THE HUMAN GENOME PROJECT: OPPORTUNITIES, CHALLENGES AND CONSEQUENCES FOR POPULATION SCREENING

John Christodoulou

Western Sydney Genetics Program, The Children's Hospital at Westmead, Sydney, & School of Pediatrics & Child Health, University of Sydney, NSW, Australia

Abstract. It is widely agreed that the Human Genome Project represents one of the most successful international collaborative research efforts of the last millennium. The fruits of the Human Genome Project have paved the way for a revolution in future health care, built on the foundations of an improved understanding of the molecular pathogenesis of human disease, and new and emerging technologies that will allow the rapid throughput of sophisticated gene-based testing. In this review some practical examples of these advances will be outlined, touching on such areas as pharmacogenetics, pharmacogenomics, and microarray technologies. In addition, some of the legal, ethical and social issues that have arisen from the Human Genome Project will be discussed. It is argued that in embracing the new knowledge afforded by the Human Genome Project, we must maintain a cautious approach to help minimise the potential risks in order to reap the potentially enormous benefits to society.

INTRODUCTION

Few would argue that the Human Genome Project has been the most successful international collaborative research effort of the last millennium, that will have far-reaching effects in the way we all practice medicine in the future. The excitement spawned by this venture, and the controversies associated with it, have been exploited by the lay media with varying degrees of veracity, sometimes resulting in a frenzy of hyperbole and wild speculations. It is perhaps timely, therefore, that balanced views be presented outlining the opportunities afforded and potential threats posed by the postgenomic era in the new millennium. In this review an overview of the Human Genome Project will be outlined, highlighting some of the successes to date, describing some of the potential areas of future study, and bringing into focus some of the challenges, both technical and ethical, that lie ahead.

The Human Genome Project had its seeds in the mid 1980's, when the US Department of Energy was directed to study the health effects in humans of exposure to radiation. In 1990, the goals were expanded with a major aim being to describe all genetic material by 2005, but with considerable scepticism and debate as to the value or practicality of such a venture (for an historical perspective see *http://www.ornl.gov/hgmis/project/hgp.html*). Initial efforts focused on generating accurate genetic and physical maps of the human genome, with large-scale sequencing delayed until faster and cheaper technologies could be developed. As a consequence of considerable global intellectual and financial investment by many public research institutions, and competition fuelled by the private sector led by Craig Venter, there was a dramatic escalation in progress. Spectacular claims and counterclaims as to the progress were made by the public research team under the banner of the Human Genome Project and Craig Venter representing the Celera Genomics Corporation. This culminated in a "truce" of sorts in June 2000, when there was a joint announcement that a working draft of the human genome sequence had been completed (Marshall, 2000), and was published in Nature (Lander *et al*, 2001) and Science (Venter *et al*, 2001) respectively in February 2001. It was recently reported that 99% of the human genome sequence is now "finished" (each base pair has been sequenced 8 – 10 times, giving an error rate of less than 1 in 10,000 bp), and it is highly contiguous (the only remaining gaps in the sequence cannot be resolved using current technologies; *http://www.ornl.gov/TechResources/Human_Genome/project/50yr/press4_2003.htm).*

Along the way, the genomes of a number of microbial organisms have been sequenced in their entirety, commencing with *Haemophilus* (Fleischmann *et al*, 1995), the yeast *Saccharomyces cerevisiae* (Goffeau *et al*, 1997), the nematode *Caenorhabditis elegans* (The C Elegans Consortium, 1998), the fruitfly *Drosophila melanogaster*

(Adams *et al*, 2000), and the plant *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000), the human malaria parasite *Plasmodium falciparum* (Gardner *et al*, 2002) to name a few. In addition, there are now major efforts under way tackling the genomes of mouse, rat, pig, chimpanzee and other vertebrate animals, which will help in the development of a clearer idea of what makes us genetically different from our vertebrate "cousins", but will also help us to understand the functions of our genes, paving the way for the very powerful research tool of comparative genomics. It is perhaps somewhat sobering to know that the human genome is only roughly twice as complex, at least in terms of the number of genes, as those of the nematode and the fruitfly (Pennisi, 2001). However, a recent comparison of the Celera and public consortium versions of the human genome sequence revealed only about a 50% overlap, suggesting that the initial estimate of the number of genes was perhaps an underestimate (Hogenesch *et al*, 2001).

Armed with this wealth of genomic data, and new techniques for dissecting it, the approaches used in studying human disease are already undergoing major paradigm shifts (Peltonen and McKusick, 2001). For instance, once a gene's sequence and structure has been determined (structural genomics), attention very rapidly shifts to understanding the function of the gene (functional genomics). Researchers are now moving from studying often rare single-gene disorders, to studying multifactorial disorders, where the subtle interplay of multiple genes, environment and other epigenetic factors may contribute to a particular disease phenotype. We are moving from specific DNA (mutation) diagnosis to using molecular genetic approaches to monitor for susceptibility to disease, and from the painstaking analysis of each gene one by one, to the simultaneous analysis of multiple genes in gene families or metabolic pathways.

A NEW ERA IN GENETIC DIAGNOSIS

A good example of how the Human Genome Project has led to rapid insights into disease etiology is afforded by the Rett syndrome story. This disorder, first described in the 1960s (Rett, 1966), almost exclusively affects girls, rarely affects more than one individual in a family, and is the second most common cause of severe intellectual disability after Down syndrome (Leonard *et al*, 1997). It is a neurodevelopmental disorder, associated with progressive loss of intellectual functioning and fine and gross motor skills. There are no biochemical or histological markers of the disorder, and because of the unusual genetics of Rett syndrome, the identification of the responsible gene defied all efforts until 1999, when a chance finding by a group led by Huda Zoghbi at Baylor College led them to interrogate the public genomic databases annotating the distal end of the X chromosome. Mutations in a gene, methyl-CpG-binding protein 2 (*MECP2*) were identified (Amir *et al*, 1999), and there followed a flurry of mutation screening activity, with now in excess of 150 mutations having been identified, as well as efforts examining possible genotype-phenotype correlations (Amir *et al*, 2000; Amir and Zoghbi, 2000; Hoffbuhr *et al*, 2001; Huppke *et al*, 2000; Weaving *et al*, 2003; Christodoulou *et al*, 2003). Despite these advances, little is understood in terms of the pathophysiology of Rett syndrome, with no prospect for the development of specific therapeutic strategies until there is a much deeper understanding of the biological consequences of mutations in the *MECP2* gene. The development of mouse models (Guy *et al*, 2001) and the use of microarray technology to study the downstream effects of mutations in *MECP2* (Colantuoni *et al*, 2001) will pave the way to the development of rational therapies in this era of modern molecular medicine.

PHARMACOGENETICS: A MOLECULAR BASIS TO INDIVIDUAL RESPONSES TO MEDICAL TREATMENTS

It has been reported that over 100,000 patients die and over 2 million are harmed each year as a result of adverse reactions to their medications (Lazarou *et al*, 1998). Many of these incidents are a consequence of individual variation in drug responsiveness, and much of this individuality has a genetic basis. Pharmacogenetics is the research field devoted to identifying the genetic factors predisposing one to adverse drug reactions, and has immense potential in improving the safety of drug prescription and the efficacy of drug treatments. By identifying genes and functional variations in those genes which alter drug responsiveness (either a lack of therapeutic effect or an exaggerated clinical response), it will be possible to alter prescribing habits to the benefit of patients at the individual level. For example gain-of-function mutations in the liver-specific cytochrome P450 enzyme CYP2D6, which is responsible for the oxidation (clearance) of drugs such as codeine, dextromethorphan and nortryptiline, would make such individuals less responsive to usual dosages, whilst mutations reducing the activity of this enzyme, would make these individuals more susceptible to toxicity at otherwise standard dosages (Wolf *et al*, 2000). Similarly, the systematic screening for polymorphisms in candidate genes which could be associated with disease causation could allow one to determine which drug might best suit a particular patient, as was reported for the drug clozapine in patients with schizophrenia (Arranz *et al*, 2000). One could envisage a future scenario where population-based screening for a

range of pharmacogenetically important genetic alterations could be performed, allowing clinicians to move away from empiric prescribing towards individualised drug therapies, much like the pharmacogenetic advice that we now provide to sufferers of glucose-6-phosphate dehydrogenase deficiency (Luzzatto *et al*, 2001). Similarly, the recognition that genomic variations can influence the susceptibility and progression of infectious diseases, e.g. allelic variants of the chemokine receptor CCR5, altering the susceptibility to HIV (Kaslow and McNicholl, 1999), will permit different approaches to prevention and treatment of infection, such as alterations in lifestyle behaviours.

PHARMACOGENOMICS: THE DEVELOPMENT OF "DESIGNER" DRUGS

As stated earlier, with the Human Genome Project has come the opportunity to understand normal biological processes more deeply than was ever imagined possible. A corollary of this is that the fundamental mechanisms responsible for monogenic and multifactorial diseases will be steadily unravelled, and lead to the direct development of novel pharmaceutical approaches to their prevention and treatment. The ability to dissect the cascade of events unleashing cancers, for instance, has led to the opportunity to design specific treatments aimed at these molecular targets (Gibbs, 2000), and the use of microarray techniques to develop "functional fingerprints" of specific tumours, will permit an individualised approach to therapy more finely tuned than ever before (Alizadeh *et al*, 2000). Similarly, DNA vaccines (DNA that encodes for a peptide being an antigen of interest) may come to the fore in disorders of global magnitude such as malaria and HIV/AIDS (Seder and Gurunathan, 1999), and the complete annotation of the genomes of disease-causing bacteria, will permit the development of novel classes of antimicrobials (Rosamond and Allsop, 2000).

GENETIC ADVANCES, BUT AT WHAT COST TO SOCIETY?

Few would argue against the statement that the discoveries stemming from the Human Genome Project will have unprecedented and far-reaching societal effects. The early recognition of this potential by the US Department of Energy and National Institutes of Health led them to devoting up to 5% of their annual Human Genome Project budgets towards examining the ethical, legal and social issues consequent upon these new molecular discoveries.

Unravelling the complex interactions between genes and environment for common disorders like diabetes mellitus, obesity, hypertension and cardiovascular disease will be a massive long-term challenge, and will need very large collections of community-based clinical and sequencing data, as well as expertise with epidemiological, biostatistical and bioinformatics methodologies. Public interests must be balanced against the important need for individual patient rights and confidentiality. Without governmental legislation to protect the individual's genetic privacy, there is a real risk that there could be unfair use of such genetic information by insurers, employers, courts, law enforcement agencies, adoption agencies or the military, to name a few. Indeed, there is already evidence of genetic discrimination by insurers (Barlow-Stewart and Keays, 2001), raising serious questions as to who should have access to such personal information and how it should be used.

Serious ethical questions relating to whether parents have the right to have their minor children tested for adult-onset diseases, or how genetic information about an individual might affect society's perception of that individual, the definition of normality versus disability, and whether in searching for a cure for a genetic disorder we are demeaning the lives of individuals currently affected by that disorder, all need to be considered. Society is in general in favour of gene therapy for genetic disorders or cancer. But what about using this technology for genetic enhancement to supply a characteristic that a parent might want in a child, such as height or physical strength, but which does not involve the prevention or treatment of a disease? The use of complex genetic technologies like mutational analysis or pre-implantation genetic diagnosis, which are often expensive and not freely available, recently brought into sharp relief by the patenting of the breast cancer genes *BRCA1* and *BRCA2* (Balter, 2001; Dickson, 1996), could potentially lead to the creation of a genetic "under-class" (those who cannot afford to have access to these expensive technologies), potentially widening the apartheid between rich and poor. These, and many other ethical issues, need open and rational debate in public fora, with the facts presented in understandable terms, with all relevant groups having an opportunity to state their case.

In summary, the fruits of the Human Genome Project could yield enormous potential benefits to humanity, but also pose significant risks. As stated by Professor Gordon Duff of the University of Sheffield: "...preventative medicine is an economic necessity, and genomic medicine represents the best route we have to preventative medicine..." (Richards, 2001). However, the courts will face many novel, challenging and often disturbing challenges, with society currently barely keeping pace with scientific progress. It behoves us all to work towards

striking the right balance, by maintaining a cautiously optimistic approach, and in doing so minimising the potential risks to society. We do indeed live in interesting times.

REFERENCES

Adams, MD, Celniker SE, Holt RA, *et al.* The genome sequence of *Drosophila melanogaster*. *Science* 2000; 287: 2185-95.

Alizadeh AA, Eisen MB, Davis RE, *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403: 503-11.

Amir RE, Van den Veyver IB, *et al.* Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes. *Ann Neurol* 2000; 47: 670-9.

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nature Genet* 1999; 23: 185-8.

Amir RE, Zoghbi HY. Rett syndrome: Methyl-CpG-binding protein 2 mutations and phenotype-genotype correlations. *Am J Med Genet* 2000; 97: 147-152.

Arranz MJ, Munro J, Birkett J, *et al.* Pharmacogenetic prediction of clozapine response. *Lancet* 2000; 355: 1615-16.

Balter D. Transatlantic war over *BRCA1* patent. *Science* 2001; 292: 1818.

Barlow-Stewart K. and Keays D. Genetic discrimination in Australia. *J Law Med* 2001; 8: 250-62.

Christodoulou J, Grimm A, Maher T, and Bennetts B. RettBASE: The IRSA *MECP2* variation database-a new mutation database in evolution. *Hum Mut* 2003; 21: 466-72.

Colantuoni C, Jeon OH, Hyder K, *et al.* Gene expression profiling in postmortem Rett syndrome brain: differential gene expression and patient classification. *Neurobiol Dis* 2001; 8: 847-65.

Dickson D. 'Leak' rumours fuel debate on gene patent. *Nature* 1996; 379: 574.

Fleischmann RD, Adams MD, White O, *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269: 496-512.

Gardner MJ, Hall N, Fung E, *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002; 419(6906): 498-511.

Gibbs JB. Mechanism-based target identification and drug discovery in cancer research. *Science* 2000; 287: 1969-73.

Goffeau A, Aert R, Agostini-Carbone ML, *et al.* The yeast genome directory. *Nature* 1997; 387: 1-105.

Guy J, Hendrich B, Holmes M, Martin JE and Bird A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature Genet* 2001; 27: 322-6.

Hoffbuhr K, Devaney JM, LaFleur B, *et al.* MeCP2 mutations in children with and without the phenotype of Rett syndrome. *Neurology* 2001; 56: 1486-95.

Hogenesch JB, Ching KA, Batalov S, *et al.* A comparison of the Celera and Ensembl gene sets reveals little overlap in novel genes. *Cell* 2001; 106.

Huppke P, Laccone F, Krämer N, Engel W, Hanefeld F. Rett syndrome: analysis of MECP2 and clinical characterization of 31 patients. *Hum Mol Genet* 2000; 9: 1369-75.

Kaslow RA, McNicholl JM. Genetic determinants of HIV-1 infection and its manifestations. *Proc Assoc Am Physicians* 1999; 111: 299-307.

Lander ES, Linton LM, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860-921.

Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalised patients: a meta-analysis of prospective studies. *JAMA* 1998; 279: 1200-5.

Leonard H, Bower C, English D. The prevalence and incidence of Rett syndrome in Australia. *Eur Child Adolesc Psych* 1997; 6: S8-10.

Luzzatto L, Mehta A, Vulliamy T. Glucose-6-phosphate dehydrogenase deficiency. In: Scriver CR, Beaudet AL, Sly WS and Valle D, eds. The metabolic and molecular bases of inherited disease. 8th ed. McGraw-Hill, New York, 2001 Vol. 3: 4517-53.

Marshall E. Rival genome sequencers celebrate a milestone together. *Science* 2000; 288: 294-5.

Peltonen L, McKusick VA. Dissecting human disease in the postgenomic era. *Science* 2001; 291: 1224-9.

Pennisi E. The human genome. *Science* 2001; 291: 1177-80.

Rett A. Uber ein eigartiges hirnatrophisches Syndrom bei Hyperammoniamie in Kindesalter. *Wien Med Wochenschr* 1966; 116: 723-38.

Richards T. Three views of genetics: the enthusiast, the visionary and the sceptic. *Br Med J* 2001; 322: 1016-18.

Rosamond J and Allsop A. Harnessing the power of the genome in the search for new antibiotics. *Science* 2000; 287: 1973-6.

Seder RA and Gurunathan S. DNA vaccines - designer vaccines for the 21st century. *N Engl J Med* 1999; 341.

The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000; 408: 796-815.

The C Elegans Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* 1998; 282: 2012-18.

Venter JC, Adams MD, Myers EW, *et al*. The sequence of the human genome. *Science* 2001; 291: 1304-51.

Weaving LS, Williamson SL, Bennetts B, *et al*. Effects of *MECP2* mutation type, location and X-inactivation in modulating Rett syndrome phenotype. *Am J Med Genet* 2003; 118A: 103-14, 2003.

Wolf CR, Smith G and Smith RL. Pharmacogenetics. *Br Med J* 2000; 320: 987-90.