



# « Le NGS pour les nuls »

## Les neutropénies congénitales à l'ère du NGS

**Christine Bellanné-Chantelot**

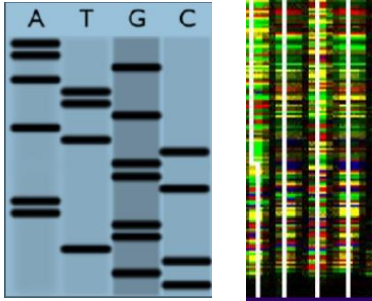
*Département de Génétique, Hôpital Pitié-Salpêtrière*

*UMR1170, Gustave Roussy, Villejuif*



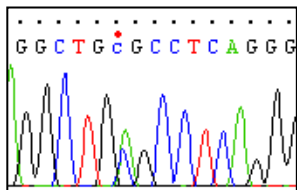
*Juin 2019*

# Du Sanger au **NGS** : une révolution technologique

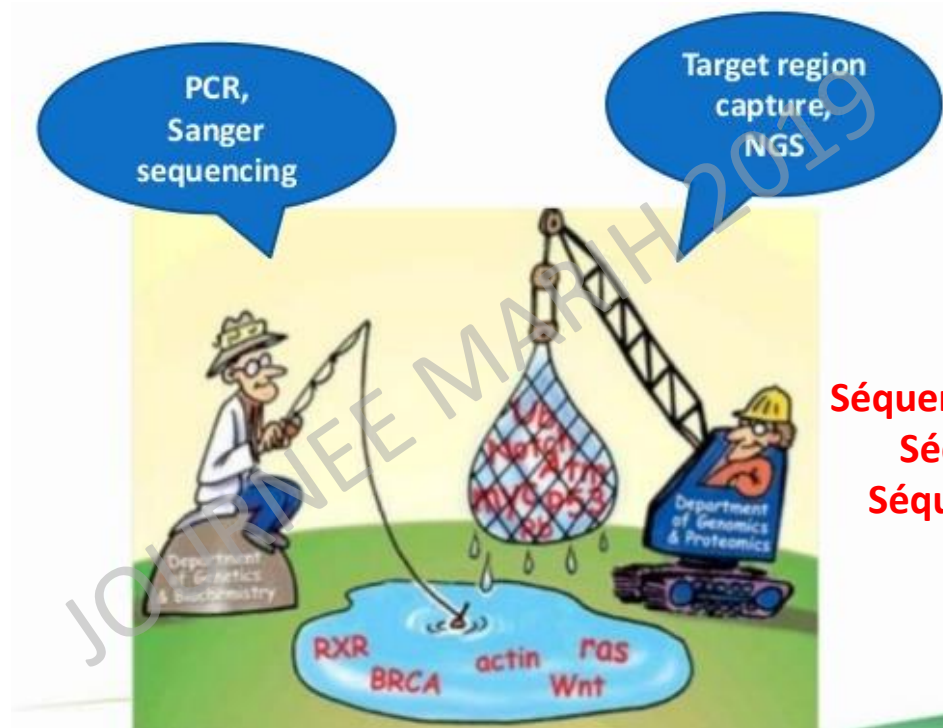


1977

1998



2004

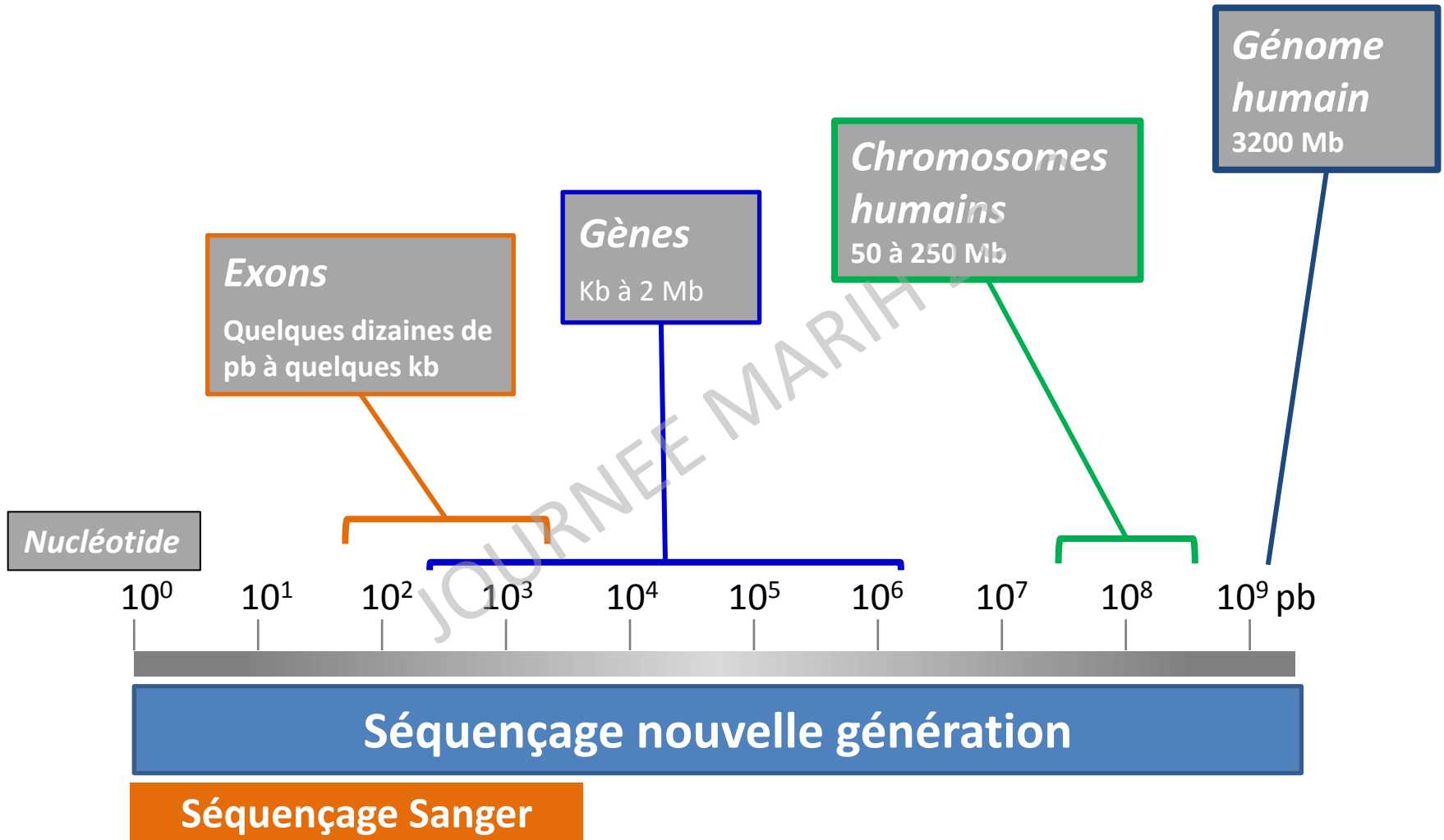


2005

**NGS**

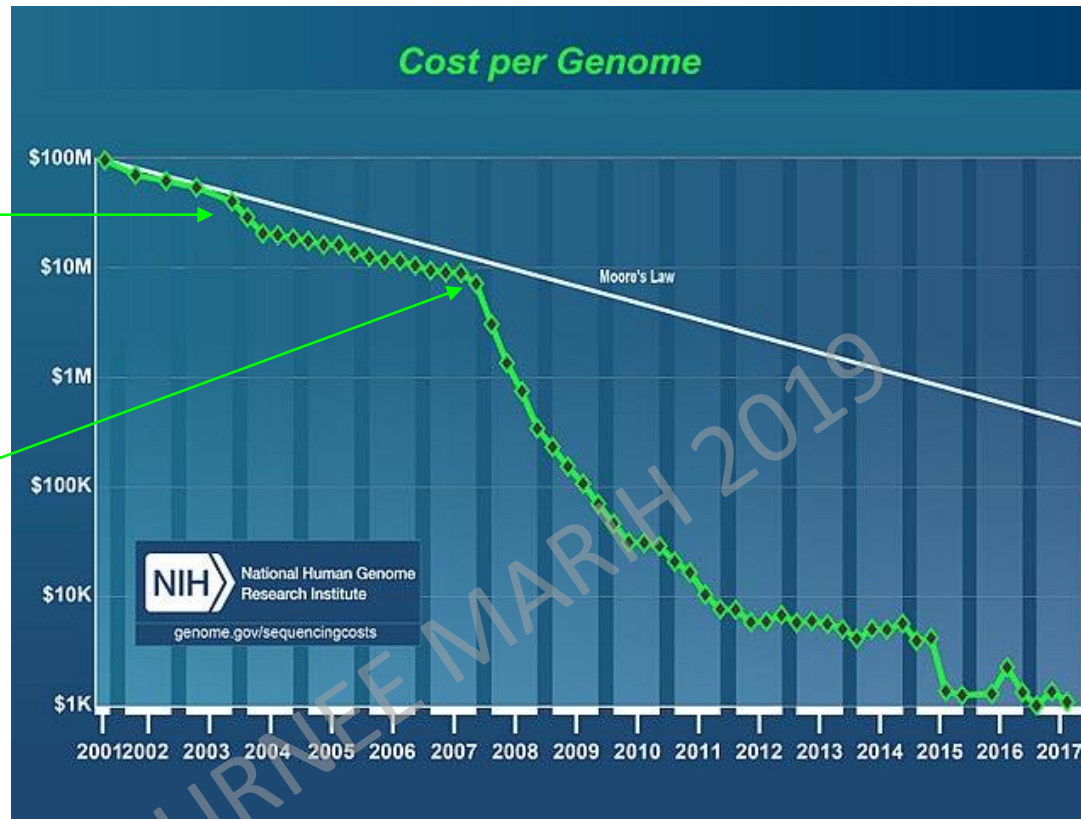
**Séquençage nouvelle génération**  
**Séquençage à haut débit**  
**Séquençage massif parallèle**

# NGS : un changement d'échelle





NGS



**2003**  
Séquençage du génome  
humain  
**Sanger**  
3 milliards \$, 13 ans

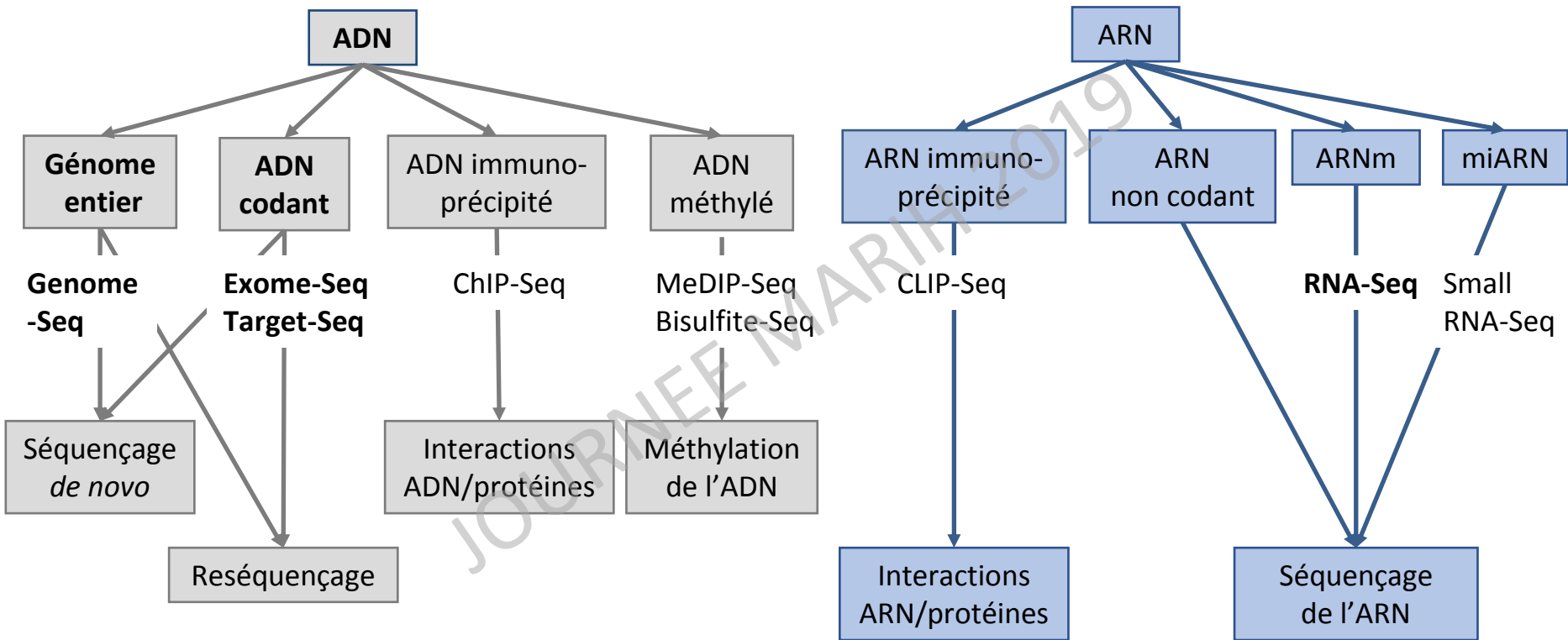


**2008**  
Séquençage d'un génome  
humain  
**NGS**  
1 million \$, 2 mois



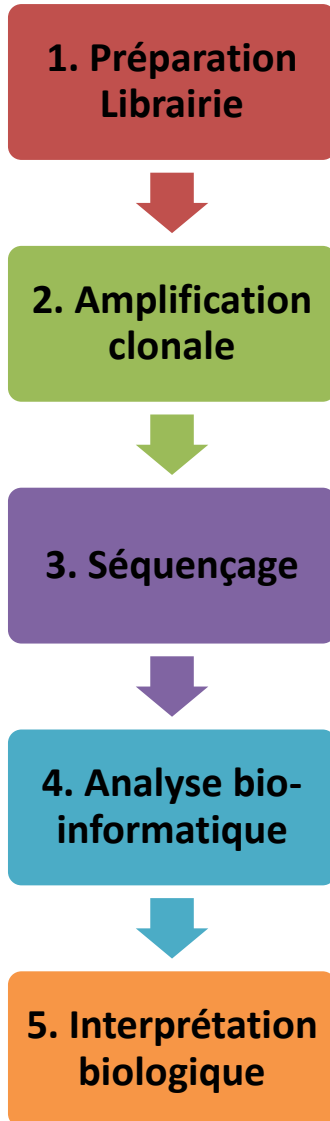
**2017**  
Séquençage d'un génome  
humain  
**NGS**  
1000 \$, 2 jours

# Les applications du NGS



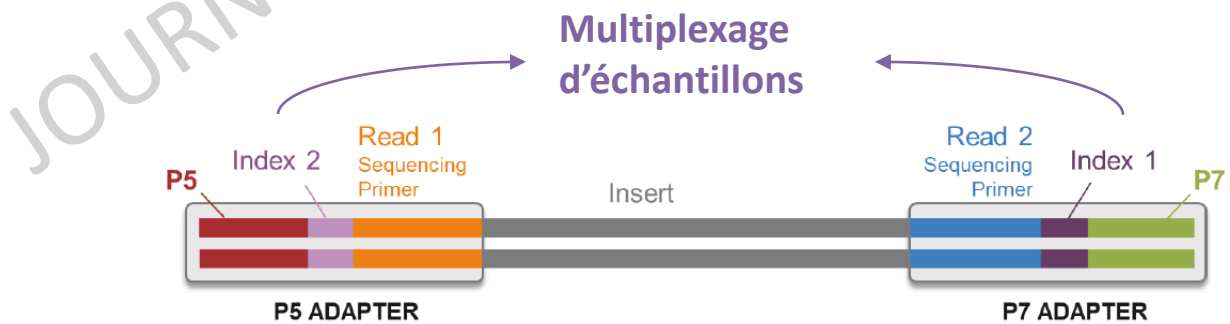
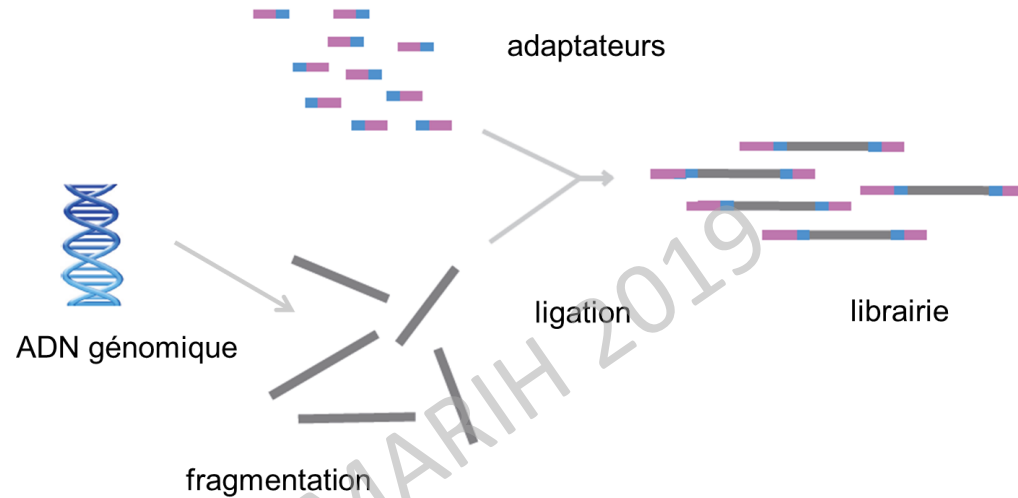
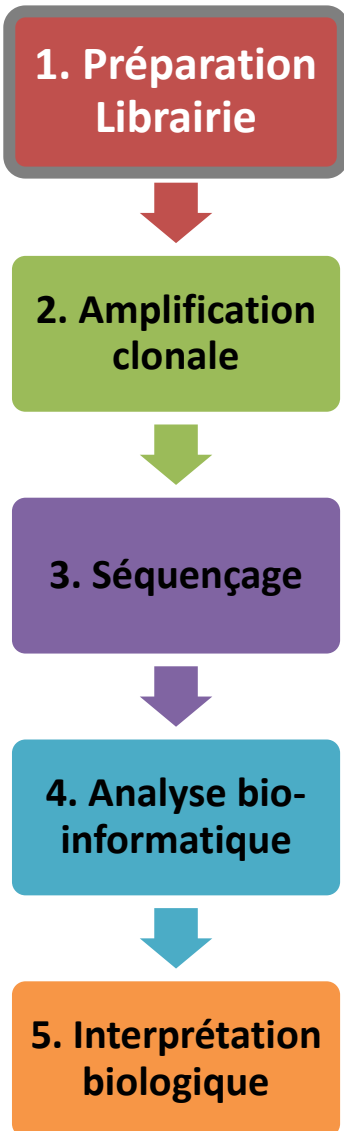
# Les 5 étapes du NGS

---

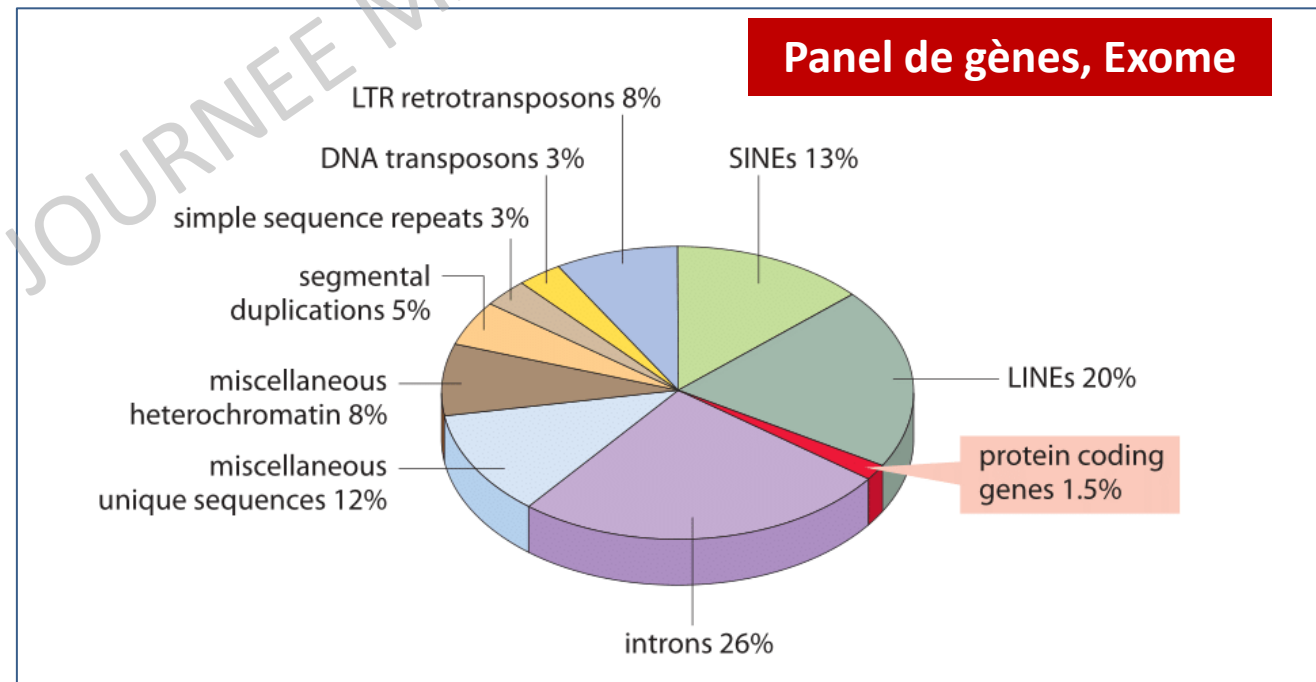
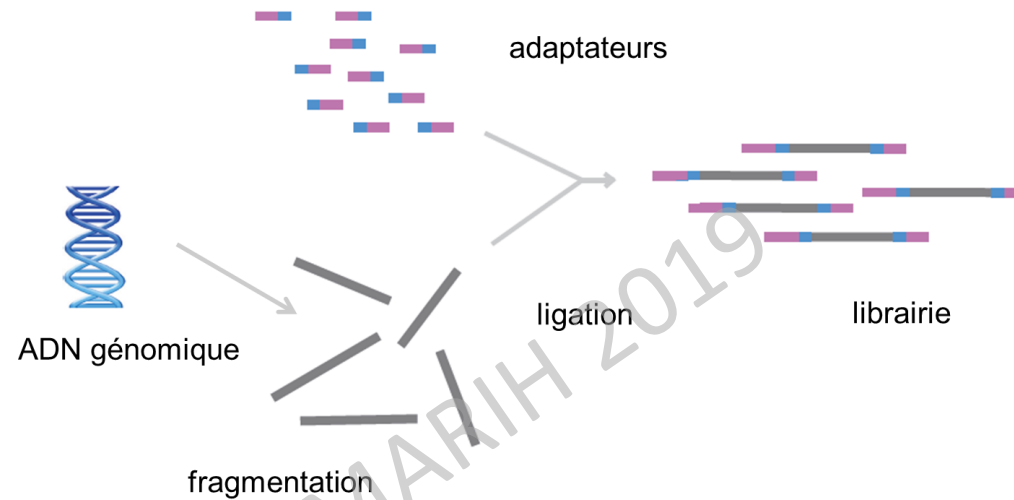
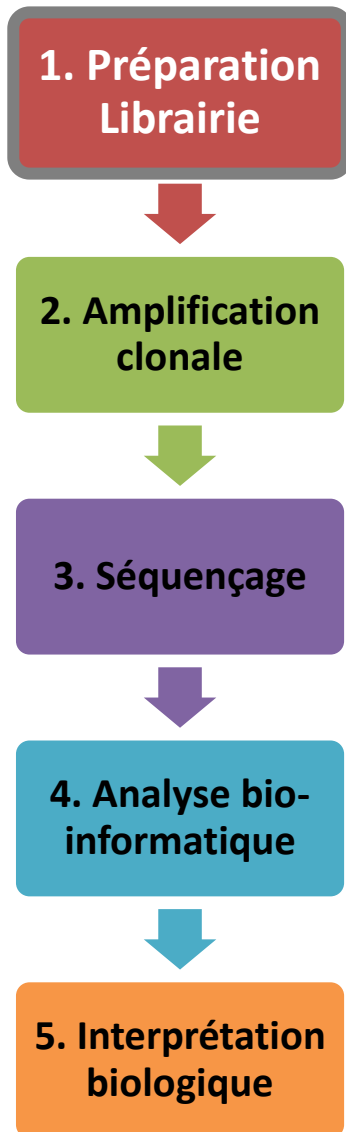


JOURNEE MARIH 2019

# La librairie

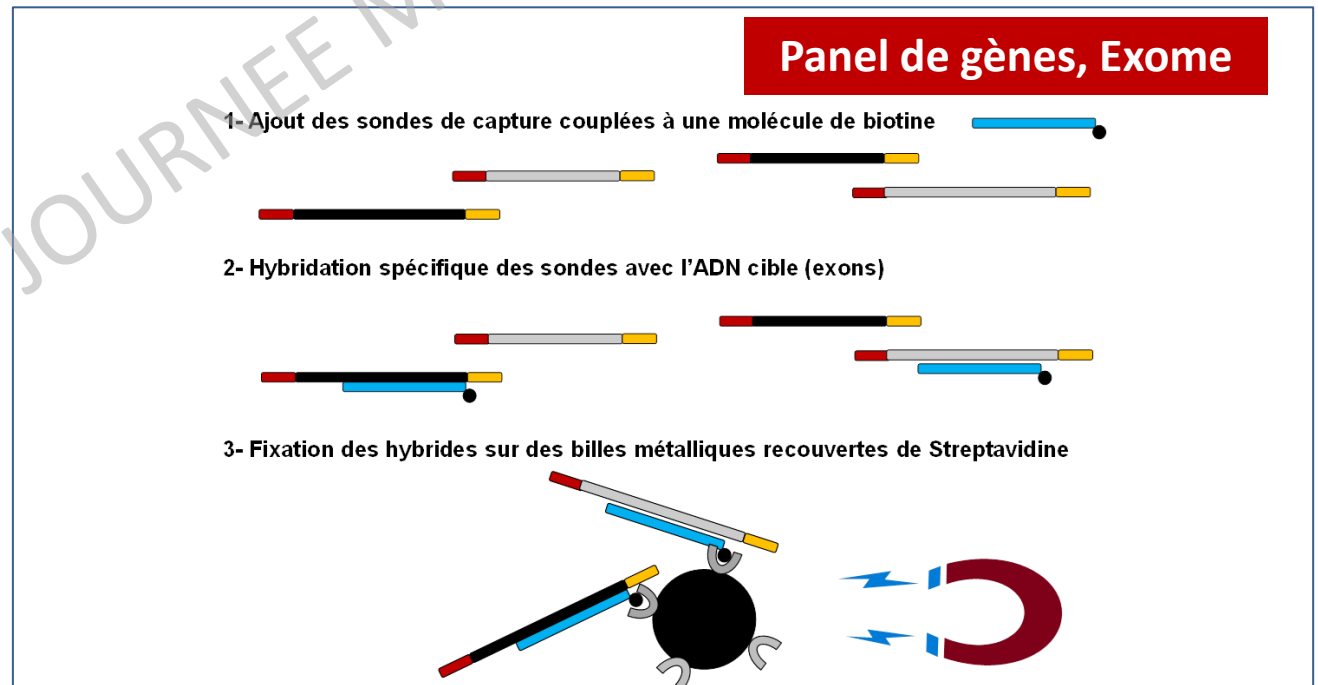
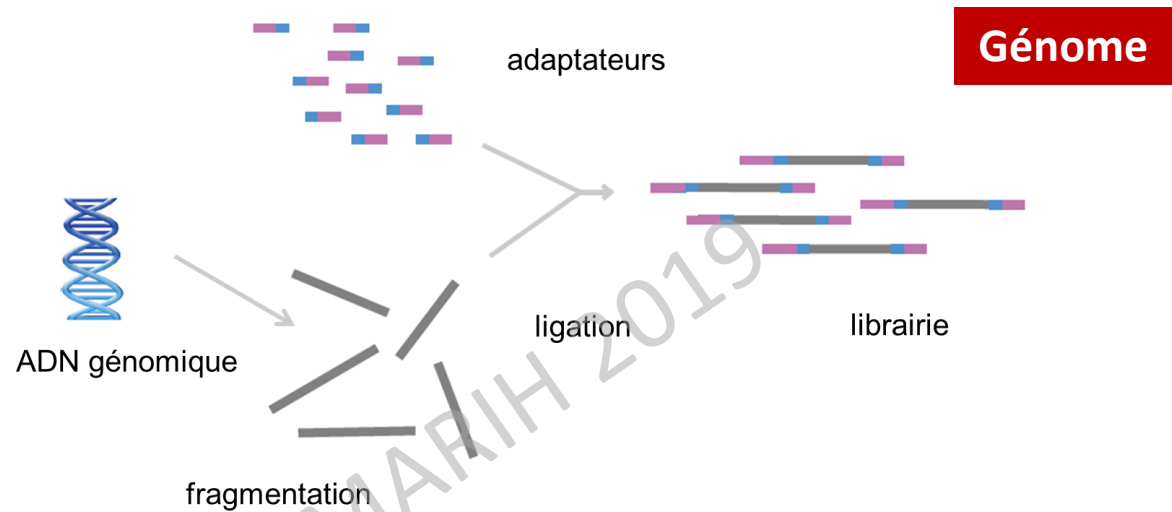
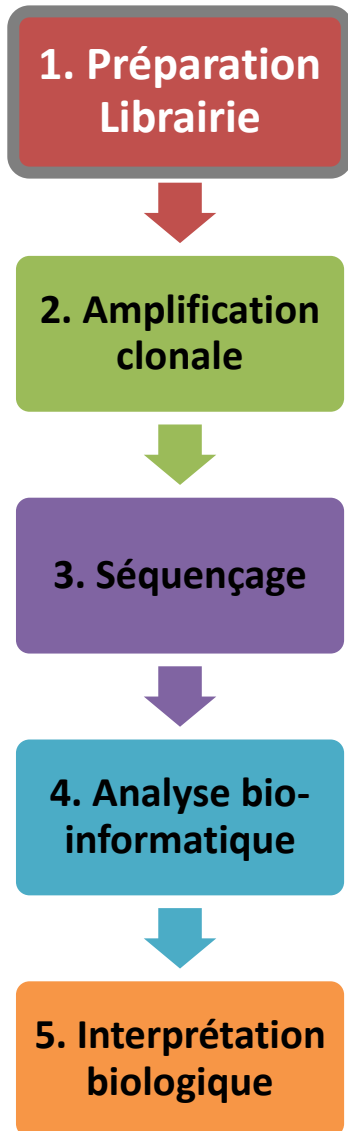


# L'enrichissement en régions d'intérêts

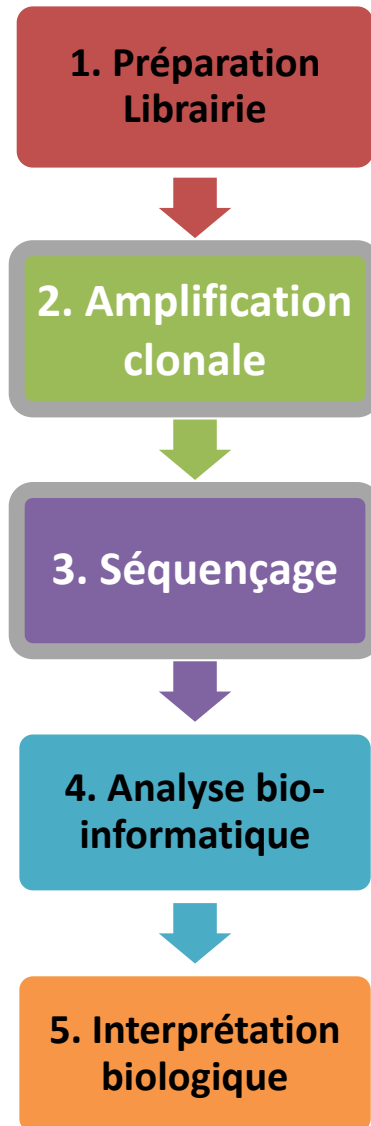




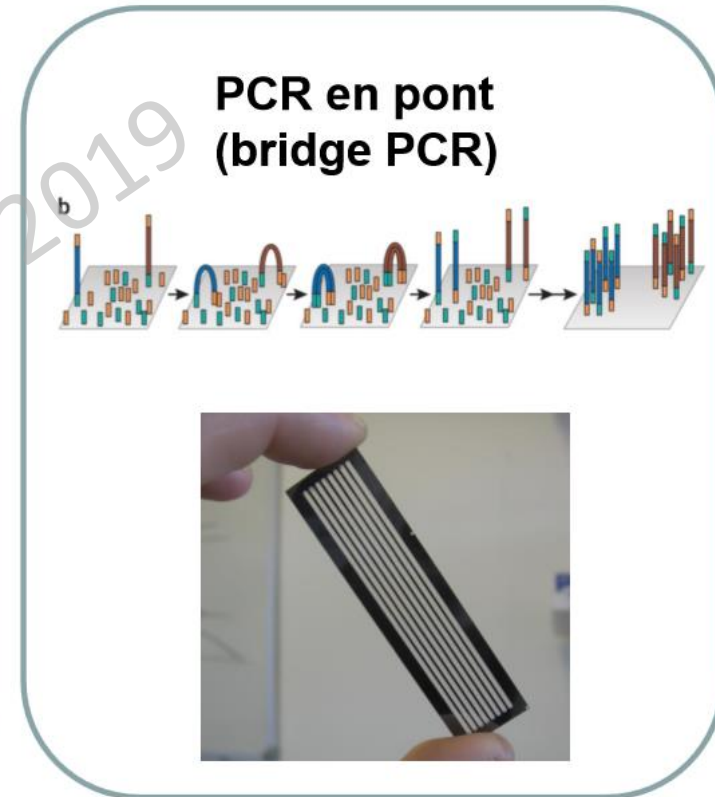
# L'enrichissement par capture



# Le séquençage

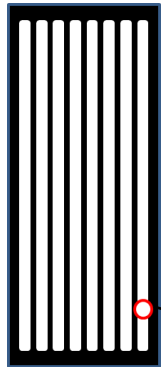


**Ion torrent  
ThermoFisher**

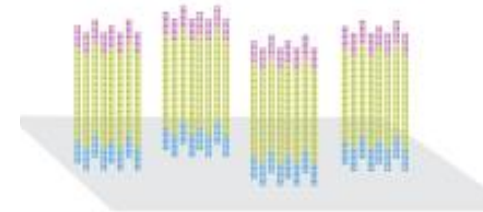
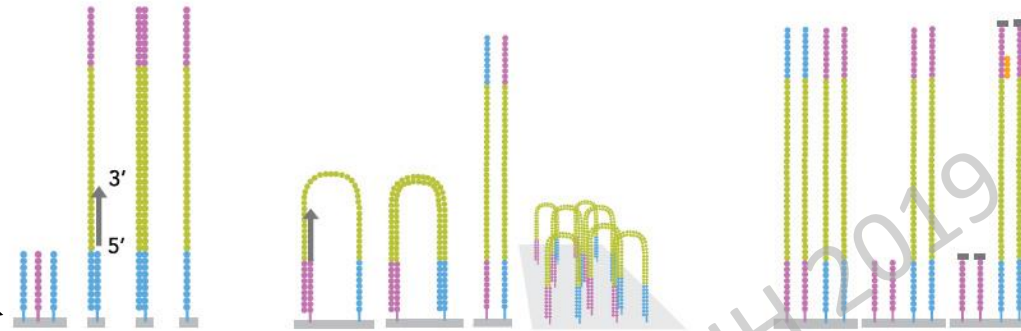


**Solexa  
Illumina**

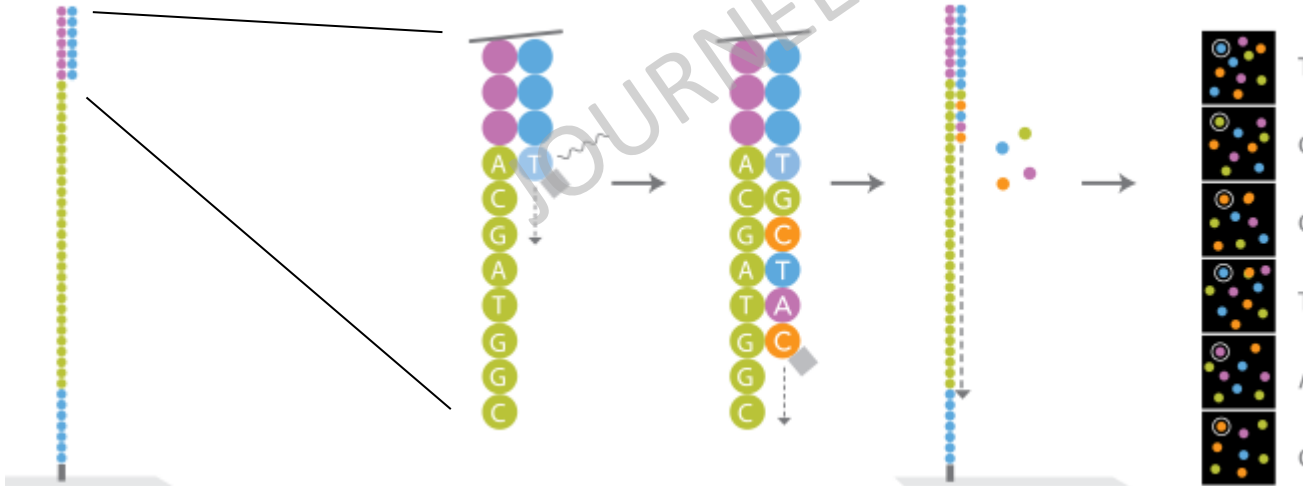
# Chimie Illumina (Solexa)



Flow cell



Formation des «clusters» pour séquençage





	<b>MiSeq (v3)</b>	<b>NextSeq 500</b>	<b>HiSeq 2500 (haut débit v4)</b>	<b>HiSeq X (2 flow cells)</b>
<b>Lectures</b>	150-600 pb	75-300 pb	50-250 pb	300 pb
<b>Millions de lectures / run</b>	22	400	2,000	6,000
<b>Débit</b>	3.3-13.2 Gpb	30-120 Gpb	100-500 Gpb	1,800 Gpb

# L'analyse bio-informatique des données

1. Préparation  
Librairie



2. Amplification  
clonale



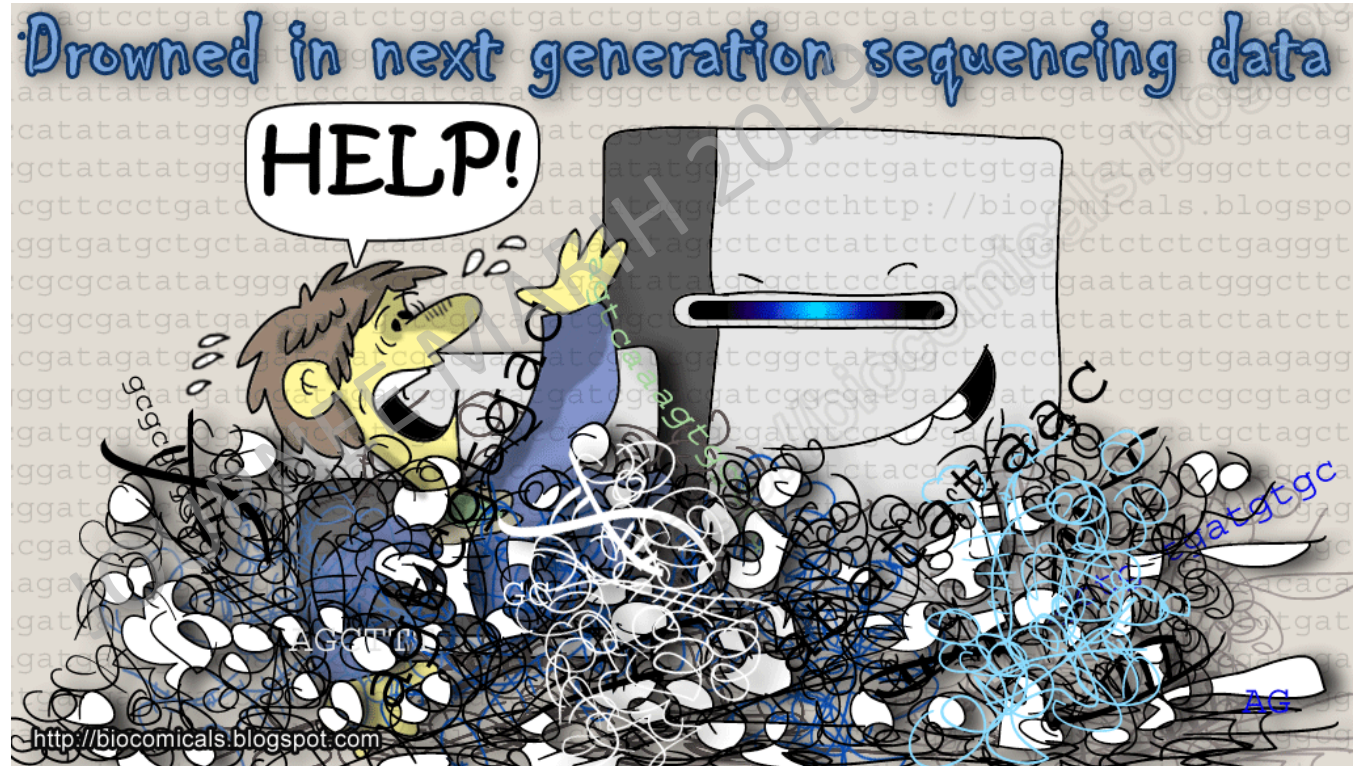
3. Séquençage



4. Analyse bio-  
informatique

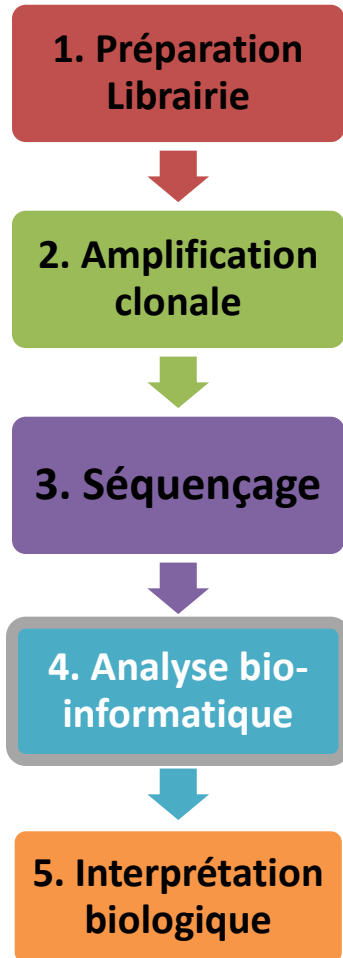


5. Interprétation  
biologique



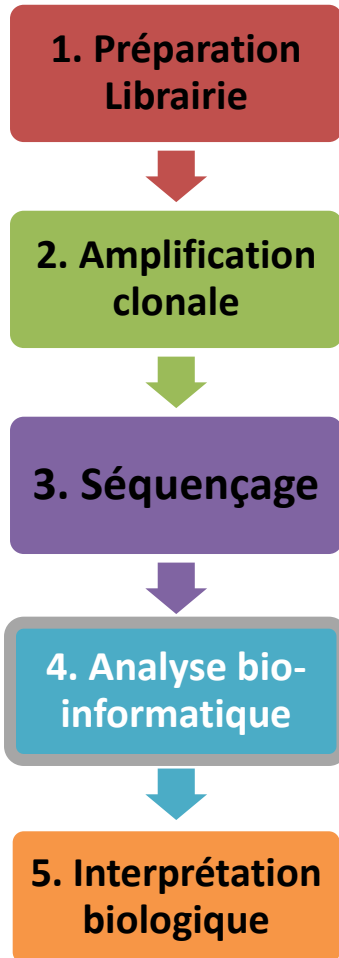
# L'analyse bio-informatique des données

---



1. **Détermination de la séquence** « Base calling » (Fastq)
2. **Alignement** de la séquence sur le génome de référence « Mapping » (BAM/SAM)
3. **Détection des variants** « Variant calling » (VCF)
4. **Annotation** des variants (VCF)
5. **Filtrage des variants**

# L'analyse bio-informatique des données

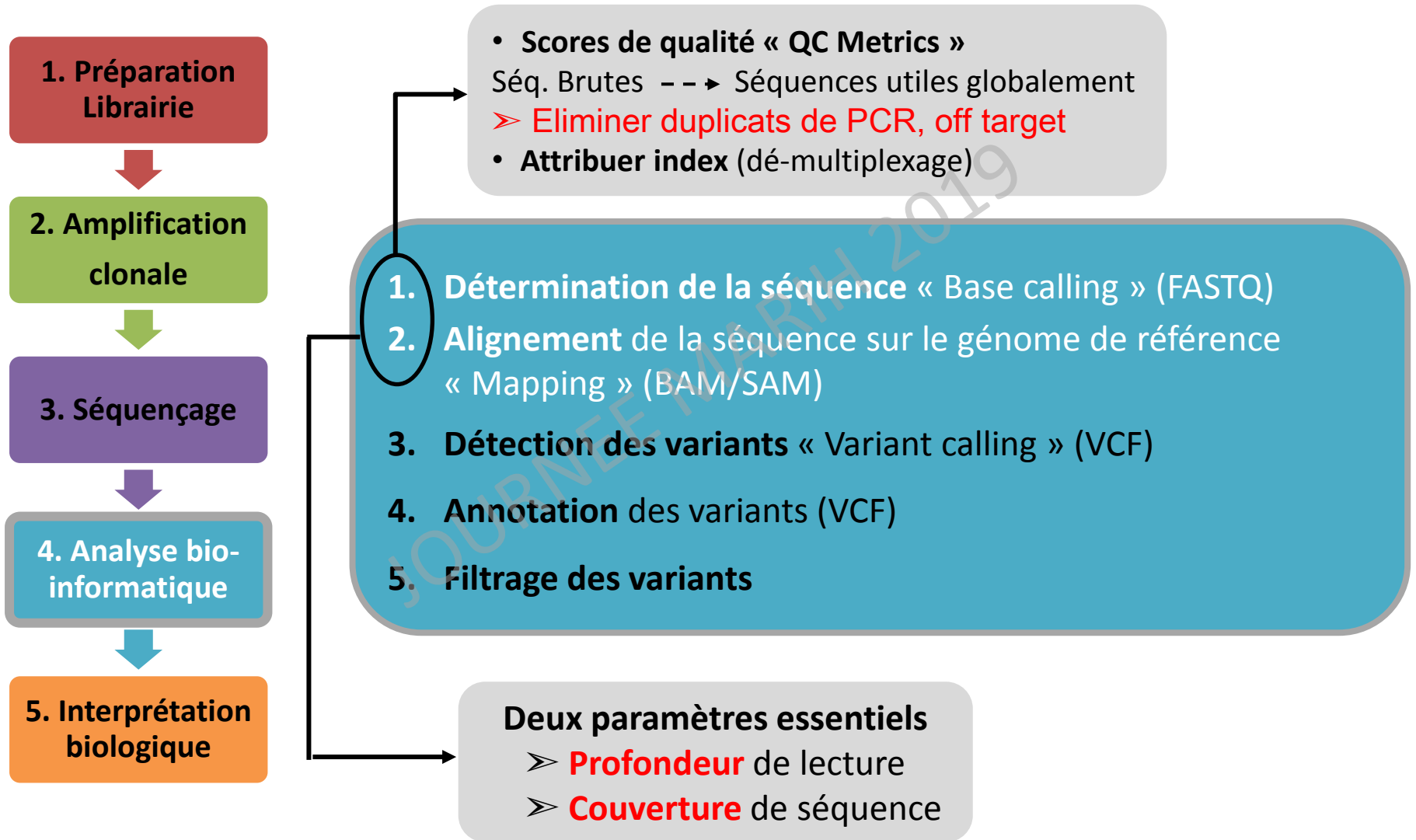


- Scores de qualité « QC Metrics »  
Séq. Brutes --> Séquences utiles globalement  
➤ **Eliminer duplicats de PCR, off target**
- Attribuer index (dé-multiplexage)

1. Détermination de la séquence « Base calling » (Fastq)
2. Alignement de la séquence sur le génome de référence « Mapping » (BAM/SAM)
3. Détection des variants « Variant calling » (VCF)
4. Annotation des variants (VCF)
5. Filtrage des variants

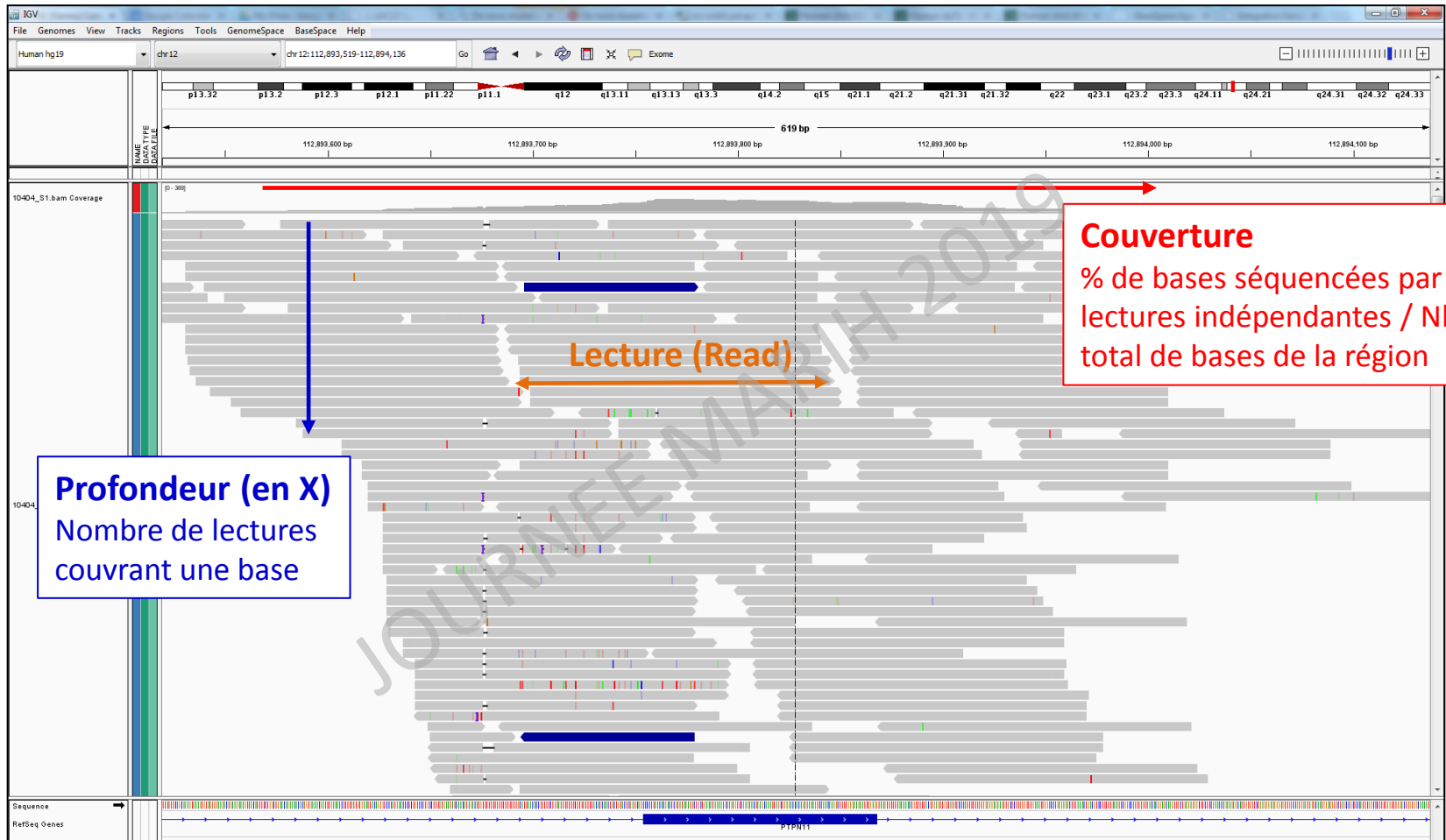


# L'analyse bio-informatique des données





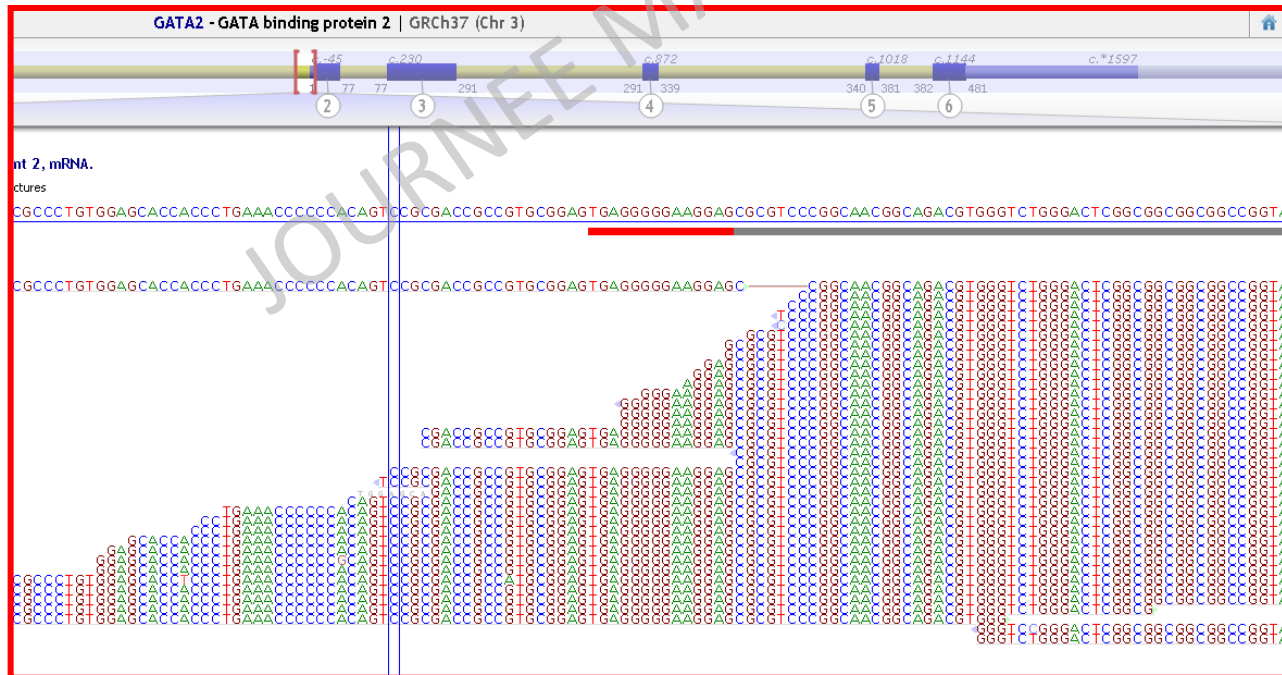
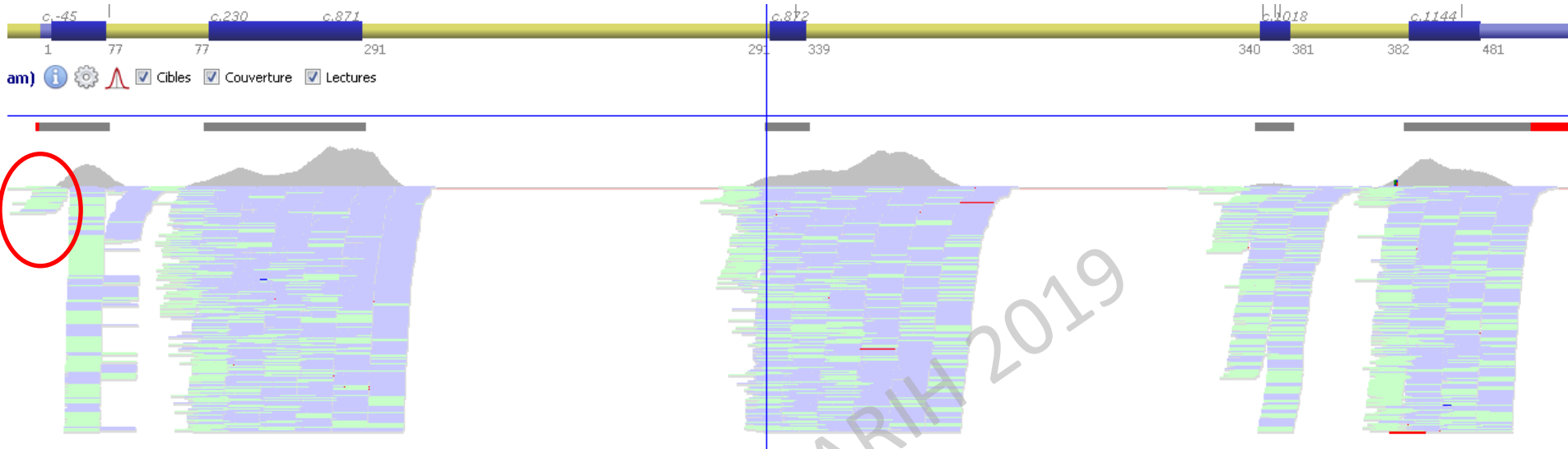
# Deux impératifs de qualité de séquence



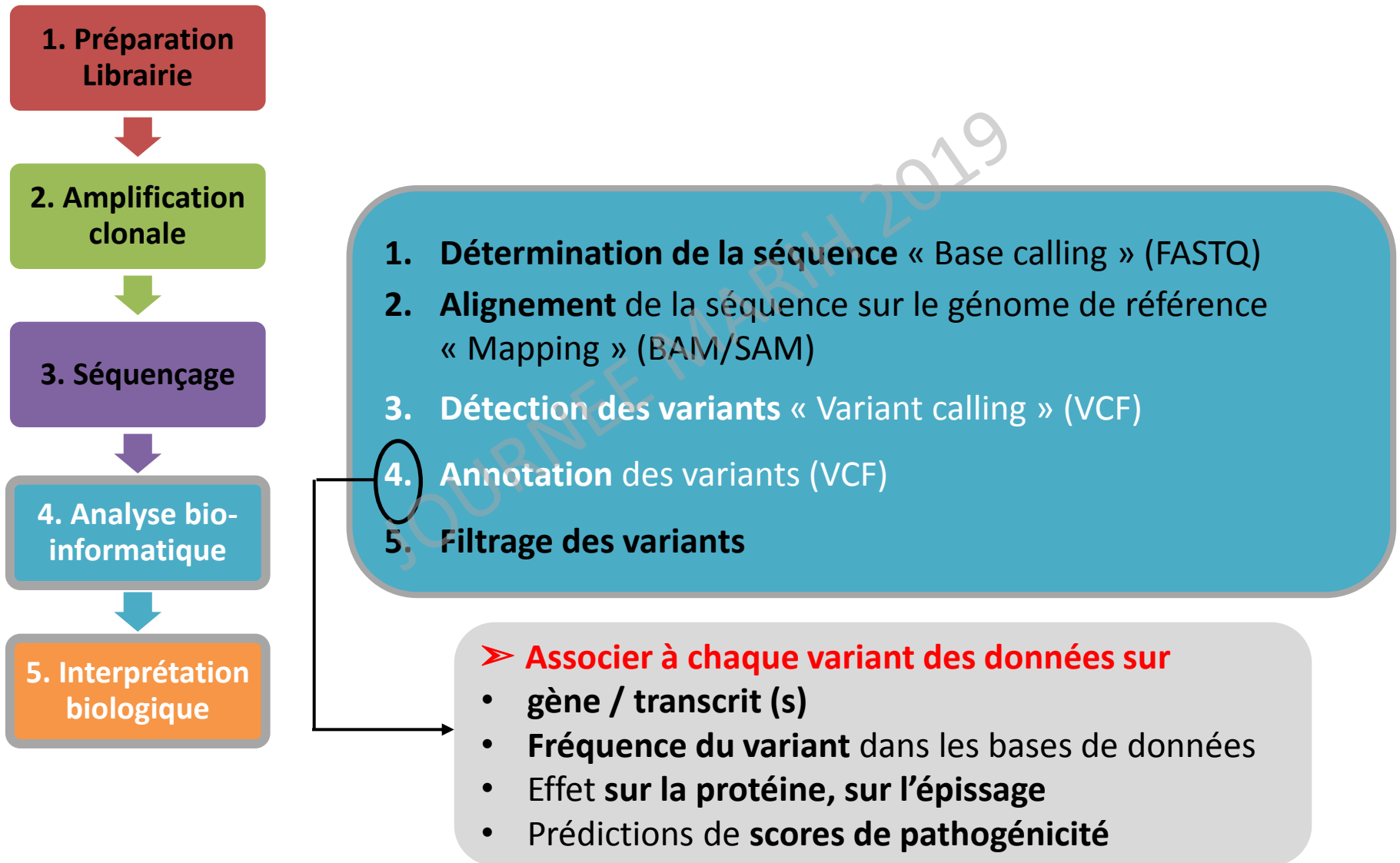
## Exigences :

- Variant germlinal : 100 % couverture avec une profondeur > 20X
- Variant somatique : profondeur > 1000X pour détecter un variant à 1%

# GATA2



# L'analyse des données



# L'analyse des données

---

1. Préparation  
Librairie



2. Amplification  
clonale



3. Séquençage



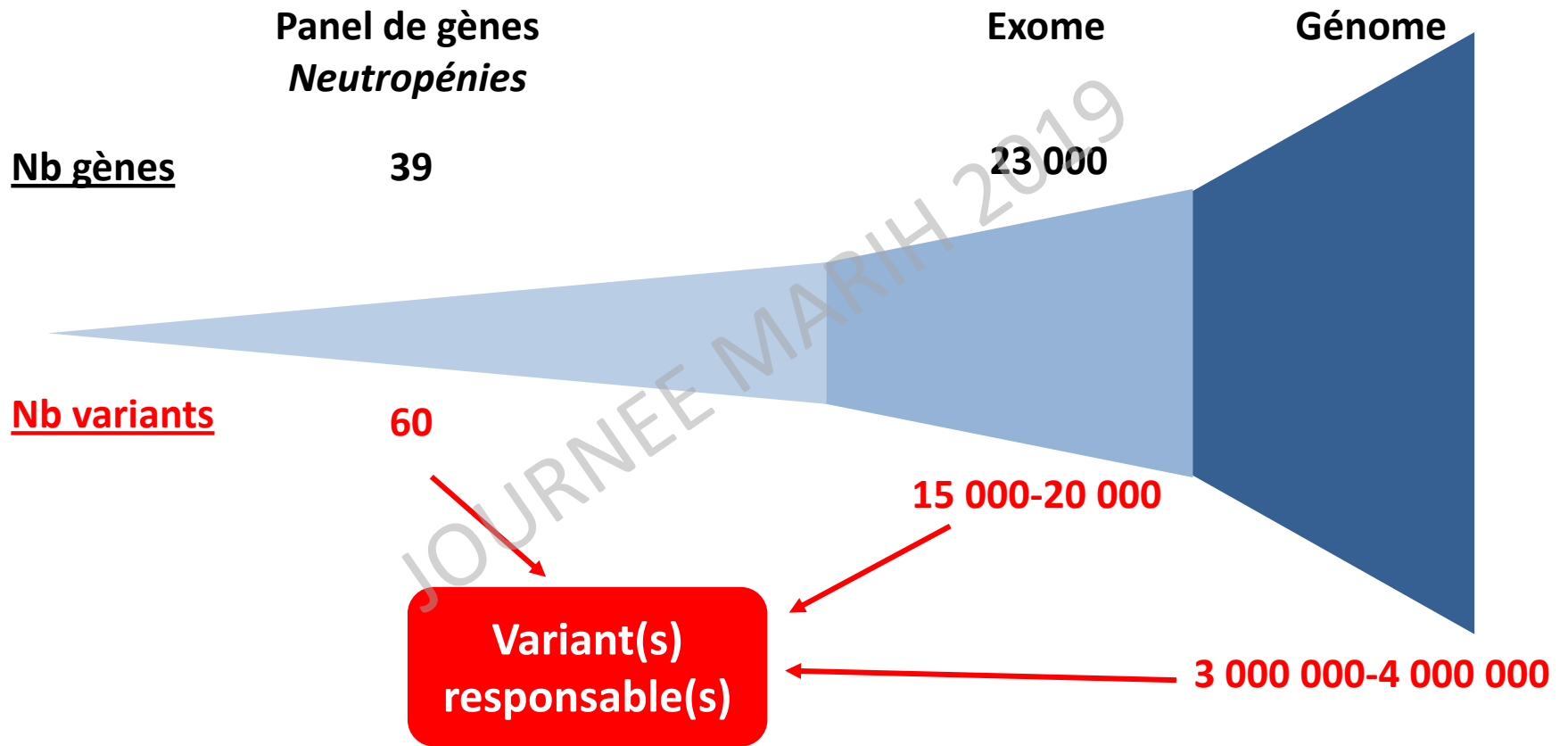
4. Analyse bio-  
informatique



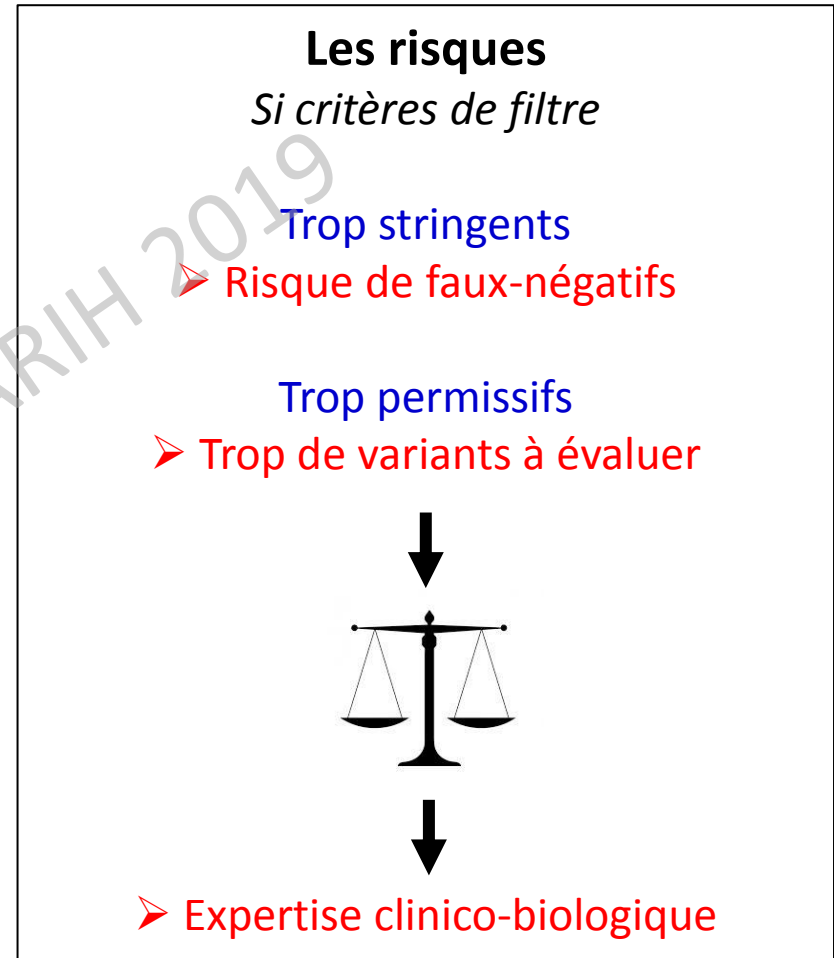
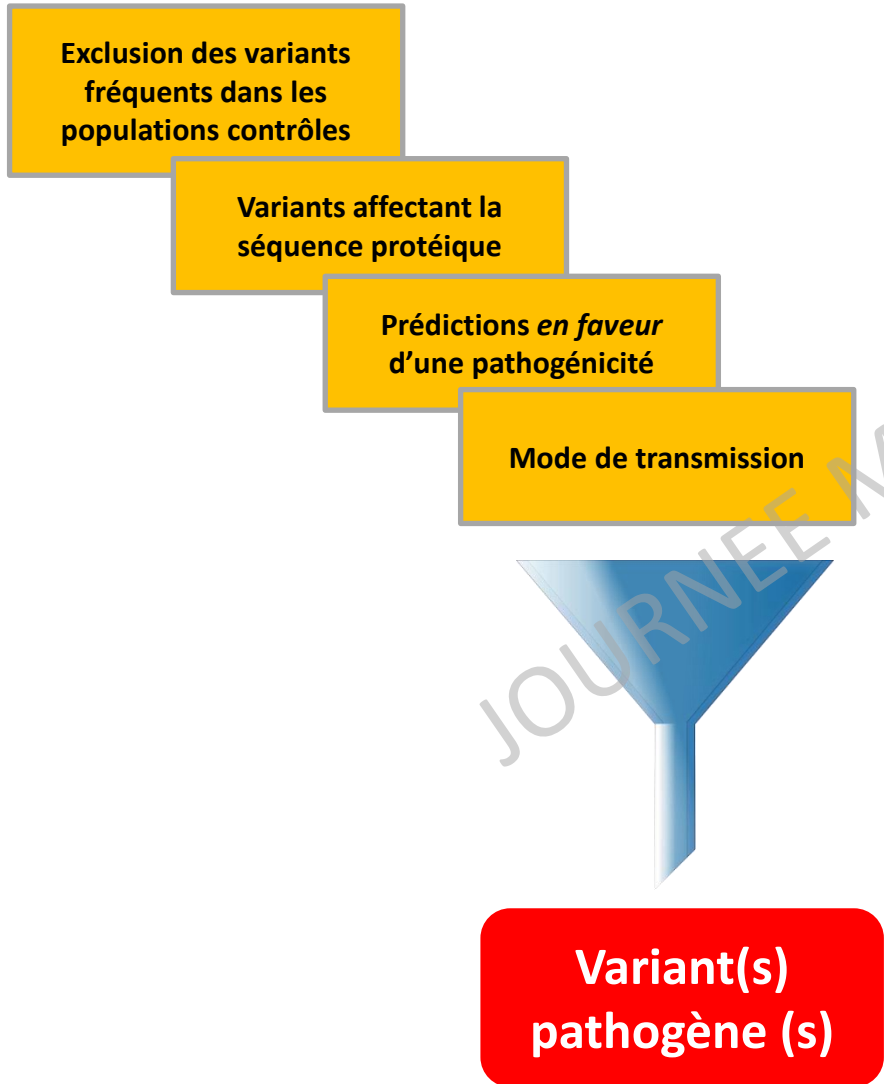
5. Interprétation  
biologique

1. **Détermination de la séquence** « Base calling » (FASTQ)
2. **Alignement** de la séquence sur le génome de référence « Mapping » (BAM/SAM)
3. **Détection des variants** « Variant calling » (VCF)
4. **Annotation** des variants (VCF)
5. **Filtrage / « priorisation »** des variants

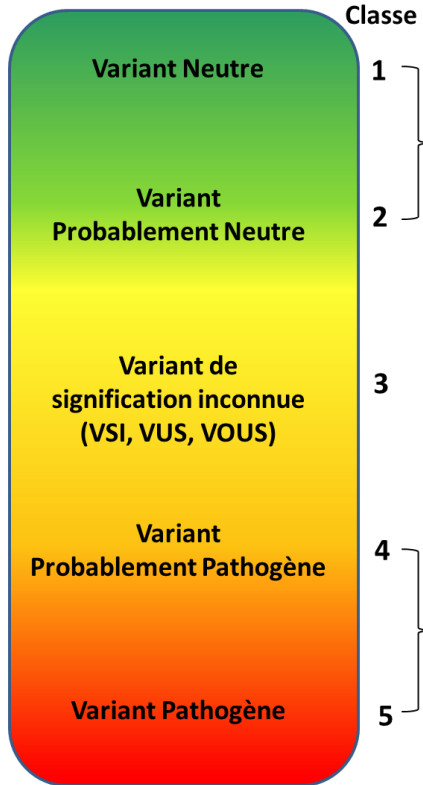
# Défi : nombre et interprétation des variants



# Défi : nombre et interprétation des variants



# Aide à l'interprétation des variants

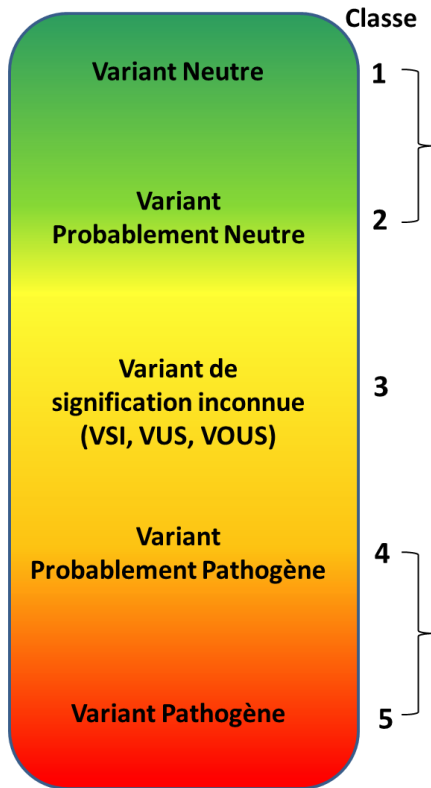


- Données épidémiologiques**
  - Bases de données de populations
- Effet du variant sur la fonction de la protéine**
  - Nature du variant nucléotidique
  - Prédiction *in silico* de pathogénicité
  - Données fonctionnelles
- Ségrégation familiale**
- Phénotypes**

**Base gnomAD**  
120 000 exome+génomés

g.856145C>T <Exomes+Genomes>		
Filter: <b>PASS</b>		
Max AF in populations: <b>1.63%</b>		
	Allele Freq	Hom Fre
<b>All</b>	<b>0.15%</b>	<b>0.0015%</b>
African	1.63%	0.017%
Latino	0.049%	
Ashkenazi Jewish		
East Asian		
South Asian	0.0033%	
Eur. (Non-Finnish)	0.0032%	
Eur. (Finnish)		
Other	0.031%	

# Aide à l'interprétation des variants



Classe

1

2

3

4

5

## ☐ Données épidémiologiques

- Bases de données de populations

## ☐ Effet du variant sur la fonction de la protéine

- Nature du variant nucléotidique
- Prédiction *in silico* de pathogénicité
- Données fonctionnelles

## ☐ Ségrégation familiale

## ☐ Phénotypes

### Classe 5 : Variant pathogène

1 **PVS** ET  $\geq 1$  **PS** OU  
 $\geq 2$  **PM** OU  
 1 **PM** ET 1 **PP** OU  
 $\geq 2$  **PP** OU

$\geq 2$  **PS** OU  
 1 **PS** ET  $\geq 3$  **PM** OU  
 2 **PM** ET  $\geq 2$  **PP** OU  
 1 **PM** ET  $\geq 4$  **PP**

### Classe 4 : Variant probablement pathogène

1 **PVS** ET 1 **PM** OU  
 1 **PS** ET 1-2 **PM** OU  
 1 **PS** ET  $\geq 2$  **PP** OU  
 $\geq 3$  **PM** OU  
 2 **PM** ET  $\geq 2$  **PP** OU  
 1 **PM** ET  $\geq 4$  **PP**

**PVS/S/M/P** : Argument très fort/fort/moyen/faible en faveur de la pathogénicité du variant.¶

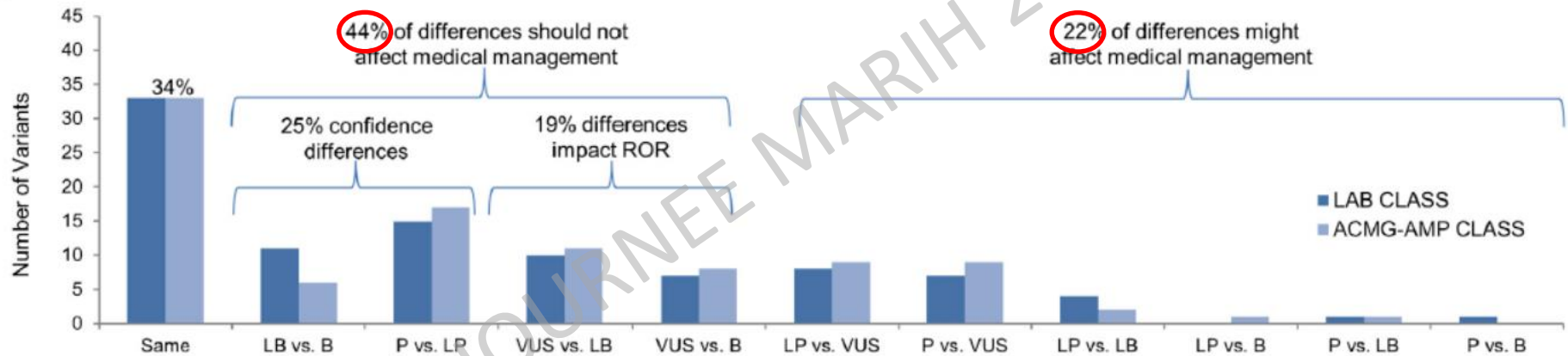


*Amendola et al, Am J Hum Genet , 2016*

- 9 laboratoires américains
- Interpréter 99 variants

## Comparaison de l'interprétation **inter-laboratoire**

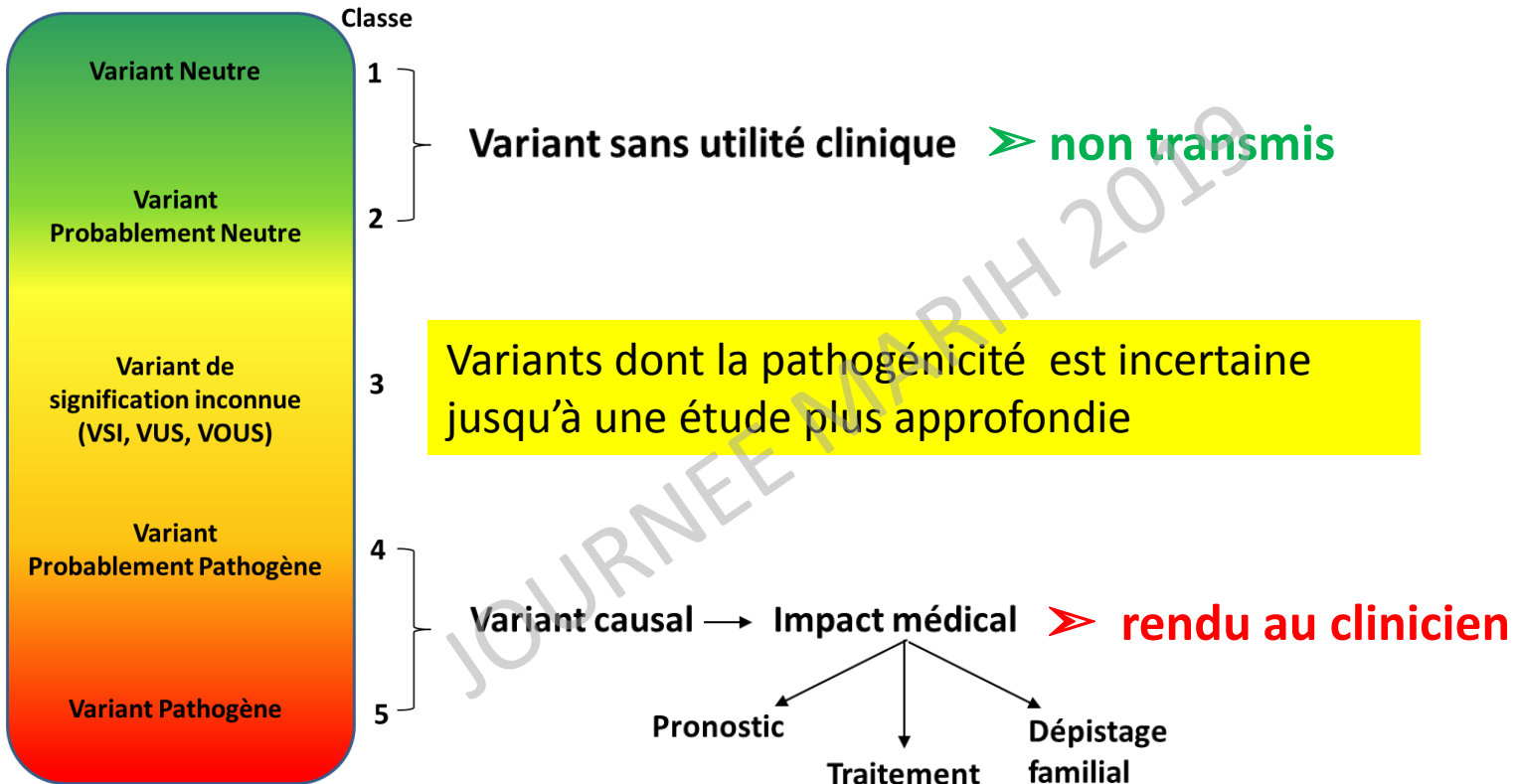
**B**



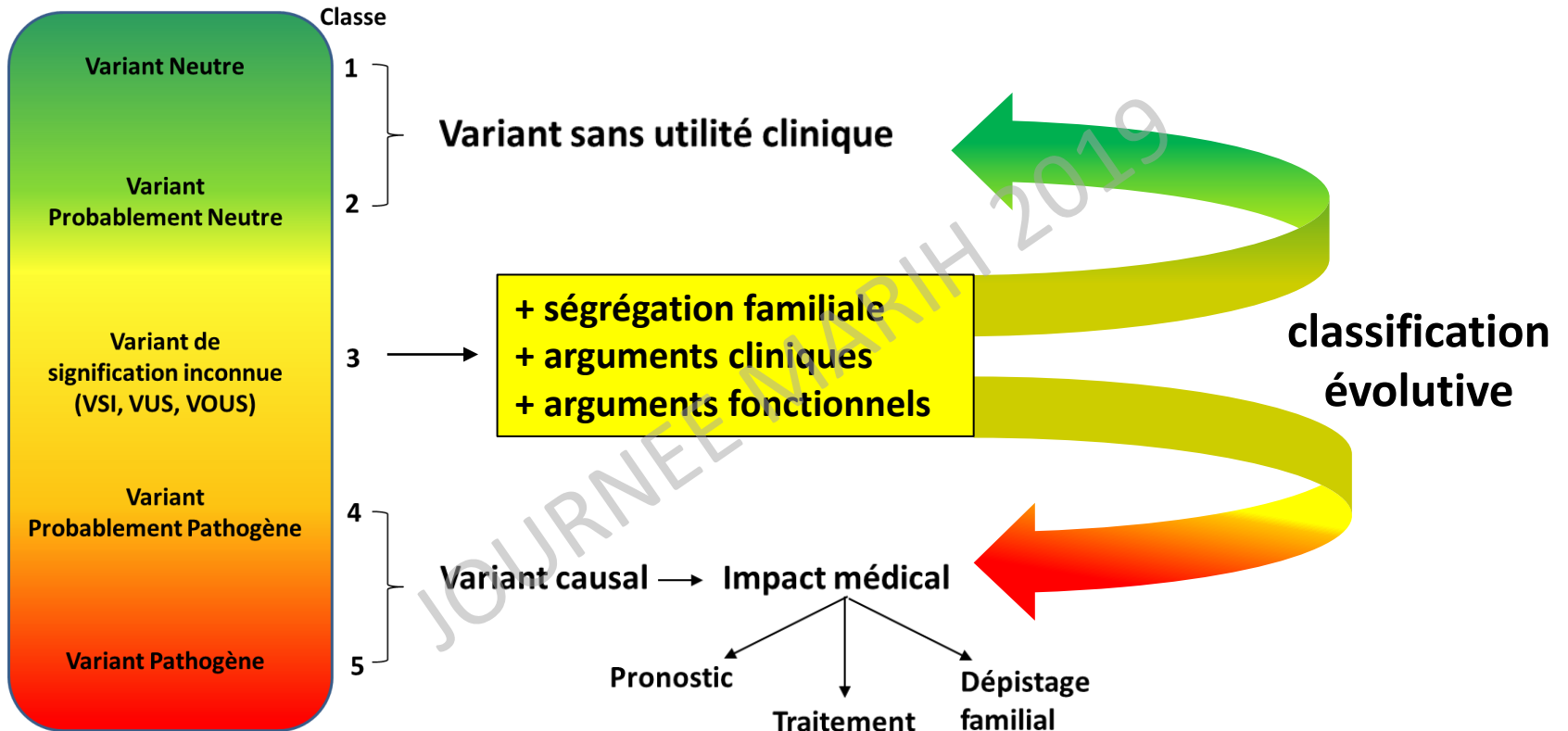
Concordance **34%** quand comparaison inter-laboratoire de l'interprétation

Après discussions, concordance **71%** ➤ réduction des discordances de 66% à 29%

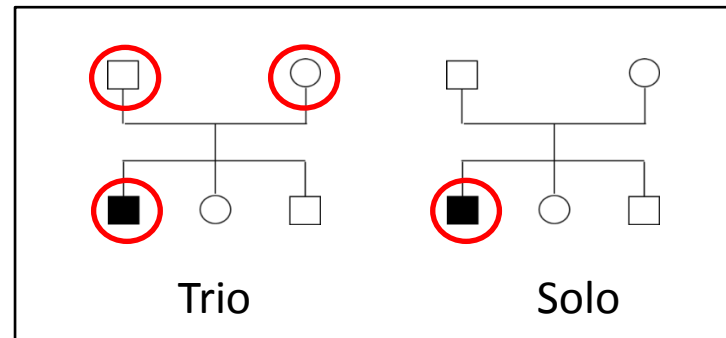
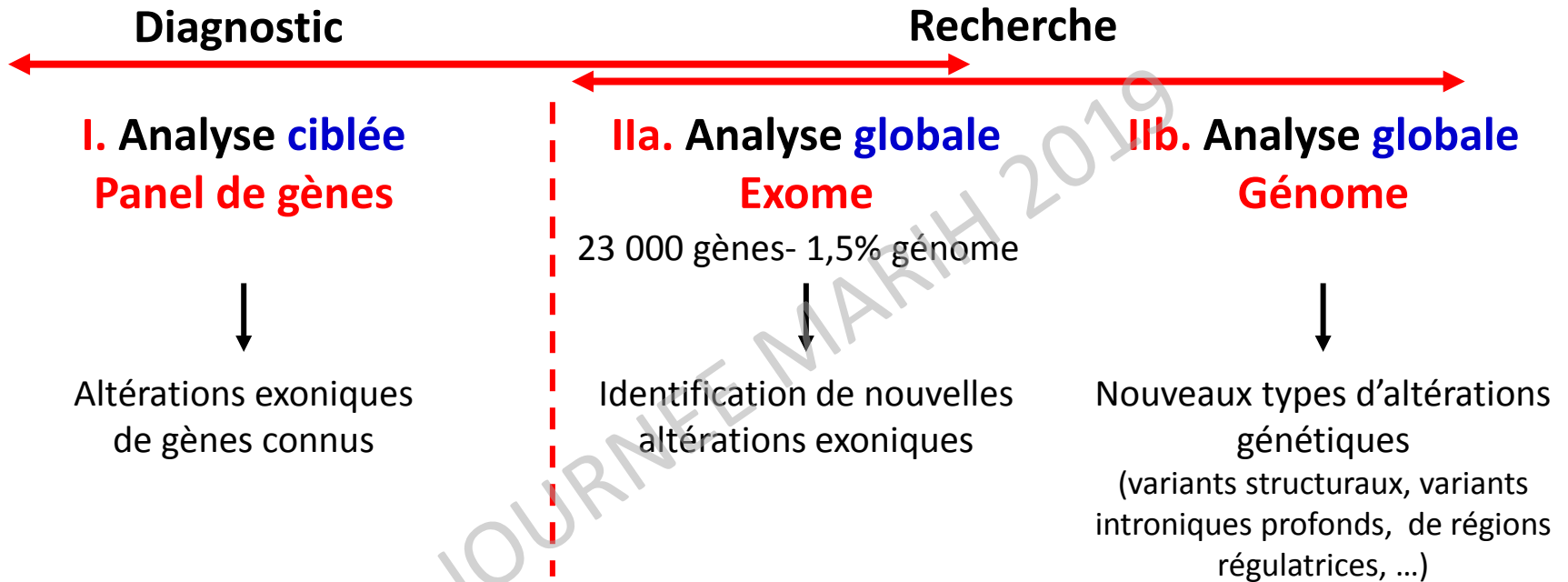
# Complexité du diagnostic : **variants de signification inconnue**



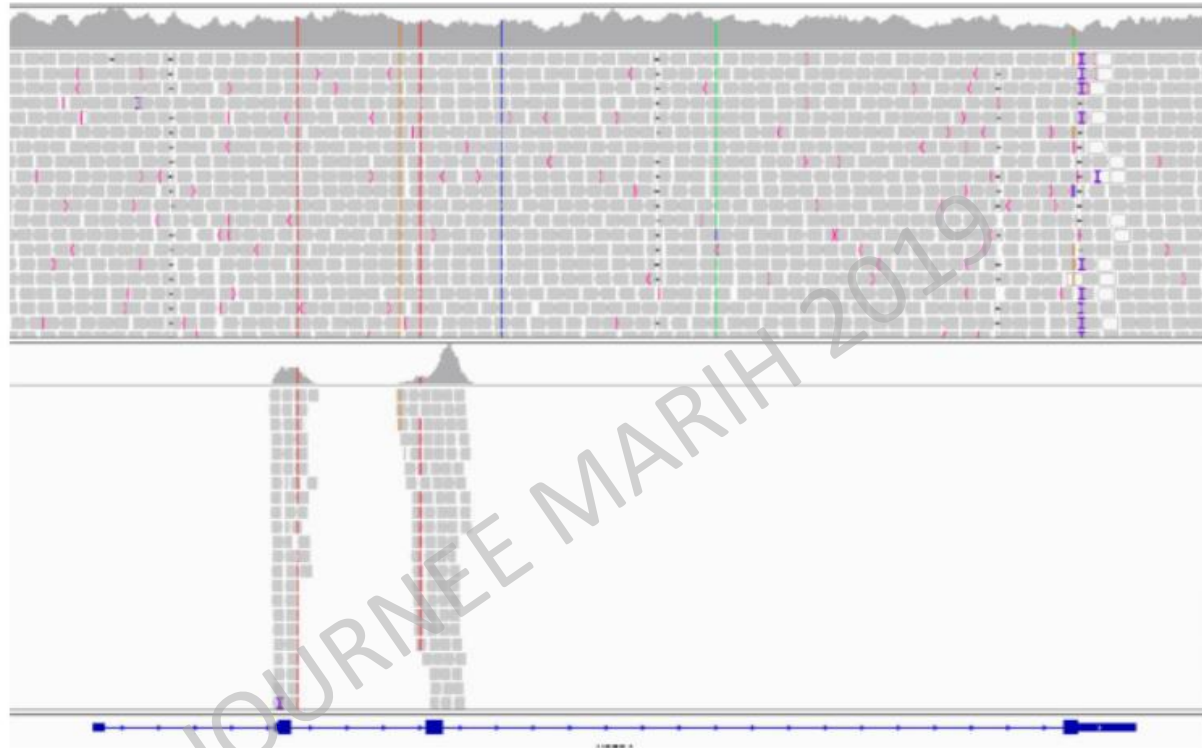
# Complexité du diagnostic : **variants de signification inconnue**



# Impacts sur les stratégies diagnostiques



# De l'exome au Génome



## Avantages

- Pas d'étape d'enrichissement
- Pas (peu) de régions non couvertes
- Homogénéité de la profondeur de lecture
- Améliore détection des CNV

## Limites

- Coût du séquençage
- Volume de stockage des données
- Analyse bioinformatique plus complexe

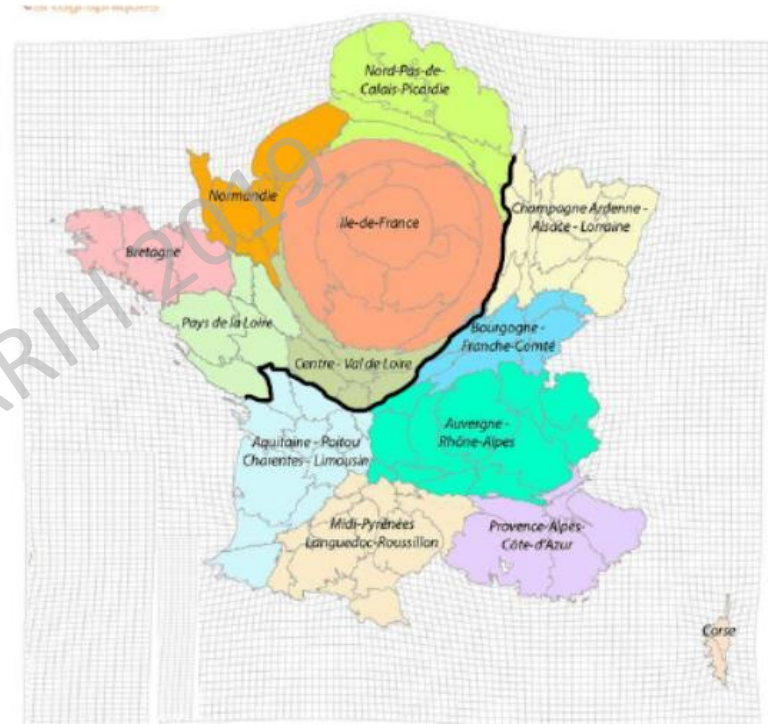
# SeqOIA

LA PLATEFORME GÉNOMIQUE  
DE PARIS RÉGION

- Ile-de-France, Bretagne, Pays de la Loire, Normandie, Hauts de France, Centre Val de Loire



- Auvergne Rhône Alpes, Bourgogne Franche Comté, Grand Est, Languedoc Roussillon, Nouvelle Aquitaine, Occitanie, Provence Alpes Côte d'Azur, Corse, Outre-Mer



# Limites inhérentes au NGS 2<sup>nd</sup>e génération

---

- ❑ **Variants non (mal) détectés par NGS**
  - Expansions de triplets
  - Variants de structure
  
- ❑ **Pseudogènes : difficultés pour l'alignement**
  - *SBDS* : 98% homologie avec pseudogène
  
- ❑ **Régions difficiles à capturer /amplifier**
  - Régions riches en GC

# Futur : Séquenceurs de 3<sup>ème</sup> et 4<sup>ème</sup> génération

Séquençage en temps réel **sur molécule unique sans amplification**

Séquençage **long read**



PACIFIC  
BIOSCIENCES®

bionano®  
GENOMICS

Oxford  
NANOPORE  
Technologies

10X™  
GENOMICS

- **ADN**

- Séquençage des régions riches en GC
- Séquençage des régions répétées
  - Distinction gène / pseudogène
  - Séquençage d'expansions
- Détection des variants de structure
- Reconstruction d'haplotypes

- **ARN**

- Séquençage des longs ARN sur même read



# Impacts du NGS sur le diagnostic des neutropénies congénitales



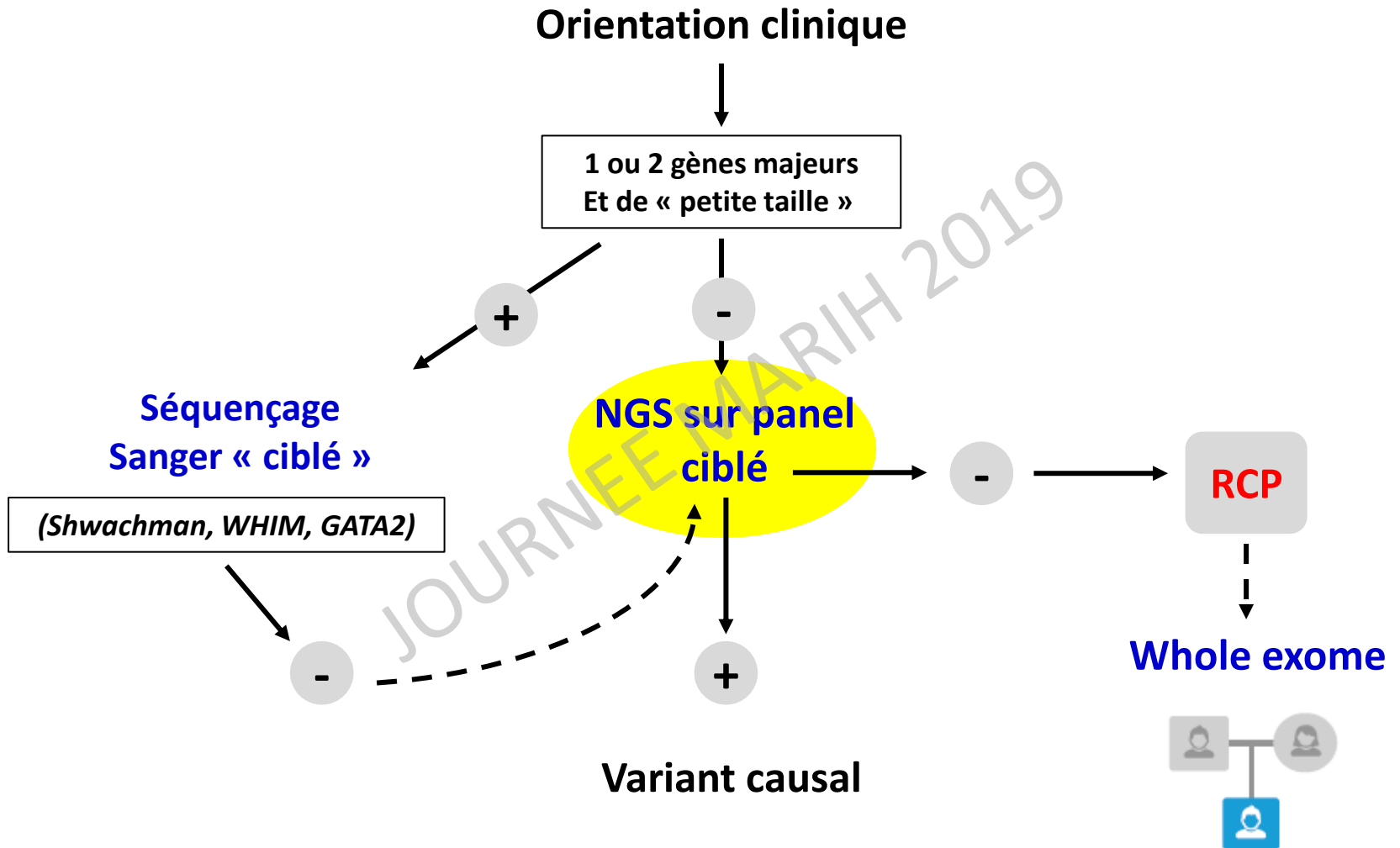
# Contexte

- Entités **très rares** : 1-2 /100 000
- Hétérogénéité clinique
- Hétérogénéité génétique : environ 30 gènes

	ELANE	GFI1	CSF3R	CXCR2	WAS	CXCR4	SBDS	HAX1	G6PC3	JAGN1	GATA2	TCIRG1	VPS45	DNAJC21	EFL1	AP3B1	VPS13B	TAZ	USB1	STK4	CLPB	SMARCD2	LAMTOR2	SLC374A	EIF2AK3	
Avec blocage maturation	■		■	■			■	■	■	■				■	■	■	■	■	■	■	■	■	■	■	■	■
Sans blocage maturation		■		■		■	■			■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Monocytopénie					■	■					■			■	■	■	■	■	■	■	■	■	■	■	■	■
Thrombopénie							■				■			■	■	■	■	■	■	■	■	■	■	■	■	■
Pancréas							■		■	■				■	■	■	■	■	■	■	■	■	■	■	■	■
Tractus digestif								■						■	■	■	■	■	■	■	■	■	■	■	■	■
Coeur					■	■		■						■	■	■	■	■	■	■	■	■	■	■	■	■
Peau						■		■		■				■	■	■	■	■	■	■	■	■	■	■	■	■
Os						■			■				■	■	■	■	■	■	■	■	■	■	■	■	■	■
Appareil uro-génital								■					■	■	■	■	■	■	■	■	■	■	■	■	■	■
SNC						■	■						■	■	■	■	■	■	■	■	■	■	■	■	■	■
Muscle								■					■	■	■	■	■	■	■	■	■	■	■	■	■	■
Surdité	■										■		■	■	■	■	■	■	■	■	■	■	■	■	■	■
Dysmorphie						■							■	■	■	■	■	■	■	■	■	■	■	■	■	■
Inf. non bactériennes					■					■			■	■	■	■	■	■	■	■	■	■	■	■	■	■

*Autosomique dominant, autosomique récessif, lié à l'X*

# La stratégie diagnostique



# Quels gènes inclure dans les panels ?

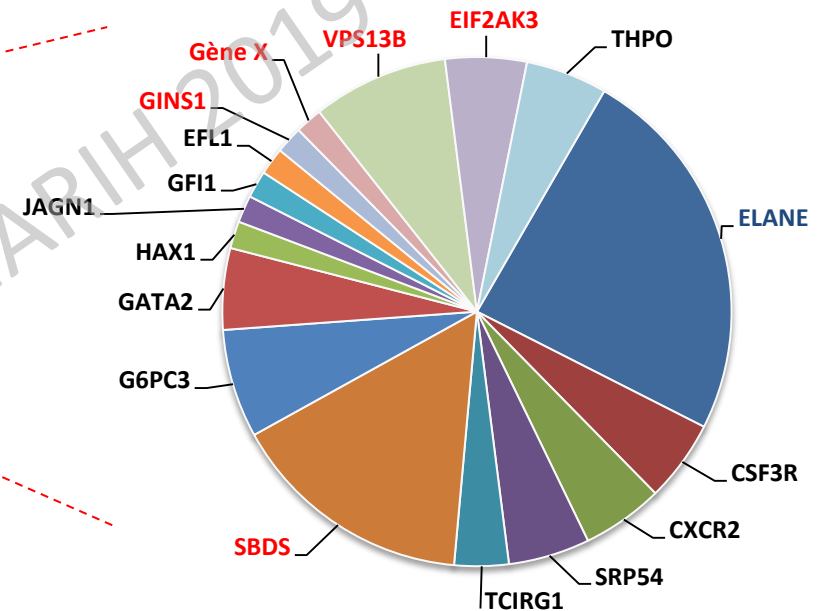
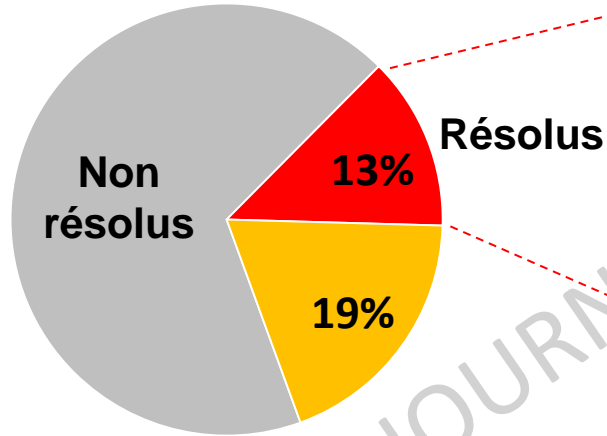
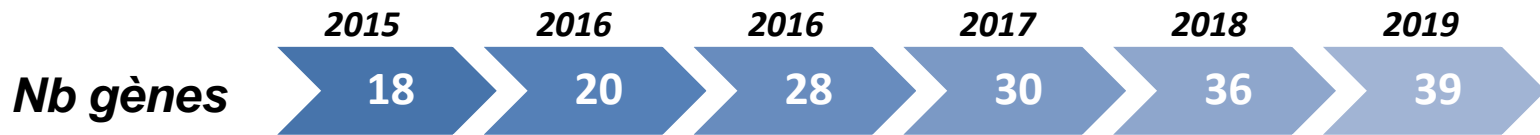
---

- Gènes **d'utilité clinique**
    - Implication dans la physiopathologie démontrée
    - Gènes « actionnables »
  - Gènes **associés à des entités syndromiques**
    - présentation atypique
    - Variants à pénétrance incomplète
  - Gènes **associés à des phénotypes chevauchants**
  - Gènes **identifiés dans le cadre d'exomes**
    - Souvent inclus en pratique pour affirmer l'implication d'un gène
    - Mais résultat rendu uniquement si étude fonctionnelle
- **Discussions généticiens / cliniciens / CRMR – Filières**
- **Panels évolutifs**



# Résultats de l'analyse NGS

- 454 patients (2015 – Mai 2019) – Porte d'entrée « neutropénie chronique »

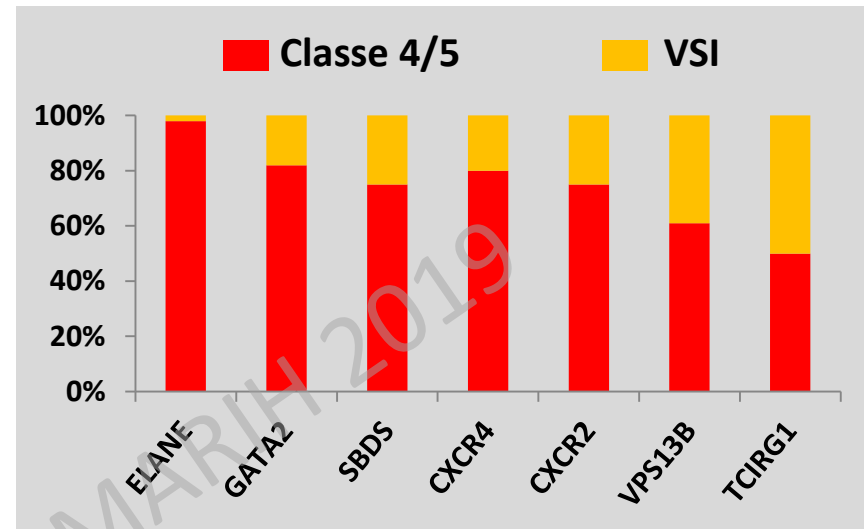
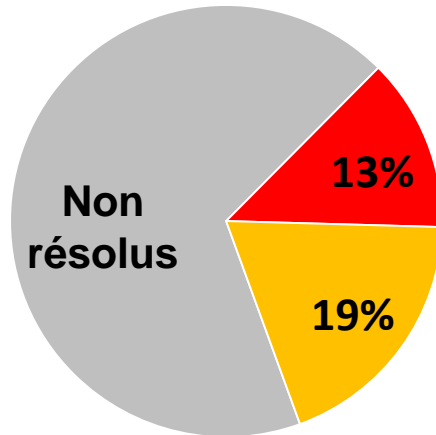


Variants pathogènes (classe 4, 5)

Variants de signification inconnue

- 17 gènes
- 6 gènes (35%) : 1 seul cas identifié
- Découverte fortuite entité syndromique (20%)

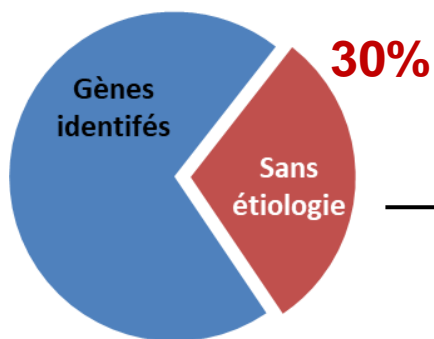
# Rendu et suivi des VSI



## COMPTE RENDU

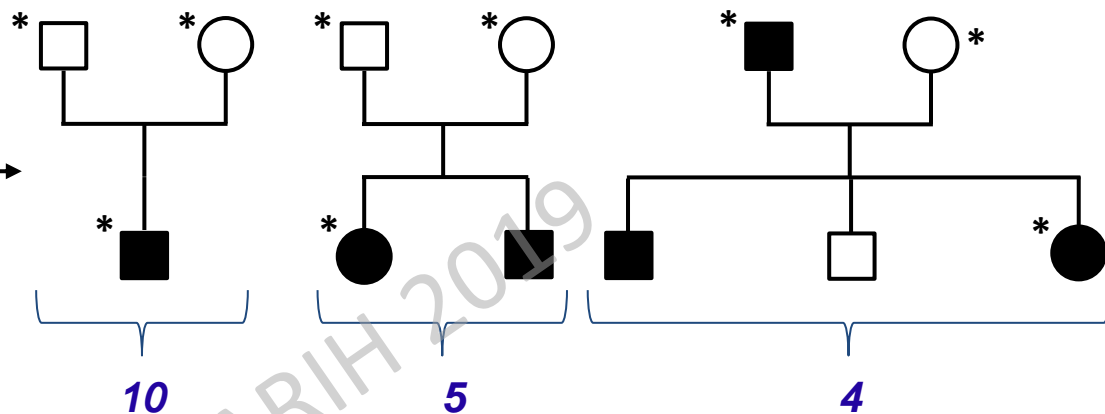
- **Résultat** : « **aucun variant clairement pathogène** »
- **Commentaire** : « **identification d'un VSI** »
  - Poursuivre l'investigation génétique
    - Analyse de ségrégation familiale
  - Discuter les résultats avec les cliniciens
    - Arguments cliniques additionnels ?
  - Poursuivre avec une investigation fonctionnelle
    - Conséquence sur le transcrit
    - Conséquence sur la fonction de la protéine

# Du panel ciblé à l'exome

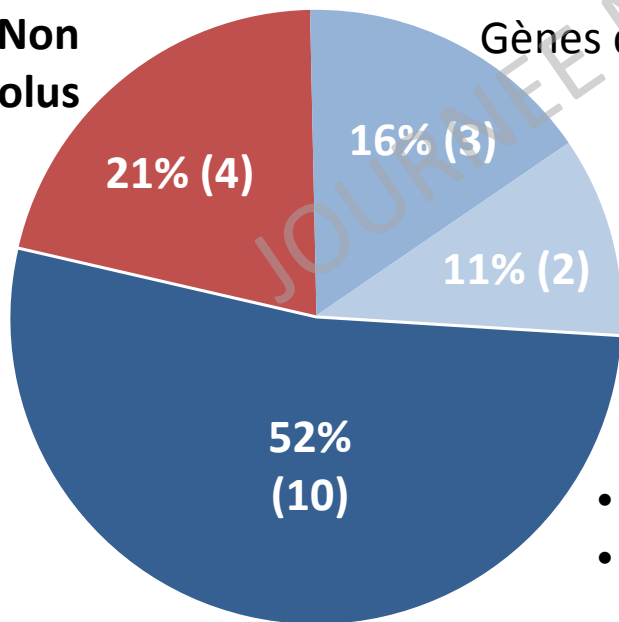


Registre Neutropénies

WES  
trio \*



Non résolu

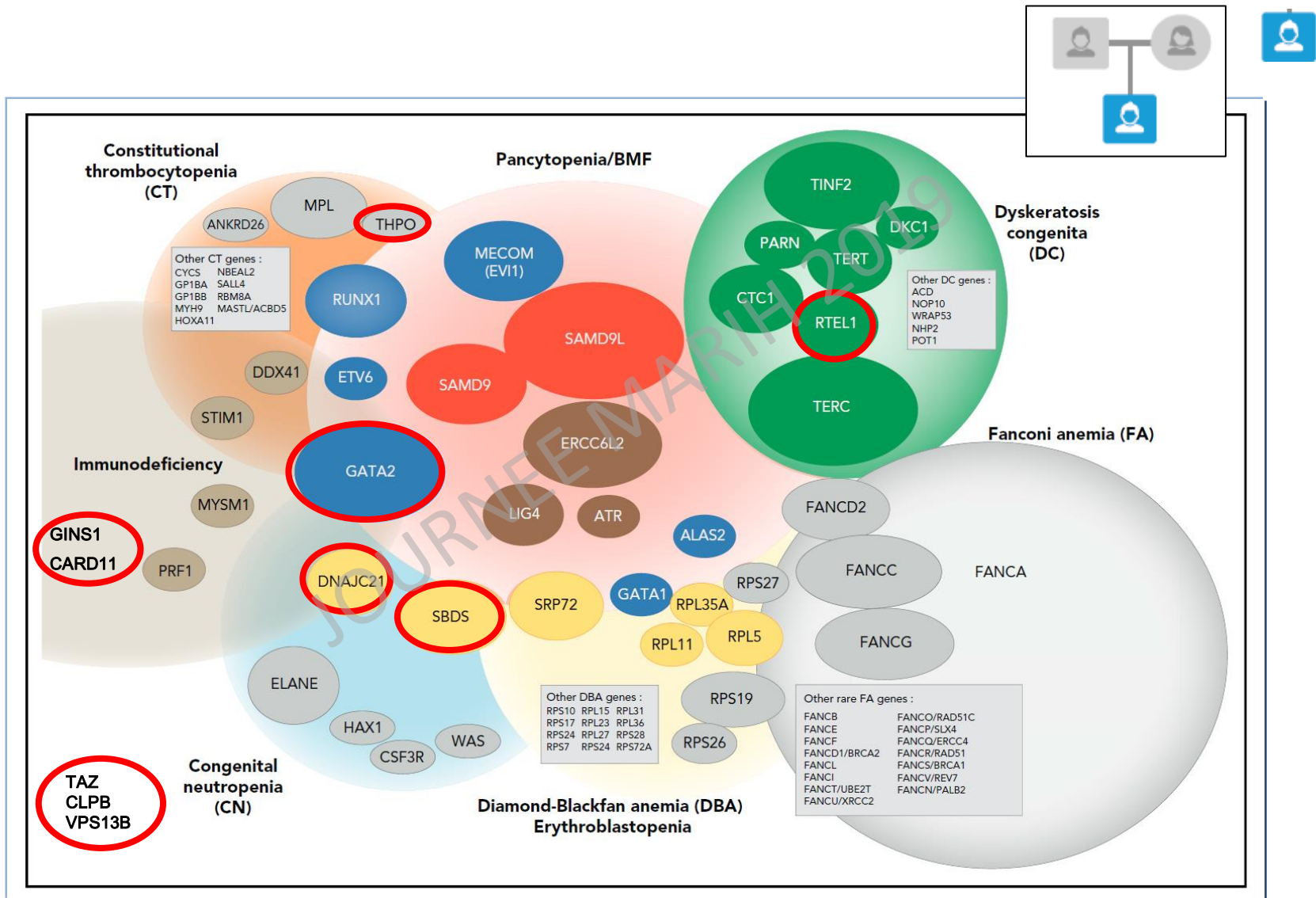


Gènes d'entités **syndromiques**  
(*TAZ*, *CLPB*)

Gènes d'entités **chevauchantes**  
(*RTEL1*, *CARD11*)

- ***SRP54*** (5 familles)
- **5 nouveaux gènes** – 1 cas identifié / gène

# Du panel ciblé au panel virtuel sur exome





# En conclusion

---

- **Vers une harmonisation des tests diagnostiques**
  - Panel de gènes ➤ Panel virtuel /exome / Génome
  - Phase de transition en termes de développements technologiques
- **Complexité du diagnostic par NGS : le nombre et l'interprétation des variants**
  - Interprétation sur un faisceau d'arguments
  - Interprétation évolutive des variants
  - Discuter les résultats avec les cliniciens et les laboratoires « experts » du gène
  - Difficultés de poursuivre les analyses (études familiales, tests fonctionnels) pour tous les variants et gènes « privés »
- **Diagnostic des neutropénies congénitales par NGS**
  - Augmentation du taux diagnostique
  - Ré-évaluation de la fréquence des gènes
  - Découverte fortuite d'entités syndromiques, entités chevauchantes



**Hôpital Pitié-Salpêtrière**  
*Département de Génétique*

**Julien Buratti**  
**Florence Bellanger**  
**Séverine Clauin**  
**Gwendoline Leroy**

**Hôpital Trousseau, Paris**  
*CRMR Neutropénies Chroniques*

**Jean Donadieu**  
**Blandine Beaupain**  
**Hélène Lapillonne**

**Tous les cliniciens des services d'Hématologie**

*Gustave Roussy, Villejuif*  
**INSERM U1170**

**Isabelle Plo**  
**Caroline Marty**  
**Barbara Schmaltz-Panneau**

**Hôpital Robert Debré**  
*Laboratoire d'Hématologie*  
**Odile Fenneteau**

**Sorbonne Université, Paris**  
**CNRS UMR7590**  
**Isabelle Callebaut**

**Hôpital Saint-Antoine**  
*Laboratoire d'Hématologie*  
**François Delhommeau**  
**Pierre Hirsch**

**Institut Imagine**  
**UMR1163**  
**Patrick Revy**

**Université Paris-Saclay**  
**UMR996**  
**Françoise Bachelerie**



Filière de santé Maladies Rares Immuno-Hématologiques Reconnue par le Ministère de la Santé