



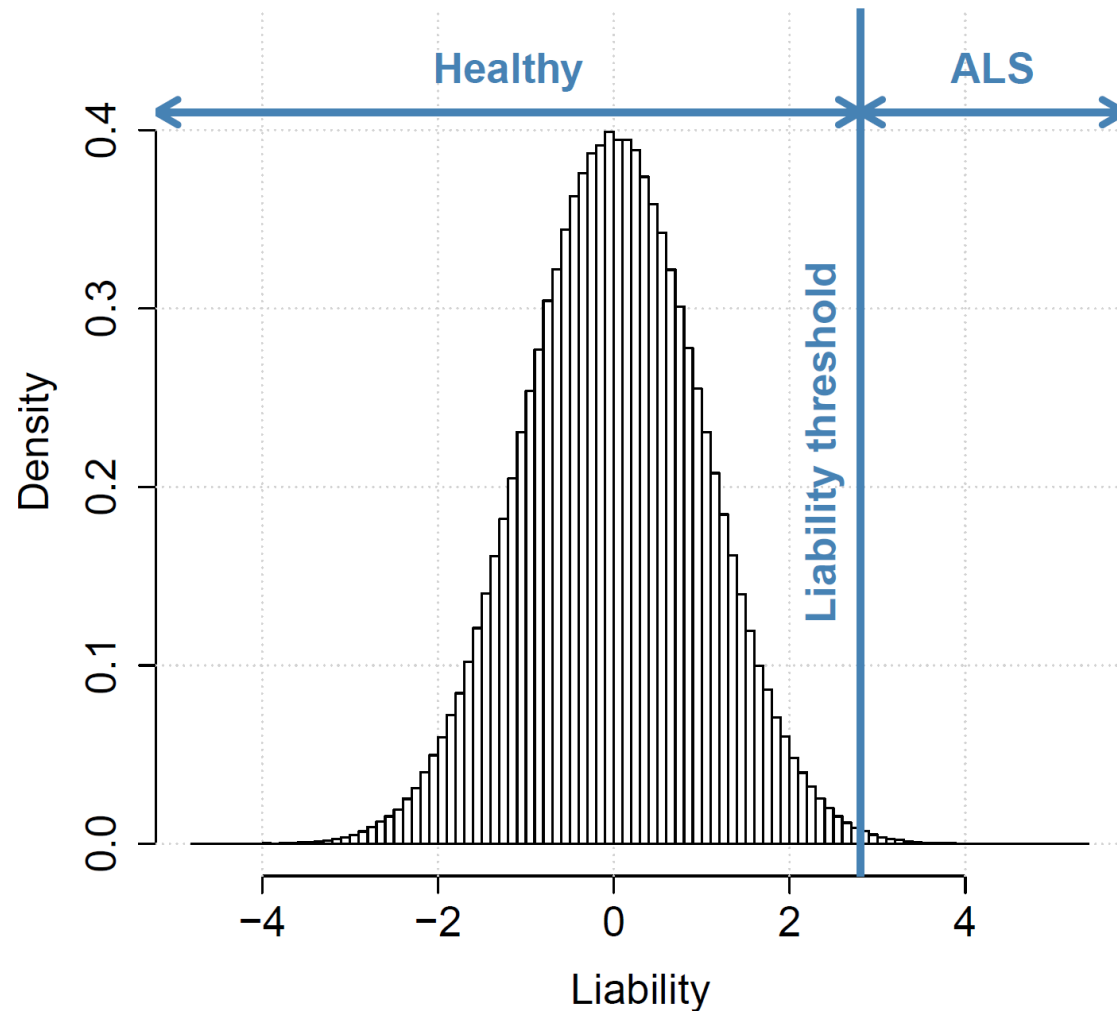
The panorama of ALS genomics

ENCALS 2017 | Ljubljana | 19.05.2017

Russell McLaughlin | mclaugr@tcd.ie | [@RSLMcL](https://twitter.com/RSLMcL)

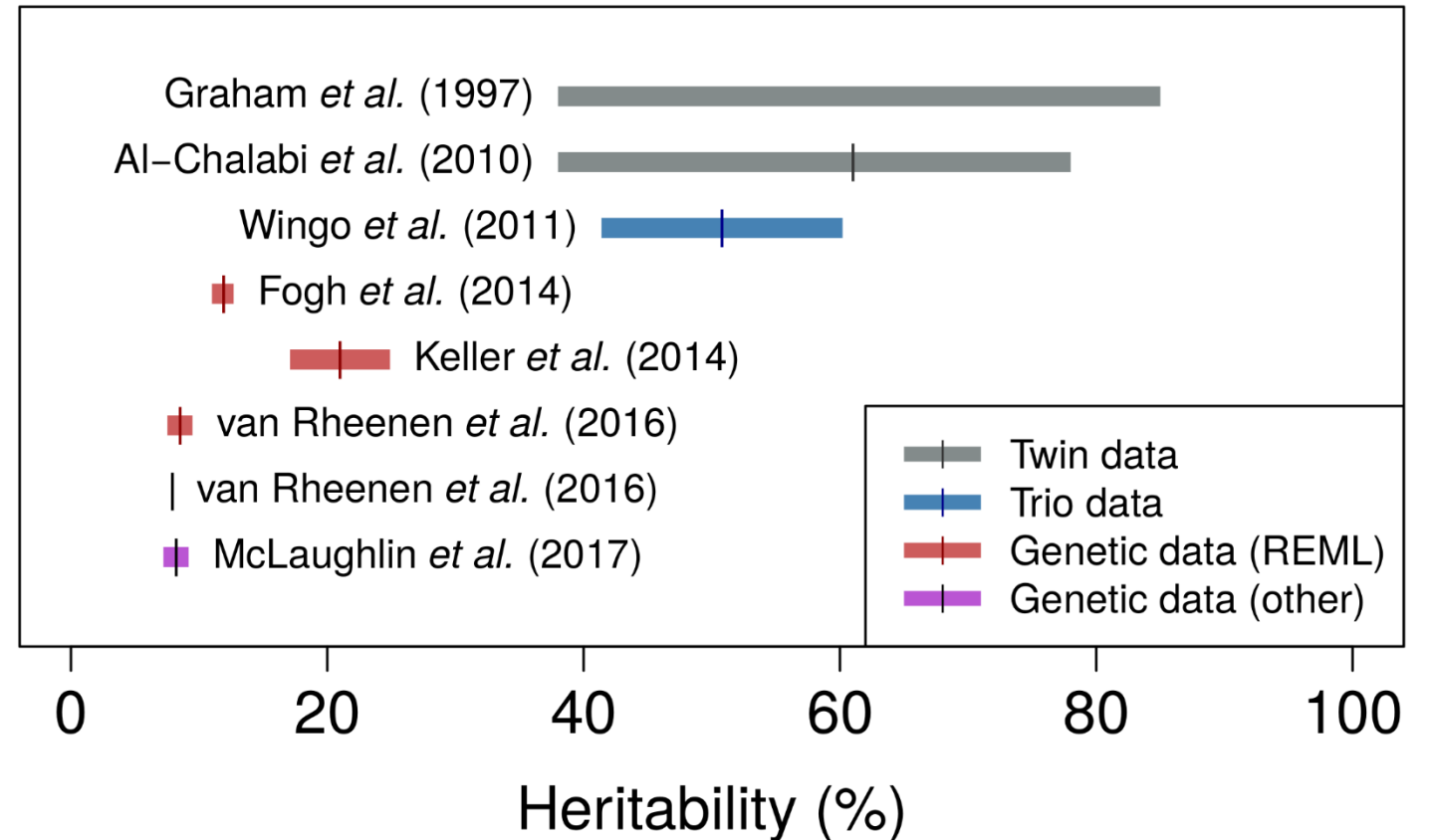
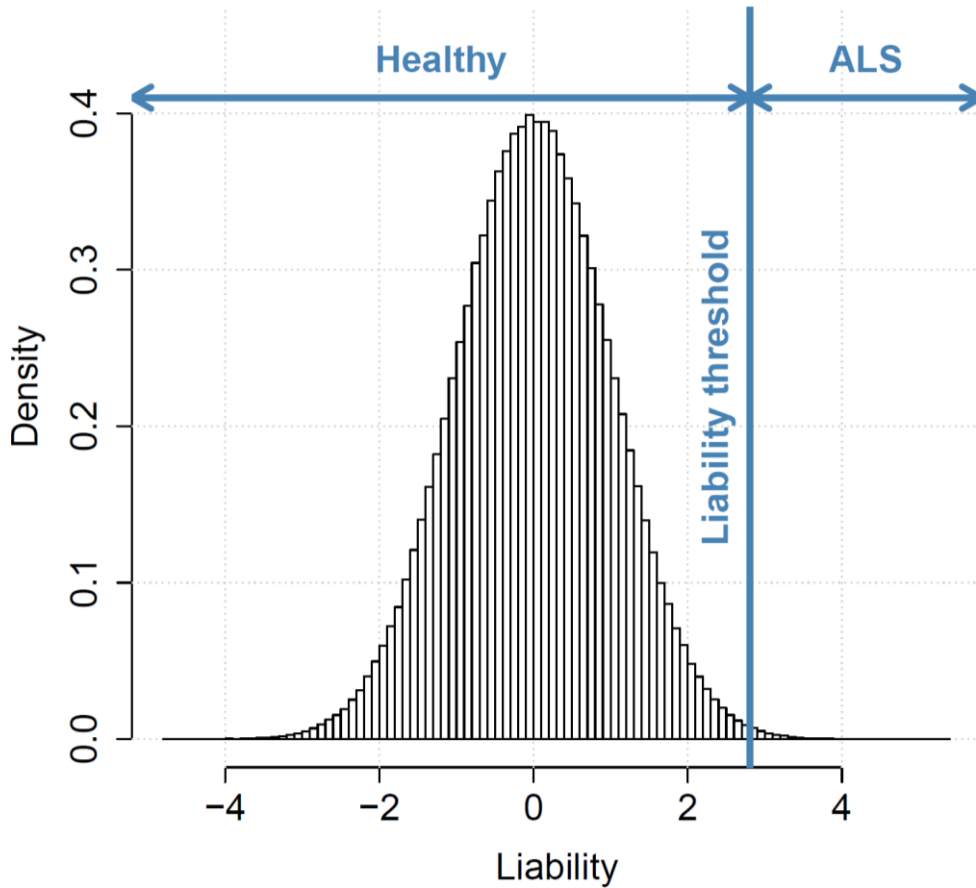
Is ALS genetic?

Liability to develop ALS - explain liability on this slide



Is ALS genetic?

Heritability: the proportion of variance in liability conferred by genetic variation

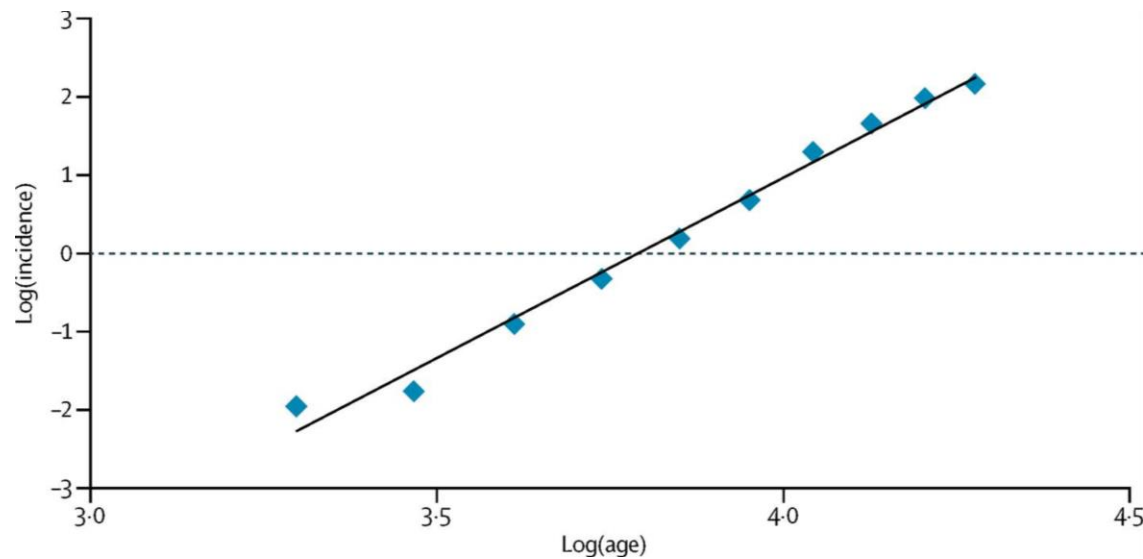


Non-genetic risk

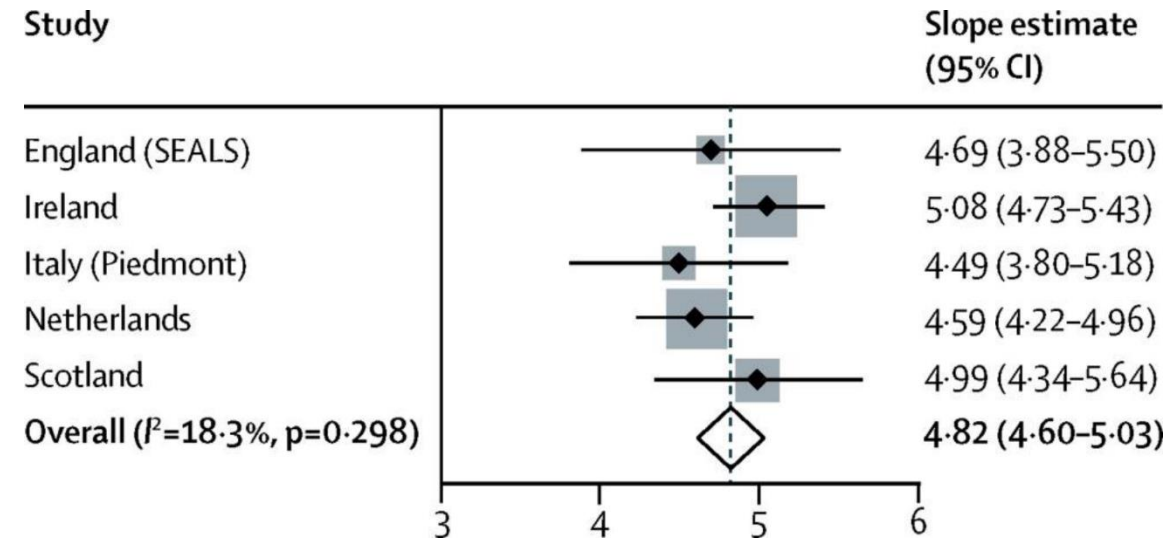
Multiple lifetime steps/exposures are required to develop ALS

Age-specific incidence Steps Time (age)

$$\log i = (n - 1) \log t + \log \prod_{i=1}^n u_n$$

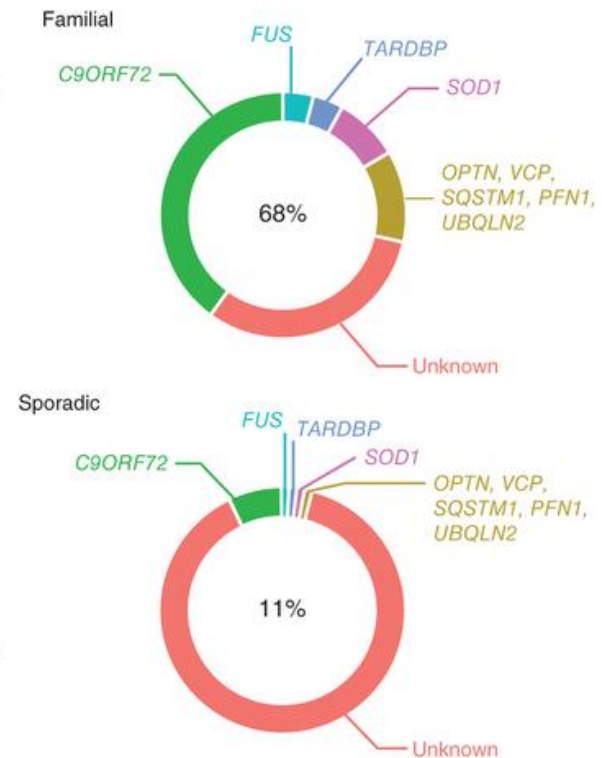
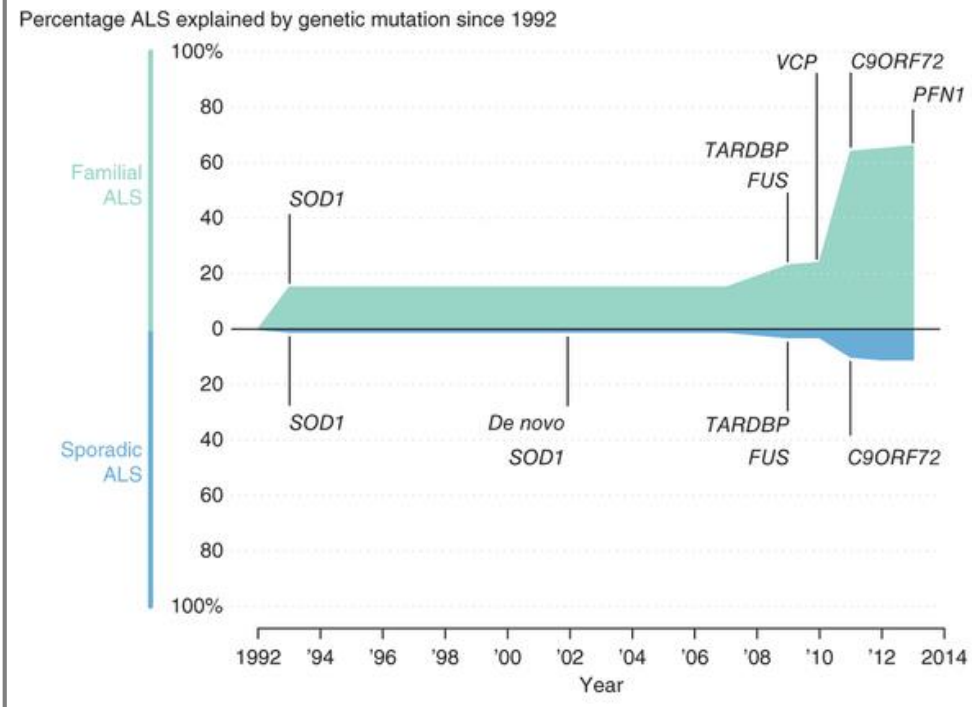
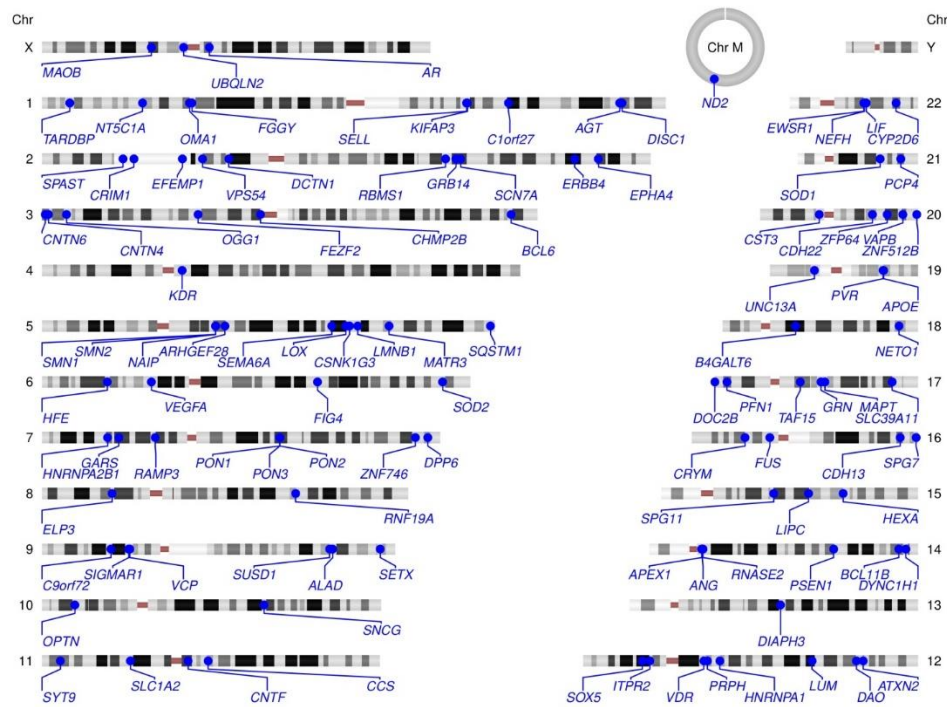


Al-Chalabi *et al.* (2014) *Lancet Neurol* 13(11):1108-1113



Genetics of ALS

Genes that have been investigated in ALS



Renton *et al.* (2013) Nature Neuroscience 17:17-23

McLaughlin *et al.* (2015) In *Movement Disorder Genetics*, Springer

How do we discover new ALS genes?

Genome-wide association study (GWAS)

Person 1 ... TCAGCCATGCTACT**C**GATCGACTAAG**G**CG ... (maternal)
 ... TCAGCCATGCTACT**C**GATCGACTAATCG ... (paternal)

Person 2 ... TCAGCCATGCTACT**C**GATCGACTAAG**G**CG ... (maternal)
 ... TCAGCCATGCTACT**T**GATCGACTAATCG ... (paternal)

Person 3 ... TCAGCCATGCTACT**T**GATCGACTAATCG ... (maternal)
 ... TCAGCCATGCTACT**T**GATCGACTAATCG ... (paternal)

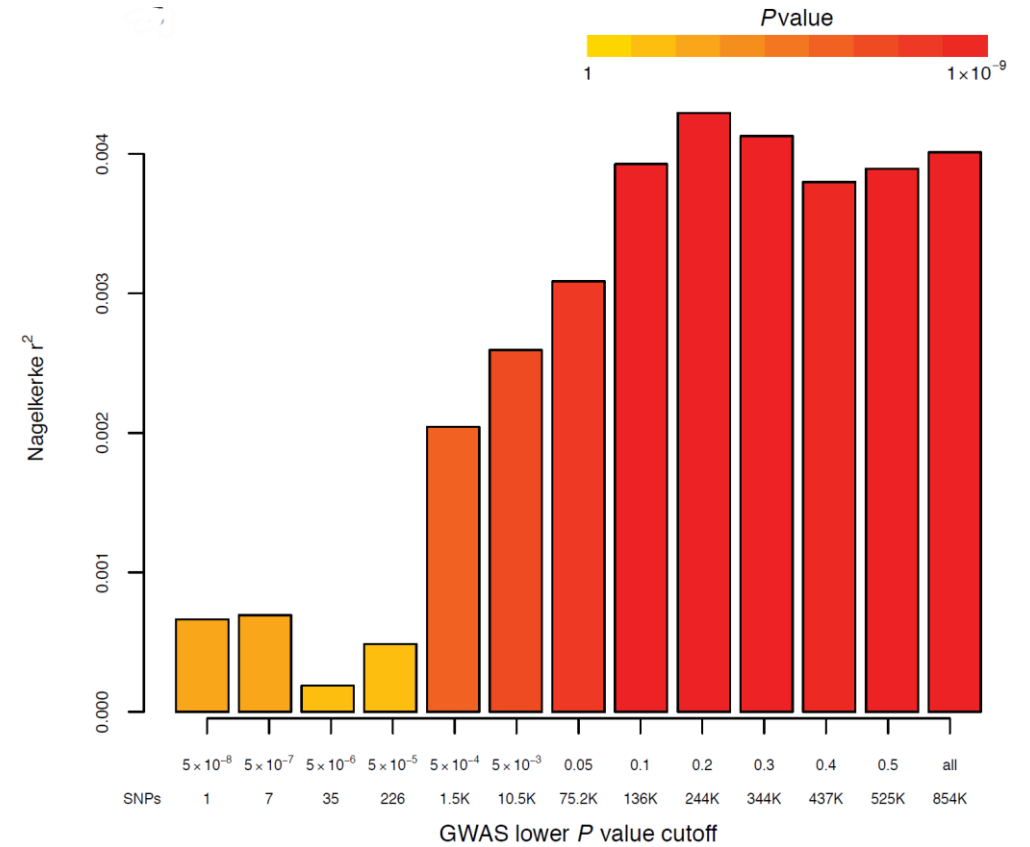
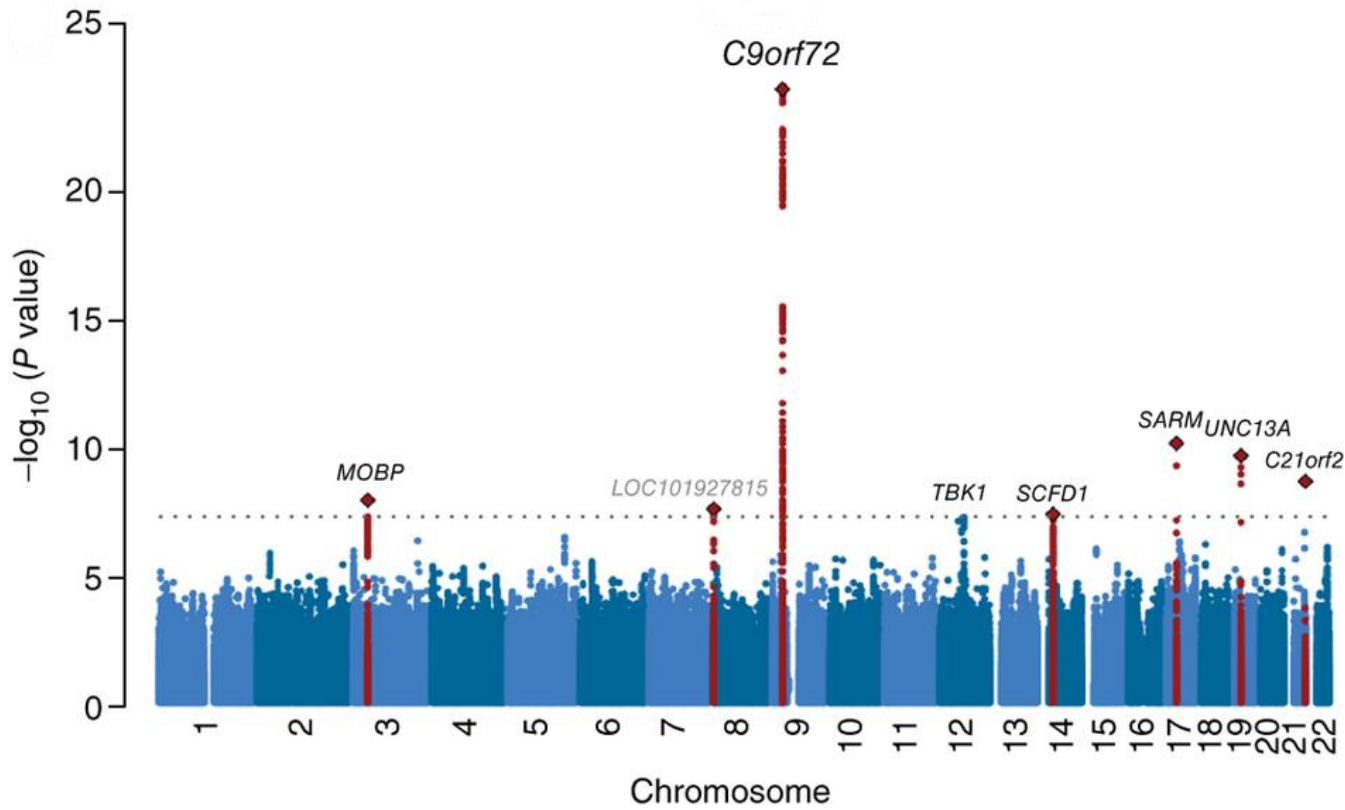
↑
 Single nucleotide polymorphism
 (SNP)

If only present in one individual, or if rare	If rare and observed in disease context	If common in the population
Variant or single nucleotide variant (SNV)	Mutation	Single nucleotide polymorphism (SNP)

SNP1	SNP2	SNP...
Cases	Cases	<i>Repeat for all SNPs</i>
Count of G: 2104 of 4000	Count of G: 1648 of 4000	
Frequency of G: 52.6%	Frequency of G: 41.2%	
Controls	Controls	
Count of G: 2676 of 6000	Count of G: 2532 of 6000	
Frequency of G: 44.6%	Frequency of G: 42.2%	
P-value: $5.0 \cdot 10^{-15}$	P-value: 0.33	

2016 ALS GWAS

13 countries; 12,577 ALS cases; 23,475 healthy controls



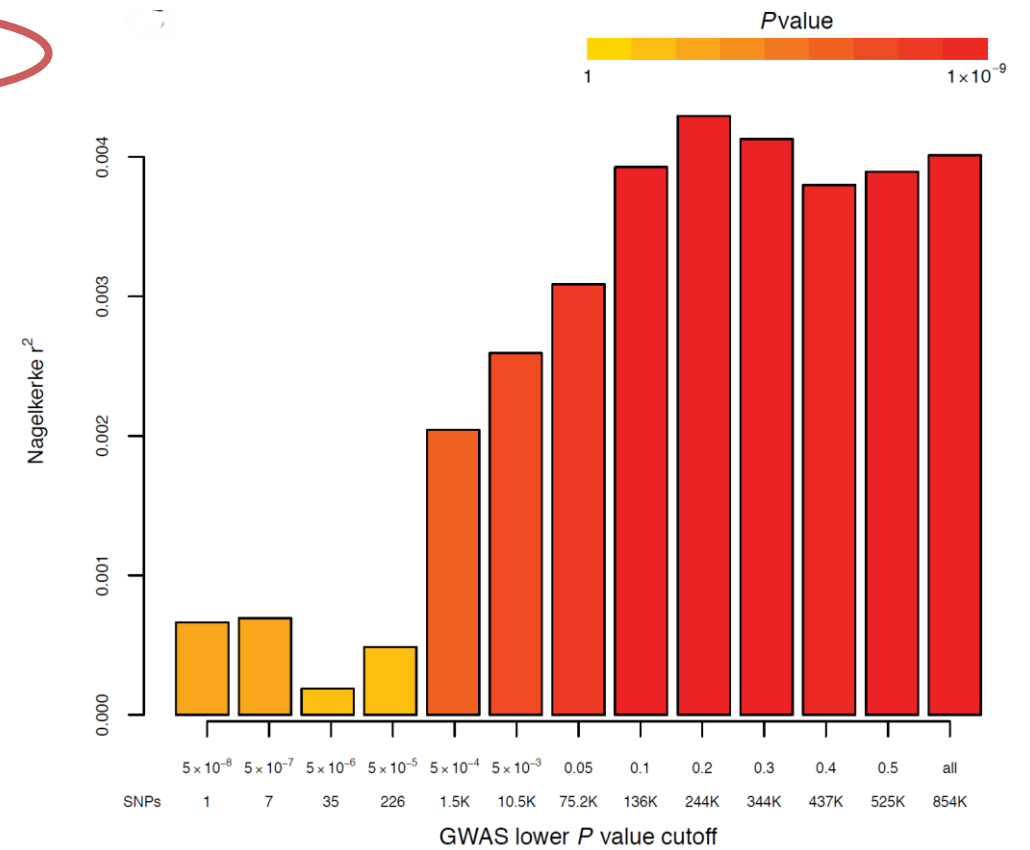
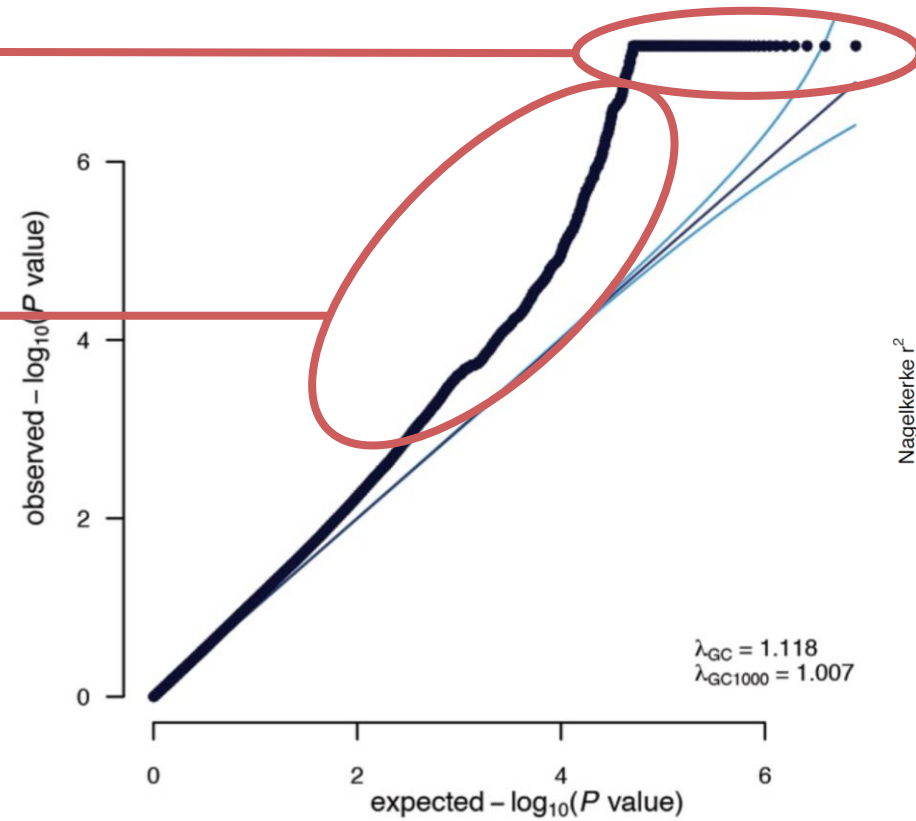
van Rheenen *et al.* (2016) *Nat Genet* 48(9):1043-8

2016 ALS GWAS

13 countries; 12,577 ALS cases; 23,475 healthy controls

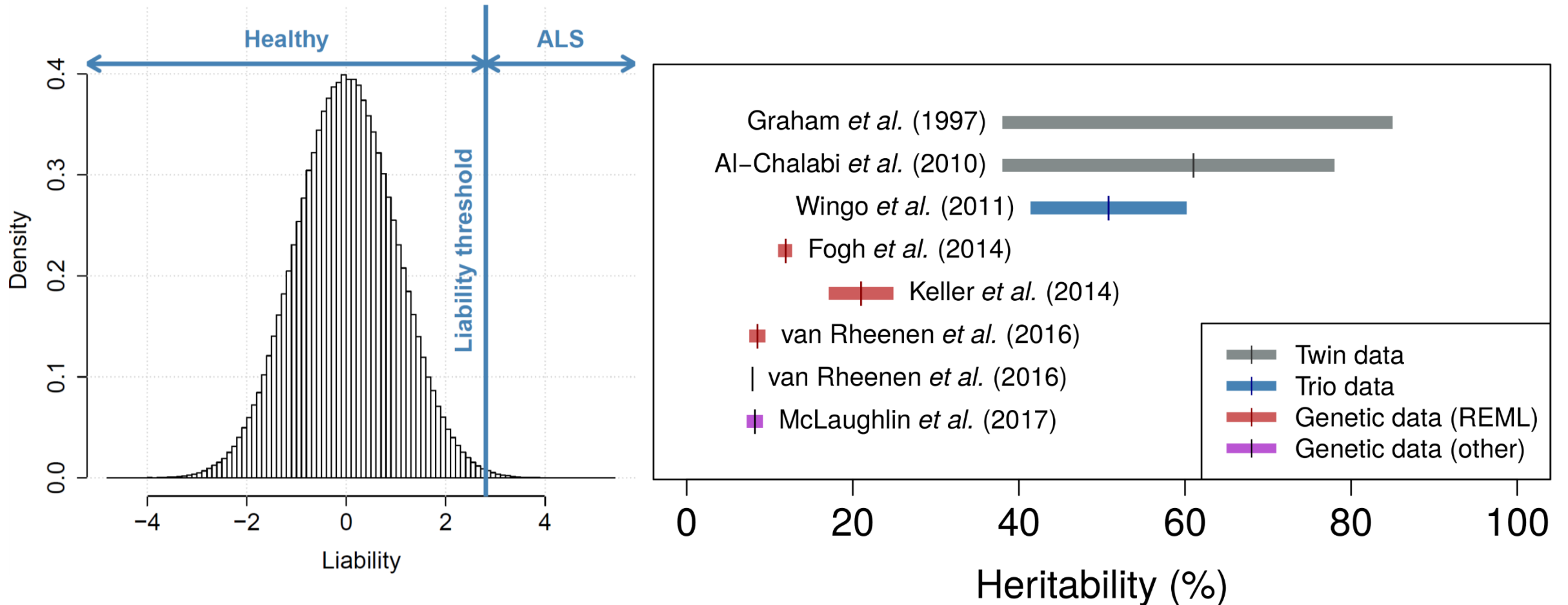
Significantly-associated SNPs
($p < 5 \times 10^{-8}$)

Sub-threshold SNPs
(more associated than they should be)



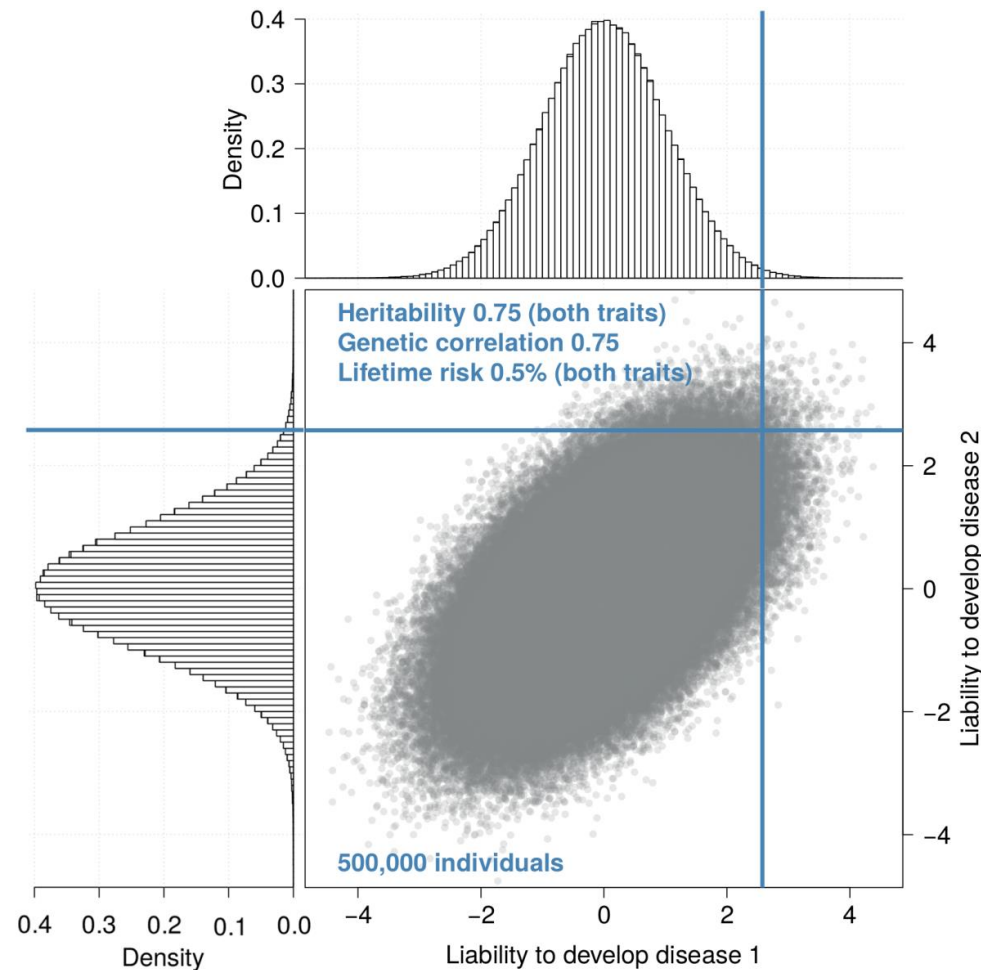
Heritability revisited

SNP-based heritability (polygenic risk)



Genetic correlation

Covariance between two traits due to shared genetic variation



ORIGINAL ARTICLE

Aggregation of Neurologic and Neuropsychiatric Disease in Amyotrophic Lateral Sclerosis Kindreds: A Population-Based Case–Control Cohort Study of Familial and Sporadic Amyotrophic Lateral Sclerosis

Susan Byrne, PhD,^{1,2} Mark Heverin, MSc,² Marwa Elamin, PhD,² Peter Bede, MD,^{1,2} Catherine Lynch, MSc,¹ Kevin Kenna, BSc,³ Russell MacLaughlin, PhD,¹ Cathal Walsh, PhD,⁴ Ammar Al Chalabi, PhD,⁵ and Orla Hardiman, FRCPI^{1,2}

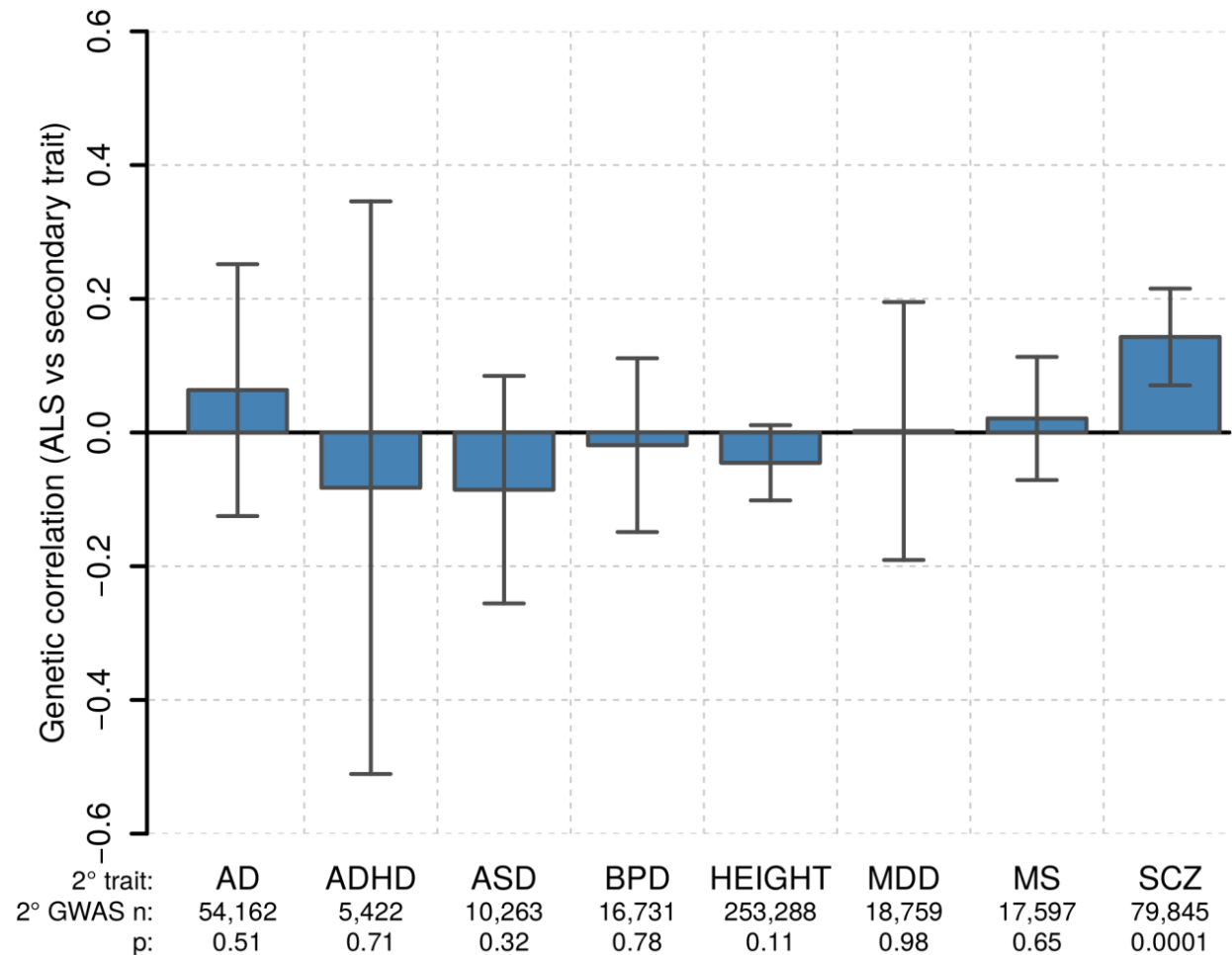
TABLE 5. Relatives of C9-Positive Cases and C9-Negative Cases Compared to Controls in a Cox Regression Proportional Model

Disease	Relatives	HR	95% CI	<i>p</i>
Parkinson disease	Relatives of C9-positive patients	1.3	0.5–3.7	0.570
	Relatives of C9-negative patients	0.7	0.4–1.1	0.126
Dementia	Relatives of C9-positive patients	1.6	1.1–2.4	0.017 ^a
	Relatives of C9-negative patients	1.2	0.9–1.4	0.100
Depression	Relatives of C9-positive patients	3.3	1.6–7.0	0.002 ^a
	Relatives of C9-negative patients	0.6	0.3–1.1	0.075
Schizophrenia/psychotic illness	Relatives of C9-positive patients	9.9	4.8–20.5	<0.0001 ^a
	Relatives of C9-negative patients	3.9	2.4–6.5	<0.0001 ^a
Suicide	Relatives of C9-positive patients	16.6	5.6–49.4	<0.0001 ^a
	Relatives of C9-negative patients	5.1	2.2–12.1	<0.0001 ^a

^aStatistically significant.
CI = confidence interval; HR = hazard ratio.

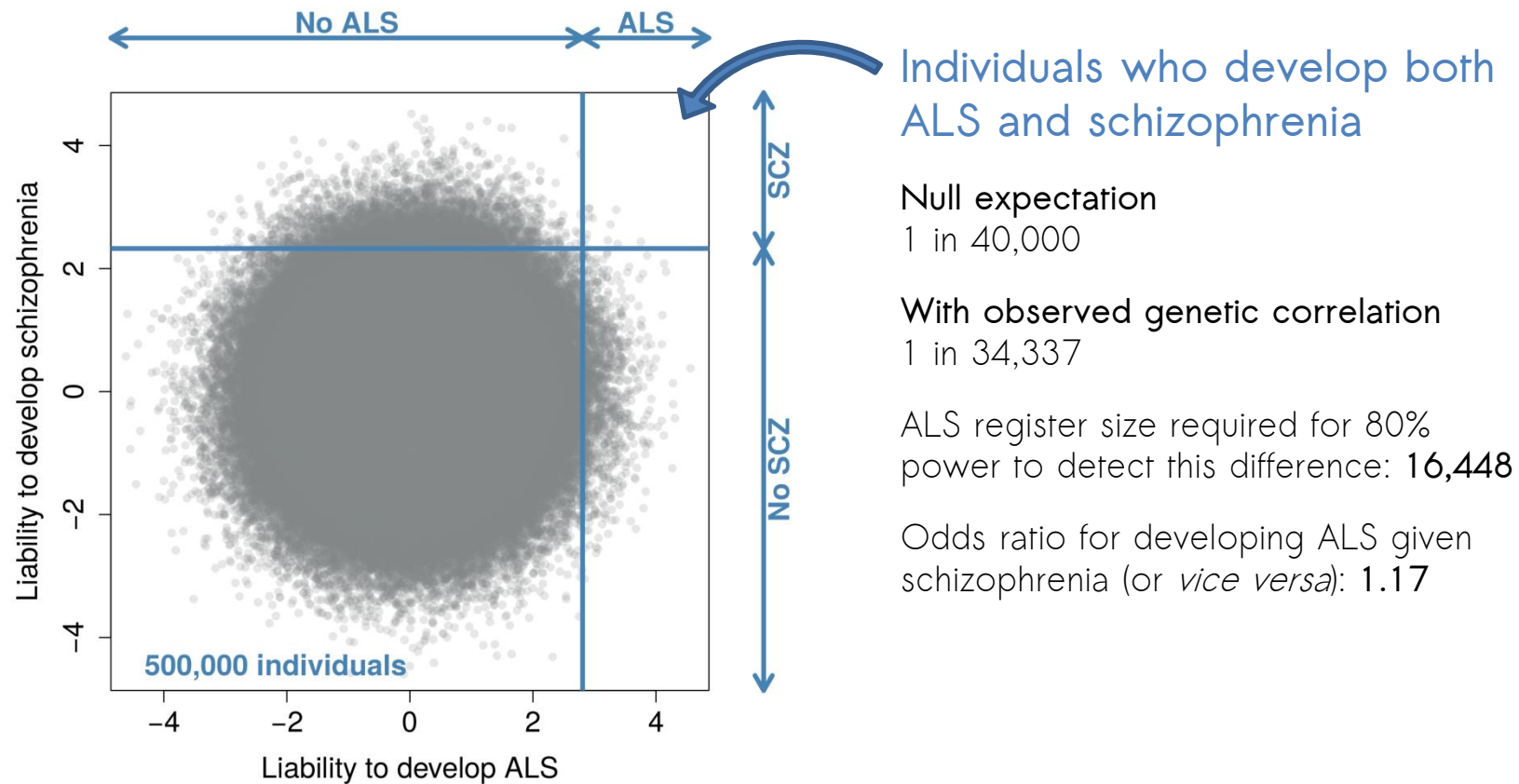
ALS and schizophrenia are genetically correlated

Genetic correlation of 14.3% refers to polygenic components of both diseases



ALS and schizophrenia are genetically correlated

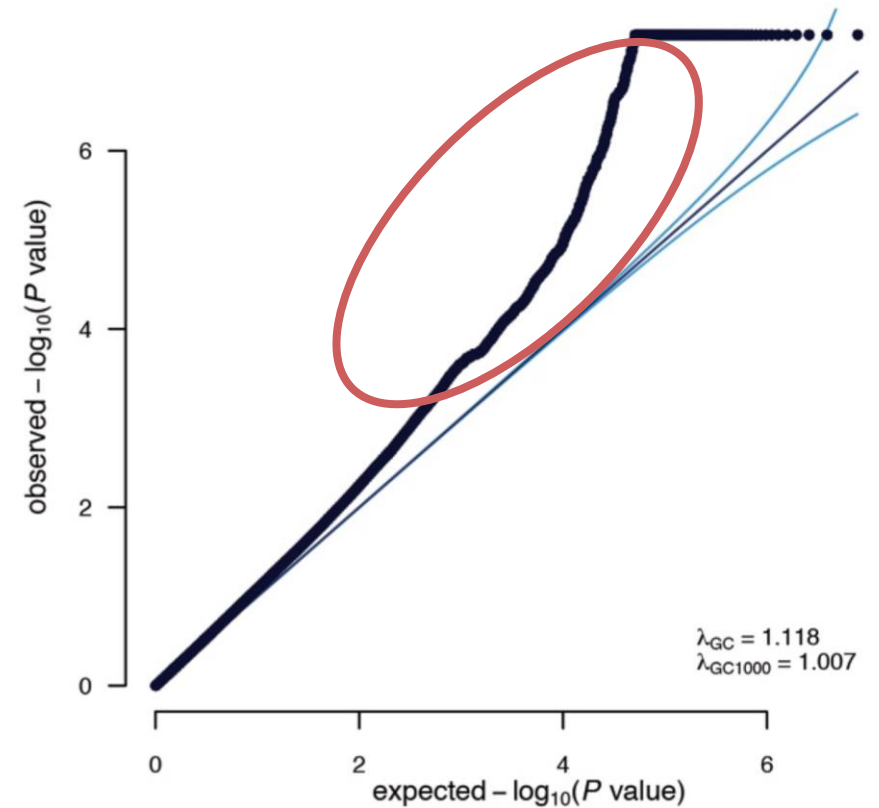
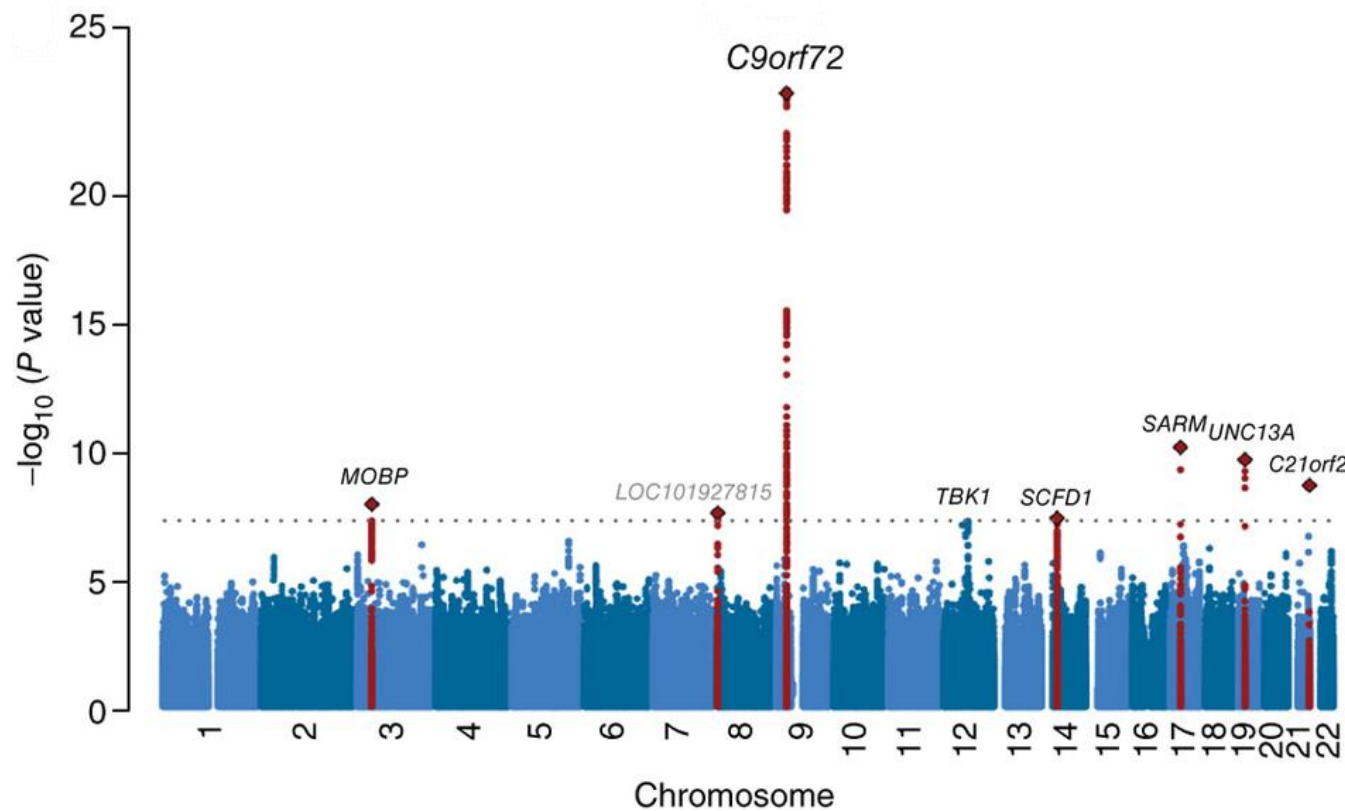
Implications



GWAS in ALS

General conclusions and further considerations

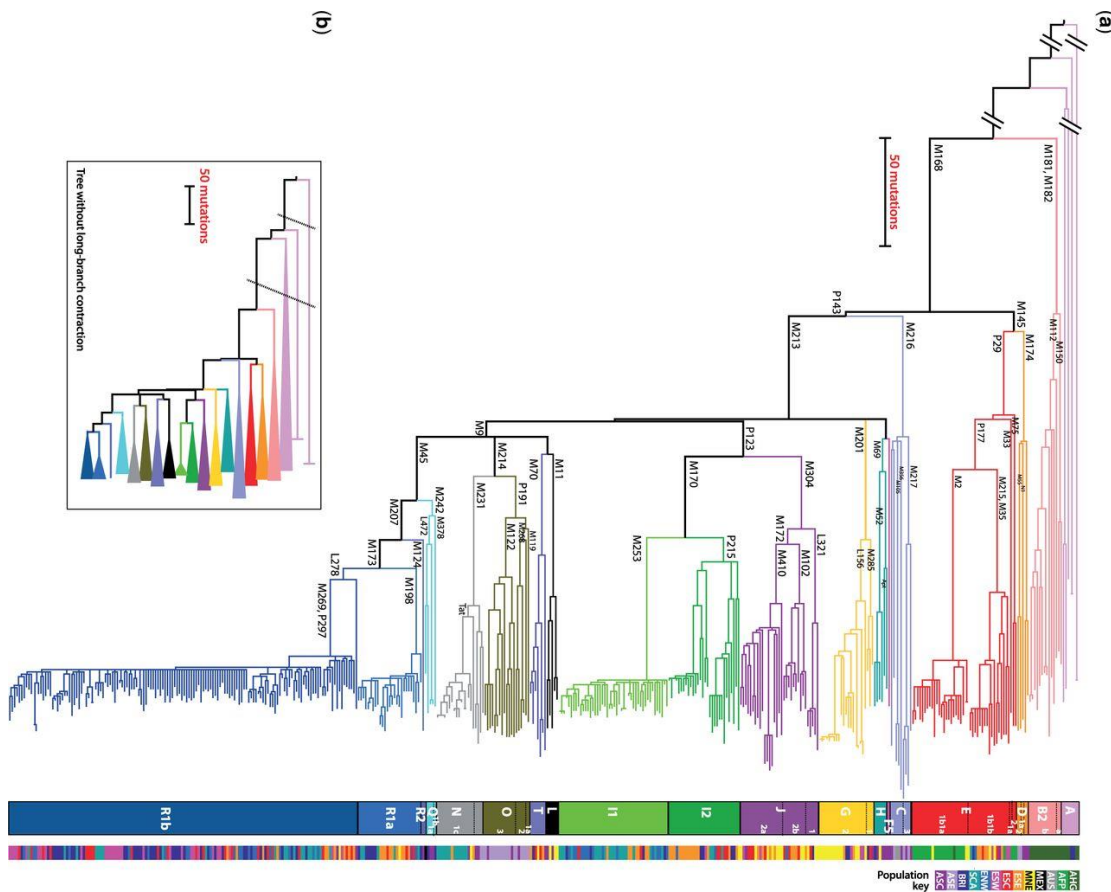
1. Bigger GWAS will discover more disease loci



GWAS in ALS

General conclusions and further considerations

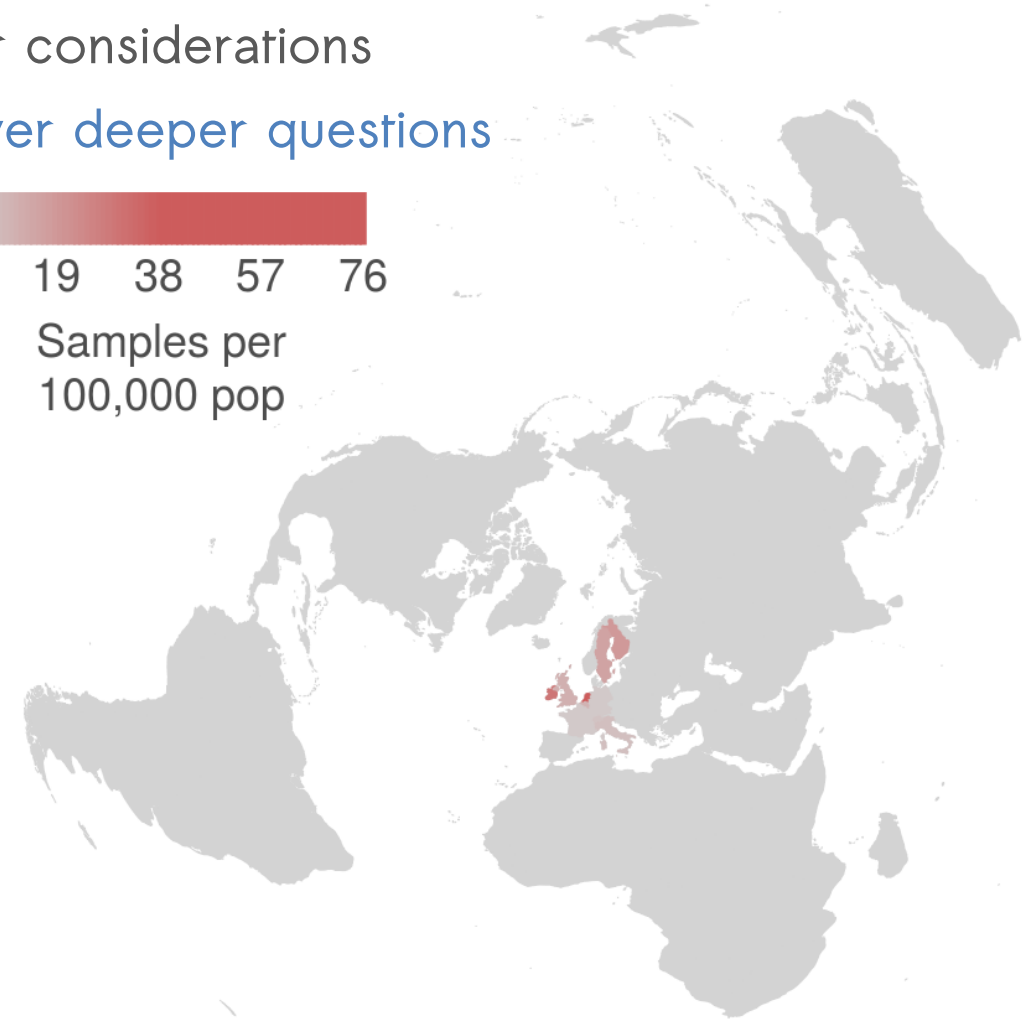
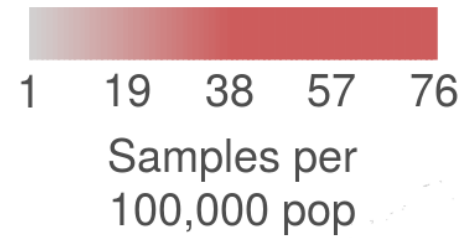
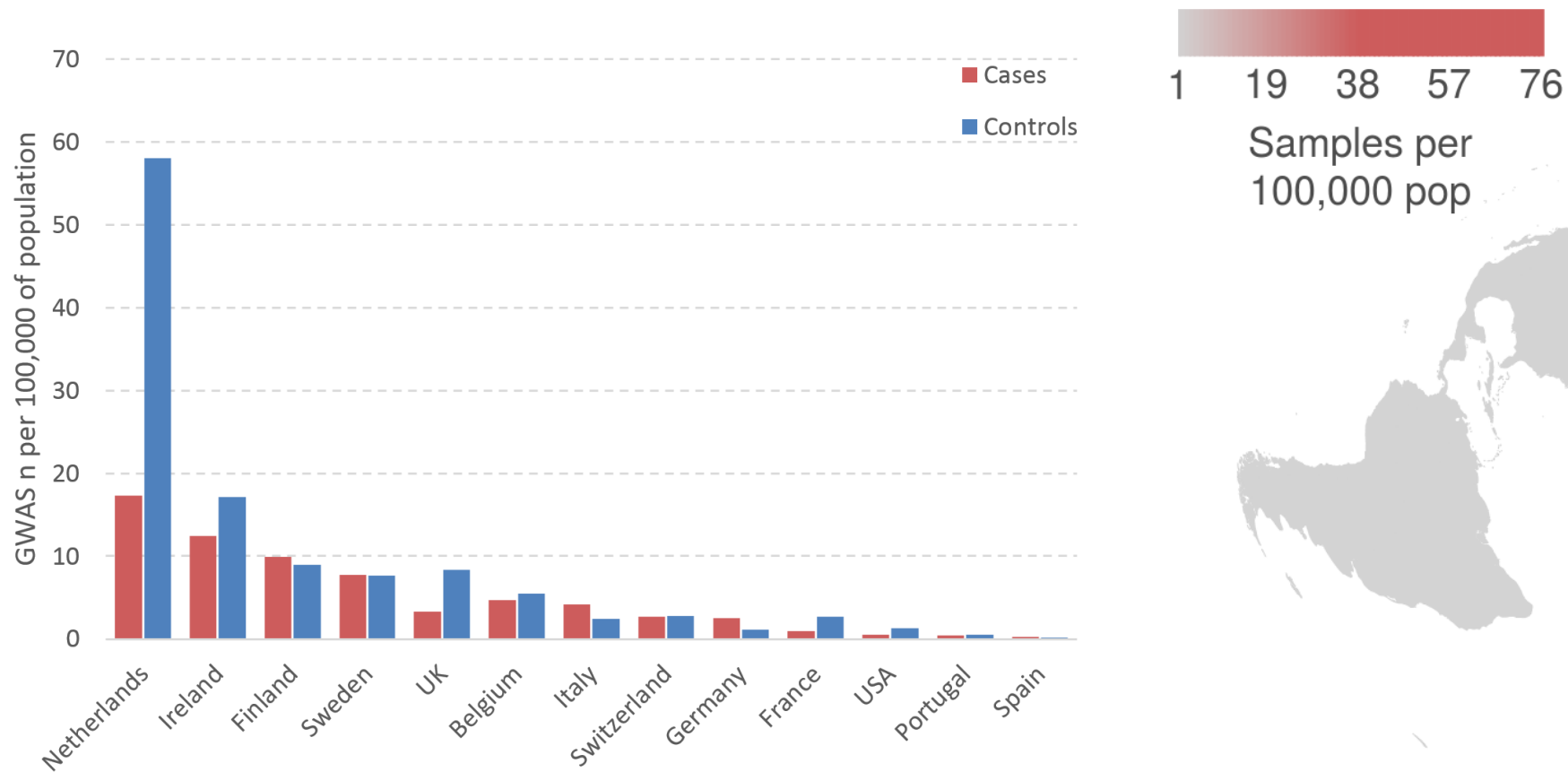
2. Involving more countries will answer deeper questions



GWAS in ALS

General conclusions and further considerations

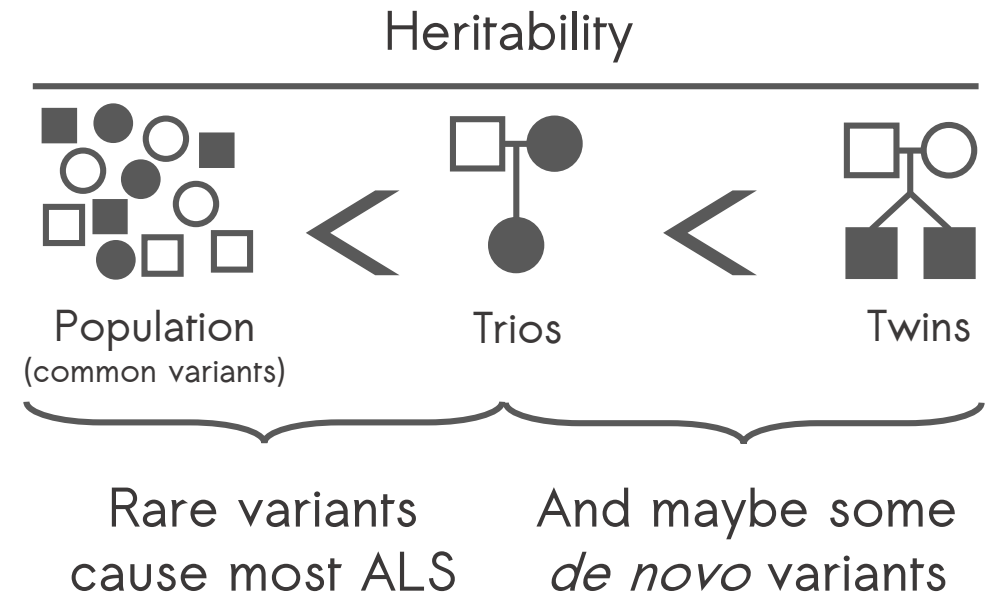
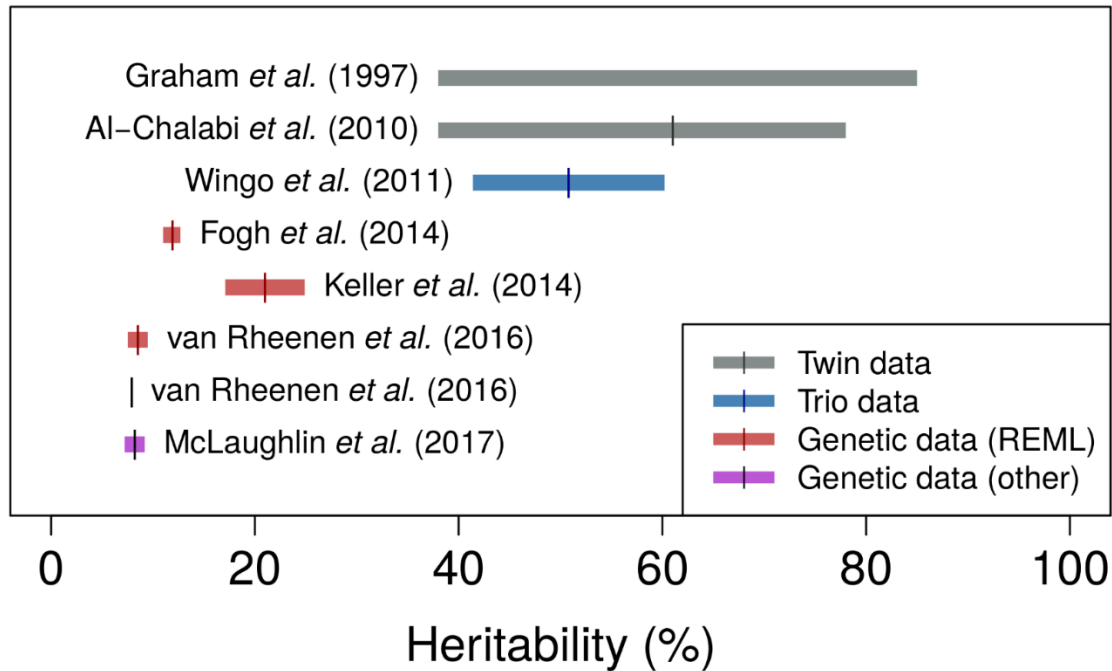
2. Involving more countries will answer deeper questions



GWAS in ALS

General conclusions and further considerations

3. GWAS has helped us to better understand the genetic architecture of ALS



Finding rare variants

Exome sequencing

Previous approach: SNP chips (GWAS)

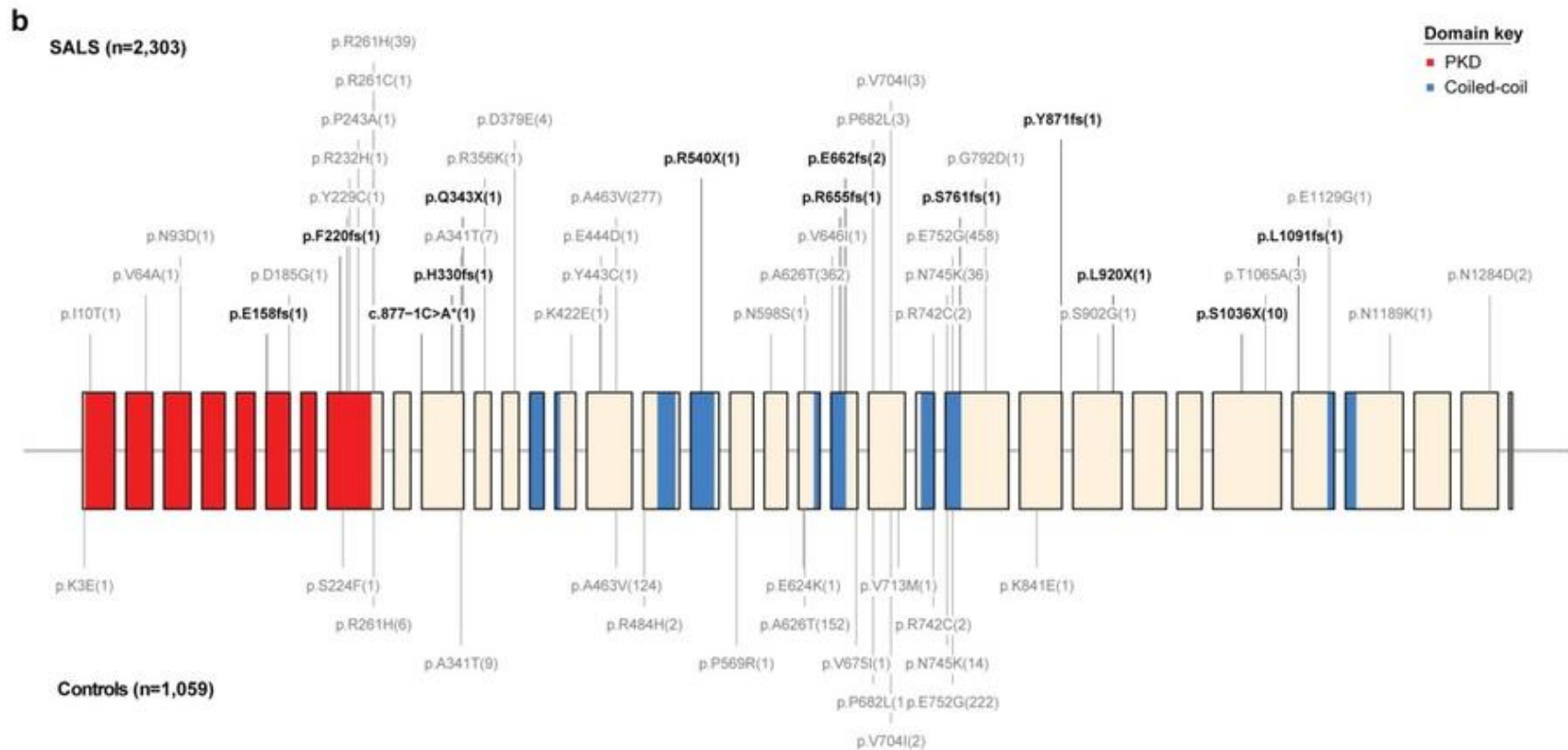
```
-----|exon|-----|intron|----- < gene exons/introns
CTGCTAGCTAGTCTATCGTGCTAGCTAGCTAGCTAGCTAGCGCGTATCGATGAGTCAGCCGTAG < reference genome
... .. C ... .. A ... .. < paternal chromosome
... .. T ... .. A ... .. < maternal chromosome
```

New approach: exome sequencing

```
-----|exon|-----|intron|----- < gene exons/introns
CTGCTAGCTAGTCTATCGTGCTAGCTAGCTAGCTAGCTAGCGCGTATCGATGAGTCAGCCGTAG < reference genome
... .. .TCCATCGTGCTAGCTAGCTAGCTAGCTAGCGC. ... .. < paternal chromosome
... .. .TCTATCGTGCTAGCTAGTAGCTAGCTAGCGC. ... .. < maternal chromosome
```

Implicating rare variants

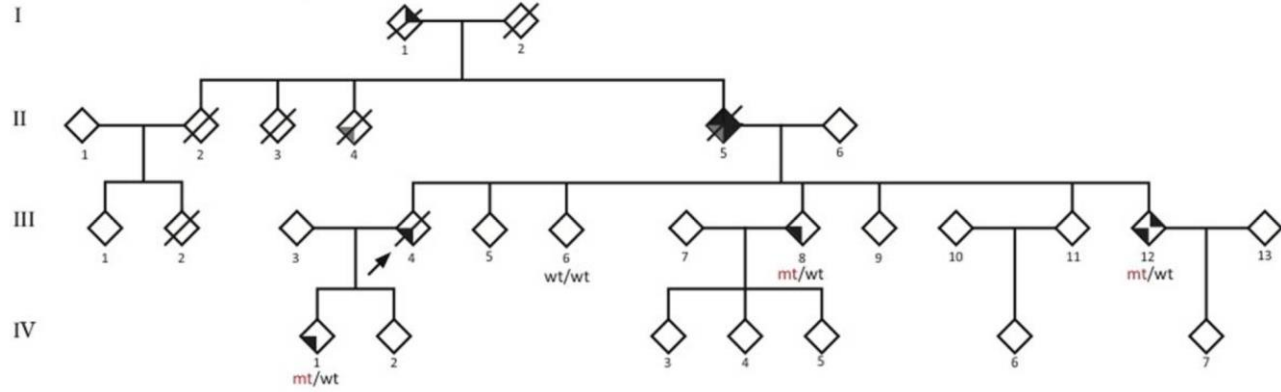
Assessing significance by burden testing



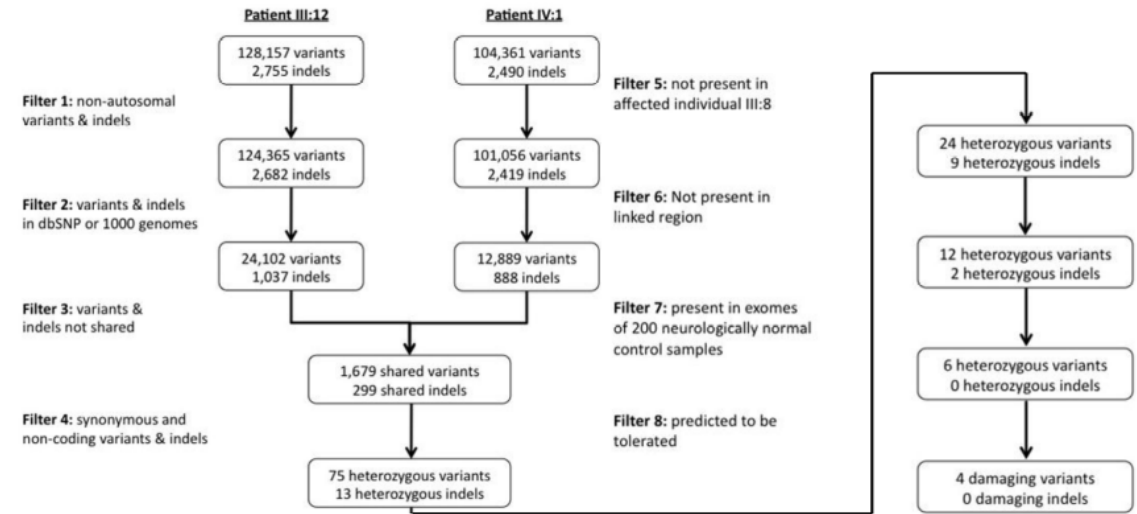
Implicating rare variants

Exome sequencing with pedigrees

A. ITALS#1 (p.R191Q)

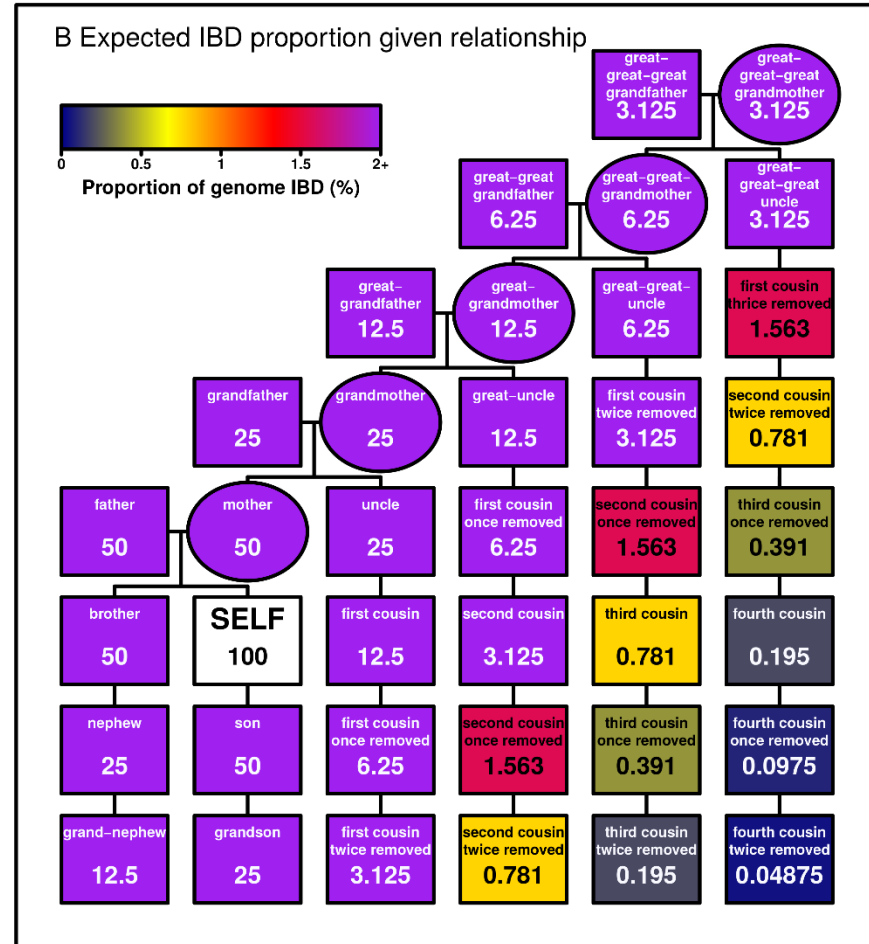
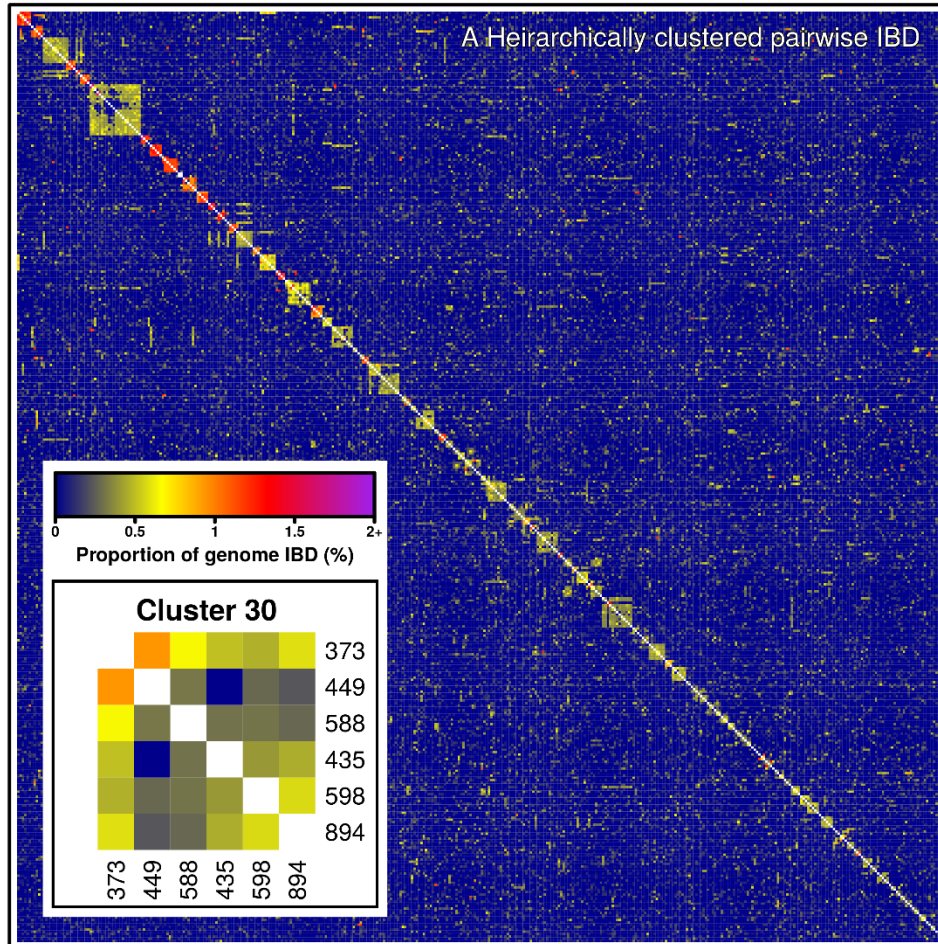


Johnson *et al.* (2010) *Neuron* 68(5):857-64



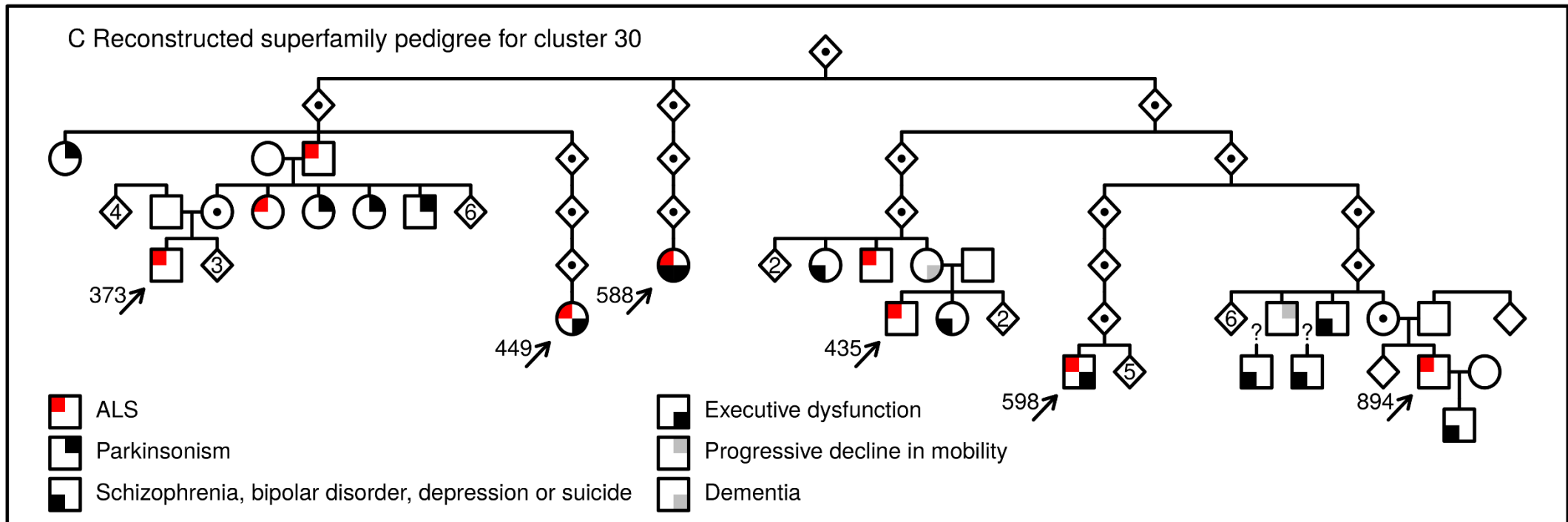
Implicating rare variants

Pseudofamily analysis (IBD = *identity by descent*, ie relatedness)



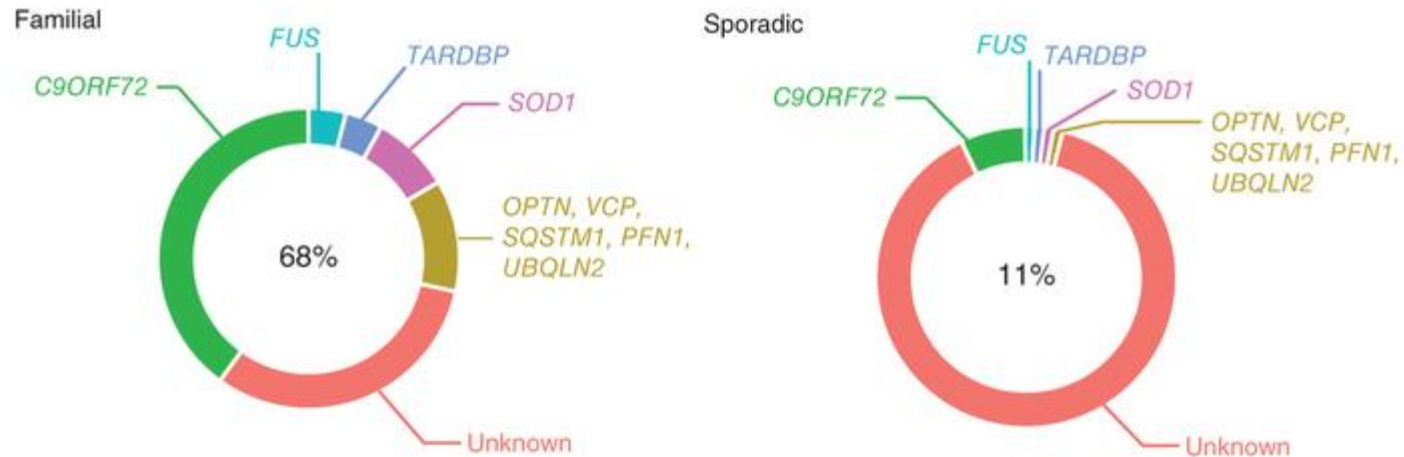
Implicating rare variants

Pseudofamily analysis (IBD = *identity by descent*, ie relatedness)



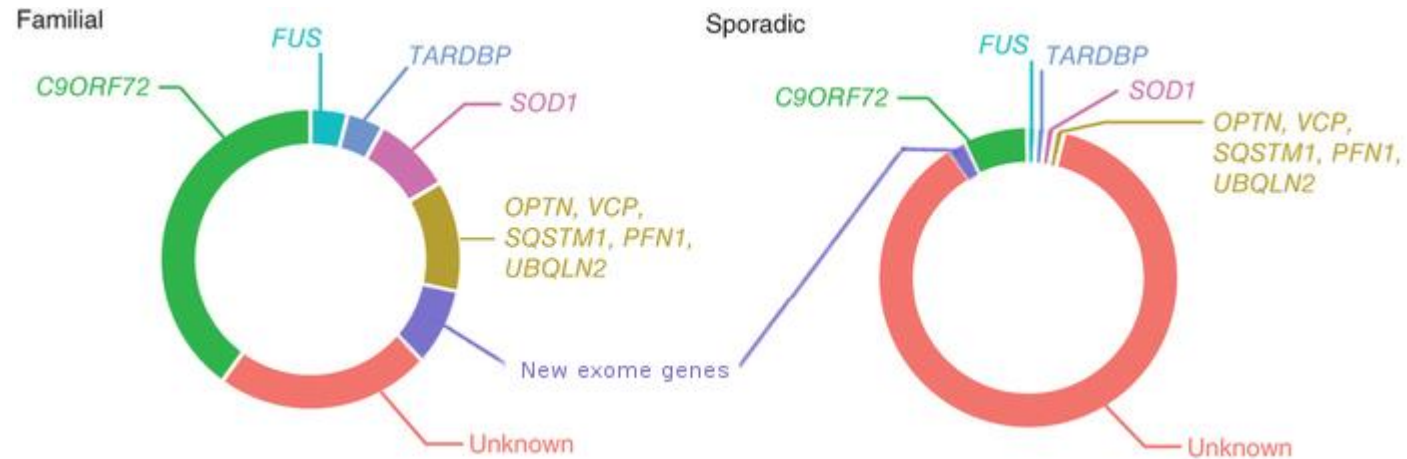
Updating ALS genomics

One gene at a time



Updating ALS genomics

One gene at a time



From exome to whole-genome sequencing

Original approach: SNP chips (GWAS)

```
-----|||----- < gene exons/introns
CTGCTAGCTAGTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCGCGTATCGATGAGTCAGCCGTAG < reference genome
... .. C ... .. A ... .. < paternal chromosome
... .. T ... .. A ... .. < maternal chromosome
```

New approach: exome sequencing

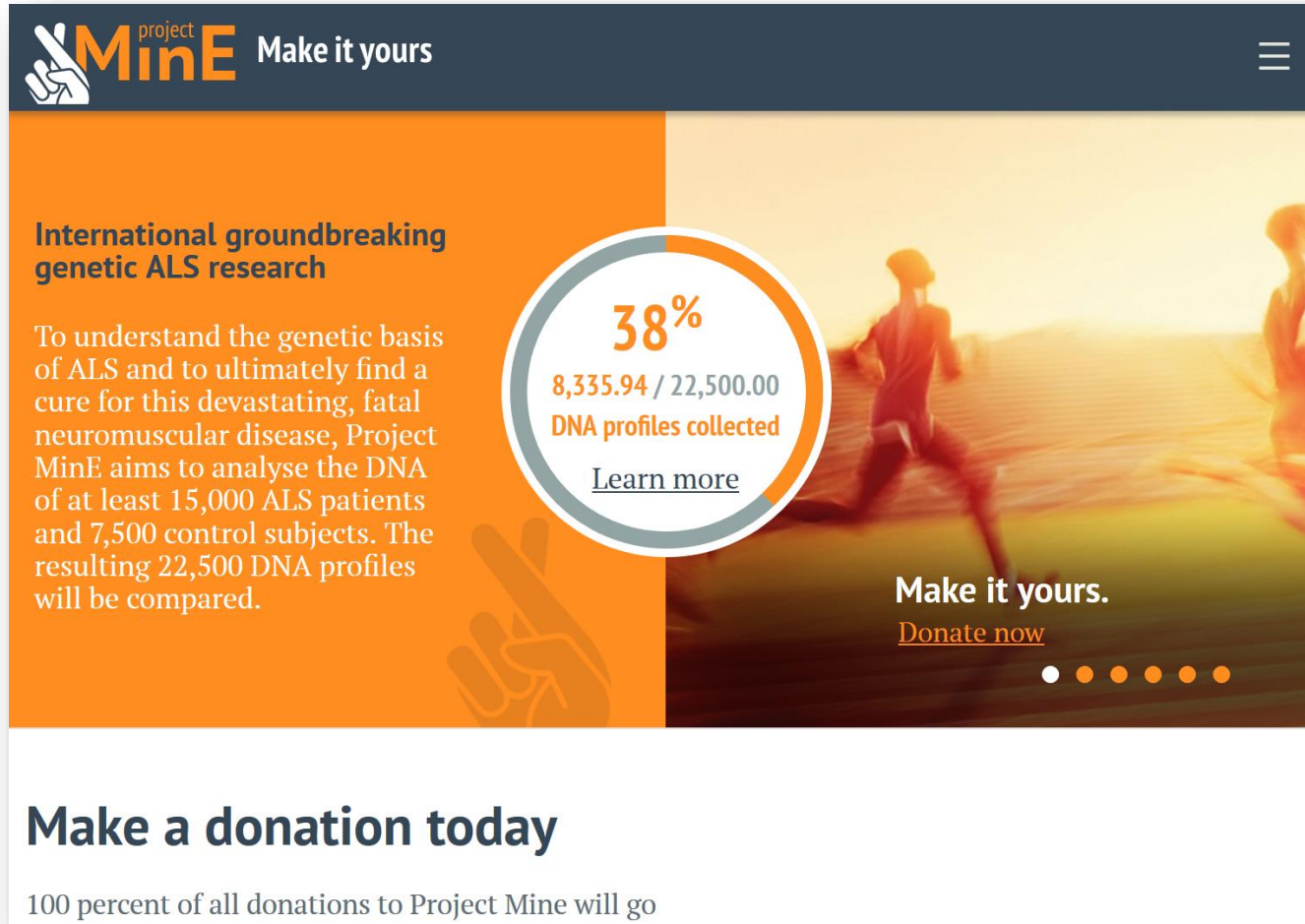
```
-----|||----- < gene exons/introns
CTGCTAGCTAGTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCGCGTATCGATGAGTCAGCCGTAG < reference genome
... .. .TCCATCGTGCTAGCTAGCTAGCTAGCTAGCGC. ... .. < paternal chromosome
... .. .TCTATCGTGCTAGCTAGTAGCTAGCTAGCGC. ... .. < maternal chromosome
```

Newer approach: whole-genome sequencing

```
-----|||----- < gene exons/introns
CTGCTAGCTAGTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCGCGTATCGATGAGTCAGCCGTAG < reference genome
CTGCTAGCTAGTCATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCGCGTATAGATGACTCAGCCGTAG < paternal chromosome
CTGCTAGCTAGTCATCGTGCTAGCTAGTAGCTAGCTAGCGCGTATAGATGAGTCAGCCGTAG < maternal chromosome
```

Whole-genome sequencing in large populations

Project MinE

A banner for Project MinE with a dark blue header containing the logo and the slogan 'Make it yours'. The main content area features a background image of runners and a circular progress indicator showing 38% completion. Text on the left describes the project's goal to analyze DNA of ALS patients and control subjects. A 'Learn more' link is provided. At the bottom, there is a 'Make it yours. Donate now' call to action with a progress bar of five dots, the first of which is white and the others are orange.

project MinE Make it yours

International groundbreaking genetic ALS research

To understand the genetic basis of ALS and to ultimately find a cure for this devastating, fatal neuromuscular disease, Project MinE aims to analyse the DNA of at least 15,000 ALS patients and 7,500 control subjects. The resulting 22,500 DNA profiles will be compared.

38%
8,335.94 / 22,500.00
DNA profiles collected

[Learn more](#)

Make it yours.
[Donate now](#)

Make a donation today

100 percent of all donations to Project Mine will go



A hypothesis

(to explain why so much is still unexplained in ALS genetics)

A multitude of rare **repeat expansions** cause a substantial proportion of ALS

A multitude of rare repeat expansions cause a substantial proportion of ALS

C9orf72 repeat expansion

Human reference genome (chr9:27,573,516-27,573,556)
 ...GACCACGCCCCGGCCCCGGCCCCGGCCCCCTAGCGCGCGACT...

Healthy individual (typical)
 ...GACCACGCCCCGGCCCC-----GGCCCCCTAGCGCGCGACT...

C9orf72-positive ALS
 ...GACCACGCCCCGGCCCC(G₂C₄)_nGGCCCCCTAGCGCGCGACT...

(where n = potentially >1,000)

Table 1 Repeat expansions that cause neurodegeneration

Gene	Disease	Repeat motif	(Non)coding	Pathogenic range
<i>AFF2/FMR3</i>	FRAXE mental retardation syndrome	CCG	Noncoding	>200
<i>AR</i>	Spinal and bulbar muscular atrophy	CAG	Coding	40-62
<i>ARX</i>	X-linked mental retardation	GCG	Coding	17-23
<i>ATN1</i>	Dentatorubral-pallidoluyisian atrophy	CAG	Coding	49-88
<i>ATXN1</i>	Spinocerebellar ataxia type 1	CAG	Coding	39-83
<i>ATXN10</i>	Spinocerebellar ataxia type 10	ATTCT	Noncoding	280-4500
<i>ATXN2</i>	Spinocerebellar ataxia type 2	CAG	Coding	34-59
	Amyotrophic lateral sclerosis	CAG	Coding	27-33
<i>ATXN3</i>	Spinocerebellar ataxia type 3	CAG	Coding	55-84
<i>ATXN7</i>	Spinocerebellar ataxia type 7	CAG	Coding	34->300
<i>ATXN8</i>	Spinocerebellar ataxia type 8	CAG/CTG	Both	80-1300
<i>C9orf72</i>	ALS/FTD	GGGGCC	Noncoding	>30
<i>CACNA1A</i>	Spinocerebellar ataxia type 6	CAG	Coding	21-30
<i>CNBP</i>	Myotonic dystrophy type 2	CCTG	Noncoding	75-11000
<i>CSTB</i>	Epilepsy progressive myoclonia	(C) ₄ G(C) ₄ GCG	Noncoding	30-75
<i>DIP2B</i>	FRA12A mental retardation syndrome	CGG	Noncoding	>23
<i>DMPK</i>	Myotonic dystrophy type 1	CTG	Noncoding	50-6500
	Fragile X mental retardation type 1	CGG	Noncoding	>200
<i>FMR1</i>	Fragile X-associated tremor ataxia syndrome	CGG	Noncoding	55-200
	Fragile X mental retardation type 2	CGG	Noncoding	200-900
<i>FXN</i>	Friedreich's ataxia	GAA	Noncoding	66-1700
<i>HTT</i>	Huntington's disease	CAG	Coding	>35
<i>JPH3</i>	Huntington's disease-like 2	CAG/CTG	Noncoding	>41
<i>NOP56</i>	Spinocerebellar ataxia type 36	GGCCTG	Noncoding	1500-2500
<i>PABPN1</i>	Oculopharyngeal muscular dystrophy	GCG	Coding	11-17
<i>PPP2R2B</i>	Spinocerebellar ataxia type 12	CAG/CTG	Noncoding	55-78
<i>TBP</i>	Spinocerebellar ataxia type 17	CAG	Coding	49-66
<i>TK2-BEAN</i>	Spinocerebellar ataxia type 31	TGGAA	Noncoding	500-760

Pathogenic range is the number of repeats required to manifest disease

Some questions in ALS genomics

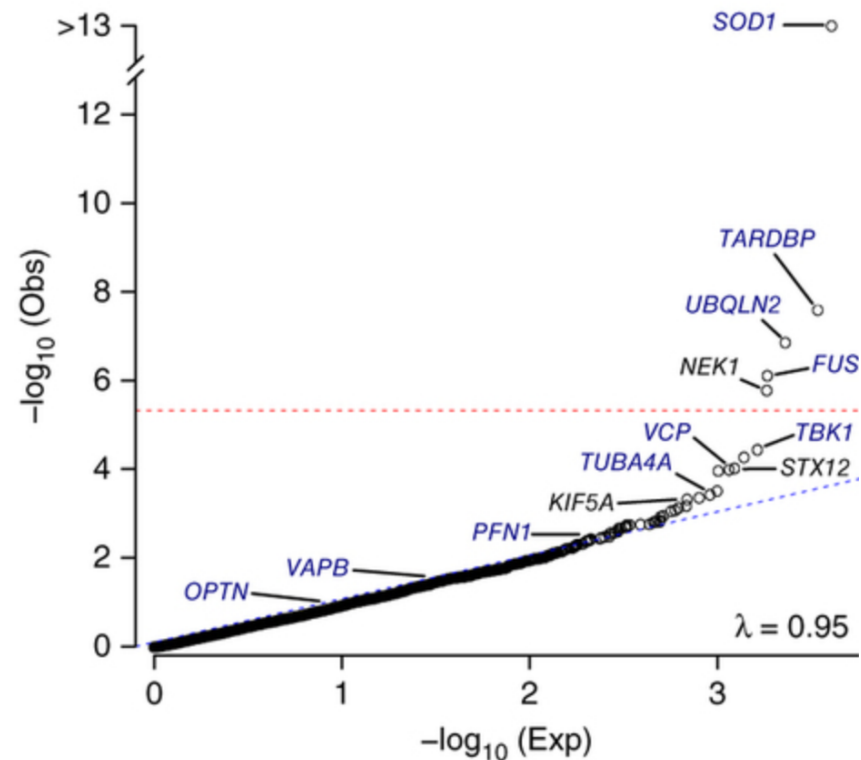
(and some possible answers)

1. Why have **so few** loci been discovered by GWAS (despite >36,000 individuals)?
 - a) Rare repeat expansions not tagged by GWAS SNPs

Some questions in ALS genomics

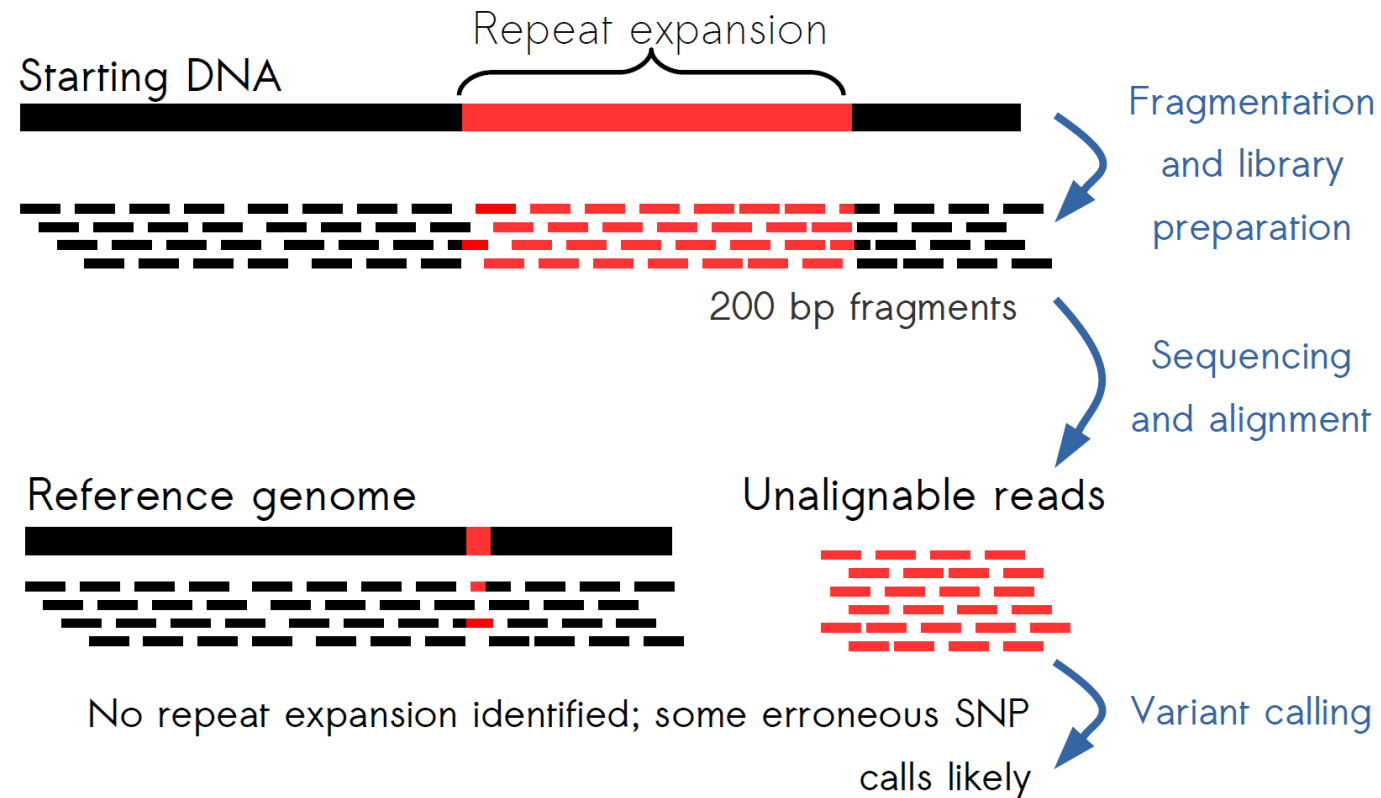
(and some possible answers)

1. Why have **so few** loci been discovered by GWAS (despite >36,000 individuals)?
 - a) Rare repeat expansions not tagged by GWAS SNPs
2. If it's all rare, why are exome sequencing studies not more **inflated**?



Repeat expansions are hard to sequence

Next-generation sequencing: a simplified overview



Some questions in ALS genomics

(and some possible answers)

1. Why have **so few** loci been discovered by GWAS (despite >36,000 individuals)?
 - a) Rare repeat expansions not tagged by GWAS SNPs
2. If it's all rare, why are exome sequencing studies not more **inflated**?
 - a) Next-generation sequencing wouldn't natively discover repeat expansions

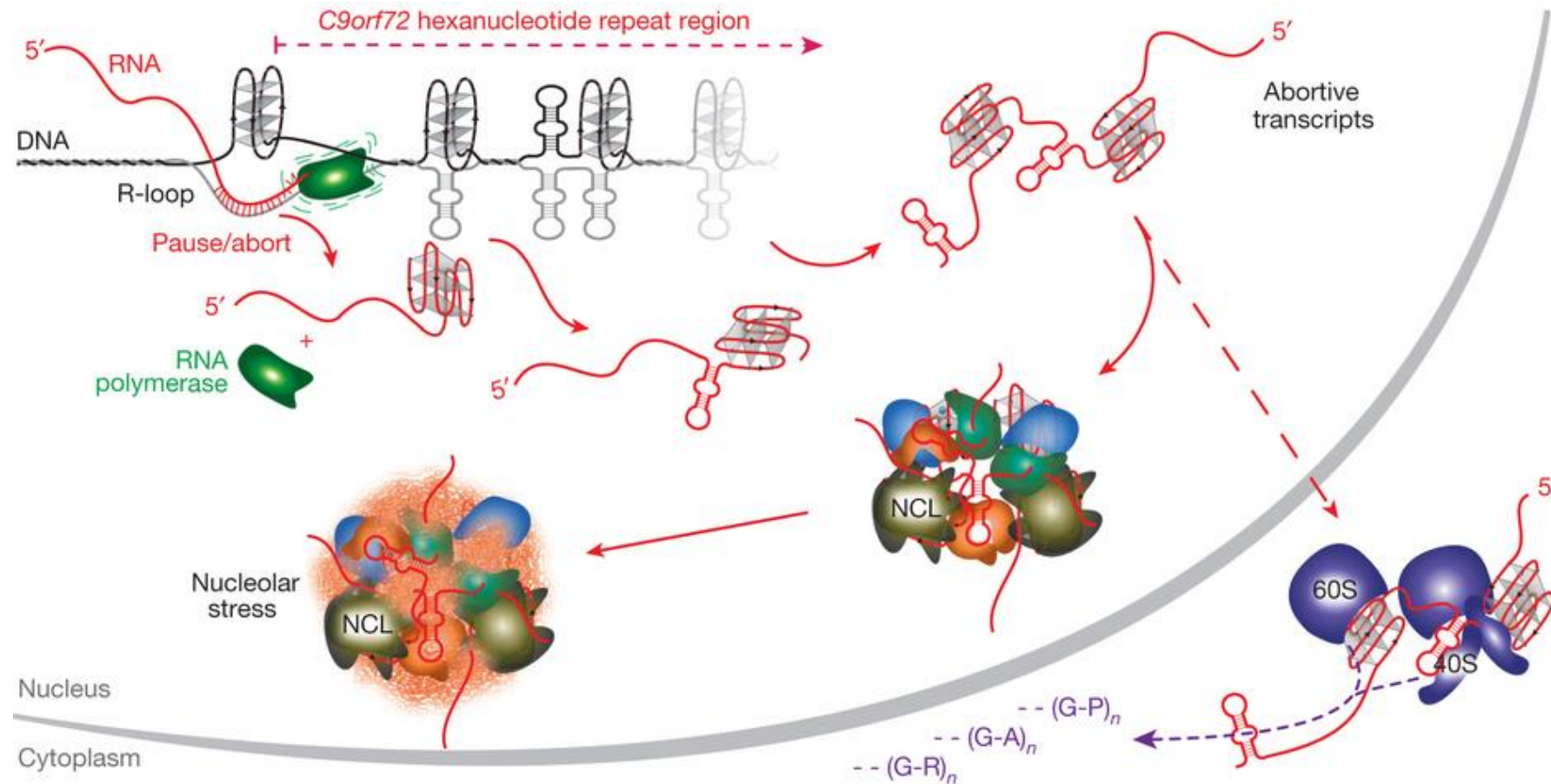
Some questions in ALS genomics

(and some possible answers)

1. Why have **so few** loci been discovered by GWAS (despite >36,000 individuals)?
 - a) Rare repeat expansions not tagged by GWAS SNPs
2. If it's all rare, why are exome sequencing studies not more **inflated**?
 - a) Next-generation sequencing wouldn't natively discover repeat expansions
3. Rare implies **numerous different genes**. How do so many genes confer same(ish) phenotype?

C9orf72 repeat expansion

Mechanisms of disease



Haeusler *et al.* (2014) Nature 507:195-200

Some questions in ALS genomics

(and some possible answers)

1. Why have **so few** loci been discovered by GWAS (despite >36,000 individuals)?
 - a) Rare repeat expansions not tagged by GWAS SNPs
2. If it's all rare, why are exome sequencing studies not more **inflated**?
 - a) Next-generation sequencing wouldn't natively discover repeat expansions
3. Rare implies **numerous different genes**. How do so many genes confer same(ish) phenotype?
 - a) Same disease mechanism (eg RAN translation) on different transcripts

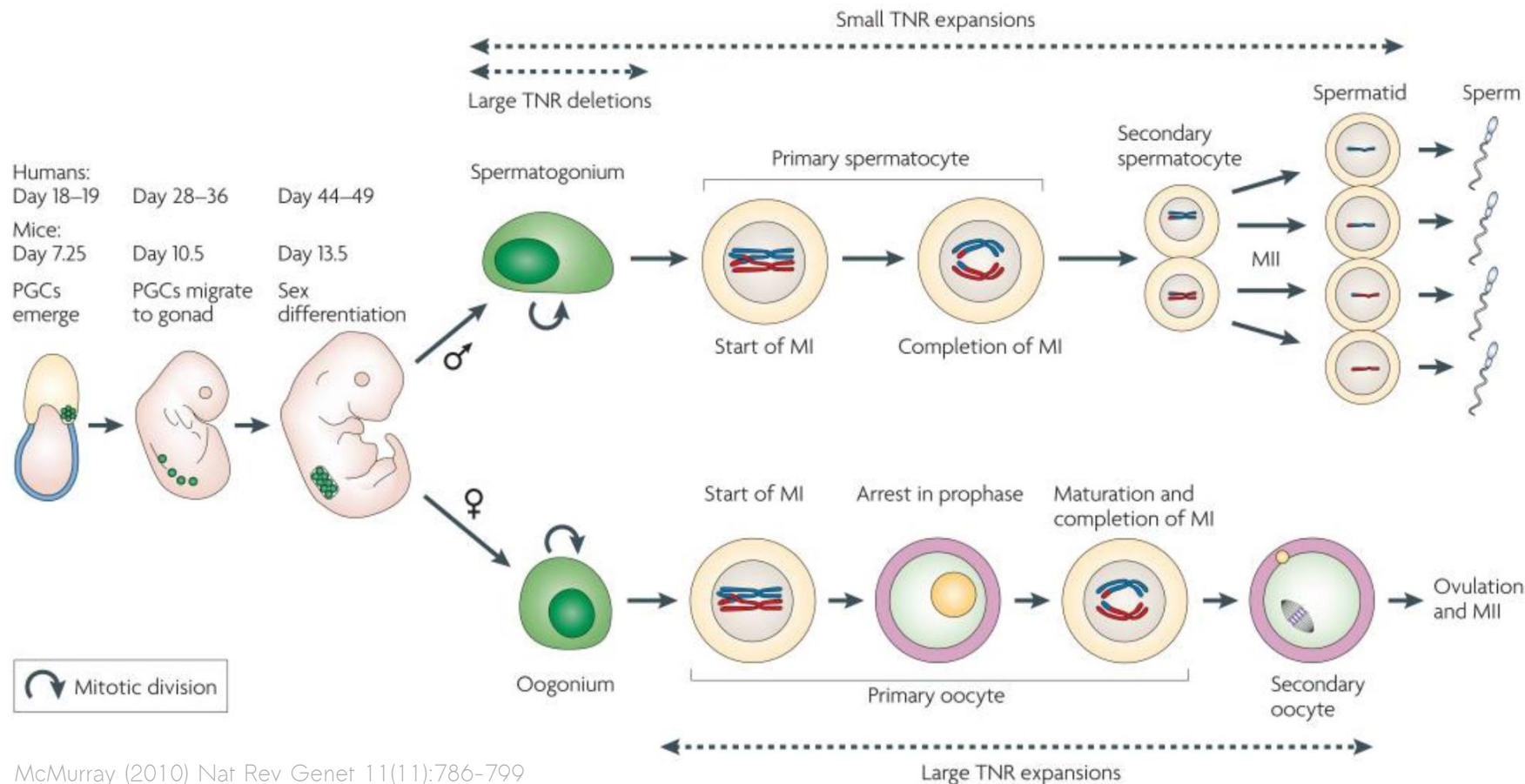
Some questions in ALS genomics

(and some possible answers)

1. Why have **so few** loci been discovered by GWAS (despite >36,000 individuals)?
 - a) Rare repeat expansions not tagged by GWAS SNPs
2. If it's all rare, why are exome sequencing studies not more **inflated**?
 - a) Next-generation sequencing wouldn't natively discover repeat expansions
3. Rare implies **numerous different genes**. How do so many genes confer same(ish) phenotype?
 - a) Same disease mechanism (eg RAN translation) on different transcripts
4. Rare implies **selectively disadvantageous**. How can this happen in ALS (a late onset disease)?

Developmental dynamics of repeat expansion stability

Different mechanisms in mother and father



TNR = trinucleotide repeat

Prone to:

- de novo* expansion during maternal oocyte development
- contraction during paternal spermatogonium development
- de novo* expansion during paternal spermatocyte differentiation

McMurray (2010) Nat Rev Genet 11(11):786-799

Some questions in ALS genomics

(and some possible answers)

1. Why have **so few** loci been discovered by GWAS (despite >36,000 individuals)?
 - a) Rare repeat expansions not tagged by GWAS SNPs
2. If it's all rare, why are exome sequencing studies not more **inflated**?
 - a) Next-generation sequencing wouldn't natively discover repeat expansions
3. Rare implies **numerous different genes**. How do so many genes confer same(ish) phenotype?
 - a) Same disease mechanism (eg RAN translation) on different transcripts
4. Rare implies **selectively disadvantageous**. How can this happen in ALS (a late onset disease)?
 - a) Developmental dynamics eg repeat contraction lowers mutation allele frequency

Some questions in ALS genomics

(and some possible answers)

1. Why have **so few** loci been discovered by GWAS (despite >36,000 individuals)?
 - a) Rare repeat expansions not tagged by GWAS SNPs
2. If it's all rare, why are exome sequencing studies not more **inflated**?
 - a) Next-generation sequencing wouldn't natively discover repeat expansions
3. Rare implies **numerous different genes**. How do so many genes confer same(ish) phenotype?
 - a) Same disease mechanism (eg RAN translation) on different transcripts
4. Rare implies **selectively disadvantageous**. How can this happen in ALS (a late onset disease)?
 - a) Developmental dynamics eg repeat contraction lowers mutation allele frequency
5. Heritability disparity suggests some *de novo* mutations. Why don't we see increased paternal age?

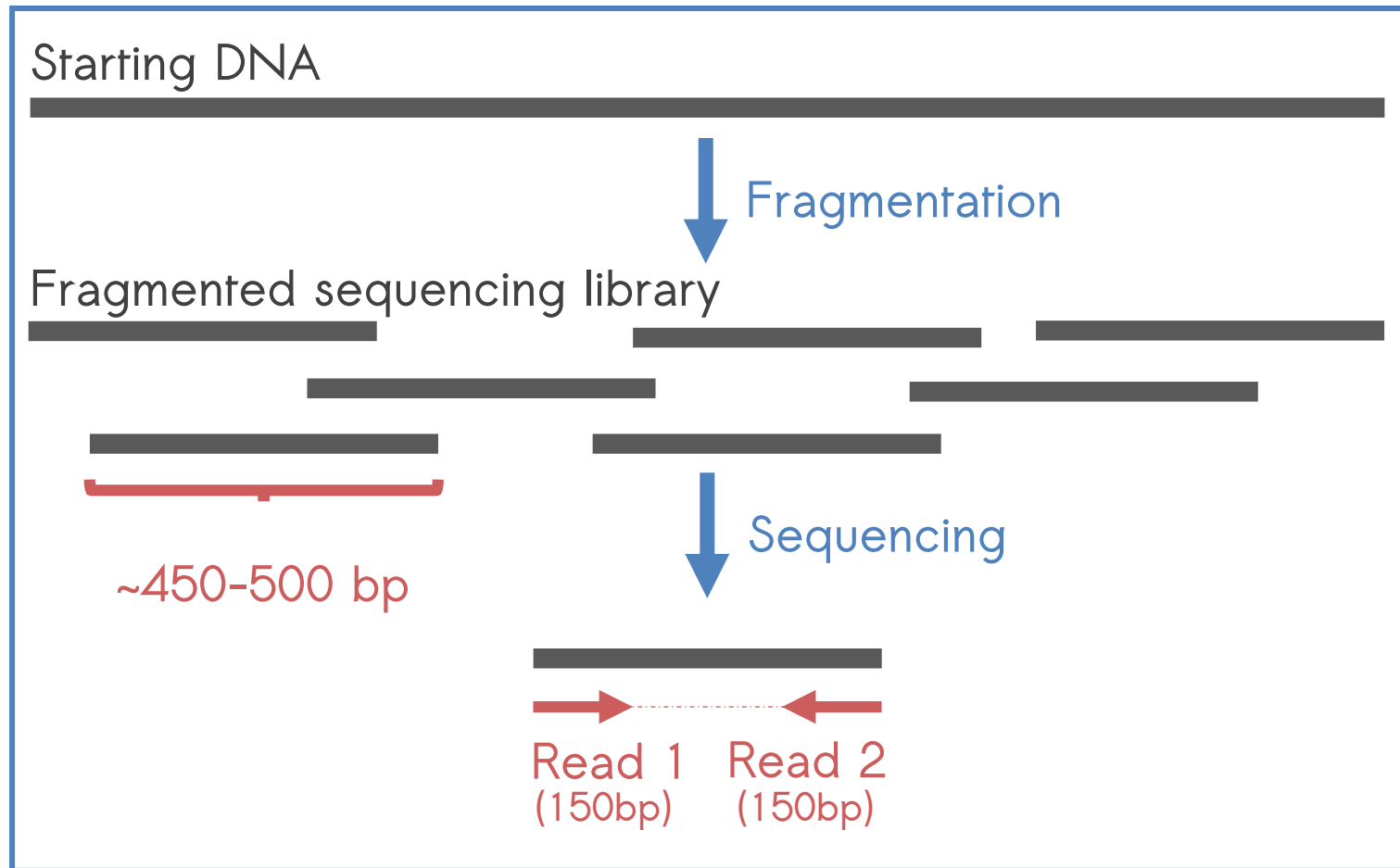
Some questions in ALS genomics

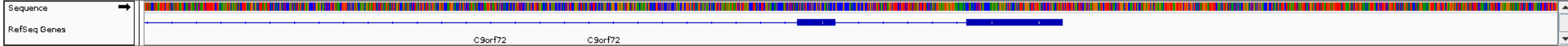
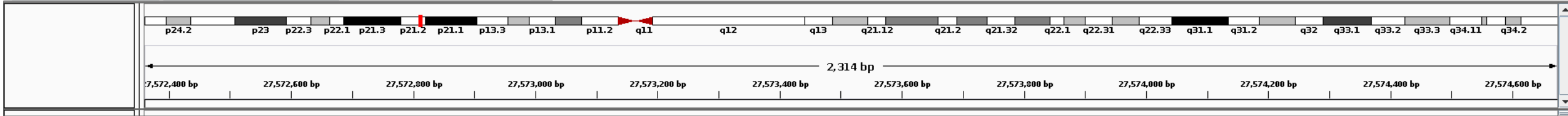
(and some possible answers)

1. Why have **so few** loci been discovered by GWAS (despite >36,000 individuals)?
 - a) Rare repeat expansions not tagged by GWAS SNPs
2. If it's all rare, why are exome sequencing studies not more **inflated**?
 - a) Next-generation sequencing wouldn't natively discover repeat expansions
3. Rare implies **numerous different genes**. How do so many genes confer same(ish) phenotype?
 - a) Same disease mechanism (eg RAN translation) on different transcripts
4. Rare implies **selectively disadvantageous**. How can this happen in ALS (a late onset disease)?
 - a) Developmental dynamics eg repeat contraction lowers mutation allele frequency
5. Heritability disparity suggests some *de novo* mutations. Why don't we see increased paternal age?
 - a) Repeat expansions are prone to *de novo* instability and are not tied to paternal age

How can we find (novel) repeat expansions?

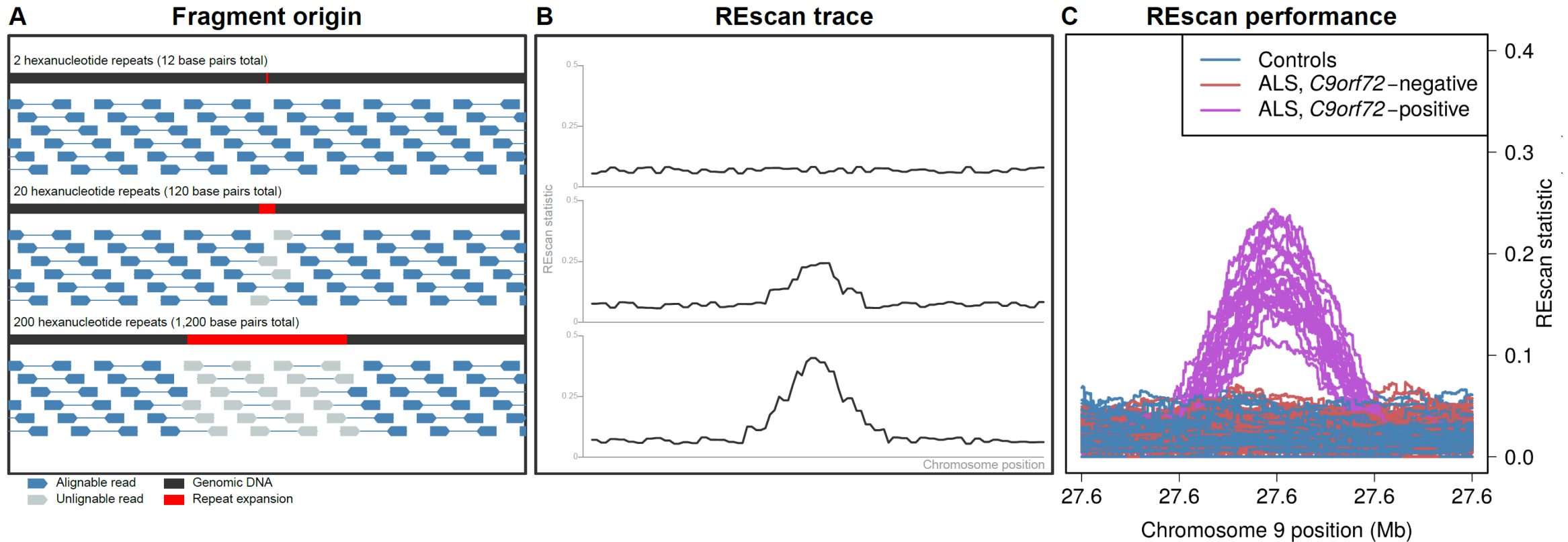
Paired-end next-generation sequencing





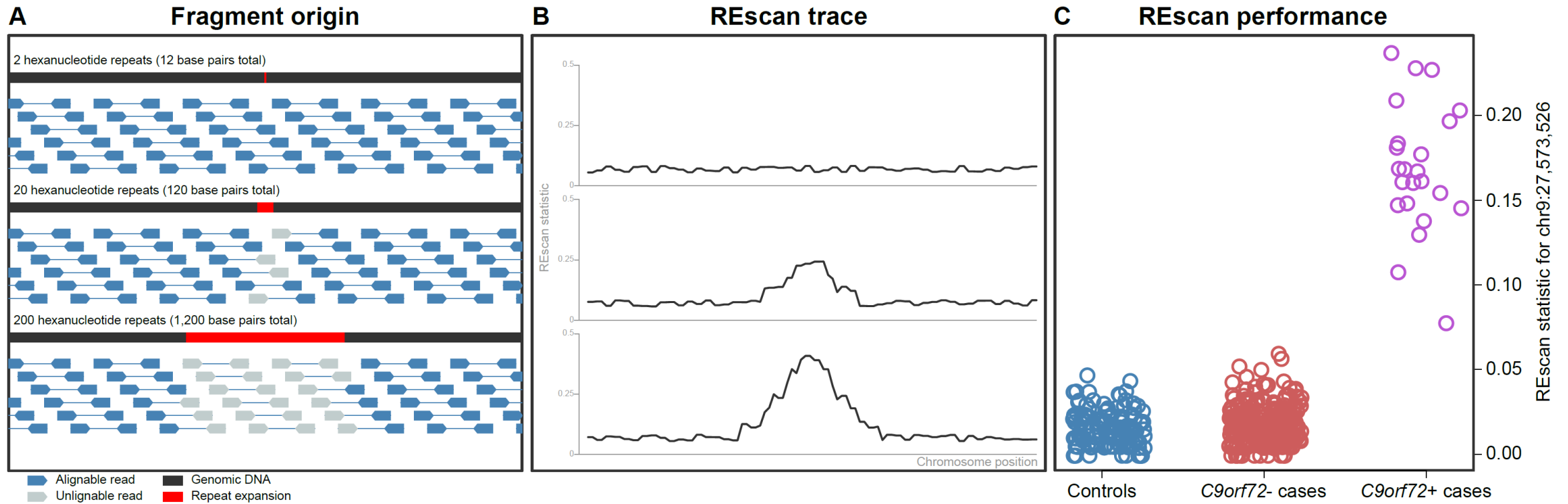
Finding repeat expansions

Using paired-end next-generation sequencing data



Finding repeat expansions

Using paired-end next-generation sequencing data



Finding repeat expansions

ExpansionHunter

peer reviewed) is the author/funder. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

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

Egor Dolzhenko^{1,2}, Joke J.F.A. van Vugt^{2,3}, Richard J. Shaw^{3,4}, Mitchell A. Bekritsky⁵, Marka van Blitterswijk⁶, Zoya Kingsbury⁷, Sean J. Humphray³, Raymond D. Schellevis², William J. Brands², Matt Baker³, Rosa Rademakers³, Maarten Kooyman⁸, Gijs H.P. Tazelaar², Michael A. van Es², Russell McLaughlin^{7,8}, William Sproviero⁹, Aleksey Shatunov⁹, Ashley Jones⁵, Ahmad Al Khleifat⁹, Alan Pittman¹⁰, Sarah Morgan¹⁰, Orla Hardiman^{7,8}, Ammar Al-Chalabi⁹, Leonard H. van den Berg², David R. Bentley³, Michael A. Eberle^{1,2,*} and Jan H. Veldink^{2,*}

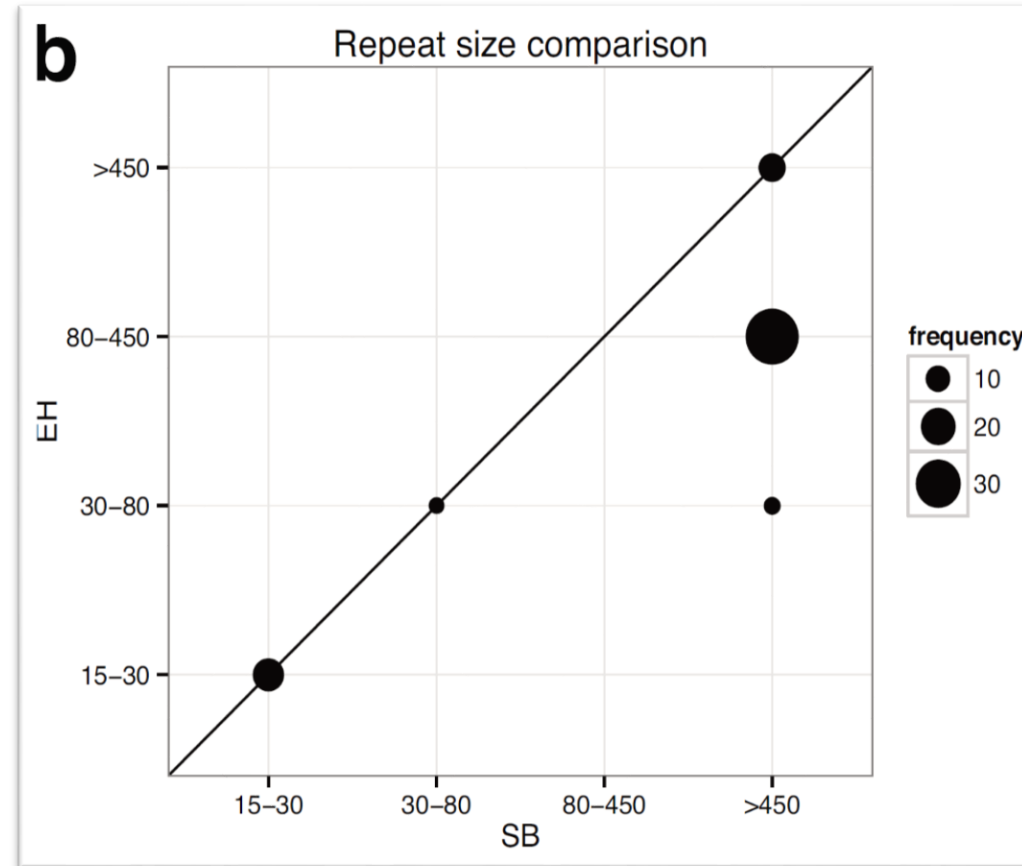
¹ Illumina Inc., 5200 Illumina Way, San Diego, CA, USA
² Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
³ Illumina Cambridge Ltd., Chesterford Research Park, Little Chesterford, UK
⁴ Reproductive Ltd., Future Business Centre, Kings Hedges Rd, Cambridge, UK
⁵ Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA
⁶ Streeklara, Science Park 340, Amsterdam, The Netherlands
⁷ Academic Unit of Neurology, Trinity College Dublin, Trinity Biomedical Sciences Institute, Dublin, Republic of Ireland
⁸ Department of Neurology, Beaumont Hospital, Dublin, Republic of Ireland
⁹ Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
¹⁰ Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK

*These authors contributed equally
To whom correspondence should be addressed

Abstract

Accurately identifying large repeat expansions including those that cause amyotrophic lateral sclerosis (ALS) and Fragile X syndrome is challenging for short-read (100-150bp) whole genome sequencing (WGS) data. A solution to this problem is an important step towards integrating WGS into precision medicine. We have developed a research tool called ExpansionHunter that, using PCR-free WGS data, can identify repeat expansions at the locus of interest, even if the expansion is larger than the read length. We applied our algorithm to WGS data from 3,001 ALS patients who have been tested for the presence of the *C9orf72* repeat expansion with repeat-primed PCR (RP-PCR). Southern blot and fragment length analysis were applied on a subset of samples to confirm the presence or absence of the repeat expansion. Compared to the RP-PCR results, our WGS-based method identified pathogenic repeat expansions (>30 GGCCCC repeats) with 98.1% sensitivity and 99.7% specificity. Further inspection identified that 11 of the 12

1



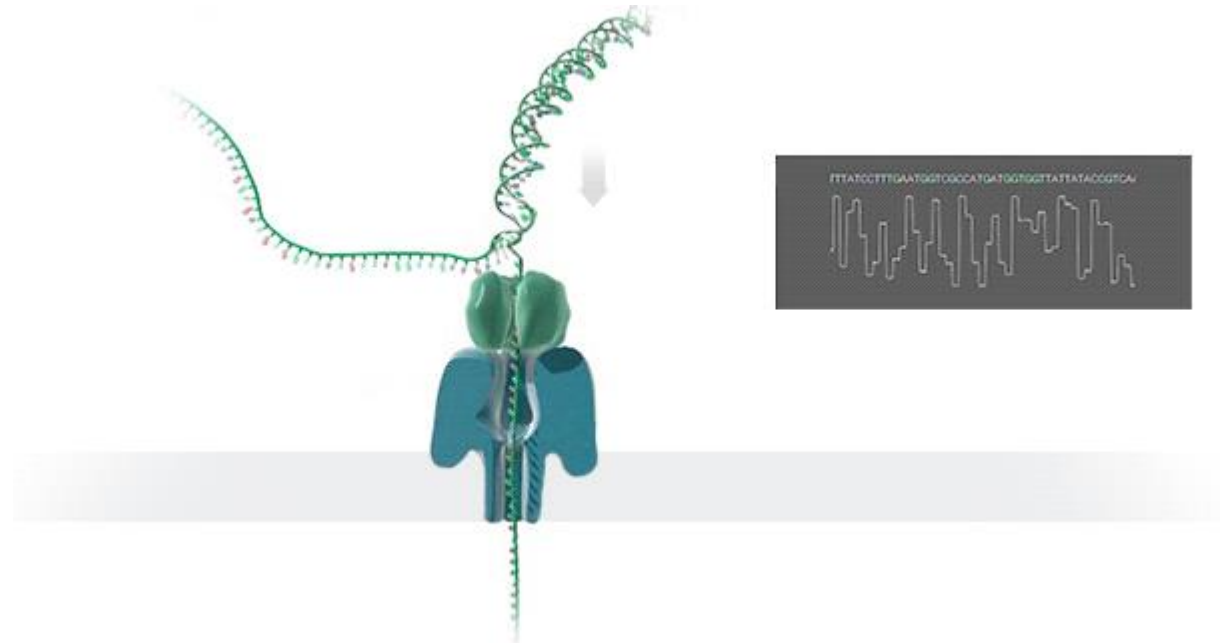
Can we directly measure repeat expansions?

From the *next* generation to the *third* generation



Can we directly measure repeat expansions?

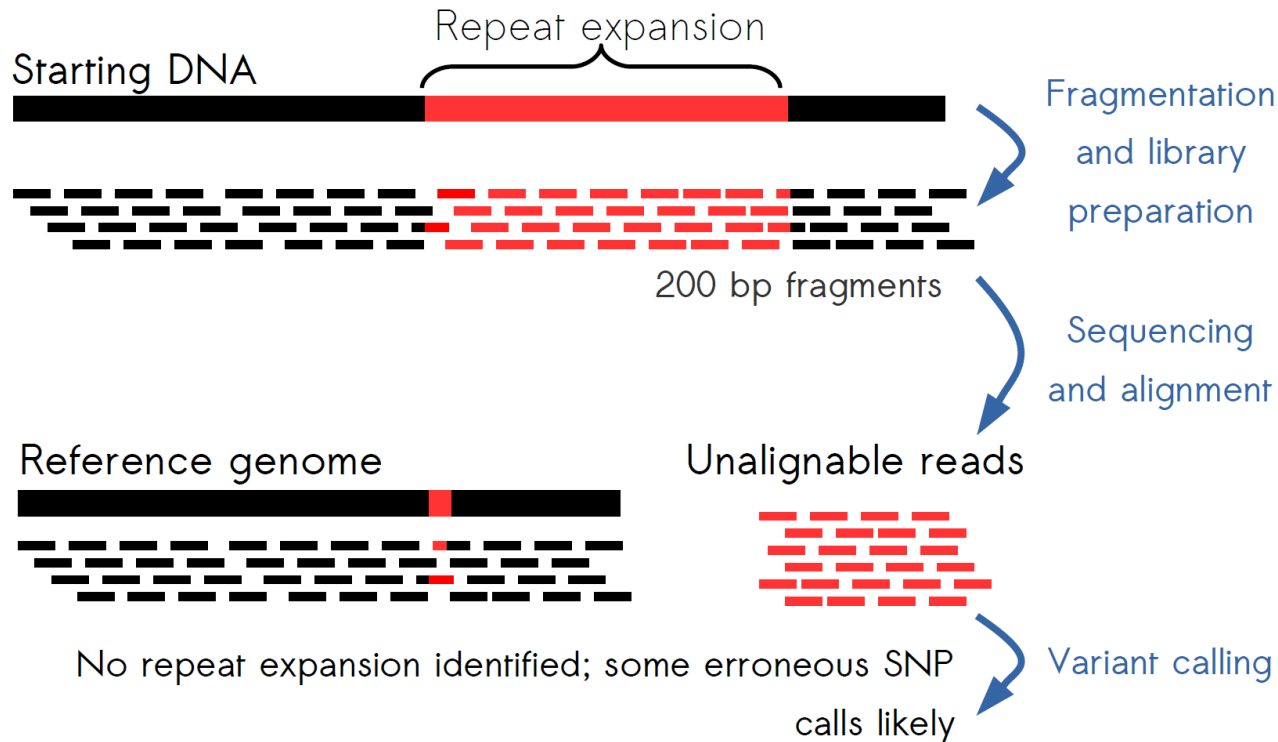
3rd-generation sequencing with ultra-long reads using Oxford Nanopore MinION



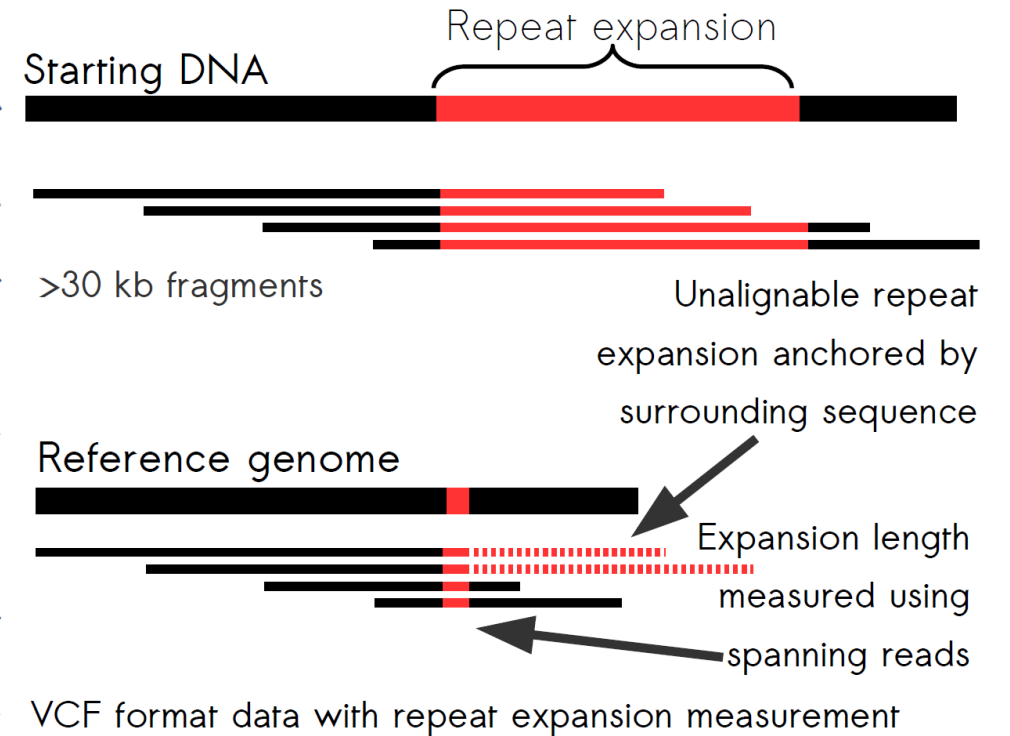
NGS vs 3GS

Spanning repeat expansions

Next-generation sequencing

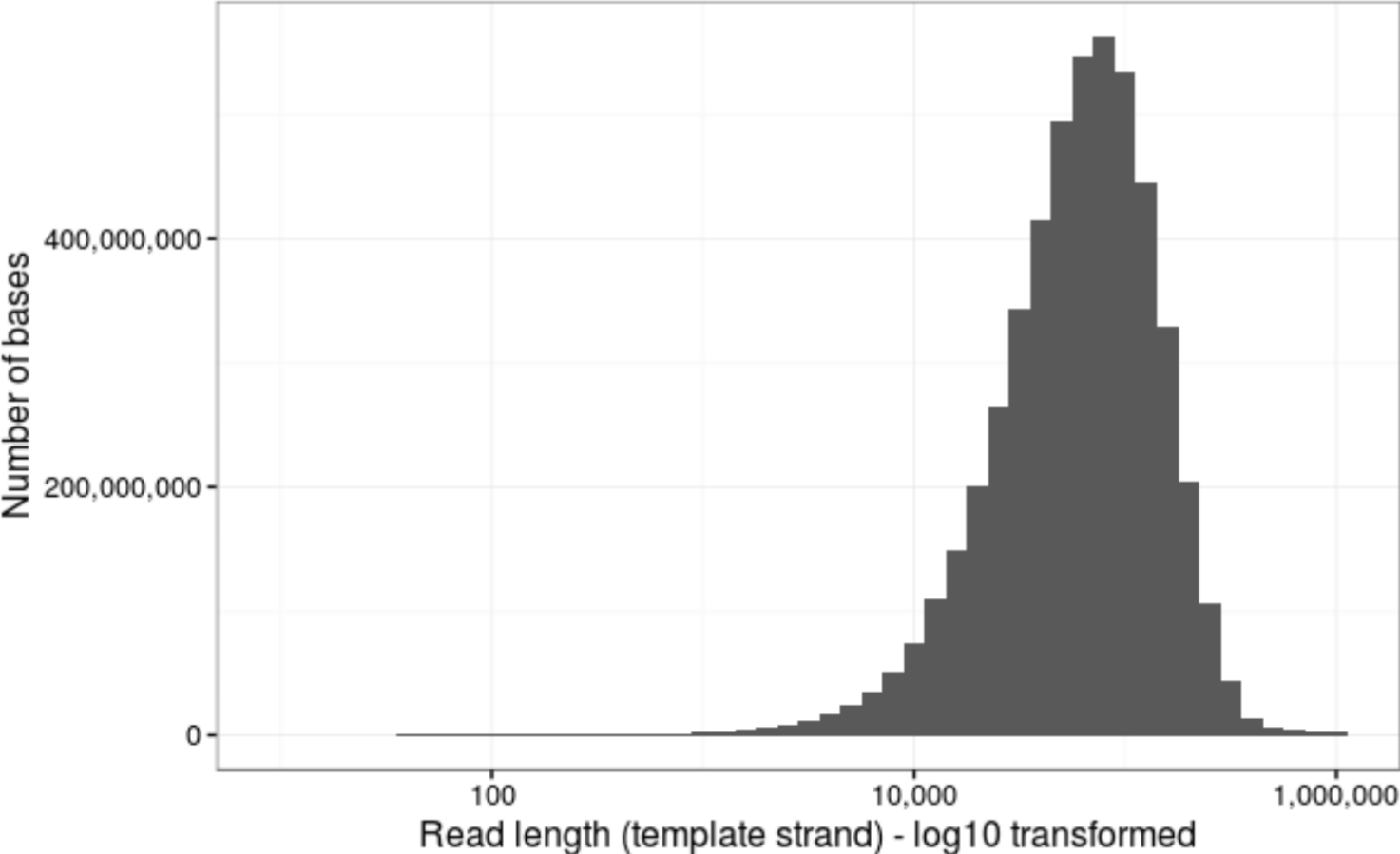


3rd-generation sequencing



How long is a long read?

Answer: *very*

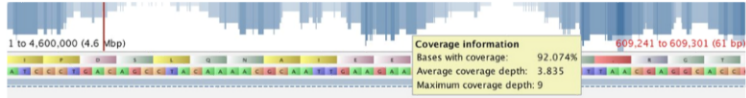


Thar she blows! Ultra long read method for nanopore sequencing

09 Mar 2017

tl;dr version

- Ultra long reads (up to 882 kb and indeed higher) can be achieved on the Oxford Nanopore MinION using traditional DNA extraction techniques and minor changes to the library preparation protocol, without the need for size selection



Nick Loman
@pathogenomenick



Not bad .. 92% genome coverage of E. coli, average depth of 3.8x. But from just 43 reads. github.com/nickloman/mass... (gt350kb.fasta)

9:30 PM - 2 Mar 2017

59 109

How long is a long read?

Answer: *very*



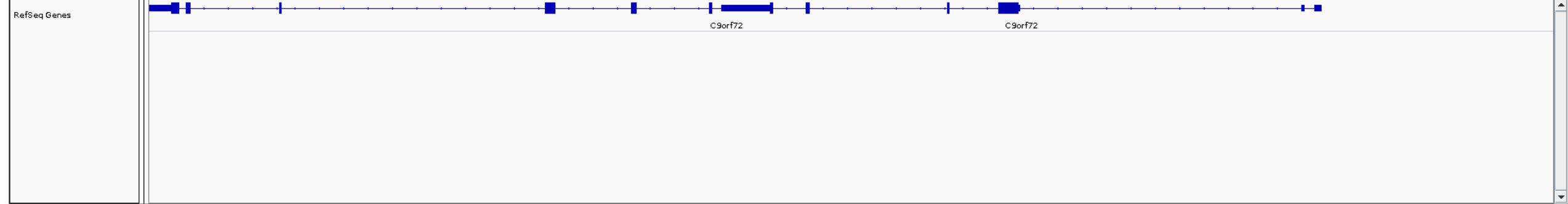
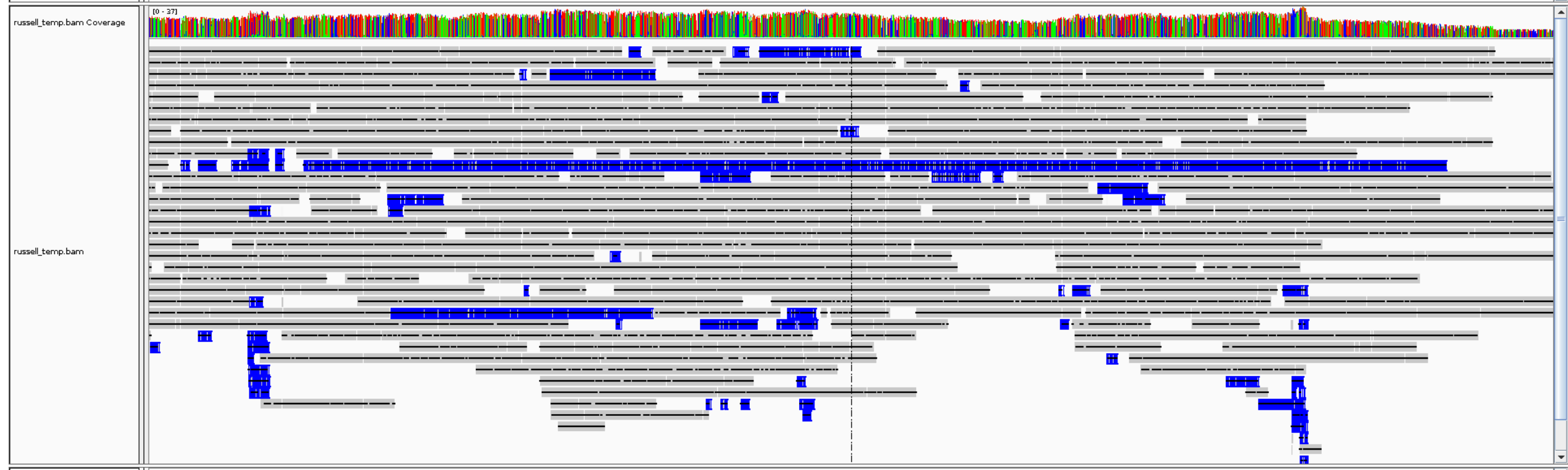
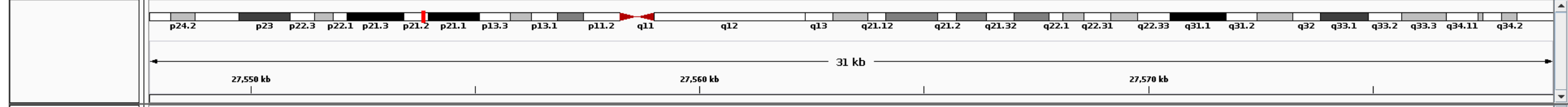
150 bp Illumina read
(banana for scale)



950,000 bp Oxford Nanopore read
(Burj Khalifa for scale)

C9orf72 locus with Oxford Nanopore





Acknowledgements

Trinity College Dublin

Neurology

Orla Hardiman & team
Peter Bede
Susan Byrne
Emma Corr
Mark Heverin
James Rooney
Marie Ryan
Alice Vajda

Population Genetics

Dan Bradley & team
Matthew Teasdale

Complex Trait Genomics

Ross Byrne
Mark Doherty
Jenny Hengeveld

UMC Utrecht

Jan Veldink
Leonard van den Berg
Wouter van Rheenen
Sara Pulit
Michael van Es
Gijs Tazelaar
Jurjen Luykx
Dick Schijven

King's College London

Ammar Al-Chalabi
William Sproviero

University of Massachusetts

John Landers
Kevin Kenna & family



✉ mclaugr@tcd.ie

🌐 <http://bioinf.gen.tcd.ie/ctg>

🐦 @RSLMcL

🌀 [rlmcl](#)