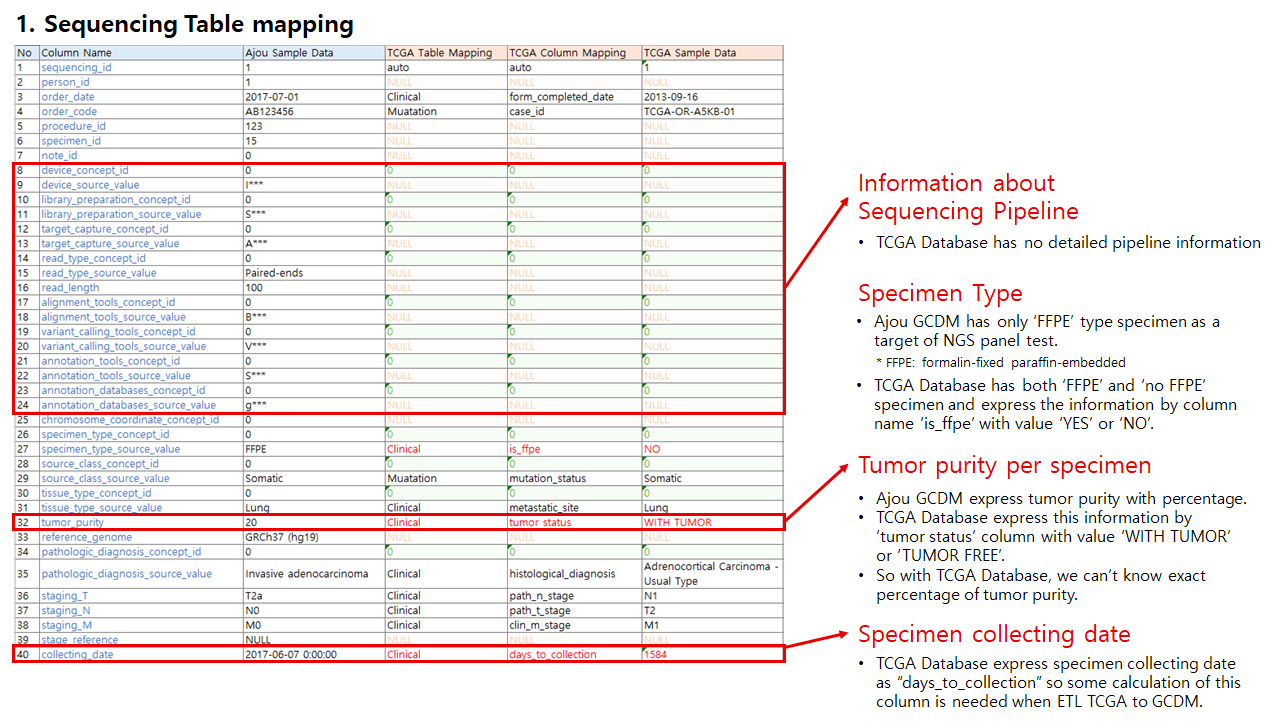
GCDM Issue List for Everyone

updated: 2018-01-12

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Date | Suggestor | Issue |
| 1 | 2018-01-04 | Seojeong Shin | Sequencing Table |



1-1. Information about Sequencing Pipeline

: TCGA Database has no detailed pipeline information

1-2. Specimen Type

: Ajou GCDM has only ‘FFPE’ type specimen as a target of NGS panel test.

(\* FFPE: formalin-fixed paraffin-embedded)

: TCGA Database has both ‘FFPE’ and ‘no FFPE’ specimen and express the information by column name ‘is\_ffpe’ with value ‘YES’ or ‘NO’.

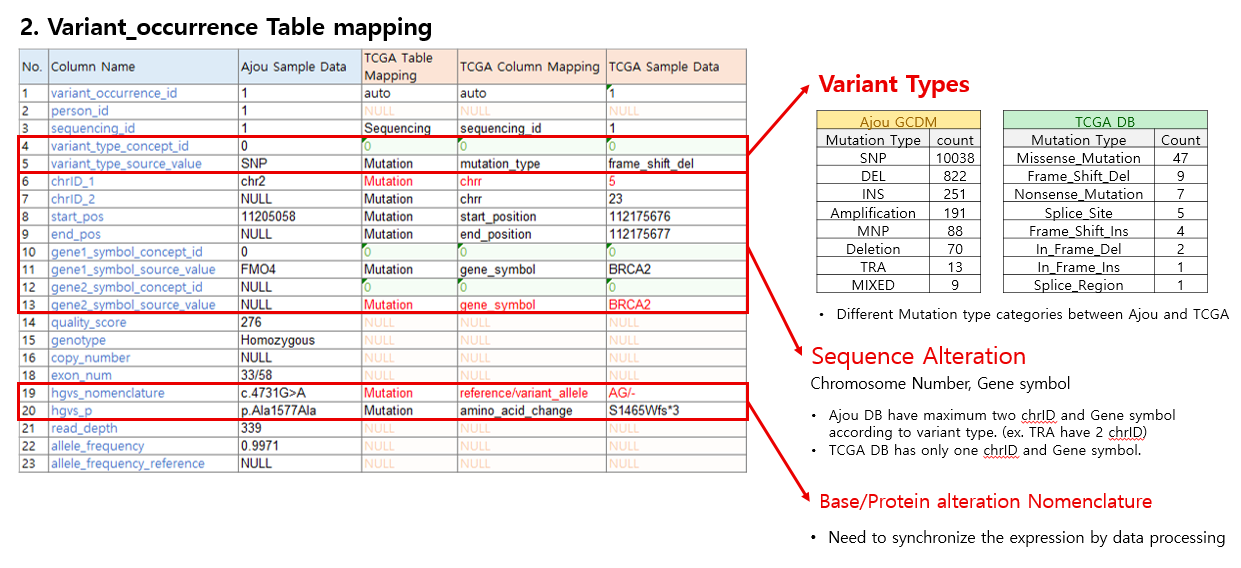
1-3. Tumor purity per specimen

: Ajou GCDM express tumor purity with percentage. TCGA Database express this information by ‘tumor status’ column with value ‘WITH TUMOR’ or ‘TUMOR FREE’. So with TCGA Database, we can’t know exact percentage of tumor purity.

1-4. Specimen collecting date

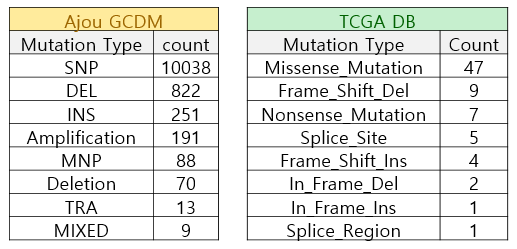
: TCGA Database express specimen collecting date as “days\_to\_collection” so some calculation of this column is needed when ETL TCGA to GCDM.

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Date | Suggestor | Issue |
| 2 | 2018-01-04 | Seojeong Shin | Variant\_occurrence Table |



2-1. Variant Types

: Different Mutation type categories between Ajou and TCGA



: Clear and common categories should be defined for types of variant.

2-2. Sequence Alteration

: Chromosome Number, Gene symbol

: Ajou DB have maximum two chrID and Gene symbol according to variant type.

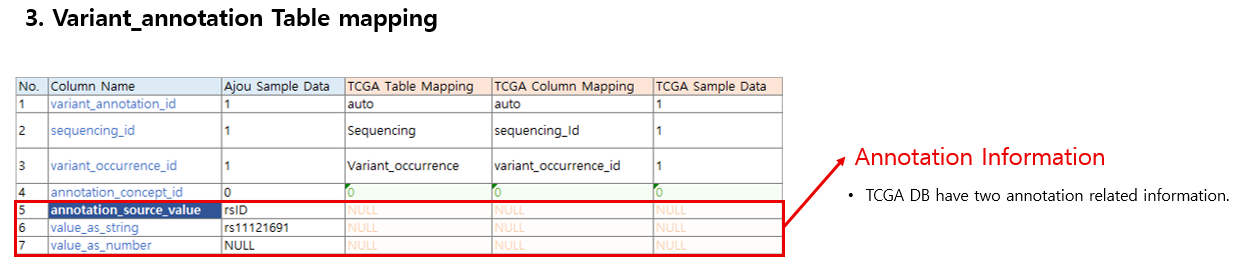
(ex. TRA have 2 chrID)

: TCGA DB has only one chrID and Gene symbol.

2-3. Base/Protein alteration Nomenclature

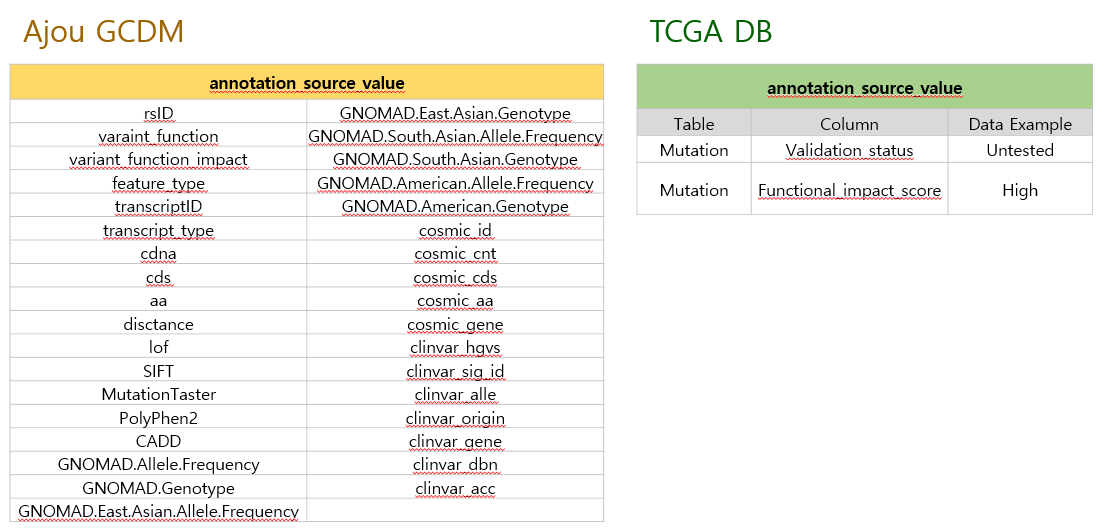
: Need to synchronize the expression by data processing

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Date | Suggestor | Issue |
| 3 | 2018-01-04 | Seojeong Shin | Variant\_annotation Table |



3-1. Annotation Information

: TCGA DB have two annotation related information.



|  |  |  |  |
| --- | --- | --- | --- |
| No. | Date | Suggestor | Issue |
| 4 | 2018-01-12 | Seojeong Shin | Column Adding |

4-1. Sequencing Center

: Sequencing Center column will important to know NGS pipeline and regional information.

: The Column should be added into ‘Sequencing’ Table.

4-2. Total target gene list

: Target gene list should be added to ‘Sequencing’ Table. It’ll be used for knowing which gene is not mutated in each patient.

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Date | Suggestor | Issue |
| 5 | 2018-01-31 | Clair | Review G-CDM created by Chan & Seojeong, Ajou Univ (2018.1.31) |

-Participants: Clair, Chan, Seojeong, Qi, Sunah, Mina, Michel, David

* Modification of Related table
  + specimen table
    - Should specimen table instead of sequencing table include **specimen type**? (FFPE vs non-FFPE)

⇒ We’ll propose to add ‘specimen\_preservation\_concept\_id’ to ‘specimen’ table

* + - Should specimen table instead of sequencing table include **tumor purity per specimen**

⇒ disease\_status\_concept\_id??

* + - Should specimen table instead of sequencing table include **collecting date for specimen**

⇒ collecting\_date in sequencing table was removed

* + Oncology expansion table
    - **Pathologic diagnosis** or **staging** should be included in oncology expansion table (I’ll wait for solution of oncology working group)

⇒ Oncology working group will suggest the oncology extension model soon. We’re waiting for it.

* Sequencing pipeline
  + Sequencing table cannot capture the targeting gene set (for targeted NGS). How can we fix this?
  + Shouldn’t we generate a database in OHDSI for each sequencing pipeline?

⇒ It isn’t easy work. We have not reached an agreement yet.

* Change of the table schema
  + Where functional mutation information should be stored in variant\_occurrence table or variant annotation table?  
    eg. SNP, DEL vs. Missense Mutation, Nonsense mutation

⇒ SNP, insertion, deletion should be stored in ‘variant\_type\_concept\_id’

⇒ **‘Variant\_function\_concept\_id’** is added to variant\_occurrence table (eg. missense, nonsense, intron, nonsense, splice region variant….)

* Essential information for Variant\_annotation

⇒ We need more experiences and your suggestions.

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Date | Suggestor | Issue |
| 6 | 2018-02-14 | Chan&ShinSJ | Modification of Related table and GCDM schema (2018.02.14) |

-Participants: Clair, Chan, Seojeong, KP, Frank, David, Qi, Sunah

* Modification of Related table
  + specimen table
    - We’ll propose to add **‘specimen\_preservation\_concept\_id’** to ‘specimen’ table. (FFPE vs non-FFPE)
    - **‘collecting\_date’** column in sequencing table was removed.
* Change of the table schema
  + Sequencing table
    - **‘Sequencing\_date’** column was added.
  + Variant\_occurrence table and Variant annotation table
    - SNP, insertion, deletion should be stored in **‘variant\_type\_concept\_id’** in variant\_occurrence table.
    - **‘variant\_function\_concept\_id’** is added to variant\_occurrence table (eg. missense, nonsense, intron, nonsense, splice region variant….) for expressing functional change of each variant based on reference genome.
* [**GCDM Data Overview (Ajou DB)**](https://docs.google.com/spreadsheets/d/1ijXcWd7Soc13EbunNvQgBCZWxBa_PzPQUeIxDs3nMB0/edit?usp=sharing)
  + Based on the NGS data set of Ajou University Hospital, the data composition of each column has been shown graphically.