



trackViewer

LOLLIPOP/DANDELION PLOTS FOR
METHYLATION STATUS AND MUTATION DATA

Jianhong Ou

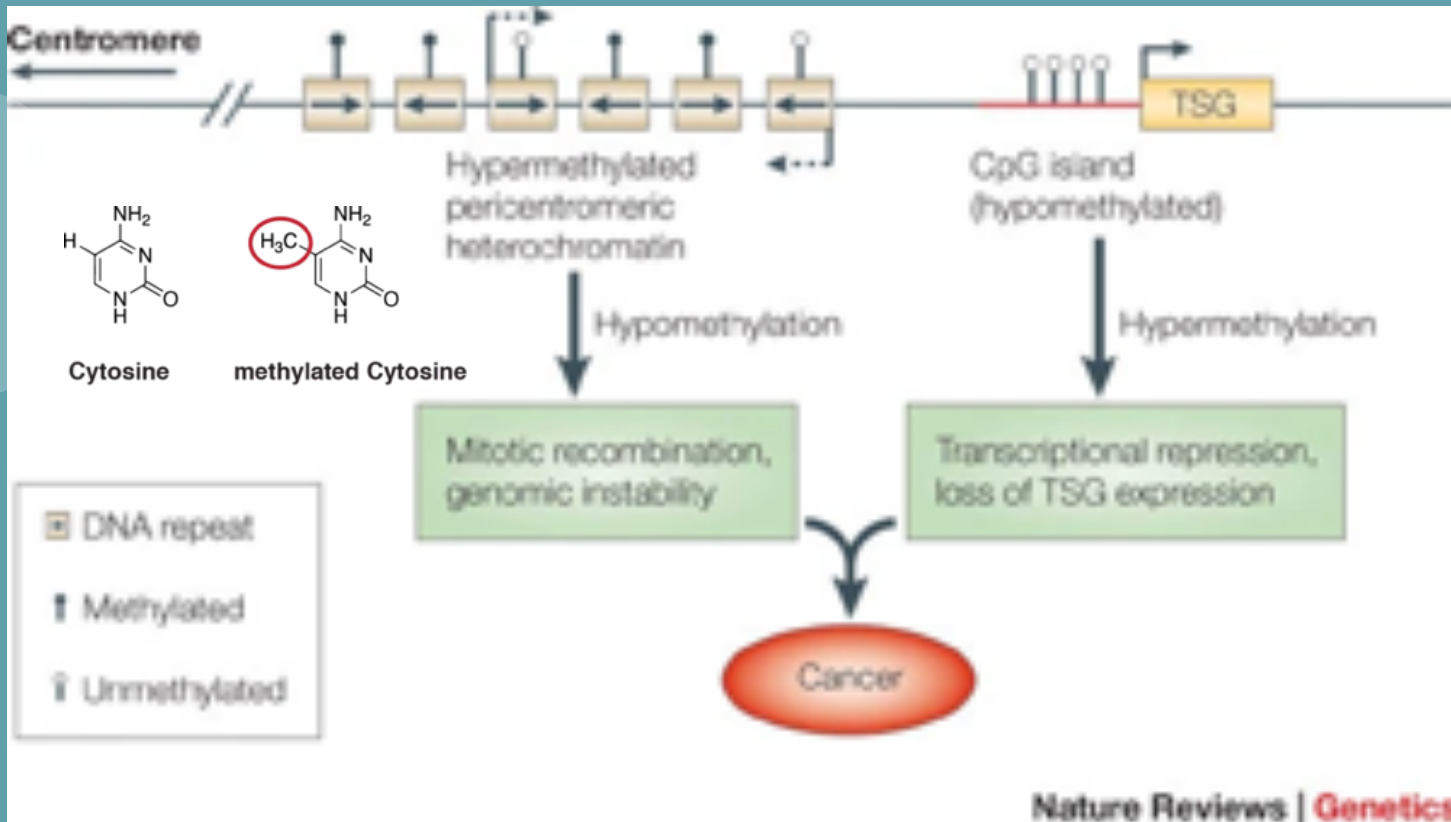
INSTALL THE WORKSHOP PKG

```
## set the working directory,  
## replace "~/Downloads/workshop2020" by your path  
wd <- "~/Downloads/workshop2020"  
dir.create(wd)  
setwd(wd)  
library(BiocManager)  
install("jianhong/trackViewerBiocAsia2020Workshop",  
        build_vignettes = TRUE)  
vignette("trackViewer", package="trackViewerBiocAsia2020Workshop")
```

<https://github.com/jianhong/trackViewerBiocAsia2020Workshop>
<https://jianhong.github.io/trackViewerBiocAsia2020Workshop/articles/trackViewer.html>
<https://bioconductor.org/packages/trackViewer>
<https://www.nature.com/articles/s41592-019-0430-y>
Slides:
<https://github.com/jianhong/trackViewerBiocAsia2020Workshop/blob/master/inst/extdata/slides.pdf>



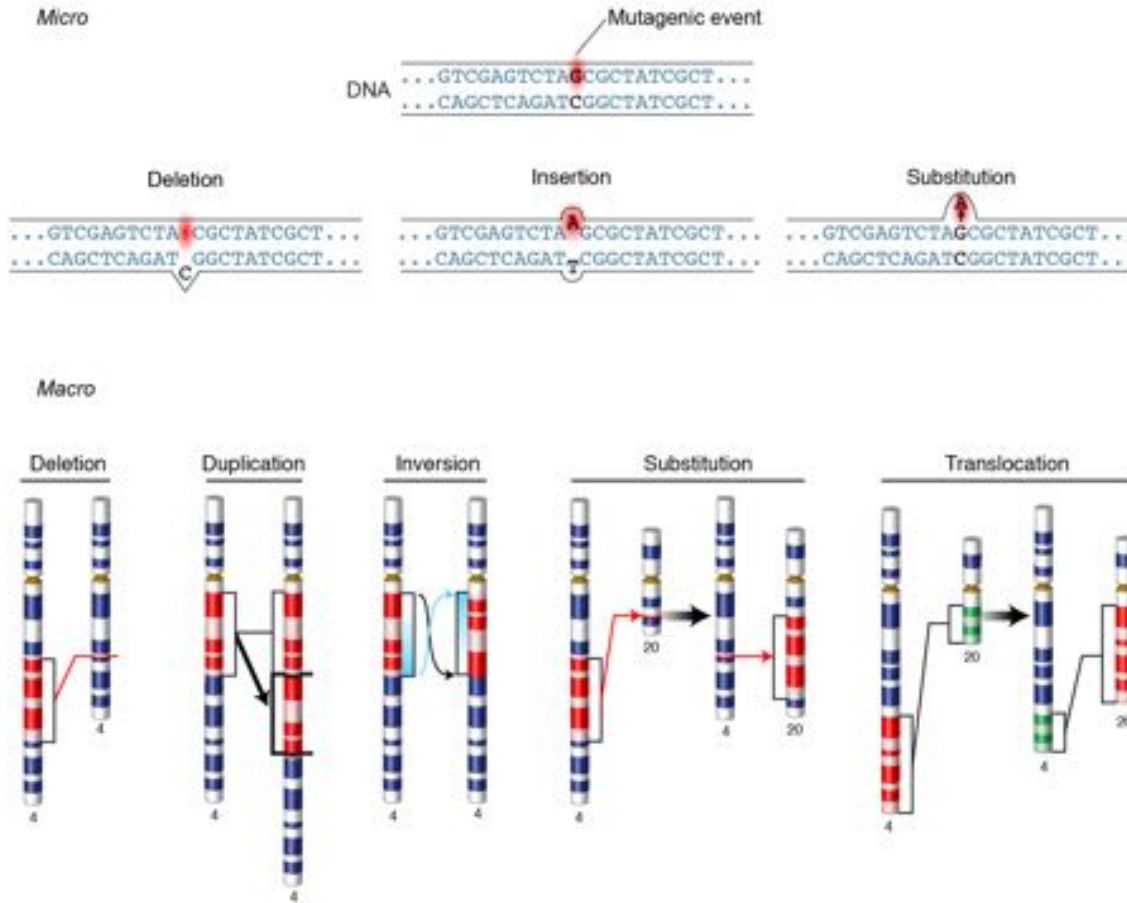
DNA METHYLATION



Robertson, K. 2005. doi: 10.1038/nrg1655
<http://www.roadmapepigenomics.org/>







<https://www.genome.gov/genetics-glossary/Mutation>

MUTATION

7/13/2020 COSMIC Genome Browser 17:7668266..7668623

COSMIC
Catalogue of somatic mutations in cancer

Genome Browser

Available Tracks

filter by text

☐ GRCh38 (reference Sequence)

Genes 5

☒ Cosmic Genes
☐ Ensembl Transcripts
☐ Non-Coding RNA
☐ Regulatory Features
☐ miRNA

Variants 8

☐ Coding Mutations
☐ Copy Number
☐ Gene Expression
☒ Methylation
☐ Non-Coding Variants
☒ SNPs (noise reduction filtering)
☐ Structural Variants
☒ dbSNP b150

0 10,000,000 20,000,000 30,000,000 40,000,000 50,000,000

7,668,375 7,668,623

Cosmic Genes

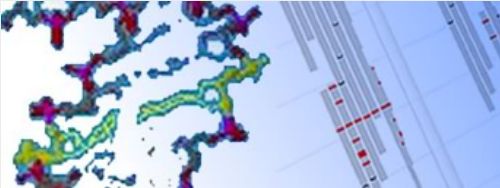
TP53_ENST00000617185
 TP53
 TP53_ENST00000620739
 TP53_ENST00000610292
 TP53_ENST00000504937
 TP53_ENST00000504290
 TP53_ENST00000610623
 TP53_ENST00000619186
 TP53_ENST00000510385
 TP53_ENST00000618944
 TP53_ENST00000610538
 TP53_ENST00000622645
 TP53_ENST00000420246
 TP53_ENST00000455263
 TP53_ENST00000445888
 TP53_ENST00000619485

rs147976274 rs17884947
 rs568709906
 rs1794294

rs17881782 rs78378222
 rs172577432

<https://cancer.sanger.ac.uk/cosmic>

https://cancer.sanger.ac.uk/browse/index.html?tracks=cosmic_gene%2CdbSNP%2Cmeth%2Ccopy%2Ccosmic&loc=17%3A7668266..7668623&data=23j&on...

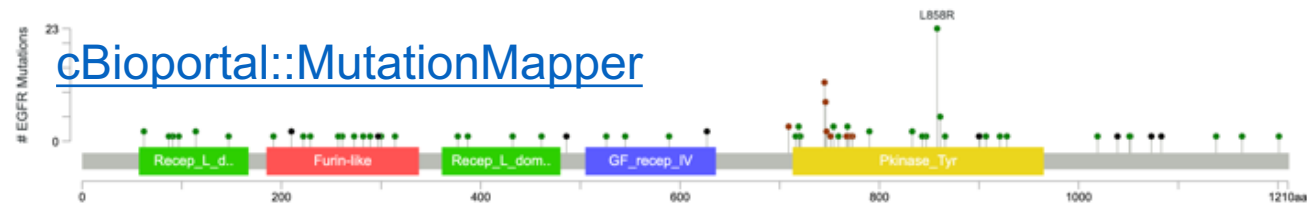


dbSNP

dbSNP contains human single nucleotide variations, microsatellites, and small-scale insertions and deletions along with publication, population frequency, molecular consequence, and genomic and RefSeq mapping information for both common variations and clinical mutations.

GENOME BROWSERS AND VIEWERS

LOLLIPOP/NEEDLE PLOTS



[pfam::lollipops](#)



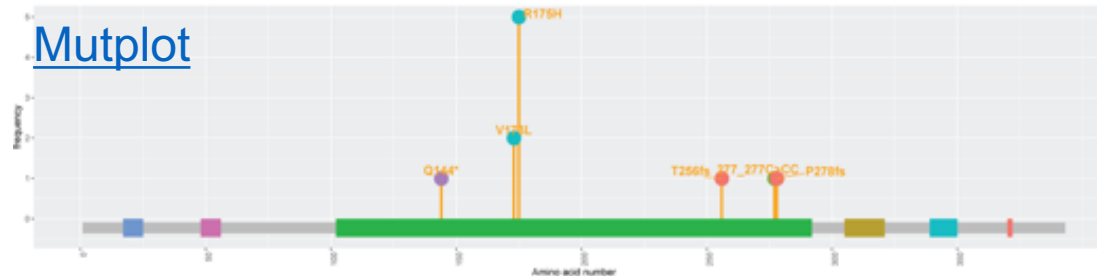
[lollipops](#)



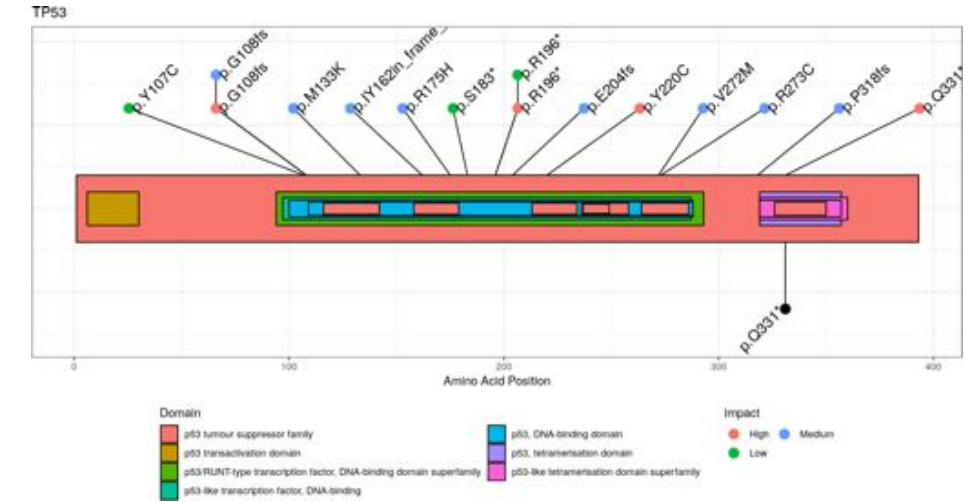
[Plot Protein](#)



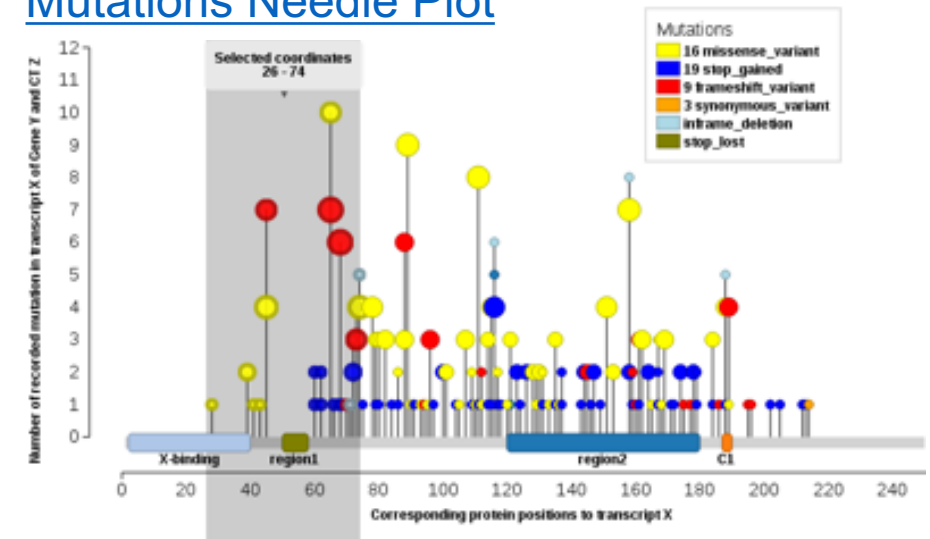
[Mutplot](#)



[GenVisR](#)



[Mutations Needle Plot](#)



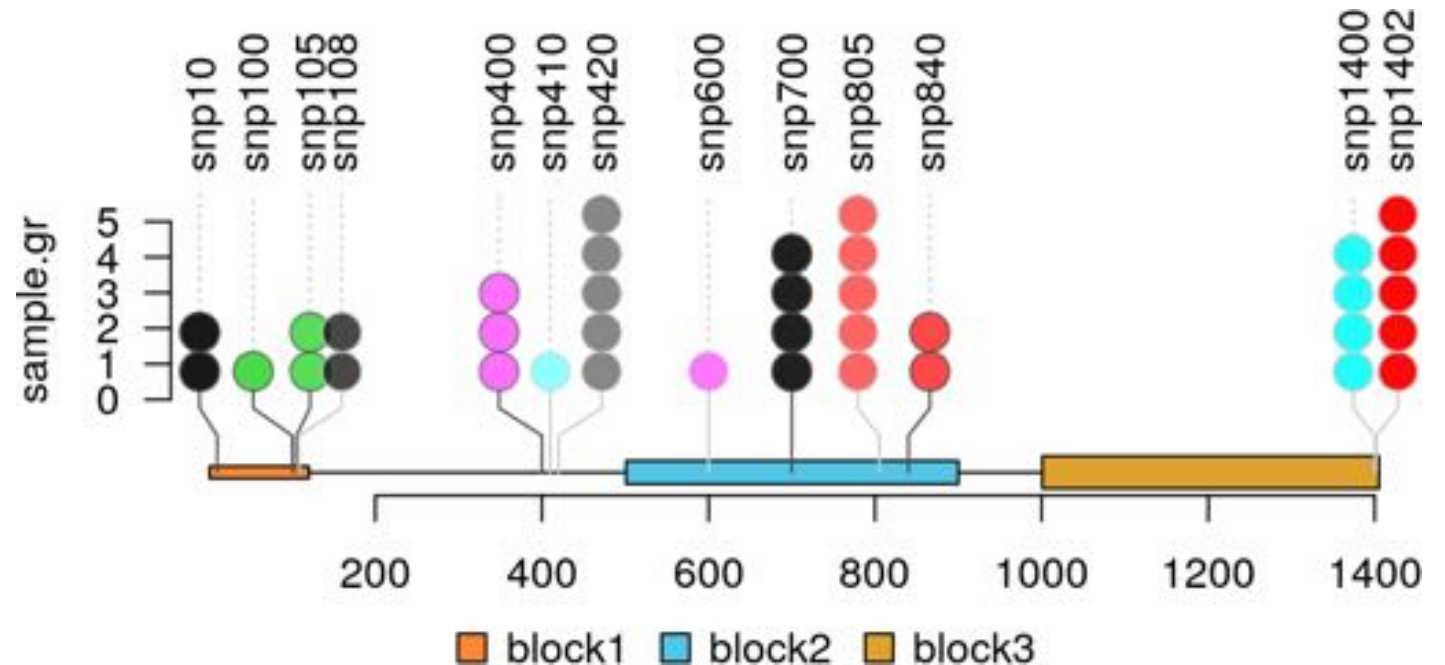
WHY trackViewer

pros:

- minimized: simple, clean design for the publication
- highly customizable: labels, symbols, colors, opacity and size
- powerful: handle high dense data

cons:

- learning curve
- no embedded database



SNPs/INDELS

SNP: Single Nucleotide Polymorphism

INDEL: INsertion/DEletion

Name	Input formats	Output formats	Called variants
Atlas-SNP2	BAM	VCF	SNPs
GATK (UnifiedGenotyper)	BAM	VCF	SNPs, InDels
SAMtools (samtools mpileup)	BAM	VCF	SNPs, InDels
SNVer	BAM	VCF	SNPs, InDels
SOAPsnp	SOAP	Text format, GLFv2	SNPs
Varscan2	Pileup/mpileup	Text format, VCF	SNPs, InDels, CNA

VCF Variant Call Format

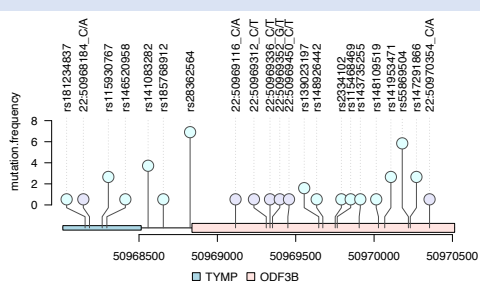
```
##fileformat=VCFv4.3
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA000001 NA000002 NA000003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

VCF/BED/BedGraph

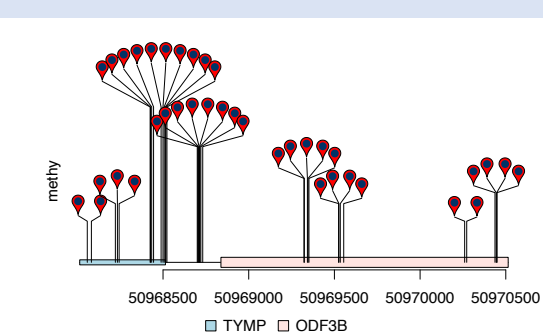
TxDb annotation package

VISUALIZATION OF METHYLATION AND MUTATION DATA IN DIFFERENT STYLES USING trackViewer

Low dense lollipop plot



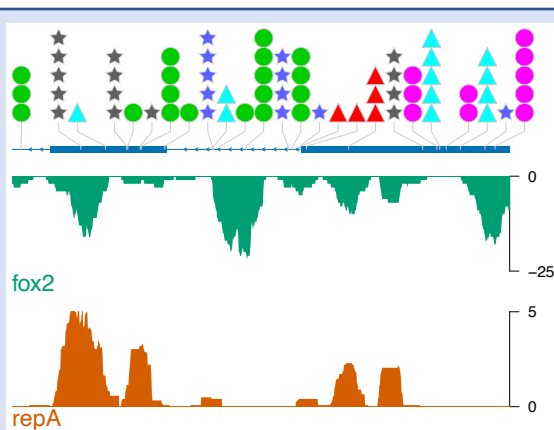
High dense dandelion plot



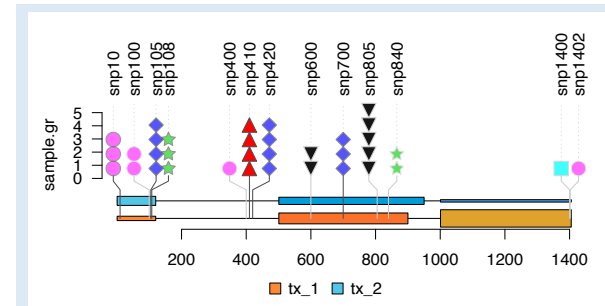
Coverage plot

WIG/BigWig/
BED/BedGraph/Bam

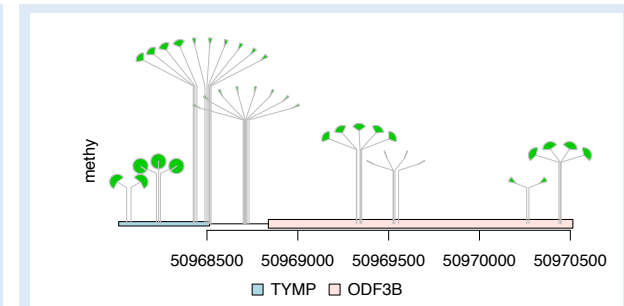
Integrated plot



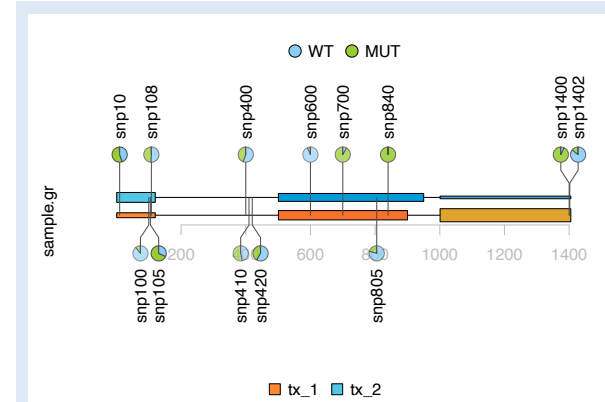
Lollipop plot with different symbols



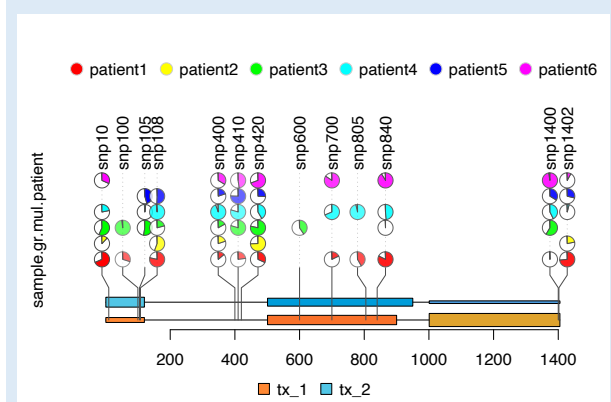
Dandelion plot for depicting dense methylation data



Lollipop plot in caterpillar layout



Lollipop plot in pie.stack layout



trackViewer

select TxDb package
TxDb.Hsapiens.UCSC.hg19.knownGene

select Org package
org.Hs.eg.db

chromosome
22

start
50968014

end
50970514

☒ include transcripts track

Add lollipop plot track from file

select file
chr22_mutation_frequency.bed

file format
BED

sample name
chr22_mutation_frequency.bed

lollipop plot type
circle

gene model
☒ none
☐ use TxDb package
☐ load from the following file

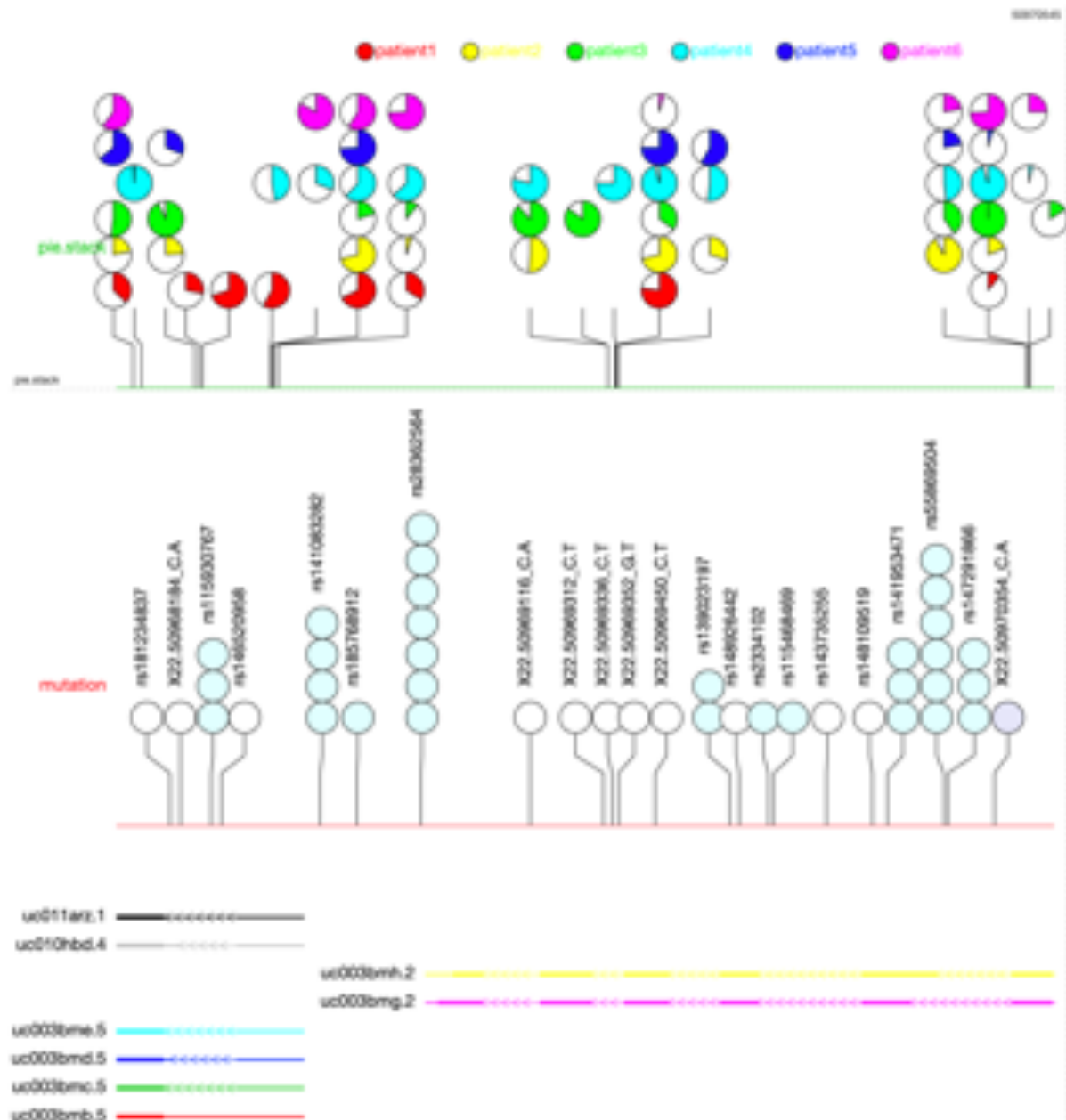
select transcript file
chr17_amps_for_dandelion.bed

transcript file format
BED

☒ remove above track

Add lollipop plot track from file

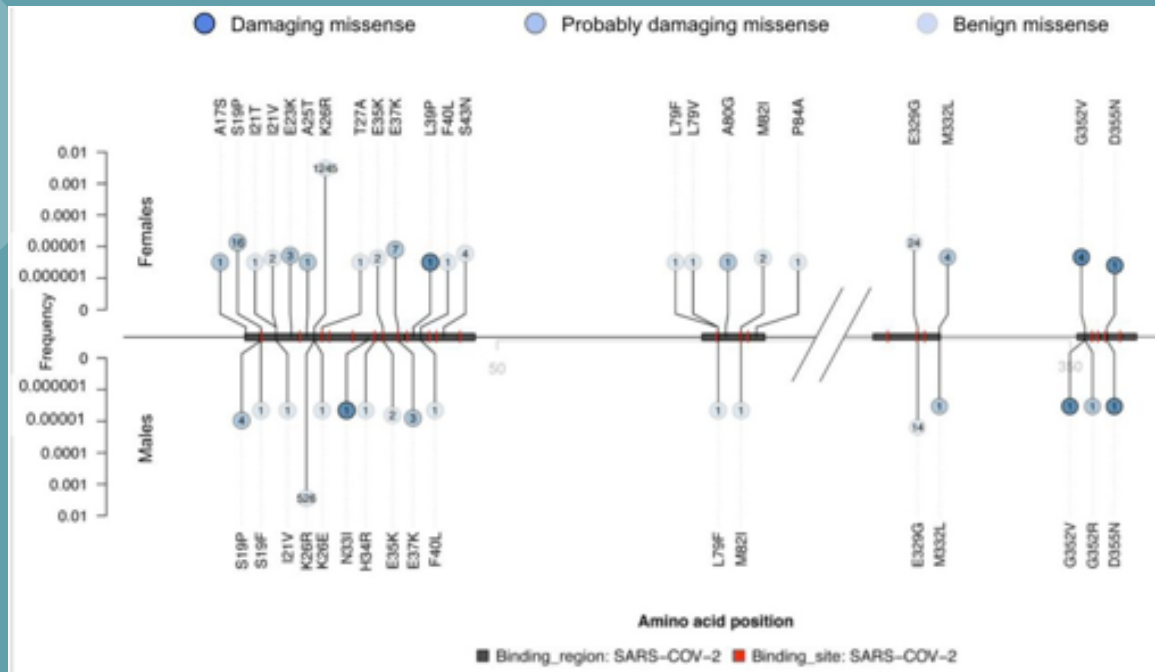
select file



<https://github.com/jianhong/trackViewer.documentation/tree/master/trackViewerShinyApp>

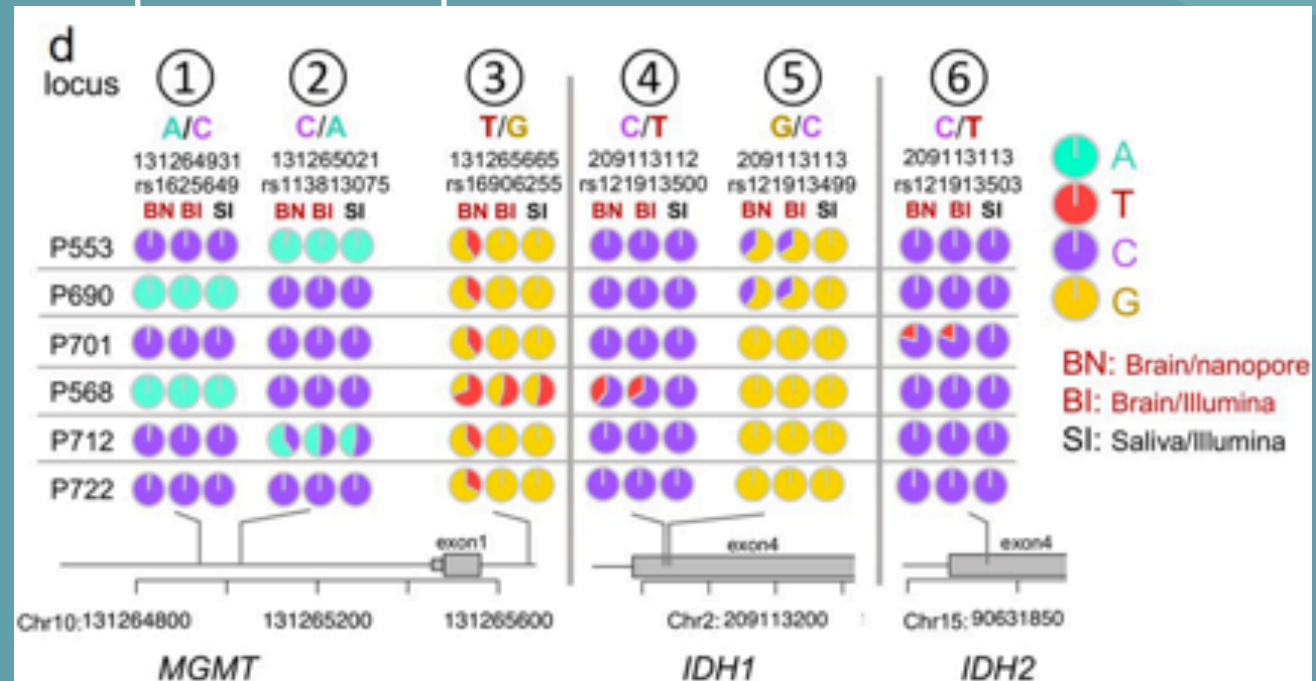
EXAMPLES USING THE trackViewer PACKAGE

Lolliplot of missense variants in the regions of ACE2 that bind to SARS-CoV-2.



Cirulli et.al., 2020. doi: 10.1101/2020.04.07.030544

SNVs in MGMT and IDH1/2 in tumor and saliva samples from 6 patients



Wongsurawat et.al., 2020. doi: 10.1186/s40478-020-00963-0

GRanges Class

`IRanges(start=NULL, end=NULL, width=NULL, names=NULL, ...)`: Creates an IRanges object.

sequence names

coordinates and names

strand information: '+', '-', '*'

`GRanges(seqnames=NULL, ranges=IRanges(), strand=NULL, ...,
seqinfo=NULL, seqlengths=NULL)`: Creates a GRanges object.

chromosome information:
length, is circular or not

chromosome length

GOTO VIGNETTE

<https://jianhong.github.io/trackViewerBiocAsia2020Workshop/>

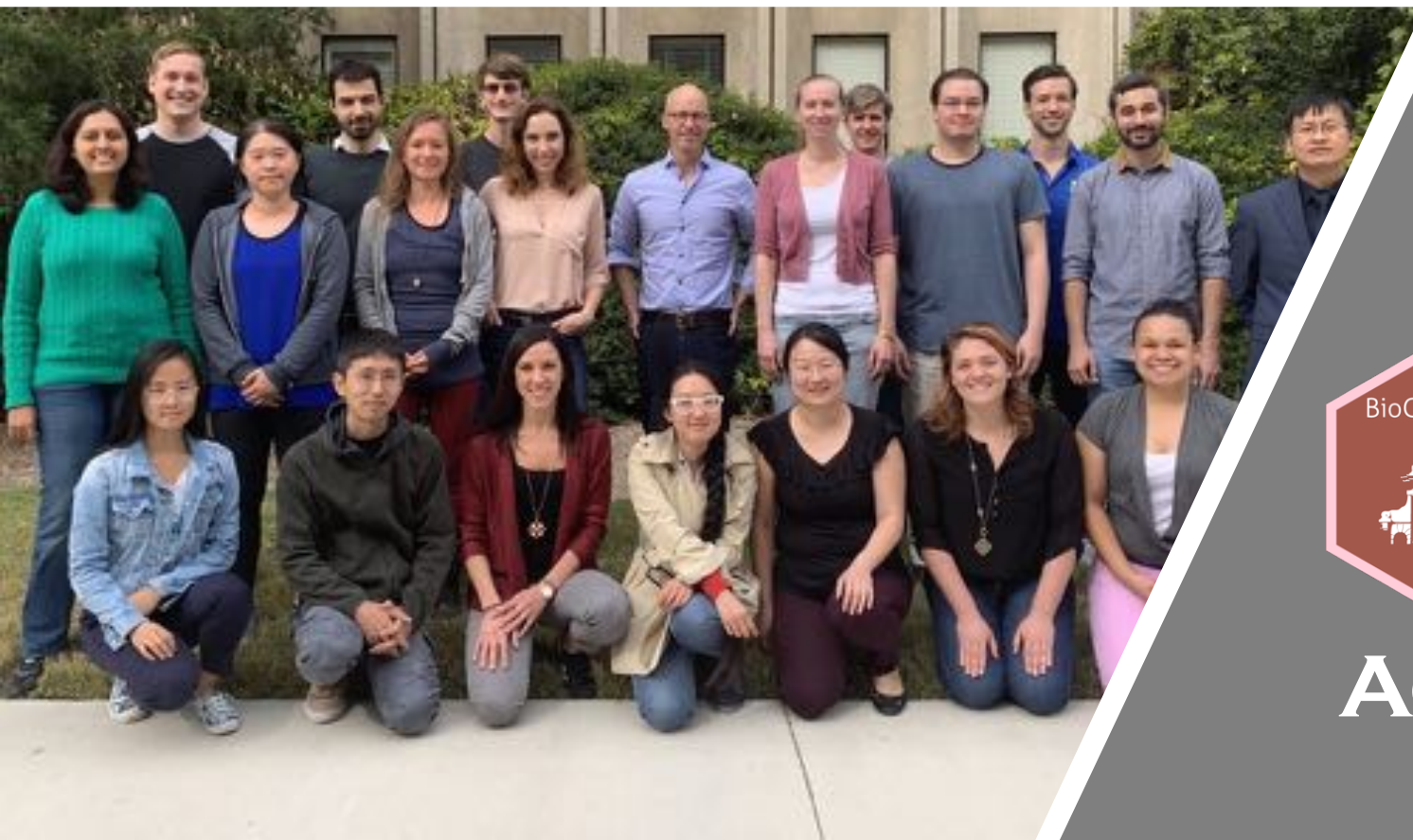
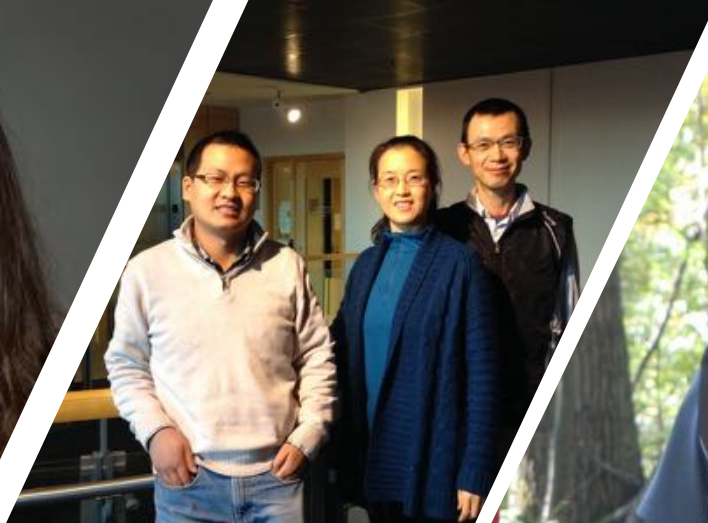
trackViewer **CAN ...**

generate lollipop plot to depict the methylation and SNP/mutation status

visualize regular read coverage tracks

be easily integrated into standard analysis pipeline for various high-throughput sequencing dataset such as ChIP-seq, RNA-seq, methylation-seq or DNA-seq

produce highly customized figure for publication



ACKNOWLEDGEMENT