TAGC — Implementation of whole genome longreads sequencing to decode somatic genotypes in T-ALL

30 October 2024

Dr. Thomas Steimlé (MD)

— Thèse co-dirigée par Pr. Vahid Asnafi (PU-PH Necker) et Dr. Salvatore Spicuglia (DR1 Inserm)

Necker ENFANTS MALADES HÔPITAL UNIVERSITAIRE

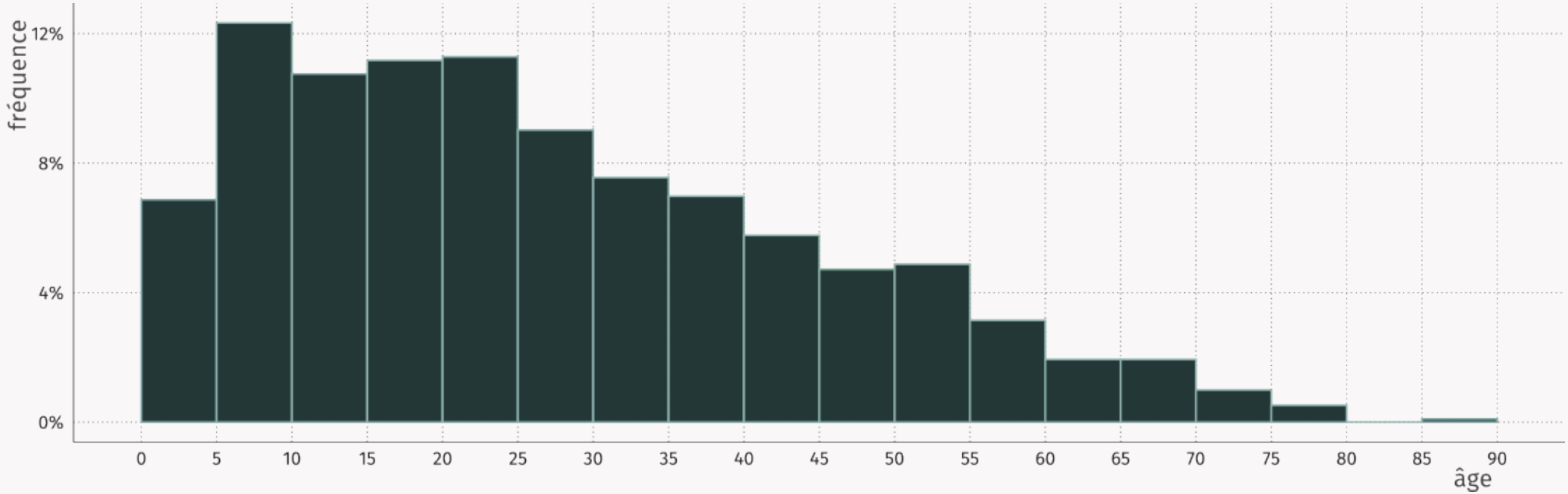
ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

❓ Main Hypothesis

Tumor proliferations harbor within their genome intergenic somatic mutations that disrupts the expression of oncogenes.

👨 Model — T-ALL

Acute lymphoblastic leukemia is a rare disease (≃ 120/yr in Fr) resulting from a tumoral process driven by the clonal **proliferation of immature T lymphocytes**.

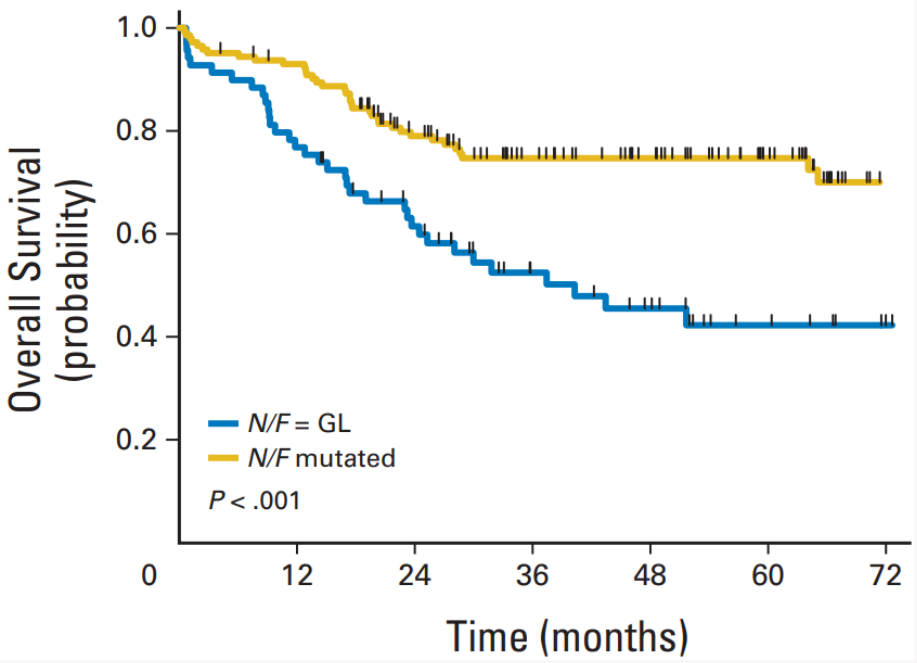


Internal unpublished data

👨‍⚕️ Model — T-ALL

Acute lymphoblastic leukemia is a rare disease (≃ 120/yr in Fr) resulting from a tumoral process driven by the clonal proliferation of **immature T lymphocytes**.

- Adults OS 3 years: 67% (GRAALL-2005)

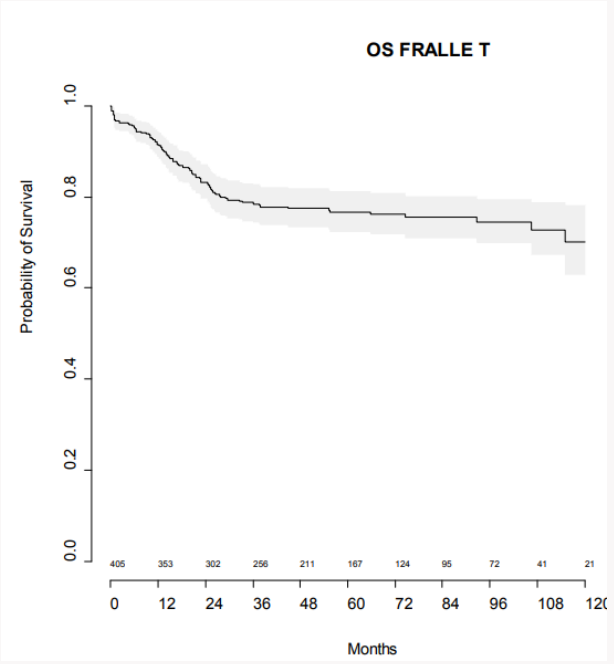


Trinquand A et al. J Clin Oncol. 2013;31(34):4333-4342

Model — T-ALL

Acute lymphoblastic leukemia is a rare disease (≃ 120/yr in Fr) resulting from a tumoral process driven by the clonal proliferation of **immature T lymphocytes**.

- Adults OS 3 years: 67% (GRAALL-2005)
- Youths OS 5 years: 77% (FRALLE)



Petit A et al. Blood. 2018;131(3):289-300.

👨‍⚕️ Model — T-ALL

Refractory cases to standard chemotherapy as well as **relapses** (UKALL12: adults at 5 years 42%) have a poor prognosis.

👨‍⚕️ Model — T-ALL

Refractory cases to standard chemotherapy as well as **relapses** (UKALL12: adults at 5 years 42%) have a poor prognosis.

We need new treatments to address these cases !

👨‍⚕️ Model — T-ALL

Refractory cases to standard chemotherapy as well as **relapses** (UKALL12: adults at 5 years 42%) have a poor prognosis.

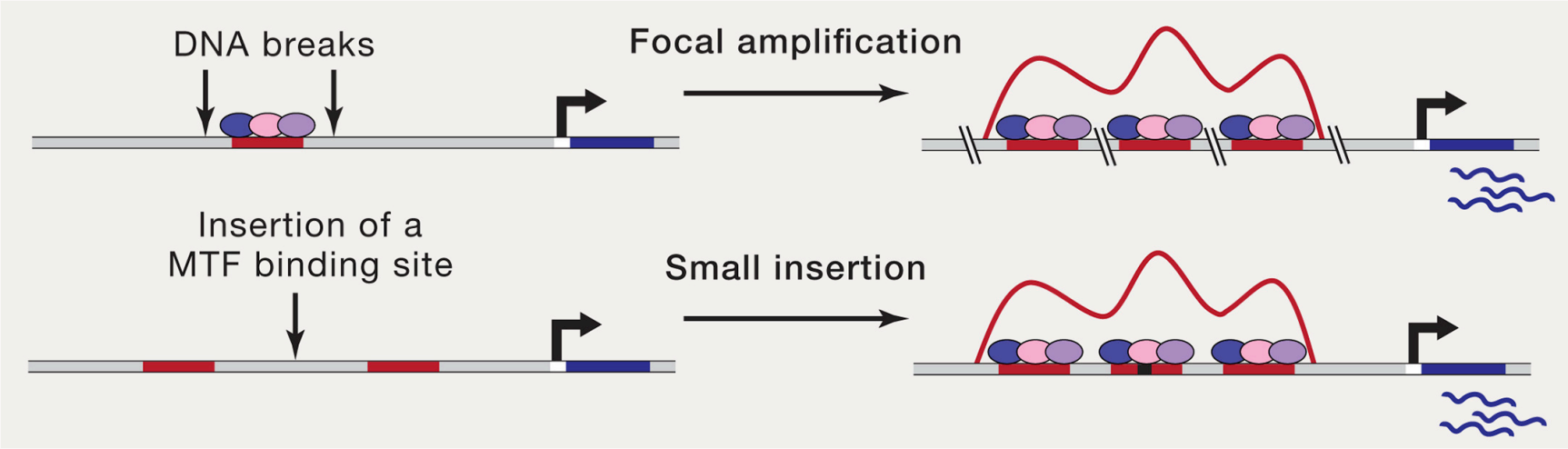
We need new treatments to address these cases !

The tumor phenotype emerges from alterations in their genotype.

⇨ A better description of the genotype will lead to a better understanding of oncogenic mechanisms and to the discovery of more **effective therapies tailored to specific alterations**.

Intergenic alterations

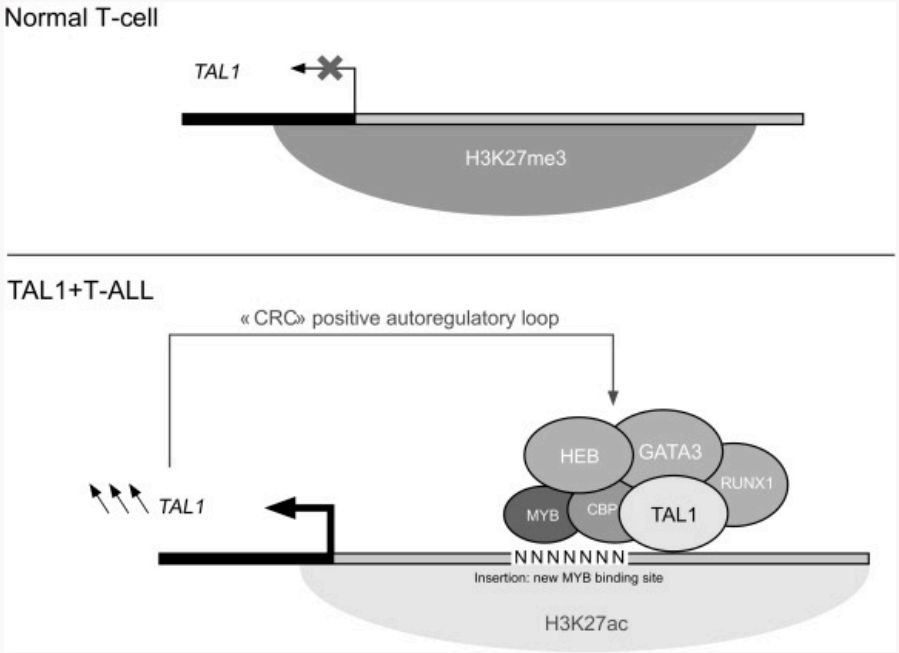
Somatic intergenic alterations are responsible of the deregulation of oncogenes.



Bradner JE, et al. Cancer. Cell. 2017 Feb9 ;168(4):629-643

Preliminary results

- Our laboratory has shown that a somatic insertion upstream of *TAL1* leads to the formation of a **neo-enhancer** and thus leads to the overexpression of *TAL1*.



Smith, C et al. "TAL1 activation in T-cell acute lymphoblastic leukemia: a novel oncogenic 3′ neo-enhancer." Haematologica vol. 108,5 1259-1271. 1 May. 2023

🎯 Goals

- **Implement** a method for sequencing the **whole tumoral genome**.
-
-
-

🎯 Goals

- **Implement** a method for sequencing the **whole tumoral genome**.
- **Detect** structural variations (SV) and SNV with good sensitivity/specificity.
-
-

🎯 Goals

- **Implement** a method for sequencing the **whole tumoral genome**.
- **Detect** structural variations (SV) and SNV with good sensitivity/specificity.
- **Describe** a set of somatic intergenic alterations likely responsible for the deregulation of known oncogenes.
-

🎯 Goals

- **Implement** a method for sequencing the **whole tumoral genome**.
- **Detect** structural variations (SV) and SNV with good sensitivity/specificity.
- **Describe** a set of somatic intergenic alterations likely responsible for the deregulation of known oncogenes.
- **Discover** similar alterations deregulating genes not known to be oncogenes.

🎯 Goals

- **Implement** a method for sequencing the **whole tumoral genome**.
- **Detect** structural variations (SV) and SNV with good sensitivity/specificity.
- **Describe** a set of somatic intergenic alterations likely responsible for the deregulation of known oncogenes.
- **Discover** similar alterations deregulating genes not known to be oncogenes.

⇨ Implementation of the **Oxford Nanopore sequencing method** and integrate multi-omics data.

Infrastucture

Computer A for processing and analyzing the signal generated by the sequencer (MAD).

- 2x Intel Xeon Platinum 8380 CPU 2.30GHz
- 160 cores
- 503 GB RAM
- 4 nVidia A100 GPU

Infrastucture

Computer A for processing and analyzing the signal generated by the sequencer (MAD).

- 2x Intel Xeon Platinum 8380 CPU 2.30GHz
- 160 cores
- 503 GB RAM
- 4 nVidia A100 GPU

Computer B for data archiving and sharing.

- 52 To fo HD in raidz1 mode (redundancy)
- LTO: magnetic tape drive (raw data archiving)

🧪 Wet lab

- Intermediate difficulty level.
- 3 half-days of work.

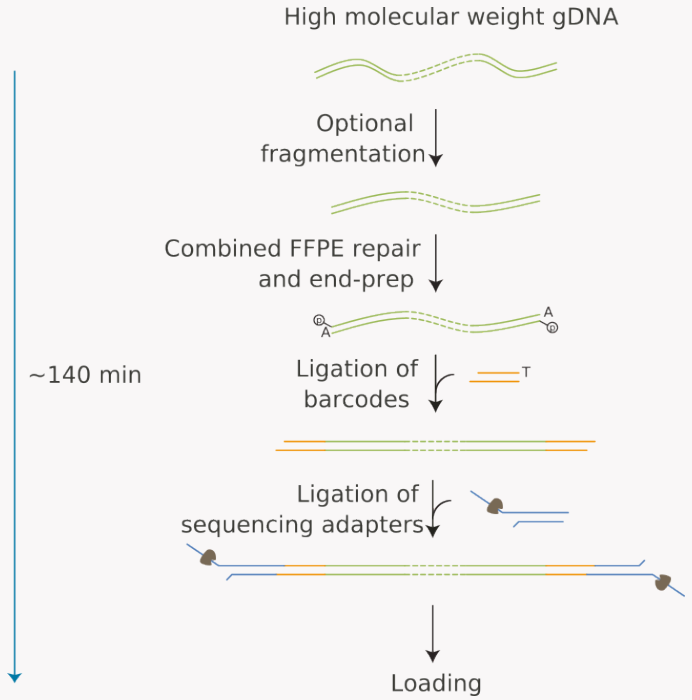
Wet lab

- Intermediate difficulty level.
- 3 half-days of work.

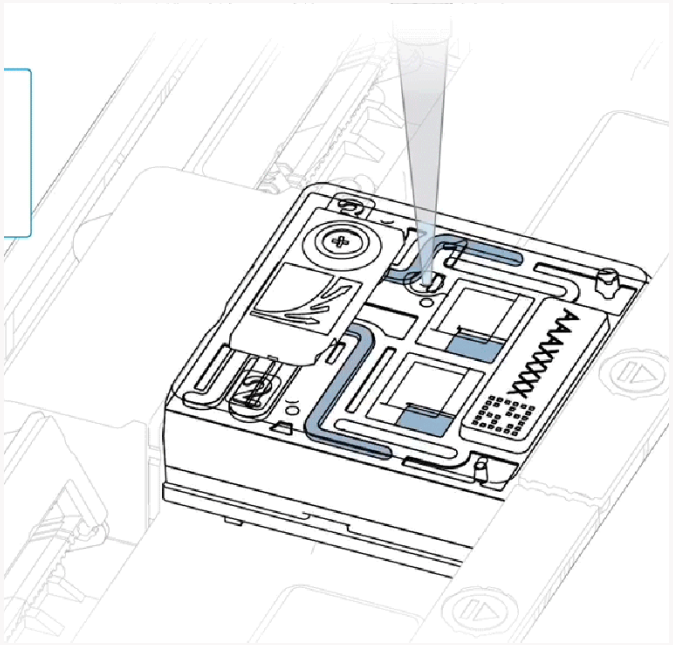
1. Genomic DNA **shearing** Covaris g-TUBE (3µg, 8,000rpm, 1min)

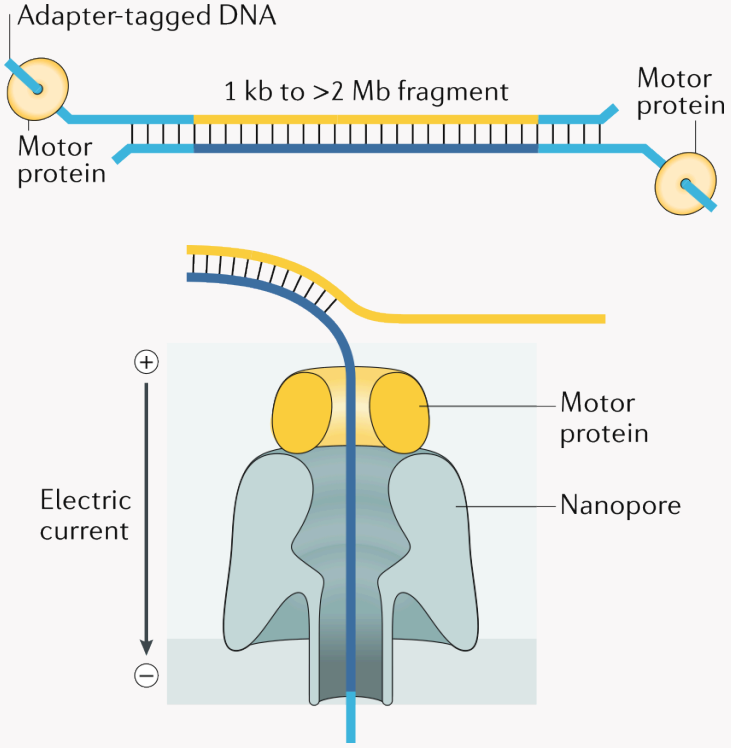
2. Preparation of the sequence library

DNA **repair** and end-prep ⇨ **barcode** ligation ⇨ **adapter** ligation



3. Flowcells **loading** and **sequencing run**

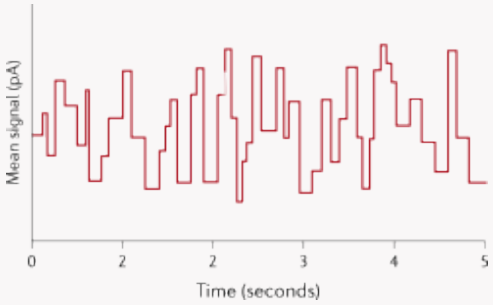




Bioinformatic Pipeline

1. **Base calling** and **Alignement** on hs1 (T2T).

Latest version of Dorado 0.8.1 with latest AI model of 5mC 5hmC modified basecalling (v5.0).



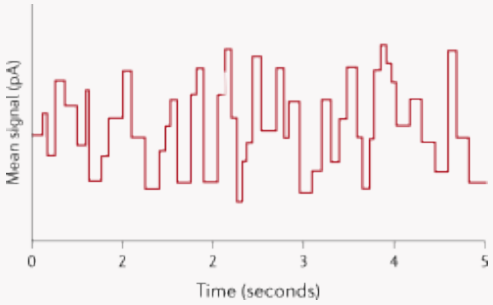
→

```
059b2db8-416c-4fdf-b7eb-3cdd3390a3bd 0 chr1 2 1
1075S5M1D8M1D36M1D4M1D116M1D111M1I485M3I595M...69M1I12M1I16M1D48M1641S * 0 0
AAGGTTAAAACCAAGACTCGCTGTGC 8;88;;<<<:::98...87877676645654456798997 qs:i:22 du:f:14.
              137 ns:i:70685 ts:i:904 mx:i:2 ch:i:1307 st:Z:2023-12-14T18:30:59.187+00:00
rn:i:36917 fn:Z:PAS34492_pass_barcode
              09_2e65ae3a_a7dfa7d7_4228.pod5 sm:f:-764.172 sd:f:0.00798761 sv:Z:pa dx:i:0
RG:Z:a7dfa7d727c0e04ecb6e9c5c3dac8080fe7cffaa_dna_r10.4.1_e8.2_400bps_sup@v4.3.0 MN:i:5392
MM:Z:C+h.,1,0,1,0,0,4,5,117,36,0,93,99,56,0,65...; ML:B:C,152,3,5,3,3,1,...; NM:i:1
              71 ms:i:4485 AS:i:4438 nn:i:0 de:f:0.0502045 tp:A:P cm:i:11 s1:i:82 s2:i:148
MD:Z:5^A8^C36^C4^A116^C...; rl:i
              :3576 SA:Z:chr8,146252402,-,1604M1D3788S,36,10;
```

Bioinformatic Pipeline

1. **Base calling** and **Alignement** on hs1 (T2T).

Latest version of Dorado 0.8.1 with latest AI model of 5mC 5hmC modified basecalling (v5.0).



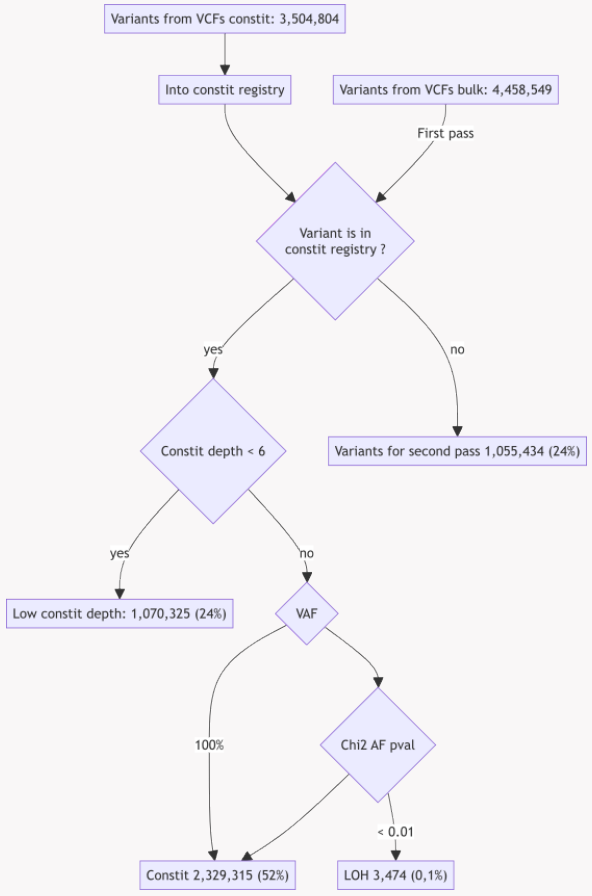
→

```
059b2db8-416c-4fdf-b7eb-3cdd3390a3bd 0 chr1 2 1
1075S5M1D8M1D36M1D4M1D116M1D111M1I485M3I595M...69M1I12M1I16M1D48M1641S * 0 0
AAGGTTAAAACCAAGACTCGCTGTGC 8;88;;<<<:::98...87877676645654456798997 qs:i:22 du:f:14.
                137 ns:i:70685 ts:i:904 mx:i:2 ch:i:1307 st:Z:2023-12-14T18:30:59.187+00:00
rn:i:36917 fn:Z:PAS34492_pass_barcode
                09_2e65ae3a_a7dfa7d7_4228.pod5 sm:f:-764.172 sd:f:0.00798761 sv:Z:pa dx:i:0
RG:Z:a7dfa7d727c0e04ecb6e9c5c3dac8080fe7cffaa_dna_r10.4.1_e8.2_400bps_sup@v4.3.0 MN:i:5392
MM:Z:C+h.,1,0,1,0,0,4,5,117,36,0,93,99,56,0,65...; ML:B:C,152,3,5,3,3,1,...; NM:i:1
                71 ms:i:4485 AS:i:4438 nn:i:0 de:f:0.0502045 tp:A:P cm:i:11 s1:i:82 s2:i:148
MD:Z:5^A8^C36^C4^A116^C...; rl:i
                :3576 SA:Z:chr8,146252402,-,1604M1D3788S,36,10;
```

2. **Variant calling**
   - DeepVariant (Google v1.6.1) on constit and tumoral BAMs.
   - ClairS (HK-UBAL v0.1.7) takes both constit and tumoral BAMs.
   - ~~Sniffles (Fritz Sedlazeck v2.2) on constit and tumoral BAMs.~~
   - NanoMonSV (Yuichi Shiraishi v0.7.2) takes both constit and tumoral BAMs.
   - Exogene (Z. Stephens v15) viral integration.

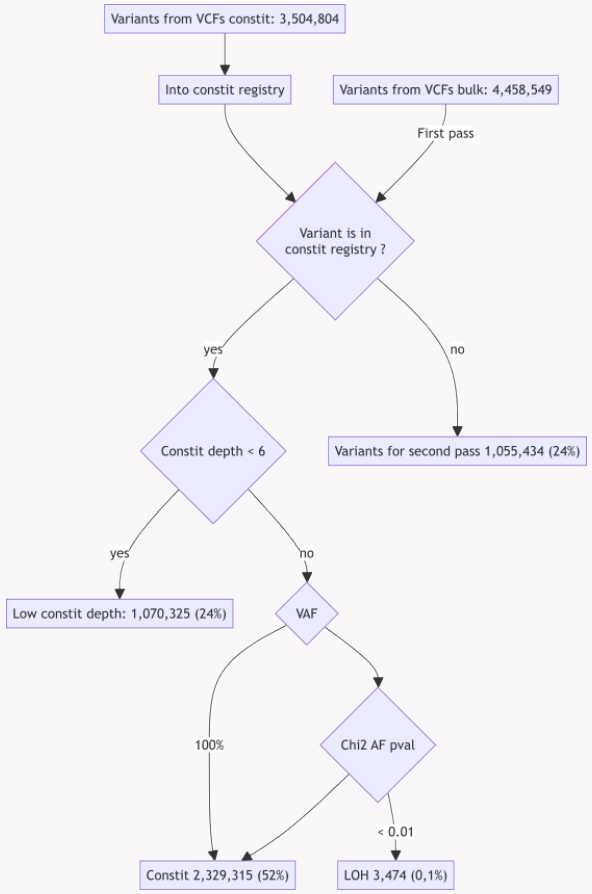
Bioinformatic Pipeline — Aggregation

1. VCF filter

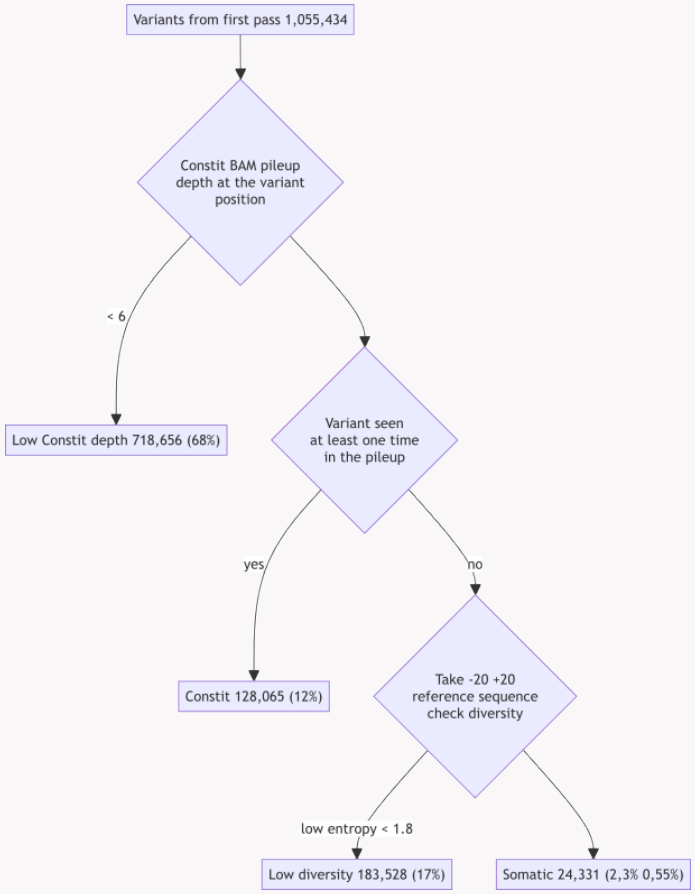


Bioinformatic Pipeline — Aggregation

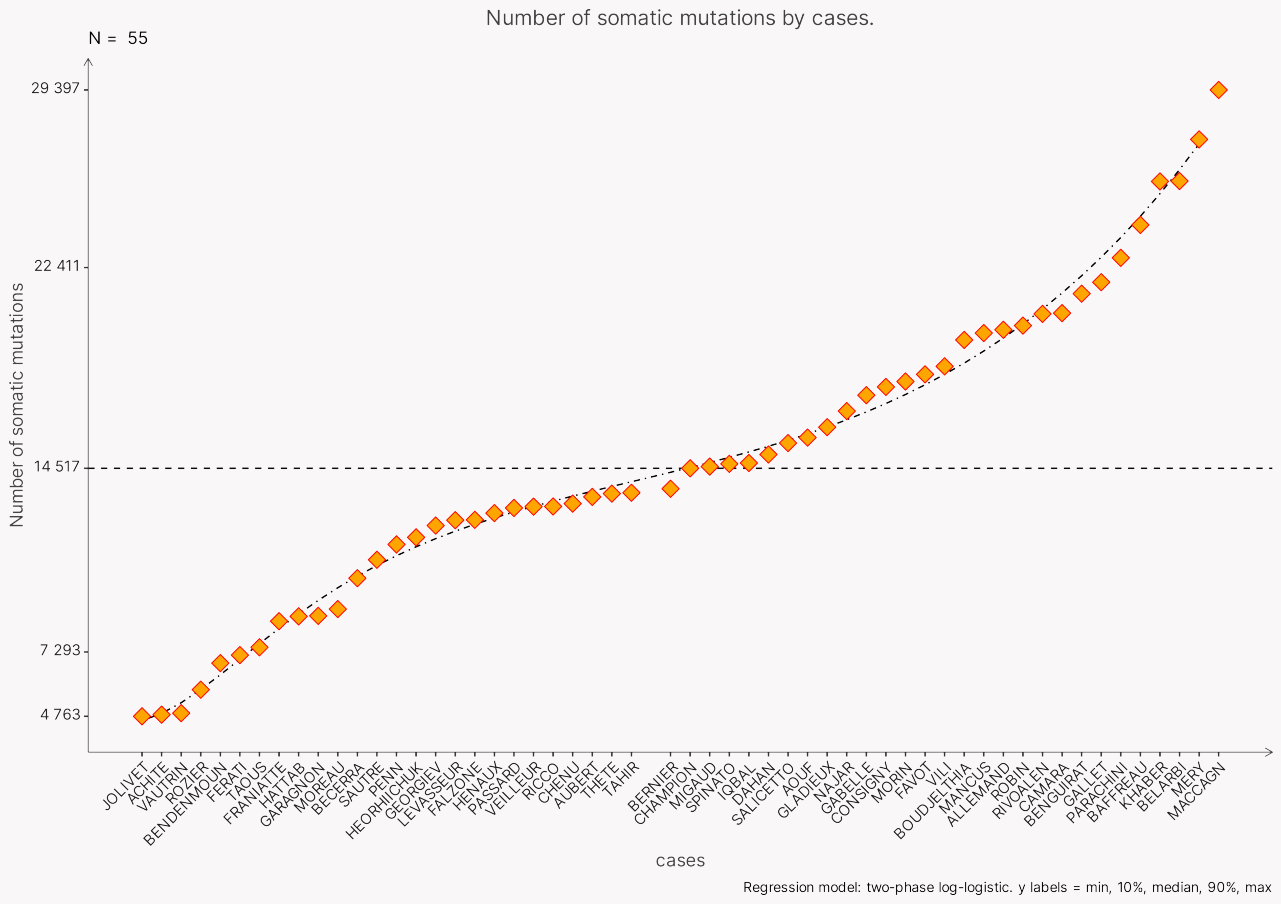
1. VCF filter



2. BAM filter



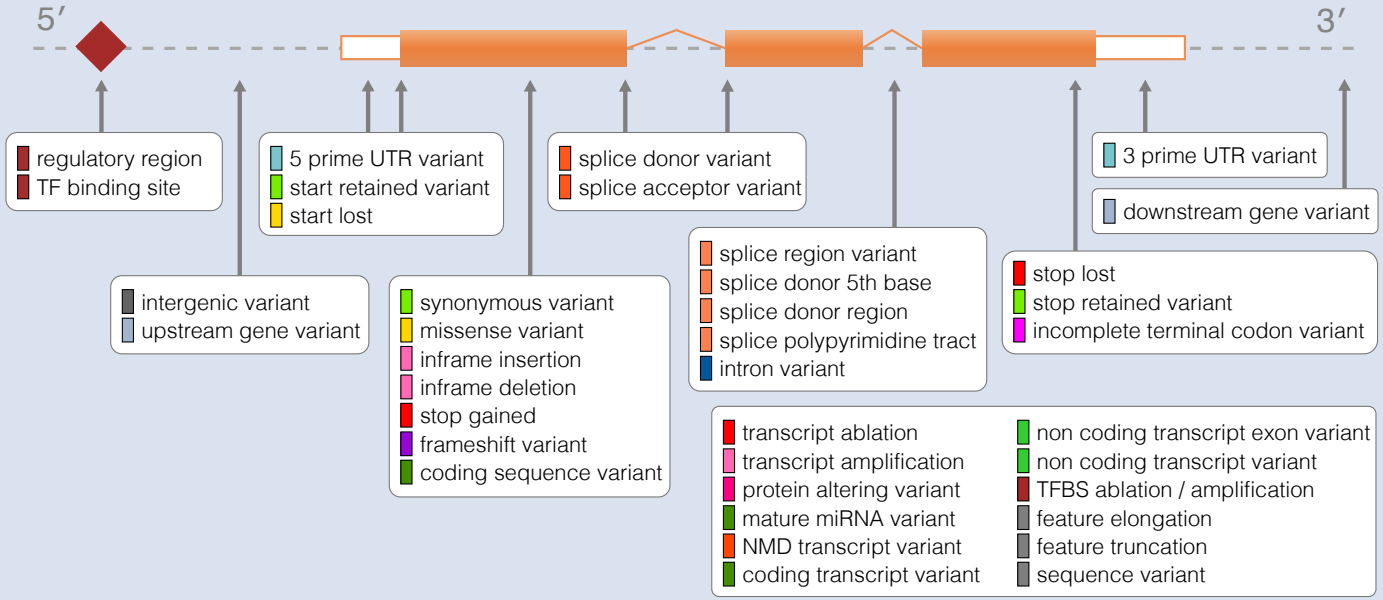
Results — Number of somatic alterations



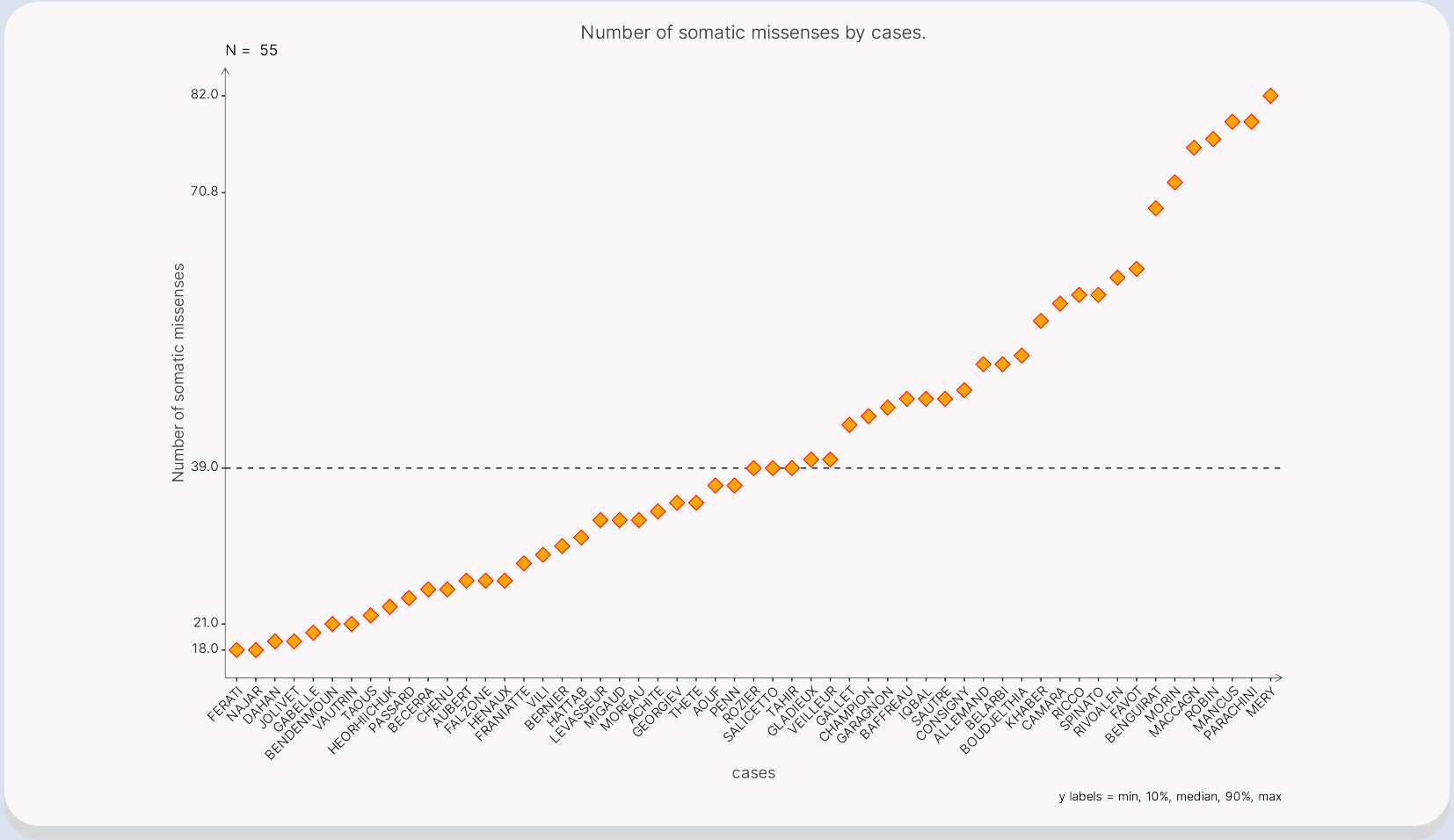
Bioinformatic Pipeline — Annotation

3. Annotations

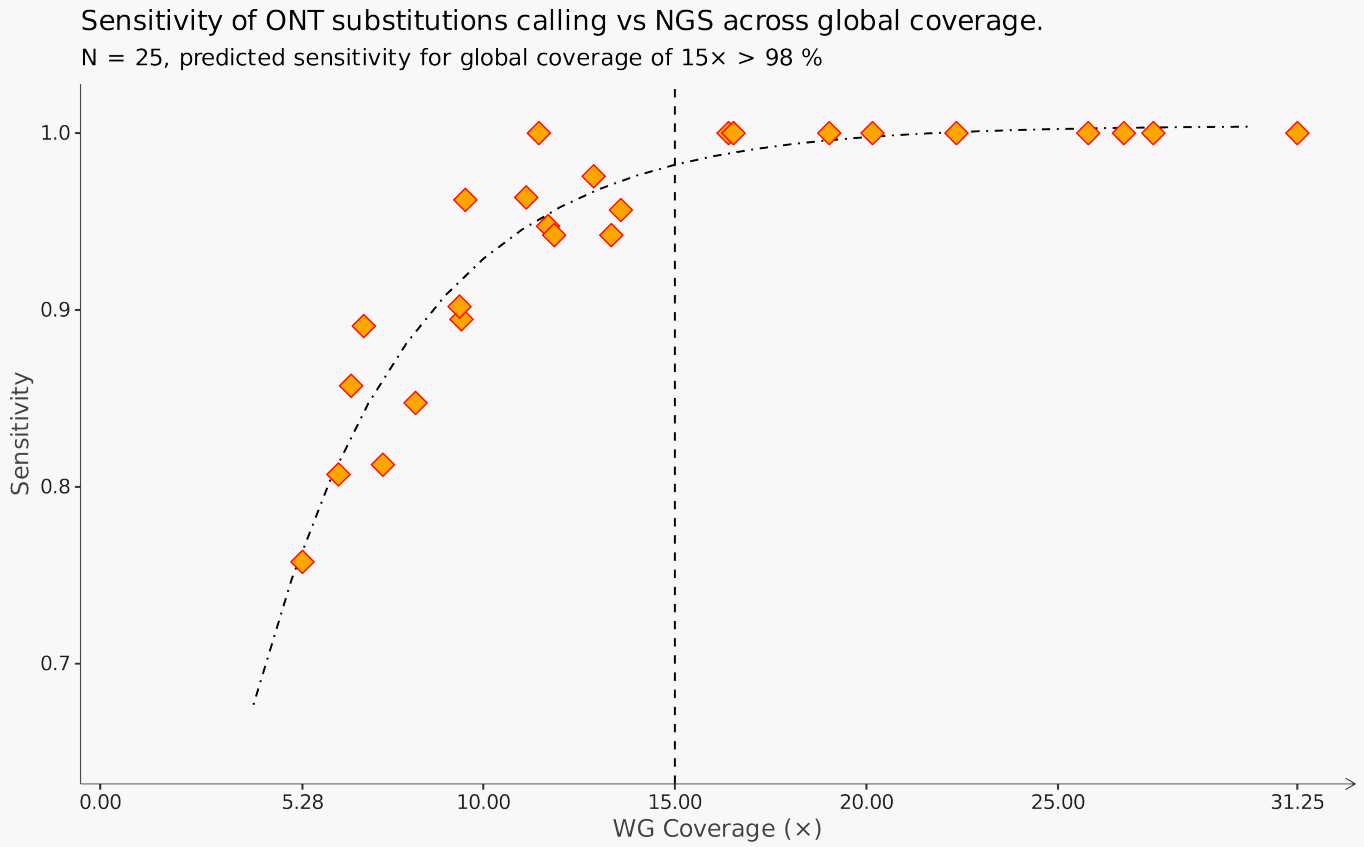
- VEP for variant consequence prediction (ensembl v112)
- Cosmic DB (latest)
- dbSNP
- NCBI genomic regions (latest)



Results — Number of somatic missense alterations



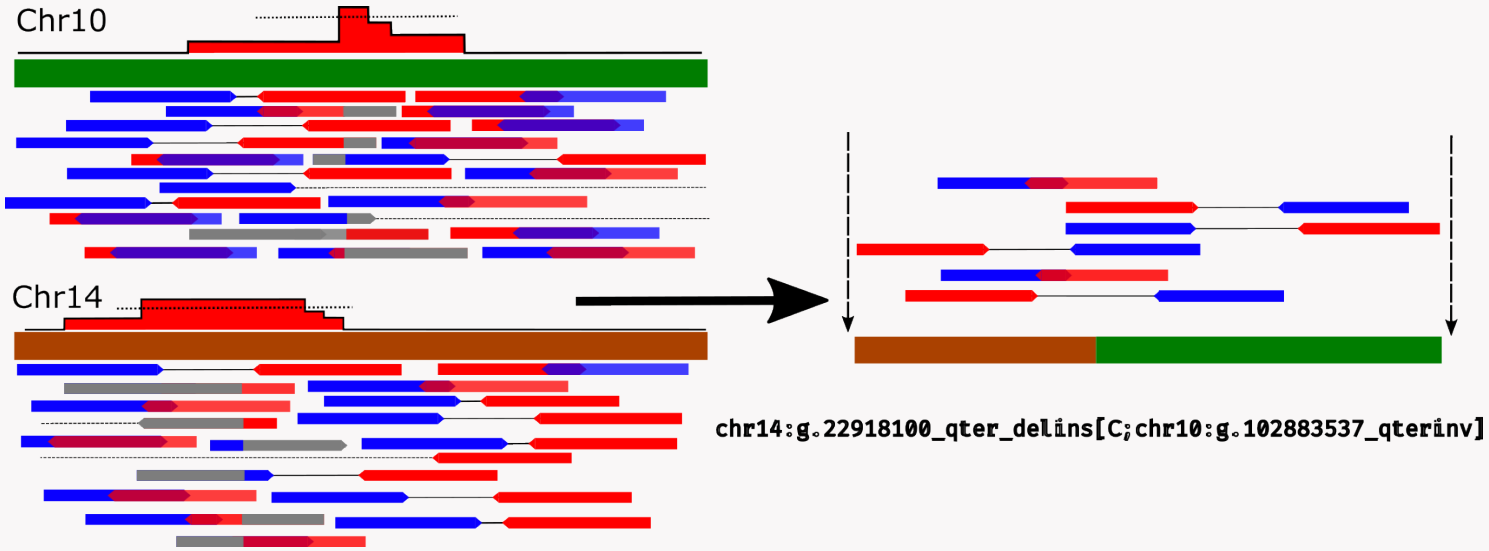
Performances — Substitution calling



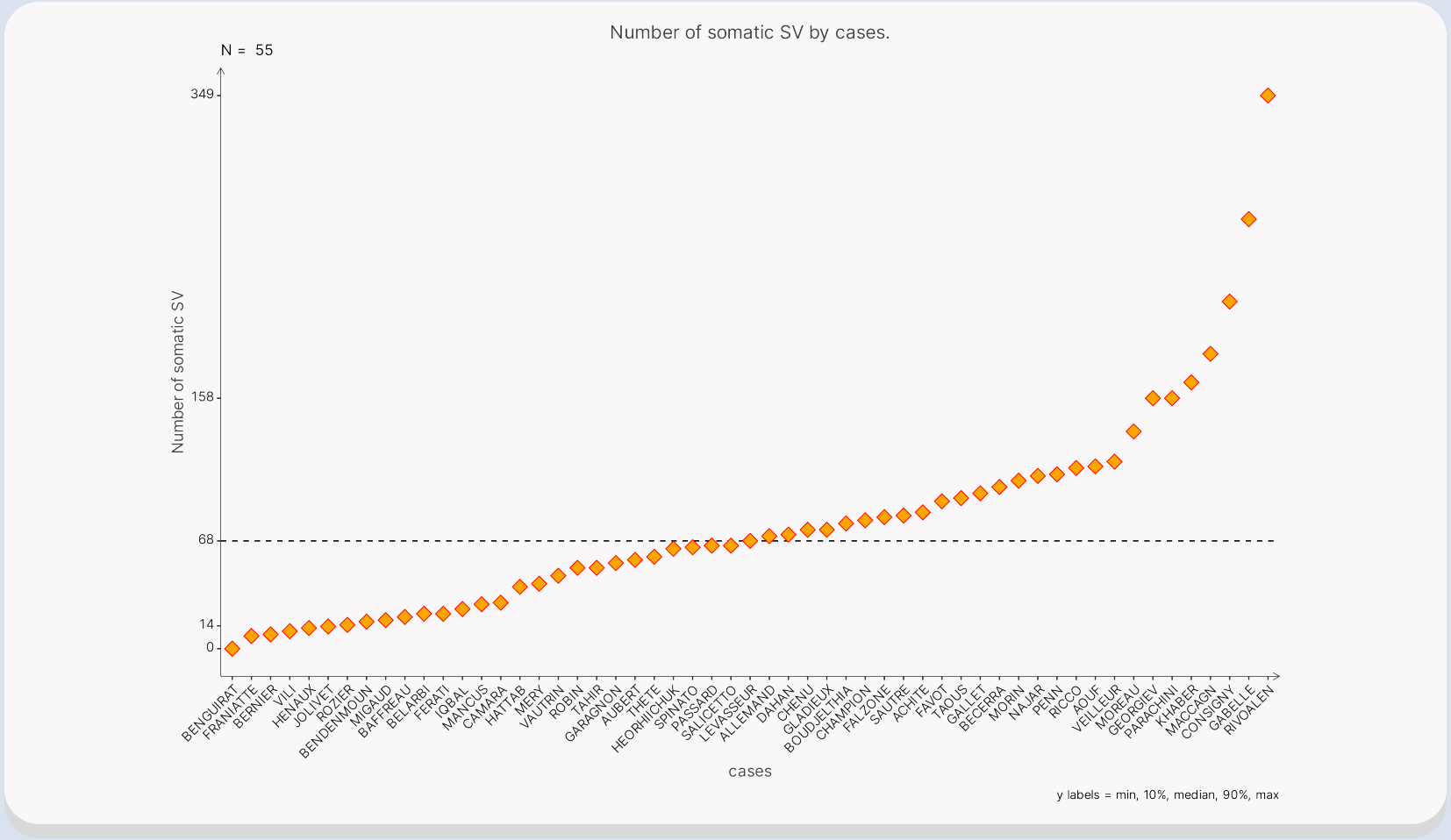
Bioinformatic Pipeline — *de novo*

Implementation of *de novo assemblage* for LRS:
Inspired by **SV-finder** (local de novo assembly)

- **Scan** alignements and select locally misaligned reads (outliers detection).
- **Assemble** them together (wtdbg2 v2.5 and spades v4.0).
- **Describe** the resulting consensus sequence (Blast, minimap2).



Results — Number of somatic SV



🔬 Visualization and interpretation of results

Development of a web service (HTMX + Bun) for **sharing, visualization, and interpretation of the results**.

DEMO (C.... ex.: *PHF6* et *AEBP2*)

Reporting

After the manual interpretation of the results and the redaction of a conclusion.

Reporting

After the manual interpretation of the results and the redaction of a conclusion.

The system generates a detailed PDF report that seamlessly integrates:

1. Detailed quality metrics (with graphics generation)
2. Interpreted genetic mutations (Pathogenic, ...)
3. Analytical conclusions

Reporting

After the manual interpretation of the results and the redaction of a conclusion.

The system generates a detailed PDF report that seamlessly integrates:

1. Detailed quality metrics (with graphics generation)
2. Interpreted genetic mutations (Pathogenic, ...)
3. Analytical conclusions

Example...

👥 Cohorts

Initially, to investigate our hypothesis, we decided to sequence T-ALLs harboring deregulation of the expression of known frequent oncogenes in T-ALL:

- ***TAL1*** (cis-deregulated),
- ***HOXA9*** (RT-MLPA neg),
- ***TLX1*** deregulated without genetic explanation (FISH neg and NGS panel neg).

👥 Cohorts

Initially, to investigate our hypothesis, we decided to sequence T-ALLs harboring deregulation of the expression of known frequent oncogenes in T-ALL:

- ***TAL1*** (cis-deregulated),
- ***HOXA9*** (RT-MLPA neg),
- ***TLX1*** deregulated without genetic explanation (FISH neg and NGS panel neg).

We also decided to sequence a cohort of pediatric T-ALLs and an adult one with deregulation of *TLX3* (Pediac/Manon project). As well as a cohort of T-ALLs **< 3 years**.

👥 Bilan des flowcells

Bilan WGS Nanopore par projet

	Fait WGS Nanopore	
projet	patients	echantillions
<3ans	20	40
	20 D+M	
TLX3 ped	13	24
	11 D+M; 2 D	
TLX3 ad	13	25
	12 D+M; 1 D	
TAL1	14	20
	6 D+M; 8 D	
TLX1	17	28
	11 D+M; 6 D	
HOXA	17	34
	17 D+M	
LBL-T	9	16
	7 D+M; 2 D	
ZFP36L2	1	2
	1 D+M	
TOTAL	104	189
TOTAL*	98	179
	81 D+M; 17 D	

* 6 patients comuns entre les projets

	Reste a faire	
projet	patients	echantillions
<3ans	0	0
TLX3 ped	1	2
	1 D+M	
TLX3 ad	0	0
TAL1	3	5
	2 D+M ; 1 D	
TLX1	0	0
HOXA	8	16
	8 D+M	
LBL-T	8	16
	8 D+M	
ZFP36L2	1	2
	1 D+M	
TOTAL	21	41

commandes FC WGS :

1- Commande en Mai de 40 pack de FC (soit 160 FC)

a. Réception de 5 pack (20 FC) en Juin – reste 0

b. Réception de 4 pack (16 FC) en septembre - reste 16

2- Réception prévu en décembre : 13 pack (52 FC) – reste 52

3- Réception prévu en Mars 2025 : 17 pack (68 FC) – reste 68

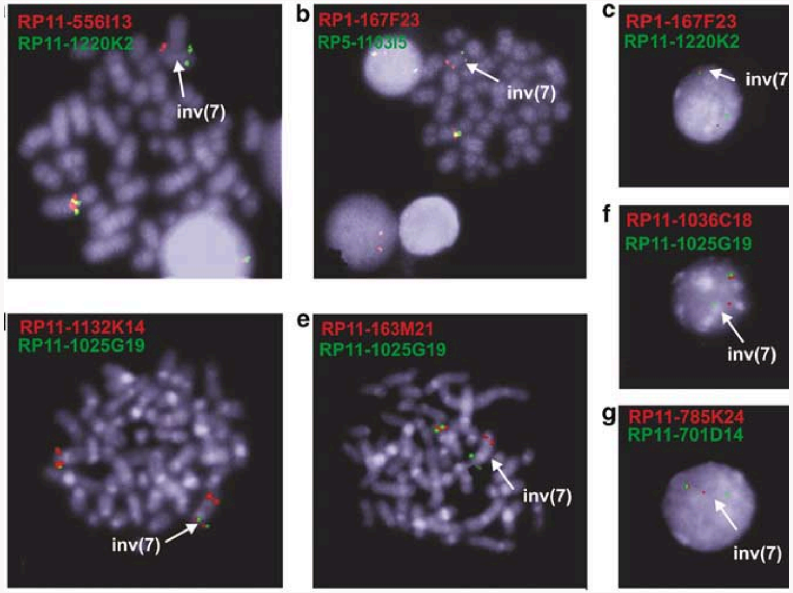
👥 Cohort *HOXA* N = 14

id	X	mécanisme	interpretation	info
ALLEMAND	11	Multiples mutations de *MLLT3* + hotspot CSF3R	❓	
BELARBI	13	Altération de GATA3 + KMT2C	❓	
BERNIER	15	DPP10::SET	✅	
BOUDJELTHIA	13	KMT2E ter	❓	JAK/STAT
GARAGNON	13	TRB::HOXA inv7	✅	JAK/STAT
HATTAB	11	AFDN::KMT2A	✅	
KENNOUCHE	*en cours*			
LAVIDALE	*en cours*			
MANCUSO	10	TNRC18::KMT2A	✅	
MERY	16	TRB::HOXA inv7	✅	JAK/STAT
MICHELAS	*en cours*			
MIGAUD	26	mutation EN1 (homeobox)	❓	
MORIN	22	TRB::HOXA inv7	✅	JAK/STAT
SAUTRE	28	AFDN::KMT2A	✅	

Interesting results: *HOXA9*

ME

- **Inv(7)(p15q34)** TRB/HOXA10
- `chr7:27,287,813_delins[ATGGGGGGGG_chr7:144,128,326inv]`
- `chr7:27,313,165_delins[GATGG_chr7:144,161,537inv]`



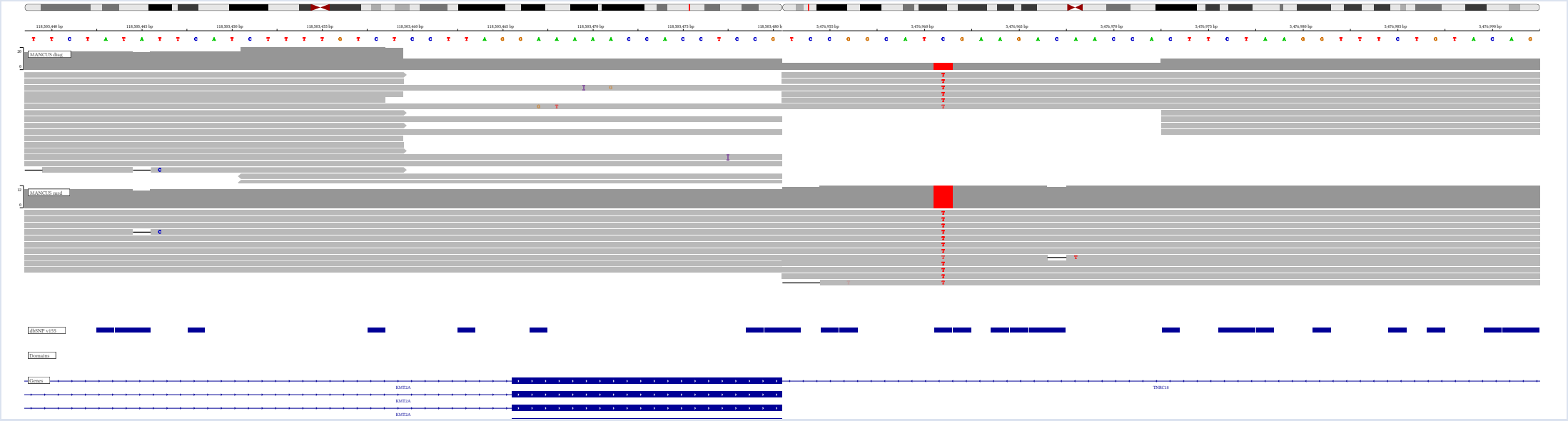
Résultats FISH break apart.

Speleman, F et al. "A new recurrent inversion, inv(7)(p15q34), leads to transcriptional activation of HOXA10 and HOXA11 in a subset of T-cell acute lymphoblastic leukemias." Leukemia vol. 19,3 (2005): 358-66.

Interesting results: *HOXA9*

MA

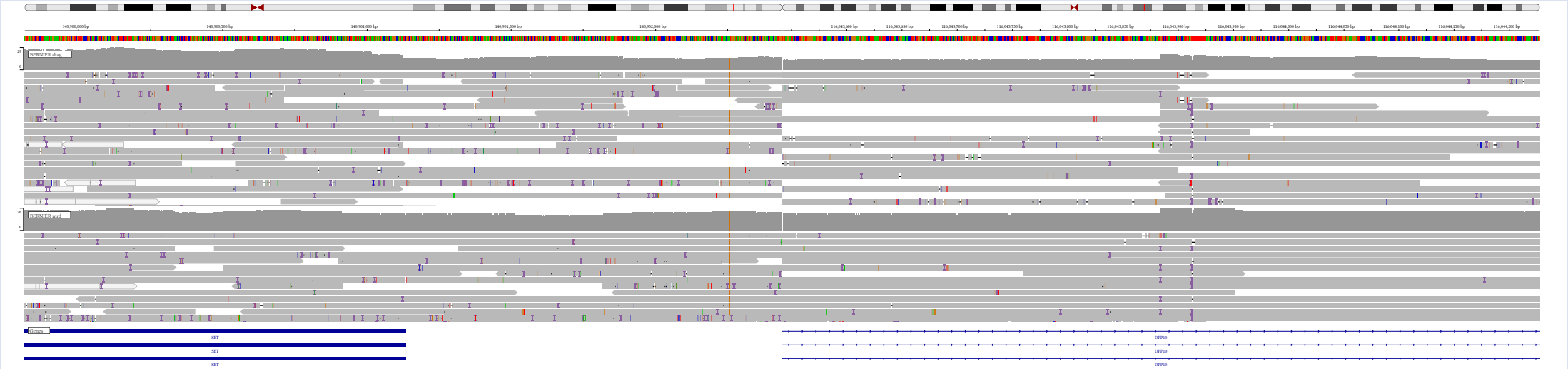
- t(7;11)(p22;q23) likely ***KMT2A::TNRC18***
- `chr11:118,503,451(KMT2A)_delins[CCCC_chr7:5476973(TNRC18)]`
- Known fusion transcript (2 pediatric cases / 759 LAL. Meyer, C et al. Leukemia 2009)
- RT-MLPA probes w/o TNRC18



Interesting results: *HOXA9*

BE

- Translocation t(2;9)(q14.1;q34.11) likely fusion transcript ***SET::DPP10***
- `chr9:140,901,114(SET)_delins[GAACATAAAGAAAAAAA_chr2:116,043,899(DPP10)]`
- **New fusion transcript in T-ALL** (only described one time in CRC: Xia, Li C et al. “Identification of large rearrangements in cancer genomes with barcode linked reads.” Nucleic acids research vol. 46,4 2018)



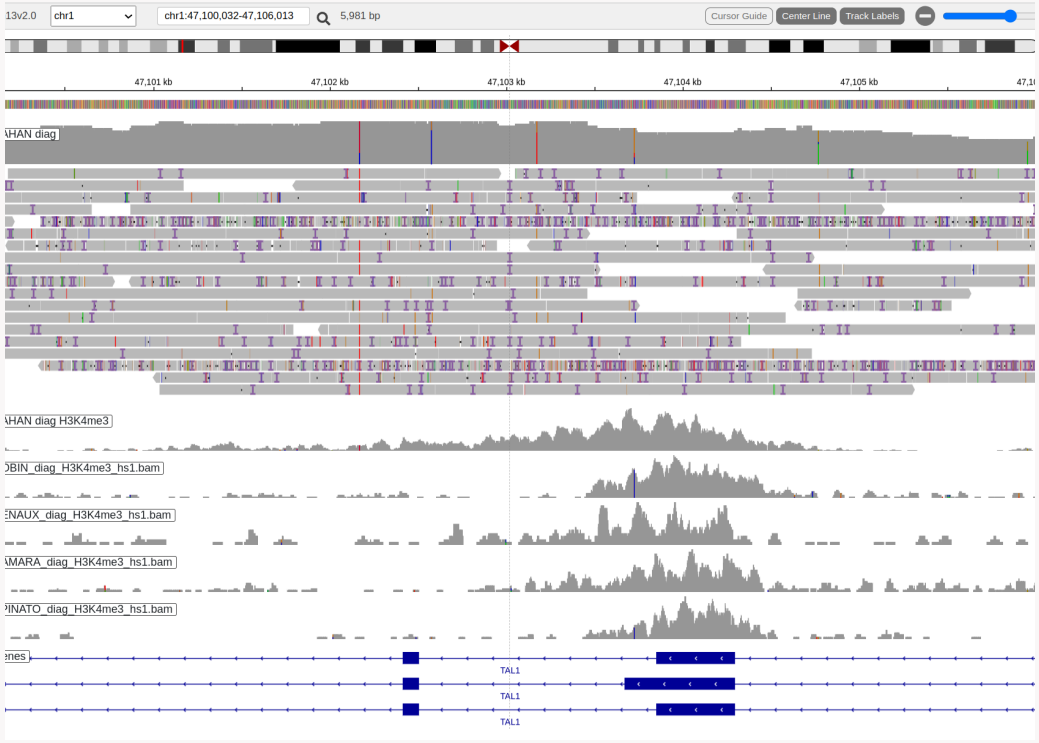
👥 Cohort *TAL1* N = 8

id	X	mécanisme	interpretation
ACHITE	12	t(1;17) *TAL1::RPTOR*	✅
BECERRA	12	*TCR::LMO2*	❓
CAMARA	20	insertion dans *MYB*	❓
COLLE	7		
DAHAN	13	duplication intronique *TAL1* (ex2-3)	❓
DOUYERE	22 pas mrd	(*BCL11B::TLX3*)	❓
FALLE	*en cours*		
GILLOUX	15 pas mrd	pas d'altération locus *TAL1*	❓
HENAUX	12	ins 9 nt ex3-4	❓
JEANSELME	*en cours*		
ROBIN	14	*LMO2::RAG1* (del 11q)	❓
SCHOLZ	16		

Interesting results: *TAL1*

DAH

- Mono-allelic surexpression of TAL1.
- Somatic tandem duplication between exon 2 and 3 of TAL1.
- Also observed in ChIP-seq data showing broad H3K4me3 coverage.



👥 Cohort *TLX1* N = 12

id	X	mécanisme	interpretation
AUBERT	17	del 10q (10mb) *TLX1*	✅
CHAMPION	32	inv10	✅
LEVASSEUR	20	inv10 (avec 5' UTR ralongé)	✅
PARACHINI	17	t(10;14)	✅
PASSARD	18	chromotripsis du 10	✅
PAYSAN	*en cours*		
RIVOALEN	19	inv10	✅
SALICETTO	12 pas mrd	inv10	✅

Interesting results: TLX1

LEV

- Surexpression of TLX1.
-
-

Interesting results: TLX1

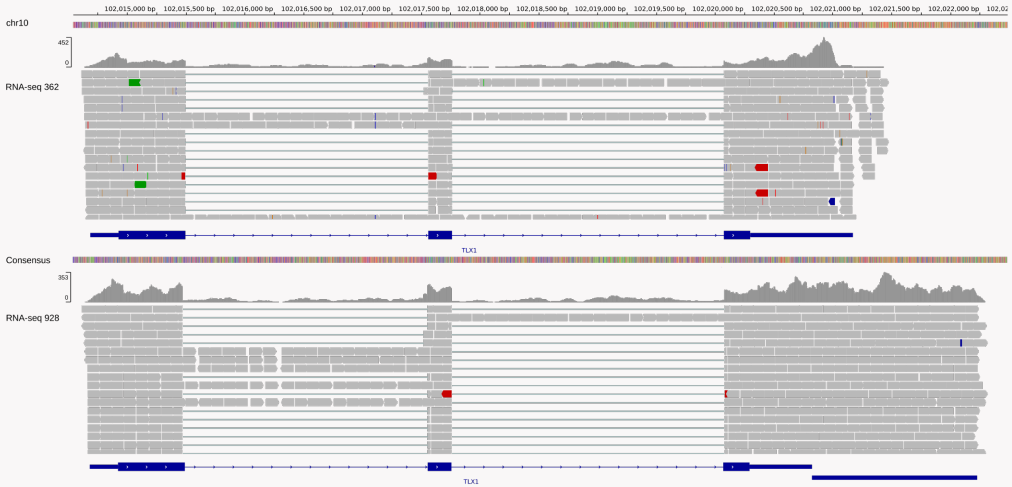
LEV

- Surexpression of TLX1.
- Inv10
-

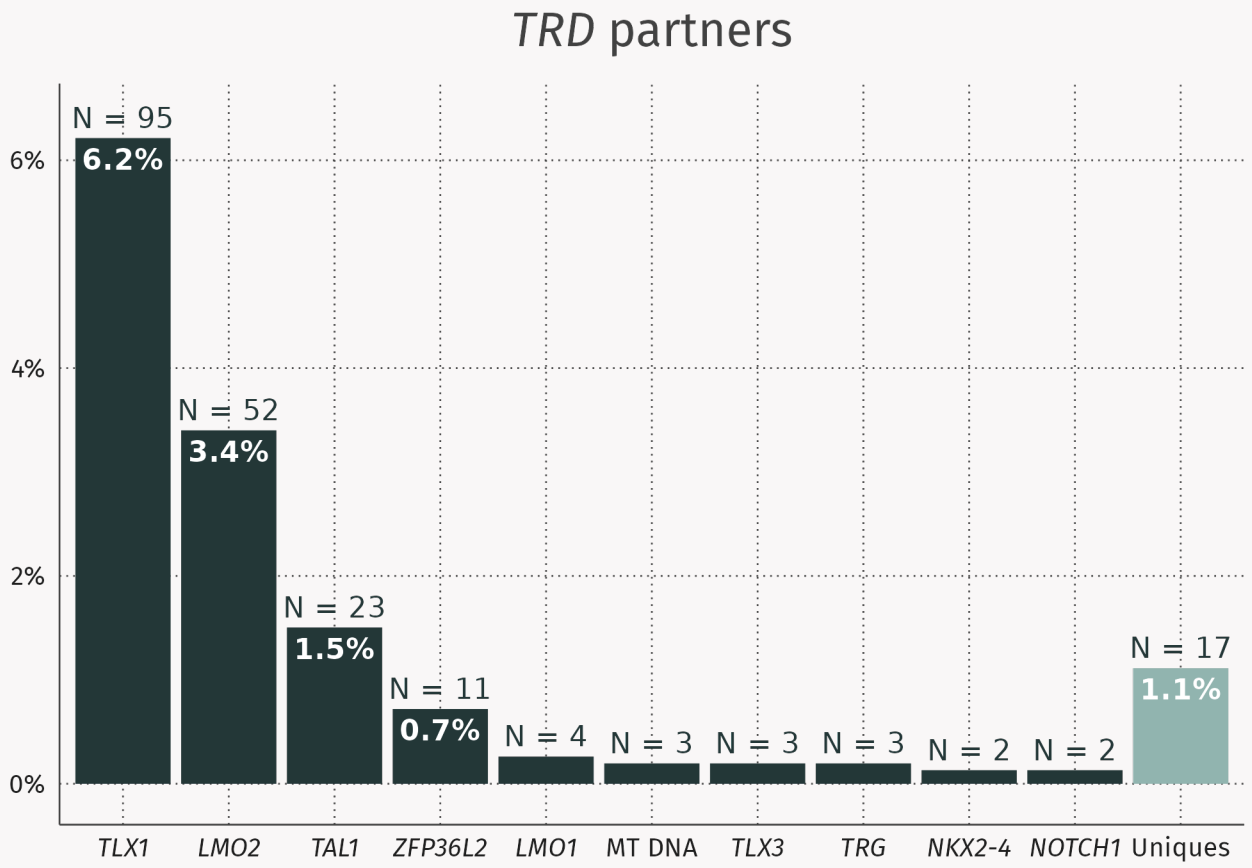
Interesting results: TLX1

LEV

- Surexpression of TLX1.
- Inv10
- Modification of 3′ UTR.



Discussion


TRD partners
6%
4%
2%
0%
N = 95
6.2%
N = 52
3.4%
N = 23
1.5%
N = 11
0.7%
N = 4
N = 3
N = 3
N = 3
N = 2
N = 2
N = 17
1.1%
TLX1
LMO2
TAL1
ZFP36L2
LMO1
MT DNA
TLX3
TRG
NKX2-4
NOTCH1
Uniques

🏁 Conclusions

What is **done**

- ✅ The implementation of a robust and more informative whole-genome sequencing method.
- ✅ The development of a pipeline for detecting somatic alterations (SNV, SV, viral insertion) as well as a simple way to visualize and interpret the results.
- ✅ Compare missense/indels variant calling with NGS panel.
- ✅ The integration, if available, of RNAseq / ChIPseq.
- ✅ Develop a de novo assembly pipeline (better accuracy and visualization of SV).
- ✅ Automate reporting.

🏁 Conclusions

What is **done**

- ✅ The implementation of a robust and more informative whole-genome sequencing method.
- ✅ The development of a pipeline for detecting somatic alterations (SNV, SV, viral insertion) as well as a simple way to visualize and interpret the results.
- ✅ Compare missense/indels variant calling with NGS panel.
- ✅ The integration, if available, of RNAseq / ChIPseq.
- ✅ Develop a de novo assembly pipeline (better accuracy and visualization of SV).
- ✅ Automate reporting.

TODO

- ❌ Functional experiments ?
- ❌ Complete sequencing of cases (> 200).
- ❌ Aggregate results by cohorts and tag recurrency (at gene and mutation levels).

Remerciements

The Necker team

Pr. Vahid Asnafi

Dr. Patrick Villarese

Dr. Agata Cieslak

Coline Lefevre

Necker ENFANTS MALADES HÔPITAL UNIVERSITAIRE

Aix*Marseille université

The TAGC team

Dr. Salvatore Spicuglia

Dr. Guillaume Charbonnier

Gaëlle Farah