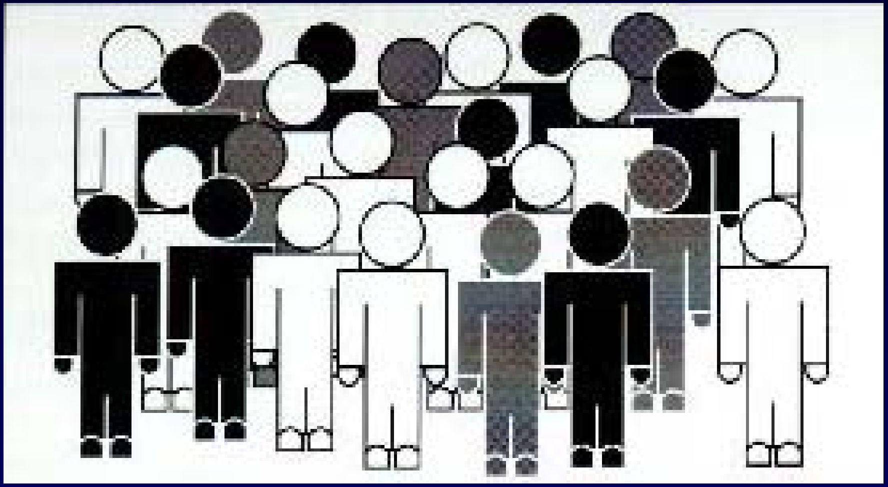
Human genetic variation



CHEW Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Human Genetic Variation

Variants contribute to rare and common diseases

Variants can be used to trace human origins

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Human Genetic Variation

- What types of variants exist?
- How are variants found?
- How are variants scored?
- How are variants used?

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Human Genetic Variation

- Sequence repeats
- Single nucleotide polymorphisms
- Insertion/deletion
  - Nucleotide(s)
  - Alu element

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

LINES (Long interspersed elements)

The human genome contains over 500,000 LINES (representing some 16% of the genome).

LINES are long DNA sequences that represent reverse-transcribed RNA molecules originally transcribed by RNA polymerase II; that is, messenger RNAs.

Lacking introns as well as the necessary control elements like promoters, these genes are not expressed. They are called pseudogenes. However, some LINES do encode a functional reverse transcriptase and/or integrase.

These enable them to mobilize not only themselves but also

- other, otherwise nonfunctional, LINES and
- Alu sequences.

Because transposition is done by copy-paste, the number of LINES can increase in the genome. The diversity LINES between individual human genomes make them useful markers for DNA “fingerprinting”.

SINES (Short interspersed elements)

SINES are short DNA sequences that represent reverse-transcribed RNA molecules originally transcribed by RNA polymerase III; that is, molecules of tRNA, 5S rRNA, and some other small nuclear RNAs.

The most abundant SINES are the **Alu elements**. There are about one million copies in the human genome (representing about 11% of the total DNA).

Alu elements consist of a sequence of 300 base pairs containing a site that is recognized by the restriction enzyme AluI. They appear to be reverse transcripts of 7S RNA, part of the signal recognition particle.

SINES do not encode any functional molecules and (like LINES) their presence in the genome is a mystery. Like LINES, they seem to represent only "junk" or "selfish" DNA.

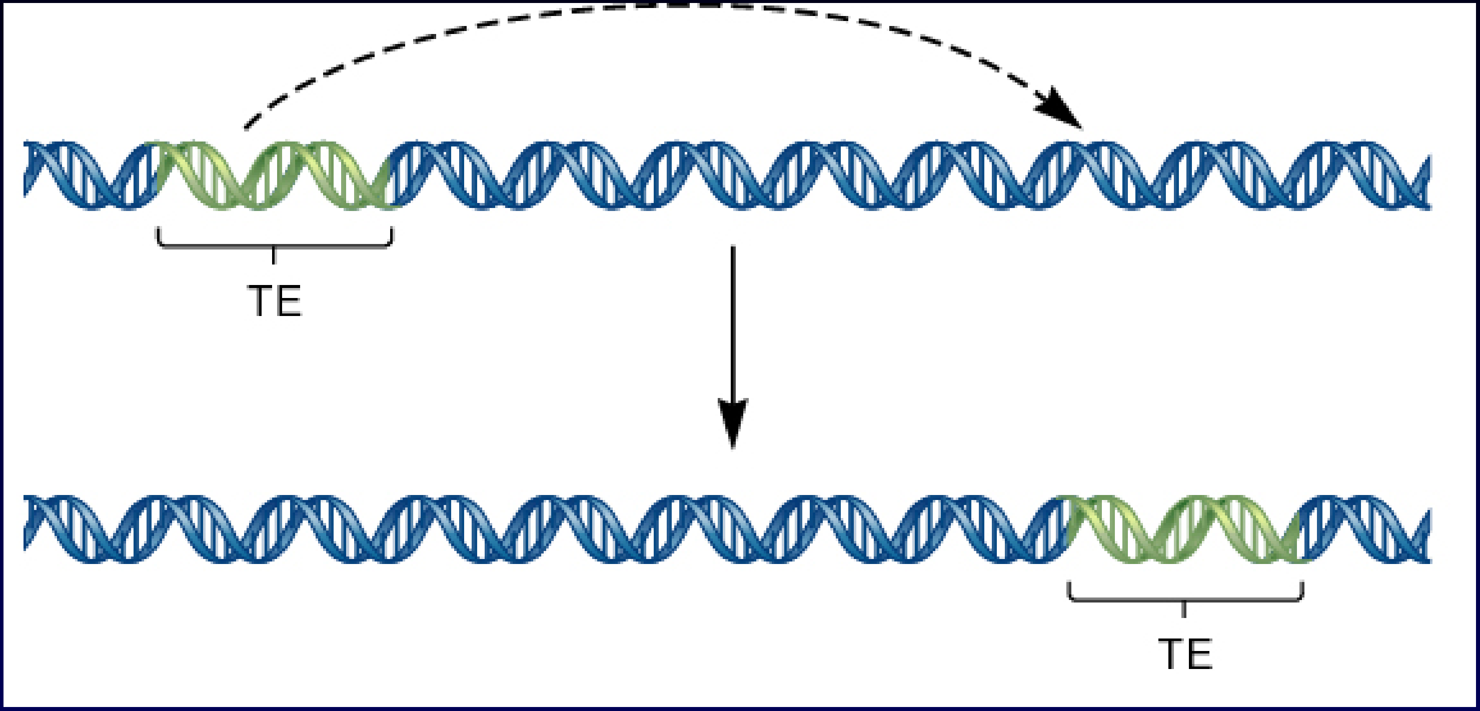
Transposable elements

- Many transposable elements have been found in bacteria, fungi, plant and animal cells

- Three general types of transposition pathways have been identified
  - 1. Simple transposition
  - 2. Replicative transposition
  - 3. Retrotransposition

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

1. Simple transposition

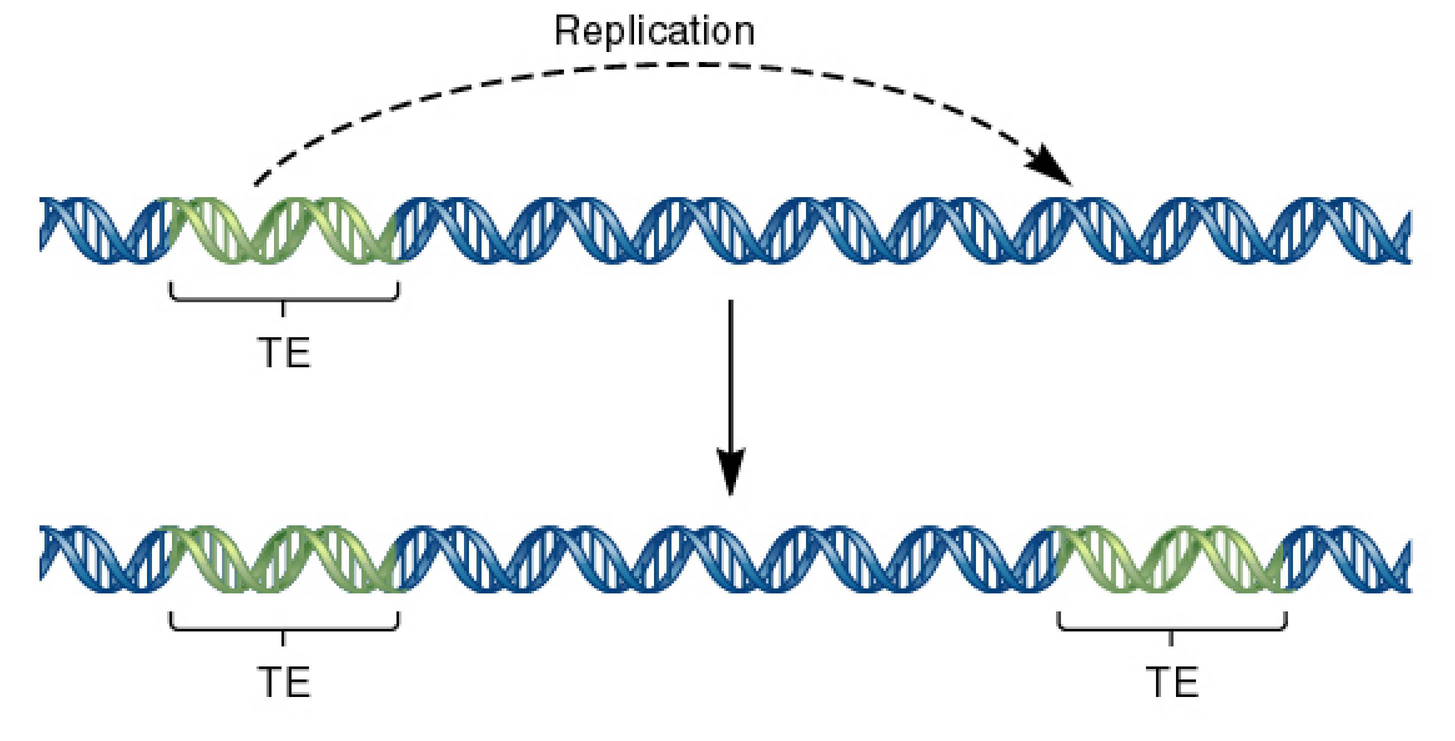


- This mechanism is also called a “cut-and-paste”
- It is widely found in bacteria and eukaryotes

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

2. Replicative transposition

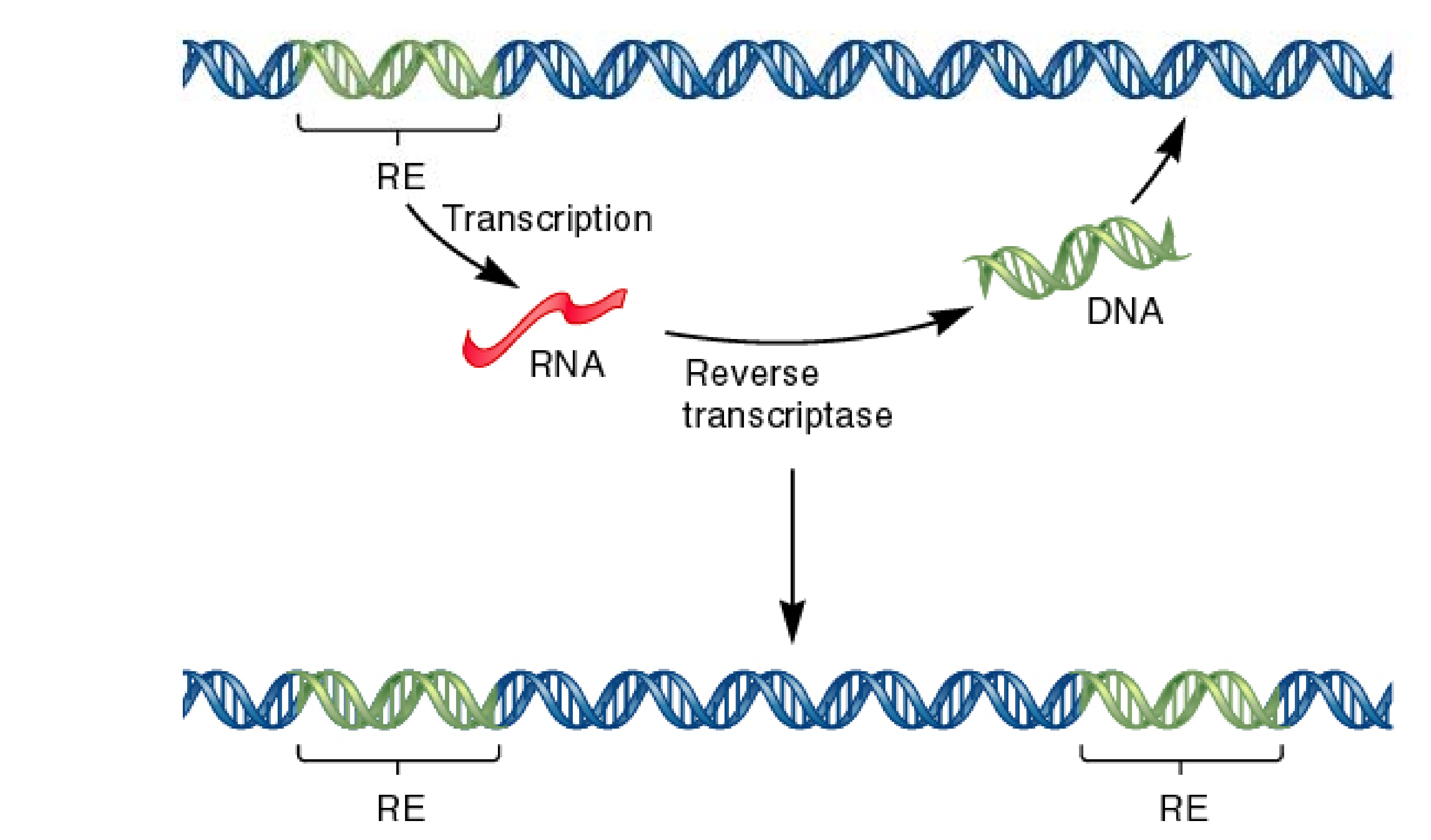


(b) Replicative transposition

- This mechanism involves replication of the TE and insertion of the copy into another chromosomal location

- It is relatively uncommon and only found in bacteria

3. Retrotransposition



- This mechanism is very common but only found in eukaryotes

- These types of elements are termed retroelements, retrotransposons, or retroposons

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Human Genetic Variation

- Sequence repeats
- Single nucleotide polymorphisms
- Insertion/deletion
  - Nucleotide(s)
  - Alu element

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

A typical sequence from the human genome…

```
GGCATCTTTGTGTTACTCTGCTCAACATTCAAAGTCCCAGGGGAGAATATTATTAGTTGGGCTTAGGTCACATGCCCACATGGCTGTACTGGGATGAGA
GAGAAGGAATCCGATGAAAGGAGCCCACAGTAACCCTTCTGCTTCTGTTATTTGGGGGCAAGACACACCAATCTGTCATACACCAGTCTGAAAACAATG
GGGGAGAGGATTTCCTAAAAGGAAACTAGGATGTTATTTACTTATTTTTATTTTTATTTTTTTGAGATGGAGTCTTGCTCTGTCGCCCAGGCTGGAGTG
CAGTGGTGCAATTTCAGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGCCTCAGCCTCCCCCCATAGCTGGAATTACAGGCATGTGCC
ACCATGCCCAGCTAATTTTTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCCA
CCTCGGCCTCCCAGAGTGCTGGGATTACAGTTGTGAGCCACCATGTCCGGCCCTAGGATATTTTCAATTAAGAAAAGAATGCTGGATAGCCAAAGTGAA
AATACACACACACACACACACACACACACACACACACACACAAAACCCCGTCCATAAAAACTGGAGCTCAAATAATTCGTAATTATTTAATAAAAGAAA
AACATCAGAATCTTTCATCTTTGAAGGCACAAAGAGTTAGTATTCACAGAGGATAGCTATCTTATCTCTCCTCTCTGGAGGGTTCAGAAAATGTTTGAT
CTCATCCTGGGGAAAGCCAGATGATAACGTTCAATGGAGCAAAGAAAAGGTGCACACAAATTGAGGTGTCTTACAAAACAAATGGAAGTTTCATATCCT
GCTACAAAGGGCCAGAGGAATATTTCCCATAAAAGCATTGTTGCGAGGGATGAATGAGATAAGGATGTAGACCTCTGAGTATGATAAATGGTTAGTTCT
TCCTATTAGTTGTTGTTTCTGATGTAGAAACAGCGTCTTTCTCCCTATATCTGGTCTAAAATCCAACCTGATAGGAGACGTTTTCGTTTGGGATTATGG
AAAGATACAACAGTTCTGGGGGTTGAGTTCAGGGCTAATTTTCTGAAGGATAAGAGAGCAAGCCCCAGCCAAGAGCCAAGAGAAAGCAATGATGAGGAA
GCGGGCAGTAGCAGCCATTTAGACTGGTTGCTTTGTGGGACTCCCTTCTATTTGTACATTATTAGGCTTTCCAACAGGGGACAATAAACAGTATGAATC
CAGACAGGATGAGGGTGGGTTGCACAAGCAGCTGGGCCCACTGAACTAGAGCCTGACTCAAAAAAGGAAGGAGGCTGGGCGCAGTGGCTCACACCTGTA
ATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCACGAGGTCTGGAGTTCGAGACAAGCCTGGCCAATATGGTGAAACCCCATAGCTACTAAAAATAC
AAAAATTAGCCAGGCATGGTGGCAGGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTG
AGCTGAGATTGTGCCACTGCACTCCAGCCTGGTGACAGAGCAAGACTCCATCTCAAAAAAAAAAAAAAAAAAAGAAGGAAGATCTGCCATGGTGTTAGGA
CCCACCATCCGTTCCTTCTGGTCGAGTCAGGCTGTGTCCCCATTGACTGGGGCATGATTGCACTTCTTGTGATCCGGTAGCATGTTCCCAGGCCCAGGG
AGTGTCCAGGCAGTGCATCAGATTATCAGGCATTGACCAGAGATACCTATAAGCTGAGAGCTACAGCCATTTTGGCAAGCTCTGAAAACCCAGAGTTGG
CGCTGTTCATGGGGGAGGGATCTGCATGGTGACTCGCTGAGCCGATGGTTTTTGTGTTCTGTTTGGAAAGCCTACACATATGTGTTTAAACCATCCCTA
TGCATCATTAGCCTGCT
```

…from sequence on chromosome 3 stretching from base positions 212,378,797 to 212,380,793 of the UCSC August 2001 assembly.

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Microsatellite

```
GGCATCTTTGTGTTACTCTGCTCAACATTCAAAGTCCCAGGGGAGAATATTATTAGTTGGGCTTAGGTCACATGCCCACATGGCTGTACTGGGATGAGA
GAGAAGGAATCCGATGAAAGGAGCCCACAGTAACCCTTCTGCTTCTGTTATTTGGGGGCAAGACACACCAATCTGTCATACACCAGTCTGAAAACAATG
GGGGAGAGGATTTCCTAAAAGGAAACTAGGATGTTATTTACTTATTTTTATTTTTATTTTTTTGAGATGGAGTCTTGCTCTGTCGCCCAGGCTGGAGTG
CAGTGGTGCAATTTCAGCTCACTGCAACCTCTGCCTCCCAGGTTCAAGTGATTCTCCTGCCTCAGCCTCCCCCCATAGCTGGAATTACAGGCATGTGCC
ACCATGCCCAGCTAATTTTTTTTTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCGCCCA
CCTCGGCCTCCCAGAGTGCTGGGATTACAGTTGTGAGCCACCATGTCCGGCCCTAGGATATTTTCAATTAAGAAAAGAATGCTGGATAGCCAAAGTGAA
AATA CACACACACACACACACACACACACACACACACACACACACA
AAACCCCGTCCATAAAAACTGGAGCTCAAATAATTCGTAATTATTTAATAAAAGAAAAACATCAGAATCTTTCATCTTTGAAGGCACAAAGAGTTAGTA
TTCACAGAGGATAGCTATCTTATCTCTCCTCTCTGGAGGGTTCAGAAAATGTTTGATCTCATCCTGGGGAAAGCCAGATGATAACGTTCAATGGAGCAA
AGAAAAGGTGCACACAAATTGAGGTGTCTTACAAAACAAATGGAAGTTTCATATCCTGCTACAAAGGGCCAGAGGAATATTTCCCATAAAAGCATTGTT
GCGAGGGATGAATGAGATAAGGATGTAGACCTCTGAGTATGATAAATGGTTAGTTCTTCCTATTAGTTGTTGTTTCTGATGTAGAAACAGCGTCTTTCT
CCCTATATCTGGTCTAAAATCCAACCTGATAGGAGACGTTTTCGTTTGGGATTATGGAAAGATACAACAGTTCTGGGGGTTGAGTTCAGGGCTAATTTT
CTGAAGGATAAGAGAGCAAGCCCCAGCCAAGAGCCAAGAGAAAGCAATGATGAGGAAGCGGGCAGTAGCAGCCATTTAGACTGGTTGCTTTGTGGGACT
CCCTTCTATTTGTACATTATTAGGCTTTCCAACAGGGGACAATAAACAGTATGAATCCAGACAGGATGAGGGTGGGTTGCACAAGCAGCTGGGCCCACT
GAACTAGAGCCTGACTCAAAAAAGGAAGGAGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCACGAGGT
CTGGAGTTCGAGACAAGCCTGGCCAATATGGTGAAACCCCATAGCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGCACCTGTAGTCCCAGC
TACTCGGGAGGCTGAGGCAGAAGAATCACTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCTGAGATTGTGCCACTGCACTCCAGCCTGGTGACAGAGCA
AGACTCCATCTCAAAAAAAAAAAAAAAAAAAAGAAGGAAGATCTGCCATGGTGTTAGGACCCACCATCCGTTCCTTCTGGTCGAGTCAGGCTGTGTCCCCA
TTGACTGGGGCATGATTGCACTTCTTGTGATCCGGTAGCATGTTCCCAGGCCCAGGGAGTGTCCAGGCAGTGCATCAGATTATCAGGCATTGACCAGAG
ATACCTATAAGCTGAGAGCTACAGCCATTTTGGCAAGCTCTGAAAACCCAGAGTTGGCGCTGTTCATGGGGGAGGGATCTGCATGGTGACTCGCTGAGC
CGATGGTTTTTGTGTTCTGTTTGGAAAGCCTACACATATGTGTTTAAACCATCCCTATGCATCATTAGCCTGCT
```

A dinucleotide marker named AFM059XA9 and D3S1262 is located at position 212,379,395.

Microsatellite

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

More typical sequence…

```
GAAATAATTAATGTTTTCCTTCCTTCTCCTATTTTGTCCTTTACTTCAATTTATTTATTTATTATTAATATTATTATTTTTTGAGACGGAGTTTCACTCTTGT
TGCCAACCTGGAGTGCAGTGGCGTGATCTCAGCTCACTGCACACTCCGCTTTCTGGTTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACA
GTCACACACCACCACGCCCGGCTAATTTTTGTATTTTTAGTAGAGTTGGGGTTTCACCATGTTGGCCAGACTGGTCTCGAACTCCTGACCTTGTGATCCGCCA
GCCTCTGCCTCCCAAAGAGCTGGGATTACAGGCGTGAGCCACCGCGCTCGGCCCTTTGCATCAATTTCTACAGCTTGTTTTCTTTGCCTGGACTTTACAAGTC
TTACCTTGTTCTGCCTTCAGATATTTGTGTGGTCTCATTCTGGTGTGCCAGTAGCTAAAAATCCATGATTTGCTCTCATCCCACTCCTGTTGTTCATCTCCTC
TTATCTGGGGTCACATATCTCTTCGTGATTGCATTCTGATCCCCAGTACTTAGCATGTGCGTAACAACTCTGCCTCTGCTTTCCCAGGCTGTTGATGGGGTGC
TGTTCATGCCTCAGAAAAATGCATTGTAAGTTAAATTATTAAAGATTTTAAATATAGGAAAAAAGTAAGCAAACATAAGGAACAAAAAGGAAAGAACATGTAT
TCTAATCCATTATTTATTATACAATTAAGAAATTTGGAAACTTTAGATTACACTGCTTTTAGAGATGGAGATGTAGTAAGTCTTTTACTCTTTACAAAATACA
TGTGTTAGCAATTTTGGGAAGAATAGTAACTCACCCGAACAGTGTAATGTGAATATGTCACTTACTAGAGGAAAGAAGGCACTTGAAAAACATCTCTAAACCG
TATAAAAACAATTACATCATAATGATGAAAACCCAAGGAATTTTTTTAGAAAACATTACCAGGGCTAATAACAAAGTAGAGCCACATGTCATTTATCTTCCCT
TTGTGTCTGTGTGAGAATTCTAGAGTTATATTTGTACATAGCATGGAAAAATGAGAGGCTAGTTTATCAACTAGTTCATTTTTAAAAGTCTAACACATCCTAG
GTATAGGTGAACTGTCCTCCTGCCAATGTATTGCACATTTGTGCCCAGATCCAGCATAGGGTATGTTTGCCATTTACAAACGTTTATGTCTTAAGAGAGGAAA
TATGAAGAGCAAAACAGTGCATGCTGGAGAGAGAAAGCTGATACAAATATAAATGAAACAATAATTGGAAAAATTGAGAAACTACTCATTTTCTAAATTACTC
ATGTATTTTCCTAGAATTTAAGTCTTTTAATTTTTGATAAATCCCAATGTGAGACAAGATAAGTATTAGTGATGGTATGAGTAATTAATATCTGTTATATAAT
ATTCATTTTCATAGTGGAAGAAATAAAATAAAGGTTGTGATGATTGTTGATTATTTTTTCTAGAGGGGTTGTCAGGGAAAGAAATTGCTTTTTTTCATTCTCT
CTTTCCACTAAGAAAGTTCAACTATTAATTTAGGCACATACAATAATTACTCCATTCTAAAATGCCAAAAAGGTAATTTAAGAGACTTAAAACTGAAAAGTTT
AAGATAGTCACACTGAACTATATTAAAAAATCCACAGGGTGGTTGGAACTAGGCCTTATATTAAAGAGGCTAAAAATTGCAATAAGACCACAGGCTTTAAATA
TGGCTTTAAACTGTGAAAGGTGAAACTAGAATGAATAAAATCCTATAAATTTAAATCAAAAGAAAGAAACAAACTGAAATTAAAGTTATTATACAAGAATATG
GTGGCCTGGATCTAGTGAACATATAGTAAAGATAAAACAGAATATTTCTGAAAAATCCTGGAAAATCTTTTGGGCTAACCTGAAAACAGTATATTTGAAACTA
TTTTTAAAATGCAGTGATACTAGAAATATTTTAGAATCATATGTA
```

…from sequence on chromosome 7 stretching from base positions 54,020,442 to 54,022,443.

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Single nucleotide polymorphisms (SNPs)

```
GAAATAATTAATGTTTTCCTTCCTTCTCCTATTTTGTCCTTTACTTCAATTTATTTATTTATTATTAATATTATTATTTTTTGAGACGGAGTTTCACTCTTGT
TGCCAACCTGGAGTGCAGTGGCGTGATCTCAGCTCACTGCACACTCCGCTTTCCGGTTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACA
GTCACACACCACCACGCCCGGCTAATTTTTGTATTTTTAGTAGAGTTGGGGTTTCACCATGTTGGCCAGACTGGTCTCGAACTCCTGACCTTGTGATCCGCCA
GCCTCTGCCTCCCAAAGAGCTGGGATTACAGGCGTGAGCCACCGCGCTCGGCCCTTTGCATCAATTTCTACAGCTTGTTTTCTTTGCCTGGACTTTACAAGTC
TTACCTTGTTCTGCCTTCAGATATTTGTGTGGTCTCATTCTGGTGTGCCAGTAGCTAAAAATCCATGATTTGCTCTCATCCCACTCCTGTTGTTCATCTCCTC
TTATCTGGGGTCACCTATCTCTTCGTGATTGCATTCTGATCCCCAGTACTTAGCATGTGCGTAACAACTCTGCCTCTGCTTTCCCAGGCTGTTGATGGGGTGC
TGTTCATGCCTCAGAAAAATGCATTGTAAGTTAAATTATTAAAGATTTTAAATATAGGAAAAAAGTAAGCAAACATAAGGAACAAAAAGGAAAGAACATGTAT
TCTAATCCATTATTTATTATACAATTAAGAAATTTGGAAACTTTAGATTACACTGCTTTTAGAGATGGAGATGTAGTAAGTCTTTTACTCTTTACAAAATACA
TGTGTTAGCAATTTTGGGAAGAATAGTAACTCACCCGAACAGTGTAATGTGAATATGTCACTTACTAGAGGAAAGAAGGCACTTGAAAAACATCTCTAAACCG
TATAAAAACAATTACATCATAATGATGAAAACCCAAGGAATTTTTTTAGAAAACATTACCAGGGCTAATAACAAAGTAGAGCCACATGTCATTTATCTTCCCT
TTGTGTCTGTGTGAGAATTCTAGAGTTATATTTGTACATAGCATGGAAAAATGAGAGGCTAGTTTATCAACTAGTTCATTTTTAAAAGTCTAACACATCCTAG
GTATAGGTGAACTGTCCTCCTGCCAATGTATTGCACATTTGTGCCCAGATCCAGCATAGGGTATGTTTGCCATTTACAAACGTTTATGTCTTAAGAGAGGAAA
TATGAAGAGCAAAACAGTGCATGCTGGAGAGAGAAAGCTGATACAAATATAAATGAAACAATAATTGGAAAAATTGAGAAACTACTCATTTTCTAAATTACTC
ATGTATTTTCCTAGAATTTAAGTCTTTTAATTTTTGATAAATCCCAATGTGAGACAAGATAAGTATTAGTGATGGTATGAGTAATTAATATCTGTTATATAAT
ATTCATTTTCATAGTGGAAGAAATAAAATAAAGGTTGTGATGATTGTTGATTATTTTTTCTAGAGGGGTTGTCAGGGAAAGAAATTGCTTTTTTTCATTCTCT
CTTTCCACTAAGAAAGTTCAACTATTAATTTAGGCACATACAATAATTACTCCATTCTAAAATGCCAAAAAGGTAATTTAAGAGACTTAAAACTGAAAAGTTT
AAGATAGTCACACTGAACTATATTAAAAAATCCACAGGGTGGTTGGAACTAGGCCTTATATTAAAGAGGCTAAAAATTGCAATAAGACCACAGGCTTTAAATA
TGGCTTTAAACTGTGAAAGGTGAAACTAGAATGAATAAAATCCTATAAATTTAAATCAAAAGAAAGAAACAAACTAAAATTAAAGTTATTATACAAGAATATG
GTGGCCTGGATCTAGTGAACATATAGTAAAGATAAAACAGAATATTTCTGAAAAATCCTGGAAAATCTTTTGGGCTAACCTGAAAACAGTATATTTGAAACTA
TTTTTAAAATGCAGTGATACTAGAAATATTTTAGAATCATATGTA
```

Three SNPs are located at positions 54,020,598, 54,020,971 and 54,022,268.

SNPs

- Less polymorphic/informative
- More stable inheritance
- ~1 SNP / 1,250 kb between any

  two genomes
- 2.5 million between two genomes
- Exist in coding regions

Human Genetic Variation

- What types of variants exist?
- How are variants found?
- How are variants scored?
- How are variants used?

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Microsatellite identification

- Databases
  - Marshfield Clinic

    http://research.marshfieldclinic.org/genetics/
  - Genome DataBase

    http://gdbwww.gdb.org/
  - Cooperative Human Linkage Center

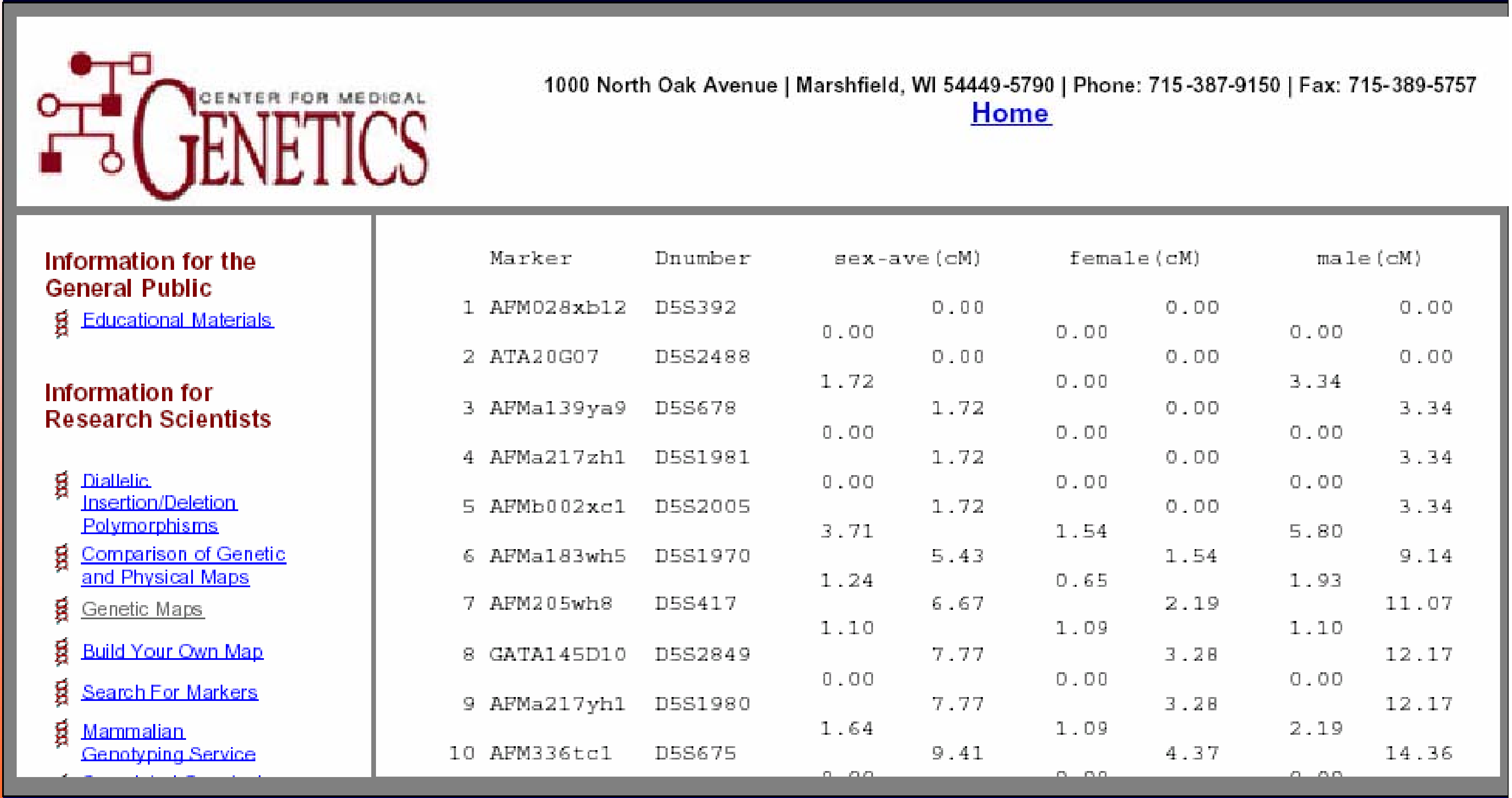
    http://lpg.nci.nih.gov/CHLC/
  - Genethon

    http://www.genethon.fr/eng/indeng.html
  - Hapmap for human SNP distribution and profile

    http://www.hapmap.org

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Microsatellite identification: database



1000 North Oak Avenue | Marshfield, WI 54449-5790 | Phone: 715-387-9150 | Fax: 715-389-5757

Home

Information for the General Public

- Educational Materials

Information for Research Scientists

- Diallelic Insertion/Deletion Polymorphisms
- Comparison of Genetic and Physical Maps
- Genetic Maps
- Build Your Own Map
- Search For Markers
- Mammalian Genotyping Service

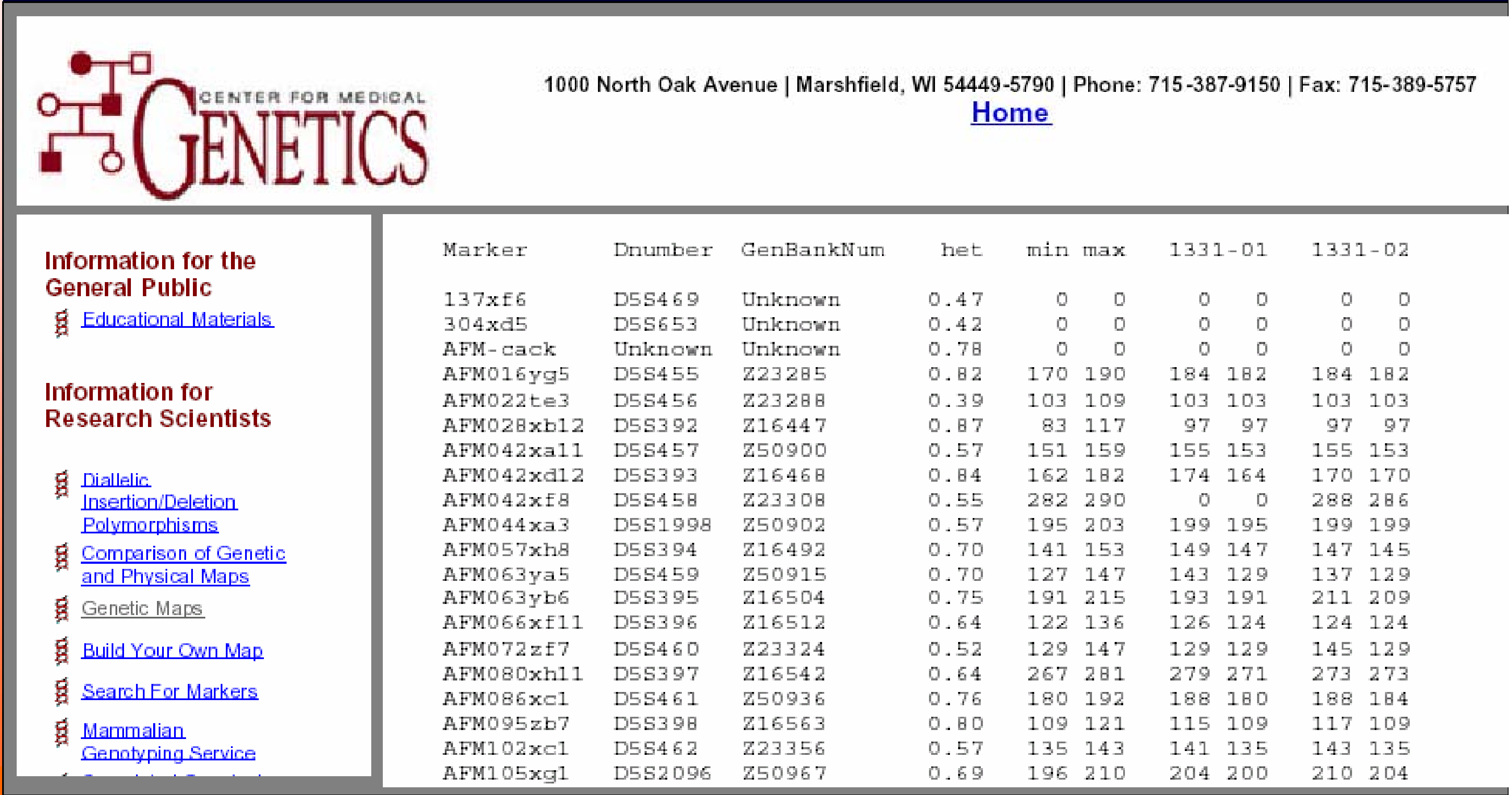
	Marker	Dnumber	sex-ave(cM)	female(cM)	male(cM)
1	AFM028xb12	D5S392	0.00	0.00	0.00
			0.00	0.00	0.00
2	ATA20G07	D5S2488	0.00	0.00	0.00
			1.72	0.00	3.34
3	AFMa139ya9	D5S678	1.72	0.00	3.34
			0.00	0.00	0.00
4	AFMa217zh1	D5S1981	1.72	0.00	3.34
			0.00	0.00	0.00
5	AFMb002xc1	D5S2005	1.72	0.00	3.34
			3.71	1.54	5.80
6	AFMa183wh5	D5S1970	5.43	1.54	9.14
			1.24	0.65	1.93
7	AFM205wh8	D5S417	6.67	2.19	11.07
			1.10	1.09	1.10
8	GATA145D10	D5S2849	7.77	3.28	12.17
			0.00	0.00	0.00
9	AFMa217yh1	D5S1980	7.77	3.28	12.17
			1.64	1.09	2.19
10	AFM336tc1	D5S675	9.41	4.37	14.36
			0.00	0.00	0.00

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Microsatellite identification: database

(probe name, locus name, GenBank accession number, heterozygosity, allele size range and genotypes for CEPH individuals 1331-01 and 1331-02, for each marker).

***C**entre d'**E**tude du **P**olymorphisme **H**umain* (CEPH)



CENTER FOR MEDICAL GENETICS

1000 North Oak Avenue | Marshfield, WI 54449-5790 | Phone: 715-387-9150 | Fax: 715-389-5757

Home

Information for the General Public

- Educational Materials

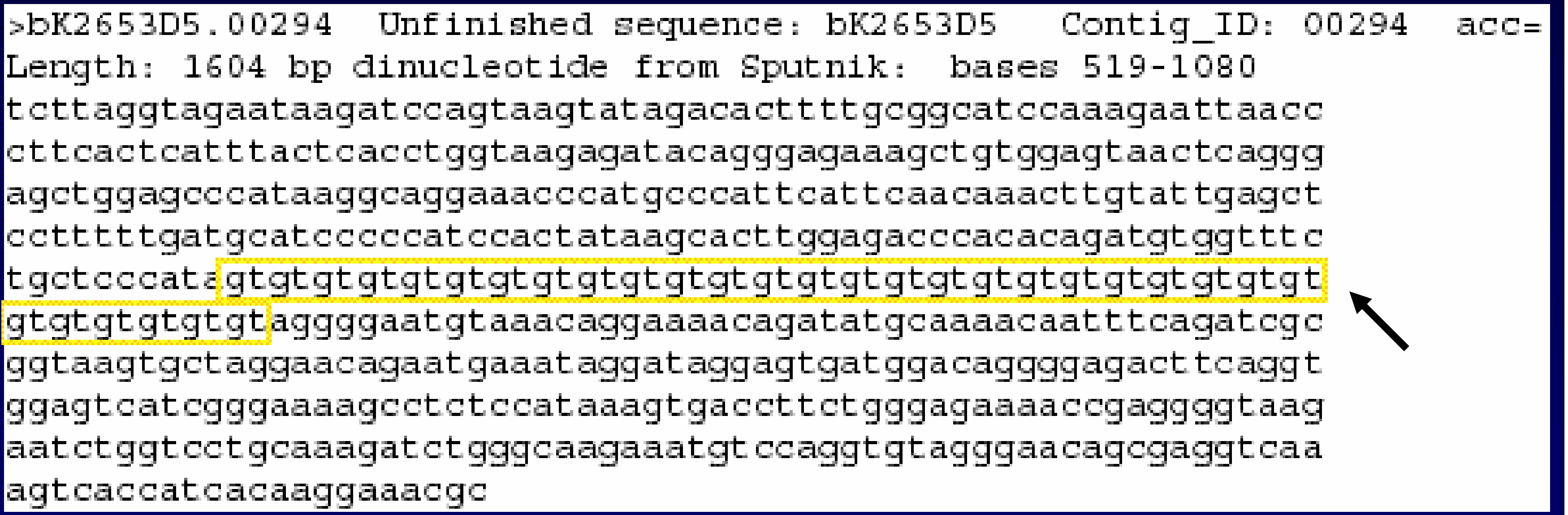
Information for Research Scientists

- Diallelic Insertion/Deletion Polymorphisms
- Comparison of Genetic and Physical Maps
- Genetic Maps
- Build Your Own Map
- Search For Markers
- Mammalian Genotyping Service

Marker	Dnumber	GenBankNum	het	min	max	1331-01		1331-02	
137xf6	D5S469	Unknown	0.47	0	0	0	0	0	0
304xd5	D5S653	Unknown	0.42	0	0	0	0	0	0
AFM-cack	Unknown	Unknown	0.78	0	0	0	0	0	0
AFM016yg5	D5S455	Z23285	0.82	170	190	184	182	184	182
AFM022te3	D5S456	Z23288	0.39	103	109	103	103	103	103
AFM028xb12	D5S392	Z16447	0.87	83	117	97	97	97	97
AFM042xa11	D5S457	Z50900	0.57	151	159	155	153	155	153
AFM042xd12	D5S393	Z16468	0.84	162	182	174	164	170	170
AFM042xf8	D5S458	Z23308	0.55	282	290	0	0	288	286
AFM044xa3	D5S1998	Z50902	0.57	195	203	199	195	199	199
AFM057xh8	D5S394	Z16492	0.70	141	153	149	147	147	145
AFM063ya5	D5S459	Z50915	0.70	127	147	143	129	137	129
AFM063yb6	D5S395	Z16504	0.75	191	215	193	191	211	209
AFM066xf11	D5S396	Z16512	0.64	122	136	126	124	124	124
AFM072zf7	D5S460	Z23324	0.52	129	147	129	129	145	129
AFM080xh11	D5S397	Z16542	0.64	267	281	279	271	273	273
AFM086xc1	D5S461	Z50936	0.76	180	192	188	180	188	184
AFM095zb7	D5S398	Z16563	0.80	109	121	115	109	117	109
AFM102xc1	D5S462	Z23356	0.57	135	143	141	135	143	135
AFM105xg1	D5S2096	Z50967	0.69	196	210	204	200	210	204

Microsatellite identification: from sequence

Sputnik: searches DNA sequence files in Fasta format for microsatellite repeats.



```
>bK2653D5.00294  Unfinished sequence: bK2653D5   Contig_ID: 00294   acc=
Length: 1604 bp dinucleotide from Sputnik:  bases 519-1080
tcttaggtagaataagatccagtaagtatagacactttgcggcatccaaagaattaacc
cttcactcatttactcacctggtaagagatacagggagaaagctgtggagtaactcaggg
agctggagcccataaggcaggaaacccatgcccattcattcaacaaacttgtattgagct
ccttttgatgcatccccatccactataagcacttggagacccacacagatgtggtttc
tgctcccatagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
gtgtgtgtgtgtaggggaatgtaaacaggaaaacagatatgcaaaacaatttcagatcgc
ggtaagtgctaggaacagaatgaaataggataggagtgatggacaggggagacttcaggt
ggagtcatcgggaaaagcctctccataaagtgaccttctgggagaaaaccgaggggtaag
aatctggtcctgcaaagatctgggcaagaaatgtccaggtgtagggaacagcgaggtcaa
agtcaccatcacaaggaaacgc
```

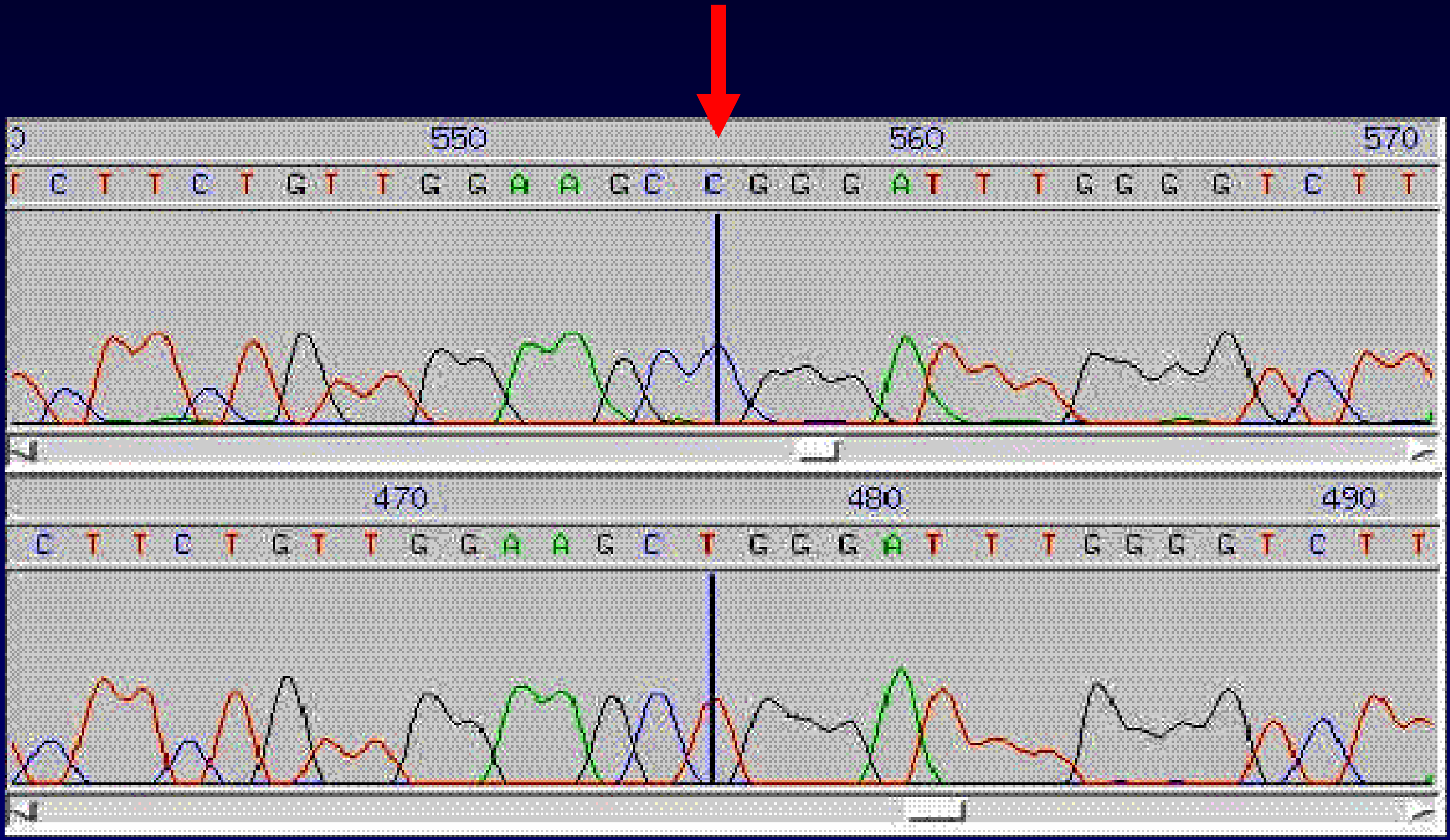
Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

SNP identification

- Sequencing
- Databases

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

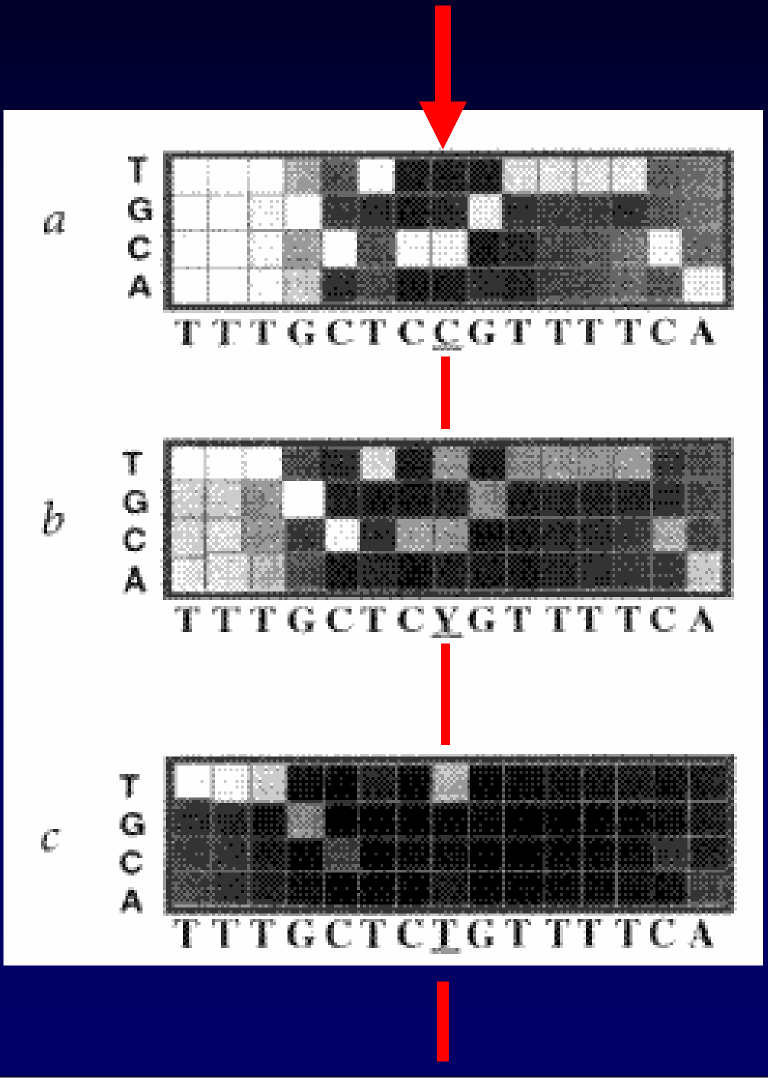
SNP identification: sequencing


550
560
570
T C T T C T G T T G G A A G C C G G G A T T T G G G G T C T T
470
480
490
C T T C T G T T G G A A G C T G G G A T T T G G G G T C T T

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

SNP identification: Sequencing chip

...GCTCCGTTT...
...GCTCTGTTT...

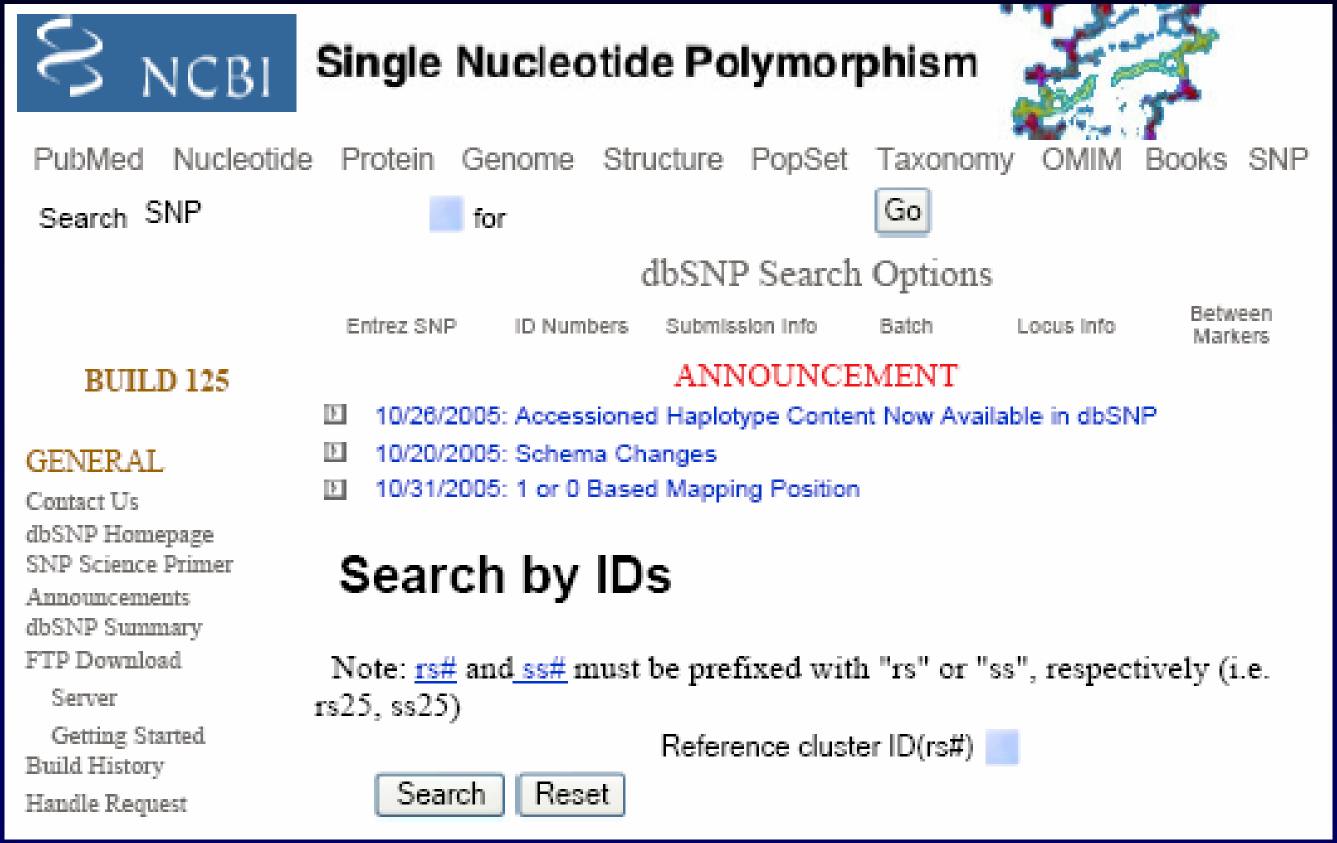


SNP identification: databases

- dbSNP
  - >27,189,291 submitted; 4,236,590 reference
- **T**he **S**NP **C**onsortium (TSC)
- Human Gene Variation base (HGVbase)
- CGAP Genetic Annotation Initiative (CGAP-GAI)
- Japanese SNPs (JSNP)

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

SNP identification: databases



NCBI **Single Nucleotide Polymorphism**

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP
Search SNP for
Go

dbSNP Search Options

BUILD 125

ANNOUNCEMENT

- 10/26/2005: Accessioned Haplotype Content Now Available in dbSNP
- 10/20/2005: Schema Changes
- 10/31/2005: 1 or 0 Based Mapping Position

GENERAL

Search by IDs

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

Reference cluster ID(rs#)

Searching dbSNP:

- by gene name/nomenclature association
- by map location
- as a BLAST operation on dbSNP using a candidate sequence

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Conclusions from TSC data

2.3M SNPs: 1,992,262M unique in map

Intergenic:	1,668,651	(84%)
Intragenic:	323,611	(16%)
Exonic	33,405	(1.7%)
Intronic	290,206	(14.5%)
Splice	130	

Conclusions from TSC data

Of 1,500 coding SNPs examined:

Silent	45%	1/600bp	2 / gene
Conservative	16%		
Non-conserved	38%	1/600bp	2 / gene
Nonsense	1%		

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

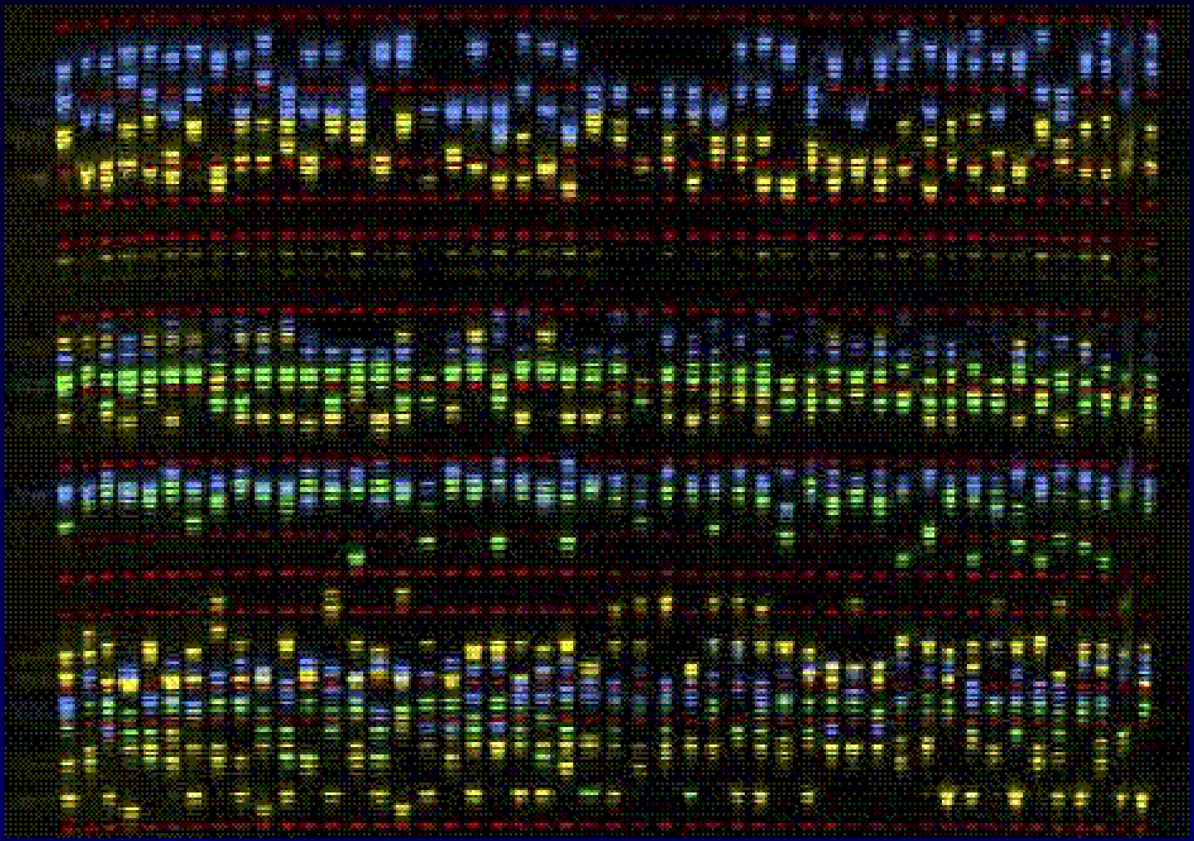
NUS National University of Singapore

Human Genetic Variation

- What types of variants exist?
- How are variants found?
- How are variants scored?
- How are variants used?

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Scoring Microsatellites



Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Scoring SNP

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

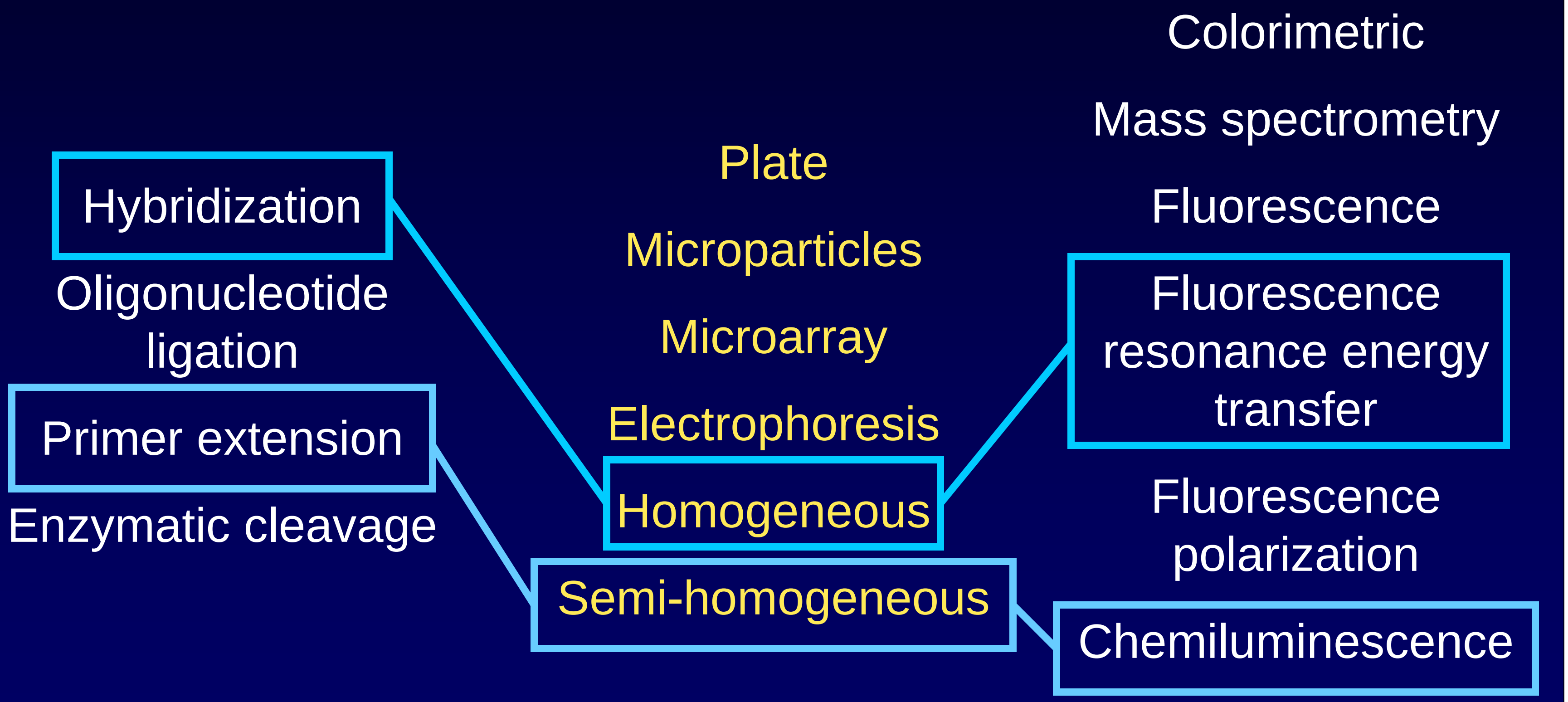
Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Scoring SNP

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

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Overview of SNP typing methods



Hybridization

Oligonucleotide ligation

Primer extension

Enzymatic cleavage

Plate

Microparticles

Microarray

Electrophoresis

Homogeneous

Semi-homogeneous

Colorimetric

Mass spectrometry

Fluorescence

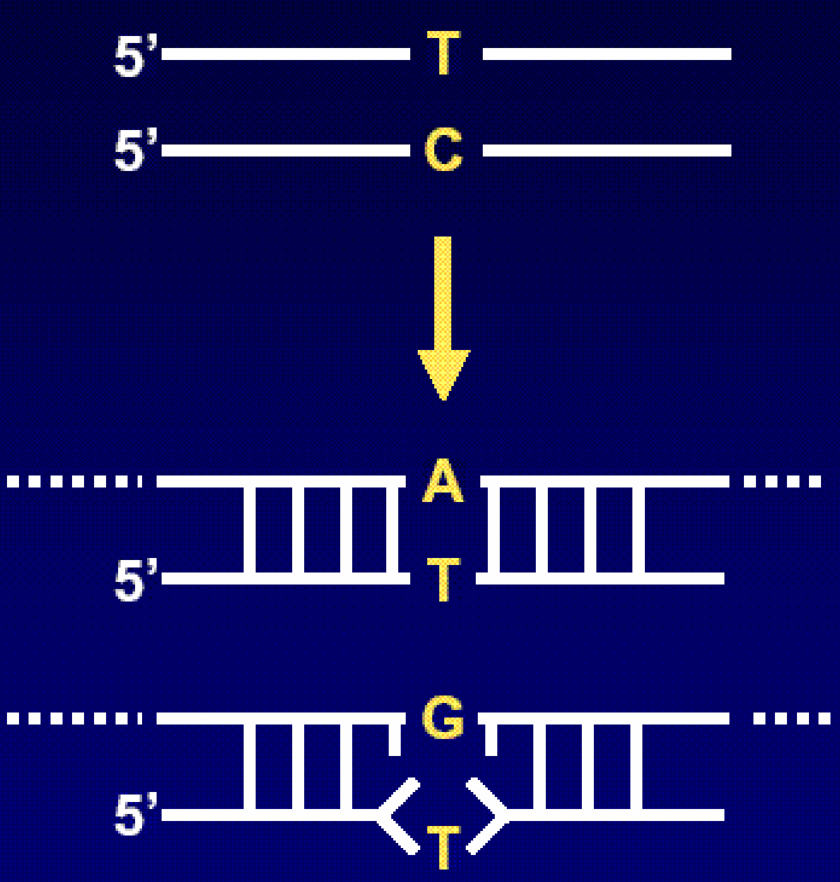
Fluorescence resonance energy transfer

Fluorescence polarization

Chemiluminescence

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Hybridization



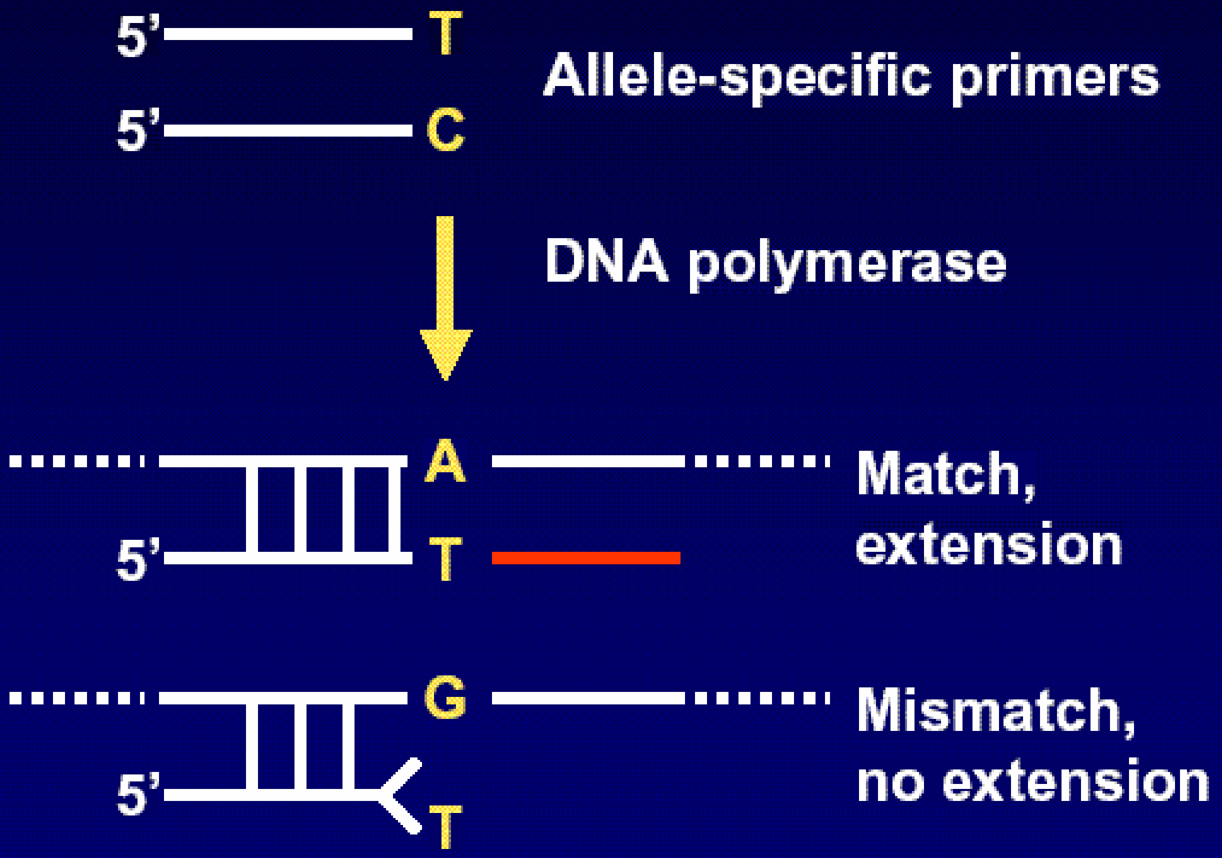
Allele-specific oligonucleotide probes

Match, stable

Mismatch, unstable

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Allele specific PCR



Allele-specific primers

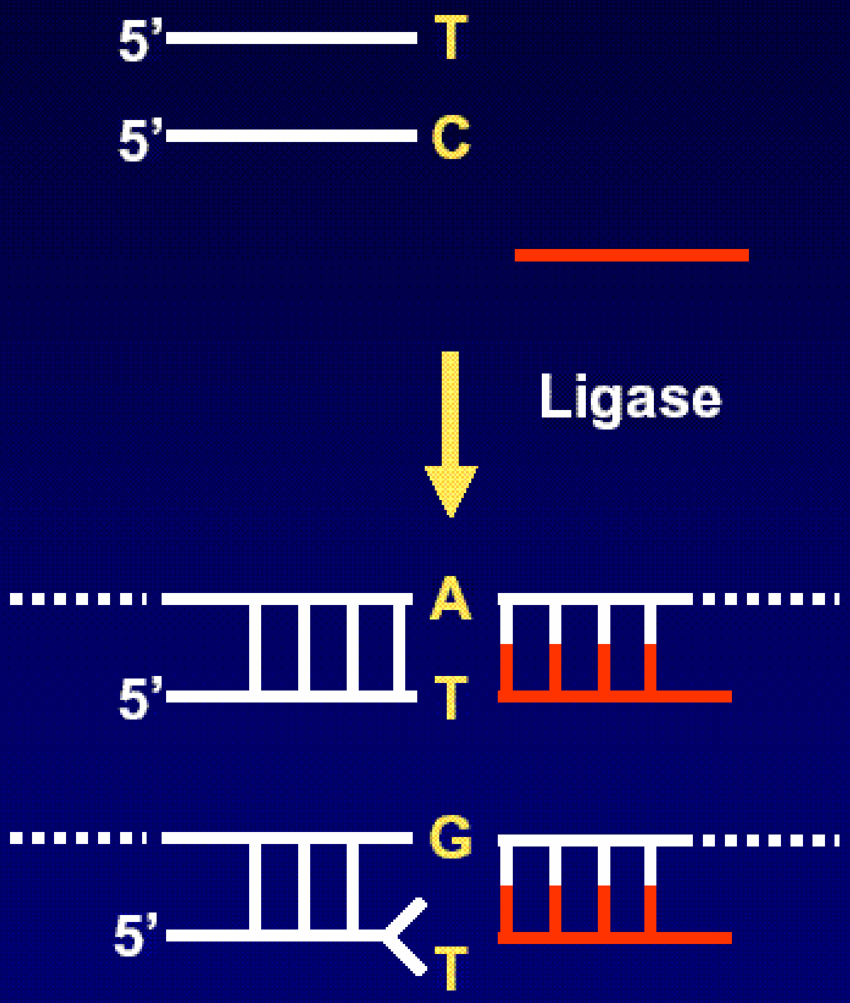
DNA polymerase

Match,
extension

Mismatch,
no extension

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Oligoneuclotide Ligation Assay (OLA)



Allele-specific ligation probes

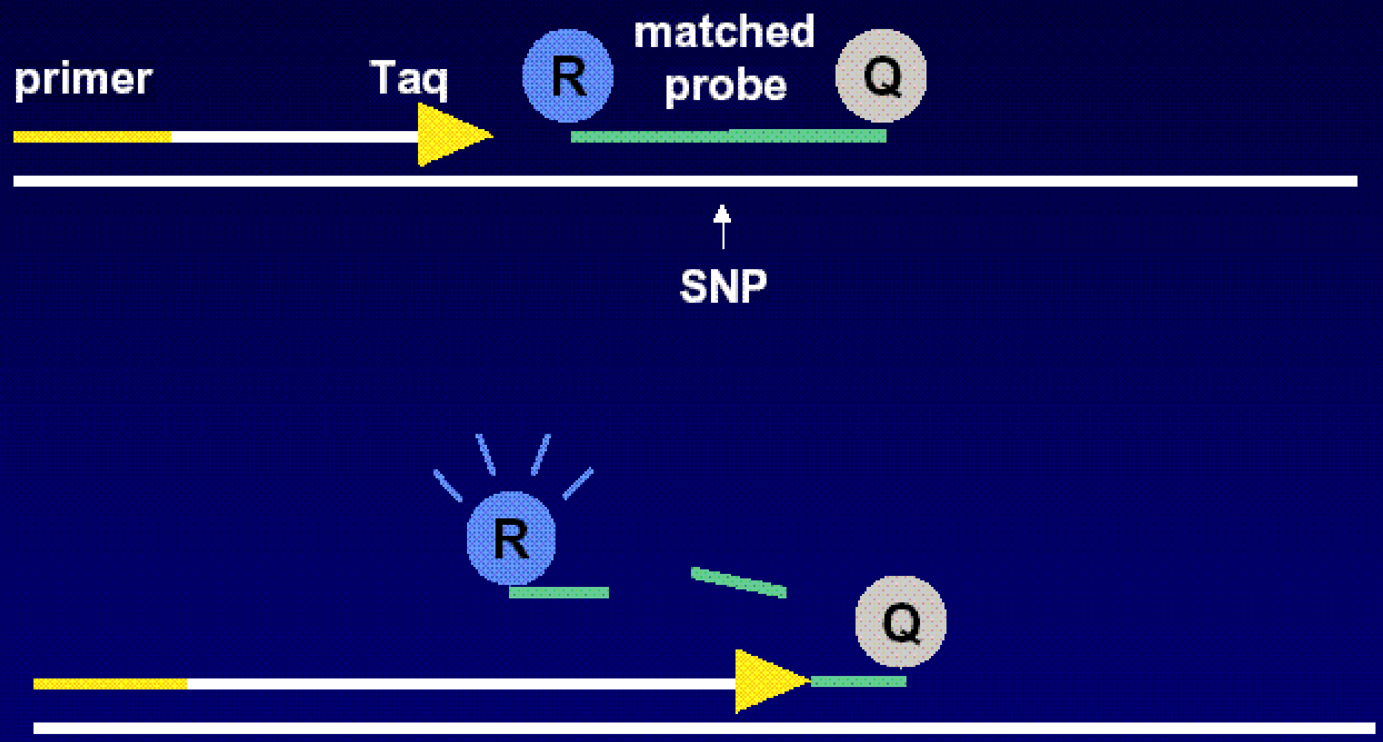
Adjacent ligation probe

Ligase

Match, ligation

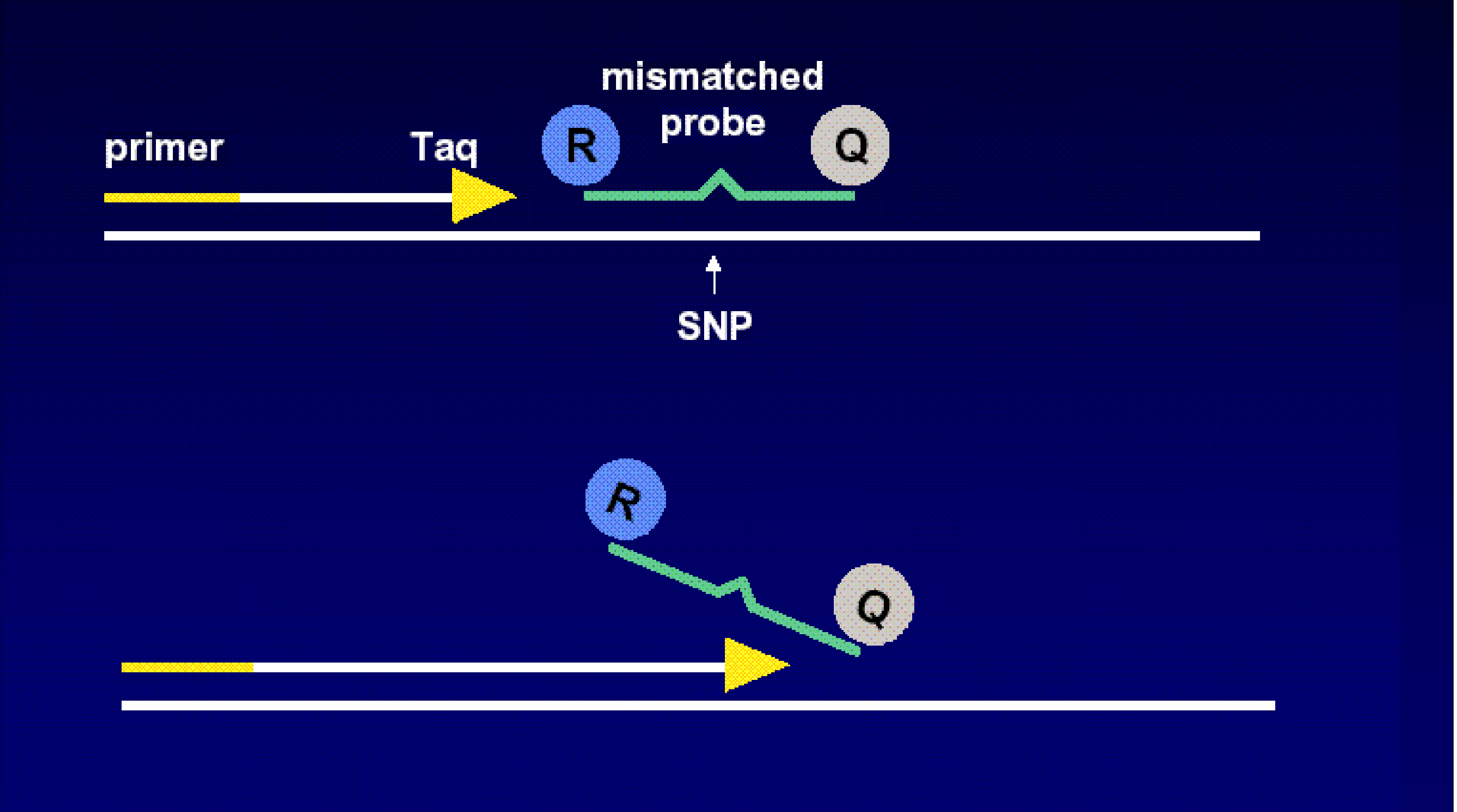
Mismatch, no ligation

Fluorescence resonance energy transfer (FRET)

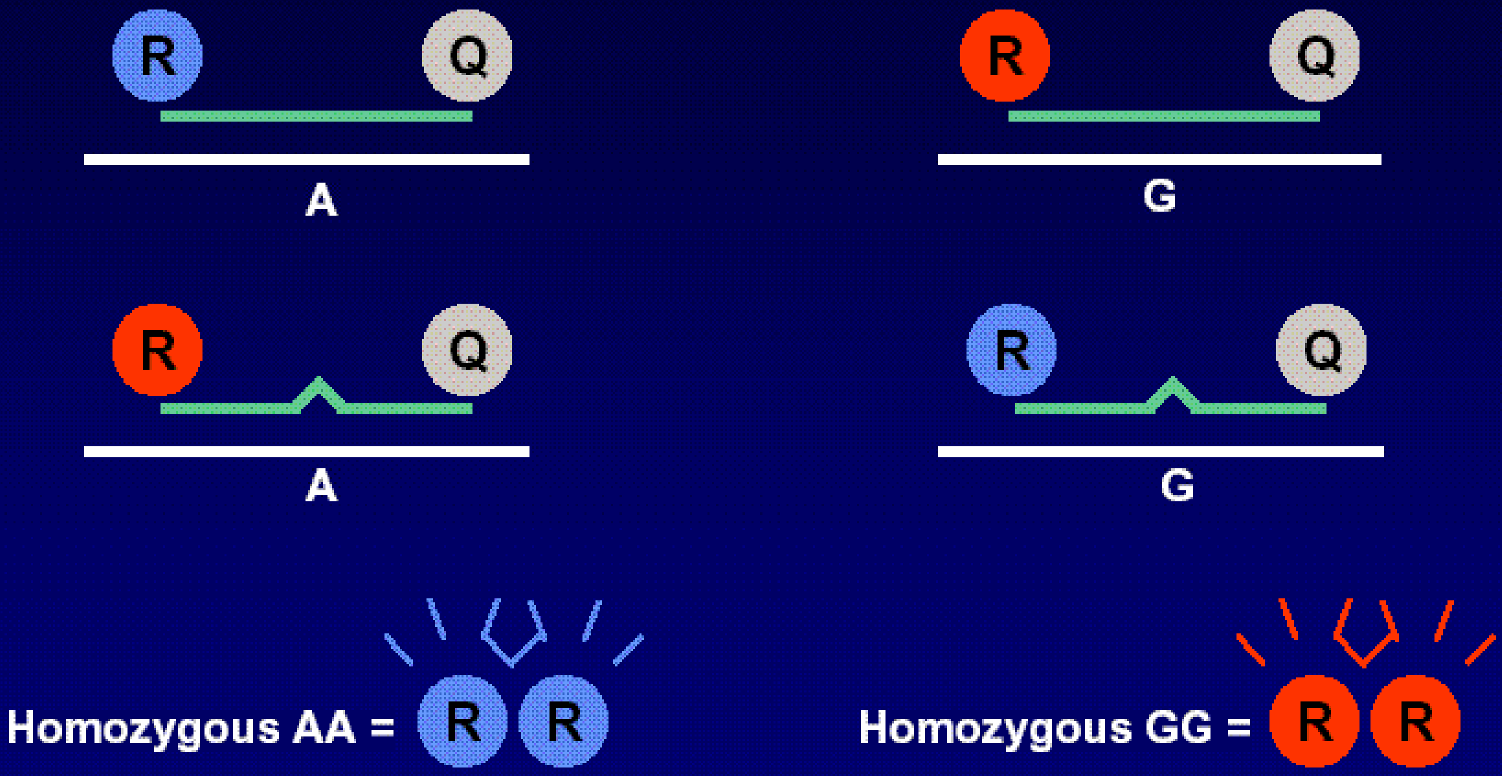


Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Fluorescence resonance energy transfer (FRET)

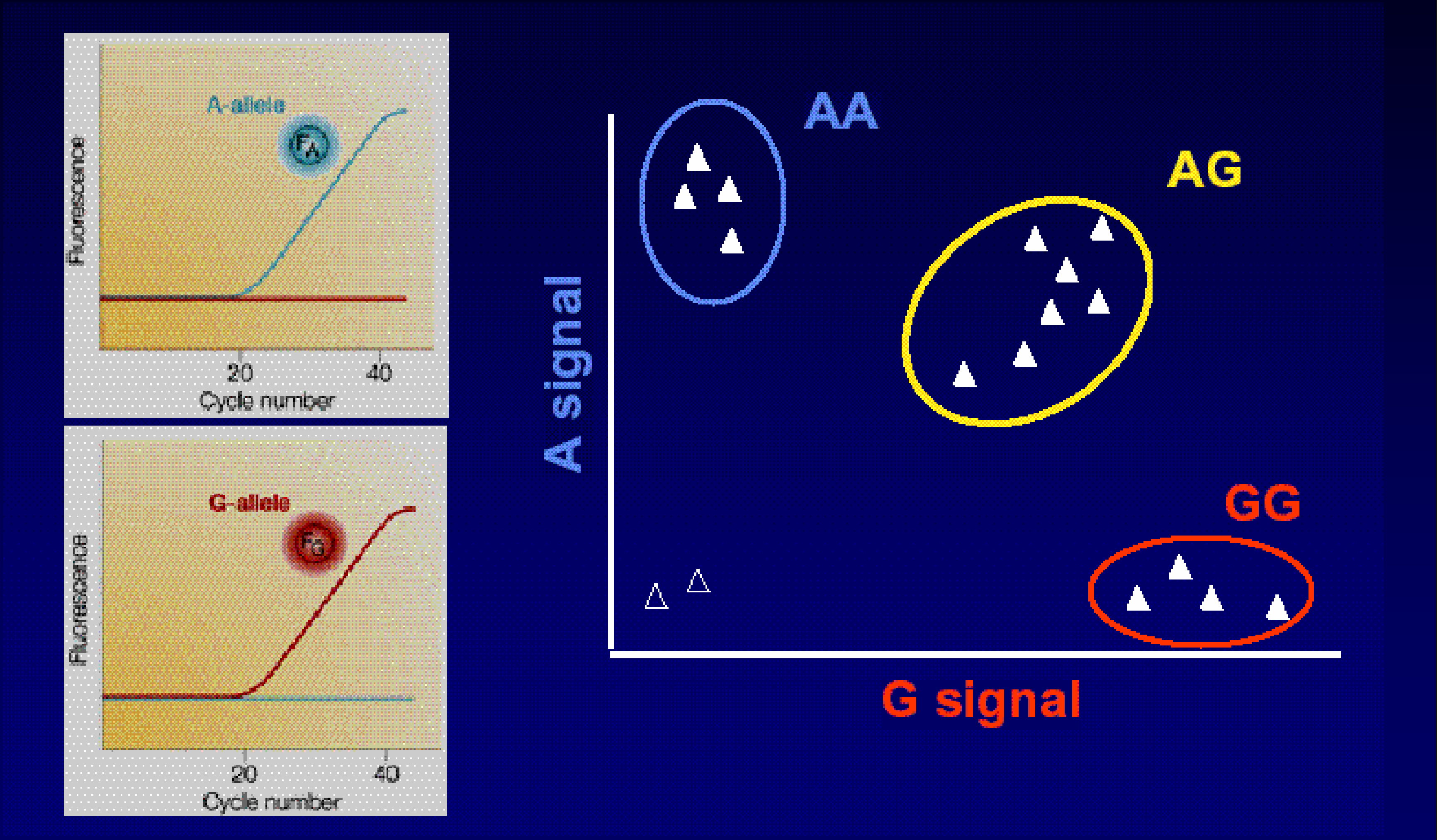


TaqMan competing probes



Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

TaqMan genotype scoring

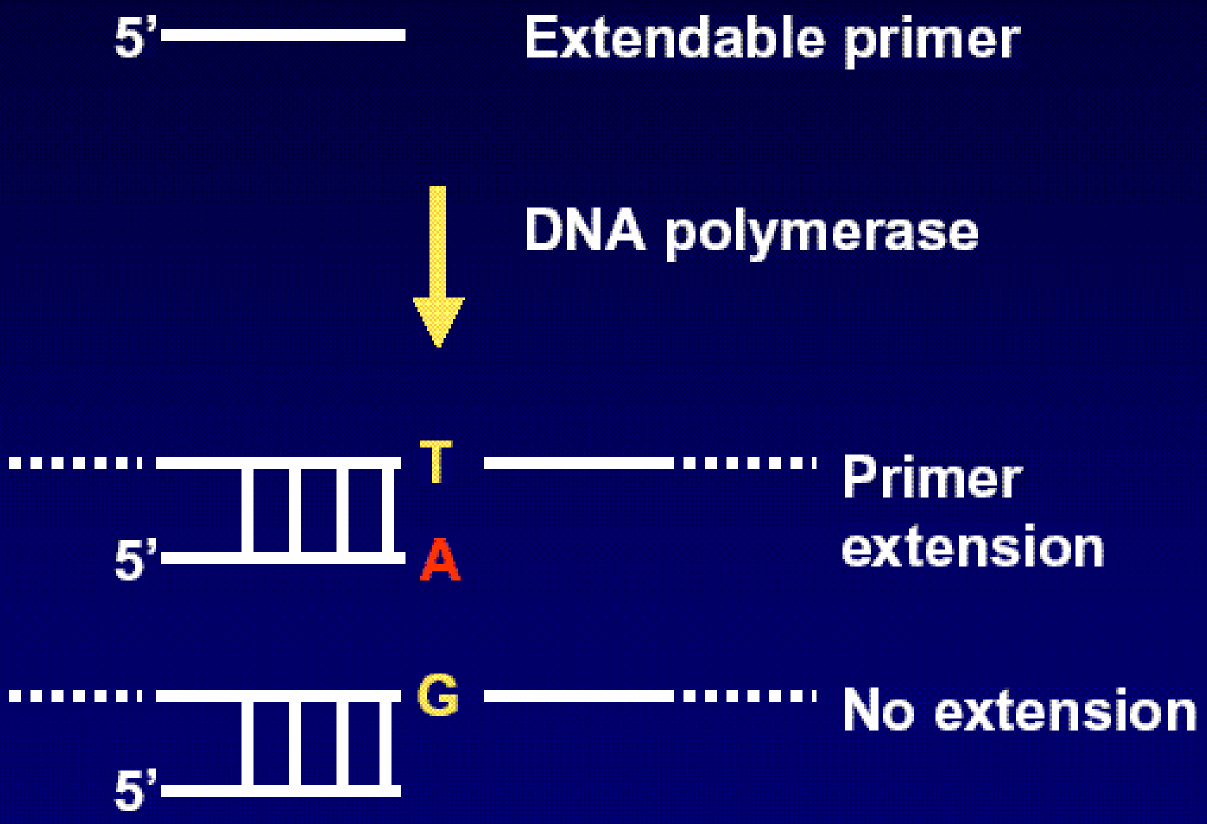


Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

TaqMan

- Advantages:
  - Simple to perform
  - Closed-tube system
  - Accurate quantification
- Disadvantages
  - Expensive probes
  - Assays require optimization

Primer extension = Minisequencing



5'

Extendable primer

DNA polymerase

T

5'

A

Primer extension

G

5'

No extension

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

NUS National University of Singapore

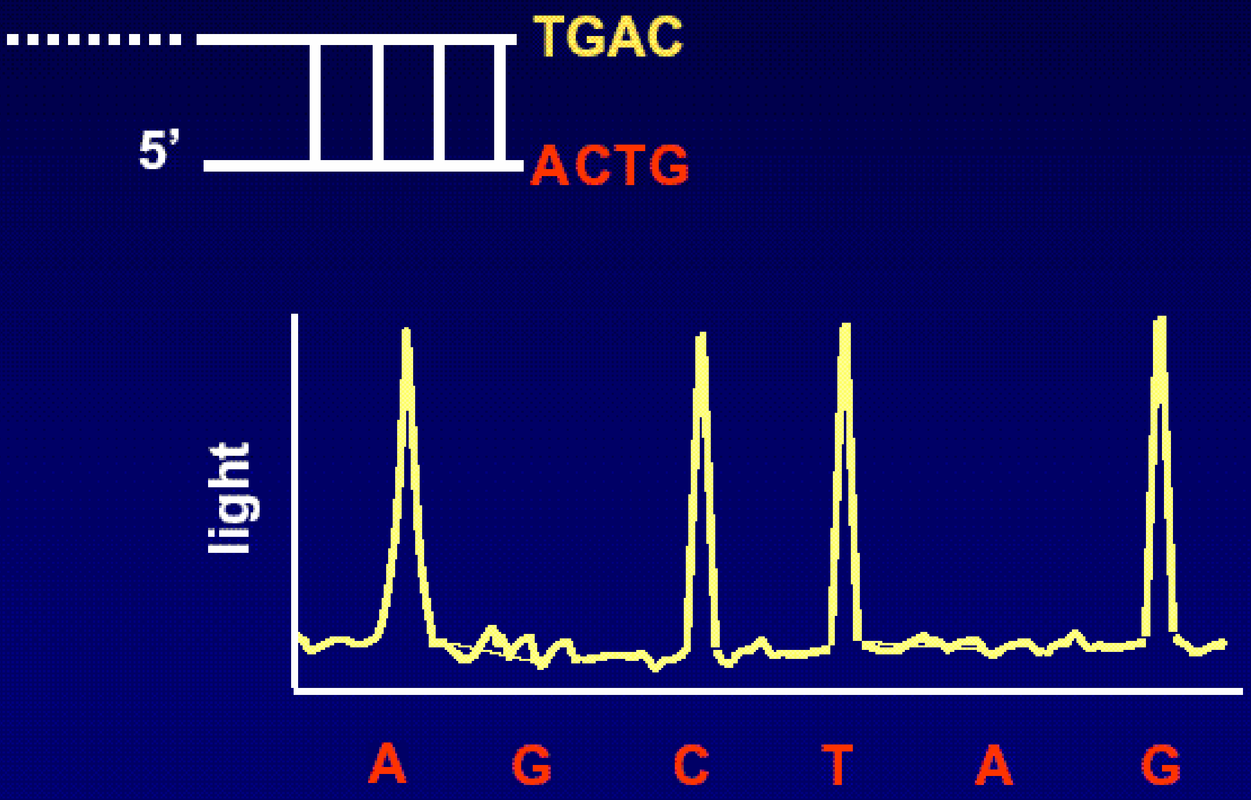
Pyrosequencing

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

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Pyrosequencing

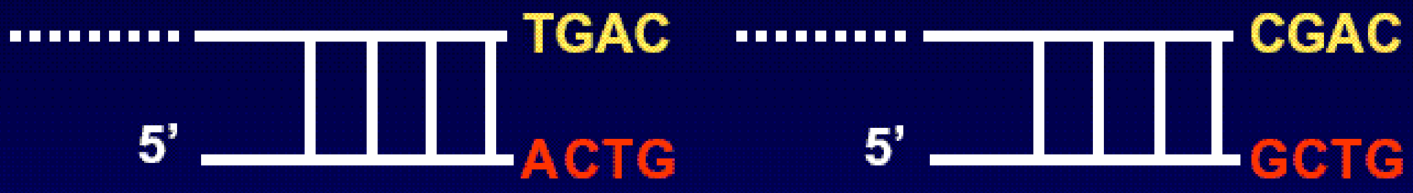
…[T/C]GAC…

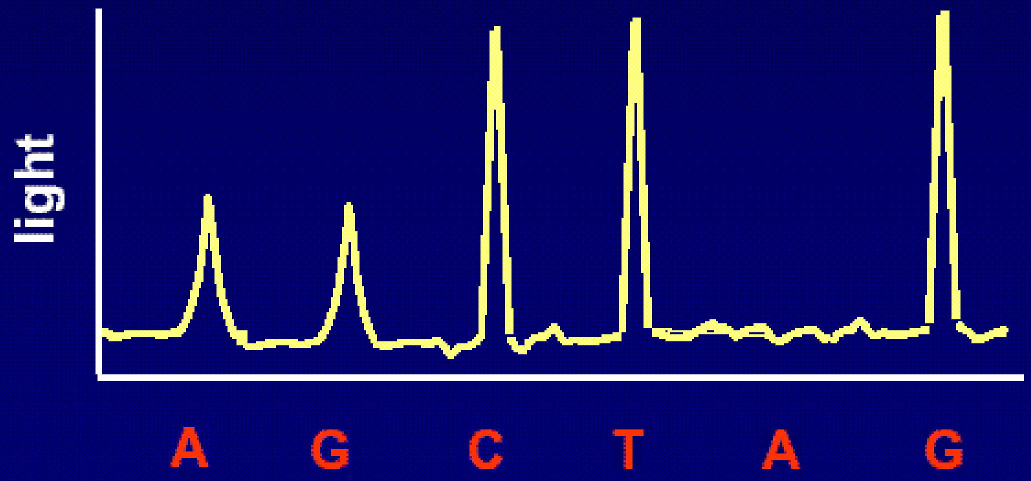


Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Pyrosequencing

…[T/C]GAC…





Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

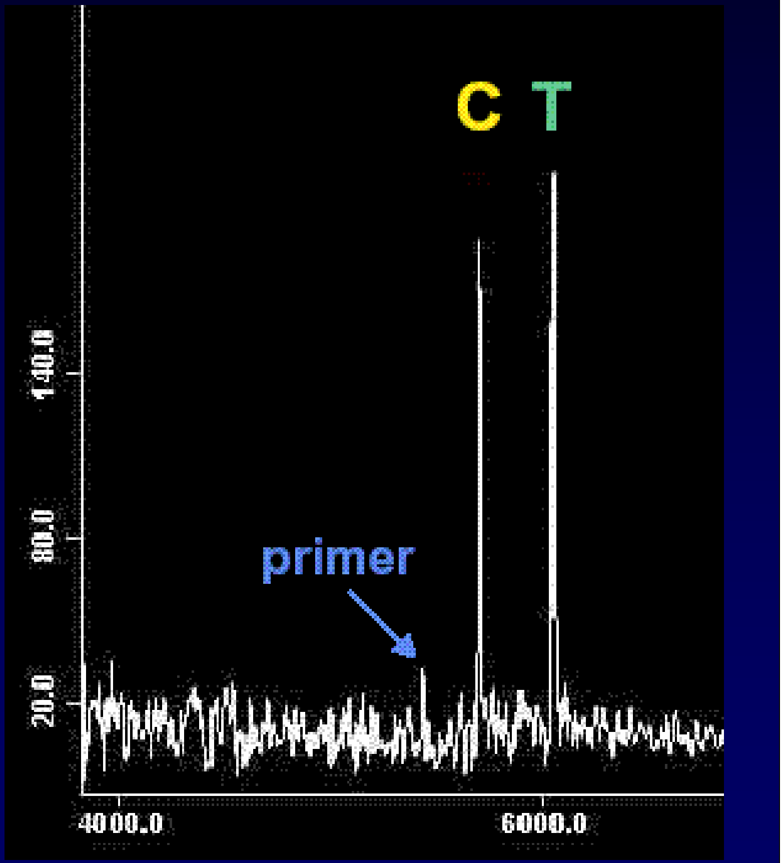
Pyrosequencing

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

Primer extension: mass spectrometry

Primer extension reactions designed to generate different sized products

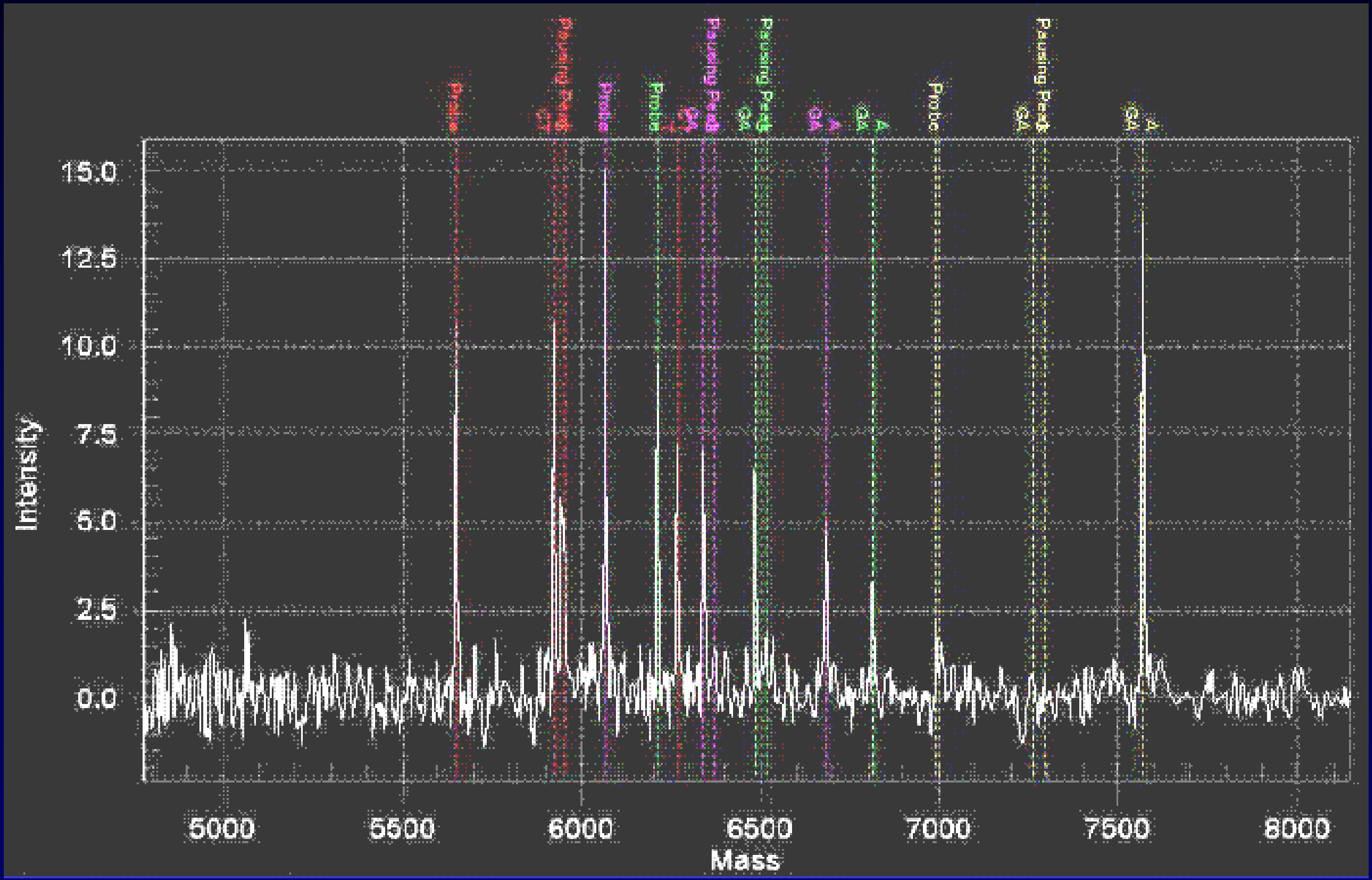
	Mass in Daltons
GGACCTGGAGCCCCCACC	**5430.5**
GGACCTGGAGCCCCCACCC	**5703.7**
GGACCTGGAGCCCCCACCTG	**6047.9**



Primer extension: mass spectrometry

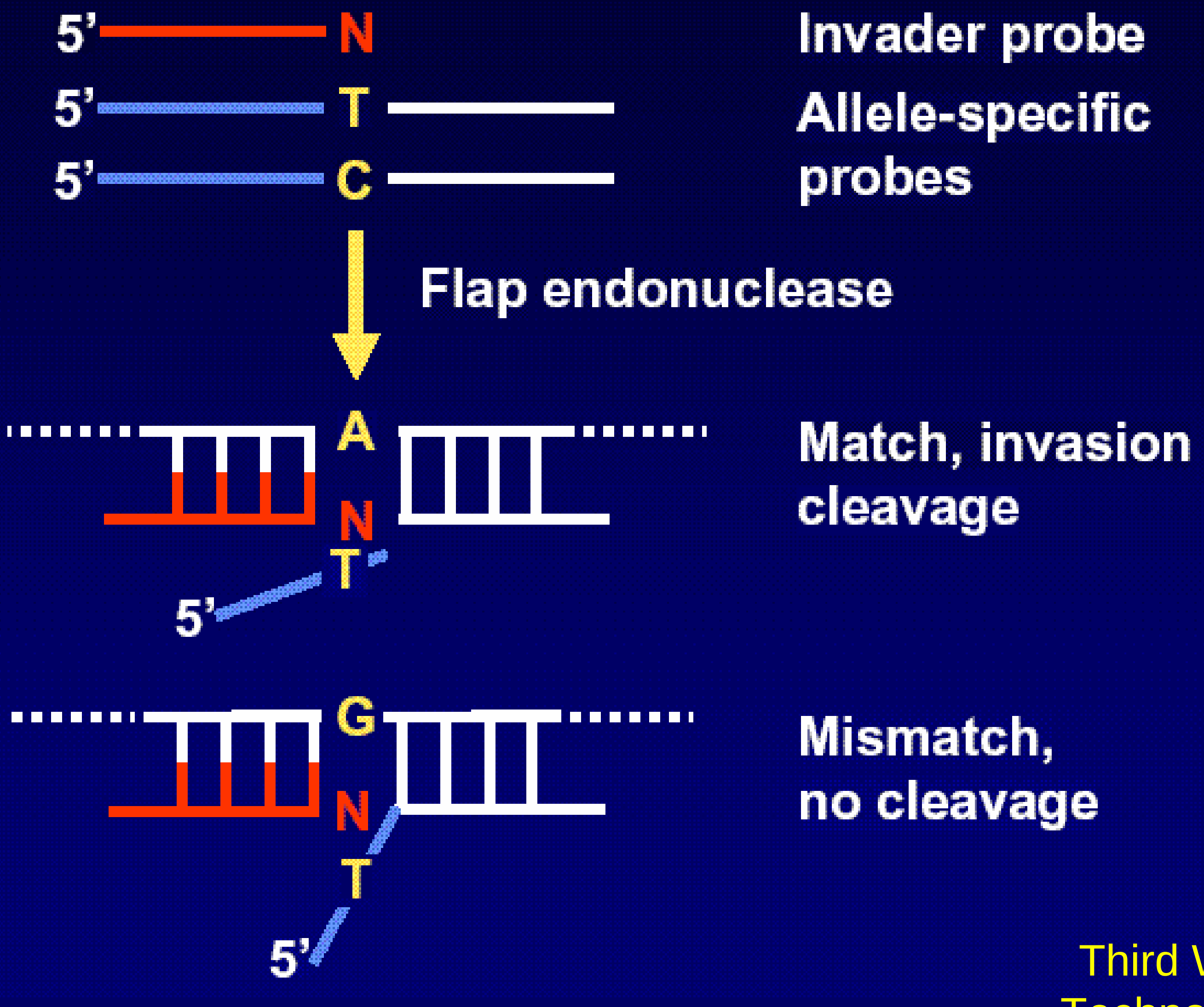
- Advantages:
  - Accurate
  - Automated assay design
  - Fast automated data collection
  - Multiplexing capacity
- Disadvantages
  - Expensive instruments, consumables
  - Extensive post-PCR processing

Mass spectrometry multiplexing



Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Invasive cleavage of oligo probes



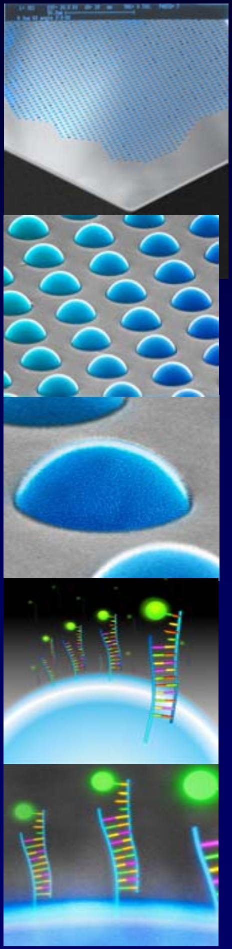
Third Wave Technologies

Invasive cleavage of oligo probes

- Advantages
  - Avoids need for PCR
- Disadvantages
  - Still requires larger amount of DNA
  - Tricky probe design

Illumina Solutions to Whole Genome Genotyping

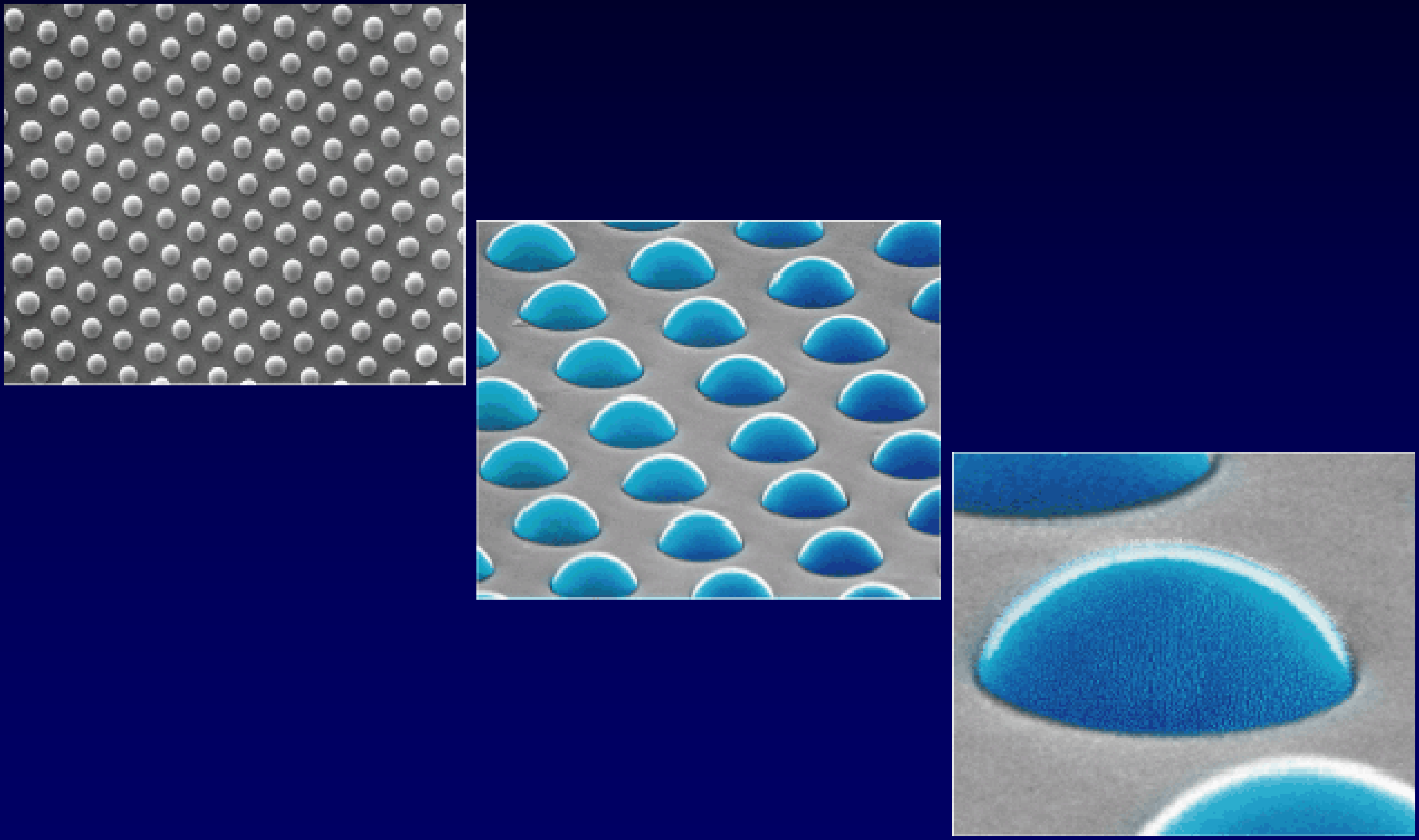
illumina®



Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

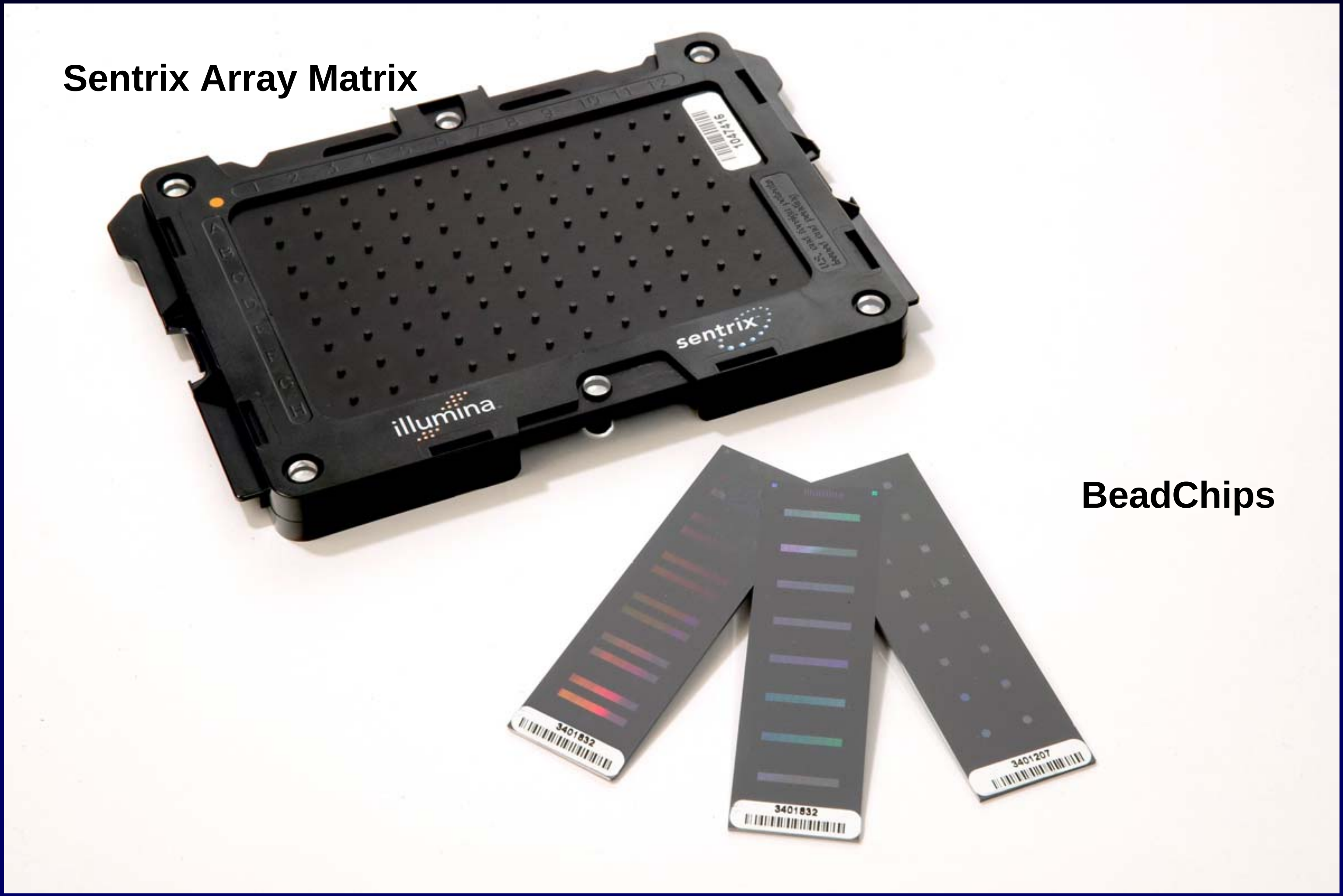
NUS National University of Singapore

Bead Arrays: Oligo coated Beads in Wells



Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Array Formats



Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

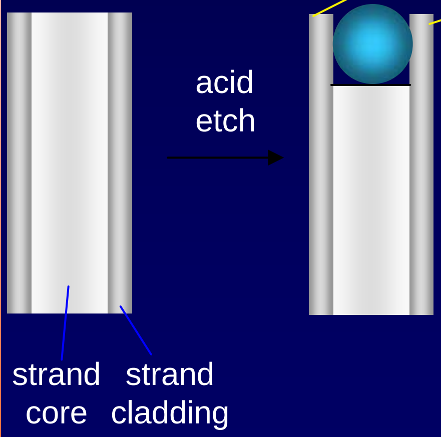
NUS National University of Singapore

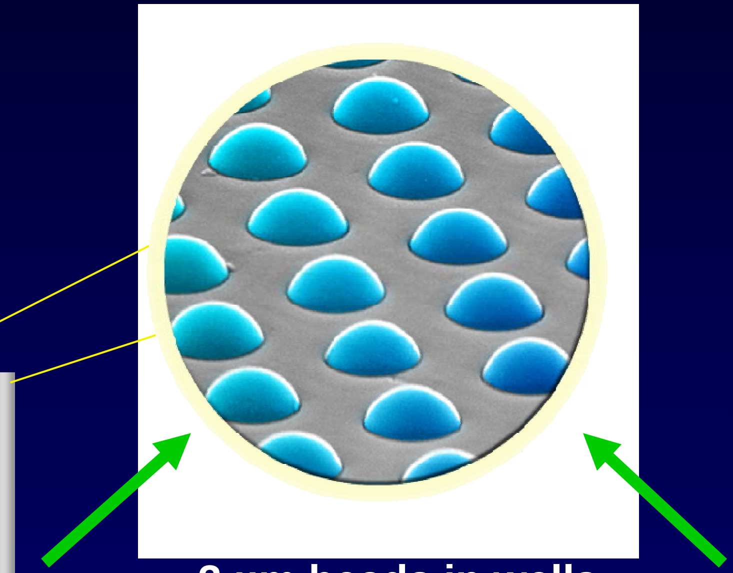
BeadArray: Microwell Fabrication

SAM

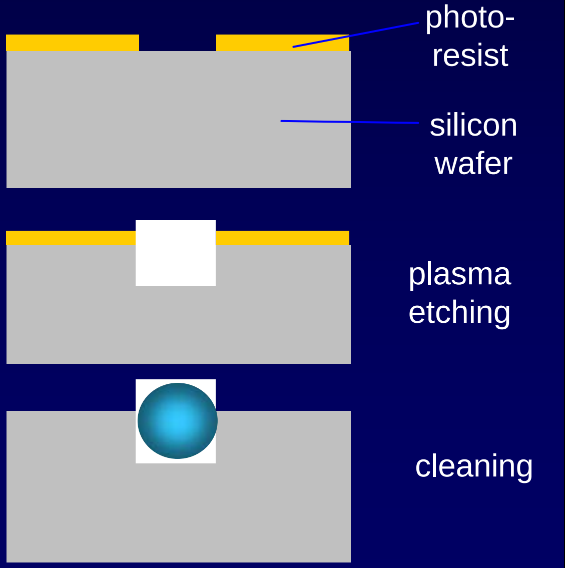
BeadChip

Optical fiber





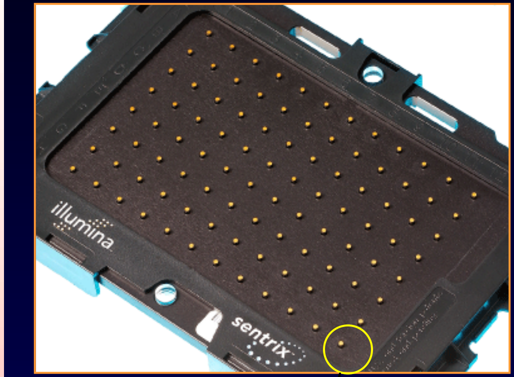
3 µm beads in wells

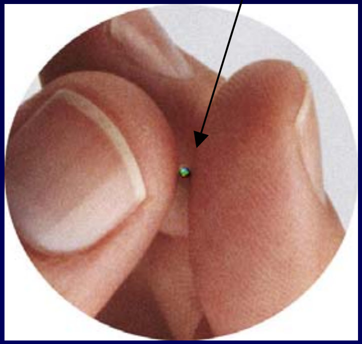


Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Sentrix™ Array Matrix and BeadChip Formats

Sentrix Array Matrix

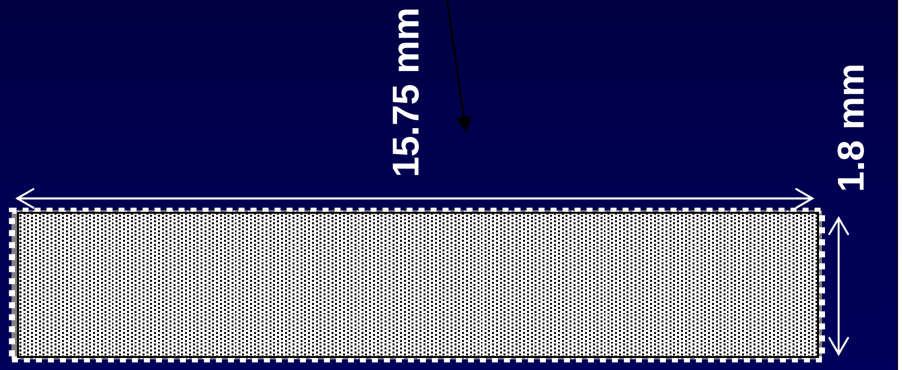




~50,000 Beads

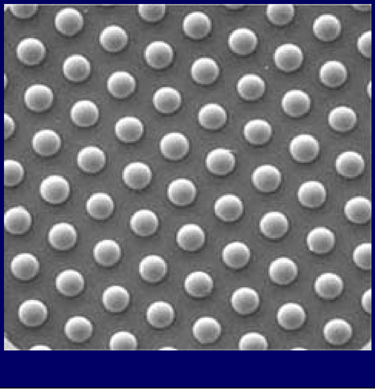
50,000/30 = 1666 types (genes)

Sentrix BeadChip



~900,000 Beads

900,000/30 = 30,000 types (genes)



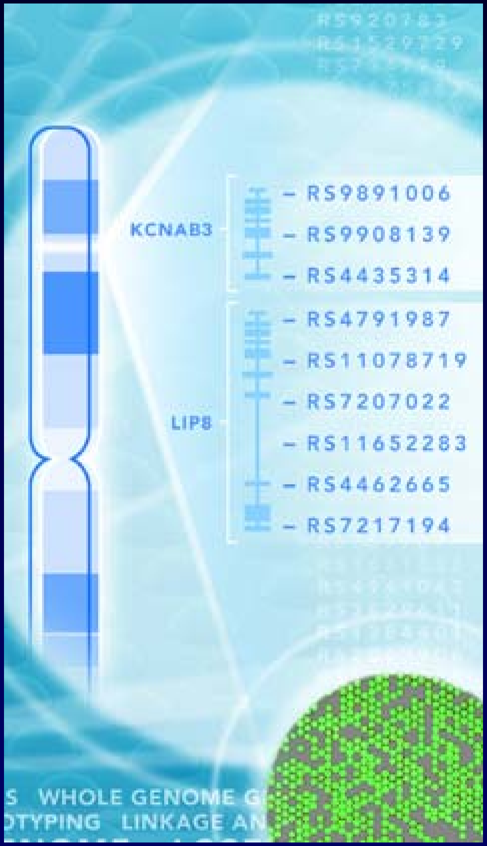
Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

NUS National University of Singapore

Whole Genome Association Studies

What are the key needs?

- Genotype 100,000's of loci accurately
  - High locus selectivity
  - High specificity for allelic discrimination
- Ability to assay SNPs of interest, access to vast majority of genome
- A robust means of processing many samples easily and efficiently
- A technician-friendly automatable process that reduces possibility of sample tracking error

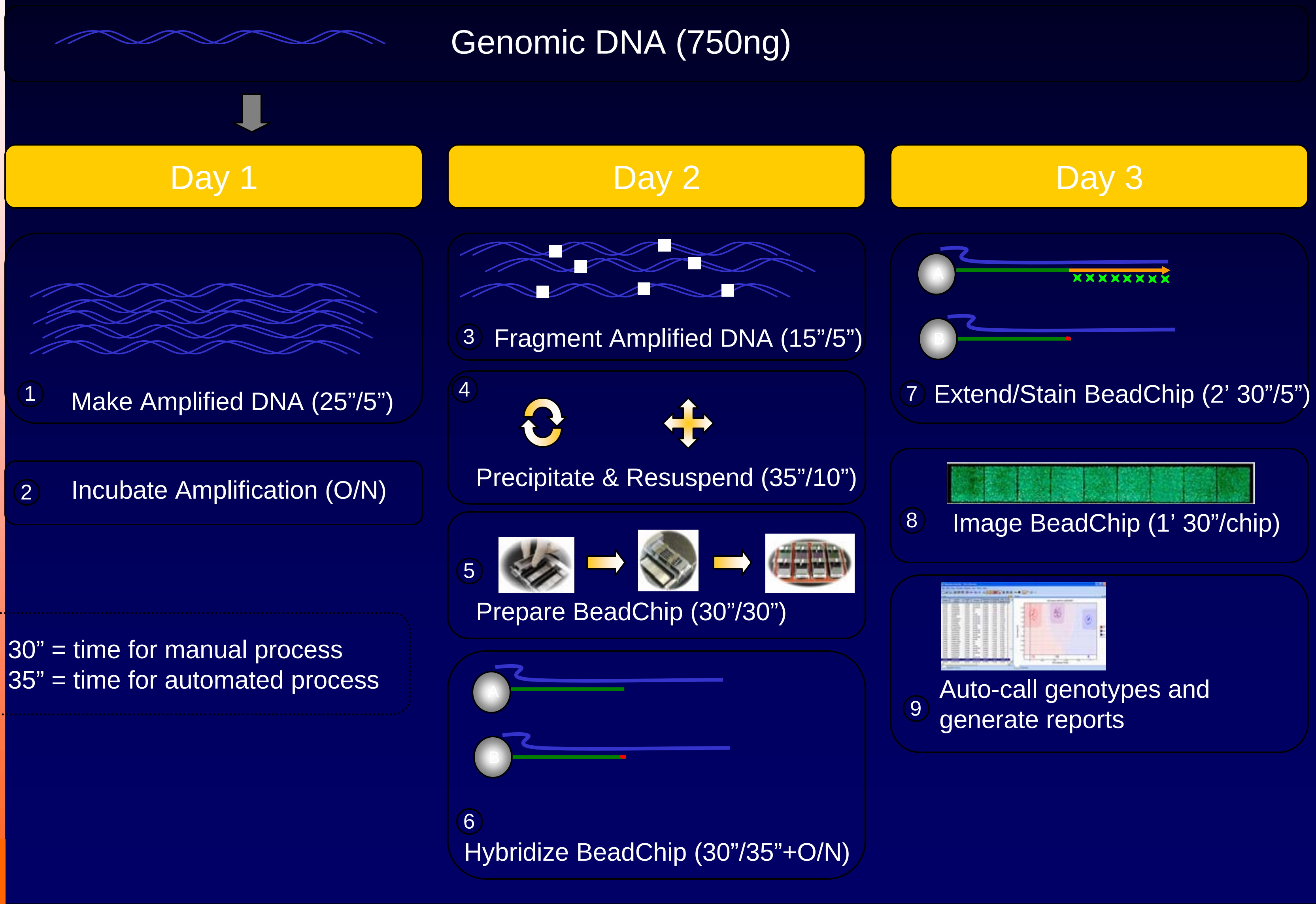


Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Infinium Whole Genome Genotyping

- Flexible BeadChip design
  - ➢ High density architecture
  - ➢ Easily configured for different content and sample numbers
- Flexible SNP selection
- BeadArray™ technology
  - ➢ 100% QC on 100% of arrays
  - ➢ Average 30-fold redundancy

Infinium I
Whole Genome Genotyping Workflow



Genomic DNA (750ng)

Day 1

1 Make Amplified DNA (25”/5”)

2 Incubate Amplification (O/N)

30” = time for manual process
35” = time for automated process

Day 2

3 Fragment Amplified DNA (15”/5”)

4 Precipitate & Resuspend (35”/10”)

5 Prepare BeadChip (30”/30”)

6 Hybridize BeadChip (30”/35”+O/N)

Day 3

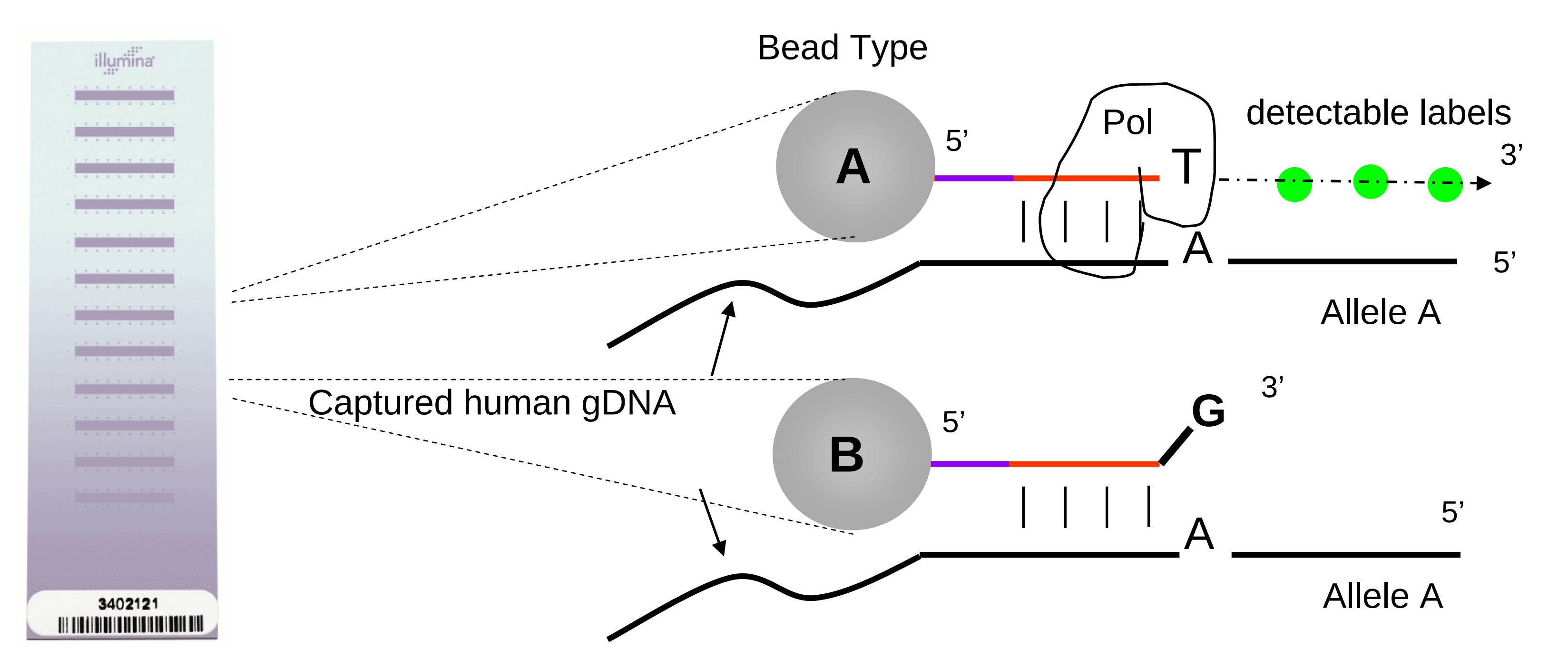
7 Extend/Stain BeadChip (2’ 30”/5”)

8 Image BeadChip (1’ 30”/chip)

9 Auto-call genotypes and generate reports

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

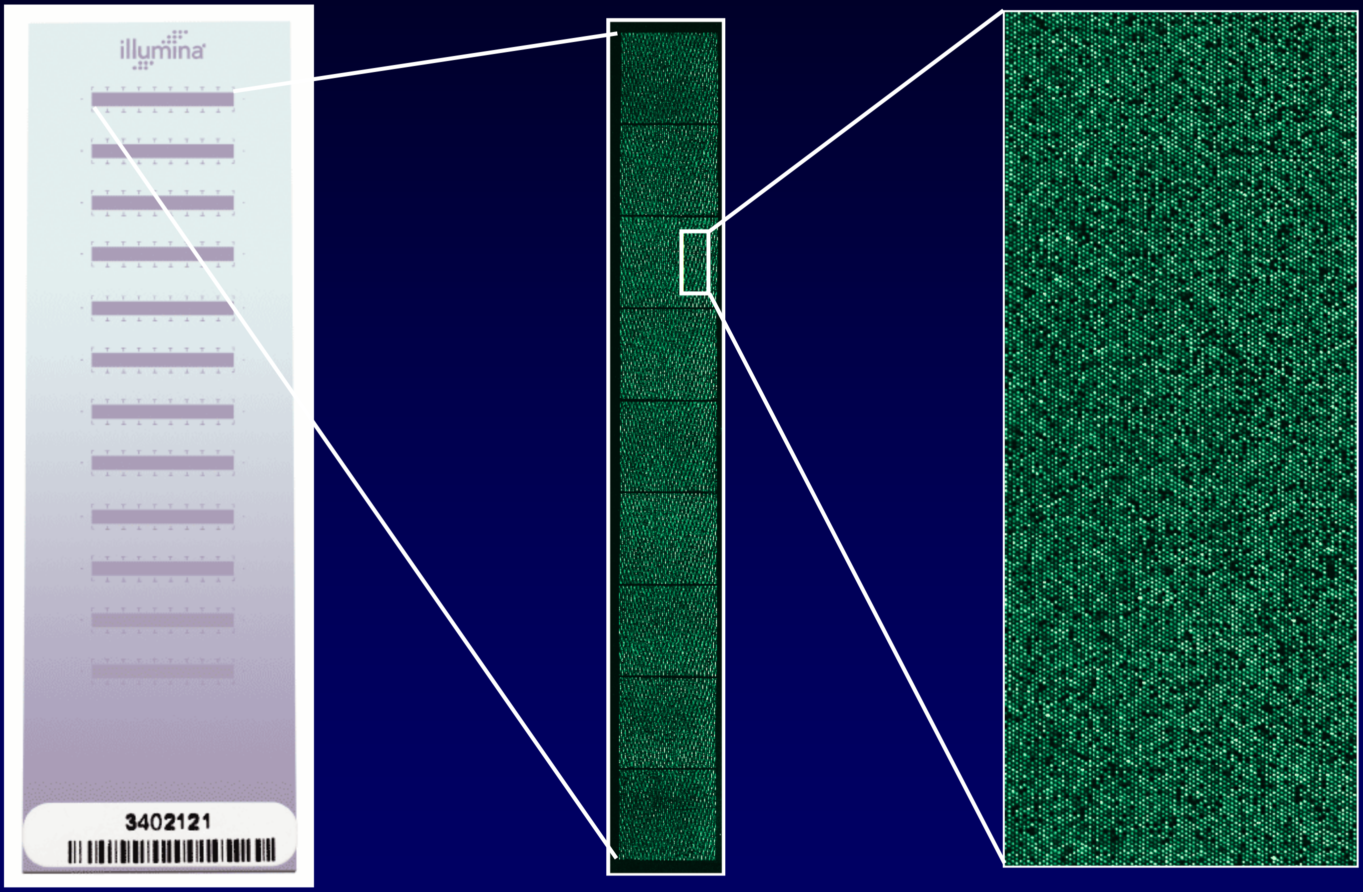
Infinium I: Allele-Specific Primer Extension



- Freedom to choose SNPs.

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Scan → Image Registration → Intensity Data



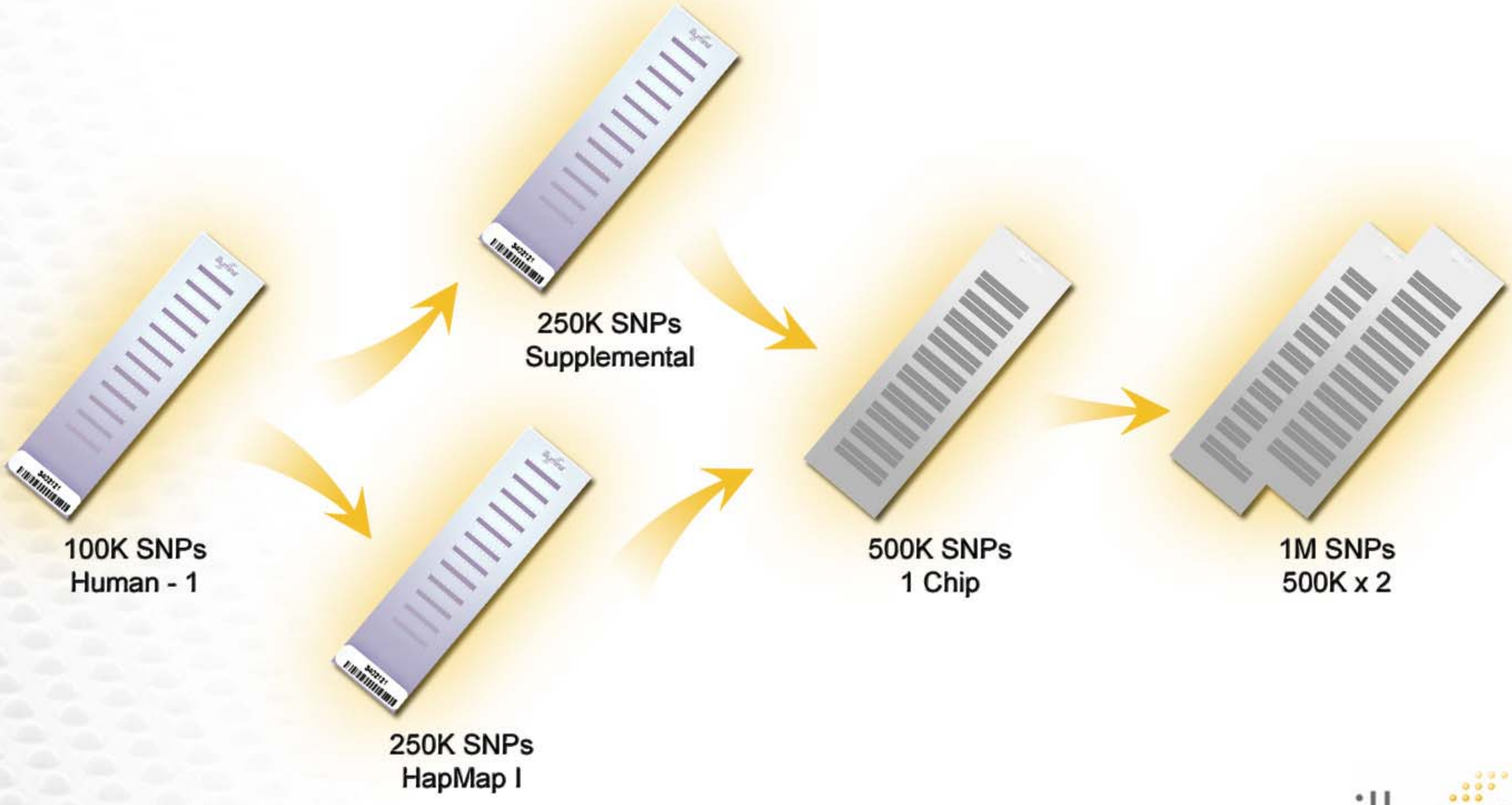
12 Sections

>890,000 Beads per Section

Average 30 fold redundancy

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Whole Genome Genotyping Product Evolution



100K SNPs
Human - 1

250K SNPs
Supplemental

250K SNPs
HapMap I

500K SNPs
1 Chip

1M SNPs
500K x 2

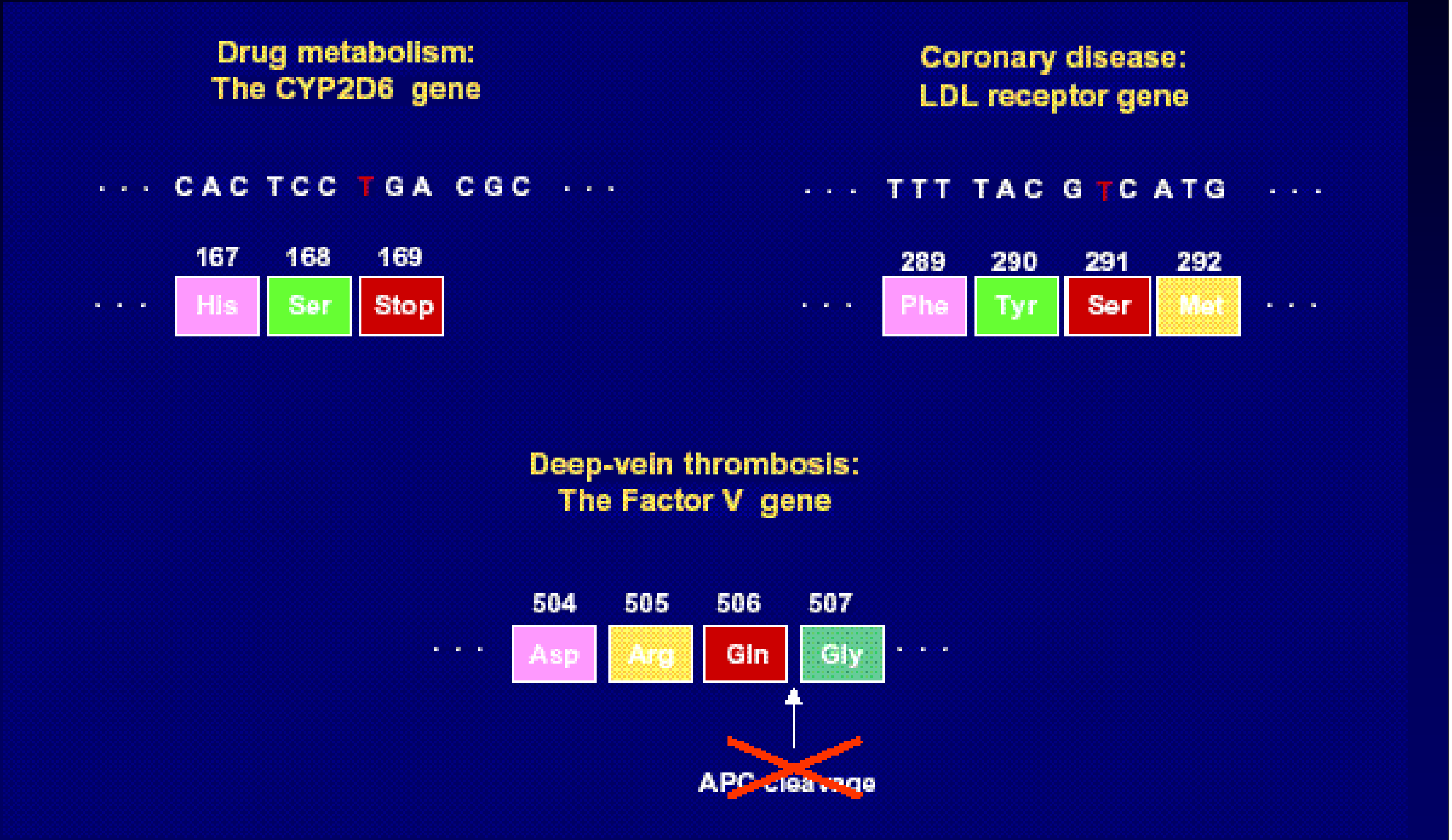
Human Genetic Variation

- What types of variants exist?
- How are variants found?
- How are variants scored?
- How are variants used?

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

NUS National University of Singapore

Functional variants



Drug metabolism:
The CYP2D6 gene

... CAC TCC TGA CGC ...

Coronary disease:
LDL receptor gene

... TTT TAC GTC ATG ...

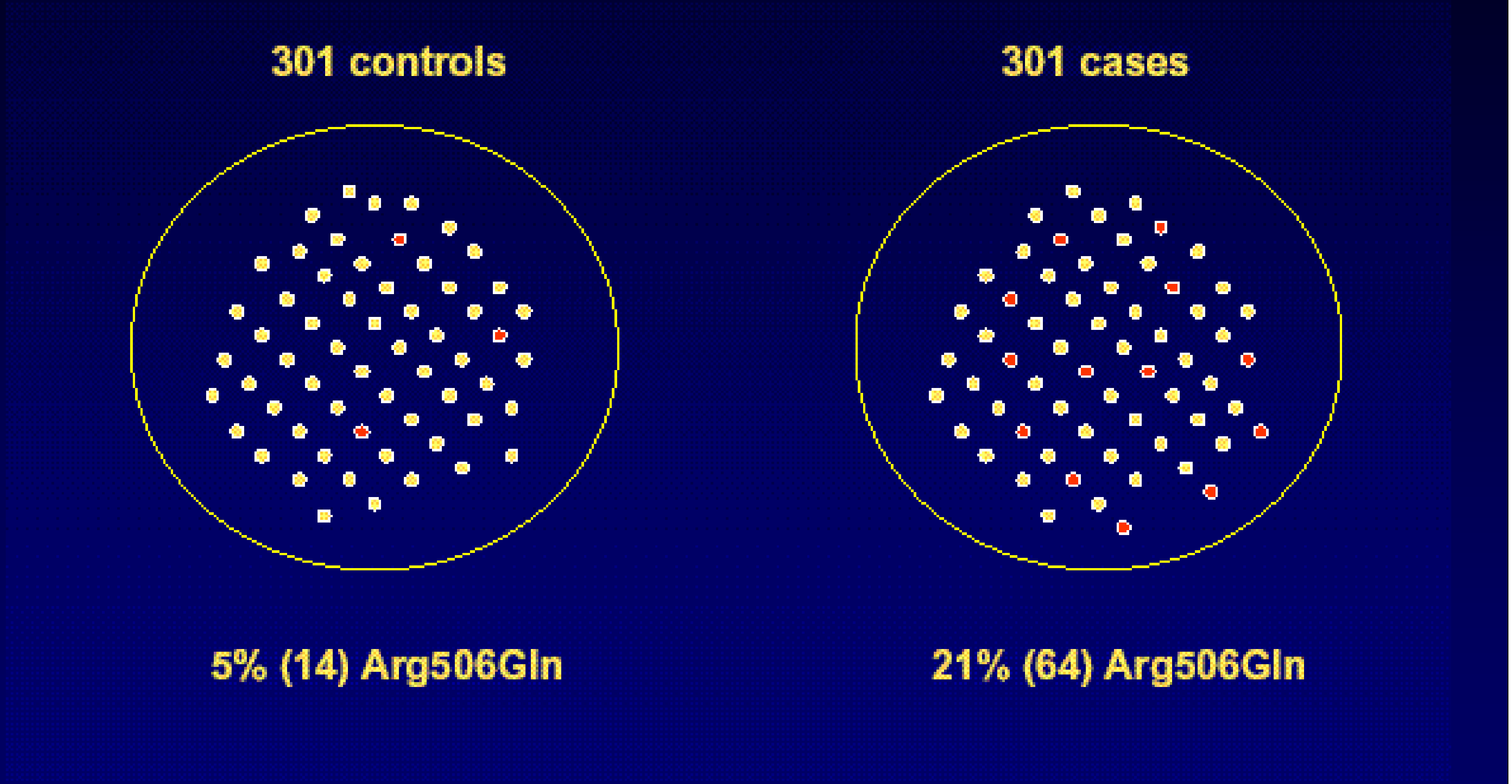
Deep-vein thrombosis:
The Factor V gene

~~APC cleavage~~

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

NUS National University of Singapore

Factor V $^{\text{Leiden}}$ association study



301 controls

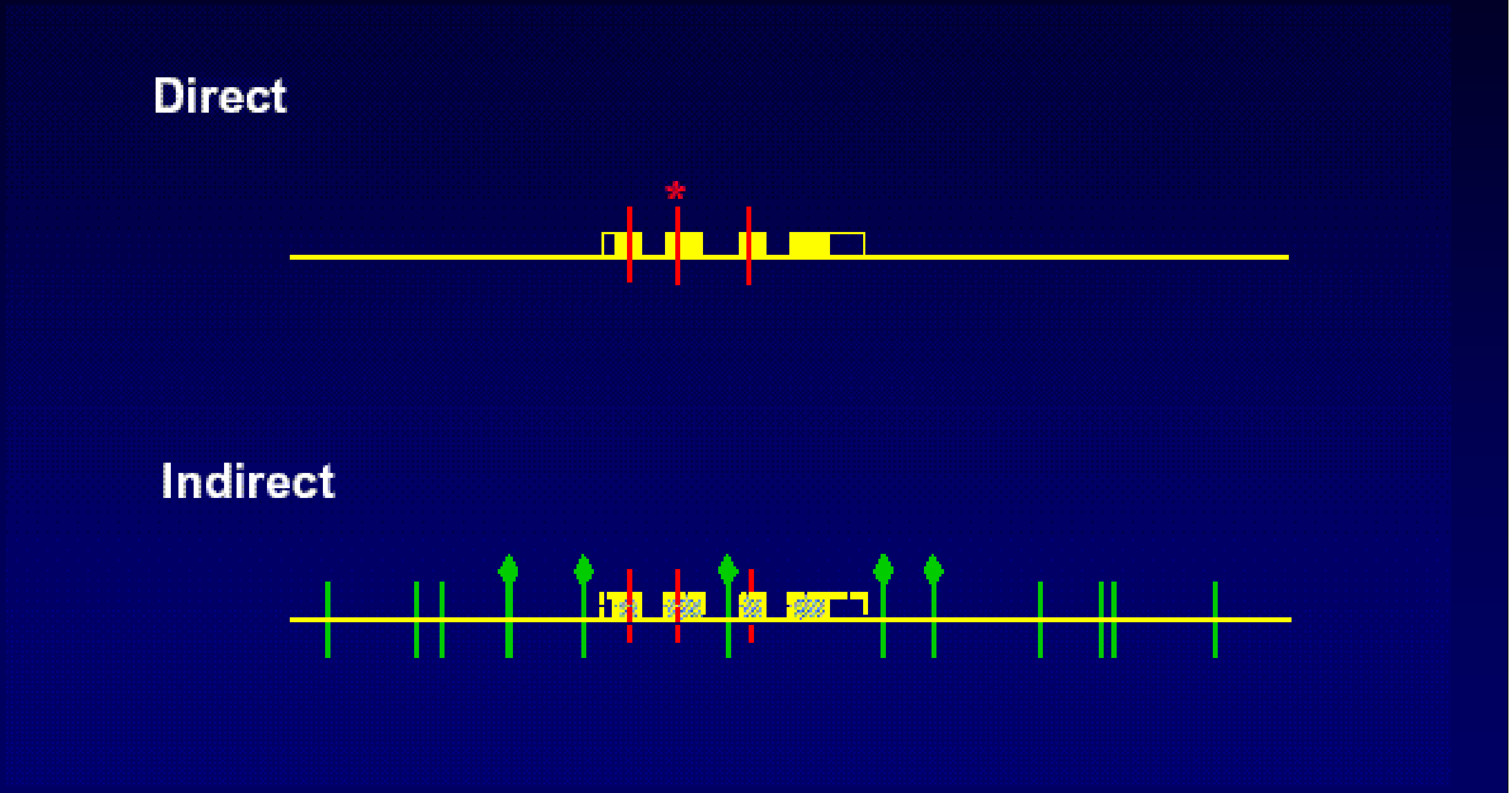
301 cases

5% (14) Arg506Gln

21% (64) Arg506Gln

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Association Studies



Example: case-control association study

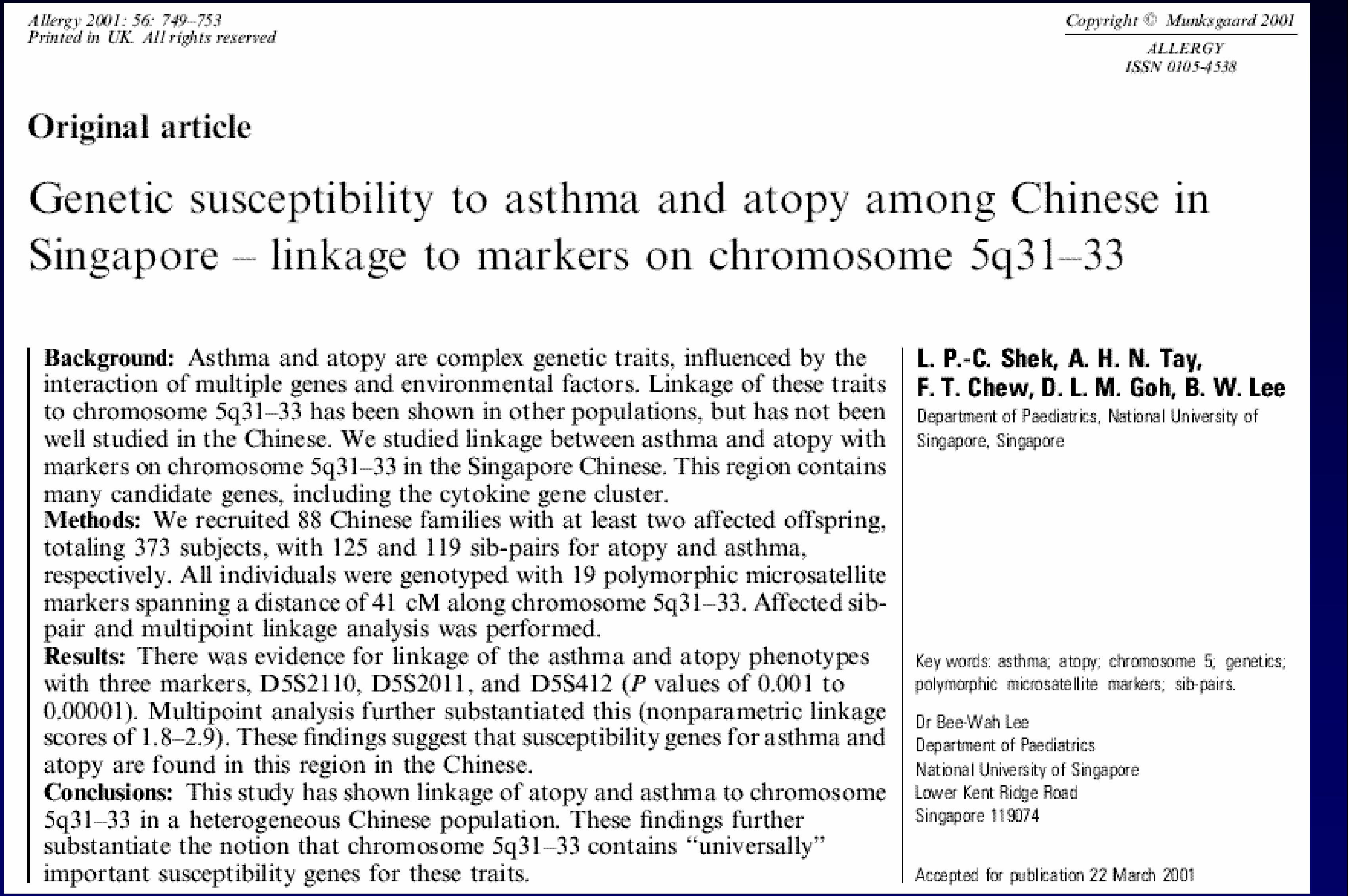
500 cases
500 controls

Prior evidence suggests
10 Mb candidate region

In 10Mb, expect ~10,000 SNPs, ~100 genes

Need:
Efficient way to screen SNPs
Knowledge of most useful SNPs

Asthma among Chinese Singaporeans linked to markers on chromosome 5q31-33



Allergy 2001: 56: 749–753
Printed in UK. All rights reserved

Copyright © Munksgaard 2001
ALLERGY
ISSN 0105-4538

Original article

Genetic susceptibility to asthma and atopy among Chinese in Singapore – linkage to markers on chromosome 5q31–33

Background: Asthma and atopy are complex genetic traits, influenced by the interaction of multiple genes and environmental factors. Linkage of these traits to chromosome 5q31–33 has been shown in other populations, but has not been well studied in the Chinese. We studied linkage between asthma and atopy with markers on chromosome 5q31–33 in the Singapore Chinese. This region contains many candidate genes, including the cytokine gene cluster.
Methods: We recruited 88 Chinese families with at least two affected offspring, totaling 373 subjects, with 125 and 119 sib-pairs for atopy and asthma, respectively. All individuals were genotyped with 19 polymorphic microsatellite markers spanning a distance of 41 cM along chromosome 5q31–33. Affected sib-pair and multipoint linkage analysis was performed.
Results: There was evidence for linkage of the asthma and atopy phenotypes with three markers, D5S2110, D5S2011, and D5S412 (*P* values of 0.001 to 0.00001). Multipoint analysis further substantiated this (nonparametric linkage scores of 1.8–2.9). These findings suggest that susceptibility genes for asthma and atopy are found in this region in the Chinese.
Conclusions: This study has shown linkage of atopy and asthma to chromosome 5q31–33 in a heterogeneous Chinese population. These findings further substantiate the notion that chromosome 5q31–33 contains "universally" important susceptibility genes for these traits.

L. P.-C. Shek, A. H. N. Tay, F. T. Chew, D. L. M. Goh, B. W. Lee
Department of Paediatrics, National University of Singapore, Singapore

Key words: asthma; atopy; chromosome 5; genetics; polymorphic microsatellite markers; sib-pairs.

Dr Bee-Wah Lee
Department of Paediatrics
National University of Singapore
Lower Kent Ridge Road
Singapore 119074

Accepted for publication 22 March 2001

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Screen SNPs using pooled DNA

500 cases	one pool
500 controls	one pool
10,000 SNPs	

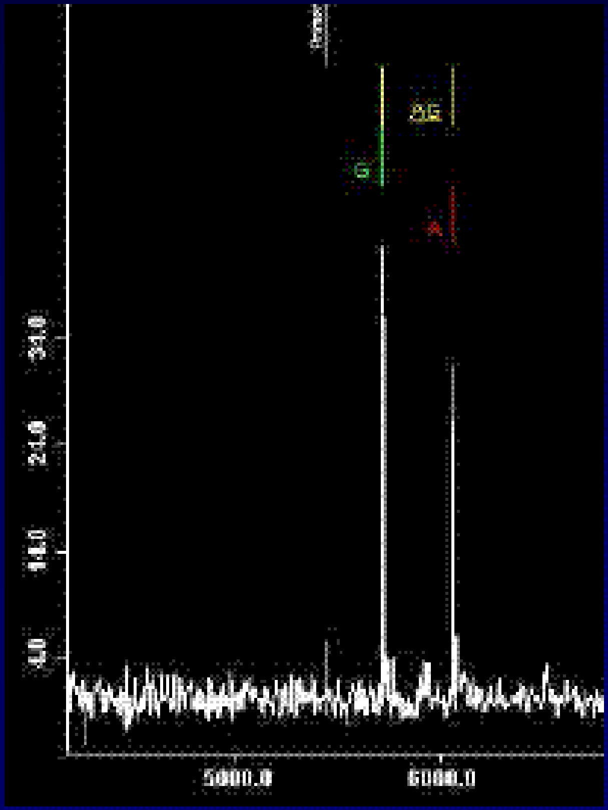
Direct analysis: 10,000,000 genotypes

Pooled DNA analysis: 20,000 genotypes

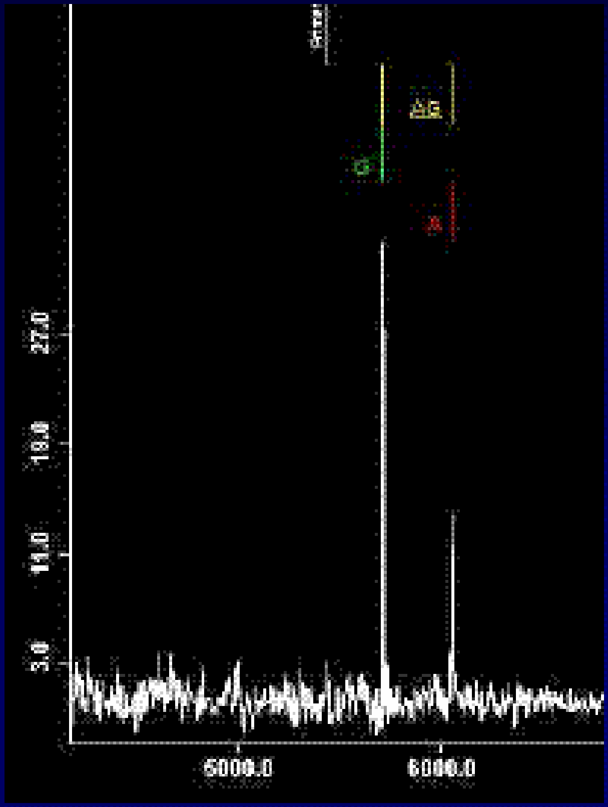
Genotyping of DNA pools

- Create equimolar pools of individual DNAs
- Type SNP and determine relative allele frequencies

Affected



Unaffected controls



Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Example: case-control association study

500 cases
500 controls

Prior evidence suggests
10 Mb candidate region

In 10Mb, expect ~10,000 SNPs, ~100 genes

Need:
Efficient way to screen SNPs
Knowledge of most useful SNPs

Variation at adjacent sites tends to correlate

```
GAAATAATTAATGTTTTCCTTCCTTCTCCTATTTTGTCCTTTACTTCAATTTATTTTTATTATTAATATTATTATTTTTTGAGACGGAGTTTCACTCTTGT
TGCCAACCTGGAGTGCAGTGGCGTGATCTCAGCTCACTGCACACTCCGCTTTC [C/T] GGTTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGAC
TACAGTCACACACCACCACGCCCGGCTAATTTTTGTATTTTTAGTAGAGTTGGGGTTTCACCATGTTGGCCAGACTGGTCTCGAACTCCTGACCTTGTGATCC
GCCAGCCTCTGCCTCCCAAAGAGCTGGGATTACAGGCGTGAGCCACCGCGCTCGGCCCTTTGCATCAATTTCTACAGCTTGTTTTCTTTGCCTGGACTTTACA
AGTCTTACCTTGTTCTGCCTTTGGATATTTGTGTGGTCTCATTCTGGTGTGCCAGTAGCTAAAAATCCATGATTTGCTCTCATCCCACTCCTGTTGTTCATCT
CCTCTTATCTGGGGTCAC [A/C] TATCTCTTCGTGATTGCATTCTGATCCCCAGTACTTAGCATGTGCGTAACAACTCTGCCTCTGCTTTCCCAGGCTGTTGA
TGGGGTGCTGTTCATGCCTCAGAAAAAATGCATTGTAAGTTAAATTATTAAAGATTTTAAATATAGGAAAAAAAGTAAGCAAACATAAGGAACAAAAAAGGAAAGA
ACATGTATTCTAATCCATTATTTATTATACAATTAAGAAATTTGGAAACTTTAGATTACACTGCTTTTAGAGATGGAGATGTAGTAAGTCTTTTACTCTTTAC
AAAATACATGTGTTAGCAATTTTGGGAAGAATAGTAACTCACCCGAACAGTGTAATGTGAATATGTCACTTACTAGAGGAAAGAAGGCACTTGAAAAAACATCT
CTAAACCGTATAAAAACAATTACATCATAATGATGAAAACCCAAGGAATTTTTTTAGAAAACATTACCAGGGCTAATAACAAAGTAGAGCCACATGTCATTTA
TCTTCCCTTTGTGTCTGTGTGAGAATTCTAGAGTTATATTTGTACATAGCATGGAAAAAATGAGAGGCTAGTTTATCAACTAGTTCATTTTTAAAAGTCTAACA
CATCCTAGGTATAGGTGAACTGTCCTCCTGCCAATGTATTGCACATTTGTGCCCAGATCCAGCATAGGGTATGTTTGCCATTTACAAACGTTTATGTCTTAAG
AGAGGAAATATGAAGAGCAAAACAGTGCATGCTGGAGAGAGAAAGCTGATACAAATATAAATGAAACAATAATTGGAAAAAATTGAGAAACTACTCATTTTCTA
AATTACTCATGTATTTTCCTAGAATTTAAGTCTTTTAATTTTTGATAAATCCCAATGTGAGACAAGATAAGTATTAGTGATGGTATGAGTAATTAATATCTGT
TATATAATATTCATTTTCATAGTGGAAGAAATAAAATAAAGGTTGTGATGATTGTTGATTATTTTTTCTAGAGGGGTTGTCAGGGAAAGAAATTGCTTTTTTT
CATTCTCTCTTTCCACTAAGAAAGTTCAACTATTAATTTAGGCACATACAATAATTACTCCATTCTAAAATGCCAAAAAGGTAATTTAAGAGACTTAAAACTG
AAAAGTTTAAGATAGTCACACTGAACTATATTAAAAAAATCCACAGGGTGGTTGGAACTAGGCCTTATATTAAAGAGGCTAAAAATTTAATAAGACCACAGGC
TTTAAATATGGCTTTAAACTGTGAAAGGTGAAACTAGAATGAATAAAATCCTATAAATTTAAATCAAAAGAAAGAAACAAACT [A/G] AAATTAAAGTTATTA
TACAAGAATATGGTGGCCTGGATCTAGTGAACATATAGTAAAGATAAAACAGAATATTTCTGAAAAATCCTGGAAAATCTTTTGGGCTAACCTGAAAACAGTA
TATTTGAAACTATTTTTAAAATGCAGTGATACTAGAAATATTTTAGAATCATATGTA
```

[C/T] [A/C] [A/G]

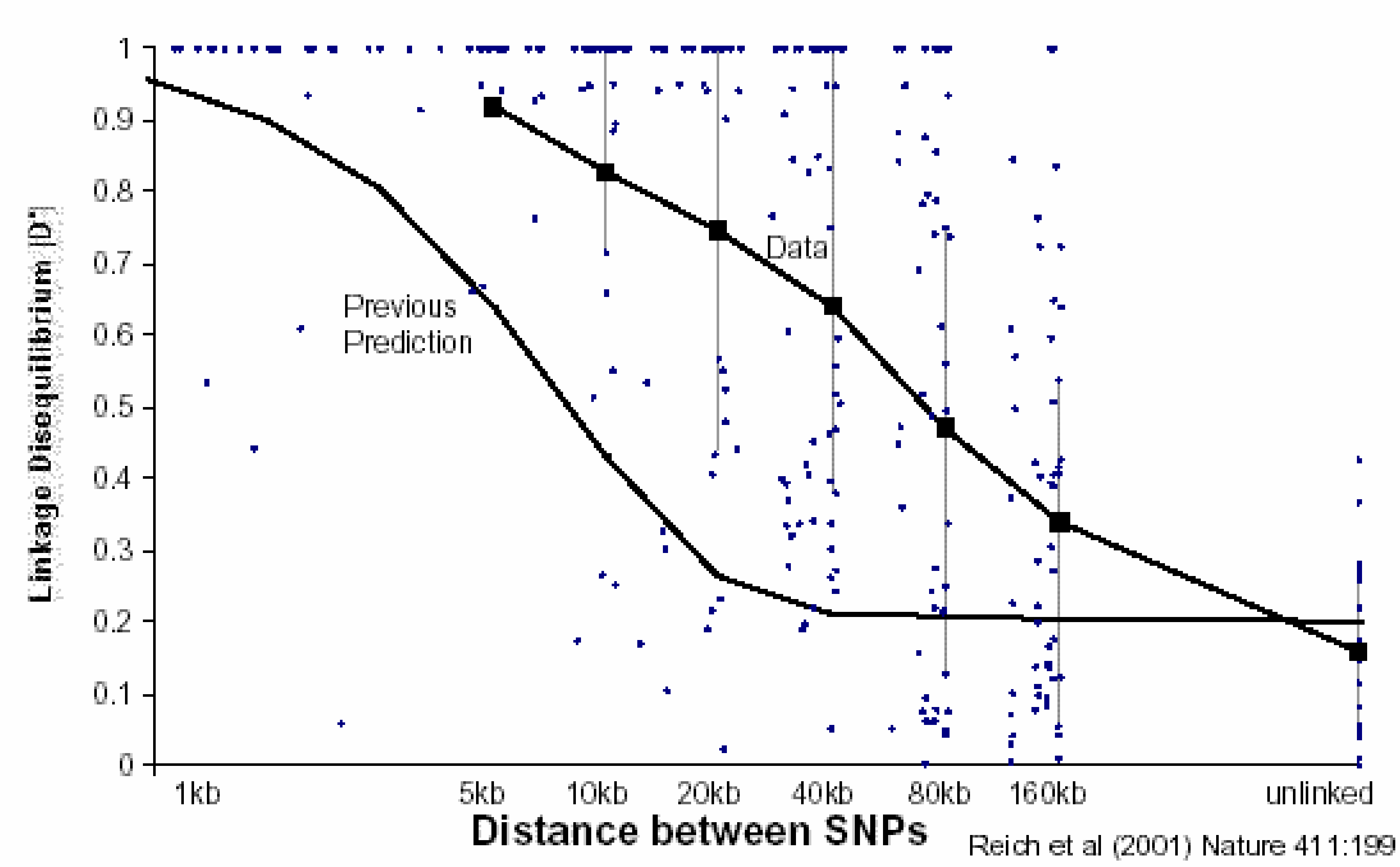
Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Linkage disequilibrium

How large are the conserved segments?

3 kb? 30 kb? 300 kb??

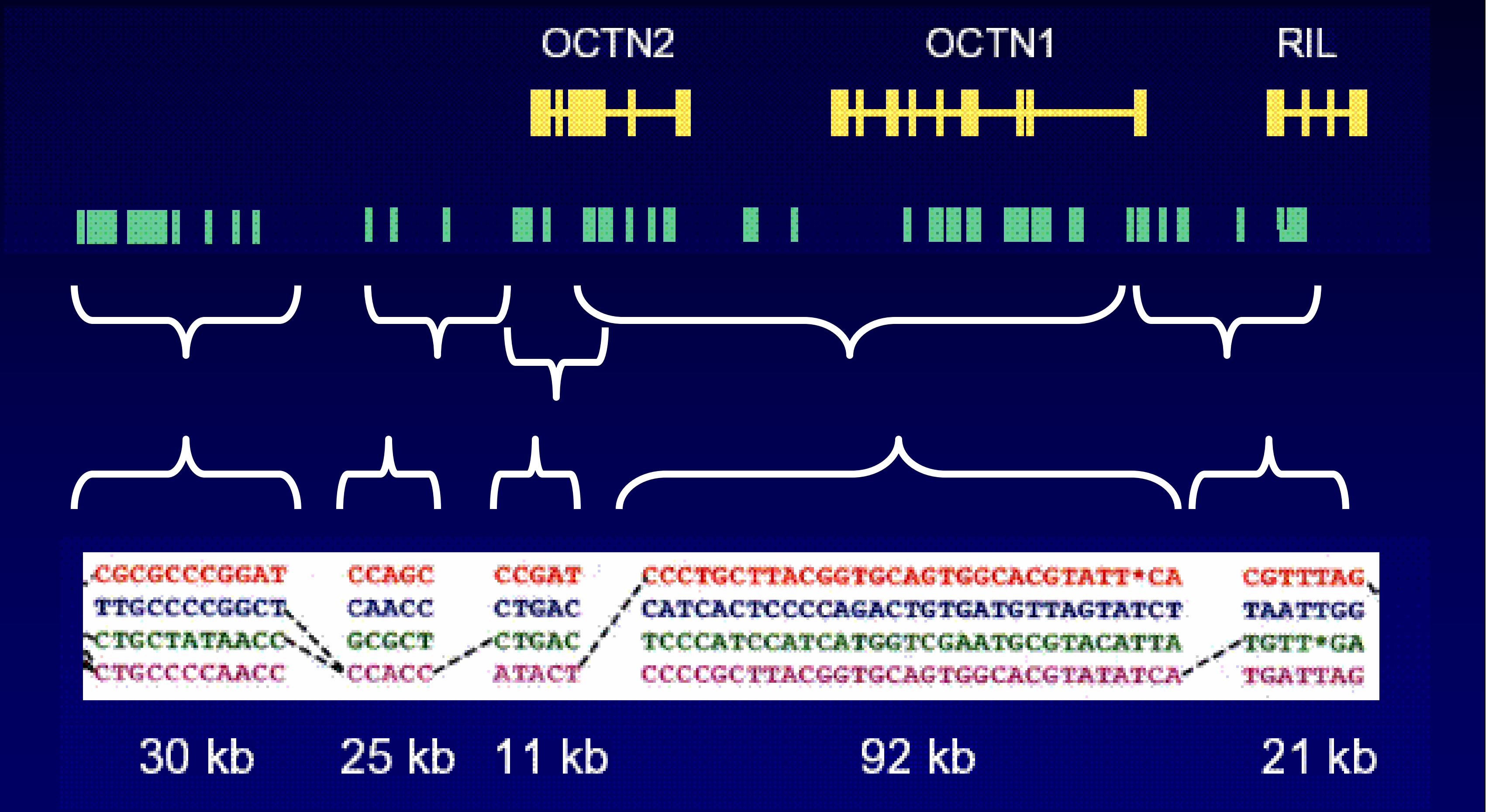
In North Europeans, linkage disequilibrium extends 60 kb in each direction



Reich et al (2001) Nature 411:199

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

Haplotypes from 258 chromosomes on 5q31



30 kb	25 kb	11 kb	92 kb	21 kb
CGCGCCCGGAT	CCAGC	CCGAT	CCCTGCTTACGGTGCAGTGGCACGTATT*CA	CGTTTAG
TTGCCCCGGCT	CAACC	CTGAC	CATCACTCCCCAGACTGTGATGTTAGTATCT	TAATTGG
CTGCTATAACC	GCGCT	CTGAC	TCCCATCCATCATGGTCGAATGCGTACATTA	TGTT*GA
CTGCCCCAACC	CCACC	ATACT	CCCCGCTTACGGTGCAGTGGCACGTATATCA	TGATTAG

Daly *et al* (2001) *Nature Genetics* 29:229

Linkage disequilibrium

How large are the conserved segments?

Average block size perhaps ~20 kb

Genotype only the most useful SNPs

500 cases	one pool
500 controls	one pool
~~10,000~~ SNPs	
1,000 ‘haplotype tag’ SNPs	

Direct analysis:	10,000,000 genotypes
Pooled DNA analysis:	20,000 genotypes
Selected SNPs:	2,000 genotypes

Chew Fook Tim
Functional Genomics Laboratories
Department of Biological Sciences
National University of Singapore

NUS
National University of Singapore

Future

- Continued identification of SNPs
- Faster, cheaper, easier genotyping
- Genome haplotype map
- SNP panel(s) for association studies
- Discovery of new functional variants