

VCF-DART

(VCF Diagnostic Annotation and Reporting Tool)

Dr Miles Benton

Senior Scientist Bioinformatics, ESR

(QRW Bioinformatics - 30th-31st August 2018)

Advocate for reproducible research...

Where possible my presentations and code are available online



github
SOCIAL CODING



sirselim.github.io/presentations



ESR (2nd July 2018 -)

Wellington (Porirua), NZ

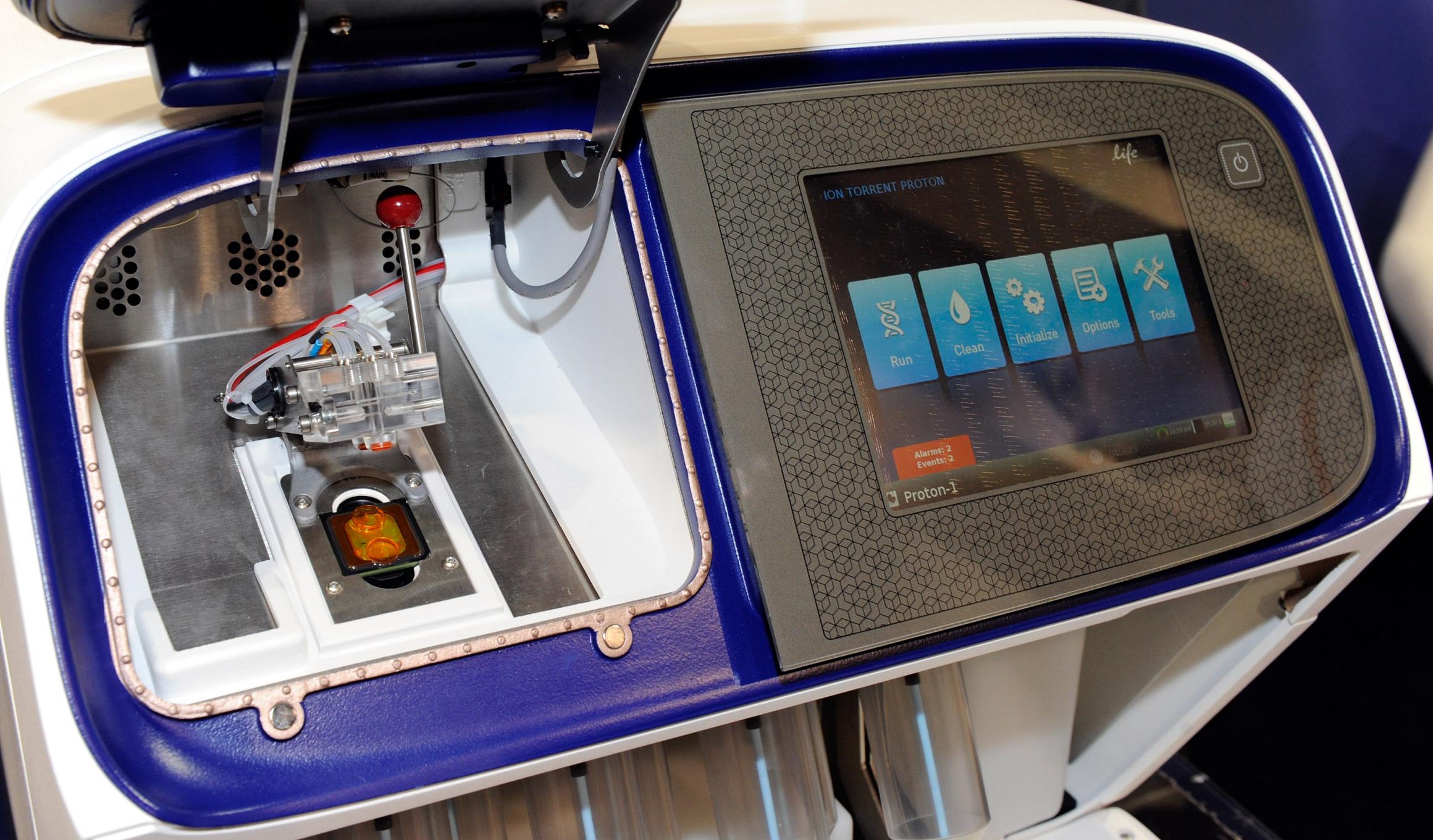
Senior Scientist Bioinformatics and Data Science



QUT (February 2014 - June 2018)

Brisbane (QLD), Australia

PostDoc Research Fellow



Diagnostics Lab

Gene panel and Exome Sequencing (NATA accredited - neurological disorders)

Sanger -> Gene panel -> Exome

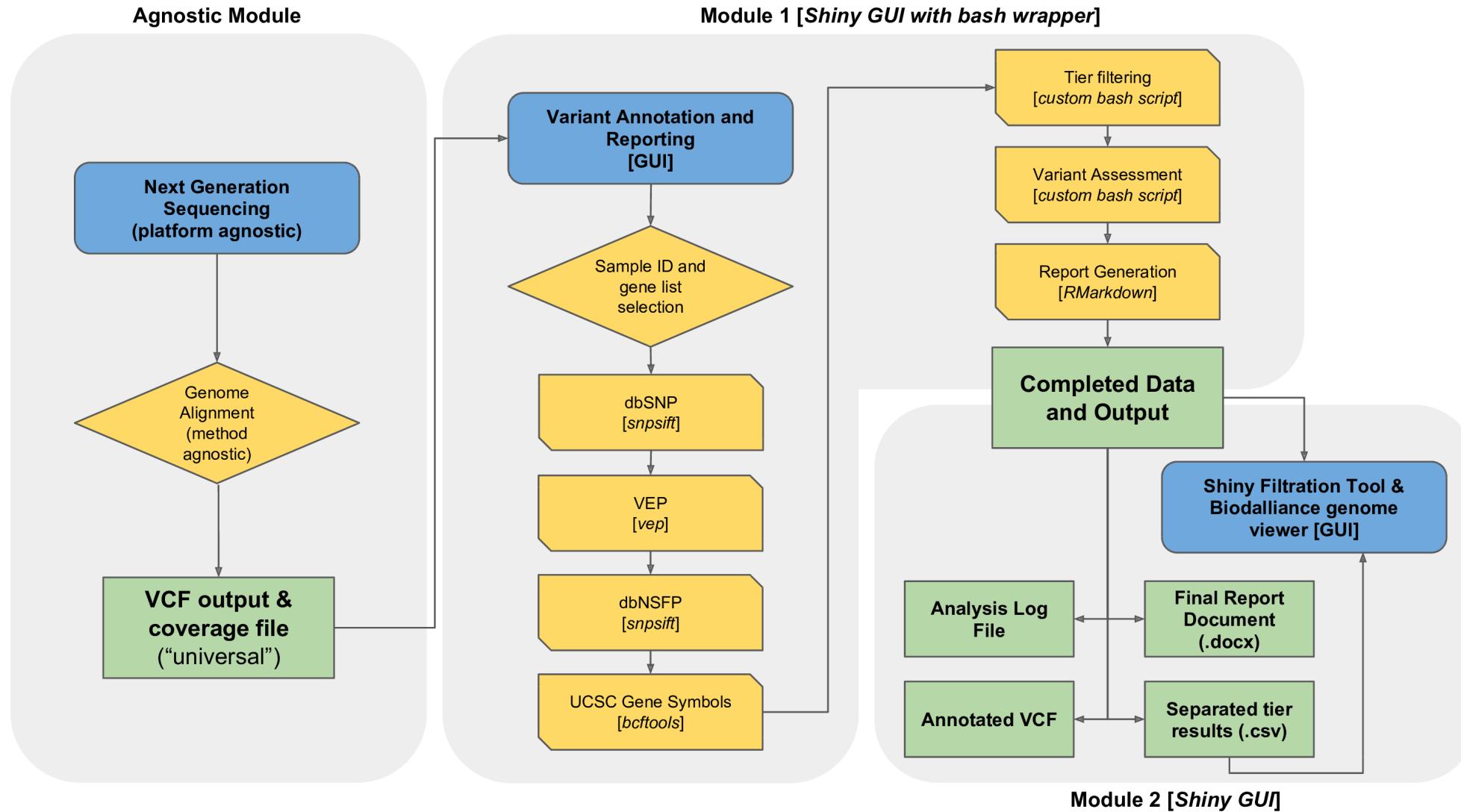
- used to manually work through 30-60K rows of data in Excel!
- could take 4-8 weeks to identify and validate variant(s)

VCF-DART - what it's not...

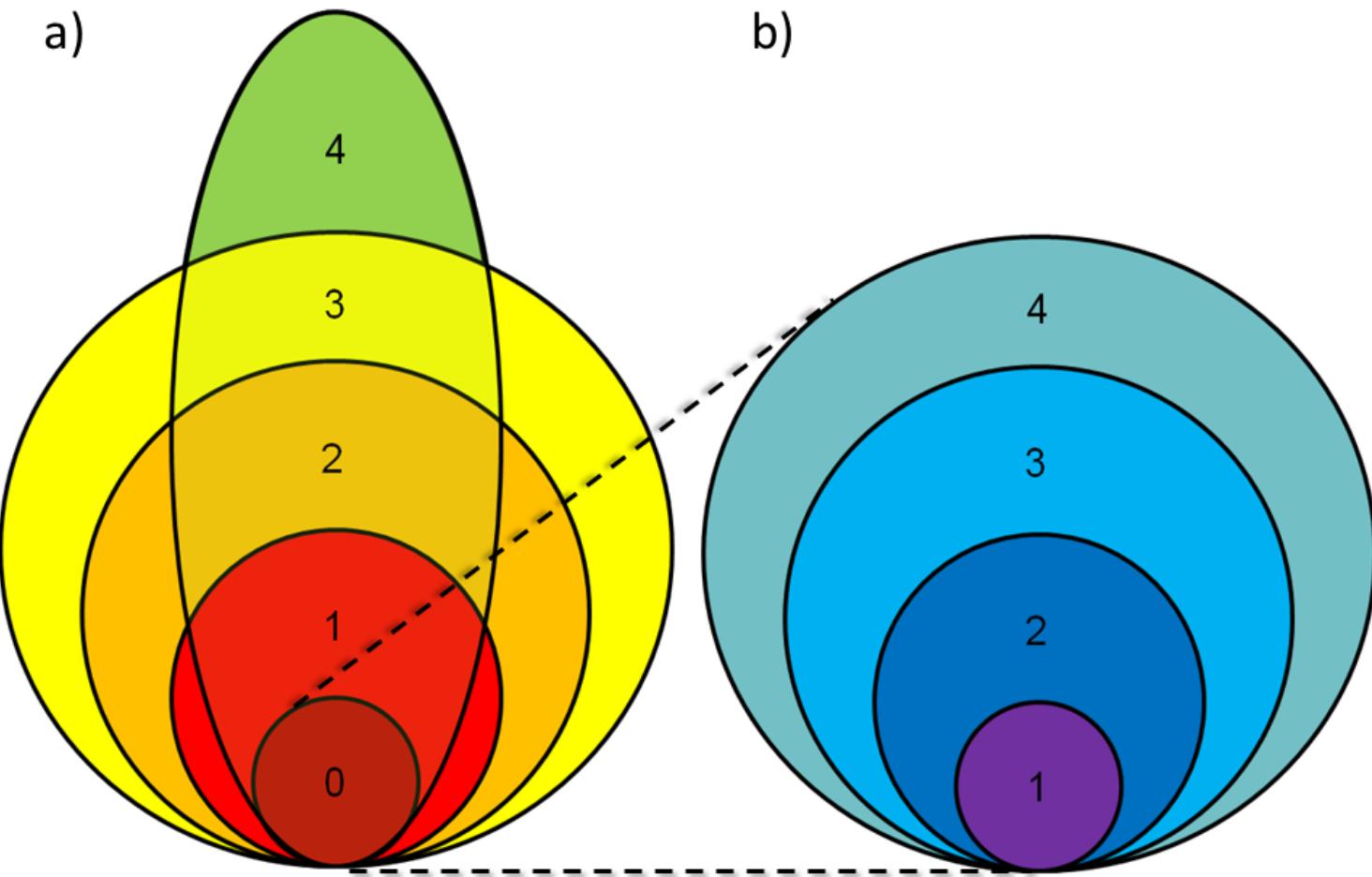
It is **NOT** a variant calling and quality control pipeline

... it's about data

Overview



Variant tiers



"Module 1" - VCF-DART

VCF-DART (VCF Diagnostics Annotation and Reporting Tool)

Enter details for annotation run and report generation.

Home Directory (location of data)

User Name

Sample ID

Label (i.e. barcode)

Run ID

Genome Build

Choose Tier 0 gene list

Upload complete

Choose Tier 1 gene list

Upload complete

Choose Tier 2 gene list

Upload complete

Output Directory Name

Update details
Click to update values displayed in the main panel.

VCF-DART **User upload**

Please review the details you entered below before proceeding.

Selected home dir: /home/ubuntu/
Selected user: Miles Benton
Selected sample: HG00103
Selected barcode: 103
Selected run ID: test_run_001
Selected genome build: hg19
Selected tier 0 gene list: gene_lists/diagnostics_tier0_panel.txt
Selected tier 1 gene list: gene_lists/diagnostics_tier1_list.txt
Selected tier 2 gene list: gene_lists/diagnostics_tier2_pathways_list.txt
Selected output directory: HG00101_20180830

If you are happy with the above please continue.

Go! Click the button to start analysis.

VCF-DART (VCF Diagnostics Annotation and Reporting Tool)

User upload panel...[under construction...]

VCF-DART

User upload

Upload data to server

Choose a VCF file to upload:

Browse...

No file selected

Choose a txt file to upload (QC/coverage information):

Browse...

No file selected

Runtime

30-50K variants (avg Proton exome VCF):

- 6-10 mins (24 core 256GB RAM)
- 17-23 mins (4 core 12GB RAM)

"Module 2" - VCF-DART Viewer

...live demonstration...

?

VCF-DART Viewer: visual results discovery for NGS variant data

Enter a SampleID and use the Tier tabs to filter exome variants based on the below variables.

Enter SampleID

HG00103

Columns in results to show:

- location
- genotype
- ref
- alt
- dbSNP
- gene
- transcript
- coding
- AAchange
- MutationTaster
- SIFT
- Polyphen2
- coverage
- ref_freq
- alt_freq
- GnomAD

Tier0 variants

Tier1 variants

Tier2 variants

Tier3 variants

MutationAssessor variants

GO search

Biodalliance Genome Viewer

Downloads

Help

NOTE: This tier consists of genes in metabolic pathways known to be involved in causing the conditions being investigated.

Show 10 entries

Search: Copy Print Download

location	genotype	dbSNP	gene	coding	AAchange	MutationTaster	SIFT	Polyphen2	coverage	ref_freq	alt_freq	GnomAD
chr1:46500345	G/A	rs141937880	MAST2	c.4004G>A	p.Arg1335His	D	D	D	97 G(45) A(51)	0.9994	0.000599	GnomAD Link
chr7:55268045	G/A	rs144496976	EGFR	c.2885G>A	p.Arg962His	D	D	P;D	99 G(54) A(45)	0.9994	0.000599	GnomAD Link
chr8:128750540	A/G	rs4645959	MYC	c.32A>G	p.Asn11Ser	D	D	D	59 A(31) G(27)	0.9848	0	GnomAD Link
chr9:124091559	C/T	rs144434647	GSN	c.1931C>T	p.Thr644Met	D	D	D	50 C(22) T(28)	0.9992	0.0007987	GnomAD Link
chr9:140777306	C/G	rs4422842	CACNA1B;AK128414	c.501C>G	p.Asn167Lys	D	D	D	130 C(65) G(58)	0.846	0.154	GnomAD Link
chr10:18828486	C/G	rs61733968	CACNB2	c.1732C>G	p.Arg578Gly	D	D	B;P;D	83 C(41) G(41)	0.9956	0.004393	GnomAD Link
chr12:21457434	T/G	rs11568563	SLCO1A2	c.516A>C	p.Glu172Asp	D	D	D	69 T(38) G(30)	0.9736	0.02636	GnomAD Link
chr14:68159269	C/T	rs80140987	RDH11	c.235G>A	p.Glu79Lys	D	D;T	P;D	198 C(103) T(93)	0.9892	0.01078	GnomAD Link
chr15:65113687	A/G	rs118062397	PIF1	c.850T>C	p.Cys284Arg	D	D	D	80 A(44) G(35)	0.9842	0.01577	GnomAD Link

All

.

A

All

(X)

p (X)

D (X)

(X)

(X)

(X)

(X)

(X)

(X)

All

Table 3: A list of variants passing the filter criteria for Tier2.

Showing 1 to 9 of 9 entries (filtered from 2,950 total entries)

Previous

1

Next

This is a beta test of the redesigned gene/region pages. If you would like to report bugs or provide suggestions, please post an issue on [github](#) or [email us](#). Your input is greatly appreciated!

Variant: 10:18828486 C / G

Note: This variant is multiallelic! The other alt alleles are:

- 10-18828486-C-T

	Exomes	Genomes	Total
Filter	Pass	Pass	
Allele Count	2236	305	2541
Allele Number	246074	30804	276878
Allele Frequency	0.009087	0.009901	0.009177
dbSNP	rs61733968		
UCSC	10-18828486-C-G ↗		
ClinVar	Click to search for variant in Clinvar ↗		

Genotype Quality Metrics

Site Quality Metrics

Report this variant

Annotations

This variant falls on 11 transcripts in 2 genes:

missense

- CACNB2

Transcripts ↴

intron

- RP11-499P20.2 - ENST00000425669 *

ENST00000282343 (p.Arg578Gly)

Polyphen: probably_damaging; SIFT: tolerated_low_confidence

ENST00000324631 * (p.Arg606Gly)

Polyphen: possibly_damaging; SIFT: tolerated_low_confidence

ENST00000352115 (p.Arg582Gly)

Polyphen: possibly_damaging; SIFT: tolerated_low_confidence

ENST00000377315 (p.Arg558Gly)

Polyphen: probably_damaging; SIFT: tolerated_low_confidence

ENST00000377319 (p.Arg513Gly)

Polyphen: possibly_damaging; SIFT: tolerated_low_confidence

ENST00000377328 (p.Arg356Gly)

Polyphen: probably_damaging; SIFT: tolerated

ENST00000377329 (p.Arg552Gly)

Polyphen: probably_damaging; SIFT: tolerated

ENST00000377331 (p.Arg554Gly)

Polyphen: possibly_damaging; SIFT: tolerated_low_confidence

ENST00000396576 (p.Arg551Gly)

Polyphen: possibly_damaging; SIFT: tolerated

Read Data

This interactive

Note: These are re

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (Non-Finnish)	2022	126512	12	0.01598
Other	70	6460	1	0.01084
Latino	177	34418	0	0.005143
European (Finnish)	124	25784	1	0.004809
African	64	23964	0	0.002671
Ashkenazi Jewish*	27	10138	0	0.002663
South Asian	57	30780	0	0.001852
East Asian	0	18822	0	0.000
Total	2541	276878	14	0.009177

Include: Exomes Genomes

* For detailed analysis of Ashkenazi Jewish frequency see the IBD Exomes Browser.

represent what HaplotypeCaller was seeing when it called this variant.

chr10:18,828,446-18,828,526



81 bp

hide labels



VCF-DART Viewer: visual results discovery for NGS variant data

This tab allows you to enter a gene, or list of genes (comma separated), and retrieve a list of GO terms associated with each gene.

i.e. input: "DDO, FTO, ATXN1" and press submit.

Note: as this is accessing the GO servers it can take a min or two, please be patient.

Tier0 variants Tier1 variants Tier2 variants Tier3 variants MutationAssessor variants GO search Biocatalliance Genome Viewer Downloads Help

Enter gene symbol

CACNB2

Submit

Your selected gene(s):

[1] "CACNB2"

Show 10 ▾ entries

Search:

Copy

Print

Download

gene	description	uniprot	goterm	pubmed	ontology
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0005245	1309651	voltage-gated calcium channel activity
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0005262	9594024	calcium channel activity
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0005515	17525370	protein binding
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0005886		plasma membrane
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0005887	9594024	integral component of plasma membrane
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0005891	17224476	voltage-gated calcium channel complex
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0007268		chemical synaptic transmission
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0007528	8494331	neuromuscular junction development
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0007601		visual perception
CACNB2	Voltage-dependent L-type calcium channel subunit beta-2	Q08289	GO:0051015	25533460	high voltage-gated calcium channel activity

All

All

All

All

All

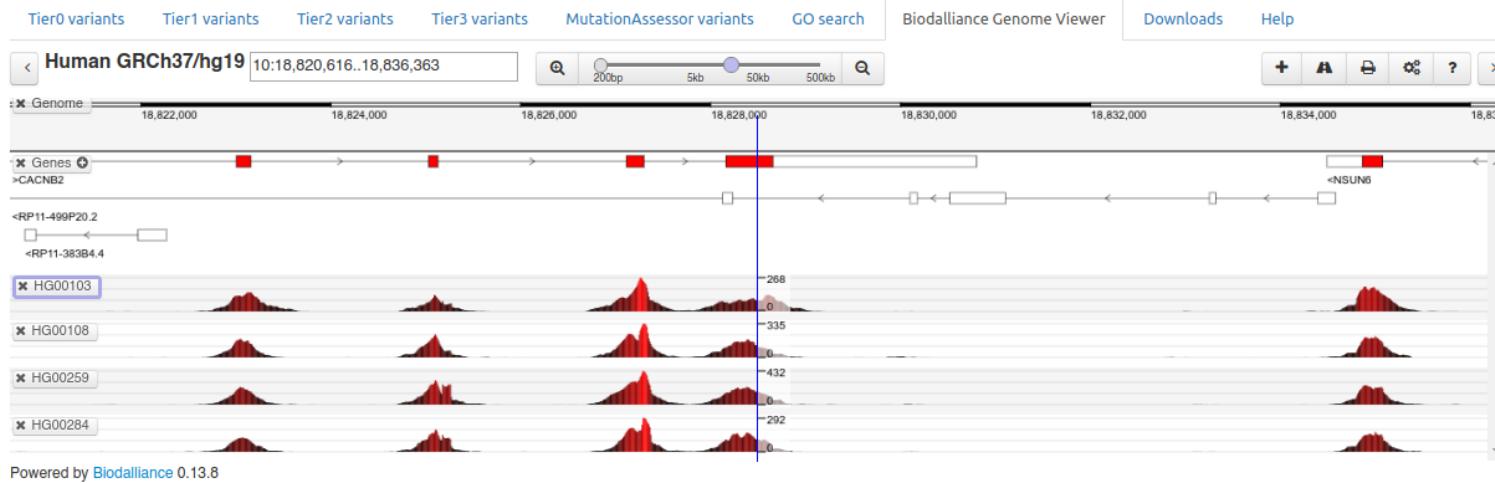
All

Table 6: A list of GO terms for selected gene(s).

VCF-DART Viewer: visual results discovery for NGS variant data

Biodalliance Embedded Genome Viewer.

NOTE: this is now functioning but still under development.



VCF-DART Viewer: visual results discovery for NGS variant data

Biodalliance Embedded Genome Viewer.

NOTE: this is now functioning but still under development.



VCF-DART Viewer: visual results discovery for NGS variant data

Downloads section (TESTING).

NOTE: this section is still under development.

[Tier0 variants](#)[Tier1 variants](#)[Tier2 variants](#)[Tier3 variants](#)[MutationAssessor variants](#)[GO search](#)[Biodalliance Genome Viewer](#)[Downloads](#)[Help](#)

Clinical Report Document

Click below to download the generated clinical report (.docx) for this sample.

 [Download Report](#)

Run Log File

Click below to download the log file (.log) for this samples annotation run.

 [Download Log File](#)

All files (compressed)

Click below to download a compressed file (tar.gz) containing the run directory with all data and results.

 [Download All Data](#)



Spun up VM to give reviewers a chance to ~~break~~ test things.

- using 6 public exomes (1000G)



nectarcloud

Spun up VM to give reviewers a chance to ~~break~~ test things.

- using 6 public exomes (1000G)

Allocated VM:

- Ubuntu 18.08
- 4 cores
- 12GB RAM
- 30GB root
- 120GB ephemeral disk



nectarcloud

Spun up VM to give reviewers a chance to ~~break~~ test things.

- using 6 public exomes (1000G)

Allocated VM:

- Ubuntu 18.08
- 4 cores
- 12GB RAM
- 30GB root
- 120GB ephemeral disk

...so it can scale down to quite reasonable specs...

VCF-DART in the lab

- now can have a short list of variants in **1 day**
- validation in **<2 weeks**

to-do

- submit manuscript - out for review (Genetics in Medicine)
- release to the wild (anyone interested can help out)
- finish implementing ClinGen links
- option for GnomAD-beta
- refactoring main code base (bpipe)
- complete docker version
- continue developing documentation!
- work with clinicians to further develop and refine

acknowledgements



Dr Robert Smith (QUT)



Dr Donia Macartney-Coxson (ESR)

Craig Windell (QUT)

Sam Beardman

QRISCloud / Nectar

Thank you!