PROGRAMME

Assemblée Générale de France Génomique 24-25 Novembre 2015
Hôtel Novotel, 22 rue Voltaire, 94270 LE KREMLIN BICETRE, France

http://www.novotel.com/fr/hotel-5586-novotel-paris-porte-d-italie/index.shtml

Mardi 24 Novembre 2015

12h30-13h30	Accueil des Participants	
13h30 -14h	Introduction, Bilan et Perspectives**, Pierre Le Ber**	*Salle Pasteur*
14h-14h45	Faits marquants Wetlab 2015, **Denis Milan**	*Salle Pasteur*
14h45-15h30	Fait marquants Bioinfo 2015, **Valentin Loux**	*Salle Pasteur*

15h30-16h *Pause café*

16h -18h Atelier/groupe de travail (les 3 ateliers sont en parallèle)

Long Read (salle Pasteur)	**Epigénétique (salle Gobelins)**	**Single Cell - Petite Qté RNA (Salle Massena)**
Animateurs : Cécile Donnadieu et Stéfan Engelen *60 personnes inscrites*	Animateurs : Sophie Chantalat et Nizar Touleimat *22 personnes inscrites*	Animateurs : Marie Agnes Dillies et Pascal Barbry *45 personnes inscrites*

A partir de 18h Installation des posters — *18h à 19h session management I* *Massena ou Gobelin*

19h30 Départ pour le dîner
Le transfert en car sera assuré à compter de 19h30 au départ du Novotel et à destination de Vincennes,

20h Dîner au Chalet du Lac, http://www.chaletdulac.fr/
Le retour vers l'hôtel est prévu vers 22h15/22h30 pour l'intégralité des convives.

Mercredi 25 Novembre

8h00-9h00 *Café d'accueil*

9h- 9h30 Bilan des actions : coordination, animation et communication, **Karine Hugot, Mathilde Clément** *Salle Pasteur*

9h30-10h15 Retour et bilan sur les groupes de travail *Salle Pasteur*

10h15-10h45 *Pause café*

10h45 -11h45 Session résultats Grands Projets de Séquençage FG « Biodiversité » *Salle Pasteur*

10h 45-11h15 Projet : Metagenomics of ancient artic soils, **Chantal Abergel**, Laboratoire Informations génomique et Structurale, Institut de Microbiologie de la Méditerranée, Marseille

11h15-11h45 Projet : 1002 yeast genome, **Joseph Schacherer**, Génétique Moléculaire Génomique Microbiologie, département Microorganismes, Génomes, Environnement, Université de Strasbourg

11h45-13h00 Session poster — *11h45-12h45 session management II* *Massena ou Gobelin*

13h00-14h00 *Pause déjeuner*

14h-15h Session résultats Grands Projets de Séquençage FG « Génétique humaine » *Salle pasteur*

14h-14h30 Projet : Les leucémies myélomonocytaires chroniques, **Eric Solary** , Institut Gustave Roussy, Villejuif

14h30-15h Projet : Myocapture, novel genes for myopathies, **Raphaël Schneider** Département Médecine Translationnelle et Neurogénétique, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg

15h-16h30 Session résultats projets des plateformes France Génomique *Salle Pasteur*

15h-15h30 Apports du séquençage à haut débit dans le rôle des hormones thyroïdiennes lors de la métamorphose des amphibiens. **Laurent Sachs**, Département Régulation, Développement et Diversité Moléculaire, Muséum National d'Histoire Naturelle, Paris

15-30-16h Neuronal identity genes regulated by super-enhancers are preferentially down-regulated in the striatum of Huntington's disease mice. **Karine Mérienne**, Laboratoire de Neurosciences Cognitives et Adaptatives, Université de Strasbourg

16h00 -16h30 Pateamine A-sensitive ribosome profiling reveals the scope of translation in mouse embryonic stem cells **Alexandra Popa**, Plateforme de Génomique Fonctionnelle, Institut de Pharmacologie Moléculaire et Cellulaire, Sophia Antipolis

16h30-17h00 Discussion générale, conclusion, **Pierre Le Ber** *Salle Pasteur*

Résumé des posters

Plateforme de génomique fonctionnelle Nice

N°1 Development of a cost effective single cell RNA seq approach

Marie-Jeanne Arguel, Kevin Lebrigand, Agnès Paquet, Laure-Emmanuelle Zaragosi, Rainer Waldmann, Pascal Barbry

The standard protocol for single cell transcriptome profiling with the Fluidigm C1 96 cell chip requires off chip barcoding and library preparation of 96 samples what is both labor intensive and costly. Here we developed a method for single cell gene expression profiling that uses on chip indexing for reagent and time savings. We add the sequencing indexes during cDNA synthesis on the microfluidic device. This allows us to pool the amplified cDNAs from the 96 cells and to prepare just one library instead of the 96 libraries required by the standard protocol. Our approach sequences preferentially the 5' ends of transcripts and provides also information on transcript start sites. To address amplification bias that is inherent to ultra-low input approaches, we introduce short random sequences ("unique molecule identifiers", UMIs) during cDNA synthesis that allow molecule counting instead of just a simple read counting.

N°2 Comparative genomic analysis of Drechmeria coniospora reveals core and specific genetic requirements for fungal endoparisitism of nematodes

Kevin Lebrigand, Le D. He, Marie-Jeanne Arguel, Nishant Thakur, Jérôme Gouzy, Bernard Henrissat, Eric Record, Ghislaine Magdelenat, Valérie Barbe, Sylvain Raffaele, Pascal Barbry and Jonathan J. Ewbank.

Drechmeria coniospora is an obligate fungal parasite that infects nematodes via the adhesion of specialized spores to the host cuticle. D. coniospora is frequently found associated with Caenorhabditis elegans in environmental samples. It has been developed as a model for the study of the host response to fungal infection. Full understanding of this bipartite interaction requires knowledge of the pathogen's genome, analysis of its gene expression program and a capacity for genetic engineering. We report here the acquisition of all three. We undertook a phylogenetic analysis that placed D. coniospora close to the entomopathogenic Metarhizium fungi, which are characterized by a broad host range, and Metacordyceps chlamydosporia, a facultative nematophagous fungus. Ascomycete nematopathogenicity is polyphyletic; D. coniospora represents a branch that has not been molecularly characterized. A detailed in silico functional analysis, comparing D. coniospora to 9 fungal species, revealed genes and gene family potentially involved in virulence and showed it to be a highly specialized pathogen. A targeted comparison with nematophagous fungi highlighted D. coniospora-specific genes and a core set of genes associated with nematode parasitism. We conducted a comparative gene expression analysis of fungal spores and mycelia, and also identified genes specifically expressed during infection of C. elegans, giving a molecular genetic view of the different stages of the D. coniospora lifecycle. Further, we present methods for the transformation of D. coniospora, allowing gene knock-out and the production of fungus that expresses GFP. Our high-quality annotated genome for D. coniospora gives insights into the evolution and virulence of nematode-destroying fungi. Coupled with genetic transformation, it opens the way for molecular dissection of D. coniospora physiology, and will allow both sides of the interaction between D. coniospora and C. elegans, as well as the evolutionary arms race that exists between pathogen and host, to be studied.

N°3 Pateamine A-sensitive ribosome profiling reveals the scope of translation in mouse embryonic stem cells

Alexandra Popa, Kevin Lebrigand, Pascal Barbry , Rainer Waldmann

Background. Open reading frames are common in long noncoding RNAs (lncRNAs) and 5'UTRs of protein coding transcripts (uORFs). The question of whether those ORFs are translated was recently addressed by several groups using ribosome profiling. Most of those studies concluded that certain lncRNAs and uORFs are translated, essentially based on computational analysis of ribosome footprints. However, major discrepancies remain on the scope of translation and the translational status of individual ORFs. In consequence, further criteria are required to reliably identify translated ORFs from ribosome profiling data.

Results. We examined the effect of the translation inhibitors pateamine A, harringtonine and puromycine on murine ES cell ribosome footprints. We found that pateamine A, a drug that targets Eif4A, allows a far more accurate identification of translated sequences than previously used drugs and computational scoring schemes. Our data show that at least one third but less than two thirds of ES cell lncRNAs are translated. We also identified translated uORFs in hundreds of annotated coding transcripts including key pluripotency transcripts, such as dicer, lin28, trim71, and ctcf.

Conclusion. Pateamine A inhibition data clearly increase the precision of the detection of translated ORFs in ribosome profiling experiments. Our data show that translation of lncRNAs and uORFs in murine ES cells is rather common although less pervasive than previously suggested. The observation of translated uORFs in several key pluripotency transcripts suggests that translational regulation by uORFs might be part of the network that defines mammalian stem cell identity.

TGML Marseille

N°4 Stop looking for a needle in a haystack : how to cath it without fire !

Nicolas Fernandez (a,b), Fabrice Lopez (a,b), Béatrice Loriod (a,b), Laurent Vanhille (b), Lan Dao (b), Phillipe Naquet (c), Salvatore Spicuglia (b)

(a) Plateforme TGML
(b) UMR1090 TAGC
(c) CIML

La plateforme TGML est souvent confrontée à des questions exploratoires de la part des chercheurs avec qui elle collabore, pour lesquelles seule une petite partie du génome est concernée. Ainsi, la recherche de variants exomiques, la détermination de sites d'insertion de plasmides dans des lignées cellulaires transgéniques, ou bien l'étude d'une région précise du génome sont autant de situations pour lesquelles le séquençage après enrichissement s'avère être une stratégie intéressante, tant sur le plan scientifique que d'un point de vue économique. Le fait de ne séquencer qu'une zone restreinte permet de diminuer drastiquement le bruit dû à des alignements non spécifiques, tout en assurant une profondeur de séquençage importante avec un faible nombre de reads. Ce dernier point nous permet de traiter un plus grand nombre d'échantillons par run, ce qui entraine un coût par échantillon réduit, et donc la possibilité pour les chercheurs de traiter un plus grand nombre d'échantillons à budget constant. Cependant, les technologies de séquençage ciblé impliquent la synthèse d'amorces ou de sondes de capture en grand nombre afin d'être rentables, limitant ces approches aux moyennes et grandes études allant de quelques dizaines à plusieurs milliers d'échantillons.
Une étape particulièrement cruciale lors d'un reséquençage ciblé est le design des sondes de capture. La diversité scientifique des projets traités sur la plateforme TGML implique beaucoup de souplesse dans la réalisation de ce design, avec une étude à façon quasi systématique. Nous utilisons pour cela la technologie "Agilent SureSelect DNA Capture Array" dont nous concevons les sondes sur mesure grâce à l'application web eArrays (Agilent). Les microarrays personnalisés haute-fidélité sont ensuite synthétisés en utilisant la plate-forme de fabrication SurePrint (Agilent). Nous vous présentons ici 4 exemples de projets réalisés en 2015 au sein de la plateforme TGML et impliquant chacun l'utilisation de cette technologie d'enrichissement par capture : la recherche d'insertion dans le génome d'une souris transgénique, l'étude de lincRNA, la mise au point du protocole CapStarr-seq et l'étude des réagencements dans les régions VDJ.

N°5 Activité générale de la plateforme TGML

Fabrice Lopez, Nicolas Fernandez

N°6 Développement de workflows pour l'analyse ChIP-seq sous Snakemake et déploiement d'une machine virtuelle sur le cloud IFB.

Claire Rioualen [1,2], Lucie Khamvongsa [1,2], Christophe Blanchet [3], Jacques van Helden [1,2]

1. INSERM, U1090 TAGC, Marseille F-13288, France.
2. Aix-Marseille Université, U1090 TAGC, Marseille F-13288, France.
3. CNRS, UMS 3601, Institut Français de Bioinformatique, IFB-core, Gif-sur-Yvette F-91190 , France.

Le workpackage 2.6 de France Génomique s'articule autour de la régulation de l'expression des gènes, et s'inscrit dans une volonté d'encourager les efforts collectifs dans l'évalutation des outils existants, le développement de nouveaux outils et pipelines ainsi que la maintenance des ressources et leur mise à disposition pour le public. Ceci doit permettre, à terme, d'éviter la duplication des travaux et de capitaliser les connaissances et bonnes pratiques acquises ce faisant.

Snakemake est une librairie basée sur le langage python, permettant de construire des workflows et utilisant la logique des règles et des cibles de GNU make. Elle permet de chaîner des opérations, de les paralléliser ou encore de réaliser du benchmarking, et ceci en utilisant les langages python, R ou shell.

Dans le cadre du WP2.6, nous avons développé un catalogue de règles collaboratif permettant de construire, brique par brique, des workflows ChIP-seq personnalisés afin de répondre à différentes questions biologiques. Celui-ci inclut des opérations allant du contrôle qualité au peak-calling, en passant par différents algorithmes de mapping. Une recherche de motifs peut également être effectuée via la suite logicielle RSAT implémentée. Le workflow a d'ores et déjà été appliqué à des cas d'étude variés tels que bactéries, levures ou drosophile.

Ce travail collaboratif a donné lieu à un projet git hébergé le service SourceSup/Renater, et comprend également un catalogue de règles adaptées au RNA-seq, avec l'ojectif de permettre des analyses intégrées.
Un manuel est en cours de rédaction pour faciliter l'installation et l'utilisation de Snakemake. Ceci permettra également de favoriser l'analyse en production et le benchmarking d'outils de peak-calling. Enfin une machine virtuelle, actuellement développée sur le cloud de l'Institut Français de Bioinformatique (IFB), permettra dès son déploiement l'usage des workflows par tout utilisateur en garantissant leur portabilité.

URGI Versailles

N°7 Benchmarks for transcript assembly and differential gene expression (new isoforms) analysis in the context of transposable elements

T. ALAEITABAR [1], N. Francillonne[1], M. LOAEC [1], H. QUESNEVILLE[1], J. AMSELEM[1]

[1] INRA, UR1164 URGI - Research Unit in Genomics-Info, INRA de Versailles, Route de Saint-Cyr, Versailles, 78026, France

The existence of multiple TSS for a gene is a key event to create diversity and flexibility in the regulation of gene expression under differential conditions (biotic or abiotic). The genomes of most eukaryotes are composed of transposable elements (TEs). TEs are also reported to be carrier of significant signals for the initiation of RNA synthesis and processing. Thus, the presence/absence of TE near the 5' region of a gene may have a role in creation of new TSS leading to the expression of a new isoform. In this context we develop a pipeline to analyze RNA-seq data in the context of gene expression associated with structural changes (transcript isoforms) in different conditions, especially the study of hypothesis that TEs inserted in the vicinity of genes may affect structural changes of transcripts. In the last few years, a number of transcriptome assemblers have been developed, but the real challenge is to choose one of the existing assemblers that perform well enough for all data with different transcriptome complexities.

We will present here preliminary results of our benchmarking to compare three different RNA assemblers, Cufflinks, Trinity and Grit. Then, we will also discuss about some cases studies of relation between TE and new TSS depending on the experimental condition tested.

Institut Curie Paris

N°8 The ICGex Next Generation Sequencing Platform : New Developments in Molecular Characterization of Tumors

Sylvain Baulande, Virginie Bernard, Mylène Bohec, Sonia Lameiras, Patricia Legoix-Né, Virginie Raynal, Elisabeth Hess, Alain Nicolas and Olivier Delattre

By decreasing the cost and increasing the throughput, Next-Generation Sequencing (NGS) has profoundly transformed genomic research. Whole genome and exome sequencing, transcriptome analysis, DNA methylation or other epigenetics studies can be efficiently performed, providing priceless information for researchers. In addition, NGS is also entering into the clinics for molecular diagnosis. The NGS platform of the Institut Curie is dedicated to support research teams involved in multiple aspects of basic, translational and clinical research. The flexibility of our facility allows to provide a wide range of genomic solutions to study molecular biology of normal and cancer cells. Our close relationship with the clinics and the growing need for molecular characterization of tumors in personalized medicine imply the development of innovating approaches. Sequencing of tumors DNA from clinical samples is routinely performed in our platform to generate molecular reports for clinicians in patient care but also in the context of clinical trials. Among on going developments, we are improving throughput of cancer panel sequencing for retrospective studies and also plan to implement whole exome sequencing (WES) on FFPE samples. Recently, we started a collaboration with “Cambridge Epigenetix”, a biotech developing solutions in epigenetics that will be used to study DNA modifications in tumor genomes (methylation, hydroxymethylation …). We also dispose of a C1 system allowing to focus on single cell genomics, an important way to address tumorigenic mechanisms at a finer resolution, bypassing issues due to the heterogeneous nature of tumors.

N°9 Personalized, Mechanistic and Functional Signatures of Human Papilloma Virus in Cervical Carcinomas

Allyson Holmes[1], Sonia Lameiras[1], Emmanuelle Jeannot[2], Yannick Marie[3], Laurent Castera[4], Xavier Sastre-Garau[2] and Alain Nicolas[1]

[1] Recombination and Genetic Instability, Institut Curie Centre de Recherche, CNRS UMR 3244, Université Pierre et Marie Curie, 75248 Paris Cedex 05, France
[2] Department of Biopathology, Institut Curie Hôpital, 75248 Paris Cedex 05, France
[3] Genotyping and Sequencing Platform, Institut du Cerveau et de la Moelle épinière (ICM), Pitié-Salpêtrière Hôpital, Paris Cedex 13, France
[4] Department of Genetics, Centre François Baclesse, 14076 Caen, France

To identify new personal biomarkers for the improved diagnosis, prognosis and biological follow-up of human papillomavirus (HPV)-associated carcinomas, we developed a generic and comprehensive Capture-HPV method followed by Next Generation Sequencing (NGS). Starting from biopsies or circulating DNA samples, our double-Capture NGS approach rapidly identifies the HPV genotype, HPV status (integrated, episomal or absence), the viral-host DNA junctions and the associated genome rearrangements (A. Holmes *et al.*, submitted). Thus, our patient-to-patient analysis of 46 cervical carcinomas identified five HPV signatures. The first two signatures contain two hybrid chromosomal-HPV junctions whose orientations are co-linear (2J-COL) or non-linear (2J-NL), revealing two modes of viral integration associated with chromosomal deletion or amplification events, respectively. The third and fourth signatures exhibit 3-12 hybrid junctions, either clustered in one locus (MJ-CL) or scattered at distinct loci (MJ-SC) while the fifth signature consists of episomal HPV genomes (EPI). Similar pathological cervical carcinomas, without HPV sequences are also identified. Cross analyses between the HPV signatures and the clinical and virological data outline the relevance of this new classification according to the HPV genotype, patient age and disease outcome. Overall, our findings establish a rational and cost effective approach for the molecular detection of HPV-associated cervical, anal and head-and-neck carcinoma from biopsies or ctDNA and provide ultimate sequence information to develop sensitive and specific biomarkers for each patient.

N°10 NGS activities at the Institut Curie Bioinformatics Platform

Alban Lermine, Nicolas Servant, Jocelyn Brayet, Vivien Deshaies, Mandy Cadix, Elodie Girard, Aurélie Teissandier, Ivaylo Vassilev, Windy Rondof, Emmanuel Barillot, Philippe Hupé

Institut Curie's NGS bioinformatics platform main missions consist in the bioinformatic support for the NGS sequencing platform activities, the data-management of all NGS data generated by Institut Curie's collaborators, the pre-processing of the data, their quality controls and the generation of run reports. We offer a collaborative support for NGS data analyses, may developed/are able to develop new tools to respond to new biological challenges. We also propose internal and external trainings (training courses) for both bioinformaticians and biologists. The IT group support data management, storage and secured data availability. Moreover, a data quality check is realized as well as a standard alignment on reference genome. Preliminary analysis are summarized into a report transferred to the sequencing platform and to the project leader. The Analyses group handles downstream analyses for a large variety of NGS applications (DNA-seq, RNA-seq, mRNA, Hi-C, ChIP-seq, WGBS, etc ...). In this aim, tools and pipelines have been developed in close collaboration with research and/or medical groups to respond initially to biological needs/queries that are unsatisfied. All of them have been published (or are submitted) to international journals and are made available to the scientific community through widely used framework such as Galaxy or Docker, as well as academic computing infrastructure (IFB cloud). Moreoover, our HPC works on the optimisation of NGS pipeline to speed-up comptation with computing cluster architecture with parallel paradigms (MPI, HTC,...). We also propose publics complete analyses solution available for anyone (tools and computing ressources), under two Galaxy servers, http://nebula.curie.fr, dedicated to ChIP-seq data analyses and https://galaxy-public.curie.fr for all type of NGS data.

N°11 Advances in bioinformatics analysis of sRNA-seq, ChIP-seq and Hi-C sequencing data

Jocelyn Brayet, Céline Hernandez, Claire Rioualen, Alban Lermine, Jacques Van Helden, Morgane Thomas-Chollier and Nicolas Servant

The France Genomique WP2.6 is dedicated to regulation analysis and therefore concerns several types of high-throughput sequencing data such as sRNA-seq, whole-genome bisulfite or ChIP-seq analysis. Here, we will present our recent advances for sRNA-seq, ChIP-seq and Hi-C data analysis. The ncPRO-seq pipeline for small RNA analysis was updated and packaged to ease its sharing among partners. It is now available through the Galaxy web interface, the IFB Cloud or as a Docker file application. More recently, we focus our efforts on the development of our ChIP-seq Galaxy instance, Nebula. The Nebula instance was therefore updated with the RSAT tools and a new appliance is under construction on the IFB Cloud. Finally, we also developed a new pipeline for Hi-C data processing, named HiC-Pro. HiC-Pro is currently one of the most efficient, complete and flexible solution for Hi-C data processing. The first version is available as a command-line pipeline.

N°12 Analysis of Whole Genome Sequencing with MPI on HPC Architecture

Frederic Jarlier

In this communication we present recent works the Institut Curie has undertook in the field of whole genome analysis. After an extensive study of existing NGS pipelines we propose an approach, grounded on high performance computing technics, to leverage problems of scalabilities, memory and storage. We focus our development on early stages of analysis: the alignment, the sorting and the discovering of structural variations. To accelerate the pace of analysis we combine parallelism (MPI) and cut-edge algorithms (Modified Bruck, Doubling recursion,...) to produce significant speed-up and computing efficiency.

MGX Montpellier

N°13 Genome-wide DNA methylation profiling by reduced representation bisulfite sequencing

Maurine BONABAUD, Emeric DUBOIS, Samia GUENDOUZ, Hugues PARRINELLO, Stéphanie RIALLE, Marine ROHMER, Dany SEVERAC, Sophie VIVIER and Laurent JOURNOT

DNA methylation is a chemical modification of cytosine bases that is pivotal for gene regulation, cellular specification and cancer development. Reduced Representation Bisulfite Sequencing (RRBS) is a common technique for measuring DNA methylation. This technique combines restriction enzymes and bisulfite sequencing in order to enrich for the areas of the genome that have a high CpG content. The MGX core facility proposes a complete pipeline, from the library construction to the annotation of differentially methylated cytosines (DMC) or regions (DMR).

IG CNG Evry

N°14 From Exome to Whole Genome bio-informatic pipeline: Evolution of Varscan.

Florian SANDRON Lilia MESROB Ghislain SEPTIER Stéphane MESLAGE Aurélie LEDUC Delphine BACQ Nicolas WIART Vincent MEYER François ARTIGUENAVE

Au cours des ces dernières années le profil des projets de reséquençage du génome humain a évolué de manière drastique. Le tournant vers l'avènement des projets de reséquençage de génome entier est aujourd'hui amorcé et accompagné de nombreux challenges techniques et scientifiques. Au sein de ce poster nous présentons un bilan technique associé à cette transition ainsi que les outils développés nous permettons d'envisager l'analyse à haut débit des données issues des dernières générations de séquenceurs.

N°15 Evaluation & setting of a BS-seq data pipeline.

Xavier Benigni, Nizar Touleimat, François Artiguenave

Le laboratoire de Bioinformatique du CNG a été missionné par le Consortium France Génomique pour répondre au besoin d'outils bioinformatiques pour le traitement de données BS-seq, dans le cadre du workpage 2.6.3 (WP2.6.3). Au terme de 12 mois de contrat, avec l'aide de l'ensemble des participants du WP2.6.3 nous avons pu recenser les principaux outils de traitement de données BS-seq existant et en effectuer une première évaluation. Cela a permis de sélectionner deux outils d'intérêt : BS-seeker et Bismark dont le laboratoire de Bioinformatique du CNG a évalué les performances de façon plus approfondie sur différents jeux de données BS-seq artificiels et réels. Nous avons appréhendé à la fois des performances informatiques (rapidité, consommation mémoire) et des performances bioinformatiques (contrôle de la qualité des données, qualité de l'alignement des fragments séquencés, etc.) Nous avons démontré à l'issue de cette évaluation que le pipeline Bismark présentait, dans le cadre de nos infrastructures, les meilleures performances et pouvait facilement être implémenté et utilisé au sein de celles du CCRT. Nous avons intégré le pipeline Bismark au sein d'un pipeline global avec l'ajout d'étapes de contrôle qualité et de filtrage des données de séquençage. Notre apport majeur a été de proposer une implémentation d'un pipeline BS-seq permettant de paralléliser son exécution sur les nœuds de calcul du CCRT. Cette parallélisation se base sur un découpage structuré des données pour les traiter de façon indépendante celles qui le sont, puis de les fusionner pour l'analyse finale. Cette procédure offre un gain de temps pouvant atteindre l'équivalent de deux unités logarithmiques en fonction de la disponibilité des ressources du CCRT. Le pipeline BS-seq implémenté au CCRT est dès maintenant accessible à l'ensemble du Consortium France Génomique.

N°16 Combining METagenomic And GENomics in severe OBesity

CNG: Centre national de genotypage: Béatrice Segurens, Jean-Francois Deleuze
ICAN: Institut of Cardiometabolism and nutrition: Karine Clément, Edi Prifti, Jean Daniel Zucker
MGP/ MetaGenoPolis: Emmanuelle le Chatelier, Dusko Ehrlich, Joel Doré

The obesity epidemic of the late 20th and 21st centuries remains one of the most prominent life risks in today's world. From 2005, the genome wide scan association studies (GWAS) allowed to identify and confirm many loci contributing to obesities (including severe and morbid obesities) and related phenotypes in large populations.

The human gut microbiome has also been linked to overweight and obesity in humans and mice (Le Chatelier et al., Nature, 2013).

Up to date, no studies have explored the interactions between host genomic variability and that of our gut microbiome in the context of severe obesity disease, which is highly dependent on environmental conditions and partly heritable at birth.

We aim to test the hypothesis according to which specific associations between microbiome and genome diversities could form a combined molecular basis of obesity. To test this hypothesis we propose to exome-sequence 400 severely obese (BMI>35kg/m2) patients and 100 lean controls benefiting from an existing project (EU-MetaCardis) in which 80% of participants have their gut metagenomes currently in sequencing. Our consortium (CNG, ICAN, MGP) has the right set of expertise (clinical, genome, metagenome, data analysis and integration) to conduct the **project C-METAGENOB**.

N°17 RNA design : méthodologie pour un comparatif des approches d'étude du transcriptome (de la qPCR au RNAseq en passant par les microarrays)

Derbois C., Palomares M-A., Battail C., Deleuze J-F., Olaso R.

N°18 RNA design :Microarray

Derbois C., Palomares M-A., Battail C., Deleuze J-F., Olaso R.

N°19 RNA design : RNAseq

Palomares M-A., Derbois C., Battail C., Deleuze J-F., Olaso R.

L'étude de l'expression des gènes est une étape essentielle dans la compréhension des mécanismes physiologiques et patho-physiologiques. Les approches pour étudier le transcriptome se sont diversifiées ces dernières décennies (de l'ancien Northern Blot au récent séquençage de l'ARN).
Afin d'appréhender au plus près, les avantages et les limites des méthodes actuelles, le CNG mène un projet comparatif.
Utilisant le même matériel de départ, l'objectif est de comparer les données obtenues à partir de puces à ADN (ou microarray) et de séquençage ARN (ou RNAseq). Nous présentons un triptyque qui décline l'étude sur 3 posters.
Le poster central présente la méthodologie employée, qui permet d'établir un set d'ARN de départ qui peut être reproduit dès que besoin, ainsi que les approches technologiques utilisées (Microarray, RNAseq et qPCR).
Sont associés à ce poster central, deux autres posters. L'un présente les premiers résultats obtenus à partir des expériences de Microarray ; l'autre, les résultats obtenus à partir des expériences de RNAseq.
Le travail se poursuit afin de mettre à disposition de la "communauté France Génomique" l'ensemble des éléments permettant un choix éclairé de technologie lors d'un projet d'étude du transcriptome.

IG Genoscope Evry

N°20 Genome assembly using Nanopore-guided Long ans Error_free DNA reads

Mohammed-Amin Madoui1, Stefan Engelen,CorinneCruaud, Caroline Belser, Laurie Bertrand, Adriana Alberti ,Arnaud Lemainque, Patrick Wincker and Jean-Marc Aury

Background: Long-read sequencing technologies were launched a few years ago, and in contrast with short-read sequencing technologies, they offered a promise of solving assembly problems for large and complex genomes. Moreover by providing long-range information, it could also solve haplotype phasing. However, existing long-read technologies still have several limitations that complicate their use for most research laboratories, as well as in large and/or complex genome projects. In 2014, Oxford Nanopore released the MinION® device, a small and low-cost single-molecule nanopore sequencer, which offers the possibility of sequencing long DNA fragments. Results: The assembly of long reads generated using the Oxford Nanopore MinION® instrument is challenging as existing assemblers were not implemented to deal with long reads exhibiting close to 30% of errors. Here, we presented a hybrid approach developed to take advantage of data generated using MinION® device. We sequenced a well-known bacterium,Acinetobacter baylyi ADP1 and applied our method to obtain a highly contiguous (one single contig) and accurate genome assembly even in repetitive regions, in contrast to an Illumina-only assembly. Our hybrid strategy was able to generate NaS (Nanopore Synthetic-long) reads up to 60 kb that aligned entirely and with no error to the reference genome and that spanned highly conserved repetitive regions. The average accuracy of NaS reads reached 99.99% without losing the initial size of the input MinION® reads. Conclusions: We described NaS tool, a hybrid approach allowing the sequencing of microbial genomes using the MinION® device. Our method, based ideally on 20x and 50x of NaS and Illumina reads respectively, provides an efficient and cost-effective way of sequencing microbial or small eukaryotic genomes in a very short time even in small facilities. Moreover, we demonstrated that although the Oxford Nanopore technology is a relatively new sequencing technology, currently with a high error rate, it is already useful in the generation of high-quality genome assemblies.

N°21 TASKMANAGER : Massively parallelizable workflow management

Artem KOURLAIEV1, Carole DOSSAT1, Stefan ENGELEN1, Jean-Marc AURY1
1CEA/Institut de Génomique/Genoscope/LIS/RDBIOSEQ, 2 rue Gaston Crémieux, 91000, Evry, Cedex France
Auteur à contacter : akourlai@genoscope.cns.fr

TaskManager permet de développer des workflows génériques en Perl, potentiellement exécutables dans tous les environnements existants et avec différents degrés de parallélisation. Il est capable d'exécuter plusieurs milliers de tâches en parallèle selon les ressources mises à sa disposition en utilisant différents "batch manager" (Slurm ou Lsf). L'approche diviser pour régner est particulièrement adapté à TaskManger pour réduire les temps de calculs des analyses de données NGS. Il génère des logs ayant la même structure dans tous les environnements d'exécution facilitant les analyses et les reprises sur erreurs. Il est développé en utilisant la programmation orienté objet facilitant ainsi son évolution, l'ajout de nouvelles fonctionnalités et d'environnements d'exécution. Avec son utilisation le développeur aura toujours la même manière d'implémenter les workflows quelque soit l'environnement de la plateforme et le niveau de parallélisation. TaskManager a été utilisé pour le portage de pipeline d'alignement du Genoscope au TGCC (centre HPC) afin de traiter des projets impliquant des big data (TARA Oceans).

TaskManager allows to develop generic workflows in Perl, potentiallyexecutable in various environments and with different degrees of parallelization. TaskManager can execute thousands tasks in parallel, depending on the available resources, usingdifferent "batch manager" (Slurm or Lsf). The divide and conquer approach isparticularly well adapted to TaskManagerto reduce the computing time of NGS data analysis.TaksManagergenerates logs with the same structure in all execution environments that facilitate analysis and error recovery. TaskManager is developed using object-orientedprogramming that facilitate future evolution, adding functionality and new runtime environments. Thanks to TaskManager, the developer will always have the same way of implementing workflows on different plateforms and different level of parallelization. TaskManager has been used to export mapping workflows from Genoscope to TGCC (HPC center) in order to treat big data projects (TARA oceans).

N°22 MaGuS: a tool for map-guided scaffolding and quality assessment of genome assemblies

Mohammed-Amin Madoui[(1)], Carole Dossat[(1)], Léo d'Agata[(1)], Jan van Oeveren[(2)], Edwin van der Vossen[(2),] Jean-Marc Aury[(1)]

[(1)]CEA, DSV, Institut de Génomique, Genoscope, 2 rue Gaston Crémieux, CP5706, 91057 Evry, France

[(2)]Keygene NV, Agro Business Park 90, 6708 PW, Wageningen, The Netherlands

Scaffolding is a crucial step in the genome assembly process. Current methods based on large fragment paired-end reads or long reads allow an increase in continuity but often lack consistency in repetitive regions, resulting in fragmented assemblies. Here, we describe a novel tool to link assemblies to a genome map to aid complex genome reconstruction by detecting assembly errors and allowing scaffold ordering and anchoring.

We present MaGuS (map-guided scaffolding), a modular tool that uses a draft genome assembly, a genome map, and high-throughput paired-end sequencing data to estimate the quality and to enhance the continuity of an assembly. We generated several assemblies of the Arabidopsis genome using different scaffolding programs and applied MaGuS to select the best assembly using quality metrics. Then, we used MaGuS to perform map-guided scaffolding to increase continuity by creating new scaffold links in low-covered and highly repetitive regions where other commonly used scaffolding methods lack consistency.

MaGuS is a powerful reference-free evaluator of assembly quality and a map-guided scaffolder that is freely available at https://github.com/institut-de-genomique/MaGuS. Its use can be extended to other high-throughput sequencing data (e.g., long-read data) and also to other map data (e.g., genetic maps) to improve the quality and the continuity of large and complex genome assemblies.

N°23 WP 1.1, e-infrastructure France-Genomique, airain.ccc.cea.fr, TGCC

C. Scarpelli & N. Wiart

L'e-infrastructure France-Genomique implantée sur le calculateur Airain du TGCC est en production depuis maintenant 2 ans et demi. Nous présentons ici quelques éléments chiffrés sur l'utilisation du calculateur, ainsi que des informations sur l'environnement logiciel de base pour la génomique.

N°24 A new BAC ends sequencing approach to improve wheat 1B chromosome assembly

Laura Brinas, Caroline Belser, Adriana Alberti, Céline Orvain, Karine Labadie, Laurie Bertrand, Arnaud Couloux, Jean-Marc Aury, Valérie Barbe, Frédéric Choulet, Etienne Paux, Patrick Wincker

Bacterial artificial chromosome (BAC) libraries are still a valuable tool for de novo assembly of complex genomes, such as plants genomes. Shotgun sequencing of BACs, individually or by pools, produces first assemblies which usually need further improvement towards finished quality. We developed a new approach to obtain BAC ends libraries for Illumina sequencing (BES), overcoming the expensive and time consuming BAC ends Sanger sequencing. We applied this protocol to improve the initial assembly of wheat 1B long arm chromosome. To produce a reference sequence, a total of 6023 BAC clones representing the minimal tiling path of the 1B long arm physical map were sequenced by Illumina technology: paired end data from pools of about 10 BACs were combined to 5 kb mate pair reads to ensure the assembly of large sequence scaffolds. Contigs were generated by Newbler assembler using merged paired end data; then, they were scaffolded using SSPACE and mate pair sequences. Average scaffold size was 62 Kb and N50 about 231 Kb. We tested the BES approach by preparing libraries from pools containing 384 BACs each. Illumina paired end reads were generated and used for scaffolding improvement with SSPACE. We observed that 62% of the previous assemblies benefited from this approach: scaffolds average size and N50 increased significantly (respectively 76 Kb and 354 kb on average). Therefore, the new method revealed useful for improving de novo assembly, especially in the case of challenging highly repeated genomes. Here we present the library protocol in details and discuss further assembly analyses.

IGBMC Strasbourg

N°25 Low input ChIPseq analyses

B. Hamelin, T. Ye, H. Neyret-Kahn, P. Laurette, G. Davidson, V. Alunni, S. Vicaire, C. Thibault, F. Radvanyi, I. Davidson, S Le Gras et B. Jost

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is one of the main activities of our platform. In the past few years, we prepared thousands of ChIP-seq samples with different protocols, but none of them was suitable for small quantify of input DNA. Recently, we have tested the MicroPlex Library Preparation Kit v2 from Diagenode. Ten ng, 1 ng or 0.1 ng anti-H3k4me3 ChIP DNA samples were processed with this protocol and the sequencing results have been compared with that from our standard protocol, BiooScientific Nextflex ChIP-seq preparation kit. MicroPlex kit produced high yield and quality libraries using 10 and 1ng DNA comparable to libraries generated with Nextflex kit using 10 ng sample. Since fewer PCR amplification cycles are needed using MicroPlex kit, the number of unique reads was significantly increased. When starting with 0.1ng DNA, the proportion of duplicated reads was significantly increased. We still have to test the performance of this kit on transcription factor ChIP-seq samples.

N°26 Development and integration into Galaxy of a suite of workflows dedicated to RNAseq data analysis.

Amandine Velt, Serge Uge, Stephanie Le Gras, Julien Seiler and Céline Keime IGBMC - CNRS UMR 7104 - INSERM U 964, Université de Strasbourg, Illkirch, France

The exponential growth of high-throughput Omics data has posed a great technical challenge to experimentalists who lack bioinformatics skills and computing power. Moreover, integrative analysis of data from various sources is needed to provide biological insights into biological systems. That is why we have developed our own Galaxy instance, GalaxEast, an open and powerful public web-based platform for integrative analysis of Omics data (www.galaxeast.fr). In order to provide researchers with a friendly and comprehensive resource for analyzing RNA-seq data, we implemented several workflows in GalaxEast and produced an associated documentation providing advice and guidance for these analyses. We will present the three Galaxy workflows we implemented: the first performing quality assessment, mapping and quantification of gene expression, the second enabling gene expression data merging, normalization and annotation and the last allowing to perform a differential gene expression analysis. We will also describe how we adapt Galaxy tools, downloaded from the Galaxy Tool Shed, to our needs and the methods we used to package our own scripts and tools for Galaxy. We already shared these modified tools with IFB and we would be pleased to share these tools, workflows and practices with other platforms. In order to complete these workflows we are currently working on the implementation of a workflow dedicated to alternative splicing analysis using DEXseq package.

IBENS Paris

N°27 Présentation du jeu de donnée servant à valider les protocoles de fabrication des banques de la plateforme génomique de l'IBENS.

Fanny Coulpier; Sophie Lemoine; Corinne Blugeon; Marie-Noëlle Rossignol; Asha Baskaran; Stéphane Lecrom

Nous sommes confrontés sur la plateforme génomique, à la profusion de nouveaux protocoles proposés par les fournisseurs de produits de biologie moléculaire pour permettre la fabrication des banques en vue du séquençage à haut débit des échantillons de génomique fonctionnelle. En tant que plateforme, nous nous devons de nous assurer que les protocoles que nous utilisons fonctionnent correctement et que les résultats que nous rendons à nos utilisateurs sont fiables. C'est la raison pour laquelle nous avons décidé d'obtenir des échantillons qui soient reproductibles dans le temps et dont nous connaissions bien le modèle biologique afin de pouvoir valider nos protocoles expérimentaux ainsi que les différents modes de séquençage utilisés. Nous avons pour cela choisi le modèle expérimental du laboratoire de Patrick Charnay à l'IBENS en comparant deux conditions avec 3 réplicats chacun afin de permettre les analyses statistiques des résultats obtenus. Depuis maintenant 5 ans, nous avons utilisé ce dessin expérimental pour tester nos protocoles de RNA-Seq : directionnel et non directionnel, purification poly A et déplétion ribosomique ainsi que l'amplification des petites quantités de matériel, tout cela avec des méthodes manuelles ou automatisées. Soit un total de 11 protocoles de fabrication de banques différents séquencés sur notre séquenceur HiSeq 1500 en mode haut débit, en mode rapide et sur notre nouveau NextSeq 500. Les données que nous présentons ici, rassemblent toutes ces résultats. Ce jeu de données, que nous souhaitons mettre à disposition de la communauté scientifique, est l'occasion de comparer les données obtenues sur un même dessin expérimental avec des protocoles de fabrication de banques différents et d'analyser les résultats sur un modèle biologique connu.

N°28 An automated and modular output quality control pipeline for Illumina sequencers

Sandrine Perrin, Sophie Lemoine, Stéphane Le Crom and Laurent Jourdren

Aozan[1] est un pipeline de traitement des données post-séquençage, il prend en charge automatiquement la détection des runs, le transfert des données, le démultiplexage des lectures avec bcl2fastq2 [2] et le contrôle qualité avec FastQC [3] et Fastq-Screen [4], parallèlement il surveille les espaces disques disponibles. Aozan peut être déployé en utilisant la technologie Docker [5] qui offre un passage en production très simple et rapide. Une fois installé et configuré, Aozan fonctionne sans nécessité d'intervention humaine et il informe les utilisateurs par courriel du déroulement des étapes. Très flexible, Aozan convient aux petites comme aux grandes plateformes possédant un ou plusieurs séquenceurs Illumina (HiSeq ou NextSeq). Afin d'optimiser les temps de traitement dans un environnement multi-serveur, il est possible, par exemple, d'assigner un ou plusieurs serveurs à chaque étape. Le rapport de contrôle qualité produit est totalement personnalisable à l'aide d'un unique fichier de configuration. Les données des fichiers de contrôle et de sortie de séquençage (InterOp Illumina, rapport FastQC, fichiers xml...) sont compilées dans plusieurs sections optionelles: runs, lignes, projets et échantillons. Afin de répondre aisément aux besoins des utilisateurs, le système de plugins permet d'intégrer facilement de nouvelles fonctionnalités, comme l'ajout d'un traitement complémentaire sur les reads. En conclusion, Aozan est totalement autonome et modulable. Il assure une parfaite tracabilité des données de séquençage et du contrôle qualité en adéquation avec les attentes d'une certification qualité ISO 9001.

N°29 Gestion automatisée des annotations dans le contexte d'une plateforme ouverte

Sophie Lemoine, Laurent Jourdren, Stéphane Le Crom

La plateforme génomique de l'Ecole normale supérieure est une infrastructure ouverte spécialisée en génomique fonctionnelle. Nous accueillons des projets quelque soit l'espèce eucaryote étudiée pourvu qu'un génome et qu'une annotation GFF3 [1] cohérente puissent être fournis à notre pipeline d'analyse Eoulsan [2].

Les formats d'annotations disponibles sont très différents. Cette diversité implique la validation de chaque couple génome/annotation avant toute utilisation en production dans Eoulsan. Cette procédure est assez simple mais les étapes et les choix qui la précèdent posent plusieurs questions. En effet, il est difficile de n'avoir qu'une seule source de données si l'on ne souhaite pas restreindre notre catalogue, mais en privilégiant Ensembl [3] quand l'espèce y est déposée, il est possible de : (i) interroger directement Ensembl via son API pour générer un gff3 natif, (ii) contrôler les versions de génome et d'annotations selon de la version de l'API, (iii) bénéficier de BioMart [4] pour agréger des annotations complémentaires. A chaque mise à jour d'Ensembl, notre pipeline nous permet de : (i) rapatrier génomes, fichiers GFF3 et annotations BioMart automatiquement, (ii) lancer la simulation de lectures sur le génome à tester, (iii) exécuter Eoulsan pour valider génomes, annotations GFF3 et annotations complémentaires.

Cette chaîne de validation de l'annotation nous permet la réalisation rapide et systématique de cette étape cruciale pour tout projet de génomique fonctionnelle en évitant son coté fastidieux.

1- Eilbeck K., Lewis S.E., Mungall C.J., Yandell M., Stein L., Durbin R., Ashburner M, The Sequence Ontology: A tool for the unification of genome annotations, Genome Biology (2005) 6:R44
2- Laurent Jourdren, Maria Bernard, Marie-Agnès Dillies and Stéphane Le Crom. Bioinformatics (2012) 28 (11): 1542-1543
3- Paul Flicek et al, Ensembl 2014, Nucleic Acids Research 2014 42 Database issue:D749-D755
4- Jonathan M. Guberman et al, BioMart Central Portal: an open database network for the biological community, Database 2011: bar041

N°30 A French Galaxy Tool Shed to federate the national infrastructures and offering quality assessed tools like SARTools

Loraine BRILLET-GUÉGUEN[1], Christophe CARON[1], Valentin LOUX[2] and the French Galaxy Working Group[3]

[1] ABIMS, FR2424 CNRS-UPMC, Station Biologique, Place Georges Teissier, 29680, Roscoff, France

[2] UR1404 Mathématiques et Informatique Appliquées du Génome à l'Environnement, INRA, F-78352 Jouy-en-Josas, France

[3] Institut Français de Bioinformatique [ANR-11-INBS-0013], France Génomique [ANR-10-INBS-0009] and MetaboHUB [ANR-11-INBS-0010]

Auteur à contacter : gtgalaxy@groupes.france-bioinformatique.fr

L'environnement *Galaxy* dédié notamment à l'activité de bio-analyse connaît un succès croissant au sein des communautés bio-informatiques et biologistes. **L'*Institut Français de Bioinformatique*** (IFB) a missionné en 2013 un Groupe de Travail autour de la plateforme Galaxy (GT Galaxy). Ce groupe rassemble plusieurs plateformes nationales, et pilote des actions d'animation (Galaxy Day, écoles thématiques, etc.) et de structuration (formation, guides des bonnes pratiques, etc.) des communautés utilisateurs et développeurs.

SARTools est un package R dédié à l'analyse différentielle de données RNA-seq et développé à l'Institut Pasteur dans le cadre d'un work package ***France Génomique***. Cet outil, disponible depuis mai 2015 sur le Tool Shed IFB, est plébiscité par la communauté des bio-analystes et a été utilisé lors de l'école thématique de bioinformatique AVIESAN en septembre 2015.

Le *Tool Shed IFB* (dépôt commun d'outils et de workflows) participe à la mise en œuvre d'une stratégie de fédération de la communauté avec la diffusion de *bonnes pratiques* d'intégration d'outils dans Galaxy et la *formation* des ingénieurs des plateformes concernées. Un effort particulier est porté sur la *qualité* des intégrations d'outils et de workflows proposées avec la mise en place de tests fonctionnels et de procédures de validation. Ce projet a permis de centraliser et promouvoir les outils de bio-analyse de la communauté française au travers d'une collaboration entre deux infrastructures nationales : l'IFB et France Génomique.

POPS Orsay

N°31 UltraLow input from micro-dissected samples

Taconnat Ludivine, Yansouni Jennifer, Delannoy Etienne , Balzergue Sandrine, INRA, POPS - Transcriptomic Platform, IPS2
Sakai Kaori, Borrega Nero, Lepiniec Loïc, Faure Jean Denis, Dubreucq Bertrand, INRA, IJPB
Brunaud Véronique, Martin Magniette Marie-Laure, INRA, Genomic Network team, IPS2

Advances in RNA-seq methodologies **from limiting amounts of total RNA** have facilitated the characterization of singular cell-types in various biological systems. However, RNA sequencing still remains challenging and expensive when working with limited material. Furthermore total RNA isolated from micro-dissections can often be of low quality too. Consequently, conformity of results and quality of data can be affected in low input samples. Here we report **technical developments in the construction and analysis of RNA-seq libraries prepared from Arabidopsis laser micro-dissection of embryo epidermis tissues**. The associated scientific purpose allows us to corroborate the protocol **starting from 100pg of micro-dissected plant total RNA** ; we recently tested and validate the protocol with 75pg and 50pg of starting total RNA (data not shown).

Institut Pasteur Paris

N°32 SARTools: a DESeq2- and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data

Hugo Varet, Jean-Yves Coppée, Marie-Agnès Dillies

We present the SARTools R package dedicated to the differential analysis of RNA-seq data for experiments with a simple design, i.e. experiments comparing several conditions of the same biological factor. SARTools provides tools to generate descriptive and diagnostic graphs, to run the differential analysis with one of the well known DESeq2 or edgeR packages and to export the results into easily readable tab-delimited files. It also facilitates the generation of a final HTML report which displays all the figures produced, explains the statistical methods and gives the results of the differential analysis. Note that SARTools does not intend to replace DESeq2 or edgeR: it simply provides an environment to go with them. Moreover, the vignette of the package contains extensive help on how to use the workflow and gives some advices to detect potential problems such as the presence of a batch effect within the experiment or inversions of samples. SARTools is available on GitHub and is distributed with two R script templates (one for DESeq2 and one for edgeR) which use functions of the package. It is also possible to use it on Galaxy.

N°33 Pipelines and tools for comparative genomes analysis of large bacterial populations

A Villain, Alexandre Almeida, Christiane Bouchier, Phillippe Glaser and Pierre Lechat

The sequencing of a large number of bacterial strains has become possible thanks to the improvement of Next-Generation Sequencing (NGS) technologies. Bioinformatics analysis evolved accordingly, and the need for standardized and accessible workflows as well as visualization tools to display data in a meaningful way is stronger than ever. Here we use the comparative genomics analysis of 50 bovine isolated Streptococcus agalactiae strains as a practical application for our variant calling pipeline and genome browser SynTView (Lechat, 2013). Our goal is to characterize the strains coming from different farms and understand their relation and evolution through a SNP typing done by Illumina sequencing. An existing CRISPR typing can be used as a comparison, and various phenotypic information about the strains are available to broaden.

N°34 The transcript isoform quantification conundrum: an overview

Thibault DAYRIS, Oussema SOUIAI, Coline BILLEREY, Claire TOFFANO-NIOCHE, and Daniel GAUTHERET

Résumé Dans l'ensemble des domaines du vivant, les gènes sont transcrits en de multiples isoformes. Le RNA-Seq est aujourd'hui la technologie privilégiée pour leur identification et leur quantification. Cependant, déduire correctement leur structure reste un défi, compte tenu de la petite taille des reads et de la diversité des événements de transcription (débuts, terminaisons, maturations et dégradations alternatifs) qui limitent notre capacité à quantifier exactement ces isoformes à partir d'une librairie RNA-Seq. Dans cette étude, nous considérons une trentaine d'outils de quantification d'isoformes. Nous soulignons leur diversité en terme de méthodes de représentation et de quantification des événements alternatifs, ce qui nous amène à définir différents domaines d'applications tels que la détection d'événements spécifiques, la détection de nouveaux événements, ou la réalisation d'analyse d'expression différentielle. Nous montrons que ces outils ne sont ni interchangeables ni équivalents lors de leur utilisation. Enfin, nous nous concentrons sur les logiciels qui mesurent l'expression différentielle d'isoformes entre deux conditions. Notre objectif est de comparer ces outils selon un critère objectif (i.e. f-score), via le développement de jeux de données RNA-Seq simulés aux proportions d'événements d'épissage alternatif contrôllées. Nous présenterons l'avancement de notre projet et nos premiers résultats sur des données RNA-Seq « réelles ». Summary In all domains of life, genes are transcribed into multiple transcript isoforms. RNA-seq is currently the preferred technology for the identification and quantification of these isoforms. However, correctly inferring isoforms is very challenging due to the small read size in RNA-seq data and the diversity of isoforms that result from multiple events, including variations in transcription start, termination, processing and degradation. This limitation strongly impacts our ability to properly quantify isoforms based on a given RNA-seq library. In this study, we review about 30 isoform quantification tools. We highlight their diversity in terms of methods for representing and quantifying “events”. In turn this defines different ranges of application such as detecting specific event classes, detecting novel events, or performing differential expression analysis. We show that those tools are neither equivalent nor interchangeable in their use. We then focus on software that measure differential isoform expression between pairs of conditions. With the aim of benchmarking and rating these tools through an objective value (such as an f-measure), we are developing simulated RNA-seq datasets with fully controlled proportions of alternative events. We will present these ongoing developments, as well as results of our initial comparisons on “real” RNA-seq datasets.

Migale Jouy-en-Josas

N°35 View and synchronize several genotypes using IGV

Marie-Laure Franchinard (Migale/INRA) Frédéric Sapet (Biogemma) Sandra Dérozier (Migale/INRA) Franck Samson (Migale/INRA) Jean-François Gibrat (Migale/INRA)

Résumé: IGV (Integrative Genomics Viewer) est un outil graphique écrit en Java très efficace pour visualiser et explorer facilement une très grande variété de données génomiques, mais sur un seul génome à la fois. Or pour des besoins de génomique comparative, il peut-être très utile d'observer différents types de données simultanément sur plusieurs génotypes. Dans le cadre du projet BioDataCloud (Programme des Investissements d'Avenir), une collaboration entre la plateforme Migale et l'entreprise Biogemma a été mise en place pour répondre à ce besoin. Conformément au cahier des charges établi par Biogemma, nous avons ajouté une nouvelle fonctionnalité à IGV permettant un "saut" vers un nouveau génotype à partirde différents types de données (gènes, régions dans la séquence génomique, marqueurs génétiques) sélectionnés par l'utilisateur sur le génome de référence. Ce saut se traduit par l'ouverture d'une nouvelle fenêtre IGV sur les données sélectionnées à condition qu'elles soient disponibles sur le nouveau génotype. Cette fenêtre conserve toutes les fonctionnalités d'IGV et se synchronise simultanément avec la fenêtre principale. Tous les sauts peuvent être sauvegardés dans un fichier session d'IGV ce qui permet de restaurer rapidement les génotypes et les données utilisés ou de les partager avec d'autres utilisateurs. Le nombre de sauts réalisables et donc de génotypes observables simultanément ne dépend que des capacités matérielles utilisées, notamment en mémoire vive, et de la disponibilité des données correspondantes. Le déploiement de cette nouvelle version d'IGV dans une machine virtuelle duCloud de l'IFB, l'appliance BioDataCloud-IGV, garantit donc un gain certain en performance et accessibilité.

N°36 Mapdecode : inventory and benchmark of read mapping tools

Compain Jérome [2],Heriveau Claudia [1], Jullien Renaud [1], Nandy Sivasangari [1],Collin Olivier [1],Gibrat Jean-François [2],Loux Valentin [2], Martin Véronique [2],Schbath Sophie[2]

[1] CNRS, UMR6074 Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, France

[2] INRA, UR1044 Unité Mathématiques et Informatique Appliquées du Génome à l'Environnement, Jouy-en-Josas, France

We will present mapdecode, an inventory of published mapping tools, associated with various benchmatks. The inventory of the mapping tools and the benchmark results for the one tested are available on the Mapdecode website (http://mapdecode.france-genomique.org).

Bilille Lille

N°37 Fast and easy identification of microRNAs in plant genomes with miRkwood

Isabelle Guigon, Sylvain Legrand, Jean-Frederic Berthelot, Mohcen benmounah, Helene Touzet

MicroRNAs (miRNAs) play a crucial role in the post-transcriptional regulation of eukaryotic gene expression, in plants and animals. Many aspects of the biogenesis and evolution of miRNAs in animals and plants differ. For example, unlike miRNAs of animals, which are mainly found in introns or exons from protein coding genes, most plant miRNAs are encoded by discrete genes. Moreover, miRNAs are released from their precursors using distinct pathways in the two kingdoms. Also, miRNA precursors are more heterogeneous in plants than in animals, varying greatly in size and structure. These differences have justified dedicated approaches for miRNA gene finding. However although several prediction tools are available for metazoan genomes, the number of tools dedicated to plants is relatively limited. Considering this gap, we have developed miRkwood, a user-friendly web server specifically designed for plant miRNAs. miRkwood is able to face the diversity of plant pre-miRNAs and allows the prediction of precursors of both conserved and non-conserved miRNAs. miRkwood can deal with both full small RNA sequencing reads and short genomic sequences (up to 100 000 nt). Moreover, it offers an intuitive and comprehensive user interface to navigate in the data, as well as many export options (GFF, CSV, FASTA, ODT) to allow the user to conduct further analyses on a local computer. It is accessible at http://bioinfo.lifl.fr/mirkwood.

Genotoul bioinfoToulouse

N°38 De novo RNA-seq Assembly Pipeline

Cédric Cabau, Frédéric Escudié, Anis Djari, Yann Guiguen, Julien Bobe and Christophe Klopp

Short read RNASeq de novo assembly is a well established method to study transcription of organisms lacking a reference genome sequence. Available software packages such as Trinity and Oases have proven to be able to build high quality contigs from short reads. But there is still room for improvement on different points such as: compactness: they often produce different contigs which are included in one another or overlapping one another, chimerism: the contigs contain different kinds on chimera such as duplicated open reading frames, substitution, insertion, deletion errors: the consensus sequences build by the assembler contain errors which can be partly corrected using the read alignments. DRAP includes three modules: runDrap chains an Oases or Trinity assembly of reads from a given sample with several compaction and correction steps. It produces several assembly files with different FPKM threshold for total contigs or contigs comprising an open reading frame. A report file presents the resulting assembly and alignment metrics. runMeta gathers all the samples assemblies and fusions the results in a unique representative contig set. It also removes the redundancy between sets and produces a general reports including assembly and alignment metrics. runAssessment processes different contigs sets build from the same read sets to generate assembly and alignment metrics which are collected in report. It helps to choose the best assembly.

Get-PlaGe

N°39 Making Our Lives Easier NGS Goes Automatic

Gaëlle VILCHEZ[1], Claire KUCHLY[1], Jérôme MARIETTE[3], Frédéric ESCUDIE[3], Ibouniyamine NABIHOUDINE[3], Olivier BOUCHEZ[2], Diane ESQUERRE[2], Céline VANDERCASTEELE[2], Johanna BARBIERI[1], Céline JEZIORSKI[1], Sophie VALIERE[1], Clémence GENTHON[1], Marie VIDAL[1,] Alain ROULET[1], Sandra FOURRE[2], Catherine ZANCHETTA[1], Adeline CHAUBET[1], David RENGEL[4], Denis MILAN[1], Cécile DONNADIEU[2] and Gérald SALIN[2]

[1] DGA, UAR1209 INRA, Plateforme GeT-PlaGe, 31326 Castanet-Tolosan, France
[2] GenPhySE, UMR1388 INRA, Plateforme GeT-PlaGe, 31326 Castanet-Tolosan, France
[3] MIAT, UR875 INRA, Plateforme bioinformatique, 31326, Castanet-Tolosan, France
[4] LIPM, UMR441 INRA, 31326 Castanet-Tolosan, France

Due to the constant increase of second generation sequencers (as HiSeq or MiSeq) throughput and the multiplication of their applications, a challenge is to sequence, at the same time, in a single run, ever more samples of different types, while being able to verify the quality of the produced data. At the INRA GeT-PlaGe facility, we have automated both library production and data quality control steps, in partnership with the Genotoul Bioinformatics facility, for protocols such as whole genome sequencing, Amplicon sequencing (e.g. 16S sequencing on MiSeq for metagenomics studies), stranded RNA-seq, Mate-Pair or whole genome bisulfite sequencing. Having acquired a solid expertise in library preparation and data quality control of short fragments, our challenge for the coming months will be to integrate data from 3rd generation sequencers in our automated quality control processes, dealing with the specificity of long fragments, in partnership with the bioinformatics community of Toulouse and France Génomique

N°40 First results on different projects on PACBIO RSII

L'ensemble de l'équipe GeT-PlaGe
L'équipe Tournesol du LIPM INRA de Toulouse
L'équipe Bioinformatique du LIPM de Toulouse
La sociéte Libragen

1- de novo complex genome sequencing: 405 SMRTcell result assembling to resolve complexity of sunflowers's genome**: "SUNRISE projet"**

2- epigenetic analysis :Methylation adaptation of Ralstonia solanacearum in differents plant host : "**GENEPIA project"**

3- metaganomic: Identification and quantification of mock bacterial community with the full length 16S**: "Libragen project"**

I2BC Gif-sur-Yvette

N°41 Systematic comparison of small RNA library preparation protocols for next-generation sequencing

Cloelia Dard-Dascot, Yves d'Aubenton-Carafa, Karine Alix, Claude Thermes and Erwin van Dijk

Next-generation sequencing approaches have revolutionized the study of small RNAs (sRNAs) on a genome-wide scale. However, classical small RNA library preparation protocols suffer from serious bias, mainly introduced during adapter ligation steps. Several types of sRNA including plant microRNAs (miRNAs), piwi-interacting RNAs (piRNAs) in insects, nematodes and mammals, and small interfering RNAs (siRNAs) in insects and plants contain a 2'-O-methyl modification at their 3' terminal nucleotide. This inhibits 3' adapter ligation and makes library preparation particularly challenging. Randomized or "High Definition" (HD) adapters have been shown to reduce bias among different RNA sequences and the addition of PEG improved ligation efficiency for a 2'-O-methyl modified RNA in experiments using a single isolated sRNA. However, the effects of HD adapters and PEG on the representation of 2'-O-methyl modified RNAs in complex sRNA libraries have not been investigated. Here we evaluate the performance of two different sRNA preparation kits, the classical Illumina kit using standard adapters without PEG and a novel kit from BIOO Scientific using HD adapters with PEG, with regard to bias among different RNA sequences and against 2' O-methyl RNAs. We modified both protocols to dissect the roles of HD adapters and PEG in bias reduction. In addition, we tested a variant of a recently developed type of adapters, designated MidRand-like (MRL) adapters in the background of both kits. Our results show that: (1) with the classical Illumina protocol there is overall strong bias against 2' O methyl RNAs but this bias heavily depends on the RNA sequence and may vary from 3 fold to more than a 100 fold under-representation, (2) HD adapters and PEG both reduce bias but may also promote a loss of sequences due to the formation of adapter dimers. (3) MRL adapters further reduce bias among different sequences and against 2' O-methyl RNAs in general. The best overall results were obtained with MRL adapters in the presence of PEG allowing a significantly improved detection of 2' O-methyl sRNAs.

ProfileXpert Lyon

N°42 Single cell analysis pipeline to capture intra-tumoural heterogeneity

Magali Roche1, Catherine Rey2, Séverine Croze2, Clement Delestre1, Audrey Kadouri1, Anne Wierinckx, Catherine Legras-Lachuer1-3, Joel Lachuer3-4

1 Centre de Recherche en Cancérologie de Lyon, INSERM U1052/CNRS UMR 5286 Centre Léon Bérard, Lyon, France
2 Université de Lyon, Université Lyon 1, Lyon, France
3 ViroScan3D, Trévoux, France
4 ProfileXpert, SFR-Est, CNRS UMR-S3453 - INSERM US7, Lyon, France
5 UMR CNRS 5557 UCBL USC INRA 1193 ENVL, Dynamique microbienne et transmission virale, Lyon, France

Intratumoural heterogeneity is described to be the major mechanism involved in tumour plasticity and treatment resistance. Indeed, most of the anti-cancer treatments are designed on the basis of molecular analyses performed on the tumour. Nevertheless, it was described that although the majority of the cells constituting the tumour mass are destroyed by treatment some rare cells are resistant and are at the origin of tumour resurgence. Thus, in order to understand the molecular mechanisms that sustain tumour progression and treatment resistance, it is necessary to distinguish the different cell populations constituting the tumour and to characterize their individual properties.

Until recently, molecular analyses of single cell were extremely difficult as isolation and amplification methods were not optimized to allow the analysis of one cell but rather a reduced group of cells thus limiting the accurate identification of tumour cell subpopulations. To answer this technical challenge, we have set up and evaluated a pipeline specifically designed to perform the three critical steps for molecular analyses of unique cell that are the capture, the isolation and the whole genome or transcriptome amplification. This pipeline is based on a very innovative technology the C1 system that uses microfluidic to isolate a single cell in individual chamber and perform the lysis and enzymatic reactions necessary for transcriptomic or genomic analyses in picoliters.

To evaluate the accuracy of the C1 system we have performed a whole RNA sequencing experiment on 288 isolated stem cells on HiSeq 2500 Illumina. Here we present the results of this technical evaluation. Each critical step of the pipeline was evaluated thanks to precise quality control criteria. Statistical controls for technical bias evaluation were performed on each single cell transcriptome.

Here we show that the C1 system is able to perform accurate transcriptomic profiling on a single cell, thus highlighting subcellular populations. Nevertheless, the main drawback of this system is that the starting cell population should be globally homogeneous in size. This pipeline could be significantly improved with the addition of the DEParray from Silicon Biosystem that allows the selection of cells from a really heterogeneous population and on more parameters than C1 system.

ATGC Montpellier

N°43 Long read based assembly and impact of error correction

Eric RIVALS et Amal MAKRINI

Recent technological advances gave rise to the 3rd generation of sequencing techniques, which generate long reads. The Pacific Biosciences technology ('PacBio') and Oxford Nanopore Technology ('MinIon') are the most prevalent long-read technologies today. In theory, long reads should ease genome assembly by resolving the structure of repeated regions, and help distinguishing RNA isoforms in complex eukaryotic transcriptomes. Yet, this promise is hampered by a high error rate and lower coverage With error rates superior to 15\%, long reads are not amenable to direct assembly. Hence, the challenge of long reads is to error correct them. Two error correction strategies have been proposed: either self-correction using only long reads, or hybrid correction using high quality set of short reads. LoRDEC is a hybrid error correction tool for long reads that can handle very large sets of short and long reads. It aligns the long reads on the paths of a de Bruijn graphs of the short reads. LoRDEC achieves a good level of correction while it outperforms existing other error correction software in terms of time and memory. The objectives of this study is to : i) Assess the error correction accuracy by LoRDEC1 off both PacBio and MinIon data, and ii) investigated the impact of error correction on the assembly.

IGS Marseille

N°44 Assignation taxonomique et détection de chimères dans les données génomiques complexes et les métagénomes: prototype d'un outil interactif

Desmarais Damien, Poirot Olivier, Claverie Jean-Michel

L'étude des données de métagénomique est rendue complexe par le nombre d'espèces diffétrentes dans l'échantillon mais aussi l'accumulation des problèmes liés au séquençage de cellule unique. Nous avons ici développé un outils d'aide à la décision lors de l'assignation taxonomique des contigs issus de l'assemblage de jeu de données métagénomique

PRABI LYON

N°45 Towards an automatic compilation of a compact, phylogenetically driven and taxonomy compliant set of prokaryotic 16S rRNA

Jean-François Taly, Christine Oger, Jean-Pierre Flandrois et Guy Perrière

A universal prokaryotic taxonomy is necessary for fundamental research on evolution or for quantifying the microbial diversity of an environment. Estimating the phylogenetic position of micro-organisms by genetic sequence comparisons is considered as the gold-standard in taxonomy. This is also a way to approach the geno-species composition of a sample by comparison with an annotated dataset. The quality of the reference database used in such analyses is crucial: the database must reflect the up-to-date nomenclature but also must contain currently not described geno-species. Due to advance in sequencing technology, sequence and taxonomy databases are updated daily with new or corrected material. However, the massive amount of duplicated, misannotated or erroneous sequences makes it difficult to decipher the phylogenetic signal from the noise. Databases are either taxonomy driven -losing most of the genomic variability information- or highly relaxed -retaining most of the heterogeneity but also technical noise. In this work we propose to build a reference database from a weekly automatic refinement of the 16S ribosomal RNA collections stored in Genbank and RefSeq. The aim is to better take into account the genetical variability of those markers thanks to a phylogenetic reorganization of the diversity around proven nodes of the NCBI taxonomy. In practical, each species (or species-equivalent nodes) is first summarized by a selection of the most representative sequences in term of phylogenetical signal. We then reproduce the same strategy for the next two taxonomic/phylogeny levels. We obtain thus an up-to-date reference database with considerable reduction of the sequence number without limiting the genomic variability information and an optimal level of taxonomic information.

Prénom	Nom	Organisme	Ville	mail
Chantal	ABERGEL	IGS	Marseille	Chantal.Abergel@igs.cnrs-mrs.fr
Tina	ALAEITABAR	URGI	Versailles	tina.alaeitabar@versailles.inra.fr
Adriana	ALBERTI	IG Genoscope	Evry	aalberti@genoscope.cns.fr
Joelle	AMSELEM	URGI	Versailles	joelle.amselem@versailles.inra.fr
Marie-Jeanne	ARGUEL	Plateforme de génomique fonctionnelle	Nice-Sophia Antipolis	arguel@ipmc.cnrs.fr
Francois	ARTIGUENAVI	IG CNG	Evry	artiguenave@cng.fr
Hélène	AUGER	I2BC	Gif-sur-Yvette	helene.auger@i2bc.paris-saclay.fr
Jean-Marc	AURY	IG Genoscope	Evry	jmaury@genoscope.cns.fr
Valerie	BARBE	IG Genoscope	Evry	vbarbe@genoscope.cns.fr
Pascal	BARBRY	Plateforme de génomique fonctionnelle	Nice-Sophia Antipolis	barbry@ipmc.cnrs.fr
Christophe	BATTAIL	IG CNG	Evry	christophe.battail@cea.fr
Sylvain	BAULANDE	Institut Curie	Paris	sylvain.baulande@curie.fr
Xavier	BENIGNI	IG CNG	Evry	xavier.benigni@cng.fr
Aurélie	BÉRARD	IG EPGV	Evry	berard@cng.fr
Wahiba	BERRABAH	IG Genoscope	Evry	berrabah@genoscope.cns.fr
Alexis	BERTRAND	IG Genoscope	Evry	abertrand@genoscope.cns.fr
Celine	BESSE	IG CNG	Evry	besse@cng.fr
Marie -Thérès	BIHOREAU	IG CNG	Evry	bihoreau@cng.fr
Corinne	BLUGEON	IBENS	Paris	blugeon@biologie.ens.fr
Anne	BOLAND	IG CNG	Evry	boland@cng.fr
Maurine	BONABAUD	MGX	Montpellier	maurine.bonabaud@mgx.cnrs.fr
Christiane	BOUCHIER	Institut Pasteur	Paris	bouchier@pasteur.fr
Remi	BOUNON	IG CNG	Evry	bounon@cng.fr
Laurent	BOURI	GenOuest	rennes	laurent.bouri@irisa.fr
Jocelyn	BRAYET	Institut Curie	Paris	jocelyn.brayet@curie.fr
Véronique	BRUNAUD	POPS	Orsay	brunaud@evry.inra.fr
Dominique	BRUNEL	IG CNG	Evry	dbrunel@versailles.inra.fr
Christophe	CARON	ABiMS	Roscoff	christophe.caron@sb-roscoff.fr
Smahane	CHALABI	IG CNG	Evry	schalabi@cng.fr
Sophie	CHANTALAT	IG CNG	Evry	chantalat@cng.fr
Mathilde	CLEMENT	France Génomique	Nice-Sophia Antipolis	mathilde.clement@sophia.inra.fr
Jean-Yves	COPPEE	Institut Pasteur	Paris	jycoppee@pasteur.fr
Fanny	COULPIER	IBENS	Paris	coulpier@biologie.ens.fr
Corinne	CRUAUD	IG Genoscope	Evry	cruaud@genoscope.cns.fr
Stéphane	CRUVEILLER	IG Genoscope	Evry	scruveil@genoscope.cns.fr
Martine	DA ROCHA	Plateforme de génomique fonctionnelle	Nice-Sophia Antipolis	mdarocha@sophia.inra.fr
Léo	D'AGATA	IG Genoscope	Evry	ldagata@genoscope.cns.fr
Delphine	DAIAN	IG CNG	Evry	bacq@cng.fr
Maëlle	DAUNESSE	IBENS	Paris	daunesse@biologie.ens.fr
Irwin	DAVIDSON	IGBMC	Strasbourg	irwin@igbmc.fr
Thibault	DAYRIS	eBio	Orsay	thibault.dayris@i2bc.paris-saclay.fr
Jean -François	DELEUZE	IG CNG	Evry	deleuze@cng.fr
Céline	DERBOIS	IG CNG	Evry	derbois@cng.fr
Marie-Agnès	DILLIES	Institut Pasteur	Paris	marie-agnes.dillies@pasteur.fr
Cécile	DONNADIEU	GeT PlaGe	Toulouse	cecile.donnadieu@toulouse.inra.fr
Marion	DUBARRY	IG Genoscope	Evry	mdubarry@genoscope.cns.fr
Stefan	ENGELEN	IG Genoscope	Evry	sengelen@genoscope.cns.fr
Sébastien	FAYE	IG Genoscope	Evry	sfaye@genoscope.cns.fr
Nicolas	FERNANDEZ	TGML	Marseille	nicolas.fernandez-nunez@inserm.fr
Cyril	FIRMO	TGML	Marseille	firmo.cyril@gmail.com
Marie-Laure	FRANCHINARI	Migale	Jouy-en-Josas	marie-laure.franchinard@jouy.inra.fr
Jean Guillaum	GARNIER	IG CNG	Evry	garnier@cng.fr
Daniel	GAUTHERET	eBIO	Orsay	daniel.gautheret@u-psud.fr
Jean-François	GIBRAT	IFB-core	Gif-sur-Yvette	jean-francois.gibrat@france-bioinformatique.fr
Romain	GLANDIER	IG Genoscope	Evry	rglandie@genoscope.cns.fr
Alexis	GROPPI	CBiB	Bordeaux	alexis.groppi@u-bordeaux.fr
Samia	GUENDOUZ-S	MGX	Montpellier	samia.guendouz@mgx.cnrs.fr
Isabelle	GUIGON	CRIStAL	Lille	isabelle.guigon@univ-lille1.fr
Claudia	HERIVEAU	GenOuest	Rennes	claudia.heriveau@irisa.fr
Claire	HOEDE	GenoToul	Toulouse	claire.hoede@toulouse.inra.fr
Karine	HUGOT	Plateforme de génomique fonctionnelle	Nice-Sophia Antipolis	karine.hugot@paca.inra.fr
Philippe	HUPÉ	Institut Curie	Paris	philippe.hupe@curie.fr
Frederic	JARLIER	Institut Curie	Paris	frederic.jarlier@curie.fr
Yan	JASZCZYSZYN	I2BC	Gif-sur-Yvette	Yan.JASZCZYSZYN@i2bc.paris-saclay.fr
Céline	JEZIORSKI	GeT PlaGe	Toulouse	celine.jeziorski@toulouse.inra.fr
Bernard	JOST	IGBMC	Strasbourg	jost@igbmc.fr
Laurent	JOURDREN	IBENS	Paris	jourdren@biologie.ens.fr
Laurent	JOURNOT	MGX	Montpellier	laurent.journot@mgx.cnrs.fr
Renaud	JULLIEN	GenOuest	Rennes	renaud.jullien@irisa.fr
Céline	KEIME	IGBMC	Strasbourg	keime@igbmc.fr
Sean	KENNEDY	Institut Pasteur	Paris	sean.kennedy@pasteur.fr
Artem	KOURLAIEV	IG Genoscope	Evry	akourlai@genoscope.cns.fr
Claire	KUCHLY	GeT-PlaGe	Toulouse	claire.kuchly@toulouse.inra.fr
Karine	LABADIE	IG Genoscope	Evry	klabadie@genoscope.cns.fr

Joel	LACHUER	ProfileXpert	Lyon	lachuer@univ-lyon1.fr
Sonia	LAMEIRAS	Institut Curie	Paris	sonia.lameiras@curie.fr
Pierre	LE BER	IG	Evry	pleber@genoscope.cns.fr
Stéphane	LE CROM	IBENS	Paris	lecrom@biologie.ens.fr
Edith	LE FLOCH	IG CNG	Evry	edith.lefloch@cng.fr
Marie-Christir	LE PASLIER	IG CNG	Evry	le.paslier@cng.fr
Kevin	LEBRIGAND	Plateforme de génomique fonctionnelle	Nice-Sophia Antipolis	lebrigand@ipmc.cnrs.fr
Doris	LECHNER	IG CNG	Evry	lechner@cng.fr
Aurélie	LEDUC	IG CNG	Evry	leduc@cng.fr
Rachel	LEGENDRE	Institut Pasteur	Paris	rachel.legendre@pasteur.fr
Arnaud	LEMAINQUE	IG Genoscope	Evry	alemainque@genoscope.cns.fr
Sophie	LEMOINE	IBENS	Paris	slemoine@biologie.ens.fr
Alban	LERMINE	Institut Curie	Paris	alban.lermine@curie.fr
Fabrice	LOPEZ	TGML	Marseille	fabrice.lopez@inserm.fr
Valentin	LOUX	Migale	Jouy-en-Josas	valentin.loux@jouy.inra.fr
Laurence	MA	Institut Pasteur	Paris	laurence.ma@pasteur.fr
Virginie	MAGNONE	Plateforme de génomique fonctionnelle	Nice-Sophia Antipolis	magnone@ipmc.cnrs.fr
Christophe	MALABAT	Institut Pasteur	Paris	christophe.malabat@pasteur.fr
Elodie	MARQUAND	IG CNG	Evry	emarquand@versailles.inra.fr
Veronique	MARTIN	MIGALE	Jouy-en-Josas	Veronique.Martin@jouy.inra.fr
Claudine	MÉDIGUE	IG Genoscope	Evry	cmedigue@genoscope.cns.fr
Karine	MERIENNE	LNCA	Strasbourg	karine.merienne@unistra.fr
Lilia	MESROB	IG CNG	Evry	lilia.mesrob@cng.fr
Vincent	MEYER	IG CNG	Evry	vmeyer@cng.fr
Denis	MILAN	GeT PlaGe	Toulouse	milan@toulouse.inra.fr
Dario	MONACHELLC	POPS	Orsay	dario.monachello@evry.inra.fr
Delphine	NAQUIN	I2BC	Gif-sur-Yvette	delphine.naquin@i2bc.paris-saclay.fr
Catherine	NGUYEN	TGML	Marseille	catherine.nguyen@inserm.fr
Alain	NICOLAS	Institut Curie	Paris	alain.nicolas@curie.fr
Benjamin	NOEL	IG Genoscope	Evry	bnoel@genoscope.cns.fr
Christine	OGER	PRABI	Lyon	christine.oger@univ-lyon1.fr
Robert	OLASO	IG CNG	Evry	olaso@cng.fr
Marie-Ange	PALOMARES	IG CNG	Evry	palomares@cng.fr
Hugues	PARRINELLO	MGX	Montpellier	hugues.parrinello@mgx.cnrs.fr
Benjamin-Edo	PEIGNEY	IG CNG	Evry	peigney@cng.fr
Guy	PERRIÈRE	PRABI	Lyon	guy.perriere@univ-lyon1.fr
Sandrine	PERRIN	IFB	Gif-sur-Yvette	sandrine.perrin@france-bioinformatique.fr
Alexandra	POPA	Plateforme de génomique fonctionnelle	Nice-Sophia Antipolis	popa@ipmc.cnrs.fr
Juliette	POUCH	IBENS	Paris	pouch@biologie.ens.fr
Corinne	RANCUREL	Plateforme de génomique fonctionnelle	Nice-Sophia Antipolis	corinne.rancurel@sophia.inra.fr
Richard	REDON	Biogenouest	Nantes	richard.redon@univ-nantes.fr
Béatrice	REGNAULT	Institut Pasteur	Paris	beatrice.regnault@pasteur.fr
Florence	RIBIERRE	IG CNG	Evry	ribierre@cng.fr
Claire	RIOUALEN	TGML	Marseille	claire.rioualen@inserm.fr
Eric	RIVALS	ATGC	Montpellier	rivals@lirmm.fr
Diana	RUSSO	IG CNG	Evry	diana.russo@cng.fr
Laurent	SACHS	MNHN	Paris	laurent.sachs@mnhn.fr
Gérald	SALIN	GeT-PlaGe	Toulouse	gerald.salin@toulouse.inra.fr
Florian	SANDRON	IG CNG	Evry	sandron@cng.fr
Sebastien	SANTINI	IGS	Marseille	santini@igs.cnrs-mrs.fr
Claude	SCARPELLI	IG Genoscope	Evry	claude@genoscope.cns.fr
Joseph	SCHACHERER	GMGM	Strasbourg	schacherer@unistra.fr
Raphaël	SCHNEIDER	IGBMC	Strasbourg	raphaelschneider@unistra.fr
Béatrice	SEGURENS	IG CNG	Evry	segurens@cng.fr
Marine	SÉJOURNÉ	IG Genoscope	Evry	msejour@genoscope.cns.fr
Nicolas	SERVANT	Institut Curie	Paris	Nicolas.Servant@curie.fr
Dany	SEVERAC	MGX	Montpellier	dany.severac@mgx.cnrs.fr
Pascal	SIRAND-PUGN	PGTB	Bordeaux	sirand@bordeaux.inra.fr
Odile	SISMEIRO	Institut Pasteur	Paris	osisme@pasteur.fr
Eric	SOLARY	Institut Gustave Roussy	Villejuif	Eric.SOLARY@gustaveroussy.fr
Ludivine	SOUBIGOU-TA	POPS	Orsay	soubigou@evry.inra.fr
Jean-Francois	TALY	PRABI	Lyon	jftaly@gmail.com
Francois-Xavie	THEODULE	TGML	Marseille	theodule@tagc.univ-mrs.fr
Claude	THERMES	I2BC	Gif-sur-Yvette	claude.thermes@i2bc.paris-saclay.fr
Christelle	THIBAULT-CA	IGBMC	Strasbourg	thibault@igbmc.fr
Nizar	TOULEIMAT	IG CNG	Evry	nizar.touleimat@cng.fr
Hélène	TOUZET	Bilille	Lille	helene.touzet@univ-lille1.fr
Erwin	VAN DIJK	I2BC	Gif-sur-Yvette	erwin.vandijk@i2bc.paris-saclay.fr
Hugo	VARET	Institut Pasteur	Paris	hugo.varet@pasteur.fr
Amandine	VELT	IGBMC	Strasbourg	velt@igbmc.fr
Sophie	VIVIER	MGX	Montpellier	Sophie.Vivier@mgx.cnrs.fr
Nicolas	WIART	IG CNG	Evry	nicolas.wiart@cng.fr
Patrick	WINCKER	IG Genoscope	Evry	pwincker@genoscope.cns.fr
Tao	YE	IGBMC	Strasbourg	yetao@igbmc.fr