



## Plateforme de Recherche de Mutations



- **La plateforme mutation:**

- Gis Institut des maladies rares et IbiSA.



- Les instituts thématiques de l'INSERM (Génétique et développement, Neurosciences, neurologie et psychiatrie).



- L'Association Française contre les Myopathies.



- Le Genoscope - Centre national de séquençage du CEA.



- **Les intervenants techniques:**

- Capture des séquences: Laboratoire de Ressources Génomiques (LRG).

**Gabòr Gyapay**

- Séquençage: Laboratoire de Séquençage.

**Patrick Wincker**

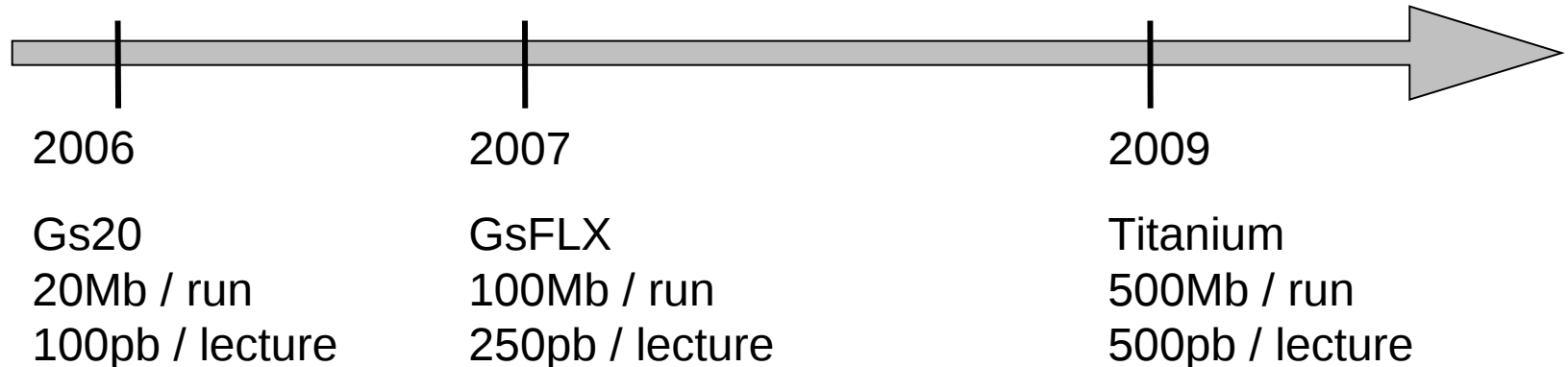
- Traitements et analyses des données: Laboratoire d'Analyse Bioinformatique des Séquences. (LABiS)

**François Artiguenave**

## → Quels types de projets?:

- Maladies génétiques rares (dermatologie, neurologie, etc...)
- Cancérologie
- Autres thématiques venant d'appels de proposition du Génoscope.  
(Analyse du génome humain et d'autres mammifères, etc...)

- Des séquenceurs en constante évolution.



- Taille minimale de la région à séquencer → > 500kb.
- Sur 1 processeur l'étape de mapping ~ 1h/run → Cluster de 100 processeurs.
- Données générées: 500Mo/run → Stockage de l'ordre de 20 To.

### → Les limites rencontrées?

- L'accompagnement des collaborateurs au cours de l'interprétation biologique.
- Nécessite encore le développement de nouvelles méthodes d'analyse.

Conseil et expertise.  
Évaluation des projets.



1



- Appels de proposition « Maladies rares » à « Guichet ouvert ».

- Conseil et expertise de la plateforme:

➔ Proposer la stratégie la plus adaptée aux porteurs de projet (redéfinition des régions d'intérêt, multiplexage, paired-end, ...)

- Evaluation du coût:

➔ De nombreux aspects à prendre en compte (Puces, Taille de la séquence, Multiplexage, ...)

- Passage devant le conseil scientifique désigné par l'INSERM.

➔ Réunion du conseil tous les 3 mois.



Conseil et expertise.

Évaluation des projets.

Préparation des échantillons.

**Validation des régions de capture.**

1mois



2

## • Préparation des échantillons:

➔ Les échantillons (ADN) sont apportés par les collaborateurs accompagnés de plusieurs contrôles qualités:

1) 21µg d'ADN à une concentration comprise entre 250 et 500ng/µl

2) Ratio  $A_{260}/A_{280} > 1.8$  et  $A_{260}/A_{230} > 1.9$

3) Contrôle sur gel d'agrose

➔ Étape de validation interne.

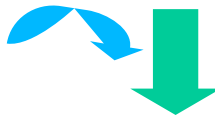
➔ Nebulisation de l'ADN en fragments de 300-1000mer

➔ Amplification par LM-PCR (3-600 x)

- **Validation des régions de capture:**

1) Délimitation des régions d'intérêt en accord avec les porteurs de projet.

17 38530280 38530657  
17 38530814 38530994



2) Soumission des régions aux fabricants de puces (NimbleGen).



3) Redéfinition des régions d'intérêt masquées par le constructeur.



4) Validation finale des régions de capture en accord avec les porteurs de projet.



Conseil et expertise.

Évaluation des projets.

Préparation des échantillons.

**Validation des régions de capture**

Capture + Séquençage.

**Étude du polymorphisme.**

1 semaine



3

- **Capture + Séquençage:**

- ➔ S'appuie sur la technologie de séquençage du Génoscope, (454 GS-FLX).

- ➔ Réalisée sur des oligonucléotides longs (385 000 oligos de 60 à 80 mer).

- ➔ Contrôles qualité de la capture.

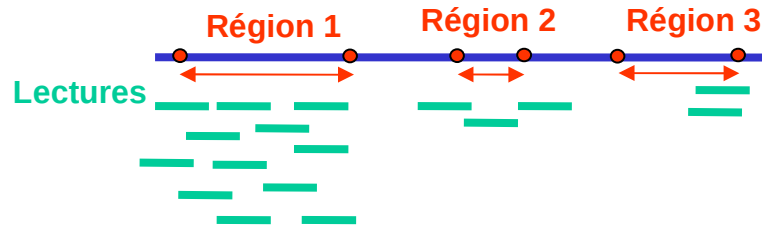
- ➔ Enrichissement de 500kb à 5Mb.

- ➔ Une profondeur moyenne comprise entre 5 et 10X.



- ➔ Mapping des lectures sur le Genome Humain (build 36).

Séquence référence



- ➔ Identification des différences avec la séquence de référence...

Séquence référence

Lectures

```

AAATGATGCATTACTACTACGCTTGG
AA-TGATGCATGACTACTACGCTAGG
---TGCATGACTA-TCCGCTGGG
AATTGATGCATGACTACTCCGCTGGG
AA-TGATGCATGACTACTC--CTGGG
  
```

- ➔ Isoler les variations en les confrontant aux données connues de dbSNP et HAPMAP.
- ➔ Contextualisation génomique, annotation fonctionnelle, ...



Évaluation des projets.

Préparation des échantillons.

**Validation des régions de capture.**

Séquençage.

**Étude du polymorphisme.**

4

3 semaines

**Mise à disposition et visualisation des données.**

**Accompagnement au cours de l'interprétation des données.**

**Stockage des données.**



## • Mise à disposition:

Mise à disposition d'un emplacement privé et propre à chaque projet consultable par internet, qui donnera accès à diverses informations :

- 1) Les séquences produites, les valeurs qualités, flowgram.
- 2) Sensibilité et spécificité de la capture par individu.
- 3) Liste des variations associées à leur score de significativité (p-valeur).
- 4) Interface web permettant de visualiser l'ensemble des régions étudiées.

- **Assistance des collaborateurs:**

- ➔ Mise à disposition d'un tutorial en ligne (en développement)

- ➔ Accueil des collaborateurs pour:

- 1) La présentation des données et formation aux outils d'interprétation.

- 2) L'explication des résultats.

- 3) Un développement spécifique aux thématiques abordées.

# Stockage.

## • Stockage:

- Données conservées jusqu'à leur publication.
- Seront ensuite déposées dans les banques de données (EBI, NCBI)

The screenshot shows the EMBL-EBI website interface. At the top, there is a search bar with the text "All Databases" and "Enter Text Here". Below the search bar is a navigation menu with categories like "Databases", "Tools", "EBI Groups", "Training", "Industry", "About Us", and "Help". The main content area is divided into several sections:

- Data Resources & Tools:** A grid of links including EMBL-BANK, UniProt, ArrayExpress, Ensembl, InterPro, PDB, Genomes, Nucleotide Sequences, Protein Sequences, Macromolecular Structures, Small Molecules, Gene Expression, Molecular Interactions, Reactions & Pathways, Protein Families, Enzymes, Literature, Taxonomy, Ontologies, Sequence Similarity & Analysis, Pattern & Motif Searches, Structure Analysis, Text Mining, Downloads, and Web Services.
- European Bioinformatics Institute:** A central banner with a building image and the text "European Bioinformatics Institute".
- About the EBI:** Links for Research, PhD Studies, Training, Industry Support, Group & Team Leaders, and EBI Funders.
- EBI hosted EU Project Websites:** Links for BioSapiens, E-MAP, ELDIR, EMBRACE, EMERALD, ENFIN, FELICS, and SYMBIOmatics.
- Training:** Information about hands-on training at EBI, including dates (16-18 March 2009) and a registration deadline (Fri 13 Feb 2009).
- Research Highlights:** A section titled "How bacteria survive environmental change" dated 03 Nov 2008, mentioning an article in *Genome Research* about metabolic protein production regulation in the model bacterium *E. coli*.
- Latest News:** A section titled "Re-write the textbooks: transcription is bidirectional" dated 26 Jun 2009, mentioning a genome-wide study of transcription in yeast.

<http://ebi.ac.uk>

The screenshot shows the NCBI Short Read Archive website. At the top, there is a navigation bar with "NCBI", "Site map", "All databases", "PubMed", and "Search". Below this is the "Short Read Archive" header with sub-links for "Main", "Browse", "Submit", "Download", "Documentation", "Trace Archive", "Trace Assembly", "Trace Home", and "Trace BLAST". The main content area includes:

- News, Events and Notifications:** A section titled "News, Events and Notifications" with a sub-link for "Provisional SRA Tracking".
- Announcement:** A paragraph stating "The Short Read Archive (SRA) 1.0 will be deployed gradually up to Feb 4, 2009. Various data will be loaded and new facilities made operational without any particular announcement. New capabilities may appear on the web site. These may not work fully when you attempt to use them."
- Provisional SRA:** A paragraph stating "During this period, the Provisional SRA will co-exist and deposits there will continue to be available to the public. The Provisional SRA will remain in operation until SRA 1.0 is fully operational and tested."
- Upcoming capabilities:** A list of capabilities that will soon appear, including "Run Browser", "Study/Sample/Experiment/Analysis browsers", "Download facility", "Search capability (search by RUN)", "Entrez indexing", "Interactive submissions facility", and "Automated submissions".
- Facilities:** A paragraph stating "We are delaying full product release announcement until the entire site is deployed. However, we welcome your visits and comments, and please try out a few of the new facilities."
- Accession:** A paragraph stating "As before, we are working to accession and serve short read submissions we have already received."
- Footer:** Links for "Write to the Help Desk", "Privacy Notice", "Disclaimer", and "Accessibility". It also includes the "National Center for Biotechnology Information" and "U.S. National Library of Medicine" logos, along with the "FIRST GOV" logo.

<http://dtd.nlm.nih.gov>



contact: [pfm@genoscope.cns.fr](mailto:pfm@genoscope.cns.fr)

Instituts  
thématiques



**Inserm**

Institut national  
de la santé et de la recherche médicale

