Studying Genetic Variation II: Laboratory Techniques

Karen Mohlke, PhD
Department of Genetics
University of North Carolina

Genetic variation in other lectures

- **Population genetics, patterns of human genetic variation, linkage disequilibrium, HapMap, genome-wide association studies - Lynn Jorde**
- **Linkage analysis, genome-wide linkage studies, haplotype analysis, susceptibility to cancer- Elaine Ostrander**
- **Origins of genetic variants, types of variants, discovery methods, use of databases, HapMap, linkage disequilibrium - Jim Mullikan**

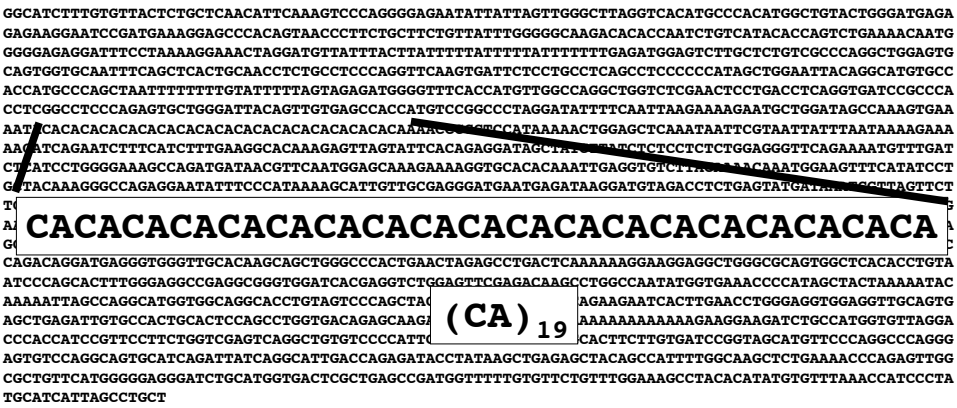
Human Genetic Variation

- Types of variants
- Methods for scoring variants
- Genome-wide scoring of SNPs
- Structural variants

Human Genetic Variation

- Sequence repeats
- Single nucleotide polymorphisms
- Insertions and deletions
- Other structural variation

Microsatellite

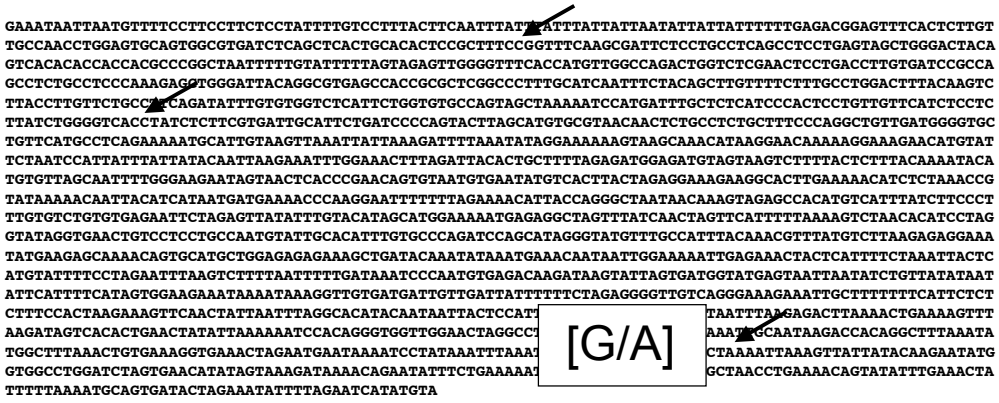


Example dinucleotide marker named
AFM059XA9 and D3S1262

Microsatellites

- **Many alleles, highly informative**
- **>50,000 in human genome**
- **Relatively high mutation rate**
- **Used to build first framework map**

Single nucleotide polymorphisms (SNPs)



Three SNPs are located at positions 49,719,887, 49,720,260 and 49,721,557.

SNPs

- **Less polymorphic/informative**
- **More stable inheritance**
- **~1 SNP with frequency greater than 1% per 300 nucleotides (10 million in genome)**
- **Mutation at CpG 10-fold higher rate**
- **Exist in coding regions**

Deletion/insertion polymorphisms (indels)

- **One to many nucleotides present or not**
- **Example:**

AGTATCTTCACAGAAATGACCATA

AGTATCTTCACAAGAAATGACCATA

AGTATCTTCACA[-/A]GAAATGACCATA

Indel polymorphisms

Another example:

CAGACTCAATAAGCATGTTTTTACAGACTCAATAAGCATGTTT
TTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCA
GGCTGGAGTGCAGTGGCGCGATCTCGGCTCACTGCAAGCTC
CGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGA
GTAGCTGGGACTACAGGCTCCCGCCACCACGCCCGGCTAAT
TTTTTGTATTTTTAGTAGAGACGGGGTTAGCATGTTTTT

CAGACTCAATA[LARGEINSERTION/-]AGCATGTTTTT

Structural variation

- **Includes deletions, insertions, duplications, inversions, translocations**
- **~1 million > 1 bp, at least 1500 > 1kb**
- **Many small indels are in linkage disequilibrium with nearby SNPs**
- **Some deletions and rearrangements recur between repeated sequences**

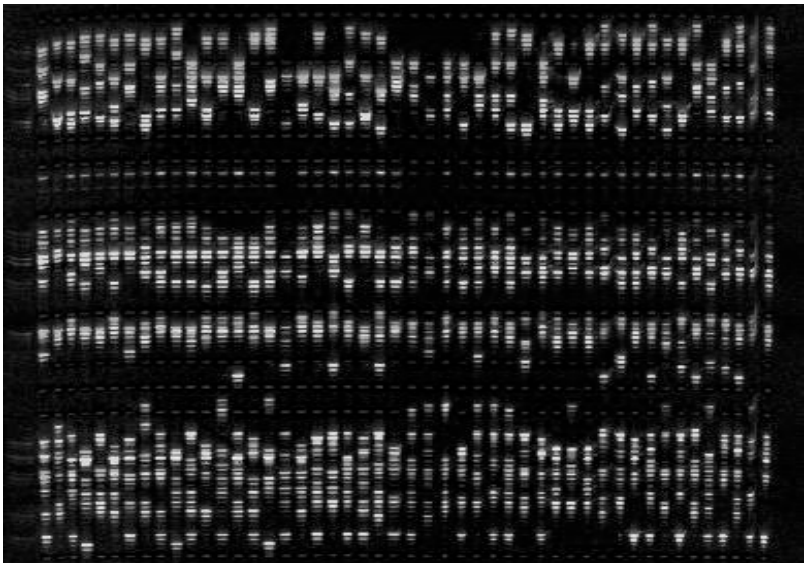
Human Genetic Variation

- **Types of variants**
- **Methods for scoring variants**
- **Genome-wide scoring of SNPs**
- **Structural variants**

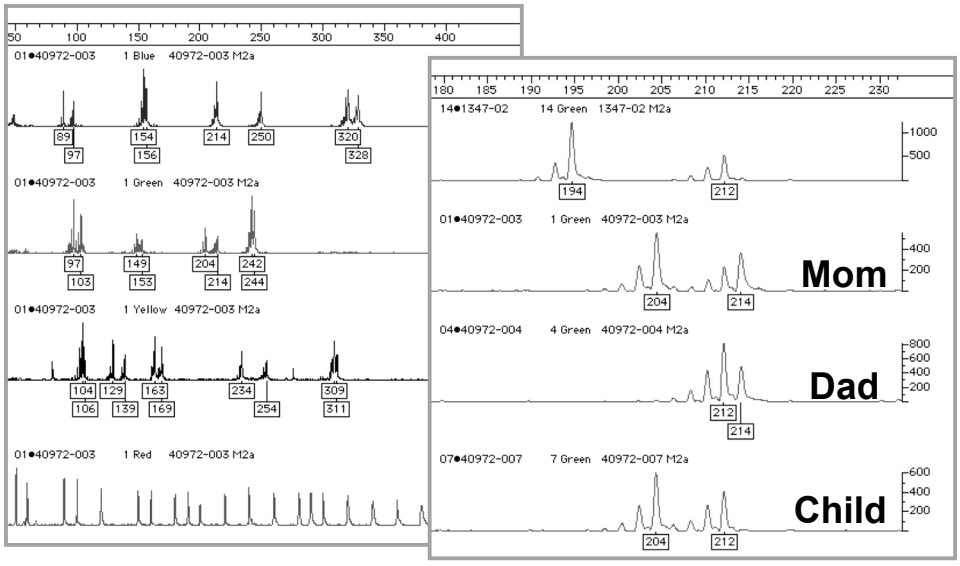
Scoring Variants

- **Scoring = genotyping = typing**
- **Laboratory technique depends on**
  - **Type of variant**
  - **Fixed or custom set of variants**
  - **Number of variants**
  - **Number of samples**

Scoring Microsatellites



Scoring Microsatellites



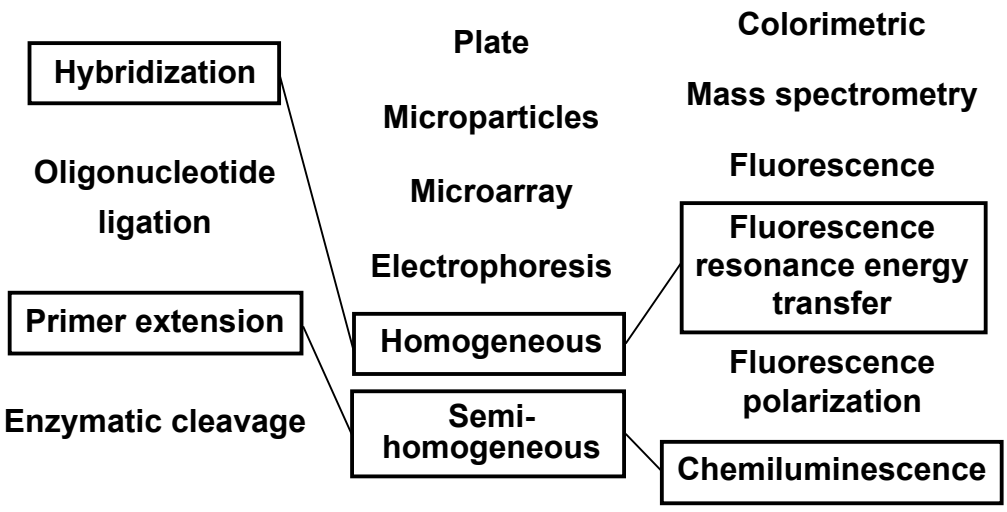
Scoring SNPs

- **Genotype accuracy**
- **Cost of assays and specialized instrument(s)**
- **Assay development time and ease**
- **Ability to automate**

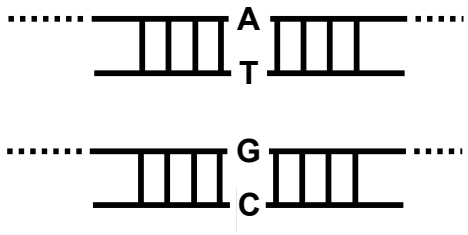
Scoring SNPs (2)

- **Time to perform assays**
- **Ability to multiplex**
- **Data accumulation and analysis**
- **Allele frequency quantification**

Overview of SNP typing methods

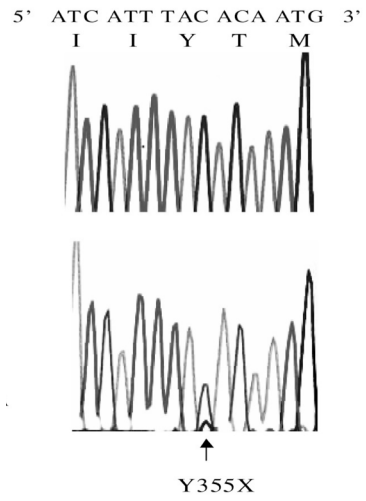


Example SNP



Sequencing

PPARG Y355X



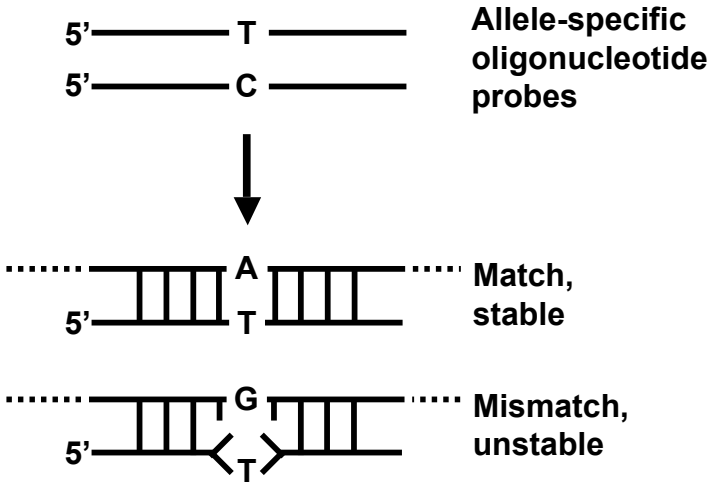
Francis et al. BMC Med Gen 2006 7:3

Sequencing

- **Advantages:**
  - **Instrumentation widely available**
  - **Easy and fast for small studies**

- **Disadvantages**
  - **Expensive for many SNPs or samples**
  - **Local sequence affects success**

Hybridization



Affymetrix Custom Sequencing Array

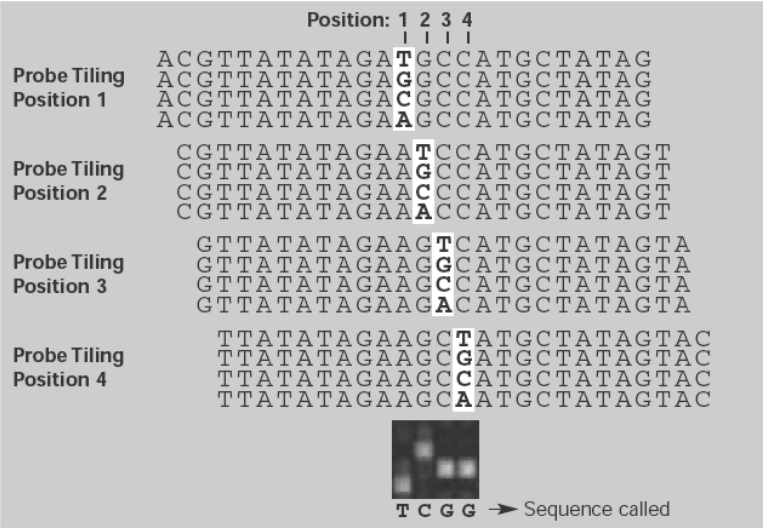


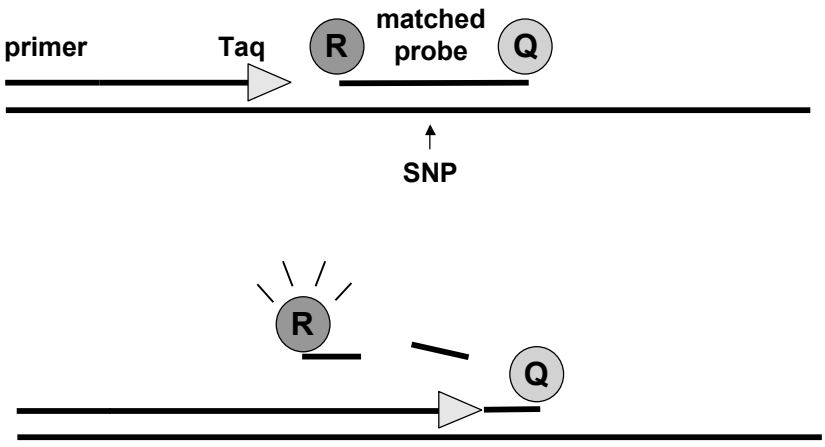
Figure 1: CustomSeq™ arrays tile four probes per strand for each individual base. The central position of each probe varies to incorporate each of the four possible nucleotides—A, C, G, or T.

images from affymetrix.com

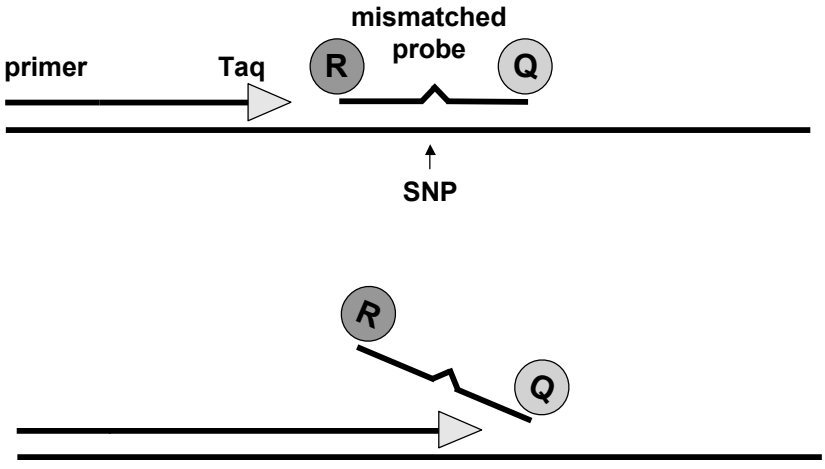
Hybridization to Oligonucleotide Arrays

- **Advantages:**
  - **Simple to perform**
  - **Highly multiplexed**
  - **Automated analysis**
- **Disadvantages**
  - **Custom chip expensive to design/create**
  - **Local sequence affects success**

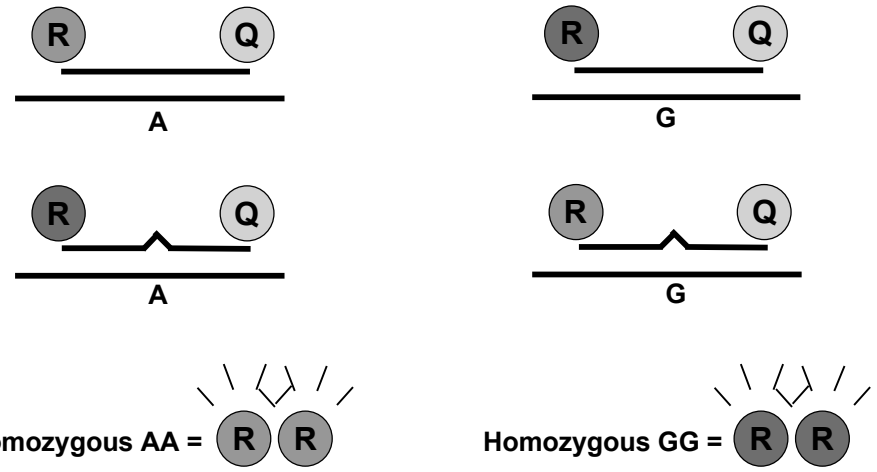
Fluorescence resonance energy transfer (FRET)



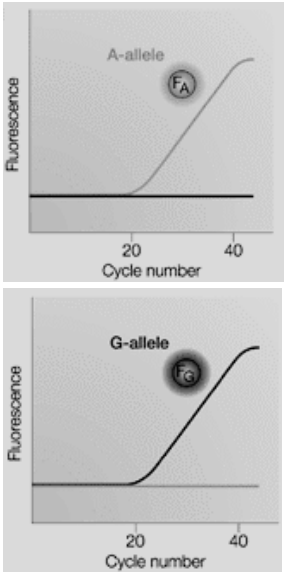
Fluorescence resonance energy transfer (FRET)

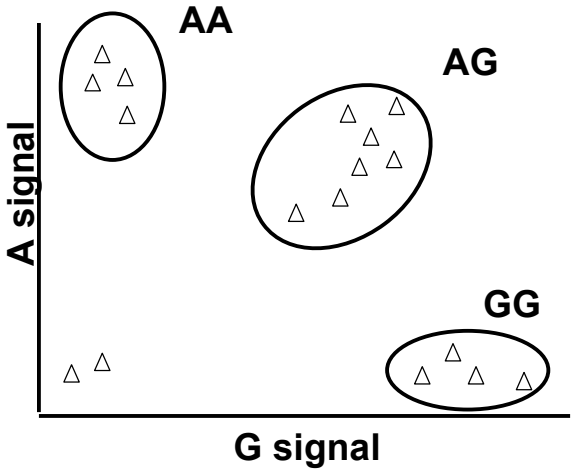


TaqMan competing probes



TaqMan genotype scoring



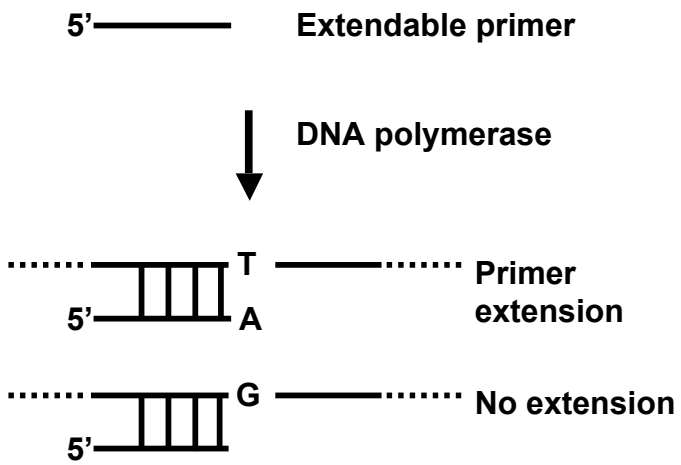


TaqMan

- **Advantages:**
  - **Simple to perform**
  - **Closed-tube system**
  - **Accurate quantification**

- **Disadvantages**
  - **Expensive probes**
  - **No multiplexing**
  - **Assays require optimization**

Primer extension = Minisequencing

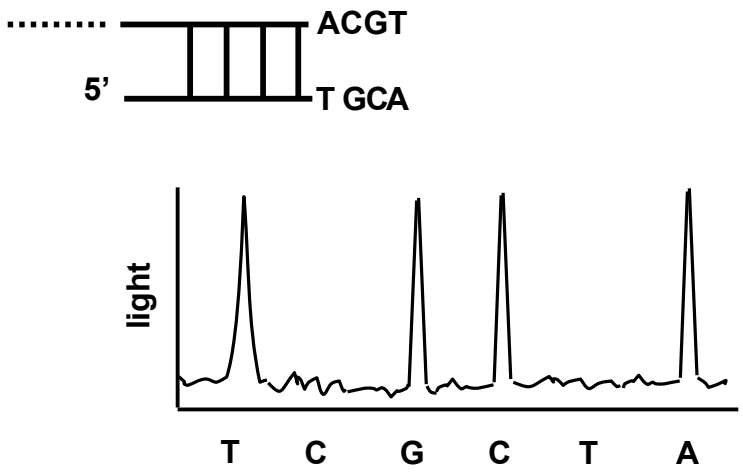


Pyrosequencing

- **Four enzymes**
  - **DNA polymerase**
  - **ATP sulfurylase--converts pyrophosphate to ATP**
  - **Luciferase--converts ATP to light**
  - **Apyrase--degrades excess nucleotides**
- **Nucleotides added sequentially**

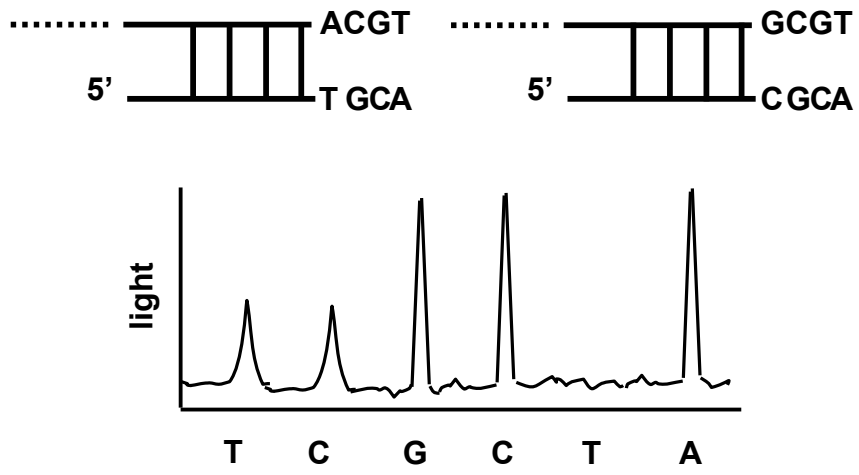
Pyrosequencing

…[A/G]CGT…



Pyrosequencing

…[A/G]CGT…



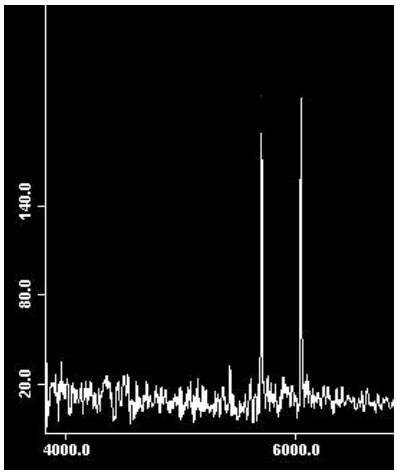
Pyrosequencing

- **Advantages:**
  - **Accurate**
  - **Accurate allele frequency estimation**
  - **Robust for closely spaced SNPs**
- **Disadvantages**
  - **Expensive reagents**
  - **Requires post-PCR processing**

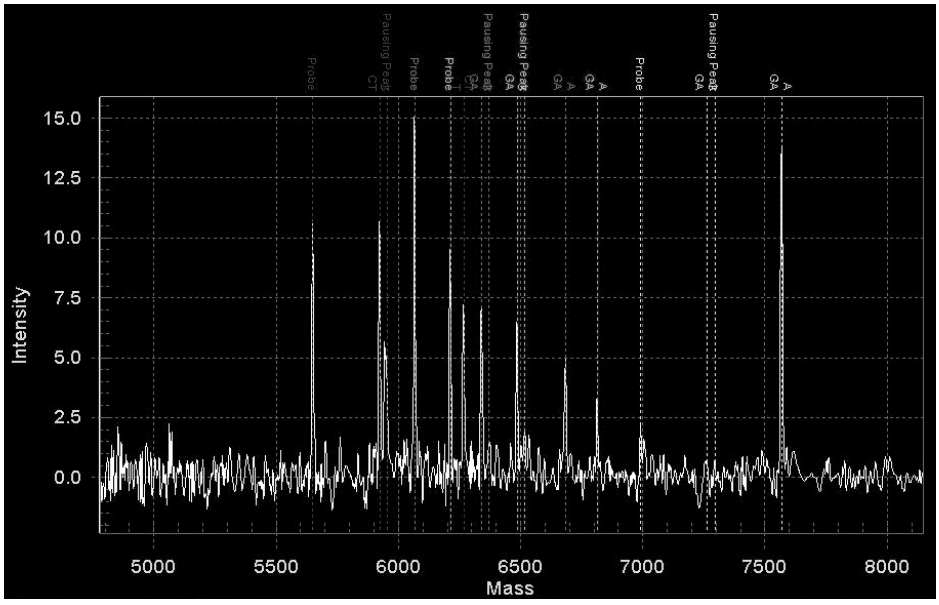
Primer extension mass spectrometry

Primer extension reactions designed to generate different sized products

	Mass in Daltons
GGACCTGGAGCCCCCACC	**5430.5**
GGACCTGGAGCCCCCACCC	**5703.7**
GGACCTGGAGCCCCCACCTG	**6047.9**

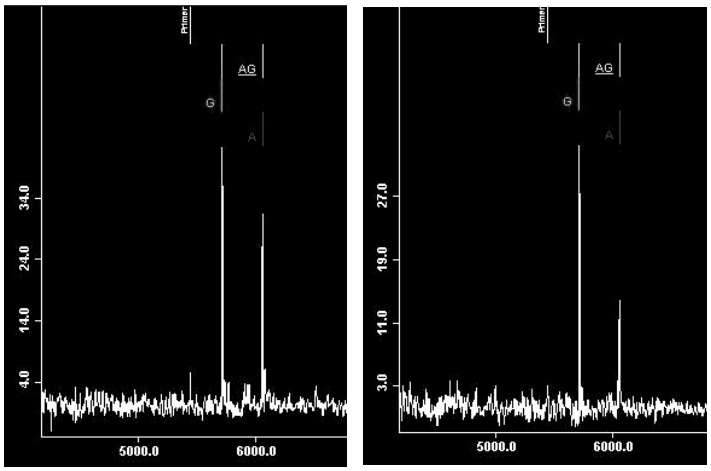


Mass spectrometry multiplexing



Allelic quantification

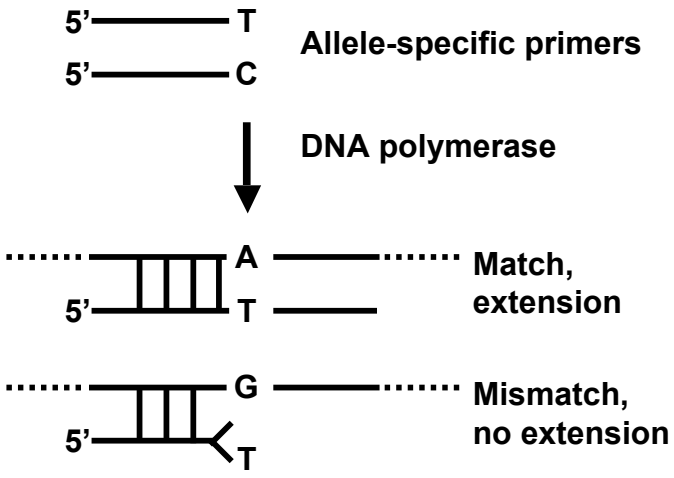
- **Using cDNA or DNA pools or tumor sample**
- **Type SNP and determine relative allele frequencies**



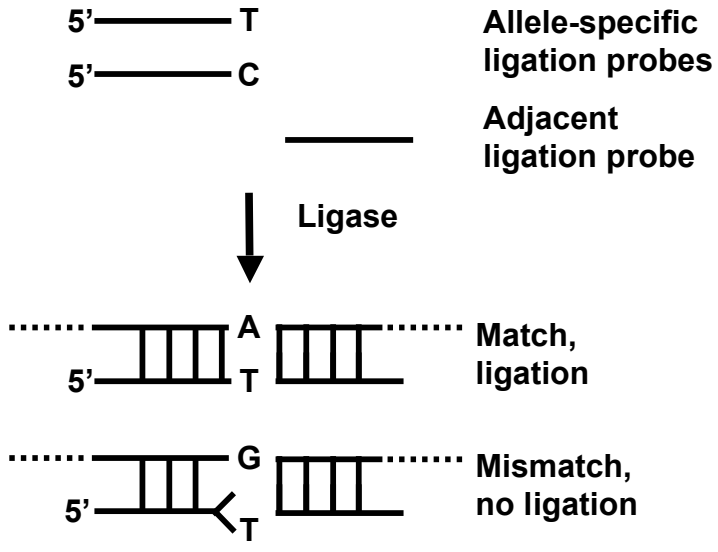
Primer extension mass spectrometry

- **Advantages:**
  - **Accurate**
  - **Automated assay design**
  - **Fast automated data collection**
  - **Multiplexing capacity**
- **Disadvantages**
  - **Expensive instruments, consumables**
  - **Extensive post-PCR processing**

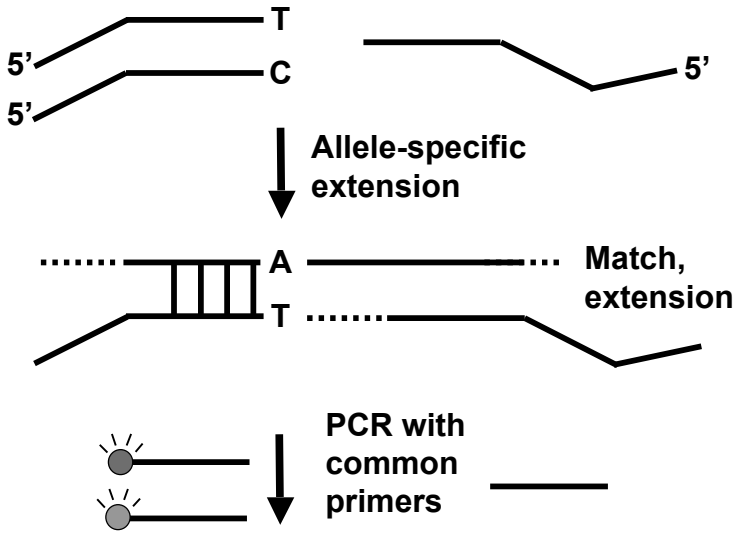
Allele-specific PCR



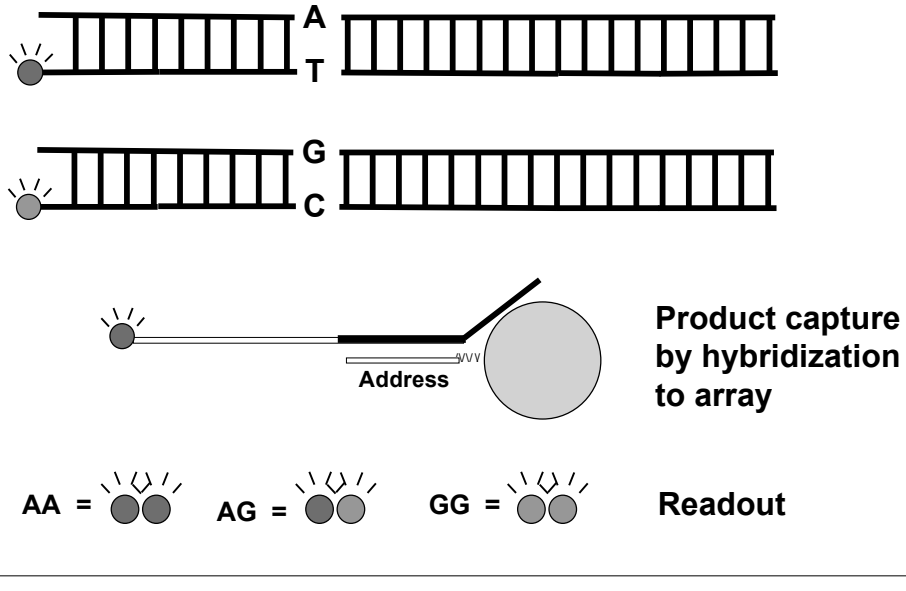
Oligonucleotide Ligation Assay (OLA)



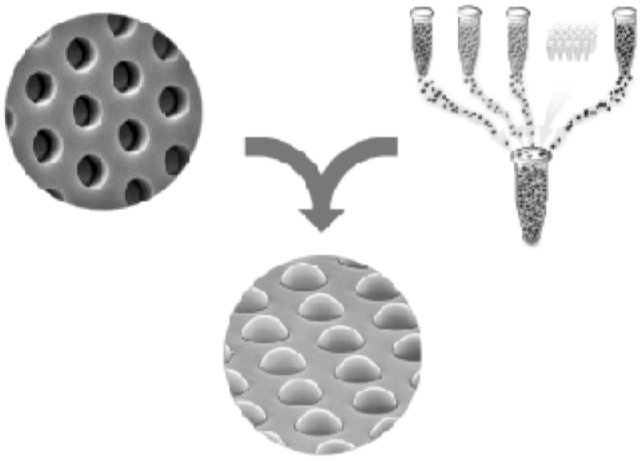
GoldenGate: Allele-specific extension



GoldenGate: Allele-specific extension



GoldenGate genotyping technology



Illumina GoldenGate

- **Advantages:**
  - **Very highly multiplexed**
  - **Accurate**
  - **Low cost per genotype**
- **Disadvantages**
  - **Not cost-effective for small studies**
  - **Limits to SNPs that can be designed**

Quality control of genotype data

- **High genotype success**
- **Accurate duplicate genotypes**
- **No genotypes in no DNA controls**
- **Allele frequencies similar to databases**
- **Accurate on a second platform**

Quality control of genotype data

- **Test whether data are consistent with Hardy-Weinberg Equilibrium (HWE): $p^2 + 2pq + q^2 = 1$**
- **Calculate observed frequencies p and q**
- **Use p and q to calculate expected genotype frequencies**
- **Compare observed and expected genotype frequencies by X^2 test with 1 degree of freedom**

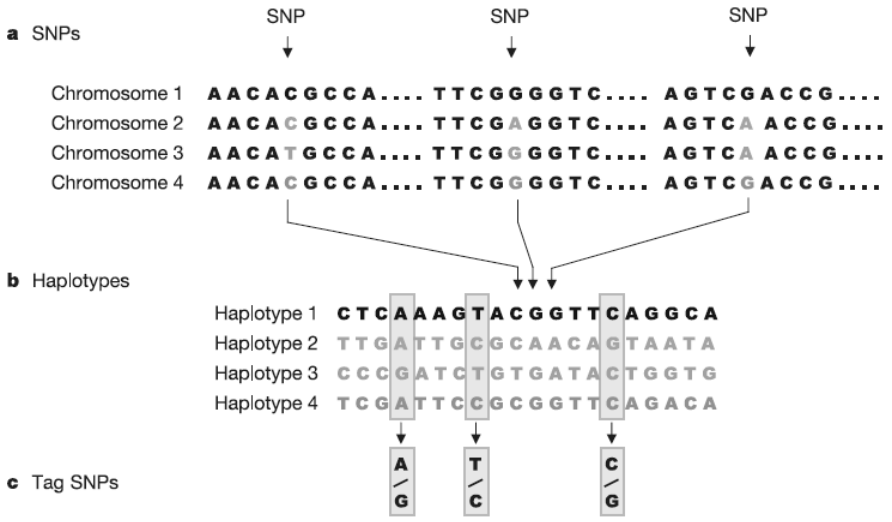
Human Genetic Variation

- **Types of variants**
- **Methods for scoring variants**
- **Genome-wide scoring of SNPs**
- **Structural variants**

Genome-wide SNP panels

- **10,000 - 650,000+ SNPs per experiment**
- **Affymetrix, Illumina, Parallele, Perlegen**
  - **Random SNPs**
  - **Selected haplotype tag SNPs**
  - **Coding or nonsynonymous SNPs**

Selecting 'haplotype tag' SNPs



International HapMap Consortium (2003) Nature 426:789

Affymetrix GeneChip Array

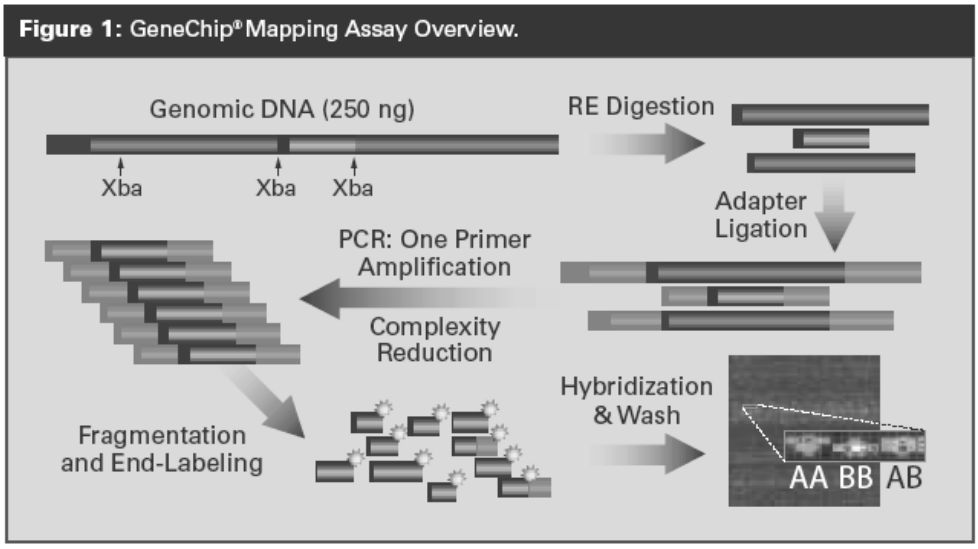
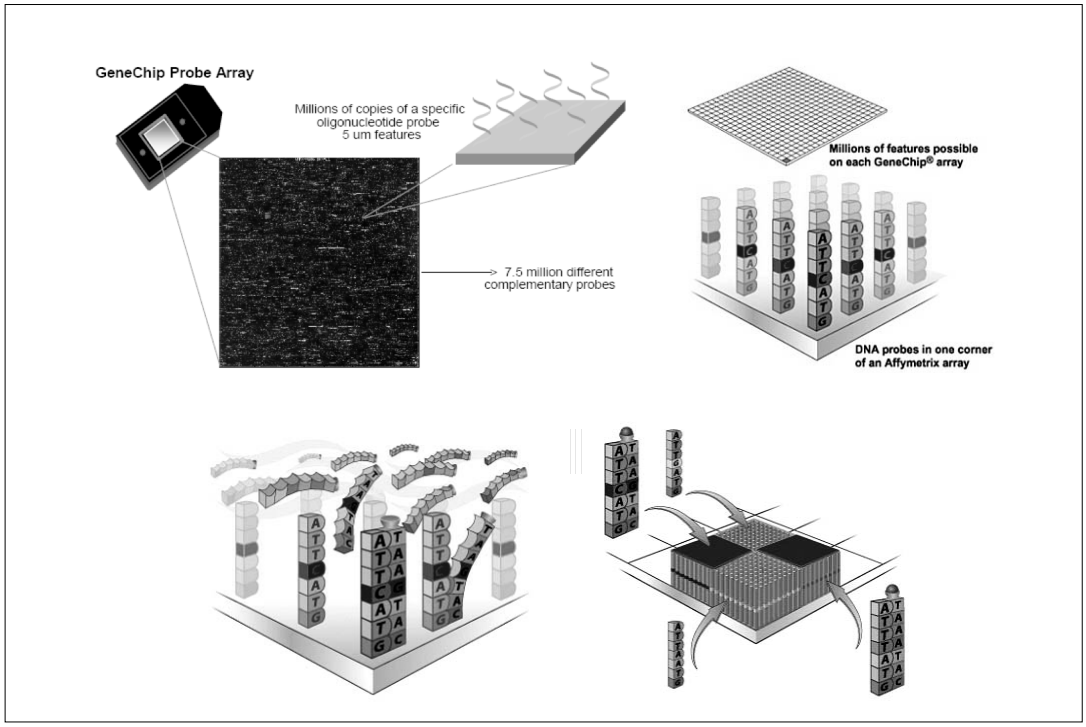
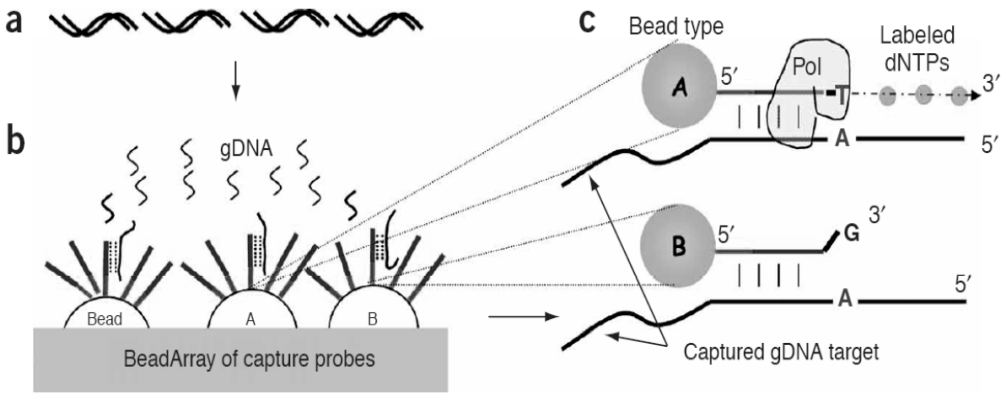


image from affymetrix.com

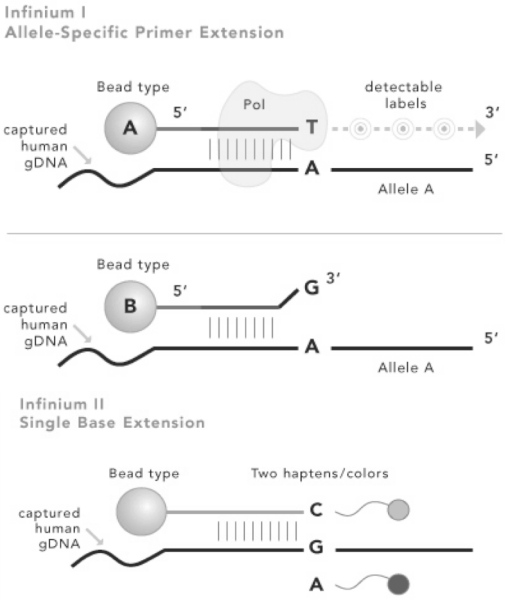


Illumina Infinium Assay



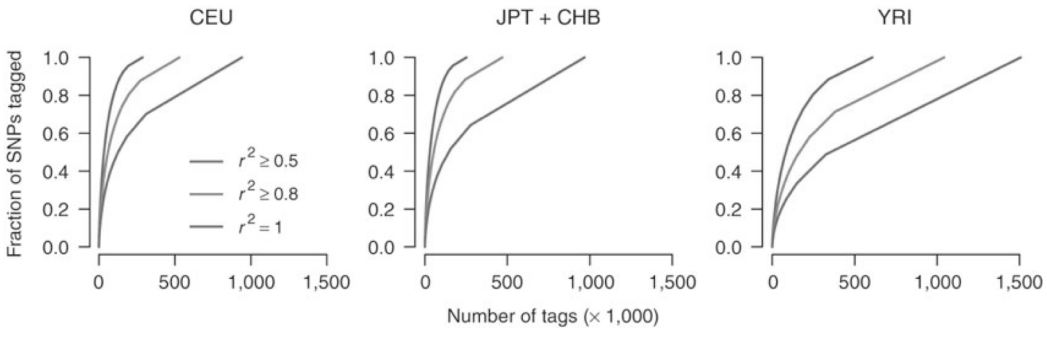
Gunderson et al. (2005) NatGen 37:549

Illumina Infinium Assays



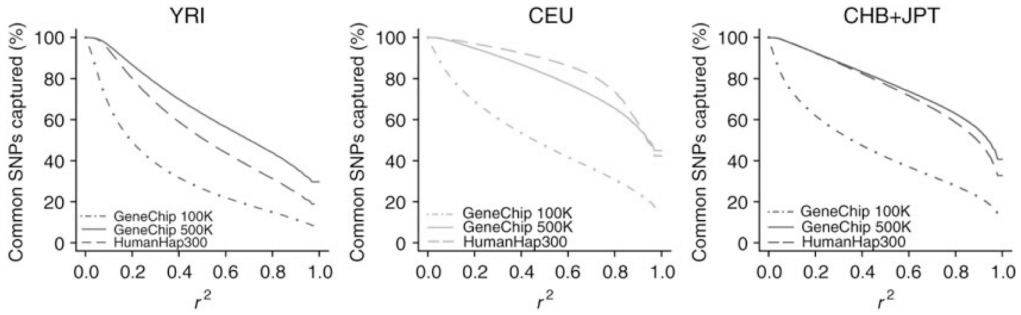
Illumina.com

Genomic coverage: maximally efficient tag SNP sets



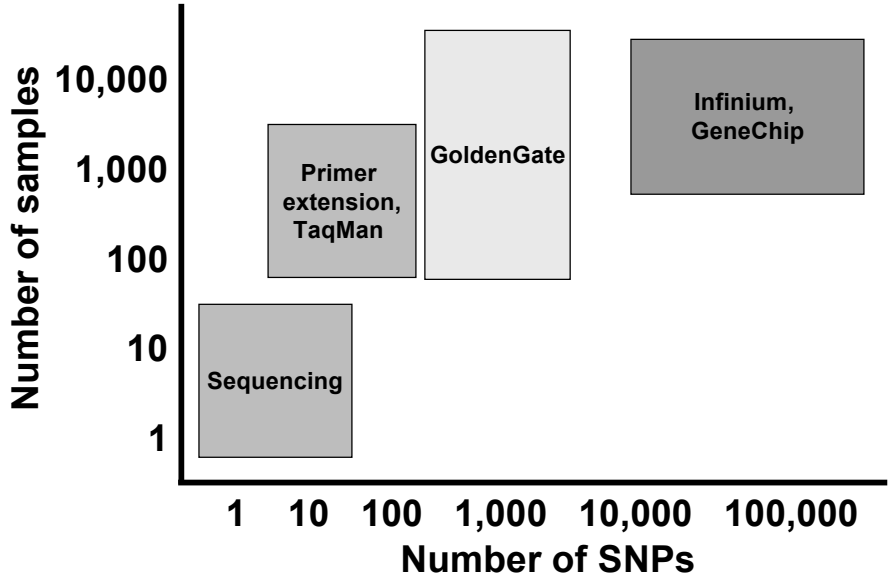
Barrett et al. (2006) NatGen 38:659

Coverage of genome-wide panels



Pe'er et al. (2006) NatGen 38:663

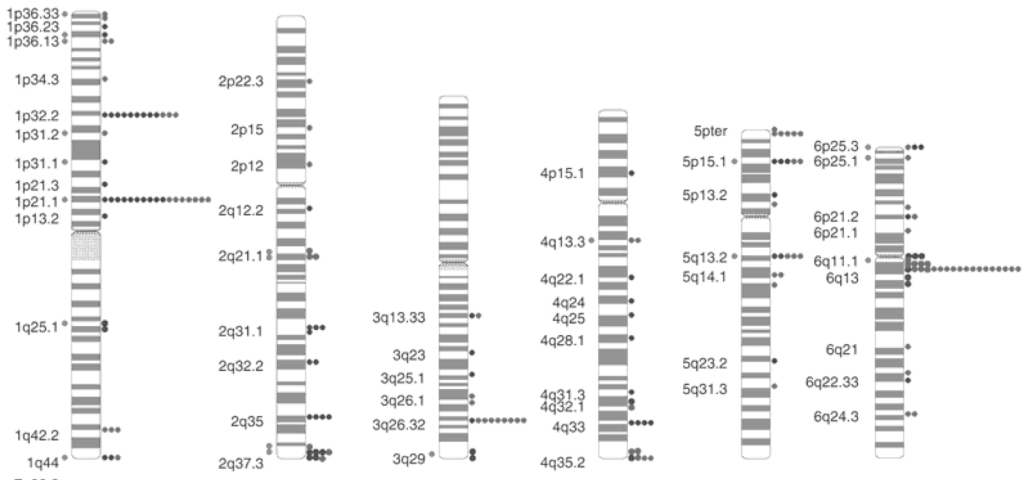
Which SNP genotyping method?



Human Genetic Variation

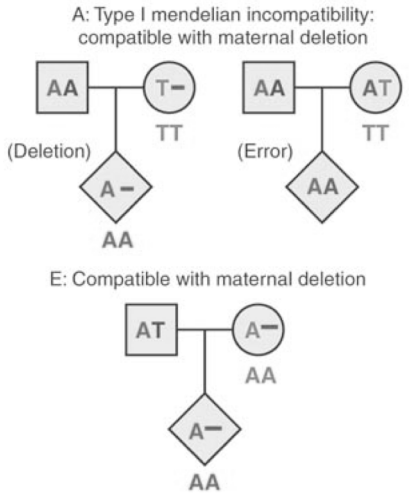
- **Types of variants**
- **Methods for scoring variants**
- **Genome-wide scoring of SNPs**
- **Structural variants**

Structural variants span the genome



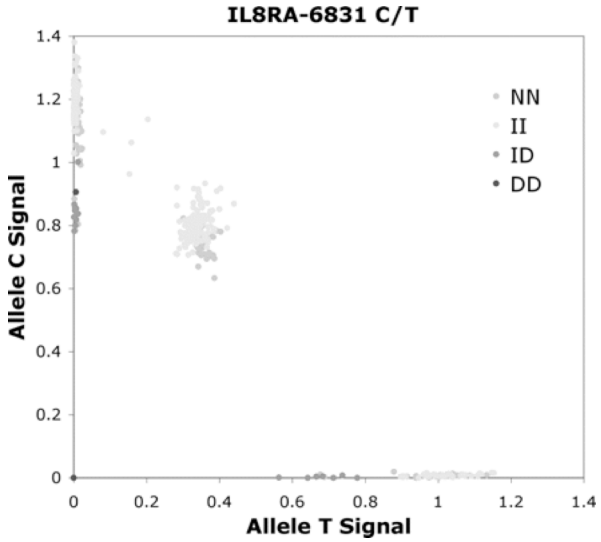
Iafrate et al. (2004) NatGen 36:949

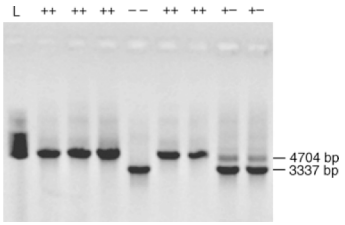
Detecting deletions from SNP data



Conrad et al. (2006) NatGen 38:75

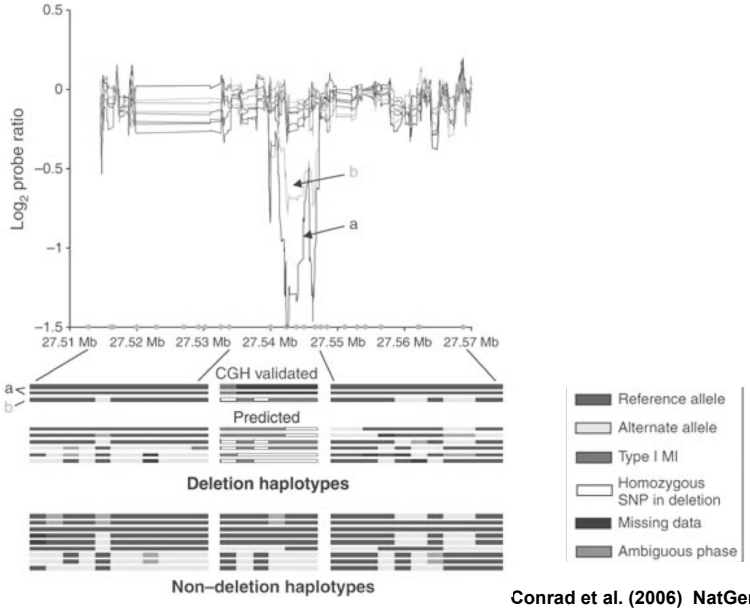
Allele Intensity in SNP Genotyping





Carlson et al. (2006) HMG 15:1931

Comparative genomic hybridization



Conrad et al. (2006) NatGen 38:75

Future

- **Faster, cheaper, easier genotyping**
- **More SNP panels for genome-wide association studies**
- **Genome maps of structural variants**
- **Discovery of new susceptibility genes for complex traits**

References

SNP Genotyping

Syvanen (2001) Nat Review Genet 2:930

Kwok (2001) Ann Rev Genomics Hum Genet 2:235

Gut (2001) Human Mutation 17:475

Genome-wide SNP Genotyping

Matsuzaki (2004) Genome Research 14:414

Matsuzaki (2004) Nature Methods 1:109

Gunderson (2005) Nature Genetics 37:549

Copy Number Variation

Feuk (2006) Human Molecular Genetics 15:R57

Eichler (2006) Nature Genetics 38:9