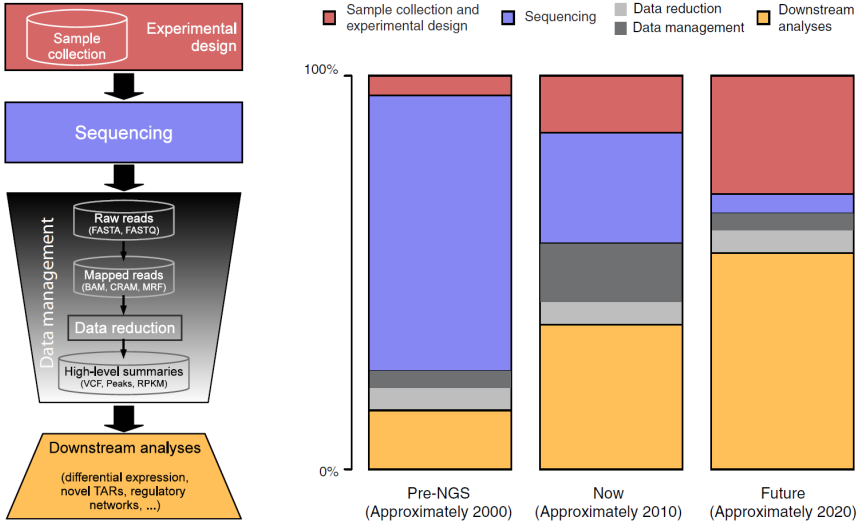
Priorisation de variants

Antony Le Béchec

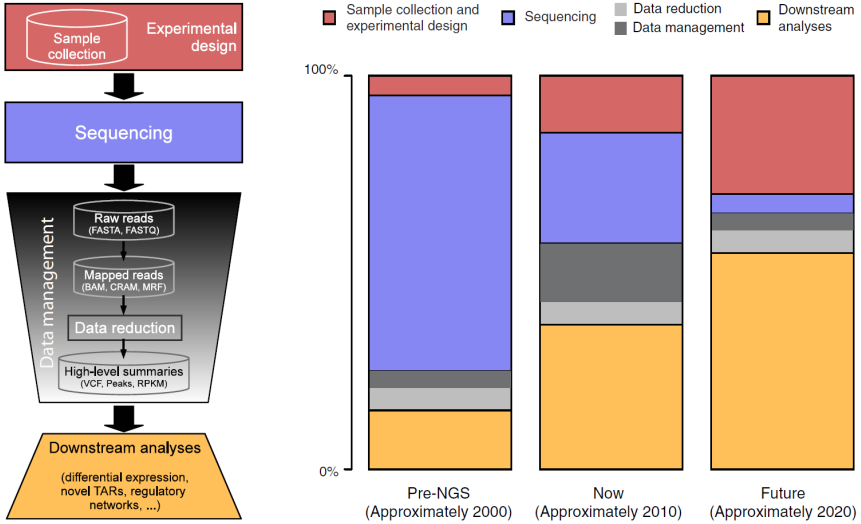
15/05/2024

Le futur d’avant…



Sboner et al. (2011)

Le futur d'avant… jusqu'aux défis d'aujourd'hui



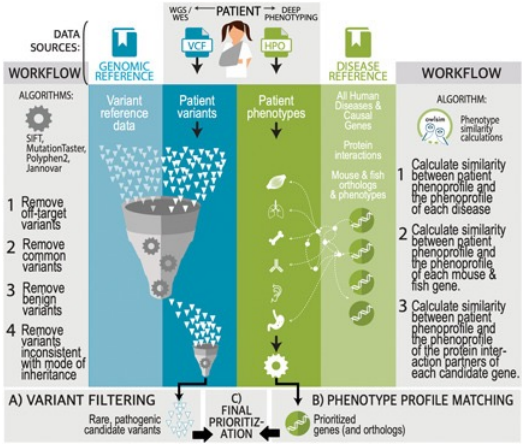
Sboner et al. (2011)

- Volume élevé de **données** de séquençage (exome/génome)
- Volume élevé d'**informations** et de **connaissances** (databases énormes)
- Intégration de données **cliniques** (phénotypes)
- Variabilité entre **individus** (médecine personnalisée)
- Interprétation des **variants non classés** (zone grise)
- Manque de consensus sur les **critères de priorisation**, les **méthodes** et les **sources**

Méthodes de priorisation

Best practices for the interpretation and reporting of clinical whole genome sequencing

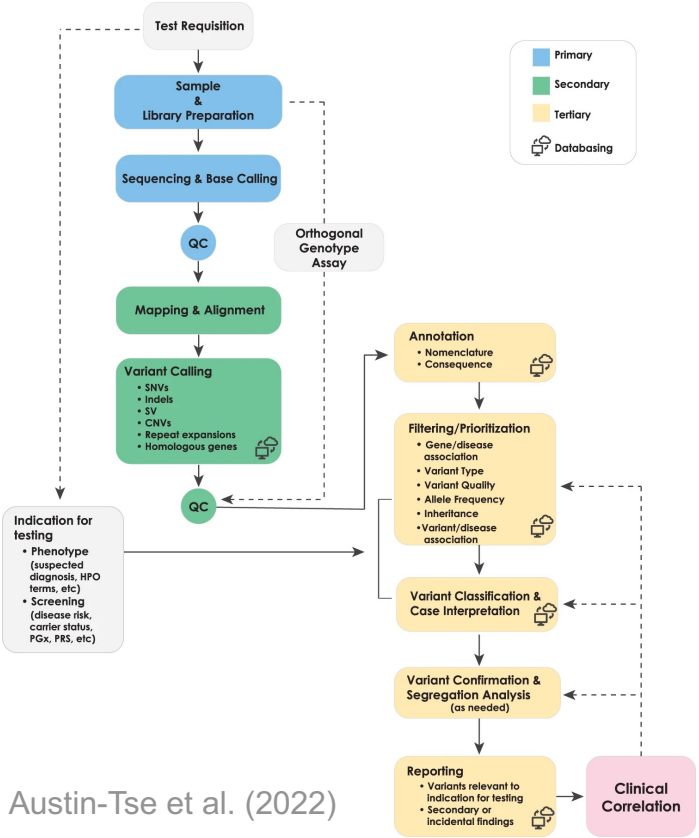
Phenotype-driven approaches to enhance variant prioritization and diagnosis of rare disease



Jacobsen et al. (2022)

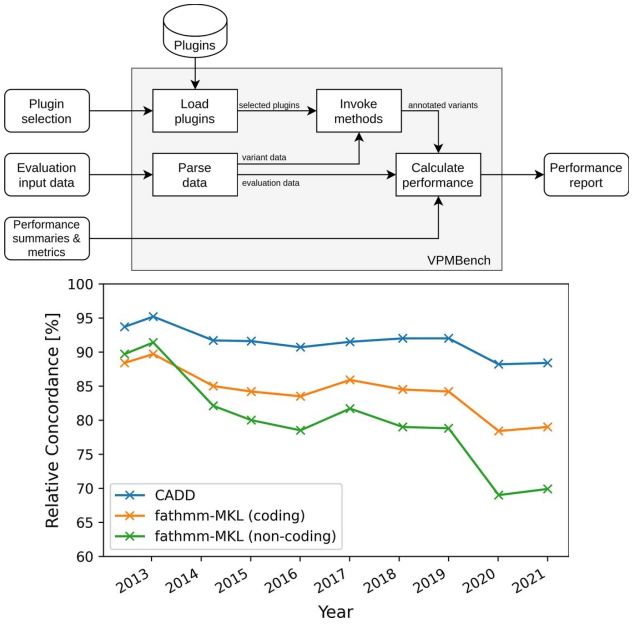
ACMG: American College of Medical Genetics and Genomics

ESHG (ABC System) : variant classification of any type of genetic variant

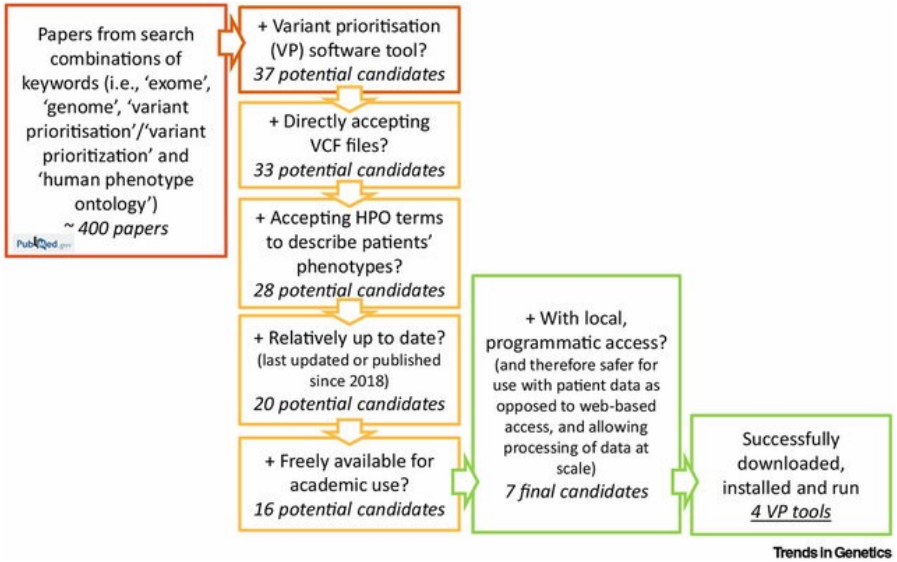


Austin-Tse et al. (2022)

Méthodes de priorisation - benchmark



Ruscheinski et al. (2021)
VPMBench: a test bench for variant prioritization methods



Kelly et al. (2022)
Phenotype-aware prioritisation of rare Mendelian disease variants

Outils de priorisation

Software	Low-throughput web access	High-throughput programmatic access	GRCh38 analysis	Family-based analysis
Exomiser framework (Smedley et al., 2015) including PhenIX (Zemojtel et al., 2014) and Genomiser (Smedley et al., 2016)	Yes	Yes	Yes	Yes
AMELIE (Birgmeier et al., 2020)	Yes	Yes	No	Yes
AnnotSV (Geoffroy et al., 2021)	Yes	Yes	Yes	No
SvAnna (Danis et al., 2021a)	No	Yes	Yes	No
LIRICAL (Robinson et al., 2020)	No	Yes	Yes	No
xRare (Q. Li, Zhao, et al., 2019)	No	Yes	No	No
VARPP (Anderson et al., 2019)	No	Yes	No	No
DeepPVP (Boudellioua et al., 2019)	No	Yes	No	No
MutationDistiller (Hombach et al., 2019)	Yes	No	No	No
GenIO (Koile et al., 2018)	Yes	No	No	No
wAnnovar (H. Yang & Wang, 2015)	Yes	No	Yes	No
QueryOR (Bertoldi et al., 2017)	Yes	No	No	Yes
BierApp (Alemán et al., 2014)	Yes	No	No	Yes
OVA (Antanaviciute et al., 2015)	Yes	No	No	No

Jacobsen et al. (2022)

Phenotype-driven approaches to enhance variant prioritization and diagnosis of rare disease

Selection of phenotype-aware variant prioritisation (VP) software tools based on five suitability criteria [a]

VP software tool	Directly accepting VCF files	Accepting HPO terms	Last updated or published since 2018	Freely available	Local, programmatic access	Refs
Exomiser[b]	√	√	√	√	√	[8]
LIRICAL[b]	√	√	√	√	√	[35]
VARPP[b]	√	√	√	√	√	[11]
Xrare[b]	√	√	√	√	√	[30]
Phenoxome[b]	√	√	√	√	√	[42]
DeepPVP[b]	√	√	√	√	√	[16]
PhenIX[b]	√	√	√	√	√	[44]
VINYL		√	√	√	√	[18]
eDiva		√	√	√	√	[15]
VarSight		√	√	√	√	[20]
AMELIE	√	√	√	√	No local installation	[14]
GeneTerpret	√	√	√	√	Web-based only	[32]
PhenoPro	√	√	√	√	Web-based only	[31]
MutationDistiller	√	√	√	√	Web-based only	[22]
GenIO	√	√	√	√	Web-based only	[28]
PhenoVar	√	√	√	√	Web-based only	[40]
GEM	√	√	√	Commercial	Web-based only	[3]
EVIDENCE	√	√	√	Commercial	Web-based only	[36]
VarElect	√	√	√	Commercial	Web-based only	[39]
Phevor (now Phevor2)		√	√	√	Web-based only	[38]
Moon	√	√	√	Commercial	Code available on request	[34]
wAnnovar	√	√		√	Web-based only	[43]
OVA	√	√		√	Web-based only	[12]
BierApp	√	√		√	Web-based only	[10]

Kelly et al. (2022)

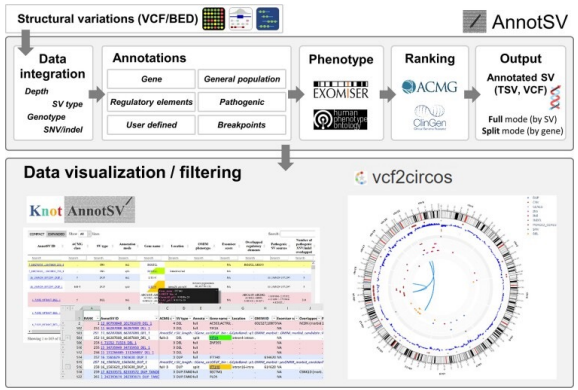
Phenotype-aware prioritisation of rare Mendelian disease variants

Outils de priorisation - le choix

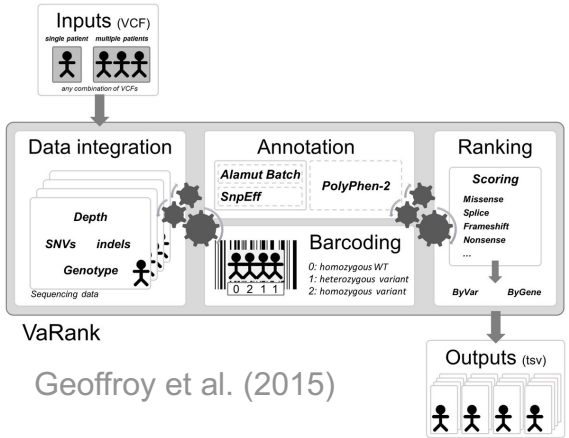
Software
Exomiser framework (Smedley et al., 2015) including PhenIX (Zemojtel et al., 2014) and Genomiser (Smedley et al., 2016)
AMELIE (Birgmeier et al., 2020)
AnnotSV (Geoffroy et al., 2021)
SvAnna (Danis et al., 2021a)
LIRICAL (Robinson et al., 2020)
xRare (Q. Li, Zhao, et al., 2019)
VARPP (Anderson et al., 2019)
DeepPVP (Boudellioua et al., 2019)
MutationDistiller (Hombach et al., 2019)
GenIO (Koile et al., 2018)
wAnnovar (H. Yang & Wang, 2015)
QueryOR (Bertoldi et al., 2017)
BierApp (Alemán et al., 2014)
OVA (Antanaviciute et al., 2015)

Jacobsen et al. (2022)

Baud et al. (2021)



Geoffroy et al. (2023)



Geoffroy et al. (2015)

Sources d’annotations des variants

- **Multiplicité** des sources de données : Gérer et intégrer plusieurs bases de données (avec des critères différents).
- **Redondance** dans les bases de données : Choix des sources données pour éviter les sur-représentations (e.g. agrégateurs CADD, REVEL).
- **Contexte** de l'annotation : Sélectionner le bon transcrit pour une annotation précise.
- **Evolution** des bases de données : Suivre les mises à jour, évaluer la qualité et la pertinence des nouvelles bases de données (e.g. Prédictions).
- **Complexité** et taille des bases de données : Gérer les ressources informatiques pour traiter les données volumineuses (e.g. gnomAD, dbNSFP).

Contexte / Evidence / Sens

Critères de priorisation

```
## Exomiser Job Template.
# The job is split into three sections:
#    sample: describes the proband
#    analysis: details the steps exomiser will take to analyse the sample
#    outputOptions: specifies what files to output and where with filtering options for the
#
# These can be input separately on the command line e.g. --sample sample.yml --analysis anal
# In this example the 'sample' is replaced with a 'phenopacket' see https://phenopacket-sche
---
phenopacket:
  id: manuel
  subject:
    id: manuel
    sex: MALE
  phenotypicFeatures:
    - type:
        id: HP:0001159
        label: Syndactyly
    - type:
        id: HP:0000486
        label: Strabismus
    - type:
        id: HP:0000327
        label: Hypoplasia of the maxilla
    - type:
        id: HP:0000520
        label: Proptosis
    - type:
        id: HP:0000316
        label: Hypertelorism
    - type:
        id: HP:0000244
        label: Brachyturricephaly
  htsFiles:
    - uri: examples/Pfeiffer.vcf
      htsFormat: VCF
      genomeAssembly: hg19
  metaData:
    created: '2019-11-12T13:47:51.948Z'
    createdBy: julesj
    resources:
      - id: hp
        name: human phenotype ontology
        url: http://purl.obolibrary.org/obo/hp.owl
        version: hp/releases/2019-11-08
        namespacePrefix: HP
        iriPrefix: 'http://purl.obolibrary.org/obo/HP_'
    phenopacketSchemaVersion: 1.0
analysis:
  #FULL or PASS_ONLY
  analysisMode: PASS_ONLY
  # In cases where you do not want any cut-offs applied an empty map should be used e.g. inh
  # These are the default settings, with values representing the maximum minor allele freque
  # allele to be considered as a causative candidate under that mode of inheritance.
  # If you just want to analyse a sample under a single inheritance mode, delete/comment-out
  # or X_RECESSIVE ensure *both* relevant HOM_ALT and COMP_HET modes are present.
  inheritanceModes: {
    AUTOSOMAL_DOMINANT: 0.1,
    AUTOSOMAL_RECESSIVE_COMP_HET: 2.0,
    AUTOSOMAL_RECESSIVE_HOM_ALT: 0.1,
    X_DOMINANT: 0.1,
    X_RECESSIVE_COMP_HET: 2.0,
    X_RECESSIVE_HOM_ALT: 0.1,
    MITOCHONDRIAL: 0.2
  }
```

```
  # Possible frequencySources:
  # UK10K - http://www.uk10k.org/ (UK10K)
  # gnomAD - http://gnomad.broadinstitute.org/ (GNOMAD_E, GNOMAD_G)
  # note that as of gnomAD v2.1 1000 genomes, ExAC are part of gnomAD
  # as of gnomAD v4 TOPMed & ESP are also included in gnomAD
  frequencySources: [
    UK10K,

    GNOMAD_E_AFR,
    GNOMAD_E_AMR,
    #    GNOMAD_E_ASJ,
    GNOMAD_E_EAS,
    #    GNOMAD_E_FIN,
    GNOMAD_E_NFE,
    #    GNOMAD_E_OTH,
    GNOMAD_E_SAS,

    GNOMAD_G_AFR,
    GNOMAD_G_AMR,
    #    GNOMAD_G_ASJ,
    GNOMAD_G_EAS,
    #    GNOMAD_G_FIN,
    GNOMAD_G_NFE,
    #    GNOMAD_G_OTH,
    GNOMAD_G_SAS
  ]
  # Possible pathogenicitySources: (POLYPHEN, MUTATION_TASTER, SIFT), (REVEL, MVP), CADD, RE
  # REMM is trained on non-coding regulatory regions
  # *WARNING* if you enable CADD or REMM ensure that you have downloaded and installed the C
  # and updated their location in the application.properties. Exomiser will not run without
  pathogenicitySources: [ REVEL, MVP ]
  #this is the standard exomiser order.
  #all steps are optional
  steps: [
    #hiPhivePrioritiser: {},
    #priorityScoreFilter: {priorityType: HIPHIVE_PRIORITY, minPriorityScore: 0.500},
    #intervalFilter: {interval: 'chr10:123256200-123256300'},
    # or for multiple intervals:
    #intervalFilter: {intervals: ['chr10:123256200-123256300', 'chr10:123256290-123256358']}
    # or using a BED file - NOTE this should be 0-based, Exomiser otherwise uses 1-based coo
    #intervalFilter: {bed: /full/path/to/bed_file.bed},
    #genePanelFilter: {geneSymbols: ['FGFR1','FGFR2']},
    # geneBlacklistFilter: { },
      failedVariantFilter: { },
    #qualityFilter: {minQuality: 50.0},
      variantEffectFilter: {
        remove: [
            FIVE_PRIME_UTR_EXON_VARIANT,
            FIVE_PRIME_UTR_INTRON_VARIANT,
            THREE_PRIME_UTR_EXON_VARIANT,
            THREE_PRIME_UTR_INTRON_VARIANT,
            NON_CODING_TRANSCRIPT_EXON_VARIANT,
            UPSTREAM_GENE_VARIANT,
            INTERGENIC_VARIANT,
        REGULATORY_REGION_VARIANT,
        CODING_TRANSCRIPT_INTRON_VARIANT,
        NON_CODING_TRANSCRIPT_INTRON_VARIANT,
        DOWNSTREAM_GENE_VARIANT
        ]
      },
    # removes variants represented in the database
    #knownVariantFilter: {},
    frequencyFilter: {maxFrequency: 2.0},
    pathogenicityFilter: {keepNonPathogenic: true},
    # inheritanceFilter and omimPrioritiser should always run AFTER all other filters have c
    # they will analyse genes according to the specified modeOfInheritance above- UNDEFINED
```

Beaucoup de sources

Beaucoup de formats

Beaucoup de critères

Beaucoup de paramètres…

Robinson et al. (2014)

Improved exome prioritization of disease genes through cross-species phenotype comparison

Valeurs de priorisation

- **Filtre** (boolean) : Méthode Yes/No

- **Score** (integer) : Valeur numérique possiblement normalisée [0-1]

- **Tags** (enum) : Deleterious/Tolerant/Neutral

- **Classement** (liste ordonnée) : ACMG

Valeurs de priorisation et leur modalités de calcul

- **Filtre** (boolean) : Méthode Yes/No
- **Score** (integer) : Valeur numérique possiblement normalisée [0-1]
- **Tags** (enum) : Deleterious/Tolerant/Neutral
- **Classement** (liste ordonnée) : ACMG

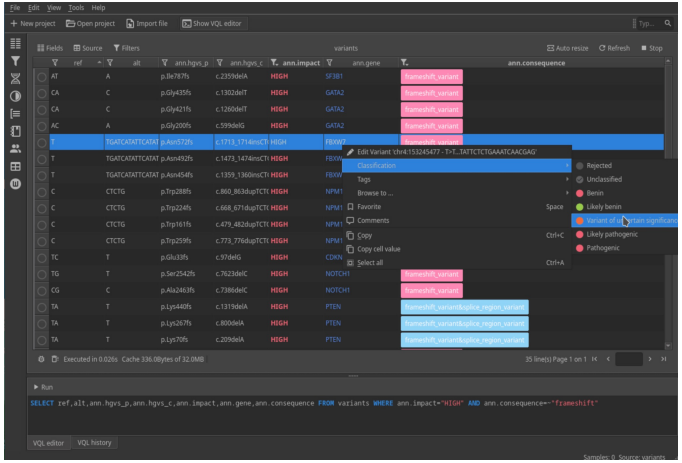
- **A la main**: définition des critères
- **Algorithms**: définition de paramètres
- **Machine learning**: définition du contexte
- **IA**: définition de rien du tout

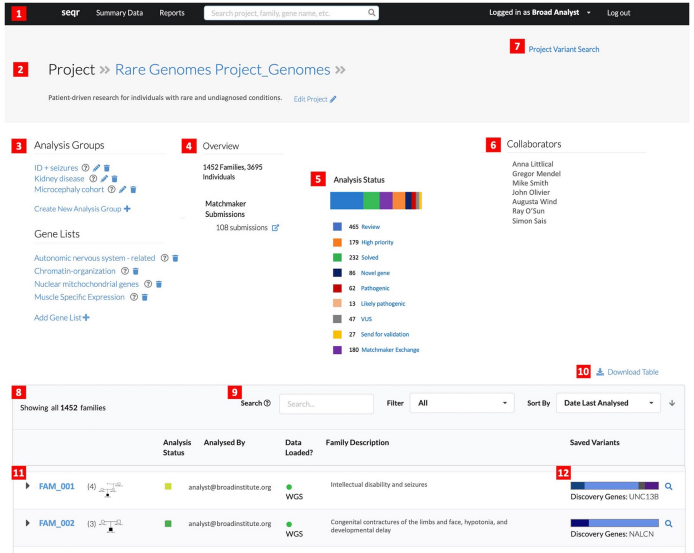
Exploration et validation

GLEAVES

DIAGHO

PORTAIL D'INTERPRÉTATION GÉNOMIQUE DU GRAND OUEST





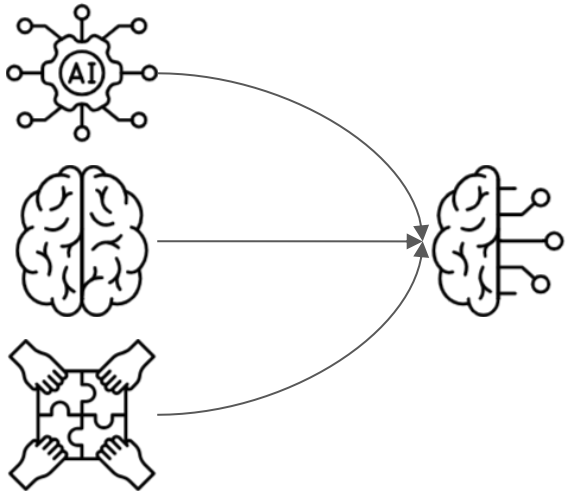
Intelligence ?

Intelligences ABC

AI Intelligence Artificielle

BI Intelligence Bio/BioInfo

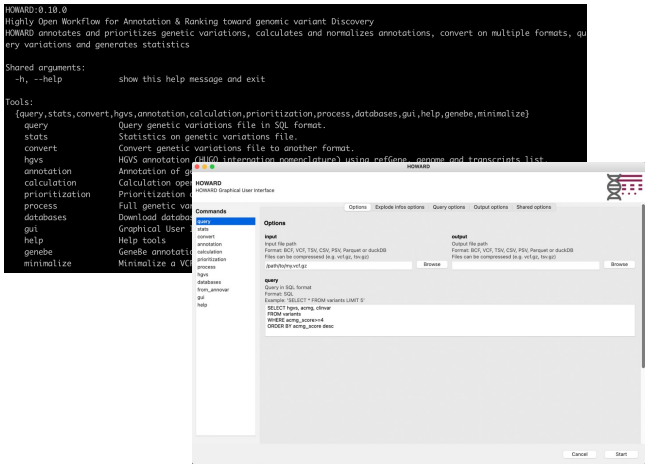
CI Intelligence Collective

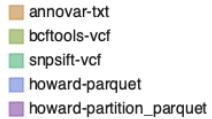


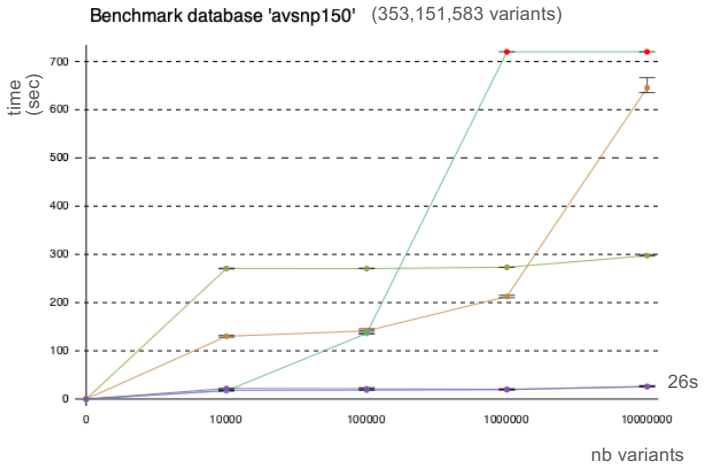
HOWARD

Highly Open Workflow for Annotation & Ranking toward genomic variant Discovery

Workflow complet : Annotation / Calculation / Prioritization / Classification / Filtration / Conversion





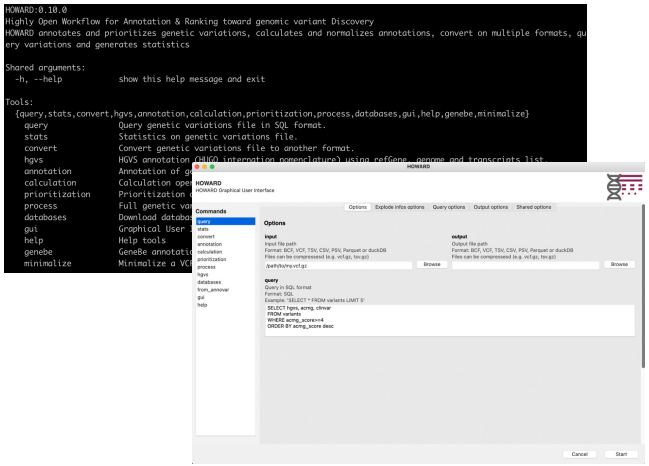


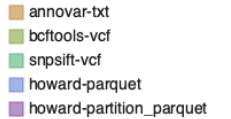
bioinfo-chru-strasbourg/howard

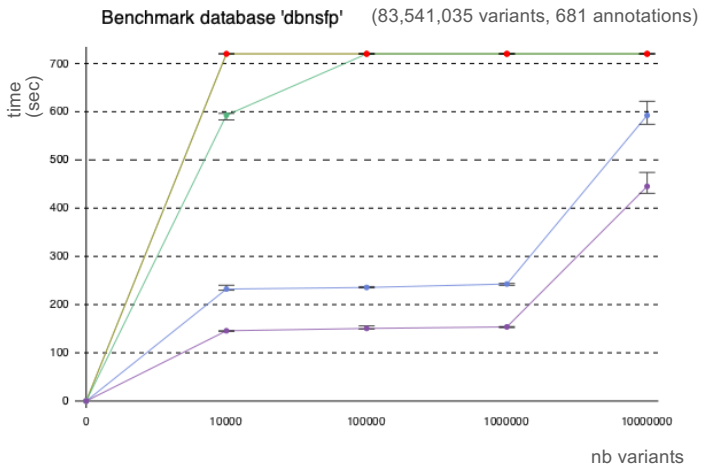
HOWARD

Highly Open Workflow for Annotation & Ranking toward genomic variant Discovery

Workflow complet : Annotation / Calculation / Prioritization / Classification / Filtration / Conversion







bioinfo-chru-strasbourg/howard

Merci !!!

Hôpitaux Universitaires de Strasbourg

Bioinformatique médicale appliquée au diagnostic

bioinfo@chru-strasbourg.fr

- Jean Muller
- Thomas Lavaux
- Samuel Nicaise
- Mateusz Rauch
- Jean-Baptiste Lamouche