RNA variants

position and possible consequences

everything what can go wrong, will go wrong

Johan den Dunnen

Gene < > phenotype

- **gene function**

  gene function 'explains' phenotype
  combination should make 'sense'
  for analysis use affected tissue
  RNA, protein analysis

- **expression**

  gene expressed in affected tissue
  select major reference transcript

RNA

..the neglected molecule

under-appreciated

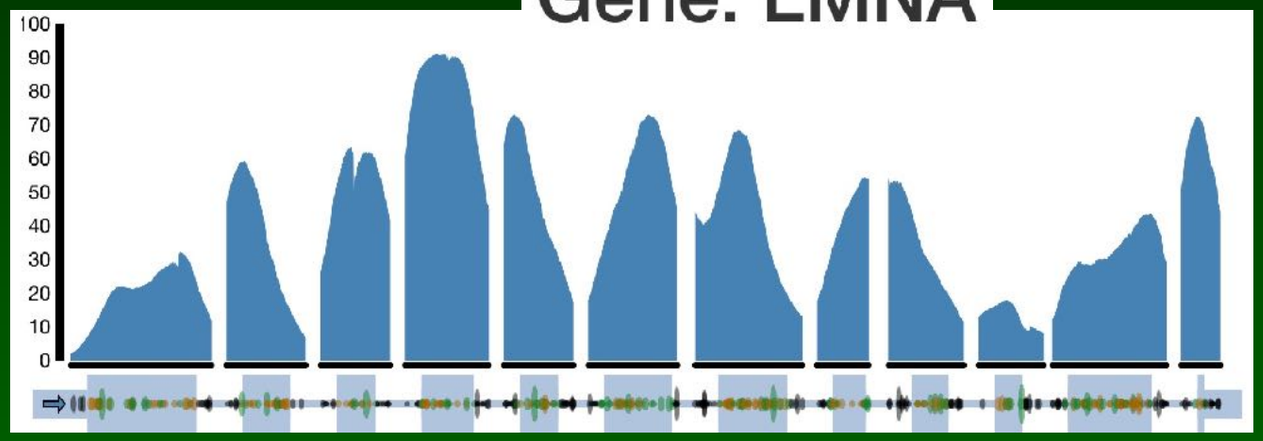
most go blindly DNA > protein

..there is much more

gnomAD / ExAC

ExAC Browser (Beta) | Exome Aggregation Consortium

Gene: LMNA



1:156084751 G / A	1	156084751	n.80-4G>A~	PASS	splice region		1	27826	0	0.00003594
1:156084753 A / C	1	156084753	n.80-2A>C†	PASS	splice acceptor	LC LoF	1	29400	0	0.00003401
1:156084756 C / A	1	156084756	p.Ala16Asp	PASS	missense		2	31032	0	0.00006445
1:156084760 C / T (rs11549668)	1	156084760	p.Ser17Ser	PASS	synonymous		813	33894	8	0.02399
1:156084783 G / T	1	156084783	p.Arg25Leu	PASS	missense		2	44832	0	0.00004461
1:156084787 C / T	1	156084787	p.Ile26Ile	PASS	synonymous		5	46600	0	0.0001073
1:156084841 C / G	1	156084841	p.Val44Val	PASS	synonymous		1	51050	0	0.00001959
1:156084851 C / G	1	156084851	p.Arg46Gly	PASS	missense		2	50862	0	0.00003932
1:156084856 G / A	1	156084856	n.355G>A†	PASS	splice region		1	51262	0	0.00001951
1:156084858 G / A	1	156084858	n.356+1G>A†	PASS	splice donor	LC LoF	3	51416	0	0.00005835
1:156084859 C / T	1	156084859	n.356+2C>T†	PASS	splice donor	LC LoF	6	52036	0	0.0001153
1:156084862 C / T	1	156084862	n.356+5C>T†	PASS	splice region		1	52004	0	0.00001923

Gene: CAV3

Constraint from ExAC	Expected no. variants	Observed no. variants	Constraint Metric
Synonymous	29.9	32	z = -0.24
Missense	59.8	41	z = 1.19
LoF	4.2	1	pLI = 0.34
CNV	2.4	5	z = -0.47

RNA?

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
African	121	5274	2	0.02294
Other	2	532	0	0.003759
Latino	7	6052	0	0.001157
European (Non-Finnish)	1	39690	0	2.52e-05
East Asian	0	5584	0	0
European (Finnish)	0	4052	0	0
South Asian	0	10756	0	0
Total	**131**	**71940**	**2**	**0.001821**

ClinVar DMD

825.	NM_004006.2(**DMD**):c.3637A>G (p.Lys1213Glu) *GRCh37:* ChrX:32466722 *GRCh38:* ChrX:32448605	**DMD**	not specified	Uncertain significance (Feb 1, 2016)	criteria provided, single submitter
826.	NM_004006.2(**DMD**):c.3631G>T (p.Glu1211Ter) *GRCh37:* ChrX:32466728 *GRCh38:* ChrX:32448611	**DMD**	Becker muscular dystrophy	Pathogenic (Nov 1, 1997)	no assertion criteria provided
827.	NM_004006.2(**DMD**):c.3604-12T>A *GRCh37:* ChrX:32466767 *GRCh38:* ChrX:32448650	**DMD**	not specified	Likely benign (Aug 12, 2016)	
828.	NM_004006.2(**DMD**):c.3604-14T>C *GRCh37:* ChrX:32466769 *GRCh38:* ChrX:32448652	**DMD**	not specified	Likely benign (Nov 16, 2016)	
829.	NM_004006.2(**DMD**):c.3603+15dupA *GRCh37:* ChrX:32472764 *GRCh38:* ChrX:32454647	**DMD**	not specified	Likely benign (Apr 28, 2015)	

no RNA

seems blind
DNA > protein

(linked paper may have RNA)

NCBI Resources How To

ClinVar | ClinVar | Search ClinVar for gene symbols, HGVS expressio

Advanced

Home | About | Access | Help | Submit | Statistics | FTP

NM_004006.2(DMD):c.3631G>T (p.Glu1211Ter)

Variation ID: 11280

Review status: ☆ ☆ ☆ ☆ (0/4) no assertion criteria provided

Interpretation Go to:

Clinical significance: Pathogenic

Last evaluated: Nov 1, 1997

Number of submission(s): 1

Condition(s): Becker muscular dystrophy [MedGen - Orphanet - OMIM]

See supporting ClinVar records

Allele(s) Go to:

NM_004006.2(DMD):c.3631G>T (p.Glu1211Ter)

LOVD DMD

27	c.3628_3665del	-	r.3628_3665del	p.Lys1210*
27	c.3630delA	-	r.(?)	p.(Glu1211Lysfs*4)
27	c.3631G>T	3839G>T	r.[3631g>u, 3604_3786del, 3604_4071del]	p.[Glu1211*; Arg1202_1262del; Arg1202_1357del]
27	c.3679C>T	-	r.(?)	p.(Gln1227*)
27	c.3697delC			p.(Gln1233Lysfs*4)
27	c.3700G>T			p.(Glu1234*)
27	c.3705C>T (Reported 2 times)			p.(=)

Patient data (#0006974)	
Phenotype	muscular dystrophy, Becker (BMD)
Phenotype additional	-
Reference	Japan:Kobe
Remarks	-
Geographic origin	Japan
Ethnic origin	-
Gender	M
Inheritance	unknown
Consanguinity	-
Fam_Pat	-
# reported	1
CK level	-
Protein data	-
Submitter	Masafumi Matsuo

Variant data	
Allele	Parent #1
Reported pathogenicity	Pathogenic
Concluded pathogenicity	Unknown
Exon	27
DNA change	c.3631G>T (View in UCSC Genome Browser, Ensembl)
Var_pub_as	3839G>T
RNA change	r.[3631g>u, 3604_3786del, 3604_4071del]
Protein change	p.[Glu1211*; Arg1202_1262del; Arg1202_1357del]
DB-ID	DMD_00074
Variant remarks	10% diff.splice
Genet_ori	germline (inherited)
Segregation	-
Reference	Shiga, Takeshima 2010, (OMIM 0074)
Template	DNA, RNA
Technique	RT-PCR, SEQ, SSCA
Frequency	-
RE-site	-

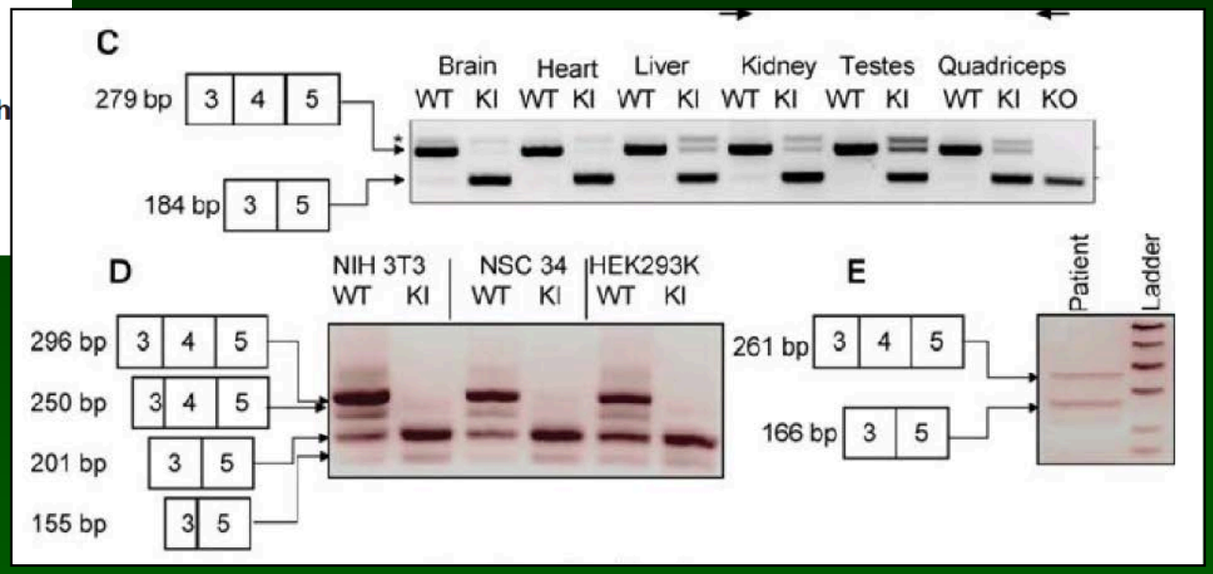
825.	NM_004006.2(DMD):c.3637A>G (p.Lys1213Glu) GRCh37: ChrX:32466722 GRCh38: ChrX:32448605	DMD	not specified	Uncertain significance (Feb 1, 2016)	criteria provided, single submitter
826.	NM_004006.2(DMD):c.3631G>T (p.Glu1211Ter) GRCh37: ChrX:32466728 GRCh38: ChrX:32448611	DMD	Becker muscular dystrophy	Pathogenic (Nov 1, 1997)	no assertion criteria provided
827.	NM_004006.2(DMD):c.3604-12T>A GRCh37: ChrX:32466767 GRCh38: ChrX:32448650	DMD	not specified	Likely benign (Aug 12, 2018)	criteria provided, single submitter
828.	NM_004006.2(DMD):c.3604-14T>C GRCh37: ChrX:32466769 GRCh38: ChrX:32448652	DMD	not specified	Likely benign (Nov 16, 2018)	criteria provided, single submitter
829.	NM_004006.2(DMD):c.3603+15dupA GRCh37: ChrX:32472764 GRCh38: ChrX:32454647	DMD	not specified	Likely benign (Apr 28, 2015)	criteria provided, single submitter

Not like this

Human Molecular Genetics, 2012, Vol. 21, No. 4 **811–825**
doi:10.1093/hmg/ddr512
Advance Access published on November 7, 2011

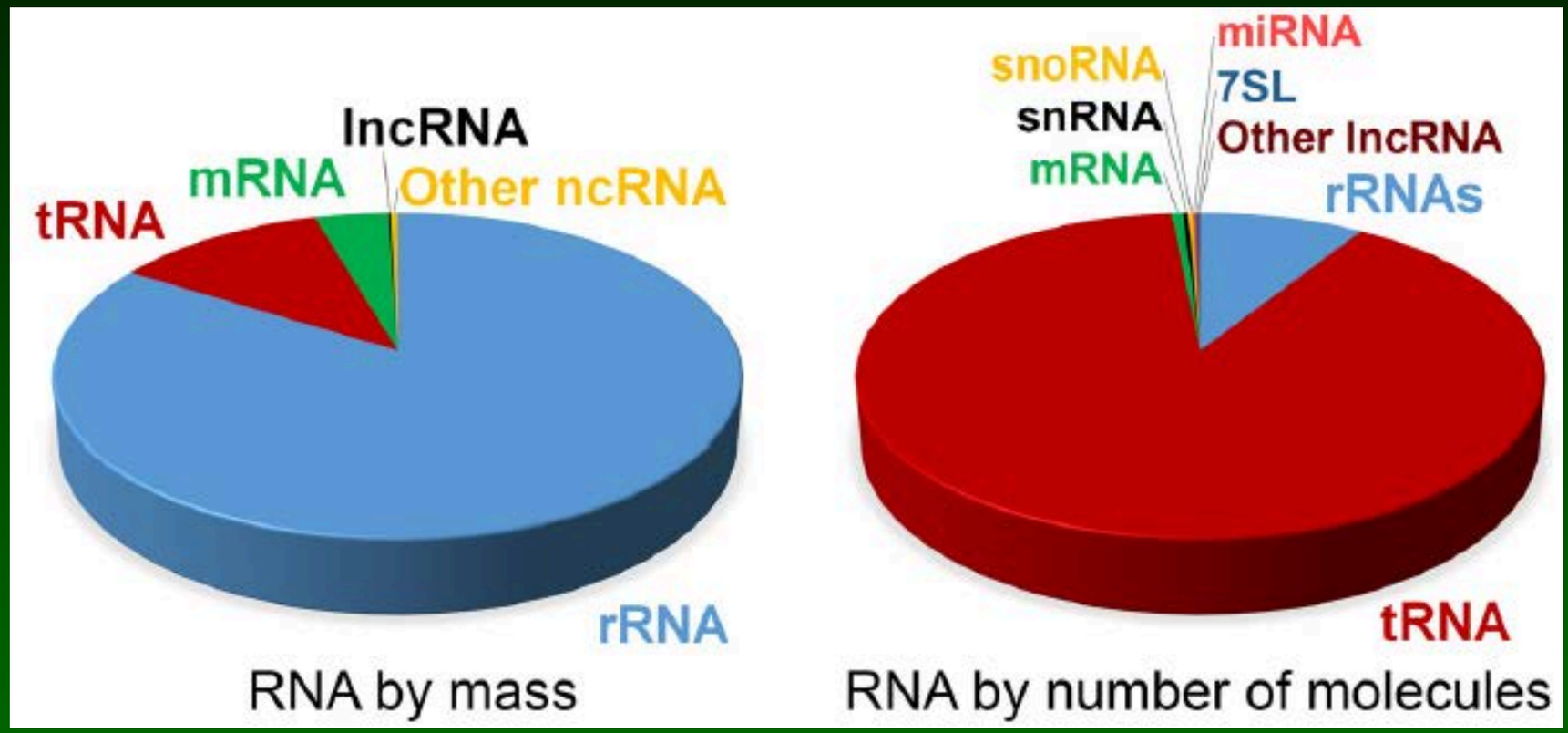
Modeling the human *MTM1* p.R69C mutation in murine *Mtm1* results in exon 4 skipping and a less severe myotubular myopathy phenotype

Christopher R. Pierson[1,2,4,*], Ashley N. Dulin-Smith[1], Ashley N. Durban[1], Morgan L. Marsh[1], Jordan T. Marshall[1], Andrew D. Snyder[1], Nada Naiyer[1], Jordan T. Gladman[1], Dawn S. Chandler[1,3,4], Michael W. Lawlor[5,†], Anna Buj-Bello[6], James J. Dowling[7] and Alan H. Beggs[5,*]



- **unclear missense effect**
  - ***mouse model generated***
  - ***> no protein in mouse***
    - ***> RNA analysis shows splice effect***
      - ***> confirmed in human***
        - ***> explains severity phenotype***

RNA types



Palazzo & Lee (2015) Front Genet 6:1-11

- **coding (mRNA) > protein**
- **non-coding > regulation**

amount: RNA stability, translation

1 gene / many RNAs

- **transcription initiation**
  different promoter / first exon
  stability, uORF, cap

where, when, how much

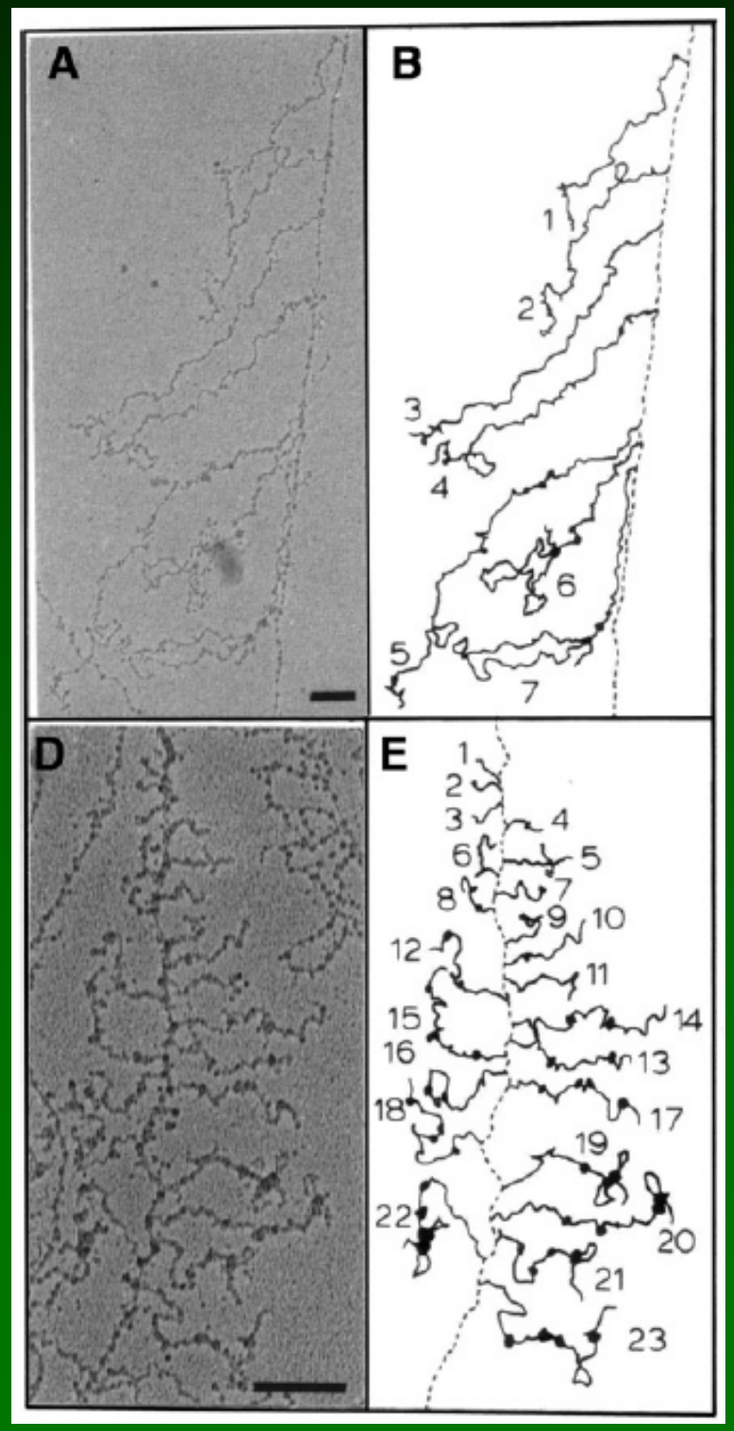
- **splicing**
  inclusion / exclusion exons
  encoding different proteins

structure

- **polyA addition**
  alternative last exons
  alternative polyA sites
  +/- RNA regulatory sequences

stability

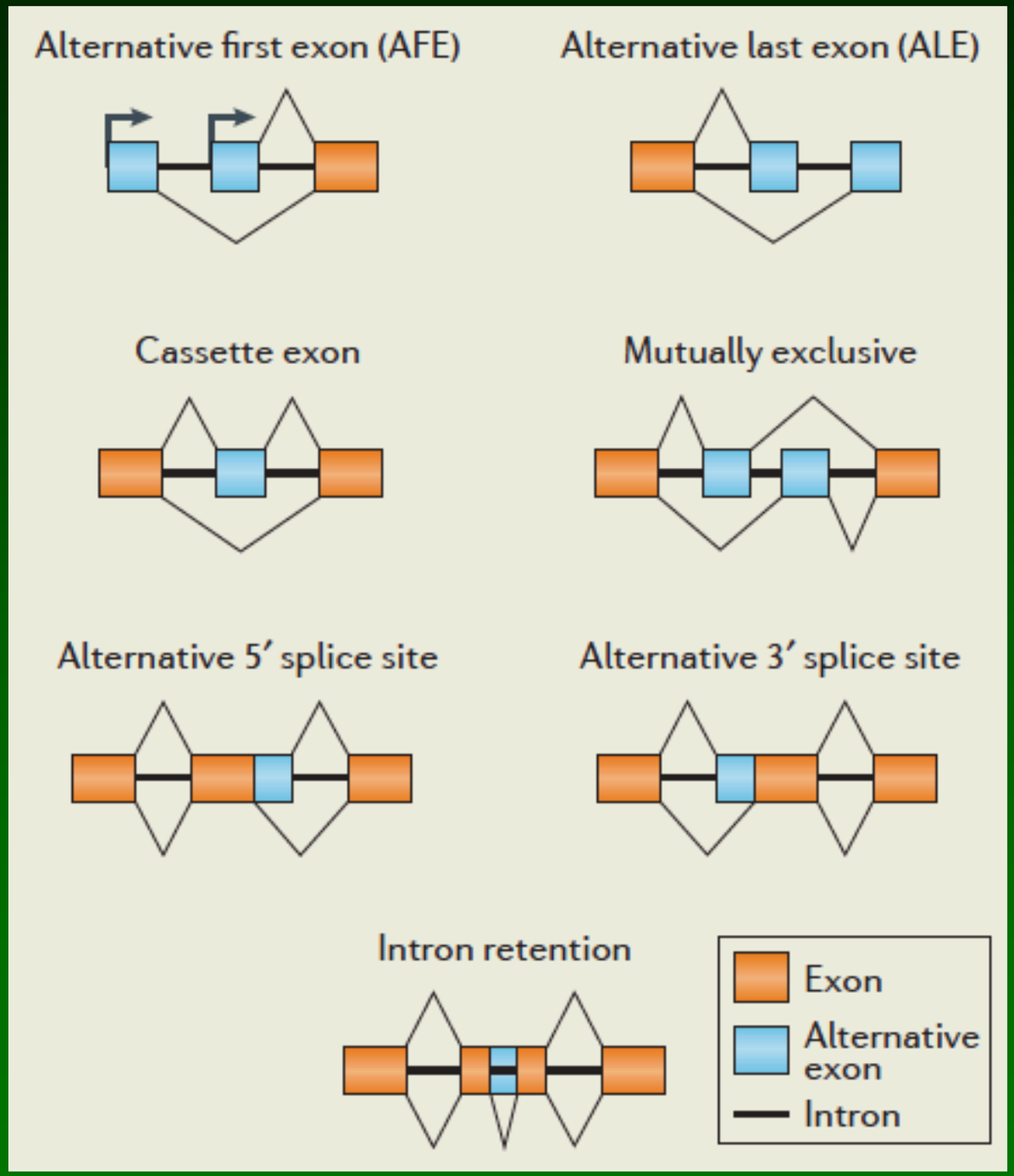
Transcription



EM transcription

Drosophila

1 gene / many RNAs



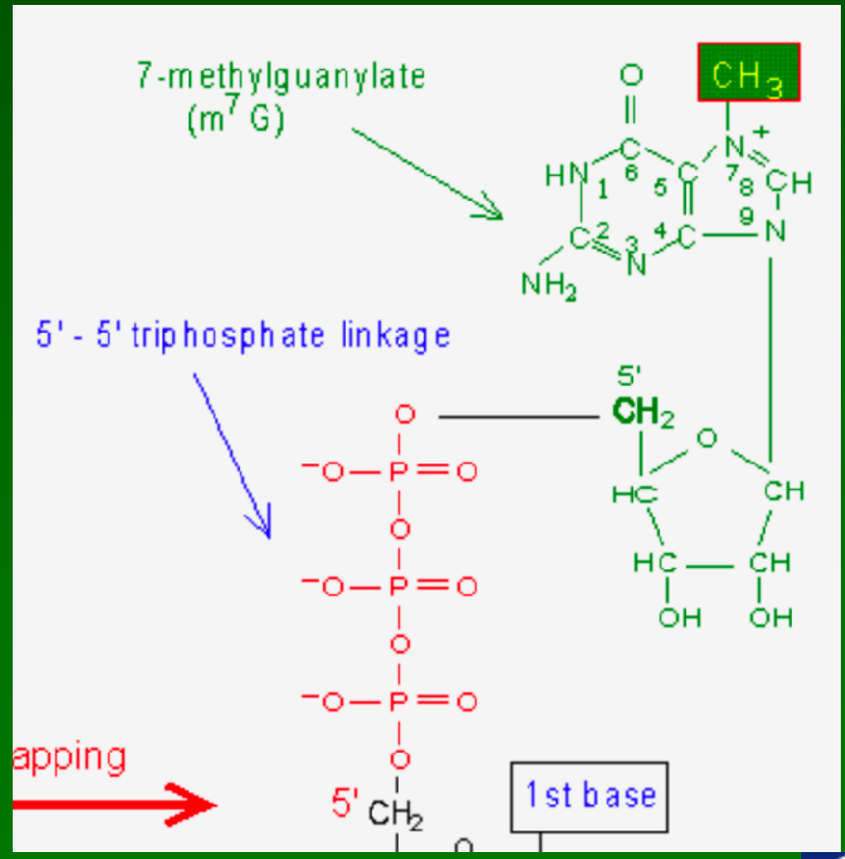
Scotti & Swanson (2016) Nat Rev Genet 17:19-32

Capping

- **5’ cap**
  added immediately after transcription starts

- **block 5’ end with 7-methylguanosine**

- **function**
  - ***prevent degradation***
    - ***by exonucleases***
  - ***regulation nuclear export***
  - ***promote translation***
    - ***loops with polyA***

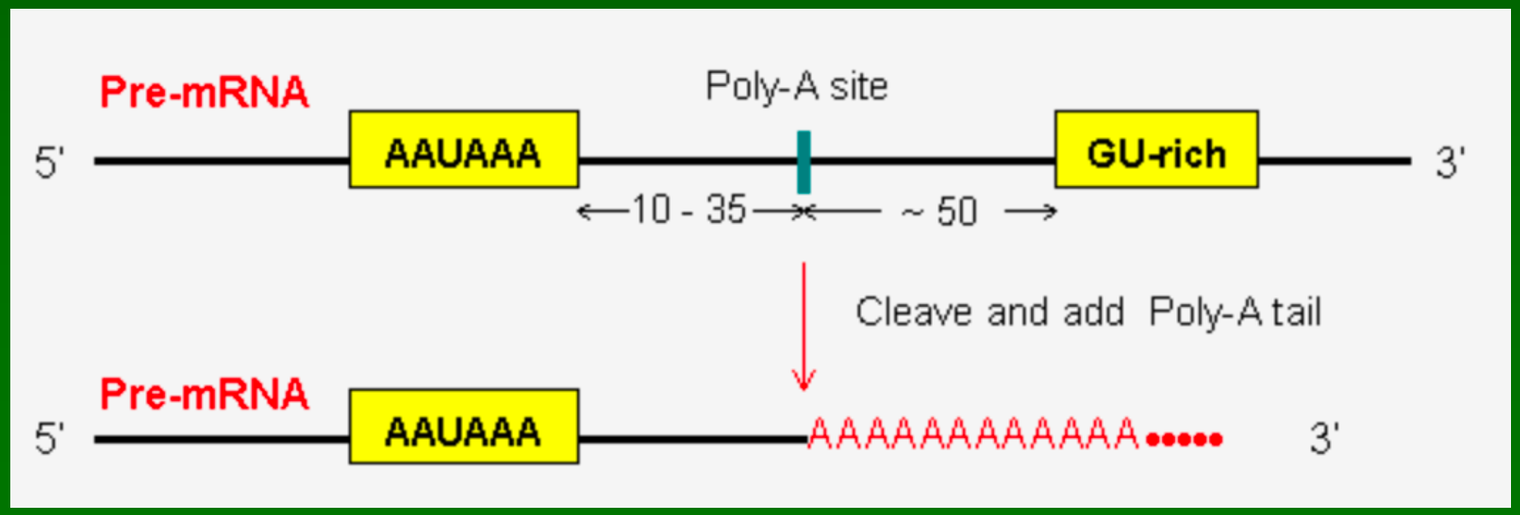


www.web-books.com/MoBio/Free/Ch5A.htm

PolyA addition

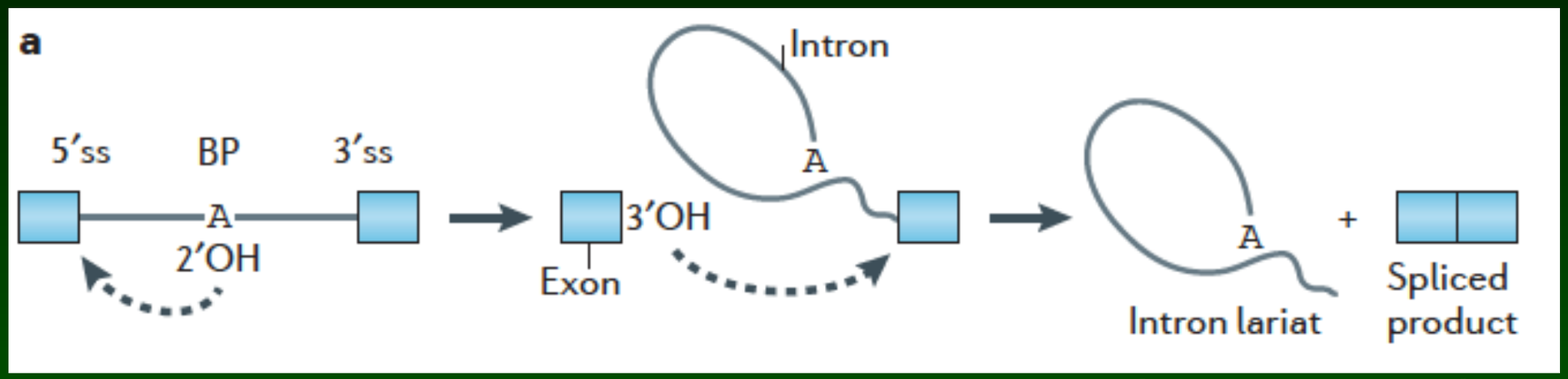
- **3’ polyA-tail**
  added at 3’ end many transcripts
- **one gene often several 3’ ends**
- **function**

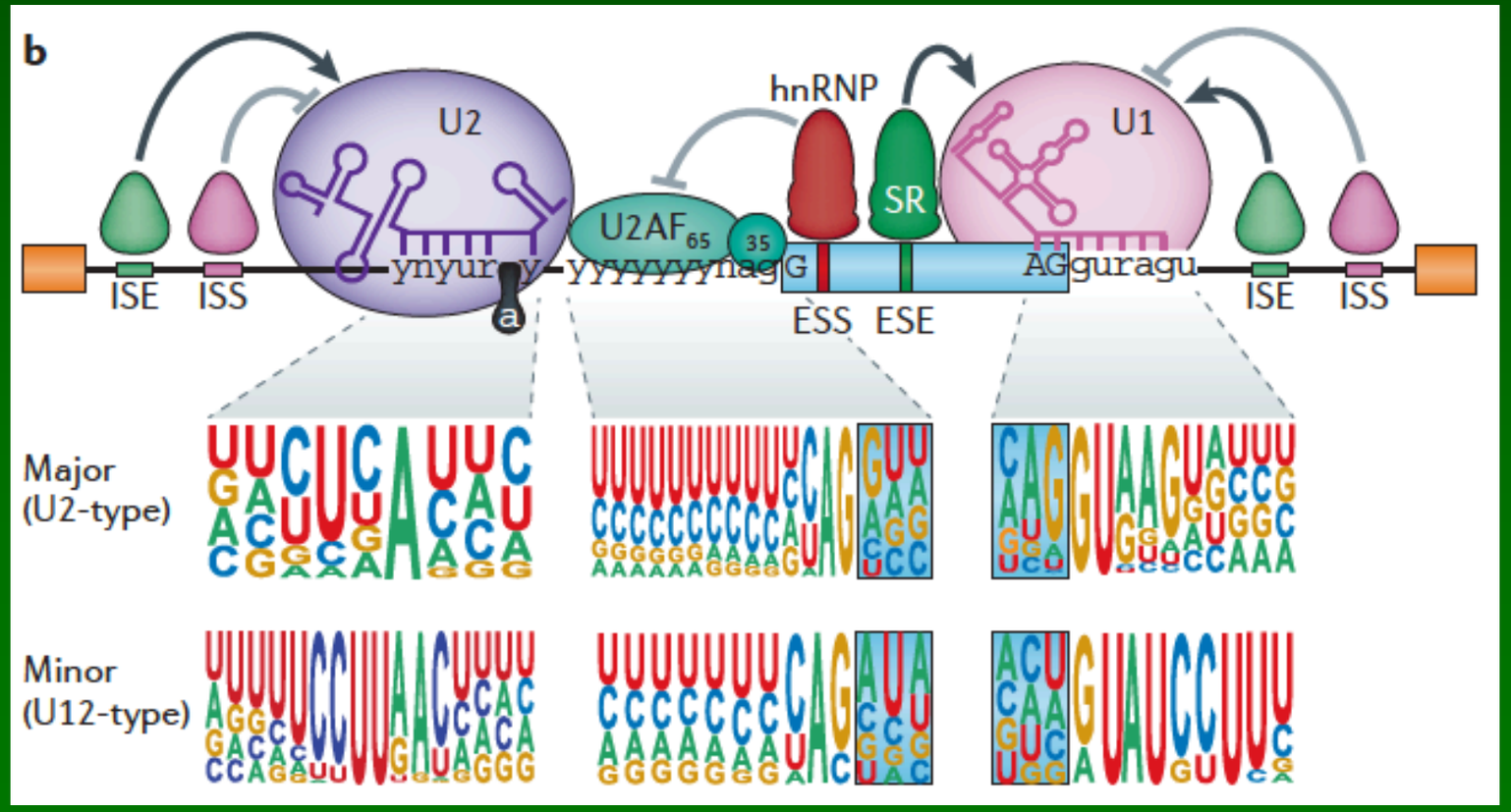
prevent degradation
by exonucleases
stability, influences translation



www.web-books.com/MoBio/Free/Ch5A.htm

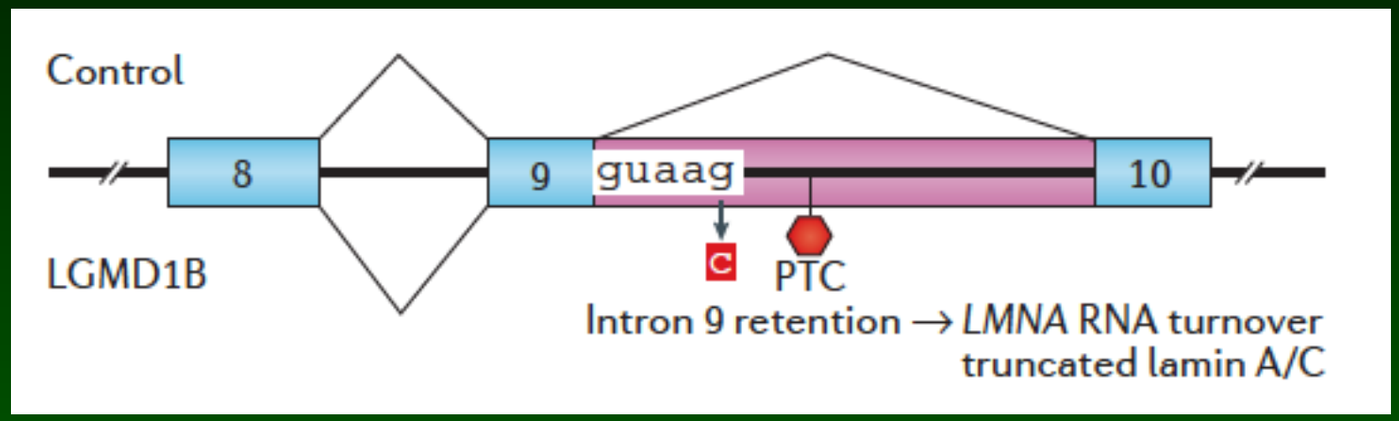
Splicing





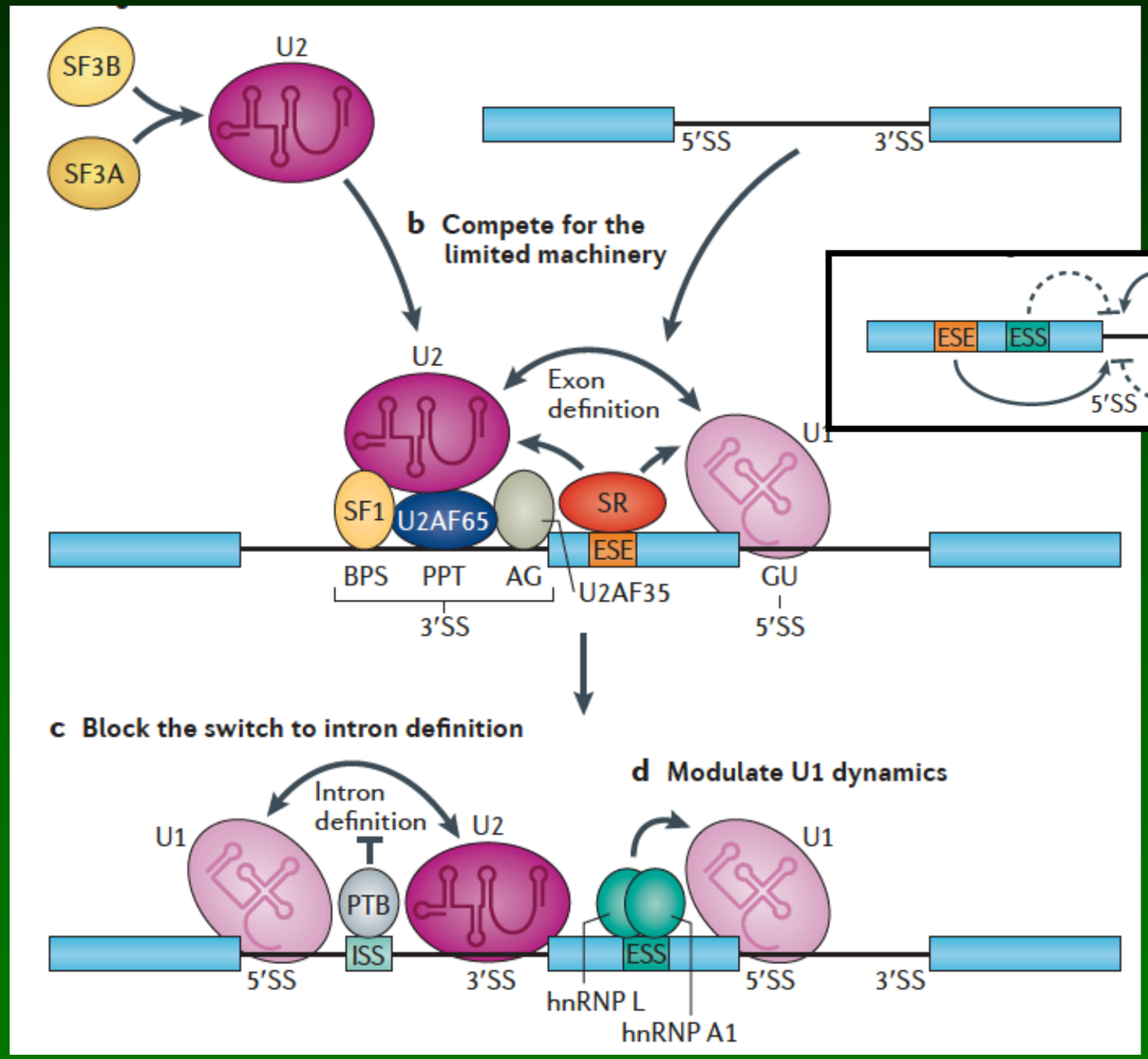
Scotti & Swanson (2016) Nat Rev Genet 17:19-32

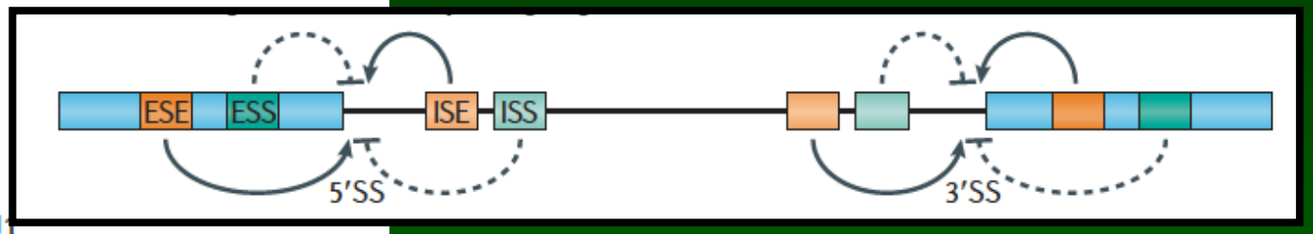
Variants & splicing



Scotti & Swanson (2016) Nat Rev Genet 17:19-32

Splicing regulation



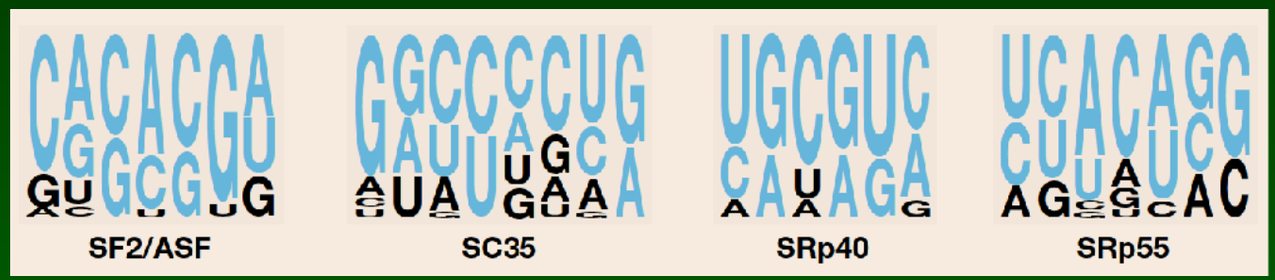


Fu & Ares (2014) Nat Rev Genet 15:689-701

Splicing regulation

- **standard signals**
  - *splice donor (GT)*
  - *splice acceptor (AG)*
  - *branch point (A)*

Cartegni (2002) Nat Rev Genet 3:285-298

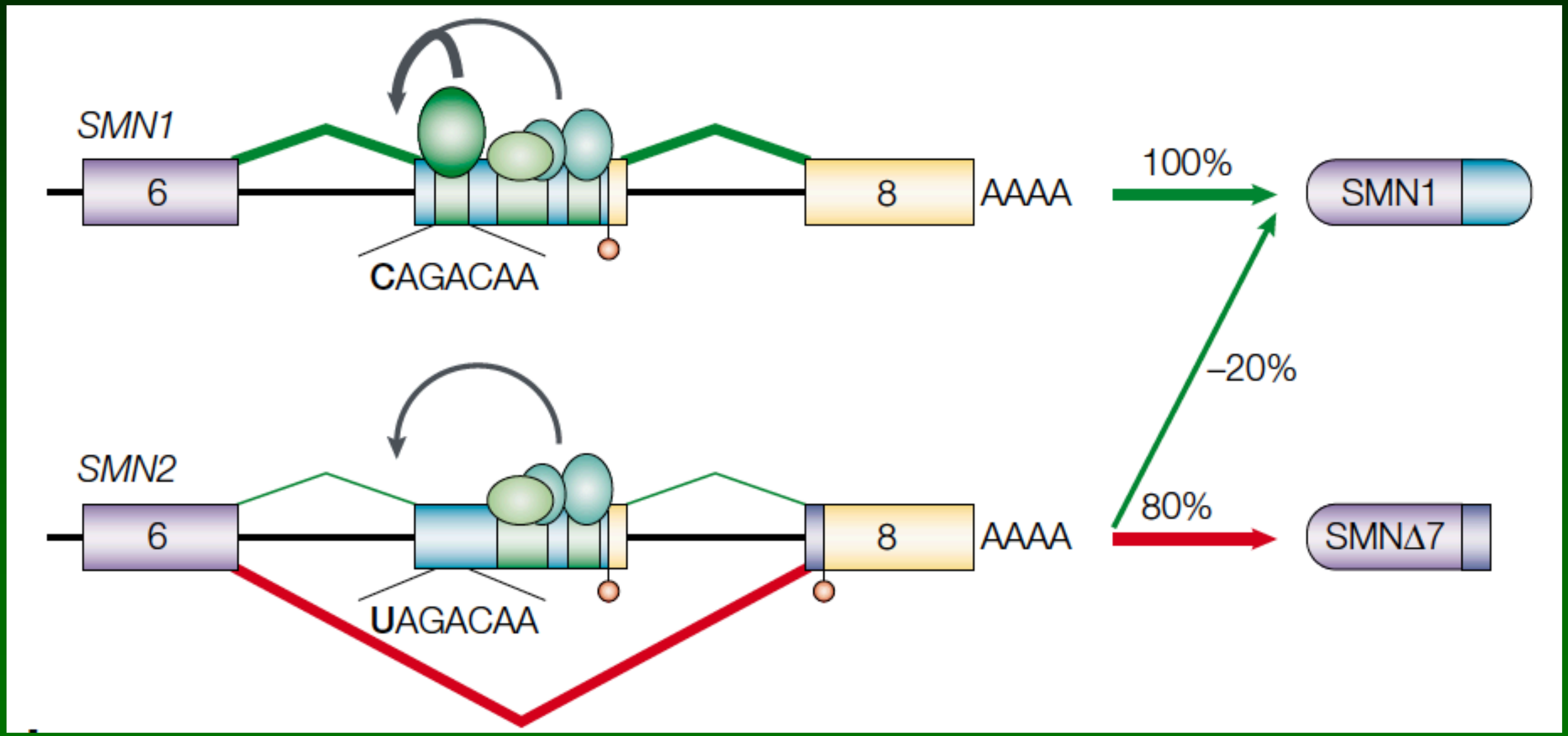


- **exonic**
  - *ESE - exonic splice enhancer*
  - *ESS - exonic splice silencer (ESR - repressor)*

- **intronic**
  - *ISE - intronic splice enhancer*
  - *ISS - intronic splice silencer*

Splicing regulation

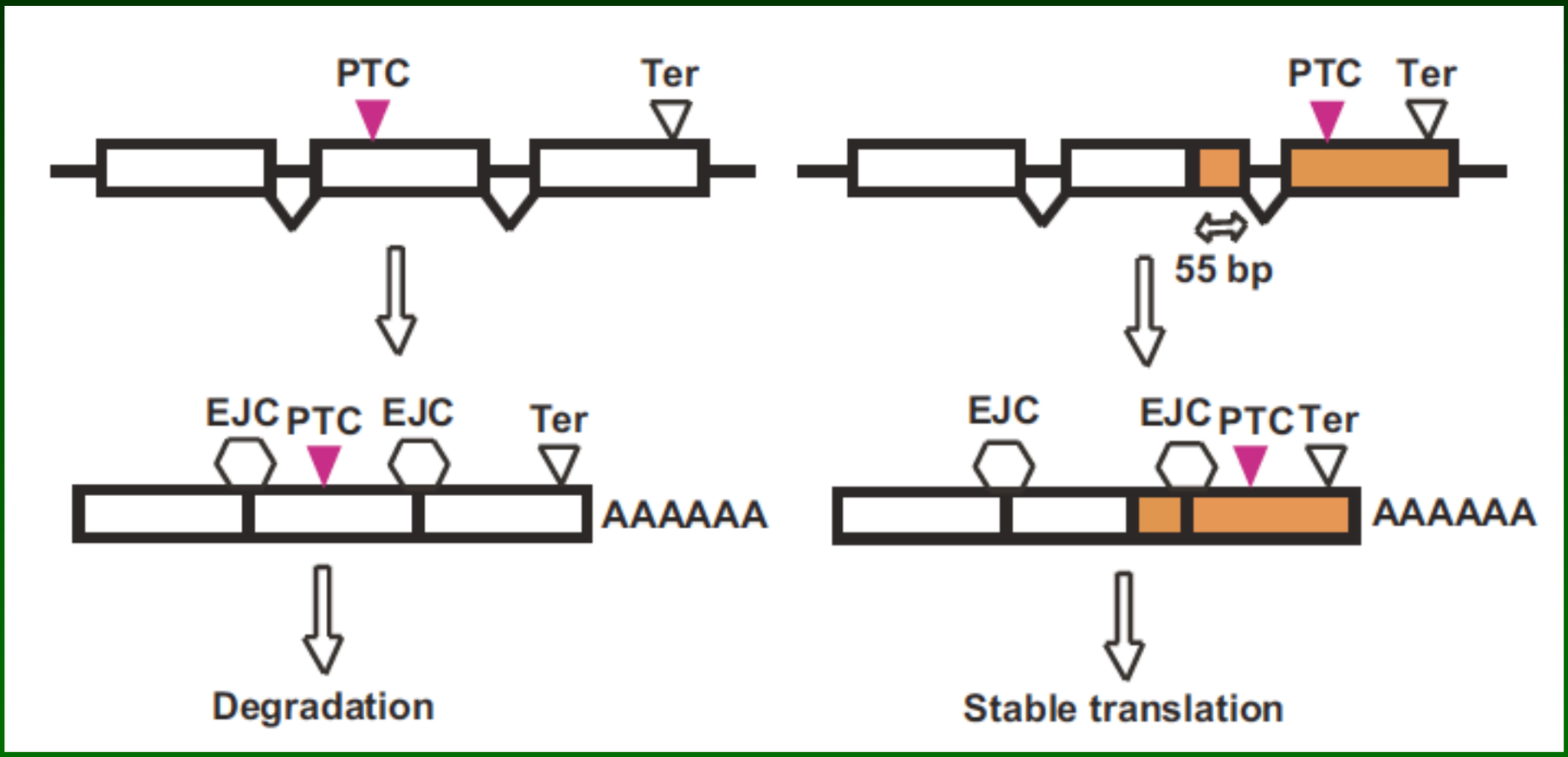
Cartegni (2002) Nat Rev Genet 3:285-298



SMA: masking ISS used in therapy

NMD

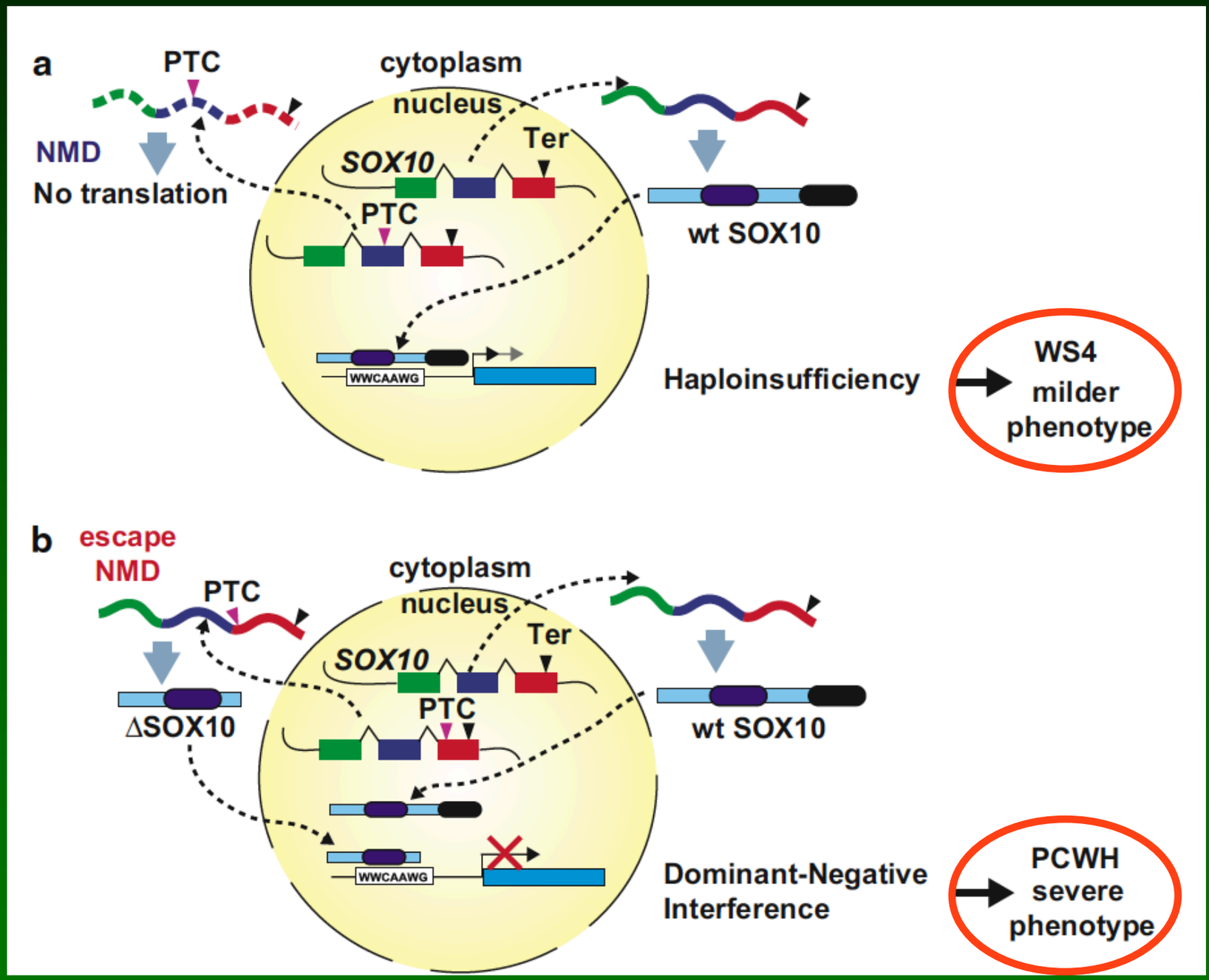
nonsense mediated mRNA decay



EJC = exon junction complex
PTC = premature termination codon
Ter = termination codon

Khajavi (2006) Eur j Hum Genet 14:1074-1081

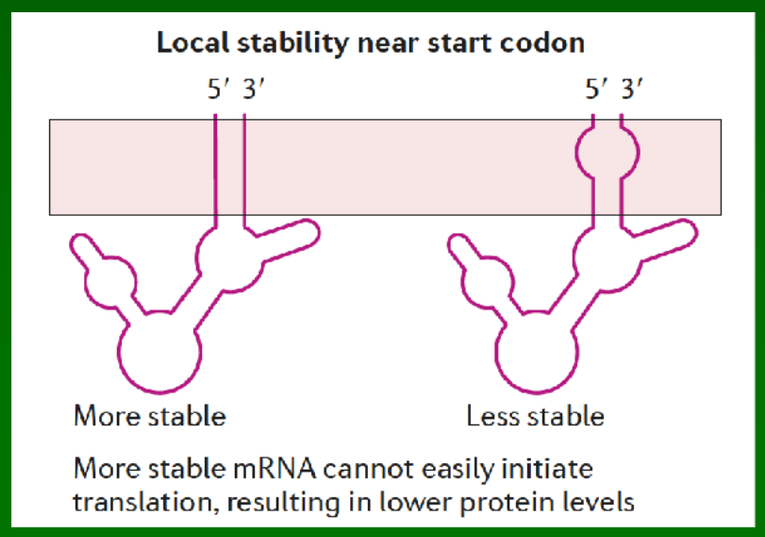
NMD

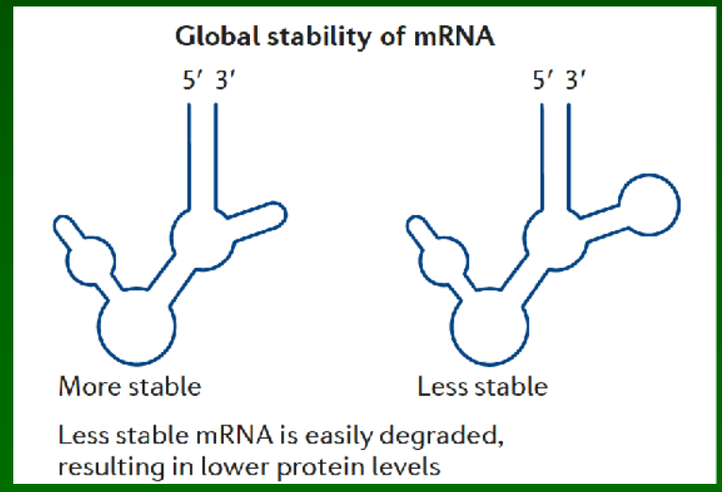


Khajavi (2006) Eur j Hum Genet 14:1074-1081

mRNA stability

- **single stranded RNA unstable**
  - ***quickly degraded***
  - ***nonsense mediated decay***
  - ***protected by cap, polyA tail, folding***
- **variants will affect stability**
  - ***...and binding of RNAs and proteins***





Sauna & Kimchi-Sarfaty (2011) Nat Rev Genet 12:683-691

RNA variants

- **change of encoded protein**
- **affecting processing**
  splicing, capping, polyA addition
- **change in stability**
  less stable (NMD) / more stable
  folding (hairpin structures)
- **change in interactions**
  binding other molecules
  miRNA, proteins
  change in cellular localisation
  altered traffickin, subcellular localisation

RNA...

- **amount**
  too much / too little
  stability (protein turnover)
- **timing**
  expression at wrong time (development)
  in reaction to wrong stimulus
- **place**
  wrong tissue

Dosage

- **dosage-sensitive genes**
  too much/little protein is deleterious
    incl. at wrong time, wrong place

- **dosage-insensitive genes**
  missing/extra copy not deleterious
    while variants in a copy are

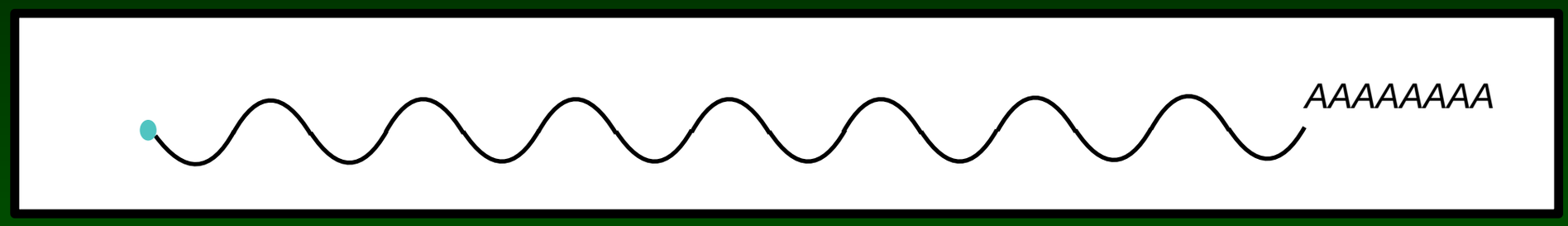
ClinGen is working on a list of proven links

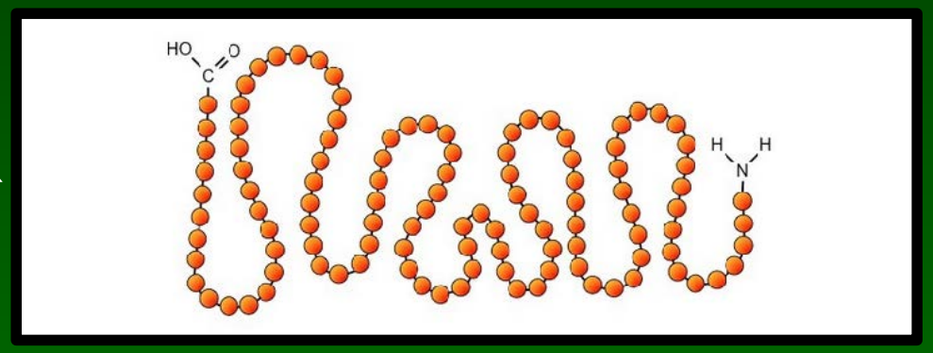
- **dosage**
  deletion / duplication
    deletion often more deleterious

- **partial deletion / duplication**
  intragenic
  incl. 5' or 3' end (gene / transcript / translation)
    leaves one normal copy

Translation

mRNA

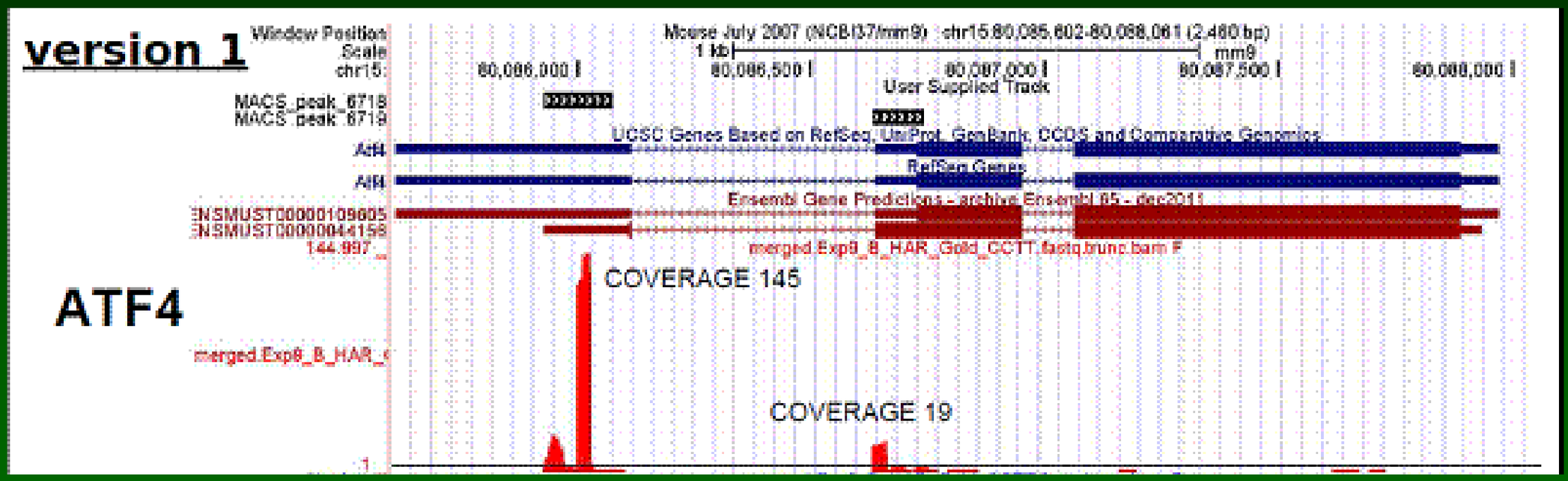




translation initiation site (start codon)
Kozak sequence
translation termination site (stop codon)
uORF
codon usage > translation speed > protein folding
RNA/protein binding > stability, amount of protein
…

uORFs

untranslated OpenReadingFrame



uORFs translation start (ATG) not annotated

RNA assays

- **expression cloning**
  clone exon(s) in splice construct
  transfect and express (cell line)
  analyse splice products generated

- **NOTE**
  artificial situation
  not complete gene/intron
  expression in other cell type (tissue specificity)

TEMPERATURE

- **RNA of patient preferable**
  cell line, biopsy, IPS cells, ...

RNA > protein

- **gene expression profiling**

  (m)RNA ≠ protein
  extensive regulation of translation
  stored RNA

- **to confirm RNA results**

  Western blot, stain cells (IHC), …
  proteomics (mass spec)

  ribosome profiling
  bound = translated

Protein variant

- **p.Phe159Leu**

  on DNA

  c.477C>A
  c.477C>G
  c.479T>C

in vitro test of p.Phe159

are all three variants identical ?

Protein variant

- **p.Phe159Leu**

  on DNA

  c.477C>A
  c.477C>G
  c.479T>C

in vitro test of p.Phe159

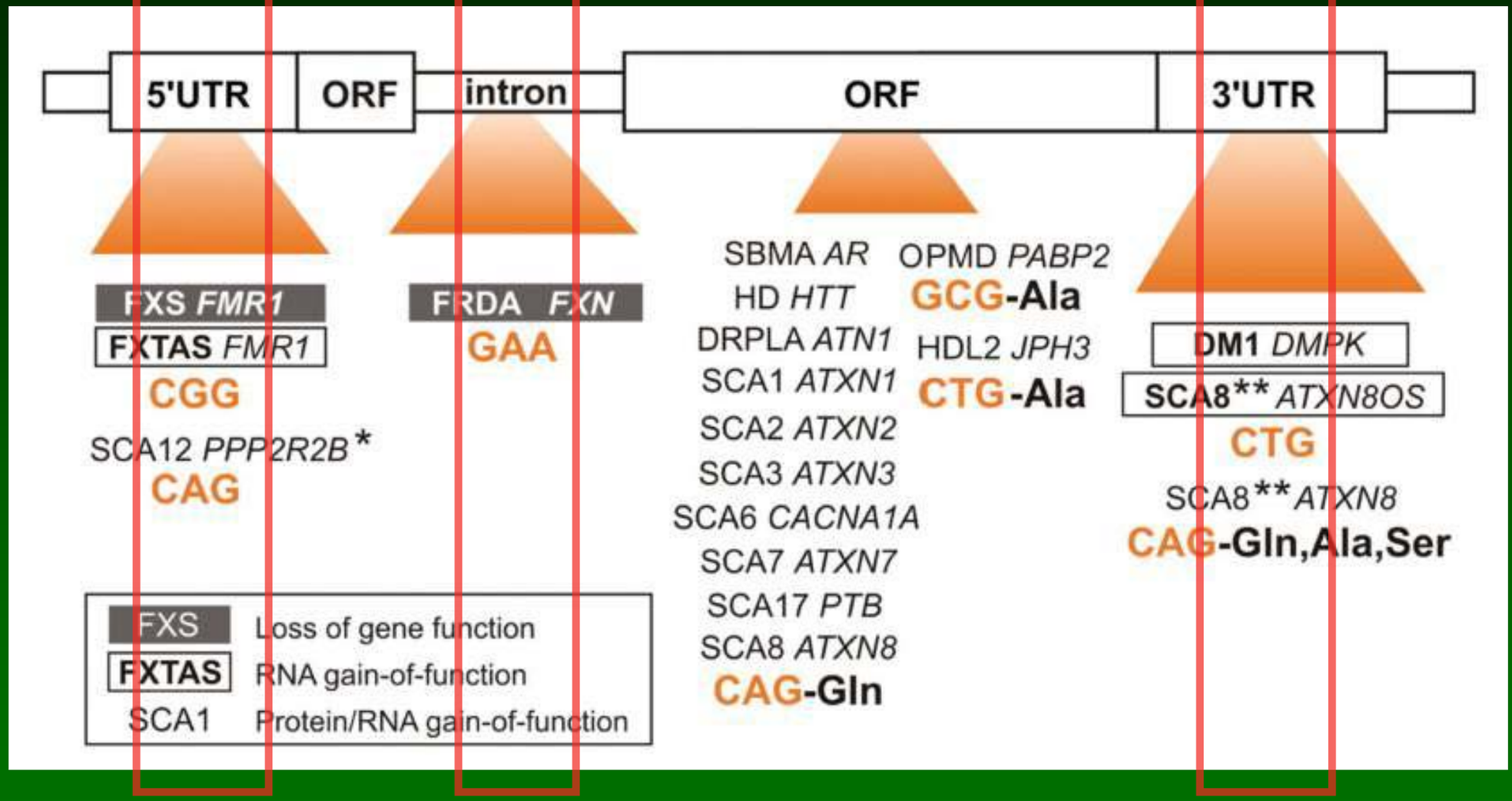
are all three variants identical ?

NO

check potential consequences on RNA
RNA sequence changes affect
stability RNA, RNA folding, binding motifs
codon changes > influence translation speed

be careful with classification of one variant based on other report giving same AA substitution

Repeat expansions



not all consequences understood,
nor the size range

Fusion RNA

- **from specific rearrangements**
  translocation (leukemia)
- **consider when 3' end gene missing**
  transcription needs to end somewhere
  deletion might be in upstream gene (EPCAM)
- **in general deleterious consequences**

RNA in practice

RNA +5 splice site

Splicing mutations in DMD/BMD detected by RT-PCR/PTT: detection of a 19AA insertion in the cysteine rich domain of dystrophin compatible with BMD

J Med Genet 1996;33:935–939

Pauline A M Roest, Mattie Bout, Astric C van der Tuijn, Ieke B Ginjaar, Egbert Bakker, Frans B L Hogervorst, Gert Jan B van Ommen, Johan T den Dunnen

proof of effect

in-frame insertion
> BMD

culture cells in with cycloheximide
(inhibit nonsense-mediated mRNA decay)

X-linked TOD

- **Terminal Osseous Dysplasia**
  pigmentary anomalies skin
  skeletal abnormalities limbs
  recurring digital fibromatosis childhood

- **X-linked** ***(Xq25-ter)***
  dominant
    male lethal
    female skewed X_i

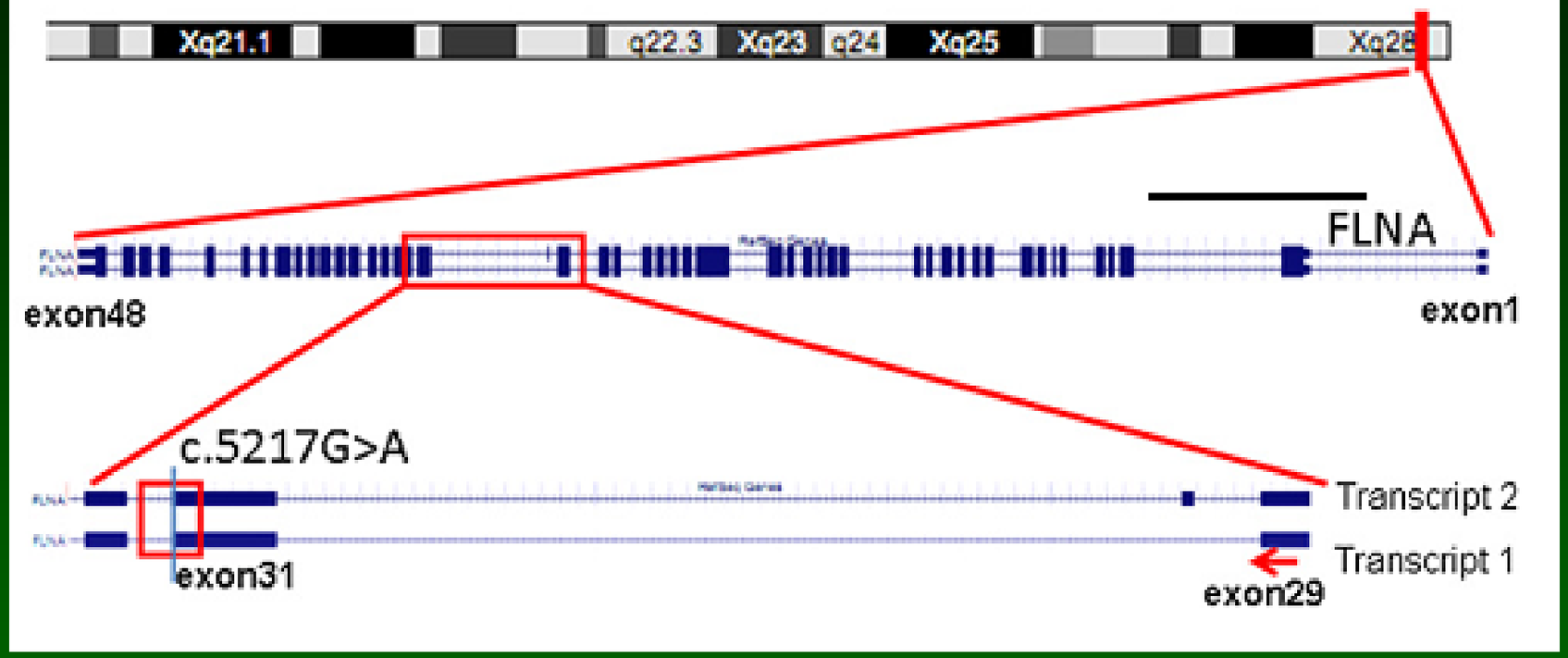
American Journal of Medical Genetics 94:91–101 (2000)

New Syndrome?

Recurrent Digital Fibroma, Focal Dermal Hypoplasia, and Limb Malformations

M.H. Breuning,[1]* A.P. Oranje,[2] R.A.Th.M. Langemeijer,[3] S.E.R. Hovius,[4] A.F.M. Diepstraten,[5] J.C. den Hollander,[6] N. Baumgartner,[7] J.R. Dwek,[8] A. Sommer,[9] and H. Toriello[7]

TOD X-exome$_2$

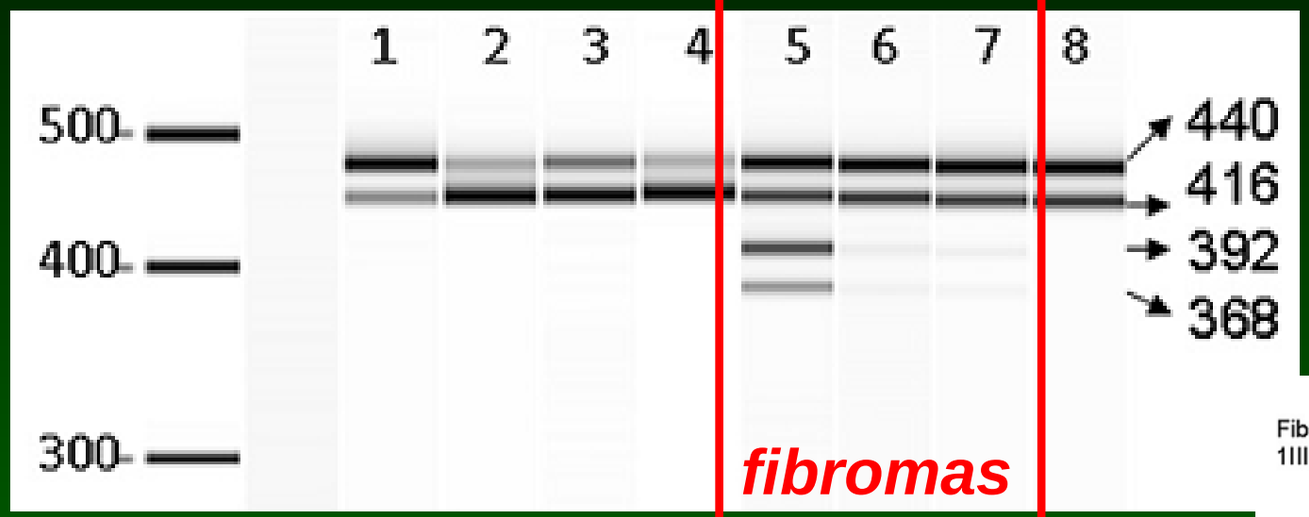


- **variant last nucleotide exon**
  alters splicing !?

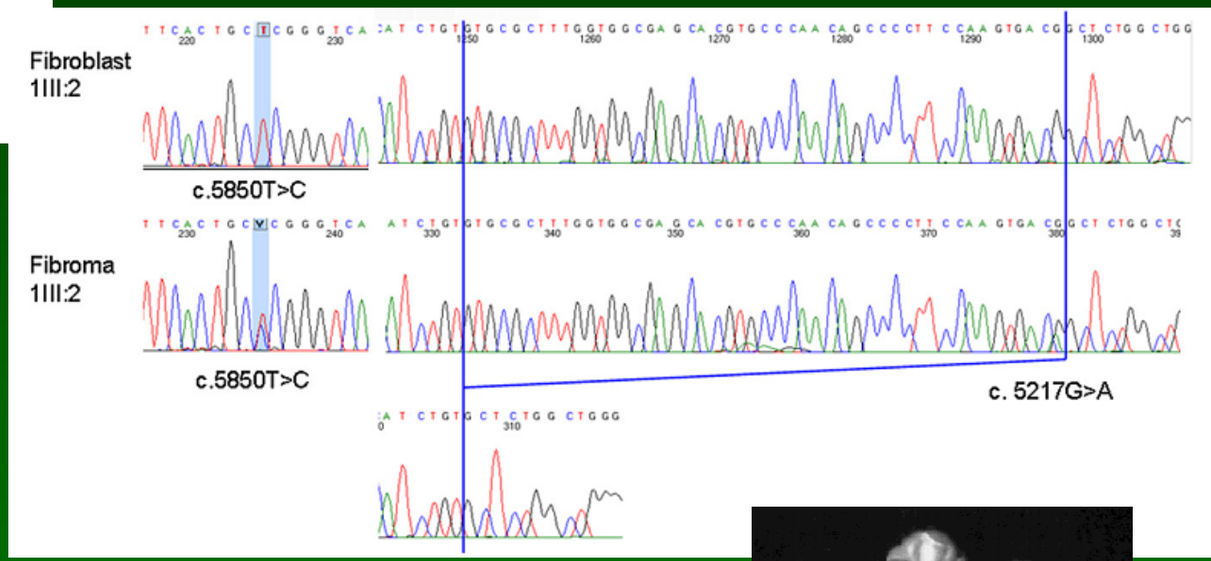
Sun et al. 2010
Am.J.Hum.Genet. 87: 146

- **RNA expression**
  cultured cells / blood
    100% X_i , only normal allele expressed
  X_i "affected chromosome"

TOD X-exome$_3$



RNA



- **archived fibroma tissue**
  (15 year old)
  both alleles expressed
  activated cryptic exonic splice site

Sun et al. 2010 Am.J.Hum.Genet. 87: 146

Aarskog-Scott syndrome

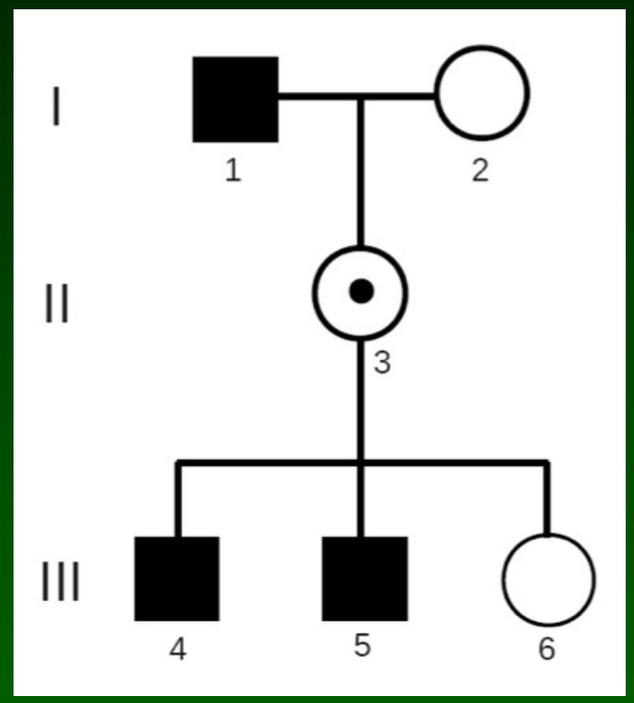
- **Aarskog-Scott syndrome**
  AAS (OMIM305400)
  faciogenital dysplasia
  short proportionate stature, short limbs, broad hands/feet, genital hypoplasia, facial dysmorphisms

- **X-linked recessive form**
  FGD1 gene
  Xp11.21

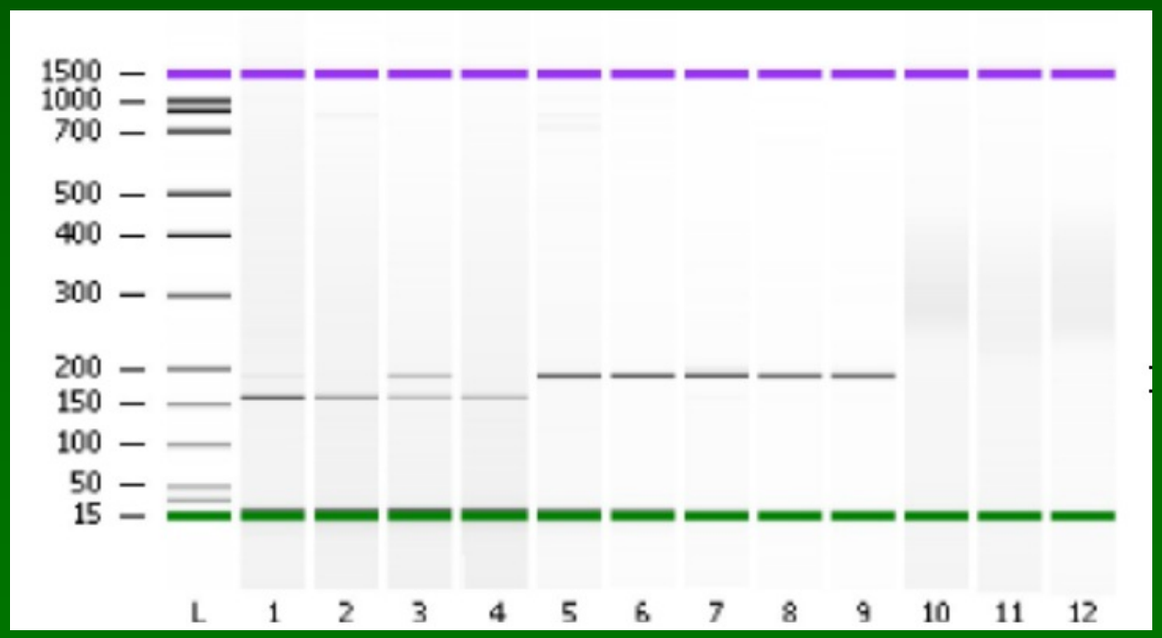
- **other forms**
  autosomal dominant & recessive genes involved ?

Aarskog-Scott syndrome

- **FGD1 gene variant**
  intron -35delA
  predicted branch site



- **RNA analysis**
  expressed in blood / fibroblasts



P P C P controls

©Yu Sun
Emmelien Aten

Aarskog-Scott syndrome

- **why FGD1 variant missed ?**
  - *primer on variant site*
    - *not standard to screen to -50*
- **exome capture**
  - *lower coverage into intron*
  - *variant filtering to -10*
    - *many additional variants*
    - *difficult to confirm*
- **few branch site variants**
  - *rare, easily missed,*
  - *difficult to proof*


BRIEF REPORT

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

Exome Sequencing Identifies A Branch Point Variant in Aarskog–Scott Syndrome

Emmelien Aten,[1†] Yu Sun,[1†] Rowida Almomani,[1] Gijs W.E. Santen,[1] Tobias Messemaker,[1] Saskia M. Maas,[2] Martijn H. Breuning,[1] and Johan T. den Dunnen[1*]


LOVD

Mental Retardation database

FYVE, RhoGEF and PH domain containing 1 (FGD1)

Curators: Johan den Dunnen and Curator vacancy

LOVD v.2.0 Build 31 [Current LOVD status]

Welcome, Johan den Dunnen

Your account | Log out

Home | Variants | Submitters | Submit | Configuration | Setup | Documentation

FGD1 homepage | Switch gene

LOVD Gene homepage

General information	
Gene name	FYVE, RhoGEF and PH domain containing 1
Gene symbol	FGD1
Chromosome Location	Xp11.22
Database location	www.LOVD.nl/MR
Curator	Johan den Dunnen and Curator vacancy
PubMed references	View all (unique) PubMed references in the FGD1 database
Date of creation	March 05, 2009
Last update	June 08, 2011
Version	FGD1 110608
Add sequence variant	Submit a sequence variant
First time submitters	Register here
Reference sequence	Genomic reference sequence for describing sequence variants
GenBank reference	FGD1_NG_008054.1.gb
Exon/intron information	Exon/intron information table
Total number of unique DNA variants reported	34
Total number of individuals with variant(s)	41
Total number of variants reported	41
Subscribe to updates of this gene	
NOTE	The work leading to the establishment of these LSDDs was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 200754 - the GEN2PHEN project.

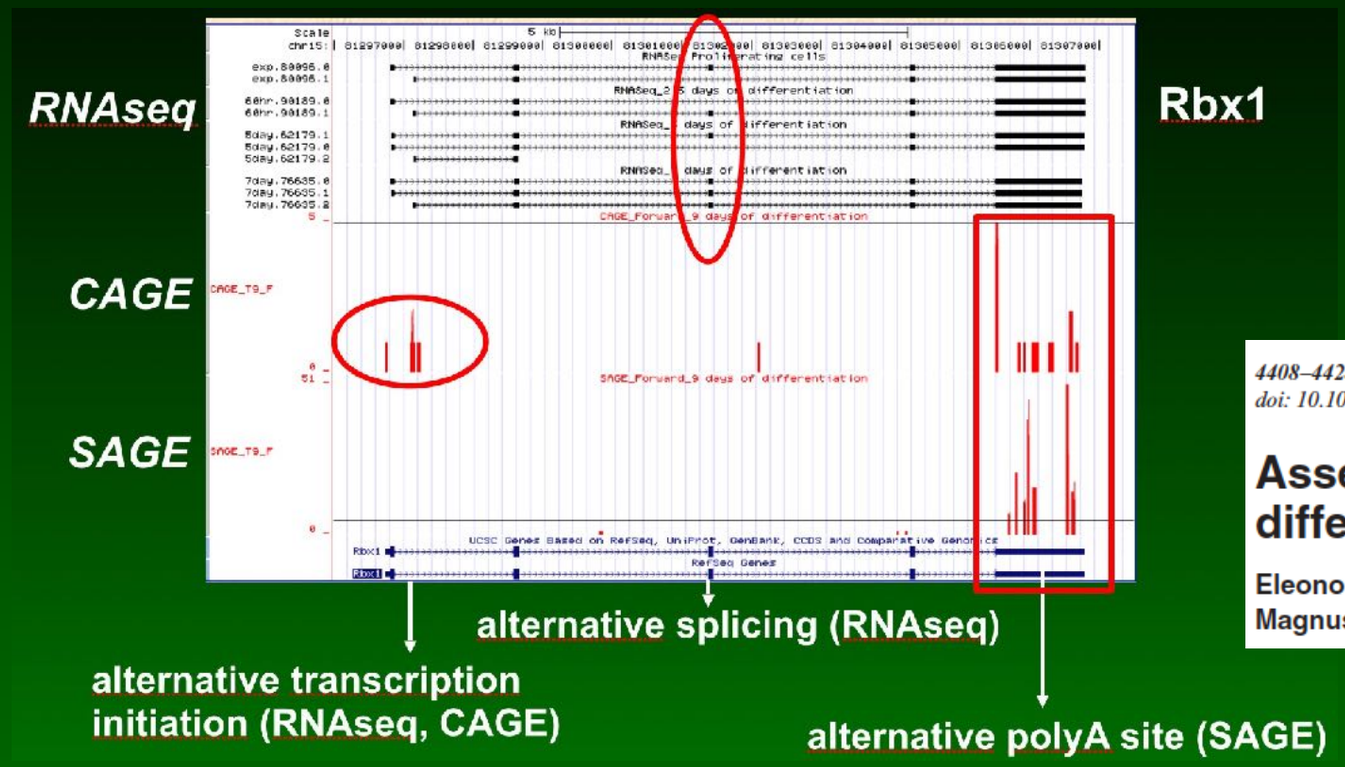
www.LOVD.nl/FGD1

exome performed, RNA analysis would be simpler ...& much cheaper

Exome + RNAseq

- **next to exome perform RNAseq**
  - ***5 - 50 million reads***
- **analyse RNA**
  - ***variants & allelic expression differences***
  - ***splice variants***
- **expression profiling** *(blood)*
  - ***compare to reference database***
    - ***5,000 control samples***
  - ***expression changes***
    - ***individual genes***
    - ***disturbed pathways***

Gene annotation



detailed RNA studies reveal many new transcripts

4408–4428 Nucleic Acids Research, 2015, Vol. 43, No. 9 Published online 14 April 2015
doi: 10.1093/nar/gkv281

Assessing the translational landscape of myogenic differentiation by ribosome profiling

Eleonora de Klerk[1], Ivo F.A.C. Fokkema[1,†], Klaske A.M.H. Thiadens[2,†], Jelle J. Goeman[3], Magnus Palmblad[4], Johan T. den Dunnen[1], Marieke von Lindern[2] and Peter A.C. 't Hoen[1,*]

Published online 6 July 2012 Nucleic Acids Research, 2012, Vol. 40, No. 18 9089–9101
doi:10.1093/nar/gks655

Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation

Eleonora de Klerk[1], Andrea Venema[1], S. Yahya Anvar[1], Jelle J. Goeman[2], OuHua Hu[1], Capucine Trollet[3,4], George Dickson[3], Johan T. den Dunnen[1], Silvère M. van der Maarel[1], Vered Raz[1,*] and Peter A. C. 't Hoen[1,*]

Cell. Mol. Life Sci. (2014) 71:3537–3551
DOI 10.1007/s00018-014-1637-9 Cellular and Molecular Life Sciences

REVIEW

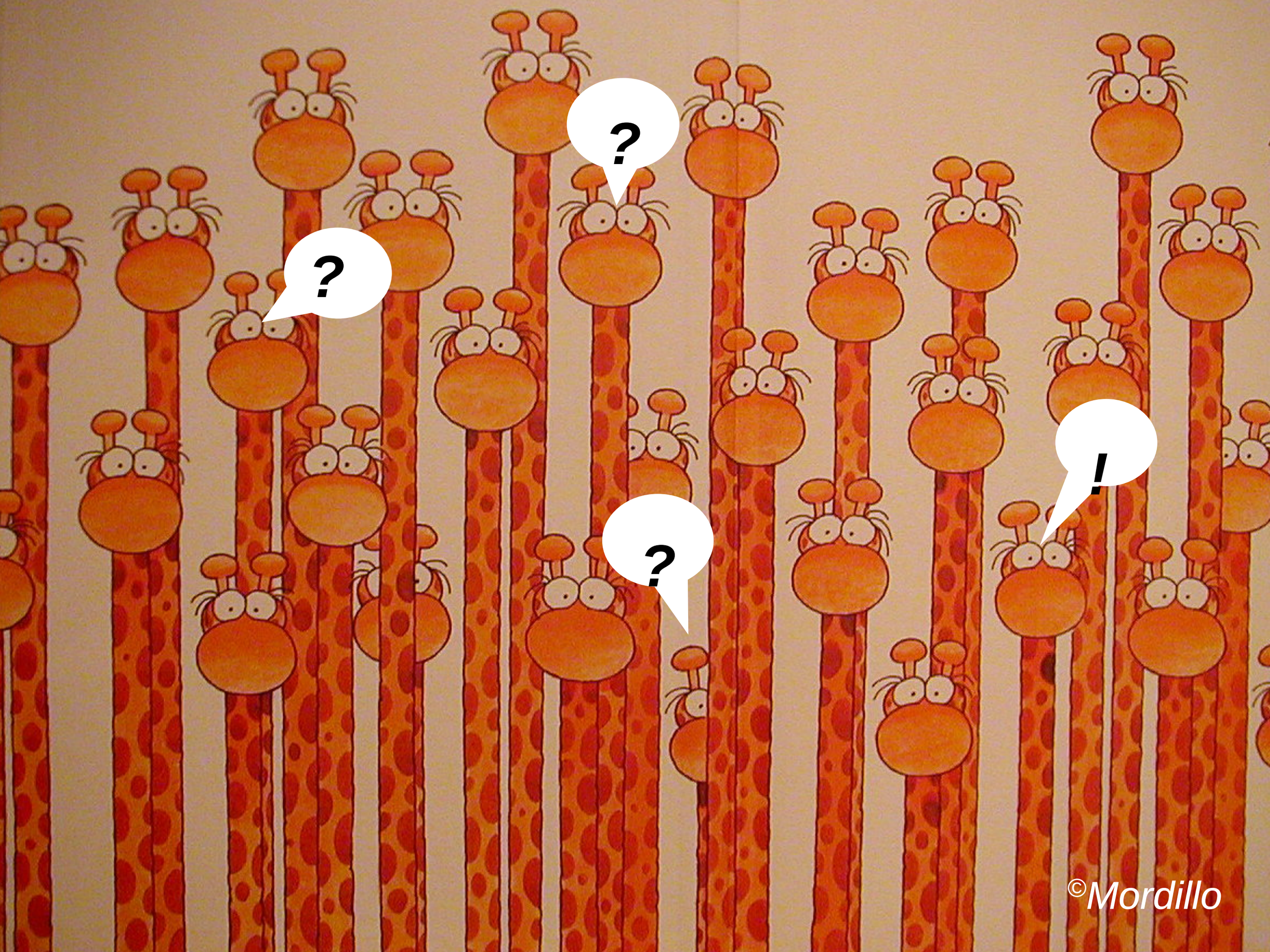
RNA sequencing: from tag-based profiling to resolving complete transcript structure

Eleonora de Klerk · Johan T. den Dunnen · Peter A. C. 't Hoen

Published online 7 July 2010 Nucleic Acids Research, 2010, Vol. 38, No. 16 e165
doi:10.1093/nar/gkq602

Tissue-specific transcript annotation and expression profiling with complementary next-generation sequencing technologies

Matthew S. Hestand[1,2], Andreas Klingenhoff[3], Matthias Scherf[3], Yavuz Ariyurek[2], Yolande Ramos[4], Wilbert van Workum[5], Makoto Suzuki[6], Thomas Werner[3], Gert-Jan B. van Ommen[1], Johan T. den Dunnen[1,2], Matthias Harbers[6] and Peter A.C. 't Hoen[1,*]


?
?
?
!
©Mordillo

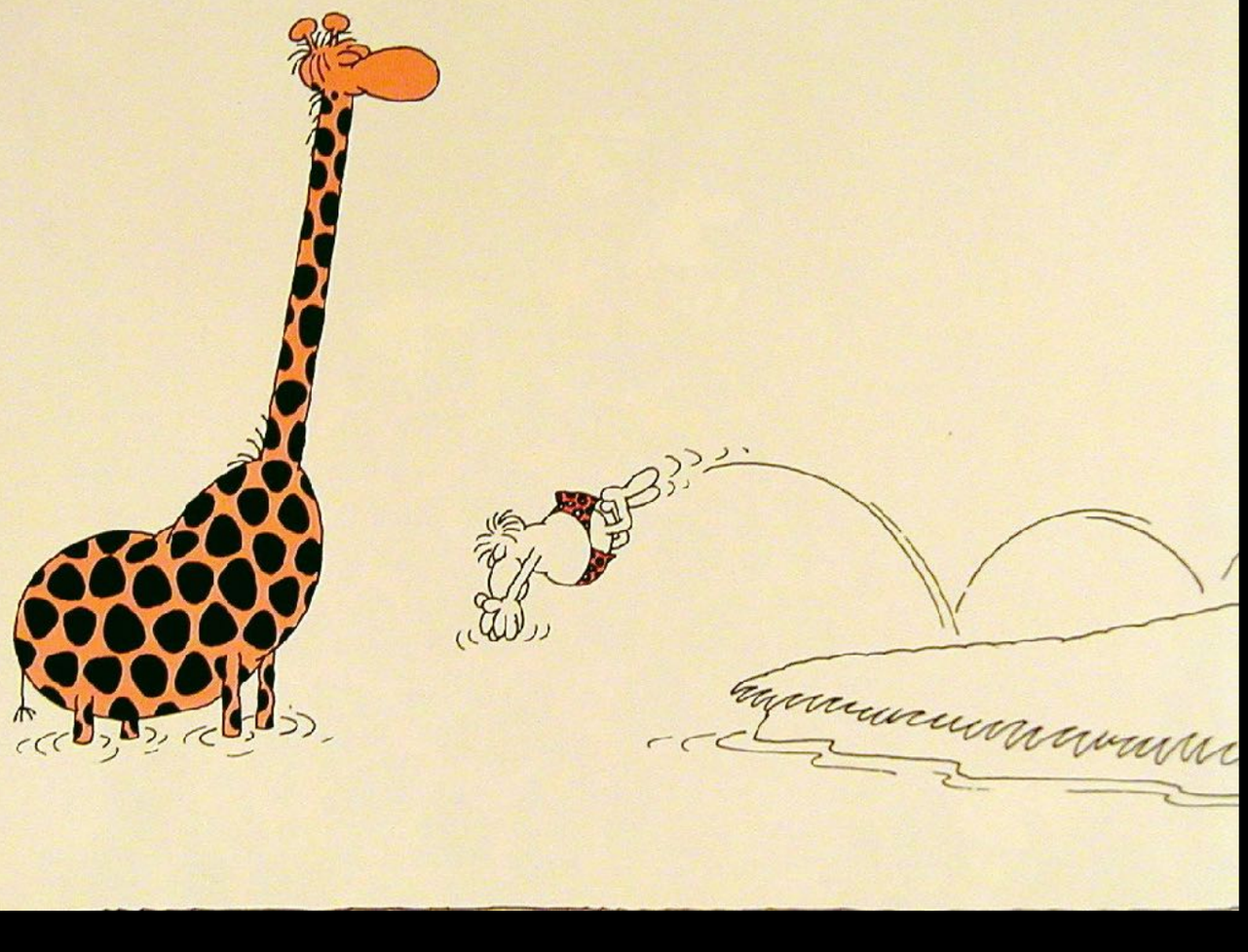
Acknowledgement

Presentation prepared by:

Johan den Dunnen

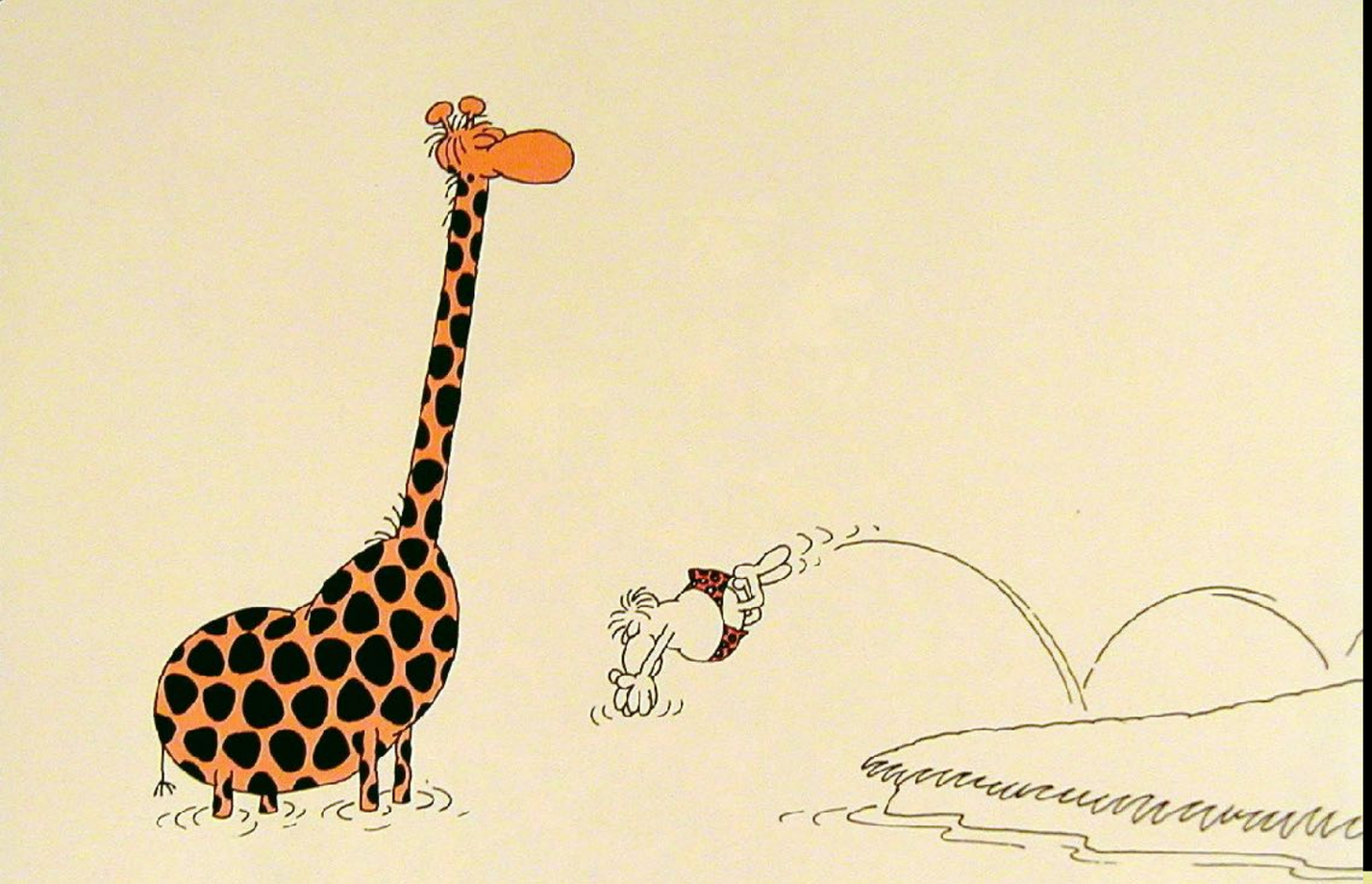
Human Genetics & Clinical Genetics
Leiden University Medical Center
Leiden, Nederland

date: April 2019



variant classification

without RNA analysis



variant classification

without RNA analysis