NCGM
分析服務及註解介紹

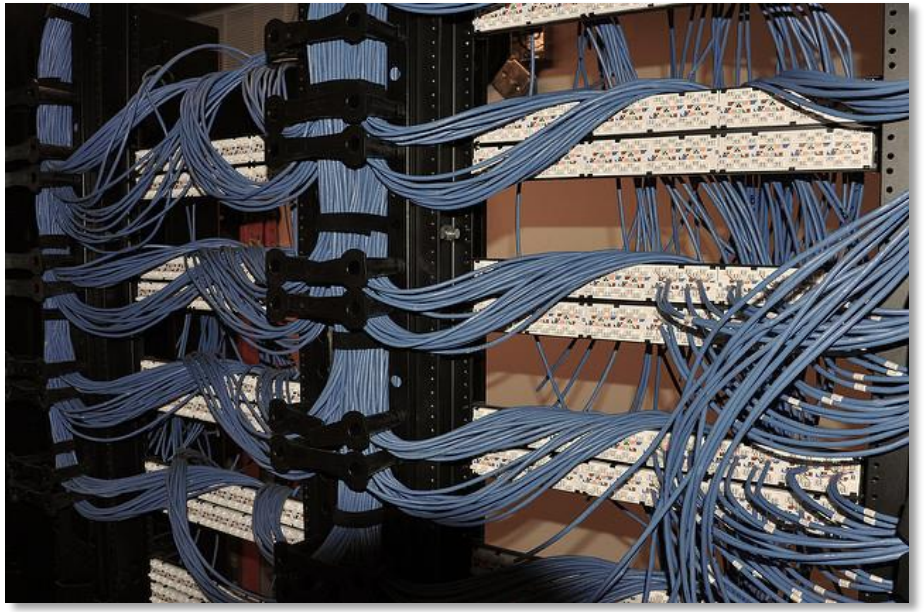
吳行中

Agenda

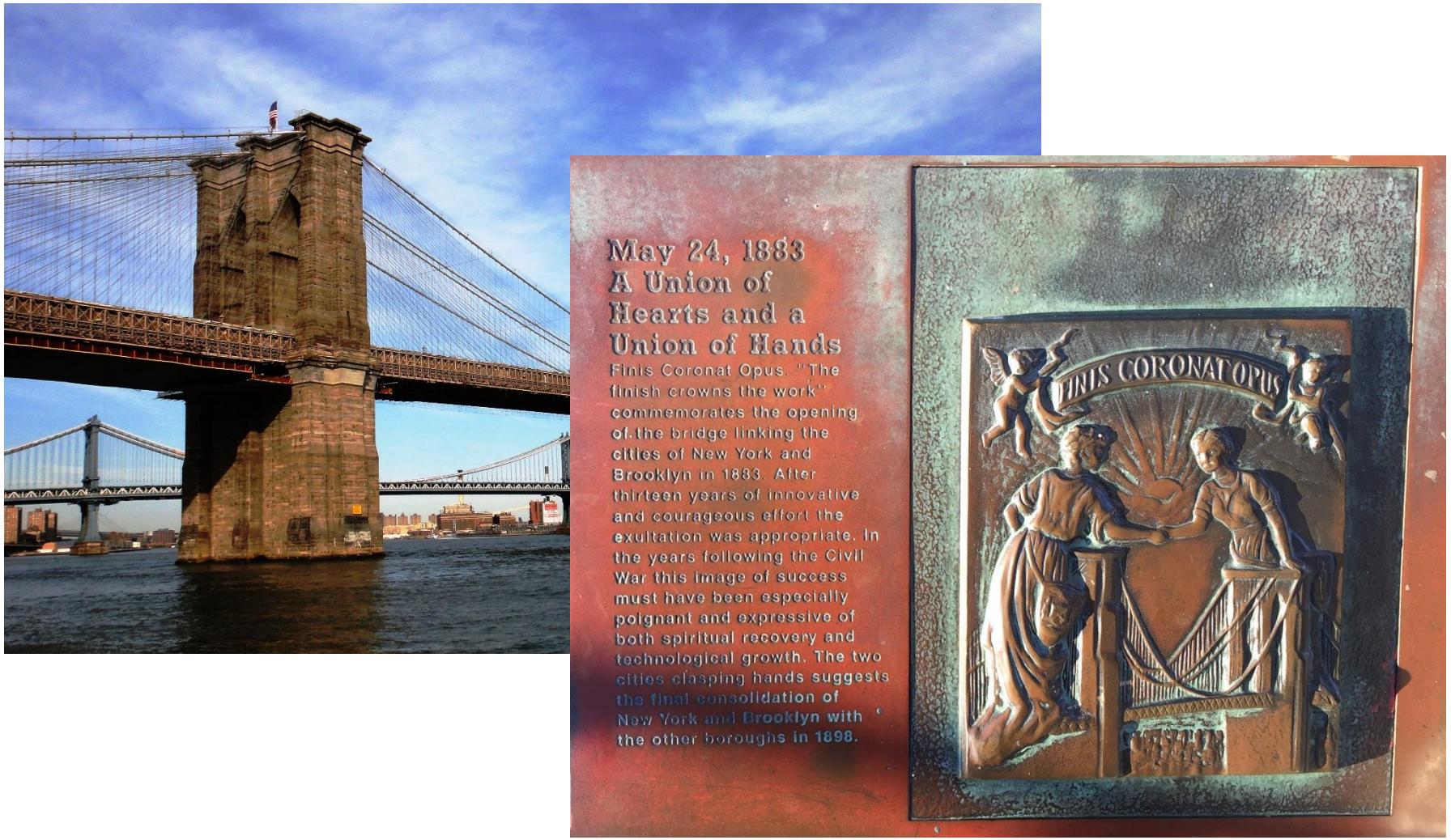
- Introduction
- System
  - HiPipe Basic
  - HiPipe Professional
- Annotation
  - VarioWatch
  - ClinVar
  - Cosmic
- Customized Analysis
- Conclusion

Wet Lab and Dry Lab





A Union of Hearts and a Union of Hands

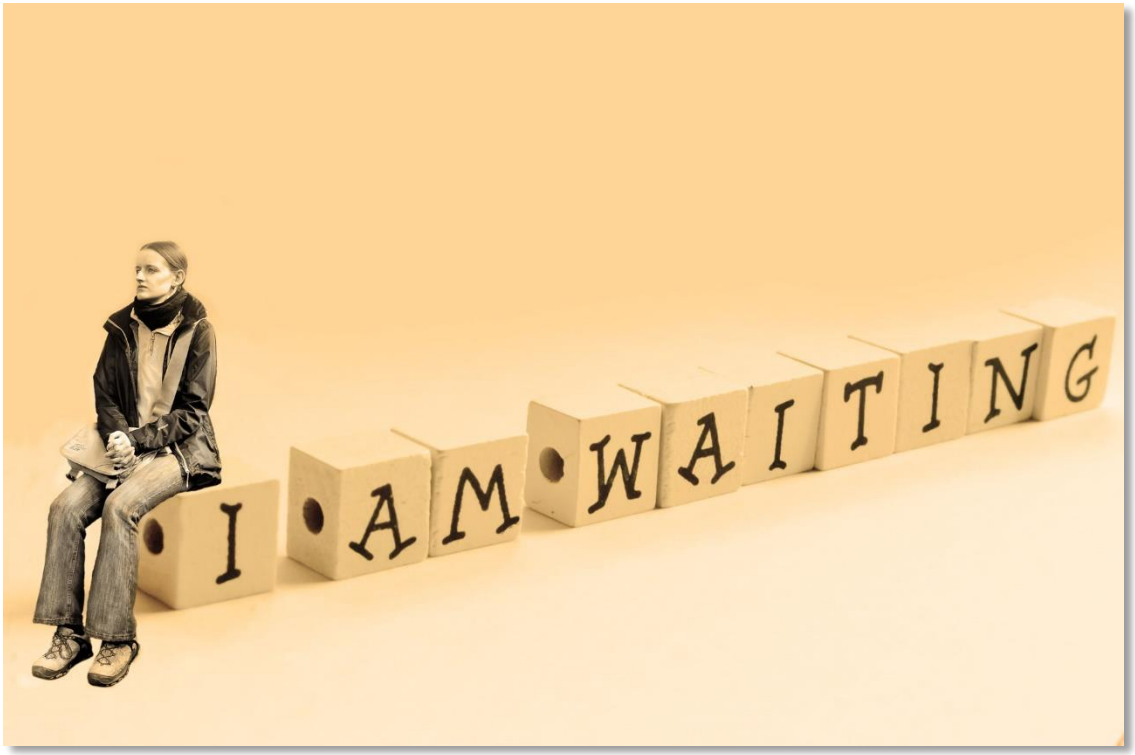


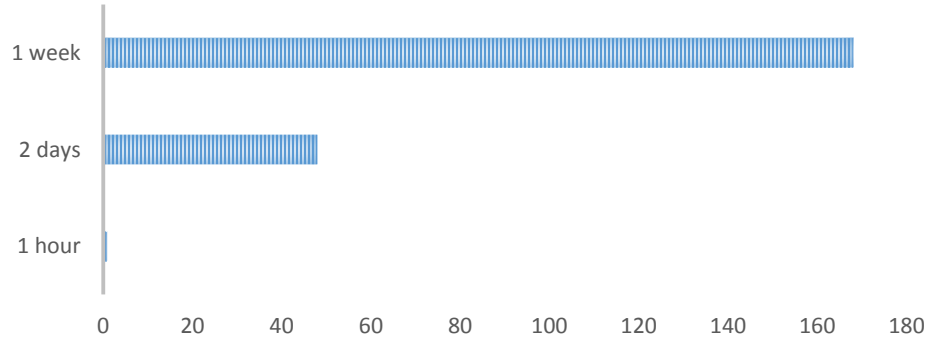
Most analysis tools run only in the Linux environment



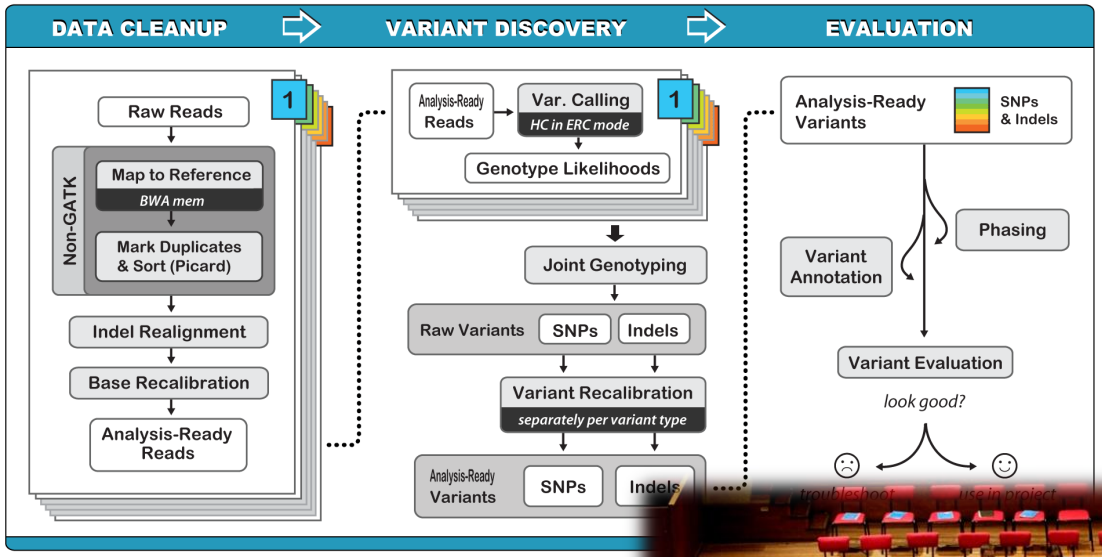

```
[root@localhost ~]# ping -q fa.wikipedia.org
PING text.pmtpa.wikimedia.org (208.80.152.2) 56(84) bytes of data.
^C
--- text.pmtpa.wikimedia.org ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 0ms
rtt min/avg/max/mdev = 540.528/540.528/540.528/0.000 ms
[root@localhost ~]# pwd
/root
[root@localhost ~]# cd /var
[root@localhost var]# ls -la
total 72
drwxr-xr-x. 18 root root 4096 Jul 30 22:43 .
drwxr-xr-x. 23 root root 4096 Sep 14 20:42 ..
drwxr-xr-x.  2 root root 4096 May 14 00:15 account
drwxr-xr-x. 11 root root 4096 Jul 31 22:26 cache
drwxr-xr-x.  3 root root 4096 May 18 16:03 db
drwxr-xr-x.  3 root root 4096 May 18 16:03 empty
drwxr-xr-x.  2 root root 4096 May 18 16:03 games
drwxrwx--T.  2 root gdm  4096 Jun  2 18:39 gdm
drwxr-xr-x. 38 root root 4096 May 18 16:03 lib
drwxr-xr-x.  2 root root 4096 May 18 16:03 local
lrwxrwxrwx.  1 root root   11 May 14 00:12 lock -> ../run/lock
drwxr-xr-x. 14 root root 4096 Sep 14 20:42 log
lrwxrwxrwx.  1 root root   10 Jul 30 22:43 mail -> spool/mail
drwxr-xr-x.  2 root root 4096 May 18 16:03 nis
drwxr-xr-x.  2 root root 4096 May 18 16:03 opt
drwxr-xr-x.  2 root root 4096 May 18 16:03 preserve
drwxr-xr-x.  2 root root 4096 Jul  1 22:11 report
lrwxrwxrwx.  1 root root    6 May 14 00:12 run -> ../run
drwxr-xr-x. 14 root root 4096 May 18 16:03 spool
drwxrwxrwt.  4 root root 4096 Sep 12 23:50 tmp
drwxr-xr-x.  2 root root 4096 May 18 16:03 yp
```

It takes a long time to get a result





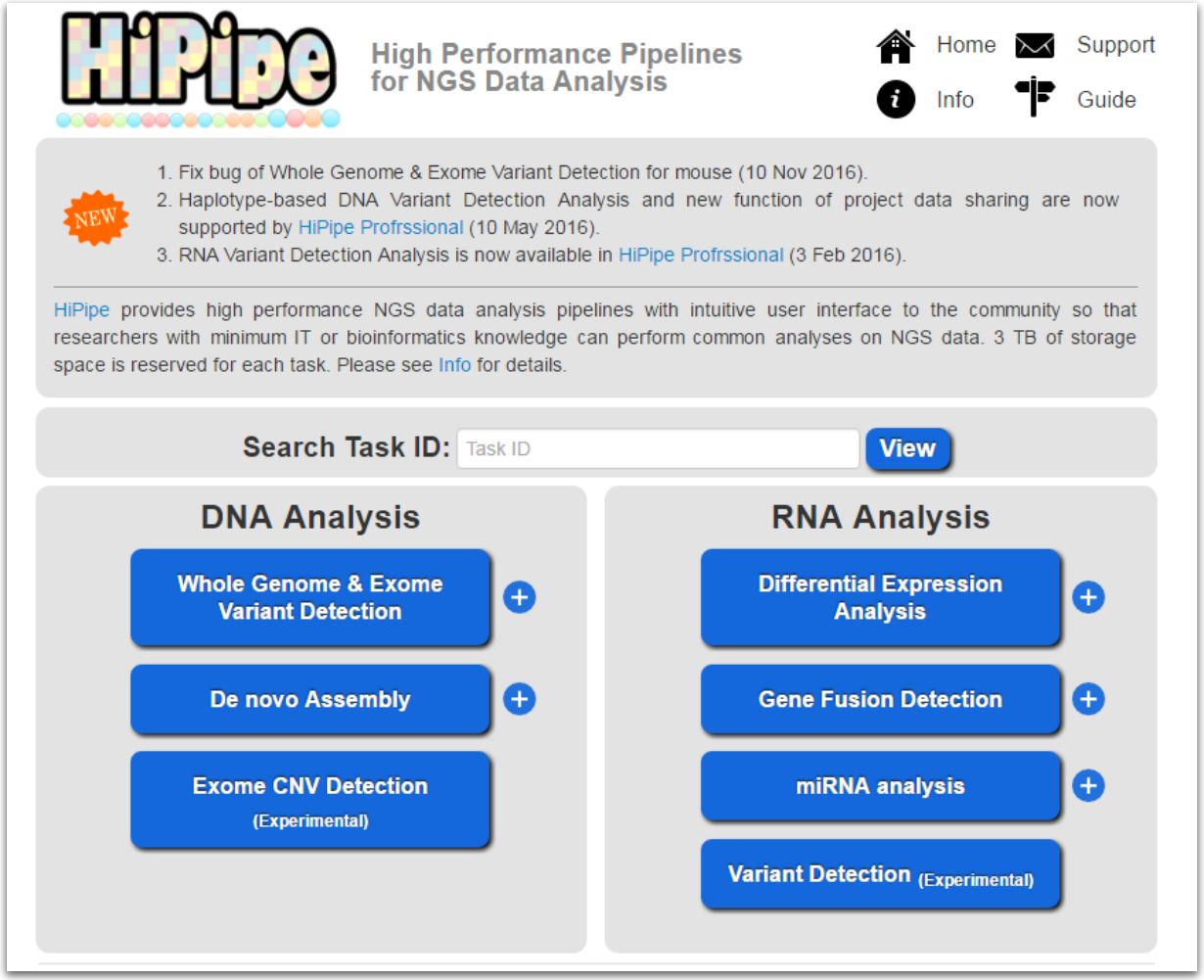
It is error-prone to orchestrate different analysis tools



(http://cdn.vanillaforums.com/gatk.vanillaforums.c
a74b64ba47b02d1bae.png)

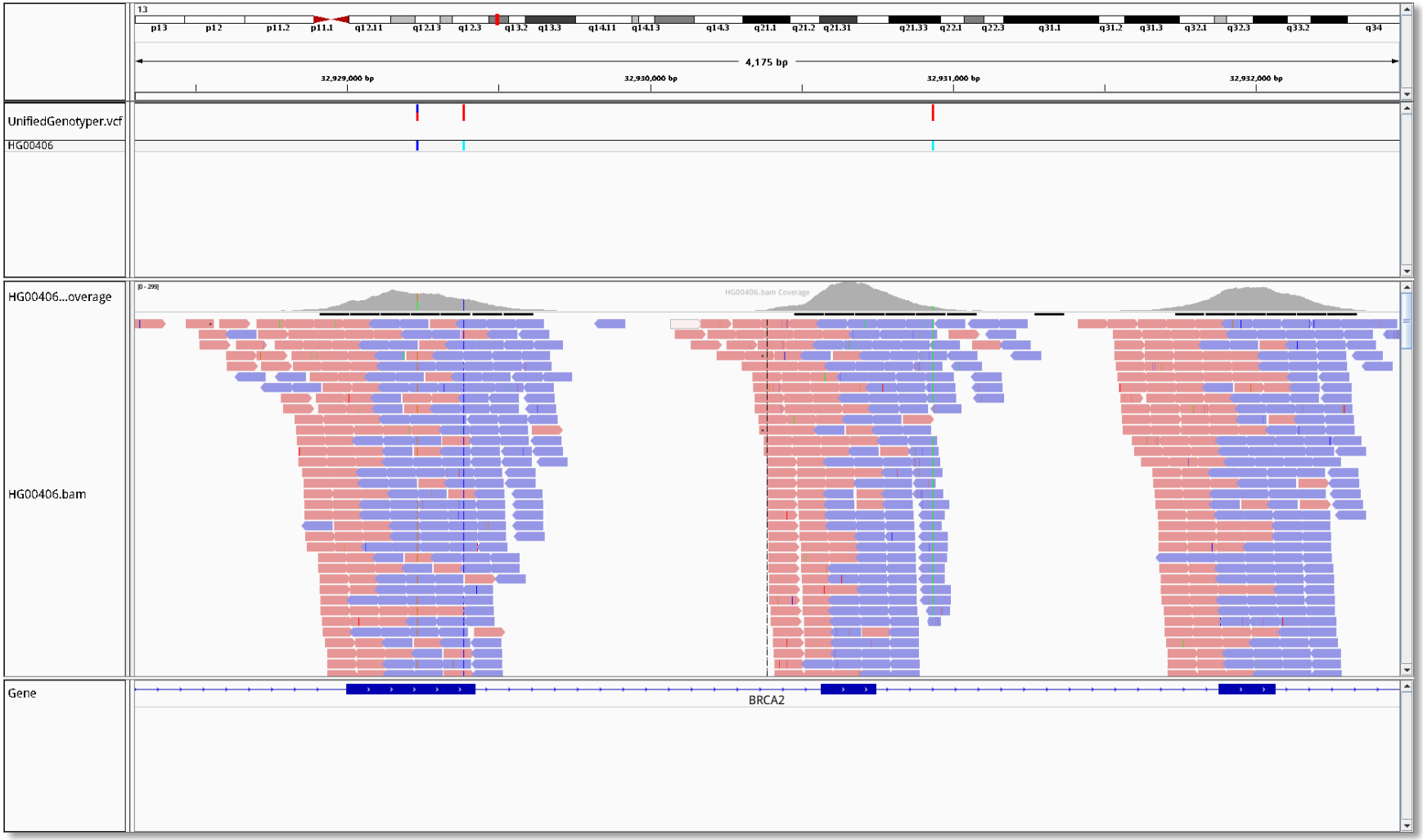


Analysis Online

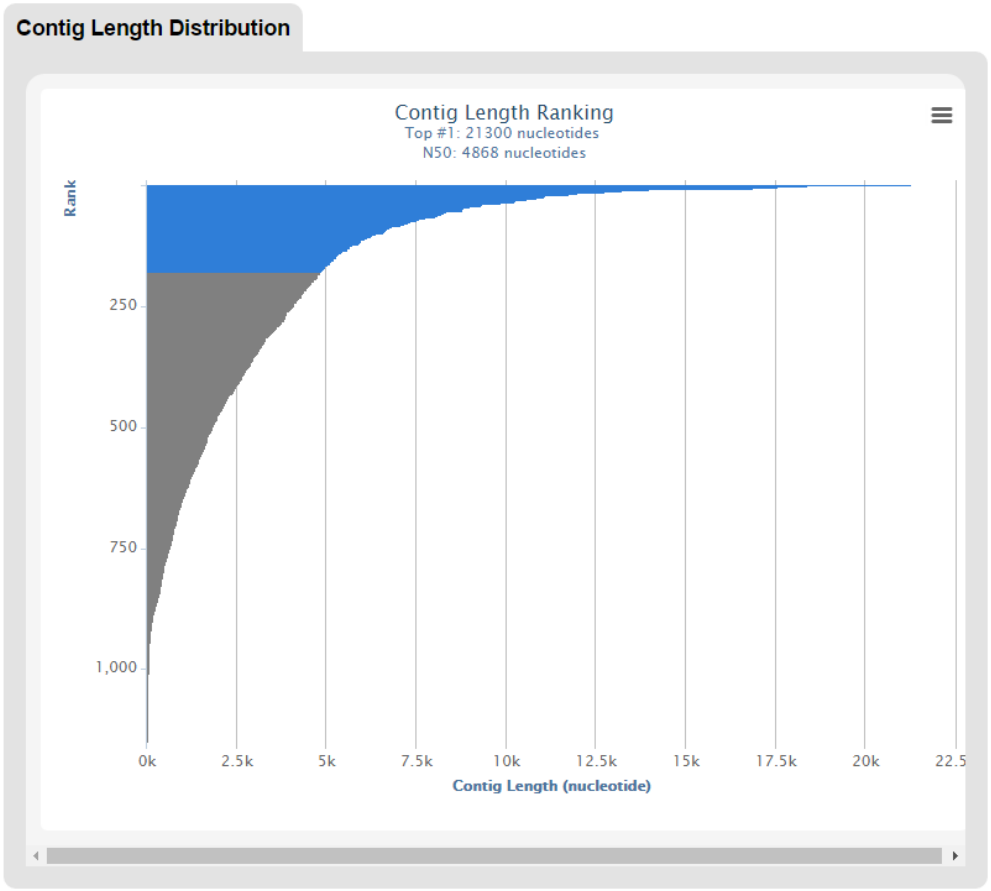


(http://hipipe.ncgm.sinica.edu.tw/)

DNA Variant Detection



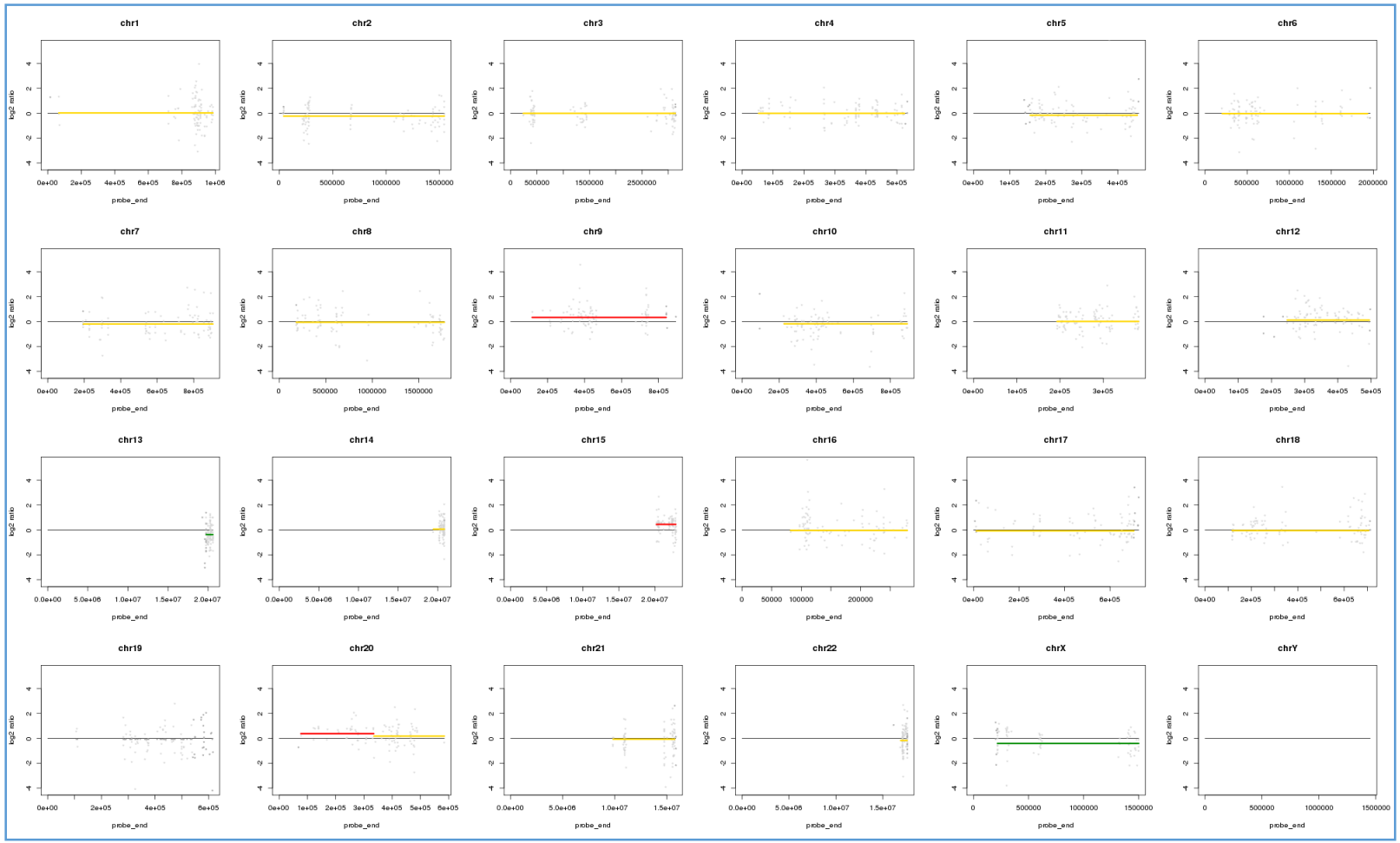
DNA: De Novo Assembly



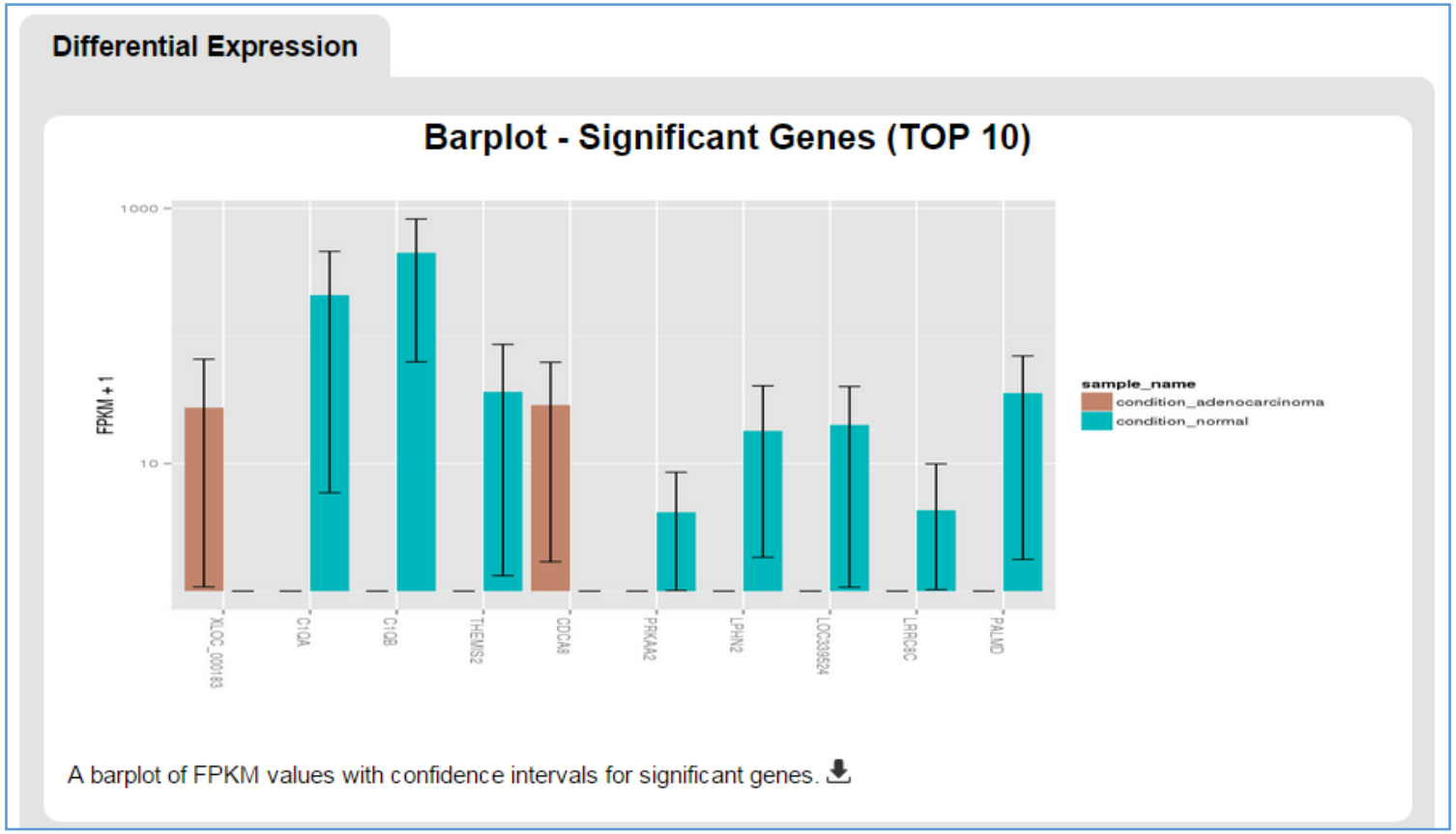
contigs.fa

```
>NODE_4_length_3695_cov_11.429500
CCATGCGCAGCACTTCTCTTTAAATTGACATTGGCACAAACGCCACCACCAAAAAGGATT
AAGATTGCAGTTCCTAGGAATTCTGCTAAATATACATTCATATTAAATGTCCTCCCAAAA
AGACAAAGAGATTCCTACACTAGCAGTTTGGCTCGTAGAAATCTCCCGTCTCTATTCTAT
GTATTAACTTTTATTATATTTGTTATGATACTATCGATATGAAAGCGTTGTCAATGGTTT
TTGTAAAAATTTTGTCAAATTTAATTTTTCAAGTCTTTATAAAAGTAGTTTAATTTGAAA
ACTAGAAATACCAATCCTAACTAATTTATTTATGACGCTCTTTTGTTTAATTGACATTCT
TCAGCCATATTTTTTATGACTAGTTGCATTTTTTACTAATAAACACCTCTAAAACTTTAA
TGATTTTAATCGTTTTAGAGGTGCTAATTATTTTATTTGATTATTTTTTGTTTGATACCT
ACTGCATATCCCATATGAAAACGGCTTTTTTATTATGTTATATGACTAAATCTCGTGAAA
AATGAAATTTTTGCAGACTTATGATTTACCAAAGTTTATCATAACTAGTAGTTACATATT
TCGCTCCATTTTTAACAGCTTCATTGACTTCATCTATTGTATTAATTAGGCCACCTGCAA
TGACTTGTGTGTTTGTTTCTTTCTGAATATGATGAATCGCTTTACTCGCAACACCTGGAA
GTACTTCAACAAAATCAGGTTCAACTTTTTTTATCAAATCTATACTGCGTTTCAATGCTT
GACTATCAATAATAAATACTCTAAAAATCGTTAAAGTATTTAATGATTTAGCTTTTTTTA
TTACTTTAGATTTAGTCGATACGATACCTTTTGGCTTGTATTGCTGAATAATAAATTCAC
TTGCAAATTCATCGTGGCTTAAACCTTTTATCAAATCTATATGAATAAAACACTCTATAT
GATTTTGCTTCAGCAATTCCATAATACTTTTTATATGTCCTATATGCATATCTAGAAGCA
CACACATTTTATAGTCTGTTTTAATCAGTTTCTCTAAATCTTTAATGTTTCTTATAGCAG
GCAATATGTTGTTATTCACTTGATTCATCCTCTCTACATCACACGCTTAAATAATTTTTC
TAATTCGTATTTTGAAAAATTAATGATTATTGGACGACCATGTGGACATGTAAATGGATC
TTCCGCTTCTCTTAATTGATCAATTAAATCAGACATTTCATGTTTTTGTAAATAATGATT
CGCTTTAATGGATTTTTTACATGACATCATGATTGCTACATCTTCACGTAATTTTTTGAT
ATCTACTTTTTTCTCTTCCAAAATTAGCTCAATCATATCTTTAATAATTTCTTCTACTTC
ATCTTTAGGGAACCAAACTGGATAGCTACTTACAATATAATCATGACCACCAAAATGTTC
TAACATGATACCTACTTGTTGAAGCTCATTTTTATATTGATCAATGACTAATTGTTCATC
TTTTGAAAAATGAAATGTTAACGGGATTAATAAATCTTGTACTTCATTGGTAACCTCACC
TATTTTATCTCGAAAATATTCATATTTTATTCTTTCTTGAGCTGCATGCTGGTCAATCAT
GTACATGCCAAATTCATTTTGAGCAATAATATACGTTCCATGTACTTGTCCAACAATTTC
CATATATGGAATTCTTCGTTTAGGGTCTTTACTTTTTGTTCCTTTAATATCGTTTACATC
ATAATAATCATCAGACGCTTTATTCTCATAATTTTGCACAGTTGTACCATCATTTGAATC
AAATGTTTGATCCATTTCTTGTAAAATTTCTTGTTGTTTTTTGAAGTAGTCTTCATCTTT
TTCACGTGTATTATTATAACCAGAAGATTCATCATCAGCTATTAAACTTTCAGATGTTTT
CGTTACGAATGGATTATATGAATCTTCTTTTATAACTATCTCATCGTTTTGATTTTCTTC
CATAAATGGCTTACTGTTACTTTCTTCAGATGAAAACGTCTTCTCTTGGTTATTCTCTGT
GTTTTGTCTTTGTTCAAATTCGATTTTTTGTTGTTCGAATGAATGTAATACTTTATTTTT
TTTTGGCACATAATCCAAGTTATTTTTAGGAATTAATATACGGTCTTTAAATGCTTCTTG
TATTTTGCTCACAATCAATTGGTATAGTTGCTCTTCTTTTGATAAACGCACTTCTAGTTT
TGTTGGATGAACATTTACGTCTACTAAGATTGGATCCATTTCAATATTAATATAACAAAT
CGGGAACCTACCTATTGTTAAGAGTGTATGATAGCCTTCTAAAATCGCTTTATTTAGCAT
```

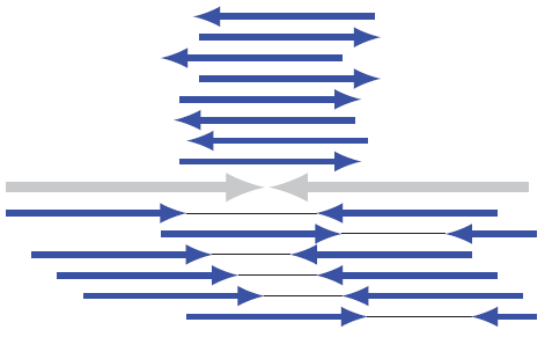
DNA: Exome CNV



RNA: Differential Expression



Gene Fusion



(a) supporting reads in blue and contradicti

chr20-chr

Left Side		
chr	pos	gene symbol
chr20	49411709	BCAS4(exon1) 49411465-49411709

Seq: CTCTTCCTGACCCCCGAGCCTGGGGCCGAG

Dep: 99999999999999999999999999999

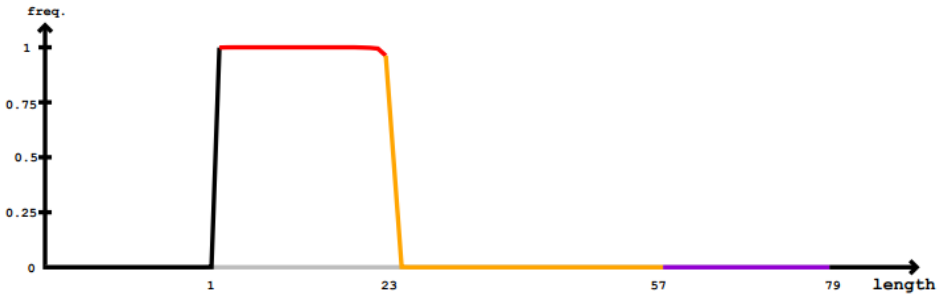
Spanning reads: 103
Spanning mate pairs: 113
Spanning mate pairs where one end spans a fusion: 17
Fusion score: 551.41

```
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGGCCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411676 59445703 34M59445687F16M    CGCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAG
chr20 chr17 49411677 59445704 33M59445687F17M     GCGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGC
chr20 chr17 49411678 59445705 32M59445687F18M      CGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCG
chr20 chr17 49411678 59445705 32M59445687F18M      CGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCG
chr20 chr17 49411678 59445705 32M59445687F18M      CGCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCG
chr20 chr17 49411679 59445706 31M59445687F19M       GCTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGT
chr20 chr17 49411680 59445707 30M59445687F20M        CTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTG
chr20 chr17 49411680 59445707 30M59445687F20M        CTCTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTG
chr20 chr17 49411680 59445707 30M59445687F20M        CTCTTCCCGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTG
chr20 chr17 49411682 59445709 28M59445687F22M          CTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGAC
chr20 chr17 49411682 59445709 28M59445687F22M          CTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGAC
chr20 chr17 49411682 59445709 28M59445687F22M          CTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGAC
chr20 chr17 49411682 59445709 28M59445687F22M          CTTCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGAC
chr20 chr17 49411684 59445711 26M59445687F24M            TCCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCC
chr20 chr17 49411685 59445712 25M59445687F25M             CCTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCT
chr20 chr17 49411686 59445713 24M59445687F26M              CTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCCG
chr20 chr17 49411686 59445713 24M59445687F26M              CTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTG
chr20 chr17 49411686 59445713 24M59445687F26M              CTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTG
chr20 chr17 49411686 59445713 24M59445687F26M              CTGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTG
chr20 chr17 49411687 59445714 23M59445687F27M               TGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGC
chr20 chr17 49411687 59445714 23M59445687F27M               TGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGC
chr20 chr17 49411687 59445714 23M59445687F27M               TGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGC
chr20 chr17 49411687 59445714 23M59445687F27M               TGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGC
chr20 chr17 49411687 59445714 23M59445687F27M               TGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGC
chr20 chr17 49411687 59445714 23M59445687F27M               TGACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGC
chr20 chr17 49411688 59445715 22M59445687F28M                GACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCT
chr20 chr17 49411688 59445715 22M59445687F28M                GACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCA
chr20 chr17 49411688 59445715 22M59445687F28M                GACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCC
chr20 chr17 49411688 59445715 22M59445687F28M                GACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCT
chr20 chr17 49411689 59445716 21M59445687F29M                 ACCCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTG
chr20 chr17 49411691 59445718 19M59445687F31M                   CCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGA
chr20 chr17 49411691 59445718 19M59445687F31M                   CCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGA
chr20 chr17 49411691 59445718 19M59445687F31M                   CCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGA
chr20 chr17 49411691 59445718 19M59445687F31M                   CCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGA
chr20 chr17 49411691 59445718 19M59445687F31M                   CCCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGA
chr20 chr17 49411692 59445719 18M59445687F32M                    CCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAG
chr20 chr17 49411692 59445719 18M59445687F32M                    CCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAG
chr20 chr17 49411692 59445719 18M59445687F32M                    CCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAG
chr20 chr17 49411692 59445719 18M59445687F32M                    CCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAG
chr20 chr17 49411692 59445719 18M59445687F32M                    CCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGGCCCTGCTGGAG
chr20 chr17 49411692 59445719 18M59445687F32M                    CCCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAG
chr20 chr17 49411693 59445720 17M59445687F33M                     CCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGG
chr20 chr17 49411693 59445720 17M59445687F33M                     CCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGG
chr20 chr17 49411693 59445720 17M59445687F33M                     CCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGG
chr20 chr17 49411693 59445720 17M59445687F33M                     CCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGG
chr20 chr17 49411693 59445720 17M59445687F33M                     CCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGG
chr20 chr17 49411693 59445720 17M59445687F33M                     CCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGG
chr20 chr17 49411693 59445720 17M59445687F33M                     CCGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGG
chr20 chr17 49411694 59445721 16M59445687F34M                      CGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGGT
chr20 chr17 49411694 59445721 16M59445687F34M                      CGATCCAGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGGT
chr20 chr17 49411694 59445721 16M59445687F34M                      CGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGGT
chr20 chr17 49411694 59445721 16M59445687F34M                      CGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGGT
chr20 chr17 49411694 59445721 16M59445687F34M                      CGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGGT
chr20 chr17 49411694 59445721 16M59445687F34M                      CGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGGT
chr20 chr17 49411694 59445721 16M59445687F34M                      CGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGGT
chr20 chr17 49411694 59445721 16M59445687F34M                      CGATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGAGGT
chr20 chr17 49411696 59445723 14M59445687F36M                        ATCCTGGGGCCGAG GTACCTTTGACAGGAGCGTGACCCTGCTGGCGGTGT
```

miRNA

Provisional ID	: 9_3536
Score total	: 319108.4
Score for star read(s)	: 3.9
Score for read counts	: 319097.6
Score for mfe	: 2.2
Score for randfold	: 1.6
Score for cons. seed	: 3
Total read count	: 625908
Mature read count	: 625898
Loop read count	: 0
Star read count	: 10



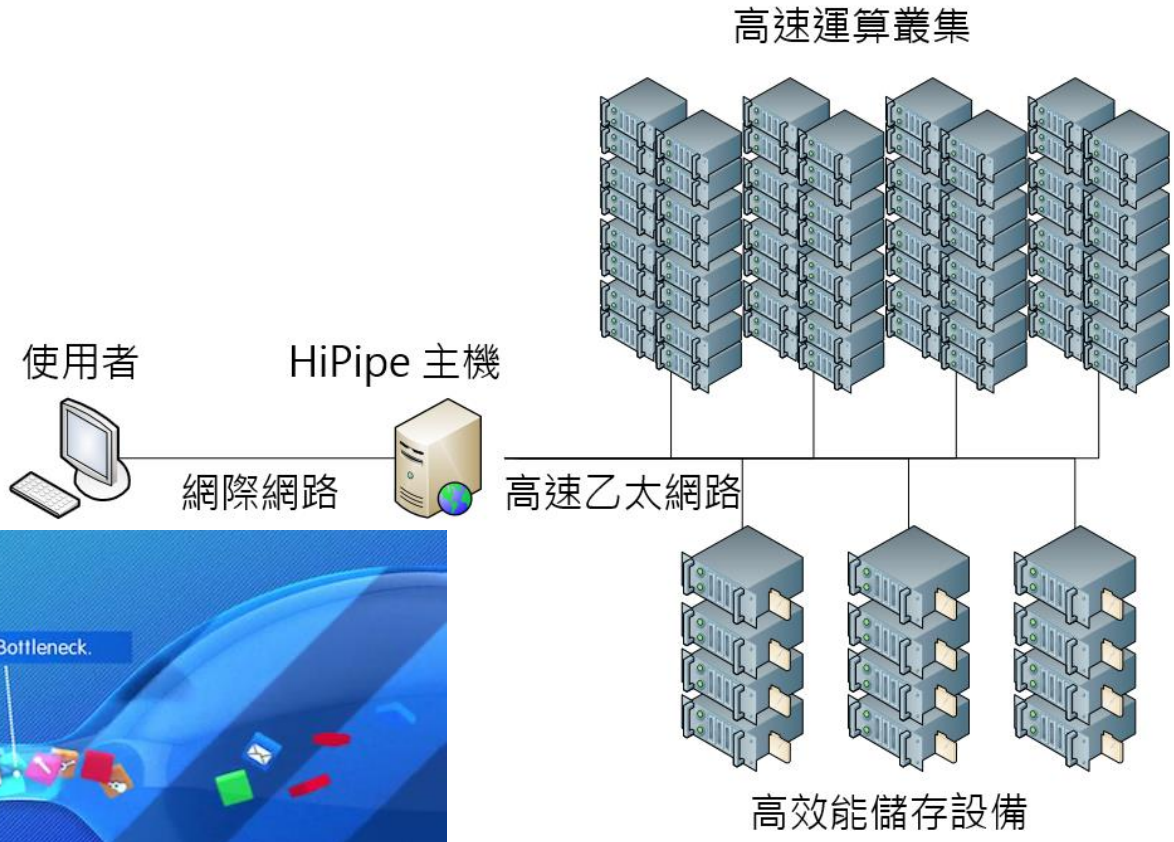


Mature Star

```
5'- aaaacauugcucuaucagagugagguaguagauuguauaguugugggguagugauuuuacccuguucaggagauaacuauacaaucuauugccuucccugaggaguag -3'   obs
   aaaacauugcucuaucagagugagguaguagauuguauaguugugggguagugauuuuacccuguucaggagauaacuauacaaucuauugccuucccugaggaguag       exp
   aaaacauugcucuaucagagugagguaguagauuguauaguugugggguagugauuuuacccuguucaggagauaacuauacaaucuauugccuucccugaggaguag       known
   .......((((((.((((.(..(((((((((((((((((((((((((((((.....)))))))..........)))))))))))))))))))))..)))))))))).
```

	reads	mm	sample
..aAugagguaguagauuguauag..	1	1	seq
..aAugagguaguagauuguauagu..	1	1	seq
..aAugagguaguagauuguauaguu..	17	1	seq
..Cugagguaguagauuguaua..	1	1	seq
..Uugagguaguagauuguaua..	1	1	seq
..Uugagguaguagauuguauag..	9	1	seq
..Cugagguaguagauuguauagu..	6	1	seq
..Uugagguaguagauuguauagu..	56	1	seq
..Uugagguaguagauuguauaguu..	4	1	seq
..Cugagguaguagauuguauaguu..	221	1	seq
..Cugagguaguagauuguauaguug..	1	1	seq

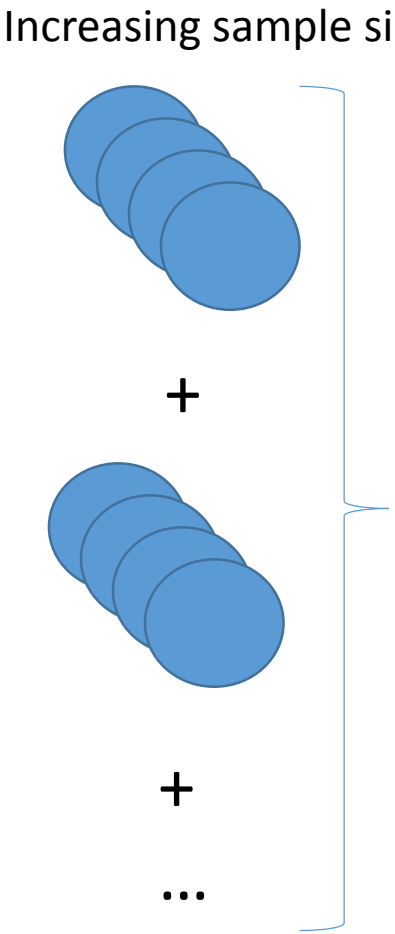
HiPipe後端的硬體架構



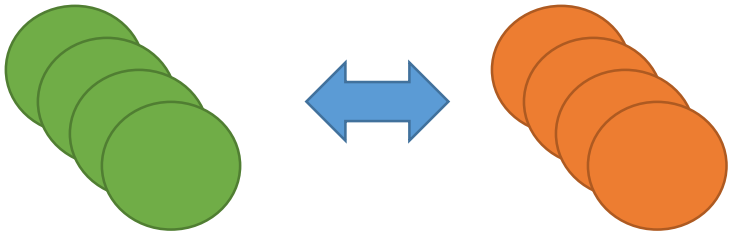


(http://www.vidalcasey.com/wp-content/uploads/2014/07/speed-bottleneck.jpg)

Multiple Samples



Joint calling



Replicates in conditions



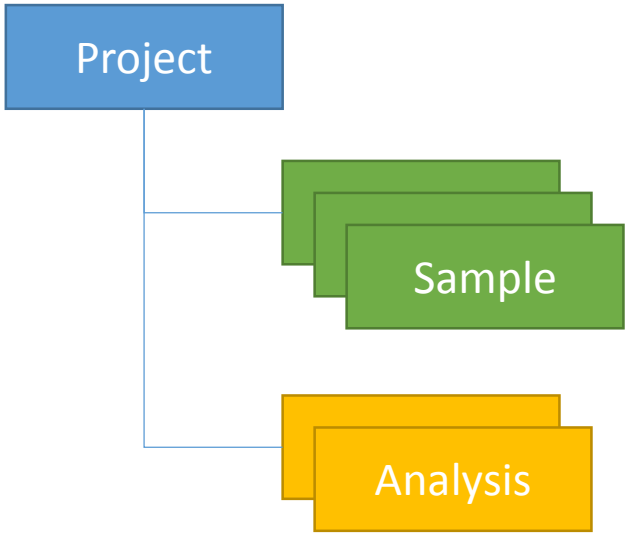
Differential Expression

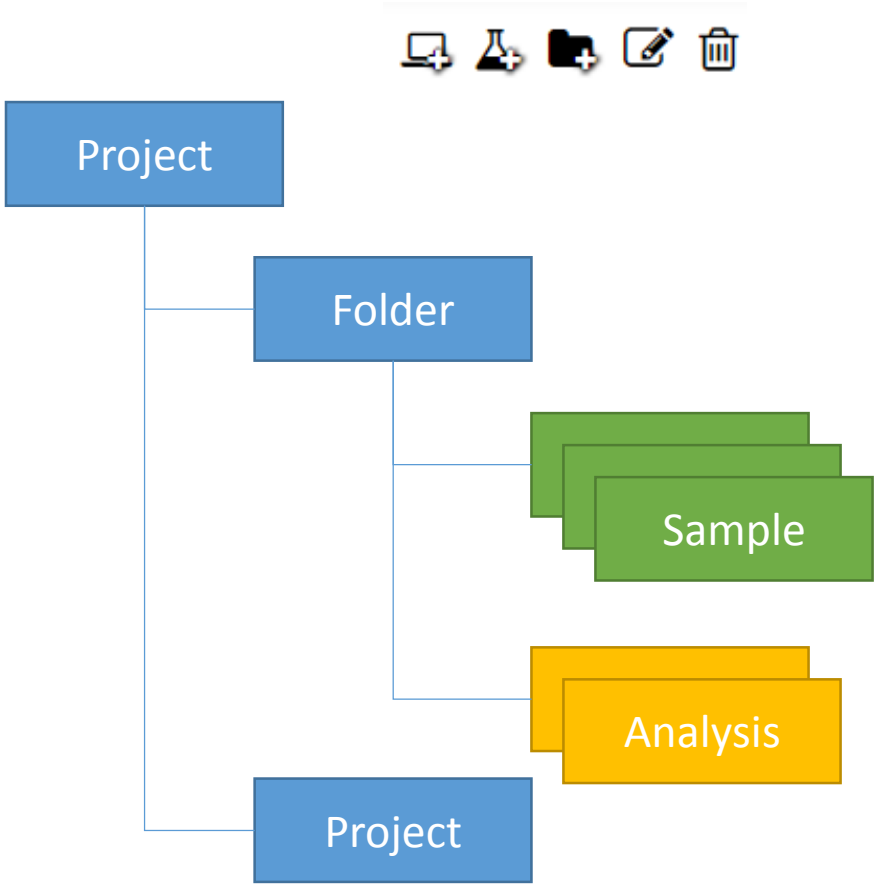
HiPipe Professional

- Multi-samples
- Management, Sharing and Security

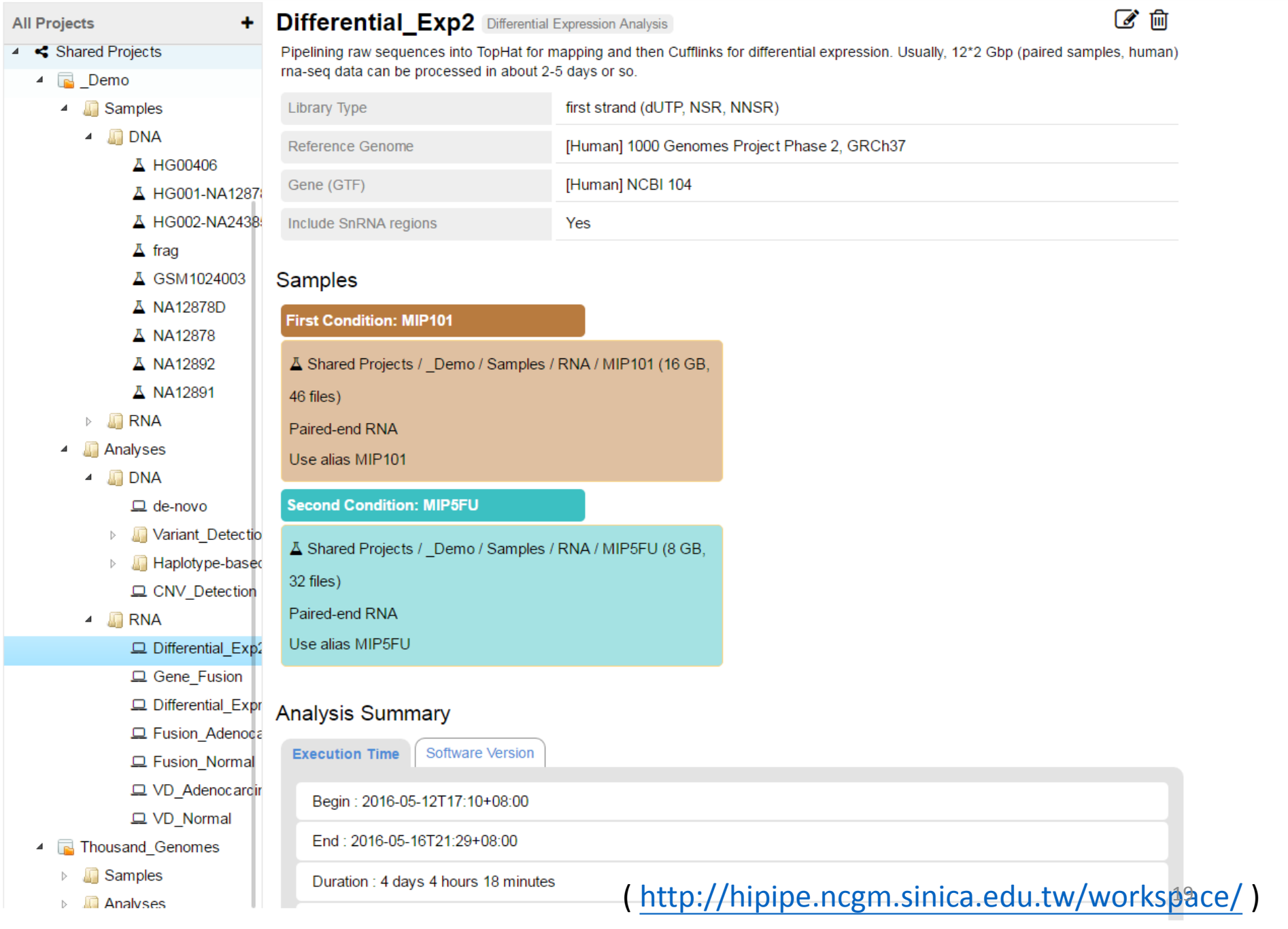


Structure


Project
Sample
Analysis


Project
Folder
Sample
Analysis
Project

HiPipe Professional Beta



(http://hipipe.ncgm.sinica.edu.tw/workspace/)

Roles & Authorities



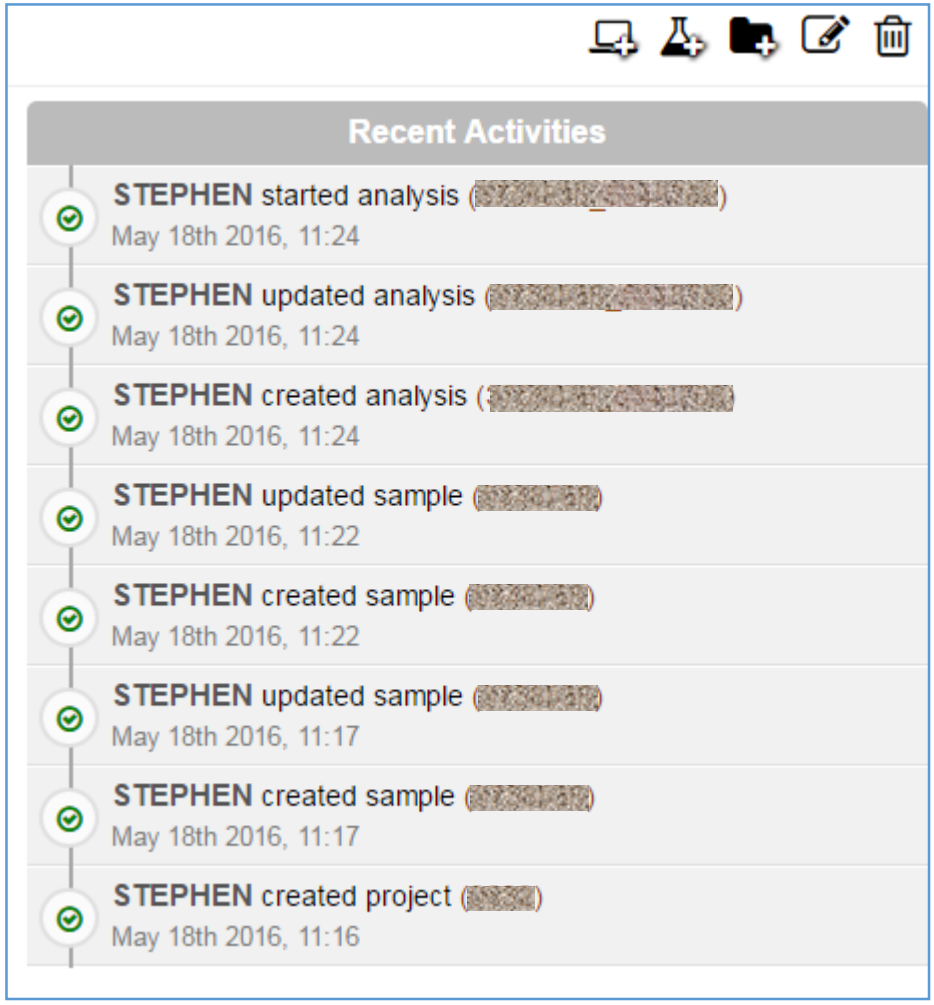
(http://www.rsc.org/eic/2015/03/mooc-massive-open-onli

- Authentication
- Authorization

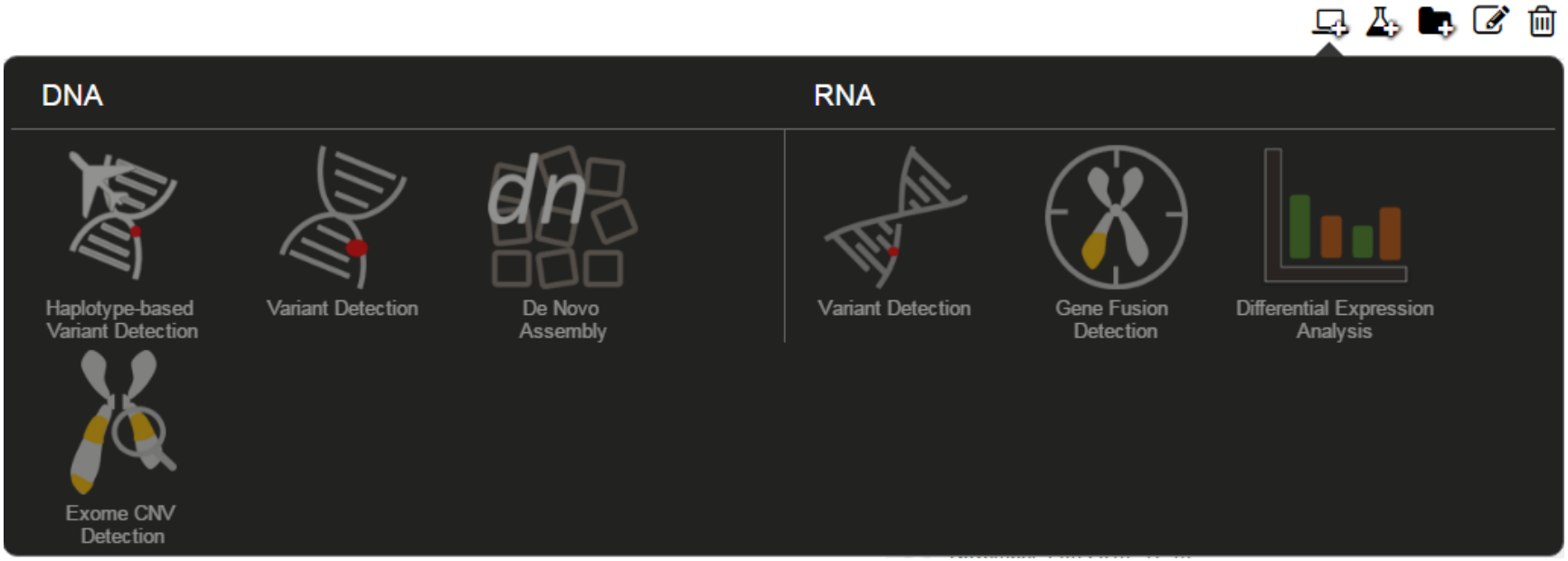
Authority	Manager	Member	Visitor
add member & viewer	o		
add sample/analysis	o	o	
delete	o	o	
edit	o	o	
view	o	o	o
download	o	o	o

Security

- Logging
  - Who
  - When
  - What



Create Analysis

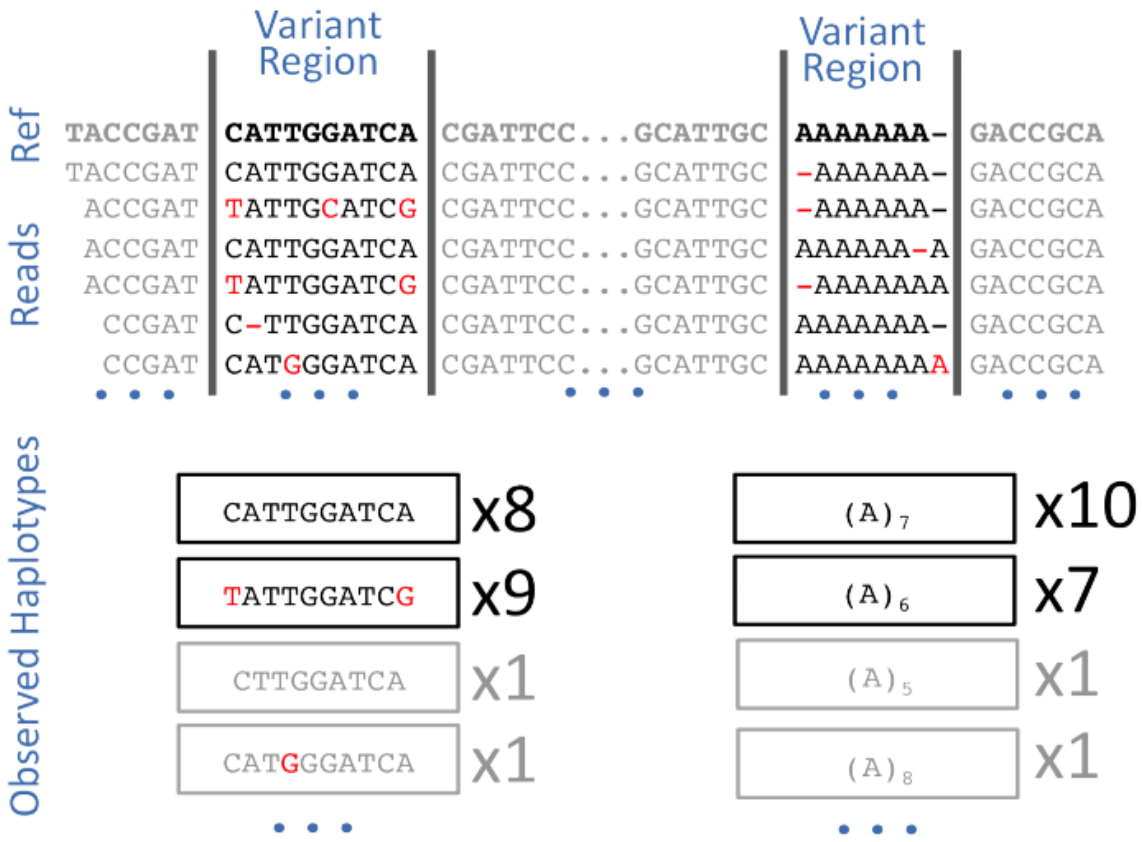


Haplotype-based variant detection from short-read sequencing

Erik Garrison and Gabor Marth

January 12, 2016

With genomic variant dete
genome. Widely-used metho
information is essential for t
directly from the primary se
use of haplotypes does not
large improvement in detecti
freebayes Bayesian variant de
model to allow for arbitrary g
numbers of alleles. To furth
of our confidence that the lo
experimental data. These adv
at the detection of SNPs, ind



Scalability

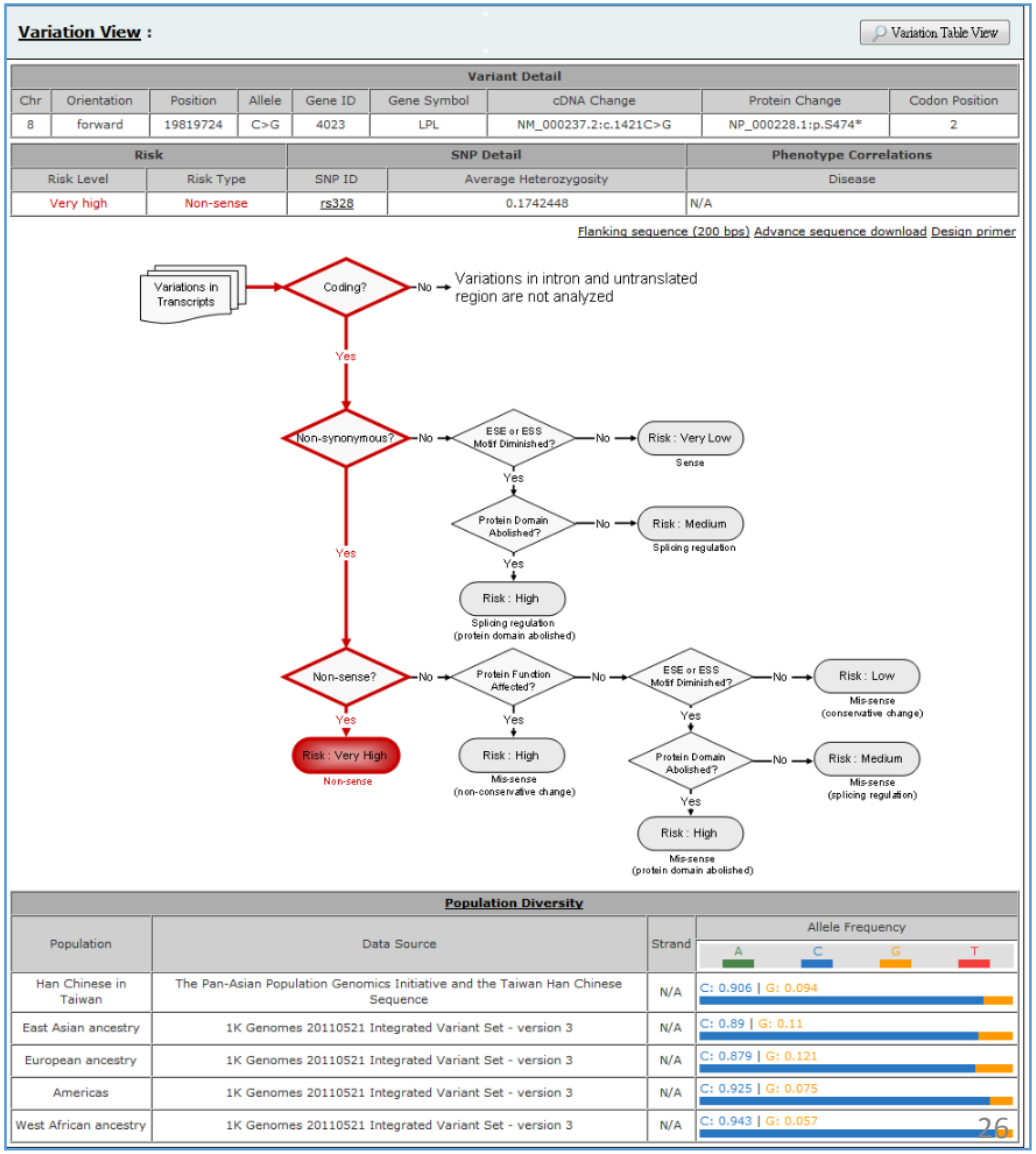
	HiPipe	HiPipe Professional				
	Basic	Cloud	Cluster	Quad	Duo	Uno
storage	cloud	cloud	local	local	local	local
nodes	10	10	10	1	1	1
CPU	640	640	640	64	16	6
variant detection 40X human whole genome	56m	56m	56m	3.5h	8.4h	24.7h
user	具大量快速分析需求的單位			需要小量快速分析的單位	研究所醫院	單一實驗室

New Features

	HiPipe	HiPipe Professional				
	HiPipe Basic	HiPipe Cloud	HiPipe Cluster	HiPipe Quad	HiPipe Duo	HiPipe Uno
單一樣本分析	O	O	O	O	O	O
上傳 > 2GB檔案	O	O	O	O	O	O
分析結果儲存方案	1個月	O	O	O	O	O
分析檔案選擇方式	每次上傳	勾選	勾選	勾選	勾選	勾選
檔案總管方式管理		O	O	O	O	O
檔案續傳功能		O	O	O	O	O
多樣本合併分析		O	O	O	O	O
帳號權限管理		O	O	O	O	O
在客戶端分析儲存			O	O	O	O

Annotation

- Risk level
- Risk Type
- Population Allele Frequencies
- …
- ClinVar
- Cosmic



Variation View : Variation Table View

Variant Detail

Chr	Orientation	Position	Allele	Gene ID	Gene Symbol	cDNA Change	Protein Change	Codon Position
8	forward	19819724	C>G	4023	LPL	NM_000237.2:c.1421C>G	NP_000228.1:p.S474*	2

Risk		SNP Detail		Phenotype Correlations
Risk Level	Risk Type	SNP ID	Average Heterozygosity	Disease
Very high	Non-sense	rs328	0.1742448	N/A

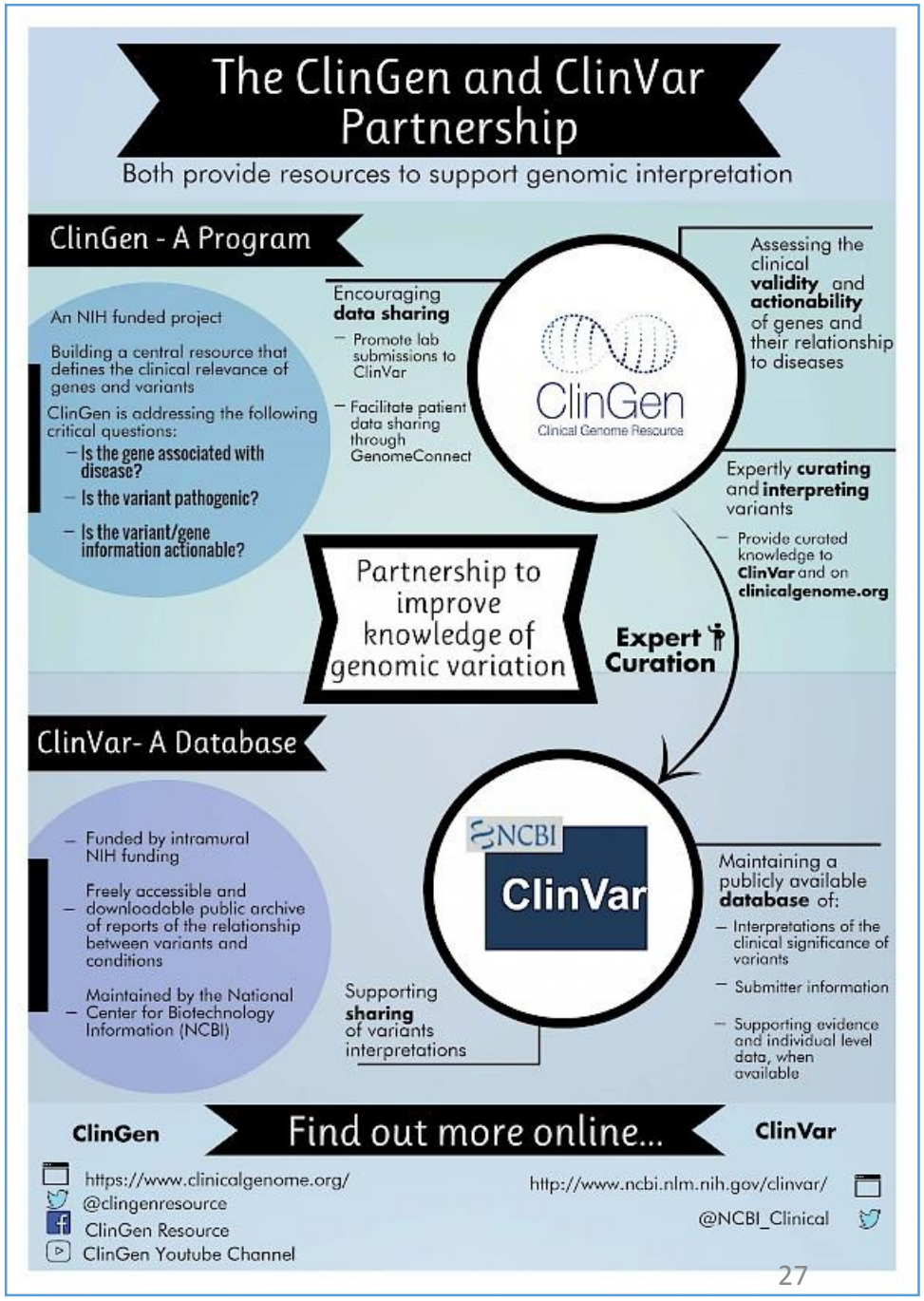
Flanking sequence (200 bps) Advance sequence download Design primer

Population Diversity

Population	Data Source	Strand	Allele Frequency (A C G T)
Han Chinese in Taiwan	The Pan-Asian Population Genomics Initiative and the Taiwan Han Chinese Sequence	N/A	C: 0.906 | G: 0.094
East Asian ancestry	1K Genomes 20110521 Integrated Variant Set - version 3	N/A	C: 0.89 | G: 0.11
European ancestry	1K Genomes 20110521 Integrated Variant Set - version 3	N/A	C: 0.879 | G: 0.121
Americas	1K Genomes 20110521 Integrated Variant Set - version 3	N/A	C: 0.925 | G: 0.075
West African ancestry	1K Genomes 20110521 Integrated Variant Set - version 3	N/A	C: 0.943 | G: 0.057

ClinVar

- an archival database that aggregates information about genomic variation and its relationship to human health.



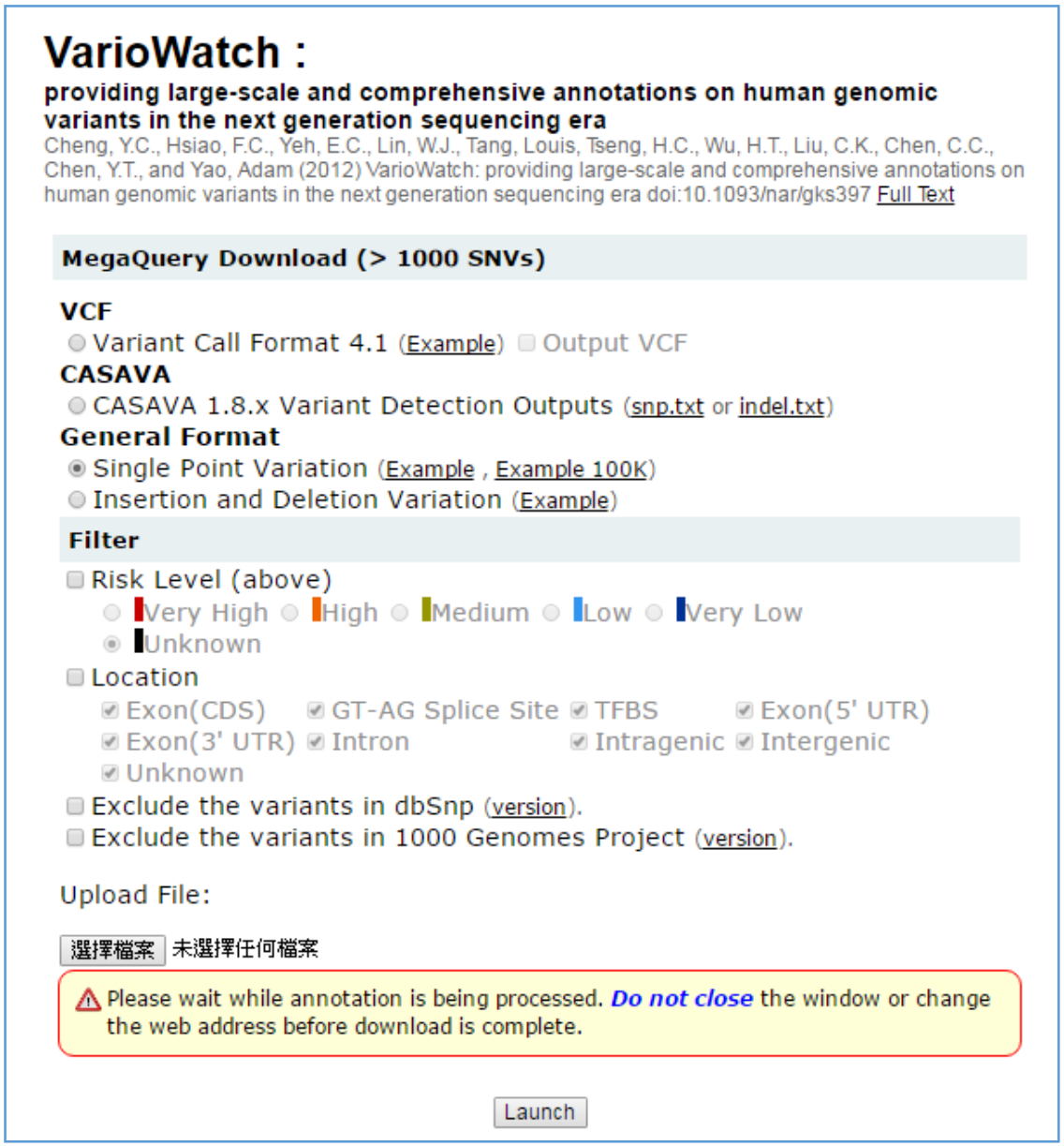
ClinVar doesn't currently support batch query

How can I retrieve batches of data from ClinVar?

ClinVar does not currently support a batch query interface, but there are several approaches that might still meet your needs:

Approaches to process batches of data

Use case	Possible solutions
Variant-specific data for a list of genes	Query ClinVar by listing the genes using the boolean OR, and download the results interactively or usin http://www.ncbi.nlm.nih.gov/clinvar?term=spred1[gene]%20OR%20shoc2[gene]%20OR%20raf1[gene]
	Download the file ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/tab_delimited/variant_summary.txt.gz and proces
	Download the full data extract ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/xml/ClinVarFullRelease_00-latest.xn
Variant-specific data for a list of conditions	Use the same approach as for genes, but use instead a list of identifiers such as MIM numbers or Cond identifiers for phenoytpes, but not their names.
Gene-disease relationships	ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/gene_condition_source_id reports gene-disease relationships used GeneReviews, and some from NCBI staff curation. This report does include all genes with variants ass

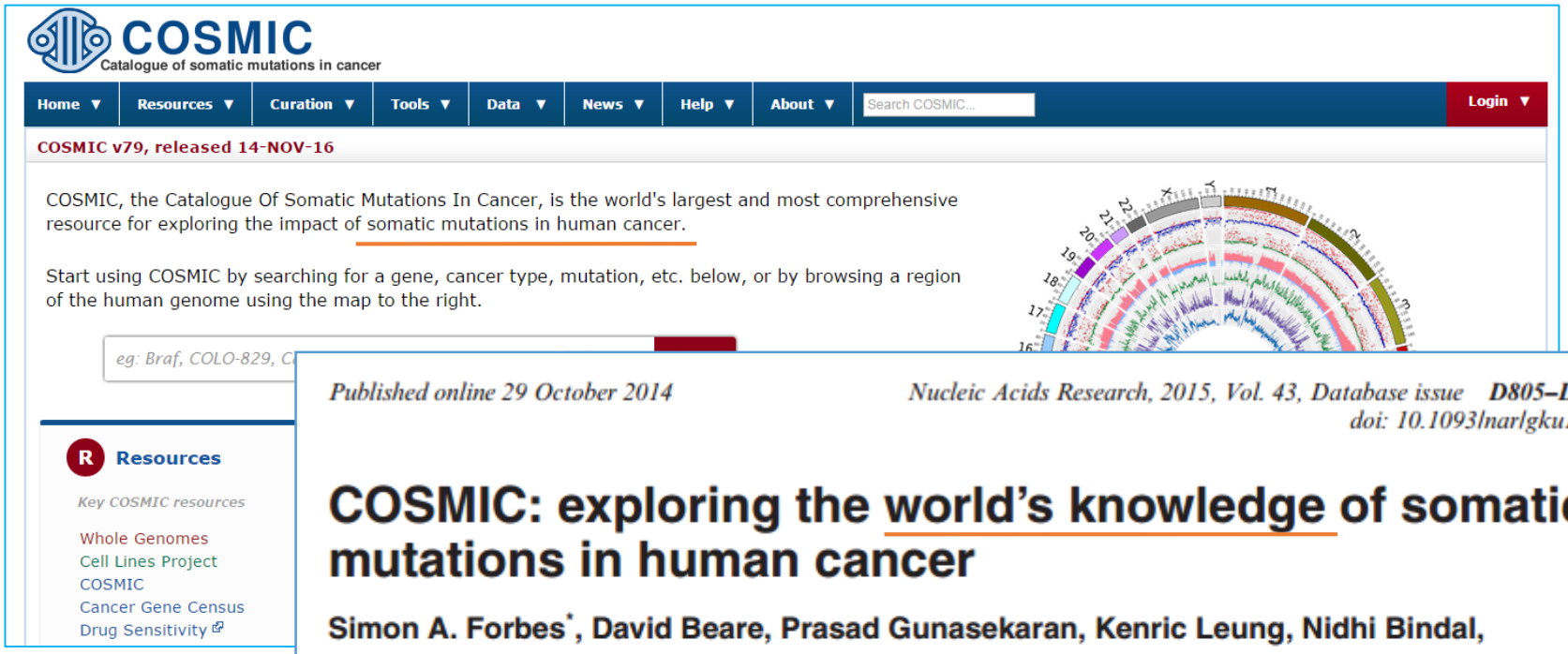


(http://genepipe.ncgm.sinica.edu.tw/variowatch/main.do?mega=true)

#CHROM	POS	ID	REF	ALT	clinical significance	condition	URL
8	90990479	rs1805794	C	G	Benign	Hereditary cancer-predisposing syndrome	http://www.ncbi.nlm.nih.gov/clinvar/variation/134876
17	7579472	rs1042522	G	C	drug response	antineoplastic agents response - Efficacy, Toxicity/ADR	http://www.ncbi.nlm.nih.gov/clinvar/variation/12351
13	32913655	rs80359490	CAG	C	Pathogenic	Breast-ovarian cancer, familial 2	http://www.ncbi.nlm.nih.gov/clinvar/variation/51791
4	46329723	rs279845	T	A	risk factor	Alcoholism, susceptibility to	http://www.ncbi.nlm.nih.gov/clinvar/variation/16219

- Benign 良性
- Likely benign 可能良性
- Uncertain significance 不確定的意義
- Likely pathogenic 可能致病
- Pathogenic 致病
- Drug response 藥物反應
  - affects a drug response, not a disease.
- Confers sensitivity 賦予敏感性
- Risk factor 風險因素
- Association 有關
- Protective 保護的
  - For variants that decrease the risk of a disorder, including infections.
- Has conflicts 有衝突
  - Only for submissions from a consortium, where groups within the consortium have **conflicting intepretations** of a variant but provide a single submission to ClinVar.

Cosmic



(http://cancer.sanger

Published online 29 October 2014

Nucleic Acids Research, 2015, Vol. 43, Database issue ***D805–D811***
doi: 10.1093/nar/gku1075

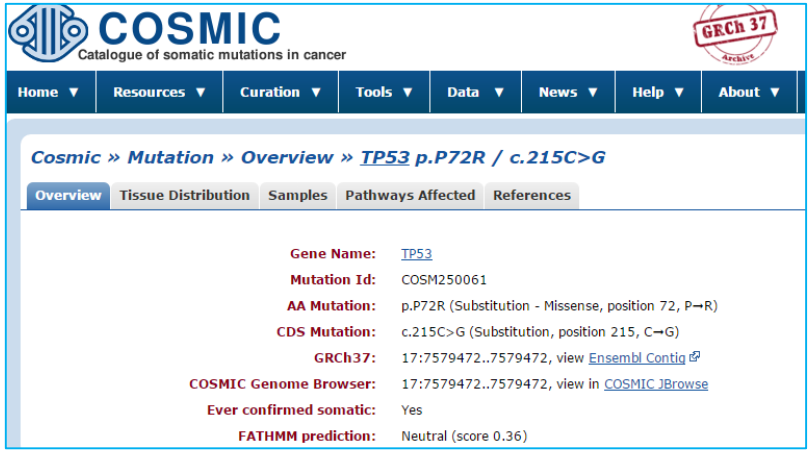
COSMIC: exploring the world's knowledge of somatic mutations in human cancer

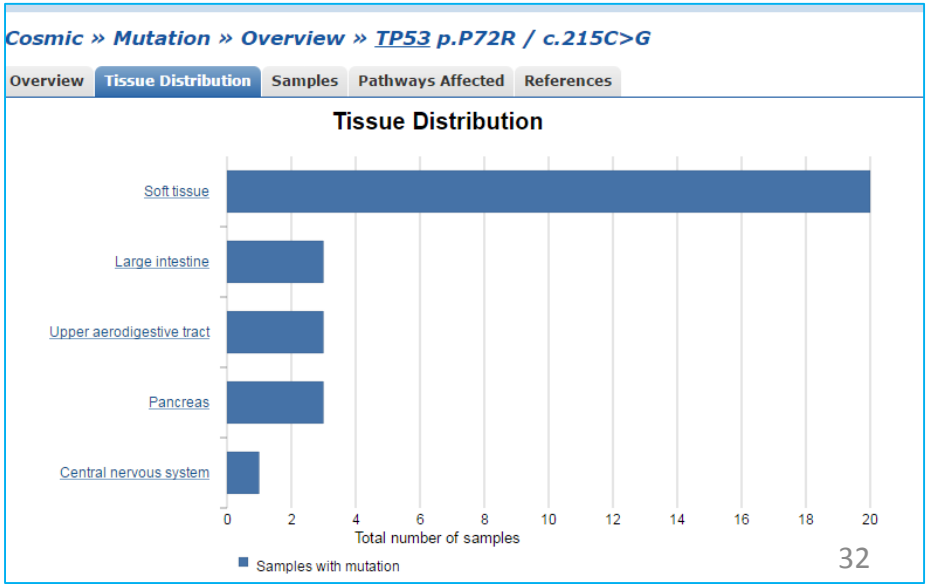
Simon A. Forbes*, David Beare, Prasad Gunasekaran, Kenric Leung, Nidhi Bindal, Harry Boutselakis, Minjie Ding, Sally Bamford, Charlotte Cole, Sari Ward, Chai Yin Kok, Mingming Jia, Tisham De, Jon W. Teague, Michael R. Stratton, Ultan McDermott and Peter J. Campbell

Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK, CB10 1SA.

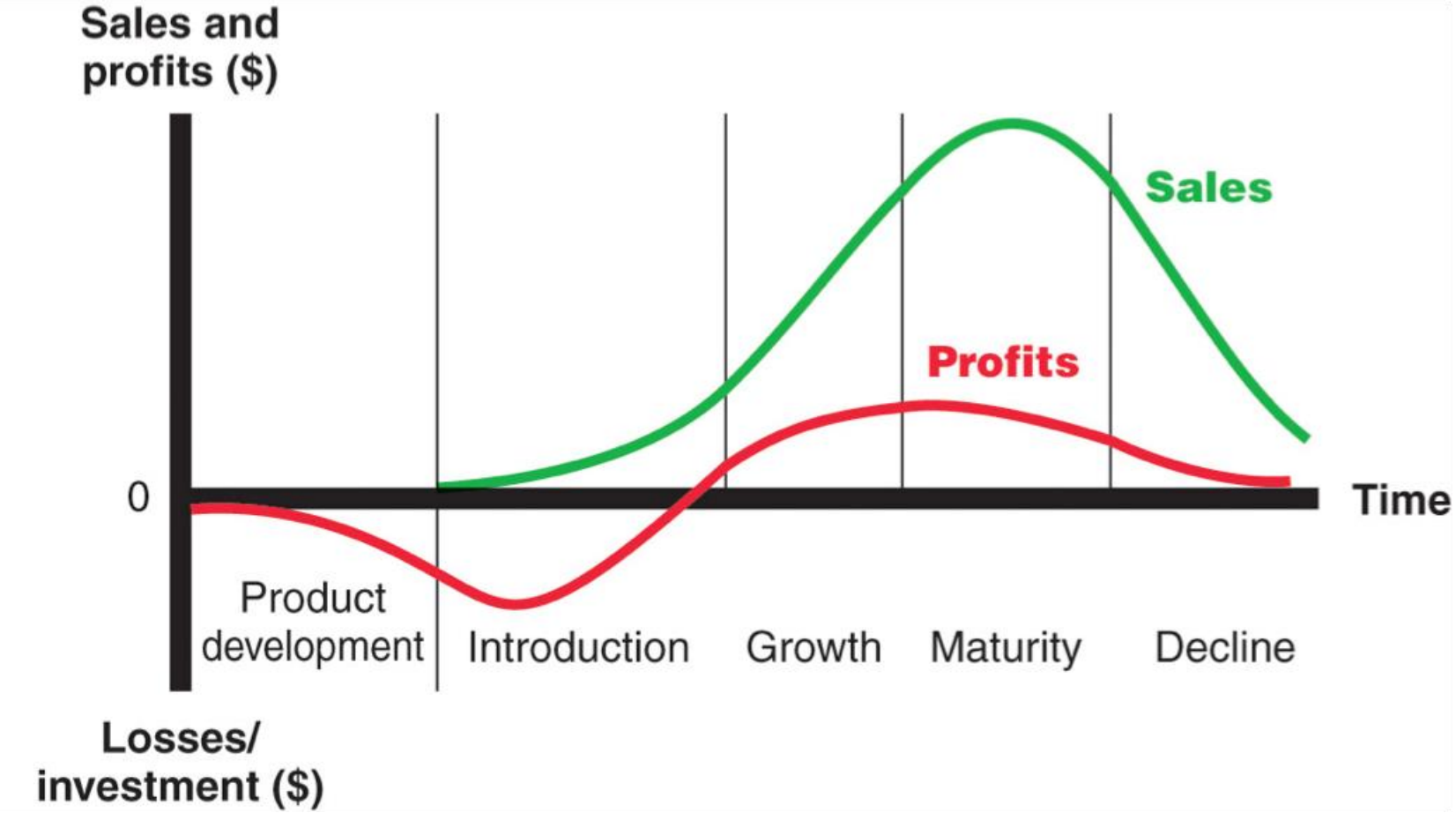
Received October 03, 2014; Revised October 14, 2014; Accepted October 16, 2014

#CHROM	POS	ID	REF	ALT	somatic status	primary tissue	histology	URL
8	90990479	rs1805794	C	G	Confirmed Somatic	central_nervous_system	glioma	http://grch37-cancer.sanger.ac.uk/cosmic/mutation/overview?id=250061
					Confirmed Somatic	large_intestine	carcinoma	
					Confirmed Somatic	liver	carcinoma	
					Confirmed Somatic	soft_tissue	haemangioblastoma	
					Previously Reported	upper_aerodigestive_tract	carcinoma	





Customized Analysis



(https://atitus707.wordpress.com/product-life-cycle-stages-and-strategies/)

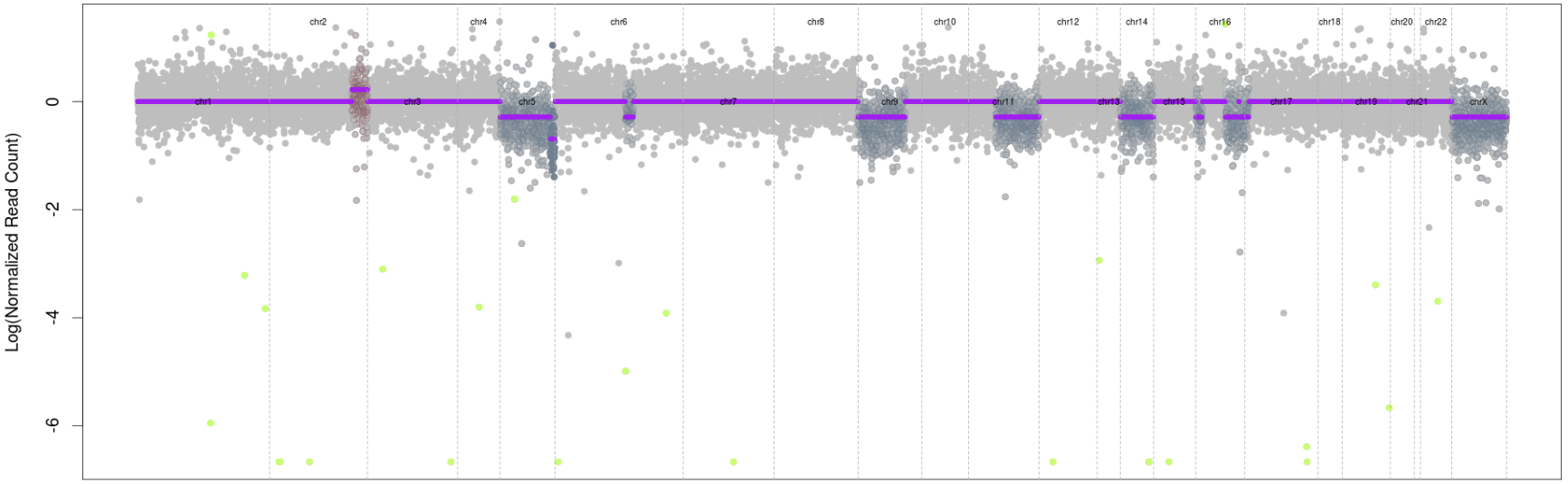
PCR-based DNA Amplification

- Pros
  - deep sequencing
  - identification of low allelic-fraction variants.
  - low cost
- Variant detection
  - SNP
  - indel

CNV Detection for Amplicons

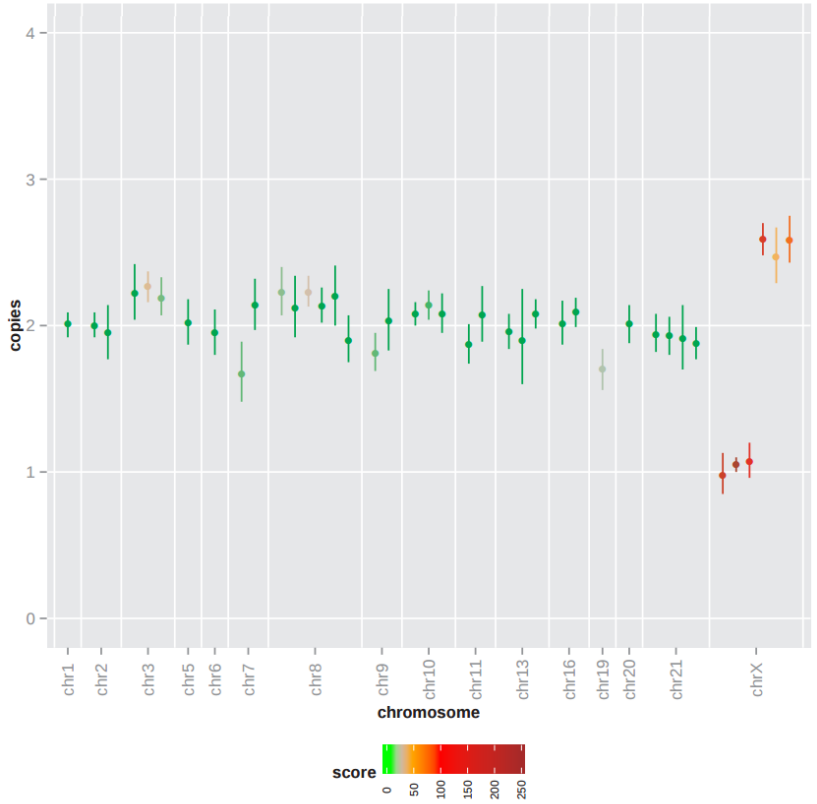
- Cons
  - PCR的過程會扭曲 tumor 最初的 copy number 的資訊
  - very small fraction of the genome
- ONCOCNV
- quandico

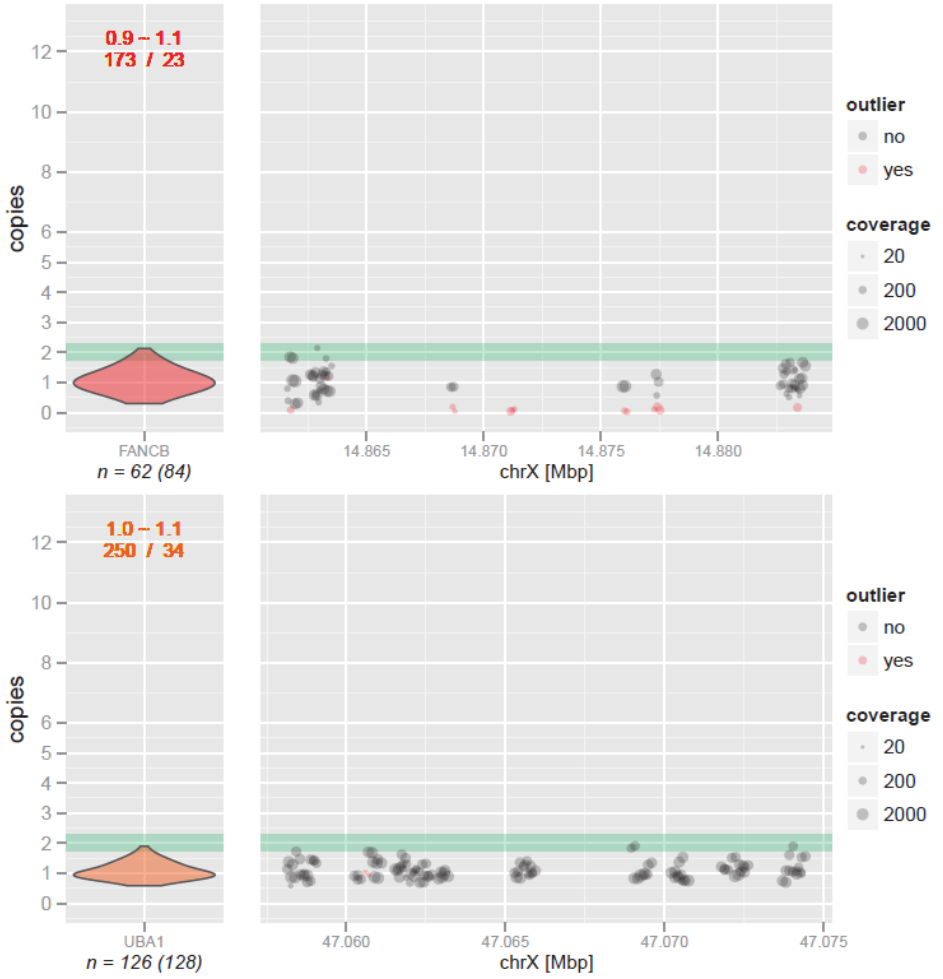
ONCOCNV



- GREEN one-point-**outlier**
- DARK GREY SURROUNDINGS frequent one-point-outlier
- BROWN >1 level **gain**
- BROWN SURROUNDINGS 1-level gain
- BLUE >1 level **loss**
- BLUE SURROUNDINGS 1-level loss

quandico

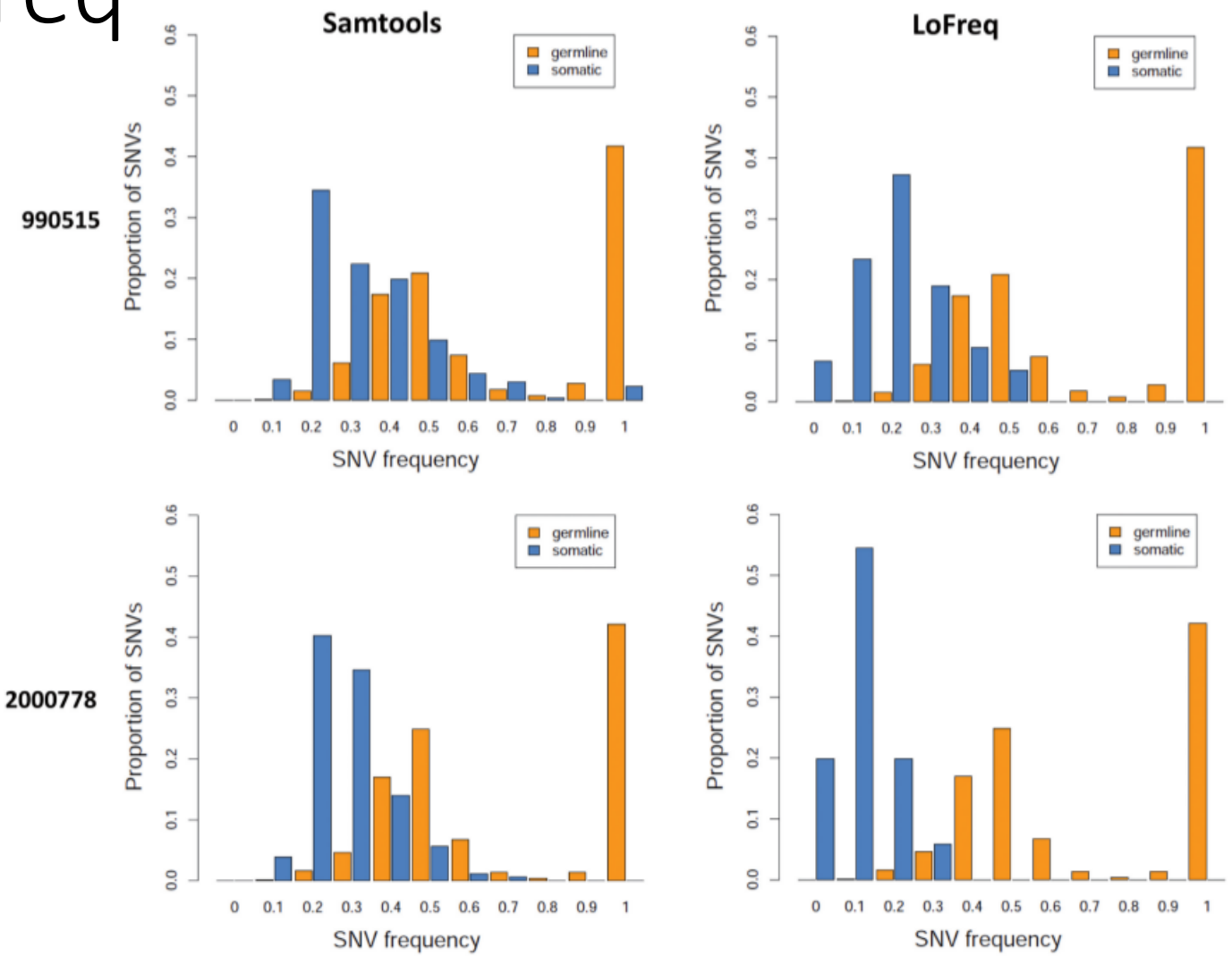

copies
chromosome
chr1
chr2
chr3
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr13
chr16
chr19
chr20
chr21
chrX
score
0
50
100
150
200
250


0.9 – 1.1
173 / 23
copies
FANCB
n = 62 (84)
14.865
14.870
14.875
14.880
chrX [Mbp]
outlier
no
yes
coverage
20
200
2000
1.0 – 1.1
250 / 34
UBA1
n = 126 (128)
47.060
47.065
47.070
47.075

Low Frequency Variant Detection

- population perspective
  - viral quasi species
  - intra-host variation
  - bacterial sub-populations
  - sub-clonal evolution in cancer
    - Several recent cancer sequencing studies have further emphasized the functional role of rare sub-populations and variants in aspects such as tumor growth, drug resistance and metastasis and the need for computational tools to study them.
- the majority of variant-calling methods
  - diploid calls with discrete frequencies of interest
    - 0, 0.5 and 1

LoFreq


Samtools
LoFreq
990515
2000778
germline
somatic
Proportion of SNVs
SNV frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Other Customized Analysis

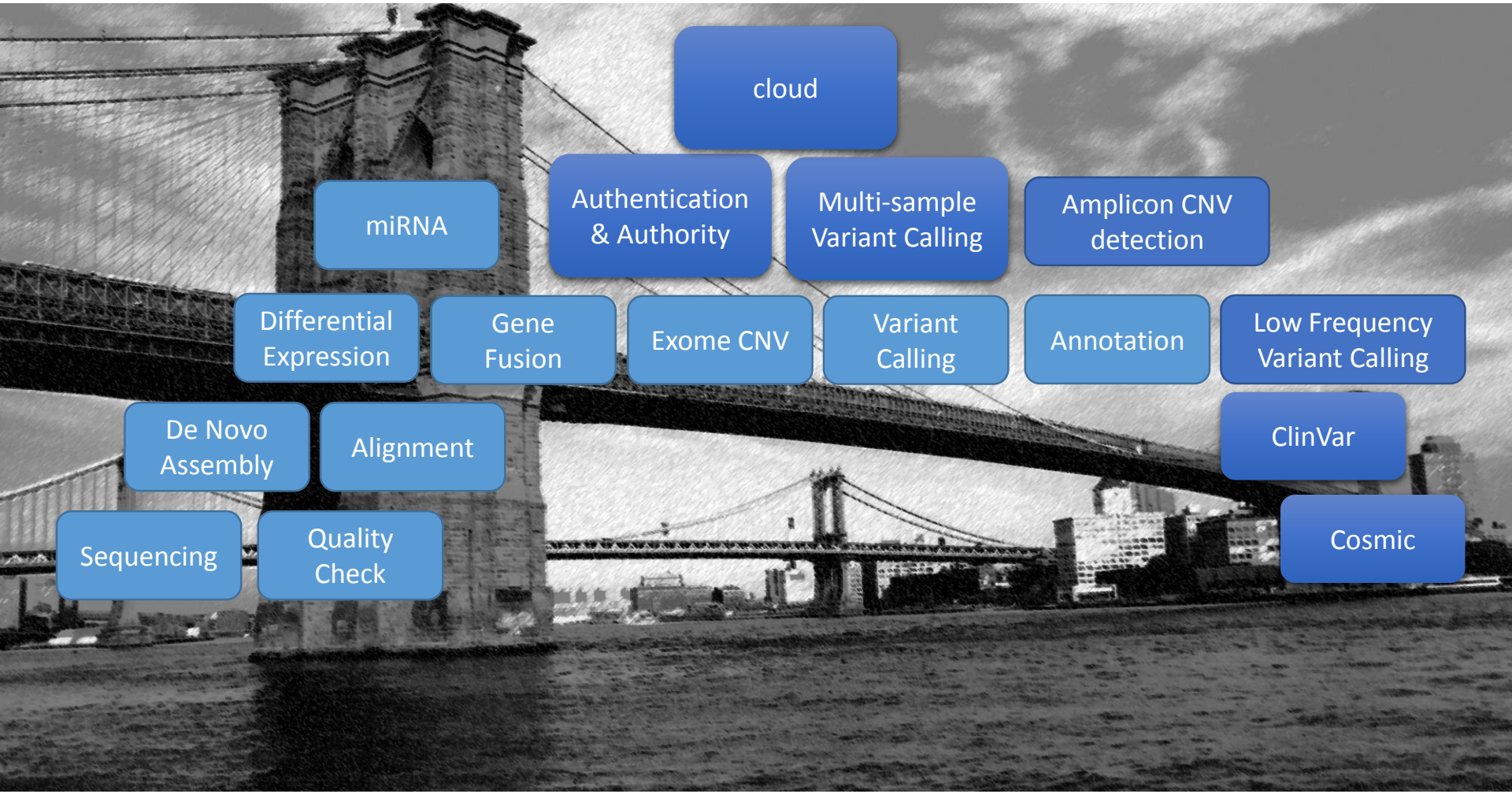
- ChIP-Seq
  - Peak-calling
  - annotation
- Aptamer
  - Barcode
  - aptamer
  - clustering
- …



Khalid K. Alam, Jonathan L. Chang & Donald H. Burke "FASTAptamer: A Bioinformatic Toolkit for High-Throughput Sequence Analysis of Combinatorial Selections." Molecular Therapy — Nucleic Acids. 2015. 4, e230; DOI: 10.1038/mtna.2015.4



- Finis Coronat Opus
- The finish crowns the work


cloud
miRNA
Authentication & Authority
Multi-sample Variant Calling
Amplicon CNV detection
Differential Expression
Gene Fusion
Exome CNV
Variant Calling
Annotation
Low Frequency Variant Calling
De Novo Assembly
Alignment
ClinVar
Sequencing
Quality Check
Cosmic

NGS分析服務收費標準

Code	Service	Service Unit	Academia (NT$)	Industry (NT$)
C2-22	生物資訊分析、軟體開發與諮詢	Per Hour	444	665
C2-25	生物資訊分析、軟體開發與諮詢之系統使用費率	Per Hour	0	1,929
C2-38	NGS全基因體分析	Per Sample	444	8,771
C2-39	NGS Exome分析	Per Sample	444	4,718
C2-40	NGS RNA分析	Per Sample	444	8,771
C2-41	LOH偵測	Per Sample	444	4,718
C2-49.1	NGS運算分析服務	Two Core Hours	2	4
C2-49.2	NGS資料儲存服務	Ten Gigabytes per Month	4	13

(http://ncgm.sinica.edu.tw/bioit_service_03.html)

Team Members

- 軟體工具的熟悉與使用 (行中、亭均)
- 分析工作之自動化與排程 (Louis、德華)
- 叢集與分散式系統之控制 (方智)
- 效能調教
  - 軟體平行最佳化 (耀德、仁屏)
  - 硬體 (與 MIS組同仁協同開發)
    - 網路頻寬
    - 磁碟讀寫速度
- 系統分析與使用者介面
  (爾瞻、萬嘉、中饋、蕙棻)





Thank you for Your Attention!

(Painters working on the bridge, October 1914 (Eugene de Salignac)

https://en.wikipedia.org/wiki/Brooklyn_Bridge#/media/File:Painters_suspended_on_cables_of_the_Brooklyn_Bridge,_on_07_October_1914.png)