

Comité GeT

5 novembre 2010

Cécile Donnadiou
PlaGe

Plateforme Génome et transcriptome
<http://get.genotoul.fr>



Plateforme GeT : 5 sites sur Toulouse



Présentation de la PlaGe

- Plateforme en place dès **2000** (création de la Génopole)
- Plateforme reconnue par **IBiSA 2008**
- Plateforme **stratégique INRA**, soutenue par le **CPER**
- Plateforme certifiée **ISO 9001:2000** en octobre **2008**



- **Bâtiment 250 m² en 2008, qui sera étendu à 500 m² en 2011**
- **Un partenariat avec la plateforme Bioinformatique**

- **Activités :**

- **Extraction d'ADN**
- **PCR quantitative**
- **Génotypage**
- **Séquençage**



Extraction & stockage d'ADN + traçabilité

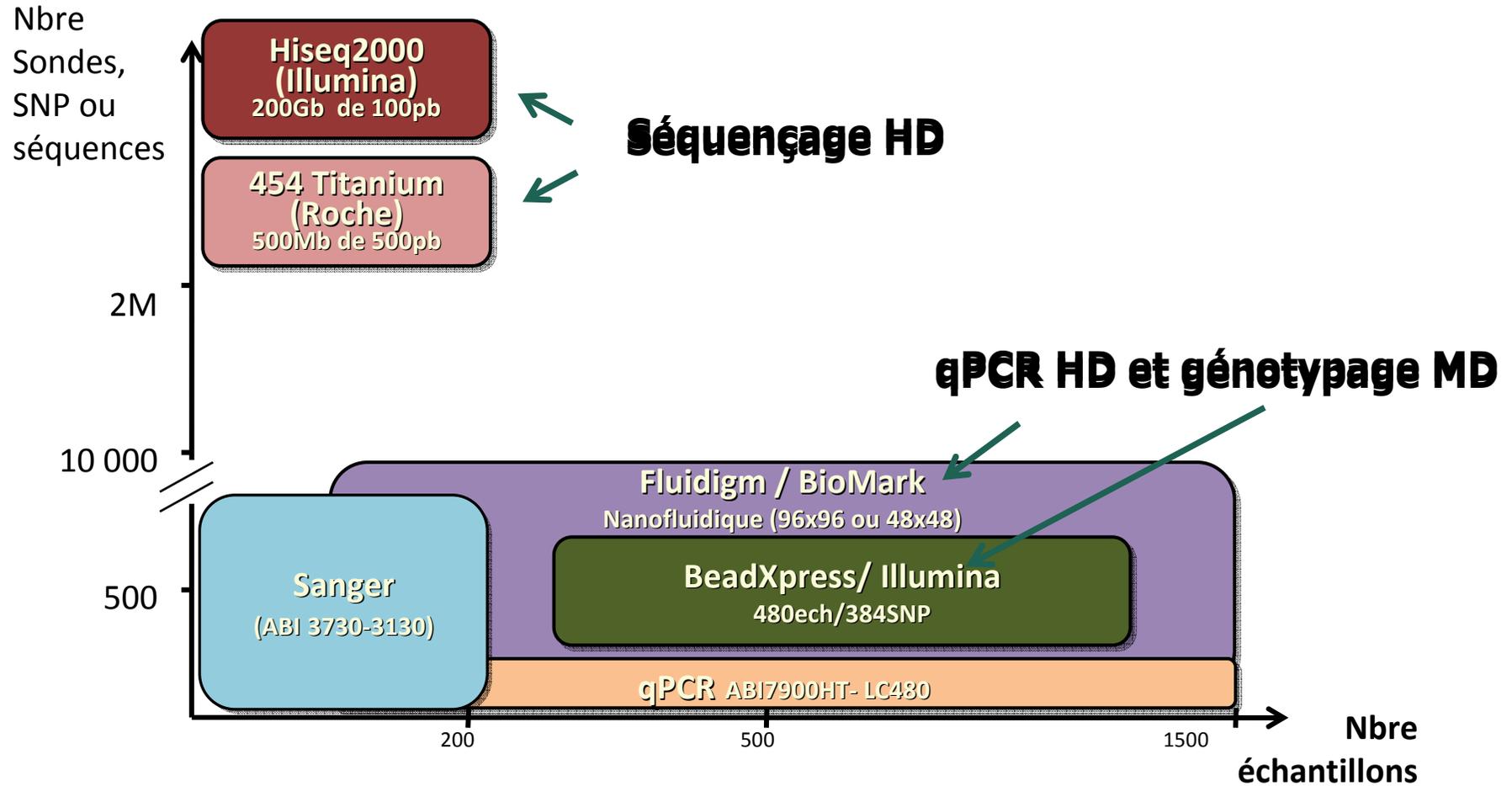


Matériel :

- Extracteur d'ADN de qualité Genomic **ExtraGene**, 72 échantillons/24h
- Extractions rapides en plaques sur TECAN
- **Congélateurs** -20°C avec suivi des températures
- Imprimantes et Lecteurs de **codes-barres** 1 D et 2 D

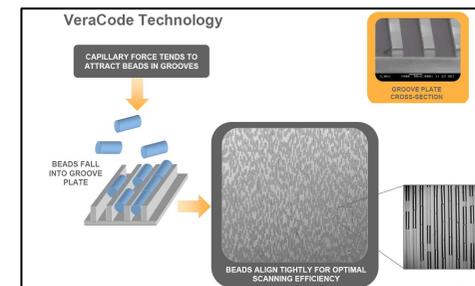
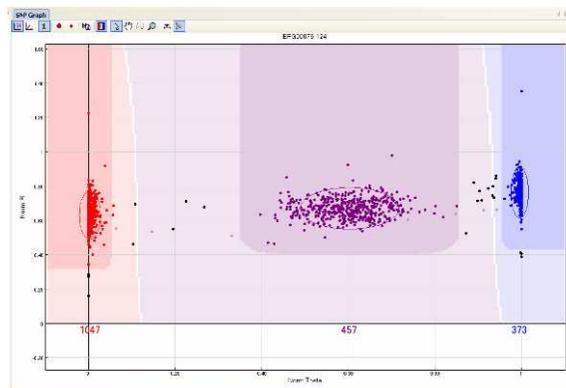


Technologies disponibles



BeadXPress – Génotypage SNP

- Depuis 2008, mise en place du **BeadXPress** :
 - Génotypage de jeux de **96** ou **384 SNP** sur des jeux de **480 individus** (0.03 à 0.10 € / génotype)
 - Technique très **robuste** basée sur PCR allèle spécifique et ligation



Biomark de Fluidigm : PCR Quant & SNP

- **Spécifications :**
 - PCR en **nanovolume** (7 nl).
 - Chargement automatisé jusqu'à **96 x 96**
 - **Versatilité** des jeux définis lors du chargement.
 - Sondes **TaqMan** ou **Evagreen** pour la qPCR
 - Sondes **TaqMan** ou **Kaspar** pour les SNP
- **Avantages :**
 - **Réduction des coûts** de réactifs, du plastique et du travail
 - **Grande productivité**

 **BIOMARK**
GENETIC ANALYSIS BY FLUIDIGM

The BioMark System

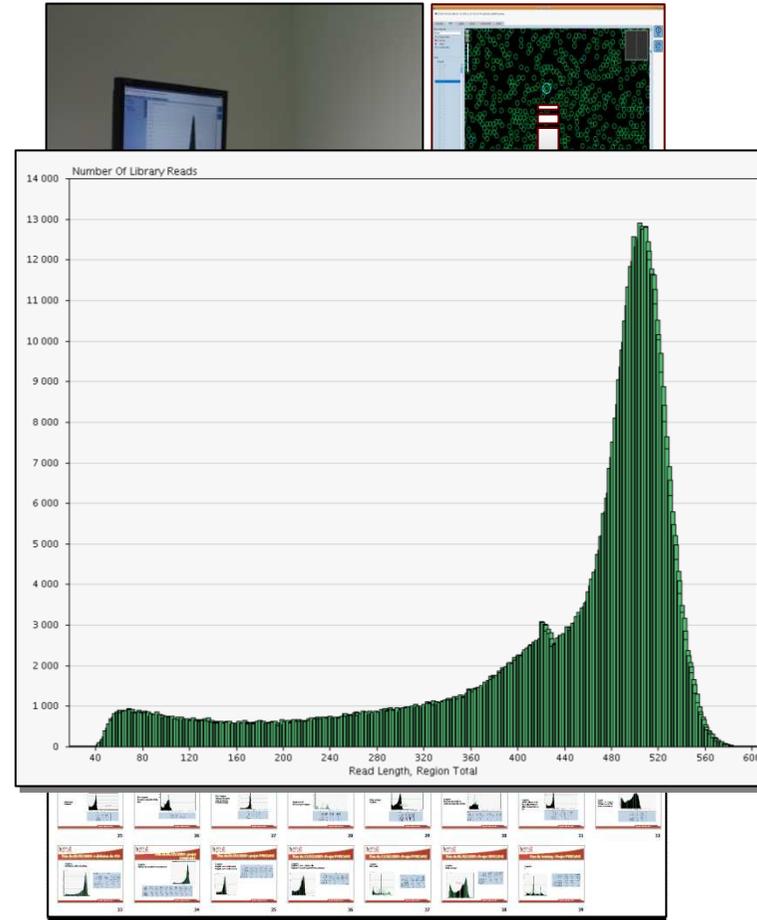
- BioMark System
- IFC Controller HX/MX
- Integrated Fluidic Circuits (IFCs)



For Research Use Only

Séquençage HD : Roche GS FLX

- **Spécifications :**
 - **1.2 Million** de séquences/run
 - Longueur moyenne des séquences : **450 pb**
 - **500 Mb**/run (1/6^e génome mammifère)
 - Lames 2, 4, 8 et 16 régions
 - Possibilité de réaliser des **multiplexages**
- **Applications :**
 - Séquençage **de novo**
 - **Reséquençage** de régions d'intérêt (par LR-PCR ou capture de séquences Nimblegen)
 - **Transcriptomique**
 - Analyses **Epigénétiques** (Etudes de Méthylation)
 - Analyses **Métagénomiques**
- **Expérience acquise :**
 - **60** runs (dont 5 communs) pour **30** laboratoires



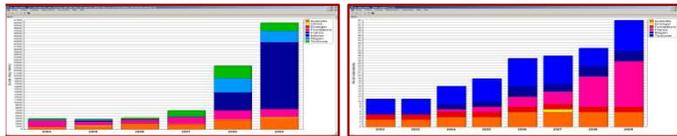
Séquençage THD : Hiseq 2000 Illumina

- **Spécifications :**
 - **500 Millions** de séquences/run (x2 lames)
 - Longueur moyenne des séquences : **100 pb**
 - **100 Gb à 200 Gb**/run (15-30 x génome mammifère)
 - Run de **4 jours** (1 extrémité) ou **8 jours** (2 extrémités)
 - Lames **8 régions**
 - Possibilité de réaliser des **multiplexages (x12)**
- **Applications :**
 - séquençage **de novo**
 - **re-séquençage** génomes entiers/régions candidates
 - analyses **épigénétiques**
 - **ChiP-seq**
 - analyses **transcriptomiques**
 - identification et quantification de **petits ARN**
- **Expérience acquise :**
 - En cours de validation



**3ème machine
en France**

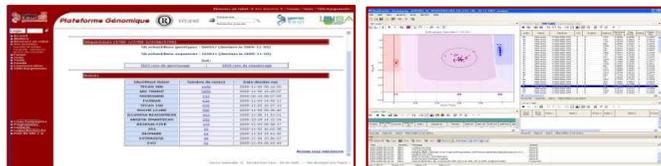
Activité



Portail web



Interface de gestion de la base de données



Récupération des données



Traitement des données

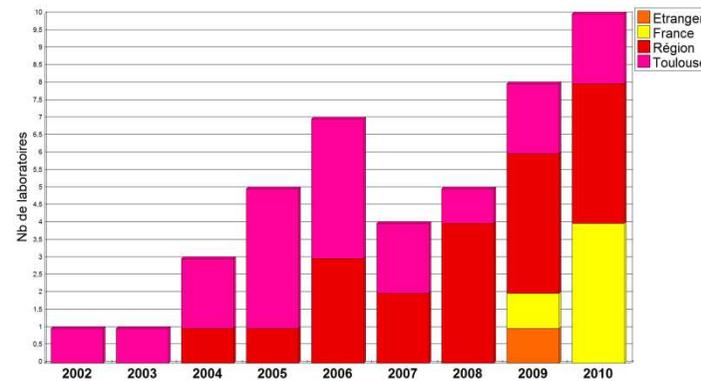
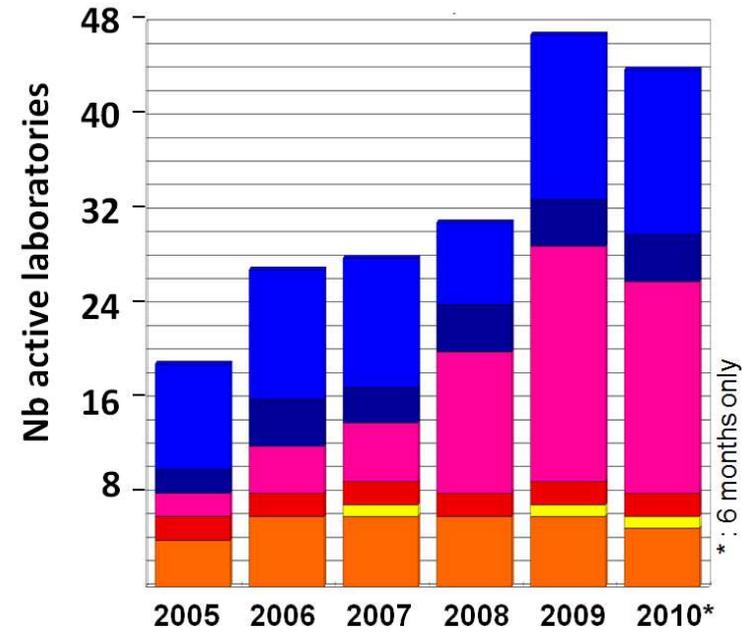
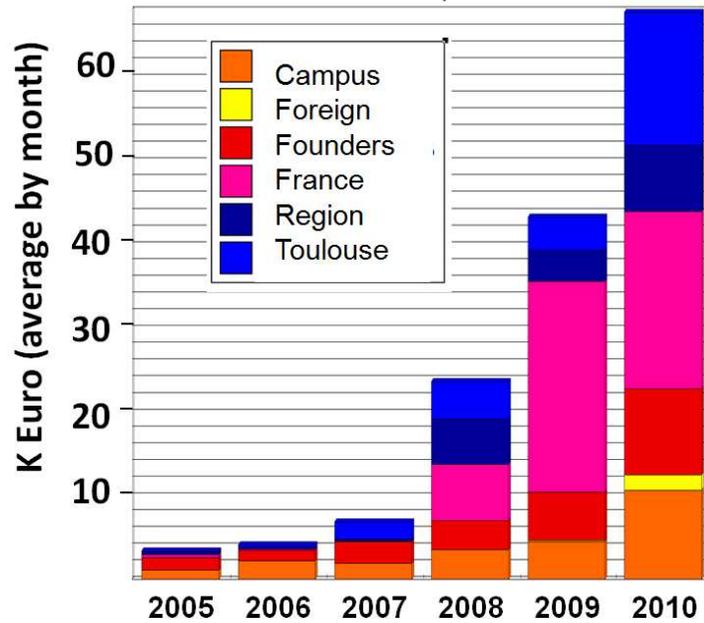


Authentification de l'utilisateur et chargement des données



Traçabilité des objets par code-barres

Une activité en croissance régulière



La suite...

- La mise en place de l'HISEQ
- De nouveaux outils - CPER 2007-2012
- Un bâtiment de 500m² à l'horizon 2013
- Le développement de notre expertise NGS
 - partenariat avec la plateforme Bioinformatique
 - développement de la préparation des librairies
 - développement des technologies d'enrichissement (exome capture, RainDance, Access array)
- Le dépôt d'un projet Grand emprunt « infrastructures distribuées » – InDiGen
- Le dépôt de plusieurs projets Grand Emprunt dans lesquels la plateforme est impliquée (ex: PHENEX)
 - Continuer la Veille Machine de 3ème Génération en 2012?

L'équipe de la PlaGe

Denis MILAN (INRA) Directeur
Cécile DONNADIEU (INRA) Directrice adjointe
Gérald SALIN (INRA) Resp Informatique
Olivier BOUCHEZ (INRA) Resp NGS
Marion DANEYROLE (INRA) Gestionnaire



Diane ESQUERRE (INRA) Projets génétique animale

Julien FREMEZ (CDD) Informaticien
Eugénie ROBE (CDD) Biologiste
Jérôme LLUCH (CDD) Biologiste
Frédéric MARTINS (CDD) Biologiste
Sophie RUZAFKA (CDD) Biologiste

Ils vont nous rejoindre en 2011

Nathalie Marsaud (BioPuce) NGS transcriptome
Jean-José Maoret (INSERM) Projets INSERM QPCR



Séquençage THD : Réalisation et analyse



Séquençage réalisé par la
Plateforme Génomique avec
l'équipe de recherche



Transfert des données et analyses
qualitatives sur la Plateforme
Bioinformatique



Analyse des séquences par la
Plateforme Bioinformatique

Collaboration Equipe de
Recherche/Plateforme
Bioinfo

Pipeline de traitement des données

PROJETS RUNS recherche OK

Projets > TESTBACS > CS_17311 > filterAssembly

Analyse filterAssembly : Filtre les contigs assemblés

Résumé du nettoyage :

Taille finale de l'assemblage : 13 contigs avec une taille de 144089 pb.

Liste des contigs retenus dans l'assemblage final :

- ⊕ contig00020 : profondeur de 16.0(longueur du contig : 22998 pb, somme des longueurs des lectures : 387958 pb avec 1219 de lectures assemblées.)
- ⊕ contig00017 : profondeur de 31.0(longueur du contig : 752 pb, somme des longueurs des lectures : 23607 pb avec 68 de lectures assemblées.)
- ⊕ contig00018 : profondeur de 18.0(longueur du contig : 6010 pb, somme des longueurs des lectures : 110261 pb avec 344 de lectures assemblées.)
- ⊕ contig00019 : profondeur de 54.0(longueur du contig : 801 pb, somme des longueurs des lectures : 43908 pb avec 123 de lectures assemblées.)
- ⊕ contig00007 : profondeur de 12.0(longueur du contig : 519 pb, somme des longueurs des lectures : 6734 pb avec 21 de lectures assemblées.)
- ⊕ contig00006 : profondeur de 50.0(longueur du contig : 1944 pb, somme des longueurs des lectures : 98825 pb avec 279 de lectures assemblées.)
- ⊕ contig00005 : profondeur de 37.0(longueur du contig : 5973 pb, somme des longueurs des lectures : 226384 pb avec 649 de lectures assemblées.)
- ⊕ contig00004 : profondeur de 17.0(longueur du contig : 35871 pb, somme des longueurs des lectures : 612024 pb avec 1855 de lectures assemblées.)
- ⊕ **contig00014(*)** : profondeur de 18.0(longueur du contig : 22863 pb, somme des longueurs des lectures : 416862 pb avec 1282 de lectures assemblées.)
- ⊕ contig00022 : profondeur de 17.0(longueur du contig : 12652 pb, somme des longueurs des lectures : 216883 pb avec 658 de lectures assemblées.)
- ⊕ **contig00015(*)** : profondeur de 17.0(longueur du contig : 11838 pb, somme des longueurs des lectures : 206065 pb avec 624 de lectures assemblées.)
- ⊕ contig00439 : profondeur de 8.0(longueur du contig : 676 pb, somme des longueurs des lectures : 5615 pb avec 22 de lectures assemblées.)
- ⊕ contig00016 : profondeur de 17.0(longueur du contig : 21192 pb, somme des longueurs des lectures : 377659 pb avec 1142 de lectures assemblées.)

(*) Contigs d'extrémités.

Fichiers fasta & qual résultats : [validated_contigs.fasta](#), [validated_contigs.qual](#),

Données en entrée issue de l'assemblage :

Données en entrée :

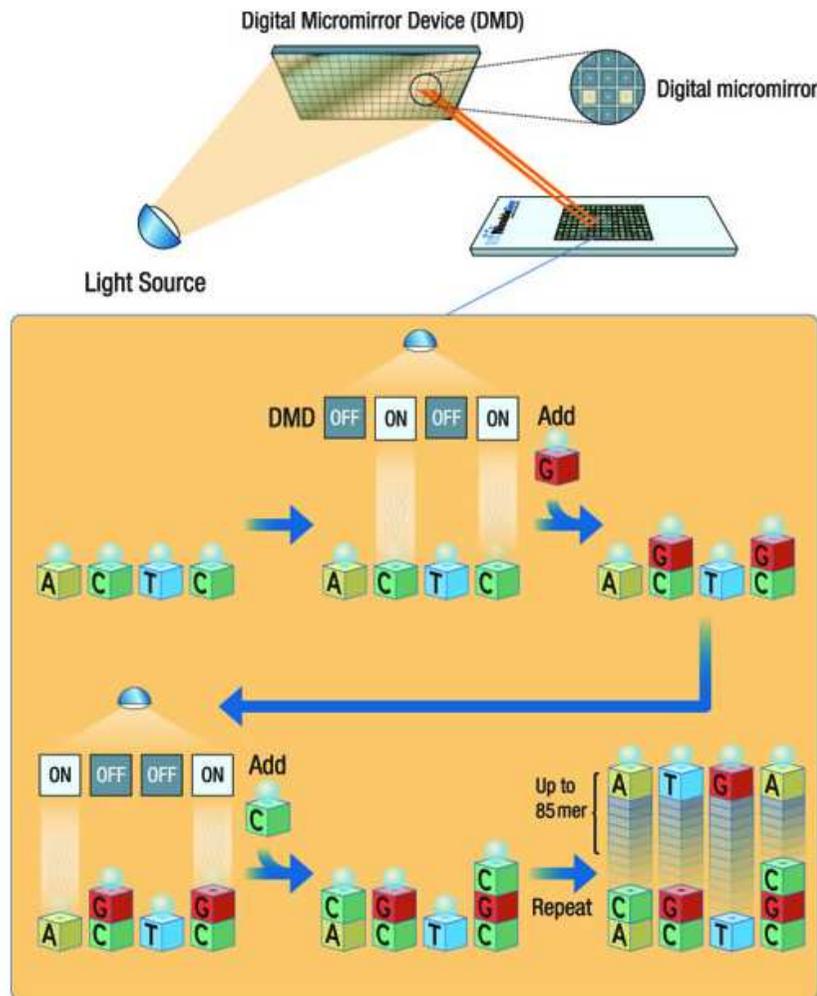
- ⊕ file : 454Reads.MID3.clean.sff
- ⊕ 14622, 14622 lectures (*)
- ⊕ 4614107, 4608534 bases (*)

* : premier chiffre correspond au nombre de lectures/bases présentes dans le fichier sff et le second chiffre correspond au nombre de lectures/bases utilisées après trimming des primers, de la qualité... (cf doc Roche: [GS_FLX_Software_Manual.pdf](#))

Informations générales sur l'assemblage :

- ⊕ 10763, 73.61% lectures alignées
- ⊕ 3228231, 70.05% bases alignées
- ⊕ 1.00%, 32274 pourcentage du nombre total de différences
- ⊕ 9263 lectures assemblées
- ⊕ 1500 lectures partiellement assemblées
- ⊕ 3853 singletons

Microarray Nimblegen



● = photolabile protecting group



1-plex
1 x 385 K

4-plex
4 x 72 K

1-plex
1 x 2.1 M
3-plex
3 x 720 K

12-plex
12 x 135 K

Microarray Agilent

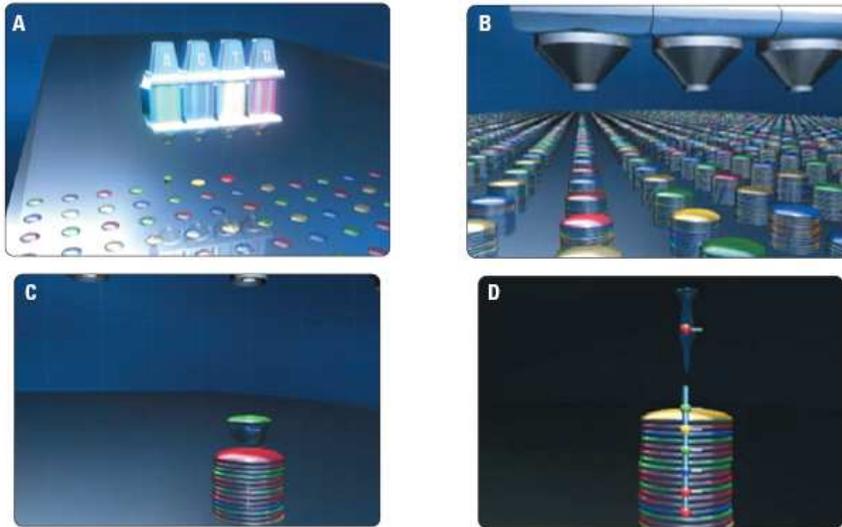
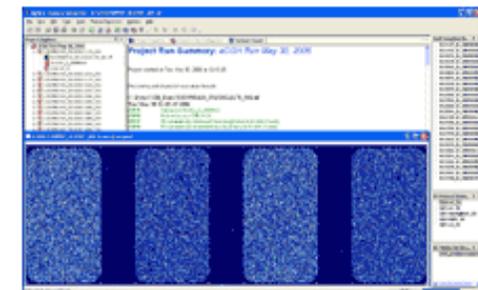
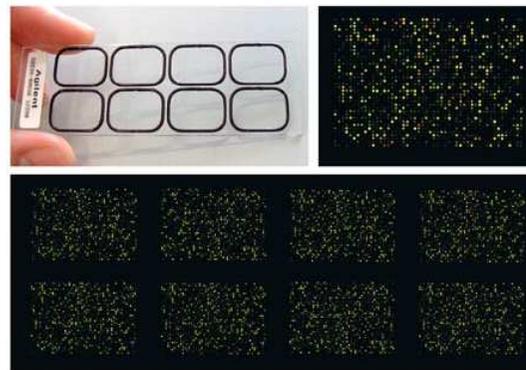
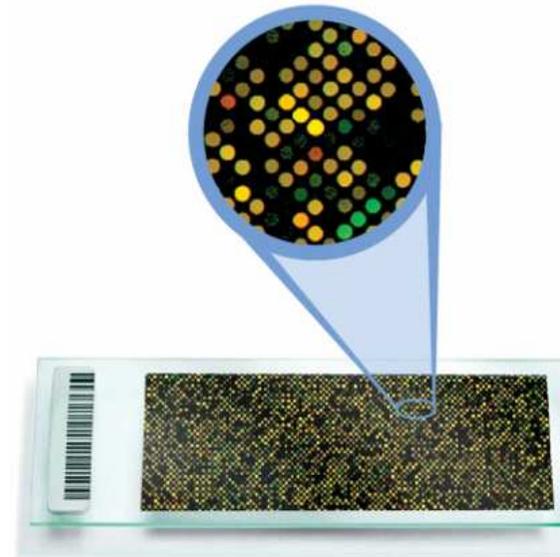


Figure 3. These four images communicate the general mechanism for oligo synthesis via inkjet printing. **A** shows the first layer of nucleotides being deposited on the activated microarray surface. **B** shows the growth of the oligos after multiple layers of nucleotides have been precisely printed. **C** is a close-up of one oligo as a new base is being added to the chain, which is shown in figure **D**.

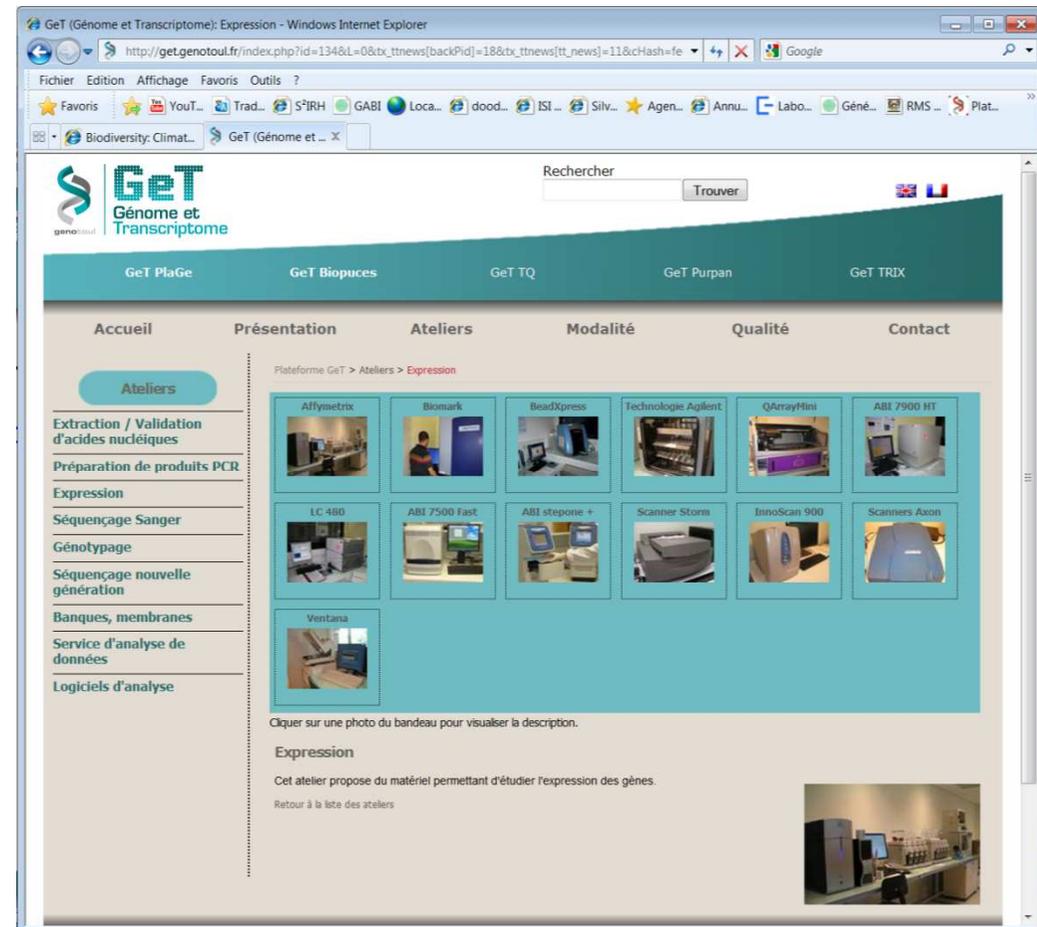


Feature extraction

Domaines d'application des microarray

- **Expression (mRNA - miRNA)**
- **Génotypage (CGH - hybridation génomique comparative ou CNV et SNP)**
- **ChIP on chip (Regulation de l'expression des gènes: épigénétique)**
- **Capture de séquence**

- Une adresse commune : get@genotoul.fr
- <http://get.genotoul.fr>



Modalités d'accès

- **Projets en collaboration (ANR, Région, Europe...)**
- **Prestation de service:**
 - **Conseils** scientifiques et techniques pour la **réalisation** de vos projets
 - Mise en place des projets – design expérimental
- **Mise à disposition de machines:**
 - Formation d'un **responsable technique** par équipe ou laboratoire
 - **Mise à disposition** de matériels opérationnels
- **Séquençage commun :**
 - Réalisation des réactions de séquence (labos publics)
 - Séquençage commun sur le **Roche 454 GS FLX**