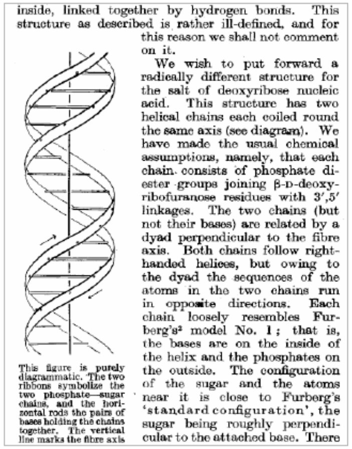
The Birth of Molecular Biology: DNA Structure



inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

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

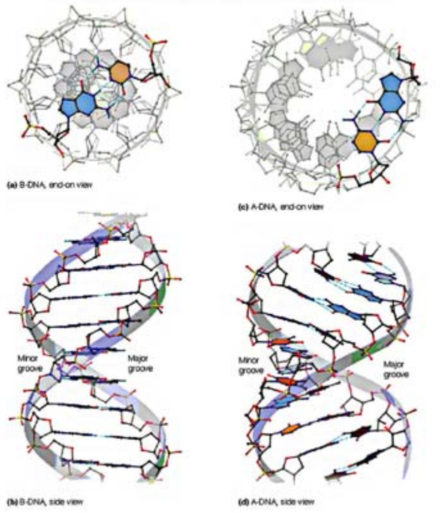
This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

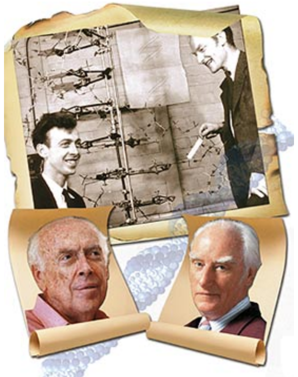
Nature – 1953

15 February 2001
nature
www.nature.com
the human genome
Nuclear fission
Five-dimensional energy landscapes
Seafloor spreading
The view from under the Arctic ice
Career prospects
Sequence creates new opportunities

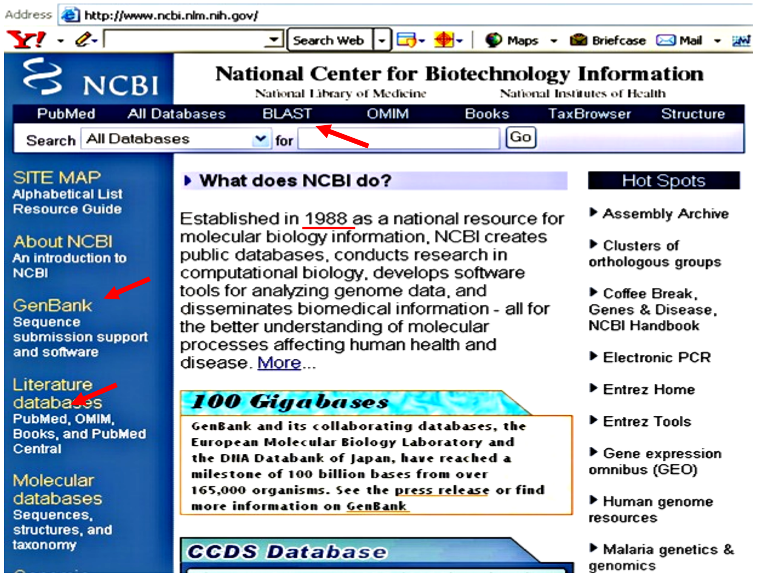
Nature – 2001

A and B Double Helices





Fall 2007



Address: http://www.ncbi.nlm.nih.gov/

National Center for Biotechnology Information

National Library of Medicine — National Institutes of Health

SITE MAP
Alphabetical List
Resource Guide

About NCBI
An introduction to NCBI

GenBank
Sequence submission support and software

Literature databases
PubMed, OMIM, Books, and PubMed Central

Molecular databases
Sequences, structures, and taxonomy

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...

100 Gigabases

GenBank and its collaborating databases, the European Molecular Biology Laboratory and the DNA Databank of Japan, have reached a milestone of 100 billion bases from over 165,000 organisms. See the press release or find more information on GenBank

CCDS Database

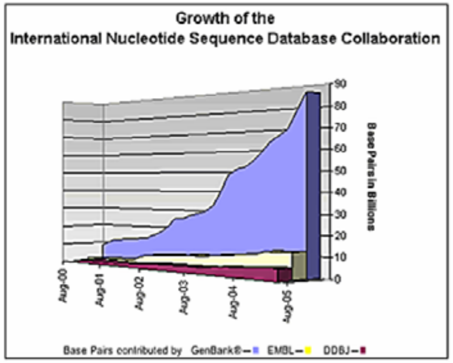
Hot Spots

- Assembly Archive
- Clusters of orthologous groups
- Coffee Break, Genes & Disease, NCBI Handbook
- Electronic PCR
- Entrez Home
- Entrez Tools
- Gene expression omnibus (GEO)
- Human genome resources
- Malaria genetics & genomics

International sequence databases exceed 100 gigabases

In August 2005, the INSDC announced the DNA sequence database exceeded 100 gigabases. GenBank is proud of its contributions toward this milestone. We thank all the scientists who have worked through the submission process at GenBank and made their sequence data available to the world. See the related press release.

>100,000,000,000 bases



> 200,000 organisms!!

Sequencing DNA

Prior to the mid-1970’s no method existed by which DNA could be directly sequenced. Knowledge about gene and genome organization was based upon studies of prokaryotic organisms and the primary means of obtaining DNA sequence was so-called reverse genetics in which the amino acid sequence of the gene product of interest is back-translated into a nucleotide sequence based upon the appropriate codons.

- •Maxam-Gilbert DNA Sequencing
- •Sanger (didexoy) DNA Sequencing

Maxam-Gilbert DNA Sequencing

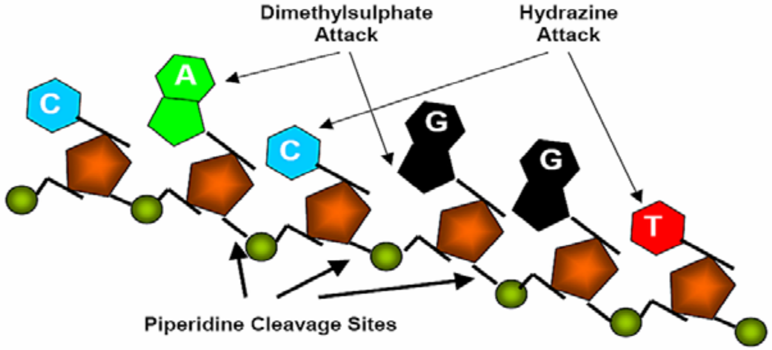


Figure 1. Chemical targets in the Maxam-Gilbert DNA sequencing strategy. Dimethylsulphate or hydrazine will attack the purine or pyrimidine rings respectively and piperidine will cleave the phosphate bond at the 3’ carbon.

http://www.idtdna.com/support/technical/TechnicalBulletinPDF/DNA_Sequencing.pdf

INNOVATION & PRECISION IN NUCLEIC ACID SYNTHESIS IDT® INTEGRATED DNA TECHNOLOGIES

IDTutorial: DNA Sequencing

Allan Maxam / Walter Gilbert DNA Sequencing

Sequencing single-stranded DNA

Two-step catalytic process:

1) Break glycoside bond between the ribose sugar and the base / displace base

Purines react with dimethyl sulfate

Pyrimidines react with hydrazine

2) Piperidine catalyzes phosphodiester bond cleavage where base displaced

---

“G” - dimethyl sulfate and piperidine

“A + G” - dimethyl sulfate and piperidine in formic acid

“C” - hydrazine and piperidine in 1.5M NaCl

“C + T” - hydrazine and piperidine

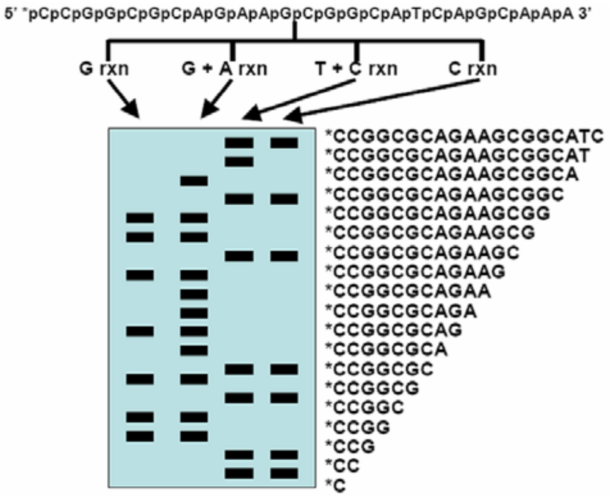


Figure 2. The Maxam-Gilbert manual sequencing scheme. The target DNA is radiolabeled and then split into the four chemical cleavage reactions. Each reaction is loaded onto a polyacrylamide gel and run. Finally, the gel is autoradiographed and base calling proceeds from bottom to top.

Maxam-Gilbert DNA Sequencing

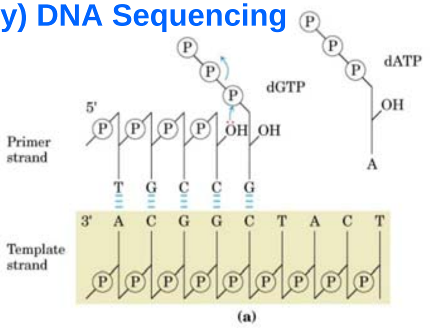
- 200-300 bases of DNA sequence every few days
- Use large amounts of radioactive material, 35S or 32P
- Constantly pouring large, paper thin acrylamide gels
- Hydrazine is a neurotoxin

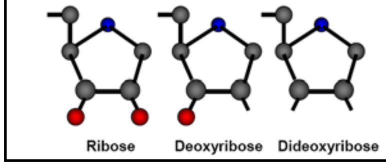
Early Benefits -

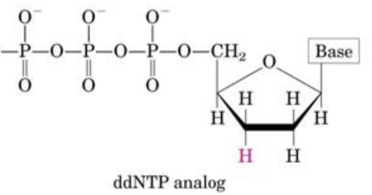
Discovery that the gene for ovalbumin in chicken and the gene encoding β-globin in rabbit contained non-coding gaps in the coding regions. These gaps< were flanked by the same dinucleotides in the two genes; GT on the 5' end of the gaps and AG on the 3' end of the gaps. Soon, the terms intron and exon were added to the genetic lexicon to describe the coding and non-coding regions of eukaryotic genes (1977).

Fred Sanger (dideoxy) DNA Sequencing

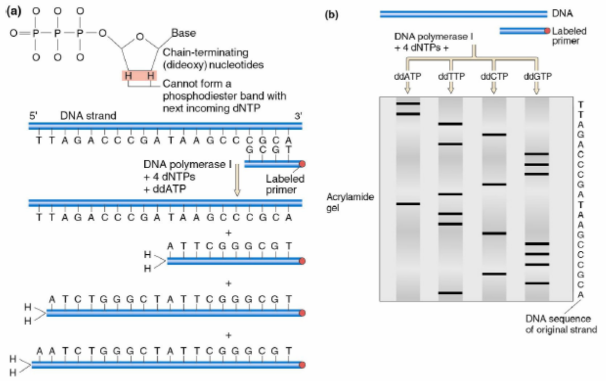
Sanger knew that, whenever a dideoxynucleotide was incorporated into a polynucleotide, the chain would irreversibly stop, or terminate. Thus, the incorporation of specific dideoxynucleotides in vitro would result in selective chain termination.

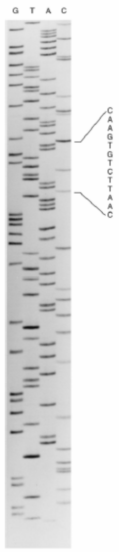






Sanger (dideoxy) DNA Sequencing


(a)
Base
Chain-terminating (dideoxy) nucleotides
Cannot form a phosphodiester band with next incoming dNTP
5'
DNA strand
3'
T T A G A C C C G A T A A G C C C G C A
G C G T
DNA polymerase I + 4 dNTPs + ddATP
Labeled primer
T T A G A C C C G A T A A G C C C G C A
+
A T T C G G G C G T
+
A T C T G G G C T A T T C G G G C G T
+
A A T C T G G G C T A T T C G G G C G T
H
(b)
DNA
DNA polymerase I + 4 dNTPs +
Labeled primer
ddATP
ddTTP
ddCTP
ddGTP
Acrylamide gel
T T A G A C C C G A T A A G C C C G C A
DNA sequence of original strand



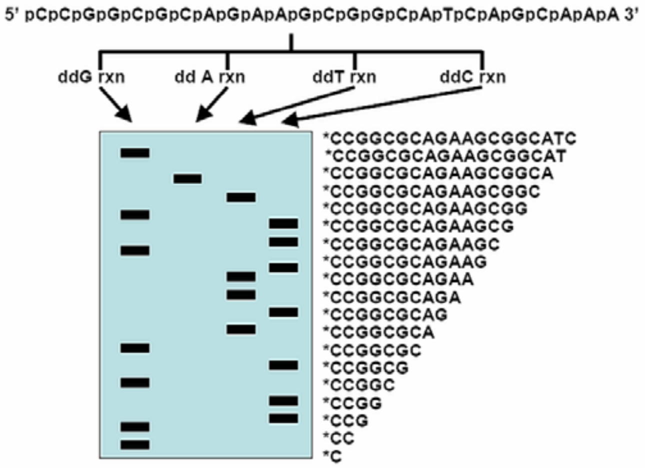


Figure 4. A Sanger sequencing scheme. Here, a sequencing primer is radiolabeled and the reaction involves generation of sequence fragments that are the complement of the template DNA. The sequence fragments are resolved on a polyacrylamide gel and the gel is autoradiographed. Base calling for the template is the complement of the gel band.

Advantages of dideoxy DNA Sequencing

- Elimination of dangerous chemicals (hydrazine)
- Greater efficiency (>3x)

Taq polymerase makes DNA strands off of a template at rate of about 500 bases per minute

Chemical synthesis of a 25-mer oligonucleotide takes more than two hours.

→ High Throughput Methods (Human Genome Project)

Automated Fluorescence Sequencing

In 1986, Leroy Hood and colleagues reported on a DNA sequencing method in which the radioactive labels, autoradiography, and manual base calling were all replaced by fluorescent labels, laser induced fluorescence detection, and computerized base calling.

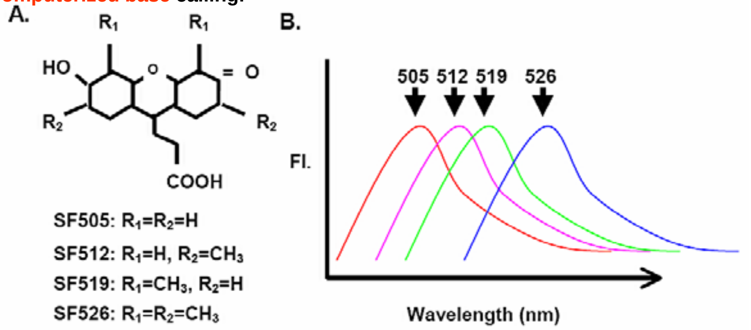
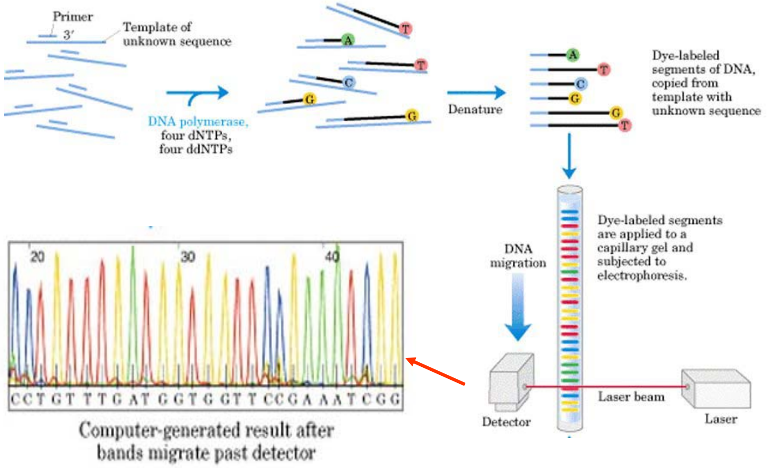


Figure 5. A. Chemical structure of the four succinylfluorescein dyes developed at DuPont. B. Normalized fluorescence emission spectra for each of the four dyes following excitation at 488nm. Shifts in the spectra were achieved by changing the side groups R_1 and R_2.

Automated DNA sequencing


Primer
3′
Template of unknown sequence
DNA polymerase, four dNTPs, four ddNTPs
A
T
T
C
G
G
Denature
A
T
C
G
G
T
Dye-labeled segments of DNA, copied from template with unknown sequence
Dye-labeled segments are applied to a capillary gel and subjected to electrophoresis.
DNA migration
Detector
Laser beam
Laser
20
30
40
CCTGTTTGATGGTGGTTCCGAAATCGG
Computer-generated result after bands migrate past detector

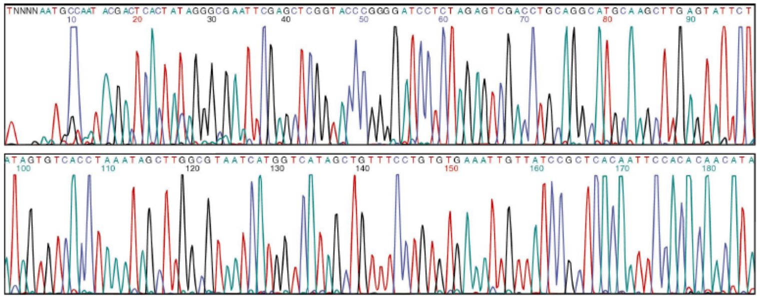
Automated dye-terminator sequencing

4-fluorescently labelled dideoxy dye terminators

ddATP
ddGTP
ddCTP
ddTTP

} pool and load in a single well or capillary

- scan with laser + detector specific for each dye
- automated base calling
- very long reads (~ 1000 bases)/run



Human Genome Project

Begun formally in 1990, the U.S. Human Genome Project was a 13-year effort coordinated by the U.S. Department of Energy and the National Institutes of Health. The project originally was planned to last 15 years, but rapid technological advances accelerated the completion date to 2003.

Project goals:

- ***identify*** **all the approximately 20,000-25,000 genes in human DNA,**
- ***determine*** **the sequences of ~3 billion chemical base pairs of human DNA,**
- ***store*** **this information in databases,**
- ***improve*** **tools for data analysis,**
- ***transfer*** **related technologies to the private sector, and**
- ***address*** **the ethical, legal, and social issues (ELSI) from the project.**
- ***sequence 500 Mb/year at < $0.25 per finished base***

  (Sequenced >1,400 Mb/year at <$0.09 per finished base)
- ***complete genome sequences of*** ***E. coli, S. cerevisiae, C. elegans, D. melanogaster***
- ***develop genomic-scale technologies*** ***(oligo syn, DNA microarrays, 2-hybrid sys)***

HGP Hero - Jim Kent (research scientist at UC Santa Cruz)

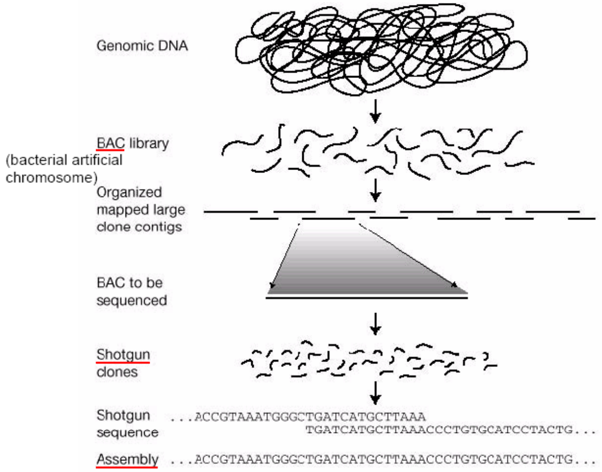
The human genome project was ultimately a race between **Celera Genomics** and the **public** effort, with the final push being a bioinformatics problem to put all of the sequence reads together into a draft genome sequence. **Jim Kent was a grad student at UCSC**, who worked for weeks developing the algorithm to put all of this together, **beating Celera by 3 days** to an assembled human genome sequence.

His efforts ensured that the human genome data remained in the public domain and were not patented into private intellectual property.

Kent built a grid of cheap, commodity PC's running the Linux operating system and other Freeware to beat Celera's, what was thought of then as the, world's most powerful civilian computer. In **June 2000**, thanks to the work done by Kent and several others, the **Human Genome Project** was able to publish its data in the Public Domain just hours ahead of Celera.

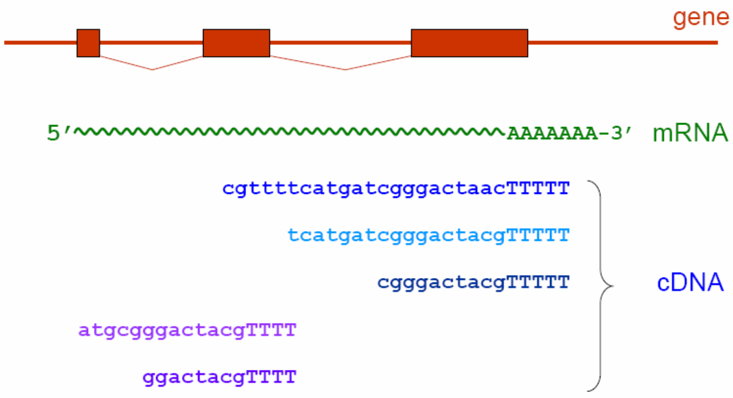
Kent went on to write BLAT and the UCSC Human Genome Browser to help analyze important genome data, receiving his PhD in biology in 2002. He remained at UCSC to work primarily on web tools to help understand the human genome. He helped maintain and upgrade the browser, and worked on projects such as comparative genomics and Parasol.

Physical mapping and sequencing of the human genome


Genomic DNA
BAC library
(bacterial artificial chromosome)
Organized mapped large clone contigs
BAC to be sequenced
Shotgun clones
Shotgun sequence
...ACCGTAAATGGGCTGATCATGCTTAAA
TGATCATGCTTAAACCCTGTGCATCCTACTG...
Assembly
...ACCGTAAATGGGCTGATCATGCTTAAACCCTGTGCATCCTACTG...

Nature (2001) **409** p. 860-921

Genes can also be identified by sequencing cDNAs at random. The sequenced cDNAs are called ESTs (expressed sequence tags)



Finding genes in genomes

- compare to EST or cDNA sequence
- look for open reading frames
- similarity to other genes and proteins
- Gene prediction algorithms (identifying splice sites, coding sequence bias, etc.)

The BIG QUESTION:

Why do we have so few genes?

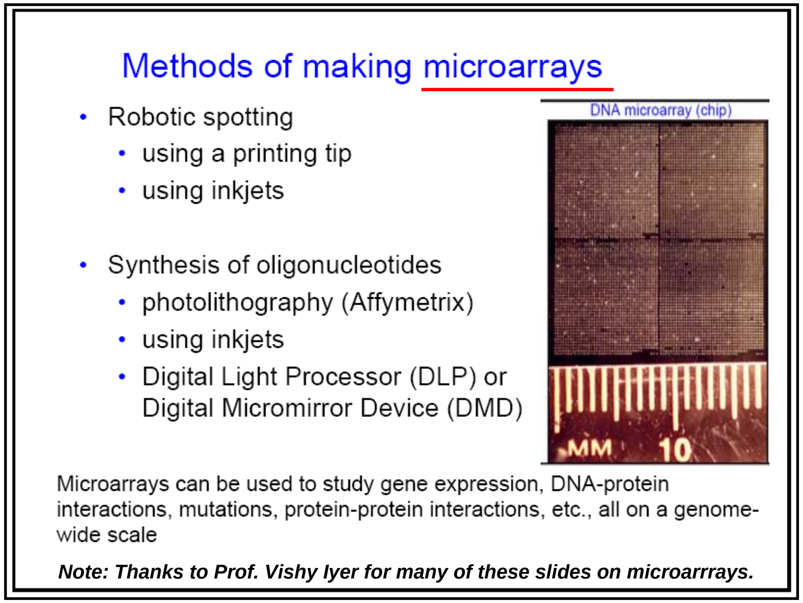
Species	Genome size	Number of genes
Human (*Homo sapiens*)	2.9 billion base pairs	25,000 - 30,000
Fruit fly (*Drosophila melanogaster*)	120 million base pairs	13,600
Worm (*Caenorhabditis elegans*)	97 million base pairs	19,000
Budding yeast (*Saccharomyces cerevisiae*)	12 million base pairs	6,000
E. coli	4.1 million base pairs	4,800

Genomics vs. Proteomics

With the completion of a rough draft of the human genome in the Spring of 2003, many researchers began looking at how genes and proteins interact to form other proteins. A surprising finding of the Human Genome Project is that there are far fewer protein-coding genes in the human genome than proteins in the human proteome (20,000 to 25,000 genes vs. about 1,000,000 proteins). The human body may contain more than 2 million proteins, each having different functions. The protein diversity is thought to be due to alternative splicing and post-translational modification of proteins. The discrepancy implies that *protein diversity cannot be fully characterized by gene expression analysis*, thus proteomics is needed for characterizing cells and tissues.

Functional genomics and proteomics

- Identify genes and proteins encoded in the genome (Gene finding)
- Measure gene expression on a genome-wide scale (microarrays)
- Identify protein function
  30-50% of the genes in a genome are of unknown function
- Identify protein interactions, biochemical pathways, gene interaction networks inside cells



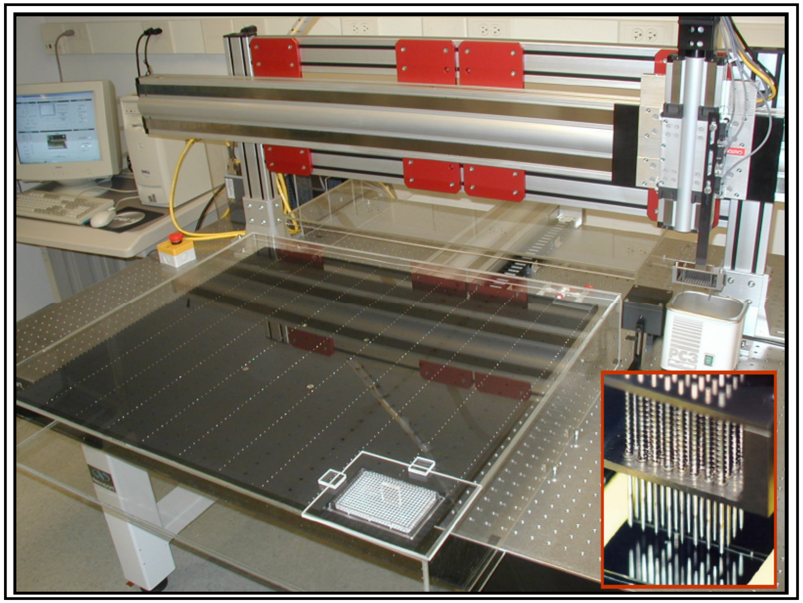
Methods of making microarrays

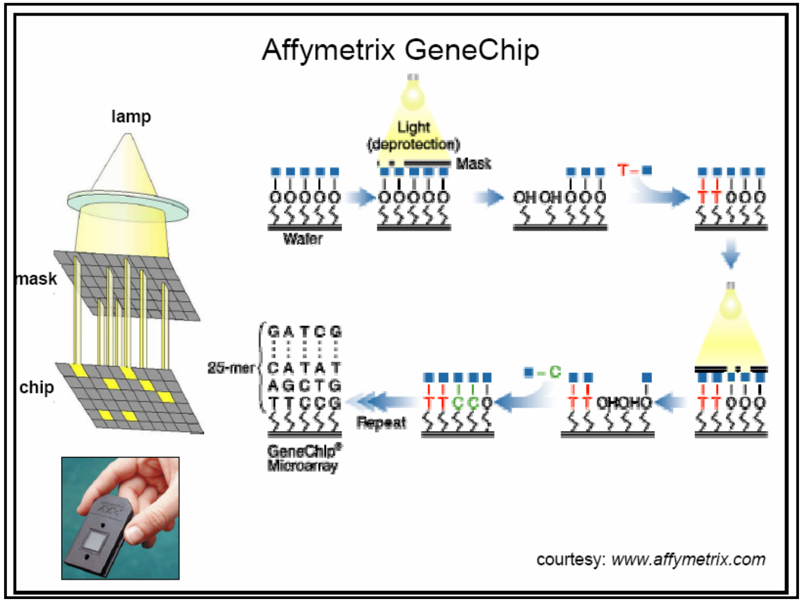
- Robotic spotting
  - using a printing tip
  - using inkjets
- Synthesis of oligonucleotides
  - photolithography (Affymetrix)
  - using inkjets
  - Digital Light Processor (DLP) or Digital Micromirror Device (DMD)

DNA microarray (chip)
MM
10

Microarrays can be used to study gene expression, DNA-protein interactions, mutations, protein-protein interactions, etc., all on a genome-wide scale

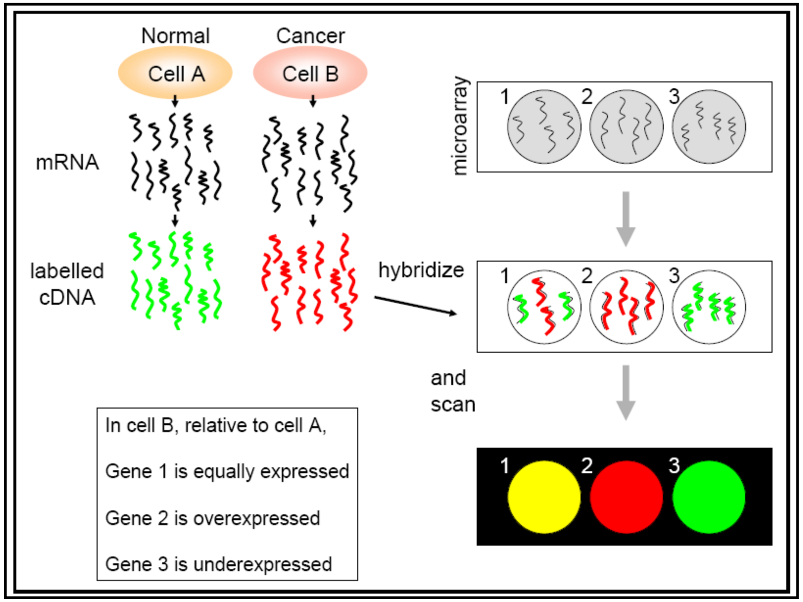
Note: Thanks to Prof. Vishy Iyer for many of these slides on microarrrays.





Affymetrix GeneChip

courtesy: *www.affymetrix.com*

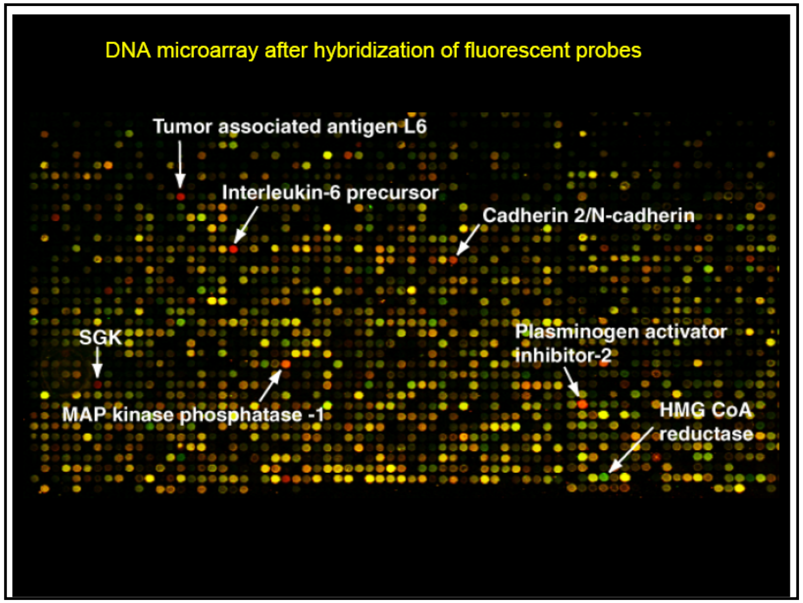


In cell B, relative to cell A,

Gene 1 is equally expressed

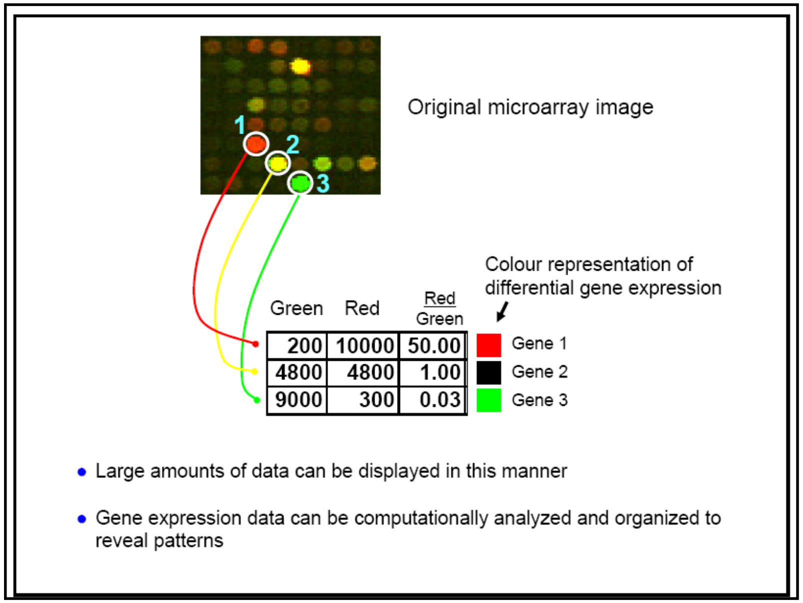
Gene 2 is overexpressed

Gene 3 is underexpressed



DNA microarray after hybridization of fluorescent probes

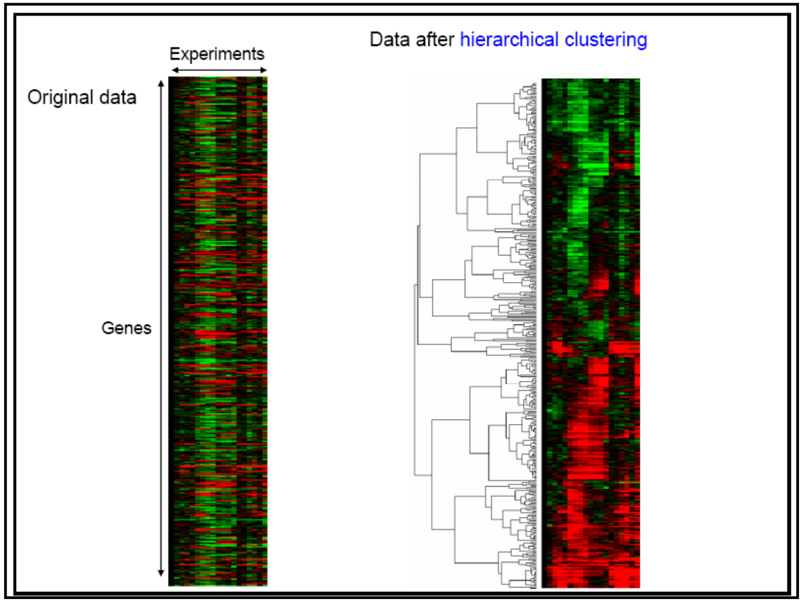
Tumor associated antigen L6
Interleukin-6 precursor
Cadherin 2/N-cadherin
SGK
Plasminogen activator inhibitor-2
MAP kinase phosphatase -1
HMG CoA reductase



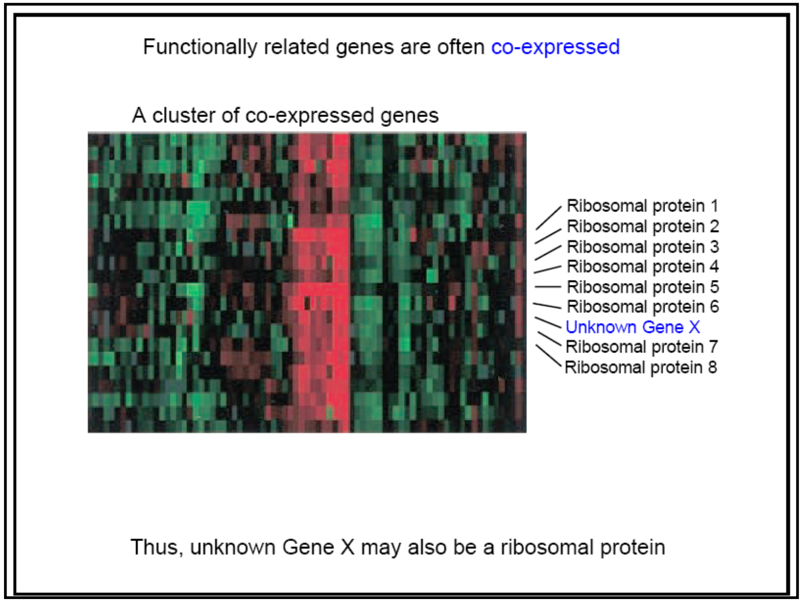
Original microarray image
1
2
3
Colour representation of differential gene expression

Green	Red	$\frac{\text{Red}}{\text{Green}}$		
200	10000	50.00	🟥	Gene 1
4800	4800	1.00	⬛	Gene 2
9000	300	0.03	🟩	Gene 3

- Large amounts of data can be displayed in this manner
- Gene expression data can be computationally analyzed and organized to reveal patterns



Data after hierarchical clustering

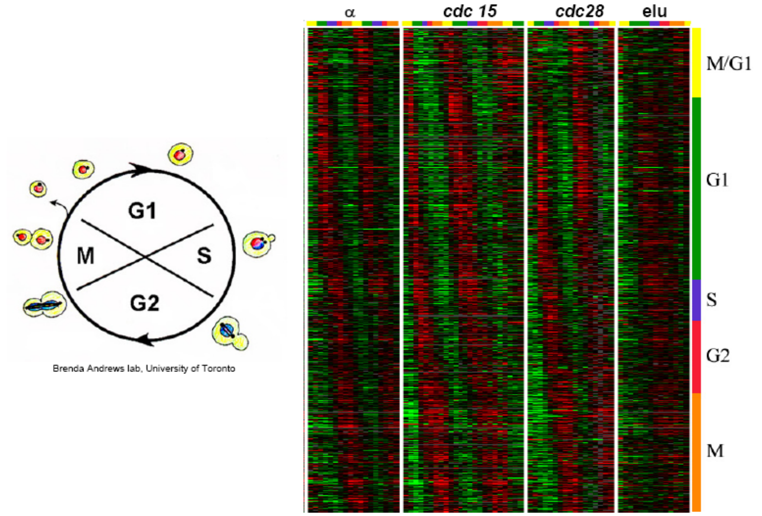


Functionally related genes are often co-expressed

A cluster of co-expressed genes

Thus, unknown Gene X may also be a ribosomal protein

S. cerevisiae mitotic cell-cycle



Brenda Andrews lab, University of Toronto

Spellman *et al*, (1998) *Mol. Biol. Cell*

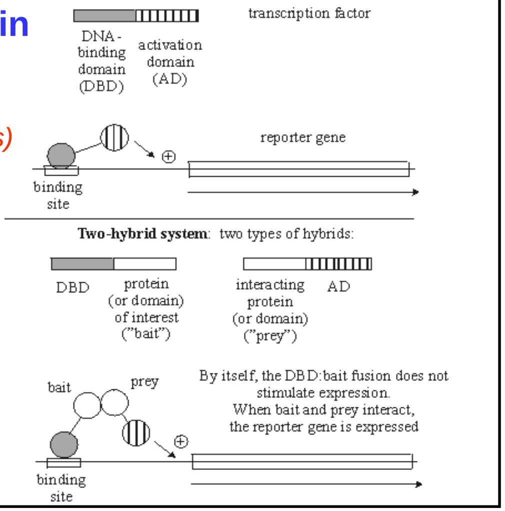
Protein – Protein Interactions

(Interaction Networks)

Also see: *Methods Mol Biol.* 2004; **270**:403-22.

Two-hybrid system

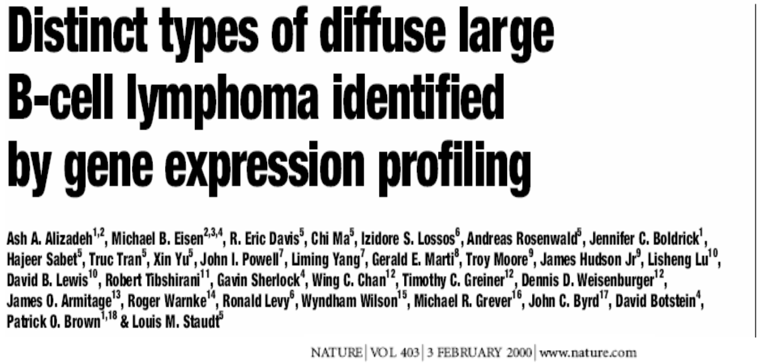
Structure-function properties of a typical **transcription factor**:


transcription factor
DNA-binding domain (DBD)
activation domain (AD)
reporter gene
binding site
⊕
Two-hybrid system: two types of hybrids:
DBD
protein (or domain) of interest ("bait")
interacting protein (or domain) ("prey")
AD
bait
prey
⊕
binding site

By itself, the DBD:bait fusion does not stimulate expression.
When bait and prey interact, the reporter gene is expressed

Some Examples / Applications

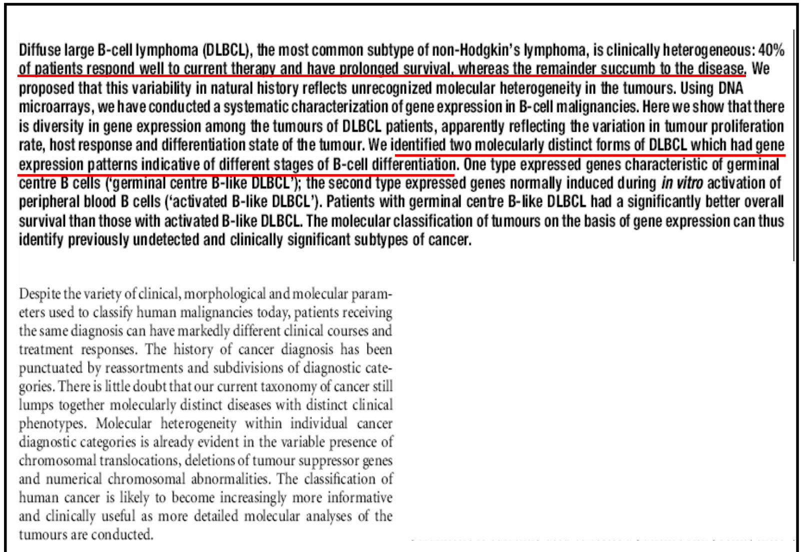
- DLBCL
- P4 - Medicine



Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling

Ash A. Alizadeh[1,2], Michael B. Eisen[2,3,4], R. Eric Davis[5], Chi Ma[5], Izidore S. Lossos[6], Andreas Rosenwald[5], Jennifer C. Boldrick[1], Hajeer Sabet[5], Truc Tran[5], Xin Yu[5], John I. Powell[7], Liming Yang[7], Gerald E. Marti[8], Troy Moore[9], James Hudson Jr[9], Lisheng Lu[10], David B. Lewis[10], Robert Tibshirani[11], Gavin Sherlock[4], Wing C. Chan[12], Timothy C. Greiner[12], Dennis D. Weisenburger[12], James O. Armitage[13], Roger Warnke[14], Ronald Levy[6], Wyndham Wilson[15], Michael R. Grever[16], John C. Byrd[17], David Botstein[4], Patrick O. Brown[1,18] & Louis M. Staudt[5]

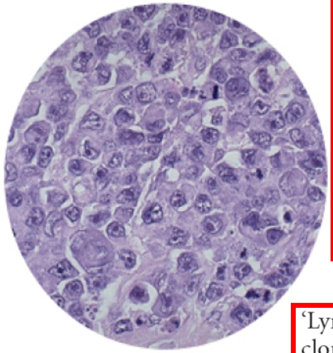
NATURE | VOL 403 | 3 FEBRUARY 2000 | www.nature.com



Diffuse large B-cell lymphoma (DLBCL), the most common subtype of non-Hodgkin's lymphoma, is clinically heterogeneous: 40% of patients respond well to current therapy and have prolonged survival, whereas the remainder succumb to the disease. We proposed that this variability in natural history reflects unrecognized molecular heterogeneity in the tumours. Using DNA microarrays, we have conducted a systematic characterization of gene expression in B-cell malignancies. Here we show that there is diversity in gene expression among the tumours of DLBCL patients, apparently reflecting the variation in tumour proliferation rate, host response and differentiation state of the tumour. We identified two molecularly distinct forms of DLBCL which had gene expression patterns indicative of different stages of B-cell differentiation. One type expressed genes characteristic of germinal centre B cells ('germinal centre B-like DLBCL'); the second type expressed genes normally induced during *in vitro* activation of peripheral blood B cells ('activated B-like DLBCL'). Patients with germinal centre B-like DLBCL had a significantly better overall survival than those with activated B-like DLBCL. The molecular classification of tumours on the basis of gene expression can thus identify previously undetected and clinically significant subtypes of cancer.

Despite the variety of clinical, morphological and molecular parameters used to classify human malignancies today, patients receiving the same diagnosis can have markedly different clinical courses and treatment responses. The history of cancer diagnosis has been punctuated by reassortments and subdivisions of diagnostic categories. There is little doubt that our current taxonomy of cancer still lumps together molecularly distinct diseases with distinct clinical phenotypes. Molecular heterogeneity within individual cancer diagnostic categories is already evident in the variable presence of chromosomal translocations, deletions of tumour suppressor genes and numerical chromosomal abnormalities. The classification of human cancer is likely to become increasingly more informative and clinically useful as more detailed molecular analyses of the tumours are conducted.

The challenge of cancer diagnosis



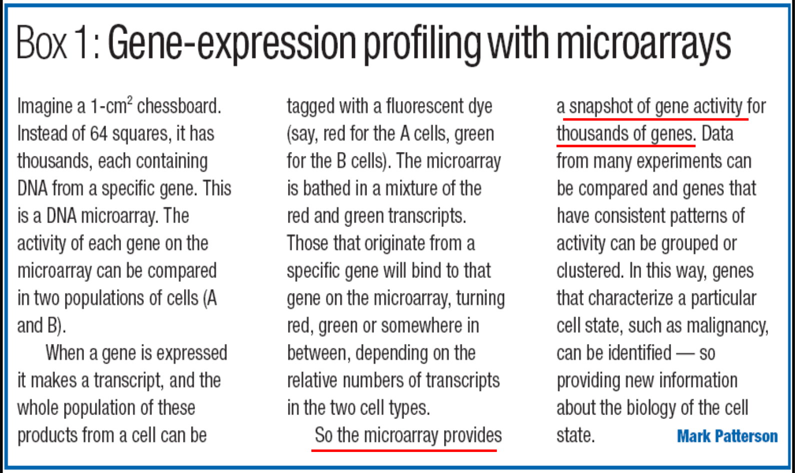
Diffuse large B-cell lymphoma is the most common subtype of non-Hodgkin's lymphoma. With current treatments, long-term survival can be achieved in only 40% of patients. There are no reliable indicators — morphological, clinical, immunohistochemical or genetic — that can be used to recognize subclasses of **DLBCL** and point to a differential therapeutic approach to patients.

'Lymphochip', a microarray carrying 18,000 clones of complementary DNA designed to monitor genes involved in normal and abnormal lymphocyte development.

What type of cancer?

What is the underlying molecular basis?

What is the optimal treatment?



Box 1: Gene-expression profiling with microarrays

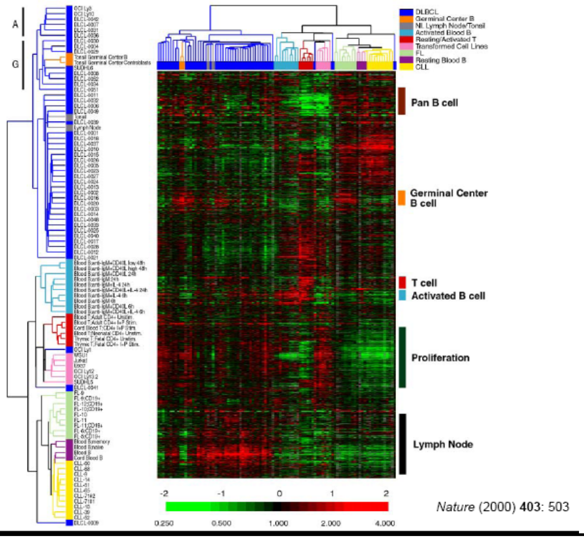
Imagine a 1-cm^2 chessboard. Instead of 64 squares, it has thousands, each containing DNA from a specific gene. This is a DNA microarray. The activity of each gene on the microarray can be compared in two populations of cells (A and B).

When a gene is expressed it makes a transcript, and the whole population of these products from a cell can be tagged with a fluorescent dye (say, red for the A cells, green for the B cells). The microarray is bathed in a mixture of the red and green transcripts. Those that originate from a specific gene will bind to that gene on the microarray, turning red, green or somewhere in between, depending on the relative numbers of transcripts in the two cell types.

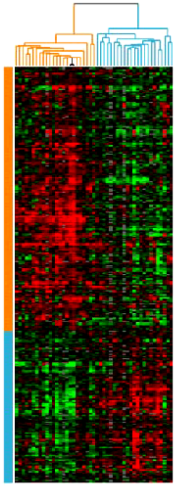
So the microarray provides a snapshot of gene activity for thousands of genes. Data from many experiments can be compared and genes that have consistent patterns of activity can be grouped or clustered. In this way, genes that characterize a particular cell state, such as malignancy, can be identified — so providing new information about the biology of the cell state.

Mark Patterson

Hierarchical clustering of gene expression data (as ratios).

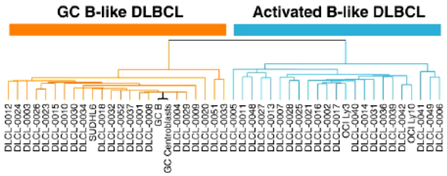


Nature (2000) **403**: 503

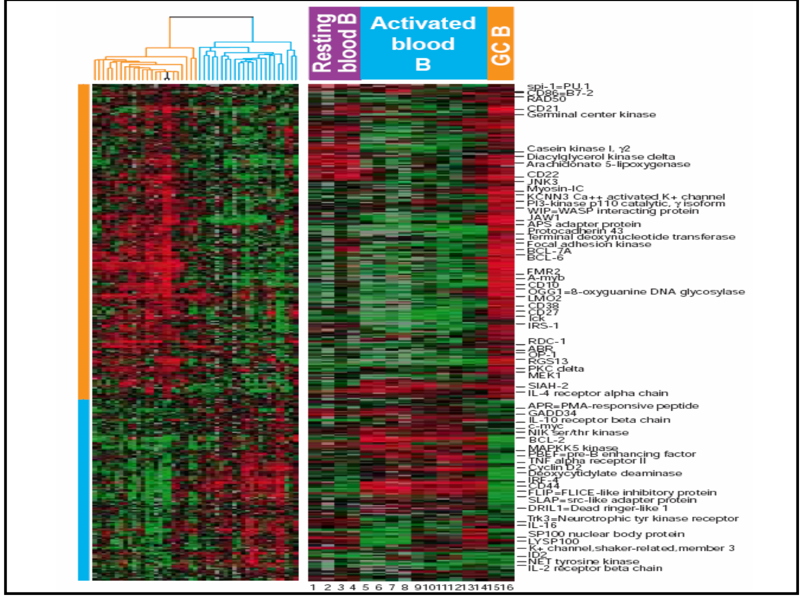


Clustering of tumour samples from cancer patients can be used for molecular classification of cancers. This may be useful for diagnosis and treatment

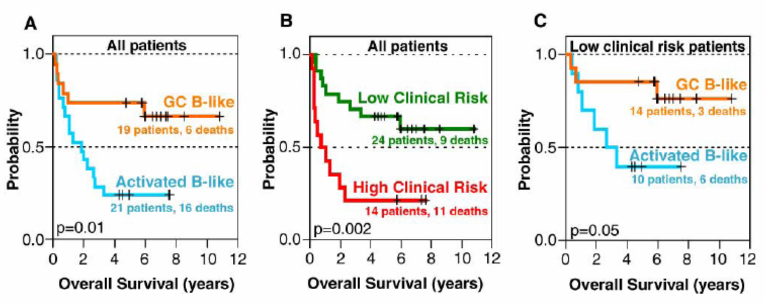
Subtypes of Diffuse Large B-Cell Lymphoma (DLBCL)



Nature (2000) **403**: 503



Using "clustering analysis," Alizadeh *et al.* could separate DLBCL into two categories, which had marked differences in overall survival of the patients concerned. The gene expression signatures of these subgroups corresponded to distinct stages in the differentiation of B cells, the type of lymphocyte that makes antibodies.



Systems Biology and Systems Medicine: Predictive, Personalized, Preventive and Participatory (P4)

Lee Hood

Institute for Systems Biology, Seattle

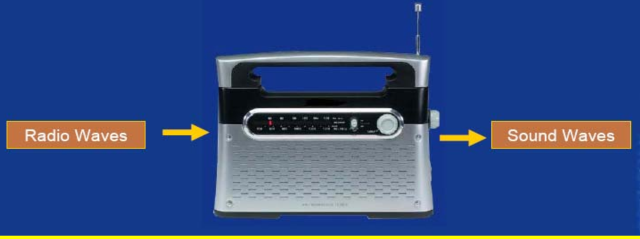
Dr. Leroy Hood
M.D., Johns Hopkins School of Medicine, 1964
Ph.D., Biochemistry, California Institute of Technology, 1968

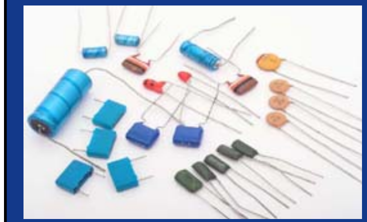
Note: The following (blue) slides were edited from a presentation by Lee Hood of the Inst. for Systems Biology to NIST on the P4 Medicine found at:

http://www.itl.nist.gov/Healthcare/conf/presentations/LH%20NIST%209-24-07.pdf

A similar lecture on P4 Medicine was presented by Dr. Hood at the 2007 Welch Conference: "Physical Biology -From Atoms to Cells"

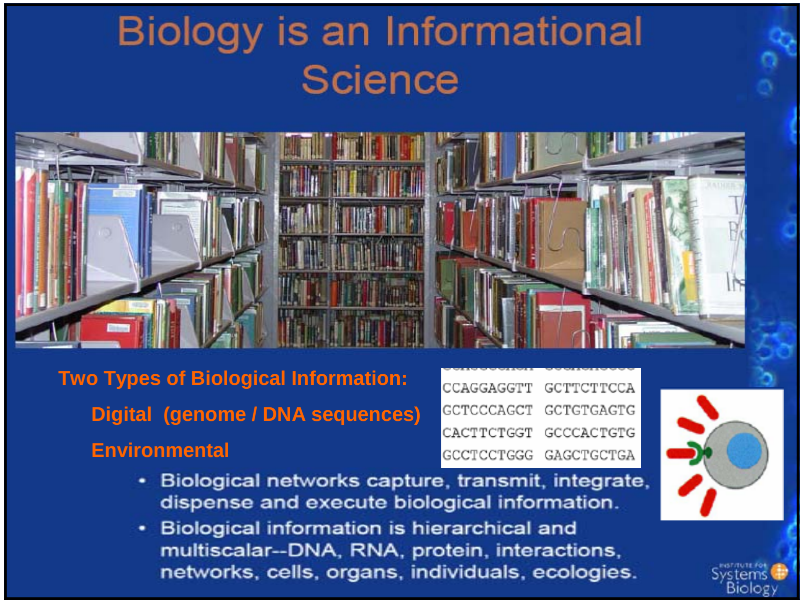
Need for a Systems Approach





~ biomolecules

~ complexes / cells



Biology is an Informational Science

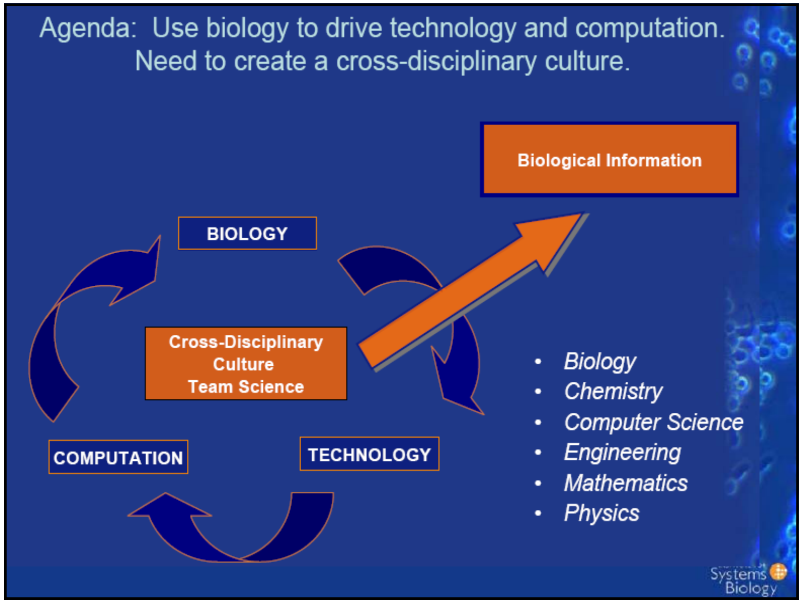
Two Types of Biological Information:

Digital (genome / DNA sequences)

Environmental

```
CCAGGAGGTT  GCTTCTTCCA
GCTCCCAGCT  GCTGTGAGTG
CACTTCTGGT  GCCCACTGTG
GCCTCCTGGG  GAGCTGCTGA
```

- Biological networks capture, transmit, integrate, dispense and execute biological information.
- Biological information is hierarchical and multiscalar--DNA, RNA, protein, interactions, networks, cells, organs, individuals, ecologies.



Agenda: Use biology to drive technology and computation. Need to create a cross-disciplinary culture.

Biological Information

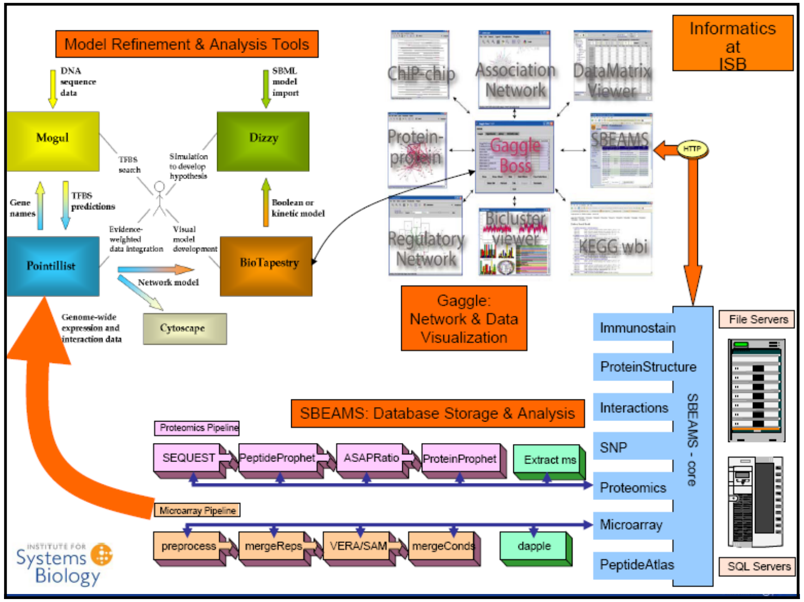
BIOLOGY

Cross-Disciplinary Culture
Team Science

COMPUTATION

TECHNOLOGY

- *Biology*
- *Chemistry*
- *Computer Science*
- *Engineering*
- *Mathematics*
- *Physics*

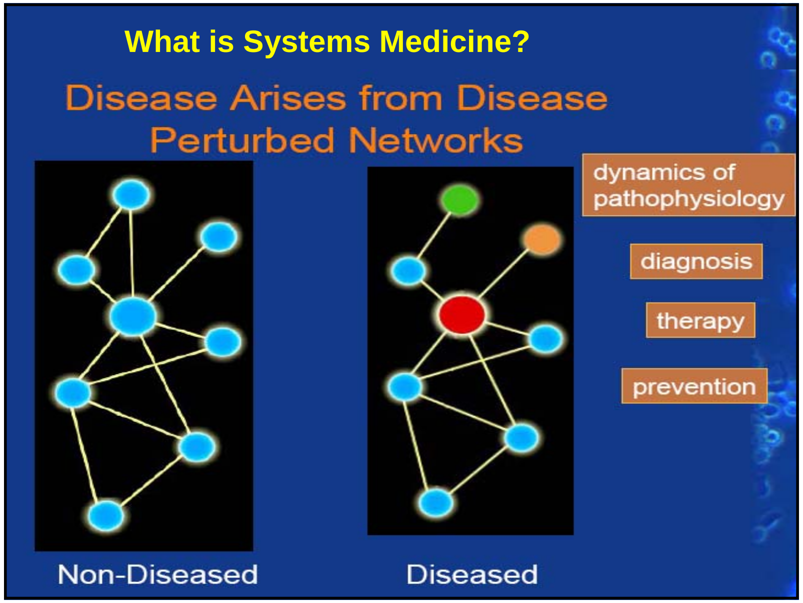


Informatics at ISB

Model Refinement & Analysis Tools

Gaggle: Network & Data Visualization

SBEAMS: Database Storage & Analysis



What is Systems Medicine?

Disease Arises from Disease Perturbed Networks

dynamics of pathophysiology

diagnosis

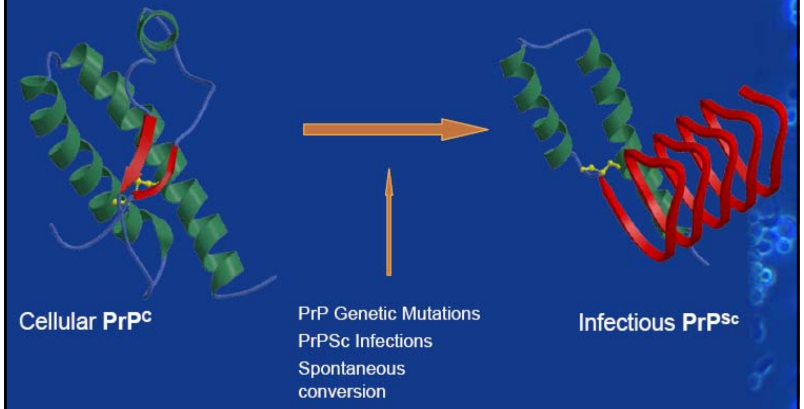
therapy

prevention

Non-Diseased

Diseased

Prion Protein Exists in Two Forms

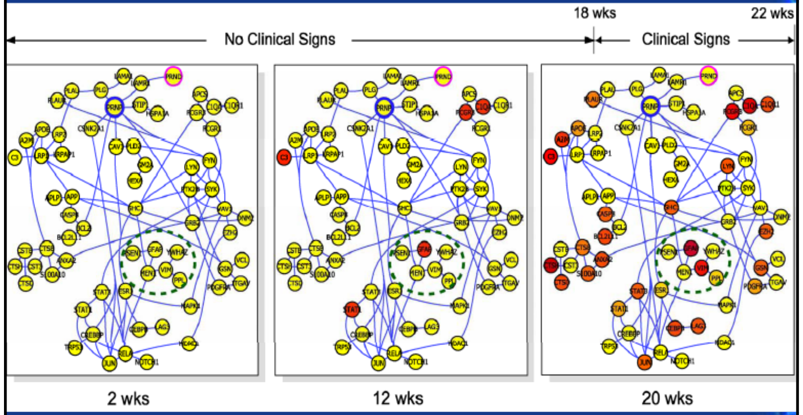


Cellular PrP^{C}

PrP Genetic Mutations
PrPSc Infections
Spontaneous conversion

Infectious PrP^{Sc}

Dynamics of a Prion Perturbed Network

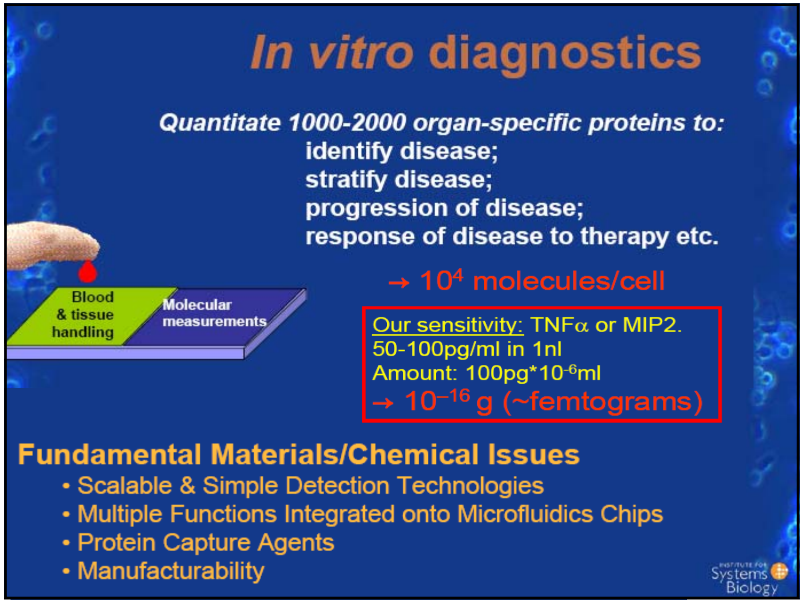


DEGs Encoding Known and Novel Prion Disease Phenotypes

- 7400 Differentially Expressed Genes (DEGs) in 5 inbred strains upon prion perturbation.
- Biological filters reduce to 924 core DEGs for prion disease
- 253/924 DEGs encode known disease phenotypes
- 671/924 DEGs encode novel disease phenotypes

Organ-Specific Blood Proteins Will Make the Blood a Window into Health and Disease

- Perhaps 50 major organs or cell types--each secreting protein blood molecular fingerprint.
- The levels of each protein in a particular blood fingerprint will report the status of that organ. Probably need 10-50 organ-specific proteins per organ.
- Need to quantify 500-2500 blood proteins from a droplet of blood.
- Key point: changes in the levels of organ-specific markers will assess all diseases or environmental challenges for a particular organ



In vitro diagnostics

Quantitate 1000-2000 organ-specific proteins to:

identify disease;
stratify disease;
progression of disease;
response of disease to therapy etc.

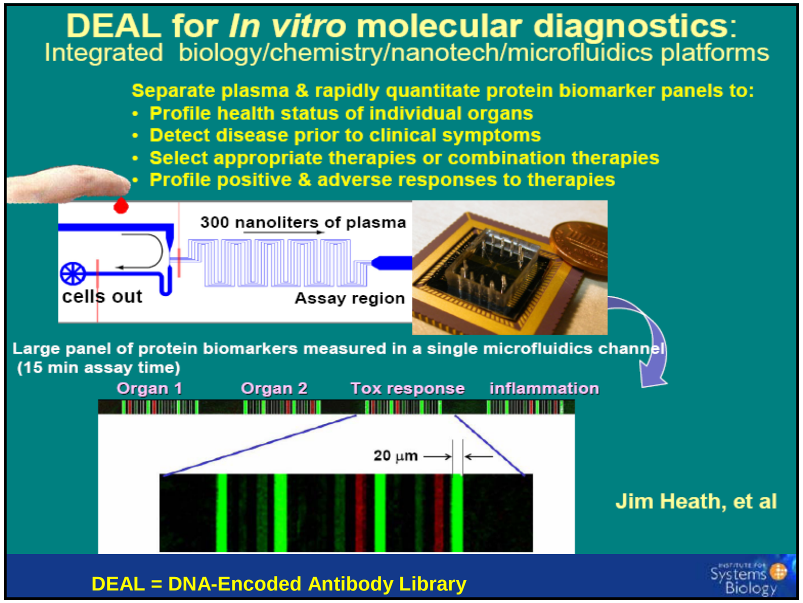
→ 10^4 molecules/cell

Our sensitivity: TNFα or MIP2.
50-100pg/ml in 1nl
Amount: 100pg*10^{-6}ml
→ 10^{-16} g (~femtograms)

Fundamental Materials/Chemical Issues

- Scalable & Simple Detection Technologies
- Multiple Functions Integrated onto Microfluidics Chips
- Protein Capture Agents
- Manufacturability

Blood & tissue handling
Molecular measurements



DEAL for *In vitro* molecular diagnostics:

Integrated biology/chemistry/nanotech/microfluidics platforms

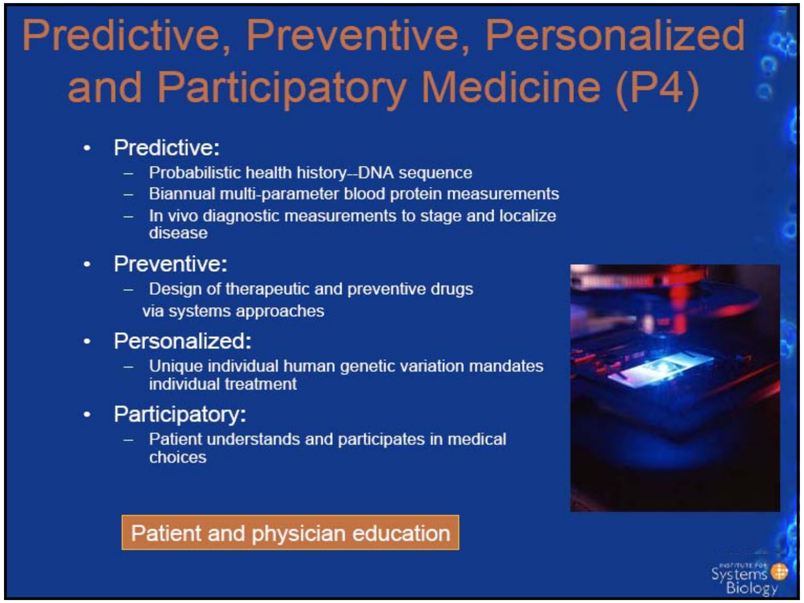
Separate plasma & rapidly quantitate protein biomarker panels to:

- **Profile health status of individual organs**
- **Detect disease prior to clinical symptoms**
- **Select appropriate therapies or combination therapies**
- **Profile positive & adverse responses to therapies**

Large panel of protein biomarkers measured in a single microfluidics channel (15 min assay time)

Jim Heath, et al

DEAL = DNA-Encoded Antibody Library



Predictive, Preventive, Personalized and Participatory Medicine (P4)

- Predictive:
  - Probabilistic health history--DNA sequence
  - Biannual multi-parameter blood protein measurements
  - In vivo diagnostic measurements to stage and localize disease
- Preventive:
  - Design of therapeutic and preventive drugs via systems approaches
- Personalized:
  - Unique individual human genetic variation mandates individual treatment
- Participatory:
  - Patient understands and participates in medical choices

Patient and physician education