My HSST research component C1 and C2

George Burghel Cohort 1 HSST

National School of Healthcare Science

MAHSE Manchester Academy for Healthcare Scientist Education

Manchester Centre for Genomic Medicine

Manchester 1824 The University of Manchester

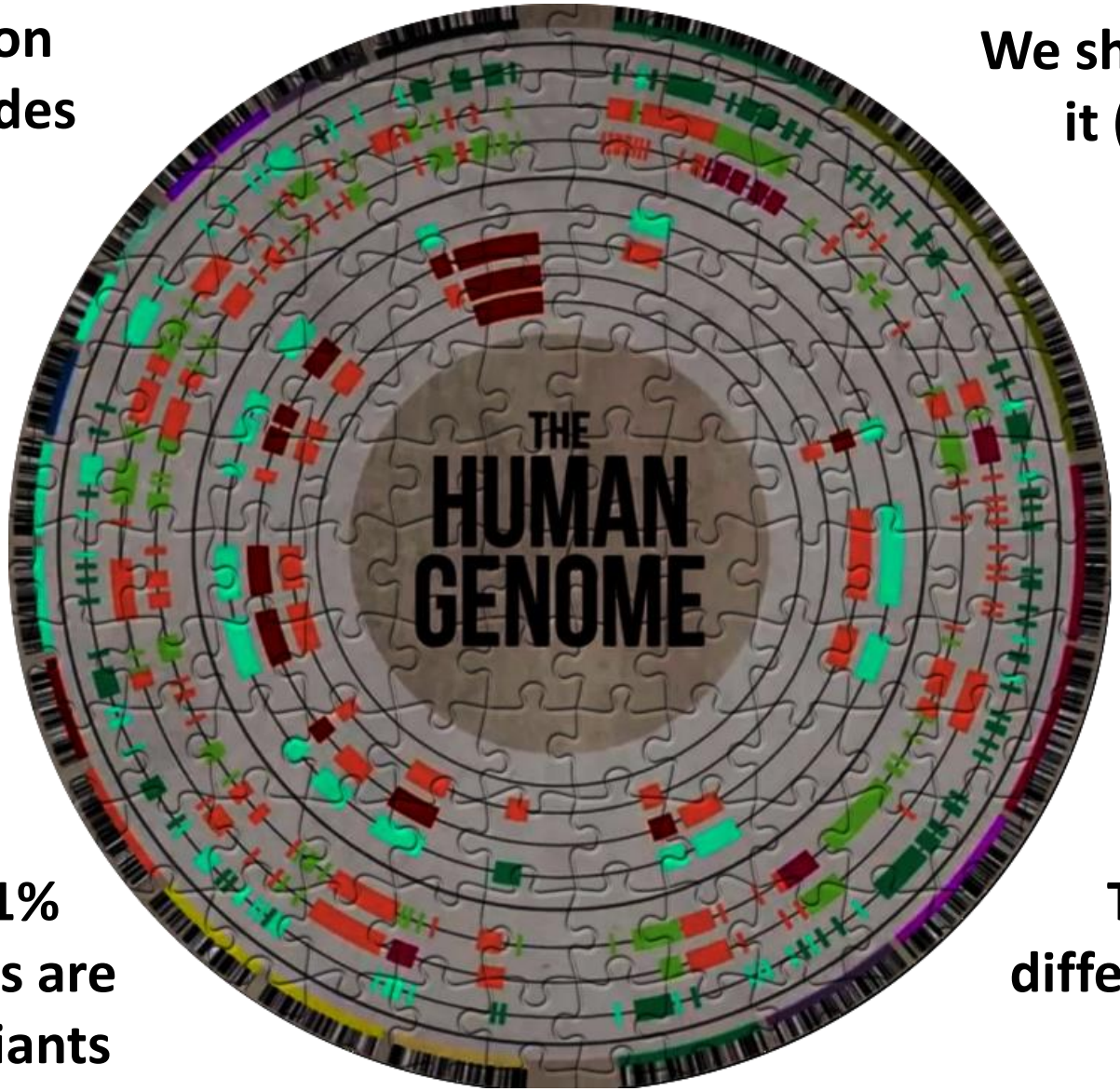
My research component

- C1 innovation proposal and literature review
- C1 presentation (Including Lay representative)
- C2 literature review (updated C1)
- C2 project
  - Copy number variants and socio economic status (manuscript nearly ready)
  - Innovative ways of interpreting copy number variants
    - Gene centric methods
    - Regulatory regions

The Human Genome

> 3 billion nucleotides

We share most of it (>99.9%)



The ~0.1% differences are called variants

There are different types of variants

DNA variants

GTGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCGCTGTGGCACTGC

GAGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCGCTGTGGCACTGC

GTGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCGCTGTGGCACTGC

GTGGCGCGAGCTTCTGAAACTA____GCAGAGGCGGAGCCGCTGTGGCACTGC

- We also have insertions, inversions and few more types
- 100,000s of SNVs and small scale variants (<50bp) in any human genome

DNA variants

GTGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCGCTGTGGCACTGC

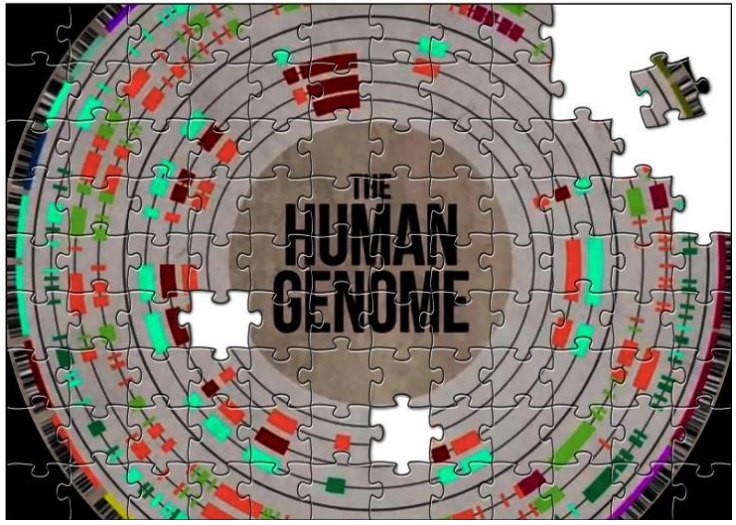
GAGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCGCTGTGGCACTGC

GTGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCGCTGTGGCACTGC

GTGGCGCGAGCTTCTGAAACTA____GCAGAGGCGGAGCCGCTGTGGCACTGC

- We also have insertions, inversions and few more types
- 100,000s of SNVs and small scale variants (<50bp) in any human genome

- Structural variants (SVs) are large type of DNA variation
- Copy Number Variants (CNVs) are the most common type of SVs



Copy number variants - CNVs

- >50bp – few Mbs
- Can be a **deletion** of DNA or a **gain**

GTGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCG
CTGTGGCACTGCTGCGCCTCTGCTGCGCCTCGGGTGTCTTT
TGCGGCGGTGGGTCGCCGCCGGGAGAAGCGTGAGGGGACAG
ATTTGTGACCGGCGCGGTTTTTGTCAGCTTACTCCGGCCAA
AAAAGAACTGCACCTCTGGAGCGGGTTAGTGGTGGTGGTAG
TGGGTTGGGACGAGCGCGTCTTCCGCAGTCCCAGTCCAGCG
TGGCGGGGGAGCGCCTCACGCCCCGGGTCGCTGCCGCGGCT
TCTTGCCCTTTTGTCTCTGCCAACCCCCACCCATGCCTGAG
AGAAAGGTCCTTGCCCGAAGGCAGATTTTCGCCAAGCAAAT
TCGAGCCCCGCCCCTTCCCTGGGTCTCCATTTCCCGCCTCC
GGCCCGGCCTTTGGGCTCCGCCTTCAGCTCAAGACTTAACT
TCCCTCCCAGCTGTCCCAGATGACGCCATCTGAAATTTCTT
GGAAACACGATCACTTTAACGGAATATTGCTGTTTTGGGGA
AGTGTTTTACAGCTGCTGGGCACGCTGTATTTGCCTTACTT
AAGCCCCTGGTAATTGCTGTATTCCGAAGACATGCTGATGG
GAATTACCAGGCGGCGTTGGTCTCTAACTGGAGCCCTCTGT
CCCCACTAGCCACGCGTCACTGGTTAGCGTGATTGAAACTA
AATCGTATGAAAATCCTCTTCTCTAGTCGCACTAGCCACGT
TTCGAGTGCTTAATGTGGCTAGTGGCACCGGTTTGGACAGC
ACAGCTGTAAAATGTTCCCATCCTCACAGTAAGCTGTTACC
GTTCCAGGAGATGGGACTGAATTAGAATTCAAACAAATTTT
CCAGCGCTTCTGAGTTTTACCTCAGTCACATAATAAGGAAT
GCATCCCTGTGTAAGTGCATTTTGGTCTTCTGTTTTGCAGA
CTTATTTACCAAGCATTGGAGGAATATCGTAGGTAAAAATG
CCTATTGGATCCAAAGAGAGGCCAACATTTTTTGAAATTTT
TAAGACACGCTGCAACAAAGCAGGTATTGACAAATTTTATA
TAACTTTATAAATTACACCGAGAAAGTGTTTTCTAAAAAA

GTGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCG
CTGTGGCACTGCTGCGCCTCTGCTGCGCCTCGGGTGTCTTT
TGCGGCGGTGGGTCGCCGCCGGGAGAAGCGTGAGGGGACAG
ATTTGTGACCGGCGCGGTTTTTGTCAGCTTACTCCGGCCAA
AAAAGAACTGCACCTCTGGAGCGGGTTAGTGGTGGTGGTAG
TGGGTTGGGACGAGCGCGTCTTCCGCAGTCCCAGTCCAGCG
TGGCGGGGGAGCGCCTCACGCCCCGGGTCGCTGCCGCGGCT
TCTTGCCCTTTTGTCTCTGCCAACCCCCACCCATGCCTGAG

Deletion

AATCGTATGAAAATCCTCTTCTCTAGTCGCACTAGCCACGT
TTCGAGTGCTTAATGTGGCTAGTGGCACCGGTTTGGACAGC
ACAGCTGTAAAATGTTCCCATCCTCACAGTAAGCTGTTACC
GTTCCAGGAGATGGGACTGAATTAGAATTCAAACAAATTTT
CCAGCGCTTCTGAGTTTTACCTCAGTCACATAATAAGGAAT
GCATCCCTGTGTAAGTGCATTTTGGTCTTCTGTTTTGCAGA
CTTATTTACCAAGCATTGGAGGAATATCGTAGGTAAAAATG
CCTATTGGATCCAAAGAGAGGCCAACATTTTTTGAAATTTT
TAAGACACGCTGCAACAAAGCAGGTATTGACAAATTTTATA
TAACTTTATAAATTACACCGAGAAAGTGTTTTCTAAAAAA

GTGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAGCCG
CTGTGGCACTGCTGCGCCTCTGCTGCGCCTCGGGTGTCTTT
TGCGGCGGTGGGTCGCCGCCGGGAGAAGCGTGAGGGGACAG
ATTTGTGACCGGCGCGGTTTTTGTCAGCTTACTCCGGCCAA
AAAAGAACTGCACCTCTGGAGCGGGTTAGTGGTGGTGGTAG
TGGGTTGGGACGAGCGCGTCTTCCGCAGTCCCAGTCCAGCG
TGGCGGGGGAGCGCCTCACGCCCCGGGTCGCTGCCGCGGCT
TCTTGCCCTTTTGTCTCTGCCAACCCCCACCCATGCCTGAG
AGAAAGGTCCTTGCCCGAAGGCAGATTTTCGCCAAGCAAAT
TCGAGCCCCGCCCCTTCCCTGGGTCTCCATTTCCCGCCTCC
GGCCCGGCCTTTGGGCTCCGCCTTCAGCTCAAGACTTAACT
TCCCTCCCAGCTGTCCCAGATGACGCCATCTGAAATTTCTT
GGAAACACGATCACTTTAACGGAATATTGCTGTTTTGGGGA
AGTGTTTTACAGCTGCTGGGCACGCTGTATTTGCCTTACTT
AAGCCCCTGGTAATTGCTGTATTCCGAAGACATGCTGATGG
GAATTACCAGGCGGCGTTGGTCTCTAACTGGAGCCCTCTGT
CCCCACTAGCCACGCGTCACTGGTTAGCGTGATTGAAACTA
AATCGTATGAAAATCCTCTTCTCTAGTCGCACTAGCCACGT
TTCGAGTGCTTAATGTGGCTAGTGGCACCGGTTTGGACAGC
ACAGCTGTAAAATGTTCCCATCCTCACAGTAAGCTGTTACC
GTTCCAGGAGATGGGACTGAATTAGAATTCAAACAAATTTT
CCAGCGCTTCTGAGTTTTACCTCAGTCACATAATAAGGAAT
GCATCCCTGTGTAAGTGCATTTTGGTCTTCTGTTTTGCAGA
CTTATTTACCAAGCATTGGAGGAATATCGTAGGTAAAAATG
CCTATTGGATCCAAAGAGAGGCCAACATTTTTTGAAATTTT
TAAGACACGCTGCAACAAAGCAGGTATTGACAAATTTTATA
TAACTTTATAAATTACACCGAGAAAGTGTTTTCTAAAAAA

CNVs and Human disease

- CNVs can be benign
- CNVs cause and contribute to a variety of diseases
  - intellectual disability
  - developmental delays
  - autism spectrum disorders
  - epilepsy and multiple congenital abnormalities

- Many methods to detect them

CNVs and Human disease

- The clinical significance of CNVs is classified into one of five categories:

Parameter	Benign (Class 1)	Likely benign (Class 2)	Variants of uncertain significance (Class 3)	Likely pathogenic (Class 4)	Pathogenic (Class 5)
Size	← Tend to be smaller		Varies	Tend to be larger →	
Gene(tic) content	← Few genes or none		Some with genes Some in non-coding regions	Several genes, disease genes →	
Reported on databases?	← Frequently seen		Not common	Listed on Disease databases →	
Internal databases	← Frequently seen		Not common	Not seen/reported in other patients →	
Published literature			Varies	Published →	

CNVs and Human disease

- The clinical significance of CNVs is classified into one of five categories:

Parameter	Benign (Class 1)	Likely benign (Class 2)	Variants of uncertain significance (Class 3)	Likely pathogenic (Class 4)	Pathogenic (Class 5)
Size	Tend to be smaller		Varies	Tend to be larger	
Gene(tic) content	Few genes or none		Some with genes Some in non-coding regions	Several genes, disease genes	
Reported on databases?	Frequently seen		Not common	Listed on Disease databases	
Internal databases	Frequently seen		Not common	Not seen/reported in other patients	
Published literature			Varies	Published	

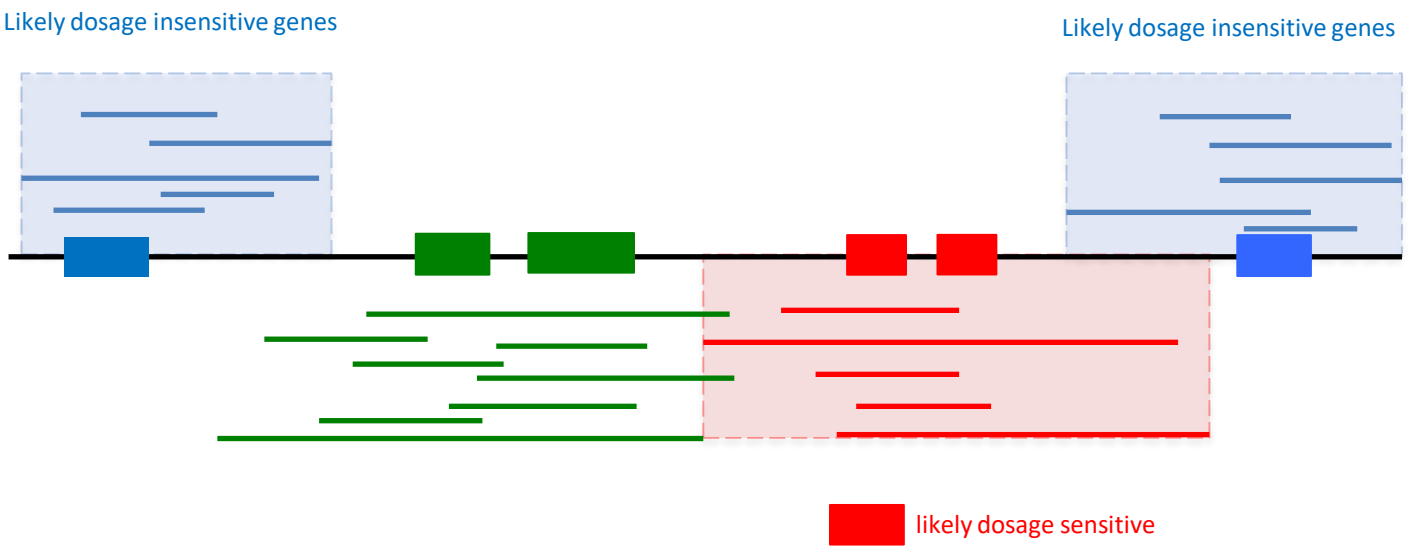
more than 40% of identified CNVs end up classified as variants of unknown significance – class 3!

Aims of the project

- Could CNVs impact SES?
- Could we improve CNV interpretation?

Analysis of dosage sensitivity

Could we improve CNV interpretation?

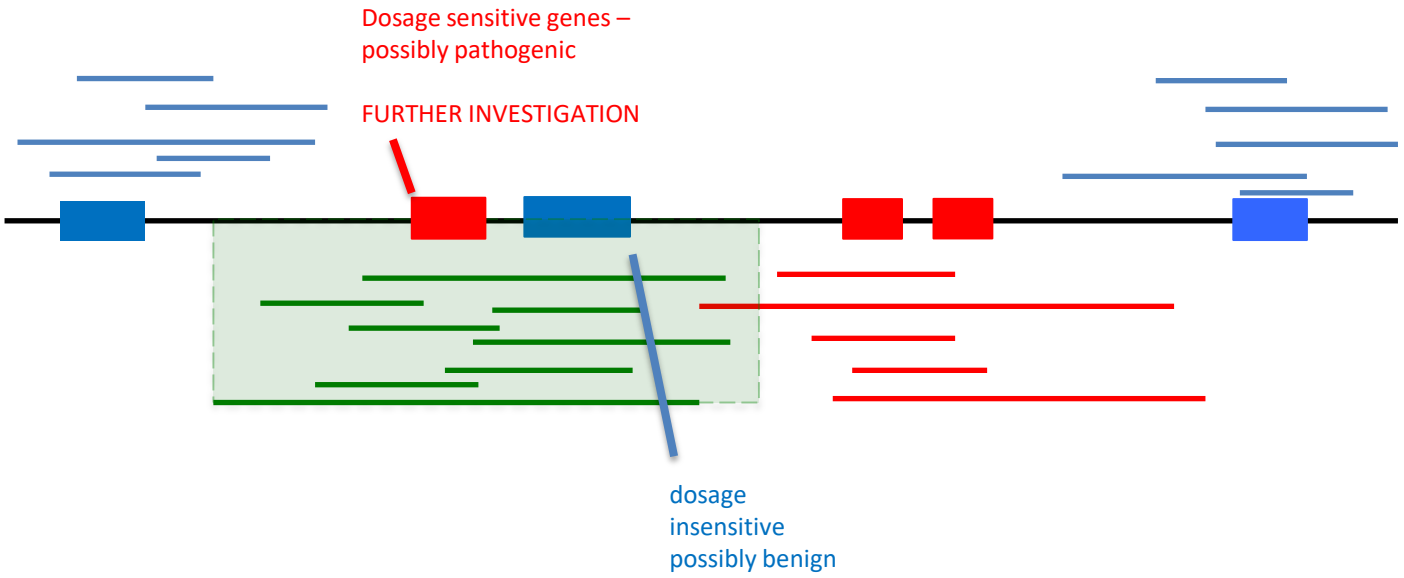


Pathogenic CNV

CNV of uncertain significance

Benign CNV

Analysis of dosage sensitivity

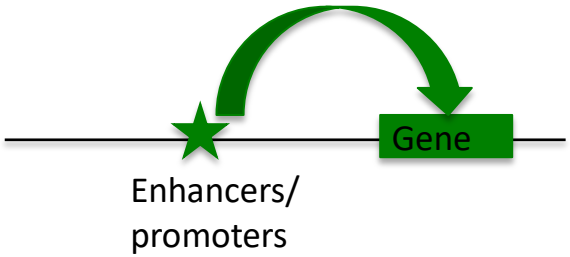


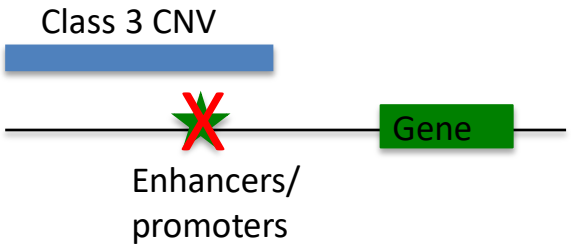
Pathogenic CNV

CNV of uncertain significance

Benign CNV

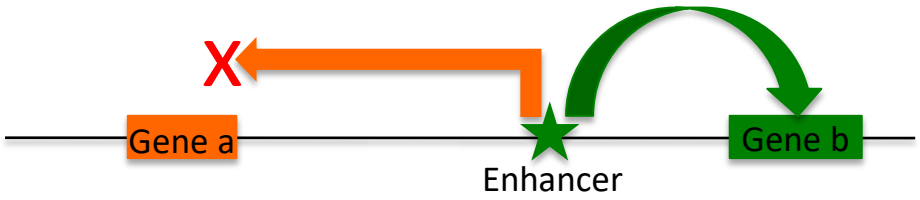
Analysis of regulatory elements



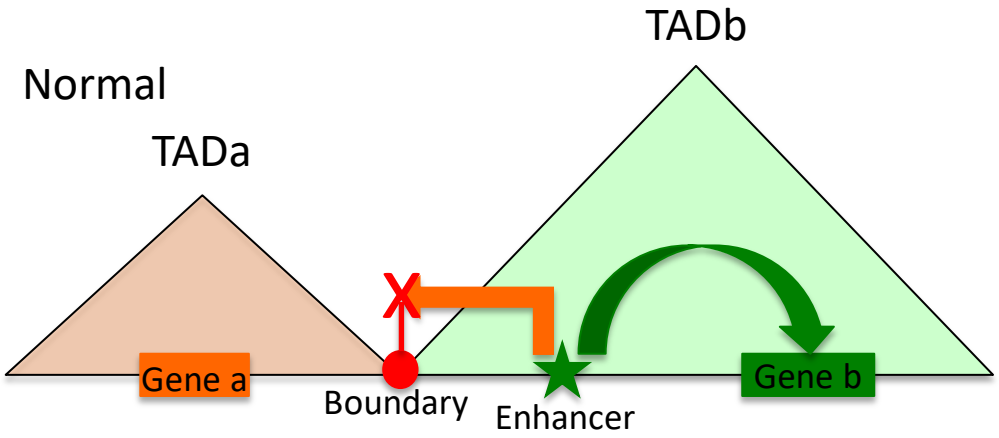


Topologically associated domains

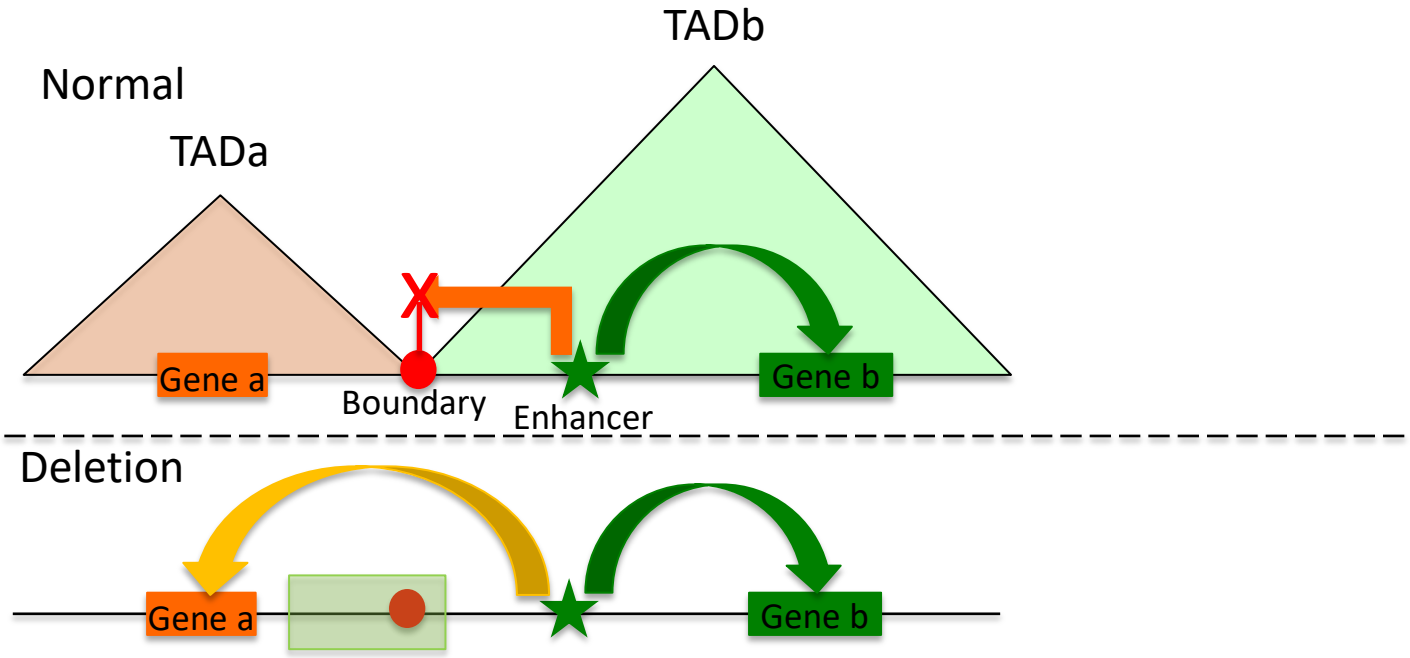
TADs are thought to guide regulatory elements to their associated genes



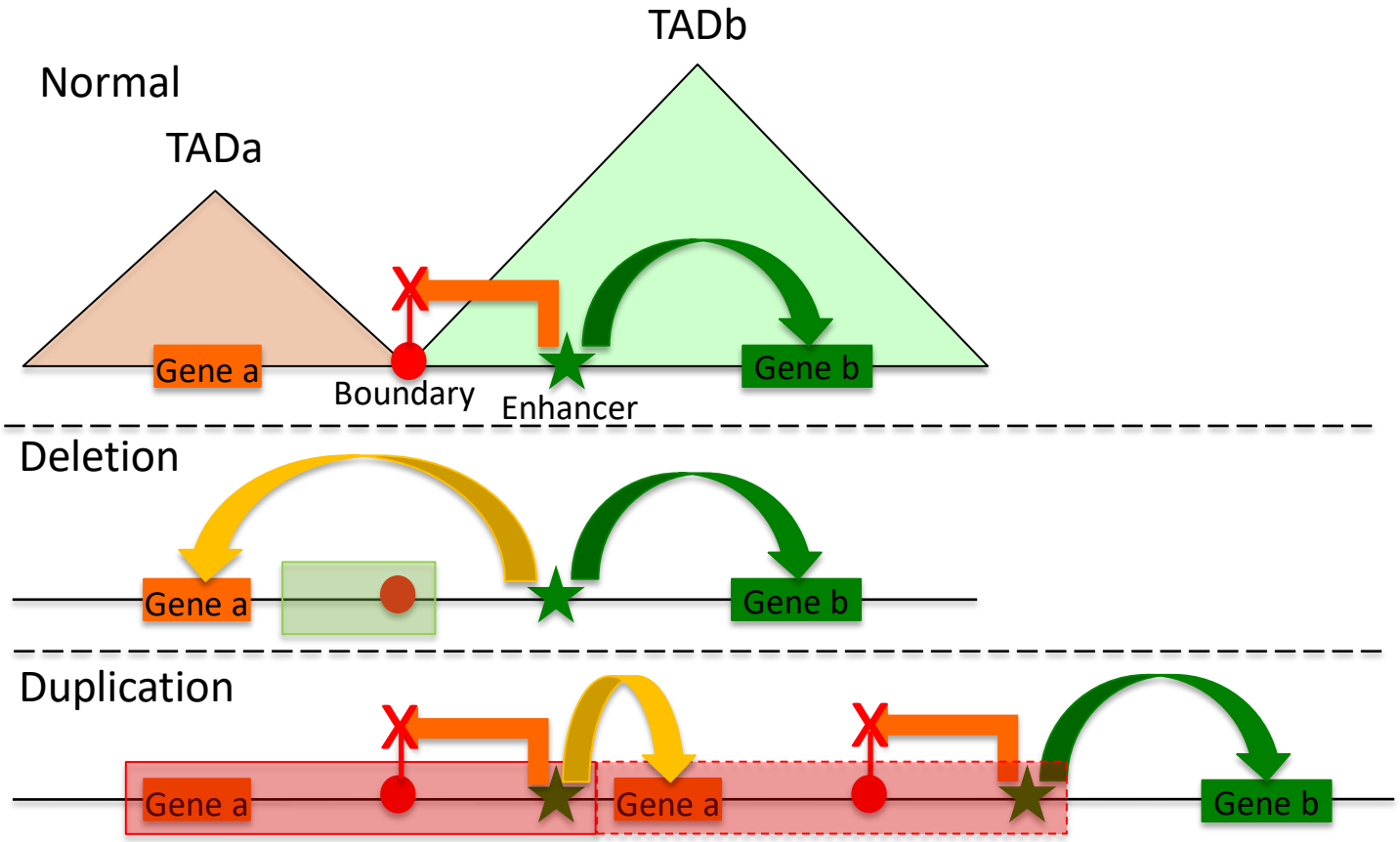
Topologically associated domains



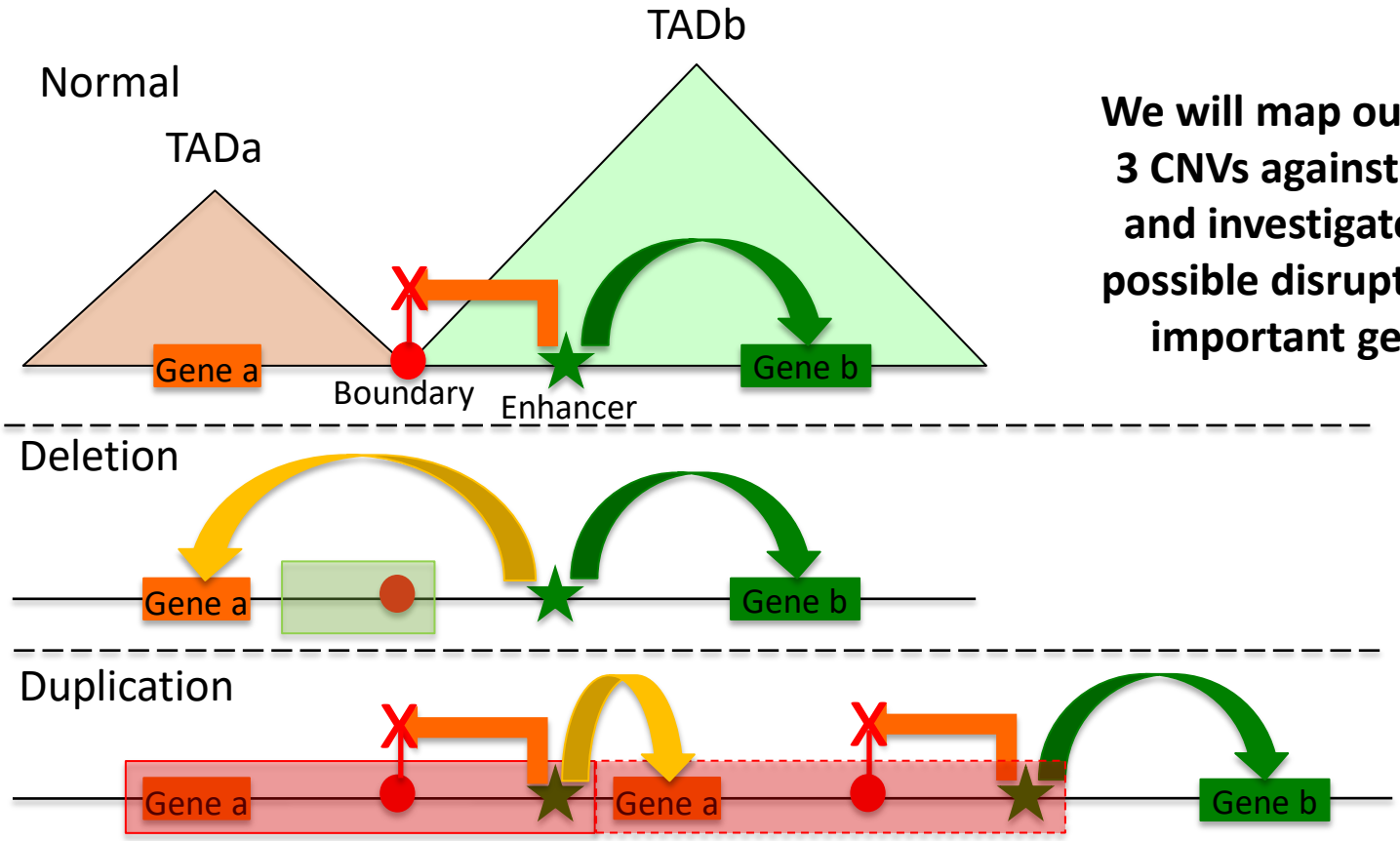
Topologically associated domains



Topologically associated domains



Topologically associated domains



We will map our class 3 CNVs against TADs and investigate any possible disruption of important genes

The good and the bad

- **The good**
  - Engaged supervisor
  - Data based – main cost is my time
  - Engaged bioinformatician
  - Made part of my project as an MSc project

- **The bad**
  - Time time time…
  - Parts of my initial plans will not be achieved (over ambitious)

- **Advice**
  - Be realistic
  - Plan lots
  - Start early
  - Be flexible, plans and project will change and evolve