

BIOINFODIAG

GT RNA-seq

GT3 Question biologique

Planning et participants:

- **Réunion 1 (14/5/2023, à Rouen)** : PA Rollat, AS Denommé-Pichon, C. Houdayer, S. Gorokhova, F. Lecoquierre, W. Carré, J. Muller, C. Van Goethem, Céline Derambure et Myriam Vezain ... liste à compléter ?!
- **Réunion 2 (06/10/2023, en visio)** : Alexandra Lespagnol, Charles Van Goethem, Christophe Habib, John Rendu, Svetlana Gorokhova, Wilfried Carré, Anne-Sophie Jourdain, Marion Larrieux, Christel Vaché, Perrine Brunelle, Raphael Leman, Jean Muller.
- **Réunion 3 (06/12/2023, en visio)** : Perrine Brunelle, Wilfrid Carré, Anne-Sophie Denommé-Pichon, Christophe Habib, Jean Muller, Pierre-Antoine Rollat-Farnier, Charles van Goethem, Svetlana Gorokhova

Objectifs : Définir les besoins du point de vue du biologiste pour les analyses RNASeq

GT3 Question biologique

Données au préalable

- Quels gènes/pseudogènes sont exprimés dans quels tissus ? (e.g. GTEx, HPA)
- Données liées au gène (isoforme spécifique, nb isoformes, taille CDS, nb exons, taille UTR, empreintes) sources : ensembl/RefSeq (pas UCSC), geneimprint.org
- Choix de l'isoforme.
- Génome de référence (GRCh38)
- Annotations des gènes partenaires ou down/upstream régulés (i.e. voies de signalisation)
- Zones complexes du génome (Mapping, RNASeq spécifique ?)
- Metadata (tissu, age, pathologie, duplicat...) (lexique?)



Rq : fort potentiel de fichiers à **fournir/partager** par BioInfoDIAG pour la communauté.

GT3 Question biologique

Metrics Qualité run/patients

- Control qualité séquenceur/run (e.g. FastQC)
- Duplicats, On target, (e.g. MultiQC), détection de la contamination ADN génomique
- Distribution de la profondeur de lecture sur le gène (e.g. kmerexplor)
- Seuil de profondeur de lecture sur le gène et sur les exons
- Identitovigilance (à voir avec kmer pour SNP/SRY) avant alignement (à voir avec Sylvain Mareschal)

Recherche de Fusions

- Événements connus et *de novo*, intra et interchromosomiques

GT3 Question biologique

Expression génique

- Gènes différentiellement exprimés (up and down) -> outliers
- Isoformes
- Signature transcriptomique (multigènes)
- Transcriptome Tumeur vs Référence, Tissu vs Référence

Anomalies épissage

- Altération partielle ou totale de jonctions exon (disparition, ajout)
- Comparaison à une référence (GTEx/interne/population générale -> SpliceVault). Recherche des outliers.



Rq : Large délétion intragénique peut donner une impression d'isoforme plus courte

GT3 Question biologique

Détection des SNV/indel

- Détection à proprement dite sur le transcriptome
- Somatique: comparaison avec le constitutionnel
- Détermination du génotype
- Détection fraction allélique faible (dilution liée aux isoformes multiples et inconnues)
- Notion de zones « dark » (Alignement ? mauvaise expression ? difficulté d'amplification ? NMD ?)
- Importance de la disponibilité (intégration ?) des CNV en constit si disponible pour interprétation du RNASeq

Expression monoallélique

- Détection de zones génomiques (limitées au gène voire plusieurs gènes) contiguës (partielles/totales)
- Nombre et statuts de variations détectées (het vs hom) pour un gène (constitutionnel vs gnomAD) avec l'idée de contrôle qualité => Analyse faisable ou pas.



Rq : anomalies de profondeur de lecture dans les gènes/zones génomiques (i.e. UTR) adjacents.

GT3 Question biologique

Visualisation(s)

- Besoin d'un outil intégré d'analyse incluant **des données d'expression et du génome**
- Besoin de l'intégration de **metadata** (tissu, âge, pathologie, duplicat...)
- Visualisation **QC** (Utilisation MultiQC à priori mature)
- Visualisation des **transcrits de fusion**
- Visualisation des **jonctions exons** type Sashimi plot
- Visualisation des données **d'expression différentielle**:
 - Type volcano plot pour la visualisation des outliers
 - Heatmap (design avec multiple conditions)



- Exemples (liste non exhaustive):
 - TGG Viewer (Broad, <https://tgg-viewer.broadinstitute.org/>)
 - IGV (<https://igv.org/>)
 - Sashimi_plot (<https://miso.readthedocs.io/en/fastmiso/sashimi.html>)
 - ggsashimi (<https://github.com/guigolab/ggsashimi/>)
 - splicejam (<https://jmw86069.github.io/splicejam/index.html>)