Standard Operating Procedure of Genetic Variation Module

SOP 2.1 in iCMDB™

Version 1.0

16-Mar-2016

CONTENTS

TABLES

FIGURES

SUMMARY

The standard operating of procedure (SOP) of genetic variation module is summarized in Table 1.

Table 1 Summary of SOP in Genetic Variation (using "FLT3" in *Homo sapiens/human* as example)

Name	Description/Format	Examples
Gene Details		
Gene*	The name of basic functional unit of DNA sequence. Follow the official name provided by Entrez Gene in NCBI.	FLT3
Category*	Select based on the type of the variation. The categories include tandem repeats, complex region-based variation, CNV, protein overexpression, complex indel, MSI, SNP, methylation, fusion, inversion, translocation, amplification, overexpress, deletion, insertion, and point mutation. Refer to Table 2 for details.	POINT MUTATION
Accession*	NCBI reference sequence mRNA accession. Refer to SOP 1.9 for details.	NM_004119
Accession Version*	NCBI Reference Sequence mRNA accession version. Record the number after '.' in NCBI accession version (NM_XXXX.X).	2
Mutation Name*	Summarized description of mutation details.	FLT3 Internal Tandem Duplication
Mutation Details		
DNA Source*	Select according to the position where DNA is extracted from. DNA sources include tissue, plasma, and serum.	Tissue
Assembly*	Release version of genomic reference sequence. GRCh37 is preferred in iCMDB™.	GRCh37
Cds Mutation Syntax*	Syntax of coding DNA sequence (CDS) mutation. Refer to SOP 1.9 for procedures. Refer to HGVS for nomenclature details. ! For complex region-based variation, write "NA".	c.2503G>T

*Field are required for every entry.

Table 1 (continued)

Name	Description/Format	Examples
Mutation Details		
Amino acid substitution*	Protein-level amino acid sequence variation according to the mutation. Refer to SOP 1.9 for procedures. Refer to HGVS for format details. ! For complex region-based variation, write "NA".	p.D835Y
Chrom*	Chromosome of the gene. Possible values are 1~22, X and Y.	13
Location	Location on the gene where this mutation occurs. Location categories include exon (need to indicate exon number), intron (need to indicate intron number), promoter, 5'UTR, and 3'UTR. Refer to "Mutation Details" procedure 6 for details. ! If the mutations region is discontinuous, multiple continuous regions should be added. See detail in "mutation details".	Exon 20
Hg position*	Starting position of the mutation on genome. Use the number provided in genomic context. ! If the mutations region is discontinuous, multiple continuous regions should be added. See detail in "mutation details".	28592642
Hg position end*	Ending position of the mutation on genome. Use the number provided in genomic context. ! If the mutations region is discontinuous, multiple continuous regions should be added. See detail in "mutation details".	28592642
Primer	A strand of short nucleic acid sequences that serves as a starting point for DNA synthesis.	
Ref seq at pos*	Original CDS for this mutation in the sense (+) strand. Refer to Table 3 for details. ! If the gene is located on the antisense (-) strand, the corresponding CDS in the sense (+) strand will be the reverse complement of the sequence according to the base-paring rules (A pairs with T, C pairs with G). E.g. If the original sequence is "ATG", the reverse complement will be "CAT".	C

*Field are required for every entry.

Table 1 (continued)

Name	Description/Format	Examples
Mutation Details		
Alt seq at pos*	Alternative CDS for this mutation in the sense (+) strand. Refer to Table 3 for details. ! If the gene is located on the antisense (-) strand, the corresponding CDS in the sense (+) strand will be the reverse complement of the sequence according to the base-paring rules (A pairs with T, C pairs with G). E.g. If the original sequence is "ATG", the reverse complement will be "CAT".	A
RefSNP	RefSNP is an identifying number of a SNP recorded in NCBI. If the mutation doesn't have a corresponding RefSNP, leave it as blank.	rs121913488
Mutation Description*	Summarized description of mutation details. ! This only exists in complex region-based variation.	EGFR exon 19 deletion
Mechanism of variant*	Description about the variant. It should include the following information: 1. Location of the variant in the gene/protein. 2. Structural or functional importance of this location. 3. Functional effect of the mutation on protein structure, enzyme activity, downstream pathway or biological process. Refer to Table 6 for format and requirements for annotation.	*Location* This mutation is located in the highly conserved activation loop of FLT3, (to be continued) *Structural or functional importance* (Continued) which encodes part of the kinase domain (catalytic domain). *Functional effect* Mutations at D835 are an activating mutation and can result in a constitutive active kinase activity (FLT3, ERK1/2, AKT and STAT3) and constitutive signaling through downstream pathways leading to uncontrolled cellular proliferation and enhanced survival (Yamamoto et al., 2001; Fröhling et al., 2007; Tan et al., 2012).

*Field are required for every entry.

Table 1 (continued)

Name	Description/Format	Examples
Mutation Details		
Mechanism of variant (Chinese)*	Description of the mechanism of variant in Chinese. Follow the Chinese translation standards in Table 5. Refer to Table 6 for format and requirements for annotation.	该突变位于 FLT3 的高保守活化环内，该区域编码部分激酶结构域（催化结构域）。在 D835 发生的突变是活化突变，会通过下游通路导致激酶持续活化（FLT3, ERK1/2, AKT 和 STAT3）和信号持续传导，导致不受控制的细胞增殖并增加存活率（Yamamoto et al., 2001; Fröhling et al., 2007; Tan et al., 2012）。
Reference link*	References used in "Mechanism of variant". Only iCMDB™ reference ID is used and separated by comma without space.	7756,7757,7050
Annotation Details		
Annotation Category*	Select the category of the annotation. Annotation categories include biotherapy (subtypes include targeted therapy, immunotherapy, hormonotherapy), chemotherapy, radiotherapy, prognosis, pioneer study, diagnosis.	Biotherapy Targeted Therapy
Target Drug*	Select the drug described in the annotation. If the drug is not included in the database, add the drug first following SOP 1.2.	Crenolanib
Associated Disease*	Select the disease described in the annotation.	Acute Myeloid Leukemia
Drug Sensitivity	Categorize the effect of the drug based on the criteria in Table 4. The drug sensitivity categories include significantly resistant, partially resistant, partially sensitive, completely sensitive, highly toxic, and lowly toxic.	Completely Sensitive

*Field are required for every entry.

Table 1 (continued)

Name	Description/Format	Examples
Annotation Details		
Evidence Level	Select the highest level of evidence among all studies mentioned in the annotation. Evidence levels include randomized controlled trial, cohort study, case control study, case report, animal assay, and cell line study. Definition of each evidence level can be found in SOP 1.4 Table 3.	Animal Assay
Reference Link*	References used in "Mutation Description". Only iCMDB™ reference ID is used and separated by comma without space.	8475,8474,5472
Mutation Description*	1. For treatment with drug, describe the effect of this mutation on the above mentioned disease and drug. It should include the following information: a) Description of the drug and its targets. Please indicate the approval of FDA or recommendation by NCCN (if exist). b) Results from preclinical studies. Describe the effect of the drug on cells/animals harboring this mutation. c) Results from clinical studies (if exist). Describe the effect of the drug on patients with the above mentioned disease harboring this mutation. 2. For prognosis, describe the prognostic effect of this mutation on patients with a specific disease. It should include the following information:	*1. Treatment with drug* *Drug Information* Crenolanib (CP-868596) is a PDGFR inhibitor with potential antineoplastic activity. Crenolanib binds to and inhibits PDGFR, which may result in the inhibition of PDGFR-related signal transduction pathways, and therefore the inhibition of tumor angiogenesis and tumor cell proliferation (NCI Drug Dictionary). *Results from in vitro studies* In vitro studies showed that crenolanib was significantly more potent than quizartinib (Galanis et al., 2014) and sorafenib (Zimmerman et al., 2013) against the D835Y mutant cells (crenolanib vs quizartinib: D835Y IC50=6.9 vs 33.7 nM; ITD/D835Y IC50=8.7 vs 93.1 nM; kinase binding for crenolanib vs sorafenib: D835Y Kd=0.14 vs 24 nM).

*Field are required for every entry.

Table 1 (continued)

Name	Description/Format	Examples
Annotation Details		
Mutation Description* **(continued)**	(continued) a) Prognosis impact. Describe the prognosis impact of the mutation on patients with a specific disease. b) Results from clinical studies. Describe the clinical results with supporting data. Refer to Table 6 for format and requirements for annotation.	(continued) *Results from in vivo studies* An in vivo study showed that mice engrafted with FLT3-ITD/D835Y cells had a significant increased survival compared with vehicle-treated mice when treated with crenolanib (median survival=27, 23, and 20 days for crenolanib twice daily, crenolanib once daily, and vehicle treatment groups, respectively; P<0.0001). In contrast, sorafenib treatment produced no survival advantage over vehicle-treated controls (median survival=20 days) (Zimmerman et al., 2013). *2. Prognosis (using NPM1 as an example)* *Prognosis impact* NPM1 mutations have favorable prognostic impact in older patients with cytogenetically normal acute myeloid leukemia, especially those age ≥70 years. *Results from clinical studies* It has been reported that patients with NPM1 mutations (56%) had higher complete remission (CR) rates (84% vs 48%; P<0.001) and longer disease-free survival (DFS; P=0.047; 3-year rates, 23% vs 10%) and overall survival (OS; P<0.001; 3-year rates, 35% vs 8%) than NPM1 wild-type patients (Heiko et al., 2009). In a systematic review and meta-analysis, cytogenetically normal AML with mutated NPM1 without FLT3-ITD is associated with a favorable outcome in young patients (Port et al., 2014; Döhner et al., 2015).

*Field are required for every entry.

Table 1 (continued)

Name	Description/Format	Examples
Annotation Details		
Mutation Description (Chinese) *	Mutation description in Chinese. Refer to Table 5 for Chinese translation of common terminology. Refer to Table 6 for format and requirements for annotation.	*1. Treatment with drug* Crenolanib（CP-868596）是一种 PDGFR 抑制剂，具有潜在的抗肿瘤活性。Crenolanib 结合并抑制 PDGFR，这可能导致 PDGFR 相关的信号转导途径被抑制，因此导致抑制肿瘤血管生成和肿瘤细胞增殖（NCI Drug Dictionary）。体外研究表明，crenolanib 对 D835Y 突变的细胞比 quizartinib（Galanis et al., 2014）和索拉非尼（Zimmerman et al., 2013）显著更加有效（crenolanib vs quizartinib：D835Y IC50=6.9 vs 33.7 nM；ITD/D835Y IC50=8.7 vs 93.1 nM；crenolanib vs 索拉非尼激酶结合：D835Y Kd=0.14 vs 24 nM）。体内研究表明，crenolanib 治疗的移植了 FLT3-ITD/D835Y 细胞的小鼠与载体对照的小鼠相比生存期显著增加（每天两次 crenolanib 给药，每天一次 crenolanib 给药，和载体对照组的中值存活期分别为：27，23，20 天；P<0.0001）。与此相反，索拉非尼治疗与载体对照相比没有产生存活优势（中值存活期=20 天）（Zimmerman et al., 2013）。 *2. Prognosis (using NPM1 as an example)* 在细胞基因型正常的急性髓系白血病患者中，尤其是年龄高于或等于 70 岁的患者中，携带 NPM1 突变的患者往往有良好的预后效果。有报道称 NPM1 突变携带病人（56%）与无突变的病人相比，有更高的完全缓解率（84% vs 48%；P<0.001），更长的无疾病生存期（DFS；P=0.047；3 年缓解率，23% vs 10%），和更长的总生存期（OS；P<0.001；3 年缓解率，35% vs 8%）。在系统回顾整合分析中，细胞基因型正常的急性白血病 NPM1 突变携带者且无 FLT3-ITD 的年轻病人和良好的预后相关（Port et al., 2014; Döhner et al., 2015）。

*Field are required for every entry.

Figure 1 Example of Genetic Variation entry

Figure 1.1 Gene Details

Figure 1.2 Mutation Details

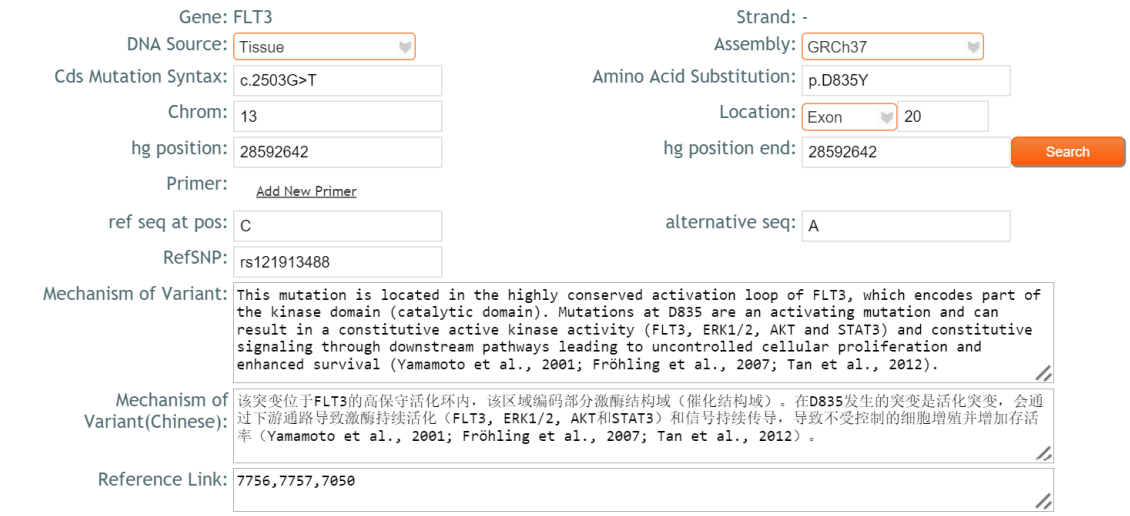


Figure 1.3 Annotation Details

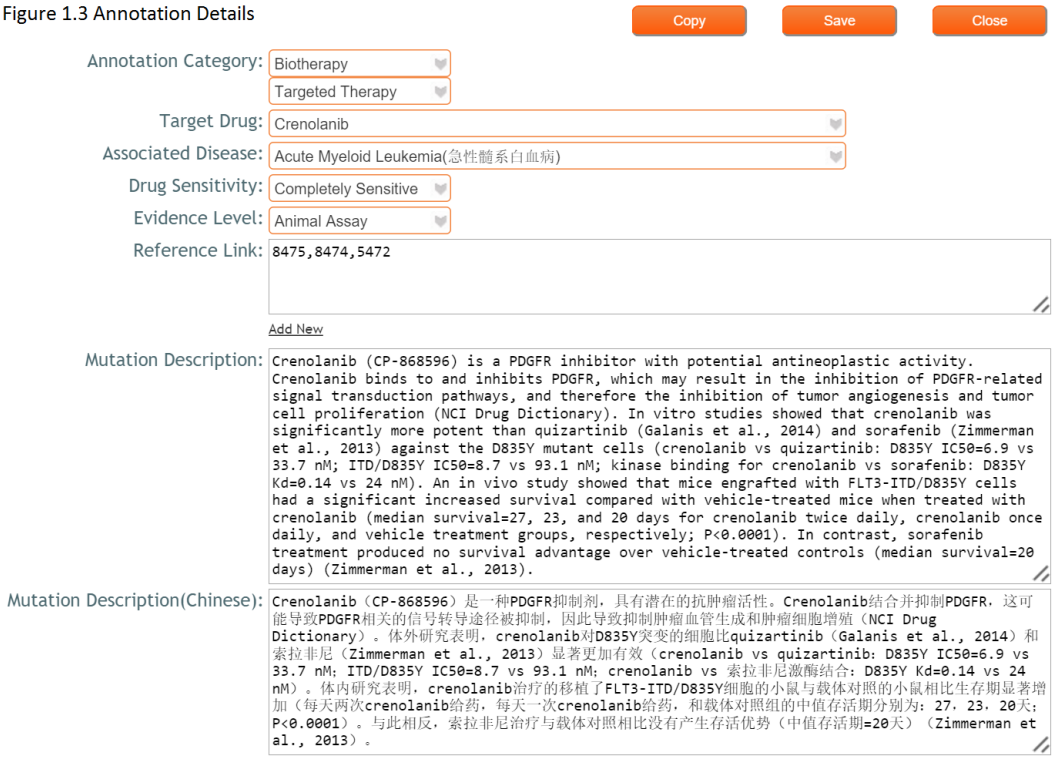
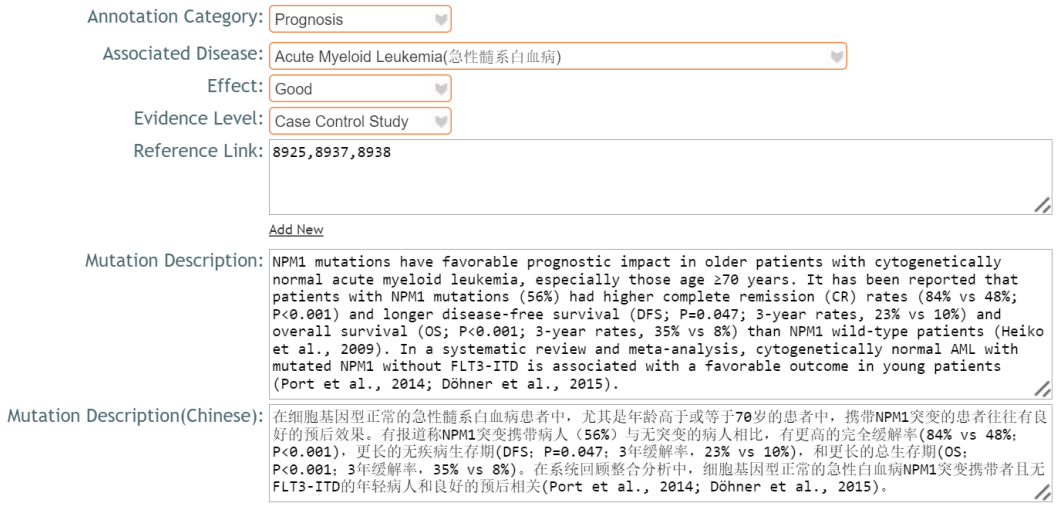


Figure 1 (continued)

Figure 1.3 Annotation Details



INTRODUCTION AND PURPOSE

SOP 2.1is a standard operation procedure to guarantee the accuracy and conciseness of data for "Genetic Variation" module in iCMDB™. It mainly uses NCBI, NCCN, FDA, ClinicalTrials, OMIM, COSMIC, Ensembl, PharmGKB as reference.

The Purpose of SOP 2.1 is to provide detailed guidelines and standards for data search and data entry of "Genetic Variation" in "Oncology" module.

RESPONSIBILITIES

Whoever creates the record is responsible for the accuracy of the data in "Genetic Variation" module. Review process is conducted weekly using incorporated check flag function and review records are kept for tracing. The procedures of SOP preparation and update follows SOP 10.2.

SPECIFIC PROCEDURE

Gene Details

1. Go to "Genetic Variation" module in iCMDB™ backend database.
2. Choose the corresponding "Gene Symbol" and click "Search" button in the "Search for Biomarker Mutation" page. If cannot find the correspond gene in the "Gene Symbol" list, add the gene into our database first following SOP1.3.
3. If the corresponding mutation category is not listed in the search result, click "Add New" hyperlink near "Search" button. A "Mutation Details" page will pop-up.
4. Record "Gene". Value in this variable is the same as "Gene Symbol" in the "Search for Biomarker Mutation" page.
5. Select "Category" following the criteria in Table 2.
6. Record "Accession" and "Accession Version". Refer to SOP 1.9 for details.
7. Record "Mutation Name". It will be filled automatically if "Gene" and "Category" are selected. If the category is "Complex Region-based Variation", should give a mutation description rather than using the auto-generated one (e.g. EGFR exon 20 insertion).
8. Click "Save" button.

Mutation Details

1. Click "Add New Biomarker Mutation" hyperlink. A new "Mutation Details" page will pop-up.
2. Select "DNA Source" and "Assembly" accordingly. GRCh37 is preferred in iCMDB™.
3. Record "Cds Mutation Syntax". This information can be found in papers or COSMIC (see Figure 2 as an example). The nomenclature of mutation follows HGVS standards. Things to notice:
   a) Duplicating insertions are described as duplications, not as insertions. e.g. "ACTTTGTGCC" to "ACTTTGTG**G**CC" is described as "c.8dupG", not as "c.8_9insG".
   b) If the sequence has more than or equal to 9 nucleotides, use the length of the sequence to represent the string of nucleotides. e.g. "c.2307_2308insATGGCCAGCGTGGAC" is written as "c.2307_2308ins15"; "c.2235_2246delGGAATTAAGAGA" is written as "c.2235_2246del12".
   c) Complex indels are described using "delins". The suggestion to use ">" to indicate "delins" in frame shift descriptions has been retracted. e.g. "c. 112_117delinsTG" should not be written as "c. 112_117AGGTCA>TG".

4. Record "Amino Acid Substitution". This information can be found in papers or COSMIC (see Figure 2 as an example). The nomenclature of mutation follows HGVS standards. Things to notice:
   a) If the sequence has more than or equal to 9 amino acids, use the length of the sequence to represent the string of amino acids. e.g. "p.W182_Q183insINGNNYVYIDPTQL" is written as "p.W182_Q183ins14".
   b) Complex indels are described using "delins". The suggestion to use ">" to indicate "delins" in frame shift descriptions has been retracted. e.g. "p.C28_K29delinsW" should not be written as "p.C28_K29>W".
5. Record "Chrom", "hg position", "hg position end". This information can be found in COSMIC. See Figure 2 as an example.

If the mutations region is discontinuous, multiple continuous regions (containing "location" "start position" and "end position") should be added. In this case, "hg position" = the min start position, "hg position end" = the max end position (see Figure 3 as an example).

6. Record "Location". This information can be found in papers or database such as ensembl. If the mutation is located in intron/exon, please indicate the number of the intron/exon. Procedures for finding intron/exon position in ensembl database are as follows (See Figure 4 as an example):
   a) Go to ensembl database and search for the gene (e.g. EGFR).
   b) Look for the correct accession number (same with the one in iCMDB™) in "RefSeq" and click the corresponding "Transcript ID" hyperlink.
   c) Click on "Exons" in the filters on the left hand side of the page. The hg position of the mutation should fall in the "Start" and "End" region of the corresponding intron/exon.
7. Record "ref seq at pos" and "alternative seq". Refer to Table 3 for details.

If the gene is located on the antisense (-) strand, use the reverse complement of the sequence (e.g. the reverse complement of "ATG" is "CAT").

8. Record "RefSNP" if the mutation has one. This information can be found in papers or databases such as ClinVar and dbSNP.
9. Describe "Mechanism of Variant" in 3 aspects:
   a) Location of the variant in the gene/protein.
      Describe the location of the mutation in a sentence. E.g. "This mutation is located in the XXX region of XX".
   b) Structural or functional importance of this location.
      Describe the structure or function of the mutated position in a sentence. E.g. "XXX stabilize/activate/induce/initiate XXX, leading to XXX".
   c) Functional effect of the mutation.
      Describe the effect of the mutation on gene/protein function in a sentence. E.g. "XXX result in/lead to XXX".

Format and requirements for annotation are given in Table 6.

10. Record "Mechanism of Variant (Chinese)" and the corresponding "Reference Link" iCMDB™ ID.

Chinese translations of common terminologies are given in Table 5.

Annotation Details

1. To add an annotation (prognosis/drug), click the "Add New Drug" hyperlink.
2. Select the correct "Annotation Category". Definition of all categories are as follows:
   a) Biotherapy: A type of treatment that uses substances made from living organisms to treat disease. These substances may occur naturally in the body or may be made in the laboratory. Subtypes includes:
      i. Targeted Therapy: A type of treatment that uses drugs or other substances to identify and attack specific types of cancer cells with less harm to normal cells. Targeted therapy blocks the action of certain enzymes, proteins, or other molecules involved in the growth and spread of cancer cells.
      ii. Immunotherapy: A type of biological therapy that uses substances to stimulate or suppress the immune system to help the body fight cancer, infection, and other diseases.
      iii. Hormonotherapy: Treatment with hormones or hormone antagonists.
   b) Chemotherapy: A type of treatment that uses chemical substances, especially one or more anti-cancer drugs (chemotherapeutic agents) that are given as part of a standardized chemotherapy regimen.
   c) Radiotherapy: The use of high-energy radiation from x-rays, gamma rays, neutrons, protons, and other sources to kill cancer cells and shrink tumors.
   d) Prognosis: The likely outcome or course of a disease; the chance of recovery or recurrence.
   e) Pioneer Study: A study that is the first or among the earliest in the field of inquiry or progress. A study that begins something new or takes part in the early development of something.
   f) Diagnosis: The process of identifying a disease, condition, or injury from its signs and symptoms.
3. Select "Target Drug". If the drug is not existed in the database, add the drug first according to SOP1.2.
4. Select "Associated Disease". If the disease is not existed in the database, add the disease first according to SOP1.1.
5. Select the correct "Drug Sensitivity" following the criteria given in Table 4.
6. Select the highest "Evidence Level" among all studies mentioned in the annotation. Definition of each evidence level can be found in SOP 1.4 Table 3.
7. Describe "Mutation Description".
   a) For treatment with a drug, describe "Mutation Description" in the following 3 aspects:
      i. Drug information.
      Describe the drug and its targets. Mention in the text if it's approved by FDA or recommended in NCCN. For drugs that are still under investigation and not approved by FDA yet, drug information can be found in NCI drug dictionary or papers. E.g. "[drug XXX] is a [gene/protein XXX] inhibitor. It was approved by FDA in [year XXXX]".
      ii. Results from preclinical studies.
      Describe the effect of the drug on cells/animals harboring this mutation. E.g. In vivo/vitro studies showed that, [drug XXX] significantly inhibited/prolonged [cell/animal XXX] growth/survival with IC50/OS=XXX".

iii. Results from clinical studies.
Describe the effect of the drug on patients with the above mentioned disease harboring this mutation. E.g. In a clinical/cohort study, [disease] patients with [mutation details] treated with [drug] had a significantly increased/decreased survival/blast count ([data from result])".

b) For prognosis, describe "Mutation Description" in the following 2 aspects:

i. Prognosis impact.
Describe the prognosis impact of the mutation on patients with a specific disease. E.g. "[gene XXX mutation] has a significant favorable/poor prognostic impact in patients with [disease]".

ii. Results from clinical studies.
Describe the clinical results with supporting data. E.g. "It has been reported that patients with [gene XXX mutation] had better/worse survival/remission than [control group] ([data from result])".

Format and requirements for annotation are given in Table 6.

8. Record "Mutation Description (Chinese)" and the corresponding "Reference Link" iCMDB™ ID.

Chinese translations of common terminologies are given in Table 5.

9. Click "Save" button.

Table 2 Type and description of genetic variation category

Type	Description
Tandem Repeats	Tandem repeats occur in DNA when a pattern of one or more nucleotides is repeated and the repetitions are directly adjacent to each other.
Complex Region-based Variation	Mutation in certain region of the gene.
CNV	A copy number variation (CNV) is when the number of copies of a particular gene varies from one individual to the next.
Protein Overexpression	Make too many copies of a protein. Overexpression of certain proteins or other substances may play a role in cancer development.
Complex Indel	A mutation resulting in both an insertion of nucleotides and a deletion of nucleotides which results in a net change in the total number of nucleotides, where both changes are nearby on the DNA.
MSI	Microsatellite instability (MSI) is a hyper-mutable phenotype caused by the loss of DNA mismatch repair activity.
SNP	A single nucleotide polymorphism (SNP) is a variation at a single position in a DNA sequence among individuals.
Methylation	A chemical reaction in which a small molecule called a methyl group is added to other molecules. Methylation of proteins or nucleic acids may affect how they act in the body.
Fusion	A gene made by joining parts of two different genes.
Inversion	An inversion is a chromosome rearrangement in which a segment of a chromosome is reversed end to end. An inversion occurs when a single chromosome undergoes breakage and rearrangement within itself.
Translocation	A genetic change in which a piece of one chromosome breaks off and attaches to another chromosome. Sometimes pieces from two different chromosomes will trade places with each other. Translocations may lead to medical problems such as leukemia, breast cancer, schizophrenia, muscular dystrophy, and Down syndrome.
Amplification	A selective increase in the number of copies of a gene coding for a specific protein without a proportional increase in other genes.
Overexpress	Excessive expression of a gene by producing too much of its effect or product.
Deletion	Mutation in which a section of DNA is lost, or deleted.
Insertion	Mutation in which one or more base pairs is added to a DNA sequence.
Point Mutation	Mutation that causes a single nucleotide base substitution.

Table 3 Format and example for reference sequence and alternative sequence

Mutation Category	Mutation example	Format		Example	
		ref seq at pos	alternative seq	ref seq at pos	alternative seq
Point Mutation	EGFR c.2155G>A	[original sequence]	[alternative sequence]	G	A
Insertion	EGFR c.2310_2311insGGT	.	[inserted nucleotides]	.	GGT
Deletion	FLT3* c.2503_2505delGAT	[deleted nucleotides]	.	ATC	.
Complex Indel	FLT3* c.2503_2504GA>TT	[original sequence]	[alternative sequence]	TC	AA
SNP	EGFR (+) rs1050171 c.2361G>A	[original sequence]	[alternative sequence]	G	A
Tandem Repeats	FLT3* ITD	.	NN	.	NN
Complex Region-Based Variation	Point Mutation	NN	NN	NN	NN
	Insertion	.	NN	.	NN
	Deletion	NN	.	NN	.
	Mutation	NA	NA	NA	NA
CNV	TBD	TBD	TBD	TBD	TBD
Protein Overexpression	TBD	TBD	TBD	TBD	TBD
MSI	TBD	TBD	TBD	TBD	TBD
Methylation	TBD	TBD	TBD	TBD	TBD
Fusion	TBD	TBD	TBD	TBD	TBD
Inversion	TBD	TBD	TBD	TBD	TBD
Translocation	TBD	TBD	TBD	TBD	TBD
Amplification	TBD	TBD	TBD	TBD	TBD
Overexpress	TBD	TBD	TBD	TBD	TBD

*Gene is on the antisense (-) strand. Need to use the reverse complement of the original sequence.
TBD: To be determined and discussed.

Table 4 Type and description of drug sensitivity

Type	Description
Significantly Resistant	It is mentioned in the text (e.g. highly resistant, ineffective) or mutation carriers (cell lines, animals, or patients) showed no response to the drug. E.g. Growth of cells with this mutation could not be inhibited by the drug; animals or patients with this mutation showed no response to the drug.
Partially Resistant	It is mentioned in the text (e.g. mild resistant) or mutation carriers (cell lines, animals, or patients) have decreased sensitivity to the drug. E.g. IC50 for cells with this mutation was much higher than the control group; animals or patients with this mutation showed some response to the drug at first but then became refractory.
Partially Sensitive	It is mentioned in the text (e.g. partial response, moderate activity) or mutation carriers (cell lines, animals, or patients) have decreased resistance to the drug. E.g. IC50 for cells with this mutation was higher than the control group, but still at a relative low level; animals or patients with this mutation showed some response to the drug but did not reach statistical significance.
Completely Sensitive	It is mentioned in the text (e.g. highly effective) or mutation carriers (cell lines, animals, or patients) showed significant response to the drug. E.g. Growth of cells with this mutation was inhibited by the drug at very low concentration; animals or patients with this mutation showed a significantly prolonged survival.
Highly Toxic	Patients taking this drug have serious adverse reactions. The drug is toxic at low concentration.
Lowly Toxic	Patients taking this drug have mild adverse reactions. The drug is toxic at high concentration.

Table 5 Chinese Translation of Common Terminology

English	Abbreviation	Chinese
Overall survival	OS	总生存期
Disease-free survival	DFS	无疾病生存期
Event-free survival	EFS	无事件生存期
Progression-free survival	PFS	无进展生存期
Confidence interval	CI	置信区间
Complete remission	CR	完全缓解
Partial remission	PR	部分缓解
Hazard ration	HR	风险比
Median XXX (e.g. median survival)		中位 XXX(例如，中位生存期)

Table 6 Format and Requirements for Annotation

Scenario	Format	Requirement	Example
Gene symbol	Use the official symbol	The gene symbol should be consistent throughout the annotation. For some papers, the symbol of the gene is written as its synonyms (e.g. TP53 written as p53, CSF1R written as FMS). Remember to change it into the official gene symbol when using these papers.	TP53 (do not use a mixture of TP53 and p53) CSF1R (do not use a mixture of CSF1R and FMS)
P value	P=0.XXX P<0.XXX P>0.XXX	1. Write "P" in capital letter. 2. Do not leave any space before or after "=", "<", or ">". 3. There should be a "0" before the decimal point.	P=0.12 P<0.12 P>0.12
"=", "<", ">"	[parameter]=XX [parameter]<XX [Parameter]>XX	Do not leave any space before or after "=", "<" or ">".	HR=1.2 95% CI=1.2-3.4 …in older patients (>60 years)
Unit	XX [unit]	1. Leave a space between the number and the unit. 2. For units of time, do not use the abbreviation. I.e. use "days", "months", and "years" instead of "d", "m", "y". 3. For percentage, do not leave any space between the number and "%".	IC50=1.2 nM OS=3.4 days CR=56%
Interval	XX-XX	1. Use "-" rather than "~". 2. Do not leave any space before or after "-". 3. If the number has a unit, write the unit after the second number only. 4. For interval of percentage, write "%" after the second number only.	IC50=1.2-3.4 nM CI=1.2-3.4%

Table 6 (continued)

Scenario	Format	Requirement	Example
Comparison	XX vs XX	1. Use "vs" rather than "vs.". 2. Leave a space before and after "vs". 3. If the number has a unit, write the unit after the second number only. 4. For comparison between percentages, write "%" after both numbers.	IC50=1.2 vs 3.4 nM CR=12% vs 34%
Number of patients	N=XX	1. Write "N" in capital letter. 2. Do not leave any space before or after "=".	N=12
Chinese translation		1. Do not use a mixture of Chinese and English words (except for abbreviations and proper names). 2. Use the Chinese punctuations (comma"，", full stop"。", brackets "（", "）", etc.) and do not leave any space before or after these symbols in the main text. 3. For reference citation, follow the same format as in English annotation. 4. Use the generic name of the drug.	TET2 基因突变对于骨髓增生异常综合征（MDS）病人是一个有利的预后因子（HR=5.2，P=0.005）（Kosmider et al., 2009）。

Figure 2 Example of Mutation Position

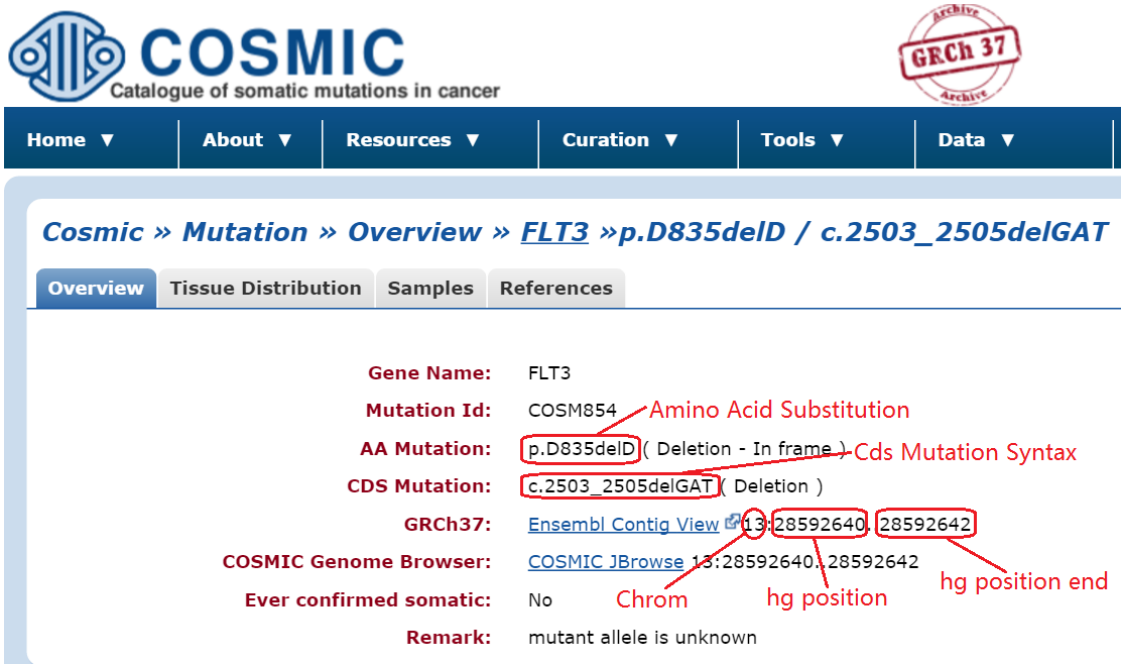


Figure 3 Example of Discontinuous Region

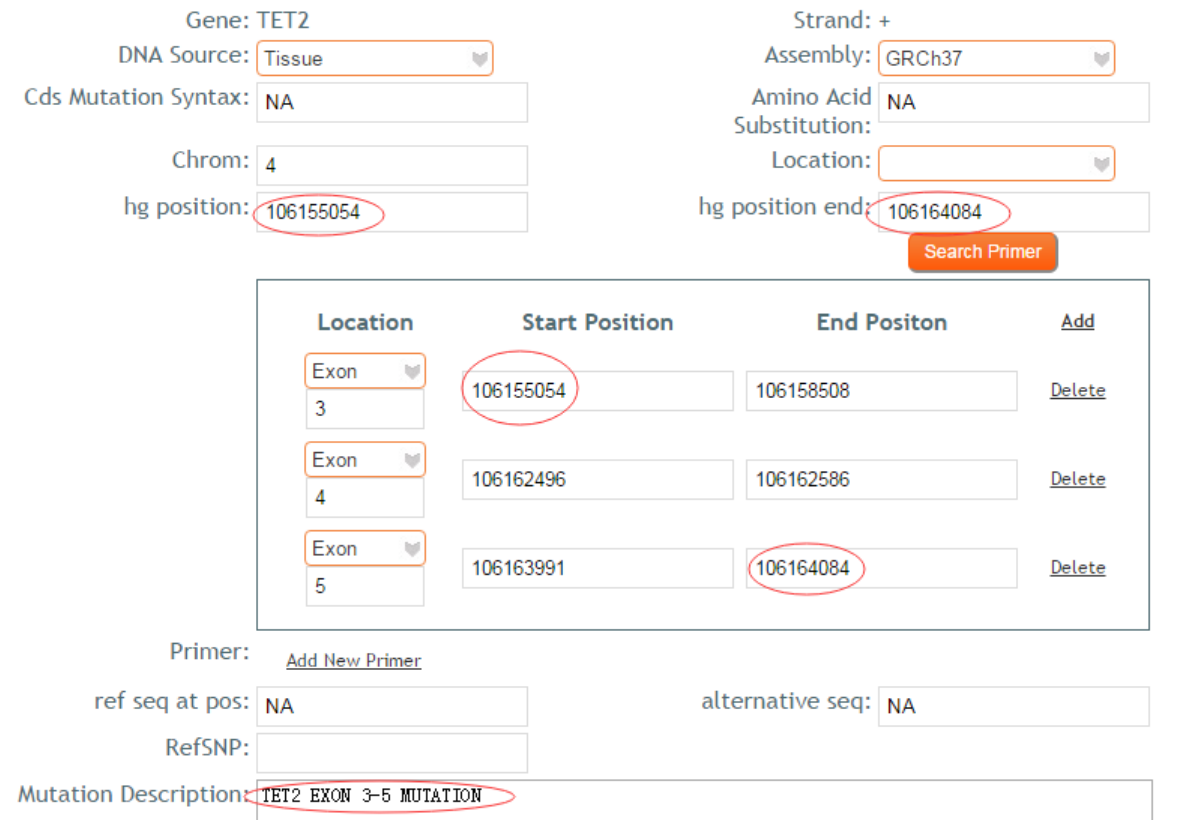


Figure 4 Example of Intron/Exon Position

Figure 4.1 Click the "Transcript ID" hyperlink corresponding to the correct accession number in "RefSeq"

Figure 4.2 Click "Exons"

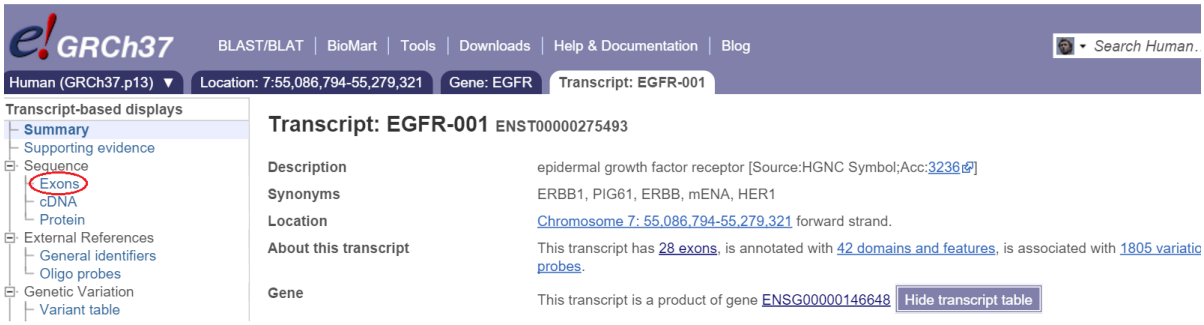
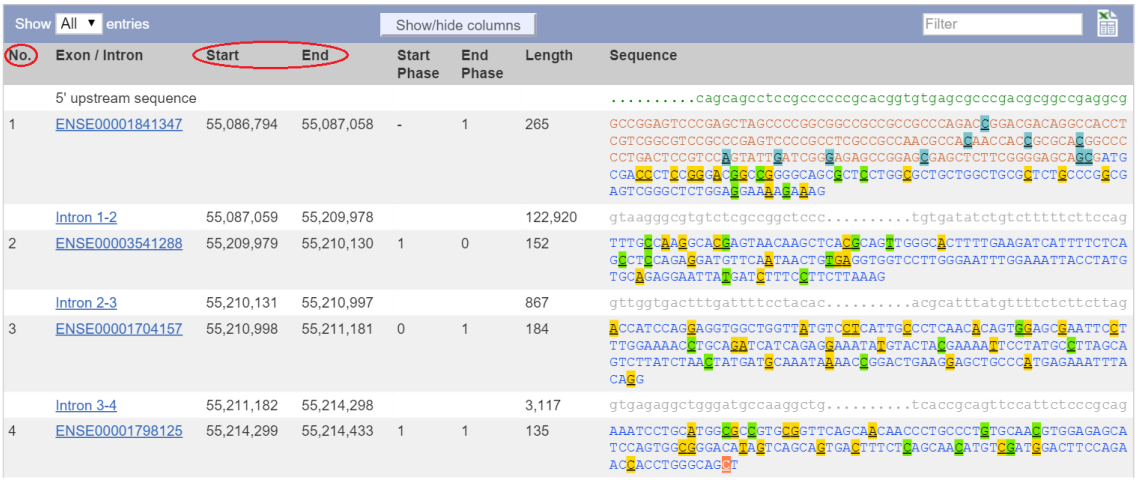


Figure 4.3 The hg position of the mutation should fall in the region of the corresponding intron/extron



REFERENCE

Internal References

SOP No.	Version	Description
1.3	1.0	Biomarker
1.4	1.0	Reference
1.8	1.0	Reference Selection
1.9	0.1	Determination of cds syntax and amino acid syntax

External References

No.	Resource	Description
1	Entrez Gene	Entrez Gene (http://www.ncbi.nlm.nih.gov/gene) is National Center for Biotechnology Information (NCBI)'s database for gene-specific information. Entrez Gene maintains records from genomes which have been completely sequenced, which have an active research community to submit gene-specific information, or which are scheduled for intense sequence analysis. Records in Entrez Gene are assigned unique, stable and tracked integers as identifiers (PMCID: PMC3013746).
2	NCCN	National Comprehensive Cancer Network (NCCN, http://www.nccn.org/) is an alliance of the world's leading cancer centers. It provides authoritative source of comprehensive cancer care.
3	FDA	Food and Drug Administration (FDA, http://www.fda.gov/) is a federal agency of U.S. It is responsible for protecting the public health by assuring the safety, efficacy and security of human and veterinary drugs, biological products, medical devices, our nation's food supply, cosmetics, and products that emit radiation. This website updates all related information released by U.S government.
4	ClinicalTrials	ClinicalTrials (https://www.clinicaltrials.gov/) is a registry and results database of publicly and privately supported clinical studies of human participants conducted around the world.
5	OMIM	Online Mendelian Inheritance in Man (OMIM, http://www.ncbi.nlm.nih.gov/omim) is a National Center for Biotechnology Information (NCBI)'s database. OMIM is a continuously updated catalog of human genes and genetic disorders and traits, with a particular focus on the gene-phenotype relationship.
6	COSMIC	Catalogue Of Somatic Mutations In Cancer (COSMIC, http://grch37-cancer.sanger.ac.uk/cosmic) is an online database of somatically acquired mutations found in human cancer. Somatic mutations are those that occur in non-germline cells that are not inherited by children.
7	Ensembl	Ensembl (http://grch37.ensembl.org/index.html) is a joint project between EMBL-EBI and the Wellcome Trust Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.

External References (continued)

No.	Resource	Description
8	PharmGKB	PharmGKB (https://www.pharmgkb.org/) is a comprehensive resource that curates knowledge about the impact of genetic variation on drug response for clinicians and researchers.
9	NCI Drug Dictionary	The NCI Drug Dictionary (http://www.cancer.gov/publications/dictionaries/cancer-drug) contains technical definitions and synonyms for drugs/agents used to treat patients with cancer or conditions related to cancer. Each drug entry includes links to check for clinical trials listed in NCI's List of Cancer Clinical Trials.
10	ClinVar	ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) is a National Center for Biotechnology Information (NCBI)'s database. ClinVar aggregates information about genomic variation and its relationship to human health.
11	dbSNP	dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) is a is a National Center for Biotechnology Information (NCBI)'s database. dbSNP is a database of single nucleotide polymorphisms (SNPs) and multiple small-scale variations that include insertions/deletions, microsatellites, and non-polymorphic variants.
12	Nucleotide	Nucleotide (http://www.ncbi.nlm.nih.gov/nuccore) is a National Center for Biotechnology Information (NCBI)'s database. The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.
13	HGVS	Human Genome Variation Society (HGVS, http://www.hgvs.org/) is a society aims to foster discovery and characterization of genomic variations including population distribution and phenotypic associations. Promote collection, documentation and free distribution of genomic variation information and associated clinical variations. Endeavor to foster the development of the necessary methodology and informatics.
14	Chinese Clinical Trial Registry, ChiCTR	Chinese Clinical Trial Register (ChiCTR) is hosted on Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University.The Chinese Clinical Trial Register is a non-profit organisation, is established according to both the WHO International Clinical Trials Register Platform Standard and Ottawa Group Standard.Chinese Clinical Trial Register provides the services include register for trials, consultation for trial design, central randomization for an allocation sequence, peer review for draft articles and training for peer reviewers

CHANGE HISTORY

Revision	Date	Significant Changes
0.1	11 Dec. 2015	Initial version
1.0	16 Mar 2016	First release to department of database