

APPORT DU SÉQUENÇAGE ONT DANS LA DÉTECTION DE VARIANTS DE STRUCTURE ISSUS D'HÉMOPATHIES MYÉLOÏDES MALIGNES

HCL
HOSPICES CIVILS
DE LYON

02/04/25

QUENTIN TESTARD

Cellule de Bioinformatique et service d'Hématologie Biologique et d'Hémostase Clinique

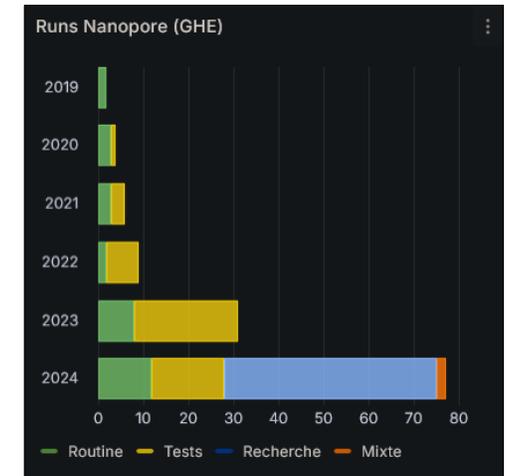
www.chu-lyon.fr

LE SÉQUENÇAGE ONT

ANALYSE DE DONNÉES NANOPORE AU SEIN DE LA CELLULE BIOINFORMATIQUE DES HCL

- En plein essor depuis 2019 aux HCL
- Données issues de **Flongle**, **MinION**, **GridION**, **PromethION** ...

	Flongle	MinION	GridION (5 flow cells)	PromethION (48 flow cells)
				
Maximum run time	16 hours	72 hours	72 hours	64 hours
Theoretical 1D maximum yield	Up to 3.3 Gb	Up to 40 Gb	Up to 200 Gb	Up to 15 Tb
Current 1D maximum yield	Up to 2 Gb	Up to 30 Gb	Up to 150 Gb	Up to 8.6 Tb
Available channels	Up to 126	Up to 512	Up to 2,560	Up to 144,000



Nb de runs ONT traités par la cellule / an

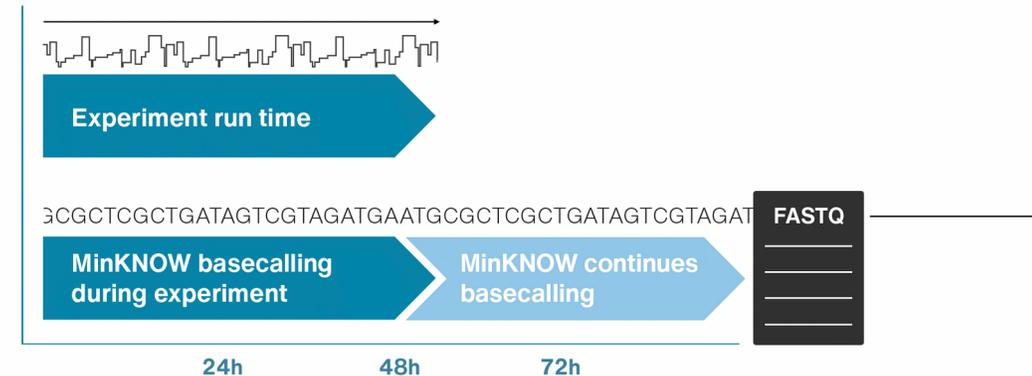
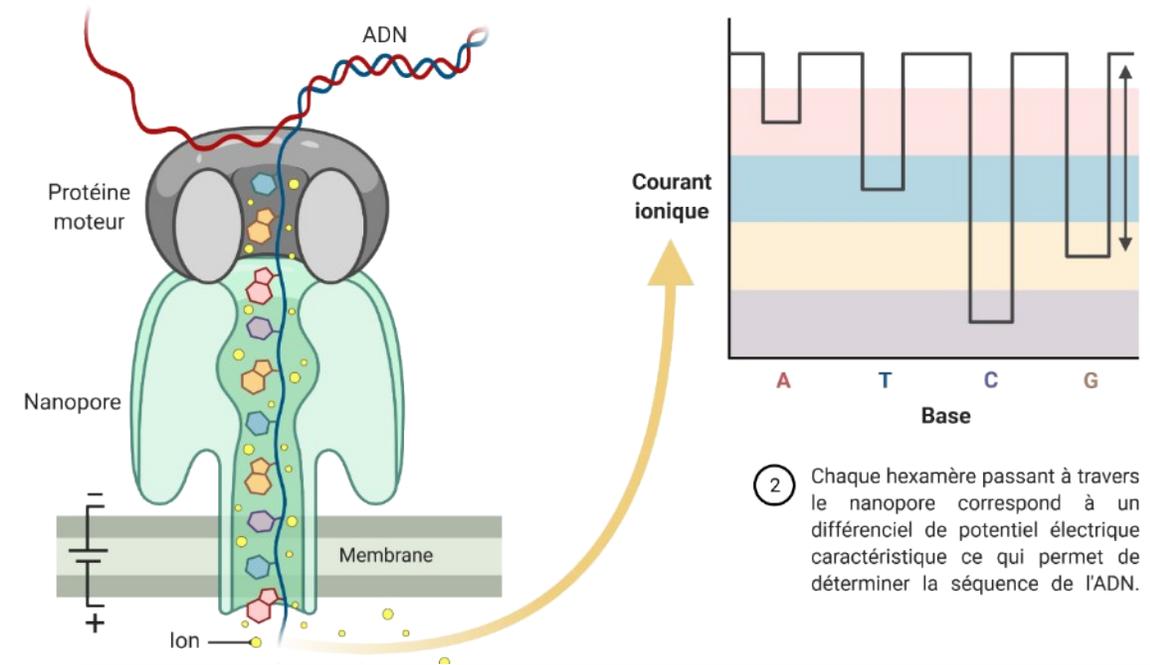
- Un **PromethION 24** depuis septembre 2024

LE SÉQUENÇAGE ONT

DIFFICULTÉS TECHNIQUES SPÉCIFIQUES

- **Données brutes** : fichiers de potentiel électrique (FAST5, POD5...)
 - **Lourdes** (1,5To POD5 flowcell PromethION)
 - **GPU** pour **basecalling** en **FASTQ**
 - **Outils, versions et modèle** de basecalling en **évolution**
- **Garder les données brutes** pour profiter des dernières améliorations ?
- **Limite de temps** pour interopérabilité

① L'ADN est "dézippé" par la protéine motrice. Le brin est guidé à travers le pore vers le côté positif de la membrane.

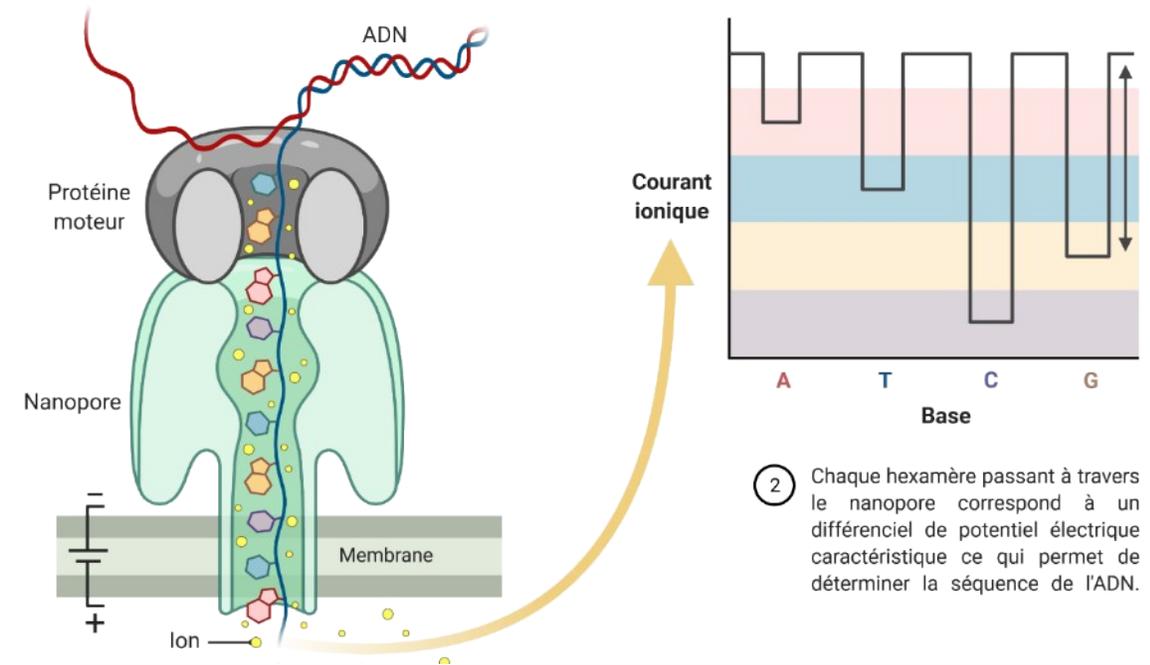


LE SÉQUENÇAGE ONT

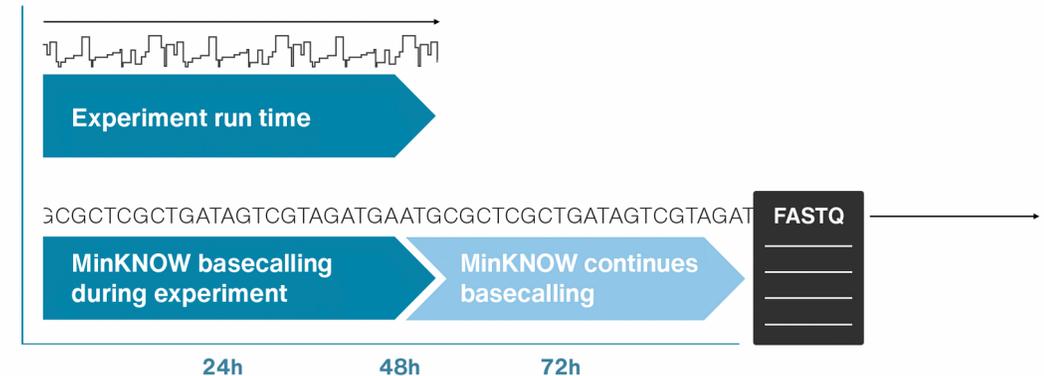
DIFFICULTÉS TECHNIQUES SPÉCIFIQUES

- **Beaucoup de données différentes**
 - Nécessitent **différents outils**
 - **Peu d'outils** validés par la communauté
 - Encore **buggés** / en **développement**
 - **Optimisation de paramètres**
- **Technologie nouvelle avec encore des erreurs systématiques :**
 - Utilisation **modèles d'apprentissage** (parfois **IA**)
 - Nécessite **GPU**
 - Pas toujours la bonne version du modèle

① L'ADN est "dézippé" par la protéine motrice. Le brin est guidé à travers le pore vers le côté positif de la membrane.

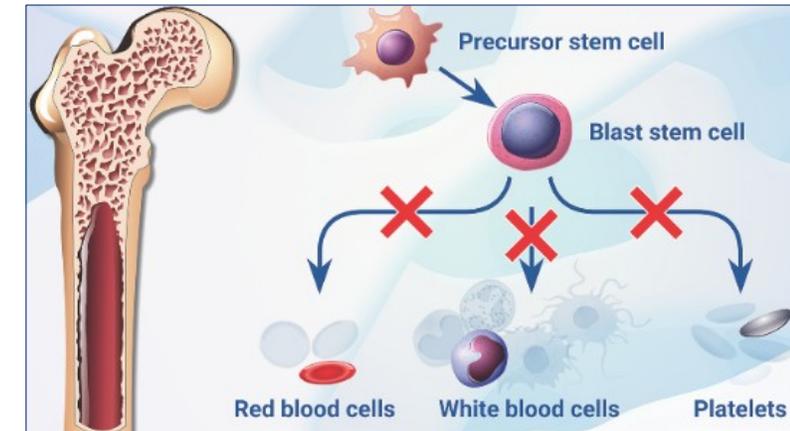


② Chaque hexamère passant à travers le nanopore correspond à un différentiel de potentiel électrique caractéristique ce qui permet de déterminer la séquence de l'ADN.



DÉFINITION

- **Cancers du sang affectant cellules souches myéloïdes de la moelle osseuse** (cellules progénitrices sanguines).
- Plusieurs types dont les **Syndromes Myélodysplasiques (SMD)** ou les **Leucémies Aiguës Myéloïdes (LAM)**
- **Syndromes myélodysplasiques (SMD)** caractérisés par :
 - **Prolifération clonale** de cellules myéloïdes anormales **non différenciées (blastes < 20%)**
 - Atteinte médullaire avec retentissement sur l'hémogramme (cytopénies ou hyperleucocytose).
 - **Anomalies de nombre de copies (CNV)**
- **Risque évolutif vers des formes plus agressives : SMD => LAM**



Source : MDS fondation

LEUCÉMIES AIGUËS MYÉLOÏDES (LAM)

- Caractérisées par :
 - **> 20% blastes**
 - **Dysfonctionnement de la production des cellules sanguines** (anémie, neutropénie, thrombopénie)
 - **Présence d'anomalies génétiques :**
 - **Anomalies de nombre de copies (CNV)**
 - **Variants structuraux (SV) spécifiques classifiants** : permettent la définition de la sous-entité, même **< 20 % blastes**

Table 7. Acute myeloid leukaemia.

Acute myeloid leukaemia with defining genetic abnormalities
Acute promyelocytic leukaemia with <i>PML::RARA</i> fusion
Acute myeloid leukaemia with <i>RUNX1::RUNX1T1</i> fusion
Acute myeloid leukaemia with <i>CBFB::MYH11</i> fusion
Acute myeloid leukaemia with <i>DEK::NUP214</i> fusion
Acute myeloid leukaemia with <i>RBM15::MRTFA</i> fusion
Acute myeloid leukaemia with <i>BCR::ABL1</i> fusion
Acute myeloid leukaemia with <i>KMT2A</i> rearrangement
Acute myeloid leukaemia with <i>MECOM</i> rearrangement
Acute myeloid leukaemia with <i>NUP98</i> rearrangement
Acute myeloid leukaemia with <i>NPM1</i> mutation
Acute myeloid leukaemia with <i>CEBPA</i> mutation
Acute myeloid leukaemia, myelodysplasia-related
Acute myeloid leukaemia with other defined genetic alterations
Acute myeloid leukaemia, defined by differentiation
Acute myeloid leukaemia with minimal differentiation
Acute myeloid leukaemia without maturation
Acute myeloid leukaemia with maturation
Acute basophilic leukaemia
Acute myelomonocytic leukaemia
Acute monocytic leukaemia
Acute erythroid leukaemia
Acute megakaryoblastic leukaemia

DES GAINS EN PRATICITÉ, RÉOLUTION ET EN TEMPS

Méthode	Résolution	Types d'anomalies détectées	Durée	Culture cellulaire	PCR
Caryotype	Faible (5 à 10 Mbp)	Grandes anomalies chromosomiques (translocations, délétions, duplications)	7 à 22 jours	Oui	Non
FISH	Moyenne (~100 kbp)	Anomalies ciblées (requiert une sonde spécifique)	3 à 7 jours	Oui	Oui
Séquençage Nanopore	Haute (1bp à plusieurs Mbp)	CNVs de toute taille, mutations ponctuelles, insertions/délétions, fusions de gènes, anomalies de méthylation	Extraction d'ADN + préparation de librairie ~ 1 jour, séquençage temps réel 1 à 2 jours, transfert et analyse bioinfo ~ 1 jour (~ 3 – 4 jours total)	Non	Non

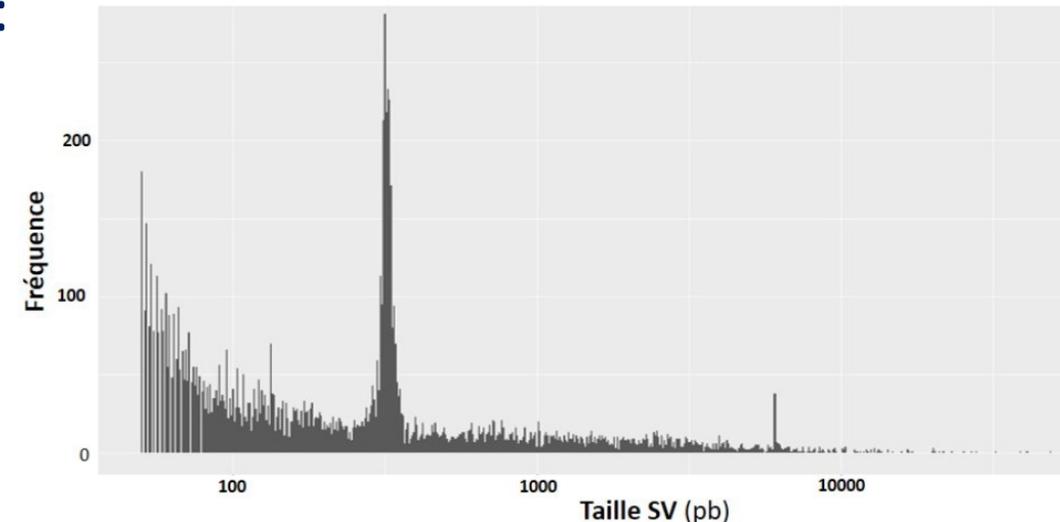
D'ABORD SUR GRIDION

- Détection de **CNV** pour **Syndromes Myélodysplasiques**
- Détection de **CNV** et de **SV** pour **Leucémies Aiguës Myéloïdes**
- Essais d'abord sur **GridION** :
 - **WGS** basse couverture pour **CNV** => **OK**
 - **Adaptive Sampling** sur GridION (**SV** sur liste de **gènes ciblés**) => **OK**
- **Pipeline « historique »** des équipes bioinformatiques de l'**Hôpital de la Charité de Berlin** :
 - Environnement conda, scripts bash, script R, exécutables
 - **NGMLR** + **Sniffles** pour SV
 - **ACE** pour CNV
 - Lancé sur machines « modeste » => 24h de calcul / run
- Anticipation d'**activité 1500+ patients / an** => **PromethION**
 - Pipeline **peu maintenable / optimisé** pour **PromethION**
 - **Redévelopper** un **pipeline** aux **standards actuels**

BENCHMARKS PRÉALABLES

- **HG002 GIAB**, set de référence **SV** en hg19
- Données **PromethION** 05/23 Nanopore **R10.4.1**
- Alignement 64X
 - **Sous échantillonnage** FASTQ seqtk => 10, 15, 20, 30X
- Benchmark d'outils de **SV calling** « généralistes » :
 - **Sawfish** (PacBio)
 - **Sniffles**
 - **SVIM**
- Benchmark d'outils de **CNV calling** :
 - **HiFiCNV** (PacBio)
 - **Spectre**

Taille des SV (pb)	Nb SV
50 à 99	2955
100 à 299	2349
300 à 999	3057
1000 à 5 000	1019
5 000 à 10 000	208
10 000 à 50 000	53
50 000+	5
Total	9646
Médiane	232
Moyenne	706



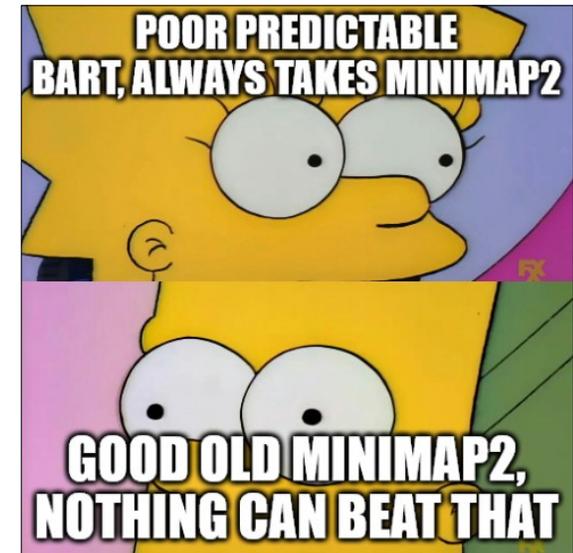
Description des SV contenus dans le set de référence HG002 GIAB

LE CAS SAWFISH



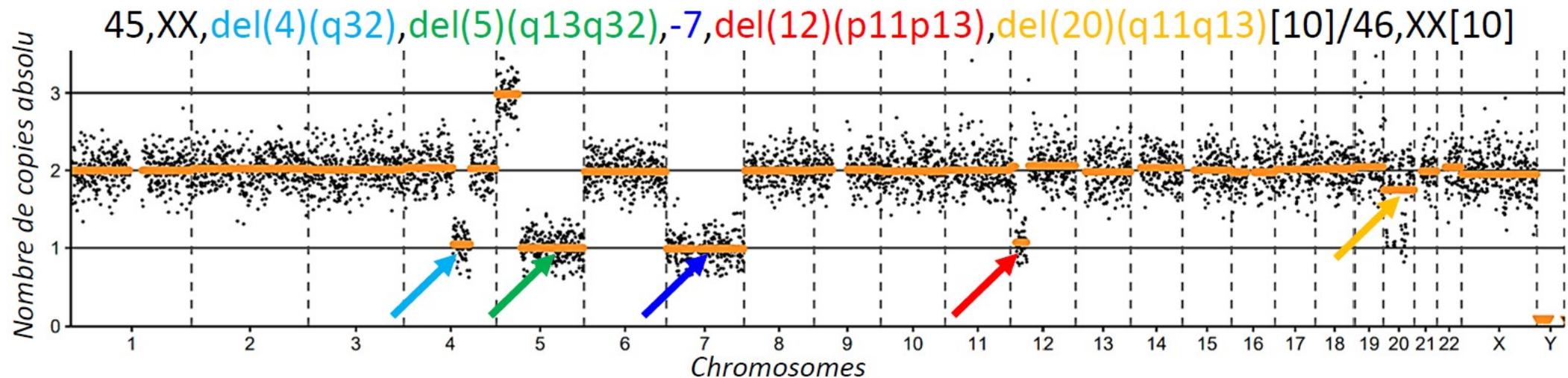
Logo Sawfish

- Nouvel outil de SV calling de PacBio
- Validé avec **pbmm2** (wrapper **minimap2** de PacBio), ne fonctionne pas avec *minimap2 -ax map-ont*
- **Sawfish** > **Sniffles** et **SVIM** avec **pbmm2**
- Problème : impossible d'avoir des **BAM** issus de séquenceurs ONT avec des **tags** de **méthylation** avec **pbmm2**, contrairement à **minimap2**
- **Sawfish** fonctionne en théorie avec n'importe quel aligner dont alignements supplémentaires sans *hardclip* => *Minimap_SV*
- **Sawfish** > **Sniffles** et **SVIM** avec **Minimap_SV**



ACE (ABSOLUTE COPY NUMBER ESTIMATION)

- Détection de **CNV** dans données **WGS** basse couverture
- Package Bioconductor basé sur **QDNAseq**
- Initialement dans le **pipeline** de l'**Hôpital de la Charité** => **bonnes performances**



Profil du nombre de copies absolu pour chaque segment de 500kbp, obtenu par nanopore pour un patient atteint de SMD, en comparaison à la formule chromosomique obtenue en cytogénétique. Thèse A. Salotti

PROMETHION

- **VDM** détection de **CNV** sur échantillon **Syndromes Myélodysplasiques** :
 - **25 runs de séquençage** prévus
 - **11 séquencés et analysés**
- Analyse sur serveur **64 coeurs** 128 threads, **886Go RAM** :
 - **Moins de 2 heures** pour 10/11 runs **sans phasing depuis FASTQ**
 - 24+ heures avec phasing

GRIDION

- Réanalyse de **données GridION** pour **SV + CNV** données **Leucémies Aiguës Myéloïdes**
 - **9 runs de GridION réanalysés** en **moins de 2 heures sans phasing depuis FASTQ**
 - 6 en Adaptive Sampling
 - 3 WGS basse couverture

DÉTECTION DES ANOMALIES POUR DIAGNOSTIC DE SMD

- **11 runs lus** par l'équipe médicale, **70 patients recrutés**
- **99 anomalies connues** préalablement car **détectées par caryotype**

➤ **88 anomalies retrouvées (89%)**

➤ **11 manquantes**

❑ **5 patients avec moins de 5 Gbases séquencées**

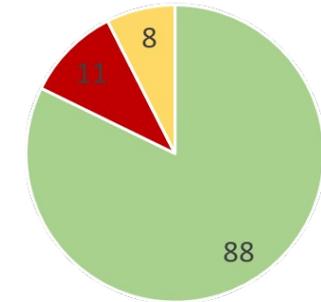
❑ **8 avec caryotype < 5 mitoses / 20**

❑ **1 avec caryotype < 6 mitoses / 20**

❑ **1 del(3q) + del(3p) = anomalie équilibrée**

- **8 anomalies supplémentaires**

Répartition des anomalies



■ Trouvées ■ Manquantes ■ Supplémentaires

Source M. Bonnot, interne HCL

➤ SV à faible %, limite de sensibilité

VDM PROMETHION

ANALYSE DE L'IMPACT PATIENT

- **70 patients**
 - **56** pour lesquels les **mêmes anomalies** sont retrouvées, donc **même conclusion**
 - **8** pour lesquels il **manque une anomalie** ou plus
 - ❑ **Pas d'impact cliniquement significatif** pour les **8 patients**
 - **6** avec **une anomalie supplémentaires** ou plus
 - ❑ Seul **1 changement de pronostic** : défavorable => très défavorable.
 - ❑ Le reste : **caryotype déjà complexe**

Impact patient ?	Anomalie technique ?
Nb de bases insuffisant, aurait été repassé, pas d'impact (+ caryotype retenu normal par les cliniciens)	Oui
Pas d'impact	
Problème de run, aurait été repassé, pas d'impact	Oui
Evolution donale, pas d'impact	
Problème de run, aurait été repassé, pas d'impact	Oui
Problème de run, aurait été repassé, pas d'impact	
Pas d'impact	
Pas d'impact	

Source M. Bonnot, interne HCL

DIAGNOSTIC LAM

- Adaptive Sampling
- Réanalyse en cours
- Résultats pipeline historique
- Certains SV loupés par NGMLR + Sniffles vus par Minimap2_SV + Sawfish (et inversement)
- Tous les CNV détectés sauf un
- Deux SV manqués car absents de la liste de gènes d'intérêt

N° access	Caryotype / anomalie à rechercher	Résultat (anomalie détectée 1/0)
APL256a	46,XY,t(15;17)(q22;q12-21)[13]/46,idem,del(7)(q21q31)[6]/46,XY[1].nuc ish(PMLx3,RARAx3)PMLconRARAx2)[97/100] BCR3 (exon 3- exon 3)	1
P1456a	t(9;22)	1
L3656a	inv(16) atypique 47,XX,del(2)(q33),del(3)(q13q23),+8,der(16)inv(16)(p13q22)del(16)(p13)del(16)(q22q23)[14]/46,XX[6].	1 sur bam uniquement
L3746a	Echec / MLL-dup / muté IDH2 (vaf 46%) et DNMT3A (vaf 45%)	1
L3186a	46,XX,t(8;21)(q21;q22)[6]/45,idem,-X[4]/47,idem,+15[4]/48,idem,+15,+m[6]/46,XX[3]. nucish(RUNX1T1x3,RUNX1x3,RUNX1T1conRUNX1x2)[95/100], (MECOM,KMT2A,MYH11,CBFB)x2[200]	1
L2959a	46,XX[20]. FLT3-ITD : taux 14 ; taille 63 / NPM1 (vaf=52%)	1
M9290a/L1278a	FIP1-PDGFR	1
L3804a	46,XY,inv(16)(p13q22)[8]/46,idem,t(4;17)(q28;q12)[7]/49,idem,+13,+14,+22[2]/46,XY[3]	inv(16)=1 sur bam uniquement (très bien visible) t(4;17)=0 CNV=0
L1823a	45,XX,-7,der(10)(11pter->11p13::?:10p12->10qter),der(11)t(10;11)(p12;p13)[18]/46,XX[2]. Présence d'un clone hypodiploïde présentant une monosomie 7 associée à une translocation complexe t(10;11) entraînant un réarrangement atypique de KMT2A.	1 déséquilibre 10/11
L3428a	K normal mais en RNAseq TF USP42 ::RUNX1 qui serait issu d'une t(7;21). En FISH, présence d'un faible signal de RUNX1 supplémentaire dans 88% des noyaux analysés, en faveur d'un réarrangement cryptique de ce gène.	1
L3866a	46,XX,t(11;19)(q23;p13)[19]/46,XX[1]	1 sur bam uniquement
L2543b	possible réarrangement de KMT2A au caryotype = t(11;17)? 46,XY,t(11;?17)(q23;?q25)[2]/46,idem,add(9)(q21)[6]/46,XY[10]	Pas de réarrangement de KMT2A
L4143a	Caryotype très complexe présentant un 5q- associé à une translocation déséquilibrée t(5;18) des nullosomie 6, une t(6;9), un remaniement d'un chromosome 15 et 2 marqueurs chromosomiques sur 4 mitoses. Présence d'un sous clone avec une monosomie 7 associée aux précédentes anomalies sur 6 mitoses. L'analyse en FISH interphasique confirme la présence d'un 5q- (8,5% des noyaux) et d'une monosomie 7 (3% des noyaux).	0 (fusion inconnue probablement impossible à voir si gènes pas dans le bed)
L4152a	Présence d'une inversion péricentrique du chromosome 16 sur 17 des 20 mitoses analysées. L'analyse en FISH interphasique confirme la présence d'une fusion CBFB/MYH11 sur 90% des noyaux analysés.	inv(16)=1 sur bam uniquement

CONCLUSION

DES RÉSULTATS ENCOURAGEANTS



- **Détection de la plupart des CNV sur données PromethION pour diagnostic SMD**
 - **Pas de changement d'impact**
 - **Détection de CNV supplémentaires**
 - **Problème sensibilité**
 - **Évaluation de HiFiCNV + Spectre + ACE vs ACE seul**
 - **Toujours une explication biologique ou technique pour les CNV manqués**
- **Détection de la plupart des SV et CNV sur données GridION pour diagnostic LAM**
 - **Amélioration prévue avec nouveau pipeline**
 - **Élargissement de la liste de gènes d'intérêt ?**
 - **Évaluation de l'apport de Sawfish + SVIM + Sniffles vs Sniffles seul**

CONCLUSION

UN TRAVAIL EN COURS

- Analyse et interprétation de **14 runs** de **VDM PromethION** restants
- Relecture de la **totalité** des résultats **GridION**
- Encore du travail sur le pipeline
- En attente de l'ensemble des résultats pour se prononcer sur l'équivalence ou la supériorité du séquençage Nanopore sur les techniques de cytogénétique de référence
- Conditionné au fait que le **PromethION 24** reste aux **HCL**
- Travail au niveau des **HCL** sur la **gestion du volume** grandissant des données **ONT** (**matériel, informatique, bioinformatique et humain**)



MERCI!

Service d'Hématologie biologique et d'hémostase clinique :

Auriane Salotti
Isabelle Mosnier
Aurélie Charansol
Marine Blanchet
Kaddour Chabane
Mathilde Bonnot
Sarah Huet
Pierre Sujobert

Service de Cytogénétique :

Audrey Labalme
Flavie Diguët
Nicolas Chatron

IAI :

Quentin Semanas
Richard Chalvignac
Christophe Ginevra
Noémie Fessy
Éva Oddoux

Cellule de Bioinformatique :

Thomas Simonet
Pierre-Antoine Rollat-Farnier
Sylvain Mareschal
Valentin Wucher
Genna Ben-Hassen
Corentin Molitor
Claire Bardel-Danjean

Charité de Berlin :

Lars Bullinger
Anna Dolnik
Eric Strang



SV CALLING

VERSIONS D'OUTIL

- Minimap2 2.2.28
 - Sniffles 2.4
 - SVIM 2.0.0
 - Sawfish v0.12.6
 - HiFiCNV 1.0.0
 - Spectre 0.2.1
- PBMM2 1.14.99
- Truvari 4.2.2

SV CALLING

TEST D'OUTILS D'APPEL DE SV « GÉNÉRAUX »

- Minimap2
 - Sniffles
 - SVIM (filtration)
 - cuteSV
 - Sniffles + cuteSV (Jasmin)
 - Sniffles + SVIM (filtration)

- PBMM2
 - Sniffles
 - pbsv
 - Sniffles + pbsv (Jasmin)