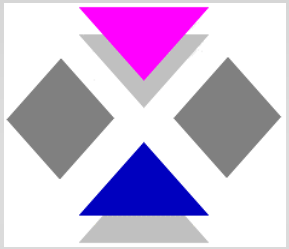
Molecular Basis of Disease



WESTERN
GENERAL
HOSPITALS
NHS TRUST

EDINBURGH

Wayne Lam

Consultant Clinical Geneticist

Molecular Medicine Centre

Western General Hospital

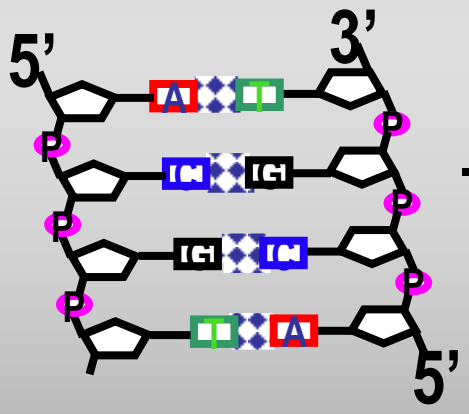
Edinburgh EH4 2XU

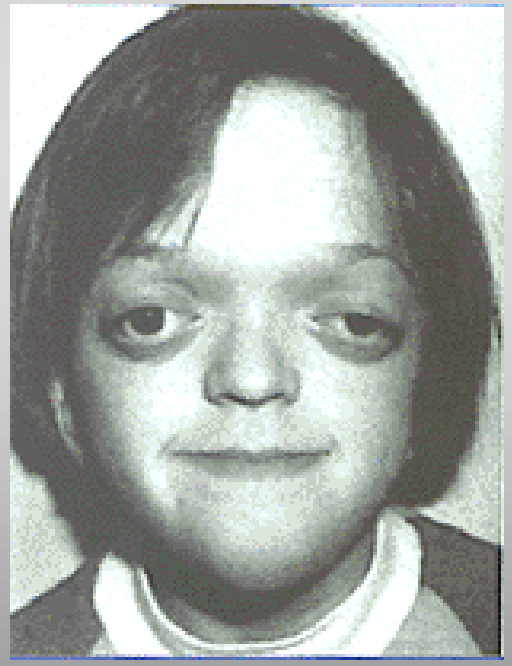
wayne.lam@ed.ac.uk

Genetics:
Some basic molecular definitions

- Central dogma of genetics

DNA ↔ RNA ↔ Protein ↔ Phenotype





What is a pathogenic mutation?

Most important question in molecular genetics

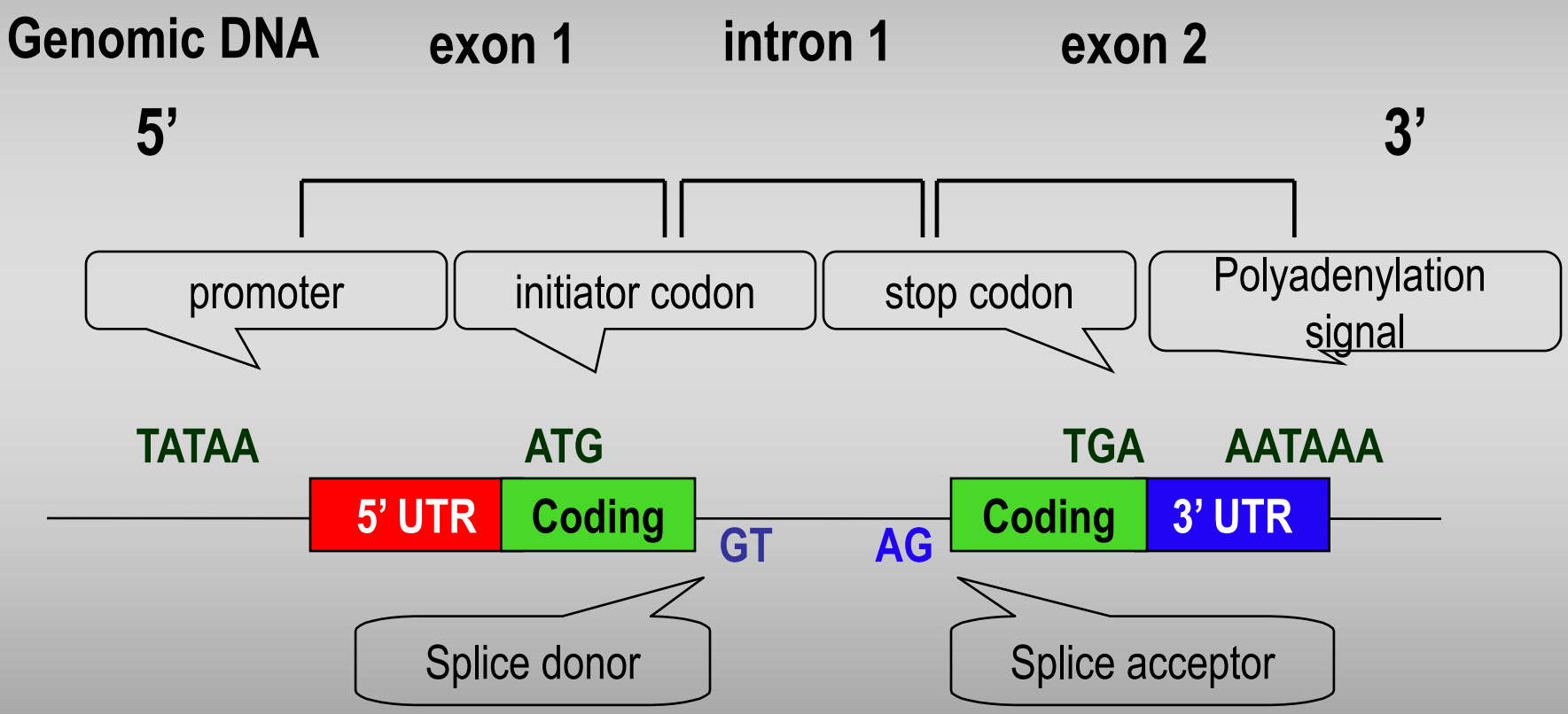
Is this gene change pathogenic?

i.e. is this the cause of the disorder

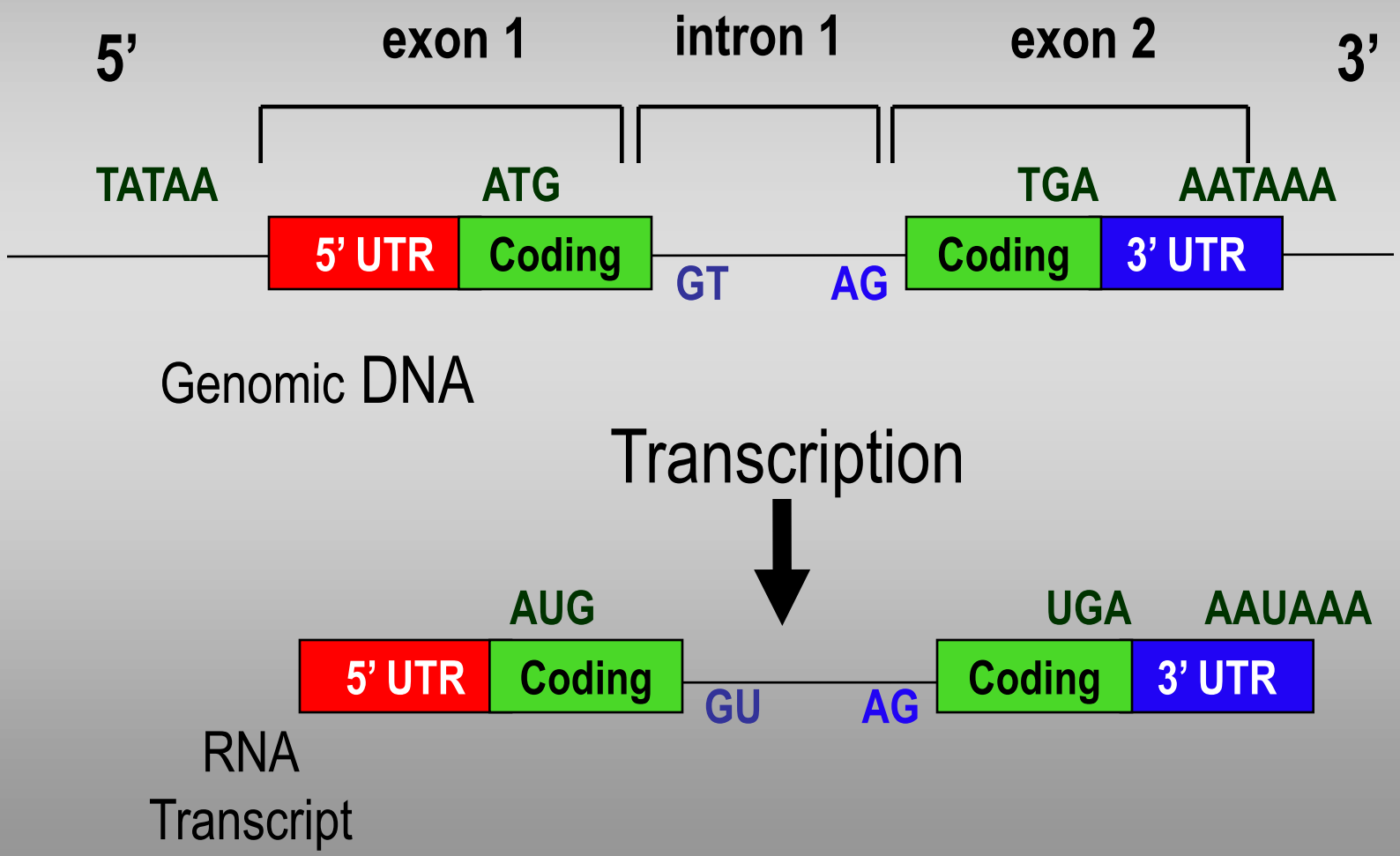
Mutation Criteria

- Does it affect the function of the protein ?
- It is in a conserved region of the protein ?
  - Does it co-segregate with the disorder in the family ?
- Is the change seen in the normal population ?

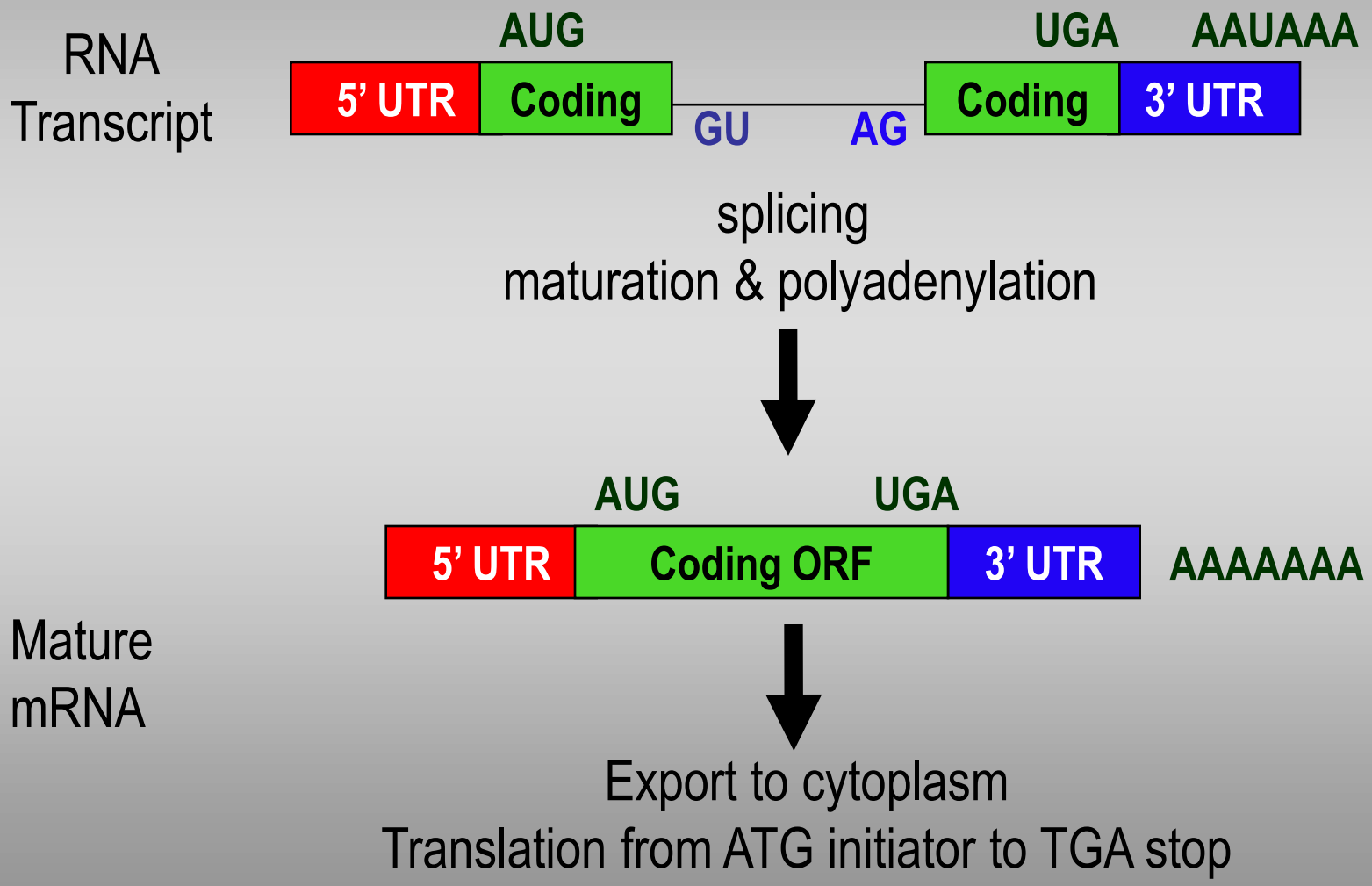
Anatomy of a Gene

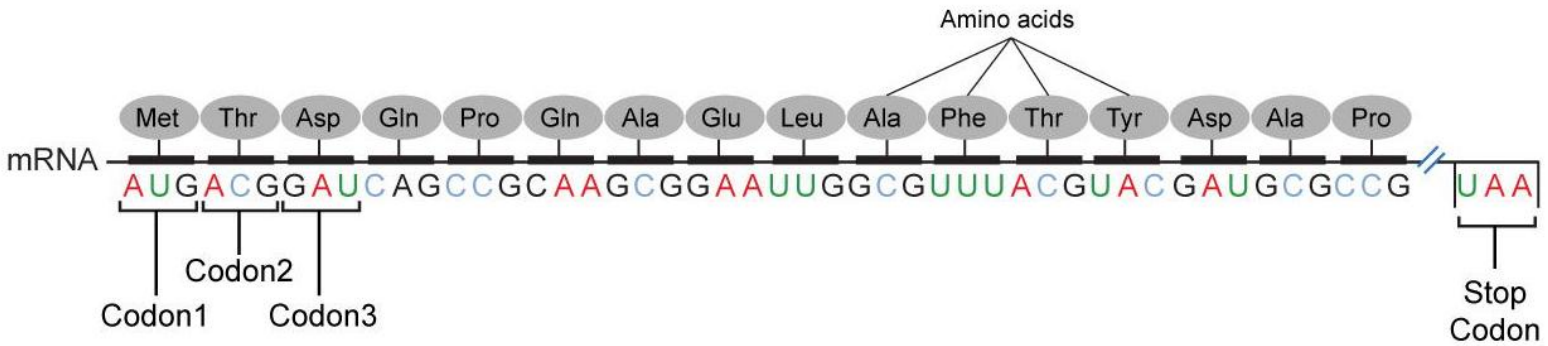


Gene Structure Transcription



Post transcriptional modifications





Types of mutation in DNA sequences

Deletions

- Ranges from 1bp to megabases

Insertions

- Ranges vary can be as small as 1bp up to megabases
- Duplication and inversions

Single base pair substitutions

- **Missense mutations**
  - Replaces one amino acid with another
- **Nonsense mutations**
  - Replaces an amino acid codon with a stop codon
- **Splice site mutations**
  - Creates or destroys signals/coding for exon-intron splicing

Types of mutation in DNA sequences

- **Frameshifts**
  - Caused by deletions, insertions or splicing errors
- **Dynamic mutations**
  - Tandem repeats

Types of mutation in DNA sequences

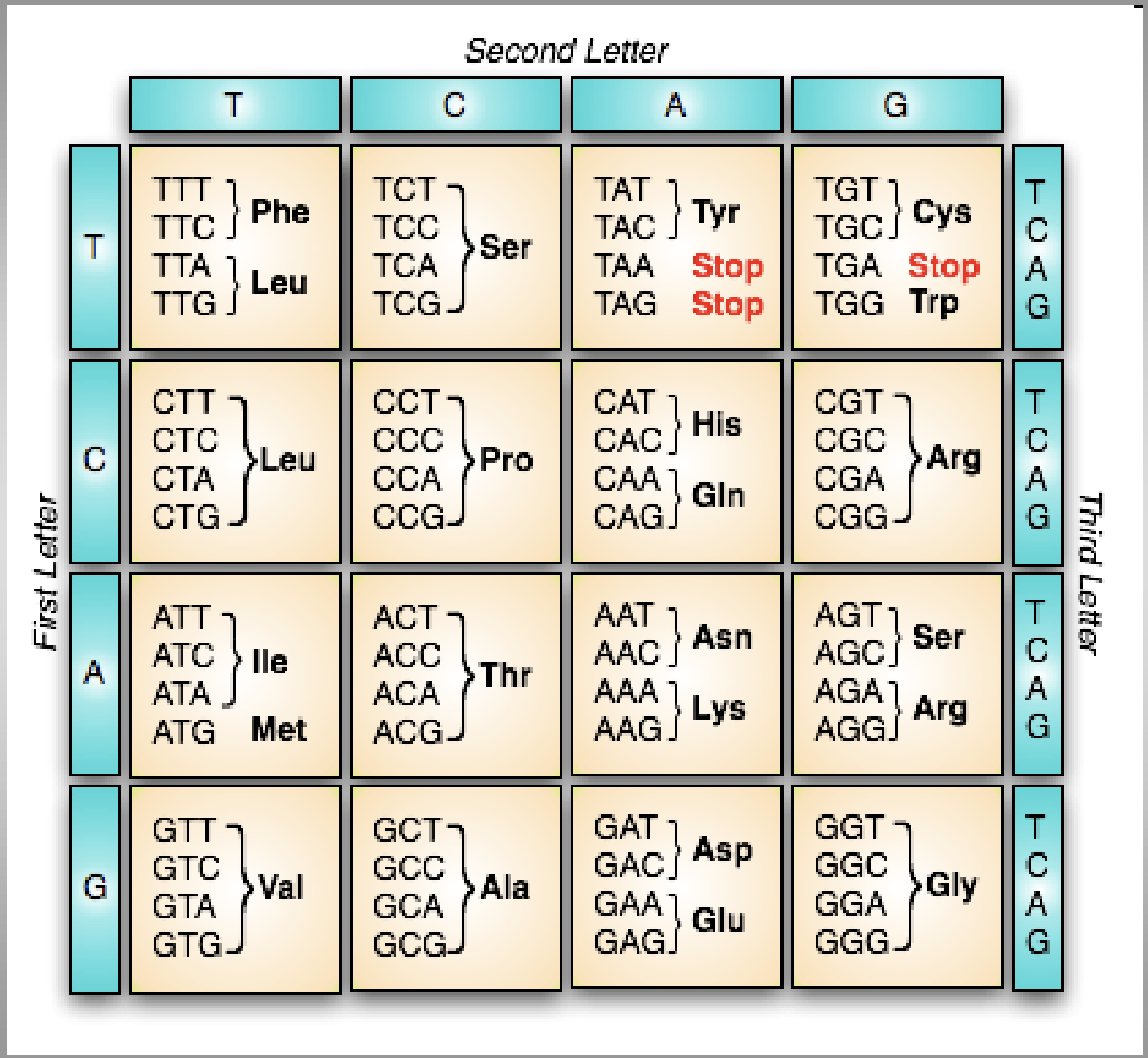
The cat sat on the mat	**Wild type**
The cat spa to nth ema t	**Insertion**
The cas ato nt hem at	**Deletion**
	Frameshifts
The cat	**Stop / Nonsense**
The car sat on the mat	**Missense**

Types of mutation in DNA sequences

The cat sat on the mat	**Wild type**
The cat spa to nth ema t	**Insertion**
The cas ato nt hem at	**Deletion**
	Frameshifts
The cat	**Stop / Nonsense**
The car sat on the mat	**Missense**
The **cat cat** sat on the mat	**Triplet expansion** (Dynamic mutation)
The **tas tac** on the mat	**Inversion**

Missense mutations
(within exon)

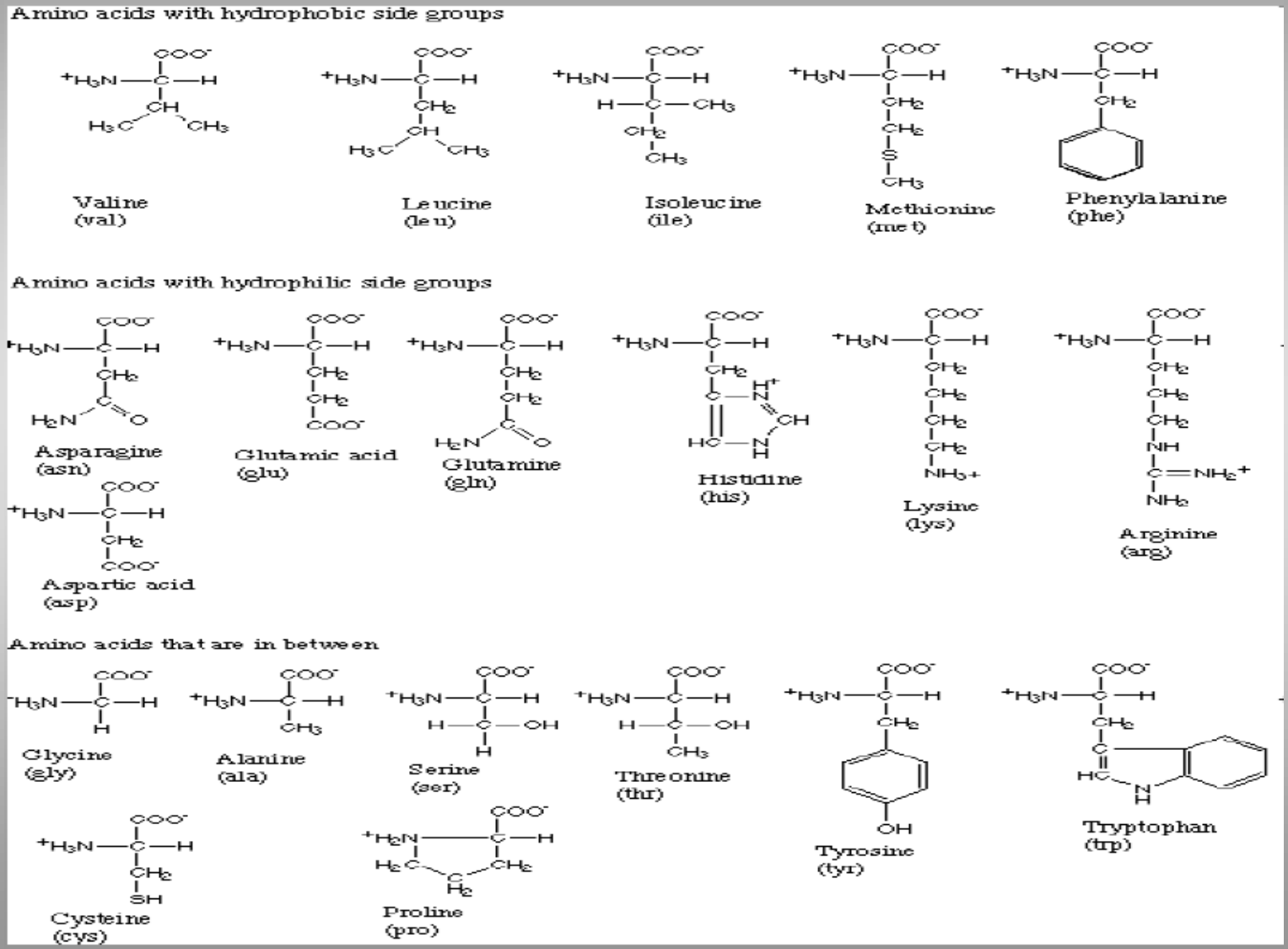
- Has it caused a change in amino acid?
  - Some redundancy in the genetic code
  - 20 amino acids and 64 possible codons



Missense mutations
(within exon)

- Where there has been a change in amino acid
- Has it caused a conserved or non-conservative change in amino acid
  - Change in polarity
  - Change in hydrophobicity

Missense mutations



Grantham Matrix

From: Grantham R. Amino acid difference formula to help explain protein evolution. Science 185:862–4 (1974)

Arg	Leu	Pro	Thr	Ala	Val	Gly	Ile	Phe	Tyr	Cys	His	Gln	Asn	Lys	Asp	Glu	Met	Trp	
110	145	74	58	99	124	56	142	155	144	112	89	68	46	121	65	80	135	177	**Ser**
	102	103	71	112	96	125	97	97	77	180	29	43	86	26	96	54	91	101	**Arg**
		98	92	96	32	138	5	22	36	198	99	113	153	107	172	138	15	61	**Leu**
			38	27	68	42	95	114	110	169	77	76	91	103	108	93	87	147	**Pro**
				58	69	59	89	103	92	149	47	42	65	78	85	65	81	128	**Thr**
					64	60	94	113	112	195	86	91	111	106	126	107	84	148	**Ala**
						109	29	50	55	192	84	96	133	97	152	121	21	88	**Val**
							135	153	147	159	98	87	80	127	94	98	127	184	**Gly**
								21	33	198	94	109	149	102	168	134	10	61	**Ile**
									22	205	100	116	158	102	177	140	28	40	**Phe**
										194	83	99	143	85	160	122	36	37	**Tyr**
											174	154	139	202	154	170	196	215	**Cys**
												24	68	32	81	40	87	115	**His**
													46	53	61	29	101	130	**Gln**
														94	23	42	142	174	**Asn**
															101	56	95	110	**Lys**
																45	160	181	**Asp**
																	126	152	**Glu**
																		67	**Met**

- Method in calculating the significance of the amino acid substitution
- The bigger the score the more likely that the missense mutation has caused a change in the resultant protein structure

Mutation Analysis

It is in a conserved region of the protein ?

- More likely to affect function if changes in an region conserved across species (orthologs) or between members of a gene family (paralogs)
- Indicative of critical function

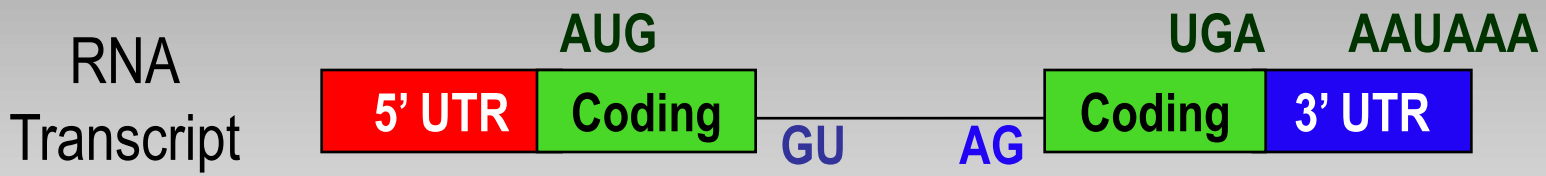
Does it co-segregate with the disorder in the family ?

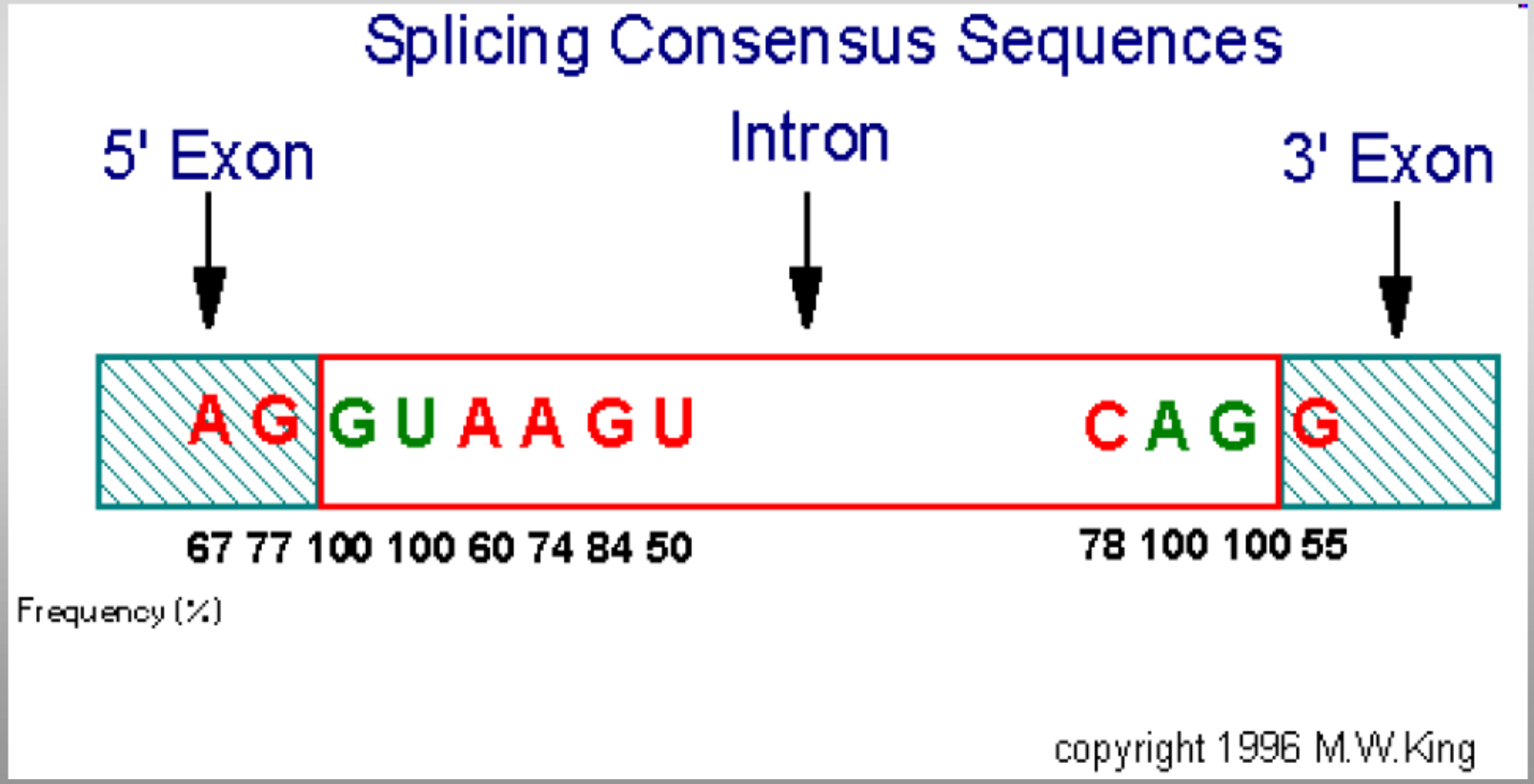
- Is the gene change only found in affected members

Is the change seen in the normal population ?

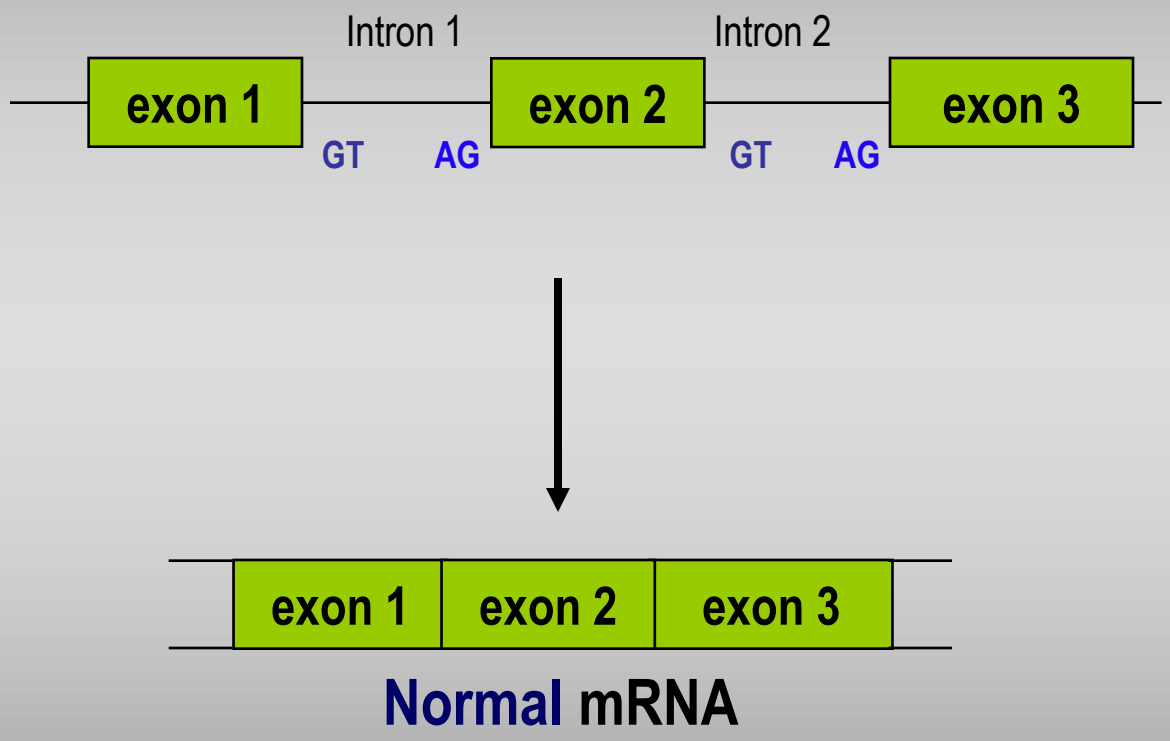
- Has a sample of the normal population been screen

Splice site

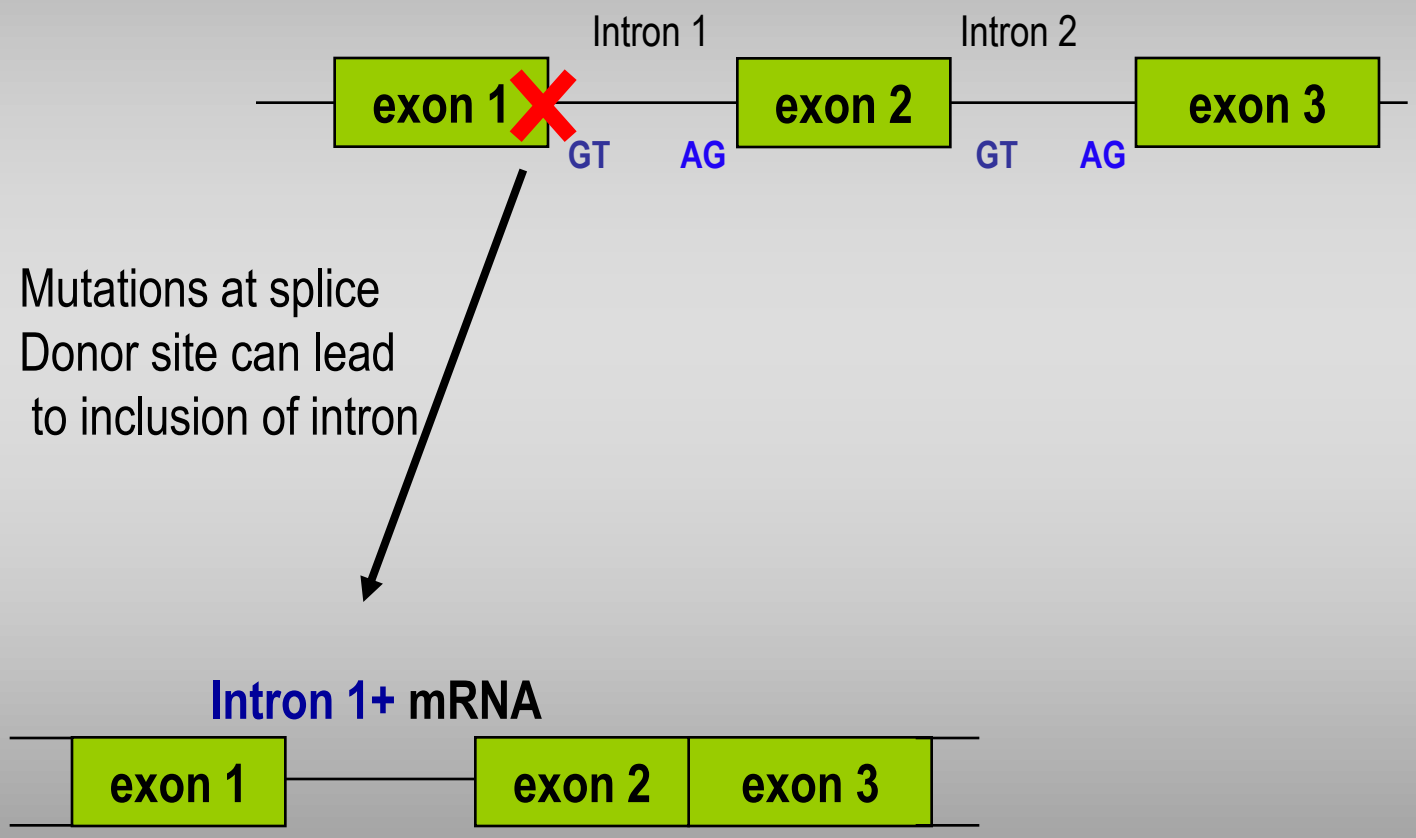

RNA
Transcript
AUG
UGA
AAUAAA
5' UTR
Coding
GU
AG
Coding
3' UTR


Splicing Consensus Sequences
5' Exon
Intron
3' Exon
A G G U A A G U
C A G G
67 77 100 100 60 74 84 50
78 100 100 55
Frequency (%)
copyright 1996 M.W.King

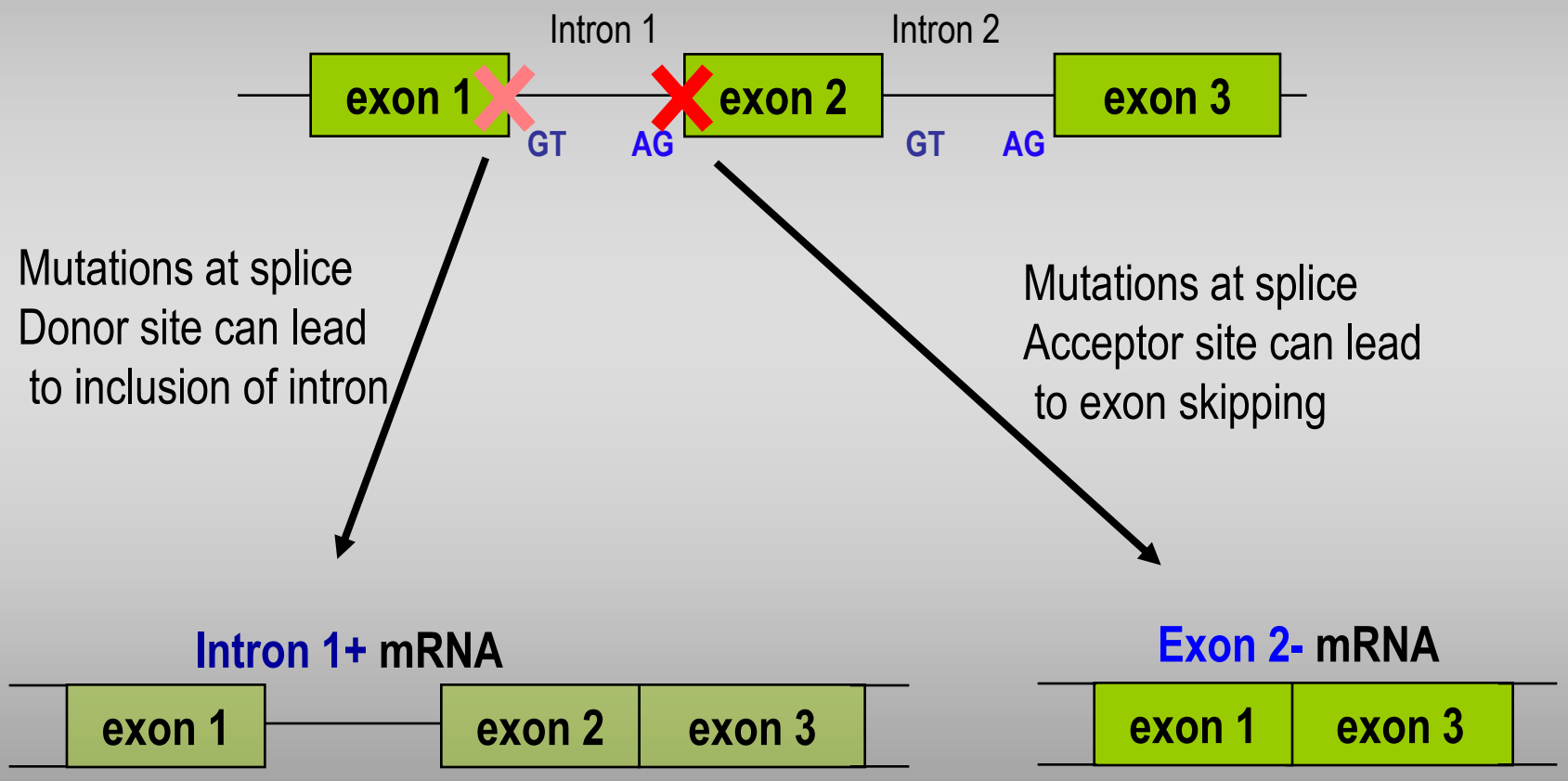
Introns are spliced out when mRNA is made

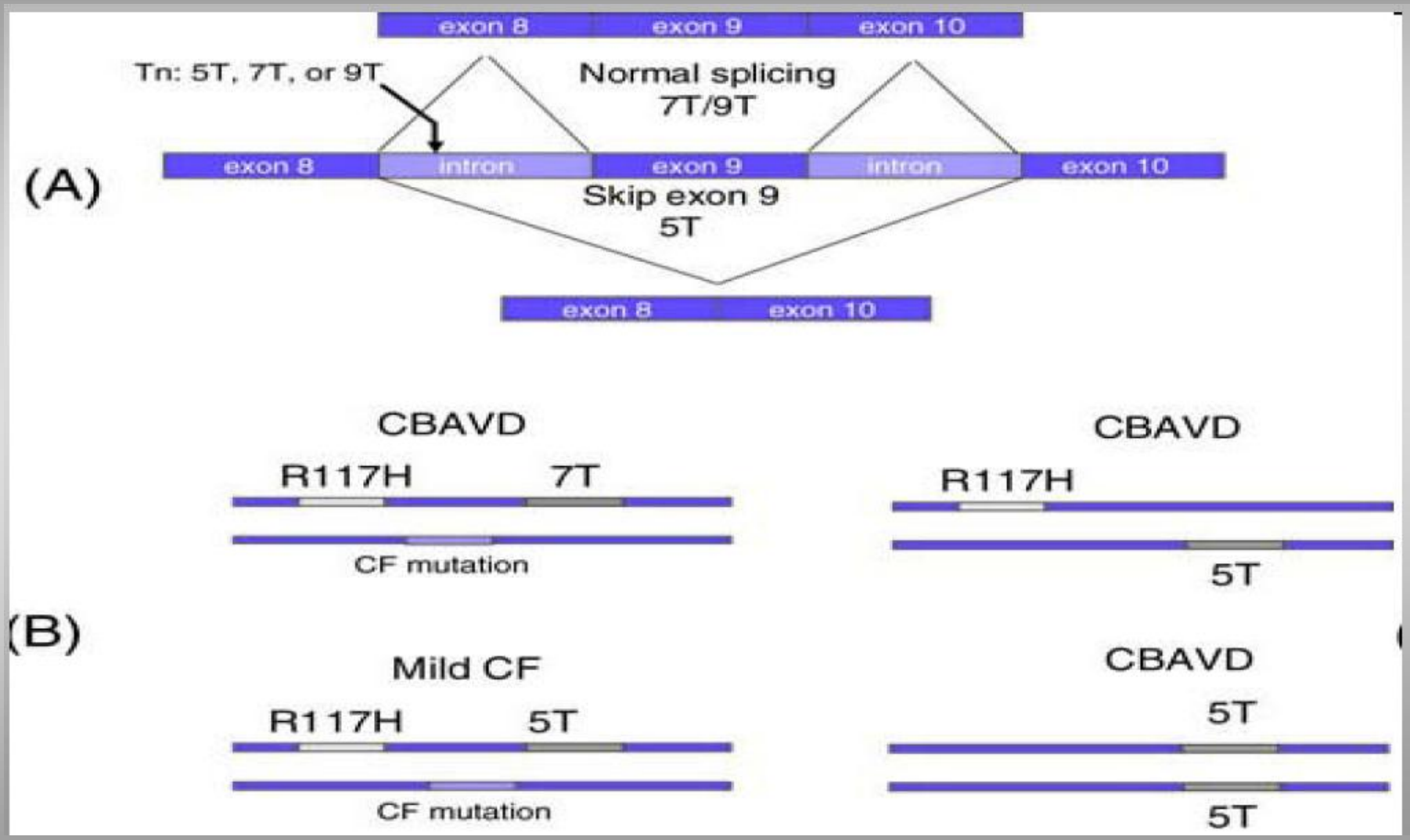


Spice site mutations



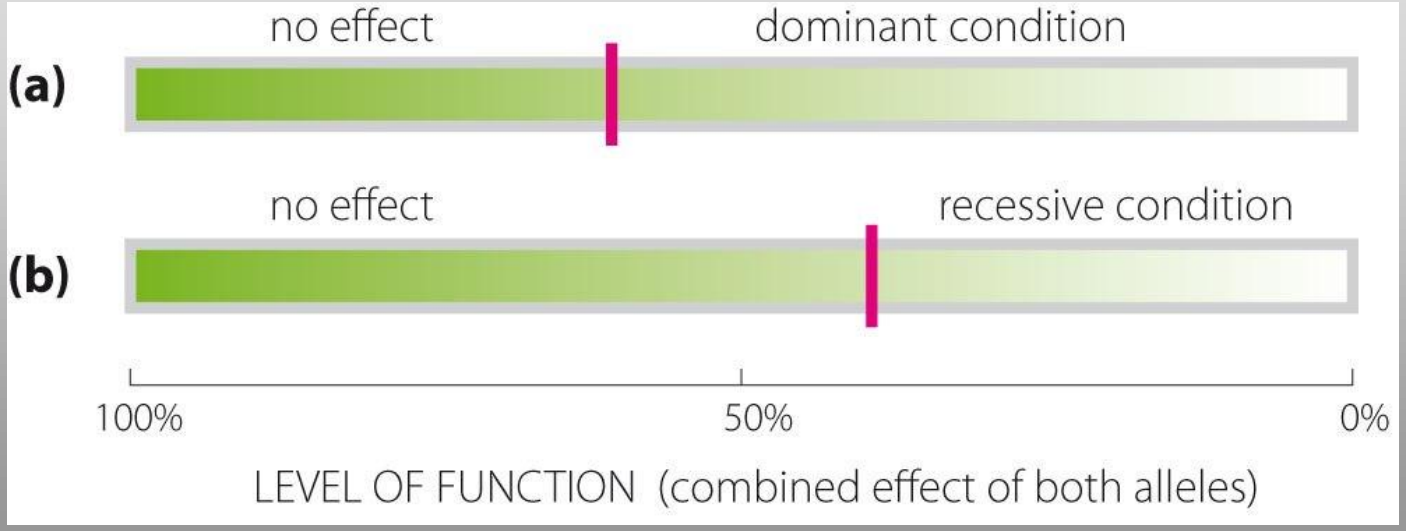
Spice site mutations




exon 8
exon 9
exon 10
Tn: 5T, 7T, or 9T
Normal splicing
7T/9T
(A)
exon 8
intron
exon 9
intron
exon 10
Skip exon 9
5T
exon 8
exon 10
CBAVD
R117H
7T
CF mutation
CBAVD
R117H
5T
(B)
Mild CF
R117H
5T
CF mutation
CBAVD
5T
5T

A Mutation Can Cause Disease by:

1. Loss of function (Abolition)
    - Due to non-functioning or truncated protein
        - Marfan syndrome, Duchennes muscular dystrophy
    - Haploinsufficiency
        - William syndrome
    - Dominant negative
        - Deafness syndromes, Collagen disorders



Dominant negative

- Special class of loss of function
  - The mutation produces a none functioning protein
  - The none functioning protein interferes with the protein of the normal functioning homologous gene
  - Resulting in no effective gene product

A Mutation Can Cause Disease by:

2. Modification
   - Creating a poorly functioning protein
     - Beckers muscular dystrophy
   - Abnormal activation of protein (overexpression)
     - Cancer genes
   - Gain of function of protein (novel function)
     - Huntington disease, cancer genes (philadelphilia chromosome-fusion protein)

Types of DNA testing

Direct testing:

The DNA from a consultand is tested to see whether or not it contains a given pathogenic mutation.

Indirect testing (gene tracking):

Linked markers are used in family studies to discover if the consultand inherited the disease carrying chromosome/allele from a parent.

Polymerase Chain Reaction (PCR) DNA Amplification:

- Very efficient at amplification of template DNA to yield products for analysis.
- DNA can be extracted from various sources blood specimens, mouthwash or tissue specimens.
- Only requires small amounts of patient genomic DNA.
- Best at amplifying small specific segments of DNA

PCR movie

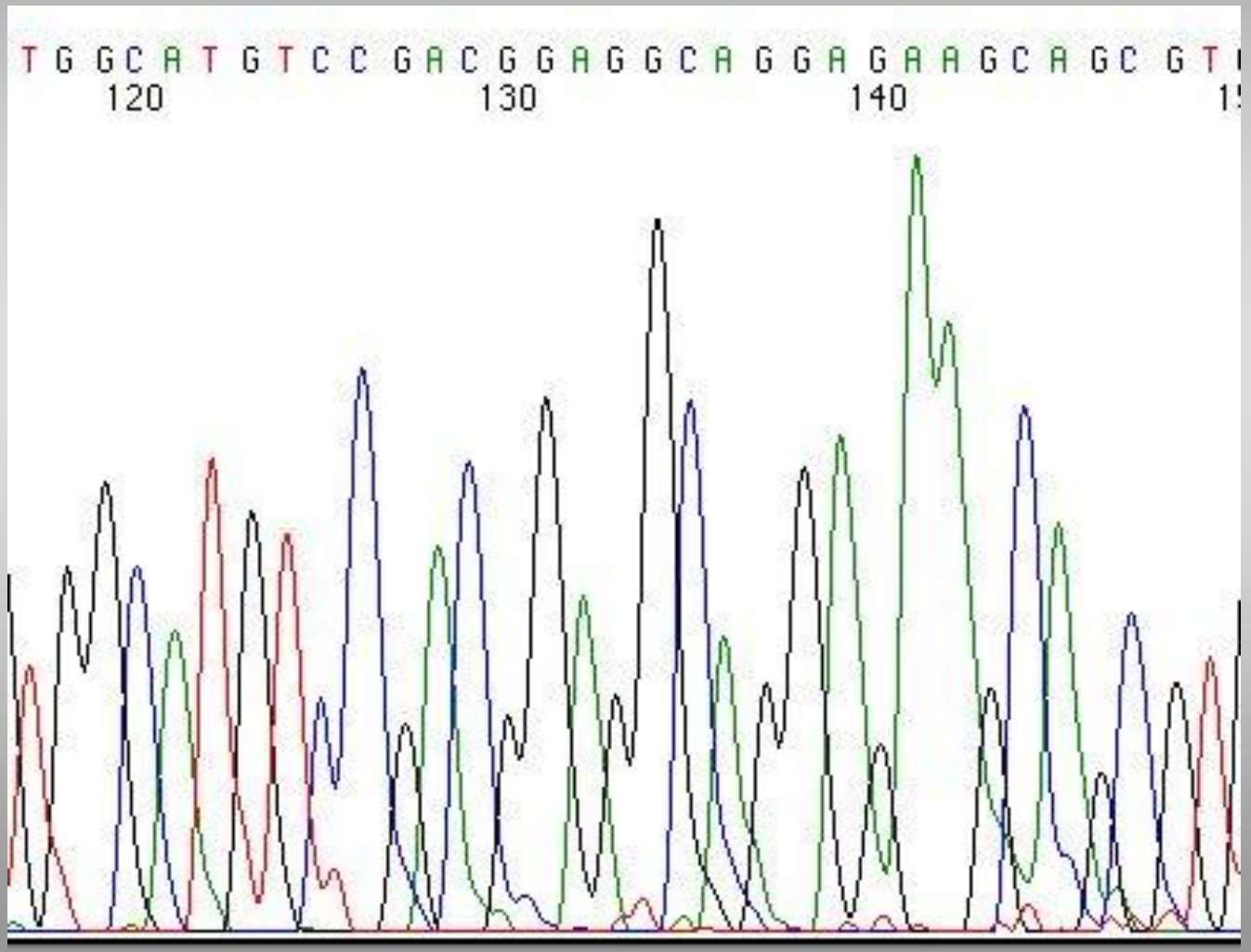
DNA amplification by PCR

- Requires knowledge of targeted sequence
- Specificity is dictated by two short (~25 bases) synthetic single stranded DNA molecules or oligonucleotides (primers).
- Mis-priming
- Preferential amplification of normal allele (PCR drop out)

Mutation Detection Techniques

- Sanger Sequencing
- Next Generation Sequencing (massive parallel sequencing)
- Gel electrophoresis

Sanger Sequencing



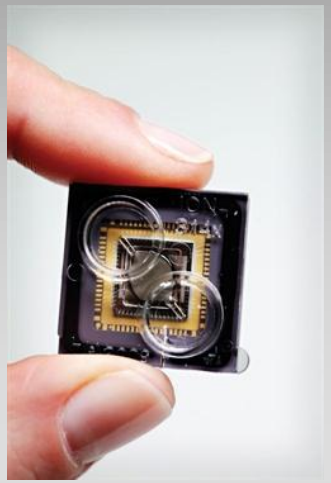
Sanger Sequencing

- Gold standard
- Well established (~20years)
- Robust

- one reaction = 1 sequencing reaction
- optimal sequencing length (500bp-900bp)
- sequencing 1 gene will require multiple reactions
- labour intensive and time consuming


ACAGGTCTAT CACCCTATTA ACCACTCACG GGAGCTCTCC ATGCATTTGG
TATTTTCGTC TGGGGGGTAT GCACGCGATA GCATTGCGAG ACGCTGGAGC
CGGAGCACCC TATGTCGCAG TATCTGTCTT TGATTCCTGC CTCATCCTAT
TATTTATCGC ACCTACGTTC AATATTACAG GCGAACATAC TTACTAAAGT
GTGTTAATTA ATTAATGCTT GTAGGACATA ATAATAACAA TTGAATGTCT
GCACAGCCAC TTTCCACACA GACATCATAA CAAAAAATTT CCACCAAACC
CCCCCTCCCC CGCTTCTGGC CACAGCACTT AAACACATCT CTGCCAAACC
CCAAAAACAA AGAACCCTAA CACCAGCCTA ACCAGATTTC AAATTTTATC
TTTTGGCGGT ATGCACTTTT AACAGTCACC CCCCAACTAA CACATTATTT
TCCCCTCCCA CTCCCATACT ACTAATCTCA TCAATACAAC CCCCGCCCAT
CCTACCCAGC ACACACACAC CGCTGCTAAC CCCATACCCC GAACCAACCA
TTCTTTCATG GGGAAGCAGA TTTGGGTACC ACCCAAGTAT TGACTCACCC
ATCAACAACC GCTATGTATT TCGTACATTA CTGCCAGCCA CCATGAATAT
TGTACGGTAC CATAAATACT TGACCACCTG TAGTACATAA AAACCCAATC
CACATCAAAA CCCCCTCCCC ATGCTTACAA GCAAGTACAG CAATCAACCC
TCAACTATCA CACATCAACT GCAACTCCAA AGCCACCCCT CACCCACTAG
GATACCAACA AACCTACCCA CCCTTAACAG TACATAGTAC ATAAAGCCAT
TTACCGTACA TAGCACATTA CAGTCAAATC CCTTCTCGTC CCCATGGATG
ACCCCCCTCA GATAGGGGTC CCTTGACCAC CATCCTCCGT GAAATCAATA
TCCCGCACAA GAGTGCTACT CTCCTCGCTC CGGGCCCATA ACACTTGGGG
GTAGCTAAAG TGAACTGTAT CCGACATCTG GTTCCTACTT CAGGGTCATA
AAGCCTAAAT AGCCCACACG TTCCCCTTAA ATAAGACATC ACGATGGATC
ACAGGTCTAT CACCCTATTA ACCACTCACG GGAGCTCTCC ATGCATTTGG
TATTTTCGTC TGGGGGGTAT GCACGCGATA GCATTGCGAG ACGCTGGAGC
CGGAGCACCC TATGTCGCAG TATCTGTCTT TGATTCCTGC CTCATCCTAT
TATTTATCGC ACCTACGTTC AATATTACAG GCGAACATAC TTACTAAAGT
GTGTTAATTA ATTAATGCTT GTAGGACATA ATAATAACAA TTGAATGTCT
GCACAGCCAC TTTCCACACA GACATCATAA CAAAAAATTT CCACCAAACC
CCCCCTCCCC CGCTTCTGGC CACAGCACTT AAACACATCT CTGCCAAACC
CCAAAAACAA AGAACCCTAA CACCAGCCTA ACCAGATTTC AAATTTTATC
TTTTGGCGGT ATGCACTTTT AACAGTCACC CCCCAACTAA CACATTATTT
TCCCCTCCCA CTCCCATACT ACTAATCTCA TCAATACAAC CCCCGCCCAT
CCTACCCAGC ACACACACAC CGCTGCTAAC CCCATACCCC GAACCAACCA
TTCTTTCATG GGGAAGCAGA TTTGGGTACC ACCCAAGTAT TGACTCACCC
ATCAACAACC GCTATGTATT TCGTACATTA CTGCCAGCCA CCATGAATAT
TGTACGGTAC CATAAATACT TGACCACCTG TAGTACATAA AAACCCAATC
CACATCAAAA CCCCCTCCCC ATGCTTACAA GCAAGTACAG CAATCAACCC
TCAACTATCA CACATCAACT GCAACTCCAA AGCCACCCCT CACCCACTAG
GATACCAACA AACCTACCCA CCCTTAACAG TACATAGTAC ATAAAGCCAT
TTACCGTACA TAGCACATTA CAGTCAAATC CCTTCTCGTC CCCATGGATG
ACCCCCCTCA GATAGGGGTC CCTTGACCAC CATCCTCCGT GAAATCAATA
TCCCGCACAA GAGTGCTACT CTCCTCGCTC CGGGCCCATA ACACTTGGGG
GTAGCTAAAG TGAACTGTAT CCGACATCTG GTTCCTACTT CAGGGTCATA
AAGCCTAAAT AGCCCACACG TTCCCCTTAA ATAAGACATC ACGATGGATC
ACAGGTCTAT CACCCTATTA ACCACTCACG GGAGCTCTCC ATGCATTTGG

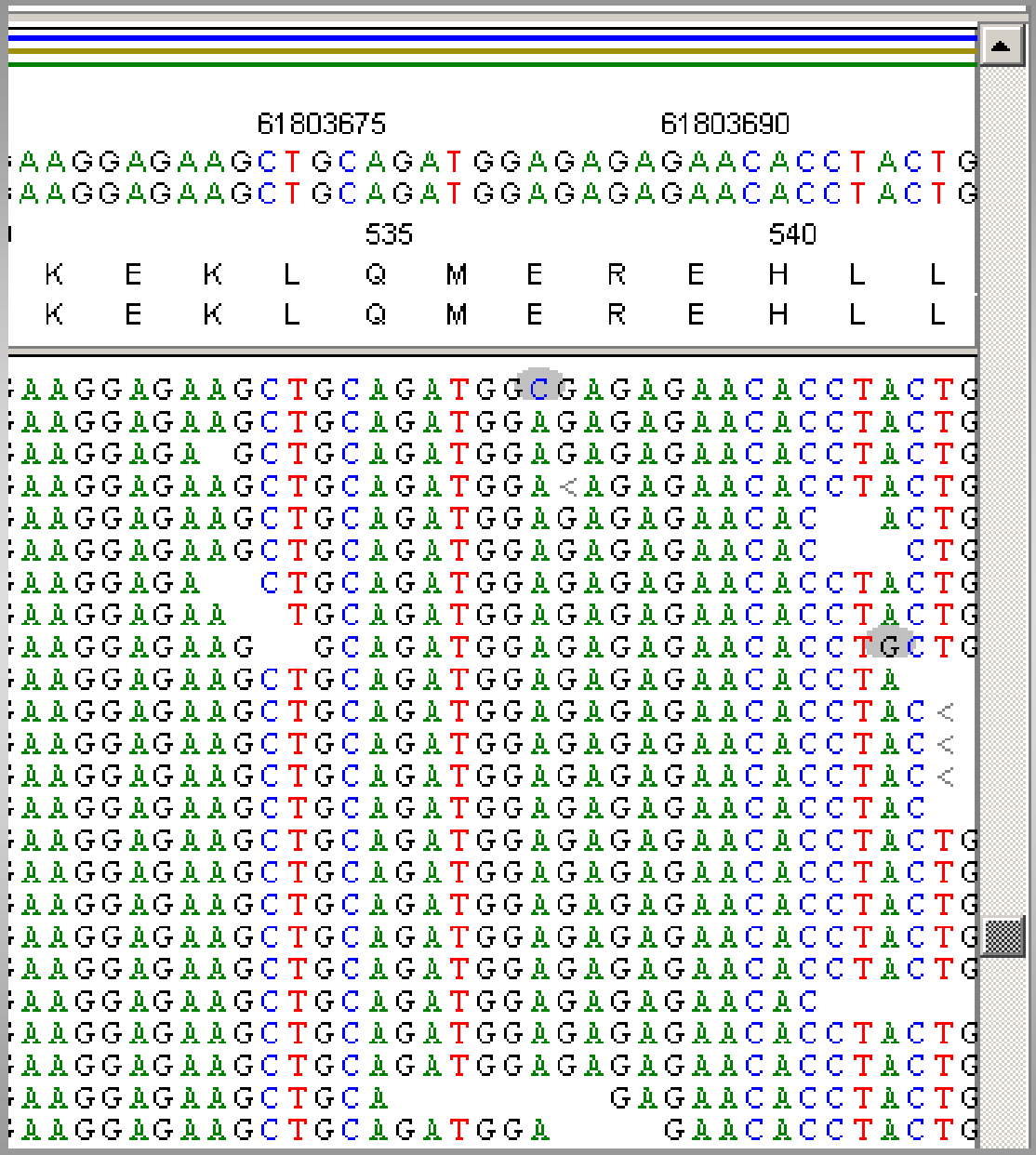
Next Generation Sequencing

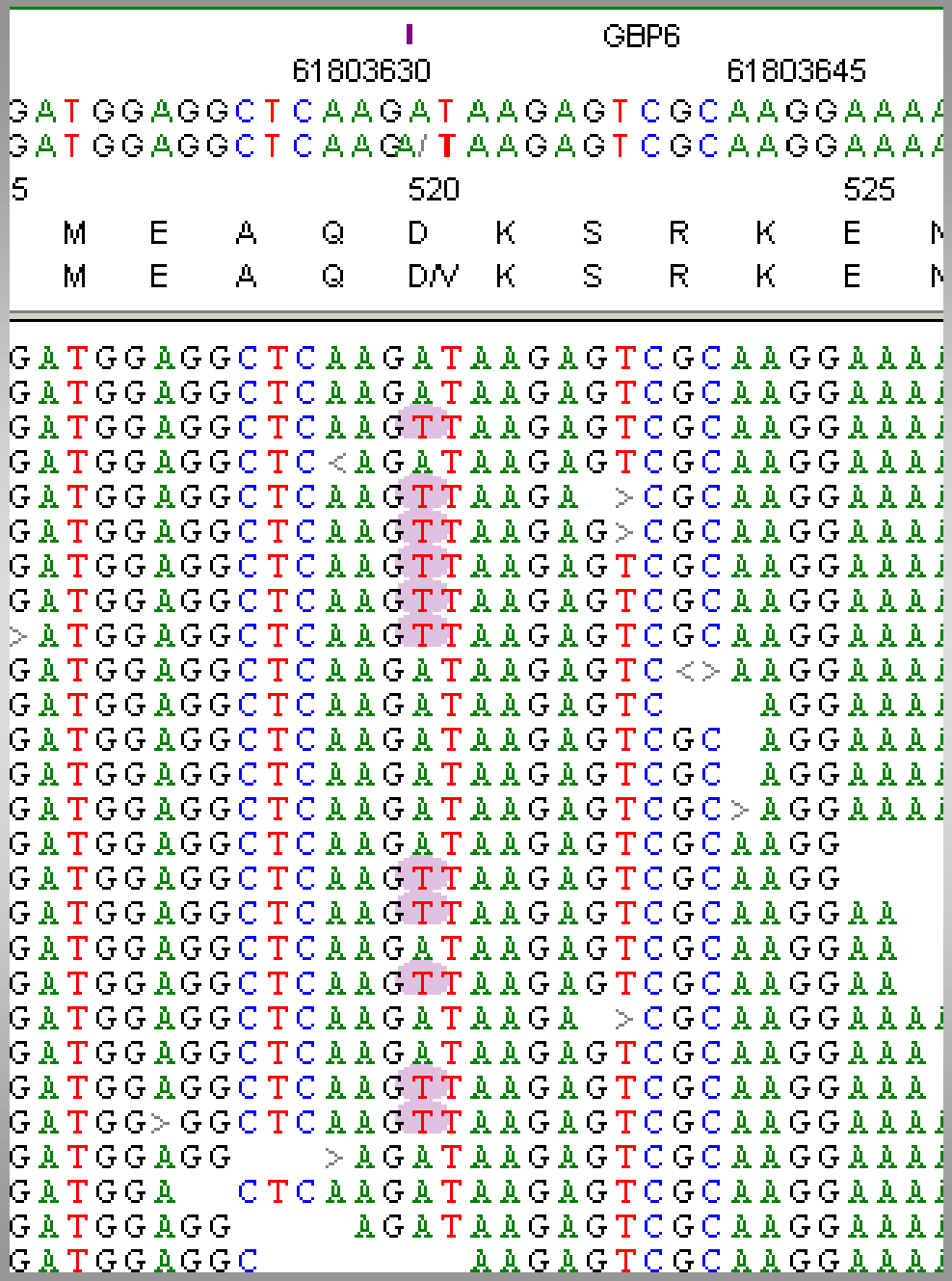


=> 1 million

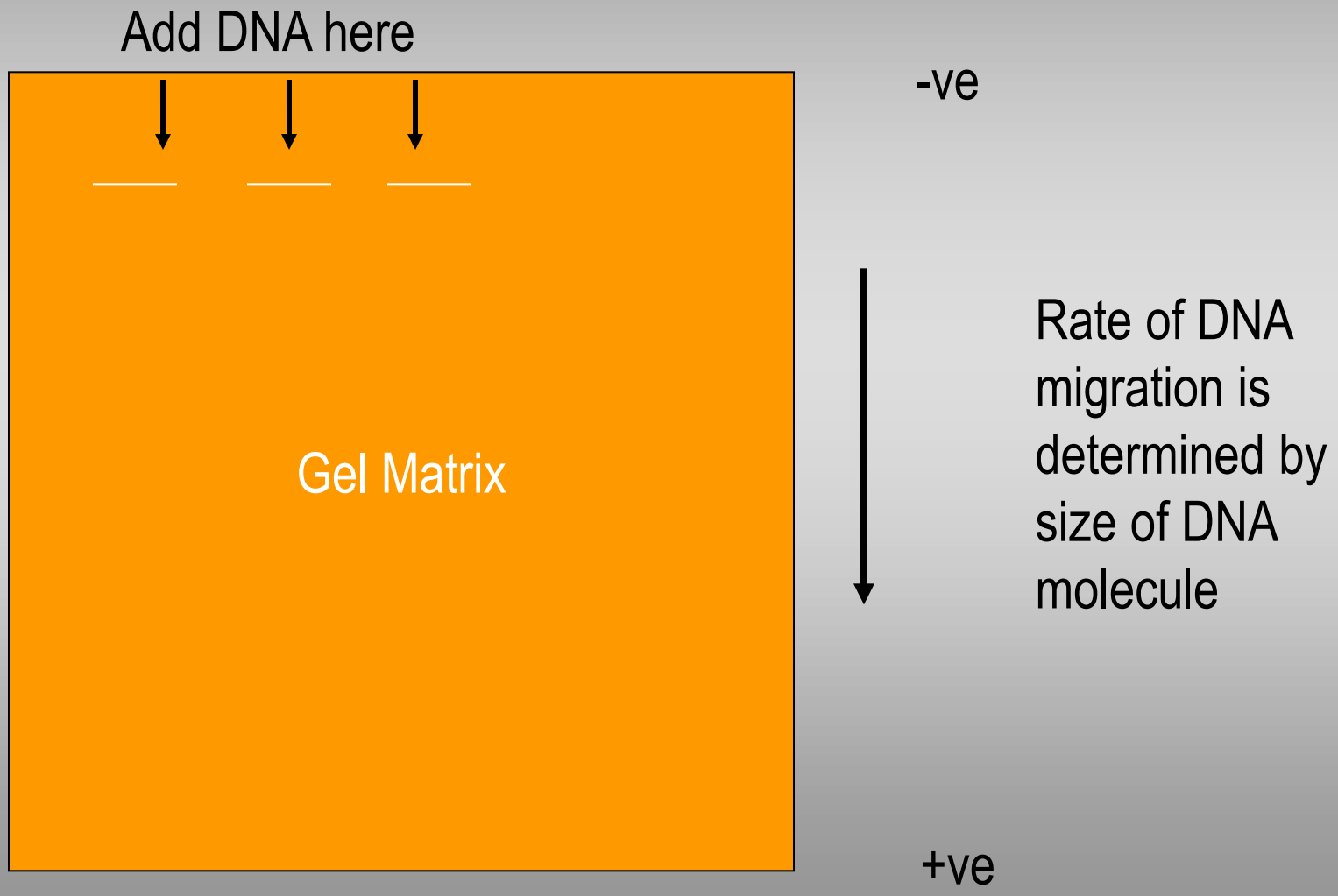
- Very expensive (getting cheaper)
- High volume of data
- High number of genetic variants of unknown significance
- Require sophisticated bio-informatics
- Good for multi-gene analysis (exome or whole genome)


GBP6
61803585
61803600
61803615
61803630
61803645
61803660
61803675
61803690
505
510
515
520
525
530
535
540
Q E L L K Q K L Q E Q Q Q Q M E A Q D K S R K E N I A Q L K E K L Q M E R E H L L
Q E L L K Q K L Q E Q Q Q Q M E A Q D/V K S R K E N I A Q L K E K L Q M E R E H L L
20
25
30
35
40
45
50
55
60
65


61803675
61803690
535
540
K E K L Q M E R E H L L
K E K L Q M E R E H L L


GBP6
61803630
61803645
5
520
525
M E A Q D K S R K E N
M E A Q D/V K S R K E N

Gel Electrophoresis



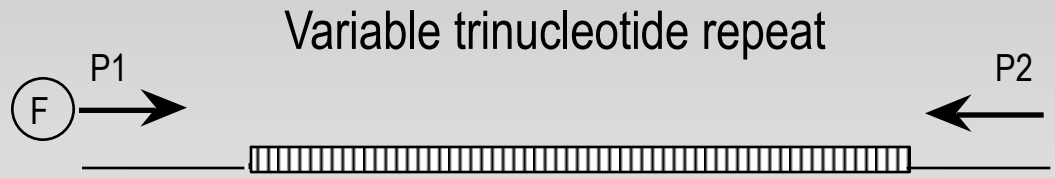
Rate of DNA migration is determined by size of DNA molecule

Triplet Repeat Analysis

Huntington's Disease

- Autosomal dominant disorder
- Incidence of 1 in 10,000
- Neurodegenerative disorder
- Triplet repeat expansion
- Onset in the third decade
- Progressive deterioration of cognitive function leading to dementia
- Associated with abnormal movement

Huntington's disease CAG PCR



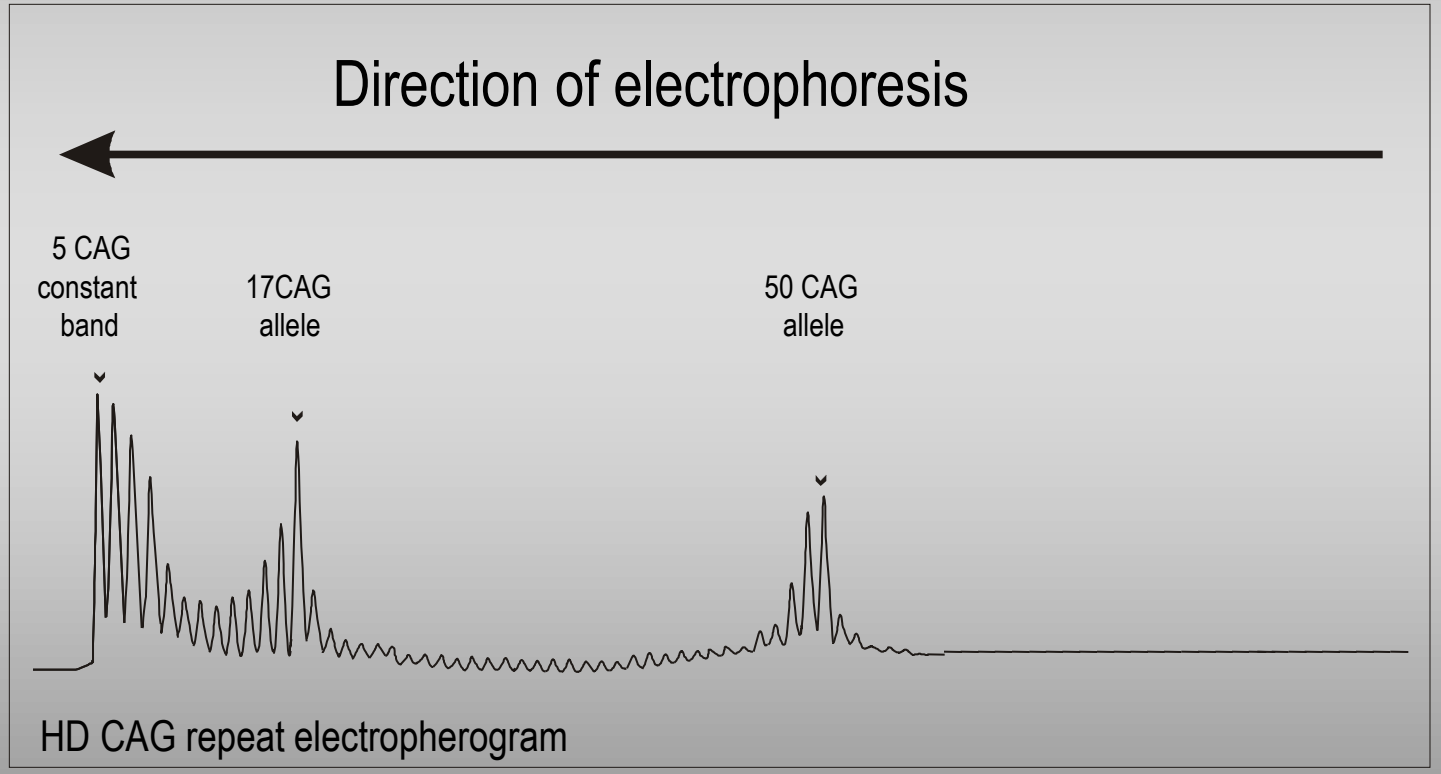
Fluorescent primer P1

PCR amplification of CAG repeat within the Huntington gene on chromosome 4

Huntington's disease CAG PCR

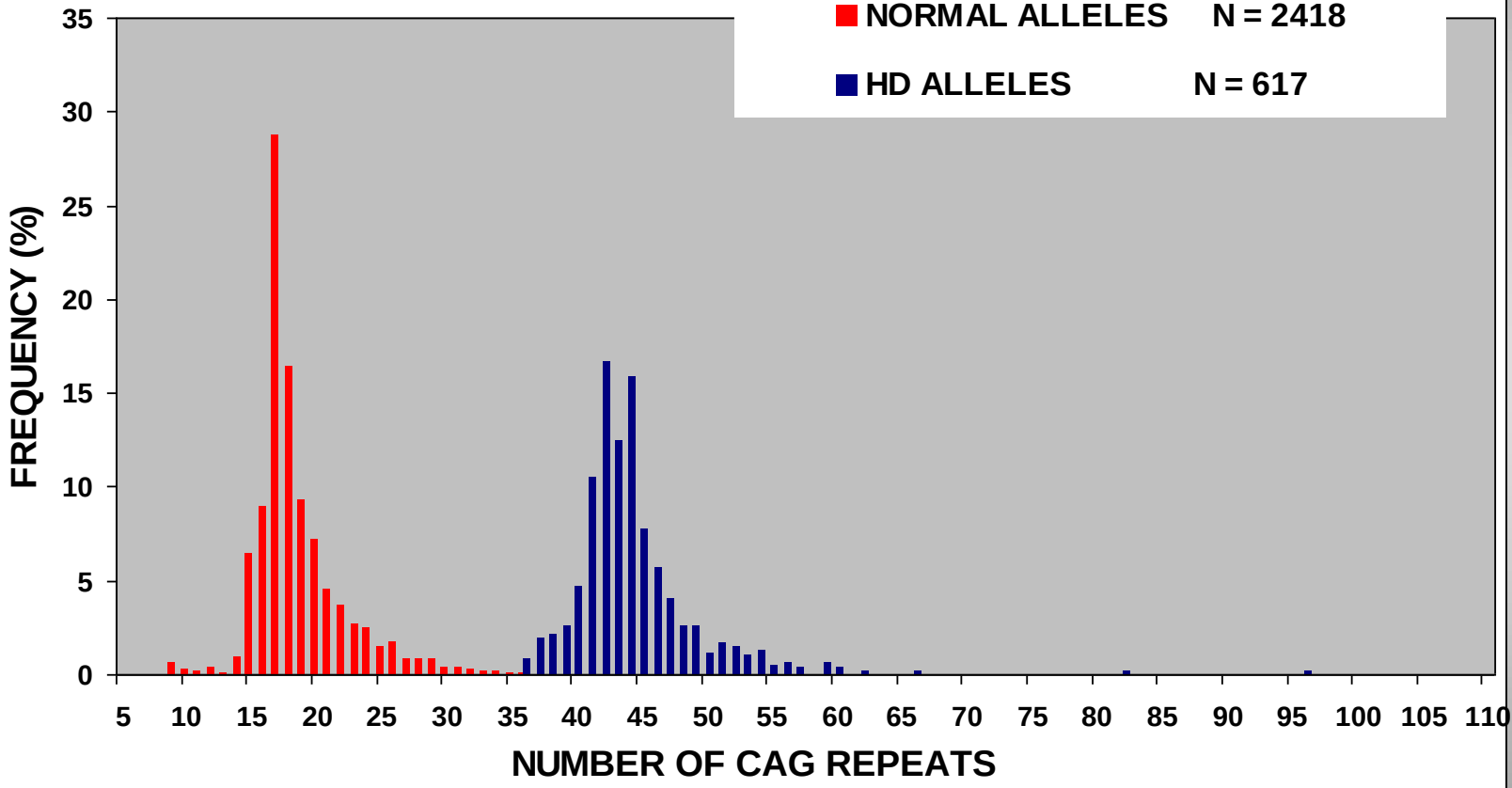
- Amplification of variable repeat gives a range of sizes or alleles.
- PCR products resolved on high resolution poly-acrylamide gel on automatic laser fluorescent sequencer for exact sizing.
- Distinct size ranges are seen for affected HD population and normal population.

Scoring of HD electropherogram



HD CAG repeat electropherogram

FREQUENCY OF CAG REPEAT ALLELES IN THE SCOTTISH NORMAL AND HUNTINGTON'S DISEASE POPULATIONS

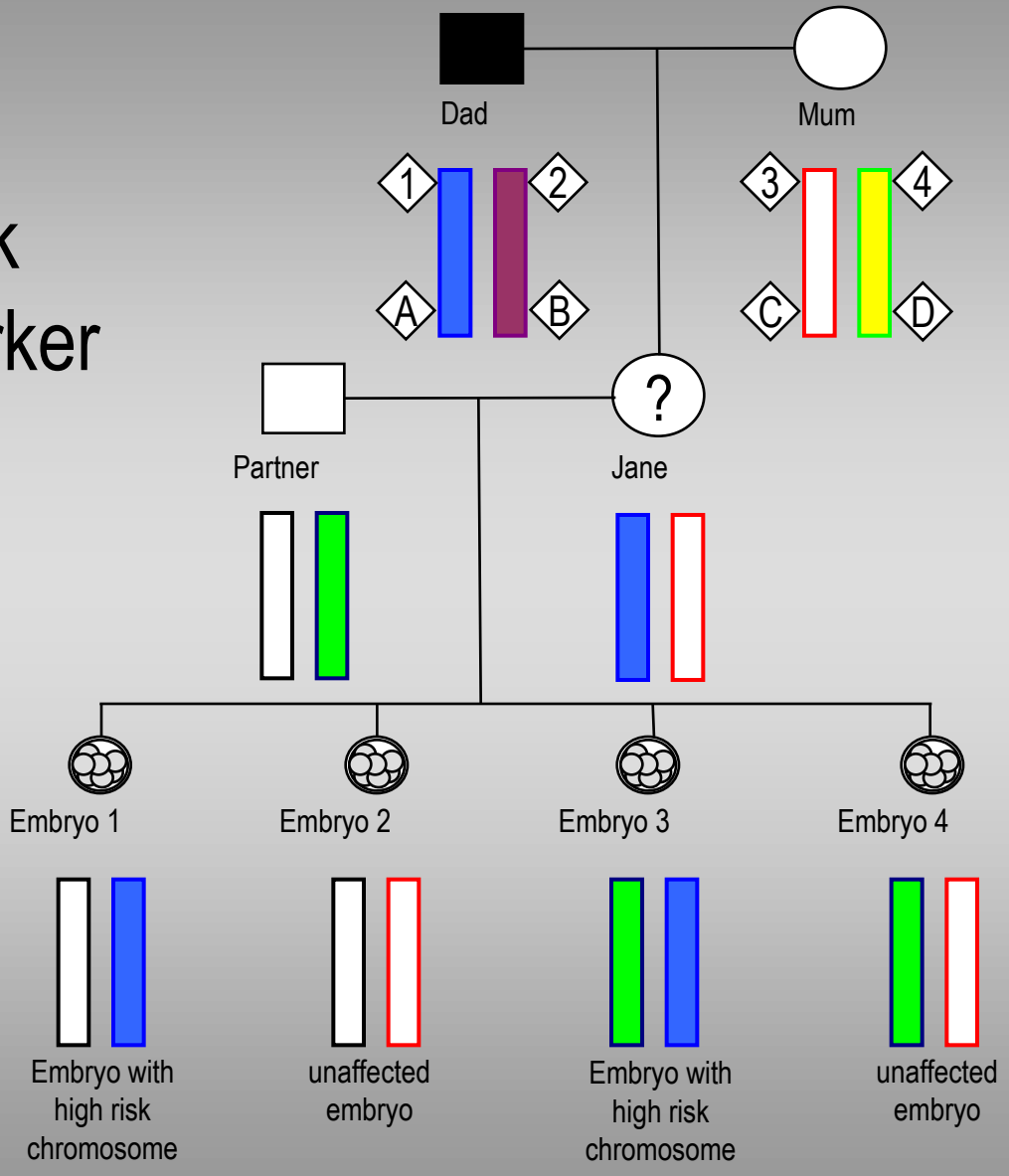


HD PCR size ranges

- Normal individuals in our population have alleles in the 8 to 35 CAG repeat range.
- A repeat size of 36 repeats or greater is diagnostic of HD.
- Alleles between 36 and 39 repeats are frequently associated with later onset of symptoms.
- Alleles between 27 and 35 repeats are potentially unstable and rare expansions into the affected range have been seen.

Identifying high risk haplotype by link marker analysis

Exclusion testing



Which technique for what type of test

- Direct DNA sequencing: PCR fragments of 150-850 bp for mutation scanning.
- Next Gen sequencing: Multi-gene analysis
- PCR then Gel electrophoresis
  - fluorescent sizing of products:
    - trinucleotide repeats
    - microsatellite repeats (up to 400bp)
- Southern blotting of digested DNA: methylation sensitivity and larger size range. 500bp to 20kb.

Case 1: Craniosynostosis

Definition: Premature closure of the fibrous joints between the bones of the skull

- Prevalence: 343 per million
- Saggittal synostosis most common account for 57% (M>F)
- Coronal synostosis accounts for 18%-29% (F>M)

Isolated craniosynostosis accounts for the majority of cases

- Most cases are sporadic
- Familial isolated craniosynostosis
  - 2% of sagittal synostosis
  - 8% of coronal synostosis
- Autosomal dominant mode of inheritance
    - Syndromic craniosynostosis
    - >90 syndromes characterised

Fibroblast Growth Factor Receptor 2
FGFR2

Chromosomal location: 10q26

Point mutations in the third immunoglobulin loop and transmembrane domain are associated with 5 cranio-synostotic syndromes.

These mutations lead to constitutive activation of the receptor.

Mutations	Syndrome
	Crouzon
	Jackson-Weiss
	Pfeiffer
	Apert
	Beare-Stevenson

Clinical heterogeneity

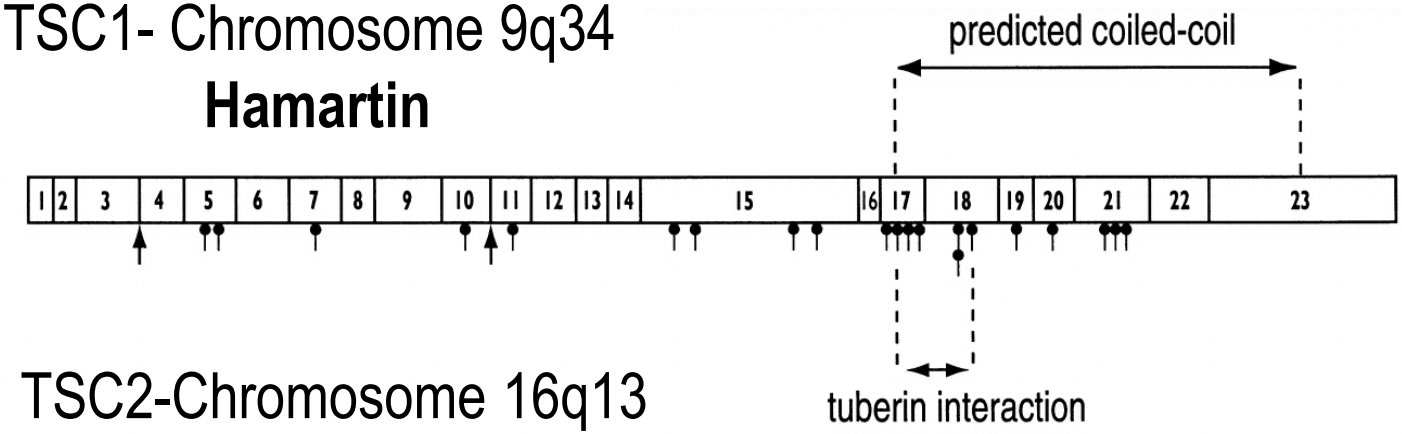
http://www.nidcr.nih.gov/NR/rdonlyres/3C4AFD26-75DB-41D8-A3CF-E00470739212/0/diagfibroblast.gif

Case 2: Tuberous Sclerosis

- Autosomal dominant condition
- Characterized by harmatomatous lesions
- Multisystem involvement
- Prevalence of 1 in 6,000
- 2/3 sporadic (new mutations)
- 1/3 familial

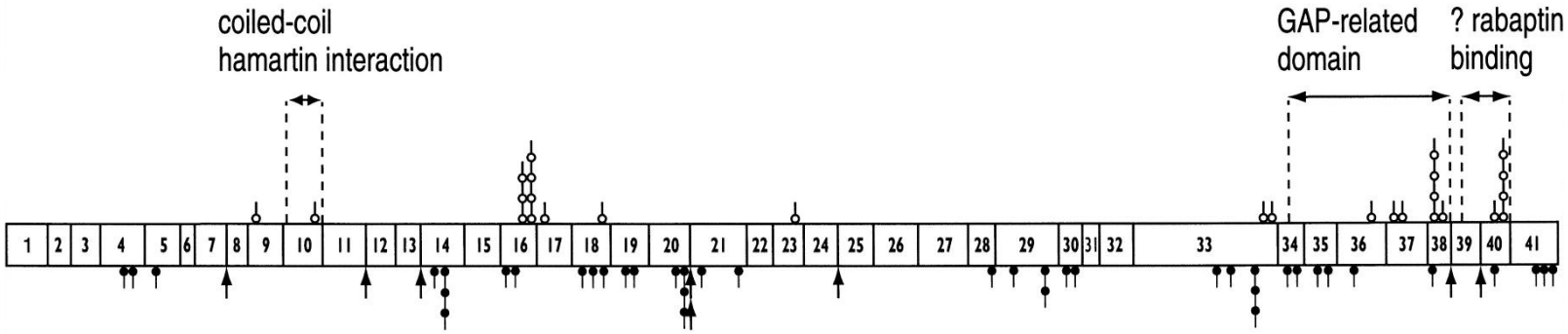
TSC1- Chromosome 9q34

Hamartin

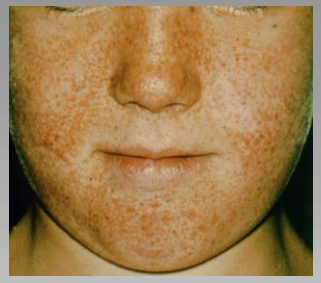


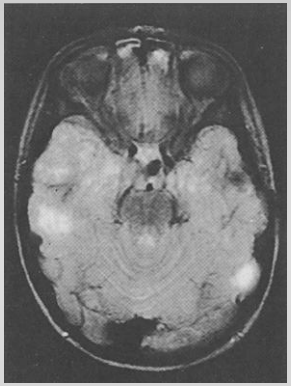
TSC2-Chromosome 16q13

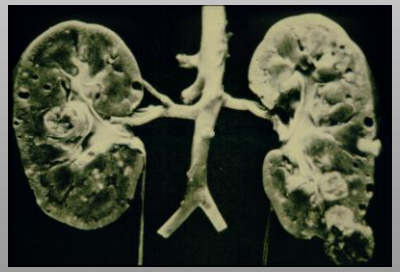
Tuberin



Gene	% of Familial cases	% of Sporadic cases
TSC1	50%	~10%
TSC2	50%	~60% - 70%







TSC1
TSC2

Epilepsy

Developmental delay

Behavioural problems

Locus heterogeneity

Summary 1
Types of mutation in DNA sequences

The cat sat on the mat	**Wild type**
The cat	**Stop / Nonsense**
The car sat on the mat	**Missense**
The cat spa to nth ema t	**Insertion**
The cas ato nt hem at	**Deletion**
The **cat cat** sat on the mat	**Triplet expansion** (Dynamic mutation)
The **tas tac** on the mat	**Inversion**

Summary 2
Mutation Criteria

- Does it affect the function of the protein ?
- It is in a conserved region of the protein ?
- Does it co-segregate with the disorder in the family ?
- Is the change seen in the normal population ?

Summary 3
A Mutation Can Cause Disease by:

Abolition (Loss of function)

- Due to non-functioning or truncated protein
- Haploinsufficiency
- Dominant negative

Modification

- Creating a poorly functioning protein
- Abnormal activation of protein (over-expression)
- Gain of function of protein (novel function)