SNP Resources: Finding SNPs Discovery and Databases

Mark J. Rieder, PhD

NIEHS Variation Workshop
January 30-31, 2006

SNP Resources: SNP discovery and cataloging

1. SNP discovery/genotyping: Genome-wide approaches
   - ✓ SNP Consortium
   - ✓ HapMap

2. The current state of SNP resources

3. Comprehensive SNP discovery
   NIEHS SNPs - Environmental Genome Project

SNP Databases - **"How to" Manual for finding SNPs**
In class - Tutorial

Genetic Markers: Overview

1. RFLPs (SNPs circa 1980)

2. Microsatellites (SSLP; di-, tri-, tetranucleotide repeats)
   - 1/50,000 bp
   - Linkage Studies - 300-400 markers (~1 Mbp)
   - Multi-allelic/High heterozygosity/informative
   - Complex genotyping assays

3. Single Nucleotide Polymorphisms (SNPs)
   - Most frequent genetic variant (base substitutions)
   - 1/1000 bp (comparing randomly selected chromosomes)
   - Biallelic/less informative
   - Simplified genotyping platforms (+/- calling)

Development of a genome-wide SNP map: How many SNPs?

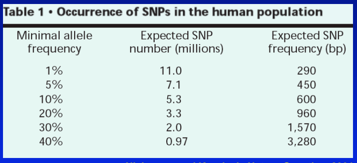


Table 1 • Occurrence of SNPs in the human population

Minimal allele frequency	Expected SNP number (millions)	Expected SNP frequency (bp)
1%	11.0	290
5%	7.1	450
10%	5.3	600
20%	3.3	960
30%	2.0	1,570
40%	0.97	3,280

Nickerson and Kruglyak, Nature Genetics, 2001

~ 10 million common SNPs (> 1- 5% MAF) - 1/300 bp

How has SNP discovery progressed toward this goal?

Finding SNPs: Marker Discovery and Methods

SNP discovery has proceeded in two distinct phases:

1 - SNP Identification
Define the alleles
Map this to a unique place in the genome

2 - SNP Characterization
Determination of the genotype in many individuals
Population frequency of SNPs

Finding SNPs: Marker Discovery and Methods

SNP Discovery has proceeded in two distinct phases:

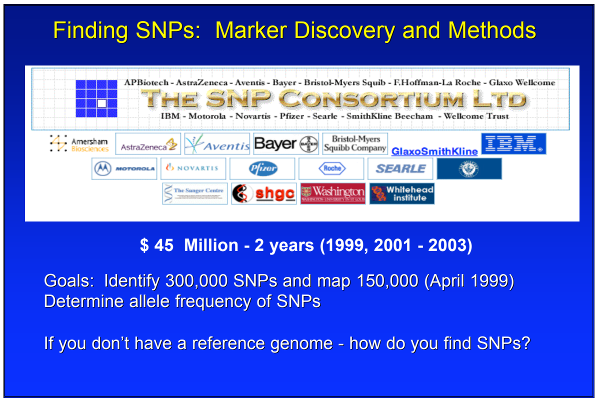
1 - SNP Discovery**/Characterization

APBiotech - AstraZeneca - Aventis - Bayer - Bristol-Myers Squib - F.Hoffman-La Roche - Glaxo Wellcome
THE SNP CONSORTIUM LTD
IBM - Motorola - Novartis - Pfizer - Searle - SmithKline Beecham - Wellcome Trust

2 - SNP Discovery/Characterization**

International HapMap Project

International
HapMap
Project
Home I About the Project I Data I Publications
中文 | English | Français | 日本語 | Yoruba



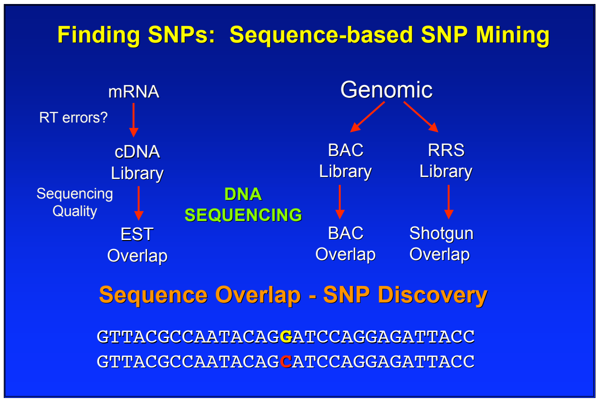
Finding SNPs: Marker Discovery and Methods

APBiotech - AstraZeneca - Aventis - Bayer - Bristol-Myers Squib - F.Hoffman-La Roche - Glaxo Wellcome
THE SNP CONSORTIUM LTD
IBM - Motorola - Novartis - Pfizer - Searle - SmithKline Beecham - Wellcome Trust

$ 45 Million - 2 years (1999, 2001 - 2003)

Goals: Identify 300,000 SNPs and map 150,000 (April 1999)
Determine allele frequency of SNPs

If you don't have a reference genome - how do you find SNPs?



Finding SNPs: Sequence-based SNP Mining

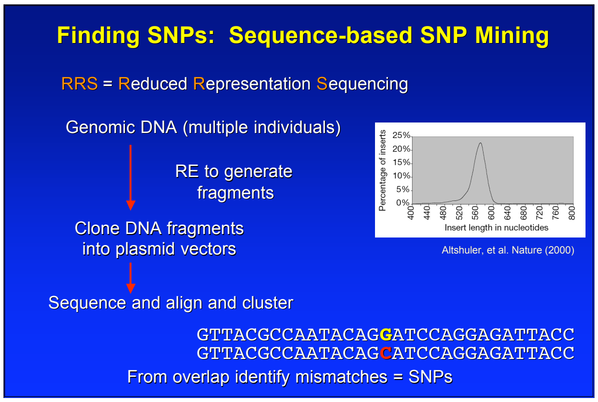
mRNA
RT errors?
cDNA Library
Sequencing Quality
EST Overlap

DNA SEQUENCING

Genomic
BAC Library
RRS Library
BAC Overlap
Shotgun Overlap

Sequence Overlap - SNP Discovery

```
GTTACGCCAATACAGGATCCAGGAGATTACC
GTTACGCCAATACAGCATCCAGGAGATTACC
```



Finding SNPs: Sequence-based SNP Mining

RRS = Reduced Representation Sequencing

Genomic DNA (multiple individuals)

RE to generate fragments

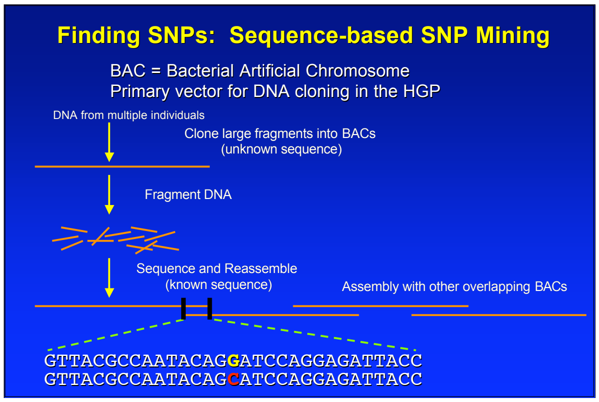
Clone DNA fragments into plasmid vectors

Sequence and align and cluster

Altshuler, et al. Nature (2000)

```
GTTACGCCAATACAGGATCCAGGAGATTACC
GTTACGCCAATACAGCATCCAGGAGATTACC
```

From overlap identify mismatches = SNPs



Finding SNPs: Sequence-based SNP Mining

BAC = Bacterial Artificial Chromosome
Primary vector for DNA cloning in the HGP

DNA from multiple individuals

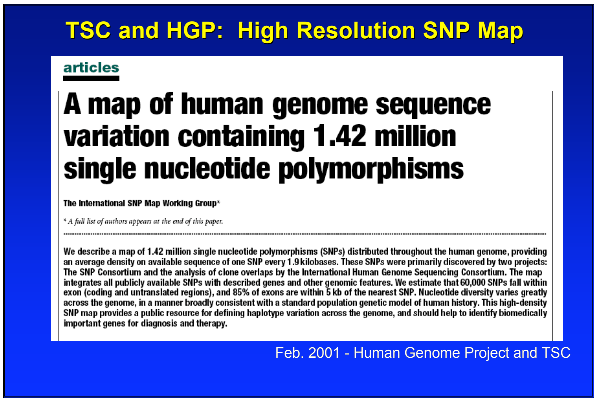
Clone large fragments into BACs
(unknown sequence)

Fragment DNA

Sequence and Reassemble
(known sequence)

Assembly with other overlapping BACs

```
GTTACGCCAATACAGGATCCAGGAGATTACC
GTTACGCCAATACAGCATCCAGGAGATTACC
```



TSC and HGP: High Resolution SNP Map

articles

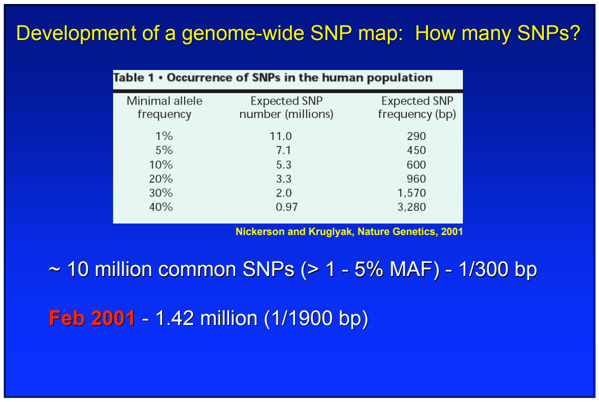
A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms

The International SNP Map Working Group*

*A full list of authors appears at the end of this paper.

We describe a map of 1.42 million single nucleotide polymorphisms (SNPs) distributed throughout the human genome, providing an average density on available sequence of one SNP every 1.9 kilobases. These SNPs were primarily discovered by two projects: The SNP Consortium and the analysis of clone overlaps by the International Human Genome Sequencing Consortium. The map integrates all publicly available SNPs with described genes and other genomic features. We estimate that 60,000 SNPs fall within exon (coding and untranslated regions), and 85% of exons are within 5 kb of the nearest SNP. Nucleotide diversity varies greatly across the genome, in a manner broadly consistent with a standard population genetic model of human history. This high-density SNP map provides a public resource for defining haplotype variation across the genome, and should help to identify biomedically important genes for diagnosis and therapy.

Feb. 2001 - Human Genome Project and TSC



Development of a genome-wide SNP map: How many SNPs?

Table 1 • Occurrence of SNPs in the human population

Minimal allele frequency	Expected SNP number (millions)	Expected SNP frequency (bp)
1%	11.0	290
5%	7.1	450
10%	5.3	600
20%	3.3	960
30%	2.0	1,570
40%	0.97	3,280

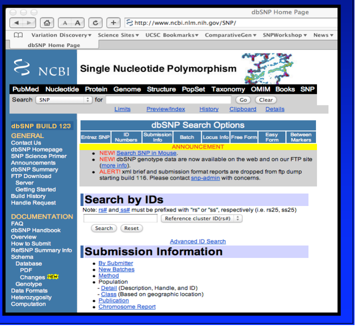
Nickerson and Kruglyak, Nature Genetics, 2001

~ 10 million common SNPs (> 1 - 5% MAF) - 1/300 bp

Feb 2001 - 1.42 million (1/1900 bp)

SNP Discovery: dbSNP database

dbSNP
-NCBI SNP database


dbSNP Home Page
http://www.ncbi.nlm.nih.gov/SNP/
Variation Discovery ▾ Science Sites ▾ UCSC Bookmarks ▾ ComparativeGen ▾ SNPWorkshop ▾ News ▾
dbSNP Home Page
NCBI
PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP
Search SNP for
Go Clear
Limits Preview/Index History Clipboard Details

Single Nucleotide Polymorphism

dbSNP BUILD 123

GENERAL
Contact Us
dbSNP Homepage
SNP Science Primer
Announcements
dbSNP Summary
FTP Download
Server
Getting Started
Build History
Handle Request

DOCUMENTATION
FAQ
dbSNP Handbook
Overview
How to Submit
RefSNP Summary Info
Schema
Database
PDF
Changes NEW!
Genotype
Data Formats
Heterozygosity
Computation

dbSNP Search Options

Entrez SNP	ID Numbers	Submission Info	Batch	Locus Info	Free Form	Easy Form	Between Markers

ANNOUNCEMENT

- NEW! Search SNP in Mouse.
- NEW! dbSNP genotype data are now available on the web and on our FTP site (more info).
- ALERT! xml brief and submission format reports are dropped from ftp dump starting build 116. Please contact snp-admin with concerns.

Search by IDs

Note: rs# and ss# must be prefixed with "rs" or "ss", respectively (i.e. rs25, ss25)

Reference cluster ID(rs#)

Search Reset

Advanced ID Search

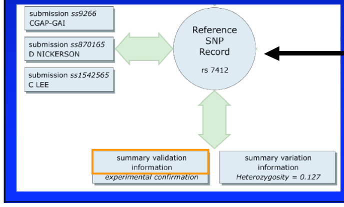
Submission Information

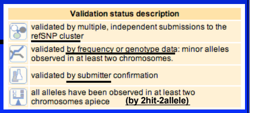
- By Submitter
- New Batches
- Method
- Population
  - Detail (Description, Handle, and ID)
  - Class (Based on geographic location)
- Publication
- Chromosome Report

SNP data submitted to dbSNP: Clustering

dbSNP processing of SNPs

SNPs submitted
By research communtiy
(submitted SNPs = ss#) → Unique mapping
to a genome location
(reference SNP = rs#)



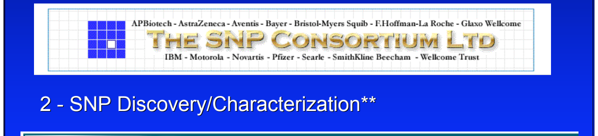


Validation status description
validated by multiple, independent submissions to the refSNP cluster
validated by frequency or genotype data: minor alleles observed in at least two chromosomes.
validated by submitter confirmation
all alleles have been observed in at least two chromosomes apiece **(by 2hit-2allele)**

Finding SNPs: Marker Discovery and Methods

SNP Discovery has proceeded in two distinct phases:

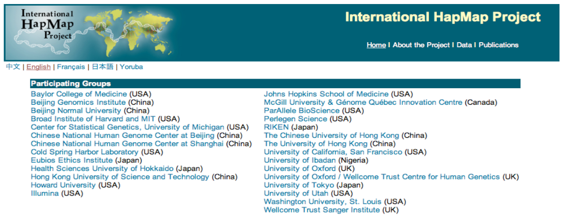
1 - SNP Identification**/Discovery


APBiotech - AstraZeneca - Aventis - Bayer - Bristol-Myers Squib - F.Hoffman-La Roche - Glaxo Wellcome
THE SNP CONSORTIUM LTD
IBM - Motorola - Novartis - Pfizer - Searle - SmithKline Beecham - Wellcome Trust

2 - SNP Discovery/Characterization**

International HapMap Project
International HapMap Project
Home I About the Project I Data I Publications
中文 | English | Français | 日本語 | Yoruba

HapMap Project Proposed: Map more SNPs and genotype


International HapMap Project
International HapMap Project
Home I About the Project I Data I Publications
中文 | English | Français | 日本語 | Yoruba

Participating Groups

Baylor College of Medicine (USA)
Beijing Genomics Institute (China)
Beijing Normal University (China)
Broad Institute of Harvard and MIT (USA)
Center for Statistical Genetics, University of Michigan (USA)
Chinese National Human Genome Center at Beijing (China)
Chinese National Human Genome Center at Shanghai (China)
Cold Spring Harbor Laboratory (USA)
Eubios Ethics Institute (Japan)
Health Sciences University of Hokkaido (Japan)
Hong Kong University of Science and Technology (China)
Howard University (USA)
Illumina (USA)
Johns Hopkins School of Medicine (USA)
McGill University & Génome Québec Innovation Centre (Canada)
ParAllele BioScience (USA)
Perlegen Science (USA)
RIKEN (Japan)
The Chinese University of Hong Kong (China)
The University of Hong Kong (China)
University of California, San Francisco (USA)
University of Ibadan (Nigeria)
University of Oxford (UK)
University of Oxford / Wellcome Trust Centre for Human Genetics (UK)
University of Tokyo (Japan)
University of Utah (USA)
Washington University, St. Louis (USA)
Wellcome Trust Sanger Institute (UK)

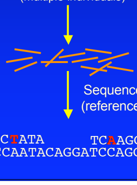
- **Increase SNP density over the first 6 - 12 months**
- **Ultimately produce a fine scale genetic map (HapMap) which would serve as a common resource for all biomedical reseseachers**
- **Genotype 600,000 SNPs genome-wide**
- **Four populations: CEPH (Europe), Yoruban (Africa), Japanese/ Chinese (Asian)**

HapMap SNP Discovery: Prior to Genotyping

Initiation of project planning (July 2001):
2.8 million SNPs (1.4 million validated) - 1/1900 bp
Nov 2003 - 5.7 million (2 million validated) - 1/1500 bp
Feb 2004 - 7.2 million (3.3 million validated) - 1/900 bp

Generate more SNPs:
Random Shotgun Sequencing

Genomic DNA
(multiple individuals)



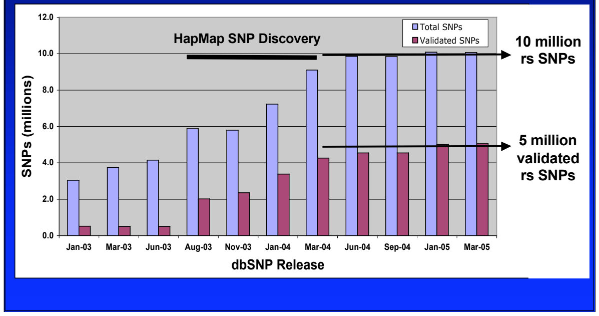
Sequence and align
(reference sequence)

Other Sources of SNPs:
Perlegen (Affymetrix chips) SNP data (chr22)
Sequence chromatograms from Celera project

```
  TACGCCTATA          TCAAGGAGAT
GTTACGCCAATACAGGATCCAGGAGATTACC
```

Draft Human Genome

HapMap Discovery Increased SNP Density and Validated SNPs



10 million rs SNPs

5 million validated rs SNPs

Development of a genome-wide SNP map: How many SNPs?

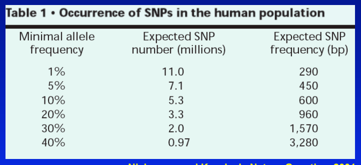


Table 1 • Occurrence of SNPs in the human population

Minimal allele frequency	Expected SNP number (millions)	Expected SNP frequency (bp)
1%	11.0	290
5%	7.1	450
10%	5.3	600
20%	3.3	960
30%	2.0	1,570
40%	0.97	3,280

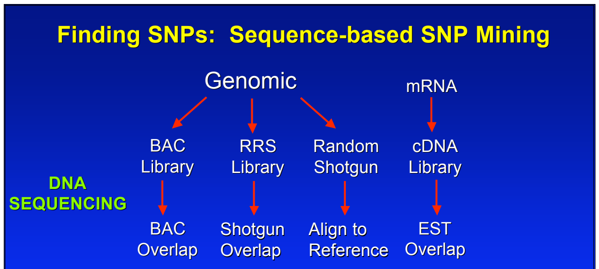
Nickerson and Kruglyak, Nature Genetics, 2001

~ 10 million common SNPs (> 1- 5% MAF) - 1/300 bp

Feb 2001 - 1.42 million (1/1900 bp)
Nov 2003 - 2.0 million (1/1500 bp)
Feb 2004 - 3.3 million (1/900 bp)
Mar 2005 - 5.0 million (validated - 1/600 bp)

When will we have them all?

Finding SNPs: Sequence-based SNP Mining

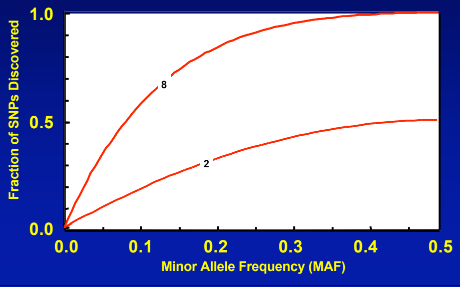


RANDOM Sequence Overlap - SNP Discovery

GTTACGCCAATACAGGATCCAGGAGATTACC
GTTACGCCAATACAGCATCCAGGAGATTACC

SNP discovery is dependent on your sample population size

2 chromosomes { GTTACGCCAATACAGGATCCAGGAGATTACC
GTTACGCCAATACAGCATCCAGGAGATTACC



SNP Characterization/Genotyping

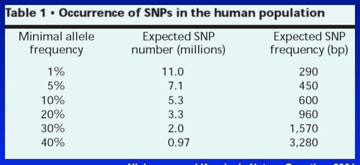


Table 1 • Occurrence of SNPs in the human population

Minimal allele frequency	Expected SNP number (millions)	Expected SNP frequency (bp)
1%	11.0	290
5%	7.1	450
10%	5.3	600
20%	3.3	960
30%	2.0	1,570
40%	0.97	3,280

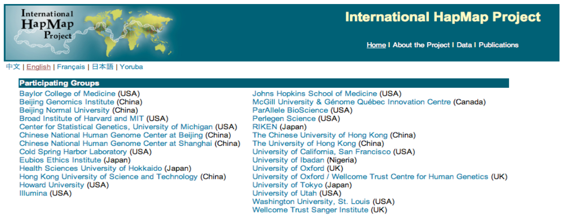
Nickerson and Kruglyak, Nature Genetics, 2001

~ 10 million common SNPs (>1- 5% MAF) - 1/300 bp

Mar 2005 - 5.0 million (validated/mapped - 1/600 bp)

5.0/10.0 = 50% of all common SNPs (validated)!

HapMap Project Proposed: Map more SNPs and genotype



International HapMap Project

Home I About the Project I Data I Publications

中文 | English | Français | 日本語 | Yoruba

Participating Groups

Baylor College of Medicine (USA)
Beijing Genomics Institute (China)
Beijing Normal University (China)
Broad Institute of Harvard and MIT (USA)
Center for Statistical Genetics, University of Michigan (USA)
Chinese National Human Genome Center at Beijing (China)
Chinese National Human Genome Center at Shanghai (China)
Cold Spring Harbor Laboratory (USA)
Eubios Ethics Institute (Japan)
Health Sciences University of Hokkaido (Japan)
Hong Kong University of Science and Technology (China)
Howard University (USA)
Illumina (USA)

Johns Hopkins School of Medicine (USA)
McGill University & Génome Québec Innovation Centre (Canada)
ParAllele BioScience (USA)
Perlegen Science (USA)
RIKEN (Japan)
The Chinese University of Hong Kong (China)
The University of Hong Kong (China)
University of California, San Francisco (USA)
University of Ibadan (Nigeria)
University of Oxford (UK)
University of Oxford / Wellcome Trust Centre for Human Genetics (UK)
University of Tokyo (Japan)
University of Utah (USA)
Washington University, St. Louis (USA)
Wellcome Trust Sanger Institute (UK)

- **Genotype 600,000 SNPs genome-wide**
- **Four populations:**
  - **CEPH (CEU) (Europe - n = 90, trios)**
  - **Yoruban (YRI) (Africa - n = 90, trios)**
  - **Japanese (JPT) (Asian - n = 45)**
  - **Chinese (HCB) (Asian - n =45)**

Finding SNPs: Genotype Data Adds Value to SNPs HapMap Genotyping

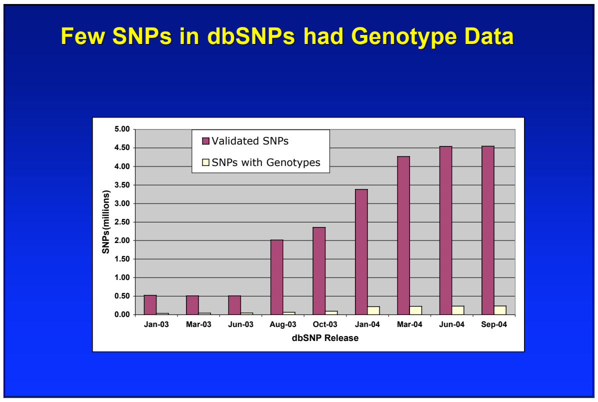
✓ Confirms SNP as “real” and “informative”

✓ Minor Allele Frequency (MAF) - common or rare

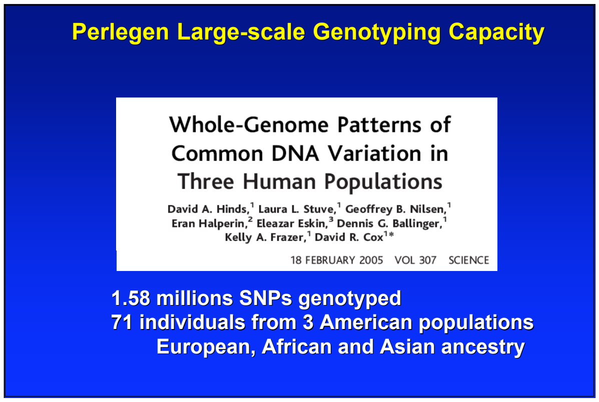
✓ MAF in different populations

✓ Detection of SNP x SNP correlations
(Linkage Disequilibrium)

✓ Determine haplotypes



Few SNPs in dbSNPs had Genotype Data



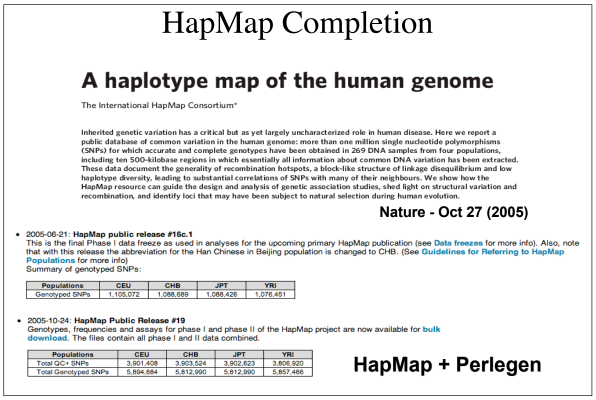
Perlegen Large-scale Genotyping Capacity

Whole-Genome Patterns of Common DNA Variation in Three Human Populations

David A. Hinds,[1] Laura L. Stuve,[1] Geoffrey B. Nilsen,[1] Eran Halperin,[2] Eleazar Eskin,[3] Dennis G. Ballinger,[1] Kelly A. Frazer,[1] David R. Cox[1]*

18 FEBRUARY 2005 VOL 307 SCIENCE

1.58 millions SNPs genotyped
71 individuals from 3 American populations
European, African and Asian ancestry



HapMap Completion

A haplotype map of the human genome

The International HapMap Consortium*

Inherited genetic variation has a critical but as yet largely uncharacterized role in human disease. Here we report a public database of common variation in the human genome: more than one million single nucleotide polymorphisms (SNPs) for which accurate and complete genotypes have been obtained in 269 DNA samples from four populations, including ten 500-kilobase regions in which essentially all information about common DNA variation has been extracted. These data document the generality of recombination hotspots, a block-like structure of linkage disequilibrium and low haplotype diversity, leading to substantial correlations of SNPs with many of their neighbours. We show how the HapMap resource can guide the design and analysis of genetic association studies, shed light on structural variation and recombination, and identify loci that may have been subject to natural selection during human evolution.

Nature - Oct 27 (2005)

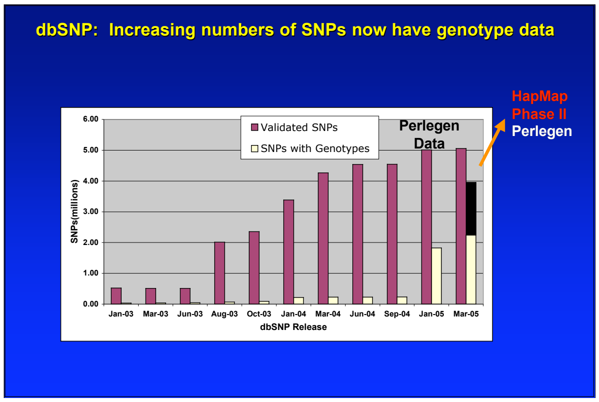
- 2005-06-21: **HapMap public release #16c.1**
  This is the final Phase I data freeze as used in analyses for the upcoming primary HapMap publication (see **Data freezes** for more info). Also, note that with this release the abbreviation for the Han Chinese in Beijing population is changed to CHB. (See **Guidelines for Referring to HapMap Populations** for more info)
  Summary of genotyped SNPs:

Populations	CEU	CHB	JPT	YRI
Genotyped SNPs	1,105,072	1,088,689	1,088,426	1,076,451

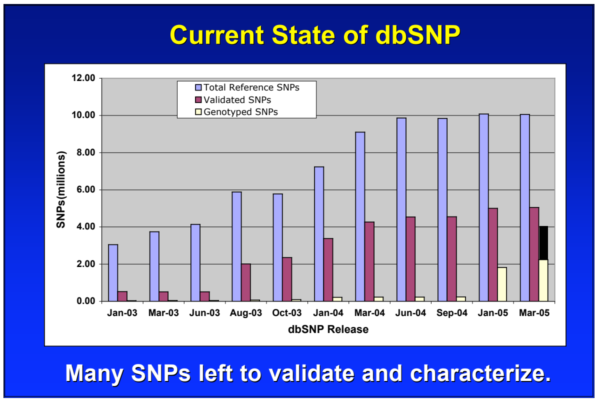
- 2005-10-24: **HapMap Public Release #19**
  Genotypes, frequencies and assays for phase I and phase II of the HapMap project are now available for **bulk download**. The files contain all phase I and II data combined.

Populations	CEU	CHB	JPT	YRI
Total QC+ SNPs	3,901,408	3,903,524	3,902,623	3,806,920
Total Genotyped SNPs	5,894,684	5,812,990	5,812,990	5,857,466

HapMap + Perlegen

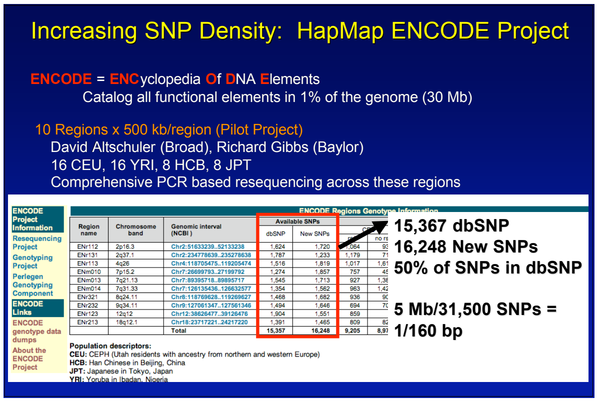


dbSNP: Increasing numbers of SNPs now have genotype data



Current State of dbSNP

Many SNPs left to validate and characterize.



Increasing SNP Density: HapMap ENCODE Project

ENCODE = ENCyclopedia Of DNA Elements
Catalog all functional elements in 1% of the genome (30 Mb)

10 Regions x 500 kb/region (Pilot Project)
David Altschuler (Broad), Richard Gibbs (Baylor)
16 CEU, 16 YRI, 8 HCB, 8 JPT
Comprehensive PCR based resequencing across these regions

ENCODE Project Information
Resequencing Project
Genotyping Project
Perlegen Genotyping Component
ENCODE Links
ENCODE genotype data dumps
About the ENCODE Project

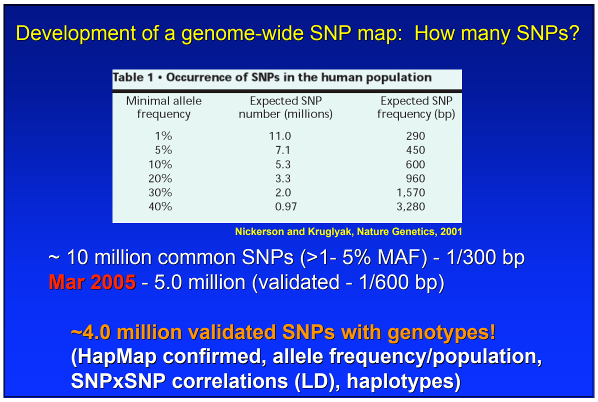
ENCODE Regions Genotype Information

Region name	Chromosome band	Genomic interval (NCBI)	Available SNPs		CE	
			dbSNP	New SNPs	rs	no rs
ENr112	2p16.3	Chr2:51633239..52133238	1,624	1,720	1,064	93
ENr131	2q37.1	Chr2:234778639..235278638	1,787	1,233	1,179	71
ENr113	4q26	Chr4:118705475..119205474	1,516	1,819	1,017	1,61
ENm010	7p15.2	Chr7:26699793..27199792	1,274	1,857	757	45
ENm013	7q21.13	Chr7:89395718..89895717	1,545	1,713	927	1,38
ENm014	7q31.33	Chr7:126135436..126632577	1,354	1,562	963	1,42
ENr321	8q24.11	Chr8:118769628..119269627	1,468	1,682	936	90
ENr232	9q34.11	Chr9:127061347..127561346	1,494	1,646	694	70
ENr123	12q12	Chr12:38626477..39126476	1,904	1,551	859	
ENr213	18q12.1	Chr18:23717221..24217220	1,391	1,465	809	82
		Total	15,357	16,248	9,205	8,97

Population descriptors:
CEU: CEPH (Utah residents with ancestry from northern and western Europe)
HCB: Han Chinese in Beijing, China
JPT: Japanese in Tokyo, Japan
YRI: Yoruba in Ibadan, Nigeria

15,367 dbSNP
16,248 New SNPs
50% of SNPs in dbSNP

5 Mb/31,500 SNPs = 1/160 bp



Development of a genome-wide SNP map: How many SNPs?

Table 1 • Occurrence of SNPs in the human population

Minimal allele frequency	Expected SNP number (millions)	Expected SNP frequency (bp)
1%	11.0	290
5%	7.1	450
10%	5.3	600
20%	3.3	960
30%	2.0	1,570
40%	0.97	3,280

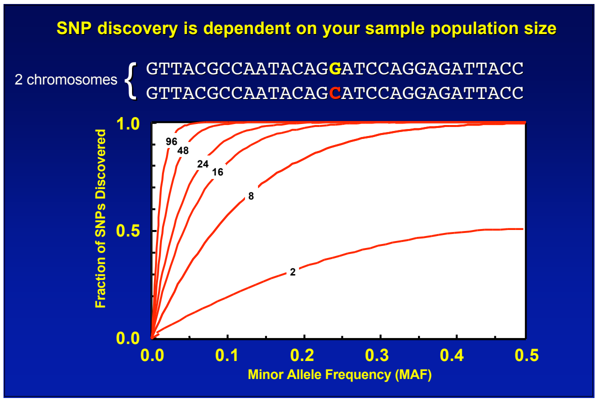
Nickerson and Kruglyak, Nature Genetics, 2001

~ 10 million common SNPs (>1- 5% MAF) - 1/300 bp

Mar 2005 - 5.0 million (validated - 1/600 bp)

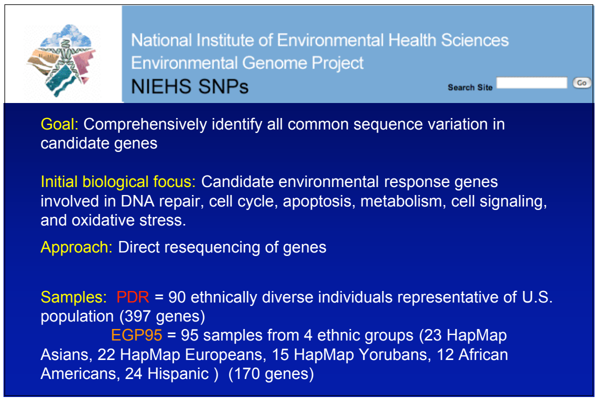
~4.0 million validated SNPs with genotypes!

(HapMap confirmed, allele frequency/population, SNPxSNP correlations (LD), haplotypes)



SNP discovery is dependent on your sample population size

2 chromosomes { GTTACGCCAATACAGGATCCAGGAGATTACC
GTTACGCCAATACAGCATCCAGGAGATTACC



National Institute of Environmental Health Sciences
Environmental Genome Project
NIEHS SNPs

Search Site
Go

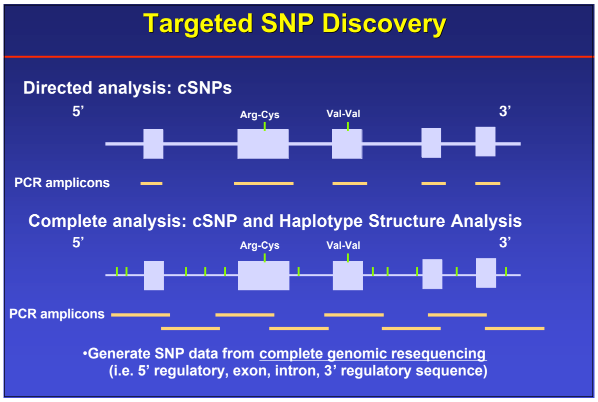
Goal: Comprehensively identify all common sequence variation in candidate genes

Initial biological focus: Candidate environmental response genes involved in DNA repair, cell cycle, apoptosis, metabolism, cell signaling, and oxidative stress.

Approach: Direct resequencing of genes

Samples: PDR = 90 ethnically diverse individuals representative of U.S. population (397 genes)

EGP95 = 95 samples from 4 ethnic groups (23 HapMap Asians, 22 HapMap Europeans, 15 HapMap Yorubans, 12 African Americans, 24 Hispanic) (170 genes)

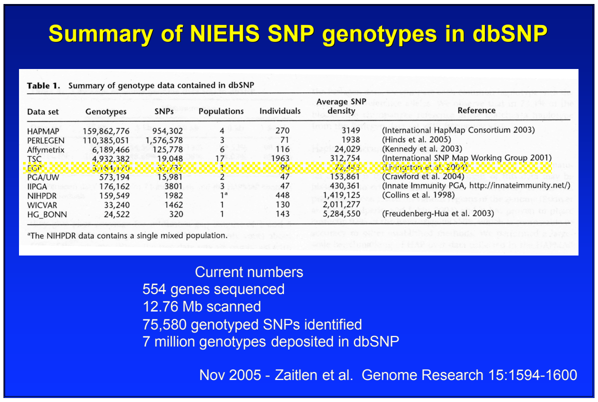


Targeted SNP Discovery

Directed analysis: cSNPs

Complete analysis: cSNP and Haplotype Structure Analysis

•Generate SNP data from complete genomic resequencing (i.e. 5' regulatory, exon, intron, 3' regulatory sequence)



Summary of NIEHS SNP genotypes in dbSNP

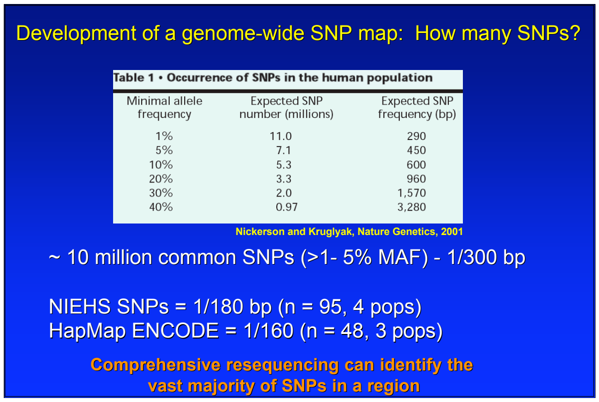
Table 1. Summary of genotype data contained in dbSNP

Data set	Genotypes	SNPs	Populations	Individuals	Average SNP density	Reference
HAPMAP	159,862,776	954,302	4	270	3149	(International HapMap Consortium 2003)
PERLEGEN	110,385,051	1,576,578	3	71	1938	(Hinds et al. 2005)
Affymetrix	6,189,466	125,778	6	116	24,029	(Kennedy et al. 2003)
TSC	4,932,382	19,048	17	1963	312,754	(International SNP Map Working Group 2001)
EGP	3,184,170	37,737	1	90	72,443	(Livingston et al. 2004)
PGA/UW	573,194	15,981	2	47	153,861	(Crawford et al. 2004)
IIPGA	176,162	3801	3	47	430,361	(Innate Immunity PGA, http://innateimmunity.net/)
NIHPDR	159,549	1982	1[a]	448	1,419,125	(Collins et al. 1998)
WICVAR	33,240	1462	1	130	2,011,277	
HG_BONN	24,522	320	1	143	5,284,550	(Freudenberg-Hua et al. 2003)

[a] The NIHPDR data contains a single mixed population.

Current numbers
554 genes sequenced
12.76 Mb scanned
75,580 genotyped SNPs identified
7 million genotypes deposited in dbSNP

Nov 2005 - Zaitlen et al. Genome Research 15:1594-1600



Development of a genome-wide SNP map: How many SNPs?

Table 1 • Occurrence of SNPs in the human population

Minimal allele frequency	Expected SNP number (millions)	Expected SNP frequency (bp)
1%	11.0	290
5%	7.1	450
10%	5.3	600
20%	3.3	960
30%	2.0	1,570
40%	0.97	3,280

Nickerson and Kruglyak, Nature Genetics, 2001

~ 10 million common SNPs (>1- 5% MAF) - 1/300 bp

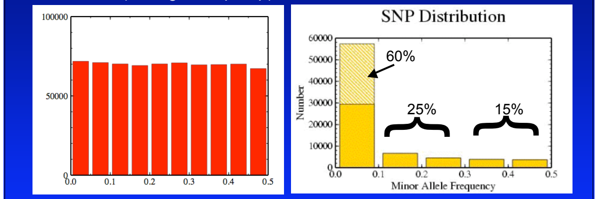
NIEHS SNPs = 1/180 bp (n = 95, 4 pops)
HapMap ENCODE = 1/160 (n = 48, 3 pops)

Comprehensive resequencing can identify the vast majority of SNPs in a region

SNP Discovery: dbSNP database

dbSNP (Perlegen/HapMap)

NIEHS SNPs



Minor Allele Freq. (MAF)

Rarer and population specific SNPs are found by resequencing

NIEHS SNPs Characterization

PDR = 90 ethnically diverse individuals representative of U.S. population (397 genes - ~55,000 SNPs)

Selection of informative (high frequency, coding, etc) SNPs to be genotyped in defined populations (~7600 SNPs)

HapMap Populations

European (CEU,n=60)
African (YRI, n =60)
Asian (HCB, n = 45 and JPT, n = 45)

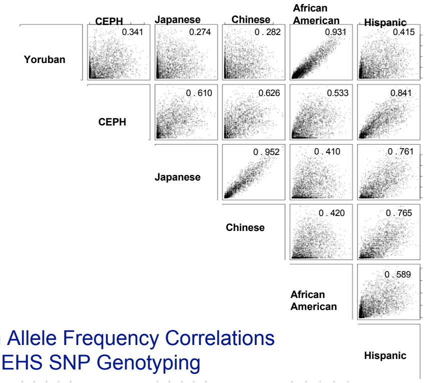
Non-HapMap Populations

Hispanic (n = 60)
African-American (n = 62)

Illumina NIEHS SNPs Genotyping

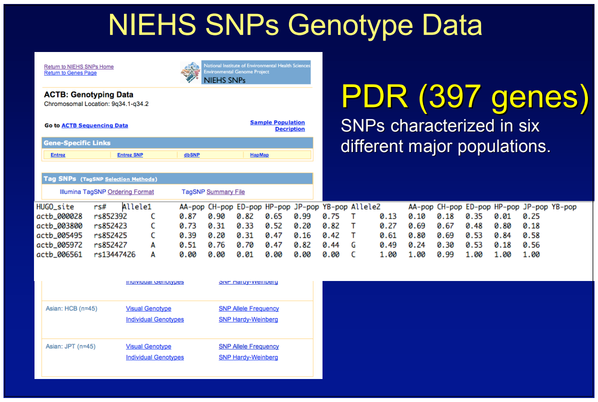
- Each well samples 1536 SNPs in one individual
- For each HapMap sample 5 x 1536 (7680 genotyped SNPs)
- 3,000,000 genotypes generated (total ~400 samples)

Array (1536)	Site Conversion Rate (%)	Average Site Coverage (%)	Concordance (%)
1	85	96.6	99.7
2	91	97.7	99.5
3	82	98.5	99.3



Population Allele Frequency Correlations
Illumina NIEHS SNP Genotyping

NIEHS SNPs Genotype Data



Return to NIEHS SNPs Home
Return to Genes Page

National Institute of Environmental Health Sciences
Environmental Genome Project
NIEHS SNPs

ACTB: Genotyping Data
Chromosomal Location: 9q34.1-q34.2

Go to ACTB Sequencing Data

Sample Population Decription

Gene-Specific Links

Entrez | Entrez SNP | dbSNP | HapMap

Tag SNPs (TagSNP Selection Methods)

Illumina TagSNP Ordering Format   TagSNP Summary File

```
HUGO_site     rs#      Allele1     AA-pop CH-pop ED-pop HP-pop JP-pop YB-pop Allele2      AA-pop CH-pop ED-pop HP-pop JP-pop YB-pop
actb_000028   rs852392       C     0.87   0.90   0.82   0.65   0.99   0.75   T       0.13   0.10   0.18   0.35   0.01   0.25
actb_003800   rs852423       C     0.73   0.31   0.33   0.52   0.20   0.82   T       0.27   0.69   0.67   0.48   0.80   0.18
actb_005495   rs852425       C     0.39   0.20   0.31   0.47   0.16   0.42   T       0.61   0.80   0.69   0.53   0.84   0.58
actb_005972   rs852427       A     0.51   0.76   0.70   0.47   0.82   0.44   G       0.49   0.24   0.30   0.53   0.18   0.56
actb_006561   rs13447426     A     0.00   0.00   0.01   0.00   0.00   0.00   C       1.00   1.00   0.99   1.00   1.00   1.00
```

Individual Genotypes   SNP Hardy-Weinberg

Asian: HCB (n=45)	Visual Genotype Individual Genotypes	SNP Allele Frequency SNP Hardy-Weinberg
Asian: JPT (n=45)	Visual Genotype Individual Genotypes	SNP Allele Frequency SNP Hardy-Weinberg

PDR (397 genes)

SNPs characterized in six different major populations.

Summary: The Current State of SNP Resources

- **SNPs have been rapidly adopted as the genetic marker of choice.**
- **Approximately 10 million common SNPs exist in the human genome (1/300 bp).**
- **Random SNP discovery processes generate many SNPs (TSC and HapMap).**
- **Random approaches to SNPs discovery have reached limits of discovery and validation (1/600 bp; 50% SNP validation)**
- **Most validated SNPs (5 million) will be genotyped by the HapMap (3 pops)**
- **Resequencing approaches continue to catalog important variants (rarer)**
- **NIEHS SNPs has generated SNP data on >550 candidate genes and 75 K SNPs**