

DE LA RECHERCHE À L'INDUSTRIE



[www.cea.fr](http://www.cea.fr)

# Gestion informatique des données de séquençage

**William Amory**  
**M1 BI-IPFB Université Paris Cité**

**Laboratoire de Bioinformatique pour la  
Génomique et la Biodiversité  
(Genoscope - LBGB)**

**Sous la responsabilité de Frédéric Gavory**

- 1 Genoscope - LBGB
- 2 Contexte et objectifs de la mission
- 4 Pipeline NGS-RG pour les séquenceurs MGI
- 5 Perspectives

## Plusieurs groupes de travail

- Evaluation des techniques de séquençage
- **Production**
- Assemblage
- Annotation

## Missions du groupe Production

- Répondre aux besoins des équipes de recherches et de productions
- Vielle technologique et évaluation de nouveaux outils
- Développer, tester et maintenir les librairies et scripts
- Mise en place et maintient de pipelines automatisant l'exécution de ces scripts pour le **Genoscope** (centre national de séquençage) et le **CNRGH** (centre national de recherche en génétique humaine)
  - Génération des fichiers de séquences
  - Contrôle qualité et nettoyage des fichiers de séquences
  - Analyses biologiques
- Mise à jour de la base de données de référence NGL  
(*Next Generation LIMS*)

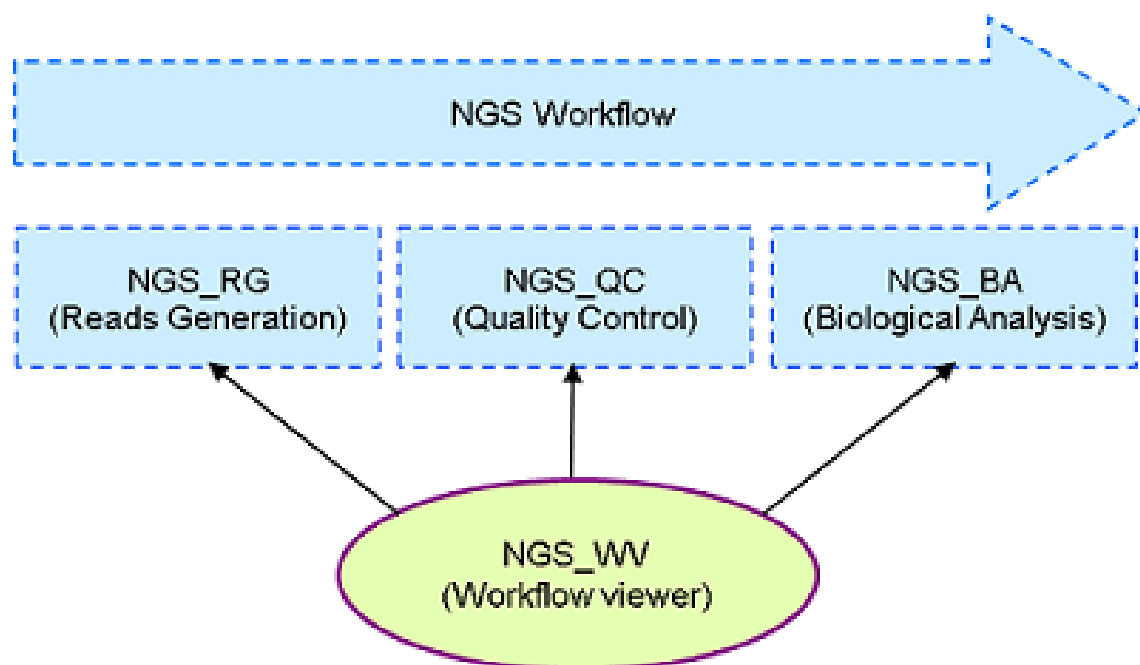


Figure 1 – Workflow de génération, de contrôle qualité et d'analyse biologique des FASTQ

<https://www.genoscope.cns.fr/rdbioseq/> consulté le 21/06/2022

## Arrivée des Séquenceurs MGI



### 2 DNBSEQ-G400

- 2 flowcell - 2/4 pistes
- 1.4 TB
- 5000 Millions de reads
- Taille max des reads :
  - 150pb PE
  - 400pb SE
- Temps moyen d'un run :
  - 24h ~ 30h



### 1 DNBSEQ-T7

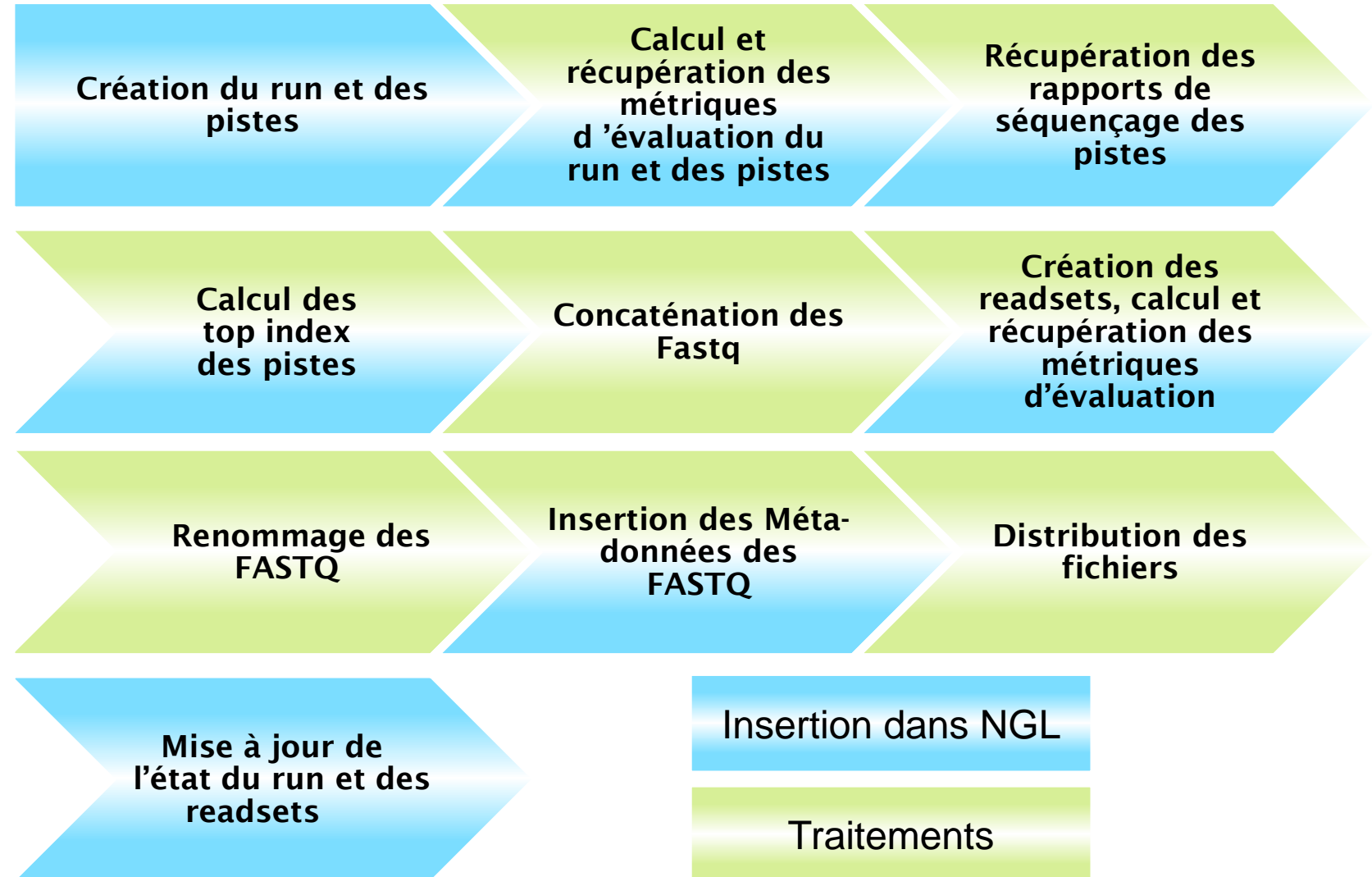
- 4 flowcell - 1 piste
- 6 TB
- 1800 Millions de reads
- Taille max des reads :
  - 200pb PE
  - 400pb SE
- Temps moyen d'un run :
  - 14h ~ 109h

<https://en.mgi-tech.com/products/> consulté le 21/06/2022

## 1 script Perl

Qui fait appel à :

- 14 librairies de traitements de run MGI
- 3 librairies communes à tous les traitements de run MGI
- 11 librairies d'interaction avec NGL pour les run MGI
- 1 librairie commune à tous les type de run



Création du run et des pistes

Calcul et récupération des métriques d'évaluation du run et des pistes

Récupération des rapports de séquençage des pistes

Calcul des top index des pistes

Concaténation des Fastq

Création des readsets, calcul et récupération des métriques d'évaluation

Renommage des FASTQ

Insertion des méta-données des Fastq

Distribution des fichiers

Mise à jour du run et des readsets

## Création du run et des pistes

### Traitements

- Création d'un répertoire temporaire de traitement du run

### NGL

- Création du run
- Création des piste

### Objectifs

- Rendre disponible les informations à propos du run et l'état de traitement de celui-ci aux utilisateurs

NGL-BI-UAT **Runs** Readsets Analyses Statistiques Archives Bilans Descriptions Aide wamory

Recherche de Runs x

220620\_MUSHU\_V300... x

## 220620\_MUSHU\_V300064083 Read generation en cours

Général [Infos workflow](#)

Code	220620_MUSHU_V300064083	Type	DNBSeq G400	Nb Cycles	Code Instrument	MUSHU
Etat	Read generation en cours	Nb Clusters (total)		Ligne Contrôle	Type d'Instrument	DNBSEQG400
Valide ?	---	% Clusters filt. (moyenne)		Code Flowcell	Date Run	20/06/2022
Comptes Rendus		Nb Clusters filt. (total)		Position Flowcell	Date fin RG	
Critères		Nb Bases (total)		Version RTA		
Évalué par	()	A conserver ? <input type="checkbox"/>		Version Flowcell		
		Supprimé	Non			

Détails évaluation

Pistes

#	Valide ?	Comptes Rendus
1	---	
2	---	
3	---	
4	---	

Création du run et des pistes

Calcul et récupération des métriques d'évaluation du run et des pistes

Récupération des rapports de séquençage des pistes

Calcul des top index des pistes

Concaténation des Fastq

Création des readsets, calcul et récupération des métriques d'évaluation

Renommage des FASTQ

Insertion des méta-données des Fastq

Distribution des fichiers

Mise à jour du run et des readsets

## Calcul et récupération des métriques d'évaluation du run et des pistes

### Traitements

- Récupération des métriques du run et de pistes
  - Nombre de cycle des reads et des index
  - Nombre de reads, de bases
  - Pourcentage de Q30
  - Pourcentage de perte après le premier démultiplexage
  - Etc.

### NGL

- Insertion des métriques du run et des pistes

### Objectifs

- Permettre l'évaluation du run et des pistes

NGL-BI-UAT

Runs

Readsets

Analyses

Statistiques

Archives

Bilans

Descriptions

Aide

wamory

Comptes Rendus

Critères

Évalué par

Nb Clusters filt.  
(total)

Nb Bases (total)

A conserver ?

Supprimé

Position Flowcell

Version RTA

Version Flowcell

A

Date fin RG

193 988 797 170

☐

Non

Détails évaluation

NGS-RG

#	Nb Cycles Utiles	Nb reads (total)	%ESR	%q30	%q20	%q10	%N	Recover Value	%Chip productivity	Nb bases	%Runon
1	100, 10, 0, 100	490 877 409	81,07	89,465	97,002	99,213	3 190 822,320	2,63	80,69	98 175 481 800	0,04
2	100, 10, 0, 100	432 878 768	71,16	88,288	96,568	99,107	5 910 649,816	2,88	71,16	86 575 753 600	0,04
3	100, 10, 0, 100	456 548 442	75,58	88,935	96,824	99,212	1 610 252,134	2,53	75,06	91 309 688 400	0,05
4	100, 10, 0, 100	446 030 287	73,33	89,835	96,979	99,178	7 291 856,438	2,89	73,33	89 206 057 400	0,04

Pistes

#	Valide ?	Comptes Rendus
1	---	
2	---	
3	---	
4	---	

Pistes		
#	Valide ?	Comptes Rendus
1	---	
2	---	
3	---	
4	---	

Création du run et des pistes

Calcul et récupération des métriques d'évaluation du run et des pistes

Récupération des rapports de séquençage des pistes

Calcul des top index des pistes

Concaténation des Fastq

Création des readsets, calcul et récupération des métriques d'évaluation

Renommage des FASTQ

Insertion des méta-données des Fastq

Distribution des fichiers

Mise à jour du run et des readsets

## Récupération des rapports de séquençage des pistes

### Traitements

- Récupération des rapport de séquençage des piste
  - Contenu du rapport html
  - Extension
  - Nom du rapport

### NGL

- Insertion des rapport de séquençage des pistes

### Objectifs

- Permettre l'évaluation des pistes
- Contient plusieurs tableaux de métriques et de graphique d'évaluation des pistes

Tab1. Summary Information

Category	Value
SoftwareVersion	1.0.8.208
TemplateVersion	0.8.0
Reference	NULL
CycleNumber	210
ChipProductivity(%)	80.69
ImageArea	432
TotalReads(M)	485.06
Q30(%)	89.57
SplitRate(%)	98.82
Runon1(%)	0.04
Runon2(%)	0.05
Lag1(%)	0.14
Lag2(%)	0.17
ESR(%)	81.07
MaxOffsetX	26.24
MaxOffsetY	20.61
InitialOffsetX	16.37
InitialOffsetY	16.48
RecoverValue(A)	2.24
RecoverValue(C)	2.93
RecoverValue(G)	2.97
RecoverValue(T)	2.37
RecoverValue(AVG)	2.63

Tab2. Biochemistry Information

Category	Value
ISW Version	1.0.0.34
Machine ID	R13040100200006
Sequence Type	PE100
Recipe Version	V1.4.0.176
Sequence Date	2020-07-21
Sequence Time	12:42:50
Reagent ID	W2006010541
Flowcell Pos	A
DNB ID	TEST_circu
Barcode Type	1~128
Barcode File	barcodeAL01.csv
Read1 Cycles	100
Read2 Cycles	100
Barcode	10
Dual Barcode	
Read1 Dark Cycles	
Read2 Dark Cycles	

Fig14. Average Quality Distribution

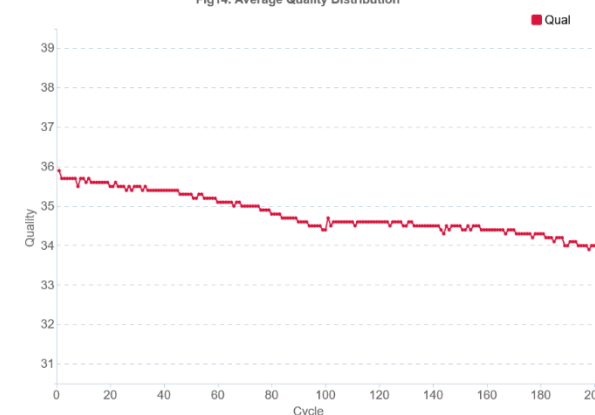
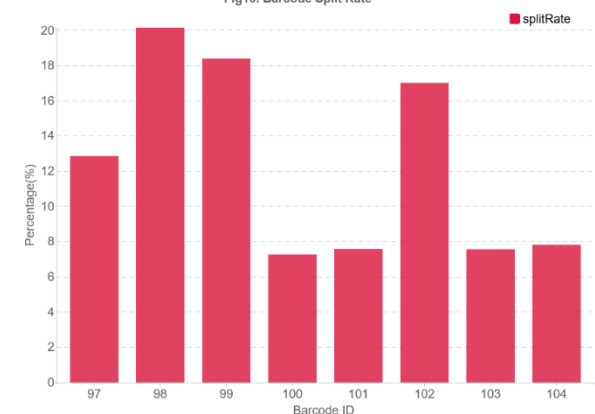


Fig10. Barcode Split Rate





Création du run et des pistes

Calcul et récupération des métriques d'évaluation du run et des pistes

Récupération des rapports de séquençage des pistes

Calcul des top index des pistes

Concaténation des Fastq

Création des readsets, calcul et récupération des métriques d'évaluation

Renommage des FASTQ

Insertion des méta-données des Fastq

Distribution des fichiers

Mise à jour du run et des readsets

## Calcul des top index des pistes

### Traitements

- Récupération des index représenté à plus de 0.01% de la pistes et des index attendus
  - Trie des index par ordre décroissant

### NGL

- Insertion top index par piste

### Objectifs

- Permettre de vérifier que les index attendus par pistes sont bien majoritairement représentés

NGL-BI-UAT	Runs	Readsets	Analyses	Statistiques	Archives	Bilans	Descriptions	Aide	wamory
NGS-RG Rapport séquençage MGI Démultiplexage MGI									
Lane 1									
barcode	count	percent							
barcode98	98 898 442	20,147							
barcode99	90 340 138	18,404							
barcode102	83 544 661	17,019							
barcode97	63 193 710	12,874							
barcode104	38 393 930	7,821							
barcode101	37 241 062	7,587							
barcode103	37 183 005	7,575							
barcode100	35 681 639	7,269							
NNNNNNNNNN	1 149 326	0,234							
barcode57	319 055	0,065							
GTTGCATCGT	244 878	0,050							
CGCCGTGAAT	212 841	0,043							
GTGCATTCGT	207 168	0,042							
ACGTCGATCT	204 918	0,042							
TTGCATTCGT	193 336	0,039							
GTTGATTCGT	135 657	0,028							
ACGCGGATCT	97 527	0,020							
barcode58	90 788	0,018							
TTCCGCGAGT	86 393	0,018							
GTTCATTCGT	71 888	0,015							

Pistes		
#	Valide ?	Comptes Rendus
1	---	
2	---	
3	---	
4	---	

Création du run  
et des pistes

Calcul et  
récupération des  
métriques  
d'évaluation du  
run et des pistes

Récupération des  
rapports de  
séquençage des  
pistes

Calcul des top  
index des pistes

Concaténation  
des Fastq

Création des  
readsets, calcul et  
récupération des  
métriques  
d'évaluation

Renommage des  
FASTQ

Insertion des  
méta-données des  
Fastq

Distribution des  
fichiers

Mise à jour du  
run et des  
readsets

## Concaténation des FASTQ

### Traitements

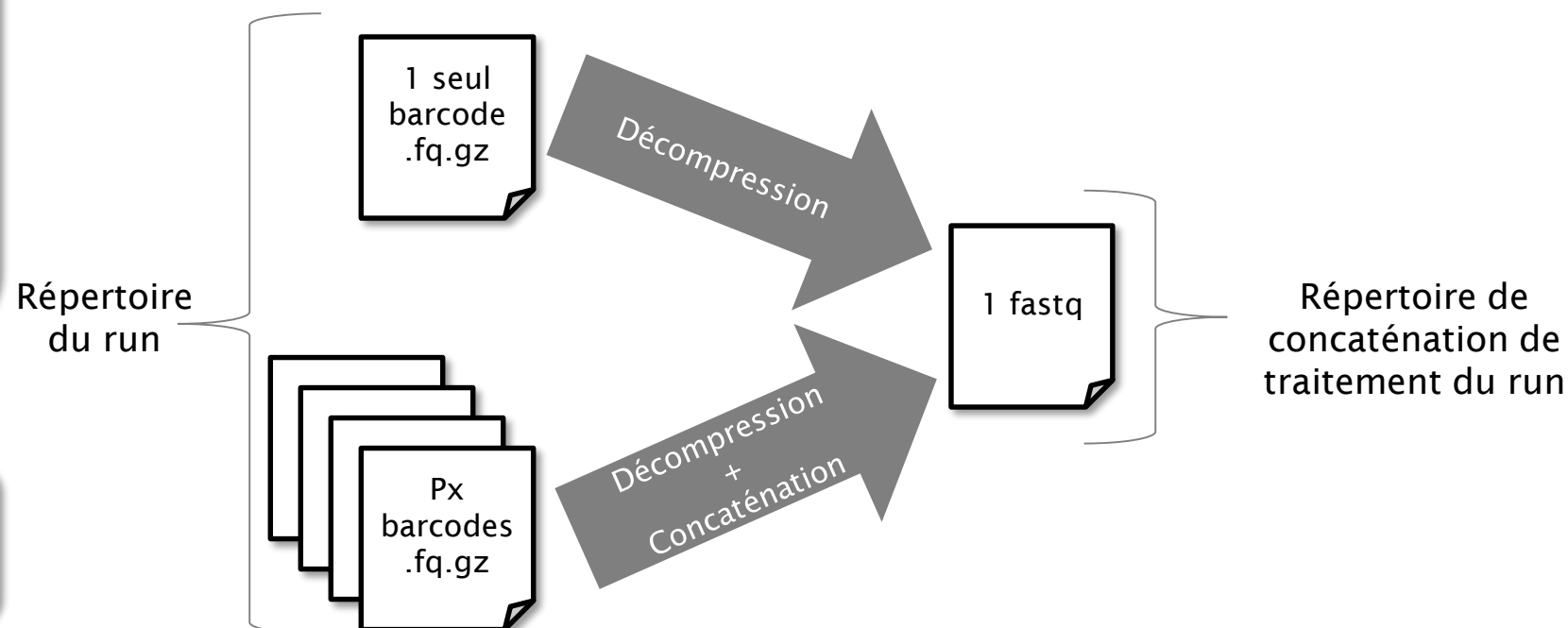
- Si un seul index :
  - Décompression et renommage du FASTQ
- Si plusieurs index :
  - Décompression, concaténation et renommage des FASTQ

La décompression et la concaténation est réalisé avec **unpigz** sur 2 threads

### Objectifs

Obtenir un seul FASTQ par readset

La technologie MGI requiert une homogénéité des bases pour chaque cycle des index  
Le démultiplexage génère un FASTQ par index connu  
Un échantillon peut être divisé en plusieurs fichiers



Création du run  
et des pistes

Calcul et  
récupération des  
métriques  
d'évaluation du  
run et des pistes

Récupération des  
rapports de  
séquençage des  
pistes

Calcul des top  
index des pistes

Concaténation  
des Fastq

Création des  
readsets, calcul et  
récupération des  
métriques  
d'évaluation

Renommage des  
FASTQ

Insertion des  
méta-données des  
Fastq

Distribution des  
fichiers

Mise à jour du  
run et des  
readsets

## Création des readsets, calcul et récupération des métriques d'évaluation

### Traitements (3 traitements)

- NGSRG
  - Nombre de reads
  - Nombre de bases
  - Qualité moyenne
  - Etc.
- Global (sera mis à jour par NGS-QC)
  - Nombre de reads
  - Nombre de bases

### NGL

- Création des readsets
- Insertion des métriques d'évaluation des readsets

### Objectifs

Permettre l'évaluation des readsets

NGL-BI-UAT   Runs ▾   **Readsets ▾**   Analyses ▾   Statistiques ▾   Archives   Bilans ▾   Descriptions ▾   Aide   wamory ▾

Général   **Avancé**   Infos échantillon   Infos workflow

Code	CTB_DA_AA_1_V300064083.BC10Nb Séquences utiles	75 576 935	Run / N° Piste	220620_MUSHU_V3000640 / 1	
Etat	Read generation en cours	Nb Bases utiles	15 115 387 000	Type de Run	RDNBG400
Valide QC ?	---	Valide BioInfo ?	---	Nb Cycles	
Comptes Rendus QC		Comptes Rendus BioInfo		Date Run	20/06/2022
Critères QC		Critères BioInfo		Date fin RG	
Évalué par	()	Évalué par	()	Date fin QC	

Détails évaluation

NGS-RG

Nb reads	% déposé	Nb bases	% séquences valides/piste
75576935	4.175	15115387000	15.3962952081993



## Renommage des FASTQ

### Traitements

- Renommage des Fastq selon le format utilisé au Genoscope et CNRGH

Le renommage est réalisé par des liens symboliques des FASTQ issue du traitement de concaténation

### Objectifs

- Tous les fichiers de séquences doivent avoir un nom unique et « parlant »
- Doit permettre d'identifier à quel échantillon, projet, flowcell, ect le fichier appartient

Le format des FASTQ finaux sont :

<project\_code>\_<bank\_type>\_<sample\_code>\_<lane\_number>\_<read\_number>\_<flowcell\_id>.BC  
<first\_barcode>.fastq

Ex : ABC\_DA\_AAAA\_1\_2\_F0123456789.BC3.fastq

Répertoire de concaténation du répertoire de traitement du run

concat\_L01\_CTB\_AA\_1.BC103.fastq

Répertoire de renommage des FASTQ du répertoire de traitement du run

CTB\_DA\_AA\_1\_1\_V300064083.BC103.fastq

Lien symbolique

Création du run et des pistes

Calcul et récupération des métriques d'évaluation du run et des pistes

Récupération des rapports de séquençage des pistes

Calcul des top index des pistes

Concaténation des Fastq

Création des readsets, calcul et récupération des métriques d'évaluation

Renommage des FASTQ

Insertion des méta-données des Fastq

Distribution des fichiers

Mise à jour du run et des readsets

## Insertion des méta-données des FASTQ

### Traitements

- Récupération de l'extension et du type d'encodage de la qualité
- Construction du chemin du répertoire des fichiers
- Construction du label

### NGL

- Insertion des méta-données des FASTQ de chaque readsets

### Objectifs

- Décrire les fichiers disponible pour un readset, ainsi que leurs emplacement dans le système de fichiers

NGL-BI-UAT Runs **Readsets** Analyses Statistiques Archives Bilans Descriptions Aide wamory

Recherche de Readsets

**CTB\_DA\_AA\_1\_V300064083.BC103** Read generation en cours

Général **Avancé** Infos échantillon Infos workflow

SSID

Date de l'archive

Chemin fichiers utiles /env/ig/atelier/mgi/cns/DEV\_MGI/TEST\_DISPATCH/Test\_Dispatch\_Fastq/proj/projet\_CTBA/AA/RunsMGI/220620\_MUSHU\_V300064083/

Localisation CNS

Envoyé Collaborateur ? ☐

Etat pour la soumission Pas associé à une soumission

Nom du fichier	Type de fichier	Utilisable	Label	Encodage ASCII	Clé codage md5	Nom fichier collaborateur
CTB_DA_AA_1_1_V300064083.BC103.fastq	RAW	Oui	READ1	33		
CTB_DA_AA_1_2_V300064083.BC103.fastq	RAW	Oui	READ2	33		

NGS-RG

Index	Nb occurrences	% de cet index dans le readset
barcode104	38 393 930	50,801
barcode103	37 183 005	49,199



## Distribution des fichiers (FASTQ attendus)

### Traitements

- Dispatch des FASTQ attendus dans le répertoire final de l'échantillon du projet du run
- Changement des droits d'accès au FASTQ (droits restrictifs)

### Objectifs

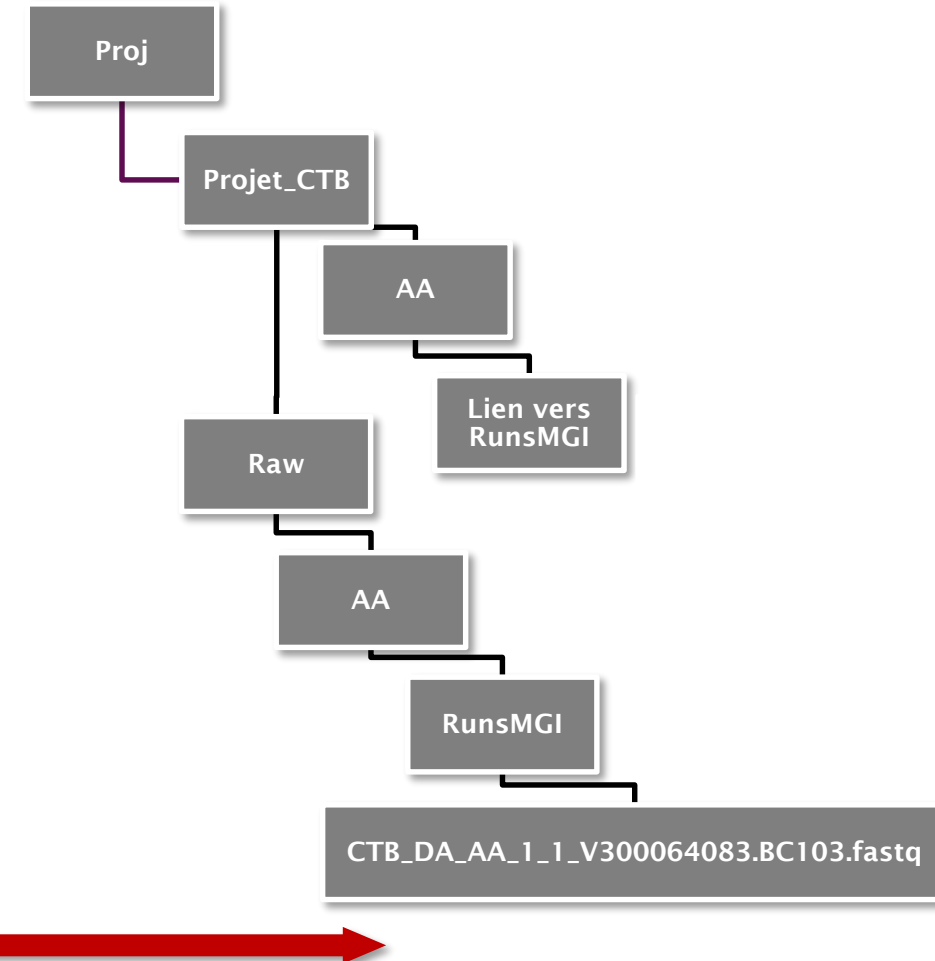
- Rendre disponible les fichiers de séquences, tous en restreignant les droits d'écriture, de lecture et d'exécution aux utilisateurs

Répertoire de renommage de traitement du run

CTB\_DA\_AA\_1\_2\_V300064083.BC103.fastq

copie + chmod

## Organisation du système de fichier des projets



Création du run  
et des pistes

Calcul et  
récupération des  
métriques  
d'évaluation du  
run et des pistes

Récupération des  
rapports de  
séquençage des  
pistes

Calcul des top  
index des pistes

Concaténation  
des Fastq

Création des  
readsets, calcul et  
récupération des  
métriques  
d'évaluation

Renommage des  
FASTQ

Insertion des  
méta-données des  
Fastq

Distribution des  
fichiers

Mise à jour du  
run et des  
readsets

## Distribution des fichiers (fichiers de statistiques)

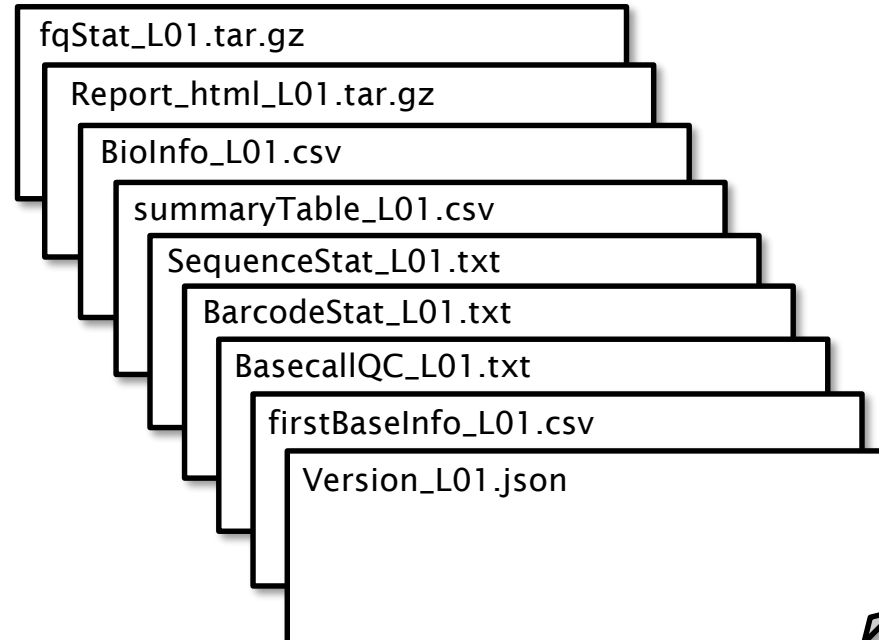
### Traitements

- Dispatch des fichiers de Statistique dans le répertoire dédié
  - 1 archive compressée pour les \*.fqStat.txt par piste
  - 1 archive compressée pour les fichiers html par piste
  - Renommage des autres fichiers de statistiques en ajoutant le numéro de la piste
- Changement des droits d'accès aux fichiers de statistiques (droits restrictifs)

### Objectifs

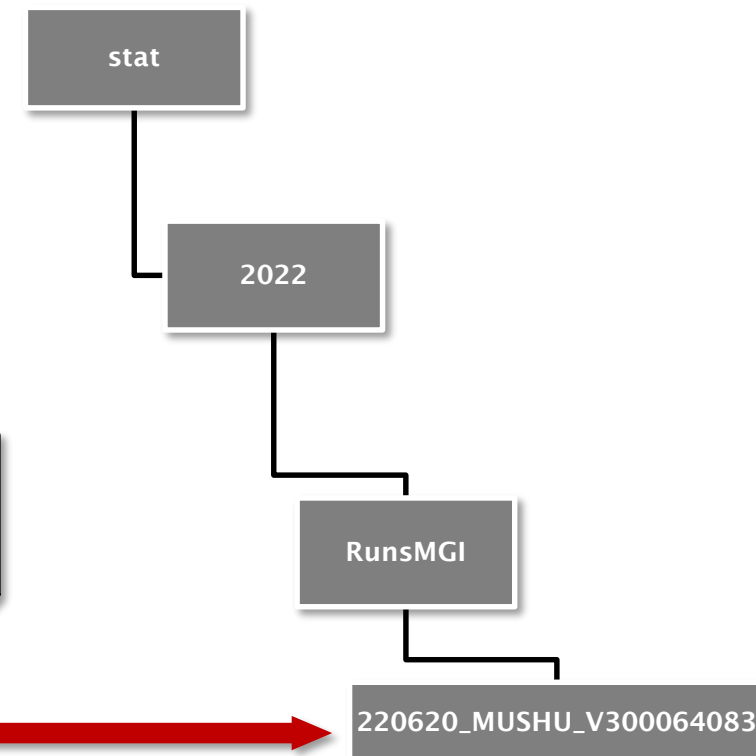
- Conserver les fichiers de statistiques du run

Répertoire de renommage et d'archivage des fichiers de statistiques de traitement du run



Copie + chmod

Organisation du système de fichier des statistiques des runs





Création du run  
et des pistes

Calcul et  
récupération des  
métriques  
d'évaluation du  
run et des pistes

Récupération des  
rapports de  
séquençage des  
pistes

Calcul des top  
index des pistes

Concaténation  
des Fastq

Création des  
readsets, calcul et  
récupération des  
métriques  
d'évaluation

Renommage des  
FASTQ

Insertion des  
méta-données des  
Fastq

Distribution des  
fichiers

Mise à jour du  
run et des  
readsets

## Distribution des fichiers (Fastq non attendus)

### Traitements

- Renommage des FASTQ non attendus selon le format utilisé au Genoscope
- Dispatch des FASTQ non attendus dans le répertoire dédié
  - 1 archive compressée des FASTQ non attendus par piste
- Changement des droits d'accès au FASTQ (droits restrictifs)

### Objectifs

- Conserver les FASTQ non-attendus du run en cas de mauvaise déclaration de la composition en barcode des readsets

### Format de renommage :

BC<barcode\_number>\_<lane\_number>\_<read\_number>\_<flowcell\_code>.fastq.gz

Pour les FASTQ de barcode non connus BC<barcode\_number> remplacé par UNKNOWN

Répertoire de renommage et d'archivage des FASTQ non attendus de traitement du run

Unexpected\_Fastq\_L01.tar.gz

Unexpected\_Fastq\_L02.tar.gz

Unexpected\_Fastq\_L03.tar.gz

Unexpected\_Fastq\_L04.tar.gz

Copie + chmod

Organisation du système de fichier des FASTQ non attendus des runs

noDispatch

2022

RunsMGI

220620\_MUSHU\_V300064083



Création du run et des pistes

Calcul et récupération des métriques d'évaluation du run et des pistes

Récupération des rapports de séquençage des pistes

Calcul des top index des pistes

Concaténation des Fastq

Création des readsets, calcul et récupération des métriques d'évaluation

Renommage des FASTQ

Insertion des méta-données des Fastq

Distribution des fichiers

Mise à jour du run et des readsets

## Mise à jour de l'état du run et des readsets

### Mise à jour NGL

- Mise à jour du run et des readset en cascade à « Fin de génération de reads »
- Mise à jour automatique du run à « Evaluation en attente »
- Mise à jour automatique des readset à « Contrôle qualité en attente »

### Objectifs

- Indiquer que l'évaluation du run peut être réalisé
- Indiquer au pipeline NGS-QC qu'il peut réaliser le contrôle qualité des readsets

NGL-BI-UAT **Runs** Readsets Analyses Statistiques Archives Bilans Descriptions Aide wamory

Runs à Evaluer  
220620\_MUSHU\_V300064083

220620\_MUSHU\_V300064083 Evaluation en attente

Général Infos workflow

Code	220620_MUSHU_V300064083	Type	DNBSeq G400	Nb Cycles	100, 10, 0, 100	Code Instrument	MUSHU
Etat	Evaluation en attente	Nb Clusters (total)	923 756 177	Ligne Contrôle		Type d'Instrument	DNBSEQG400

NGL-BI-UAT **Readsets** Analyses Statistiques Archives Bilans Descriptions Aide wamory

Recherche de Readsets  
CTB\_DA\_AA\_1\_V300064083

CTB\_DA\_AA\_1\_V300064083.BC103 Contrôle qualité en attente

Mode impression

Général Avancé Infos échantillon Infos workflow

Code	CTB_DA_AA_1_V300064083.BC103	Nb Séquences utiles	75 576 935	Run / N° Piste	220620_MUSHU_V3000640 / 1
Etat	Contrôle qualité en attente	Nb Bases utiles	15 115 387 000	Type de Run	RDNBG400

## Pipeline NGS-RG MGI

- Ajout du second démultiplexage (démidage) pour les run comportant des mid
- Benchmark du pipeline (identifier les étapes les plus longues en temps, les plus consommatrice de mémoire et les optimiser)

## Workflow NGS MGI

- Ecriture du pipeline NGS-QC MGI
- Mise en production des 2 pipelines (NGS-RG et NGS-QC MGI)

## Evaluation d'autres outils

- Outils d'assignation taxonomique
- Outils de *trimming* (pour NGS-QC)
- Intégration d'outils des autres groupes de travail dans les pipelines NGS-BA
  - Outils d'assemblage, *scaffolding*, *mapping* ...

- Impact of sequencing depth and technology on de novo RNA-Seq assembly. Patterson. 2022-01-23, *BMC Genomics*. <https://doi.org/10.1186/s12864-019-5965-x>
- Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. Jeon, S.A., Park, J.L., Park, S.J. and al. *Genes Genom* **43**, 713–724 (2021). <https://doi.org/10.1007/s13258-021-01096-x>
- Best practices for the interpretation and reporting of clinical whole genome sequencing. Austin-Tse, C.A., Jobanputra, V., Perry, D.L. and al. *npj Genom. Med.* **7**, 27 (2022). <https://doi.org/10.1038/s41525-022-00295-z>
- Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing. Hak-Min Kim and al. *GigaScience*, Volume 10, Issue 3, March 2021, giab014, <https://doi.org/10.1093/gigascience/giab014>
- Highly comparable metabarcoding results from MGI-Tech and Illumina sequencing platforms. Anslan S, Mikryukov V, and al. 2021. *PeerJ* 9:e12254 <https://doi.org/10.7717/peerj.12254>
- CoolMPS™: Advanced massively parallel sequencing using antibodies specific to each natural nucleobase. Snezana Drmanac, Matthew Callow and al. *bioRxiv preprint*. <https://doi.org/10.1101/2020.02.19.953307>
- perl - The Perl 5 language interpreter - Perldoc Browser. 2022-01-23, <https://perldoc.perl.org/perl>
- The Comprehensive Perl Archive Network. 2022-01-23, [www.cpan.org](http://www.cpan.org)

DE LA RECHERCHE À L'INDUSTRIE

cea



Université  
Paris Cité



[www.cea.fr](http://www.cea.fr)

**Merci de votre attention**

**William Amory**  
**M1 BI-IPFB Université Paris Cité**

**Laboratoire de Bioinformatique pour la  
Génomique et la Biodiversité  
(Genoscope - LBGB)**

**Sous la responsabilité de Frédérick Gavory**

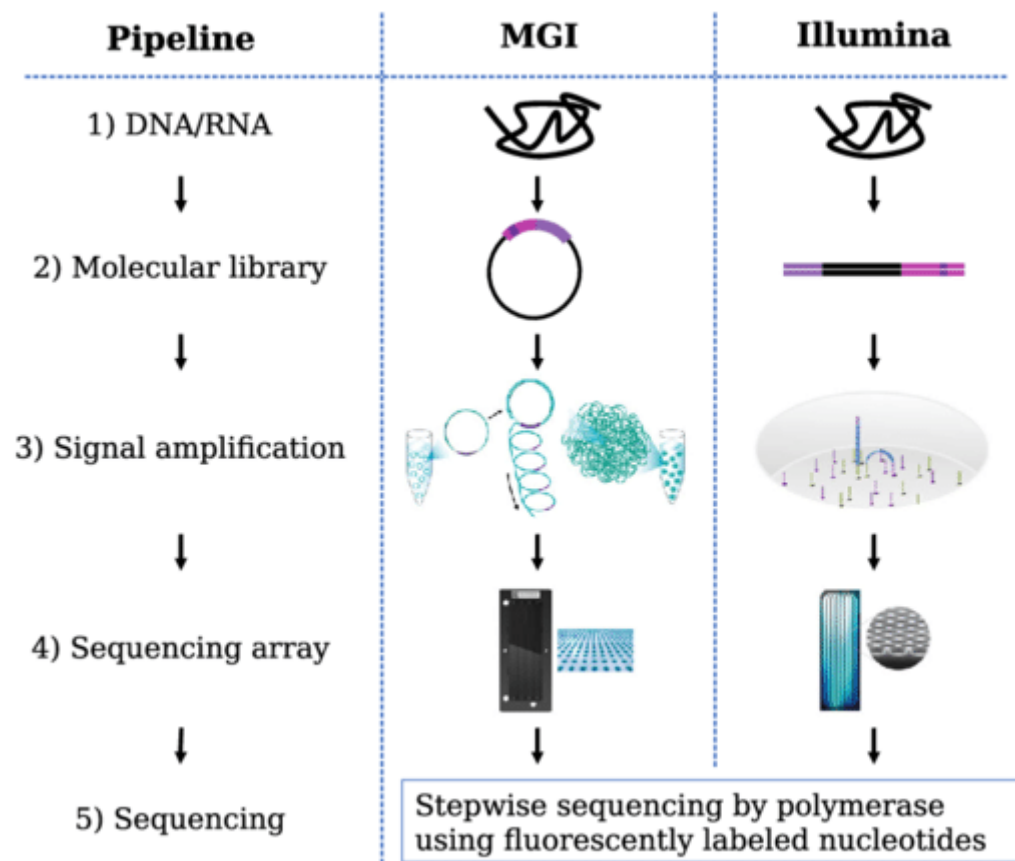


Figure 2 – Différences entre Illumina et MGI de technologie NGL

J. Patterson & all. (2019). Impact of sequencing depth and technology on de novo RNA-Seq assembly. BMC Genomics. 20. 10.1186/s12864-019-5965-x.

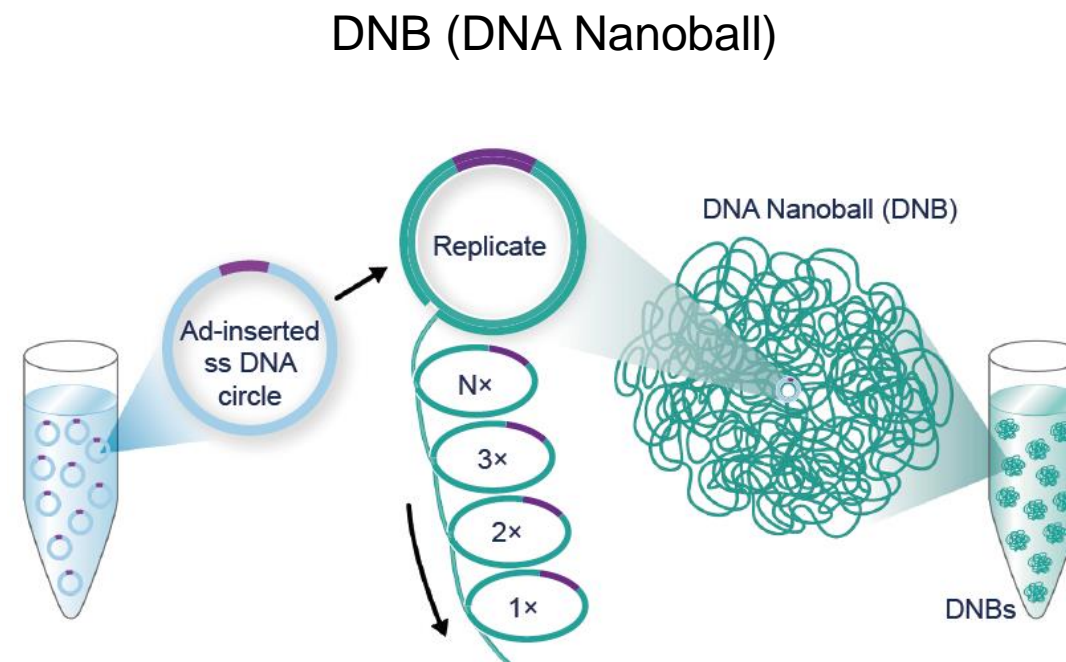


Figure 3 – Schéma de la technologie des *DNA nanoballs* de MGI

<https://en.mgi-tech.com/products/> consulté le 21/06/2022

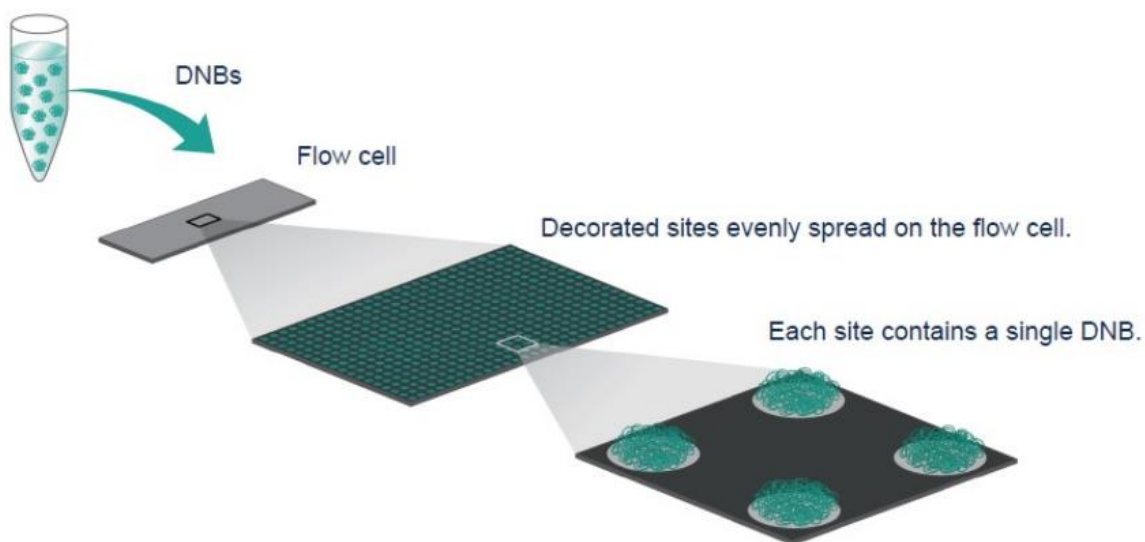


Figure 4 – Schéma d'une flowcell et des DNB dans les puits de la flowcell

<https://en.mgi-tech.com/products/> consulté le 21/06/2022

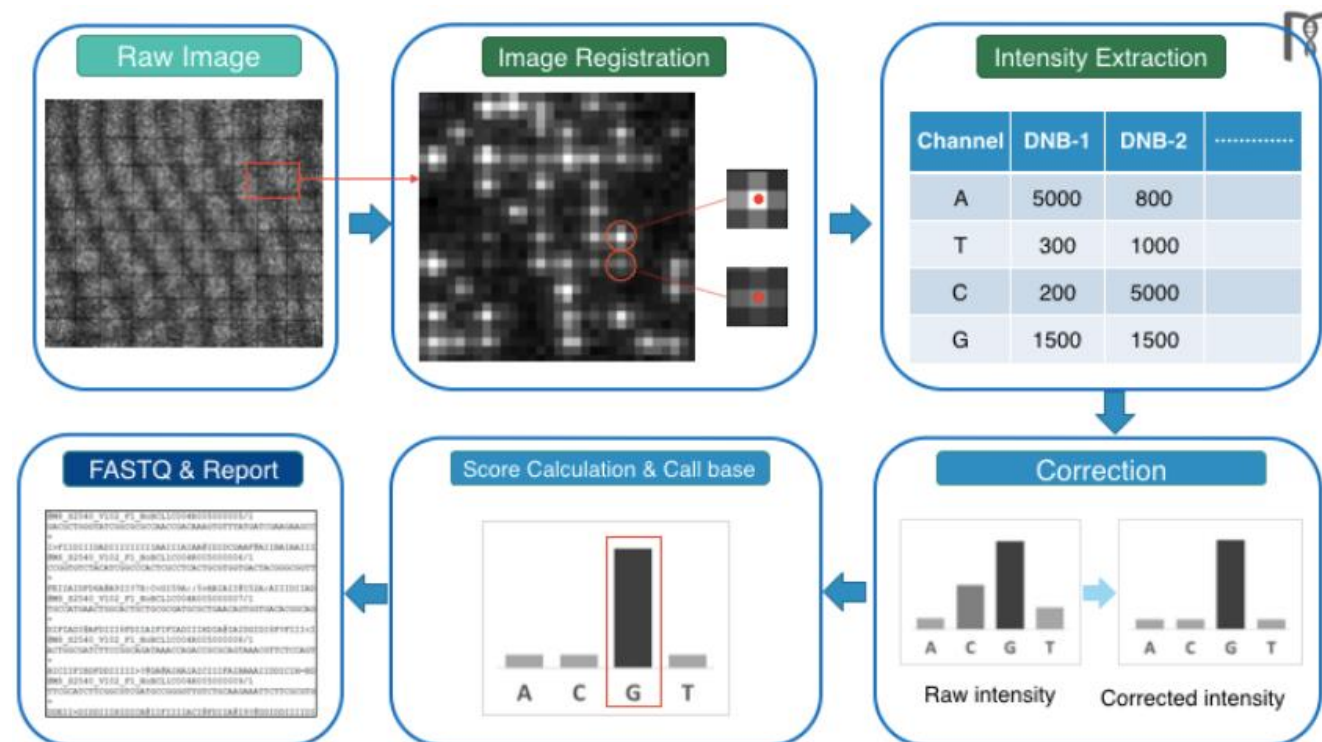


Figure 5 – Schéma de basecalling des séquenceurs MGI

<https://en.mgi-tech.com/products/> consulté le 21/06/2022



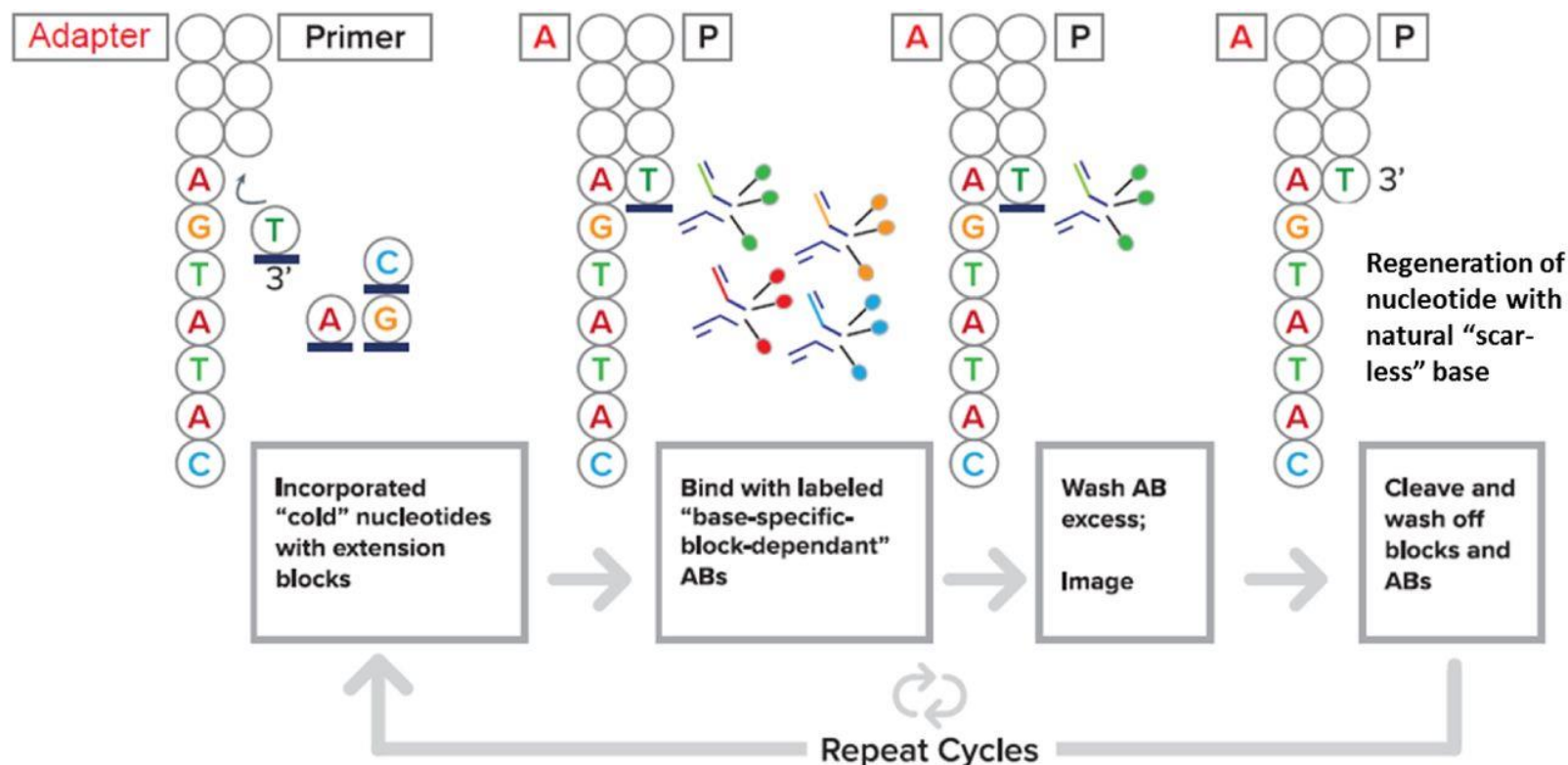


Figure 6 – Schéma de la nouvelle chimie MGI : CoolIMPS

CoolIMPS™: Advanced massively parallel sequencing using antibodies specific to each natural nucleobase. Snezana Drmanac, Matthew Callow and al. bioRxiv preprint. <https://doi.org/10.1101/2020.02.19.953307>

## Infos workflow du run et des readsets

NGL-BI-UAT **Runs** ▾ Readsets ▾ Analyses ▾ Statistiques ▾ Archives Bilans ▾ Descriptions ▾

Runs à Evaluer x

220620\_MUSHU\_V300... x

220620\_MUSHU\_V300064083

Général

Infos workflow

Etat	Date	Par
Nouveau	22/06/2022 11:44:38	ngsrg
Séquençage en cours	22/06/2022 11:44:38	ngsrg
Séquençage terminé	22/06/2022 11:44:38	ngsrg
Read generation en attente	22/06/2022 11:44:38	ngsrg
Read generation en cours	22/06/2022 11:44:38	ngsrg
Read generation terminée	22/06/2022 17:59:44	ngsrg
Evaluation en attente	22/06/2022 17:59:44	ngsrg

NGL-BI-UAT **Readsets** ▾ Analyses ▾ Statistiques ▾ Archives Bilans ▾ Descriptions ▾ Aide

Recherche de Readsets x

CTB\_DA\_AA\_1\_V30006... x

CTB\_DA\_AA\_1\_V300064083.BC103

Général

Avancé

Infos échantillon

Infos workflow

Etat	Date	Par
Nouveau	22/06/2022 15:10:05	ngsrg
Read generation en cours	22/06/2022 15:10:05	ngsrg
Read generation terminée	22/06/2022 17:59:44	ngsrg
Contrôle qualité en attente	22/06/2022 17:59:44	ngsrg

## Infos échantillon

NGL-BI-UAT **Readsets** ▾ Analyses ▾ Statistiques ▾ Archives Bilans ▾ Descriptions ▾ Aide

wamory ▾

Recherche de Readsets x

CTB\_DA\_AA\_1\_V30006... x

CTB\_DA\_AA\_1\_V300064083.BC103 Contrôle qualité en attente

Mode impression

Général

Avancé

Infos échantillon

Infos workflow

Code d'échantillon	CTB_AA	% par piste	
Ref. Collaborateur	ZymoBIOMICS_Mock_Com	Type processus banque (null)	DA - DNaseq
Taxon Id	1235509	Layout Nominal Length (pb)	-1
Taxon	synthetic metagenome	Orientation brin synthétisé (null)	undef
Type d'échantillon	ADN	Oui	META Fraction run (%)
Catégorie d'échantillon	ADN		
Code support container	V300064083		
Code container	V300064083_1		

NGS-RG Répartition des index



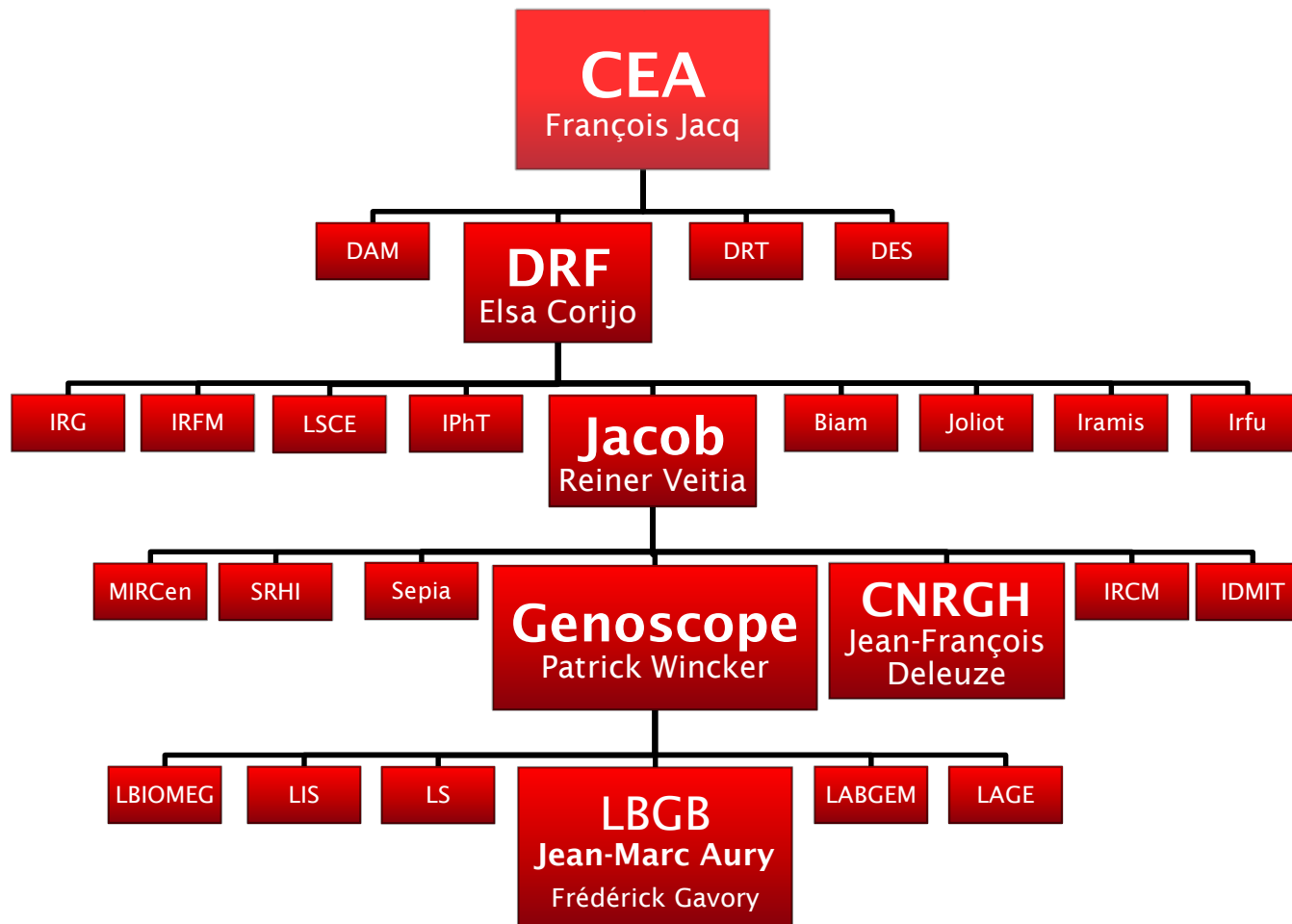


Figure 7 – Organigramme situant l'équipe du *Laboratoire de Bioinformatique pour la Génomique et la Biodiversité (LBGB)* au sein du Genoscope et du CEA (2022)

## CEA (Commissariat à l'énergie atomique et aux énergies alternatives)

- créé le 18 octobre 1945 par Charles de Gaulle
- 20 000 Salariés
- 4 directions opérationnelles et 9 directions fonctionnelles

## Genoscope (Centre National de Séquençage) Créé en 1996 - 250 salariés

- Participation au **projet Génome humain** (Séquençage du chromosome 14)
- Développement de programmes de génomiques en France
- Plus grand centre de séquençage français
- **France génomique** – unité mixte de service – regroupe les 4 principaux organismes de recherche (CEA, CNRS, TNRA, INSERM) – rassemblement de la majorité des plateformes de séquençage et de bioinformatique français
- **Projets Tara** (Pacifique – Océans – Arctique ...) - étude des écosystèmes marins
- **Projets ERGA** (European Reference Genome Atlas) – création d'une base de données de références de haute qualité des génomes d'espèces européennes