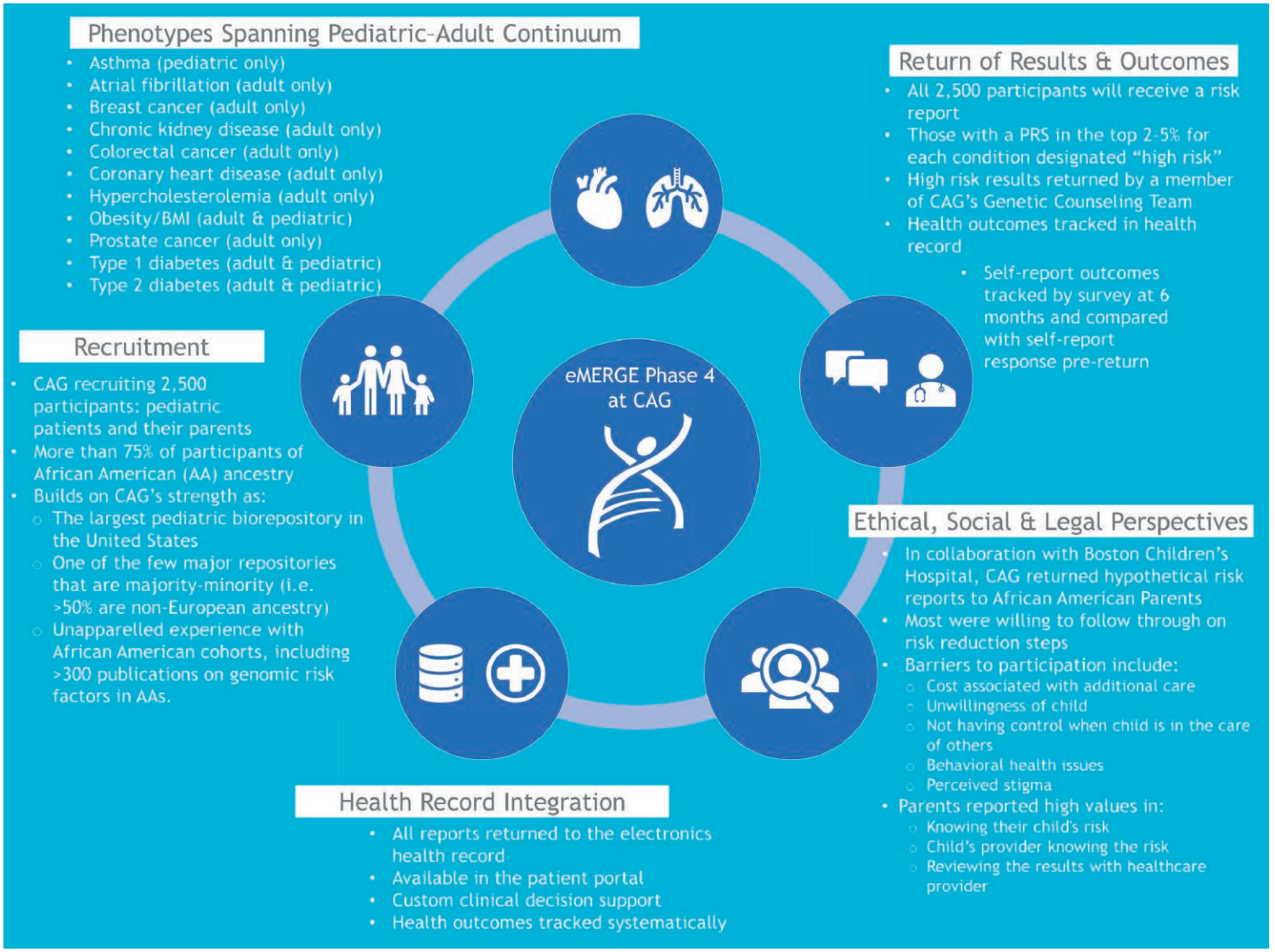


Figure 1: CAG's eMERGE Phase 4 program

CAG is the largest biorepository in the United States in which the majority of samples are from racial or ethnic minorities, primarily African Americans (AAs). The repository (n>500,000) includes more than 100,000 CHOP participants with linked electronic medical record and genome-wide genotype data, of whom >50,000 are AA. CAG's scientific output includes >300 peer-reviewed publications on genomic risk factors in AAs. Moreover, 43% of its contribution to the current eMERGE program includes data from AA subjects (n=5,640), while 90% of eMERGE whole genome sequence data (814 of 900) are from AAs.

This allowed the Network to derive environmental exposures as a potential way to power gene-environment (GxE) interaction analyses across sites. To use the asthma phenotype as example, the Network was able to derive air quality measures from the geocoding data to enable both genome-wide conditional analysis on air quality as well as a gene environment interaction analysis to identify novel asthma susceptibility variants. These data are publicly available along with all eMERGE genotypes (Phases 1-3, N> 100,000) in dbGaP (accession number: phs001584.v2. p2).

Importantly, the geocoding proximity approach offers a highly cost-effective and scalable means of addressing critical data-points often absent from traditional EMR–derived data, including environmental toxins such as heavy metals, organics and air pollution, all of which affect development and particularly neurologic development.[56-63]

Phase 4 (P4: July 2020–May 2026)

eMERGE 4 is primarily focused on developing, validating, and implementing polygenic risk scores

56. Walkowiak J, Wiener JA, Fastabend A, Heinzow B, Kramer U, Schmidt E, et al. Environmental exposure to polychlorinated biphenyls and quality of the home environment: effects on psychodevelopment in early childhood. Lancet. 2001;358(9293):1602-7. doi: 10.1016/S0140-6736(01)06654-5. PubMed PMID: 11716887.
57. Bobak M. Outdoor air pollution, low birth weight, and prematurity. Environ Health Perspect. 2000;108(2):173-6. PubMed PMID: 10656859; PubMed Central PMCID: PMC1637893.
58. Eskenazi B, Marks AR, Bradman A, Fenster L, Johnson C, Barr DB, et al. In utero exposure to dichlorodiphenyltrichloroethane (DDT) and dichlorodiphenyldichloroethylene (DDE) and neurodevelopment among young Mexican American children. Pediatrics. 2006;118(1):233-41. doi: 10.1542/peds.2005-3117. PubMed PMID: 16818570.
59. Grandjean P, Weihe P, White RF, Debes F, Araki S, Yokoyama K, et al. Cognitive deficit in 7-year-old children with prenatal exposure to methylmercury. Neurotoxicol Teratol. 1997;19(6):417-28. PubMed PMID: 9392777.
60. Mylchreest E, Cattley RC, Foster PM. Male reproductive tract malformations in rats following gestational and lactational exposure to Di(n-butyl) phthalate: an antiandrogenic mechanism? Toxicol Sci. 1998;43(1):47-60. doi: 10.1006/toxs.1998.2436. PubMed PMID: 9629619.
61. Yu XD, Yan CH, Shen XM, Tian Y, Cao LL, Yu XG, et al. Prenatal exposure to multiple toxic heavy metals and neonatal neurobehavioral development in Shanghai, China. Neurotoxicol Teratol. 2011;33(4):437-43. doi: 10.1016/j.ntt.2011.05.010. PubMed PMID: 21664460.
62. Wang SL, Chang FH, Liou SH, Wang HJ, Li WF, Hsieh DP. Inorganic arsenic exposure and its relation to metabolic syndrome in an industrial area of Taiwan. Environ Int. 2007;33(6):805-11. doi: 10.1016/j.envint.2007.03.004. PubMed PMID: 17481731.

(PRSs) for ten common diseases (listed in **Figure 1**, upper left corner (colorectal cancer evaluated only for monogenic risk)). Critically, the program will target minority and underserved populations; these groups have benefited comparatively less from genomics research to date. At CAG, the team's focus is on individuals of African ancestry, who will constitute >75% of participants. This work will build on a research program with a focus on diversity that has been foundational to its research agenda (illustrated in **Figure 1**).

Collectively, the Network aims to implement PRSs for 25,000 individuals – 2,500 at each of 10 sites. These PRSs will help patients and providers learn about their risk of developing specific diseases and generate recommendations to improve health outcomes. In order to deliver genomics-guided healthcare using PRS, the validity of the scores and phenotypes will be thoroughly assessed. The current inequalities in genomics systems, which are biased toward individuals of European ancestry, will be directly augmented. Limitations in genomics education and understanding among health care providers, as well as the related EMR systems that are not optimized for modern genomic sciences, will be carefully evaluated and supported through the eMERGE 4 initiatives. Special attention will be applied to tracking consent and long-term outcomes of pediatric cohorts, particularly ones of enhanced diversity where emphasis needs to be placed on equalizing healthcare opportunities.

P4 Legacy and Plans

Phenotypes:

We (Dr Hakonarson and his CAG investigator team) have worked extensively on PRSs for almost a decade. During that time, we have published numerous papers on PRS across multiple phenotypes.[64-72] PRSs for a range of phenotypes may be applicable to patient care and have the potential to outperform clinical predictors for several diseases, including but not limited to breast cancer, type 1 diabetes, and prostate cancer.[73-75] Many such predictors are most directly relevant to individuals of European ancestry,[76-84] who constitute a large proportion of research participants in the United States.

However, PRSs are only as valid as the GWASs on which they are based, and currently ~79% of GWASs are based on individuals of European ancestry – reflecting well-documented biases in research recruitment and infrastructure.

63. Wasserman GA, Liu X, Parvez F, Ahsan H, Factor-Litvak P, van Geen A, et al. Water arsenic exposure and children's intellectual function in Araihazar, Bangladesh. Environ Health Perspect. 2004;112(13):1329-33. PubMed PMID: 15345348; PubMed Central PMCID: PMC1247525.
64. Germine L, Robinson EB, Smoller JW, Calkins ME, Moore TM, Hakonarson H, et al. Association between polygenic risk for schizophrenia, neurocognition and social cognition across development. Transl Psychiatry. 2016;6(10):e924. Epub 2016/10/19. doi: 10.1038/tp.2016.147. PubMed PMID: 27754483; PubMed Central PMCID: PMC5315539.
65. Maier R, Moser G, Chen GB, Ripke S, Coryell W, Potash JB, et al. Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. Am J Hum Genet. 2015;96(2):283-94. Epub 2015/02/03. doi: 10.1016/j.ajhg.2014.12.006. PubMed PMID: 25640677; PubMed Central PMCID: PMC4320268.
66. Hamshere ML, Langley K, Martin J, Agha SS, Stergiakouli E, Anney RJ, et al. High loading of polygenic risk for ADHD in children with comorbid aggression. Am J Psychiatry. 2013;170(8):909-16. Epub 2013/04/20. doi: 10.1176/appi.ajp.2013.12081129. PubMed PMID: 23599091; PubMed Central PMCID: PMC3935265.
67. Escott-Price V, Shoai M, Pither R, Williams J, Hardy J. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. Neurobiol Aging. 2017;49:214.e7-.e11. Epub 2016/09/07. doi: 10.1016/j.neurobiolaging.2016.07.018. PubMed PMID: 27595457.
68. Axelrud LK, Santoro ML, Pine DS, Talarico F, Gadelha A, Manfro GG, et al. Polygenic Risk Score for Alzheimer's Disease: Implications for Memory Performance and Hippocampal Volumes in Early Life. Am J Psychiatry. 2018;175(6):555-63. Epub 2018/03/03. doi: 10.1176/appi.ajp.2017.17050529. PubMed PMID: 29495896; PubMed Central PMCID: PMC6372950.
69. Groen-Blokhuis MM, Middeldorp CM, Kan KJ, Abdellaoui A, van Beijsterveldt CE, Ehli EA, et al. Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. J Am Acad Child Adolesc Psychiatry. 2014;53(10):1123-9.e6. Epub 2014/09/24. doi: 10.1016/j.jaac.2014.06.014. PubMed PMID: 25245356.
70. Morgan AR, Touchard S, O'Hagan C, Sims R, Majounie E, Escott-Price V, et al. The Correlation between Inflammatory Biomarkers and Polygenic Risk Score in Alzheimer's Disease. Journal of Alzheimer's disease : JAD. 2017;56(1):25-36. Epub 2016/12/03. doi: 10.3233/jad-160889. PubMed PMID: 27911318.
71. Yilmaz Z, Schaumberg K, Crowley J, Breen G, Bulik C, Micali N, et al. Anorexia Nervosa and Obsessive Compulsive Poygenic Risk Score: Associations with Adolesent Eating Disorder Phenotypes. European Neuropsychopharmacology. 2017;29(3):S818-9. doi: https://doi.org/10.1016/j.euroneuro.2017.08.065.
72. Guo Y, Wei Z, Keating BJ, Hakonarson H. Machine learning derived risk prediction of anorexia nervosa. BMC Med Genomics. 2016;9:4. Epub 2016/01/23. doi: 10.1186/s12920-016-0165-x. PubMed PMID: 26792494; PubMed Central PMCID: PMC4721143.
73. Maas P, Barrdahl M, Joshi AD, Auer PL, Gaudet MM, Milne RL, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. JAMA oncology. 2016;2(10):1295-302. Epub 2016/05/27. doi: 10.1001/jamaoncol.2016.1025. PubMed PMID: 27228256; PubMed Central PMCID: PMC5719876.
74. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 2018;50(7):928-36. Epub 2018/06/13. doi: 10.1038/s41588-018-0142-8. PubMed PMID: 29892016; PubMed Central PMCID: PMC6568012.
75. Sharp SA, Rich SS, Wood AR, Jones SE, Beaumont RN, Harrison JW, et al. Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis. Diabetes Care. 2019;42(2):200-7. Epub 2019/01/19. doi: 10.2337/dc18-1785. PubMed PMID: 30655379; PubMed Central PMCID: PMC6341291.
76. Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am J Hum Genet. 2015;97(4):576-92. Epub 2015/10/03. doi: 10.1016/j.ajhg.2015.09.001. PubMed PMID: 26430803; PubMed Central PMCID: PMC4596916.
77. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511(7510):421-7. Epub 2014/07/25. doi: 10.1038/nature13595. PubMed PMID: 25056061; PubMed Central PMCID: PMC4112379.
78. Belsky DW, Moffitt TE, Sugden K, Williams B, Houts R, McCarthy J, et al. Development and evaluation of a genetic risk score for obesity. Biodemography and social biology. 2013;59(1):85-100. Epub 2013/05/25. doi: 10.1080/19485565.2013.774628. PubMed PMID: 23701538; PubMed Central PMCID: PMC3671353.
79. Domingue BW, Belsky D, Conley D, Harris KM, Boardman JD. Polygenic Influence on Educational Attainment: New evidence from The National Longitudinal Study of Adolescent to Adult Health. AERA open. 2015;1(3):1-13. Epub 2015/07/01. doi: 10.1177/2332858415599972. PubMed PMID: 28164148; PubMed Central PMCID: PMC5291340.
80. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50(8):1112-21. Epub 2018/07/25. doi: 10.1038/s41588-018-0147-3. PubMed PMID: 30038396; PubMed Central PMCID: PMC6393768.
81. Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, et al. An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. Biol Psychiatry. 2017;81(6):470-7. Epub 2016/10/22. doi: 10.1016/j.biopsych.2016.06.028. PubMed PMID: 27765268.
82. Akiyama M, Okada Y, Kanai M, Takahashi A, Momozawa Y, Ikeda M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. Nat Genet. 2017;49(10):1458-67. Epub 2017/09/12. doi: 10.1038/ng.3951. PubMed PMID: 28892062.
83. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet. 2019;51(5):793-803. Epub 2019/05/03. doi: 10.1038/s41588-019-0397-8. PubMed PMID: 31043756.
84. Li Z, Chen J, Yu H, He L, Xu Y, Zhang D, et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nat Genet. 2017;49(11):1576-83. Epub 2017/10/11. doi: 10.1038/ng.3973. PubMed PMID: 28991256.

This bias is clearly problematic and further magnified by European-centric content of genotyping arrays and reference panels.[85-87] In the US, where ancestry and health disparities already overlap markedly, and where real and perceived health disparities are higher than other high- and middle-income countries,[88,89] it is especially important that PRS-linked phenotypes be validated in under-represented cohorts. With over 50% of its recruited samples from minority populations, CAG has the largest minority-based biorepository in the United States Hence, CAG is uniquely equipped to address this issue, as reflected in multiple publications by its team on the genetic landscape in African Americans (AAs, see, e.g.,[90-98]). This point is exemplified in eMERGE phenotypes driven by CAG, including asthma and diabetes.

Asthma

Asthma is the most common chronic disease in children and young adults with prevalence as high as 10% in the general population. Through eMERGE, CAG developed an asthma algorithm (including severity), implemented by all sites, with total accrual of 8,178 cases and 20,134 controls and positive predictive values >94%. This analysis powered a GWAS that identified 4 novel loci, including the *TGES* gene in AAs only, which encodes TIE2, a protein that has been shown to be involved in remodeling the airway wall in asthma and remained significant after conditioning by allergy.[90] CAG also highlighted a variant in the *ADRB2* that predicted increased exacerbations for patients taking inhaled corticosteroids,[42] and a Mendelian randomization analysis demonstrating low vitamin D is unlikely causative for pediatric asthma.[99]

"While no significant effect for rs1042713 (Arg16Gly) was identified, the study showed that participants homozygous for Gln27 had increased exacerbations while taking inhaled corticosteroids. This demonstrated for the first time that the Glu27 variant in ADRB2 is associated with increased frequencies of asthma exacerbations."

CAG has published more than 60 papers on genomic correlates of asthma, including the first common genetic association in children at *DENND1B*.[97] Intriguingly, the same variant is associated with an increased risk of asthma in AA children but a decreased risk in children of European ancestry, again showing the limitations of a "one size fits all" approach in complex disease.

Risk Management:

PRS could help guide asthma therapy, particularly in young children where the diagnosis is less certain, and the benefit of long-term preventive therapy vs intermittent therapy is less clear. In adults, PRS could help guide the level of therapy needed to prevent the development of an irreversible chronic obstructive lung disease (COPD). Chronic inhaled steroids therapy in children is effective in preventing exacerbations but can increase risk of long-term side effects from growth retardation.[100] Conversely, if therapy is inadequate, these children are at risk of developing chronic airway remodeling setting them up for greater risk of chronic lung disease later in life, particularly if they become smokers. There is a growing body of evidence showing that AA ancestry predisposes to more severe asthma, supporting CAG's recruitment approach.[101]

PRS Validation:

Polygenic risk scores for asthma were among the first developed;[102] in children, these scores can predict those who are likely to remit versus those requiring lifelong treatment. Furthermore, PRS has the potential to predict long-term, real work implications (e.g., school and work absenteeism, hospital admissions).[102] A small-scale GWAS based study previously showed significant polygenic prediction of lung function[103] and improved prediction of other phenotypes.[104,105] For example, wheezing is common in infants; a validated »

85. Carroll MD, Kit BK, Lacher DA, Shero ST, Mussolino ME. Trends in lipids and lipoproteins in US adults, 1988-2010. Jama. 2012;308(15):1545-54. Epub 2012/10/18. doi: 10.1001/jama.2012.13260. PubMed PMID: 23073951.
86. Kit BK, Carroll MD, Lacher DA, Sorlie PD, DeJesus JM, Ogden C. Trends in serum lipids among US youths aged 6 to 19 years, 1988-2010. Jama. 2012;308(6):591-600. Epub 2012/08/09. doi: 10.1001/jama.2012.9136. PubMed PMID: 22871871.
87. Lim E, Miyamura J, Chen JJ. Racial/Ethnic-Specific Reference Intervals for Common Laboratory Tests: A Comparison among Asians, Blacks, Hispanics, and White. Hawai'i journal of medicine & public health : a journal of Asia Pacific Medicine & Public Health. 2015;74(9):302-10. Epub 2015/10/16. PubMed PMID: 26468426; PubMed Central PMCID: PMC4578165.
88. Hero JO, Zaslavsky AM, Blendon RJ. The United States Leads Other Nations In Differences By Income In Perceptions Of Health And Health Care. Health affairs (Project Hope). 2017;36(6):1032-40. Epub 2017/06/07. doi: 10.1377/hlthaff.2017.0006. PubMed PMID: 28583961.
89. Williams DR, Priest N, Anderson NB. Understanding associations among race, socioeconomic status, and health: Patterns and prospects. Health psychology : official journal of the Division of Health Psychology, American Psychological Association. 2016;35(4):407-11. Epub 2016/03/29. doi: 10.1037/hea0000242. PubMed PMID: 27018733; PubMed Central PMCID: PMC4817358.
90. Almoguera B, Vazquez L, Mentch F, Connolly J, Pacheco JA, Sundaresan AS, et al. Identification of Four Novel Loci in Asthma in European American and African American Populations. Am J Respir Crit Care Med. 2017;195(4):456-63. Epub 2016/09/10. doi: 10.1164/rccm.201604-0861OC. PubMed PMID: 27611488; PubMed Central PMCID: PMC5378422.
91. Almoguera B, Vazquez L, Mentch F, March ME, Connolly JJ, Peissig PL, et al. Novel locus for atopic dermatitis in African Americans and replication in European Americans. J Allergy Clin Immunol. 2019;143(3):1229-31. Epub 2018/11/12. doi: 10.1016/j.jaci.2018.10.038. PubMed PMID: 30414857.
92. Chang X, Pellegrino R, Garifallou J, March M, Snyder J, Mentch F, et al. Common variants at 5q33.1 predispose to migraine in African-American children. J Med Genet. 2018;55(12):831-6. Epub 2018/09/30. doi: 10.1136/jmedgenet-2018-105359. PubMed PMID: 30266756.
93. Grant S, Glessner JT, Bradfield JP, Zhao JH, Tirone JE, Berkowitz RI, et al. Relationship of Mitochondrial Heteroplasmy and Variation With Childhood Obesity in European and African Americans. Obesity. 2011;19:S233-S. PubMed PMID: WOS:000296603100849.
94. Grant SF, Petri, M., Bradfield, J.P., Kim, C.E., Santa, E., Annaiah, K., Frackelton, E.C., Glessner, J.T., Otieno, F.O., Shaner, J.L., Smith, r., Eckert, A.W., Chiavacci, R.M., Imielinski, M., Sullivan, K.E., & Hakonarson, H. Association of the BANK1 R61H variant with systemic lupus erythematosus in Americans of European and African ancestry. The Application of Clinical Genetics 2008;2009(2):5. doi: http://dx.doi.org/10.2147/TACG.S4089.
95. Latorre V, Diskin S, Diamond M, Zhang HT, Hakonarson H, Maris J, et al. Genetic Analysis of Neuroblastoma in African American Patients. Genet Epidemiol. 2010;34(8):968-9. PubMed PMID: WOS:000284719100194.
96. Li J, Wei Z, Zhang J, Hakonarson H, Cook-Sather SD. Candidate gene analyses for acute pain and morphine analgesia after pediatric day surgery: African American versus European Caucasian ancestry and dose prediction limits. Pharmacogenomics J. 2019. Epub 2019/02/15. doi: 10.1038/s41397-019-0074-4. PubMed PMID: 30760877.
97. Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, et al. Variants of DENND1B associated with asthma in children. N Engl J Med. 2010;362(1):36-44. Epub 2009/12/25. doi: 10.1056/NEJMoa0901867. PubMed PMID: 20032318.
98. Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, et al. The landscape of recombination in African Americans. Nature. 2011;476(7359):170-5. Epub 2011/07/22. doi: 10.1038/nature10336. PubMed PMID: 21775986; PubMed Central PMCID: PMC3154982.
99. Hysinger EB, Roizen JD, Mentch FD, Vazquez L, Connolly JJ, Bradfield JP, et al. Mendelian randomization analysis demonstrates that low vitamin D is unlikely causative for pediatric asthma. J Allergy Clin Immunol. 2016;138(6):1747-9 e4. Epub 2016/08/25. doi: 10.1016/j.jaci.2016.06.056. PubMed PMID: 27554823; PubMed Central PMCID: PMC5526345.
100. Hvizdos KM, Jarvis B. Budesonide inhalation suspension: a review of its use in infants, children and adults with inflammatory respiratory disorders. Drugs. 2000;60(5):1141-78. Epub 2000/12/29. doi: 10.2165/00003495-200060050-00010. PubMed PMID: 11129126.
101. Lee DS, Gross E, Hotz A, Rastogi D. Comparison of severity of asthma hospitalization between African American and Hispanic children in the Bronx. The Journal of asthma : official journal of the Association for the Care of Asthma. 2019:1-7. Epub 2019/05/08. doi: 10.1080/02770903.2019.1609981. PubMed PMID: 31062634.
102. Belsky DW, Sears MR, Hancox RJ, Harrington H, Houts R, Moffitt TE, et al. Polygenic risk and the development and course of asthma: an analysis of data from a four-decade longitudinal study. Lancet Respir Med. 2013;1(6):453-61. Epub 2014/01/17. doi: 10.1016/s2213-2600(13)70101-2. PubMed PMID: 24429243; PubMed Central PMCID: PMC3899706.
103. McGeachie MJ, Clemmer GL, Croteau-Chonka DC, Castaldi PJ, Cho MH, Sordillo JE, et al. Whole genome prediction and heritability of childhood asthma phenotypes. Immunity, inflammation and disease. 2016;4(4):487-96. Epub 2016/12/17. doi: 10.1002/iid3.133. PubMed PMID: 27980782; PubMed Central PMCID: PMC5134727.
104. Jansen PR, Petrus NCM, Venema A, Posthuma D, Mannens M, Sprikkelman AB, et al. Higher Polygenetic Predisposition for Asthma in Cow's Milk Allergic Children. Nutrients. 2018;10(11). Epub 2018/10/31. doi: 10.3390/nu10111582. PubMed PMID: 30373230; PubMed Central PMCID: PMC6266812.
105. Dashti HS, Redline S, Saxena R. Polygenic risk score identifies associations between sleep duration and diseases determined from an electronic medical record biobank. Sleep. 2019;42(3). Epub 2018/12/07. doi: 10.1093/sleep/zsy247. PubMed PMID: 30521049; PubMed Central PMCID: PMC6424085.

PRS has the potential to differentiate wheezing from an asthma prognosis, yielding important opportunities for an earlier, targeted intervention. The proposed risk management protocol will monitor critical variables to determine the most effective use of long-term asthma management with controller medications to minimize the use of rescue therapy for chronic asthma with primary focus on AA children.

Diabetes:

CHOP's Diabetes Center is a world-leading center for care and research and CAG has long been a leader in diabetes genetics, having led landmark studies of atypical, monogenic, type-1 (T1D) and type-2 (T2D) sub-types, including the largest T1D meta-analysis study reported.[106-113] CAG was the first group to identify T1DM-risk variants in *CLEC16A* and to describe the mechanism of this risk.[110]

Risk Management:

PRS could help guide diabetes diagnosis and therapy and differentiate between diabetes subtypes, including T1D, T2D and maturity onset diabetes of the young (MODY). While T1D is typically associated with autoantibodies, a subset of T1D patients have no autoantibodies and a large number of children develop diabetes-related autoantibodies without ever developing diabetes. PRS is anticipated to have an impact in classifying children without autoantibodies and with treatment selection (insulin injection vs insulin stimulating drugs).

T2D is a rapidly growing problem in teenagers, particularly those of AA descent.[114] Identifying those subjects at highest risk will allow for earlier interventions and reduce the risk of diabetes complications, such as metabolic syndrome, leg ulcers and chronic kidney disease – all of which are extremely costly to the healthcare system and devastating to the patients and their families. This problem is amplified in adults, also targeted for PRS-mediated intervention.

> *"Importantly, the geocoding proximity approach offers a highly cost-effective and scalable means of addressing critical data-points often absent from traditional EMR–derived data, including environmental toxins such as heavy metals, organics and air pollution, all of which affect development and particularly neurologic development."*

CAG's proposed risk management protocols will consider a host of patient data sets, including family history, physical examination observations, and critical laboratory results to determine the most effective therapeutic approaches for managing T1D. The therapeutic regimen could range from insulin treatment (dose of insulin per kg and ratio of short to long-acting insulin) to diet interventions, and, for T2D with respect to diet interventions, mono vs combination therapy in the management of T2D with primary focus on AA subjects. Individuals with PRSs in between T1D and T2D can be tested for MODY and other atypical forms of diabetes.

Tackling the Need for Diversity:

The overwhelming majority of PRS studies to date focus on individuals of European ancestry. Similarly, commercial genotyping arrays are known to have a major ascertainment bias towards European ancestry. For PRS, this problem is exacerbated by linkage disequilibrium-based pruning/corrections, because they rely on reference haplotype panels that lack population diversity – and PRS derived from European-ancestry GWAS data are biased in unpredictable directions, depending on the demographics.[115] Importantly, individuals of African ancestry have significantly higher risk allele frequencies, which is higher for ancestral versus derived (i.e., more recent) risk alleles.[116] CAG is arguably among the most experienced in the world in addressing these inequities having genotyped >50,000 AAs to date and customized a range of informatics tools to address disparities.

EMR Integration:

We plan to leverage the eMERGE 4 platform as a model of prophylactic genomic medicine at CHOP, including the markedly increased integration of multidisciplinary CDS. Barriers to implementing CDS extend beyond technology to other issues resulting from organizational, environmental, process, and patient factors.[117,118] Adding to the challenge is the need to educate physicians and other organizational decision-makers about the complex sociotechnical factors that arise from genomic test results.[119]

CDS design and development:

CDS requirements from cognitive task analysis will guide development of interactive scenario based CDS mock-ups for lead phenotypes. In addition, health screening and intervention workflows are often markedly different in pediatric settings, further complicating CDS design.[120] A well-designed CDS development and validation plan is critical to addressing these barriers and

106. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet. 2011;7(9):e1002293. Epub 2011/10/08. doi: 10.1371/journal.pgen.1002293. PubMed PMID: 21980299; PubMed Central PMCID: PMC3183083.
107. Grant SF, Li M, Bradfield JP, Kim CE, Annaiah K, Santa E, et al. Association analysis of the FTO gene with obesity in children of Caucasian and African ancestry reveals a common tagging SNP. PLoS One. 2008;3(3):e1746. Epub 2008/03/13. doi: 10.1371/journal.pone.0001746. PubMed PMID: 18335027; PubMed Central PMCID: 2262153.
108. Grant SF, Qu HQ, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. Follow-up analysis of genome-wide association data identifies novel loci for type 1 diabetes. Diabetes. 2009;58(1):290-5. Epub 2008/10/09. doi: 10.2337/db08-1022. PubMed PMID: 18840781; PubMed Central PMCID: PMC2606889.
109. Grant SFA, Glessner JT, Bradfield JP, Wang K, Zhang HT, Mentch FD, et al. Genome Wide Copy Number Variation Study Reveals Loci Exclusively Impacted in Childhood Obesity Cases. Diabetes. 2010;59:A55-A. PubMed PMID: WOS:000278844600203.
110. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature. 2007;448(7153):591-4. Epub 2007/07/17. doi: 10.1038/nature06010. PubMed PMID: 17632545.
111. Hakonarson H, Qu HQ, Bradfield JP, Marchand L, Kim CE, Glessner JT, et al. A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. Diabetes. 2008;57(4):1143-6. Epub 2008/01/17. doi: 10.2337/db07-1305. PubMed PMID: 18198356.
112. Mishra R, Bradfield JP, Cousminer DL, Chesi A, Hodge KM, Hakonarson H, et al. Genetic Discrimination between LADA and Type 1 Diabetes within the MHC. Hum Hered. 2017;83(1):47-8. PubMed PMID: WOS:000448600500090.
113. Wang K, Baldassano R, Zhang H, Qu HQ, Imielinski M, Kugathasan S, et al. Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. Hum Mol Genet. 2010;19(10):2059-67. Epub 2010/02/24. doi: 10.1093/hmg/ddq078. PubMed PMID: 20176734; PubMed Central PMCID: 2860894.
114. Hasson RE, Adam TC, Pearson J, Davis JN, Spruijt-Metz D, Goran MI. Sociocultural and socioeconomic influences on type 2 diabetes risk in overweight/obese African-American and Latino-American children and adolescents. Journal of obesity. 2013;2013:512914. Epub 2013/06/14. doi: 10.1155/2013/512914. PubMed PMID: 23762538; PubMed Central PMCID: PMC3666294.
115. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet. 2017;100(4):635-49. Epub 2017/04/04. doi: 10.1016/j.ajhg.2017.03.004. PubMed PMID: 28366442; PubMed Central PMCID: PMC5384097.
116. Kim MS, Patel KP, Teng AK, Berens AJ, Lachance J. Genetic disease risks can be misestimated across global populations. Genome Biol. 2018;19(1):179. Epub 2018/11/15. doi: 10.1186/s13059-018-1561-7. PubMed PMID: 30424772; PubMed Central PMCID: PMC6234640.
117. Carayon P, Schoofs Hundt A, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Quality & safety in health care. 2006;15 Suppl 1:i50-8. Epub 2006/12/05. doi: 10.1136/qshc.2005.015842. PubMed PMID: 17142610; PubMed Central PMCID: PMC2464868.
118. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics. 2013;56(11):1669-86. Epub 2013/10/04. doi: 10.1080/00140139.2013.838643. PubMed PMID: 24088063; PubMed Central PMCID: PMC3835697.
119. Pennington JW, Karavite DJ, Krause EM, Miller J, Bernhardt BA, Grundmeier RW. Genomic decision support needs in pediatric primary care. J Am Med Inform Assoc. 2017;24(4):851-6. Epub 2017/03/25. doi: 10.1093/jamia/ocw184. PubMed PMID: 28339689.
120. Boland L, Graham ID, Legare F, Lewis K, Jull J, Shephard A, et al. Barriers and facilitators of pediatric shared decision-making: a systematic review. Implementation science : IS. 2019;14(1):7. Epub 2019/01/20. doi: 10.1186/s13012-018-0851-5. PubMed PMID: 30658670; PubMed Central PMCID: PMC6339273.

creating a CDS model that can be implemented across phenotypes and eMERGE CSs.

Based on previous work,[119] we anticipate that CDS targeted at the primary care pediatrician may include actionable steps in the EMR such as modification of the problem list, reviewing provider education, provision of patient-family education, referrals to specialists, and/or ordering a preventive medication.

Conclusions

As the eMERGE enterprise has matured, relevant projects have become increasingly expansive and ambitious in scope and output, which is in line with the goal of forging new ground in genomic medicine research. In eMERGE 2, PGx reports were issued on more than 9,000 patient participants, where clinical decision support initiatives facilitated individualized dosing. The eMERGEseq panel developed in eMERGE 3 was applied to more than 25,000 individuals and identified 202 individuals with positive diagnostic findings and 1,294 with additional/secondary findings of medical significance deemed to be returnable. In a similar vein, eMERGE 4 will recruit 25,000 individuals, all of whom will receive a genomics-informed risk report. The latest phase in particular constitutes a model of genomic medicine that is generalizable and scalable to a large proportion of the US population, and other countries with genomics capacity and access to electronic medical records.

Any research that directly impacts healthcare systems will always demand careful attention to ethics, cost, and outcomes; as such, eMERGE and related programs require similarly exacting standards for revisiting and revising our understanding of genomic variation. The eMERGE Network is focused directly on this endpoint and is delivering a promise made through the human genome project – the large-scale provision of individualized medicine. The product of this collective effort is development of a comprehensive suite of resources and know-how that support the use of genomic information in clinical care. This is particularly important in light of the eMERGE's emphasis on diversity populations, which is helping unfold the genetic architecture in under-represented individuals, thereby validating important genetic/genomic findings form an ancestry standpoint. JoPM

John J Connolly, Ph.D.

John J Connolly, Ph.D., is a researcher at the Center for Applied Genomics (CAG) at The Children's Hospital of Philadelphia (CHOP). Since joining CAG in 2010, he has led a range of neuropsychiatric studies, with a focus on autism and ADHD. He also maintained a lead role in the Philadelphia Neurodevelopmental Cohort project, which integrates deeply-characterized data from 9,500 children and young adults that includes electronic medical records (EMRs), neuroimaging, genotype, sequencing, and methylation profiles. He is currently a project leader for CHOP as part of our collaboration with the electronic Medical Records and Genomics (eMERGE) Consortium, a major NHGRI 9-Center initiative to integrate EMRs and genomics.

Dr. Connolly trained as neuropsychologist at Trinity College Dublin, where he studied stroke rehabilitation and neural correlates of ADHD. He subsequently joined Cold Spring Harbor Laboratory to lead development of Genes to Cognition Online, an online resource that examines how the brain, biochemicals, and genes interact to produce neuropsychiatric. He also created *3D Brain*, an award-winning app that has been downloaded more than 4 million times.

Patrick Sleiman

Dr Sleiman is associate professor of pediatrics at the University of Pennsylvania Medical School and the associate director and lead statistical geneticist at the Center for Applied Genomics of the Children's Hospital of Philadelphia. He received his PhD in genetics from the University of London before completing postdoctoral training at the Institute of Neurology in London where he led a study that resulted in the identification of a novel early onset Parkinson's disease gene, PINK1. His research interests lay in uncovering the genetic basis of human disease with a view towards providing more patient focused treatment through both mining of the CAG internal EMR-based patient collection and work with external collaborators across academia and the pharmaceutical industry. His work has resulted in the identification of numerous novel genes and disease loci across multiple phenotypes including eosinophilic esophagitis, schizophrenia, frontotemporal dementia, progressive supranuclear palsy, Parkinson's disease and asthma that have been reported in over 150 peer reviewed publications, garnering over 23,000 citations.

Joseph T. Glessner, Ph.D.

Technical Director, Genetics Core Facility, Center for Applied Genomics

Research Assistant Professor of Pediatrics, Perelman School of Medicine, University of Pennsylvania

Dr. Glessner's current research focuses on childhood neuropsychiatric and neurodevelopmental disorders along with the genetic architecture associated with them, including single nucleotide polymorphisms, single nucleotide variations, and copy number variations ascertained by genomic technologies.

Hakon Hakonarson, M.D., Ph.D.

Hakon Hakonarson, M.D., Ph.D., is Director of the Center for Applied Genomics, Endowed Chair in Genomics Research and Professor of Pediatrics at The University of Pennsylvania, Perelman School of Medicine. Dr. Hakonarson leads a major commitment from CHOP to genomically characterize approximately 100,000 children, an initiative that has gained nationwide attention in the Wall Street Journal, New York Times, Time Magazine, Nature and Science. Dr. Hakonarson is a Principal Investigator within the Kids First program and the TopMed genomics program funded by the NIH. Dr. Hakonarson published the first pediatric GWAS in T1D (Hakonarson, Nature, 2007) and he has over >25 T1D discovery and translational publications since then.

Dr. Hakonarson has previously held several senior posts within the biopharmaceutical industry, directing a number of genomics and pharmacogenomics projects as vice president of Clinical Sciences and Development at deCODE genetics, Inc. Dr. Hakonarson has been the principal investigator (PI) on multiple National Institute of Health-sponsored grants, and he was a principal investigator on the Neurodevelopmental Genomics: Trajectories of Complex Phenotypes, the largest research project ever supported by the National Institute of Mental Health. Dr. Hakonarson recently completed a clinical biomarker study in ADHD demonstrating strong efficacy and safety of a neuromodulator compound (NFC-1) in children with specific mutations in the glutamate metabotropic (mGluR) receptor family of genes with ADHD and Autism (www.ClinicalTrial.gov; Elia et al, *Nat Comm*, 2018).

Dr. Hakonarson has published over 750 scientific papers, including numerous high-impact papers on genomic discoveries and their translations in some of the most prestigious scientific medical journals, including *Nature, Nature Medicine, Nature Genetics, Cell and The New England Journal of Medicine*. Time Magazine listed Dr. Hakonarson's autism gene discovery project, reported in Nature in 2009, among the top 10 medical breakthroughs of that year. With over 20 years of experience in pioneering genomics research and genome-wide mapping and association studies, Dr. Hakonarson has intimate knowledge of the complexities of large-scale genomics and drug development projects, and he has put together the necessary infrastructure and workflow processes to unravel these complexities for optimized deliverables of precision medicine programs.