

## TUTORIAL

### Getting started with SNPator - Free accounts

You have received an e-mail from the **SNPator** system administrators containing a **username** and a **Password**. You are eager to begin using **SNPator** to analyze your own data. Your genotypes and info about the samples that were genotyped are waiting in your computer.

#### 1. Logging into the system

- Using any web browser, connect to the following URL:

http://www.snpator.org

Inside the SNPator home page, a login window will appear

**Login into SNPator**

User:

Password:

**Login**

You are wellcome to enter the system. Just remember that user and password are case sensitive.

[I don't have SNPator User / forgot my password](#)

- You should now type the **Username** and **Password** you have received and press the "**Login**" Button. Take care of entering upper and lower case correctly.

If everything is alright, you should be by now inside **SNPator** and your web browser screen should look like:

The screenshot shows the CEGEN Bioinformatics Division website in Microsoft Internet Explorer. The browser title is "Bioinformática CEGEN - Microsoft Internet Explorer". The address bar shows the URL: <http://bioinformatica.cegen.upf.es/privat/principal/index.php>. The website header includes the CEGEN logo, the text "Bioinformatics Division", and "Centre Nacional de Genòtip". A navigation menu on the left lists various functions: Study Selection, Filters, Data Management, Genotypes Validation, Descriptors Administration, Data Retrieval, Basic Analysis, Genomic Analysis, Disease Analysis, Public Downloads, Jobs, User Results, and Help. At the bottom of the menu, it says "ADVANCED connected" and "Logout". The main content area displays "Study : carlos\_01". The Windows taskbar at the bottom shows several open applications and the system clock indicating 16:33.

## 2. Entering data

Now, the very first thing that you need to do is to put your data inside **SNPator**. We are showing you now one of the several ways of doing this. Once you will become an advanced user of SNPator you will be able to choose the way that best suits your needs.

This data will usually consist of:

- A list of SNPs with genomic information associated to each SNP.
  - A list of samples with phenotypic information associated to each sample.
  - The genotypes obtained from the analysis of those SNPs in those samples.
- If you go to the "**Public downloads**" section in the **left menu**, you'll see something like this:

### Index of /public/downloads

<a href="#">Name</a>	<a href="#">Last modified</a>	<a href="#">Size</a>	<a href="#">Description</a>
 <a href="#">Parent Directory</a>		-	
 <a href="#">ExcelFiles/</a>	29-May-2006 10:31	-	
 <a href="#">Plate_Example/</a>	06-Jun-2006 09:41	-	

*Apache/2.2.0 Server at bioinformatica.cegen.upf.es Port 80*

- Access the "**ExcelFiles**" folder and download to your computer the ZIP file that you will find inside. This ZIP contains the latest version of the following Excel files.

 results_list.xls	34.304
 results_matrix_samples.xls	37.888
 results_matrix_snps.xls	40.448
 samples.xls	44.544
 snps.xls	50.176

- Save those files into your computer.

### 2.1. Uploading SNP information

First bit of data that gets into the system is usually SNP information.

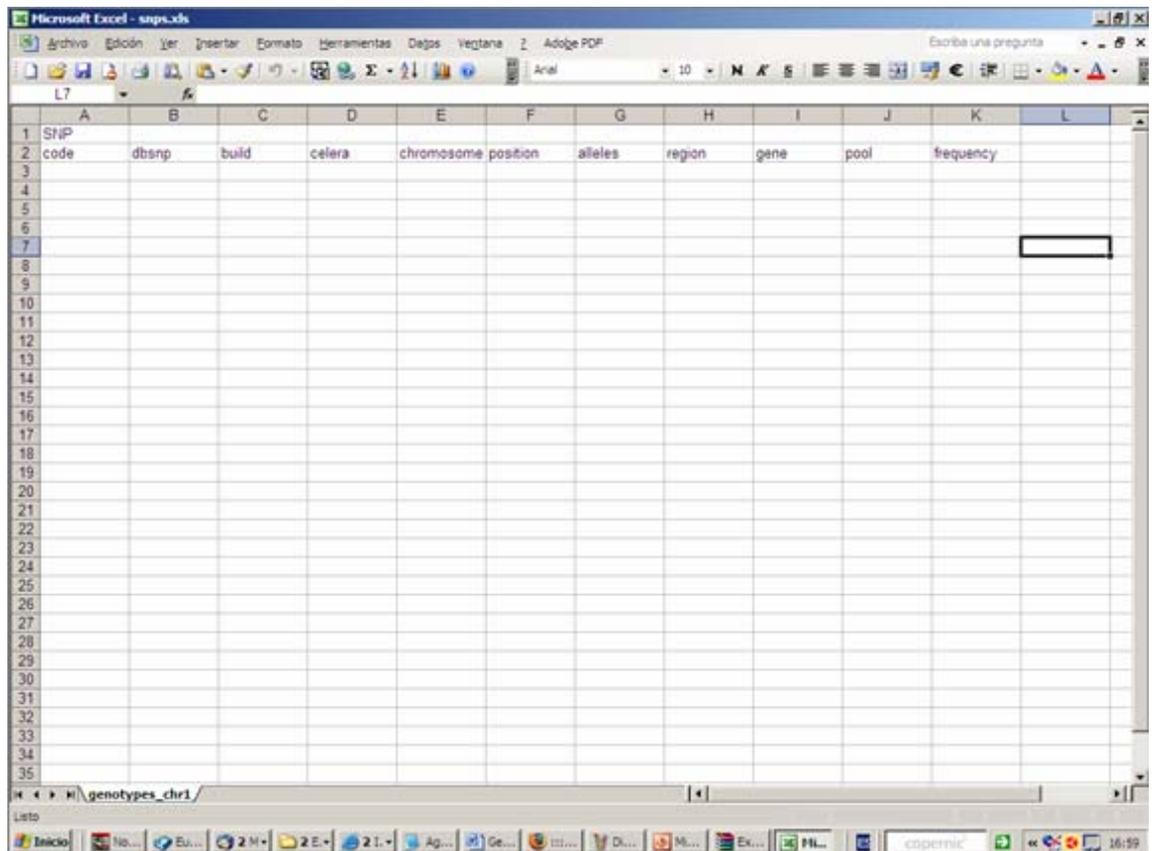
### - What if I don't have any SNP information to upload?

You can bypass this section. **SNPator** will give you the possibility of automatically retrieving information from dbSNP database for your **SNPs** later on, provided that you enter the rs# SNP code when uploading the genotypes.

It is recommended, however, to enter the **SNP** data, even if only the **SNP code** alone, as an added quality control procedure.

- Open the "**snps.xls**" file from your "**ExcelFiles**" zip:

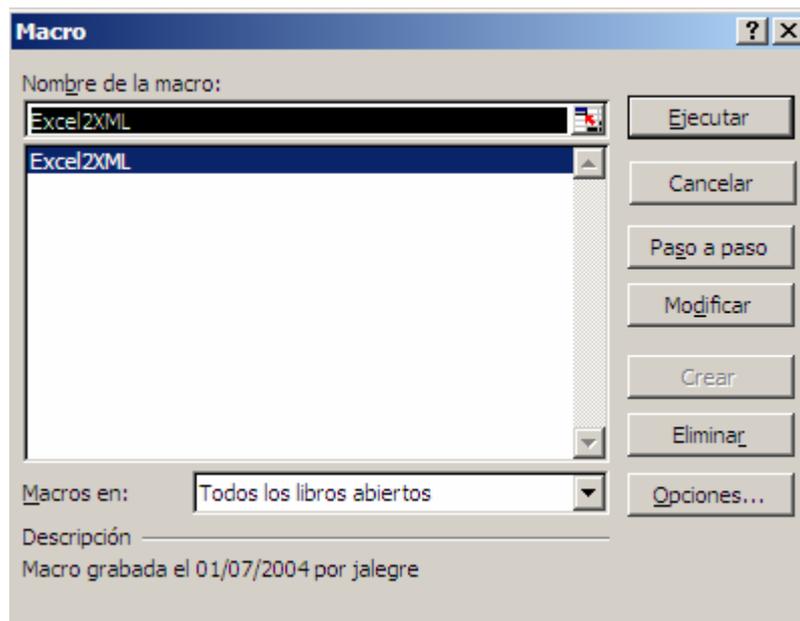
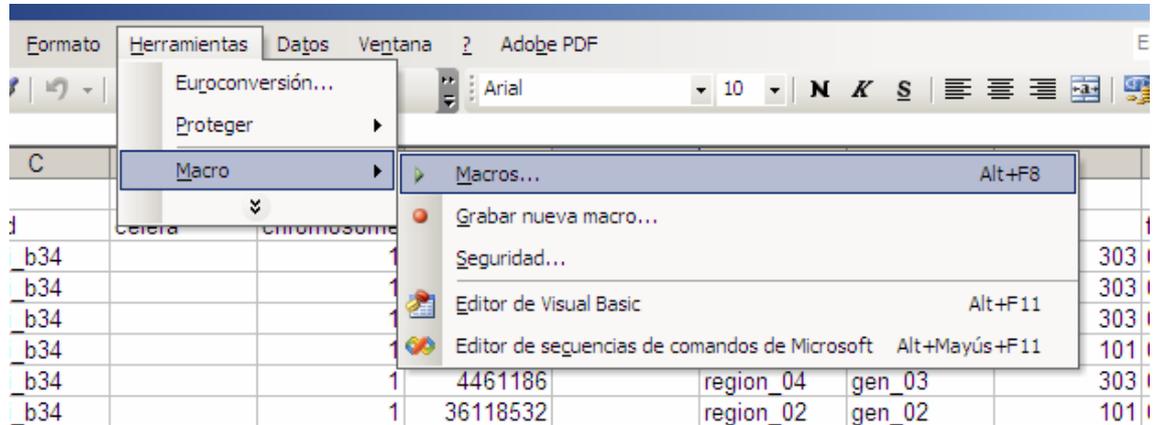
(It is important that you have your "run macros" option activated in your Excel application)



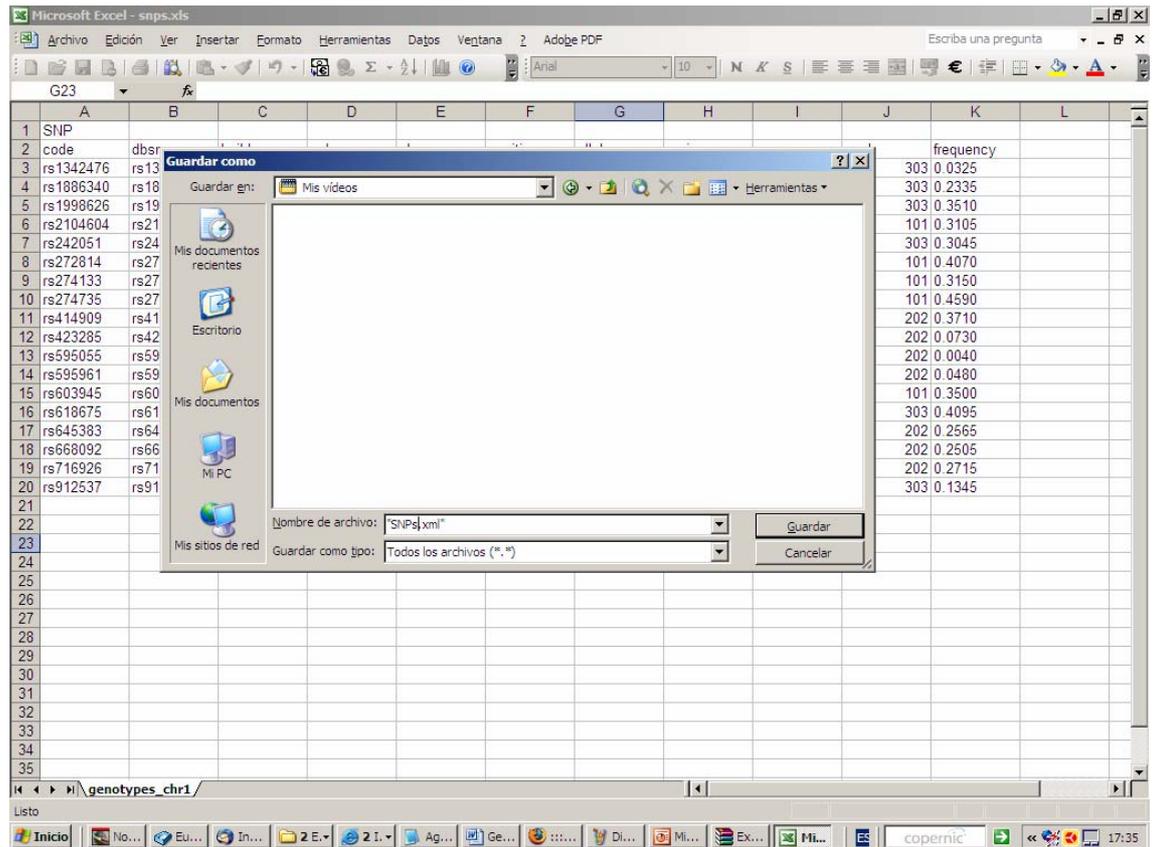
- Put your **SNP** information on the file while keeping the two first original lines .  
Only "**code**" field is compulsory. Codes have to be the same that you will use in the genotypes file.

Remember that "**build**", "**chromosome**" and "**position**" fields can be automatically retrieved afterwards in **SNPator** if the rs# dbSNP code is used in the "**code**" or "**dbSNP**" fields.

- Once all the **SNP** data is entered, you must run a built-in Macro named "Excel2XML" that generates a XML file containing all the information from the excel file (don't worry if you don't know what a XML file is, it's just a way of coding information that SNPator understands) :



- You give a name to the XML file:



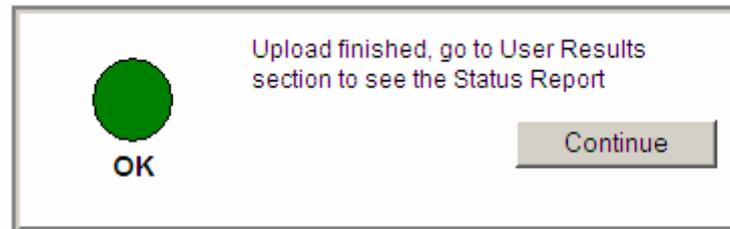
- and save it to your local drive.
- Go now to the **menu at your left** and open the **"Data management / SNPs / Upload SNPs"** section:

### Data Management / Upload / SNPs

SNPs XML File

Overwrite SNP if already exists

- Enter the XML file you just created with the Excel macro and press the **"Upload"** button. If everything is OK, you should get:



- Press "**Continue**" and it's done.
- You can check the "**User results**" section in order to get a report of possible errors during the uploading process.
- To check the **SNPs** introduced, you can use the "**Show SNPs**" button in the "**Data retrieval / SNPs / Samples**" section:

**Data Retrieval / SNPs or Samples**

---

Code	DBSNP	Build	Celera	Chromosome	Position	Alleles	Region	Gene	Pool	Frequency
rs1342476	rs1342476	ncbi_b34		1	34554620		region_05	gen_02	303	0.0325
rs1886340	rs1886340	ncbi_b34		1	34571724		region_05	gen_02	303	0.2335
rs1998626	rs1998626	ncbi_b34		1	34579291		region_05	gen_02	303	0.3510
rs2104604	rs2104604	ncbi_b34		1	101809619		region_01	gen_01	101	0.3105
rs242051	rs242051	ncbi_b34		1	4461186		region_04	gen_03	303	0.3045
rs272814	rs272814	ncbi_b34		1	36118532		region_02	gen_02	101	0.4070
rs274133	rs274133	ncbi_b34		1	36009454		region_02	gen_02	101	0.3150
rs274735	rs274735	ncbi_b34		1	35928799		region_02	gen_02	101	0.4590
rs414909	rs414909	ncbi_b34		1	4508439		region_04	gen_03	202	0.3710
rs423285	rs423285	ncbi_b34		1	4501274		region_04	gen_03	202	0.0730

## 2.2. Uploading sample information

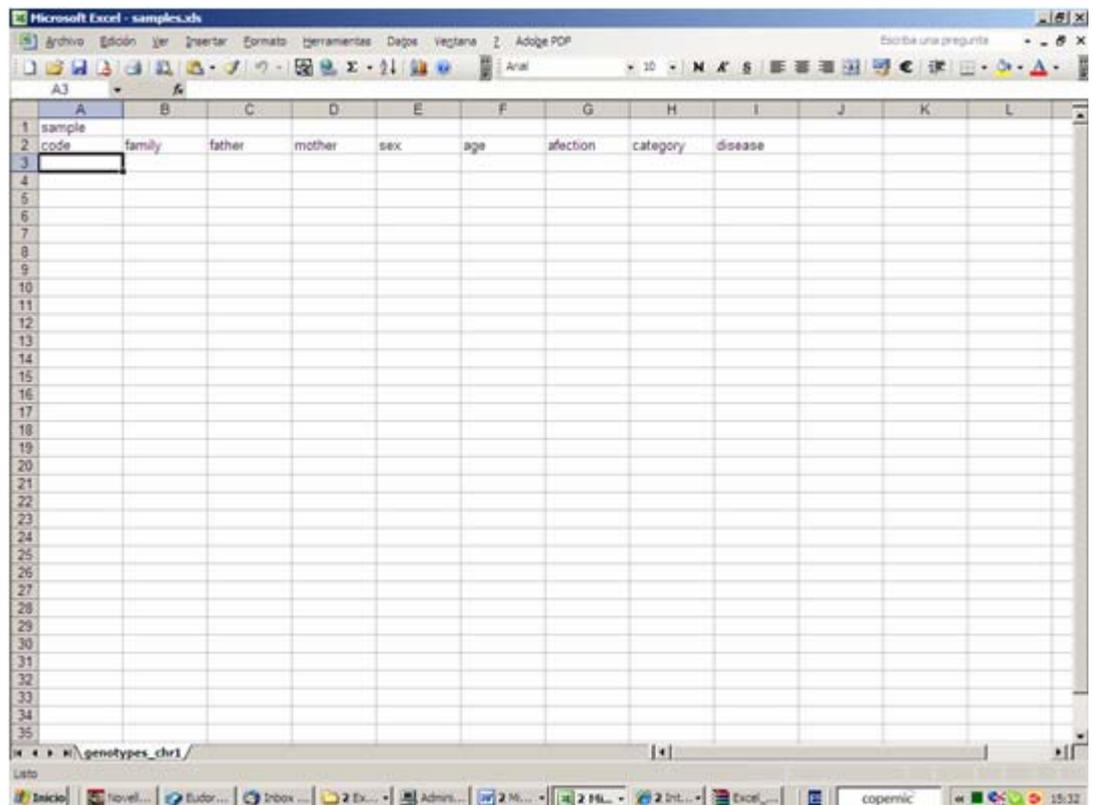
Now you proceed to upload information about your samples. This is usually their ID codes and their case/control status, but many more kinds of information can be added.

**What if I don't have any sample information to upload?**

You can bypass this section. However, it is recommended that, as an added quality control procedure, sample data are entered, even if only sample codes are available.

- Open the "**samples.xls**" file from your "**ExcelFiles**" zip:

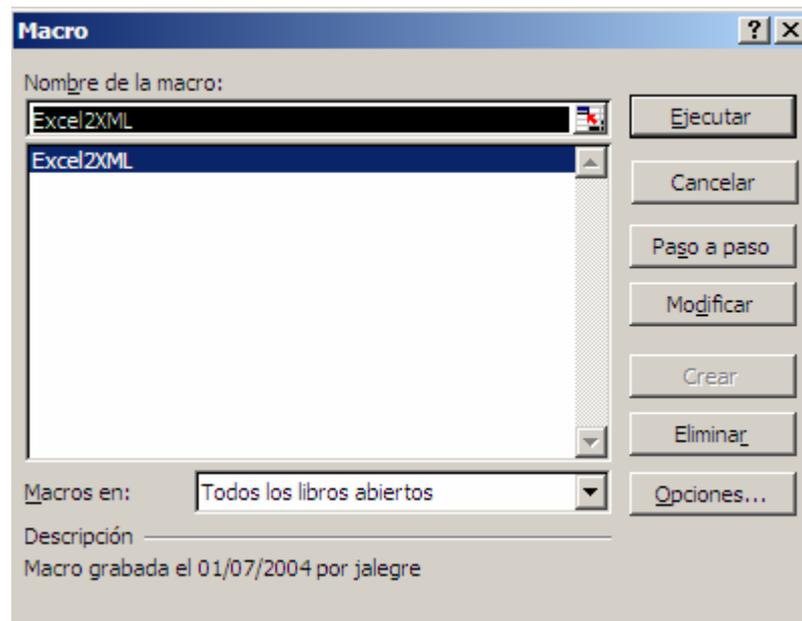
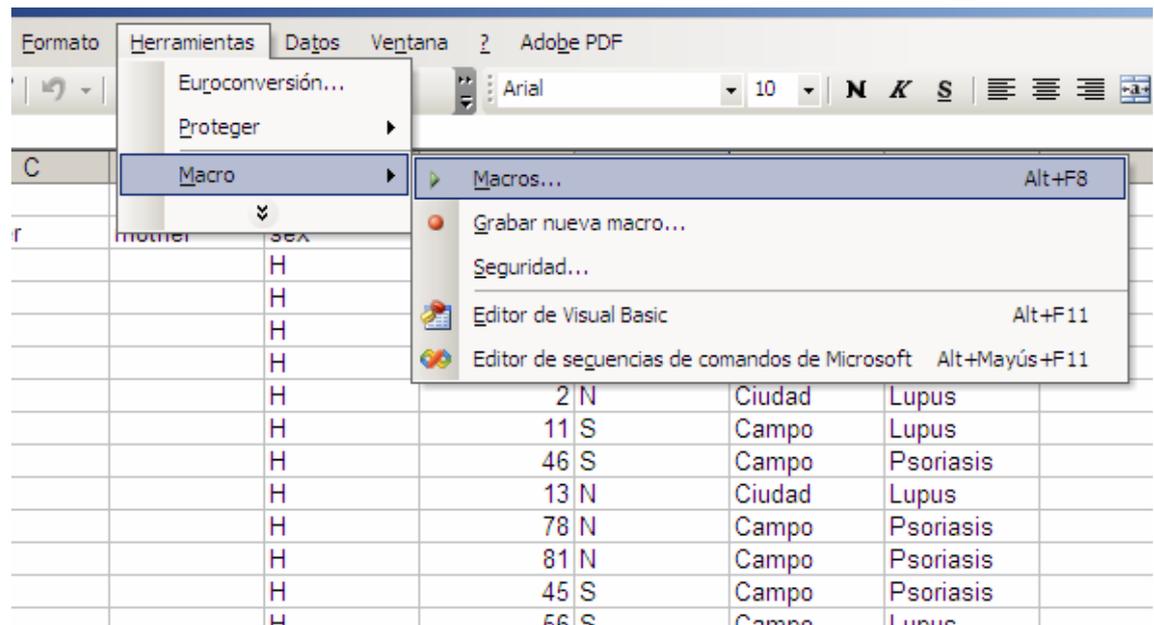
(It is important that you have your "run macros" option activated in your Excel application)



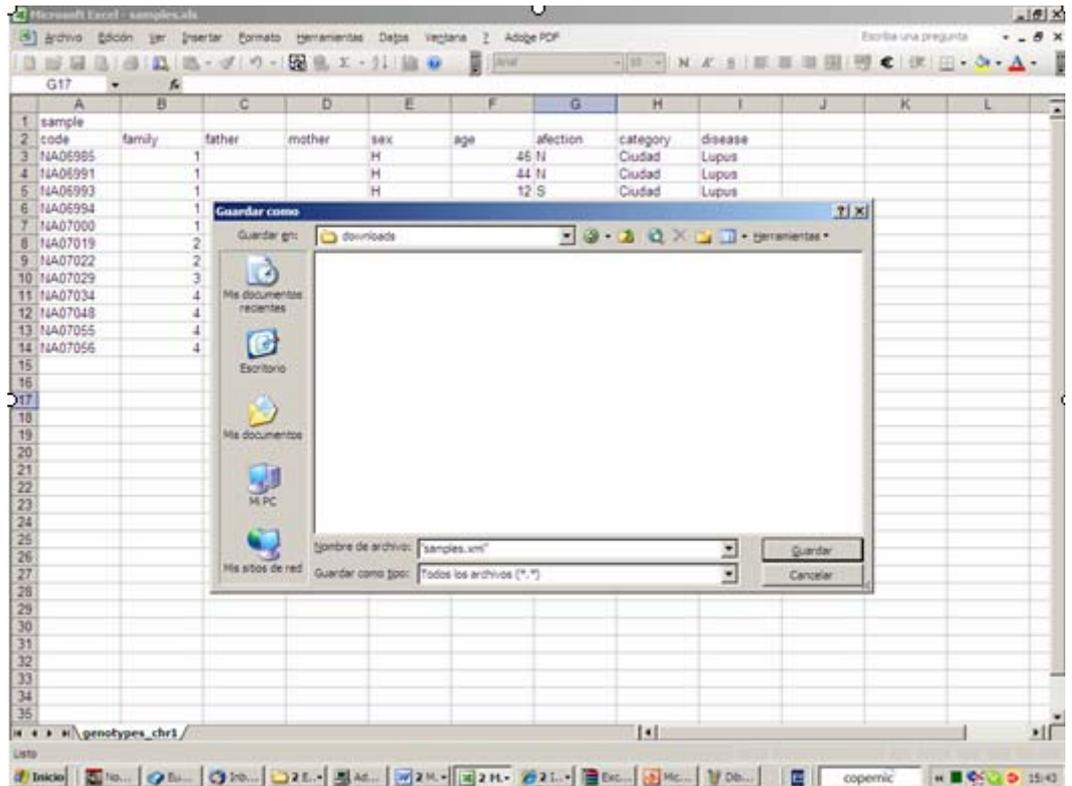
- Put your **Samples** information on the file. "Pasting" it will usually be enough.

Only "**code**" field is compulsory. "**code**" has to be the same that you will use in the genotypes file.

- Once all the sample data is entered, you must run a built-in Macro named "**Excel2XML**" that generates a XML file containing all the information from the excel file:



- You give a name to the XML file:



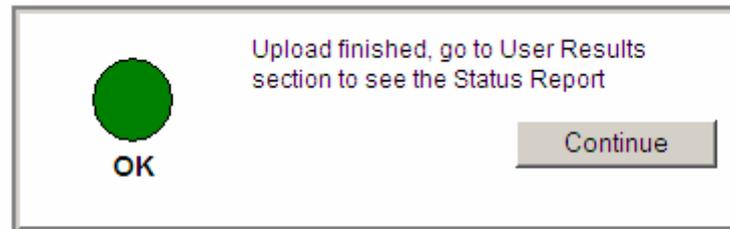
- and save it to your local drive.
- You now access the menu at your left in **SNPator** and go to the **"Data management / Samples / Upload Samples"** section:

### Data Management / Upload / Samples

Samples XML File

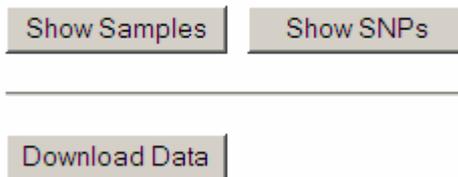
Overwrite sample if already exists

- Enter the XML file you just created with the Excel macro and press the **"Upload"** button. If everything is OK, you should get:



- Press "**Continue**" and you are done.
- You can check the "**User results**" section in order to get a report of possible errors during the process.
- You can check the Sample information you introduced with the "**Show Samples**" button in the "**Data retrieval / SNPs / Samples**" section:

### Data Retrieval / SNPs or Samples



Code	Family	Father	Mother	Sex	Age	Afeccion	Group	Disease	continent	country
NA06985	1			H	46	N	Ciudad	Lupus		
NA06991	1			M	44	N	Ciudad	Lupus		
NA06993	1	NA06985	NA06991	H	12	S	Ciudad	Lupus		
NA06994	1	NA06985	NA06991	M	8	N	Ciudad	Psoriasis		
NA07019	2			M	11	S	Campo	Lupus		
NA07022	2			M	46	S	Campo	Psoriasis		
NA07029	3			H	13	N	Ciudad	Lupus		
NA07034	4			H	78	N	Campo	Psoriasis		
NA07048	4			M	81	N	Campo	Psoriasis		
NA07055	4	NA07034	NA07048	M	45	S	Campo	Psoriasis		

Download Data

### 2.3. Uploading Genotypes information

Finally, we get to the key point. You have genotyped your SNPs in your samples and you want to get the resulting genotypes into the system.

**SNPator** allows you to run quality control procedures in your genotypes in order to make sure that genotypes uploaded into the system are consistent and don't have, for example, different results for control repetitions in your plate. Thus, uploading of genotypes is a two-step process:

1. All **genotypes** are uploaded in an intermediate table called "**Raw Genotypes**". In there, several quality control processes can be run.
2. Once your **genotypes** are free of contradictions are transferred to the "**Genotypes**" table where all analysis and data formats will be performed.

- Open the "**results\_list.xls**", the "**results\_matrix\_samples.xls**" OR the "**results\_matrix\_snps.xls**" file from your "**ExcelFiles**" zip.

All 3 archives have different data structure but they are equivalent and you can choose whichever you prefer.

- Paste the **Genotypes** information into the file

genotypes	NA06985	NA06991	NA06993	NA06994	NA07000	NA07019	NA07022	NA07029	NA07034	NA07048	NA07055	NA07056
rs5982858	GG	G	AG	AG	G	G	AG	AG	AG	AG	GG	G
rs11152548	AA	A	AA	AA	A	A	AA	AA	AA	AA	AA	A
rs2316292	AA	A	AA	AA	A	A	AA	AA	AA	AA	AA	A
rs311173	CC	C	CC	CC	C	C	CC	CC	CC	CC	CC	C
rs311183	AA	A	AA	AA	A	A	AA	AA	AA	AA	AA	A

Names of **SNPs** and samples in this file have to be identical to those entered in the **SNPs** and **Sample** files.

- Once all the **genotypes** are entered, you must run a built-in Macro that generates a XML file containing all the information from the Excel file. This is exactly what you did before for the **SNPs** and **samples** data.
- You should now access the menu at your left in **SNPator**, and go to the "**Data management / Genotypes / Plates / Upload**" section:

Data Management / Genotypes / Plates / Upload Help

Technology

Plate

Genotyping date  Processing date

File

Translation File

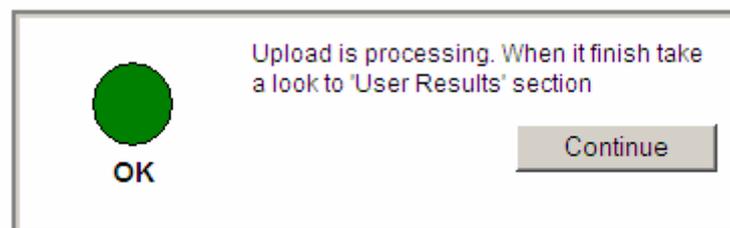
- Fill in the following fields:

**"Technology"**: "XML"

**"Plate"**: A label to identify the set of genotypes that you are entering now.

**"File"** The XML file you just have created with your data.

- Press "**Upload**"



- In the same menu at the left in the **SNPator** screen, you can now go to the "**Data management / Genotypes / Study status**" section to check that the data have been actually uploaded.

## Data Management / Genotypes / Study Status

[Help](#)

Plate	User	Date	Technology	Action	
Set 01	advanced	2007-01-09 17:50:01	XML	<a href="#">Information</a>	<a href="#">Remove</a>
Set 02	advanced	2007-01-09 17:50:14	XML	<a href="#">Information</a>	<a href="#">Remove</a>

Now, all **genotypes** are stored in the "**raw\_genotypes**" table.

- Go to "**Data management / Genotypes / Data Transfer**" section. You have several options in there:

## Data Management / Genotypes / Data Transfer

- Create associated SNP / sample if don't exist!
- Allow multiallelic SNPs
- Allow transfer with warnings
- Keep validity

- If you have previously entered your **SNP** and **sample** data, just press the button "**Transfer**". If you have not entered your **SNPs** or **samples**, the you must activate the option "**Create associated SNP / sample if they do not exist!**" before pressing "**Transfer**".
- You can check the "**User results**" section in order to get a report of possible errors during the process.

This transfer is not possible if there are contradictions in the genotypes that you uploaded. If this is the case, the report of the "**User Results**" section will tell you, and you will have to go to the "**Data Management / Genotypes / Quality control**" section to correct your **genotypes** before trying a "**Data transfer**" again.

### 3. Validating data

Now, we have all our data inside **SNPator** and we are only one step away from beginning our analysis.



### 3.2. Single kind of data

If all your **SNPs** belong to the same kind of chromosome (autosomes or pseudoautosomal regions, Y or Mitochondrial) you only have to follow these steps:

- Go to the "**Genotypes validation / Perform validation / Global**" section:

#### Genotypes Validation / Perform Validation / Global

Help

a) All genotypes belong to :

b) Validate using SNP and Sample Data:
 

Autosomes  
 Y or MT  
 X

Male :

Female :

c) Validate all genotypes with no change.

Ok

- Select, the appropriate setting in the "**All genotypes belong to**" menu.
- Press "**OK**" and check the report in the "**User results**" section to see if everything was OK
- You can check that all your data have been validated in the "**Genotypes validation / Validation Status**" section.

### 3.3. Mixed data

If you have some **genotypes** which have to be in single form ("A") and others which have to be in double form ("AA") depending on their chromosome location and on the sex of the sample, you will need **SNP** and **sample** information before performing your **validation**.

- Go to "**Data retrieval / SNPs / Samples**" and press the "**Download Data**" button.

- Go to "**User results**" section and look, in the report generated, whether you already have the required information ("**Chromosome**" in the "**SNPs**" table and, if you have some **SNPs** in the X chromosome, "**sex**" in the "**samples**" table)
- If necessary, fill up the "**chromosome**" and "**position**" fields in the "**SNPs**" table using the "**Data Management / SNPs / Load dbSNP info**". To be able to perform this you need the dbSNP code of the chromosomes either in the "**Code**" or "**dbSNP**" fields.
- To fill in the "**sex**" field in the "**samples**" table you will have to reload the **sample** data adding this information. To do this, follow the instructions of the "**uploading sample information**" section of this tutorial just activating the "**Overwrite sample if already exists**" option when you reach the "**Data management / Samples / Upload Samples**" step.
- Once all required information is inside, go to "**Genotypes validation / Perform validation / Global**" section:

### Genotypes Validation / Perform Validation / Global

[Help](#)

a) All genotypes belong to : [Select item] ▾  
[Select item]  
Autosomes  
Y or MT  
X

b) Validate using SNP and Sample Data.

Male : [Select item] ▾  
 Female : [Select item] ▾

c) Validate all genotypes with no change.

- Select option "**(b)**" declaring which value you associate with male and which with female.
- Press "**OK**" and look in the "**User results**" section for the report generated by this procedure. In there you will be able to see if everything was OK
- You can check that all your data have been validated in the "**Genotypes validation / Validation Status**" section.

#### 4. Working with SNPator

Now you are ready to work with SNPator. Every analysis and data retrieval option comes with a help page that will guide you through the process.

In case of further doubts about the system, please contact our support team at [bioinformatica.cegen@upf.edu](mailto:bioinformatica.cegen@upf.edu)