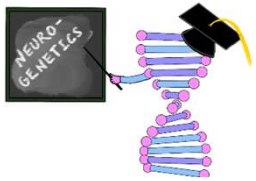
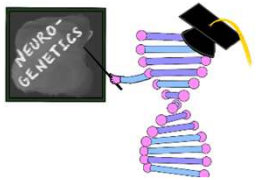
Introduction to Genetics

MODULE 0



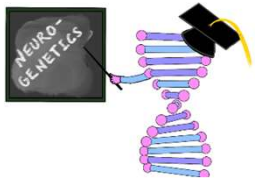
Learning Objectives

- Re-familiarize with basic genetic concepts
- Understand some of the language of genetics
- Understand the basics of testing and its limitations
- Develop strategies to understand the results of testing



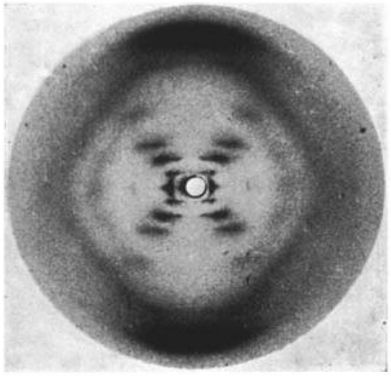
How Did We Get Here?

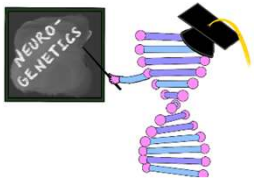
- 1822-1884: Gregor Mendel first discussed the idea that discrete inherited units give rise to observable physical characteristics
  - His work is ignored until 1889
- Hugo de Vries publishes a book called *Intracellular Pangenesis*
  - Inheritance of specific traits in organisms comes in particles he called "pangenes"
- 1909: Wilhelm Johannsen introduced the term "gene"



How Did We Get Here?

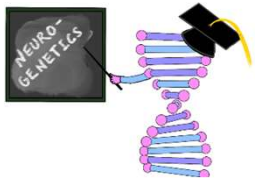
- 1940s and 50's: Deoxyribonucleic acid shown to be the molecular "substance" of genes
- Rosalind Franklin and Maurice Wilkins
  - X-ray crystallography 1952



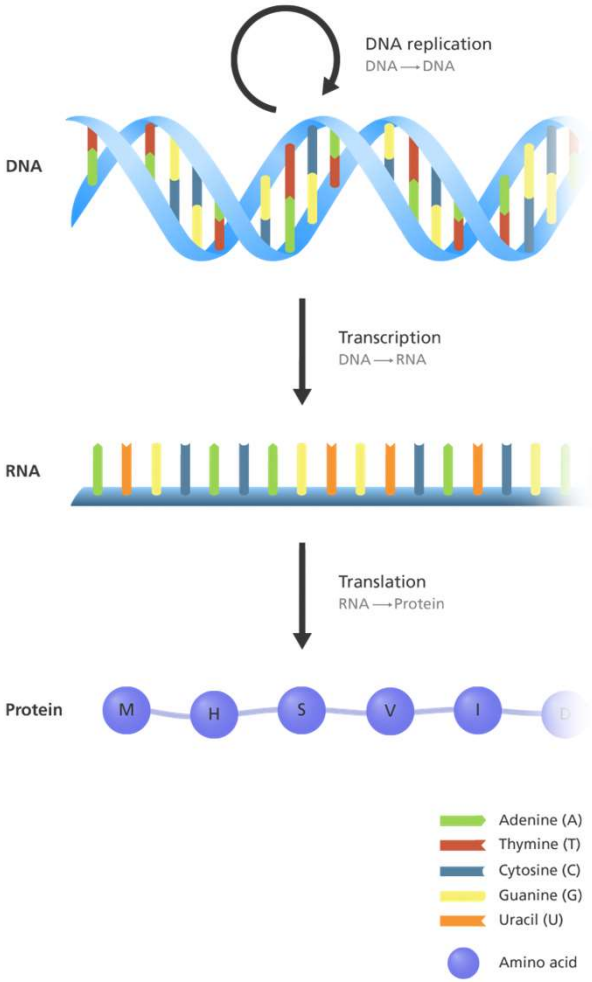


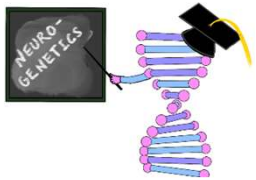
How Did We Get Here?

- 1953 - Watson and Crick publish the double helical structure of DNA
- 1956 - Modern karyotypes
- 1977 - Frederick Sanger develops chain-termination sequencing
- 1983 - First genetic disease mapped (Huntington's)
- 1990 - Human genome project officially launched
- 2001 - Working draft of human genome published
- 2003 - More complete draft of the human genome published



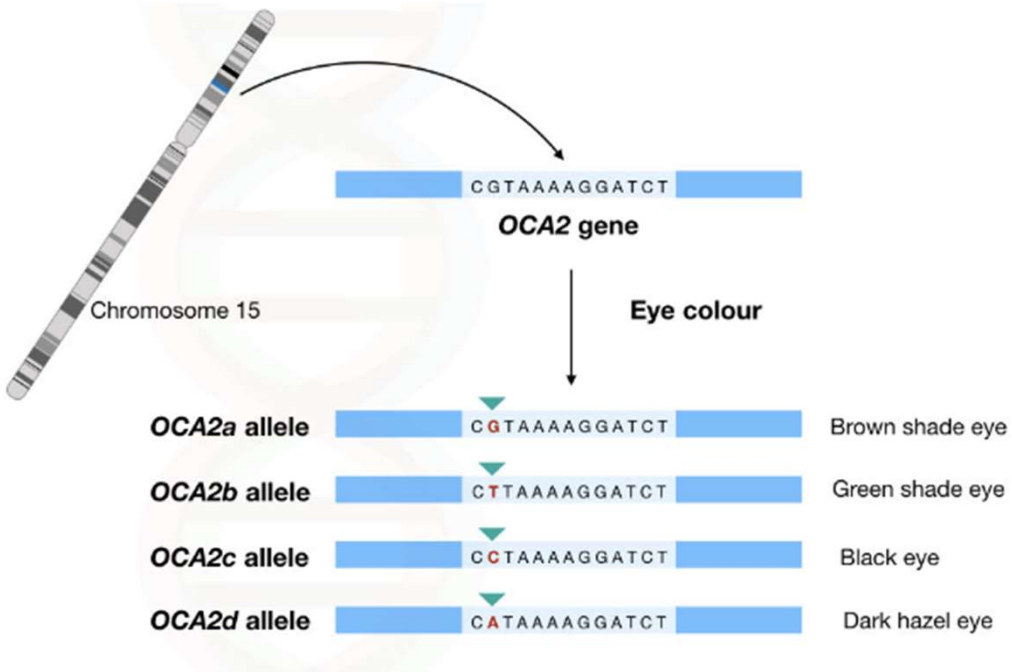
Central Dogma

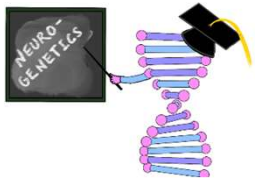




Basics About DNA

1. How many human chromosomes per cell?
2. What does 15q13.3 mean?
3. What is a gene?
4. How many human genes are there?
5. What is an allele?

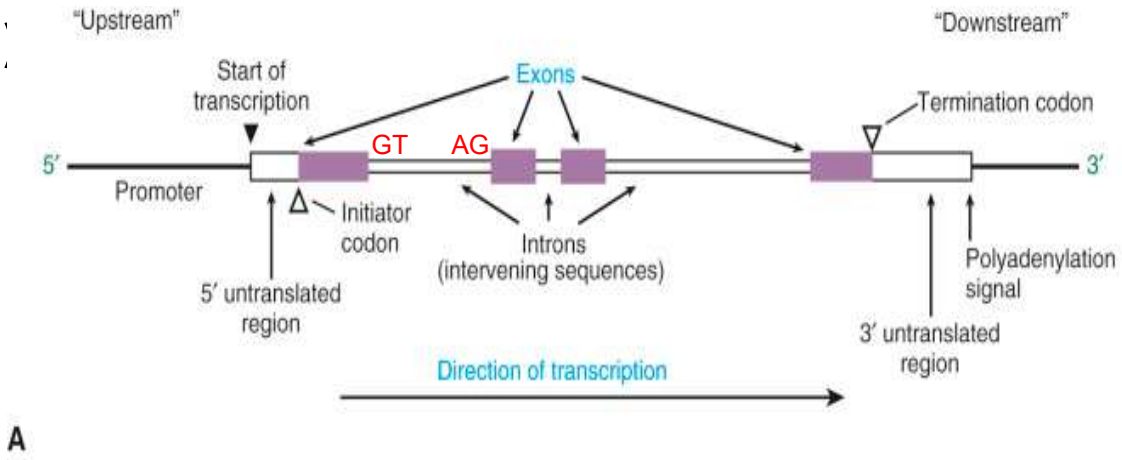




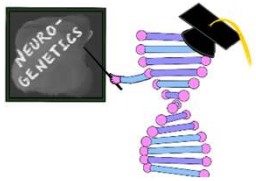
Gene vs Allele: Key Genetic Concepts Simplified (geneticeducation.co.in)

Typical Gene Structure

- Initiator codon AUG – methionine
- Type of variation on the level of single gene:
  - Single nucleotide substitution
    - Splice Site Variants
  - Insertion/Deletion (indel)
  - Trinucleotide repeat



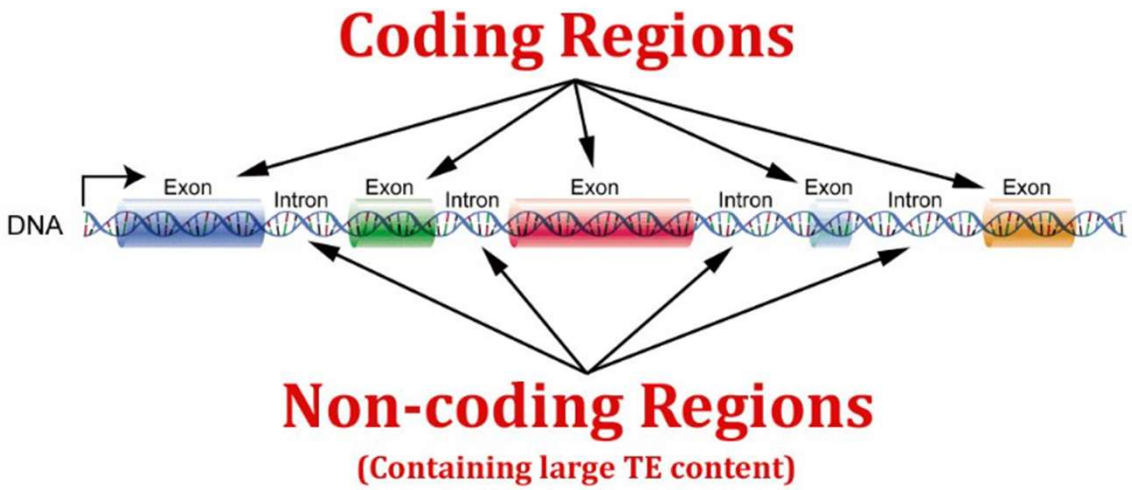
A

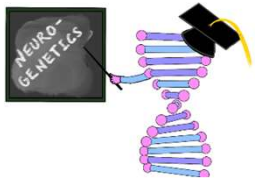


Nussbaum RL, McInnes RR, Willard HF. *Thompson and Thompson Genetics in Medicine.* Pg 17. (2007)

What Did We Learn from the Human Genome Project?

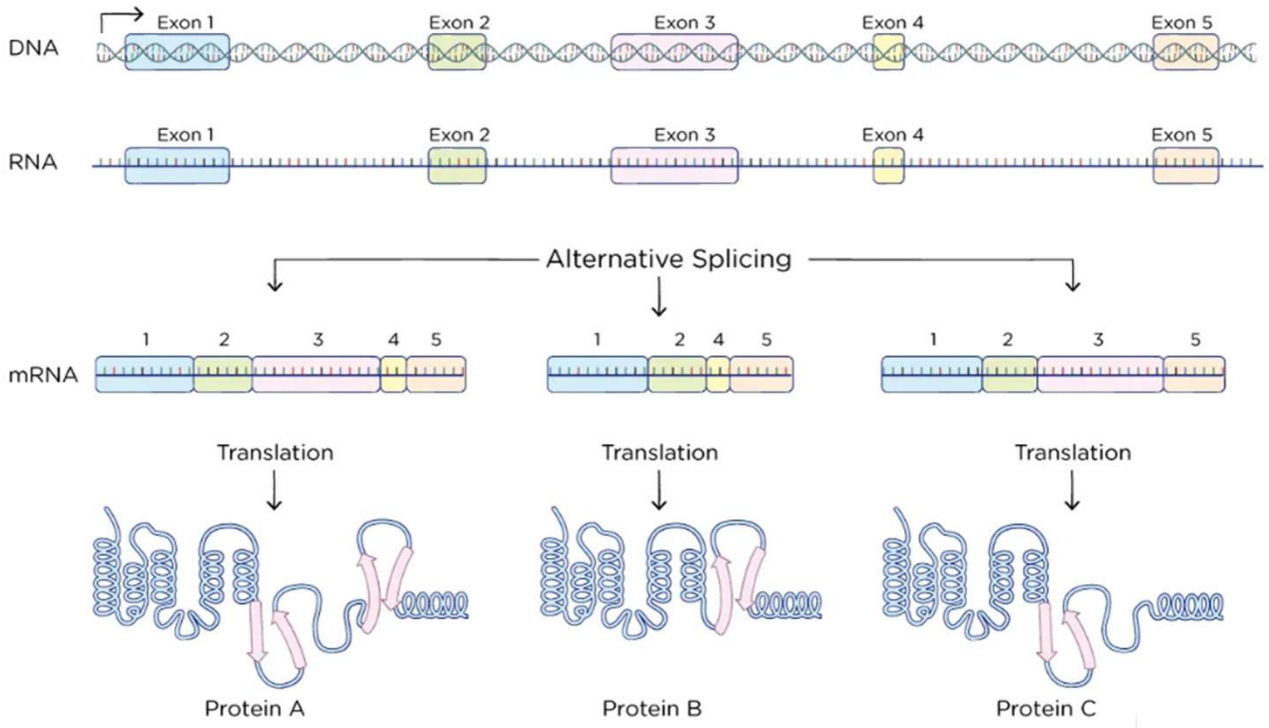
- 1.5 % of DNA is protein coding
  - Approximately 20,000 protein coding genes
  - 20,000 to 25,000 non-coding RNA genes
    - RNAs important for transcription, splicing, snoRNAs, miRNA

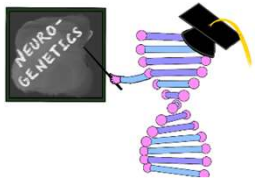




Structured Prediction and Probabilistic Graphical Models (ubc.ca)

mRNA Splicing

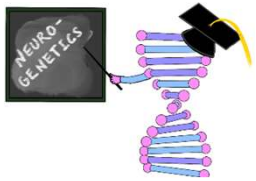




Alternative splicing - Wikipedia

The Genetic Code is Degenerate

- Single nucleotide substitutions
  - Synonymous variant
    - Ex: DNA: TTT > TTC → Amino acids: Phe > Phe
  - Missense variant
    - Ex: DNA: TTT > TTA → Amino acids: Phe > Leu
  - Nonsense variant
    - Ex: DNA: TAC > TAA → Amino acids: Tyr > STOP
- INDELs
  - ATGCCAACCTATATTAGT
    - Met-Pro-Thr-Tyr-Ile-Ser
  - ATGCCAACTATATTAGT
    - Met-Pro-Thr-Ile-Leu



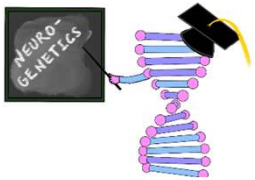
nonpolar polar basic acidic (stop codon)

1st base	2nd base U		2nd base C		2nd base A		2nd base G	
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine
	UUC	(Phe/F) Phenylalanine	UCC	(Ser/S) Serine	UAC	(Tyr/Y) Tyrosine	UGC	(Cys/C) Cysteine
	UUA	(Leu/L) Leucine	UCA	(Ser/S) Serine	UAA	Ochre (*Stop*)	UGA	Opal (*Stop*)
	UUG	(Leu/L) Leucine	UCG	(Ser/S) Serine	UAG	Amber (*Stop*)	UGG	(Trp/W) Tryptophan
C	CUU	(Leu/L) Leucine	CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine
	CUC	(Leu/L) Leucine	CCC	(Pro/P) Proline	CAC	(His/H) Histidine	CGC	(Arg/R) Arginine
	CUA	(Leu/L) Leucine	CCA	(Pro/P) Proline	CAA	(Gln/Q) Glutamine	CGA	(Arg/R) Arginine
	CUG	(Leu/L) Leucine	CCG	(Pro/P) Proline	CAG	(Gln/Q) Glutamine	CGG	(Arg/R) Arginine
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine
	AUC	(Ile/I) Isoleucine	ACC	(Thr/T) Threonine	AAC	(Asn/N) Asparagine	AGC	(Ser/S) Serine
	AUA	(Ile/I) Isoleucine	ACA	(Thr/T) Threonine	AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine
	AUG[A]	(Met/M) Methionine	ACG	(Thr/T) Threonine	AAG	(Lys/K) Lysine	AGG	(Arg/R) Arginine
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine
	GUC	(Val/V) Valine	GCC	(Ala/A) Alanine	GAC	(Asp/D) Aspartic acid	GGC	(Gly/G) Glycine
	GUA	(Val/V) Valine	GCA	(Ala/A) Alanine	GAA	(Glu/E) Glutamic acid	GGA	(Gly/G) Glycine
	GUG	(Val/V) Valine	GCG	(Ala/A) Alanine	GAG	(Glu/E) Glutamic acid	GGG	(Gly/G) Glycine

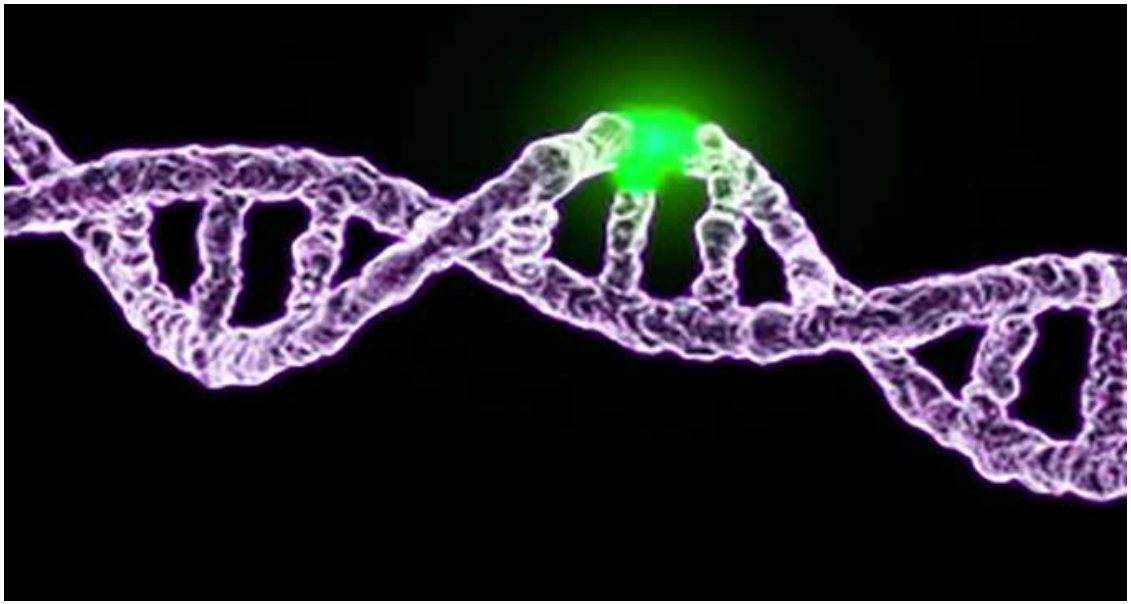
fire.biol.wwu.edu/trent/trent/primersequence.htm

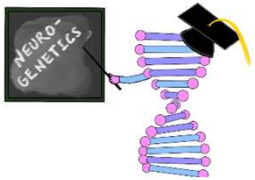
Definitions

<u>Genotype</u> – alleles at a particular locus	AA (homozygous)	Aa (heterozygous)	aa (homozygous)
<u>Phenotype</u> – observable physical characteristics			

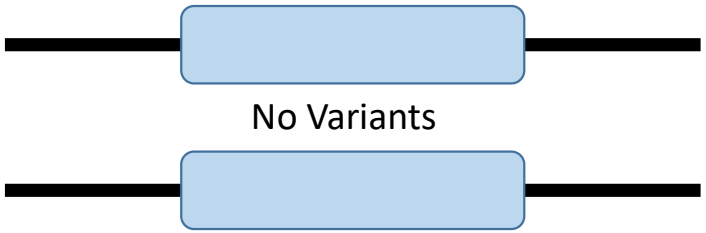


Mutation/Pathogenic Variant



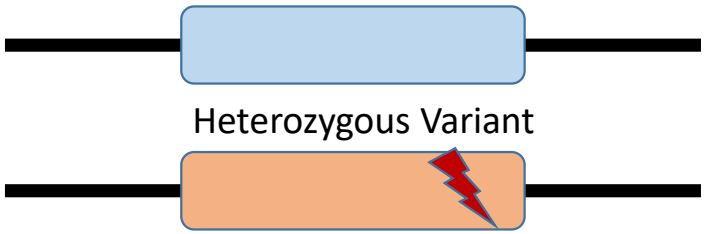


NO DISEASE



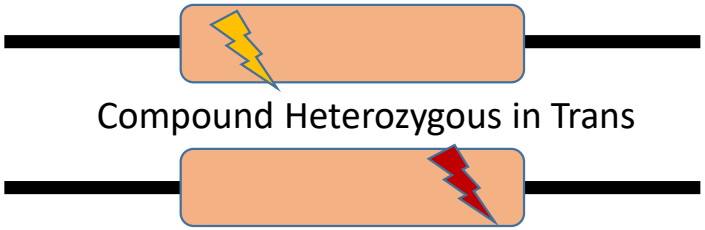
No Variants

Monoallelic/AD



Heterozygous Variant

Biallelic/AR



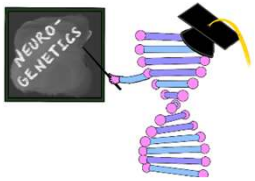
Compound Heterozygous in Trans

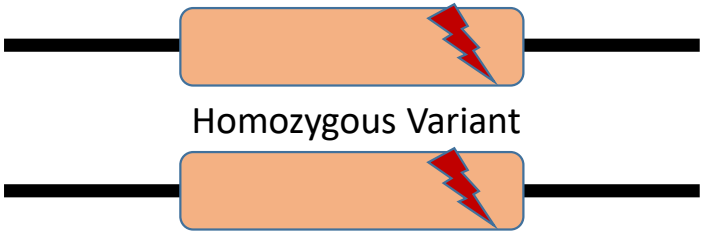
Unaffected Allele

Affected Allele

AD Autosomal Dominant

AR Autosomal Recessive

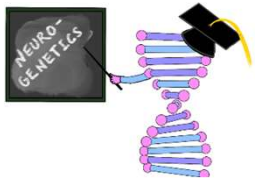




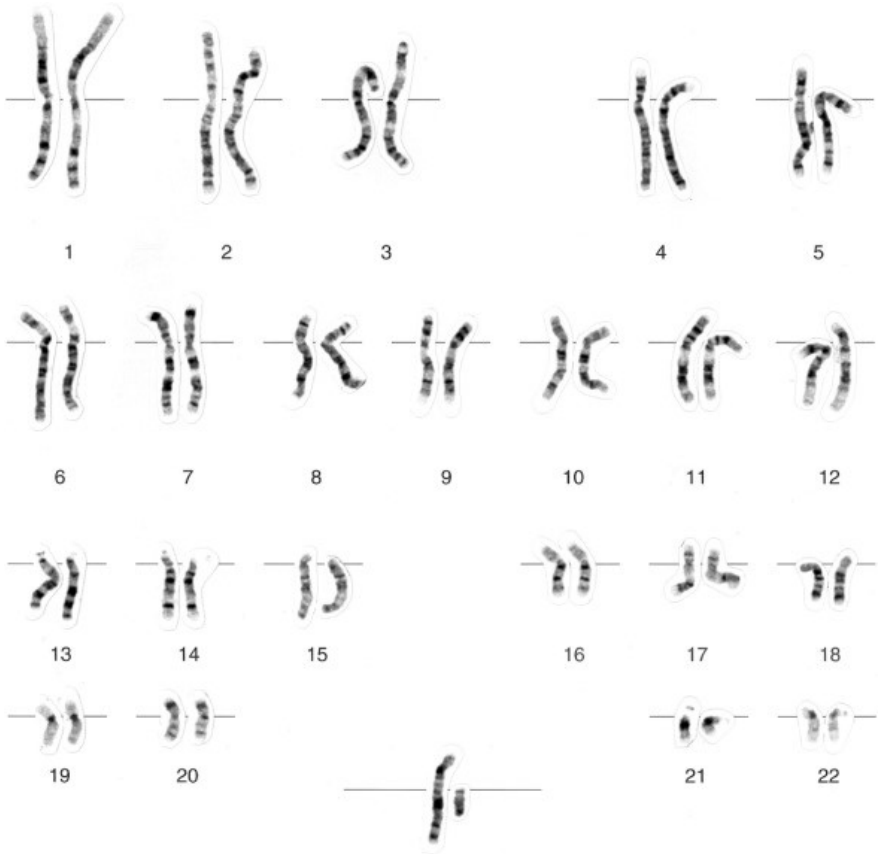
Homozygous Variant

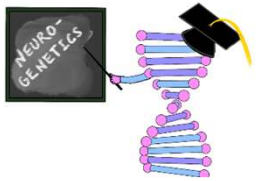
Definitions (Continued)

- Pleiotropy - when one gene influences two or more seemingly unrelated phenotypic traits
- Genetic heterogeneity - two or more genetic loci produce the same or similar phenotypes
- Penetrance – the probability of a gene or trait to be expressed
- Variable expressivity - the range of signs and symptoms that can occur in different people with the same genetic condition



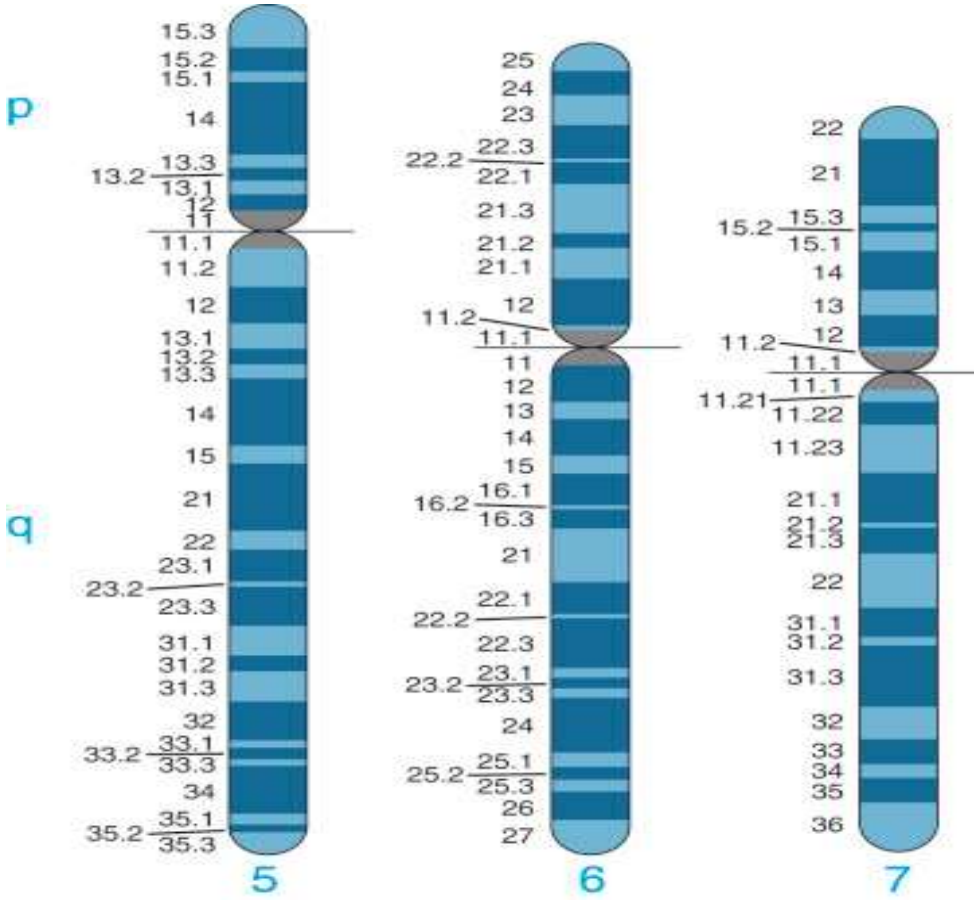
Karyotype (G-banded)

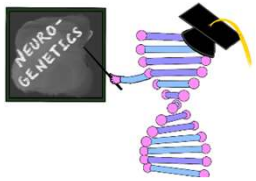




Nussbaum RL, McInnes RR, Willard HF. *Thompson and Thompson Genetics in Medicine.* Pg 17. (2007)

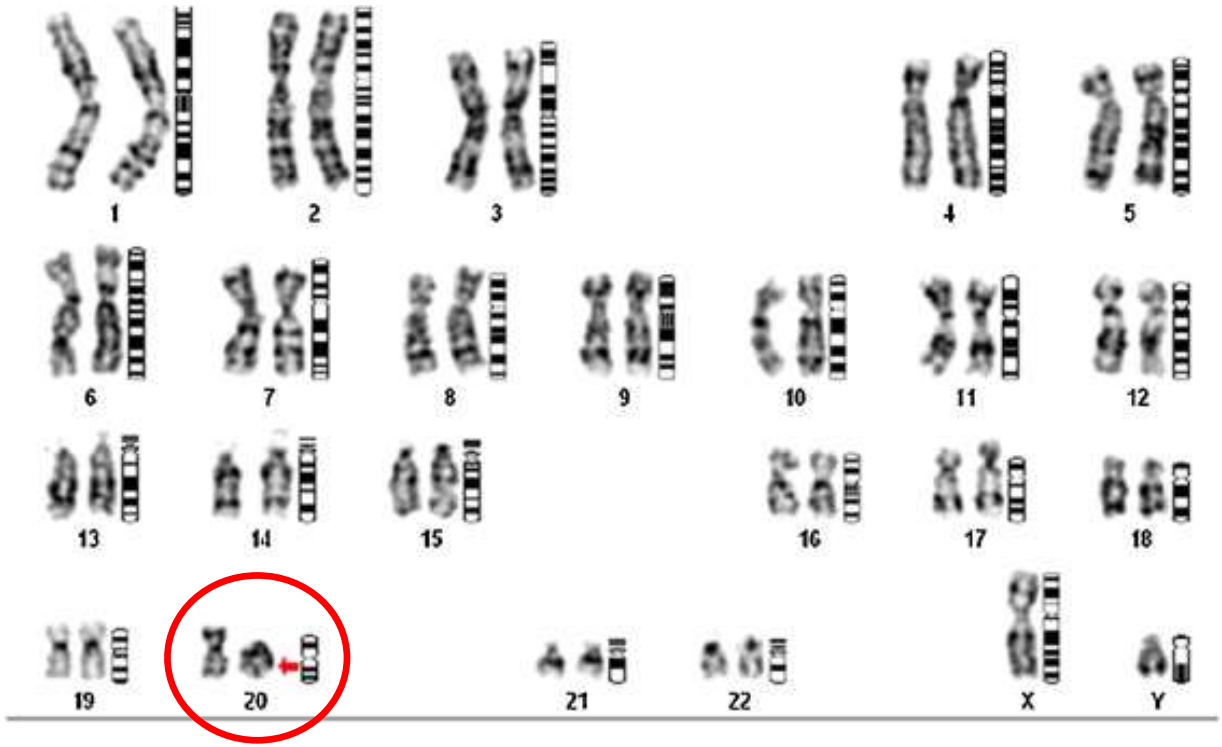
Chromosomal Nomenclature

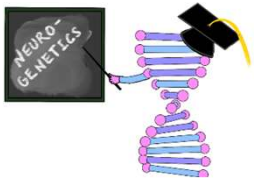




Principles of Clinical Cytogenetics and Genome Analysis - Thompson & Thompson Genetics in Medicine, 8th Edition (doctorlib.info)

Karyotypes are Best for Identifying Structural Problems





Kalane U, Datar C, Kalane S. Ring chromosome 20 syndrome-a rare chromosomal cause of refractory epilepsy in children. *International Journal of Epilepsy*. 4(2017) 87-89.

Fluorescence In Situ Hybridization (FISH)

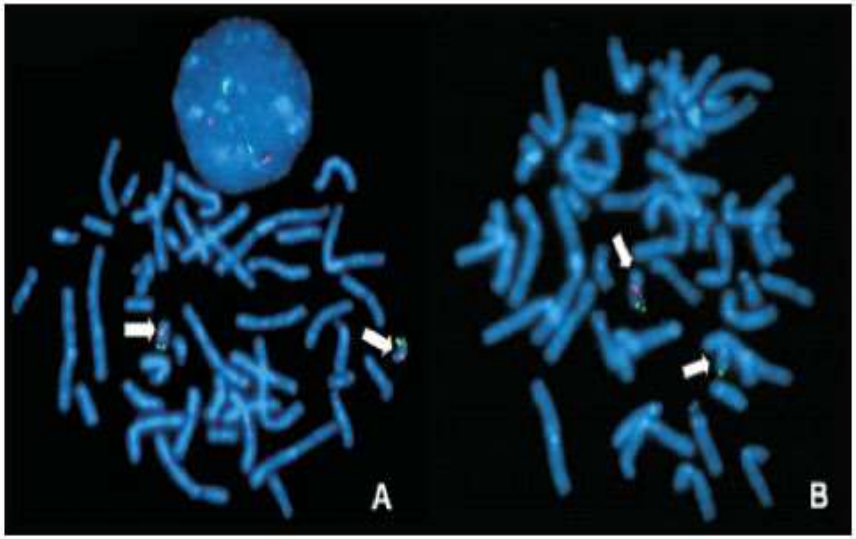
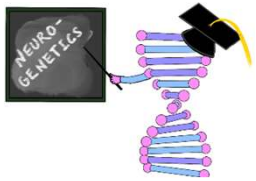
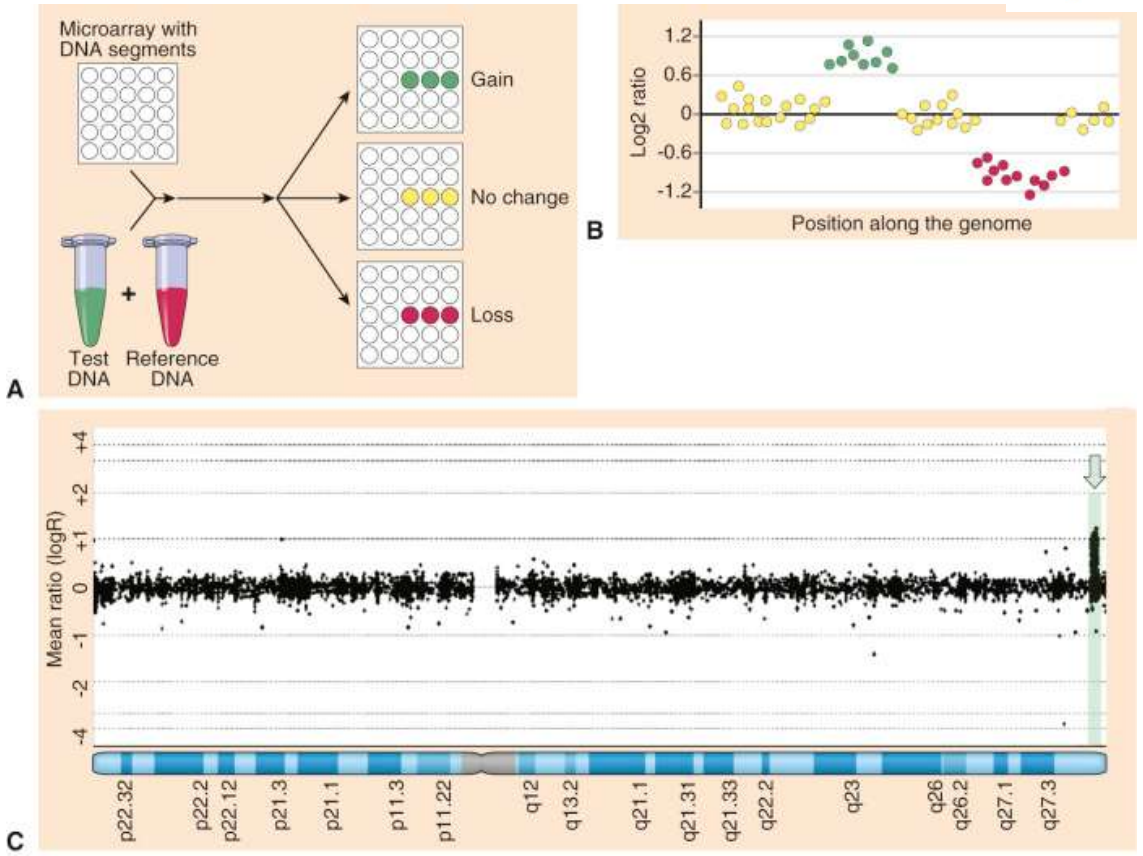


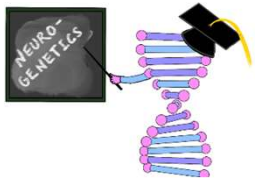
Figure 1 – FISH - metaphase plates showing (A) the expected signs in both chromosomes 22 (normal pattern) and (B) absence of the signal corresponding to region 11.2 of the long arm (q) of chromosome 22, consistent with a 22q11.2 microdeletion (the arrows indicate the chromosomes 22). In (A) note an interphase nucleus with two red signals (normal pattern)



Rosa et al. 2011

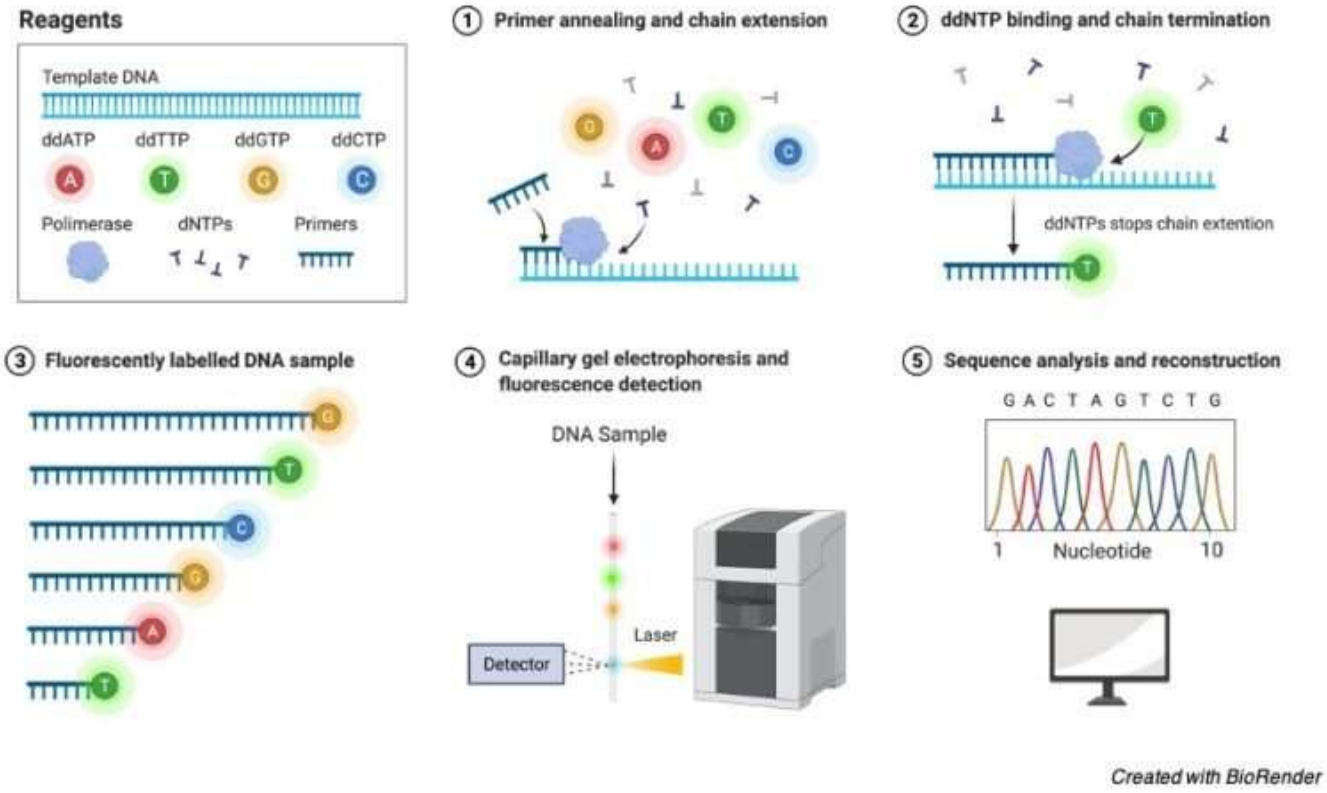
Genome Analysis Using Microarray

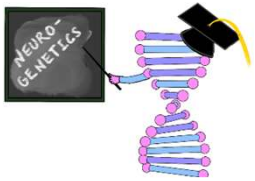




Principles of Clinical Cytogenetics and Genome Analysis - Thompson & Thompson Genetics in Medicine, 8th Edition (doctorlib.info)

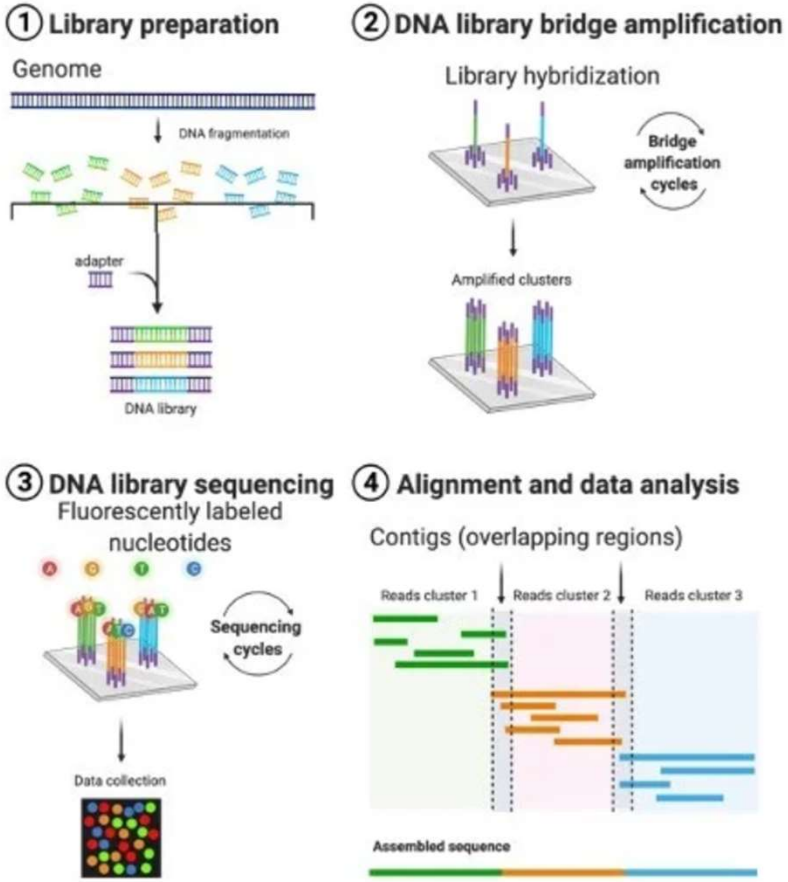
Sanger Sequencing

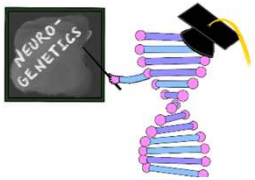




DNA Sequencing: Definition, Type, Mechanism and Significance (researchtweet.com)

Next Generation Sequencing

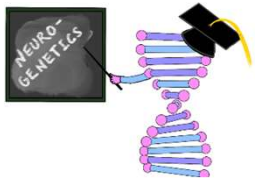




DNA Sequencing: Definition, Type, Mechanism and Significance (researchtweet.com)

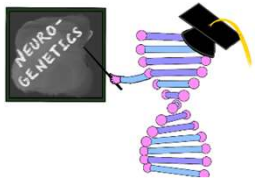
Next-Gen Sequencing

- NGS often won't identify:
  - Deletions/duplications >100 bp (newer NGS technology is better at this)
    - Specific deletion/duplication testing at a gene level available through MLPA or custom arrays
  - Trinucleotide repeats
    - PCR +Southern blot
  - Imprinting disorders
    - Methylation
  - Whether mutations are in cis or trans
    - Parental samples



Next-Gen Sequencing

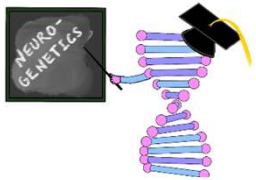
- These are the tests that use NGS:
  - Gene panels
  - Exome sequencing
  - Genome sequencing



Understanding the Report

Describe the type of genetic report and what the result is indicating:

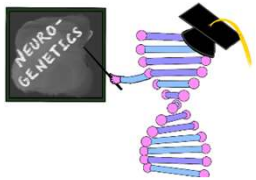
1. arr[hg19]2q33.1(198,356,789-203,491,035)X1
2. c.346G>C (p.G116R)
3. c.1945dupT (p.S649FfsX40)
4. c.847C>T (p.R283X)
5. c.487+1G>T



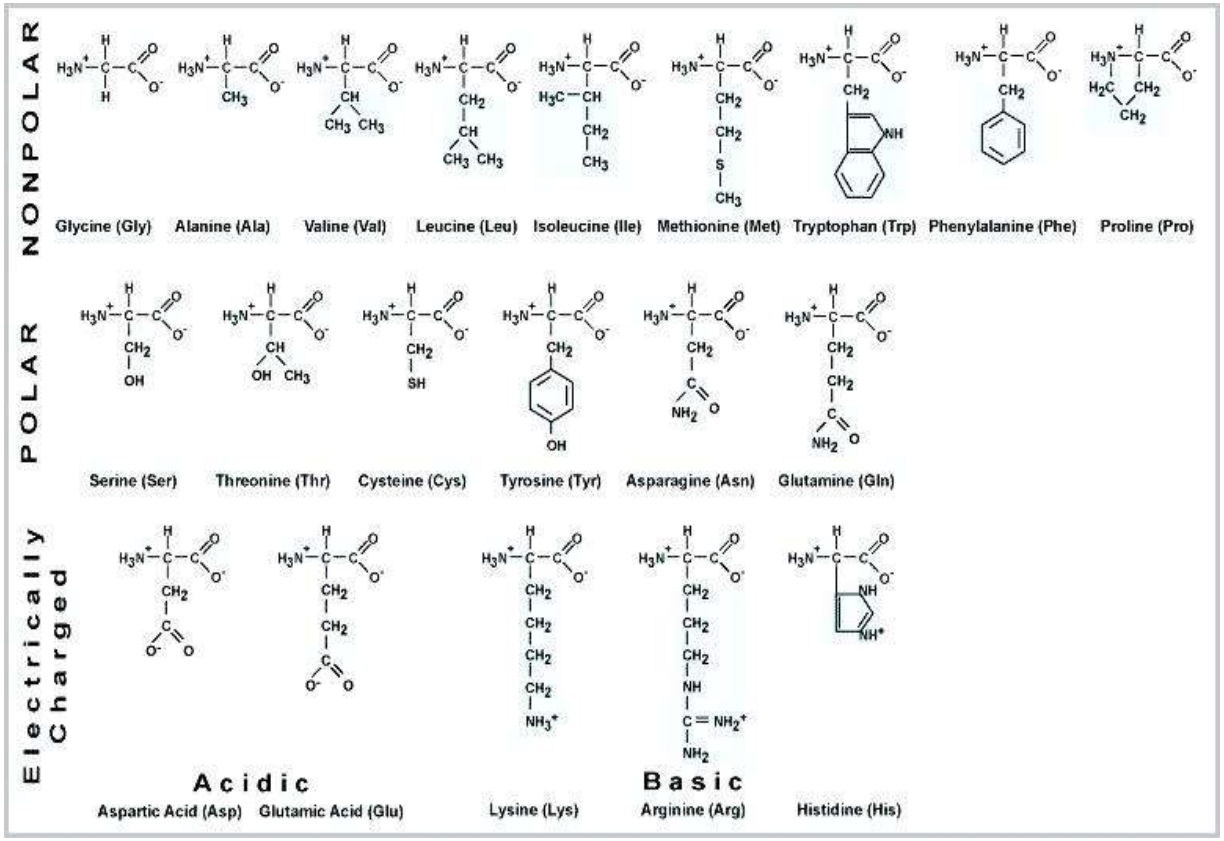
So, You Found a VUS…

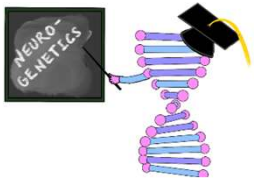
- For copy number variants:
  - What genes are included in the area?
  - How big is the area?
  - Is it inherited or de novo?
  - Is it a deletion or duplication?

- For single nucleotide changes:
  - What type of change is it?
  - Is an amino acid changed?
  - How similar is the amino acid?
  - Is it conserved?
  - Does it occur in an important part of the protein?
  - Is it found in population databases? (ExAC, 1000 genomes)
  - Has it been shown to segregate with disease?
  - Have there been any functional studies?



Physiochemical Properties of Amino Acids

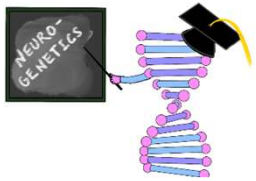




Histone H1 (residues 120-180)

```
HUMAN   KKASKPKKAASKAPTKKPKATPVKKAKKKLAATPKKAKKPKTVKAKPVKASKPKKAKPVK
MOUSE   KKAAKPKKAASKAPSKKPKATPVKKAKKKPAATPKKAKKPKVVKVKPVKASKPKKAKTVK
RAT     KKAAKPKKAASKAPSKKPKATPVKKAKKKPAATPKKAKKPKIVKVKPVKASKPKKAKPVK
COW     KKAAKPKKAASKAPSKKPKATPVKKAKKKPAATPKKTKKPKTVKAKPVKASKPKKTKPVK
CHIMP   KKASKPKKAASKAPTKKPKATPVKKAKKKLAATPKKAKKPKTVKAKPVKASKPKKAKPVK
        ***:**********:**************  ******:****  **.**********:*  **
```

NON-CONSERVED AMINO ACIDS: Conservative; Conservative; Non-conservative; Conservative; Non-conservative; Semi-conservative; Conservative; Non-conservative

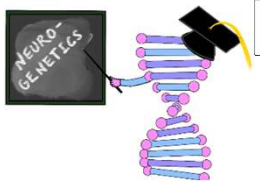


Sequence alignment - Wikipedia

ACMG 2015 Guidelines for Variant Classification

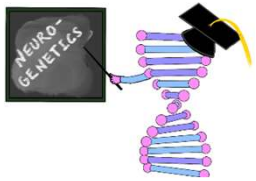
CHILD NEUROLOGY SOCIETY

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			



Suggested Reading

- Gayon J. From Mendel to epigenetics: History of genetics. *C R Biol*. 2016;339(7-8):225-230. doi:10.1016/j.crvi.2016.05.009
- Francis Crick, Rosalind Franklin, James Watson, and Maurice Wilkins | Science History Institute



Acknowledgements

Leads:
- Kuntal Sen (CNMC)
- Louis Dang (UM)

Core members:
- Amitha Ananth (UAB)
- Andrea Gropman (CNMC)
- Education
  - Rachel Gottlieb-Smith (UM)
  - Jeff Strelzik (CNMC)

Committee members:
- Daniel Calame (Baylor)
- Divakar Mithal (Northwestern)
- Christa Habela (Hopkins)
- Kristin Baranano (Hopkins)
- Lisa Emrick (Baylor)
- Margie Ream (Nationwide)
- Julie Ziobro (UM)

Additional Members:
- Alexa Taylor (CNMC)