

## The human genome and molecular medicine – promises and pitfalls

The genome is the full complement of genetic information of an organism. It occurs in the form of DNA and is inherited from our parents. DNA is a series of nucleotides (bases) abbreviated A, C, G and T. Every nucleated cell in the body contains a full complement of genetic information that is organised into 46 chromosomes (XX or XY sex chromosomes and 22 pairs of autosomes).

The working (rough) draft of the human genome was published early in 2001, and the first individual genome was sequenced in 2007. This year (2010) is the 10th anniversary of completion of draft sequence. To date, more than 25 individual genomes have been published while hundreds of completed sequences remain unpublished.

The reference human genome is contained in one copy of each of the 22 autosomes and the 2 sex chromosomes (X and Y). This is referred to as a haploid genome. Since the full complement of genetic information of an individual is contained in 46 chromosomes, this is the equivalent of 2 haploid genomes and is referred to as diploid. If one were to consider this in terms of the number of base pairs (since DNA is double-stranded), a haploid genome contains 3 billion base pairs. This means that the entire complement of genetic information of an individual is contained in 6 billion ( $6 \times 10^9$ ) base pairs. If stored in book form, with each page containing 1 000 letters and each book containing 1 000 pages, more than 6 000 books would be needed to store the information that is present in a diploid genome. And all this is contained in every nucleated cell in our bodies!

Information contained in the genome determines bodily structure and function. Since all structure and function (including memory and emotion) has a molecular basis, the genome is critical to health and disease. Genetic factors do not act in isolation but are influenced by the environment in which they are expressed. This interaction results in what is referred to as phenotype; phenotype is what we can see or measure. Variability between individuals determines their uniqueness. This variability may also result in disease. Sequencing of a full complement of genetic information in an individual assesses the nucleotide sequences of both maternally and paternally derived haploid genomes, which allows one to determine whether variations relative to a reference sequence are homozygous or heterozygous at specific bases.

Chromosomes contain genes and other (non-coding) regions. A gene is a sequence of nucleotides that encodes a specific functional product (including but not limited to proteins). With the sequencing of the human genome, it has become apparent that the genome contains in excess of 20 000 different protein coding genes. These proteins constitute the molecular building blocks of our bodies. Not all 20 000 genes are transcribed in every cell, and the unique combination of those that are (several thousand in varied permutations) provides uniqueness to the different cell types that make up our bodies. Less than 1.5% of the human genome codes for proteins.

It is interesting to note that  $6 \times 10^9$  base pairs of DNA need to be replicated, packed and segregated every time a cell divides. This is done up to  $10^{12}$  times during the development of a human being from a single fertilised cell. It is no wonder that genetic errors are introduced during this process. Even more astonishing is the extremely high degree of fidelity of the replication process, and the fact that the more than 6 billion people on this planet are alive and functioning because of this fidelity!

The information derived from human genome sequences has applications in molecular medicine, DNA forensics, archaeology, anthropology, evolution and human migration.

### Applications in molecular medicine

Potential medical applications of the knowledge obtained from human genome sequences can be considered under the headings of prevention, diagnosis and treatment.

#### Prevention

When genetic information is combined with that on lifestyle (environment), it could be of predictive value for the later development of disease. Once this factor is known, preventive measures can be put into place, including diet, exercise and pharmacological intervention to minimise the risk of developing the disease. However, one of the concerns in this nascent field is the unregulated marketing of genetic tests directly to the general public. In the absence of evidence-based data, these tests should be interpreted and acted upon with a degree of caution.

The field of pharmacogenetics depends on an understanding of the genetic predisposition to adverse drug reactions and treatment non-response. The use of this information to develop patient-specific drug regimens is a major component of the rapidly emerging field of personalised medicine.

#### Diagnosis

Information on the sequences of genes involved in the pathogenesis of a given disease will enhance the ability to diagnose that disease and will positively affect both sensitivity and specificity. The importance of pharmacogenetics in this setting is highlighted by the fact that it may assist in understanding retrospectively the origin of an adverse drug reaction or non-response to treatment.

#### Treatment

Information on a genome-wide level is being used in the field of rational drug design. Finally, when it becomes an accepted form of medical practice and is used to replace the product of a defective gene, gene therapy will also be dependent on information derived from the genome.

### Ethical issues

As with any rapidly emerging field, clearly defined ethical and legal boundaries and guidelines need to be established. Regarding the human genome, some of the areas currently being debated are outlined below.

The first is confidentiality. Who will have access to genomic information? Should employers and insurance companies have access to our genetic information? Could we have our claims refused by a medical aid company that argues that we should have known that we were susceptible to a given condition because the technology is available? In 2008, the United States government introduced the Genetic Information Nondiscrimination Act which attempts to address these issues.

The second is the suggestion that work in this area may be interfering with Mother Nature. Some feel that we may be working against evolution, and that it would be preferable to allow Mother

Nature to take her natural course. However, ignorance breeds fear, hence the tendency to enter into a state of denial and avoid confronting issues of this nature. But can we afford to remain ignorant when we see the emergence around us of designer organisms? Then there is the issue of reproductive cloning, which in humans is universally accepted as being unethical and is banned.

And thirdly is the issue of patents and commercialisation. Who owns the data generated from DNA sequencing? Should the individuals from whom this information was obtained be entitled to royalties from DNA-based products and technologies? This notion of benefit-sharing is particularly important, given the size of the global biotechnology industry emanating from work on the human genome.

## The future

Although the excitement in this rapidly emerging field is tangible, one may well ask 'Who will benefit from knowledge accrued from sequencing information?' If we accept the notion that no single life is better than another, then everyone stands to benefit. National, local, ethnic and gender-specific differences will nonetheless need to be embraced in order to provide our patients with objective and optimal care. However, there should be no distinction between the 'haves' and the 'have-nots', although for practical reasons (including human and financial resources), priority may have to be given to common v. rare diseases.

In parallel with discoveries on the human genome has come the realisation that cellular structure and function is mediated by complex networks of molecules that function both within and outside the cell. If one combines the concept that several thousand genes are transcribed in every cell and that the resulting proteins interact in complex intra- and extracellular networks, one quickly realises that our current understanding of how cells function is only in its infancy. This is a very humbling thought, and forces us to question every element of our present understanding of how the body works. For example, previous studies might have implicated a particular molecule in a given disease process. However, we now discover that there are many more molecules (literally tens or even hundreds) that are involved in the process we are studying. The question then arises as to what importance we should ascribe to each molecule in

the disease process, i.e. where can we rank the molecule in which we are interested in order of its relative importance. From a genetic perspective, this is currently only possible for monogenic/mendelian disorders (such as cystic fibrosis, Duchenne muscular dystrophy or osteogenesis imperfecta) but not for the more common so-called multifactorial disorders such as diabetes, obesity and ischaemic heart disease, which arise from variants in several genes in a given individual and their interaction with the environment.

To conclude this line of thought, I suggest that a great deal of caution must be applied when attempting to implicate a particular molecular mechanism/pathway in a given disease. Information obtained from sequencing the human genome has highlighted the fact that we have relatively little knowledge about the pathogenesis of human disease from a molecular perspective. The 'acid test' of a given hypothesis must remain objective clinical data. This can only be obtained from studies that adhere to the principles of evidence-based medicine in which data are derived from cohorts of patients that are large enough to be statistically and biologically significant and in which the information has been carefully recorded using techniques/instruments that have been adequately validated.

Finally, with the exception of molecular diagnostics for mendelian disorders, the progress on the translation of genome data to the clinic has been relatively slow, although we can expect this trend to change with an increasing number of applications in the near future. The potential of this rapidly expanding field is unquestioned, and we should be ready to embrace the vast amount of new information obtained from sequencing the human genome as it reaches our shores.

### Michael S Pepper

*Department of Immunology  
Faculty of Health Sciences  
University of Pretoria, and  
Department of Genetic Medicine and Development  
University Medical Centre  
Geneva  
Switzerland*

**Corresponding author:** M Pepper ([michael.pepper@up.ac.za](mailto:michael.pepper@up.ac.za))