

le cnam



L'implémentation d'outils de priorisation dans un pipeline d'analyse de données NGS

Méline WERY, Licence Professionnelle BioInformatique

Responsables : Pr Alain MOREL , Jonathan DAUVÉ

Laboratoire d'Oncopharmacologie

Institut de Cancérologie de l'Ouest – Site Paul PAPIN

Le cancer en quelques chiffres

- Selon le rapport de l'INCa en 2015 :
 - Plus de 380 000 nouveaux cas en France
 - Les plus fréquents :
 - Prostate, Poumon, Côlon chez l'homme
 - Sein, Côlon, Poumon chez la femme
 - Près de 150 000 décès recensés
- Principaux traitements :
 - Chirurgie
 - Radiothérapie
 - Chimiothérapie
 - Thérapie ciblée (nombreux essais cliniques)

=> Déterminer les cibles les plus pertinentes

Les axes du laboratoire

Génétique constitutionnelle

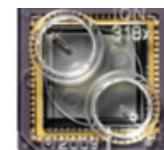
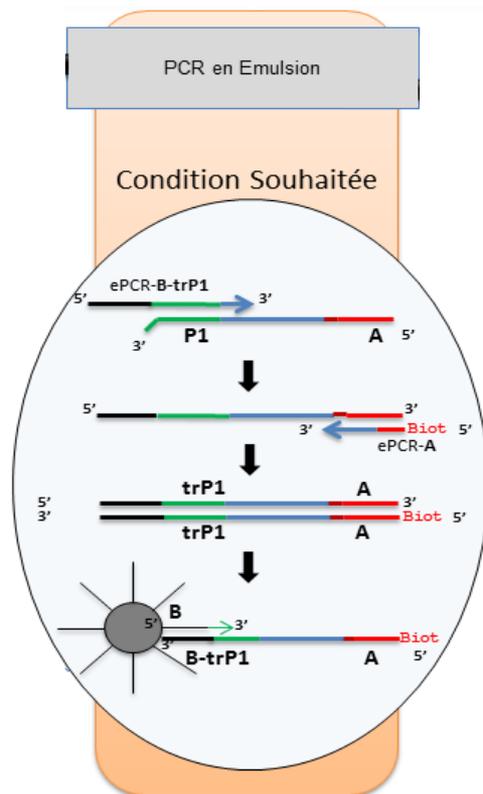
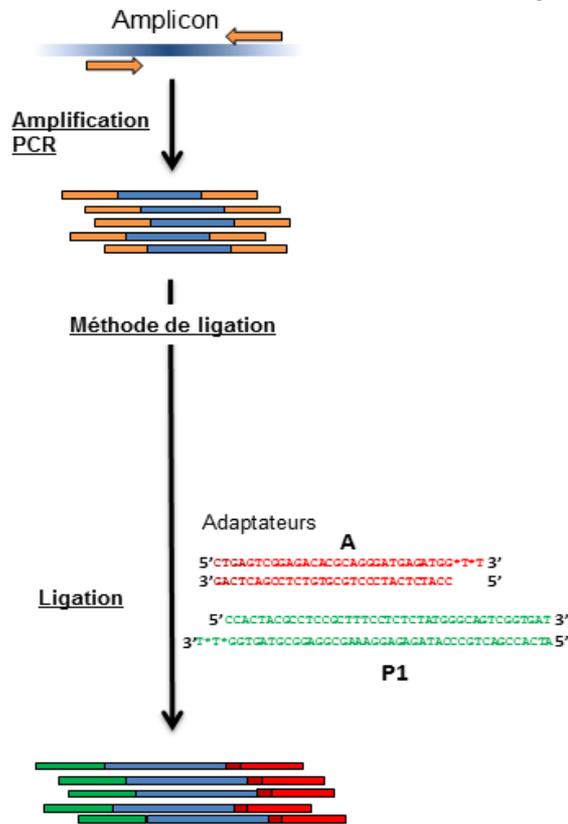
- Profil du patient
- Adaptation du traitement
- Analyse de protéines impliquées dans la métabolisation du principe actif (gène et activité)

Génétique somatique

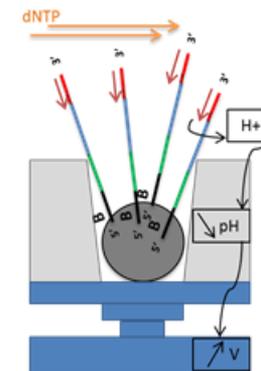
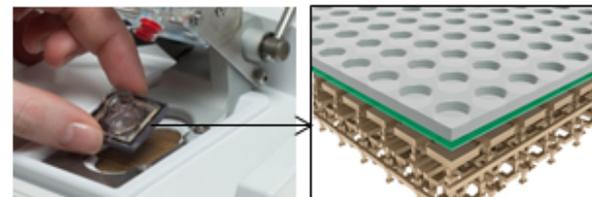
- Profil de la tumeur
- Mise en place de thérapie ciblée
- Analyse de protéines clés de la tumorigénèse (analyse du génome de la tumeur)

L'analyse du génome : NGS (1)

Séquençage par PGM IonTorrent



Puce 318
1Gb
(~11 M puits)



Préparation de la
bibliothèque



Amplification par
PCR en émulsion



Séquençage

L'analyse du génome : NGS (2)

Base Calling

- Association à une base
- Score de qualité

Torrent Mapping Alignment Program

- Génome humain (hg19)
- Fichier BAM (fichier binaire)

Torrent Variant Caller

- Recherche de variants
- Filtre (fréquence, couverture)
- Fichier VCF (position variant, nucléotides, valeurs filtre)

Annotation via pipeline d'analyse (Python)

- Annovar
- Liste d'informations (gène, mutation, référence COSMIC, dbSNP)

OBJECTIF

Ajouter au pipeline d'analyse des outils permettant de prioriser les variants

Les outils utilisés

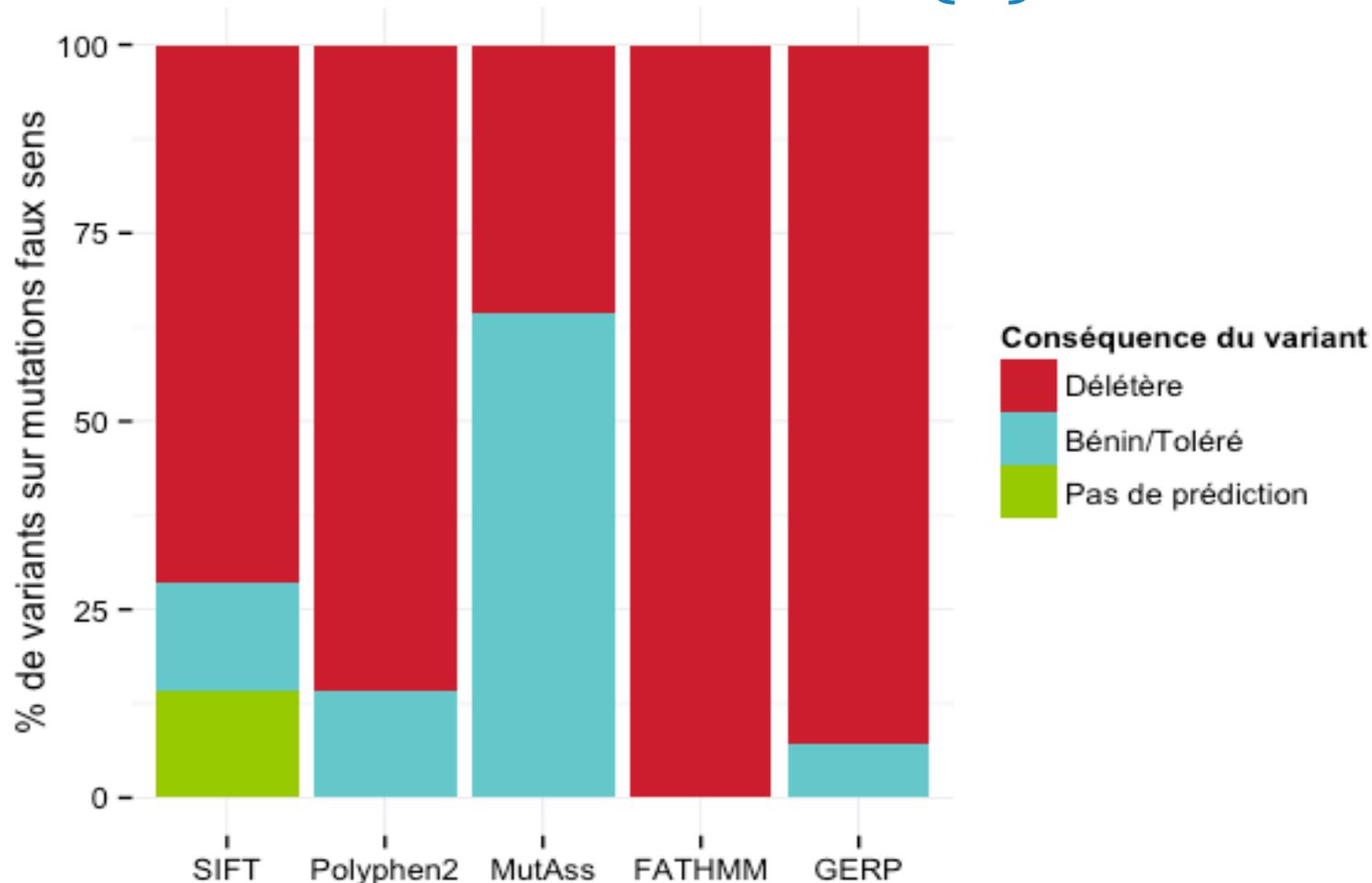
Score de conservation

- Alignement type BLAST :
 - SIFT
 - PolyPhen2
 - MutationAssessor
- Alignement via HMM :
 - FATHMM
 - Utilisation des HMM, de bases de données (HGMD, Uniprot)
 - GERP
 - Utilisation arbre phylogénétique

Base de données

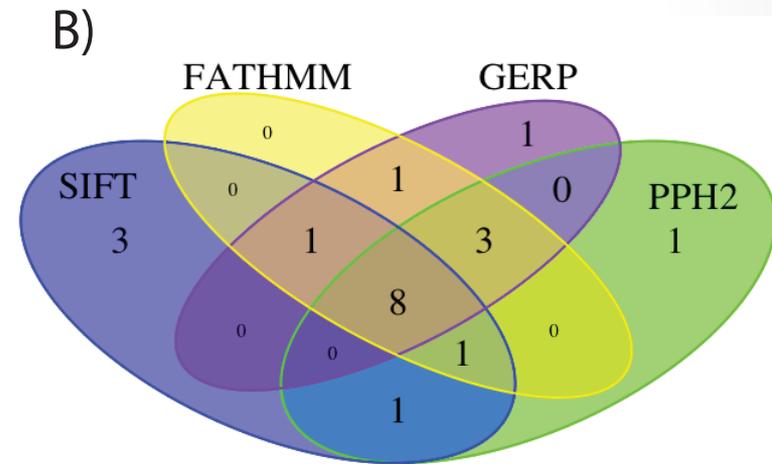
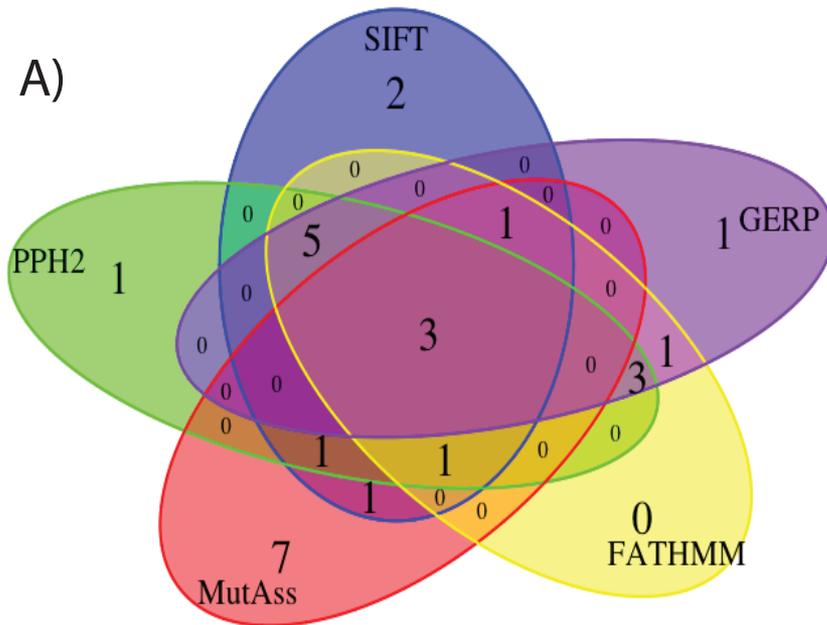
- ClinVar
 - Basée sur les données observées en clinique
- FunSeq2
 - Environnement de données (base de données)
 - Annotation des variants selon des caractéristiques
 - Donne une liste d'informations sur chaque variant

Vérification des outils sur le panel Côlon-Poumon (1)



Proportion des outils de conservation sur mutation faux sens

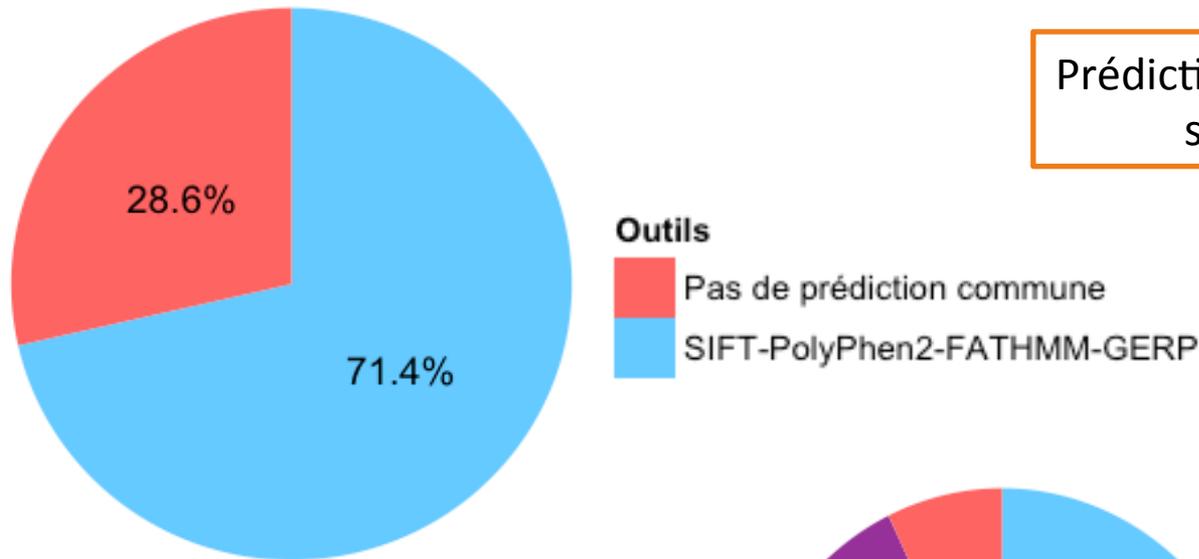
Vérification des outils sur le panel Côlon-Poumon (2)



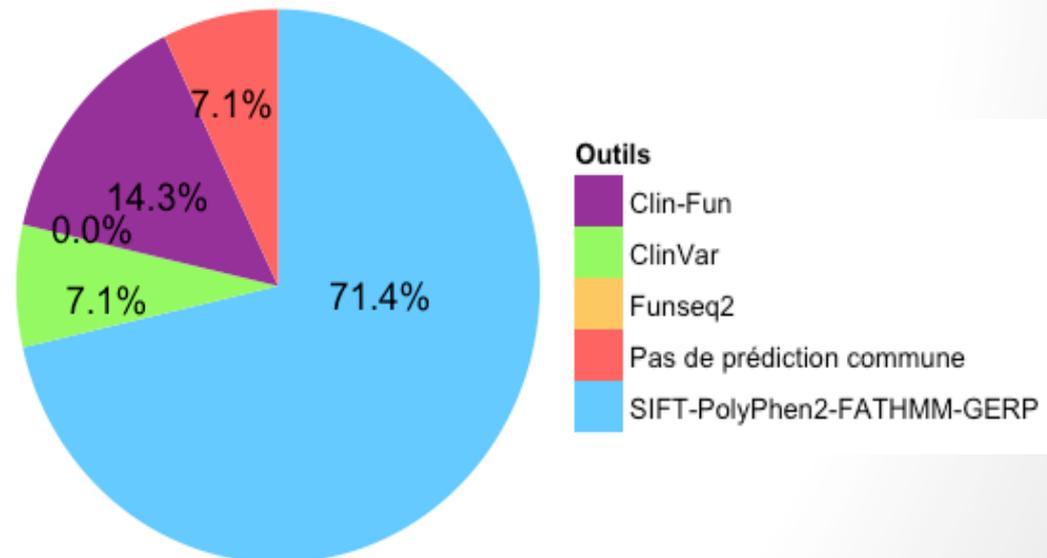
Comparaison des outils de conservation sur mutation faux sens (14 variants)
Représentation en diagramme de Venn

Vérification des outils sur le panel Côlon-Poumon (3)

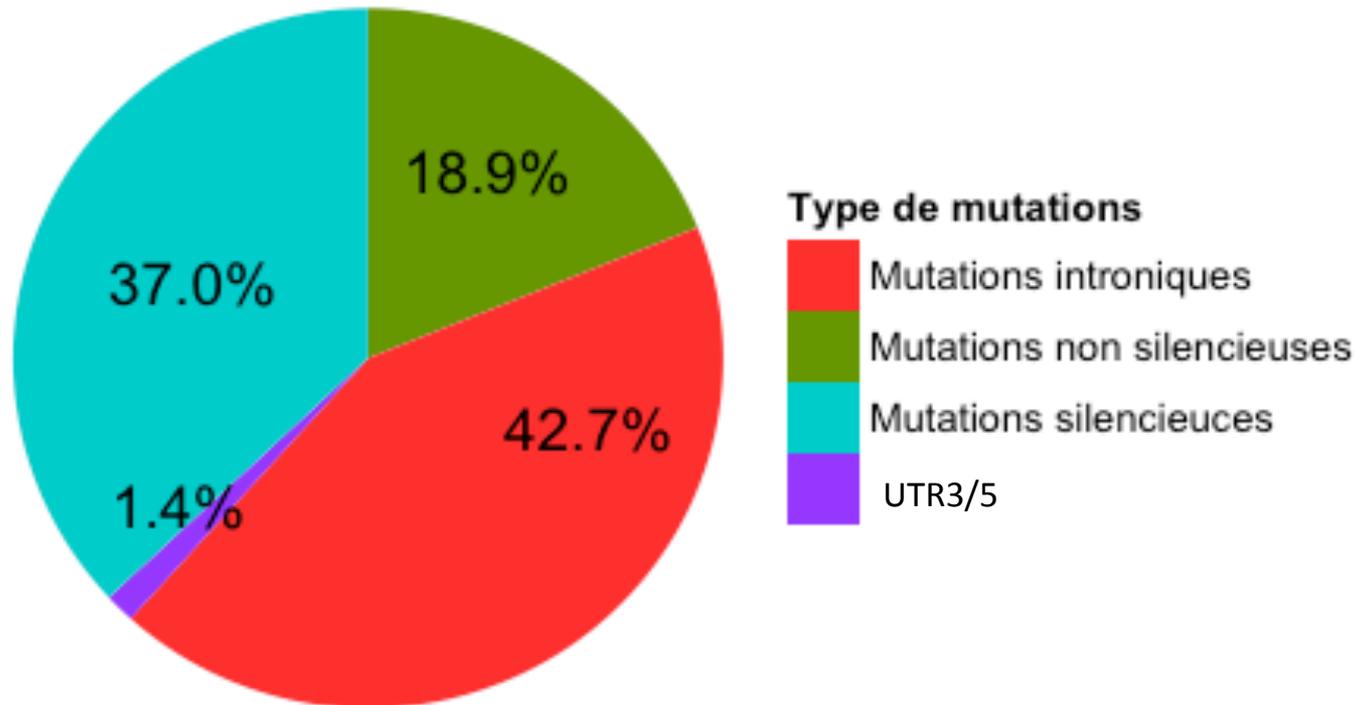
Prédiction des outils de conservation
sur mutations faux sens



Prédiction de tous les outils
sur mutations faux sens

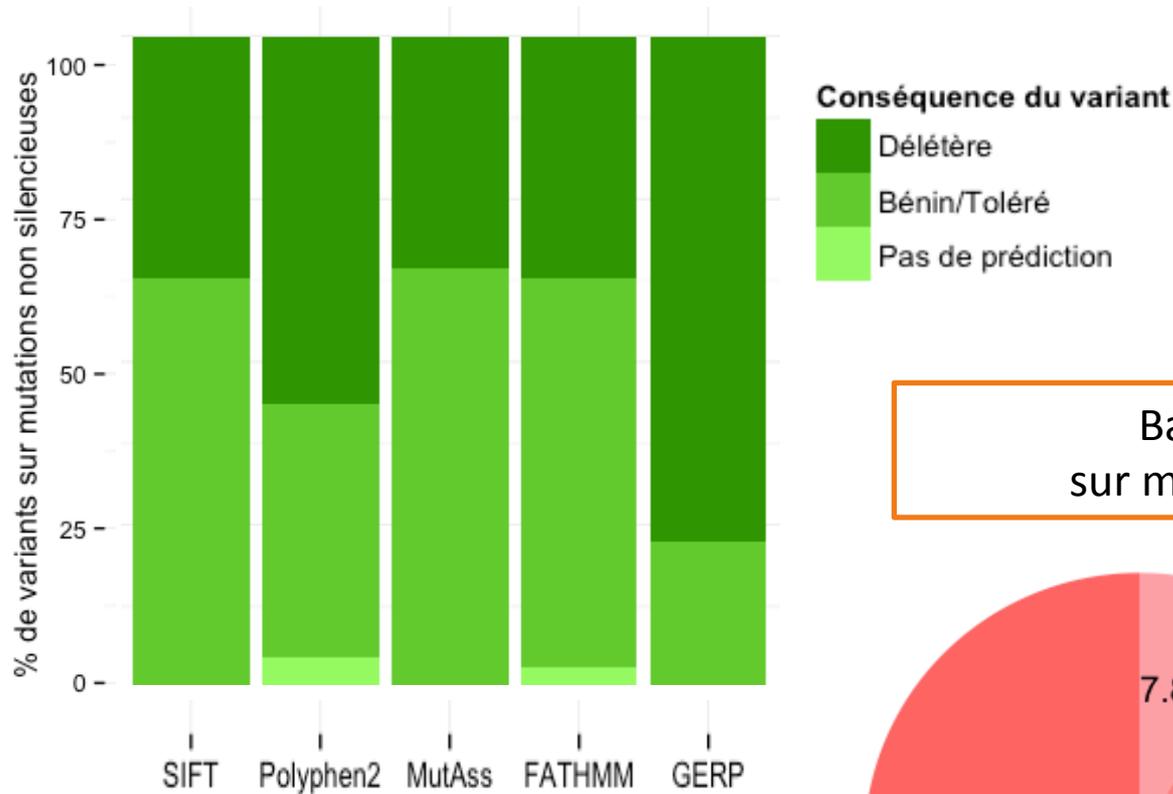


Test des outils sur le panel SAFIR02 (1)



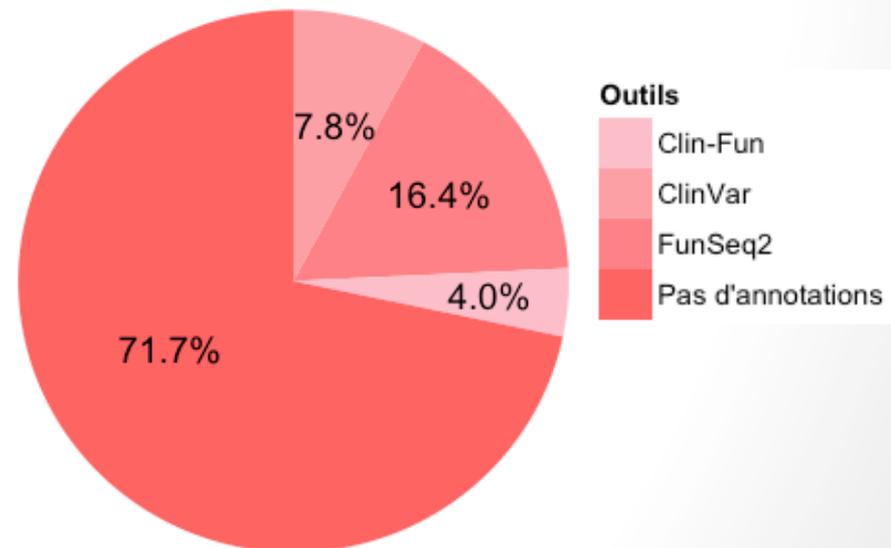
Proportion du type de mutations
dans les échantillons SAFIR02

Test des outils sur le panel SAFIR02 (2)

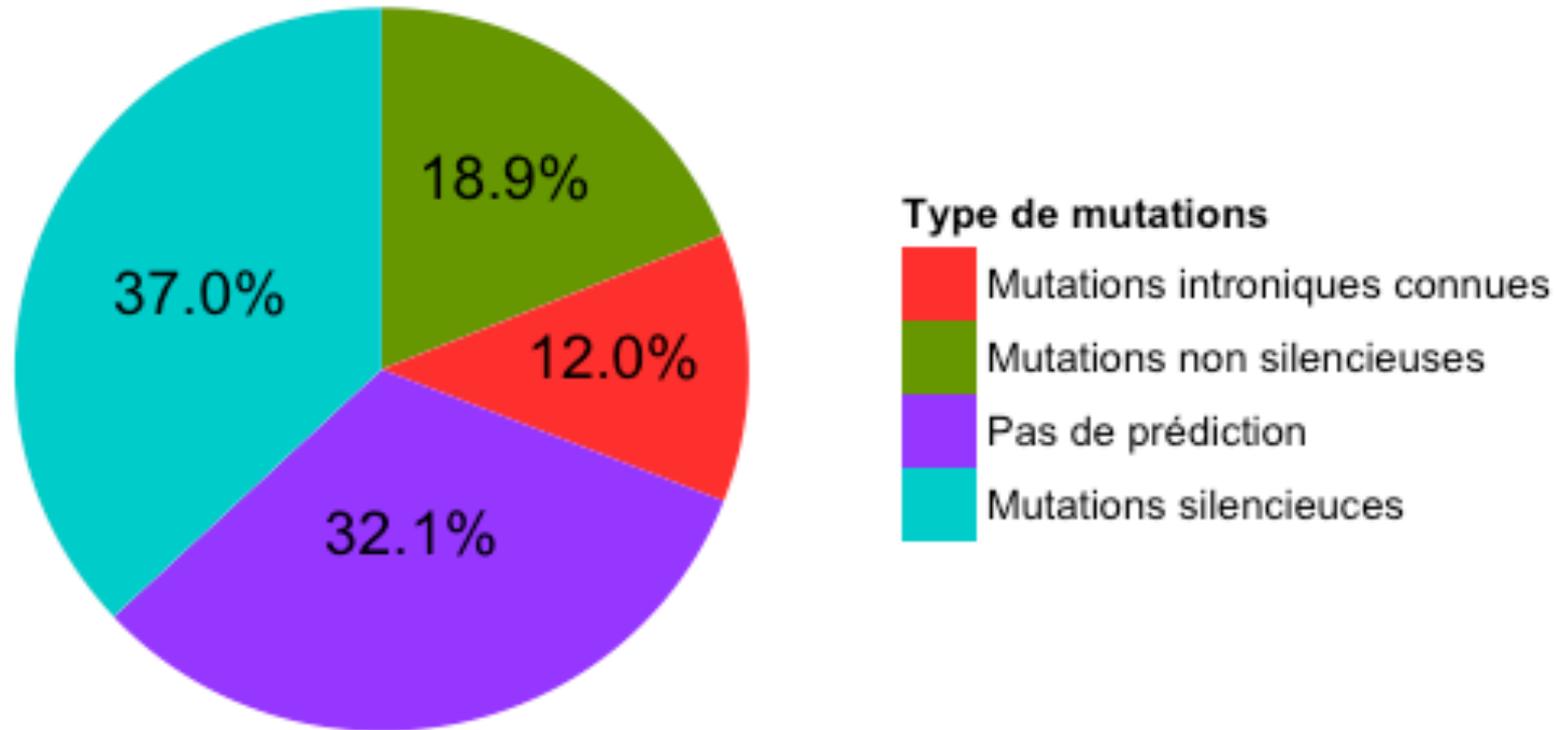


Base de données
sur mutation intronique

Outils de conservation
sur mutations non silencieuses



Test des outils sur le panel SAFIR02 (3)



Proportion de l'information concernant les types de mutations dans les échantillons SAFIR02

Conclusion - Discussion

- MutationAssessor et GERP sont en décalage dans les prédictions
- FATHMM semble plus précis via sa méthode de calcul
- Outils de calcul de score ne suffisent pas
- Nécessite d'avoir des données cliniques

Perspective

=> Implémentation :

- D'autres outils comme CADD et iCAGES
- Des bases de données sur la régulation (RegulomeDB)
- Des outils d'annotations sur les insertions et délétions



Vérifier les mises à jour

Sélectionner les variants les plus pertinents

→ un patient, une tumeur, un traitement

Plan

- Introduction
 - Le cancer
 - Les cancers du côlon et du poumon
 - Les axes du laboratoire
 - Analyse par NGS
- Résultats
 - Analyse sur panel Côlon-Poumon
 - Analyse sur panel SAFIR02
- Conclusion

Le cancer du côlon et du poumon

Cancer colorectal

- Plus de 43 000 nouveaux cas
- Près de 18 000 décès
- Développement de projet de dépistage : Diminution de la mortalité.

Cancer du poumon

- Plus de 45 000 nouveaux cas
- Plus de 30 000 décès
- Le plus meurtrier dans le monde (19,4% des décès par cancer)
- Le tabac est le principal facteur de risque

=> Déterminer les cibles les plus pertinentes