

Genetic Output Analysis Tool



Présentée par :

Beatriz Kanzki

Laboratoire du Dr. Hamet

Superviseurs :

Dr. Pavel Hamet

Dr. Michael Philips

Dr. Alain April



“Tout n’est pas écrit dans les livres, quelquefois on détecte quelque chose par expérience, par bon sens, ou par la logique. Mais d’autres fois, il faut prendre du recul pour réanalyser et rechercher ce qu’on n’a pas remarqué avant.”

“Not everything is written in books, sometimes you detect some things with experience, common sense, or logic, but other times you have to take a step back and reanalyze to see what exactly did you miss.”

Dr Adrien Andre
Orthopedic surgeon, Traumatology
Haiti’s general hospital

Plan de la presentation

1. Introduction et contexte du projet.
2. Présentation du projet realisée:
 1. Historique
 2. But
 3. Hypothèse
 4. Outils utilisés
 5. Fonctionnement.
3. Conclusion

Introduction et presentation du contexte

- Objectif de la recherche:
 - Accroître la connaissance,
 - Faire reculer l'incertitude.
- Dans ce domaine il est cruciale d'avoir accès aux résultats déjà obtenus.
- **GOAT (Genetic Output Analysis Tool)** est inspiré d'un outil du laboratoire du Dr Hamet.

Historique

- Problèmes d'efficacité
- Problèmes de performance
- Problèmes d'accès

Tableau de resultats

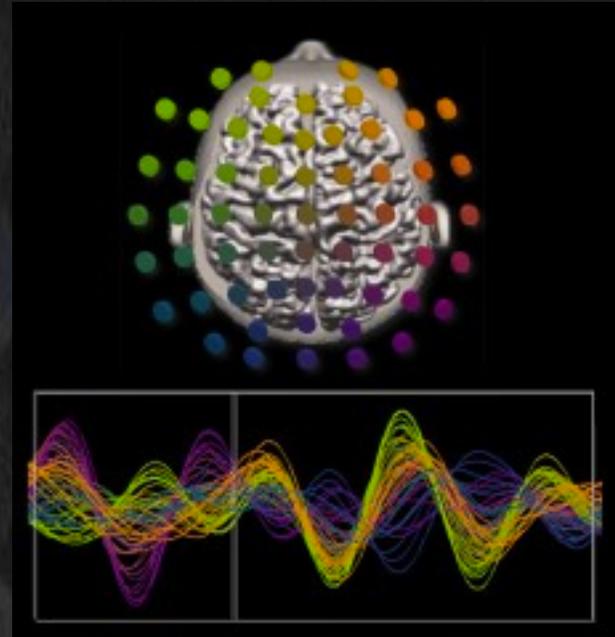
Metaanalysis									
snp	allele1	allele2	effet	se	pvalue	direction	hetisq	chromosome	position
CKD overall with ADVANCE ^{^top^}									
snp	allele1	allele2	effet	se	pvalue	direction	hetisq	chromosome	position
rs4606726	a	g	-7.93e-2	1.86e-2	1.92e-5	-----+-----+---+	9.6	16	20383700
rs12708631	a	t	-0.14	4.21e-2	5.87e-4	-----?-?-?-+---?+	0.0	16	20365697
rs9928757	c	g	-0.19	2.51e-2	1.14e-13	-----+-----?-----	0.0	16	20352863
rs4293393	a	g	0.20	2.47e-2	1.33e-15	+++++-----+-----+	3.6	16	20364588
rs9928936	a	g	-0.19	2.52e-2	1.42e-13	-----+-----?-----	35.2	16	20353049
rs12446492 ^[b]	a	t	-0.11	1.98e-2	1.45e-8	-----+-----+-----+	26.4	16	20408377
rs4297685	t	c	9.84e-2	2.05e-2	1.67e-6	+++-----+-----+?+++++---	0.0	16	20368564
rs13329952	t	c	0.20	2.55e-2	1.80e-15	+++++-----+?+++++?+++++?+	16.9	16	20366507
rs11647727	a	g	-0.11	2.17e-2	2.35e-7	+++-----?-----+	6.1	16	20356165
rs7185940	a	g	-0.11	2.07e-2	2.86e-7	---+-----?-----+	2.6	16	20410488
rs9928648	a	g	0.11	2.08e-2	2.98e-7	+++-----+-----+?+++++?+++++?+	0.0	16	20410669
rs7185389	t	g	-0.11	2.07e-2	3.44e-7	---+-----?-----+	0.0	16	20410263
rs9926580	t	c	-0.11	2.07e-2	3.45e-7	---+-----?-----+	0.0	16	20410547
rs12922822	t	c	-0.20	2.56e-2	4.67e-15	---+-----?-----	17.9	16	20367645
rs12917707	t	g	-0.20	2.57e-2	6.16e-15	---+-----?-----	21.8	16	20367690
rs13333226	a	g	0.20	2.46e-2	7.71e-16	+++++-----+-----+	7.9	16	20365654
rs13335818	t	c	-0.20	2.48e-2	7.75e-16	-----+-----	12.0	16	20359831
rs11864909	t	c	-0.15	2.15e-2	9.15e-12	-----+-----?-----+++	32.5	16	20400839
CKD overall without ADVANCE ^{^top^}									
snp	allele1	allele2	effet	se	pvalue	direction	hetisq	chromosome	position
rs4606726	a	g	-8.09e-2	1.90e-2	2.01e-5	-----+-----+---+	13.4	16	20383700

Hypothèse

Serait-il possible qu'une interface facile d'utilisation, conçu avec des logiciels puissent présenter de l'information complexe de manière plus efficace, et de regrouper toutes les ressources externes pour ce champ de recherche en un seul endroit ?

Outils disponible

- Plusieurs outils de visualisation ont déjà été conçu avec des gratuits.
- Pour ne citer que quelques uns:
 - LocusZoom
 - GWAS pipeline
- Beaucoup de gratuits offrent une alternative aux logiciels d'analyses de données.



FAU Neuroscientist Receive Patent for New 5D Method to understand big Data

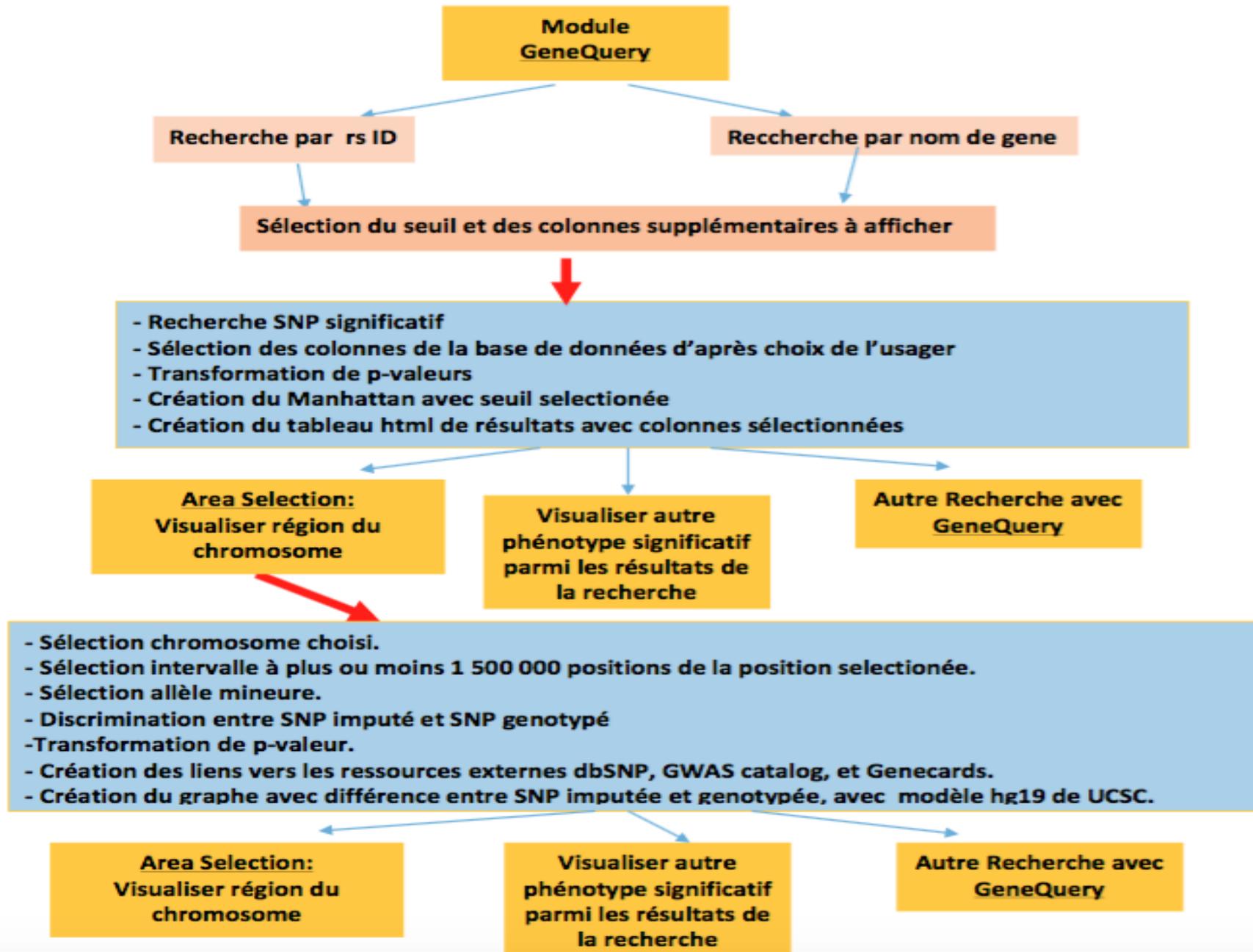
But de GOAT

1. Rendre l'information de la base de données disponibles.
2. Intégrer des outils de visualisation pour ces derniers.
3. Rendre la recherche de résultats plus efficace.

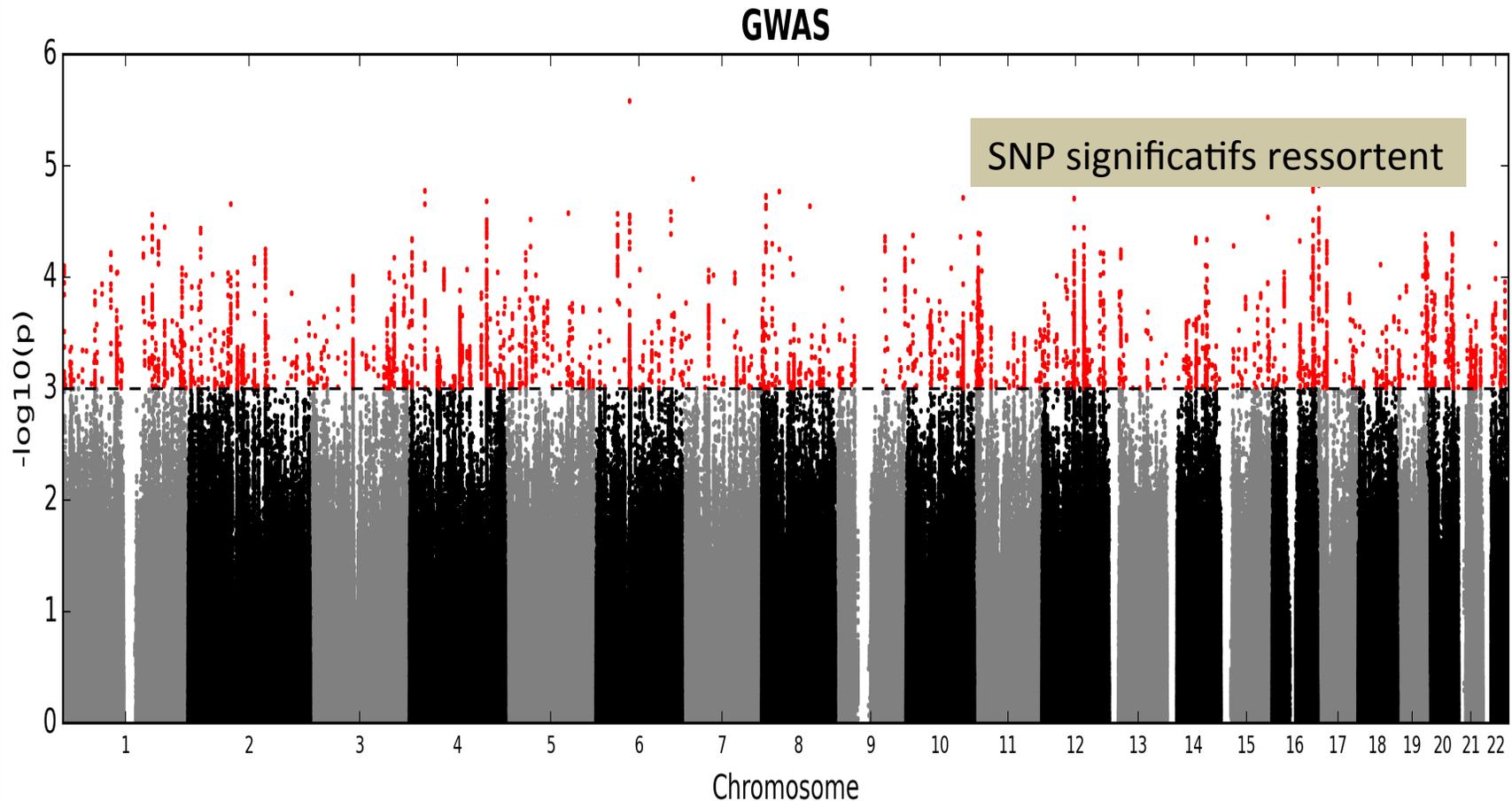
Outils

- Librairie de Biopython: Bokeh, Numpy, Pandas, Blaze, Flask, Matplotlib avec Django (Framework)
- MySQL pour les requetes dans la base de données.

Fonctionnement du 1er module de **GOAT**



Graphe du phenotype “All cause death”



Données Obtenues de la base de données du Dr Hamet, et créé avec Biopython matplotlib, du groupe Anaconda de Continuum Analytics.

Tableau vue générale

[Download CSV File](#)

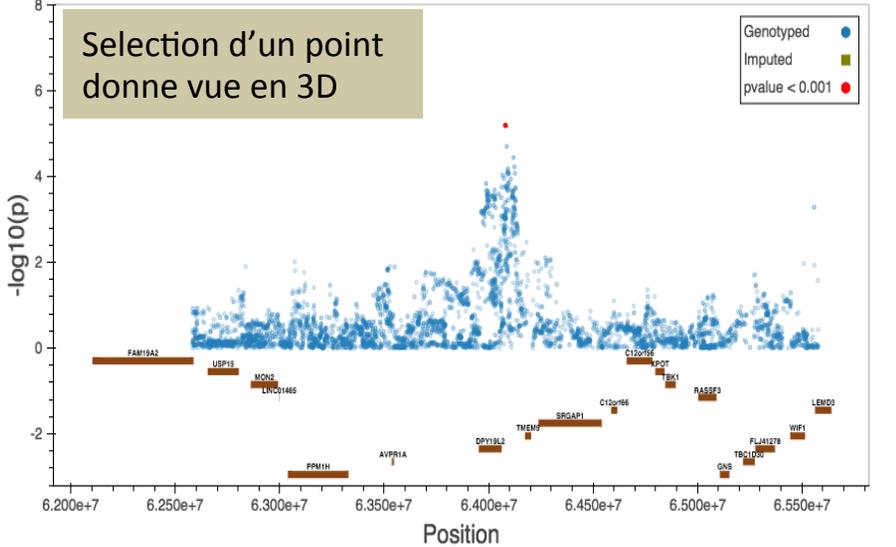
rs_id_assoc	chromosome	pos	pvalue_assoc	allele_A	allele_B	covariates	beta_assoc	maf	all_OR
rs7543737	1	3009969	0.000112	G	A	age sex pc1 pc2	0.31853	0.258920	1.30230
rs760565	1	3010317	0.000308	A	T	age sex pc1 pc2	0.28900	0.394010	1.23670
rs4648361	1	3010327	0.000093	G	A	age sex pc1 pc2	0.32502	0.260900	1.30020
rs734373	1	3010703	0.000090	G	T	age sex pc1 pc2	0.32746	0.259890	1.30230
rs734372	1	3010853	0.000080	G	A	age sex pc1 pc2	0.33010	0.260310	1.30310
rs734374	1	3010969	0.000145	A	C	age sex pc1 pc2	0.32191	0.255850	1.28690
rs2981874	1	3011286	0.000087	C	G	age sex pc1 pc2	0.33420	0.258630	1.28830
rs2981873	1	3011329	0.000428	T	C	age sex pc1 pc2	0.28653	0.390460	1.22270
rs10449220	1	3011549	0.000852	G	A	age sex pc1 pc2	0.30516	0.226220	1.24330
rs10449221	1	3011558	0.000778	G	A	age sex pc1 pc2	0.30709	0.226870	1.24470
rs2993491	1	3011610	0.000709	T	C	age sex pc1 pc2	0.30975	0.228860	1.24190
rs61768911	1	3748982	0.000826	C	G	age sex pc1 pc2	0.49624	0.071450	1.62730

Colonnes apparaissant par défaut en utilisant le Pandas dataframe de Biopython de Continuum Analytics

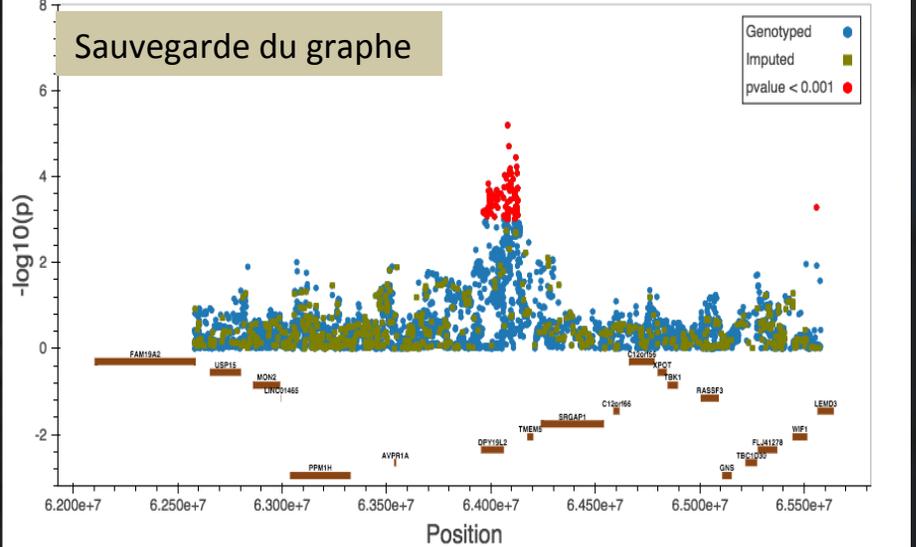


Vue spécifique "Area Selection"

CHROMOSOME :12 BETWEEN POSITIONS 62580518 AND 65580518



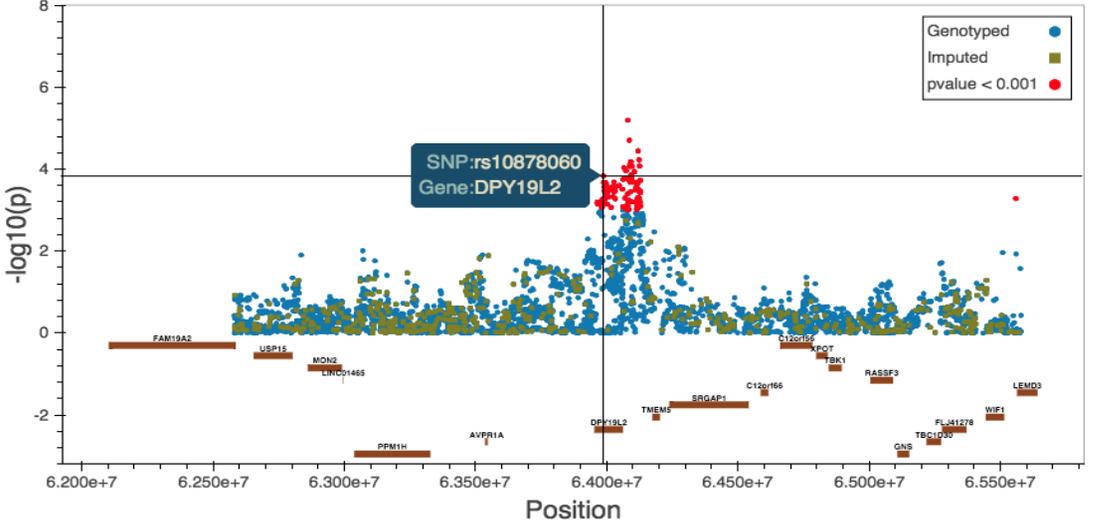
CHROMOSOME :12 BETWEEN POSITIONS 62580518 AND 65580518



Vue avec interaction
(zoom, selection de points, nom de gene)



CHROMOSOME :12 BETWEEN POSITIONS 62580518 AND 65580518



Généré avec Bokeh de Continuum Analytics. Une librairie permettant le partage de graphe sur le web.

Tableau vue spécifique

bkanzki Gene Query Area Selection Select Significant Phenotype to display : All cause death

[Download CSV File](#)

rs_ID	Chr	Position	Gene before	Gene	Gene after	P-value	Imputed	Allele A	Allele B	Allele mineure	Cohort	dbSNP	GWAS_Catalog	Genecards
rs12823385	12	65020085	MIR548C	RASSF3	GNS	0.352670	False	T	C	C	Advance_3449	dbSNP	By rsID By gene before By gene By gene after	By gene before By gene By gene after
rs35860561	12	65020484	MIR548Z	RASSF3	GNS	0.953360	False	G	A	A	Advance_3449	dbSNP	By rsID By gene before By gene By gene after	By gene before By gene By gene after

Informations contenues sur le graphe plus l'allèle mineure, la cohort qui avait été utilisé et les liens vers ressources externes. Obtenue Avec Pandas dataframe de Continuum Analytics.



Conclusion

La création de GOAT a été possible grâce à la demande des utilisateurs du Laboratoire du Dr Hamet qui voulaient d'un outils plus efficace pour une meilleure visualisation des résultats de la base de données.

Un grand effort a été fait pour créer une interface simple, qui utilisent les ressources computationnelles disponibles pour la visualisation des résultats.

Mon intention est de continuer à lui ajouter d'autres modules tels que : "Phenotype comparison", "Results pairing", "Statistical Analysis", "Interaction Analysis", "GWAS" tirant profit des technologies BIG DATA.

Remerciement

- Dr Pavel Hamet et Dre Johanne Tremblay pour leur support dans ce projet.
- Dr Michael Philips pour son appui et ses conseils.
- Dr Alain April et son équipe pour son appui et la coordination du projet.
- L'équipe du Dr Hamet pour leur collaboration et support.
- Remerciement à l'équipe de Continuum Analytics.



Références

- P. Joly, **La recherche médicale et sa valorisation**, La santé et la médecine à l'aube du XXI siècle, Magazine N°562 Février 2001.
- E. HalperinandDietrich, A. Stephan, **SNP Imputation in association studies**, Nature Biotechnology 27, 349-351 (2009), doi: 10.1038/nbt0409-349.
- J.Y.Dai, I. Ruczinski, M. Leblanc, C. Kooperberg, **Imputation Methods to improve Inference in SNP Association Studies**, Genet. Epidemiol. DOI 10.1002/gepi
- Chatterjee, S., Gopidi, N., Kyasa, R., and Prashanth Ravi, P., **Evaluation of Open Source Tools and development Platforms for Data Analysis in Engine Development**, SAE Technical Paper 2015-26-0076, 2015, doi: 10.4271/2015-26-0076.
- C. Rossant, K. D. Harris, **Hardware-accelerated interactive data visualization for neuroscience in Python**, Frontiers Neuroinformatics 2013;7:36, PMCID: PMC3867689, doi: 10.3389/fninf.2013.00036

Références (suite)

- V.M. Ayer, S. Miguez, B. H. Toby, **Why scientists should learn to program in Python**, Crystallography Education, Advanced Photon source, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, Illinois 60439-4814, doi: 10.1017/S0885715614000931.
- R. Wang, Y. Perez-Riverl, H. Hermjakob, J. A. Vizcaino, **Open source librairies and frameworks for biological data visualization: A guide for developpers**, Proteomics, 5 FEB 2015, DOI: 10.1002/epmic.201400377.
- D. Farrell, S. V. Gordon, **Epitopemap: a web application for integrated whole proteome epitope prediction**, BMC Bioinformatics 2015, 16:221, doi: 10.1186/s12859-015-0659-0.