The 'G' That Never WAS

Friedrich C. Luft

Experimental and Clinical Research Center, Charité Medical Faculty and Max Delbrück Center for Molecular Medicine, Berlin, Germany

The Man Who Never Was is a 1953 nonfiction book based on 'Operation Mincemeat', a 1943 British Intelligence plot to deceive the Nazis into believing that the invasion of Sicily would take someplace else. Briefly, the subterfuge involved dressing up a cadaver as a British major and dumping him into the sea off the coast of Spain. The gullible Germans swallowed the story and directed their attention to Sardinia instead. The lesson for us is to remain ever wary and skeptical. Flashy claims and trapings do not necessarily translate into solid science. The current 'breakthrough' in medical science that we need to regard very critically is the genome-wide association study (GWAS).

GWAS are generally case-control (cohort or trio designs are also possible) association studies that are based on the 'common disease, common variant' hypothesis [1]. This assumption maintains that relatively few common genetic variants can identify genetic risk for a common disease. GWAS have in part replaced the myriads of 'candidate gene' association studies, in which one or several polymorphisms in a gene of interest are tested for statistical association with a trait or condition. Examples of the latter are the 5,000 or so published papers on a favorite *Alu* repeat in the angiotensin converting enzyme gene (*ACE*), the 'insertion-deletion' polymorphism. GWAS are generally much larger (although not always) and use high throughput technology to define thousands of polymorphisms, generally single nucleotide polymorphisms (SNP). We have about 3,000,000 of these variants in our

3,000,000,000 bp (fig. 1). Gene chips that can determine 900,000 SNP scattered across the entire genome (genome-wide) in each individual subject are now common. The principle is to seek genetic, i.e. SNP, variants that frequently occur in the group with the disease (or interme-

Fig. 1. We are blessed with an SNP about every 1,000 bp. GWAS has emerged as a hypothesis-free method of screening the entire genome for disease-related genetic variants through the analysis of SNP. We now have the technology to determine relevant SNP for genetic purposes.

diary phenotype) compared to the control population. If genetic variants are more common in people with the disease, the variations are said to be 'associated' with the disease. The associated genetic variants are then possible indicators suggesting that the disease-causing problem resides nearby on the genome. Hypothesis-driven methods start with the hypothesis that a particular gene may be associated with a particular disease and try to find the association. GWAS are non-hypothesis-driven studies that scan the entire genome.

Oddly, most of the SNP variants associated with disease are not in the region of DNA that codes for a protein. Instead, they are commonly in the large non-coding regions on the chromosome between genes or in intron sequences that are edited out of the DNA sequence when proteins are processed. These variants presumably are, or are close to, sequences of DNA that control other genes. However, their function is largely unknown. About 1,000 GWAS studies have been reported. Since the 'players' here are generally investigators in control of large sample numbers, commonly from clinical trials or from epidemiological cohorts, GWAS have frequently catapulted trialists and epidemiologists into becoming molecular geneticists. Essentially, if you are into 'networking' and the word 'hypothesis' causes you to break out in a rash, this kind of research may be for you.

In this issue, Murea et al. [2] report the results of a GWAS involving African-American dialysis patients with type 2 diabetes mellitus. The breakdown variable was survival on dialysis. The endpoint was death from any cause since the start of dialysis. The analysis involved 610 such patients. About 5 years later, 2/3 of the patients were dead. They died of the usual reasons for death in dialysis patients. The 130 survivors allowed a glance into genomic regions associated with survival. The authors list their top 12 SNP findings; almost all were intronic; 1 resided in a promoter region. However, the location has little or nothing to do with their probable relevance. The authors also provide a table of intergenic SNP. Interesting genes appear here as well. However, what are we to do with such results? Dialysis patients have many different potential causes of death. Moreover, type 2 diabetic patients do not all die of cardiovascular disease or nephropathy. A recent report pointed out the very important fact that type 2 diabetic patients have an increased risk of dying of cancer, but are also at an increased risk of dying of non-cardiovascular/non-cancer causes, including trauma and self harm [3]. The African-Americans analyzed by Murea et al. [2] were 'self-reported', which is politically correct but scientifically imprecise, since the Afri-

can continent is a heterogeneous place. Population stratification and disease heterogeneity are the bane of GWAS. Thus, there are many confounding variables in this study.

Some potential limitations of GWAS are false-positive and false-negative results, insensitivity to rare variants and structural variants, requirements for large sample sizes, genotyping errors (not the major problem with the chip used here), ignorance of gene function, and biases when selecting cases and controls. Here, the grim reaper did the selections, but we cannot trust him either. What needs to be done? A robust sample size for a disease such as type 2 diabetes mellitus is said to be about 10,000 individuals. Furthermore, after a disease-associated region has been identified, a replication cohort is generally required to provide credibility. Next, some attempt at identifying functional relevance should be made. The variants can be transfected into cell models and tested for function, mouse models can be developed etc. Such efforts are expensive, require expertise that cannot solely be provided by a few individuals, and oftentimes are just not feasible.

GWAS statistics are not for the faint-hearted. For ease of display, association statistics are typically shown as the $-\log_{10}$ of the p value [4]. The p value remains the probability of the observed association arising by chance alone, so that $p = 0.01$ would be plotted as '2' on the y-axis and $p = 10^{-7}$ as '7'. Murea et al. [2] provide such a plot showing 5 SNP residing above 6. However, it is difficult for us to discern what this value means in terms of explaining the genetic variance on the phenotype (dying earlier on dialysis). Such displays also often plot a matrix of r^2 values for each pair of SNP in the region, with larger r^2 values more intensely shaded. These plots can be used to identify linkage disequilibrium blocks containing SNP associated with disease, allowing estimation of the independence of the SNP associations observed. Murea et al. [2] also show us a Q-Q plot. The Q-Q plot is used to assess the number and magnitude of observed associations between genotyped SNP and the disease or trait under study, compared to the association statistics expected under the null hypothesis of no association [3]. Their black line does not vary from the red line that represents the expected line under a null distribution. A strong deviation from the null line would suggest a very heavily associated or heavily genotyped locus or differences in population structures. Neither was present here.

For a little more practice, we could take a brief look at another GWAS published recently that should interest nephrologists [5]. The UK Prospective Diabetes Study (UKPDS) consortium (remember those papers?) recently

published a GWAS for glycemic response to metformin in 1,024 Scottish individuals with type 2 diabetes with replication in 2 cohorts including 1,783 Scottish individuals and 1,113 individuals from UKPDS. You will recall that metformin was about the only hypoglycemic treatment of value in that study. The group performed a combined meta-analysis, suggesting that individual cohorts were not robust alone. They identified a SNP associated with treatment success ($p = 2.9 \times 10^{-9}$). This finding represented an OR = 1.35 at a locus containing the ataxia-telangiectasia-mutated (*ATM*) gene. It is not a bank machine. The *ATM* gene encodes a protein that recognizes and corrects errors in duplicating DNA when cells divide and is responsible for destroying cells when the errors cannot be corrected. Metformin activates the AMP-activated protein kinase (AMPK) by inhibiting the mitochondrial respiratory chain and increasing cellular AMP. Why was *ATM* a candidate? Patients with ataxia telangiectases happen to have an increased risk of diabetes. An inhibitor for the *ATM*-encoded protein (KU-55933) is available. The investigators next used a rat hepatoma cell line and found that inhibition of *ATM* with KU-55933 attenuated the phosphorylation and activation of AMP-activated protein kinase in response to metformin. The investigators were very fortunate. Their GWAS worked out;

the confirmatory cohorts worked out as well. Patients receiving metformin as a monotherapy confirmed the findings. A candidate gene arose that is associated with the disease. A cell model at least supports the investigators' assumption. So, the group was justified in concluding that *ATM*, a gene involved in DNA repair and cell cycle control, could play a role in the effect of metformin upstream of AMP-activated protein kinase. Variations in this gene appear to alter the glycemic response to metformin. Much work remains ahead to bring one mechanism to another.

Murea et al. [2] were less fortunate and I would argue that their problem is even more difficult to solve. They need a replication cohort or two. This request is not impossible, given the many dialysis patients in the USA and elsewhere and the frequency of type 2 diabetes. They may need to delineate causes of death precisely because presumably, type 2 diabetes increases the all-cause risk for death in dialysis patients as it does in the general population. People die for many reasons. Therefore, asking for a mechanism of death may be 'going overdrawn'. Was this effort an 'operation mincemeat'? Perhaps not, but you need not memorize the findings for the next nephrology boards.

References

- 1 Manolio TA: Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010;363:166–176.
- 2 Murea M, Lu L, Ma L, Hicks PJ, Divers J, McDonough CW, Langefeld CD, Bowden DW, Freedman BI: Genome-wide association scan for survival on dialysis in African-Americans with type 2 diabetes. *Am J Nephrol* 2011;33:502–509.
- 3 Seshasai SR, Kaptoge S, Thompson A, Di Angelantonio E, Gao P, Sarwar N, Whincup PH, Mukamal KJ, Gillum RF, Holme I, Njølstad I, Fletcher A, Nilsson P, Lewington S, Collins R, Gudnason V, Thompson SG, Sattar N, Selvin E, Hu FB, Danesh J: Diabetes mellitus, fasting glucose, and risk of cause-specific death. *Emerging Risk Factors Collaboration. N Engl J Med* 2011;364:829–841.
- 4 Pearson TA, Manolio TA: How to interpret a genome-wide association study. *JAMA* 2008;299:1335–1344.
- 5 GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group, Wellcome Trust Case Control Consortium 2, Zhou K, Bellenguez C, Spencer CC, Bennett AJ, Coleman RL, Tavendale R, Hawley SA, Donnelly LA, Schofield C, Groves CJ, Burch L, Carr F, Strange A, Freeman C, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin A, Craddock N, Deloukas P, Dronov S, Duncanson A, Edkins S, Gray E, Hunt S, Jankowski J, Langford C, Markus HS, Mathew CG, Plomin R, Rautanen A, Sawcer SJ, Samani NJ, Trembath R, Viswanathan AC, Wood NW, MAGIC investigators, Harries LW, Hattersley AT, Doney AS, Colhoun H, Morris AD, Sutherland C, Hardie DG, Peltonen L, McCarthy MI, Holman RR, Palmer CN, Donnelly P, Pearson ER: Common variants near *ATM* are associated with glycemic response to metformin in type 2 diabetes. *Nat Genet* 2011;43:117–120.