Snipper Documentation
Version 1.2
Author: Ryan Welch (welchr@umich.edu)

Contents

- Purpose
- Requirements
- Installation
  - Windows
  - Source (Linux/Unix)
- Using Snipper (command line)
  - Synopsis
  - Options
- Using Snipper (GUI)
- Output
  - Console output
  - HTML output
- Building your own position database
- Examples
- License

Purpose

Snipper is a research tool for investigating genes near associated loci from GWAS studies. The user can supply a SNP or list of SNPs, a list of genes, and/or a list of chromosomal regions, after which Snipper will:

- Create a gene list, where genes are added by:
  - For each SNP, find genes nearby up to a certain distance specified by the user, or those genes whose expression is known to be associated with the SNP
  - For each region (ex: chr9:1911-939393), find genes within or overlapping the region
  - Include each gene specified by the user
- Retrieve annotations for each gene from NCBI Entrez Gene, OMIM, and the Michigan Molecular Interactions database (MiMI)
- If the user supplies search terms: search PubMed for each combination of search term and gene
- Search annotation information on each gene for user's search terms
- Create an HTML (or console) report containing all of the available information for each gene (including where search terms matched and how often)

Snipper is designed to handle a modest number of loci (25-50), but has been tested to handle up to 100. Submitting a large number of SNPs beyond this is not recommended, as the program may require extreme amounts of time and memory. We have typically used a handful of search terms (less than 10) - using substantially more than this may also require very large amounts of time for Snipper to finish.

Installation

Installing on Windows (Vista/7)

To install under Windows, simply run the installer exe downloaded from our website. This will put Snipper under your appropriate \Program Files\ directory, and create a desktop icon to launch the GUI.

If you wish to use the command line instead of the GUI, you will need to:

1. Install Python 2.6 or greater (but **not** the 3.x branch). You can download this from http://www.python.org/.
2. Install Snipper from the setup executable downloaded from our website
3. Navigate to where Snipper was installed (usually something like C:\Program Files (x86)\Snipper\, the installer will tell you this)
4. Run the bin/setup_snipper.py script.
5. Use the command line script bin/snipper.py to run your queries.

Installing from source (Linux/Unix)

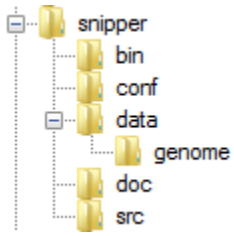
To install Snipper, you will need the following:

- Python version 2.6 or greater (but **not** the 3.x branch!)
- A working internet connection

To use the GUI for Snipper, you will also need the proper Tk package installed on your system. Under Ubuntu Linux, you can install it by using:

```
sudo apt-get install python-tk
```

1. Extract Snipper to the location of your choice. Note that you must extract the files into the directory structure given in the archive. It should follow this tree structure:



   You can do this by running the following command:

```
tar zxf snipper_tarball.tar.gz
```

2. Next, navigate to snipper/bin, and run the **setup_snipper.py** script. This script will install dependencies for Snipper, and ensure that they were installed correctly.

```
python setup_snipper.py
```

   You should first see dependencies being installed. It takes roughly a minute or so to install. The packages are installed to a virtual python environment under snipper/pyenv, and are not installed globally on your system (therefore, you do not need admin privileges.)

3. Snipper is ready to run, and can be launched by navigating to snipper/bin and running **snipper.py**. For simplicity, you could create a shortcut to this by doing the following:

```
ln –s snipper/bin/snipper.py /usr/local/bin/snipper
```

Using Snipper (command line)

Synopsis

Typical usage of Snipper will be of the form:

```
snipper --snpfile <file containing SNPs>
```

If one had a file containing SNPs, wanted to search 250kb away from each SNP for genes:

```
snipper --snpfile <file containing SNPs> -d 250kb
```

A user will generally want to include search terms with their query, for example:

```
snipper --snpfile <file containing SNPs> -d 250kb --terms "glucose,insulin"
```

Snipper can include genes explicitly requested by the user:

```
snipper --gene "TCF7L2,P53,BRCA1"
```

Or, the program can be run with chromosomal regions:

```
snipper --regions "chr#:start-end"
```

All of these can be mixed together, for example:

```
snipper -s "rs7903146,rs1002227" --regions "chr3:12393001-12475854" --gene
"RB1,PDE8B"
```

You can verify that Snipper has installed correctly by running our test example. Simply change directory into the example/ directory, and execute "run_example.py". This script runs a simple test using a few SNPs, genes, and chromosomal regions. The script will explain what it is doing, as well as showing the command line used to run Snipper. A directory called "example_results" will be created, containing the HTML output. There is also a "precompiled_results" directory, which gives the output if the program were to execute successfully. You can compare the two outputs to ensure that the program is working as expected (though we note that as databases and minor revisions to the HTML report format take place, the two may be slightly out of sync.)

Options

Snipper supports a wide variety of command line arguments for tailoring what and how much information is retrieved. Please see the table below for a full listing.

Argument	**Description**
-o, --out <string>	Specify output directory for report. The output directory will contain directories with the following: • HTML report (what you'll likely want to read) • RST files (used to generate the HTML report) • Text output (what would have been printed in plaintext using –console) • A readme file that explains the directories created, along with the command line used to run Snipper

--console	Write output to the console instead of creating a report directory.
Options for specifying SNPs, genes, and regions	
-s, --snp <string>	Lookup information for a list of SNPs - these must be separated by commas, surrounded by quotes (whitespace ignored.) Example: -s "rs1002227, rs35712349"
--snpfile <string>	Provide a list of SNPs to lookup from a file. The file may have *ANY* format, provided the file contains plain text. The program will pattern match rs### identifiers from your file. If you have SNPs in the 1000G format (e.g. chr4:9393) they must be specified on the command line using the -s option for now.
--build <string>	Select build to use for finding the positions of SNPs and genes. Snipper comes with support for hg19 by default, though other databases can be built. See "Building your own position database" for more information.
-g, --gene <string>	Lookup information for a list of gene symbols - these must be separated by commas, surrounded by quotes (whitespace ignored)
--genefile <string>	A file of genes to include in the Snipper report. Genes should be the primary HGNC gene symbol. One per line.
-r, --regions <string>	Provide a list of chromosomal regions. Genes within these regions will be included in the report. Example: --regions "chr4:19141-939393,chrX:9191-939393" The position numbers themselves should not contain commas.
-d <string>	Distance away from SNP to search, default is 1000000. If a distance is specified, the program will return *ALL* genes within the distance you specify, not just the default of 1. To specify a new distance, but still only return 1 gene (or arbitrary number of genes), use -n <number>. Distances can be specified using a kb or mb suffix, or as a raw distance. Examples: 500kb, 0.5MB, 1.4MB,834141.
-n <int>	Number of genes to return per SNP, default is 1. Note that this works in conjunction with the –d parameter listed above. For example, if you specify a distance of 10MB, but set –n 3, the program will search within 10 megabases of your SNP and return the 3 nearest genes.
Options related to PubMed: search terms, how to search, and how many articles to return	
--terms <string>	Comma-delimited string of terms, enclosed in quotes, to use in searching the literature. This will execute a search, per gene, for any of the search terms. For example: Genes: RB1, TCF7L2

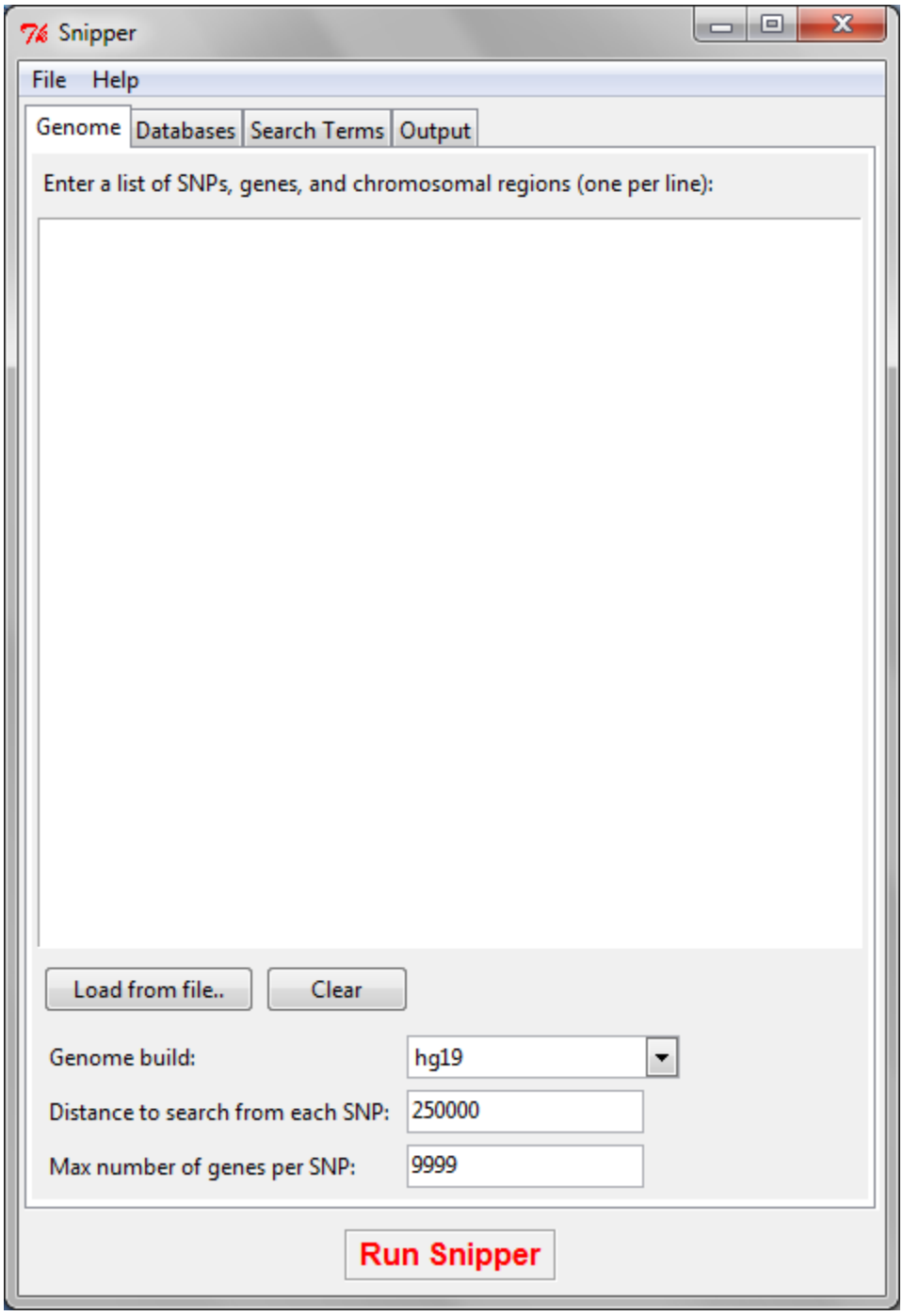
	Search terms: "glucose,retinoblastoma" What happens: -- Search literature for RB1 AND (glucose OR retinoblastoma) -- Search literature for TCF7L2 AND (glucose OR retinoblastoma) The information for genes RB1 and TCF7L2 will contain a list of PubMed articles that matched at least 1 of your search terms, all of which will be lumped together. Searching for terms with spaces is possible, but the entire argument must be enclosed in quotation marks. For example: --terms "type 2 diabetes, insulinemia, metabolic syndrome"
--each-term	When specified, the program will search each gene x search term pair, instead of lumping together search terms. For example: Genes: RB1, TCF7L2 Search terms: "glucose,retinoblastoma" What happens: -- Search literature for RB1 AND glucose -- Search literature for RB1 AND retinoblastoma -- Search literature for TCF7L2 AND glucose -- Search literature for TCF7L2 AND retinoblastoma The information for genes RB1 and TCF7L2 will have sections of PubMed articles that matched each search term individually. While this makes it more apparent why each PubMed article was returned, it also requires sending more queries to NCBI, and therefore increases the runtime of the program significantly. This is disabled by default.
--papernum <int>	Number of recent papers to display, default is 5
Options for disabling various databases (all are enabled by default)	
--no-generif	Disable GeneRIFs.
--no-omim	Disable OMIM.
--no-pubmed	Disable PubMed.
Options related to ScanDB (eQTL database)	
--no-scandb	Disables use of ScanDB for finding eQTLs connecting user defined SNPs to genes.
--scandb-pval	Change the p-value threshold for calling an eQTL association as “significant.” The default is 1.0E-06.

Using Snipper (GUI)

To launch the GUI:

- Double-click the “Snipper” icon on your desktop (or in the Start Menu, if you chose not to create a desktop icon) (Windows)
- Run bin/snipper.py without any command line arguments (Linux; see Installation to make sure you have the dependencies installed to run the GUI)

When the GUI launches, you should see the following screen:



The main text area is where SNPs, genes, and chromosomal regions can be listed. You can mix these interchangeably, so long as you keep them one per line. For example, you could try:

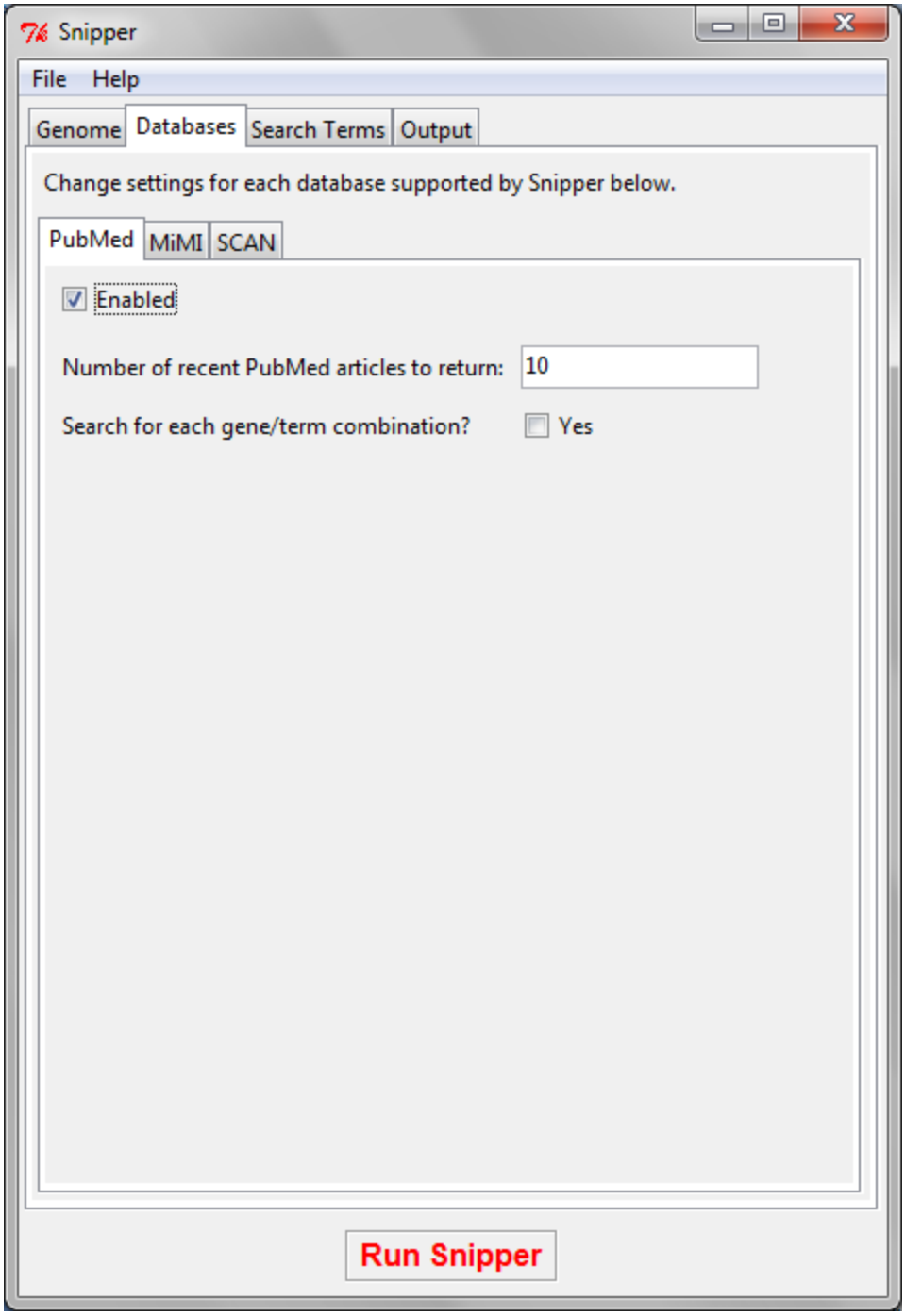
rs7903146
PDE8B

chrX:70608087-70685854
KCNJ11
rs1002227

Alternatively, you could click "**Load from file..**" to read a list of SNPs, genes, and chromosomal regions from a file.

Towards the bottom, there are a few additional options. **Genome build** denotes which human genome build to use when looking the positions of SNPs and genes. **Distance to search from each SNP** specifies the maximal distance that will be searched from each SNP when looking for genes. **Max number of genes per SNP** controls the number of genes that will be returned near a SNP (prioritized by distance, so if there were 5 genes but you desired only the nearest, you would set this to 1.) Leave the value at 9999 to return all genes near a SNP.

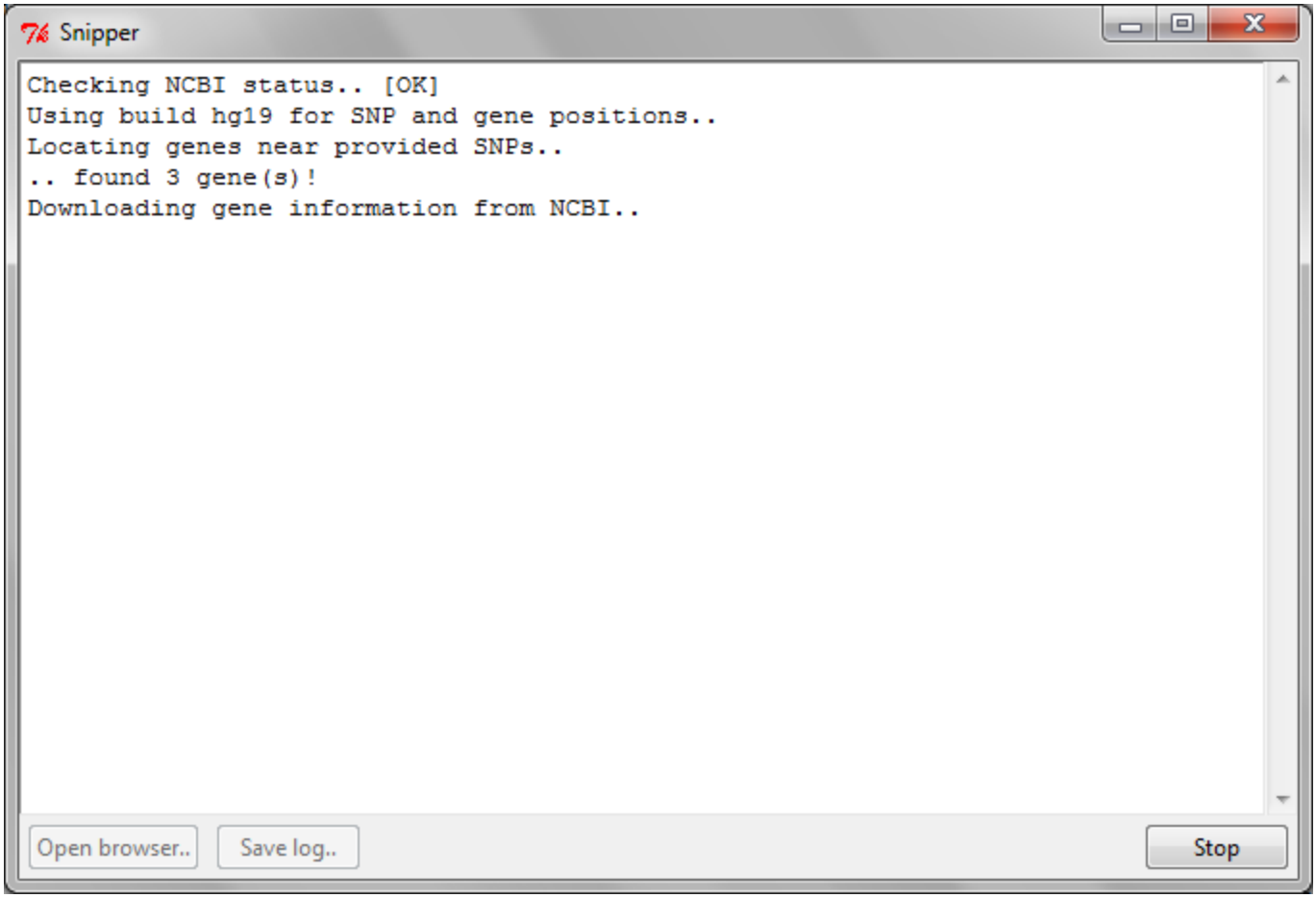
On the "**Databases**" tab there are various options for enabling or disabling databases for querying, as well as changing options that may be specific to a particular one:



The "**Search Terms**" tab is where you can list words or phrases that will be used to search within information returned from OMIM, PubMed, and Entrez Gene. Enter one word or phrase per line.

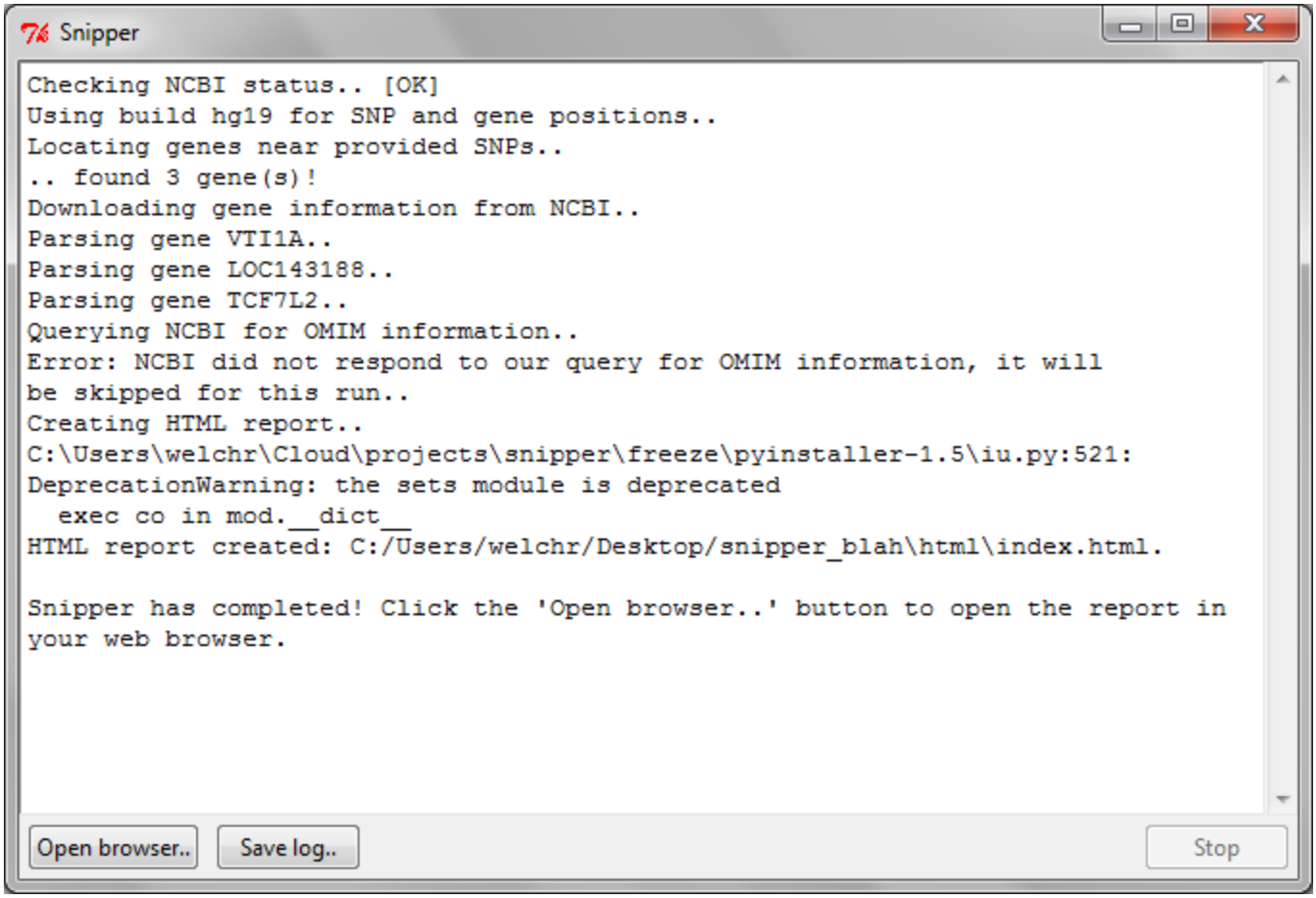
The "**Output**" tab allows the user to change where the snipper report directory will be located. If the directory already exists,

Once you're satisified with your settings, click the "**Run Snipper**" button at the bottom, or select File → Run Snipper. You'll see a console window appear:



This window shows what Snipper is currently doing, and is very similar to what would be seen in command line mode. At the bottom right, the **Stop** button can be used to terminate the current run.

Once the run has completed, you should see the following:



The "**Open browser..**" button will directly open the HTML report in your browser. This is the best way to quickly view the report. You could alternatively navigate to the directory (shown in the log file above) and open the "index.html" file.

If you wish to save the log, you can do so by clicking "**Save log..**" and selecting where the console text will be written.

To run Snipper again, simply close the console, change your settings, and click "**Run Snipper**" again.

Output

Snipper Console Output

Below, we describe the anatomy of the snipper console output. This mode is less preferable than the HTML output, but can be used for quick inspection via the command line. To activate console output, use the `--no-html` parameter. In this particular example, we've searched near known type 2 diabetes SNPs, and returned the nearest gene within 250kb for each SNP.

Gene	# SNPs	# Terms	Total Pubmed
CDKAL1	1	4	64
KCNJ11	1	4	198
IGF2BP2	1	4	58
NOTCH2	1	2	76
JAZF1	1	2	42
TCF7L2	1	4	253
KCNQ1	1	4	207
HNF1B	1	4	88
SLC30A8	1	4	83
WFS1	1	4	89

This table gives a listing for each gene identified near a SNP given as input by the user.

#SNPs - number of SNPs given by the user that were near this gene. This can be > 1 when you have 2 SNPs very close by each other.

Terms - the number of user-defined search terms that were found in the information for this gene.

```
THADA       1          2          20
FTO         1          4          133
HHEX        1          4          89
MTNR1B      1          4          34
ADAMTS9     1          2          26
CDC123      1          2          24
PPARG       1          4          938
TSPAN8      1          2          26
CDKN2B      1          4          165
KIAA1486    1          0          4

SNP           Gene/Aliases
---           ------------
rs11899863    THADA/GITA/FLJ77530/FLJ44876/FLJ44016/KIAA1767
rs7903146     TCF7L2/TCF4/TCF-4
rs1387153     MTNR1B/MT2/MEL-1B-R
rs849134      JAZF1/DKFZp761K2222/ZNF802/TIP27
rs4430796     HNF1B/TCF2/FJHN/HNF1beta/HPC11/VHNF1/MODY5/HNF
rs6795735     ADAMTS9/FLJ42955/KIAA1312
rs3802177     SLC30A8/ZNT8/ZnT-8
rs10923931    NOTCH2/AGS2/hN2
rs1801214     WFS1/WFS/FLJ51211/WOLFRAMIN/WFRS
rs163184
KCNQ1/FLJ26167/JLNS1/LQT/KVLQT1/Kv1.9/KCNA9/SQT2/RWS/LQT1/W
rs10965250    CDKN2B/MTS2/TP15/P15/p15INK4b/CDK4I/INK4B
rs11642841    FTO/KIAA1752/MGC5149
rs1470579     IGF2BP2/IMP2/p62/IMP-2/VICKZ2
rs10440833    CDKAL1/FLJ46705/MGC75469/FLJ20342
rs12779790    CDC123/D123/C10orf7/FLJ13863
rs5015480     HHEX/HMPH/HEX/PRH/PRHX/HOX11L-PEN
rs7578326     KIAA1486
rs4760790     TSPAN8/CO-029/TM4SF3
rs5215        KCNJ11/IKATP/TNDM3/PHHI/HHF2/KIR6.2/MGC133230/BIR
rs13081389    PPARG/PPARgamma/GLM1/PPARG2/PPARG1/CIMT1/NR1C3
```

For each SNP provided by the user, the table to the left gives the following information:

- The SNP itself
- A row for each gene found near the SNP (this particular example only has 1 gene per SNP, there could be more depending on your settings)
- The primary gene symbol is listed first, followed by the gene's aliases, i.e.:

  TCF7L2/TCF4/TCF-4

  TCF7L2 is the primary symbol, TCF4 and TCF-4 are aliases.

From this point forward, Snipper lists genes found near SNPs. It will first list genes known to be associated with SNPs (eQTLs), and then subsequently list the remaining genes in order by distance to SNP. If the user is interested in a particular gene, most text editors can search quickly by using CTRL+F.

For this particular example, we list only 1 gene - CDKAL1.

```
================================================================================

[+] GENE: potassium inwardly-rectifying channel, subfamily J
[+] Entrez Gene UID: 3767
[+] Location: 11p15.1
[+] Type: protein-coding
[+] Synonyms: IKATP TNDM3 PHHI HHF2 KIR6.2 MGC133230 BIR

[+] Search terms matched:
-- Location: Gene summary    Terms: insulin, diabetes
-- Location: GO Term         Terms: insulin, glucose
-- Location: GeneRIF         Terms: glucose, insulin, diabet
-- Location: Pubmed          Terms: any
-- Location: Phenotype       Terms: diabetes
-- Location: KEGG Pathway    Terms: diabetes
-- Location: OMIM Text       Terms: glucose, insulin, diabet

[+] Associated SNPs:
SNP: rs5215          Distance (bp): 0    Direction: Within
```

The gene's full name, primary symbol, and synonyms. Also the UID (Entrez Gene's identifier for the gene), chromosomal location, and type of gene.

This section shows which search terms provided by the user were found in this gene's information, and also shows where they matched.

SNP provided by user to search near, and the distance to this SNP.

[+] Summary: Potassium channels are present in most mammalian cells, where they participate in a wide range of physiologic responses. The protein encoded by this gene is an integral membrane protein and inward-rectifier potassium channel. The encoded protein, which has a greater tendency to allow potassium to flow into a cell rather than out of a cell, is controlled by G-proteins and is found associated with the sulfonylurea receptor SUR. Mutations in this gene are a cause of familial persistent hyperinsulinemic hypoglycemia of infancy (PHHI), an autosomal recessive disorder characterized by unregulated insulin secretion. Defects in this gene may also contribute to autosomal dominant non-insulin-dependent diabetes mellitus type II (NIDDM), transient neonatal diabetes mellitus type 3 (TNDM3), and permanent neonatal diabetes mellitus (PNDM). [provided by RefSeq]

Summary of the gene (provided by NCBI Entrez Gene.)

[+] Phenotypes: "Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome wide association data" "Diabetes mellitus, permanent neonatal, with neurologic features, 606176, {Diabetes mellitus, type 2, susceptibility to" "Diabetes mellitus, type 2, susceptibility to" "Diabetes, permanent neonatal" "Hyperinsulinemic hypoglycemia, familial, 2" "Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes" "Diabetes mellitus, transient neonatal, 3" "Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes" "A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants" "Diabetes mellitus, permanent neonatal, with neurologic features" "Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels"

Phenotypes associated with the gene (provided by NCBI Entrez Gene.)

[+] KEGG Pathways:
Type II diabetes mellitus [http://www.genome.jp/dbget-bin/show_pathway?hsa04930+3767]

KEGG pathways.

[+] GO Terms: "regulation of membrane potential" "ATP-activated inward rectifier potassium channel activity" "potassium ion import" "endoplasmic reticulum" "voltage-gated ion channel activity" "ATP-sensitive potassium channel complex" "plasma membrane" "microsome" "negative regulation of insulin secretion" "ion transport" "glucose metabolic process" "protein C-terminus binding" "T-tubule" "response to drug" "ATP binding" "mitochondrion" "potassium ion binding" "neurological system process" "response to ATP"

Gene ontology terms.

[+] OMIM: [600937] Link: http://www.ncbi.nlm.nih.gov/entrez/
[+] OMIM Text: The KCNJ11 gene encodes a subunit of an inwardly rectifying ATP-sensitive potassium channel. I(KATP) channels were discovered in cardiac muscle and later found in pancreatic beta cells, pituitary tissue, skeletal muscle, brain, and vascular and nonvascular smooth muscle. I(KATP) currents function in secretion and muscle contraction by coupling metabolic activity to membrane potential ({13:Inagaki et al., 1995}).

OMIM ID, link to the OMIM article itself, and the OMIM summary for the gene.

[+] Gene references into function for KCNJ11:
-- Observational study of gene-disease association. (HuGE Navigator) PMID: 19685080
-- Mutations in the pore-forming K(ATP) channel subunit cause neonatal diabetes & discusses recent advances in understanding of clinical features of neonatal diabetes, its underlying molecular mechanisms & impact on treatment[review] PMID: 18566517
-- mutations in the slide helix of Kir6.2 (V59G) influence the channel kinetics, providing evidence that this domain is involved in Kir channel gating PMID: 15583126
-- Case of an 18-month-old infant with permanent neonatal diabetes due to an activating KCNJ11 mutation who successfully transitioned from subcutaneous insulin therapy to oral sulfonylurea therapy in the outpatient setting. PMID: 18221420
-- the MDR-like core of SUR is linked with the K(IR) pore in KATP channels PMID: 12213829
-- caveolin-3 negatively regulates Kir6.2/SUR2A channel function. PMID: 19481058
-- The prevalent Glu23Lys polymorphism in the potassium inward rectifier 6.2

GeneRIFs (Gene References Into Function), provided by NCBI Entrez Gene.

This can be enabled by either --generif or the --all option.

```
   (KIR6.2) gene is associated with impaired glucagon suppression in response to
   hyperglycemia. PMID: 12196481
-- the common E23K genetic variant at the KCNJ11 gene locus was significantly
   associated with cardiovascular function PMID: 17720745

[+] Top Pubmed articles linked to gene KCNJ11, by date:
-- Zhao J et al. "Examination of type 2 diabetes loci implic
   birth weight gene." Diabetes. 2009 Oct;58(10):2414-8. PMI
-- Salanti G et al. "Underlying genetic models of inheritanc
   type 2 diabetes associations." Am J Epidemiol. 2009 Sep 1
   PMID: 19602701
-- Schulze MB et al. "Use of Multiple Metabolic and Genetic
   the Prediction of Type 2 Diabetes: the European Prospecti
   into Cancer and Nutrition (EPIC)-Potsdam study." Diabetes
   PMID: 19720844
-- Reyes S et al. "K(ATP) channel Kir6.2 E23K variant overre
   heart failure is associated with impaired exercise stress
   Genet. 2009 Aug 14;. PMID: 19685080
-- Yoshida T et al. "Association of genetic variants with ch
   in individuals with different lipid profiles." Int J Mol
   Aug;24(2):233-46. PMID: 19578796

[+] Top Pubmed articles linked to gene KCNJ11 matching any s
-- Gach A et al. "Neonatal diabetes in a child positive for
   antibodies at onset and Kir6.2 activating mutation." Diab
   2009 Nov;86(2):e25-e27. PMID: 19692135
-- 't Hart LM et al. "A Combined Risk Allele Score of Eight
   Genes Is Associated With Reduced First Phase Glucose Stim
   Secretion During Hyperglycemic Clamps." Diabetes. 2009 Oc
-- StancÃ¡kovÃ¡ A et al. "Association of 18 confirmed suscep
   type 2 diabetes with indices of insulin release, proinsul
   insulin sensitivity in 5,327 nondiabetic Finnish men." Di
   Sep;58(9):2129-36. PMID: 19502414
-- Salanti G et al. "Underlying genetic models of inheritanc
   type 2 diabetes associations." Am J Epidemiol. 2009 Sep 1
   PMID: 19602701
-- Nikolac N et al. "Metabolic control in type 2 diabetes is
   sulfonylurea receptor-1 (SUR-1) but not with KCNJ11 polym
   Res. 2009 Jul;40(5):387-92. PMID: 19766903
-- Ting WH et al. "Improved diabetic control during oral sul
   in two children with permanent neonatal diabetes mellitus
   Endocrinol Metab. 2009 Jul;22(7):661-7. PMID: 19774848
```

Lists the most recent PubMed articles linked to this gene.

This information is enabled by either --pubmed or --all.

The number of articles returned is dependent on the --papernum option.

Lists PubMed articles that matched the user's search terms AND are linked to the gene listed here.

If multiple search terms are provided, this section will show papers that matched ANY of the search terms AND the gene.

For more in-depth searching, use --each-term. A section for each search term will then appear, showing the PubMed articles (listed by most recent first) linked to the gene and only that particular search term.

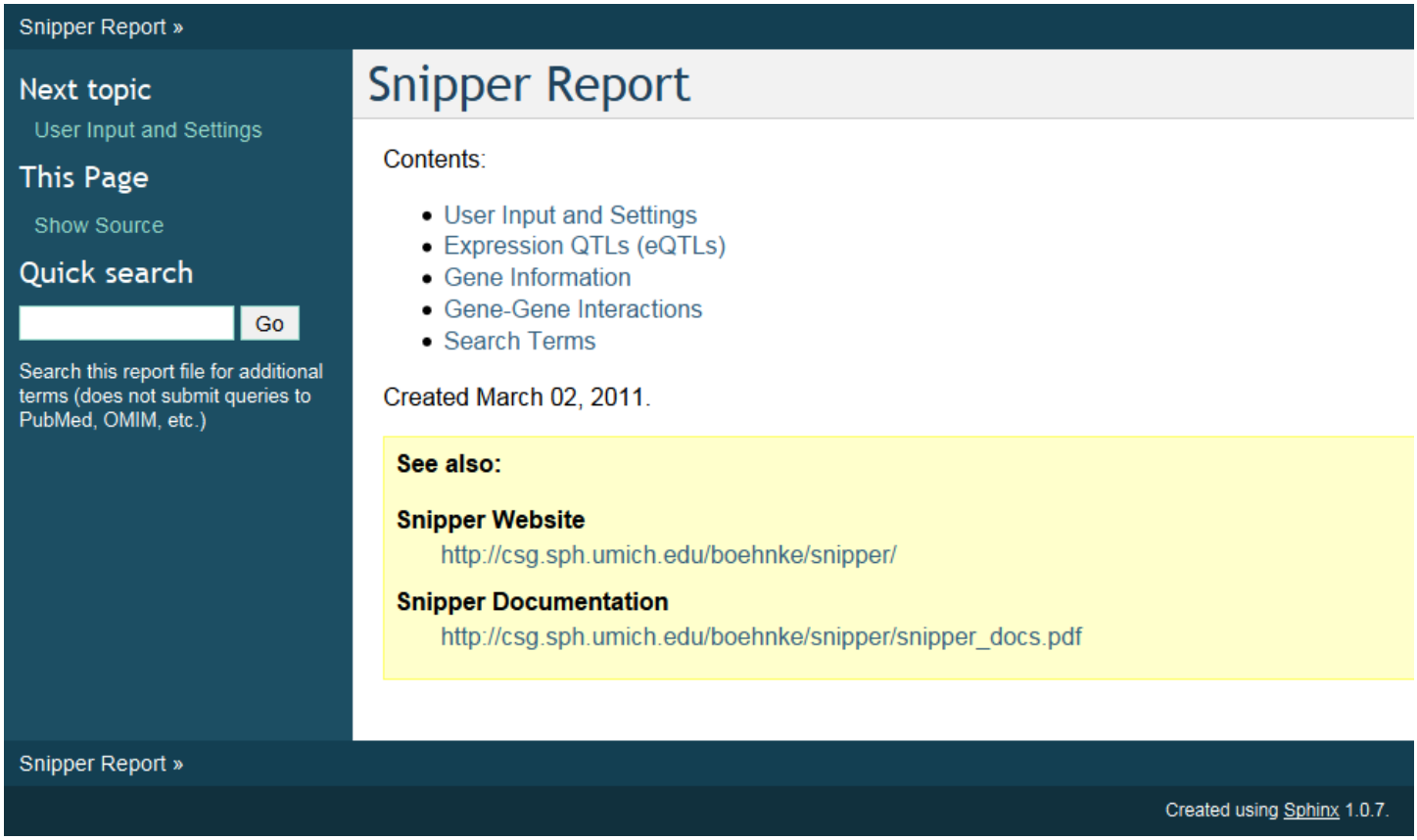
Snipper HTML output

This mode is the preferred method of running Snipper, and is also the default. Snipper will produce a formatted HTML report file in a directory of your choosing by doing the following:

```
snipper --snpfile yoursnps.txt –o my_html_directory
```

If no directory is given using –o, a directory called “snipper_report” is created by default.

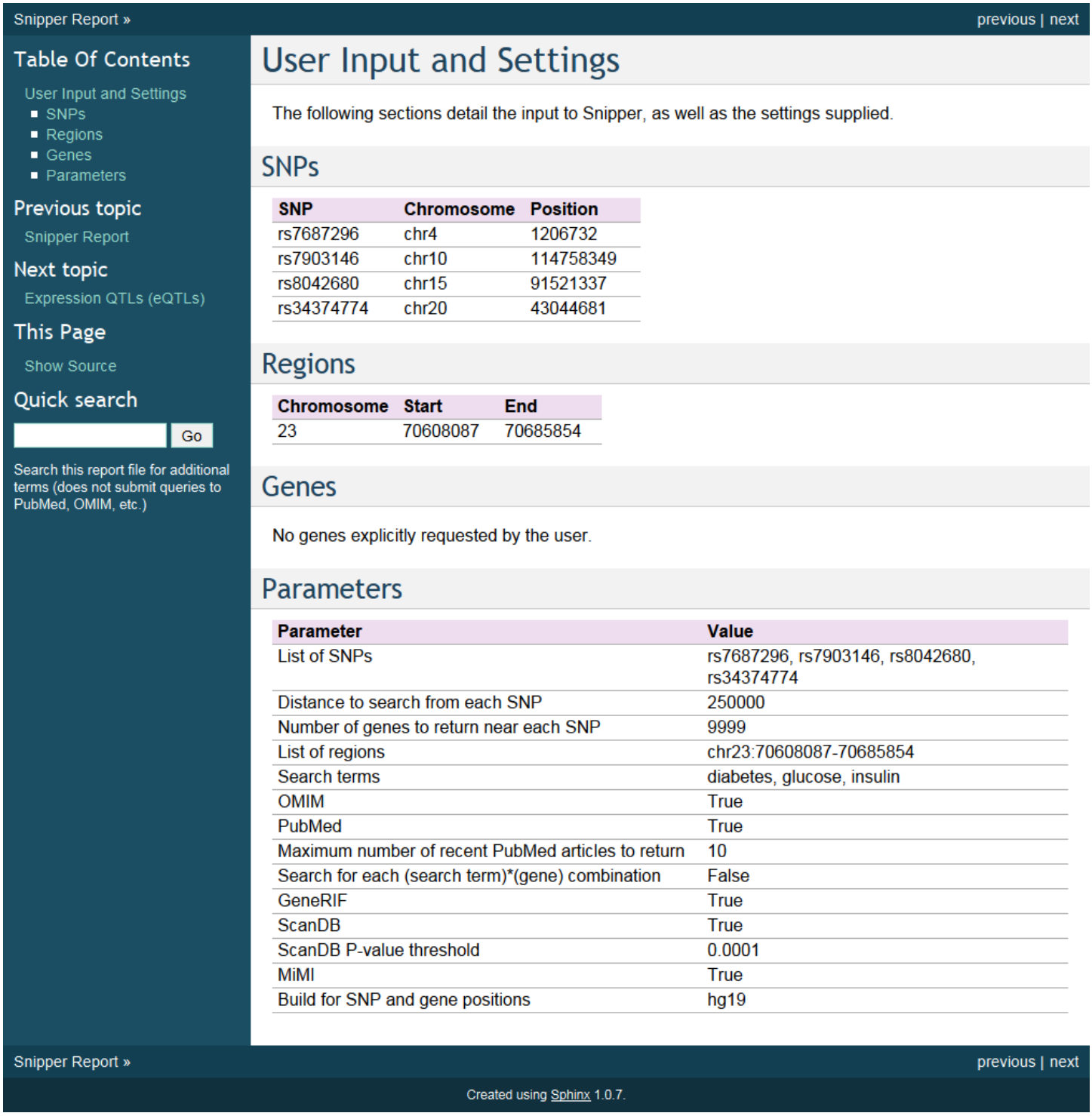
The HTML report begins with “index.html”, which has the table of contents. An image of this table (cropped to remove whitespace) is shown below.



The report contains 5 main sections:

- User Input and Settings: lists all of the input SNPs and regions as given by the user, as well as all command line options that were specified
- Expression QTLs: contains information on genes whose expression is associated with SNPs given by the user
- Gene Information: information on each gene found near the input SNPs/regions
- Gene-Gene Interactions: lists direct interactions between all genes found near input SNPs/regions
- Search Terms: lists each search term given by the user, and where they matched within each gene's information

An example of the input section:



Snipper Report » previous | next

Table Of Contents

User Input and Settings
- SNPs
- Regions
- Genes
- Parameters

Previous topic

Snipper Report

Next topic

Expression QTLs (eQTLs)

This Page

Show Source

Quick search

Go

Search this report file for additional terms (does not submit queries to PubMed, OMIM, etc.)

User Input and Settings

The following sections detail the input to Snipper, as well as the settings supplied.

SNPs

SNP	Chromosome	Position
rs7687296	chr4	1206732
rs7903146	chr10	114758349
rs8042680	chr15	91521337
rs34374774	chr20	43044681

Regions

Chromosome	Start	End
23	70608087	70685854

Genes

No genes explicitly requested by the user.

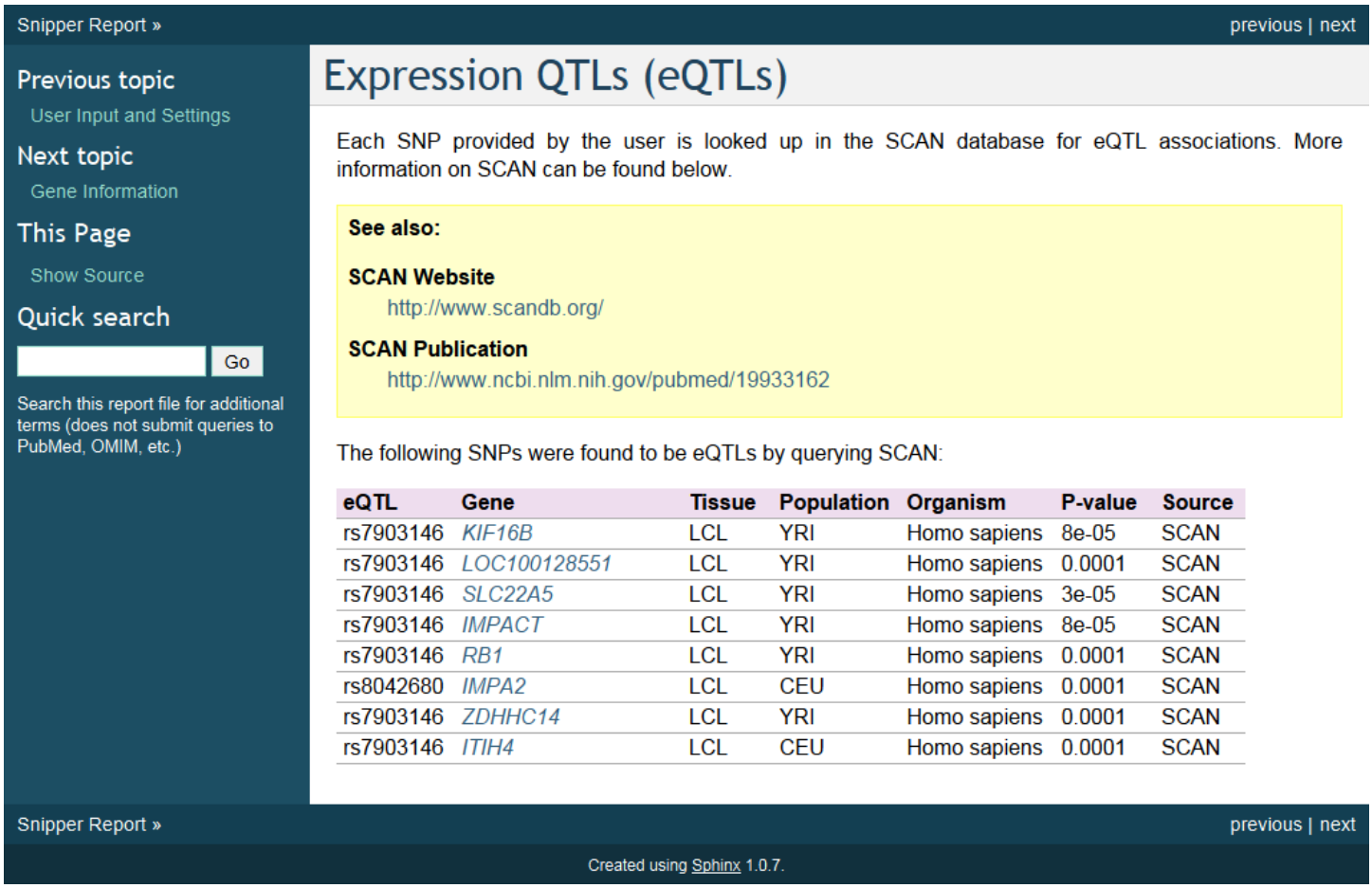
Parameters

Parameter	Value
List of SNPs	rs7687296, rs7903146, rs8042680, rs34374774
Distance to search from each SNP	250000
Number of genes to return near each SNP	9999
List of regions	chr23:70608087-70685854
Search terms	diabetes, glucose, insulin
OMIM	True
PubMed	True
Maximum number of recent PubMed articles to return	10
Search for each (search term)*(gene) combination	False
GeneRIF	True
ScanDB	True
ScanDB P-value threshold	0.0001
MiMI	True
Build for SNP and gene positions	hg19

Snipper Report » previous | next

Created using Sphinx 1.0.7.

The eQTL report contains a list of genes that were associated with user input SNPs. Each gene is listed along with the information for the association as retrieved from the SCAN database.



Snipper Report » previous | next

Previous topic
User Input and Settings
Next topic
Gene Information
This Page
Show Source
Quick search
Go
Search this report file for additional terms (does not submit queries to PubMed, OMIM, etc.)

Expression QTLs (eQTLs)

Each SNP provided by the user is looked up in the SCAN database for eQTL associations. More information on SCAN can be found below.

See also:

SCAN Website
http://www.scandb.org/

SCAN Publication
http://www.ncbi.nlm.nih.gov/pubmed/19933162

The following SNPs were found to be eQTLs by querying SCAN:

eQTL	Gene	Tissue	Population	Organism	P-value	Source
rs7903146	*KIF16B*	LCL	YRI	Homo sapiens	8e-05	SCAN
rs7903146	*LOC100128551*	LCL	YRI	Homo sapiens	0.0001	SCAN
rs7903146	*SLC22A5*	LCL	YRI	Homo sapiens	3e-05	SCAN
rs7903146	*IMPACT*	LCL	YRI	Homo sapiens	8e-05	SCAN
rs7903146	*RB1*	LCL	YRI	Homo sapiens	0.0001	SCAN
rs8042680	*IMPA2*	LCL	CEU	Homo sapiens	0.0001	SCAN
rs7903146	*ZDHHC14*	LCL	YRI	Homo sapiens	0.0001	SCAN
rs7903146	*ITIH4*	LCL	CEU	Homo sapiens	0.0001	SCAN

Snipper Report » previous | next

Created using Sphinx 1.0.7.

The gene report contains two sections: a table listing each gene found near SNPs, or those found in regions given by the user, and finally an individual section for each gene. An example:

Table Of Contents

- Gene Information
  - Region Table
  - Genes
    - CTBP1
    - LOC100130872
    - SPON2
    - C4orf42
    - MAEA
    - TMED11P
    - RNF212
    - KIAA1530
    - CRIPAK
    - FGFRL1
    - IDUA
    - SLC26A1
    - DGKQ
    - TCF7L2
    - LOC143188
    - VTI1A
    - SLC22A5
    - KIF16B
    - IMPACT
    - LOC100128551
    - RB1
    - ZDHHC14
    - ITIH4
    - PRC1
    - RCCD1
    - VPS33B
    - UNC45A
    - HDDC3
    - MAN2A2
    - FES
    - FURIN
    - SV2B
    - BLM
    - IMPA2
    - HNF4A
    - MIR3646
    - TTPAL
    - R3HDML
    - SERINC3

Gene Information

Region Table

The table below lists each user-provided SNP or chromosomal region, sorted by position in the genome.

For each SNP, nearby genes are listed first, sorted by distance to the SNP. eQTL genes are listed after nearby genes.

Search terms that match information other than PubMed searches are listed individually. If at least 1 term matches an article in PubMed, the word "pubmed" is listed in the "Search Terms Matched" column.

SNP/Region	Chrom	Nearby Gene	eQTL Gene	Search Terms Matched	PubMed Articles
rs7687296	4	*CTBP1*		pubmed, glucose	113
rs7687296	4	*LOC100130872*			3
rs7687296	4	*SPON2*		pubmed	13
rs7687296	4	*C4orf42*			3
rs7687296	4	*MAEA*		pubmed	6
rs7687296	4	*TMED11P*			3
rs7687296	4	*RNF212*			7
rs7687296	4	*KIAA1530*			3
rs7687296	4	*CRIPAK*			3
rs7687296	4	*FGFRL1*		pubmed	22
rs7687296	4	*IDUA*		pubmed	46
rs7687296	4	*SLC26A1*			6
rs7687296	4	*DGKQ*			14
rs7903146	10	*TCF7L2*		insulin, glucose, pubmed, diabetes	349
rs7903146	10	*LOC143188*			1
rs7903146	10	*VTI1A*		pubmed	13
rs7903146	10		*SLC22A5*	pubmed, diabetes	110
rs7903146	10		*KIF16B*		11
rs7903146	10		*IMPACT*	pubmed	8
rs7903146	10		*LOC100128551*		0
rs7903146	10		*RB1*	glucose, pubmed, insulin	629
rs7903146	10		*ZDHHC14*		4
rs7903146	10		*ITIH4*	pubmed	20

TCF7L2

General information regarding this gene:

Full gene name:	transcription factor 7-like 2 (T-cell specific, HMG-box)
Entrez Gene ID:	6934
Location:	10q25.3
Synonyms:	TCF4, TCF-4
Type:	protein-coding

SNPs given by the user that are near or inside this gene:

SNP	Distance (bp)	Direction
rs7903146	0	within

Summary

This gene encodes a high mobility group (HMG) box-containing transcription factor that plays a key role in the Wnt signaling pathway. The protein has been implicated in blood glucose homeostasis. Genetic variants of this gene are associated with increased risk of type 2 diabetes. Several transcript variants encoding multiple different isoforms have been found for this gene.

OMIM Summary

The TCL7L2 gene product is a high mobility group (HMG) box-containing transcription factor implicated in blood glucose homeostasis. The study of Yi et al. (2005) suggested that TCL7L2 acts through regulation of proglucagon (138030) through repression of the proglucagon gene in enteroendocrine cells via the Wnt signaling pathway. [OMIM ID 602228]

Phenotypes

- Diabetes mellitus, type 2, susceptibility to
- Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia.
- Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.
- Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population.
- A variant in CDKAL1 influences insulin response and risk of type 2 diabetes.
- Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium.
- A genome-wide association study of the metabolic syndrome in Indian Asian men
- A genome-wide association study identifies novel risk loci for type 2 diabetes.
- A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.
- New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.
- OMIM

The example is abbreviated and only includes a single gene, TCF7L2. Within this gene, every occurrence of the words “glucose”, “insulin”, and “diabetes” are highlighted (these are search terms supplied with --terms.)

The next section displays a list of direct interactions between all genes (those both near SNPs and within regions), along with their details as downloaded from MiMI. An example:

Table Of Contents

Gene-Gene Interactions
- Direct Interactions

Previous topic

Gene Information

Next topic

Search Terms

This Page

Show Source

Quick search

Go

Search this report file for additional terms (does not submit queries to PubMed, OMIM, etc.)

Gene-Gene Interactions

The following sections list different sets of interactions between genes found near input SNPs/regions. These interactions are collected from NCIBI's MiMI database.

See also:

Michigan Molecular Interactions (MiMI)
http://mimi.ncibi.org/

Direct Interactions

The following table lists direct interactions between the protein products of genes found near any input region:

Gene 1	Gene 2	Types	Provenance	PMIDs
RB1	*HNF4A*	• bidirectional	• BIND	
CTBP1	*RB1*	• bidirectional • Invivo • in vivo • Affinity Capture-MS	• CCSB • GRID • HPRD	• 11583618 • 16189514
TAF1	*RB1*	• acts_on • in vitro • Invivo • in vivo • Invitro • bidirectional	• BIND • DIP • IntAct • GRID • HPRD	• 9242374 • 9858607 • 7724524
RB1	*HNF4A*	• bidirectional	• BIND	
CTBP1	*RB1*	• bidirectional • Invivo • in vivo • Affinity Capture-MS	• CCSB • GRID • HPRD	• 11583618 • 16189514
CTBP1	*TCF7L2*	• PPrel • bidirectional	• BIND	
TAF1	*RB1*	• acts_on • in vitro • Invivo • in vivo • Invitro • bidirectional	• BIND • DIP • IntAct • GRID • HPRD	• 9242374 • 9858607 • 7724524
CTBP1	*TCF7L2*	• PPrel • bidirectional	• BIND	

For these interactions, the following ontology terms were found:

Gene 1	Gene 2	Components	Processes	Functions
RB1	*HNF4A*	• nucleus [GO:0005634]	• transcription [GO:0006350] • positive regulation of transcription from RNA polymerase II promoter [GO:0045944]	• transcription factor activity [GO:0003700]
CTBP1	*RB1*	• nucleus [GO:0005634]	• negative regulation of cell proliferation [GO:0008285]	• transcription factor binding [GO:0008134]
TAF1	*RB1*	• nucleus [GO:0005634]	• transcription [GO:0006350] • cell cycle [GO:0007049] • regulation of transcription, DNA-dependent [GO:0006355]	
RB1	*HNF4A*	• nucleus [GO:0005634]	• transcription [GO:0006350] • positive regulation of transcription from RNA polymerase II promoter [GO:0045944]	• transcription factor activity [GO:0003700]
CTBP1	*RB1*	• nucleus	• negative regulation of cell proliferation	• transcription factor

And finally, a section listing each search term provided by the user, and where each term matched within the information for each gene:

Previous topic
Gene-Gene Interactions
This Page
Show Source
Quick search
Go
Search this report file for additional terms (does not submit queries to PubMed, OMIM, etc.)

Search Terms

any search term

- *SLC22A5 - Pubmed*
- *IMPACT - Pubmed*
- *RB1 - Pubmed*
- *IMPA2 - Pubmed*
- *ITIH4 - Pubmed*
- *HNF4A - Pubmed*
- *PRC1 - Pubmed*
- *TAF1 - Pubmed*
- *TCF7L2 - Pubmed*
- *CTBP1 - Pubmed*
- *SPON2 - Pubmed*
- *VPS33B - Pubmed*
- *MAN2A2 - Pubmed*
- *MAEA - Pubmed*
- *FES - Pubmed*
- *FURIN - Pubmed*
- *FITM2 - Pubmed*
- *PKIG - Pubmed*
- *SV2B - Pubmed*
- *BLM - Pubmed*
- *VTI1A - Pubmed*
- *FGFRL1 - Pubmed*
- *ADA - Pubmed*
- *IDUA - Pubmed*

diabetes

- *SLC22A5 - GeneRIF*
- *HNF4A - Gene summary*
- *HNF4A - GeneRIF*
- *HNF4A - Phenotype*
- *HNF4A - Pathway*
- *HNF4A - OMIM Text*
- *PRC1 - Phenotype*
- *TCF7L2 - Gene summary*
- *TCF7L2 - GeneRIF*
- *TCF7L2 - Phenotype*
- *TCF7L2 - OMIM Text*
- *ADA - GeneRIF*
- *ADA - OMIM Text*

glucose

- *RB1 - GeneRIF*
- *HNF4A - GeneRIF*
- *HNF4A - OMIM Text*
- *TCF7L2 - GO Term*
- *TCF7L2 - Gene summary*
- *TCF7L2 - GeneRIF*
- *TCF7L2 - Phenotype*
- *TCF7L2 - OMIM Text*
- *CTBP1 - GeneRIF*
- *SV2B - OMIM Text*
- *ADA - GeneRIF*

insulin

- *RB1 - GeneRIF*
- *HNF4A - Gene summary*
- *HNF4A - GeneRIF*
- *HNF4A - OMIM Text*
- *TCF7L2 - GO Term*
- *TCF7L2 - GeneRIF*
- *TCF7L2 - Phenotype*
- *TCF7L2 - OMIM Text*
- *SV2B - OMIM Text*

The “any” term is used when search terms are globbed together as 1 query, that is: “term1 OR term2 OR term3”. This makes searching PubMed much faster, but also not quite as specific. The user can supply the --each-term parameter, which forces Snipper to submit independent queries to PubMed for each search term + gene pair.

Examples

We list below a few examples of running Snipper by giving both a word description of what the program is doing, along with the command line parameters.

- Using Snipper on a single SNP (-s)
- Search 500 kb away (-d 500kb) from the SNP for genes
- List more PubMed articles than the default (--papernum 10)

```
snipper -s "rs1002227" -d 500kb --papernum 10
```

- Use Snipper on a file that contains SNP (rs#) names. The file can be of arbitrary format.
- Search 1 MB from each SNP in the file for genes, but return only the 3 nearest genes (-n) for each SNP
- Add search terms "diabetes","insulin","glucose" (--terms)

```
snipper --snpfile file_with_snps.txt -d 1MB -n 3 --terms
“diabetes,glucose,insulin”
```

Building your own position database

Snipper comes pre-loaded with a database file giving the positions of SNPs and genes for human genome build hg19 (UCSC).

To build your own database, a script has been provided in the bin/ directory called "build_db.py". You simply need to execute this script with the human genome build for which you wish to build a database. For example:

```
bin/build_db.py --build hg18
```

This will create a database file called "hg18.db" in the data/genome/ directory, and add information about this newly created database to the conf file (conf/snipper.conf). To use this new database, you can use the --build parameter when running Snipper, like so:

```
snipper -s "rs7903146" --build hg18
```

The build_db.py script always downloads the latest snp and refFlat tables for a given human genome build.

Warning: building the database can take many hours (2-3 at minimum.) You should allow ample time for the script to complete!

License

along with this program. If not, see <http://www.gnu.org/licenses/>.

In addition, Snipper is bound by the licensing of those programs, databases, and services included within.

This software is subject to the restrictions and copyright information set forth by NCBI, which can be referenced at:
http://www.ncbi.nlm.nih.gov/About/disclaimer.html

Additionally, all journal abstracts, titles, and other information associated with PubMed as a service of the National Library of Medicine (NLM) may be copyrighted by the journal authors. Please read the link below for more information.
http://www.nlm.nih.gov/databases/download.html

OMIM(TM) and Online Mendelian Inheritance in Man(TM) are trademarks of Johns Hopkins University. This software complies with the licensing restrictions set forth by Johns Hopkins University, and may only be used for academic (non-commercial) research applications.
http://www.ncbi.nlm.nih.gov/Omim/restrictions.html