

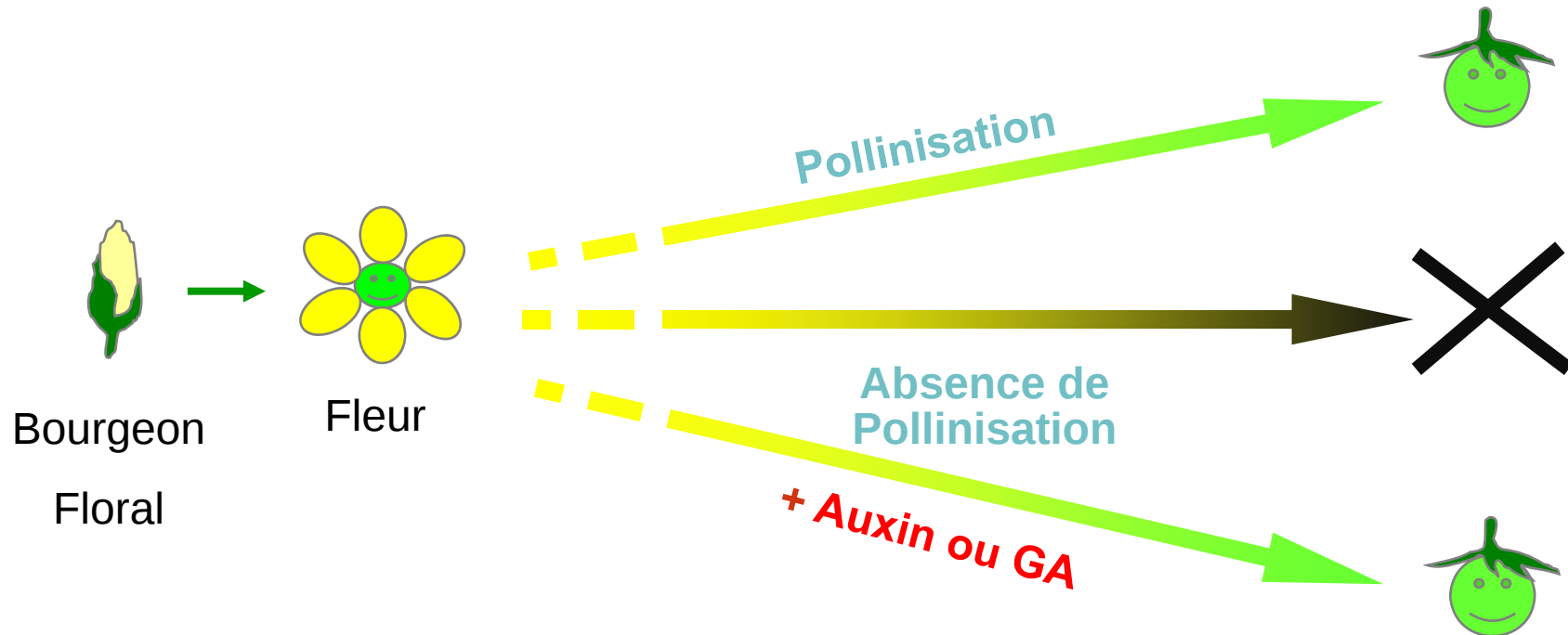


Analyse du développement précoce du fruit chez la tomate par une approche RNAseq

Mohamed Zouine
Laboratoire GBF

Le 7 octobre 2011

Le développement précoce du fruit



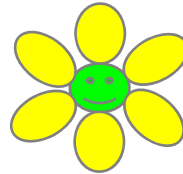
La transition Fleur-Fruit est régulée par un réseau complexe de signalisation hormonale

Questions biologiques:

Développement précoce du fruit



Bourgeon

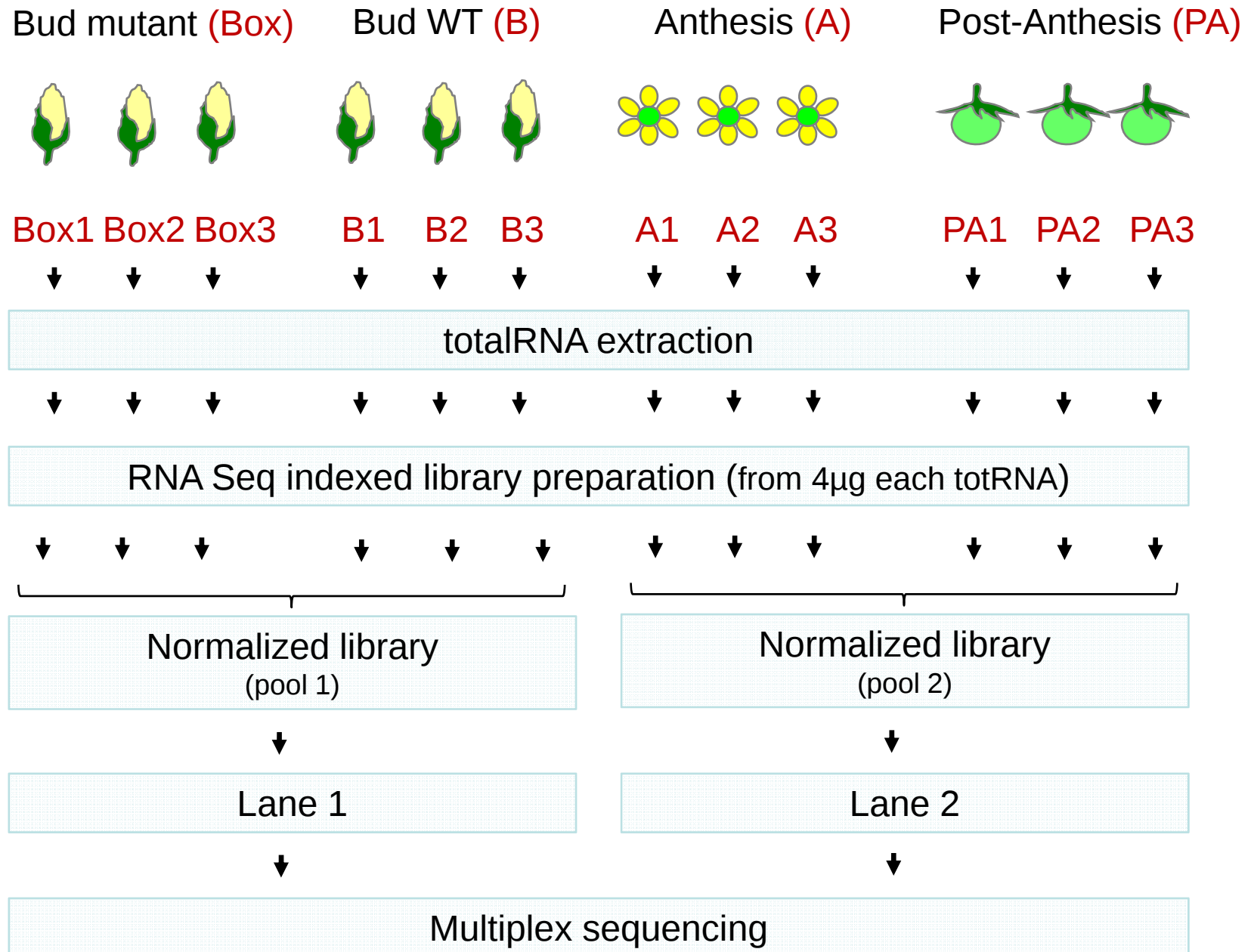


Fleur



Jeune Fruit

- Quels sont les gènes différentiellement exprimés au cours de la mise en place du fruit?
- Quels sont les gènes dérégulés dans les plantes sur-exprimant ARF8?



Raw Sequence data
Fastq files

QC by NG6
Sequence, quality, size

TopHat
Reference genome

Mapped reads on the
reference genome
BAM files

QC by
GBF

CuffLinks

Genes (loci) and their
expression levels
Expression files (FPKM)

CuffCompare
Annotation file

De novo transcripts
(genes) identification
New genes Model file (gff3)

QC by
coverageBed, GetORF,
Homology, iprscan

Cufflinks

Current and de novo
gene expression
levels in each sample
Expression files (FPKM)

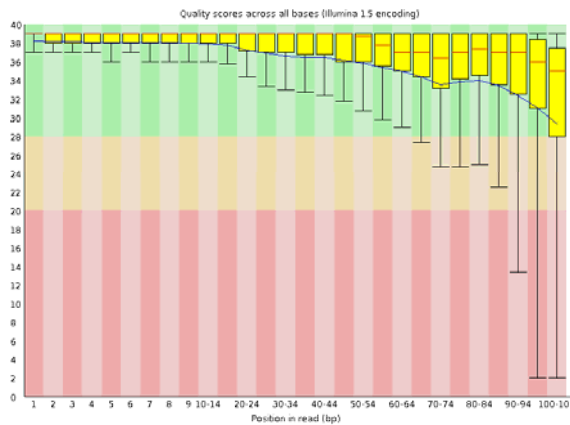
QC by
pearson correlation,
HC, BoxPlot

Cuffdiff

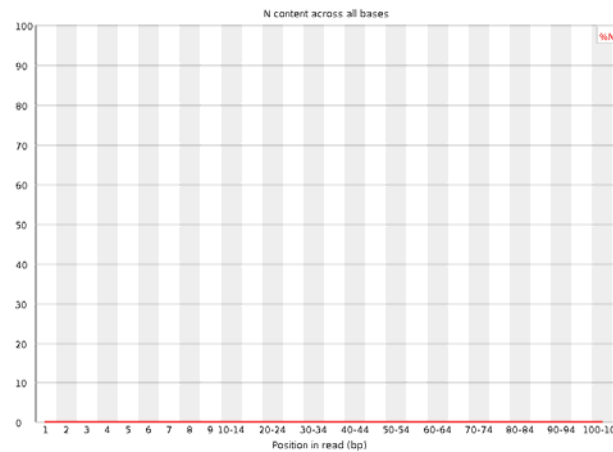
Differentially
expressed genes
(DE genes file: ratio, p-value, FDR...)

QC by
Q-PCR

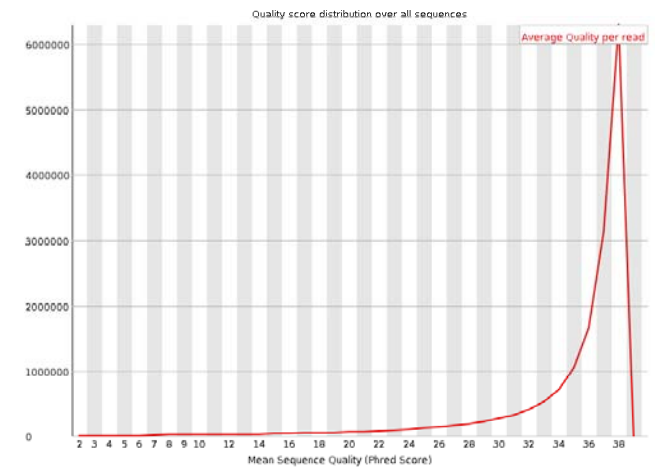
Control qualité des séquences (NG6)



Statistiques par position:
Qualité par position



Statistiques par position:
Nombre de « N »



Statistiques par séquence:
Qualité moyenne

TopHat
Reference genome

Raw Sequence data
Fastq files



QC by NG6
Sequence, quality, size

Mapped reads on the
reference genome
BAM files



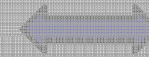
QC by
GBF

CuffLinks

Genes (loci) and their
expression levels
Expression files (FPKM)

CuffCompare
Annotation file

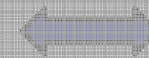
De novo transcripts
(genes) identification
New genes Model file (gff3)



QC by
coverageBed, GetORF,
Homology, iprscan

Cufflinks

Current and de novo
gene expression
levels in each sample
Expression files (FPKM)



QC by
pearson correlation,
HC, BoxPlot

Cuffdiff

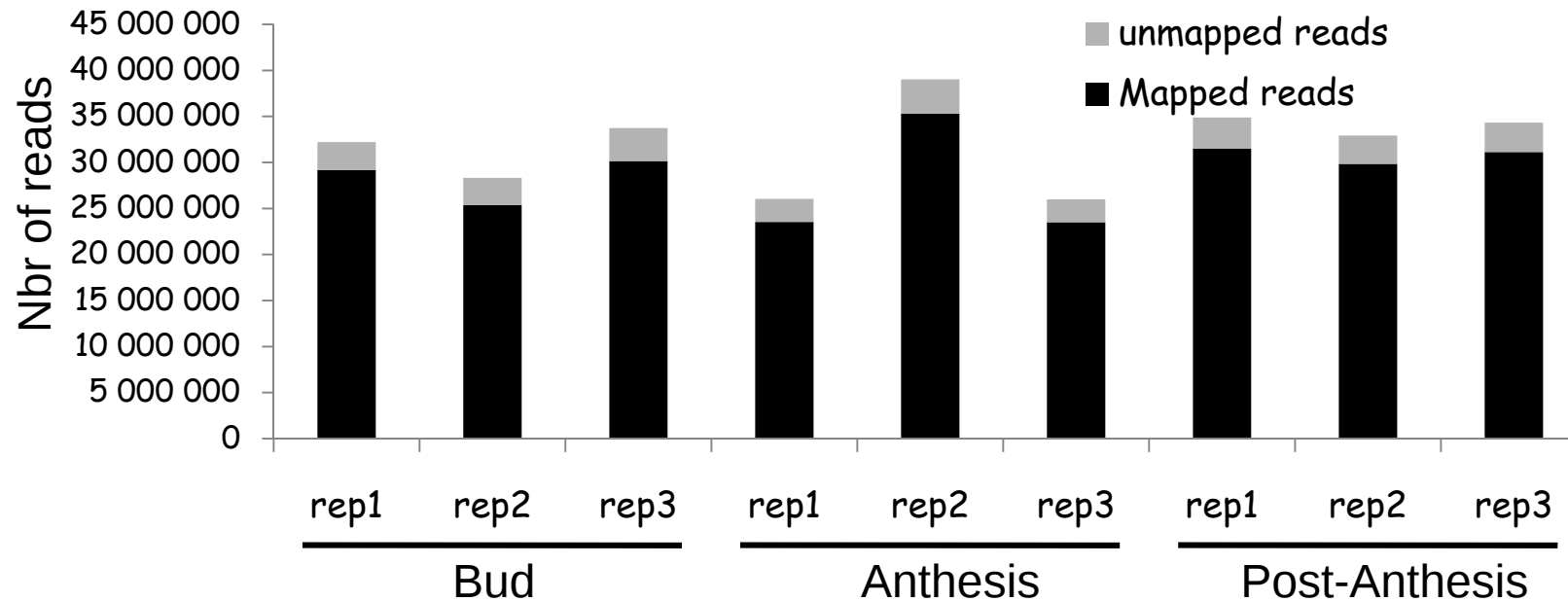
Differentially
expressed genes
(DE genes file: ratio, p-value, FDR...)

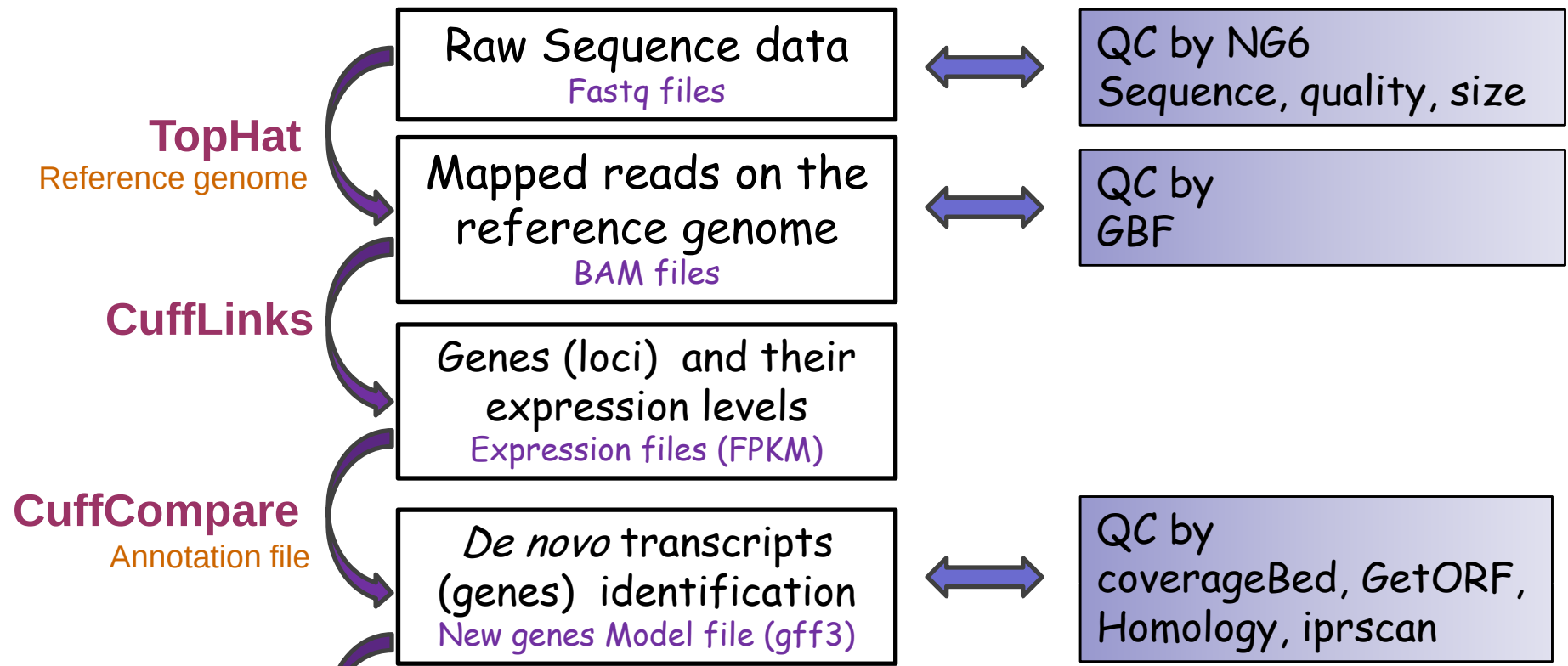


QC by
Q-PCR

Stats: Alignement des séquences par TopHat

	Bud			Anthesis			Post-Anthesis (4DPA)		
									
	B1	B2	B3	F1	F2	F3	P1	P2	P3
Nb reads	32 212 724	28 327 890	33 743 982	26 043 830	39 038 372	25 990 022	34 881 702	32 933 710	34 332 294
Nb reads mapped	29 183 046	25 369 984	30 154 705	23 537 063	35 285 876	23 509 723	31 524 530	29 852 279	31 131 389
Nb reads not mapped	3 029 678	2 957 906	3 589 277	2 506 767	3 752 496	2 480 299	3 357 172	3 081 431	3 200 905



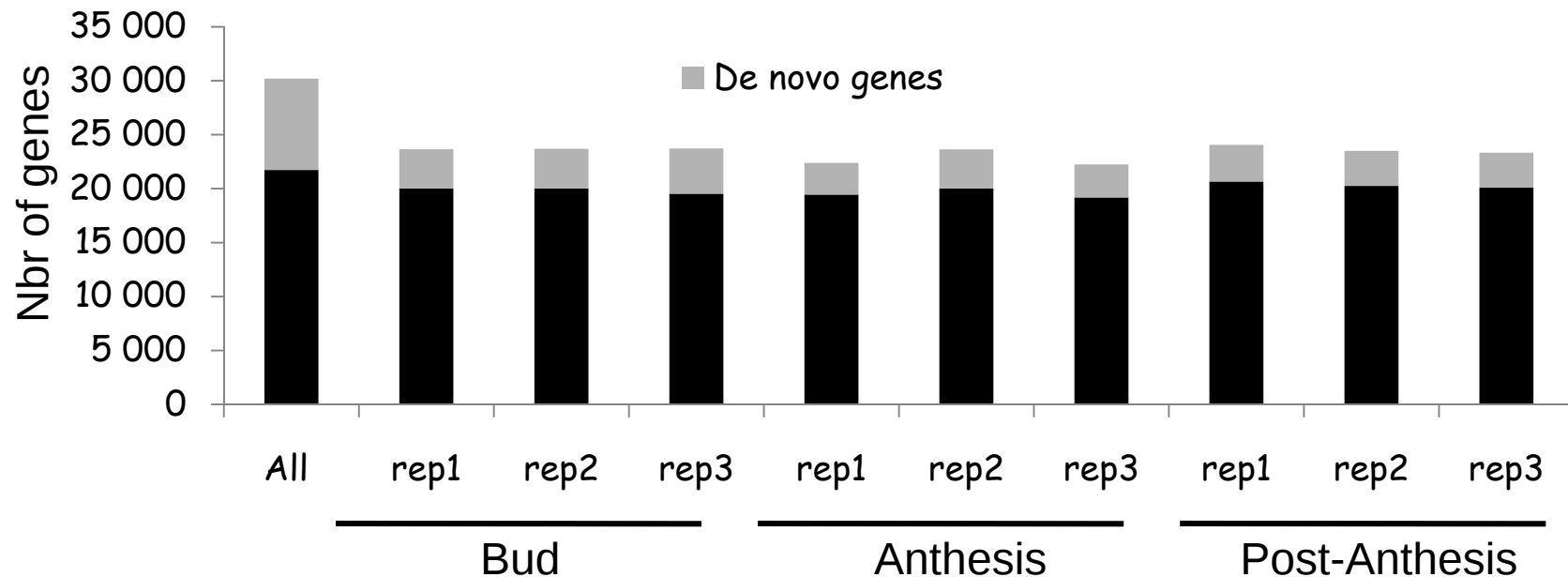


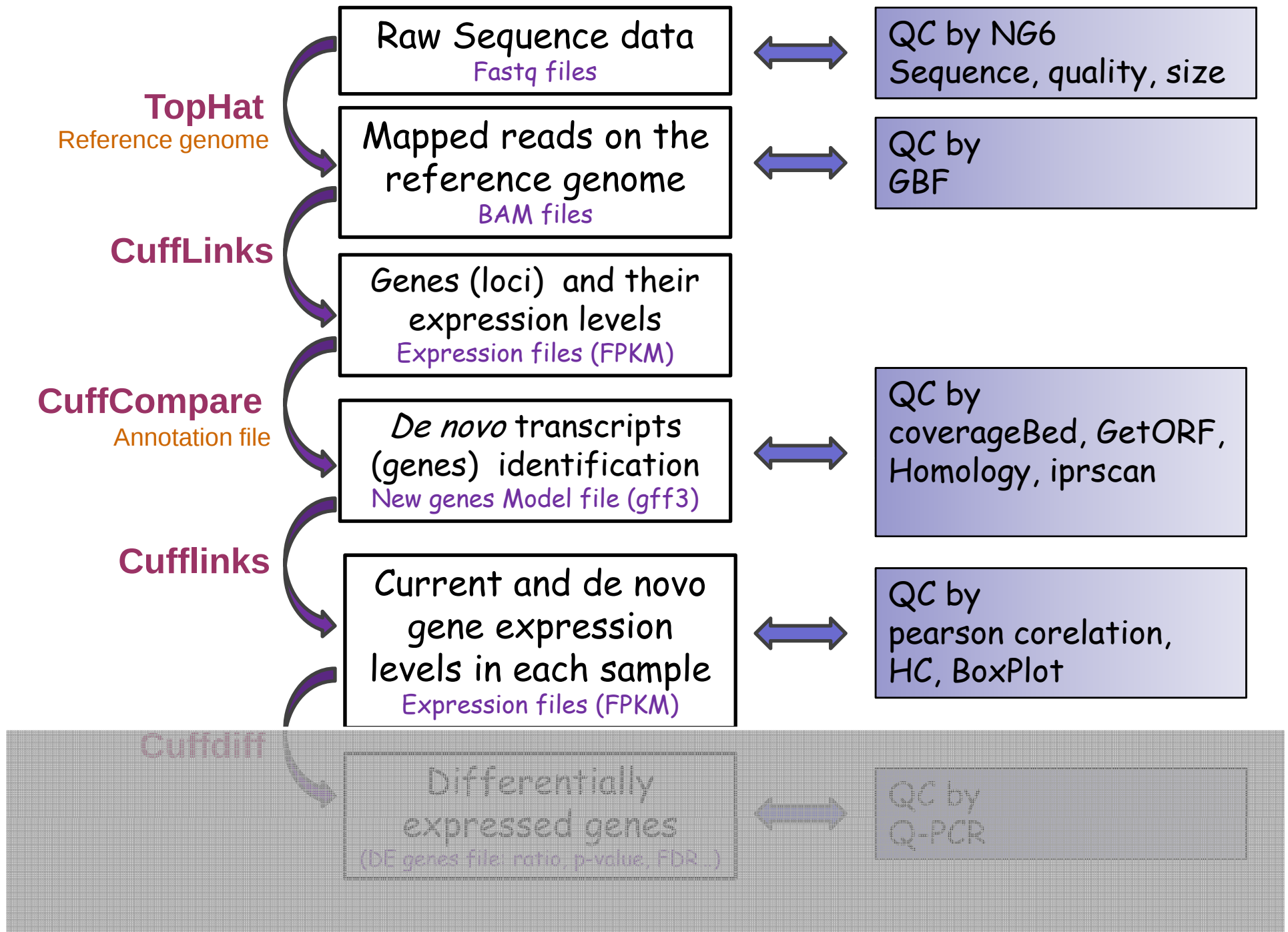
Cufflinks

Cuffdiff

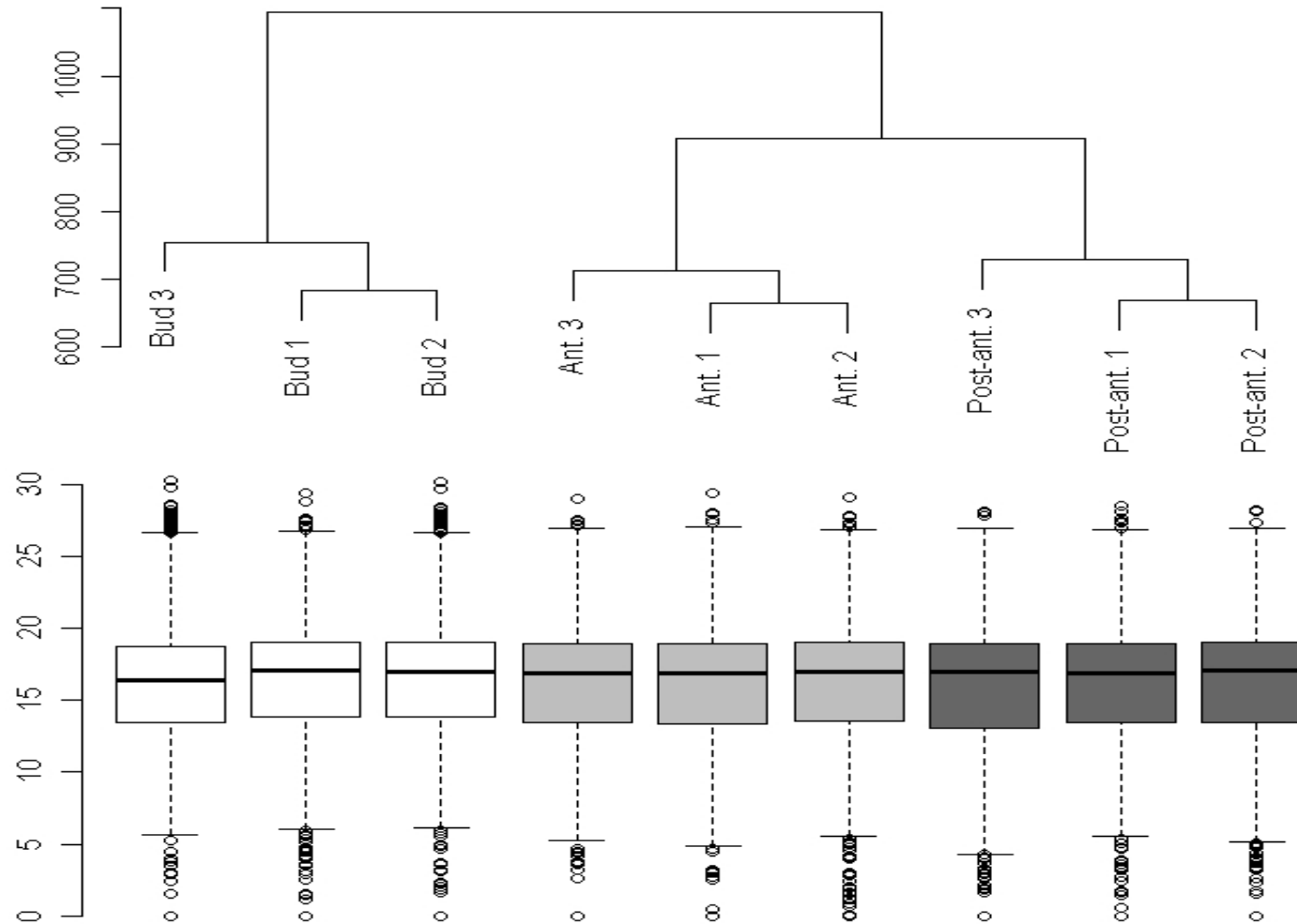
Stats: Expression des gènes par Cufflinks: Découverte de nouveaux gènes (*de novo*)

		Bud			Anthesis			Post-Anthesis		
	All	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3
Total of expressed loci	30 200	23 255	23 305	23 374	21 999	23 281	21 869	23 573	23 119	22 952
Known genes	21 735	20 015	20 026	19 531	19 443	20 019	19 182	20 667	20 252	20 095
<i>De novo</i> genes	8 462	3 638	3 672	4 199	2 949	3 618	3 059	3 393	3 252	3 244
<i>De novo</i> gens after QC	7 236	2 433	2 449	2 925	1 785	2 415	1 887	2 096	1 985	1 996
<i>De novo</i> protein coding genes	4 641	1 669	1 536	1 715	1 239	1 662	1 320	1 398	1 366	1 396

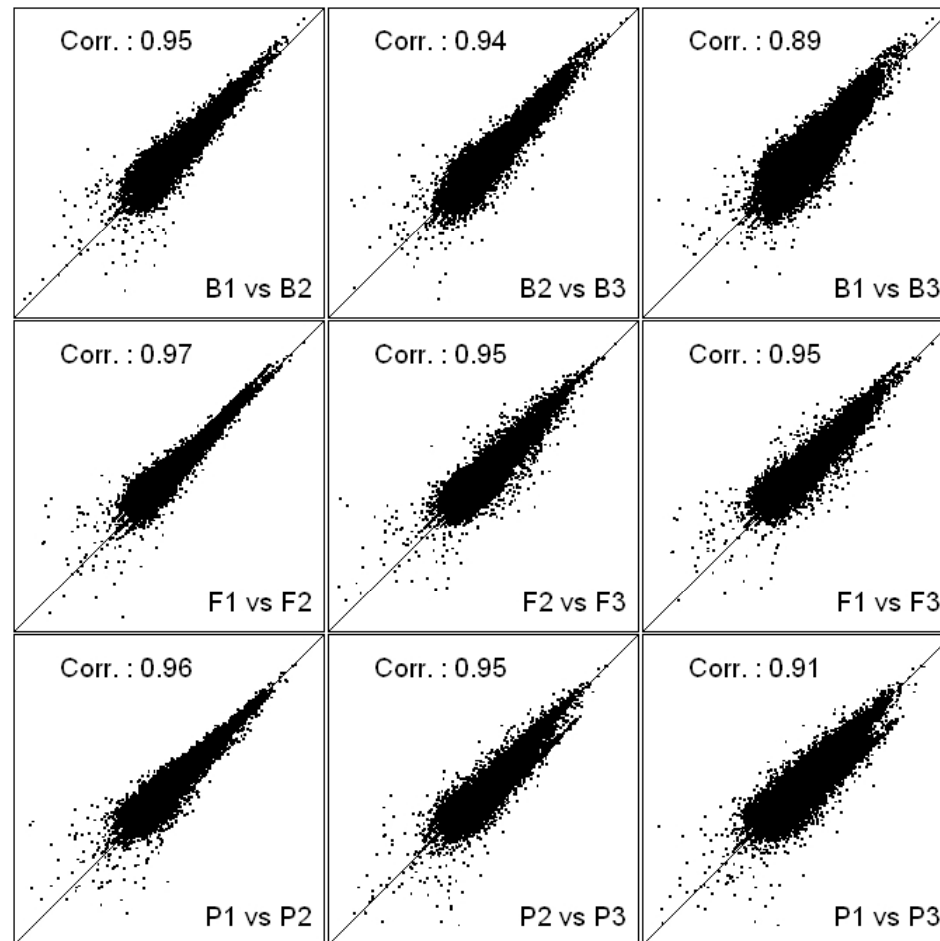


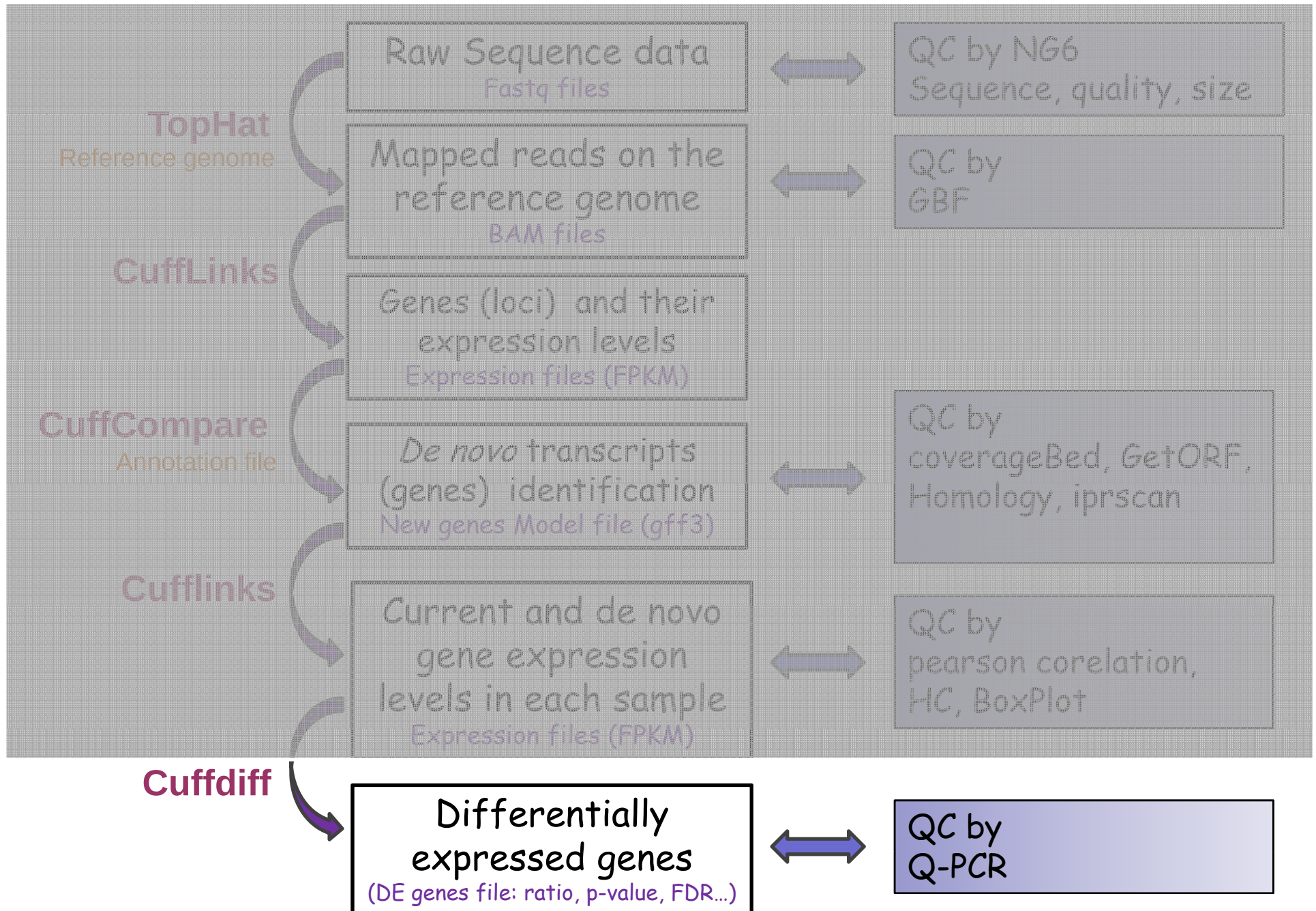


Control qualité des répétitions biologiques: Classification hiérarchique et taux d'expression globale



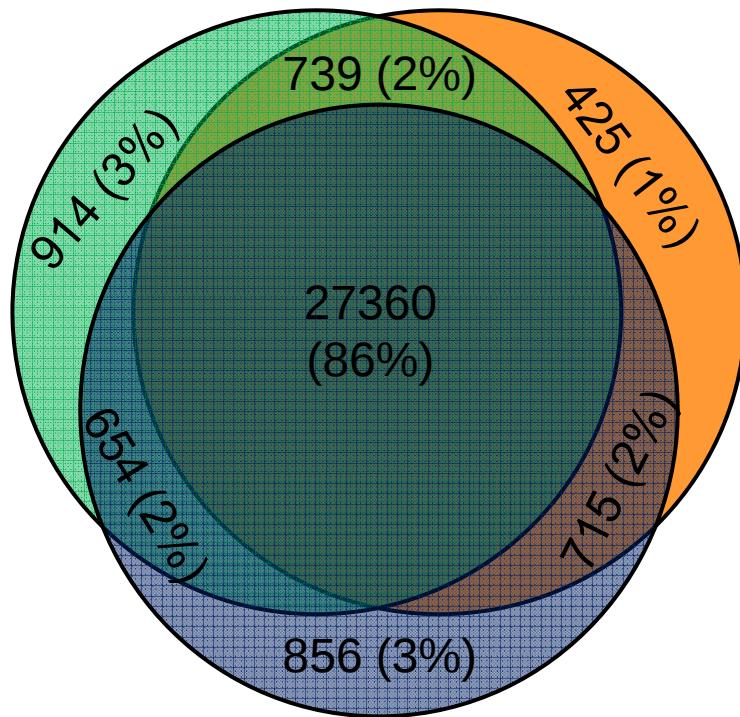
Control qualité des répétitions biologiques: Coefficient de corrélation de Pearson





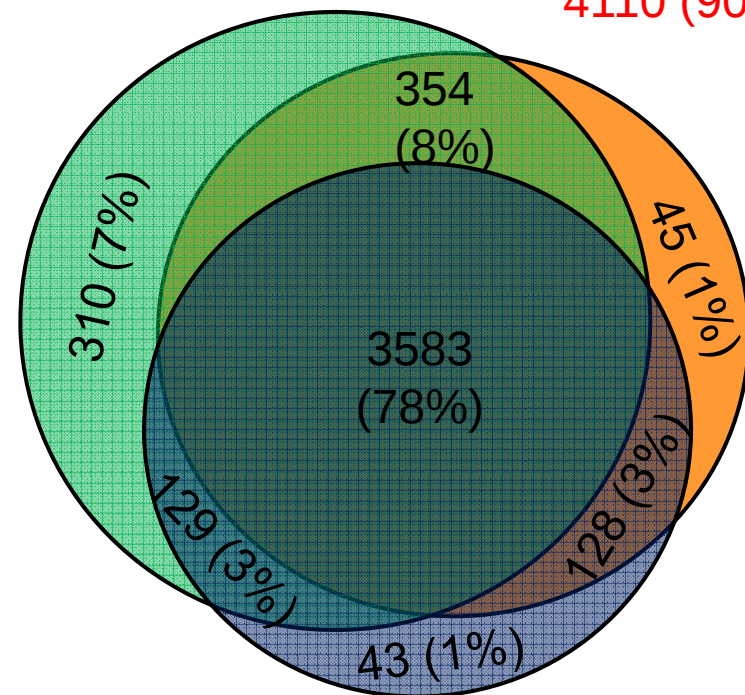
Plus de gènes identifiés dans le stade « bourgeon floral »

Bud 29667 (93%)
All genes
Anthesis 29239 (91%)



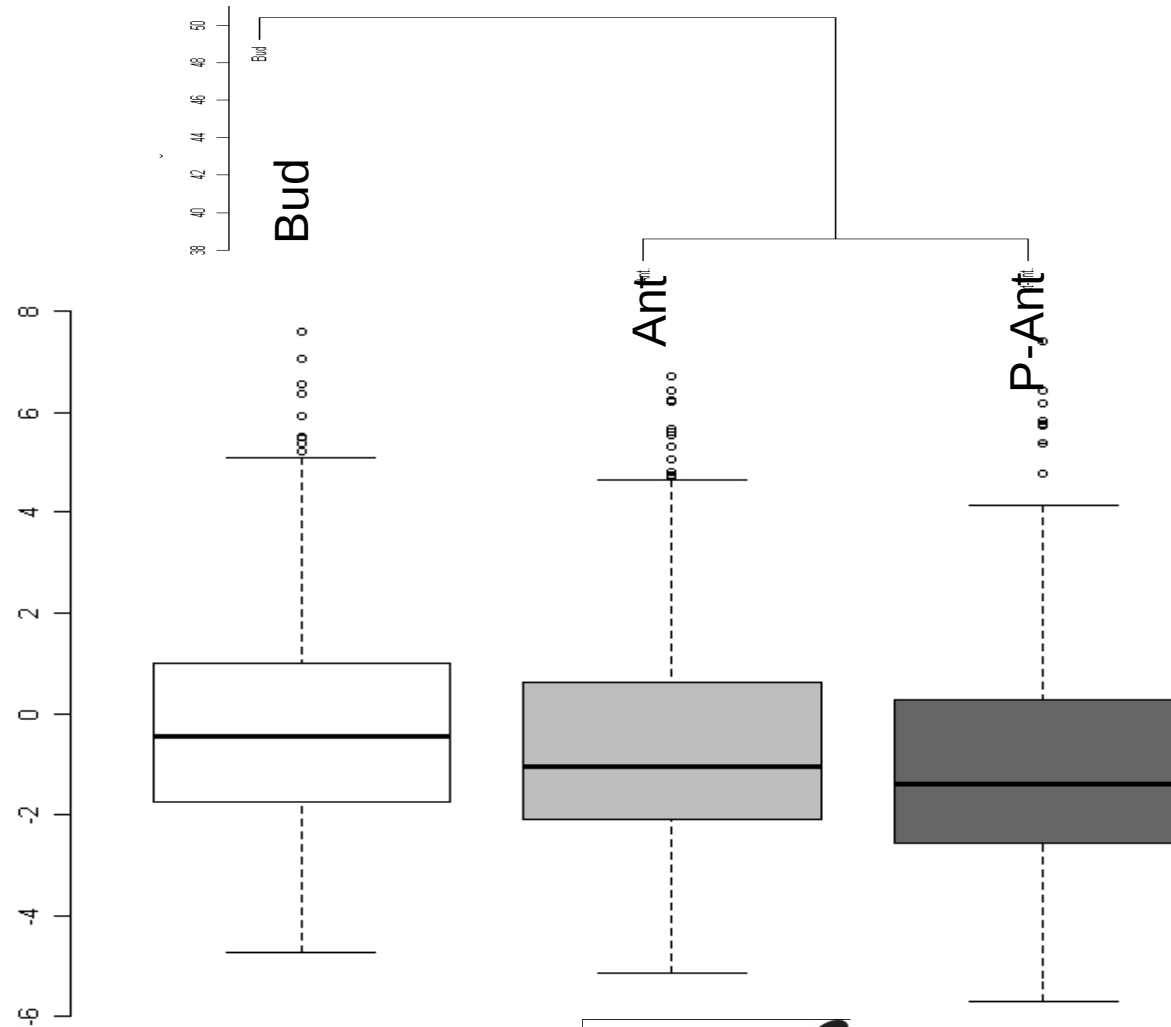
Post-anthesis
29585 (93%)

Bud 4376 (96%)
De novo
Anthesis 4110 (90%)

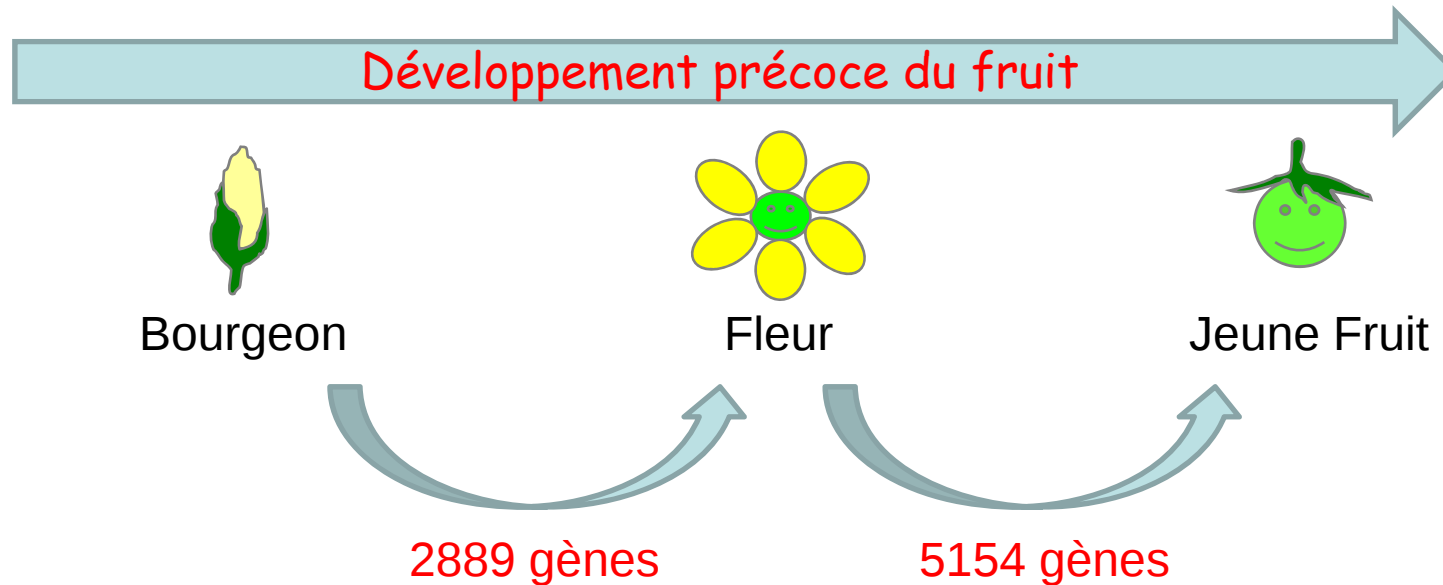


Post-anthesis
3883 (86%)

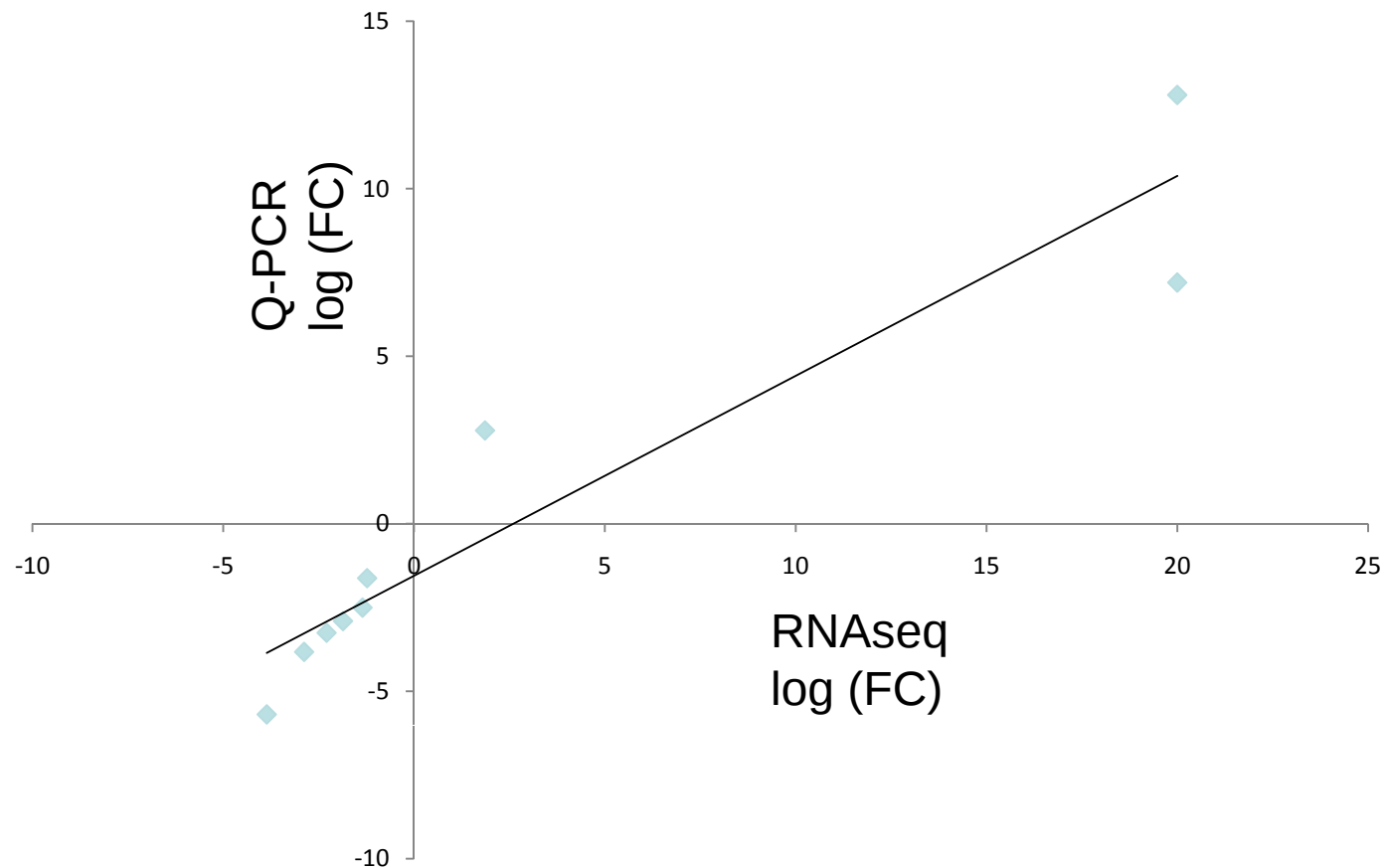
Plus forte expression des gènes dans le stade « bourgeon floral »

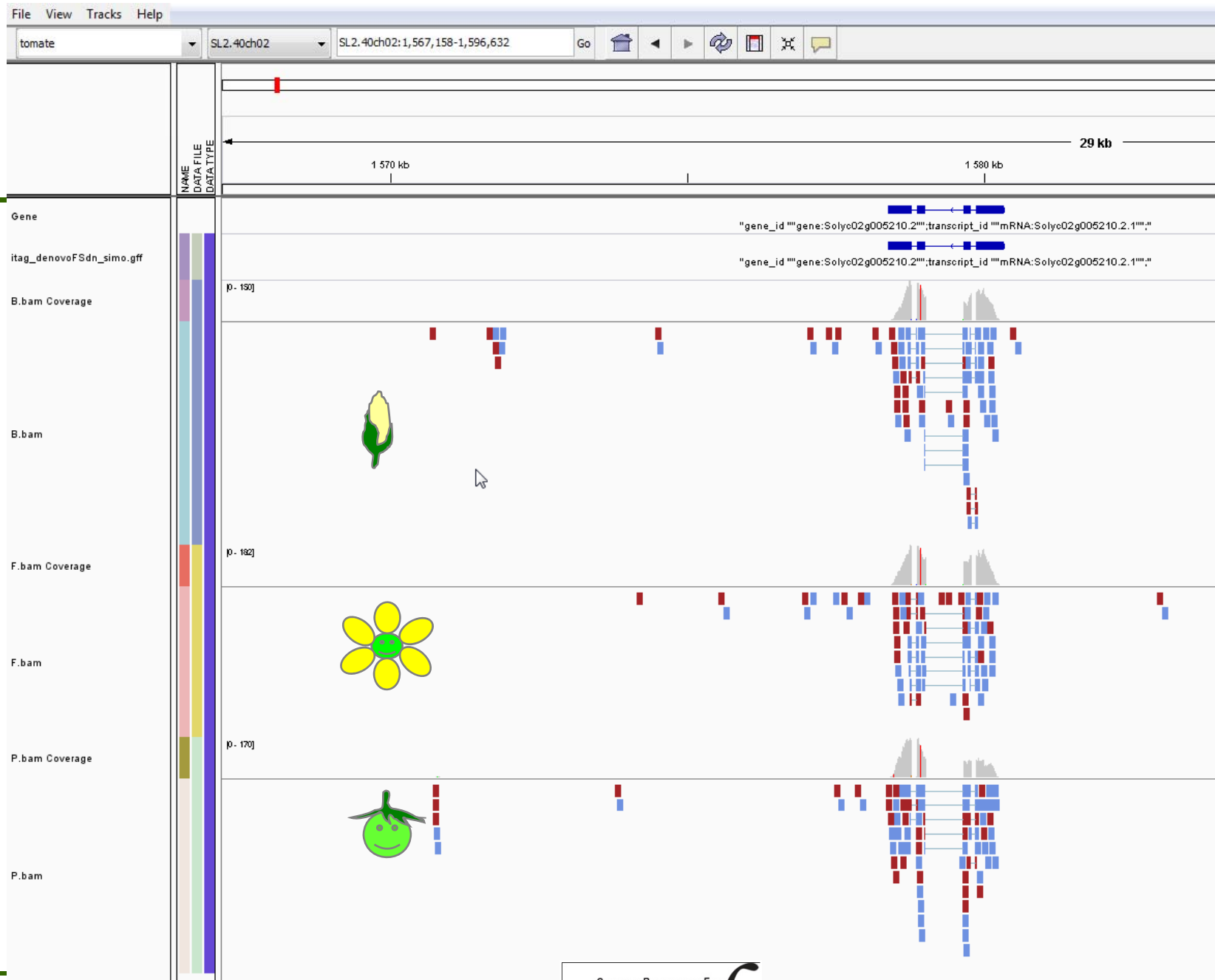


Question de départ: gènes différentiellement exprimés au cours de la mise en place du fruit (CuffDiff)

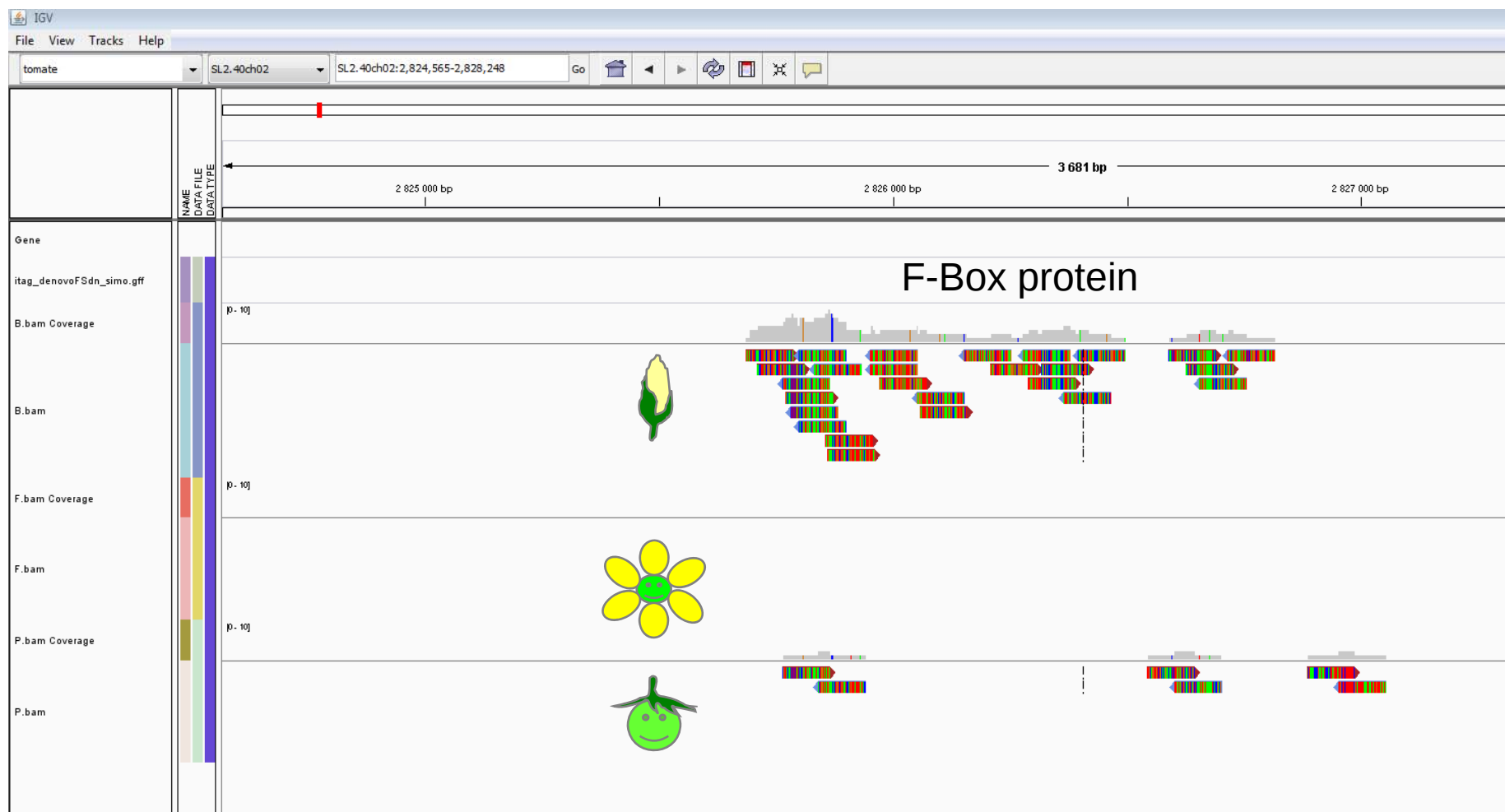


Validation par PCR quantitative

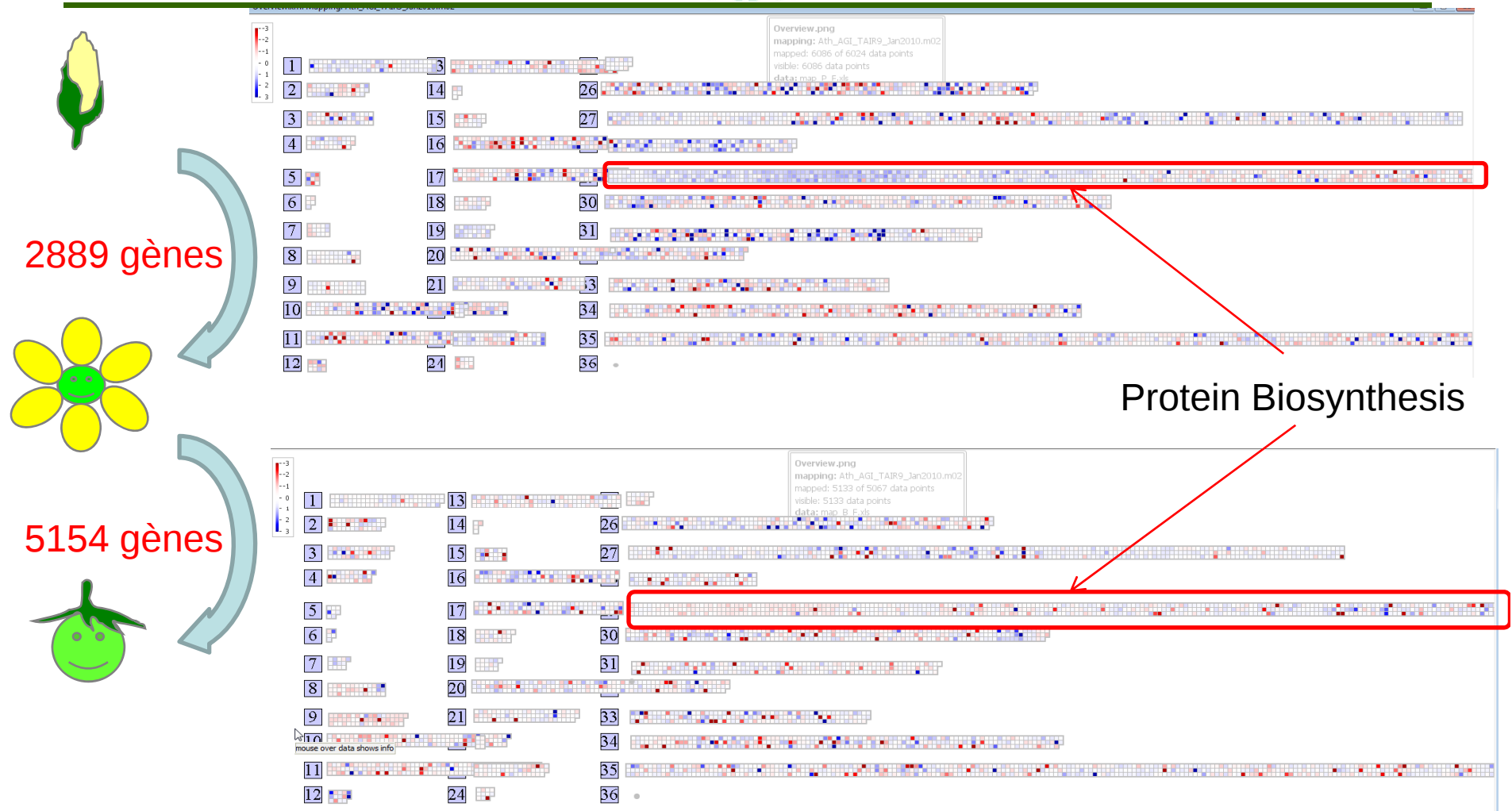




QC par IGV



Data mining par MapMan



Conclusions

- Volet séquençage
 - Aspects positifs:
 - Séquences de très bonne qualité
 - Bonne profondeur
 - Aspect Négatif
 - Séquençage non orienté
- Analyse bioinformatique
 - Evolution rapide des logiciels d'analyses

Remerciements

Laboratoire GBF

- Pierre Frasse
- Elie Maza
- Christine Rousseau
- Yougyao Fu
- Mondher Bouzayen

Plateforme GeT-PlaGe

- Nathalie Marsaud
- Olivier Boucher
- Gerald Salin

Plateforme genotoul-Bioinfo

- Christophe Klopp
- Jerome Mariette