

Supplementary File S5 : Supplementary Methods.

Specification of all CNV signatures recognized by the software CNVmap. This list is implemented in the function `get_hypotheses()`.

Each type of CNV event that is identified in this study is specified by the presence of specific peaks in its different genome-wide allele frequency profiles obtained when considering subsets of individuals associated with different 3-markers genotype classes. There are 6 such possible classes, denoted by the following 3-character strings: AAA, AHA, A-A, BBB, BHB, B-B. The second character of the string indicates the allele at the candidate marker M^* and the first respectively third character indicates the allele at the flanking Mendelian marker M_L respectively M_R (see text).

Each type of CNV event produces a specific signature for its six profiles, which can be made of complete peaks (upwards, 100 % individuals having the allele A from the first parent), complete troughs (downwards, 0 % of allele A, 100 % of allele B from the second parent), partial peaks or troughs (which do not reach the 0 % or 100 % value), or just a flat profile around 50 % of either allele (or with a similar level of segregation bias as in the whole population). Such a signature is found at the position of each locus involved in the CNV event. To encode a signature, the software specifies the behavior of the 6 profiles at each of the CNV loci *via* the character « A » for a complete peak (upwards), « C » for an incomplete peak, « B » for a complete trough (downwards), and « D » for an incomplete trough. For instance in a 1:2 CNV of the type A/B_A/- , the signature will consist of two strings of length 6 corresponding to the patterns (1) a AAA peak, a BBB trough, and a BHB trough at the position of the first locus (locus where the candidate marker was initially mapped), and (2) a BBB trough and a BHB peak at the position of the second (previously unknown) locus. Note that the profile associated with the AAA class is flat at the position of the second locus (such a behavior is denoted « E » in the encoded signature), and there are no individuals expected to belong to the class AHA (denoted « - » in the encoded signature). Moreover, for this type of event

which does not rely on missing data, the software ignores the profiles associated with classes A-A and B-B, as those classes rely on missing data, which can also be due to technical problems in the genotyping ; in that case, we use the character « i » (for ignore) to specify the « behavior » of the profile for that class. All the characters used to encode the different signatures recognized by the software are summarized in the following table :

Code	Individuals present	Complete peak or trough	Comment
-	No		
A	Yes	Peak	
B	Yes	Trough	
C	Yes	No	Incomplete peak (A-like)
D	Yes	No	Incomplete trough (B-like)
E	Yes	No	Flat profile
a	Yes or No	Peak or No	Peak dispensable (if no individuals in the class) but must be complete if individuals present
b	Yes or No	Trough or No	Trough dispensable (if no individuals in the class) but must be complete if individuals present
i	Yes or No	Yes or No	Ignore possible peaks or troughs

Note that in some particular cases involving systematic genotyping errors (such as the type A/B_A/-|HmHb), some classes may or may not have individuals, depending on the associated error probabilities. Such classes are encoded with lowercase letters « a » or « b » as shown in the table.

Given these characters encoding the behavior of a profile for a given class at a given locus, the signature of a CNV is encoded by as many strings of 6 characters as there are loci involved. Each string is of length 6 because of the 6 genotypic classes which we order as : AAA, AHA, A-A, BBB, BHB, B-B. For instance the signature of a CNV of the type A/B_A/- will be associated with the string A-iBBi at the first locus, and the string E-iBAi at the second locus.

The table below lists all types of CNV events recognized by the software, and provides for each of them the encoding of their signature (that is the list of associated character strings, one for each

locus involved). Finally, the last two columns show the expected apparent genotype of the parents of the mapping population for the candidate markers of the considered type of CNV.

Index	Type	Locus 1	Locus 2	Locus 3	Parent A	Parent B
1	A/B	A--B--	NA	NA	A	B
2	A/B Ah	aA-B--	NA	NA	A	B
3	A/B Am	a-AB--	NA	NA	A	B
4	A/B Bh	A--bB-	NA	NA	A	B
5	A/B Bm	A--b-B	NA	NA	A	B
6	A/B_A/-	A-iBBi	E-iBAi	NA	A	B
7	A/B_A/- Hm	A-iB-B	E-iB-A	NA	A	B
8	A/B_A/- HmHb	A-iBbB	E-iDaA	NA	A	B
9	A/B_-/A	A-iBBi	E-iABi	NA	A	H
10	A/B_-/A Hm	A-iB-B	E-iA-B	NA	A	H
11	A/B_-/A HmHb	A-iBbB	E-iCbB	NA	A	H
12	A/B_B/-	AAiB-i	BAiE-i	NA	H	B
13	A/B_B/- Hm	A-AB-i	B-AE-i	NA	H	B
14	A/B_B/- HmHa	AaAB-i	DaAE-i	NA	H	B
15	A/B_-/B	AAiB-i	ABiE-i	NA	A	B
16	A/B_-/B Hm	A-AB-i	A-BE-i	NA	A	B
17	A/B_-/B HmHa	AaAB-i	CbBE-i	NA	A	B
18	A/B_-/A_-/A	A-iBBi	E-iADi	E-iADi	A	H
19	A/B_A/-_A/-	A-iBBi	E-iBCi	E-iBCi	A	B
20	A/B_-/B_-/B	AAiB-i	ADiE-i	ADiE-i	A	B
21	A/B_B/-_B/-	AAiB-i	BCiE-i	BCiE-i	H	B
22	A/B_-/A_A/-	A-iBBi	E-iADi	E-iBCi	A	H
23	A/B_A/-_-/A	A-iBBi	E-iBCi	E-iADi	A	H
24	A/B_-/A_-/B	AAiBBi	EEiABi	ABiEEi	A	H
25	A/B_-/B_-/A	AAiBBi	ABiEEi	EEiABi	A	H
26	A/B_-/A_B/-	AAiBBi	EEiABi	BAiEEi	H	H
27	A/B_B/-_-/A	AAiBBi	BAiEEi	EEiABi	H	H
28	A/B_A/-_-/B	AAiBBi	EEiBAi	ABiEEi	A	B
29	A/B_-/B_A/-	AAiBBi	ABiEEi	EEiBAi	A	B
30	A/B_A/-_B/-	AAiBBi	EEiBAi	BAiEEi	H	B
31	A/B_B/-_A/-	AAiBBi	BAiEEi	EEiBAi	H	B
32	A/B_-/B_B/-	AAiB-i	ADiE-i	BCiE-i	H	B
33	A/B_B/-_-/B	AAiB-i	BCiE-i	ADiE-i	H	B