DNA and the genome

A major news story over recent years has been the announcement of the genome sequence for humans. In fact, this project reached a symbolic completion point in April 2003. But this human genome work is just part of a much bigger story -- which includes a list of many completed genomes, for microbes, plants and animals. All this genome work is just the beginning; genome information alone does not solve anything in particular; it is a big resource that will make further biological work easier.

Introduction

Two major news stories of 2003 set the background for this discussion. One is the 50th anniversary of the announcement of the double helical structure of DNA. The other is the announcement of the completed DNA sequence for the human genome. We discussed the development of the DNA structure. A key idea that emerged from this is the complementarity of the two DNA strands. This complementarity immediately suggests how DNA replicates -- by the two strands separating and each serving as a template for a new strand. The resulting "daughter" DNA molecules have one "old" strand and one "new" strand. A physical test of replicated DNA, showing this characteristic, was key in "proving" the basic DNA model. There is much chemical complexity to DNA and much biochemical complexity to how DNA really replicates, but the basic logic of a double stranded structure held together by complementarity still holds.

We then discussed DNA sequencing. We started by looking at some simple DNA sequencing results -- and showed how easy it is to actually read the sequence. Of course, what we looked at is the end step of a lengthy series of steps. We discussed an example of how one might generate the pattern we saw on the sequencing film; our example was not what is actually done, but was a simpler variation to illustrate the logic. The main problem with this basic sequencing procedure is that it works for only about 500 bases. Thus sequencing larger genomes requires some additional work, but it is still based on the same classical procedure that we started with. For large genomes, the process is highly automated, including the
use of lasers to read dye-coded bases. Further, tremendous computer capability is needed to keep track of the data from the millions of pieces of DNA that are individually sequenced.

We discussed the gene count for humans. It is rather low -- and also uncertain. It is uncertain because we actually have considerable difficulty recognizing genes simply from DNA sequences, especially for complex organisms. The low gene count is forcing us to emphasize complexities in gene function, such as splicing and editing, that allow more than one protein to be made from a gene. We then discussed applications of genome information, especially of genome differences between individuals. These include applications such as forensic testing and paternity testing, which were developed some time ago. We discussed some drugs which are chosen based on specific genetic characteristics -- either of the individual, or even of the particular cancer. We then discussed more recent work, using gene chips (microarrays), where analysis of many genes allows leukemia (or leprosy) sub-types to be recognized. The specific figure that I showed was from a recent supplement to The Scientist: New Frontiers in Cancer Research, Sept 22, 2003. One topic that came up during general discussion was prions; I now have a page on prions.

An introduction to DNA: basic structure and how it replicates

The human genome is made of DNA -- as is the genome of almost all organisms. (A few viruses use the closely related chemical, RNA, for their genome; RNA operates by the same basic principles as DNA in this role.) A major milestone in the history of DNA is being celebrated in 2003 (the year this page was started)... It was fifty years ago, April 1953, that Watson & Crick announced that they had determined the structure of DNA -- a structure that in fact "made clear" how it works.
The Fig at the left is a diagram of the general structure of DNA. It shows the famous overall double helix. And it shows the four bases (A, T, G and C) -- which are the "information". At each rung along the DNA ladder is a base pair. Each pair is either A with T or G with C; that is, one strand precisely determines the other strand -- and that indeed is the key to how DNA replicates. See next Fig.

This Fig is from the Glossary of the NIH genome site, [http://www.genome.gov/glossary/index.cfm?](http://www.genome.gov/glossary/index.cfm?). Choose deoxyribonucleic acid (DNA). Also see next Fig.
The Fig at the left is a diagram of DNA replicating. The top of the Fig shows a "parental" DNA molecule; the bottom shows two "daughters". During DNA replication, the two parental strands separate, and each serves as the template for a new strand, which is made by those simple base pairing rules (A-T and G-C), which were mentioned with the Fig above.

In this Fig, the "replication fork" (the site and apparatus for making new DNA) is moving upward.

This Fig is also from the Glossary of the NIH genome site, http://www.genome.gov/glossary/index.cfm? Choose DNA replication.

Good overview of DNA, by David Goodsell. This is a "Molecule of the Month" feature at the Protein Data Bank. http://www.rcsb.org/pdb/101/motm.do?momID=23. For more, see http://ndbserver.rutgers.edu/education/index.html. This is from the educational resources of The Nucleic Acid Database Project at Rutgers.

Commemorations of the 50th anniversary of the Watson-Crick DNA structure

The double helix structure was published by Watson and Crick in 1953 in the journal Nature. 2003 is the 50th anniversary of that landmark, and there are many commemorations. The January 23, 2003, issue of Nature has a big feature on this. It includes an introductory article (Nature 421:310), copies of the original papers on DNA structure, and many articles discussing various aspects of the DNA story. And then there is more in the April 24, 2003, issue. This includes an article (Nature 422:835) by Francis Collins et al on the future of the human genome project. Fig 1 of that article is a fold-out
timeline "Landmarks in Genetics and Genomics"; this is available as a pdf file from the Nature web site. At least some of this material could be usefully read or browsed by those with little background in the field.

* Nature is available online at [http://www.nature.com/index.html](http://www.nature.com/index.html).
* The Nature "web focus" Double helix: 50 years of DNA ... [http://www.nature.com/nature/dna50/index.html](http://www.nature.com/nature/dna50/index.html).

Among other web sites that resulted from the commemoration of the DNA anniversary...

- The Cold Spring Harbor Laboratory (long headed by Watson) proclaimed itself the Official Site of the 50th Anniversary of the DNA Double Helix. They no longer maintain the anniversary celebration page, but they have much about DNA... Go to the Dolan DNA Learning Center: [http://www.dnalc.org/websites/](http://www.dnalc.org/websites/). Sections of interest there include DNA from the Beginning and DNA Interactive -- and more (from "websites" list at the left). Also, choose the "Resources" section (top menu) and find The Biology Animations Library, with DNA methods such as PCR and Southern blotting. DNA from the Beginning is also available in Chinese, Danish, French, Icelandic, Italian, Portuguese (with German, Spanish promised soon); it is also listed for Molecular Biology Resources: Methods. Another section, Inside Cancer, is listed for BITN Resources: Miscellaneous -- Cancer.
- And from our local Exploratorium: [http://www.exploratorium.edu/origins/](http://www.exploratorium.edu/origins/). Choose Cold Spring - DNA.

**The human genome**

The human genome was officially announced in February 2001 by two groups.

- The work of Celera Genomics, headed by Craig Venter, was published in Science for Feb 16, 2001. That entire issue is available online, with free access, at [http://www.sciencemag.org/content/291/5507.toc](http://www.sciencemag.org/content/291/5507.toc).

The main genome articles are probably too technical for most, but the issues contain many news stories dealing with various aspects of the project.

The Human Genome. A genome site from the Burroughs Wellcome Trust, which supported much of the British part of the genome project. [http://genome.wellcome.ac.uk/](http://genome.wellcome.ac.uk/). Includes a range of information at various levels, including for the general public.

Nature: Human Genome Collection. [http://www.nature.com/nature/supplements/collections/humangenome/index.html](http://www.nature.com/nature/supplements/collections/humangenome/index.html). Links to all human genome work from Nature journals. Much consists of the technical articles, but there are also news stories and discussions.

Neandertal genome. February 2009 brings the announcement of a genome sequence from a 38,000 year old Neandertal. It is actually fairly rough at this point, but it is a remarkable achievement to get this far. There is little to conclude for now, except that the genome evidence so far provides no evidence for interbreeding between Neandertals and modern man (Homo sapiens).

• An excellent longer story from Science. It includes basic background information on Neandertals, and on the genome project. Neandertal Genomics: Tales of a Prehistoric Human Genome. Science 323:866, 2/12/09. It is online at: [http://www.sciencemag.org/content/323/5916/866.summary](http://www.sciencemag.org/content/323/5916/866.summary).

Genome results are so important and fascinating that rodents have been seen scrutinizing their genome data. [http://news.bbc.co.uk/2/hi/science/nature/424076.stm](http://news.bbc.co.uk/2/hi/science/nature/424076.stm). (My main purpose in giving this link is for the Figure, for fun. But the work described there is an example of moving a gene from one organism to another, and using that as a tool to learn about the characteristics of an organism.)

Examples of how genome information is useful

As noted earlier, the genome is just data. It is not the magic solution to anything in particular. Because the genome data is fairly new, in fact few practical advances can be directly attributed to it. So, much of what I do here is to show how genome info might be used.

Pharmacogenomics and nutrigenomics. Traditional recommendations about proper nutrition and medicine assume that the population is uniform. Data is collected about population averages and this is used to guide medical treatments and nutritional advice. But we are not all the same. In fact, some examples of genetic differences in how we respond to drugs or nutrients have been found, more or less accidentally, in the past. The availability of complete genome information will allow such knowledge to come more rapidly. Briefly, pharmacogenomics is the customization of drug usage depending on an individual's genetic makeup; nutrigenomics is the analogous customization of nutrition information depending on an individual's genetic makeup.

The following two items are major nutrigenomics sites:

* The Center of Excellence for Nutritional Genomics at UC Davis, supported by the NCMHD (National Center for Minority Health and Health Disparities, part of the NIH): [http://nutrigenomics.ucdavis.edu](http://nutrigenomics.ucdavis.edu).

* The European Nutrigenomics Organisation (NuGO): [http://www.nugo.org/everyone/](http://www.nugo.org/everyone/). In particular, see their page [http://www.nugo.org/nip/](http://www.nugo.org/nip/) for the Nutrigenomics Information Portal, then choose Research. Also, they have an electronic newsletter. You can read it online, or sign up to receive it by email; choose NutriAlerts from the "NuGO sites" menu at the left (of either of those pages).

The two sites above are also listed on my page Further reading: Medical topics, under Web Sites. A specific page of the NuGO site, on Adipose Tissue, is listed for Organic/Biochemistry Internet resources, under Lipids.


Race. Is "race" a useful criterion for guiding medical treatment? The important point for us here is that genomics is offering new insight into this socially-charged question. At this point, genetic analysis suggests that there are some genes that reflect "geographical origin", but that the variability of human genomes within any "race" is far more than the genetic differences between "races". Of course, this information will be of more practical use as details emerge.

The following New York Times article discusses a clinical trial of a drug that is being targeted to and tested with only one racial group -- with the approval of the FDA. U.S. to Review Heart Drug Intended for One Race, June 2005. http://www.nytimes.com/2005/06/13/business/13cardio.html.

The following two short essays are by scientists discussing the race issue:


Personalized medicine. There are now companies that will take your DNA (and some money) and report back to you your risk for certain diseases. A good idea in principle, but how good is it in practice. Genome pioneer Craig Venter and colleagues have evaluated a couple of these companies, and offer some suggestions. As a general perspective, they think the companies are doing high quality work, technically, but the quality and usefulness of the information is questionable. It is true that your DNA contains information about disease susceptibility, but current knowledge of that is limited -- more limited than the companies want to admit. The paper is: P C Ng et al, An agenda for personalized medicine. Nature 461:724, 10/8/09. The paper seems to be freely available via the web site of the Venter Institute. Go to their page of press releases: http://www.jcvi.org/cms/press/press-releases/. Scroll down to the item for October 7, 2009. Click on its link; it takes you directly to the article at Nature. This probably means that the article is freely available directly from Nature.

Added May 7, 2011. There are many Musings posts in the broad area of personalized medicine. One of the first was: Personalized medicine: Getting your genes checked (10/27/09). It links to several others in the area.

Miscellaneous (books, web sites, comments)

An Introduction to Genomics: The Human Genome and Beyond, and related educational materials on the how and why of sequencing. From the Joint Genome Institute, a US DOE lab in Walnut Creek, CA. http://www.jgi.doe.gov/education/index.html.


story, most famously as co-discoverer of the DNA double helical structure and as the first head of the US Human Genome Project. Here he discusses the history and future of the human genome project. He is a fine writer -- clear, and provocative enough to be fun. This book is for the general public. The science in it is good, and well-explained, with helpful artwork. The history is broadly good. And it is Watson's style to tell you what he thinks about controversial issues; agree or disagree, he makes for lively reading. For two -- very different -- reviews: Lindee, Science 300:432, 4/18/03; Singer, Nature 422:809, 4/24/03. Lindee concludes that "[Watson's] latest promotional brochure is not worth anyone's time." Singer says that the public and even scientists "can learn a great deal from the book, and enjoy doing so." I recommend it -- without endorsing all of his opinions.

Online video. A conversation with Jim Watson. Go to the Caltech theater listings for Science and Technology: http://today.caltech.edu/theater/list?subset=science&story%5fcount=end. Scroll down the list to this item, dated May 5, 2003. The conversation is with David Baltimore, (then) president of Caltech and himself a Nobel prize winner (for his discovery of the enzyme reverse transcriptase, the enzyme that copies RNA into DNA).

Book. B Maddox, Rosalind Franklin - The dark lady of DNA. Harper/Collins, 2002. One of the dark parts of the DNA story is the lack of recognition of the role of Rosalind Franklin, who made the very fine X-ray pictures that Watson & Crick used as part of developing the double-helix structure. This lack of recognition was magnified by Watson's poor treatment of Franklin, especially in his earlier book, The Double Helix. Brenda Maddox's new biography has received wide praise as being fair and accurate; she had access to many materials that were previously unavailable. This is a biography, not a science book -- though you will certainly get a good sense of how the DNA story was developed. Highly recommended, but don't expect to come away declaring winners and losers; it's not that simple, but it is a good story, and it certainly enhances our understanding of an important scientist. (One part of the controversy, to some, is why Franklin did not share in the Nobel prize for the DNA work. It is a sufficient answer to that question that she died a few years before the DNA Nobel, 1962; posthumous Nobels are not allowed. Note that this point does not address the merits of her contributions, but does address one question which often comes to the forefront.)

There is a short essay about Franklin, in the general spirit of the book, online in the Mill Hill collection: K Rittinger & A Pastore, Rosalind Franklin - The dark lady of DNA... http://www.nimr.mrc.ac.uk/mill-hill-essays/rosalind-franklin-the-dark-lady-of-dna. For more about the Mill Hill essays, see the note on the BITN main page, under Web sites.

Recent items, briefly noted

Coumadin (warfarin) is a widely prescribed medication to reduce blood clotting. The dosage must be carefully controlled, and people vary in how they respond. The FDA has announced a new labeling of coumadin that encourages testing the patient for two known genetic factors that affect the metabolism of the drug. A brief version of the announcement is at http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm152972.htm.

A small trial has been reported showing that such testing is beneficial. So far, all we have is a news story summarizing the key findings. Gene test cuts complications from blood thinner warfarin (3/16/10). http://www.usatoday.com/news/health/2010-03-16-warfarin-gene_N.htm.
Sequencing technology -- and cost. The human genome project cost about $3 billion. Much technology was developed along the way; as the project wrapped up, it was estimated that one could sequence a person's genome for a few million dollars. There is a dream -- and goal -- of sequencing an individual's genome for a thousand dollars. That may still be a way off, but the cost of sequencing has been declining, in large part due to fundamentally new approaches to sequencing. 2009 brings a report of a complete human genome for $50,000. A news story on this: Cost of Decoding a Genome Is Lowered. A Stanford engineer has invented a new technology for decoding DNA and used it to decode his own genome for less than $50,000. August 10, 2009. http://www.nytimes.com/2009/08/11/science/11gene.html.

Using genetic information to assess risk and guide screening. Most genes that affect disease susceptibility have only a small effect. How do we use such information? A paper in the New England Journal of Medicine lays out a model. Although there is probably much to quibble with, the model is clear enough, and may be a useful reference point for discussion. They start with the current UK recommendation that women be screened for breast cancer starting at age 50. Accepting this as the starting point, they note that this is the point at which a woman has a 2.3% chance of breast cancer within the next 10 years. They then argue that by a simple test for some known genetic variants, they can mark some women for screening at age 40 -- because with their genetic makeup that is the age at which they now have a 2.3% risk of breast cancer within 10 years. Similarly, women with other genetic variants have lower risk, and their screening can be delayed. The result is the same use of resources, but more effectively deployed. A news story about this work: Cancer gene test 'for all women', June 26, 2008. Online: http://news.bbc.co.uk/2/hi/health/7475312.stm. The paper is P D P Pharoah et al, Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer. N Engl J Med 358:2796, 6/26/08. Free online: http://www.nejm.org/doi/full/10.1056/NEJMs0708739.

Tradeoff. We sometimes dream of finding "the gene" that causes a particular disease -- so we can counteract that gene. But among the complications... It may be that the same gene is good in one way and bad in another. Recent work suggests such a tradeoff may occur between diabetes and prostate cancer. In fact, two genes with this tradeoff have been found. News story: Genetic variants may be 'trading' one illness with another using new genes, Oxford research shows. Online: http://www.timesonline.co.uk/tol/news/science/article3649020.ece.

Genome ethics. Genome work is raising a new set of ethics questions -- especially since there is so much uncertainty what the genome information means at this point. A group of bioethicists has proposed a set of guidelines for doing genome research, published as: T Caulfield et al, Research Ethics Recommendations for Whole-Genome Research: Consensus Statement. PLOS Biology 6, e73, 3/08. The paper is free online: http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0060073.

Ancestry. An interesting subject is tracing human lineages by genetic tests. This is indeed a proper area of study, and has yielded insights into human migrations. It has also entered the popular arena. There are commercial tests that claim to reveal your ancestry. Unfortunately, the quality of this testing is questionable at this point. A "Policy Forum" article about this appeared in Science, and a news story about the work and that article appeared in the UC Berkeley news. The Science article: D A Bolnick et al, Genetics: The science and business of genetic ancestry testing. Science 318:399, 10/19/07. The UC Berkeley news story, featuring co-author Kimberly TallBear: Researchers caution against genetic ancestry testing; October 18, 2007. http://www.berkeley.edu/news/media/releases/2007/10/18_genetictesting.shtml.

Craig Venter is one of the pioneers of genome work. He is also the first person to have his entire DNA -- the diploid chromosome set -- completely sequenced and reported. Importance? Well, for now it is a technical milestone and something of a curiosity. However, as more complete genomes become available -- and as the cost comes down -- the
usefulness will increase. For example, they note how he has specific alleles that both favor and disfavor heart disease. At this point, that is too little info to be useful. At some point, with more information, it will be useful. I doubt that many will want to read this in detail, but simply browsing the Introduction and Discussion sections will give the flavor. And it is a historic paper. The paper -- by Venter, about Venter, and from the Venter Institute -- is: S Levy et al, The Diploid Genome Sequence of an Individual Human. PLoS Biol 5(10): e254. 9/4/07. It is open access at [http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0050254](http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0050254).

M May, Pharmacogenetics lurches forward. The Scientist 8/2/04, p 26. This article discusses several specific examples of how drugs may affect individuals differently, depending on their genetics. It includes the recent genetic analysis of why Iressa works for some patients and not others.

**Contributors**

- Robert Bruner ([http://bbruner.org](http://bbruner.org))

This page viewed 6179 times

The BioWiki has 6743 Modules.