

The
Economist

A special report on the human genome

Marathon man

Genomics has not yet delivered the drugs, but it will

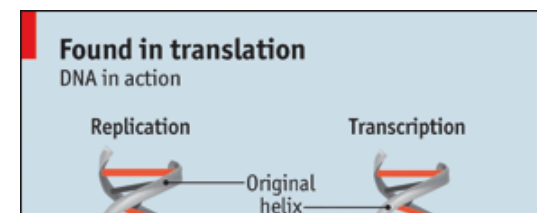
Jun 17th 2010

“WHERE’S the beef?” is always a reasonable question to ask. For the human genome it can be rephrased slightly as “where are the drugs?” It is a question that does not exactly make genomicists squirm, but it puts them on the defensive.

By now, if you had believed the more bullish pronouncements made at the time the human-genome project was coming to fruition, the pipelines of pharmaceutical companies would have been bursting with aspiring treatments for everything from Alzheimer’s disease to Zollinger-Ellison syndrome, as the genes involved in these illnesses were identified and drug molecules that could correct malfunctions of those genes were discovered. In fact, the pipelines are empty; company analysts often seem to regard research as a drain on the balance-sheet, rather than an asset; and drug companies seem to be reinventing themselves as marketing firms for established products. The explanation is a toxic mix of science and economics, but the result is an industry ripe for disruption.

Don’t count your chickens

In 1990, when the human-genome project began, everybody thought they knew what a gene was. It was a stretch of DNA that could be transcribed by an enzyme called polymerase into a chemically similar molecule known as RNA. The RNA acted as a

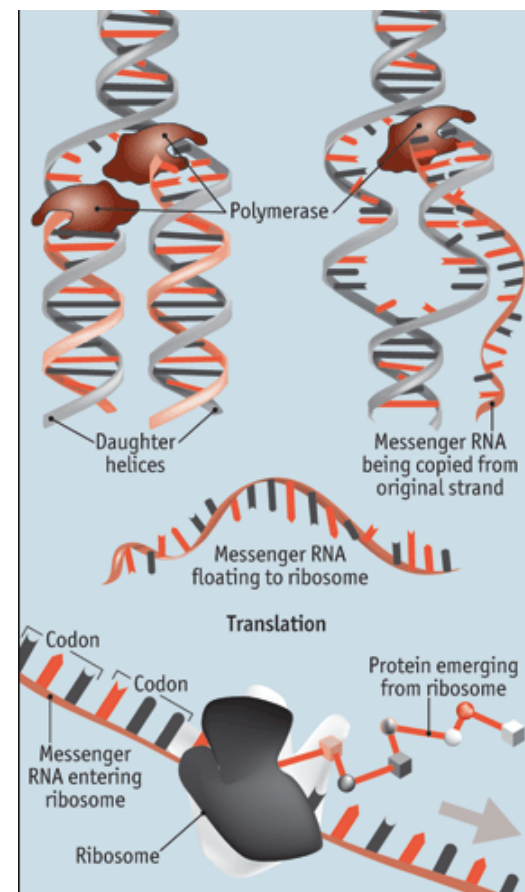


messenger that was itself translated into protein molecules in sub-cellular factories called ribosomes. The translation code was a series of three-letter “words”, called codons, each standing for one of the 20 amino-acid molecules that form the components of proteins. The codons were written in a four-letter alphabet, A, C, G, T, that abbreviated the names of the chemicals of which DNA is made.

It was all very neat. Nobel prizes were awarded in abundance and, except for a few specialised genes whose RNA was directly involved in the protein-manufacturing process, it was understood that the “central dogma” of biology (so described by Francis Crick, co-discoverer of the structure of DNA) was that one gene equals one RNA messenger molecule equals one protein. Proteins are the workhorses of cells, acting as enzymes, ion channels, signalling molecules and structural elements. And some proteins act as transcription factors, regulating the output of the genes themselves. The system made perfect sense. There were a few oddities. Most notably, the best estimate for the amount of DNA that encoded proteins was only 3% of all the DNA in the genome. In the rush of self-congratulation, however, no one paid too much attention to that fact. The non-gene DNA was dismissively labelled “junk”.

Someone should have taken note, though. A sizeable amount of the junk, it turns out, is transcribed into RNA even though it does not make proteins. Instead, the RNA itself is busy doing jobs that were once thought to be the prerogative of proteins: regulating the transcription of other genes, protecting cells from viral attack and even keeping control of bits of DNA that really are junk (or, more accurately, are parasitic on the whole genomic apparatus).

No one knows how many “RNA-only” genes there are, but there could well be more than 100,000 of them. Ten years ago, only a handful were known. By contrast, the current estimate of the



number of protein-coding genes in the human genome is 23,000. The RNA-only genes are, moreover, medically significant. They keep popping up, for example, in cancers.

The second layer of complexity—not completely unexpected, but certainly underappreciated—is called epigenetics. This is the process by which DNA is chemically altered by the addition of a methyl group (a carbon atom and three hydrogens) to genetic letter C. Epigenesis is yet another way of regulating transcription (the methylation stops this happening). It is, however, more permanent than the on/off switching provided by transcription factors and RNA-only genes. Indeed, it is so permanent that it can sometimes be passed down the generations, leading to a lot of excitable talk about the inheritance of acquired characteristics—normally regarded as a Darwinian no-no.

Such talk is premature. More permanent is not the same as indelible, and epigenetic changes are not passed on indefinitely. Nevertheless, they may help explain patterns of disease such as late-onset diabetes. This, some researchers hypothesise, might be encouraged by children inheriting epigenetic patterns appropriate to the diets of their parents but inappropriate to the different, more calorific diets those children are enjoying thanks to the abundance of modern life.

The third layer of complexity is one that is only now starting to be explored. Biologists and laymen alike think of the genome as linear. DNA is, indeed, a long-chain molecule. It is so long, though, that if the 3 billion base pairs were linked together and pulled out straight, the result would be a metre in extent. In reality, DNA is twisted and folded up inside the cell nucleus, with the result that bits of the molecule that seem far apart on a map are actually next to each other in the nucleus.

How much this matters is almost completely obscure. What is known is that there are often active zones of DNA transcription within a nucleus that seem to be much bigger than the width of a strand of DNA and its associated proteins. This suggests that genes apparently a long way from one another are actually, in some sense, collaborating.

All this biological complexity would be bad enough by itself for drugmakers seeking a quiet life. The other problem, though, was a quite monumental naivety about the ease of linking newly discovered genes to diseases and disease processes. This has actually proved fiendishly difficult.

Lighten our darkness

The favoured approach has been the genome-wide association study, or GWAS. Hundreds of these have been carried out over the past five years or so. The idea sounds sensible: gather samples from people with and without particular diseases and look for associations between those diseases and particular genetic mutations, in the form of SNPs. The practice, however, has not really come up with the goods.

The thinking behind GWAS was that it would expose multigenic diseases. These are conditions that seem to run in families but do not obey the clear-cut laws of inheritance laid down in the 19th century by Gregor Mendel. Those diseases that do behave in a Mendelian way—haemophilia and sickle-cell anaemia, for example—are closely tied to the mutational failure of individual genes (a blood-clotting factor and one of the genes for haemoglobin, respectively, for these two diseases). The tendency of people in some families to suffer heart disease, strokes, late-onset diabetes, Alzheimer's disease and so on is, by contrast, less clear-cut. Environmental factors are obviously involved. But mutations are, too—just not single, large-effect mutations like those that cause haemophilia and sickle-cell anaemia. Instead, the pattern of inheritance suggests that many mutations of small individual effect come together to produce a risk rather than a certainty.

GWAS has not been a total failure. It has revealed lots of mutations of small effect. On average, though, these add up to only 10% of the total heritability of any given disease. Mendelian effects add about another 1%. The rest, in a phrase that geneticists have borrowed from physicists, is referred to as "dark matter". These mutations appear to be tremendously important, yet neither Mendelian nor GWAS techniques can detect them. Mendelian mutations are noticed because they are rare and powerful. GWAS mutations are seen because, though puny, they are common. The dark matter lies in the middle: too rare for GWAS but not powerful enough to leave a clear Mendelian signal. Bigger GWAS, with more statistical power, may help a bit, but clearly new methods are needed. One will be to deploy whole-genome sequencing more widely, now that it is becoming so much cheaper. And here the study of one particular sort of disease, cancer, is leading the way.

Compare and contrast

Cancer is at the vanguard of genomic medicine for two reasons. One is that oncologists and their patients (and also the regulators of medical practice) are often willing to take risks that would be unacceptable if the alternative were not a horrible death. The other is that cancer is now known unequivocally to be a genetic disease. Its environmental correlates (smoking, for example) act not by poisoning cells directly but by promoting mutations in those cells' DNA. Such somatic mutations, as those in body cells are known, can cause chaos in an individual's organs, but are not passed to his or her offspring.

In the case of cancer, an accumulation of somatic mutations causes a breakdown of the regulatory mechanisms that stop a cell from multiplying uncontrollably. With the brakes off, the cycle of division, growth and further division continues unabated until the body can no longer support both healthy tissue and tumour.



One lesson that genomics taught oncology early on is that cancers which look similar under the microscope can have completely different genetic causes and thus require different treatments. That general observation should soon be reinforced in detail by a project run by the International

Cancer Genome Consortium (ICGC), a collaboration of researchers in 11 countries. The plan is to take advantage of the falling cost of sequencing to collect full DNA sequences from 500 people suffering from each of 50 types of cancer. Not only will the cancerous tissue be sampled, so will healthy tissue from each patient.

Comparing the healthy and the cancerous tissue in each individual will reveal the somatic mutations which that individual has undergone. Comparing cancerous tissue from different individuals will show which mutations are important. This is necessary because in cancer patients the genes which control the proofreading of new DNA strands often become damaged, so that mutations accumulate much faster. That means crucial mutations are more likely to happen, but also that in any given cancer there is a lot of mutational "noise".

Until now, this has made it more difficult to discover which mutations are important and which merely incidental. The result of the ICGC study should be a near-complete understanding of cancer at the genetic level. That will help diagnosis and treatment (allowing doctors to choose appropriate drugs the first time round, rather than employing trial and error) and, with luck, should promote the development of new treatments.

But identifying the dodgy genes is only the first step to such treatments. Not all gene products are, in the argot, "drugable". And this is where the economics comes in.

Todd Golub of the Broad Institute, in Cambridge, Massachusetts, reckons drug firms have got rather lazy about pursuing leads. For example, many oncogenes, as those whose breakage causes cancer are known, encode proteins called kinases. These are enzymes which are involved in intracellular signalling pathways. A lucky break some years ago revealed a systematic way of attacking kinases with small molecules that block their activation. Researchers with putative anti-kinase drugs are thus welcomed by venture-capital firms. The odds of success are understood and the time to market is tolerable. That is in marked contrast to, say, drugs that might control transcription factors. A failed transcription-factor gene is as common a cause of cancer as a failed kinase gene. Transcription factors, though, are not regarded as drugable. No systematic way of dealing with them has yet been discovered.

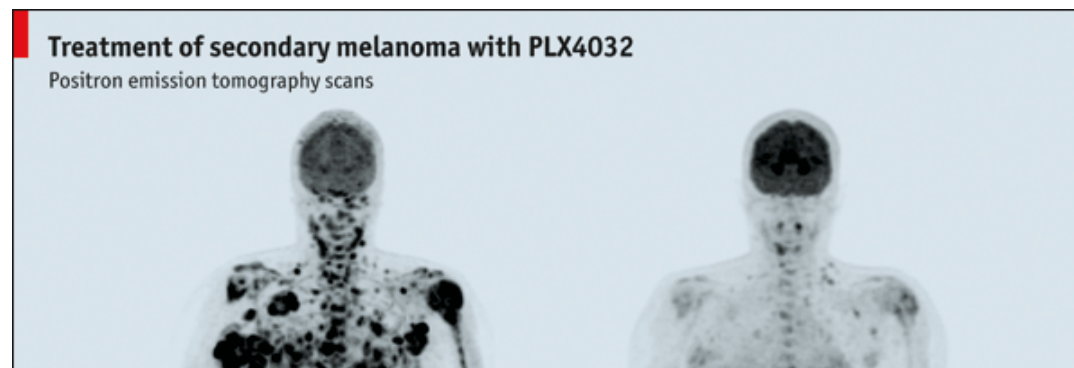
That is not the venture capitalists' fault. Is it the drug companies' fault? They might argue that

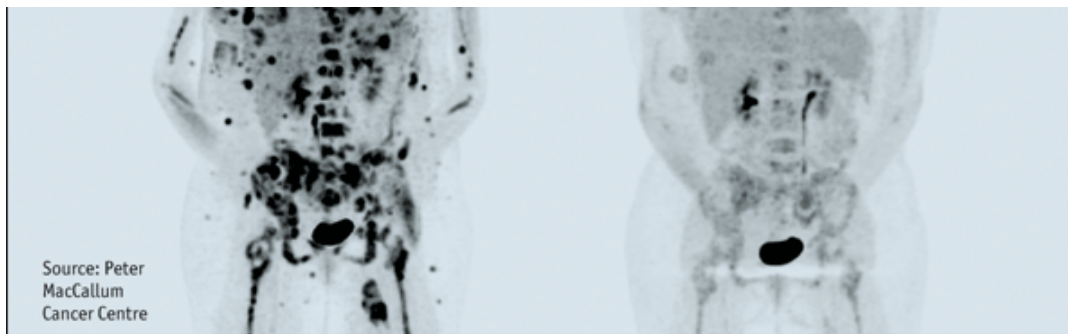
they are not in the business of basic research. On the other hand, a breakthrough in this area would create a whole new line of business. However, if that breakthrough were a conceptual one that could not be protected by patent rather than, say, an individual molecule that could be patented, then other firms would be able to freeride on the discoverer's expensive research.

Fair shares

Dr Golub has a suggestion to break the impasse. Independent laboratories like the Broad could act as honest brokers for general research paid for by a cabal of all the big drug companies. Having paid equally, all would benefit equally. This being basic research, openly published, such collaboration would probably be permitted by antitrust laws. Something similar was tried at the beginning of SNP studies (though admittedly those have not yet led to much in the way of medicine). At the moment the drug firms do not seem interested. Perhaps that will change as their pipelines empty.

Despite such obstacles, genomics has already led to some successes in cancer treatment. The astonishing possibilities can be seen in the two photographs on this page. They are of the same individual before and after treatment with a molecule code-named PLX4032. The shadows are tumours from secondary melanoma, one of the most aggressive cancers known. PLX4032 cleared them almost completely. It was designed specifically to interact with the protein produced by a particular mutated version of a gene called *B-RAF*. This mutation, called V600E, has been found to be involved in 60% of cases of malignant melanoma and, less commonly, in other cancers. PLX4032 inhibits the activity of the mutated protein and causes cells containing it to die.





In this case, the system has worked as it is supposed to. The protein encoded by *B-RAF* is a kinase (and therefore familiar to venture capitalists). The initial development was done by a small biotech firm called Plexxikon, co-founded by Joseph Schlessinger of Yale University, one of the early researchers on *B-RAF*. The molecule has now been picked up by a big drug company, Roche, which is paying for phase III trials, the last stage before a drug is offered to the authorities for approval. If all goes well PLX4032, no doubt sporting a more friendly name, will soon be available for those suffering from melanoma, and will also be undergoing trials in other sorts of tumour in which V600E is implicated

There is a sting in the tail. For the moment the protective effects of PLX4032 last only for six months or so. Presumably, further mutations bypass the V600E—precisely the sort of question that the ICGC project is designed to address. Once those mutations are identified, the hope is that drugs against them can be developed, too. If that proves possible, all of the pathways that lead to cancer could be blocked. That would, in effect, be a cure.

As a demonstration of what genomics can do, PLX4032 is impressive. The question is, can this sort of thing be done with other sorts of disease? One of Dr Schlessinger's colleagues at Yale, Richard Lifton, thinks it can. He points to a number of recently discovered genes that are now the subject of investigation by drug companies. *PCSK9*, which encodes an enzyme involved in cholesterol metabolism, is a target for the prevention of heart disease. People with mutated versions of *ROMK*, the gene for a type of potassium-ion channel, have abnormally low blood pressure so the search is on for a drug that tweaks the unmutated version of the channel, to lower the pressure of people with hypertension. Those with mutated versions of *SCN9A*, which encodes a particular sodium-ion channel, are insensitive to pain. Tinkering with this might produce a

superior analgesic. And *BACE*, the gene for an enzyme called beta secretase, is involved in Alzheimer's disease. Inhibiting its action may delay the progress of that condition.

A little knowledge

This handful of promising candidates, though, shows up the drug companies' real gripe about genomics. It is one thing to find a gene in the genome; it is quite another to find out what it does; and another still to understand whether that knowledge has any medical value. Until these points are dealt with, the drugmaking machine that genomics once promised to become cannot be built.

Thinking on a grander scale is needed. One bold thinker is George Church of the Harvard Medical School. Dr Church's Personal Genome Project (PGP) proposes to collect samples and medical data from 100,000 people and use the newly emerging mass-sequencing techniques to record the entire genomes of each of them. In a way, the PGP will be competing with a number of commercial operations (see [article](#)). The two differences are that it will be free to enter and that all the information will be publicly available.

That open access is a bold idea—and PGP is made bolder by the fact that Dr Church hopes to use new techniques to convert some sample cells into stem cells, from which all of the body's tissue types can be grown. This will enable the project's researchers to do genetic investigation on a tissue-by-tissue basis.

The United Kingdom's Biobank is even more ambitious. It is a government-run project that aims to collect tissue samples and medical data from 500,000 people in Britain. That will allow an analysis of long-term correlations between genes (as well as lifestyles) and health, though in contrast to the PGP the participants' anonymity will be preserved.

"Retrofitting" existing data-collection projects is also yielding results. As an example of what can be achieved, Dr Lifton points to the Framingham Heart Study. This began in 1948 and has followed several generations in the town (near Boston) to try to disentangle the causes of heart disease and strokes. It has revealed a number of genes that may prove classic examples of dark matter.

If present as a single copy inherited from either mother or father, certain mutations of these genes protect against heart disease and strokes. Those mutations are not, however, common enough to be caught by conventional GWAS. On the other hand, individuals who inherit two copies of them, one from each parent, will be stillborn and so they cannot be detected by classic Mendelian counts either.

This sort of result suggests the logical thing to do would be to throw everyone's medical records into the genetic maw. The bigger the sample, the more robust the eventual result. The main objection to that is privacy, and the extent to which this really is an issue may be shown by the success or otherwise of Dr Church's approach. The PGP began recruiting in earnest in March, so all should soon be clear. If the project is a success, the view that a person's DNA is his own business may fade away.

Even when the data from the biobanks and the personal genome projects are in, though, they will still have to be turned into knowledge that can be converted into pharmaceuticals. Turning data into knowledge is the job of people like Aviv Regev of the Broad Institute. It will be a hard slog, involving yet more sequencing (of RNA, rather than DNA, to see which genes are actually active in particular sorts of cell) and lots of computing (to show up the correlations in activity which indicate that particular molecules are parts of the same biological pathway).

The way Dr Regev describes it is very much like electronics. First, the components (the biological equivalents of transistors, diodes and resistors) must be identified. That might now be thought of as "classical" genomics. Then those components need to be assembled into modules (the equivalent of a computer's logic gates). That is where the RNA sequencing (along with a host of other tools) comes in. Lastly, the modules can be linked up as circuits and the whole apparatus of the cell should become clear.

The practical advantage of this knowledge will be that the cell's circuitry can be altered to bypass broken bits rather than fixing the break itself. That opens up a whole new way of thinking about drug development. Add that to the plethora of new targets, in the form not only of the extra protein-coding genes discovered by the original genome project but also of the RNA-only genes and the epigenome, and the long-term opportunities for pharmaceuticals ought to be bright. Who

will take advantage of those opportunities, though, remains to be seen. For if it is true that the rich world's established pharmaceutical companies have become stodgy, the genomic future may lie elsewhere.

Special reports

Copyright © The Economist Newspaper Limited 2010. All rights reserved.