Musings

Musings on genome medicine: genome wide association studies

David G Nathan and Stuart H Orkin

Address: Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA.

Correspondence: David G Nathan. Email: david_nathan@dfci.harvard.edu

Published: 20 January 2009

*Genome **Medicine*** 2009, **1**:3 (doi:10.1186/gm3)

The electronic version of this article is the complete one and can be found online at http://genomemedicine.com/content/1/1/3

© 2009 BioMed Central Ltd

We are grateful to the editors of *Genome Medicine*, who have rather incautiously invited us to write a monthly commentary on the exciting events that have occurred in this burgeoning field. We have decided to write the column together because one of us (DGN) is a clinical investigator and the other (SHO) is a basic geneticist and developmental biologist. Together, we believe we can do justice to the field and eschew cant. We begin with a discussion of the controversial role of genome-wide association studies in clinical medicine.

The development of practical approaches to DNA sequencing in the 1990s produced a remarkable scientific challenge - a proposal to establish the complete (or near complete) sequence of the human genome. Although most members of the scientific community and the media hailed the 2001 announcement of the project's initial success [1,2] as a huge intellectual and technical breakthrough, there were other voices [3]. One of us (SHO) was a member of the original US National Research Council panel that evaluated the proposal. The panel was initially highly skeptical but ended its deliberations with unbridled enthusiasm. Some leading scientists grumbled that the genome project, as it was called, was a quagmire and a money sump that had drained funds from individual investigators and provided a jumble of DNA bases the sequences of which would shed very little light on the human condition. The naysayers particularly emphasized their doubts that any medical benefit would be derived from most of the data. Indeed, when most of the human DNA sequence data had been collected, the laboratories that had accomplished the feat began to use their considerable resources to sequence the DNA of one animal species after another [4,5], with the questionable assumption that knowledge of DNA evolution would be useful and not a mere intellectual and technical exercise. Doubters began to wonder whether a large proportion of the biomedical research budget would be wasted in an effort to keep sequencing machines humming. The doubts were, in fact, quite loud in some quarters, despite the obvious fact that the project has provided investigators with ready access to all genes and facilitated positional cloning (see below).

Responding to the criticism, and always ebulliently optimistic, Francis Collins, the guiding spirit of the public effort to sequence the human genome, simply changed the subject. He proposed the human HapMap project [6] to replace laborious and relatively crude restriction enzyme maps. Now the National Institutes of Health was to finance a study of all or most of the common variations, rather than just the bases, in the human genome. Single nucleotide polymorphisms (SNPs) were to become the lingua franca of medical genetics. The HapMap project brought forth a bonanza for companies such as Affymetrix and Illumina, as common SNP detection moved into a broad base of laboratories and became a cottage industry.

The obvious potential application of the HapMap project to medicine lay in disease gene detection. This approach was initiated by YW Kan, who had shown that the sickle cell mutation in the first exon of the β globin gene could be predicted from a restriction enzyme polymorphism well downstream from the gene itself [7]. Kan's findings were based on David Botstein's proposal to use restriction enzymes as a tool for linkage mapping in humans [8]. Disease gene detection was then dramatically advanced by Louis Kunkel, SHO and their associates, who used what they termed 'reverse genetics' to detect a common muscular dystrophy gene and the most frequent chronic granulomatous disease gene [9,10]. They firmly established that disease genes could be defined by direct analysis of DNA without reliance on the availability of the offending protein. Others began to use a candidate gene approach. For example, if we did not already know the mutation that causes sickle cell anemia, it would be

rational to probe the α and β globin genes to find it because the disease is obviously due to a mutation in hemoglobin A. But sickle cell disease is a monogenic disorder, a perfect candidate for a candidate gene approach. What about the polygenic disorders such as obesity and diabetes? There could be 20 or more genes that interact to produce those syndromes. How can they be found - and how valuable would it be to find them?

Enter the HapMap project, carrying a huge assumption - that the disease genes that cause such common illnesses as diabetes and obesity have not been deleted by natural selection because such disorders so often occur after procreation has been achieved. Surely the sickle cell gene would have disappeared if it had not offered partial protection from the ravages of infantile falciparum malaria. So, the argument continued, the genes that contribute to obesity and diabetes could probably be detected by the association of the diseases with particular common SNPs.

Before one launches into a critique of genome-wide association studies, it is important to recognize that all of medicine is, as emphasized by Jerome Groopman, practiced by association [11]. The sacrosanct history and physical examination is almost totally based on association. When one reads the opening sentence of a patient's chart such as "This six year old African American male enters the hospital with a chief complaint of chest pain", several associations leap to the fore. He is only six. We do not associate coronary artery disease with that age. He is African American. We associate that race with sickle cell disease. He has chest pain. We associate that symptom with lung disease. We already wonder whether he has sickle cell disease and concomitant pneumonia. Even the physical examination is performed by association. We look for a tower skull and prominent maxillae because they are associated with sickle cell disease. We listen to breath sounds and associate each different sound with a unique pathology. Now, we could be entirely wrong. He might be a child with a genetic defect in the coagulation system who has had a pulmonary embolism. Reliance on association is clearly dangerous, but without association, the practice of medicine is crippled. Great clinicians such as the late Samuel A Levine, one of the fathers of clinical cardiology, associated large ears and light-colored hair with pernicious anemia. We don't know how many normal serum vitamin B12s DGN measured before he gave up on that association. Levine went to his grave convinced of its veracity.

The results of genome-wide association studies, in which common and complex diseases such as obesity and diabetes are associated with common SNPs, have been controversial for five main reasons: first, because the assumption that such SNPs actually exist has not been accepted in many quarters [12]; second, because, as Walter Bodmer has recently emphasized [13], much more powerful and historically established associations, such as the association of stomach ulcer and cancer with blood group A, have never been pathophysiologically explained; third, because the studies as currently constituted can detect only common variants and not rare ones; fourth, because the current studies have explained only a small proportion of the heritability of common multigenic diseases; and finally, because it is not at all clear that any useful therapy can emerge from the associations. Those who pursue the associations argue that we can determine the risk of such diseases and ward them off. But the arcane statistics used to establish the associations give rise to relative risks of such low order that it would seem foolhardy to use the weak data in a burst of what is called personalized medicine. Many of the studies are underpowered; still more are not reproducible. It would seem just as rational to advise men with long ears and light-colored hair to inject themselves with vitamin B12.

Then there is the growing use of genome-wide association studies to determine clinically important aspects of pharmacogenetics. There are certainly genetic bases for drug sensitivity or resistance; the relationship of warfarin sensitivity to cytochrome p450 polymorphism is an example [14]. But will widespread adoption of SNP analysis of two such genes really contribute to the management of warfarin therapy? There are many other acquired causes of warfarin sensitivity or resistance. Although the test has gained Food and Drug Administration approval, the jury is out on its utility. It certainly increases the costs of treatment: that's all we know right now.

More problematic applications of genome-wide association studies can be found in attempts to wander down the genome with gun and camera and relate common SNPs to clinical severity. Sickle cell disease is a case in point [15]. We know that there are globin and non-globin genes that modify the severity of sickle cell disease. The level of fetal hemoglobin is a massive modifier. Not surprisingly, the relationship of red cell membrane area to volume is a modifier because sickle red cells, like normal cells, must squeeze through tiny apertures. Therefore, concomitant α-thalassemia is a modifier. The red cell water content is a critical modifier because sickling is closely related to hemoglobin concentration. These are hugely productive areas to investigate that would surely lead to better treatment. To spend valuable research dollars wandering around the genome to find 'modifiers' with a 1% or 2% effect seems ridiculous on the surface. There must be some research priorities. Fads such as genome-wide association studies do have an important role in certain circumstances when modifiers are unknown. But in sickle cell disease the important modifiers are already known. Let's use our increasingly limited ammunition to go after the obvious opportunities.

Ours is perhaps a dour view. But we are seasoned hands who have seen many biomedical fads appear, take the front of the stage and then return to the wings as other actors enter from

behind an arras. We are conditioned by our experience to go after the obvious and the doable. But, despite our doubts, we remain hopeful that useful candidate genes and genetic pathways are likely to emerge from genome-wide association studies if the studies are performed under stringent conditions, are sufficiently powered and are thoroughly reproduced [16,17]. Indeed, one of us (SHO) has recently used data from genome-wide association studies performed by others to explore genes that might modify fetal hemoglobin expression in the hemoglobinopathies [18]. So there is surely something to be gained from this new approach, but we need to keep it in perspective and always focus most of our research resources on experiments that, in the end, are most likely to contribute to biomedical science and patient care, now and in the future.

References

1. **Human genomes, public and private.** *Nature* 2001, **409:**745.
2. **The human genome. Science genome map.** *Science* 2001, **291:**1218.
3. **Great 15-year project to decipher genes stirs opposition.** [http://query.nytimes.com/gst/fullpage.html?res=9C0CEEDD1139F936A35755C0A966958260]
4. Chowdhary BP, Raudsepp T: **The horse genome.** *Genome Dyn* 2006, **2:**97-110.
5. Cockett NE: **The sheep genome.** *Genome Dyn* 2006, **2:**79-85.
6. The International HapMap Consortium: **The International HapMap Project.** *Nature* 2003, **426:**789-796.
7. Kan YW, Dozy AM: **Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation.** *Proc Natl Acad Sci USA* 1978, **75:**5631-5635.
8. Botstein D, White RL, Skolnick M, Davis RW: **Construction of a genetic linkage map in man using restriction fragment length polymorphisms.** *Am J Hum Genet* 1980, **32:**314-331.
9. Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, Cole FS, Curnutte JT, Orkin SH: **Cloning the gene for an inherited human disorder—chronic granulomatous disease—on the basis of its chromosomal location.** *Nature* 1986, **322:**32-38.
10. Monaco AP, Bertelson CJ, Middlesworth W, Colletti CA, Aldridge J, Fischbeck KH, Bartlett R, Pericak-Vance MA, Roses AD, Kunkel LM: **Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment.** *Nature* 1985, **316:**842-845.
11. Groopman J: *How Doctors Think*. New York: Houghton Mifflin; 2007.
12. **A dissenting voice as the genome is sifted to fight disease.** [http://www.nytimes.com/2008/09/16/science/16prof.html?partner=permalink&exprod=permalink]
13. Bodmer W, Bonilla C: **Common and rare variants in multifactorial susceptibility to common diseases.** *Nat Genet* 2008, **40:**695-701.
14. Cooper GM, Johnson JA, Langaee TY, Feng H, Stanaway IB, Schwarz UI, Ritchie MD, Stein CM, Roden DM, Smith JD, Veenstra DL, Rettie AE, Rieder MJ: **A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose.** *Blood* 2008, **112:**1022-1027.
15. Sebastiani P, Nolan VG, Baldwin CT, Abad-Grau MM, Wang L, Adewoye AH, McMahon LC, Farrer LA, Taylor JG 4th, Kato GJ, Gladwin MT, Steinberg MH: **A network model to predict the risk of death in sickle cell disease.** *Blood* 2007, **110:**2727-2735.
16. McCarthy MI, Hirschhorn JN: **Genome-wide association studies: past, present and future.** *Hum Mol Genet* 2008, **17:**R100-R101.
17. Altshuler D, Daly MJ, Lander ES: **Genetic mapping in human disease.** *Science* 2008, **322:**881-888.
18. Lettre G, Sankaran VG, Bezerra MA, Araújo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN, Orkin SH: **DNA polymorphisms at the BCL11A, HBS1L-MYB, and β-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease.** *Proc Natl Acad Sci USA* 2008, **105:**11869-11874.