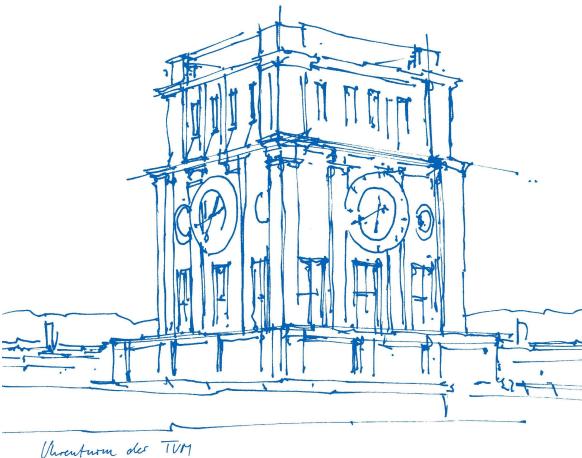


Genetic diagnosis of Mendelian disorders via RNA sequencing

Daniel M Bader

Gagneur lab - Computational biology
Twitter @gagneurlab

May 29, 2017



I have no commercial disclosure.

Limitations of genome sequencing

- Exome sequencing
 - ~2% of genome covered
 - ~50% patients not diagnosed
 - Genome sequencing
 - detection of all variants
 - difficult prioritization and interpretation
 - Many variants of unknown significance
→ synonymous or non-coding
 - Knowledge gap between coding and regulatory sequence
- Chong 2015 AJHG review
Wortmann 2015 J Inherit Metab Dis
Retterer 2016 Genetics in Medicine
- Taylor 2015 Nature genetics

Limitations of genome sequencing

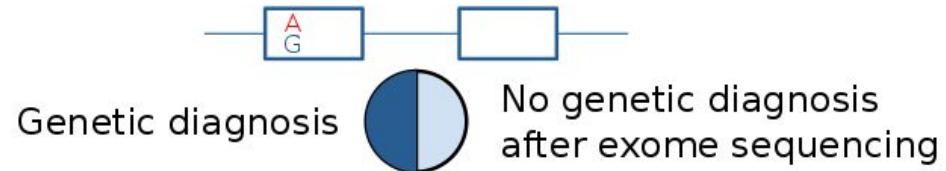
- Exome sequencing
 - ~2% of genome covered
 - ~50% patients not diagnosed
 - Genome sequencing
 - detection of all variants
 - difficult prioritization and interpretation
 - Many variants of unknown significance
→ synonymous or non-coding
 - Knowledge gap between coding and regulatory sequence
- Chong 2015 AJHG review
Wortmann 2015 J Inherit Metab Dis
Retterer 2016 Genetics in Medicine
- Taylor 2015 Nature genetics

⇒ RNA sequencing!

Using the power of RNA-seq to investigate Mendelian disorders of mitochondria

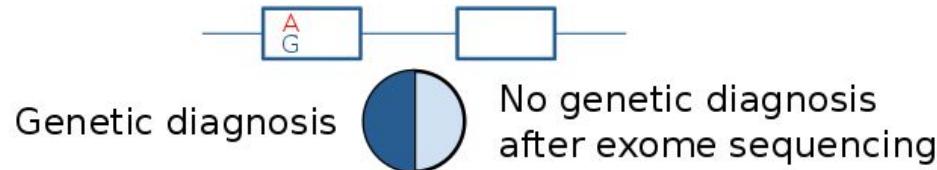
- Birth prevalence **2 in 10,000**
- Large genetic basis with causative defects identified in more than **250 genes**
 - mtDNA
 - Nuclear DNA
- Test for **respiratory activity** and its rescue using fibroblast cell lines

Gorman 2016 Nat Rev Dis Primers
Mayr 2015 J Inherit Metab Dis



1. Patient fibroblasts (n=105)



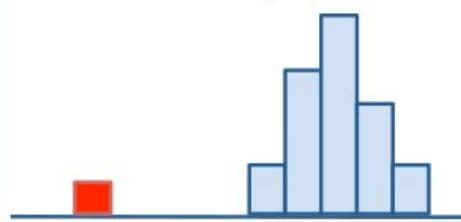


1. Patient fibroblasts (n=105)



2. RNA sequencing

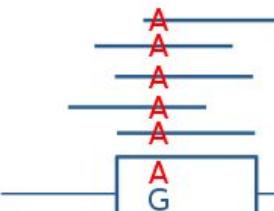
Aberrant expression

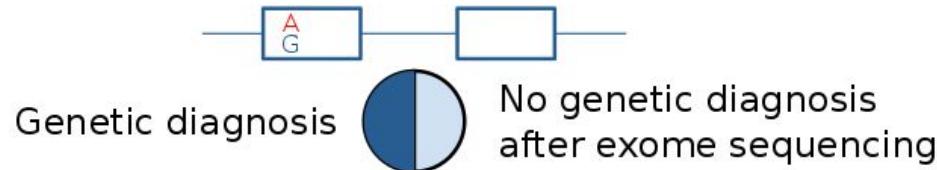


Aberrant splicing



Mono-allelic expression



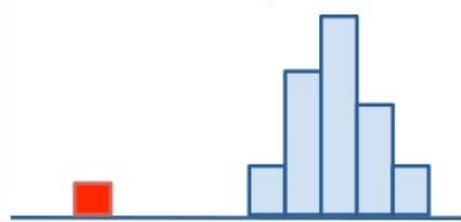


1. Patient fibroblasts (n=105)



2. RNA sequencing

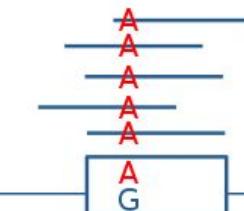
Aberrant expression



Aberrant splicing

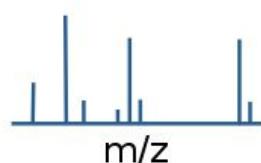


Mono-allelic expression

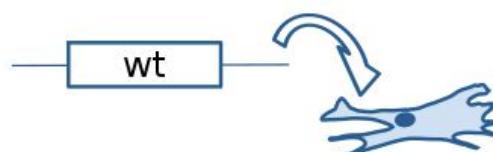


3. Functional and biochemical validation

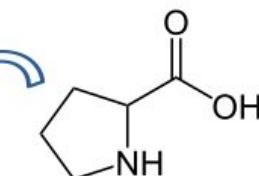
Proteomics



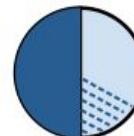
Complementation



Supplementation



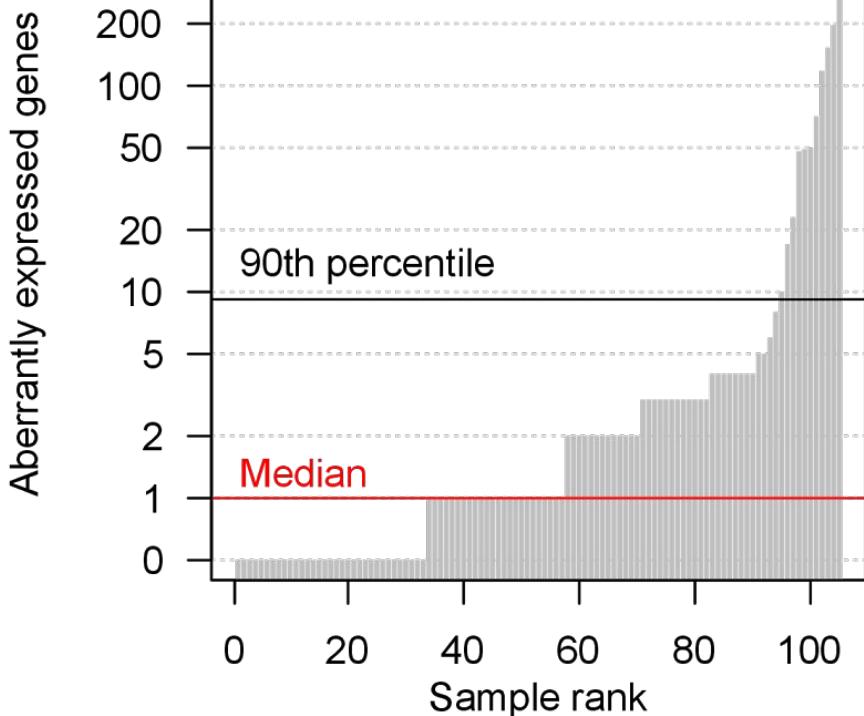
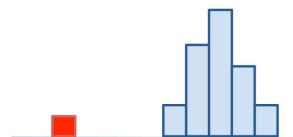
Genetic diagnosis



No genetic diagnosis

New genetic diagnosis

Detection of aberrant expression

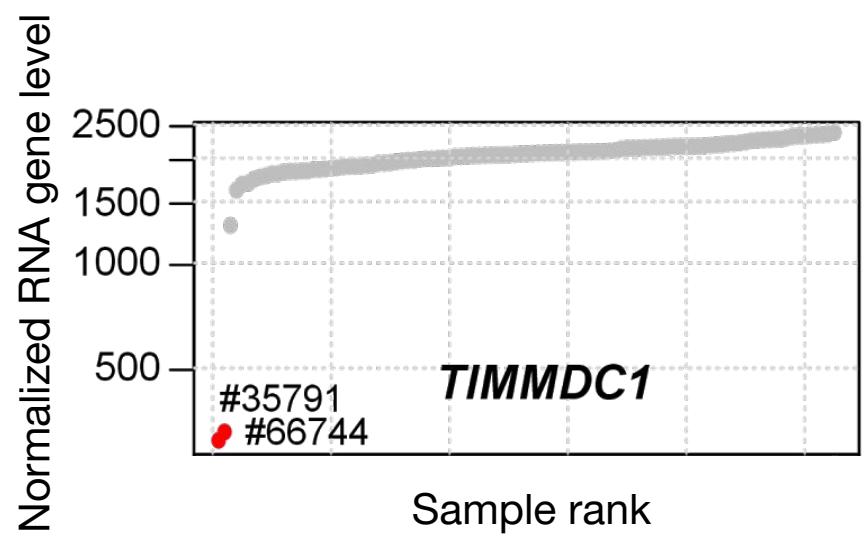
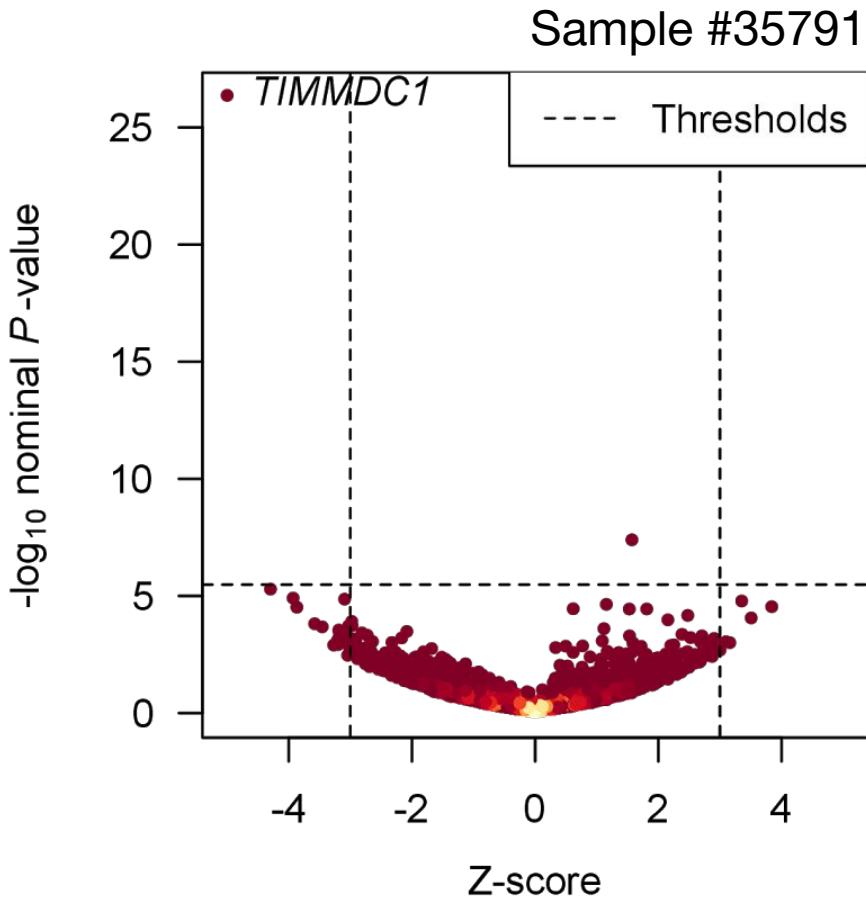
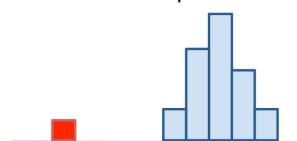


- Gene-wise read counts
- Normalize for sample effects
- Test 1 versus rest
- Adjusted P-value < 0.05 & $|Z\text{-score}| > 3$

Z-score =
difference to mean /
standard deviation

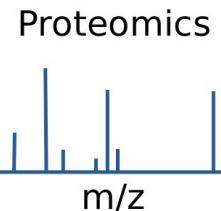
TIMMDC1 as example for aberrant expression

Aberrant expression

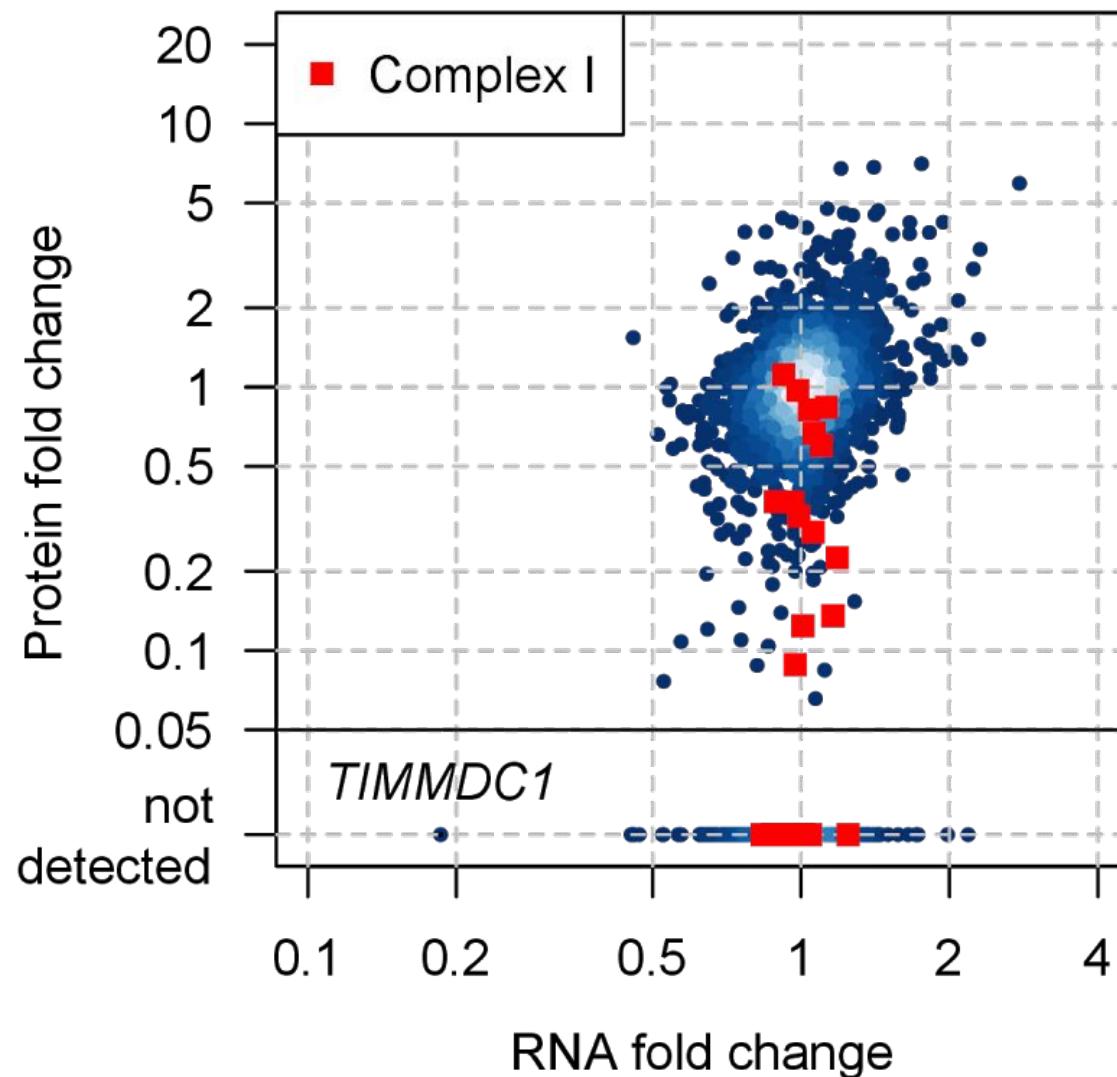


TIMMDC1: Translocase of Inner Mitochondrial Membrane Domain Containing 1;
Complex I assembly factor

Proteomic validation

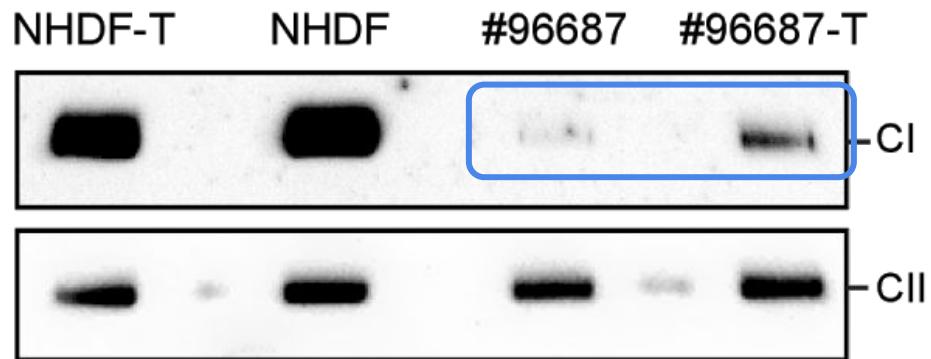
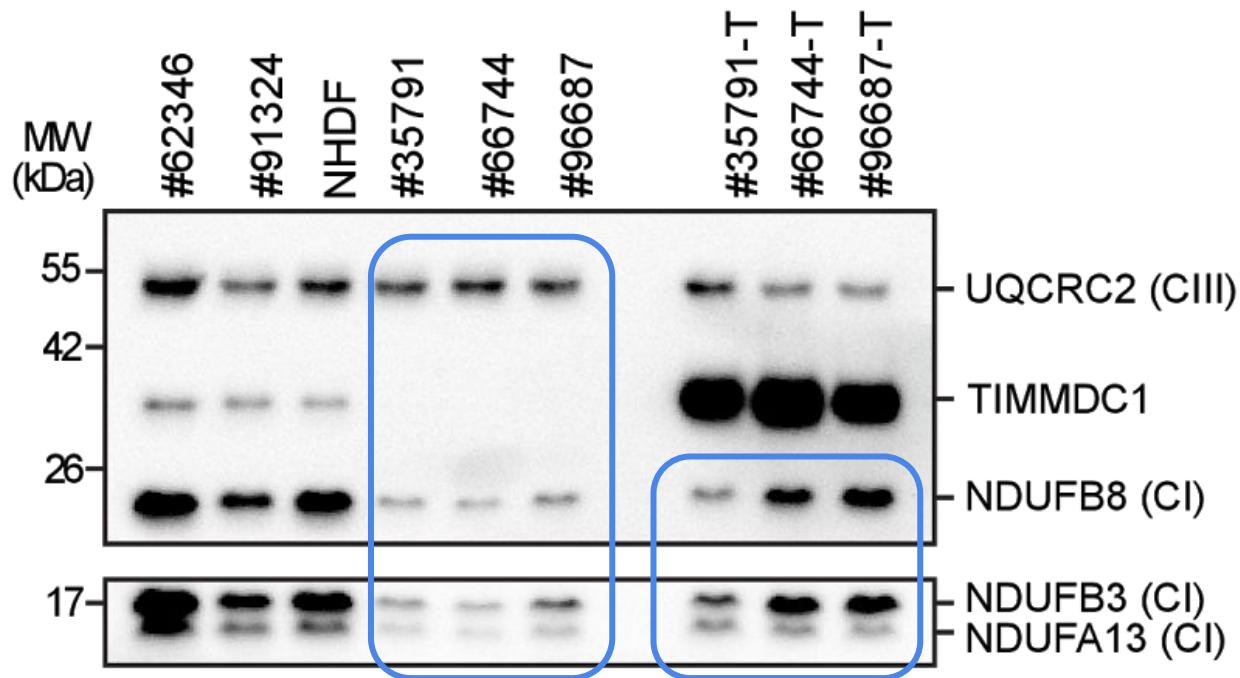
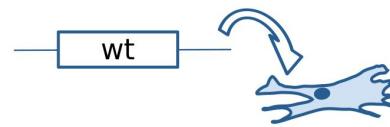


Sample #35791



Rescue complex I assembly by wildtype complementation

Complementation

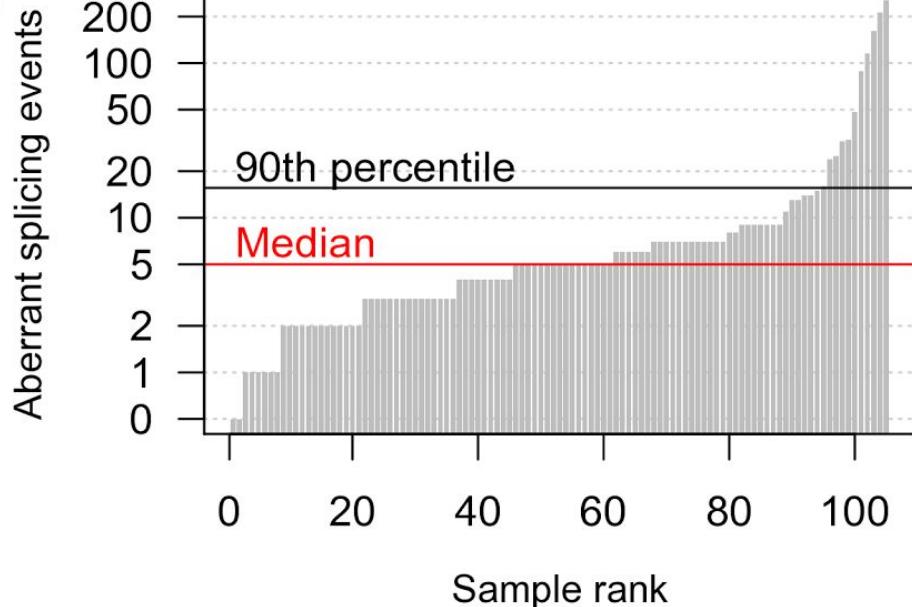


Aberrant splicing

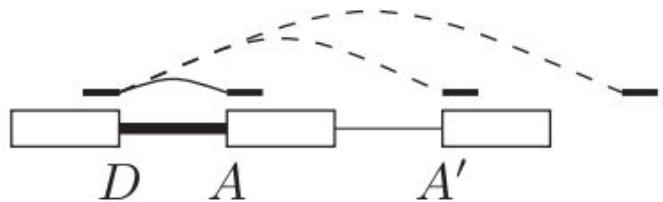


Aberrant splicing

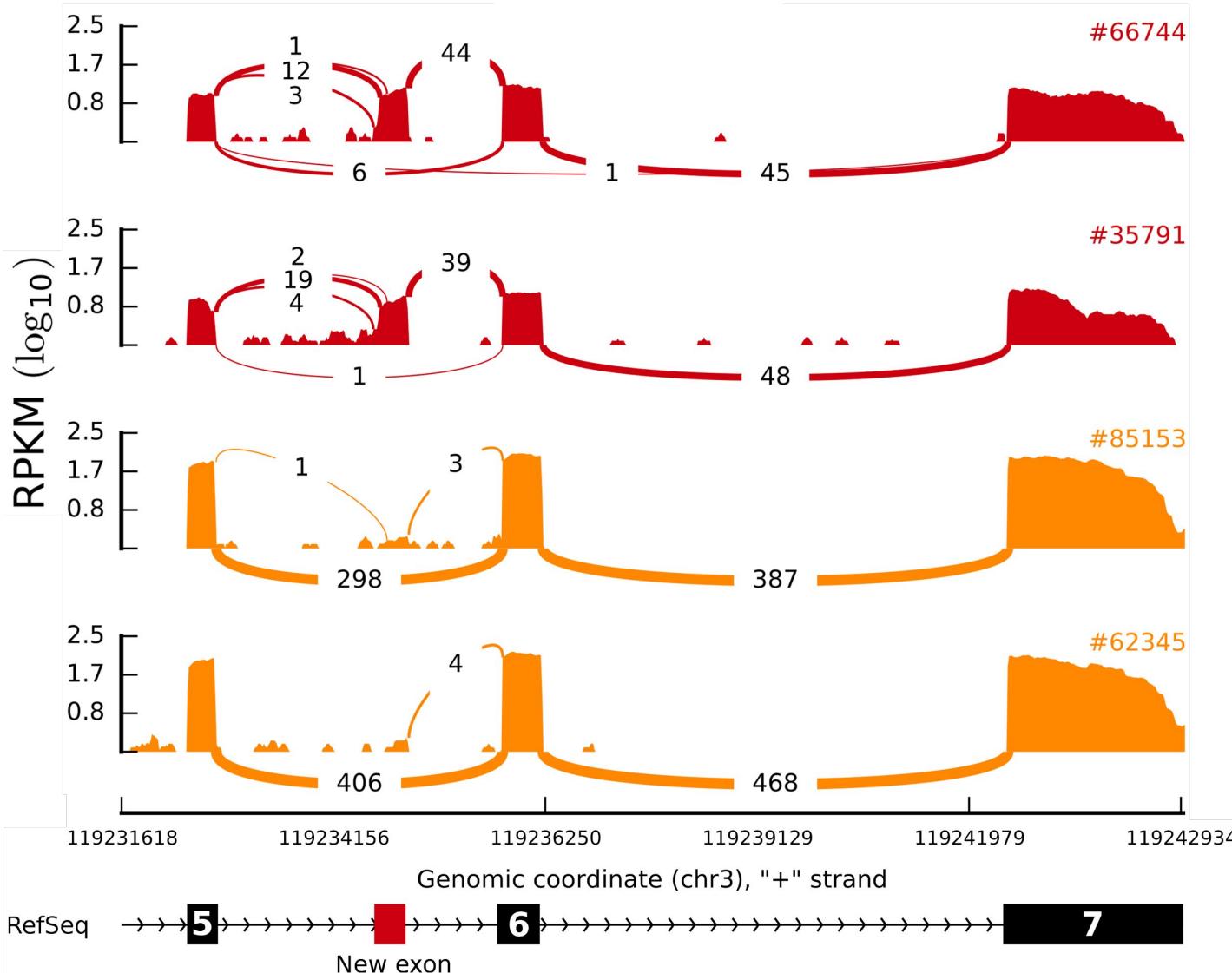
Detection of aberrant splicing



- Based on Leafcutter
 - Exon-junction read counts
 - Statistics on percent spliced in (PSI Ψ) values
- Method adapted for testing 1 vs rest by pseudo counts
- Adjusted P-value < 0.05



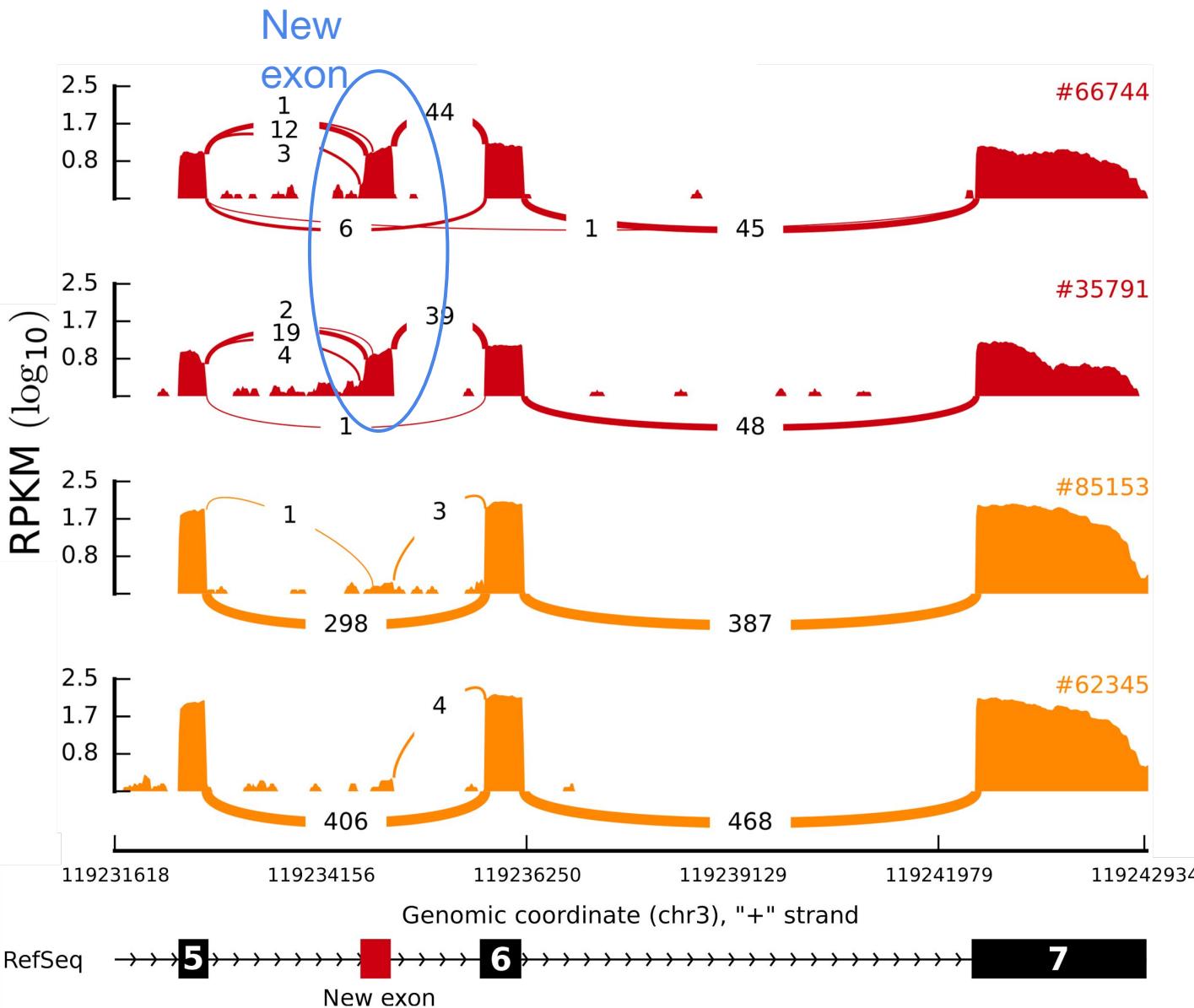
TIMMDC1 - new exon creation



Samples with
TIMMDC1
defect

Other
samples

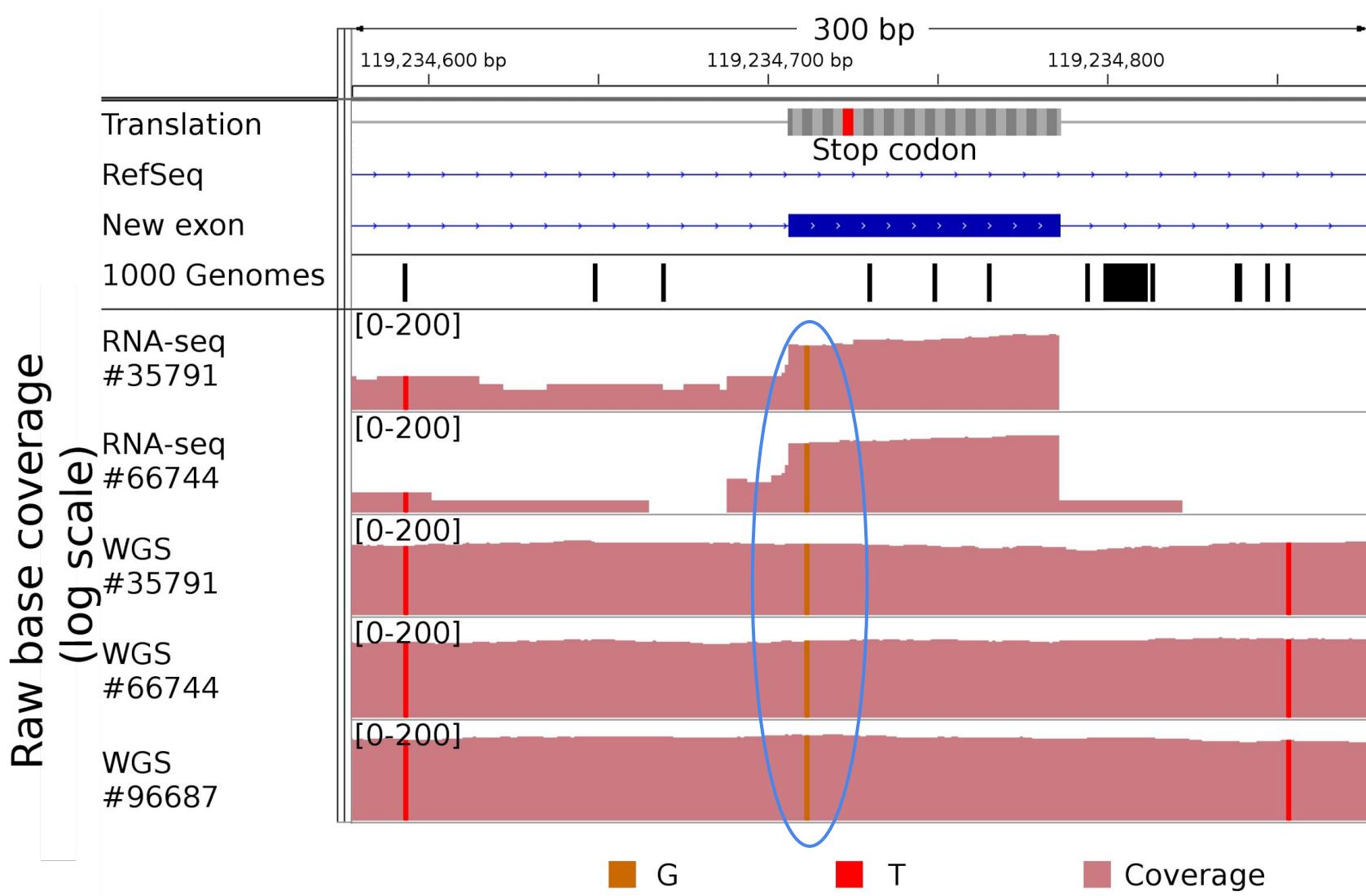
TIMMD1C - new exon creation



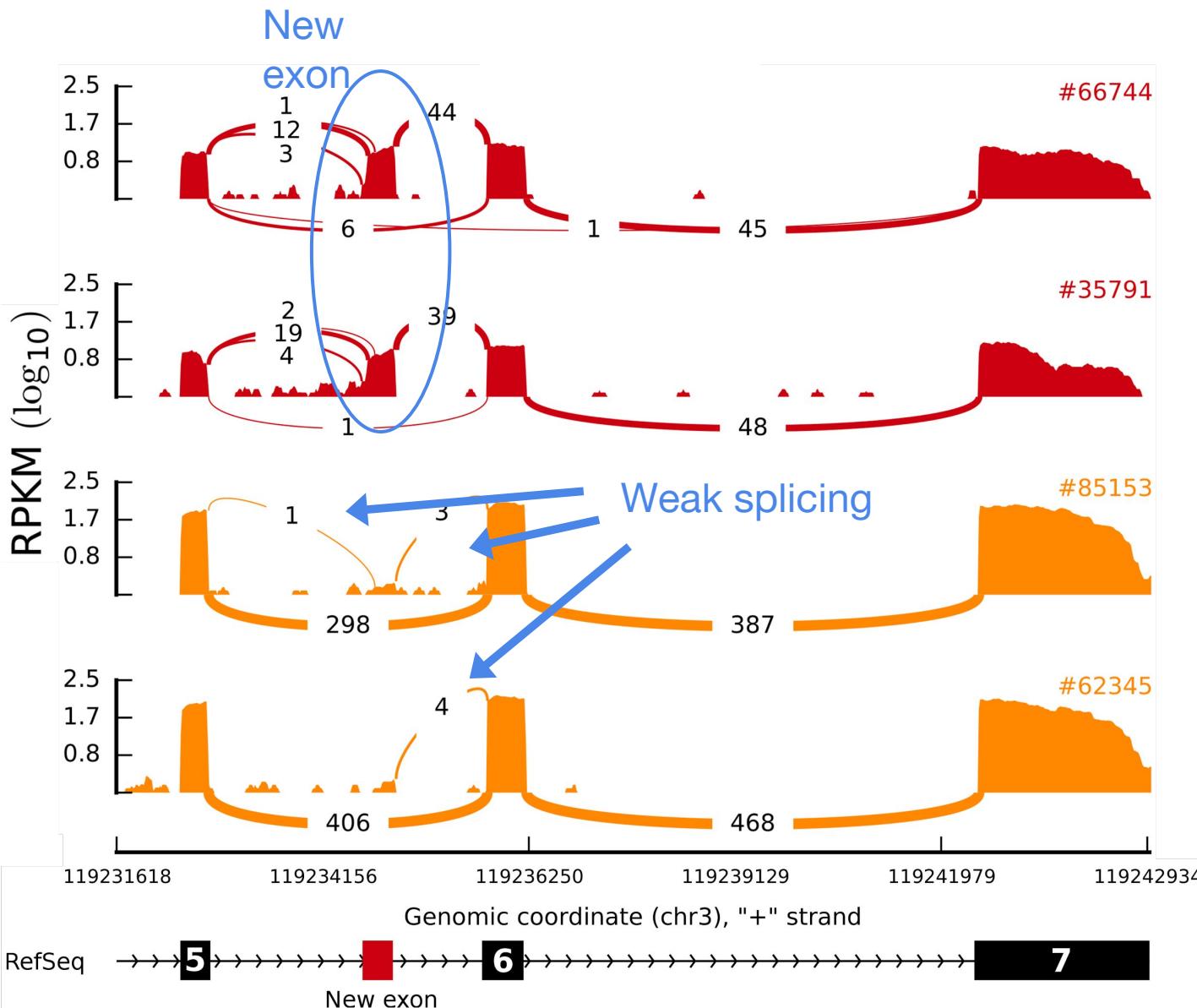
Samples with
TIMMD1
defect

Other
samples

Rare deep intronic mutation inside the new exon



TIMMD1C - new exon creation

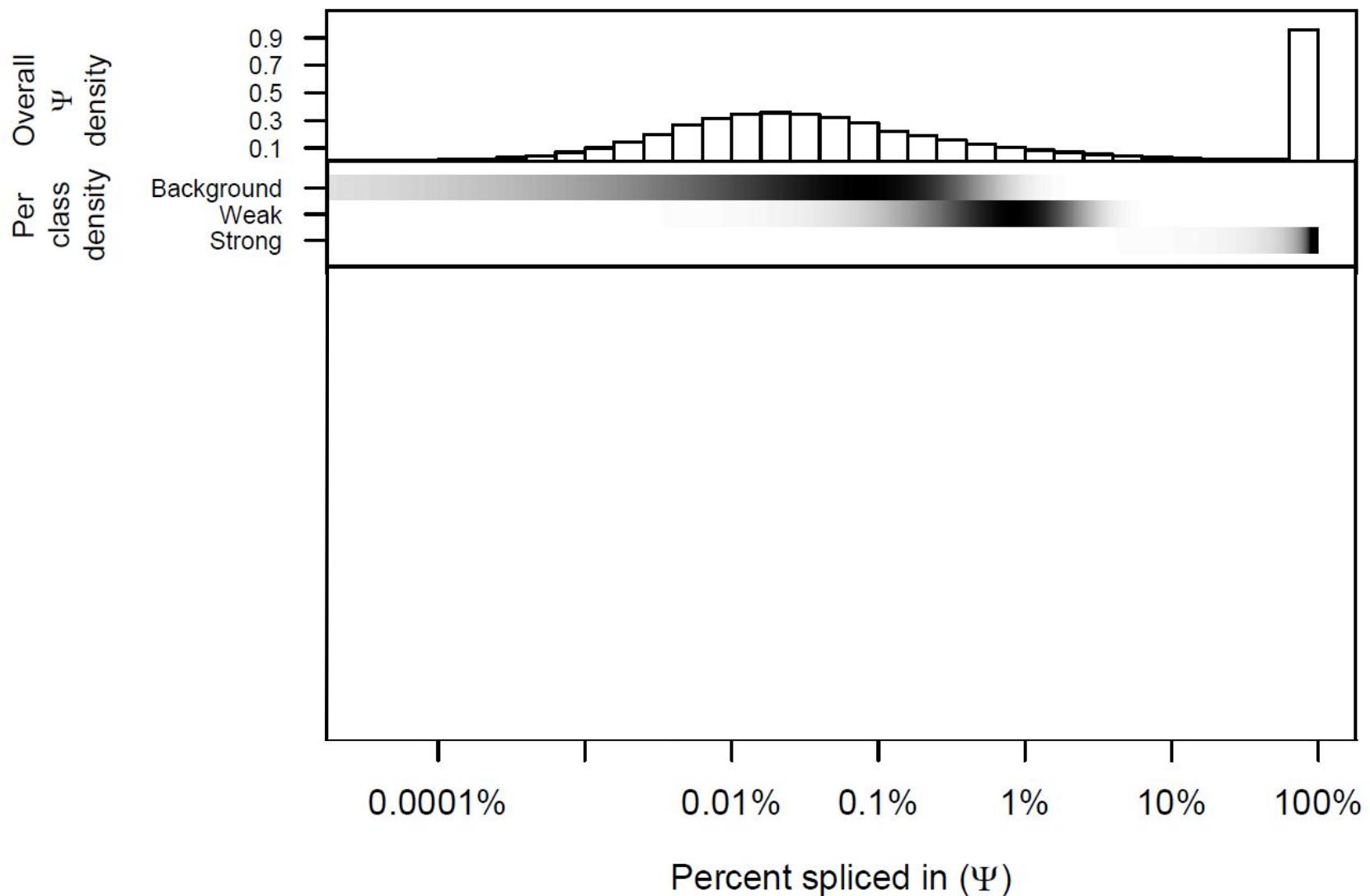


Samples with
TIMMD1
defect

Other
samples

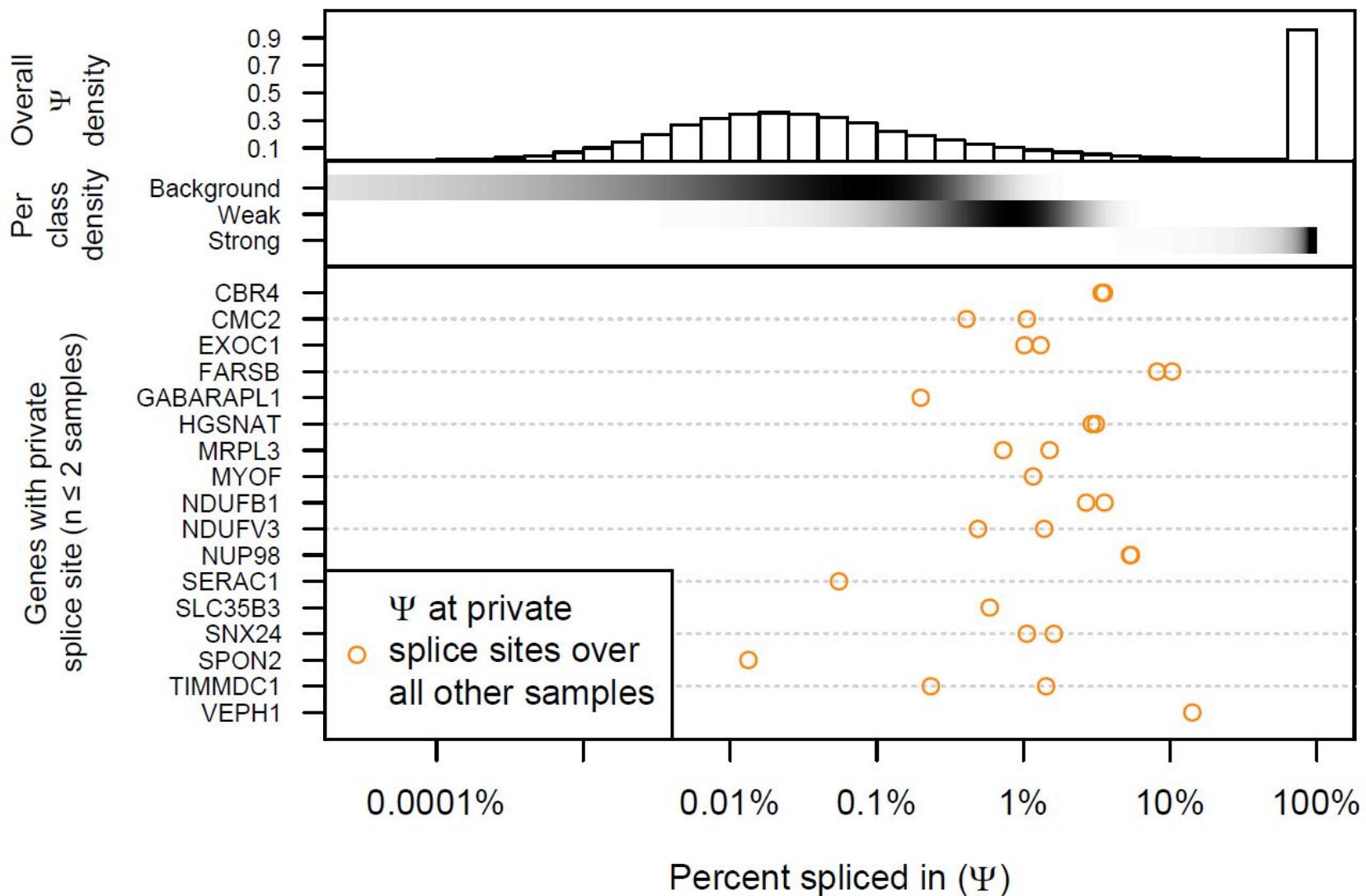
Weak splice sites as origin for new exons

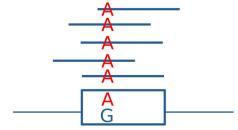
Aberrant splicing



Weak splice sites as origin for new exons

Aberrant splicing

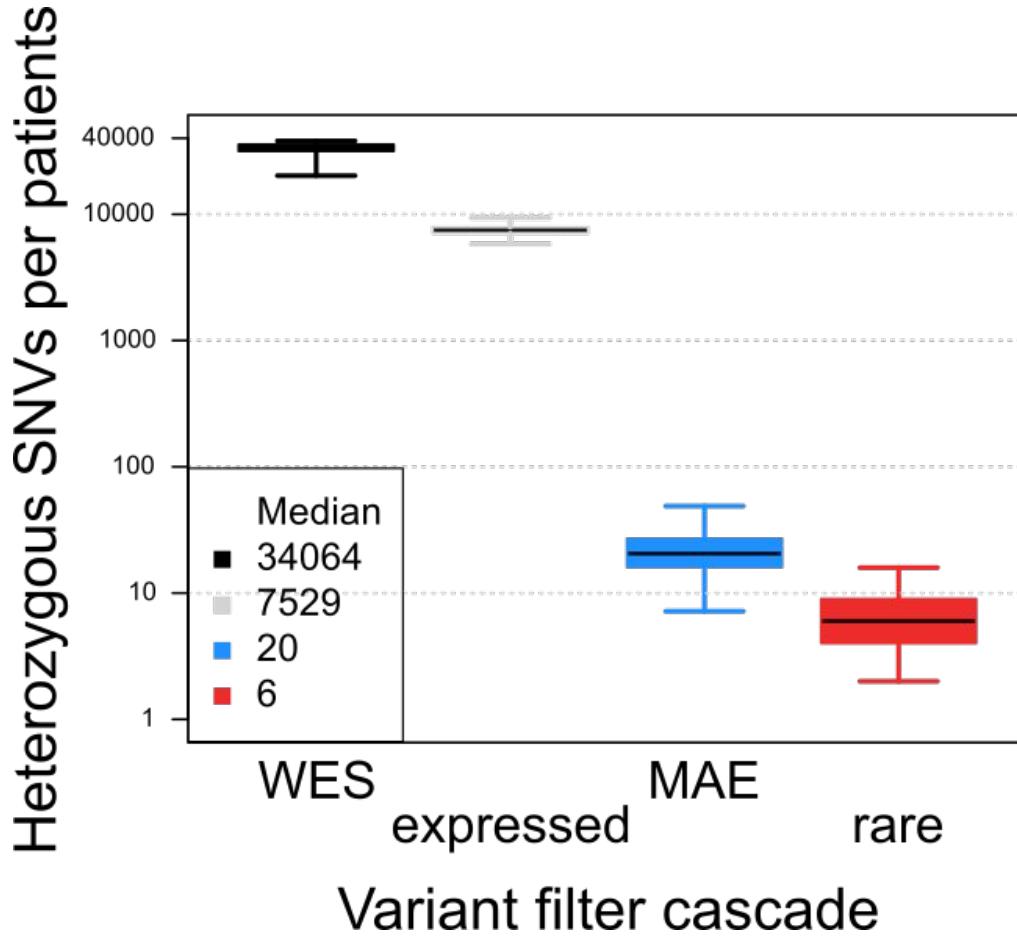
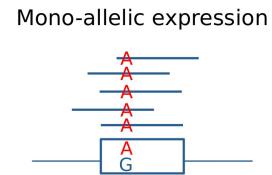




Mono-allelic expression

of the alternative allele

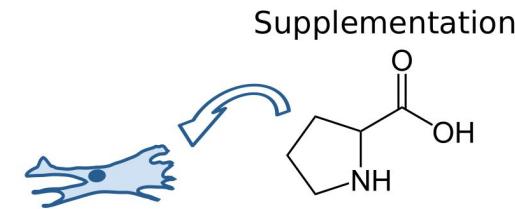
Detection summary mono-allelic expression



- MAE := Mono-allelic expression
- Read counts per variant
- Test for MAE within each sample
- MAE := P-value < 0.05 & Allele-frequency ≥ 0.8

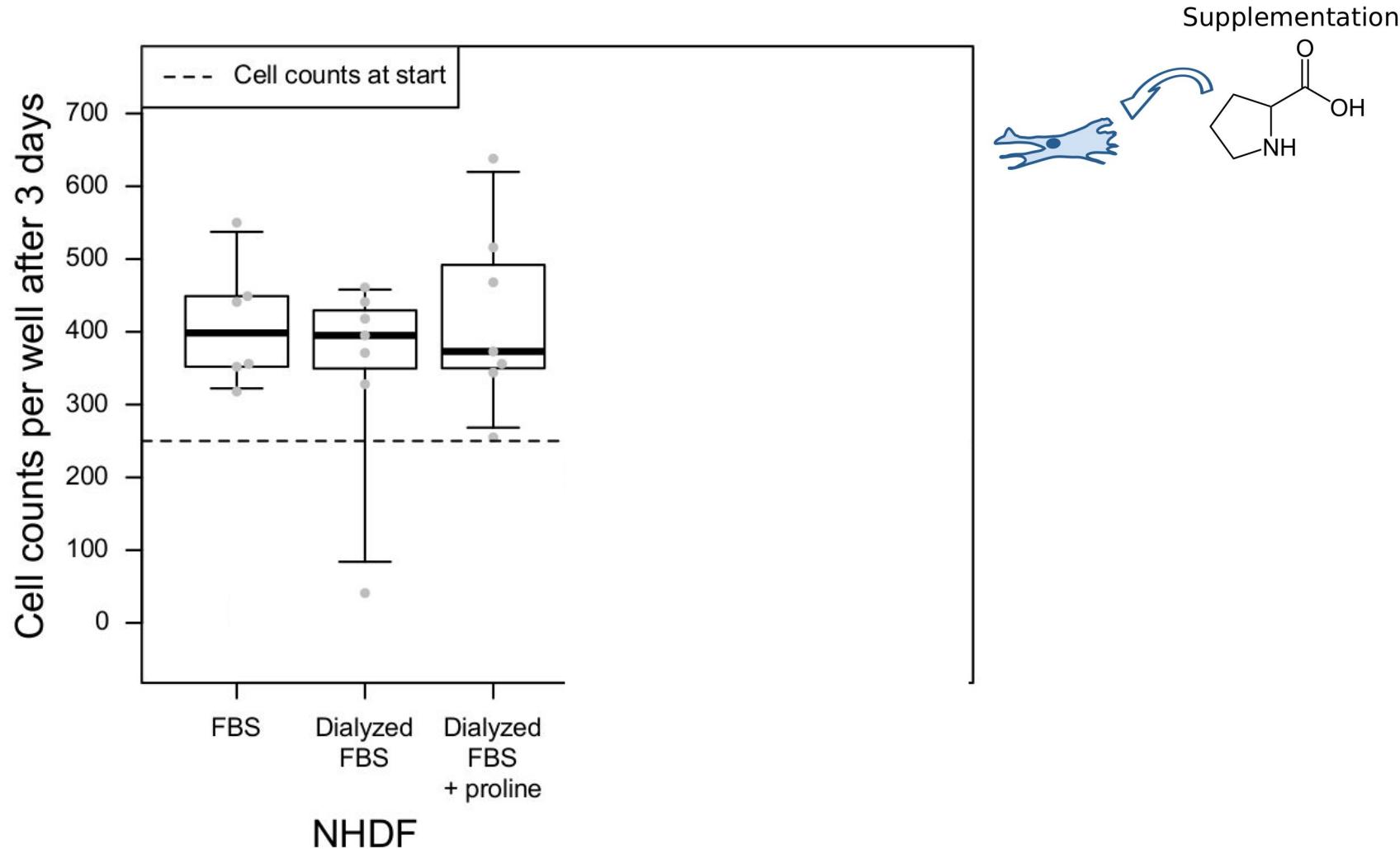
⇒ rare MAE in [ALDH18A1](#) for sample #80256

Rescue through proline supplementation



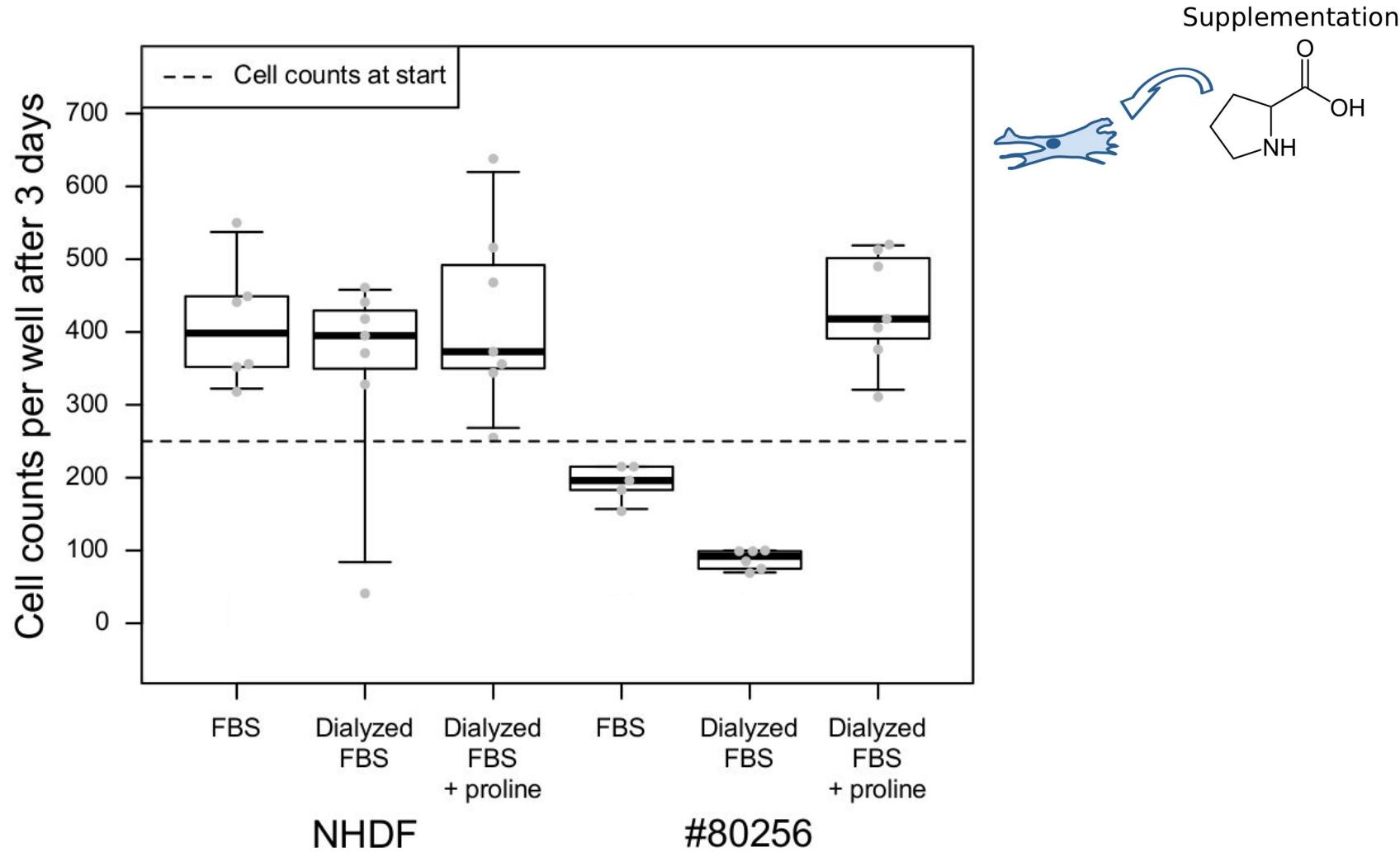
ALDH18A1: protein in [mitochondria](#); involved in the [biosynthesis of proline](#)

Rescue through proline supplementation



ALDH18A1: protein in [mitochondria](#); involved in the [biosynthesis of proline](#)

Rescue through proline supplementation



ALDH18A1: protein in [mitochondria](#); involved in the [biosynthesis of proline](#)

Conclusion

- RNA-Seq improves the diagnosis rate of WES by at least 10%
→ Similar results by Cummings et al. 2017 SciTranslMed
- RNA-Seq can detect
 - Aberrant expression
 - Aberrant splicing
 - Mono-allelic expression
- Weak (“cryptic”) splice sites are susceptible to variation

	TIMMD1	MGST1	CLPP	ALDH18A1	MCOLN1	Candidates per sample
Detected by						
Aberrant expression	✓	✓	-	✓	✓	1± 1
Aberrant splicing	✓	-	✓	-	✓	5± 3
Mono-allelic expression	-	-	-	✓	✓	6± 3
Validated by						
Proteomics/Western blot	✓	✓	✓	✓	-	
m/z						
Complementation	✓	-	-	-	-	
Supplementation	-	-	-	✓	-	
Disease associated variant detected	✓	-	✓	✓	✓	

Cryptic splicing

- Deep intronic variant for TIMMDC1
 - Not detected by WES
 - Hard to interpret with WGS
- Private exons arise from weak splice sites (PSI ~1%)
- Rare variants at cryptic splice sites
→ aberrant splicing

New Results

Genetic diagnosis of Mendelian disorders via RNA sequencing

Laura S Kremer,  Daniel M Bader, Christian Mertes, Robert Kopajtich, Garwin Pichler, Arcangela Iuso, Tobias B Haack, Elisabeth Graf, Thomas Schwarzmayr, Caterina Terrile, Eliska Konafikova, Birgit Repp, Gabi Kastenmüller, Jerzy Adamski, Peter Lichtner, Christoph Leonhardt, Benoit Funalot, Alice Donati, Valeria Tiranti, Anne Lombes, Claude Jardel, Dieter Gläser, Robert W Taylor, Daniele Ghezzi, Johannes A Mayr, Agnes Rötig, Peter Freisinger, Felix Distelmaier, Tim M Strom, Thomas Meitinger, Julien Gagneur, Holger Prokisch

doi: <http://dx.doi.org/10.1101/066738>



Laura S Kremer



Christian Mertes



Julien Gagneur



Holger Prokisch

In press at [Nature communications...](#)

Acknowledgements



HelmholtzZentrum münchen
German Research Center for Environmental Health



SOUND

Gagneur Lab:

- gagneurlab.in.tum.de
- Technical University of Munich
- Graduate School
Quantitative Biosciences Munich

The groups of Holger Prokisch and Tim Strom

mitOomics

Bioinformatische und statistische Analyse genomer Daten von Patienten mit mitochondrialen Krankheiten zur Identifizierung kausaler Mutationen und Pathways

Statistical Multi-Omics Understanding
Grant Agreement no. 633974

Conclusion

- RNA-Seq improves the diagnosis rate of WES by at least 10%
→ Similar results by Cummings et al. 2017 SciTranslMed
- RNA-Seq can detect
 - Aberrant expression
 - Aberrant splicing
 - Mono-allelic expression
- Weak (“cryptic”) splice sites are susceptible to variation

	TIMMD1	MGST1	CLPP	ALDH18A1	MCOLN1	Candidates per sample
Detected by						
Aberrant expression	✓	✓	-	✓	✓	1 ± 1
Aberrant splicing	✓	-	✓	-	✓	5 ± 3
Mono-allelic expression	-	-	-	-	✓	6 ± 3
Validated by						
Proteomics/Western blot	✓	✓	✓	✓	-	
m/z						
Complementation	✓	-	-	-	-	
Supplementation	-	-	-	✓	-	
Disease associated variant detected	✓	-	✓	✓	✓	

Supplement

Genetic diagnosis

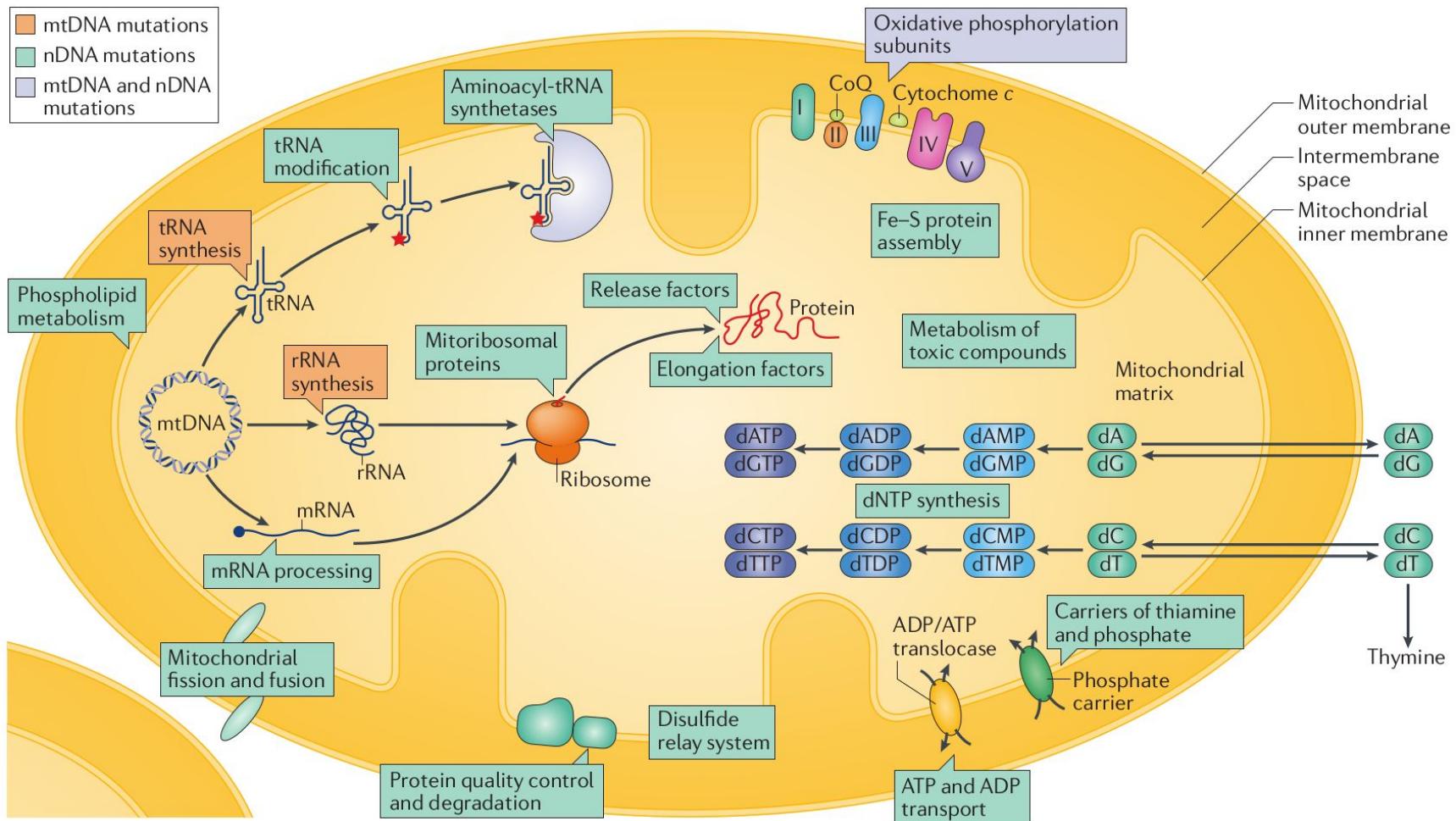
Unaffected tissue: fibroblasts

- Byproducts of muscle biopsies (routine in clinic)
→ biochemical diagnosis of mitochondrial disorders with enzymatic assays
- Limited accessibility of affected tissue,
e.g. brain, heart, skeletal muscle or liver
- RNA defects detectable
 - physiological consequences on fibroblasts might be negligible
 - advantage that the regulatory consequences on other genes are limited
- Perform perturbation and complementation tests in cell lines
→ rapid demonstration of candidate variant's role

Recovery rate of diagnosed patients

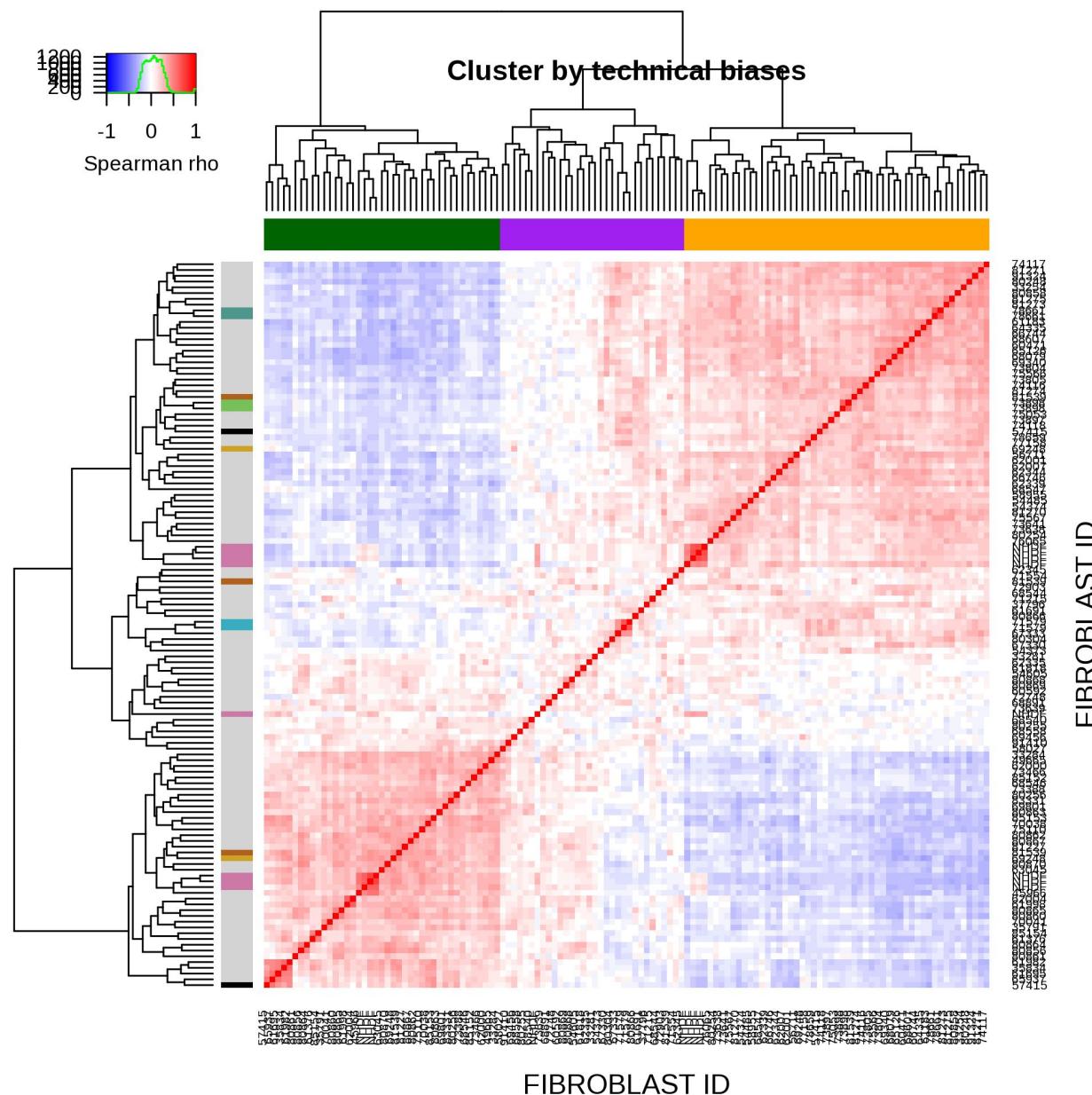
- 40 patients diagnosed before our study with WES & RNAseq available
- 7/8 splice variants detected by aberrant splicing
- 3/6 missense + stop or frame-shift by MAE
- 3/4 homozygous stop by aberrant expression
- 1/9 frame-shift variants with RNA defect (MAE)
- 0/14 homozygous missense variants with detected RNA defect

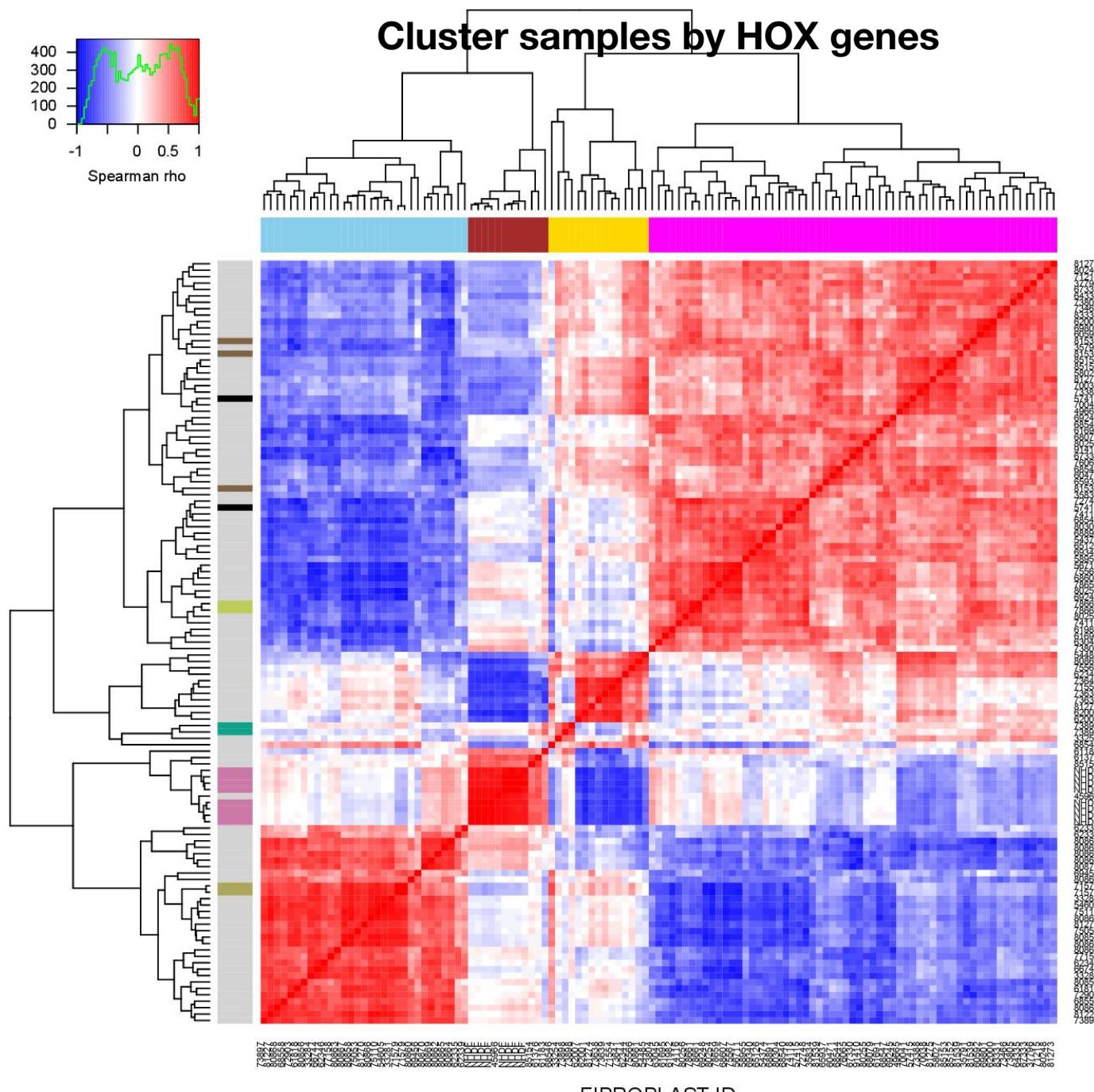
Mitochondrial disorders



Gorman 2016 Nat Rev Dis Primers

RNA normalization of confounding effects





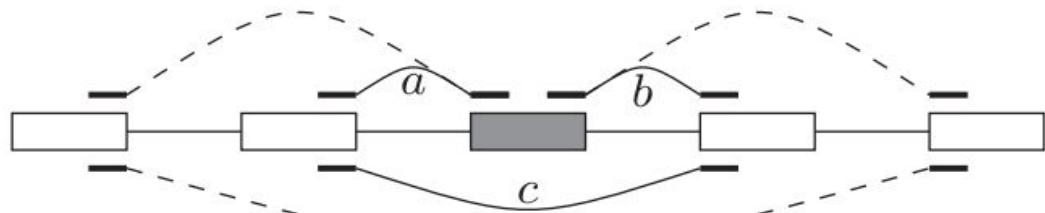


Leafcutter adaptation

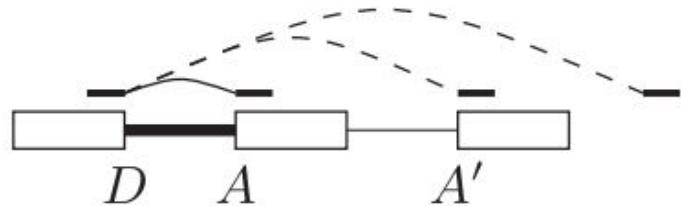
- Exon-junction reads per sample
- Cluster junctions over all samples
 - Keep low expressed junctions ($\Psi > 0.0001$)
 - Keep junctions supported by ≥ 10 reads by ≥ 1 sample
- Test for differential splicing
 - 1 versus rest (all other samples)
 - Include a pseudo sample in control group to have all junctions represented (conservative)

How to measure splicing: Percentage Spliced In (PSI - Ψ)

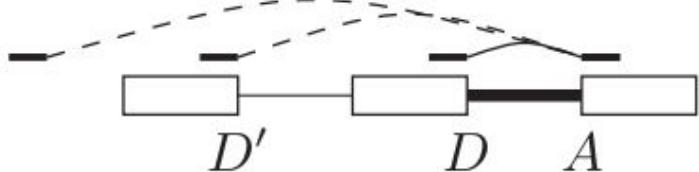
Aberrant splicing



$$\Psi = \frac{a + b}{a + b + 2c}$$

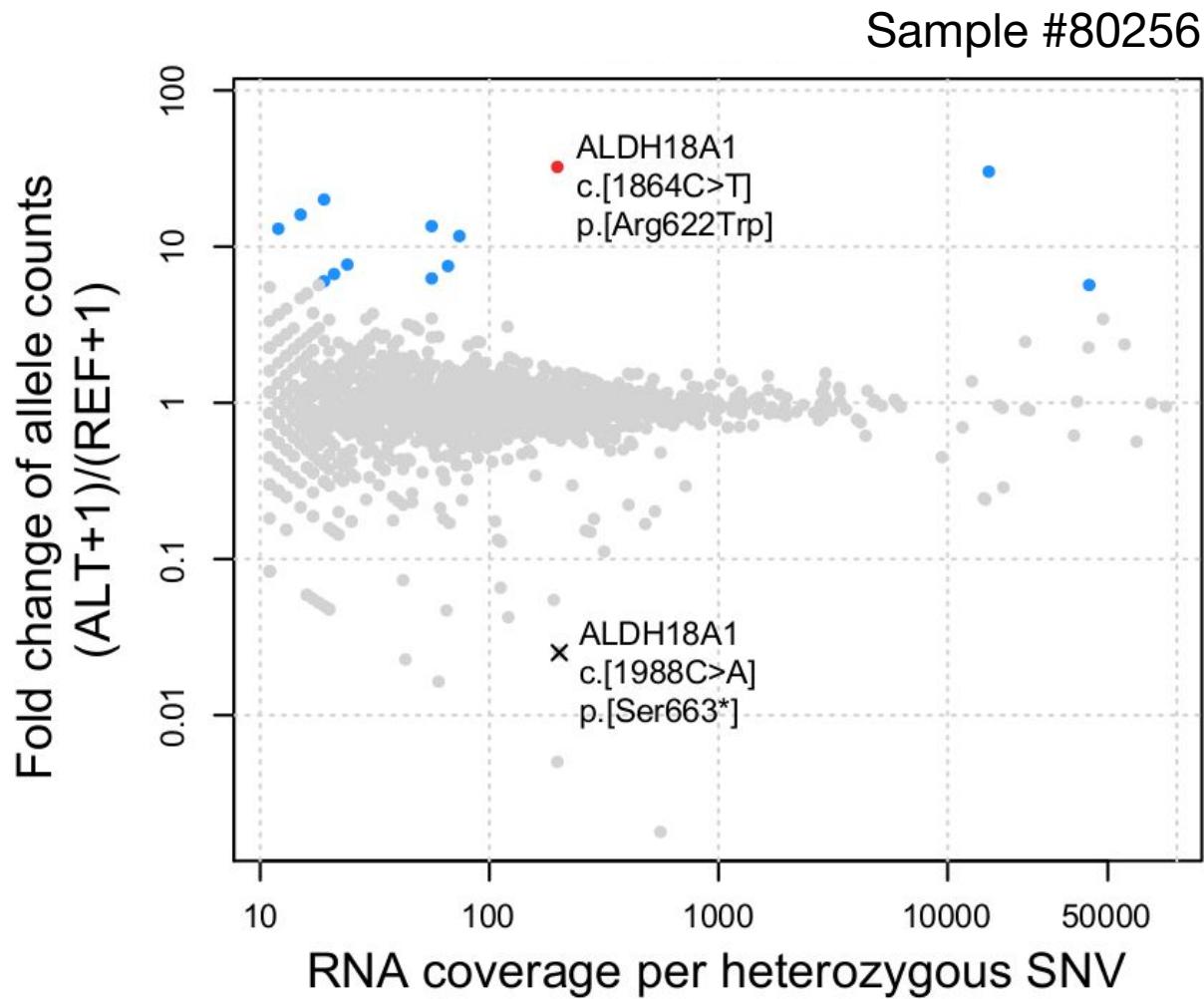
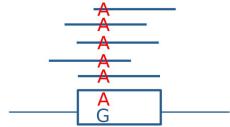


$$\psi_5(D, A) = \frac{n(D, A)}{\sum_{A'} n(D, A')}$$

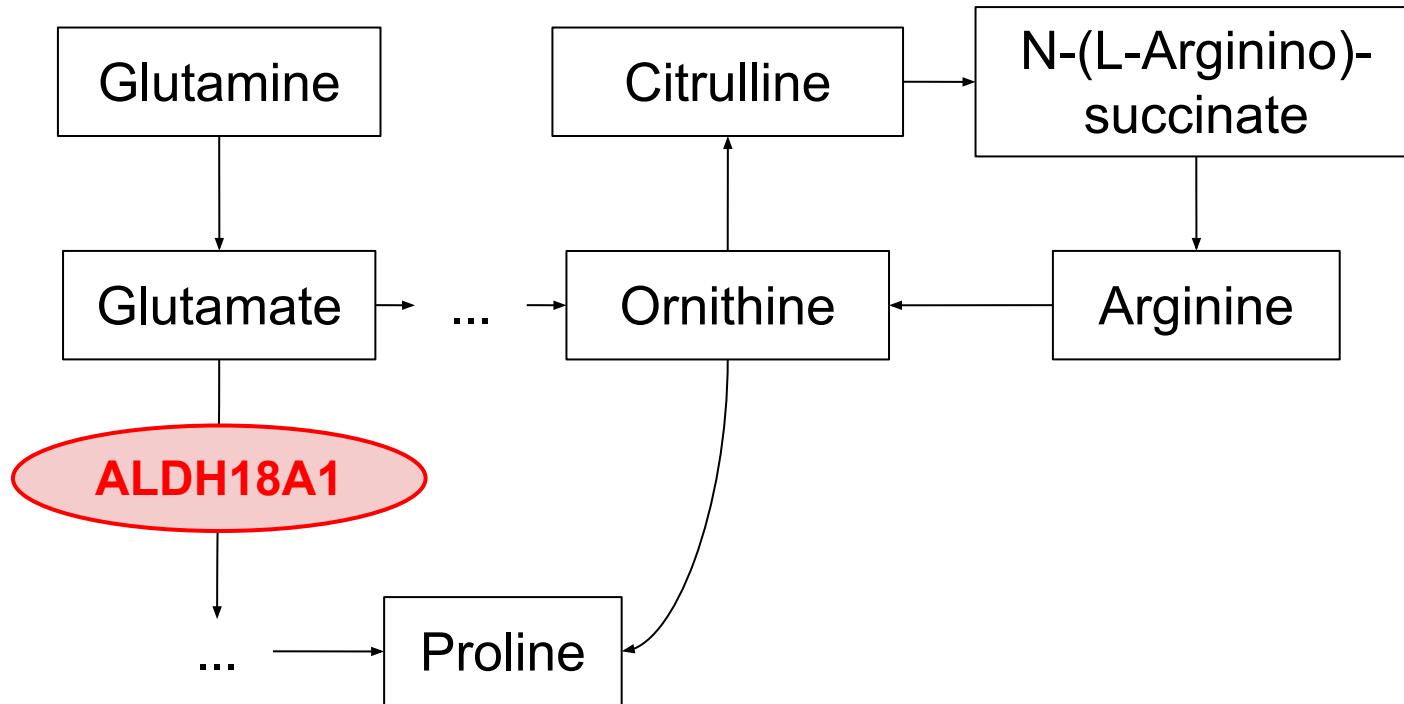


$$\psi_3(D, A) = \frac{n(D, A)}{\sum_{D'} n(D', A)}$$

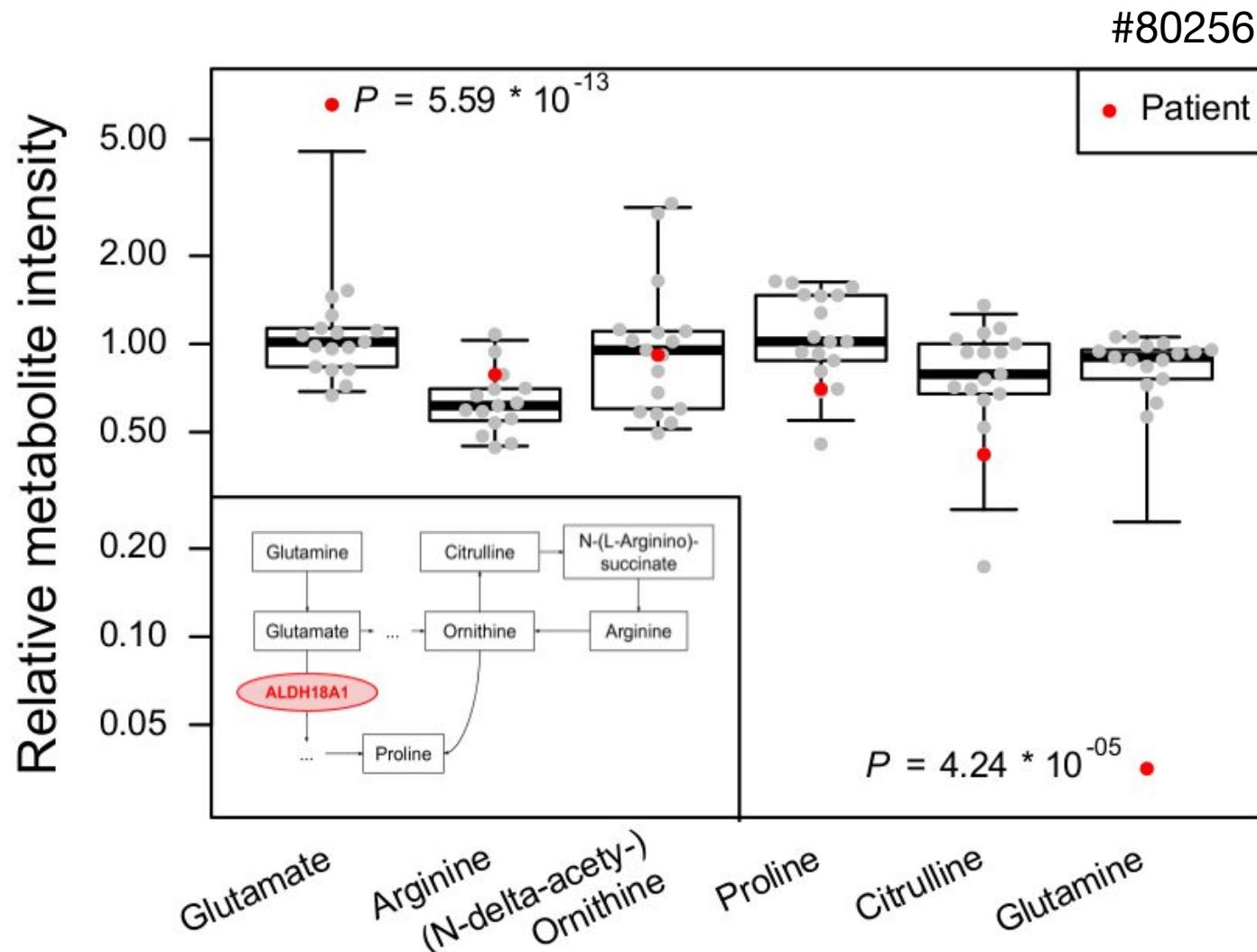
Example for mono-allelic expression



Proline synthesis pathway

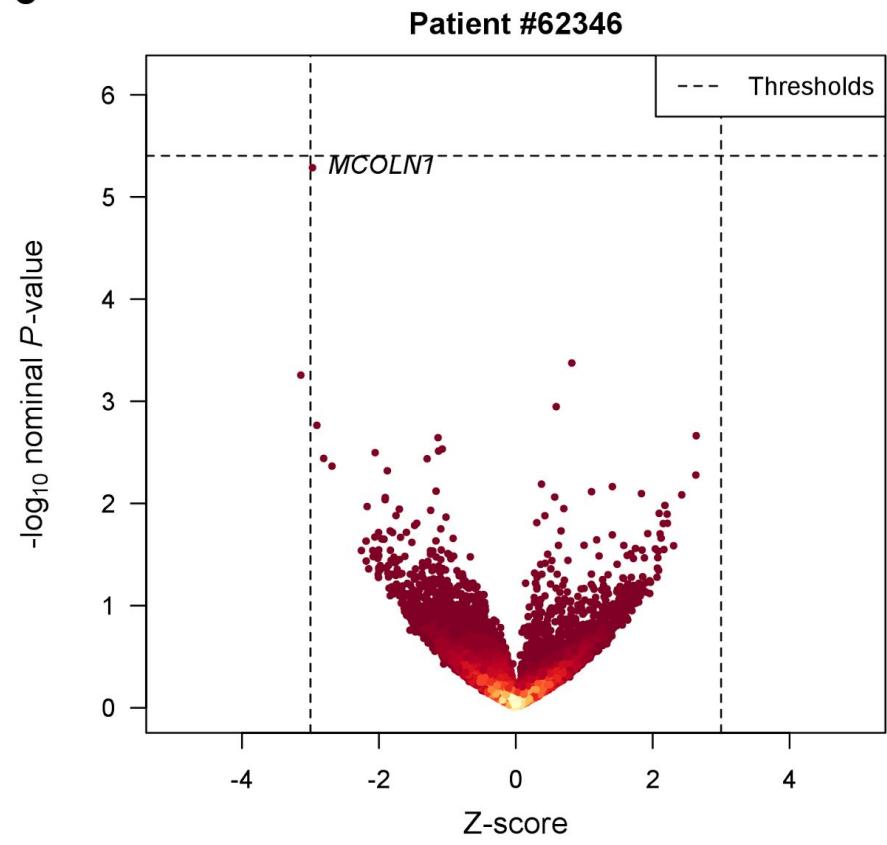


Metabolites associated with ALDH18A1

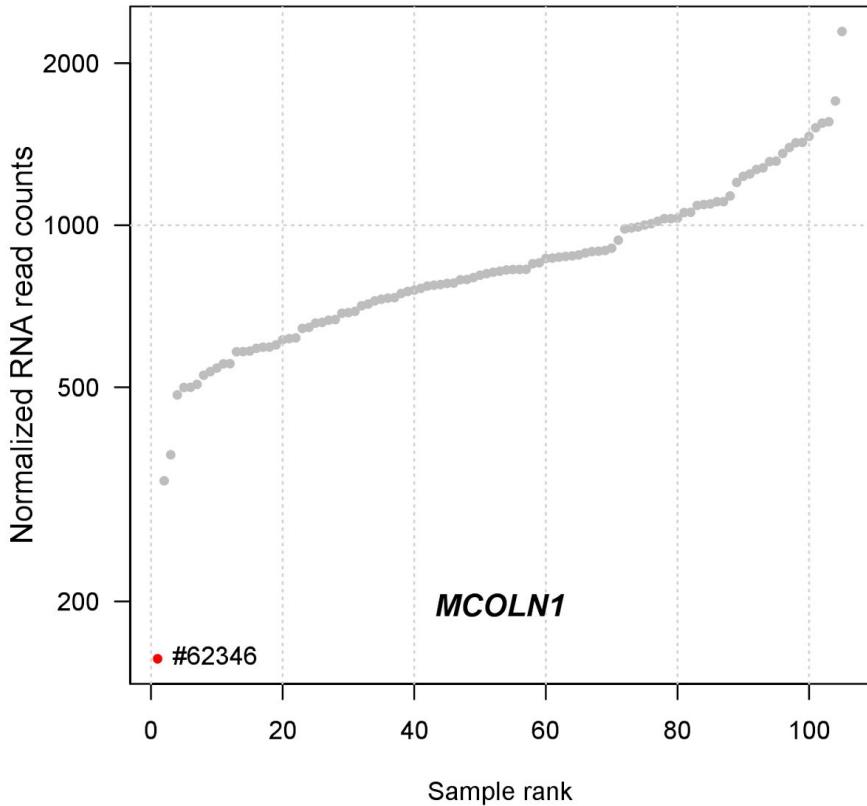


MCOLN1

c



d



MCOLN1: Diseases associated with Mucolipidosis IV and Cerebral Palsy, Ataxic.