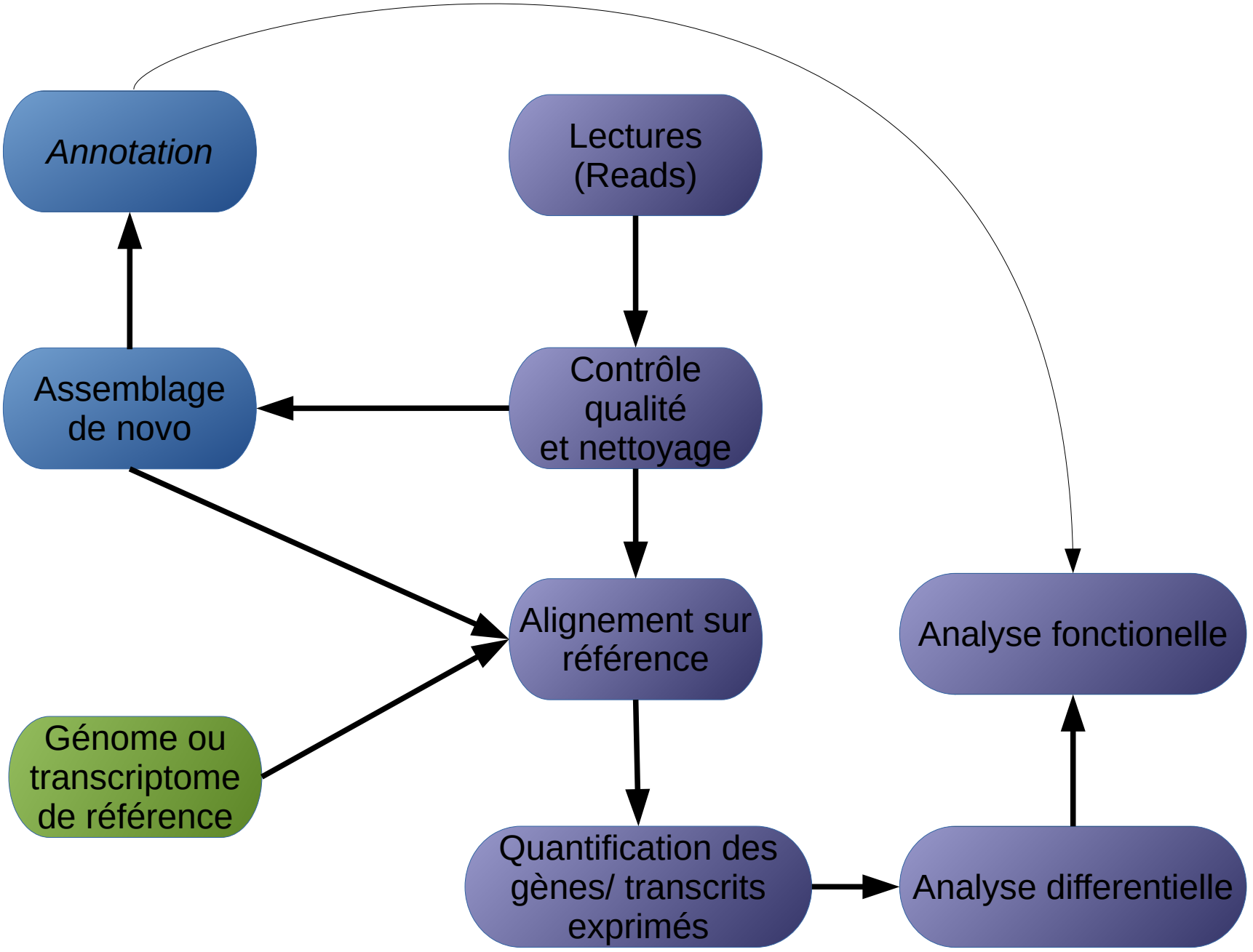
# TRAITEMENT BIOINFORMATIQUE DE DONNÉES RNA-Seq

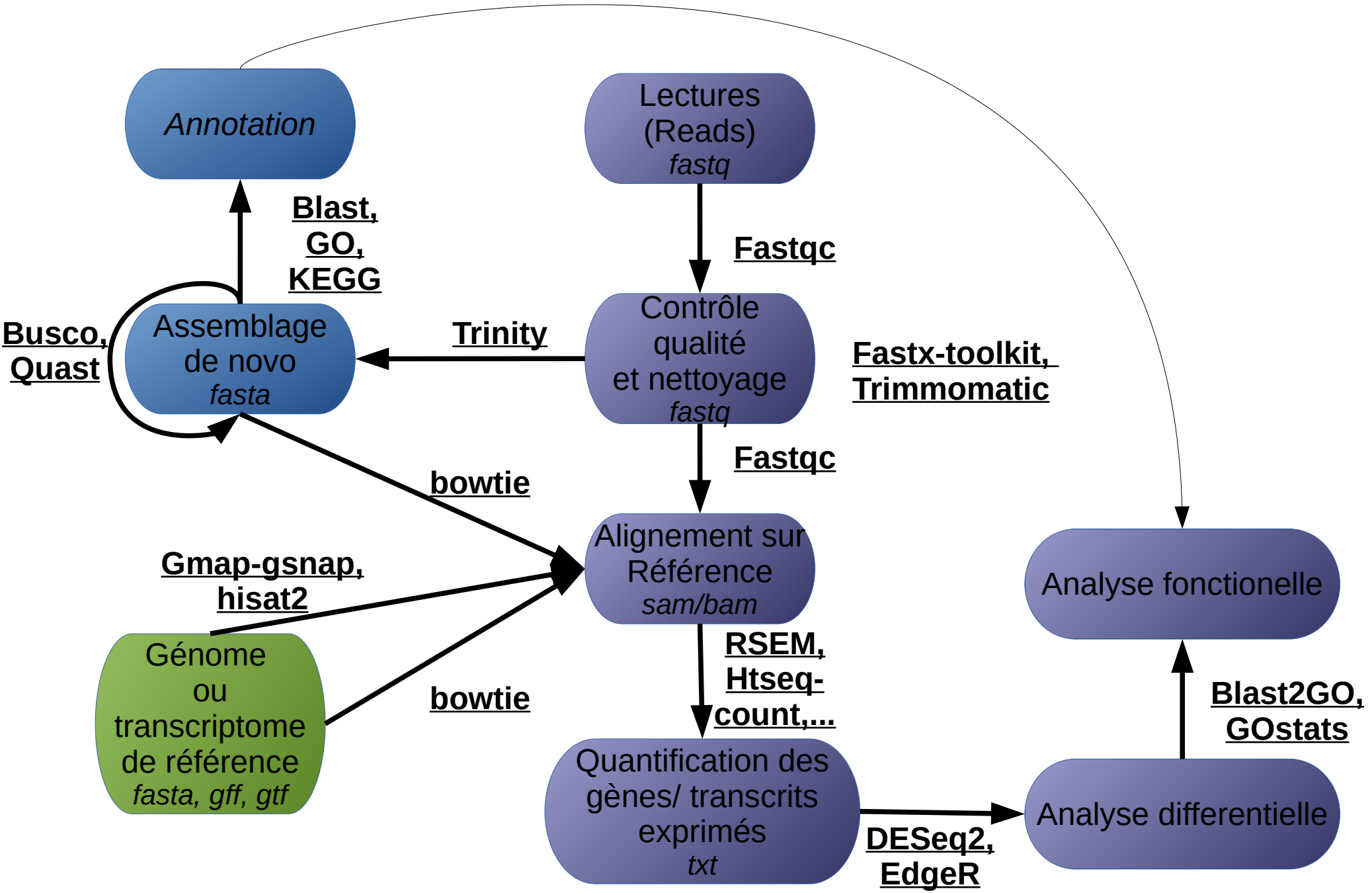
Anamaria NECSULEA & Adil EL FILALI

18/10/2023

# Workflow d'analyse RNA-Seq



# Workflow d'analyse RNA-Seq



# Outils RNA-seq

## Data preprocessing

- Experimental design — 19 tools
- Base calling — 33 tools
- Read quality control — 112 tools
- Error correction — 114 tools
- Adapter trimming — 108 tools
- k-mer counting — 36 tools
- Depth of coverage — 22 tools
- Read clustering — 24 tools
- Demultiplexing — 28 tools
- Barcode removal — 4 tools

## Data processing

- Spliced read alignment — 98 tools
- Read alignment — 251 tools
- Read realignment — 3 tools
- Indel realignment — 4 tools
- Alignment evaluation — 33 tools
- De novo transcriptome assembly — 55 tools
- Reference-based transcriptome assembly — 46 tools
- Assembly evaluation — 10 tools
- Base quality recalibration — 3 tools
- Duplicate read removal — 42 tools

## Quantification

- Read count — 63 tools
- Known transcript quantification — 135 tools
- Novel transcript quantification — 46 tools
- Alignment-free transcript quantification — 12 tools
- Batch effect correction — 13 tools
- Normalization — 174 tools

## Data visualization

- Alternative splicing visualization — 29 tools
- Gene expression visualization — 67 tools
- Transcriptome visualization — 21 tools
- Gene fusion visualization — 10 tools
- Circular RNA visualization — 1 tool
- Genome visualization — 162 tools
- Read alignment visualization — 46 tools
- Heatmap generation — 36 tools
- Venn diagram creation — 29 tools
- Variant visualization — 44 tools
- Gene visualization — 21 tools
- Sequence alignment visualization — 55 tools
- Tree visualization — 1[illegible] tools
- Network visualization — 210 tools

## Classification/Clustering

- Gene expression clustering — 174 tools
- Gene expression classification — 64 tools

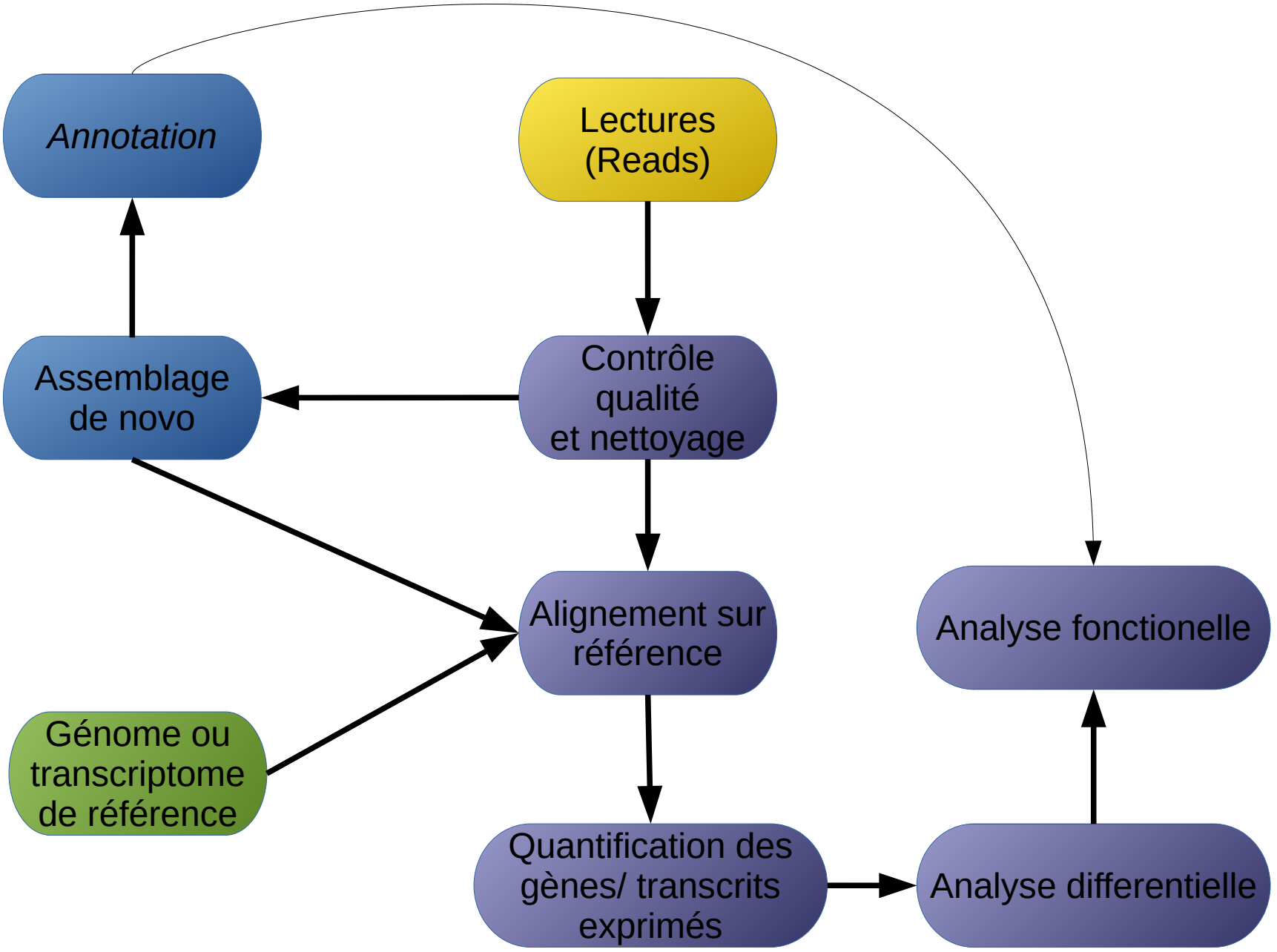
## Network analysis

- Gene co-expression prediction — 74 tools
- Differential co-expression analysis — 16 tools
- miRNA regulatory modules inference — 23 tools
- Gene regulatory network inference — 325 tools

## Enrichment analysis

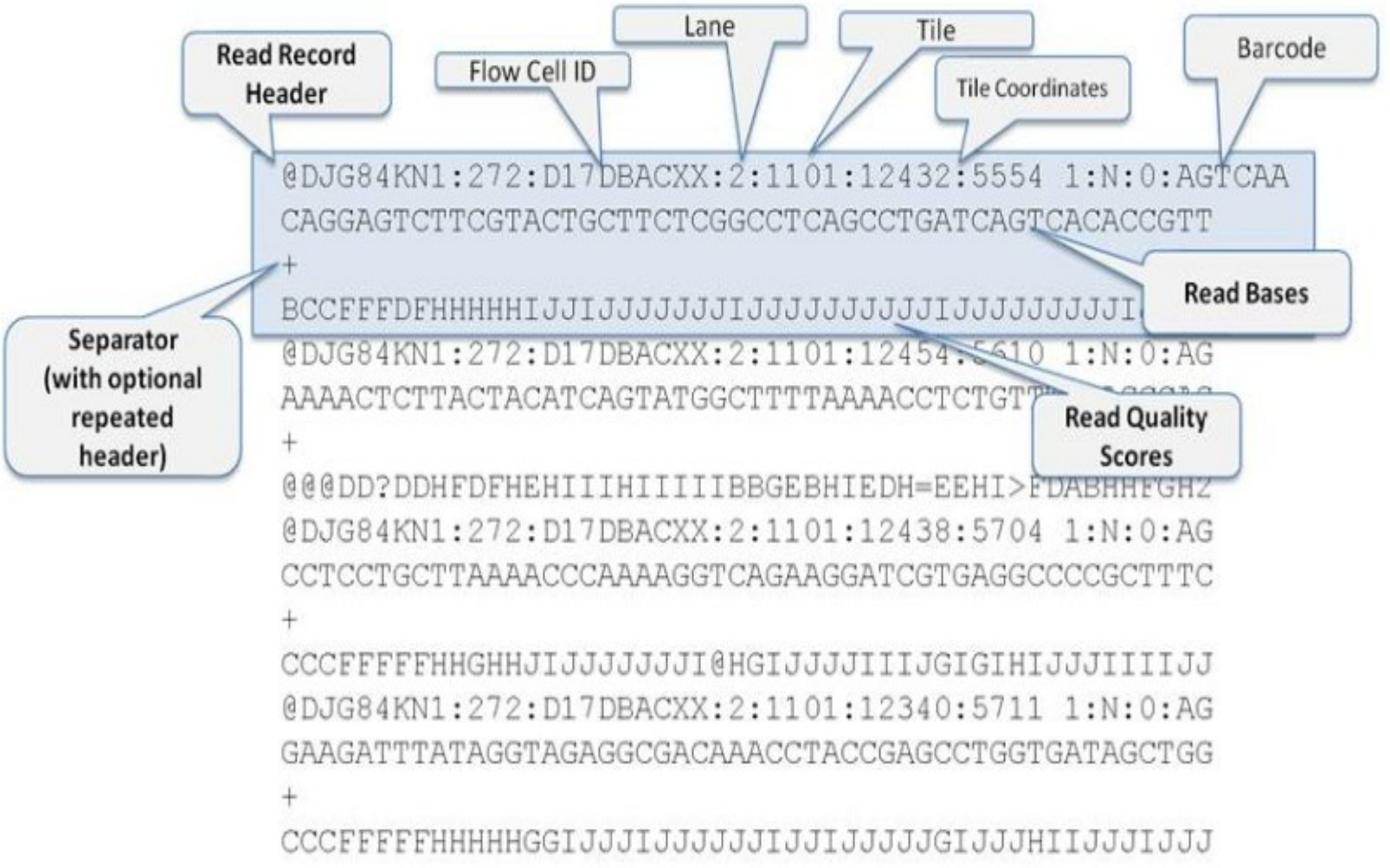
- RBP motif enrichment — 2 tools
- Over-representation analysis — 48 tools
- Gene set enrichment analysis — 205 tools
- Topology enrichment analysis — 70 tools

# Workflow d'analyse RNA-Seq



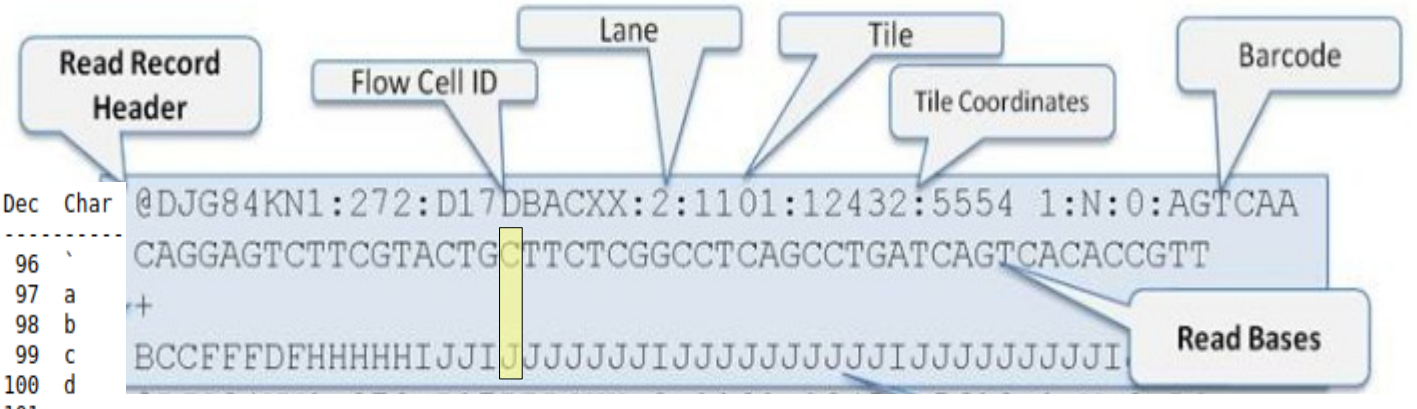
# Lectures (*fastq)*

2 FASTQ files in case of a paired-end experiment



# Lectures (*fastq)*

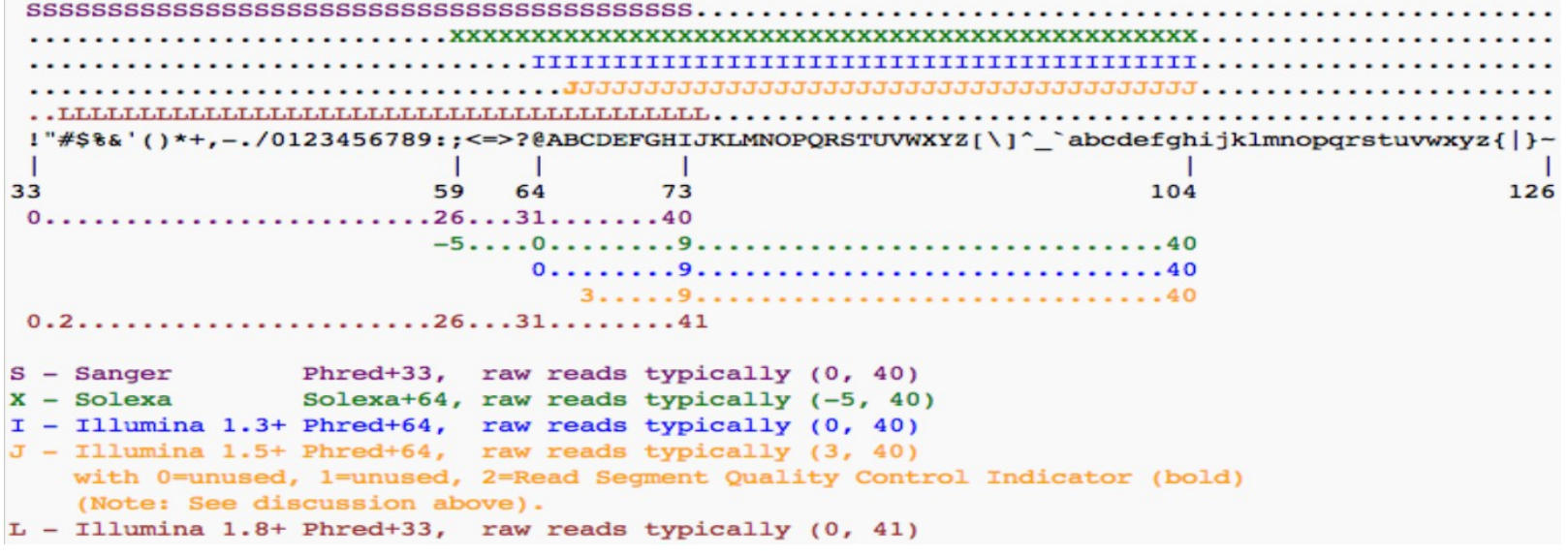
## 2 FASTQ files in case of a paired-end experiment




| Dec | Char | Dec | Char | Dec | Char |
|---|---|---|---|---|---|
| 32 | SPACE | 64 | @ | 96 | ` |
| 33 | ! | 65 | A | 97 | a |
| 34 | " | 66 | B | 98 | b |
| 35 | # | 67 | C | 99 | c |
| 36 | $ | 68 | D | 100 | d |
| 37 | % | 69 | E | 101 | e |
| 38 | & | 70 | F | 102 | f |
| 39 | ' | 71 | G | 103 | g |
| 40 | ( | 72 | H | 104 | h |
| 41 | ) | 73 | I | 105 | i |
| 42 | * | 74 | J | 106 | j |
| 43 | + | 75 | K | 107 | k |
| 44 | , | 76 | L | 108 | l |
| 45 | - | 77 | M | 109 | m |
| 46 | . | 78 | N | 110 | n |
| 47 | / | 79 | O | 111 | o |
| 48 | 0 | 80 | P | 112 | p |
| 49 | 1 | 81 | Q | 113 | q |
| 50 | 2 | 82 | R | 114 | r |
| 51 | 3 | 83 | S | 115 | s |
| 52 | 4 | 84 | T | 116 | t |
| 53 | 5 | 85 | U | 117 | u |
| 54 | 6 | 86 | V | 118 | v |
| 55 | 7 | 87 | W | 119 | w |
| 56 | 8 | 88 | X | 120 | x |
| 57 | 9 | 89 | Y | 121 | y |
| 58 | : | 90 | Z | 122 | z |
| 59 | ; | 91 | [ | 123 | { |
| 60 | < | 92 | \ | 124 | \| |
| 61 | = | 93 | ] | 125 | } |
| 62 | > | 94 | ^ | 126 | ~ |
| 63 | ? | 95 | _ | 127 | DEL |

En Phred+33 : score de qualité = nombre ASCII – 33

74 - 33 = 41



# Lectures (*fastq)*

Score de qualité = Phred Score

- Pour chaque base, mesure la probabilité que la base assignée soit fausse

- Le score d'une base donné (Q) est donné par l'équation suivante :

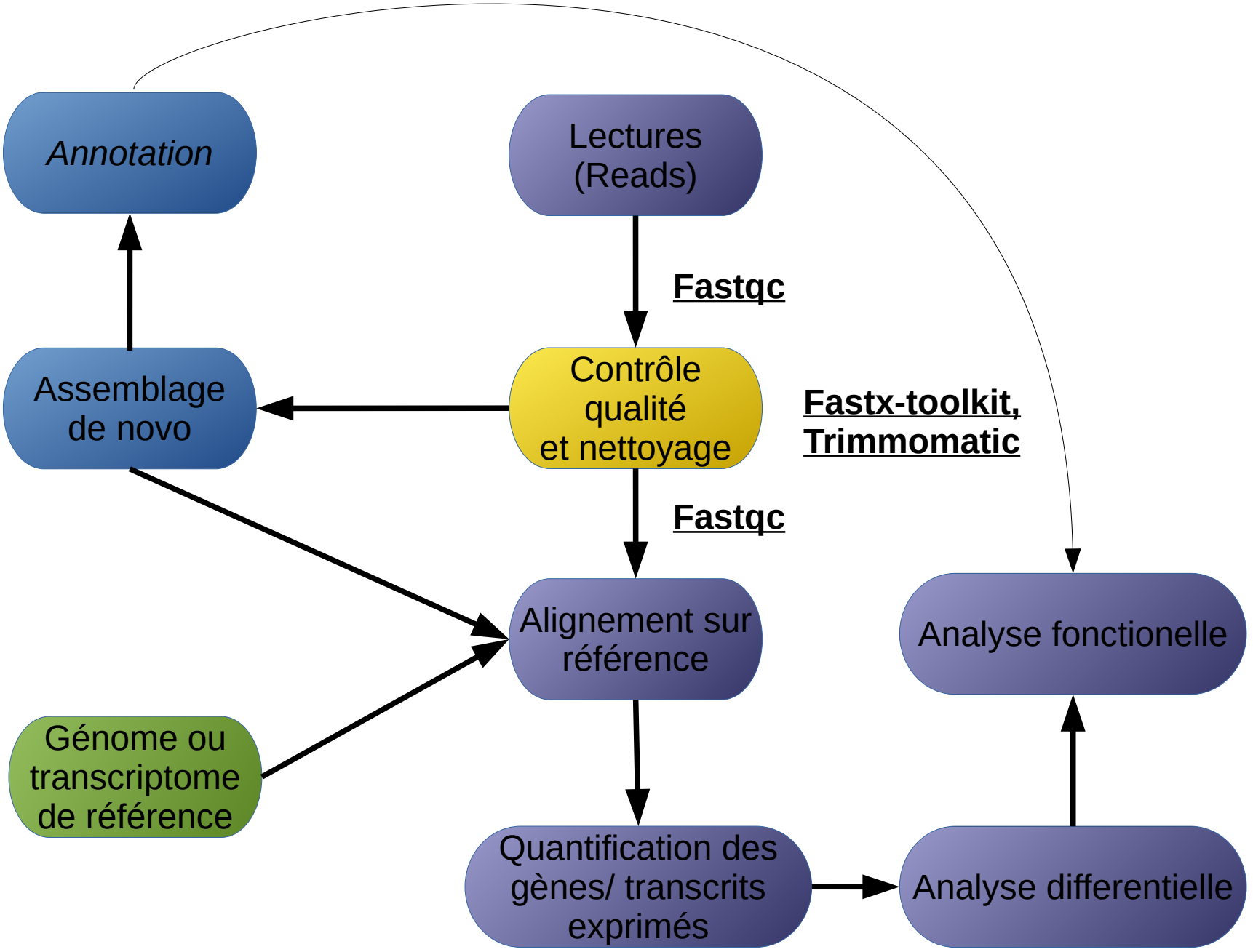
$$Q = -10\log10(p)$$

p: Probabilité estimée que la base donnée soit fausse

Donc un haut score indique une plus petite probabilité d'erreur.

| Base Quality | $P_{error}$(obs. base) |
|---|---|
| 3 | 50 % |
| 5 | 32 % |
| 10 | 10 % |
| 20 | 1 % |
| 30 | 0.1 % |
| 40 | 0.01 % |

# Workflow d'analyse RNA-Seq



# FastQC

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ce n'est pas un outil de correction mais un outil de visualisation de la qualité des données génomique.

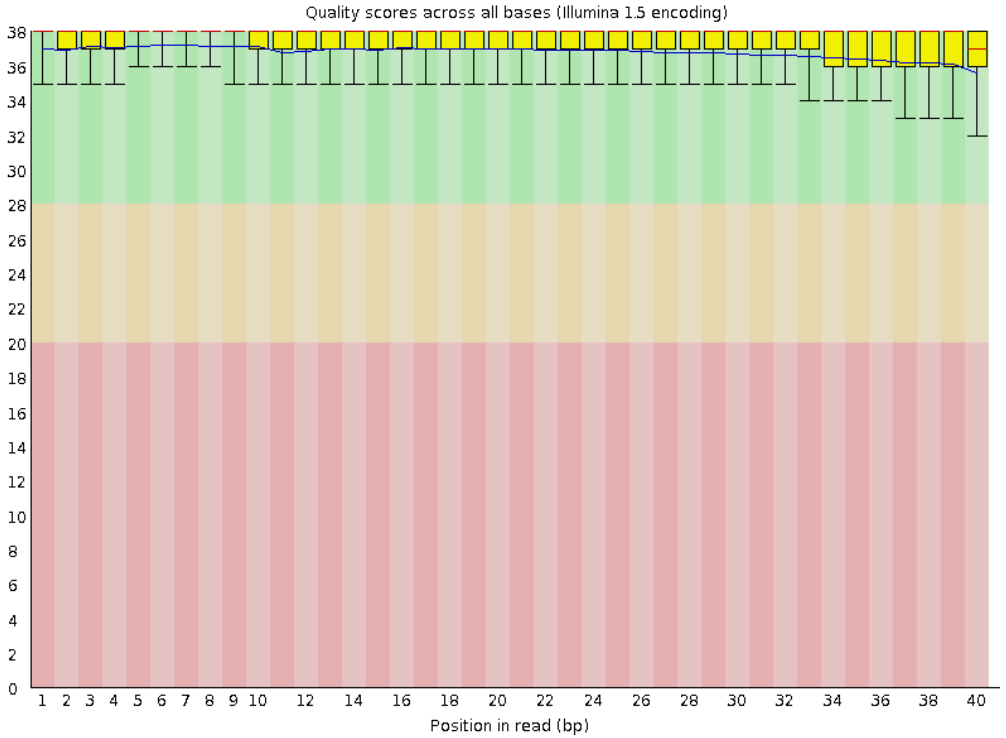
## Summary

- Basic Statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

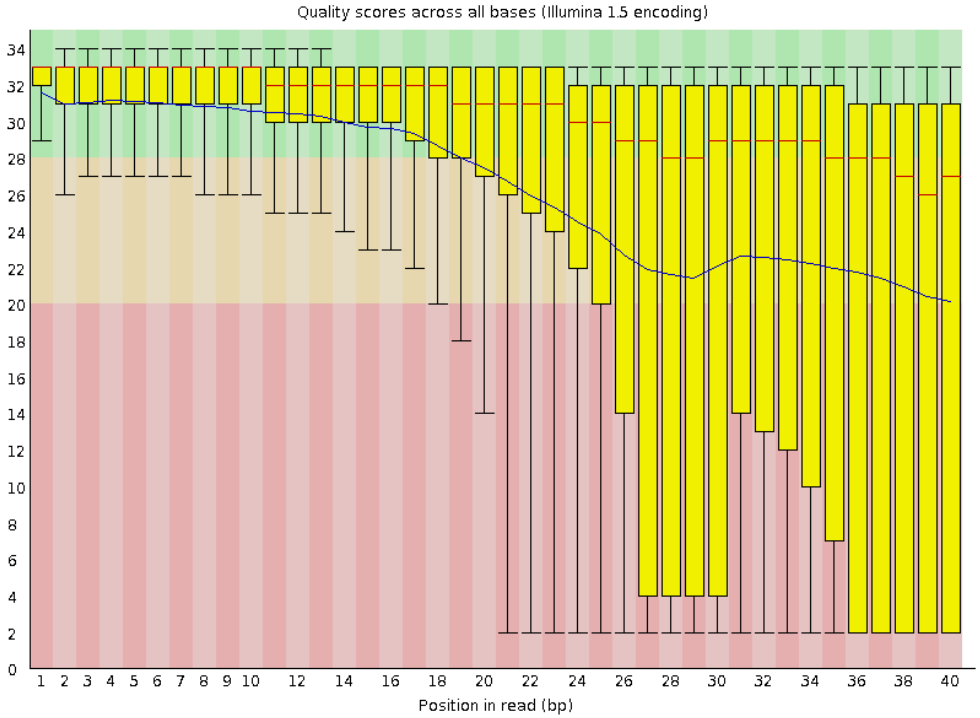
| Measure | Value |
|---|---|
| Filename | 454_SRR073599.fastq.gz |
| File type | Conventional base calls |
| Encoding | Sanger / Illumina 1.9 |
| Total Sequences | 125000 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 44-2042 |
| %GC | 43 |

# FastQC

**Per base sequence quality**

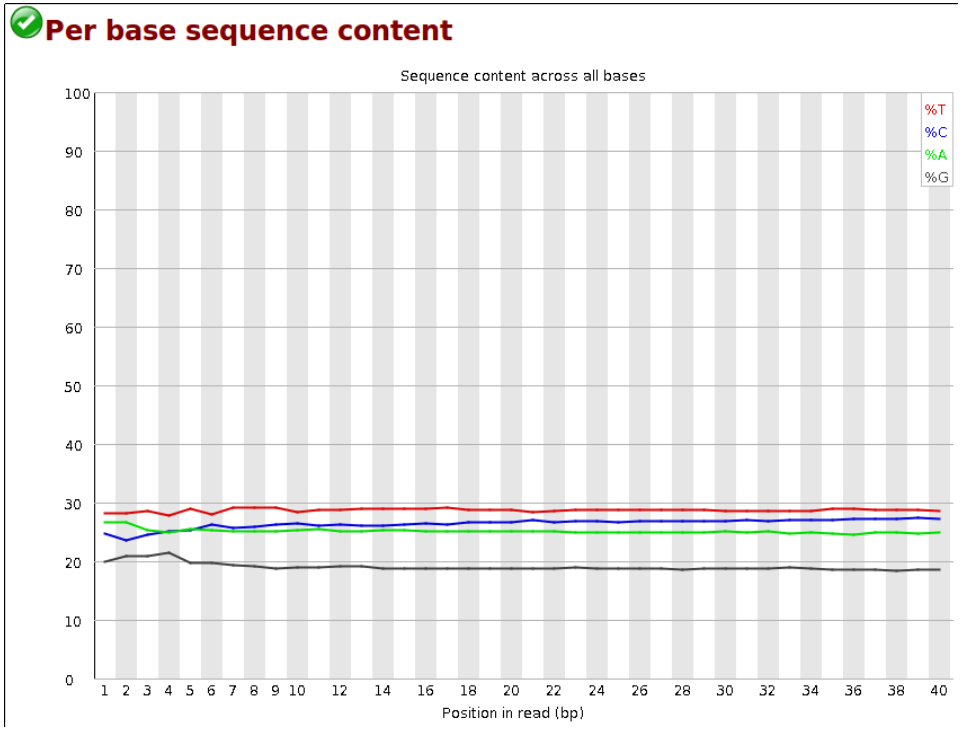


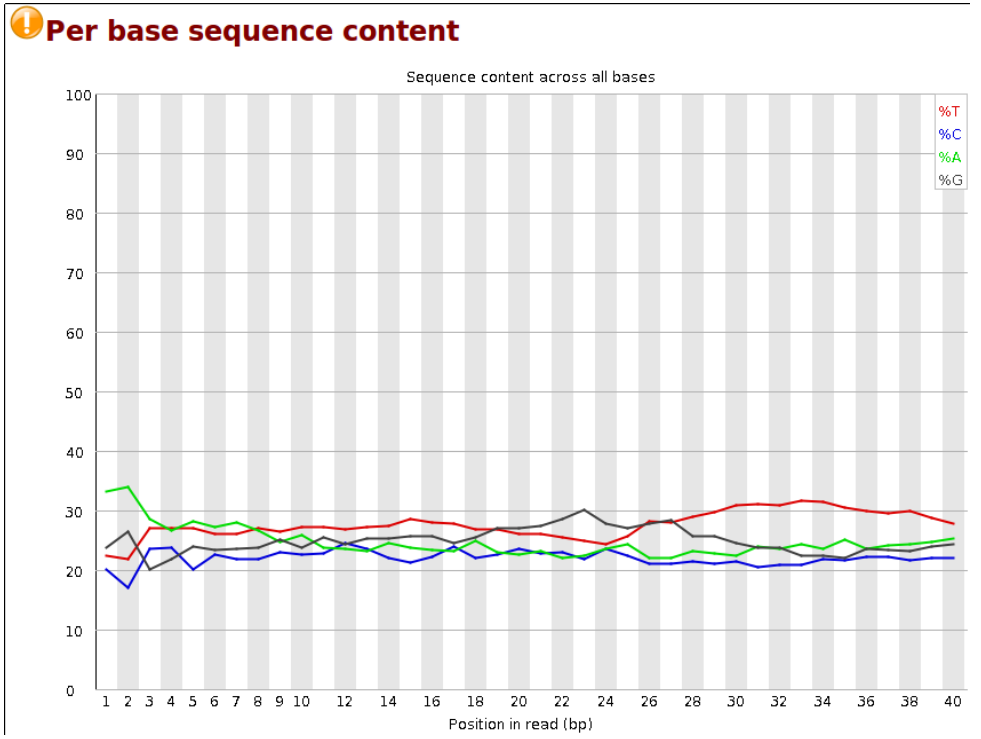

**Per base sequence quality**




- La ligne rouge centrale correspond a la médiane
- La boxplot jaune represente l'inter-quartile (25-75%)
- Les lignes inférieures et supérieures representent la limite de 10 % et 90 %
- La courbe bleue correspond a la moyenne de la qualité

# FastQC




- Indication de la proportion en base (A/T/C/G) par position
- Dans un jeu de données aléatoire, peu de différence entre bases ⟶ lignes quasi-paralleles
- Dans un jeu de données biaisé, différence dans la proportion des bases

⟶ Sur-representation d une séquence contaminant le jeu de donnée

# Biais : random hexamer priming

- Fort biais de composition des 13 premières nucléotides en 5’
  - spécificité de séquence de la polymérase

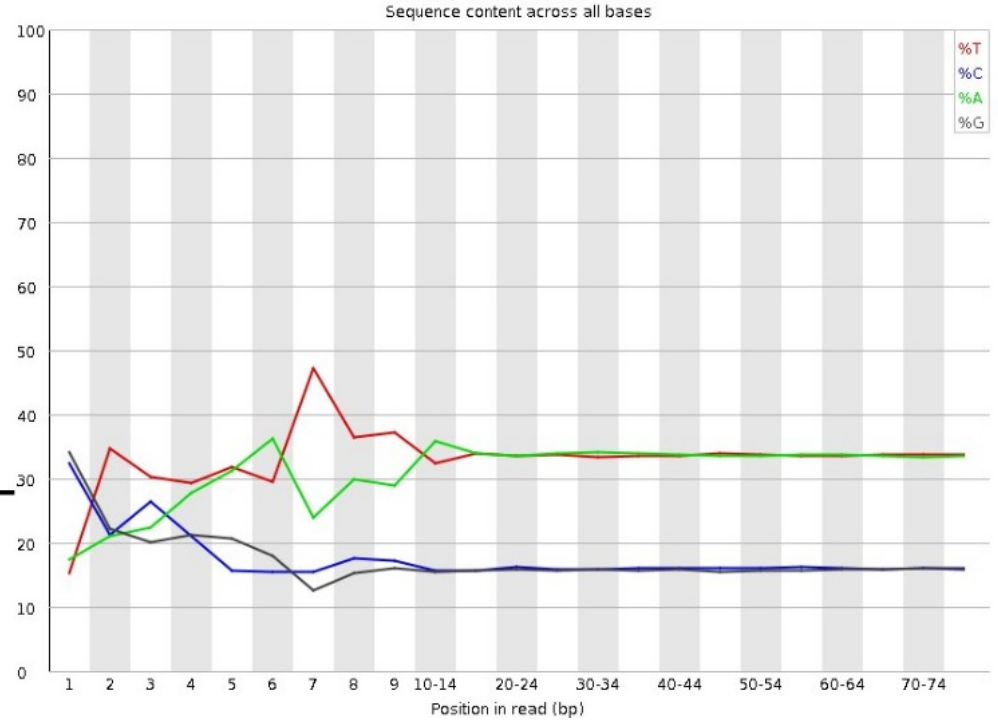
*Published online 14 April 2010* *Nucleic Acids Research, 2010, Vol. 38, No. 12 e131*
*doi:10.1093/nar/gkq224*

**Biases in Illumina transcriptome sequencing caused by random hexamer priming**

Kasper D. Hansen[1,*], Steven E. Brenner[2] and Sandrine Dudoit[1,3]

**ABSTRACT**

Generation of cDNA using random hexamer priming induces biases in the nucleotide composition at the beginning of transcriptome sequencing reads from the Illumina Genome Analyzer. The bias is independent of organism and laboratory and impacts the uniformity of the reads along the transcriptome. We provide a read count reweighting scheme, based on the nucleotide frequencies of the reads, that mitigates the impact of the bias.

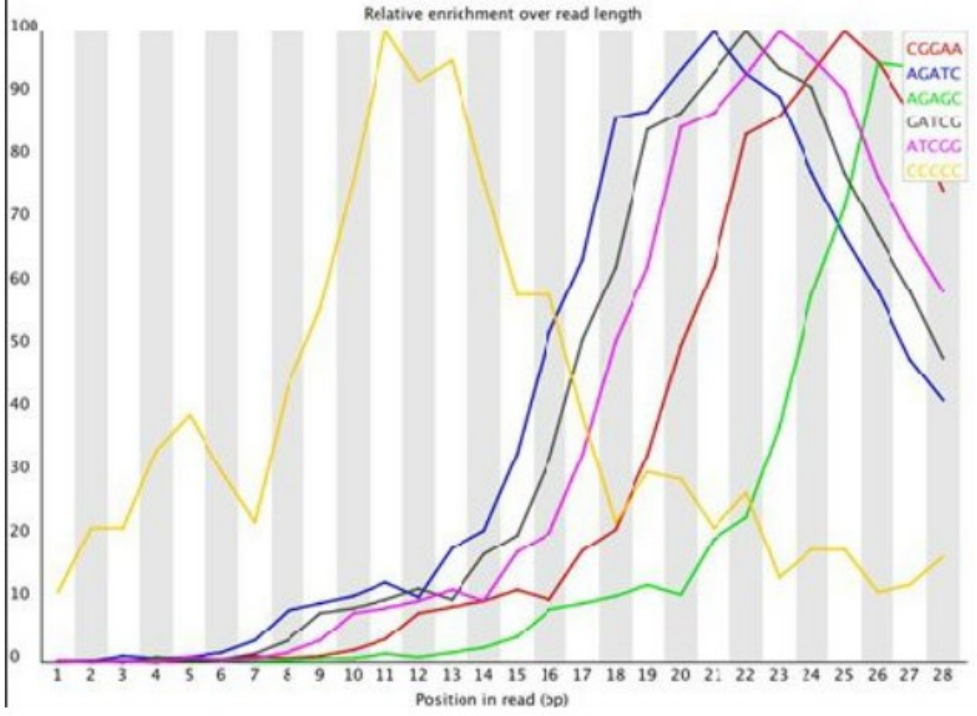


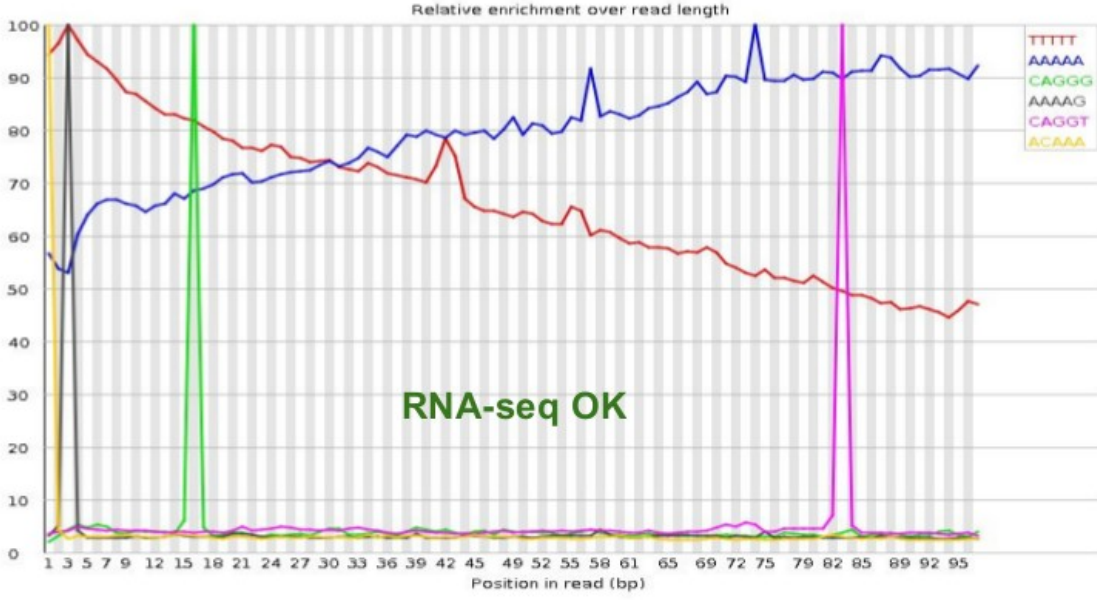
# FastQC

Surpresentation des k-mers

- Permet de repérer les séquences surreprésentées, donner une bonne impression de toute contamination

- vérifier la présence d'adaptateur ou de séquence poly-A (ou poly-T)




| Sequence | Count | Obs/Exp Overall | Obs/Exp Max | Max Obs/Exp Position |
|---|---|---|---|---|
| TTTTT | 47499960 | 4.84021 | 7.2762637 | 3 |
| AAAAA | 18101385 | 4.2297845 | 5.3006034 | 74 |
| CAGGG | 12486915 | 2.3769662 | 49.03375 | 16 |
| AAAAG | 10728075 | 2.3667703 | 56.233307 | 3 |

# Nettoyage

But :

Retirer les adaptateurs
Eliminer les reads de faibles qualités

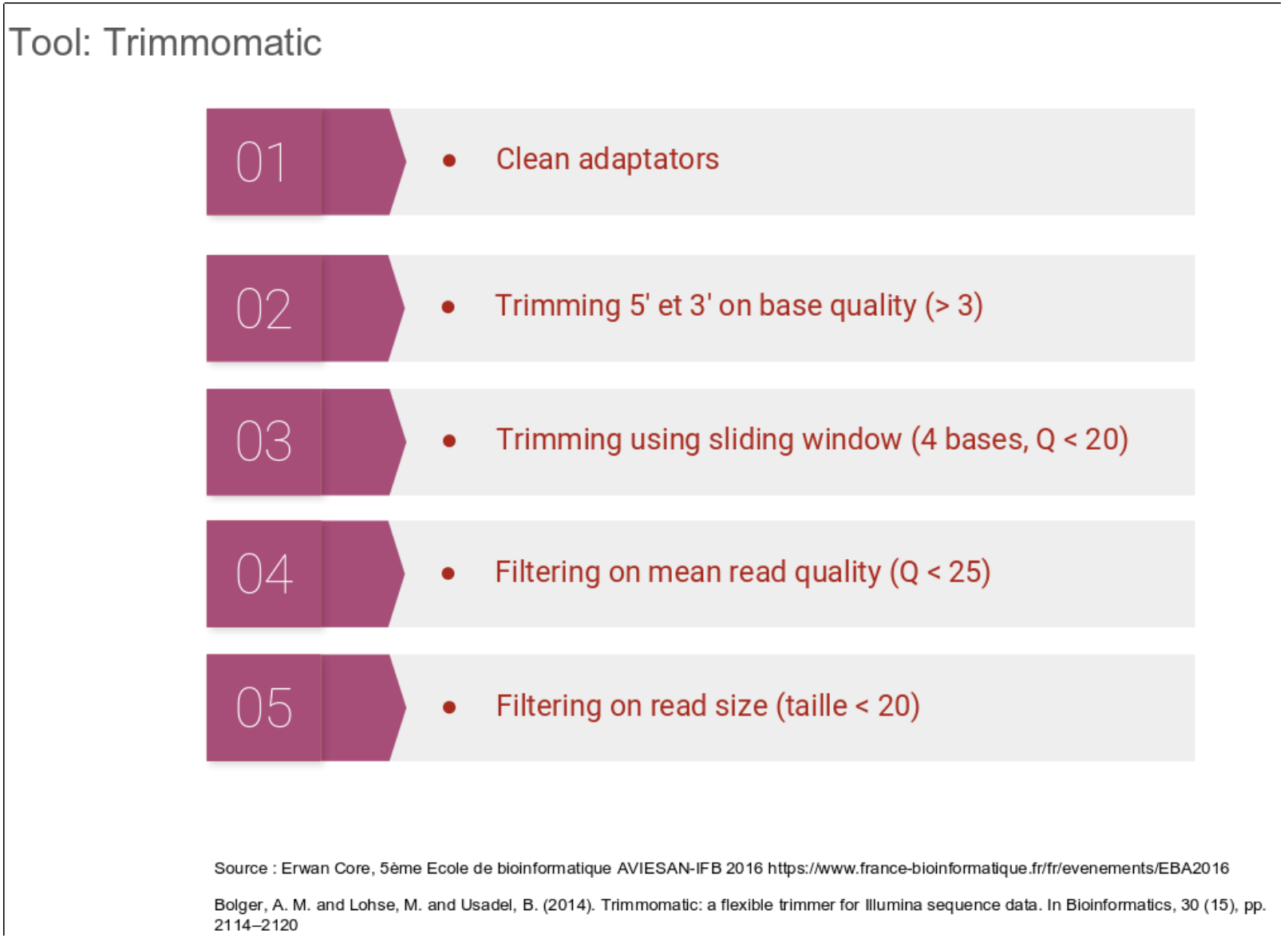
Software :

BBTools - https://jgi.doe.gov/data-and-tools/bbtools/
Bignorm - https://git.informatik.uni-kiel.de/axw/Bignorm
Centrifuge - https://github.com/DaehwanKimLab/centrifuge
cutadapt - https://github.com/marcelm/cutadapt
Falco - https://github.com/smithlabcode/falco
fastp - https://github.com/OpenGene/fastp
FastQC - https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
khmer - https://github.com/dib-lab/khmer
Kraken2 - https://github.com/DerrickWood/kraken2
MultiQC - https://multiqc.info
rCorrector - https://github.com/mourisl/Rcorrector
SortMeRNA - https://github.com/biocore/sortmerna
TrimGalore - https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Trimmomatic - https://github.com/usadellab/Trimmomatic

# Nettoyage

**Trimmomatic** est un programme qui effectue une variété de trimming et de filtrage sur les reads (en single-end ou en paired-end), Il requière Java.

http://www.usadellab.org/cms/?page=trimmomatic




Source : Erwan Core, 5ème Ecole de bioinformatique AVIESAN-IFB 2016 https://www.france-bioinformatique.fr/fr/evenements/EBA2016

Bolger, A. M. and Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. In Bioinformatics, 30 (15), pp. 2114–2120

# Nettoyage

Cutadapt

- développé en Python
- semi-global alignment algorithm
- détecte et supprime les séquences adaptatrices, les amorces, les queues poly-A et d'autres types de séquences

$ cutadapt -a ADAPTER-SEQUENCE input.fastq > output.fastq

$ cutadapt -b TGAGACACGCA -b AGGCACACAGGG input.fastq > output.fastq

# Définition d une bonne qualité pour les reads

- Le nombre de reads produites correspondant au nombre attendu

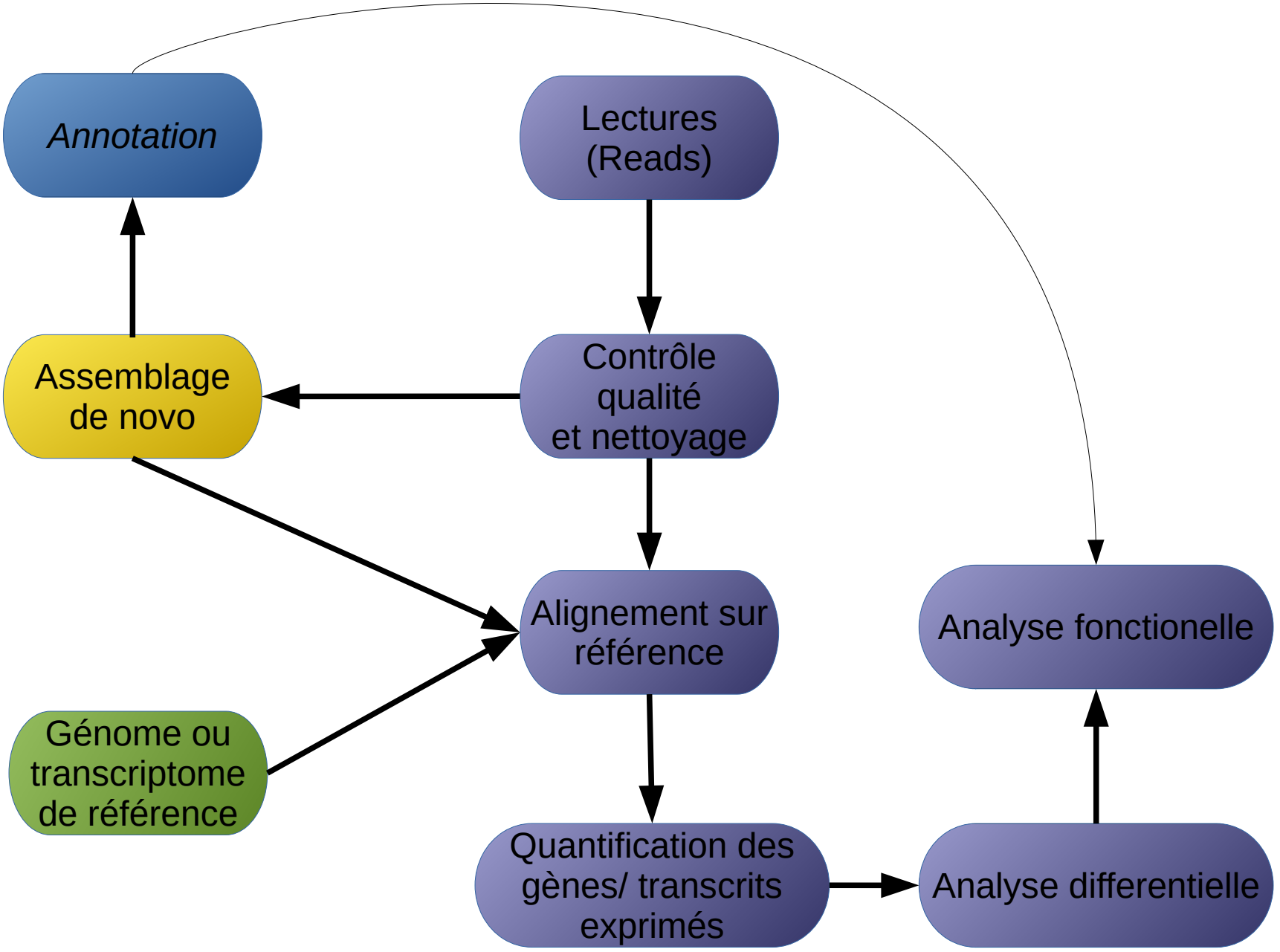
- Pas de contaminations (K-mer de polyA, polyT, adaptateurs …)

- Longueur des reads corrects (100-150 pb)

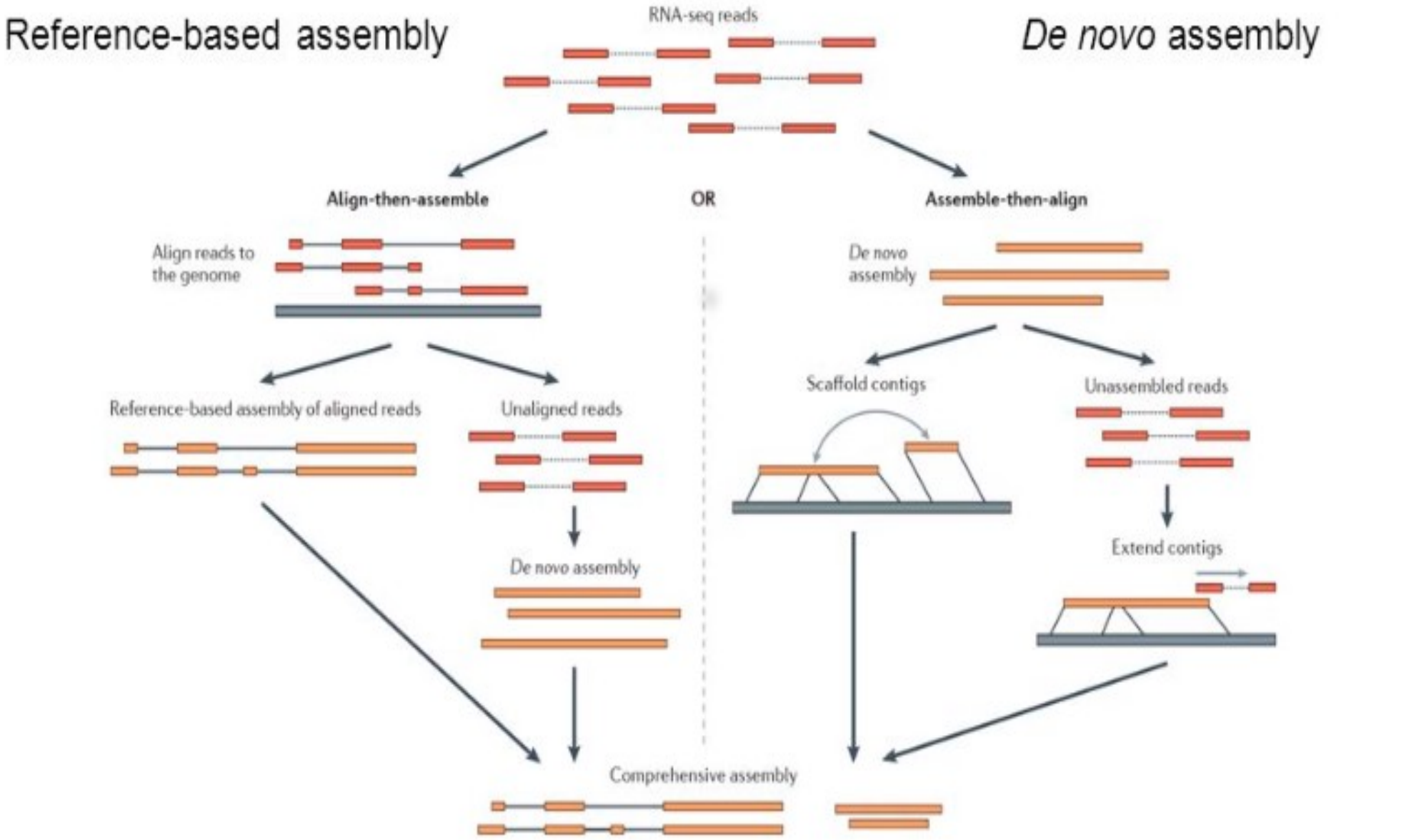
- Bonne qualité

- Bon alignement (re-séquencage avec génome de référence « propre ») : peu de reads non-alignés

# Workflow d'analyse RNA-Seq

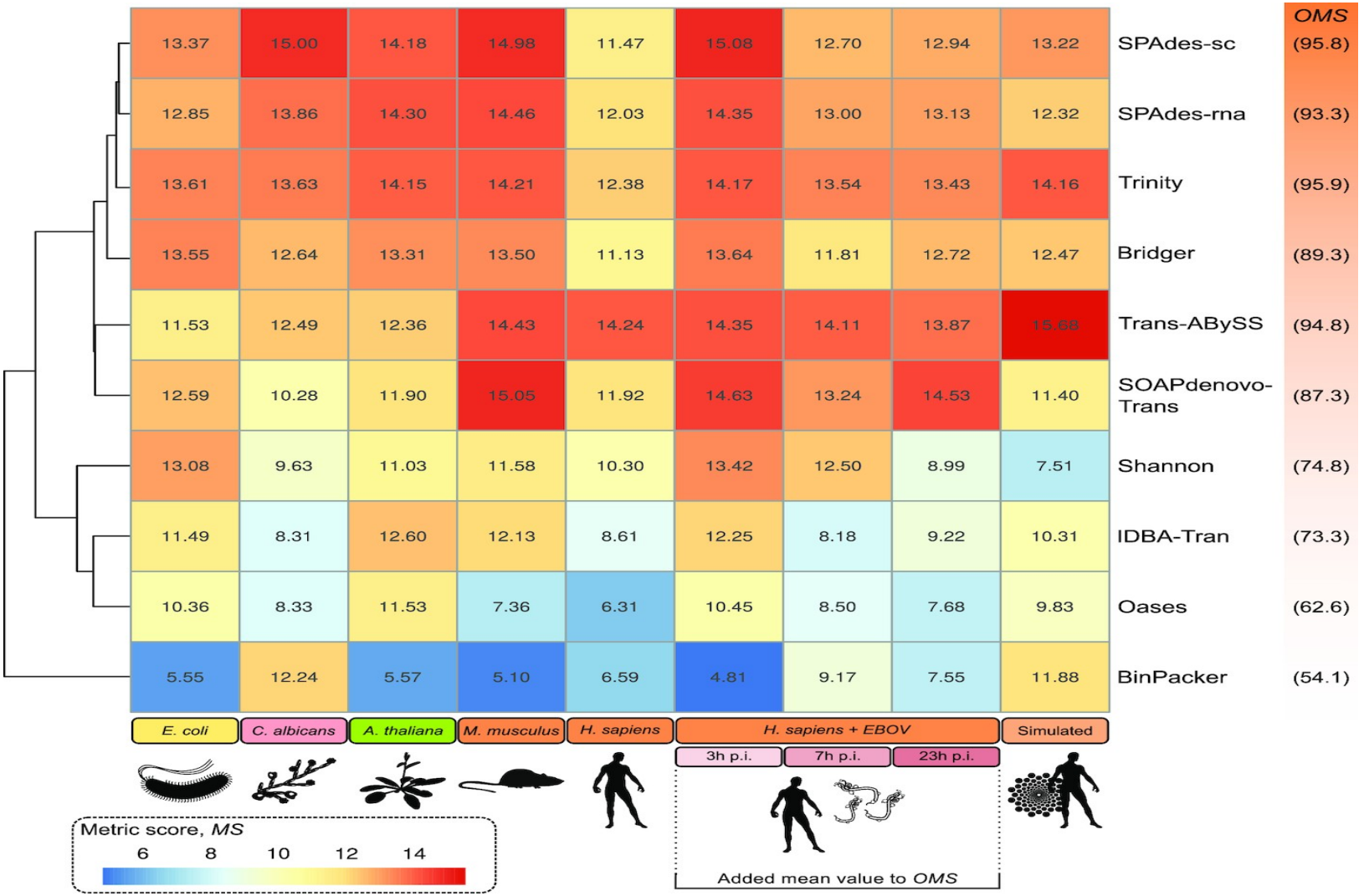


# Assemblage




Martin J.A. and Wang Z., Nat. Rev. Genet. (2011) 12:671–682

# Assemblage de-novo




De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. Hölzer M, Marz M. Gigascience. 2019 May 1;8(5)

# Assemblage de-novo
# Trinity

- Inchworm :
    - coupure des reads en k-mer (Jellyfish)
    - reconstruction des isoformes a partir des k-mer (gourmand en mémoire vive)
    - resultat : ensemble des contigs


- Chrysalis :
    - clusterise les contigs se chevauchant sur une longueur de k-1
    - construction d'un graphe de Bruijn pour chaque cluster


- Butterfly :
    - traitement des graphes individuellements
    - reconciliation des graphes avec les reads (pairés ou non)
    - resultat : gènes et isoformes

# Assemblage de-novo
# Trinity

Désavantages :

- basée sur une heuristique

- surestimation des transcrits (DRAP)

- gourmand en mémoire vive

Avantages :

- intègre toute les étapes bioinformatiques, du traitement des données brutes RNA-seq (trimmomatic, quantification, normalisation, analyse différentielle, annotation)

- choix du k-mer

- nombreux paramètres (sur toutes les étapes de l'assemblage)

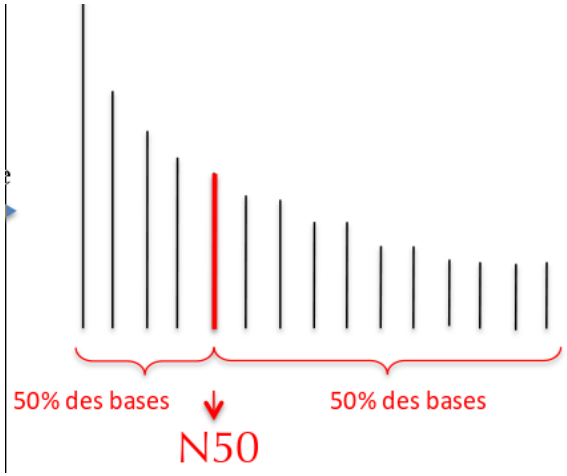
# Evaluation de l’assemblage de novo

| Tool | metrics |
|---|---|
| HISAT2 | Overall mapping rate |
| rnaQUAST | Transcripts ≥1,000 nt, Misassemblies, Mismatches per transcript, Average alignment length, 95%-assembled isoforms, Duplication ratio |
| Trinity/Salmon | Ex90N50 |
| Blastx | Full-length transcripts |
| TransRate | Reference coverage, Mean ORF percentage, Optimal score, Percentage bases uncovered, Number of ambiguous bases |
| DETONATE | Nucleotide F1, Contig F1, KC score, RSEM-EVAL |
| BUSCO | Complete BUSCOs, Missing BUSCOs |

# Evaluation de l'assemblage de novo : Quast

http://bioinf.spbau.ru/quast

- Nombre de contigs
- Nombre des longs contigs (i.e. > 1000 bp)
- Longueur du plus long contigs
- Longueur total de l'assemblage
- N50 : est la taille du scaffold (ou contig) tel que 50% des bases de l'assemblage sont comprises dans des scaffolds de taille supérieures à cette taille
- L50 : Le nombre minimum X tel que X contigs les plus longs couvrent au moins 50% de l'assemblage




All statistics are based on contigs of size >= 500 bp, unless otherwise noted

Short report

| Basic statistics | L_uniseq_100 | R_uniseq_100 |
|---|---|---|
| # contigs | 60 973 | 48 220 |
| # contigs (>= 0 bp) | 135 904 | 115 018 |
| # contigs (>= 1000 bp) | 10 517 | 10 586 |
| Largest contig | 5384 | 10 707 |
| Total length | 47 553 103 | 40 485 523 |
| Total length (>= 0 bp) | 72 636 910 | 63 616 017 |
| Total length (>= 1000 bp) | 14 543 268 | 15 807 198 |
| N50 | 749 | 827 |
| N75 | 601 | 617 |
| L50 | 21 334 | 15 476 |
| L75 | 39 258 | 29 950 |
| **Misassemblies** | | |
| **Unaligned** | | |
| **Genome statistics** | | |
| GC (%) | 41.04 | 41.44 |
| # N's | 8528 | 3280 |
| # N's per 100 kbp | 17.93 | 8.1 |
| **Aligned statistics** | | |

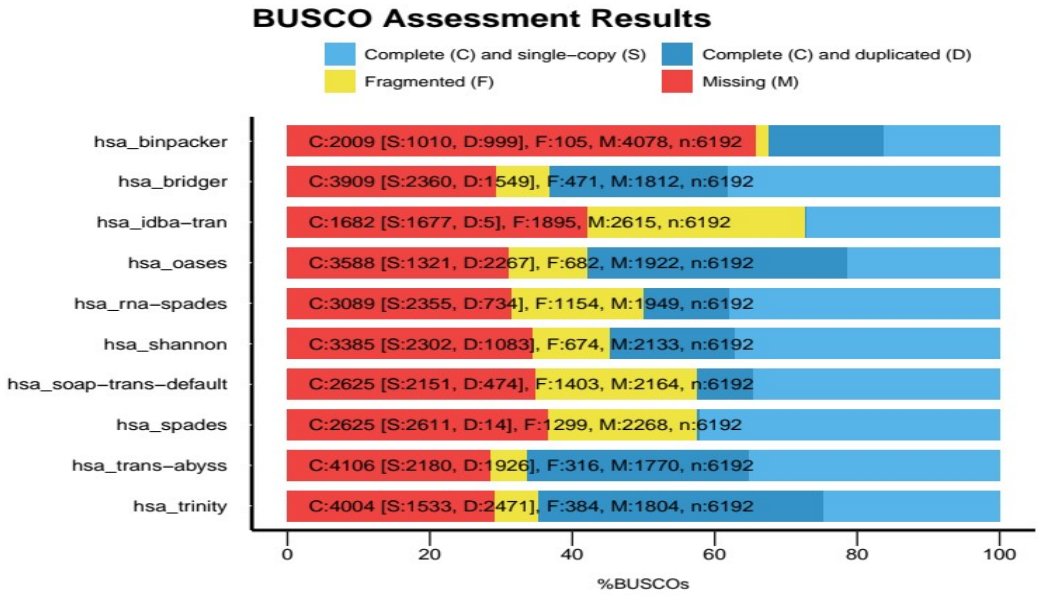
# Evaluation de l'assemblage de novo : BUSCO

BUSCO : Benchmarking Universal Single-Copy Orthologs

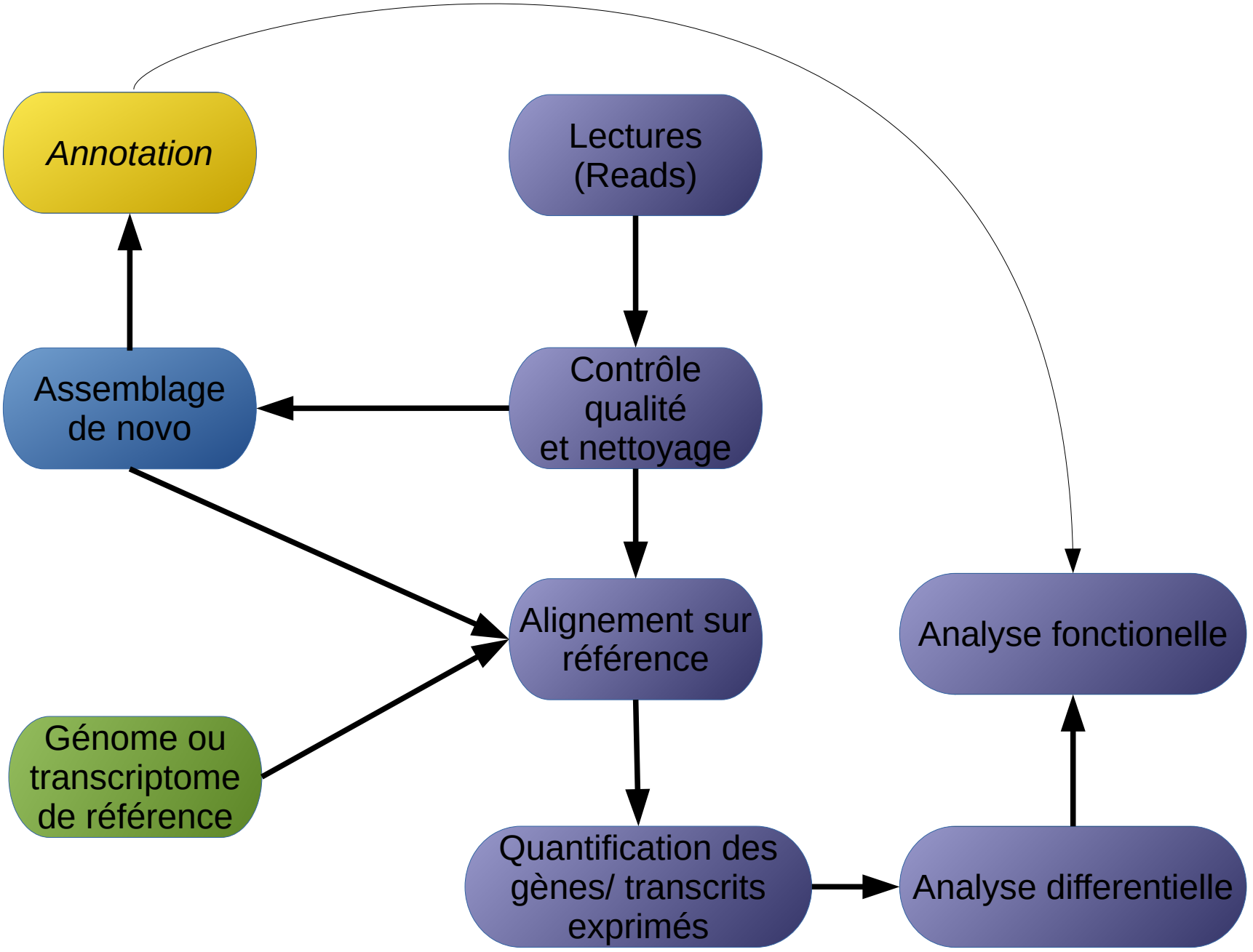
- évaluation de l'assemblage des génomes/transcriptomes en comparant les contigs à des séquences conservées simples copies et universels

- Le processus général de BUSCO est le suivant:

  - tTBLASTn contre des séquences consensus BUSCO.
  - Prévision de la structure des gènes en utilisant Augustus avec les profils de blocs BUSCO.
  - Assignation : "complete", "duplicated", "fragmented", ou "missing" (si il n y a pas de match).



# Workflow d'analyse RNA-Seq



# Annotation fonctionnelle

- Annotation fonctionnelle : attribuer une fonction potentielle à une séquence

- Comment : comparaison à des séquences, dont la fonction biologique est connue ou supposée, disponibles dans les bases de données (ex : nr, nt, swissprot...)

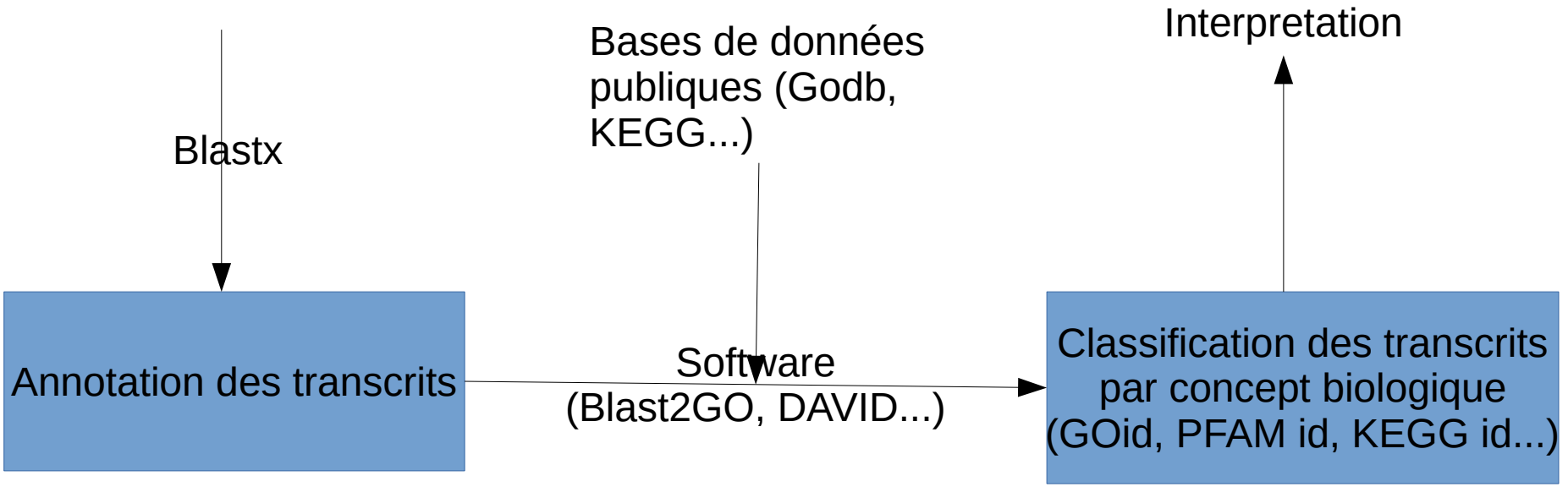
- Programmes :
  - blastn : comparaison à une banque nucléotidique (nt)
  - blastx : comparaison à une banque protéique (nr), traduction de la séquence dans les 6 cadre de lectures et comparaison à la banque

# Classification fonctionnelle

Classification fonctionnelle : attribuer à chaque séquence unique une classe de fonction

L'intérêt est de pouvoir cibler certains gènes dans une classe de fonction particulière. Elle vient en complément de l'annotation.

Comment : utilisation de base de données qui relient à un gène donné d'une espèce donnée une (ou des) classes(s) de fonction particulier(s).



# Classification fonctionnelle








diamond

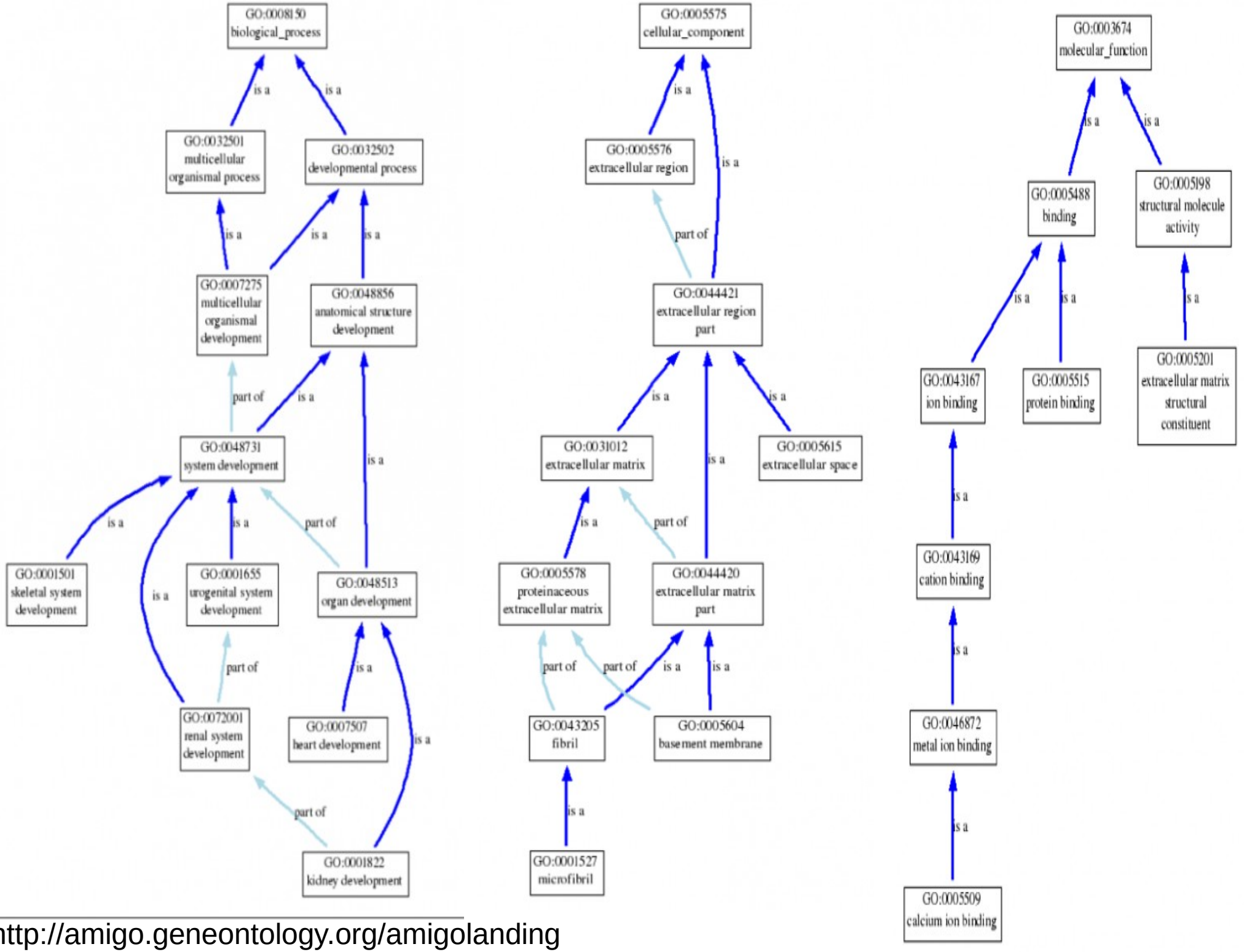
PSICQUIC

PANNZER2


Le Gene Ontology Consortium est un projet visant à créer une base de données décrivant les organismes de manière commune, permettant aux chercheurs de comparer plus facilement les données entre les espèces

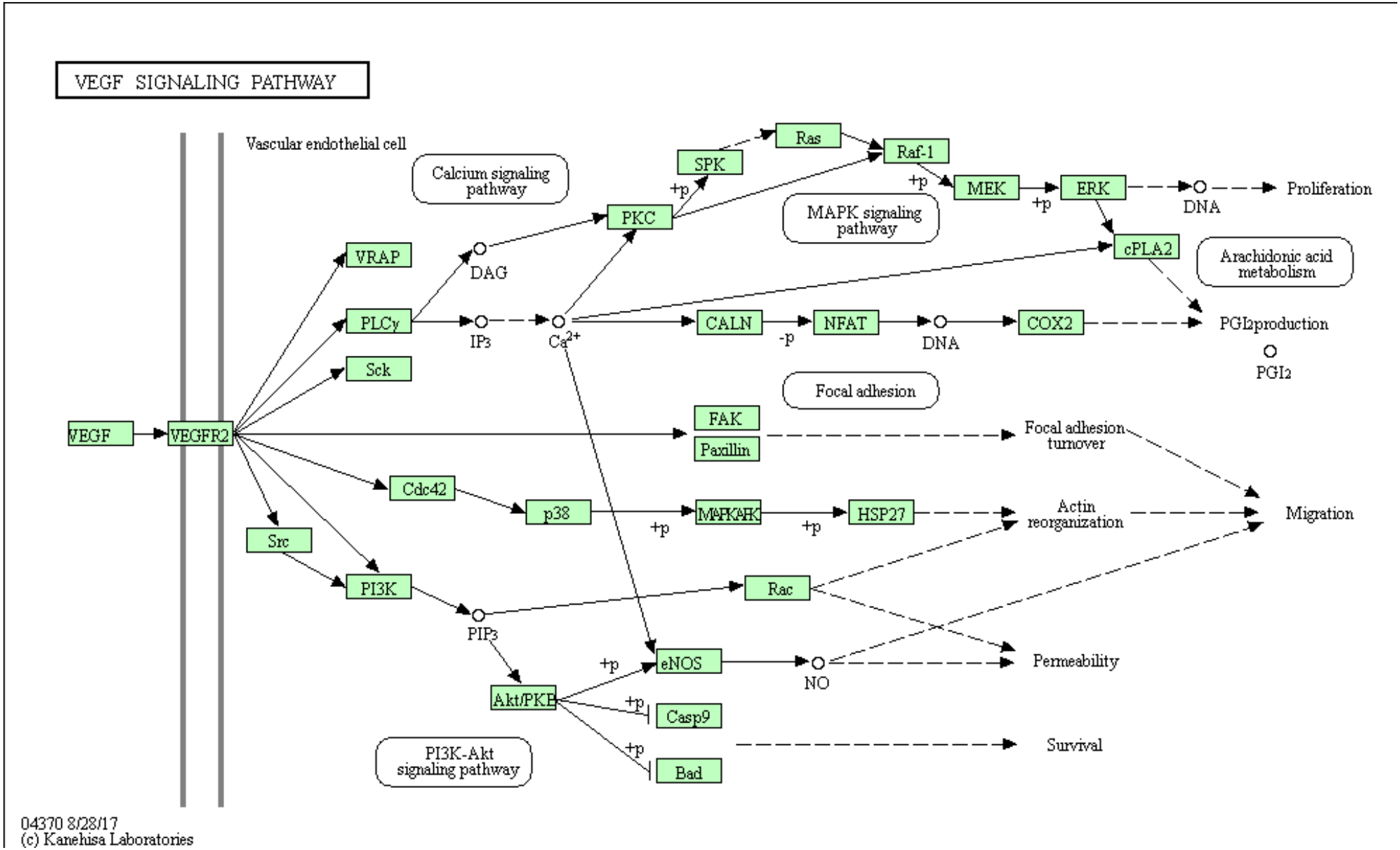
Les ontologies génétiques classent les informations selon trois niveaux différents:

fonction moléculaire (960900 ids)
processus biologique (934523 ids)
composant cellulaire (1019136 ids)

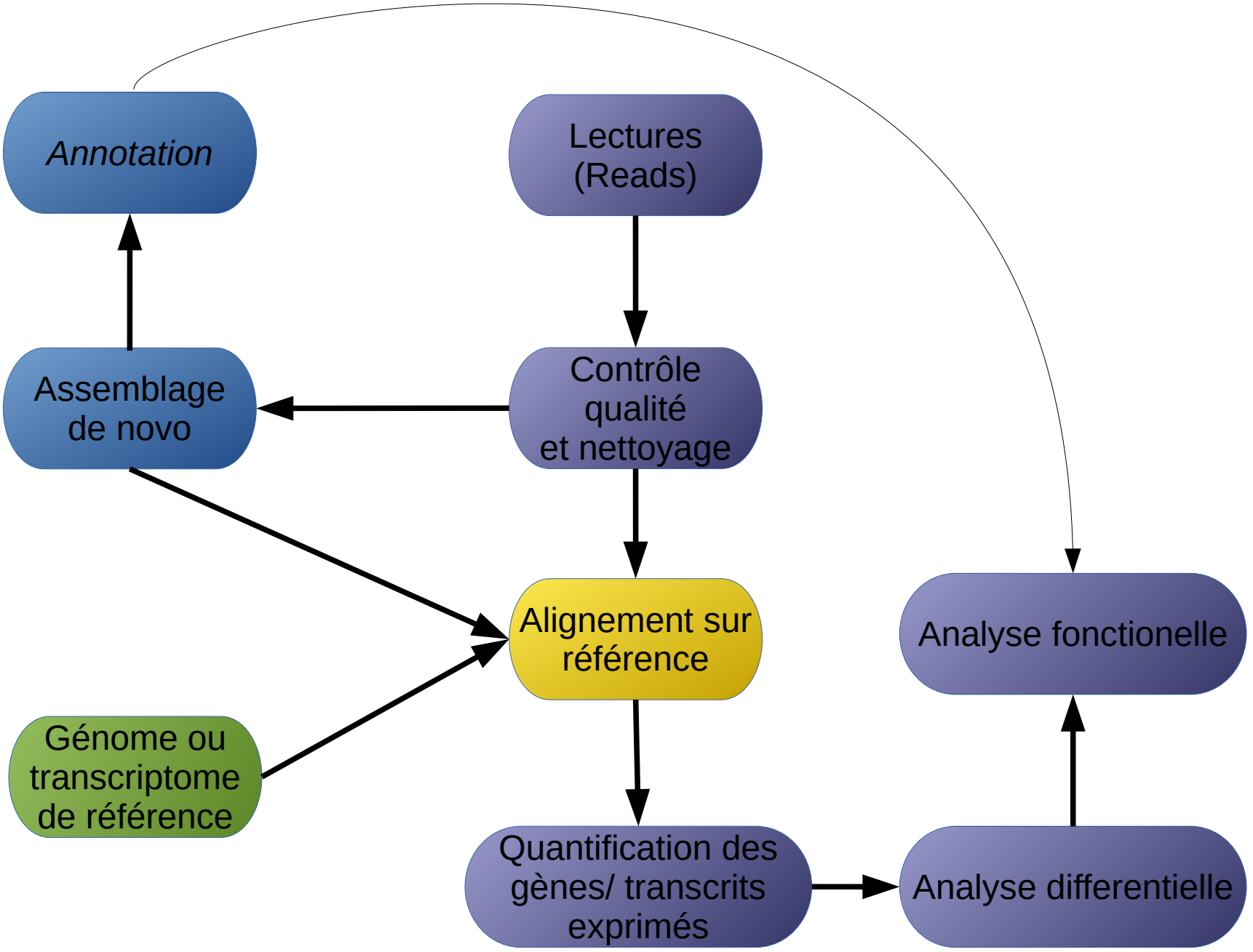



http://amigo.geneontology.org/amigolanding

# KEGG


VEGF SIGNALING PATHWAY
Vascular endothelial cell
Calcium signaling pathway
MAPK signaling pathway
Arachidonic acid metabolism
Focal adhesion
PI3K-Akt signaling pathway
VEGF
VEGFR2
VRAP
PLCγ
Sck
Src
PI3K
Cdc42
p38
FAK
Paxillin
SPK
PKC
DAG
IP3
Ca2+
Ras
Raf-1
MEK
ERK
DNA
Proliferation
cPLA2
CALN
NFAT
COX2
PGI2production
PGI2
MAPKAPK
HSP27
Actin reorganization
Focal adhesion turnover
Migration
Rac
PIP3
eNOS
NO
Permeability
Akt/PKB
Casp9
Bad
Survival
+p
-p
04370 8/28/17
(c) Kanehisa Laboratories

# Workflow d'analyse RNA-Seq



# Définition

Le mapping est la prédiction du locus dont est originaire la lecture.

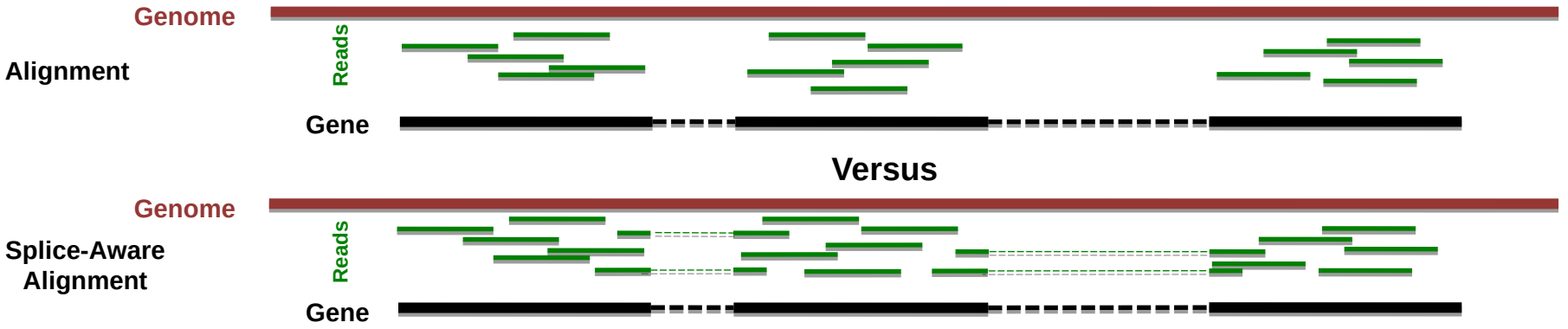
○ Locus : le résultat est un ensemble de positions génomiques
(ex.: chr1:100..150)

○ Mapping ARN ≠ Mapping ADN

Aligner les lectures issues du séquençage de dscDNA (transcrits) sur le génome, en tenant compte de l'épissage alternatif

Être capable d'exploiter les listes des jonctions exons-exons connues, mais également d'en détecter de nouvelles

Tout cela dans un temps raisonnable...



# Choix de l'aligneur

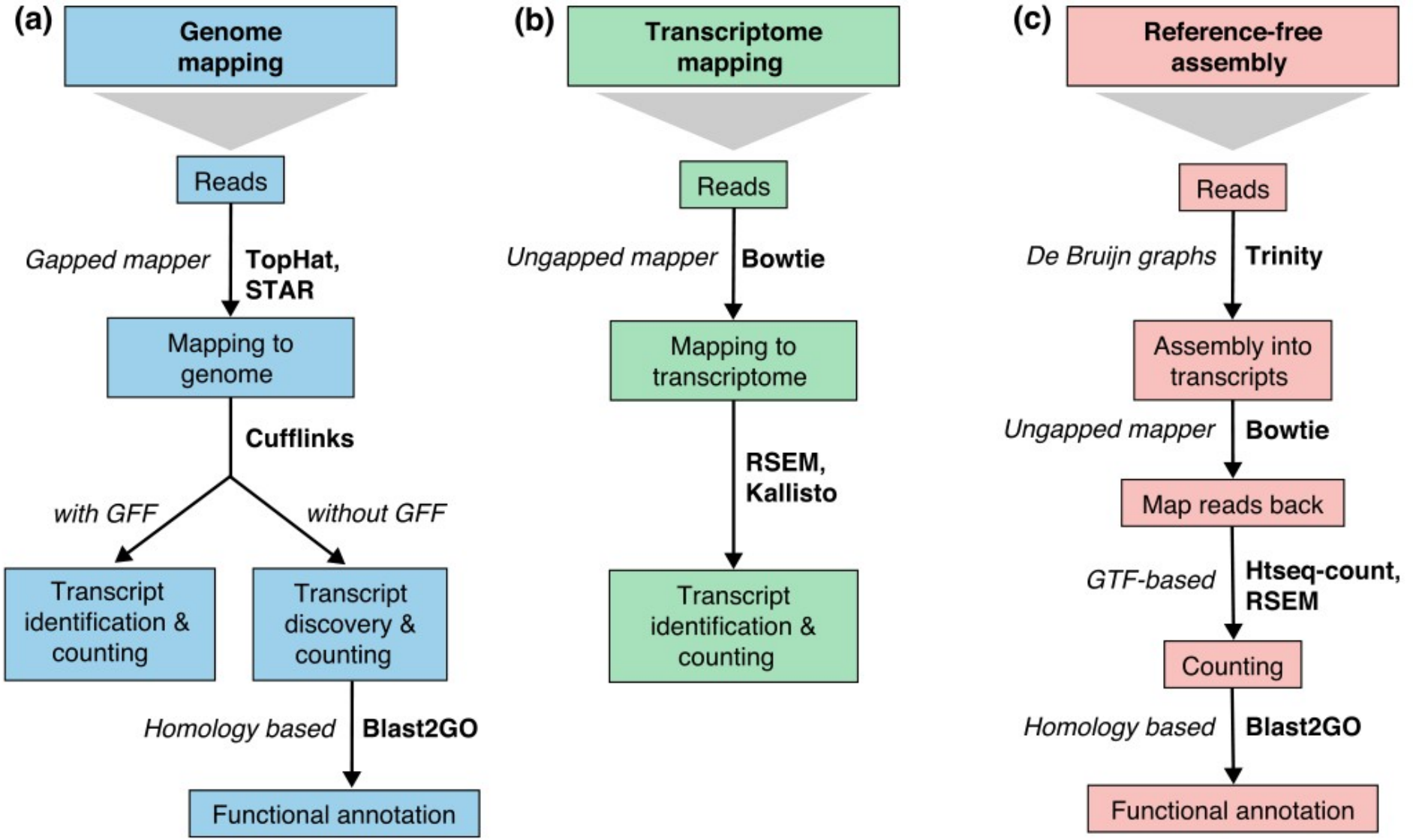
Nous devons aligner les données de séquence sur notre génome d'intérêt

- Si vous alignez les données RNA-Seq sur le génome, choisissez toujours un aligneur considérant l’épissage (sauf si c'est un génome bactérien!)

HiSat2, STAR, MapSplice, SOAPSplice, SpliceMap, GSNAP, HMMSplicer …

- Il existe d'excellents aligneurs disponibles qui ne sont pas sensibles à l’épissages. Ceci est idéal pour les génomes bactériens

BWA, Novoalign (not free), Bowtie2, SOAPaligner, HiSat2 ...




Conesa *et al.* (2016) Genome Biology

# Choix de l'aligneur

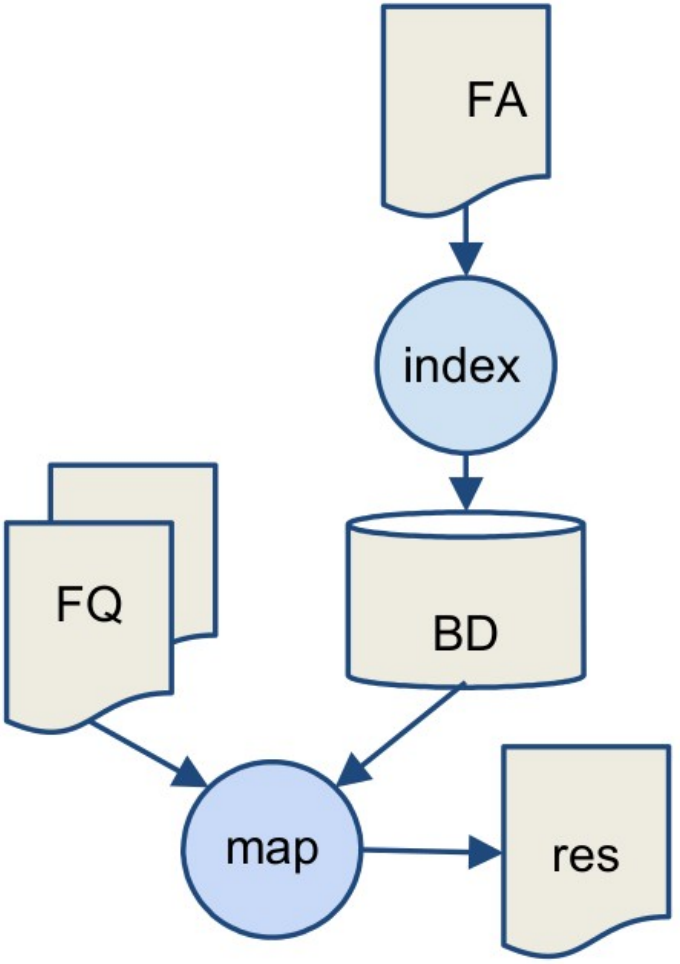
Autres considérations a prendre en compte dans le choix de l aligneur :

- Comportement de l’aligneur par rapport à l’alignement multiple
- Prise en charge d’alignement en paired-end ou single-end
- Le nombre de mismatches qu il autorise entre les reads et la reference
- Prise en charge de long-reads ou short-read ou des reads de tailles différentes

...

# Étapes de mapping

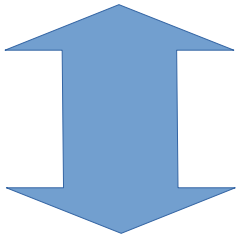
- Indexation du génome une fois pour toutes
- Mapping des lectures en utilisant l'index



# SAM (Sequence Alignment Map)

```
Coor     12345678901234  5678901234567890123456789012345
ref      AGCATGTTAGATAA**GATAGCTGTGCTAGTAGGCAGTCAGCGCCAT

+r001/1        TTAGATAAAGGATA*CTG
+r002         aaaAGATAA*GGATA
+r003       gcctaAGCTAA
+r004                     ATAGCT..............TCAGC
-r003                           ttagctTAGGC
-r001/2                                      CAGCGGCAT
```



```
@HD VN:1.5 SO:coordinate
@SQ SN:ref LN:45
r001   99 ref  7 30 8M2I4M1D3M = 37  39 TTAGATAAAGGATACTG *
r002    0 ref  9 30 3S6M1P1I4M *  0   0 AAAAGATAAGGATA    *
r003    0 ref  9 30 5S6M       *  0   0 GCCTAAGCTAA       * SA:Z:ref,29,-,6H5M,17,0;
r004    0 ref 16 30 6M14N5M    *  0   0 ATAGCTTCAGC       *
r003 2064 ref 29 17 6H5M       *  0   0 TAGGC             * SA:Z:ref,9,+,5S6M,30,1;
r001  147 ref 37 30 9M         =  7 -39 CAGCGGCAT         * NM:i:1
```

Header section

Alignment section

# SAM (Sequence Alignment Map)

| Name | Description |
|---|---|
| qname | Identifiant du read |
| flag | Valeur numerique qui encode les détails extra-alignement (voir samtools man page) |
| rname | Séquence de reference |
| strand | Brin d'alignement |
| pos | Coordonnée « start » de l'alignement |
| qwidth | Longueur du read |
| mapq | Qualité de l'alignement |
| cigar | Nomenclature CIGAR qui donne une vision globale de l'alignement |
| mrnm/rnext | reference chromosome/sequence to which opposite read pair aligned |
| mpos/pnext | alignment start coordinate to which opposite read pair aligned |
| isize/tlen | Taille de l'insert (paired-end) |
| seq | Séquence read qui s'aligne |
| qual | Score de la qualité |
| tags | Informations supplementaires sur l'alignement (optionnel) TAG:TYPE:VALUE |

# SAM (CIGAR)

| Symbole | Description |
|---|---|
| M | Alignement « matchant » |
| I | Insertion dans le read |
| D | délétion dans le read |
| N | Base illisible dans le read |
| S | Soft clipping (clipped sequences present in SEQ) |
| H | Hard clipping (clipped sequences NOT present in SEQ) |
| P | Padding (silent deletion from padded reference) |
| = | Séquence match |
| X | Séquence mismatch |

```
Exemple :

RefPos:     1  2  3  4  5  6  7     8  9 10 11 12 13 14 15 16 17 18 19
Reference:  C  C  A  T  A  C  T     G  A  A  C  T  G  A  C  T  A  A  C
Read:                  A  C  T  A  G  A  A     T  G  G  C  T

POS: 5
CIGAR: 3M1I3M1D5M
```

# SAM (Sequence Alignment Map)

| name | flag | chr | start | mapQ | cigar | nNext | sNext | tlen | seq | qual | tags |
|---|---|---|---|---|---|---|---|---|---|---|---|
| r001 | 99 | ref | 7 | 30 | 8M2I4M1D3M | = | 37 | 39 | TTAGATAAAGGATACTG | * | |
| r002 | 0 | ref | 9 | 30 | 3S6M1P1I4M | * | 0 | 0 | AAAAGATAAGGATA | * | |
| r003 | 0 | ref | 9 | 30 | 5S6M | * | 0 | 0 | GCCTAAGCTAA | * | SA:Z:ref,29,-,6H5M,17,0; |
| r004 | 0 | ref | 16 | 30 | 6M14N5M | * | 0 | 0 | ATAGCTTCAGC | * | |
| r003 | 2064 | ref | 29 | 17 | 6H5M | * | 0 | 0 | TAGGC | * | SA:Z:ref,9,+,5S6M,30,1; |
| r001 | 147 | ref | 37 | 30 | 9M | = | 7 | -39 | CAGCGGCAT | * | NM:i:1 |

| # | Decimal | Description of first read |
|---|---|---|
| 1 | 1 | Read paired |
| 2 | 2 | Read mapped in proper pair |
| 3 | 4 | Read unmapped |
| 4 | 8 | Mate unmapped |
| 5 | 16 | Read reverse strand |
| 6 | 32 | Mate reverse strand |
| 7 | 64 | First in pair |
| 8 | 128 | Second in pair |
| 9 | 256 | Not primary alignment |
| 10 | 512 | Read fails platform/vendor quality checks |
| 11 | 1024 | Read is PCR or optical duplicate |
| 12 | 2048 | Supplementary alignment |
| **Sum** | 99 | |

| Decimal | Description of second read |
|---|---|
| 1 | Read paired |
| 2 | Read mapped in proper pair |
| 4 | Read unmapped |
| 8 | Mate unmapped |
| 16 | Read reverse strand |
| 32 | Mate reverse strand |
| 64 | First in pair |
| 128 | Second in pair |
| 256 | Not primary alignment |
| 512 | Read fails platform/vendor quality checks |
| 1024 | Read is PCR or optical duplicate |
| 2048 | Supplementary alignment |
| 147 | |

Common flags*

One of the reads is unmapped:
73, 133, 89, 121, 165, 181, 101, 117, 153, 185, 69, 137

Both reads are unmapped:
77, 141

Mapped within the insert size and in correct orientation:
99, 147, 83, 163

Mapped within the insert size but in wrong orientation:
67, 131, 115, 179

Mapped uniquely, but with wrong insert size:
81, 161, 97, 145, 65, 129, 113, 177

* Collected from here

http://www.samformat.info/sam-format-flag

# BAM (Binary Alignment/Map)

BAM (Binary Alignment/Map) format:

- version compressée et binarisée de SAM
- samtools: reading, writing, and manipulating BAM Files
- la plupart des outils exploitant les fichiers bam, requière des bam triés et indexés


Autres formats : CRAM

# Visualisation des alignements via IGV