Invited Article

Molecular Biology at the Bedside:
The Impact of Genomics on the Practice of Medicine

Alex E. Felice

Imagine being a newborn baby, discharged home after delivery, with a most unusual gift, a compact disc (CD) carrying all data on "My Genome". The parents would be most anxious to play it on their home PC. Soon, they could discover whether their child has been spared those major single gene disorders which uncle John had before he died young. The doctors had told them they could be selected out when they had opted for pre-gestational diagnosis and selective fertilisation. That problem solved, what about those complex disorders that seem to run in the family; the hyper-cholesterolaemia, obesity and hypertension that persist despite expensive diets, drugs and gym subscriptions, and nanna's premature osteoporosis gene(s). Ah, but look at this other sequence! Now we know why auntie could never take that medication for her arthritis without serious complication. Oops! Those T lymphocyte receptor sequences look funny. Better be careful with vaccinations. These globin genes on chromosome 11 are not quite normal either. Reminds us of that pesky brother in law who speaks of a strange anaemia with a name he could never pronounce well. Better get everyone checked for thalassaemia, and while we are at it for phenylketonuria (PKU) too. These HLA genes seem interesting. Perhaps at last, there is a transplant available for your little brother's leukaemia. All this might appear far-fetched, but one might also say "all this and more" As genomics technologies improve and the costs of DNA sequencing decrease the fictitious scenario could be quite realistic in the near future, perhaps in one to two decades.

Genomics is a relative newcomer to the vocabulary of life science in general and of medicine in particular. Genomics deals with information. Thus far, it has mainly referred to the collection of gene maps and complete nucleotide sequencing of the DNA of various organisms. Modern high throughput sequencing has made the collection of these data possible in the last decade. Powerful computational tools (bio-informatics) have been essential to compile complete structures. Structural genomics is now evolving to functional genomics or molecular physiology and pathology by seeking to annotate the biological and clinical significance of each and every gene or network of gene products in development, homeostasis and disease. A nearly complete draft of the human genome sequence has been available for over a year. The genomes of many other organisms are also in various stages of completion.

The Human Genome consists of 23 long linear polymers made up of any one of the four building blocks, the nucleotides adenine (A), cytosine (C), guanine (G) and thymine (T) in a specific sequence that collectively amounts to around three billion (3×10^9) of them. It is crucial to understand the difference between composition and sequence. The composition of most genomes is roughly around 50% in GC content ([G + C] / [G + C + A + T]). Sequence, i.e. the order with which any consecutive position in the DNA is occupied by either an A / C / G / T, is different and establishes uniqueness of the individual genome.

It is amazing that only 3% of sequence accounts for all of about 30,000 – 40,000 human genes. Initial estimates, mostly based on protein expression profiles, had assumed that there should be three to four times as many. The apparent inconsistency may be explained by weakness in identifying genes, alternate expression of genes, the occurrence of genes within genes and complex modifications of proteins after expression (glycation in particular). With the exception of the erythrocytes, every cell of the human organism contains a nucleus and a mitochondrion with an exact copy of the individual's genome. However, only a selective part of the sequence is actually expressed due to differentiation or developmental control. In the Lymphocytes, genomic re-arrangements among the Immunoglobulin gene super-family (i.e. immunoglobulin, T-Cell Receptor, Cellular Adhesion Molecules and others) account for immunological diversity. Each gene sequence contains all regulatory and coding sequence necessary for transcription into RNA, most of which, albeit not all (i.e. excluding the ribosomal, transfer, small nuclear and riboregulatory RNA), is translated into protein. Nascent proteins are subject to post-translational biochemical modifications that signal traffic by targeting to intra-cellular organelles or the membrane surface, or for extra–cellular secretion into the extra-cellular matrix and plasma. Admittedly, no single molecule, supra-molecular structure, organelle or cell acts on its own. Rather, they participate in huge and intricate network interactions. Our understanding of molecular physiology cannot advance without parallel progress in computational physiology. There are two important discovery platforms. One is to discover genes and their linkage to physiology and pathology or pathogenesis leading to enhanced biomedical science and molecular diagnostics, and the other is to profile expression and uncover valid drug targets for

Alex. E. Felice M.D., Ph.D.
Laboratory of Molecular Genetics,
Department of Physiology and Biochemistry,
University of Malta
Email: alex.felice@um.edu.mt

therapeutics.

The 3% of the human genome sequence that is accounted for as genes discovered so far follows classical triplet nucleotide coding rules. In general, genes may be classified into two large groups; the housekeeping genes which function constitutively and whose products are required for the structure and function of any cell type, and, in contrast those specialised molecules which are developmentally regulated with tissue specificity under stringent controls. For instance, eight different globin genes are expressed from embryonic to foetal and postnatal development. They are organised on two loci with the embryonic zeta and alpha globin genes being on chromosome 16 and the embryonic epsilon, the foetal gamma and the adult delta and beta being on chromosome 11. They are relatively small genes of about 1,500 nucleotides (compared to the Cystic Fibrosis gene and others which are over 100,000 nt long) and expressed only in the erythroblasts and no other cell type. Each coding sequence (exon) is interrupted by two transcribed but un-translated intervening sequences (introns / IVS) with strict rules governing the interface between the two. Regulatory sequences, which determine developmental and tissue specific expression can be found upstream, downstream and distant from the genes. Various types of sequence abnormalities (mutations) account for the production of abnormal haemoglobins, such as sickle cell disease, and regulatory disorders such as beta thalassaemia.

The significance of the residual 97% of the DNA in between the 30,000 genes (inter-genic DNA) is, to say the least, intriguing. It has been called "Junk DNA" because a clear function cannot as yet be assigned to it. Alternatively, it has been called "selfish DNA" because it seems to drift by homologous recombination without natural selection. Some have tried to read rules with which the sequence could be interpreted differentially from the classical triplet codons found in the exons of genes. Others have considered the possibility of a structural role that serves to anchor the DNA to the matrix of the nucleus. The loops of DNA in between Matrix Attachment Regions could result in functional domains necessary for gene control. Much of it is made up of repetitive sequences that differ in complexity from simple dinucleotide repeats (e.g. [AC]n) to variable or hypervariable repeats of a specific sequence (e.g. Alu repeats) One may also find sequences related to viral or even bacterial genomes which may have been acquired by horizontal transmission in the course of historical infection and now become fixed in the human genome. This part of the sequence could also be considered a relic of human and mammalian evolution, although it does not any longer bear resemblance to the parental genomes due to drift without natural selection.

Intergenic DNA is also characterised by a high degree of diversity among individuals. Estimates of Single Nucleotide Polymorphisms (SNPs) range from 1:300 to 1:1,000 of every nucleotide being replaced by another one without any apparent functional defect. Although the functional correlates remain mysterious, blocks of sequence within a certain distance of each other and SNPs are transmitted through generations together more often than the rate which could be predicted by chance alone (haplotypes / linkage disequilibrium) and they follow classical Mendelian rules of transmission. Consequently they are useful as anonymous markers in the determination of human identity as in forensic DNA fingerprinting, paternity testing and in certain instances also in genetics medicine, or gene discovery research.

A small fraction of the genome is found in multiple copies of a small circular molecule in the mitochondria (mtDNA). It codes for components of mitochondrial translation (ribosomes and tRNA) and for enzymes of the oxidative phosphorylation complexes (oxyphos) It acquires mutations faster than the nuclear genome due to exposure to high levels of toxic Reactive Oxygen Species (ROS). ROS are the by-products of oxidative phosphorylation. Tissues with high-energy demand such as heart, skeletal muscle and brain are particularly susceptible to degradation due to random inactivation of mtDNA (heteroplasmy) by ROS. Inherited mutations cause neuro-degenerative disease. mtDNA is transmitted down the maternal lineage.

Structural genomics, as described above, is a tremendously powerful tool with which to understand physiology. Genomics has spawned a number of "omics" domains with which to profile the different levels of gene expression in plasma and each type of cell in the course of development, homeostasis and disease including cancer; ribosomics refers to the profile of RNA molecules, proteomics to that of proteins, glycomics to that of carbohydrate, in particular the patterns of tertiary modification on proteins, and metabolomics refers to profiles of small metabolites. Comparative genomics refers to the re-sequencing of genomes from different individuals seeking mainly to discover new diversity in SNPs and link them to the occurrence of common complex disease due to the inheritance of multiple traits with quantitative effects (obesity, hypertension, diabetes, thrombophilia, osteoporosis, polycystic ovary etc.) It also refers to sequencing genomes of other organisms seeking to relate structural diversity with functional determination of molecular networks. The original Human Genome Project was conceived to include the sequencing of other model organisms, the mouse, the common laboratory microbe E.coli, yeast, the worm C. elegans and the common weed A. thaliana because they have been important experimental models and much was already documented of their physiology. Since then, however, a substantial number of other organisms have been started, partly completed or are under consideration.

Contrary to expectations, the human genome is not among the largest known. Certain plants such as trees can have genomes ten times as large. Smaller genomes belong to most prokaryotes including pathogenic micro-organisms and simpler eukaryotes such as yeast. The genomes of higher apes differ only marginally from humans, perhaps by not more than 2%. The critical difference may lie in a few genes with widespread post-translational effects on many cell surface molecules, which influence intercellular connectivity particularly in the brain and social behaviour. The genomic substrate and genetic background of behaviour, mental health and psychiatric disease is, as yet, a huge black box A combination of advanced neuro-imaging procedures with genomics may be necessary to move this line of research forward.

One may legitimately ask, to what extent is the human condition pre-determined by one's genome? Undoubtedly, the inheritance of mutations in one critical gene inevitably destines

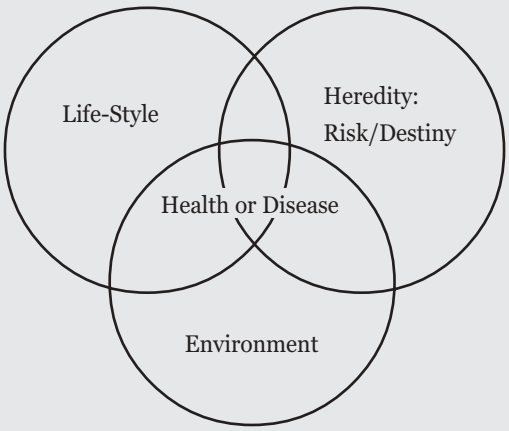
one to suffer a specific disease (e.g. CF VIII and haemophilia, dystrophin and muscular dystrophy) The second possibility is to inherit "risk" or susceptibility to acquire a class of complex disorders which may be infectious, metabolic or oncogenic, through the assembly of a genetic haplotype. The health outcome of any single human genome is likely expressed by interplay with the genomes of other organisms (Fig 1]. A huge number of organisms, which come into intimate contact with man can be found in the environment. Some, plants and animals, provide food and nutrients while others are actual or potential pathogens. Food, processed or not, and its associated bacteria and viruses influences the composition of the microbial flora that establishes itself in the gastro-intestinal tract and provide essential micro-nutrients. The bronchi and skin are constantly exposed to potential allergens. There is considerable genomic heterogeneity in human susceptability to infection, inflammation and allergy. The same degree of genomic variability influences the pathogenicity of micro-organisms. There is concern that, whereas genomes evolved very slowly over millennia, feeding habits and the patterns of exposure to other organisms have changed very rapidly over a few decades. The social and economic changes in the last century have drastically modified diets and exposed humans to new environments through urban sprawl and extensive travel. Many aspects of behaviour and life-style choice increase risk of exposure to new pathogenic genomes such as HIV, Lyme disease and rare viruses (e.g. Ebola, Western Nile) from unusual habitats.

Mammalian and human evolution has been largely shaped by genetic selection due to infectious micro-organisms such as malaria (Haemoglobinopathies), perhaps Leishmania, and diarrhoeal agents (Cystic Fibrosis) Natural selection works on blocks of genomes rather than on single genes resulting in the "co-evolution" of sequences which regulate the production of powerful inflammatory cytokines of body defence mechanisms [Interferons, Interleukins, Colony Stimulating Factors / Haematopoietins). The genome of contemporary humans may be considered to be in "over-gear" to defend itself from infections to which it may no longer be exposed in the contemporary environment. Immune derived "Danger Signals" raise levels of inflammatory molecules with long term degeneration in joints (arthritis), bone (osteoporosis) and blood vessels (atherosclerosis), which afflict the health of the senior citizenry.

Public health and infectious disease medicine will increasingly be able to employ molecular biology tools, such as large micro-arrays, to monitor environmental pathogens (e.g. Legionella in water systems); to diagnose infectious disease specifically and rapidly without needing to resort to time consuming microbial cultures, thus closing the immunological window in patients (e.g. H.I.V.) and animals (e.g. Brucella), permitting prevention by vaccination; and for surveillance of genetically modified organisms. Haplotyping of viruses like HIV and Hep C may guide therapeutics.

In biomedical science, one of the main goals of genomic research is to link sequence diversity with the expression of disease by DNA mapping among large families or selected populations. Although, small isolated populations with good gynecological and health records have been thought most useful for gene discovery, the results have fallen short of expectations, because it seems that linkage (and linkage disequilibrium)

Figure 1: *The genomic pillars of health and disease*



depends on demographic history. In Malta, the rapid growth of the population with input from distant lands in the second half of the millennium contrasts with its small size and disproportionately large inputs from neighbouring countries in the first half. The population structure may be suitable for the discovery of multiples genes, which interact and result in complex disease. Pharmacogenomics is closely tied to this line of discovery research because genome sequence diversity can also be linked with variability in the response to drugs and lead to the discovery of new pharmacological targets.

The pharmaceutical industry (Pharma) was fast in establishing strategic alliances with small genomics based biotechnology companies or academic research groups by investing heavily to expand its product pipeline. New bio-pharmaceuticals such as Human Erythropoietin, Coagulation factors VIII and IX, Interferon and Insulin have already been on the market and in routine use. Others such as Monoclonal Antibodies and Anti-Sense oligonucleotides should appear soon. These products and a vast array of others in various stages of development are proteins or nucleic acids. They are difficult and costly to produce and they have to be administered parentally. Pharma is more comfortable seeking the mass production of small organic chemicals, which can be administered orally. In this case, the recombinant proteins are used as targets for the discovery of small organic pharmaceuticals which are more specific, efficient and with diminished side-effects. It is possible that pharmacogenomics will help to develop different compounds best suitable for groups of patients with the same disease, but individualised on the basis of their genomes. Alternatively, pharmacogenomics will help to produce one compound that is safe, and effective independently of the patients' genomes. Strictly on economic grounds, the latter is a more likely picture of the future of pharmaceuticals and therapeutics.

After genomics, the face of genetics medicine will never be the same, although, it must be said that genetics therapeutics lags markedly behind (molecular) diagnostics. In principle, now, any hereditary disease, which is known to result from any kind of mutation in any gene can be diagnosed in the genomic DNA. Over 15,000 loci and a larger number of associated phenotypes are catalogued in McKusick's Mendelian Inheritance of Man, which is also accessible on line. The small beta-globin gene (1500 nts) is associated with over 600

biochemical and clinical phenotypes. Even a classical single gene disorder such as Cystic Fibrosis is heterogeneous in clinical expression, possibly due to the co-inheritance of "modifier genes". Sickle Cell Disease (Hb S homozygotes) is considered a syndrome due to the inheritance of any one of at least 12 genotypes.

The entire DNA sequence is also available on line and powerful IT tools that efficiently link the two databases are in rapid evolution. Only minute samples of blood or any nucleated cell are needed for the isolation of a proband's DNA since the target test sequences can be amplified by the Polymerase Chain Reaction and sequenced rapidly with automated capillary electrophoresis / whole genome micro-arrays (ASOH: Allele Specific Oligo-Nucleotide Hybridisation). In the near future, whole genome scans will provide detail on the inheritance of any mutation in any clinically relevant gene. Clinical bio-informatics will aid physicians to absorb the significance of this huge amount of information for the benefit of patients and their families. Prediction of disease before the appearance of signs or symptoms and the evaluation of risk will improve.

Genomics adds a new dimension to population testing programs at the level of family planning, maternal ante-natal care and newborn testing. The recruitment of newborns with hereditary disease into ad hoc clinical care will be accelerated. Newborn testing for hemoglobinopathies (Sickle Cell Disease and Thalassaemia) and amino-acidopathies (e.g. PKU) are excellent examples of good practice. The identification of "couples at risk" and who may benefit from counselling and "genomics assisted reproduction procedures" (e.g. pre-gestational diagnosis by polar body biopsy) will be enhanced. Newborn testing by genomic sequencing or scanning is efficient for gene discovery research and contributes to define molecular epidemiology in different populations.

The development of gene therapy has been slow because stem cells are difficult to transfect, vectors are somewhat inefficient and as yet occasionally unsafe, while gene control is complex. There is all anticipation that the technical obstacles will be overcome in a reasonable period of time, and that gene therapy will eventually enter into practice too. In the meantime, recombinant DNA derived replacement biotherapeutics have been produced for certain conditions such as CF VIII / IX in haemophilia A and B. A class of compounds such as Hydroxyurea, Short Chain Fatty Acids (Butyrate) or Azacytidine which act at gene level are being explored with some success to stimulate the replacement of Hb S with Hb F in Sickle Cell Disease with possible applications also in beta-thalassaemia.

The practice of Pathology and Laboratory Medicine will be challenged to adopt the new molecular biology procedures. Tissue arrays and molecular cytology, will replace their morphological roots to generate expression profiles of tissues and cancer cells. It is possible that the uniqueness of every cancer will be documented and associated with individualised therapy.

Clearly, genomics will bring about profound changes in the practice of medicine with no speciality being spared. Most will have to confront new jargon and new procedures, in which the computer and bio-informatics will play increasingly demanding roles. Medical curricula are undergoing reform to replace old knowledge with new knowledge. The clients too ought to be better informed on the impact of the new science on society, the economy and their health. It seems desirable to seek higher levels of science literacy in the population at large.

Another aspect to consider is that genomics reflects the essence of man and cannot be conducted within an ethical or legal vacuum. Genetics medicine, in particular, should be practiced within the context of a code of good conduct, which respects, among others, rules of equity, individual autonomy and privacy. They are not additional burdens but tools for good practice. It is not an exaggeration to state that the practice of medicine beyond 2000 is at a new frontier, exceeding in impact the introduction of antibiotics or chemotherapy and resulting in a paradigm change in which it is practiced. Society and the economy too will be influenced in a variety of ways. Already biotechnology is a large business sector.

One wonders whether there will come a time when prenuptial agreements will be complemented by a comparison of genomes on CD before tying the knot.

Bibliography

Electronic-database Information:

- The Human Genome, special issue, Science, 16 February 2001, Vol. 291, №. 5507 http://www.sciencemag.org/content/vol291/issue5507/ all articles available online
- The Human Genome issue, Nature, 15 February 2001, Vol. 409, №. 6822, pgs 745-964 http://www.nature.com/genomics/human/ all articles available online.
- See also; http://www.nature.com/naturegenetics, a user's Guide to the Human Genome, FREE online. The guide is a hands-on manual for browsing and analyzing publicly available sequence data produced by various systematic sequencing efforts.
- http://www.ncbi.nlm.nih.gov/; The National Centre for Biotechnology Information, provides links to a number of scientific resources.
- http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed Search engine for scientific abstracts, has links to a few full text articles.
- http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM Online Mendelian Inheritance in Man (McKusick) is a catalogue of all known inherited disorders, providing clinical and genetic information as well as references and links. The hard copy is published by The Johns Hopkins University Press.
- http://genome.ucsc.edu/ The Santa Cruz human and mouse genome DNA sequence databases and resources.
- http://www.interscience.wiley.com/jpages/002-3417/ Journal of Pathology website, including free access to special issue on genomics and pathology.
- http://www.ornl.gov/hgmis/project/about.html, gives information on, the human genome project.

Other

Evans GA; The Human Genome Project: applications in the diagnosis and treatment of neurologic disease. Arch Neurol. 1998 Oct; 55(10):1287-90.

Ezzell C; Beyond The Human Genome. Scientific American, July, 52 – 57, 2001

Honore B; Genome- and proteome-based technologies: status and applications in the postgenomic era. Expert Rev Mol Diagn. 2001 Sep;1(3):265-74.

Journal of Pathology; J Pathol 2001; 195: 1-2. DOI: 10.1002/path.922; A Special issue dedicated to Molecular Pathology.

Liefers GJ, Tollenaar RA, Nakamura Y, van de Velde CJ; Genetic cancer syndromes and large-scale gene expression analysis: applications in surgical oncology. Eur J Surg Oncol. 2001 Jun;27(4):343-8.

Marshall T, Williams KM; Proteomics and its impact upon biomedical sciences, Br J Biomed Sci. 2002; 59(1):47-64.

Rindfleisch TC, Brutlag DL, Directions for clinical research and genomic research into the next decade: implications for informatics, J Am Med Inform Assoc. 1998 Sep-Oct;5(5):404-11.

Cauchi MN, (Editor), Bioethical Issues at the Beginning and End of Life. The Bioethics Consultative Committee, Government Press, Malta, 2002.