# grasp Documentation

Release 0.3.0b1

**Michael Dacre** 

### Contents

I	bubmodules	
	.1 grasp.tables	
	.2 grasp.db	
	.3 grasp.config	
	.4 grasp.query	
	.5 grasp.info	
	.6 grasp.ref	. 1
2	Console Script	1
3	ndices and tables	1
Ру	on Module Index	1
In	$\mathbf{x}$	1

A Simple GRASP (grasp.nhlbi.nih.gov) API based on SQLAlchemy and Pandas.

Author	Michael D Dacre <mike.dacre@gmail.com></mike.dacre@gmail.com>
License	MIT License, made at Stanford, use as you wish.
Version	0.2.0b

For an introduction see the github readme

For table information see the wiki

Contents 1

2 Contents

### **Submodules**

# 1.1 grasp.tables

InMiRNA

GRASP table descriptions in SQLAlchemy ORM.

These tables do not exist in the GRASP data, which is a single flat file. By separating the data into these tables querying is much more efficient.

This submodule should only be used for querying.

```
class grasp.tables.SNP (**kwargs)
     Bases: sqlalchemy.ext.declarative.api.Base
     An SQLAlchemy Talble for GRASP SNPs.
     Study and phenotype information are pushed to other tables to minimize table size and make querying easier.
     Table Name: snps
     Columns: Described in the columns attribute
     int
         The ID number of the SNP, usually the NHLBIkey
     str
         SNP loction expressed as 'chr:pos'
     hvgs_ids
         A list of HGVS IDs for this SNP
     columns
         A dictionary of all columns 'column_name'=>('type', 'desc')
     ConservPredTFBS
     CreationDate
     Eqt1MethMetabStudy
     HUPfield
     HumanEnhancer
     InGene
     InLincRNA
```

```
InMiRNABS
LSSNP
LastCurationDate
NHLBIkey
NearestGene
ORegAnno
PolyPhen2
RNAedit
SIFT
UniProt
chrom
columns = OrderedDict([('id', ('BigInteger', 'NHLBIkey')), ('snpid', ('String', 'SNPid')), ('chrom', ('String', 'chr')), ('pring', 'snpid', ('String', 'snpid', ('snpid', ('String', 'snpid', ('snpid', ('snpi
                 A description of all columns in this table.
dbSNPClinStatus
dbSNPMAF
dbSNPfxn
dbSNPinfo
dbSNPvalidation
display_columns (display_as='table', write=False)
                Return all columns in the table nicely formatted.
                 Display choices: table: A formatted grid-like table tab: A tab delimited non-formatted version of table
                               list: A string list of column names
                               Parameters
                                             display_as - {table,tab,list}
                                             • write – If true, print output to console, otherwise return string.
                               Returns A formatted string or None
get_columns (return_as='list')
                Return all columns in the table nicely formatted.
```

Display choices: list: A python list of column names dictionary: A python dictionary of name=>desc long\_dict: A python dictionary of name=>(type, desc)

Parameters return\_as - {table,tab,list,dictionary,long\_dict,id\_dict}

**Returns** A list or dictionary

```
get_variant_info (fields='dbsnp', pandas=True)
```

Use the myvariant API to get info about this SNP.

Note that this service can be very slow. It will be faster to query multiple SNPs.

#### **Parameters**

• fields - Choose fields to display from: 'docs.myvariant.info/en/latest/doc/data.html#available-

fields'\_ Good choices are 'dbsnp', 'clinvar', or 'gwassnps' Can also use 'grasp' to get a

different version of this info.

```
• pandas – Return a dataframe instead of dictionary.
             Returns A dictionary or a dataframe.
     hvgs_ids
         The HVGS ID from myvariant.
     id
     paper_loc
     phenotype_cats
     phenotype_desc
     population
     population_id
     pos
     pval
     snp loc
         Return a simple string containing the SNP location.
     snpid
     study
     study_id
     study_snpid
class grasp.tables.Phenotype (**kwargs)
     Bases: sqlalchemy.ext.declarative.api.Base
     An SQLAlchemy table to store the primary phenotype.
     Table Name: phenos
     Columns: phenotype: The string phenotype from the GRASP DB, unique. alias: A short representation of the
          phenotype, not unique. studies: A link to the studies table.
     int
          The ID number.
          The name of the phenotype.
     alias
     id
     phenotype
     studies
class grasp.tables.PhenoCats(**kwargs)
     Bases: sqlalchemy.ext.declarative.api.Base
     An SQLAlchemy table to store the lists of phenotype categories.
     Table Name: pheno_cats
```

1.1. grasp.tables 5

```
not unique. snps: A link to all SNPs in this category. studies: A link to all studies in this category.
     int
          The PhenoCat ID
     str
          The category name
     alias
     category
     id
     snps
     studies
class grasp.tables.Platform(platform)
     Bases: sqlalchemy.ext.declarative.api.Base
     An SQLAlchemy table to store the platform information.
     Table Name: platforms
     Columns: platform: The name of the platform from GRASP. studies: A link to all studies using this platform.
     int
          The ID number of this platform
     str
          The name of the platform
     id
     platform
     studies
class grasp.tables.Population (population)
     Bases: sqlalchemy.ext.declarative.api.Base
     An SQLAlchemy table to store the platform information.
     Table Name: populations
     Columns: population: The name of the population. studies: A link to all studies in this population. snps: A
          link to all SNPs in this populations.
     int
          Population ID number
     str
          The name of the population
     id
     population
```

**Columns:** category: The category from the grasp database, unique. alias: An easy to use alias of the category,

# 1.2 grasp.db

Functions for managing the GRASP database.

get\_session() is used everywhere in the module to create a connection to the database. initialize\_database() is used to build the database from the GRASP file. It takes about an hour 90 minutes to run and will overwrite any existing database.

```
grasp.db.get_session(echo=False)
```

Return a session and engine, uses config file.

**Parameters** echo – Echo all SQL to the console.

#### Returns

A SQLAlchemy session and engine object corresponding to the grasp database for use in querying.

Return type session, engine

grasp.db.initialize\_database (study\_file, grasp\_file, commit\_every=250000, progress=False) Create the database quickly.

**Study\_file** Tab delimited GRASP study file, available here: github.com/MikeDacre/grasp/blob/master/grasp\_studies.txt

**Grasp\_file** Tab delimited GRASP file.

Commit\_every How many rows to go through before committing to disk.

**Progress** Display a progress bar (db length hard coded).

### 1.3 grasp.config

Manage a persistent configuration for the database.

```
grasp.config.config = <configparser.ConfigParser object>
```

A globally accessible ConfigParger object, initialized with CONFIG\_FILE.

```
grasp.config.CONFIG_FILE = '/Users/dacre/.grasp'
```

The PATH to the config file.

grasp.config.init\_config(db\_type, db\_file='', db\_host='', db\_user='', db\_pass='')
Create an initial config file.

### Parameters

- db\_type 'sqlite/mysql/postgresql'
- **db\_file** PATH to sqlite database file
- **db\_host** Hostname for mysql or postgresql server
- **db\_user** Username for mysql or postgresql server
- **db\_pass** Password for mysql or postgresql server (not secure)

Returns NoneType

Return type None

```
grasp.config.init_config_interactive()
```

Interact with the user to create a new config.

Uses readline autocompletion to make setup easier.

```
grasp.config.write config()
```

Write the current config to CONFIG\_FILE.

1.3. grasp.config 7

### 1.4 grasp.query

A mix of functions to make querying the database and analyzing the results faster.

Return a list of studies filtered by phenotype and population.

There are two ways to query both phenotype and population.

**Phenotype:** GRASP provides a 'primary phenotype' for each study, which are fairly poorly curated. They also provide a list of phenotype categories, which are well curated. The problem with the categories is that there are multiple per study and some are to general to be useful. If using categories be sure to post filter the study list.

Note: I have made a list of aliases for the phenotype categories to make them easier to type. Use pheno\_cats\_alias for that.

**Population:** Each study has a primary population (list available with 'get\_populations') but some studies also have other populations in the cohort. GRASP indexes all population counts, so those can be used to query also. To query these use *has\_* or *only\_* (exclusive) parameters, you can query either discovery populations or replication populations. Note that you cannot provide both *has\_* and *only\_* parameters for the same population type.

For doing population specific analyses most of the time you will want the excl\_disc\_pop query.

**Argument Description:** Phenotype Arguments are 'primary\_phenotype', 'pheno\_cats', and 'pheno\_cats\_alias'.

Only provide one of pheno\_cats or pheno\_cats\_alias

Population Arguments are primary\_pop, has\_disc\_pop, has\_rep\_pop, only\_disc\_pop, only\_rep\_pop.

primary pop is a simple argument, the others use bitwise flags for lookup.

The easiest way to use the following parameters is with the \_ref.PopFlag object. It uses py-flags. For example:

```
pops = _ref.PopFlag.eur | _ref.PopFlag.afr
```

In addition you can provide a list of strings correcponding to PopFlag attributes.

Note: the *only*\_ parameters work as ANDs, not ORs. So only\_disc\_pop='eurlafr' will return those studies that have BOTH european and african discovery populations, but no other discovery populations. On the other hand, *has*\_ works as an OR, and will return any study with any of the spefified populations.

#### **Parameters**

- **primary\_phenotype** Phenotype of interest, string or list of strings.
- pheno\_cats Phenotype category of interest.
- pheno\_cats\_alias Phenotype category of interest.
- primary\_pop Query the primary population, string or list of strings.
- has\_disc\_pop Return all studies with these discovery populations
- has\_rep\_pop Return all studies with these replication populations
- only\_disc\_pop Return all studies with ONLY these discovery populations

- only\_rep\_pop Return all studies with ONLY these replication populations
- query Return the query instead of the list of study objects.
- count Return a count of the number of studies.
- **dictionary** Return a dictionary of title->id for filtering.
- pandas Return a dataframe of study information instead of the list.

**Returns** A list of study objects, a query, or a dataframe.

```
grasp.query.get_snps (studies, pandas=True)
```

Return a list of SNPs in a single population in a single phenotype.

**Studies** A list of studies.

Pandas Return a dataframe instead of a list of SNP objects.

Returns Either a DataFrame or list of SNP objects.

grasp.query.get\_variant\_info (snp\_list, fields='dbsnp', pandas=True)
Get variant info for a list of SNPs.

#### **Parameters**

- snp\_list A list of SNP objects or SNP rsIDs
- fields Choose fields to display from: 'docs.myvariant.info/en/latest/doc/data.html#available-fields'\_ Good choices are 'dbsnp', 'clinvar', or 'gwassnps' Can also use 'grasp' to get a different version of this info.
- pandas Return a dataframe instead of dictionary.

**Returns** A dictionary or a dataframe.

Return a list of phenotypes that are present in all populations.

Can only provide one of primary\_pops or pop\_flags. pop\_flags does a bitwise lookup, primary\_pops quries the primary string only.

By default this function returns a list of phenotype categories, if you want to check primary phenotypes instead, provide check='primary'.

#### **Parameters**

- primary\_pops A string or list of strings corresponding to the tables. Study. phenotype column
- pop\_flags A ref.PopFlag object or list of objects.
- **check** cat/primary either check categories or primary phenos.
- **pop\_type** disc/rep Use with pop\_flags only, check either discovery or replication populations.
- exclusive Use with pop\_flags only, do an excusive rather than inclusion population search
- list\_only Return a list of names only, rather than a list of objects

**Returns** A list of *table.Phenotype* or *table.PhenoCat* objects, or a list of names if *list\_only* is specified.

1.4. grasp.query 9

```
\begin{tabular}{ll} $\tt grasp.query.collapse\_dataframe (\it df, mechanism='median', pvalue\_filter=None, protected\_columns=None) \end{tabular} $$pvalue\_filter=None, protected\_columns=None) $$
```

Collapse a dataframe by chrom:location from get\_snps.

Will use the mechanism defined by 'mechanism' to collapse a dataframe to one indexed by 'chrom:location' with pvalue and count only.

This function is agnostic to all dataframe columns other than:

```
['chrom', 'pos', 'snpid', 'pval']
```

All other columns are collapsed into a comma separated list, a string. 'chrom' and 'pos' are merged to become the new colon-separated index, snpid is maintained, and pval is merged using the function in 'mechanism'.

#### **Parameters**

- df A pandas dataframe, must have 'chrom', 'pos', 'snpid', and 'pval' columns.
- mechanism A numpy statistical function to use to collapse the pvalue, median or mean are the common ones.
- pvalue\_filter After collapsing the dataframe, filte to only include pvalues less than this cutoff.
- **protected\_columns** A list of column names that will be maintened as is, although all duplicates will be dropped (randomly). Only makes sense for columns that are identical for all studies of the same SNP.

#### Returns

**Indexed by chr:pos, contains flattened pvalue column, and** all original columns as a comma-separated list. Additionally contains a count and stddev (of pvalues) column. stddev is nan if count is 1.

Return type DataFrame

Plot all SNPs that overlap between two pvalue series.

### Parameters

- series A pandas series object
- names A list of two names to use for the resultant dataframes
- **stats** Print some stats on the intersection
- plot Plot the resulting intersection

**Returns** with the two series as columns

Return type DataFrame

# 1.5 grasp.info

Little functions to pretty print column lists and category info.

get\_{phenotypes,phenotype\_categories,popululations} all display a dump of the whole database.

get\_population\_flags displays available flags from PopFlag.

display {study,snp} columns displays a list of available columns in those two tables as a formatted string.

get\_{study,snp}\_columns return a list of available columns in those two tables as python objects.

grasp.info.display\_snp\_columns (display\_as='table', write=False)

Return all columns in the SNP table as a string.

**Display choices:** table: A formatted grid-like table tab: A tab delimited non-formatted version of table list: A string list of column names

#### **Parameters**

- display\_as {table,tab,list}
- write If true, print output to console, otherwise return string.

**Returns** A formatted string or None

grasp.info.display\_study\_columns (display\_as='table', write=False)

Return all columns in the Study table as a string.

**Display choices:** table: A formatted grid-like table tab: A tab delimited non-formatted version of table list: A string list of column names

#### **Parameters**

- display\_as {table,tab,list}
- write If true, print output to console, otherwise return string.

**Returns** A formatted string or None

grasp.info.get\_phenotype\_categories (list\_only=False, dictionary=False, table=False)
Return all phenotype categories from the PhenoCats table.

**List\_only** Return a simple text list instead of a list of Phenotype objects.

**Dictionary** Return a dictionary of phenotype=>ID

**Table** Return a pretty table for printing.

grasp.info.get\_phenotypes (list\_only=False, dictionary=False, table=False)

Return all phenotypes from the Phenotype table.

**List\_only** Return a simple text list instead of a list of Phenotype objects.

**Dictionary** Return a dictionary of phenotype=>ID

Table Return a pretty table for printing.

grasp.info.get population flags (list only=False, dictionary=False, table=False)

Return all population flags available in the PopFlags class.

**List\_only** Return a simple text list instead of a list of Phenotype objects.

**Dictionary** Return a dictionary of population=>ID

Table Return a pretty table for printing.

 $\verb|grasp.info.get_populations|| (\textit{list\_only=False}, \textit{dictionary=False}, \textit{table=False})|$ 

Return all populations from the Population table.

**List\_only** Return a simple text list instead of a list of Phenotype objects.

**Dictionary** Return a dictionary of population=>ID

**Table** Return a pretty table for printing.

1.5. grasp.info

```
grasp.info.get_snp_columns(return_as='list')
```

Return all columns in the SNP table.

**Display choices:** list: A python list of column names dictionary: A python dictionary of name=>desc long\_dict: A python dictionary of name=>(type, desc)

Parameters return\_as - {table,tab,list,dictionary,long\_dict,id\_dict}

**Returns** A list or dictionary

```
grasp.info.get_study_columns(return_as='list')
```

Return all columns in the SNP table.

**Display choices:** list: A python list of column names dictionary: A python dictionary of name=>desc long\_dict: A python dictionary of name=>(type, desc)

Parameters return\_as - {table,tab,list,dictionary,long\_dict,id\_dict}

**Returns** A list or dictionary

# 1.6 grasp.ref

*ref.py* holds some simple lookups and the *PopFlags* classes that don't really go anywhere else. Holds reference objects for use elsewhere in the module.

```
class grasp.ref.PopFlag
     Bases: flags.Flags
```

A simplified bitwise flag system for tracking populations.

# **Console Script**

### A Simple GRASP (grasp.nhlbi.nih.gov) API based on SQLAlchemy and Pandas

Author	Michael D Dacre <mike.dacre@gmail.com></mike.dacre@gmail.com>
Organization	Stanford University
License	MIT License, use as you wish
Created	2016-10-08
Version	0.3.0b1

Last modified: 2016-10-17 00:18

This is the front-end to a python grasp api, intended to allow easy database creation and simple querying. For most of the functions of this module, you will need to call the module directly.

```
usage: grasp [-h] {search,conf,info,init} ...
```

### **Sub-commands:**

search (s, lookup) Query database for variants by location or id

Query for SNPs in the database. By default returns a tab-delimeted list of SNPs with the following columns: 'id', 'snpid', 'study\_snpid', 'chrom', 'pos', 'phenotype', 'pval'

The -extra flag adds these columns: 'InGene', 'InMiRNA', 'inLincRNa', 'LSSNP'

The -study-info flag adds these columns: 'study\_id (PMID)', 'title'

The –db-snp flag uses the myvariant API to pull additional data from db\_snp.

#### **Positional arguments:**

query rsID, chrom:loc or chrom:start-end

**Options:** 

--extra Add some extra columns to output
 --study-info Include study title and PMID
 --db-snp Add dbSNP info to output
 --pandas Write output as a pandas dataframe
 -o, --out File to write to, default STDOUT.

--path

PATH to write files to

### conf (config) Manage local config

#### **Options:**

**--db** Set the current database platform.

Possible choices: sqlite, postgresql, mysql

--get-path Change the sqlite file path--set-path Change the sqlite file path

**--init** Initialize the config with default settings. Will ERASE your old

config!

#### info Display database info

Write data summaries (also found on the wiki) to a file or the console.

Choices: all: Will write everything to separate rst files, ignores all other flags except '-path' phenotypes: All primary phenotypes. Phenotype\_categories: All phenotype categories. populations: All primary populations. population\_flags: All population flags. snp\_columns: All SNP columns. study\_columns: All Study columns.

#### **Positional arguments:**

**display** Choice of item to display, if all, results are written to independent

rst files, and optional args are ignored

Possible choices: study\_columns, populations, population\_flags,

snp\_columns, phenotypes, all, phenotype\_categories

**Options:** 

**-0, --out** File to write to, default STDOUT.

**--path** PATH to write files to

#### init Initialize the database

```
usage: grasp init [-h] [-n] study_file grasp_file
```

#### **Positional arguments:**

study\_file GRASP study file from: github.com/MikeDacre/grasp/blob/master/grasp2\_studies.txt

**grasp\_file** GRASP tab delimeted file

**Options:** 

**-n, --no-progress** Do not display a progress bar

# CHAPTER 3

# Indices and tables

- genindex
- modindex
- search

Python Module Index

### g

```
grasp.config,7
grasp.db,6
grasp.info,10
grasp.query,8
grasp.ref,12
grasp.tables,3
```

18 Python Module Index

A	get_studies() (in module grasp.query), 8		
alias (grasp.tables.PhenoCats attribute), 6 alias (grasp.tables.Phenotype attribute), 5	get_study_columns() (in module grasp.info), 12 get_variant_info() (grasp.tables.SNP method), 4 get_variant_info() (in module grasp.query), 9		
C	grasp.config (module), 7		
category (grasp.tables.PhenoCats attribute), 6 chrom (grasp.tables.SNP attribute), 4 collapse_dataframe() (in module grasp.query), 9 columns (grasp.tables.SNP attribute), 3, 4 config (in module grasp.config), 7	grasp.db (module), 6 grasp.info (module), 10 grasp.query (module), 8 grasp.ref (module), 12 grasp.tables (module), 3		
CONFIG_FILE (in module grasp.config), 7	Н		
ConservPredTFBS (grasp.tables.SNP attribute), 3 CreationDate (grasp.tables.SNP attribute), 3	HumanEnhancer (grasp.tables.SNP attribute), 3 HUPfield (grasp.tables.SNP attribute), 3		
D	hvgs_ids (grasp.tables.SNP attribute), 3, 5		
dbSNPClinStatus (grasp.tables.SNP attribute), 4 dbSNPfxn (grasp.tables.SNP attribute), 4 dbSNPinfo (grasp.tables.SNP attribute), 4 dbSNPMAF (grasp.tables.SNP attribute), 4 dbSNPvalidation (grasp.tables.SNP attribute), 4 display_columns() (grasp.tables.SNP method), 4 display_snp_columns() (in module grasp.info), 11 display_study_columns() (in module grasp.info), 11  E EqtlMethMetabStudy (grasp.tables.SNP attribute), 3  F find_intersecting_phenotypes() (in module grasp.query), 9	id (grasp.tables.PhenoCats attribute), 6 id (grasp.tables.Phenotype attribute), 5 id (grasp.tables.Platform attribute), 6 id (grasp.tables.Population attribute), 6 id (grasp.tables.SNP attribute), 5 InGene (grasp.tables.SNP attribute), 3 init_config() (in module grasp.config), 7 init_config_interactive() (in module grasp.config), 7 initalize_database() (in module grasp.db), 7 InLincRNA (grasp.tables.SNP attribute), 3 InMiRNA (grasp.tables.SNP attribute), 3 InMiRNABS (grasp.tables.SNP attribute), 3 int (grasp.tables.PhenoCats attribute), 6 int (grasp.tables.Phenotype attribute), 5		
G	int (grasp.tables.Platform attribute), 6		
get_columns() (grasp.tables.SNP method), 4 get_phenotype_categories() (in module grasp.info), 11 get_phenotypes() (in module grasp.info), 11	int (grasp.tables.Population attribute), 6 int (grasp.tables.SNP attribute), 3 intersect_overlapping_series() (in module grasp.query) 10		
get_population_flags() (in module grasp.info), 11 get_populations() (in module grasp.info), 11	L		
get_session() (in module grasp.db), 7 get_snp_columns() (in module grasp.info), 11 get_snps() (in module grasp query), 9	LastCurationDate (grasp.tables.SNP attribute), 4 LSSNP (grasp.tables.SNP attribute), 4		

# Ν NearestGene (grasp.tables.SNP attribute), 4 NHLBIkey (grasp.tables.SNP attribute), 4 0 ORegAnno (grasp.tables.SNP attribute), 4 paper\_loc (grasp.tables.SNP attribute), 5 PhenoCats (class in grasp.tables), 5 Phenotype (class in grasp.tables), 5 phenotype (grasp.tables.Phenotype attribute), 5 phenotype\_cats (grasp.tables.SNP attribute), 5 phenotype\_desc (grasp.tables.SNP attribute), 5 Platform (class in grasp.tables), 6 platform (grasp.tables.Platform attribute), 6 PolyPhen2 (grasp.tables.SNP attribute), 4 PopFlag (class in grasp.ref), 12 Population (class in grasp.tables), 6 population (grasp.tables.Population attribute), 6 population (grasp.tables.SNP attribute), 5 population id (grasp.tables.SNP attribute), 5 pos (grasp.tables.SNP attribute), 5 pval (grasp.tables.SNP attribute), 5 R RNAedit (grasp.tables.SNP attribute), 4 S SIFT (grasp.tables.SNP attribute), 4 SNP (class in grasp.tables), 3 snp\_loc (grasp.tables.SNP attribute), 5 snpid (grasp.tables.SNP attribute), 5 snps (grasp.tables.PhenoCats attribute), 6 str (grasp.tables.PhenoCats attribute), 6 str (grasp.tables.Phenotype attribute), 5 str (grasp.tables.Platform attribute), 6 str (grasp.tables.Population attribute), 6 str (grasp.tables.SNP attribute), 3 studies (grasp.tables.PhenoCats attribute), 6 studies (grasp.tables.Phenotype attribute), 5 studies (grasp.tables.Platform attribute), 6 study (grasp.tables.SNP attribute), 5 study\_id (grasp.tables.SNP attribute), 5 study\_snpid (grasp.tables.SNP attribute), 5 U UniProt (grasp.tables.SNP attribute), 4 W

write\_config() (in module grasp.config), 7

20 Index