



**GeT**  
Génome et  
Transcriptome

# Toulouse Genomic Core facility

**Cécile Donnadieu**

**GeT Animator**

**GeT-PlaGe Manager**

**<http://get.genotoul.fr>**

**@GeT\_Genotoul**

# Who are we?

- **Genomics and transcriptomics core facility spread on**

- 5 sites GeT in Toulouse
- 1 site Oncopole



- **National Infrastructure within the « France Génomique » program**

- **IBISA Label**

- **INRA strategic core-facility**

- **ISO9001 et NFX0800 Certification**

- **Partnership with Genotoul Bioinfo**



# The Core facility missions

- **To provide innovating technologies for genome analysis to the scientific community**
  - Sequencing / Genotyping
  - Gene expression
- **To Develop new protocols, new methodologies, acquire expertise and train in those technologies**
- **To animate workshop for user network**



# Team and Expertise

- **A team of 30 people with:**

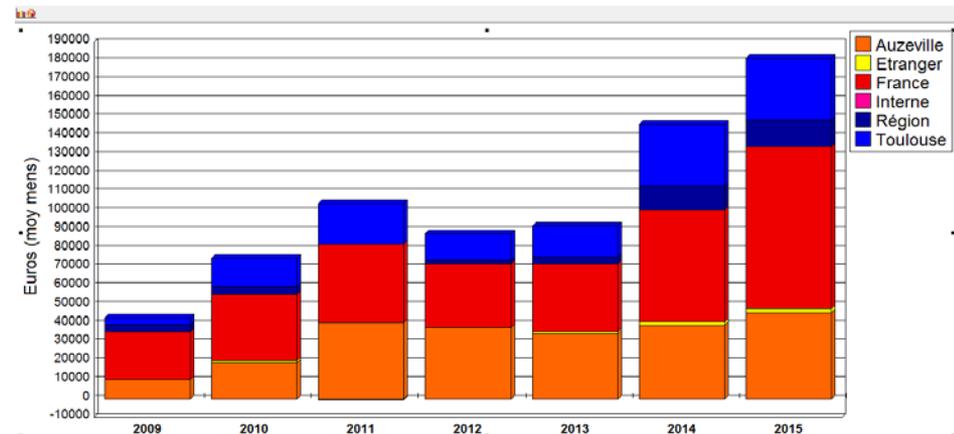
- **Competence in biology, bioinformatics, biostatistics**
- **Experts in Agronomy, Environment, Microbiology, Health**
- **Technical Specialty and scientific community by site**



**ACTIVITY**

# Activity per year

- **15 R&D projects**
  - Mate paired, chipseq, metylation, GBS, 3G NGS, capture, HIC...
- **More than 100 laboratories (INRA, CNRS, INSA, INSERM, CHU, CIRAD ...)**
  - More than 160 research teams
  - More than 250 projects
  - 2M€ of activity



# Collaboration and partnership

- **17 projects in collaboration (France – Europe)**

- **CARTOSEQ** (2010-2014) INRA : Identification en masse des variants génétiques influençant les caractères d'élevage chez les trois principales races laitières françaises
- **FUNHYMAT** (2011-2015) Université Pau: Structure et fonctionnement de tapis microbiens contaminés avec des hydrocarbures
- **Domesticick** (2013-2016) INRA: De la génomique du genre Gallus à l'histoire de la domestication du poulet
- **EFFECTOORES** (2013-2016) : Exploitation des connaissances sur les effecteurs des Oomycetes pour la recherche de résistances durables aux maladies chez les plantes cultivées
- **IMPACT** (2014-2016) INRA: Identification of Matrix Proteins Affecting CalciteTexture in chicken and guinea fowls eggshells
- **PigLetBiota** (2014-2019): Une étude de biologie intégrative de l'influence du microbiote intestinal sur la robustesse des porcelets
- **AgENCODE** (2015-2016): a French pilot project to enrich the annotation of livestock genomes
- **Treasure** (2015-2019- H2020): Treatment and Sustainable Reuse of Effluents in semiarid climates
- **Bovano** (2015-2017): IDENTIFICATION AND FUNCTIONAL STUDY OF CATTLE DELETERIOUS MUTATIONS
- **Feed-A-Gene** (2015-2020 – H2020) : projet européen pour améliorer l'efficacité alimentaire des monogastriques
- **Vaisseaux et Cancer** (2013-2016 - INCA) : caractérisation moléculaire des vaisseaux qui contribuent à l'inhibition de la croissance tumorale
- **SELGEN - GenSSeq**: développement et la mise en œuvre des méthodes à haut débit permettant d'estimer la valeur génétique des animaux et végétaux
- **Agri-Métatranscriptomique** – diversité: Nouvelles perspectives dans l'étude des communautés microbiennes complexes
- **HeliOr** (2015-2018) : Séquençage du Génome de l'Orobanche et 2<sup>ème</sup> génotype Tournesol dans le cadre de SUNRISE
- **Meta-Pac** (2015-2016): Mise au point des analyses de Metagénomique long read
- **B-TB** (2013 – 2016 ANR) Role of B cells tuberculosis immunity and inflammation.



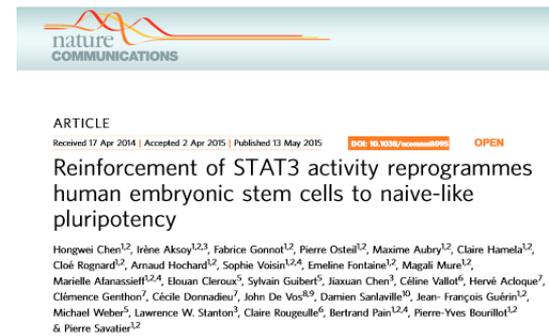
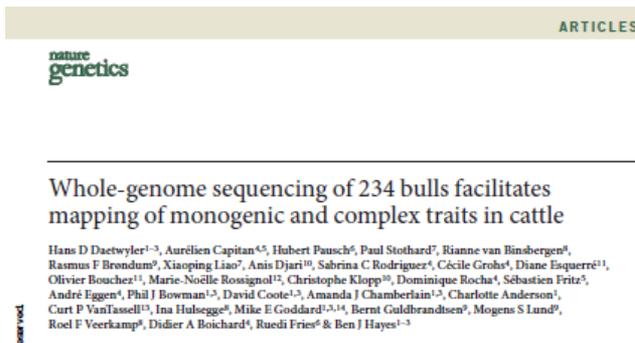
# Collaboration and partnership

- **46 publications as co-author since 2012**

2012 (14) : PLoS ONE, NAR, Gene, Current Biology, Developmental and Comparative Immunology, Gene 500, Ecotoxicology, N, Biochimieew Phytologist, FEMS, Ecotoxicology, Nucleic Acids Research, PLOS Genetics, Hepatology

2013 (13) : Biotechnology for biofuel, BMC Genomics, Mol Ecol ressources, infection genetics and evolution, PLoS pathogene, BMC, molecular ecology, genome amouncements, molecular phyloenetics and evolution, FEMS, Gastroenterology & Hepatology, Biochem Biophys Res Commun, Front. Microbio

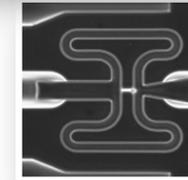
2014 (19): nature genetics, BMC Genomics, JAS, Exp and molecular pathology, Gene Nar, Molecular Biology ressources, EMBOJ, Emerging Infectious Diseases, molecular cell, PLoS ONE, PLoS Biol, Poultry science, FEMS, Biochemistry & Molecular Biology, Toxicology, Medecine, Research & Experimental, J Am Soc Nephrol, Pathologie Biologie, Journal of Biotechnology



**TECHNOLOGIES**

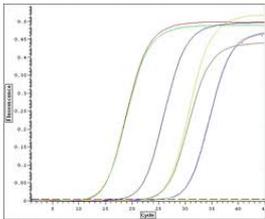
# Tools to improve the activity

- **Sample and library quality controls**
- **Single cell capture (C1 Fluidigm)**
- **Pipetting platforms for sample preparation**
  - Partnership with Tecan (4 Evo + 1 genesis), Agilent Bravo
  - Acces array (fluidigm)



# Tools to analyse gene expression and to genotype

## PCR quantitative



**Vii7, QuantStudio,  
ABI7900HT, ...**



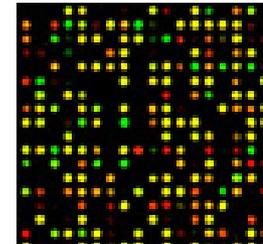
## PCR quantitative microfluidic



**BioMark (Fluidigm)**



## Microarray



**Affymetrix – Agilent  
iScan**



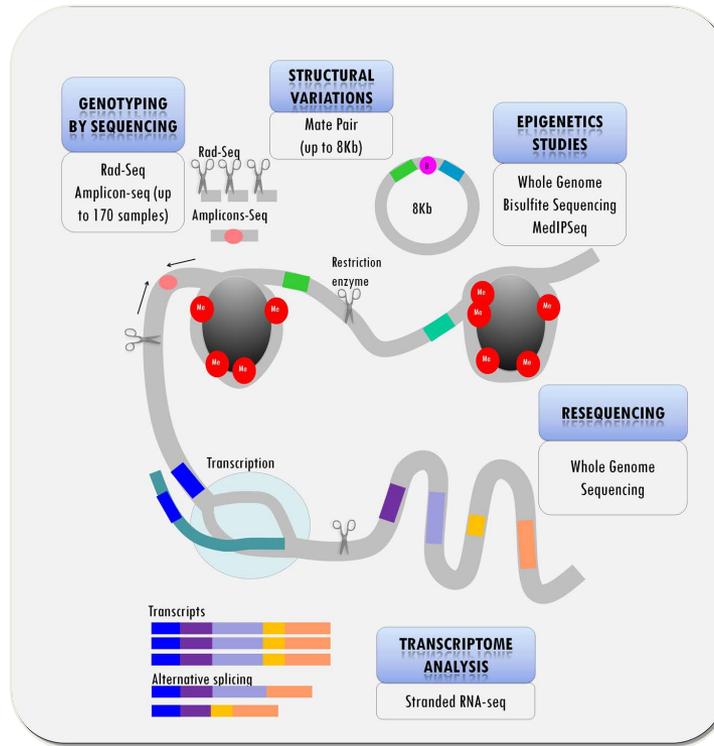
# From Sanger to short read NGS revolution



400 pb  
1Gb



200 pb  
13 Gb



2x 300 pb  
15 Gb



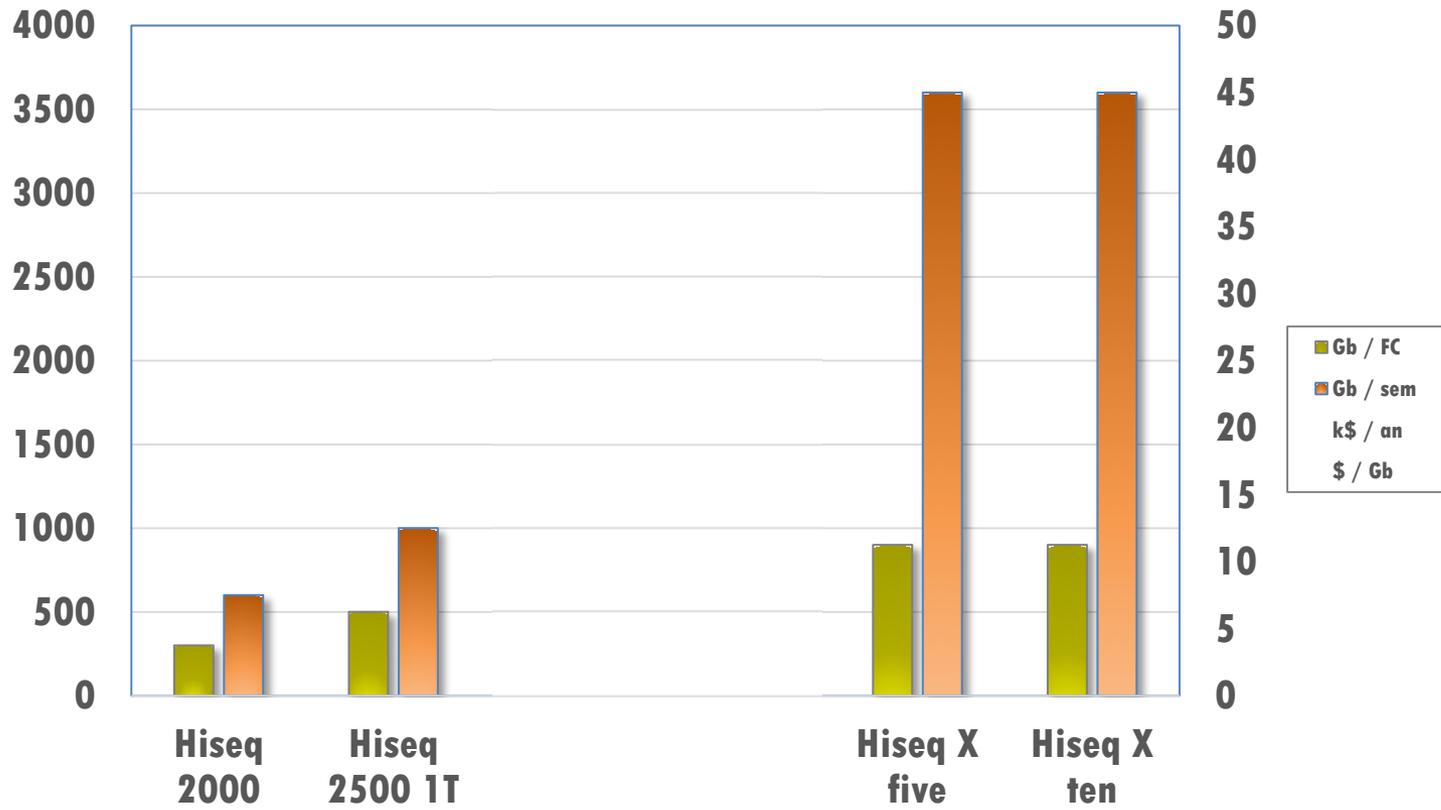
2x150 pb  
700 Gb

From 1 human genome sequencing to....

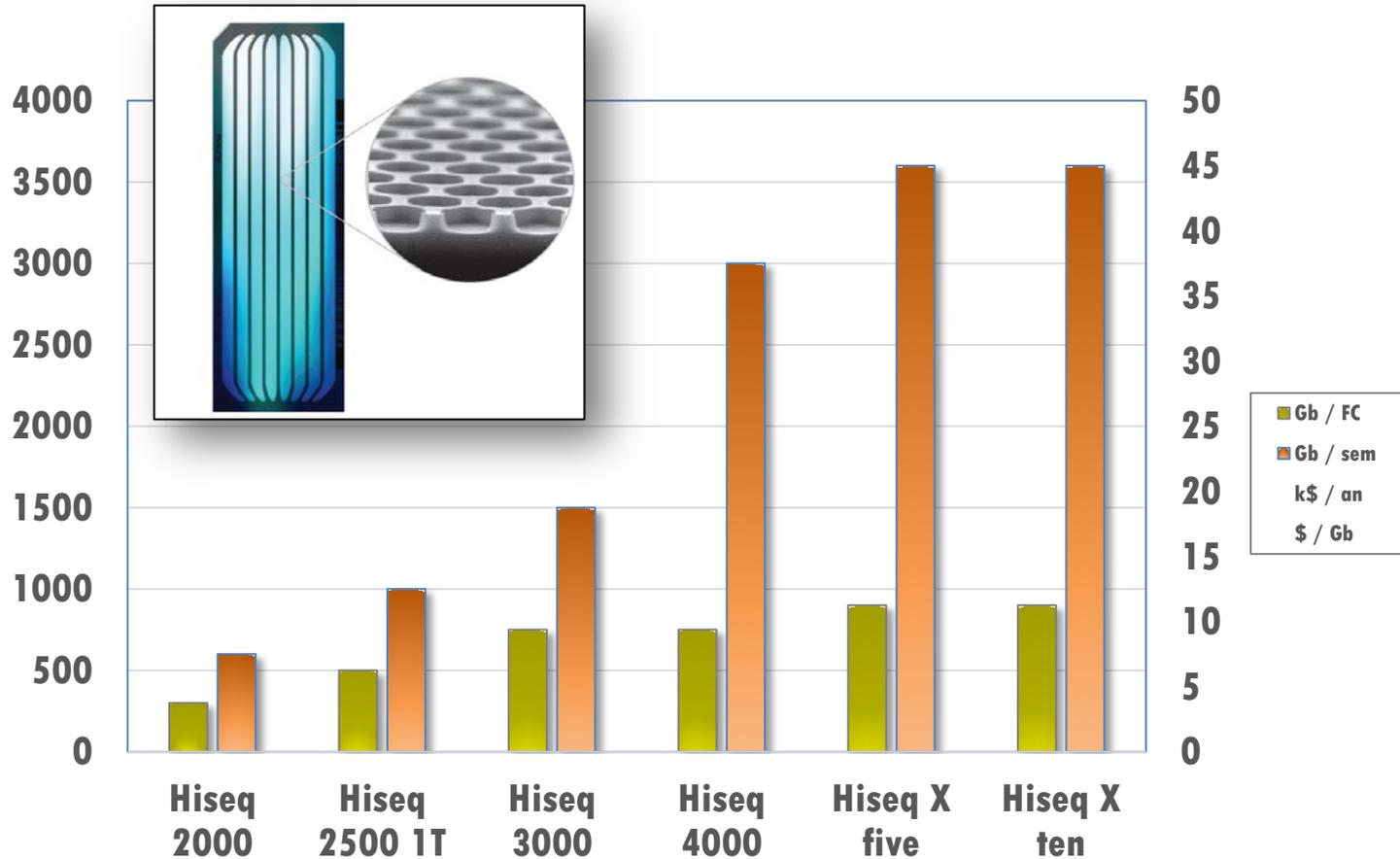
... « 1000 genomes » projects for all species

# Productivité des Hiseq & coûts

2015 : Une année riche en nouveautés



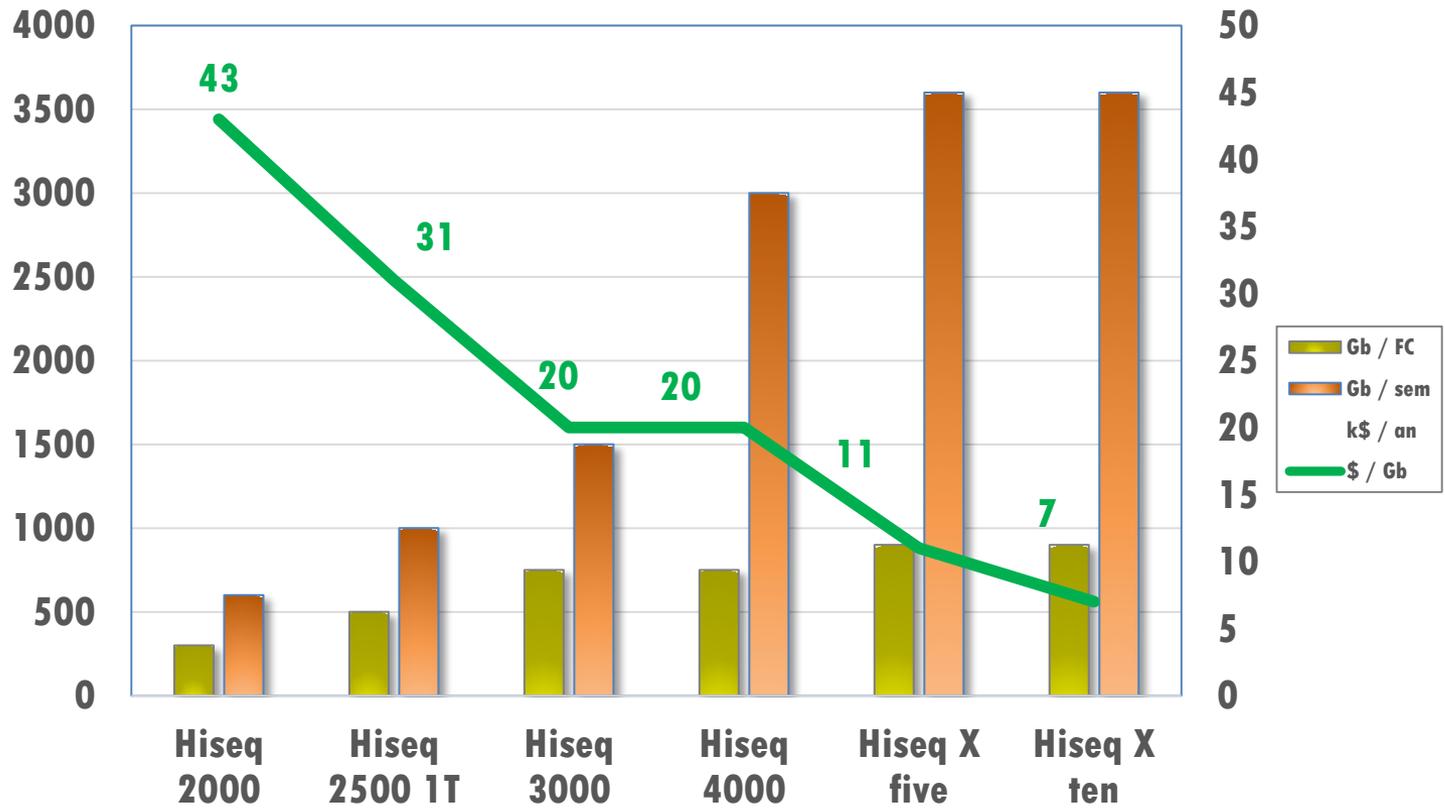
# Productivité des Hiseq & coûts



**Janvier 2015 : Sortie des HiSeq 3000 & HiSeq 4000**  
 (basé sur les Flow Cells de l'HiSeq X)

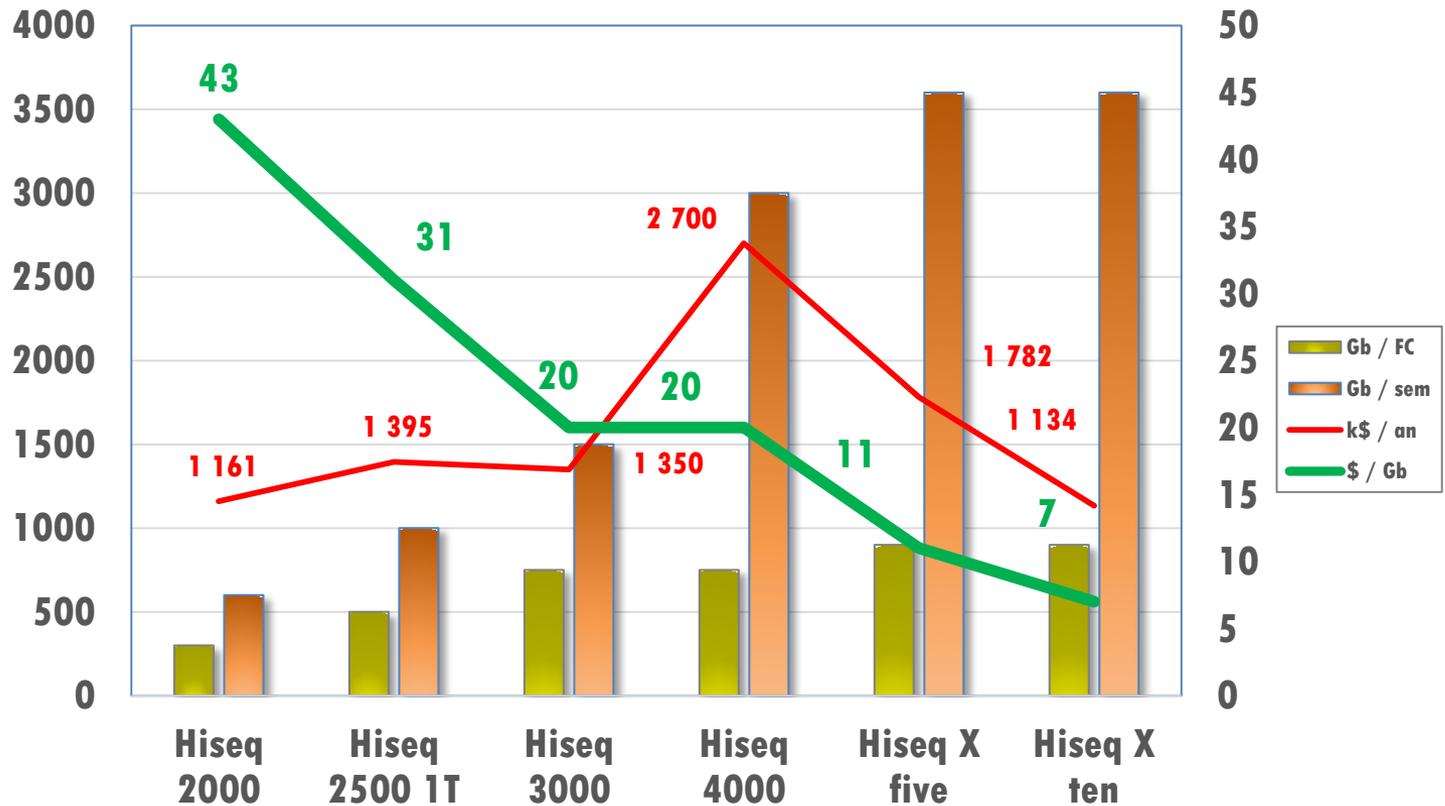
# Productivité des Hiseq & coûts

Coûts réactifs Illumina \$ / Gb



# Productivité des Hiseq & coûts

**Coûts réactifs k\$ /an pour une machine à ≈ 100 % 45 sem /an**

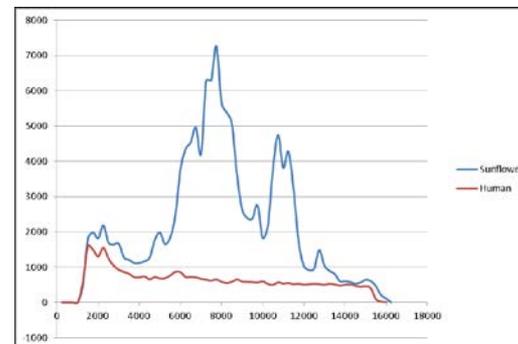


**Octobre 2015 : Ouverture des Hiseq X au non humain (en 30x)**  
 (s'achète par 5 ou plus)

- **La chimie P6/C4 n'améliore pas la qualité intrinsèque, mais permet le séquençage de plus grands fragments**
- **Achat d'un PacBio RSII à Toulouse sur crédits Région MP/FEDER, en raison d'un intérêt tout particulier pour les génomes végétaux.**



**Tournesol : 30 % de séquences répétées (LTR)**  
**Homme : 8.8 % de séquences répétées**




**SOFIPROTEOL**  
 L'engagement durable





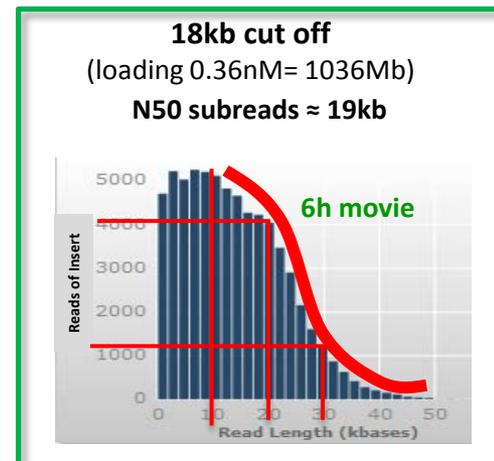
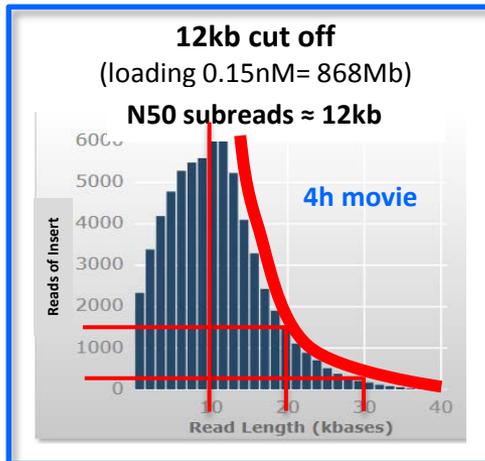
**LA DIFFICULTE : avoir suffisamment de séquences passant les grandes répétitions**



overview-of-smrt-technology.mp4

# LE FILM

# Amélioration de la longueur des lectures



**15kb cut off**

**20kb cut off**

**20kb cut off**

		NUM	MAX	N50 BP	N50 NUM	MEAN	MEDIAN	BP/SMRTcell	(b)
<b>IGM (San Diego, USA) 202 SMR cells</b>									
	moyenne	98666	45457	12211	28413	9176	9032	0,906 M	b
	max	146374	52725	12981	41602	9997	9809	1,36Gb	
<b>Lausanne University (Swiss) 59 SMRT cells</b>									vie
	moyenne	106800	46800	15172	28371	10773	9821	1,15Gb	
	max	144358	53253	16132	38325	11436	10568	1,6GB	
<b>Get-PlaGe (France) 146 SMRT cells</b>									
	moyenne	77086,6301	52317,4932	15365,4795	19705,5822	10326,6773	9152,5137	800Mb	
	max	126777	80974	20507	33133	13635	12295	1,3Gb	

Improvements of the molecular biology steps have increased the length of the Pacio Sequences (B. Mayjonade)

# Amélioration de la longueur des lectures

## Moyennes sur 147 SMRT 1<sup>er</sup> Génome Tournesol :

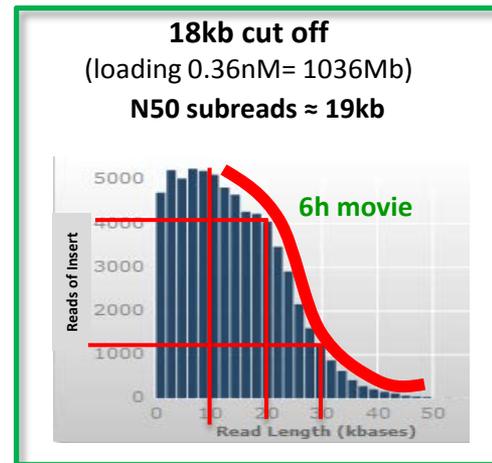
- **N50 moyen de 15365 (19705 reads > N50)**
- **800 Mb / SMRT cell**

## Moyennes sur 103 SMRT 2<sup>e</sup> Génome Tournesol :

- **N50 moyen de 18510 (20989 reads > N50)**
- **1 041 Mb / SMRT cell (max 1 445 Mb)**

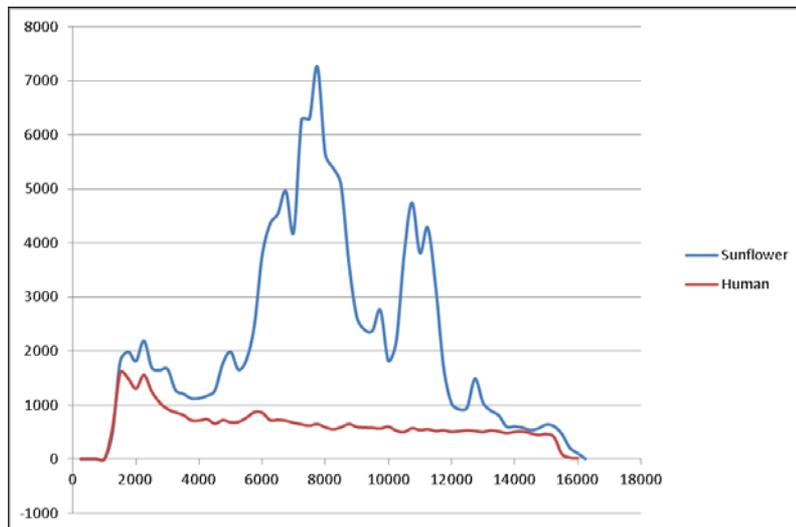
### Top 10 of our longest subreads

80974 bp  
 79860 bp  
 79834 bp  
 78105 bp  
 77481 bp  
 76881 bp  
 76558 bp  
 76355 bp  
 75569 bp  
 75559 bp



# Séquençage du Tournesol

**Tournesol : 30 % de séquences répétées (LTR)**  
**Homme : 8.8 % de séquences répétées**



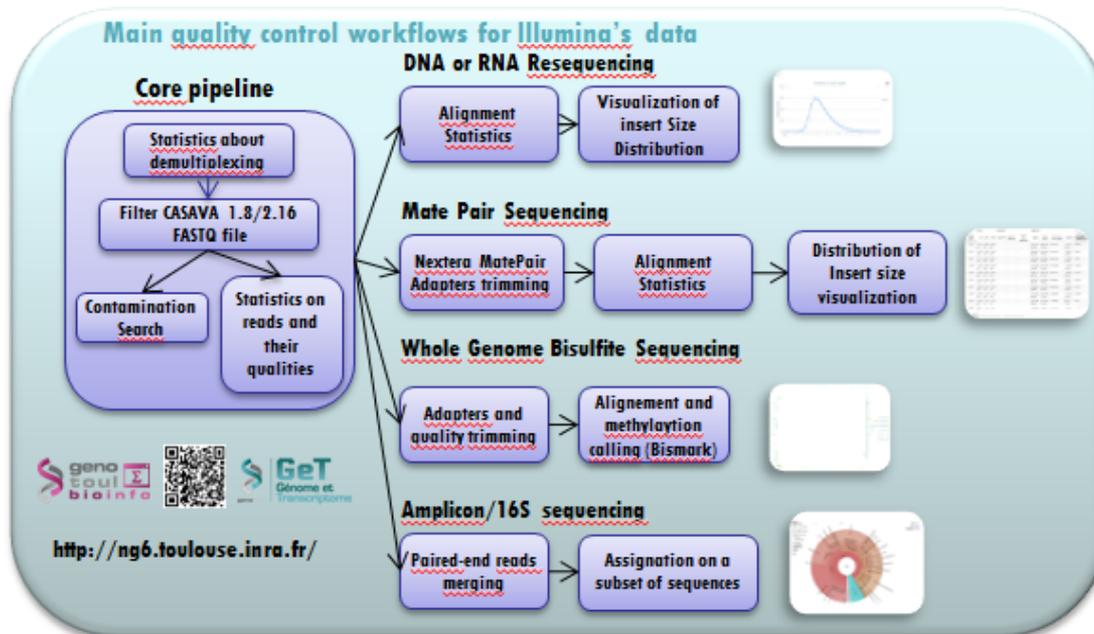
**LA DIFFICULTE : avoir suffisamment de séquences passant les grandes répétitions**

- INRA (équipe Tournesol) :**  
**HiSeq en 127 X**  
 → **43 % de couverture**
- Effort international :**  
**454, HiSeq, Carte Génétique & Physique (BAC fingerprintés)**  
 → **63 % de couverture**
- INRA (équipe Tournesol) :**  
**PacBio en 107 X (407 SMRT)**  
 → **84 % de couverture**  
**13124 contigs avec N50 = 498 kb**  
 → **+90 % des séquences ancrées**

**LIMS**

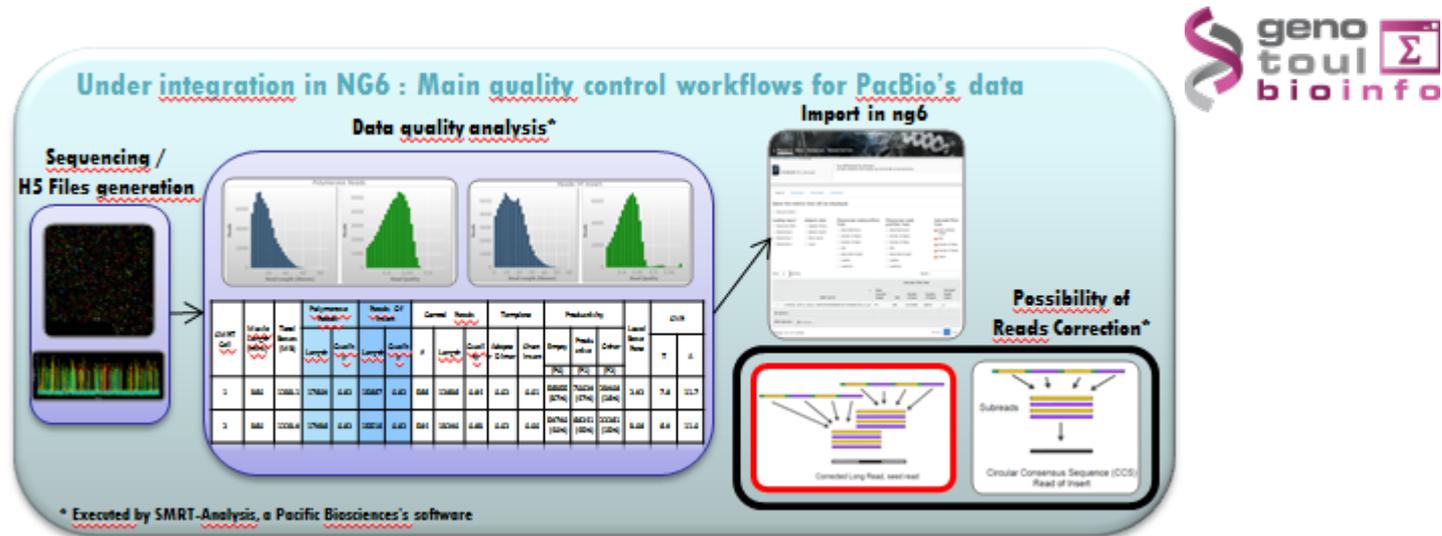
# Storage of raw reads and quality control results

- Each step of the process is **tracked** by various modules of our in-house information system.
- Biological samples are uniquely identified using **barcodes**.
- NGS pipelines are provided by **NG6**, an information system built upon the jflow workflow management system



# Storage of raw reads and quality control results

- **Current integration in NG6 : Main quality control workflows for PacBio's data**



- **Upcoming : A new LIMS for NGS samples, sequencing and analysis tracking**



**MORE DETAILS**

# L'annonce du Sequel

- L'annonce d'une sortie du Sequel :
  - Machine **2 x moins chère** et **2 x plus petite**
  - Machine **7 x plus productive** (1 M de ZMW = puits)
  - Coûts au **Gb divisé par 2-4**
- Disponibilité :
  - aux US en Q4 2015
  - En Europe en 2016
- Complémentaire du RSII ?



# THANKS

## GeT-PlaGe Team

Denis Milan  
Gerald Salin  
Olivier Bouchez  
Diane Esquerre  
Anne Fleurbe  
Sandra Fourre  
Isabelle Hochu  
Céline Vendecastel  
Dounia Ben Salah  
Claire Kuchly  
Gaëlle Vilchez  
Catherine Zanchetta  
Marie Vidal  
Pauline Heuillard  
Céline Jeziorski  
Clémence Genthon  
Maarten Pirson  
Alain Roulet  
Céline Roques  
Adeline Chaubet  
Frédéric Toppan

## GeT-BioPuce

Marie-Ange Teste  
Lidwine Trouilh  
Nathalie Marsaud  
Delphine Labourdette  
Sophie Lamarre  
Matthieu Guionnet

## GeT-TRIX Team

Yannick Lippi  
Claire Naylies

## GeT-TQ Team

Jean-José Maoret  
Frédéric Martins

## GeT-Purpan Team

Nicolas Borot  
Nathalie Jonca  
Emeline Lhuillier  
Marion Roy

