

universal input file format

ChildID	FatherID	MotherID	PregID	Index
ab27	aaa	bbb	27	1
ab11	aaa	bbb	11	2
up36	uuu	0	36	3
eg82	eee	ggg	82	4

.....

- The header is present
- Fixed order of columns
- ChildID, FatherID, MotherID - all are SentraxIDs
- ChildID and PregID contain unique never missing values
- FatherID and MotherID contains non-unique values or “0”
- Index is the sequential order, i.e., from 1 to nrow
- Pregnancy-centered (=> parents can be repeated)
- Includes multi-child families (different pregnancies, same parents)
- Includes pregnancies without one parent (missing ID is “0”)
- All mentioned IDs have genotypes in imputation output file
- File naming (m12 and m24 separately):

haploPrepFam_m12_n6636_19c7de9_20171201.txt
haploPrepFam_m24_n4271_19c7de9_20171201.txt

output file format 1

Pos	Ref	Alt	Index=1			Index=2			Index=3		
			kid	mat	pat	kid	mat	pat	kid	mat	pat
			M P	T U	T U	M P	T U	T U	M P	T U	T U
1234567	G	A	0 0	0 0	0 1	0 1	0 0	1 0	0 1	0 .	1 1
2345678	C	A	0 1	0 0	1 1	1 0	1 .	0 0
3456789	A	G	1 1	1 1	1 0	1 1	1 1	1 0	1 1	1 .	1 1
4567890	T	G	1 1	1 0	1 0	0 1	0 1	1 0	0 1	0 .	1 0
5678901	T	C	1 0	1 0	0 1	0 1	0 1	1 0	1 0	1 .	0 1
6789012	C	A	0 1	0 1	1 0	1 1	1 0	1 0	0 0	0 .	0 1

.....

- Standard vcf header with “#” (ID = *PregID_IndID*)
- First columns: standard vcf
- M = maternally derived, P = paternally derived
- T = transmitted, U = untransmitted
- Three columns correspond to one row in input
- Pipe symbol (|) to separate haplotypes
- Same Index order as in input
- “.” for missing data or phasing/Mendel error
- If genotype for one parent is missing, he/she still gets assigned the transmitted allele (rewritten from child)
- File naming (per chip, per chromosome):
eg., *haploFullTrio_m24_n4271_19c7de9_chr22.txt*
- Output file should contain the same hash as the input

output file format 2

Pos	Id	Ref	Alt	Index=		
				1	2	3
1234567	rs9876	G	A	0	0	.
2345678	rs8765	C	A	0	.	.
3456789	rs7654	A	G	1	1	.
4567890	rs6543	T	G	0	1	.
5678901	rs5432	T	C	0	1	.
6789012	rs4321	C	A	1	0	.

.....

- No header
- Only untransmitted maternal haplotype
- One column in output corresponds to one row in input
- Same Index order as in input
- “.” for missing data or phasing/Mendel error
- File naming (per chip, per chromosome): eg., *haploMomUnT_m24_n4271_19c7de9_chr22.txt*
- Output file should contain the same hash as the input file (automatically)