🧬 TREC mediated oncogenesis in human immature T lymphoid malignancies preferentially involves *ZFP36L2* — Molecular Cancer

Dr. Thomas Steimlé

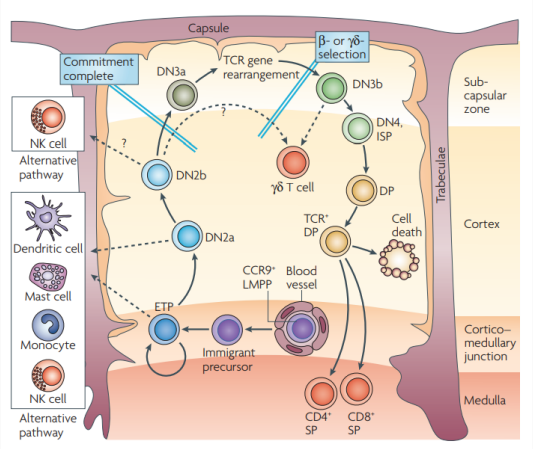
June 29, 2023

Aix*Marseille université

umr U 1090 TAGC theory and approaches of genomic complexity

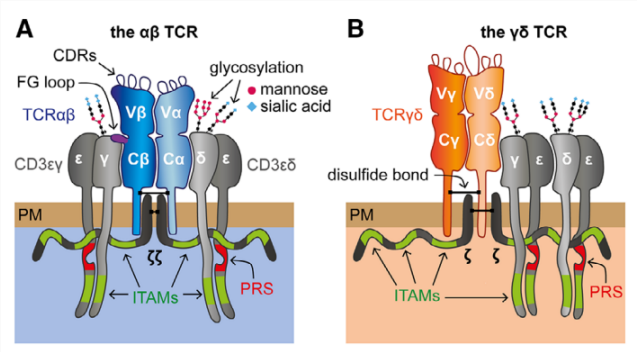
Background — Thymopoiesis

- During thymopoiesis (HSC ⇒ T-cell), the phenotypic diversity of the antigen receptor (TCR) is acquired.



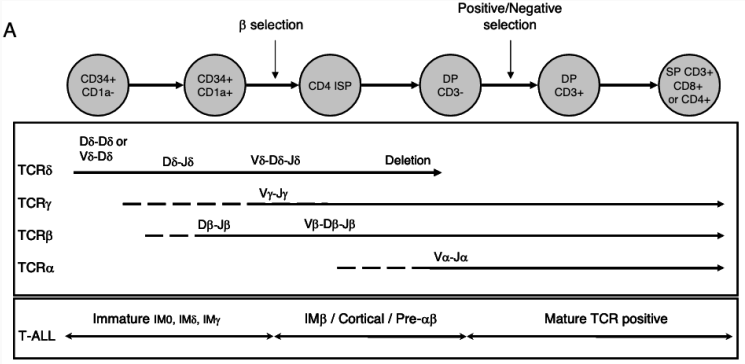
Background — Thymopoiesis

- During thymopoiesis (HSC $\Rightarrow$ T-cell), the phenotypic diversity of the antigen receptor (TCR) is acquired.



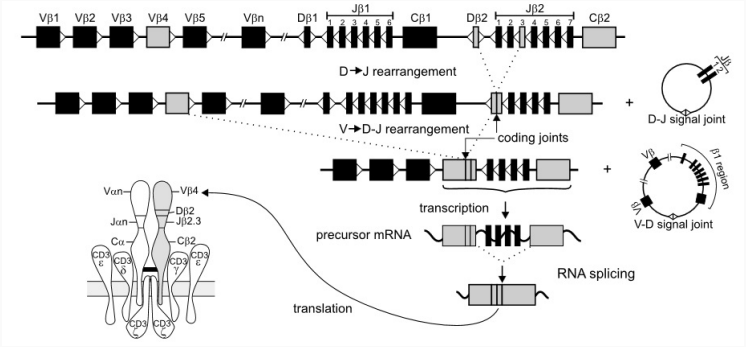
Background — V(D)J recombination

- This process involves a series of genome recombination of the different TCR loci followed by proliferation and selection of functionnal non self-reacting TCR.



Background — V(D)J recombination

- This process involves a series of genome recombination of the different TCR loci followed by proliferation and selection of functionnal non self-reacting TCR.



Background — V(D)J recombination

- V(D)J recombination is a threat to genomic stability, prone to induce DSB occurring in genes outside of the TCR loci, followed by erroneous repair resultating in SV.
- This oncogenetic process is responsible of well known genetic alterations in T-ALL (particulary translocations accountable of ectopic expression of oncogenes *TLX1*, *TAL1* etc...)[1].

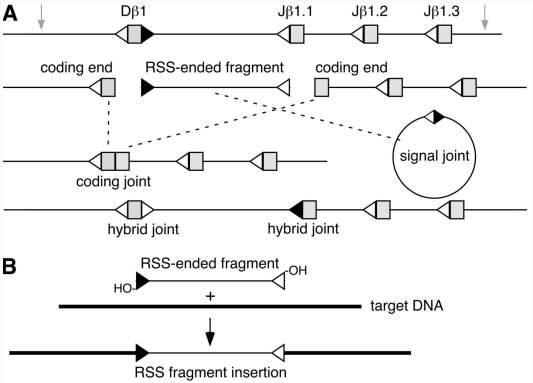
Table 1. TCR-oncogene translocation occurrence in T-ALL.

Protein family	Gene	Chromosome band	Chromosomal Aberrations*	Occurrence in T-ALL
Basic helix-loop- helix transcription factors (bHLH)	*TAL1*	1p32	t(1;14)(p32;q11)/ t(1;7)(p32;q34)	~4%
	TAL2	9q34	t(7;9)(q34;q34)	~2%
	LYL1	19p13	t(7;19)(q34;p13)	~7%
	OLIG2	21q22	t(14;21)(q11;q22)	2%
	MYC	8q24	t(8;14)(q24;q11)	~1%
Lim only domain (LMO) proteins	*LMO2*	11p13	t(11;14)(p13;q11)/ t(7;11)(q34;p15)	~6%
	LMO1	11p15	t(11;14)(p15;q11)/ t(7;11)(q34;p13)	~2%
	LMO3	12p12	t(7;12)(q34;p12)	<1
Homeobox proteins	*TLX1*	10q24	t(10;14)(q24;q11)/ t(7;10)(q34;q24)	5-10%C, ~30%U
	TLX3	5q35	t(5;14)(q35;q11)	20-25%C,~5%U
	HOXA cluster	7p15	t(7;14)(p15;q11)/ inv(7)(p15;q34)	~3%
	NKX2.1	14q13	t(7;14)(q34;q13)/ inv(14)(q13;q32)	<1%
	NKX2.4	20p11	t(20;14)(p11;q11)	<1%
	NKX2.5	5q35	t(5;14)(q35;q32)	<1%
Other	*NOTCH1*	9q32	t(7;9)(q34;q34)	<1%
	CCND2	12p13	t(7;12)(q34;p13)/ t(12;14)(p13;q11)	<1%
	MYB	6q23	t(6;7)(q23;q34)	~3%
	LCK	1p34	t(1;7)(p34;q34)	<1%
	BCL11B	14q32	inv(14)(q11;q32)	<1%
	TCL1A	14q32	t(7;14)(q34;q32)/ inv(14)(q11;q32)	<1%
	BMI1	10p12	t(7;10)(q34;p12)	<1%

*Chr 14q11: TCRD locus. Chr 7q34: TCRB locus. **Larmonie et al., *unpublished data, 2013*. C:Childhood, A:Adulthood.

1. Larmonie, Nicole S D et al. "Breakpoint sites disclose the role of the V(D)J recombination machinery in the formation of T-cell receptor (TCR) and non-TCR associated aberrations in T-cell acute lymphoblastic leukemia." Haematologica vol. 98,8 (2013): 1173-84

Background — T-cell receptor excision circles (TRECs)



- During recombination deleted parts of the loci are circulized into TRECs.
- Similar to transposons, the reintegration of TRECs has been implicated in the deregulation or inactivation of targeted genes.

[O] Curry, John D et al. "Chromosomal reinsertion of broken RSS ends during T cell development." The Journal of experimental medicine vol. 204,10 (2007): 2293-303. doi:10.1084/jem.20070583

Hypothesis

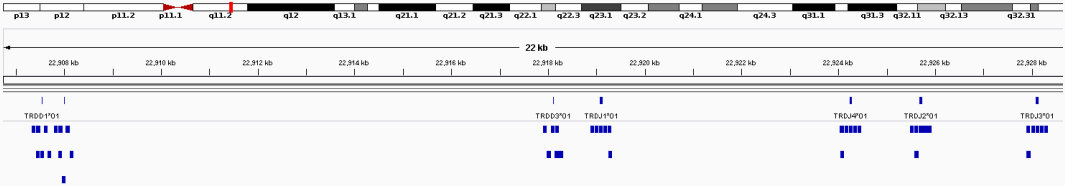
- In T-ALL, we could find with molecular biology tools insertions of those TRECs.
- With the same tools we could also find all the translocations which involves the TCR.

Material & Method

- We used our extensed collection of T-ALL samples at diagnostic n = 1533.

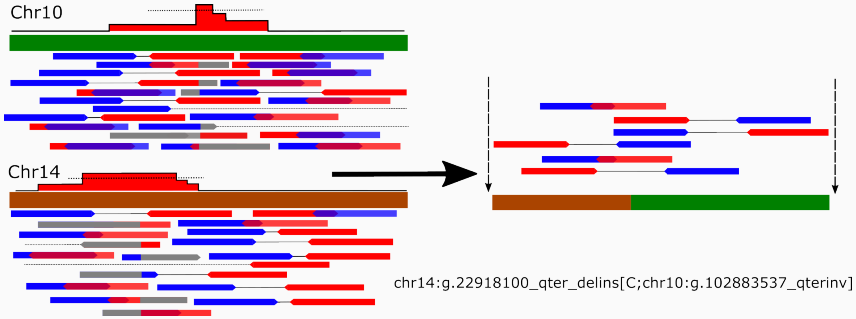
Material & Method

- We used our extensed collection of T-ALL samples at diagnostic n = 1533.
- We designed a NGS capture assay with capture probes mapped at multiple parts of de TCR δ locus.



Material & Method

- We used our extensed collection of T-ALL samples at diagnostic n = 1533.
- We designed a NGS capture assay with capture probes mapped at multiple parts of de TCR δ locus.
- We developed a specific software to analysed aligned reads and call SV `https://github.com/Dr-TSteimle/sv-finder`.



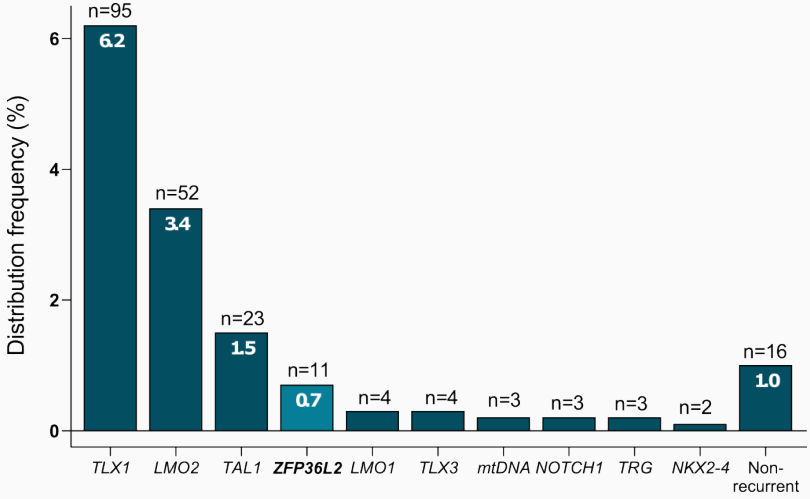
Results — *TRD* translocations — Validation cohort

- To validate our method, we used a previously published cohort of 264 cases analysed with *TRD* dual-color FISH probe[1].
- Se = 98.1% [95% CI 96-99] and Sp = 97.7%
- The 4 FN cases are in fact TRECs insertions inside *ZFP36L2* that couldn't have been seen with FISH !

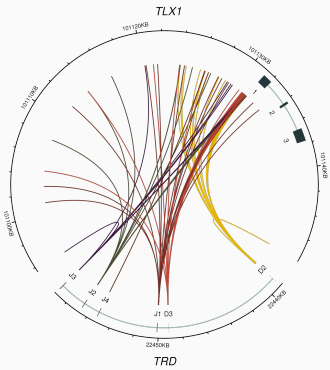
		FISH	
		Positive	Negative
NGS	Positive	43	4
	Negative	1	216

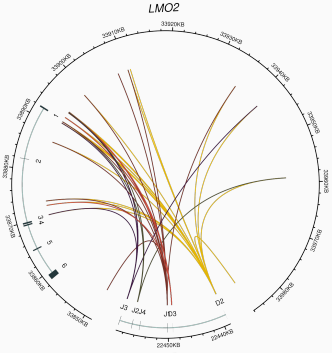
[1] Le Noir, Sandrine et al. "Extensive molecular mapping of TCRα/δ- and TCRβ-involved chromosomal translocations reveals distinct mechanisms of oncogene activation in T-ALL." Blood vol. 120,16 (2012): 3298-309. doi:10.1182/blood-2012-04-425488

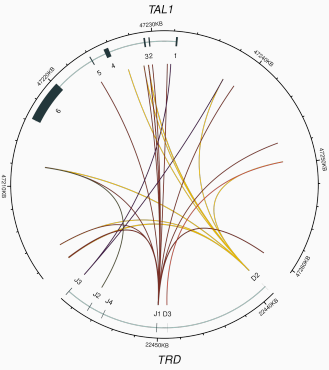
Results — *TRD* translocations — Discovery cohort



🎯 Results — *TRD* translocations — Discovery cohort

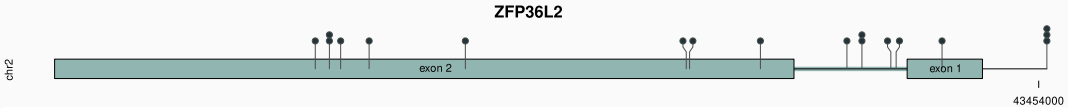




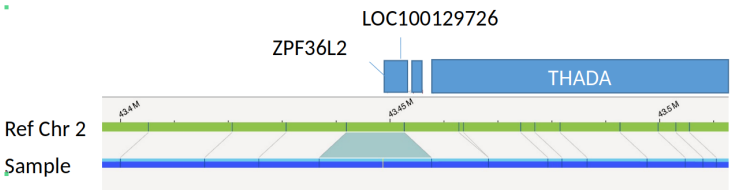


Results — *TRD* translocations — Discovery cohort

We confirmed all TRECs insertions with sanger sequencing and OGM.



Patient ID	ZFP36L2	N	heptamer	spacer 23bp	nonamer		nonamer	spacer 12bp	heptamer	N	ZFP36L2
			Dδ2 23RSS				Dδ3 12RSS				
UPN 631	...TTGTTTTTTAACAGA	ACCCCGAT	...ACAG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...	AGTTTTTGT	AAAGCTCTGTAG	CACTGTG	-	AGGGGGGCGGCGGCT...
UPN 1044	...CGCAAGGCTTCCAGA	-	CACACAG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...	-	-	-	-	-
UPN 1402	...GGCAGCCGGGGCGGA	TGGC	CACACAG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...	-	-	-	-	-
UPN 1061	...GGGAGGGGCGTCCCC	-	CACACAG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...GCATCTG	-	-	-	AATTTCGTCCCGCCA	GAAGCCATGCGCGAA...
UPN 1310	...ACCAAGTAACCAGTA	-	CACACAG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...	AGTTTTTGT	AAAGCTCTGTAG	CACTGTG	-	CAGTATGGACCTTTG...
UPN 1007	...TGTCTCAGACCGGTG	-	CACACAG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...	AGTTTTTGT	AAAGCTCTGTAG	CACTGTG	-	CGGTGTAAACAAACC...
UPN 1115	...CGGGAAGGCACGGCT	-	CACACAG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...	AGTTTTTGT	AAAGCTCT....	.....	A	CGGCAGTGGAGAAAC...
LLT-152	...TTGTAGTTCTTGGAC	CT	.....AG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...	AGTTTTTGT	AAAGCTCTGTAG	CACTGTG	-	AGCCGTGCCTTCCCG...
LLT-211	...GGGCGAGGGGCGGGG	-	CACACAG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...	-	-	-	-	-
LLT-79	-	-	-	-	-	...	AGTTTTTGT	AAAGCTCTGTAG	CACTGTG	-	ACCCCTCGCCATCTC...
LLT-244	...CTGCCGGGGCTGGGC	-	CACACAG	GTTGGAGTGCATTAAGCCTTTGT	CCAAAAACA	...	AGTTTTTGT	AAAGCTCTGTAG	CAC....	CCTGT	GCACTTTTCGCCGTA...



Conclusions

- We developed a highly accurate and sensitive method that enables precise characterization of *TRD* structural variations. This method can be applied at diagnosis without incurring any additional costs.
- Our findings provide confirmation that recurrent TRECs insertions are present in cases of T-ALL.
- Using our method, we have observed that these recurrent TRECs insertions predominantly disrupt the *ZFP36L2* gene.
- The tumor suppressor *ZFP36L2* is well-established to be involved in V(D)J recombination, but further clarifications are needed regarding its specific role in oncogenesis. Our findings will help in the characterization of this tumor suppressor.

🤔 My PhD — Epigenetic Rationnal

- The aforementioned results contribute to the understanding of the dysregulation of proto-oncogenes such as *TAL1*, *TLX1* or *TLX3*.

[1] Bradner JE, et al. Cancer. Cell. 2017 Feb 9;168(4):629-643

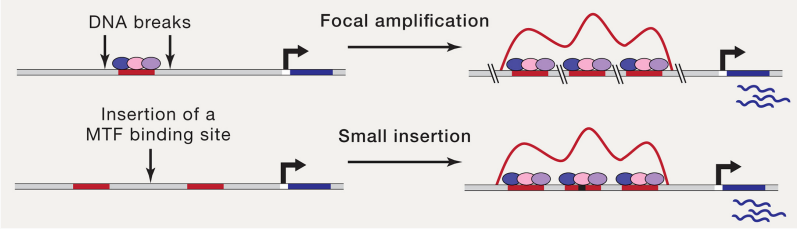
🤔 My PhD — Epigenetic Rationnal

- The aforementioned results contribute to the understanding of the dysregulation of proto-oncogenes such as *TAL1*, *TLX1* or *TLX3*.
- Despite extensive investigation, the molecular mechanisms underlying the dysregulation of these oncogenes, remain elusive in many cases.

[1] Bradner JE, et al. Cancer. Cell. 2017 Feb 9;168(4):629-643

🤔 My PhD — Epigenetic Rationnal

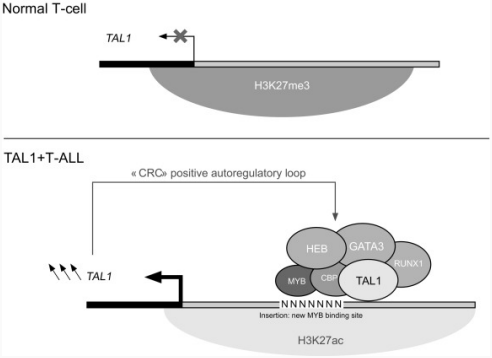
- The aforementioned results contribute to the understanding of the dysregulation of proto-oncogenes such as *TAL1*, *TLX1* or *TLX3*.
- Despite extensive investigation, the molecular mechanisms underlying the dysregulation of these oncogenes, remain elusive in many cases.
- It has been demonstrated that tumor cells acquire enhancers[1] through intergenic sequence mutations that enable binding of transcription factors.



[1]Bradner JE, et al. Cancer. Cell. 2017 Feb 9;168(4):629-643

🤔 My PhD — Epigenetic Rationnal

- The presence of upstream indels in *TAL1* leads to the formation of a neo-enhancer[1].
- It has also been shown that the transcription factor MYB can bind to this neo-enhancers[2].



[1] Navarro JM et al. Nat Commun. 2015;6:6094

[2] 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

Hypothesis

A By taking a pan-genomic approach, it is possible to identify numerous intergenic alterations correlated with the cis deregulation of adjacent genes (correlation between transcriptome RNA-seq and ChIP-seq).

Hypothesis

A By taking a pan-genomic approach, it is possible to identify numerous intergenic alterations correlated with the cis deregulation of adjacent genes (correlation between transcriptome RNA-seq and ChIP-seq).

B Some of these genes are expected to be known oncogenes, while others have the potential to be novel oncogenes (discovery).

Hypothesis

A By taking a pan-genomic approach, it is possible to identify numerous intergenic alterations correlated with the cis deregulation of adjacent genes (correlation between transcriptome RNA-seq and ChIP-seq).

B Some of these genes are expected to be known oncogenes, while others have the potential to be novel oncogenes (discovery).

C It is likely that these specific alterations are the causative factors for the cis deregulation of adjacent genes (functional experiments).

Hypothesis

A By taking a pan-genomic approach, it is possible to identify numerous intergenic alterations correlated with the cis deregulation of adjacent genes (correlation between transcriptome RNA-seq and ChIP-seq).

B Some of these genes are expected to be known oncogenes, while others have the potential to be novel oncogenes (discovery).

C It is likely that these specific alterations are the causative factors for the cis deregulation of adjacent genes (functional experiments).

D Based on these alterations, it is possible to stratify patients into groups with similar prognoses.

Hypothesis

A By taking a pan-genomic approach, it is possible to identify numerous intergenic alterations correlated with the cis deregulation of adjacent genes (correlation between transcriptome RNA-seq and ChIP-seq).

B Some of these genes are expected to be known oncogenes, while others have the potential to be novel oncogenes (discovery).

C It is likely that these specific alterations are the causative factors for the cis deregulation of adjacent genes (functional experiments).

D Based on these alterations, it is possible to stratify patients into groups with similar prognoses.

E The characterization of the deregulation mechanisms and the discovered oncogenes should help identify vulnerabilities that can be targeted by treatment.

Hypothesis

A By taking a pan-genomic approach, it is possible to identify numerous intergenic alterations correlated with the cis deregulation of adjacent genes (correlation between transcriptome RNA-seq and ChIP-seq).

B Some of these genes are expected to be known oncogenes, while others have the potential to be novel oncogenes (discovery).

C It is likely that these specific alterations are the causative factors for the cis deregulation of adjacent genes (functional experiments).

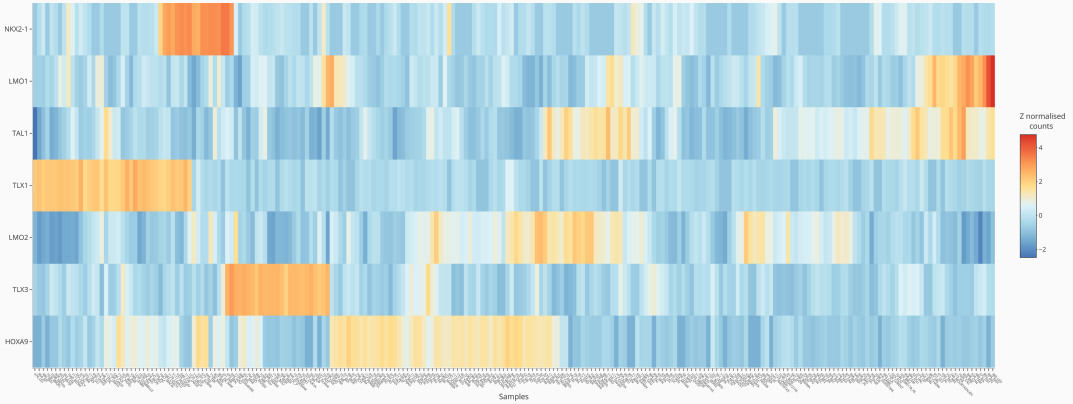
D Based on these alterations, it is possible to stratify patients into groups with similar prognoses.

E The characterization of the deregulation mechanisms and the discovered oncogenes should help identify vulnerabilities that can be targeted by treatment.

F This treatment may prove to be more effective with fewer side effects compared to the currently prescribed polychemotherapy.

✅ What's already done

✅ Alignement and copy number analysis of more than 260 RNA-seq



✅ What's already done

- ✅ Alignement and copy number analysis of more than 260 RNA-seq
- ✅ Alignement and copy number analysis of 72 ChIP-seq (H3K4me4 and H4K27ac).

✅ What's already done

- ✅ Alignement and copy number analysis of more than 260 RNA-seq
- ✅ Alignement and copy number analysis of 72 ChIP-seq (H3K4me4 and H4K27ac).
- ✅ We have developed a high-performance tool that optimizes memory usage and speed for correlating variably expressed genes with depth of ChIP-seq peaks. Specifically, our tool efficiently handles a large dataset consisting of 10,036 gene expressions and 84,839 ChIP-seq positions, resulting in a total of 851,444,204 correlations.

✅ What's already done

- ✅ Alignement and copy number analysis of more than 260 RNA-seq
- ✅ Alignement and copy number analysis of 72 ChIP-seq (H3K4me4 and H4K27ac).
- ✅ We have developed a high-performance tool that optimizes memory usage and speed for correlating variably expressed genes with depth of ChIP-seq peaks. Specifically, our tool efficiently handles a large dataset consisting of 10,036 gene expressions and 84,839 ChIP-seq positions, resulting in a total of 851,444,204 correlations.
- ✅ Calling of genetic alterations sequenced by ChIP-seq.

🎯 First results

With these filters:

- Recurrent mutations in the same ChIP-seq peak (> 1 case) with enrichment of the alternative allele (AF > 0.6).
- Cases with the mutations should have correlated genes (Pearson coefficient > 0.7) significantly upregulated (t-test p value < 0.05)

😵 Caveats

- Most of intergenic alterations are SNPs.

😵 Caveats

- Most of intergenic alterations are SNPs.
- Complexe alterations like indels and SV are difficult to call with ChIP-seq small reads.

😵 Caveats

- Most of intergenic alterations are SNPs.
- Complexe alterations like indels and SV are difficult to call with ChIP-seq small reads.

→ We are implementing longreads sequencing with the Oxford Nanopore Promethion for resolving complex genomic regions, detecting structural variations, and studying repetitive elements.

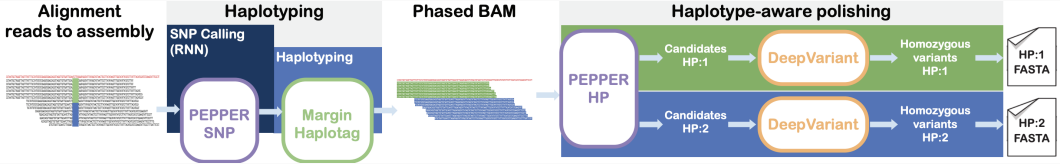
Longreads pipeline

- We will conduct a whole-genome sequencing of 150 T-ALL cases along with their corresponding constitutional samples.

[1] Kolmogorov, Mikhail et al. “Assembly of long, error-prone reads using repeat graphs.” Nature biotechnology vol. 37,5 (2019): 540-546. doi:10.1038/s41587-019-0072-8

Longreads pipeline

- We will conduct a whole-genome sequencing of 150 T-ALL cases along with their corresponding constitutional samples.
- Our objective is to implement a pipeline capable of assembling long reads and generating a diploid reference genome for each of the cases.



[1] Kolmogorov, Mikhail et al. "Assembly of long, error-prone reads using repeat graphs." Nature biotechnology vol. 37,5 (2019): 540-546. doi:10.1038/s41587-019-0072-8

Longreads pipeline

- We will conduct a whole-genome sequencing of 150 T-ALL cases along with their corresponding constitutional samples.
- Our objective is to implement a pipeline capable of assembling long reads and generating a diploid reference genome for each of the cases.
- The alignment of somatic reads on it and the subsequent calling of somatic variants, especially structural variations (SV), will be of significant interest.

Longreads pipeline

- We will conduct a whole-genome sequencing of 150 T-ALL cases along with their corresponding constitutional samples.
- Our objective is to implement a pipeline capable of assembling long reads and generating a diploid reference genome for each of the cases.
- The alignment of somatic reads on it and the subsequent calling of somatic variants, especially structural variations (SV), will be of significant interest.
- By aligning our current RNA-seq and ChIP-seq data using this approach, we will be able to phase gene expression and identify allele-enriched epigenetic marks more efficiently.

Longreads pipeline

- We will conduct a whole-genome sequencing of 150 T-ALL cases along with their corresponding constitutional samples.
- Our objective is to implement a pipeline capable of assembling long reads and generating a diploid reference genome for each of the cases.
- The alignment of somatic reads on it and the subsequent calling of somatic variants, especially structural variations (SV), will be of significant interest.
- By aligning our current RNA-seq and ChIP-seq data using this approach, we will be able to phase gene expression and identify allele-enriched epigenetic marks more efficiently.
- We will also have access to phased methylation of CpG islands with the same technic.

Thank you for listening !
Questions ?