DIU Génétique et Reproduction

Exercices NGS Séquençage ciblé

16 Septembre 2021
Philippe Dessen
Institut Gustave Roussy, Villejuif

dessen@igr.fr

Abdelkader Heddar
Hôpital Bicêtre

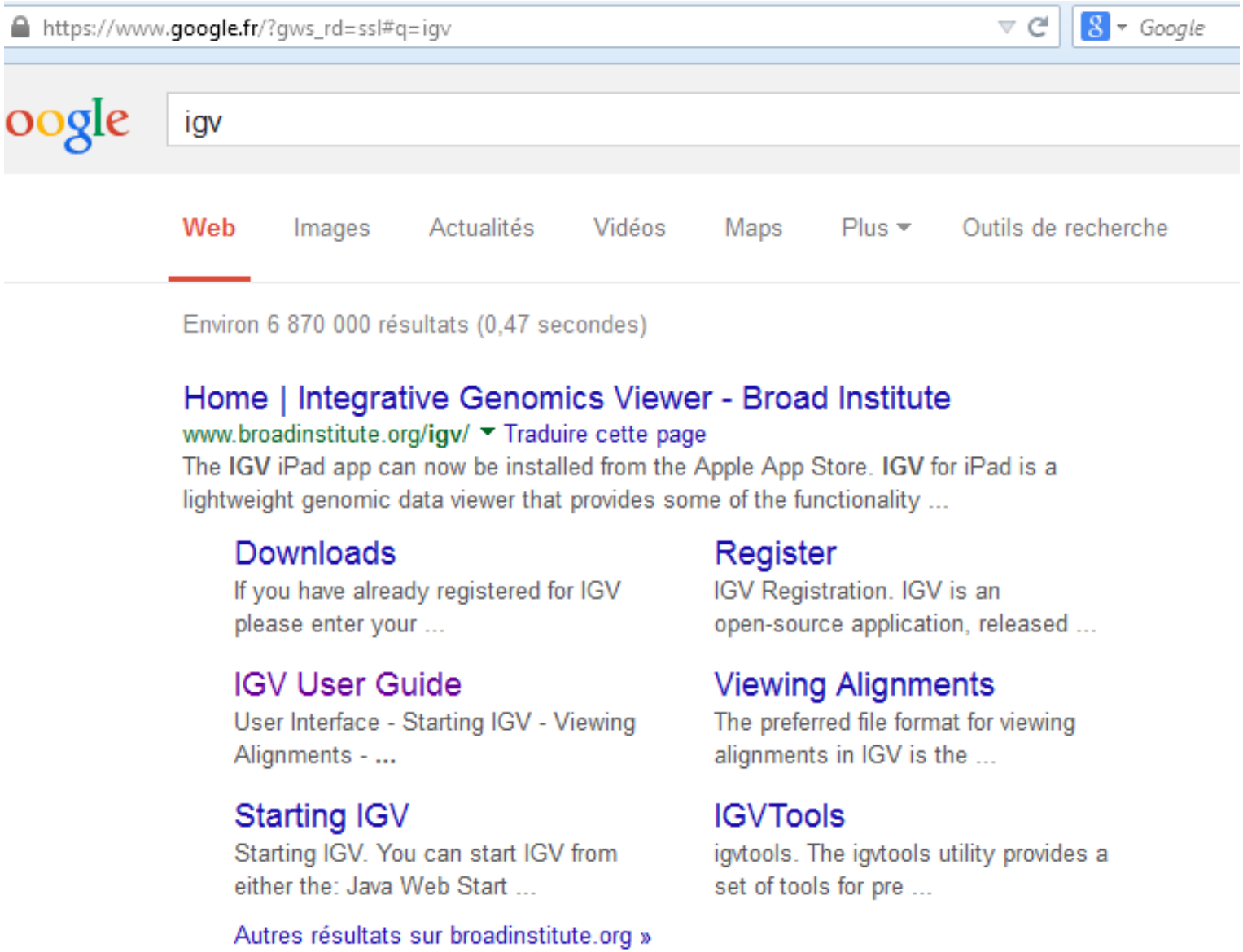
abdelkader.heddar@aphp.fr

Exercices

Analyse d’exomes

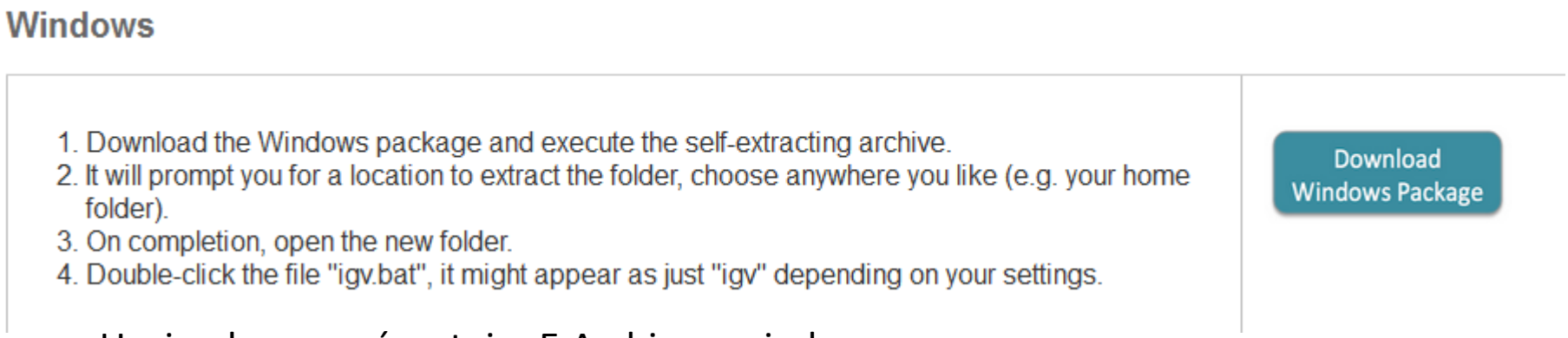
Récupérer sur http://pdessen.free.fr/KB

IGV Integrative Genomic Viewer



http://software.broadinstitute.org/software/igv/

Install IGV :
http://software.broadinstitute.org/software/igv/download



Unzip dans un répertoire E:Archives_windows
IGV_2.3.82_jre.exe

Exécuter en cliquant sur IGV.bat

On peut aussi (voir page suivante) charger la Java Web Start

Par ailleurs
Récupérer une archive contenant les fichiers bam sur :
http://pdessen.free.fr/KB/BAM.zip
Désarchiver ce fichier (clic droit et extraction) sur le bureau

Java Web Start (All Platforms)

The buttons below use Java Web Start (JWS) to install and launch IGV directly from our web site.

*__Mac Users__: The Java Web Start option does not work for some users due to security settings. The recommended solution is to use the bundled Mac App from the link above. Alternatively you can try to work around this by right-clicking on the buttons and saving the "jnlp" file, then right-clicking on the saved "jnlp" file and select "Open With > Java Web Start".

Chrome: Chrome does not automatically launch the Java Web Start files by default. Instead, the launch buttons below will download a "jnlp" file. This should appear in the lower left corner of the browser. Double-click the downloaded file to run, or if on a Mac right-click and select "Open With > Java Web Start"

Windows users: To run with more than 1.2 GB of memory you must install 64-bit Java. ***Most Windows installs do not include 64-bit Java by default, even if the operating system is 64-bit***. Attempting to use the 2GB or greater launch options with 32-bit Java will result in the error *"could not create virtual machine"*.

			
Launch with 750 MB	Launch with 1.2 GB Maximum usable memory for Windows OS with 32-bit Java.	Launch with 2 GB Maximum usable memory for 32-bit MacOS.	Launch with 10 GB For large memory machines with 64-bit Java.

BAM.zip (sur http://pdessen.free.fr/KB/BAM)

Unzip BAM.zip

BAM

- test_HCV:
  - B00G39Y_chr16.bam
  - B00G39Y_chr16.bam.bai
  - B00G39Y_chr5.bam
  - B00G39Y_chr5.bam.bai
- test_Tumor_Ctrl:
  - control
    - control_sorted.bam
    - control_sorted.bam.bai
    - QC_filtered
    - QC_raw
  - tumor
    - QC_filtered
    - QC_raw
    - tumor_sorted.bam
    - tumor_sorted.bam.bai

Mutations dans des gènes impliqués dans la reconnaissance du virus HCV (étude le M. MISRAHI)

Pour des questions de volume on a limité les données d'exome à 2 bandes chromosomiques 5q13.2 et 16p13.3

Rappel : Seuls les zones couvrant les CDS et les bornes sont Couvertes par des reads.

Un clic droit de souris fait apparaitre un grand nombre d'options

Dans IGV : **Sélectionner le build hg19**
Charger le fichier B00G39Y_chr5.bam

Sachant qu il y a des reads d un exome autour de la bande 5q13.2
balayer cette zone pour trouver des gènes séquencés en suivant la piste bam coverage

Quels gène(s) ? (autour de 68.800 Mb)
Combien d exons ? (passer la souris sur le gène)
Sens du gène ?
Combien de transcrits ? (option squished sur refSeq)
Zoomer pour regarder les reads ?

Est ce que tous les exons sont couverts ? CDS ?
Profondeur ?
Qualité ?
Trouver la (les) mutations et localisez les à la base près
Proportion d allèles REF et ALT
Repérer la couleur différente sur le coverage . Visualiser les fréquences
Variant homozygote ou hététzyote ?

Examiner les possibilités avec la touche Ctrl sur une position
Comment extraire une séquence locale ?

Dans IGV : Charger le fichier B00G39Y_chr16.bam

Sachant qu il y a des reads d un exome autour de la bande 16p13.3
Balayer cette zone pour trouver des gènes séquencés en suivant la piste bam coverage

Quels gène(s) ?
Combien d exons ?
Sens du gène ?

Zoomer pour regarder les reads

Est ce que tous les exons sont couverts ? CDS ?
Profondeur ?
Qualité ?
Trouver la (les) mutations et localisez les à la base pres
Proportion d allèle REF et ALT
Repérer la couleur différente sur le coverage . Visualiser les frequences
Variant homozygote ou hétérozyote ?

Ctrl sur une position ?
Comment extraire une séquence locale
Une autre mutation ?

Regarder l'impact de ces mutations

Dans dbSNP https://www.ncbi.nlm.nih.gov/SNP/

1000 Genomes http://www.internationalgenome.org/

EVS http://www.internationalgenome.org/

ExAC http://exac.broadinstitute.org/
gnomAD http://gnomad.broadinstitute.org/

Fréquence et présence dans les populations (1000G)
Nature de l'impact ?

Tester PolyPhen2 http://genetics.bwh.harvard.edu/pph2/
Tester Sift http://sift.jcvi.org/www/SIFT_chr_coords_submit.htm

Voir les annotations dans UniProt http://www.uniprot.org

Sélection par CTRL de la séquence consensus

chr16:3065097-3065673

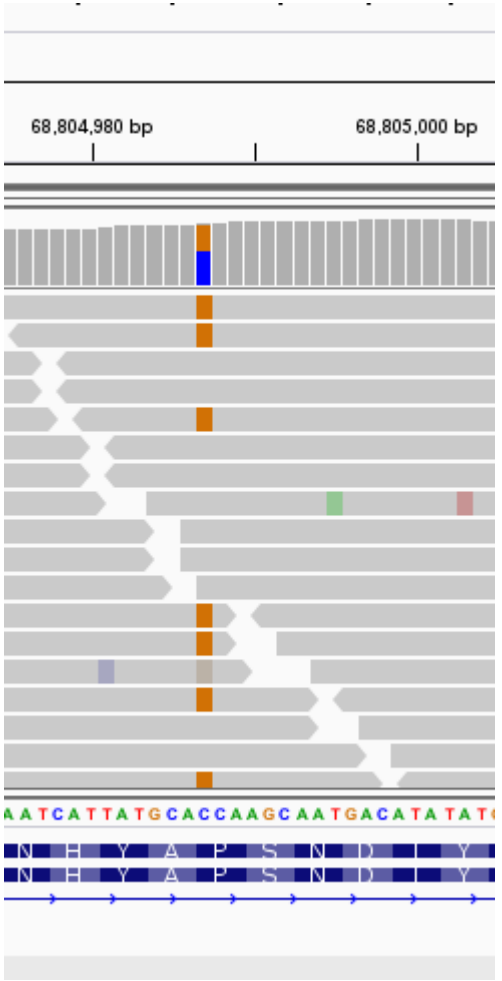
NNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNAAATGAATTTTAAATATC
CTCAGTCAAAAGAAAAATAGCAGGAGGCAGAAA
CAAAAGGTACGAACCCATCCCAAAGCTGTTGGGC
ACTGCCACTTCTGGATGGCTCTAGCGCCAGCGGA
GCCCCCATTCCCCTCCACGTCAGACGTAATTCTTG
GTAGGGTACTCAGAGGGCCCCYGAGAGATGGCA
GGGGCAGATGTTGAGTAGCGGGCCATGTAATGGC
TGGGGCCCTGGGACCCCCCCGAGGGGCAAGTGC
AGCACAGCAACCCCCCACCCAGCAACAAAAGGCC
TGAGGCCGCCCAGCCCAAGTAGAGGGAGGCCCC
CAGCTCCCGCTTTTGGGCCTCAGCCACCAGGGGG
TTATAGAAGTCCCGGATGACGGCATGCGCCGTCCA
GCACACGGGGATTAGCGTCAGGACCCCTGAGATG
ACAAAGACAATCCCAGAGGTGAGCACCA

Clic droit sur le polymorphisme
chr5:68804985 C>G Pro

chr5:68788119-68853931
OCLN
chr5:68788119-68853931
id = NM_002538

Exon number: 3
Amino acid coding number: 24
chr5:68804968-68805646

chr5:68,804,987
<hr>Total count: 144
A : 0
C : 80 (56%, 58+, 22-)
G : 64 (44%, 40+, 24-)
T : 0
N : 0

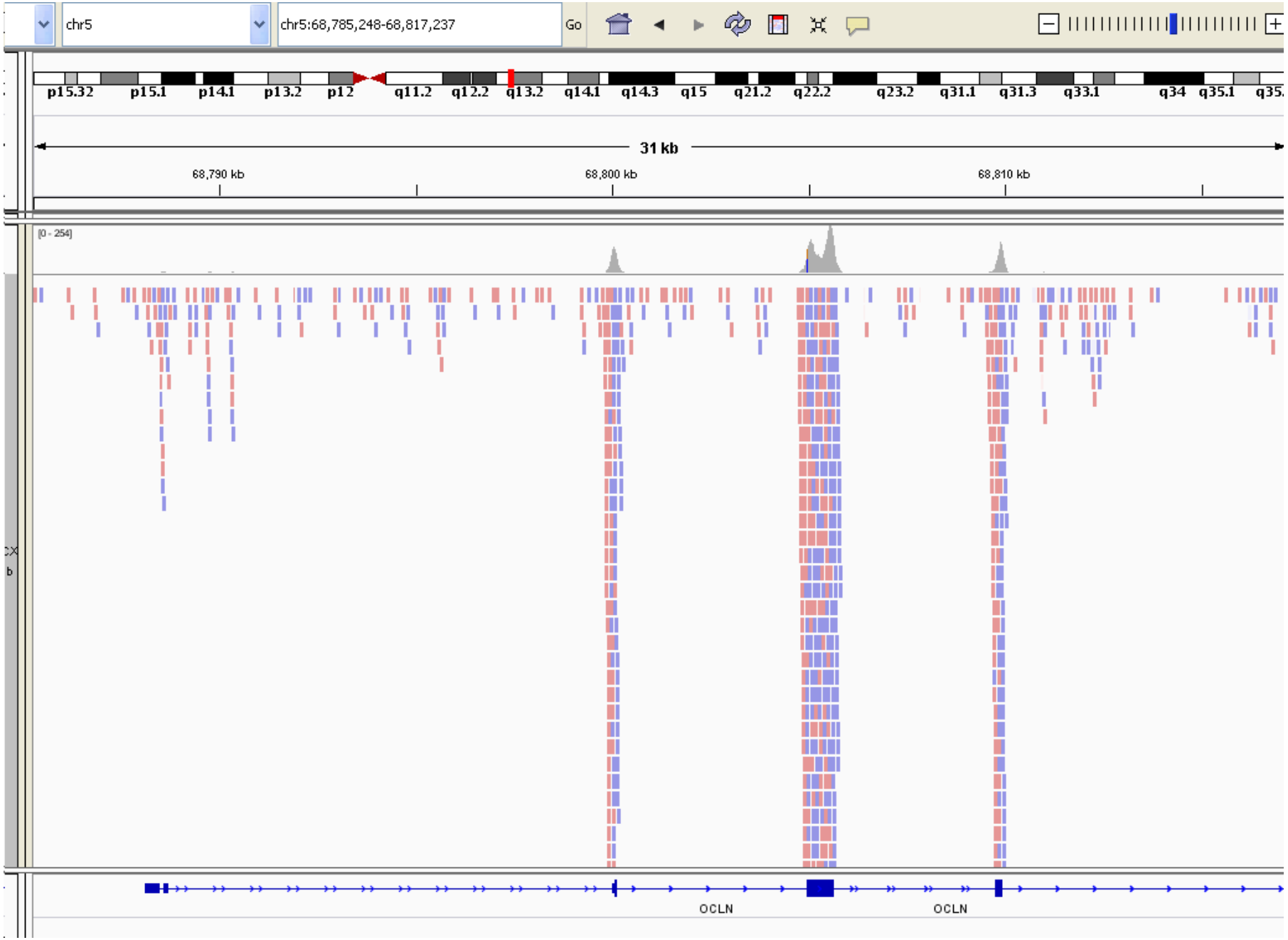


chr	beg	rc_snp	REF	ALT	EFFeffect	EFFimpact	EFFclass	EFFcodon	EFFaa	EFFsymb	B00G39X	B00G39Y
3	190030989	rs10513846	G	A	EXON	MODIFIER				CLDN1	1	0
16	3065397	rs149605777	C	T	NON_SYNONYMOUS_CODING	MODERATE	MISSENSE	cGg/cAg	R209Q	CLDN6	0	1
16	3065635	rs138377300	C	G	NON_SYNONYMOUS_CODING	MODERATE	MISSENSE	Gtc/Ctc	V130L	CLDN6	0	0
16	3065596	rs2257295	T	C	NON_SYNONYMOUS_CODING	MODERATE	MISSENSE	Atc/Gtc	I143V	CLDN6	1	2
16	3063149	rs2231608	T	G	START_GAINED	LOW				CLDN9	0	0
5	68804987	rs147125035	C	G	NON_SYNONYMOUS_CODING	MODERATE	MISSENSE	Cca/Gca	P24A	OCLN	0	1
5	68849586	.	G	A	EXON	MODIFIER				OCLN	0	0
5	68830632	.	C	T	NON_SYNONYMOUS_CODING	MODERATE	MISSENSE	Cgg/Tgg	R13W	OCLN	0	0

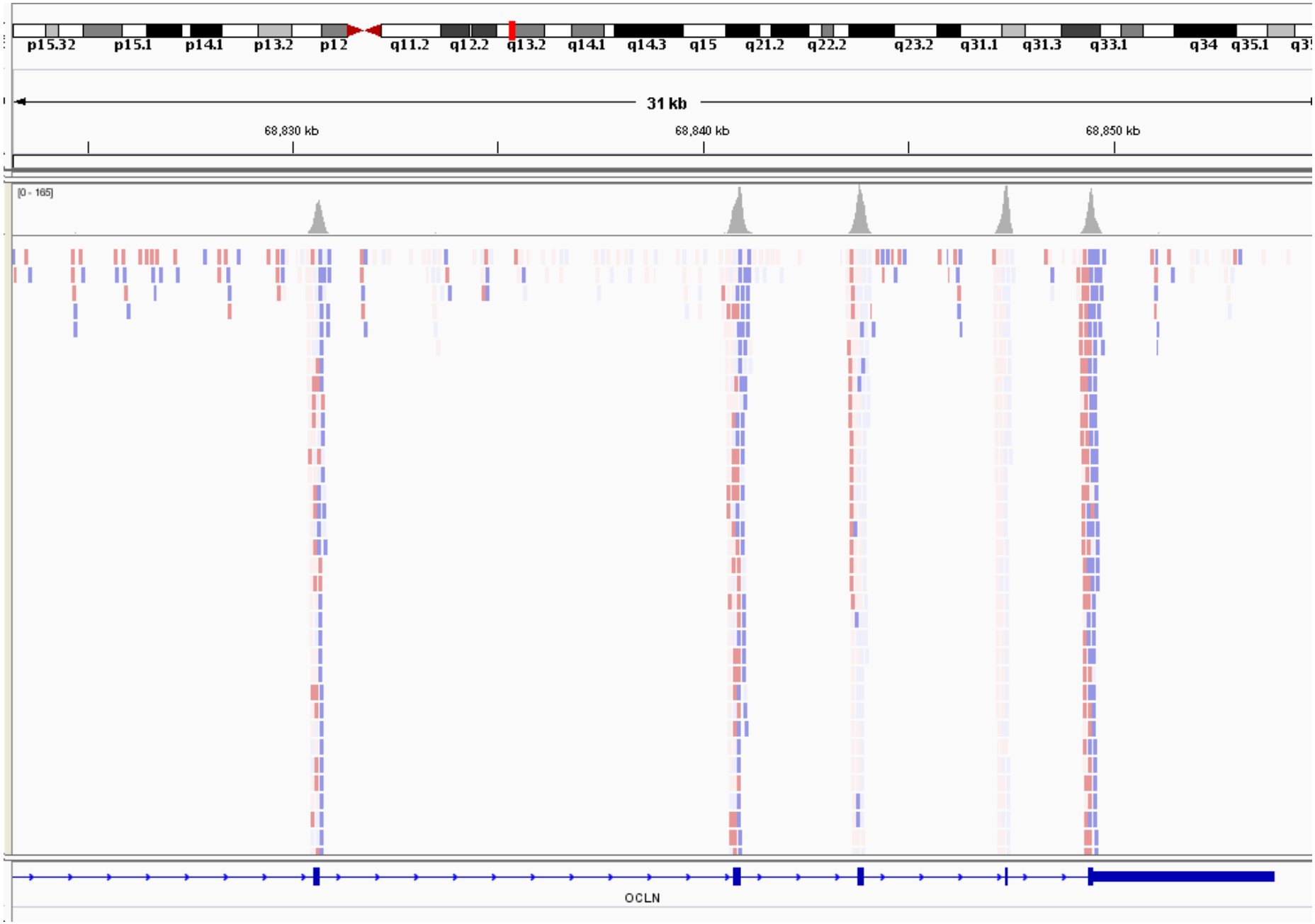
Visualisation des mutations
(validation sur les gènes déjà séquencés en Sanger)

gène	mutation	dbSNP rs#ID		code
CLDN6	c.626 G>A (R209Q)	rs149605777	chr3:3065397	B00G39Y
OCLN	c.70 C>G (P24A)	rs147125035	chr5:68804987	B00G39Y
SR-B1	c.403 G>A	rs5891	chr12:125299542	B00G39Z
SR-B1	NC+23 C>T	NA	en 3'UTR ?	B00G39X
SR-B1	c.1056 G>A (V35I)	rs372663606	chr12:125284742	B00G39Z

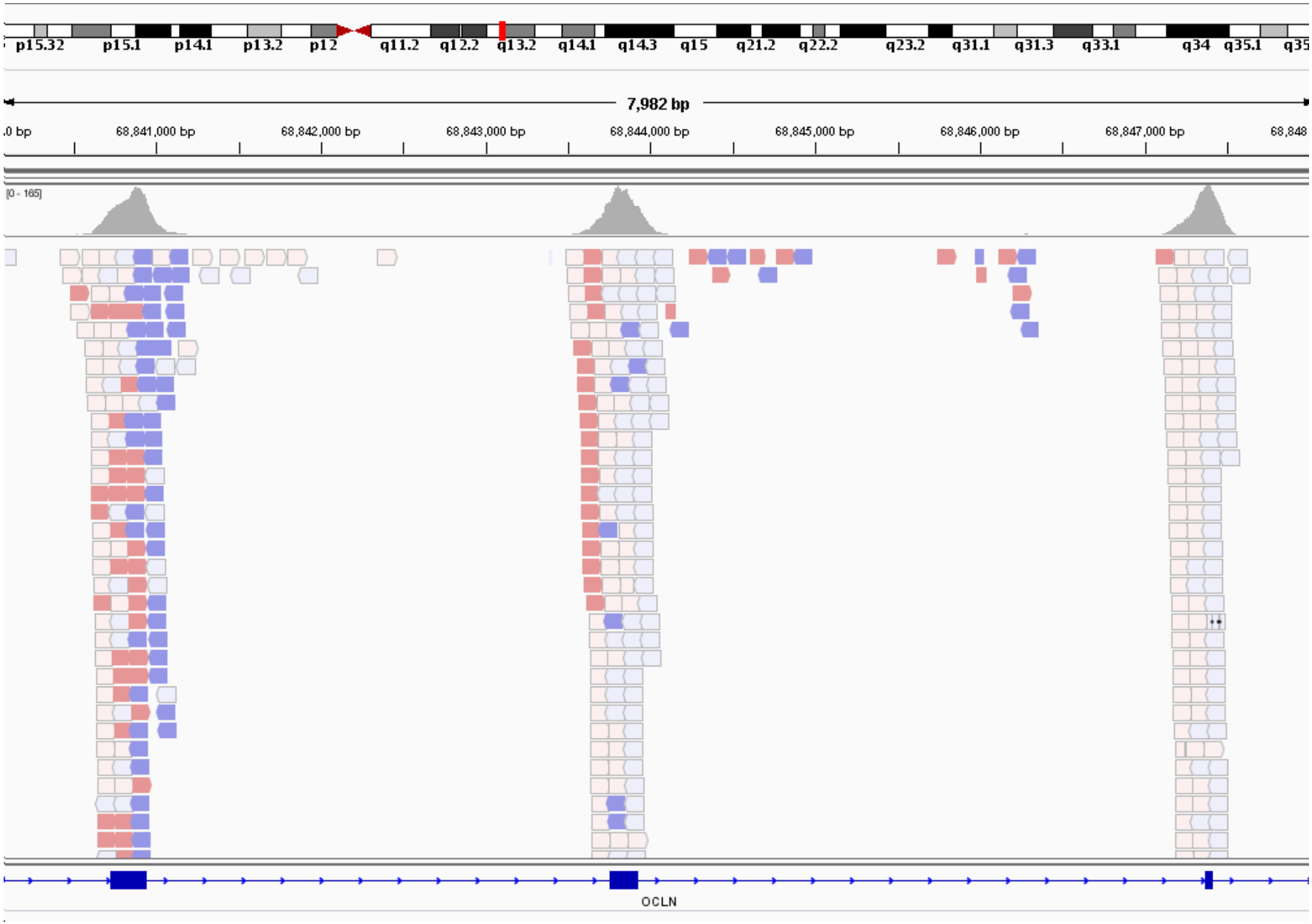
B00G39Y OCLN



B00G39Y OCLN



B00G39Y OCLN



Couverture de capture Agilent v5

A partir du fichier bed : annotation des exons présents

```
$ more refg.txt
chr01:000011874-000012227__DDX11L1__NR_046018__exon__ex_01
chr01:000012613-000012721__DDX11L1__NR_046018__exon__ex_02
chr01:000013221-000014409__DDX11L1__NR_046018__exon__ex_03
chr01:000014362-000014829__WASH7P__NR_024540__exon__ex_11
chr01:000014970-000015038__WASH7P__NR_024540__exon__ex_10
```

820772 exons couverts

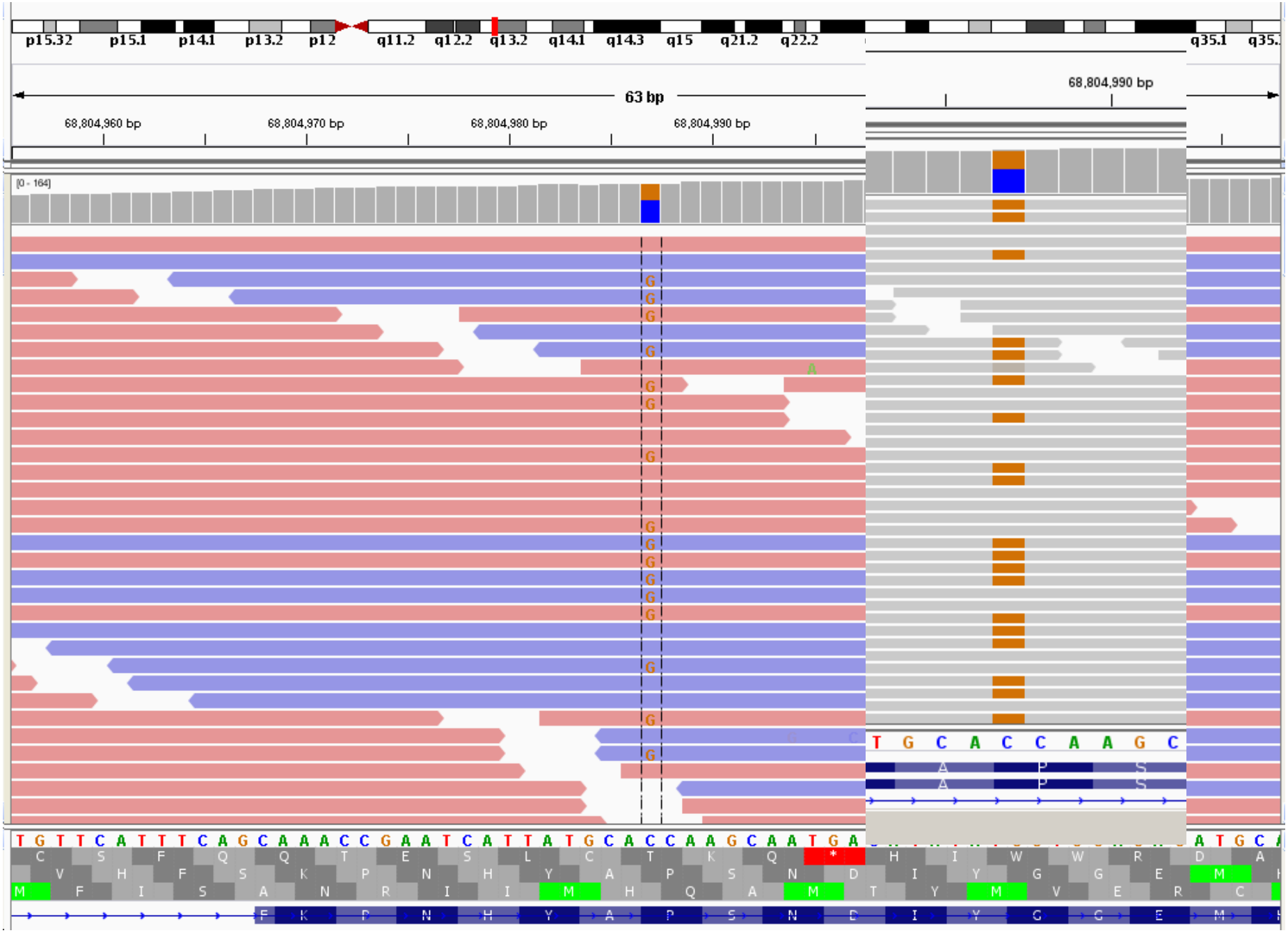
Dans refGene (hg19)

```
grep exon /db/GoldenPath/current/database/refGene.gff | wc -l
  467234
grep exon /db/GoldenPath/current/database/refGene.gff | awk '{ print $2}' |sort -u |wc -l
   25981    nbre de genes
```

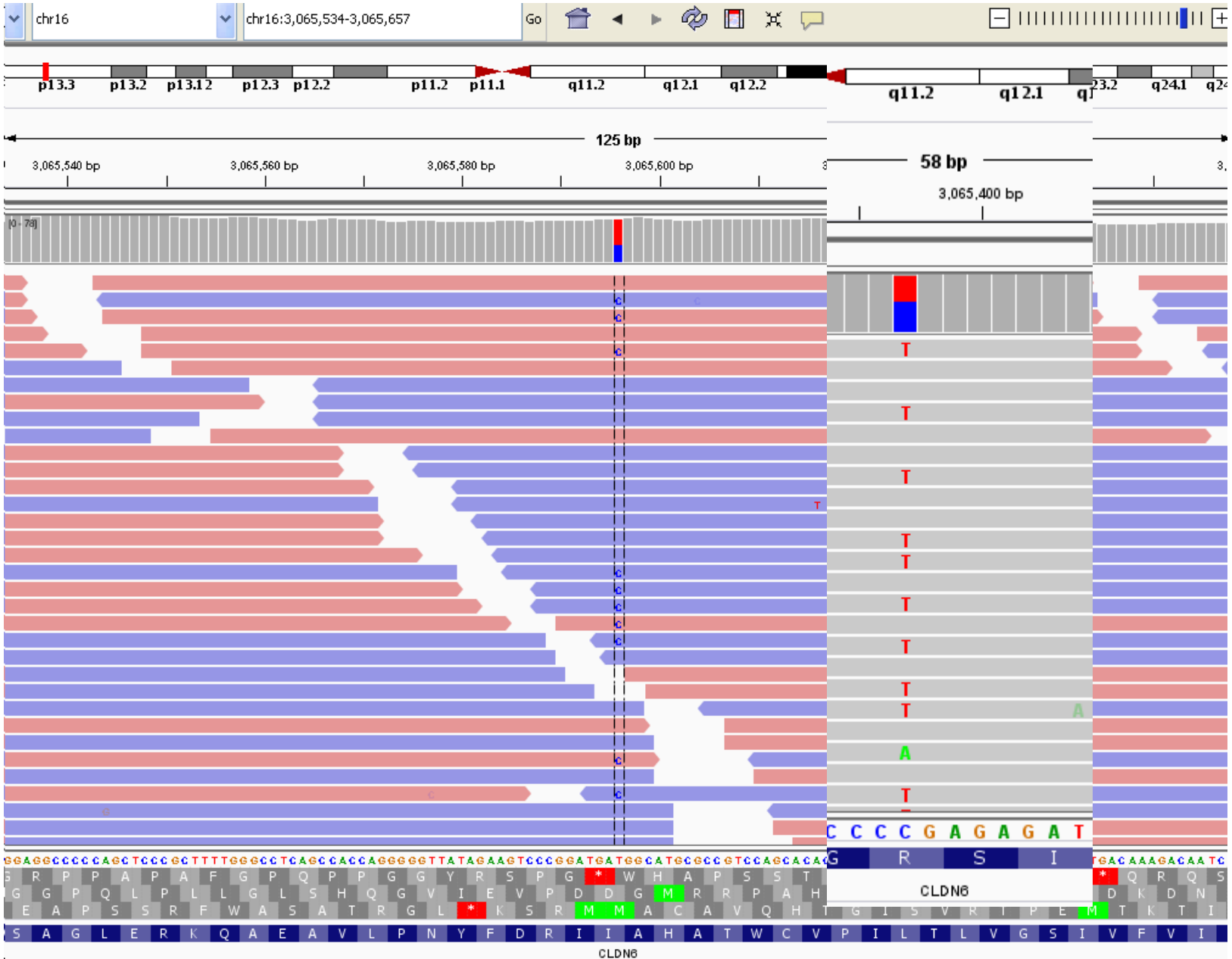
Genes absents

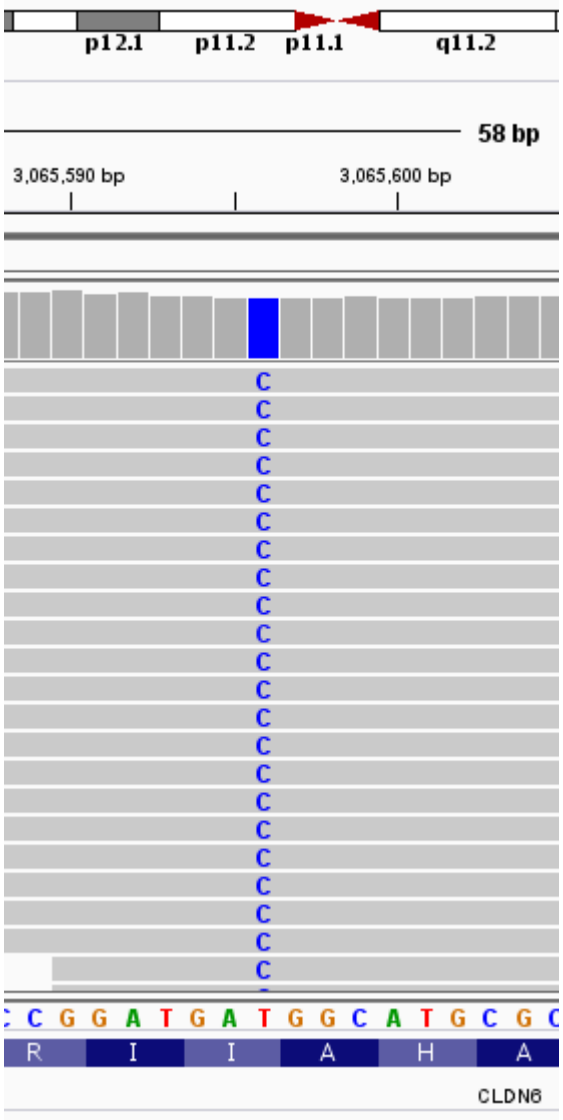
```
 wc list_genes_CDS_absent.txt
    900     list_genes_CDS_absent.txt                    (sur un seul sample)
```

B00G39Y OCLN C>G P24A chr5:68804987



B00G39Y CLDN6 chr16:3076596 T>C 28/42




p12.1
p11.2
p11.1
q11.2
58 bp
3,065,590 bp
3,065,600 bp
C G G A T G A T G G C A T G C G C
R I I A H A
CLDN6

Visualisation d'une mutation du gène PI3KCA sur un couple tumeur/normal

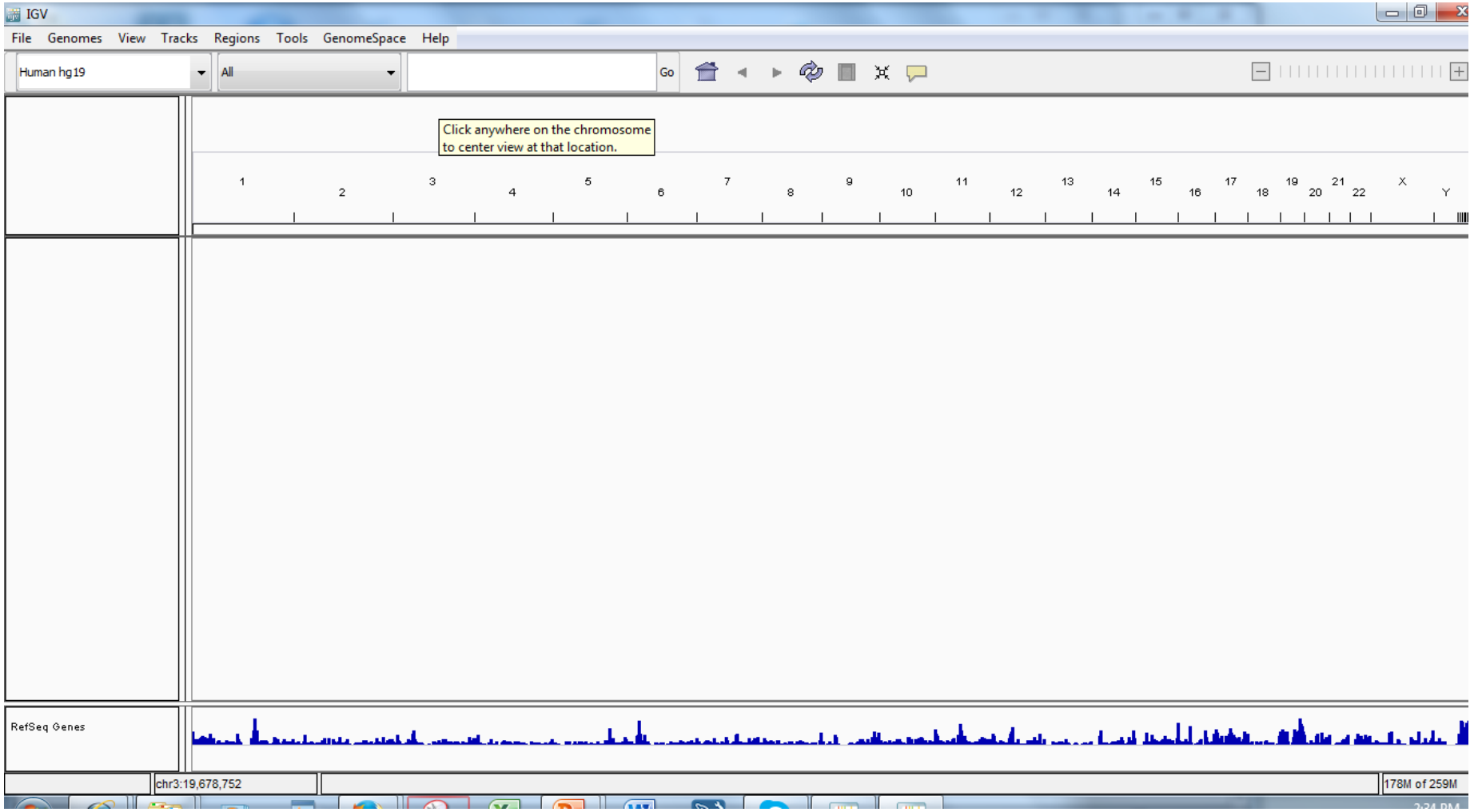
Sur IGV

Fichiers :
control.bam et tumor.bam

Sequencage ciblé autour du gène PIK3CA (chr3:178Mb)

Menu genome : choisir human hg19

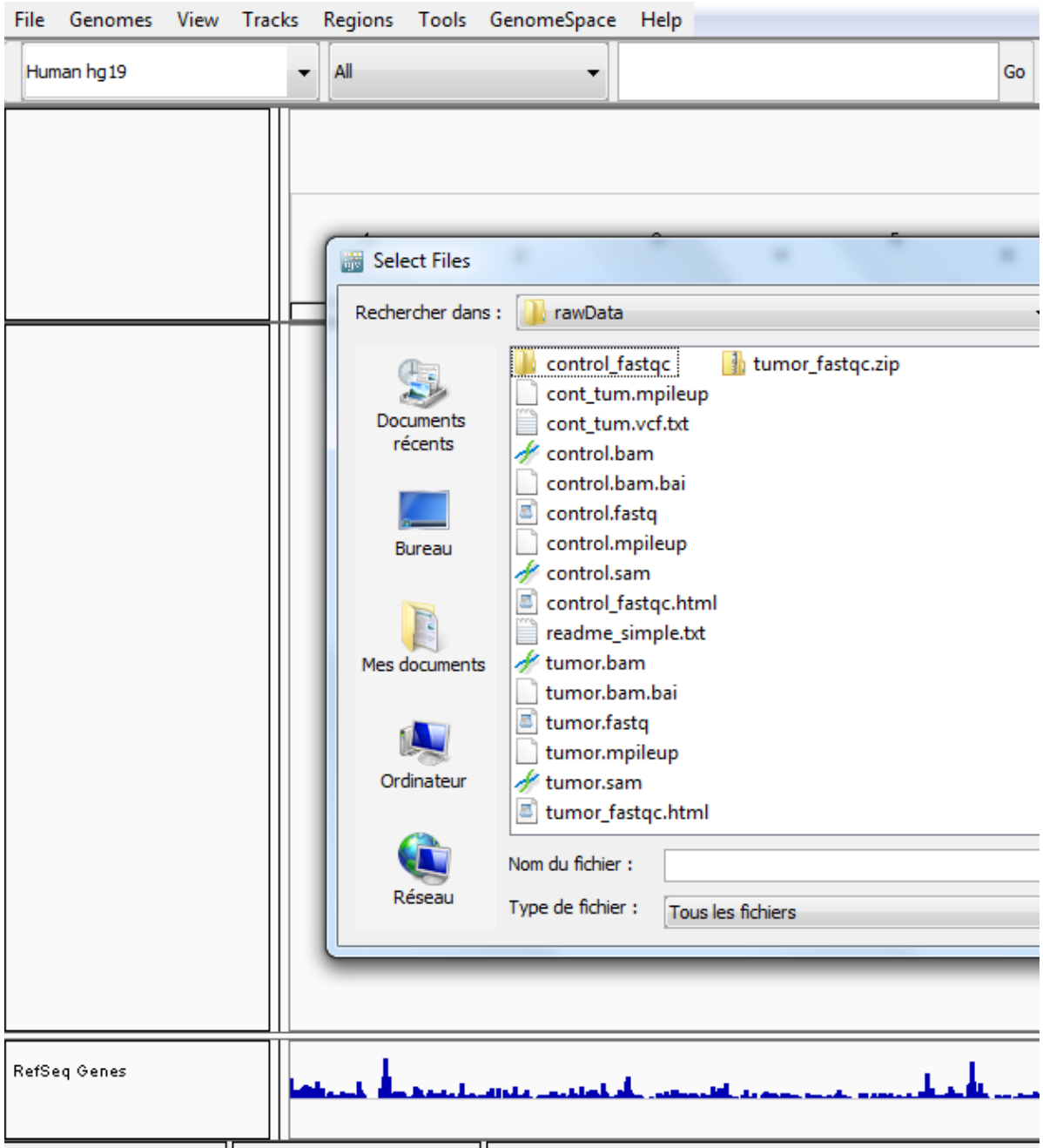
Puis les fichiers bam (indexés en bai)



Piste de référence : RefSeq sur hg19 : refGene.gff

NM_006218	PIK3CA	mRNA	chr3	178866311	178952497	.	+	.	21
NM_006218	PIK3CA	exon	chr3	178866311	178866391	ex_01	+	-1	.
NM_006218	PIK3CA	exon	chr3	178916538	178916965	ex_02	+	0	.
NM_006218	PIK3CA	exon	chr3	178917478	178917687	ex_03	+	1	.
NM_006218	PIK3CA	exon	chr3	178919078	178919328	ex_04	+	1	.
NM_006218	PIK3CA	exon	chr3	178921332	178921577	ex_05	+	0	.
NM_006218	PIK3CA	exon	chr3	178922291	178922376	ex_06	+	0	.
NM_006218	PIK3CA	exon	chr3	178927383	178927488	ex_07	+	2	.
NM_006218	PIK3CA	exon	chr3	178927974	178928126	ex_08	+	0	.
NM_006218	PIK3CA	exon	chr3	178928219	178928353	ex_09	+	0	.
NM_006218	PIK3CA	exon	chr3	178935998	178936122	ex_10	+	0	.
NM_006218	PIK3CA	exon	chr3	178936984	178937065	ex_11	+	2	.
NM_006218	PIK3CA	exon	chr3	178937359	178937523	ex_12	+	0	.
NM_006218	PIK3CA	exon	chr3	178937737	178937840	ex_13	+	0	.
NM_006218	PIK3CA	exon	chr3	178938774	178938945	ex_14	+	2	.
NM_006218	PIK3CA	exon	chr3	178941869	178941975	ex_15	+	0	.
NM_006218	PIK3CA	exon	chr3	178942488	178942609	ex_16	+	2	.
NM_006218	PIK3CA	exon	chr3	178943750	178943828	ex_17	+	1	.
NM_006218	PIK3CA	exon	chr3	178947060	178947230	ex_18	+	2	.
NM_006218	PIK3CA	exon	chr3	178947792	178947909	ex_19	+	2	.
NM_006218	PIK3CA	exon	chr3	178948013	178948164	ex_20	+	0	.
NM_006218	PIK3CA	exon	chr3	178951882	178952497	ex_21	+	2	.

Charger les fichiers .bam control et tumor

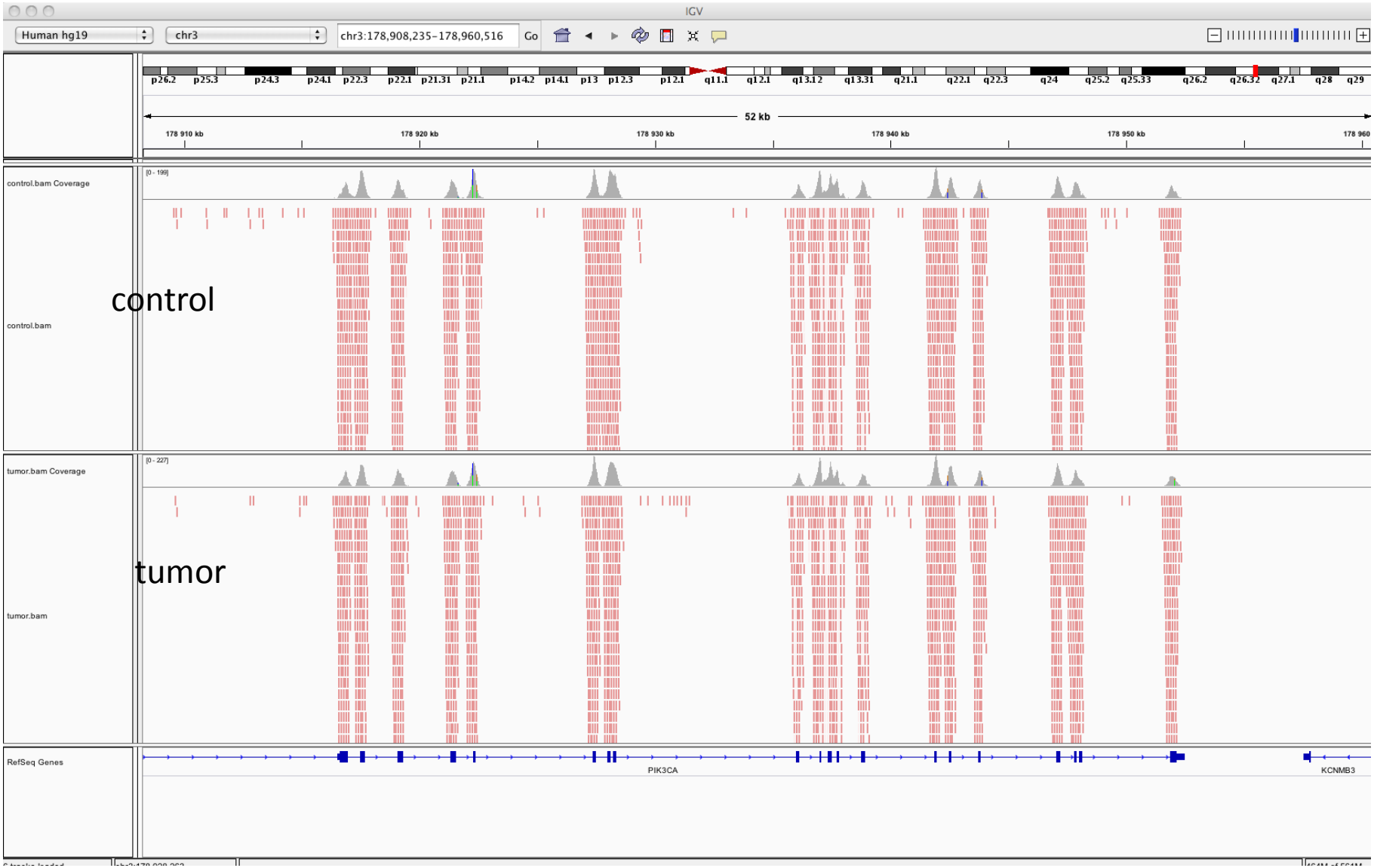


Regarder les menus
Utiliser le clic droit pour
Les options de visualisation

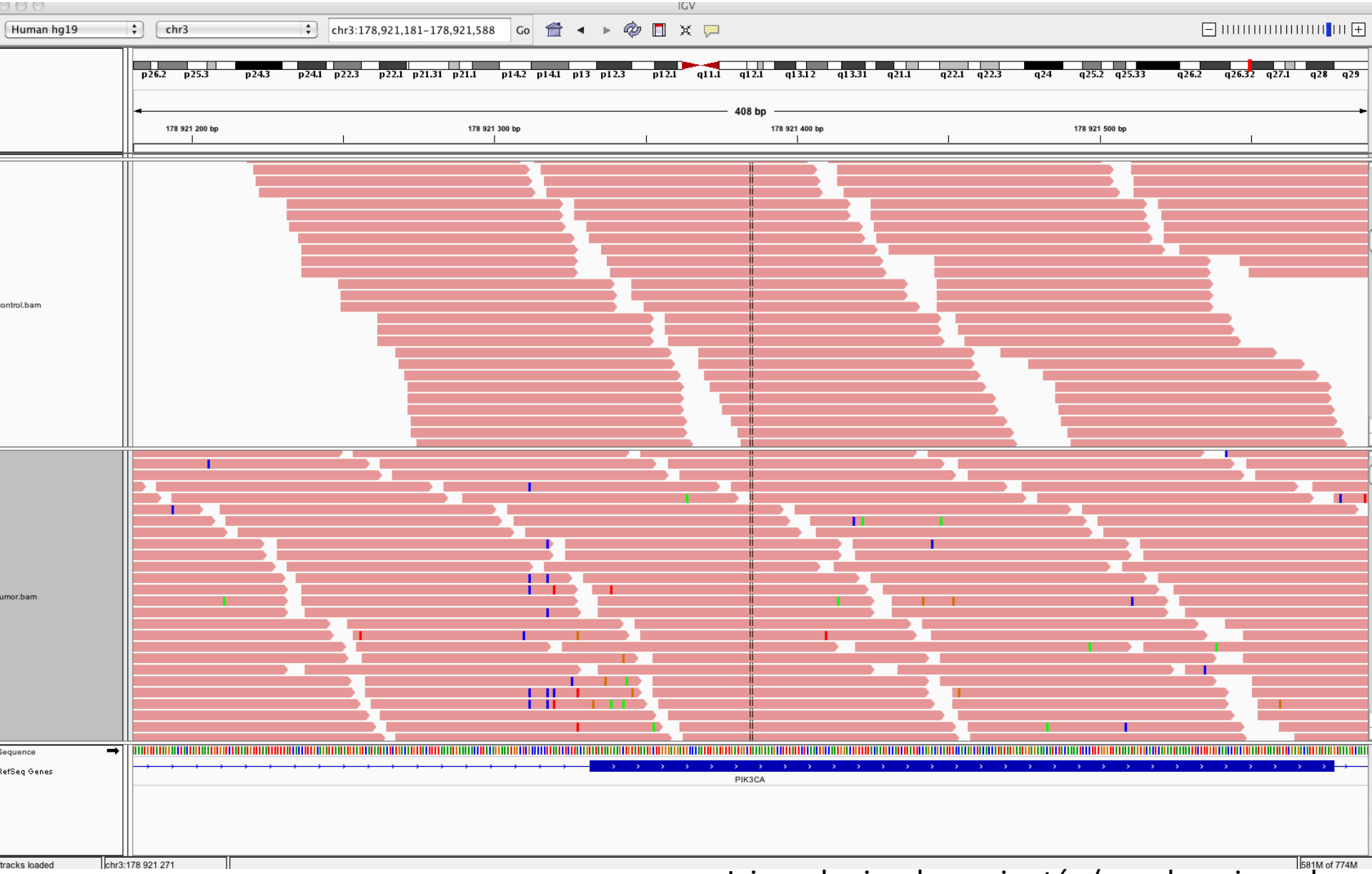
Déplacement des corrdonnées
Zoom

Les reads ne s affichent que si
Le zoom est suffisant

IGV 2.3 (hg19)

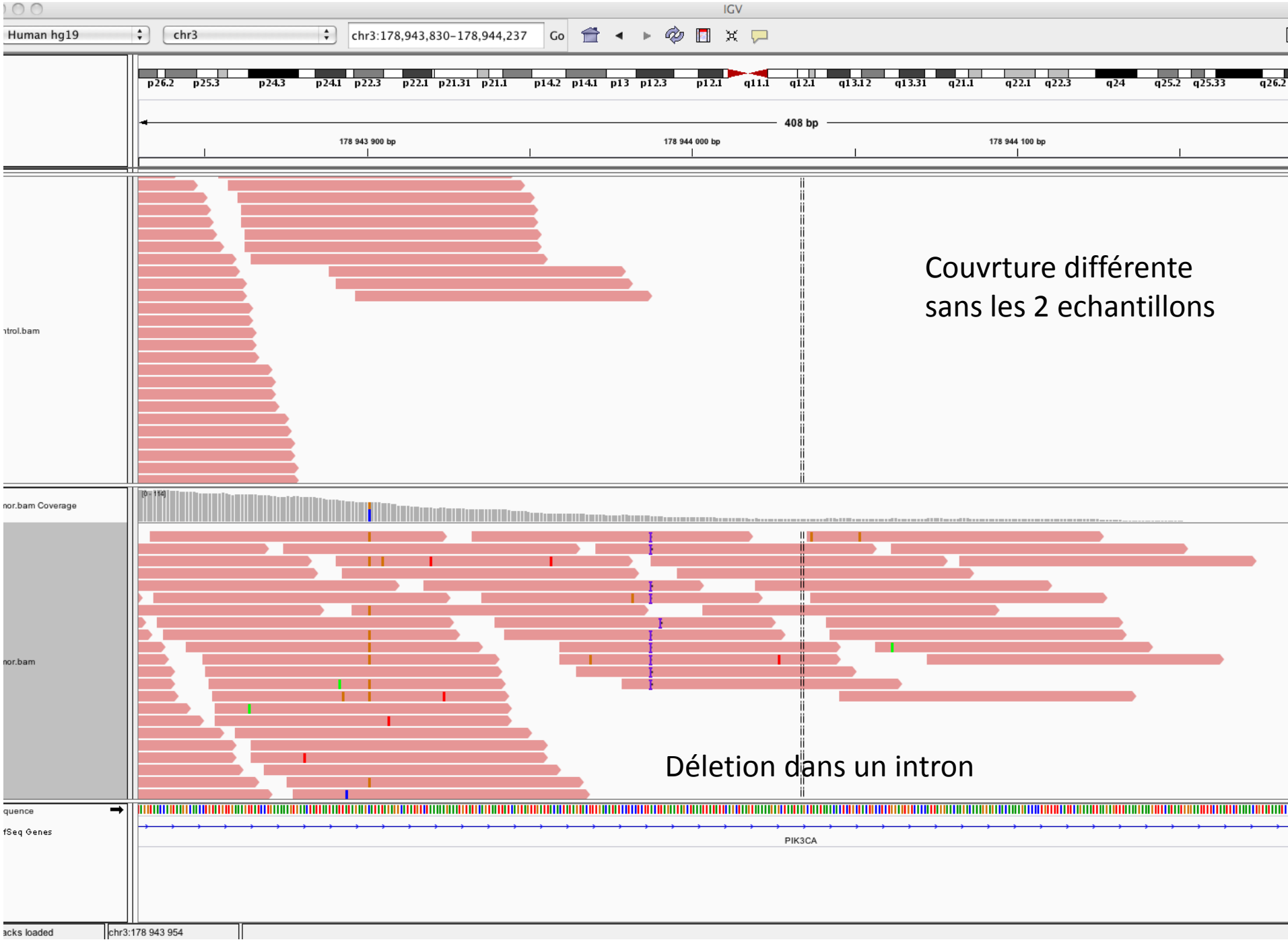


Exons du gène PIK3CA chr3:178900000

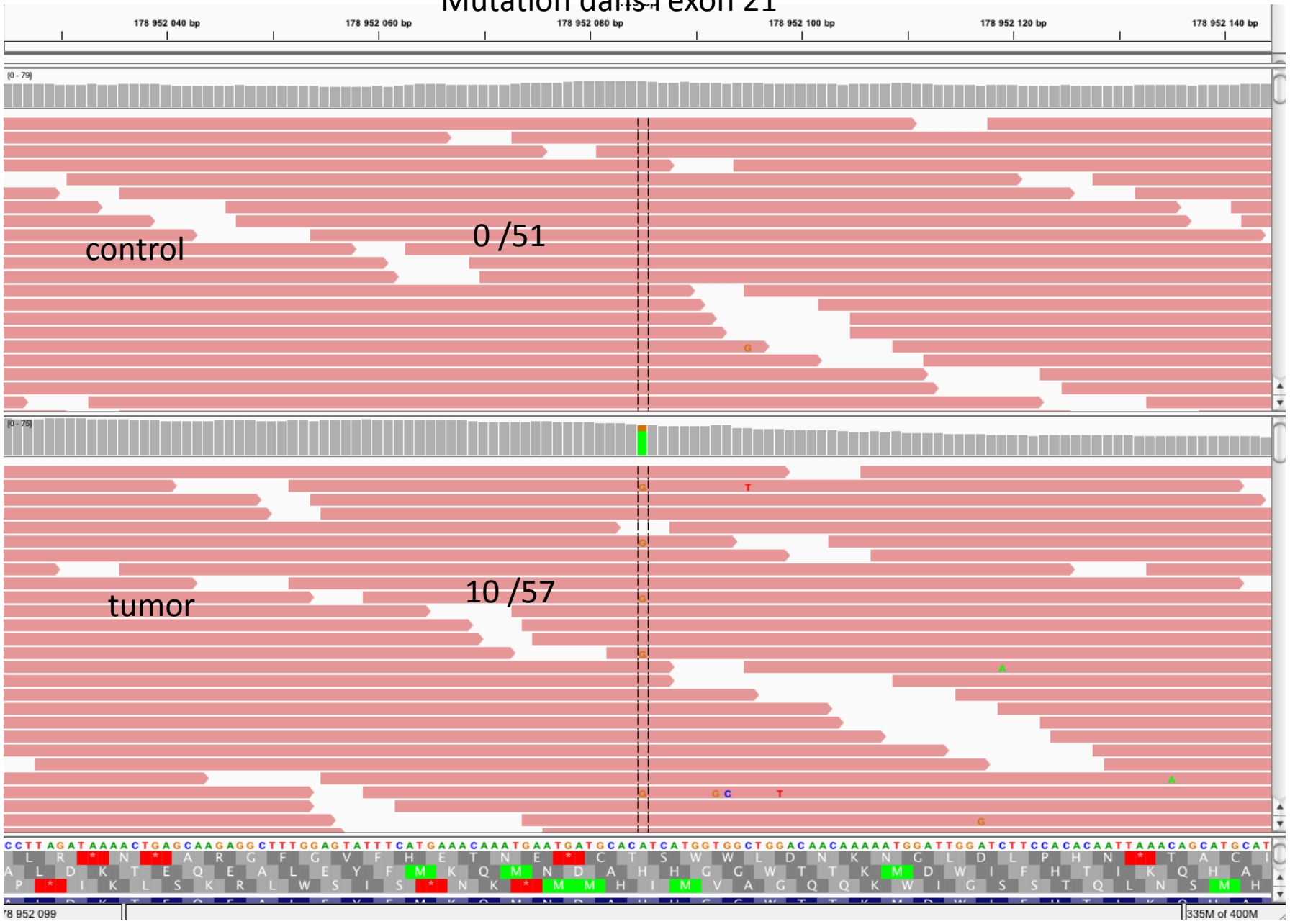


En 5' de l exon 5

Ici reads simples orientés (par de pair-end : 1 seule couleur rouge)


IGV
Human hg19
chr3
chr3:178,943,830-178,944,237
Go
p26.2 p25.3 p24.3 p24.1 p22.3 p22.1 p21.31 p21.1 p14.2 p14.1 p13 p12.3 p12.1 q11.1 q12.1 q13.12 q13.31 q21.1 q22.1 q22.3 q24 q25.2 q25.33 q26.2
408 bp
178 943 900 bp
178 944 000 bp
178 944 100 bp
ntrol.bam
Couvrture différente
sans les 2 echantillons
nor.bam Coverage
[0 - 114]
nor.bam
Déletion dans un intron
quence
fSeq Genes
PIK3CA
acks loaded
chr3:178 943 954

Mutation dans l'exon 21



Chr3: 178952085 Mutation A > G (CAT > CGT) H 1047 R

Test statistique de Fisher
pour évaluer la pertinence de la mutation

Utilisation de la fonction fisher.test de R

Test exact de Fisher

Le Test exact de Fisher est un test statistique utilisé pour l'analyse des tables de contingence. Ce test est utilisé en général avec des faibles effectifs mais il est valide pour toutes les tailles d'échantillon. Il doit son nom à son inventeur Ronald Fisher. C'est un test qualifié d'exact car les probabilités peuvent être calculées exactement plutôt qu'en s'appuyant sur une approximation qui ne devient correcte qu'asymptotiquement comme pour le test du χ^2 utilisé dans les tables de contingence.

Les calculs à la main ne sont raisonnables que pour les tables 2 × 2 mais le principe du test peut s'étendre au cas général et certains logiciels de statistique permettent le calcul pour le cas général.

Exemple [modifier | modifier le code]

Soit un échantillon d'adolescents on sépare l'échantillon entre filles et garçons et entre ceux qui suivent un régime et ceux qui n'en suivent pas et nous supposons que la proportion de filles qui suivent un régime est supérieure à celle des garçons, et nous voulons tester si la différence de proportions observées est significative. Voici les données

	Garçons	**Filles**	*total ligne*
régime	**1**	**9**	*10*
non régime	**11**	**3**	*14*
total colonne	*12*	*12*	*24*

Ces données ne sont pas adaptées pour une analyse par un test du χ^2, parce que les valeurs attendues (théoriques) dans la table sont inférieures à 10, et dans une table de contingence 2 × 2, le nombre de degrés de liberté est toujours égal à 1.

La question que l'on se pose à propos de ces données est : sachant que 10 de ces 24 adolescents pratiquent un régime et que 12 sont des filles, quelle est la probabilité que ces 10 qui pratiquent un régime soient répartis de manière équilibrée entre les filles et les garçons ? Si on choisit 10 adolescents au hasard, quelle est la probabilité que 9 d'entre eux soient parmi les 12 filles et seulement 1 parmi les 12 garçons ?

Avant de passer au test de Fisher nous introduisons quelques notations. On représente les cellules par les lettres *a, b, c* et *d* et on note n le total général. La table se présente ainsi :

	Garçons	**Filles**	*Total*
Régime	*a*	*b*	*a + b*
Non-régime	*c*	*d*	*c + d*
Totaux	*a + c*	*b + d*	*a + b + c + d (=n)*

Fisher a montré que la probabilité d'obtenir un tel ensemble de valeurs était donnée par la loi hypergéometrique :

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$$

où $\binom{n}{k}$ est le coefficient binomial et le symbole ! indique la factorielle.

Dans l'exemple la probabilité d'obtenir le tableau croisé observé, avec les totaux marginaux donnés, est donc:

$$p = \frac{10!14!12!12!}{1!9!11!3!24!} = \frac{40}{29716} = 0,001346...$$

De manière à calculer si des données observées sont significativement éloignées de l'indépendance, c'est-à-dire la probabilité d'observer des données aussi ou plus éloignées que celles observées si l'hypothèse nulle (indépendance) est satisfaite, il faut calculer les valeurs de *p* pour ces tables et les ajouter. Cela donne un test unilatéral; pour un test bilatéral on doit considérer les tables qui sont extrêmes mais dans l'autre direction. Malheureusement classer les tables pour savoir si elles sont ou non aussi extrêmes est problématique. Une approche utilisée dans R avec la fonction "fisher.test" calcule la valeur p en sommant les probabilités de toutes les tables ayant une probabilité inférieure ou égale à celle de la table observée.

Voici la commande permettant d'entrer le tableau croisé de l'exemple présenté plus haut comme une matrice de R et de calculer la valeur de la probabilité.

fisher.test(matrix(c(1,9,11,3),nrow=2))

Et voici la valeur de la probabilité obtenue:

p-value = 0.002759

Cette valeur se calcule aussi assez facilement sans logiciel puissant avec le principe exposé. Il y a en tout quatre tables aussi éloignées ou plus éloignées de l'indépendance que la table observée. Il n'y a en tout que onze tables possibles sachant qu'il y a seulement dix adolescents qui suivent un régime, le nombre de garçons suivant un régime peut varier de 0 jusqu'à 10. Une fois que le nombre de garçons suivant un régime est connu tout le reste de la table est connu, ce qui correspond intuitivement au degré de liberté valant 1. Les quatre tables extrêmes correspondent aux valeurs 0,1,9 et 10 pour le nombre de garçons suivant un régime. La probabilité des deux tables plus éloignées de l'indépendance est très faible elle se calcule comme précédemment:

$$p = \frac{10!14!12!12!}{0!10!12!2!24!} = \frac{1}{29716} = 0,0000336519...$$

En ajoutant les probabilités des quatre tables on trouve $p = \frac{82}{29716} = 0,0027594561...$

Le test permet de rejeter l'indépendance entre le sexe et le fait de faire un régime.

Fisher's Exact Test for Count Data

Usage

```
fisher.test(x, y = NULL, workspace = 200000, hybrid = FALSE,
        control = list(), or = 1, alternative = "two.sided",
        conf.int = TRUE, conf.level = 0.95,
        simulate.p.value = FALSE, B = 2000)
```

Arguments

x	either a two-dimensional contingency table in matrix form, or a factor object.
y	a factor object; ignored if x is a matrix.
workspace	an integer specifying the size of the workspace used in the network algorithm. In units of 4 bytes. Only used for non-simulated p-values larger than *2 by 2* tables.
hybrid	a logical. Only used for larger than *2 by 2* tables, in which cases it indicated whether the exact probabilities (default) or a hybrid approximation thereof should be computed. See 'Details'.
control	a list with named components for low level algorithm control. At present the only one used is "mult", a positive integer ≥ *2* with default 30 used only for larger than *2 by 2* tables. This says how many times as much space should be allocated to paths as to keys: see file 'fexact.c' in the sources of this package.
or	the hypothesized odds ratio. Only used in the *2 by 2* case.
alternative	indicates the alternative hypothesis and must be one of "two.sided", "greater" or "less". You can specify just the initial letter. Only used in the *2 by 2* case.
conf.int	logical indicating if a confidence interval should be computed (and returned).
conf.level	confidence level for the returned confidence interval. Only used in the *2 by 2* case if conf.int = TRUE.
simulate.p.value	a logical indicating whether to compute p-values by Monte Carlo simulation, in larger than *2 by 2* tables
B	an integer specifying the number of replicates used in the Monte Carlo test.

A list with class "htest" containing the following components:

p.value	the p-value of the test.
conf.int	a confidence interval for the odds ratio. Only present in the *2 by 2* case if argument conf.int = TRUE.
estimate	an estimate of the odds ratio. Note that the *conditional* Maximum Likelihood Estimate (MLE) rather than the unconditional MLE (the sample odds ratio) is used. Only present in the *2 by 2* case.
null.value	the odds ratio under the null, or. Only present in the *2 by 2* case.
alternative	a character string describing the alternative hypothesis.
method	the character string "Fisher's Exact Test for Count Data".
data.name	a character string giving the names of the data.

	control	tumor	
A	a	b	a+b
G	c	d	c+d
Total	a+c	b+d	a+b+c+d
	control	tumor	
A	51	47	98
G	0	10	10
Total	51	57	108

```
> m <- matrix(c(51,47,0,10), byrow=T,
nrow = 2,  dimnames = list(c("A","G"),
c("Control", "Tumor")))
> m
  Control Tumor
A      51     47
G       0     10

> fisher.test(m)$p.value
[1] 0.001445162
```

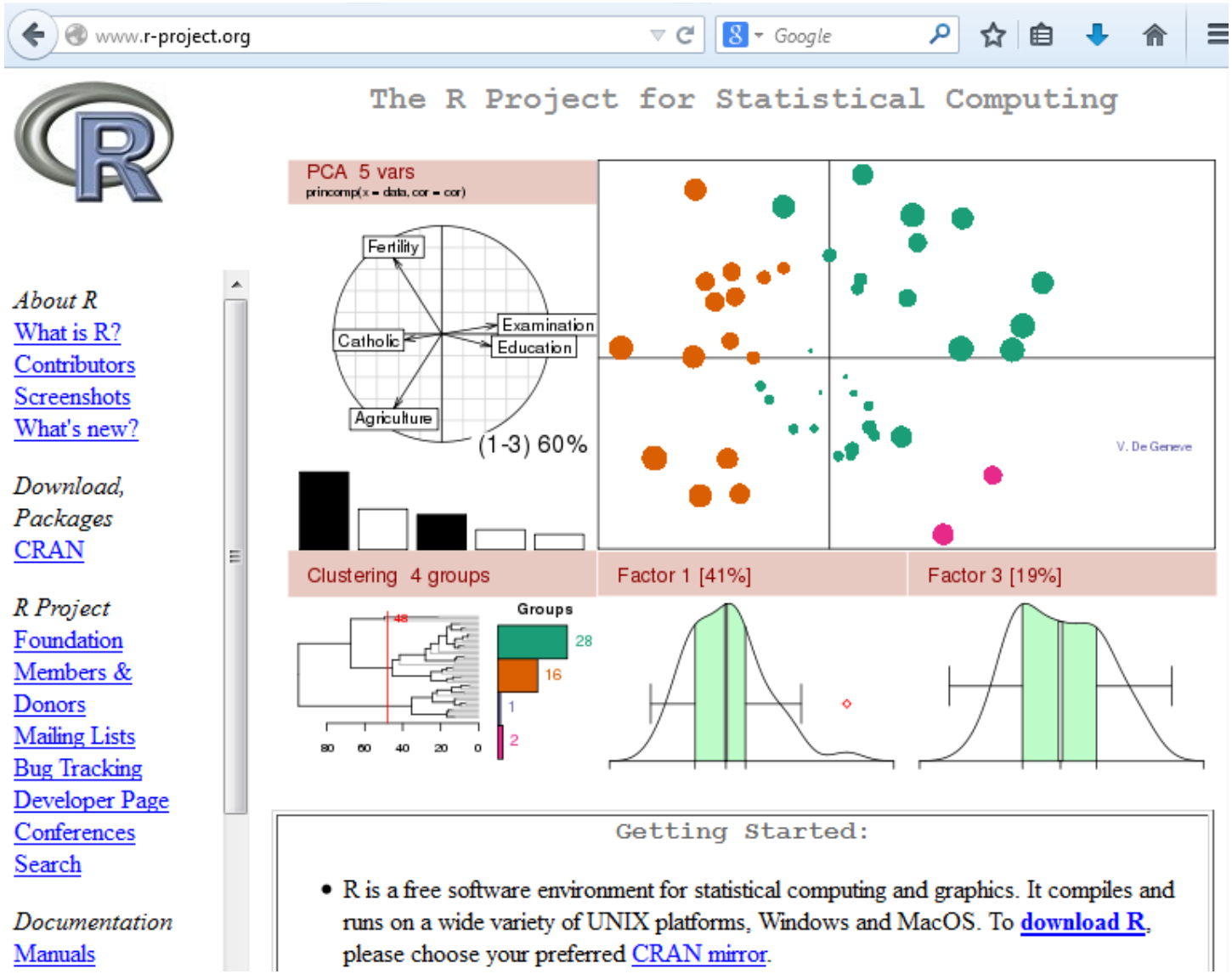
ou plus simplement (en une seule ligne) :

```
fisher.test(matrix(c(51,47,0,10),byrow=T,nr
ow = 2))$p.value
```

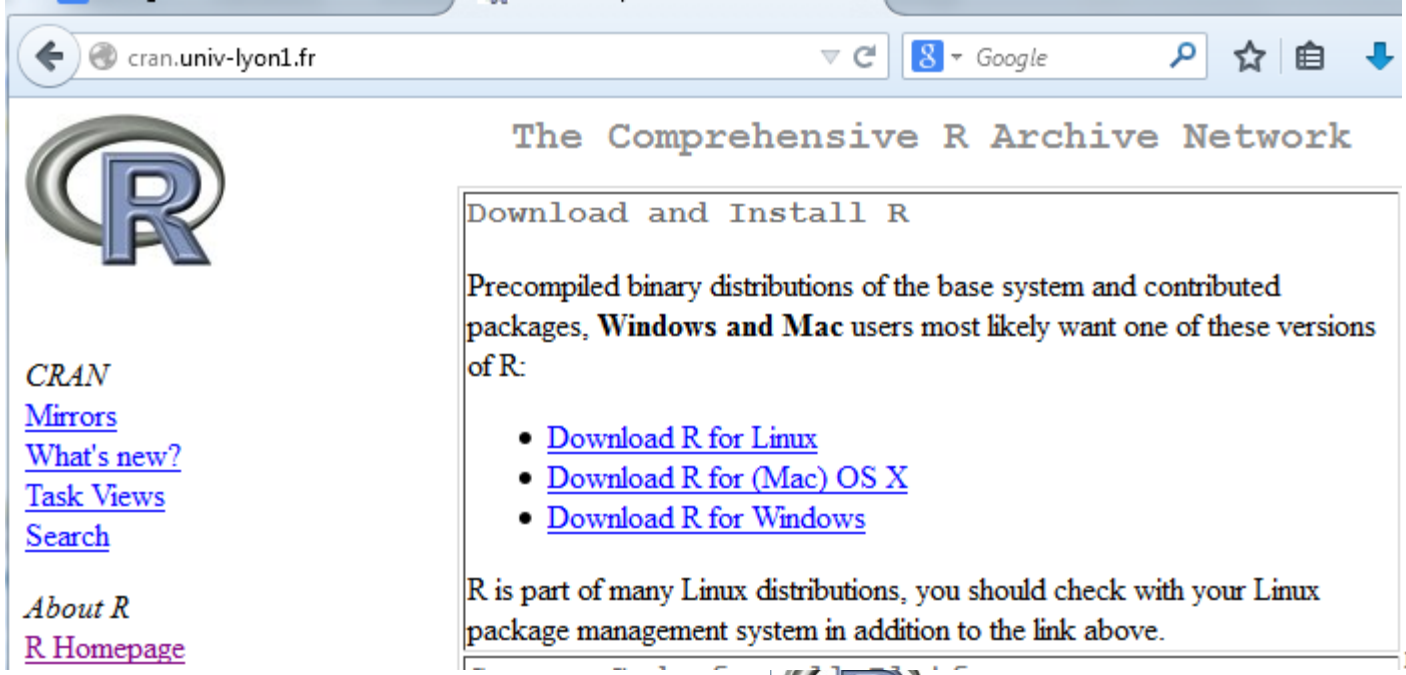
La probabilité de la mutation A>G est dans ces conditions de 1.4 10-3
G n'est pas présent dans le control. Il apparaît à 10/57 dans la tumeur
La mutation n affecte pas 100% des reads .
Probablement mutation hétérozygote … et % cellules tumorales < 100%

En général on filtrera les variants entre normal et tumoral à > 0.01.

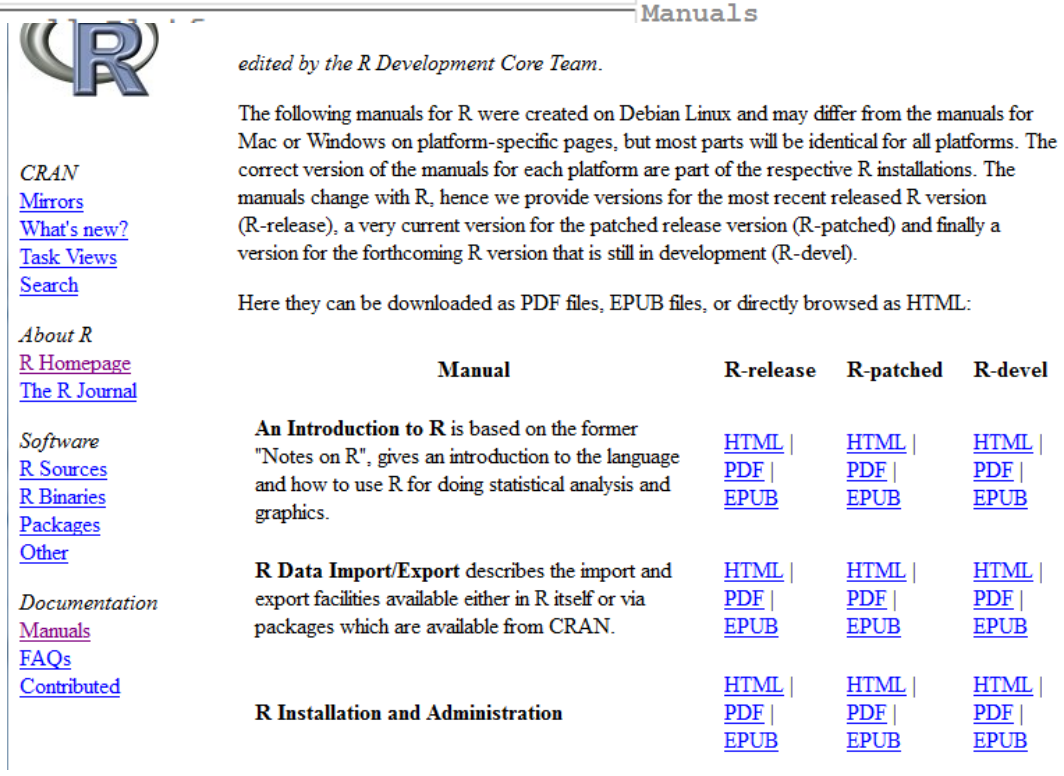
Digression sur le logiciel R



Logiciel (sur PC, Mac Unix) d analyse statistique et graphique



https://pbil.univ-lyon1.fr/CRAN/



"R for Beginners" by Emmanuel Paradis (PDF).
http://cran.univ-lyon1.fr/doc/contrib/Paradis-rdebuts_en.pdf

"Using R for Data Analysis and Graphics - Introduction, Examples and Commentary" by John Maindonald
http://cran.univ-lyon1.fr/doc/contrib/usingR.pdf

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

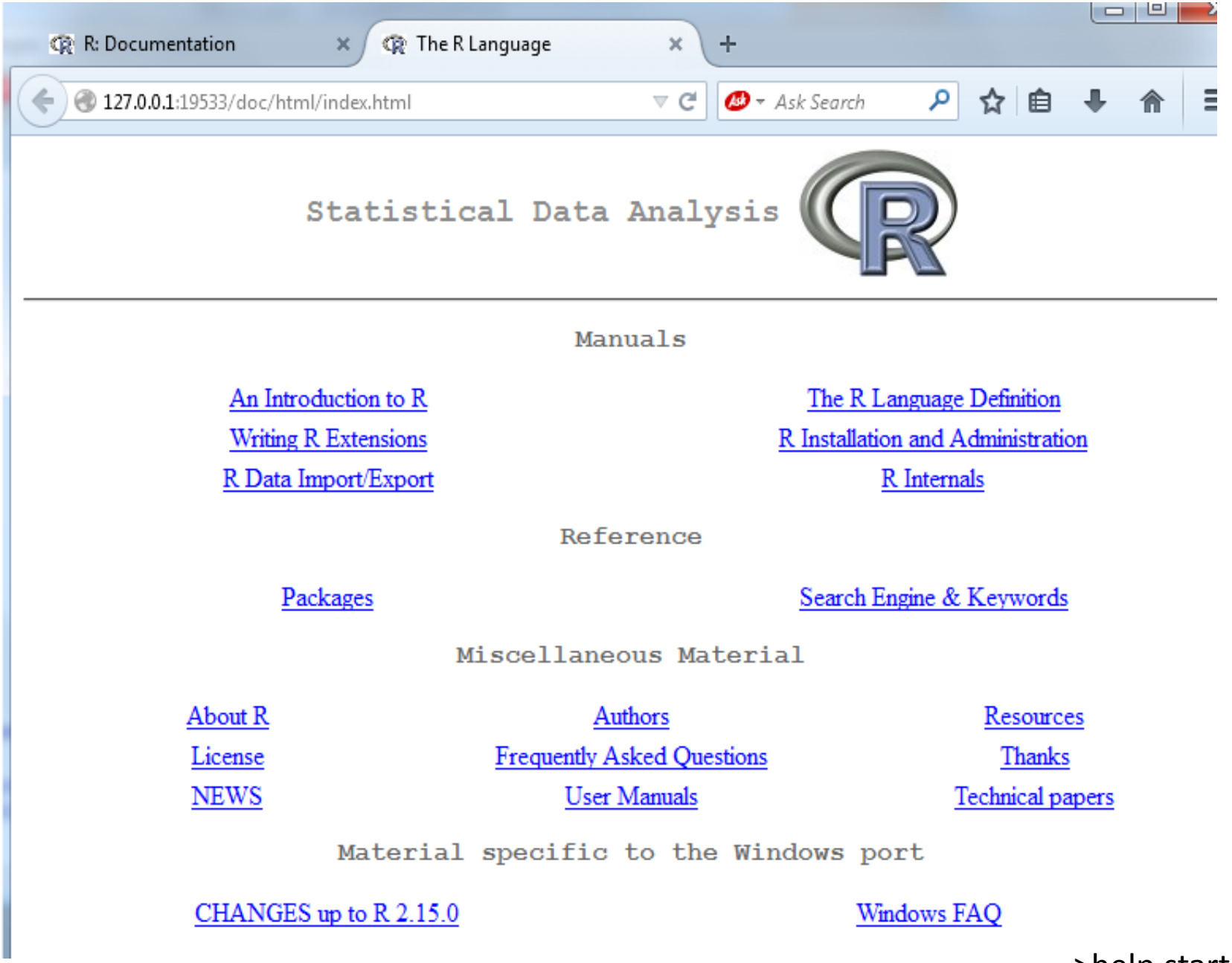
```
> 2+2
[1] 4
> x <- 10
> x
[1] 10
> x <- rep(1:10,3)
> mean(x)
[1] 5.5
> sd(x)
[1] 2.921384

> fisher.test(matrix(c(51,47,0,10), byrow=T, nrow = 2))$p.value
```

Ou

```
fisher.test(matrix(c(51,0,47,10), nrow = 2))$p.value
>help.start()

> help(plot)
```


R: Documentation
The R Language
127.0.0.1:19533/doc/html/index.html
Ask Search
Statistical Data Analysis
Manuals
An Introduction to R
The R Language Definition
Writing R Extensions
R Installation and Administration
R Data Import/Export
R Internals
Reference
Packages
Search Engine & Keywords
Miscellaneous Material
About R
Authors
Resources
License
Frequently Asked Questions
Thanks
NEWS
User Manuals
Technical papers
Material specific to the Windows port
CHANGES up to R 2.15.0
Windows FAQ

>help.start()

Caractérisation de la mutation H 1047 R dans la PIK3CA dans les bases de données

COSMIC

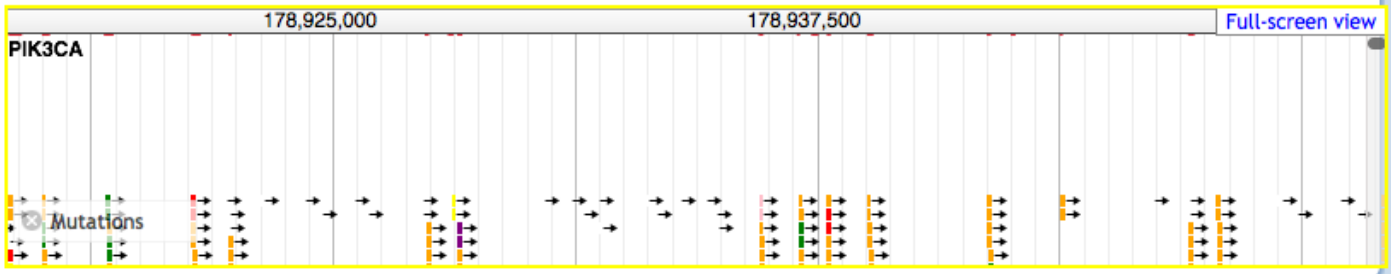
Catalogue of somatic mutations in cancer

Cosmic » Gene » Overview » PIK3CA

Overview | Sequence | Fusions | Studies | References

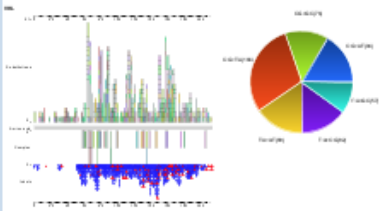
Gene name	PIK3CA
Genomic Coordinates	3:178916614..178952152
Gene Id	COSG14
Synonyms	MGC142161,MGC142163,PI3K,p110-alpha,P42336,ENSG00000121879
Drug Sensitivity Data:	Mutations in PIK3CA are associated with altered sensitivity to the following drug(s) : AZD6482, Bexarotene, AKT inhibitor VIII, WO2009093972, AMG-706, CI-1040, PAC-1, WZ-1-84, Lapatinib, AZD6482, AKT inhibitor VIII, AZD6482, Methotrexate, VX-702, PAC-1 Click here for PIK3CA all drug sensitivity data
No.of Samples	Total number of unique samples: 72354 Unique samples with mutations: 7418
Alternative Transcripts	PIK3CA_ENST00000263967

Cosmic Genome Browser



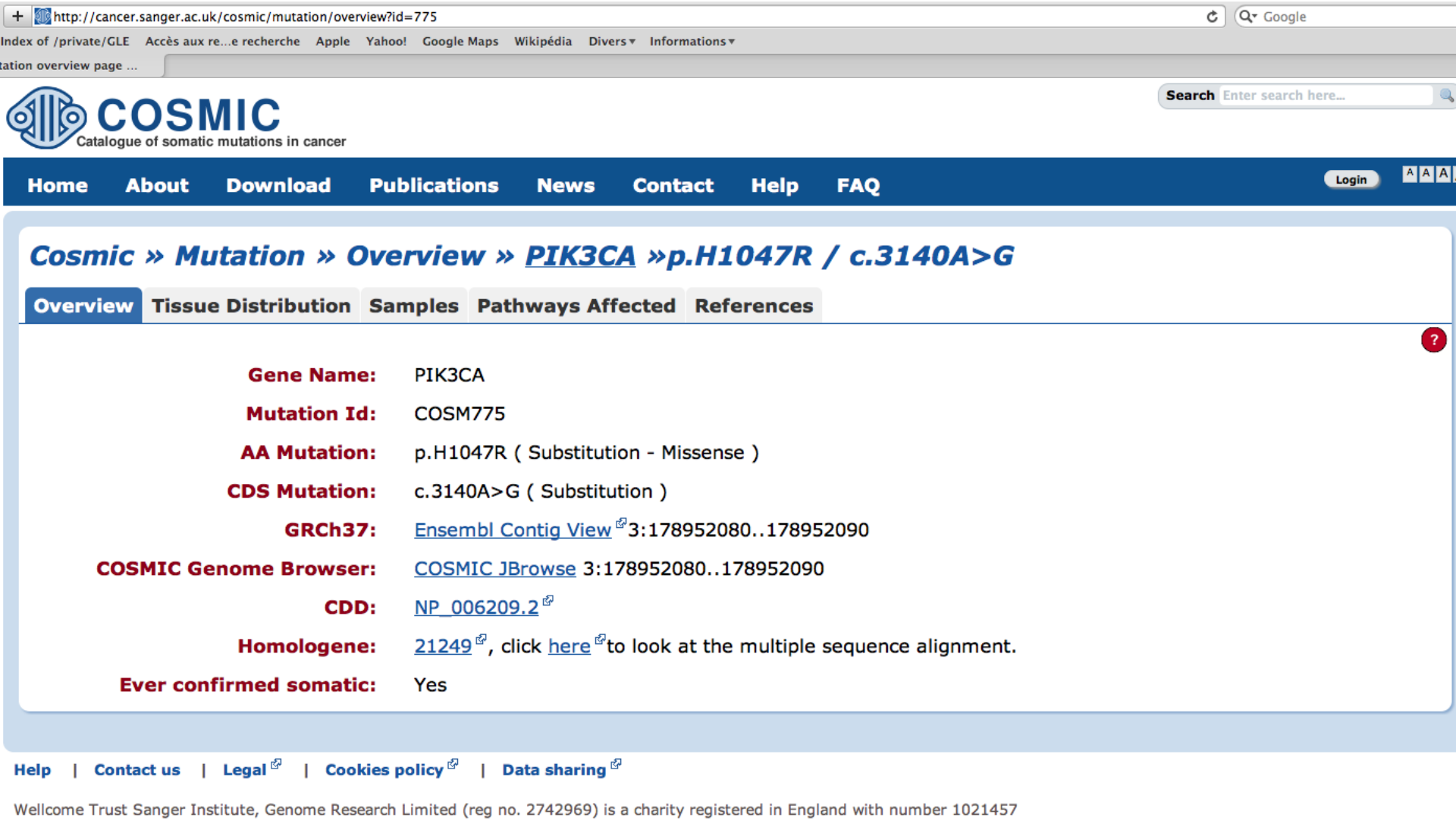
Mutation Analysis

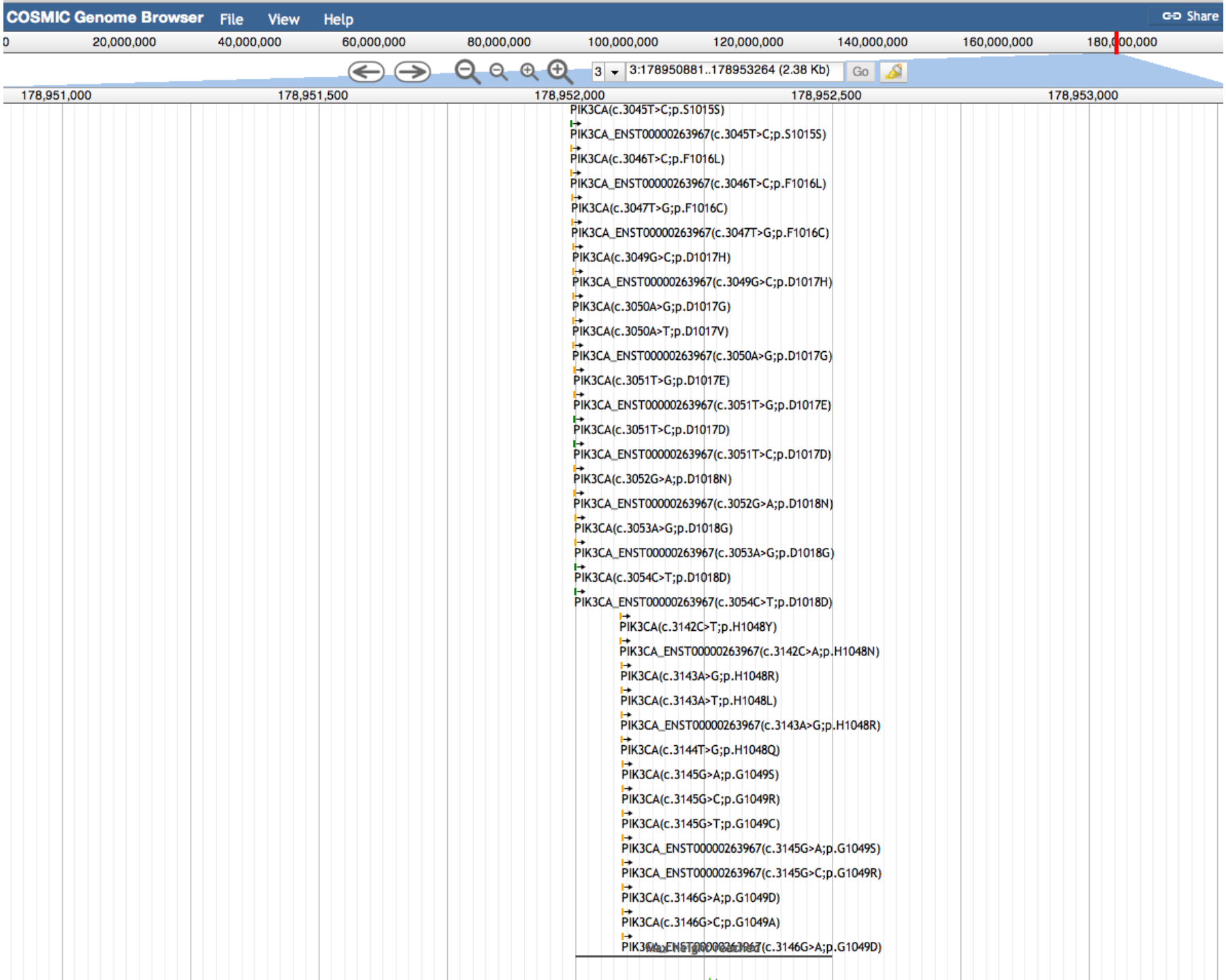
Histogram | Distribution



External Links

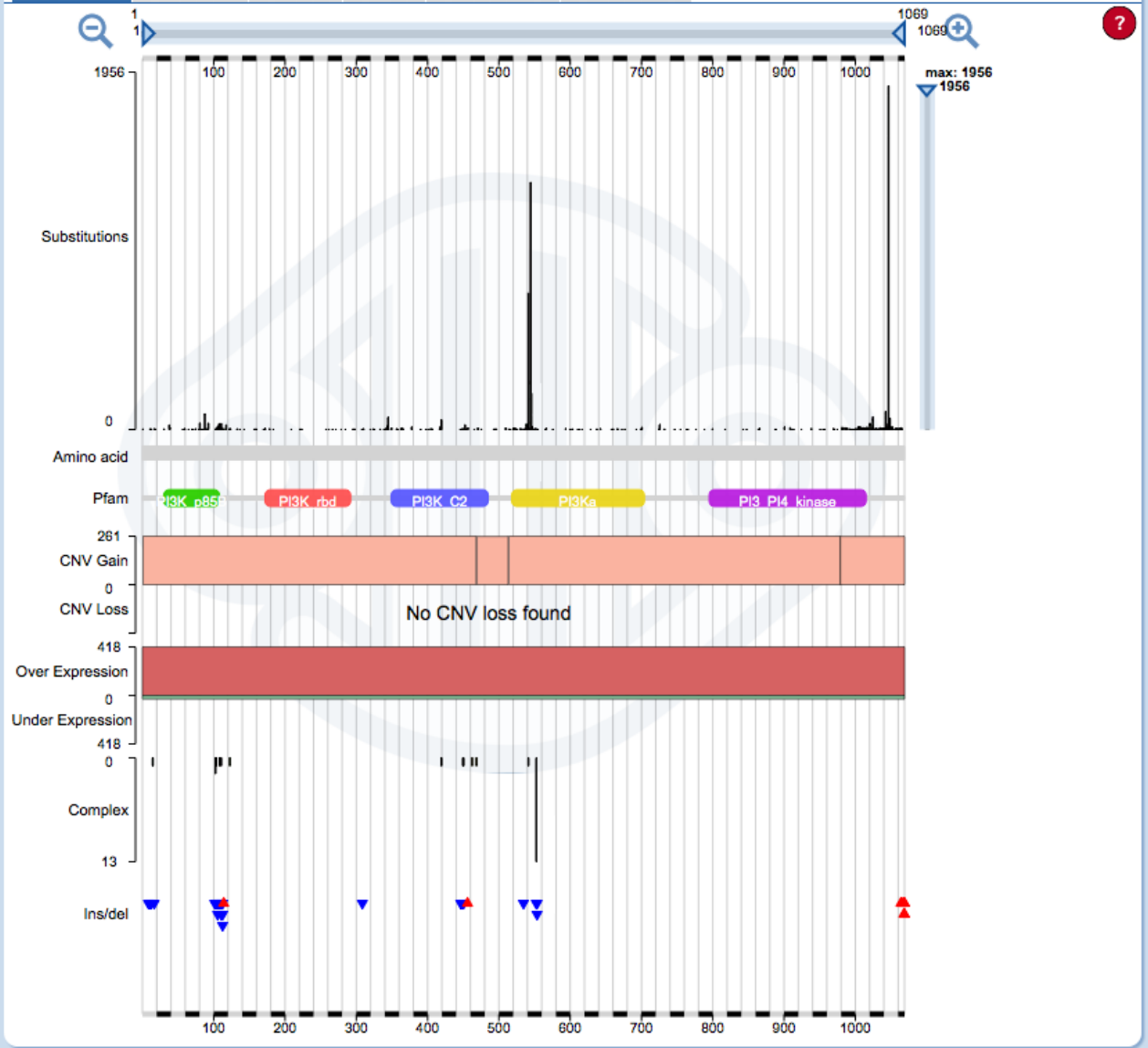
OMIM:	171834
Transcript:	NM_006218.1
Ensembl Contig View:	PIK3CA
UCSC Browser:	PIK3CA
Copy Number:	CONAN
Pfam:	P42336
Atlas Genetic Oncology:	GC_PIK3CA
HGNC:	8975


http://cancer.sanger.ac.uk/cosmic/mutation/overview?id=775
Google
Index of /private/GLE
Accès aux re...e recherche
Apple
Yahoo!
Google Maps
Wikipédia
Divers
Informations
COSMIC
Catalogue of somatic mutations in cancer
Search
Enter search here...
Home
About
Download
Publications
News
Contact
Help
FAQ
Login
Cosmic » Mutation » Overview » PIK3CA »p.H1047R / c.3140A>G
Overview
Tissue Distribution
Samples
Pathways Affected
References
Gene Name: PIK3CA
Mutation Id: COSM775
AA Mutation: p.H1047R (Substitution - Missense)
CDS Mutation: c.3140A>G (Substitution)
GRCh37: Ensembl Contig View 3:178952080..178952090
COSMIC Genome Browser: COSMIC JBrowse 3:178952080..178952090
CDD: NP_006209.2
Homologene: 21249, click here to look at the multiple sequence alignment.
Ever confirmed somatic: Yes
Help | Contact us | Legal | Cookies policy | Data sharing
Wellcome Trust Sanger Institute, Genome Research Limited (reg no. 2742969) is a charity registered in England with number 1021457


COSMIC Genome Browser
File
View
Help
Share
0
20,000,000
40,000,000
60,000,000
80,000,000
100,000,000
120,000,000
140,000,000
160,000,000
180,000,000
3
3:178950881..178953264 (2.38 Kb)
Go
178,951,000
178,951,500
178,952,000
178,952,500
178,953,000
PIK3CA(c.3045T>C;p.S1015S)
PIK3CA_ENST00000263967(c.3045T>C;p.S1015S)
PIK3CA(c.3046T>C;p.F1016L)
PIK3CA_ENST00000263967(c.3046T>C;p.F1016L)
PIK3CA(c.3047T>G;p.F1016C)
PIK3CA_ENST00000263967(c.3047T>G;p.F1016C)
PIK3CA(c.3049G>C;p.D1017H)
PIK3CA_ENST00000263967(c.3049G>C;p.D1017H)
PIK3CA(c.3050A>G;p.D1017G)
PIK3CA(c.3050A>T;p.D1017V)
PIK3CA_ENST00000263967(c.3050A>G;p.D1017G)
PIK3CA(c.3051T>G;p.D1017E)
PIK3CA_ENST00000263967(c.3051T>G;p.D1017E)
PIK3CA(c.3051T>C;p.D1017D)
PIK3CA_ENST00000263967(c.3051T>C;p.D1017D)
PIK3CA(c.3052G>A;p.D1018N)
PIK3CA_ENST00000263967(c.3052G>A;p.D1018N)
PIK3CA(c.3053A>G;p.D1018G)
PIK3CA_ENST00000263967(c.3053A>G;p.D1018G)
PIK3CA(c.3054C>T;p.D1018D)
PIK3CA_ENST00000263967(c.3054C>T;p.D1018D)
PIK3CA(c.3142C>T;p.H1048Y)
PIK3CA_ENST00000263967(c.3142C>A;p.H1048N)
PIK3CA(c.3143A>G;p.H1048R)
PIK3CA(c.3143A>T;p.H1048L)
PIK3CA_ENST00000263967(c.3143A>G;p.H1048R)
PIK3CA(c.3144T>G;p.H1048Q)
PIK3CA(c.3145G>A;p.G1049S)
PIK3CA(c.3145G>C;p.G1049R)
PIK3CA(c.3145G>T;p.G1049C)
PIK3CA_ENST00000263967(c.3145G>A;p.G1049S)
PIK3CA_ENST00000263967(c.3145G>C;p.G1049R)
PIK3CA(c.3146G>A;p.G1049D)
PIK3CA(c.3146G>C;p.G1049A)
PIK3CA_ENST00000263967(c.3146G>A;p.G1049D)

Cosmic » Gene » Analysis » PIK3CA

Histogram | Mutations | Fusions | Tissue | Distribution | CNV & Expr



Filters

Gene: PIK3CA

Position
Start: 1
End: 1069

Sequence Type:
cDNA
Amino Acid

- Systematic Screen
- Somatic Status
- Tumour Source
- Mutation Type
- Copy Number Variation
- Gene Expression

Apply Reset

Cosmic » Gene » Analysis » PIK3CA

Histogram | Mutations | Fusions | Tissue | Distribution | CNV & Expr

Show 10 entries Search: Export: CSV TSV

Position (AA)	Mutation (CDS)	Mutation (Amino Acid)	Mutation ID (COSM)	Count	Mutation Type
1	c.1A>G	p.M1V	COSM1041443	4	Substitution - Missense
9	c.24_56del33	p.E9_R19del	COSM39613	1	Deletion - In frame
10	c.28_48del21	p.L10_M16del	COSM35324	1	Deletion - In frame
11	c.30_53del24	p.W11_P18delWGIHLMPP	COSM1041445	1	Deletion - In frame
11	c.32G>T	p.W11L	COSM582520	1	Substitution - Missense
12	c.35G>A	p.G12D	COSM27495	5	Substitution - Missense
13	c.36_56del21	p.I13_R19del	COSM166338	1	Deletion - In frame
13	c.38T>C	p.I13T	COSM1420724	1	Substitution - Missense
15	c.42_64>TCCAA	p.L15_V22>PM	COSM87211	1	Complex - deletion inframe
17	c.49C>T	p.P17S	COSM27494	1	Substitution - Missense

Showing 1 to 10 of 703 entries First Previous 1 2 3 4 5 Next Last

Filters

Gene PIK3CA

Position
Start 1
End 1069

Sequence Type:
cDNA
Amino Acid

Systematic Screen
Somatic Status
Tumour Source
Mutation Type
Copy Number Variation
Gene Expression

Apply Reset

UNIPROT : acc P42336

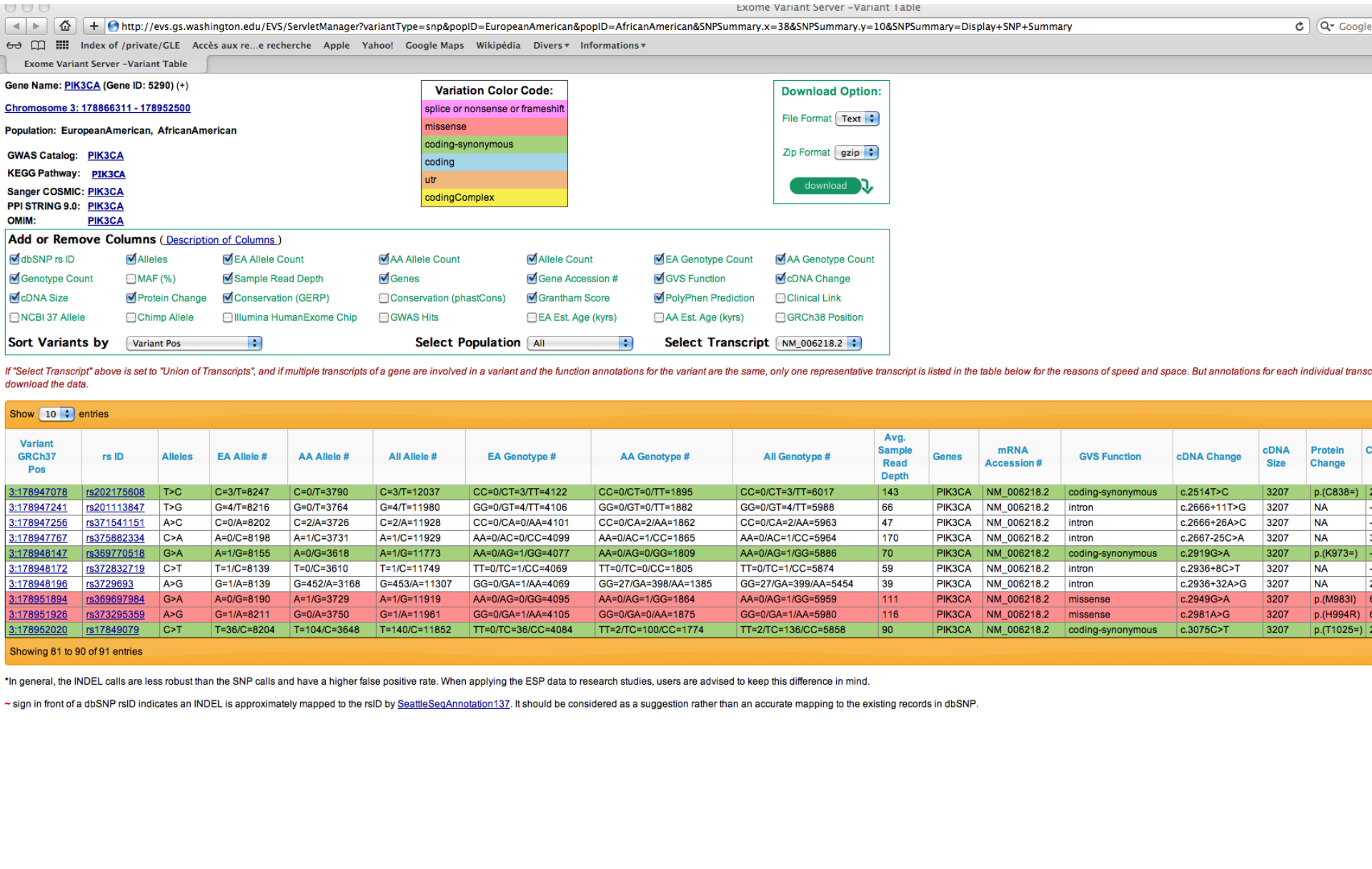
			activity. 2 Publications			
Natural variant[i]	1047 – 1047	1	H → R in CLOVE, KERSEB, CRC, BC and OC; also found in an endometrial carcinoma sample; shows an increase in lipid kinase activity; oncogenic in vivo; requires binding to p85 regulatory subunit to induce cellular transformation but not interaction with RAS; may mimic the conformatitonal change triggered by the interaction with RAS; enhances invadopodia-mediated extracellular matrix degradation and invasion in breast cancer cells; increases lipid kinase activity; may alter the interaction of the PI3K/PI4K kinase domain with the cell membrane. 12 Publications		VAR_026192	

Breast cancer (BC) [MIM:114480]: A common malignancy originating from breast epithelial tissue. Breast neoplasms can be distinguished by their histologic pattern. Invasive ductal carcinoma is by far the most common type. Breast cancer is etiologically and genetically heterogeneous. Important genetic factors have been indicated by familial occurrence and bilateral involvement. Mutations at more than one locus can be involved in different families or even in the same case. 1 Publication

Note: Disease susceptibility is associated with variations affecting the gene represented in this entry.

Exome variant Server

http://evs.gs.washington.edu/EVS/



http://www.cbioportal.org/public-portal/index.do

Index of /private/GLE | Accès aux re...e recherche | Apple | Yahoo! | Google Maps | Wikipédia | Divers | Informations

cBioPortal for Cancer Genomics

cBioPortal for Cancer Genomics

Visualize, analyze, discover.

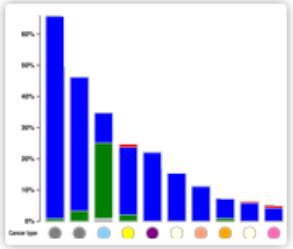
Memorial Sloan Kettering Cancer Center

HOME | DATA SETS | WEB API | R/MATLAB | TUTORIALS | FAQ | NEWS | TOOLS | ABOUT | JOBS | VISUALIZE YOUR DATA

The cBioPortal for Cancer Genomics provides **visualization**, **analysis** and **download** of large-scale **cancer genomics** data sets.

Please adhere to the TCGA publication guidelines when using TCGA data in your publications.

Please cite Gao et al. *Sci. Signal.* 2013 & Cerami et al. *Cancer Discov.* 2012 when publishing results based on cBioPortal.



Query | **Download Data**

Select Cancer Study: All Cancer Studies

Select Data Type Priority: (•) Mutation and CNA ( ) Only Mutation ( ) Only CNA

Enter Gene Set: Advanced: Onco Query Language (OQL)

User-defined List

PIK3CA

All gene symbols are valid.

Submit

http://www.cbioportal.org/index.do

What's New

- We are hiring a **Front End Engineer**
- We are hiring a **Data Curator**
- **New data and features released**

Sign up for low-volume email news alerts:

Subscribe

Or follow us *@cbioportal* on Twitter

Data Sets

The Portal contains data for **17584 tumor samples from 69 cancer studies.** [Details.]

Example Queries

RAS/RAF alterations in colorectal cancer

BRCA1 and BRCA2 mutations in ovarian cancer

POLE hotspot mutations in endometrial cancer

cBioPortal for Cancer Genomics

Visualize, analyze, discover.

Memorial Sloan Kettering Cancer Center

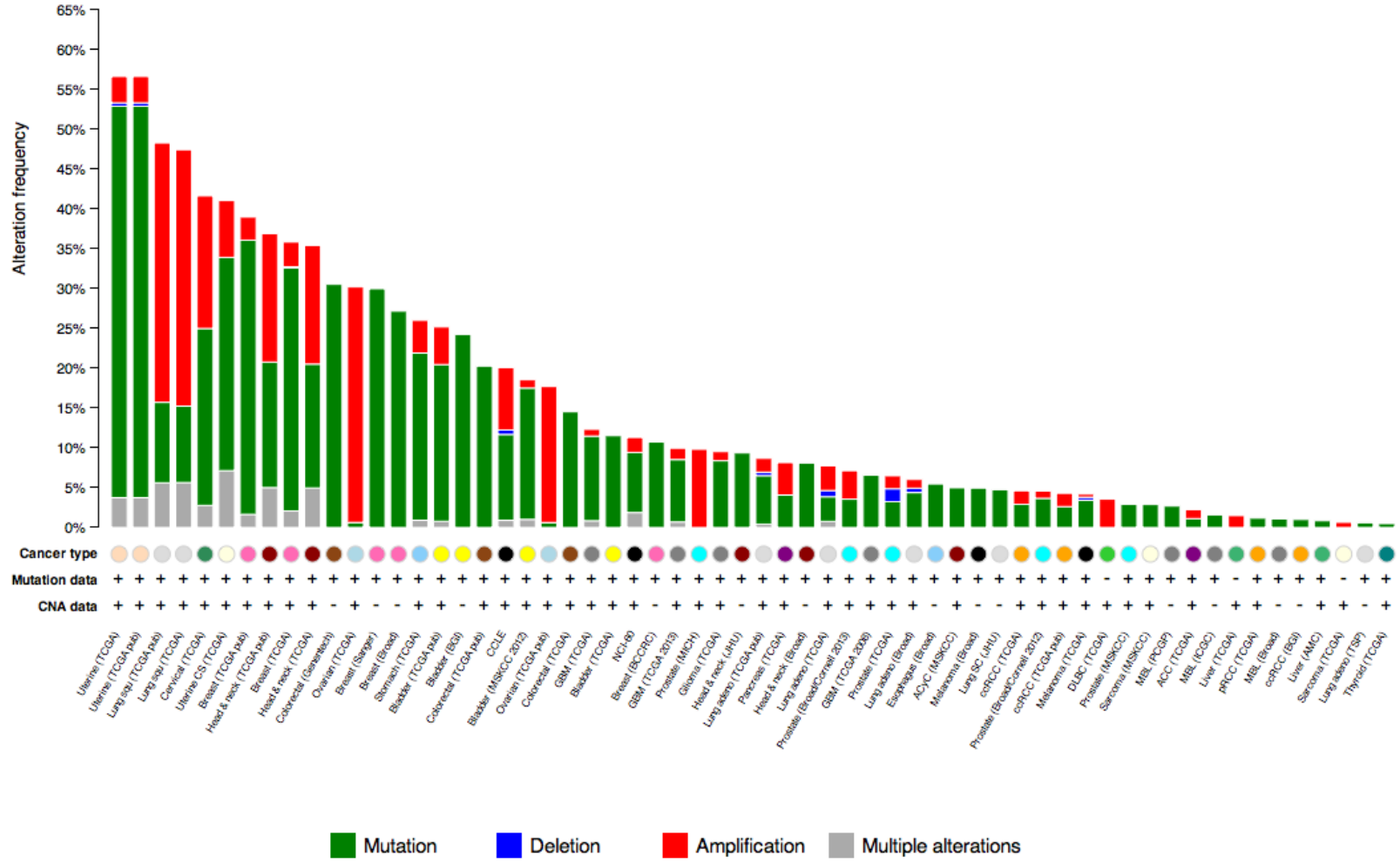
HOME | DATA SETS | WEB API | R/MATLAB | RESULTS | TUTORIALS | FAQ | NEWS | TOOLS | ABOUT | JOBS | VISUALIZE YOUR

▸ **Modify Query**

Overview | Mutations | Download | Bookmark

Cross-cancer alteration summary for PIK3CA (69 studies / 1 gene)

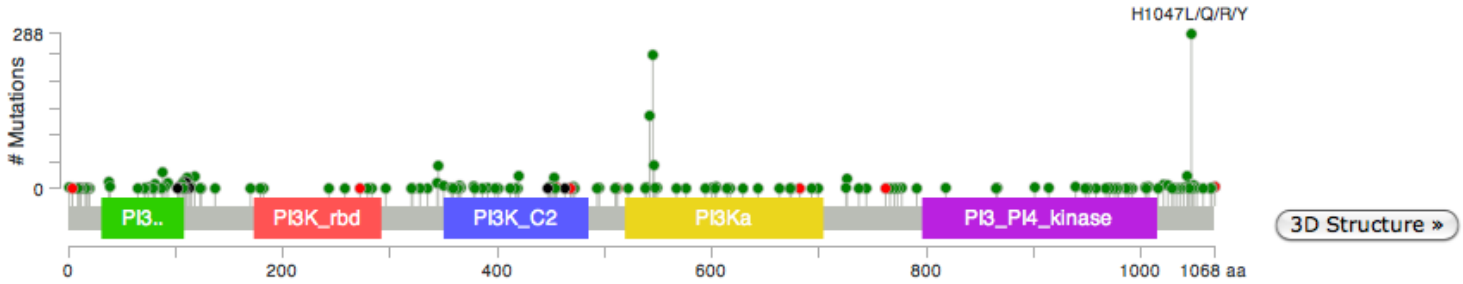
PDF | SVG | Customize histogram



PIK3CA

PIK3CA:

PK3CA_HUMAN



3D Structure »

Show / hide columns

Showing 1,750 mutation(s)

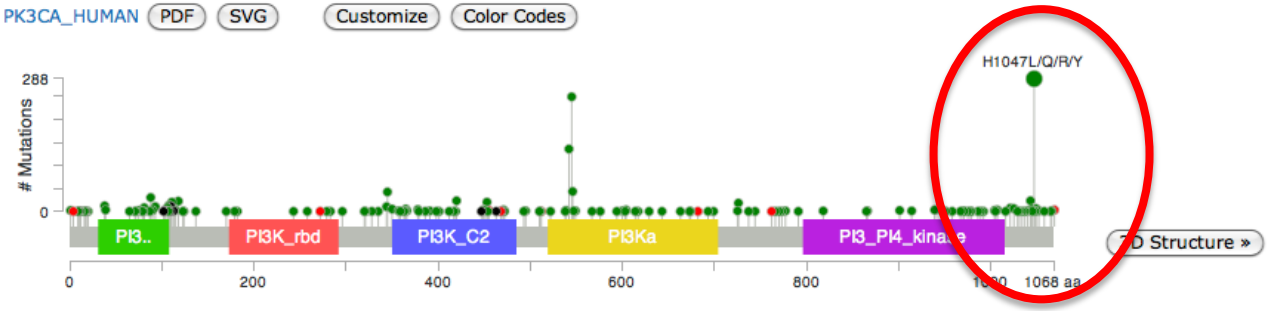
Search:

Case ID	Cancer Study	AA change		Type	Copy #	COSMIC	VS	Mutation Assessor	Allele Freq (T)	#Mut in Sample
SA217	Breast (BCCRC)	*103_106EPVG>D*	3D	IF del	NA	2	U		NA	16
TCGA-A2-A0EV	Breast (TCGA pub)	*469_469E>DK*	3D	IF ins	gain	1	U		NA	46
TCGA-B6-A0WY	Breast (TCGA pub)	*109_112EEKI>D*	3D	IF del	gain	1	U		NA	30
TCGA-A2-A0EV	Breast (TCGA)	*469_469E>DK*	3D	IF ins	gain	1	U		NA	46
KYSE70_OESOPHAGUS	CCLE	**1069W*		Nonstop	AMP	1	U		0.86	63
IGROV1_OVARY	CCLE	**1069W*		Nonstop	diploid	1	U		0.52	276
TCGA-32-1979	GBM (TCGA 2013)	**1069W*		Nonstop	diploid	1	U		0.35	53
TCGA-06-0879	GBM (TCGA 2013)	*455_460LLNPIG>F*	3D	IF del	diploid		U		0.15	39
TCGA-02-0046	GBM (TCGA 2008)	*17_18PP>P*	3D	IF del	diploid		V		NA	4
TCGA-32-1979	GBM (TCGA)	**1069W*		Nonstop	diploid	1	U		NA	71
TCGA-06-0879	GBM (TCGA)	*455_460LLNPIG>F*	3D	IF del	diploid		U		NA	49
TCGA-73-7499	Lung adeno (TCGA pub)	*123_124MP>IA*	3D	Missense	gain	1	U		0.36	77
IGROV1	NCI-60	**1069W*		Nonstop	diploid	1	U		NA	620
PT143MRC	Sarcoma (MSKCC)	*MUTATED*		Missense	diploid		V		NA	1
PT195SYN	Sarcoma (MSKCC)	*MUTATED*		FS del	diploid		V		NA	2
PT173PL	Sarcoma (MSKCC)	*MUTATED*		FS del	gain		V		NA	2
PT149MRC	Sarcoma (MSKCC)	*MUTATED*		Missense	diploid		V		NA	2
PT138MRC	Sarcoma (MSKCC)	*MUTATED*		Missense	diploid		V		NA	1
PT158MRC	Sarcoma (MSKCC)	*MUTATED*		Missense	diploid		V		NA	1
TCGA-BG-A0MT	Uterine (TCGA)	*463_465GSN>D*	3D	IF del	gain	1	U		NA	41
TCGA-AP-A051	Uterine (TCGA)	*e1-1*	3D	Splice	diploid		U		0.50	6542
TCGA-BG-A0MT	Uterine (TCGA pub)	*463_465GSN>D*	3D	IF del	gain	1	U		NA	43

PIK3CA

PIK3CA:

PK3CA_HUMAN PDF SVG Customize Color Codes



Current view shows filtered results. Click here to reset all filters.

Show / hide columns

Showing 400 mutation(s) (out of 1,750 total mutations)

Search:

Case ID	Cancer Study	AA change		Type	Copy #	COSMIC	VS	Mutation Assessor	Allele Freq (T)	#Mut in Sample
TCGA-OR-A5LK	ACC (TCGA)	H1047R	3D	Missense	gain	1878	U	Neutral	NA	82
BL018	Bladder (MSKCC 2012)	H1047R	3D	Missense	diploid	1878	V	Neutral	NA	1
BL067	Bladder (MSKCC 2012)	H1047R	3D	Missense	diploid	1878	V	Neutral	NA	1
B78	Bladder (BGI)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	100
TCGA-FD-A3SJ	Bladder (TCGA pub)	H1047L	3D	Missense	gain	1878	U	Neutral	0.43	202
TCGA-HT-7478	Glioma (TCGA)	H1047L	3D	Missense	diploid	1878	U	Neutral	0.13	27
TCGA-FG-8187	Glioma (TCGA)	H1047R	3D	Missense	diploid	1878	U	Neutral	0.08	14
SA223	Breast (BCCRC)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	15
SA218	Breast (BCCRC)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	126
SA214	Breast (BCCRC)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	193
SA215	Breast (BCCRC)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	12
SA056	Breast (BCCRC)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	14
BR-V-003	Breast (Broad)	H1047R	3D	Missense	NA	1878	V	Neutral	NA	13
BR-V-020	Breast (Broad)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	15
BR-V-024	Breast (Broad)	H1047L	3D	Missense	NA	1878	V	Neutral	NA	26
BR-M-083	Breast (Broad)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	17
BR-M-050	Breast (Broad)	H1047R	3D	Missense	NA	1878	V	Neutral	NA	24
BR-M-036	Breast (Broad)	H1047R	3D	Missense	NA	1878	V	Neutral	NA	11
BR-M-026	Breast (Broad)	H1047R	3D	Missense	NA	1878	V	Neutral	NA	27
BR-M-116	Breast (Broad)	H1047R	3D	Missense	NA	1878	V	Neutral	NA	102
BR-V-064	Breast (Broad)	H1047L	3D	Missense	NA	1878	U	Neutral	NA	17
BR-M-158	Breast (Broad)	H1047R	3D	Missense	NA	1878	V	Neutral	NA	16
BR-M-121	Breast (Broad)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	26
BR-M-098	Breast (Broad)	H1047L	3D	Missense	NA	1878	V	Neutral	NA	29
BR-M-184	Breast (Broad)	H1047R	3D	Missense	NA	1878	V	Neutral	NA	11
BR-V-052	Breast (Broad)	H1047R	3D	Missense	NA	1878	V	Neutral	NA	28
PD4124a	Breast (Sanger)	H1047L	3D	Missense	NA	1878	U	Neutral	NA	28
PD4601a	Breast (Sanger)	H1047R	3D	Missense	NA	1878	U	Neutral	NA	151

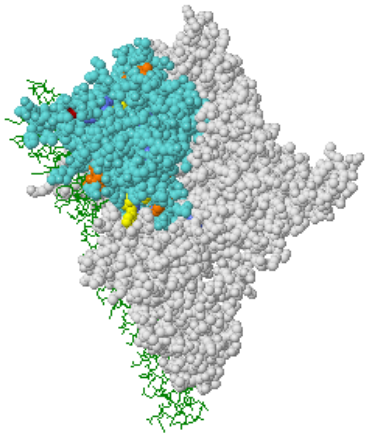
BLAST mapping of all PDB structures to region PK3CA_HUMAN/797-1068
Only mappings with position of current variant H1047R shown

	PDB	e-value	length	identity	resolution	ligand	small.mol name	ligand protein ID	title
1	4a55:A	0	265	96.2%	3.5				crystal structure of p110alpha in complex with ish2 of p85alp the inhibitor pik-108
2	2rd0:A	0	254	96.1%	3.05				structure of a human p110alpha/p85alpha complex
3	3hhm:A	0	266	95.1%	2.8				crystal structure of p110alpha h1047r mutant in complex with nish2 of p85alpha and the drug wortmannin
4	3hiz:A	0	266	88.3%	3.3				crystal structure of p110alpha h1047r mutant in complex with nish2 of p85alpha

Mapped PDB chain shown with spacefill, all other chains - with wireframe. Structure centered on variant residue.

```
PK3CA_HUMAN/797-1061 : MMLGSGMPELQSFDDIAYIRKTLALDKTEQEALEYFMKQMNDAHHGGWTTKMDWIFHT
             midline : MMLGSGMPELQSFDDIAYIRKTLALDKTEQEALEYF KQMNDAHHGGWTTKMDWIFHT
      4a55:A/745-1001 : MMLGSGMPELQSFDDIAYIRKTLALDKTEQEALEYFTKQMNDAHHGGWTTKMDWIFHT
```

specificity
conserved
neutral
unmapped
hetero
variant



Imol

1000 Genomes

http://phase3browser.1000genomes.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000121879;r=3:178865902-178957881#missense_variant_tablePanel

rs121913279	3:178952085	A/G/**T**	-	SNP	dbSNP		?	Missense variant	H/L	1047	0.75	0.085	ENST00000263967
rs121913279	3:178952085	A/**G**/T	-	SNP	dbSNP		?	Missense variant	H/R	1047	0.36	0.529	ENST00000263967
COSM776	3:178952085	A/T	-	somatic_SNV	COSMIC	-	-	Missense variant	H/L	1047	0.75	0.085	ENST00000263967
COSM775	3:178952085	A/G	-	somatic_SNV	COSMIC	-	-	Missense variant	H/R	1047	0.36	0.529	ENST00000263967
COSM249874	3:178952085	A/C	-	somatic_SNV	COSMIC	-	-	Missense variant	H/P	1047	0.21	0.825	ENST00000263967
COSM94987	3:178952085	A/T	-	somatic_SNV	COSMIC	-	-	Missense variant	H/L	1047	0.75	0.085	ENST00000263967
COSM94986	3:178952085	A/G	-	somatic_SNV	COSMIC	-	-	Missense variant	H/R	1047	0.36	0.529	ENST00000263967
COSM776	3:178952085	A/T	-	somatic_SNV	COSMIC	-	-	Missense variant	H/L	1047	0.75	0.085	ENST00000263967
COSM775	3:178952085	A/G	-	somatic_SNV	COSMIC	-	-	Missense variant	H/R	1047	0.36	0.529	ENST00000263967
COSM249874	3:178952085	A/C	-	somatic_SNV	COSMIC	-	-	Missense variant	H/P	1047	0.21	0.825	ENST00000263967
COSM94987	3:178952085	A/T	-	somatic_SNV	COSMIC	-	-	Missense variant	H/L	1047	0.75	0.085	ENST00000263967
COSM94986	3:178952085	A/G	-	somatic_SNV	COSMIC	-	-	Missense variant	H/R	1047	0.36	0.529	ENST00000263967

COSM94986 SOMATIC SNV

Original source	Somatic mutations found in human cancers from the COSMIC project (release 67) I View in COSMIC
Alleles	**A/G** I Ancestral: **A** I Ambiguity code: **R**
Location	Chromosome **3:178952085** (forward strand) I View in location tab
Co-located	with **COSMIC** COSM776 (A/T), COSM775 (A/G), COSM249874 (A/C), COSM94987 (A/T) ; **dbSNP** rs121913279 (A/G/T)
HGVS names ⊟	This variation has **3** HGVS names - click the plus to show 3:g.178952085A>G ENST00000263967.3:c.3140A>G ENSP00000263967.3:p.His1047Arg

Explore this variation

COSM94986 SOMATIC SNV

Original source	Somatic mutations found in human cancers from the COSMIC project (release 67) I View in COSMIC
Alleles	**A/G** I Ancestral: **A** I Ambiguity code: **R**
Location	Chromosome **3:178952085** (forward strand) I View in location tab
Co-located	with **COSMIC** COSM776 (A/T), COSM775 (A/G), COSM249874 (A/C), COSM94987 (A/T) ; **dbSNP** rs121913279 (A/G/T)
HGVS names ⊟	This variation has **3** HGVS names - click the plus to show 3:g.178952085A>G ENST00000263967.3:c.3140A>G ENSP00000263967.3:p.His1047Arg

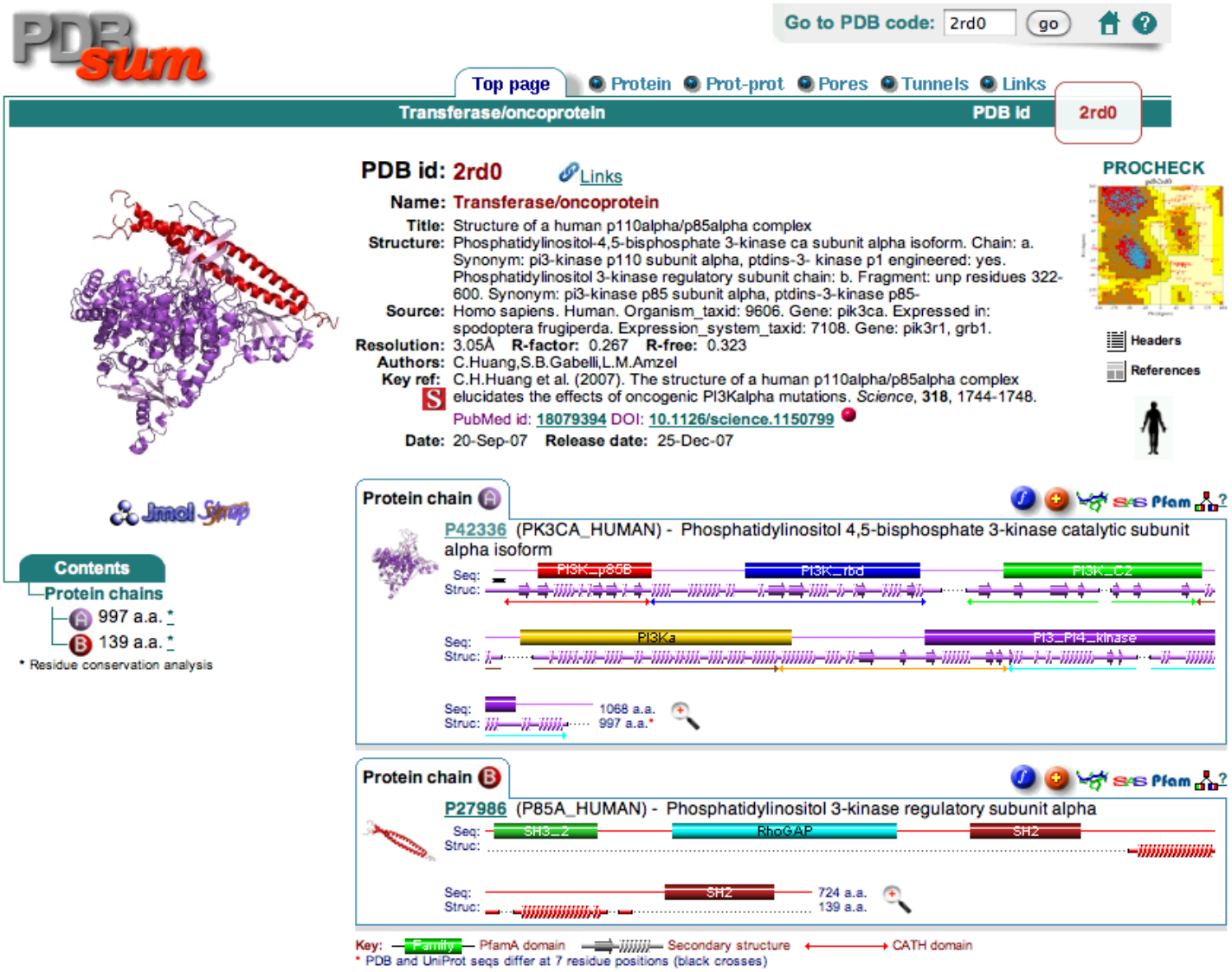
Phenotype Data

Significant association(s)

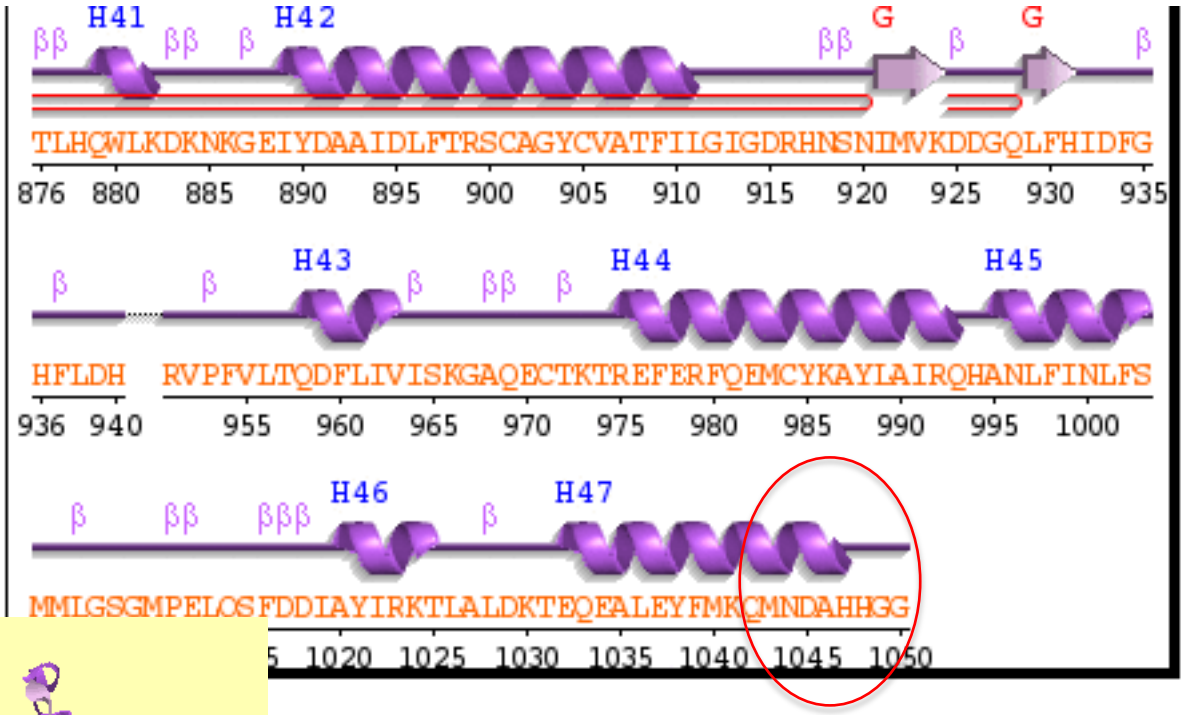
Show All entries | Show/hide columns | Filter

Disease/Trait	Source(s)	Tumour site	Reported gene(s)	Associated variant(s
COSMIC tumour site: liver [View on Karyotype]	[COSMIC]	liver	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: skin [View on Karyotype]	[COSMIC]	skin	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: endometrium [View on Karyotype]	[COSMIC]	endometrium	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: breast [View on Karyotype]	[COSMIC]	breast	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: oesophagus [View on Karyotype]	[COSMIC]	oesophagus	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: stomach [View on Karyotype]	[COSMIC]	stomach	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: kidney [View on Karyotype]	[COSMIC]	kidney	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: lung [View on Karyotype]	[COSMIC]	lung	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: pancreas [View on Karyotype]	[COSMIC]	pancreas	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: central nervous system [View on Karyotype]	[COSMIC]	central nervous system	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: prostate [View on Karyotype]	[COSMIC]	prostate	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: ovary [View on Karyotype]	[COSMIC]	ovary	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: large intestine [View on Karyotype]	[COSMIC]	large intestine	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: urinary tract [View on Karyotype]	[COSMIC]	urinary tract	PIK3CA_ENST00000263967	COSM94986
COSMIC tumour site: upper aerodigestive tract [View on Karyotype]	[COSMIC]	upper aerodigestive tract	PIK3CA_ENST00000263967	COSM94986

PDBsum entry 2rd0



https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html



Helices labelled H1, H2, ... and strands by their sheets A, B, ...

rand