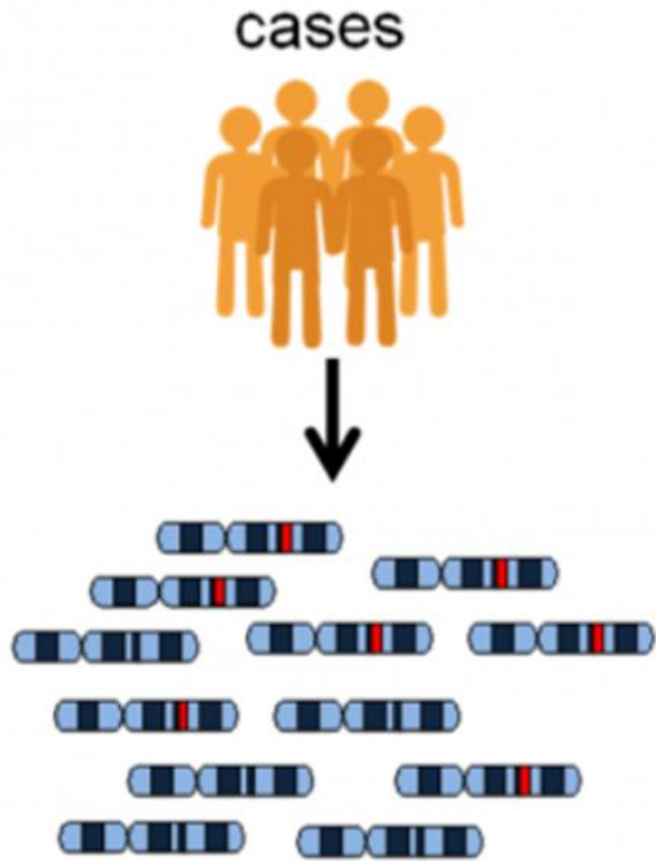


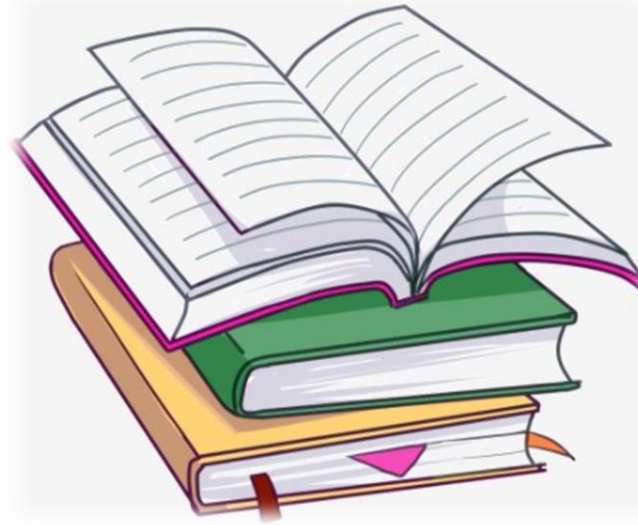
QIAGEN Clinical Insight 變異位點闡釋工具

2021.09.09

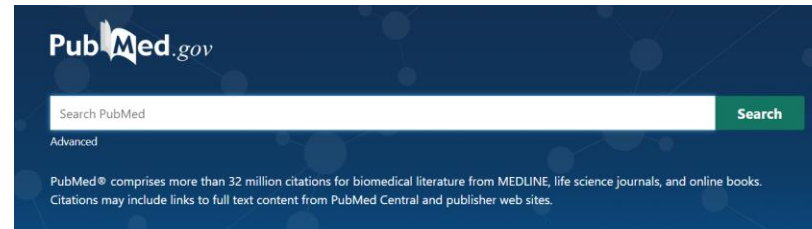
Clair Tsai 蔡宜庭 業務副理
Willis Cheng 鄭耀璋 專案主任



<https://reurl.cc/OXIYvv>

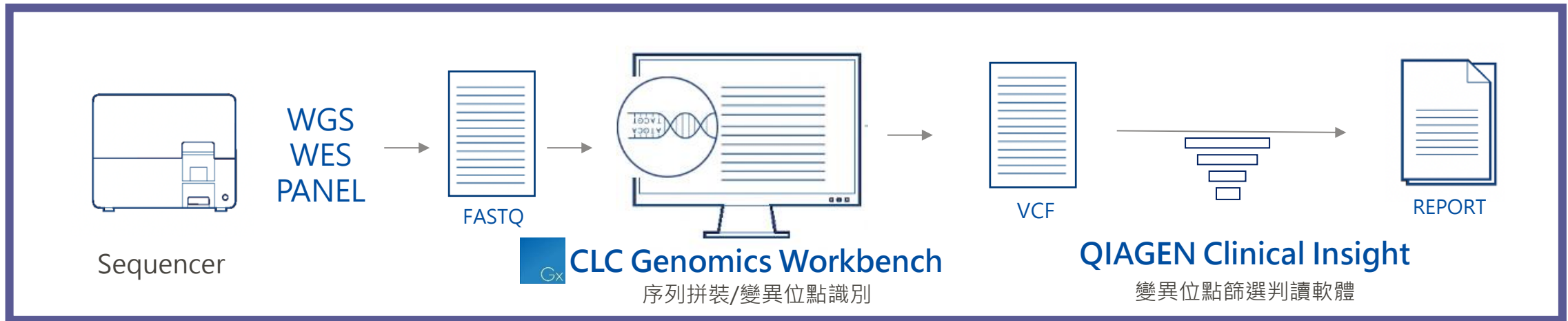


Papers & References



Pubmed or Google

After sequencing, What's Next?



WES全外顯子約有
60,000,000bp

*WGS有30億個鹼基對

與人體參考基因組比對
約有30,000個變異位點

如何找到關鍵制病位點?

1. 常見位點篩選
2. 位點信心程度篩選
3. 資料庫評斷位點制病性
4. 位點與疾病的相關程度
5. 位點用藥資訊
6. 制病位點報告生成

NGS Variant Analysis Service

Variants
number

1: sequencing & base call [FastQ]

```

@PCC078PAC00:2:1101:1184:3385#ACTCAA/2
GGGGCATCAATGATAGTCACATAGTACTTGTGTCTCAAATTTCCACAAGGAGATCAATGATGATACCAGCTTCCAGCTTTC
+
_!_beeeeeeegfgfgehhgfghhhfsgihiiiiiihhchfhiisihhiiiiigdegfhdfghihfdddffgdihdgfgpeddc-dddcd
@PCC078PAC00:2:1101:1162:3439#ACTCAA/2
CTGCTTTCCCTTCATATTCGGAATCCAGCAGTTCTTCTGTGTTTTGCGAGACACAGATATCTTATTTGGGTACACCA
+
!_ceeeeeeceeg'beadf'gbg'chf@bbhiihibd_'cefbbbe]aYaefffthafff[ 'odgd@bd]decdcccc'Z'bbbbBB
@PCC078PAC00:2:1101:1210:3447#ACTCAA/2
GGCCCGCATCACTGTGCTGCTTGAAGGTATAAGCCGCTGAAGTCCCCCGAGTGAAGACATGGCCCTGACTGCACCTCAATCTGGATGCCCG
+
a_bceeeceegpghaeethff]]ebcf'bggfthie]fafh[beghi.fgg0'ZZZ_'Z'dbdecc']bbcbc]bc]bbb@bY_'ba_
:
    
```

2: Variants calling[VCF]

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT
chr1	978603	rs3588118	CCT	C	2883.6	PASS	AC=1;AF=GT;AD:DP:GQ:PL	
chr1	1163811	rs1476947	G	GGACA	2118.64	PASS	AC=1;AF=GT;AD:DP:GQ:PL	
chr1	1177918	rs3835300	CT	C	1614.6	PASS	AC=1;AF=GT;AD:DP:GQ:PL	
chr1	1223465	rs5808105	GC	G	375.6	PASS	AC=1;AF=GT;AD:DP:GQ:PL	
chr1	1247578	.	T	TGG	13	PASS	P=0.9462 GT:VR:RR:DF:GQ	

3: variants
annotation

[annotation file]

Gene	Type	Position	AChange	dbsnp	AlleleFreq	gnomAD_e	gnomAD_c	primaryACMGEvid	primaryACMGL
TSHR	nonsynony	chr14:816	NM_000369.3:exon9:c.	rs1895064	0.0019230	0 22	0 3	pm1 pm2 pm5 pp3	Likely pathogeni
LZTR1	nonsynony	chr22:213	NM_006767.4:exon14:c.	rs1331354	5.57165e-4	0 1	.	pm1 pm2 pm5	Likely pathogeni
C7	frameshift	chr5:4095	NM_000587.4:exon11:c.	rs7458262	0.0006426	0 5	0 1	pvs1 pm2	Likely pathogeni
CACNA1A	nonsynony	chr19:133	NM_001127221.1:exon.	rs5741943	0.0008409	0 15	.	pp2 pm1 pm2 pp3	Likely pathogeni
CACNA1A	nonframes	chr19:133	NM_001127222.2:exon.	pm2 bp3	VUS
HSD17B4	nonsynony	chr5:1188	NM_000414.4:exon4:c.	rs5444551	0.0006524	0 12	.	pm1 pm2 pp3	VUS
COL11A1	nonsynony	chr1:1035	NM_001854.4:exon3:c.	rs1158320	.	.	.	pm1 pm2 bp4	VUS
KMT2C	stopgain	chr7:1519	NM_170606.3:exon18:c.	rs5852856	.	.	.	pvs1 pm2	Likely pathogeni

4: Clinical
report

Gene	Position	Transcript	HGVS(.c) (Exon)	HGVS(.p)	Type	dbSNP	MAF	Heterozygosity	Clinical Level	Disease and Inheritance
MYH7	chr14:2390286 5-23902865	NM_00025 7.3	c.77C>T (exon3)	p.A26V	nonsynony mous SNV	rs186964 570	0.0069	het	Likely path ogenic	Cardiomyopathy, hypertrophic, 1 AD

Format

FASTQ

VCF

Excel / TXT

Excel / TXT

PDF

VCF(Variant Calling Format)

VCF header

```
##fileformat=VCFv4.0
##fileDate=20100707
##source=VCFtools
##reference=NCBI36
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality (phred score)">
##FORMAT=<ID=GL,Number=3,Type=Float,Description="Likelihoods for RR,RA,AA genotypes (R=ref,A=alt)">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##ALT=<ID=DEL,Description="Deletion">
##INFO=<ID=SVTYPE,Number=1,Type=String,Description="Type of structural variant">
##INFO=<ID=END,Number=1,Type=Integer,Description="End position of the variant">
```

Mandatory header lines (points to ##fileformat=VCFv4.0)

Optional header lines (meta-data about the annotations in the VCF body) (points to ##INFO=...)

Body

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE1	SAMPLE2
1	1	.	ACG	A,AT	.	PASS	.	GT:DP	1/2:13	0/0:29
1	2	rs1	C	T,CT	.	PASS	H2;AA=T	GT:GQ	0 1:100	2/2:70
1	5	.	A	G	.	PASS	.	GT:GQ	1 0:77	1/1:95
1	100	.	T		.	PASS	SVTYPE=DEL;END=300	GT:GQ:DP	1/1:12:3	0/0:20

Reference alleles (GT=0) (points to 0/0:29)

Alternate alleles (GT>0 is an index to the ALT column) (points to 1/1:95)

Deletion (points to)

SNP (points to A,AT)

Large SV (points to SVTYPE=DEL;END=300)

Insertion (points to T,CT)

Other event (points to H2;AA=T)

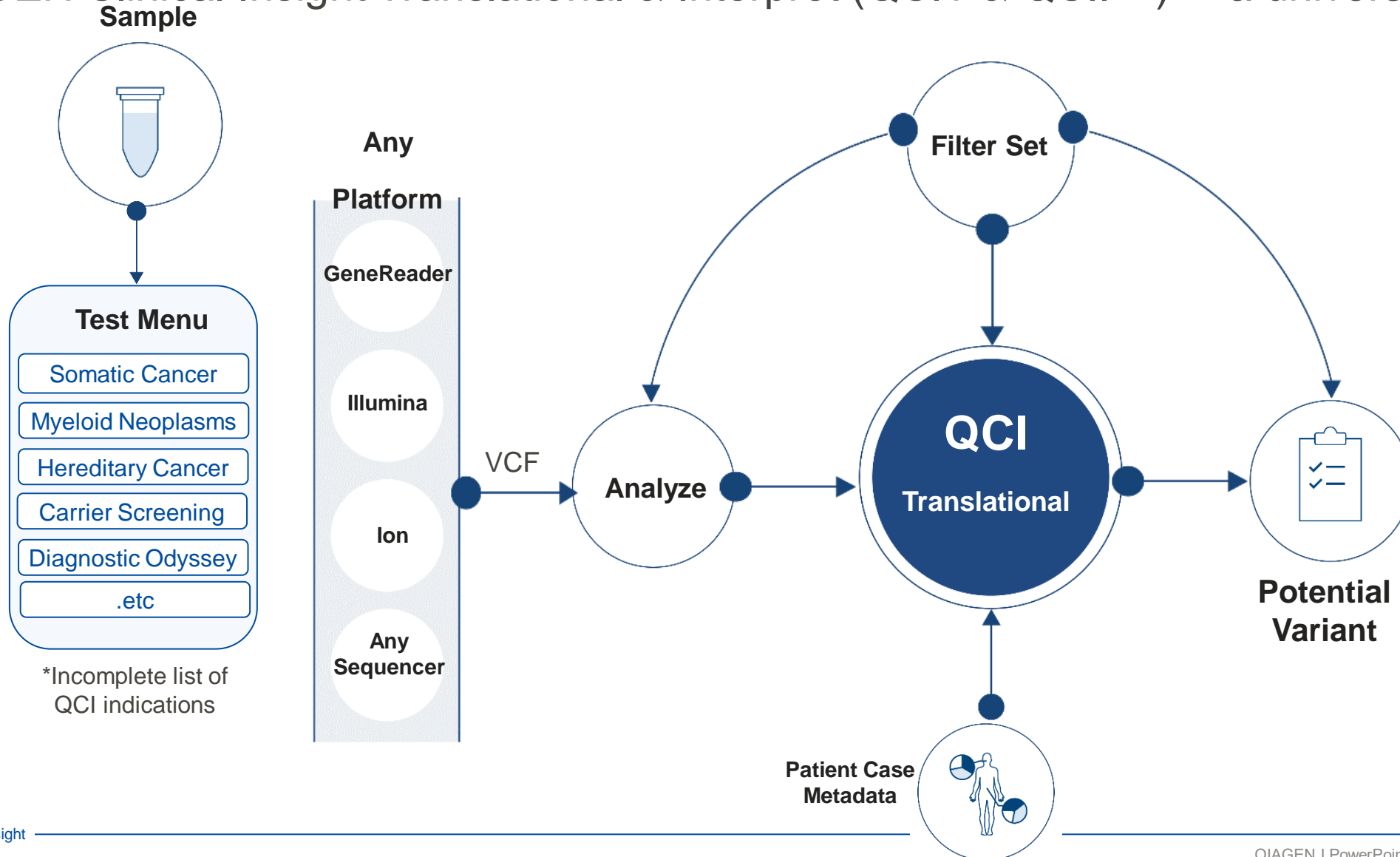
Phased data (G and C above are on the same chromosome) (points to 0|1:100)

<https://weitinglin.com/2017/05/29/vcfvariant-call-format-%E5%9F%BA%E5%9B%A0%E7%AA%81%E8%AE%8A%E8%B3%87%E6%96%99%E5%84%B2%E5%AD%98%E6%A0%BC%E5%BC%8F/comment-page-1/>

Database in QCII/QCIT (Free access in QCII/QCIT software)

Public Databases	Licensed Databases	QIAGEN Databases	Clinical Guidelines	Others
TCGA		Clinical Cases	ACMG	CADD
Clinvar		Clinical & Functional Studies	AMP	Polyphen
dbSNP	COSMIC	HGMD	NCCN	SIFT
1000 Genome	ICGC	PGMD	ASCO	
ESP	OMIM	Curated variants	CAP	PhyloP
gnomAD	BIC	Pathway and causal network	ESMO	Blosum
ExAC	CentO MD	Allele Frequency Community	FDA	MaxEntScan
Clinical Trials			EMA	Mutation Taster

QIAGEN Clinical Insight Translational & Interpret (QCIT & QCII™) – a universal solution



QCII/QCIT difference

	QCIT(偏研究使用)	QCII(偏臨床使用)
功能	註解工具與篩選位點	註解工具與篩選位點 臨床與藥物資料提供臨床判讀使用
應用	<ul style="list-style-type: none"> • 可單一或多樣本分析，或是家族(trio)分析 • 可做群組分析(cohort study) 	<ul style="list-style-type: none"> • 出具臨床報告
優勢	<ul style="list-style-type: none"> • Qiagen內建database(含ACMG guideline) • 有權限管理系統 • 可設定多種分析流程 	<ul style="list-style-type: none"> • Qiagen內建database(含ACMG&AMP guideline) • 可客製報告模板 • 有權限管理系統 • 可設定多種分析流程(TPP) • 可設定報告簽核系統
輸出	Excel表格(註釋資料)	Excel表格與檢測報告

How to log in QCI System

全部 新聞 圖片 地圖 影片 更多 工具

約有 8,220,000 項結果 (搜尋時間 : 0.43 秒)

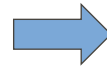
提示 : 只顯示繁體中文搜尋結果。您可以在使用偏好中指定搜尋語言

<https://digitalinsights.qiagen.com> > produ... ▾ 翻譯這個網頁

Product Log in - QIAGEN Digital Insights

Login to IVA, IPA, QCI, MyCLC. ... Log in. Ingenuity Pathway Analysis. Ingenuity Pathway Analysis for China. QCI Interpret Translational for USA ...

您曾多次瀏覽這個網頁。上次瀏覽日期 : 2021/8/30



	QCI Interpret Translational for USA QCIT
	QCI Interpret Translational for China
	QCI Interpret for USA QCII

Sample & Test List

Search... 169 Tests

< 1 2 3 4 >

Create Test

Accession ID	Test Product Code	State	Filter By	Days
220259257_104.all_variants	ABC - Somatic somatic	In Review genechen@gga.asia Aug 29, 2021 11:30:19 AM		3 Days Aug 29, 2021 9:45:32 AM
201945505_103.all_variants	ABC - Somatic somatic	Pending genechen@gga.asia		3 Days Aug 29, 2021 9:44:12 AM
220291795_109.all_variants	ABC - Somatic somatic	Pending genechen@gga.asia		3 Days Aug 29, 2021 9:42:51 AM
225876580_109.all_variants	ABC - Somatic somatic	Pending genechen@gga.asia		3 Days Aug 29, 2021 9:41:18 AM
22000503_107.all_variants	ABC - Somatic somatic	Pending genechen@gga.asia		3 Days Aug 29, 2021 9:41:18 AM

Test List
(Your Analysis)

Search... 210 samples

< 1 2 3 4 5 >

Create Test Upload Sample

Sample Name	Subject ID	Project	State	Filter By	Date
220478986_109.all_variants	220478986_109.all_variants		Active genechen@gga.asia		Aug 29, 2021 9:25:50 AM
220478986_106.all_variants	220478986_106.all_variants		Active genechen@gga.asia		Aug 29, 2021 9:24:44 AM
220006970_109.all_variants	220006970_109.all_variants		Active genechen@gga.asia		Aug 29, 2021 9:15:02 AM
220169946_109.all_variants	220169946_109.all_variants		Active		Aug 29, 2021

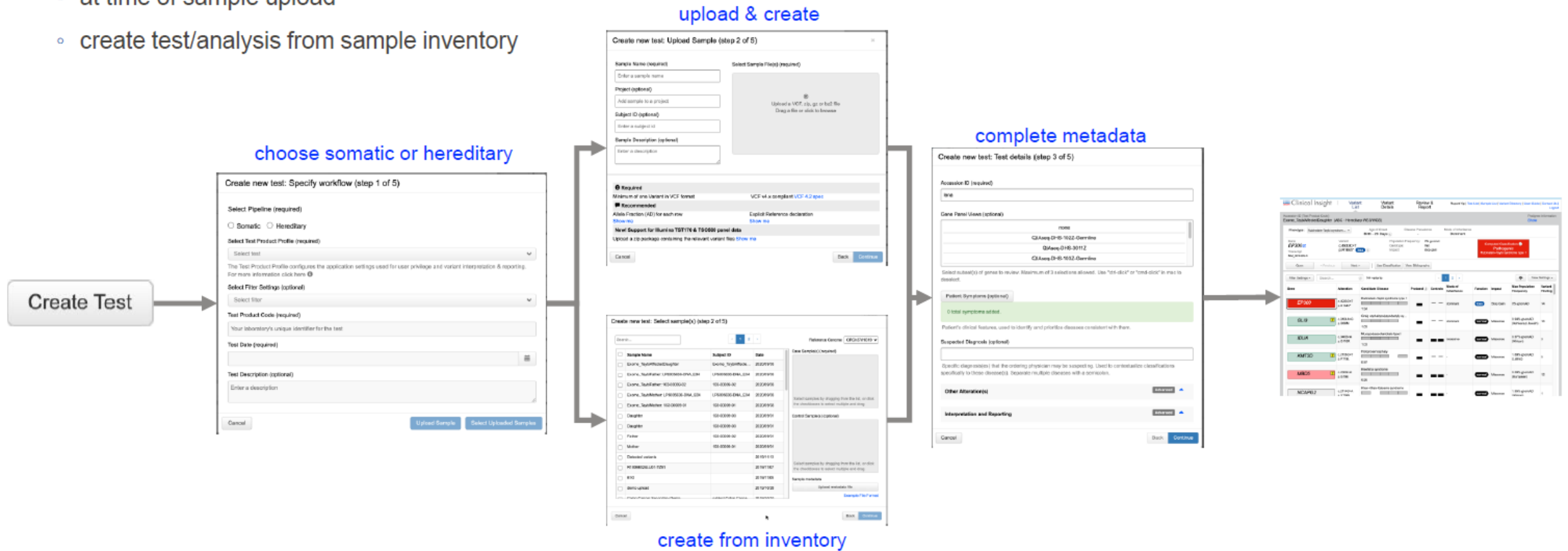
Sample List
(Your Uploaded Samples)

Create Your Variant Analysis Strategically

New Create Test button to start test creation workflow

Create new test/analysis

- at time of sample upload
- create test/analysis from sample inventory



QIAGEN Clinical Insight Interface

Pathogenic

Likely Pathogenic

VUS

Likely Benign

Benign

Actionability

AMP/ASCO/CAP Guidelines – Somatic Testing**

Tier 1A Strong clinical significance
Tier 1B

Tier 2C Potential clinical significance
Tier 2D

Tier 3 Unknown clinical significance

Tier 4 Likely benign or benign

Clinical Insight | Variant List | Variant Details | Review & Report

Gene Chen | Test List | Sample List | Variant Directory | User Guide

Accession ID (Test Product Code)
TestA: A4 (ABC - Somatic)

1 Variant Basic Information

Sex
Female

Ethnicity
-

Phenotype: Breast cancer | Age of Onset: 61 Years | Gene Prevalence: 20% | Disease Prevalence: 1/77

Gene: *PIK3CA* | Variant: c.1624G>A p.E542K gain | Somatic Frequency: 2.53% | Population Frequency: 0% gnomAD | Allele Fraction: 35% (of 60 reads) | Impact: missense

2 ACMG & AMP Guideline

Computed Classification
Tier 1A
Pathogenic
Breast cancer

New Assessment
Tier 1A
Pathogenic
for Breast cancer
Reportable

3 Filter Setting

4 View Variant List

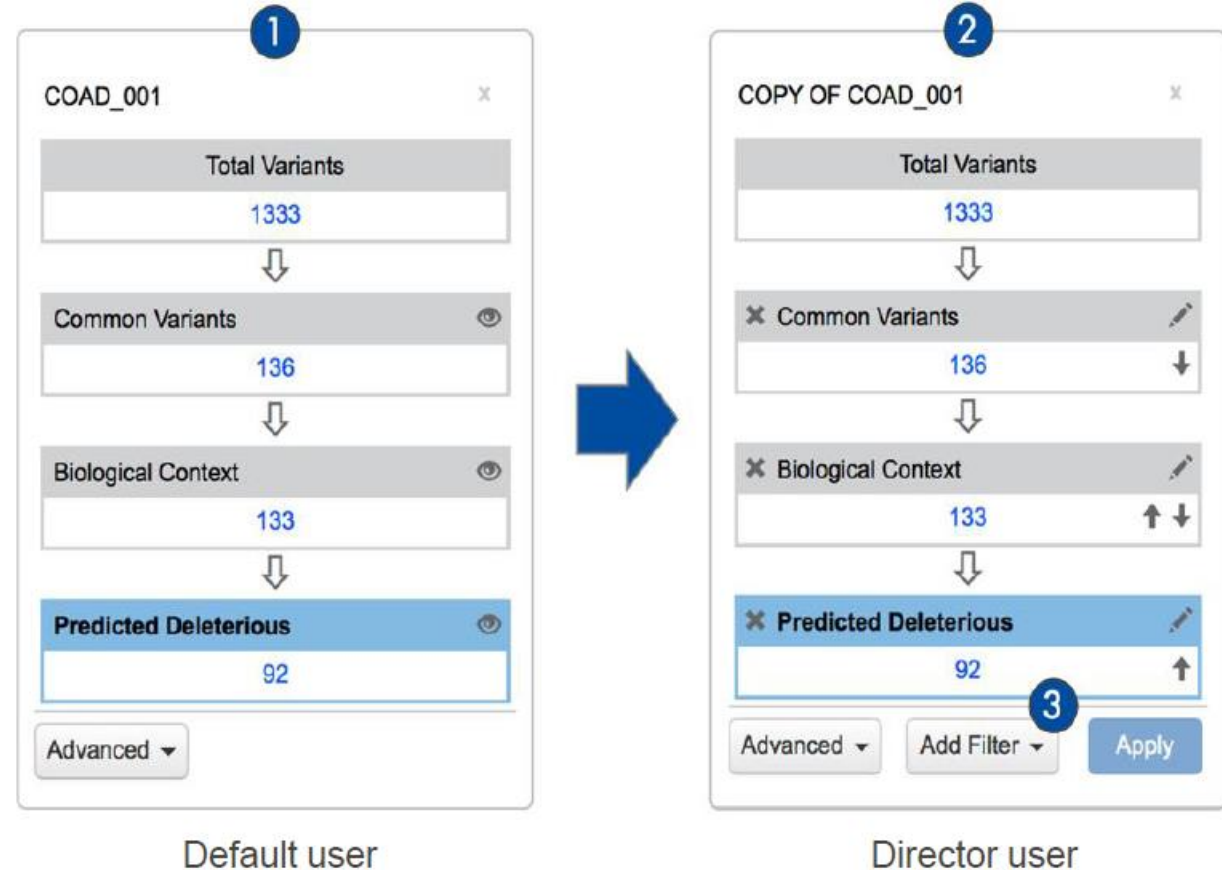
5 View Setting

Filter Settings | Search... | 39 variants

Biomarker	Alteration	Function	Impact	Case - Quantity	Somatic Frequency	Max Population Frequency
<i>PIK3CA</i> 1A Pathogenic	c.1624G>A p.E542K	gain	Missense	35% (of 60 reads)	2.53%	0% gnomAD
<i>ESR1</i> 1B Pathogenic	c.1610A>C p.Y537S	gain	Missense	24% (of 74 reads)	0.30%	0% gnomAD
2C <i>FANCD2</i>	c.1278+3_1278+6delAAGT	loss	-	14% (of 74 reads)	0%	0.001% gnomAD (European)
2C <i>ATRX</i>	c.2671G>C p.E891Q	loss	Missense	72% (of 50 reads)	0%	0% gnomAD
3 <i>CYP2D6</i>	c.1457G>C p.S486T	loss	Missense	63% (of 40 reads)	0%	0% gnomAD
3 <i>HLA-DRB1</i>	c.115C>T p.Q39*	loss	Stop Gain	26% (of 39 reads)	0%	0% gnomAD
3 <i>PRSS1</i>	c.47C>T p.A16V	loss	Missense	32% (of 44 reads)	0%	4.32% gnomAD (African)

Characterizing the Potential Variants by Databases and Filter Set

- QCI users can view the filtering strategy to see how and where variants are filtered out
- Clicking on blue numbers updates the variant table with list of filtered variants



The image illustrates the difference in variant filtering capabilities between a Default user and a Director user. A large blue arrow points from the Default user interface to the Director user interface.

Default user (labeled 1): The interface for 'COAD_001' shows a hierarchy of filters. The 'Total Variants' section displays 1333. Below it, 'Common Variants' shows 136, 'Biological Context' shows 133, and 'Predicted Deleterious' shows 92. The 'Predicted Deleterious' section is highlighted in blue. At the bottom, there is an 'Advanced' dropdown menu.

Director user (labeled 2): The interface for 'COPY OF COAD_001' shows the same hierarchy. The 'Total Variants' section displays 1333. The 'Common Variants' and 'Biological Context' sections are marked with an 'X' and include edit and expand/collapse icons. The 'Predicted Deleterious' section is also marked with an 'X' and includes an edit icon and an expand/collapse icon. At the bottom, there is an 'Advanced' dropdown menu, an 'Add Filter' dropdown menu, and an 'Apply' button. A blue circle with the number 3 is positioned over the 'Add Filter' dropdown.

Variant Filtering Sets

The following 4 filters are introduced to QCI variant filtering:

- **Common Variants**
 - remove/include variants based on population frequency cut off
- **Confidence**
 - remove/include variants based on variant calling metrics
- **Predicted Deleterious**
 - select or variants based on ACMG, HGMD, ClinVar , literature evidence, or in silico predictions
- **Biological Context**
 - based on relationships (direct/indirect) between gene and biological terms within QIAGEN ontology/QKB

Test Performed: Somatic Panel

Report Date Nov 8, 2020
Status -

Patient
Patient Name Michelle Doe
Date of Birth
Age
Sex Female
Ethnicity
Diagnosis Breast Cancer

Client
Client General Hospital
Client ID ABC123
Physician Dr. E Smith
Pathologist Dr. R Jones

Specimen
Specimen TestA: A4
Specimen biopsy
Collection Nov 9, 2020
Accession Nov 9, 2020
Primary Tumor Site Breast

Result: Positive

2 Clinically Significant Variants
5 Therapies Associated with Resistance
8 Therapies with Potential Clinical Benefit
22 Clinical Trials

Report Summary

PIK3CA E542K was identified and is associated with an available treatment. One alteration is associated with resistance to aromatase inhibitor therapy.

#####

Actionable Variants With Associated Therapies

Gene / Variant	Allelic Fraction	Approved Therapies			Clinical Trials
		Breast Cancer	Other Indications	Associated With Resistance	
PIK3CA c.1624G>A p.E542K g.179218294G>A Tier 1A Pathogenic	35.0% (of 60 reads)	alpelisib alpelisib /fulvestrant lapatinib /letrozole letrozole	-	-	19
ESR1 c.1610A>C p.Y537S g.152098788A>C Tier 1B Pathogenic	24.0% (of 74 reads)	fulvestrant neratinib tamoxifen toremifene	-	anastrozole aromatase inhibitor fulvestrant letrozole tamoxifen	3

Gene / Variant	Trial Title Trial ID	Treatments	Trial Phase	Location / Contact
ESR1 p.Y537S g.152098788A>C Tier 1B Pathogenic	A Phase 1 Study of SY 5609, an Oral, Selective CDK7 Inhibitor, in Adult Patients With Select Advanced Solid Tumors NCT04247126	SY-5609 fulvestrant	Phase 1	United States: MI, OK, PA, TN, TX Kimberley Caliri; kcaliri@syros.com; 617-674-9053;
ESR1 p.Y537S g.152098788A>C Tier 1B Pathogenic	INTERACT- Integrated Evaluation of Resistance and Actionability Using Circulating Tumor DNA in HR Positive Metastatic Breast Cancers NCT04256941	anastrozole letrozole ribociclib abemaciclib /letrozole letrozole /palbociclib palbociclib abemaciclib fulvestrant letrozole /ribociclib	Phase 2	United States: TX Senthilkumar Damodaran; sdamodaran@mdanderson.org; 713-792-2817;

Individual Variant Interpretations

Gene PIK3CA Exon 10 Nucleotide NM_006218.4: g.179218294G>A c.1624G>A Amino Acid p.E542K Function gain Allelic Fraction 35.0% (of 60 reads) Classification Tier 1A Assessment Pathogenic	Interpretation PIK3CA encodes the protein p110-alpha, which is the catalytic subunit of phosphatidylinositol 3-kinase (PI3K). The PI3K pathway is involved in cell signaling that regulates a number of critical cellular functions, including cell growth, proliferation, differentiation, motility, and survival [16, 6]. PIK3CA mutations are not mutually exclusive with EGFR or KRAS or BRAF mutations, and are associated with increased PI3K signaling and increased activation of Akt [22, 10]. Activating missense mutations in PIK3CA, including the E542K, E545K, and H1047R hotspot mutations, predominantly occur in the helical (exon 9) and kinase (exon 20) domains [17, 2].
Gene ESR1 Exon 10 Nucleotide NM_001122742.1: g.152098788A>C c.1610A>C Amino Acid p.Y537S Function gain Allelic Fraction 24.0% (of 74 reads) Classification Tier 1B Assessment Pathogenic	Interpretation ESR1 encodes estrogen receptor alpha (ER-alpha), one of the major estrogen receptor isoforms in humans; binding of estrogen to ER-alpha promotes its translocation to the nucleus and the transcriptional activation of genes involved in cell cycle progression and survival [14]. Mutation or amplification of ESR1 and activation of ER-alpha may result in the upregulation of genes involved in cell cycle progression and survival, and ER-alpha signaling has been implicated in a number of cancer types [14, 12, 4, 21, 15]. However, ER-alpha may act as a tumor suppressor in some cancers [23, 3, 1, 5].

CASE STUDY

1. 針對單一病人表型之癌症病患篩選位點並提供用藥資訊
2. 針對同群體病人表型來判讀變異位點資訊
 - Test-demo: 4 samples
 - Breast cancer

test_demo.vcf

```
test_demo: test7
```

```
test_demo: test4
```

```
test_demo: test3
```

```
test_demo: test1
```

Sample Upload

- Samples in multiple VCF files or in single VCF file will be separate into individual sample record.

Upload Sample
✕

Sample Name (required)

Project (optional)

Subject ID (optional)

Sample Description (optional)

Select Sample File(s) (required)

⊕

Upload a vcf, zip, gz, csv, tsv or bz2 file
Drag a file or click to browse

test_demo.vcf ✕

1 files selected (maximum 200)

[Submit](#)

ⓘ Required

Minimum of one Variant in VCF format [VCF v4.x compliant VCF 4.2 spec](#)

📌 Recommended

Allele Fraction (AD) for each row [Explicit Reference declaration](#)

[Show me](#) [Show me](#)

New! Support for Illumina TST170 & TSO500 panel data

Upload a zip package containing the relevant variant files [Show me](#)

[Close](#)



Sample Name	File name	Status	Date
test_demo: test1	test_demo.vcf	queued <small>ATTG</small>	09/02/2021
test_demo: test3	test_demo.vcf	queued <small>ATTG</small>	09/02/2021
test_demo: test4	test_demo.vcf	queued <small>ATTG</small>	09/02/2021
test_demo: test7	test_demo.vcf	queued <small>ATTG</small>	09/02/2021

Set a new analysis

Create new test: Specify workflow (step 1 of 5)

Select Pipeline (required)

Hereditary Somatic

Select Test Product Profile (required)

QCI Interpret Somatic Default

The Test Product Profile configures the application settings used for user privilege and variant interpretation & reporting. For more information click here

Select Filter Settings (optional)

Select or search filter

Test Product Code (required)

ABC - Somatic

Test Date (required)

Test Description (optional)

Enter a description



Create new test: Select sample(s) (step 2 of 5)

Search...

1
2

Reference Genome GRCh38/HG38

	Sample Name	Subject ID	Date
<input type="checkbox"/>	220478986_109.all_variants	220478986_109.all...	2021/08/29
<input type="checkbox"/>	220478986_106.all_variants	220478986_106.all...	2021/08/29
<input type="checkbox"/>	220006970_109.all_variants	220006970_109.all...	2021/08/29
<input type="checkbox"/>	220169946_109.all_variants	220169946_109.all...	2021/08/29
<input type="checkbox"/>	220000503_107.all_variants	220000503_107.all...	2021/08/29
<input type="checkbox"/>	225876580_109.all_variants	225876580_109.all...	2021/08/29
<input type="checkbox"/>	220291795_109.all_variants	220291795_109.all...	2021/08/29
<input type="checkbox"/>	201945505_103.all_variants	201945505_103.all...	2021/08/29
<input type="checkbox"/>	220259257_104.all_variants	220259257_104.all...	2021/08/29
<input type="checkbox"/>	S221820555_109.all_variants	S221820555_109.all...	2021/08/23
<input type="checkbox"/>	NGSO-210416001_S1	NGSO-210416001_S1	2021/05/10
<input type="checkbox"/>	N532	N532	2021/04/28
<input type="checkbox"/>	Identified_variants-Yang3_S3 (paired)	Identified_variants-Y...	2021/04/15
<input type="checkbox"/>	PROMDVariants_passing_filters-NGSO-20103...	PROMDVariants_pa...	2020/12/28

Case Sample(s) (required)

test_demo: test1

test_demo: test3

test_demo: test4

test_demo: test7

Select samples by dragging from the list, or click the checkboxes to select multiple and drag

Control Sample(s) (optional)

Select samples by dragging from the list, or click the checkboxes to select multiple and drag

Cancel

Back

Continue

Create new test: Test details (step 3 of 5)

Accession ID (required)

Keep only confidently detected variants (optional)
 Keep only variants that pass upstream pipeline quality filtering (VCF FILTER column is PASS or '!')

Gene Panel Views (optional)

<input type="checkbox"/>	QIAact-BRCA-UMI-FFPE_QIAGEN
<input type="checkbox"/>	Illumina® TruSight™ Tumor 170
<input type="checkbox"/>	Illumina® TruSight™ Oncology 500
<input type="checkbox"/>	QIAact-Lung-Plasma-Track

Select subset(s) of genes to review. Maximum of 3 selections allowed. Use "ctrl-click" or "cmd-click" in mac to deselect.

Diagnosis (required)

This diagnosis will be used to match treatments and trials and for display on the report.

Diagnosis Stage (optional)

This is used to contextualize clinical trials to the selected stage. If left unselected, all clinical trials pertinent to the diagnosis will be displayed.

Create new test: Patient Information (step 4 of 5)

Information about the subject

Sex (optional)

Male Female

This is used to limit to gender-appropriate clinical trials.

Ethnicity (optional)

Date of Birth (optional)

Age (recommended)

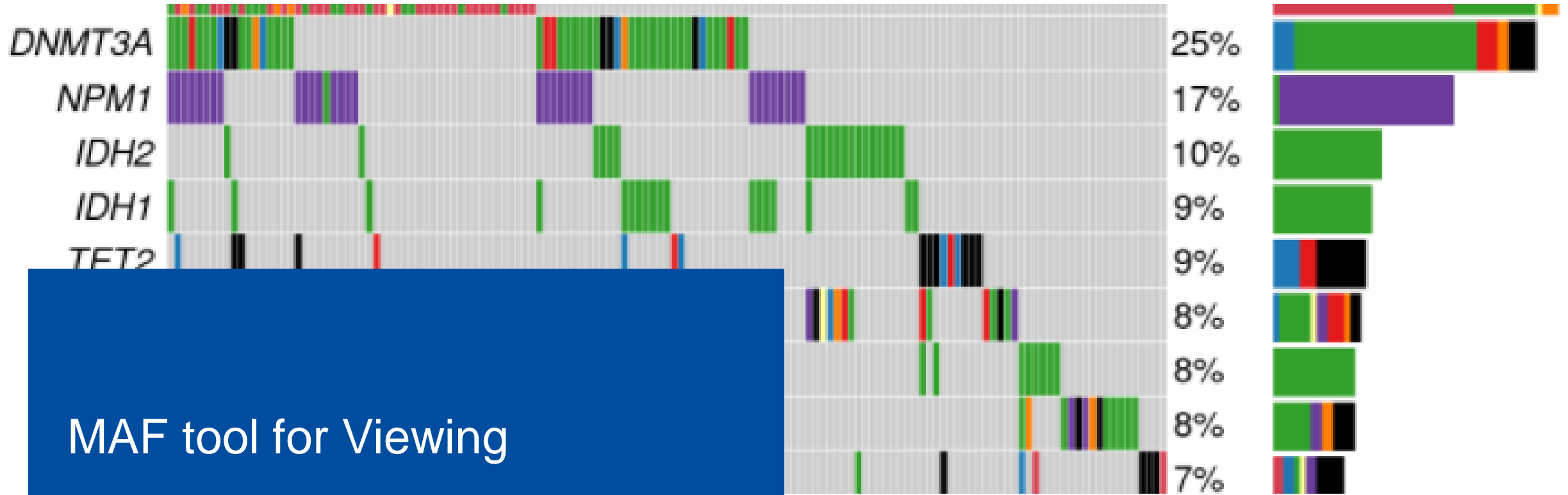
This is used to limit to age-appropriate clinical trials (e.g. pediatric for children).

Create new test: Reporting (step 5 of 5)

The information below is used only on the final report. For validation samples, you can use the data that has been pre-filled for you below, or hit Clear Data and enter your own information.

Subject Name (optional)

Client Name (optional)



MAF tool for Viewing

MAF Format

- MAF (Mutation Annotation Format)
 - A Mutation Annotation Format (MAF) file (.maf) is a **tab-delimited text file** that lists mutations. The format originates from The Cancer Genome Atlas (TCGA) project and is described in detail [here](#). As such, the format pertains to human genomes.
 - In the context of human cancer, MAF files come in two types--protected and somatic. These two types extend conceptually to (1) mutation files that contain all sequenced mutations--however mutations are defined--e.g. against matched normal tissue, against a reference much like for the [VCF format](#), or against another tumor stage, and (2) mutation files that filter the mutation types that are listed based on set criteria, e.g. only somatic mutations.
 - <https://software.broadinstitute.org/software/igv/MutationAnnotationFormat>

Requirement of MAF File

Field	Description	Allowed values
Hugo_Symbol	HUGO gene symbol	-
Chromosome	Chromosome no.	1-22, X, Y
Start_Position	Event start position	Numeric
End_Position	Event end position	Numeric
Reference_Allele	Positive strand reference allele	A, T, C, G
Tumor_Seq_Allele2	Primary data genotype	A, T, C, G
Variation_Classification	Translational effect of variant allele	Frame_Shift_Del, Frame_Shift_Ins, In_Frame_Del, In_Frame_Ins, Missense_Mutation, Nonsense_Mutation, Silent, Splice_Site, Translation_Start_Site, Nonstop_Mutation, 3'UTR, 3'Flank, 5'UTR, 5'Flank, IGR, Intron, RNA, Targeted_Region, De_novo_Start_InFrame, De_novo_Start_OutOfFrame, Splice_Region, Unknown
Variation_Type	Variation Type	SNP, DNP, INS, DEL, TNP and ONP
Tumor_Sample_Barcode	Sample ID	Either a TCGA barcode, or for non-TCGA data, a literal SAMPLE_ID as listed in the clinical data file

<https://github.com/umccr/MAF-summary#maf-field-requirements>

Example of MAF

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Hugo_Sym	Entrez_Ge	Center	NCBI_Bu	Chromoso	Start_Posit	End_Posit	Strand	Variant_C	Variant_T	Reference	Tumor_Se	Tumor_Se	Tumor_Sa	Protein_Cli	TumorV	i_transcript_name	
2	AHDC1			38	1	27552023	27552024		Frame_Sh	INS			CC	A1	p.P32fs*86	37.84		
3	DPYD			38	1	97883329	97883329		Missense_SNP		A	A	G	A1	p.C29R	51.16		
4	FANCD2			38	3	10046723	10046726			DEL	AGTA	AGTA		A1		58.9		
5	TREX1; ATRIP-TREX1; ATR			38	3	48467186	48467186		synonymo	SNP	T	T	C	A1	p.Y38Y; p	100		
5	FRG2C			38	3	75664413	75664413		Missense_SNP		T	T	C	A1	p.C12R	84.21		
7	FRG2C			38	3	75665147	75665147		Missense_SNP		A	A	G	A1	p.Q93R; p	18.75		
3	FRG2C			38	3	75665186	75665186		Missense_SNP		T	T	G	A1	p.I105S; p	20		
9	ZNF717; MIR4273			38	3	75737105	75737105		Nonstop_1	SNP	T	T	A	A1	p.K840*; j	60.76		
0	ZNF717; MIR4273			38	3	75737127	75737127		Missense_SNP		A	A	C	A1	p.H832Q; j	58.57		
1	ZNF717; MIR4273			38	3	75737676	75737676		Nonstop_1	SNP	A	A	T	A1	p.C599*; j	53.85		
2	ZNF717; MIR4273			38	3	75737948	75737949		Frame_Sh	DEL	CT	CT		A1	p.E560fs*	28.81		
3	HLA-DRB1			38	6	32584174	32584174		Missense_SNP		G	G	C	A1	p.A102G	100		
4	HLA-DRB1			38	6	32584364	32584364		Nonstop_1	SNP	G	G	A	A1	p.Q39*	74.07		
5	CYP3A7-CYP3A51P; CYP3A7			38	7	99709062	99709062		Missense_SNP		G	G	C	A1	p.T409R	41.43		
6	PRSS1			38	7	1.43E+08	1.43E+08		Missense_SNP		C	C	T	A1	p.A16V	29.87		
7	LRRC6			38	8	1.33E+08	1.33E+08		Nonstop_1	SNP	G	G	A	A1	p.R60*; p	37.23		

How to Make MAF File

- Straight Forward:
 - Coding
 - No matter you use Python, R, Perl, or other languages.
 - The goal is to make a tab-delimited text file for the data structure of MAF format
 - The only limitation is that almost the profiles are from different styles of data by users
 - By Program from other' programmers
 - Format specific!
 - Using Excel (Maybe another nice choice)

Install R



[Home]

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing Lists](#)

[Developer Pages](#)

[R Blog](#)

R Foundation

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

Help With R

[Getting Help](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- [R version 4.1.1 \(Kick Things\)](#) has been released on 2021-08-10.
- [R version 4.0.5 \(Shake and Throw\)](#) was released on 2021-03-31.
- Thanks to the organisers of useR! 2020 for a successful online conference. Recorded tutorials and talks from the conference are available on the [R Consortium YouTube channel](#).
- You can support the R Foundation with a renewable subscription as a [supporting member](#)

News via Twitter



Available Packages

Currently, the CRAN package repository features 18000 available packages.

<https://cran.csie.ntu.edu.tw/>

Download and Install R


Precompiled binary distributions of the base system and contribut R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)


Install Rstudio (Option)

<https://www.rstudio.com/products/rstudio/>

There are two versions of RStudio:



RStudio Desktop
Run RStudio on your desktop



RStudio Server
Centralize access and computation

https://joe11051105.gitbooks.io/r_basic/content/environment_settings/install_RStudio.html

https://joe11051105.gitbooks.io/r_basic/content/environment_settings/RStudio_introduction.html

Installation of maftools

- On R console:
- Code:

```
if (!require("BiocManager"))  
  install.packages("BiocManager")  
BiocManager::install("maftools")
```

Coding Approach

- Key point:
 - Using data frame (in Python) or data table (in R) to sort out all the required information from VCF files or other text files.
 - Using String processing method to retrieve the best information what you want
 - General key word on languages (R, python)
 - split
 - re (regular expression)
 - **Make sure the information is fitted to the data structure and value.**
 - E.g.:
 - Reference allele \neq Tumor_Sample_allele1
 - The value of Allele frequency is split from original attribute in VCF (GT)
 - Export tab-delimited format and the extension of file is “.maf”

Note on R

- rbind, cbind, or '\$' operator to generate column
 - DataTable
 - Combine columns to make your maf file
 - Save as txt files

Make your MAF Format by yourself

```

library("maftools")

### Make MAF from your csv
# Read csv files from QCII

df <- read.csv("XXXXXXXXXXXXXXXXXXXXXXXXXXXX/test_demo_QCII.csv", sep=";", header=TRUE)

### Retrieve the key columns from dataframe

maf <- df['Gene.symbol']
column?(maf) <- "HUGO"
maf$Entrez_Gene_Id <- NA
maf$Center <- NA
maf$NCBI_Build <- 38
maf$Chromosome <- df['Chromosome']
maf$Start_Position <- df['Position']
maf$End_Position <- df['End_Position']
maf$Strand <- NA
maf$Variant_Classification <- df['Translation.Imp?ct']
maf$Variant_Type <- df['Variation.Type']

```

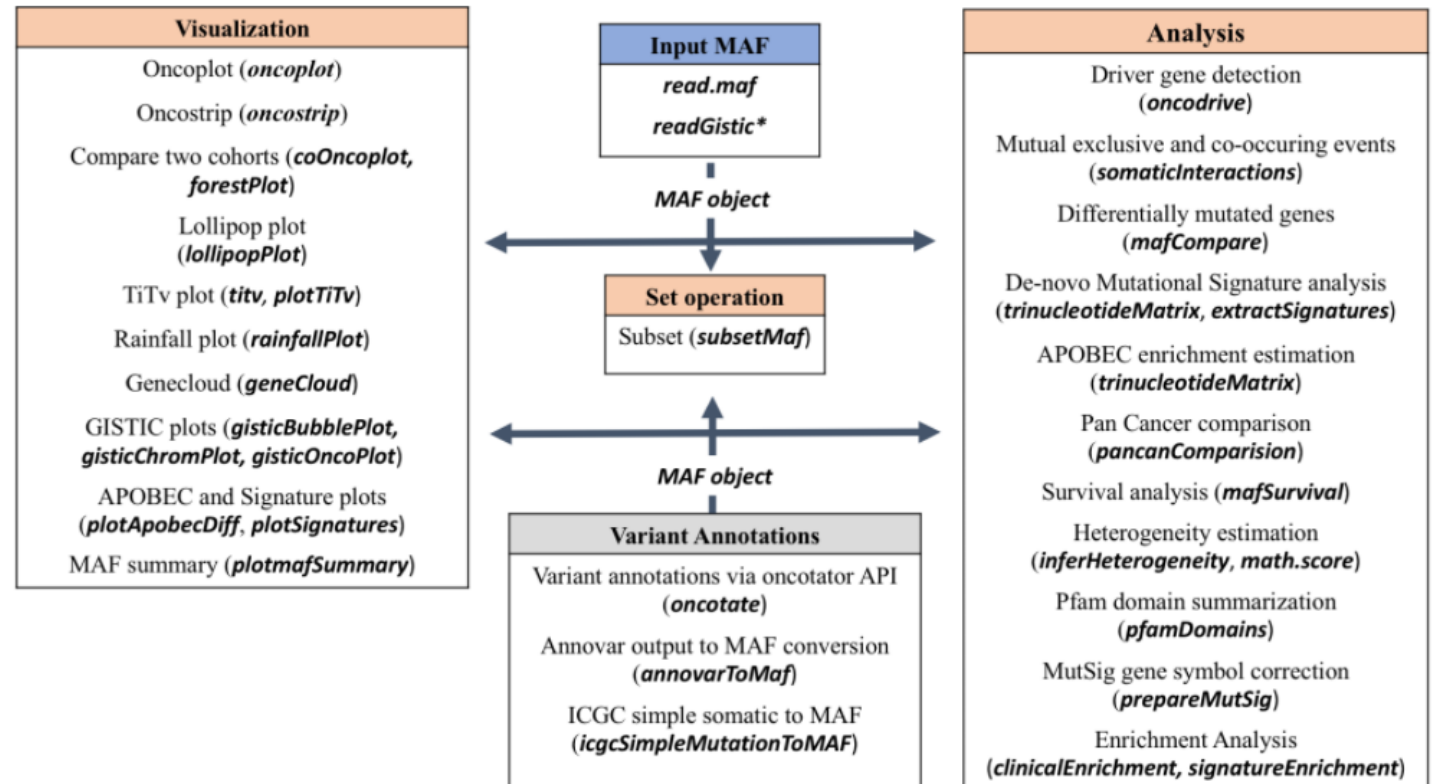
Let's see the script of R!

Regular Expression

- A sequence of **characters** that define a search **pattern**
 - Be useful to search the pattern of word in your data
- Example:
 - `re.findall('(\\d+):',str(df['Location']))`
 - `(\\d+):` means:
 - I want the **number of INTEGER or FLOAT (or more) before the pattern ‘:’** and the data is from the attribute of **Location** in the DataFrame called **df**
- Disadvantages:
 - Not easy to learn... very complicated!

maftools on Bioconductor of R

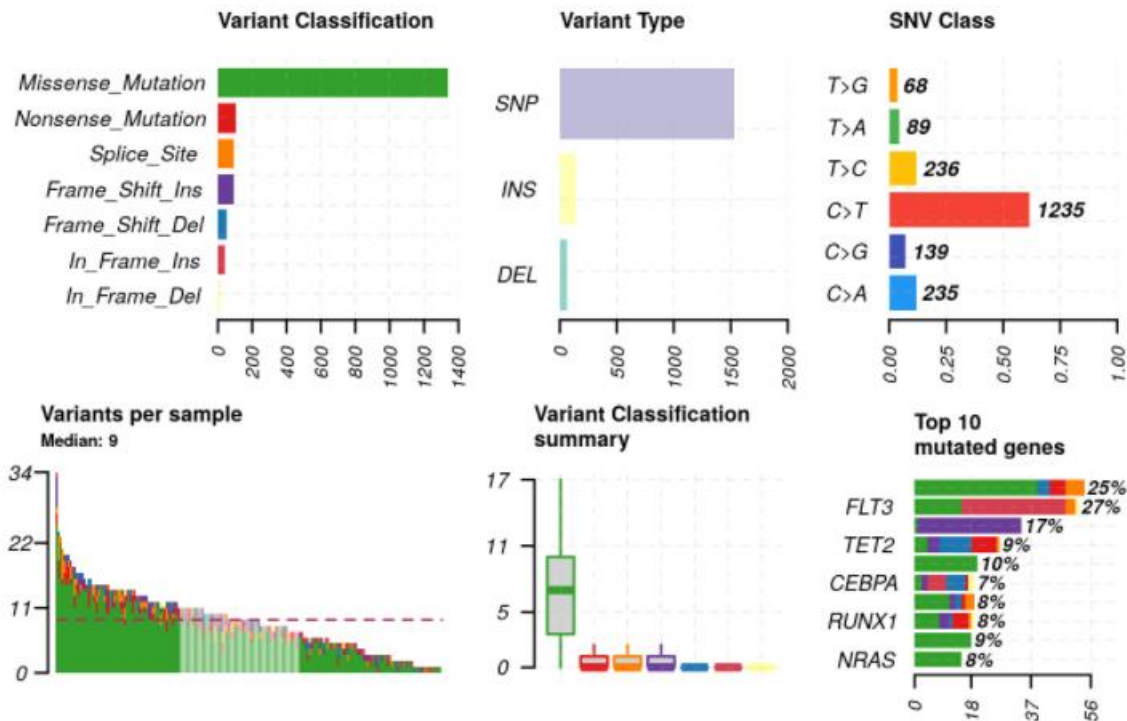
- The common visualization way to visualize the variants in your data
- The example of visualization
 - Oncoplot
 - Rainfall plot
 - Statistics plot (MAF summary)
 - Bar plot



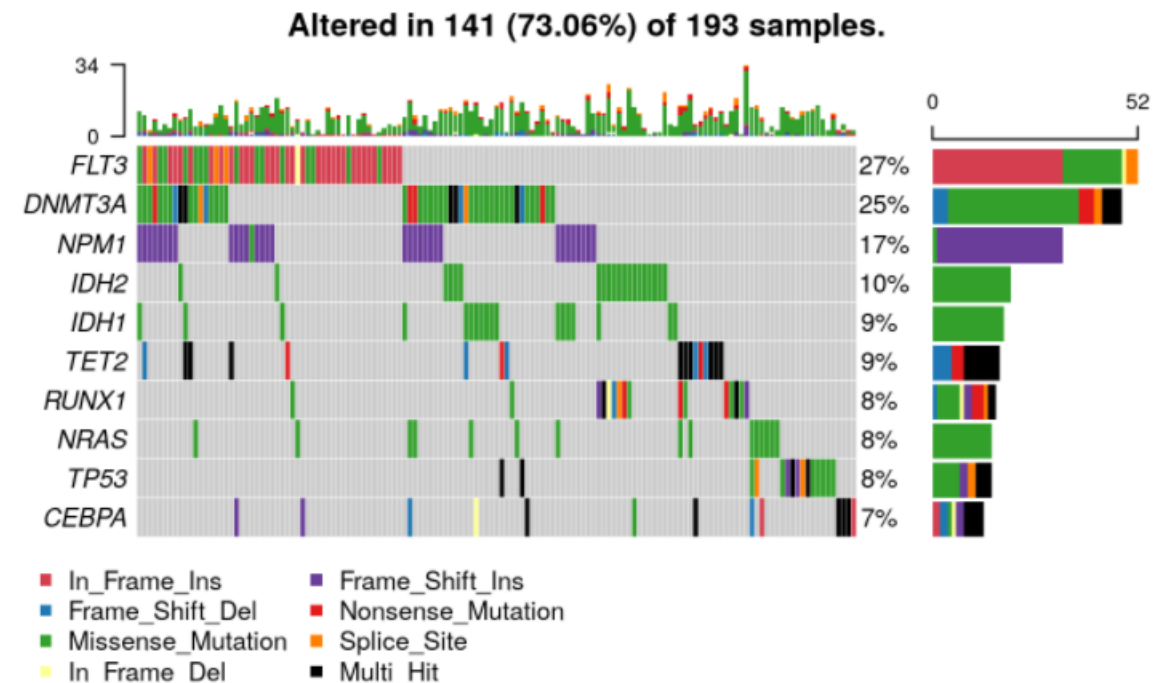
- <https://www.bioconductor.org/packages/release/bioc/vignettes/maftools/inst/doc/maftools.html>

The output is generated by maftools

MAF Summary



Oncoplot



Key Points on maftools

- No other defined words in Variant Classification section
- No **synonymous** variants in your OncoPlot
- If you have multiple-hit on the specific genes, it will be classified as “Multi-Hit”
- Samples are defined in the **Tumor Sample Barcode** attribute

- References:
 - <http://tardis.cgu.edu.tw/maftools.html>
 - <https://github.com/umccr/MAF-summary#maf-field-requirements>

Example small script on R

```
library(maftools)

df <- data.table::fread("D:/willischeng/Client/share/all_ABC_specific.maf")

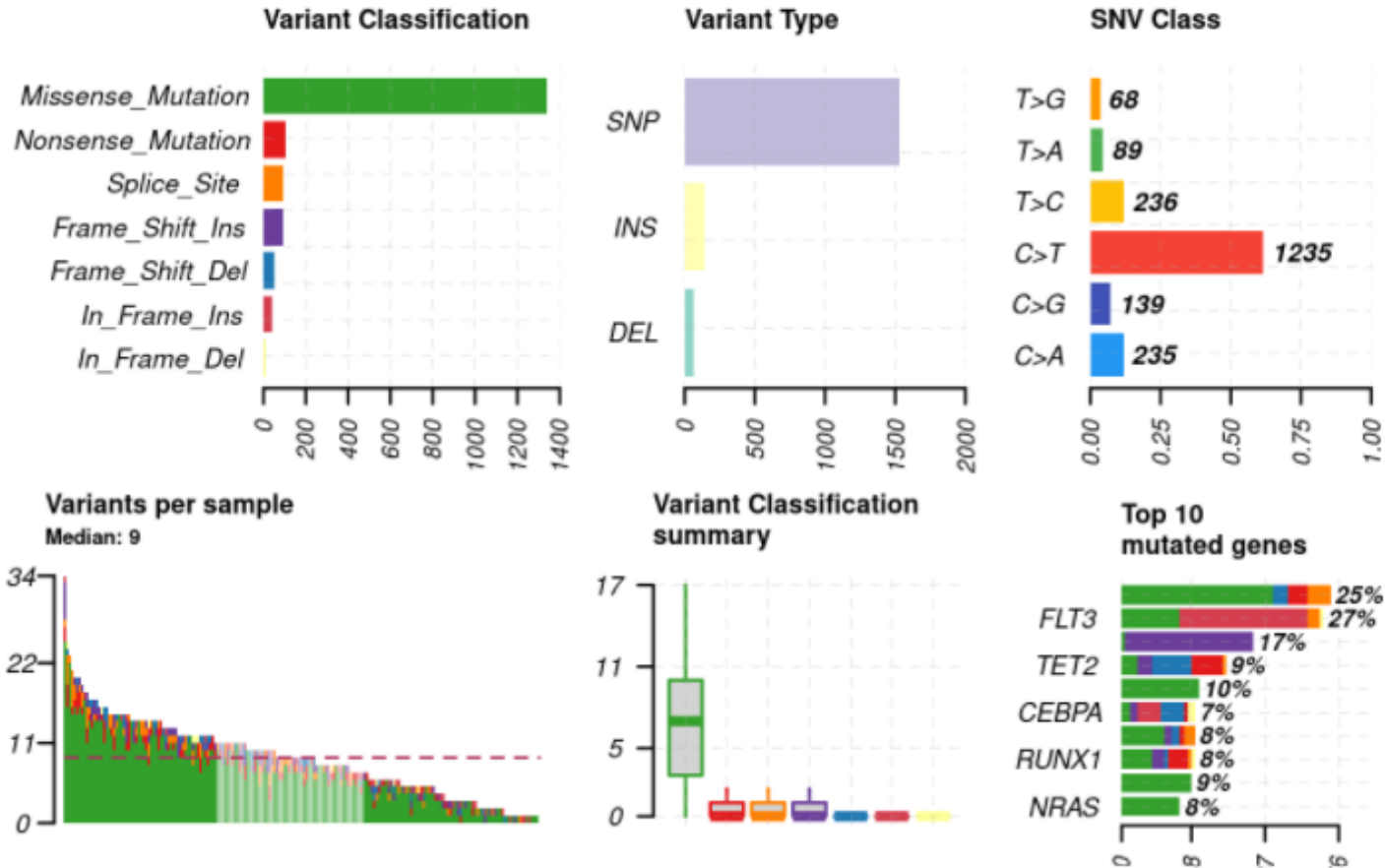
syn <- c("synonymous", NA, "", "-")
vc <- names(table(df$variant_classification))
nonSyn <- setdiff(vc, syn)
colors <- rainbow(length(nonSyn))
names(colors) <- nonSyn
ALL.maf = read.maf(maf = "D:/willischeng/Client/share/all_ABC_specific.maf", vc_nonSyn = nonSyn, verbose=TRUE)

plotmafsummary(ALL.maf, rmOutlier = TRUE, addStat = 'median', dashboard = TRUE, titvRaw = FALSE, color = colors)
#
vc_cols = RColorBrewer::brewer.pal(n = 8, name = 'Paired')
nonSyn[8]="Multi_Hit"
names(vc_cols) = nonSyn

oncoplot(maf = ALL.maf, colors=vc_cols, clinicalFeatures = "Tumor_Sample_Barcode", top=20, removeNonMutated = FALSE)
ALL.titv = titv(maf = ALL.maf, plot = FALSE, useSyn = TRUE, )
plotTitv(res = ALL.titv)
```

plotmafSummary

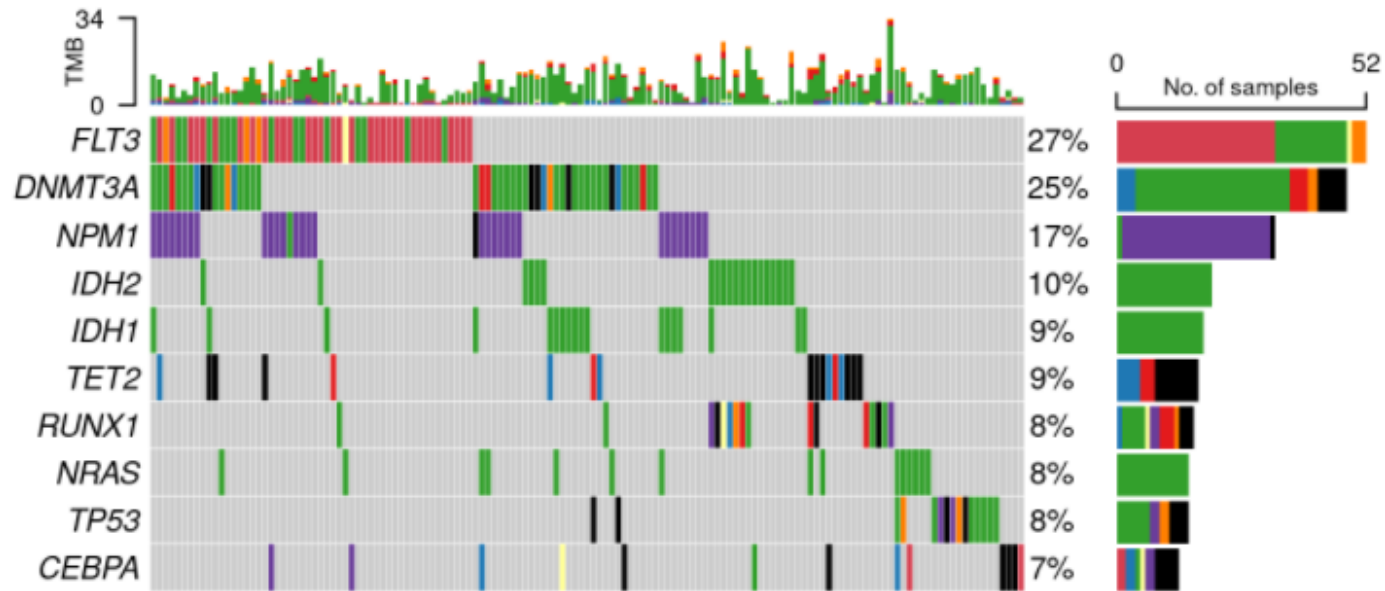
```
plotmafSummary(maf = lam1, rmOutlier = TRUE, addStat = 'median', dashboard = TRUE, titvRaw = FALSE)
```



Oncoplot

```
#oncoplot for top ten mutated genes.
oncoplot(maf = lam1, top = 10)
```

Altered in 141 (73.06%) of 193 samples.



- In_Frame_Ins
- Frame_Shift_Del
- Missense_Mutation
- In Frame Del
- Frame_Shift_Ins
- Nonsense_Mutation
- Splice_Site
- Multi Hit

TiTv plot

```
laml.titv = titv(maf = laml, plot = FALSE, useSyn = TRUE)
#plot titv summary
plotTiTv(res = laml.titv)
```

