Auxiliary scripts

`AgileROHFinder` and `AgileROHFilterer` are designed to be used as part of a pipeline. The first two scripts are designed to show how they can be automated in a Bash script

AgileROHFinder_bash_script.sh

This script highlights how to `AgileROHFinder` can be automated. Rather that manually enter a series of commands to process each file in a folder of VCF data files, this script iterates through the files, and instructs `AgileROHFinder` to save the regions data to a file with the same name as the input file, but with a 'txt' file extension.

The scripts command line structure is:

```
bash /script/AgileROHFinder_bash_script.sh /program/AgileROHFinder.exe /data/Files/ format
```

Where

- /script/AgileROHFinder_bash_script.sh: the name with location of this bash script
- /program/AgileROHFinder.exe: the name and location of the program to use
- /data/Files/: the name and location of the folder of the VCF files (the script expects the file extension to be lower case 'vcf')
- format: The format of the regions file (**-a**, **-t** or **-b**). If you which to process SNPs without an RS id, the **-Y** option could be combined by adding the **-Y** option to the format option and putting them all in speech marks i.e. "**-t -Y**". Note space between the **-t** and the **-Y**. Similarly, the ***-V option can be included:*** "***-V -Y***" will process variants without a RS is, using the genotypes in VCF file

AgileROHFilterer_bash_script.sh

Like the previous script this script show how to `AgileROHFilterr` can be automated. This script iterates through the files, and instructs `AgileROHFinder` to save the regions data to a file with the same name as the input file, but with a 'txt' file extension and save the filtered variants to a file with the same name as the input VCF file, but with '*filtered*' *inserted before the file extension. As before the format option can be modified to include the* ***-Y*** and/or ***-V*** option to process SNPs without an RS id.

The scripts command line structure is:

```
bash /script/AgileROHFilterer_bash_script.sh /program/AgileROHFinder.exe /data/Files/ format
```

Where

- /script/AgileROHFilterer_bash_script.sh: the name with location of this bash script
- /program/AgileROHFilterer.exe: the name and location of the program to use
- /data/Files/: the name and location of the folder of the VCF files (the script expects the file extension to be lower case 'vcf')
- 500000: any positive integer by which the autozygous regions are extended tp include variants close to the regions

- format: The format of the regions file (**-a**, **-t** or **-b**). If you which to process SNPs without an RS id, the **-Y** option could be combined by adding the **-Y** option to the format option and putting them all in speech marks i.e. "**-t -Y**". Note space between the **-t** and the **-Y**.

p_FindCommonRegions.py script

Neither program attempt to combine the results of two are more analysis to identify common autozygous regions. This python script will attempt to read all the text files (*.txt) created by `AgileROHFinder` or `AgileROHFilterer` with the -t format option in a folder and identify the location of the autozygous regions and then create a minimum list of regions present in all the files. Since the analysis files could originate from different data types and different families with a similar phenotype, the script makes no attempt to match the haplotype of common regions. Similarly, it does not exclude regions present in unaffected relatives as they may be autozygous for a different haplotype.

Note

The folder should only contain results files: If other text files are present the script will attempt to read then and probably crash as they will not contain the expected data/format.

The scripts command line structure is:

```
python /script/p_FindCommonRegions.py /data/ResultFiles/
```

Where

- /script/p_FindCommonRegions.py : the name with location of this python script
- /data/ResultFiles/: the name and location of the folder of the txt files (the script expects the file extension to be lower case 'txt')