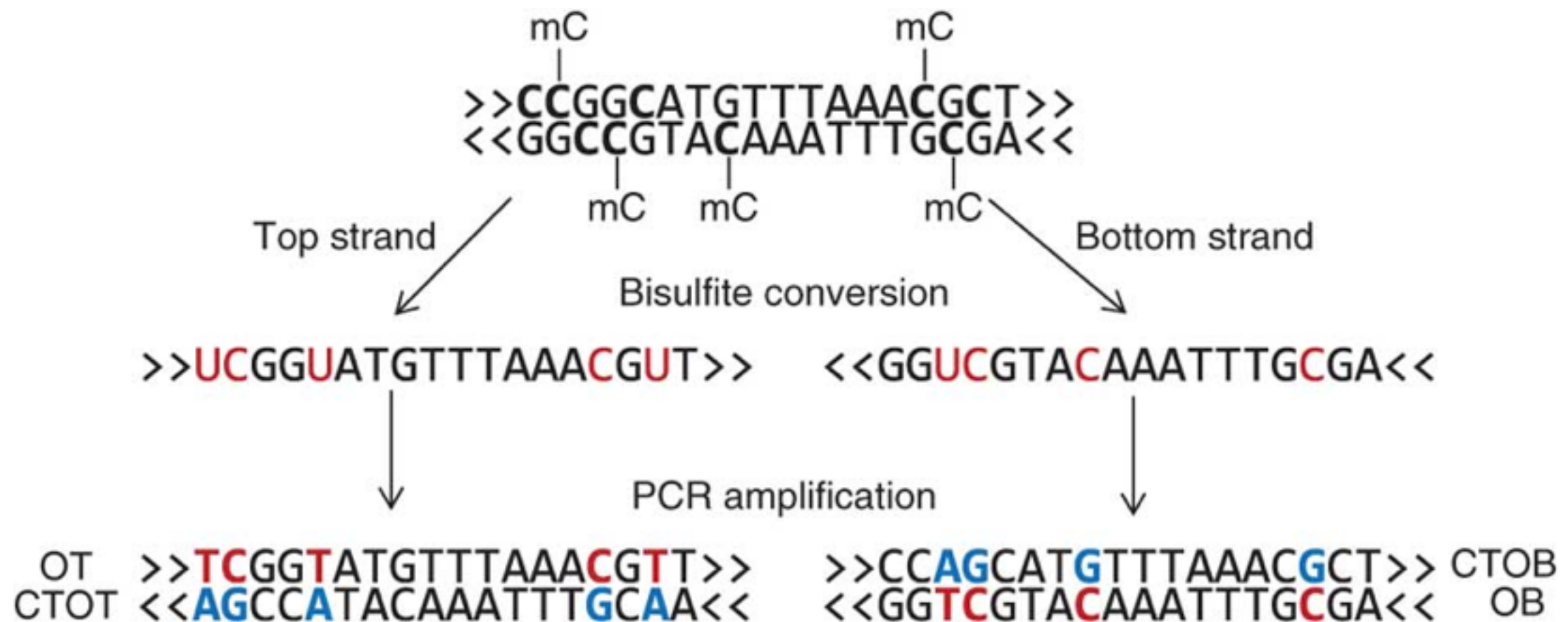


BiohackUdeM

Team #Meth (and then #SVM)

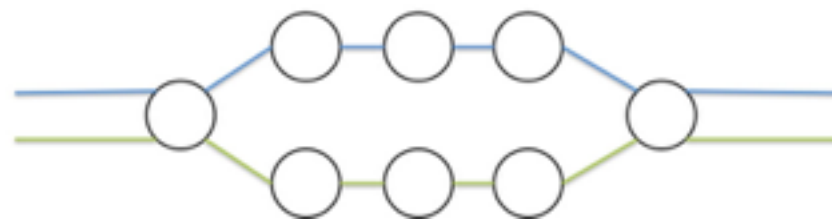
Méthylation



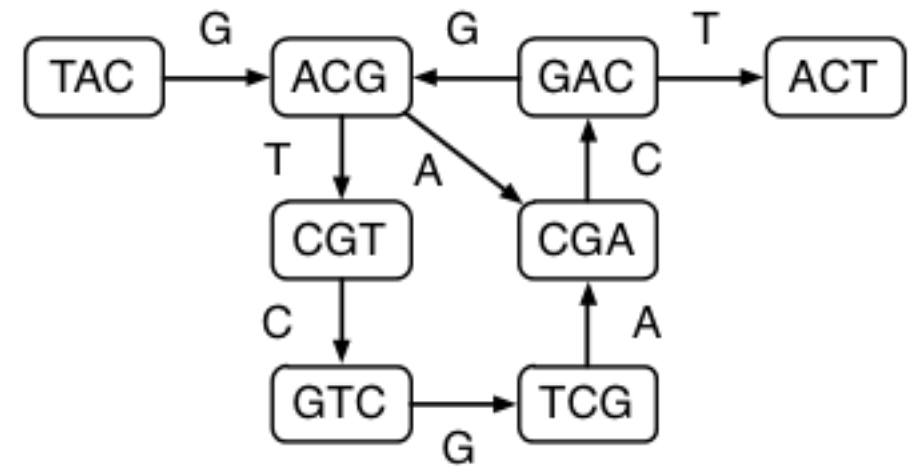
Détection de SNV sans alignement



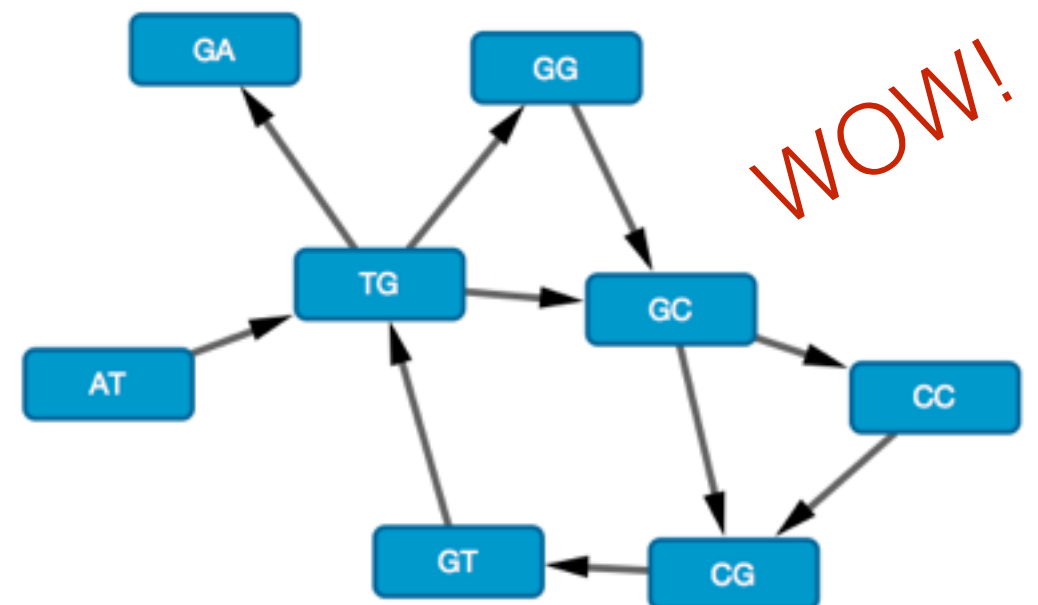
Sample 1
Sample 2



bubulles



De Bruijn graph



Nouvelle idée

FLASH

Fast Length Adjustment of Short reads

ACTTCGTACTTACGTAAT

TGCATTATTAGCATTAA

ACTTCGTACTTACGTAATAATCGTAATT

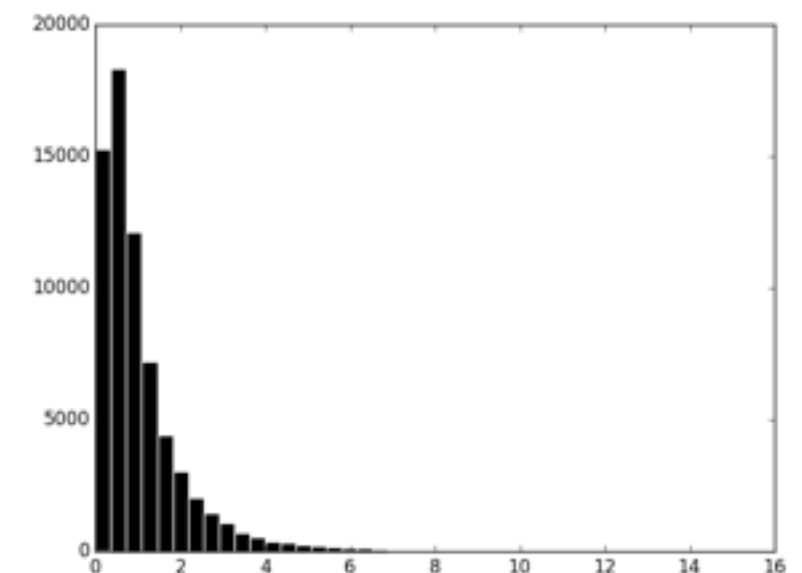
EMBL-EBI



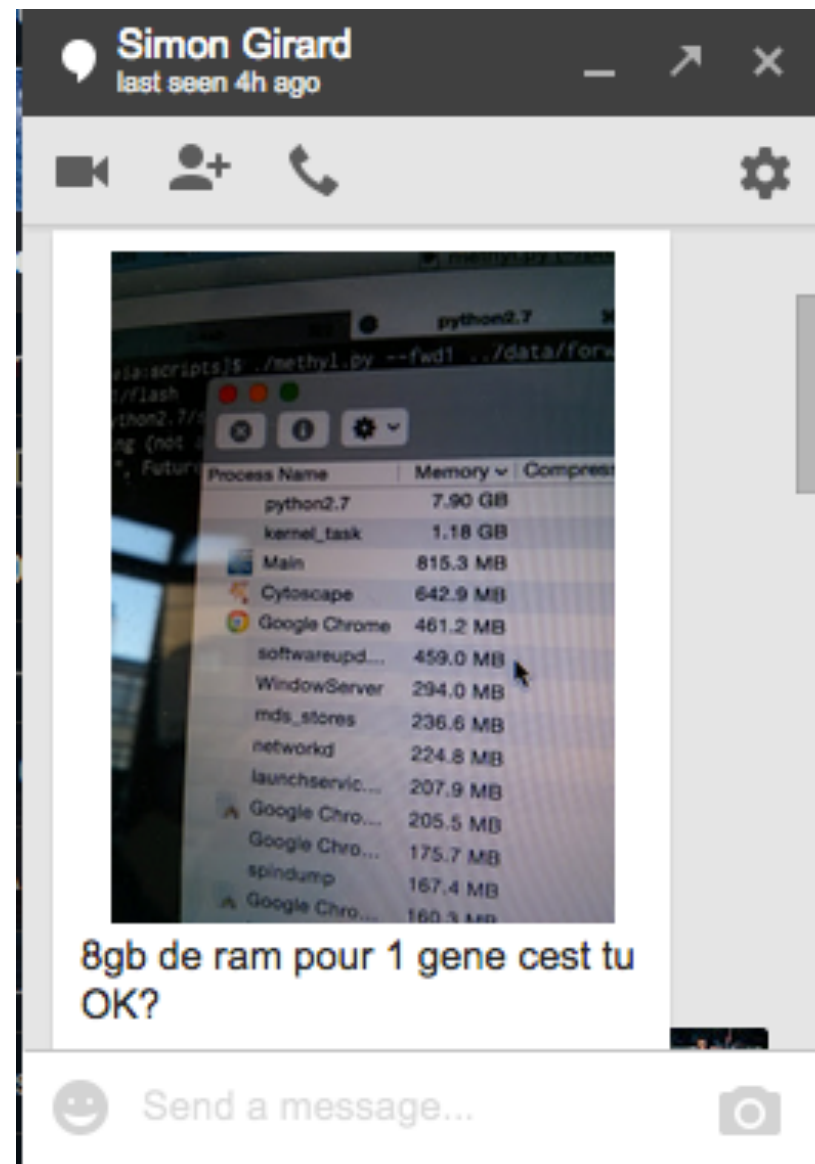
Velvet

Sequence assembler for very short reads

Identification de contigs *fwd* et *rev*
“homologues”



à 8h a.m.



Classification

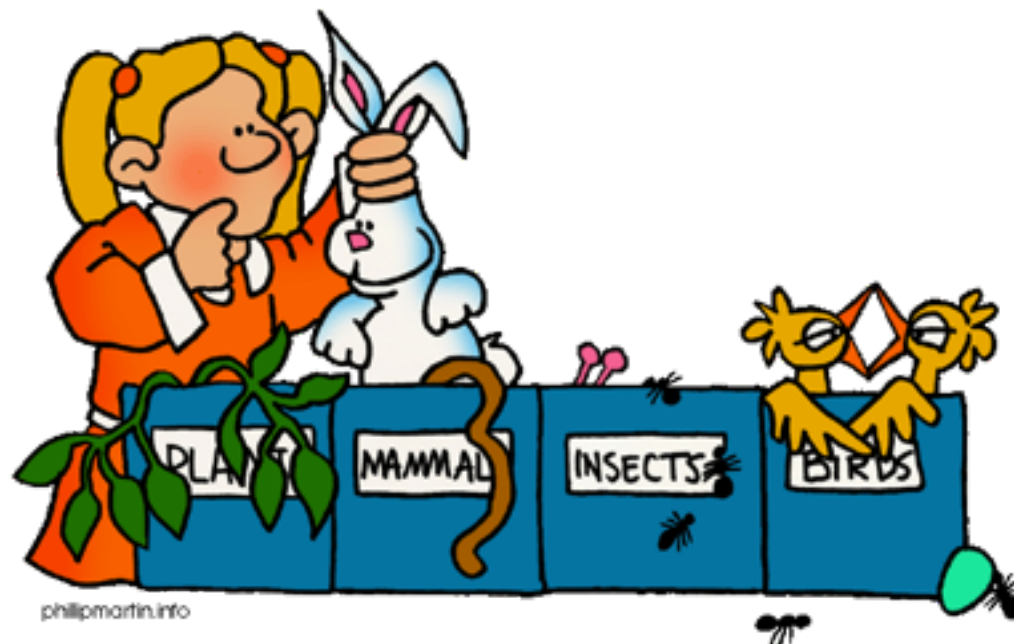
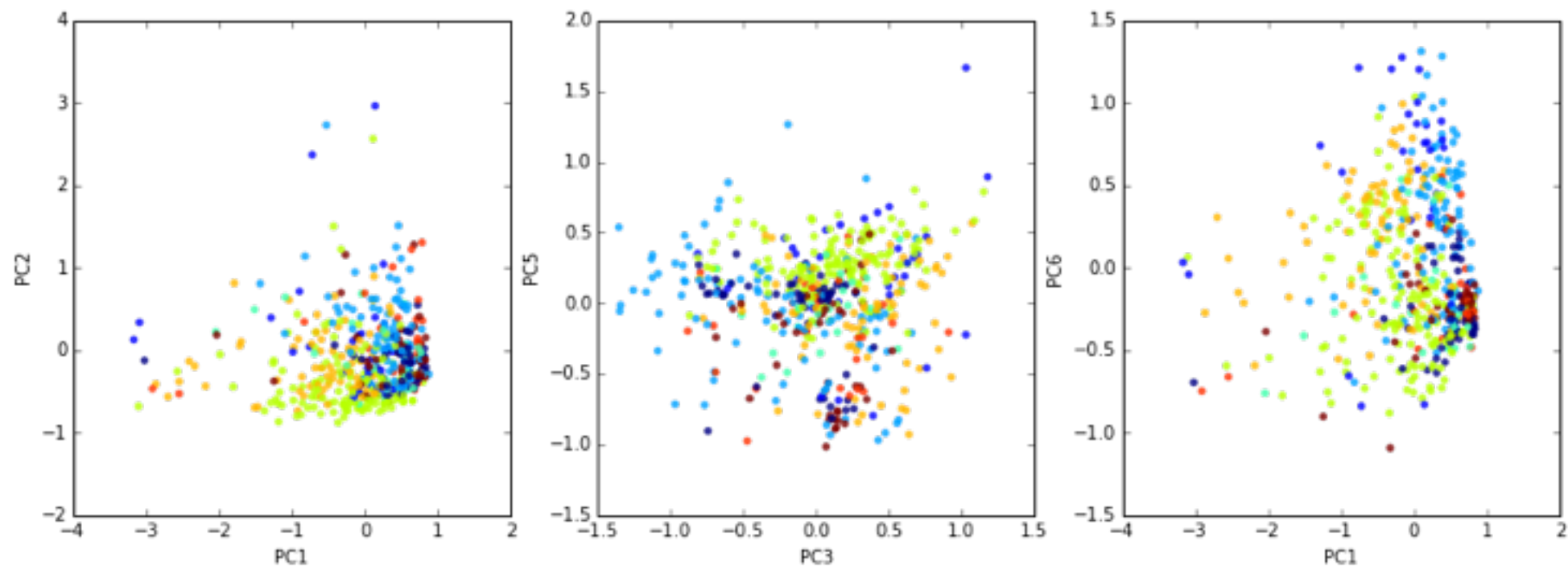
- 8 tissus (outcome)
- Gènes porteurs de SNV, insertions et délétions (prédicteurs)



- Gènes tissus-spécifiques (<3 tissus)
- Gènes majoritairement représentés (top 450)
- Total: 11,102 / 16,253 gènes
- Présence/absence de insertion/délétion par chromosome

Dimensionality reduction

- De 11,150 features à 900 composantes principales



Classification: SVM (linéaire)

Pénalité L2

Test set (500 nouveaux samples)

Accuracy: 78%

(8 classes)

Accuracy de 74% sur le
dataset "secret"

Merci!