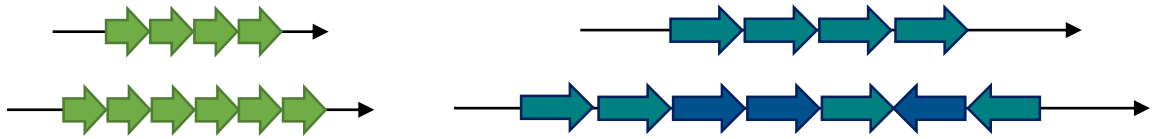
Repeat expansions in neurological disorders

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken
Universitätsmedizin Essen
Universitätsklinikum

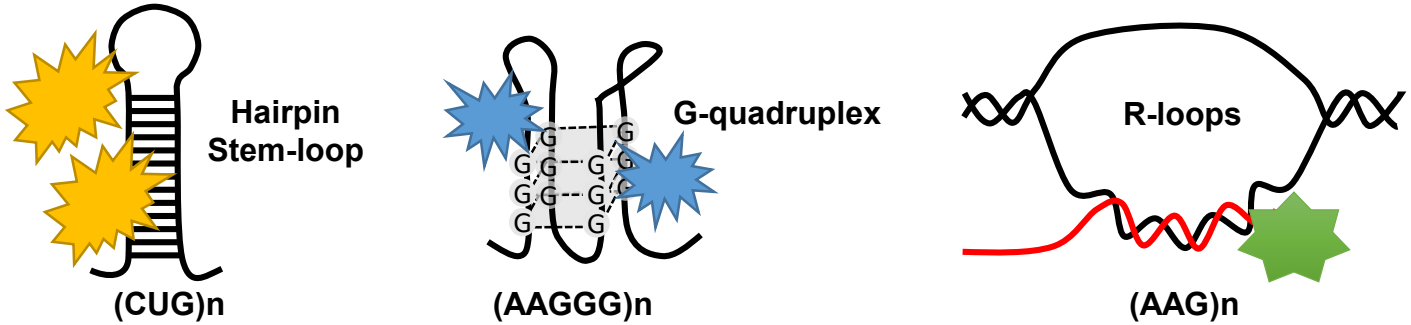
Christel Depienne
Institut für Humangenetik
Universitätsklinikum Essen

Tandem repeats

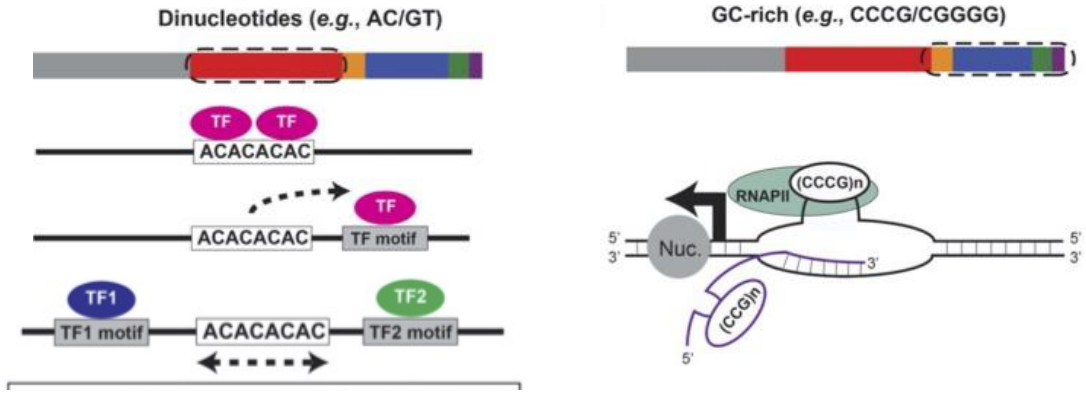
- Microsatellite (Short Tandem Repeat) 2 - 6 bp
- Minisatellite (VNTR) 7 - 100 bp



- STRs have one of the highest mutational rates
  - STRs 5 x 10^{-5} / microsatellite / generation
  - SNVs 1-2 × 10^{-8} / bp / generation
- They form stable secondary structures



- They can regulate gene expression



Kristmundsdottir et al. 2023; Jónsson et al. 2017; Malik et al. 2021; Fotsing et al. 2019

Tandem repeat expansions

(CAG)n

<36

Huntington's disease, SCA1, 2, 3, 6, 7, 17

≥36

(CGG)n

≤50

>50

>200

polyG

RAN translation

FXTAS

DNA hypermethylation

CH_3 CH_3 CH_3 CH_3 CH_3 CH_3 CH_3 CH_3 CH_3 CH_3

Fragile X syndrome

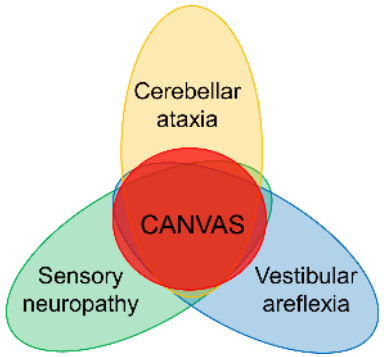
HIGHLY VARIABLE

(AAAAG)n (AAAGG)n (AAGAG)n

>200-400

(**AAGGG**)exp

CANVAS
(biallelic)



VARIABLE

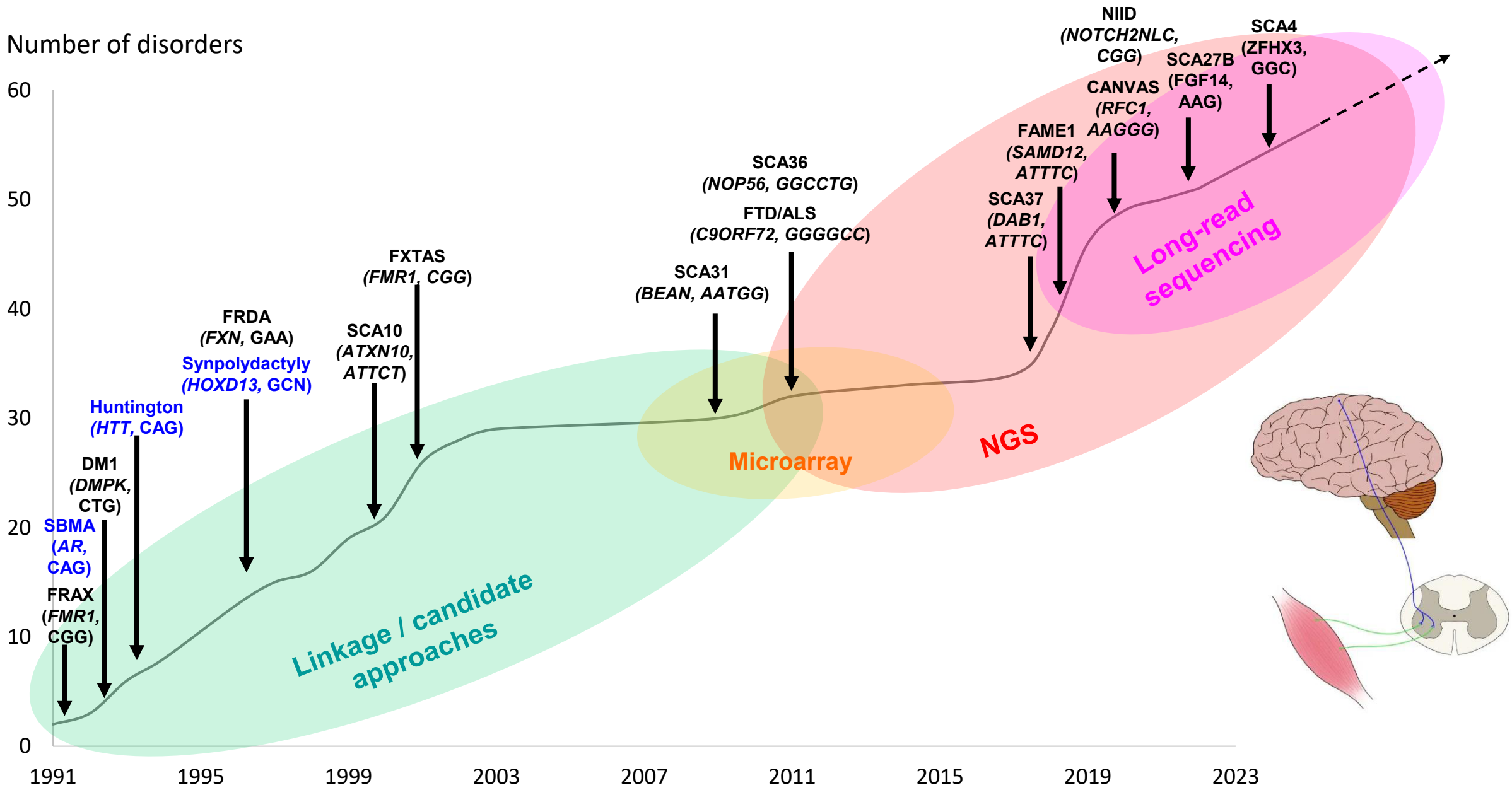
REPEAT INSERTION

(TTTTA)n (TTTTA)n

(TTTTA/**TTTCA**)exp

Familial Adult Myoclonic Epilepsy (FAME), SCA37, SCA10, SCA31

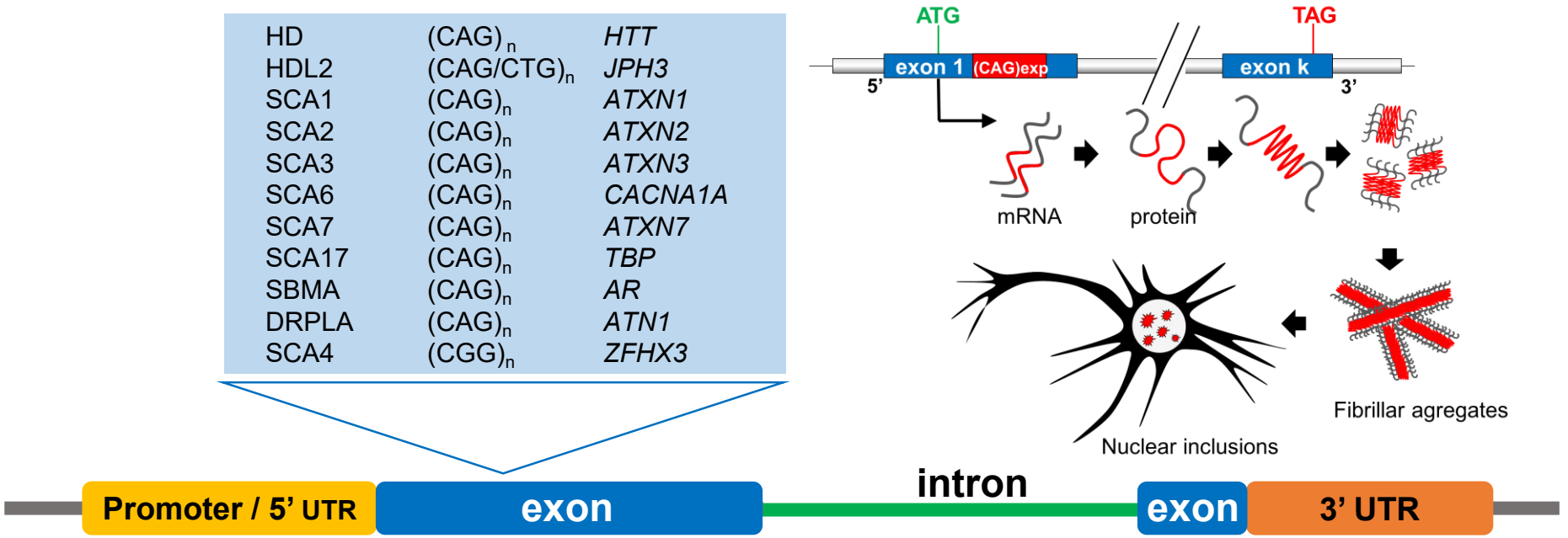
Repeat Expansion discovery



Adapted from Depienne & Mandel, 2021

Diversity of pathogenic Repeat Expansions & Mechanisms

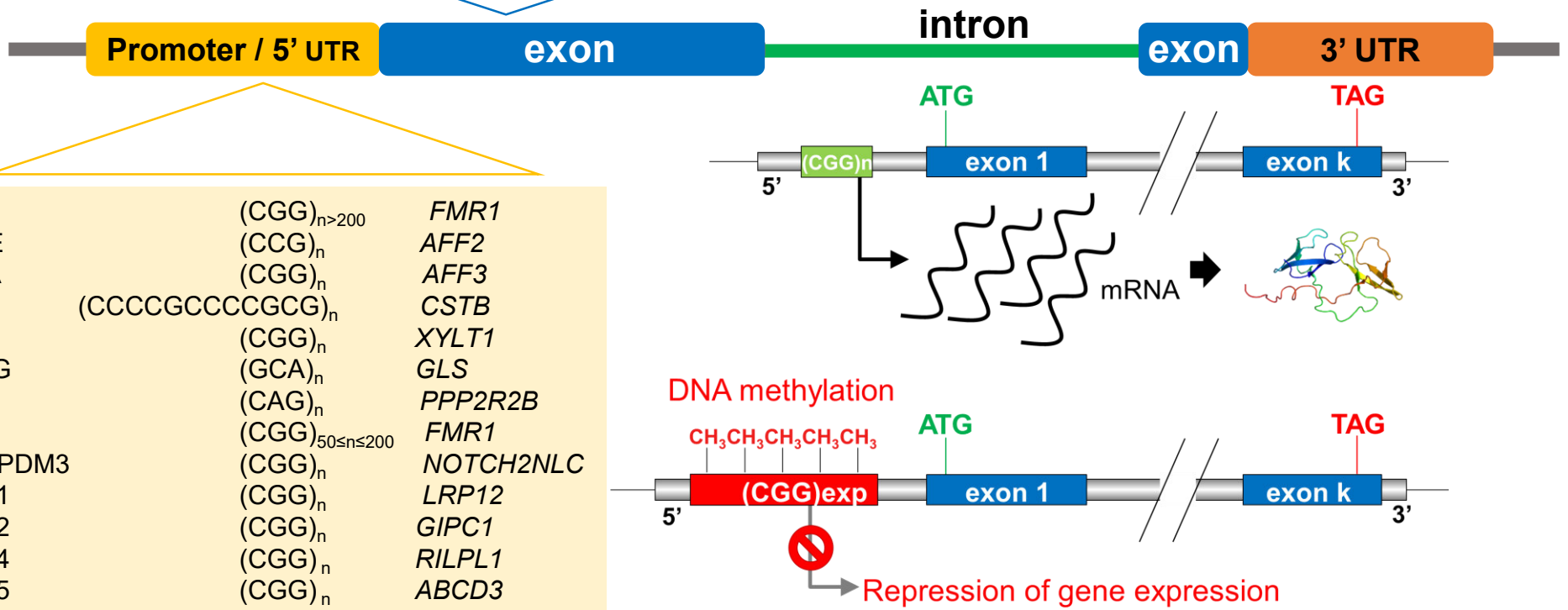
HD	$(CAG)_n$	*HTT*
HDL2	$(CAG/CTG)_n$	*JPH3*
SCA1	$(CAG)_n$	*ATXN1*
SCA2	$(CAG)_n$	*ATXN2*
SCA3	$(CAG)_n$	*ATXN3*
SCA6	$(CAG)_n$	*CACNA1A*
SCA7	$(CAG)_n$	*ATXN7*
SCA17	$(CAG)_n$	*TBP*
SBMA	$(CAG)_n$	*AR*
DRPLA	$(CAG)_n$	*ATN1*
SCA4	$(CGG)_n$	*ZFHX3*



Adapted from Depienne & Mandel, 2021

Diversity of pathogenic Repeat Expansions & Mechanisms

HD	$(CAG)_n$	*HTT*
HDL2	$(CAG/CTG)_n$	*JPH3*
SCA1	$(CAG)_n$	*ATXN1*
SCA2	$(CAG)_n$	*ATXN2*
SCA3	$(CAG)_n$	*ATXN3*
SCA6	$(CAG)_n$	*CACNA1A*
SCA7	$(CAG)_n$	*ATXN7*
SCA17	$(CAG)_n$	*TBP*
SBMA	$(CAG)_n$	*AR*
DRPLA	$(CAG)_n$	*ATN1*
SCA4	$(CGG)_n$	*ZFHX3*



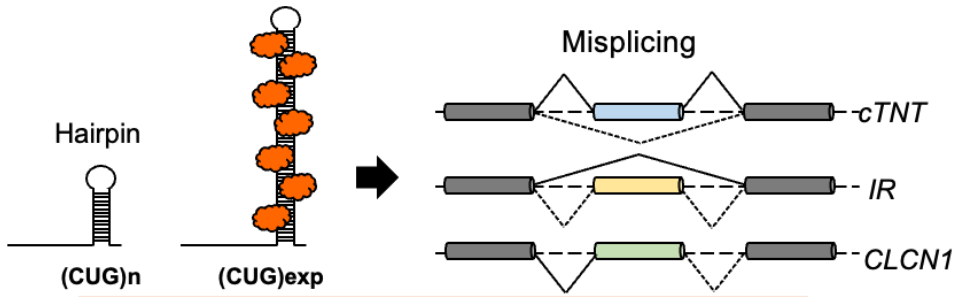
FXS	$(CGG)_{n>200}$	*FMR1*
FRAXE	$(CCG)_n$	*AFF2*
FRA2A	$(CGG)_n$	*AFF3*
EPM1	$(CCCCGCCCCGCG)_n$	*CSTB*
BSS	$(CGG)_n$	*XYLT1*
GDPAG	$(GCA)_n$	*GLS*
SCA12	$(CAG)_n$	*PPP2R2B*
FXTAS	$(CGG)_{50\leq n\leq 200}$	*FMR1*
NIID/OPDM3	$(CGG)_n$	*NOTCH2NLC*
OPDM1	$(CGG)_n$	*LRP12*
OPDM2	$(CGG)_n$	*GIPC1*
OPDM4	$(CGG)_n$	*RILPL1*
OPDM5	$(CGG)_n$	*ABCD3*

Adapted from Depienne & Mandel, 2021

Diversity of pathogenic Repeat Expansions & Mechanisms

HD	$(CAG)_n$	*HTT*
HDL2	$(CAG/CTG)_n$	*JPH3*
SCA1	$(CAG)_n$	*ATXN1*
SCA2	$(CAG)_n$	*ATXN2*
SCA3	$(CAG)_n$	*ATXN3*
SCA6	$(CAG)_n$	*CACNA1A*
SCA7	$(CAG)_n$	*ATXN7*
SCA17	$(CAG)_n$	*TBP*
SBMA	$(CAG)_n$	*AR*
DRPLA	$(CAG)_n$	*ATN1*
SCA4	$(CGG)_n$	*ZFHX3*

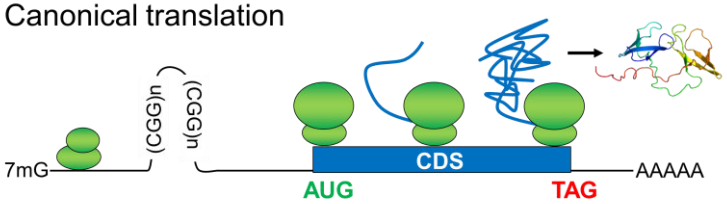
MBNL protein sequestration



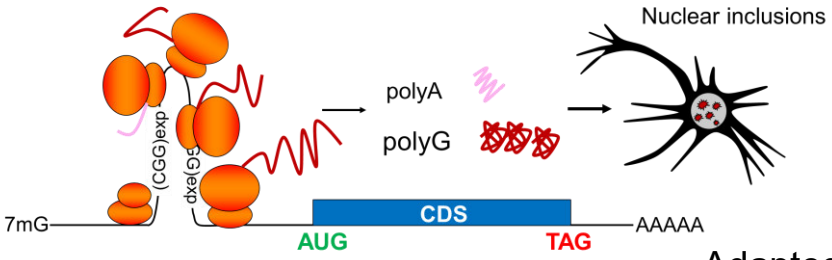
DM1	$(CTG)_n$	*DMPK*
SCA8	$(CTG/CAG)_n$	*ATXN8OS/ATXN8*

FXS	$(CGG)_{n>200}$	*FMR1*
FRAXE	$(CCG)_n$	*AFF2*
FRA2A	$(CGG)_n$	*AFF3*
EPM1	$(CCCCGCCCCGCG)_n$	*CSTB*
BSS	$(CGG)_n$	*XYLT1*
GDPAG	$(GCA)_n$	*GLS*
SCA12	$(CAG)_n$	*PPP2R2B*
FXTAS	$(CGG)_{50\leq n\leq 200}$	*FMR1*
NIID/OPDM3	$(CGG)_n$	*NOTCH2NLC*
OPDM1	$(CGG)_n$	*LRP12*
OPDM2	$(CGG)_n$	*GIPC1*
OPDM4	$(CGG)_n$	*RILPL1*
OPDM5	$(CGG)_n$	*ABCD3*

Canonical translation



RAN translation



Adapted from Depienne & Mandel, 2021

Diversity of pathogenic Repeat Expansions & Mechanisms

HD	$(CAG)_n$	*HTT*
HDL2	$(CAG/CTG)_n$	*JPH3*
SCA1	$(CAG)_n$	*ATXN1*
SCA2	$(CAG)_n$	*ATXN2*
SCA3	$(CAG)_n$	*ATXN3*
SCA6	$(CAG)_n$	*CACNA1A*
SCA7	$(CAG)_n$	*ATXN7*
SCA17	$(CAG)_n$	*TBP*
SBMA	$(CAG)_n$	*AR*
DRPLA	$(CAG)_n$	*ATN1*
SCA4	$(CGG)_n$	*ZFHX3*

DM1	$(CTG)_n$	*DMPK*
SCA8	$(CTG/CAG)_n$	*ATXN8OS/ATXN8*

FXS	$(CGG)_{n>200}$	*FMR1*
FRAXE	$(CCG)_n$	*AFF2*
FRA2A	$(CGG)_n$	*AFF3*
EPM1	$(CCCCGCCCCGCG)_n$	*CSTB*
BSS	$(CGG)_n$	*XYLT1*
GDPAG	$(GCA)_n$	*GLS*
SCA12	$(CAG)_n$	*PPP2R2B*
FXTAS	$(CGG)_{50\leq n\leq 200}$	*FMR1*
NIID/OPDM3	$(CGG)_n$	*NOTCH2NLC*
OPDM1	$(CGG)_n$	*LRP12*
OPDM2	$(CGG)_n$	*GIPC1*
OPDM4	$(CGG)_n$	*RILPL1*
OPDM5	$(CGG)_n$	*ABCD3*

ALS/FTD	$(GGGGCC)_n$	*C9ORF72*
FECD3	$(CTG)_n$	*TCF4*
DM2	$(CAG)_n$	*CNBP*
FRDA	$(GAA)_n$	*FXN*
SCA27B	(GAA)n	*FGF14*
SCA10	$(ATTCT)_n$	*ATXN10*
SCA31	$(TGGAA)_n$	*BEAN*
SCA36	$(GGCCTG)_n$	*NOP56*
CANVAS	$(AAGGG)_n$	*RFC1*
SCA37	$(ATTTT/ATTTC)_n$	*DAB1*
FAME1/6/7	$(ATTTT/ATTTC)_n$	*SAMD12, TNRC6A, RAPGEF2*
FAME2/3/4/8	$(ATTTT/ATTTC)_n$	*STARD7, MARCHF6, YEATS2, RAI1*
XDP	$(CCCTCT)_n$	*TAF1* (SVA)

Adapted from Depienne & Mandel, 2021

Diversity of pathogenic Repeat Expansions & Mechanisms

HD	$(CAG)_n$	*HTT*
HDL2	$(CAG/CTG)_n$	*JPH3*
SCA1	$(CAG)_n$	*ATXN1*
SCA2	$(CAG)_n$	*ATXN2*
SCA3	$(CAG)_n$	*ATXN3*
SCA6	$(CAG)_n$	*CACNA1A*
SCA7	$(CAG)_n$	*ATXN7*
SCA17	$(CAG)_n$	*TBP*
SBMA	$(CAG)_n$	*AR*
DRPLA	$(CAG)_n$	*ATN1*
SCA4	$(CGG)_n$	*ZFHX3*

AD (autosomal dominant)
AR (autosomal recessive)
XL (X-linked)

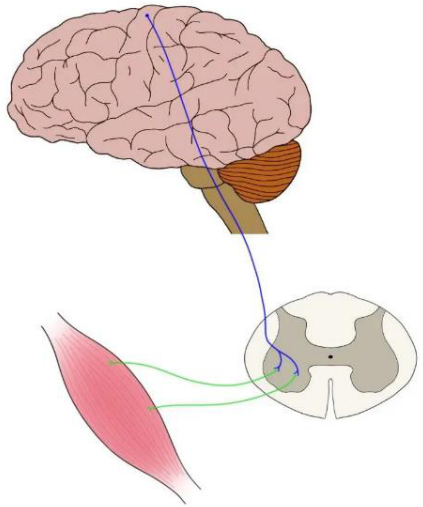
DM1	$(CTG)_n$	*DMPK*
SCA8	$(CTG/CAG)_n$	*ATXN8OS/ATXN8*

FXS	$(CGG)_{n>200}$	*FMR1*
FRAXE	$(CCG)_n$	*AFF2*
FRA2A	$(CGG)_n$	*AFF3*
EPM1	$(CCCCGCCCCGCG)_n$	*CSTB*
BSS	$(CGG)_n$	*XYLT1*
GDPAG	$(GCA)_n$	*GLS*
SCA12	$(CAG)_n$	*PPP2R2B*
FXTAS	$(CGG)_{50\leq n\leq 200}$	*FMR1*
NIID/OPDM3	$(CGG)_n$	*NOTCH2NLC*
OPDM1	$(CGG)_n$	*LRP12*
OPDM2	$(CGG)_n$	*GIPC1*
OPDM4	$(CGG)_n$	*RILPL1*
OPDM5	$(CGG)_n$	*ABCD3*

ALS/FTD	$(GGGGCC)_n$	*C9ORF72*
FECD3	$(CTG)_n$	*TCF4*
DM2	$(CAG)_n$	*CNBP*
FRDA	$(GAA)_n$	*FXN*
SCA27B	(GAA)n	*FGF14*
SCA10	$(ATTCT)_n$	*ATXN10*
SCA31	$(TGGAA)_n$	*BEAN*
SCA36	$(GGCCTG)_n$	*NOP56*
CANVAS	$(AAGGG)_n$	*RFC1*
SCA37	$(ATTTT/ATTTC)_n$	*DAB1*
FAME1/6/7	$(ATTTT/ATTTC)_n$	*SAMD12, TNRC6A, RAPGEF2*
FAME2/3/4/8	$(ATTTT/ATTTC)_n$	*STARD7, MARCHF6, YEATS2, RAI1*
XDP	$(CCCTCT)_n$	*TAF1* (SVA)

Adapted from Depienne & Mandel, 2021

Diversity of pathogenic Repeat Expansions & Mechanisms



HD	$(CAG)_n$	*HTT*
HDL2	$(CAG/CTG)_n$	*JPH3*
SCA1	**$(CAG)_n$**	***ATXN1***
SCA2	**$(CAG)_n$**	***ATXN2***
SCA3	**$(CAG)_n$**	***ATXN3***
SCA6	**$(CAG)_n$**	***CACNA1A***
SCA7	**$(CAG)_n$**	***ATXN7***
SCA17	**$(CAG)_n$**	***TBP***
SBMA	$(CAG)_n$	*AR*
DRPLA	$(CAG)_n$	*ATN1*
SCA4	**$(CGG)_n$**	***ZFHX3***

Half of known repeat expansions lead to spinocerebellar ataxia

DM1	$(CTG)_n$	*DMPK*
SCA8	**$(CTG/CAG)_n$**	***ATXN8OS/ATXN8***

Promoter / 5' UTR | **exon** | **intron** | **exon** | **3' UTR**

FXS	$(CGG)_{n>200}$	*FMR1*
FRAXE	$(CCG)_n$	*AFF2*
FRA2A	$(CGG)_n$	*AFF3*
EPM1	$(CCCCGCCCCGCG)_n$	*CSTB*
BSS	$(CGG)_n$	*XYLT1*
GDPAG	**$(GCA)_n$**	***GLS***
SCA12	**$(CAG)_n$**	***PPP2R2B***
FXTAS	**$(CGG)_{50\leq n\leq 200}$**	***FMR1***
NIID/OPDM3	$(CGG)_n$	*NOTCH2NLC*
OPDM1	$(CGG)_n$	*LRP12*
OPDM2	$(CGG)_n$	*GIPC1*
OPDM4	$(CGG)_n$	*RILPL1*
OPDM5	$(CGG)_n$	*ABCD3*

ALS/FTD	$(GGGGCC)_n$	*C9ORF72*
FECD3	$(CTG)_n$	*TCF4*
DM2	$(CAG)_n$	*CNBP*
FRDA	$(GAA)_n$	*FXN*
SCA27B	**(GAA)n**	***FGF14***
SCA10	**$(ATTCT)_n$**	***ATXN10***
SCA31	**$(TGGAA)_n$**	***BEAN***
SCA36	**$(GGCCTG)_n$**	***NOP56***
CANVAS	**$(AAGGG)_n$**	***RFC1***
SCA37	**$(ATTTT/ATTTC)_n$**	***DAB1***
FAME1/6/7	$(ATTTT/ATTTC)_n$	*SAMD12, TNRC6A, RAPGEF2*
FAME2/3/4/8	$(ATTTT/ATTTC)_n$	*STARD7, MARCHF6, YEATS2, RAI1*
XDP	$(CCCTCT)_n$	*TAF1* (SVA)

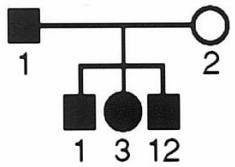
Adapted from Depienne & Mandel, 2021

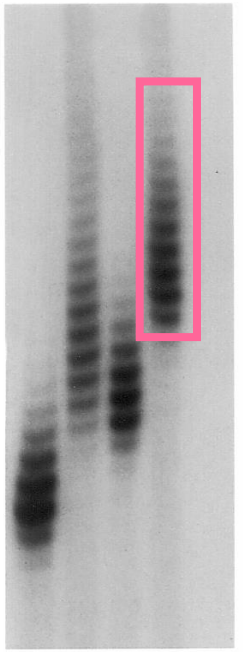
Characteristics of repeat expansions: instability

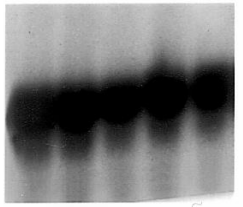
- Pure large expanded alleles are unstable (somatic mosaicism)
- Instability during transmission
- Parent of origin effect
- Mosaicism in the brain

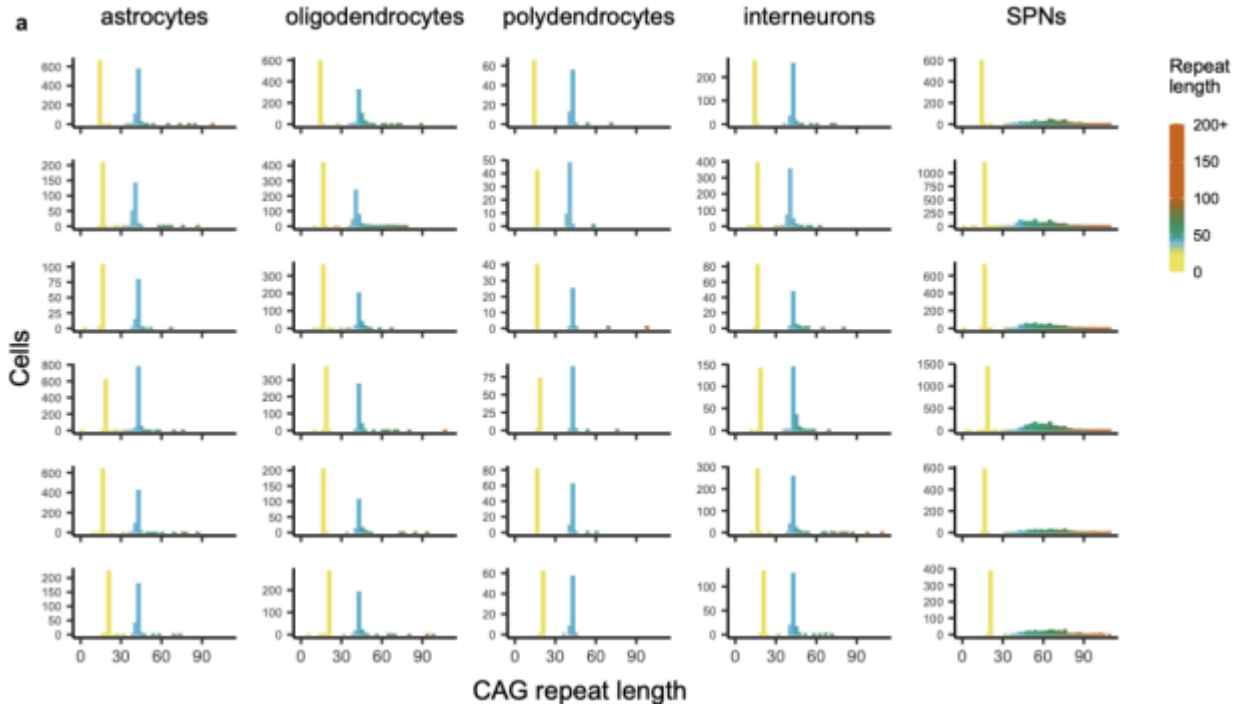
e.g., maternal inheritance of large expansions in *DMPK, FMR1*
paternal inheritance of large expansions in *HTT*

e.g., SCA2









Cold Spring Harbor Laboratory

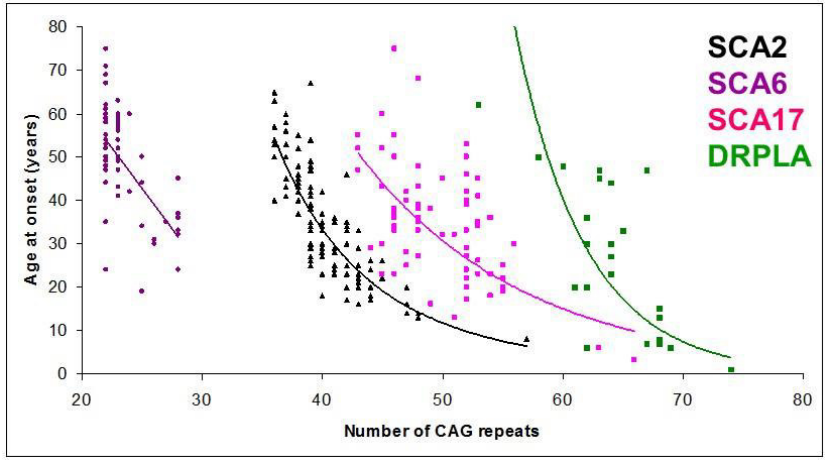
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

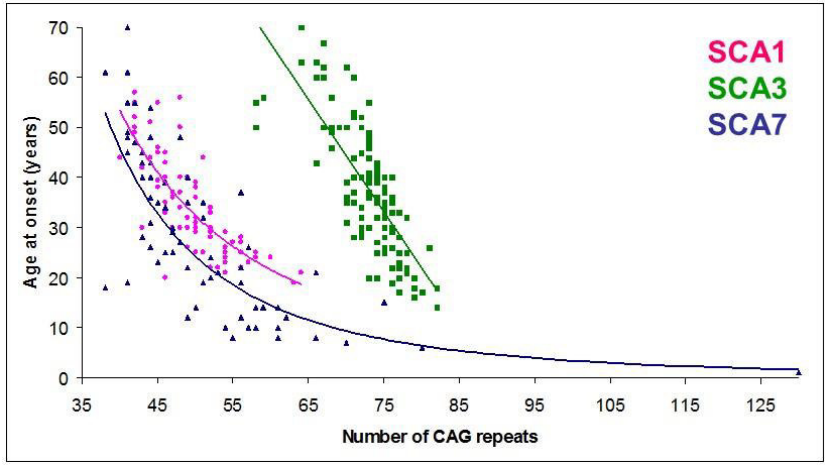
Long somatic DNA-repeat expansion drives neurodegeneration in Huntington disease

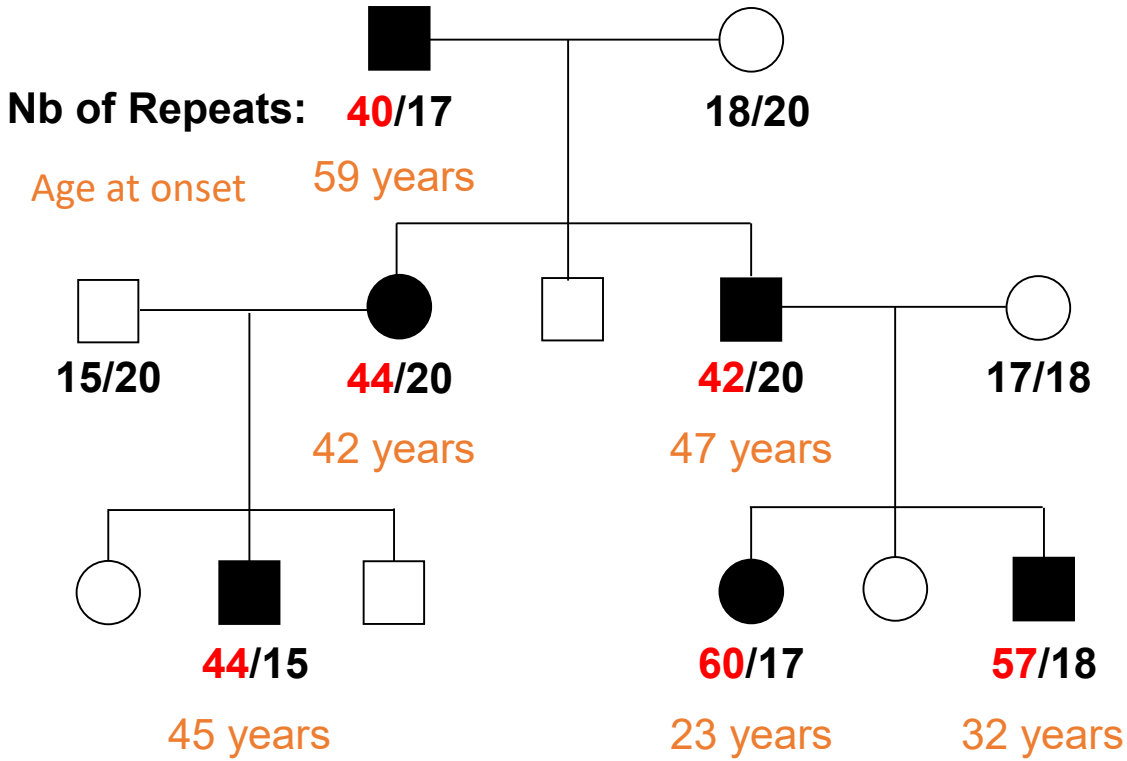
Robert E. Handsaker [1,2,§], Seva Kashin [1,2,§], Nora M. Reed [1,2,§], Steven Tan [1,2], Won-Seok Lee [1,2], Tara M. McDonald [1,2], Kiely Morris [3], Nolan Kamitaki [1,2], Christopher D. Mullally [1,2], Neda Morakabati [3], Melissa Goldman [1,2], Gabriel Lind [1,2], Rhea Kohli [1,2], Elisabeth Lawton [3], Marina Hogan [1,2], Kiku Ichihara [1,2], Sabina Berretta [1,3-5,*], Steven A. McCarroll [1,2,5,*]

Characteristics of repeat expansions: anticipation

- Inverse correlations between age at onset and repeat number
- Expansion from one generation to the next: anticipation





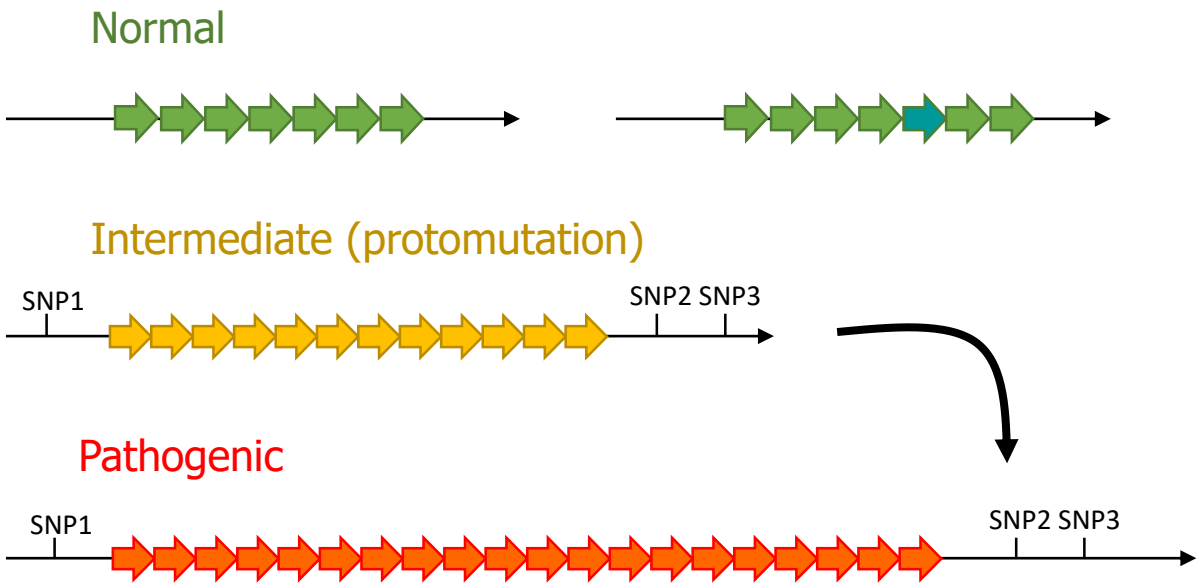


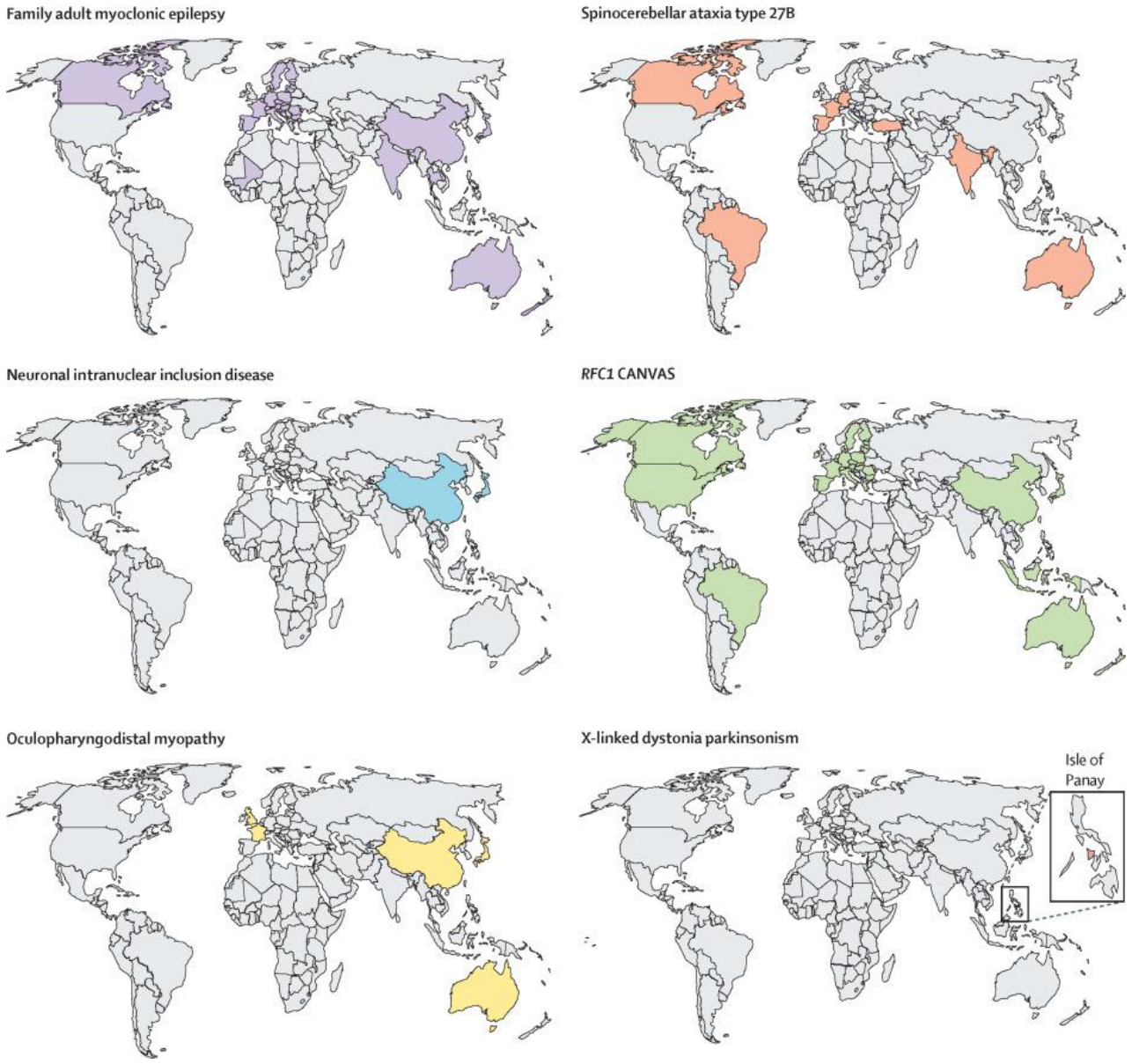
Characteristics of repeat expansions: founder effects

- Most expansions have a founder effect

(they have occurred on a specific haplotype)

- Intermediate alleles are a reservoir of larger expansions
- Geographic distribution of specific repeat expansions





Vegezzi et al. Lancet Neurology 2024

Familial Adult Myoclonus Epilepsy (FAME)

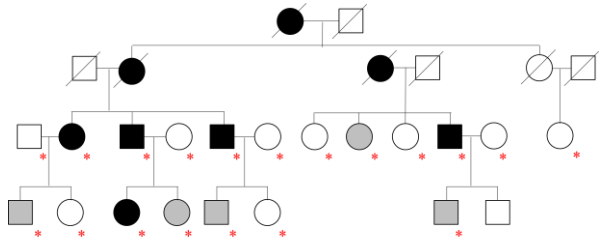
Neurogenetics (2008) 9:69–71
DOI 10.1007/s10048-007-0107-z

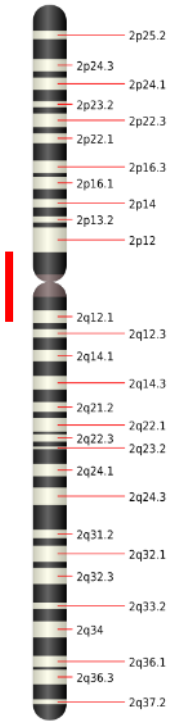
LETTER TO THE EDITORS

2008

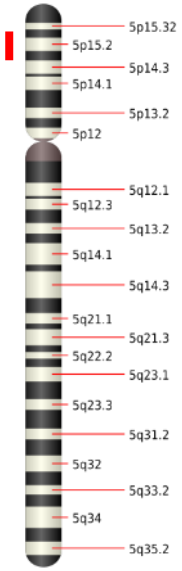
Refinement of the 2p11.1-q12.2 locus responsible for cortical tremor associated with epilepsy and exclusion of candidate genes

Cécile Saint-Martin · Delphine Bouteiller · Giovanni Stevanin · Cyprian Popescu · Céline Charon · Merle Ruberg · Stéphanie Baulac · Eric LeGuern · Pierre Labauge · Christel Depienne





2p11.1-q12.2
>159 genes

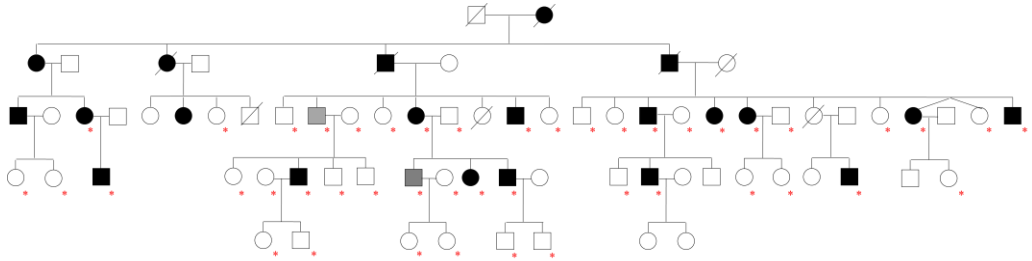


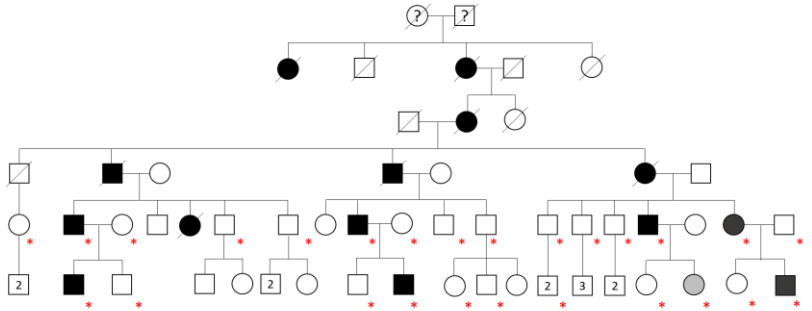
C. Depienne, PhD*
E. Magnin, MD*
D. Bouteiller, MS
G. Stevanin, PhD
C. Saint-Martin, PhD
M. Vidailhet, MD
E. Apartis, MD
E. Hirsch
E. LeGuern, MD, PhD
P. Labauge, MD, PhD
L. Rumbach, MD, PhD

Familial cortical myoclonic tremor with epilepsy

The third locus (FCMTE3) maps to 5p

2010





5p15
47genes

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], Aki Mitsue[1], Masaki Tanaka[1], Yaeko Ichikawa[9], Yuji Takahashi[10], Hidetoshi Date[1], Takashi Matsukawa[1], Junko Kanda[1], Fumiko Kusunoki Nakamoto[1], Mana Higashihara[11], Koji Abe[12], Ryoko Koike[13], Mutsuo Sasagawa[14], Yasuko Kuroha[13], Naoya Hasegawa[15], Norio Kanesawa[16], Takayuki Kondo[17], Takefumi Hitomi[17,18], Masayoshi Tada[19], Hiroki Takano[20], Yutaka Saito[21], Kazuhiro Sanpei[22], Osamu Onodera[19], Masatoyo Nishizawa[23], Masayuki Nakamura[24], Takeshi Yasuda[25], Yoshio Sakiyama[26], Mieko Otsuka[27], Akira Ueki[43], Ken-ichi Kaida[28], Jun Shimizu[1], Ritsuko Hanajima[29], Toshihiro Hayashi[1], Yasuo Terao[30], Satomi Inomata-Terada[1], Masashi Hamada[1], Yuichiro Shirota[1], Akatsuki Kubota[1], Yoshikazu Ugawa[31], Kishin Koh[32], Yoshihisa Takiyama[32], Natsumi Ohsawa-Yoshida[33], Shoichi Ishiura[33,34], Ryo Yamasaki[35], Akira Tamaoka[36], Hiroshi Akiyama[37], Taisuke Otsuki[38], Akira Sano[24], Akio Ikeda[39], Jun Goto[40], Shinichi Morishita[2] and Shoji Tsuji[1,41,42]*

→ **Sequencing of the candidate intervals (coding and non-coding)**

→ **Short-read genome sequencing / RNAseq (lymphoblasts)**

No pathogenic mutation identified

Familial Adult Myoclonus Epilepsy (FAME)

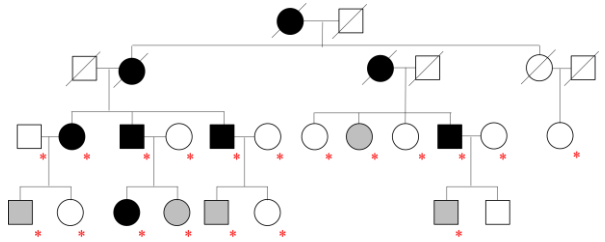
Neurogenetics (2008) 9:69–71
DOI 10.1007/s10048-007-0107-z

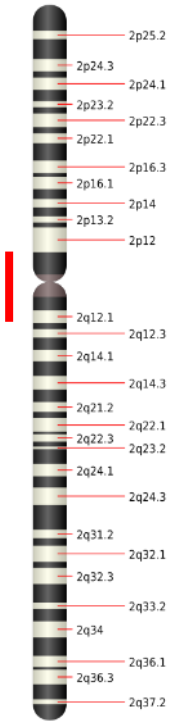
LETTER TO THE EDITORS

2008

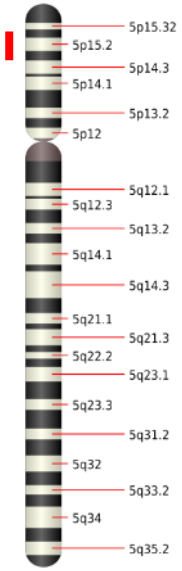
Refinement of the 2p11.1-q12.2 locus responsible for cortical tremor associated with epilepsy and exclusion of candidate genes

Cécile Saint-Martin · Delphine Bouteiller · Giovanni Stevanin · Cyprian Popescu · Céline Charon · Merle Ruberg · Stéphanie Baulac · Eric LeGuern · Pierre Labauge · Christel Depienne





2p11.1-q12.2
>159 genes

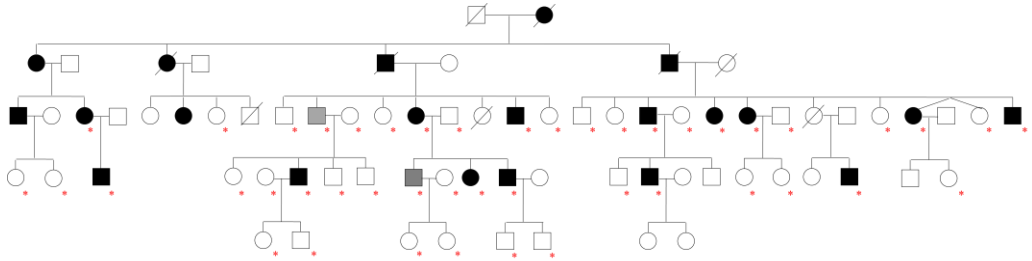


C. Depienne, PhD*
E. Magnin, MD*
D. Bouteiller, MS
G. Stevanin, PhD
C. Saint-Martin, PhD
M. Vidailhet, MD
E. Apartis, MD
E. Hirsch
E. LeGuern, MD, PhD
P. Labauge, MD, PhD
L. Rumbach, MD, PhD

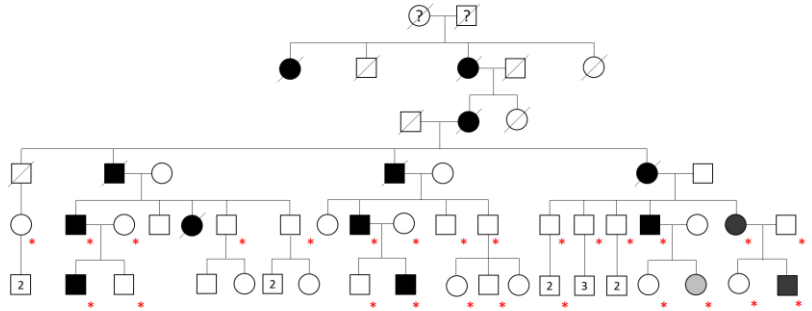
Familial cortical myoclonic tremor with epilepsy

The third locus (FCMTE3) maps to 5p

2010



5p15
47genes



nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-12671-y OPEN

Intronic ATTTC repeat expansions in *STARD7* in familial adult myoclonic epilepsy linked to chromosome 2

Mark A. Corbett et al.#

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-12763-9 OPEN

Unstable TTTTA/TTTCA expansions in *MARCH6* are associated with Familial Adult Myoclonic Epilepsy type 3

Rahel T. Florian et al.#

Familial Adult Myoclonus Epilepsy (FAME)

1965-1967
clinical description
Vogel & Diebold and Castaigne

1990
Cortical tremor (Ikeda et al)

1999
FAME1 linked to chr8q (Mikami et al)

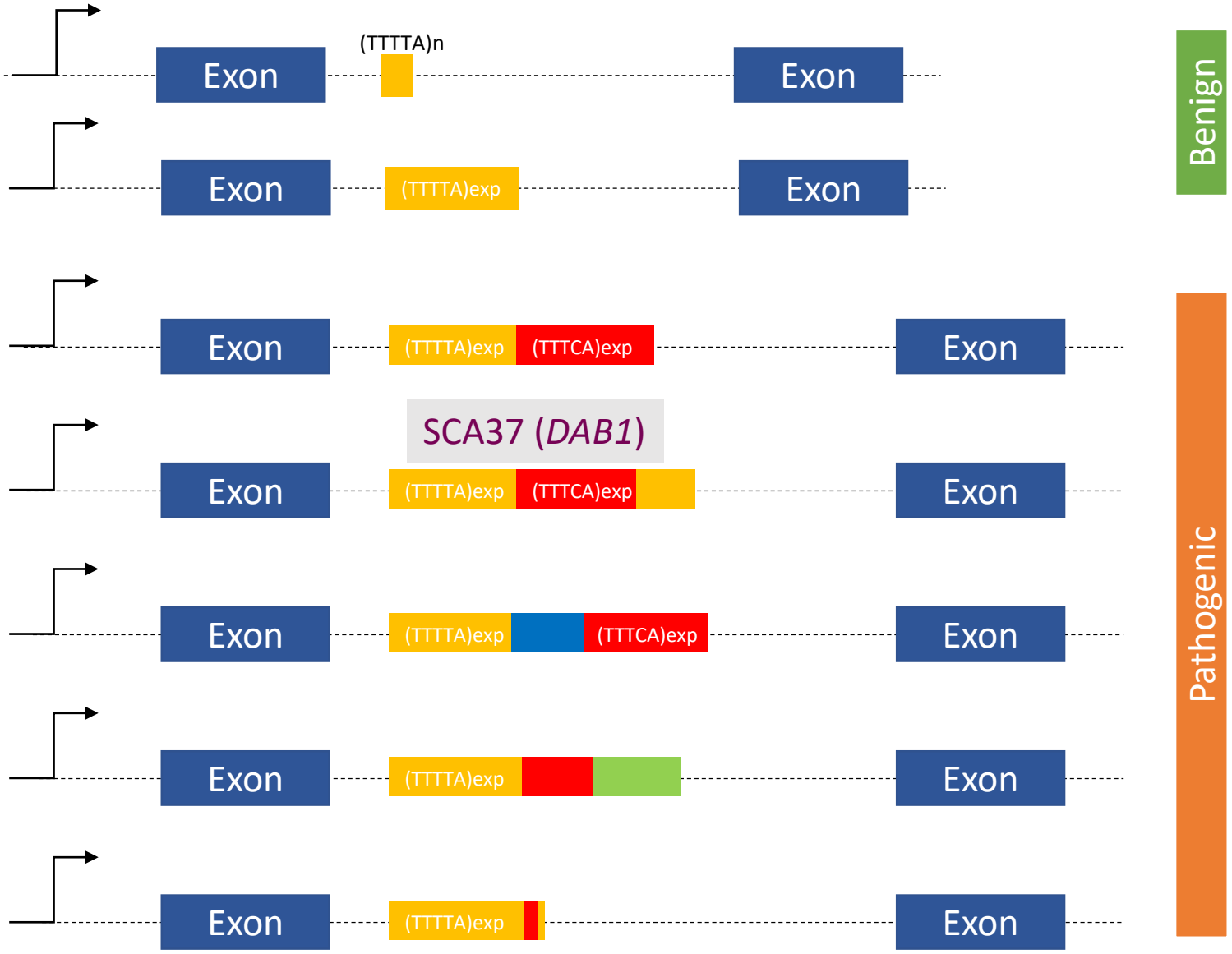
2002
FAME2 linked to chr2 (Mikami et al)

2010
FAME3 linked to chr5p (Mikami et al)

2018 Ishiura et al
Identification of TTTTA/TTTCA expansions
in FAME1 (***SAMD12***), FAME6 (***TNRC6A***) and FAME7 (***RAPGEF2***)

2019 Corbett et al; Florian et al: Yeetong et al.
Identification of TTTTA/TTTCA expansions
in FAME2 (***STARD7***), FAME3 (***MARCHF6***) and FAME4 (***YEATS2***)

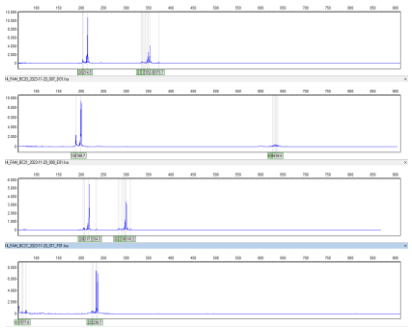
2024 Yeetong et al.
Identification of TTTTA/TTTCA expansions
in FAME8 (***RAI1***)



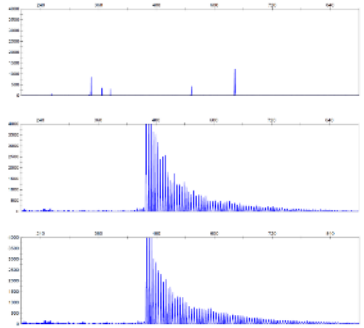
Adapted from Corbett et al. and Depienne et al. Epilepsia 2023

Methods to identify and/or characterize repeat expansions

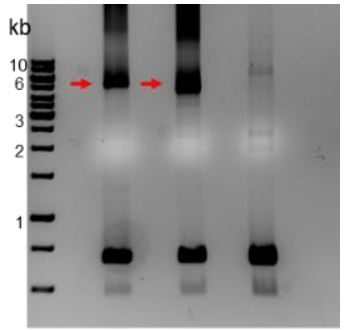
Fragment analysis (Genescan)



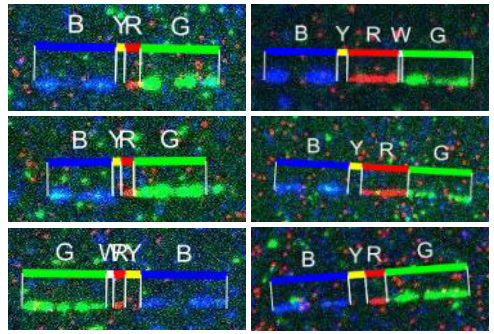
Repeat-primed PCR (RP-PCR)



Amplification (long-range PCR)

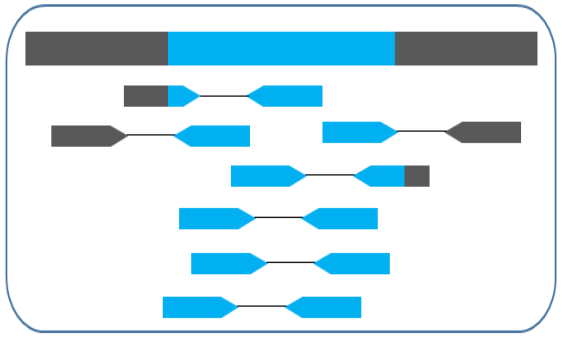


Molecular combing (Fiber FISH)

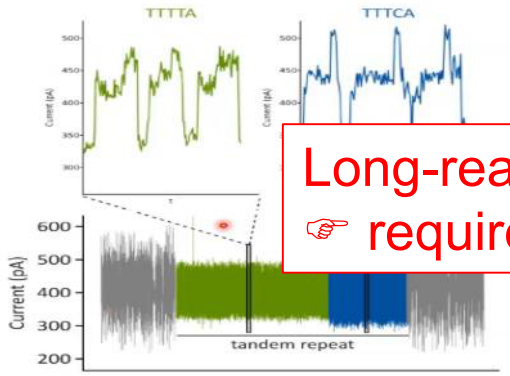


	Fragment analysis (Genescan)	Repeat-primed PCR (RP-PCR)	Amplification (long-range PCR)	Molecular combing (Fiber FISH)
Length	<800 bp			> 1 kb
Sequence				

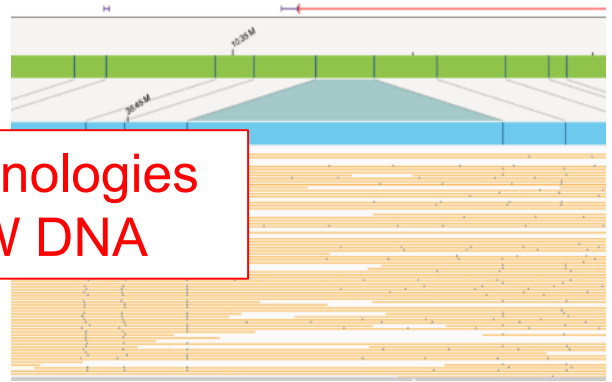
Short-read genome sequencing



Nanopore sequencing



Optical genome mapping



Long-read technologies
☞ require HMW DNA

- Yes
- No
- Partial

	Short-read genome sequencing	Nanopore sequencing	Optical genome mapping
Length			>500 bp
Sequence			

Identification of tandem repeat EXPAnsions in unsolved Neurological Disorders (EXPAND)

Short read genome sequencing

- 250 short-read genomes
- phenotypes including
  - **cerebellar ataxia** (~60)
  - **tremor** (~30)
  - **dystonia** (~10)
  - **leukencephalopathy** (~10)
  - Epilepsy (~30)
  - Aicardi-syndrome (4)
  - Septo-optic dysplasia
  - Tourette syndrome (~20)
  - Motor neuron disease
  - Neurodevelopmental disorders

- Strelka (SNVs, indels), Manta (SVs)
- STR analysis

Long read nanopore sequencing

- 100 long-read genomes (/130)
- Optimization of HMW DNA extraction
- >15X, N50 > 40kb

- wf-human-variation pipeline (Snakemake)
- Clair3 (SNVs, indels), Sniffles (SVs)
- Straglr (STR)
- methylation, phasing

RNA sequencing

DFG Deutsche Forschungsgemeinschaft

Cologne Center for Genomics

WEST GERMAN GENOME CENTER Düsseldorf

Identification of repeat expansion from short-read genome data

✓ Requires specific bioinformatic tools in addition of standard SNV/CNV/SV detection pipelines

Targeted (locus and motif known)

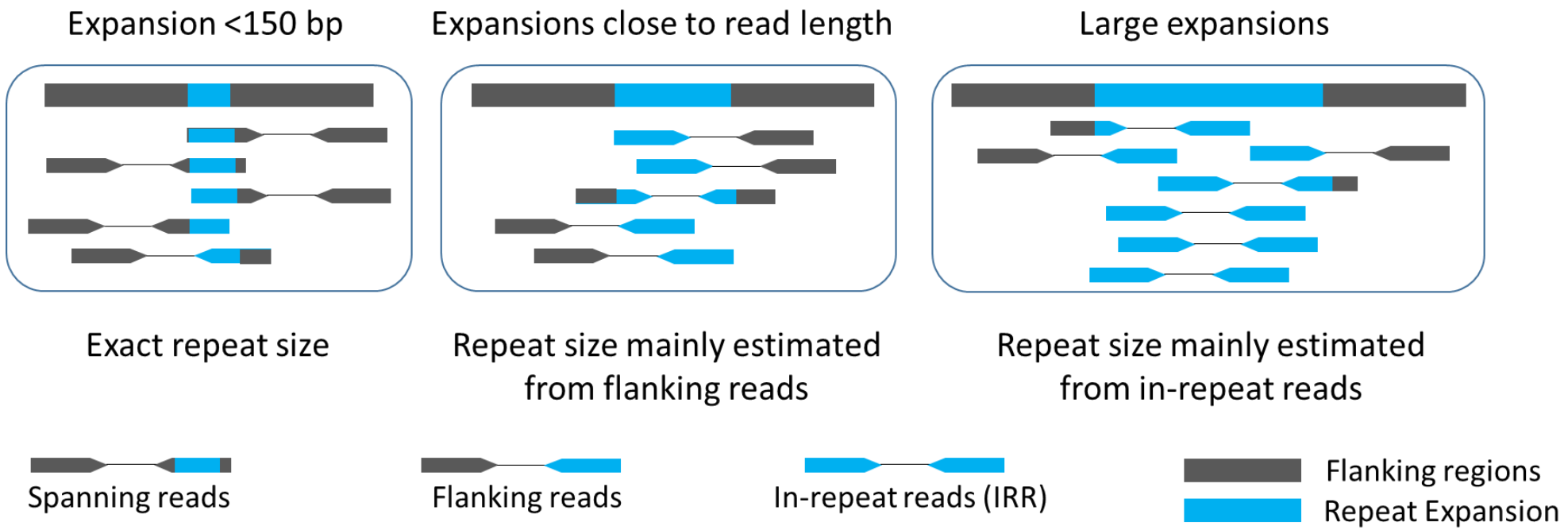
- HipSTR (2017)
- TREDPARSE (2017)
- ExpansionHunter (2017)
- STRetch (2018)
- GangSTR (2019)

Locus / motif - Independent

- TRhist (2013)
- exSTRa (2018)
- ExpansionHunter de novo (2020)
- STRling (2022)

✓ Underestimates large expansions

✓ Does not detect hidden repeat insertions



Tanudisastro et al. Nat Rev Genet 2024

Adapted from Dolzhenko et al. 2017

Identification of repeat expansion from short-read genome data

✓ Requires specific bioinformatic tools in addition of standard SNV/CNV/SV detection pipelines

Targeted (locus and motif known)

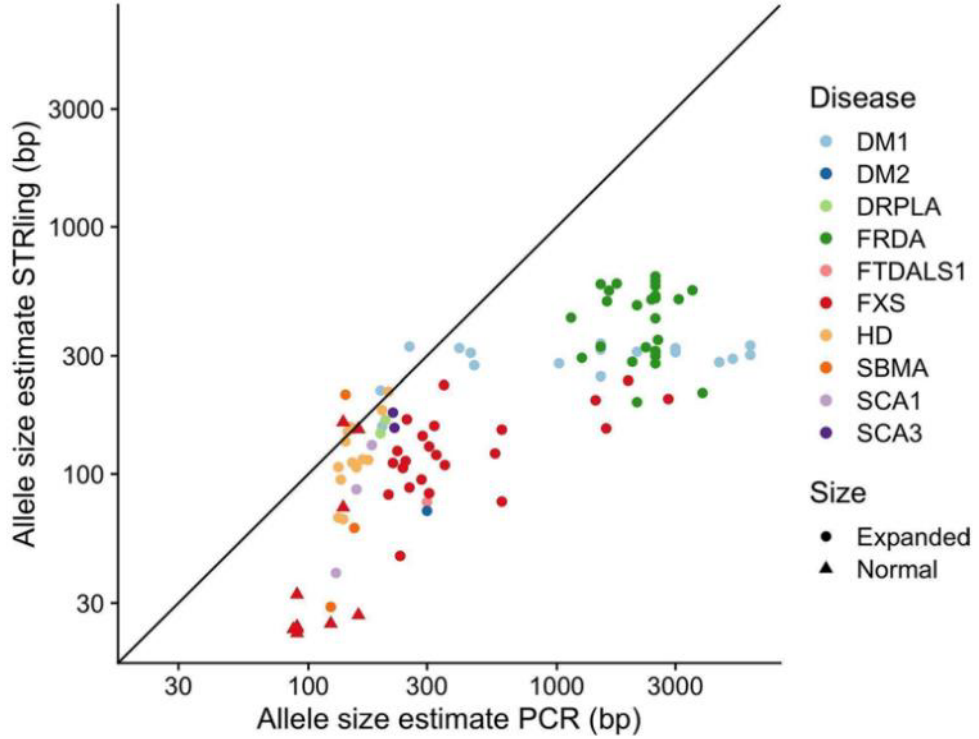
- HipSTR (2017)
- TREDPARSE (2017)
- ExpansionHunter (2017)
- STRetch (2018)
- GangSTR (2019)

Locus / motif - Independent

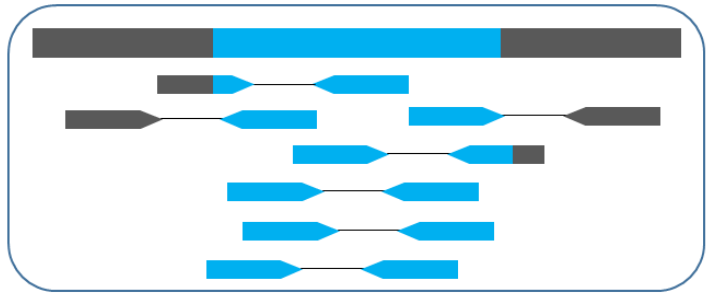
- TRhist (2013)
- exSTRa (2018)
- ExpansionHunter de novo (2020)
- STRling (2022)

✓ Underestimates large expansions

✓ Does not detect hidden repeat insertions



Large expansions



Repeat size mainly estimated from in-repeat reads

Tanudisastro et al. Nat Rev Genet 2024

Dashnow et al. 2020

Identification of tandem repeat EXPAnsions in unsolved Neurological Disorders (EXPAND)

Short read genome sequencing → Long read nanopore sequencing → RNA sequencing

- 250 short-read genomes
- phenotypes including
  - **cerebellar ataxia** (~60)
  - **tremor** (~30)
  - **dystonia** (~10)
  - **leukencephalopathy** (~10)
  - Epilepsy (~30)
  - Aicardi-syndrome (4)
  - Septo-optic dysplasia
  - Tourette syndrome (~20)
  - Motor neuron disease
  - Neurodevelopmental disorders

- Strelka (SNVs, indels), Manta (SVs)
- STR analysis

DFG Deutsche Forschungsgemeinschaft

Method

Detection of long repeat expansions from PCR-free whole-genome sequence data

Egor Dolzhenko,[1,18] Joke J.F.A. van Vugt,[2,18] Richard J. Shaw,[3,4] Mitchell A. Bekritsky,[3] Marka van Blitterswijk,[5] Giuseppe Narzisi,[6] Subramanian S. Ajay,[1] Vani Rajan,[1] Bryan R. Lajoie,[1] Nathan H. Johnson,[1] Zoya Kingsbury,[3] Sean J. Humphray,[3] Raymond D. Schellevis,[2] William J. Brands,[2] Matt Baker,[5] Rosa Rademakers,[5] Maarten Kooyman,[7] Gijs H.P. Tazelaar,[2] Michael A. van Es,[2] Russell McLaughlin,[8,9] William Sproviero,[10] Aleksey Shatunov,[10] Ashley Jones,[10] Ahmad Al Khleifat,[10] Alan Pittman,[11] Sarah Morgan,[11] Orla Hardiman,[8,9] Ammar Al-Chalabi,[10] Chris Shaw,[10] Bradley Smith,[10] Edmund J. Neo,[10] Karen Morrison,[12] Pamela J. Shaw,[13] Catherine Reeves,[6] Lara Winterkorn,[6] Nancy S. Wexler,[14,15] The US–Venezuela Collaborative Research Group,[16] David E. Housman,[17] Christopher W. Ng,[17] Alina L. Li,[17] Ryan J. Taft,[1] Leonard H. van den Berg,[2] David R. Bentley,[3] Jan H. Veldink,[2,18] and Michael A. Eberle[1,18]

Dashnow *et al. Genome Biology* (2022) 23:257
https://doi.org/10.1186/s13059-022-02826-4

Genome Biology

METHOD Open Access

STRling: a k-mer counting approach that detects short tandem repeat expansions at known and novel loci

Harriet Dashnow[1], Brent S. Pedersen[1,2], Laurel Hiatt[1], Joe Brown[1], Sarah J. Beecroft[3,4], Gianina Ravenscroft[4], Amy J. LaCroix[5], Phillipa Lamont[6], Richard H. Roxburgh[7], Miriam J. Rodrigues[7,8], Mark Davis[9], Heather C. Mefford[5], Nigel G. Laing[4,9] and Aaron R. Quinlan[1*]

- >10,000 repeat expansions in one genome detected using STRling
- >300 repeat expansions of motifs ≥ 3 per individual
- Half of families with ataxia had a possible repeat expansion in *FGF14* !

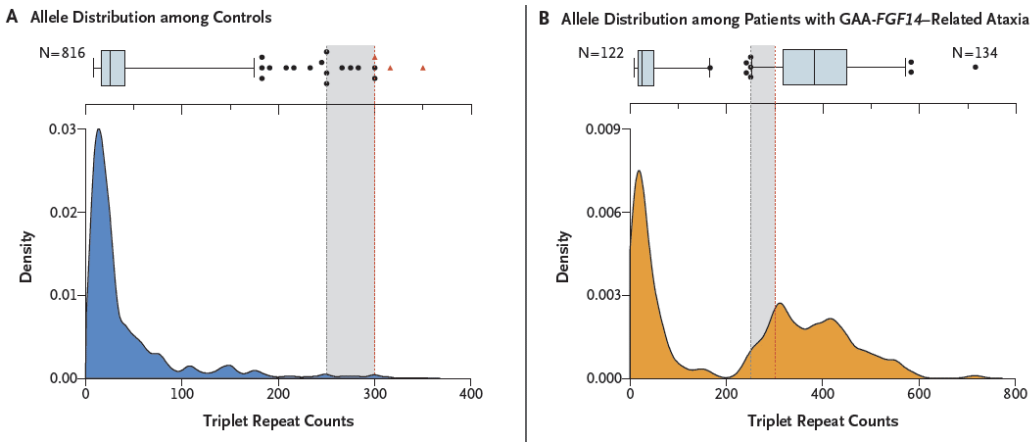
FGF14 repeat expansions (SCA27B)

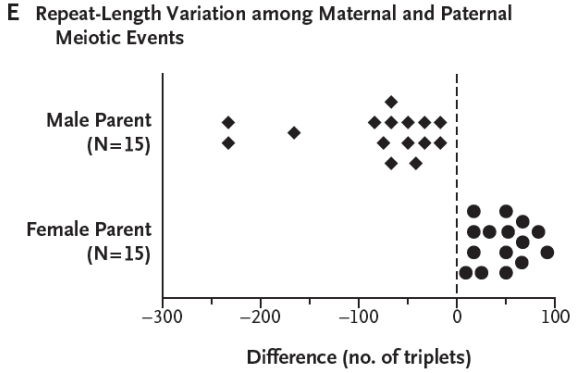
The NEW ENGLAND JOURNAL of MEDICINE

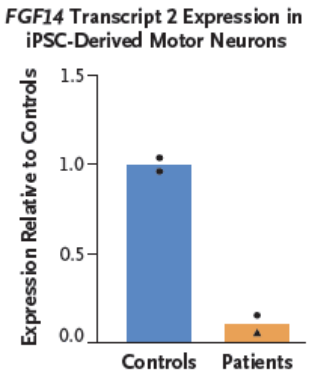
ORIGINAL ARTICLE

Deep Intronic *FGF14* GAA Repeat Expansion in Late-Onset Cerebellar Ataxia

D. Pellerin, M.C. Danzi, C. Wilke, M. Renaud, S. Fazal, M.-J. Dicaire, C.K. Scriba, C. Ashton, C. Yanick, D. Beijer, A. Rebelo, C. Rocca, Z. Jaunmuktane, J.A. Sonnen, R. Larivière, D. Genís, L. Molina Porcel, K. Choquet, R. Sakalla, S. Provost, R. Robertson, X. Allard-Chamard, M. Tétreault, S.J. Reiling, S. Nagy, V. Nishadham, M. Purushottam, S. Vengalil, M. Bardhan, A. Nalini, Z. Chen, J. Mathieu, R. Massie, C.H. Chalk, A.-L. Lafontaine, F. Evoy, M.-F. Rioux, J. Ragoussis, K.M. Boycott, M.-P. Dubé, A. Duquette, H. Houlden, G. Ravenscroft, N.G. Laing, P.J. Lamont, M.A. Saporta, R. Schüle, L. Schöls, R. La Piana, M. Synofzik, S. Zuchner, and B. Brais



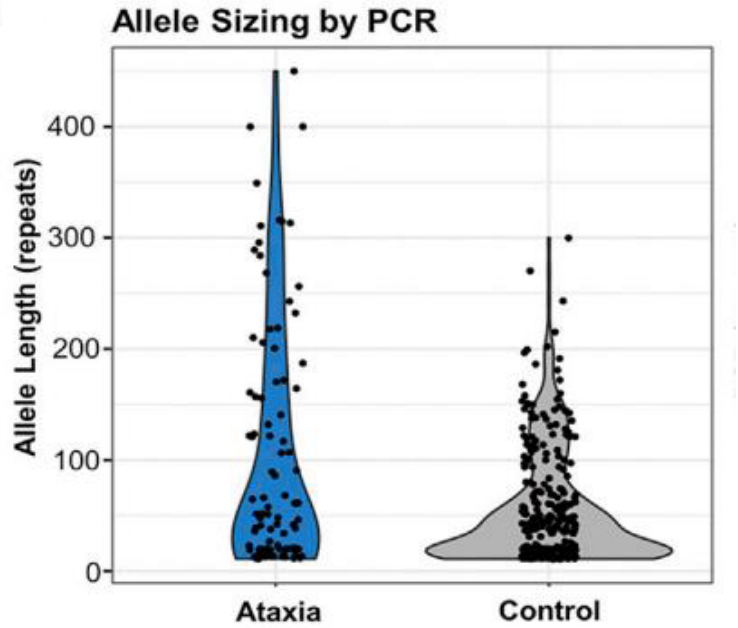




ARTICLE

An intronic GAA repeat expansion in *FGF14* causes the autosomal-dominant adult-onset ataxia SCA50/ATX-FGF14

Haloom Rafehi,[1,2,30] Justin Read,[3,4,30] David J. Szmulewicz,[5,6,30] Kayli C. Davies,[3,4] Penny Snell,[3] Liam G. Fearnley,[1,2,3] Liam Scott,[1] Mirja Thomsen,[7] Greta Gillies,[3] Kate Pope,[3] Mark F. Bennett,[1,2,8] Jacob E. Munro,[1,2] Kathie J. Ngo,[9] Luke Chen,[10] Mathew J. Wallis,[11,12,13] Ernest G. Butler,[14] Kishore R. Kumar,[15,16,17] Kathy HC. Wu,[18,19,20,21] Susan E. Tomlinson,[21,22] Stephen Tisch,[18,22] Abhishek Malhotra,[23] Matthew Lee-Archer,[24] Egor Dolzhenko,[25] Michael A. Eberle,[25] Leslie J. Roberts,[26] Brent L. Fogel,[9,28] Norbert Brüggemann,[7,27] Katja Lohmann,[7] Martin B. Delatycki,[3,4,29] Melanie Bahlo,[1,2,31,*] and Paul J. Lockhart[3,4,31,*]

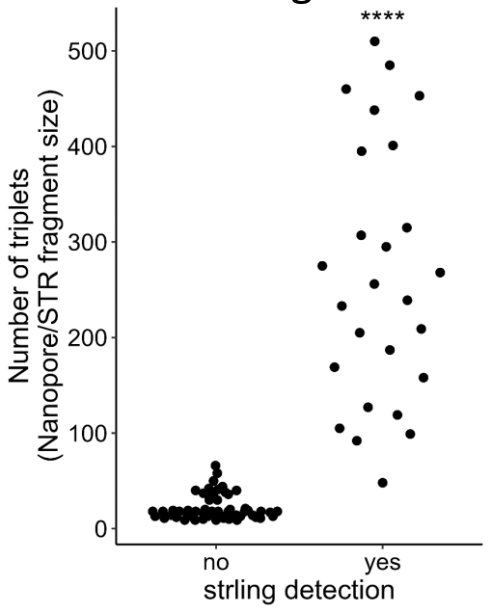


Pellerin et al. NEJM 2023

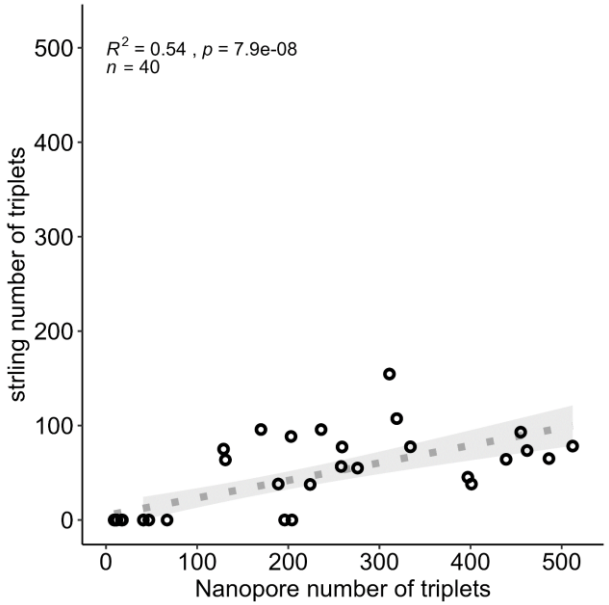
Rafehi et al. AJHG 2023

Comparaison of tools and methods to detect *FGF14* expansions

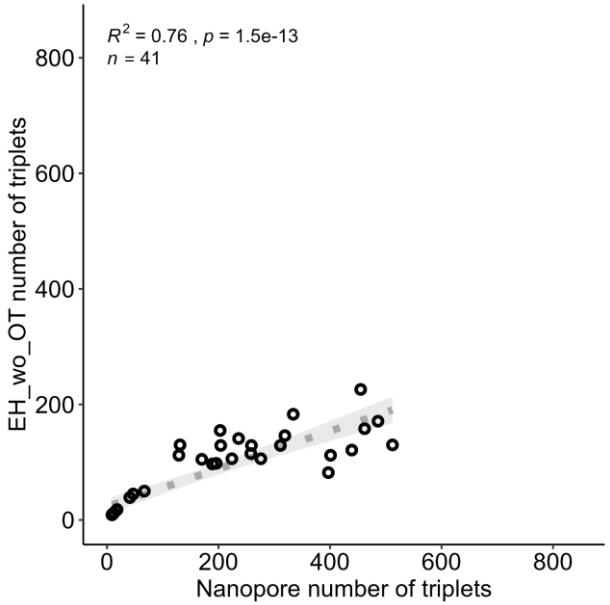
STRling

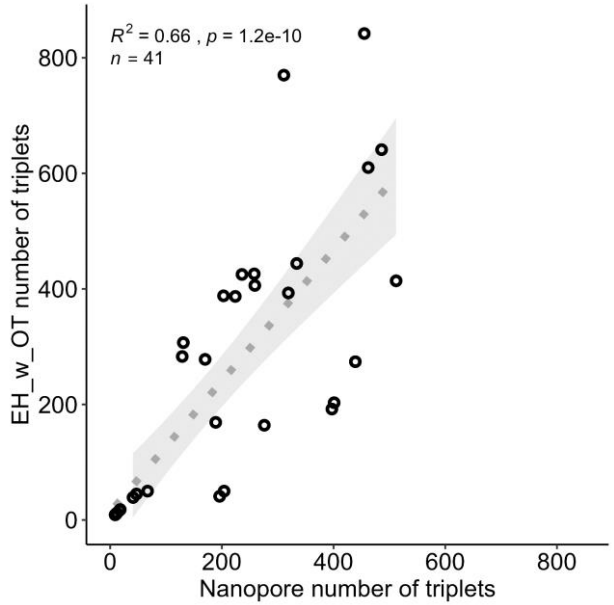


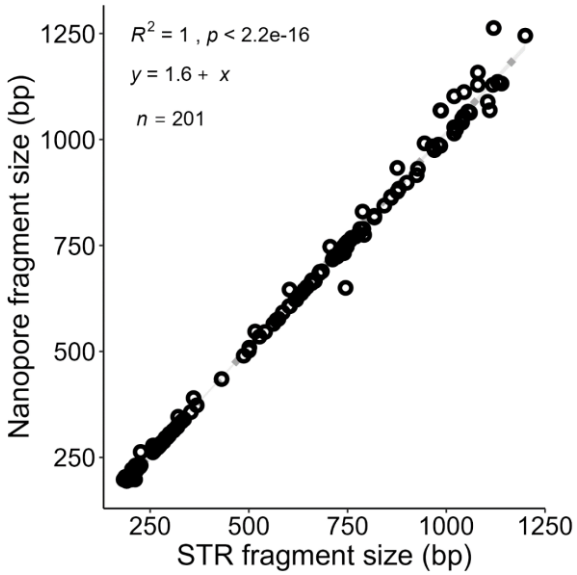
STRling



ExpansionHunter

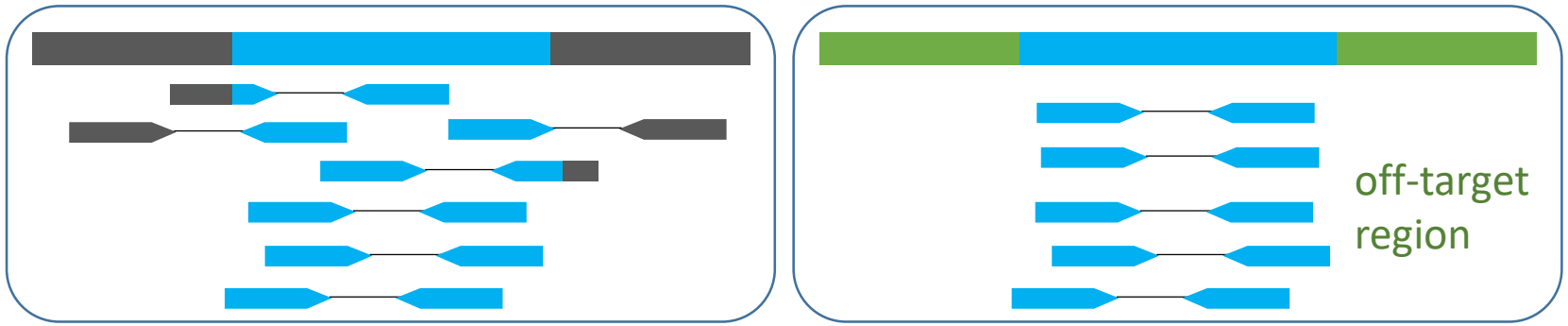






Without off-target

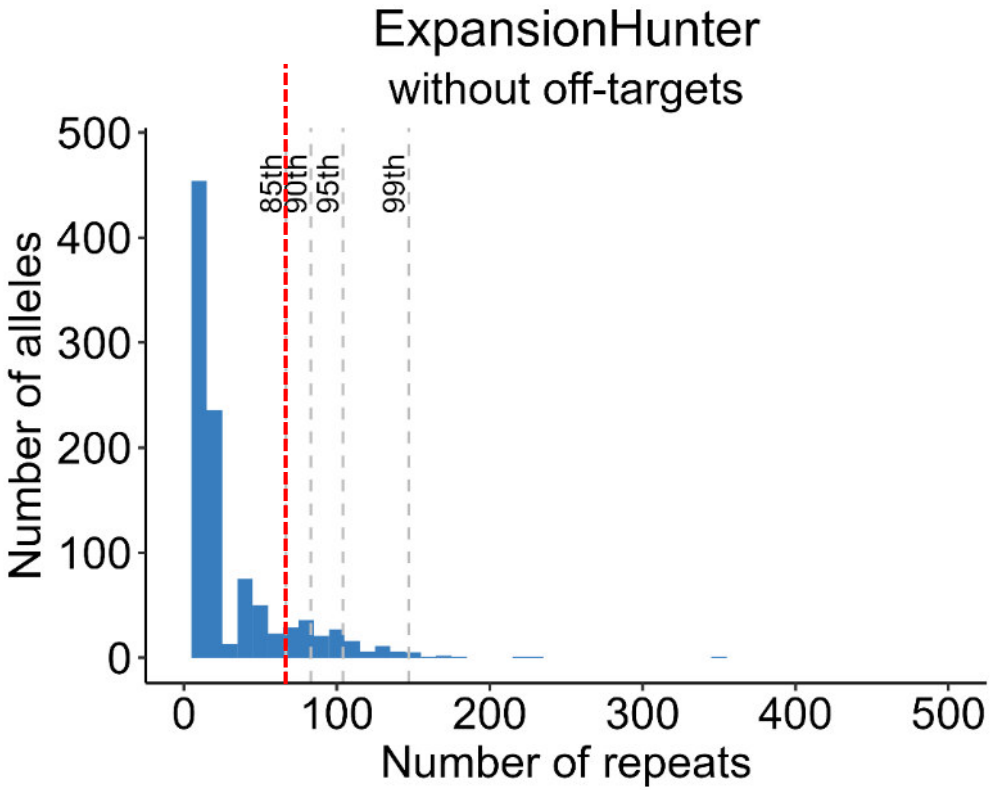
With off-target

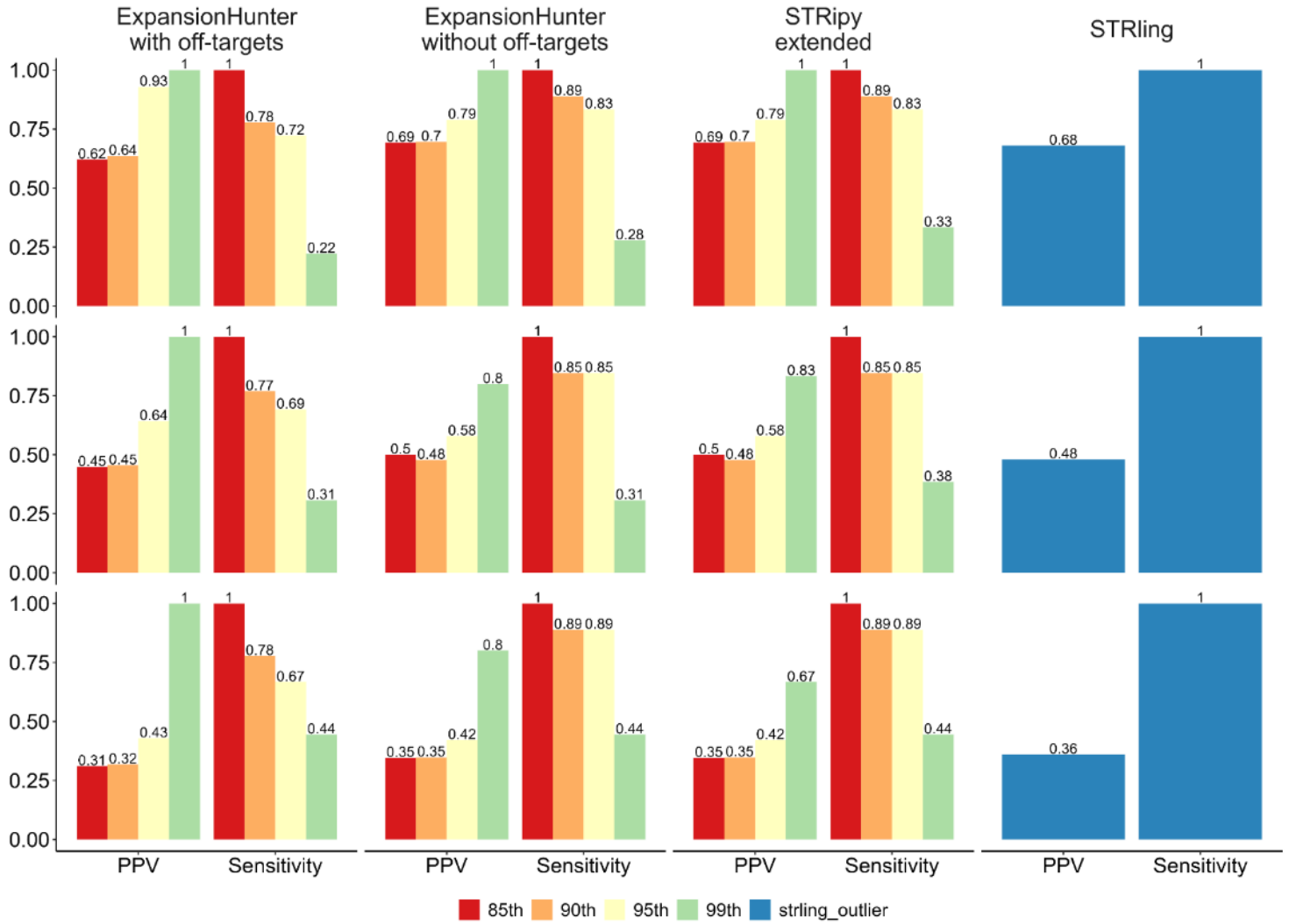


Mohren*, Erdlenbruch*, Kilpert*, Leitao*, *et al*. Nat Commun 2024

Identification of *FGF14* expansions using an outlier approach

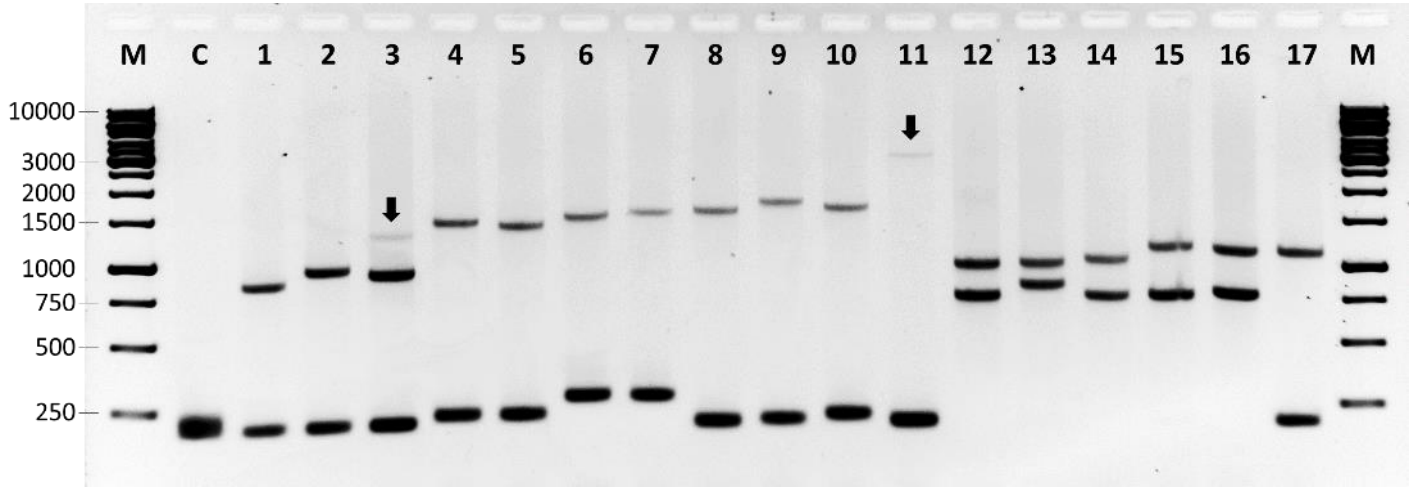
Positive predictive value (PPV) = TP/(TP+FP)
Sensitivity = TP/(TP+FN)
TP : true positives
FP : false positives
FN: false negatives

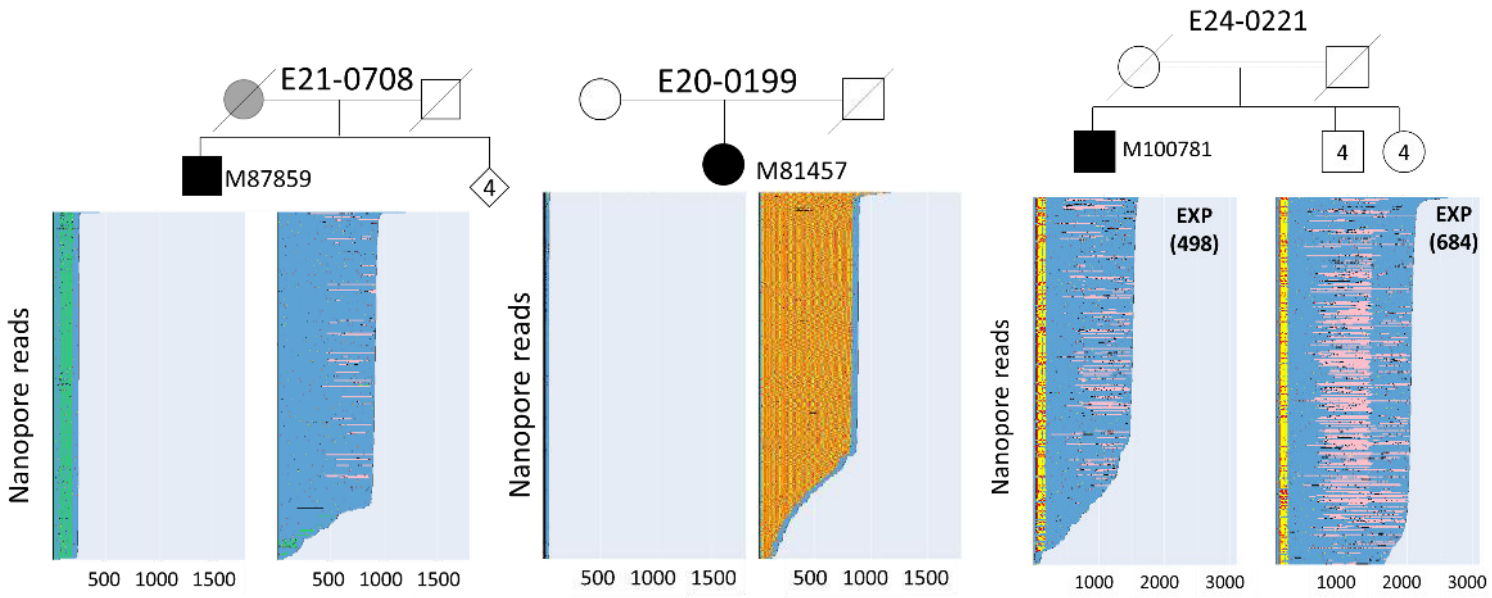


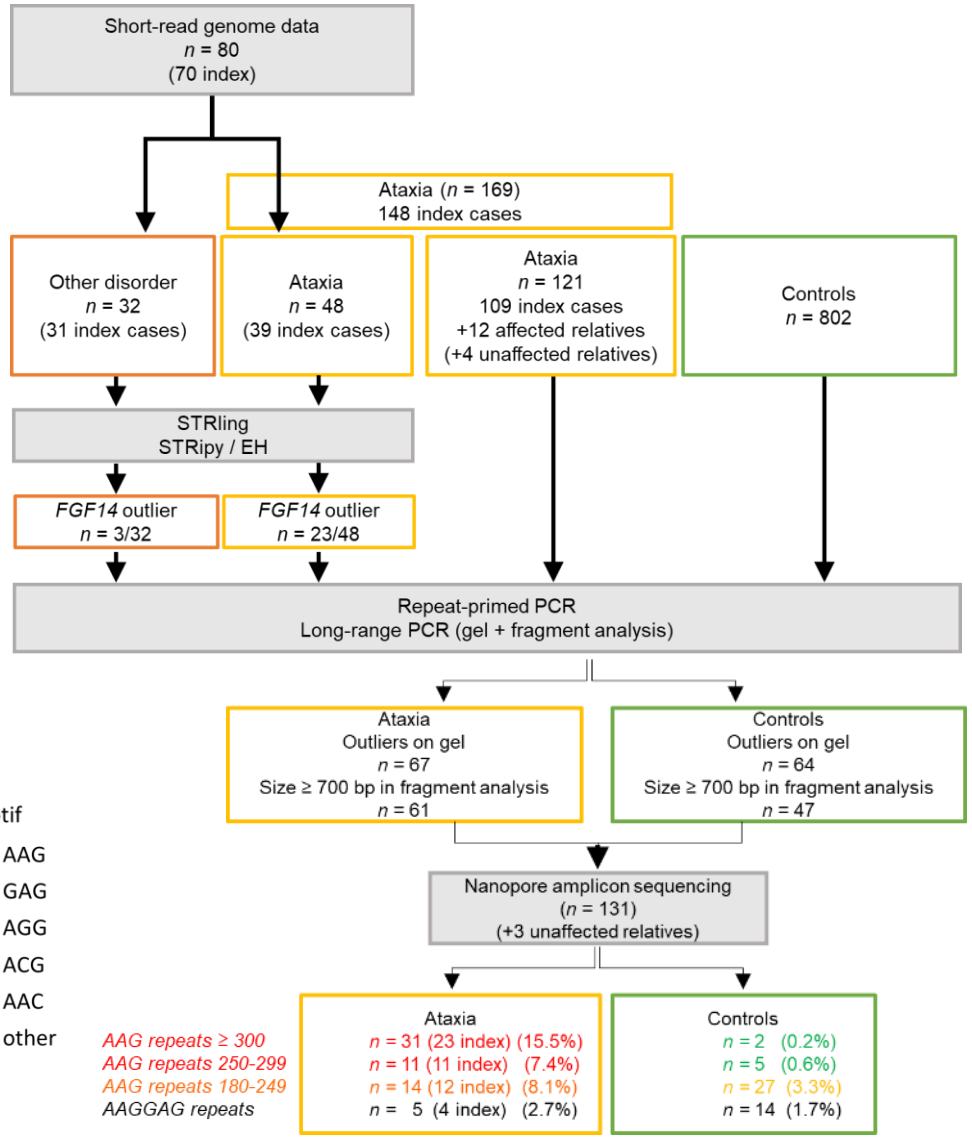


Mohren*, Erdlenbruch*, Kilpert*, Leitao*, *et al*. Nat Commun 2024

Confirmation of *FGF14* expansions by targeted nanopore sequencing

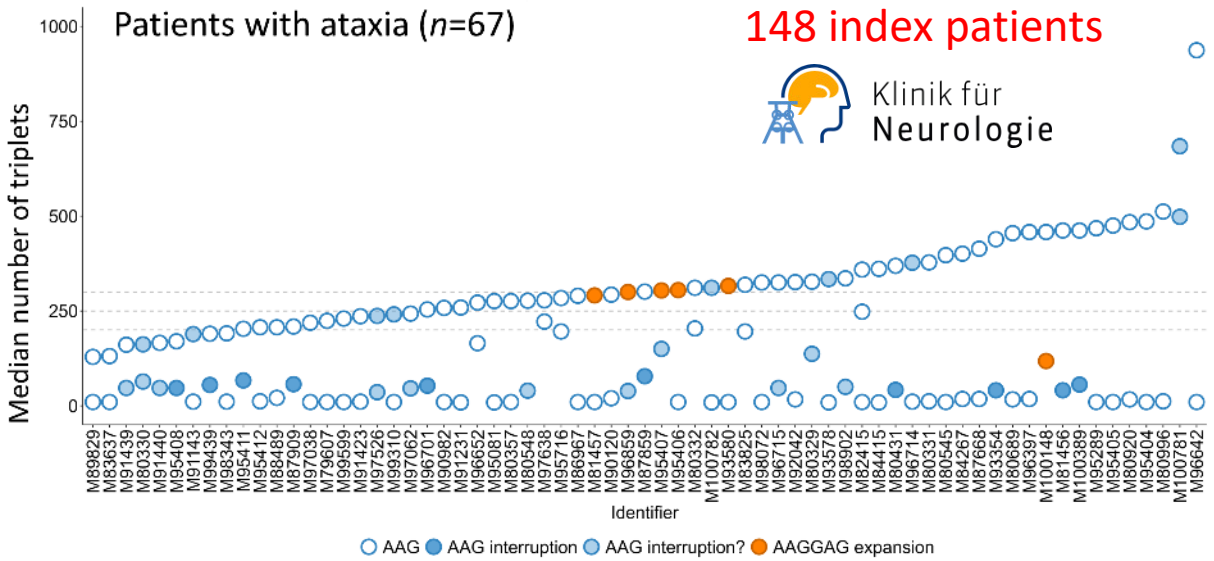


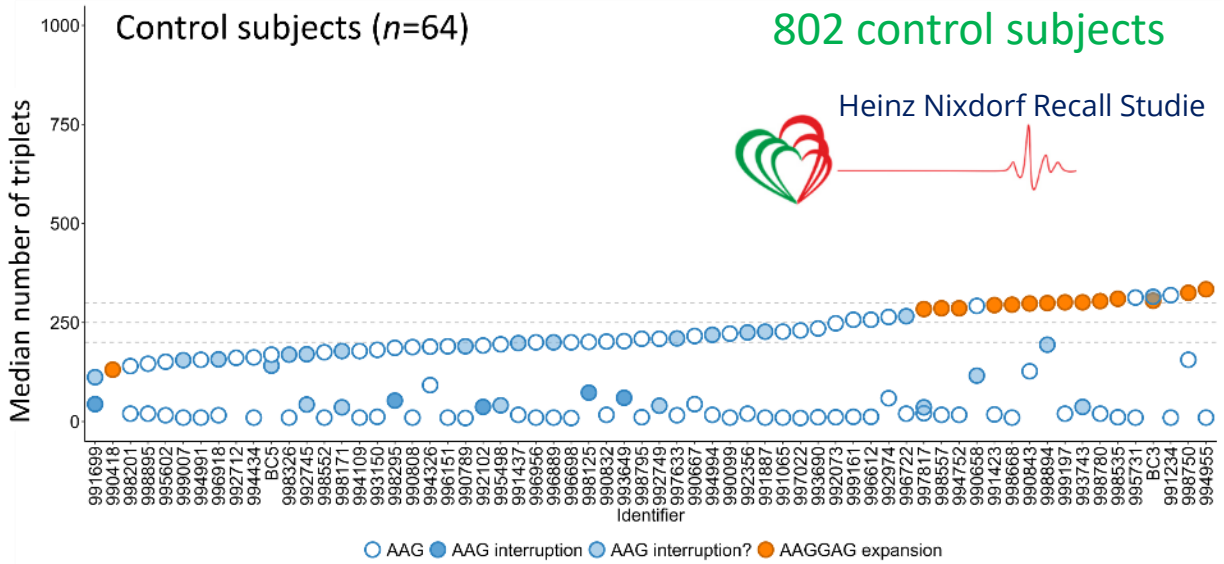


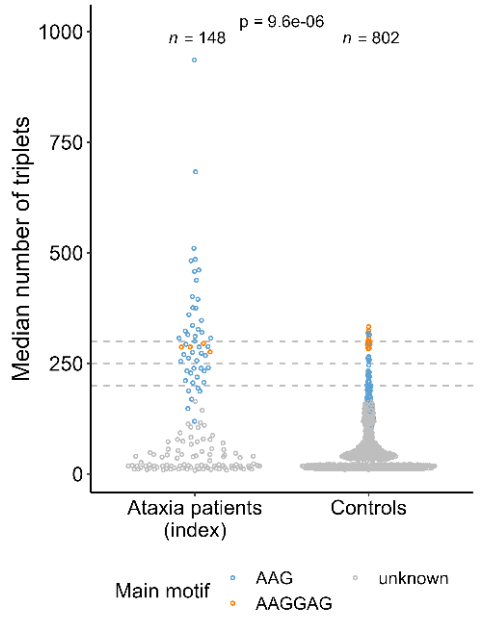


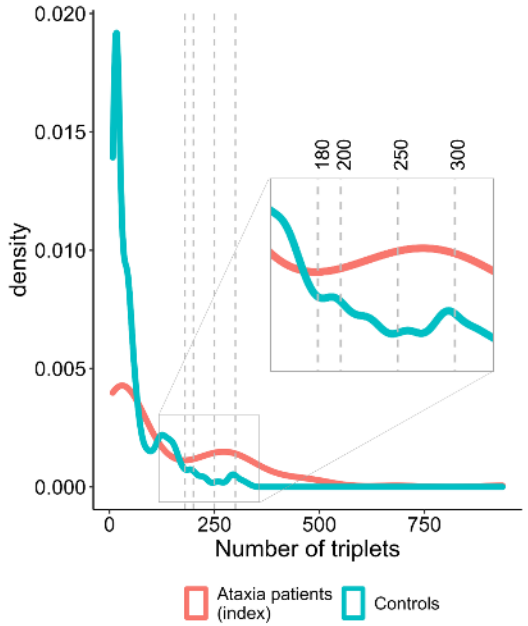
Mohren*, Erdlenbruch*, Kilpert*, Leitao*, *et al*. Nat Commun 2024

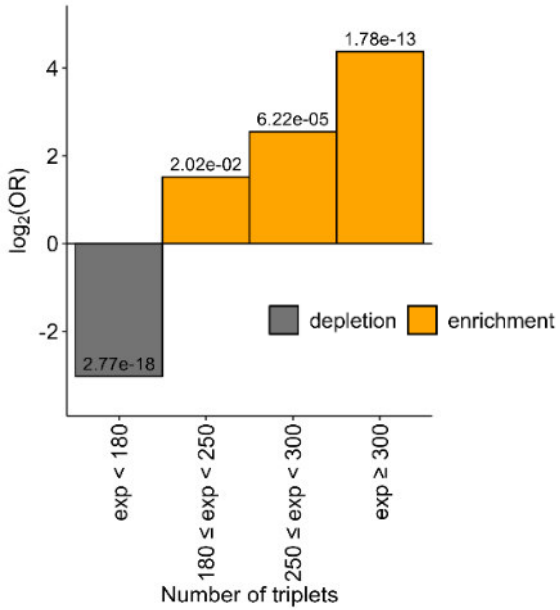
Distribution of *FGF14* alleles in patients and controls





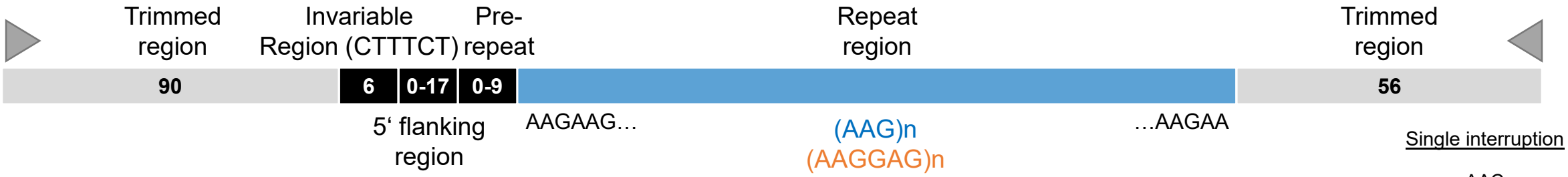




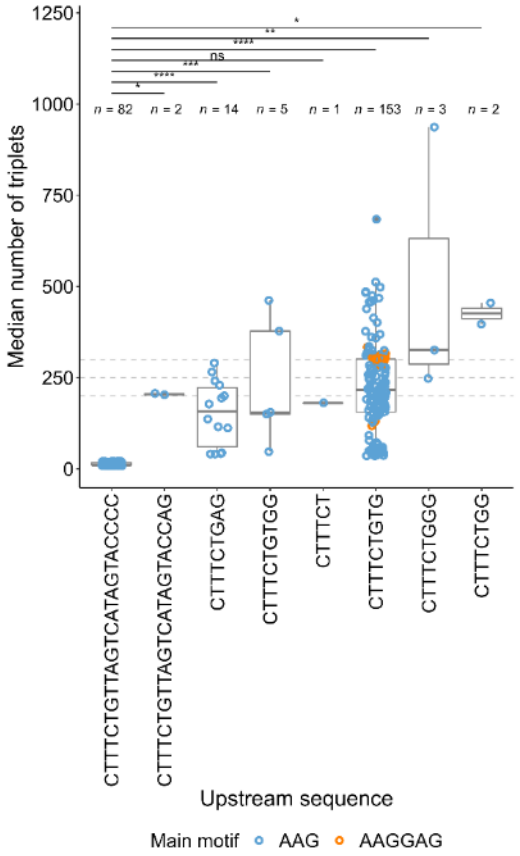


Mohren*, Erdlenbruch*, Kilpert*, Leitao*, *et al*. Nat Commun 2024

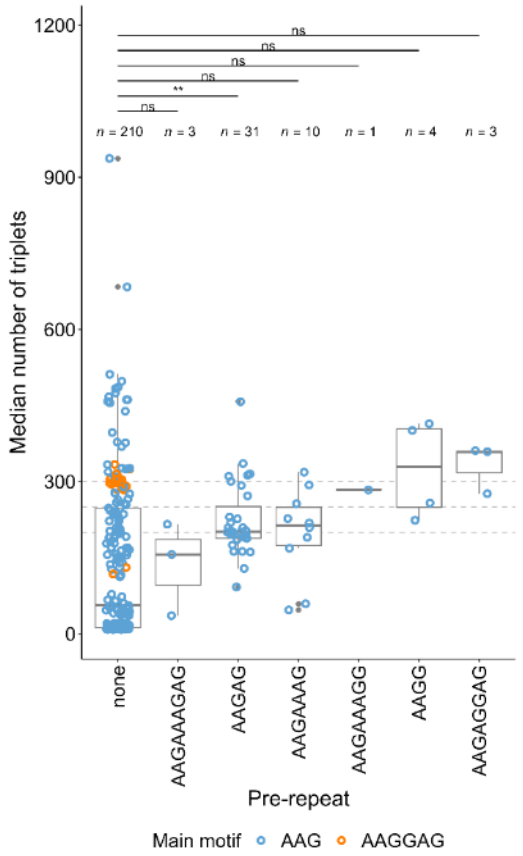
Impact of flanking regions and pre-repeats



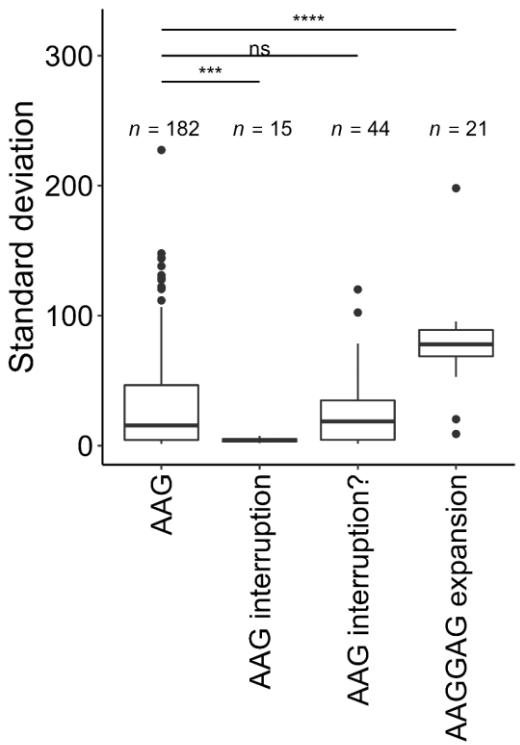
5’ flanking region

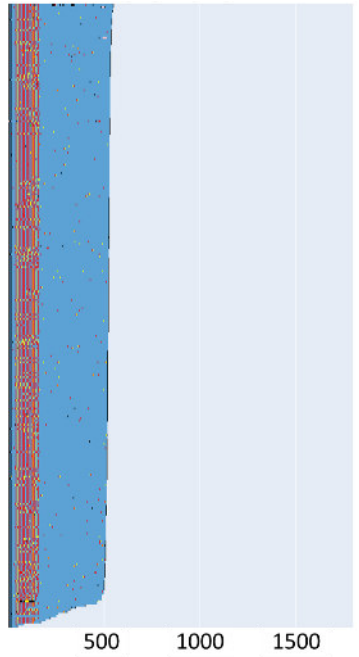


Pre-repeat



Interruptions





Single interruption

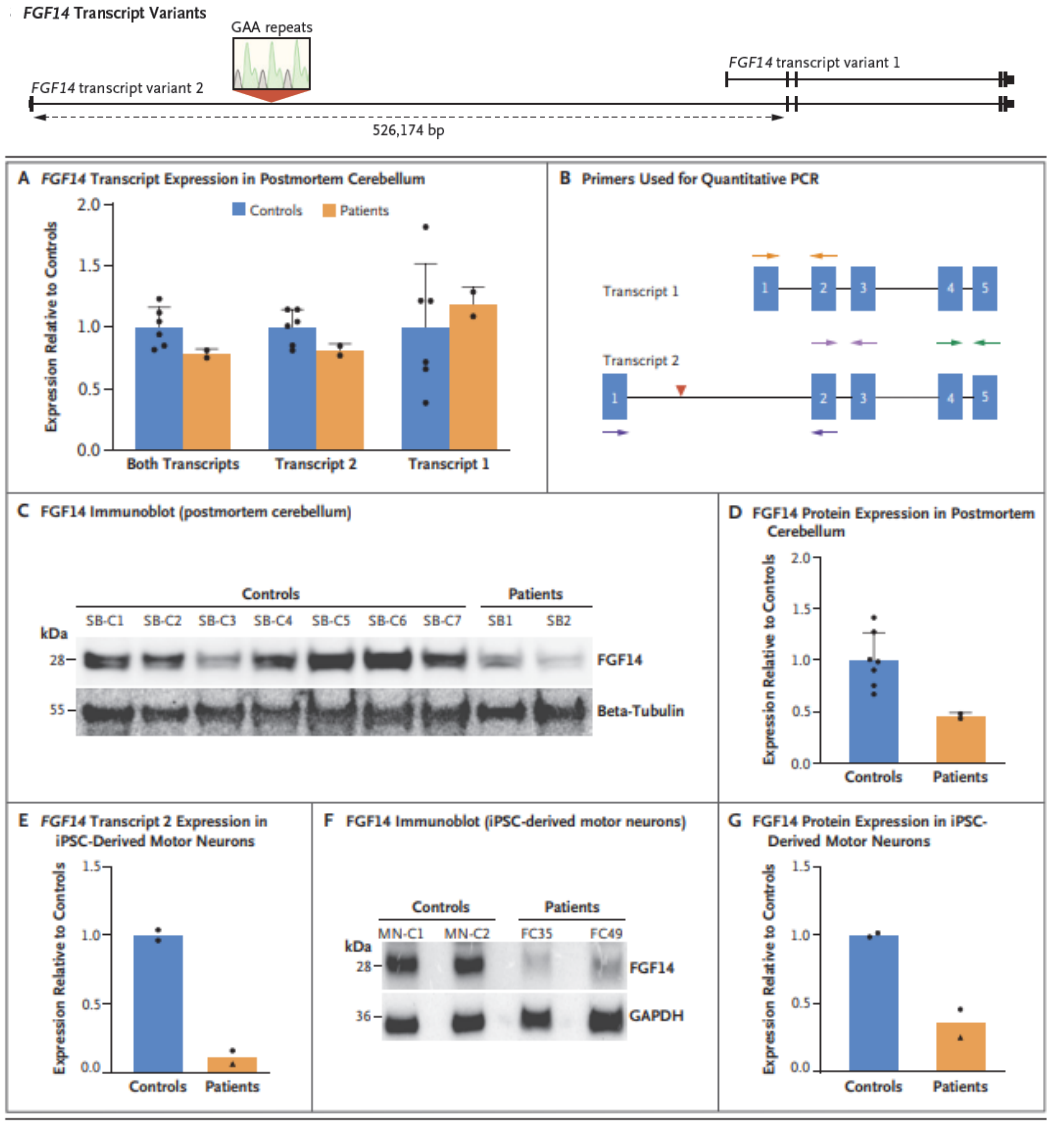
AAC
AAAAG
AAGAAGAAAG
ACG
AAAG
AGG
GAG
AAAGAAGAAGG
AAAGAAAAG
AAGG
(none)

Additional repeat

AAGAGG
AAGAAGACG
AAGAAAG
AAGACG
GAG
AAGAAGAAAG
AGG
AAGAAGAGG
AAGAAGAAGG
AAGGAG
AAGAAGAAGAGG
(none)

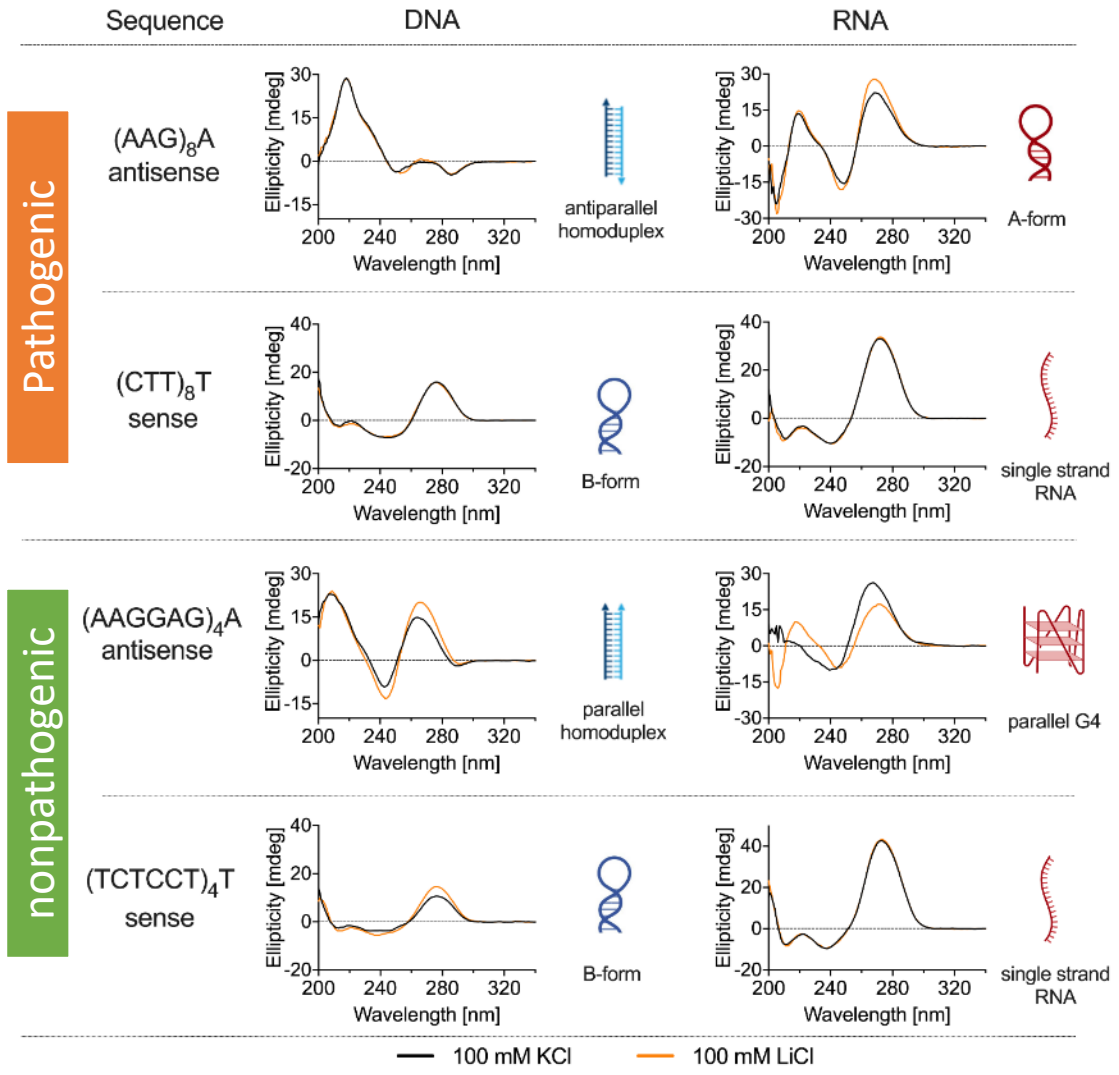
Mohren*, Erdlenbruch*, Kilpert*, Leitao*, *et al*. Nat Commun 2024

AAG & AAGGAG expansions form different secondary structures



Pellerin et al. NEJM 2022

Circular dichroism spectroscopy



Sebastian Hönes

Mohren*, Erdlenbruch*, Kilpert*, Leitao*, *et al*. Nat Commun 2024

Repeat expansions detected in EXPAND

Yes

No

Difficult

HD	**$(CAG)_n$**	***HTT***	✓
HDL2	$(CAG/CTG)_n$	*JPH3*	
SCA1	$(CAG)_n$	*ATXN1*	
SCA2	$(CAG)_n$	*ATXN2*	
SCA3	$(CAG)_n$	*ATXN3*	
SCA6	$(CAG)_n$	*CACNA1A*	
SCA7	$(CAG)_n$	*ATXN7*	
SCA17	$(CAG)_n$	*TBP*	
SBMA	**$(CAG)_n$**	***AR***	✓
DRPLA	$(CAG)_n$	*ATN1*	
SCA4	**$(CGG)_n$**	***ZFHX3***	✓

Dashnow et al. Genome Biology (2022) 23:257
https://doi.org/10.1186/s13059-022-02826-4

Genome Biology

METHOD **Open Access**

STRling: a k-mer counting approach that detects short tandem repeat expansions at known and novel loci

Harriet Dashnow[1], Brent S. Pedersen[1,2], Laurel Hiatt[1], Joe Brown[1], Sarah J. Beecroft[3,4], Gianina Ravenscroft[4], Amy J. LaCroix[5], Phillipa Lamont[6], Richard H. Roxburgh[7], Miriam J. Rodrigues[7,8], Mark Davis[9], Heather C. Mefford[5], Nigel G. Laing[4,9] and Aaron R. Quinlan[1*]

DM1	**(CTG)n**	***DMPK***
SCA8	**$(CTG/CAG)_n$**	***ATXN8OS/ATXN8***

FXS	**$(CGG)_{n>200}$**	***FMR1***
FXTAS	$(CGG)_{50\leq n\leq 200}$	*FMR1*
FRAXE	$(CCG)_n$	*AFF2*
FRA2A	$(CGG)_n$	*AFF3*
EPM1	**$(CCCCGCCCCGCG)_n$**	***CSTB***
SCA12	$(CAG)_n$	*PPP2R2B*
NIID/OPDM3	**(CGG)n**	***NOTCH2NLC***
OPDM1	$(CGG)_n$	*LRP12*
OPDM2	$(CGG)_n$	*GIPC1*
OPDM4	$(CGG)_n$	*RILPL1*
OPDM5	$(CGG)_n$	*ABCD3*
BSS	$(CGG)_n$	*XYLT1*
GDPAG	$(GCA)_n$	*GLS*

ALS/FTD	$(GGGGCC)_n$	*C9ORF72*
FECD3	**$(CTG)_n$**	***TCF4***
DM2	**$(CAG)_n$**	***CNBP***
FRDA	**$(GAA)_n$**	***FXN***
SCA27B	**(GAA)n**	***FGF14***
SCA10	$(ATTCT)_n$	*ATXN10*
SCA31	**$(TGGAA)_n$**	***BEAN***
SCA36	**$(GGCCTG)_n$**	***NOP56***
CANVAS	**$(AAGGG)_n$**	***RFC1***
SCA37	**$(ATTTT/ATTTC)_n$**	***DAB1***
FAME1/6/7	$(ATTTT/ATTTC)_n$	*SAMD12, TNRC6A, RAPGEF2*
FAME2/3/4/8	**$(ATTTT/ATTTC)_n$**	***STARD7, MARCHF6***, *YEATS2, RAI1*
XDP	$(CCCTCT)_n$	*TAF1* (SVA)

Adapted from Depienne & Mandel, 2021

Identification of tandem repeat EXPAnsions in unsolved Neurological Disorders (EXPAND)

Short read genome sequencing

- 250 short-read genomes
- phenotypes including
  - **cerebellar ataxia** (~60)
  - **tremor** (~30)
  - **dystonia** (~10)
  - **leukencephalopathy** (~10)
  - Epilepsy (~30)
  - Aicardi-syndrome (4)
  - Septo-optic dysplasia
  - Tourette syndrome (~20)
  - Motor neuron disease
  - Neurodevelopmental disorders

- Strelka (SNVs, indels), Manta (SVs)
- STR analysis

Long read nanopore sequencing

- 100 long-read genomes (/130)
- Optimization of HMW DNA extraction
- >15X, N50 > 40kb

- wf-human-variation pipeline (Snakemake)
- Clair3 (SNVs, indels), Sniffles (SVs)
- Straglr (STR)
- methylation, phasing

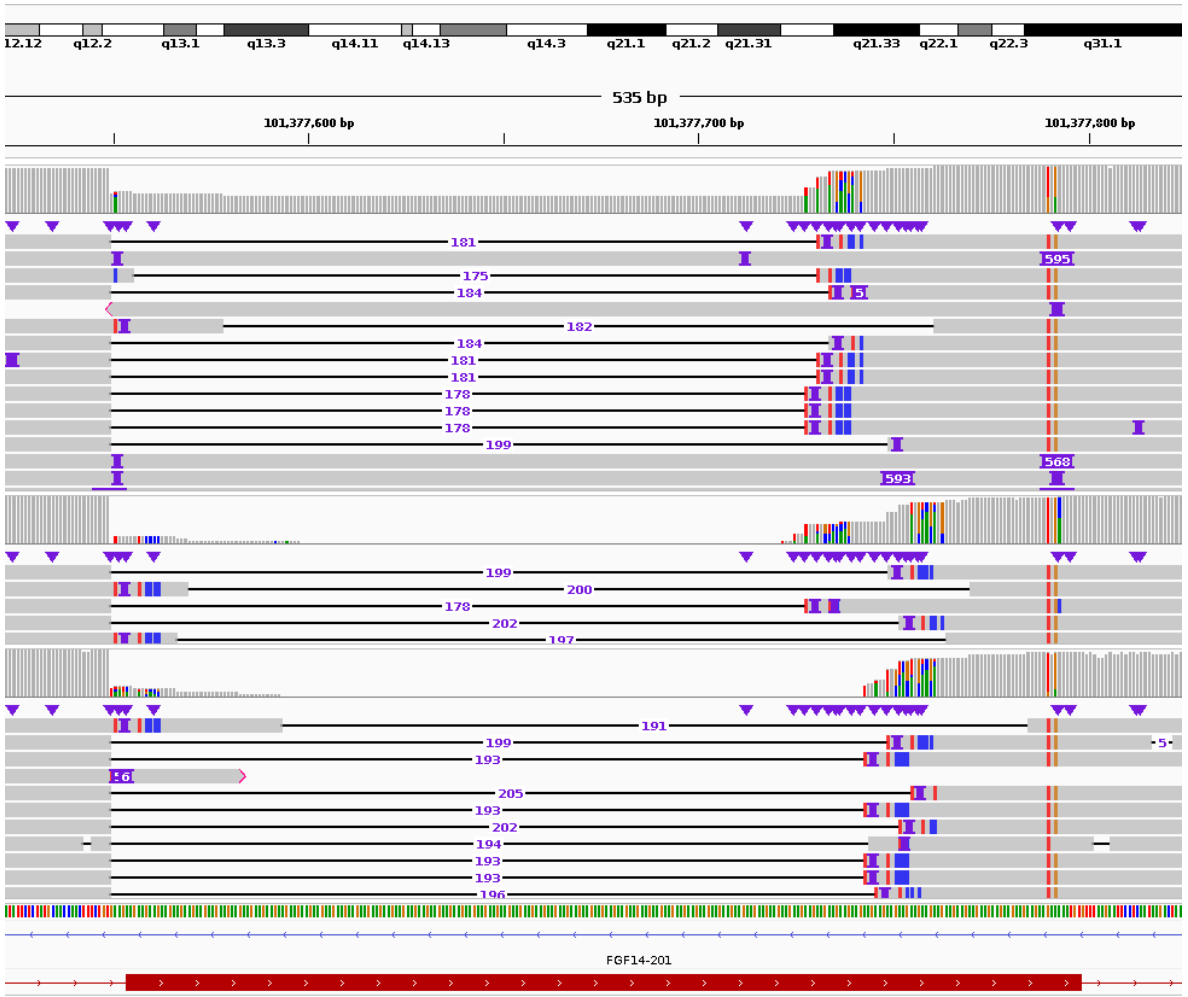
RNA sequencing

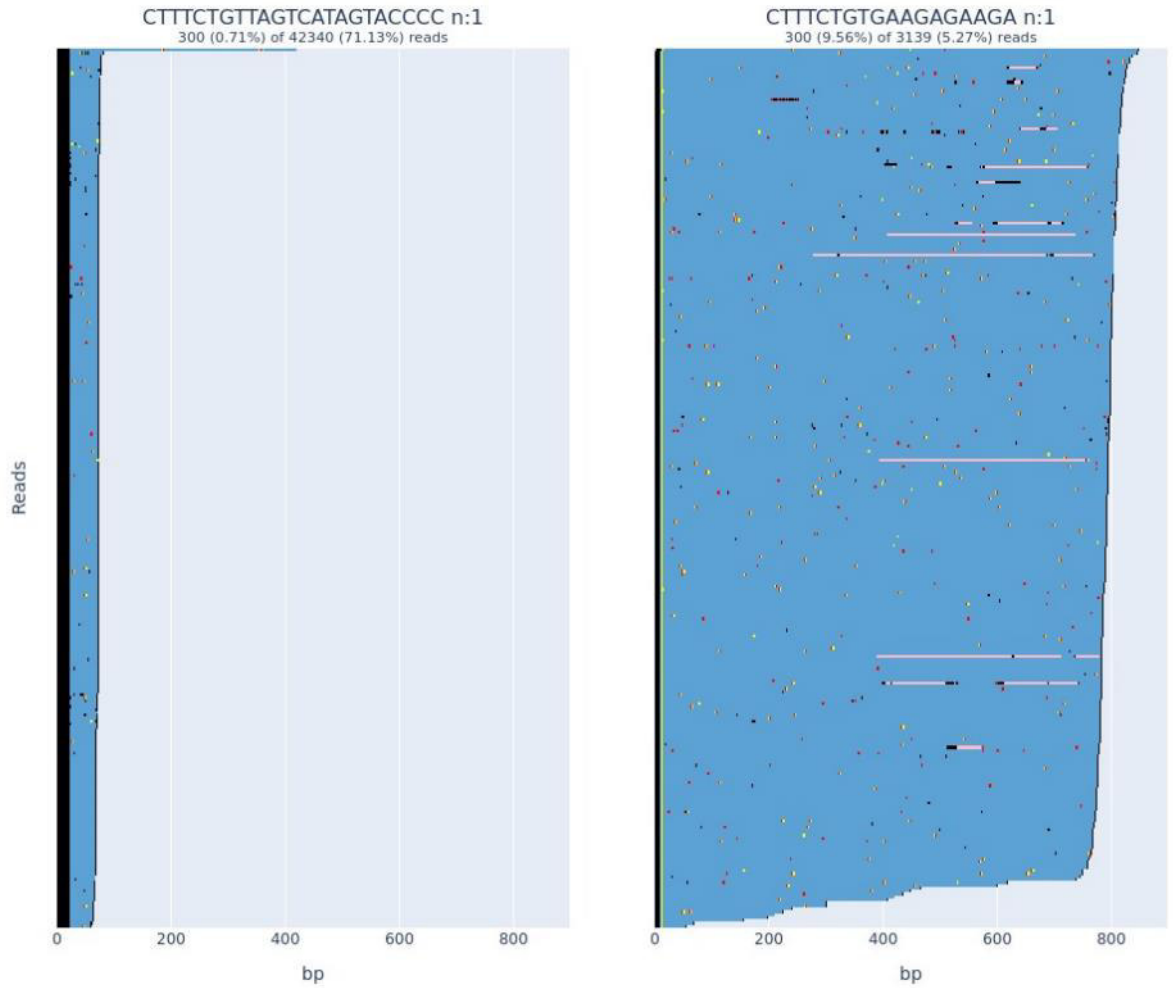
DFG Deutsche Forschungsgemeinschaft

Cologne Center for Genomics

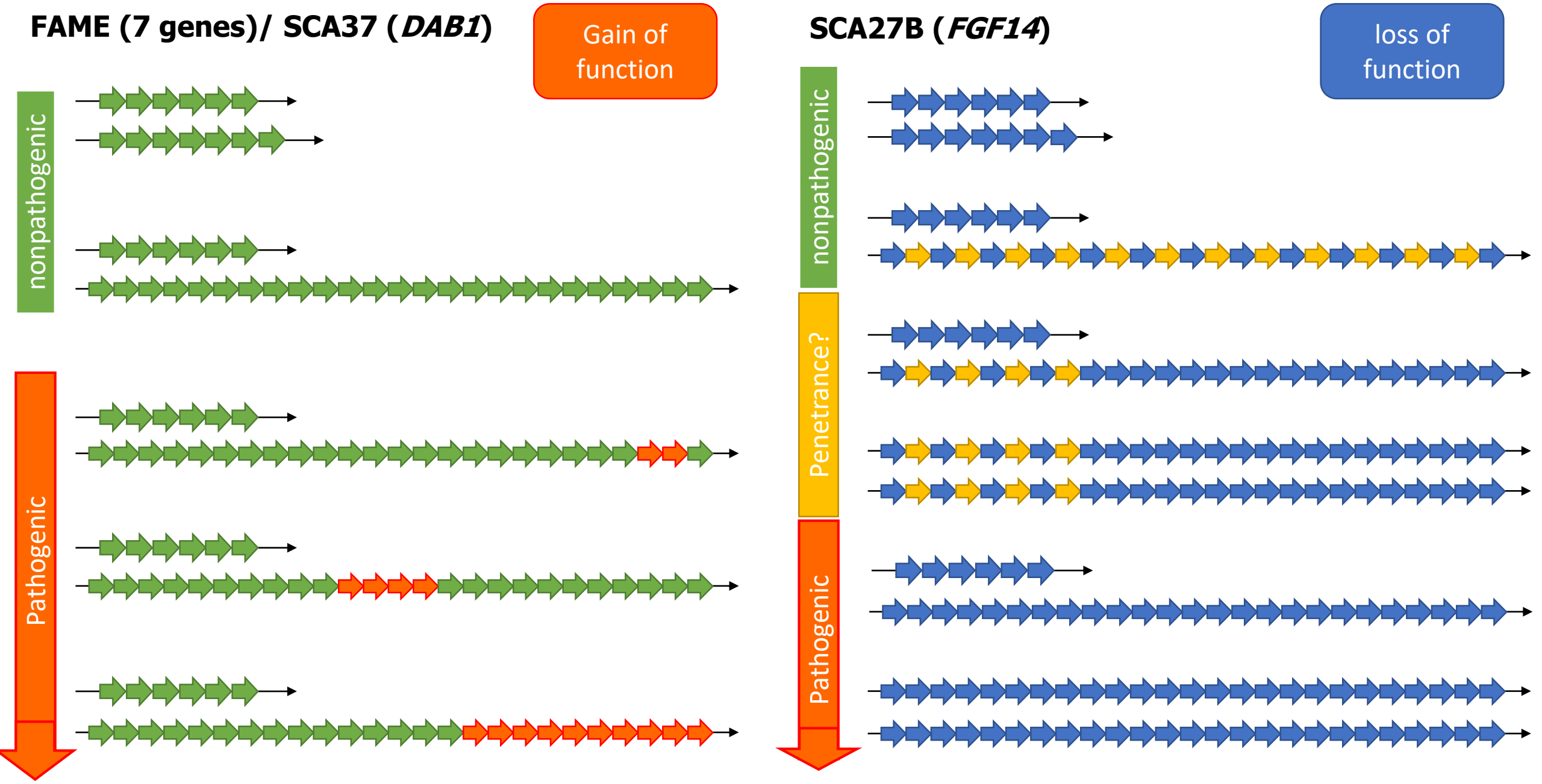
WEST GERMAN GENOME CENTER Düsseldorf

Detection of *FGF14* expansions by nanopore genome sequencing

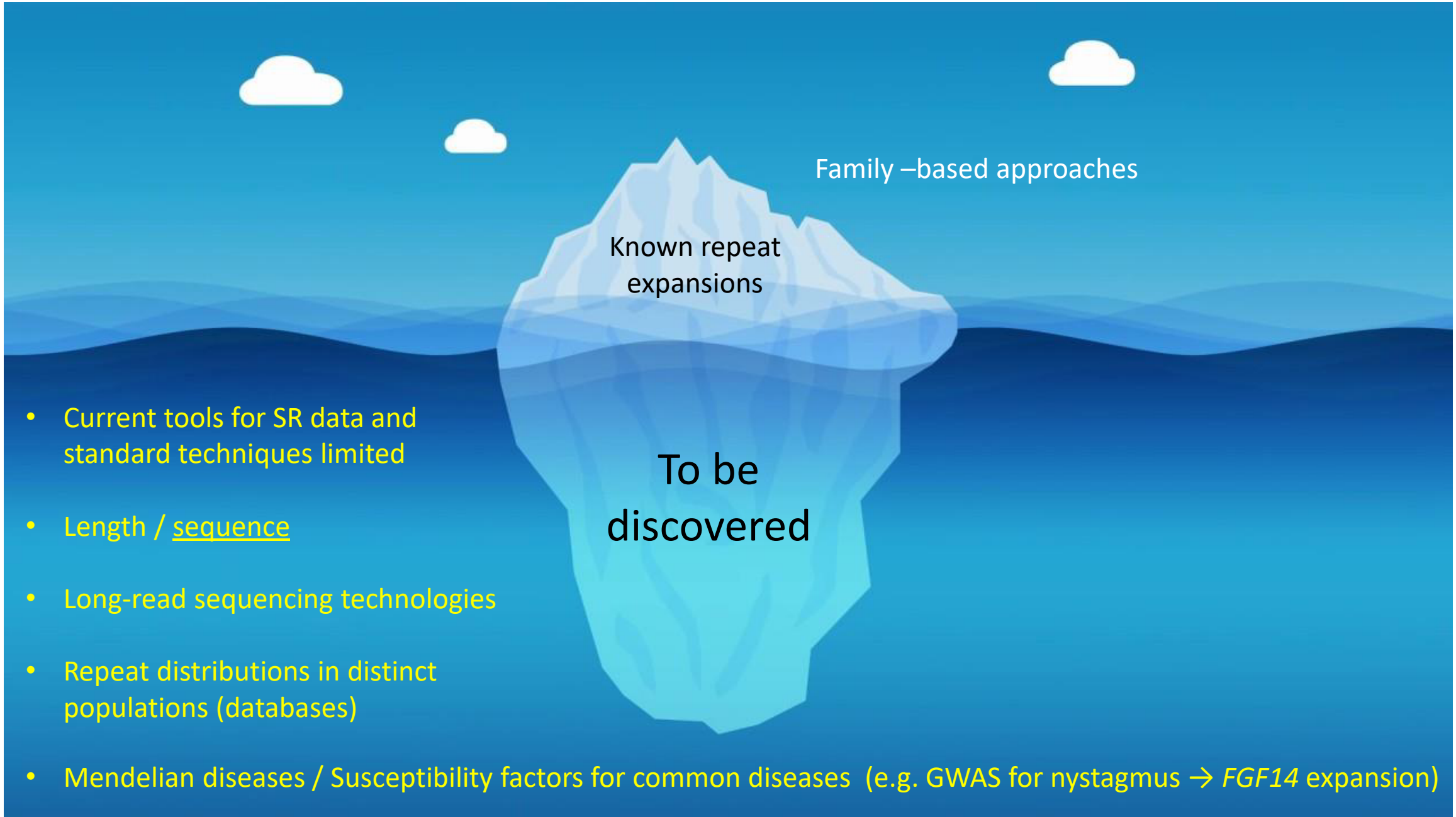




Conclusions (1)



Conclusions (2)



- Current tools for SR data and standard techniques limited
- Length / sequence
- Long-read sequencing technologies
- Repeat distributions in distinct populations (databases)
- Mendelian diseases / Susceptibility factors for common diseases (e.g. GWAS for nystagmus → *FGF14* expansion)