Curso Técnicas Ómicas en el Diagnóstico de Enfermedades Raras

TRANSCRIPTÓMICA

Marc Dabad
Gerard Muñoz-Pujol

17.11.2022

Contents:

1. Introduction

Improved diagnostics rates using omics profiling

Patient (Genetic disease)

WES/WGS
Variant Calling & Annotation
20.000 genes
Filtering & Interpretation

Phenotype Correlation

Functional Studies

Molecular Diagnosis
Diagnostic Yield of 32-62%

Limitations:

- WES misses many genomic regions → not detecting known disease-causal variants
  Also… copy-number and structural variants are not well captured
- WGS detects >3.4M SNVs → variant prioritization
- Complex pathomechanisms or phenotypes, multiple modes of inheritance, …

To improve the diagnosis rate, complementary approaches to DNA-seq need to be used.

No candidate

Periodic re-analysis with discovery of novel disease genes

Multi-Omic discovery studies

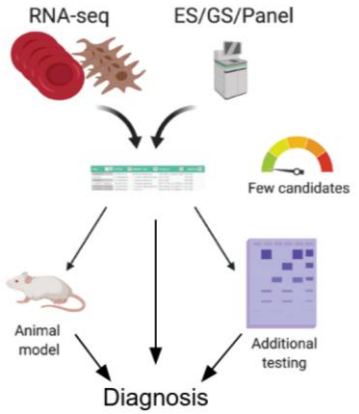
Transcriptomics
Aberrant Expression, Aberrant Splicing, MAE

Proteomics
Aberrant Abundance

Epigenomics
Episignatures, Epivariation

Molecular Diagnosis
Increased diagnostic yield

Transcriptome-directed Genomic Analysis

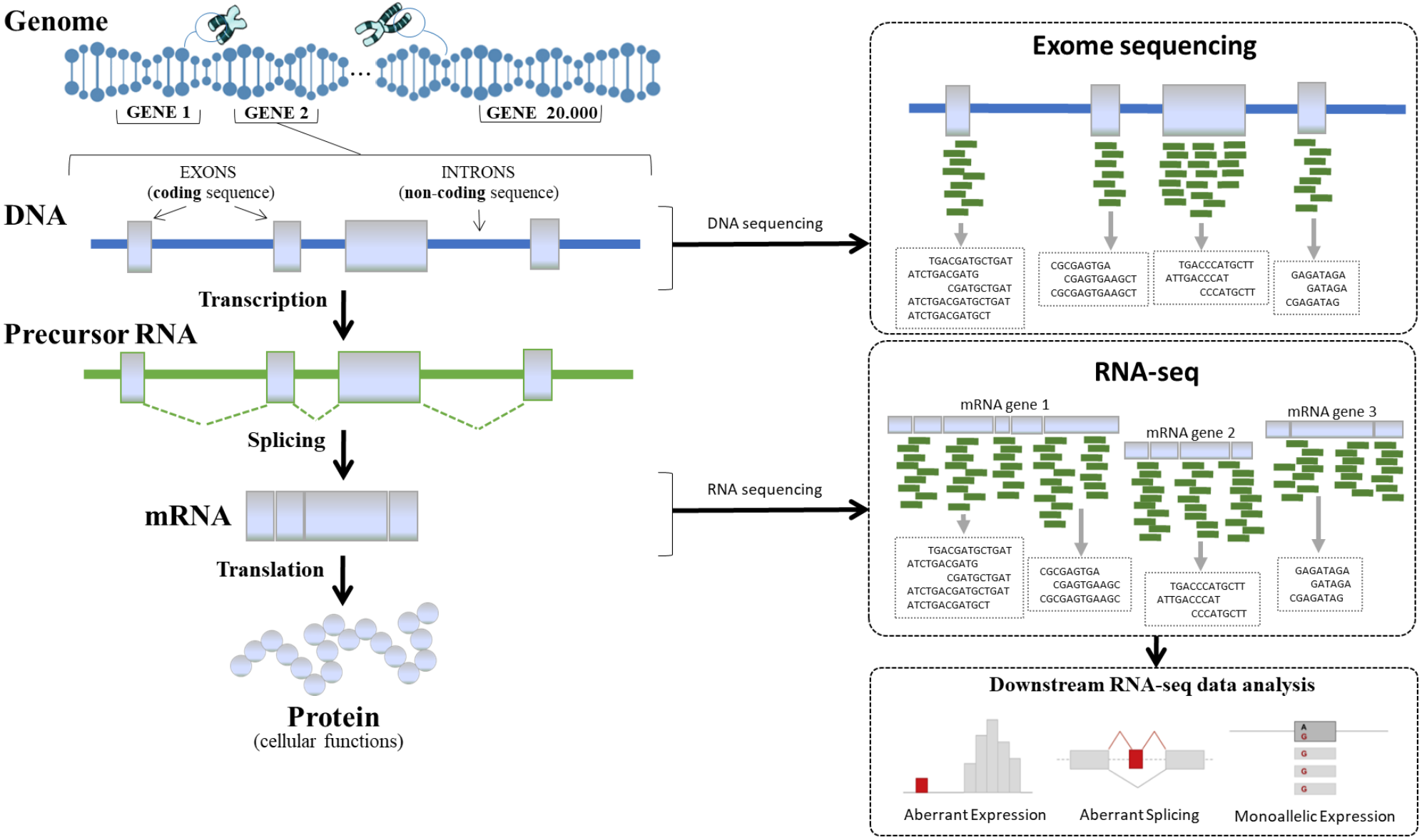


(Murdock, DR et al., 2021)

- **30%** disease-causing variants impact RNA and fall within non-coding regions (1/3 are splicing defects)
- **2000 VUS** are located in direct splice sites → **functional evidence**

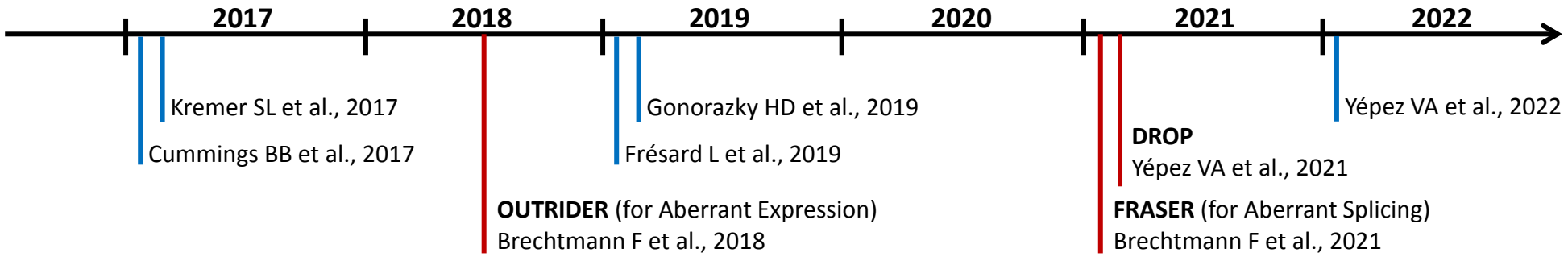
1. Introduction

NGS also allowed the advent of RNA sequencing (RNA-seq)



2. Timeline of RNA-seq implementation in diagnostics

RNA-seq has shown diagnostic utility in multiple rare disease cohorts



	N	Tissue	Disease Group	Methods	Diagnostic rate
Cummings BB et al., 2017	50		Neuromuscular	AS	**34% (17 cases)**
Kremer SL et al., 2017	48		Mitochondrial	AS, AE, MAE	**10% (5 casos)**
Frésard L et al., 2019	94		Mendelian	AS, AE, MAE	**8% (6 cases)**
Gonorazky HD et al., 2019	25		Neuromuscular	AS, AE, MAE	**36% (9 cases)**
Yépez VA et al., 2022	217		Mitochondrial	**DROP:** AS, AE, MAE	**15% (33 cases)**

RNA-seq can increase the diagnosis rates over DNA sequencing alone by 8–36%, depending on the disease entity and tissue probed.

3. DROP, *Detection of RNA Outliers Pipeline*

An automated RNA-seq computational workflow

nature protocols

PROTOCOL

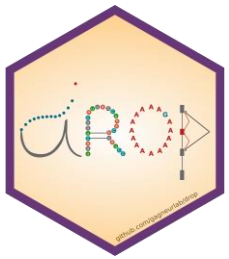
https://doi.org/10.1038/s41596-020-00462-5

Check for updates

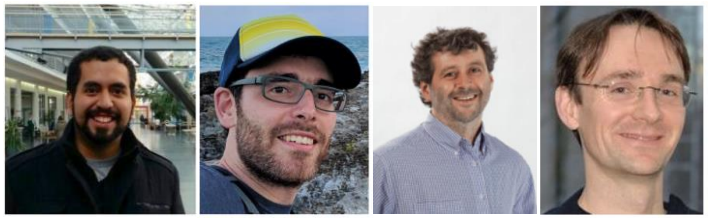
Detection of aberrant gene expression events in RNA sequencing data

Vicente A. Yépez[1,2,3], Christian Mertes[1], Michaela F. Müller[1], Daniela Klaproth-Andrade[1], Leonhard Wachutka[1], Laure Frésard[4], Mirjana Gusic[3,5,6], Ines F. Scheller[1,7], Patricia F. Goldberg[1], Holger Prokisch[3,5] and Julien Gagneur[1,3,7]✉

RNA sequencing (RNA-seq) has emerged as a powerful approach to discover disease-causing gene regulatory defects in individuals affected by genetically undiagnosed rare disorders. Pioneering studies have shown that RNA-seq could increase the diagnosis rates over DNA sequencing alone by 8–36%, depending on the disease entity and tissue probed. To accelerate adoption of RNA-seq by human genetics centers, detailed analysis protocols are now needed. We present a step-by-step protocol that details how to robustly detect aberrant expression levels, aberrant splicing and mono-allelic expression in RNA-seq data using dedicated statistical methods. We describe how to generate and assess quality control plots and interpret the analysis results. The protocol is based on the detection of RNA outliers pipeline (DROP), a modular computational workflow that integrates all the analysis steps, can leverage parallel computing infrastructures and generates browsable web page reports.



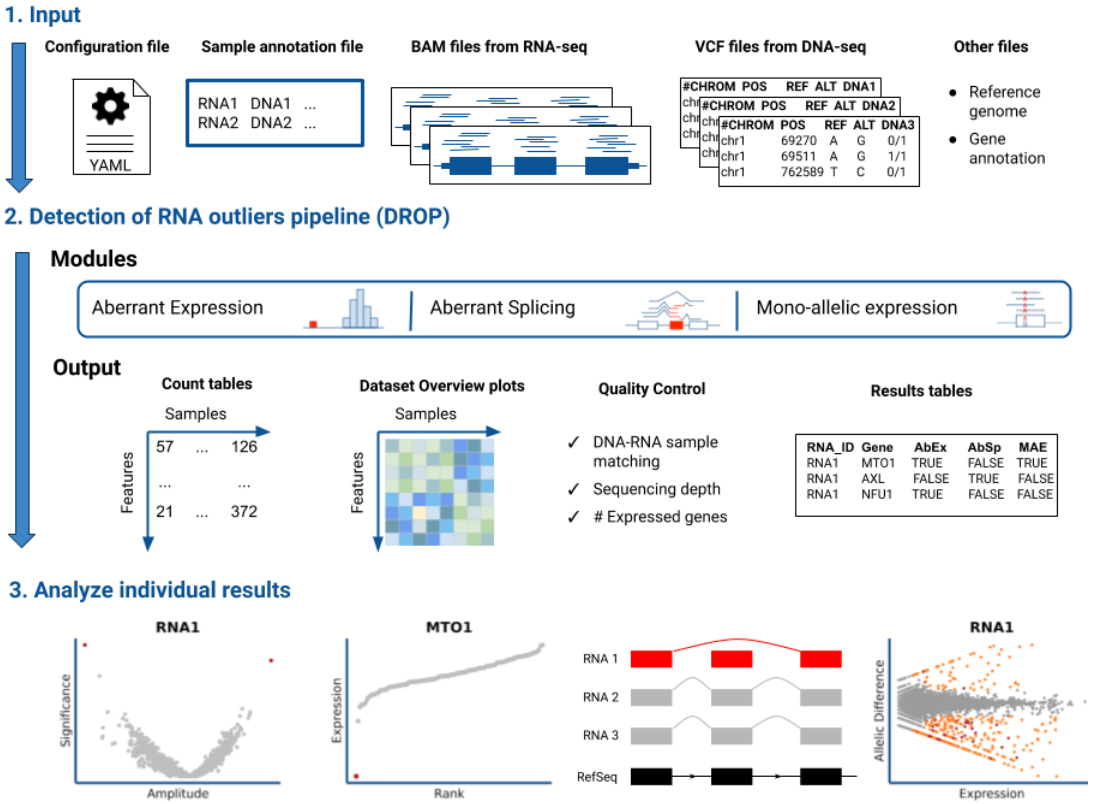
https://github.com/gagneurlab/drop



Technical University of Munich

HelmholtzZentrum münchen

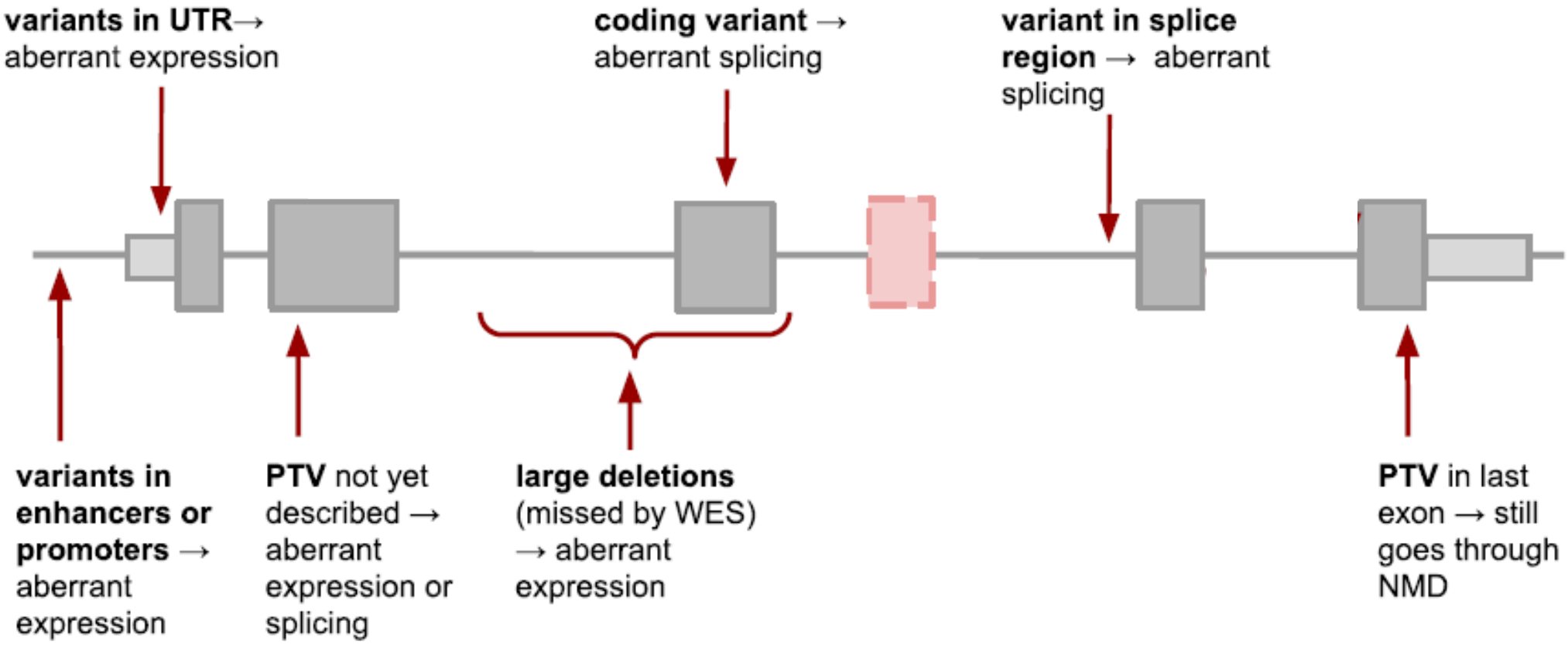
German Research Center for Environmental Health



3.1 Aberrant Expression

DROP

Aberrant expression is defined as an expression that **significantly deviates** from the normal physiological range.



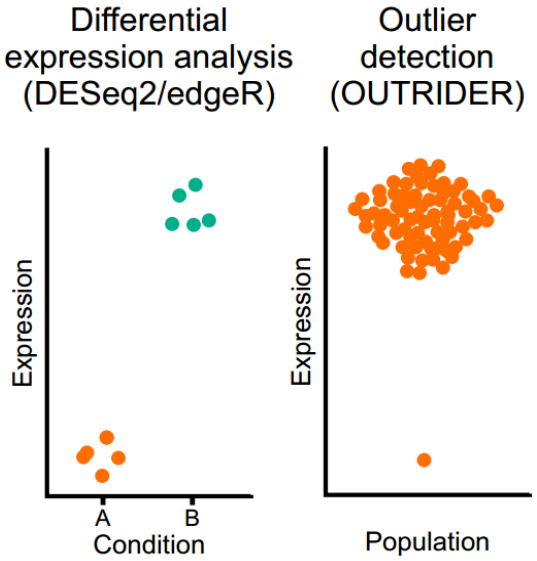
3.1 Aberrant Expression

DROP

Aberrant expression is defined as an expression that **significantly deviates** from the normal physiological range.

Aberrant expression <u>outlier</u> are genes whose expression in a sample is aberrant with respect to other samples from the same population.

Outlier calling is not a differential expression analysis!



Kremer SL et al., 2017

Frésard L et al., 2019

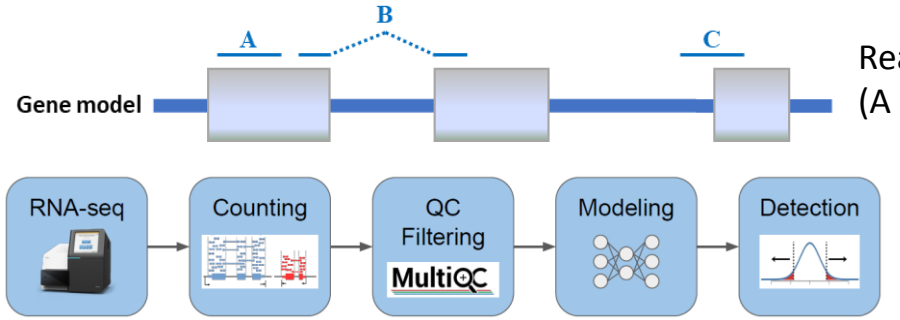
Gonorazky HD et al., 2019

z-Score approach

Yépez VA et al., 2022

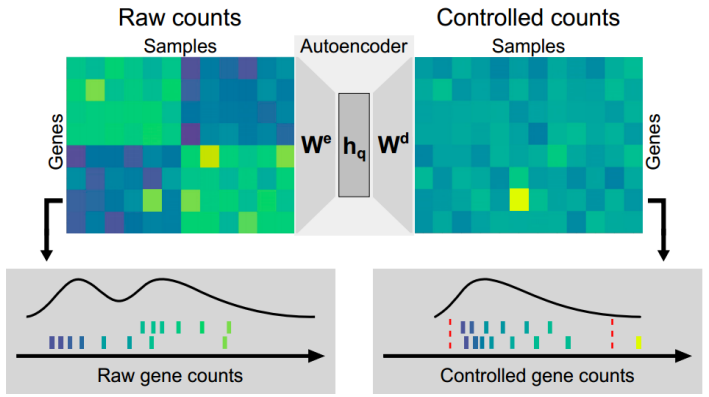
OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data

Felix Brechtmann,[1,5] Christian Mertes,[1,5] Agnė Matusevičiūtė,[1,5] Vicente A. Yépez,[1,2] Žiga Avsec,[1,2] Maximilian Herzog,[1] Daniel M. Bader,[1] Holger Prokisch,[3,4] and Julien Gagneur[1,2,*]



Reads that fully overlap exonic regions (A and B) are aggregated by gene.

Context-dependent outlier detection



OUTRIDER (OUTlier in RNA-seq fInDER) is a tool for finding aberrantly expressed genes in RNA-seq samples. It does so by fitting a <u>negative binomial model</u> to RNA-seq read counts, correcting for variations in sequencing depth and apparent co-variations across samples (*denoising autoencoders*).

Read counts that significantly deviate from the distribution are detected as outliers.

https://bioconductor.org/packages/release/bioc/html/OUTRIDER.html

3.1 Aberrant Expression

DROP

Aberrant expression is defined as an expression that **significantly deviates** from the normal physiological range.

Aberrant expression <u>outlier</u> are genes whose expression in a sample is aberrant with respect to other samples from the same population.

<u>RESULTS VISUALIZATION:</u>

	Annotation		Significance		Effect Size		Raw Data		Distribution parameters	
	geneID	sampleID	pValue	padjust	zScore	l2fc	rawcounts	normcounts	meanCorrected	theta
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>	<dbl>
1	ZNF549	NA12341	2.794264e-20	4.465636e-15	-9.36	-4.90	4	5.46	202.81	32.67
2	CAT	NA18873	1.277582e-17	2.041760e-12	-9.18	-5.49	24	26.68	1242.71	14.33
3	FAM127A	NA06984	2.171443e-16	3.470280e-11	-8.73	-2.80	65	61.87	436.61	37.30
4	TXN2	HG00103	9.124427e-16	1.458215e-10	-8.38	-1.48	852	974.13	2712.67	89.84
5	PARP4	NA18916	4.215504e-14	6.736984e-09	-8.16	-2.21	833	572.89	2648.97	39.61
6	PKP4	NA12717	1.351949e-13	2.160610e-08	-8.55	-4.80	6	24.30	774.88	15.92

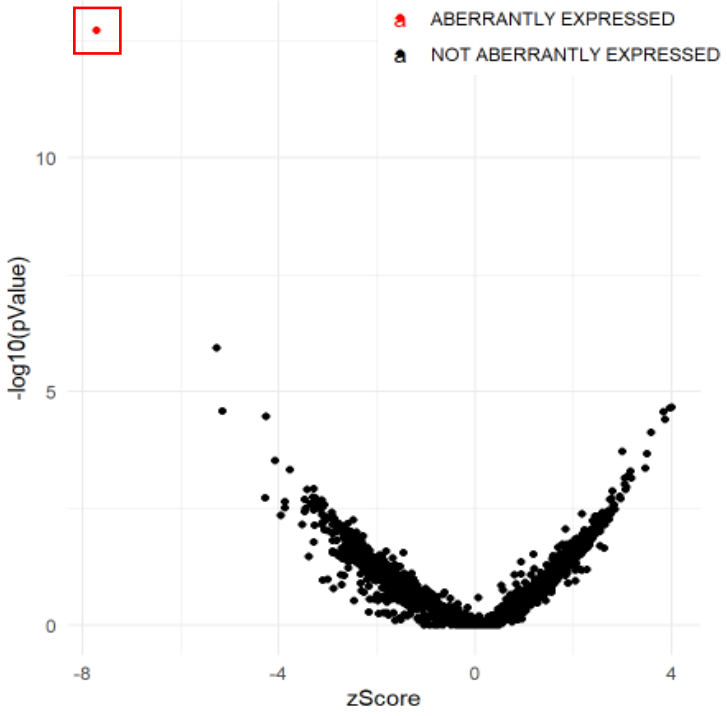
3.1 Aberrant Expression

DROP

Aberrant expression is defined as an expression that **significantly deviates** from the normal physiological range.

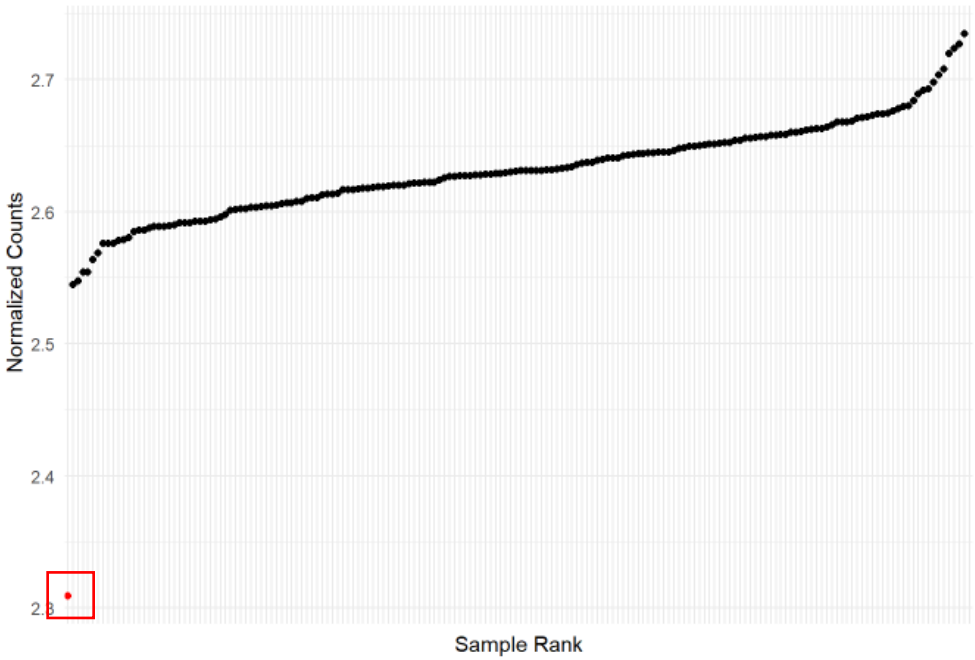
Aberrant expression <u>outlier</u> are genes whose expression in a sample is aberrant with respect to other samples from the same population.

<u>RESULTS VISUALIZATION:</u>



Volcano plot

Expression for a given gene:



Expression rank plot

3.1 Aberrant Splicing

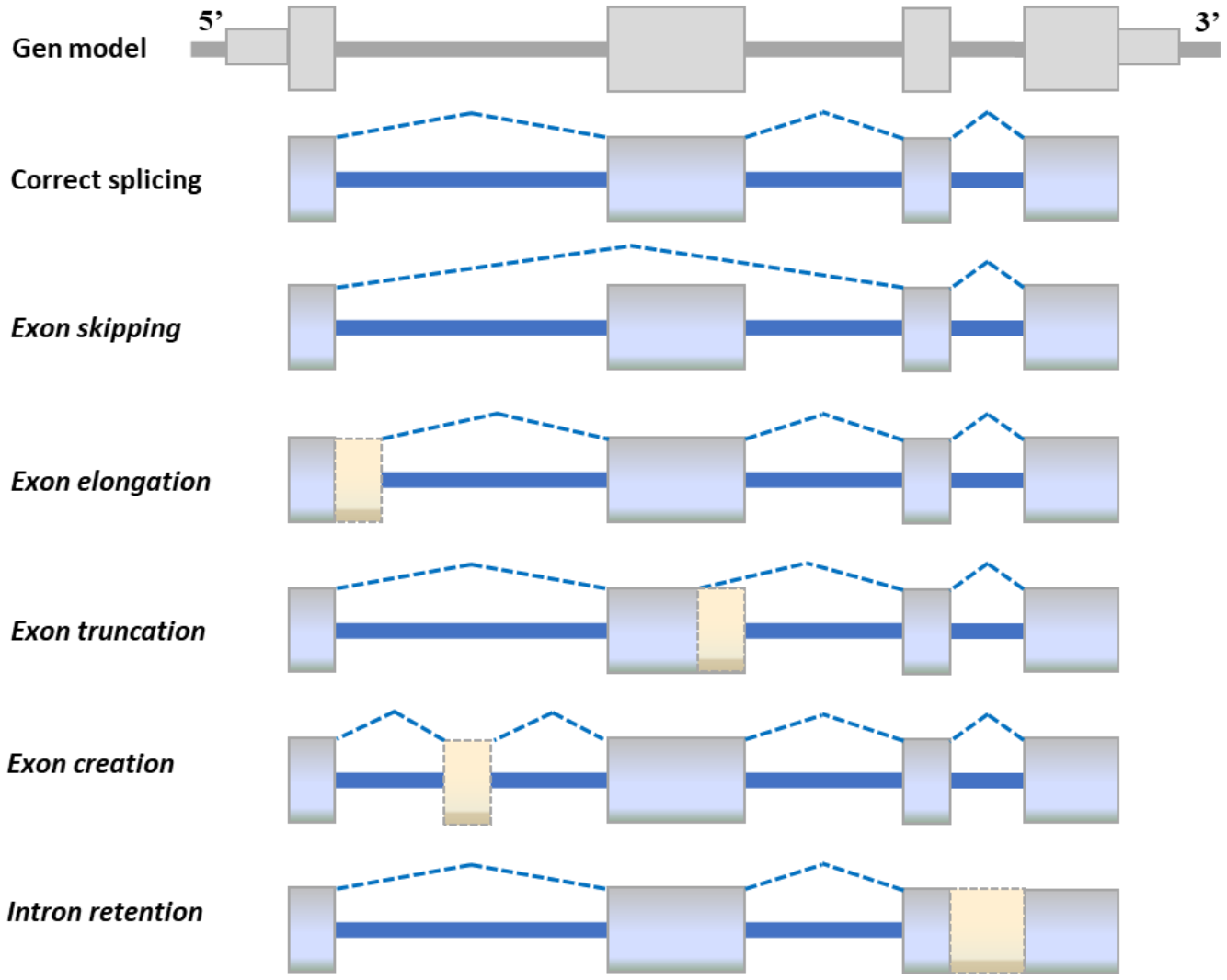
DROP

Splicing defects are involved in numerous genetic diseases.

Aberrant splicings can be caused by variants in the **canonical splice sites**, but also by variants in the less defined **splicing regulatory sequences** such as the exonic and intronic splicing enhancers.

RNA-seq can pinpoint disease-causing variants not covered by WES when are affecting a splicing.

Also… can provide functional evidence of how transcripts are being processed given a particular variant.



3.1 Aberrant Splicing

DROP

Kremer SL et al., 2017 → LeafCutter

Cummings BB et al., 2017

Gonorazky HD et al., 2019

cutoff based approach

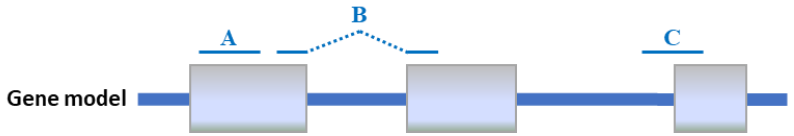
Frésard L et al., 2019 → z-Score approach

Yépez VA et al., 2022 → FRASER

	Control for covariation	Explicit control for known confounders	Stat. significance	Count data distribution
Cutoff-based Cummings et al. Sci Trans Med 2017	✗	✗	✗	✗
Leafcutter Kremer et al. Nat Comm 2017	✗	✗	✓	✓
SVA + z-score Fresard et al. Nat Med 2019	✓	✓	✗	✗
FRASER Mertes et al. bioRxiv 2019	✓	✗	✓	✓

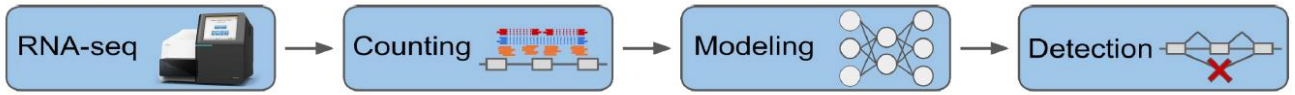
Detection of aberrant splicing events in RNA-seq data using FRASER

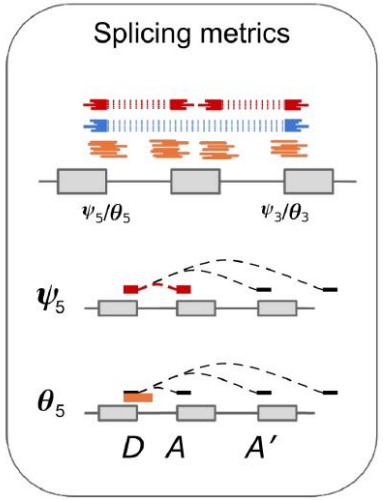
Christian Mertes[1,6], Ines F. Scheller[1,2,6], Vicente A. Yépez[1,3], Muhammed H. Çelik[1], Yingjiqiong Liang[1], Laura S. Kremer[4,5], Mirjana Gusic[4,5], Holger Prokisch[4,5] & Julien Gagneur[1,2,5]✉

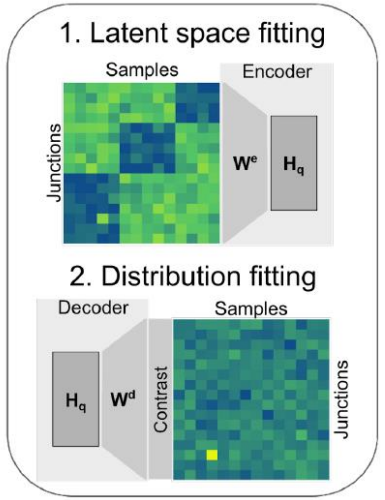


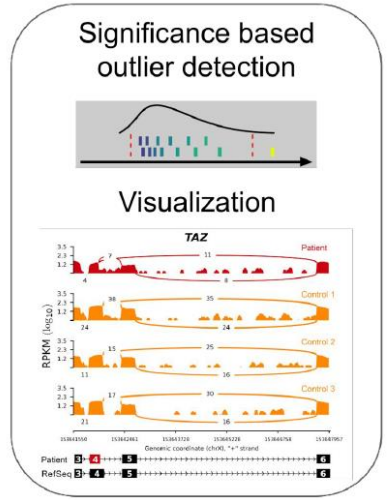
Reads spanning from one exon to another (**split reads**) and reads overlapping an exon-intron boundary (**non-split reads**) can also be quantified and aggregated by junction.

FRASER computes splicing outliers from BAM files.









Beta-binomial distribution

https://bioconductor.org/packages/release/bioc/html/FRASER.html

3.1 Aberrant Splicing

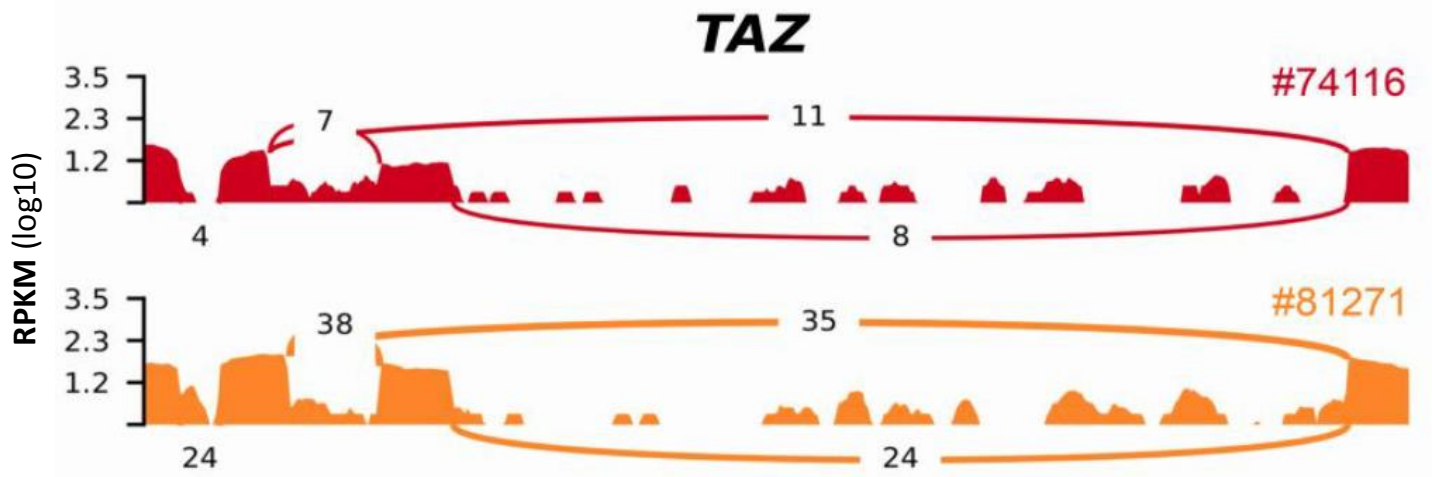
DROP

RESULTS VISUALIZATION:

Genomic position | Annotation | Significance | Effect Size | Raw Data

A data.table: 24 × 17

seqnames	start	end	width	strand	sampleID	hgncSymbol	type	pValue	padjust	zScore	psiValue	deltaPsi	meanCounts	meanTotalCounts	counts	totalCounts
<fct>	<int>	<int>	<int>	<fct>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>
chr3	53276259	53289769	13511	-	HG00116	TKT	psi3	1.5630e-09	0.00011289	5.05	0.16	0.16	1.87	457.43	56	350
chr3	53276259	53289851	13593	-	HG00116	TKT	psi3	1.5630e-09	0.00011289	-4.94	0.84	-0.16	455.50	457.43	294	350
chr19	1475249	1478772	3524	-	HG00257	C19orf25	psi5	3.3359e-07	0.02405900	4.93	0.59	0.54	0.83	33.53	23	39
chr19	1475258	1478772	3515	-	HG00257	C19orf25	psi5	3.3359e-07	0.02405900	-2.57	0.33	-0.44	24.60	33.53	13	39
chr19	35505292	35506727	1436	+	HG00275	GRAMD1A	psi3	3.7653e-07	0.02719500	-3.66	0.56	-0.41	86.97	89.07	54	96
chr19	35505325	35506727	1403	+	HG00275	GRAMD1A	psi3	3.7653e-07	0.02719500	3.99	0.33	0.31	1.13	89.07	32	96
chr19	18390985	18390986	2	*	HG00338	JUND	psiSite	5.8347e-07	0.04924300	3.92	0.98	0.46	150.10	162.17	159	163
chr19	18391017	18391018	2	*	HG00338	JUND	psiSite	8.8078e-07	0.04924300	3.79	0.84	0.52	157.73	174.73	176	210

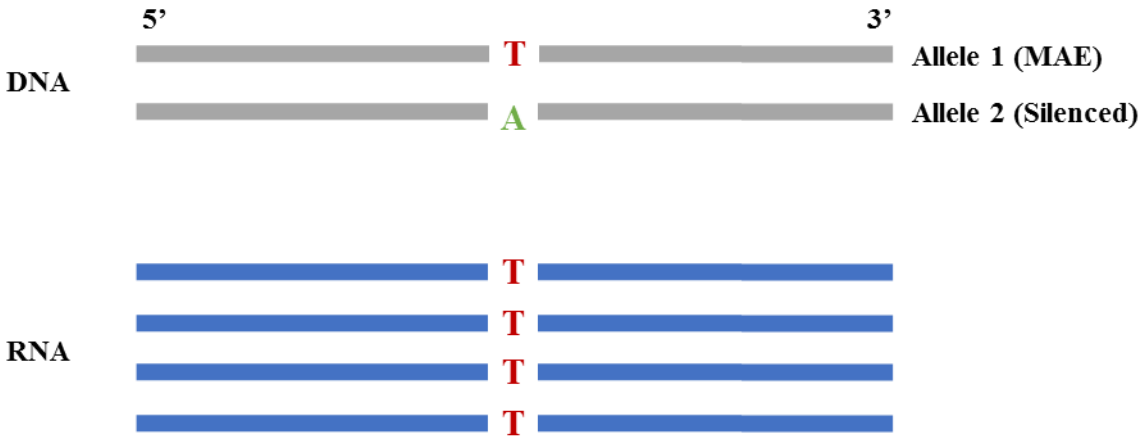


Sashimi plots visualize splice junctions from aligned RNA-seq data and a gene annotation track.

Genomic coordinates are plotted on x-axis and read density (whose value is configurable via IGV) on y-axis.

3.1 Monoallelic Expression

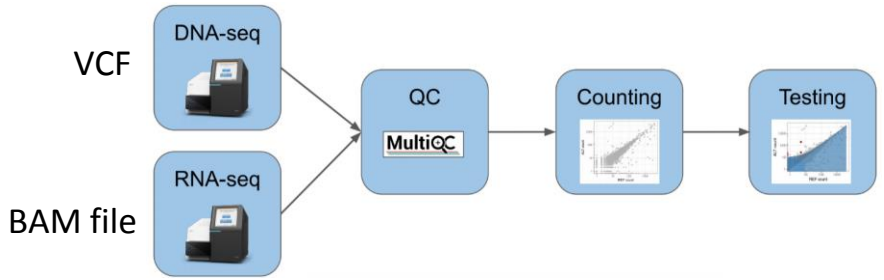
DROP



MAE refers to the expression of a single allele out of the two alleles of a gene, which could be due to genetic or epigenetic silencing of the other allele.

DROP --> Method developed in **Kremer SL et al., 2017**

- **Negative binomial test**

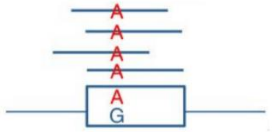


How to obtain allelic counts

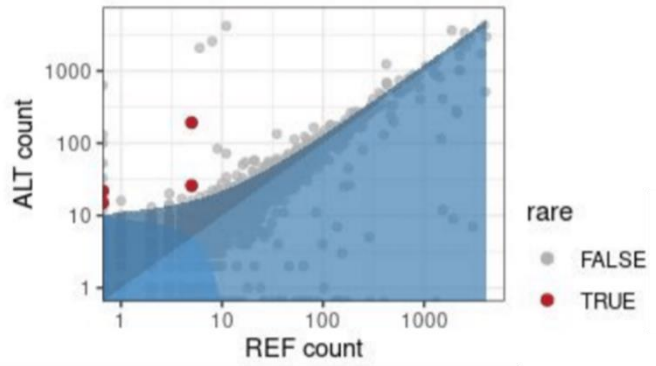
1. Call variants on either WES or WGS

2. Subset for heterozygous SNVs only
   - Homozygous variants don't provide allele-specific information

3. Count RNA reads assigned to each allele
   - `ASEReadCounter` from GATK
   - Discard:
     - low quality reads and variants
     - duplicated reads



Negative binomial distribution



3.4 Dataset design – limitations of RNA-seq

General considerations

- **Sample size**
- **Datasets are noisy**
- **Genes are expressed in a tissue specific manner**
- **The disease tissue is not always available**

3.4 Dataset design – limitations of RNA-seq

General considerations

- **Sample size**

What is the minimum number of samples needed to properly detect aberrant events?

Power analyses have suggested analyzing groups of at least **50 samples** for **aberrant expression** and at least **30 samples** for **aberrant splicing**.

3.4 Dataset design – limitations of RNA-seq

General considerations

- **Sample size**

What is the minimum number of samples needed to properly detect aberrant events?

Power analyses have suggested analyzing groups of at least **50 samples** for **aberrant expression** and at least **30 samples** for **aberrant splicing**.

- **How to boost sample size?**
  - Integrate control samples from GTEx and GEUVADIS
  - Integrate with Kremer et al., 2017 (non-strand specific fibroblasts) and Murdock et al., 2020 (strand-specific blood and fibroblasts) datasets, available through DROP

Datasets

The following publicly-available datasets of gene counts can be used as controls. Please cite as instructed for each dataset.

- 154 non-strand specific fibroblasts, build hg19, Technical University of Munich: DOI 10.5281/zenodo.4646822
- 269 strand specific fibroblasts, build hg19, Technical University of Munich: DOI 10.5281/zenodo.4646826
- 49 tissues, each containing hundreds of samples, non-strand specific, build hg19, GTEx: DOI 10.5281/zenodo.5596755
- 49 tissues, each containing hundreds of samples, non-strand specific, build hg38, GTEx: DOI 10.5281/zenodo.6078396
- 139 strand specific fibroblasts, build hg19, Baylor College of Medicine: DOI 10.5281/zenodo.3963473
- 125 strand specific blood, build hg19, Baylor College of Medicine: DOI 10.5281/zenodo.3963470

If you want to contribute with your own count matrices, please contact us: yepez at in.tum.de

3.4 Dataset design – limitations of RNA-seq

General considerations

- **Sample size**

What is the minimum number of samples needed to properly detect aberrant events?

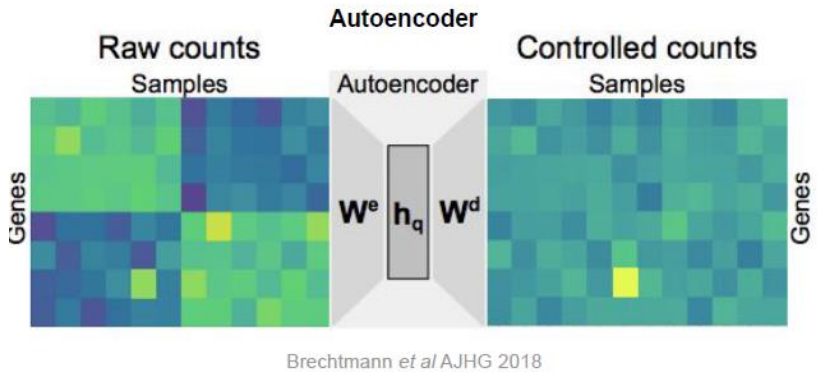
Power analyses have suggested analyzing groups of at least **50 samples** for **aberrant expression** and at least **30 samples** for **aberrant splicing**.

  - **How to boost sample size?**
    - Integrate control samples from GTEx and GEUVADIS
    - Integrate with Kremer et al., 2017 (non-strand specific fibroblasts) and Murdock et al., 2020 (strand-specific blood and fibroblasts) datasets, available through DROP

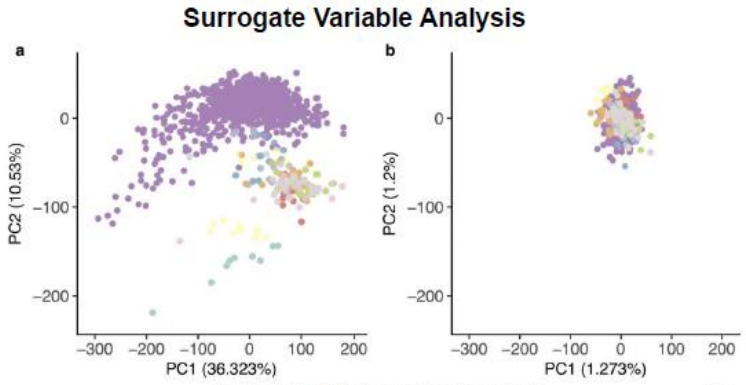
- **Datasets are noisy --> Correcting for sample co-variations is necessary**

Many variables are affecting gene expression; some of them are biological, such as sex, age, general health…

Other sources of noise are technical, such as sequencing batch, library size, technician, room temperature, RNA quality, …



Brechtmann *et al* AJHG 2018



Fresard *et al* Nature Medicine 2019

25

Combined samples must be from similar/same tissues, sequenced with similar protocol, reads aligned using the same aligner and parameters.

3.4 Dataset design – limitations of RNA-seq

General considerations

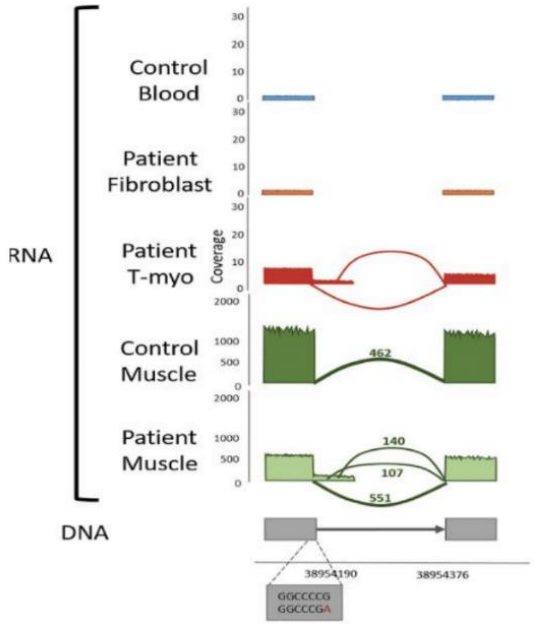
Is it possible to combine RNA-seq samples with controls from other centers, technologies, or tissues?

- **Genes are expressed in a tissue specific manner**
  & the disease tissue is not always available

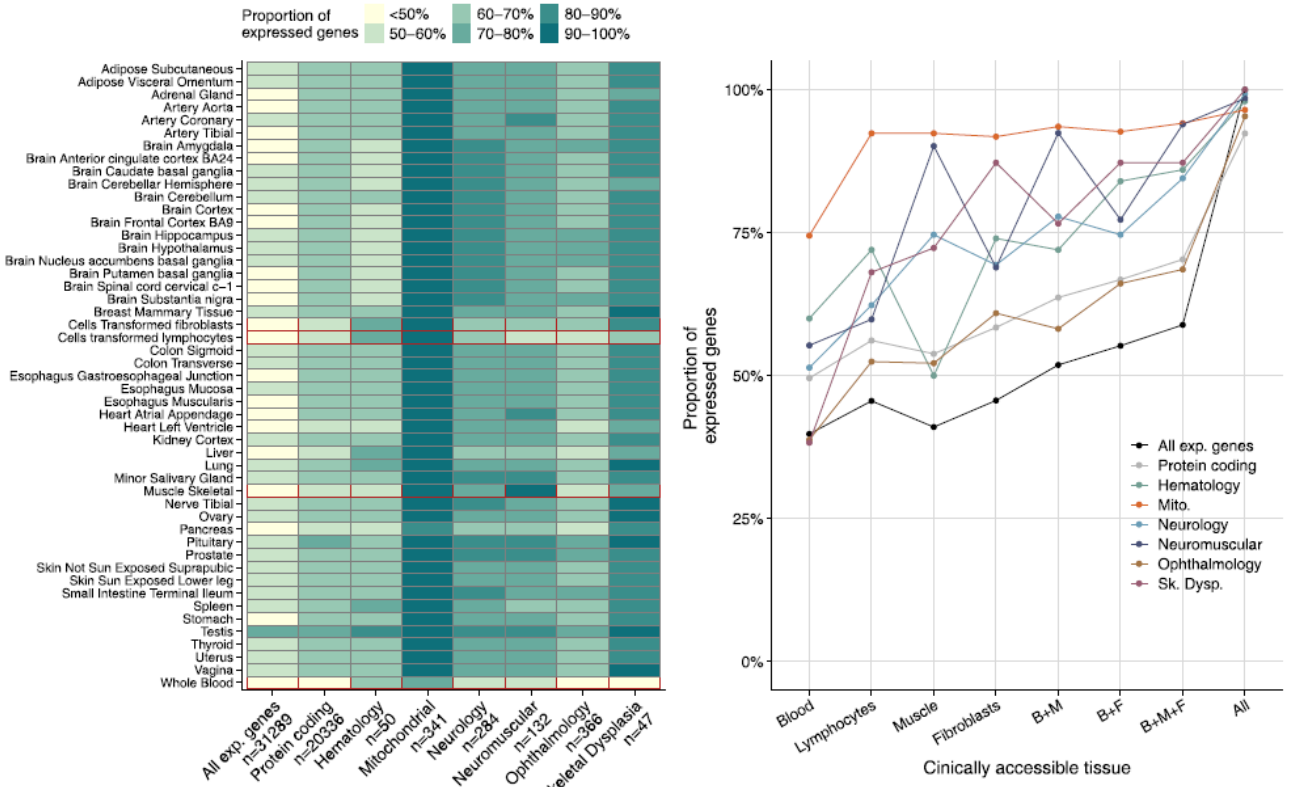
The affected tissue is not always accessible

Surrogate tissues are an alternative but causal genes might not be expressed

Availability of large expression datasets across tissues/cell lines (GTEx, i2QTL, BIOS) is of great importance to better predict/confirm effects seen on patients



(Gonorazky HD et al., 2019)



(Yépez VA et al., 2022)

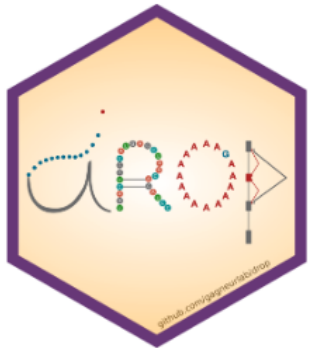
Documentación

Detection of RNA Outlier Pipeline

Build passing | release v1.2.2 | docs passing

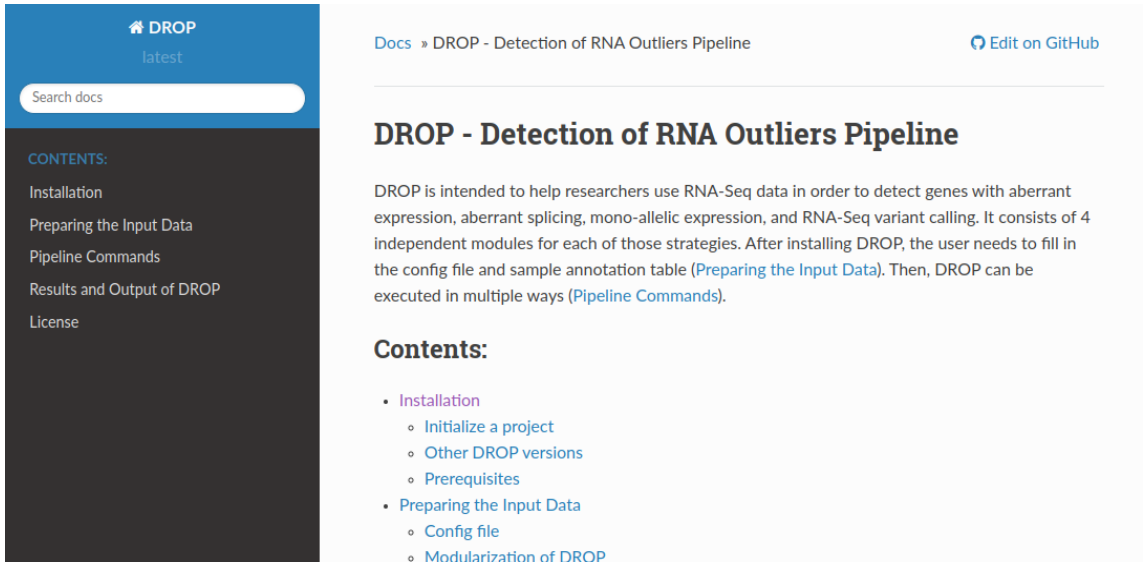
The detection of RNA Outliers Pipeline (DROP) is an integrative workflow to detect aberrant expression, aberrant splicing, and mono-allelic expression from raw sequencing files.

The manuscript is available in Nature Protocols. SharedIt link.



https://github.com/gagneurlab/drop

https://gagneurlab-drop.readthedocs.io/



Instalación

Instalable a través de los repositorios de bioconda:


```
mamba create \
        -n drop_env \
        -c conda-forge \
        -c bioconda \
        drop \
        --override-channels
```

Preparación de la Pipeline

Preparación de la Pipeline
Input

Yépez et al, 2021

Preparación de la Pipeline
Input - Configuración

```
projectTitle: Detection of RNA Outlier Pipeline
htmlOutputPath: /home/groups/funcgen/mdabad/test/Output/html
indexWithFolderName: true
root: /home/groups/funcgen/mdabad/test/Output
sampleAnnotation: /home/groups/funcgen/mdabad/test/Data/sample_annotation.tsv
geneAnnotation:
  v29: /home/groups/funcgen/mdabad/test/Data/gencode_annotation_trunc.gtf
genomeAssembly: hg19
genome:
  ncbi: /home/groups/funcgen/mdabad/test/Data/chr21_ncbi.fa
  ucsc: /home/groups/funcgen/mdabad/test/Data/chr21.fa
hpoFile: null
random_seed: true
exportCounts:
  geneAnnotations:
  - v29
  excludeGroups:
  - mae
  - outrider_external
  - fraser_external
```

Preparación de la Pipeline
Input - Configuración

```
aberrantExpression:
  run: true
  groups:
  - outrider
  - outrider_external
  fpkmCutoff: 1
  implementation: autoencoder
  padjCutoff: 1
  zScoreCutoff: 0
  maxTestedDimensionProportion: 3
  dassie:
    tssWindow: 500
    pasWindow: 1000
aberrantSplicing:
  run: true
  groups:
  - fraser
  - fraser_external
  recount: true
  longRead: false
  keepNonStandardChrs: false
  filter: false
  minExpressionInOneSample: 20
  minDeltaPsi: 0.05
  implementation: PCA
  padjCutoff: 1
  zScoreCutoff: 0
  deltaPsiCutoff: 0.05
  maxTestedDimensionProportion: 6
```

```
mae:
  run: true
  groups:
  - mae
  gatkIgnoreHeaderCheck: true
  padjCutoff: 0.5
  allelicRatioCutoff: 0.7
  addAF: false
  maxAF: 0.001
  maxVarFreqCohort: 1
  qcVcf: /home/groups/funcgen/mdabad/test/Data/qc_vcf_1000G.vcf.gz
  qcGroups:
  - mae
```

Preparación de la Pipeline
Input - Anotación de las muestras

A	B	C	D
RNA_ID	RNA_BAM_FILE	DNA_VCF_FILE	DNA_ID
HG00096	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00096_ncbi.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21_ncbi.vcf.gz	HG00096
HG00103	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00103.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21.vcf.gz	HG00103
HG00106	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00106.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21.vcf.gz	HG00106
HG00111	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00111.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21.vcf.gz	HG00111
HG00116	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00116.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21.vcf.gz	HG00116
HG00126	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00126.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21.vcf.gz	HG00126
HG00132	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00132.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21.vcf.gz	HG00132
HG00149	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00149.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21.vcf.gz	HG00149
HG00150	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00150.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21.vcf.gz	HG00150
HG00176	/home/groups/funcgen/mdabad/test/Data/rna_bam/HG00176.bam	/home/groups/funcgen/mdabad/test/Data/dna_vcf/demo_chr21.vcf.gz	HG00176
HG00178			
HG00181			
HG00191			
HG00201			

E	F	G	H	I	J
DROP_GROUP	PAIRED_END	COUNT_MODE	COUNT_OVERLAPS	STRAND	HPO_TERMS
outrider,fraser,mae,batch_0	True	IntersectionStrict	True	no	HP:0009802,HP:0010896
outrider,fraser,mae,batch_1	True	IntersectionStrict	True	no	HP:0004582,HP:0031959
outrider,outrider_external,fraser,fraser_external,mae,batch_1	True	IntersectionStrict	True	no	HP:0002895,HP:0006731
outrider,outrider_external,fraser,fraser_external	True	IntersectionStrict	True	no	HP:0100491,HP:0100871
outrider,outrider_external,fraser,fraser_external	True	IntersectionStrict	True	no	HP:0030613,HP:0012767
outrider,outrider_external,fraser,fraser_external	True	IntersectionStrict	True	no	HP:0000290,HP:0000293
outrider,outrider_external,fraser,fraser_external	True	IntersectionStrict	True	no	HP:0006489,HP:0006490
outrider,outrider_external,fraser,fraser_external	True	IntersectionStrict	True	no	HP:0000014,HP:0000020,HP:0032663
outrider,outrider_external,fraser,fraser_external	True	IntersectionStrict	True	no	HP:0030809,HP:0006144
outrider,outrider_external,fraser,fraser_external	True	IntersectionStrict	True	no	HP:0005215,HP:0010234
outrider_external					
outrider_external					
fraser_external					
fraser_external					

RNAseq BAM

- Mismo ID quee el ID RNA
- STAR two-pass

VCF DNA

- Sample ID igual que DNA_ID
- Todas las anotaciones que puedas (VEP)
  - GNOMAD, CLINVAR, ...

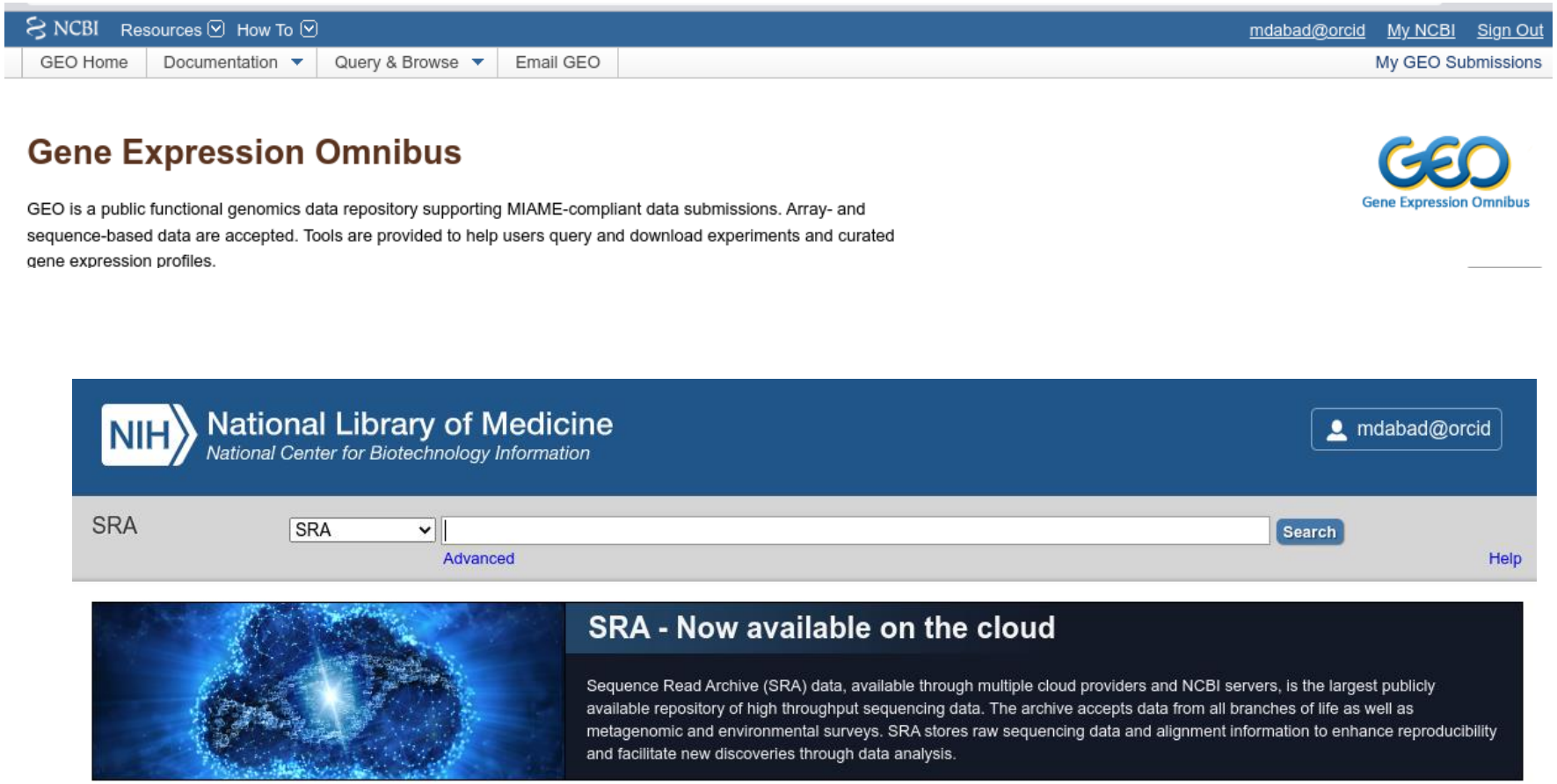
Datasets

The following publicly-available datasets of gene counts can be used as controls. Please cite as instructed for each dataset.

- 154 non-strand specific fibroblasts, build hg19, Technical University of Munich: DOI 10.5281/zenodo.4646822
- 269 strand specific fibroblasts, build hg19, Technical University of Munich: DOI 10.5281/zenodo.4646826
- 49 tissues, each containing hundreds of samples, non-strand specific, build hg19, GTEx: DOI 10.5281/zenodo.5596755
- 49 tissues, each containing hundreds of samples, non-strand specific, build hg38, GTEx: DOI 10.5281/zenodo.6078396
- 139 strand specific fibroblasts, build hg19, Baylor College of Medicine: DOI 10.5281/zenodo.3963473
- 125 strand specific blood, build hg19, Baylor College of Medicine: DOI 10.5281/zenodo.3963470

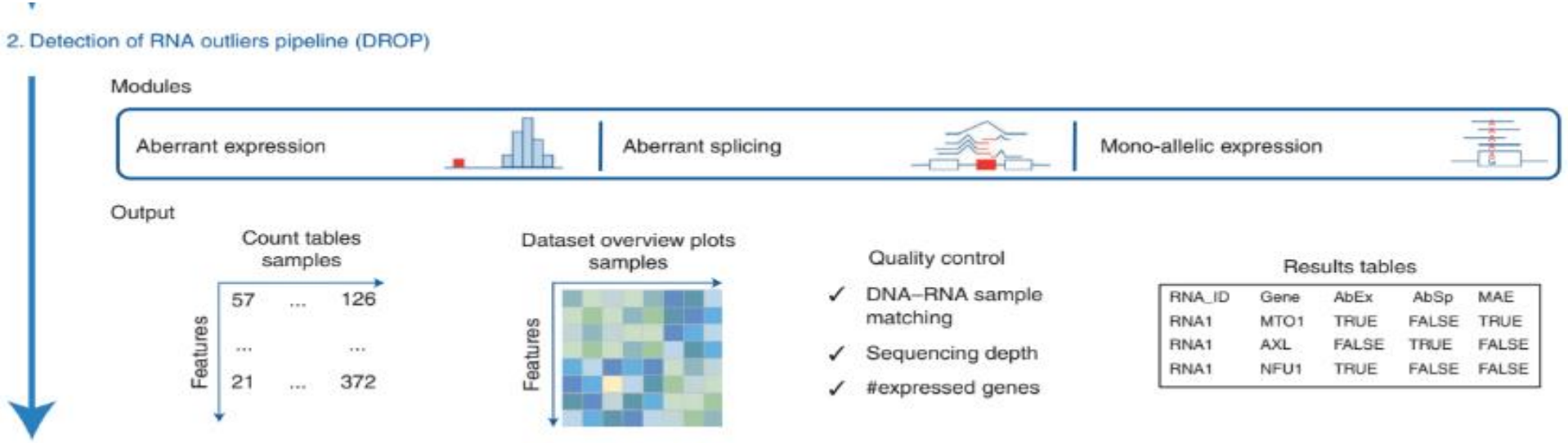
If you want to contribute with your own count matrices, please contact us: yepez at in.tum.de

Preparación de la Pipeline
Input – Datos externos



- Si son datos de secuenciación:
  - Mismo protocolo
  - Misma tecnología
  - Misma *strandness*
- Si son matrices de cuantificación:
  - Mismo genoma y anotación

Ejecución



Yépez et al, 2021

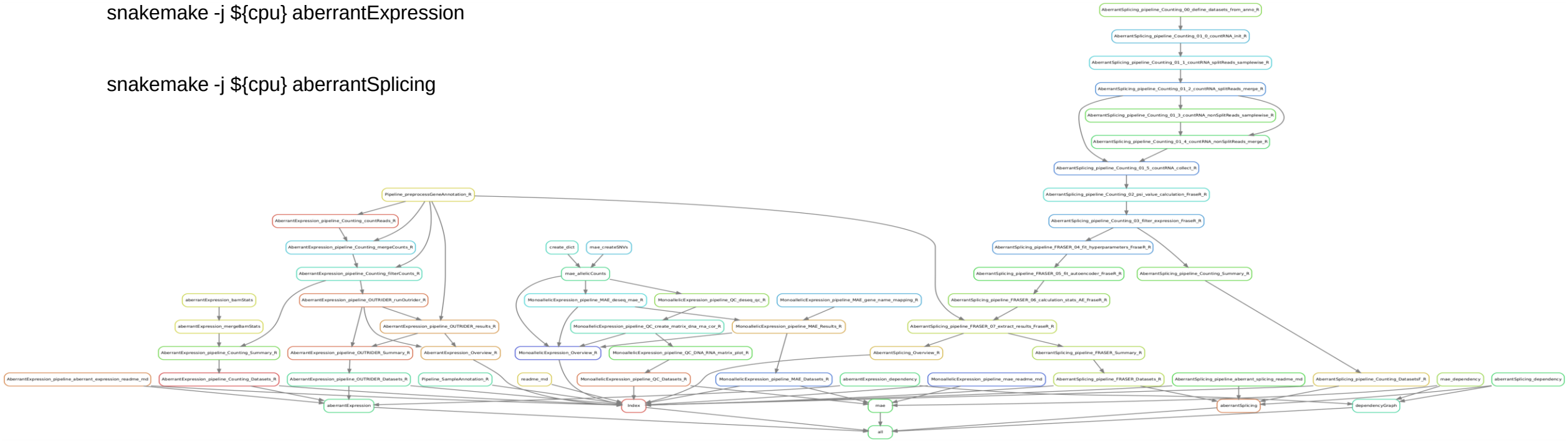
Ejecución

snakemake -j ${cpu} all

snakemake -j ${cpu} mae

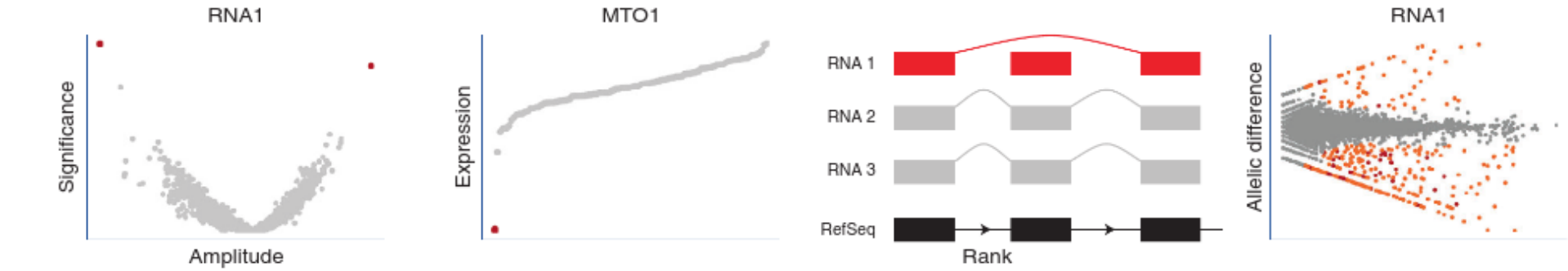
snakemake -j ${cpu} aberrantExpression

snakemake -j ${cpu} aberrantSplicing



Resultados

3. analyze individual results



Proyectos con DROP

>20K pacientes

A fibroblasts' transcriptome cohort experience

Inherited Metabolic Disorders Group (Hospital Clinic – IDIBAPS – CIBERER)

IDIBAPS Institut D'Investigacions Biomèdiques August Pi i Sunyer

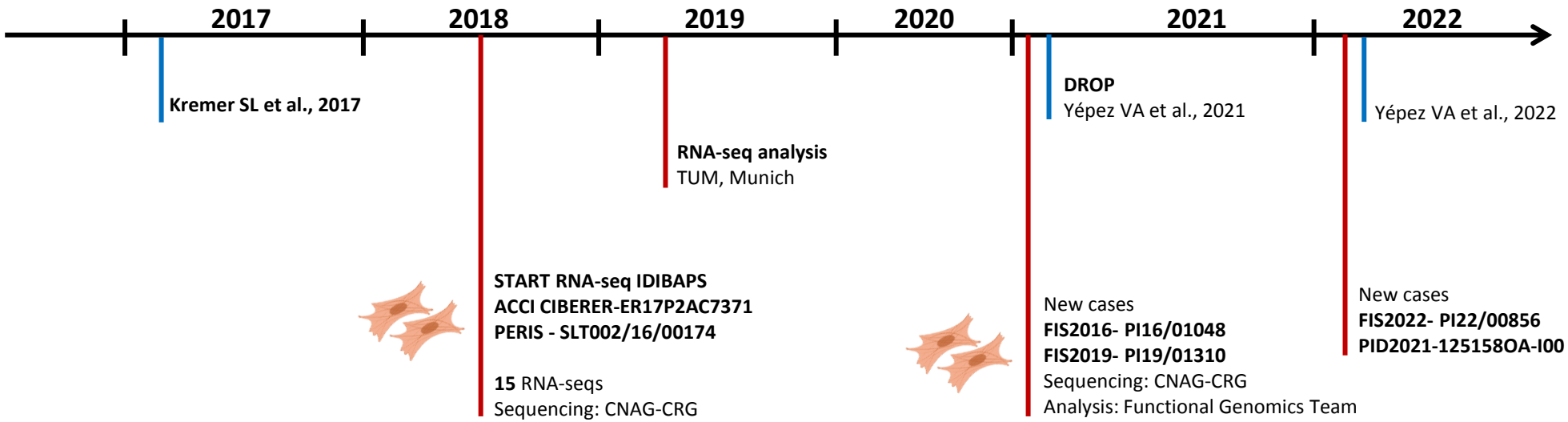
CLÍNIC BARCELONA Hospital Universitari


Research | Open Access | Published: 05 April 2022

Clinical implementation of RNA sequencing for Mendelian disease diagnostics

Vicente A. Yépez, Mirjana Gusic, Robert Kopajtich, Christian Mertes, Nicholas H. Smith, Charlotte L. Alston, Rui Ban, Skadi Beblo, Riccardo Berutti, Holger Blessing, Elżbieta Ciara, Felix Distelmaier, Peter Freisinger, Johannes Häberle, Susan J. Hayflick, Maja Hempel, Yulia S. Itkis, Yoshihito Kishita, Thomas Klopstock, Tatiana D. Krylova, Costanza Lamperti, Dominic Lenz, Christine Makowski, Signe Mosegaard, Michaela F. Müller, Gerard Muñoz-Pujol, Agnieszka Nadel, Akira Ohtake, Yasushi Okazaki, Elena Procopio, Thomas Schwarzmayr, Joël Smet, Christian Staufner, Sarah L. Stenton, Tim M. Strom, Caterina Terrile, Frederic Tort, Rudy Van Coster, Arnaud Vanlander, Matias Wagner, Manting Xu, Fang Fang, Daniele Ghezzi, Johannes A. Mayr, Dorota Piekutowska-Abramczuk, Antonia Ribes, Agnès Rötig, Robert W. Taylor, Saskia B. Wortmann, Kei Murayama, Thomas Meitinger, Julien Gagneur & Holger Prokisch — Show fewer authors

Genome Medicine **14**, Article number: 38 (2022) | Cite this article



RNA-seq cohort IDIBAPS	Solved	RNA-seq contribution	Candidate yield
45	**12**	**8**	**10**
(WES/WGS + RNA-seq)	(26%)	(17%)	

Case 2

SUMMARY:

CLINICAL PRESENTATION
- Female, 42 yo
- Retinitis pigmentosa
- Macular oedema
- Cataracts
- Hearing loss
- Abnormal facies
- Persistent mild-intensity neutropenia

GENOMIC STUDIES
- **WES analysis --> *PEX1*, c.1842del**
- OMIM #602136 PEROXISOME BIOGENESIS FACTOR 1; *PEX1*

TRANSCRIPTOMIC STUDIES
- **RNA-seq from fibroblasts --> *PEX1*, c.1240-1551A>G**

FUNCTIONAL STUDIES
- Pathogenicity validation of the rare deep intronic variant (splicing defect).

SOLVED CASE: *PEX1*, c.[1842del];[1240-1551A>G]

PEROXISOME BIOGENESIS FACTOR 1; PEX1

Alternative titles; symbols

PEROXIN 1

HGNC Approved Gene Symbol: PEX1

Cytogenetic location: 7q21.2 Genomic coordinates (GRCh38): 7:92,487,025-92,528,520 (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype (View Clinical Synopses)	Phenotype MIM number	Inheritance	Phenotype mapping key
7q21.2	Heimler syndrome 1	234580	AR	3
	Peroxisome biogenesis disorder 1A (Zellweger)	214100	AR	3
	Peroxisome biogenesis disorder 1B (NALD/IRD)	601539	AR	3

RNA-seq CONTRIBUTION:
- Variant impacts mRNA processing → AE → Functional evidence
- The second disease-causing variant fell within the non-coding region → variant prioritization challenge
- Second variant was not covered by WES but called in RNA-seq

International Journal of *Molecular Sciences*

MDPI

Article

Diagnostic Odyssey in an Adult Patient with Ophthalmologic Abnormalities and Hearing Loss: Contribution of RNA-Seq to the Diagnosis of a PEX1 Deficiency

Gerard Muñoz-Pujol [1], Socorro Alforja-Castiella [2], Ricardo Casaroli-Marano [2], Blai Morales-Romero [1], Judit García-Villoria [1], Vicente A. Yépez [3,4], Julien Gagneur [3,4], Mirjana Gusic [3,5], Holger Prokisch [3,5], Frederic Tort [1,*] and Antonia Ribes [1,*]

Case 3

SUMMARY:

CLINICAL PRESENTATION

- Leukoencephalopathy
- Cerebral dysmyelination
- Neuropathy
- Sensorineural hearing loss
- Nystagmus

GENOMIC STUDIES

- **WES analysis -->** Several candidates with homozygous variants in ROH, but inconclusive
- **WES reinspection after RNA-seq --> *UFM1*,** c.−273_−271delTCA

TRANSCRIPTOMIC STUDIES

- **RNA-seq from fibroblasts --> *UFM1* 50%** down-regulation

SOLVED CASE:

***UFM1*, c.[c.−273_−271delTCA];[c.−273_−271delTCA]**

* 610553

UBIQUITIN-FOLD MODIFIER 1; UFM1

HGNC Approved Gene Symbol: UFM1

Cytogenetic location: 13q13.3 *Genomic coordinates (GRCh38): 13:38,349,851-38,363,619* (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key
13q13.3	Leukodystrophy, hypomyelinating, 14	617899	AR	3

RNA-seq CONTRIBUTION:

- Variant impacts mRNA expression → AE → Functional evidence
- Reinspection of WES revealed an initially overseen 3-bp homozygous deletion in the promoter region (c.−273_−271delTCA).
- The variant was located in a poorly defined genomic region → variant prioritization challenge

Research | Open Access | Published: 05 April 2022

Clinical implementation of RNA sequencing for Mendelian disease diagnostics

Vicente A. Yépez, Mirjana Gusic, Robert Kopajtich, Christian Mertes, Nicholas H. Smith, Charlotte L. Alston, Rui Ban, Skadi Beblo, Riccardo Berutti, Holger Blessing, Elżbieta Ciara, Felix Distelmaier, Peter Freisinger, Johannes Häberle, Susan J. Hayflick, Maja Hempel, Yulia S. Itkis, Yoshihito Kishita, Thomas Klopstock, Tatiana D. Krylova, Costanza Lamperti, Dominic Lenz, Christine Makowski, Signe Mosegaard, Michaela F. Müller, Gerard Muñoz-Pujol, Agnieszka Nadel, Akira Ohtake, Yasushi Okazaki, Elena Procopio, Thomas [...]

Acknowledgements

IDIBAPS Institut D'Investigacions Biomèdiques August Pi i Sunyer

CLÍNIC BARCELONA Hospital Universitari

+ Hospitals col·laboradors

cnag centre nacional d'anàlisi genòmica centro nacional de análisis genómico

CRG Centre for Genomic Regulation

Functional Genomics Team

Anna Esteve-Codina

Beatriz Martín-Mur

Mariona Alberola

Marc Dabad

Funding agencies:

ciberer Centro de Investigación Biomédica en Red Enfermedades Raras

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Instituto de Salud Carlos III

FONDO EUROPEO DE DESARROLLO REGIONAL Una manera de hacer Europa

FONDO SOCIAL EUROPEO El FSE invierte en tu futuro

Unión Europea

AGAUR Agència de Gestió d'Ajuts Universitaris i de Recerca