Variants

position and possible consequences

everything what can go wrong, will go wrong

tinyurl.com/ Avans201904b

Johan den Dunnen

based on lecture Jan Traeger-Synodinos
VEPTC Prague 2017

Human & Clinical Genetics

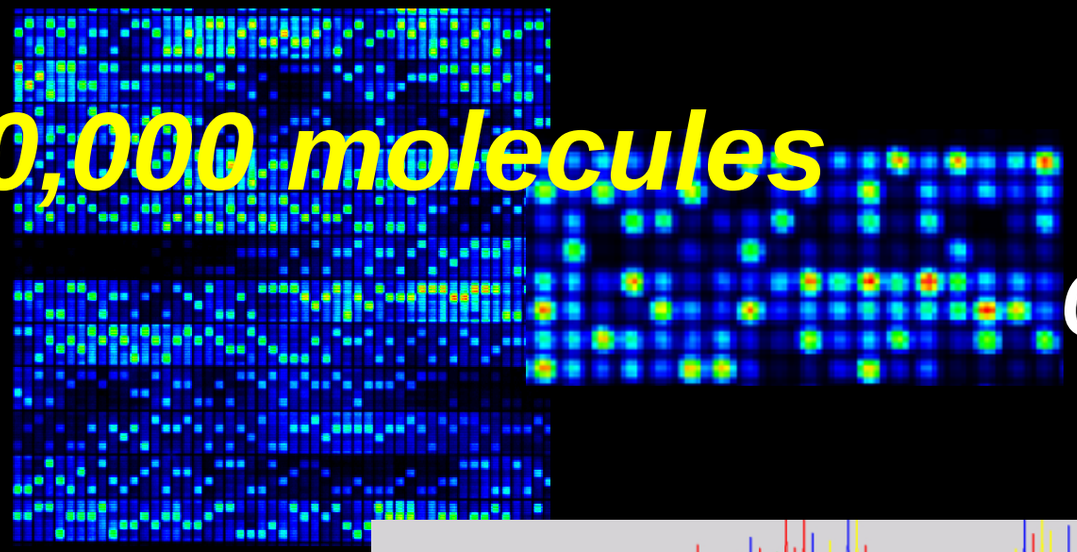
(Leiden University Medical Center)

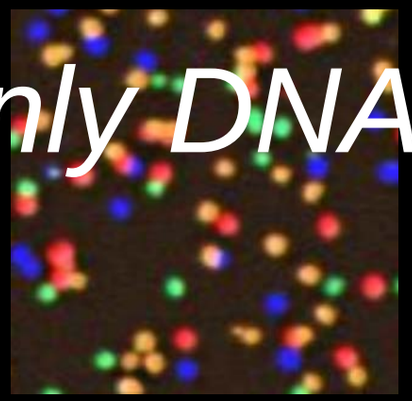
- **Genetic Disease**
  - ***neuromuscular disorders***
  - ***http:// www.DMD.nl***
  - ***diagnosis***
  - ***treatment / therapy***
  - ***exon skipping DMD***

- **Genome Technology**
  - ***try and apply***
  - ***facilitate***
  - ***Leiden Genome Technology Center***
  - ***http:// www.LGTC.nl /***

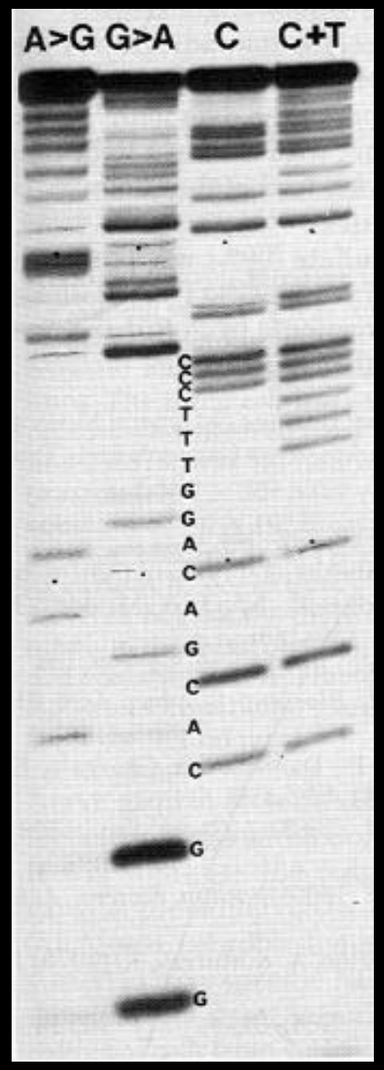
My CV

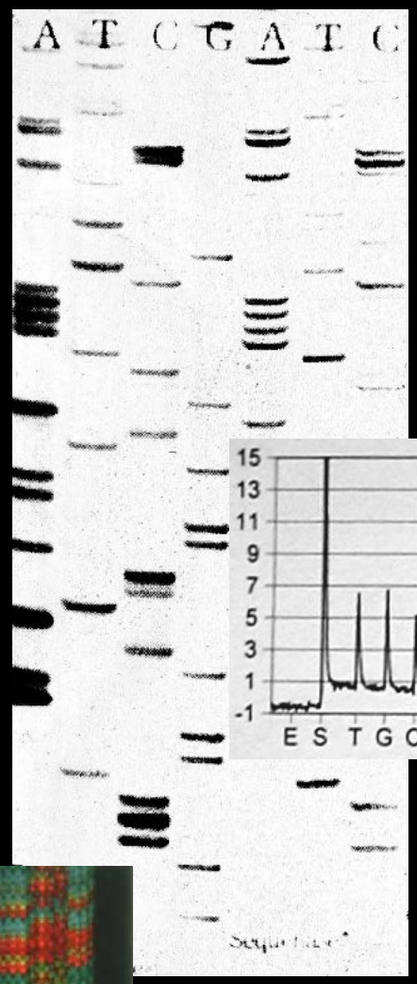
1,000,000,000 molecules

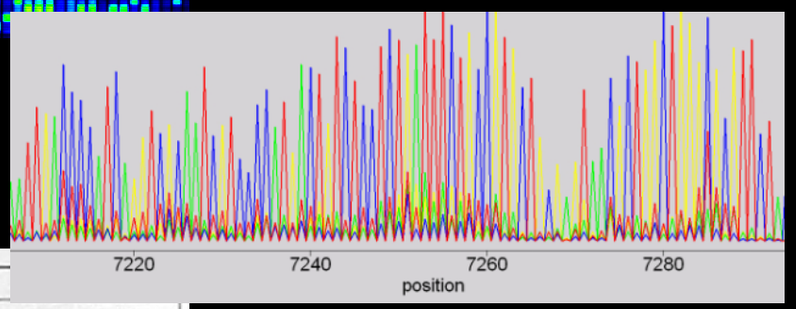


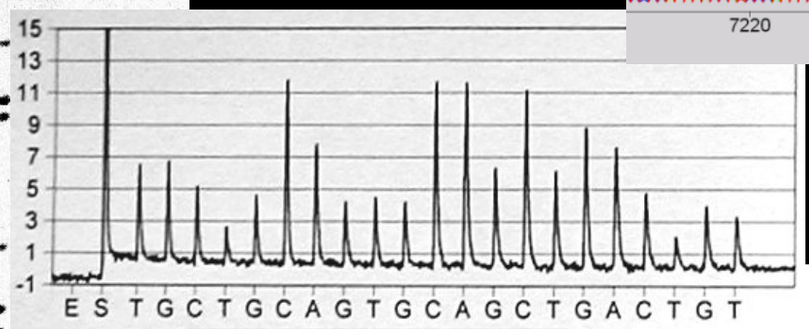


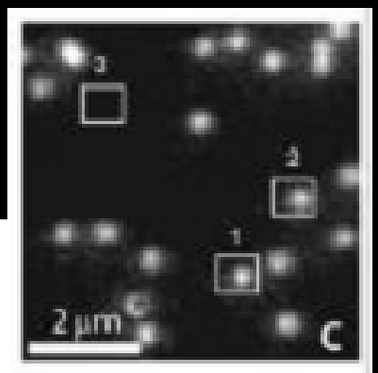
only DNA

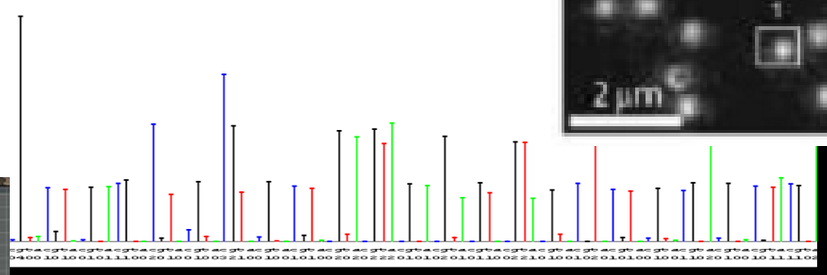


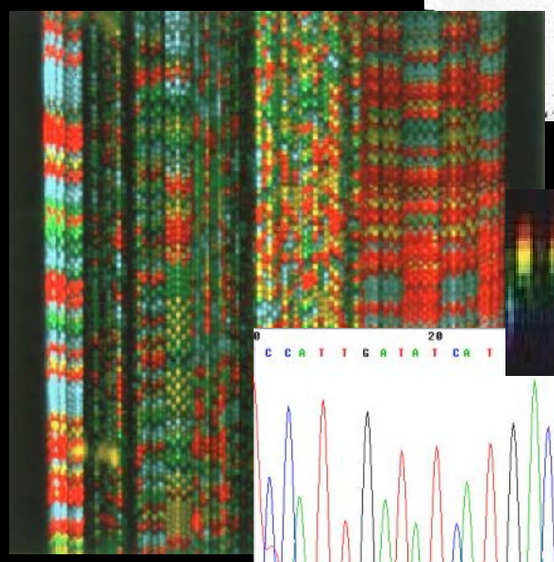




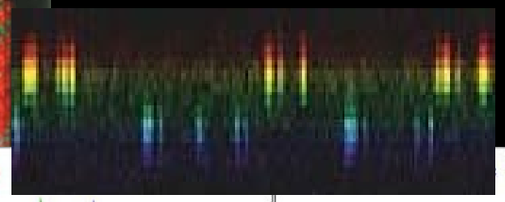


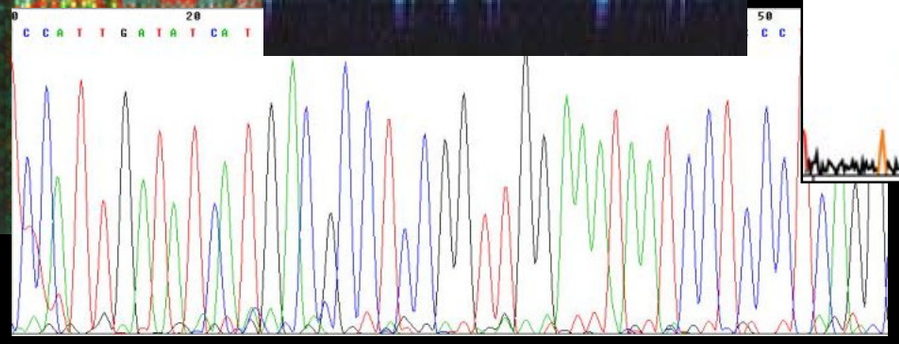


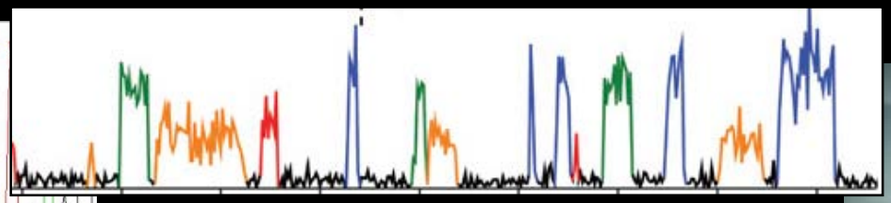


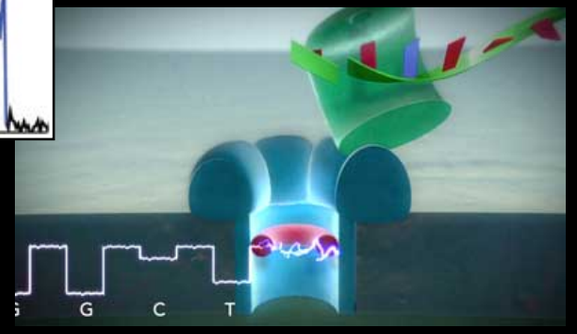


1 molecule





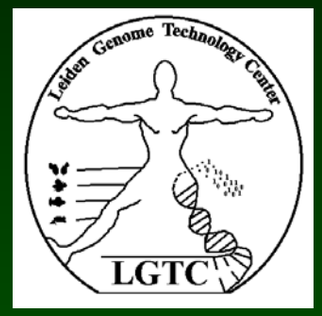




Subjects

Variants

- basic types



All possibilities & examples

- DNA
- RNA
- protein

everything what can go wrong, will go wrong

Use your imagination

which variants did you find / can you think of?

Variant < > phenotype

- **combination should make sense**

  gene function should "explain" phenotype
  gene expressed in affected tissue
  for analysis use affected tissue
  RNA, protein analysis

- **do not give up**

  check options you can not exclude
  did you detect ALL possible variants
  did you consider all variant to problem options

Terminology

- **prevent confusion**
  - ***mutation meaning, …***
    - ***-biology: change***
    - ***-medical: disease-causing change***
  - ***polymorphism, meaning …***
    - ***-biology: change in >1% population***
    - ***-medical: not disease causing change***

- **use neutral terms**
  - ***sequence variant***
  - ***alteration***
  - ***CNV*** ***(Copy Number Variant)***
  - ***SNV*** ***(not SNP)***

would you like to be a mutant?

Variant types ?

Variant types ?

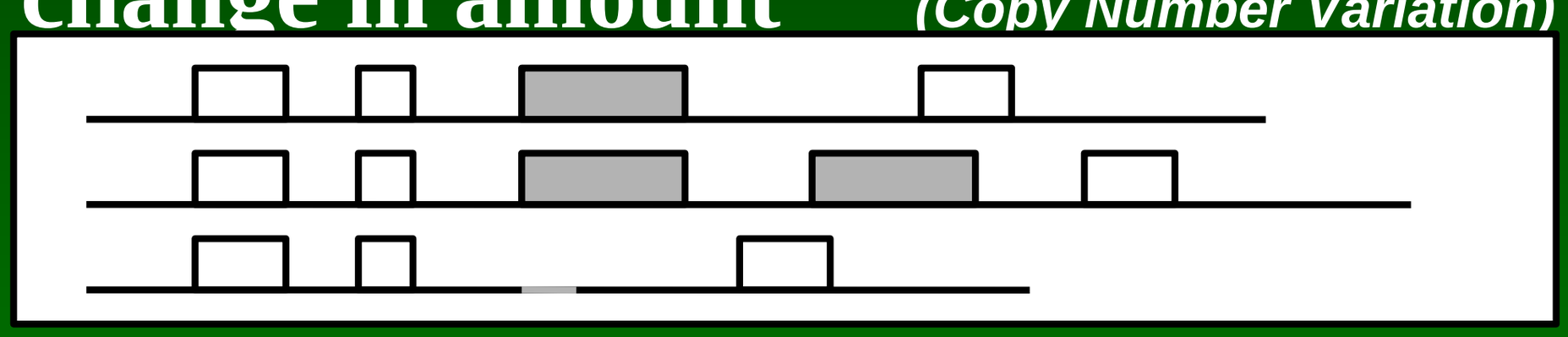
- error in sequence
- too much / too little
- wrong position
  place, time

Variant types

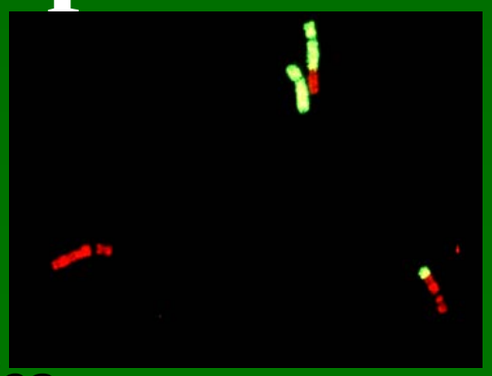
- **change in sequence**

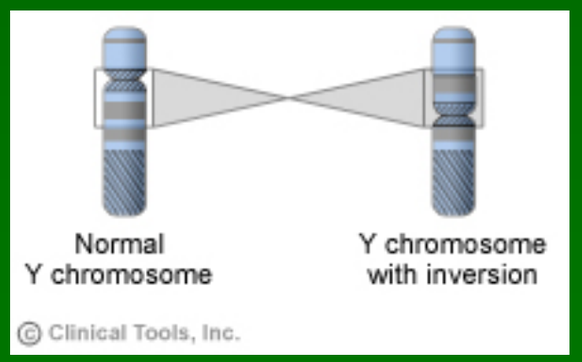
```
ACATCAGGAGAAGATGTTC  GAGACTTTGCCA
ACATCAGGAGAAGATGTTT  GAGACTTTGCCA
ACATCAGGAGAAGATGTT   GAGACTTTGCCA
ACATCAGGAGAAGATGTTCCGAGACTTTGCCA
```

- **change in amount** *(Copy Number Variation)*



- **change in position**





Structural Variation (SV)

Variants

- **small changes**

  ...

- **larger changes**

  ...

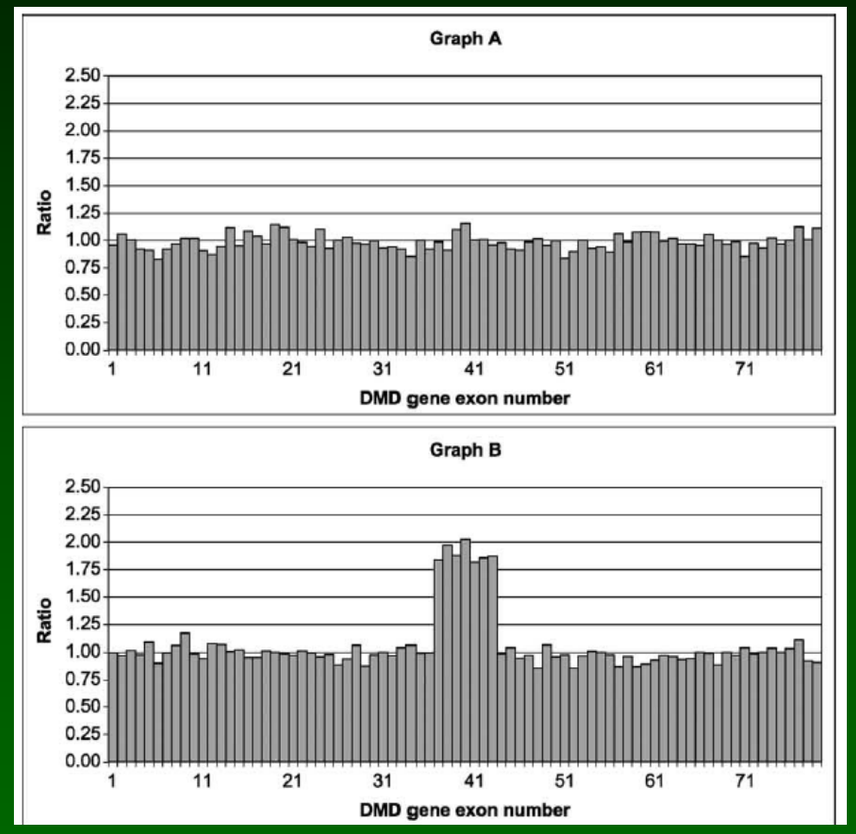
Variants

- **small changes**
  - ***substitution***
  - ***few nucleotide changes***
    - ***deletion, duplication, insertion, "indel"***
- **larger changes**
  - ***structural variants (SV)***
    - ***translocation***
    - ***insertion***
    - ***transposition***
    - ***copy number variant (CNV)***
      - ***deletion***
      - ***duplication (insertion)***
    - ***inversion***
    - ***conversion***
    - ***complex***

protein coding variant
"easy" to explain, others
often difficult to proof

MLPA result

- **result of an MLPA**
  what did the assay detect ?



the assay detects an extra copy
where the extra copy is remains to be determined, could be anywhere in the genome

some genes are dosage sensitive, some are not
deletions are more likely to be deleterious then duplications
*determining the exact break point is **critical** (incl. RNA analysis)*
*the **amount** of product can also be deleterious*

Origin of variants

- **during cell cycle**
  DNA replication *(copying information)*
  meiosis / mitosis *(dividing information)*
  different rates in female / male
    large rearrangements / SNVs

- **damage**
  environmental
    radiation, chemicals (smoke), UV, …
    ~100,000 per cell daily

consequence?

Germline (inherited)	Somatic (acquired)
present from birth	*acquired during life*
present in every cell	*limited to certain cells*
can be transmitted	*non transmitted*

De novo variants

- **genome sequencing**
  78 mother-father-child trios

- **de novo variants**
  ~60 small scale changes
  ~15 maternal *stable over time*
  ~45 paternal *variable with age*

- **effect paternal age**
  20y > 25 variants
  40y > 65 variants
  2 extra variants / year

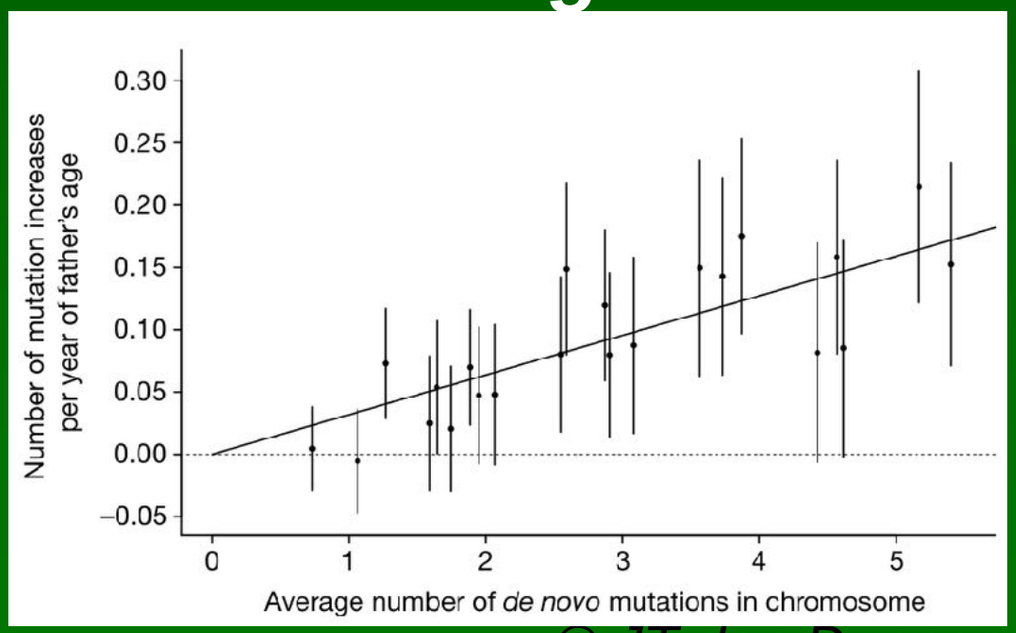
ARTICLE

23 AUGUST 2012 | VOL 488 | NATURE 471

Rate of *de novo* mutations and the importance of father's age to disease risk

Augustine Kong[1], Michael L. Frigge[1], Gisli Masson[1], Soren Besenbacher[1,2], Patrick Sulem[1], Gisli Magnusson[1], Sigurjon A. Gudjonsson[1], Asgeir Sigurdsson[1], Aslaug Jonasdottir[1], Adalbjorg Jonasdottir[1], Wendy S. W. Wong[3], Gunnar Sigurdsson[1], G. Bragi Walters[1], Stacy Steinberg[1], Hannes Helgason[1], Gudmar Thorleifsson[1], Daniel F. Gudbjartsson[1], Agnar Helgason[1,4], Olafur Th. Magnusson[1], Unnur Thorsteinsdottir[1,5] & Kari Stefansson[1,5]

Kong et al. 2012 Nature 488: 471

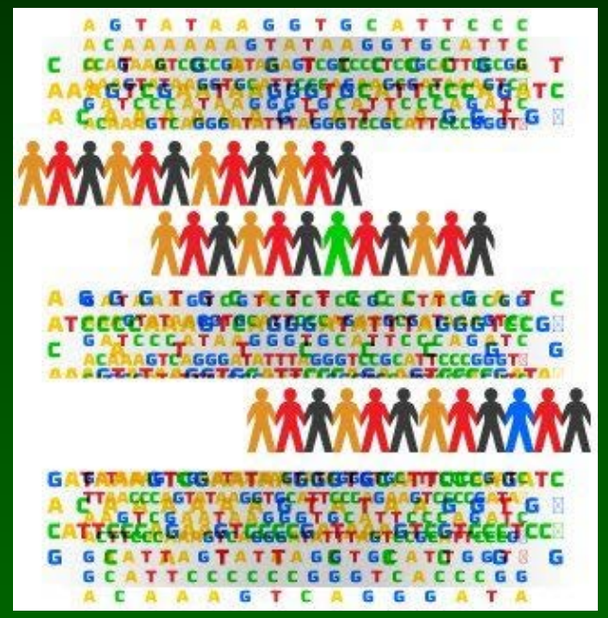


How many

based on 1000 Genomes Project Consortium

- **4.1-5.0 million small changes**

  SNVs, short indels



- **2,100-2,500 other variants**

  1,000 large deletions
  160 CNVs
  915 Alu / 128 L1 insertions
  51 composite SINE/VNTR/Alu insertions
  4 nuclear mitochondrial DNA variants
  10 inversions

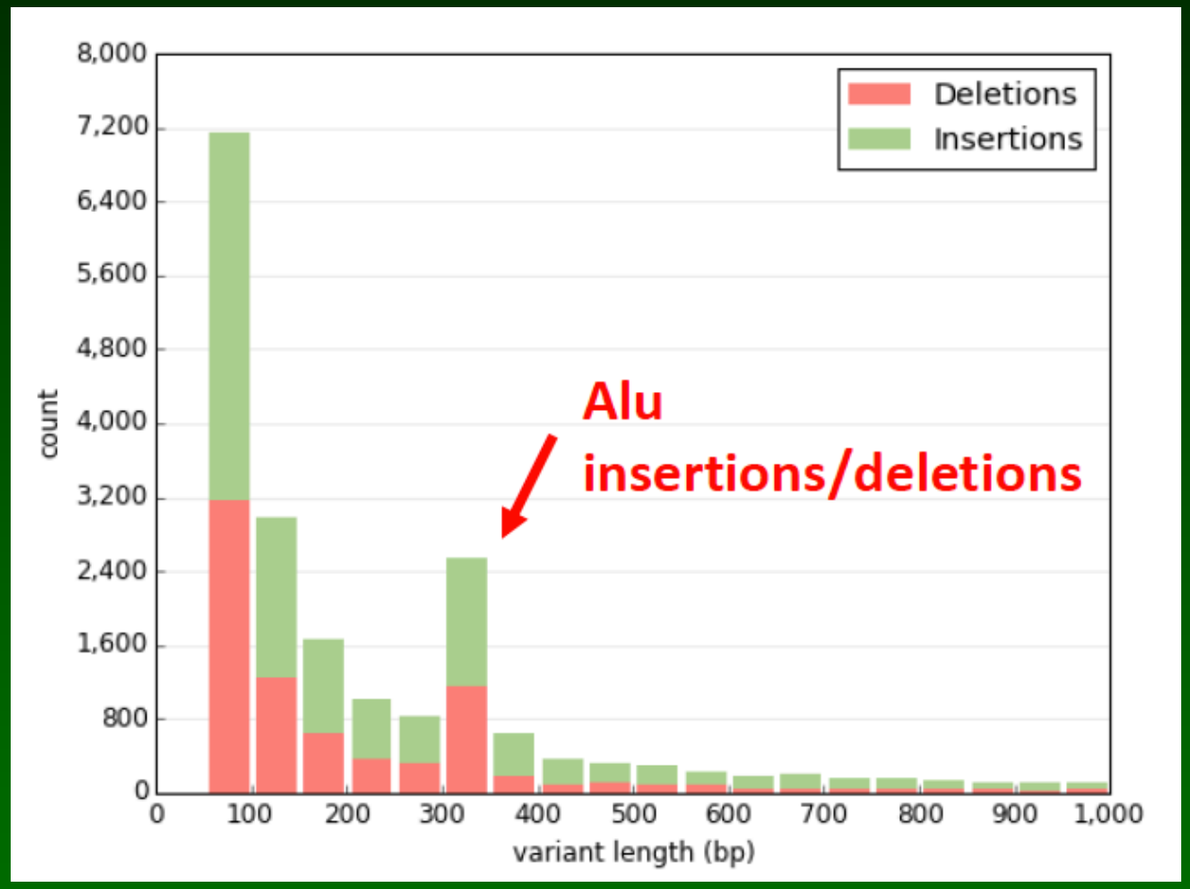
  (includes upto 30 Mb's of sequence present in you not your neighbour)

Long-read sequencing

http://www.pacb.com/wp-content/uploads/2017-EMEA-UGM-A-Hoischen-Long-read-Sequencing.pdf

A. Hoischen (Nijmegen)

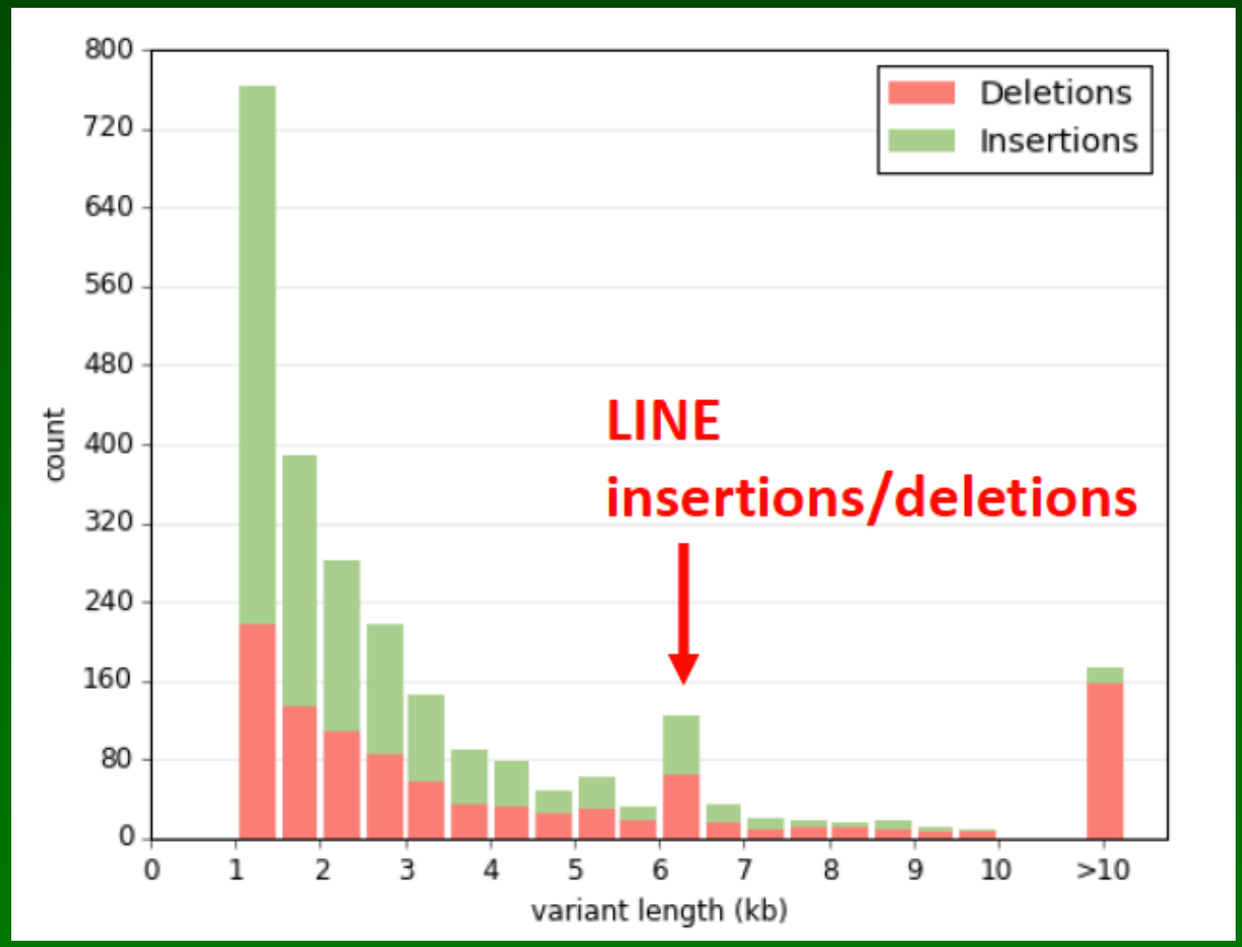
Radboudumc



small SVs

~25,000 SVs per genome

large SVs



Modern approach

sequence, everything!

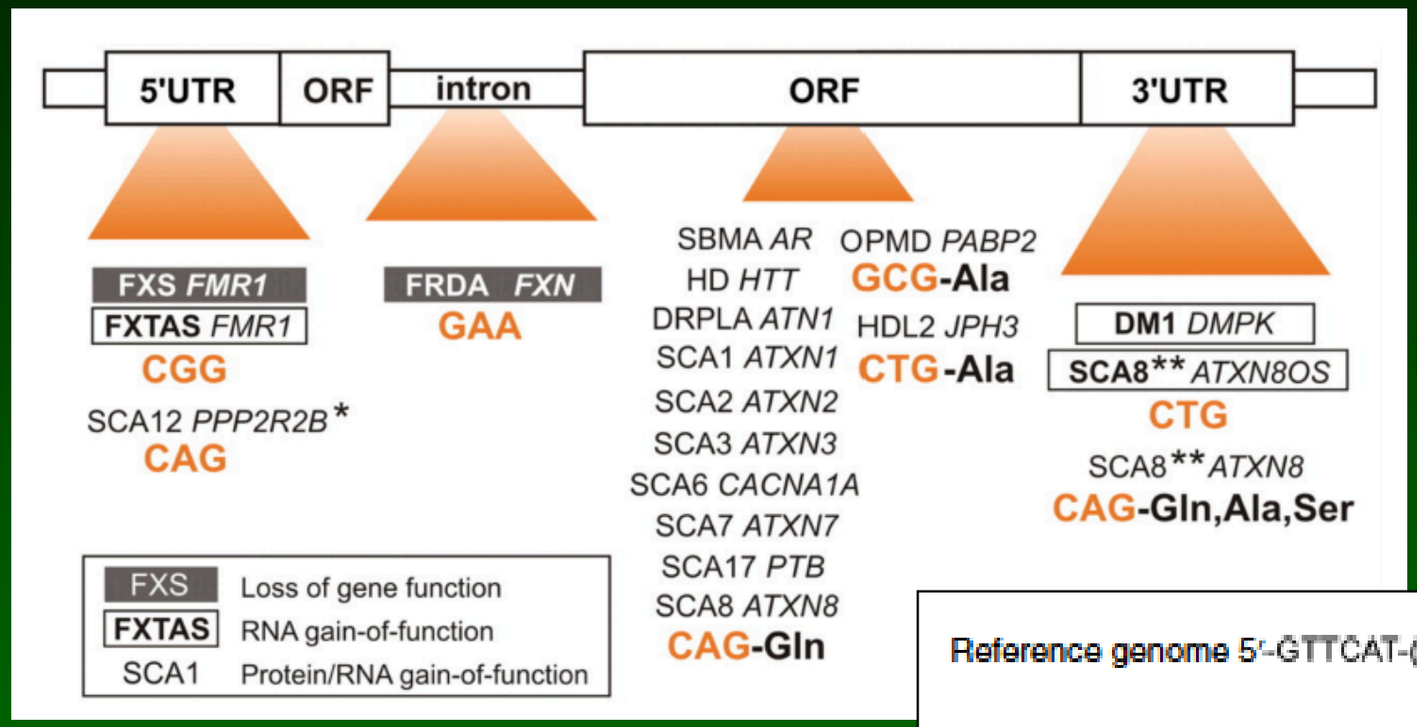
NGS: one fits all ?

- NGS fails to detect…

NGS: one fits all ?

- **targeted assays remain valuable**
  - ***clear cause: cost effective direct assay***
  - ***why sequence everything?***
    - ***protein coding variants "easy", others difficult to proof***
- **NGS fails to detect**
  - ***most structural variants (SV)***
    - ***incl. deletions / duplications***
  - ***variants in repeated/variable length sequences***
    - ***MUC genes***
  - ***variants in highly variable regions***
    - ***HLA***
  - ***repeat expansions***
    - ***trinucleotide expansions***
  - ***rearrangements covering more then read length***

Repeat expansions



unmapped
too many mismatches

Reference genome 5'-GTTCAT-$(TTTTA)_7$TTA$(TTTTA)_{13}$-TTTGA

F6906 (II-6) 5'-GTTCAT-$(TTTTA)_2$TTTTTTA$(TTTTA)_{104}$TTTA$(TTTTA)_3$TTTA$(TTTTA)_{10}$$(TTTCA)_{66}$-TTGA

F6115 (I-2) 5'-GTTCAT-$(TTTTA)_3$TTTTTTTA$(TTTTA)_{60}$ $(TTTCA)_{61}$$(TTTTA)_{73}$TTTATTTTA-TTTGA

nature genetics
ARTICLES
https://doi.org/10.1038/s41588-018-0067-2

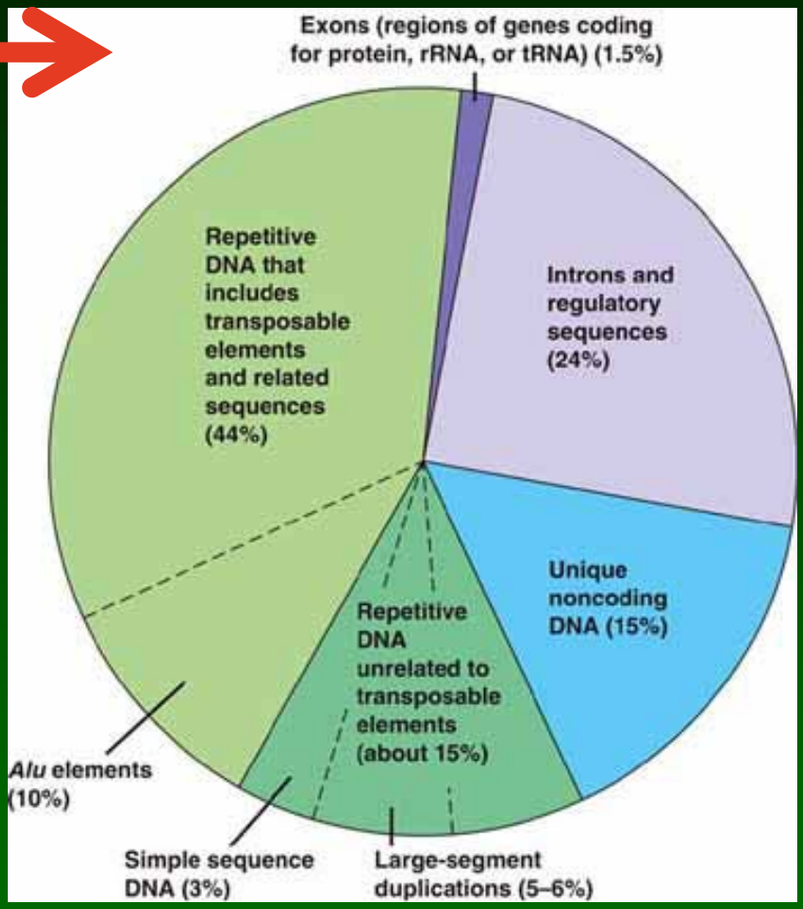
Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy

Hiroyuki Ishiura[1], Koichiro Doi[2], Jun Mitsui[1], Jun Yoshimura[2], Miho Kawabe Matsukawa[1], Asao Fujiyama[3], Yasuko Toyoshima[4], Akiyoshi Kakita[4], Hitoshi Takahashi[4], Yutaka Suzuki[5], Sumio Sugano[6], Wei Qu[2], Kazuki Ichikawa[2], Hideaki Yurino[7], Koichiro Higasa[8], Shota Shibata[1]

Human genome

version 27 (Jan.2017, freeze hg38)
- Ensembl 90

GENCODE



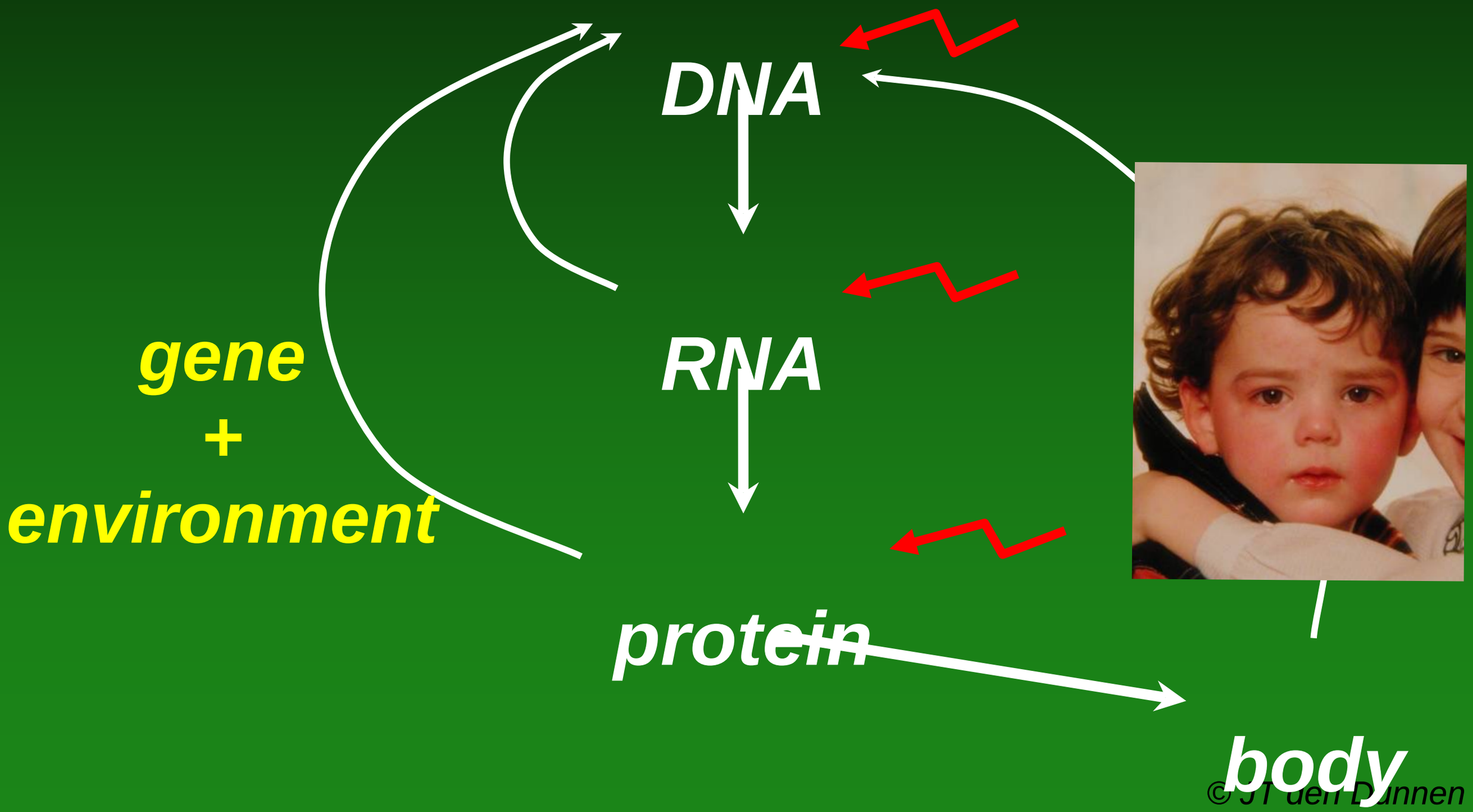
- **58,288 genes**

19,836 protein coding
15,788 long non-coding RNA
7,269 small non-coding RNA
14,694 pseudogenes
644 immunoglobulin / T-cell receptor gene segments

5-10% genes is present in multiple copies

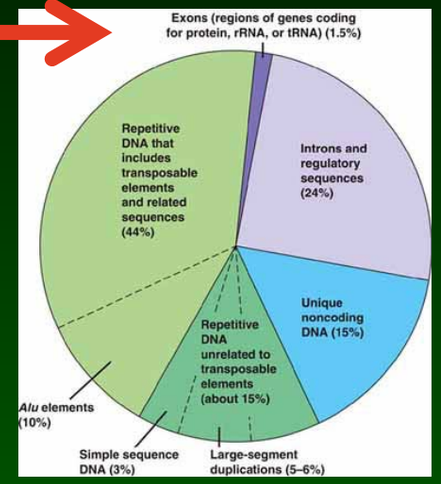
not all genes are critical

Information



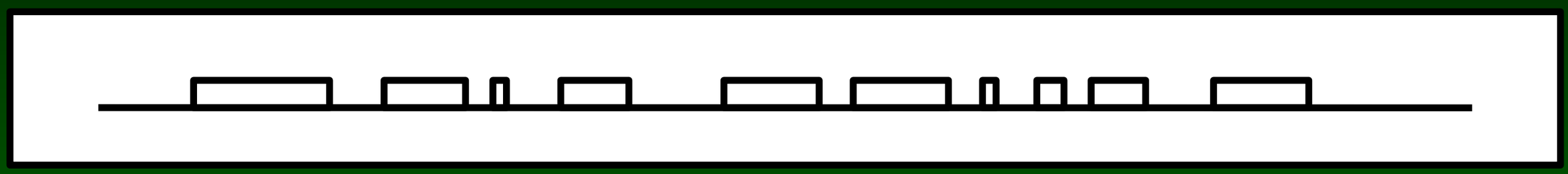
Translation variants

- **silent**
  no amino acid change
- **substitution**
  amino acids changes to another amino acid
- **nonsense**
  amino acids changes to stop codon
- **frame shift**
  translation shifts to another reading frame change
- **in-frame**
  deletion, duplication, insertion
- **other**
  extension (upstream initiation, no stop), ATG codon, ...



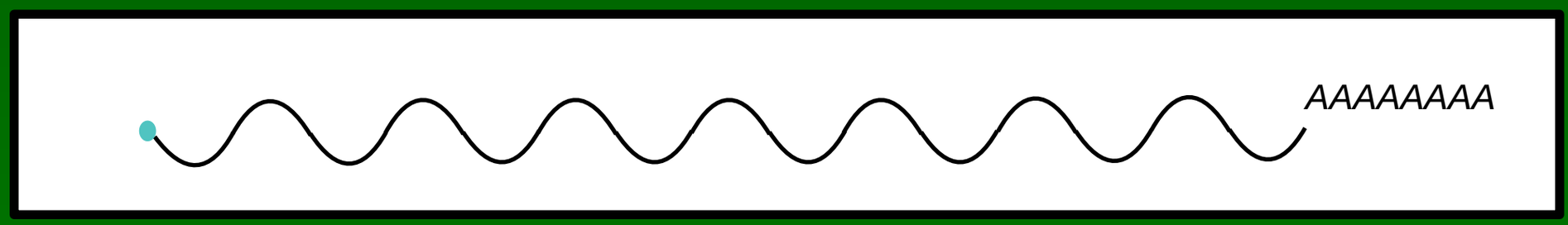
Transcription

a gene



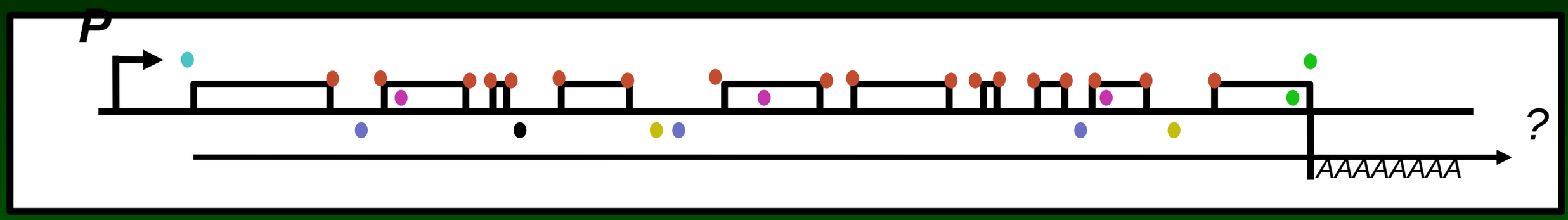
required ?

a transcript



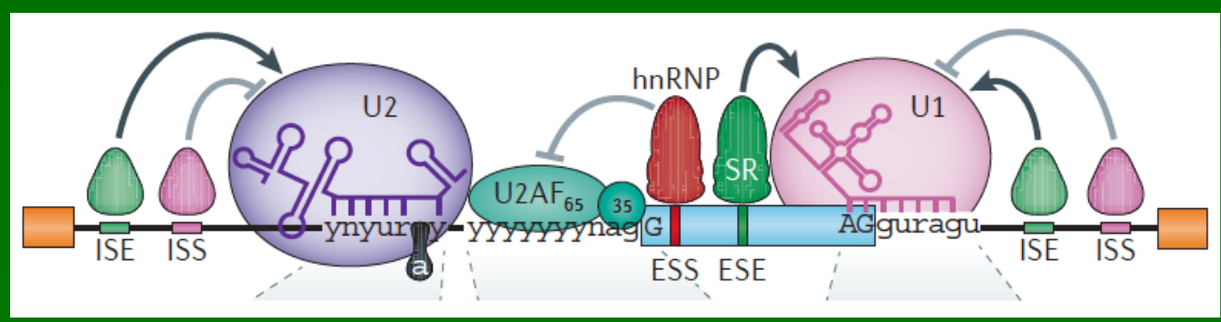
Transcription

a gene



promoter

- *transcription initiation site (cap site)*
- *transcription termination*
- *polyA-addition site, polyA-addition signal*
- *splice donor / splice acceptor site*
- *exonic splice enhancer / silencer (ESE / ESS)*
- *intronic splice enhancer / silencer (ISE / ISS)*
- *…*

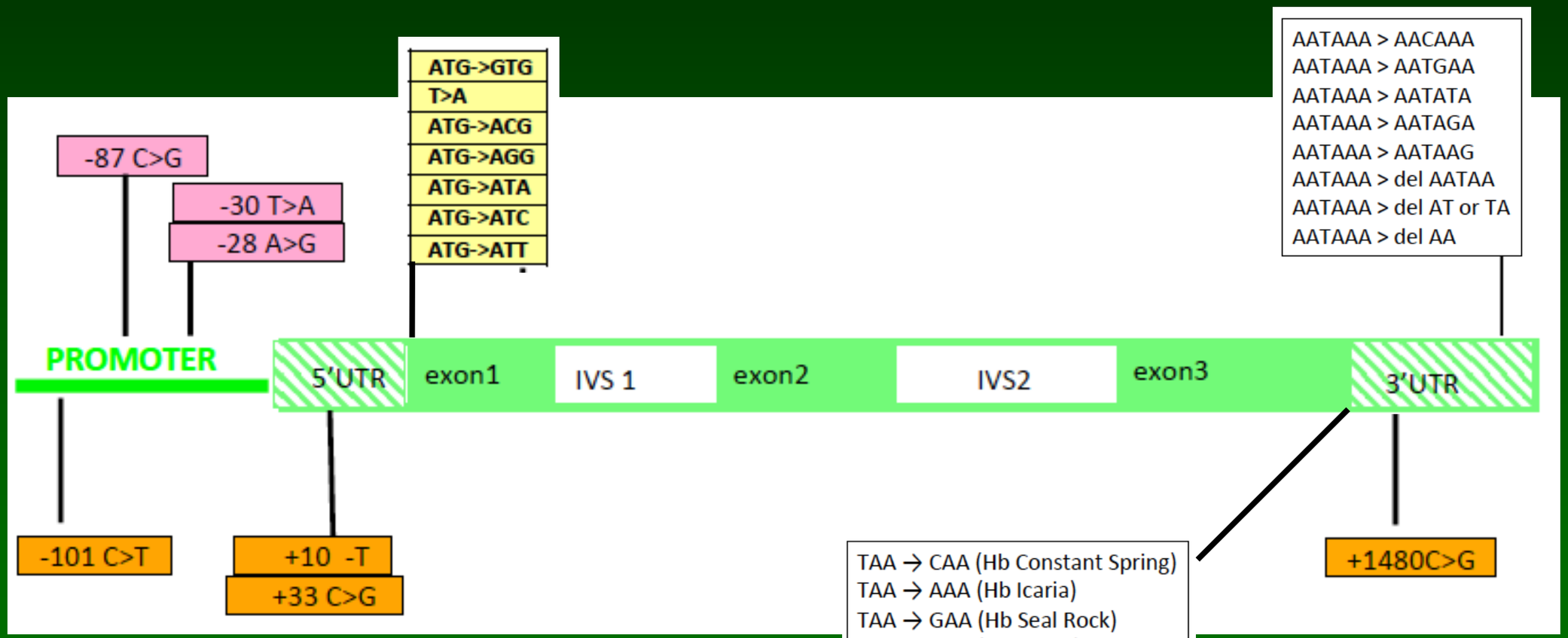


Scotti & Swanson, Nat.Rev.Genet. 2016

Consequences: RNA

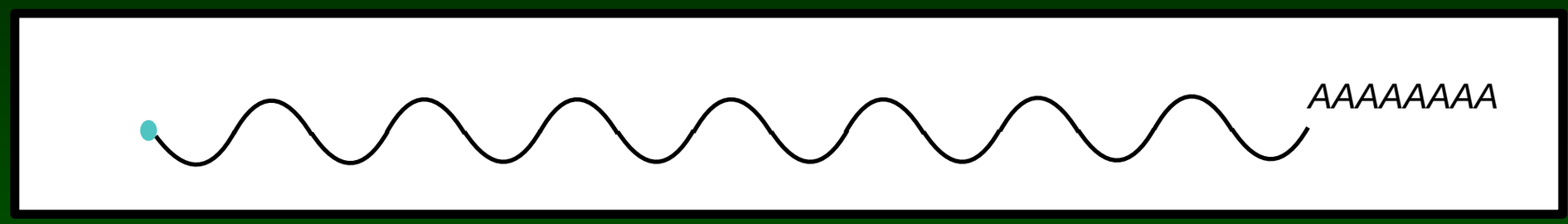
globin genes

(see HbVar for details)



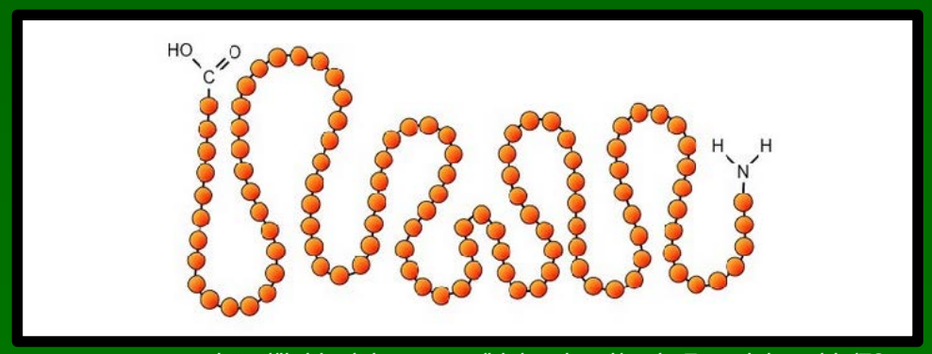
Translation

mRNA



required ?

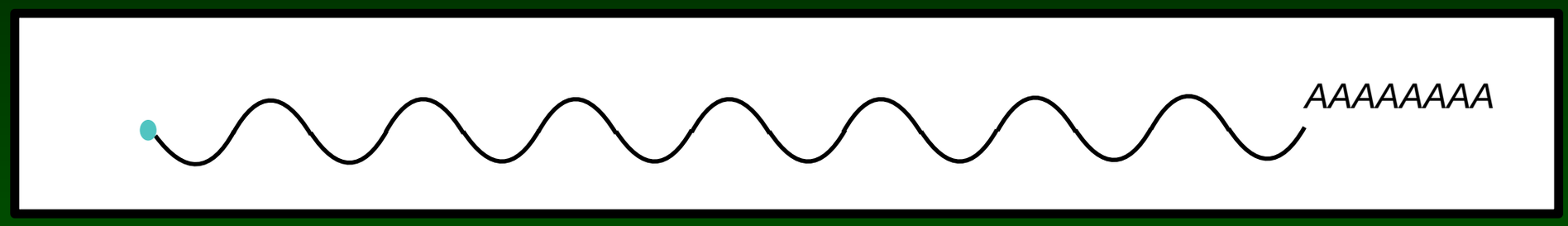
protein

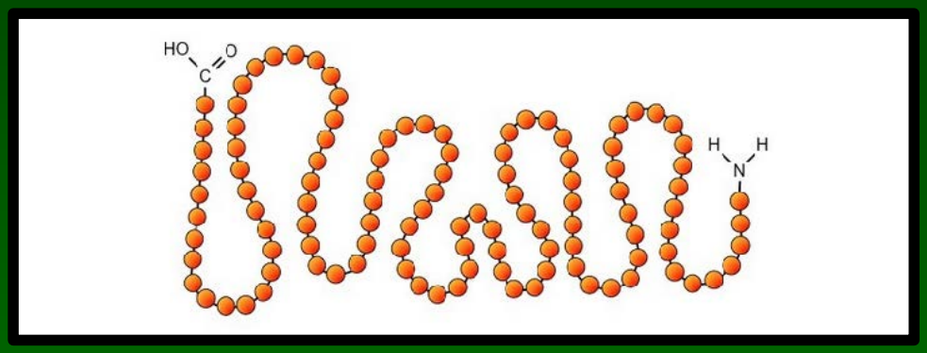


http://ib.bioninja.com.au/higher-level/topic-7-nucleic-acids/73-translation/protein-structure.html

Translation

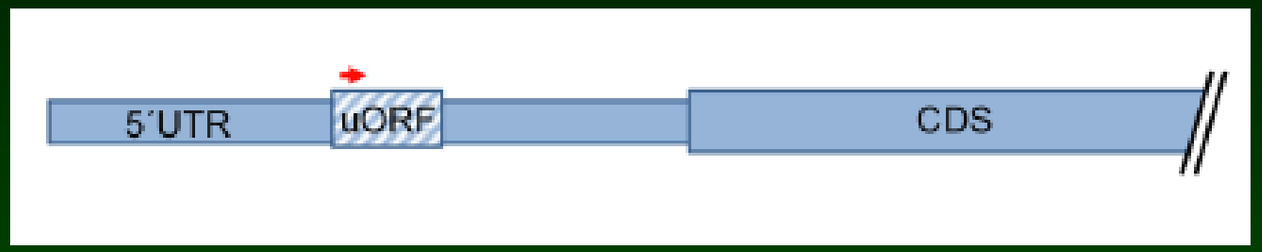
mRNA

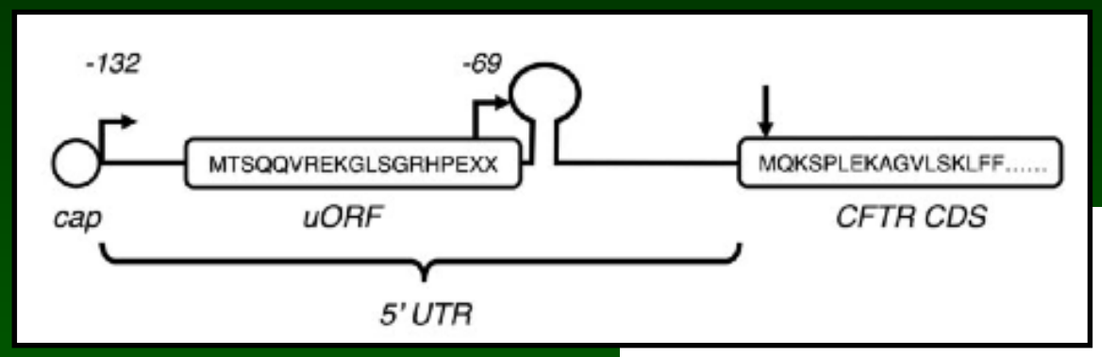




translation initiation site
(start codon, Kozak sequence)
translation termination site (stop codon)
uORF
codon usage > translation speed > protein folding
RNA/protein binding > stability, amount of protein
…

uORF





Human Molecular Genetics, 2015, Vol. 24, No. 4 **899–912**
doi:10.1093/hmg/ddu501
Advance Access published on September 30, 2014

CFTR mRNA expression is regulated by an upstream open reading frame and RNA secondary structure in its 5′ untranslated region

Samuel W. Lukowski[1,2,†,*], Joseph A. Rothnagel[1] and Ann E. O. Trezise[1,2]

influence process > increase protein expression

> treat CF-patients

Consequences: protein

- others signals present in protein?

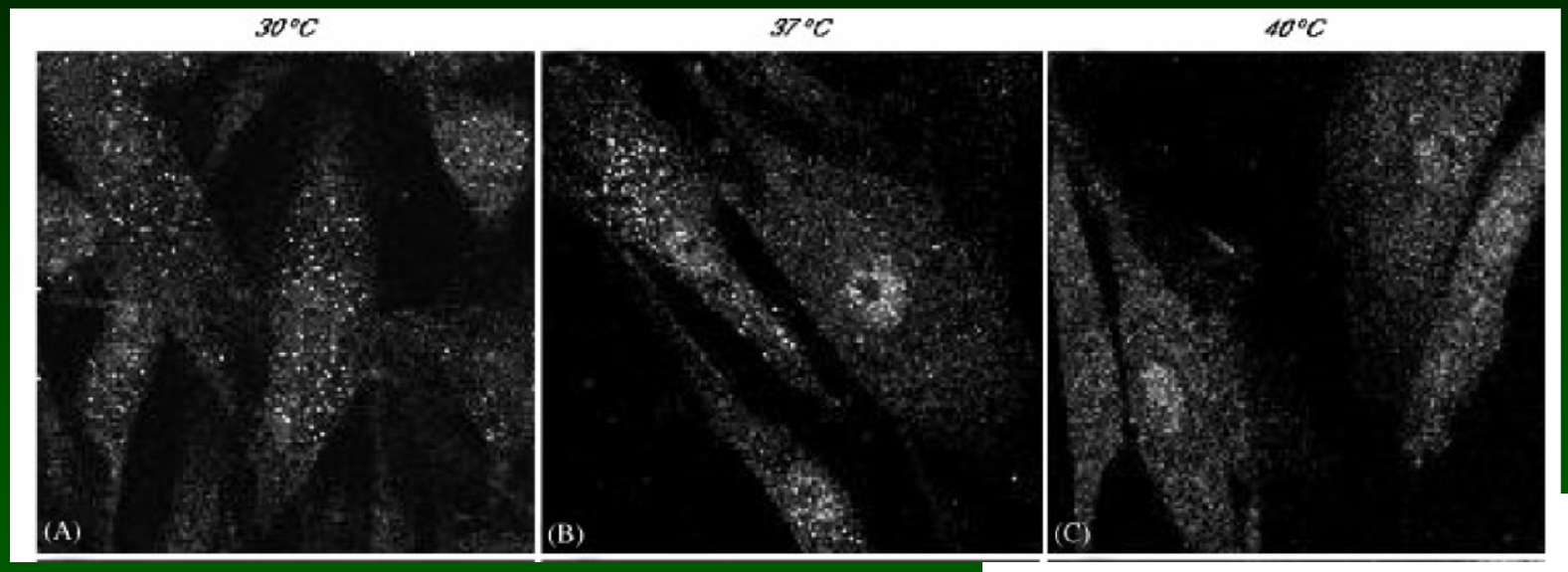
Consequences: protein

- **sequence**
  - *activity - enzymes*
  - *cleavage sites - to mature protein*
    (cell export/import)
  - *domains, signals - NLS*
  - *modification - phosphorylation, S-S bridges,*
    glycosylation
- **protein folding**
  - *secondary, tertiary, protein-protein interaction*
- **amount**
  - *too much / little*
  - *stability (protein turnover)* degration speed of protein
- **…**

...expression

- **place**
  expression in wrong tissue
- **timing**
  expression at the wrong time
  during development
  responding to incorrect trigger
- ...

T-sensitive



low T: dampens
high T: exaggerates

HUMAN MUTATION 24:130–139 (2004)

RESEARCH ARTICLE

Identification of the Molecular Defect in Patients With Peroxisomal Mosaicism Using a Novel Method Involving Culturing of Cells at 40°C: Implications for Other Inborn Errors of Metabolism

Jeannette Gootjes,[1] Frank Schmohl,[1] Petra A.W. Mooijer,[2] Conny Dekker,[2] Hanna Mandel,[3] Meral Topcu,[4] Martina Huemer,[5] M. von Schütz,[6] Thorsten Marquardt,[7] Jan A. Smeitink,[8] Hans R. Waterham,[2] and Ronald J.A. Wanders[1,2*]

temprature change impacts slicing

Consequences: DNA

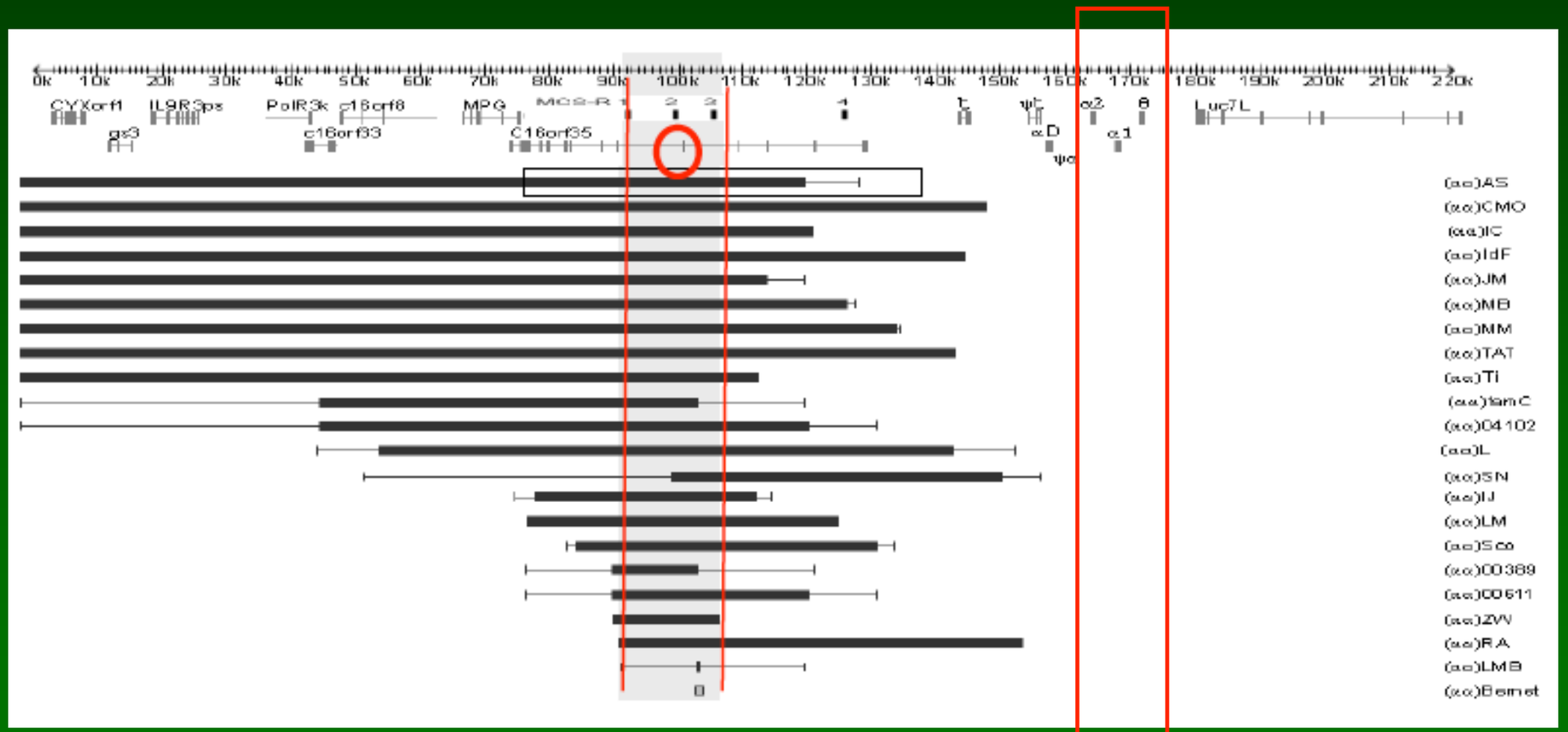
- **signals present in DNA?**
  destroyed or created

Consequences: DNA

- **replication origin**
- **TAD**
  topologically active domain (expression control)
- **expression**
  enhancer (LCR), transcription factor binding, …
- **protein binding**
  required for any process
- **DNA modification** *(methylation)*
  X-inactivation, imprinting, gene expression
- **…**
  …

LCR

- **expression**
  - ***LCR - locus control region***
    - ***regulatory elements often kb's to sometimes Mb's upstream***

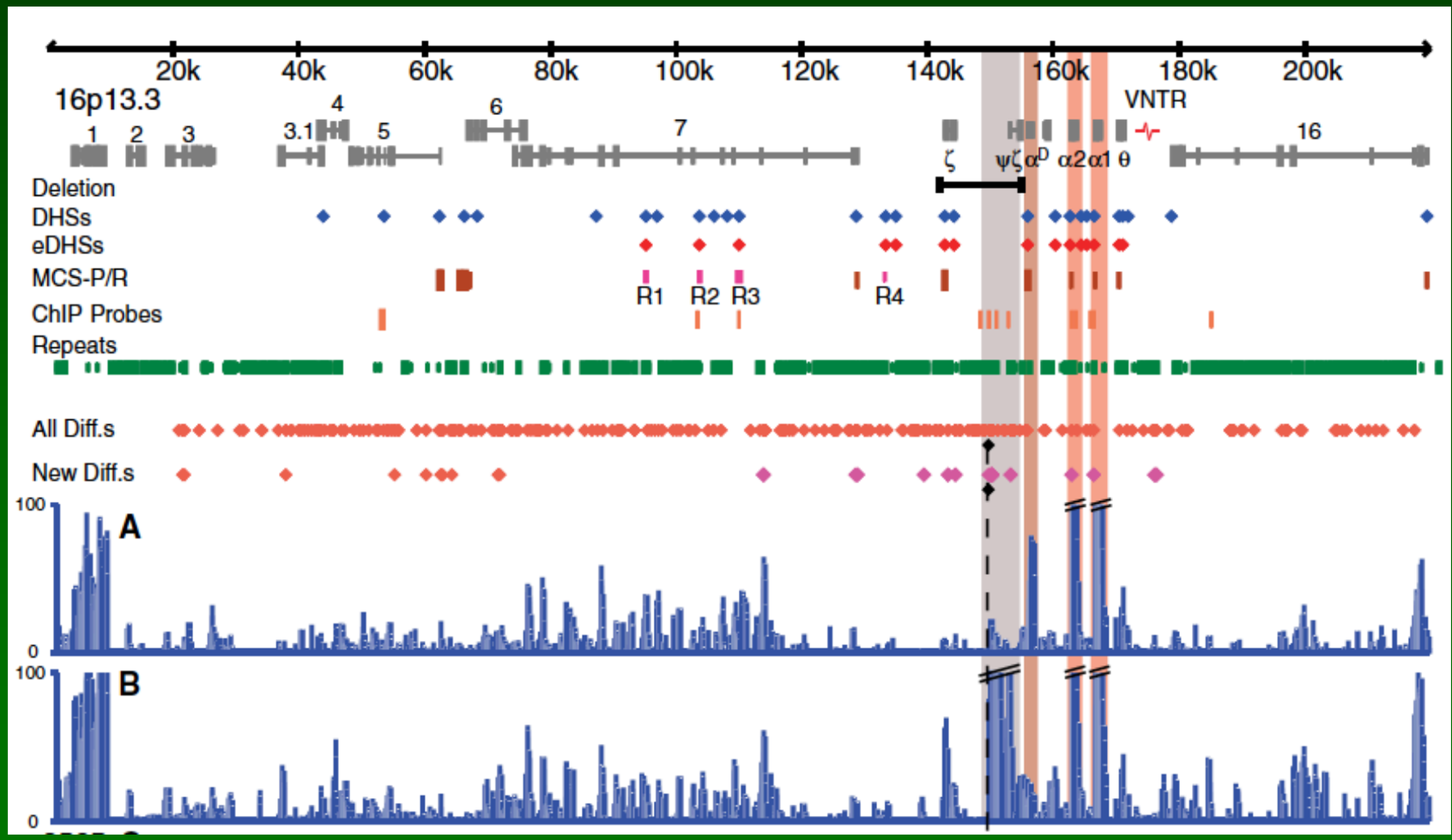


Higgs, Cambridge University Press (2009)

HBA gene cluster

New promoter

- **expression**
  - ***new promoter***
    - ***SNV creates TFB site, activating transcription, silencing downstream genes***



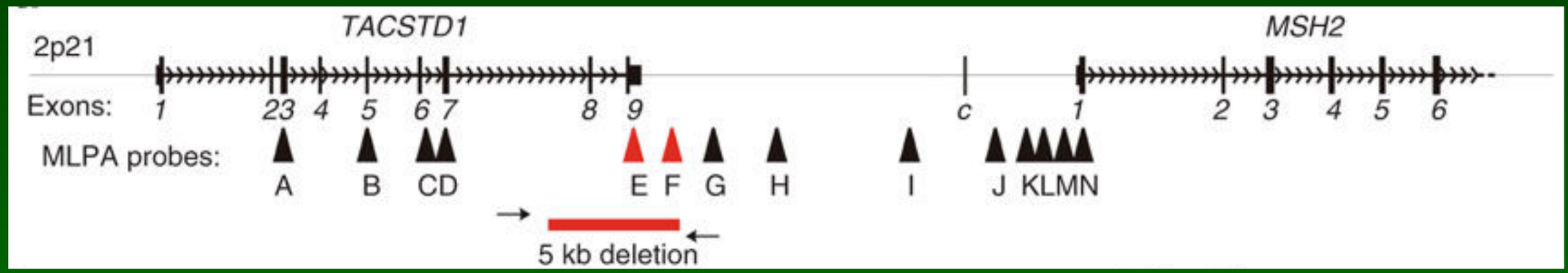
HBA genes

De Gobbi, Science (2006) 312:1215

Indirect silencing

- **expression**
  - ***gene silencing***
    - ***read-through silences downstream gene***

Ligtenberg, Nature Genet. (2009) 41:112.



EPCAM gene *MSH2 gene*

deletion 3’end EPCAM > transcription continues into MSH2

fusion transcript (out-frame), silencing MSH2 promoter > no MSH2

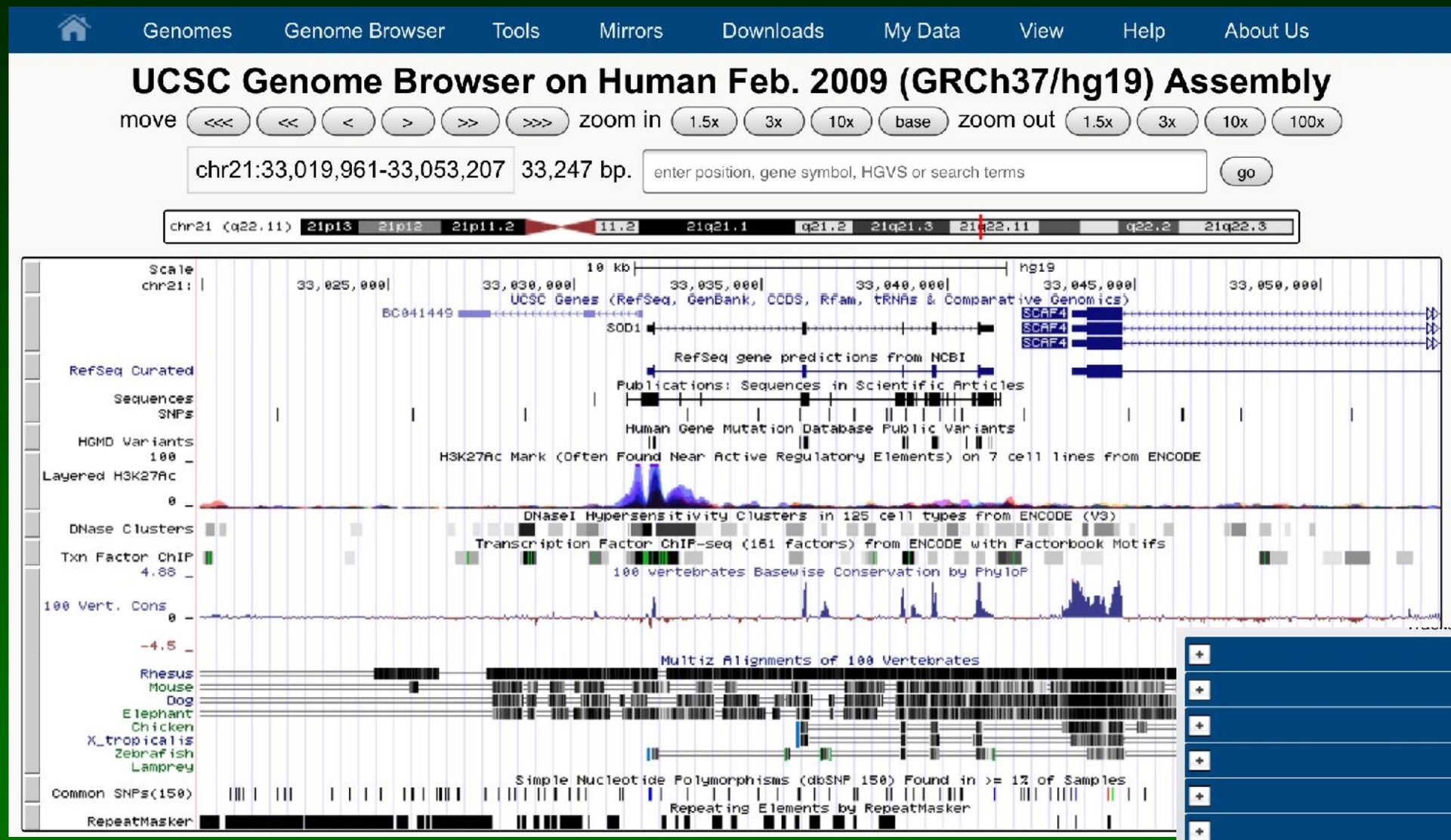
Non-coding genes

- **variants in non-coding genes**
  probably many still unknown

Genomic Element	Name	Disorder
MIR	MIR96	DFNA50 (Autosomal Dominant deafness 50)
	MIR184	EDICT syndrome
	MIR17HG	Feingold syndrome 2
Long ncRNA	TERC	AD dyskeratosis congenita; susceptibility to aplastic anemia
	RMRP	CHH (cartilage hair hypoplasia) syndrome; anauxetic dysplasia; metaphyseal dysplasia without hypotrichosis
	CISTR-ACT lncRNA	Type E polydactyly
	HELLP lincRNA	HELLP (Hemolysis, Elevated Liver enzymes, Low Platelets) syndrome
	ATXN8/ATXN8OS	Spinocerebellar ataxia 8 (SCA8)
Small ncRNA	snRNA RNU4ATAC	Microcephalic Osteodysplastic Primordial Dwarfism, type I (MOPD I)

Makrythanasis and Antonarakis, Clin.Genet. (2013)

Where to find



conservation, activity

Listing variants

- **be specific**
  list per level: DNA, RNA, protein

where to list a DNA substitution,
altering RNA splicing (insertion),
resulting in a frame shift

- **so**

DNA

substitution, deletion, duplication, insertion, delins, inversion, ...
(large deletion, duplication, insertion, ...)

RNA

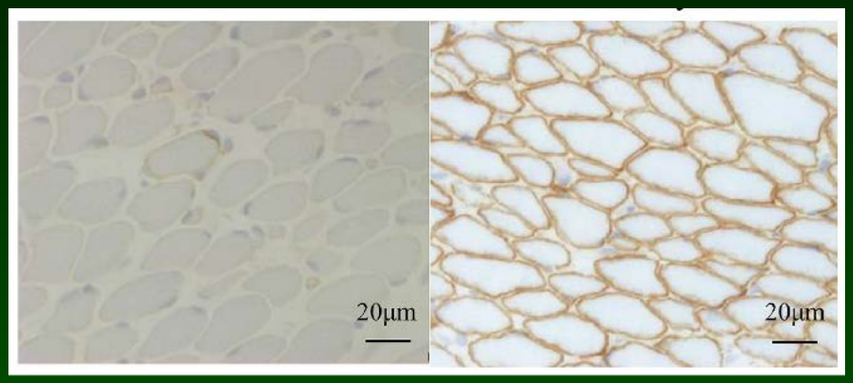
substitution, deletion, duplication, insertion, delins, ...
(splice deletion or insertion)

protein

missense, nonsense, frame shift, in frame, deletion / duplication / insertion / delins

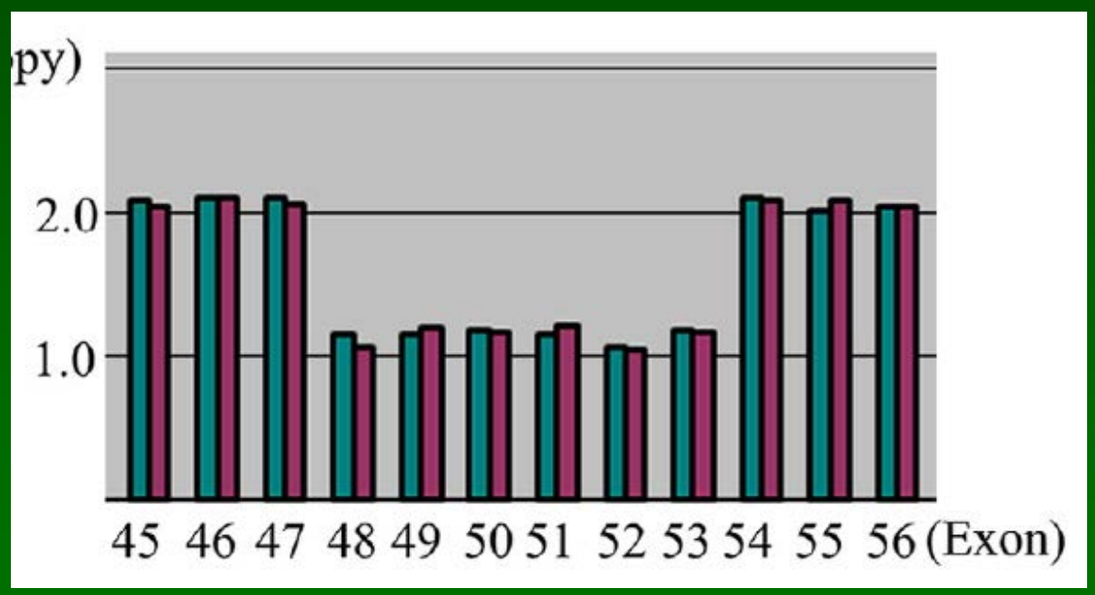
Female DMD

- **muscle biopsy**
  no dystrophin staining



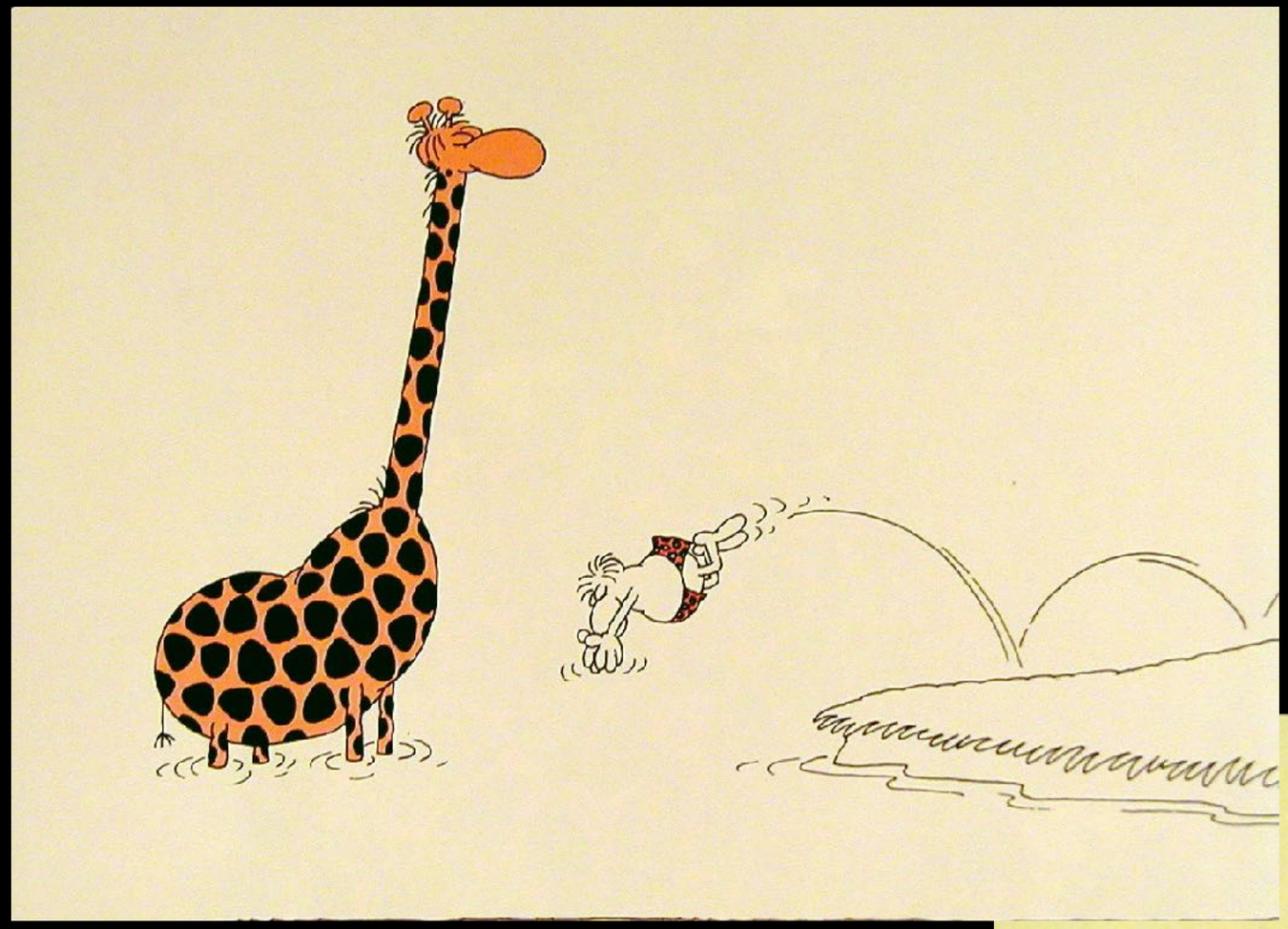
patient *control*

- **MLPA**
  deletion exons 48-53
  in-frame

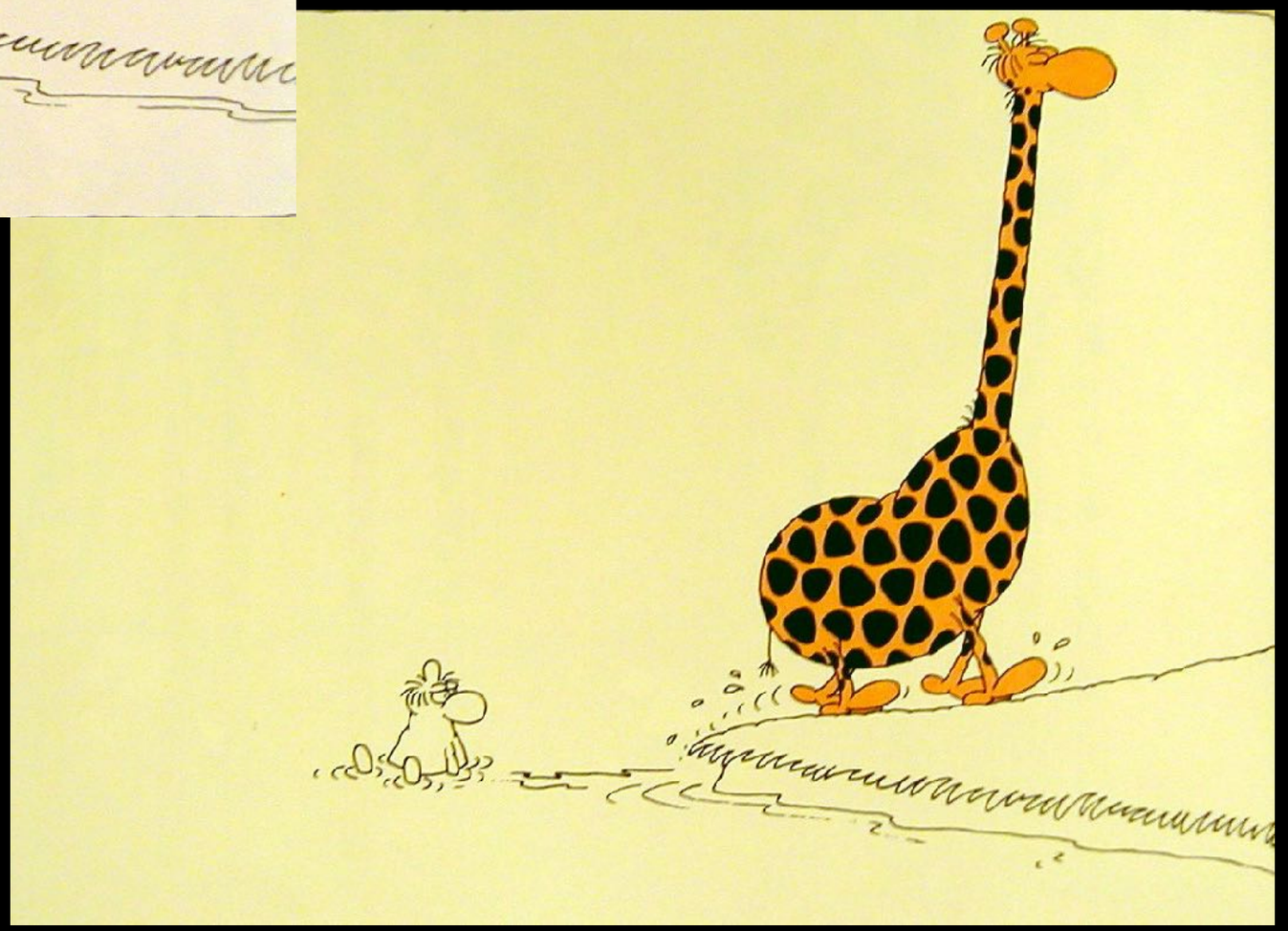


- **sequencing**
  no deleterious variants
- **X-inactivation**
  random

please explain



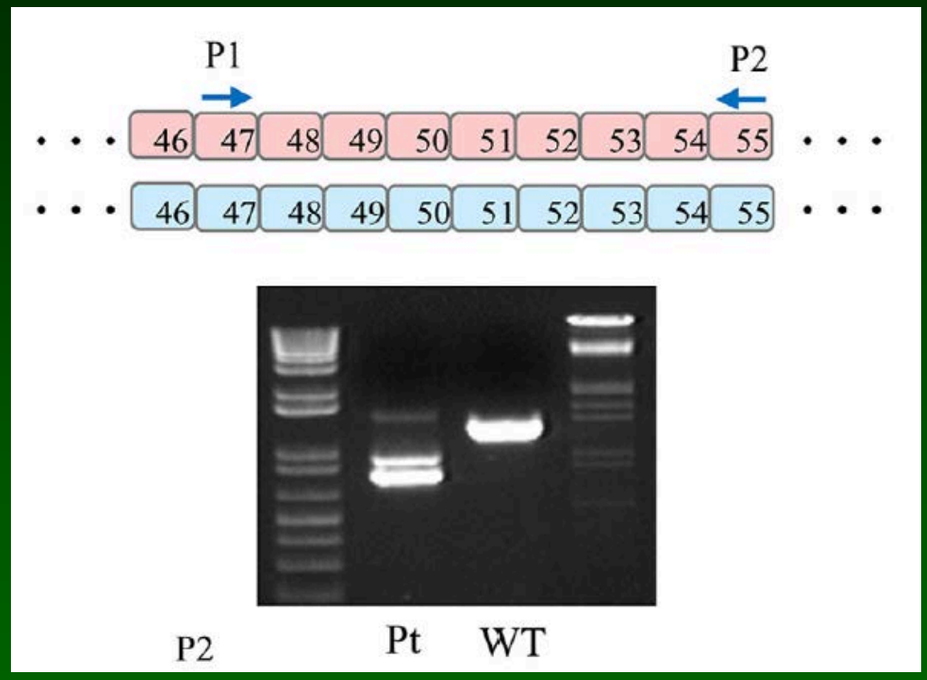
Is my conclusion right ?



Female DMD

- **RNA analysis**

RNA analysis



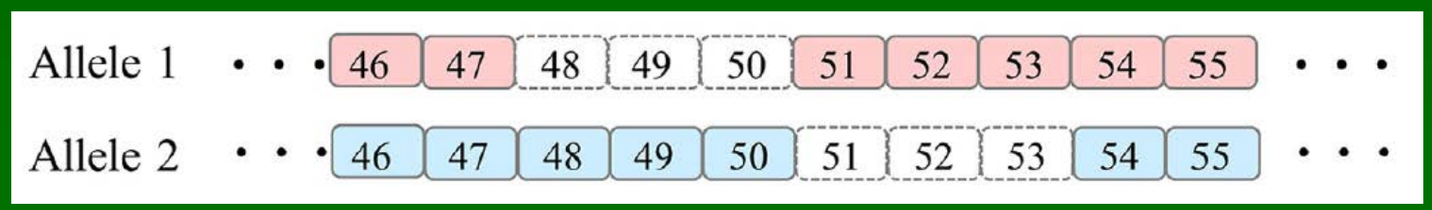
ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
NMD
Neuromuscular Disorders 27 (2017) 569–573
www.elsevier.com/locate/nmd

Case report

Duchenne muscular dystrophy in a female with compound heterozygous contiguous exon deletions

Eri Takeshita [a,*], Narihiro Minami [b,c], Kumiko Minami [c], Mikiya Suzuki [d], Takeya Awashima [a], Akihiko Ishiyama [a], Hirofumi Komaki [a], Ichizo Nishino [c,e], Masayuki Sasaki [a]

two different deletions, both frame shifting



Your example

- …

Acknowledgement

Presentation prepared by:

Johan den Dunnen

Human Genetics & Clinical Genetics
Leiden University Medical Center
Leiden, Nederland

date: April 2019

based on lecture Jan Traeger-Synodinos
VEPTC Prague 2017