



The postgenomic era and complex disease



J A Witkowski
Cold Spring Harbor
Laboratory,
Banbury Center,
PO Box 534,
Cold Spring Harbor,
NY 11724-0534 USA

‘One of the great hopes for studying complex disorders is that the new tools of human genomics will help in dissecting the multiple factors underlying individual responses to those factors.’

The articles published in this special issue of *Pharmacogenomics* culminate from a multidisciplinary group of researchers, demonstrating that data integration and multidisciplinary collaboration can yield novel approaches for handling large, complex data sets and reveal new insight and relevance to a complex illness such as chronic fatigue syndrome (CFS). Data were generated from a 2-day clinical study of CFS and comprised clinical, genomic and genetic information. The data, including over 500 clinical and epidemiological measurements and 20,000 gene expression measurements, was shared with 20 investigators who were challenged with integrating the data to delineate the heterogeneity of the study population and identify biological correlates of CFS.

As Suzanne Vernon and William Reeves describe in their overview of the Wichita (KS, USA) CFS project [1], diseases such as CFS are a major challenge to public health and to investigators trying to understand the causes of such disorders. One of the great hopes for studying complex disorders is that the new tools of human genomics will help in dissecting the multiple factors underlying individual responses to those factors. Genetic and biochemical individuality was recognized long ago, and this editorial discusses pharmacogenomics in the historical context of the development of human genetics.

Just over 100 years ago, the first step was taken in understanding the genetic basis of human disorders. Archibald Garrod, then a young physician working at St Bartholomew's Hospital in London, UK, observed a patient with urine that turned black on exposure to air. This was not a new observation, as was clear from the monograph that Garrod published in

1909, *Inborn Errors of Metabolism* [2]. What was new was Garrod's recognition of the inherited nature of alkaptonuria and his suspicion that its pattern of inheritance might conform to the then newly discovered research of Gregor Mendel. Garrod wrote to William Bateson, the champion and promoter of Mendel's work, for his opinion. Bateson included a footnote in one of his reports to the Evolution Committee of the Royal Society noting that, indeed, Garrod's findings and his review of the literature on alkaptonuria indicated that alkaptonuria behaved like a Mendelian recessive trait [3]. In his 1908 paper in the *Lancet*, Garrod pointed out that alkaptonuria was unique, and that albinism and cystinuria were also chemical 'sports' [4].

What is remarkable in this paper was Garrod's insistence that there might be many chemical differences between individuals that were not evident by simple inspection. Moreover, in their response to 'different drugs and infecting organisms, the members of the various genera and species manifest peculiarities which presumably have a chemical basis.' Here, surely, is the key insight of the uniqueness of individuals in their responses to the environment, an insight that 100 years later has reappeared as pharmacogenomics and personalized medicine.

Gradually, through the 1950s, human genetics re-emerged as clinical genetics. The human genetic disorder sickle cell anemia was the first molecular disease to be defined through the work of Linus Pauling, Harvey Itano and Vernon Ingram [5-7]. However, it was technical developments in human cytogenetics that led the way. First, Tjio and Levan demonstrated that there are 46 human chromosomes, and not 48 as had been thought for over 20 years [8]. Lejeune's finding that there are three, rather than two, copies of a chromosome in individuals with Down's syndrome led to a proliferation of similar studies [9]. Other trisomies were discovered and chromosomal abnormalities were found in Turner's syndrome and Klinefelter syndrome. Now there was a demonstrable utility for human clinical genetics. The biochemical bases of inherited metabolic disorders such as phenylketonuria, maple syrup urine disease,

Tay-Sachs disease, and Hurler and Hunter syndromes were determined and diagnostic tests developed. However, these were exceptions, and a molecular understanding of the hundreds of human genetic disorders cataloged by Victor McKusick in his *Mendelian Inheritance in Man* [10] remained elusive.

The next major steps came with the recombinant DNA revolution of the 1970s. For the first time, geneticists could examine human genes at the molecular level and, for example, determine the nucleotide mutations in β -globin that caused the thalassemias. Some years later, Botstein, White, Skolnick and Davis ushered in a new era in genetics by developing a general approach to finding genes involved in human genetic disorders [11–13]. Linkage analysis using restriction fragment length polymorphisms (RFLPs) opened the way to finding the genes involved in Duchenne muscular dystrophy, cystic fibrosis and Huntington's disease. Through the 1980s and 1990s, hunting disease genes using RFLPs and other genetic markers, such as microsatellites, became a major industry, and the genes involved in many genetic disorders were located, cloned and analyzed.

‘It is dangerous to put all one’s faith in the ‘technological fix’, but it is undoubtedly true that the molecular tools to study diseases have become much more sophisticated.’

However, there was a group of disorders that resisted the efforts of many investigators. Most notably, linkage analysis of mental disorders such as schizophrenia, manic depression and autism found many chromosomal locations that were associated with these disorders. For example, ‘regions of interest’ were found on chromosomes 4, 12, 16, 21 and X for manic depression; 1, 6, 8, 10, 13, and 22 for schizophrenia; and 2, 3, 7, 13, 16, and 17 for autism. It was not clear how to interpret these findings. Some of the linkages did not reach the level of significance proposed by Lander and Kruglyak [14], and the locations on a chromosome were not necessarily found to be the same in different studies. Replication of a linkage result in these disorders proved the exception, rather than the rule. An alternative strategy of selecting candidate proteins based

on function and studying the relevant genes in affected individuals has, by and large, also proved unrewarding.

It is clear that a number of different factors contributed to these disappointing results. There are no biomarkers to identify patients, and as diagnostic criteria are subjective, the populations being studied may not be homogeneous. The disorders are genetically complex – all the identified regions for each disorder may contain genes contributing to schizophrenia, manic depression and autism, but any one gene may have only a moderate or small effect. Furthermore, the impact of environmental factors is difficult to assess and hard to control for in the populations used for study.

There are some parallels with studies of CFS. The diagnosis is difficult, the genetic factors are complex, and environmental factors, while suspected, are largely unidentified. What hope then is there for unraveling the causes of these complex disorders? It is dangerous to put all one’s faith in the ‘technological fix’, but it is undoubtedly true that, as Vernon and Reeves stated, “the molecular tools to study diseases have become much more sophisticated” [1]. Microarray chips enable the simultaneous measurement of the expression of thousands of genes, and new sequencing techniques promise the ‘\$1000 genome,’ making it possible to sequence very large sections of human genome in thousands of individuals. Most importantly, the HapMap project will provide a vast number of genetic markers that can be used for genetic analysis, and in formats that will enable very high sample throughput (and hence less expensive) analysis.

Such genetic analysis may not be enough in the absence of reliable diagnostic criteria. How can this be done? There is a circularity in that accurate tests are needed. However, these require validation, which requires accurate identification of affected individuals, which requires accurate tests, and so on. The Wichita study of CFS attempts to break this cycle by performing a wide-ranging set of analyses – medical evaluation, psychiatric evaluation, laboratory tests and gene expression profiling. However, such a study produces a large quantity of data of different types, which has to, in some way, be reconciled. The studies on CFS reported in this issue of *Pharmacogenomics* are an heroic attempt to do so, and demonstrate how complex problems require the participation of individuals from very diverse

backgrounds. While large-scale biology projects do produce vast amounts of data, issues of data storage, manipulation and analysis on far larger scales have been faced successfully by other scientific disciplines. There is every reason to believe that continuing technical and intellectual advances will enable us to understand the basis of complex disorders such as CFS and, more importantly, lead to ways of improving the lives of those afflicted by them.

Bibliography

1. Vernon SD, Reeves WC: The challenge of integrating disparate high-content data: epidemiological, clinical and laboratory data collected during an in-hospital study of chronic fatigue syndrome. *Pharmacogenomics* 7(3), 345–354 (2006).
2. Garrod AE: *Inborn Errors of Metabolism* (reprinted by Oxford University Press, 1963) Oxford, UK (1909).
3. Bateson W, Saunders ER: Report I. In: *Reports to the Evolution Committee of the Royal Society*. Harrison, London, UK, 1–160 (1902).
4. Garrod AE: The croonian lectures of the royal college of physicians. *Lancet* ii, 1–7, 73–79, 142–148, 214–220 (1908).
5. Pauling L: The molecular basis of genetics. *Am. J. Psychiatry* 113(6), 492–495 (1956).
6. Itano HA, Neel JV: A new inherited abnormality of human hemoglobin. *Proc. Natl Acad. Sci. USA* 36(11), 613–617 (1950).
7. Ingram VM: Gene mutations in human hemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* 180(4581), 326–328 (1957).
8. Tjio TH, Levan A: The chromosome number of man. *Hereditas* 42, 1 (1956).
9. Lejeune J: Mongolism, regressive trisomy. *Ann. Genet.* 2, 1–38 (1960).
10. McKusick VA: *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders*. 12th Edition. The Johns Hopkins University Press, Baltimore, MD, USA (1998).
11. 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.* 32(3), 314–331 (1980).
12. Bishop DT, Cannon LA, Quigley AC, Skolnick MH: Strategies for efficient linkage analysis: example of Huntington's disease pedigrees. *Genet. Epidemiol. (Suppl.1)*, 217–222 (1986).
13. Davis AE 3rd: C1 inhibitor and hereditary angioneurotic edema. *Annu. Rev. Immunol.* 6, 595–628 (1988).
14. Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* 11(3), 241–247 (1995).