Snipper Documentation
Version 1.1, Released 11/10/2010

Contents

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)

Requirements

Snipper requires Python version 2.6 or greater, but not the 3.0 branch.

In addition, Snipper requires the following python packages, which are installed by the setup script:

- Sphinx, a python documentation generator (http://sphinx.pocoo.org/)

To run the setup script itself, you will need:

- An internet connection, for downloading data files from the UCSC FTP site, and python packages

Snipper has been tested on both Windows Vista/7 and Ubuntu Linux.

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 also be run from a gene-centric view, instead of for a set of SNPs or regions:

```
snipper –-gene “TCF7L2,P53,BRCA1”
```

Or, the program can be run with chromosomal regions:

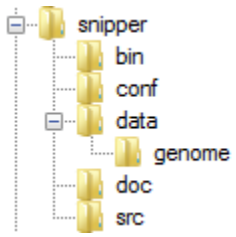
```
snipper –-regions “chr#:start-end”
```

SNPs and regions can be mixed together:

```
snipper -s "rs7903146,rs1002227" –-regions “chr#:start-end”
```

Installation

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:



   In Unix, you can do this by:

   ```
   tar zxf snipper_release.tgz
   ```

   In Windows, you will need to use a program such as WinZip or WinRAR.

2. Next, navigate to snipper/bin, and run the **setup_snipper.py** script. This script handles two tasks: install dependencies, and download information from UCSC.

   ```
   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.)

Snipper comes pre-loaded with data from UCSC hg19: both refFlat, and snp131 positions. The script checks to see if both of these are up to date, and if they are, it simply does nothing. If either are new (for example, snp132 is released), the script will download the table and create the database for Snipper to use.

You can also directly specify the build and snp table you would like to use. For example:

```
python setup_snipper.py --build hg17 --snptable snp125
```

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

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

In Windows, the easiest solution for quickly launching Snipper would be to add snipper/bin to your path.

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 or file for report. This specifies a directory for writing when using HTML output (the default), or a file when using console output.
--no-html	Use console output. In this case, -o <file> specifies a file to use for writing plain text.
Options for specifying SNPs, distance to search, and how many genes to return	
-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. The program will pattern match rs### identifier from your file. This is the most commonly used way of running Snipper.
-g, --gene <string>	Lookup information for a list of gene symbols - these must be separated by commas, surrounded by quotes (whitespace ignored)
-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)
--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 This is a much more in-depth search, at the cost of running time - NCIBI limits to 1 query / 3 seconds. If you have a very large set of genes and search terms, this can take a VERY long time to run!
--papernum <int>	Number of 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.

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
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
```

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.

Total Pubmed - the total number of PubMed articles linked to this gene. This gives the user an idea of how well-researched this particular gene is.

```
SNP           Gene/Aliases
---           ------------
rs11899863    THADA/GITA/FLJ77530/FLJ44876/FLJ44016/KIAA1767/FLJ21873
rs7903146     TCF7L2/TCF4/TCF-4
rs1387153     MTNR1B/MT2/MEL-1B-R
rs849134      JAZF1/DKFZp761K2222/ZNF802/TIP27
rs4430796     HNF1B/TCF2/FJHN/HNF1beta/HPC11/VHNF1/MODY5/HNF2/LFB3/TCF-2
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/WRS/KCNA9/KVLQT1/KV1.9/KCNQ1
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

[+] Summary: Potassium channels are present in most mammalia
participate in a wide range of physiologic responses. The pr
this gene is an integral membrane protein and inward-rectifi
channel. The encoded protein, which has a greater tendency t
flow into a cell rather than out of a cell, is controlled by
found associated with the sulfonylurea receptor SUR. Mutatio
a cause of familial persistent hyperinsulinemic hypoglycemia
an autosomal recessive disorder characterized by unregulated
Defects in this gene may also contribute to autosomal domina
dependent diabetes mellitus type II (NIDDM), transient neona
mellitus type 3 (TNDM3), and permanent neonatal diabetes mel
[provided by RefSeq]

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

[+] KEGG Pathways:
Type II diabetes mellitus [ http://www.genome.jp/dbget-bin/s

[+] GO Terms: "regulation of membrane potential" "ATP-activa
potassium channel activity" "potassium ion import" "endoplas
"voltage-gated ion channel activity" "ATP-sensitive potassiu
"plasma membrane" "microsome" "negative regulation of insuli
transport" "glucose metabolic process" "protein C-terminus b
"response to drug" "ATP binding" "mitochondrion" "potassium
"neurological system process" "response to ATP"

[+] OMIM: [600937] Link: http://www.ncbi.nlm.nih.gov/entrez/
[+] OMIM Text: The KCNJ11 gene encodes a subunit of an inwar
sensitive potassium channel. I(KATP) channels were discovere
and later found in pancreatic beta cells, pituitary tissue,
brain, and vascular and nonvascular smooth muscle. I(KATP) c
secretion and muscle contraction by coupling metabolic activ
potential ({13:Inagaki et al., 1995}).
```

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 of the gene (provided by NCBI Entrez Gene.)

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

KEGG pathways.

Gene ontology terms.

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 Na
   19685080
-- Mutations in the pore-forming K(ATP) channel subunit caus
   &amp;amp; discusses recent advances in understanding of clini
   neonatal diabetes, its underlying molecular mechanisms &a
   treatment[review] PMID: 18566517
-- mutations in the slide helix of Kir6.2 (V59G) influence t
   providing evidence that this domain is involved in Kir ch
   15583126
-- Case of an 18-month-old infant with permanent neonatal di
   activating KCNJ11 mutation who successfully transitioned
   insulin therapy to oral sulfonylurea therapy in the outpa
   18221420
-- the MDR-like core of SUR is linked with the K(IR) pore in
   12213829
-- caveolin-3 negatively regulates Kir6.2/SUR2A channel func
-- The prevalent Glu23Lys polymorphism in the potassium inwa
   (KIR6.2) gene is associated with impaired glucagon suppre
   hyperglycemia. PMID: 12196481
-- the common E23K genetic variant at the KCNJ11 gene locus
   associated with cardiovascular function PMID: 17720745
```

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

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

```
[+] 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
```

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.

```
[+] 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 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.

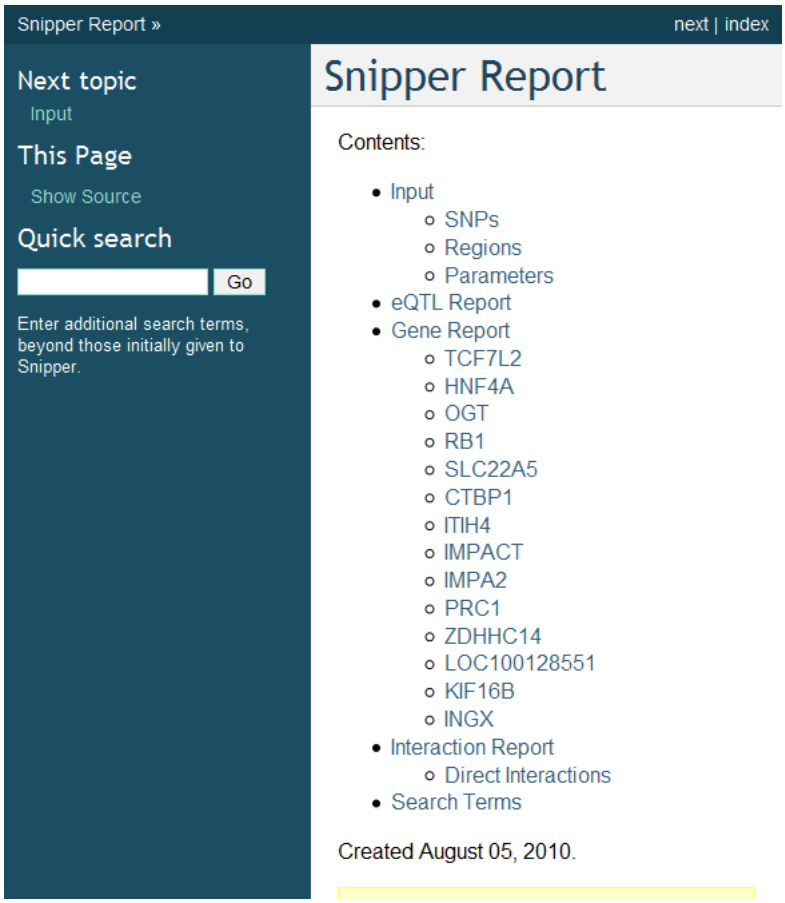
Snipper HTML output

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.

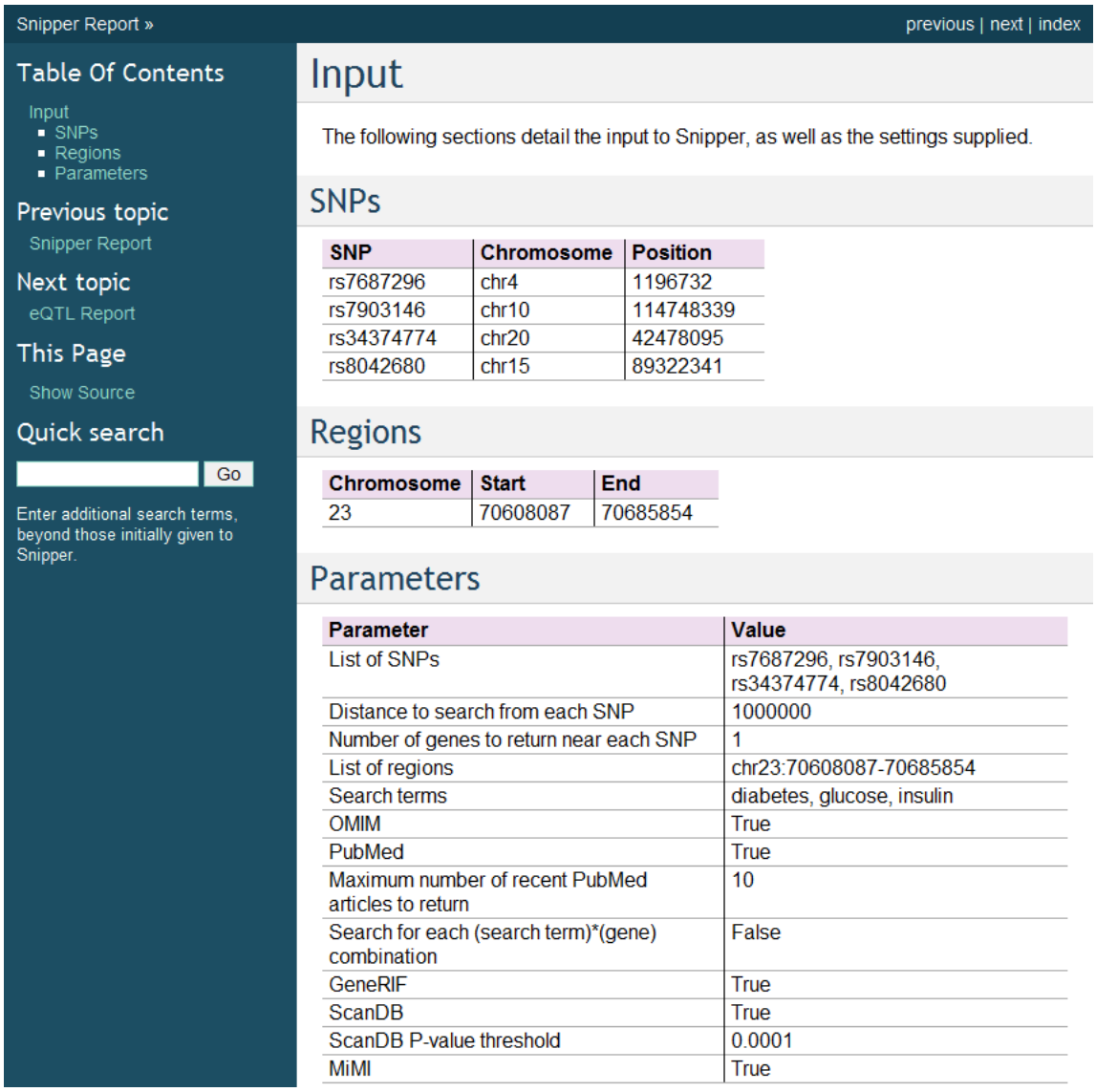
The HTML report begins with “index.html”, which has the table of contents. An image of this table (cropped to remove whitespace) is shown below.



The report contains 5 main sections:

- Input: lists all of the input SNPs and regions as given by the user, as well as all command line options that were specified
- eQTL report: contains information on genes whose expression is associated with SNPs given by the user
- Gene report: information on each gene found near the input SNPs/regions
- Interaction report: 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 | index

Table Of Contents

Input
- SNPs
- Regions
- Parameters

Previous topic
Snipper Report

Next topic
eQTL Report

This Page
Show Source

Quick search
Go
Enter additional search terms, beyond those initially given to Snipper.

Input

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

SNPs

SNP	Chromosome	Position
rs7687296	chr4	1196732
rs7903146	chr10	114748339
rs34374774	chr20	42478095
rs8042680	chr15	89322341

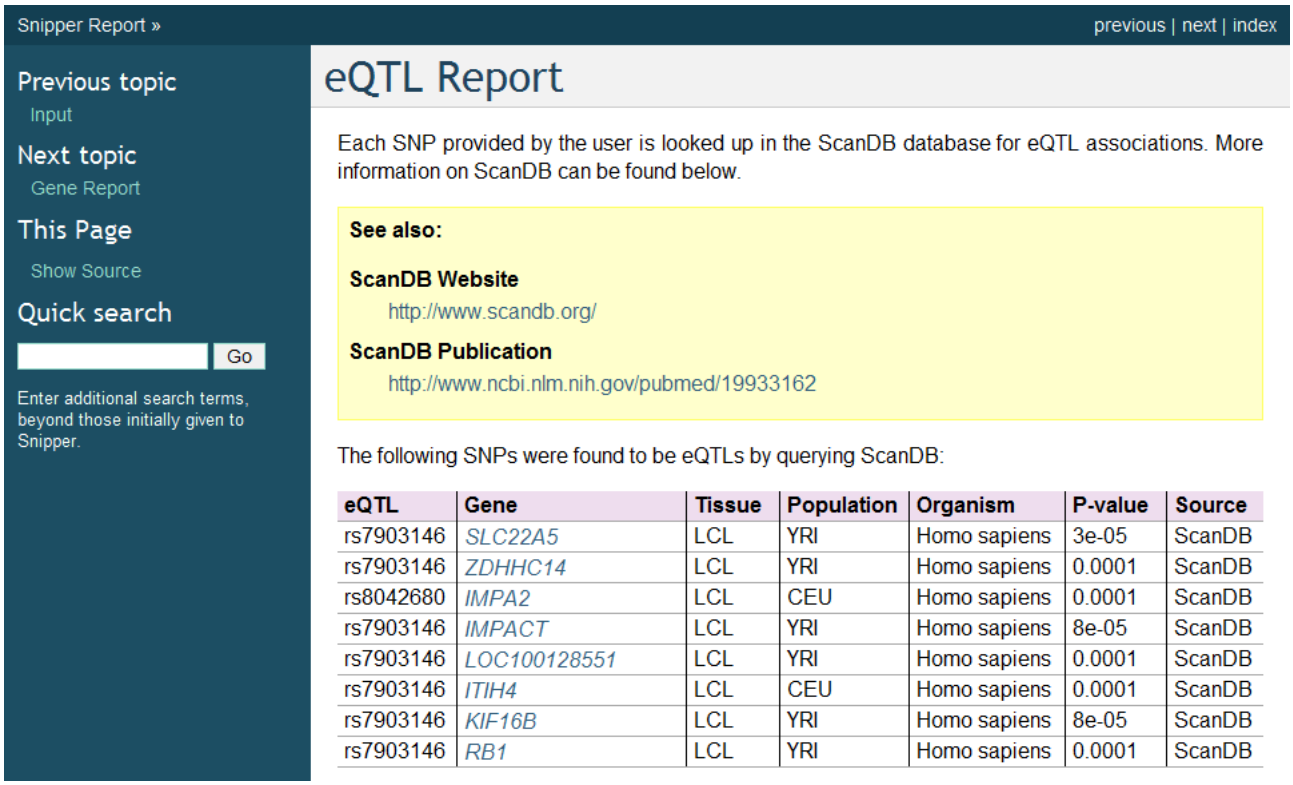
Regions

Chromosome	Start	End
23	70608087	70685854

Parameters

Parameter	Value
List of SNPs	rs7687296, rs7903146, rs34374774, rs8042680
Distance to search from each SNP	1000000
Number of genes to return near each SNP	1
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

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 | index

Previous topic
Input
Next topic
Gene Report
This Page
Show Source
Quick search
Go
Enter additional search terms, beyond those initially given to Snipper.

eQTL Report

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

See also:

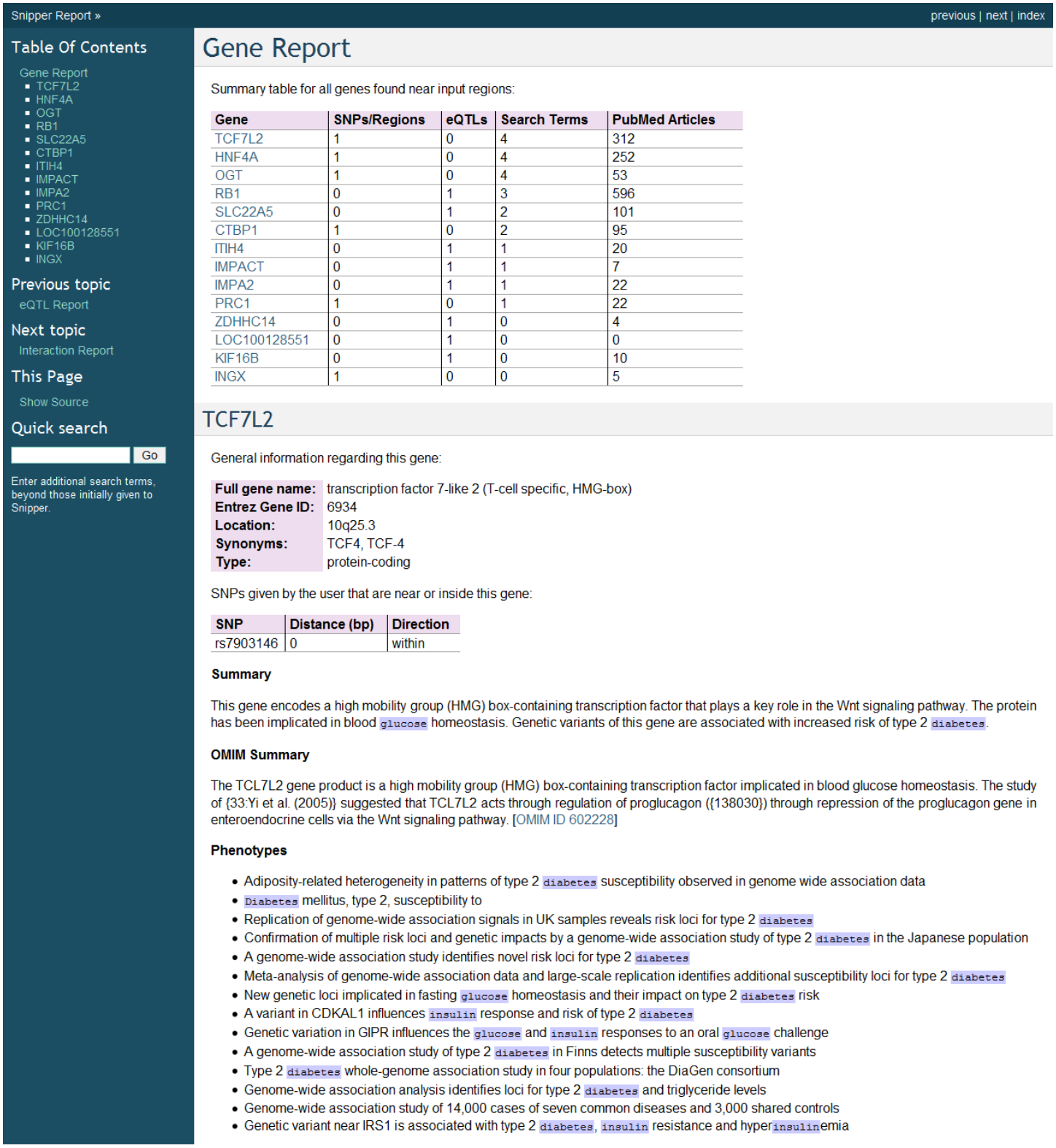
ScanDB Website
http://www.scandb.org/

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

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

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

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:



Snipper Report »

previous | next | index

Table Of Contents

Gene Report
- TCF7L2
- HNF4A
- OGT
- RB1
- SLC22A5
- CTBP1
- ITIH4
- IMPACT
- IMPA2
- PRC1
- ZDHHC14
- LOC100128551
- KIF16B
- INGX

Previous topic
eQTL Report

Next topic
Interaction Report

This Page
Show Source

Quick search
Go

Enter additional search terms, beyond those initially given to Snipper.

Gene Report

Summary table for all genes found near input regions:

Gene	SNPs/Regions	eQTLs	Search Terms	PubMed Articles
TCF7L2	1	0	4	312
HNF4A	1	0	4	252
OGT	1	0	4	53
RB1	0	1	3	596
SLC22A5	0	1	2	101
CTBP1	1	0	2	95
ITIH4	0	1	1	20
IMPACT	0	1	1	7
IMPA2	0	1	1	22
PRC1	1	0	1	22
ZDHHC14	0	1	0	4
LOC100128551	0	1	0	0
KIF16B	0	1	0	10
INGX	1	0	0	5

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.

OMIM Summary

The TCL7L2 gene product is a high mobility group (HMG) box-containing transcription factor implicated in blood glucose homeostasis. The study of {33: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

- Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome wide association data
- Diabetes mellitus, type 2, susceptibility to
- Replication of genome-wide association signals in UK samples reveals risk 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 genome-wide association study identifies novel risk loci for type 2 diabetes
- Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes
- New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk
- A variant in CDKAL1 influences insulin response and risk of type 2 diabetes
- Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge
- A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants
- Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium
- Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels
- Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls
- Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia

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:

Quick search

Go

Enter additional search terms, beyond those initially given to Snipper.

Interaction Report

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

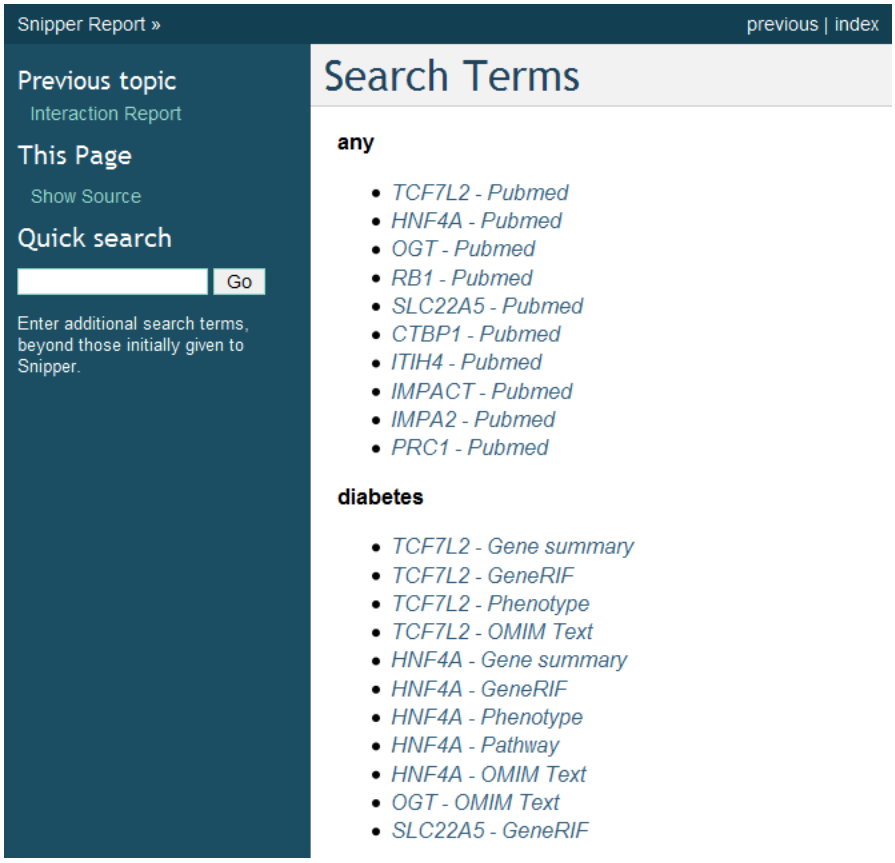
The following table lists direct interactions between the protein products of genes found near any input region:

Gene 1	Gene 2	Types	Provenance	PMIDs
RB1	*HNF4A*	• bidirectional	• BIND	
RB1	*CTBP1*	• in vivo [reverse] • Affinity Capture-MS [reverse] • bidirectional [reverse] • Invivo [reverse]	• CCSB • GRID • HPRD	• 11583618 • 16189514
RB1	*HNF4A*	• bidirectional	• BIND	
RB1	*CTBP1*	• in vivo [reverse] • Affinity Capture-MS [reverse] • bidirectional [reverse] • Invivo [reverse]	• CCSB • GRID • HPRD	• 11583618 • 16189514
CTBP1	*TCF7L2*	• PPrel • bidirectional	• BIND	
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]
RB1	*CTBP1*	• nucleus [GO:0005634]	• negative regulation of cell proliferation [GO:0008285]	• transcription factor binding [GO:0008134]
RB1	*HNF4A*	• nucleus [GO:0005634]	• transcription [GO:0006350] • positive regulation of transcription from RNA polymerase II promoter [GO:0045944]	• transcription factor activity [GO:0003700]
RB1	*CTBP1*	• nucleus [GO:0005634]	• negative regulation of cell proliferation [GO:0008285]	• transcription factor binding [GO:0008134]
CTBP1	*TCF7L2*	• nucleus [GO:0005634]		
CTBP1	*TCF7L2*	• nucleus [GO:0005634]		

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



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”
```

License

(at your option) any later version.

Snipper is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this
program. If not, see <http://www.gnu.org/licenses/>.