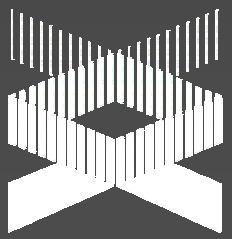
Implications of the Human Genome Project for Medical Research

Christopher Austin, M.D.
National Human Genome Research Institute
National Institutes of Health

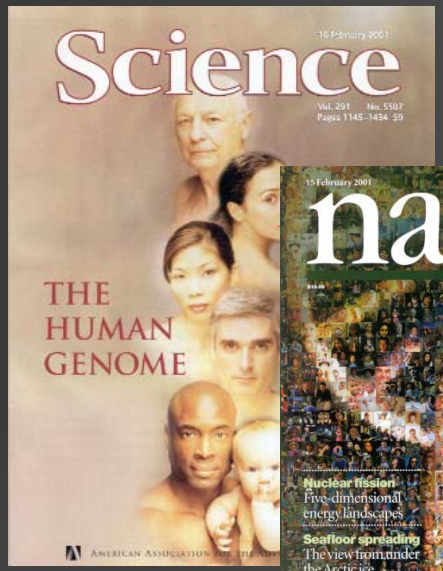


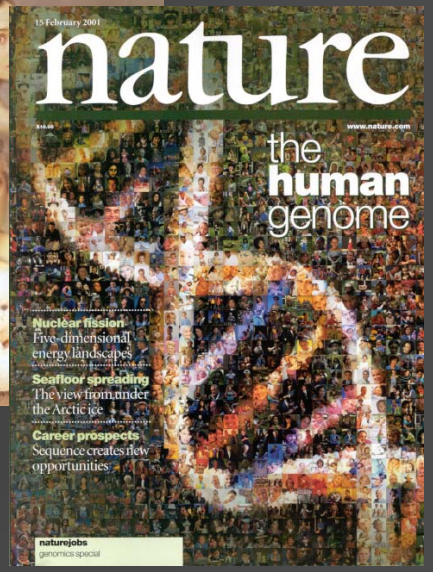
The Human Genome will be completed in April 2003

- All clonable euchromatin (>95% of the total genome) with error rate < 1/10,000 bp
- Sequencing will cease as of this time and all “draft” sequence will have been converted to “finished” sequence
- Sequencers will move on to finish mouse, rat, honeybee, chicken, and chimpanzee
- Next organisms to have their genomes sequenced will be cow, dog, sea urchin, and several fungi

Wasn't the human genome completed before?

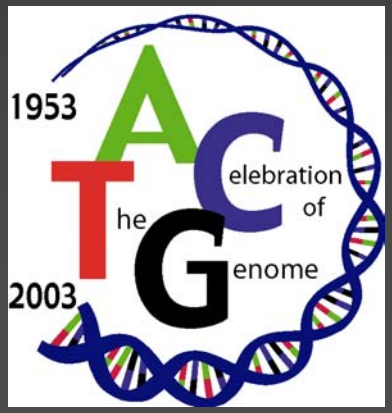
June 26, 2000: First Draft



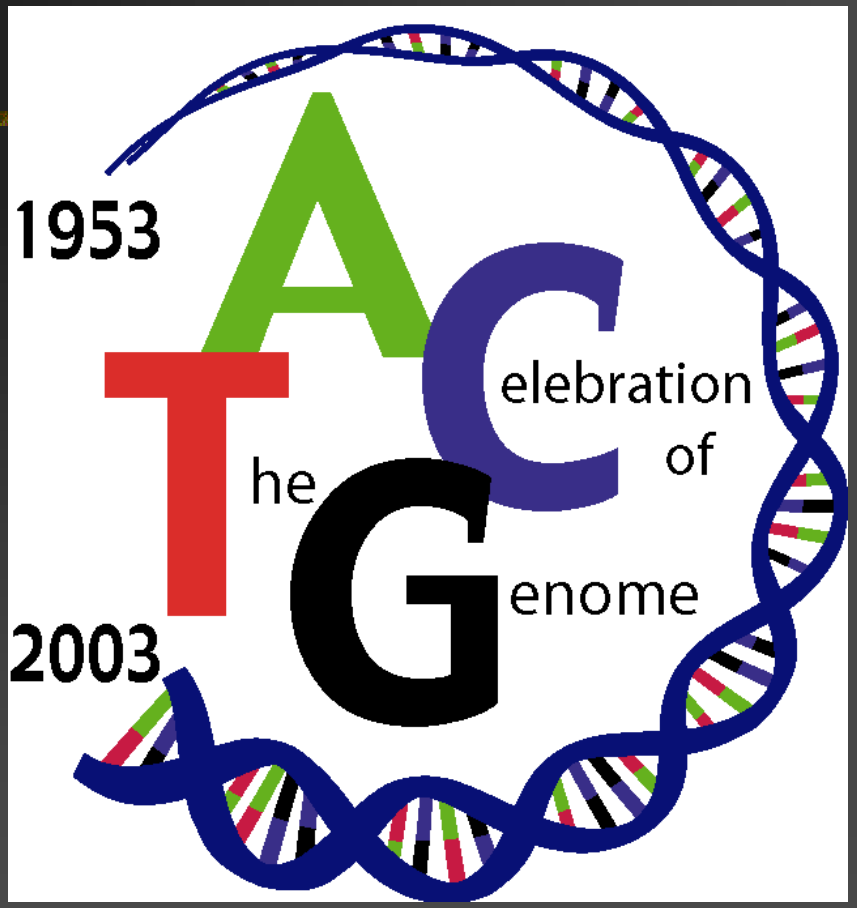


February 15, 2001: Working Draft

April 2003: Full finished sequence



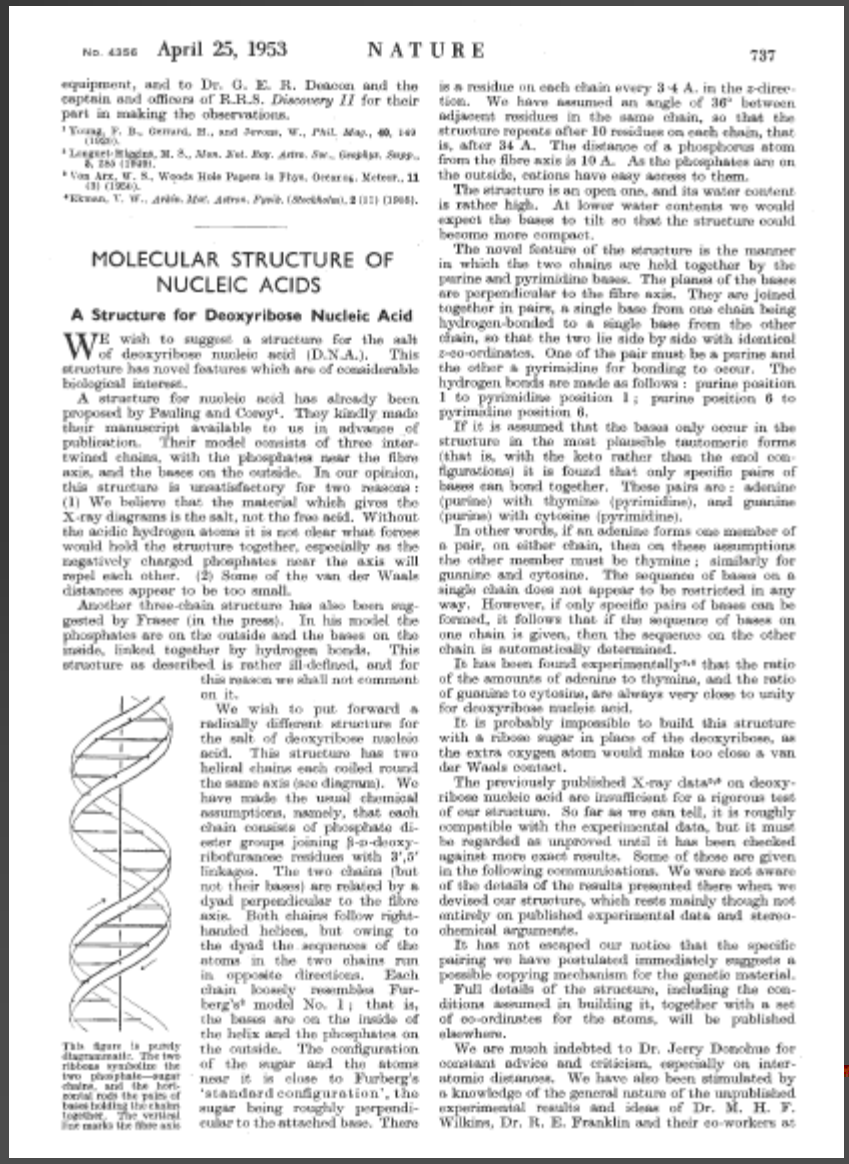
April 2003



50 Years of DNA: From Double Helix to Health

Big Events in April 2003

- 50th Anniversary of discovery of DNA structure by Watson and Crick
- Completion of the sequence of all the human chromosomes
- Announcement of bold new research plan for genomics



No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1928).

[2] Leegret-Higgins, H. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

[3] Van Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).

[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three interwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[5,6] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

Genome Celebration Public Events

April 14-15

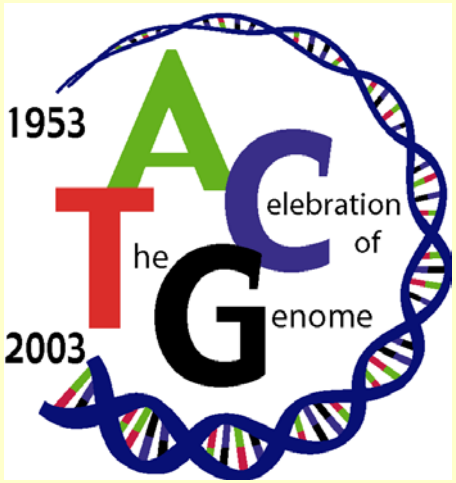
From Double Helix to Human Sequence - and Beyond

Scientific Symposium at The National Institutes of Health

April 15

Bringing the Genome to You

Public Symposium at The National Museum of Natural History



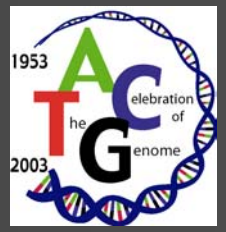
www.genome.gov/About/April2003

April 25

National DNA Day

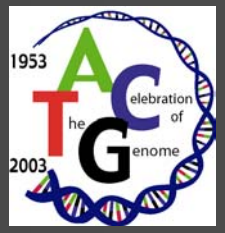
A teachable moment for educators & students across the nation

Public Symposium – April 15



Opening Remarks	James Watson Francis Crick (recorded)
The Human Genome Project	Eric Lander
HGP to Medicine	Wylie Burke
Media's View of the Genome	Robert Krulwich

Public Symposium – April 15



- The Human Genome Project to Society

  Moderator: Robert Krulwich

  - Genetic Policy — Members of Congress
  - Ethics — Tom Murray
  - A Consumer's View — Kay Jamison
  - Health Disparities — Harold Freeman
  - Disabilities — Paul Miller
  - Historical issues — Vanessa Gamble
- The Human Genome Project and the Future

  - Francis Collins

What now for the Human Genome?

- “A Vision for Genome Research” to be published April 2003
- Genome to Biology
  - structural and functional components,networks and pathways
  - heritable variation
- Genome to Health
  - genetic contributions to disease and drug response
  - genome-based diagnostic approaches
  - new therapeutic approaches to disease
- Genome to Society
  - how genetic risk information is conveyed and used in clinical settings
  - genetic discrimination, privacy – HIPAA
  - ethical boundaries

Genetics vs Genomics

- Critical and often misunderstood difference between single gene and multiple gene diseases
  - Single gene: mutation *causes* disease (100%)
    - e.g., Huntington's disease, cystic fibrosis, thalassemias
    - Are of great importance to individuals and families with them
    - But, even when added together, are relatively rare
    - Most people not directly affected
    - Thus, genetics played small role in health care (and in society)

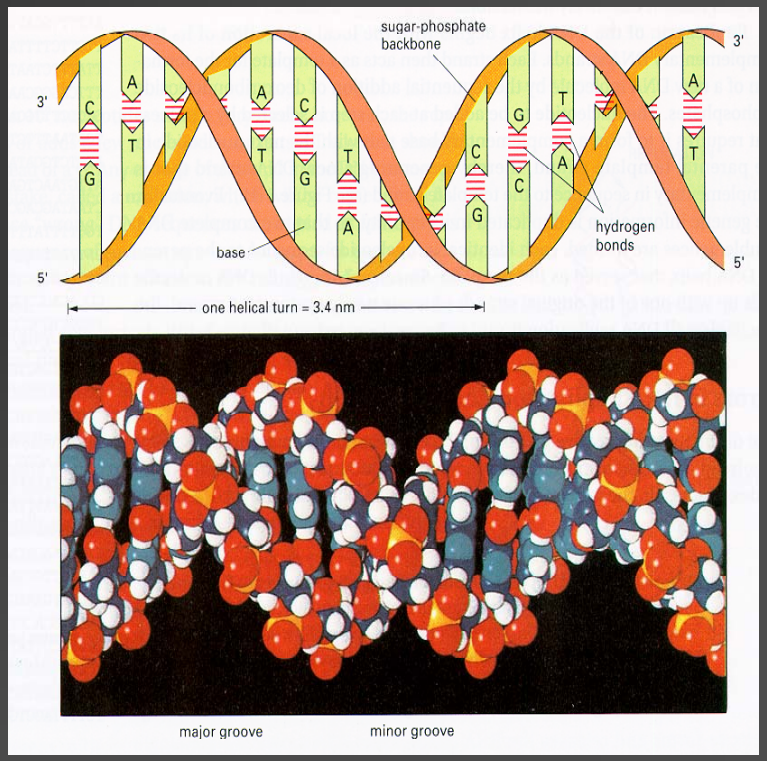
Genetics vs Genomics

- Multiple genes: mutation *predisposes to* disease (5-50%)
  - a.k.a., ‘polygenic’, ‘common’, ‘complex’, ‘genomic’ diseases
  - e.g., heart disease, hypertension, diabetes, obesity, cancer, Alzheimer’s disease, schizophrenia
    - ApoE (Alzheimer’s disease)
    - BRCA1 & 2 (breast & ovarian Ca)
    - CCR5 (HIV/AIDS resistance)
- Most common diseases have heritable (genetic) component
- Other part of disease susceptibility is environmental (e.g., diet, exercise, smoking)
- Most people directly affected
- Thus, genomics will play a large role in health care (and in society)

The Human Genome

- 3 billion nucleotide base pairs
  - Adenine (A)
  - Cytosine (C)
  - Guanine (G)
  - Thymine (T)

  on a sugar-phosphate backbone
- 99.9% identical in all humans
  - 1/1000 bp variant between individuals (3 million total)
  - 1/300 bp variant among population (10 million total)
- A *single* variant can cause disease

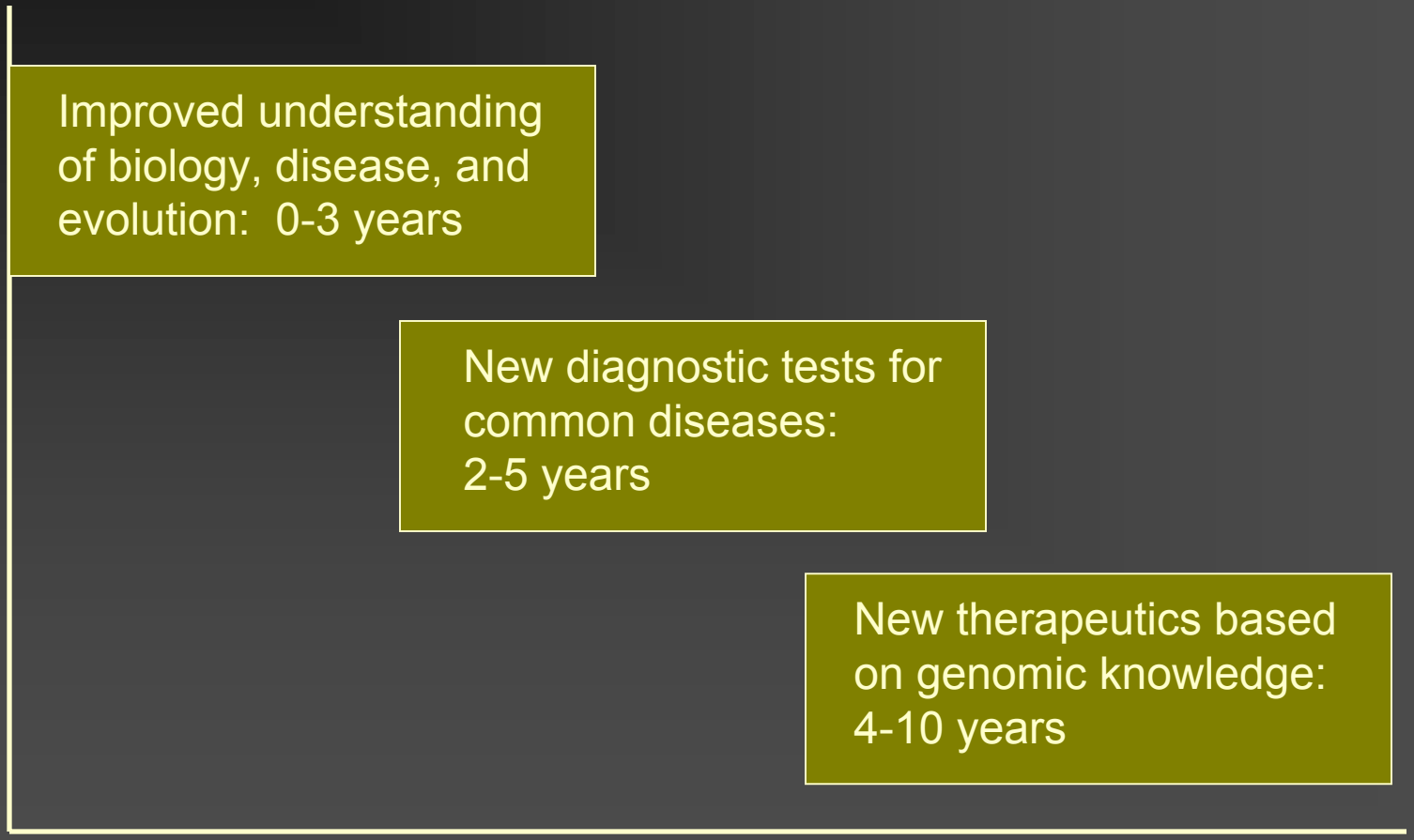


Great (Genomic) Expectations

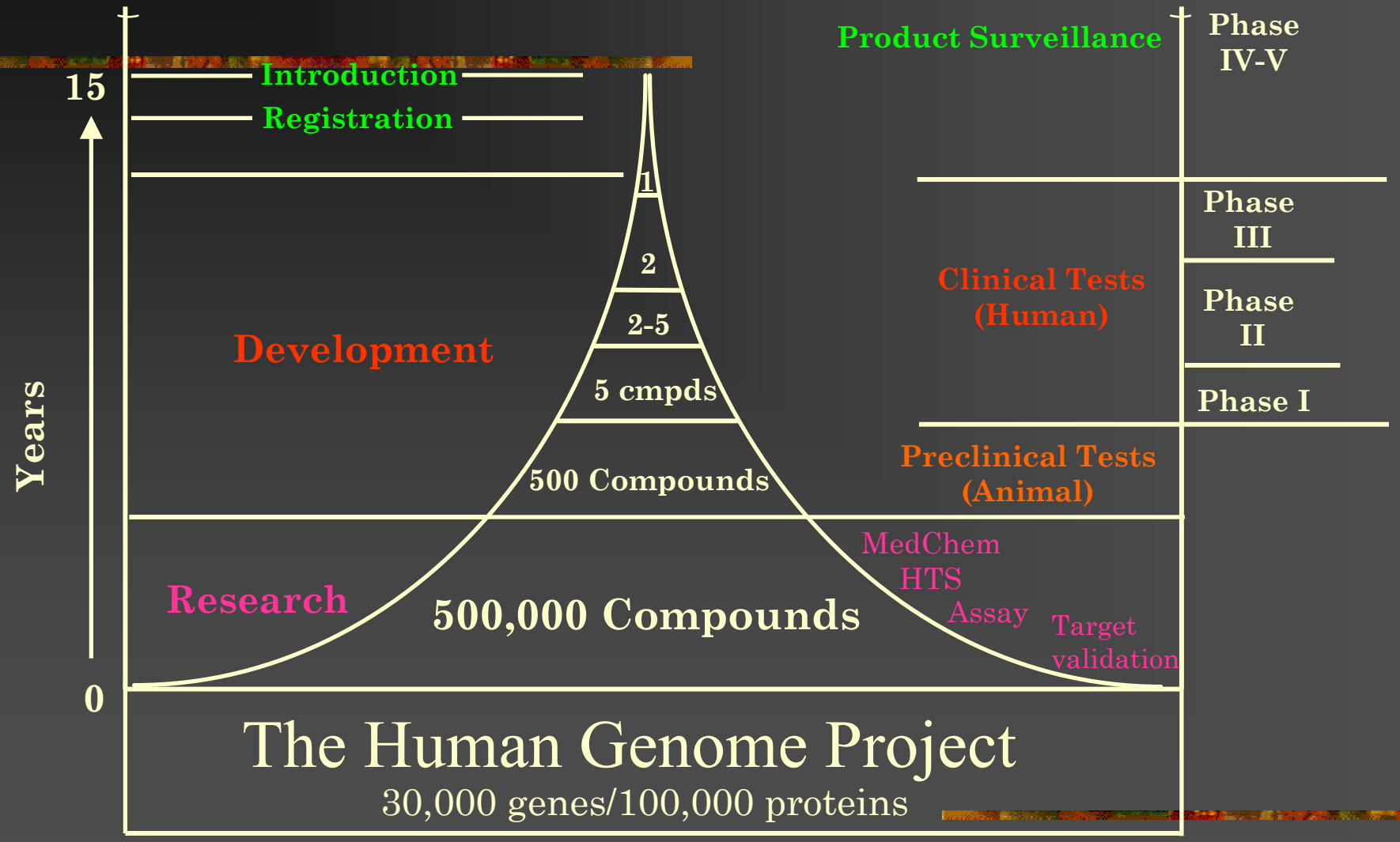
- Genomics holds great promise for improving human health, but short term expectations are outsized
- "Genomics (will) lead to short-term increases in R+D spending and little increase in productivity…the industry could go bankrupt trying to innovate"

  - McKinsey and Co. report "The Fruits of Genomics", 2001
- Issue is mismatch between data and information

Where and when can impact on medicine from the HGP be expected to begin?



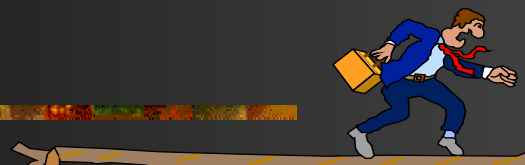
Development of a novel drug



DNA Sequences vs. Drug Targets

- Total number of human genes ~30,000
- Total number of human proteins ~100,000 (?)
- Current drug targets: ~500
- Gene identification is only the start to determining function and any therapeutic potential
- Total number of targets estimated at 10% of total, or ~3,000 $\Rightarrow$ 90% of potential remains
- “Validation”
  - Definition of sequence function, role in disease
  - Demonstration of manipulability of gene product
  - Transforms gene product into drug target

Turning a Gene into a Drug Target



DRUG TARGET

DRUG TARGET

DISEASE ASSOCIATION

PHYSIOLOGY

BIOCHEMICAL FUNCTION

Multigenic: Schizophrenia
Diabetes
Cancer
Single-Gene: Macular Degeneration
Cystic Fibrosis
Cancer

Cancer
Hypertension
Alzheimer's Disease

"Druggable" Gene Products
GPCRs
Ion Channels
Proteases
Nuclear Receptors
Kinases/Phosphatases
Secreted Proteins

Genomic Medicine

- Molecular, rather than historical/clinical, taxonomy of disease
- Individual prospective risk assessment will allow:
  - Individualized screening, e.g., mammography schedule, colonoscopy, prostate specific antigen
  - Presymptomatic medical therapies, e.g., antihypertensive agents before hypertension develops, anti-colon cancer agents before cancer occurs

Drug development in the genome era

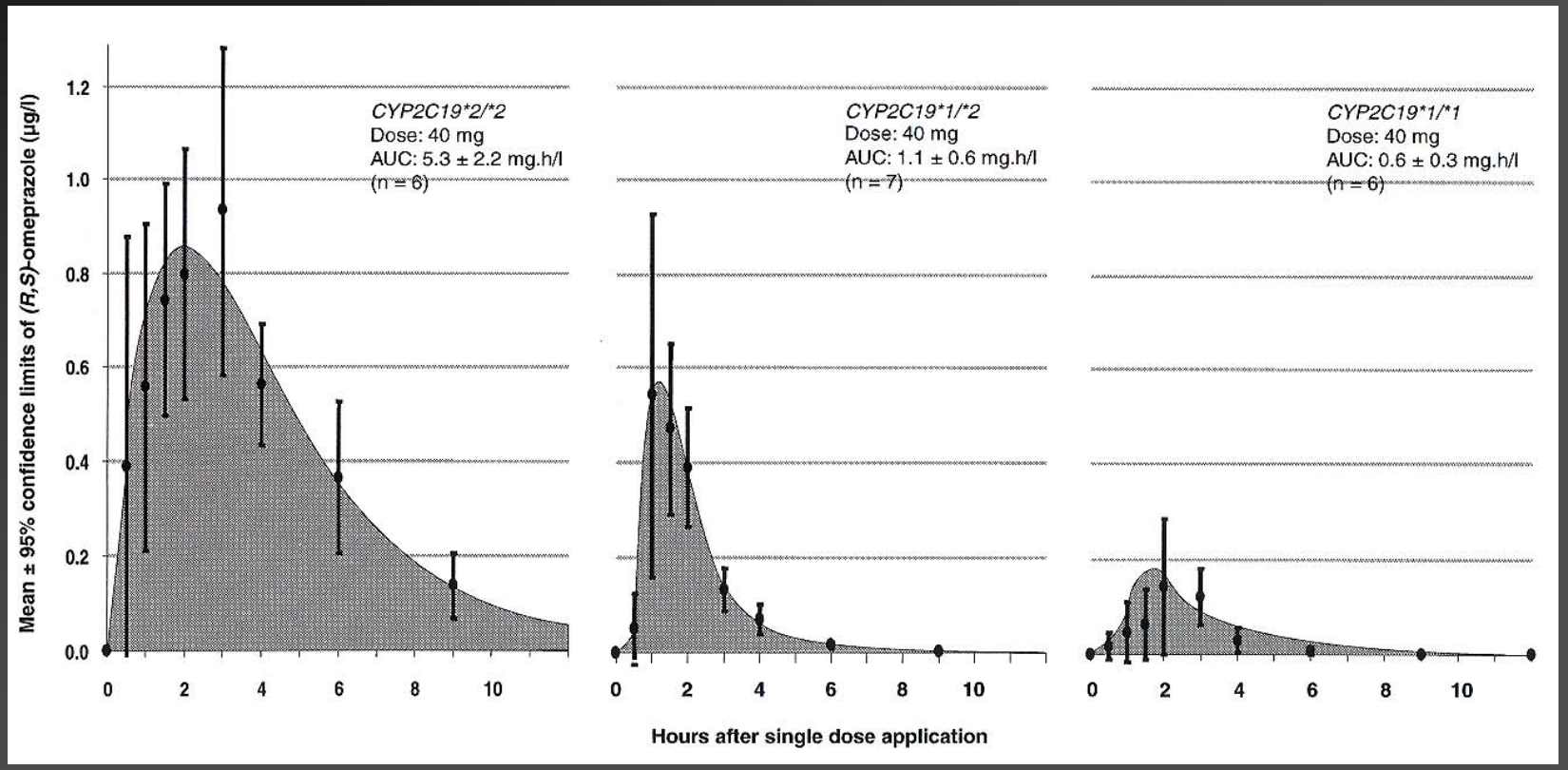
- "Parts list" of human development and function will allow
  - More intelligent chosing of targets for therapeutic development
  - Choosing among all possibilities rather than taking what's available
  - Comprehensive definition of gene interactions and pathways, critical to understanding common polygenic diseases
- Magnitude of task of functionating the genome will require
  - Shift in tasks undertaken by public vs private sectors, with more target evaluation being done in public sector
  - Better community-wide understanding of the value of early research findings
  - Resolution of IP issues surrounding gene and other research tool patents

Applications of genetic variation to drug development

- Target Identification/Prioritization
  - Association of SNPs in potential targets with disease
    - β2 adrenergic receptor – Asthma, Heart failure
    - Angiotensin II receptor - Hypertension
    - PPARγ - Diabetes
    - ACE - Peripheral/Carotid artery disease, LVH
- Target Biology
  - characterization of variability in novel targets
    - predict variability in clinical response/safety
- Screening
  - determination of correct/most prevalent allele for HTS

Genetic variation influencing drug metabolism

Improved DMPK studies, dose finding



Pharmacogenomics 2000; 1:131

- CYP2C19 SNPs affect Prilosec levels: AUCs vary 10-fold with genotype
- CYP2C9 SNPs predict warfarin and phenytoin levels

Applications of genetic variation to clinical research

- Drug Metabolism/Clinical Pharmacology
- Clinical trials
  - Improved uniformity of subjects by characterizing genetic markers $\rightarrow$ increased power
  - Post-hoc analysis of non-responders, subjects with adverse events
  - Fragmenting of markets is holding back utilization
  - Examples now in medical
    - Herceptin for breast cancer (somatic mutation)
    - Ziagen for HIV/AIDS (viral mutation)
    - 6-Mercaptopurine for pediatric leukemia (TPMT test)

Genetic variation associated with drug response
Focus drug treatment, avoid AEs

Gene polymorphism	Drug Response Affected
LTC_4 synthase	montelukast, zafirlukast
β2 adrenergic receptor	albuterol
ACE	ACE inhibitors
Cholesterol ester transfer protein	pravastatin
Potassium channels	AF, drug-induced QT prolongation

Applications of genetic variation to clinical practice

- Improved diagnosis, “splitting” of diseases
- Customization of medication dose, therapy
  - Bring into line with other consumer products
  - Decrease AE rates/costs, increase compliance (?)
- Being promoted with little regulation