

DIAGHO

PORTAIL D'INTERPRÉTATION GÉNOMIQUE DU GRAND OUEST

**Un outil académique
d'aide à l'interprétation**

Marie de Tairac
CHU Rennes



02/04/2025

1.

Contexte projet

LES ORIGINES

Un projet du Grand-Ouest à
visée collaborative

Projet stratégique du **GCS HUGO**



Un outil fédérateur catalyseur de savoir-faire

Une visée industrielle

Un descendant REGOVAR

ASPECTS OPÉRATIONNELS

Une nouvelle équipe qui
avance vite

Un début complexe

Un développement consolidé depuis 2023

Équipe stabilisée au sein du CHU de Rennes

Un chef de projet et quatre développeurs (2 back, 2 front)

Nouveaux enjeux IVDR

2.

Philosophie

AUTONOMIE DU BIOLOGISTE

Offrir la possibilité de travailler librement et permettre aux bioinfos de se focaliser sur leurs expertises

Facilité l'accès

Charger ou recharger les données et rester libre de ses recherches

Liberté de traitement

Accès à la totalité des données sans pré-filtrage

Maîtrise de l'analyse

Autonomie sur la mise en œuvre des filtres

Personnalisation

Interface, filtres favoris, formats d'exports...

Automatisation

Systematiser les actions routinières et se focaliser sur l'interprétation

MAÎTRISE DU LABORATOIRE

S'adapter aux pratiques des laboratoires et non l'inverse

Agnostique

Annotation libre pour conserver les pipelines en place

Compagnon

Facilite l'organisation des dossiers

Seamless

Du pipeline à l'interprétation sans action manuelle

Configurable

Choix de configurations pour homogénéiser les pratiques

Accessible

Pas de boîte noire et Python pour une appropriation simplifiée

Sécurisé

Contrôle des accès aux projets par équipes et/ou praticien

PORTAIL UNIQUE

Centraliser pour gagner en
efficacité !

Une seule interface

Ne plus alterner entre UI
ou fichiers selon les
données

Tous types d'analyses

Singleton, trios, familles,
non apparentés,
cohortes...

Toutes sources de data

DNA-seq short et long
reads, CNV, SV. Bientôt
RNA-Seq, STR

Approches combinées

Facilite le croisement des
données entre elles

Évolutif et généralisable

Laisse ouvert l'intégration
des technologies de
demain

La clinique à portée de clic

Organise les données
patients autour des
données génomiques

AUTOMATISER, POTENTIALISER

Permettre aux biologistes et bioinformaticiens de se focaliser sur leurs expertises pour consacrer plus de temps à l'essentiel

Chargement

Interprétations disponibles sans action humaine

Ré-annotation

Possible librement si non gérée dans le pipeline

Pré-filtrages et workflows

Optimiser l'organisation et le temps de travail

Liens IGV

Visualisez vos BAM en un clic

Supports de RCP

Exports prêts à l'emploi en fin d'interprétation et personnalisable par labo

Interopérabilité

Des connecteurs pour une organisation de bout en bout

DÉPLOIEMENT EFFICACE

Simplifier l'installation pour un utilisateur seul comme pour plusieurs équipes

Conteneurisée

Déployable en un simple
docker compose

Du laptop au cloud

Ordinateur seul, VM
interne à l'organisation,
installation multi-labo

Portail web

Pas d'installation
nécessaire sur le poste de
l'utilisateur

Léger

Déjà testé sur machine de
4CPU + 16Go RAM

OUVERT ET COLLABORATIF

Être fondamentalement tourné vers le partage et la collaboration

Académique

Initié par le CHU de Rennes

Inter-régional

Porté par un GCS inter-régional HUGO

Code ouvert

Codebase prévue en libre accès

Langage commun

Backend en Python, commun chez les bioinfos

Licence à acter

Libre dans CHU, CLCC et universités si auto-hébergé

Modèle économique à faire

Assurer la pérennité des développements et de la maintenance

3.

Technique

TECH STACK

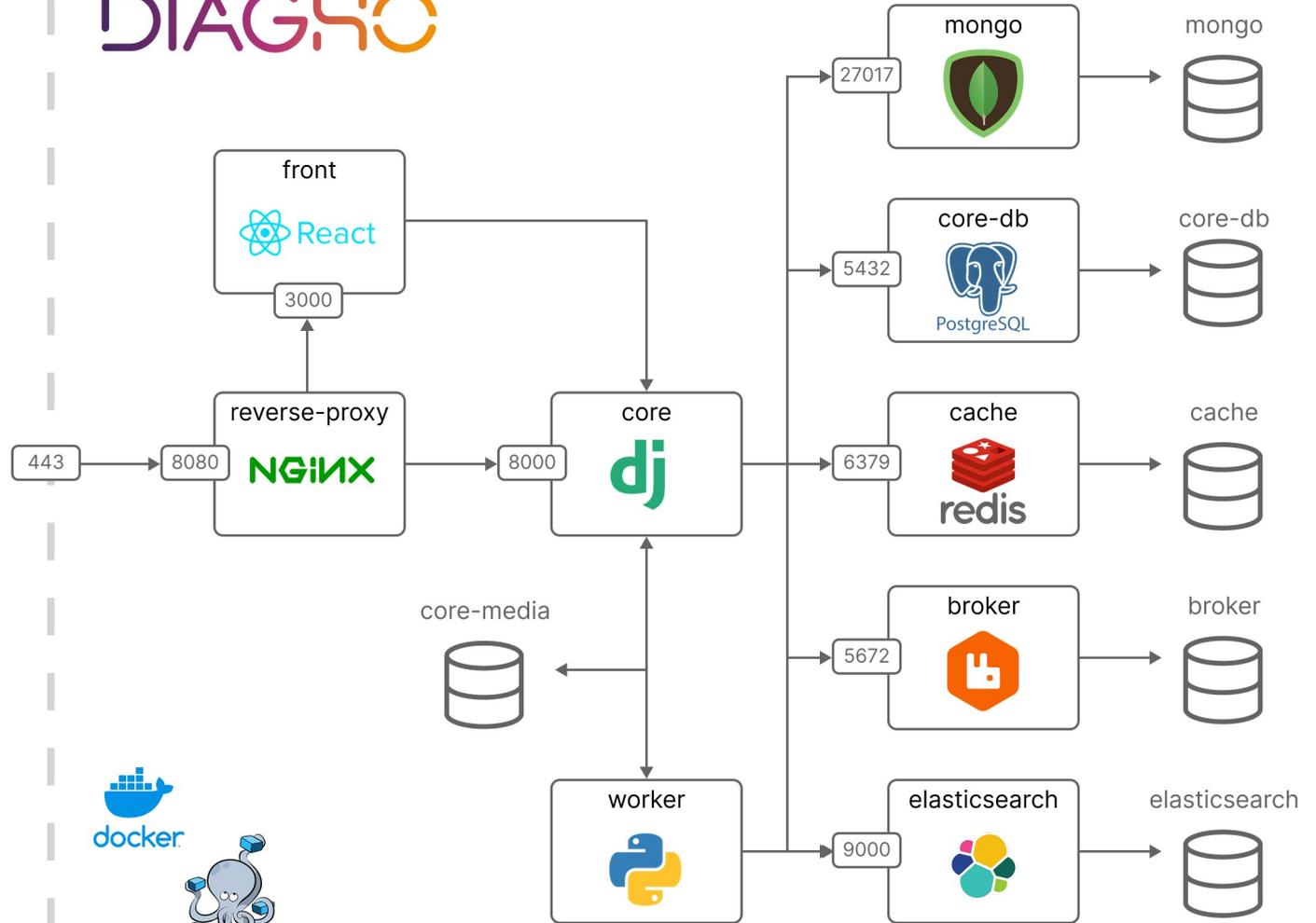
- Conteneurisation **Docker**
- **Python** et ORM **Django**
- Front **React**, SSR **Remix**, CSS **Tailwind**
- BDD **Postgres**
- BDD **MongoDB** pour les variations
- Message broker : **RabbitMQ**
- Task queue : **Celery**
- Caching : **Redis**
- Indexation des dossiers : **ElasticSearch**



DOCKER COMPOSE



DIAGHO



4.

Fonctionnalités

ORGANISATION LIBRE PAR PROJET

Des projets et sous-projets pour gérer les accès et configurations à appliquer par défaut selon les indications considérées



INTERPRÉTATIONS EN FAMILLE OU COHORTES

Interpréter les variants sur un groupe de sujets simultanément

Filter by Tags Clinvar 4/5 Go To Column								
	SYMBOL	TAGS	INDEX GT	INDEX AB	MOTHER GT	MOTHER AB	FATHER GT	SYMBOL
>G	LIPT1	Clinvar 4/5	0/1	0.485	0/1	0.5	0/0	GNRH2
C	ATG7	Clinvar 4/5	0/1	0.381	0/0	0	0/1	MSH6
>A	ACAD11	Clinvar 4/5	0/1	0.3	0/0	0	0/1	ALMS1
CTT>C	C7	Clinvar 4/5	0/1	0.462	0/0	0	0/1	UGT1A6
>T	COL13A1	Clinvar 4/5	0/1	0.619	0/1	0.333	0/0	SMN1
>CA	ANO5	Clinvar 4/5	0/1	0.655	0/1	0.359	0/0	CFTR
>T	LOXHD1	Clinvar 4/5 High Spli	0/1	0.432	0/1	0.486	0/0	PRKDC
>GC	ARSA	Clinvar 4/5	0/1	0.455	0/1	0.614	0/0	PRKD

Filter by Tags Clinvar Go To Column								
	SAMPLE_NAME	GT	AB	DP	CONSEQUENCE	CLINVAR_CLNSIG		
	L8_LL_B3-3_400	0/1	0.556	9	downstream_gene_variant	Conflicting_cl		
	L8_LL_A3-0_400	0/1	0.333	9	missense_variant	Conflicting_cl		
	L8_LL_A3-0_400	0/1	0.5	4	frameshift_variant	Conflicting_cl		
	L8_LL_A3-0_400	0/1	0.714	7	intron_variant	Conflicting_cl		
	L8_LL_A3-0_400	0/1	0.333	15	downstream_gene_variant	Conflicting_cl		
	L8_LL_A5-1_400	0/1	0.8	10	missense_variant	Conflicting_cl		
	L8_LL_A6-2_400	0/1	0.3	10	intron_variant	Conflicting_cl		
	L8_LL_A5-1_400	0/1	0.579	19	intron_variant	Conflicting_cl		

FILTRAGE LIBRE

- Possible sur toutes colonnes peu importe l'annotation
- Gestion des types de données
- Filtres rapides s'adaptant aux annotations présentes
- Extensibles aux annotations in house

Filters

Quick Complex

Variations avec note
 Variations d'intérêt
 Variations non pertinentes

^ Consequence

Clinvar Patho/Likely patho

Impact

Related consequences: inframe insertion or deletion, missense and protein altering variant

MODERATE 
 MODIFIER 

Type

Indel
 Missense
 Nonsense
 Splicing
 Synonymous

^ Location

^ Quality

^ Population frequencies

gnomAD AF  < 0,01
 gnomAD homalt  ≤ 2

^ In silico prediction

Submit

Filters

Quick Complex

Reset / Load / Manage

from Autosomique dominant

^ AND  

AB > 0,15
DP ≥ 4

^ OR   

GNOMADG_AF < 0,0001
GNOMADG_AF is empty

^ OR   

IMPACT in HIGH MODERATE
SPLICEAI > 0,2

tag results

Apply Save

FILTRES PARTAGÉS ET VERSIONNÉS

Saved Filters

×

🔍 Search

🌐 Public Filters 📁 My Filters

FAV	NAME	DESCRIPTION	TYPE	OWNER	VERSION	ACTIONS	APPLY TO INTERPRETATION
☆	AD	Autosomiques dominants connus dans OMIM	Consequence, Population frequencies	Administrateur Labo	2	  Make a copy of this filter	✓ Apply
☆	AR	Autosomiques récessifs connus dans OMIM	Consequence, Population frequencies	Administrateur Labo	2	 	✓ Apply
	AR	Autosomiques récessifs connus dans OMIM	sample_ab > 0.85 dp >= 4 impact in HIGH MODERATE omim contains recessive AND gnomadg_af < 0.001 OR gnomadg_af is empty				
☆	Clinvar	Clinvar 4/5 ou déjà rendu comme tel en cas de conflit	-	Administrateur Labo	1	 	✓ Apply
☆	Global PGC v1	Ensemble des filtres retenus pour Periquenomed	-	Administrateur Labo	1	 	✓ Apply
☆	XL	Lié à l'X	Consequence, Population frequencies, Location	Administrateur Labo	2	 	✓ Apply

PRÉ-TAG DES VARIATIONS

Application automatique de filtres
routiniers pour rentrer directement
dans l'interprétation

The screenshot shows a user interface for pre-tagging genetic variants. At the top, it displays 'Variants' with a count of 580975. Below this, a 'Pre-tagging' section lists several filters with corresponding counts and buttons to apply them:

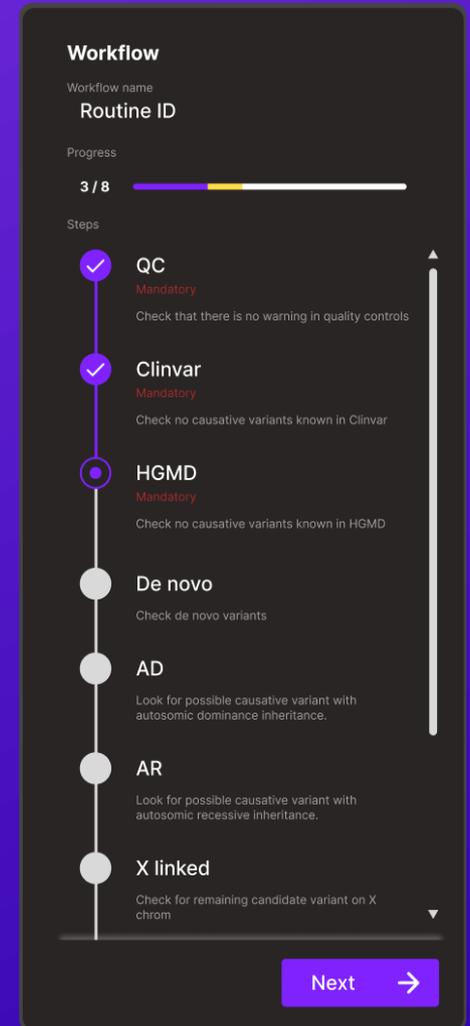
- Autosomique dominant : 19 variants (AD →)
- Autosomique récessif : 7 variants (AR →)
- Liés à l'X : 1 variant (XL →)
- Clinvar 4/5 + conflict patho : 10 variants (Clinvar →)
- Liste 1 : 288733 variants (Liste 1 →)
- Liste 2 : 265800 variants (Liste 2 →)

At the bottom of the interface, there is a prominent blue button labeled 'Go to interpretation →'.

WORKFLOWS

Applicable au niveau projet pour permettre par défaut à chaque interprétation :

- une automatisation des filtrages en routine
- un suivi d'avancement
- une assistance pour les juniors ou pour de nouvelles indications
- un protocole strict si souhaité



RÈGLE FORMATAGE

Configuration possible au niveau des colonnes :

- colorisation pour une lecture plus fluide (exactes et gradients)
- URLs vers sites et bases de données externes

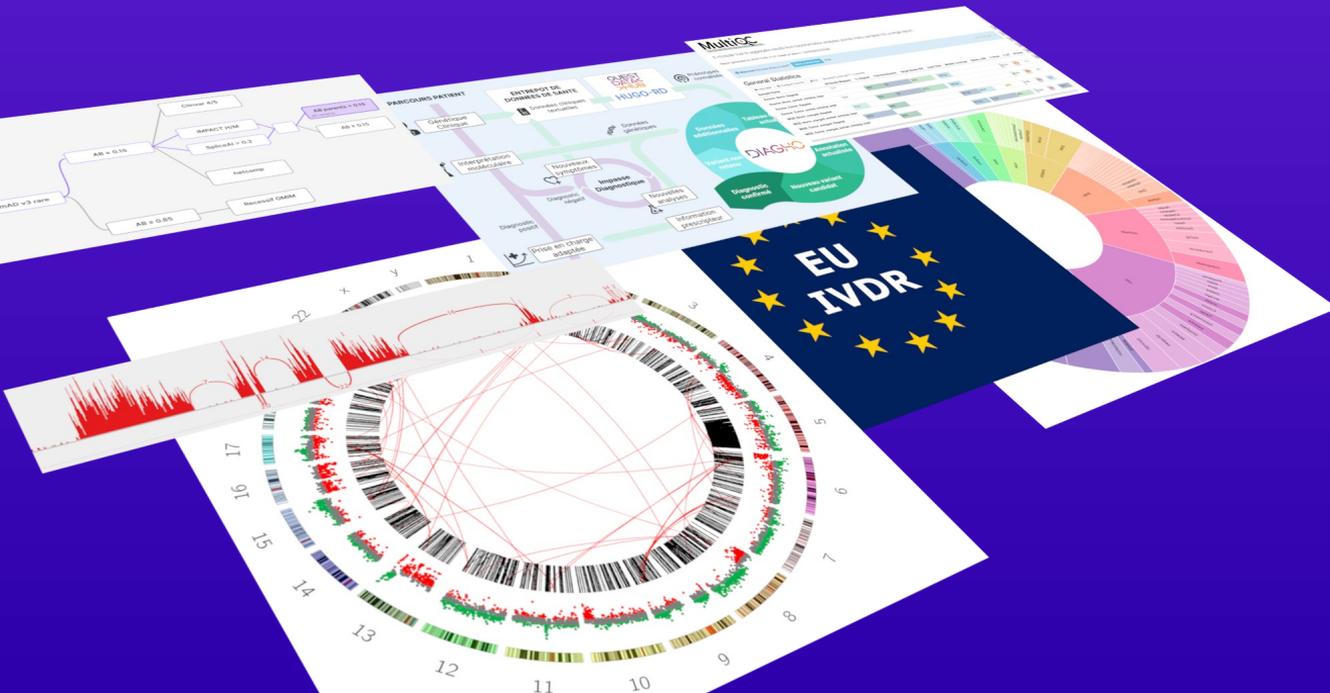
AM_CLASS	DANN_SCORE	FATHMM-MKL_CODING_PRED	FATHMM-XF_CODING_PRED
	0.984	D	N
likely_pathogenic	0.998	D	D
	0.996	D	N
	0.998	D	N

POS.	SYMBOL
9:124862400 A>T	ARPC5L
11:58844433 C>T	ATL2
14:77181894 TG>T	EM63C
15:64355987 C>T	NK1G1
21:36699922 G>A	12

- hgnc
- decipher
- varsome
- hpo
- omim
- pubmed
- panelapp
- GTEx
- LOVD

ET BIEN D'AUTRES CHOSES À VENIR

- Arbres de filtrages
- Requête par run
- Plus de données (RNA, STR...)
- Database locale
- Ploting intégré
- QC intégrées
- Panels versionnés
- Cadrage IVDR
- Ré-analyse auto cas négatifs



2025, ANNÉE DE L'OUVERTURE

- **Rennes** – A. Lokchine : analyse centaine d'exomes pour publication sur les maladies du fer
- **Baclesse** – C. Renaud : analyse de plus de 40 patientes en long reads sur les prédispositions au cancer du sein et de l'ovaire dans les régions non-codantes (200+ prévues)
- **PERIGENOMED** : DDN sur 800 gènes traitables
- **Appel à manifestation d'intérêt** : club utilisatrices (nécessite NDA et validation en attendant licence)

REMERCIEMENTS

- Anne-Sophie Denommé-Pichon
- Mathieu Chopelet
- Maximilian Blacher
- Grégoire Hoarau
- Boris Tostivint
- Christophe Marsollier
- Alexandre Cuffel
- Cassandra Anquetil

- Bénédicte Nouyou
- Wilfrid Carré
- Anna Lokchine
- Paul Rollier
- Florent Denoual

CLCC François Baclesse - Caen

- Laurent Castéra
- Thibaut Lavolé
- Anthony Dumas
- Antoine Chouteau
- Alexandre Atkinson

GCS HUGO – GEM-Excell

- Laurence Jay-Passot
- Fanny Gaudin
- Amandine Charreton
- Emmanuelle Lecommandeur
- Sylvie Odent
- Stéphane Bézieau

- Marie Faoucher
- Houda Hamdi Rozé
- Christèle Dubourg
- Marie Beaumont
- Laura Do Souto Ferreira
- Louis Legoff
- Patrick Vourc'h
- Céline Bris
- Simon Bouzy
- Audrey Menard

- Éric Mallevialle
- Alexis Eugène
- Cédric Quilleveré

