Accelerating Medicines Partnership: Parkinson's Disease (AMP PD)

Transcriptomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
alignment_summary_metrics	participant_id	STRING	null	The Participant ID.
alignment_summary_metrics	sample_id	STRING	null	The Sample ID.
alignment_summary_metrics	CATEGORY	STRING	null	One of either UNPAIRED (for a fragment run), FIRST_OF_PAIR when metrics are for only the first read in a paired run, SECOND_OF_PAIR when the metrics are for only the second read in a paired run or PAIR when the metrics are aggregated for both first and second reads in a pair.
alignment_summary_metrics	TOTAL_READS	INTEGER	null	The total number of reads including all PF and non-PF reads. When CATEGORY equals PAIR this value will be 2x the number of clusters.
alignment_summary_metrics	PF_READS	INTEGER	null	The number of PF reads where PF is defined as passing Illumina's filter.
alignment_summary_metrics	PCT_PF_READS	INTEGER	null	The fraction of reads that are PF (PF_READS / TOTAL_READS)
alignment_summary_metrics	PF_NOISE_READS	INTEGER	null	The number of PF reads that are marked as noise reads. A noise read is one which is composed entirely of A bases and/or N bases. These reads are marked as they are usually artifactual and are of no use in downstream analysis.
alignment_summary_metrics	PF_READS_ALIGNED	INTEGER	null	The number of PF reads that were aligned to the reference sequence. This includes reads that aligned with low quality (i.e. their alignments are ambiguous).
alignment_summary_metrics	PCT_PF_READS_ALIGNED	FLOAT	null	The percentage of PF reads that aligned to the reference sequence. PF_READS_ALIGNED / PF_READS
alignment_summary_metrics	PF_ALIGNED_BASES	INTEGER	null	The total number of aligned PF bases. Non-primary alignments are not counted. Bases in aligned reads that do not correspond to reference (e.g. soft clips, insertions) are not counted.
alignment_summary_metrics	PF_HQ_ALIGNED_READS	INTEGER	null	The number of PF reads that were aligned to the reference sequence with a mapping quality of Q20 or higher signifying that the aligner estimates a 1/100 (or smaller) chance that the alignment is wrong.
alignment_summary_metrics	PF_HQ_ALIGNED_BASES	INTEGER	null	The number of bases aligned to the reference sequence in reads that were mapped at high quality. Will usually approximate PF_HQ_ALIGNED_READS * READ_LENGTH but may differ when either mixed read lengths are present or many reads are aligned with gaps.
alignment_summary_metrics	PF_HQ_ALIGNED_Q20_BASES	INTEGER	null	The subset of PF_HQ_ALIGNED_BASES where the base call quality was Q20 or higher.
alignment_summary_metrics	PF_HQ_MEDIAN_MISMATCHES	INTEGER	null	The median number of mismatches versus the reference sequence in reads that were aligned to the reference at high quality (i.e. PF_HQ_ALIGNED READS).
alignment_summary_metrics	PF_MISMATCH_RATE	FLOAT	null	The rate of bases mismatching the reference for all bases aligned to the reference sequence.
alignment_summary_metrics	PF_HQ_ERROR_RATE	FLOAT	null	The fraction of bases that mismatch the reference in PF HQ aligned reads.
alignment_summary_metrics	PF_INDEL_RATE	FLOAT	null	The number of insertion and deletion events per 100 aligned bases. Uses the number of events as the numerator, not the number of inserted or deleted bases.
alignment_summary_metrics	MEAN_READ_LENGTH	FLOAT	null	The mean read length of the set of reads examined. When looking at the data for a single lane with equal length reads this number is just the read length. When looking at data for merged lanes with differing read lengths this is the mean read length of all reads.
alignment_summary_metrics	READS_ALIGNED_IN_PAIRS	INTEGER	null	The number of aligned reads whose mate pair was also aligned to the reference.
alignment_summary_metrics	PCT_READS_ALIGNED_IN_PAIRS	FLOAT	null	The fraction of reads whose mate pair was also aligned to the reference. READS_ALIGNED_IN_PAIRS / PF_READS_ALIGNED
alignment_summary_metrics	PF_READS_IMPROPER_PAIRS	INTEGER	null	The number of (primary) aligned reads that are **not** properly" aligned in pairs (as per SAM flag 0x2)."
alignment_summary_metrics	PCT_PF_READS_IMPROPER_PAIRS	FLOAT	null	The fraction of (primary) reads that are *not* properly" aligned in pairs (as per SAM flag 0x2). PF_READS_IMPROPER_PAIRS / PF_READS_ALIGNED"
alignment_summary_metrics	BAD_CYCLES	INTEGER	null	The number of instrument cycles in which 80% or more of base calls were no-calls.
alignment_summary_metrics	STRAND_BALANCE	FLOAT	null	The number of PF reads aligned to the positive strand of the genome divided by the number of PF reads aligned to the genome.
alignment_summary_metrics	PCT_CHIMERAS	FLOAT	null	The fraction of reads that map outside of a maximum insert size (usually 100kb) or that have the two ends mapping to different chromosomes.

Accelerating Medicines Partnership: Parkinson's Disease (AMP PD)

Transcriptomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
alignment_summary_metrics	PCT_ADAPTER	FLOAT	null	The fraction of PF reads that are unaligned and match to a known adapter sequence right from the start of the read.
alignment_summary_metrics	SAMPLE	STRING	null	The sample to which these metrics apply. If null, it means they apply to all reads in the file.
alignment_summary_metrics	LIBRARY	STRING	null	The library to which these metrics apply. If null, it means that the metrics were accumulated at the sample level.
alignment_summary_metrics	READ_GROUP	STRING	null	The read group to which these metrics apply. If null, it means that the metrics were accumulated at the library or sample level.
feature_counts	participant_id	STRING	NULLABLE	The Participant ID.
feature_counts	sample_id	STRING	NULLABLE	The Sample ID.
feature_counts	Geneid	STRING	NULLABLE	Ensembl ID
feature_counts	Chr	STRING	NULLABLE	Chromosome name
feature_counts	Start	STRING	NULLABLE	Gene chromosomal start position.
feature_counts	End	STRING	NULLABLE	Gene chromosomal end position
feature_counts	Strand	STRING	NULLABLE	Strand
feature_counts	Length	INTEGER	NULLABLE	The total number of non-overlapping bases in exons belonging to the same gene for each gene.
feature_counts	Value	INTEGER	NULLABLE	Counts
genome_check_HW_MAF	sample_id_1	STRING	REQUIRED	Sample ID for first sample
genome_check_HW_MAF	sample_id_2	STRING	REQUIRED	Sample ID for second sample
genome_check_HW_MAF	FID1	STRING	REQUIRED	Family ID for first sample
genome_check_HW_MAF	IID1	STRING	REQUIRED	Individual ID for first sample
genome_check_HW_MAF	FID2	STRING	REQUIRED	Family ID for second sample
genome_check_HW_MAF	IID2	STRING	REQUIRED	Individual ID for second sample
genome_check_HW_MAF	RT	STRING	REQUIRED	Relationship type inferred from .fam/.ped file
genome_check_HW_MAF	EZ	INTEGER	REQUIRED	IBD sharing expected value, based on just .fam/.ped relationship
genome_check_HW_MAF	Z0	FLOAT	REQUIRED	P(IBD=0)
genome_check_HW_MAF	Z1	FLOAT	REQUIRED	P(IBD=1)
genome_check_HW_MAF	Z2	FLOAT	REQUIRED	P(IBD=2)
genome_check_HW_MAF	PI_HAT	FLOAT	REQUIRED	Proportion IBD, i.e. P(IBD=2) + 0.5*P(IBD=1)
genome_check_HW_MAF	PHE	INTEGER	REQUIRED	Pairwise phenotypic code (1, 0, -1 = AA, AU, and UU pairs, respectively)
genome_check_HW_MAF	DST	FLOAT	REQUIRED	IBS distance, i.e. (IBS2 + 0.5*IBS1) / (IBS0 + IBS1 + IBS2)
genome_check_HW_MAF	PPC	FLOAT	REQUIRED	IBS binomial test
genome_check_HW_MAF	RATIO	FLOAT	NULLABLE	HETHET : IBS0 SNP ratio (expected value 2)
insert_size_metrics	participant_id	STRING	null	The Participant ID.
insert_size_metrics	sample_id	STRING	null	The Sample ID.
insert_size_metrics	MEDIAN_INSERT_SIZE	INTEGER	null	The MEDIAN insert size of all paired end reads where both ends mapped to the same chromosome.
insert_size_metrics	MODE_INSERT_SIZE	INTEGER	null	The MODE insert size
insert_size_metrics	MEDIAN_ABSOLUTE_DEVIATION	INTEGER	null	The median absolute deviation of the distribution. If the distribution is essentially normal then the standard deviation can be estimated as ~1.4826 * MAD.
insert_size_metrics	MIN_INSERT_SIZE	INTEGER	null	The minimum measured insert size. This is usually 1 and not very useful as it is likely artifactual.
insert_size_metrics	MAX_INSERT_SIZE	INTEGER	null	The maximum measure insert size by alignment. This is usually very high representing either an artifact or possibly the presence of a structural re-arrangement.

Accelerating Medicines Partnership: Parkinson's Disease (AMP PD)

Transcriptomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
insert_size_metrics	MEAN_INSERT_SIZE	FLOAT	null	The mean insert size of the core" of the distribution. Artefactual outliers in the distribution often cause calculation of nonsensical mean and stdev values. To avoid this the distribution is first trimmed to a "core" distribution of +/- N median absolute deviations around the median insert size. By default N=10
insert_size_metrics	STANDARD_DEVIATION	FLOAT	null	Standard deviation of insert sizes over the core" of the distribution."
insert_size_metrics	READ_PAIRS	INTEGER	null	The total number of read pairs that were examined in the entire distribution.
insert_size_metrics	PAIR_ORIENTATION	STRING	null	The pair orientation of the reads in this data category.
insert_size_metrics	WIDTH_OF_10_PERCENT	INTEGER	null	The width" of the bins centered that encompass 10% of all read pairs."
insert_size_metrics	WIDTH_OF_20_PERCENT	INTEGER	null	The width" of the bins centered that encompass 20% of all read pairs."
insert_size_metrics	WIDTH_OF_30_PERCENT	INTEGER	null	The width" of the bins centered that encompass 30% of all read pairs."
insert_size_metrics	WIDTH_OF_40_PERCENT	INTEGER	null	The width" of the bins centered that encompass 40% of all read pairs."
insert_size_metrics	WIDTH_OF_50_PERCENT	INTEGER	null	The width" of the bins centered that encompass 50% of all read pairs."
insert_size_metrics	WIDTH_OF_60_PERCENT	INTEGER	null	The width" of the bins centered that encompass 60% of all read pairs."
insert_size_metrics	WIDTH_OF_70_PERCENT	INTEGER	null	The width" of the bins centered that encompass 70% of all read pairs. This metric divided by 2 should approximate the standard deviation when the insert size distribution is a normal distribution."
insert_size_metrics	WIDTH_OF_80_PERCENT	INTEGER	null	The width" of the bins centered that encompass 80% of all read pairs."
insert_size_metrics	WIDTH_OF_90_PERCENT	INTEGER	null	The width" of the bins centered that encompass 90% of all read pairs."
insert_size_metrics	WIDTH_OF_95_PERCENT	INTEGER	null	The width" of the bins centered that encompass 95% of all read pairs."
insert_size_metrics	WIDTH_OF_99_PERCENT	INTEGER	null	The width" of the bins centered that encompass 100% of all read pairs."
insert_size_metrics	SAMPLE	STRING	null	The sample to which these metrics apply. If null, it means they apply to all reads in the file.
insert_size_metrics	LIBRARY	STRING	null	The library to which these metrics apply. If null, it means that the metrics were accumulated at the sample level.
insert_size_metrics	READ_GROUP	STRING	null	The read group to which these metrics apply. If null, it means that the metrics were accumulated at the library or sample level.
quantification_genes	participant_id	STRING	NULLABLE	The Participant ID.
quantification_genes	sample_id	STRING	NULLABLE	The Sample ID.
quantification_genes	Name	STRING	NULLABLE	Name of the target transcript provided in the input transcript database (FASTA file).
quantification_genes	Length	FLOAT	NULLABLE	Length of the target transcript in nucleotides.
quantification_genes	EffectiveLength	FLOAT	NULLABLE	Computed effective length of the target transcript. It takes into account all factors being modeled that will effect the probability of sampling fragments from this transcript, including the fragment length distribution and sequence-specific and gc-fragment bias (if they are being modeled).
quantification_genes	TPM	FLOAT	NULLABLE	This is salmon's estimate of the relative abundance of this transcript in units of Transcripts Per Million (TPM). TPM is the recommended relative abundance measure to use for downstream analysis.
quantification_genes	NumReads	FLOAT	NULLABLE	Salmon's estimate of the number of reads mapping to each transcript that was quantified. It is an "estimate" insofar as it is the expected number of reads that have originated from each transcript given the structure of the uniquely mapping and multi-mapping reads and the relative abundance estimates for each transcript.
quantification_transcripts	participant_id	STRING	NULLABLE	The Participant ID.
quantification_transcripts	sample_id	STRING	NULLABLE	The Sample ID.
quantification_transcripts	Name	STRING	NULLABLE	Name of the target transcript provided in the input transcript database (FASTA file).
quantification_transcripts	Length	INTEGER	NULLABLE	Length of the target transcript in nucleotides.
quantification_transcripts	EffectiveLength	FLOAT	NULLABLE	Computed effective length of the target transcript. It takes into account all factors being modeled that will effect the probability of sampling fragments from this transcript, including the fragment length distribution and sequence-specific and gc-fragment bias (if they are being modeled).

Accelerating Medicines Partnership: Parkinson's Disease (AMP PD)

Transcriptomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
quantification_transcripts	TPM	FLOAT	NULLABLE	Salmon's estimate of the relative abundance of this transcript in units of Transcripts Per Million (TPM). TPM is the recommended relative abundance measure to use for downstream analysis.
quantification_transcripts	NumReads	FLOAT	NULLABLE	Salmon's estimate of the number of reads mapping to each transcript that was quantified. It is an "estimate" insofar as it is the expected number of reads that have originated from each transcript given the structure of the uniquely mapping and multi-mapping reads and the relative abundance estimates for each transcript.
rna_quality_metrics	sample_id	STRING	NULLABLE	The Sample ID.
rna_quality_metrics	Specimen_Quantity	FLOAT	NULLABLE	Quantity of RNA received by sequencing facility from biorepository.
rna_quality_metrics	Concentration	FLOAT	NULLABLE	Concentration of RNA in sample received by sequencing facility from biorepository.
rna_quality_metrics	Submitted_Volume__ul_	FLOAT	NULLABLE	Volume of sample received by sequencing facility from biorepository.
rna_quality_metrics	Normalization_Volume__30ng_ul_	FLOAT	NULLABLE	Volume of sample following RNA concentration normalization.
rna_quality_metrics	Total_Volume__ul_	FLOAT	NULLABLE	Sample total volume.
rna_quality_metrics	Input_RNASeq__ng_	FLOAT	NULLABLE	RNA quantity used as input for RNAseq
rna_quality_metrics	Input_miRNA__ng_	FLOAT	NULLABLE	RNA quantity used as input for smallRNA project
rna_quality_metrics	_260_230_Ratio	FLOAT	NULLABLE	Ratio of nanodrop spectrophotometer absorbance measurements at 260 nm and 280 nm.
rna_quality_metrics	_260_280_Ratio	FLOAT	NULLABLE	Ratio of nanodrop spectrophotometer absorbance measurements at 260 nm and 230 nm.
rna_quality_metrics	RIN_Value	FLOAT	NULLABLE	RNA integrity number
rna_quality_metrics	Box	STRING	NULLABLE	Box used to ship sample from biorepository to sequencing facility.
rna_quality_metrics	Plate	STRING	NULLABLE	Library preparation plate ID.
rna_quality_metrics	Position	INTEGER	NULLABLE	Sample position on library preparation 96 well plate.
rna_seq_metrics	participant_id	STRING	null	The Participant ID.
rna_seq_metrics	sample_id	STRING	null	The Sample ID.
rna_seq_metrics	PF_BASES	INTEGER	null	The total number of PF bases including non-aligned reads.
rna_seq_metrics	PF_ALIGNED_BASES	INTEGER	null	The total number of aligned PF bases. Non-primary alignments are not counted. Bases in aligned reads that do not correspond to reference (e.g. soft clips, insertions) are not counted.
rna_seq_metrics	RIBOSOMAL_BASES	INTEGER	null	Number of bases in primary alignments that align to ribosomal sequence.
rna_seq_metrics	CODING_BASES	INTEGER	null	Number of bases in primary alignments that align to a non-UTR coding base for some gene, and not ribosomal sequence.
rna_seq_metrics	UTR_BASES	INTEGER	null	Number of bases in primary alignments that align to a UTR base for some gene, and not a coding base.
rna_seq_metrics	INTRONIC_BASES	INTEGER	null	Number of bases in primary alignments that align to an intronic base for some gene, and not a coding or UTR base.
rna_seq_metrics	INTERGENIC_BASES	INTEGER	null	Number of bases in primary alignments that do not align to any gene.
rna_seq_metrics	IGNORED_READS	INTEGER	null	Number of primary alignments that are mapped to a sequence specified on command-line as IGNORED_SEQUENCE. These are not counted in PF_ALIGNED_BASES, CORRECT_STRAND_READS, INCORRECT_STRAND_READS, or any of the base-counting metrics. These reads are counted in PF_BASES.
rna_seq_metrics	CORRECT_STRAND_READS	INTEGER	null	Number of aligned reads that are mapped to the correct strand. 0 if library is not strand-specific.
rna_seq_metrics	INCORRECT_STRAND_READS	INTEGER	null	Number of aligned reads that are mapped to the incorrect strand. 0 if library is not strand-specific.
rna_seq_metrics	NUM_R1_TRANSCRIPT_STRAND_READS	INTEGER	null	The number of reads that support the model where R1 is on the strand of transcription and R2 is on the opposite strand.
rna_seq_metrics	NUM_R2_TRANSCRIPT_STRAND_READS	INTEGER	null	The fraction of reads that support the model where R2 is on the strand of transcription and R1 is on the opposite strand.

Accelerating Medicines Partnership: Parkinson's Disease (AMP PD)

Transcriptomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
rna_seq_metrics	NUM_UNEXPLAINED_READS	INTEGER	null	The fraction of reads for which the transcription strand model could not be inferred.
rna_seq_metrics	PCT_R1_TRANSCRIPT_STRAND_READS	FLOAT	null	The fraction of reads that support the model where R1 is on the strand of transcription and R2 is on the opposite strand. For unpaired reads, it is the fraction of reads that are on the transcription strand (out of all the reads).
rna_seq_metrics	PCT_R2_TRANSCRIPT_STRAND_READS	FLOAT	null	The fraction of reads that support the model where R2 is on the strand of transcription and R1 is on the opposite strand. For unpaired reads, it is the fraction of reads that are on opposite strand than that of the the transcription strand (out of all the reads).
rna_seq_metrics	PCT_RIBOSOMAL_BASES	FLOAT	null	Fraction of PF_ALIGNED_BASES that mapped to regions encoding ribosomal RNA, RIBOSOMAL_BASES/PF_ALIGNED_BASES
rna_seq_metrics	PCT_CODING_BASES	FLOAT	null	Fraction of PF_ALIGNED_BASES that mapped to protein coding regions of genes, CODING_BASES/PF_ALIGNED_BASES
rna_seq_metrics	PCT_UTR_BASES	FLOAT	null	Fraction of PF_ALIGNED_BASES that mapped to untranslated regions (UTR) of genes, UTR_BASES/PF_ALIGNED_BASES
rna_seq_metrics	PCT_INTRONIC_BASES	FLOAT	null	Fraction of PF_ALIGNED_BASES that correspond to gene introns, INTRONIC_BASES/PF_ALIGNED_BASES
rna_seq_metrics	PCT_INTERGENIC_BASES	FLOAT	null	Fraction of PF_ALIGNED_BASES that mapped to intergenic regions of genomic DNA, INTERGENIC_BASES/PF_ALIGNED_BASES
rna_seq_metrics	PCT_MRNA_BASES	FLOAT	null	Sum of bases mapped to regions corresponding to UTRs and coding regions of mRNA transcripts, PCT_UTR_BASES + PCT_CODING_BASES
rna_seq_metrics	PCT_USABLE_BASES	FLOAT	null	The fraction of bases mapping to mRNA divided by the total number of PF bases, (CODING_BASES + UTR_BASES)/PF_BASES.
rna_seq_metrics	PCT_CORRECT_STRAND_READS	FLOAT	null	Fraction of reads corresponding to mRNA transcripts which map to the correct strand of a reference genome = CORRECT_STRAND_READS/(CORRECT_STRAND_READS + INCORRECT_STRAND_READS). 0 if library is not strand-specific.
rna_seq_metrics	MEDIAN_CV_COVERAGE	FLOAT	null	The median coefficient of variation (CV) or stdev/mean for coverage values of the 1000 most highly expressed transcripts. Ideal value = 0.
rna_seq_metrics	MEDIAN_5PRIME_BIAS	FLOAT	null	The median 5 prime bias of the 1000 most highly expressed transcripts. The 5 prime bias is calculated per transcript as: mean coverage of the 5 prime-most 100 bases divided by the mean coverage of the whole transcript.
rna_seq_metrics	MEDIAN_3PRIME_BIAS	FLOAT	null	The median 3 prime bias of the 1000 most highly expressed transcripts, where 3 prime bias is calculated per transcript as: mean coverage of the 3 prime-most 100 bases divided by the mean coverage of the whole transcript.
rna_seq_metrics	MEDIAN_5PRIME_TO_3PRIME_BIAS	FLOAT	null	The ratio of coverage at the 5 prime end to the 3 prime end based on the 1000 most highly expressed transcripts.
rna_seq_metrics	SAMPLE	STRING	null	The sample to which these metrics apply. If null, it means they apply to all reads in the file.
rna_seq_metrics	LIBRARY	STRING	null	The library to which these metrics apply. If null, it means that the metrics were accumulated at the sample level.
rna_seq_metrics	READ_GROUP	STRING	null	The read group to which these metrics apply. If null, it means that the metrics were accumulated at the library or sample level.
rna_seq_samples	participant_id	STRING	REQUIRED	The Participant ID.
rna_seq_samples	sample_id	STRING	REQUIRED	The Sample ID.
rna_seq_samples	visit_month	FLOAT	REQUIRED	Numeric encoding of a visit.
rna_seq_samples	BAM	STRING	REQUIRED	Location of .bam file in GCS.
rna_seq_samples	BAI	STRING	REQUIRED	Location of .bai file in GCS.
rna_seq_samples	QUANT_SF	STRING	REQUIRED	Location of quant.sf file in GCS.
rna_seq_samples	QUANT_GENES_SF	STRING	REQUIRED	Location of quant.genes.sf file in GCS.

Accelerating Medicines Partnership: Parkinson's Disease (AMP PD)

Transcriptomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
rna_seq_samples	FEATURECOUNTS_TSV	STRING	REQUIRED	Location of featureCounts.tsv file in GCS.
star_metrics	participant_id	STRING	REQUIRED	The Participant ID.
star_metrics	sample_id	STRING	REQUIRED	The Sample ID.
star_metrics	num_GCAG_splices	INTEGER	REQUIRED	Number of splices: GC/AG
star_metrics	num_splices	INTEGER	REQUIRED	Number of splices: Total
star_metrics	started_job_on	TIMESTAMP	REQUIRED	Started job on
star_metrics	uniquely_mapped_percent	FLOAT	REQUIRED	Uniquely mapped reads %
star_metrics	insertion_length	FLOAT	REQUIRED	Insertion average length
star_metrics	deletion_length	FLOAT	REQUIRED	Deletion average length
star_metrics	unmapped_tooshort_percent	FLOAT	REQUIRED	% of reads unmapped: too short
star_metrics	avg_mapped_read_length	FLOAT	REQUIRED	Average mapped length
star_metrics	deletion_rate	FLOAT	REQUIRED	Deletion rate per base
star_metrics	started_mapping_on	TIMESTAMP	REQUIRED	Started mapping on
star_metrics	mismatch_rate	FLOAT	REQUIRED	Mismatch rate per base, %
star_metrics	avg_input_read_length	INTEGER	REQUIRED	Average input read length
star_metrics	num_ATAC_splices	INTEGER	REQUIRED	Number of splices: AT/AC
star_metrics	num_annotated_splices	INTEGER	REQUIRED	Number of splices: Annotated (sjdb)
star_metrics	num_GTAG_splices	INTEGER	REQUIRED	Number of splices: GT/AG
star_metrics	uniquely_mapped	INTEGER	REQUIRED	Uniquely mapped reads number
star_metrics	multimapped_toomany	INTEGER	REQUIRED	Number of reads mapped to too many loci
star_metrics	unmapped_mismatches_percent	FLOAT	REQUIRED	% of reads unmapped: too many mismatches
star_metrics	mapping_speed	FLOAT	REQUIRED	Mapping speed, Million of reads per hour
star_metrics	total_reads	INTEGER	REQUIRED	Number of input reads
star_metrics	num_chimeric_reads	INTEGER	REQUIRED	Number of chimeric reads
star_metrics	insertion_rate	FLOAT	REQUIRED	Insertion rate per base
star_metrics	unmapped_other_percent	FLOAT	REQUIRED	% of reads unmapped: other
star_metrics	pct_chimeric_reads	FLOAT	REQUIRED	% of chimeric reads
star_metrics	multimapped_percent	FLOAT	REQUIRED	% of reads mapped to multiple loci
star_metrics	multimapped	INTEGER	REQUIRED	Number of reads mapped to multiple loci
star_metrics	num_noncanonical_splices	INTEGER	REQUIRED	Number of splices: Non-canonical
star_metrics	finished_on	TIMESTAMP	REQUIRED	Finished on
star_metrics	multimapped_toomany_percent	FLOAT	REQUIRED	% of reads mapped to too many loci
star_metrics	unmapped_mismatches	INTEGER	REQUIRED	Derived from multiqc.
star_metrics	unmapped_tooshort	INTEGER	REQUIRED	Derived from multiqc.
star_metrics	unmapped_other	INTEGER	REQUIRED	Derived from multiqc.

Accelerating Medicines Partnership: Parkinson's Disease (AMP PD)

Transcriptomics Data Dictionary

Table Name (Tier 2)	Column Name	Data Type	Required	Description
Clinically_Reported_Genetic_Status	participant_id	STRING	REQUIRED	Study Subject ID
Clinically_Reported_Genetic_Status	GUID	STRING	NULLABLE	Global Unique ID (USUBJID)
Clinically_Reported_Genetic_Status	visit_name	STRING	REQUIRED	Visit name: M - in months, SC - screening visit, LOG - records without visit; #2 or #3 define repated records at the same visit or repeated visit
Clinically_Reported_Genetic_Status	visit_month	FLOAT	NULLABLE	Numeric visit in months; for visits prior baseline -1, -2 is an order of screening visits
Clinically_Reported_Genetic_Status	genetic_status_enrollment	STRING	NULLABLE	Participant genetic status based on PD-associated mutations (LRRK2, GBA, or SNCA) at enrollment
Clinically_Reported_Genetic_Status	genetic_status_wgs	STRING	NULLABLE	Participant genetic status based on the determined by Whole Genome Sequencing selected PD-associated variants (LRRK2 G2019S, LRRK2 R1441G, GBA N370S, and SNCA A53T)
amp_pd_participant_mutations	participant_id	STRING	null	null
amp_pd_participant_mutations	has_known_GBA_mutation_in_WGS	STRING	null	null
amp_pd_participant_mutations	has_known_LRRK2_mutation_in_WGS	STRING	null	null
amp_pd_participant_mutations	has_known_SNCA_mutation_in_WGS	STRING	null	null
amp_pd_participant_mutations	has_known_PD_Mutation_in_WGS	STRING	null	null