# Supplementary Materials

|  |  |  |
| --- | --- | --- |
| Command | Input, Output | Description |
| **Transform Functions** |  |  |
| bior\_overlap | TJSON, TJSON | Extract annotations from a catalog based on genomic location overlap. The overlap is computed from the Start and End genomics position of a variant. |
| bior\_same\_variant | TJSON, TJSON | Extract annotations from a catalog based on variant position, reference and alternate allele definition. |
| bior\_lookup | TJSON, TJSON | Extract annotations from a catalog based on matching values of an identifier. |
| bior\_snpeff | TJSON, TJSON | Use SNPEffect1 to annotate variants. Chromosome ID, Start and Stop genomics position, reference and alternate allele of the variant is required . |
| bior\_vep | TJSON, TJSON | Use VEP2 to annotate variants. Chromosome ID, Start and Stop genomics position, reference and alternate allele of the variant is required. |
| bior\_drill | TJSON, TJSON | Extract an element from nested JSON string. |
| bior\_compress | TJSON, TJSON | Compress entries from provided set of identifiers into a single entry with each value separated by a delimiter. |
| **Utility Functions** |  |  |
| bior\_index\_catalog | identifier, index | Index the specified identifier in a catalog. Indices a stored in a separate index file. |
| bior\_create\_catalog | TJSON, catalog | Convert a text tabulated file into a catalog. Chromosome ID, Start and End genomics position fields have to be explicitly named. |
| bior\_ create\_catalog\_props | catalog, property | Create property files from the metadata extracted from a catalog. Property files are needs for proper metadata handling. |
| bior\_create\_config\_for\_tab\_to\_tjson | TSV,config | Create a configuration file that describes column description. This file is needed when uploading a tab delimited file. |
| **Input/Output Functions** |  |  |
| bior\_vcf\_to\_tjson | VCF, TJSON | Load a VCF file and convert to TJSON format. |
| bior\_tjson\_to\_vcf | TJSON, VCF | Convert TJSON to VCF format for file output. |
| bior\_bed\_to\_tjson | BED, TJSON | Load a BED file and convert to TJSON format. |
| bior\_tab\_to\_tjson | TSV, TJSON | Load a tab-delimited file and convert to TJSON format. |
| bior\_pretty\_print | TJSON, STDOUT | Convert TJSON in a readable format for screen or file output. |
| **Miscellaneous Functions** |  |  |
| bior\_annotate | VCF, TJSON | Append to the VCF ‘info’ field a set of commonly used annotations. |

Table S1: List of commands available in the BioR Toolkit. Detailed description and example is displayed when executing the command with the –h flag.   
1Cingolani, P. et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 6(2) :p. 80-92.  
2McLaren W et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. BMC Bioinformatics 26(16):2069-70

|  |  |  |
| --- | --- | --- |
| ‘Golden Identifier’ | Functions | Definition |
| \_landmark | bior\_overlap, bior\_same\_variant | Chromosome, or sequence ID where the interval is located |
| \_minBP | bior\_overlap, bior\_same\_variant | Minimum 1-based position (e.g. NCBI coordinates) on the landmark sequence |
| \_maxBP | bior\_overlap, bior\_same\_variant | Maximum 1-based position on the landmark sequence |
| \_refAllele | bior\_same\_variant | REF as in VCF standard |
| \_altAlleles | bior\_same\_variant | ALT as in VCF standard |

Table S2: List of indexed ‘golden identifiers’ that are used to accelerate coordinate-base searches and variant matching

|  |  |  |
| --- | --- | --- |
| Datasource | URL | Version |
| 1000Genomes | <http://www.1000genomes.org/category/ftp> | 20110521 |
| BGI | http://soap.genomics.org.cn/soapsnp.html | hg19 |
| COSMIC | http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/ | V63 |
| dbSNP | http://www.ncbi.nlm.nih.gov/snp/ | 137 |
| ESP6500 | https://esp.gs.washington.edu/drupal/ | build37 |
| HapMap | http://hapmap.ncbi.nlm.nih.gov | 2010-08\_phaseII+III |
| HGNC | http://www.genenames.org | 2012\_08\_12 |
| miRBase | http://www.mirbase.org | 8\_12\_12 |
| NCBIGene | http://www.ncbi.nlm.nih.gov/gene | GRCh37\_p10 |
| OMIM | http://www.omim.org | 2013\_02\_27 |
| PharmGKB | http://www.[pharmgkb.org/downloads/](http://www.pharmgkb.org/downloads/) | June 2013 |
| DrugBank | http://www.drugbank.ca/downloads | 3.0 |
| Therapeutic Target Database | http://bidd.nus.edu.sg/group/cjttd/TTD\_Download.asp | 4.3.02 |
| UCSC | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/  (note catalogs were created for each UCSC track) | hg19 |

Table S3: list of data sources from which BioR catalogs are derived. A description of the catalog is available at [http://bioinformaticstools.mayo.edu](http://bioinformaticstools.mayo.edu/)