CANCER GENOMICS CONSORTIUM

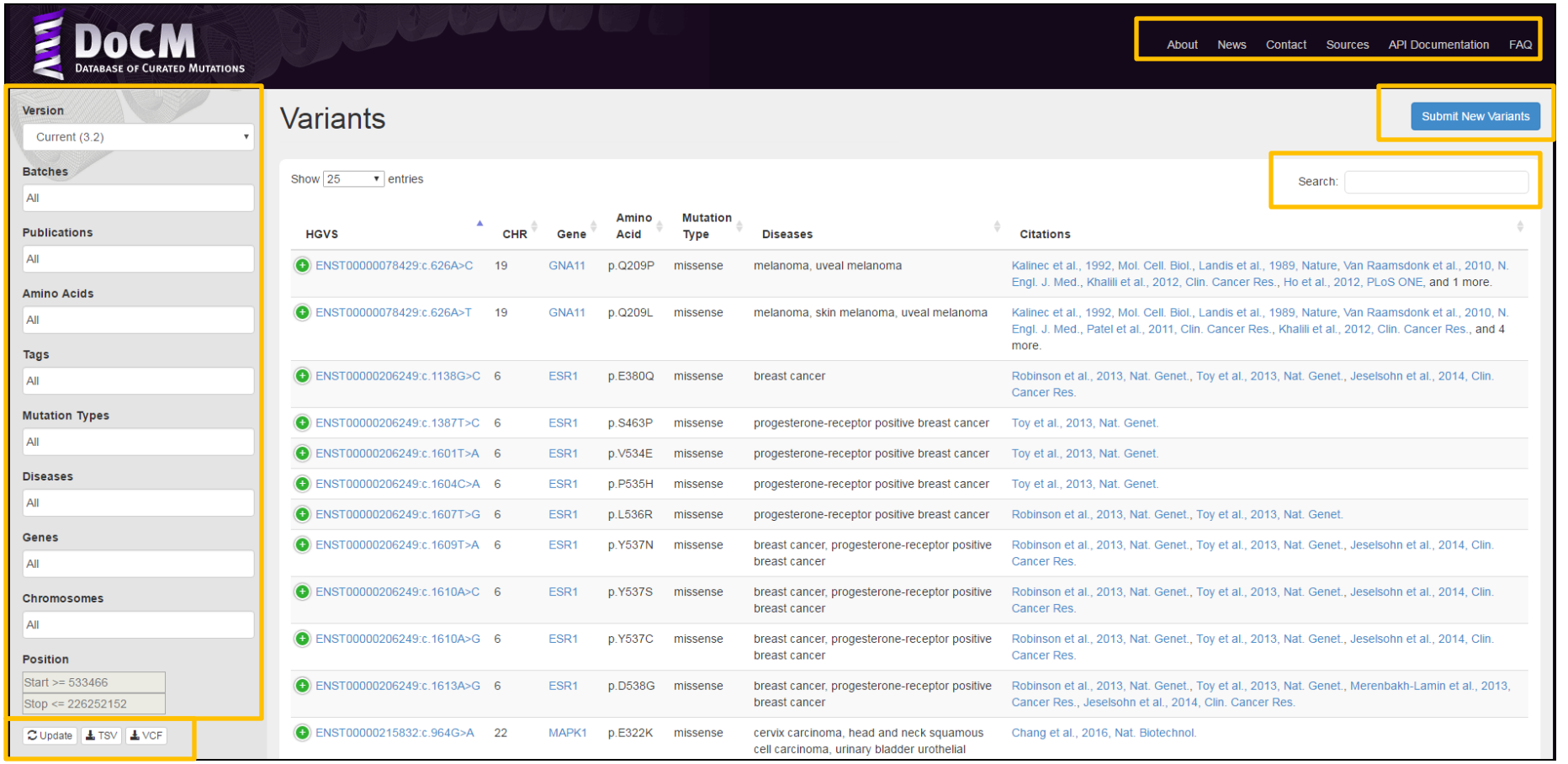
Educating for Best Practices in Clinical Cancer Genomics

Database of Curated Mutations (DoCM)

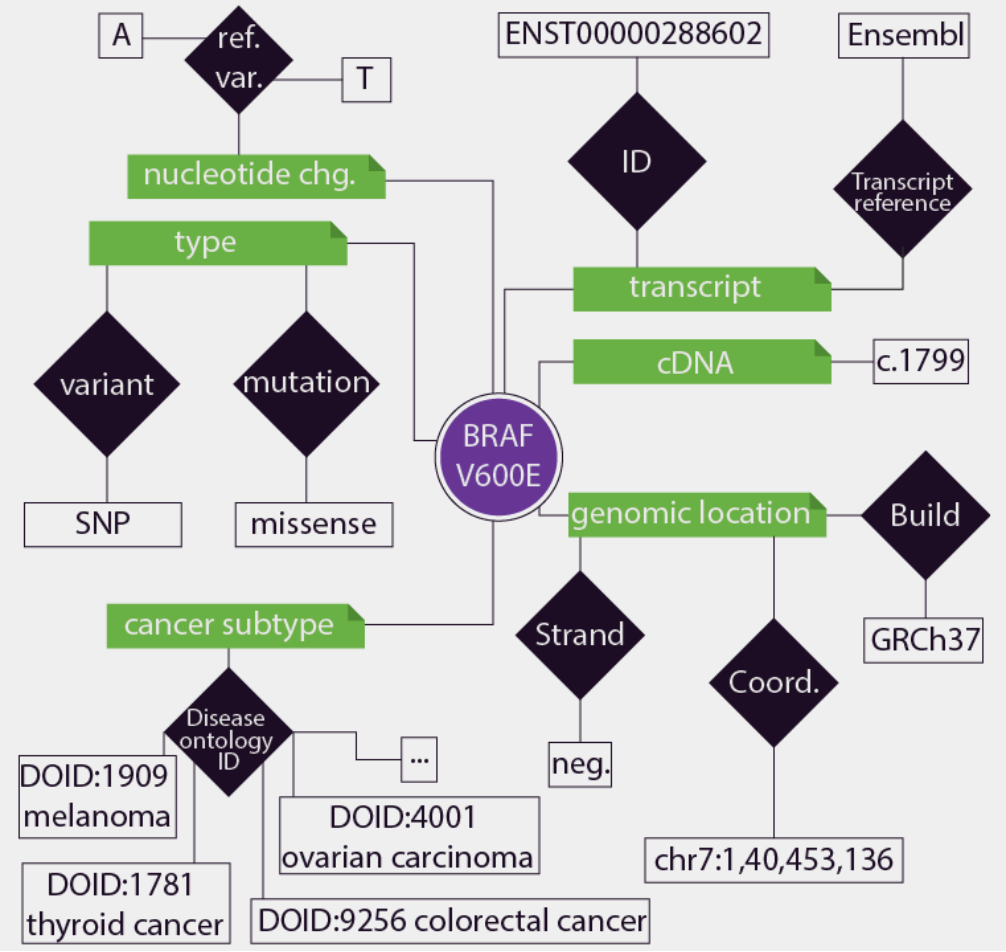
http://docm.genome.wustl.edu/

http://www.nature.com/nmeth/journal/v13/n10/full/nmeth.4000.html

Home Page


DoCM
Database of Curated Mutations
About
News
Contact
Sources
API Documentation
FAQ
Version
Current (3.2)
Batches
All
Publications
All
Amino Acids
All
Tags
All
Mutation Types
All
Diseases
All
Genes
All
Chromosomes
All
Position
Start >= 533466
Stop <= 226252152
Update
TSV
VCF
Variants
Submit New Variants
Show 25 entries
Search:
HGVS
CHR
Gene
Amino Acid
Mutation Type
Diseases
Citations
ENST00000078429:c.626A>C 19 GNA11 p.Q209P missense melanoma, uveal melanoma Kalinec et al., 1992, Mol. Cell. Biol., Landis et al., 1989, Nature, Van Raamsdonk et al., 2010, N. Engl. J. Med., Khalili et al., 2012, Clin. Cancer Res., Ho et al., 2012, PLoS ONE, and 1 more.
ENST00000078429:c.626A>T 19 GNA11 p.Q209L missense melanoma, skin melanoma, uveal melanoma Kalinec et al., 1992, Mol. Cell. Biol., Landis et al., 1989, Nature, Van Raamsdonk et al., 2010, N. Engl. J. Med., Patel et al., 2011, Clin. Cancer Res., Khalili et al., 2012, Clin. Cancer Res., and 4 more.
ENST00000206249:c.1138G>C 6 ESR1 p.E380Q missense breast cancer Robinson et al., 2013, Nat. Genet., Toy et al., 2013, Nat. Genet., Jeselsohn et al., 2014, Clin. Cancer Res.
ENST00000206249:c.1387T>C 6 ESR1 p.S463P missense progesterone-receptor positive breast cancer Toy et al., 2013, Nat. Genet.
ENST00000206249:c.1601T>A 6 ESR1 p.V534E missense progesterone-receptor positive breast cancer Toy et al., 2013, Nat. Genet.
ENST00000206249:c.1604C>A 6 ESR1 p.P535H missense progesterone-receptor positive breast cancer Toy et al., 2013, Nat. Genet.
ENST00000206249:c.1607T>G 6 ESR1 p.L536R missense progesterone-receptor positive breast cancer Robinson et al., 2013, Nat. Genet., Toy et al., 2013, Nat. Genet.
ENST00000206249:c.1609T>A 6 ESR1 p.Y537N missense breast cancer, progesterone-receptor positive breast cancer Robinson et al., 2013, Nat. Genet., Toy et al., 2013, Nat. Genet., Jeselsohn et al., 2014, Clin. Cancer Res.
ENST00000206249:c.1610A>C 6 ESR1 p.Y537S missense breast cancer, progesterone-receptor positive breast cancer Robinson et al., 2013, Nat. Genet., Toy et al., 2013, Nat. Genet., Jeselsohn et al., 2014, Clin. Cancer Res.
ENST00000206249:c.1610A>G 6 ESR1 p.Y537C missense breast cancer, progesterone-receptor positive breast cancer Robinson et al., 2013, Nat. Genet., Toy et al., 2013, Nat. Genet., Jeselsohn et al., 2014, Clin. Cancer Res.
ENST00000206249:c.1613A>G 6 ESR1 p.D538G missense breast cancer, progesterone-receptor positive breast cancer Robinson et al., 2013, Nat. Genet., Toy et al., 2013, Nat. Genet., Merenbakh-Lamin et al., 2013, Cancer Res., Jeselsohn et al., 2014, Clin. Cancer Res.
ENST00000215832:c.964G>A 22 MAPK1 p.E322K missense cervix carcinoma, head and neck squamous cell carcinoma, urinary bladder urothelial Chang et al., 2016, Nat. Biotechnol.

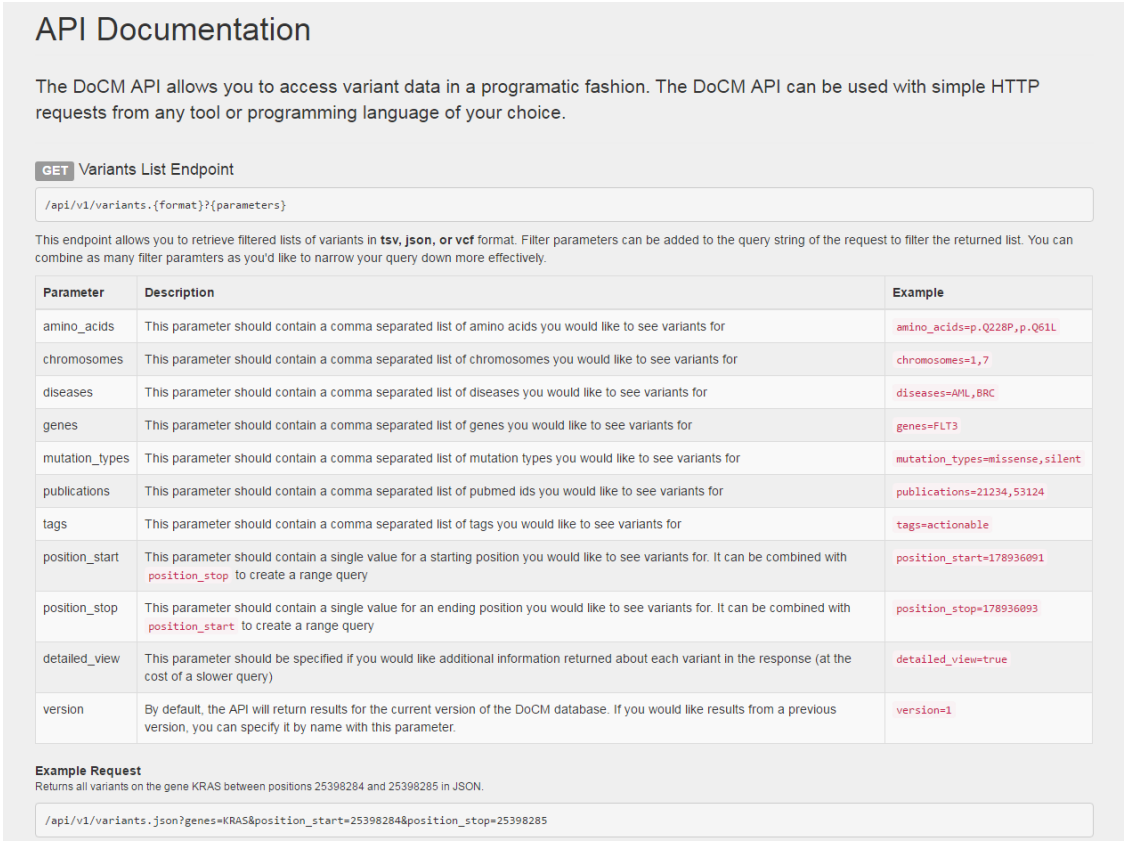
Information in DoCM



- DoCM uses many data sources to compile master list – relevant links on Sources page
  - Kin-Driver
  - WashU hematologic malignancy mutation list
  - Literature – 876 publications currently
  - Drug Gene Knowledge Database
  - CIViC
  - Pan-cancer recurrent hotspots
  - My Cancer Genome
  - Oncomap Variants

Open Source API

- API Documentation is accessible through the link on home page.



API Documentation

The DoCM API allows you to access variant data in a programatic fashion. The DoCM API can be used with simple HTTP requests from any tool or programming language of your choice.

GET Variants List Endpoint

```
/api/v1/variants.{format}?{parameters}
```

This endpoint allows you to retrieve filtered lists of variants in **tsv, json, or vcf** format. Filter parameters can be added to the query string of the request to filter the returned list. You can combine as many filter paramters as you'd like to narrow your query down more effectively.

Parameter	Description	Example
amino_acids	This parameter should contain a comma separated list of amino acids you would like to see variants for	`amino_acids=p.Q228P,p.Q61L`
chromosomes	This parameter should contain a comma separated list of chromosomes you would like to see variants for	`chromosomes=1,7`
diseases	This parameter should contain a comma separated list of diseases you would like to see variants for	`diseases=AML,BRC`
genes	This parameter should contain a comma separated list of genes you would like to see variants for	`genes=FLT3`
mutation_types	This parameter should contain a comma separated list of mutation types you would like to see variants for	`mutation_types=missense,silent`
publications	This parameter should contain a comma separated list of pubmed ids you would like to see variants for	`publications=21234,53124`
tags	This parameter should contain a comma separated list of tags you would like to see variants for	`tags=actionable`
position_start	This parameter should contain a single value for a starting position you would like to see variants for. It can be combined with `position_stop` to create a range query	`position_start=178936091`
position_stop	This parameter should contain a single value for an ending position you would like to see variants for. It can be combined with `position_start` to create a range query	`position_stop=178936093`
detailed_view	This parameter should be specified if you would like additional information returned about each variant in the response (at the cost of a slower query)	`detailed_view=true`
version	By default, the API will return results for the current version of the DoCM database. If you would like results from a previous version, you can specify it by name with this parameter.	`version=1`

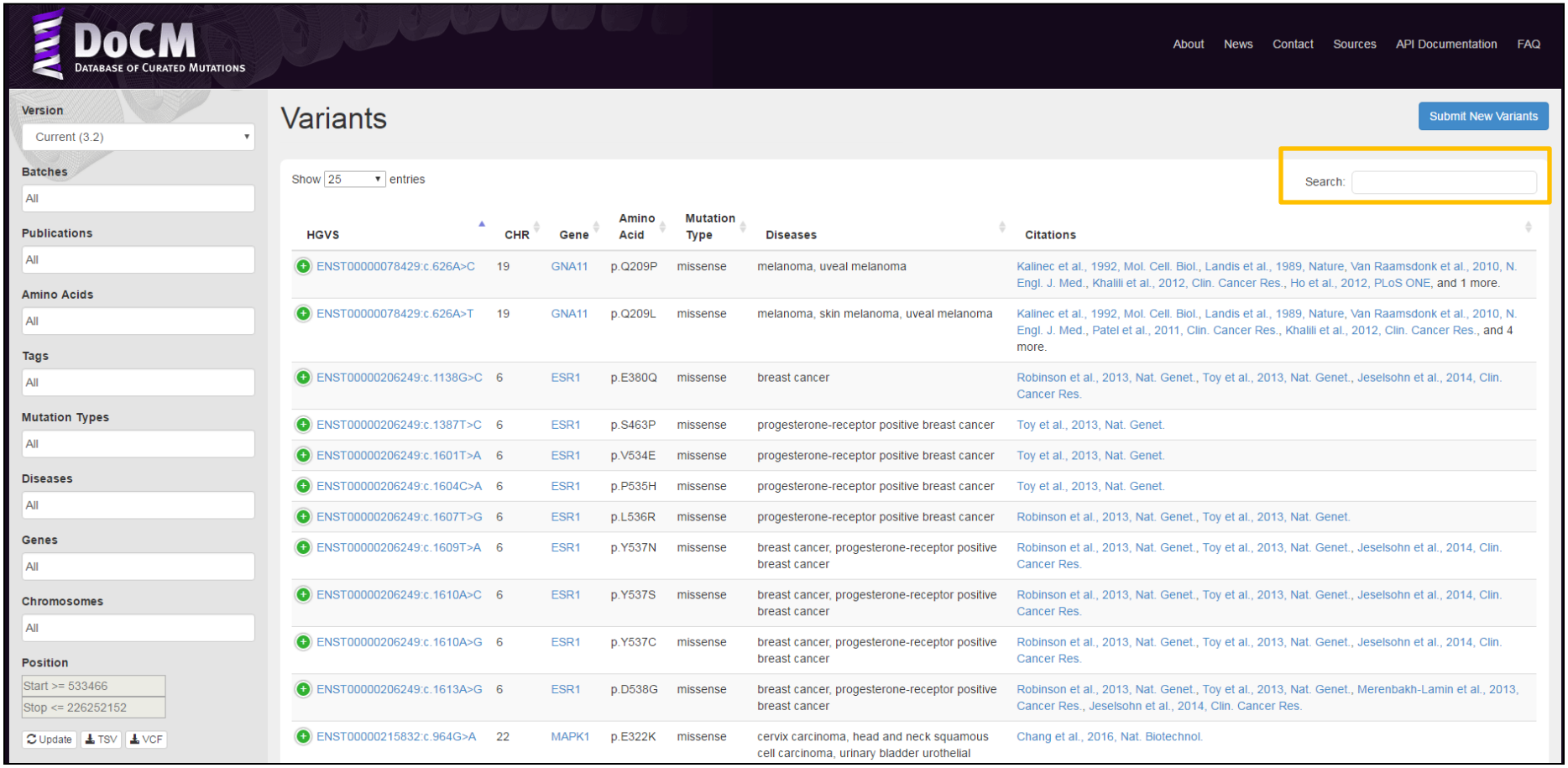
Example Request

Returns all variants on the gene KRAS between positions 25398284 and 25398285 in JSON.

```
/api/v1/variants.json?genes=KRAS&position_start=25398284&position_stop=25398285
```

DoCM Downloads

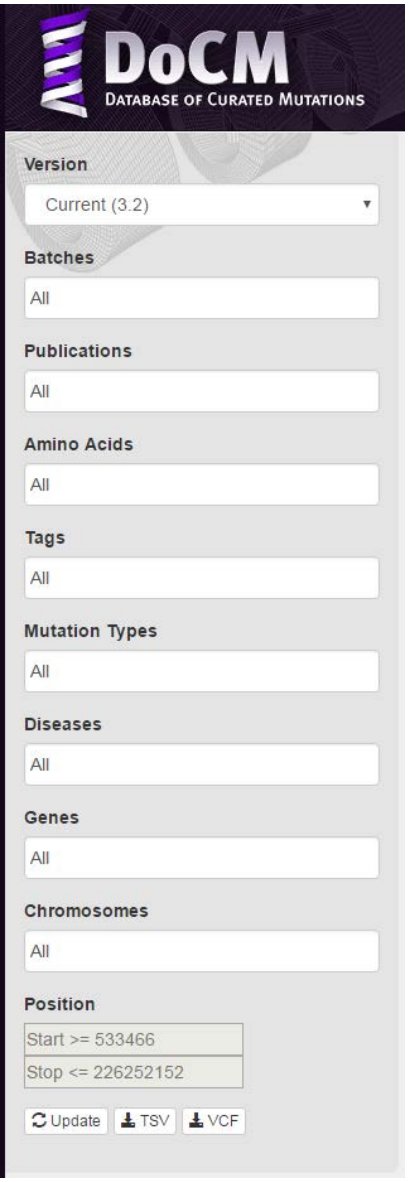
- At bottom of home page is a link for "Downloads"
- Flat file downloads as .TSV file
  - Same output as export TSV file on home page with no search/filter criteria applied
- ClinVar spreadsheets
  - More columns than flat file to better coincide with ClinVar data



- Use Search field to limit results to specific HGVS, chromosome, gene, amino acid (include “p.” when applicable), mutation type, or disease

Submit a New Variant

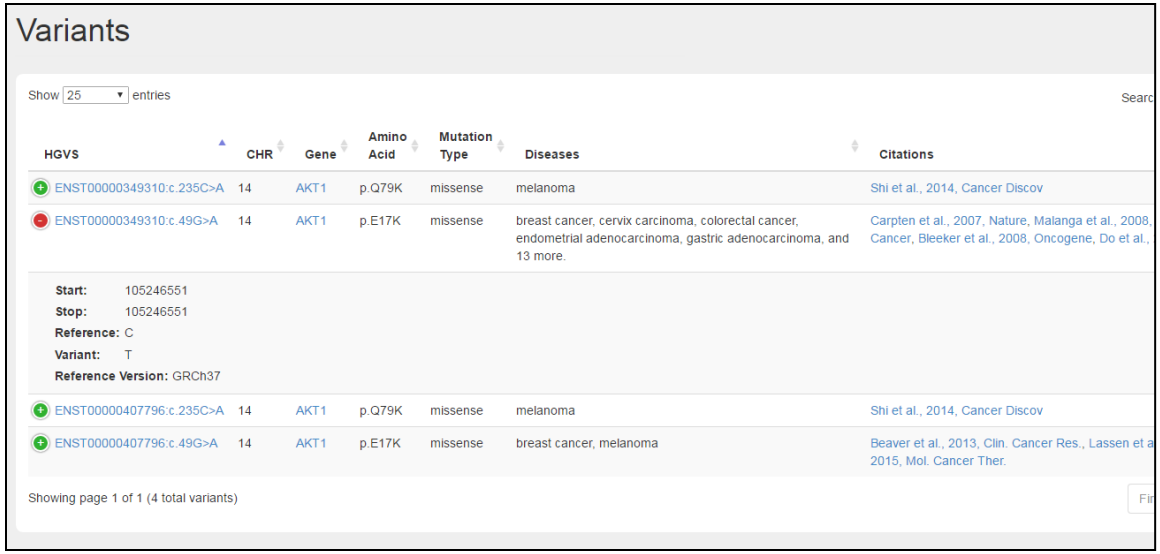
- Click on Submit New Variants button on home page and right click "batch submission help" to open in a new window
  - Click on "our criteria for inclusion" to ensure variants are suitable for submission
  - Format tab delimited file with columns as specified and upload along with submission form.



Search criteria options

- Version – search any version of the database
- Batches – search any knowledgebase/resource used to compile DoCM
- Publications – see all variants from a given publication
- Amino Acid – search by amino acid change
- Tags – select activating, inactivating, pathogenic, or likely pathogenic
- Mutation types – frameshift, inframe, missense, protein altering variant, start lost, stop lost, synonymous.
- Diseases – although filed by Disease Ontology ID (DOID), searchable by common disease name
- Genes – HGNC gene symbol, also can search using Ensembl annotation in search bar.
- Chromosomes – 1-22, X
- Buttons to Update search results and Download variants list as TSV or VCF.
  - Will download only results from current query

Variant Details



- Expand selection by clicking on green "+" sign.
  - Drop down will supply coordinates in GRCh37/hg19
- Click on HGVS to link to variant page in DoCM
- Click on gene name to link to Ensembl page
- Click on citations to link to PubMed

Variant Page

AKT1 (p.E17K)

Variant Data

Location

HGVS: ENST00000349310:c.49G>A
Reference Version: GRCh37
Chromosome: 14
Start: 105246551
Stop: 105246551
Strand: -1
Transcript: ENST00000349310 (ensembl - 74_37)
Gene: AKT1 (View drug interactions on DGIdb)

Information

Reference: C
Variant: T
Amino Acid: p.E17K
Mutation Type: missense
Variant Type: SNV (SO:0001483)
cDNA Change: c.49
Tags: pathogenic | likely pathogenic

Disease Data

Disease	Source	Batch	Tags	External Links
colorectal cancer	Carpten et al., 2007, Nature	My Cancer Genome (View variants)	pathogenic	My Cancer Genome
non-small cell lung carcinoma	Carpten et al., 2007, Nature	My Cancer Genome (View variants)	pathogenic	My Cancer Genome
breast cancer	Carpten et al., 2007, Nature	My Cancer Genome (View variants)	pathogenic, likely pathogenic	My Cancer Genome
non-small cell lung carcinoma	Malanga et al., 2008, Cell Cycle	My Cancer Genome (View variants)	pathogenic	My Cancer Genome
colorectal cancer	Kim et al., 2008, Br. J. Cancer	My Cancer Genome (View variants)	pathogenic	My Cancer Genome
non-small cell lung carcinoma	Bleeker et al., 2008, Oncogene	My Cancer Genome (View variants)	pathogenic	My Cancer Genome
non-small cell lung carcinoma	Do et al., 2008, BMC Res Notes	My Cancer Genome (View variants)	pathogenic	My Cancer Genome

Drug Interaction Data

Therapeutic Context	Pathway	Effect	Association	Status	Evidence	Source
allosteric AKT inhibitors	activation	gain-of-function		preclinical	emerging	21464312, 17611497

Variant Data

- Potential links to DGIdb (Gene), CIViC (Variant), Sequence Ontology (Variant Type)

Disease Data

- Links to PubMed, Batch Information, External Links

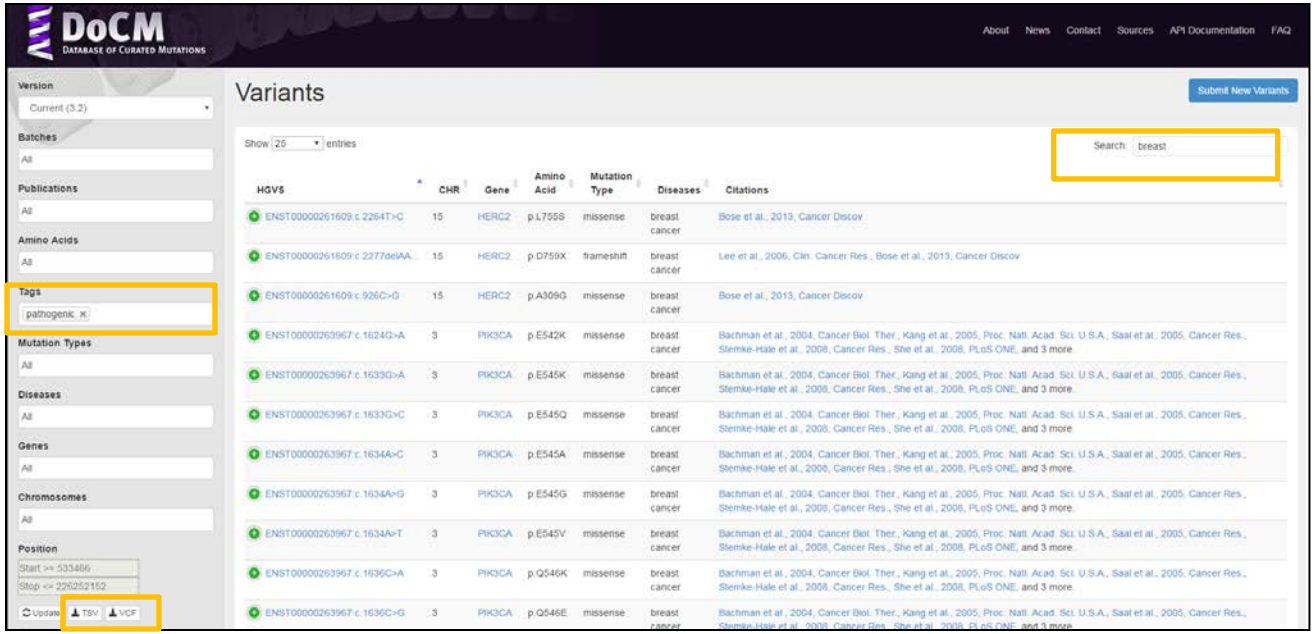
Drug Interaction Data

- Includes PMID

Scenario #1

- You are looking to make a list of genes with known pathogenic variants in breast cancer.
  - Search for breast cancer and add the "pathogenic" tag into search criteria.

Pathogenic Variants in Breast Cancer Gene List



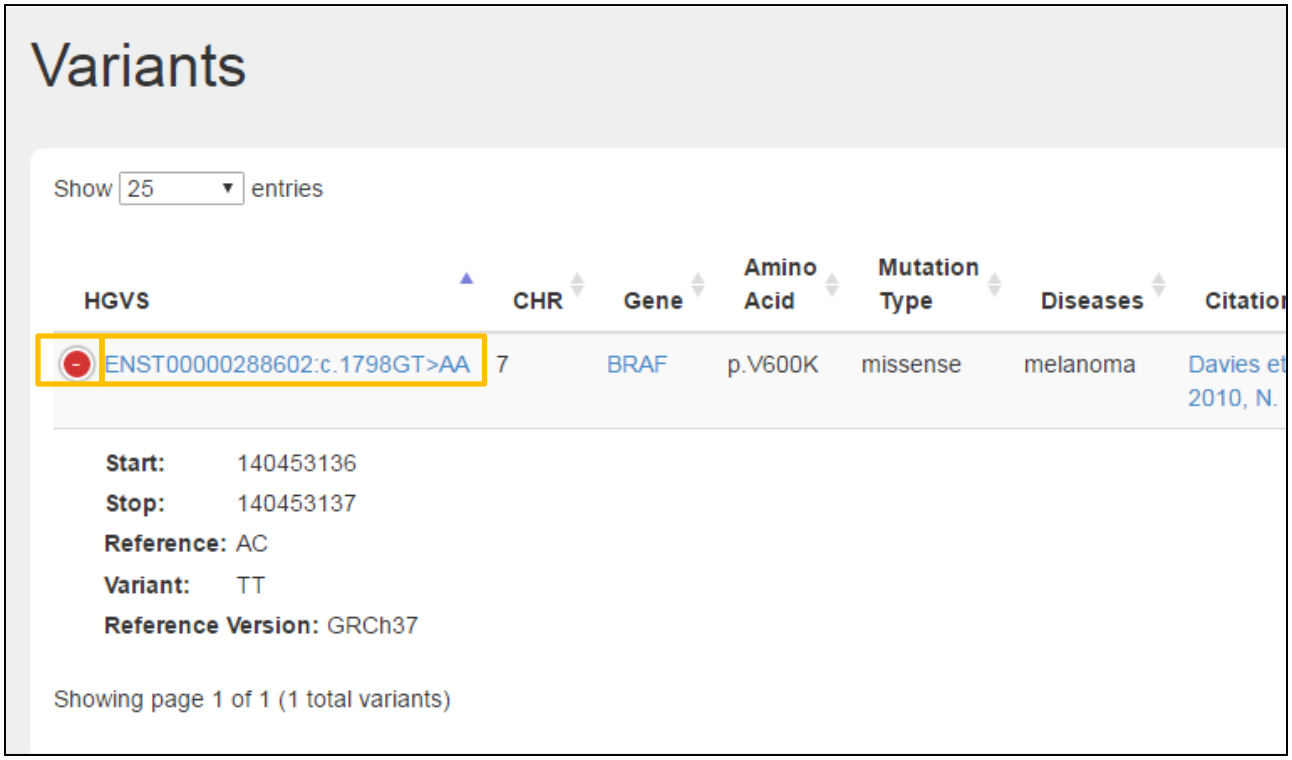
- Enter disease name into search bar
- Add “pathogenic” tag
- Export results in convenient format

CANCER GENOMICS CONSORTIUM

Educating for best practices in clinical cancer genomics

Scenario #2

- You are reporting a case with a BRAF V600K variant in melanoma.
  - Primary literature search
  - Any potential therapeutic direction
  - Start with variant search from home page



- Search for "p.V600K" in search bar on home page.
- Check sequence variant by clicking on the green (+). Marker will turn into red (-) once data is expanded.
- If this matches your data click on HGVS link.
- Links to PubMed are in citations column

Variant page

- List of Sources under "Disease Data"
- Links to PubMed in Source column

Disease Data

Disease	Source
melanoma	Davies et al., 2002, Nature
melanoma	Maldonado et al., 2003, J. Natl. Cancer Inst.
melanoma	Rubinstein et al., 2010, J Transl Med
melanoma	Flaherty et al., 2010, N. Engl. J. Med.
melanoma	Chapman et al., 2011, N. Engl. J. Med.
melanoma	Kirkwood et al., 2012, Clin. Cancer Res.
melanoma	Sosman et al., 2012, N. Engl. J. Med.
melanoma	Lovly et al., 2012, PLoS ONE
melanoma	Falchook et al., 2012, Lancet
melanoma	Flaherty et al., 2012, N. Engl. J. Med.
melanoma	Hauschild et al., 2012, Lancet
melanoma	Falchook et al., 2012, Lancet Oncol.
melanoma	Patel et al., 2013, Cancer
melanoma	Dienstmann et al., 2015, Cancer Discov
melanoma	Ascierto et al., 2013, J. Clin. Oncol.
melanoma	McArthur et al., 2014, Lancet Oncol.
melanoma	MacConaill et al., 2014, J Mol Diagn

- Link to DGIdb in Variant Data
  - Results from DGIdb can be exported as a TSV file
- Drug Interaction Data field at bottom of page has more information

Drug Interaction Data

Therapeutic Context	Pathway	Effect	Association	Status	Evidence	Source
trametinib, dabrafenib + trametinib	activation	gain-of-function		FDA-approved	consensus	FDA
vemurafenib	activation	gain-of-function		late trials	emerging	22356324

Scenario #3

- You are tasked with using DoCM to contribute to the creation of a Pan Cancer gene list in your lab.

Home Page

- From home page, you can download a TSV file of all variants in DoCM
  - See column H for gene names

	A	B	C	D	E	F	G	H	I	J	K	L
1	hgvs	chromosome	start	stop	read	variant	reference_version	gene	mutation_t	amino_aci	diseases	pubmed_sources
2	ENST00000361445:c.7514G>C	1	11169361	11169361	C	G	GRCh37	MTOR	missense	p.R2505P	renal carcinoma	24631838
3	ENST00000361445:c.7500T>G	1	11169375	11169375	A	C	GRCh37	MTOR	missense	p.I2500M	renal clear cell carcinoma,gastric adenoca	26619011
4	ENST00000361445:c.7498A>T	1	11169377	11169377	T	A	GRCh37	MTOR	missense	p.I2500F	uterine corpus endometrial carcinoma,bre	26619011
5	ENST00000361445:c.7255G>A	1	11174420	11174420	C	T	GRCh37	MTOR	missense	p.E2419K	urinary bladder urothelial carcinoma	24625776
6	ENST00000361445:c.6667C>A	1	11182179	11182179	G	T	GRCh37	MTOR	missense	p.Q2223K	renal carcinoma	24622468
7	ENST00000361445:c.6644C>T	1	11184573	11184573	G	A	GRCh37	MTOR	missense	p.S2215F	renal clear cell carcinoma,uterine corpus (	26619011
8	ENST00000361445:c.6644C>A	1	11184573	11184573	G	T	GRCh37	MTOR	missense	p.S2215Y	papillary renal cell carcinoma,renal carcin	###############
9	ENST00000361445:c.6643T>A	1	11184574	11184574	A	T	GRCh37	MTOR	missense	p.S2215T	papillary renal cell carcinoma,skin melanc	26619011
10	ENST00000361445:c.6637C>T	1	11184580	11184580	G	A	GRCh37	MTOR	missense	p.P2213S	melanoma	26490311
11	ENST00000361445:c.6324C>A	1	11187094	11187094	G	T	GRCh37	MTOR	missense	p.F2108L	thyroid carcinoma	25295501
12	ENST00000361445:c.6040G>A	1	11187857	11187857	C	T	GRCh37	MTOR	missense	p.E2014K	urinary bladder urothelial carcinoma	24625776
13	ENST00000361445:c.5902C>T	1	11188519	11188519	G	A	GRCh37	MTOR	missense	p.H1968Y	melanoma	26490311
14	ENST00000361445:c.4449C>G	1	11217229	11217229	G	C	GRCh37	MTOR	missense	p.C1483W	renal clear cell carcinoma,glioblastoma m	26619011
15	ENST00000361445:c.4448G>A	1	11217230	11217230	C	T	GRCh37	MTOR	missense	p.C1483Y	glioblastoma multiforme,breast cancer,ren	26619011
16	ENST00000361445:c.4448G>T	1	11217230	11217230	C	A	GRCh37	MTOR	missense	p.C1483F	renal clear cell carcinoma,breast cancer,g	26619011
17	ENST00000361445:c.4447T>C	1	11217231	11217231	A	G	GRCh37	MTOR	missense	p.C1483R	renal clear cell carcinoma,breast cancer,g	26619011
18	ENST00000361445:c.4379T>C	1	11217299	11217299	A	G	GRCh37	MTOR	missense	p.L1460P	renal carcinoma	24631838
19	ENST00000376592:c.665C>T	1	11856378	11856378	G	A	GRCh37	MTHFR	missense	p.A222V	stomach cancer	18704422
20	ENST00000358432:c.2162G>A	1	16458722	16458722	C	T	GRCh37	EPHA2	missense	p.R721Q	cortical senile cataract	19649315
21	ENST00000358432:c.1171G>A	1	16464489	16464489	C	T	GRCh37	EPHA2	missense	p.G391R	lung squamous cell carcinoma	20360610
22	ENST00000358432:c.1171G>C	1	16464489	16464489	C	G	GRCh37	EPHA2	missense	p.G391R	lung squamous cell carcinoma	20360610
23	ENST00000374632:c.2359G>A	1	23235518	23235518	G	A	GRCh37	EPHB2	missense	p.G787R	colorectal cancer	18682749
24	ENST00000373103:c.1853C>T	1	36933434	36933434	G	A	GRCh37	CSF3R	missense	p.T618I	chronic myeloid leukemia,acute myeloid l	###############
25	ENST00000373103:c.1843A>G	1	36933444	36933444	T	C	GRCh37	CSF3R	missense	p.T615A	chronic myeloid leukemia	23656643
26	ENST00000372470:c.1513A>T	1	43814978	43814978	A	T	GRCh37	MPL	missense	p.S505C	chronic myeloproliferative disease	21228032

Scenario #4

- You have written a paper that has just been published in a peer-reviewed journal detailing 20 patients with a specific mutation not listed in DoCM. How can you help the knowledgebase effort?

Submitting Evidence

- Submit variant in CIViC with Google account
  - CIViC variants are automatically lifted over to DoCM
- Use batch submission page and a tab delimited file. See submission form here: http://docm.genome.wustl.edu/batch_submission_help
  - Make sure your variant complies with the criteria for inclusion (link also available on URL above)

DoCM Contact Information

- Malachi Griffith: mgriffit@wustl.edu
- Obi Griffith: obigriffith@wustl.edu