

Fonctionnement et recommandations de la plateforme GeT-Biopuces

Laboratoire d'Ingénierie des Systèmes
Biologiques et des Procédés (LISBP)

INSA

Bâtiment 39 -Bio 5

135 Avenue de Ranguel
31077 Toulouse Cedex 04

Directrice de la plateforme:

Marie Ange Teste : Marie-Ange.Teste@insa-toulouse.fr,

Responsables technologies microarray et séquençage:

Lidwine Trouilh : lidwine.trouilh@insa-toulouse.fr

Nathalie Marsaud : nathalie.marsaud@insa-toulouse.fr

Responsable analyses de données :

Delphine Labourdette : delphine.labourdette@insa-toulouse.fr

Mail plateforme : biopuces@insa-toulouse.fr

Téléphone : +33 (0)5-61-55-96-87

REMARQUES IMPORTANTES

- Durant le projet et selon leur type, les échantillons seront stockés par la plateforme au congélateur ou au réfrigérateur. **Le demandeur est invité à récupérer ses échantillons dans un délai d'un mois après la fin de la réalisation de la prestation. Passé ce délai, nous ne garantissons plus leur conservation.** Bien que non reliés à un système d'alarme, la température de nos congélateurs et réfrigérateurs est relevée hebdomadairement.
- Nos micro-pipettes sont contrôlées annuellement par une société extérieure.
- A partir du moment où **le contrôle qualité des échantillons est validé et s'il n'y a pas de problème au cours de la prestation**, les expériences sont **généralement réalisées dans un délai de 4 semaines.**
- **La durée de stockage des données informatiques**, scan, analyse d'image, analyses statistiques et analyses de séquençage **est de 1 an (sauf avis contraire de votre part).**
- Les lames de verre spottées sont envoyées par courrier postal. Les résultats sous forme de données sont mis à disposition sur le site web de la plateforme via l'adresse qui vous sera envoyée par mail. **Nous vous remercions de bien vouloir accuser réception de vos données et/ou échantillons.**
- Les informations relatives à toutes les manipulations faites sur vos échantillons sont archivées dans votre dossier client à la plateforme.
- **Nous vous rappelons que le devis est valable pour les prestations et quantités mentionnées dans celui-ci, toutes les expériences refaites ou supplémentaires feront l'objet d'une nouvelle facturation.**
- **Toute prestation facturée et non réalisée dans un délai de 12 mois ne sera pas remboursée.** Passé ce délai, si vous souhaitez faire la prestation, elle sera réévaluée en fonction des tarifs, consommables et personnels en vigueur et un devis complémentaire sera réalisé.
- Merci de prendre connaissance du plan d'évacuation de la plateforme GeT Biopuces

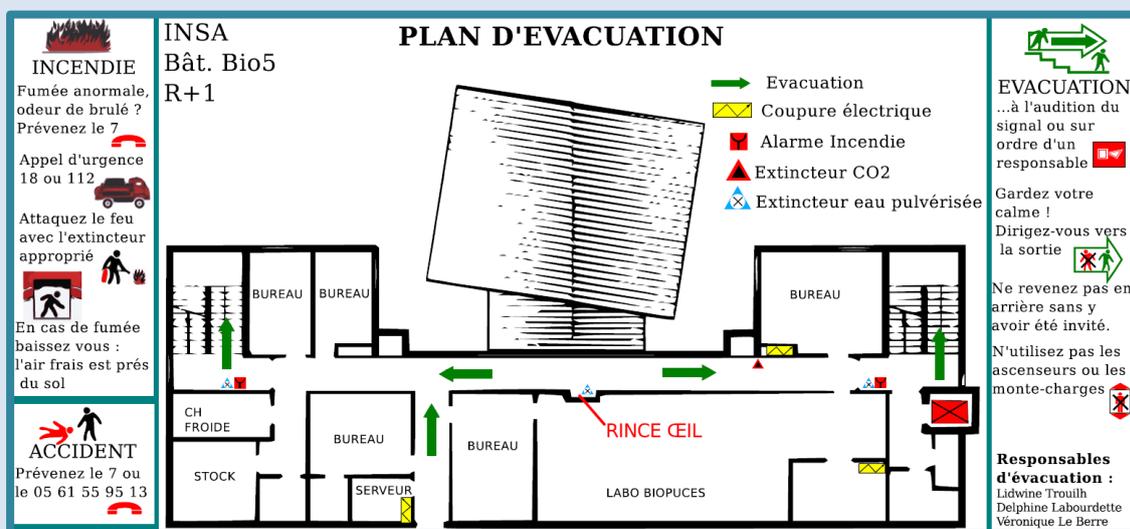


TABLE DES MATIERES

1. GENERALITES.....	4
2. SPOTTING	4
2.1. FABRICATION DES PUCES	4
2.2. CONTROLE QUALITE DU SPOTTING	5
2.3. STOCKAGE DES LAMES	5
3. CONTROLE QUALITE DES ARNS ET ADNS.....	5
3.1. ADNS	6
3.2. ARNs.....	6
4. MICROARRAY	7
4.1. DESIGN DE SONDAS	7
4.2. MARQUAGE ET CONTROLE DU MARQUAGE.....	8
4.3. HYBRIDATION	9
4.4. SCAN.....	9
5. SEQUENÇAGE	9
6. ANALYSES	11
6.1. ANALYSE D'IMAGES DES PUCES A ADN	11
6.2. CONTROLE QUALITE HYBRIDATIONS DES PUCES A ADN OU SEQUENÇAGE	11
6.3. ANALYSE BIOINFORMATIQUE DES DONNEES DE SEQUENÇAGE (QUELLE QUE SOIT LEUR ORIGINE).....	11
6.4. ANALYSE STATISTIQUES DES DONNEES (<i>QUELLE QUE SOIT LEUR ORIGINE</i>)	12
6.4.1. <i>Puces à ADN</i>	12
6.4.2. <i>RNAseq</i>	13
6.5. ENRICHISSEMENT FONCTIONNEL	13
7. RENDU DES RESULTATS.....	14
8. REGLES DE CONFIDENTIALITE, DE PROPRIETE INTELLECTUELLE ET DE VALORISATION	14

1. GENERALITES

La plateforme GeT-Biopuces réalise vos expériences depuis le design des sondes jusqu'à l'analyse des données. Elle met à votre disposition l'équipement nécessaire pour réaliser vos expériences de puces à ADN et de séquençage ainsi que le personnel pour analyser vos données quelle que soit leur provenance.

Avant tout démarrage de projet, il est nécessaire de contacter soit :

- la responsable de la plateforme Marie Ange Teste (Marie-Ange.Teste@insa-toulouse.fr)
- les responsables microarray et séquençage, Nathalie Marsaud (nathalie.marsaud@insa-toulouse.fr) ou Lidwine Trouilh (lidwine.trouilh@insa-toulouse.fr)
- la responsable analyse de données, Delphine Labourdette (delphine.labourdette@insa-toulouse.fr) pour le design des sondes, les analyses bioinformatiques de données de séquençage, l'analyse statistique des données de microarray et RNAseq ainsi que les analyses d'enrichissement fonctionnel.

Afin de répondre au mieux à votre demande, nous définissons ensemble le contenu des expériences à réaliser. A l'issue de notre discussion, nous vous communiquons une proposition de facture ainsi que la fiche prestation. Cette dernière (accompagnée du devis signé et du bon de commande associé) doit être complétée, signée et envoyée à la Plateforme GeT-Biopuces avant le début des expériences. Vous pouvez nous l'envoyer par courrier avec vos échantillons ou par mail après l'avoir scannée.

Une fois reçus, vos échantillons seront contrôlés. S'ils ne répondent pas à nos critères qualités, nous vous en informons et nous attendons que vous nous fassiez parvenir des échantillons permettant de réaliser votre demande dans de meilleures conditions. Après validation des échantillons, vos expériences sont planifiées puis réalisées et vos résultats vous sont envoyés dans un délai moyen de 1 mois.

Bien entendu, si à une étape nous rencontrons des difficultés, nous vous en faisons part et nous décidons ensemble de la suite à mener.

2. SPOTTING

2.1. Fabrication des puces

Nous pouvons réaliser vos puces à façon après avoir défini ensemble leur contenu, le système biologique, le type de sondes, le type de plaques de dépôt, le nombre de dépôts, le choix du tampon de dépôt, le type et la quantité de lames à produire.

Vous pouvez aussi nous fournir les plaques contenant les sondes à spotter ainsi que les lames support des puces pour votre organisme d'étude.

Le robot spotteur utilisé pour le dépôt des oligonucléotides ou protéines sur lames de verre est le « QArray mini » (Genetix). Il est équipé au maximum de 48 aiguilles, creuses (Telechem SMP3) ou plates (Telechem SSP015).

Les lames que nous utilisons possèdent un code-barres permettant une identification unique et l'orientation de chaque lame. En routine, les lames sont spottées avec le code-barres vers le bas.

Une fois le spotting terminé, le fichier « *.gal » est créé par le personnel de la plateforme et est envoyé par mail. Un bon de livraison vous sera remis ou envoyé avec les lames. Pour toute demande relative au fichier « *.gal », veuillez-vous adresser à la responsable microarray et séquençage Lidwine Trouilh (lidwine.trouilh@insa-toulouse.fr) ou à la responsable analyse de données Delphine Labourdette (delphine.labourdette@insa-toulouse.fr).

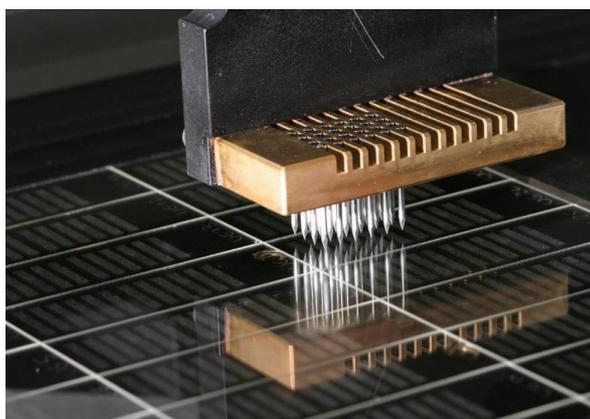


Fig 1: Robot en cours de dépôt sur lame de verre

2.2. Contrôle qualité du spotting

La qualité du spotting est déterminée par scan, 24h après la fin du dépôt. Ce scan est réalisé sur chaque lame qui a précédé et suivi le rechargement des aiguilles. Le lot de lames est validé lorsque 95% des spots attendus sont présents.

2.3. Stockage des lames

Nous recommandons de conserver les lames dans un dessiccateur à température ambiante et de les utiliser dans l'année suivant la réception.

3. CONTROLE QUALITE DES ARNS ET ADNS

Nous ne réalisons pas les extractions des ARNs ni des ADNs.

Ils doivent nous être livrés, congelés dans de l'eau.

Les contrôles qualités sont réalisés sur puces RNA 6000 Lab-on-Chip ou DNA 12000 Lab-on-Chip de chez Agilent (Bioanalyzer), sur le Nanodrop ou sur le Qubit. Ils permettent une

validation qualitative et quantitative des ARNs et des ADNs.

3.1. ADNs

Les ADNs sont d'abord contrôlés sur le Nanodrop et sur le Qubit (pour le séquençage). Le minimum requis est :

- un **ratio 260nm/230nm et 260nm/280nm \geq à 1,8**
- une **concentration minimale d'échantillon** nécessaire pour l'expérience choisie (celle-ci est fournie dans les recommandations d'utilisation du fournisseur du kit).

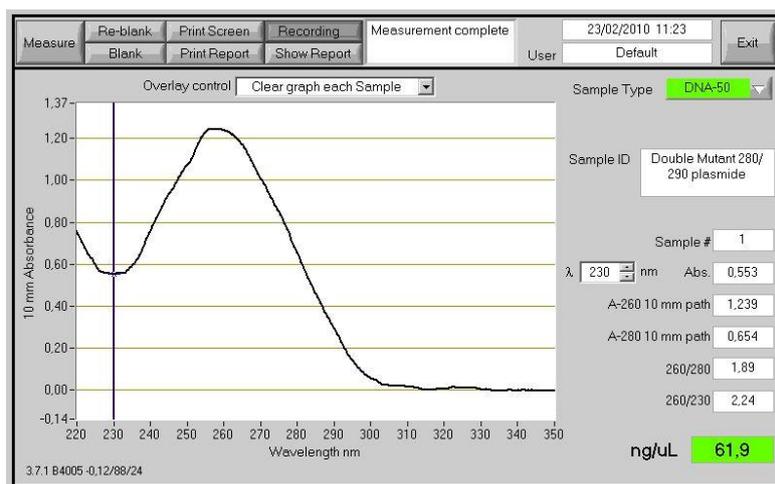


Fig 2: Profil d'ADN correct après lecture au nanodrop

Ensuite, nous demandons à ce qu'une migration des ADNs sur gel d'agarose soit réalisée. Un ADN de qualité correcte et utilisable dans une expérience de puce à ADN ou de séquençage ne doit présenter qu'une seule bande de migration dans les grandes tailles et non un smear.

Pour le ChIPseq, la taille des fragments d'ADN doit être centrée autour de 150pb.

Au vu de la complexité des ADN destinés au séquençage métagénomique, les critères qualités cités ci-dessus ne s'appliquent pas mais une quantification au Nanodrop et/ou au Qubit est nécessaire.

3.2. ARNs

En transcriptomique, nous travaillons avec des triplicats biologiques c'est à dire qu'il faut extraire les ARNs à partir de 3 individus indépendants ou à partir de 3 cultures cellulaires indépendantes. Les ARNs sont contrôlés d'abord sur le Nanodrop.

Le minimum requis est :

- un **ratio 260nm/230nm et 260nm/280nm \geq à 1,8**
- une **concentration minimale d'échantillon** nécessaire pour l'expérience choisie (celle-ci est fournie dans les recommandations d'utilisation du fournisseur du kit).

Après mesure au Nanodrop, les ARNs sont contrôlés au Bioanalyzer sur puces AGILENT RNA 6000 Lab-on-Chip.

Le minimum requis est :

- un ratio 18S/28S \geq à 1,7 pour les Eucaryotes ou un ratio 16S/23S \geq à 1,2 pour les Procaryotes
- un RIN (RNA Integrity Number) \geq 8.5.

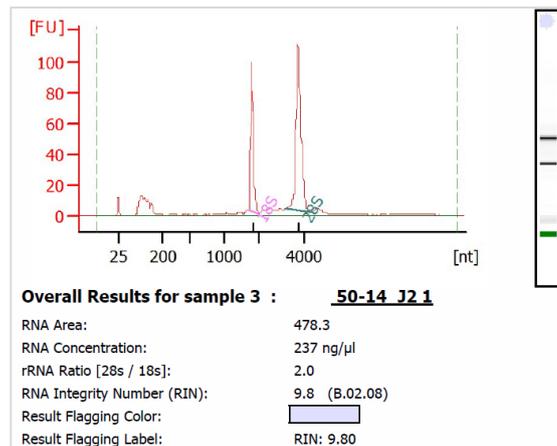


Fig 3: Profil d'ARN correct après lecture au Bioanalyzer

Un mail vous sera envoyé avec les résultats des QC. Si les ARNs ou ADNs passent nos critères qualité, la prestation est poursuivie. Dans le cas contraire, la décision de continuer ou pas, vous incombe.

4. MICROARRAY

4.1. Design de sondes

Si les puces catalogue disponibles chez les différents fournisseurs ne concernent pas votre organisme d'étude ou que vous souhaitez rajouter des sondes pour certains gènes d'intérêt, nous pouvons designer vos sondes.

Il faut tout d'abord que l'organisme que vous souhaitez étudier soit séquencé et correctement annoté de façon à avoir le maximum de chance de pouvoir interpréter efficacement les résultats. Il faut nous fournir son transcriptome ou son génome sous format « *.fasta » ou « *.txt ».

La technologie permettant de faire des puces à façon à moindre coût est la technologie Custom Gene Expression Microarray d'Agilent (synthèse directe des sondes offrant une bonne reproductibilité expérimentale).

Ce fournisseur offre une large gamme de format (entre 15000 et 1 million de sondes par zone). Le design à façon de sondes consiste à concevoir des sondes de 60 bases, spécifiques à la séquence complémentaire des gènes déterminés en amont à l'aide d'outils bio-informatiques tel que le logiciel de design en ligne eArray (<https://earray.chem.agilent.com/earray/>). Les sondes doivent, de ce fait, répondre à des critères de spécificité et de thermodynamique stricts pour garantir une bonne spécificité dans les mêmes conditions d'hybridation, en évitant la formation de structures secondaires. Les sondes dessinées seront alors testées in-silico afin de sélectionner pour chaque gène les meilleures sondes en garantissant une spécificité de reconnaissance.

Nous commandons ensuite les puces dont la livraison est assurée avec un délai de 2 mois. Une première expérience permettra de valider le design avant d'aller plus avant dans la prestation.

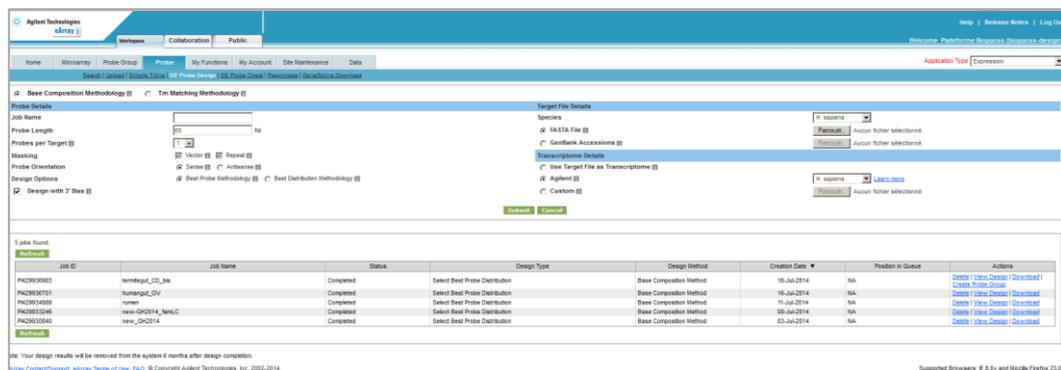


Fig 4: Interface eArray

4.2. Marquage et contrôle du marquage

Nous utilisons des kits spécifiques suivant la technologie choisie :

- pour le marquage Agilent, nous travaillons en routine avec 100ng d'ARN (concentration mini 50ng/μl) ou de 500ng d'ADNg (concentration mini 150ng/μl) pour les CGH array.
- pour Affymetrix, en routine, il faut 100ng d'ARN (concentration minimale 40ng/μl) et pour les tilling array, il faut 9 μg d'ADN amplifié (concentration minimale 200ng/μl). Nous pouvons également travailler avec de petites quantités, dans ce cas, il nous faut 500 pg d'ARN (concentration minimale 200 pg/μL).

Chaque synthèse d'ADN ou marquage est contrôlé au Nanodrop afin de déterminer la quantité de matériel produite et si besoin la quantité de fluorochrome incorporé.

S'ils ne passent pas les critères qualité minimum requis par les fournisseurs, nous vous contactons afin que vous preniez la décision de continuer ou pas.

4.3. Hybridation

Les hybridations réalisées sur la plateforme sont faites en four Agilent ou Affymetrix selon les spécifications de ces distributeurs. Si vous réalisez vous même vos hybridations manuelles dans votre laboratoire, merci de vous référer au manuel d'instruction du kit de marquage utilisé.



Fig 5: Puces Agilent (à gauche) cartouche puce gene Affymetrix (à droite)

4.4. Scan

Pour les lames Agilent et Affymetrix, nous nous référons aux spécifications fournies par les distributeurs concernés. Les scanners utilisés sont le MS200 (Tecan) pour les lames Agilent et le GS3000 (Affymetrix) pour les lames Affymetrix.

Pour les lames spottées à façon (hybridées en double couleur), nous utilisons le scanner Innoscan700 (Innopsys). Pour ce genre de lame, nous vous recommandons de scanner en choisissant les paramètres du scanner tels que les courbes d'intensités dans le canal rouge et le canal vert soient le plus superposées possible et que ces courbes « s'éteignent » autour de 50000 unités arbitraires de fluorescence sur l'axe des abscisses. Ceci évite la saturation tout en ayant une gamme dynamique de fluorescence la plus large possible.

5. SEQUENÇAGE

Le séquençage est effectué par le Ion S5 (ThermoFischer) selon les spécifications du distributeur.

Les quantités minimales de matériel au départ pour les applications majeures sont :

- pour le DNAseq : 10ng à 1µg d'ADN à une concentration entre 1 et 30ng/µl
- pour le RNAseq : 5µg d'ARN total minimum voire 10µg si possible à une concentration minimale 250ng/µL. Ou si vous vous chargez de la sélection des polyA ou de la ribodéplétion: 1 à 500 ng d'ARN-poly(A) ou 10 à 500ng d'ARN déplétés en ARN ribosomaux (concentration minimale de 1-50ng/µL)

- pour le miRNAseq : 0.5-20 µg d'ARN total soit une concentration de 10ng/µL ou 100ng de small RNA (taille <200bp) soit une concentration de 35ng/µL.
- pour l'Ampliseq : de 10 à 40 ng d'ADN génomique minimum (variable selon les kits)
- pour le ChIPseq : 10ng d'ADN fragmenté à 200pg/µL minimum
- pour la métagénomique : 10µL d'ADN extrait, peu importe la quantité car en métagénomique l'ADN procaryote est souvent plus ou moins contaminé par de l'ADN eucaryote.

Après contrôle qualité des échantillons, les étapes précédant le séquençage sont :

- la synthèse de librairie qui est la fixation des adaptateurs de séquençage sur les fragments d'ADN ou d'ARN rétrotranscrit.
- la PCR en émulsion qui permet de fixer et d'amplifier la librairie à séquencer (ou le pool de librairies à séquencer dans le cas de multiplexage) sur des billes
- L'enrichissement des billes « actives » (celles qui possèdent des fragments d'ADN)

Chaque étape est validée par un contrôle qualité au « bioanalyzer » sur puces Agilent et/ ou au Qubit afin de déterminer la qualité, la taille des fragments et la quantité de matériel produite.

La puce de séquençage chargée avec la polymérase et les billes « actives ». Nous disposons de trois formats de puces permettant de générer plus ou moins de séquences en 200 bp ou 400 bp.

	Output	Reads	Read Length
Ion 520™ Chip	1-2 Gb	3-5 M	Up to 400bp
Ion 530™ Chip	3-5 Gb	15-20 M	Up to 400bp
Ion 540™ Chip	10-15 Gb	60-80 M	Up to 200bp

Fig 6: Types de puces de séquençage sur le S5

La puce est alors insérée dans le Ion S5 pour être séquençée avec le programme approprié à l'application demandée.

A la fin du séquençage, un rapport du run permettant de vérifier le bon déroulement du séquençage est édité au format .pdf.

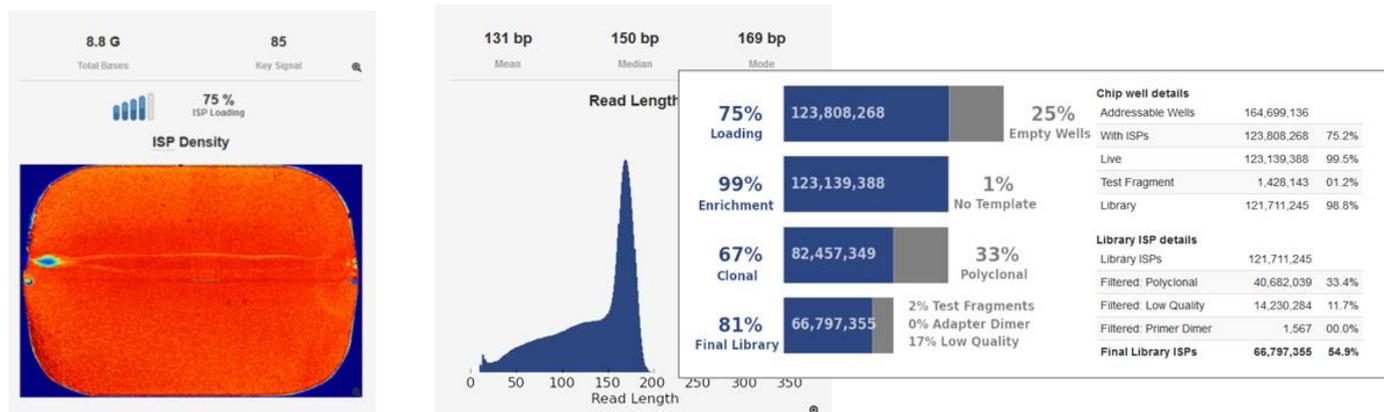


Fig 7: Puce chargée, taille des fragments séquencés, contrôle qualité du séquençage (de gauche à droite)

6. ANALYSES

6.1. Analyse d'images des puces à ADN

Les analyses d'images de puces à ADN peuvent être réalisées sur la plateforme par les responsables microarray et séquençage.

Les logiciels utilisés sont Mapix (Innopsys), Feature Extraction (Agilent), et Command Console (Affymetrix). Chacun permet d'éditer à partir des images de scan, des fichiers résultats qui peuvent être exploités avec le logiciel de votre choix.

6.2. Contrôle qualité hybridations des puces à ADN ou séquençage

Pour les puces « expression », un contrôle qualité est effectué après les hybridations et le scan. Un rapport QC vous est ensuite envoyé par mail accompagné des données brutes des hybridations et de graphiques au format png.

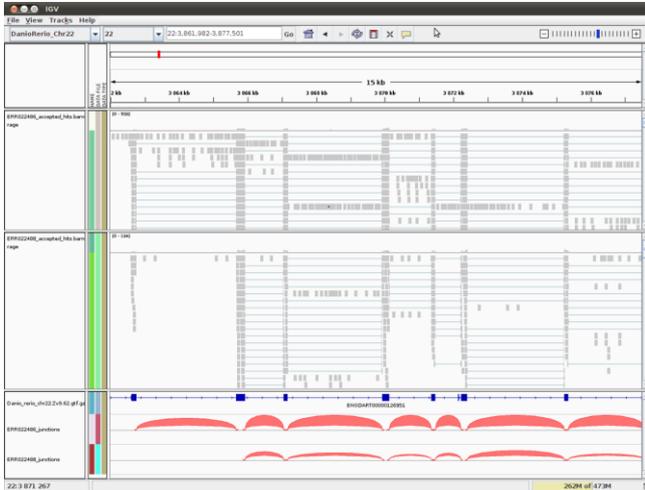
Pour les puces « génomiques », à la fin de la prestation, nous vous envoyons dans un mail détaillé, l'adresse du lien où l'ensemble des données résultats sont téléchargeables.

Un contrôle qualité est également effectué après le séquençage. Les fichiers FastQC vous sont envoyés et vos données peuvent être récupérées sur un disque dur (non fourni par la plateforme). Les données sont conservées un an sur la plateforme.

6.3. Analyse bioinformatique des données de séquençage (quelle que soit leur origine)

Quelle que soit la provenance de vos données de séquençage (GeT-Biopuces ou organisme extérieur), l'analyse des données est faite à façon.

La prestation peut comprendre l'analyse qualité des données, un assemblage *de novo* de votre génome séquencé, le mapping et la reconstruction de transcrits sur un génome de référence ainsi qu'une mesure « brute » de l'expression au niveau transcrit/gène/exon selon la demande, la recherche de SNP sur un génome ou des régions ciblées.



Les alignements sont remis au format bam (fichiers compressés), le listing des transcrits/gènes au format gtf (fichiers textes), les quantifications au format texte et les analyses de variants au format vcf.

Fig 8: Exemple de résultats d'analyse bioinformatique de RNAseq

6.4. Analyse statistiques des données (*quelle que soit leur origine*)

6.4.1. Pucés à ADN

Quelle que soit la provenance de vos données de puces ADN (GeT-Biopuces ou organisme extérieur), l'analyse statistique est réalisée à l'aide du logiciel R et des packages Bioconductor. Le résultat de l'analyse des puces (Agilent, Affymetrix) est remis sous la forme d'un rapport contenant le protocole expérimental, les contrôles qualités avant et après normalisation, une explication sur le choix de la normalisation effectuée, la liste des gènes différentiellement exprimés ou non ainsi que les graphiques de l'analyse exploratoire (Analyse en Composante Principale et Heatmap).

De plus, les graphiques sont remis au format png ou pdf, les données brutes et les données normalisées et annotées, ainsi que la liste des gènes différentiellement exprimés ou non au format Excel et pdf.

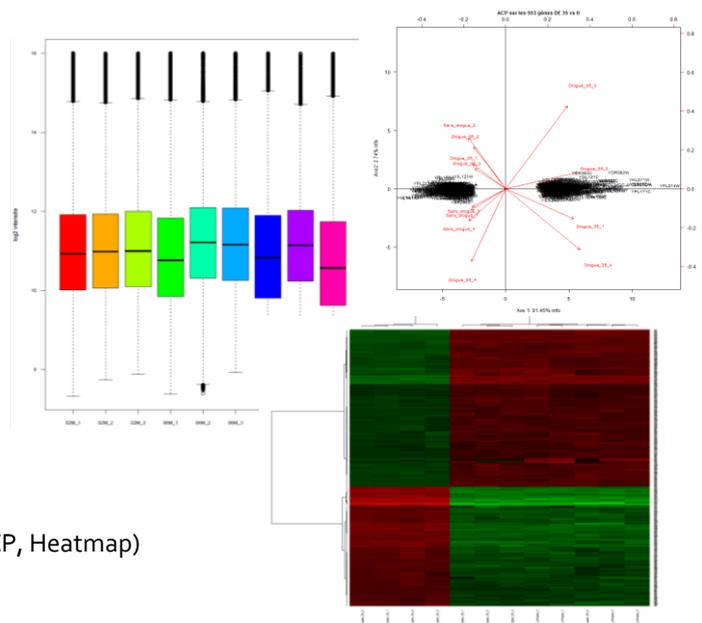


Fig 9: Exemple de graphes (Boxplots, ACP, Heatmap)

6.4.2. RNAseq

Quelle que soit la provenance de vos données RNAseq (GeT-Biopuces ou organisme extérieur), l'analyse statistique des données est réalisée par à l'aide du logiciel R et des packages Bioconductor. Le résultat de l'analyse est remis sous la forme d'un rapport contenant le protocole expérimental, les contrôles qualités avant et après normalisation, une explication sur le choix de la normalisation effectuée, la liste des gènes différentiellement exprimés ou non ainsi que les graphiques de l'analyse exploratoire (Analyse en Composante Principale et Heatmap).

De plus, les graphiques sont remis au format png ou pdf, les données brutes et les données normalisées, ainsi que la liste des gènes différentiellement exprimés ou non au format Excel et pdf.

6.5. Enrichissement fonctionnel

L'enrichissement fonctionnel est réalisé partir des listes de gènes différentiellement exprimés obtenus à partir des données de microarray, selon la demande sur les trois grandes familles de la Gene Ontology : « biological process », « cellular component », « molecular function » et sur KEGG. Cette partie est réalisée à l'aide du logiciel R et des packages Bioconductor. Le résultat de l'analyse est délivré sous la forme d'un rapport expliquant la procédure et les résultats. Les graphiques sont également remis sous format png, Excel, html, ... selon la demande du client.

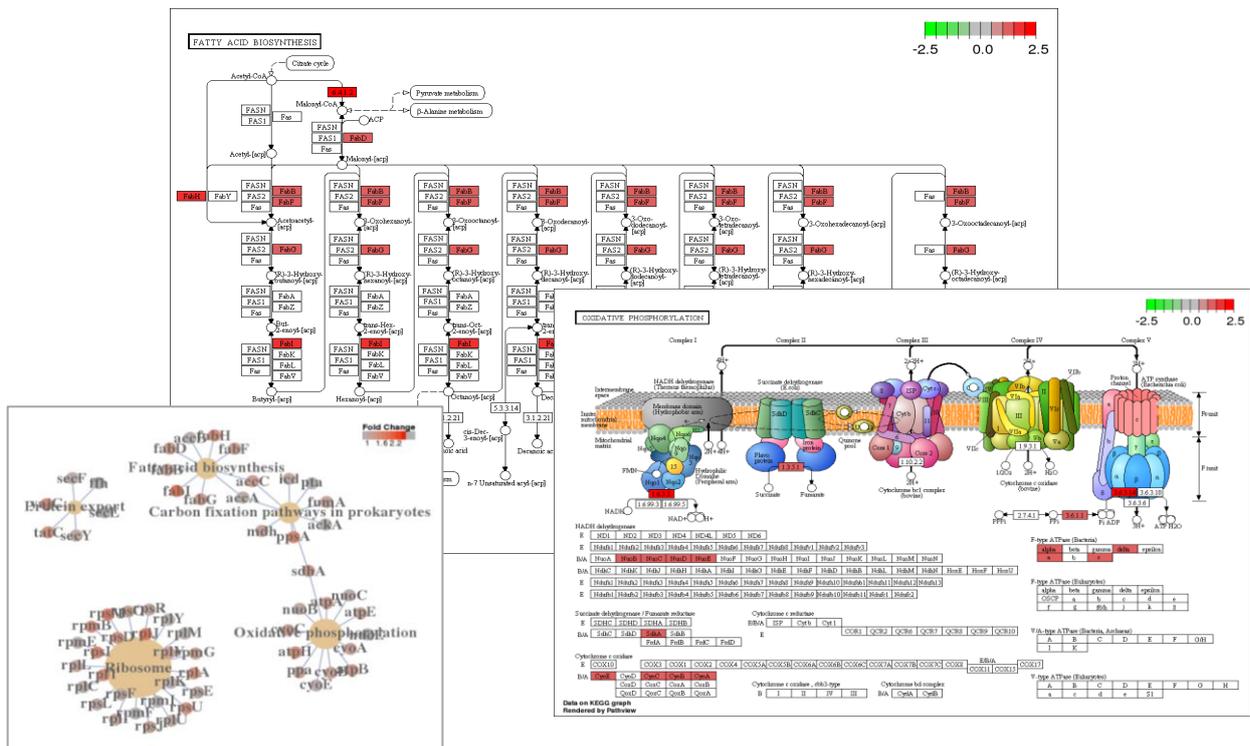


Fig 10: Exemple de résultats d'enrichissement fonctionnel KEGG

7. RENDU DES RESULTATS

Excepté pour l'analyse bioinformatique des données de séquençage, l'ensemble des éléments décrits dans les paragraphes analyse sont mis à disposition sur le site web de la plateforme, par le biais d'un lien sécurisé. Toutes les informations nécessaires ainsi que l'adresse de l'accès à vos données vous seront envoyées par mail.

Pour information, les fichiers sont seulement téléchargeables pendant 7 jours à compter de leur mise à disposition à l'URL indiquée.

Attention, il arrive que malgré de bons contrôles qualités, les analyses ne permettent pas d'obtenir de liste de gènes différentiellement exprimés. Ceci ne remet pas en cause la qualité des résultats obtenus mais cela peut être le reflet d'une réalité biologique des échantillons. De même, il peut arriver que l'enrichissement fonctionnel ne fasse ressortir aucune catégorie fonctionnelle significativement enrichie avec une p value correcte.

8. REGLES DE CONFIDENTIALITE, DE PROPRIETE INTELLECTUELLE ET DE VALORISATION

Dans le cadre d'une prestation de service, vous avez la propriété exclusive de vos résultats ainsi que leur maîtrise et leur valorisation. Cependant, la plateforme GeT-Biopuces de la Génopole Toulouse Midi Pyrénées doit être citée *a minima* dans les remerciements des publications et travaux de thèse qui découlent de l'utilisation des ressources de la plateforme. L'adresse à utiliser pour l'ensemble des publications est : LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France. Une copie des papiers présentant les travaux citant la plateforme sera appréciée.

Afin de protéger la confidentialité de vos projets, dès la première étape de discussion, nous sommes en mesure avec l'aide des juristes du SAIC de l'INSA de signer des accords de confidentialité.

De plus le SAIC négocie et exécute les accords et conventions à caractère industriel et commercial, en particulier, les contrats d'essais, de recherche, d'études, d'analyses, de conseils et d'expertises que nous pourrions effectuer pour le compte de tiers. Les entreprises peuvent bénéficier du « crédit d'impôt recherche »

Dans la mesure où votre projet nécessite une part de développement (technologique, protocole, développement de programmes etc...) effectué par le personnel de la plateforme, lors de toutes formes de valorisation (publications, brevets, posters, etc...) des résultats obtenus, ce personnel sera cité dans les auteurs.