Next Generation GWAS: Time to Focus on Phenotype?

Calum A. MacRae[1] and **Ramachandran S. Vasan**[2]

[1]Brigham and Women's Hospital and Harvard Medical School

[2]Boston University School of Medicine and The NHLBI's Framingham Heart Study

As investigators plan the next round of genome-wide association studies (GWAS), cohorts of over one hundred thousand individuals are being proposed as the solution to the 'missing heritability' from first generation studies[1-6]. While such studies will undoubtedly reveal many additional common alleles contributing to human disease, consideration of the intrinsic design of GWAS, our knowledge of the genetic architecture of disease and the results of GWAS to date suggest that other complementary strategies may well be necessary to reveal the hitherto undiscovered heritability.

First generation GWAS

The outcomes of GWAS to date might largely have been anticipated[1]. By design the approach offers an unbiased assessment of the role of common alleles in disease syndromes, and there are numerous examples of successful GWAS where multiple loci have been unequivocally associated with specific diseases. Even in genetically heterogeneous syndromes with major Mendelian contributions, shared common alleles that operate downstream to affect trait expression may be detectable[7, 8]. In less heritable traits, GWAS may offer the only evidence of any genetic contribution. Many of the most successful studies to date have explained only a proportion of the heritability, and for some key disease phenotypes no loci have been identified that achieve genome-wide statistical significance[3, 7-9]. Serendipitous identification of large-effect size common alleles has occurred, but these may be true modifiers of multiple Mendelian genes, genes with large interactions with environmental factors or just a single very homogeneous phenotype[10, 11].

While the 'small effect' alleles identified in GWAS may be important at a population level, a rigorous understanding the role of such contributions will require much more comprehensive understanding of the genetic architecture of each trait. Epistatic interactions among common alleles, or between common alleles and a range of rarer Mendelian or intermediate alleles that vary between individuals, may well contribute substantially to the unexplained heritability[12]. For the majority of human traits information on genetic architecture is scant, due the absence of unbiased family studies.

The simultaneous identification of multiple novel GWAS loci has highlighted a need for innovative approaches both to define the causal gene(s) at each locus, and to explore the fundamental mechanisms of disease[3, 4]. Pathway entry has not yet been attained for most GWAS loci as a result of the barriers to identification of the causal genes, including: the lack of definitive locus boundaries, small effect sizes and limited understanding of the regulatory functions of intergenic regions in *cis* or *trans*[13]. Even where biologic effects have been

Correspondence to: Calum MacRae, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, cmacrae@partners.org, Phone: 857-307-0301, Fax: 857- 307-0300.

Disclosures: None.

identified, it can be difficult to define the proportion of the effect at each locus that is attributable to the mechanism in question[14].

These predictable outcomes from GWAS serve to reinforce the implication from both human Mendelian genetics and clinical medicine that there may be substantial unidentified etiologic heterogeneity underlying many common disorders. Clearly, this genetic heterogeneity must be resolved if the comprehensive architecture of these traits is to be understood[1, 4, 15, 16]. In this context, increasing the size of many human disease cohorts is likely only to scale the heterogeneity in parallel. Thus, while such studies will offer the statistical power to detect larger numbers of alleles of even smaller effect size, they are not likely to be powerful enough to delineate gene-gene or gene-environment interactions.

Next generation phenotypes

At the core of many of the problems with GWAS and missing heritability is the issue of *phenotype resolution*[5, 6]. Most common diseases suffer from the lack of resolution and precision of their defining phenotypes. Indeed, phenotype resolution is likely a major determinant of the success or failure of GWAS to date. For traits where the phenotype is precise or serendipitously homogeneous (e.g., macular degeneration) large effect sizes are observed for select loci, and small cohorts have sufficed for performing 'successful' GWAS[10]. In contrast, in situations where the phenotype is less precise or subject to greater confounding (e.g., blood pressure), GWAS have yielded limited success even with impeccable design and large cohorts[3]. Clearly selection pressures also play a role in the genetic architecture for a given phenotype, and for phenotypes where there is little selection pressure on reproductive efficiency, genetic heterogeneity is less likely, while in the setting of more stringent selection even quite precise phenotypes may exhibit extensive genetic and allelic heterogeneity.

New diagnostic tools to discriminate homogeneous disease subsets, and to identify causally related but more penetrant 'endophenotypes' are urgently required[17-20]. 'Unbiased' phenotypes are emerging in model organism genetics, but large scale phenotype discovery in human disease has not been undertaken[21-23]. For many disease syndromes, the addition of even a limited number of orthogonal phenotypes may suffice to resolve the underlying heterogeneity[15, 19, 20, 24]. Functional genomic profiling of serum or tissue offers the possibility of discrete phenotypes, but to date the main successes have been in clonal neoplastic disorders. Innovative approaches to define etiology-specific functional assays range from tomographic electrical mapping to metabolomics and cellular profiling [25-27]. Maximizing the information content of phenotypic assays through dynamic testing will not only enable more rigorous genetics, but will also facilitate systems level analyses in biology and medicine[28]. Importantly, in order to redefine disease candidate phenotypes must be readily scalable and robust enough for routine clinical use. Understanding the expanding 'phenome' will require careful correlation with classic disease entities in well-characterized kindreds and in extended populations [18]. This new wave of clinical investigation will exploit translational research facilities and major population studies such as the NHLBI-funded Framingham Heart Study, ideally capturing multiple phenotypic axes in parallel. Many different phenotypes might potentially be used in downstream integrative studies, but the feasibility of translation to and from model systems is likely to be a critical attribute for the empiric definition of genetic architecture and for pathway entry or exploration[29].

Phenotypic homogeneity may be conferred by simple orthogonal features, such as the presence of an additional associated trait or the restriction to very specific demographic subgroup. Indeed, where genetic homogeneity exists, surprisingly small cohort studies may be sufficient to detect an underlying common allele. In this issue of *Circulation:*

Cardiovascular Genetics, Horne et al present a tantalizing report of GWAS in a cohort of only 40 patients with peripartum cardiomyopathy (PPCM)[30]. In this small cohort they were able to identify a locus with genome-wide significance on Chromosome 12 near the gene encoding parathyroid hormone-like hormone. This locus was validated in a second series of PPCM patients as well as in a cohort of individuals with pregnancy-associated cardiomyopathy that did not meet criteria for PPCM. Should we ignore these results simply based on the size of the discovery cohort? While it is conceivable that in this condition and in this population a single common allele is present and detectable in a study of this size, there are many reasons for caution. The limited number of unusual subjects certainly brings with it high false positive rates in GWAS, risks of population stratification, and problems with the identification of robust controls. Superimposed on these confounders is existing evidence that PPCM is not a homogeneous entity. Several reports suggest that PPCM is found in the context of several distinct autosomal dominant forms of dilated cardiomyopathy, and the data presented here are not inconsistent with a founder effect with reduced penetrance[31]. All of these issues may be directly addressed by additional replication studies in diverse cohorts with PPCM, but it may not be possible to convince skeptics without robust mechanistic modeling.

Next generation GWAS

Larger GWAS studies will definitely be performed in the near future, but the trade-off between statistical power and the effect size of alleles may limit their utility. Complementing these studies in the search for missing heritability will be next generation Mendelian studies where small kindreds will be subjected to whole genome sequence analysis. However, in order to fully define the genetic architecture of disease, it will be critical to employ hybrid approaches capable of identifying alleles with intermediate effect size, and ultimately of estimating gene-gene and gene environment interactions. Deconvoluting the genetic basis of complex disease at this resolution will require the integration of population and family-based studies in kin-cohort type designs[32], as well as novel multidimensional dynamic phenotypes that can be translated efficiently to and from model organisms.

Acknowledgments

References

1. Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet. 2001; 2:91–9. [PubMed: 11253062]

2. Bentley DR. The Human Genome Project--an overview. Med Res Rev. 2000; 20:189–96. [PubMed: 10797463]

3. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–78. [PubMed: 17554300]

4. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. Hum Mol Genet. 2008; 17:R156–65. [PubMed: 18852205]

5. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010; 11:446–50. [PubMed: 20479774]

6. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL,

Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–53. [PubMed: 19812666]

7. Ellinor PT, Yoerger DM, Ruskin JN, MacRae CA. Familial aggregation in lone atrial fibrillation. Hum Genet. 2005; 118:179–84. [PubMed: 16133178]

8. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, Palsson A, Blondal T, Sulem P, Backman VM, Hardarson GA, Palsdottir E, Helgason A, Sigurjonsdottir R, Sverrisson JT, Kostulas K, Ng MC, Baum L, So WY, Wong KS, Chan JC, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, Smith EE, Rosand J, Hillert J, Ma RC, Ellinor PT, Thorgeirsson G, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K. Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature. 2007; 448:353–7. [PubMed: 17603472]

9. Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M, West K, Kashuk C, Akyol M, Perz S, Jalilzadeh S, Illig T, Gieger C, Guo CY, Larson MG, Wichmann HE, Marbán E, O'Donnell CJ, Hirschhorn JN, Kääb S, Spooner PM, Meitinger T, Chakravarti A. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. Nat Genet. 2006; 38:644–51. [PubMed: 16648850]

10. Edwards AO, Ritter R 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. Science. 2005; 308:421–4. [PubMed: 15761121]

11. Genovese G, Friedman DJ, Ross MD, Lecordier L, Uzureau P, Freedman BI, Bowden DW, Langefeld CD, Oleksyk TK, Uscinski Knob AL, Bernhardy AJ, Hicks PJ, Nelson GW, Vanhollebeke B, Winkler CA, Kopp JB, Pays E, Pollak MR. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science. 2010; 329:841–5. [PubMed: 20647424]

12. Shao H, Burrage LC, Sinasac DS, Hill AE, Ernest SR, O'Brien W, Courtland HW, Jepsen KJ, Kirby A, Kulbokas EJ, Daly MJ, Broman KW, Lander ES, Nadeau JH. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. Proc Natl Acad Sci U S A. 2008; 105:19910–4. [PubMed: 19066216]

13. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science. 2008; 322:881–8. [PubMed: 18988837]

14. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Koteliansky V, Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S, Rader DJ. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010; 466:714–9. [PubMed: 20686566]

15. Anttila V, Kallela M, Oswell G, Kaunisto MA, Nyholt DR, Hamalainen E, Havanka H, Ilmavirta M, Terwilliger J, Sobel E, Peltonen L, Kaprio J, Farkkila M, Wessman M, Palotie A. Trait components provide tools to dissect the genetic susceptibility of migraine. Am J Hum Genet. 2006; 79:85–99. [PubMed: 16773568]

16. Terwilliger JD, Hiekkalinna T. An utter refutation of the "Fundamental Theorem of the HapMap". Eur J Hum Genet. 2006; 14:426–37. [PubMed: 16479260]

17. Singer E. "Phenome" project set to pin down subgroups of autism. Nat Med. 2005; 11:583. [PubMed: 15937456]

18. Freimer N, Sabatti C. The human phenome project. Nat Genet. 2003; 34:15–21. [PubMed: 12721547]

19. Cannon TD. The inheritance of intermediate phenotypes for schizophrenia. Curr Opin Psychiatry. 2005; 18:135–40. [PubMed: 16639165]

20. Cannon TD, Gasperoni TL, van Erp TG, Rosso IM. Quantitative neural indicators of liability to schizophrenia: implications for molecular genetic studies. Am J Med Genet. 2001; 105:16–9. [PubMed: 11424984]

21. Walhout AJ, Reboul J, Shtanko O, Bertin N, Vaglio P, Ge H, Lee H, Doucette-Stamm L, Gunsalus KC, Schetter AJ, Morton DG, Kemphues KJ, Reinke V, Kim SK, Piano F, Vidal M. Integrating interactome, phenome, and transcriptome mapping data for the C. elegans germline. Curr Biol. 2002; 12:1952–8. [PubMed: 12445390]

22. Rual JF, Ceron J, Koreth J, Hao T, Nicot AS, Hirozane-Kishikawa T, Vandenhaute J, Orkin SH, Hill DE, van den Heuvel S, Vidal M. Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. Genome Res. 2004; 14:2162–8. [PubMed: 15489339]

23. Ge H, Walhout AJ, Vidal M. Integrating 'omic' information: a bridge between genomics and systems biology. Trends Genet. 2003; 19:551–60. [PubMed: 14550629]

24. Garver DL, Holcomb JA, Christensen JD. Heterogeneity of response to antipsychotics from multiple disorders in the schizophrenia spectrum. J Clin Psychiatry. 2000; 61:964–72. quiz 73. [PubMed: 11206608]

25. Shaw SY, Westly EC, Pittet MJ, Subramanian A, Schreiber SL, Weissleder R. Perturbational profiling of nanomaterial biologic activity. Proc Natl Acad Sci U S A. 2008; 105:7387–92. [PubMed: 18492802]

26. Weckwerth W, Morgenthal K. Metabolomics: from pattern recognition to biological interpretation. Drug Discov Today. 2005; 10:1551–8. [PubMed: 16257378]

27. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB, Gerszten RE. Metabolite profiles and the risk of developing diabetes. Nature medicine. 2011; 17:448–53.

28. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011; 12:56–68. [PubMed: 21164525]

29. Deo RC, Macrae CA. The zebrafish:scalable in vivo modeling for systems biology. Wiley Interdiscip Rev Syst Biol Med. 2011; 3:335–46. [PubMed: 20882534]

30. Horne BD, Rasmusson KD, Alharethi R, Budge D, Brunisholz KD, Metz T, Carlquist JF, Connolly JJ, Porter TF, Lappé DL, Muhlestein JB, Silver R, Stehlik J, Park JJ, May HT, Bair TL, Anderson JL, Renlund DG, Kfoury AG. Genome-wide Significance and Replication of the Chromosome 12p11.22 Locus Near the PTHLH Gene for Peripartum Cardiomyopathy. Circulation Cardiovascular Genetics. 2011; 4:XXX–XXX.

31. Morales A, Painter T, Li R, Siegfried JD, Li D, Norton N, Hershberger RE. Rare variant mutations in pregnancy-associated or peripartum cardiomyopathy. Circulation. 2010; 121:2176–82. [PubMed: 20458009]

32. Gabriel SB, Salomon R, Pelet A, Angrist M, Amiel J, Fornage M, Attié-Bitach T, Olson JM, Hofstra R, Buys C, Steffann J, Munnich A, Lyonnet S, Chakravarti A. Segregation at three loci explains familial and population risk in Hirschsprung disease. Nat Genet. 2002; 31:89–93. [PubMed: 11953745]