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

Genomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
gatk_all_variants	reference_name	STRING	null	Reference name.
gatk_all_variants	start_position	INTEGER	null	Start position (0-based). Corresponds to the first base of the string of reference bases.
gatk_all_variants	end_position	INTEGER	null	End position (0-based). Corresponds to the first base after the last base in the reference allele.
gatk_all_variants	reference_bases	STRING	null	Reference bases.
gatk_all_variants	alternate_bases	RECORD	REPEATED	One record for each alternate base (if any).
gatk_all_variants	names	STRING	REPEATED	Variant names (e.g. RefSNP ID).
gatk_all_variants	quality	FLOAT	null	Phred-scaled quality score (-10log10 prob(call is wrong)). Higher values imply better quality.
gatk_all_variants	filter	STRING	REPEATED	List of failed filters (if any) or PASS" indicating the variant has passed all filters."
gatk_all_variants	AN	INTEGER	null	Total number of alleles in called genotypes
gatk_all_variants	BaseQRankSum	FLOAT	null	Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities
gatk_all_variants	ClippingRankSum	FLOAT	null	Z-score From Wilcoxon rank sum test of Alt vs. Ref number of hard clipped bases
gatk_all_variants	DB	BOOLEAN	null	dbSNP Membership
gatk_all_variants	DP	INTEGER	null	Approximate read depth; some reads may have been filtered
gatk_all_variants	DS	BOOLEAN	null	Were any of the samples downsampled?
gatk_all_variants	ExcessHet	FLOAT	null	Phred-scaled p-value for exact test of excess heterozygosity
gatk_all_variants	FS	FLOAT	null	Phred-scaled p-value using Fisher's exact test to detect strand bias
gatk_all_variants	HaplotypeScore	FLOAT	null	Consistency of the site with at most two segregating haplotypes
gatk_all_variants	InbreedingCoeff	FLOAT	null	Inbreeding coefficient as estimated from the genotype likelihoods per-sample when compared against the Hardy-Weinberg expectation
gatk_all_variants	MQ	FLOAT	null	RMS Mapping Quality
gatk_all_variants	MQRankSum	FLOAT	null	Z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities
gatk_all_variants	NEGATIVE_TRAIN_SITE	BOOLEAN	null	This variant was used to build the negative training set of bad variants
gatk_all_variants	POSITIVE_TRAIN_SITE	BOOLEAN	null	This variant was used to build the positive training set of good variants
gatk_all_variants	QD	FLOAT	null	Variant Confidence/Quality by Depth
gatk_all_variants	RAW_MQ	FLOAT	null	Raw data for RMS Mapping Quality
gatk_all_variants	ReadPosRankSum	FLOAT	null	Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias
gatk_all_variants	SOR	FLOAT	null	Symmetric Odds Ratio of 2x2 contingency table to detect strand bias
gatk_all_variants	VQSLOD	FLOAT	null	Log odds of being a true variant versus being false under the trained gaussian mixture model
gatk_all_variants	culprit	STRING	null	The annotation which was the worst performing in the Gaussian mixture model, likely the reason why the variant was filtered out
gatk_all_variants	partition_date_please_ignore	DATE	null	This field is used for BigQuery clustering and contains no useful information
gatk_all_variants	hom_ref_call	STRING	REPEATED	Name of the call for genotype 0/0
gatk_all_variants	no_call	STRING	REPEATED	Name of the call for genotype -1/-1
gatk_all_variants	call	RECORD	REPEATED	One record for each call.
gatk_passing_variants	reference_name	STRING	null	Reference name.
gatk_passing_variants	start_position	INTEGER	null	Start position (0-based). Corresponds to the first base of the string of reference bases.
gatk_passing_variants	end_position	INTEGER	null	End position (0-based). Corresponds to the first base after the last base in the reference allele.
gatk_passing_variants	reference_bases	STRING	null	Reference bases.
gatk_passing_variants	alternate_bases	RECORD	REPEATED	One record for each alternate base (if any).
gatk_passing_variants	names	STRING	REPEATED	Variant names (e.g. RefSNP ID).
gatk_passing_variants	quality	FLOAT	null	Phred-scaled quality score (-10log10 prob(call is wrong)). Higher values imply better quality.
gatk_passing_variants	filter	STRING	REPEATED	List of failed filters (if any) or PASS" indicating the variant has passed all filters."
gatk_passing_variants	AN	INTEGER	null	Total number of alleles in called genotypes
gatk_passing_variants	BaseQRankSum	FLOAT	null	Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities
gatk_passing_variants	ClippingRankSum	FLOAT	null	Z-score From Wilcoxon rank sum test of Alt vs. Ref number of hard clipped bases
gatk_passing_variants	DB	BOOLEAN	null	dbSNP Membership
gatk_passing_variants	DP	INTEGER	null	Approximate read depth; some reads may have been filtered

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

Genomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
gatk_passing_variants	DS	BOOLEAN	null	Were any of the samples downsampled?
gatk_passing_variants	ExcessHet	FLOAT	null	Phred-scaled p-value for exact test of excess heterozygosity
gatk_passing_variants	FS	FLOAT	null	Phred-scaled p-value using Fisher's exact test to detect strand bias
gatk_passing_variants	HaplotypeScore	FLOAT	null	Consistency of the site with at most two segregating haplotypes
gatk_passing_variants	InbreedingCoeff	FLOAT	null	Inbreeding coefficient as estimated from the genotype likelihoods per-sample when compared against the Hardy-Weinberg expectation
gatk_passing_variants	MQ	FLOAT	null	RMS Mapping Quality
gatk_passing_variants	MQRankSum	FLOAT	null	Z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities
gatk_passing_variants	NEGATIVE_TRAIN_SITE	BOOLEAN	null	This variant was used to build the negative training set of bad variants
gatk_passing_variants	POSITIVE_TRAIN_SITE	BOOLEAN	null	This variant was used to build the positive training set of good variants
gatk_passing_variants	QD	FLOAT	null	Variant Confidence/Quality by Depth
gatk_passing_variants	RAW_MQ	FLOAT	null	Raw data for RMS Mapping Quality
gatk_passing_variants	ReadPosRankSum	FLOAT	null	Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias
gatk_passing_variants	SOR	FLOAT	null	Symmetric Odds Ratio of 2x2 contingency table to detect strand bias
gatk_passing_variants	VQSLOD	FLOAT	null	Log odds of being a true variant versus being false under the trained gaussian mixture model
gatk_passing_variants	culprit	STRING	null	The annotation which was the worst performing in the Gaussian mixture model, likely the reason why the variant was filtered out
gatk_passing_variants	partition_date_please_ignore	DATE	null	This field is used for BigQuery clustering and contains no useful information
gatk_passing_variants	hom_ref_call	STRING	REPEATED	Name of the call for genotype 0/0
gatk_passing_variants	no_call	STRING	REPEATED	Name of the call for genotype -1/-1
gatk_passing_variants	call	RECORD	REPEATED	One record for each call.
gatk_variant_calling_detail_metrics	participant_id	STRING	null	The Participant ID.
gatk_variant_calling_detail_metrics	sample_id	STRING	null	The Sample ID.
gatk_variant_calling_detail_metrics	HET_HOMVAR_RATIO	FLOAT	null	(count of hets)/(count of homozygous non-ref) for this sample
gatk_variant_calling_detail_metrics	PCT_GQ0_VARIANTS	FLOAT	null	The percentage of variants in a particular sample that have a GQ score of 0.
gatk_variant_calling_detail_metrics	TOTAL_GQ0_VARIANTS	INTEGER	null	The total number of variants in a particular sample that have a GQ score of 0.
gatk_variant_calling_detail_metrics	TOTAL_HET_DEPTH	INTEGER	null	Total number of reads (from AD field) for passing bi-allelic SNP hets for this sample.
gatk_variant_calling_detail_metrics	TOTAL_SNPS	INTEGER	null	The number of high confidence SNPs calls (i.e. non reference genotypes) that were examined.
gatk_variant_calling_detail_metrics	NUM_IN_DB_SNP	INTEGER	null	The number of high confidence SNPs found in dbSNP
gatk_variant_calling_detail_metrics	NOVEL_SNPS	INTEGER	null	The number of high confidence SNPS called that were not found in dbSNP
gatk_variant_calling_detail_metrics	FILTERED_SNPS	INTEGER	null	The number of SNPs that are also filtered
gatk_variant_calling_detail_metrics	PCT_DBSNP	FLOAT	null	The percentage of high confidence SNPs in dbSNP
gatk_variant_calling_detail_metrics	DBSNP_TITV	FLOAT	null	The Transition/Transversion ratio of the SNP calls made at dbSNP sites.
gatk_variant_calling_detail_metrics	NOVEL_TITV	FLOAT	null	The Transition/Transversion ratio of the SNP calls made at non-dbSNP sites.
gatk_variant_calling_detail_metrics	TOTAL_INDELS	INTEGER	null	The number of high confidence Indel calls that were examined
gatk_variant_calling_detail_metrics	NOVEL_INDELS	INTEGER	null	The number of high confidence Indels called that were not found in dbSNP
gatk_variant_calling_detail_metrics	FILTERED_INDELS	INTEGER	null	The number of indels that are also filtered
gatk_variant_calling_detail_metrics	PCT_DBSNP_INDELS	FLOAT	null	The percentage of high confidence Indels in dbSNP
gatk_variant_calling_detail_metrics	NUM_IN_DB_SNP_INDELS	INTEGER	null	The number of high confidence Indels found in dbSNP
gatk_variant_calling_detail_metrics	DBSNP_INS_DEL_RATIO	FLOAT	null	The Insertion/Deletion ratio of the Indel calls made at dbSNP sites
gatk_variant_calling_detail_metrics	NOVEL_INS_DEL_RATIO	FLOAT	null	The Insertion/Deletion ratio of the Indel calls made at non-dbSNP sites
gatk_variant_calling_detail_metrics	TOTAL_MULTIALLELIC_SNPS	INTEGER	null	The number of high confidence multiallelic SNP calls that were examined.
gatk_variant_calling_detail_metrics	NUM_IN_DB_SNP_MULTIALLELIC	INTEGER	null	The number of high confidence multiallelic SNPs found in dbSNP
gatk_variant_calling_detail_metrics	TOTAL_COMPLEX_INDELS	INTEGER	null	The number of high confidence complex Indel calls that were examined
gatk_variant_calling_detail_metrics	NUM_IN_DB_SNP_COMPLEX_IN	INTEGER	null	The number of high confidence complex Indels found in dbSNP
gatk_variant_calling_detail_metrics	SNP_REFERENCE_BIAS	FLOAT	null	The rate at which reference bases are observed at ref/alt heterozygous SNP sites.

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

Genomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
gatk_variant_calling_detail_metrics	NUM_SINGLETONS	FLOAT	null	For summary metrics, the number of variants that appear in only one sample. For detail metrics, the number of variants that appear only in the current sample.
gatk_variant_calling_summary_metrics	participant_id	STRING	null	The Participant ID.
gatk_variant_calling_summary_metrics	sample_id	STRING	null	The Sample ID.
gatk_variant_calling_summary_metrics	TOTAL_SNPS	INTEGER	null	The number of high confidence SNPs calls (i.e. non reference genotypes) that were examined.
gatk_variant_calling_summary_metrics	NUM_IN_DB_SNP	INTEGER	null	The number of high confidence SNPs found in dbSNP
gatk_variant_calling_summary_metrics	NOVEL_SNPS	INTEGER	null	The number of high confidence SNPS called that were not found in dbSNP
gatk_variant_calling_summary_metrics	FILTERED_SNPS	INTEGER	null	The number of SNPs that are also filtered
gatk_variant_calling_summary_metrics	PCT_DBSNP	FLOAT	null	The percentage of high confidence SNPs in dbSNP
gatk_variant_calling_summary_metrics	DBSNP_TITV	FLOAT	null	The Transition/Transversion ratio of the SNP calls made at dbSNP sites.
gatk_variant_calling_summary_metrics	NOVEL_TITV	FLOAT	null	The Transition/Transversion ratio of the SNP calls made at non-dbSNP sites.
gatk_variant_calling_summary_metrics	TOTAL_INDELS	INTEGER	null	The number of high confidence Indel calls that were examined
gatk_variant_calling_summary_metrics	NOVEL_INDELS	INTEGER	null	The number of high confidence Indels called that were not found in dbSNP
gatk_variant_calling_summary_metrics	FILTERED_INDELS	INTEGER	null	The number of indels that are also filtered
gatk_variant_calling_summary_metrics	PCT_DBSNP_INDELS	FLOAT	null	The percentage of high confidence Indels in dbSNP
gatk_variant_calling_summary_metrics	NUM_IN_DB_SNP_INDELS	INTEGER	null	The number of high confidence Indels found in dbSNP
gatk_variant_calling_summary_metrics	DBSNP_INS_DEL_RATIO	FLOAT	null	The Insertion/Deletion ratio of the Indel calls made at dbSNP sites
gatk_variant_calling_summary_metrics	NOVEL_INS_DEL_RATIO	FLOAT	null	The Insertion/Deletion ratio of the Indel calls made at non-dbSNP sites
gatk_variant_calling_summary_metrics	TOTAL_MULTIALLELIC_SNPS	INTEGER	null	The number of high confidence multiallelic SNP calls that were examined.
gatk_variant_calling_summary_metrics	NUM_IN_DB_SNP_MULTIALLELIC	INTEGER	null	The number of high confidence multiallelic SNPs found in dbSNP
gatk_variant_calling_summary_metrics	TOTAL_COMPLEX_INDELS	INTEGER	null	The number of high confidence complex Indel calls that were examined
gatk_variant_calling_summary_metrics	NUM_IN_DB_SNP_COMPLEX_INI	INTEGER	null	The number of high confidence complex Indels found in dbSNP
gatk_variant_calling_summary_metrics	SNP_REFERENCE_BIAS	FLOAT	null	The rate at which reference bases are observed at ref/alt heterozygous SNP sites.
gatk_variant_calling_summary_metrics	NUM_SINGLETONS	FLOAT	null	For summary metrics, the number of variants that appear in only one sample. For detail metrics, the number of variants that appear only in the current sample.
preBqsr_selfSM	participant_id	STRING	null	The Participant ID.
preBqsr_selfSM	sample_id	STRING	null	The Sample ID.

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

Genomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
preBqsr_selfSM	SEQ_ID	STRING	null	SAMPLE ID in the sequence file
preBqsr_selfSM	RG	STRING	null	ReadGroup ID of sequenced lane. For [outPrefix].selfSM and [outPrefix].bestSM, these values are ALL""
preBqsr_selfSM	CHIP_ID	STRING	null	Sample ID compared to in the genotype file. For [outPrefix].selfRG and [outPrefix].selfSM, these values should be identical to [SEQ_SM] or NA" if the genotype of sequenced samples are unavailable. For [outPrefix].bestRG and [outPrefix].bestSM
preBqsr_selfSM	SNPS	INTEGER	null	# of SNPs passing the criteria from the VCF file
preBqsr_selfSM	READS	INTEGER	null	Total # of reads loaded from the BAM file
preBqsr_selfSM	AVG_DP	FLOAT	null	Average sequencing depth at the sites in the VCF file
preBqsr_selfSM	FREEMIX	FLOAT	null	Sequence-only estimate of contamination (0-1 scale)
preBqsr_selfSM	FREELK1	FLOAT	null	Maximum log-likelihood of the sequence reads given estimated contamination under sequence-only method
preBqsr_selfSM	FREELK0	FLOAT	null	Log-likelihood of the sequence reads given no contamination under sequence-only method
preBqsr_selfSM	FREE_RH	FLOAT	null	Estimated reference bias parameter Pr(refBase|HET) (when --free-refBias or --free-full is used)
preBqsr_selfSM	FREE_RA	FLOAT	null	Estimated reference bias parameter Pr(refBase|HOMALT) (when --free-refBias or --free-full is used)
preBqsr_selfSM	CHIPMIX	FLOAT	null	Sequence+array estimate of contamination (NA if the external genotype is unavailable) (0-1 scale)
preBqsr_selfSM	CHIPLK1	FLOAT	null	Maximum log-likelihood of the sequence reads given estimated contamination under sequence+array method (NA if the external genotypes are unavailable)
preBqsr_selfSM	CHIPLK0	FLOAT	null	Log-likelihood of the sequence reads given no contamination under sequence+array method (NA if the external genotypes are unavailable)
preBqsr_selfSM	CHIP_RH	FLOAT	null	Estimated reference bias parameter Pr(refBase|HET) (when --chip-refBias or --chip-full is used)
preBqsr_selfSM	CHIP_RA	FLOAT	null	Estimated reference bias parameter Pr(refBase|HOMALT) (when --chip-refBias or --chip-full is used)
preBqsr_selfSM	DPREF	FLOAT	null	Depth (Coverage) of HomRef site (based on the genotypes of (SELF_SM/BEST_SM), passing mapQ, baseQual, maxDepth thresholds.
preBqsr_selfSM	RDPHET	FLOAT	null	DPHET/DPREF, Relative depth to HomRef site at Heterozygous site.
preBqsr_selfSM	RDPALT	FLOAT	null	DPHET/DPREF, Relative depth to HomRef site at HomAlt site.
raw_wgs_metrics	participant_id	STRING	null	The Participant ID.
raw_wgs_metrics	sample_id	STRING	null	The Sample ID.
raw_wgs_metrics	GENOME_TERRITORY	INTEGER	null	The number of non-N bases in the genome reference over which coverage will be evaluated.
raw_wgs_metrics	MEAN_COVERAGE	FLOAT	null	The mean coverage in bases of the genome territory, after all filters are applied.
raw_wgs_metrics	SD_COVERAGE	FLOAT	null	The standard deviation of coverage of the genome after all filters are applied.
raw_wgs_metrics	MEDIAN_COVERAGE	INTEGER	null	The median coverage in bases of the genome territory, after all filters are applied.
raw_wgs_metrics	MAD_COVERAGE	INTEGER	null	The median absolute deviation of coverage of the genome after all filters are applied.
raw_wgs_metrics	PCT_EXC_MAPQ	FLOAT	null	The fraction of aligned bases that were filtered out because they were in reads with low mapping quality (default is < 20).
raw_wgs_metrics	PCT_EXC_DUPE	FLOAT	null	The fraction of aligned bases that were filtered out because they were in reads marked as duplicates.
raw_wgs_metrics	PCT_EXC_UNPAIRED	FLOAT	null	The fraction of aligned bases that were filtered out because they were in reads without a mapped mate pair.
raw_wgs_metrics	PCT_EXC_BASEQ	FLOAT	null	The fraction of aligned bases that were filtered out because they were of low base quality (default is < 20).
raw_wgs_metrics	PCT_EXC_OVERLAP	FLOAT	null	The fraction of aligned bases that were filtered out because they were the second observation from an insert with overlapping reads.
raw_wgs_metrics	PCT_EXC_CAPPED	FLOAT	null	The fraction of aligned bases that were filtered out because they would have raised coverage above the capped value (default cap = 250x).
raw_wgs_metrics	PCT_EXC_TOTAL	FLOAT	null	The total fraction of aligned bases excluded due to all filters.
raw_wgs_metrics	PCT_1X	FLOAT	null	The fraction of bases that attained at least 1X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_5X	FLOAT	null	The fraction of bases that attained at least 5X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_10X	FLOAT	null	The fraction of bases that attained at least 10X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_15X	FLOAT	null	The fraction of bases that attained at least 15X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_20X	FLOAT	null	The fraction of bases that attained at least 20X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_25X	FLOAT	null	The fraction of bases that attained at least 25X sequence coverage in post-filtering bases.

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

Genomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
raw_wgs_metrics	PCT_30X	FLOAT	null	The fraction of bases that attained at least 30X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_40X	FLOAT	null	The fraction of bases that attained at least 40X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_50X	FLOAT	null	The fraction of bases that attained at least 50X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_60X	FLOAT	null	The fraction of bases that attained at least 60X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_70X	FLOAT	null	The fraction of bases that attained at least 70X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_80X	FLOAT	null	The fraction of bases that attained at least 80X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_90X	FLOAT	null	The fraction of bases that attained at least 90X sequence coverage in post-filtering bases.
raw_wgs_metrics	PCT_100X	FLOAT	null	The fraction of bases that attained at least 100X sequence coverage in post-filtering bases.
raw_wgs_metrics	HET_SNP_SENSITIVITY	FLOAT	null	The theoretical HET SNP sensitivity.
raw_wgs_metrics	HET_SNP_Q	INTEGER	null	The Phred Scaled Q Score of the theoretical HET SNP sensitivity.
wgs_metrics	participant_id	STRING	null	The Participant ID.
wgs_metrics	sample_id	STRING	null	The Sample ID.
wgs_metrics	GENOME_TERRITORY	INTEGER	null	The number of non-N bases in the genome reference over which coverage will be evaluated.
wgs_metrics	MEAN_COVERAGE	FLOAT	null	The mean coverage in bases of the genome territory, after all filters are applied.
wgs_metrics	SD_COVERAGE	FLOAT	null	The standard deviation of coverage of the genome after all filters are applied.
wgs_metrics	MEDIAN_COVERAGE	INTEGER	null	The median coverage in bases of the genome territory, after all filters are applied.
wgs_metrics	MAD_COVERAGE	INTEGER	null	The median absolute deviation of coverage of the genome after all filters are applied.
wgs_metrics	PCT_EXC_MAPQ	FLOAT	null	The fraction of aligned bases that were filtered out because they were in reads with low mapping quality (default is < 20).
wgs_metrics	PCT_EXC_DUPE	FLOAT	null	The fraction of aligned bases that were filtered out because they were in reads marked as duplicates.
wgs_metrics	PCT_EXC_UNPAIRED	FLOAT	null	The fraction of aligned bases that were filtered out because they were in reads without a mapped mate pair.
wgs_metrics	PCT_EXC_BASEQ	FLOAT	null	The fraction of aligned bases that were filtered out because they were of low base quality (default is < 20).
wgs_metrics	PCT_EXC_OVERLAP	FLOAT	null	The fraction of aligned bases that were filtered out because they were the second observation from an insert with overlapping reads.
wgs_metrics	PCT_EXC_CAPPED	FLOAT	null	The fraction of aligned bases that were filtered out because they would have raised coverage above the capped value (default cap = 250x).
wgs_metrics	PCT_EXC_TOTAL	FLOAT	null	The total fraction of aligned bases excluded due to all filters.
wgs_metrics	PCT_1X	FLOAT	null	The fraction of bases that attained at least 1X sequence coverage in post-filtering bases.
wgs_metrics	PCT_5X	FLOAT	null	The fraction of bases that attained at least 5X sequence coverage in post-filtering bases.
wgs_metrics	PCT_10X	FLOAT	null	The fraction of bases that attained at least 10X sequence coverage in post-filtering bases.
wgs_metrics	PCT_15X	FLOAT	null	The fraction of bases that attained at least 15X sequence coverage in post-filtering bases.
wgs_metrics	PCT_20X	FLOAT	null	The fraction of bases that attained at least 20X sequence coverage in post-filtering bases.
wgs_metrics	PCT_25X	FLOAT	null	The fraction of bases that attained at least 25X sequence coverage in post-filtering bases.
wgs_metrics	PCT_30X	FLOAT	null	The fraction of bases that attained at least 30X sequence coverage in post-filtering bases.
wgs_metrics	PCT_40X	FLOAT	null	The fraction of bases that attained at least 40X sequence coverage in post-filtering bases.
wgs_metrics	PCT_50X	FLOAT	null	The fraction of bases that attained at least 50X sequence coverage in post-filtering bases.
wgs_metrics	PCT_60X	FLOAT	null	The fraction of bases that attained at least 60X sequence coverage in post-filtering bases.
wgs_metrics	PCT_70X	FLOAT	null	The fraction of bases that attained at least 70X sequence coverage in post-filtering bases.
wgs_metrics	PCT_80X	FLOAT	null	The fraction of bases that attained at least 80X sequence coverage in post-filtering bases.
wgs_metrics	PCT_90X	FLOAT	null	The fraction of bases that attained at least 90X sequence coverage in post-filtering bases.
wgs_metrics	PCT_100X	FLOAT	null	The fraction of bases that attained at least 100X sequence coverage in post-filtering bases.
wgs_metrics	HET_SNP_SENSITIVITY	FLOAT	null	The theoretical HET SNP sensitivity.
wgs_metrics	HET_SNP_Q	INTEGER	null	The Phred Scaled Q Score of the theoretical HET SNP sensitivity.
wgs_samples	participant_id	STRING	null	The Participant ID.
wgs_samples	sample_id	STRING	null	The Sample ID.
wgs_samples	CRAM	STRING	null	Location of the .cram file.
wgs_samples	CRAI	STRING	null	Location of the .crai file.

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

Genomics Data Dictionary

Table Name	Column Name	Data Type	Required	Description
wgs_samples	CRAM_MD5	STRING	null	Location of the .cram.md5 fille.
wgs_samples	VCF	STRING	null	Location of the .vcf.gz file.
wgs_samples	VCF_TBI	STRING	null	Location of the .vcf.gz.tbi file.
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_V	STRING	null	null
amp_pd_participant_mutations	has_known_LRRK2_mutation_in_	STRING	null	null
amp_pd_participant_mutations	has_known_SNCA_mutation_in_	STRING	null	null
amp_pd_participant_mutations	has_known_PD_Mutation_in_W	STRING	null	null