ParseSNP User Guide. Version 1.0.2 May 2026

Disclaimer

ParseSNP is provided for educational and research purposes only.
It is **not** a medical device, diagnostic tool, or substitute for professional medical advice.

- **No medical advice:** The risk assessments, odds ratios, and other outputs are basedf on published research and your raw DNA data. They do **not** constitute a medical diagnosis, prediction, or treatment recommendation.
- **Consult a doctor:** If you have concerns about any genetic condition or your health,f you should speak with a qualified healthcare professional. Do not disregard or delay seeking medical advice based on anything generated by ParseSNP.
- **No warranty:** ParseSNP is supplied “as is”, without any warranty of merchantability, accuracy, or fitness for a particular purpose. The software may contain errors or omissions.
- **No liability:** To the maximum extent permitted by law, the author(s) and contributors shall not be held liable for any damages or harm arising from the use or misuse of this software, including but not limited to emotional distress, financial loss, or medical outcomes.

By using ParseSNP, you acknowledge that you understand and accept these terms.

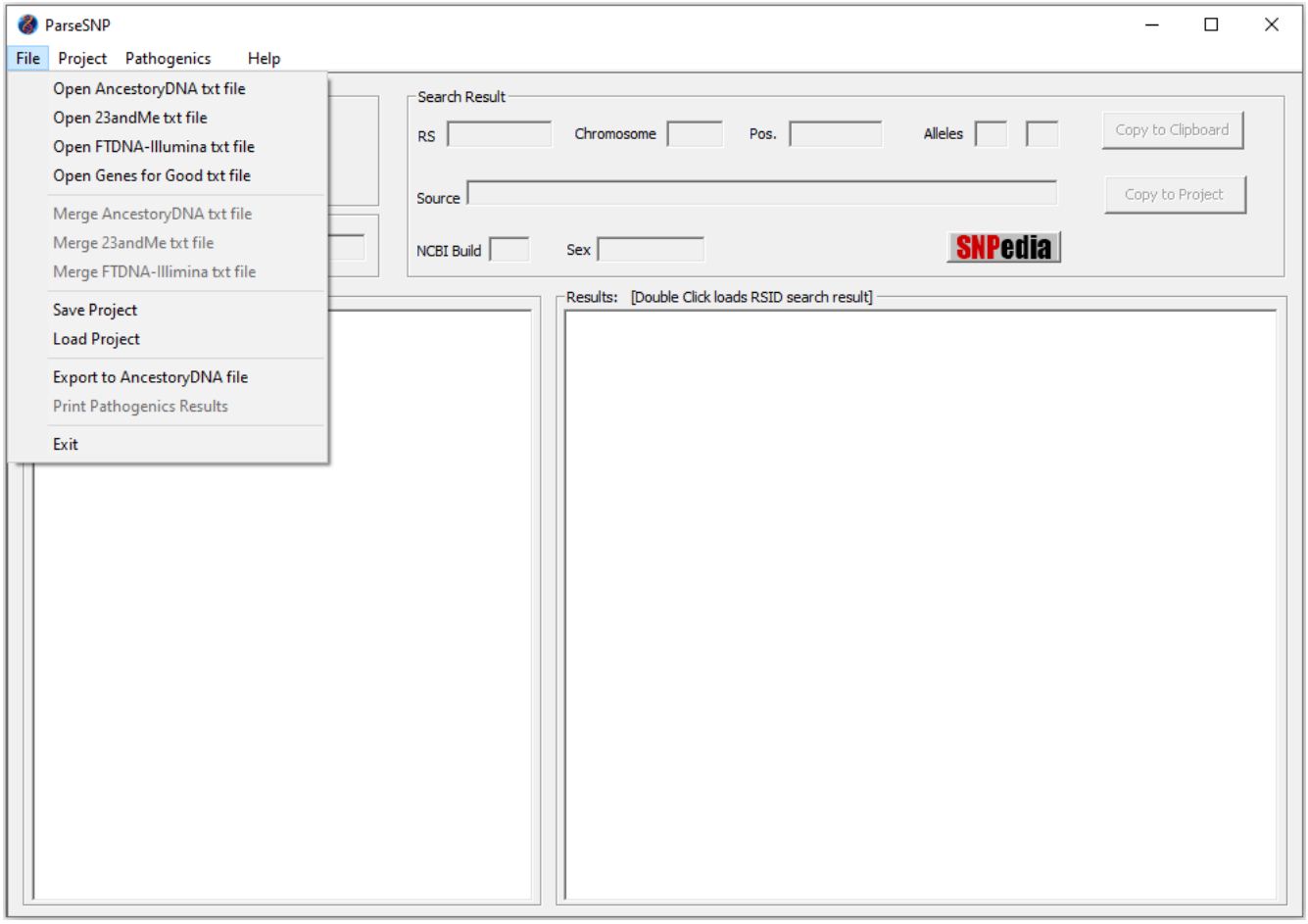
ParseSNP is a small utility application (*It would fit on a 720KB floppy disk if they still existed*) written in C++ and compiled in both 32-bit and 64-bit versions.
Its initial purpose is to allow you to search your *ancestry* raw data files for any RSID they contain and your alleles:

RSID	Chromosome	Position	Allele	Allele
rs4477212	1	82154	T	T

These values can be saved in a project window and reloaded later. This is the most basic use of ParseSNP. Deeper usage allows users to turn academic research papers’ SNP (**S**ingle **N**ucilatied **P**olymorphism) risk values into a Pathogenics file which can be loaded and your values compared against the risk data.

On the following pages I will explain some of the main available functions and their uses.

File Menu

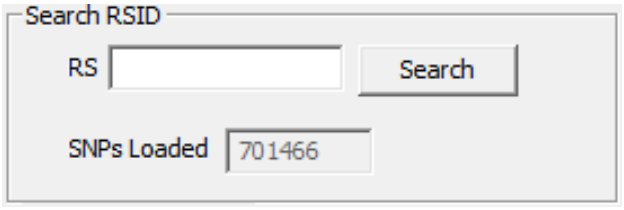


The **File** menu lets you load an Ancestry, 23andMe, FTDNA, or Genes for Good (23andMe format) raw data file into ParseSNP where it can be searched, results looked up on SNPedia, copied to the clipboard, or saved in the project window.

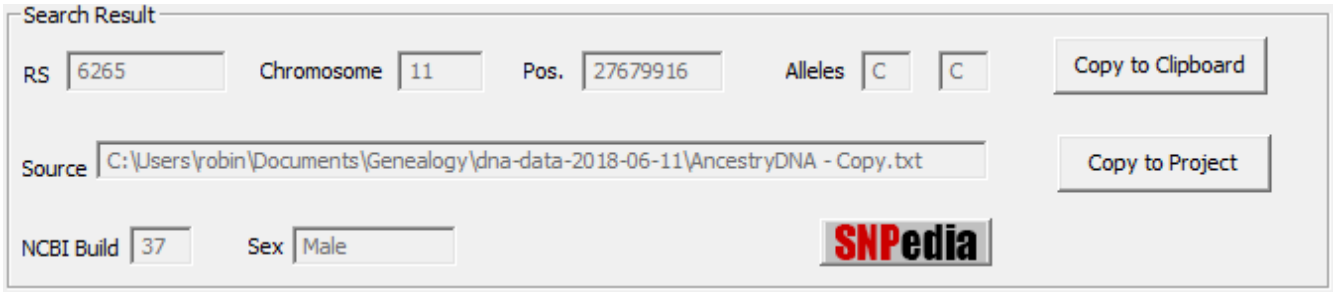
Below the open section there are options to merge any other files you have. By merging two files from different companies, you can combine the data. To keep the result of the merge (or just convert a file), you can export to an AncestryDNA file using the “Export to AncestryDNA file” function.

The “Print Pathogenics Results” prints the contents of the results window, which we will discuss in more detail later.
Load and Save Project loads or saves the project file to/from the default project path. Saving will make a copy of the raw DNA file in use; loading will restore it.

General Search functions



Once you have loaded your raw DNA file into the program, just type the RSID number you are searching for in the RS search box (*see picture above*) and click search. If it is not in your loaded data, a dialog box will let you know; otherwise, the *Search Results* area will populate with your values for that RSID:



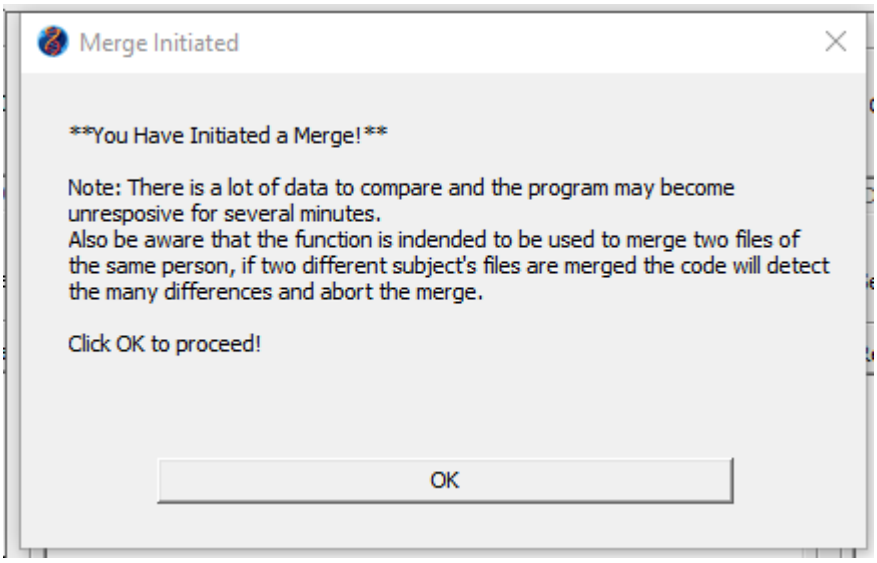
The Source field will show the loaded file and its path. The NCBI Build may be populated. Sex should be set and will not take into consideration intersex phenotypes, as commercial ancestry files do not sequence these cases they also do not sequence trisomies.
The data for your searched RSID number will populate as above: e.g. RS6265, the chromosome it is on (11), its position in the chromosome (27679916), and most importantly the alleles for a pair of genes (*chromosomes 1–22*), as SNPs are points in the genome where variations occur. These differences can increase the risk of certain medical conditions.They also account for physical differences, hair colour, attached or detached ear lobes, etc.
Search allows you to look up any SNPs you are interested in. You can copy the values to the clipboard or copy them to the project window (*more details under Project*).

Merge Option

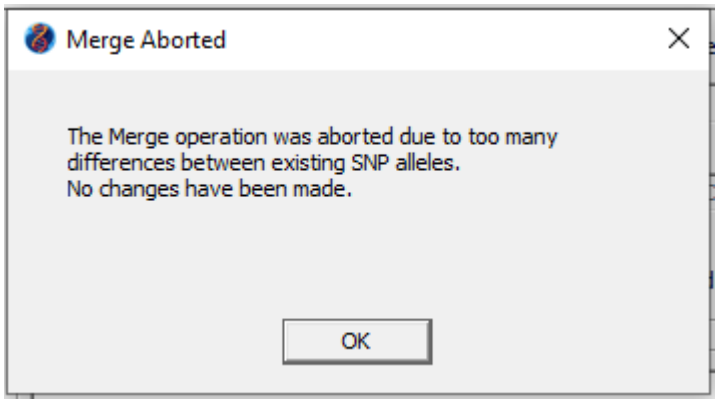
If you have raw DNA files from more than one company, you can merge the RSIDs from a second file with the data you already have loaded.

- Existing RSIDs with valid data will remain untouched.
- RSIDs that were no-reads will be replaced with valid data from the file being merged.
- RSIDs not in the original file will be added.
- Translated proprietary codes that are not present will be added.
- Untranslated proprietary codes will NOT be added, so you should never delete your original files.

As the merge process checks each entry in the original loaded data against every entry in the merge file on disk, it can take a while to complete. This is especially true on older machines with hard drives rather than SSDs. So you will see a warning screen before you initiate the merge:

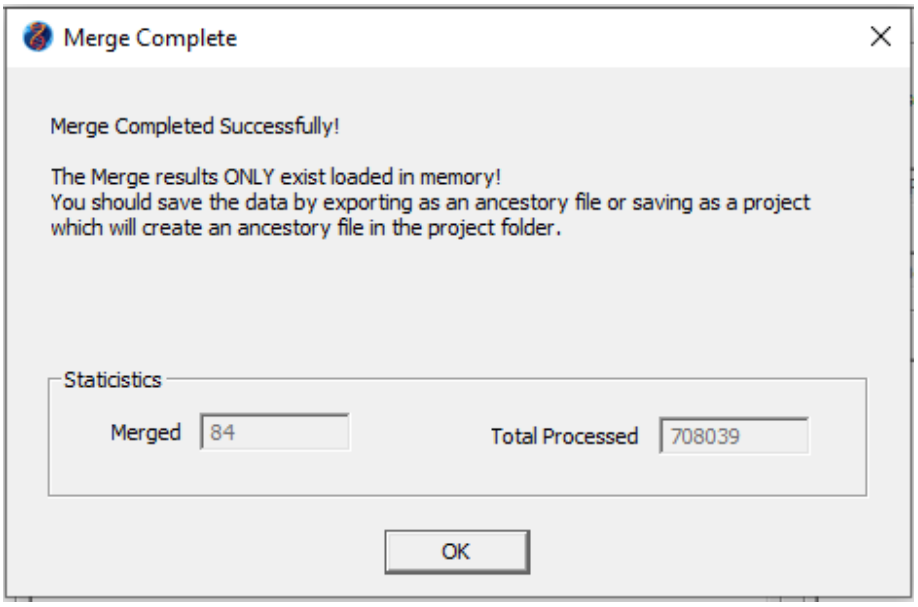


If you accidentally try to merge someone else's file, the merge routine will see too many differences in existing entries and abort with the following message:



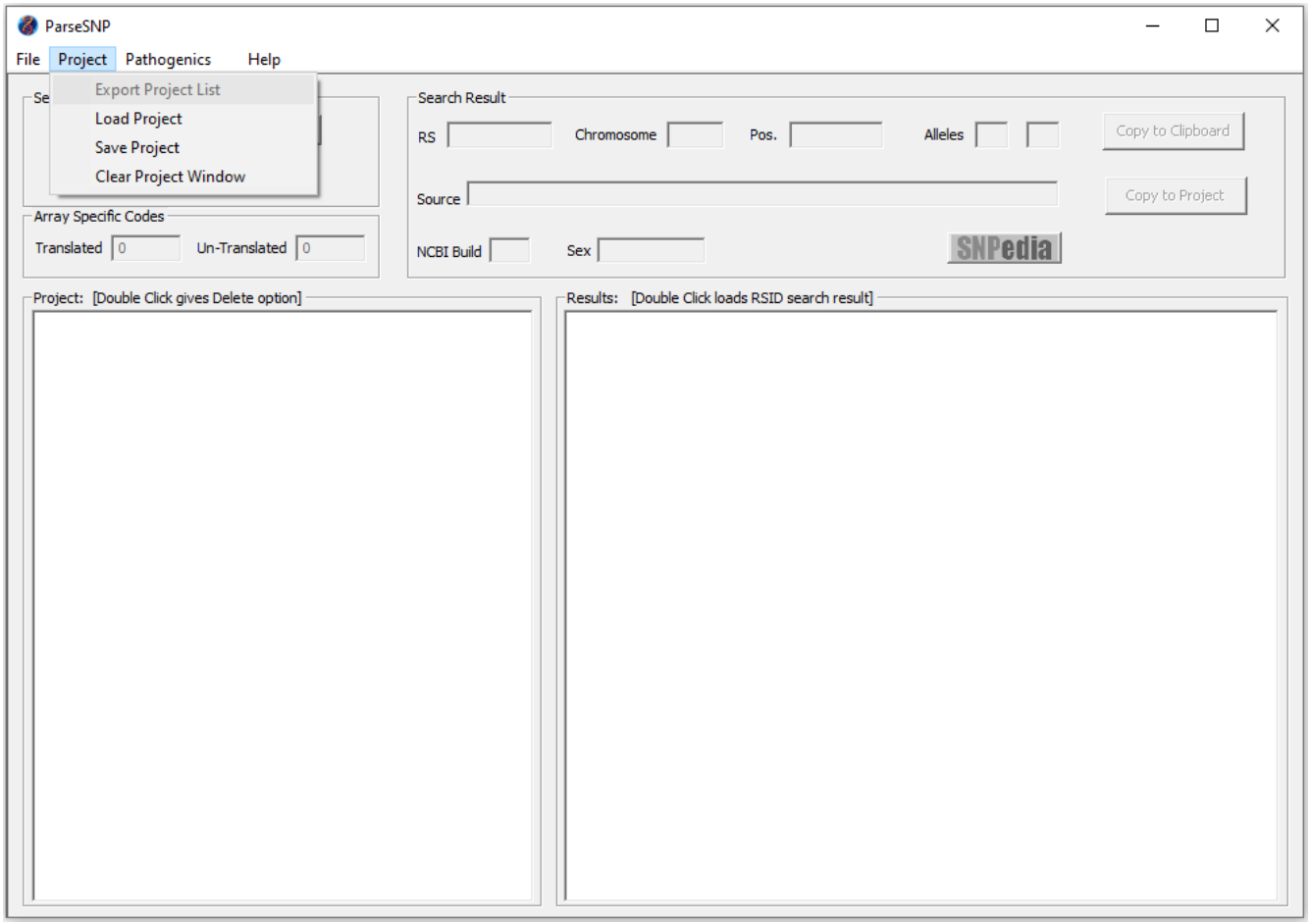
If the files are for the same person, each entry in the merge file has to be compared with every entry in the original file. Even though every match is removed from the comparison loop and the loop uses C++ pointer logic, this is a massive number of comparisons. Adding a progress indicator would only increase the time, so expect this to take several minutes during which the program will be unreposnsive!

When it completes successfully, the following screen will be shown:



You should export the merged file as an AncestryDNA file if you want to keep the changes in a new file for future use.

Project Menu

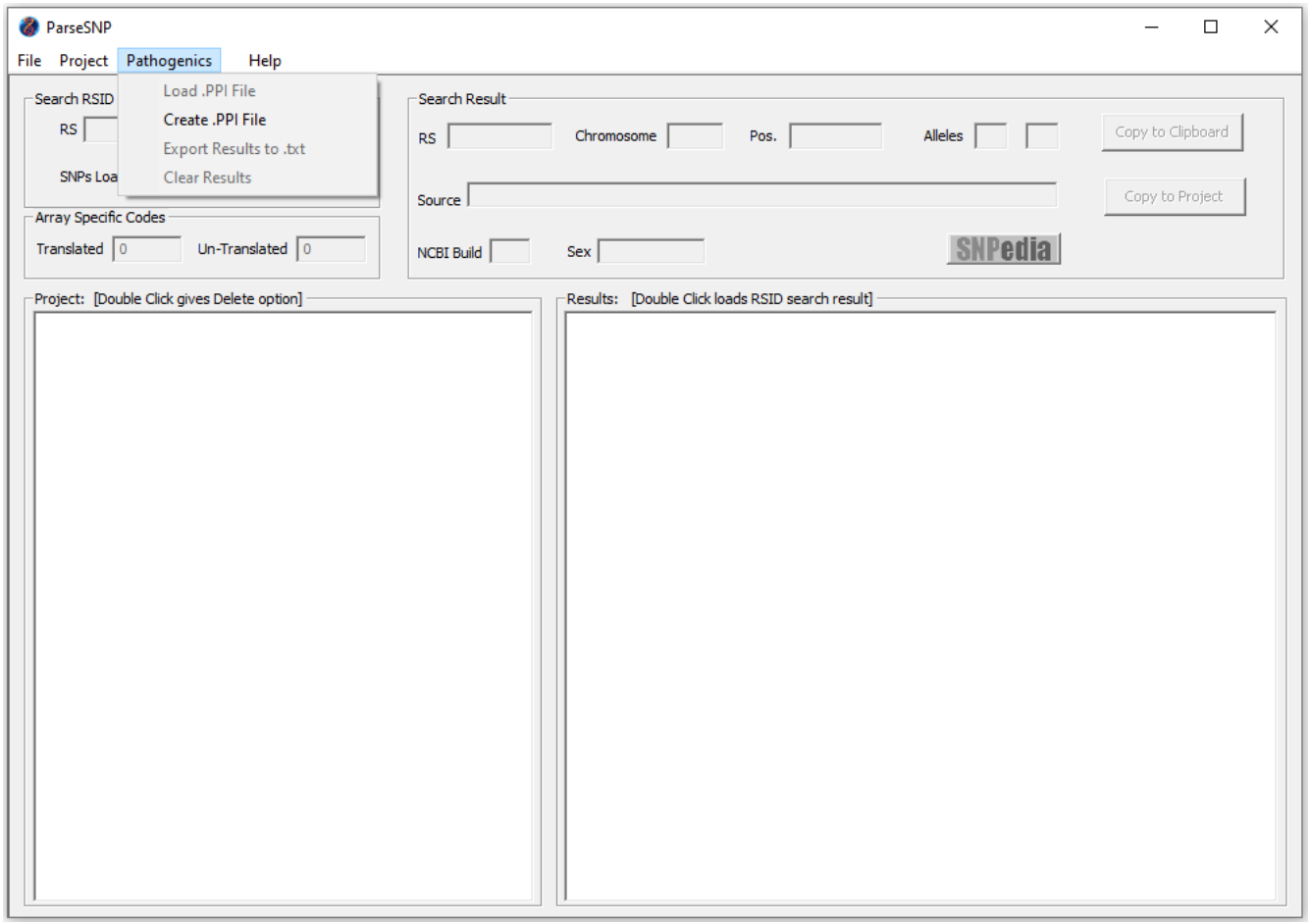


The Project Menu is pretty basic: Export Project will export the contents of the Project window to a text file whose name you choose.
Load and Save Project loads or saves the project file to/from the default project path.

Saving will save the contents of the project window and save a copy of the raw DNA file in use.

Loading will restore the contents of the poject window and load the raw DNA file copy.
Clear Project Window empties the Project window.

Pathogenics: which grew into the main purpose of the project!



The purpose of .PPI files is to compare the loaded SNP file against known risk factors for any disease that has a known genetic cause. As I was unfortunate enough to become type 2 diabetic in my mid-forties, I chose Type 2 Diabetes as my *example files* to ship with the code. I used the data from this study: https://www.hindawi.com/journals/bmri/2014/926713/ – and providing a source is mandatory when creating a .PPI file! The authors claim the results are incomplete and insufficient for predictions, but looking at my results compared to those from OpenSNP (*when we lived in a better world*) I would say they are pretty indicative. It won't tell you you're 100% safe, but it will tell you if you should make lifestyle changes! Let's just say that if I had this data in my 20s, I would have made sure I was more active (*exercise reduces the age-related drop in testosterone – the only thing protecting me from the effects of an array of compromised genes!*). I hope this utility can help people get advanced warning of many genetically caused conditions.

Odds Ratio Analysis:

At the bottom of the loaded results will be a statistical analysis and interpretation based on the odds ratio data entered at the time the .PPI file was created for each RSID risk allele. This will only show when there is sufficient odds ratio data.

```
RS10401969   19   CILP2 [C] 1.13    [T/T] No risk alleles
RS12970134   18   MC4R [A] 1.08     [A/G] Risk Heterozygous
RS7202877    16   BCAR1 [T] 1.12    [T/T] Risk Homozygous!

Your genetic risk score: 47.74%
(Percentage of theoretical maximum risk for this PPI file)
Missing data accounts for a Maximum of 13.53%
Risk level: VERY HIGH
Interpretation: Strong genetic predisposition!
Log-risk score: 4.568 (Range: 0 - 9.567)

PPI File's MD5 = 31D571C6C3DF9260E04AC81916BBB5F8
```

The first line (*genetic risk score*) is a percentage of the maximum theoretical risk (*if someone had every possible risk allele – which is almost impossible*). The second line explains this. If your raw data file did not contain all of the RSIDs in the .PPI file, the third line will give you the *maximum* the missing data could account for, so your genetic risk could increase by 0 to that *maximum*. Lines four and five give the algorithm's interpretation.

If you have any concerns about your genetic risks and have checked the paper the data was based on, you should talk to your primary care doctor, especially if your family has a history of the condition the file is checking for.

If the .PPI file does not have odds ratio data, or has more entries without odds ratio data than with it, then the analysis will not run.
See test file below:

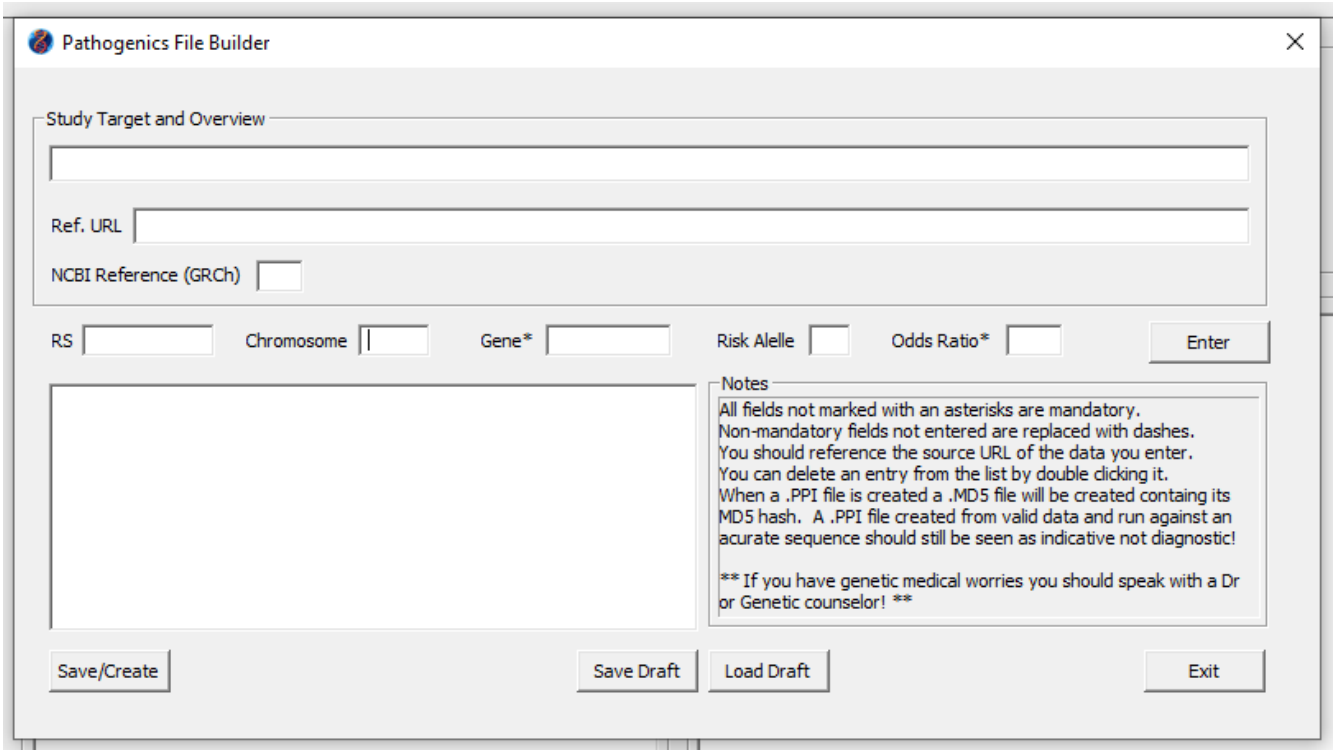
```
RS12345      1    xxx    [C]    -.--       [A/G] No risk alleles
RS12346      1    rrrr    [T]    -.--      [C/T] Risk Heterozygous
RS123457     1    eee    [G]    -.--        RSID not present in your fil


PPI File's MD5 = 2D10A888AB52EBDC5FB0870D23C14E6F
```

The last entry will be the .PPI file's MD5 cryptographic hash, so you can check that the .PPI file has not been altered by comparing it to the MD5 hash on its creator's download page.
I have type 2 diabetes, and as you can see with a `risk score of 47.74%` I was always at risk. Had I had this information at age 21, I could have made lifestyle changes that might have prevented the onset of diabetes, and even if it still happened, I would not have suffered the condition getting worse before I asked my primary doctor to check me for it!

Creating a .PPI file

Pathogenics menu → Create PPI File

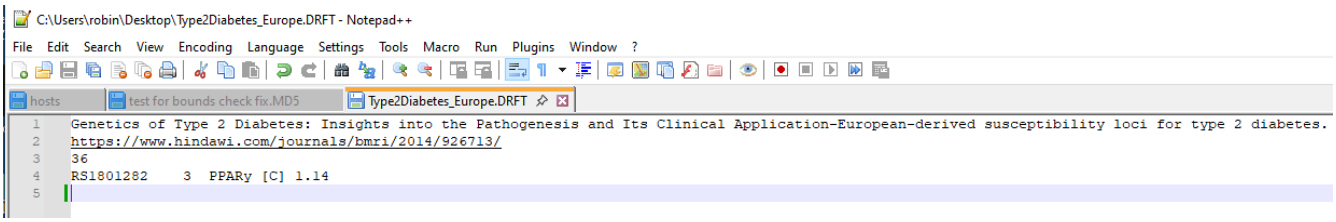


The fields are pretty clear, I hope. The top field is the title of the study or a précis of the title that clearly says what the file is checking for.
The URL: please put the link to the study so people can review it. A link is mandatory, but please take the time to find a good free link to the study.
If the study has an NCBI reference model, please enter it; it is not mandatory.
Please double-check your work as you enter the lines. You can delete a mistaken entry by double-clicking it. If you are entering a large table, you can [Save Draft] and continue it later with [Load Draft].
When you are finished, the [Save/Create] button will create two files: your .PPI file and an .MD5 file with the same name.
If you choose to share your .PPI file with others, wherever you post it you should include its MD5 hash (which you get by opening the .md5 file as a text file and copying the long hexadecimal number it contains). This hash will be recalculated when the .PPI file is run to ensure it has not been altered or corrupted.

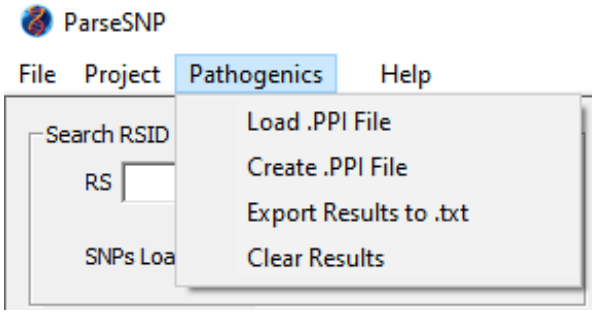
Type2Diabetes_Europe.PPI MD5 Hash: 31D571C6C3DF9260E04AC81916BBB5F8
Type2Diabetes_EastAsian.PPI MD5 Hash: 13212F699CB93921EDCD017B48BDA272
Type2Diabetes_SouthAsian.PPI MD5 Hash: B9F38C3028D6754ED0495748DDBD6C34

Quicker .PPI Creation

If you have a table with all the needed values, you can start your .PPI file creation as above, but after the first entry save it as a draft file. Open the <filename>.DRFT file as a text file in a good text editor – I use Notepad++ (https://notepad-plus-plus.org/), but you can use Windows Notepad. There you can paste in the lines and edit them to fit the file format.



When you load the .DRFT (draft file) and if it is close to the required format, it will be loaded; when saved, it will be properly formatted.



After a .PPI file has been loaded, you will have two more options in the Pathogenics menu:

Export Results to .txt

Allows you to export the contents of the Results window to a plain text file. The default Windows directory will be Documents. Note that “Print Pathogenics Results” is under the File menu, as that is a Windows convention.

Clear Results

Clears the results window.

Disclaimer

ParseSNP is provided for educational and research purposes only.
It is **not** a medical device, diagnostic tool, or substitute for professional medical advice.

- **No medical advice:** The risk assessments, odds ratios, and other outputs are based on published research and your raw DNA data. They do **not** constitute a medical diagnosis, prediction, or treatment recommendation.
- **Consult a doctor:** If you have concerns about any genetic condition or your health, you should speak with a qualified healthcare professional. Do not disregard or delay seeking medical advice based on anything generated by ParseSNP.
- **No warranty:** ParseSNP is supplied “as is”, without any warranty of merchantability, accuracy, or fitness for a particular purpose. The software may contain errors or omissions.
- **No liability:** To the maximum extent permitted by law, the author(s) and contributors shall not be held liable for any damages or harm arising from the use or misuse of this software, including but not limited to emotional distress, financial loss, or medical outcomes.

By using ParseSNP, you acknowledge that you understand and accept these terms.