

# Genomic Metadata Data Dictionary for the CP Commons

This metadata data dictionary was developed in 2019.

These metadata are designed for the CP Commons.

There are 17 fields: 6 are Mandatory.

Any problems using this form, please contact Yana Wilson ([ywilson@cerebralpalsy.org.au](mailto:ywilson@cerebralpalsy.org.au)).

Current working version: Version 1.1

## Unique IDs (2)

FIELD NAME	CATEGORY	DESCRIPTION	VALUE	NOTES /LOGIC
submitter_ID	<b>Mandatory</b>	Unique ID from submitting team	String, alphanumeric (no special characters)	Unique ID from submitting team, must be two steps removed from any personal information. The SUBMITTER_ID can only be viewed by Data Owners and CPA Data Custodians.
family_ID	<b>Mandatory</b>	Family ID	String, alphanumeric, permissible special characters ( _ )	ID code assigned by the submitting team to a family. FAMILY_ID is not necessary for singleton cases.

## Genomic Fields (15)

FIELD NAME	REQUIRED	DESCRIPTION	VALUE	NOTES /LOGIC
filename	<b>Mandatory</b>	Name of the genome datafile	String, alphanumeric, permissible special characters ( _ )	The name of the file
data_access_restriction	<b>Mandatory</b>	Data Access Restrictions	1, Open-access data 2, Controlled-access data	<u>Open</u> – immediately accessible to users <u>Controlled</u> – Users must submit a data access request to access the data.
consent	<b>Mandatory</b>	Do you have consent from the participant to share their de-identified data?	0, No 1, Yes	
consent_no	Core	If no, have you received a consent waiver to share these data without the individuals consent?	0, No 1, Yes	Response required if [consent] has value of 0.
restrictions	Recommended	Does this data have additional data use restrictions?	0, No 1, Yes	

pipeline	Core	Please select the CP Commons harmonisation pipeline	1, Pre-processing (BAM only) 2, gVCF pipeline 3, SNPs + indels
file_format	Core	Please indicate the file format for this data file:	1, VCF 2, gVCF 3, BAM 4, TXT
molecular_class	Core	Broad categorisation of the molecular data:	1, Genome 2, Methylome 3, Transcriptome
exp_strategy	Core	Please select the sequencing strategy used to generate the data file:	1, Whole Genome Sequencing (WGS) 2, Whole Exome Sequencing (WES) 3, Array-CGH (CNV) 4, SNP-array 5, Methylation array 6, CpG island array 7, Whole Genome Bisulfite-sequencing (WGBS) 8, Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) 9, Chromatin immunoprecipitation sequencing (ChIP-seq) 10, microRNA sequencing (miRNA-seq) 11, Microarray 12, Whole RNA-seq 13, Whole transcriptome shotgun sequencing (WTSS) 14, Targeted RNA-seq amplicon
instrument	Recommended	Name of platform used for sequencing:	1, Affymetrix 2, Agilent 3, Illumina 4, Ion Torrent

			5, Nimblegen 6, PacBio	
instrument_model	Recommended	Name of instrument model used for sequencing:	String, alphanumeric, permissible special characters ( )	
instrument_centre	Recommended	Where was the sequencing performed?	String, alphanumeric, permissible special characters ( )	
raw_data	Core	Is the raw data stored in an open database?	0, No 1, Yes	
raw_data_detail	Core	If yes, please provide the hyperlink:	String, alphanumeric, permissible special characters ( _ , . , / , : )	Response required on if [raw_data] has a value of (1).
md5sum	<b>Mandatory</b>	MD5 Sum:	Integer	The 128-bit hash value expressed as a 32 digit hexadecimal number (in lower case) used as a file's digital fingerprint.