

ALL RIGHTS RESERVED
<http://www.ces.genbank.com>



*"What ever will we think about now that
the genome project is almost complete?"*

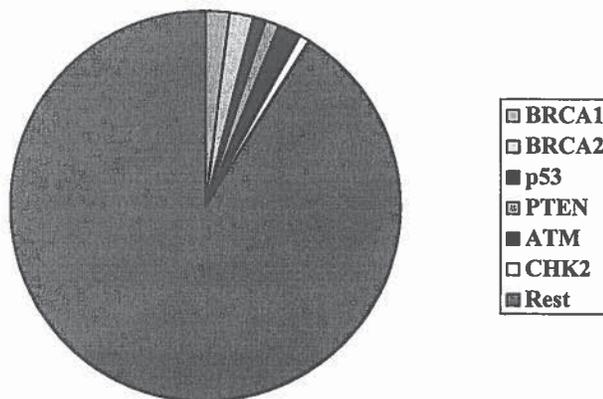
Mission of Core Genotyping Facility

**Develop an Integrated Program to
Apply Genomics to Large-scale
Population-based Epidemiology
Studies in the NCI Intramural
Program**

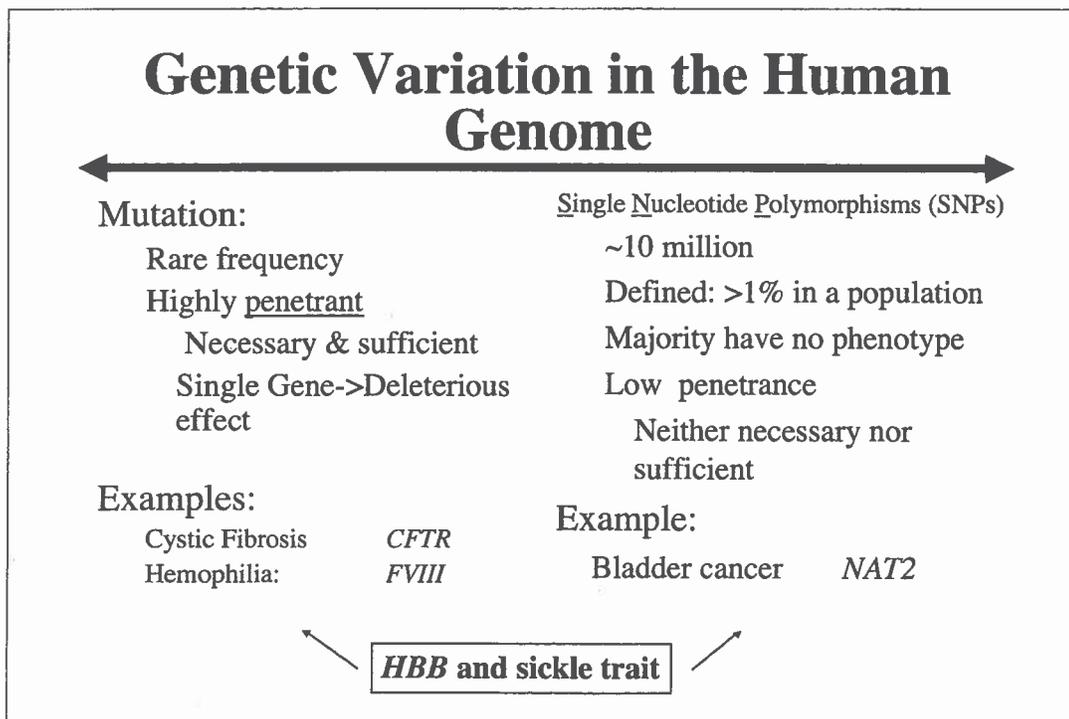
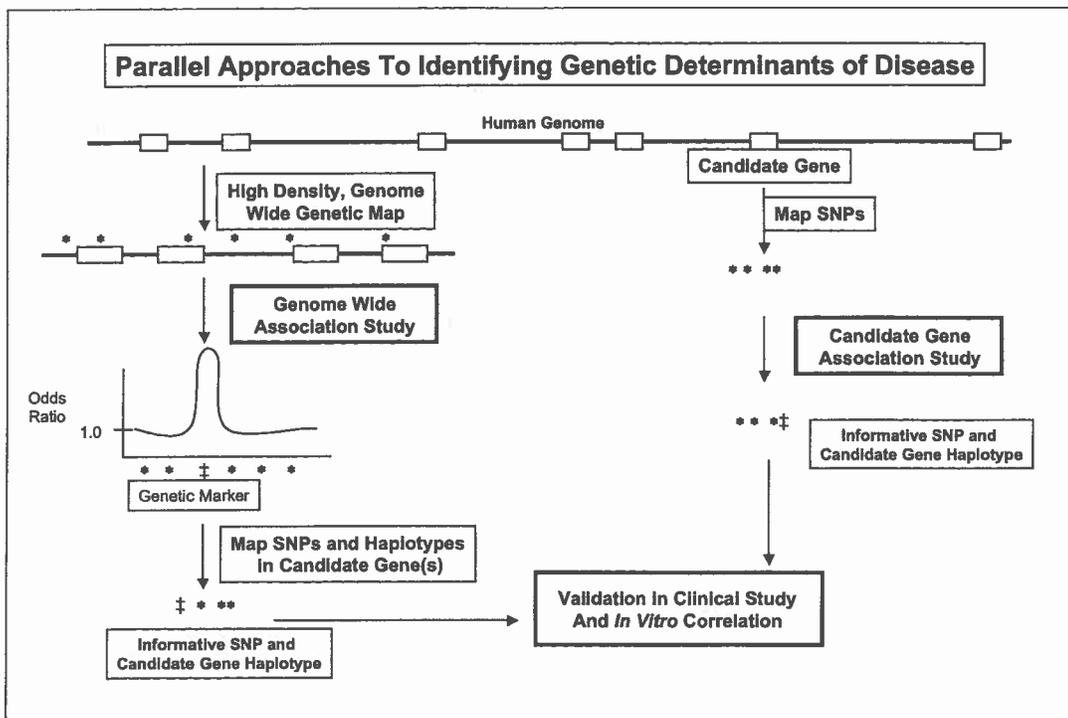
Comprehensive Approach of Core Genotyping Facility

1. Selection of candidate genes
2. Annotation of candidate genes
3. Test hypotheses in large population studies
4. Replication of studies
5. Critical resource for population genetics

PROPORTION OF BREAST CANCER ATTRIBUTABLE TO KNOWN GERMLINE GENE MUTATIONS



If genetics has a role in breast cancer, it will be as a complex disease and probably include many genetic factors

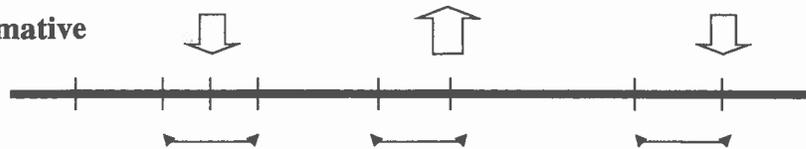


SNPs

Haplotype:

Combination of SNPs (alleles) which are inherited as a unit from parent to off-spring

Informative
SNPs



Evidence for Cancer Susceptibility in Population-based Studies

Cancer

Bladder

Gastric

HIV-related

Kaposi Sarcoma

Genetic variants

NAT2

GSTM1

IL1B

FCGR3A

Studies at the CGF

Received (in 2002)

24 studies with 25,000 samples

Analyzed (in 2002)

10 Studies with 14,000 samples

Delivered

~450,000 genotypes (SNPs)

~300,000 microsatellites (STRs)

Expect (in 2003 & 2004)

50 studies with ~80,000 samples

Studies at CGF

- Breast
- Prostate
- Lung
- Colon
- Brain
- Non-Hodgkins Lymphoma
- Ovarian
- Cervical
- Bladder
- Acute Lymphoblastic Leukemia
- Nasopharyngeal
- Melanoma
- Biliary
- Gastric
- Esophageal
- Testicular
- Kaposi's Sarcoma
- Neuroblastoma
- Osteogenic Sarcoma
- Ewing's Sarcoma

Scope of Studies at the CGF:

~50 Studies of 20 Cancers

Population-based Case-Control

Brain

Bladder

NHL (InterLymph Consortium)

Cohort Studies

Breast & Prostate (Cohort Consortium)

Family Studies

Melanoma, CLL

Bio-Informatics: The Core of CGF

Annotation of genes

Select candidate variants (SNPs & Haplotypes)

Education

Lectures

Office Hours

Web-tools (SNP500cancer of CGAP)

Drives Workflow

Tracking of Genotypes by Sample/Project/Study

Oracle Data-base

Lab Information Management System (LIMS)

Genotype Analysis: Each SNP must be looked at the 'old-fashioned way'



One small segment (amplicon) at a time.....

NATIONAL CANCER INSTITUTE **CGAP** Cancer Genome Anatomy Project
SNP500Cancer Database

NCI > CGAP > SNP500Cancer > Home

Welcome
PACKERB

Home

Search by Gene/
Chromosome/
Haplotype

Search by SNP

Links

Log In

What is SNP500Cancer?

The goal of the SNP500Cancer project is to resequence 102 reference samples to find known or newly discovered single nucleotide polymorphisms (SNPs) which are of immediate importance to molecular epidemiology studies in cancer. SNP500Cancer provides a central resource for sequence verification of SNPs. For more information, see [SNP500Cancer background](#).

Search for SNPs in SNP500Cancer database

You can search for SNPs using SNP identifier, gene symbol, gene alias, chromosome, gene ontology pathway, or haplotype.

- **by SNP Identifier**
Enter the dbSNP ID or internal SNP ID:
- **by Gene**
Enter gene symbol, alias, or GenBank ID:
or: List genes with analyzed SNPs
or: List genes with pending SNPs
- **by Chromosome**
Enter chromosome number:
- **by Gene Ontology Pathway**
Enter GO Pathway (text or numeric):
- **by Haplotype** [List of available haplotypes](#)

<http://snp500cancer.nci.nih.gov>

dbSNP ID: rs179955					
Polymorphism ID: BRCA2-04					
Gene	BRCA2				
Amino Acid Change	S2414				
HGVBase ID	SNP000002462				
Sequence of Analyzed Amplicon					
GAACATCTGACTTTGGAAAAATCTCAAGCAATTTAGCAGTTTCAGGACATC CATTTTATCAAGTTTCTGCTACAAAGAAATGAAAATGAGACACTTGATTAC TACAGGCAGACCAACCAAAGTCTTTGTCACCTTTTAAACTAAATC (A/G))CATTTCACAGAGTTGAACAGTGTGTTAGGAATATTAACTTGGAGGAAAC AGACAAAACGAAAACATTGATGGACRTGGCTCTGATGATAGTAAARAYAGA TTAATGACAATGAGATTCATCAGTTTAAACAAAACAACTCCAATCAAGCAGY AGCTGTAACCTTTCACAAAGT					
Frequency Data (102 anonymized subjects)					
Total Completed (validated by TaqMan)	Genotypic			Allelic	
100	AA 59/100 (0.590)	AG 33/100 (0.330)	GG 8/100 (0.080)	A 151/200 (0.755)	G 49/200 (0.245)
View Subpopulation Frequencies					

Bio-informatic Analysis

← SNP of interest
← Linked SNPs

Frequencies for:
Genotype
Alleles
Haplotype Estimation (E-M)

You requested the subpopulation data for rs179955:

dbSNP ID: rs179955						
Subpopulations	Genotypic			Linkage passed HWE?	Allelic	
	AA	AG	GG		A	G
Total Completed	59/100 (0.590)	33/100 (0.330)	8/100 (0.080)	-	151/200 (0.755)	49/200 (0.245)
Afr/Afr American	13/24 (0.542)	8/24 (0.333)	3/24 (0.125)	passed	34/48 (0.708)	14/48 (0.292)
Caucasian	17/30 (0.567)	11/30 (0.367)	2/30 (0.067)	passed	45/60 (0.750)	15/60 (0.250)
Hispanic	17/23 (0.739)	6/23 (0.261)	0/23 (0.000)	passed	40/46 (0.870)	6/46 (0.130)
Pacific Rim	12/23 (0.522)	8/23 (0.348)	3/23 (0.130)	passed	32/46 (0.696)	14/46 (0.304)

Why Validate SNPs by Sequence?

~15-20% "SNPs" in public databases show no variation

Assay Optimization

Population Specific Linked SNPs

Develop Reference Standard

Coriell Samples- Quality Control

Upcoming Features of SNP500Cancer

Population frequencies for SNPs

**Data from as many as 40,000 controls
genotyped at CGF**

Estimates of

Haplotypes

SNPs required for haplotype

Nucleotide Diversity

<http://snp500cancer.nci.nih.gov>

Optimization Initiatives at CGF: Looking Towards the Future

Quantification of DNA

Efficient Usage

DNA Pooling

Efficient Screening of Cases vs. Controls

Whole Genome Amplification

Utilize Precious Samples

CGF Partnerships

Academic Collaborations

DCEG

Bio-statistics

Genetic Epidemiology

MALDI-TOF Working group

NHGRI F Collins

CIDR

NIAMS D Kastans

Consultants

HLA J Hansen

Evolutionary A Hughes

Collaborative Development

ABI

TaqMan Optimization Pipeline

EPOCH Biosciences

Eclipse (3' Hybridization-
Triggered Fluorescence)

Agencourt

Serum-based gDNA extraction

Whole Genome SNP Studies

Illumina Affymetrix

Orchid ABI

Goals of the NCI-Core Genotyping Facility

Integration of genomics into molecular
epidemiology studies

Participation in design, choice, and analysis
of genetic variants

Develop SNP assays (genes, haplotypes,
pathways on CGF website)

Assess promising technologies