SNPchiMp v3.0

A multi-species database to disentangle the SNP chip jungle

Latest update: November 2014.

<u>Subscribe to the SNPchiMp rss feed to be continuously updated!</u>

Please cite:

Copyright, Disclaimer, Contacts and Important information:

The information contained in this database has not a commercial intent. Our goal is to help researchers streamline their work. The information contained in this database was supplied by publicly available content from both Illumina and Affymetrix. However, Illumina and Affymetrix are not responsible for any information contained in this website, not even in the case this database contains wrong information.

Please note that you might find a higher number of associations between commercial SNP names and rsID doing your own crosschecks over the dbSNP database. This is because *SNPchiMp* uses the publicly released information from commercial chip providers. For Affymetrix SNPs we only considered SNP name - rsID associations provided by the producer (publicly available at Affymetrix web-site). For Illumina data, we used SNP names to obtain the dbSNP associations (blasting samples of sequence information for cross-checking). Since the information is not updated in real-time, we rely on important feedback from the community to remain current.

For bug report, feedback and questions (PLEASE read the carefully this manual before sending your question) contact **ezequiel [dot] nicolazzi [at] tecnoparco [dot] org.**

General information

SNPchiMp v.3 is a free web-based tool that allows researchers to "jump" between different SNP chips and types of information in an intuitive and painless way. It has now a few new tools integrated to help researchers doing more work, but easily!

With *SNPchiMp* v.3 you can, with a few clicks:
- Know which SNPs are contained in all (or part of) commercial SNPchips currently available, for all the species contemplated.
- Get chromosome and physical position of your SNPs updated to the latest build of the genomes involved in your research (only bovine data for now!).
- Have all your product SNP/probe names converted to [rs/ss]IDs, where available.
- Have complete access to sender original information (including orientation, strand, etc).
- Obtain Interbull exchange index for all your SNP chips
- Obtain direct access to Illumina's FORWARD or TOP allele coding customized to the desired SNP chip(s), and Affymetrix's A/B coding.
- Interact with Ensembl BioMart and Variant Effect Predictor (VeP) tools directly (query and download your results without any extra work!).
- Access the open-source (or open-access for third parties) tools for basic SNP data management (like SNP allele conversion) provided in the "Tools" section.

Why *SNPchiMp*?
- Because there are two main SNPchip producers and many custom SNP chips available that have different output data options.
- Because higher density SNP chips do not always contain all the SNPs in the lower density chips.
- Because some SNPs have been referenced with different SNP names within the same community.
- Because there are and there will be updates in the reference assembly.
- Because it is difficult to find a direct link between manufacturer names and rsIDs (that are often searchable terms in all public DBs).
- Because SNP alleles in output files can reference different strands (e.g. A/B, FORWARD, TOP)
- Because for some downstream analyses, knowing the strand of the SNPs can be fundamental…
- Because with SNPchiMp you can have access to Biomart directly and extract all the genes annotated within a user-defined range up-/down-stream of your SNPs. In addition you can extract the consequence of your subset of mutations using Ensembl VeP API REST tool, now integrated entirely in the tool!

What type of analyses can benefit from the information contained in *SNPchiMp*?
- Multiple-chip / multiple-producers **genotype imputation and integration**.
- **Selection signatures** and **GWAS**: explore your results directly and link your SNPs to dbSNP and ENSEMBL (other links will be available soon).
- Easier integration of your SNPs with **sequence data**…
- **Annotation** of a subset of SNPs

Using *SNPchiMp*

This database can be interrogated entirely from the World Wide Web. This choice was made to allow using the information contained with any OS, platform, or browser available. To connect to the *SNPchiMp* simply write the following address into your browser:

http:/bioinformatics.tecnoparco.org/SNPchimp.

It is possible way of connecting directly to specific parts of the database using a browser address (e.g. helping you integrating this tool to your own programs and software!).

The web page has several menus:

Specifically:
- **Home:**
- **Info:** Well… you already got here if you're reading this!
- **Download:** In this section you can download the information available in the database for the desired SNP chip (or more than one) in the desired species.
- **Browse:** In this section you can browse a subset of SNPs in any of the available species and find out a lot of linked information inside and outside our database/webpage.
- **Data Source:** In this section you can see all the sources of the data displayed on this web tool. Please note that in general data comes directly from the producer or from dbSNP.
- **Tools:** In this menu we provide useful links to software (usually open-source) for basic SNP data management.
- **News:** In this section we'll comment the upcoming changes and progresses made on the tool. The great thing about this tool is that is evolving constantly, so ***we strongly suggest you to subscribe to the RSS FEED***.
- **Contacts:** Where you can find who we are, what we do, and how (un)handsome we are! ☺
- **FAQs:** Please read carefully these FAQs before writing an email asking help.
- **Login:** This is just for admin (us…).

Now we'll introduce in detail how to use the main two features of this tool: menus "**Download**" and "**Browse**".

A) Download:

In this section you can download some or all the information available in the database for a specific SNP chip (or more than one) for your desired species.

In fact the first choice you have to make is choosing your species. Once the choice has been made you only have to follow 3 simple self-explaining steps:

STEP 1 – Choose the chip by clicking on each commercial SNP chip available. Please note that, obviously, the number of chips available is different from species to species and that for custom chips the number of SNPs you're retrieving from our tool is lower than the number of SNPs you might have. This is simply because *SNPchiMp* only stores publicly available data.

All those SNP that are IP-protected are not in here, except if the owners explicitly share them to the whole community (which, up to now, is clearly not the case for any IP-protected SNP!). In this section you can choose 1 or more SNP chips, make your choice and go to STEP 2!

STEP 2 – Once you have made your mind of the SNP chip(s), you have to make your second choice: wether you want detailed information on the SNP chip(s) or if you want the across SNP chip table (WARNING: This latter option is off course available if you choose at least 2 SNPchips). Detailed information allows you to download... well, detailed information on your SNP chip(s). Please see STEP 3.a. for further information. The "across SNP chip table" allows you to merge the information of your SNP chips into a common table. Please see STEP 3.b. for further information.

STEP 3.a. DETAILED SNP INFORMATION: Clicking on this option, you'll have to choose exactly what you want the *SNPchiMp* tool extracts from the chip(s) you chose in STEP 1.

In detail:
- **Assembly:** With this option you can choose the assembly against which you want your chromosome and position data to be downloaded. **Please note that we coded chromosome 99 and position 0 for those SNPs without SNP-rsID link, or not mapped on the required assembly.**
  Options are: Native platform (default mapping information coming from genotyping labs), and all dbSNP builds since June 2012. For example, up to April 2014, there are other 3 assemblies available for bovine livestock: UMD3.1, BTAU4.2 and BTAU4.6 and one for Pig (Sscrofa10.2) and Horse (EquCab2.0).
- **Chromosome and position:** you can download the SNP information for either all chromosomes (default) or just a single chromosome you choose.
- **ss information (Illumina SNPs only)**: This data comes directly from Ensembl's dbSNP. By selecting this option, you download the ss-linked information for each SNP, corresponding to the SNP information of the sender used to associate the SNP name and the rsID. This is available only for Illumina SNPs, as for Affymetrix we used official (and publicly available) SNPname-rsID associations.
- **Exchange Interbull Index (for COW species)**: The official Interbull Index used to exchange gentoypes in the framework of the different genotyping consortia (e.g. InterGenomics).
- **A/B alleles forward alleles (Illumina only):** If you click on any Illumina chip, you'll have this option available with the alleles for converting A/B into an A/C/G/T alleles referencing the FORWARD strand format. The first allele corresponds to the "A" allele, the second to the "B" allele. *Please see notes below regarding this option.*
- **A/B alleles top alleles (Illumina only):** If you click on any Illumina chip, you'll have this option available with the alleles for converting A/B into an A/C/G/T alleles referencing the TOP strand format. The first allele corresponds to the "A" allele, the second to the "B" allele.
- **A/B alleles (Affymetrix only):** If you click on any Affymetrix chip, you'll have this option available with the alleles for converting A/B alleles into A/C/G/T alleles. The first allele corresponds to the "A" allele, the second to the "B" allele.

Once you have selected the options desired, click on the "QueryMe" button at the bottom right corner of the browser window. A box will be displayed with the first 25 hits of SNPs and all the information you chose.

Notes:
- Original SNP names, rs IDs and the chosen chromosome and position are included by default.

- To download the file, you have to click on the "Download" section, choosing the delimiter (comma separated or tab separated).
- Files will be automatically compressed before downloading.
- Usually you get more SNPs than the total number of SNPs for each chip. This is because the file includes also cross-reference SNPs (e.g. same SNP named in two -or more- different ways).
- Differently from Illumina TOP/BOT allele coding, **Forward allele coding** is, by definition, a changing-over-time allele coding. We do our best to keep track of the latest official allele coding(s) released by the producers. Please (Please!) note that the forward allele coding for 2 chips even of the same producer might have different Forward coding!! (e.g. Illumina LD and Illumina LD v.1.1 have 8 SNPs with different forward coding!).

STEP 3.b. Across SNPchip table:
With this option you can download an across-SNPchip table with presence/absence information for all chips you desire.
If you click on "ALL chip(s) SNPs" you will be downloading a table with SNPnames and 1's or 0's indicating the presence or absence, respectively, of all SNPs in the SNPchips you chose.
If you click on "Only SNPs in common" you'll be downloading a list of *only SNPs in common* between the SNPchip(s) you have chosen.

B) Browse:

In this section of the *SNPchiMp* you can browse a subset of SNPs. For example, you can browse a list of significant SNPs of a GWAS, or a selection signature analysis, among others. This section consists in four steps:
1) Choose your species
2) Choose the assembly required.
3) Choose the information required: similarly to the "Detailed SNP information" section in "Download" menu, you can choose one or more of the following: Present in chip (if the SNPs are present in N SNP chips, you'll find N rows/SNPs) or ss information. Please see the information in the "Download" menu for more details.
4) Write your query. Here you have 4 different ways of performing your query (only one at the time is allowed):
   a. By Position: Syntax as in Ensembl (chromosome:start..stop). In this case you can choose to obtain all the information required for a specific chromosome segment.
   b. By rs ID: If you have a list of rs IDs, comma separated, you can access all the above information just by copying it to this section and clicking "QueryMe". Syntax for this section requires the "rs", but is NOT case sensitive. Thus, "rs" and "RS" are equivalent.
   c. By ss IDs: As in b., but including a list of (not case sensitive) ss IDs.
   d. By SNP IDs: as in b., but including a list of commercial SNP IDs (e.g. SNP IDs as in Illumina or Affymetrix chips). Please note that also this section is NOT case sensitive, thus BovineHD0100000005 can be also written bovinehd0100000005.

Once you have a preview of your results, you can directly hit the external web pages links, for detailed information on the chosen SNPs. This feature is available only for the first 20 SNPs listed in the query.

B.1. Querying your data using Ensembl BioMart and VeP (Variant Effect Predictor)

The *SNPchiMp* capability of interrogating Ensembl's BioMart data mining tool has been recently put in place. Ensembl recognized the utility of this tool, allowing us to proudly bear the E!mpowered tool symbol. This option is available for any species/assembly in Ensembl.

You can now search all genes in both forward and reverse strands +- for a user-defined Kbp of the SNPs browsed. The two-step procedure will guide you through. Please be aware that the *SNPchiMp* will re-direct you to a "ready-to-run" BioMart web page with a set of default options. You can modify some of the options by clicking on: "Attributes" on the left-side of the page. You only have to click on **RESULTS** (top left, third option) on the BioMart web-page and… eureka! Use BioMart functionalities to download your data if you wish…

Similarly, also Ensembl VeP tool can be accessed directly from this tool. Obviously, no range needs to be provided and the results are exactly what you would get interrogating VeP *on* Ensembl (without the fancy graphs… ☹).

Please note:
- As for download, original SNP names, rs IDs and the chosen chromosome and position are included by default.
- external websites are referred to their own assembly, so please check if the assembly is the same you're analysing your data!!!
- you can download the data from our tool and use it, for example, for a query in Biomart (again, please check carefully the assembly version used by other databases/browsers/etc!!!)

Have fun!!!

Subscribe to the SNPchiMp rss feed to be continuously updated!