PRECISION MEDICINE: FROM DIPLOTYPES TO DISPARITIES TOWARDS IMPROVED HEALTH AND THERAPIES

DANA C. CRAWFORD

Population and Quantitative Health Sciences, Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, 44106 USA
Email: dana.crawford@case.edu

ALEXANDER A. MORGAN

Khosla Ventures, Menlo Park, CA, 94025 USA
Email: alex@khoslaventures.com

JOSHUA C. DENNY

Vanderbilt University Medical Center, Nashville, TN 37203 USA
Email: josh.denny@vanderbilt.edu

BRUCE J. ARONOW

Center for Computational Medicine, Cincinnati Children's Hospital Medical Center and the University of Cincinnati, Cincinnati, OH 45229 USA
Email: bruce.Aronow@cchmc.org

STEVEN E. BRENNER

University of California, Berkeley, CA 94720-3012 USA
Email: brenner@compbio.berkeley.edu

Precision medicine research efforts both in basic science discovery and clinical implementation are well underway and promise to provide individualized preventions and treatments, improving overall health care delivery. To achieve these goals, advances in data capture and analysis are needed spanning different types of 'omic and clinical data. The efforts to enhance precise treatments for all may accentuate healthcare disparities unless specific challenges are identified and addressed. This session of the 2018 Pacific Symposium on Biocomputing presents the latest developments in this transdisciplinary research space of genomics, medicine, and population health.

1. Introduction

Precision medicine is often described as providing a patient the optimal tailored treatment the first time as opposed to standard treatment or trial and error. Precision medicine has arguably been practiced since the emergence of modern medicine[1]. The definition of precision medicine has recently evolved from the 20th century's addition of genetic variants that impact drug response[2-4] to the 21st century's recognition that lifestyle, social, and environmental factors interact with the patient's genome[5], impacting a range of health consequences including conferring risk to disease and differential response to treatmen[6].

Although the concept of precision medicine is not new, the implementation of precision medicine

is relatively nascent. Recent key advances in genomics[7] and the emergence of electronic health records (EHRs) in part through the HITECH Act[8, 9] make it feasible to put precision medicine to practice[10]. Despite these advances, the adoption of genomic data in routine clinical practice has been slow likely due to a variety of reasons including costs or inconsistent reimbursement policies (such as changing and limited coverage by Medicare) and access to genomic testing[11].

Another major driver behind the lack of implementation is the lack of data. The lack of data is distributed across active research areas that fuel precision medicine including basic discovery and functional or biological characterization[12]. Inherent in these broad areas of research are needed expansions of informatics approaches to extract health-informative data from various big data sources coupled with advances in novel statistical and computational methodology to integrate and interpret disparate data types for predictive modeling or further discovery and functional characterization[13, 14]. Beyond the traditional boundaries of basic research are data needs related to precision medicine implementation ranging in topics from cost-effectiveness and clinical utility to clinical decision support[15, 16], among other topics[17].

Precision medicine research and implementation programs will undoubtedly evolve rapidly with the recent infusion of support from the previous administration[18]. The initial structure and recruitment sites for the Precision Medicine Initiative Cohort Program, an ambitious effort to ascertain one million participants in the United States for precision medicine research, have been established as of 2016, and recruitment and collection of these data are anticipated this year. Of note is the Program's emphasis on racial/ethnic, geographic, and economic diversity, variables that continue to be underrepresented in the basic discovery studies[19-21] despite their known influence on human health[22, 23]. The absence of these data may lead to misdiagnoses and missed opportunities for many patients not represented in discovery studies[24]. The new Cohort endeavor, now known as "All of Us," is expected to both generate new data but also to inspire new regulations and guidelines based on safety and bioethics[25-27]. Given innovative biocomputing is the engine of precision medicine's implementation vehicle, PSB is the optimal forum for the presentation, discussion, and debate of these diverse topics that eventually fuel true individualized health for all.

The investigators and research featured in this session each represent a facet of precision medicine research highlighting needs and gaps that must be addressed to achieve the goals of translational research. Topics covered in this session include the problem of finding sufficient numbers of patients or participants with similar characteristics required to achieve adequate power to identify important biomarkers that distinguish subtypes including genetic, metabolomic or other phenotypic features that have a molecular or mechanistic relationship. Without careful attention, this can lead to health disparities, as less information may be available from vulnerable groups, and thus leading to less effective diagnosis and treatment. Such challenges need to be identified and actively addressed.

This session also addresses development of incorporating social network information to recruit patients in context where they are and better understand the variables that are operative for those individuals in disadvantaged coverage and difficult complex environments, all of which opens up

for new possibilities in research and care.

2. Podium presentations

A group of four manuscripts describe novel and emerging approaches to tackle 'omic data generated by metabolomics and transcriptomics. **A Orlenko** and colleagues[28] present a case study of Tree-based Pipeline Optimization Tool (TPOT)[29], an Automated Machine Learning (AutoML) tool, applied to the clinical metabolic profiling of patients exposed to metformin. In their assessment of TPOT in 546 samples and 42 metabolites, the investigators identified a putatively novel association between increased homocysteine and long-term exposure to metformin. The investigators also developed several considerations for future studies including suggestions for adjustments for confounding features and recommendations for the ideal study design or dataset characteristics that minimize bias and improve AutoML performance.

J Westra and colleagues[30] take a different approach in the analysis of metabolomic data with the observation that these data are better represented by Gaussian mixture distributions rather than the linear models that assume normality commonly used to test for SNP-trait associations. The investigators present an adaptation of Kim et al where tests of association between copy number variants and traits were performed with a likelihood ratio test[31]. The adapted test presented here is a likelihood ratio test that can be constrained based on *a priori* biological data describing the genotype-phenotype relationship, if available. This adapted tested was applied to simulated data to evaluate performance of the test versus liner models as well as to natural data from 20,315 SNPs on chromosome 11 for 5,936 Framingham Heart Study participants with gas-chromatography-measured red blood cell fatty acid levels[32]. A total of 28 SNPs from five different regions of chromosome 11 were associated with the metabolic trait, including 19 SNPs containing the FADS gene complex known to enzymatically desaturate arachidonic acid to dihoo-gamma-linoleneic acid. While further work is needed to extend the model, the present study suggests that powerful analysis of metabolomics data may require different models compared with the traditional linear models so popular in GWAS.

Both J Berghout and S Rachid Zaim offer methods applicable to single-subject transcriptomics. **J Berghout** and colleagues[33] describe a mixture mole for transcript fold-change clustering from isogenically paired samples known as the "N-of-1 pathways MixEnrich."[34] The Gene Ontology Biological Processes (GO-BP)[35, 36] is applied to both reduce dimensionality as well as to identify functional attributes. The method was validated using a microarray dataset of inbred mouse strains exposed to different diets. The MixEnrich results for the paired mouse liver transcriptomes are compared with results from two other methods, Linear Models for Microarray (*limma*)[37] and Gene Set Enrichment Analysis[38], both of which require a minimum of three pairs as opposed to a single sample pair required by MixEnrich. Results suggest that MixEnrich reproduces GO-BP signals in similar priority order compared with the other approaches, thus offering an efficient alternative to cost prohibitive cohort-derived gold standards used for validation.

S Rachid Zaim and colleagues[39] use simulations to address limitations in using transcriptomics as a clinical biomarker. Here, the investigators assume that a transcriptional signal or association can

be detected at the pathway level regardless of patient-level heterogeneity in gene expression. The simulations are performed to assess 54 different scenarios using single-subject and cohort-based techniques describing variability at various levels including pathway gene set size, fraction of expressed gene responsive within the gene set, fraction of up and down-regulated gene expression response, and sample size. Results of these simulations suggest that single-subject pathway detection methods that include patient-level variability can detect transcriptional dysregulation at the pathway level, scenarios likely relevant to heterogeneous clinical populations expected in precision medicine.

Of all the 'omics, genomics, represented by genotyping and sequencing, is the most mature and readily available in the clinic. Whole genome sequencing has been particularly groundbreaking in diagnosing previously undiagnosed rare diseases[40-42], but these data remain an analysis challenge for more common, complex diseases. **A Gupta** and colleagues[43] note, as have others[44], that the genetic architecture of some complex diseases such as autism spectrum disorder (ASD) remains unknown despite evidence of a strong genetic component. While it is now recognized that complex diseases are often a result of many variants across multiple genes with independent and interacting effects, and sequencing methods are available to capture genome-wide variability, few statistical methods have emerged to detect these complex genetic architectures. The investigators posit that Coalition Game Theory (CGT) can be applied to genomic data to identify individual genes ("players") who improve the performance of the coalition or, in this case, the relationship to ASD. The investigators apply CGT to 2,710 whole-genome sequences from ASD multiplex families and identify eight genes with significantly elevated "player scores." All eight genes are in biological pathways known to be affected by ASD and directly interact with genes previously associated with ASD. Although further follow-up is needed, these results suggest CGT is a promising method for large-scale genome data generated for complex diseases.

Two manuscripts centered on the use of automated clinical systems in the delivery or practice of precision medicine. In the first, **C-L Chi** and colleagues[45] apply a treatment simulation and optimization approach to develop decision support for warfarin dosing. Warfarin is a commonly prescribed anti-coagulant, and variability in initial dosing has a strong and known genetic component[46, 47]. Genetically-guiding warfarin dosing was an early poster-child for pharmacogenomics in precision medicine, but lackluster clinical trials[48-50] save for one[51] among other issues have dampened enthusiasm. Logistically, the development of algorithm-based dosing delivered via EHR-automated decision support has been a challenge for this and other drugs impacted by genetics, particularly for diverse populations[52]. The investigators present results of simulations that employed the property of minimal entropy to minimize overall risks for the largest patient groups. The investigators further discuss these results through the lens of ease of implementation, a factor highly relevant for this[53] and other potential precision medicine clinical applications.

S Poole and colleagues[54] address alarm fatigue, an unfortunate consequence of easy and constant automated vital sign monitoring. The investigators aim to improve default vital sign alarm thresholds to decrease the number of unnecessary alarms. The investigators develop personalized vital sign thresholds based on a large heart rate database used to identify the 1st and 99th

percentiles of a patient's heart rate on the patient's first day of monitoring. Results suggest that these new thresholds would decrease low and heart rate alarms while preserving sensitivity and boosting specificity. Overall, the investigators suggest these thresholds will reduce alarm fatigue thereby improving both patient care and hospital costs, major goals underlying the original HITECH Act[55].

3. Posters with published papers

This year's poster session with papers published in the proceedings will feature two research groups. In the first, **A Fish** and colleagues[56] present evidence that genetic associations are modified by local ancestry transitions inferred from genome-wide association study (GWAS) fine-mapping Metabochip data available on ~10,000 African Americans with de-identified EHRs[57]. African Americans are considered admixed with an average 75-93% their genomes originating from West African ancestral populations and the remaining from European ancestral populations[58-61]. Local ancestry, as opposed to inferred genome-wide global ancestry, is the inference of ancestral state at the locus-level. Local ancestry estimates have been instrumental in admixture mapping efforts[62] as well as the estimation of recombination rates and the identification of recombination hotspots[63]. The investigators leveraged this admixed population with clinical data to identify SNP x transition interactions where the transition represent a switch in ancestral backgrounds between nearby variants. Five Bonferroni-corrected significant interactions were identified, and subsequent statistical follow-up suggested that a European to African transition modifies the association rs16890640 between mean corpuscular hemoglobin and mean corpuscular volume. Bioinformatic data coupled with model organism data suggest that alterations in the region chromatin conformation are the biological basis for the modifying effect of the ancestral transition. More broadly, this study offers an example of epistasis where the interaction, and the not variant itself, is associated with the phenotype[64]. Furthermore, this study highlights yet another example of the importance of diverse populations in the search for all genetic variants and their modifiers important in human health and health disparities.

In the second, **B Li** and colleagues[65] test two functions of PrediXcan: 1) its ability to predict gene expression and 2) its ability to prioritize GWAS results. PrediXcan is a gene-based association method rooted in the observation that phenotypic variability can be explained by regulatory variants that modulate the expression levels of genes[66]. PrediXcan uses reference transcriptome data to infer gene expression in GWAS data by estimating the genetically determined component of gene expression. This gene-based approach reduces multiple testing and proffers biological insights or mechanisms compared with standard single SNP tests of association common in GWAS. This study evaluates PrediXcan using genotypic and transcriptomic datasets available from the 1000 Genomes Project (Yoruba; YRI) and GWAS data from the AIDS Clinical Trials Group (ACTG) protocol A5202[67, 68]. To characterize the accuracy in predicting gene expression levels, the investigators compared the PrediXcan-inferred YRI data based on whole blood models and transcriptomic data from the Genotype-Tissue Expression (GTEx) Project[69] and Depression, Genes and Networks (DGN)[70] to the actual YRI expression data and found that the slopes of correlation between predicted and actual were negative for almost one-half of the genes tested. Despite these differences, PrediXcan identified 19 genes in the A5202 cohort dataset associated

with triglyceride change from baseline to 24 or 48 weeks that were not previously identified using phenome-wide association testing, attesting to PrediXcan's potential ability to prioritize GWAS findings. Of note is the poor transcriptome prediction in YRI despite the fact that the GTEx cohort includes African Americans. These data suggest that testing the limits of PrediXcan in gene-trait associations will require more, larger, and diverse populations with both GWAS and transcriptome-level data.

4. Acknowledgements

We would like to thank the members of the program committee for reviewing all submissions and providing expert critiques used in evaluating manuscripts for inclusion into the session and the PSB proceedings. We would also like to thank the PSB 2018 chairs and Tiffany Murray of Stanford University for their efforts in organizing the meeting. This work is supported in part by NIH grants including U41 HG007346 to S.E.B and by a collaborative research agreement with Tata Consultancy Services.

5. References

1. Murray JF. Personalized Medicine: Been There, Done That, Always Needs Work! *American Journal of Respiratory and Critical Care Medicine*. 2012;185:1251-1252.
2. Gurwitz D and Motulsky AG. Drug reactions, enzymes, and biochemical genetics: 50 years later. *Pharmacogenomics*. 2007;8:1479.
3. Motulsky AG. Drug reactions, enzymes, and biochemical reactions. *JAMA*. 1957;165:835-837.
4. Motulsky AG and King MC. The great adventure of an American human geneticist. *Annu Rev Genomics Hum Genet*. 2016;17:1-15.
5. Hall MA, Moore JH and Ritchie MD. Embracing Complex Associations in Common Traits: Critical Considerations for Precision Medicine. *Trends in Genetics*. 2016;32:470-484.
6. Carlsten C, Brauer M, Brinkman F, Brook J, Daley D, McNagny K, Pui M, Royce D, Takaro T and Denburg J. Genes, the environment and personalized medicine. *We need to harness both environmental and genetic data to maximize personal and population health*. 2014;15:736-739.
7. van Nimwegen KJM, van Soest RA, Veltman JA, Nelen MR, van der Wilt GJ, Vissers LELM and Grutters JPC. Is the $1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. *Clinical Chemistry*. 2016;62:1458-1464.
8. Feero W, Guttmacher AE and Collins FS. The genome gets personal--almost. *JAMA*. 2008;299:1351-1352.
9. Blumenthal D. Launching HITECH. *New England Journal of Medicine*. 2010;362:382-385.
10. Duffy DJ. Problems, challenges and promises: perspectives on precision medicine. *Briefings in Bioinformatics*. 2016;17:494-504.
11. Lynch JA, Berse B, Dotson WD, Khoury MJ, Coomer N and Kautter J. Utilization of genetic tests: analysis of gene-specific billing in Medicare claims data. *Genet Med*. 2017.
12. Green ED and Guyer MS. Charting a course for genomic medicine from base pairs to bedside. *Nature*. 2011;470:204-213.

13. Iyengar R, Altman RB, Troyanskya O and FitzGerald GA. Personalization in practice. *Science*. 2015;350:282-283.
14. Ritchie MD, Holzinger ER, Li R, Pendergrass SA and Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet*. 2015;16:85-97.
15. Schully SD and Khoury MJ. What is translational genomics? An expanded research agenda for improving individual and population health. *Applied & Translational Genomics*. 2014;3:82-83.
16. Herr T, Bielinski S, Bottinger E, Brautbar A, Brilliant M, Chute C, Cobb B, Denny J, Hakonarson H, Hartzler A, Hripcsak G, Kannry J, Kohane I, Kullo I, Lin S, Manzi S, Marsolo K, Overby C, Pathak J, Peissig P, Pulley J, Ralston J, Rasmussen L, Roden D, Tromp G, Uphoff T, Weng C, Wolf W, Williams M and Starren J. Practical considerations in genomic decision support: The eMERGE experience. *Journal of Pathology Informatics*. 2015;6:50-50.
17. Manolio TA. Implementing genomics and pharmacogenomics in the clinic: The National Human Genome Research Institute's genomic medicine portfolio. *Atherosclerosis*. 2016;253:225-236.
18. Collins FS and Varmus H. A New Initiative on Precision Medicine. *New England Journal of Medicine*. 2015;372:793-795.
19. Bustamante CD, De La Vega FM and Burchard EG. Genomics for the world. *Nature*. 2011;475:163-165.
20. Popejoy AB and Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538:161-164.
21. Adler NE and Stead WW. Patients in Context — EHR Capture of Social and Behavioral Determinants of Health. *New England Journal of Medicine*. 2015;372:698-701.
22. Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, Mountain JL, Perez-Stable EJ, Sheppard D and Risch N. The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice. *New England Journal of Medicine*. 2003;348:1170-1175.
23. Braveman P, Egerter S and Williams DR. The social determinants of health: coming of age. *Annu Rev Public Health*. 2011;32:381-398.
24. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, Margulies DM, Loscalzo J and Kohane IS. Genetic Misdiagnoses and the Potential for Health Disparities. *New England Journal of Medicine*. 2016;375:655-665.
25. Kohane IS. Ten things we have to do to achieve precision medicine. *Science*. 2015;349:37-38.
26. Sankar PL and Parker LS. The Precision Medicine Initiative/'s All of Us Research Program: an agenda for research on its ethical, legal, and social issues. *Genet Med*. 2016.
27. Thompson B and Boiani J. The Legal Environment for Precision Medicine. *Clinical Pharmacology & Therapeutics*. 2016;99:167-169.
28. Orlenko A, Moore JH, Orzechowski P, Olson RS, Cairns J, Caraballo PJ, Weinshilboum RM, Wang L and Breitenstein MK. Considerations for automated machine learning in clinical metabolic profiling: altered homocysteine plasma concentration associated wtih metformin exposure. *Pac Symp Biocomput*. 2017;23.
29. Olson RS, Bartley N, Urbanowicz RJ and Moore JH. Evaluation of a tree-based pipeline optimization tool for automating data science. *Proc Genet Evol Comput Conf*. 2016;GECCO '16:485-492.

30. Westra J, Hartman N, Lake B and Tintle N. Analyzing metabolomics data for association with genotypes using two-component Gaussian mixture distributions. *Pac Symp Biocomput*. 2017;23.
31. Kim W, Gordon D, Sebat J, Ye KQ and Finch SJ. Computing Power and Sample Size for Case-Control Association Studies with Copy Number Polymorphism: Application of Mixture-Based Likelihood Ratio Test. *PLOS ONE*. 2008;3:e3475.
32. Harris WS, Pottala JV, Lacey SM, Vasan RS, Larson MG and Robins SJ. Clinical correlates and heritability of erythrocyte eicosapentaenoic and docosahexaenoic acid content in the Framingham Heart Study. *Atherosclerosis*. 225:425-431.
33. Berghout J, Li Q, Pouladi N, Li J and Lussier YA. Single subject transcriptome analysis reproduces signed gene set functional activation singals from cohort analysis of murine response to high fat diet. *Pac Symp Biocomput*. 2017;23.
34. Li Q, Schissler AG, Gardeux V, Achour I, Kenost C, Berghout J, Li H, Zhang HH and Lussier YA. N-of-1-pathways MixEnrich: advancing precision medicine via single-subject analysis in discovering dynamic changes of transcriptomes. *BMC Medical Genomics*. 2017;10:27.
35. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H and Cherry JM. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25.
36. Consortium TGO. Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res*. 2017;45:D331-D338.
37. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and mciroarray studies. *Nucleic Acids Res*. 2015;43:e47.
38. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005;102:15545-15550.
39. Rachid Zaim S, Li Q, Schissler AG and Lussier YA. Emergence of pathway-level composite biomarkers from converging gene set signals of heterogeneous Genome-by-Environment transcriptomic responses. *Pac Symp Biocomput*. 2017;23.
40. Chong Jessica X, Buckingham Kati J, Jhangiani Shalini N, Boehm C, Sobreira N, Smith Joshua D, Harrell Tanya M, McMillin Margaret J, Wiszniewski W, Gambin T, Coban Akdemir Zeynep H, Doheny K, Scott Alan F, Avramopoulos D, Chakravarti A, Hoover-Fong J, Mathews D, Witmer PD, Ling H, Hetrick K, Watkins L, Patterson Karynne E, Reinier F, Blue E, Muzny D, Kircher M, Bilguvar K, López-Giráldez F, Sutton VR, Tabor Holly K, Leal Suzanne M, Gunel M, Mane S, Gibbs Richard A, Boerwinkle E, Hamosh A, Shendure J, Lupski James R, Lifton Richard P, Valle D, Nickerson Deborah A and Bamshad Michael J. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *The American Journal of Human Genetics*. 2015;97:199-215.
41. Biesecker LG and Green RC. Diagnostic Clinical Genome and Exome Sequencing. *New England Journal of Medicine*. 2014;370:2418-2425.
42. Willig LK, Petrikin JE, Smith LD, Saunders CJ, Thiffault I, Miller NA, Soden SE, Cakici JA, Herd SM, Twist G, Noll A, Creed M, Alba PM, Carpenter SL, Clements MA, Fischer RT, Hays JA, Kilbride H, McDonough RJ, Rosterman JL, Tsai SL, Zellmer L, Farrow EG and Kingsmore SF. Whole-genome sequencing for identification of Mendelian disorders in critically

ill infants: a retrospective analysis of diagnostic and clinical findings. *The Lancet Respiratory Medicine*. 2015;3:377-387.
43. Gupta A, Sun MW, Paskov KM, Stockham NT and Wall DP. Coalitional game theory as a promising approach to identify candidate autism genes. *Pac Symp Biocomput*. 2017;23.
44. Merikangas KR and Risch N. Genomic Priorities and Public Health. *Science*. 2003;302:599-601.
45. Chi C-L, He L, Ravvaz K, Weissert J and Tonellato PJ. Optimized decision support rules of precision warfarin treatment. *Pac Symp Biocomput*. 2017;23.
46. Rieder MJ, Reiner AP, Gage BF, Nickerson DA, Eby CS, McLeod HL, Blough DK, Thummel KE, Veenstra DL and Rettie AE. Effect of VKORC1 Haplotypes on Transcriptional Regulation and Warfarin Dose. *New England Journal of Medicine*. 2005;352:2285-2293.
47. Perera MA, Cavallari LH, Limdi NA, Gamazon ER, Konkashbaev A, Daneshjou R, Pluzhnikov A, Crawford DC, Wang J, Liu N, Tatonetti NJ, Bourgeois S, Takahashi H, Bradford Y, Burkley BM, Desnick RJ, Halperin JL, Khlifa SI, Langaee TY, Lubitz SA, Nutescu EA, Oetjens M, Shahin MH, Shitalben RP, Tector M, Rieder MJ, Scott SA, Wu AHB, Burmester JK, Deloukis P, Wagner MJ, Mushiroda T, Kubo M, Roden DM, Cox NJ, Altman RB, Klein TE, Nakamura Y and Johnson JA. Genetic variants associated wtih warfarin dose in African-American individuals: a genome-wide association study. *Lancet*. 2013;382:790-796.
48. Pirmohamed M, Burnside G, Eriksson N, Jorgensen AL, Toh CH, Nicholson T, Kesteven P, Christersson C, Wahlstr+¦m B, Stafberg C, Zhang JE, Leathart JB, Kohnke H, Maitland-van der Zee A, Williamson PR, Daly AK, Avery P, Kamali F and Wadelius M. A Randomized Trial of Genotype-Guided Dosing of Warfarin. *New England Journal of Medicine*. 2013;369:2294-2303.
49. Kimmel SE, French B, Kasner SE, Johnson JA, Anderson JL, Gage BF, Rosenberg YD, Eby CS, Madigan RA, McBane RB, Abdel-Rahman SZ, Stevens SM, Yale S, Mohler ER, Fang MC, Shah V, Horenstein RB, Limdi NA, Muldowney JAS, Gujral J, Delafontaine P, Desnick RJ, Ortel TL, Billett HH, Pendleton RC, Geller NL, Halperin JL, Goldhaber SZ, Caldwell MD, Califf RM and Ellenberg JH. A Pharmacogenetic versus a Clinical Algorithm for Warfarin Dosing. *New England Journal of Medicine*. 2013;369:2283-2293.
50. Verhoef TI, Ragia G, de Boer A, Barallon R, Kolovou G, Kolovou V, Konstantinides S, Le Cessie S, Maltezos E, van der Meer FJM, Redekop WK, Remkes M, Rosendaal FR, van Schie RMF, Tavridou A, Tziakas D, Wadelius M, Manolopoulos VG and Maitland-van der Zee AH. A Randomized Trial of Genotype-Guided Dosing of Acenocoumarol and Phenprocoumon. *New England Journal of Medicine*. 2013;369:2304-2312.
51. Gage BF, Bass AR, Lin H, Woller SC, Stevens SM, Al-Hammadi N, Li J, Rodriguez T, Miller JP, McMillin GA, Pendleton RC, Jaffer AK, King CR, Whipple BD, Porche-Sorbet R, Napoli L, Merritt K, Thompson AM, Hyun G, Anderson JL, Hollomon W, Barrack RL, Nunley RM, Moskowitz G, Davila-Roman V and Eby CS. Effect of genotype-guided warfarin dosing on clinical events and anticoagulation control among patients undergoing hip or knee arthoplasty. The GIFT Randomized Clinical Trial. *JAMA*. 2017;318:1115-1124.
52. Johnson JA, Caudle KE, Gong L, Whirl-Carrillo M, Stein CM, Scott SA, Lee MT, Gage BF, Kimmel SE, Perera MA, Anderson JL, Pirmohamed M, Klein TE, Limdi NA, Cavallari LH and Wadelius M. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update. *Clinical Pharmacology & Therapeutics*. 2017;102:397-404.

53. Emery JD. Pharmacogenomic testing and warfarin. What evidence has the GIFT Trial provided? *JAMA*. 2017;318:1110-1112.
54. Poole S and Shah N. Addressing vital sign alarm fatigue using personalized alarm thresholds. *Pac Symp Biocomput*. 2017;23.
55. Adler-Milstein J and Jha AK. HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. *Health Affairs*. 2017;36:1416-1422.
56. Fish AE, Crawford DC, Captra JA and Bush WS. Local ancestry transitions modify SNP-trait associations. *Pac Symp Biocomput*. 2017;23.
57. Crawford DC, Goodloe R, Farber-Eger E, Boston J, Pendergrass SA, Haines JL, Ritchie MD and Bush WS. Leveraging epidemiologic and clinical collections for genomic studies of complex traits. *Human Heredity*. 2015;79:137-146.
58. Bryc K, Durand E-á, Macpherson J-á, Reich D and Mountain J-á. The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States. *The American Journal of Human Genetics*. 2015;96:37-53.
59. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE and Shriver MD. Estimating African American Admixture Proportions by Use of Population-Specific Alleles. *The American Journal of Human Genetics*. 1998;63:1839-1851.
60. Reiner AP, Ziv E, Lind DL, Nievergelt CM, Schork NJ, Cummings SR, Phong A, Burchard EGí, Harris TB, Psaty BM and Kwok PY. Population Structure, Admixture, and Aging-Related Phenotypes in African American Adults: The Cardiovascular Health Study. *The American Journal of Human Genetics*. 2005;76:463-477.
61. Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ, Bustamante CD, Kenny EE, Williams SM, Aldrich MC and Gravel S. The Great Migration and African-American Genomic Diversity. *PLoS Genet*. 2016;12:e1006059.
62. Shriner D. Overview of admixture mapping. *Curr Protoc Hum Genet*. 2017;94:1.23.1-1.23.8.
63. Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, Chen GK, Wang K, Buxbaum SG, Akylbekova EL, Aldrich MC, Ambrosone CB, Amos C, Bandera EV, Berndt SI, Bernstein L, Blot WJ, Bock CH, Boerwinkle E, Cai Q, Caporaso N, Casey G, Adrienne Cupples L, Deming SL, Ryan Diver W, Divers J, Fornage M, Gillanders EM, Glessner J, Harris CC, Hu JJ, Ingles SA, Isaacs W, John EM, Linda Kao WH, Keating B, Kittles RA, Kolonel LN, Larkin E, Le Marchand L, McNeill LH, Millikan RC, Murphy, Musani S, Neslund-Dudas C, Nyante S, Papanicolaou GJ, Press MF, Psaty BM, Reiner AP, Rich SS, Rodriguez-Gil JL, Rotter JI, Rybicki BA, Schwartz AG, Signorello LB, Spitz M, Strom SS, Thun MJ, Tucker MA, Wang Z, Wiencke JK, Witte JS, Wrensch M, Wu X, Yamamura Y, Zanetti KA, Zheng W, Ziegler RG, Zhu X, Redline S, Hirschhorn JN, Henderson BE, Taylor Jr HA, Price AL, Hakonarson H, Chanock SJ, Haiman CA, Wilson JG, Reich D and Myers SR. The landscape of recombination in African Americans. *Nature*. 2011;476:170-175.
64. Ritchie MD. Finding the Epistasis Needles in the Genome-Wide Haystack. In: J. H. Moore and S. M. Williams, eds. *Epistasis: Methods and Protocols* New York, NY: Springer New York; 2015: 19-33.
65. Li B, Verma SS, Veturi YC, Verma A, Bradford Y, Haas DW and Ritchie MD. Evaluation of PrediXcan for prioritizing GWAS associations and predicting gene expression. *Pac Symp Biocomput*. 2017;23.

66. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyler AE, Denny JC, Consortium GT, Nicolae DL, Cox NJ and Im HK. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47:1091-1098.
67. Verma SS, Frase AT, Verma A, Pendergrass SA, Mahony SA, Haas DW and Ritchie MD. Phenome-wide interaction study (PheWIS) in AIDS Clinical Trials Group Data (ACTG). *Pac Symp Biocomput*. 2016:5768-68.
68. Verma A, Bradford Y, Verma SS, Pendergrass SA, Daar ES, Venuto C, Morse GD, Ritchie MD and Haas DW. Multiphenotype associaton study of patients randomized to initiate antiretroviral regimens in AIDS Clinical Trials Group protocl A5202. *Pharmacogenet Genomics*. 2017;27:101-111.
69. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalin A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struewing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Little AR, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J and Moore HF. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580-585.
70. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R, Urban AE, Montgomery SB, Levinson DF and Koller D. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Research*. 2014;24:14-24.