

Génoptole Toulouse



**Genotoul**  
GÉNOPOLE TOULOUSE

# FAITS MARQUANTS 2015 - 2016

Un réseau de compétences et de plateformes de recherche  
en sciences du vivant

Mesdames, Messieurs, Chers Collègues,

L'année 2016 a été une année de transition pour Génotoul, cette transition se poursuivra en 2017 pour mener à **des évolutions** que je souhaite effectives dès l'été.

**Transition** marquée principalement par le départ de Nic Lindley (Directeur du GIS) vers des cieux où le soleil se lève plus tôt. Nic Lindley a su prendre le relai de Claude Chevalet et maintenir une cohésion Génopole. Je tiens au nom de tous à remercier Nic pour le travail effectué pendant les 4 dernières années, travail constant pour maintenir et amplifier la reconnaissance et l'apport du réseau mis en place auprès de l'ensemble de la communauté Midi-Pyrénées.

Ce départ a induit une période d'environ 6 mois au cours de laquelle trois personnes ont accepté d'assurer la « vacance du poste de directeur » et ont permis de continuer l'animation du GIS et entamer la réflexion sur les évolutions futures. Xavier Collet, Denis Milan et Michel Rivière ont ainsi su, en plus de leurs charges habituelles, mobiliser leur énergie pour maintenir la dynamique nécessaire.

Transition qui a vu ma nomination en tant que directeur le 16 décembre par les membres du Conseil de Groupement. Je les remercie de la confiance qui m'est ainsi faite, en espérant être à la hauteur des défis que nous devons relever pour faire évoluer Génotoul.

Transition qui va se poursuivre avec le départ annoncé, en mai, de Monique Falières qui avec Claude Chevalet aura lancé Génotoul et sans qui la structure ne serait pas ce qu'elle est.

#### **Évolutions.**

Certaines ont été accomplies au cours de l'année écoulée. On peut citer en particulier, l'intégration de deux nouvelles entités, le Centre de Ressources Biologiques humaines (CRBh) et la Plateforme de Phénotypage Plantes-Microorganismes (TPMP).

D'autres sont à faire. La Génopole Toulouse a maintenant plus de 17 ans d'existence et les paysages qui nous entourent se sont largement modifiés tant au niveau local, régional, national qu'europpéen. Certaines adaptations se sont faites au cours des années, mais nous devons, sans aucun doute, réfléchir, proposer et faire aboutir des changements qui nous permettront de répondre au mieux aux attentes des différents acteurs.

Ces évolutions se font dans un cadre qui lui aussi évolue. En premier celui de la nouvelle Région Occitanie, partenaire important de Génotoul dont les contours ont changé et nous pousse à des rapprochements souhaités et souhaitables avec nos collègues d'autres agglomérations (Montpellier, Perpignan principalement).

La fin du PIA1 et les réflexions du PIA3 et des infrastructures nationales qui doivent nous ouvrir des opportunités que nous devons saisir.

C'est cette réflexion que je veux conduire avec vous au cours de la première moitié de cette année afin de nous fixer tous ensemble des objectifs ambitieux, mais atteignables que nous pourrons, pour certains, mettre en route dès avant l'été, je le souhaite.

Je veux aussi saluer et remercier le travail accompli par les responsables et les personnels des plateformes. Leur implication et leur travail tout au long de l'année sont à la base des avancées que vous découvrirez dans ce livret. L'Assemblée Générale qui s'est tenue le 2 Février a été un moment d'échanges extrêmement instructif pour moi et les propositions, suggestions émises collectivement constituent des pistes de réflexion.

Je remercie également les directeurs des unités d'adossement et les membres des différentes instances du GIS. Leur soutien est indispensable à la bonne marche des différentes entités qui constituent Génotoul et du réseau lui-même.

Vous souhaitant au travers de la lecture de ces faits marquants, d'appréhender la richesse, la qualité, la nouveauté de ce que font et réalisent les plateformes de Génotoul.

---

**Luc Pénicaud**

Directeur de la Génopole Toulouse

## Un réseau de plateformes en sciences du vivant




**PLATEFORMES en émergence**  
 PHENOTYPAGE (INRA-FR AIB, LabEx TULIP)  
 CÉSAR (Centre de ressources biologiques humaines)  
 Coordination des Structures d'Appui à la Recherche au CHU de Toulouse

**IBISA**  
 Gis - Infrastructures en Biologie Santé et Agronomie

**8000 m<sup>2</sup>** mobilisés sur les différents sites

**+ 81**  
**projets**  
 nationaux, européens,  
 internationaux

**56**  
**projets**  
 collaboratifs  
 avec des entreprises

  
**302** agents mobilisés  
 représentant  
**231** ETP

**785**  
 personnes formées  
 (public-privé)

### Nos plateformes

Investies dans des infrastructures nationales

- France Génomique,
- MetaboHub,
- Institut Français de BioInformatique,
- PROFI,
- CELPHEDIA

Et démonstrateur

- Toulouse White Biotechnology
- Equipex (ANINFIMIP)

**Toutes sont ouvertes  
 vers le monde de l'industrie**



**Interventions** dans les :

- masters,
- écoles doctorales,
- écoles d'ingénieurs.

**En France comme  
 à l'étranger**



Un fort soutien de l'État, de la Région Occitanie, de l'Europe



Données 2015 - 2016

PROJET COFINANCÉ PAR LE FONDS EUROPÉEN DE DÉVELOPPEMENT RÉGIONAL



## Séquençage d'ADN longs fragments pour l'analyse de génomes complexes

Afin d'accompagner les équipes de recherche dans l'analyse des variations structurales de génomes complexes, la Plateforme GeT a investi en mars 2015, grâce à l'aide d'un programme régional, dans une technologie unique en France : le PACBIO RSII de la société Pacific Biosciences. Cette technologie permet le séquençage de longs fragments d'ADN. Elle permet l'analyse de zones riches en séquences répétées, impossible à assembler lorsque l'on ne dispose que de petits fragments, et est donc totalement complémentaire des technologies HISEQ, MISEQ, S5 et PGM déjà présentes sur la plateforme GeT de Toulouse. Cette technologie et l'expertise acquise par l'équipe GeT a permis de répondre aux besoins urgents de la communauté scientifique. Dans le cadre du programme SUNRISE porté par l'équipe Tournesol de l'INRA et par ses partenaires, le premier projet réalisé sur le PACBIO RSII a permis de séquencer le génome du Tournesol.

Considérant que notre communauté travaille sur des génomes complexes, nous avons développé des protocoles de préparation permettant le séquençage de fragments les plus longs possibles. Dans le cadre plus particulier du génome du tournesol, dont 30 % des séquences sont répétées, le séquençage Illumina (avec une profondeur de 125X) ne permet l'assemblage que de 43 % du génome. La taille des lectures (>20kbases) obtenue grâce au développement de protocoles sur le PACBIO RSII, nous a permis d'obtenir un assemblage des données PacBio avec une couverture de 84 % du génome et plus de 90 % de séquences ancrées.

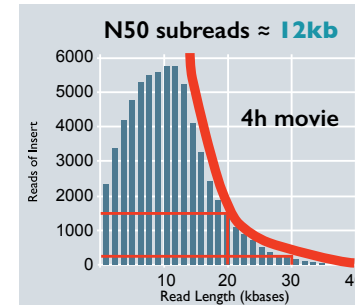
Deux autres programmes pilotes ont été réalisés : un programme en partenariat avec SOFIPROTEOL, pour l'analyse du génome de l'Orobanche cumana,

parasite du tournesol et un programme en partenariat avec Libragen pour l'analyse de méta-

génomes bactériens à partir d'amplicons 16S pleine longueur. Depuis sa mise en production, le PACBIO RSII a permis de répondre à plus de 40 projets scientifiques, dont voici quelques exemples :

1. Le projet GENEPIA : identification des bases génétiques et épigénétiques de l'adaptation à différentes plantes hôtes chez la bactérie phytopathogène *Ralstonia solanacearum*. Un dossier a été soumis à l'ANR afin de poursuivre ce projet.
2. Le projet JAVAFISH : obtention d'un génome de bonne qualité permettant d'assembler les chromosomes sexuels riches en éléments répétés. Un dossier a été soumis à l'ANR afin de poursuivre ce projet.
3. Le projet PBmetaSeed : analyse métagénomique d'un ensemble bactérien associé à une plante.
4. Le projet BACPAC : séquençage de BAC multi espèces contenant des séquences non assemblables avec les autres technologies de séquençage.
5. Le projet Tenacigeno : étude des flavobactéries marines pathogènes du poisson.
6. Le projet SalHosTrop : assemblage et analyse de méthylation de la Salmonelle en première étape du projet.
7. Le projet PHYTNESSE : étude de la diversité et du métabolisme des microalgues de la famille *Ostreococcus* (publication en cours).
8. Le projet HEPAC : caractérisation de génomes complets de l'hépatite E afin d'identifier de manière plus exhaustive la fréquence, la nature et la localisation des insertions au sein du génome viral.
9. Le projet PacificBee : assemblage de référence pour l'abeille noire, support de plusieurs projets ANR et Européens d'étude du génome des abeilles.
10. Le projet ATHERANOS : identification de nouveaux biomarqueurs d'intérêt pour le ciblage de la plaque d'athérome.

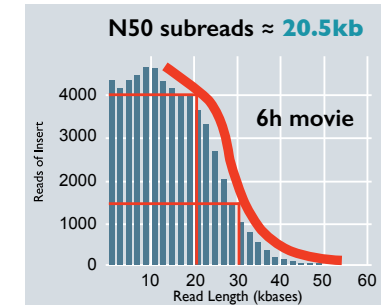
12kb cut off  
(loading 0.15nM=868Mb)



Optimisation des conditions



20kb cut off  
(loading 0.45nM=1000Mb)

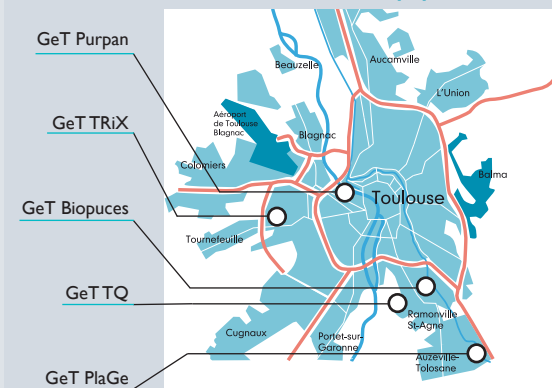


Pour les besoins de la communauté, nous continuons de développer de nouveaux protocoles ciblant les nombreuses applications. Nous travaillons notamment depuis plusieurs mois sur l'analyse transcriptomique pleine longueur permettant d'avoir les différentes formes d'épissage (Isoformes).

## PUBLICATIONS

- Le Gac et al., Evolutionary processes and cellular functions underlying divergence in *Alexandrium minutum*, *Molecular Ecology*, 2016
- Boutet et al., SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population, *BMC Genomics*, 2016
- Osteil et al., A Panel of Embryonic Stem Cell Lines Reveals the Variety and Dynamic of Pluripotent States in Rabbits, *Stem Cell Reports*, 2016
- Chen et al., Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency, *Nature Communications*, 2015
- Rupp et al., A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model, *PLoS Genetics*, 2015

### Localisation des équipements




Responsable scientifique : Denis Milan

Animateur : Cécile Donnadieu

Localisation des équipements :  
5 sites sur la région toulousaine

Contacts : [get@genotoul.fr](mailto:get@genotoul.fr),

Site web : <http://get.genotoul.fr/>

 @get\_genotoul

## Des cartes optiques pour analyser les génomes complexes

Afin de faciliter l'étude des génomes complexes dans leur globalité le CNRGV s'est doté de la technologie Irys® de BioNano Genomics (<http://bionanogenomics.com/> - Figure 1). Cette technologie permet la construction rapide de cartes physiques d'un génome entier. C'est un outil précieux pour avoir une vision globale de l'organisation des génomes. La cartographie optique est basée sur la visualisation directe de molécules d'ADN de plusieurs dizaines ou centaines de kb marquées sur des motifs de séquences spécifiques. L'ADN génomique est extrait et des molécules d'ADN longues de 100kb à 2Mb sont marquées par fluorescence par une enzyme reconnaissant une séquence spécifique (Figure 2). Les images des molécules d'ADN marquées et séparées sur une puce dans des nanochannels, sont analysées afin d'obtenir un « code-barre » spécifique pour chaque molécule (Figure 3). En comparant leurs codes-barres, les molécules d'ADN sont ensuite assemblées pour donner une carte physique du génome étudié. Une puce unique composée de deux flowcells a un débit de 50 à 100 Gb. Une couverture minimale de 50X est préconisée pour l'assemblage de novo d'un génome. Les principales applications de cette technologie sont la finition de l'assemblage des séquences génomiques et l'étude des variations structurales entre différents génotypes. La première application a été réalisée sur le génome du Tournesol en collaboration avec le LIPM - Laboratoire des Interactions Plantes Microorganismes ([http://www6.toulouse.inra.fr/lipm\\_eng/](http://www6.toulouse.inra.fr/lipm_eng/)) dans le cadre du projet SUNRISE (<http://www.sunrise-project.fr/Fr/>). Le génome du tournesol a une taille de 3,6 Gb. Nous avons construit une carte optique du tournesol avec la technologie BioNano et obtenu des molécules allant de 150kb à 2,3Mb. La carte physique résultante est

composée de 2228 scaffolds et une N50 à 1,98 Mb. Préalablement, un séquençage du génome du tournesol avait été réalisé avec la technologie PacBio qui avait permis de générer un assemblage de qualité (12318 contigs avec une N50 de 524 Kb et une taille totale de 2,93 Gb représentant une couverture de plus de 80 % du génome). Un assemblage hybride (Hybrid Scaffolding) combinant l'assemblage obtenu avec la technologie PacBio et la carte optique Bionano, a permis d'améliorer la qualité de l'assemblage final (N50 de 2,87 Mb). Cette technologie offre donc de nouvelles possibilités pour contribuer à la compréhension des génomes. Elle permettra non seulement d'améliorer l'assemblage des génomes obtenu à partir des nouvelles technologies de séquençage par un ancrage sur des molécules physiques, mais également d'explorer la diversité au sein d'une espèce en mettant en évidence des variations structurales (Insertions, Délétions, Translocations, Variation du nombre de copies – CNV).

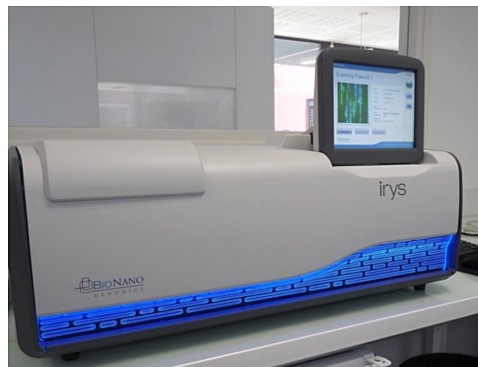


Figure 1 : le système Irys® (BioNano Genomics) installé au CNRGV

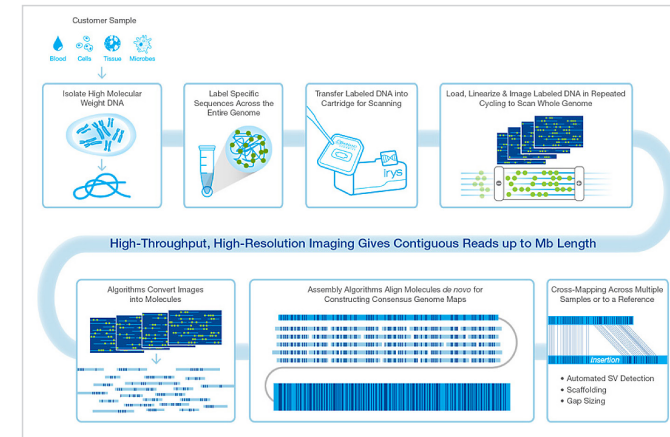


Figure 2 : Schéma des différentes étapes permettant de générer une carte optique

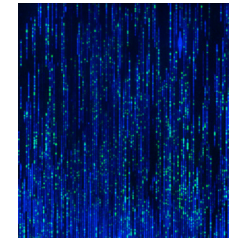


Figure 3 : Exemple d'image obtenue représentant les molécules d'ADN marquées (code-barres)

## PUBLICATIONS

- Bellec A, Courtial A, Cauet S, Rodde N, Vautrin S, Beydon G, Arnal N, Gautier N, Fourment J, Prat E, Marande W, Barriere Y and Bergès H (2016) Long Read Sequencing Technology to Solve Complex Genomic Regions Assembly in Plants. *Next Generat Sequenc & Applic.* 3:2 doi.org/10.4172/2469-9853.1000128
- Schweiger W, Steiner B, Vautrin S, Nussbaumer T, Siegwart G, Zamini M, Jungreithmeier F, Gratl V, Lemmens M, Mayer KF, Bergès H, Adam G, Buerstmayr H. (2016) Suppressed recombination and unique candidate genes in the divergent haplotype encoding Fhb1, a major Fusarium head blight resistance locus in wheat. *Theor Appl Genet.* 129(8):1607-23. doi: 10.1007/s00122-016-2727-x
- Hao DC, Vautrin S, Song C, Zhu YJ, Bergès H, Sun C, Che SL (2015) The first insight into the salvia (Lamiaceae) genome via BAC library construction and high-throughput sequencing of target BAC clones. *Pak. J. Bot.* 47(4): 1347-1357.
- Akpinar BA, Magni F, Yuce M, Lucas SJ, Šimková H, Šafař J, Vautrin S, Bergès H, Cattonaro F, Doležel J, Budak H. (2015) The physical map of wheat chromosome 5DS revealed gene duplications and small rearrangements. *BMC Genomics.* 16:453. doi: 10.1186/s12864-015-1641-y.

**Responsable scientifique :** Helene Berges

**Responsable opérationnel :** William Marande

**Localisation des équipements :**

INRA - CNRGV  
24 Chemin de Borde Rouge  
CS 52627

31326 Castanet Tolosan

**Contact :** Helene Berges

**Tel :** 33(0) 561 28 55 65

**Fax :** 33(0) 561 28 55 64

**Port :** 06 07 95 89 01

**Email :** Helene.Berges@inra.fr

**Site web :** <http://cnrgv.toulouse.inra.fr/>

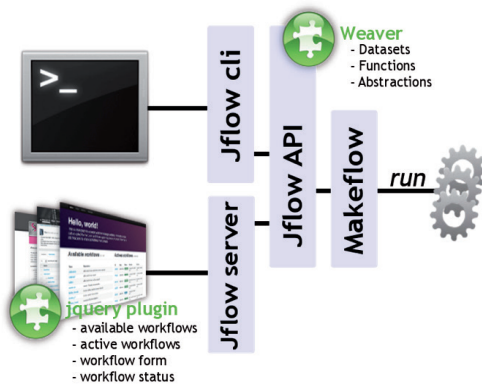
 @cnrgv

## JFlow ou comment développer de manière intégrée des applications web dédiées aux biologistes pour le traitement de données massives

Un grand nombre de traitements bio-informatiques utilisés de manière routinière nécessitent d'enchaîner plusieurs logiciels pour produire des résultats interprétables par les biologistes. Un exemple classique est la génération du fichier de comptage des séquences pour l'analyse différentielle qui enchaîne le nettoyage des séquences, leur alignement sur un génome de référence, le comptage des séquences s'alignant sur chaque

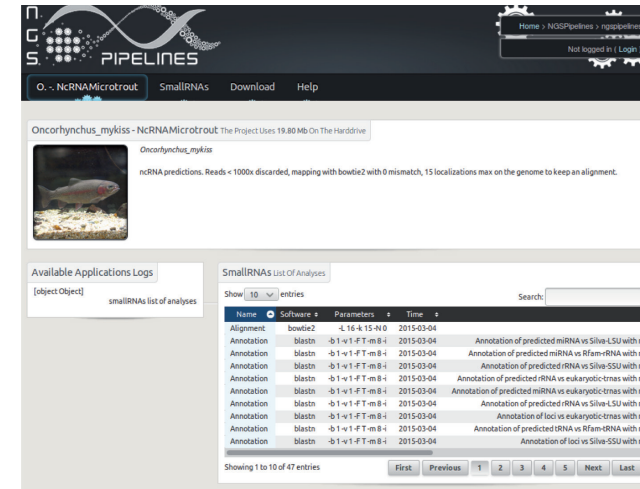
gène et la production du tableau comportant les valeurs de comptage pour tous les échantillons analysés.

JFlow (Mariette et al. 2016) est un système permettant de créer, de mettre en œuvre et de suivre le déroulé des chaînes de traitement (workflow). Il met à disposition un ensemble de modules et une bibliothèque de fonctionnalités (en Python) visant le développement intégré d'applications web.



JFlow est utilisé en routine dans les applicatifs produits par la plate-forme tels que NG6 (Mariette et al., 2012, <http://ng6.toulouse.inra.fr>), RNAbrowse (Mariette et al., 2014, Magnanou et al., 2014, Rivière et al. 2015...) et sRNAbrowse (Juanchich et al., 2016).

Il continue d'évoluer pour intégrer les fonctionnalités permettant de faciliter et optimiser le développement d'applications web intégrant des chaînes de traitements visant à répondre aux besoins des biologistes.



**Disponibilité :** JFlow est disponible sous licence GNU General Public License (GPL) à l'URL <http://bioinfo.genotoul.fr/jflow>.

**Contact :** [Jerome.Mariette@inra.fr](mailto:Jerome.Mariette@inra.fr)

## PUBLICATIONS

*Publications en lien avec les applicatifs ou leur mise en œuvre s'appuyant sur JFlow*

- Mariette J, Escudié F, Bardou P, Nabihoudine I, Noirot C, Trotard MS, Gaspin, C, Klopp C. Jflow : a workflow management system for web applications. *Bioinformatics*. 2016 Feb 1; 32(3):456-8.
- Juanchich A, Bardou P, Rué O, Gabillard JC, Gaspin C, Bobe J, Guiguen Y. Characterization of an extensive rainbow trout miRNA transcriptome by next generation sequencing. *BMC Genomics*. 2016 Mar 1; 17:164.
- Mariette J, Noirot C, Nabihoudine I, Bardou P, Hoede C, Djari A, Cabau C, Klopp C. RNAbrowse : RNA-Seq de novo assembly results browser. *PLoS One*. 2014 May 13; 9(5):e96821.
- Magnanou E, Klopp C, Noirot C, Besseau L, Falcón J. Generation and characterization of the sea bass *Dicentrarchus labrax* brain and liver transcriptomes. *Gene*. 2014 Jul 1; 544(1):56-66.
- Riviere G, Klopp C, Ibouniyamine N, Huvet A, Boudry P, Favrel P. GigaTON: an extensive publicly searchable database providing a new reference transcriptome in the pacific oyster *Crassostrea gigas*. *BMC Bioinformatics*. 2015 Dec 2; 16:401.
- Mariette J, Escudié F, Allias N, Salin G, Noirot C, Thomas S, Klopp C. NG6 : Integrated next generation sequencing storage and processing environment. *BMC Genomics*. 2012 Sep 9; 13:462.

### Responsable scientifique

Christine Gaspin : [christine.gaspin@inra.fr](mailto:christine.gaspin@inra.fr)

### Responsable technique

Christophe Klopp : [christophe.klopp@inra.fr](mailto:christophe.klopp@inra.fr)

### Localisation des équipements

INRA – CR Occitanie-Toulouse CS 52627  
31326 Castanet Tolosan cedex

### Contact

Christine Gapin : +33 5 61 28 52 82

### Site web

[bioinfo.genotoul.fr](http://bioinfo.genotoul.fr)



## Spectrométrie de Masse structurale pour l'analyse des protéines entières et des complexes protéiques macromoléculaires

La spectrométrie de masse (MS), technologie dédiée à l'origine à l'analyse des petites molécules, des peptides et de la structure primaire des protéines, a considérablement évolué au cours des 20 dernières années pour devenir à l'heure actuelle un outil supplémentaire en biologie structurale. Grâce à une grande diversité de méthodes comme la protéomique top-down, la MS native, la mobilité ionique, le pontage chimique ou l'échange hydrogène-deutérium (H/D), la MS structurale vise à collecter un maximum d'informations sur les structures secondaires, tertiaires et quaternaires des assemblages protéiques [1]. En 2016, grâce au financement de la Région Occitanie dans le cadre du

CPER (première phase 2015-2017), le site principal de l'Infrastructure Protéomique de Toulouse, localisé à l'IPBS, a acquis un nouvel ensemble d'équipement dédié aux approches de spectrométrie de masse structurale (Fig. 1). Ces équipements se composent d'un spectromètre de masse Synapt G2Si à gamme de masse étendue, équipé d'une cellule de mobilité ionique pouvant être couplée à la MS native, et de modules complémentaires pour la chromatographie d'exclusion stérique (SEC) et l'échange H/D. Ces nouvelles compétences vont permettre de nouvelles offres de services à la communauté, en particulier pour la mesure de masses et la caractérisation de protéines purifiées.



Figure 1 : Nouvel équipement à l'IPBS. Synapt G2Si et modules complémentaires pour la MS structurale.

### Protéomique Top-down pour l'identification de modifications post-traductionnelles

Les modifications post-traductionnelles (MPTs) jouent un rôle prépondérant dans la fonction et la régulation des protéines. La spectrométrie de masse représente un des outils les plus efficaces pour leur identification et leur localisation sur une séquence protéique. Un protocole classique de protéomique, incluant une étape de protéolyse enzymatique, permet d'identifier, de localiser et de quantifier ces modifications sur des peptides tryptiques. Cependant, dans le cas de modifications multiples, cela ne permet pas de déterminer si elles sont portées par la même forme modifiée de la protéine (appelée protéoforme). À l'inverse, la protéomique top-down, qui consiste à analyser et fragmenter les protéines intactes en phase gazeuse, permet d'étudier chaque protéoforme individuellement. Durant les deux dernières années, nous avons implémenté des méthodes top-down sur des appareils de type Orbitrap (Velos-ETD

et Fusion) déjà présents au laboratoire. Nous les avons mises en œuvre avec succès pour caractériser de façon exhaustive des canaux ioniques issus de *Corynebacterium glutamicum* [2], ainsi que plus de 100 protéoformes lipoglycosylées de l'antigène 19kDa de *Mycobacterium tuberculosis* [3]. Les protéines intactes sont directement injectées sur une nanocolonne C4 (phase inverse), avant d'être analysées avec un détecteur haute résolution de type Orbitrap. Les spectres obtenus sont ensuite automatiquement déconvolués, afin d'obtenir des cartes 3D de chacune des analyses effectuées (Fig. 2B). Ces « empreintes protéiques » permettent de déterminer rapidement le degré d'hétérogénéité d'un échantillon et de mettre en évidence la présence éventuelle de protéoformes tronquées ou modifiées. Dans un second temps, les protéoformes d'intérêt peuvent être sélectionnées et fragmentées afin de localiser les sites porteurs de MPTs (Fig. 2C).

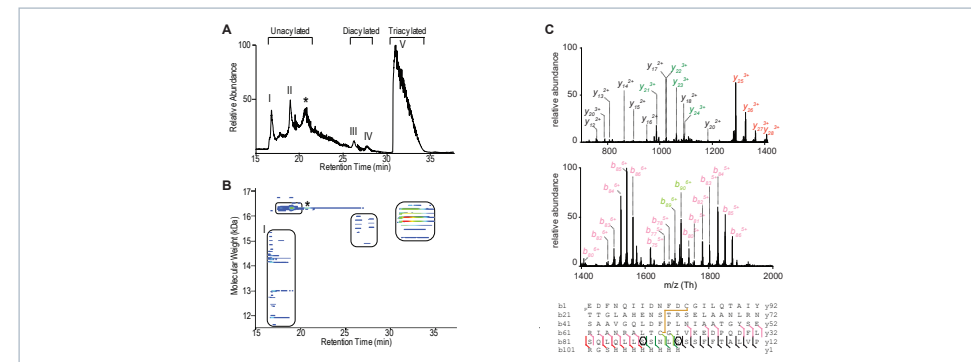


Figure 2 : A) Séparation LC-MS de différentes formes acylées de la protéine rLpQ-His exprimée chez *M. smegmatis*. Les pics I/II, III/IV et V du chromatogramme correspondent respectivement aux formes non-acylées, diacylées et triacylées de la protéine. B) Empreinte protéique de l'analyse LC-MS, indiquant les masses moléculaires des protéoformes tout au long du gradient chromatographique. L'intensité relative de chacune des espèces est représentée en couleur (du bleu au rouge). Adapté de [2]. C) Analyse MS/MS d'une protéine entière, permettant une description complète des protéoformes. Adapté de [2].

### MS Native

La MS native s'est considérablement développée depuis une vingtaine d'années et est aujourd'hui largement utilisée en biologie structurale [4]. En conservant les complexes protéiques intacts en phase gazeuse, la MS native renseigne directement sur leur stoechiométrie. Des interactions avec des ligands de nature diverse (lipides, petites molécules, peptides, sucres) peuvent également être observées grâce à cette technique. Les derniers développements technologiques ont permis d'accéder désormais à l'étude de complexes de hauts poids moléculaires (ribosomes, protéasome, capsides virales) avec de très bonnes résolutions et précision en masse.

### Échange Hydrogène-Deutérium et MS

Cette méthode permet d'observer des différences conformationnelles entre deux échantillons protéiques. Son principe repose sur la dilution des protéines, ou complexes protéiques, dans un tampon deutéré. La détermination des taux d'échange des atomes d'hydrogène des liaisons peptidiques avec les atomes de deutérium de la solution permet d'identifier à la fois les régions facilement accessibles au solvant et les régions masquées au cœur de la protéine (ou du complexe). Cette approche en pleine expansion peut être mise en œuvre pour l'étude de la dynamique des protéines, pour comparer différentes conformations protéiques, ou encore pour identifier des interfaces d'interactions protéine-protéine ou protéine-ligand.

## PUBLICATIONS

- [1] Marcoux J, Cianféran S (2015) « Towards integrative structural mass spectrometry: benefits from hybrid approaches » *Methods* 89:4-12.
- [2] Carel C\*, Marcoux J\*, Parra J, Réat V, Latgé G, Laval F, Burlet-Schiltz O, Demange P, Milon A, Daffé M, Tropis M, Renault M (2017) "Post-translational O-mycoloylation mediates protein targeting to the mycomembrane in *C. glutamicum*" *PNAS* In press
- [3] Parra J\*, Marcoux J\*, Poncin I, Cnaan S, Herrmann JL, Nigou J, Burlet-Schiltz O, Rivière M (2017) "Scrutiny of *M. tuberculosis* 19kDa antigen proteoforms provides new insights in the lipoglycoprotein biogenesis paradigm" *Scientific Reports* In Press
- [4] Marcoux J and Robinson CV (2013) "Twenty years of gas phase structural biology" *Structure* 21(9):1541-1550.

Directeur : Odile Schiltz

Responsables opérationnels

Site principal : Odile Schiltz  
Site partenaire : Frédéric Lopez

Contacts (mail, web, twitter, facebook...) :

Mail site principal : proteomique@ipbs.fr  
Mail site partenaire : Frederic.Lopez@inserm.fr

Site web : <http://proteomique.ipbs.fr>

Localisation des équipements

Site principal : IPBS, CNRS, UMR 5089,  
205 route de Narbonne,  
31077 Toulouse Cedex

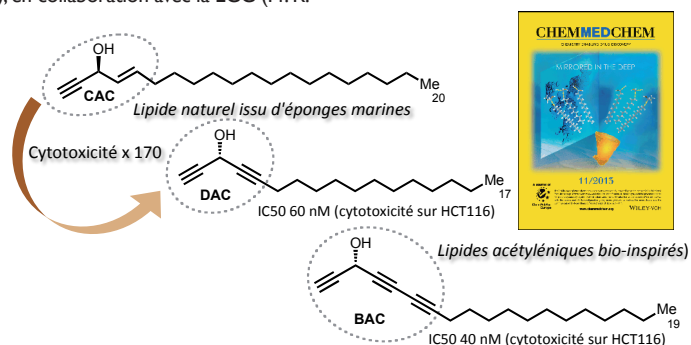
Site partenaire : CRCT, Inserm, U 1037,  
2 avenue Hubert Curien, Oncopole entrée C,  
31037 Toulouse Cedex

## PICT : vers des anti-cancéreux inspirés de la nature

Les éponges marines représentent une source illimitée de composés naturels à potentiel pharmacologique parmi lesquels se distingue une famille d'hydrocarbures hydroxylés riches en triples liaisons carbone-carbone appelés lipides acétyléniques hydroxylés. Un grand nombre de ces métabolites marins présente une cytotoxicité contre les cellules cancéreuses avec des CE50 de l'ordre du nanomolaire. Une sous-classe intrigante de lipides acétyléniques hydroxylés est caractérisée par la présence d'un motif Chiral AlcynylCarbinol (CAC) le long de squelettes aliphatiques linéaires pouvant présenter une élégante symétrie C<sub>2</sub>.<sup>1</sup> Situé en position terminale de la chaîne lipidique, ce motif CAC s'est révélé être le pharmacophore clé des métabolites les plus cytotoxiques. Malgré l'isolement et la caractérisation de plus d'une centaine de ces lipides marins, les travaux visant à déterminer les relations structure-activité et à élucider les mécanismes d'action associés sont restés limités en raison du manque d'études intégrant à la fois la chimie et la biologie. Le laboratoire SPCMIB (Dr. Y. Génisson), en collaboration avec la LCC (Pr. R.

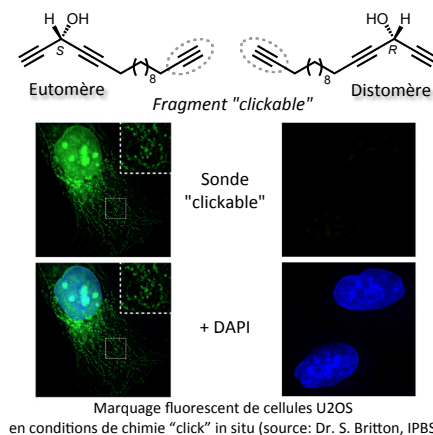
Chauvin) et l'IPBS (Dr. E. Joly/Dr. P. Calsou), a lancé un programme de recherche visant à explorer le potentiel pharmacologique de lipides acétyléniques fonctionnels bio-inspirés.

L'étude systématique du pharmacophore CAC sur la base d'une variation structurale inédite du motif naturel (S)-(4E)-en-1-yn-3-ol a été réalisée.<sup>2</sup> Une influence récurrente de la configuration absolue du centre carbinol sur la cytotoxicité a été mise en évidence. Les eutomères présentent une configuration formellement opposée à celle la plus fréquemment rencontrée dans la nature, avec des rapports eudismiques supérieurs à 40. Un produit de synthèse à chaîne tronquée présentant une unité DiAlcynylCarbinol (DAC) originale s'est révélé être non seulement le plus facile à synthétiser, mais aussi le plus actif, avec une augmentation de 170 fois la cytotoxicité par rapport à un échantillon synthétique du composé naturel. De nombreux autres analogues ont été préparés par synthèse asymétrique,<sup>3</sup> y compris dérivés sophistiqués<sup>4</sup> de symétrie C<sub>2</sub> et des sondes fluorescentes.<sup>5</sup>



En route vers l'élucidation du mécanisme d'action de ces lipides cytotoxiques, une sonde portant un fragment ω-acétylénique « clickable » a également été conçue pour l'identification de cibles protéiques par chimie « click » sur cellules.<sup>4</sup> Non seulement cette sonde conserve le niveau de cytotoxicité comparable à celle de l'analogue parent non « clickable », mais l'écart

d'activité entre ses deux énantiomères est également corrélé à un marquage cellulaire fluorescent nettement distinct.<sup>6</sup> Dernièrement, l'activité cytotoxique du pharmacophore DAC a pu être augmentée par une approche d'éthynylogation conduisant à l'analogue ButadiynylAlcynylCarbinol (BAC) de dernière génération.



Ce programme de recherche ouvre ainsi la voie à la mise au point de composés anticancéreux innovants inspirés de la Nature. Le plateau de chromatographie de haute technologie de la plateforme PICT a été indispensable pour le développement de ce projet. Une évaluation rapide, fiable et précise de l'excès énantiomérique de chacune de ces molécules chirales, tâche non triviale réalisable grâce à l'utilisation de la chromatographie ultraperformante en fluide supercritique (UPC2), a notamment été déterminante pour mettre à jour la façon dont l'activité antitumorale de ces lipides acétyléniques hydroxylés dépend de leur configuration absolue, une observation sans précédent.

## RÉFÉRENCES

- Chiral alkynylcarbinols from marine sponges: asymmetric synthesis and biological relevance.** D Listunov, V Maraval, R Chauvin and Y Génisson, Nat. Prod. Report, 2015, 32, 49-75
- Identification of chiral alkenyl- and alkynylcarbinols as pharmacophores for potent cytotoxicity.** D El Arfaoui, D Listunov, I Fabing, M Oukessou, C Frongia, V Lobjois, A Samson, F Ausseil, A Ben-Tama, E El Hadrami, R Chauvin and Y Génisson, ChemMedChem, 2013, 8, 1779-1786
- Extended structural modulation of bio-inspired chiral lipidic alkynylcarbinols as antitumor pharmacophores.** D Listunov, C Billot, E Joly, I Fabing, Y Volovenko, Y Génisson, V Maraval, R Chauvin, Tetrahedron 2015, 71, 7920-7930
- Asymmetric synthesis and biological evaluation of natural or bio-inspired cytotoxic C<sub>2</sub>-symmetrical lipids with two terminal chiral alkynylcarbinol pharmacophores.** D Listunov, I Fabing, N Saffon-Merceron, H Gaspard, Y Volovenko, V Maraval, R Chauvin, Y Génisson, J. Org. Chem. 2015, 80, 5386-5394
- Fluorophore-tagged pharmacophores for antitumor cytotoxicity: Modified chiral lipidic dialkynylcarbinols for cell imaging.** D Listunov, S Mazères, Y Volovenko, E Joly, Y Génisson, V Maraval, R Chauvin, Bioorg. Med. Chem. Lett. 2015, 25, 4652-4656
- S Britton, P Calsou, E Joly, Y Génisson, V Maraval, R Chauvin, résultats non publiés.
- Ethynylogation approach in pharmacophore design: from alkynyl- to butadiynyl-carbinols vs antitumoral cytotoxicity.** D Listunov, N Saffon-Merceron, E Joly, I Fabing, Y Génisson, V Maraval, R Chauvin, Tetrahedron 2016, 72, 6697-6704

**Directeur :** Laurent Maveyraud

**Site web :** <http://cribligand.ipbs.fr>

**E-mail :** pict@ipbs.fr

**Contact**  
Virginie Nahoum  
+ 33 5 61 17 54 48

### Localisation des équipements

CNRS - Université Paul Sabatier 205, route de Narbonne  
BP 64182 31077 Toulouse

IPBS, UMR 5089 CNRS Université Paul Sabatier Toulouse

LSPCMIB, UMR 5068 CNRS Université Paul Sabatier Toulouse

LISBP, UMR 5504 CNRS-INSAToulouse

UMR 792 INRA, INSA Toulouse



# Workflow4Metabolomics : une e-infrastructure pour le traitement des données de métabolomique

L'infrastructure on line W4M est un environnement performant unique centralisé et ergonomique permettant d'analyser les données de métabolomique, incluant des outils de pré-traitement, d'analyse statistique et d'annotation des métabolites.

Galaxy est une plateforme WEB open-source, initialement développée pour analyser les données de grande dimension issues de la recherche biomédicale, en particulier les données de séquençage « next generation » [1]. Elle a ensuite été adaptée pour l'analyse des données de protéomique. Cette plateforme fédère une large communauté comprenant à la fois des utilisateurs et des développeurs d'outils d'analyse. Cette communauté assure les mises à jour logicielles, le support et l'intégration de nouveaux outils d'analyse. Galaxy fournit une interface ergonomique : les utilisateurs sans expérience en programmation peuvent facilement spécifier les paramètres nécessaires à l'exécution des outils et chaîner ces outils dans l'ordre souhaité (workflow) pour analyser leurs données dans l'interface graphique d'édition. Les principales caractéristiques de Galaxy sont :

- la traçabilité et le stockage des résultats,
- la possibilité de partager des résultats/workflows entre utilisateurs, laboratoires...
- la possibilité d'utiliser un environnement de management de workflow.

Galaxy présente plusieurs caractéristiques intéressantes pour l'analyse des données de métabolomique par rapport à d'autres plateformes de traitement de données : pas de limitation de taille des données, la possibilité d'automatisation des « pipelines » d'analyse, et la reproductibilité des résultats. Ainsi, afin de partager des stratégies d'analyse des données de métabolomique et de centraliser les outils d'analyse et les pratiques, l'infrastructure workflow4metabolomics (W4M) basée sur Galaxy [2] a été développée dans le cadre

d'une collaboration entre l'infrastructure nationale en métabolomique et Fluxomique (Metabohub) et l'Institut Français de Bioinformatique (IFB), Fig.1.

W4M inclut des pipelines complets d'analyse des données de LC/MS (Liquid Chromatography/ Mass Spectrometry), de GC/MS (Gas Chromatography/ Mass Spectrometry) et de RMN (spectrométrie par Résonance Magnétique Nucléaire) allant du prétraitement des données à l'étape d'annotation, en passant par des étapes de normalisation, de contrôle qualité des données et d'analyse statistique (test univarié, modélisation multivariée par régression PLS/OPLS). Ces workflows modulaires et extensibles comprennent des composants existants (packages XCMS et CAMERA pour les données de LC/MS, etc.), mais également un ensemble d'outils complémentaires « maison ». Par exemple, le workflow d'analyse des données de RMN [3], illustré dans la Fig.2, comprend actuellement des outils pour l'alignement des spectres, le bucketing, la normalisation ainsi qu'un algorithme d'annotation [4].

Un compte gratuit peut être demandé via l'URL <http://informatique-mia.inra.fr/workflow4metabolomics/>.

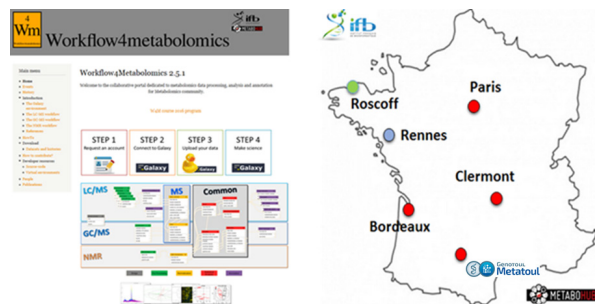


Fig.1 : Page d'accueil de l'infrastructure d'analyse des données de métabolomique W4M.

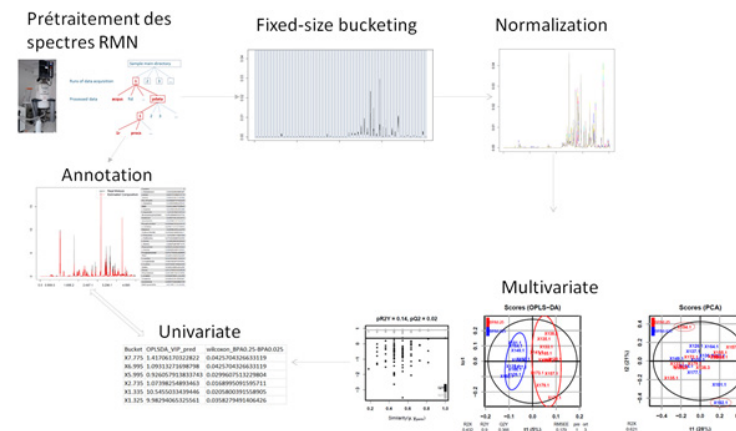


Fig.2 : Exemple de workflow d'analyse des données de métabolomique générées par spectroscopie RMN : du traitement des spectres à l'annotation des métabolites. En haut à gauche : acquisition des spectres RMN et prétraitement (fichiers Bruker); En haut au milieu : réduction des données (bucketing + normalisation); En bas à droite : analyse multivariée (score plots ACP et O-PLS-DA, tes des permutations); En bas à gauche : test univarié (buckets discriminants et significatifs); Au milieu à gauche : annotation des spectres et correspondance avec les métabolites discriminants.

## PUBLICATIONS

[1] Blankenberg D., Von Kuster G., Coraor N., Ananda G., Lazarus R., Mangan M., Nekrutenko A. and Taylor J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology*, 89:19.

[2] Giacomoni F., Le Corguillé G., Monsoor M., Landi M., Pericard P., Pétéra M., Duperier C., Tremblay-Franco M., Martin J.-F., Jacob D., Goullitquer S., Thévenot E.A. and Caron C. (2014). Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. *Bioinformatics*, <http://dx.doi.org/10.1093/bioinformatics/btu813>.

[3] Tremblay-Franco M., Landi M., MONSOOR M., GIACOMONI F., DUPERIER C., LE CORGUILLÉ G., PÉTERA M., MARTIN J.F., JACOB D., THEVENOT E., CARON C., CANLET C. A Galaxy workflow for NMR data pre-processing. 11th International Conference of the Metabolomics Society, June 29th 2015 to July 2nd 2015, San Francisco (Californie), USA.

[4] Tardivel P., Servien R. and Concordet D. Non asymptotic active set properties of lasso-type estimators in small-dimension (submitted).

### Localisation des équipements

MetaToul-Réseau : Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP) - INSA - 135, avenue de Rangueil – 31077 Toulouse cedex 4

MetaToul Lipidomique : Institut des Maladies Métaboliques et Cardiovasculaires (I2MC) – CHU Rangueil - 1 avenue Jean Poulhès - BP 84225 - 31432 Toulouse Cedex 4

MetaToul-Axiom : Laboratoire de Toxicologie Alimentaire (TOXALIM) – INRA - 180, chemin de Tournefeuille - BP 93173 - 31027 TOULOUSE cedex 3

MetaToul-Metabolite Végétaux : Laboratoire de Recherches en Sciences Végétales (LRSV) – INRA - 24 Chemin de Borde Rouge - 31326 Castanet-Tolosan

**Directeur scientifique**  
Jean Charles Portais

**Directeurs opérationnels**  
Justine Bertrand-Michel  
& Laurent Debrauwer

**Site Web**  
<http://www.metatoul.fr/>

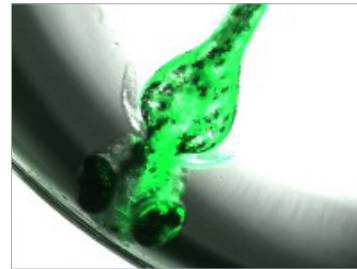
**Contact**  
[metatoul@insa-toulouse.fr](mailto:metatoul@insa-toulouse.fr)

## Explorer les pathophysiologies in vivo

La compréhension du vivant et des dysfonctionnements physiologiques qui peuvent conduire à des pathologies, un enjeu majeur en santé publique, repose sur le développement de modèles animaux adaptés et d'équipement permettant une exploration fonctionnelle précise in vivo. Au cours de ces dernières années, la plateforme Anexplo a participé au développement de nouvelles stratégies innovantes pour le développement d'animaux génétiquement modifiés et a contribué à la mise en place d'équipements de pointe permettant l'exploration fonctionnelle non-invasive d'animaux et l'expérimentation dans des environnements de niveaux de sécurité 1, 2 et 3.

### Recherche et développement

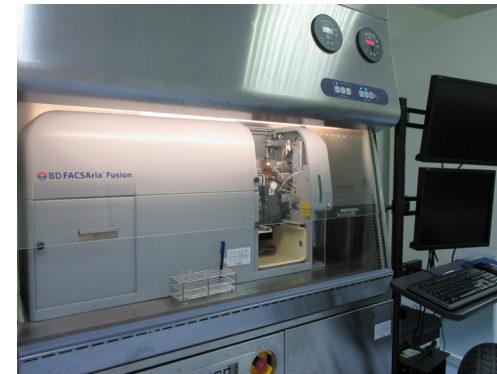
- En concertation avec Celphedia, le développement de la technologie CRISPR/Cas9 sur le poisson zèbre et la souris a été entrepris par la plateforme de transgénèse de l'US006;
- La possibilité de conserver et revivifier des souches de souris et rat génétiquement modifiés est un enjeu majeur. La plateforme de cryoconservation/décontamination a développé avec succès la congélation d'embryon de souris et de rat et la FIV chez la souris.



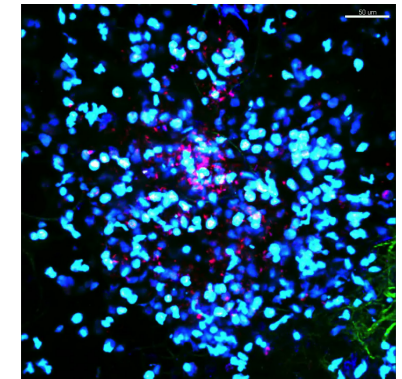
Poisson zèbre génétiquement modifié pour exprimer la protéine fluorescente verte GFP de manière ubiquitaire (Service transgénèse US006)

### Équipement et aménagement des zones de stabulation et d'expérimentation

- La compréhension des interactions hôtes/pathogènes dans des modèles intégrés est un enjeu majeur, nécessaire au développement de nouveaux traitements des maladies infectieuses émergentes ou ré-émergentes. Pour permettre le développement de projets sur ces thématiques, les zones de confinement et d'expérimentation de niveau de sécurité 2 et 3 pour des souris, rats, ovins et bovins ont été développées dans les différents sites d'expérimentation animale rattachés à Anexplo. Outre la possibilité de stabuler des animaux en milieu confiné et protégé, ces zones intègrent des équipements permettant une expérimentation directe sur les animaux.
  - IPBS : stabulation de niveau de sécurité 3 de souris adossée à un laboratoire P3 équipé de cytomètre, trieur de cellules et microscope confocal/biphoton;
  - CPTP : stabulation de niveau de sécurité 2 de souris et rats adossée à un laboratoire équipé d'un système IVIS permettant l'analyse de bioluminescence et fluorescence in vivo sur animaux vigiles et d'un laboratoire d'étude comportementale.
- Le suivi clinique de modèle animaux de différentes pathologies est essentiel à la compréhension de ces pathologies. Pour répondre à ce besoin, les plateaux techniques d'Anexplo ont acquis au cours de ces dernières années des équipements permettant une analyse non-invasive par doppler, échographie, IRM du petit animal. Plus récemment, le site de l'oncopôle de l'US006 a acquis un IRM7/SRM Tesla qui complète les SPECT/CT et TEP/CT pour l'imagerie multimodale non invasive du petit animal.



Trieur de cellules et cytomètre en flux implanté dans un environnement de niveau de sécurité 3 (A. Peixoto, IPBS)

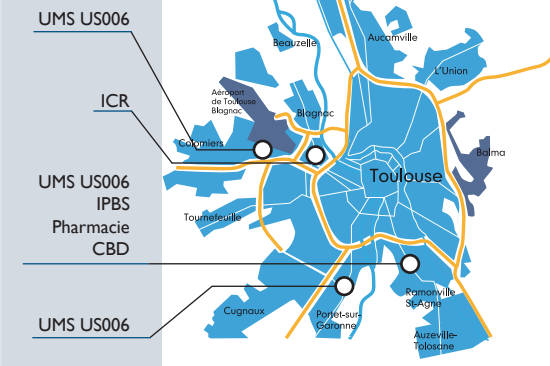


Imagerie intravitale de la formation d'une lésion granulomateuse dans le poumon d'une souris infectée par *Mycobacterium tuberculosis* (A. Peixoto, IPBS)

## PUBLICATIONS

- Troegeler A, Mercier I, Cougoule C, Pietretti D, Colom A, Duval C, Vu Manh TP, Capilla F, Poincloux R, Pingris K, Nigou J, Rademann J, Dalod M, Verreck FA, Al Saati T, Lugo-Villarino G, Lepenies B, Hudrisier D, Neyrolles O. C-type lectin receptor DCIR modulates immunity to tuberculosis by sustaining type I interferon signaling in dendritic cells. *Proc Natl Acad Sci U S A.* 2017 114(4):E540-E549
- Jungas T, Percy RT, Fawal M, Callot C, Froment C, Bulet-Schiltz O, Besson A, Davy A. Eph-mediated tyrosine phosphorylation of citron kinase controls abscission. *J Cell Biol.* 2016 214(5):555-69
- Blasco-Baque V, Garidou L, Pomié C, Escoula Q, Loubieres P, Le Gall-David S, Lemaitre M, Nicolas S, Klopp P, Waget A, Azalbert V, Colom A, Bonnaure-Mallet M, Kemoun P, Serino M, Burcelin R. Periodontitis induced by *Porphyromonas gingivalis* drives periodontal microbiota dysbiosis and insulin resistance via an impaired adaptive immune response. *Gut.* pii: gutjnl -2015-309897
- Aloulou M, Carr EJ, Gador M, Bignon A, Liblau RS, Fazilleau N, Linterman MA. Follicular regulatory T cells can be specific for the immunizing antigen and derive from naive T cells. *Nat Commun.* 2016 7:10579
- Larrue C, Saland E, Boutzen H, Vergez F, David M, Joffre C, Hospital MA, Tamburini J, Delabesse E, Manenti S, Sarry JE, Récher C. Proteasome inhibitors induce FLT3-ITD degradation through autophagy in AML cells. *Blood.* 2016 127(7):882-92

### Localisation des équipements



### Responsables scientifiques

Sylvie Guerder, Olivier Neyrolles, Eric Oswald

### Responsables opérationnels de sites

Yara Barreira, Perrine Bonneville, Daniel Cussac, Magali Jacquier, François Schelcher, Pascal Roulet

### Contacts

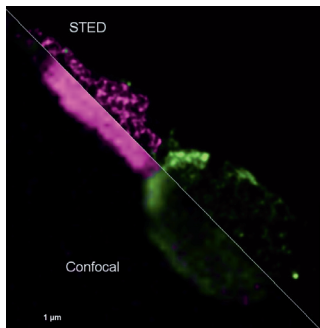
<http://anexplo.genotoul.fr/>

## À la pointe de la résolution

Les lois de l'optique, notamment celles de la diffraction de la lumière, limitent les résolutions latérale et axiale du microscope à des valeurs, respectivement, de 200 et 600 nm, ce qui ne permet pas de visualiser les composants cellulaires les plus fins. Afin de franchir cette barrière et atteindre de plus hautes résolutions, des avancées spectaculaires, regroupées sous les termes de « microscopie de super-résolution » ou « nanoscopie » ont récemment vu le jour. Ces découvertes ont valu à leurs auteurs, E. Betzig, S. Hell et W. Moerner, l'attribution du prix Nobel de Chimie en 2014 (1). Elles reposent sur des principes de type « pointillisme » (PALM/STORM),

### Équipements commerciaux - Sites CPTP, CBI et I2MC

- Microscopes dSTORM et STED au CPTP. Le premier permet des approches de type dSTORM monocouleur en microscopie TIRF, avec un laser 640 nm de 400 mW. L'originalité du second, le STED, réside dans le fait qu'il est le seul microscope de super-résolution permettant de travailler sur coupes de tissus en 3D, permettant ainsi d'accéder à la physiopathologie humaine avec une résolution très fine (30 nm).
- Un équipement similaire, le Nikon-SPT PALM/STORM est opérationnel sur la plateforme LITC du CBI. Il utilise la même technologie pointilliste que le précédent, mais permet des observations dans des longueurs d'ondes plus variées.



Mise en évidence du virus Zika dans les spermatozoïdes humains par microscopie de super-résolution STED, Tom20-StarRed (Mitochondries) et 4G2-Alexa-594 (Virus Zika). Elsa Superbielle, équipe Dunia, CPTP

d'illumination structurée (SIM) ou de déplétion laser (STED). Elles permettent désormais de localiser, in vivo, au sein même des cellules, des composés à l'échelle de la molécule unique (résolution latérale de l'ordre de 10 à 90 nm et axiale jusqu'à 30 nm).

Au cours de ces dernières années, la plateforme TRI a contribué à la mise en place de ces nouvelles approches pour l'ensemble de la communauté scientifique, sous forme de matériels soit proposés par les équipementiers, soit développés par les personnels de la plateforme.

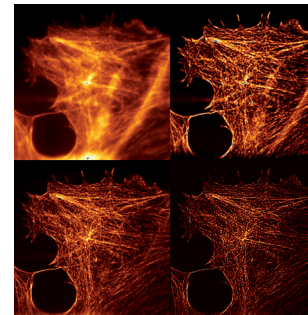
- Un équipement de type ELYRA PSI de la société Zeiss a été implanté sur le site de l'I2MC. Ce microscope de super-résolution utilise un concept d'analyse des nano-interactions au niveau des contacts entre différents types cellulaires ou entre les cellules et la matrice extracellulaire. Grâce à cet équipement, il est possible d'analyser un même échantillon par trois technologies, SIM, STORM et PALM, permettant de localiser avec précision des molécules uniques

### Technologie de R&D - Site CBI (LITC)

Un microscope a été développé intégrant les dernières innovations en imagerie par illumination structurée (SIM) issues de laboratoires de microscopie de renommée internationale (IPHT – Allemagne, Institut Fresnel – France). Elles permettent de s'affranchir des contraintes expérimentales rencontrées dans les équipements classiques de haute-résolution par éclairage structuré. Un soutien technologique provenant des sociétés françaises Oxixius (technologie laser) et AbbeLight (localisation 3D supercritique) sera effectif sur l'année 2016-2017.

Une chaîne d'illumination universelle permet d'explorer différentes techniques d'éclairage structuré conventionnel et non conventionnel sur un même setup. Cette machine est ouverte à l'ensemble de la communauté des biologistes et des physiciens. Les différentes techniques accessibles sont :

- SIM conventionnel TIRF 2D et SIM 3D à une fréquence temporelle très élevées de 20 images/s. Cette technique autorisera ainsi l'imagerie super-résolue de phénomènes rapides tels que la mitose et l'endocytose avec une résolution de 90nm.
- Illumination stochastique de type Speckle permettant d'obtenir une très haute résolution proche d'un microscope confocal parfait (pinhole infiniment petit) pour des échantillon 3D. Très récemment cette approche a permis d'atteindre un gain de résolution d'un facteur 3 avec une nouvelle approche de traitement d'image et d'éclairage stochastique [3]. Cette nouvelle technique autorise l'imagerie super-résolue sur cellule vivante de différentes épaisseurs.



Filaments d'actine images grand champ (haut gauche), Image déconvoluée (haut droite), Image déconvoluée assistée par speckle (bas gauche), super-résolution speckle (bas droite). © CBI – T. Mangeat

## PUBLICATIONS

1. Moreau M, Cochard P. De la microscopie à la nanoscopie : une révolution en résolution. Med Sci (Paris). 2014. 30(12):1169-76.
2. Mansuy JM, Suberbielle E, Chapuy-Regaud S, Mengelle C, Bujan L, Marchou B, Delobel P, Gonzalez-Dunia D, Malnou CE, Izopet J, Martin-Blondel G. Zika virus in semen and spermatozoa. Lancet Infect Dis. 2016 Oct;16(10):1106-7.
3. Labouesse S, Negash A, Idier J, Bourguignon S, Mangeat T, Liu P, Sentenac A, Allain M. Joint reconstruction strategy for structured illumination microscopy with unknown illuminations. IEEE Trans Image Process. 2017 Feb 24. doi:10.1109/TIP.2017.2675200.

### Responsable scientifique

Philippe Cochard

### Responsables opérationnels

Sophie Allart, Cécile Pouzet, Jacques Rouquette

### Contacts

tricontact@genotoul.fr  
http://trigenotoul.com

### Localisation des équipements

Centre de Physiopathologie de Toulouse-Purpan (CPTP)  
INSERM UMR 1043/CNRS UMR 5282 - Université Toulouse III  
Paul Sabatier - CHU Purpan - BP 3028 - 31024 Toulouse Cedex 03

Centre de Biologie Intégrative (CBI) Université Toulouse III  
Paul Sabatier - 118, Route de Narbonne - 31062 TOULOUSE Cedex

Institut des Maladies Métaboliques & Cardiovasculaires (I2MC)  
1 avenue Jean Poulhès - BP 84225 - 31432 Toulouse Cedex 4



## Intégration statistique de données : vers la Biologie des systèmes

Le fait marquant de l'année 2016 pour la plateforme de biostatistique est l'accroissement des sollicitations concernant la thématique de l'intégration statistique de données. En effet, de plus en plus de projets acquièrent plusieurs ensembles de données « omiques » (génomique, transcriptome, métabolome...) sur les mêmes individus ou échantillons pour ensuite mener une analyse intégrative visant à réconcilier les différents niveaux d'observations. Les outils statistiques à déployer face à ces problèmes d'intégration sont encore à développer et à tester. Face à ces demandes, la plateforme a répondu en s'impliquant dans des encadrements de thèses (co-direction ou comité de thèse), en participant au montage de projets ANR et en renouvelant son offre de formation.

Ainsi, durant l'année 2016...

... Valérie Sautron a soutenu son doctorat en biologie des systèmes (École Doctorale Sciences Ecologiques Vétérinaires Agronomiques Bioingénieries) le 27 octobre 2016. La thèse, co-financée par l'ANR (SUSoSTRESS, ANR-12-ADAP-0008) et la région Midi-Pyrénées, a été co-encadrée par deux biologistes (Elena Terenina et Pierre Mormède, unité Génétique Physiologie et Systèmes d'Élevage, GenPhySE, INRA Toulouse) et une statisticienne (Nathalie Villa-Vialaneix, unité de Mathématiques et Informatique Appliquées de Toulouse MIAT, INRA). La thèse a permis l'étude de l'impact de plusieurs types de stress chez le porc d'élevage, avec en vue l'amélioration des qualités de robustesse des animaux pour un meilleur bien être animal dans les élevages. La thèse s'est concentrée sur la prise en compte des aspects longitudinaux des expériences menées dans ce projet, lors desquelles des cinétiques de réponse ont été mesurées pour plusieurs types de données « omiques ». La thèse a permis l'intégration longitudinale de données transcriptomiques et de biologie clinique, ainsi que le développement d'une méthodologie statistique spécifique.

... Harold Duruflé a poursuivi ces travaux de thèse initiés en octobre 2014 sur l'étude de la plasticité de

la paroi d'*Arabidopsis thaliana*. La thèse, co-financée par le PRES Université de Toulouse et la Région Midi-Pyrénées, est co-encadrée par Christophe Dunand (professeur, UPS, Laboratoire de Recherche en Sciences Végétales) et Philippe Besse (professeur, INSA, Institut de Mathématiques de Toulouse). Ce co-encadrement inter-disciplinaire permet d'aborder non seulement l'acquisition de données (transcriptome, protéome, analyse des sucres, climat...) sur des échantillons d'écotypes pyrénéens, mais aussi le traitement statistique de ces données omiques hétérogènes. Ce projet propose un cadre d'application à des méthodologies récemment développées dans le package mixOmics autour de l'analyse PLS multi-block. C'est sur ce thème que le projet a également bénéficié de l'apport d'un étudiant de l'INSA (filière Génie Mathématique et Modélisation) durant un stage de 3 mois.

... Alyssa Imbert a débuté ses travaux de thèse en statistique, co-financée par la société Methodomics I et la région Midi-Pyrénées. La thèse est co-encadrée par une chercheuse biostatisticienne de l'INRA (Nathalie Villa-Vialaneix, unité MIAT, INRA Toulouse) et une chercheuse biologiste de l'INSERM (Nathalie Viguier, Institut des Maladies Métaboliques et Cardiovasculaires, INSERM Toulouse) et porte sur la prise en compte de valeurs manquantes dans l'intégration des données « omiques ». La thèse est motivée par un projet européen multicentrique, DioGenes2, qui tente d'apporter des réponses au problème de société majeur qu'est l'obésité en analysant, par une approche de biologie des systèmes, les effets d'un régime basse calorie sur l'organisme de personnes obèses. Lors de sa première année de thèse en 2016, Alyssa a développé une approche permettant de prendre en compte dans l'inférence de réseaux de gènes à partir d'une expérience transcriptomique longitudinale (RNA-seq) des individus manquants : cette approche permet d'améliorer à la fois la qualité de l'inférence, mais aussi la pertinence de la comparaison longitudinale de l'évolution des relations entre gènes.

### Formations

Ce thème de l'intégration de données a été également présent dans l'offre de formation de la plateforme avec plusieurs sessions consacrées à la librairie mixOmics pour le logiciel R.

- Septembre 2016 : co-organisation de l'école d'été internationale Multivariate data analysis methods for biological data using the R package mixOmics organisée par le réseau EU COST Action 'The quest for tolerant varieties: phenotyping at plant and cellular level' (FAI 306).
- Juillet 2016 : Intégration statistique de données : Méthodes multivariées d'analyse de données biologiques avec le package R. CATI Bios4Bio, INRA Clermont-Ferrand.
- Mars 2016 : Multivariate projection methodologies for the exploration of large biological data sets Application in R using mixOmics. Laboratoire d'Étude des Résidus et Contaminants dans les Aliments (LABERCA), École Nationale Vétérinaire, Agroalimentaire et de l'Alimentation Nantes Atlantique.

La plateforme s'investit aussi dans l'organisation de formations plus classiques, structurées en plusieurs niveaux (de débutant à avancé) :

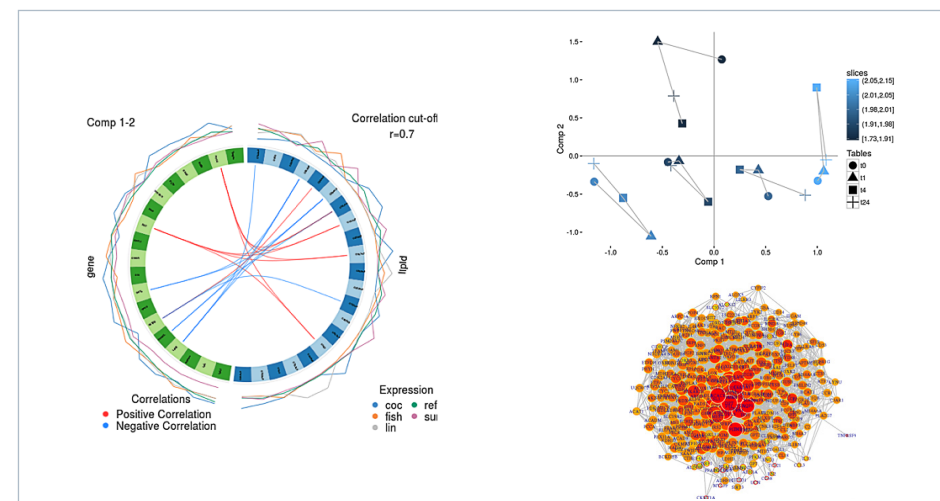
- Avril 2016 : Statistique niveau I : Introduction à la biostatistique

Ainsi que dans la participation à la formation initiale (niveau Master) :

- Novembre 2016 : Master bio-informatique et biologie des systèmes (UPS), animation d'un atelier sur l'analyse de données RNAseq
- Octobre 2016 : Master Bioingénierie (UPS), participation à un atelier génomique.

1 <http://www.methodomics.com> La SARL Methodomics a été créée en septembre 2008 pour répondre à de nouveaux besoins dans l'analyse de données biologiques.

2 <http://www.diogenes-eu.org>



Représentations graphiques issues de méthodes statistiques d'analyse intégrative

### Animateurs

Sébastien Déjean,  
David Rengel,  
Nathalie Villa-Vialaneix

### Contact

biostat@math.univ-toulouse.fr

### Localisation

Université Toulouse III Paul Sabatier  
Institut de Mathématiques de Toulouse – 118, route de Narbonne  
31400 Toulouse

### INRA Occitanie Toulouse

Unité de Mathématiques et Informatique appliquées  
et Laboratoire des Interactions Plantes-Microorganismes.  
Chemin de Borde Rouge – 31326 Castanet-Tolosan cedex

## Un carrefour de compétences sur les enjeux sociétaux des biotechnologies

Une offre diversifiée face à des questionnements et des besoins d'une importance grandissante pour les scientifiques, les institutions et les publics.

Un sujet d'actualité sociétale intense : « L'ingénierie ciblée du génome ou genome editing »

L'année 2016 a été marquée par un ensemble d'activités proposées ou menées par la plateforme SOCIÉTAL portant sur ce sujet d'actualité. Certaines ruptures technologiques font sauter des verrous et repoussent les limites à la fois de nos capacités d'exploration de systèmes biologiques, d'approfondissement des connaissances et d'applications médicales ou dans divers autres domaines. Nous vivons en direct une de ces ruptures qui font reculer nos horizons avec une nouvelle technologie d'ingénierie du génome grâce notamment au système CRISPR/Cas9.

Un **groupe de travail** commun à la Société Française de Génétique Humaine (SFGH) et à la Société Française de Thérapie Cellulaire et Génique (SFTCG) comprenant 4 membres actifs sur la plateforme a contribué au débat sur l'ingénierie ciblée du génome par une Tribune dans Le Monde, journal qui a déjà publié plusieurs contributions sur ce sujet. Ces deux sociétés présidées par un chercheur toulousain ont mis en place un groupe de travail commun sur ce sujet depuis mai 2015 dans lequel la plateforme a joué un rôle moteur. Les propositions de ce groupe visent à promouvoir une recherche responsable, en capacité d'anticiper les conséquences sociétales de ses développements, dans l'objectif d'une protection tant individuelle que collective.

L'**atelier de réflexion éthique 2016** de la plateforme a développé ce sujet sous le titre générique « Modifications ciblées des génomes et enjeux éthiques » avec pour la première fois non pas 3, mais 4 volets d'une après-midi chacun portant respectivement sur le génome humain et les thérapies, le génome animal, le génome de micro-organismes et le génome végétal mobilisant des participants et des intervenants de laboratoires et institutions variées. Les contributions et synthèses seront toutes accessibles sur le site de la plateforme après leur validation.

Des **contributions** diverses en poster, séminaires, conférences ouvertes au public se sont multipliées sur

l'ingénierie ciblée du génome, en sus de contributions à un groupe de réflexion de la Société européenne de génétique humaine, et à une audition par l'Office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST) saisi sur ce domaine, activités faisant ainsi suite à une Table ronde sur cette question réalisée lors d'un colloque organisé principalement par la plateforme en décembre 2015.

Les autres activités de la plateforme en 2016 se sont focalisées sur 2 sujets majeurs qui se poursuivront en 2017 :

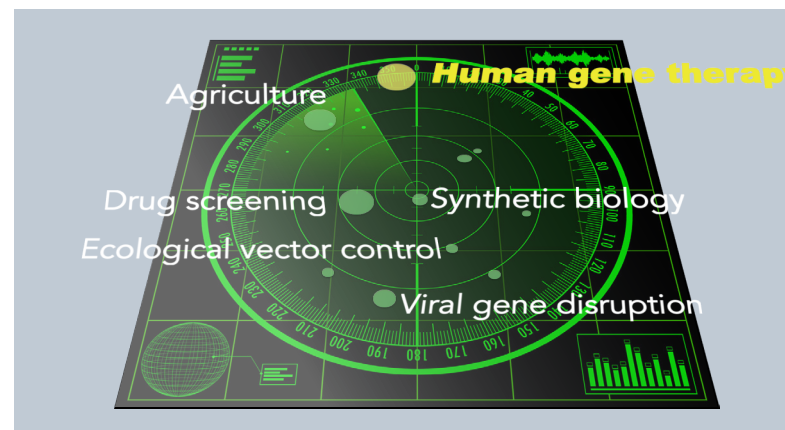
- les biobanques avec en particulier la parution d'un ouvrage sur les aspects éthiques et réglementaires avec l'infrastructure nationale et la contribution aux activités de l'infrastructure européenne BBMRI-ERIC, notamment dans le cadre de la réglementation européenne sur la protection des données (Nouveau Règlement européen paru en mai 2016);
- et les nouvelles technologies de séquençage en génomique humaine en prenant part à la préparation puis au lancement du plan français de médecine génomique 2025.



**Une plateforme précurseur pour la formation à l'éthique et à l'intégrité scientifique**

**Une reconnaissance à souligner!**

Les modules pour les doctorants proposés annuellement par la plateforme sur les enjeux éthiques et sociétaux de la recherche, lancés en 2000 se poursuivent depuis plus de 15 ans, alors que l'année 2016 est marquée par la parution de l'Arrêté du 25 mai 2016 fixant le cadre national de la formation et les modalités conduisant à la délivrance du diplôme national de doctorat, stipule que les écoles doctorales « Veillent à ce que chaque doctorant reçoive une formation à l'éthique de la recherche et à l'intégrité scientifique ».



Domaines d'application potentielle de l'ingénierie ciblée du génome

## PUBLICATIONS

- « Ingénierie du génome : il faut clarifier le cadre réglementaire », Tribune dans « Le Monde », 13 avril 2016, p8 du supplément « Science & médecine » Emmanuelle Rial-Sebbag (correspondante, Plateforme Genotoul Societal), par le Groupe de travail conjoint sur « l'ingénierie ciblée du génome de l'embryon et des cellules germinales : actualités sur les recherches et applications envisagées chez l'homme » des Société française de génétique humaine (présidente : Anne Cambon-Thomsen) et Société française de thérapie cellulaire et génique (président : Pierre Cordelier)
- Chassang G., Cambon-Thomsen A., Rial-Sebbag E. Ethique et Réglementation des biobanques de recherche. (Ouvrage, Edition et publication de BIOBANQUES, Service Éthique et réglementation), 2016
- Julia S., Bertier G., Cambon-Thomsen A. Quand l'anticipation devient plurielle : la complexité des données génomiques à l'épreuve des pratiques professionnelles. Revue Française d'Éthique Appliquée (RFEA) 2016, N° 2, 19-28
- Rial-Sebbag E, Guibet Lafaye C, Simeoni U, Junien C. [DOHaD and epigenetic information: societal challenges]. Med Sci (Paris). 2016 Jan;32(1):100-5.
- Soulier A., Cambon-Thomsen A. Promesses de biobanques : se soucier de l'avenir dans l'éthique de la recherche biomédicale. Revue Française d'Éthique Appliquée (RFEA), 2016, N° 2, 29-47

**Responsable scientifique :** Anne Cambon-Thomsen

**Responsable opérationnel :** Emmanuelle Rial-Sebbag

**Contacts (mail, web, twitter, facebook...)** : Lucie Serre - Lucie S - serres.lucie@gmail.com

**Localisation des équipements :** Faculté de médecine, 37 allées Jules Guesde 31062 Toulouse cedex 9

## Une plateforme de phénotypage microbien et végétal **nouvellement labellisée**

La **plateforme TPMP** est dédiée au phénotypage automatisé à haut débit et en conditions contrôlées d'espèces végétales modèles ou d'intérêt agronomique (à des stades précoces) et des microorganismes qui leur sont associés. Elle permet un suivi non invasif des plantes en milieu et de simuler des conditions climatiques et environnementales précises. En conditions S2 elle est compatible avec l'utilisation des OGMs et les organismes de quarantaine (avec agrément). Aujourd'hui elle comprend :



- Un robot de phénotypage en serre composé de 2 serres autonomes d'une capacité respective de 90 et 270 convoyeurs. Elles sont équipées de stations irrigation et de pesées automatisées ainsi que d'une cabine de spray. Le système est également doté de cabines d'imagerie avec caméras visible, fluorescence et proche infrarouge.
- Un automate de phénotypage en chambre de culture (Phenopsis) doté de caméra visible et à fluorescence de chlorophylle, permettant de suivre la croissance et l'activité photosynthétique (marqueur de santé) de plus de 500 plantes de petite taille.
- 6 chambres climatiques à éclairage LED, entièrement paramétrables. L'homogénéité est contrôlée par des capteurs de température, hygrométrie et lumières comme dans les serres.
- Un système Biolog pour le phénotypage microbien haut débit doté d'une caméra, permettant un criblage par mesure colorimétrique de 50 microplaques de 96 puits en parallèle, soit le suivi de 4800 expériences en simultanée.

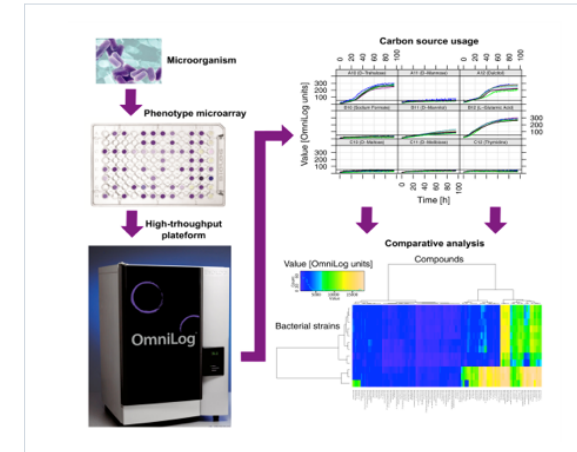


Fig. Schéma d'analyse phénotypage microbien

Le système **Biolog**, installé en 2014, a permis de réaliser des analyses pour 11 projets de recherches, 9 conduits par la communauté toulousaine et 2 pour le national (Paris, Grenoble). Deux études qui ont bénéficié des services de la plateforme ont été publiées récemment dans PLoS Pathogens (Peyraud et al. 2016 et Perrier et al. 2016). La première (Peyraud et al. 2016) a consisté en la caractérisation des capacités métaboliques de *Ralstonia solanacearum*, un pathogène des plantes dévastateur pour les cultures au niveau mondial.

Les analyses réalisées avec le robot OmniLog de la plateforme ont permis de montrer une réduction de la versatilité du pathogène lors de l'activation des fonctions de virulence, résultant en une optimisation du métabolisme sur les substrats présents dans la plante. L'étude qui combine des approches de biologie des systèmes et de phénotypage haut-débit suggère que cette réduction de versatilité peut avoir un impact important sur la dynamique des populations de pathogènes. Cette prédiction est en fait corroborée par la deuxième étude (Perrier et al. 2016). En effet, celle-ci a consisté en la caractérisation avec le robot de phénotypage de la plateforme d'un gène du réseau de régulation de *R. solanacearum*. Ce gène qui était de fonction inconnue précédemment avait été identifié lors d'expérience d'évolution expérimentale. En effet des clones évolués avaient accumulé des mutations adaptatives dans ce gène. L'analyse avec le robot OmniLog a permis de démontrer que ces mutations sont responsables d'un regain de versatilité chez ces populations évoluées comparativement à la souche sauvage initiale.

## PUBLICATIONS

- Peyraud R, Cottret L, Marmiesse L, Gouzy J, Genin S. A Resource Allocation Trade-Off between Virulence and Proliferation Drives Metabolic Versatility in the Plant Pathogen *Ralstonia solanacearum*. PLoS Pathog. 2016 Oct 12
- Perrier A, Peyraud R, Rengel D, Barlet X, Lucasson E, Gouzy J, Peeters N, Genin S, Guidot A. Enhanced in planta Fitness through Adaptive Mutations in EfpR, a Dual Regulator of Virulence and Metabolic Functions in the Plant Pathogen *Ralstonia solanacearum*. PLoS Pathog. 2016 Dec 2

**Responsable scientifique :** Julie Cullimore

**Responsable opérationnel :** Mehdi Khafif

**Contacts :** mehdi.khafif@inra.fr

**Site web :**

<http://www.fraib.fr/Plateformes-technologiques/Phenotypage-microbien-et-vegetal>

**Localisation des équipements :**

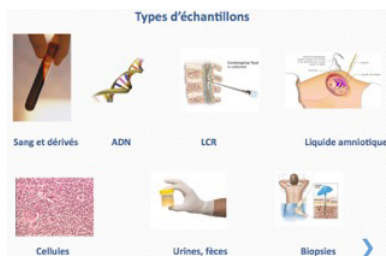
INRA – CR Toulouse, CS 52627, 31326 Castanet-Tolosan



## Un centre de ressources biologiques humaines nouvellement labellisé

Le **CRBh**, centre de ressources biologiques humaines, regroupe les 3 **CRB** mis en œuvre par le CHU de Toulouse : **Germethèque**, **CRB-Cancer** et **Toulouse Bio-Ressources**. La mission des CRB est la préparation, le conditionnement, le stockage et la mise à disposition des ressources biologiques humaines, organisées en grandes collections thématiques. Les ressources biologiques peuvent provenir de deux origines : 1°) les protocoles de recherche biomédicale, et 2°) la filière de soins, où les prélèvements destinés à l'analyse médicale peuvent être secondairement requalifiés pour la recherche. Les CRB opèrent selon un référentiel qualité spécifique (norme NF S96-900), assurant une stricte conformité réglementaire, la thermo-vigilance H24/J7, la gestion informatisée des ressources et données associées. Les spécificités des 3 CRB constitutifs de la plateforme CRBh sont présentées ci-dessous :

- **GERMETHÈQUE** comporte 11 sites répartis sur la France entière. Les ressources biologiques concernent les thématiques de la fertilité, de la procréation et du développement humain. Parmi les ressources spécifiques figurent ovocytes, spermatozoïdes, liquide folliculaire, plasma séminal, tissus germinaux, milieux de culture issus des embryons . . . . GERMETHÈQUE est partenaire de plusieurs programmes de recherche nationaux sur la procréation.
- Le **CRB-Cancer** est dédié à la pathologie tumorale et est situé à l'Institut Universitaire du Cancer de Toulouse – Oncopole(IUCT-O). Le CRB-Cancer dispose de prélèvements tumoraux et non tumoraux associés à la tumeur rassemblés en 12 collections (lymphomes, tumeurs cérébrales, tumeurs mélaniques, tumeurs colorectales, tumeurs du sein, tumeurs urologiques, tumeurs pulmonaires, tumeurs ORL, tumeurs des tissus mous, tumeurs gynécologiques, tumeurs cutanées, tumeurs osseuses et série échantillons tumoraux). Il dispose d'une plateforme d'histopathologie dédiée à la recherche et développe des techniques d'immuno-histochimie, d'hybridation in situ et de « Tissue MicroArray ». Le CRB-Cancer est impliqué dans de nombreux projets académiques et avec des industriels.
- **Toulouse Bio-Ressources** (CRB – TBR) est un CRB multithématique, comportant 40 collections développées autour de 3 axes majeurs : 1°) le vieillissement, cognitif, neurologique, cardiovasculaire, ostéoarticulaire et métabolique; 2°) les maladies du développement et de l'enfant; 3°) les pathologies infectieuses. Le CRB assure la préparation et la gestion de toute une variété d'échantillons biologiques : sang et dérivés, ADN génomique, liquide céphalo-rachidien, urines, fèces, biopsies tissulaires, cellules en culture. Le CRB est également adossé aux plateaux techniques analytiques du laboratoire de biologie médicale (LBM) du CHU de Toulouse. Le CRB – TBR est partenaire de plus de 40 protocoles en cours de recherche biomédicale, de dimension nationale et internationale.



### Projets 2016 – 2017 impliquant la plateforme CRBh :

- AzoPredhistoTEP : comparaison des paramètres du TEP/TDM entre des patients azoospermes ; Germethèque.
- FIVPHENOL : Évaluer l'impact de l'exposition au Bisphenol A sur le taux d'implantation embryonnaire après FIV/ICSI ; Germethèque.
- PARTHOM : la part de l'homme dans les fausses couches à répétition ; Germethèque.
- BACAP : base de données clinico-biologiques sur le cancer du pan créas ; CRB-Cancer.
- ACN : « Au Cœur de la Nuit », programme multicentrique national sur les conséquences métaboliques du travail posté ; CRB – TBR.
- EPAD : « European Prevention of Alzheimer Dementia », programme européen de prévention de la maladie d'Alzheimer impliquant 35 centres ; CRB-TBR.
- NOLAN : « Randomized trial of a nutritional blend to prevent cognitive decline in older adults » ; essai multicentrique européen et américain de prévention du déclin cognitif ; CRB – TBR, laboratoire central.
- RASTAT : « Treatment with HMG-CoA réductase inhibitor of growth and bone abnormalities in children with Noonan syndrome » ; PHRC national (programme hospitalier de recherche clinique) ; CRB – TBR.
- TIRATROP : « Ticagrelor in rotational atherectomy to reduce troponin enhancement », protocole thérapeutique multicentrique ; CRB – TBR.
- RODEO : « relai oral dans le traitement des endocardites multisensibles » ; CRB – TBR.
- Collection biologique VHE : virus de l'hépatite, adossée au Centre de référence national sur l'hépatite E ; CRB – TBR.

## PUBLICATIONS

- Perret B et Razat B, Les CRB du CHU de Toulouse : missions, prestations, retour d'expérience, MED'IN Toulouse de l'Invention au Patient, 22 nov 2016
- Communiqué de presse du CHU de Toulouse, un nouveau marqueur de la maladie coronarienne découvert par des équipes toulousaines, CP du 14 oct 2016
- Genoux A et al., Serum levels of mitochondrial inhibitory factor 1 are independently associated with long-term in coronary artery disease: the GENES study, BMC Med. 2016
- Nguyen MH et al., A study of aneuploidy and DNA fragmentation in spermatozoa of 3 men with sex chromosome mosaicism, Human Fertil 2015
- Camus C et al., Comparison of the effect of semen from HIV-infected and non-infected men on CD4+ T cell infection. AIDS 2016

**Responsable scientifique**  
Bertrand Perret

**Responsable opérationnel**  
Bénédicte Razat

**Contact**  
Bénédicte RAZAT

Tél : 05 67 69 03 35

Email : razat.b@chu-toulouse.fr

**Site web**  
<http://www.chu-toulouse.fr/-centres-de-ressources-biologiques-crb->

### Localisations des équipements

**Germethèque** : CECOS Midi-Pyrénées Toulouse  
Hôpital Paule de Viguière,  
330 avenue de Grande-Bretagne  
31059 Toulouse cedex 9

**CRB – Toulouse Bio-Ressources**  
Institut Fédératif de Biologie, Hôpital Purpan,  
330 avenue de Grande-Bretagne  
31059 Toulouse cedex 9

**CRB – Cancer**  
Institut Universitaire du Cancer de Toulouse – Oncopôle,  
1 avenue Irène Joliot-Curie  
31059 Toulouse cedex 9



**Genotoul**  
GENOPOLE TOULOUSE

**CENTRE DE RECHERCHE INRA  
TOULOUSE**

Chemin de Borde Rouge - CS 52627  
31326 Castanet-Tolosan

Direction  
Luc Pénicaud

Contact  
[contact.genotoul@inra.fr](mailto:contact.genotoul@inra.fr)

**PARTENAIRES**

