

Musings

Musings on genome medicine: the value of family history

Angus John Clarke

Address: Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK.
Email: clarkeaj@cardiff.ac.uk

Abstract

Will the routine availability of genome sequence information on individuals render family history information obsolete? I argue that it will not, both because the taking of a family history has other uses for the health professional, and because genome sequence data on their own omit the effects of numerous factors important for modifying risks of disease. These include information derived from factors downstream of genetic variants and from upstream epigenetic effects. Further difficulties arise with uncertainties relating to gene-gene and gene-environment interactions, which may take decades to resolve if their resolution is even possible.

Introduction

Family history information has been used for decades in the process of genetic risk assessment [1] and is used extensively to identify those at risk of inherited conditions and those at increased risk of common diseases, such as colorectal cancer, breast cancer and coronary artery disease [2]. Those at risk may then benefit from targeted genetic testing to distinguish between those who are actually at high risk of disease and those who are in fact not. Once genomic sequence information has become readily available, on patients and on healthy individuals, will this completely obviate the need for family history information, so that the taking of family histories becomes obsolete? In short, can everything important be written with just the four letters A, C, G and T?

Broader benefits of taking a family history

Undergraduate medical students are taught to take a family history for all their patients and this is also recommended in many clinical specialty training programs, including those for primary care practitioners. It not only helps to identify those at increased risk of an inherited disorder or of a single-gene (Mendelian) subset of a common disorder, but also provides useful social information about the patient and his or her family. Even if this family information were to add no further precision to estimates of disease risk, the clinician would still wish to continue collecting family histories for the additional insights they provide. One gathers not only the structure of the family

and some knowledge of the diseases that have affected the various individuals but also a rich understanding of social context, occupation, family relationships (who talks to whom), and perhaps lifestyles and habits too. The process of collecting this information allows a relationship to develop between professional and patient, which is useful to the practitioner and valued by the patient [3]. Through this, one can often sense the real concerns of the patient or client: what do they fear? Do they feel destined to follow the pattern of health and illness of one or more relatives? On whose behalf are they most apprehensive? These insights allow the practitioner to anticipate the issues that are likely to be sensitive and difficult for the patient; they are also most helpful when tackling a sensitive issue tactfully once it has arisen.

The process of taking a family history can sometimes be more problematic than this account may suggest, however, as when family and professional differ in their understandings of relevant factors or of what counts as a family history [4,5]. Similarly, although taking a family history for risk assessment has been presented as unproblematic [6], others have found that practitioners in primary care may find it difficult to switch between lay and professional terminologies and ways of speaking [7]. These difficulties underline the importance of taking a family history well, if it is to be done at all, but do not undermine the benefits of doing so.

Assessment of genetic risk in the common complex disorders

Turning to the science of the matter, one must then look at the adequacy of genomic sequence information as a secure basis for making disease risk assessments, and, therefore, as a basis for health care decisions. Here, I consider this especially in relation to the common, complex disorders such as coronary artery disease, colorectal cancer, breast cancer and diabetes mellitus. The question at issue can be formulated as, 'What obstacles and complexities do (or might) make genome sequence data unreliable for predicting long-term health outcomes, at least for today?' I suggest the following four answers to this question.

SNP, single nucleotide polymorphism.

First, the sheer volume of genomic, clinical and environmental (including lifestyle and dietary) data to be gathered and analyzed is daunting if gene-gene and gene-environment interactions of clinical relevance are to be identified and measured [8]. Such effects are likely to be of great importance but difficult to detect if the relative effects of the two principal alleles at a particular locus are roughly equivalent, when averaged over the two sexes, over the range of available environments and over all allelic combinations at other interacting loci - as will be the case for any stable polymorphism. It seems intrinsically unlikely that such interactions would be less important in humans than in *Drosophila*, for example, but of course they are much more difficult to detect in our own, so very inconvenient species [9,10]. The fact that they are there, however, is clear from the existence of modifier loci and other effects in relation to both complex and Mendelian disorders. In the case of breast cancer, for example, those who test negative for a *BRCA1* or *BRCA2* allele known to be associated with breast cancer in their relatives are still at substantially increased risk of this malignancy [11,12]. Most polymorphic alleles will have no net effect overall, except when selection is currently leading to a substantial net alteration in allele frequencies.

Just as genetic factors that modify disease susceptibility may or may not reveal themselves in overt disease - tests for such factors are not highly predictive - so family history information gives only incomplete information about the risks associated with the future health of people today. Testing to see whether a specific genetic variant has been transmitted from parent to child is possible only when the responsible variant has been identified as such, and that process is difficult even in the coding regions of many Mendelian disease genes. In the setting of genetic variants that could influence the risk of the common, complex diseases, we have barely begun the work required [13]. Full sequence information on even a large number of people with and without disease will not easily yield the sought-after risk information; indeed, the quantity and quality of phenotypic and environmental data required to optimize interpretation of the genomic data remains unknown and may well greatly exceed the complexity of the genomic data. Once a lack of genomic sequence data is no longer the immediate block to our understanding, the rate-limiting step in the generation of knowledge will be the collection of these other types of data and their joint analysis.

Second, the population(s) of origin of an individual will be very relevant to the interpretation of genome-wide association studies, because the particular set of alleles favored during times of adversity in the experience of one population in historical or pre-historical times may differ from the alleles favored in a different group, even if the initial allele frequencies and the challenges facing the two groups were much the same.

Upstream and downstream

A third answer to the question is that 'downstream' factors that mediate the effects of genetic susceptibility - such as the effects of blood pressure or serum cholesterol mediating part of the coronary artery disease risks of genetic variants - may provide more useful risk information than sequence data alone. It will be necessary to separate out the independent from the mediated effects, especially if risk assessments use the downstream data in addition to genomic data.

Finally, information about the 'upstream' modulators of gene activity, such as the methylation of specific CpG groups, histone modifications and changes in chromatin conformation, is much more complex to generate and analyze than raw genome sequence data. Such factors influencing gene expression may also vary during development and between tissues. Modified sequencing approaches can indicate the pattern of DNA methylation found in the tissues examined, but many of these epigenetic effects are likely to be missed by a strategy focused purely on sequencing. Furthermore, the specific epigenetic changes that mediate the long-term consequences of early (intra-uterine or postnatal) experience are only beginning to be recognized. It is unlikely that genome sequence data will include the markers of these epigenetic effects for many years and, until then, it will be difficult to incorporate these 'early experiences' into any genome-wide analysis of disease susceptibility.

Conclusion

The availability of genome-wide analysis of common variants - especially the single nucleotide polymorphisms (SNPs) used in numerous major studies of disease susceptibility - has led to important insights into the components of disease susceptibility that can be traced back through pre-historical times. The recognition that copy number variants are important contributors to some categories of disease - and that disease-associated CNVs will usually be of recent origin as contrasted with the prehistoric origin of most SNPs - has given the chance for some of the limitations of SNP-based genome-wide association studies to be overcome. The soon-to-be-fulfilled promise of the ready availability of human genome sequence data on large numbers of individuals will open a new era of genetic and health-related research, but it will take time before the lessons of this research allow us to assess the limits of disease risk prediction from the integrated analysis of sequence data with clinical, dietary, lifestyle and environmental information; this will most likely take decades. Our inability to measure, because of the intrinsic methodological difficulties, the various gene-gene and gene-environment interactions and the way they affect chromatin and DNA means that over this timescale our disease risk assessments based on DNA sequence information will remain, in an important sense, incomplete.

It is therefore clear that family history information will remain important in assessing disease risks for many years; we simply do not yet know whether genome sequence data will ever be able to substitute for this.

This understanding of the limitations of genetic test results will be important in the realm of health care and may be even more important for the customers of those purveyors of lifestyle guidance, who claim that their 'insights' are inspired by genomic information. These commercial operators will need to learn how to obtain and then to incorporate family history information into their genome-based risk assessments. Until this has been achieved and demonstrated to be robust and of clinical utility, let the buyer beware: *caveat emptor!* The role of the state will be important in setting enforceable standards that such commercial operators should meet.

Competing interests

The author declares that he has no competing interests.

References

1. Cole J, Conneally PM, Hodes ME, Merritt AD: **Genetic family history questionnaire.** *J Med Genet* 1978, **15**:10-18.
2. Butterworth A: *Family History as a Risk Factor for Common, Complex Disease.* Cambridge, UK: Public Health Genetics Unit; 2007.
3. Skirton H: **The client's perspective of genetic counselling- a grounded theory study.** *J Genet Couns* 2001, **10**:311-329.
4. Hunt K, Emslie C, Watt G: **Lay constructions of a family history of heart disease: potential for misunderstandings in the clinical encounter?** *Lancet* 2001, **357**:1168-1171.
5. Walter FM, Emery J: **Perceptions of family history across common diseases: a qualitative study in primary care.** *Fam Pract* 2006, **23**:472-480.
6. Scheuner MT, Wang S-J, Raffel LJ, Larabell SK, Rotter JJ: **Family history: a comprehensive genetic risk assessment method for the chronic conditions of adulthood.** *Am J Med Genet* 1997, **71**:315-324.
7. Hall R, Saukko PM, Evans PH, Quereshi N, Humphries SE: **Assessing family history of heart disease in primary care consultations: a qualitative study.** *Fam Pract* 2007, **24**:435-442.
8. Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schafer H: **Gene- environment interactions for complex traits: definitions, methodological requirements and challenges.** *Eur J Hum Genet* 2008, **16**:1164-1172.
9. Mackay TFC: **The genetic architecture of quantitative traits: lessons from *Drosophila*.** *Curr Opin Genet Dev* 2004, **14**:253-257.
10. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF: **Systems genetics of complex traits in *Drosophila melanogaster*.** *Nat Genet* 2009, **41**:299-307.
11. Smith A, Moran A, Boyd MC, Bulman M, Shenton A, Smith L, Iddenden R, Woodward ER, Laloo F, Maher ER, Evans DG: **Phenocopies in BRCA1 and BRCA2 families: evidence for modifier genes and implications for screening.** *J Med Genet* 2007, **44**:10-15.
12. Katki HA, Gail MH, Greene MH: **Breast-cancer risk in BRCA-mutation-negative women from BRCA-positive families.** *Lancet Oncol* 2007, **8**:1042-1043.
13. Janssens AC, Gwinn M, Bradley LA, Oostra BA, van Duijn CM, Khoury MJ: **A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions.** *Am J Hum Genet* 2008, **82**:593-599.

Published: 03 August 2009
doi:10.1186/gm75
© 2009 BioMed Central Ltd