

Développement d'un Tableau de bord pour la visualisation des données.

- **Description :**

Développer une plateforme dédiée à la manipulation, la visualisation, et au filtrage des variations à partir d'un fichier VCF annoté.

- **Les objectifs :**

- Upload les données de plusieurs patients.
- Analyser les types des mutations pour chaque patient.
- Filtrer les résultats par gène, type de mutation, impact génomique...
- Visualiser, en utilisant des graphes et des schémas, la distribution de chaque information (gènes, mutations (SNP/indel) , chromosomes..) pour chaque patient.
- Générer un rapport contenant les résultats de l'analyse.

- **Data :**

→ **Fichiers VCF :** Le format VCF est un format de variants génétiques (SNP, insertions/délétions) obtenu suite à un alignement de lectures de séquençage NGS sur une séquence de référence. Il s'agit d'un format tabulé qui peut être facilement parsé en Perl pour en extraire l'information et recueillir des statistiques.

- Le fichier contient plusieurs colonnes :

#CHROM	POS	ID	REF	QUAL	FILTER	INFO
1	1368647		A	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=T upstream_gene_variant(MODIFIER VWA1 transcript NM_022834.5 protein_coding c.-2482A>T 2423)
1	1659010		G	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=T intron_variant(MODIFIER CDK11B transcript NM_001787.3 protein_coding 4/19c.355+504C>A)
1	1659010		T	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=C intron_variant(MODIFIER CDK11B transcript NM_001787.3 protein_coding 4/19c.355+357A>G)
1	3397188		T	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=C _prime_UTR_variant(MODIFIER RHGEF1 transcript NM_014448.4 protein_coding 15/15c.*37T>C 37)
1	3744249		A	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=G intron_variant(MODIFIER CEP104 transcript NM_014704.4 protein_coding 15/21c.2152-941T>C)
1	3744261		C	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=A intron_variant(MODIFIER CEP104 transcript NM_014704.4 protein_coding 15/21c.2152-953G>T)
1	4646236		A	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=G downstream_gene_variant(MODIFIER LINC01646 transcript NR_147025.1 pseudogene n.*2472A>G 2472)
1	4646265		G	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=A downstream_gene_variant(MODIFIER LINC01646 transcript NR_147025.1 pseudogene n.*2501G>A 2501)
1	5002713		T	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=C intergenic_region(MODIFIER AJAP1-MIR4889 intergenic_region AJAP1-MIR4889 n.5002713T>C)
1	8409916		T	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=C intron_variant(MODIFIER ACOT7 transcript NM_183884.3 protein_coding 2/9c.292-1267A>G)
1	704325		G	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=T intron_variant(MODIFIER CAMTA1 transcript NM_015215.4 protein_coding 3/22c.235-107029G>T)
1	7399674		C	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=T intron_variant(MODIFIER CAMTA1 transcript NM_015215.4 protein_coding 5/22c.438+90188C>T)
1	7917167		T	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=G upstream_gene_variant(MODIFIER UTS2 transcript NM_021995.2 protein_coding c.-3676A>C 3616)
1	8413559		C	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=T _3_prime_UTR_variant(MODIFIER RERE transcript NM_001042681.2 protein_coding 23/23c.*1588G>A 1588)
1	8420997		A	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=G missense_variant(MODIFIER CAG transcript NM_001042681.2 protein_coding 18/23c.c.2570T>C p.Leu857Pro 31)
1	9008045		C	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=T intron_variant(MODIFIER CAG transcript NM_001042681.2 protein_coding 17/7c.80-1277C>T)
1	10748723		A	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=C intron_variant(MODIFIER CASZ1 transcript NM_001079843.3 protein_coding 4/20c.16+5208T>G)
1	10980017		T	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=C intergenic_region(MODIFIER CASZ1-C1orf127 intergenic_region CASZ1-C1orf127 n.10980017T>C)
1	10980079		G	78.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=25.36;SOR=0.693;ANN=A intergenic_region(MODIFIER CASZ1-C1orf127 intergenic_region CASZ1-C1orf127 n.10980079G>A)
1	10980083		C	78.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=28.73;SOR=0.693;ANN=T intergenic_region(MODIFIER CASZ1-C1orf127 intergenic_region CASZ1-C1orf127 n.10980083C>T)
1	12145364		G	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=A intron_variant(MODIFIER TNFRSF8 transcript NM_001243.5 protein_coding 2/14c.151+756G>A)
1	12258941		G	37.32	PASS	AC=2AF=1.00;AN=2;DP=2;ExcessHet=0.0000;FS=0.000;MLEAC=1;MLEAF=0.500;MQ=60.00;QD=18.66;SOR=0.693;ANN=T intron_variant(MODIFIER TNFRSF1B transcript NM_001066.3 protein_coding 8/9c.91-3083G>T)

La colonne **INFO** regroupe plusieurs informations concernant le nom du gène, l'impact de la mutation, la description de la région...

- **Outils et langages :** à discuter.

- **Exemple :**

Vous trouverez ci-dessous une implémentation d'une plateforme pour l'analyse de ce type de fichier :

<https://github.com/labsquare/cutevariant>

Sur la figure suivante, vous pouvez voir un exemple du type des données filtrées qui nous intéressent:

File Edit View Tools Help

+ New project Open project Import file Show VQL editor

type a gene or coordinate ...

FIELDS

17 fields

Search for a field ...

count_het

count_ref

count_none

count_tot

count_var

freq_var

count_validation_positive

count_validation_negative

count_validation_positive_sample_lock

count_validation_negative_sample_lock

control_count_hom

control_count_het

control_count_ref

case_count_hom

case_count_het

case_count_ref

FILTERS

AND (2)

ann.impact

Equal to

HIGH

ann.consequ...

Contains

frameshift

Fields Source Filters

variants

Auto resize Refresh Stop

	ref	alt	ann.hgvs_p	ann.hgvs_c	ann.impact	ann.gene	ann.consequence
<input type="radio"/>	AT	A	p.Ile787fs	c.2359delA	HIGH	SF3B1	frameshift_variant
<input type="radio"/>	CA	C	p.Gly435fs	c.1302delT	HIGH	GATA2	frameshift_variant
<input type="radio"/>	CA	C	p.Gly421fs	c.1260delT	HIGH	GATA2	frameshift_variant
<input type="radio"/>	AC	A	p.Gly200fs	c.599delG	HIGH	GATA2	frameshift_variant
<input checked="" type="radio"/>	T	TGATCATATTCATAT	p.Asn572fs	c.1713_1714insCTT	HIGH	FBXW7	frameshift_variant
<input type="radio"/>	T	TGATCATATTCATAT	p.Asn492fs	c.1473_1474insCTT	HIGH	FBXW7	frameshift_variant
<input type="radio"/>	T	TGATCATATTCATAT	p.Asn454fs	c.1359_1360insCTT	HIGH	FBXW7	frameshift_variant
<input checked="" type="radio"/>	C	CTCTG	p.Trp288fs	c.860_863dupTCTC	HIGH	NPM1	frameshift_variant
<input checked="" type="radio"/>	C	CTCTG	p.Trp224fs	c.668_671dupTCTC	HIGH	NPM1	frameshift_variant
<input checked="" type="radio"/>	C	CTCTG	p.Trp161fs	c.479_482dupTCTC	HIGH	NPM1	frameshift_variant
<input checked="" type="radio"/>	C	CTCTG	p.Trp259fs	c.773_776dupTCTC	HIGH	NPM1	frameshift_variant
<input type="radio"/>	TC	T	p.Glu33fs	c.97delG	HIGH	CDKN2A	frameshift_variant
<input type="radio"/>	TG	T	p.Ser2542fs	c.7623delC	HIGH	NOTCH1	frameshift_variant
<input type="radio"/>	CG	C	p.Ala2463fs	c.7386delC	HIGH	NOTCH1	frameshift_variant
<input type="radio"/>	TA	T	p.Lys440fs	c.1319delA	HIGH	PTEN	frameshift_variant&splice_region_variant
<input type="radio"/>	TA	T	p.Lys267fs	c.800delA	HIGH	PTEN	frameshift_variant&splice_region_variant
<input type="radio"/>	TA	T	p.Lys70fs	c.209delA	HIGH	PTEN	frameshift_variant&splice_region_variant
<input type="radio"/>	TC	T	p.Thr146fs	c.435delG	HIGH	IDH2	frameshift_variant
<input type="radio"/>	TC	T	p.Thr16fs	c.45delG	HIGH	IDH2	frameshift_variant
<input type="radio"/>	TC	T	p.Thr94fs	c.279delG	HIGH	IDH2	frameshift_variant
<input type="radio"/>	A	AG	p.Gly646fs	c.1934dupG	HIGH	ASXL1	frameshift_variant

Executed in 0.026s Cache 336.0Bytes of 32.0MB 35 line(s) Page 1 on 1

Samples: 0 Source: variants