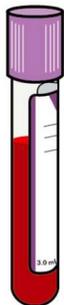


# Utilisation du datalake en RNASeq



# Diagnostic des TND



ADN

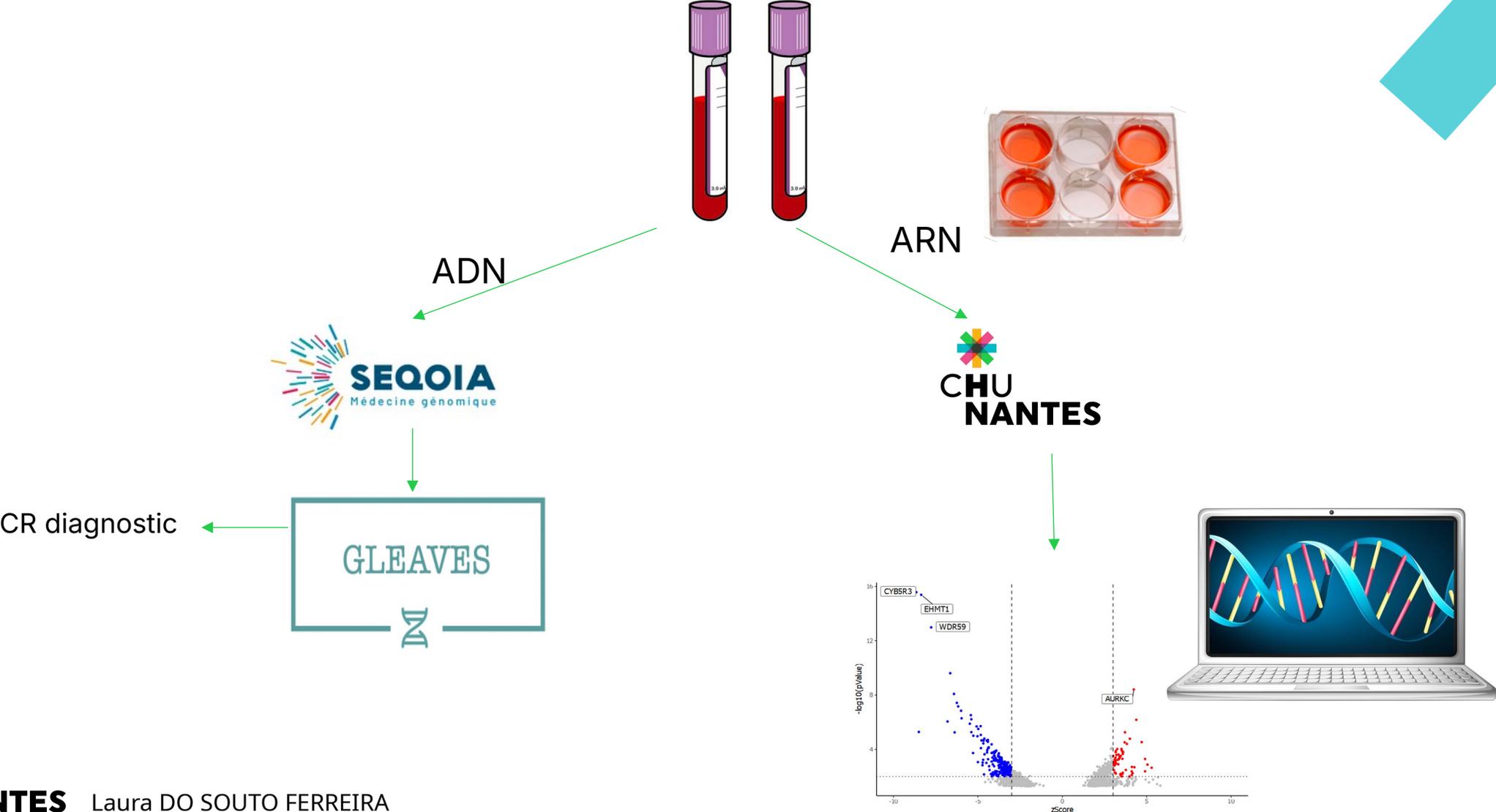


CR diagnostic

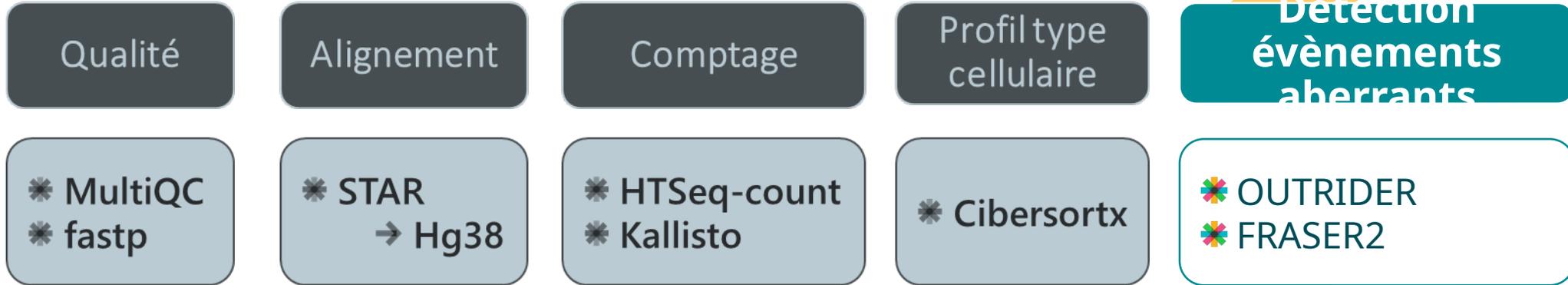


\* Benjamin Cogné  
Wallid Deb

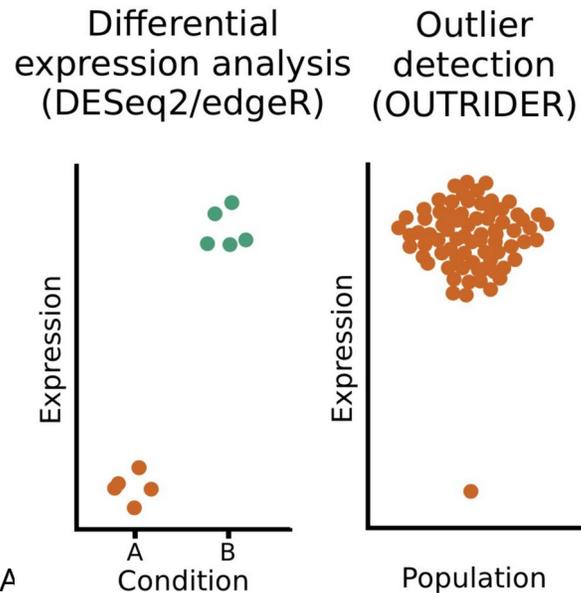
# Diagnostic des TND



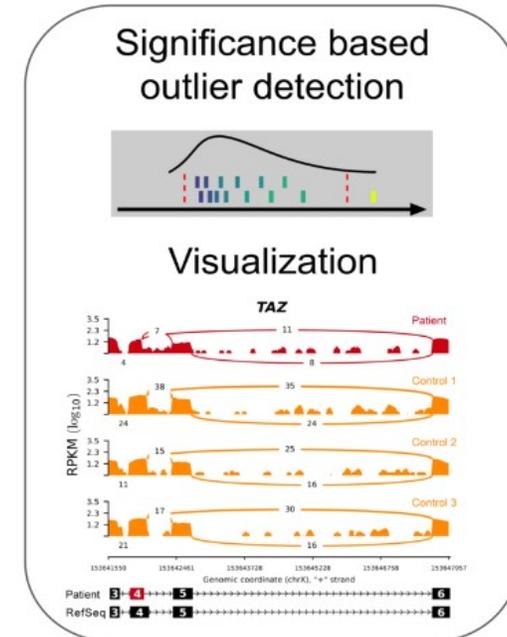
# Analyse RNASEQ



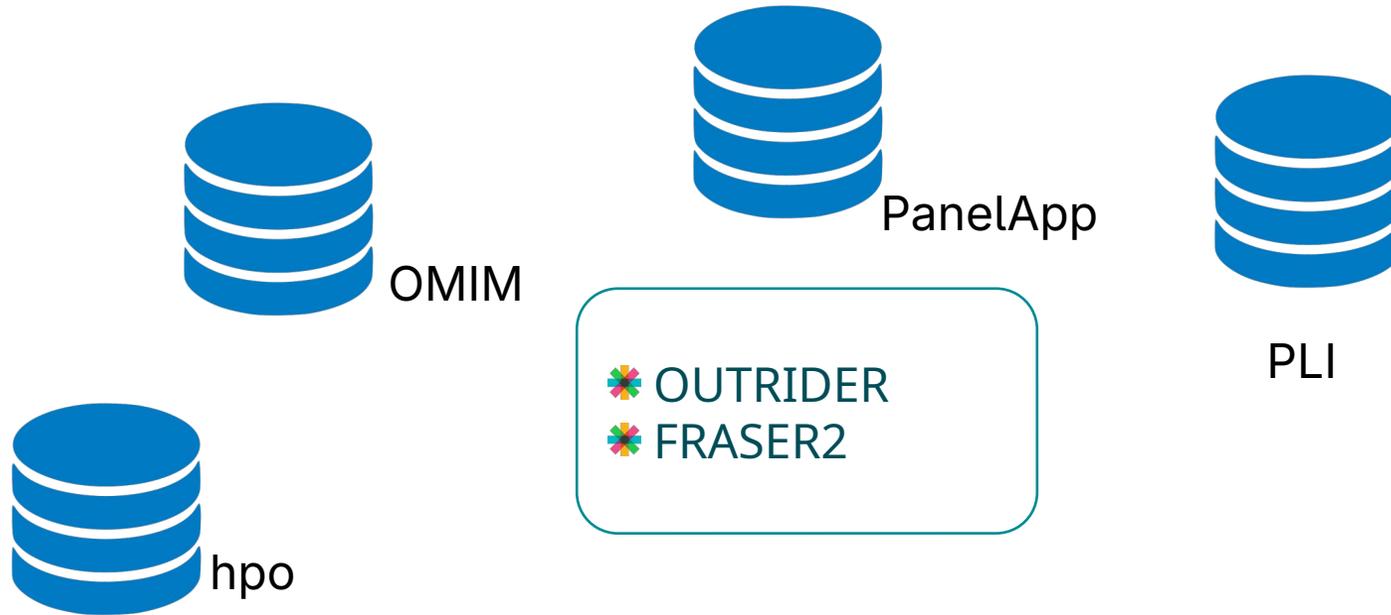
## OUTRIDER



## FRASER2

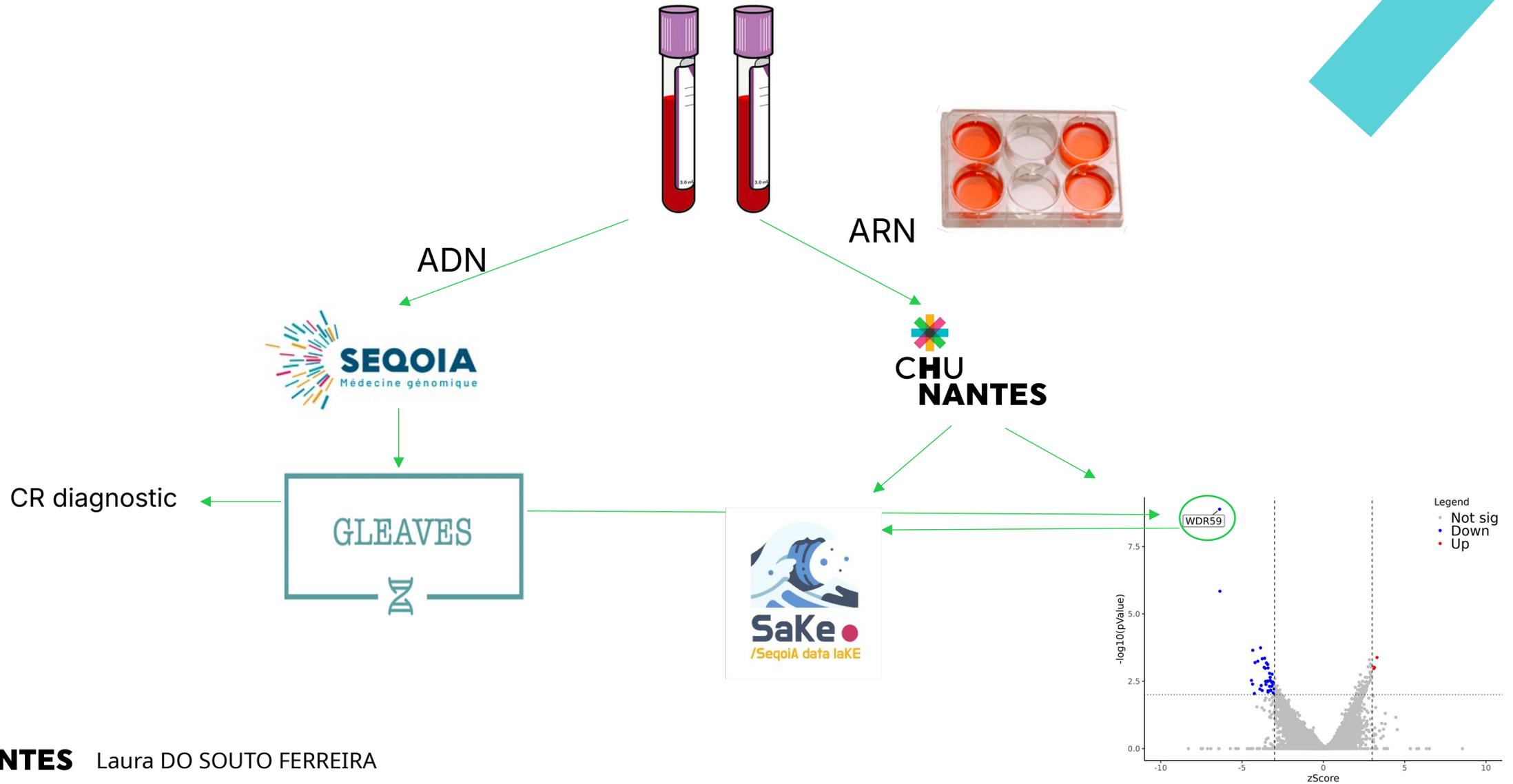


# Utilisation des bases de données

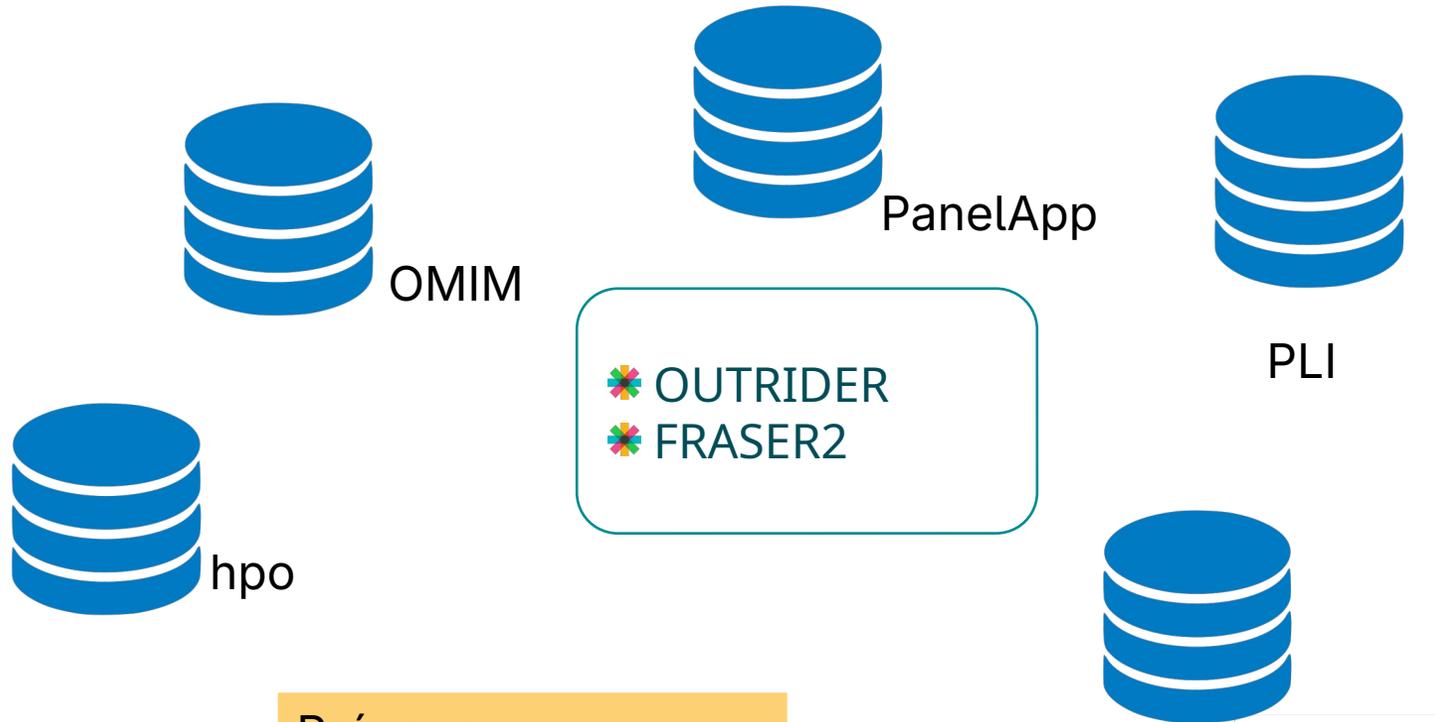


Présence ou  
absence dans l'autre  
outil

# Analyse RNASEQ



# Utilisation des bases de données

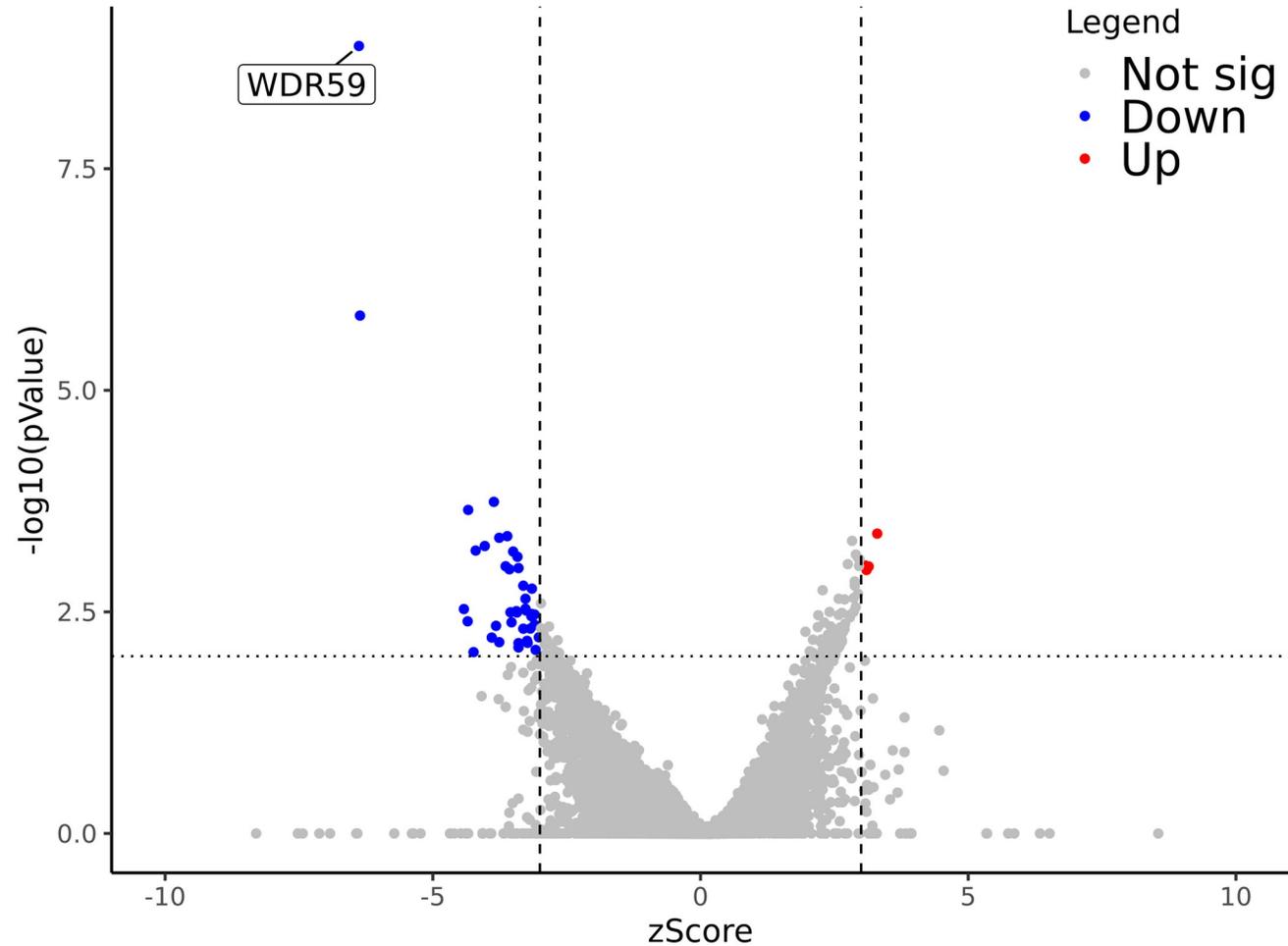


Présence ou absence dans l'autre outil



# Cas : WDR59

- Perte de fonction biallélique
- Autosomique récessif
- DI sévère
- Cardiopathie
- Présent dans une cohorte avec même phénotype



# Découpage des requêtes



Définition des paramètres

```
# define parameter
gnomad_ac_threshold = 3
sample_of_interest = ["23D30B"]
chromosome = ''
start = ''
end = ''
gene_name = "WDR59" # Set to None if you want use chromosome start end instead
```

# Découpage des requêtes



Définition des paramètres

Récupération des variants des données CHU Nantes et annotations snpeff

```
## Get chromosome scaffold
if not chromosome:
    df_chr = polars.read_csv("gene_to_chr.txt", separator="\t")
    chromosome = df_chr.filter(polars.col("Gene name") == gene_name).select("Chromosome/scaffold name").item()

# Select variant
if gene_name:
    # build snpeff query
    query = f"""
select
    snpeff.id,
    snpeff.impact as snpeff_impact, snpeff.hgvs_c as snpeff_hgvs_c, snpeff.hgvs_p as snpeff_hgvs_p,
    snpeff.gene as snpeff_gene, snpeff.geneid as snpeff_geneid
from
    read_parquet('{annotations_path} / "snpeff" / data_type / "v4.3t" / f"{chromosome}.parquet"') as snpeff
where
    snpeff.gene = '{gene_name}'
    and
    snpeff.impact in ('HIGH','MODERATE','LOW')
    """
```

# Découpage des requêtes



Définition des paramètres

Récupération des variants des données CHU Nantes et annotations snpeff

id	snpeff_impact	snpeff_hgvs_c	snpeff_hgvs_p	snpeff_gene	snpeff_geneid
2253622258503254021	MODERATE	c.1511A>G	p.Asn504Ser	WDR59	ENSG00000103091
2253572353130758148	MODERATE	c.2423G>T	p.Gly808Val	WDR59	ENSG00000103091
2253546536082341892	LOW	c.2763G>T	p.Thr921Thr	WDR59	ENSG00000103091
2253585484993265668	LOW	c.2058C>T	p.Leu686Leu	WDR59	ENSG00000103091

# Découpage des requêtes



Définition  
des  
paramètres

Récupération des  
variants des données  
CHU Nantes et  
annotations snpeff

Filtres :

- genotype
- dp
- gnomad\_a

id	snpeff_impact	snpeff_hgvs_c	snpeff_hgvs_p	snpeff_gene	snpeff_geneid	sample	gt	ad	dp	gq
2253572138382393348	LOW	c.2523C>T	p.Arg841Arg	WDR59	ENSG00000103091	5CBC1	1	10,30	40	
2253575295183355910	LOW	c.2347-8T>A		WDR59	ENSG00000103091	5CBC1	1	24,32	56	
2253572138382393348	LOW	c.2523C>T	p.Arg841Arg	WDR59	ENSG00000103091	15452	1	18,24	42	
2253576250813579271	LOW	c.2307T>C	p.Pro769Pro	WDR59	ENSG00000103091	15452	1	22,24	47	
2253784058511228932	LOW	c.42C>T	p.Phe14Phe	WDR59	ENSG00000103091	15452	1	28,33	62	

# Découpage des requêtes



Définition des paramètres

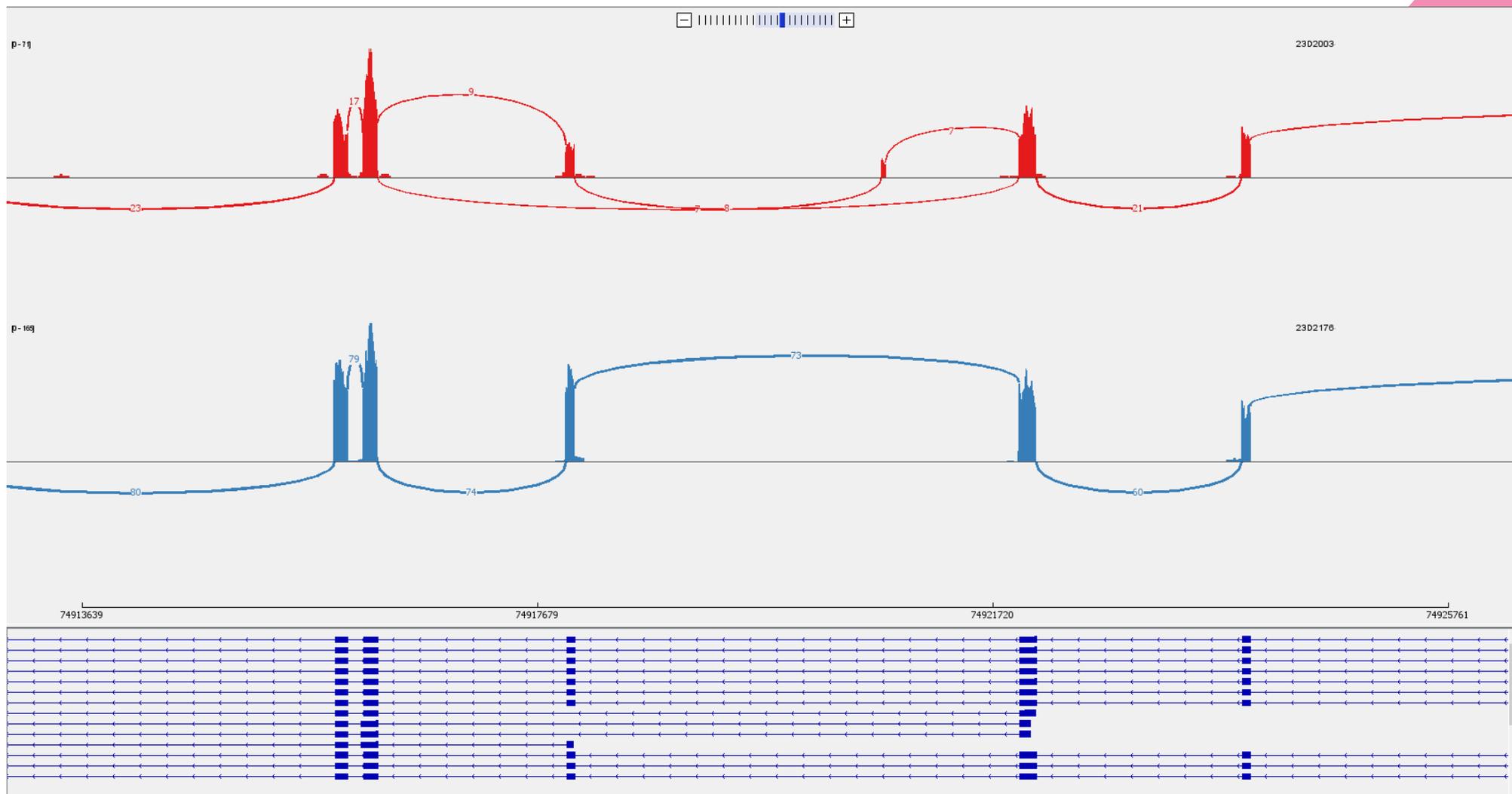
Récupération des variants des données CHU Nantes et annotations snpeff

Filtres :  
▪ genotype  
▪ dp  
▪ gnomad\_a

Filtres :  
▪ CIX  
▪ apparentés

id	snpeff_impact	snpeff_hgvs_c	snpeff_hgvs_p	snpeff_gene	snpeff_geneid	sample	gt	ad	dp	gq
2253575295183355910	LOW	c.2347-8T>A		WDR59	ENSG00000103091	23D30	1	32,19	51	99
2253640069732630533	LOW	c.966+3A>G		WDR59	ENSG00000103091	23D30	1	36,34	71	99
2253572138382393348	LOW	c.2523C>T	p.Arg841Arg	WDR59	ENSG00000103091	23D31	1	28,24	52	99
2253575295183355910	LOW	c.2347-8T>A		WDR59	ENSG00000103091	23D31	1	20,17	38	99
2253640069732630533	LOW	c.966+3A>G		WDR59	ENSG00000103091	23D31	1	16,29	45	99
2253575295183355910	LOW	c.2347-8T>A		WDR59	ENSG00000103091	23D32	1	27,25	55	99

# Résultats FRASER2



CIX

Témoi  
n

# Requête sur région



Définition  
des  
paramètres

Récupération des  
variants des données  
CHU Nantes et  
annotations snpeff

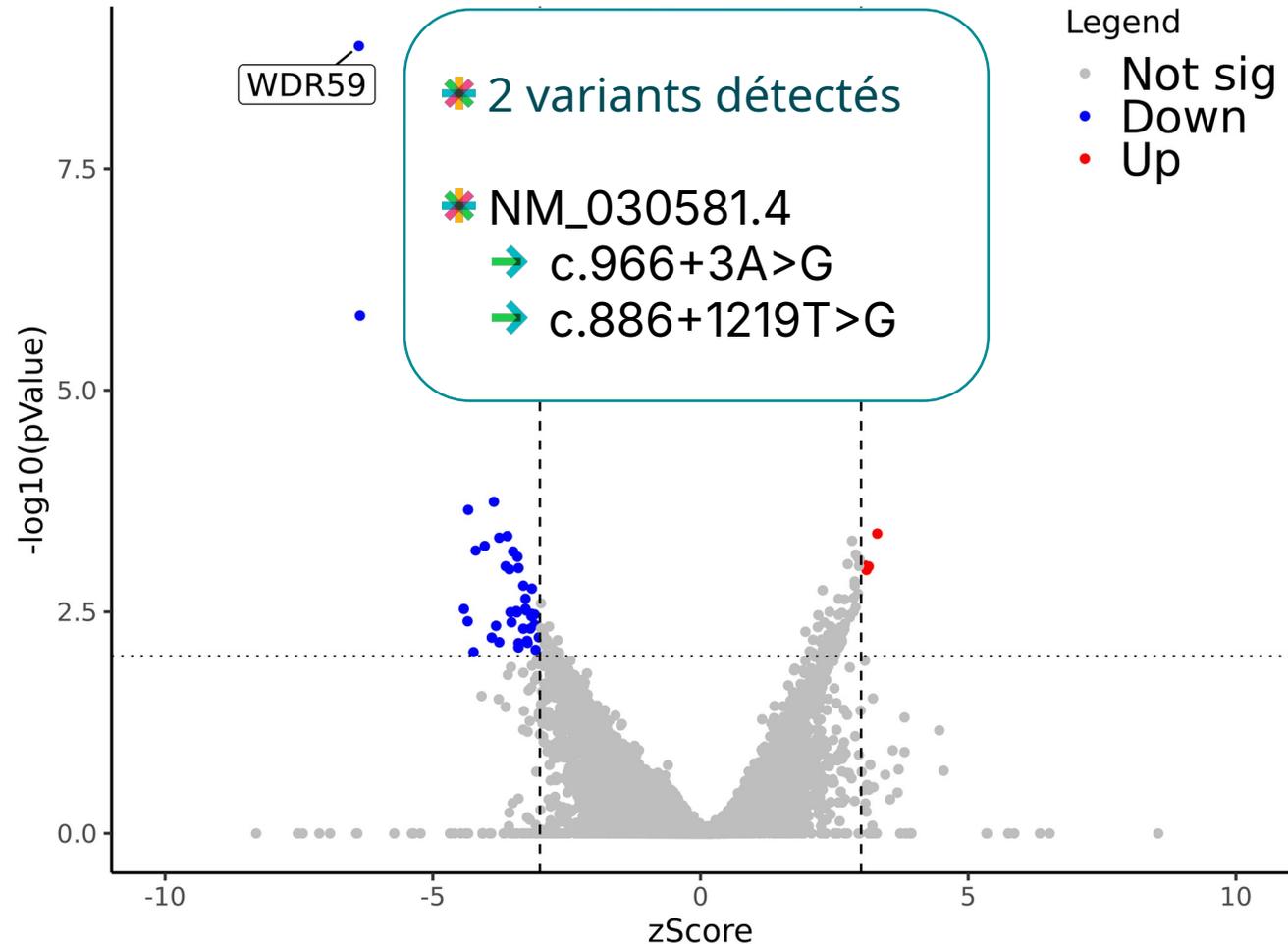
Filtres :  
▪ genotype  
▪ dp  
▪ gnomad\_a

Filtres :  
▪ CIX  
▪ apparentés

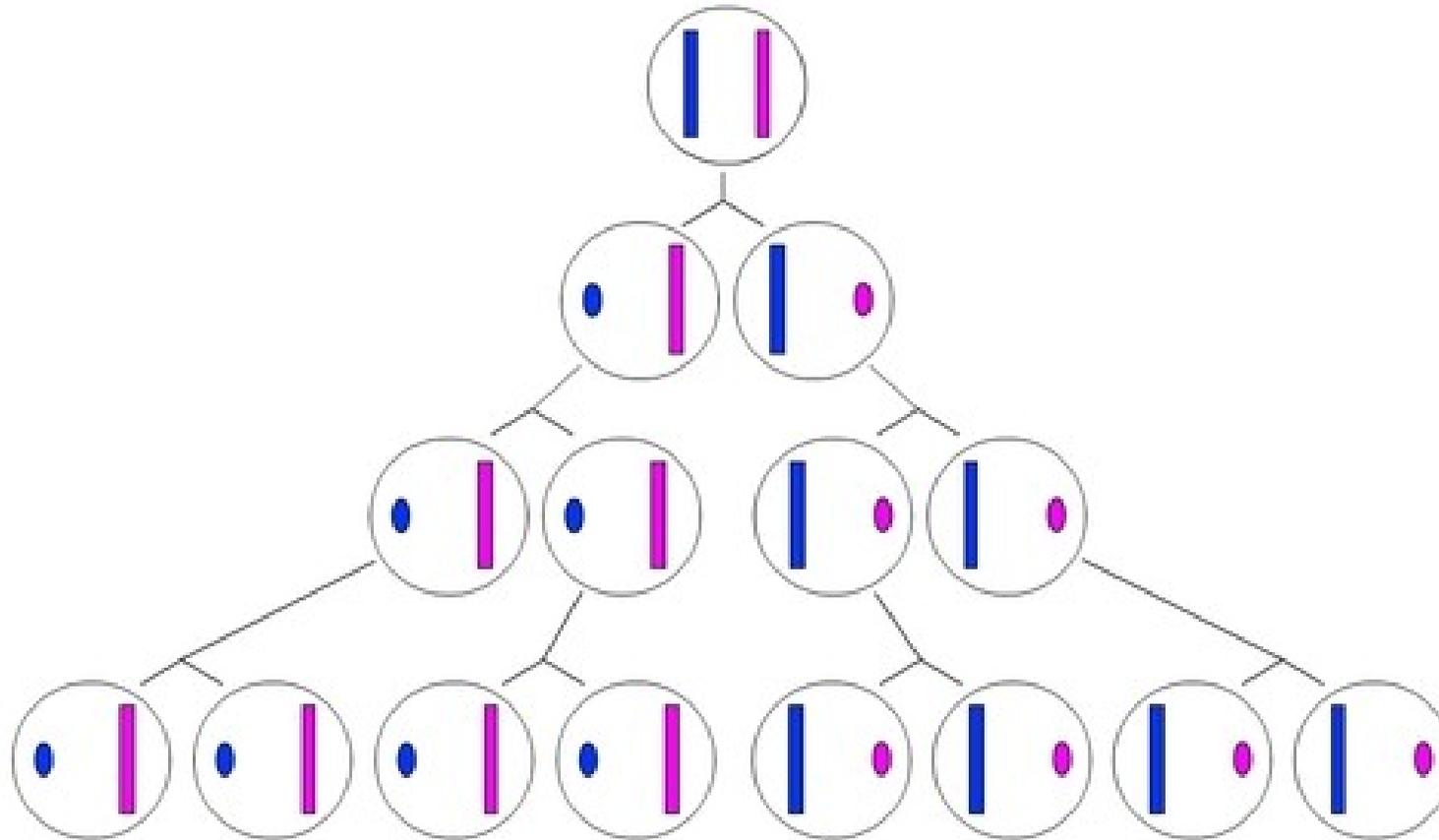
id	chr	pos	ref	alt	sample	gt	ad	dp	gq
2253644669642604553	16	74920068	CA	C	23D30	1	0,12	25	99
2253645371869757444	16	74920395	T	A	23D32	1	21,26	48	99
2253645371869757444	16	74920395	T	A	23D31	1	19,22	41	99
2253646086981812229	16	74920728	A	C	23D32	1	25,22	48	99
2253646086981812229	16	74920728	A	C	23D30	1	35,26	61	99

# Cas : WDR59 – 2 variants candidats

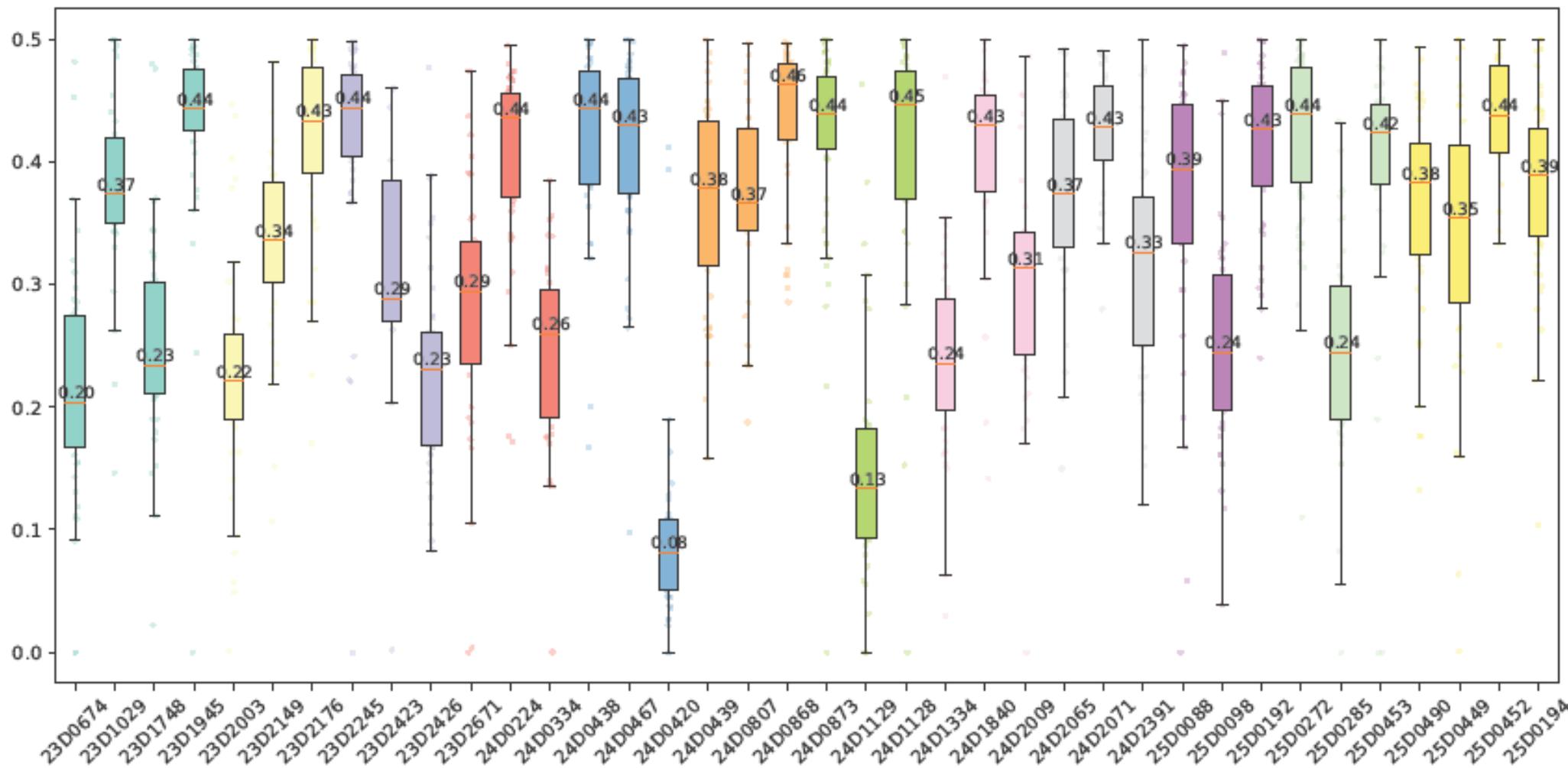
- Perte de fonction biallélique
- Autosomique récessif
- DI sévère
- Cardiopathie
- Présent dans une cohorte avec même phénotype



# *Biais : inactivation de l'X*

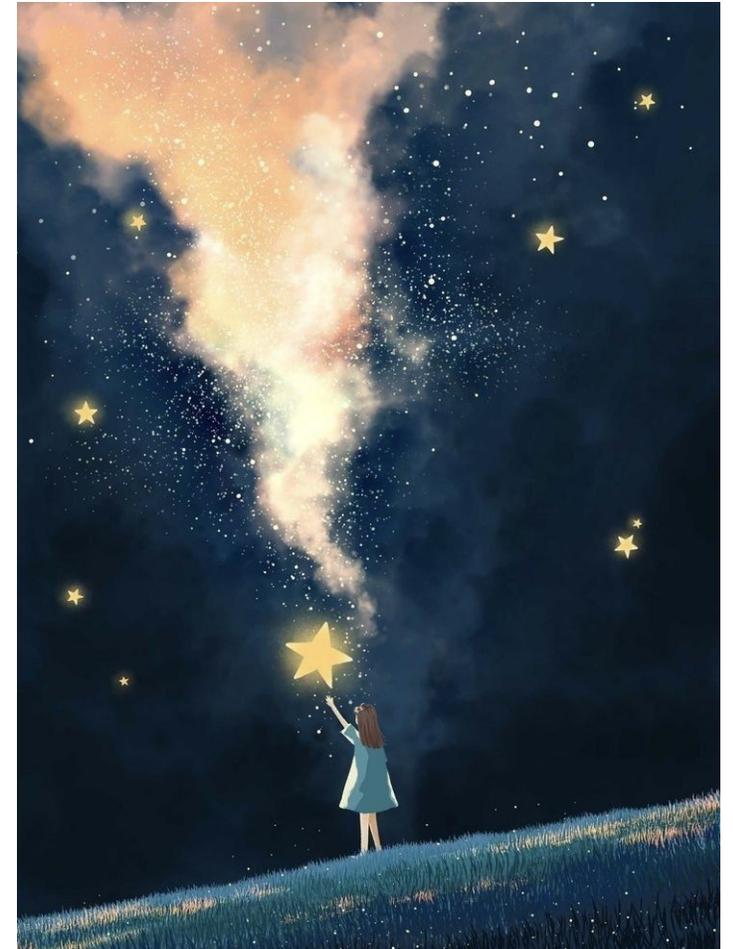


# Biais : inactivation de l'X



# Conclusion

- ✿ Sake facilite l'interprétation des données de RNASeq en mettant en parallèle les données génomiques
- ✿ Permet la recherche de variants dans des régions d'intérêt



# Remerciements



Thomas Besnard  
Patricia Talarmain  
Delphine Quinquis  
Jérôme Buscail  
Gaëlle Landeau  
Eva Trochu  
Jessy Hary  
Wallid Deb  
Stéphane Bézieau  
Benjamin Cogné

Bertrand Isidor  
Sandra Mercier  
Mathilde Nizon  
Marie Vincent  
Solène Conrad



Virginie Vignard  
Frédéric Ebstein  
Sébastien Kury



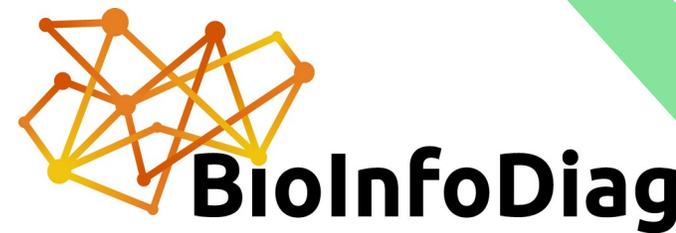
Amandine Charreton  
Emmanuelle Lecommandeur  
Sylvie Odent  
Et tous les cliniciens et  
biologistes collaborateurs



Pierre Blanc  
Alban Lermine  
Pierre Marijon  
Et leurs équipes



Laura DO SOUTO FERREIRA



Merci pour votre attention !

