Genetic **S**core **C**reat**or**

What is GenSCor?

#Big principles, objectives, what can be done with it

GenSCor is a tool for allowing non informaticians users to create their own scoring system. It's design for annotated genetic data, especially big spreadsheets data with thousands of lines and dozens of columns. With this tool, you can dynamically create a set of rules that you can apply on your data files to score each mutations and rank them.
The purpose of GenSCor is to help you define your own score, and to use this score to quickly rank your data so that the perfect variation is always #1.
You can also use GenSCor for visualizing data, but it's not the main purpose of it, and you might want to keep using excel or libreoffice instead.

What is not GenSCor?

#what can't be done with GenSCor

GenSCor is not an annotator, so it will not generate new data nor add new informations to your data.
GenSCor is not a scoring system, we don't know your data, we can't teach you how to rank them. There is no such thing like GenSCor_score, because GenSCor is here to help you create your own score which will perfectly fit your own data, not for being used as a static reference.

Why should I use GenSCor ?

#In which case GenSCor is helpful

Here is some of the case we think GenSCor for :

_ If you have annotated genetic data, and you painfully sort each of your files with excel.

_ If you have specific data, and none of the existing scoring system produce perfect results on your dataset.

_ If you can't maintain an up-to-date bibliography because there is too much different score.

_ If you want to keep track of how your analyse data on a certain so that you can reanalyse them in the future.

_ If you have a wide clinical area and you want to be able to quickly switch strategy

If you feel concerned by one or more of the above statement, you may find helpful resources in GenSCor

How powerful my config must be ?

 # and what should I install to run it

GenSCor need python3, and that's all. To launch it, just double click on it. You don't need powerful computer to run it. In fact, we tried on a 10 year old XP computer and it worked well.

Quickstart :

#I don't have time to read, where do I start ?

You can start by loading your first dataset with menu > load data.

It will display the first 15 rows of your data, and three blank rules. You can modify this rules to fit your needs. Of course you can add more rules, because three won't probably be enough.
If you have trouble with the format of your file, you can modify the parsing parameters to fit your format. For example, the default separator is tabulation, so if yours is a comma, you will need to precise it.

A rule is defined by six points :

_ The first one is the coefficient, which is a number from 0 to 5. This will coefficient the score modification. Obviously, a rule with a 0 coefficient has no influence on the score. This can be useful if you want to test your scoring system without a rule for a short time.

_ The second is a column name, based on the header of the data file, or a default one. Entry is a dropdown of all possibles choices. This will select the column where the condition will take place.

_The third is the operator. Entry is a drop-down of eight choices covering a large range of situation. Four operators are exclusively numerical ("<",">", "≥", "≤"). Operators "=" and "≠" can be used for exact matches with numbers or words, and operators "Contains" and "don't contains" are for words only.

_ The fourth is the value to compare with. The entry is a combobox, this mean that all values of the selected column will be proposed to avoid common mistakes, but you can also choose to tape in your own value, which is useful if you want to set numerical cutoff for example.

_ The fifth is the direction in which the score has to move if the condition is valid. It must be "Increase" or "Decrease". By this mean, you can use a rule to highlight interesting lines, or to lower the ranking of uninteresting ones.

_ The sixth and the last is the value to add to the score (or subtract if the direction is "decrease"). This value will be multiplied by the coefficient.

Any modification of the rules will be dynamically apply on the score display in the data section, but the ranking will not be actualised until you press refresh, to avoid useless calculation, and to easily see the results of any rule modification.
If two rules apply, both will affect score.

Detailed functionality :

#each of the menu action
File menu :
load data : for browsing your computer and choose a data file to open. Must be a flat type, like csv, tsv or txt.
save : for saving the set of rule under a chosen name.
export : export the data in the state their are : sorted with the score
load rules : load a ruleset, data will be scored and ranked
cancel : cancel last major action, like row deleting.
no visual mode : Enter the no visual mode. This mode is used to apply a ruleset on a one or more files and automatically export them without opening them in the data viewer window.

There is three buttons : the first is for loading a save containing a ruleset. The second is to load one or more file. You must hold ctrl for selecting multiples files. And the third one launch the analysis. Files will be export under the same name with “export.tsv” as a suffix. If a file already exist, a number will be add to avoid overwrite.
quit : This one is pretty straightforward : quit the software. A confirmation will be asked for avoiding mistakes.
Edit menu :
Add line : For adding an empty line to the actual rule set. Three format exists, that add one, three or five empty lines.
Add multiple : for adding a multiple rule. Such rules require multiples conditions to be true so that the score is changed. For example, you can require that A = “1” AND that B is greater than 2 for increasing score by ten points. You can add as many conditions as you want.
Sort ascending : Change the sort order, lowest score will be in the first page
Sort descending : Change the sort order, greatest score will be in the first page.
menu parameters :
Opening parameters : this action open a window for changing globals parameters :

_ Separator : Define the character used as a separator in the data file. Some classique separator character are present, such as ; or ,, but you can add one if your data format is exotic.

_ Header : Must be check if there is a header in the data (i.e. : a line with column names). If so, this line will be used to name the columns. If there is no header, default name will be used.

_ number of line to jump : The number of non data line (except the header if any) to jump before accessing the data. This can be very helpful for file with metadata in the firsts lines.

_ language : The language to be used in the interface. GenSCor currently support english and french. If you want to help translating in other language, feel free to email me.

_window mode : The way the windows must be displayed. If this option is checked, the two windows (data and rules) will be displayed separately. This is useful if you have two screens or more. If this option isn’t checked, the two windows will be displayed in a unique window.

Good to know :
#things that you might want to know if you plan to use GenSCor regularly
The saved ruleset are text files.You can modify them with any text editor, and exchange them with other user; but keep the formatting well, or you might experience troubles.
The param file is also a flat file, and it will be created where you launch GenSCore, so if you move the script, you may have to generate it again by resetting the parameters.
Among the best practice, we recommend to save a lot, as with any software. We also recommend to test the result with differents values.

FAQ :

- **Why can’t I launch it ?**

Make sure you run python 3+ and you have tkinter module installed. This module is present in the default installation of python, but can be missing on some system. Also,

- **Trouble with saves, line suppression or parameters ?**

The saving system, the parameters system and the line suppression function need to write on the disk. If you have nos space left, or no permission you may experience troubles. You can also have conflict with some antivirus, OneDrive or Dropbox, depending on your configuration.

- **Why can't I see my data ?**

Make sure the parsing parameters are matching your data, in parameters -> opening parameters.

- **My parameters weren't saved ?**

Parameters are saved in a file in the directory from where you launch the script. If you change the launching directory without the param file, default parameters will be used. You can move the param file or even switch in multiples copies if you have multiples data format.

Contact :

For any bug report or suggestion, feel free to email me :
quentin.riche.piotaix@univ-poitiers.fr