4ème édition

Séminaire BioInfoDiag

BECQUEREL

Shallow WGS appliqué aux biopsies liquides pour identifier les remaniements de nombre de copies dans les lymphomes

Pierre-Julien VIAILLY, IR, bioinformaticien

Inserm UMR1245 - CBG - Cancer and Brain Genomics, Université de Rouen Normandie

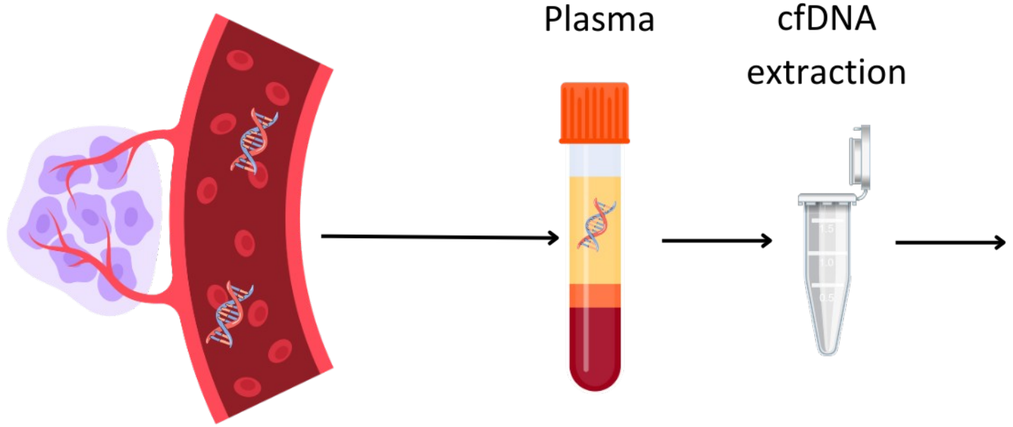
Equipe 2 - Génomique et biomarqueurs des lymphomes et des tumeurs solides

CBG CANCER & BRAIN GENOMICS

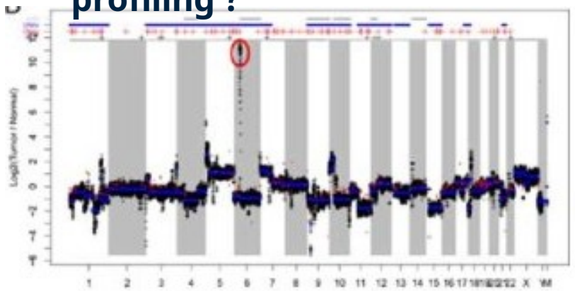
Inserm La science pour la santé From science to health

Séminaire BioinfoDiag, Montpellier– 31 Mars 2025

Summary



Copy number profiling ?

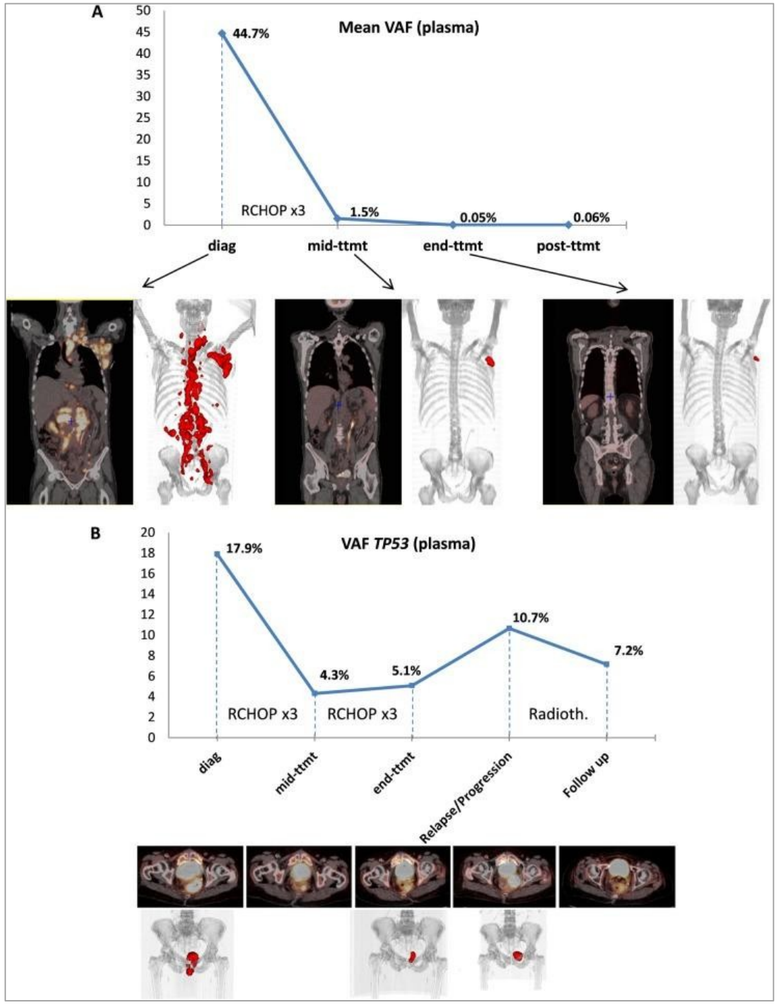


Non-invasive approach

ctDNA has a half-life of 16 minutes to 2.5 hours

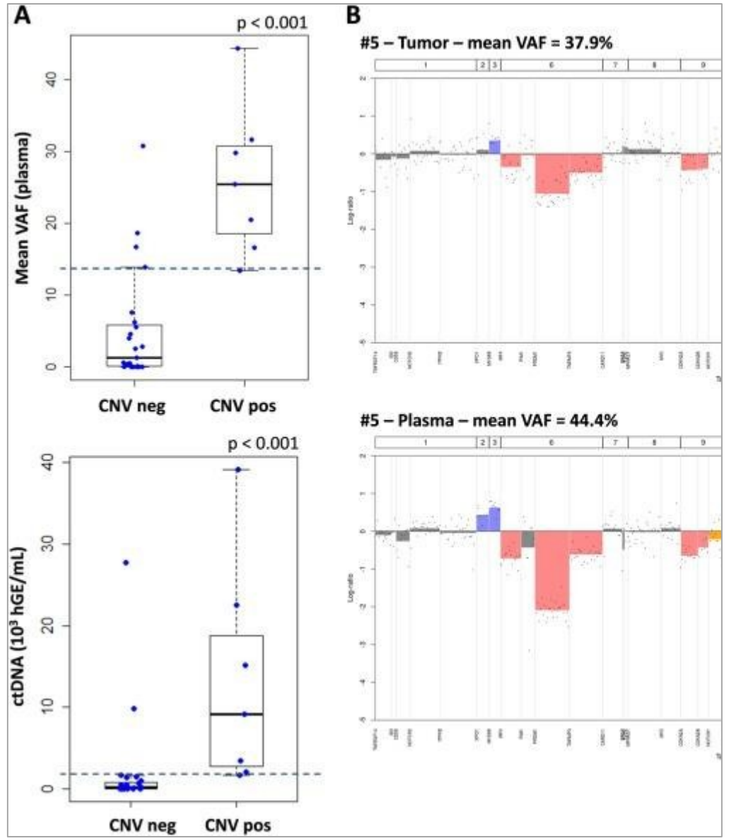
Interesting biological source for minimal residual disease monitoring (MRD) during treatment

ctDNA dynamics estimated by mutation quantification



Some copy number variations of gene detected

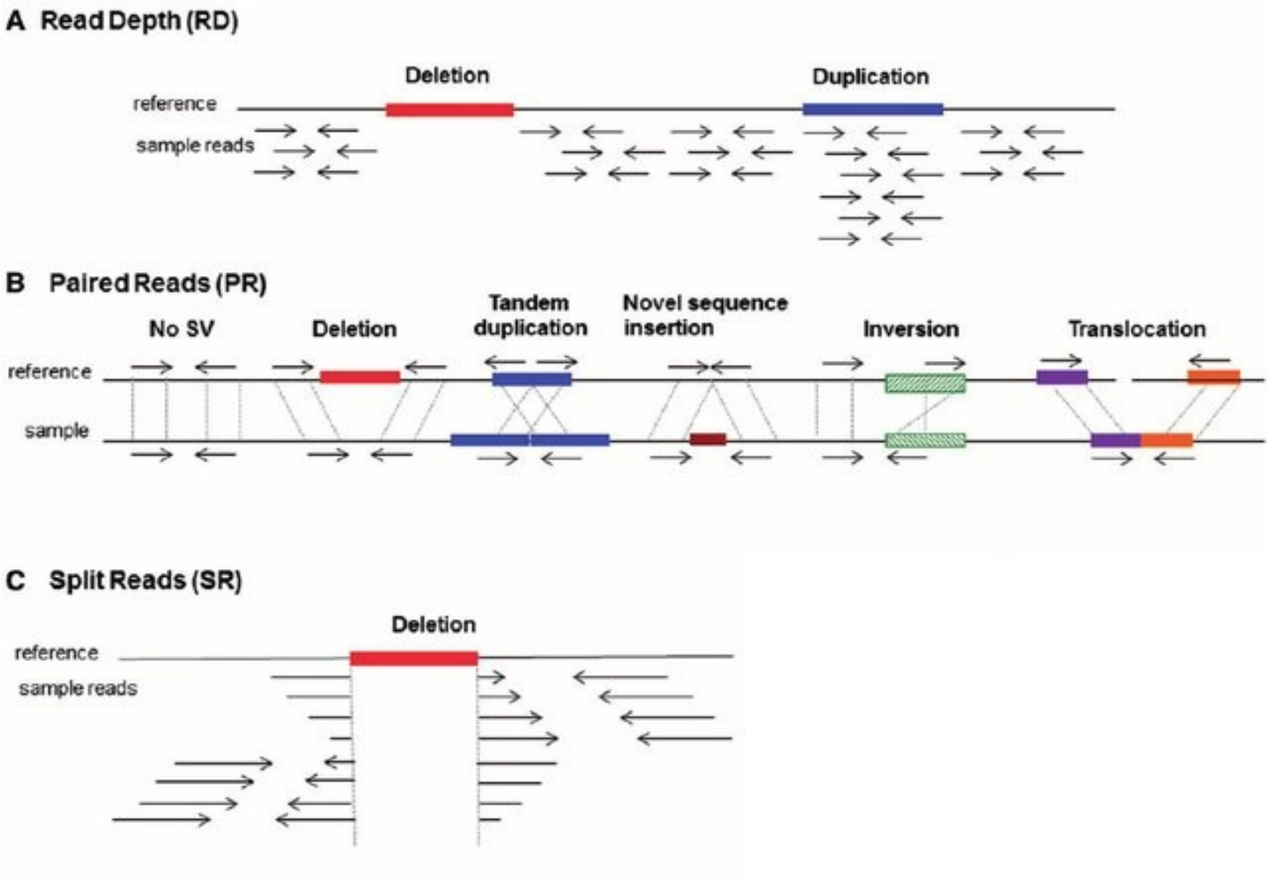
Depending on cfDNA concentration and ctDNA enrichment (VAF)



Bohers E et al. Non-invasive monitoring of diffuse large B-cell lymphoma by cell-free DNA high-throughput targeted sequencing: analysis of a prospective cohort. Blood Cancer J. 2018 Aug PMID: 30069017

3 main approaches :
- read-pair (PR)
- split-read (SR)
- **read-depth / UMI-depth (RD)**

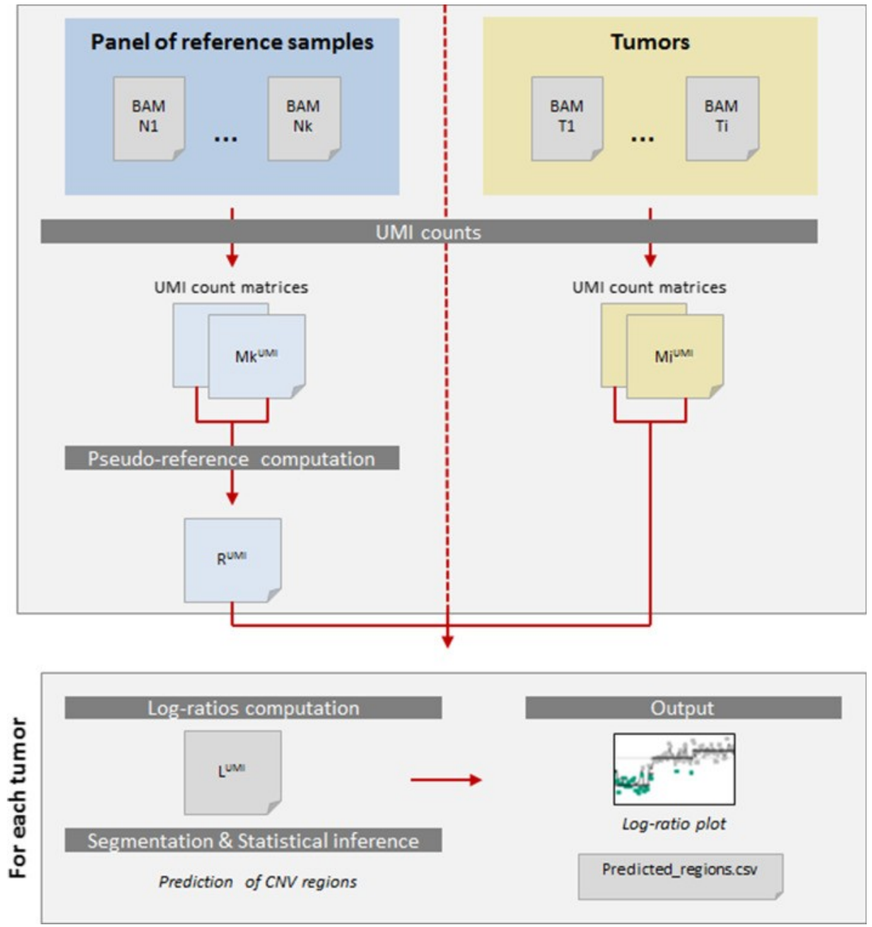
A large number of published RD algorithms :
- CNVnator,
- CNV-seq,
- ONCOCNV
- […]



Geòrgia Escaramís, Elisa Docampo, Raquel Rabionet, A decade of structural variants: description, history and methods to detect structural variation, Briefings in Functional Genomics

CNV estimation / Example of mCNA algorithm

mCNA workflow



Pierre-Julien Viailly et al., Improving high-resolution copy number variation analysis from next generation sequencing using unique molecular identifiers - BMC Bioinformatics . 2021 Mar PMID 33711922

Key points

- Sequencing depth is linked to number of reads overlapping a genomic region
- The tumor fraction in a sample will impact depth changes in the presence of CNV
- The use of UMI counts data makes it possible to overcome PCR biases (%GC, insert size...)

+1 $$L_p^{UMI} = \log_2\left(c \times \frac{3}{2} + (1-c) \times \frac{2}{2}\right)$$

-1 $$L_p^{UMI} = \log_2\left(c \times \frac{1}{2} + (1-c) \times \frac{2}{2}\right)$$

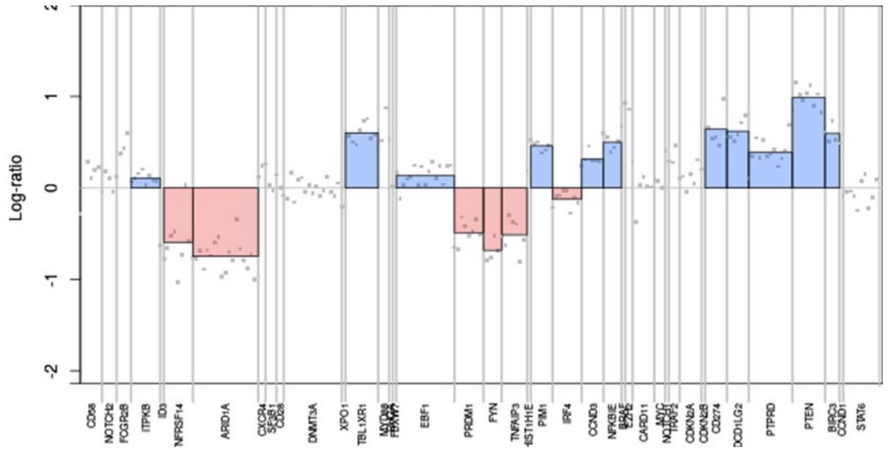
Shallow WGS (sWGS) or low pass WGS (lpWGS) is commonly defined as sequencing a genome to an average depth less than 10X coverage

Read depth algorithms for targeted sequencing

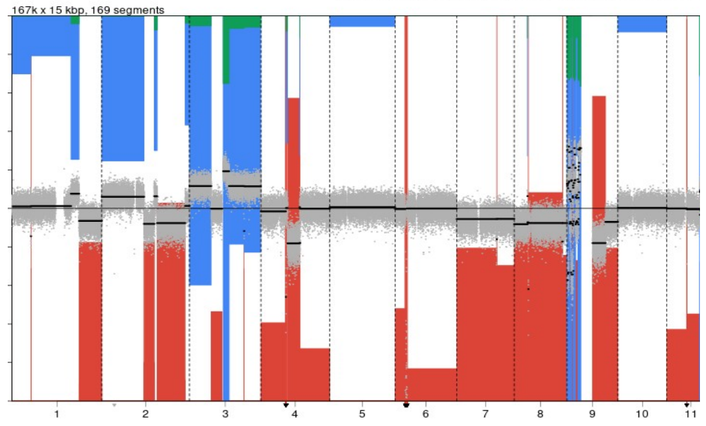
- Average coverage per base of several thousand reads
- Resolution at the level of an exon, a gene

Shallow WGS

- Less than 10 reads aligned per base
- Resolution is a parameter of the analysis (15kb, 30kb…)
- Detection of large CNV at genome scale



Mean UMI coverage : 3000x per base



Mean read coverage : 300x per bins of 15kpb

1 **Total cfDNA extraction, library preparation, sequencing**

- 10 ng of DNA minimum required (experimental threashold)
- ~5-10 millions of clusters / sample

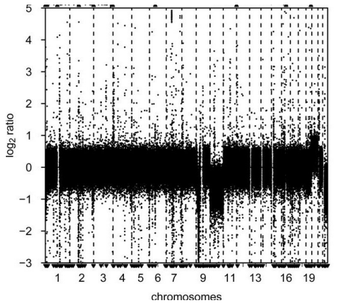
2 **Bioinformatics analyses**

- Alignment (BWA-mem)
- sWGS analysis
  - Raw coverage reads counts computing by bins
  - LOESS regression (%GC, mappability)
  - Excluding blacklisted regions (ENCODE, 1000 Genomes)
  - Circular Binary Segmentation (CBS) + Segment copy estimation

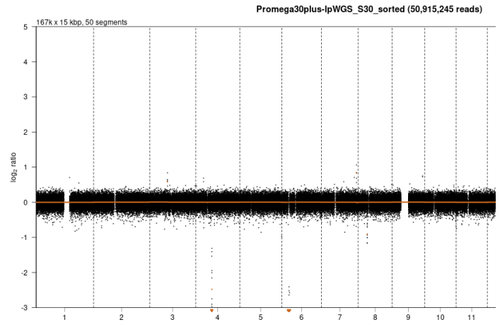
3 **Tertiary analyses**

- ctDNA/tumor comparison
- GISTIC
- Gene annotation
- [...]

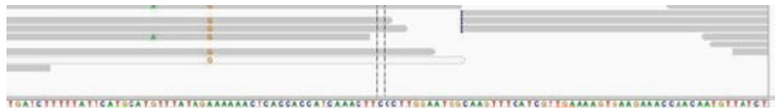
Raw signal



Processed signal



Compute read coverage per bins



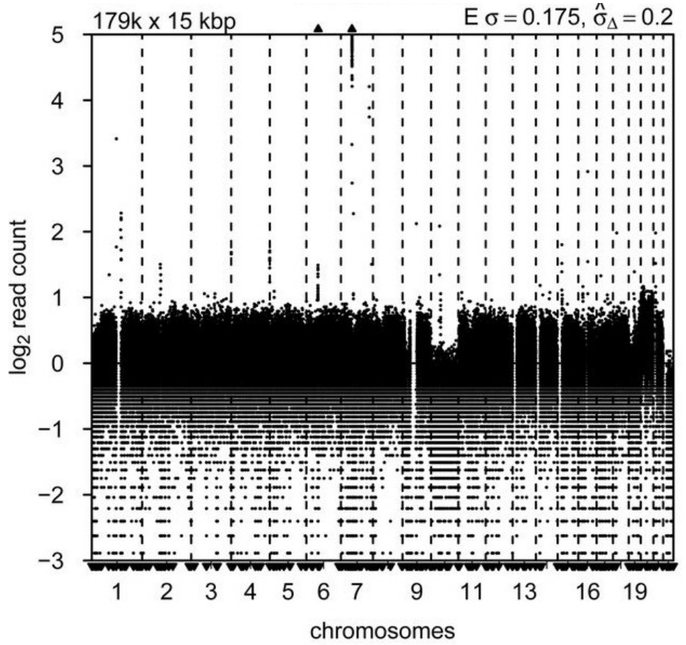
BAM file
Aligned reads

RC per bins

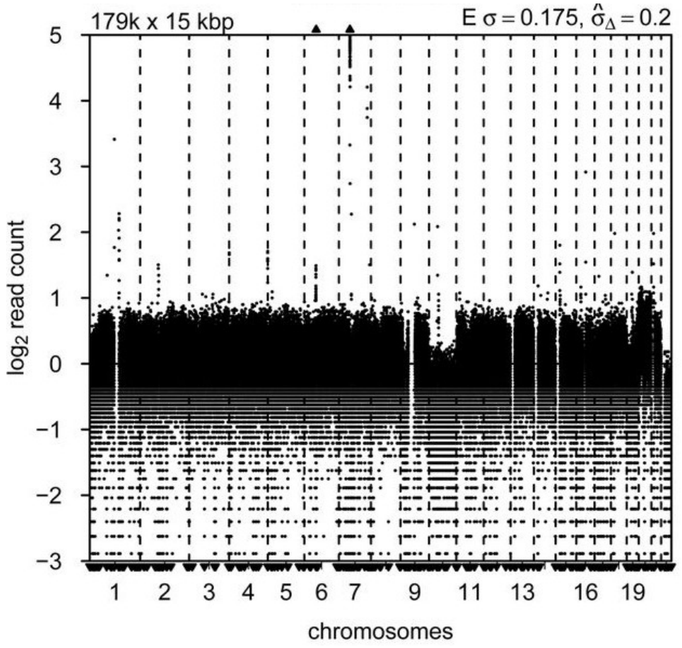
(1) The genome is divided into bins of 15 kb

→ This size defines the resolution of the experiment

(2) The number of reads overlapping each bin is computed



LOESS (**LO**cally **E**stimated **S**catterplot **S**moothing)



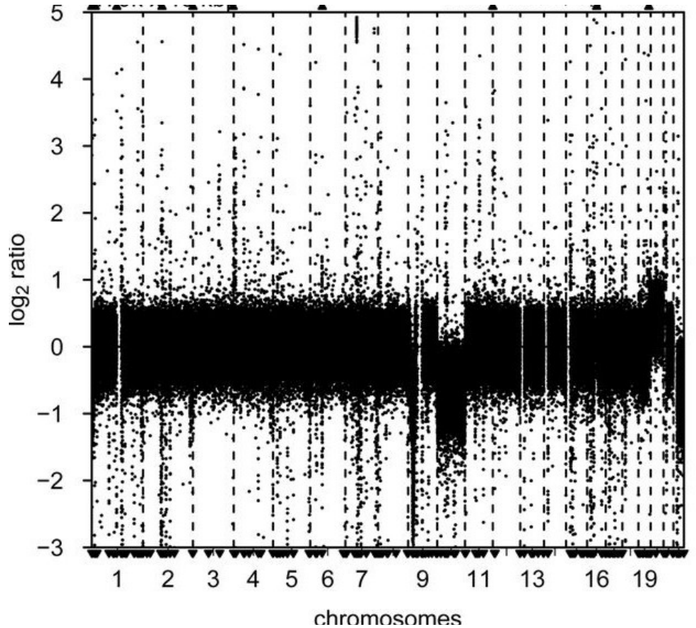
LOESS/Centering

(1) Computing the median read count of all bins with the same combinations of GC and mappability

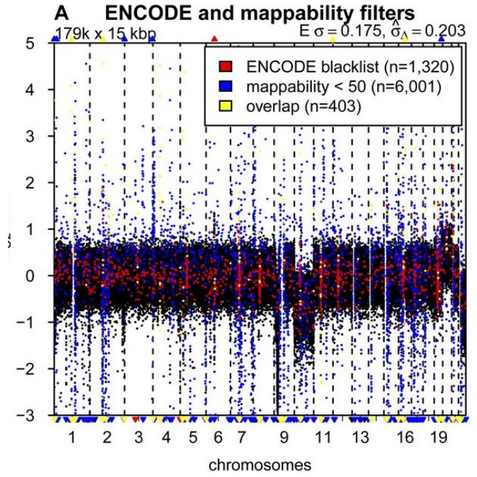
(2) Divide each read count by the LOESS value of its combination of GC and mappability

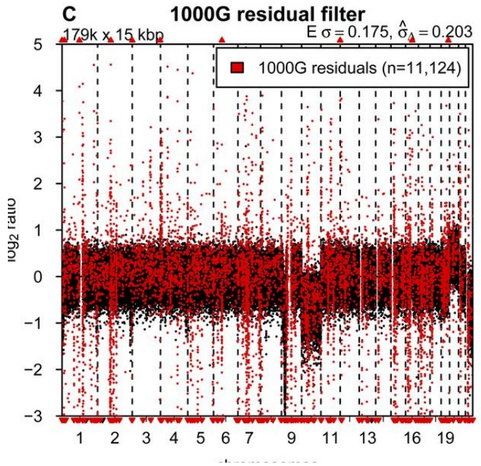
(3) LR= Log2(readCount)/mean(readCounts)

(4) Signal centering (LR=0)



Blacklist filtering

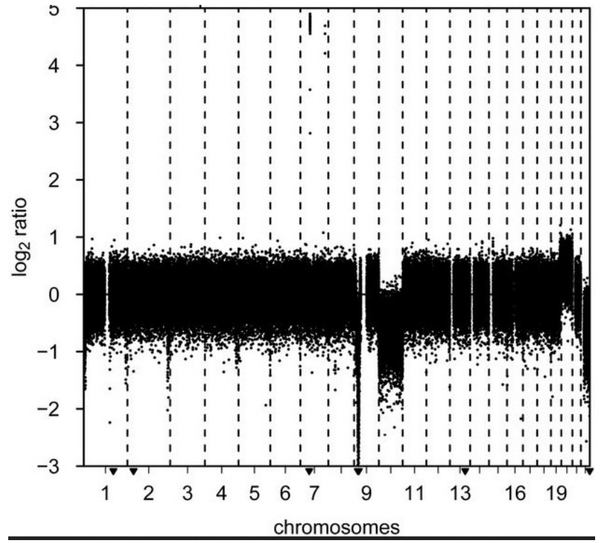




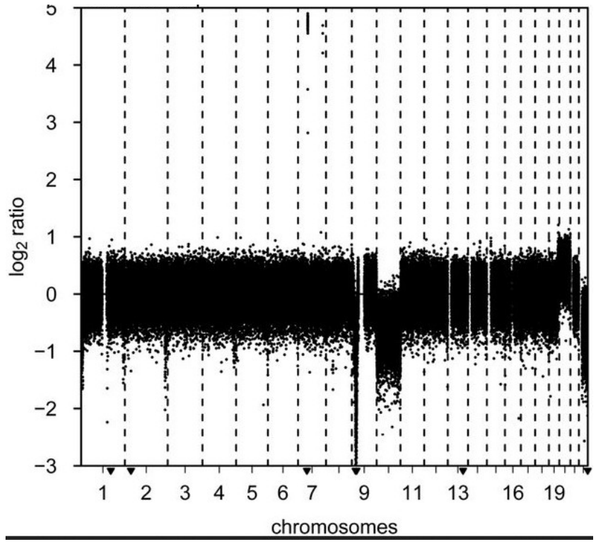
(1) Exclusion of **blacklisted regions from ENCODE** Project Consortium (satellites, centromeric and telomeric repeats…), or with low MPQ, or both

(2) Blacklist based on the **residuals of the 1000 Genomes** samples after LOESS regression

Final signal

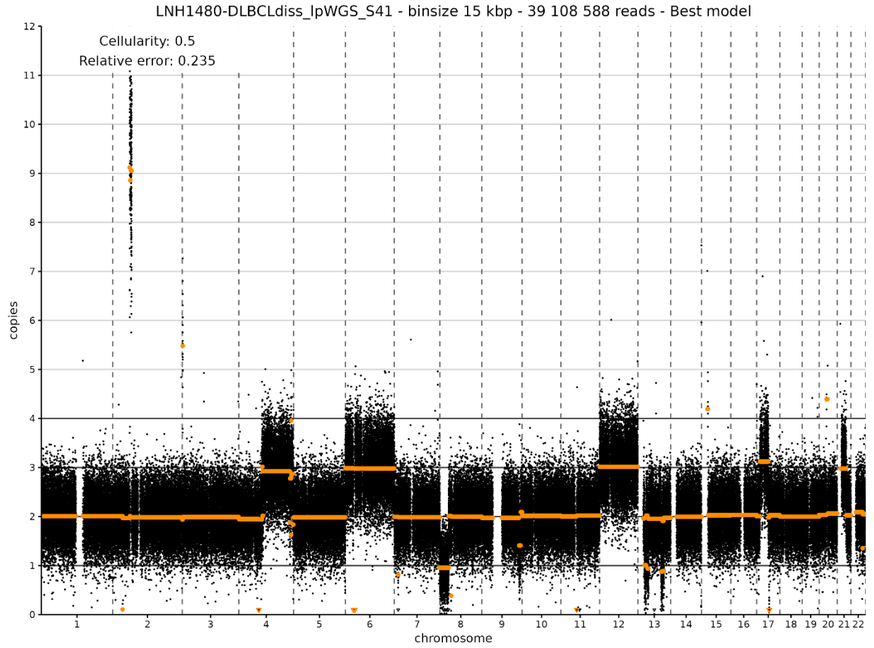


Segmentation and copy number estimation

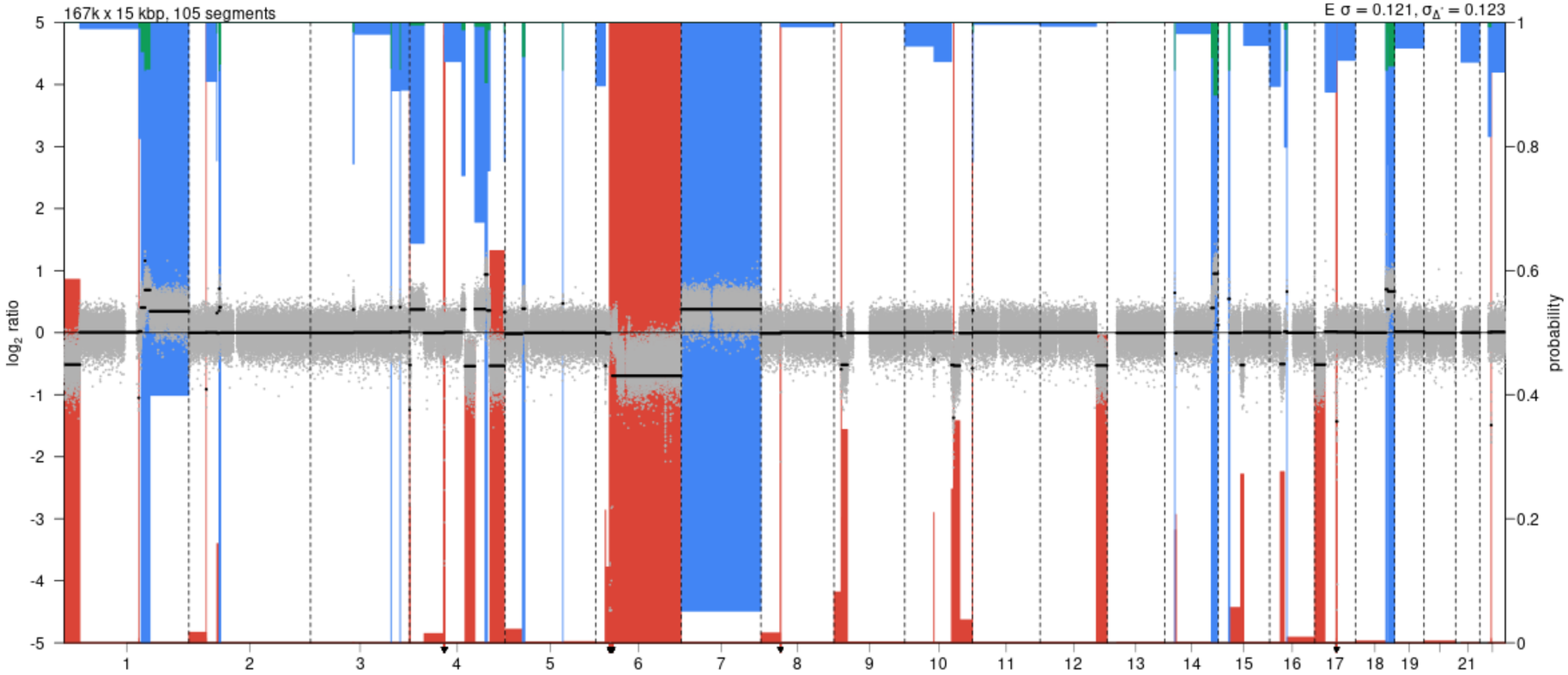


(1) **Segmentation using CBS** algorithm (Circular Binary Segmentation)

(2) **Log-ratios converted to copy number** for each segment using Absolute Copy number Estimation (ACE) algorithm



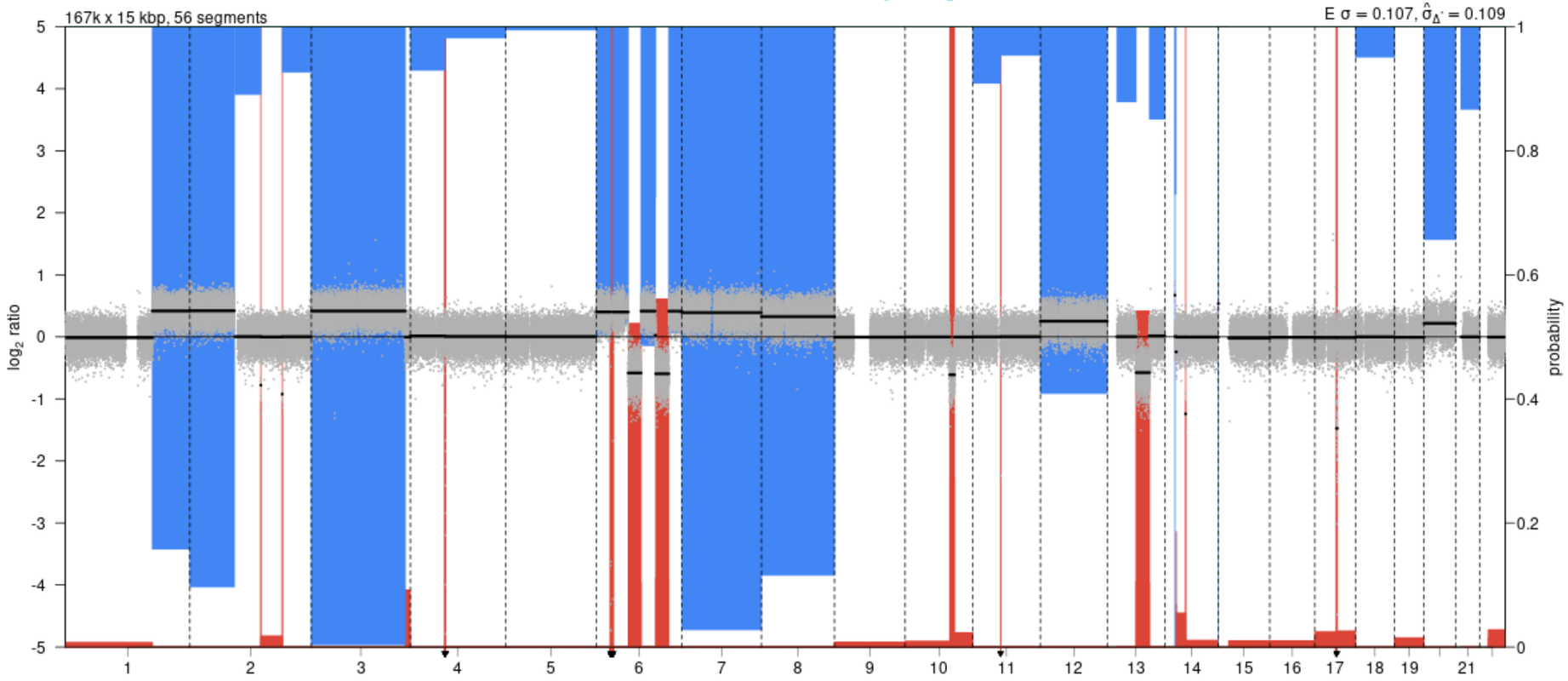
Tumor / Follicular lymphoma



Cytogenetics

46,XX,add(1)(p35)(q21) add(4)(p16)del(32) -6 +7 del(9)(p21) del(12)(q24), del(17)(p12)

Tumor / Follicular lymphoma



Cytogenetics

53,XX t(2;3)(p12;q27)+der(3) t(14;18)(q32;q21)

+6,del(6)(q14q25)x2 +7,+8

+12 del(13)(q12q14)

+20

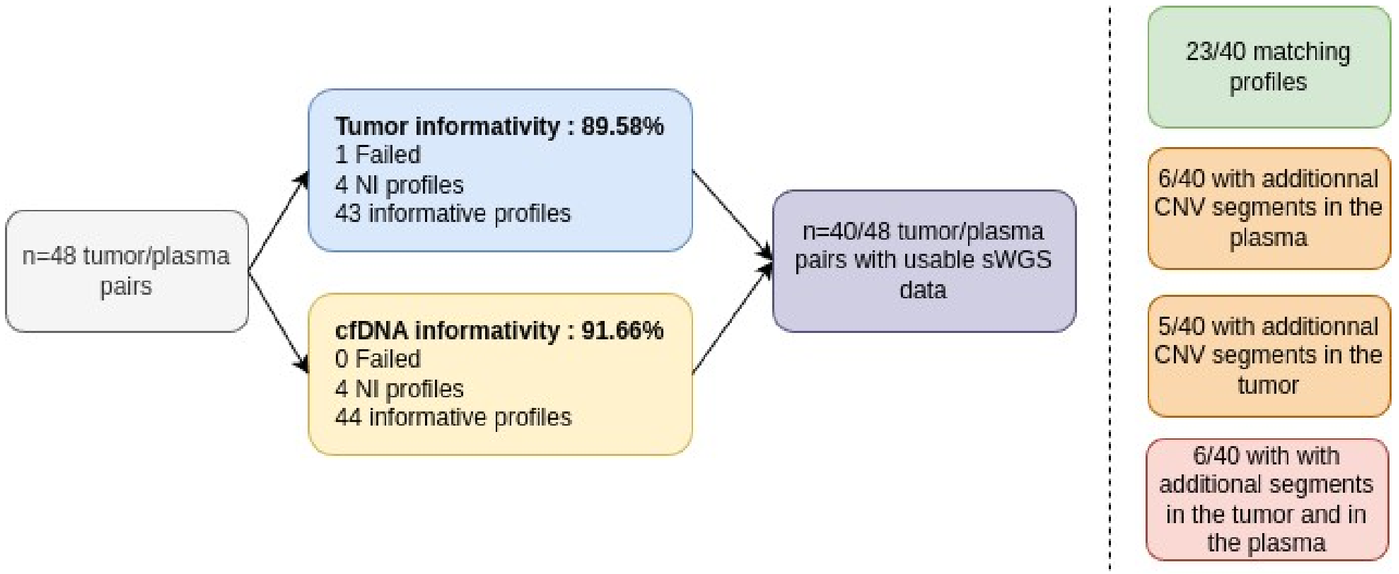
Objectives

To assess clonal heterogeneity between tumor and plasma using sWGS in a population of disseminated diffuse large B-cell lymphomas

Retrospective cohort of 48 patients with:

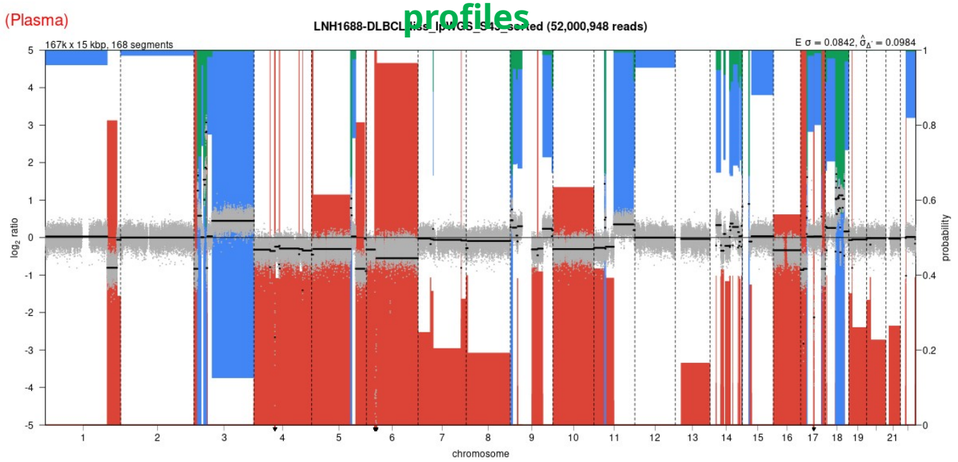
- ≥ 3 invaded extranodal sites
- Tumor/plasma available at diagnosis

sWGS informativity

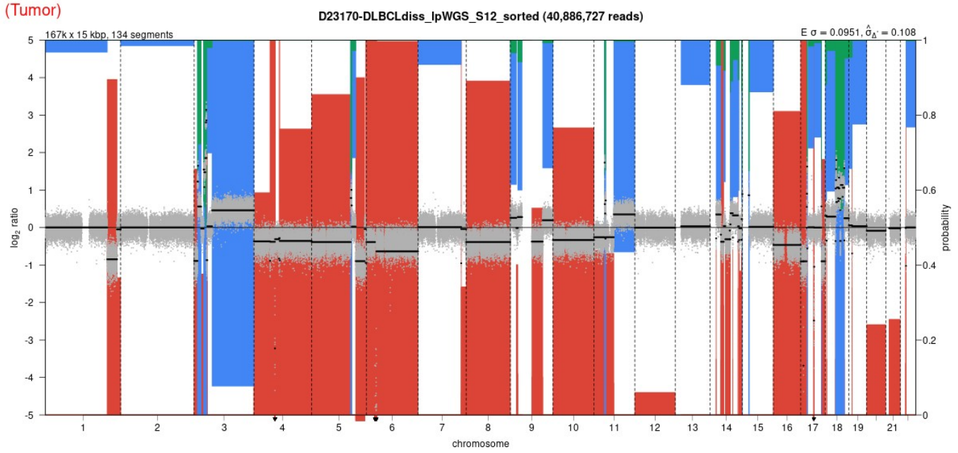


Example of concordant profiles

(Plasma)



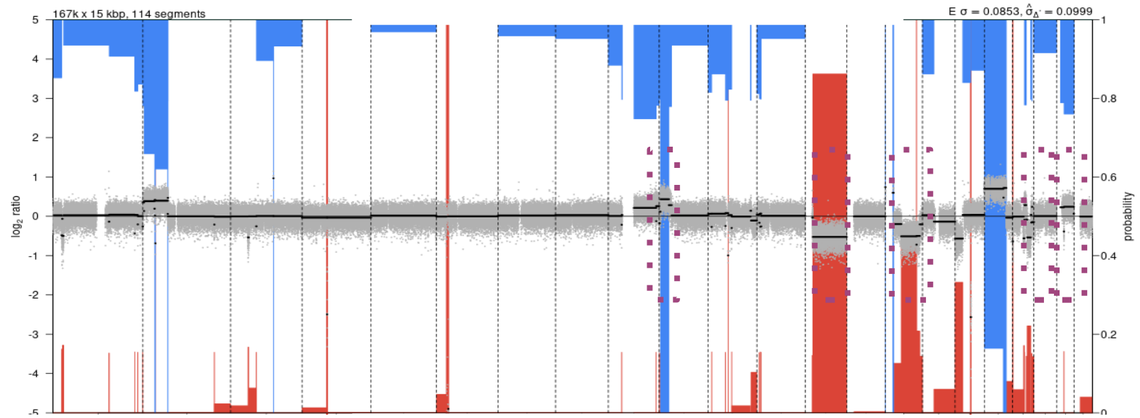
(Tumor)



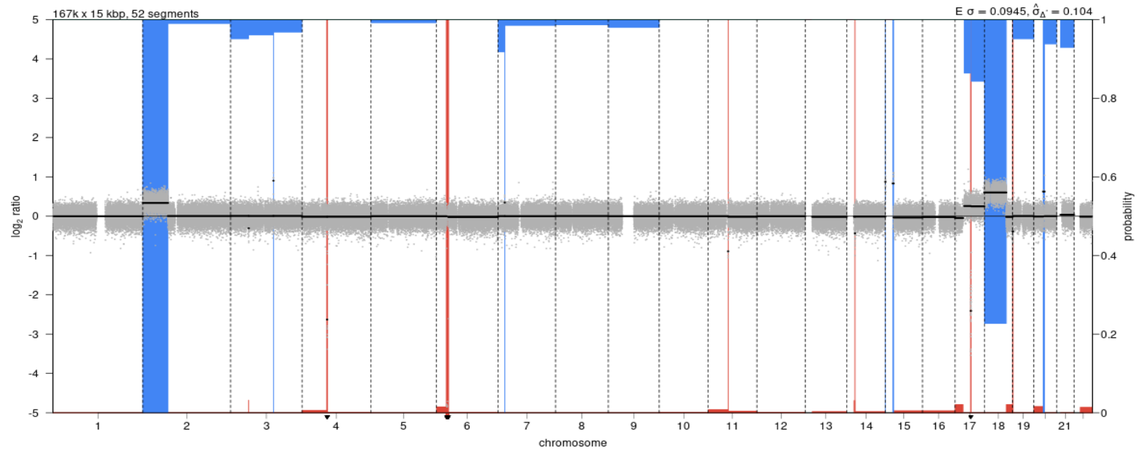
Clonal heterogeneity ?

Example of additional segments in cfDNA sample

(Plasma)



(Tumor)



Objectives

Measure the evolution of the tumor burden (ctDNA) in the plasma during treatment (RCHOP)

- (Using targeted sequencing data)
- Using sWGS
- Using fragmentomics

Predicting good responders from non-responders?

Biological material

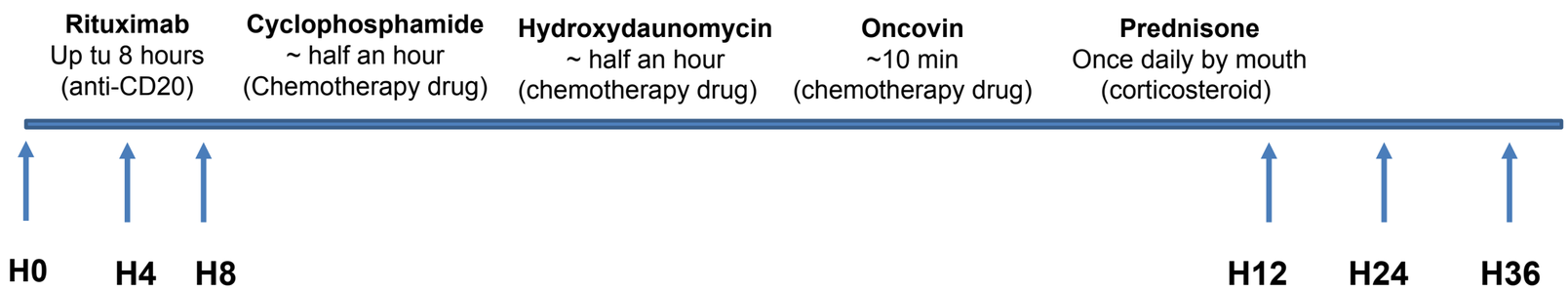
Collection of plasmas at different treatment times (H0, H4, H8, H12, H24, H36…)

N = 24 patients

sWGS at each time point

- Profil positivity
- FCS score

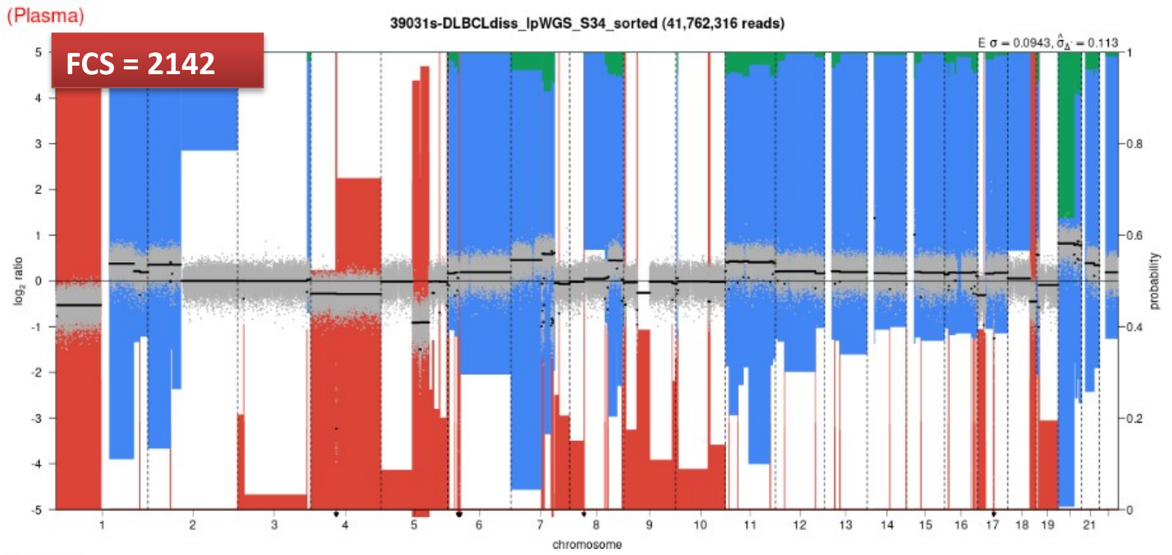
R-CHOP chemotherapy

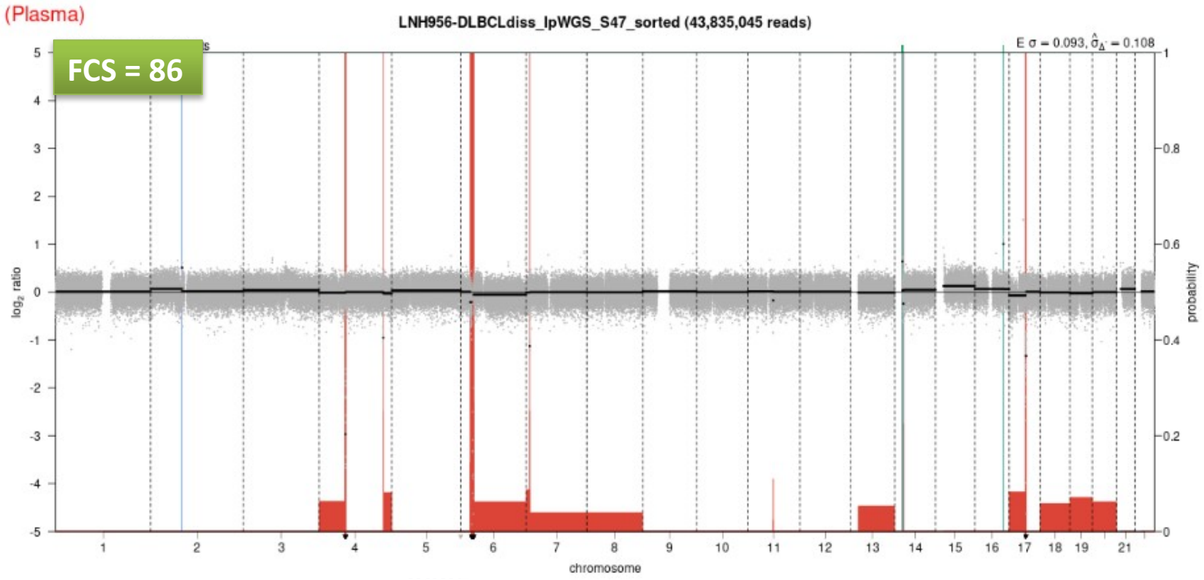


FCS score

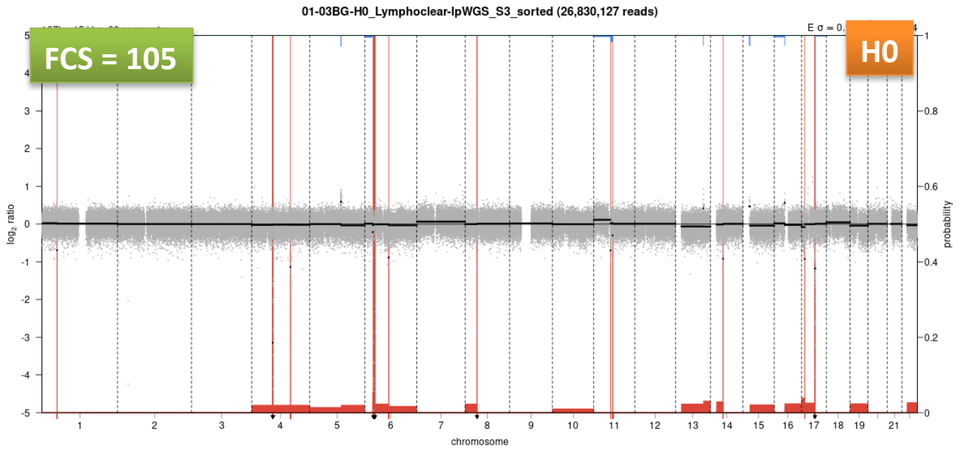
LR	CNA level	A
1	High-level gain	3
0.58	Medium-level gain	2
0.2	Low-level gain	1
– 0.2	Low-level loss	1
-1	Medium-level loss	2
– 1.74	High-level loss	3

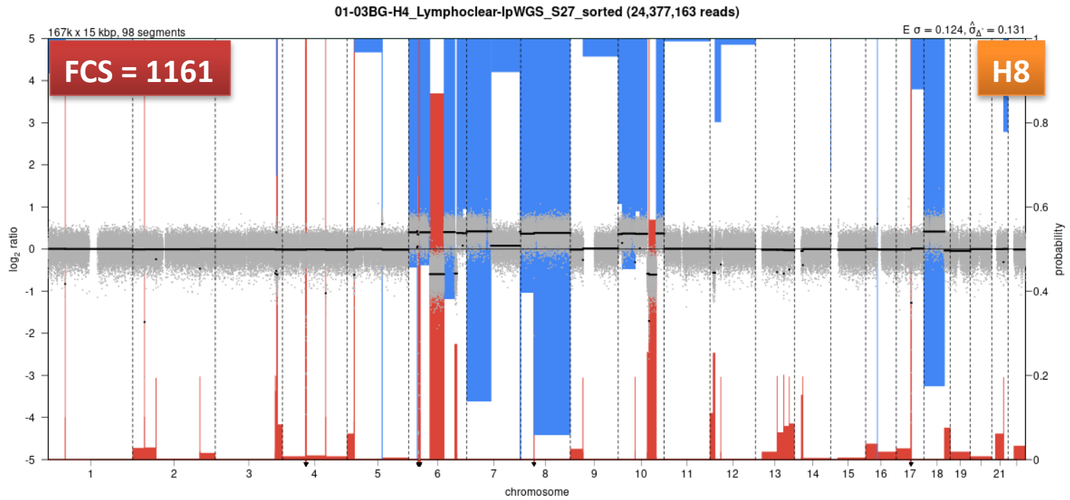
≤ 5%	1
>5% to≤15%	2
>15% to≤30%	3
>30%	4

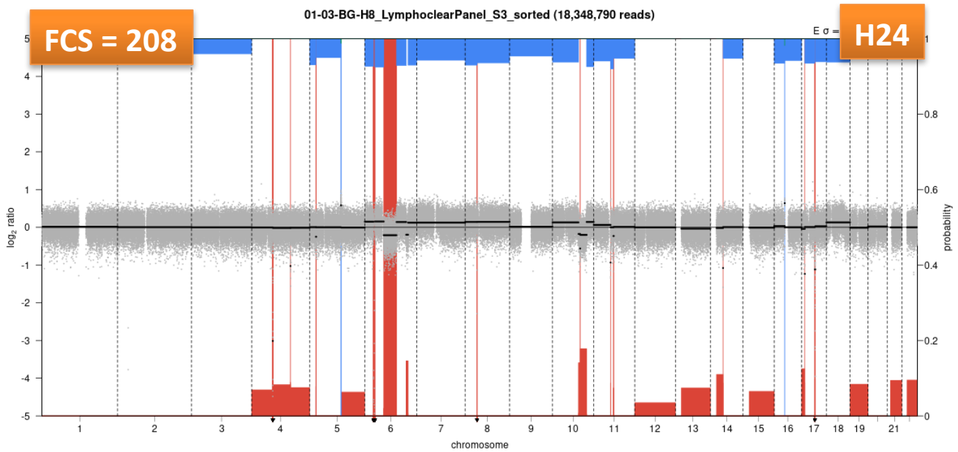


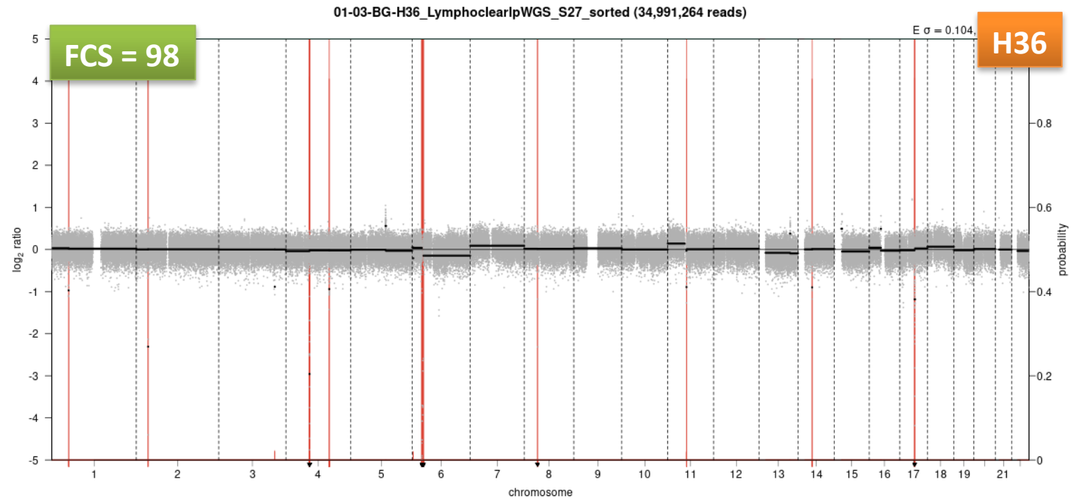


Negative ctDNA / decrease of FCS score

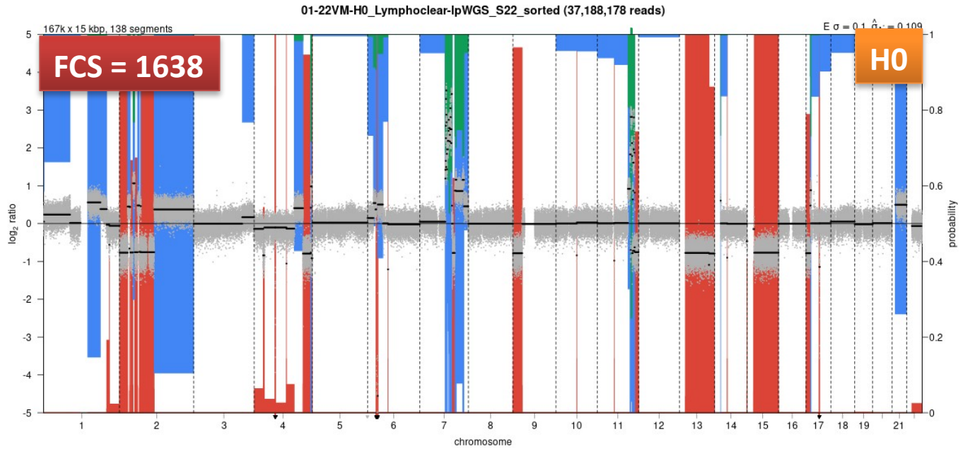


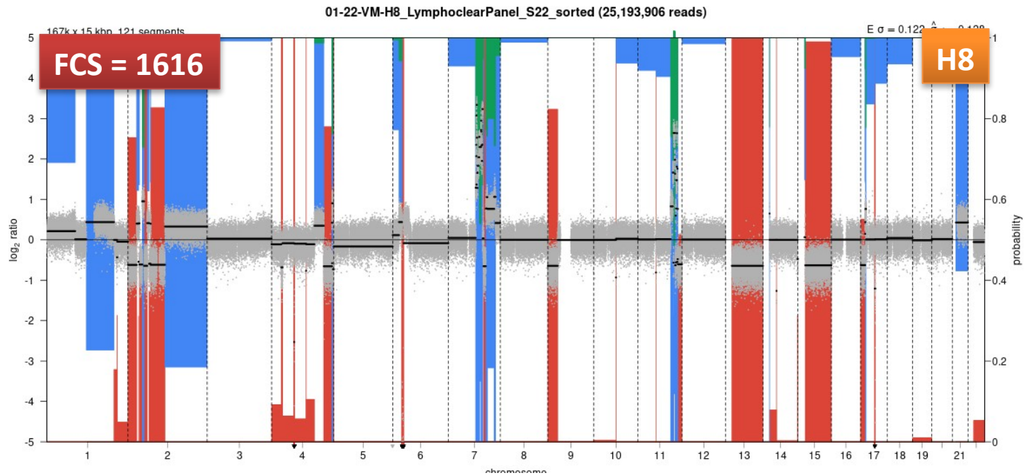


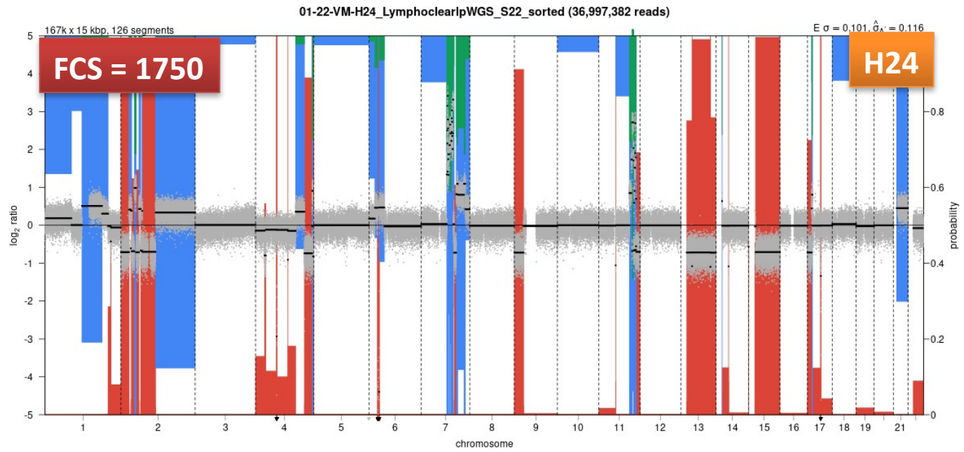


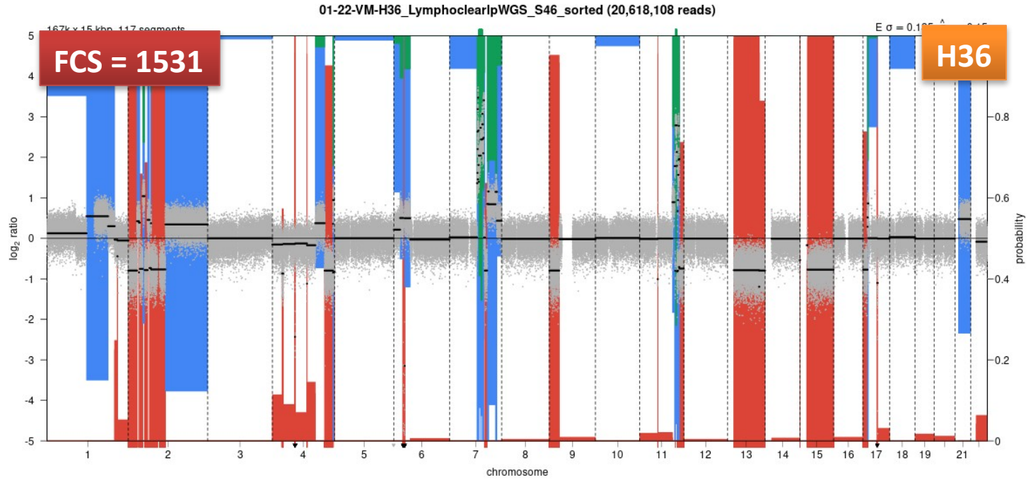


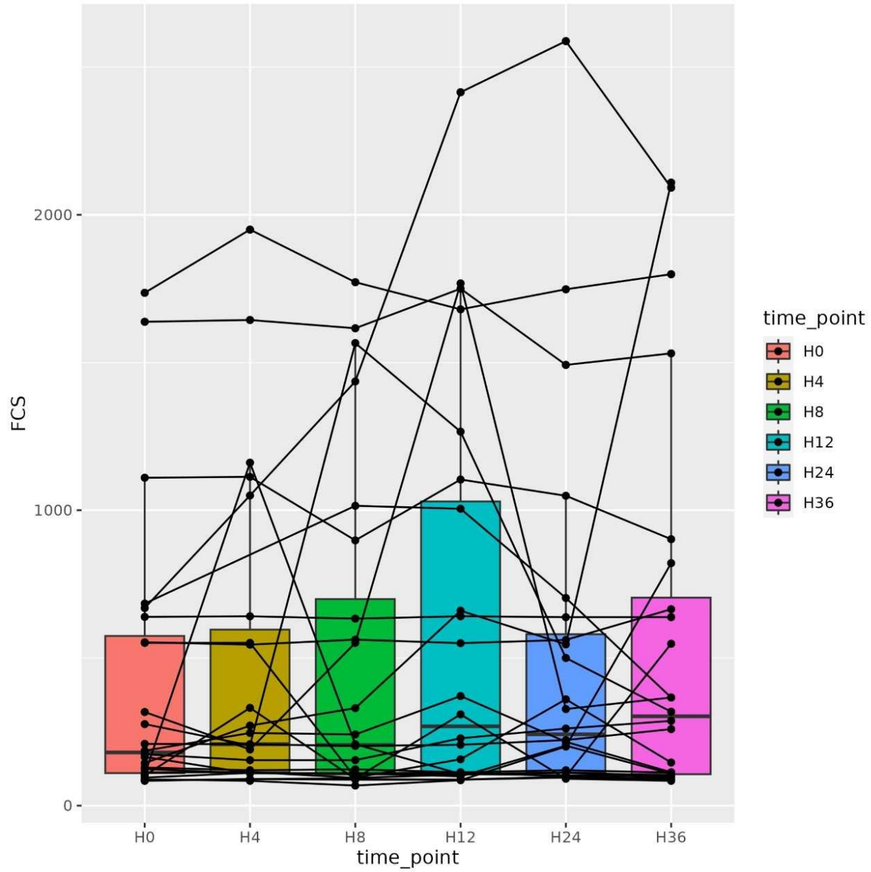
Positive ctDNA / No decrease of FCS score











Several groups

(1) Early undetectable residual sWGS segments
↓ FCS score / negative profiles

(2) Stable residual disease
≈ FCS score / positive profiles

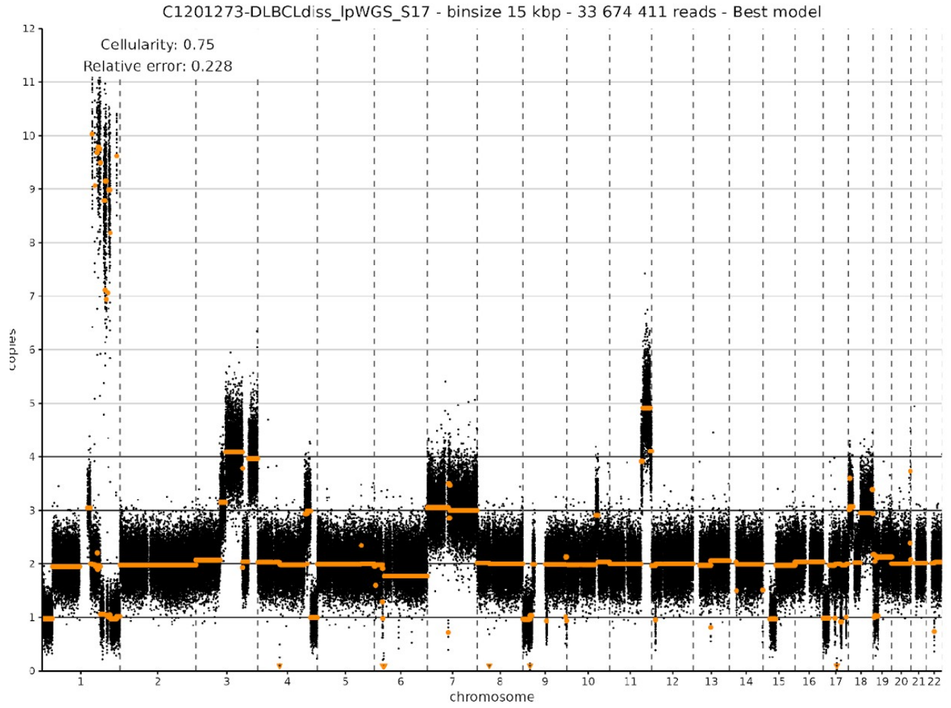
(3) Increasing residual disease
↑ FCS score / positive profiles

Results of next time points are still under analysis

Conclusion

sWGS is informative in the context of lymphomas

- ctDNA fragments are released quantitatively compared to the number of copies of each segment of the original tumors
- Liquid biopsies can be used to estimate with accuracy the CNV from the tumor of origin. LOD ?
- Some additional features in plasmas or tumors in a cohort of disseminated DLBCL : clonal evolution of the distinct sites ?
- Kinetics of ctDNA release during treatment : an early marker of response to chemotherapy? Correlation with PET-scan data ?



INSERM UMR 1245

Elodie BOHERS, Mael LOUIS, Mathieu VIENNOT, Philippe RUMINY, Vinciane RAINVILLE, Marie-Delphine LANIC, Mélody CAILLOT

Fabrice JARDIN, Hervé TILLY, Vincent CAMUS

Thank you for your attention