

A BRIEF  
**TOUR**  
OF THE  
**ENSEMBL**  
REST API

<http://beta.rest.ensembl.org/>  
<https://github.com/acbb/EnsemblRest>



- GETful REST web service
- Currently in Beta

## What is REST?

- REpresentational State Transfer
- Uses methods GET, POST, PUT, DELETE
- URL is the key to the resource, method is the verb

## And GETful?

- Read-only resource
- Only uses the GET verb



## Functions:

- Information calls
  - Get known assemblies for a given species
  - Get more detailed information about a given assembly/species
- Comparative Genomics
  - Fetch Gene Tree dumps for a given Gene Tree id
  - Fetch homology information for an Ensembl gene id
  - Fetch homology information for a given symbol/species
- Lookup
  - Given an Ensembl id, fetch the object type, species, and database



- **Cross References**
  - Fetch db cross references for a given Ensembl id
  - Fetch external information for an external ID
  - Fetch Ensembl objects for a given external symbol/species
- **Variation**
  - Fetch Variant Consequences for a given range
  - Fetch Variant Consequences for a given variation ID
- **Features**
  - Find features in a given range/species



- Sequences
  - Fetch sequence for a given Ensembl ID
  - Fetch sequence for a given range/species
- Mappings
  - Convert co-ordinates between assemblies
  - Convert from CDNA co-ordinates to genomic
  - Convert from CDS co-ordinates to genomic
  - Convert from protein co-ordinates to genomic



## Things you currently can't do? (as of v69)

- Walk the graph (ie: Transcript -> Exons)
- Feature location, length, symbol, etc
- Access older versions of the API/database
- Make more than 3 calls per second (this will hopefully be lifted once the service is out of beta)



## Example of the EnsemblRest R package:

- Find all transcripts matching GO:0051800 (PTEN)
- Find the primary ids for these transcripts in HGNC

```
> trans = xrefsBySymbol( 'GO:0051800', 'homo_sapiens', object='transcript' )
> ids    = as.character( trans$id )
> hgnc  = xrefsById( ids, external_db='HGNC' )
> unique( as.character( hgnc$primary_id ) )
[1] "9588"
```

- Find the location of the first base of this Ensembl transcript

```
> mappingCdna( ids, '1..1', 'homo_sapiens' )
GRanges with 1 range and 2 elementMetadata cols:
      seqnames          ranges strand |          gap          rank
      <Rle>            <IRanges> <Rle> | <numeric> <numeric>
[1]          10 [89622870, 89622870]   + |           0           0
---
seqlengths:
  10
 NA
```



## Alternative Resources:

If you want pombe results (for example), these are hosted on the ensemblgenomes site rather than ensembl, so with the api you can change the underlying URL:

```
> setParam( url='http://beta.rest.ensemblgenomes.org/' )
> assemblyDetails( 'I', 'schizosaccharomyces_pombe' )
is_chromosome      : TRUE
length             : 5579133
assembly_exception_type : REF
coordinate_system  : chromosome
```





## Thanks!

- Still work to be done on the R package (vectorisation, etc)
- Source currently only on Github (pull requests/issues more than welcome!)  
<https://github.com/acbb/EnsemblRest>
- Plan to submit to bioconductor when the Ensembl REST API gets out of beta
- Ensembl REST documentation can be found here:  
<http://beta.rest.ensembl.org>

