National Biobank Consortium of Taiwan
Common Data Model

Oncology Genomics Data Dictionary

NBCT CDM
Oncology Genomics Data Dictionary

Page Index

NBCT CDM

Oncology Genomics Data Dictionary

Table Overview

No	Table Name	Description
01	REPORT_PATIENT	Report_Patient
02	GENOFIND	Report_Genomic Findings
03	BIOFIND	Report_Biomarker_Findings
04	SV	Variant_Short-Variants
05	CNA	Variant_Copy-Number-Alterations
06	REARRANGEMENTS	Variant_Rearrangements
07	NONHUMAN	Variant_Non-Human-Content

NBCT CDM

Oncology Genomics Data Dictionary

01_ REPORT_PATIENT (Report_Patient)

No.	Field Name	Description	Data Type (C/N)	Format / Example
1	hash_id	FMI's unique ID for this sample (de-identified)	C	
2	submitted_diagnosis	Patient's diagnosis submitted with sample	C	Unknown primary melanoma
3	spec_site	Location from which the sample was retrieved	C	Rectum Blood
4	test_type	Name of the test performed on this sample	C	FoundationOneDx FoundationOneLiquidDx
5	coll_date	Date when the sample was collected	C	20180320
6	received_date	Date when the sample was received	C	20180330

Data Type: C (Character/String), N (Numeric)

02_ GENOFIND (Report_Genomic Findings)

No.	Field Name	Description	Data Type (C/N)	Format / Example
1	hash_id	FMI's unique ID for this sample (de-identified)	C	
2	gene_name	Name of gene	C	KDR AR
3	alteration_property_name	Name of the alteration property	C	Amplification L702H

Data Type: C (Character/String), N (Numeric)

03_ BIOFIND (Report_Biomarker_Findings)

No.	Field Name	Description	Data Type (C/N)	Format / Example
1	hash_id	FMI's unique ID for this sample (de-identified)	C	
2	report_property	Property key	C	TumorMutationBurdenScore Microsatellite status
3	value	Value of defined property	C	3 Muts/Mb MS-Stable

Data Type: C (Character/String), N (Numeric)

04_ SV (Variant_Short-Variants)

No.	Field Name	Description	Data Type (C/N)	Format / Example
1	hash_id	FMI's unique ID for this sample (de-identified)	C	
2	transcript	The accession/identifier of the reference transcript used as the basis for attributes like cds-effect and protein-effect	C	NM_002447
3	gene	The name (gene symbol) of the nearest targeted gene associated with this short variant	C	NM_000044
4	position	The genomic position (chromosome name and coordinate) of this short variant; the coordinate corresponds to the position value in the VCF representation	C	MST1R
5	cds_effect	The CDS-level effect of this short variant for the associated gene (typically the standard CDS mutant syntax, excluding the "c." prefix)	C	AR
6	protein_effect	The protein-level effect of this short variant for the associated gene (typically either the standard protein mutant syntax, excluding the "p." prefix, for coding variants, or special values starting with "splice" or "promoter")	C	chr3:49939820
7	depth	The local sequencing depth at the genomic position of this short variant	C	chrX:66931463
8	percent_reads	The percent of sequencing reads at the genomic position of this variant that support the variant allele (For short-variants, this is identical to allele-fraction except represented as a percentage.)	C	1223C>T

Data Type: C (Character/String), N (Numeric)

05_ CNA (Variant_Copy-Number-Alterations)

No.	Field Name	Description	Data Type (C/N)	Format / Example
1	hash_id	FMI's unique ID for this sample (de-identified)	C	
2	gene	The name (gene symbol) of the targeted gene associated with this copy number alteration variant	C	KDR
3	position	The genomic position (chromosome name and coordinate interval) of this copy number alteration variant	C	chr4:55896637-56038755
4	ratio	The log (base 2) of the normalized tumor/reference coverage ratio for the copy number alteration	C	1.68
5	type	The type of copy number alteration	C	amplification

Data Type: C (Character/String), N (Numeric)

06_ REARRANGEMENTS (Variant_Rearrangements)

No.	Field Name	Description	Data Type (C/N)	Format / Example
1	hash_id	FMI's unique ID for this sample (de-identified)	C	
2	othe_gene	The non-targeted or secondary targeted gene, if any, involved in this rearrangement	C	CNTNAP2
3	percent_reads	The percentage of sequencing reads at the genomic position of this variant that support the variant allele. (For rearrangements, this is related to, but not equivalent to, the allele-fraction value. The allele-fraction value is typically the value of interest.)	C	0.39
4	pos1	The genomic position (chromosome name and coordinate interval) corresponding to the target-gene breakpoint of this rearrangement variant	C	chr10:43615345
5	pos2	The genomic position (chromosome name and coordinate interval) corresponding to the other-gene breakpoint of this rearrangement variant	C	chr7:146206742
6	targeted_gene	The targeted gene (or primary targeted gene) involved in this rearrangement	C	RET

Data Type: C (Character/String), N (Numeric)

07_ NONHUMAN (Variant_Non-Human-Content)

No.	Field Name	Description	Data Type (C/N)	Format / Example
1	hash_id	FMI's unique ID for this sample (de-identified)	C	
2	organism	The name of the non-human organism detected	C	HPV-16
3	reads_per_million	The normalized abundance of sequence read data supporting the non-human content call	C	15

Data Type: C (Character/String), N (Numeric)