The panorama of ALS genomics
ENCALS 2017 | Ljubljana | 19.05.2017
Russell McLaughlin | mclaughr@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

\[ \log i = (n - 1) \log t + \log \prod_{i=1}^{n} u_{i} \]

Age-specific incidence  Steps  Time (age)

Genetics of ALS

Genes that have been investigated in ALS

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


The panorama of ALS genomics
How do we discover new ALS genes?

**Genome-wide association study (GWAS)**

Person 1
... TCAGCCATGCTACT\textcolor{red}{C}GATCGACTAA\textcolor{green}{G}CG ... (maternal)
... TCAGCCATGCTACT\textcolor{red}{C}GATCGACTAATCG ... (paternal)

Person 2
... TCAGCCATGCTACT\textcolor{red}{C}GATCGACTAA\textcolor{green}{G}CG ... (maternal)
... TCAGCCATGCTACT\textcolor{red}{T}GATCGACTAATCG ... (paternal)

Person 3
... TCAGCCATGCTACT\textcolor{red}{T}GATCGACTAATCG ... (maternal)
... TCAGCCATGCTACT\textcolor{red}{T}GATCGACTAATCG ... (paternal)

If only present in one individual, or if rare
If rare and observed in disease context
If common in the population

<table>
<thead>
<tr>
<th>Variant or single nucleotide variant (SNV)</th>
<th>Mutation</th>
<th>Single nucleotide polymorphism (SNP)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SNP1</strong> Cases</td>
<td><strong>SNP2</strong> Cases</td>
<td><strong>SNP ...</strong> Repeat for all SNPs</td>
</tr>
<tr>
<td>Count of G: 2104 of 4000</td>
<td>Count of G: 1648 of 4000</td>
<td></td>
</tr>
<tr>
<td>Frequency of G: 52.6%</td>
<td>Frequency of G: 41.2%</td>
<td></td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td><strong>Controls</strong></td>
<td></td>
</tr>
<tr>
<td>Count of G: 2676 of 6000</td>
<td>Count of G: 2532 of 6000</td>
<td></td>
</tr>
<tr>
<td>Frequency of G: 44.6%</td>
<td>Frequency of G: 42.2%</td>
<td></td>
</tr>
<tr>
<td>\textbf{P-value:} (5.0 \cdot 10^{-15})</td>
<td>\textbf{P-value:} 0.33</td>
<td></td>
</tr>
</tbody>
</table>
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)

Graham et al. (1997)
Al-Chalabi et al. (2010)
Wingo et al. (2011)
Fogh et al. (2014)
Keller et al. (2014)
van Rheenen et al. (2016)
van Rheenen et al. (2016)
McLaughlin et al. (2017)

Twin data
Trio data
Genetic data (REML)
Genetic data (other)
Genetic correlation

Covariance between two traits due to shared genetic variation

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

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

*Statistically 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

McLaughlin, Schijven et al. (2017) Nat Commun 8:14774
ALS and schizophrenia are genetically correlated

Implications

Individuals who develop both ALS and schizophrenia

Null expectation
1 in 40,000

With observed genetic correlation
1 in 34,337

ALS register size required for 80% power to detect this difference: 16,448

Odds ratio for developing ALS given schizophrenia (or vice versa): 1.17
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
3. GWAS has helped us to better understand the genetic architecture of ALS.
Finding rare variants

Exome sequencing

<table>
<thead>
<tr>
<th>Previous approach: SNP chips (GWAS)</th>
<th>New approach: exome sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>exon</td>
<td>exon</td>
</tr>
<tr>
<td>intron</td>
<td>intron</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>CTGCTAGCTAGTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGGCCTAGTCGATGAGTCAGCCGTAG</td>
<td>CTGCTAGCTAGTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCAGCCGTAG</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>C  T</td>
<td>T C</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

< gene exons/introns
< reference genome
< paternal chromosome
< maternal chromosome
Implicating rare variants

Assessing significance by burden testing
Implicating rare variants

Exome sequencing with pedigrees

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
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)

CTGCTAGCTAGCTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCGCTATCGATGAGTCAGCCGTAG

<table>
<thead>
<tr>
<th>C</th>
<th>A</th>
</tr>
</thead>
</table>

New approach: exome sequencing

CTGCTAGCTAGCTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCGCTATCGATGAGTCAGCCGTAG

<table>
<thead>
<tr>
<th>T</th>
<th>C</th>
</tr>
</thead>
</table>

Newer approach: whole-genome sequencing

CTGCTAGCTAGCTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCGCTATCGATGAGTCAGCCGTAG

<table>
<thead>
<tr>
<th>T</th>
<th>A</th>
</tr>
</thead>
</table>

CTGCTAGCTAGCTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCGCTATCGATGAGTCAGCCGTAG

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
</tr>
</thead>
</table>

CTGCTAGCTAGCTCTATCGTGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCGCTATCGATGAGTCAGCCGTAG

<table>
<thead>
<tr>
<th>C</th>
<th>A</th>
</tr>
</thead>
</table>
Whole-genome sequencing in large populations

Project MinE

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.

Make a donation today

100 percent of all donations to Project Mine will go

8,335.94 / 22,500.00 DNA profiles collected

Learn more

Make it yours.

Donate now
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)
...GACCACGCCCGGCCCCCGGCCCCGCCCCCTAGCGCGCGACT...

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

C9orf72-positive ALS
...GACCACGCCCGGCCCC\((G_2C_4)^n\)GGCCCCCTAGCGCGCGACT...

(where n = potentially >1,000)

Table 1 Repeat expansions that cause neurodegeneration

<table>
<thead>
<tr>
<th>Gene</th>
<th>Disease</th>
<th>Repeat motif</th>
<th>Non/ codig</th>
<th>Pathogenic range</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFF2/FMR3</td>
<td>FRAXE mental retardation syndrome</td>
<td>CCG</td>
<td>Noncoding</td>
<td>&gt;200</td>
</tr>
<tr>
<td>AR</td>
<td>Spinal and bulbar muscular atrophy</td>
<td>CAG</td>
<td>Coding</td>
<td>40-62</td>
</tr>
<tr>
<td>ARX</td>
<td>X-linked mental retardation</td>
<td>GCG</td>
<td>Coding</td>
<td>17-23</td>
</tr>
<tr>
<td>ATN1</td>
<td>Dentatorubral-pallidolysian atrophy</td>
<td>CAG</td>
<td>Coding</td>
<td>40-88</td>
</tr>
<tr>
<td>ATXN1</td>
<td>Spinocerebellar ataxia type 1</td>
<td>CAG</td>
<td>Coding</td>
<td>39-83</td>
</tr>
<tr>
<td>ATXN10</td>
<td>Spinocerebellar ataxia type 10</td>
<td>ATCT</td>
<td>Noncoding</td>
<td>280-4500</td>
</tr>
<tr>
<td>ATXN2</td>
<td>Spinocerebellar ataxia type 2</td>
<td>CAG</td>
<td>Coding</td>
<td>34-59</td>
</tr>
<tr>
<td>ATXN4</td>
<td>Amyotrophic lateral sclerosis</td>
<td>CAG</td>
<td>Coding</td>
<td>27-33</td>
</tr>
<tr>
<td>ATXN3</td>
<td>Spinocerebellar ataxia type 3</td>
<td>CAG</td>
<td>Coding</td>
<td>55-84</td>
</tr>
<tr>
<td>ATXN7</td>
<td>Spinocerebellar ataxia type 7</td>
<td>CAG</td>
<td>Coding</td>
<td>34-300</td>
</tr>
<tr>
<td>ATXN8</td>
<td>Spinocerebellar ataxia type 8</td>
<td>CAG/CTG</td>
<td>Both</td>
<td>80-1300</td>
</tr>
<tr>
<td>C9orf72</td>
<td>ALS/FTD</td>
<td>GGGGCC</td>
<td>Noncoding</td>
<td>&gt;30</td>
</tr>
<tr>
<td>CACNA1A</td>
<td>Spinocerebellar ataxia type 6</td>
<td>GGGGCC</td>
<td>Coding</td>
<td>21-30</td>
</tr>
<tr>
<td>CNBP</td>
<td>Myotonic dystrophy type 2</td>
<td>CCTG</td>
<td>Noncoding</td>
<td>75-11000</td>
</tr>
<tr>
<td>CSTB</td>
<td>Epilepsy progressive myoclonia (C)_G(C)_G</td>
<td>Noncoding</td>
<td>30-75</td>
<td></td>
</tr>
<tr>
<td>DIP2B</td>
<td>PRA12A mental retardation syndrome</td>
<td>CCG</td>
<td>Noncoding</td>
<td>&gt;23</td>
</tr>
<tr>
<td>DMPK</td>
<td>Myotonic dystrophy type 1</td>
<td>CTG</td>
<td>Noncoding</td>
<td>50-6500</td>
</tr>
<tr>
<td>FMR1</td>
<td>Fragile X mental retardation type 1</td>
<td>CCG</td>
<td>Noncoding</td>
<td>&gt;200</td>
</tr>
<tr>
<td>FMR2</td>
<td>Fragile X-associated tremor ataxia syndrome</td>
<td>GGG</td>
<td>Noncoding</td>
<td>55-200</td>
</tr>
<tr>
<td>FKX1</td>
<td>Friedreich's ataxia</td>
<td>GAA</td>
<td>Noncoding</td>
<td>66-1700</td>
</tr>
<tr>
<td>HTT</td>
<td>Huntington's disease</td>
<td>CAG</td>
<td>Coding</td>
<td>&gt;35</td>
</tr>
<tr>
<td>JPH3</td>
<td>Huntington's disease-like 2</td>
<td>CAG/CTG</td>
<td>Noncoding</td>
<td>&gt;41</td>
</tr>
<tr>
<td>NOP56</td>
<td>Spinocerebellar ataxia type 36</td>
<td>GGCTG</td>
<td>Noncoding</td>
<td>1500-2500</td>
</tr>
<tr>
<td>PABPN1</td>
<td>Oculopharyngeal muscular dystrophy</td>
<td>GCG</td>
<td>Coding</td>
<td>11-17</td>
</tr>
<tr>
<td>PMP22R2</td>
<td>Spinocerebellar ataxia type 12</td>
<td>CAG/CTG</td>
<td>Noncoding</td>
<td>55-78</td>
</tr>
<tr>
<td>TBP</td>
<td>Spinocerebellar ataxia type 17</td>
<td>CAG</td>
<td>Coding</td>
<td>49-66</td>
</tr>
<tr>
<td>T252-HEAN</td>
<td>Spinocerebellar ataxia type 31</td>
<td>TGGA</td>
<td>Noncoding</td>
<td>500-760</td>
</tr>
</tbody>
</table>

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

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
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?

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

19.05.2017

The panorama of ALS genomics

28
Repeat expansions are hard to sequence

Next-generation sequencing: a simplified overview

Starting DNA

Repeat expansion

Fragmentation and library preparation

200 bp fragments

Sequencing and alignment

Reference genome

Unalignable reads

No repeat expansion identified; some erroneous SNP calls likely

Variant calling
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?
   a) Same disease mechanism (e.g. RAN translation) on different transcripts

4. Rare implies selectively disadvantageous. How can this happen in ALS (a late onset disease)?
   a) Developmental dynamics (e.g. 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
**C9orf72 repeat expansion**

**Mechanisms of 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

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

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 (e.g., RAN translation) on different transcripts

4. Rare implies selectively disadvantageous. How can this happen in ALS (a late onset disease)?
   a) Developmental dynamics e.g., 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

- Starting DNA
- Fragmentation
- Fragmented sequencing library
- Sequencing
- ~450-500 bp
- Read 1 (150bp)
- Read 2 (150bp)
How can we find (novel) repeat expansions?

Paired-end next-generation sequencing

Starting DNA Fragmentation

Sequencing

Read 1

Read 2

Fragmented sequencing library

~450 - 500 bp

(150bp)

(150bp)
Finding repeat expansions

Using paired-end next-generation sequencing data

---

A. Fragment origin

- 2 hexanucleotide repeats (12 base pairs total)
- 20 hexanucleotide repeats (120 base pairs total)
- 200 hexanucleotide repeats (1,200 base pairs total)

B. REscan trace

C. REscan performance

- Controls
- ALS, C9orf72-negative
- ALS, C9orf72-positive
# Finding repeat expansions

Using paired-end next-generation sequencing data

<table>
<thead>
<tr>
<th>A</th>
<th>Fragment origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 hexanucleotide repeats (12 base pairs total)</td>
<td></td>
</tr>
<tr>
<td>20 hexanucleotide repeats (120 base pairs total)</td>
<td></td>
</tr>
<tr>
<td>200 hexanucleotide repeats (1,200 base pairs total)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>REscan trace</th>
</tr>
</thead>
<tbody>
<tr>
<td>REscan statistic</td>
<td>Chromosome position</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>REscan performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>REscan statistic</td>
<td>Controls</td>
</tr>
</tbody>
</table>

19.05.2017
Finding repeat expansions

ExpansionHunter
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**

- Starting DNA
- Repeat expansion
- 200 bp fragments
- Unalignable reads
- No repeat expansion identified; some erroneous SNP calls likely
- Variant calling
- Fragmentation and library preparation
- Sequencing and alignment
- VCF format data with repeat expansion measurement

**3rd-generation sequencing**

- Starting DNA
- Repeat expansion
- >30 kb fragments
- Unalignable repeat expansion anchored by surrounding sequence
- Expansion length measured using spanning reads

19.05.2017

The panorama of ALS genomics
How long is a long read?

Answer: very
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
The panorama of ALS genomics
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