Bioinformatics

Eight
Human Genome Variation
Beyond the HGP

Sami Khuri
Department of Computer Science
San José State University
San José, California, USA
sami.khuri@sjsu.edu
www.cs.sjsu.edu/faculty/khuri

Genomic Variations

- Allele
- Polymorphism
- Mutation
- SNP
- Tag SNP
- Haplotype
- HapMap Project

Pathway to Genomic Medicine

Human Genome Project	ENCODE Project	HapMap Project	Genomic Medicine
Sequencing of the human DNA	Interpreting the human genome sequence	Implicating genetic variants with human disease	Personalized medicine Cure for diseases

Building upon the Foundations of HGP

- As we build upon the foundation laid by the **Human Genome Project**, our ability to explore uncharted frontiers will hinge upon melding biological know-how with expertise in computer science, physics, math, clinical research, bioethics, and many other disciplines.
- A firm understanding of the powerful potential of **genomics**, **proteomics**, and **bioinformatics** will be essential to success in this amazing new world.

Discovering Genomics, Campbell, 2007 – Preface by Francis Collins

Genomics is a Way of Seeing Life

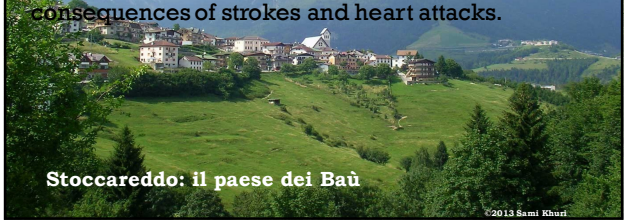
- **Genome**: the complete (haploid) DNA content of an organism.
- **Genomics**: the field of genome studies.
- **Genomics**
  - is not just a collection of methods
  - has become an enhanced way of seeing life.
- **Genomics** includes the study of interaction of molecules inside the cell:

  DNA Protein Lipids Carbohydrates
- **Genomics** requires us to analyze, hypothesize, think, and formulate models.

Wellderly: Healthy Aging

Country of the Baù

Stoccareddo is in Italy. It is known for its 400 villagers
(380 of them with the surname Baù) with great health who
tend to be able to consume fatty foods without the consequences of strokes and heart attacks.



Stoccareddo: il paese dei Baù

All known cases of **Nova Scotia Niemann-Pick Disease** were traced to a couple born in Nova Scotia in the late 1600s. It is speculated that this disease is the result of a single mutation in one of the ancestors of Yarmouth County French Acadians. In the district where most affected children lived, the heterozygote frequency was estimated to be between 10% to 26% of the population.

Cape Forchu Lighthouse in Yarmouth Harbour in Nova Scotia

Pathway to Genomic Medicine

Human Genome Project	ENCODE Project	HapMap Project	Genomic Medicine
Sequencing of the human DNA	Interpreting the human genome sequence	Implicating genetic variants with human disease	Personalized medicine Cure for diseases

ENCODE

ENCODE: **ENC**yclopedia **O**f **D**NA **E**lements

- **Goal**: Compile a Comprehensive Encyclopedia of all Functional Elements in the Human Genome
- Initial Pilot Project: 1% of Human Genome
- Apply Multiple, Diverse Approaches to Study and Analyze that 1% in a Consortium Fashion

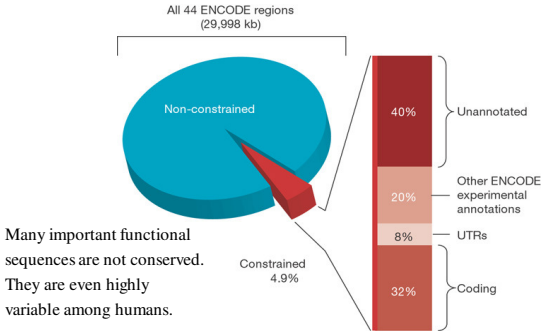
The ENCODE Project

- 44 regions of the human genome were selected, spanning 30 megabases (about 1% of the human genome).
- The **ENCODE regions** include:
  - About 50% randomly selected loci
  - About 50% containing well-known genes
    - Example: alpha and beta globins, CFTR
- The **ENCODE Project Consortium** released its findings in a 2007 article (>250 coauthors)

Major Findings of ENCODE

- The majority of all nucleotides are **transcribed** as part of
  - Coding transcripts
  - Noncoding RNAs
  - Random transcripts that may have no biological function.
- Many genes have multiple, previously undetected, **transcription start sites**
  - Regulatory sequences are as likely to be upstream as downstream of the major start sites.

Constrained Sequences



Many important functional sequences are not conserved. They are even highly variable among humans.

Human Genomic Variation

Learning outcomes:

- Understanding human **single nucleotide polymorphisms** (**SNP**s) and mining SNP databases.
- Discovering how **SNP**s can cause (are associated with) diseases.
- Examining how **SNP**s can affect medical therapies.

Genomic Variations

- Collection of genomic variations makes any person a unique human being. It contributes to that person's:
  - Potential to learn
  - Predisposition to disease
  - Predisposition to drug addiction
  - Response to pharmaceutical interventions
- There are variations within, as well as, between populations.
- The variation between individual genomes has sparked a biotech boom in the area of SNP discovery.

Reference Human Sequence

- The **reference sequence** for the human genome should not be viewed as just one long string of static characters.
- Instead, it is riddled with variable sites all along the sequence.
- Given that the number of people exceeds the number of bases in the genome, we can imagine that every base in the genome has had its chance to be different. [Baxevanis & Ouellette, 2005]

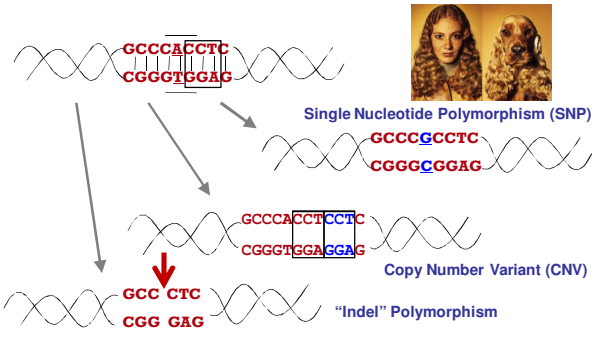
Variation in Human Genome

- How much variation is there in the human genome?
  - The biomedical field is interested in disease-causing variations.
  - What is often considered as a "simple" disease has complex genomic underpinnings.
- How are **genomic variations** used to determine the causes of complex phenotypes?
- How do **genomic variations** influence effective medical interventions?

Human Genetic Variation

- Copy Number Variation (CNV)
  - A polymorphism in which the number of repeats of a DNA sequence at a location varies from person to person
- Single Nucleotide Polymorphism (SNP)
  - Major differences between human beings
- Other structural variations
  - Includes deletions, insertions, duplications, inversions, and translocations

Types of Genomic Variations



Single Nucleotide Polymorphism

Single Nucleotide Polymorphisms are single bases at a particular locus that are different in different individuals.

```
GCATGCATGCATGCAT
||||||||||||||||
CGTACGTACGTACGTA
```

90% of all human chromosomes have the following sequence at a particular location (i.e., unique locus)

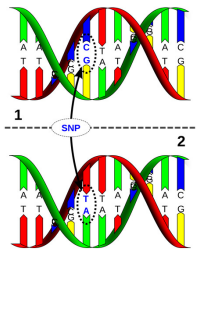
```
GCATGCAaGCATGCAT
||||||||||||||||
CGTACGTtCGTACGTA
```

But 10% of all alleles have a slightly different sequence at that particular location (i.e., unique locus)

What is a Polymorphism?

- A **polymorphism** is a difference in DNA sequence among individuals.
- **Genetic variations** occurring in more than 1% of a population would be considered useful polymorphisms for genetic analysis.
- **SNP**: position in a genome at which two or more different bases occur in the population, each with a frequency greater than 1%.

SNPs and Human Variations



To classify a variation as a SNP it should occur in at least 1% of the population.

Where do SNPs Fall?

- SNPs may fall:
  - within coding sequences of genes,
  - noncoding regions of genes, or
  - in the intergenic regions between genes.

SNPs and Haplotypes

- **SNPs** lying in close proximity in genome regions that tend to be unaffected by genomic shuffling during meiosis are usually inherited together
- Haplotypes are groups of SNPs transmitted in "blocks".
- These blocks can be characterized by a subset of their SNPs (tags).

Applications of SNPs (I)

SNPs are useful for several types of research

1) SNPs and the study of **Evolution**
   - **Example**: Different combinations of SNPs of the taste receptor gene: *Tas2R*.

2) SNPs and **Fingerprinting**
   - **Example**: Criminals and Parental Verification.

Applications of SNPs (II)

3) SNPs in **Biomedical Research**

Example: Manufacturing genotype-specific medication

Most genes contain at least one **SNP**, some of which might have functional consequences.

SNPs could be used to determine which combination of coding alleles is associated with a particular disease.

Phenylthiocarbamide (PTC)

Table 4.4 Global PTC taster SNP frequencies.
PAV is the only taster allele. Sample size for each population appears in parentheses.

SNP Combinations	European (200)	West Asian (22)	East Asian (54)	African (24)	SW Native American (18)
AVI	0.47	0.67	0.31	0.25	—
AAV	0.03	—	—	0.04	—
AAI	—	—	—	0.17	—
PVI	—	—	—	0.04	—
PAV	0.49	0.33	0.69	0.50	1.00

To some individuals the chemical compound **phenylthiocarbamide** (**PTC**) has an intensely bitter taste, while to others it is tasteless. It depends on the SNPs that are present in the receptor gene *Tas2R*.

SNPs and Evolution

- **SNP**s can be used in the study of **evolution**.
- Scientists tested 6 nonhumans primates and found that they were all tasters, in other words, they had the PAV form of *Tas2R*.
- Consequently, humans acquired (evolved) the other SNPs: AVI, AAV, AAI and PVI, after the split from our nearest relative, the chimpanzee.

Haplotypes and Evolution

- **Haplotypes** are groups of SNPs transmitted in "blocks".
- These blocks can be characterized by a subset of their SNPs (tags).
- Since they are the result of an underlying evolutionary process, they can be used to reconstruct ancestral DNA.

A National Center for Biomedical Computing, Harvard University

SNPs and Biomedical Research (I)

Two Populations

Single-locus disease

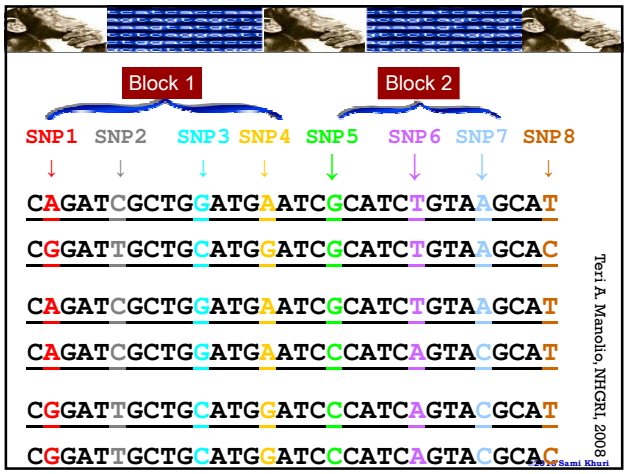
Frequency P1	Frequency P2	Genomic DNA	SNP Haplotype	Phenotype
81%	49%	SNP1 allele *A* SNP2 SNP1 allele *A* SNP2	1 – 2 1 – 2	*wt*
18%	42%	SNP1 allele *A* SNP2 SNP1' allele *a* SNP2'	1 – 2 1' – 2'	*wt*
1%	9%	SNP1' allele *a* SNP2' SNP1' allele *a* SNP2'	1' – 2' 1' – 2'	disease

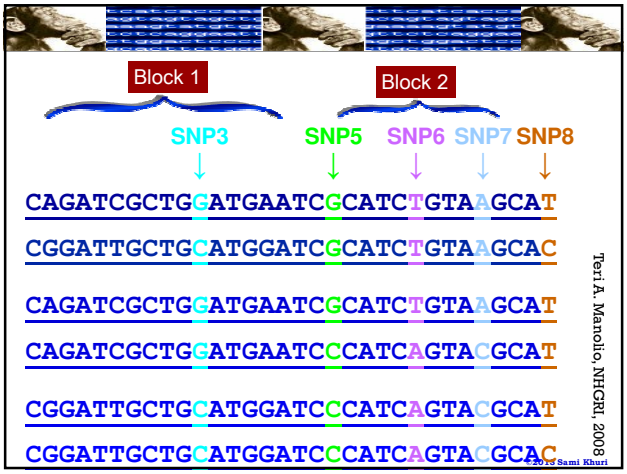
Hypothetical example in which 2 SNPs and one gene are associated with a monogenic disease.

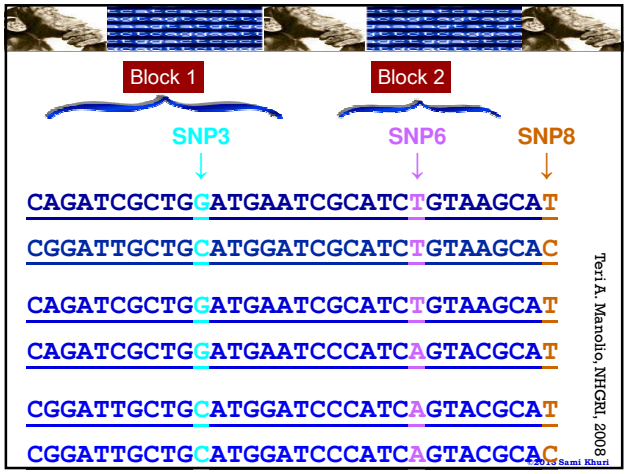
SNPs and Biomedical Research (II)

Frequency P1	Frequency P2	Genomic DNA	SNP Haplotype	Phenotype
81%	49%	SNP1 allele A SNP2 / SNP1 allele A SNP2	1 – 2 / 1 – 2	*wt*
18%	42%	SNP1 allele A SNP2 / SNP1' allele a SNP2'	1 – 2 / 1' – 2'	*wt*
1%	9%	SNP1' allele a SNP2' / SNP1' allele a SNP2'	1' – 2' / 1' – 2'	disease

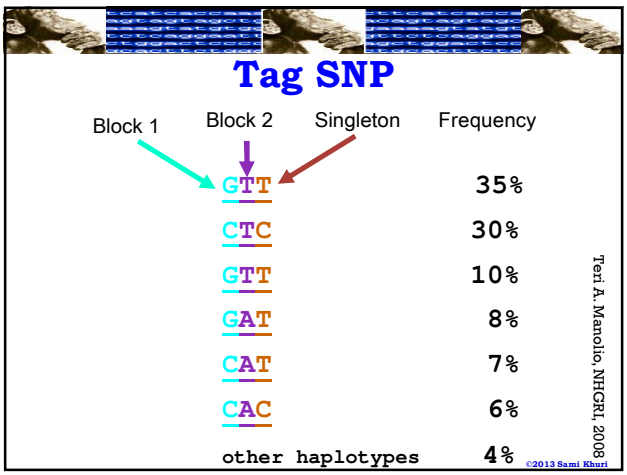
The allele and its flanking SNPs define one locus. The two SNPs: 1' and 2', and the recessive allele *a* are in **linkage disequilibrium**.



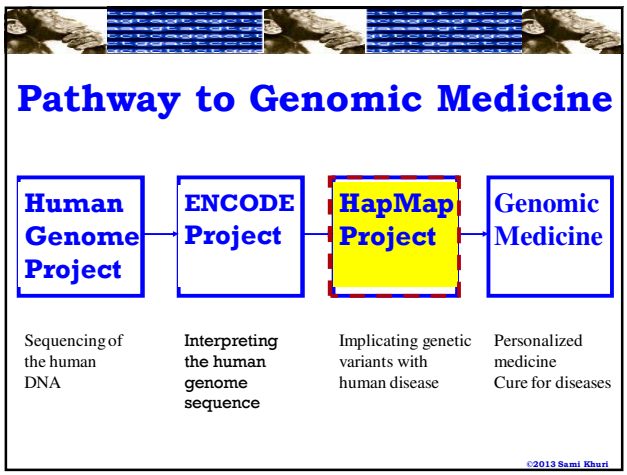




Tag SNP



Pathway to Genomic Medicine



HapMap Project

- Systematic effort to try to catalogue the common variants that exist across human populations.
- Goal: Implication (Correlation) of genetic variants (SNPs and haplotypes) with human diseases.

International Haplotype Map Project

- The goal of the International Haplotype Map Project is to develop a haplotype map of the human genome.
- The "HapMap" describes common patterns of human DNA sequence variation, and is a key source for researchers to use to find genes affecting health, disease, and responses to drugs, and environmental factors.

[Baxevanis & Ouellette, 2005]

Population of HapMap

- **International HapMap Project** analyzes DNA from populations with African, Asian, & European ancestry.
- The DNA samples came from a total of 270 people.
  - **Nigeria**: 30 sets of samples from two parents and an adult child (each such set is called a trio) from Yoruba people of Ibadan.
  - **Japan**: 45 unrelated individuals from the Tokyo area provided samples.
  - **China**: 45 unrelated individuals from Beijing provided samples.
  - **USA**: 30 trios collected in 1980 from U.S. residents with northern and western European ancestry by CEPH.

Population Descriptors

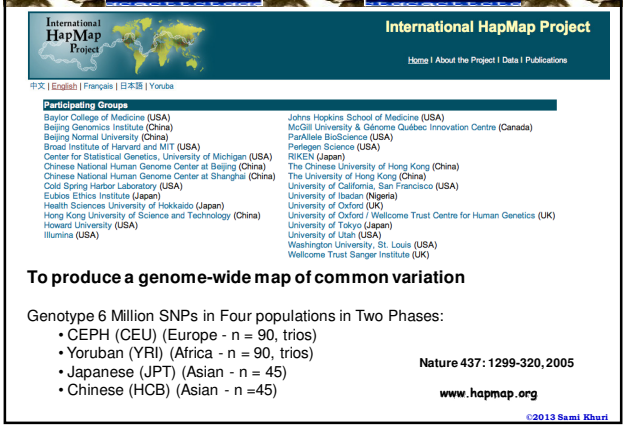
ASW	(A):	African ancestry in Southwest USA
CEU	(C):	Utah residents with Northern and Western European ancestry from the CEPH collection
CHB	(H):	Han Chinese in Beijing, China
CHD	(D):	Chinese in Metropolitan Denver, Colorado
GIH	(G):	Gujarati Indians in Houston, Texas
JPT	(J):	Japanese in Tokyo, Japan
LWK	(L):	Luhya in Webuye, Kenya
MEX	(M):	Mexican ancestry in Los Angeles, California
MKK	(K):	Maasai in Kinyawa, Kenya
TSI	(T):	Toscans in Italy
YRI	(Y):	Yoruba in Ibadan, Nigeria

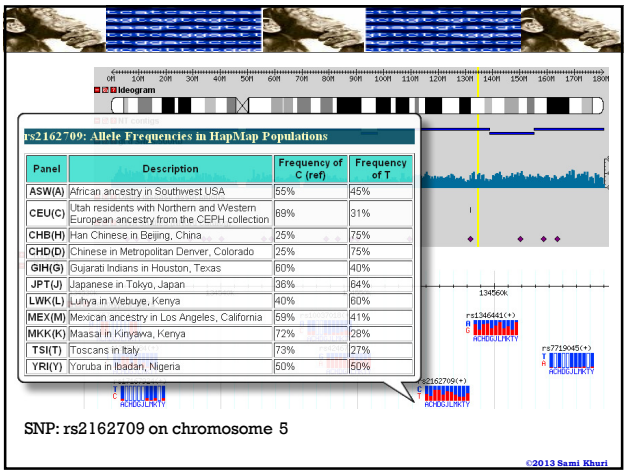
www.hapmap.org/citinghapmap.html

CEPH: Centre d'Etude du Polymorphisme Humain

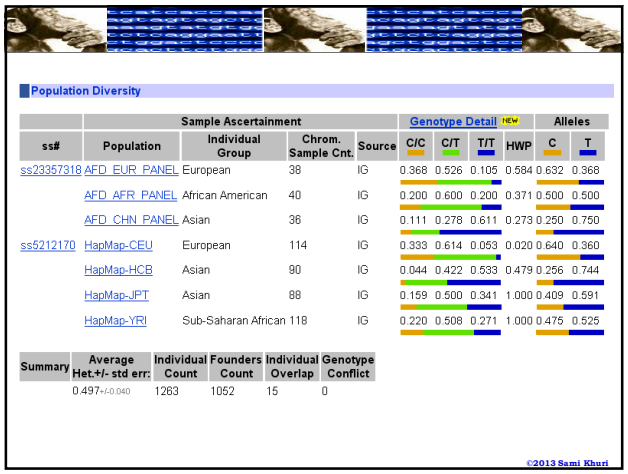
Correlation of Common SNPs

- Sequence data collected by the project confirmed that the vast majority of common SNPs are strongly correlated to one or more nearby proxies: 500,000 SNPs provide excellent power to test over 90% of common SNP variation in out-of-Africa populations, with roughly twice that number required in African populations





SNP: rs2162709 on chromosome 5



Gene Expression Profiling (I)

- Genomics is having a major impact on diagnosis precision and optimization of therapy in cancer.
- Expression profiling and clustering are used to create signatures that are used to guide diagnosis and treatment.

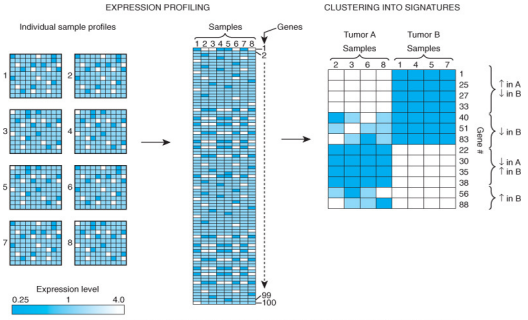
Genetics in Medicine by Nussbaum et al., 2007

Gene Expression Profiling (II)

- We have a number of tissue samples from two different cancers, A and B, and we want to develop a sensitive method to distinguish between these types of tumors in future sets of samples.
- Comparative hybridization can be used.
- Organizing the data and analyzing them to extract key information is very challenging.

Genetics in Medicine by Nussbaum et al., 2007

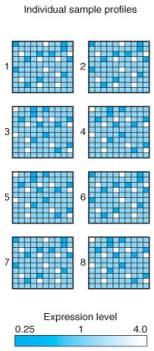
Gene Expression Profiling (III)



© Elsevier. Nussbaum et al: Thompson and Thompson's Genetics in Medicine 7e - www.studentconsult.com

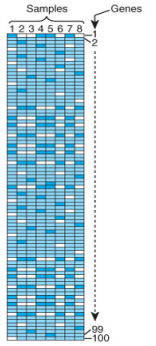
Gene Expression Profiling (IV)

- Eight tissue samples from two different cancers:
  - Cancer A: samples 1 to 4,
  - Cancer B: samples 5 to 8.
- Comparative hybridization used to measure simultaneously the level of mRNA expression in 100 genes in the tissue sample, relative to a standard sample.



800 Expression Values

- Organize the eight sample tissues with 100 gene profile expressions into a Samples/Genes matrix where the Genes are the rows and the Samples are the columns.
- Statistics and Bioinformatics tools are used to organize and analyze the data to extract key information.

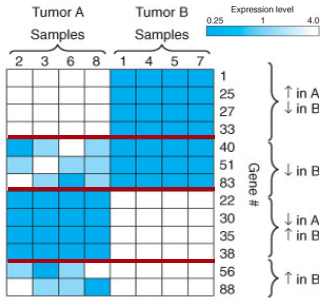


Correlation and Clustering

- Use sophisticated statistical and bioinformatics tools to find groups of genes whose expression seems to **correlate**: move up or down together, between and among the samples.
- Grouping genes by their patterns of expression across samples is termed **clustering**.
- Clusters of genes whose expression correlates with each other and with a particular set of sample constitute an **expression signature** characteristic of those samples.

Clustering into Signatures

- Only 13 genes showed correlation across subsets of samples.
- Applications:
  - Differentiate between tumor A and tumor B
    - Classify unknown tumors as A-like, B-like, neither
  - Use signature to find other tumors.
  - Revelation of previously unsuspected genes connected with same disease.



GWAS: The New Wave

Genome-Wide Association Study

- Method for interrogating all 10 million variable points across human genome.
- Variation is inherited in groups, or blocks, so not all 10 million points have to be tested.
- NIH is interested in advancing **genome-wide association studies** (GWAS) to identify common genetic factors that influence health and disease.

Teri A. Manolio, NHGRI, 2008

GWAS at NIH (I)

- NIH is interested in advancing **genome-wide association studies** (GWAS) to identify common genetic factors that influence health and disease.
- A **genome-wide association study** is defined as any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits (such as blood pressure or weight), or the presence or absence of a disease or condition.

GWAS at NIH (II)

- Whole genome information, when combined with clinical and other phenotype data, offers the potential for increased understanding of basic biological processes affecting human health, improvement in the prediction of disease and patient care, and ultimately the realization of the promise of **personalized medicine**.

GWAS at NIH (III)

- Rapid advances in understanding the patterns of human genetic variation and maturing high-throughput, cost-effective methods for genotyping are providing powerful research tools for identifying **genetic variants** that contribute to **health** and **disease**.

Testing 10 Million SNPs?

- Would Genome-Wide Association Studies require directly testing each of the nearly 10 million common variants for association to disease?
  - In other words, if only 5% of variants were tested, would 95% of associations be missed?
- Or could a subset serve as reliable proxies for their neighbors?

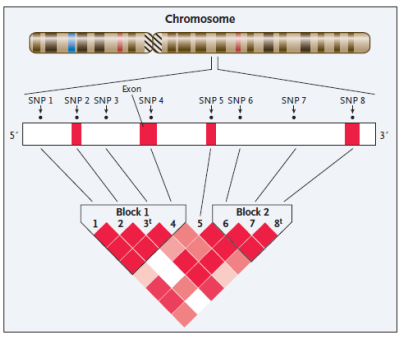
Genetic Mapping in Human Disease, Altshuler et al., 2008

Low Recombination Rates

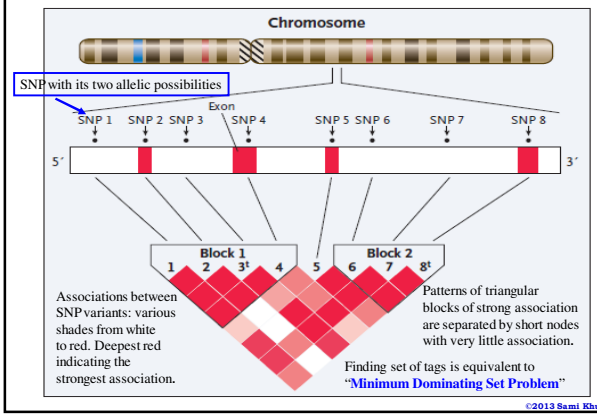
- Each disease-causing mutation arises on a particular copy of the human genome and bears a specific set of common alleles in cis at nearby loci, termed a haplotype.
- Because the recombination rate is low (about 1 crossover per 100 megabases (Mb) per generation), disease alleles in the population typically show association with nearby marker alleles for many generations, a phenomenon termed **linkage disequilibrium** (**LD**)

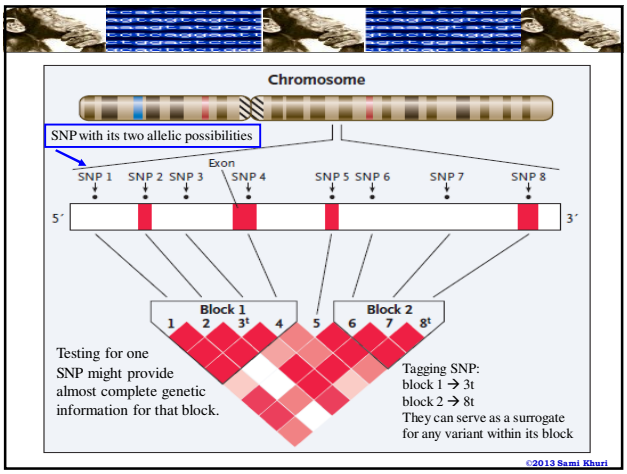
Genetic Mapping in Human Disease, Altshuler et al., 2008

Mapping Relationships Among SNPs



Christensen and Murray, 2007





Cases and Controls

- For association studies, a group of individuals are selected as "cases" and another as "controls"
- **Cases**: individuals that
  - are diagnosed with some disease
  - react to some type of medicine (allergic to penicillin)
  - especially healthy (good health & over 100 years old)
- **Controls**: individuals who do not exhibit the feature selected for the case group.

[Baxevanis & Ouellette, 2005]

©2013 Sami Khuri

Myocardial Infarction and rs1333049

Association of **alleles** of rs1333049 with Myocardial Infarction

	C N (%)	G N (%)	χ^2 (1df)	P-value
Cases	2,132 (55.4)	1,716 (44.6)	55.1	1.2×10^{-13}
Controls	2,783 (47.4)	3,089 (52.6)		

Allelic Odds Ratio = 1.38

Association of **genotypes** of rs1333049 with Myocardial Infarction

	CC N (%)	CG N (%)	GG N (%)	χ^2 (2df)	P-value
Cases	586 (30.5)	960 (49.9)	378 (19.6)	59.7	1.1×10^{-14}
Controls	676 (23.0)	1,431 (48.7)	829 (28.2)		

Heterozygote Odds Ratio = 1.47
Homozygote Odds Ratio = 1.90

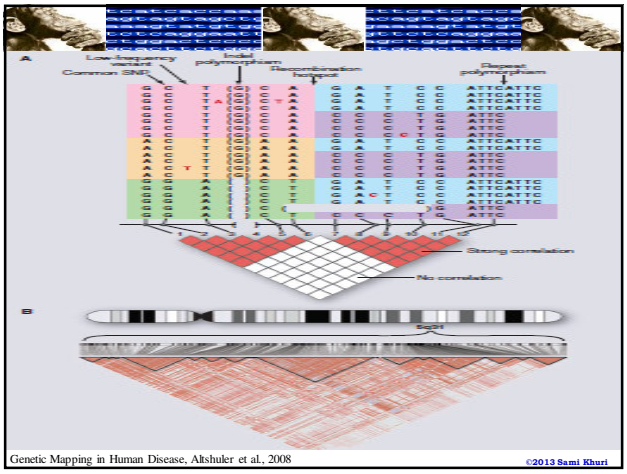
Genome-wide Association Analysis of Coronary Artery Disease, by Samani et al, *NEJM* 2007

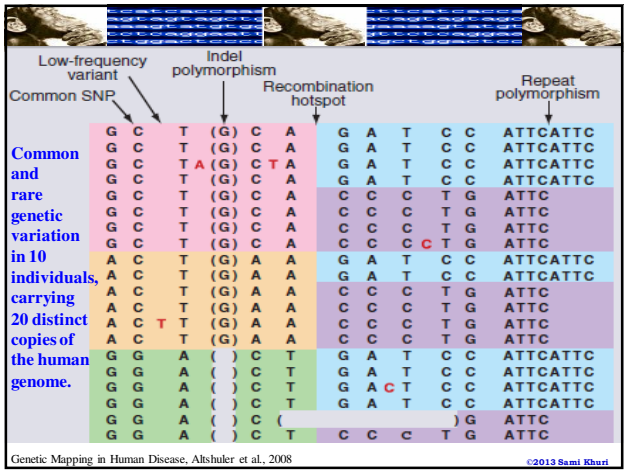
©2013 Sami Khuri

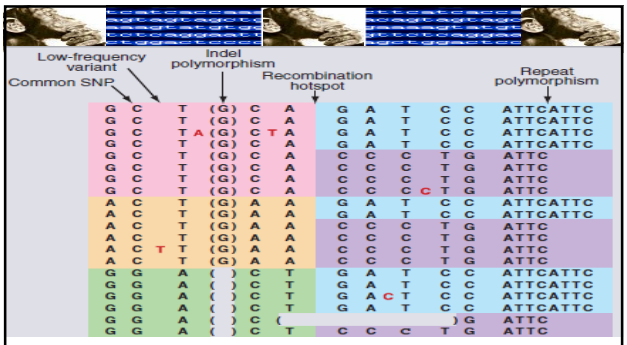
Common Disease Common Variant Hypothesis

- It is believed that genetic variations with alleles that are common in the population will explain much of the heritability of common diseases.
- These studies were made possible by
  - the sequencing of the human genome (International Human Genome Sequencing Consortium, 2004) and
  - the completion of the subsequent human haplotype mapping (HapMap) project.

©2013 Sami Khuri

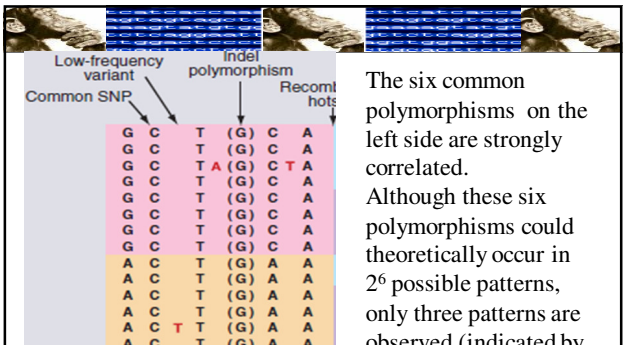






The amount of variation shown here is typical for a 5-kb stretch of genome and is centered on a strong recombination hotspot.
The 12 common variations include 10 SNPs, an insertion-deletion polymorphism (indel), and a tetranucleotide repeat polymorphism.

Genetic Mapping in Human Disease, Altshuler et al., 2008

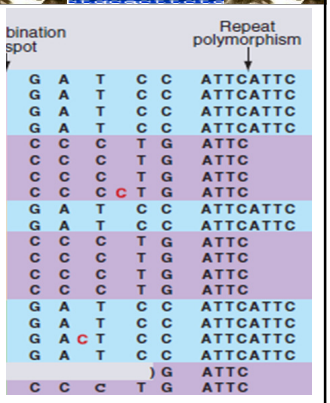


The six common polymorphisms on the left side are strongly correlated.
Although these six polymorphisms could theoretically occur in 2^6 possible patterns, only three patterns are observed (indicated by pink, orange, and green). These patterns are called haplotypes.

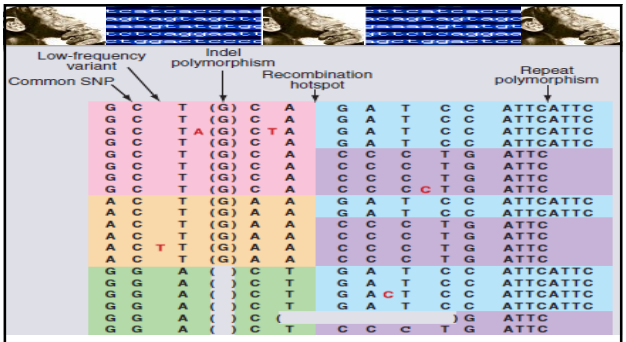
Genetic Mapping in Human Disease, Altshuler et al., 2008

Similarly, the six common polymorphisms on the right side are strongly correlated and reside on only two haplotypes.

The haplotypes are indicated by blue and purple.

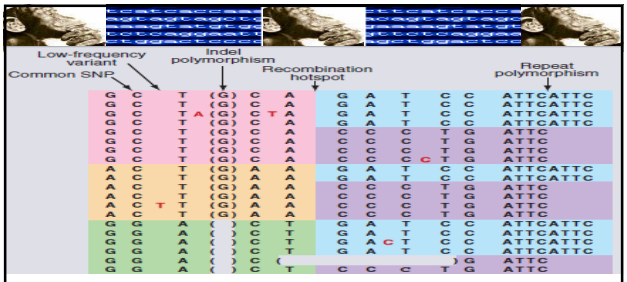


Genetic Mapping in Human Disease, Altshuler et al., 2008



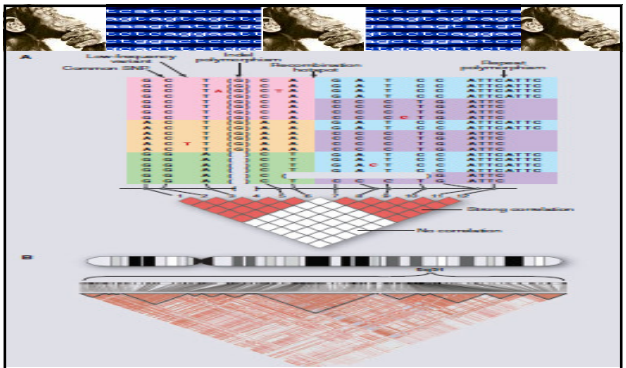
The haplotypes occur because there has not been much genetic recombination between the sites. By contrast, there is little correlation between the two groups of polymorphisms, because a hotspot of genetic recombination lies between them.

Genetic Mapping in Human Disease, Altshuler et al., 2008



In addition to the common polymorphisms, lower-frequency polymorphisms also occur in the human genome. Five rare SNPs are shown, with the variant nucleotide marked in red and the reference nucleotide not shown. In addition, on the second to last chromosome, a larger deletion variant is observed that removes several kilobases of DNA. Such larger deletion or duplicatione vents (i.e., CNVs) may be common and segregate as other DNA variants.

Genetic Mapping in Human Disease, Altshuler et al., 2008



The pairwise correlation between the common sites is shown by the red and white boxes below, with red indicating strong correlation and white indicating weak correlation.

Genetic Mapping in Human Disease, Altshuler et al., 2008

MDECODE

- **M**olecular **D**iversity and **E**pidemiology of **Co**mmon **D**isease (**MDECODE**) is a multidisciplinary and multinational project created to gain a greater understanding of the type and amount of human DNA sequence variation, its history, and the relationship of its contemporary organization to the continuous distribution of measures of human health among individuals in the population at large (such as blood pressure or plasma cholesterol levels).

http://droog.mbt.washington.edu/mdecode

Selecting Potential Markers

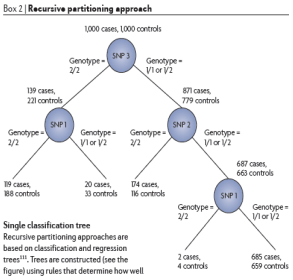
- For cases-control studies, a selection of SNPs is genotyped in both the case and the control groups, and those alleles that exhibit a higher incidence in the case group as opposed to the control group are potential markers for the observed phenotype.

[Baxevanis & Ouellette, 2005]

GWAS → Improved Health?

- Use of genetic information regarding common disease to individualize providers' approach to patients and change patients' behaviors in ways that lead to improved health ("Personalized Medicine").
- Use of genetic information regarding common disease to understand the biology of human disease to lead to improved diagnostic, therapeutic, and preventive approaches.

Methods Used for SNP Analysis

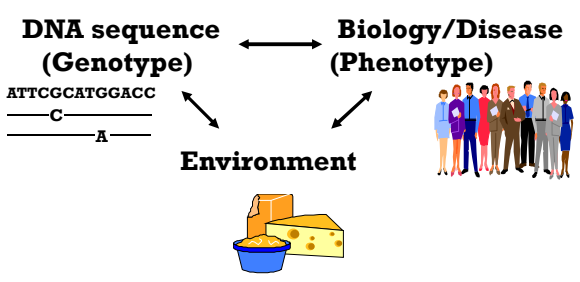


- Regression Models
- Exhaustive Searches
  - Two-Locus Interaction
  - Higher-Order Interaction
- Recursive Partitioning Approach
- Random Forest Approach
- Multifactor Dimensionality Reduction
- Bayesian model selection

Detecting gene-gene interactions that underlie human diseases" by Cordell, Nature, June 2009

The Next Challenge

Understanding the link between



Genetic Testing for the Public

23andMe

Navigenics™

deCODEme

23andme.com
decodeme.com
navigenics.com

Genetic Testing: What you Need to Know

1. Do I need genetic testing?
2. Who can help me choose the right test?
3. What information will I need to give to the laboratory?
4. What will the results tell me?
5. Will my results be confidential?
6. How can I learn more?

Center for Disease Control and Prevention
Office of Surveillance, Epidemiology, and Laboratory Services

Pathway to Genomic Medicine

Human Genome Project	ENCODE Project	HapMap Project	Genomic Medicine
Sequencing of the human DNA	Interpreting the human genome sequence	Implicating genetic variants with human disease	Personalized medicine Cure for diseases

Pharmacogenomics (I)

- Why the SNP Frenzy?
  Answer: **Pharmacogenomics**.
- **Pharmacogenomics** is the branch of pharmacology which deals with the influence of genetic variation on drug response in patients by correlating gene expression or single-nucleotide polymorphisms with a drug's efficacy or toxicity.

www.wikipedia.com

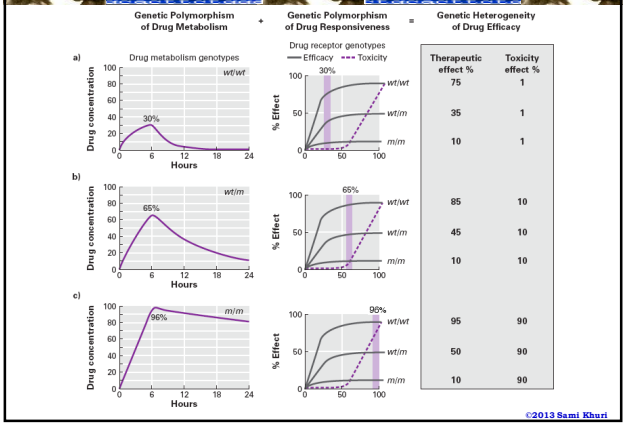
Pharmacogenomics (II)

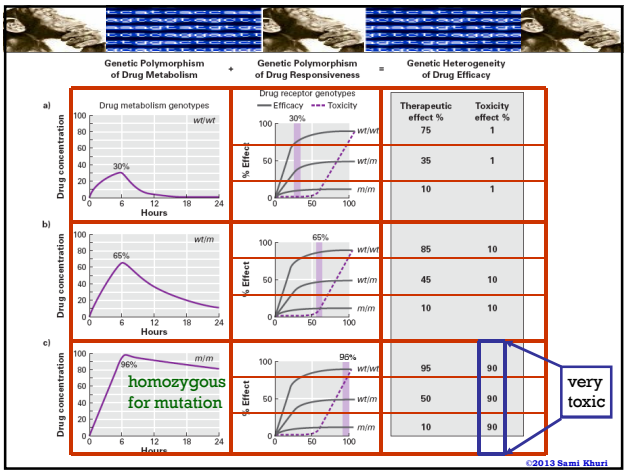
- By doing so, **pharmacogenomics** aims to develop rational means to optimize drug therapy, with respect to the patients' genotype, to ensure maximum efficacy with minimal adverse effects.
- Such approaches promise the advent of "personalized medicine"; in which drugs and drug combinations are optimized for each individual's unique genetic makeup.

www.wikipedia.com

W. Evans and M. Relling

- Evans and Relling considered the efficacy and toxicity of a drug that requires two genes:
  - An activator with 2 alleles, and
  - A binding site with 2 alleles.
- There are 9 possible genotypes.
- Therapeutic effects depend on the genotype of the drug receptors in combination with the amount of active drug in circulation.





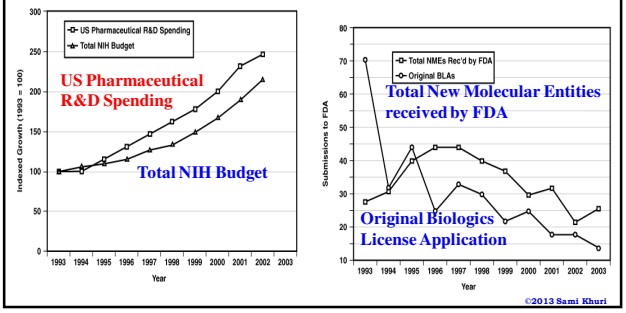
Therapeutic Effects of Drugs

- The example highlights the complex web of protein interactions that pharmacogenomics hopes to decipher.
- Drug response is polygenic, and new technologies are needed to understand the connections between relevant proteins involved in drug responses.
- If genotype-specific medication becomes viable, the physician will need to know the genotype of the ill person to determine the appropriate medication and dosage for optimal therapy.

©2013 Sami Khuri

1993-2003: Spending vs. New Drugs

Similar trends have been observed at regulatory agencies worldwide.



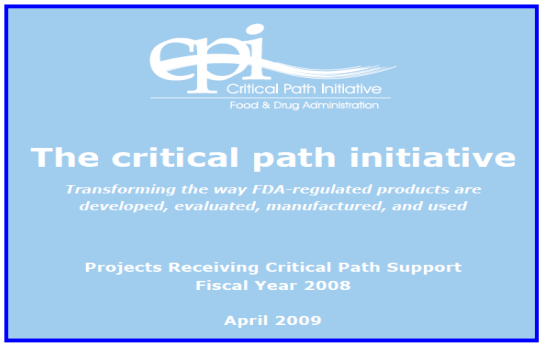
©2013 Sami Khuri

CPI: Introduction

- The Critical Path Initiative (CPI) is the Food and Drug Administration's national strategy for **modernizing the sciences** through which FDA-regulated products are developed, evaluated, manufactured, and used.
- The Initiative was launched in March 2004 to address the steep decline in the number of innovative medical products submitted for approval, despite the enormous breakthroughs being made in **biomedical science**.

©2013 Sami Khuri

The Critical Path Initiative

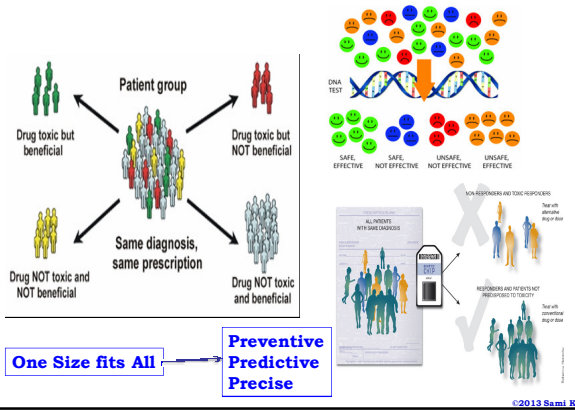


www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative
©2013 Sami Khuri

Pharmacogenomics

- **Pharmacogenomics** deals with the influence of genetic variation on drug response in patients by correlating gene expression or SNPs with a drug's efficacy or toxicity.
- **Pharmacogenomics** aims to optimize drug therapy, with respect to the patients' genotype, to ensure maximum efficacy with minimal adverse effects.
- Such approaches promise the advent of "personalized medicine" in which drugs and drug combinations are optimized for each individual's unique genetic makeup.

www.wikipedia.com

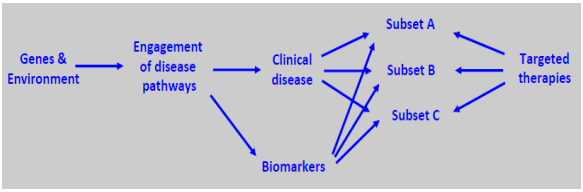


One Size fits All → **Preventive Predictive Precise**

SNPs as Biomarkers

- A lot of effort has been focused on discovering SNPs that are in the proximity of genes.
- The hope is that identifying such SNPs will lead to the diagnosis and treatment of more diseases more effectively.
- However, this task is rendered more problematic by the realization that drug effectiveness is hampered by genomic variations.

Biomarkers and Human Disease

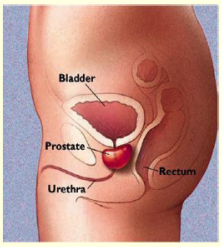


- Improve clinical trial design
- Identify disease subsets
- Guide disease selection for new therapies

Prostate Cancer Diagnosis

Risk factors 2004

- Age
- Race
- Family history



Risk factors 2008

- Age
- Race
- Family history
- Locus 1
- Locus 2
- .....
- .....
- Locus 16

cancercontrol.cancer.gov/od/phg/presentations/Xu.pdf

GWAS and Prostate Cancer

Prostate cancer risk associated variants identified from GWAS

SNPs	Chr	Position	Allele frequency Cases	Allele frequency Controls	OR (95% CI)	P
rs2660753	3p12	87,193,364	0.10	0.08	1.32 (1.13-1.54)	3.4E-04
rs9364554	6q25	160,804,075	0.33	0.31	1.12 (1.02-1.22)	0.02
rs10486567	7p15	27,749,803	0.78	0.76	1.12 (1.01-1.24)	0.03
rs6465657	7q21	97,654,263	0.51	0.47	1.16 (1.06-1.26)	6.7E-04
rs16901979	8q24 (2)	128,194,098	0.06	0.03	1.66 (1.34-2.07)	3.1E-06
rs6983267	8q24 (3)	128,482,487	0.56	0.51	1.22 (1.12-1.33)	3.6E-06
rs1447295	8q24 (1)	128,554,220	0.17	0.14	1.21 (1.08-1.36)	1.6E-03
rs10993994	10q11	51,219,502	0.43	0.39	1.15 (1.05-1.25)	1.6E-03
rs10896449	11q13	68,751,243	0.49	0.46	1.14 (1.05-1.25)	2.1E-03
rs4430796	17q12	33,172,153	0.61	0.56	1.24 (1.14-1.35)	8.5E-07
rs1859962	17q24.3	66,620,348	0.54	0.50	1.17 (1.08-1.28)	2.0E-04
rs5945619	Xp11	51,074,708	0.42	0.38	1.20 (1.06-1.36)	3.5E-03

cancercontrol.cancer.gov/od/phg/presentations/Xu.pdf

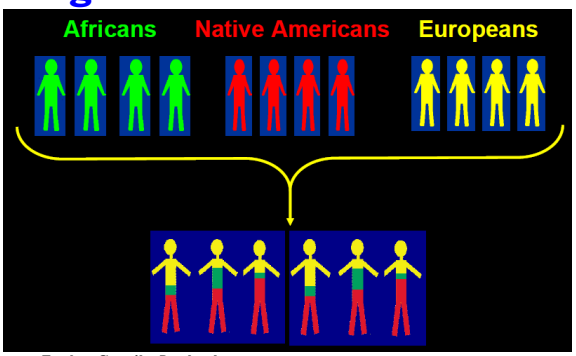
Tumors and Biomarkers

TUMOR TYPE INVESTIGATED	BIOMARKER
Breast	FGFR[2]
	EGFR[2,3]
Breast, Prostate	p53[2,4]
Colorectal	APC[5]
Colorectal, Lung	PIK3[5,3]
	BRAF[5,3]
Colorectal, Prostate	PTEN[5,4]
Lung	TPA (tissue polypeptide antigen)[8]
	TCF21[7]
	c-MET[6]
	CEA[8]
Lung, Prostate	HER-2[3,4]
Prostate	CD44[4]
	PI3K/AKT/mTOR[4]
	PSA[4]
	PSAP[4]
	PSMA[4]
	NKX3.1[4]
	EZH2[4]

Personalized medicine uses information about a person's genes, proteins, and cellular environment to prevent, diagnose, and treat disease

www.canceritspersonal.com

Origins of African Americans



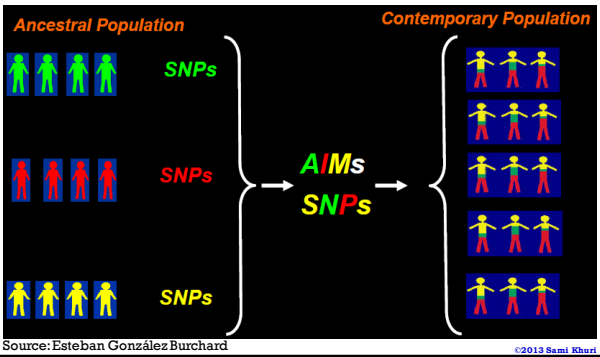
Source: Esteban González Burchard

Ancestry Informative Marker

- An **Ancestry-Informative Marker** (AIM) is a set of polymorphisms for a locus which exhibits substantially different frequencies between populations from different geographical regions.
- By using a number of **AIM**s one can estimate the geographical origins of the ancestors of an individual and ascertain what proportion of ancestry is derived from each geographical region.

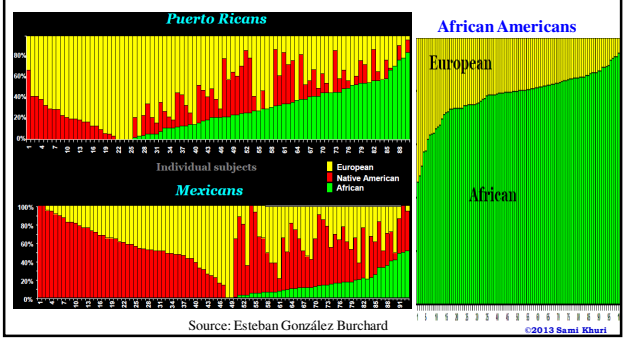
en.wikipedia.org/wiki/

SNPs and AIMs



Source: Esteban González Burchard

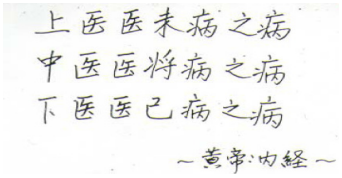
Origins of Latinos & African Amer.



Source: Esteban González Burchard

Self-Identified Race: Genetic Ancestry

The Superior Doctor

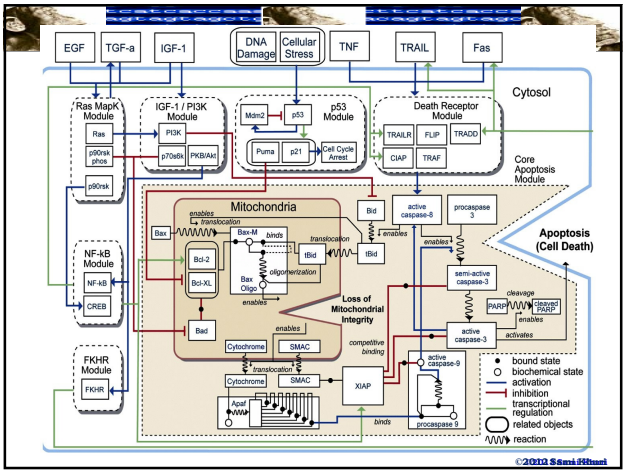


Superior doctors prevent the disease
Mediocre doctors treat the disease before evident
Inferior doctors treat the full blown disease

-Huang Dee: Nai - Ching
(2600 B.C. 1st Chinese Medical Text)

Preventive Medicine

- Prevent disease from occurring
- Identify the cause of the disease
- Treat the cause of the disease rather than the symptoms
- Genomics identifies the cause of disease
- "All medicine may become pediatrics" Paul Wise
- Effects of environment, accidents, aging, penetrance …
- Health care costs can be greatly reduced if
  - invests in preventive medicine
  - one targets the cause of disease rather than symptoms



Anatomy Lesson of Dr. Nicolaes Tulp

1632 oil painting by Rembrandt Harmenszoon van Rijn

If Rembrandt were Around Today

Source: Carlos Cordon-Cardo, Columbia University

Concluding Remarks (I)

- Every human being carries many SNPs most human genes are located near at least one SNP.
- A lot of effort has been focused on discovering such SNPs.
- The hope is that identifying such SNPs will lead to the diagnosis and treatment of more diseases more effectively.
- However, this task is rendered more problematic by the realization that drug effectiveness is hampered by genomic variations.

Concluding Remarks (II)

- Biology is becoming an information science
- Progression: **in vivo** to **in vitro** to **in silico**
- Personalized medicine remains a research concept – it is not yet ready for clinical practice!
- We are here to add what we can *to*, not to get what we can *from*, Life. William Osler
- We look to a future in which medicine will be predictive, preventive, preemptive and personalized.