Howdy GeniePool API users!

Accessing the API

REST API is available using:

http://api.geniepool.link/rest/index/$**reference**/$**coordinates**

$**reference** – Reference genomes for the chromosomal coordinates that follows. Valid inputs are hg38 or hg19.

$ **coordinates** - chromosomal coordinates according to the pattern:

chromosome:start-end

e.g. 1:12345789-123456798

Make sure that:

- Valid chromosome names are **1-22**, **X**, **Y**, **M** (but not MT) without “chr” prefix.
- For start/end, do not use commas (use 1234567 ✅ but not 1,234,567 ❎)
- When searching for a single coordinate, the “start” and “end” values should be the same, e.g. 1:123-123.
-

Examples:

http://api.geniepool.link/rest/index/hg38/7:117611791-117611791

http://api.geniepool.link/rest/index/hg38/Y:10000000-10001200

http://api.geniepool.link/rest/index/chm13v2/7:117611790-117621790

http://api.geniepool.link/rest/index/hg38/rs1045642

At the next page we will discuss how to interpret the output →

Understanding the output

The response is in JSON format, we will discuss an example that might initially look intimidating, but bear with us – everything will be understood shortly:

```
{"count":4,

"data":[

      {"pos":117611791,

      "entries":[{"ref":"T","alt":"C","impact":"Missense Ile>
      Thr","dbSNP":"rs751853765","hom":[],"het":[{"id":"SRR3938272","qual":94.64,"ad
":"8,3"}]},

      {"ref":"T","alt":"A","impact":"Missense Ile>
      Asn","hom":[],"het":[{"id":"SRR7944392","qual":367.64,"ad":"93,18"}]},

      {"ref":"T","alt":"TGC","impact":"Frameshift","hom":[],"het":[{"id":"SRR975212"
,"qual":55.6,"ad":"7,2"}]},

      {"ref":"T","alt":"TA","impact":"Frameshift","hom":[],"het":[{"id":"SRR8617690"
,"qual":103.6,"ad":"5,3"}]}]}

]

}
```

Now let us begin by breaking it to parts. We will use the “…” symbol to indicate properties that will be discussed in following steps, for visual convenience.

Let us begin →

```
{"count":4,"data":[…]}
```

The “count” property indicates how many different variants were found in total.

```
"data":[{"pos":117611791,"entries":[…]}]
```

The data property is an array of positions in which variants were found. Each item inside the array begins with a “pos” property indicating a chromosomal position, and an “entries” property with the variants themselves.

```
"entries":[
      {"ref":"T",
      "alt":"C",
      "impact":"Missense Ile>Thr",
      "dbSNP":"rs751853765",
      "hom":[…],
      "het":[{…]
      }
```

Each item the “entries” indicate a different variant. For each variant there are various attributes:

ref – The reference allele at that chromosomal position (“what was changed?”)

alt – The alternate allele in the variant (“to what it was changed to?”)

impact – Variant impact annotation by SnpEff.

dbSNP – This attribute will only appear if the variant can be found on dbSNP and will indicate its dbSNP ID.

Hom & **Het**: Information about the specific SRA samples in which the variant was found, separated by the zygosity (**Hom** for homozygotes and **Het** for heterozygotes).

```
"hom":[],

"het":[{"id":"SRR3938272","qual":94.64,"ad":"8,3"}]
```

In this example, no homozygote samples were found, however, one heterozygote sample was found. Each sample contains three attributes:

id – The SRA id of the sample.

qual – The sequencing quality phred score, taken from the sample's VCF file.

ad – The ad attribute from the VCF file, indicating two numbers: the first one (8 in our example) represents the number of alleles without the variant, and the second one (3 in our example) represents the number of alleles without the variant. To further clarify, the first out of the two numbers in homozygous variants will be 0.

Filtration

The data-lake will output a maximum of 100 variants (entries). To receive the "right" 100 variants, you can filter the results by providing additional arguments as numeric values:

?qual – Minimal quality score

?ad – Minimal sequencing coverage, the amount of reads that covered the region including the variant.

?am = Minimal AlphaMissense score. Please do take into account that by using this option, you will get only missense variants.

Example:

```
api.geniepool.link/rest/index/hg38/1:2345-6789?am=0.9?qual=100?ad=20
```

This query will return variants within the region of chromosome 1, position 2345 to 6789 (just for this example) with AlphaMissense score of at least 0.9, VCF QUAL value of 100 or more and at least read depth of 20.

Technical limitations

GeniePool can return up to 1,000 variants per query. In case more variants are present within the range, the number of variants will be given, but only the first 1,000 will be detailed.

In the unique event in which a dbSNP ID has more than one associated genomic coordinate, the one with the lowest chr:pos value will be used.

Thank you and see you on GitHub 😊

https://github.com/geniepool