

SNPflow 1.3.0

User Manual

2013-02-14

Contact:

<http://genepi-snpflow.i-med.ac.at>

hansi.weissensteiner@i-med.ac.at

GENEPI
INNSBRUCK

System requirements

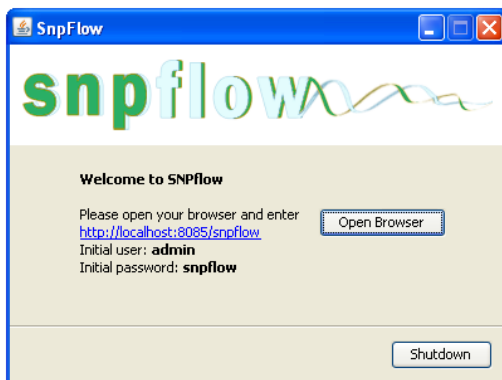
SNPflow runs on operation systems where the Java Virtual Machine (<http://java.com/en/download/index.jsp>) is installed correctly. We tested SNPflow on Windows XP, Windows 7, MacOS X and Linux-environments. To run the Server Version of SNPflow a MySQL(<http://www.mysql.com/downloads/>) database and Apache Tomcat (Version 6 or higher <http://tomcat.apache.org/>) have to be installed.

Installation of SNPflow

SNPflow can be downloaded from the projects website <http://genepi-snpflow.i-med.ac.at>. Here you find further information and the latest news / updates. You can either download the **standalone version** (snpflow-1.3.0-win.zip or snpflow-1.3.0-linux.zip) if you want to use SNPflow on a single computer or contact us to get the **Server version** for networks where several users can access the same MySQL database.

Standalone Version (zip files)

By downloading the SNPflow standalone version, you need to ensure that Java (<http://java.com/en/download/index.jsp>) is installed correctly on your operating system.



On **Windows**, extract the zip file and the client is ready for use, by clicking the snpflow.exe.

On **Linux**, perform the following steps:

1. Extract the zip file: `unzip snpflow-1.3.0-linux.zip`

2. Give execute access to the bash script: `chmod 775 snpflow.sh`

Double-click the exe file / or run the snpflow.sh on Linux for starting the application. After starting the Tomcat server and setting up of the internal database, the start-screen appears.

Now you can open the Browser of your choice and enter the URL: <http://localhost:8085/snpflow>. We tested SNPflow on the most common web-Browsers.

Server Version (war File)

The Server Version for MySQL / Apache Tomcat comes in a war file. This war file has to be copied in the Apache Tomcat webapps folder. For doing so you need to have Apache Tomcat installed and running, besides a MySQL database. The following steps are needed:

1. If MySQL is not installed → install (<http://www.mysql.com/downloads/>)
2. Open MySQL and create a user and the database (see the sql script from the website)
3. Create tables from script: `mysql -u your_mysql_user -p < workflow.sql` (get from website)
4. If not installed: install Apache Tomcat 6 or higher
5. Copy .war file in the folder `../tomcat/webapps`
6. Open the Browser and type: <http://localhost/snpflow>

Hint: you can run the Standalone Version also as a Server Version, by sharing the IP address of the computer on which SNPflow runs to others in the intranet. When this computer is shut down, all open connections to SNPflow are interrupted. Example URL: <http://192.168.10.10:8085/snpflow>

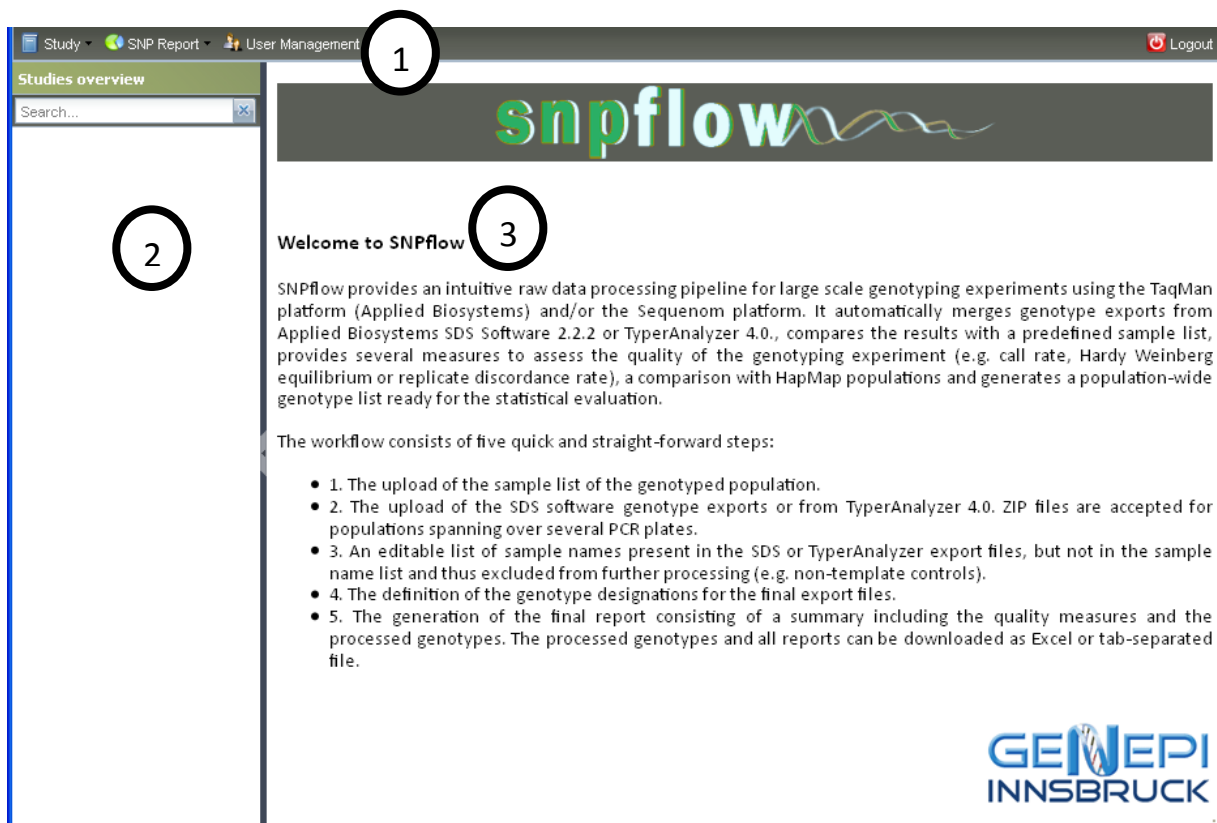
First start of SNPflow

By opening a browser and inserting SNPflows URL, a login window appears and the initial credentials for username and password are requested:

Username: **admin**

Password: **snpflow**

If everything was configured properly, SNPflow welcomes you with some basic informations for working with the application.



The screenshot shows the SNPflow web application interface. The interface is divided into three main areas:

- 1. Menu-Bar:** Located at the top, it contains a navigation menu with items like 'Study', 'SNP Report', and 'User Management', along with a 'Logout' button.
- 2. Navigation Panel:** Located on the left side, it contains a 'Studies overview' section with a search bar.
- 3. Details Window:** The main content area on the right, which displays a 'Welcome to SNPflow' message and a list of five workflow steps.

The workflow consists of five quick and straight-forward steps:

- 1. The upload of the sample list of the genotyped population.
- 2. The upload of the SDS software genotype exports or from TyperAnalyzer 4.0. ZIP files are accepted for populations spanning over several PCR plates.
- 3. An editable list of sample names present in the SDS or TyperAnalyzer export files, but not in the sample name list and thus excluded from further processing (e.g. non-template controls).
- 4. The definition of the genotype designations for the final export files.
- 5. The generation of the final report consisting of a summary including the quality measures and the processed genotypes. The processed genotypes and all reports can be downloaded as Excel or tab-separated file.

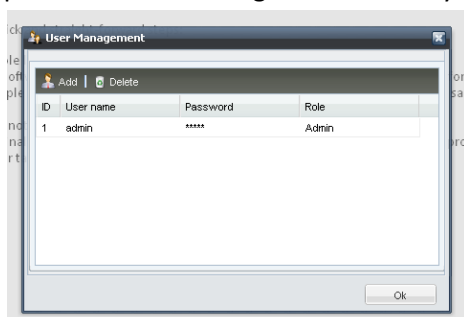
The GENIEPI INNSBRUCK logo is visible in the bottom right corner.

The screen is structured in 3 areas:

1. Menu-Bar
2. Navigation Panel
3. Details Window

User Management

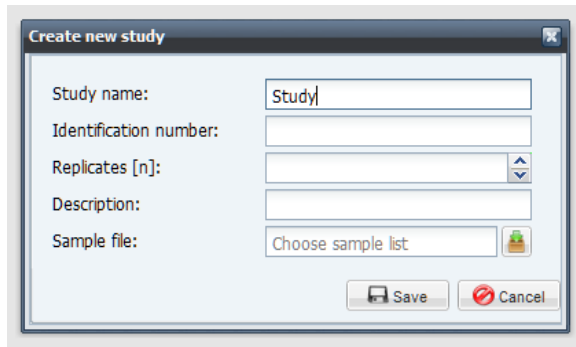
One of the first things one may do when starting SNPflow for the first time, is changing the default password or creating new users. By clicking on “User Management” in the Menu bar the small popup-window appears where all existing users are shown in a list. To add new users click on the “Add” button. A new empty row appears and the username, password as well as the user role have to be inserted. The 3 different roles are



Admin (full rights), **Lab** (right to create study, show and export results) and **Analyst** (right to show and export results).

Create new study

By starting SNPflow for the first time, one will notice that the navigation panel in the left area is empty. Later on, the different studies are listed in a tree structure as you may know from the Windows Explorer. A new study can be created by clicking on “**Study**” in the menu-bar and selecting “**Add...**”. A new window opens where several informations are needed:



1. Study name
2. Identification Number (optional)
3. Number of replicates in the study
4. Short description
5. Sample file, the list of IDs used in the study

The sample file is a text-file with one column labeled **sample_id** and the sample IDs in each row.

By clicking on **Save** the study is created and the sample list is stored in the database.

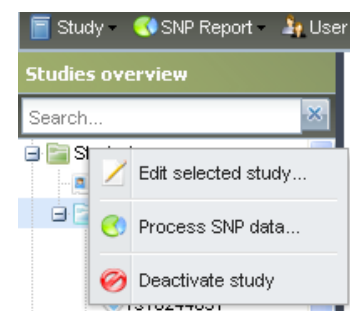
Example file for samples (see sample_file.txt):

```
sample_id
4081
4306
4432
4475
```

Edit a study

A study can be edited by clicking on the corresponding study and selecting **Study – Edit...** or by clicking on the study with the right-mouse button. A context-menu appears with several options:

1. **Edit selected study...** (same as **Study – Edit...** in the menu bar)
2. Process SNP data → see “Creating a new SNP report”
3. Deactivate study (self explanatory → see “Deactivate a Study”)



Create new SNP report

To import SNP data from either SDS files (ABI 7900HT) or TyperAnalyzer 4.0 files (Sequenom) a study has to be selected and the SNP report Wizard has to be opened through the menu-bar “**Study**” – “**Process SNP data**” or by right click on a study and selecting “**Process SNP data**”.

The SNP report Wizard is a 4 step guide where the following steps are required:

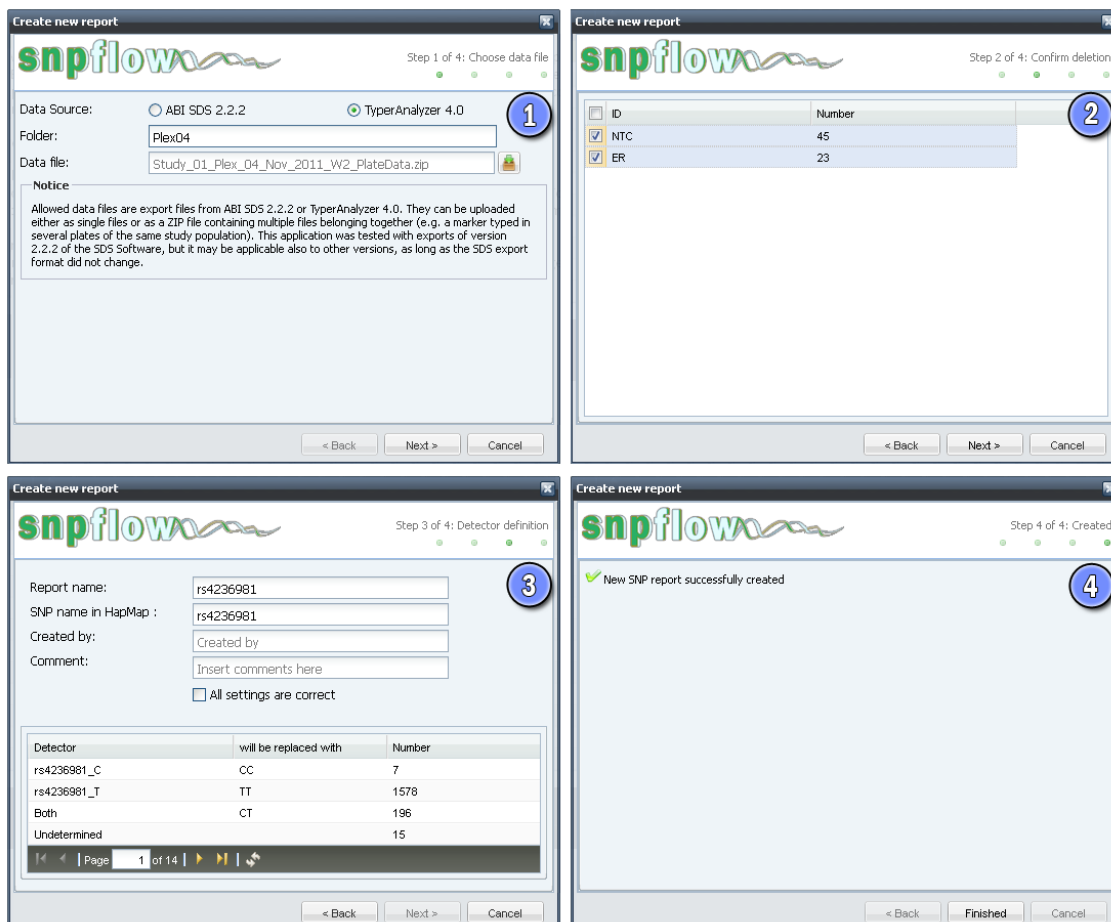
1. **Upload SDS or TyperAnalyzer file**
2. **Remove double entries**
3. **Check each SNP**
4. **Database import**

The wizard starts with the **upload** of the output files from the data analysis software (SDS or TyperAnalyzer 4.0). These files can be uploaded either as a single file or, in the case that a study spans several 384 well microplates, as a zip archive (step **1**).

SNPflow then compares the IDs found in the merged data set with the IDs stored in SNPflow’s study definitions and lists unrecognized sample IDs for removal (step **2**; first data checkpoint). Subsequently, SNPflow detects the genotype designations used in the data files and suggests the corresponding genotypes for the final result file.

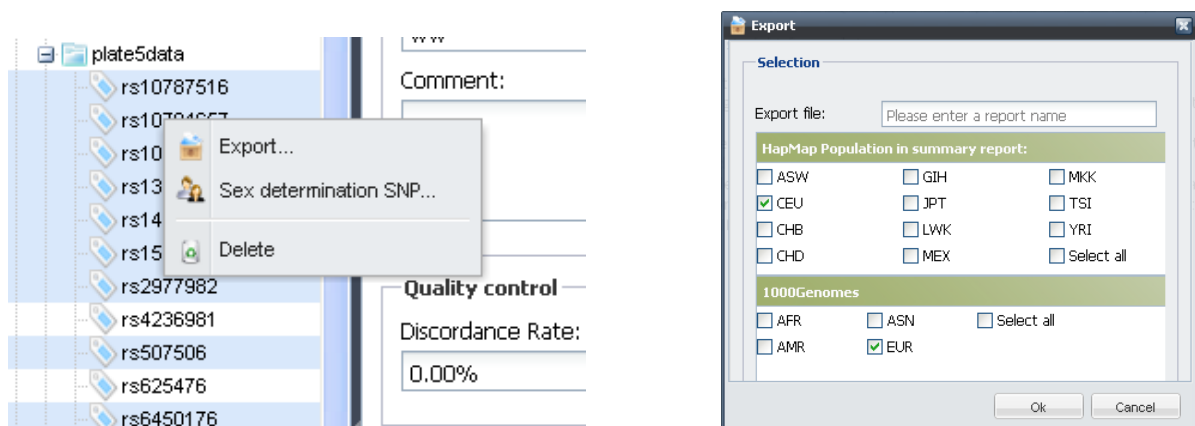
Step **3** converts the found genotype designations to standard genotypes and allows entering the name of the result, the operator name, an optional comment and the rs number of the SNP in HapMap. This is useful in case that a SNP ID has changed between the current dbSNP release and the dbSNP release used in HapMap, as happened for some SNPs. For convenience, the result name is, however, automatically pre-set as HapMap lookup. In multiplex assays pagination allows navigating through the single SNPs. For quality reasons a manual confirmation of each conversion is required (Checkbox “All settings are correct”). Please note that, since the detector definitions in the SDS software are arbitrary, we assume that the detector designations adhere to the following pattern: “<arbitrary detector name>_<allele>” (e.g. “rs328_G” or “LPL-rs328_G”). SNPflow works also with any other kind of detector definitions, but the automatic genotype detection might require manual editing. This issue is not present with MassARRAY files.

4 Finally, all QC values are calculated and the data is stored in the database.



Export SNPs

The finished genotype lists and the QC values are exported in single output files and in a merged Microsoft Excel file containing data from several independent SNP assays within a study. To do so, select the relevant SNPs by holding **CTRL-Key** and selecting the SNPs with a left mouse click. The SNP assays to be exported can be selected in any desired combination. To export the selected SNPs, either **right-mouse button** or the menu: **SNP Report -> Export** has to be selected. Moreover, the user can select, which HapMap or 1000Genomes populations have to be exported, thus avoiding the useless export of incomparable populations. In order to enhance the readability of the output files, values exceeding certain thresholds (e.g. HWE Test <0.05) are automatically highlighted in the Excel files.



Delete SNPs

If the user has the Admin or Lab role assigned, he is able to delete SNP reports, if desired. To do so, the SNPs have to be chosen by holding CTRL-button and selecting the SNPs with left mouse click. When the SNPs are selected, either go to the menu-bar SNP report and choose Delete... or use the context menu's entry by clicking with the right mouse button on one of the SNPs and choose Delete.

Deactivate studies

Projects are not meant to be deleted. It's only possible to deactivate studies. To do so, select the study in the navigation panel and by right click with the mouse select "Deactivate Study". Deactivated studies can be set to active again by right click on the deactivated project and selecting "Activate study".

Sex determination SNP

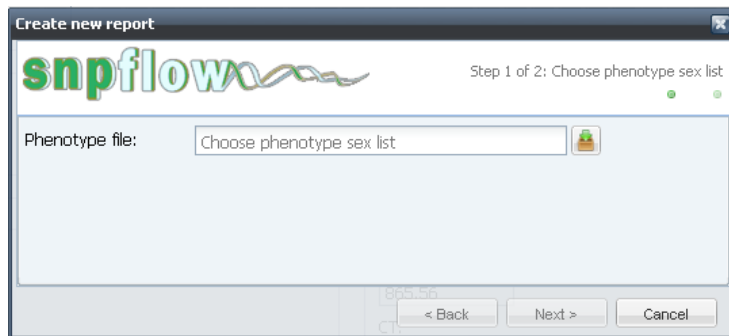
Since SNP markers for sex determination are commonly used in genetic epidemiology to check study populations for sample mix up, SNPflow provides an analysis module to check the genotyped samples for sex concordance. Any SNP stored in SNPflow can be selected by mouse right-click (same menu as export) and set as "**Sex determination SNP**". This opens a dialog box, where the user can upload a list of sample IDs with the respective sex phenotype for comparison and assign a specific sex to the genotypes.

The sex coding in the uploaded file can be selected freely and defined in the dialog box. SNPflow then compares the genotypic sex with the uploaded phenotypes and reports discordance in a separate result tab. The Sex-SNP is highlighted in green.

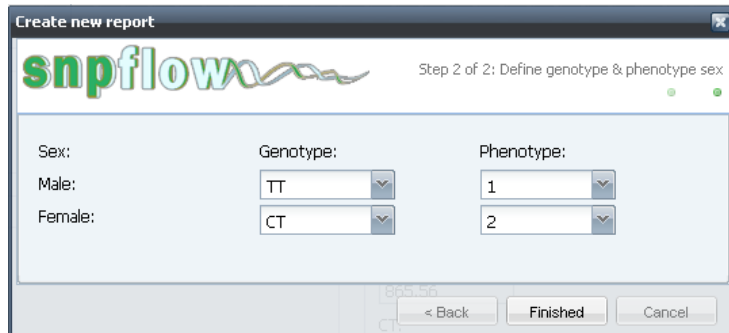
Example for sex phenotypes (see phenotype_sex_list.txt):

```

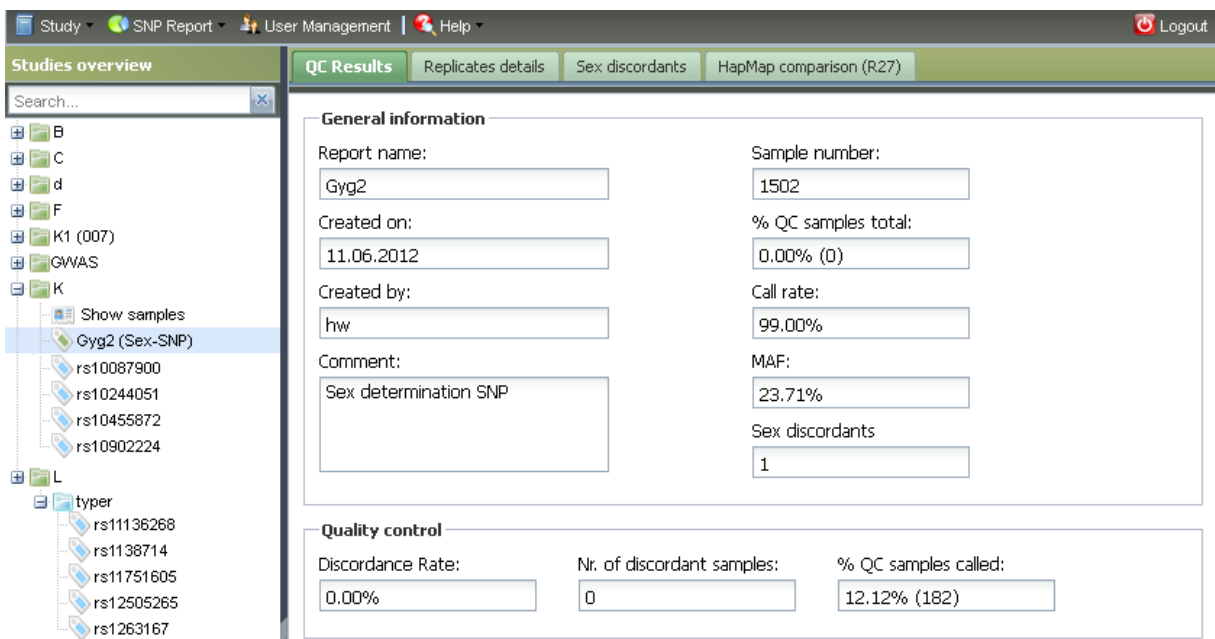
sample_id sex
4081      male
4306      female
4436      female
...
    
```



Step 1: Import sex phenotype list (Columns in text file have to be called sample ID and sex)



Step 2: Define the genotype and the phenotype to their value



The imported Sex-SNP has an additional field: sex discordants and a new tab: sex discordants (see below)

Study | SNP Report | User Management | Help | Logout

Studies overview | QC Results | Replicates details | **Sex discordants** | HapMap comparison (R27)

Search...

- ⊕ B
- ⊕ C
- ⊕ d
- ⊕ F
- ⊕ K1 (007)
- ⊕ GWAS
- ⊕ K
 - Show samples
 - Gyg2 (Sex-SNP)
 - rs10087900
 - rs10244051

| Sample ID ▲ | Phenotype | Genotype |
|------------------------------------|-----------|-------------|
| Sex discordants (1 Samples) | | |
| K319080478 | 1 (male) | TT (female) |