Documentation for Haplo2D6, version 1.0

Maria Carolina Silva de Barros Puça, Diego Mariano, Raquel Cardoso de Melo Minardi, Taís Nóbrega de Sousa[1]

November 2024

[1]Address for correspondence: René Rachou Institute, Oswaldo Cruz Foundation (FIOCRUZ), Belo Horizonte, email: haplo2d6@gmail.com, http://bioinfo.dcc.ufmg.br/Haplo2D6

How to Use Haplo2D6 to Translate CYP2D6 Genotype to Phenotype

Haplo2D6 is a powerful tool designed to predict the star alleles and enzyme activity of CYP2D6, making the analysis of genotype data more straightforward. The software is intended for researchers and clinicians who need to streamline the process of determining CYP2D6 metabolizer status based on genetic data.

1. Input Requirements

To use Haplo2D6, you need to provide:

1.1 Genotype Data: the input file supplied by the user specifies the number of individuals analyzed, the number and type of loci (biallelic or multi-allelic locus), and the genotypes for each individual.

Examples of input files containing genotype data of 9 loci on the *CYP2D6* gene for 20 individuals in a spreadsheet format*:

20																		
9																		
P 2617 4300 5222 6047 6816 7051 7189 7384 8381																		
SSSSSSSSS																		
1	C	C	C	C	C	C	G	G	C	C	C	C	G	G	G	G	G	G
2	G	G	C	C	C	C	G	G	C	C	T	T	G	G	G	G	C	C
3	C	C	C	C	C	C	G	G	A	A	C	C	G	G	G	G	G	G
4	C	C	T	T	C	C	G	G	C	C	C	C	G	G	G	G	C	C
5	C	C	C	C	T	T	G	G	C	C	T	T	G	G	G	G	C	C
6	C	C	C	C	C	C	G	G	C	C	T	T	G	G	G	G	G	G
7	C	C	C	C	C	C	G	G	C	C	T	T	A	A	G	G	C	C

20									
9									
P 2617 4300 5222 6047 6816 7051 7189 7384 8381									
SSSSSSSSS									
1	CC	CC	CC	GG	CC	CC	GG	GG	GG
2	GG	CC	CC	GG	CC	TT	GG	GG	CC
3	CC	CC	CC	GG	AA	CC	GG	GG	GG
4	CC	TT	CC	GG	CC	CC	GG	GG	CC
5	CC	CC	TT	GG	CC	TT	GG	GG	CC
6	CC	CC	CC	GG	CC	TT	GG	GG	GG
7	CC	CC	CC	GG	CC	TT	AA	GG	CC

Line 1 – Number of individuals analyzed

Line 2 – Number of loci/sites

Line 3 – P (upper case) indicates the position of locus i, relative to some arbitrary reference point; the loci must be in their physical order along the chromosome (ie, the positions must be increasing). Note. Position at NG_008376.3 (CYP2D6 RefSeqGene; reverse relative to chromosome).

Line 4 – Locus type: S for biallelic locus and M for multi-allelic locus.

* Haplo2D6 also supports the Phase default format (file.inp). For more information, see Phase documentation (https://stephenslab.uchicago.edu/phase/download.html).

1.2 Allele References: A reference file that maps genetic variants to star alleles and their respective functions.

Example of input file (comma separated) contains the variants that comprise each star allele, allele function assignments (normal function, no function, or decreased), and the allele activity value. Source material can be found on PharmGKB (https://www.pharmgkb.org/page/cyp2d6RefMaterials) and PharmVar websites (https://www.pharmvar.org/gene/CYP2D6). The positions follow the same order as in the previous table.

CCCGCCGGG,Normal,*1,1
GCCGCTGGC,Normal,*2,1
CTCACCGGC,No function,*4,0
CCCGACGGG,Reduced,*9,0.25
CTCGCCGGC,Reduced,*10,0.25
CCTGCTGGC,Reduced,*17,0.5
CCCGCTGAC,Reduced,*29,0.5
CCCGCTGGG,Normal,*34,1
CCCGCCGGC,Normal,*39,1
CCCGCTAGC,Reduced,*41,0.25

2. Analyzing Your Data

The process of analyzing your data with Haplo2D6 involves several steps:

2.1 Data Preparation: Before running Haplo2D6, ensure that your genotype data is formatted correctly and that you have the necessary allele reference file.

2.2 Haplotype Reconstruction: Haplo2D6 uses a tool called **Phase** to reconstruct haplotypes from population data. This step is crucial for accurately determining which alleles are present in each individual. Phase is an open-source software available under the following license: Phase License.

2.3 Star Allele and Enzyme Activity Prediction: Once the haplotypes are reconstructed, Haplo2D6 predicts the star alleles and corresponding enzyme activities. The output includes the predicted star alleles, allele activity value, and an associated phenotype (e.g., normal metabolizer).

Example of the output file

#	ID	Haplotype #1	Allele Functional #1	Allele #1	Activity Value #1	Haplotype #2	Allele Functional #2	Allele #2	Activity Value #2	Activity Score	Phenotype
0	1	CCCGCCGGG	Normal	*1	1.0	CCCGCCGGG	Normal	*1	1.0	2.0	gNM
1	10	CCCGCCGGG	Normal	*1	1.0	CCTGCTGGC	Reduced	*17	0.5	1.5	gNM
2	11	CTCGCCGGC	Reduced	*10	0.25	GCCGCTGGC	Normal	*2	1.0	1.25	gNM
3	12	CTCACCGGC	No function	*4	0.0	CTCACCGGC	No function	*4	0.0	0.0	gPM
4	13	CCCGCCGGC	Normal	*39	1.0	CCCGACGGG	Reduced	*9	0.25	1.25	gNM
5	14	CCCGCTGAC	Reduced	*29	0.5	GCCGCTGGC	Normal	*2	1.0	1.5	gNM
6	15	CCCGCTGGG	Normal	*34	1.0	CTCACCGGC	No function	*4	0.0	1.0	gIM
7	16	CCCGCTAGC	Reduced	*41	0.25	CCTGCTGGC	Reduced	*17	0.5	0.75	gIM

2.4 Copy Number Variation (CNV) Consideration: **Important**: Haplo2D6's initial output does not account for gene copy number variations. You must adjust the results if CNVs are detected. For instance:

- If there is only one copy of the gene, change one of the alleles to *5, which represents a complete gene deletion. For example, a genotype of *1/*1 before CNV analysis should be adjusted to *1/*5. Adjust the Total Score and Phenotype accordingly.

- If there is gene duplication, change the allele nomenclature considering the number of gene copies (eg. *1x2, *1x≥3). The phenotype may change from normal metabolizer (gNM) to ultrarapid metabolizer (gUM). For instance, *1/*1 (AS Total = 2, gNM) may change to *1/*1x2 (AS Total = 3, gUM).

2.5 Final Phenotype Adjustment: After considering CNVs, adjust the activity score and the predicted phenotype accordingly.

3. Additional Resources and Recommendations

- **Minimum Sets of Alleles for Pharmacogenomic Testing**: For guidance on which alleles to test, refer to this resource: PharmGKB Alleles to Test.
- **Gene-specific Information Tables**: These tables provide definitions for each star allele, their function (e.g., normal, no function, decreased function), and how these alleles combine to form metabolizer phenotypes. Access this resource here: PharmGKB CYP2D6 Reference Materials.
- **Copy Number Variation and Phenotype Translation Examples**: You can find examples and additional guidance on how to adjust phenotypes based on genotype data and CNVs in this template: CYP2D6 Genotyping Method and Data Template.