Sample Data

Overview

The variants query API allows users to apply filters and retrieve variants from the Variant Storage. This functionality is useful to discover variants that matches some patterns, and to make variant discovery. The API also allows to paginate in order to fetch large lists of variants.

But in case of having studies with large number of samples, it may result difficult to get information of all samples from a given variant.

The API described here fetches the sample data from a variant.

Query Params

variant

The variant to query. If the variant does not exist in the study, an empty response will be returned.

Example

```
1:24534245:A:T
```

study

Selects the study of the samples. This can be empty if only one study exists.

genotypes

Fiter samples by genotype. Only samples with any of these genotypes will be included in the query result.

Defaul value

```
0/1,1/1
```

merge

Return the sample data all together under the key "**all**". Do not group by genotype.

Default value

```
false
```

Data model

- **id** : Variant ID
- **studyId** : study identifier
- **samples** : Map from genotype to list of **SampleData** objects
  - genotype
    - **id** : sample identifier
    - **sampleData** : Map Key-Value with the FORMAT fields from the VCF
    - **fileId** : Source file identifier
- **files** : Map from fileId to **FileEntry** objects
  - fileId
    - **fileId** : File identifier
    - **call** : Original call of the variant in the input VCF file
    - **attributes** : Map Key-Value with the INFO fields from the VCF
- **stats** : Map from cohortId to **VariantStats** objects
  - cohortId
    - **alleleCount**
    - **refAlleleCount**
    - **altAlleleCount**
    - **refAlleleFreq**
    - **altAlleleFreq**
    - **genotypeCount**
    - **genotypeFreq**
    - **missingAlleleCount**
    - **missingGenotypeCount**
    - **maf**
    - **mgf**

- **mafAllele**
- **mgfGenotype**

Example


```
{
 "id" : "1:14907:A:G",
 "studyId" : "S_1",
 "samples" : {
 "0/1" : [ {
 "id" : "NA12877",
 "sampleData" : {
 "VF" : "0.616",
 "AD" : "88,141",
 "MQ" : "26",
 "GQ" : "99",
 "DP" : "229",
 "PL" : "2929,0,1104",
 "GT" : "0/1",
 "GQX" : "99"
 },
 "fileId" : "1K.end.platinum-genomes-vcf-NA12877_S1.genome.vcf.gz"
 }, {
 "id" : "NA12879",
 "sampleData" : {
 "VF" : "0.652",
 "AD" : "54,101",
 "MQ" : "24",
 "GQ" : "99",
 "DP" : "155",
 "PL" : "1965,0,692",
 "GT" : "0/1",
 "GQX" : "99"
 },
 "fileId" : "1K.end.platinum-genomes-vcf-NA12879_S1.genome.vcf.gz"
 }, {
 "id" : "NA12880",
 "sampleData" : {
 "VF" : "0.765",
 "AD" : "24,78",
 "MQ" : "26",
 "GQ" : "99",
 "DP" : "102",returned
 "PL" : "1737,0,259",
 "GT" : "0/1",
 "GQX" : "99"
 },
 "fileId" : "1K.end.platinum-genomes-vcf-NA12880_S1.genome.vcf.gz"
 } ],
 "1/1" : [ {
 "id" : "NA12878",
 "sampleData" : {
 "VF" : "0.912",returned
 "AD" : "10,104",
 "MQ" : "25",
 "GQ" : "99",
 "DP" : "114",
 "PL" : "2393,204,0",
 "GT" : "1/1",
 "GQX" : "99"
 },
 "fileId" : "1K.end.platinum-genomes-vcf-NA12878_S1.genome.vcf.gz"
 } ]
 },
 "files" : {
 "1K.end.platinum-genomes-vcf-NA12879_S1.genome.vcf.gz" : {
 "fileId" : "1K.end.platinum-genomes-vcf-NA12879_S1.genome.vcf.gz",
 "call" : null,
 "attributes" : {
```

```
"AC" : "1",
"HRun" : "1",
"MQRankSum" : "2.716",
"set" : "FilteredInAll",
"FILTER" : "TruthSensitivityTranche99.90to100.00",
"MQ" : "24",
"AF" : "0.5",
"Dels" : "0.0",
"HaplotypeScore" : "1.6006",
"BaseQRankSum" : "-3.109",
"QUAL" : "1934.62",
"DP" : "155",
"ReadPosRankSum" : "0.231",
"AN" : "2",
"FS" : "3.756",
"MQ0" : "36",
"SB" : "-976.36",
"culprit" : "DP",
"QD" : "12.48",
"VQSLOD" : "-14.8451"
}
},
"1K.end.platinum-genomes-vcf-NA12880_S1.genome.vcf.gz" : {
"fileId" : "1K.end.platinum-genomes-vcf-NA12880_S1.genome.vcf.gz",
"call" : null,
"attributes" : {
"AC" : "1",
"HRun" : "1",
"MQRankSum" : "0.842",
"set" : "FilteredInAll",
"FILTER" : "TruthSensitivityTranche99.90to100.00",
"MQ" : "26",
"AF" : "0.5",
"Dels" : "0.0",
"HaplotypeScore" : "0.0",
"BaseQRankSum" : "-1.074",
"QUAL" : "1706.81",
"DP" : "102",
"ReadPosRankSum" : "-0.197",
"AN" : "2",
"FS" : "4.133",
"MQ0" : "15",
"SB" : "-809.94",
"culprit" : "MQ",
"QD" : "16.73",
"VQSLOD" : "-5.6284"
}
},
"1K.end.platinum-genomes-vcf-NA12877_S1.genome.vcf.gz" : {
"fileId" : "1K.end.platinum-genomes-vcf-NA12877_S1.genome.vcf.gz",
"call" : null,
"attributes" : {
"AC" : "1",
"HRun" : "1",
"MQRankSum" : "2.484",
"set" : "FilteredInAll",
"FILTER" : "TruthSensitivityTranche99.90to100.00",
"MQ" : "26",
"AF" : "0.5",
"Dels" : "0.0",
"HaplotypeScore" : "1.7128",
"BaseQRankSum" : "-0.15",
"QUAL" : "2898.73",
"DP" : "229",
"ReadPosRankSum" : "0.664",
"AN" : "2",
"FS" : "0.632",
"MQ0" : "47",
"SB" : "-1283.11",
"culprit" : "DP",
"QD" : "12.66",
```

```
 "VQSLOD" : "-18.504"
 }
 },
 "1K.end.platinum-genomes-vcf-NA12878_S1.genome.vcf.gz" : {
 "fileId" : "1K.end.platinum-genomes-vcf-NA12878_S1.genome.vcf.gz",
 "call" : null,
 "attributes" : {
 "AC" : "2",
 "HRun" : "1",
 "MQRankSum" : "2.075",
 "set" : "FilteredInAll",
 "FILTER" : "TruthSensitivityTranche99.90to100.00",
 "MQ" : "25",
 "AF" : "1.0",
 "Dels" : "0.0",
 "HaplotypeScore" : "0.0",
 "BaseQRankSum" : "-0.692",
 "QUAL" : "2359.96",
 "DP" : "114",
 "ReadPosRankSum" : "-0.395",
 "AN" : "2",
 "FS" : "4.601",
 "MQ0" : "32",
 "SB" : "-819.61",
 "culprit" : "DP",
 "QD" : "20.7",
 "VQSLOD" : "-9.8329"
 }
 }
 },
 "stats" : {
 "ALL" : {
 "alleleCount" : 8,
 "refAlleleCount" : 3,
 "genotypeCount" : {
 "0/1" : 3,
 "0/0" : 0,
 "1/1" : 1
 },
 "altAlleleCount" : 5,
 "genotypeFreq" : {
 "0/1" : 0.75,
 "0/0" : 0.0,
 "1/1" : 0.25
 },
 "missingAlleleCount" : 0,
 "missingGenotypeCount" : 0,
 "refAlleleFreq" : 0.375,
 "altAlleleFreq" : 0.625,
 "maf" : 0.375,
 "mgf" : 0.0,
 "mafAllele" : "A",
 "mgfGenotype" : "0/0"
 }
 }
}
```