



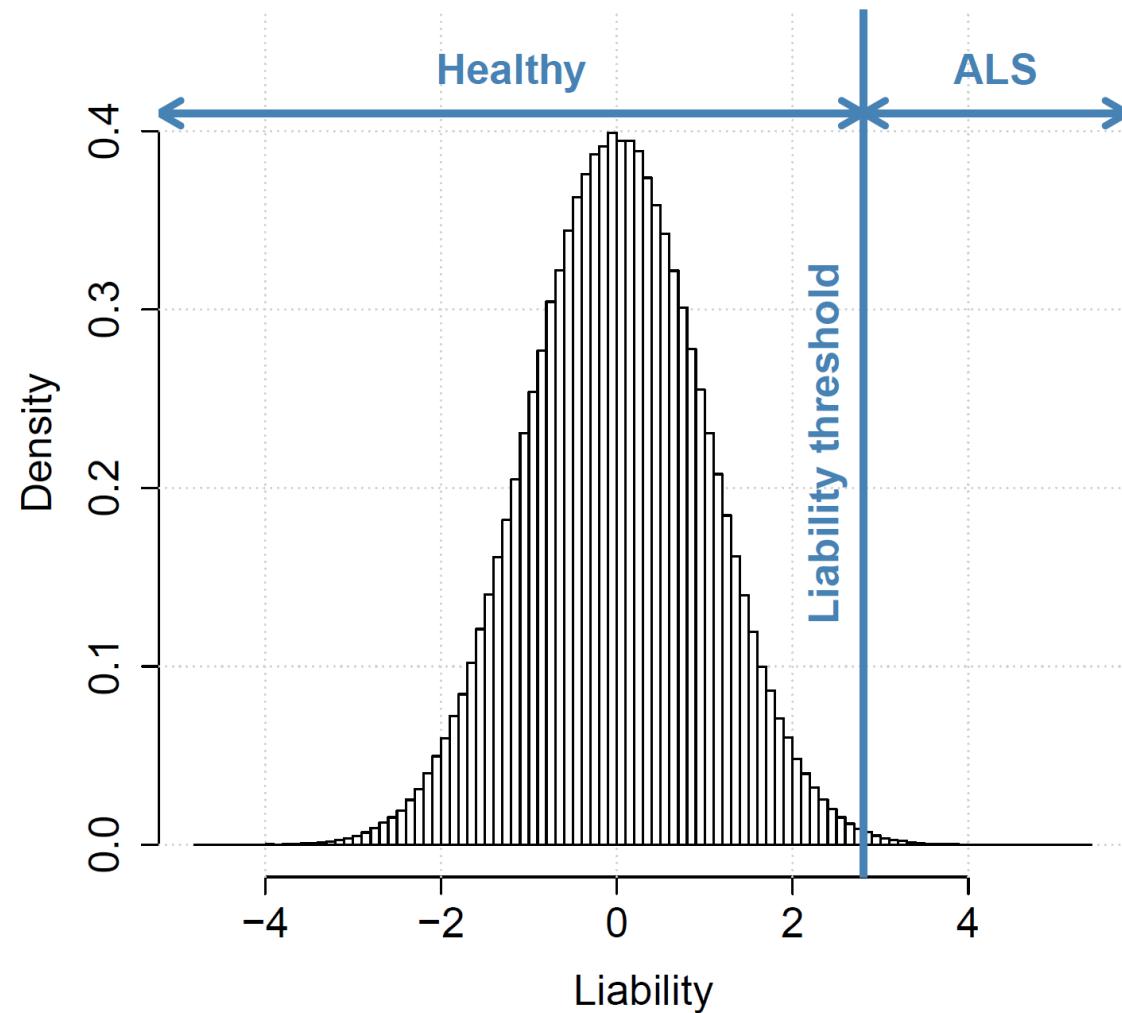
# The panorama of ALS genomics

ENCALS 2017 | Ljubljana | 19.05.2017

Russell McLaughlin | mclaugr@tcd.ie | @RSLMcL

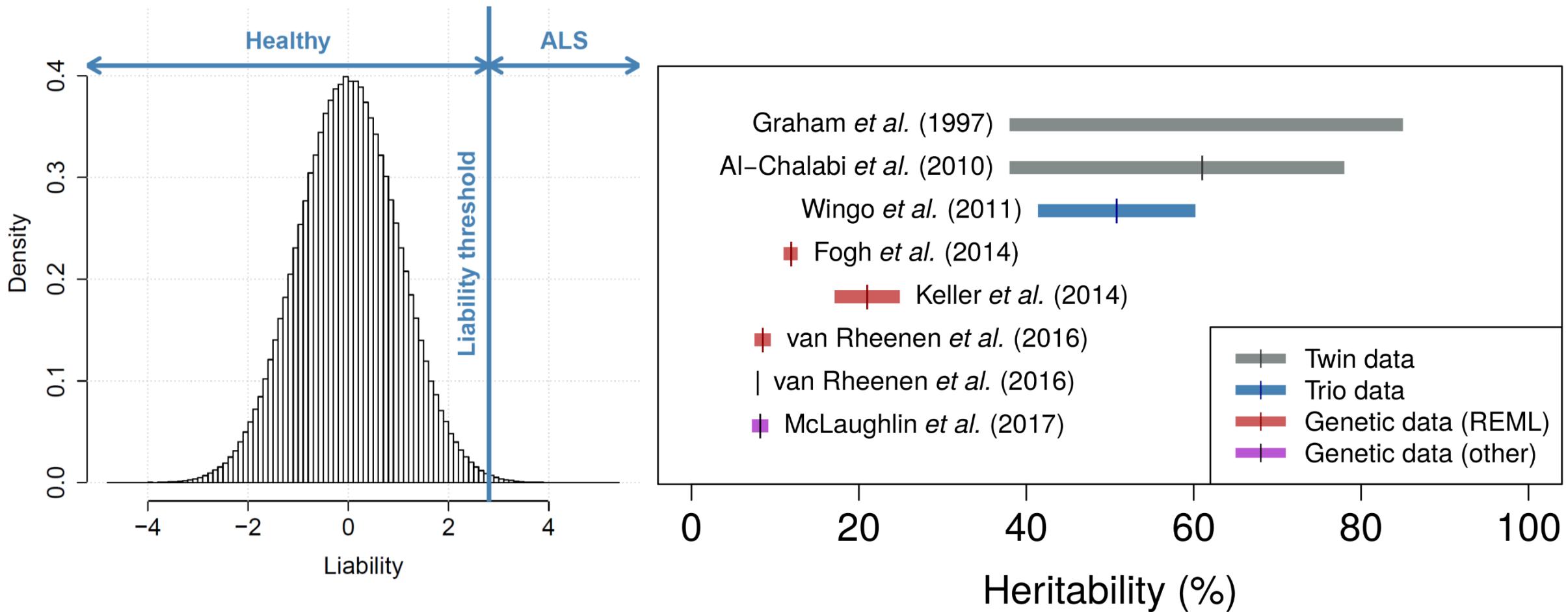
# Is ALS genetic?

Liability to develop ALS - explain liability on this slide



# Is ALS genetic?

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

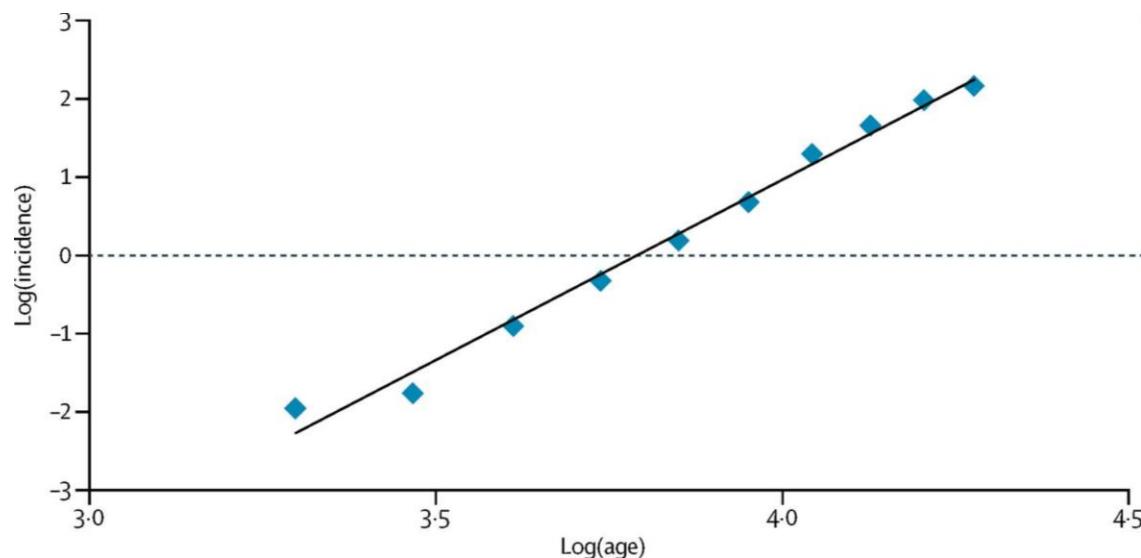


# Non-genetic risk

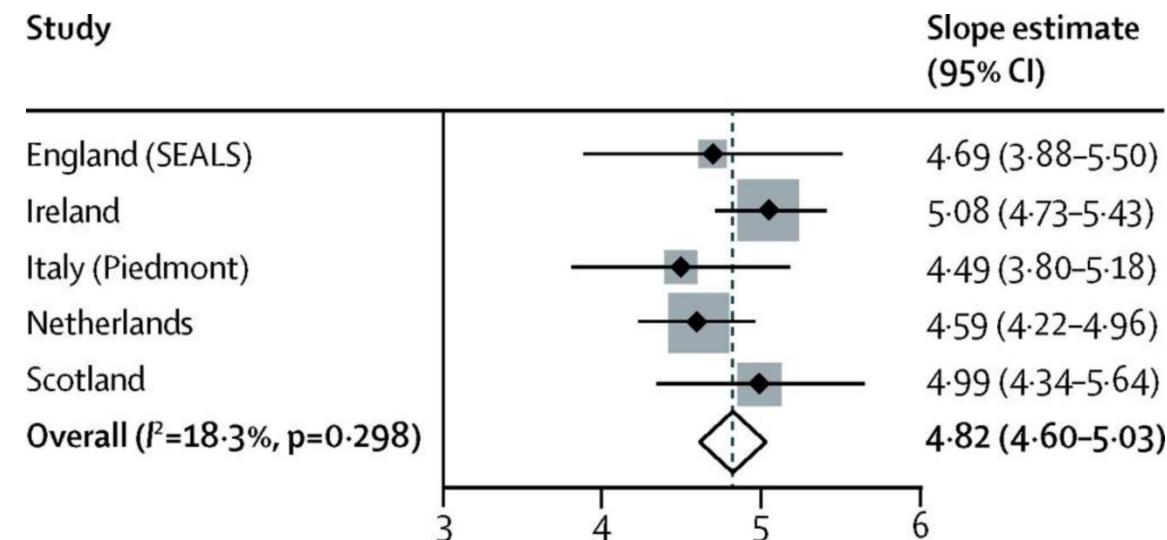
Multiple lifetime steps/exposures are required to develop ALS

Age-specific incidence      Steps      Time (age)

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

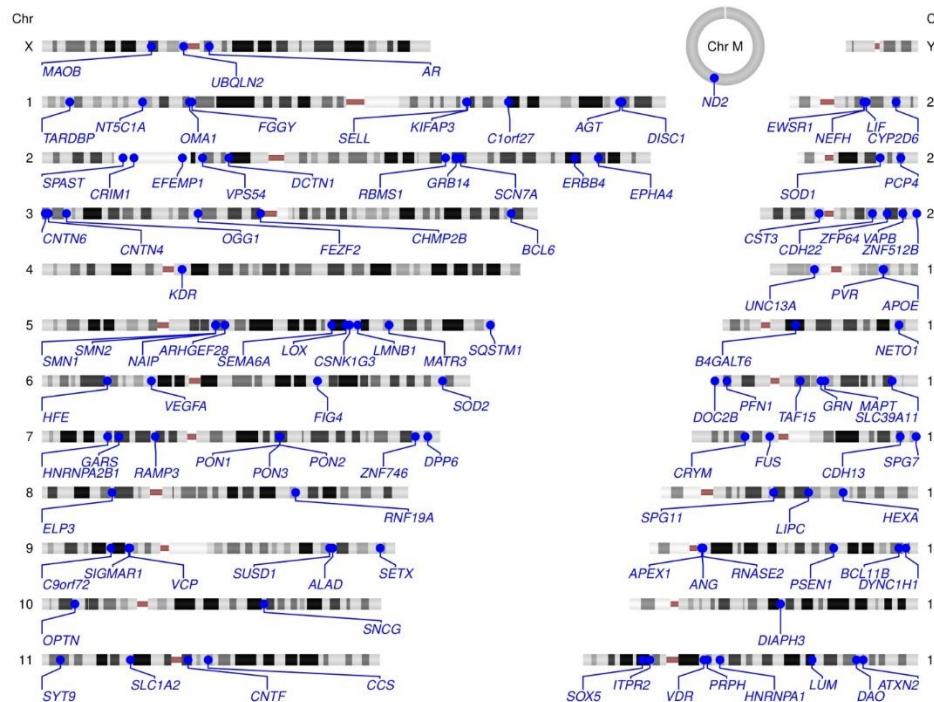


Al-Chalabi *et al.* (2014) Lancet Neuro 13(11):1108-1113

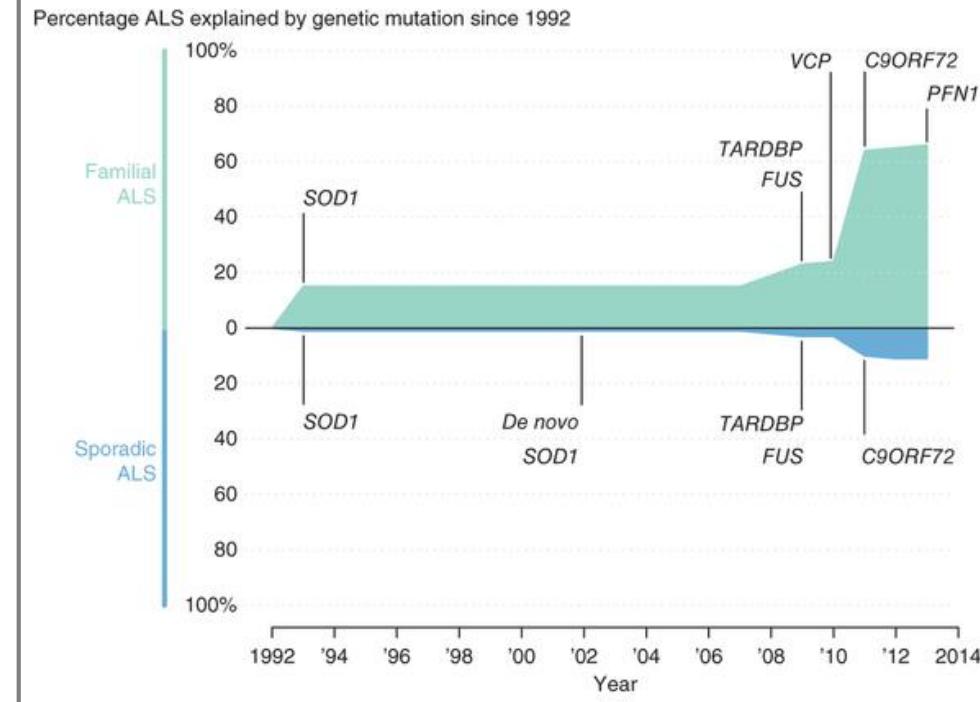


# Genetics of ALS

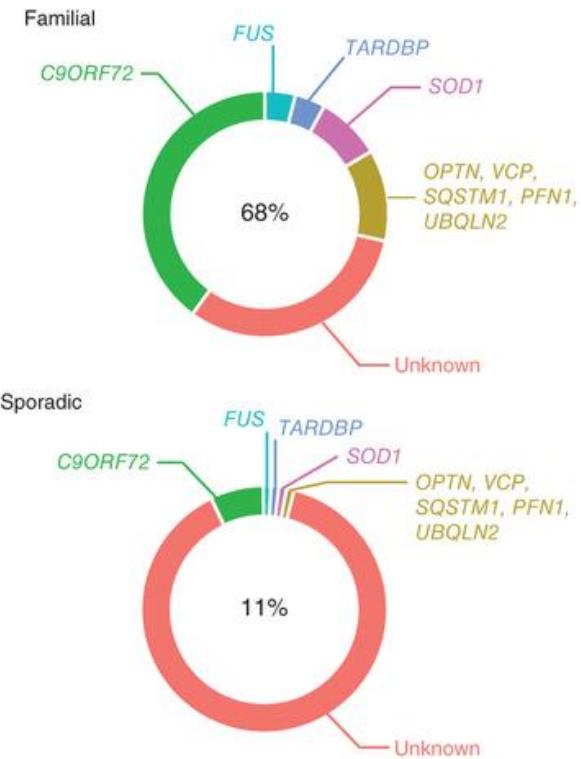
## Genes that have been investigated in ALS



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



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



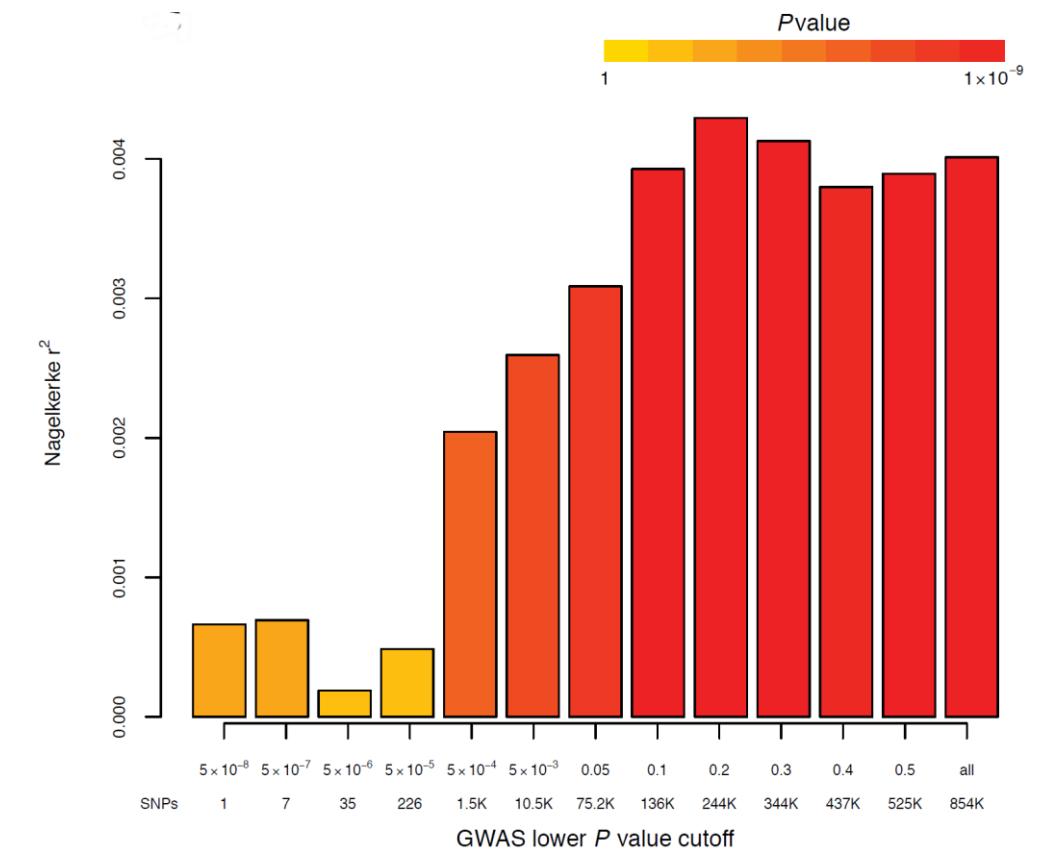
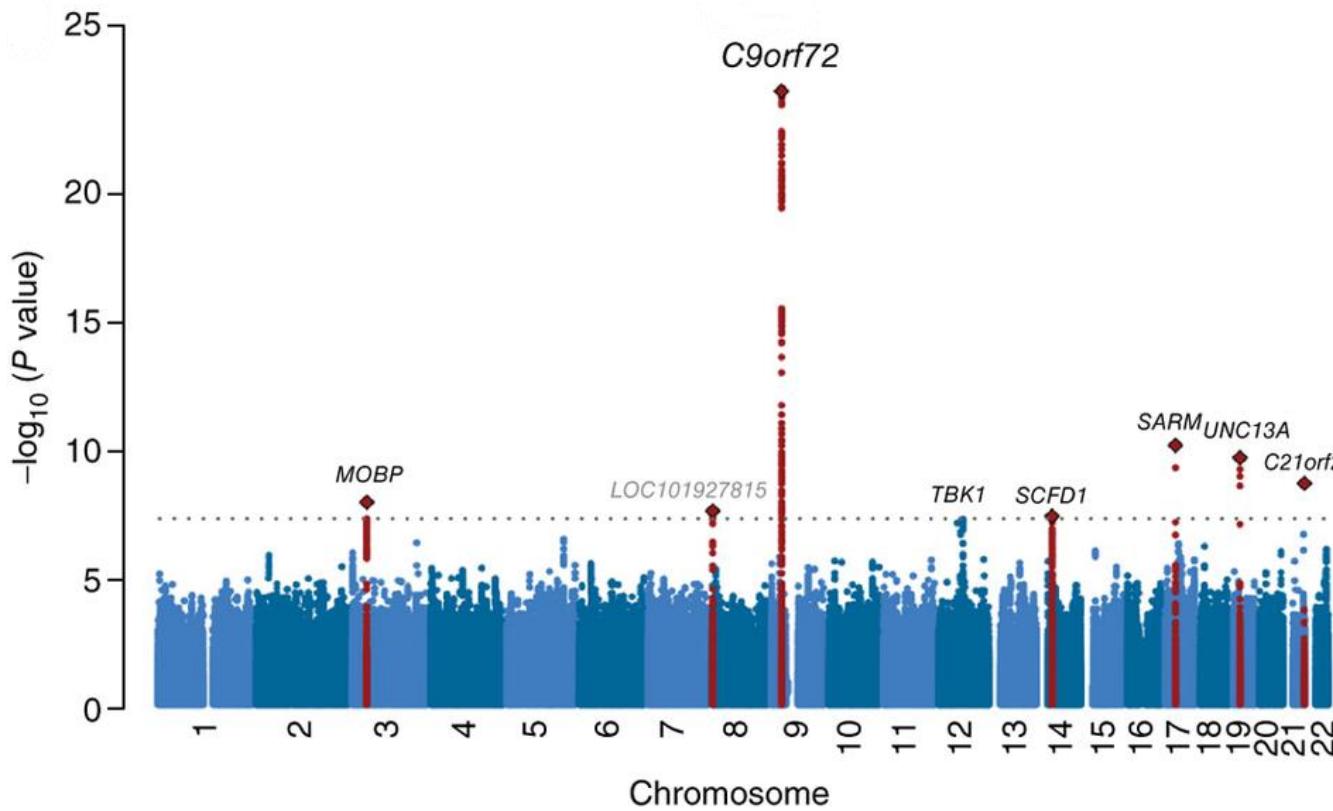
# How do we discover new ALS genes?

## Genome-wide association study (GWAS)

	Person 1	Person 2	Person 3	SNP1	SNP2	SNP...
	... TCAGCCATGCTACT <b>C</b> GATCGACTAA <b>G</b> CG ... (maternal)	... TCAGCCATGCTACT <b>C</b> GATCGACTAAATCG ... (paternal)	... TCAGCCATGCTACT <b>C</b> GATCGACTAA <b>G</b> CG ... (maternal)	<b>Cases</b> Count of G: 2104 of 4000	<b>Cases</b> Count of G: 1648 of 4000	<i>Repeat for all SNPs</i>
		... TCAGCCATGCTACT <b>T</b> GATCGACTAAATCG ... (paternal)	... TCAGCCATGCTACT <b>T</b> GATCGACTAAATCG ... (maternal)	Frequency of G: 52.6%	Frequency of G: 41.2%	
		... TCAGCCATGCTACT <b>T</b> GATCGACTAAATCG ... (maternal)	... TCAGCCATGCTACT <b>T</b> GATCGACTAAATCG ... (paternal)	<b>Controls</b> Count of G: 2676 of 6000	<b>Controls</b> Count of G: 2532 of 6000	
			↑ <b>Single nucleotide polymorphism (SNP)</b>	Frequency of G: 44.6%	Frequency of G: 42.2%	
	If only present in one individual, or if rare	If rare and observed in disease context	If common in the population	P-value:	P-value:	
	Variant or single nucleotide variant (SNV)	Mutation	Single nucleotide polymorphism (SNP)	$5.0 \cdot 10^{-15}$	0.33	

# 2016 ALS GWAS

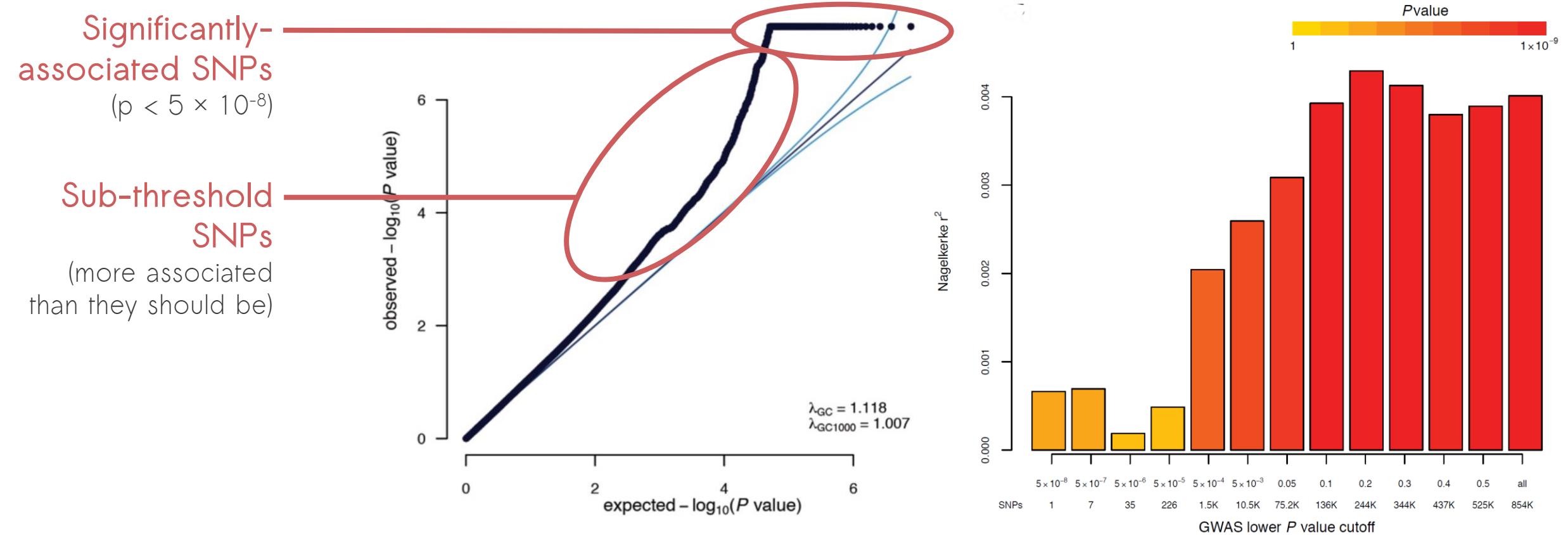
13 countries; 12,577 ALS cases; 23,475 healthy controls



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

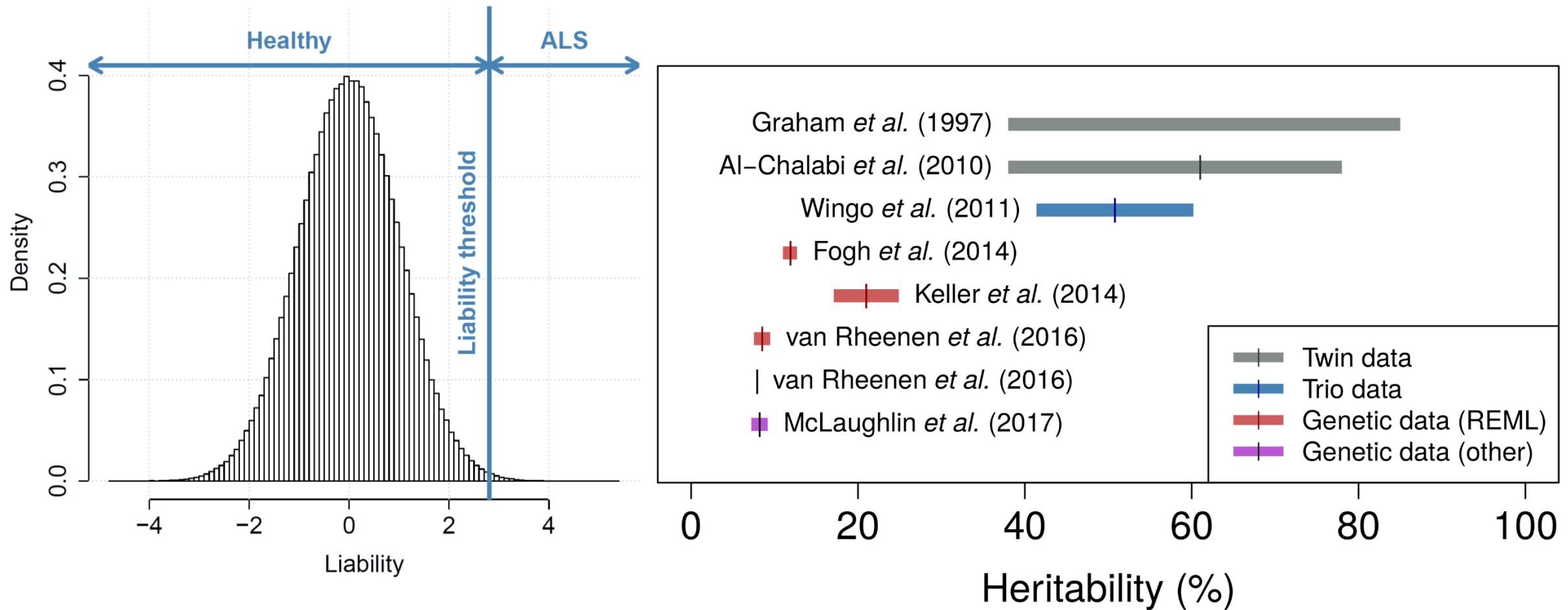
# 2016 ALS GWAS

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



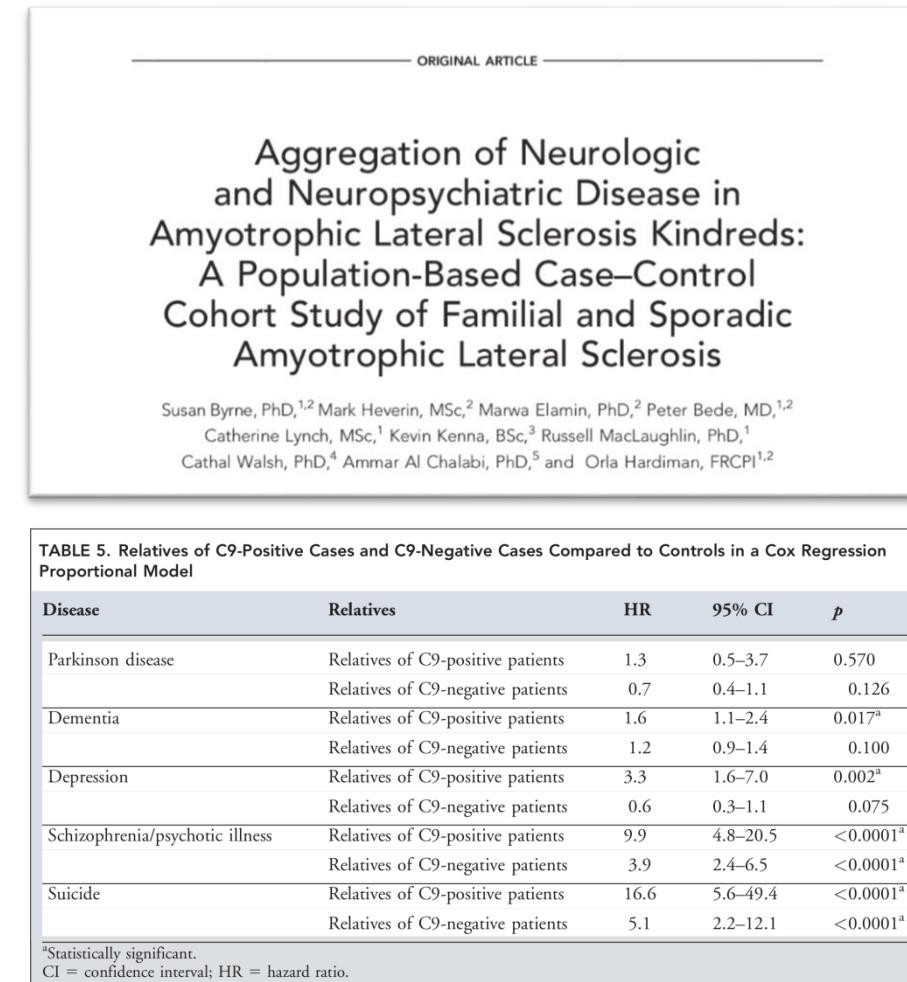
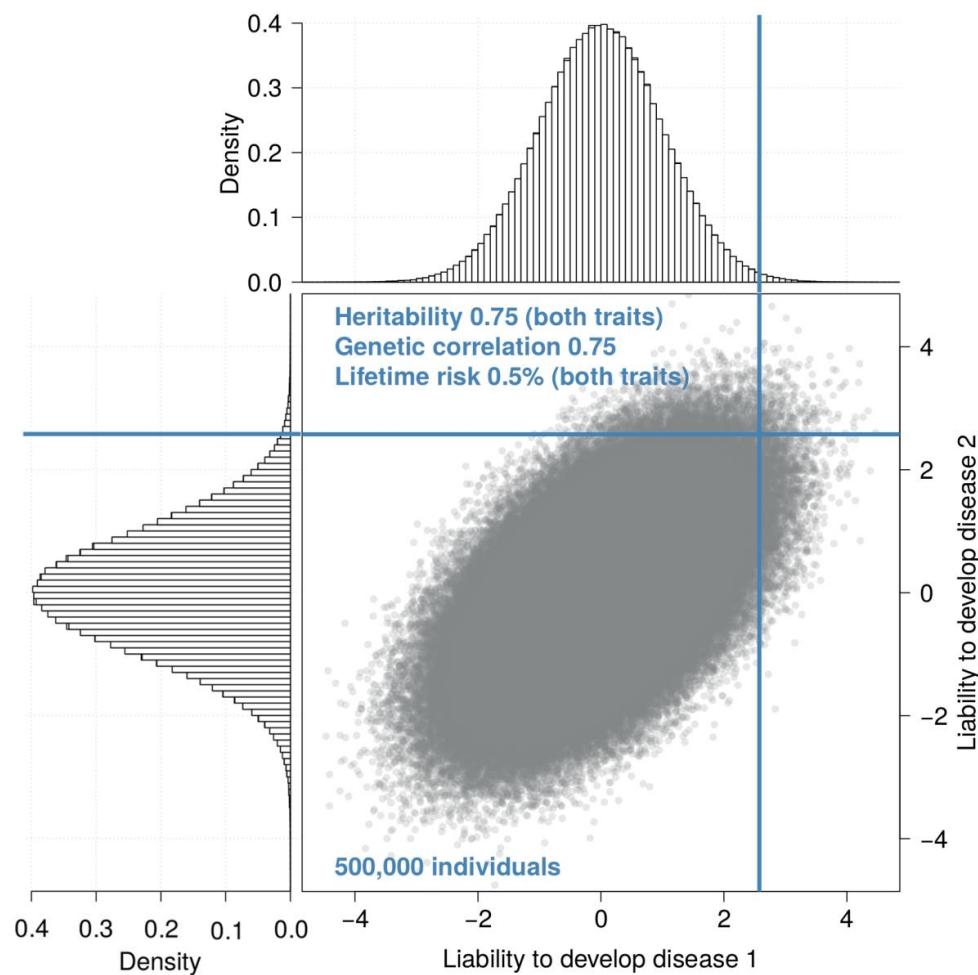
# Heritability revisited

SNP-based heritability (polygenic risk)



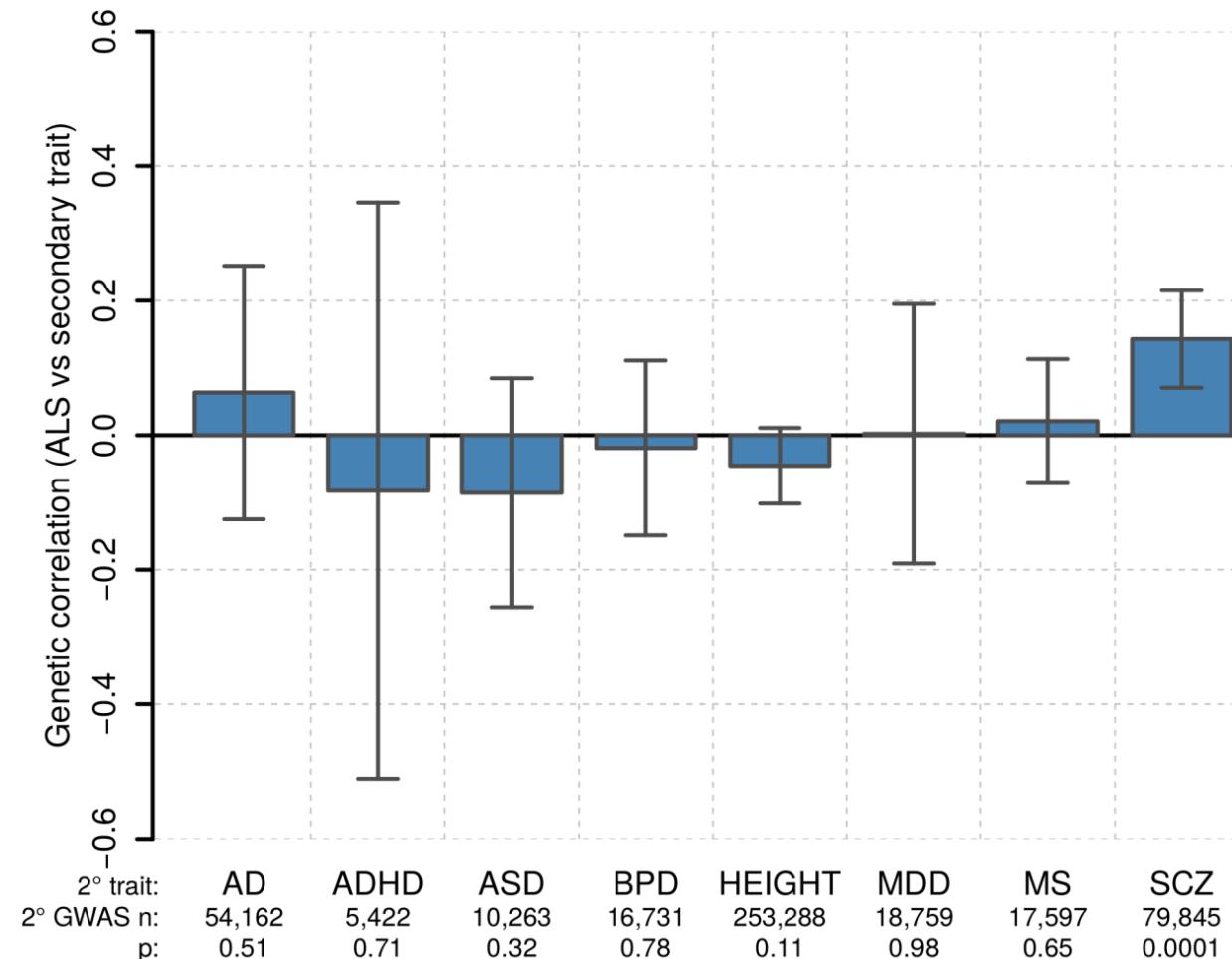
# Genetic correlation

Covariance between two traits due to shared genetic variation



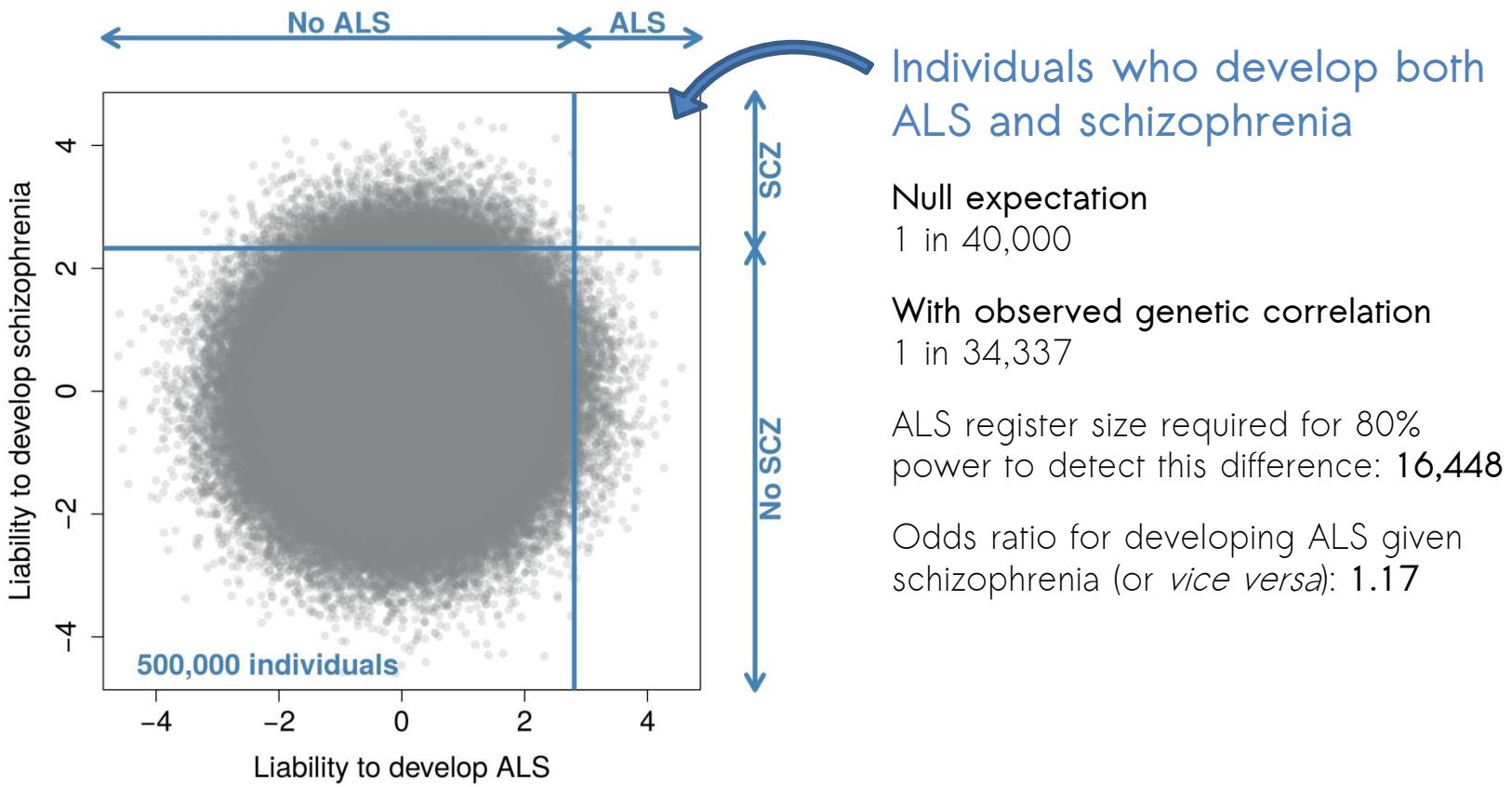
# ALS and schizophrenia are genetically correlated

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



# ALS and schizophrenia are genetically correlated

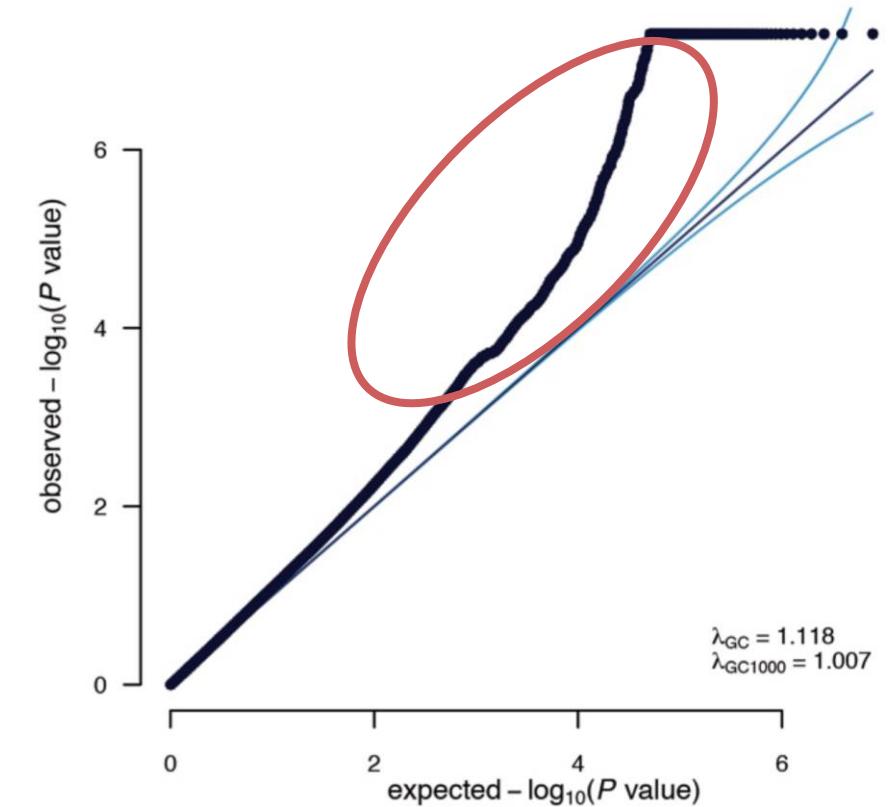
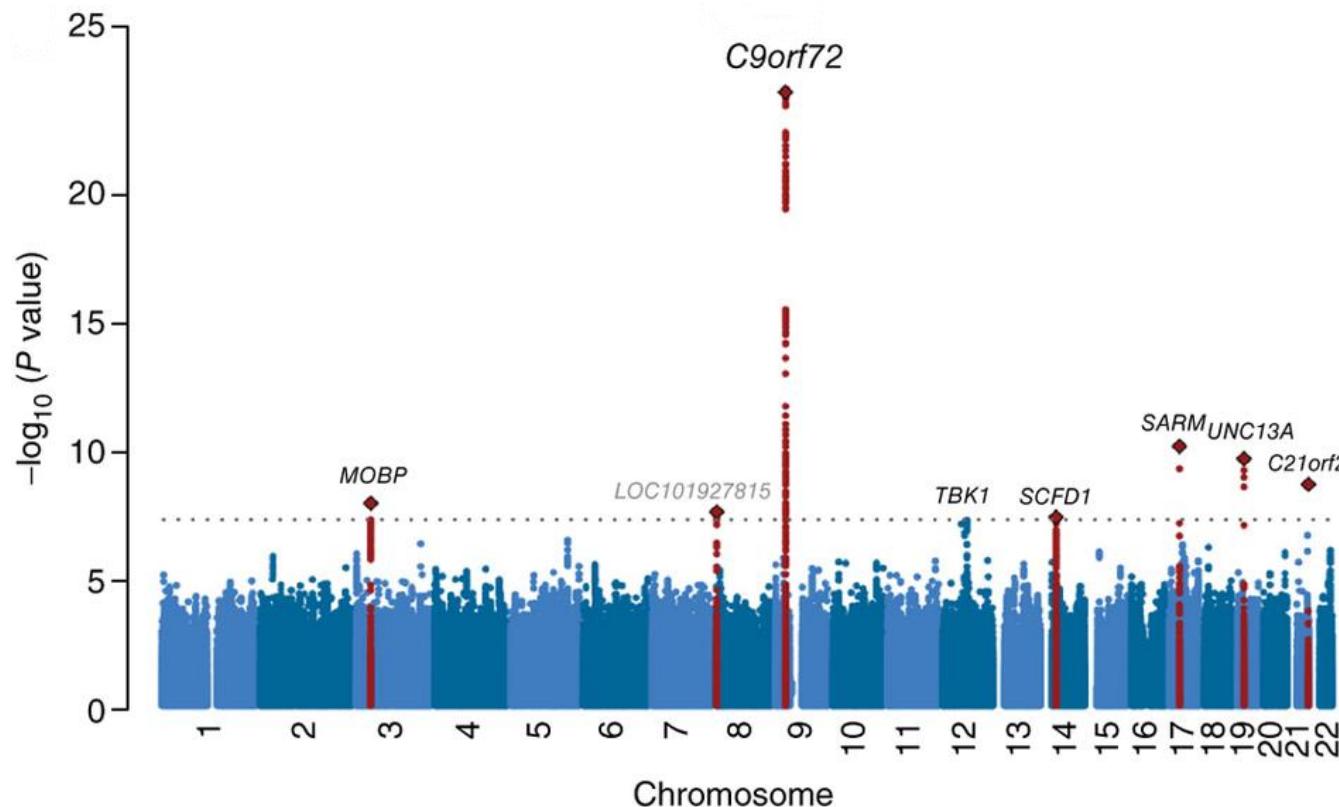
## Implications



# GWAS in ALS

General conclusions and further considerations

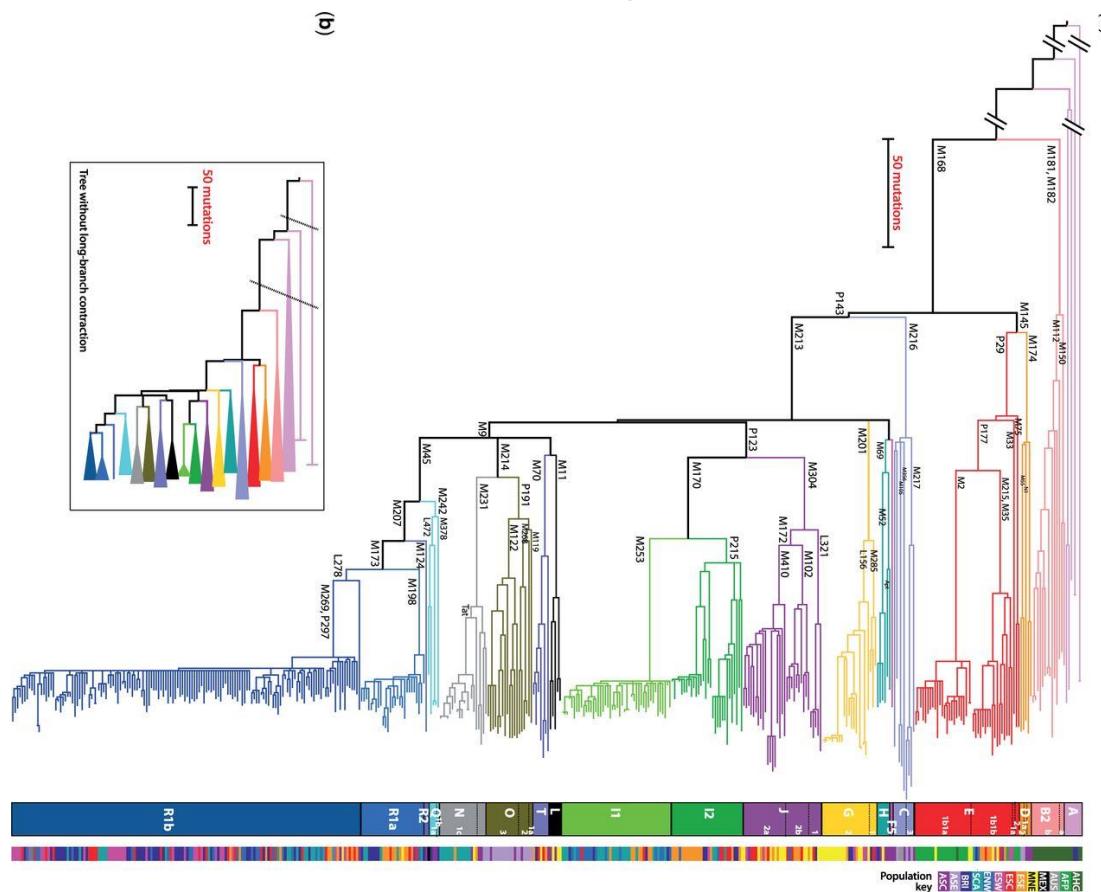
## 1. Bigger GWAS will discover more disease loci



# GWAS in ALS

General conclusions and further considerations

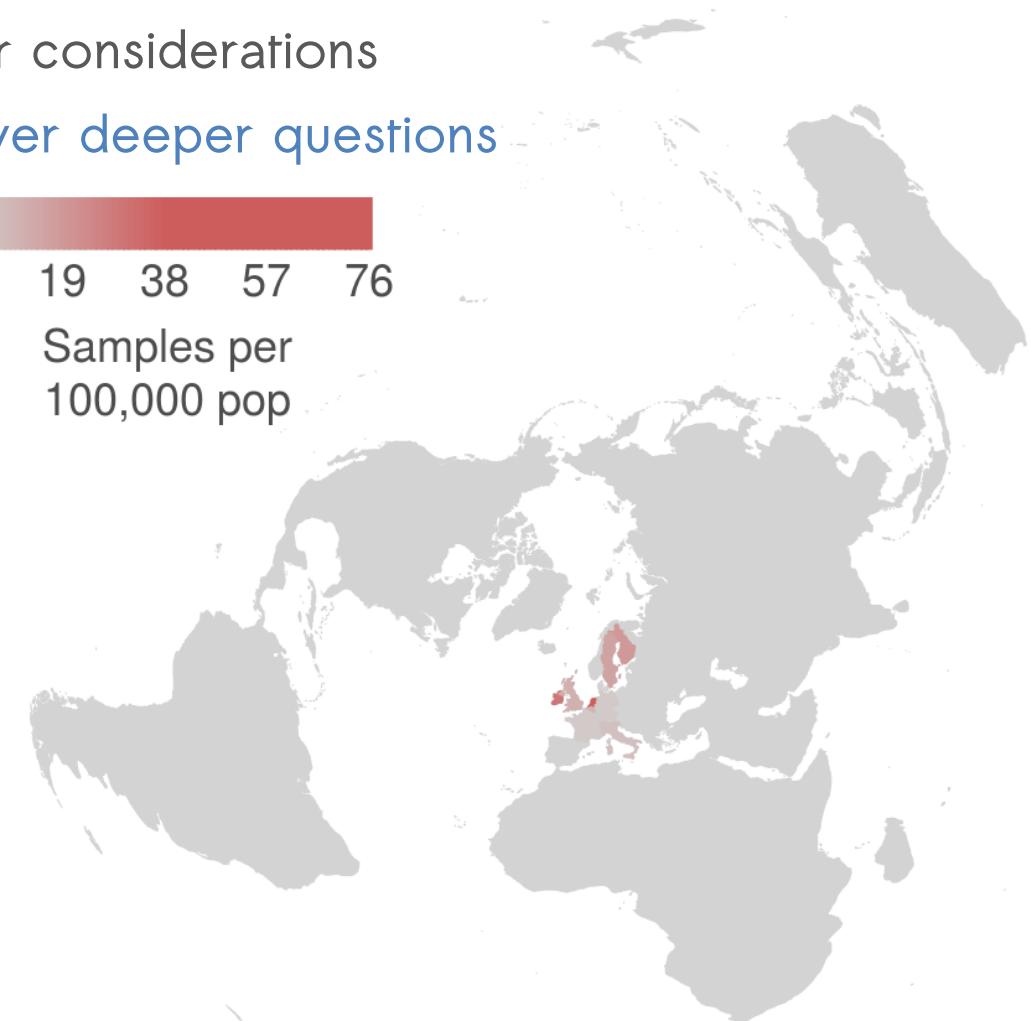
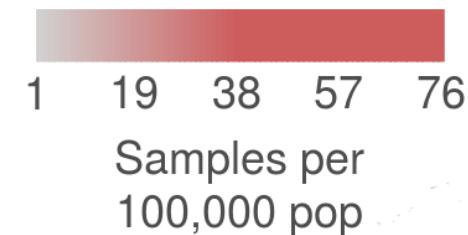
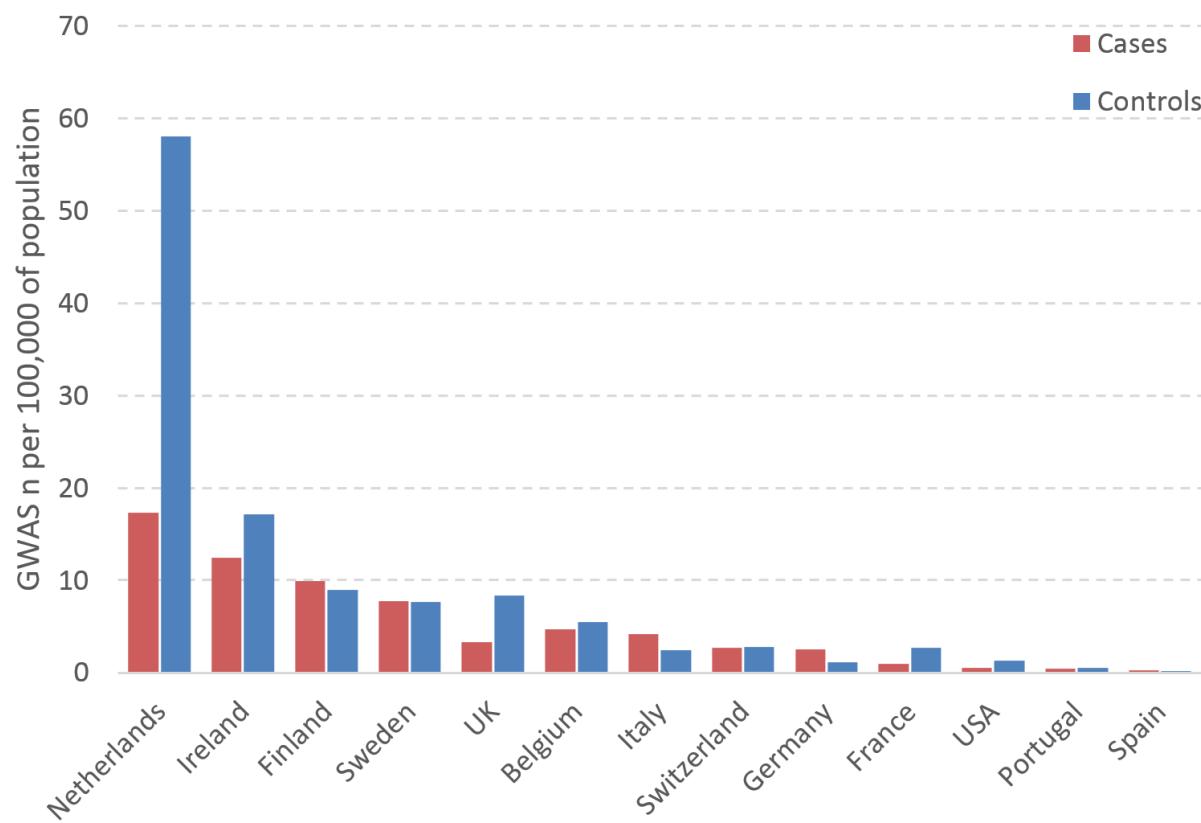
## 2. Involving more countries will answer deeper questions



# GWAS in ALS

General conclusions and further considerations

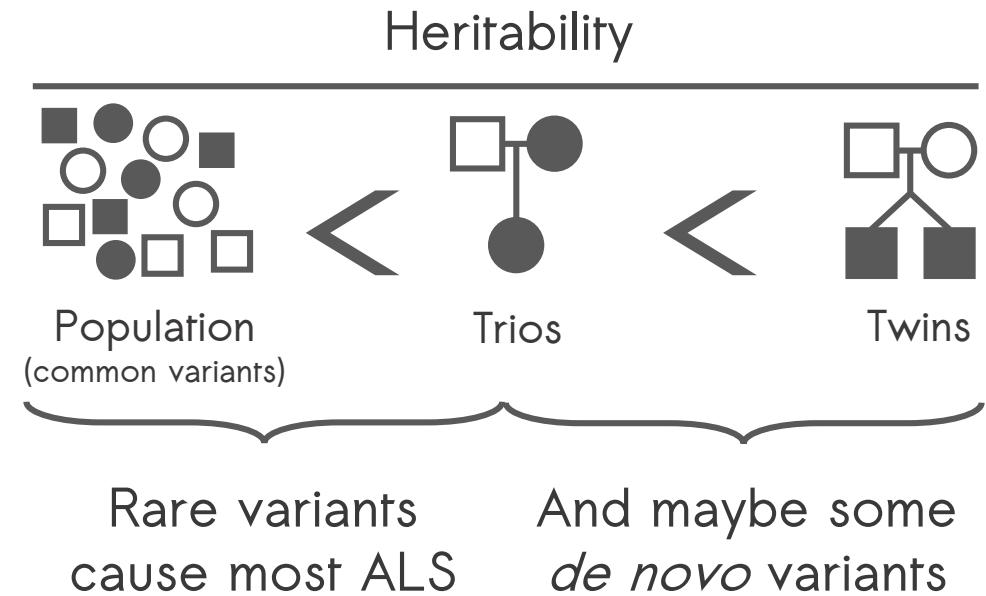
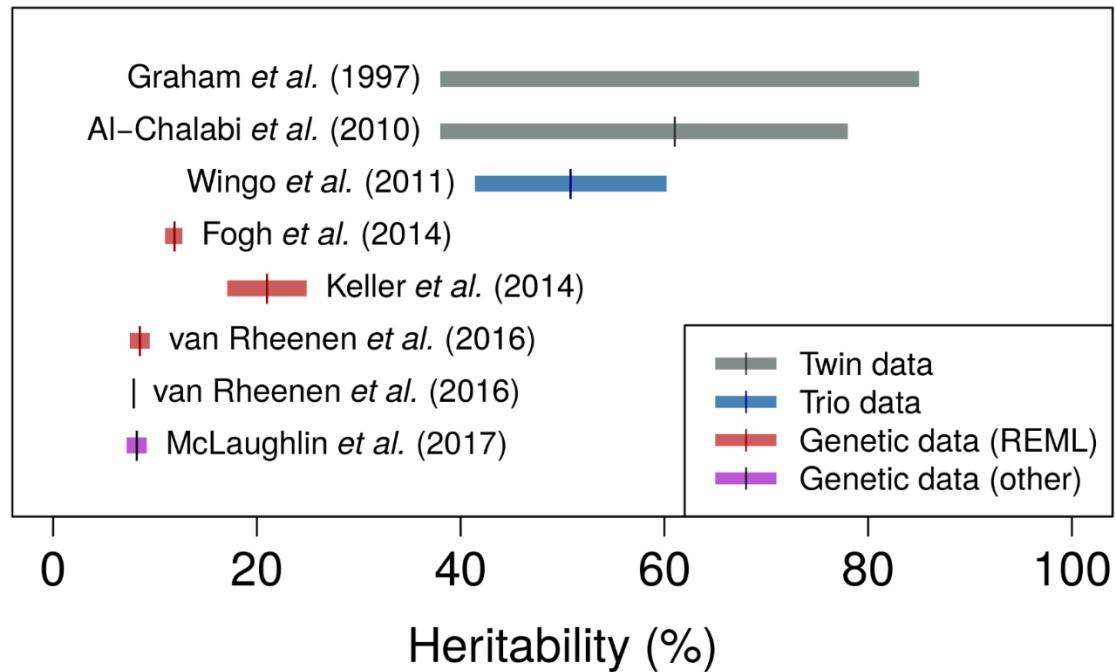
## 2. Involving more countries will answer deeper questions



# GWAS in ALS

General conclusions and further considerations

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



# Finding rare variants

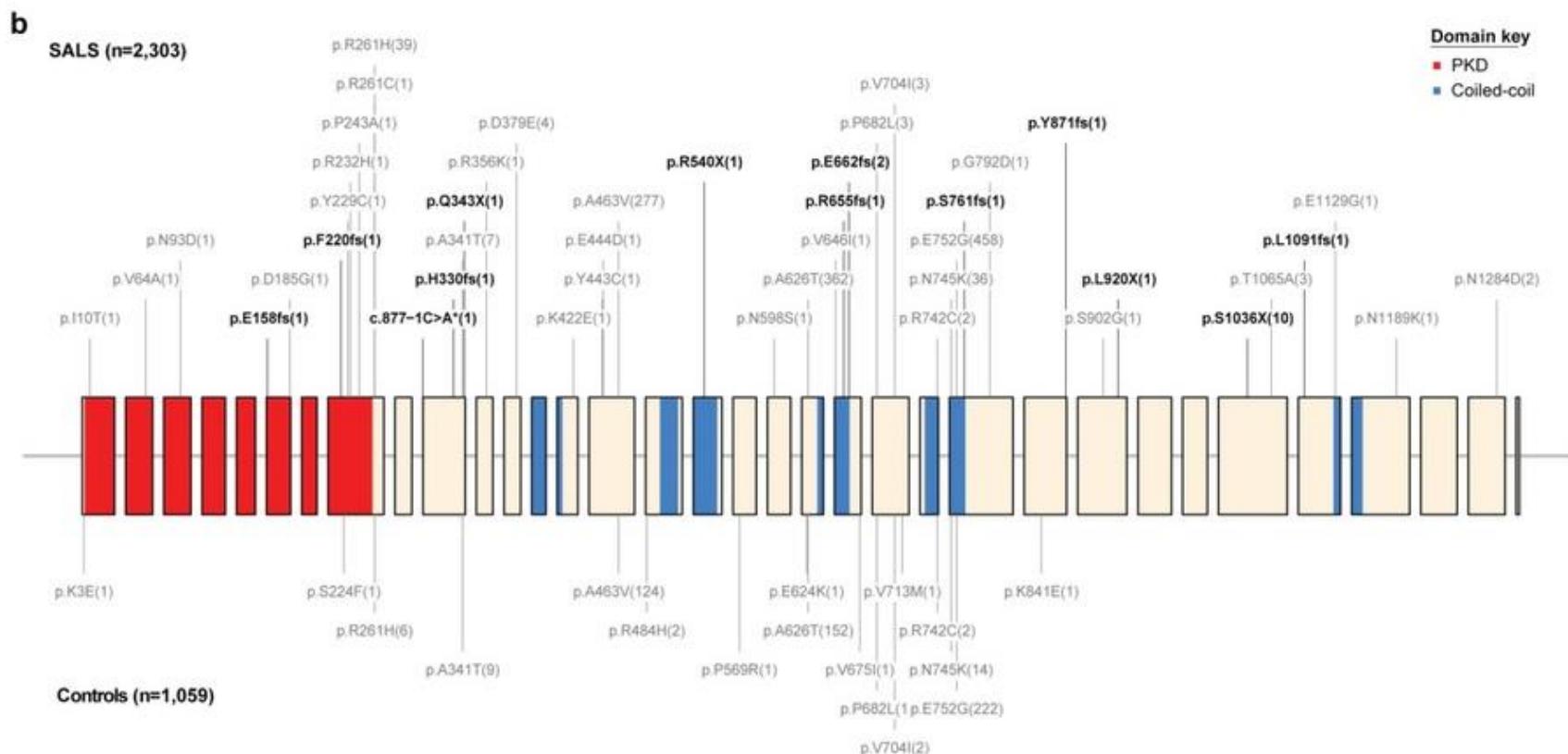
# Exome sequencing

## Previous approach: SNP chips (GWAS)

## New approach: exome sequencing

# Implicating rare variants

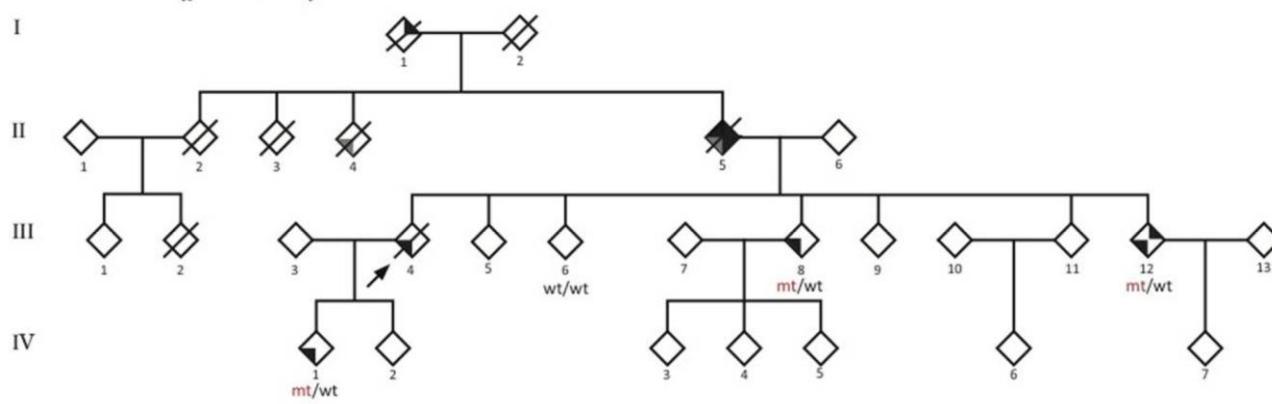
Assessing significance by burden testing



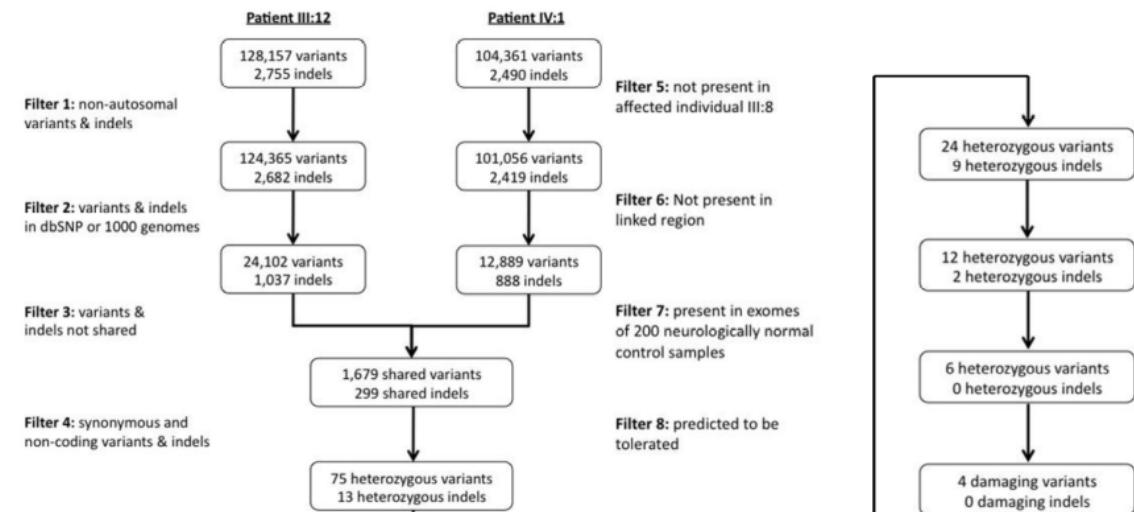
# Implicating rare variants

## Exome sequencing with pedigrees

### A. ITALS#1 (p.R191Q)

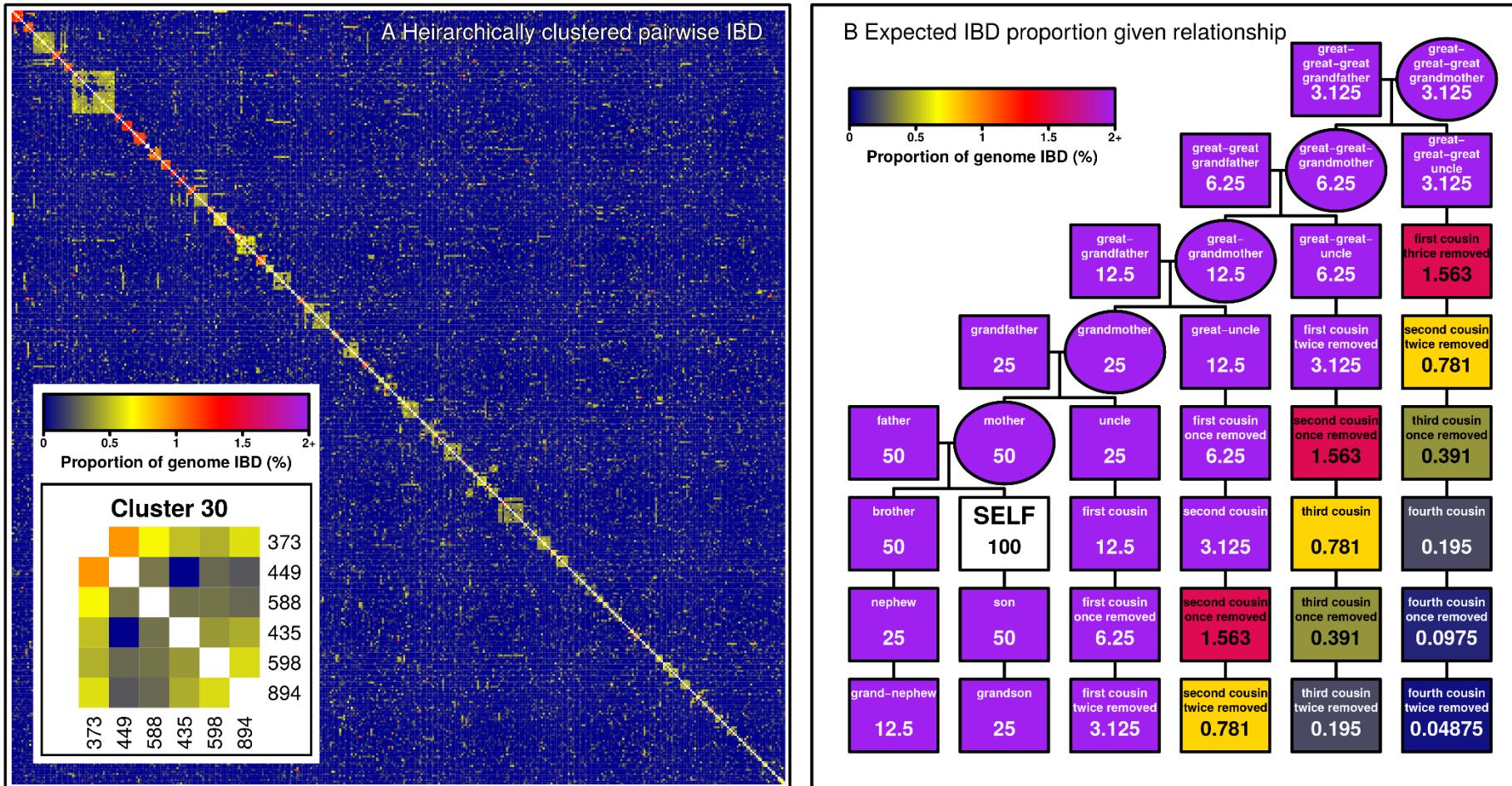


Johnson *et al.* (2010) *Neuron* 68(5):857-64



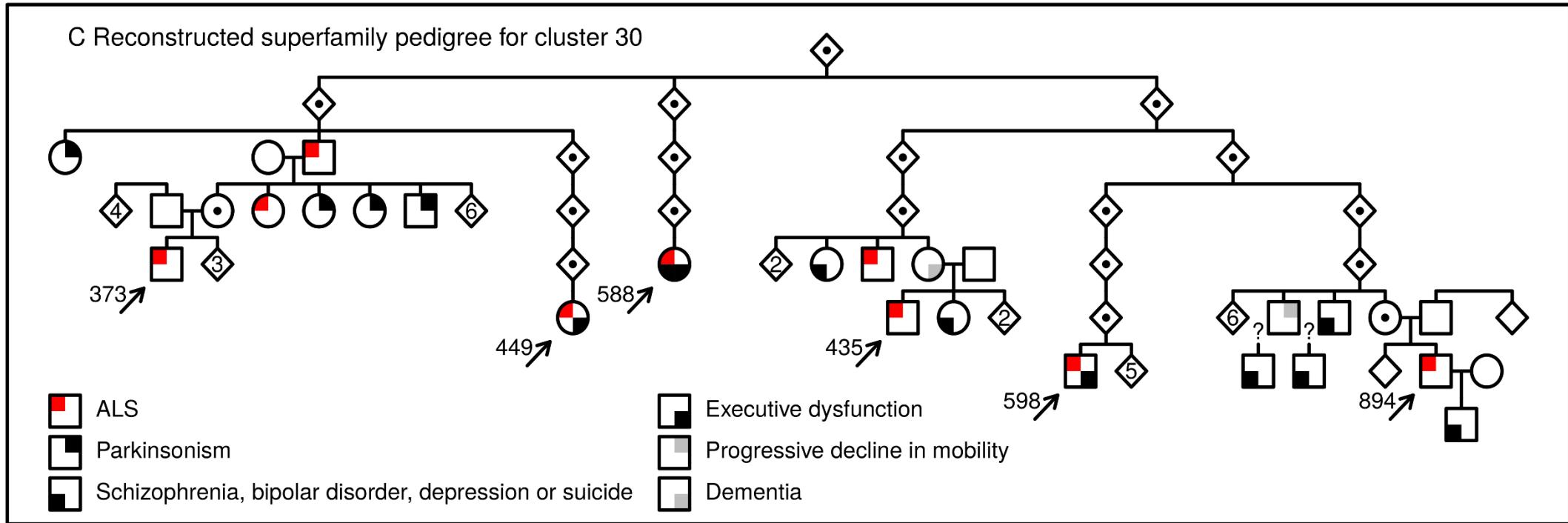
# Implicating rare variants

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



# Implicating rare variants

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



# Exome sequencing in ALS

## Genes discovered using exome sequencing

nature  
genetics



### Exome-wide Rare Variant Analysis Reveals TUBA4A Mutations Associated with Familial Amyotrophic Lateral Sclerosis

#### NEK1 variants confer susceptibility to amyotrophic lateral sclerosis

To identify genetic factors contributing to amyotrophic lateral sclerosis (ALS), we conducted whole-exome analyses of 1,022 index familial ALS (FALS) cases and 7,313 controls. In a new screening strategy, we performed gene-burden analyses trained with established ALS genes and identified a significant association between loss-of-function (LOF) NEK1 variants and FALS risk. We performed a genome-wide search for mutated variants in the Netherlands identified a NEK1 p.Arg611His variant as a candidate risk factor. Replication analyses of sporadic ALS (SALS) cases and independent control cohorts confirmed significant disease association for both p.Arg611His (10,589 samples analyzed) and NEK1 LOF variants (3,362 samples analyzed). In total, we observed NEK1 risk variants in nearly 3% of SALS cases. NEK1 has been linked to several cellular functions, including cilia formation, DNA-damage response, microtubule stability, neuronal morphology and axonal polarity. Our results provide new and important insights into ALS etiopathogenesis and genetic epidemiology.

In recent years, the combination of exome sequencing, segregation analysis and bioinformatic filtering has proven to be an effective strategy to rapidly identify new disease genes<sup>1</sup>. Unfortunately, this method can be difficult to apply to disorders such as ALS, for which late age of onset and low-to-modest variant penetrance make it difficult to obtain large informative multigenerational pedigrees. Owing to high genetic heterogeneity, ALS is also difficult to analyze using filter methods designed to exploit unrelated patient groups<sup>2</sup>. Recently, we had demonstrated the utility of exome-wide rare variant burden (RVB) analysis as an alternate approach, identifying a replicable association for the TUBA4A gene<sup>3</sup>.

In brief, RVB analysis is used to compare the combined frequency of rare variants in each gene in a case-control cohort. Candidate associations are identified by significant differences after multiple test correction. Since this initial study, we extended our data set to include complete exome sequencing for 3,736 index FALS cases and 13,883 controls. Of these, 1,022 cases and 7,315 controls met all required data, inter-relatedness and ancestral quality control criteria (Supplementary Figs. 1 and 2, and Online Methods).

Successful detection of disease associations through RVB analysis can depend heavily on the appropriate setting of test parameters.

As genetic loci often contain many alleles of no or low effect, prior filtering of variants based on minor allele frequency (MAF) and pathogenicity predictors can identify disease signatures otherwise

masked by normal genetic diversity. This strategy predicts disease-associated variants can in turn intro these limitations well-established (Fig. 1a). In our case-control cohorts (Online M expansion in the detect individual, we achieved the highest variants with MAE sense, splice-alter through hidden 3' genes exhibit corrected  $P < 2.5$  FUS/FUSC and TUBA4A (PEN1, YAP1) are disease-associated variants in population-based populations are believed to be

Extension of the coding genes identified a significant disease gene (OR = 8.2,  $P = 1.7$  (never in mutant individuals) across most analysis (nonsense and p. Table 2 and Supplementary Fig. 2) and systematic, gene-specific ascertainment biases in NEK1 8 samples carrying the association (C)

In an independent sequencing for in the Netherlands inbreeding coefficienting their high di

#### SUMMARY

Exome sequencing is an effective strategy for identifying human disease genes. However, this methodology is difficult in late-onset diseases where limited

available prohibits overcomes the need for family history with f

A full list of authors and affiliations appears at the end of the paper.

Received 25 February; accepted 24 June; published online 25 July 2014; doi:10.1038/ng.3626

NATURE GENETICS VOLUME 46 | NUMBER 9 | SEPTEMBER 2014 © Elsevier Inc.

### LETTER

#### Mutations in the profilin 1 gene as a Cause of Familial Amyotrophic lateral sclerosis

Bradley N. Smith,<sup>1,2\*</sup> Nicola Ticozzi,<sup>3,4,5</sup> Claudia Fallini,<sup>3,5,6</sup> Athina S. Emmal,<sup>1</sup> L. Scotter,<sup>1</sup> Jason Kost,<sup>4,6</sup> Claire Troakes,<sup>1</sup> Cinzia Tiloca,<sup>7</sup> Safa Al-Sarraj,<sup>1</sup> Elizabeth A. Lewis,<sup>4</sup> Viviana Pensato,<sup>7</sup> Barbara Castellotti,<sup>7</sup> Jacqueline de Belleroche,<sup>8</sup> Peter C. Sapp,<sup>9</sup> Diane McKenna-Yasek,<sup>4</sup> Russell L. McLaughlin,<sup>5</sup> Michael J. Esteban-Perez,<sup>10</sup> Jose M. Munoz-Blanco,<sup>10</sup> Michael Simpson-Wood,<sup>11</sup> Richard P. Hock,<sup>12</sup> Gianluca Giorgio,<sup>13</sup> Stefano Ghezzi,<sup>14</sup> Lucia Corradi,<sup>15</sup> Gianni Sorani,<sup>16</sup> Karin E. Morrison,<sup>17,18</sup> Kelly L. Wilcock,<sup>19</sup> Patrick A. Dion,<sup>20</sup> Claire S. Lebold,<sup>20</sup> Guy A. Rouleau,<sup>20</sup> Orla Hardiman,<sup>21</sup> Ammar Al-Chalabi,<sup>22</sup> Harden Pall,<sup>21</sup> Pamela J. Shaw,<sup>23</sup> Martin R. Turner,<sup>24</sup> Alberto Garcia-Redondo,<sup>21</sup> Zheyang Wu,<sup>25</sup> Jonathan D. Glass,<sup>10</sup> Cinzia Vassalli,<sup>26</sup> Christopher E. Shaw,<sup>1,26</sup> and John E. Landers<sup>1\*</sup>

<sup>1</sup>Centre for Neurodegenerative Disease, King's College London, Department of Neuroscience, London, SE5 8AF, UK

<sup>2</sup>Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Neurologico Carlo Besta, Italy

<sup>3</sup>Department of Pathophysiology and Transplantation, Dino Farnioli Center – University of Milan, Italy

<sup>4</sup>Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

<sup>5</sup>Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin, Ireland

<sup>6</sup>Department of Biochemistry and Molecular Biology, University of Louisville, Louisville, KY, USA

<sup>7</sup>Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione I.R.C.C.S. Neuromed, Pozzilli, Italy

<sup>8</sup>Department of Cell Biology, Imperial College London, London, WC2R 8EX, UK

<sup>9</sup>Department of Genetics and Laboratory of Molecular Medicine, National Institute of Child Health and Human Development, Bethesda, MD, USA

<sup>10</sup>Department of Cell Biology, Emory University School of Medicine, Atlanta, GA 30322, USA

<sup>11</sup>Unidad de ELA, Instituto de Investigación Hospital 12 de Octubre de Madrid, Madrid, Spain

<sup>12</sup>Unidad de ELA, Instituto de Investigación Hospital Gregorio Marañón de Madrid, Madrid, Spain

<sup>13</sup>Department of Genetics and Molecular Medicine, King's College London, London, WC2R 8QH, UK

<sup>14</sup>Department of Neuroscience, Brain Center Rudolf Magnus, Institute of Neuroscience, Utrecht University, Utrecht, The Netherlands

<sup>15</sup>3rd Neurology Unit, Fondazione IRCCS Istituto Neurologico "Carlo Besta", Milan, Italy

<sup>16</sup>Department of Medical Biotechnology and Translational Medicine - University of Milan, Italy

<sup>17</sup>Neurology Unit, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

<sup>18</sup>Experimental Neurobiology Laboratory, IRCCS C. Mondino, National Research Council, Pavia, Italy

<sup>19</sup>Department of Health Sciences, Interdisciplinary Research Center of Automatic Novara, Italy

<sup>20</sup>Department of Neurosciences, University of Padova, 35122 Padova, Italy

<sup>21</sup>School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, Queen Elizabeth Hospital, Birmingham, UK

<sup>22</sup>Australian School of Advanced Medicine, Macquarie University, Sydney, NSW, Australia

<sup>23</sup>Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada

<sup>24</sup>Department of Clinical Neuroscience, Medical Research Council Centre for Nervous System Repair, King's College London, London, WC2R 2LS, UK

<sup>25</sup>Scrieff Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK

<sup>26</sup>Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

\*Correspondence: john.landers@masmmed.edu

http://dx.doi.org/10.1016/j.neuro.2014.09.027

available prohibits overcomes the need for family history with f

### Neuron Report

#### Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS

Janet O. Johnson,<sup>1,2\*</sup> Jessica Mandrioli,<sup>4,22</sup> Michael Benatar,<sup>5,22</sup> Yevgeniya Abrahams,<sup>1</sup> John O. Trojanowski,<sup>4</sup> J. Raphael Gibbs,<sup>2,6</sup> Maura Brunetti,<sup>9</sup> Susan Gronka,<sup>3</sup> John M. Palmer-Lewis,<sup>1</sup> David Fischbeck,<sup>4</sup> Dena M. Hernandez,<sup>10</sup> Sampath Arayil,<sup>11</sup> Jeffrey Rothstein,<sup>12</sup> Francisco Garcia,<sup>13</sup> Daniel D. Wang,<sup>14</sup> Andrea Calvo,<sup>14</sup> Gianna Maria Rosaria Monzuro,<sup>15</sup> Giuliana Battistini,<sup>17</sup> Fabrizio Merello,<sup>18</sup> Rossella Spataro,<sup>19</sup> ITALSGEN Consortium,<sup>20</sup> Giovanna Galassi,<sup>21</sup> Fabrizio Merello,<sup>18</sup> Rossella Spataro,<sup>19</sup> Adriano Chiò,<sup>13,24</sup> and Bryan J. Traynor,<sup>4,10,24,\*</sup>

<sup>1</sup>Neuromuscular Diseases Research Group

<sup>2</sup>Computational Biology Core

<sup>3</sup>Medical and Molecular Genetics, University of Michigan, Ann Arbor, MI, USA

<sup>4</sup>Department of Neuroscience, Porter Neuroscience Building, NIA, NIH, Bethesda, MD 20892, USA

<sup>5</sup>Department of Neuroscience, Emory University School of Medicine, Atlanta, GA 30322, USA

<sup>6</sup>Department of Pathology and Laboratory Medicine

<sup>7</sup>University of Pennsylvania, Philadelphia, PA, USA

<sup>8</sup>Department of Molecular Neuroscience and Retina Lab Weston Laboratories, Institute of Neurology, London, WC1N 3BG, UK

<sup>9</sup>Molecular Genetics Unit, Department of Clinical Pathology, A.S.O. I.R.M.-S. Anna, 10123, Milan, Italy

<sup>10</sup>Department of Neurology, Johns Hopkins Hospital, Baltimore, MD 21287, USA

<sup>11</sup>Department of Developmental Biology, St. Jude Children's Research Hospital, Memphis, TN, USA

<sup>12</sup>Department of Neuroscience, University of Turin, 10126 Turin, Italy

<sup>13</sup>ALS Center, Salter Mauger Foundation, 20010 Milan, Italy

<sup>14</sup>Neurological Institute, Catholic University and I.C.O.M.M. Association for ALS Research, Rome, Italy

<sup>15</sup>Department of Neuroscience, Sapienza University of Rome, 00138 Rome, Italy

<sup>16</sup>Department of Neurology, San Raffaele University Hospital, Milan, Italy

<sup>17</sup>Cancer for Diagnoses and Cures of Rare Diseases, Department of Neurology, Bellaria Hospital, 50139 Florence, Italy

<sup>18</sup>Neurological Clinic, Azienda Ospedaliera Universitaria Cagliari, 09042 Monserrato-Cagliari, Italy

<sup>19</sup>Department of Neuroscience, Georgetown University, Washington, DC 20057, USA

<sup>20</sup>For these authors contributions equal to this work see the Supplemental Information

<sup>21</sup>These authors contributed equally to this work

<sup>22</sup>Correspondence: traynor@mailbox.jhu.edu

DOI 10.1016/j.neuro.2014.10.036

#### RESEARCH ARTICLE AMYOTROPHIC LATERAL SCLEROSIS

#### Mutations in the vesicular trafficking protein annexin A11 are associated with amyotrophic lateral sclerosis

Bradley N. Smith<sup>1,\*</sup>, Simon D. Topp<sup>1,\*</sup>, Claudia Fallini<sup>2,4</sup>, Hideki Shibata<sup>3,5</sup>, Han-Jou Chen<sup>1,6</sup>, Claire Troakes<sup>1</sup>, Andrew Kin... See all authors and affiliations

Science Translational Medicine 03 May 2012  
Vol. 4, Issue 388, eaad9157  
DOI: 10.1126/scitranslmed.aad9157

Article Figures & Data Info & Metrics eLetters PDF

You are currently viewing the abstract.

View Full Text ▶

#### Annexing another protein in ALS pathogenesis

Amyotrophic lateral sclerosis (ALS) causes progressive paralysis due to motor neuron degeneration. Smith *et al.* performed exome sequencing of 751 familial ALS cases and discovered six missense mutations in the *ANXA11* gene in 13 individuals, which were absent or vanishingly rare in ~70,000 healthy controls. Abundant annexin 11 protein inclusions were detected in spinal motor neurons and hippocampal axons in a patient with the p.D40G mutation. Annexin 11 is known to play a role in vesicular trafficking between the Golgi and endoplasmic reticulum. Functional studies in transfected cells revealed abnormal binding of mutant annexin 11 to calcyclin, which implicates defective intracellular protein trafficking in ALS pathogenesis.

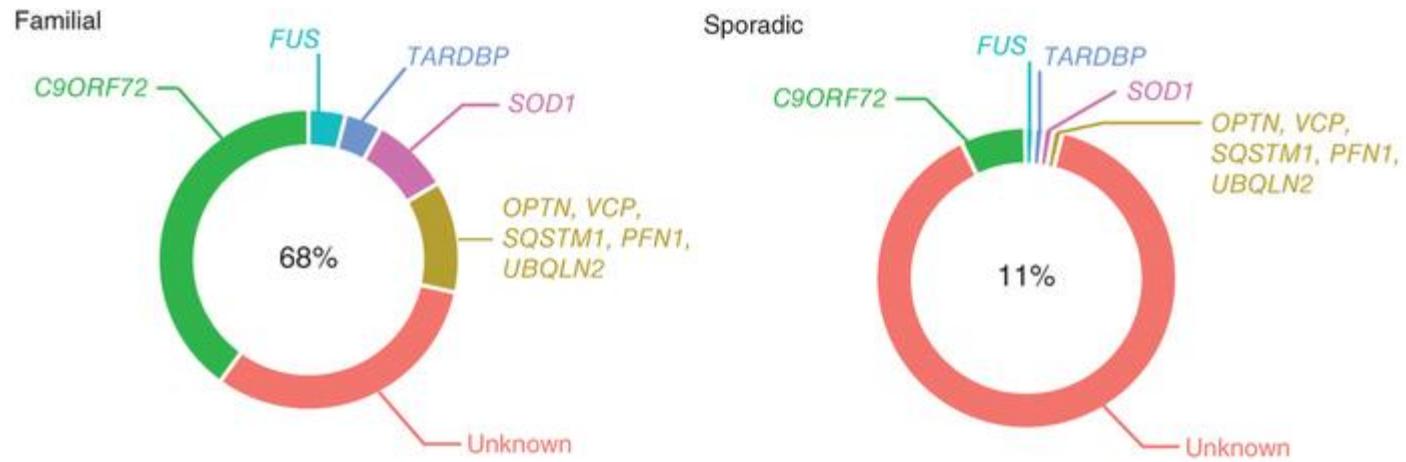
#### Abstract

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disorder. We screened 751 Italian ALS patient whole-exome sequences and identified six mutations including p.D40G in the *ANXA11* gene in 13 individuals. The p.D40G mutation was absent from ~70,000 control whole-exome sequences. This mutation segregated with disease in two kindreds and was present in two other unrelated cases ( $P = 0.0102$ ), and all mutation carriers shared a common founder haplotype. Annexin A11-positive protein aggregates were abundant in spinal cord motor neurons and hippocampal neuronal axons in an ALS patient carrying the p.D40G mutation. Transfected human embryonic kidney cells expressing *ANXA11* with the p.D40G mutation and other N-terminal mutations showed altered binding to calcyclin, and the p.R235Q mutant protein formed insoluble aggregates. We conclude that mutations in *ANXA11* are associated with ALS and implicate defective intracellular protein trafficking in disease pathogenesis.

Neuron 68, 857–866

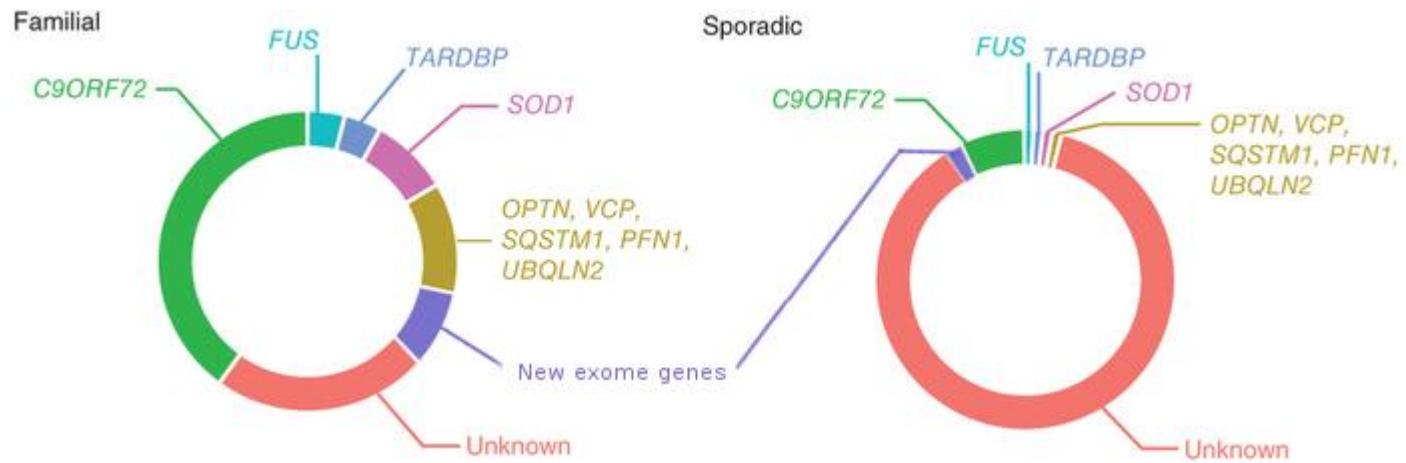
# Updating ALS genomics

One gene at a time



# Updating ALS genomics

One gene at a time



# From exome to whole-genome sequencing

## Original approach: SNP chips (GWAS)

## New approach: exome sequencing

Newer approach: whole-genome sequencing

# Whole-genome sequencing in large populations

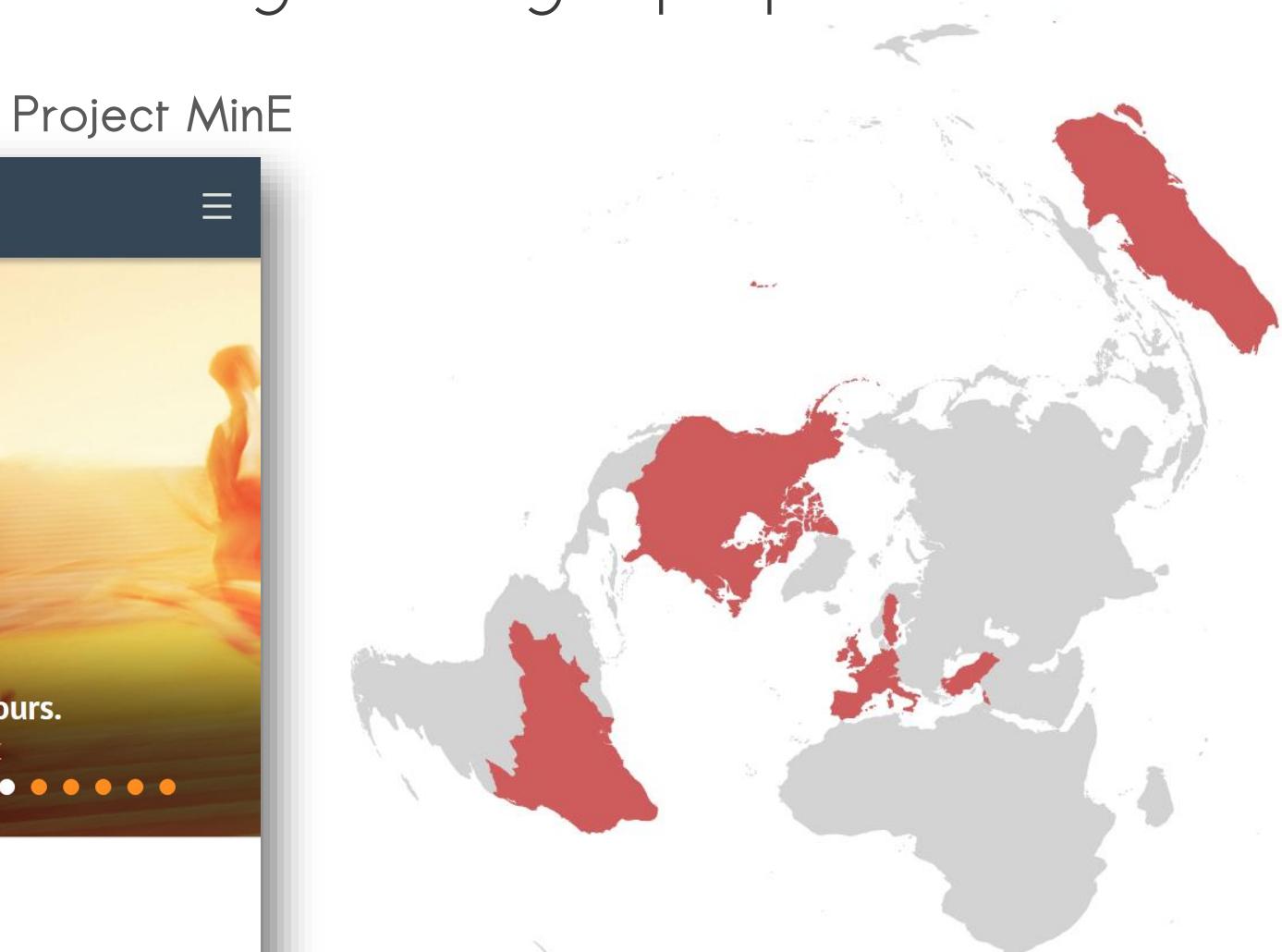
project  
**MinE** Make it yours

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)

...GACCACGCCCGGGCCCCGGCCCCCTAGCGCGCGACT...

Healthy individual (typical)

...GACCACGCCCGGGCCCC-----GGCCCTAGCGCGCGACT...

*C9orf72*-positive ALS

...GACCACGCCCGGGCCCC(G<sub>2</sub>C<sub>4</sub>)<sub>n</sub>GGCCCTAGCGCGCGACT...

(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-pallidoluysian 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
<i>ATXN3</i>	Amyotrophic lateral sclerosis	CAG	Coding	27-33
<i>ATXN7</i>	Spinocerebellar ataxia type 3	CAG	Coding	55-84
<i>ATXN8</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) <sub>4</sub> G(C) <sub>4</sub> 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
<i>FMR1</i>	Fragile X mental retardation type 1	CGG	Noncoding	>200
<i>FXN</i>	Fragile X-associated tremor ataxia syndrome	CGG	Noncoding	55-200
<i>HTT</i>	Friedreich's ataxia	GAA	Noncoding	66-1700
<i>JPH3</i>	Huntington's disease	CAG	Coding	>35
<i>NOP56</i>	Huntington's disease-like 2	CAG/CTG	Noncoding	>41
<i>PABPN1</i>	Spinocerebellar ataxia type 36	GGCCTG	Noncoding	1500-2500
<i>PPP2R2B</i>	Oculopharyngeal muscular dystrophy	GCG	Coding	11-17
<i>TBP</i>	Spinocerebellar ataxia type 12	CAG/CTG	Noncoding	55-78
<i>TK2-BEAN</i>	Spinocerebellar ataxia type 17	CAG	Coding	49-66
	Spinocerebellar ataxia type 31	TGGAA	Noncoding	500-760

Pathogenic range is the number of repeats required to manifest disease

# Some questions in ALS genomics

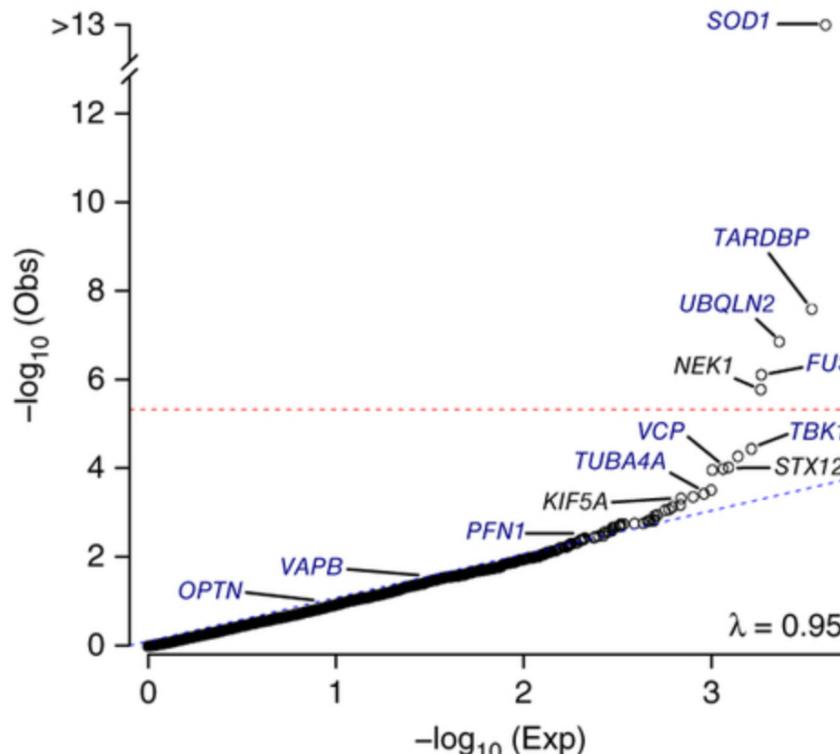
## (and some possible answers)

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

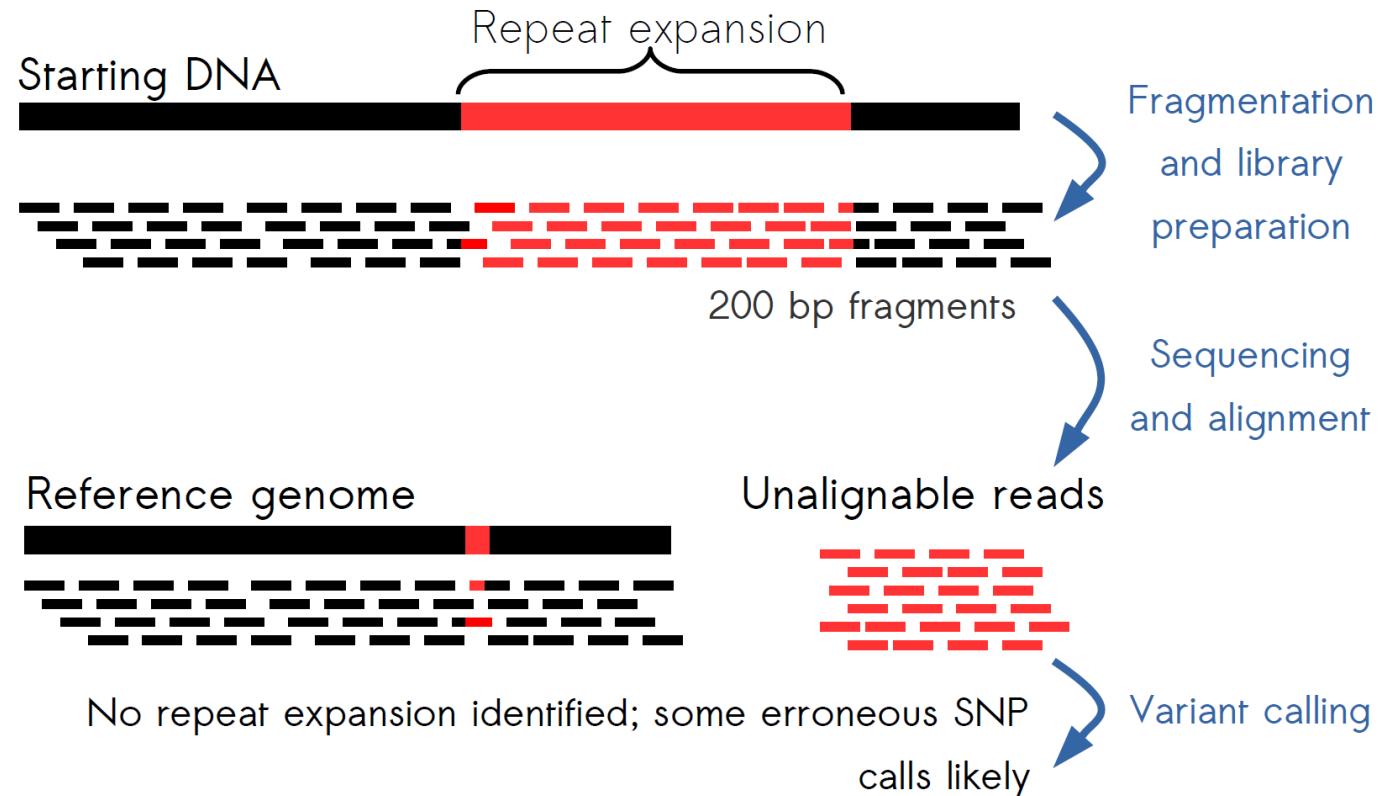
(and some possible answers)

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

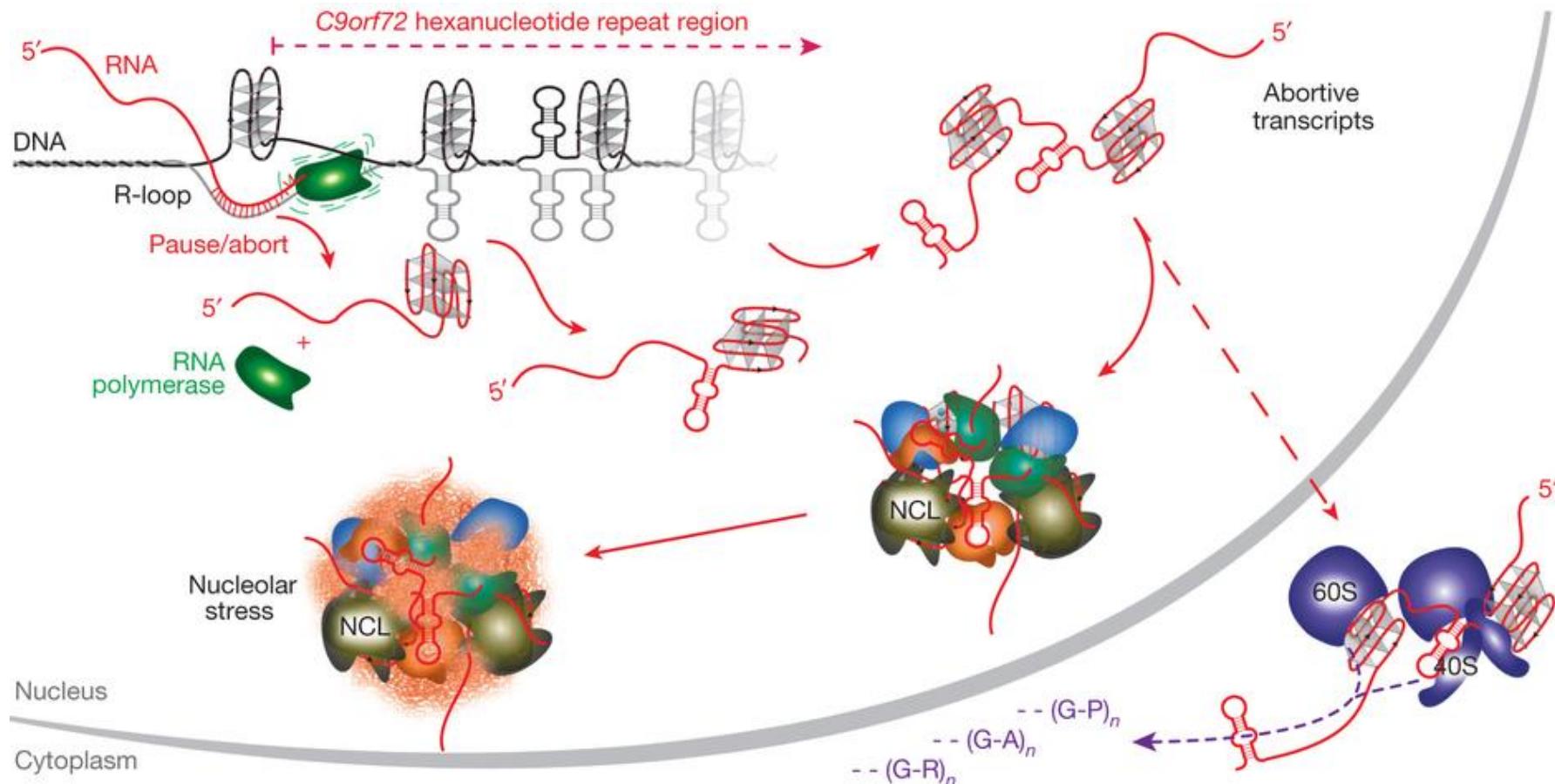
# Some questions in ALS genomics

## (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

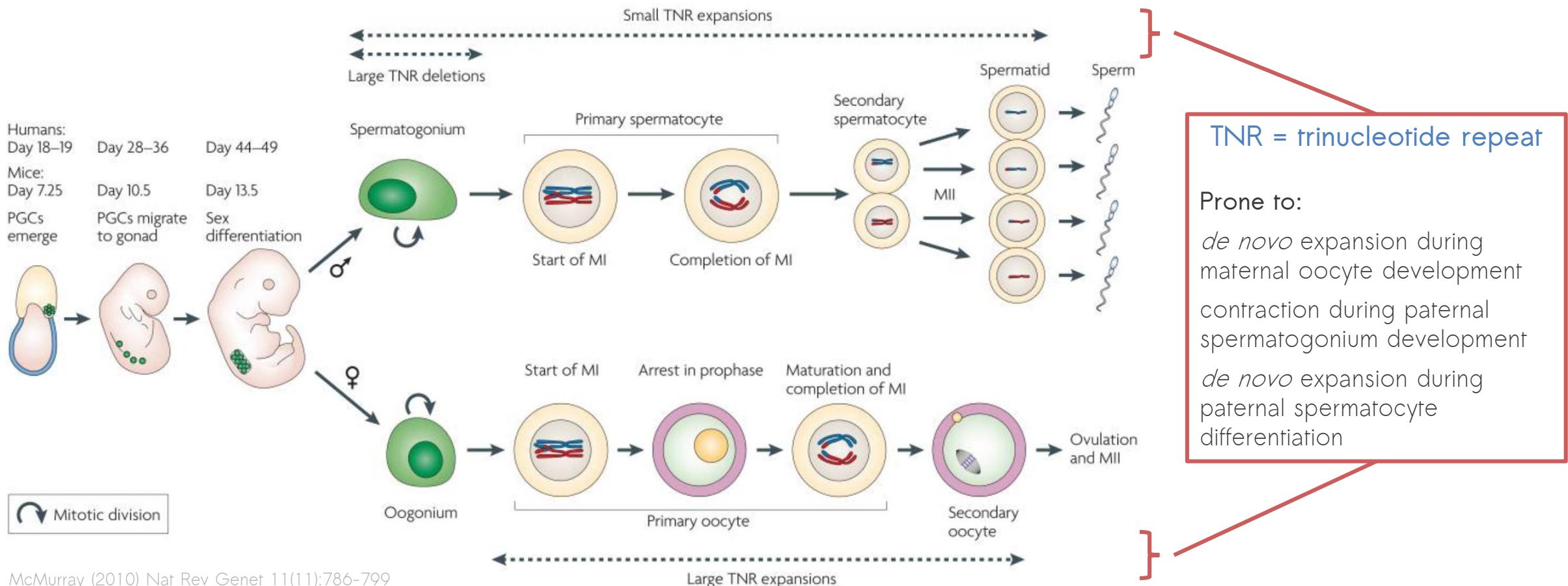
# Some questions in ALS genomics

## (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



# 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?

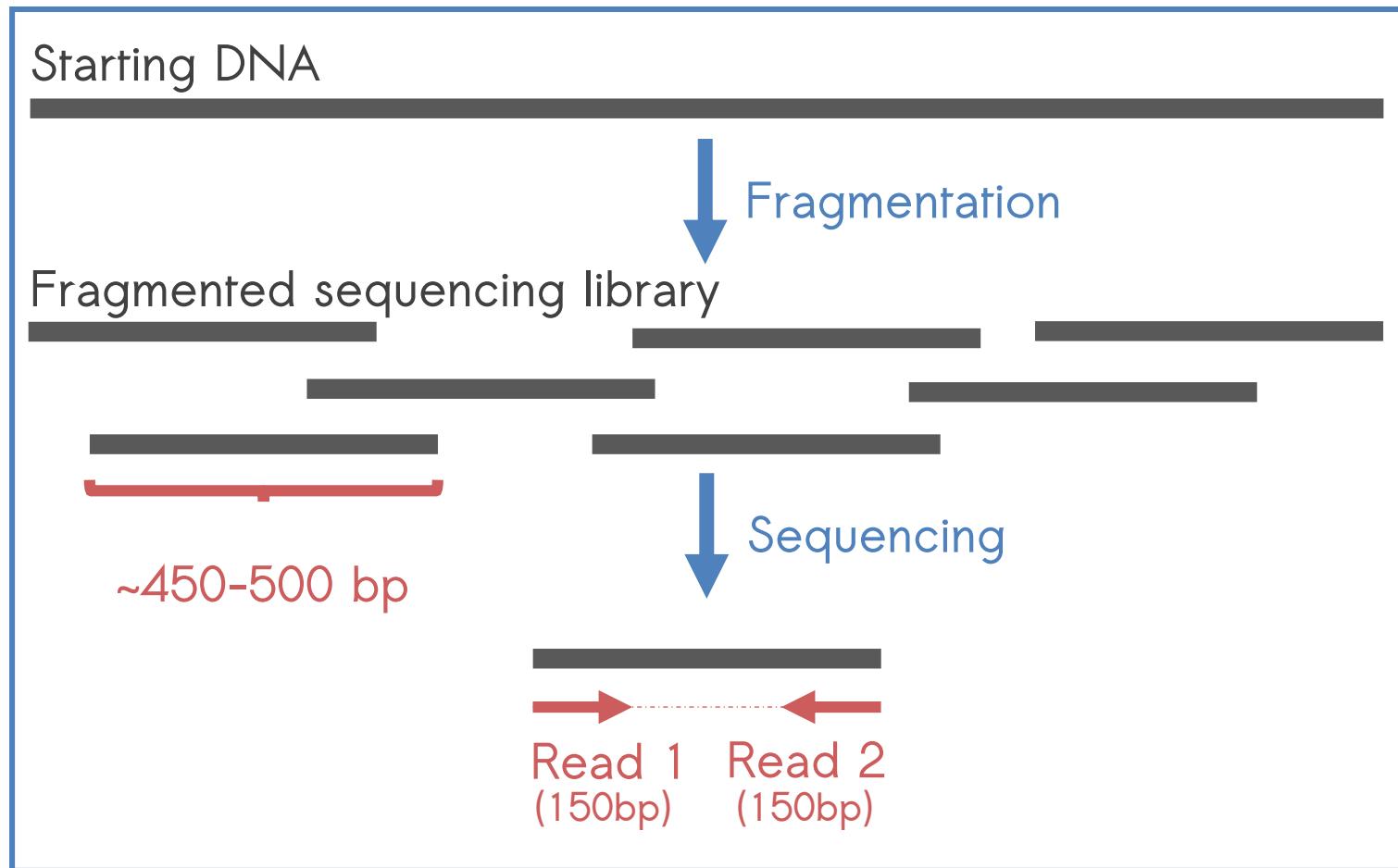
# Some questions in ALS genomics

## (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?

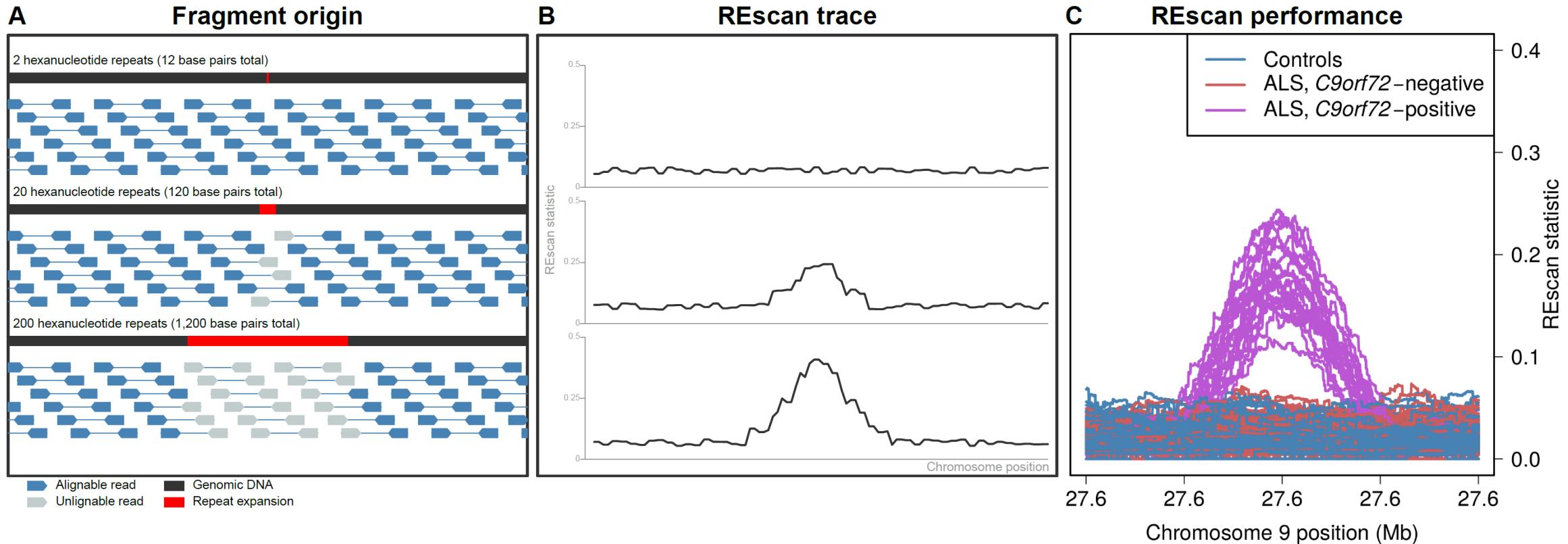
Paired-end next-generation sequencing





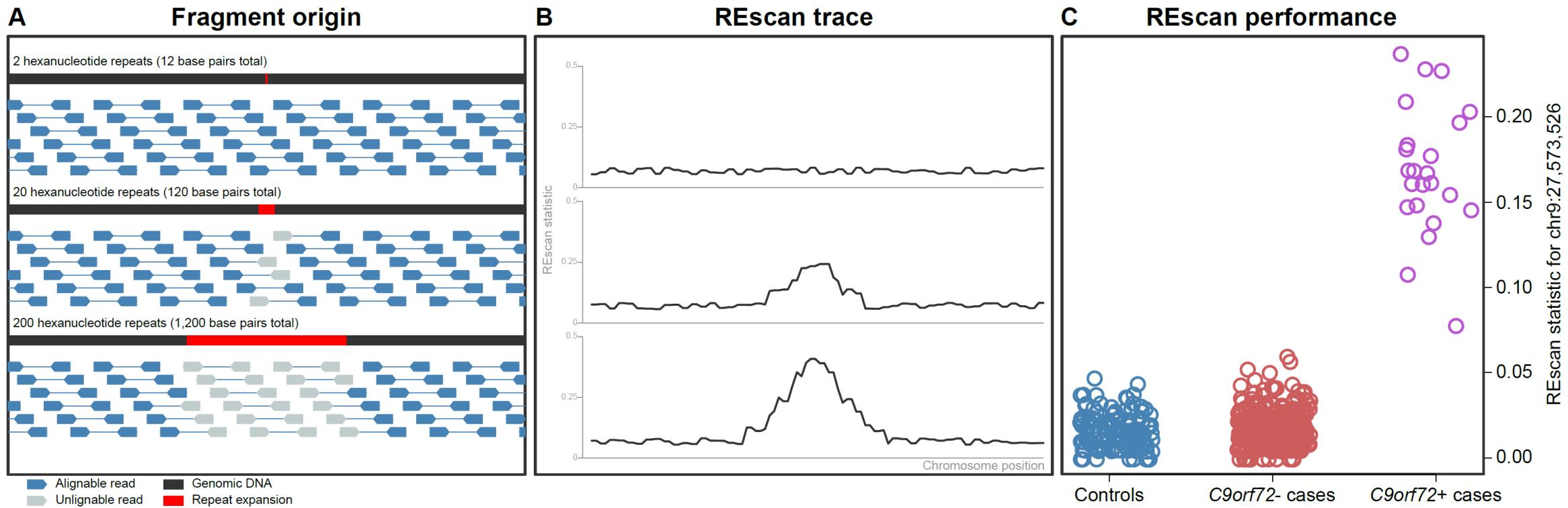
# Finding repeat expansions

Using paired-end next-generation sequencing data



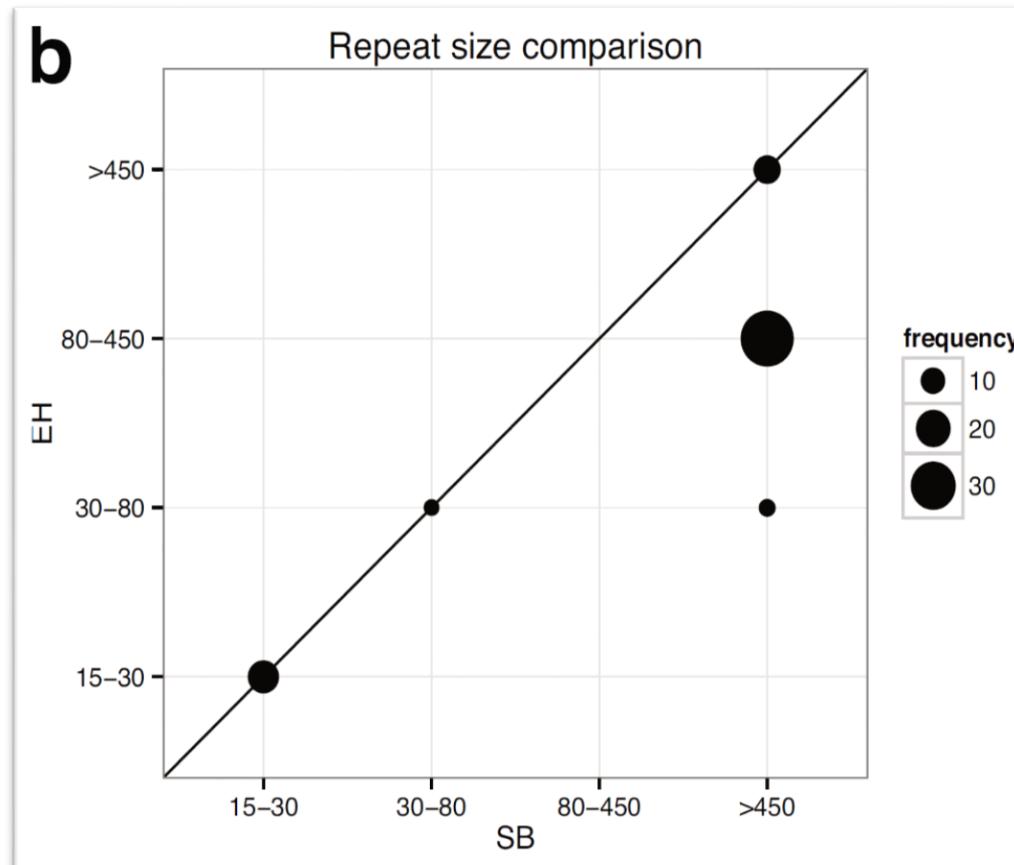
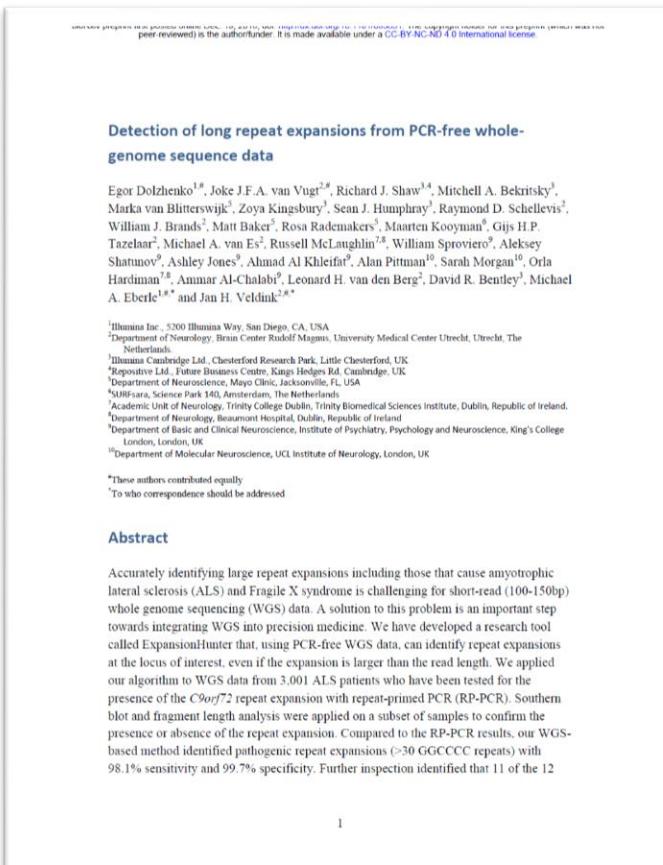
# Finding repeat expansions

Using paired-end next-generation sequencing data



# Finding repeat expansions

## ExpansionHunter



# Can we directly measure repeat expansions?

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



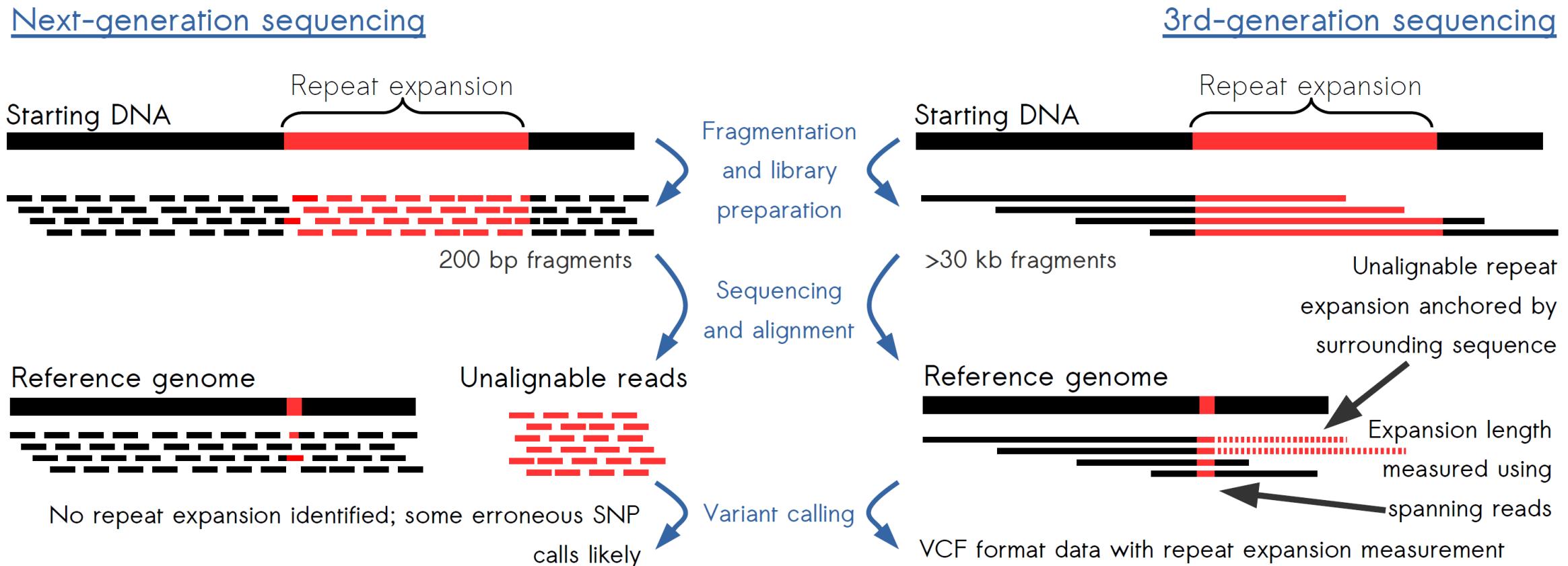
# Can we directly measure repeat expansions?

3<sup>rd</sup>-generation sequencing with ultra-long reads using Oxford Nanopore MinION



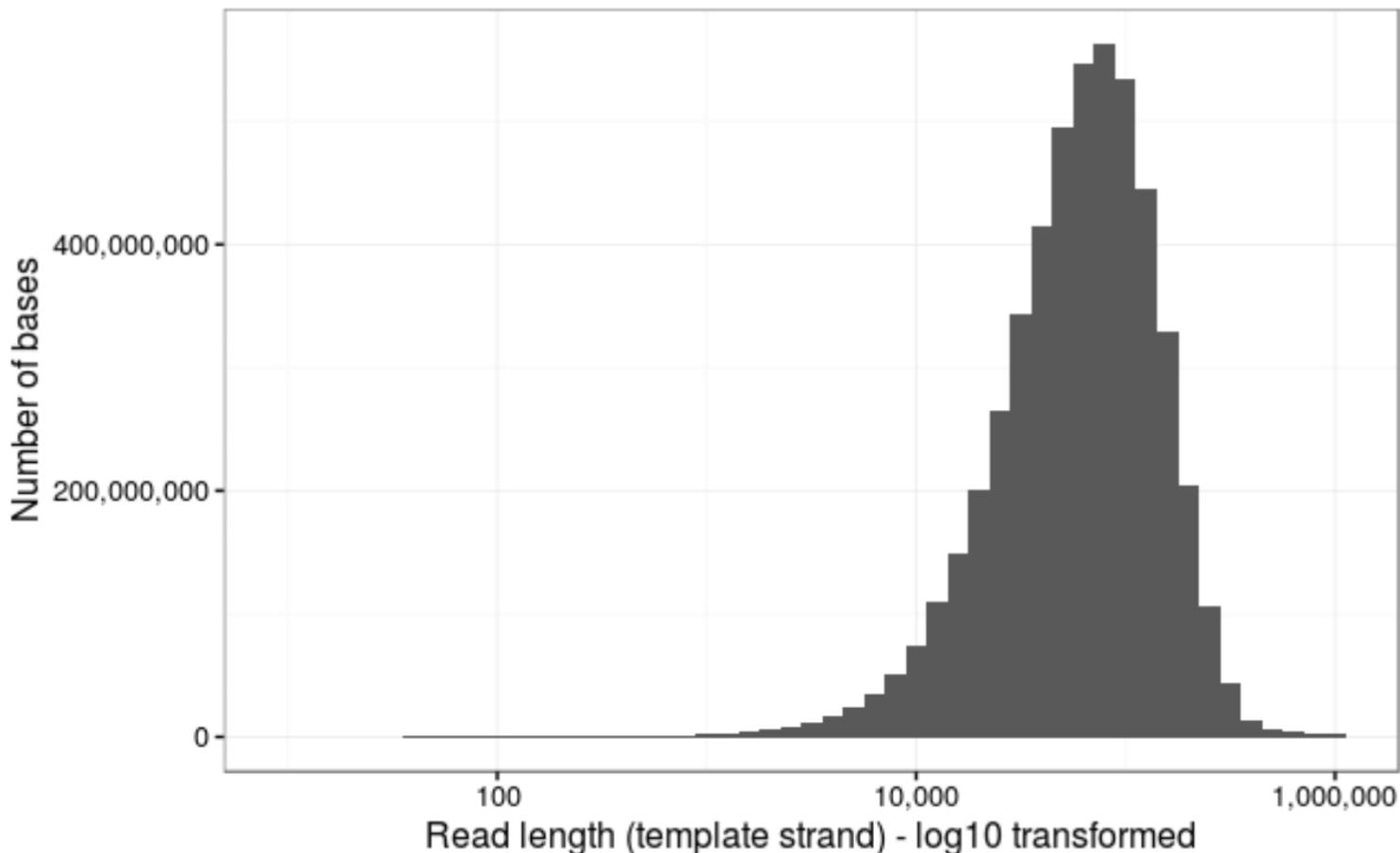
# NGS vs 3GS

## Spanning repeat expansions



# How long is a long read?

Answer: *very*

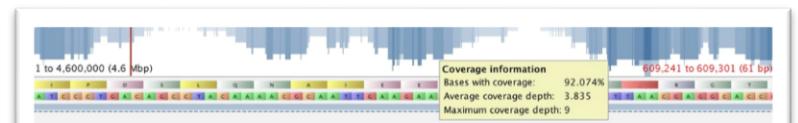


## Thar she blows! Ultra long read method for nanopore sequencing

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... \(gt350kb.fasta\)](https://github.com/nickloman/mass... (gt350kb.fasta))

9:30 PM - 2 Mar 2017

← ↗ 59 ❤ 109

# How long is a long read?

Answer: *very*



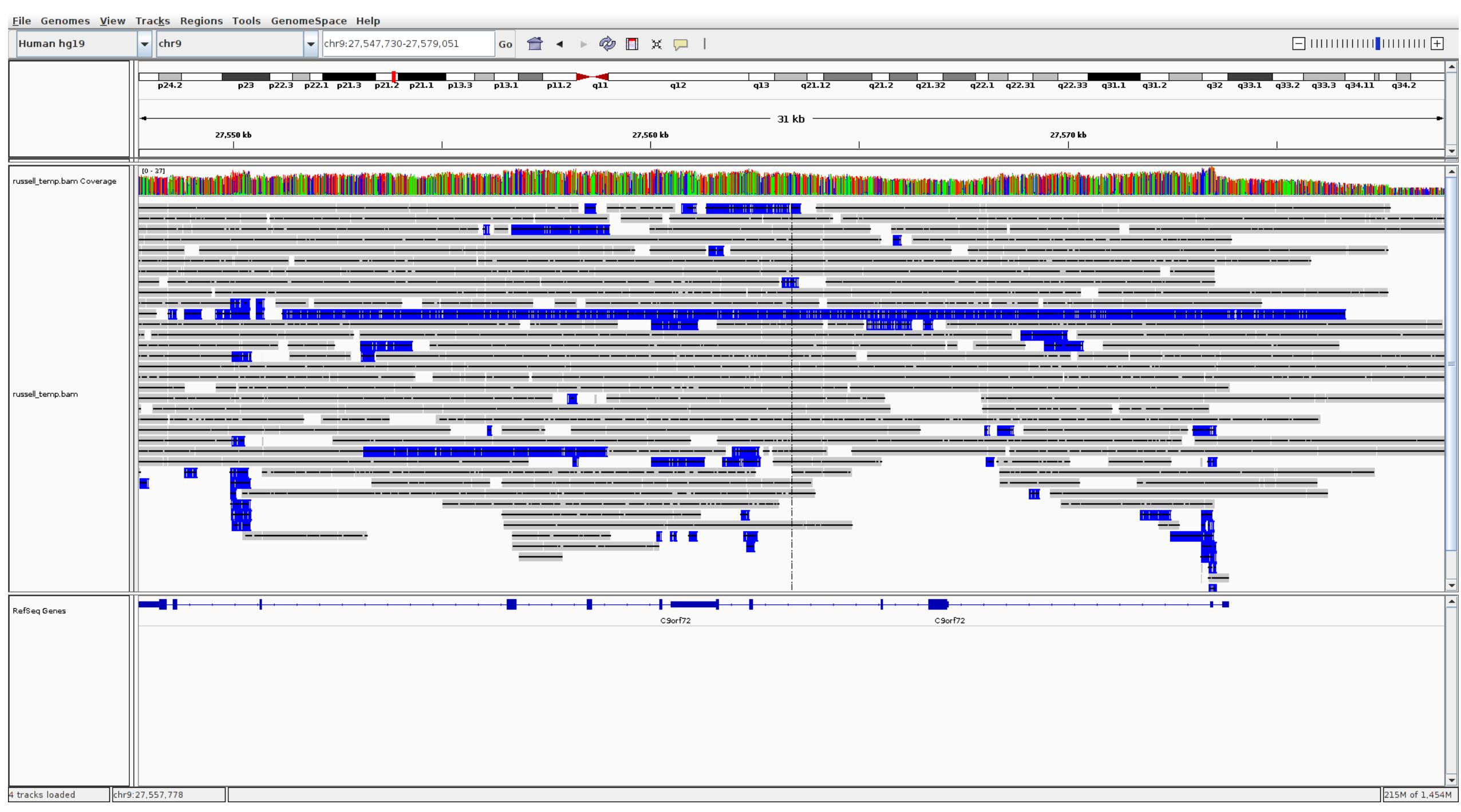
150 bp Illumina read  
(banana for scale)



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



**Trinity College Dublin**  
Coláiste na Tríonóide, Baile Átha Cliath  
The University of Dublin



mclaugh@tcd.ie

<http://bioinf.gen.tcd.ie/ctg>

@RSLMcL

rlmcl