Opening the black-box of imputation software to study the impact of the reference panel composition on performance

Supplementary Figures and Tables

T Dekeyser[1,2], E Génin[1,2], AF Herzig[1,*]

1 : Inserm, Univ Brest, EFS, UMR 1078, GGB, Brest, France
Faculté de Médecine - IBRBS 22 avenue Camille Desmoulins F-29238 BREST Cedex 3 - France
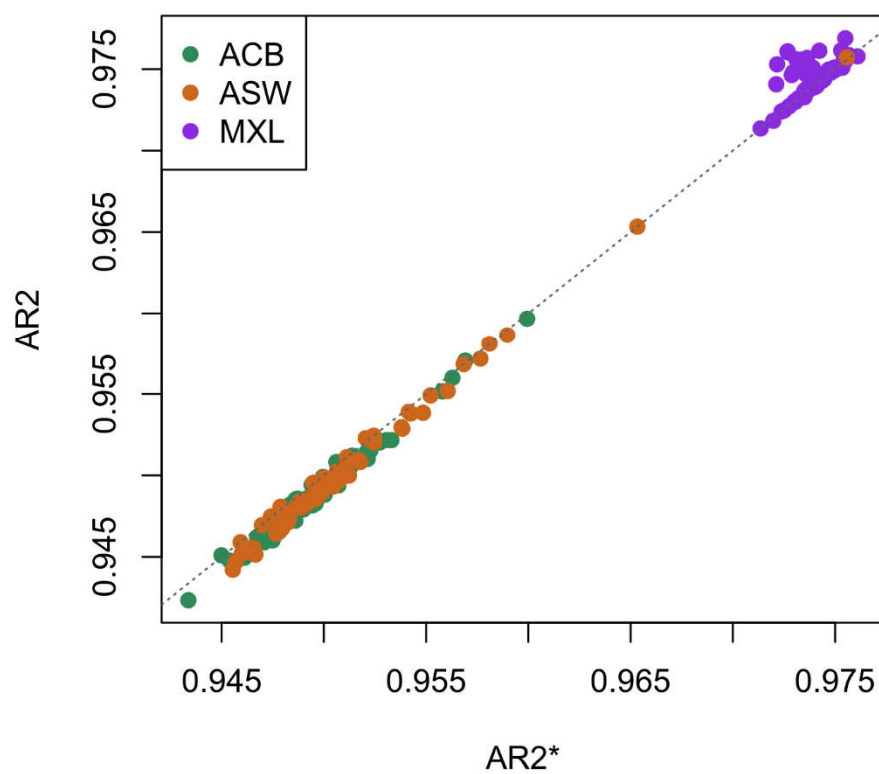**2 :** CHRU Brest, Brest, France
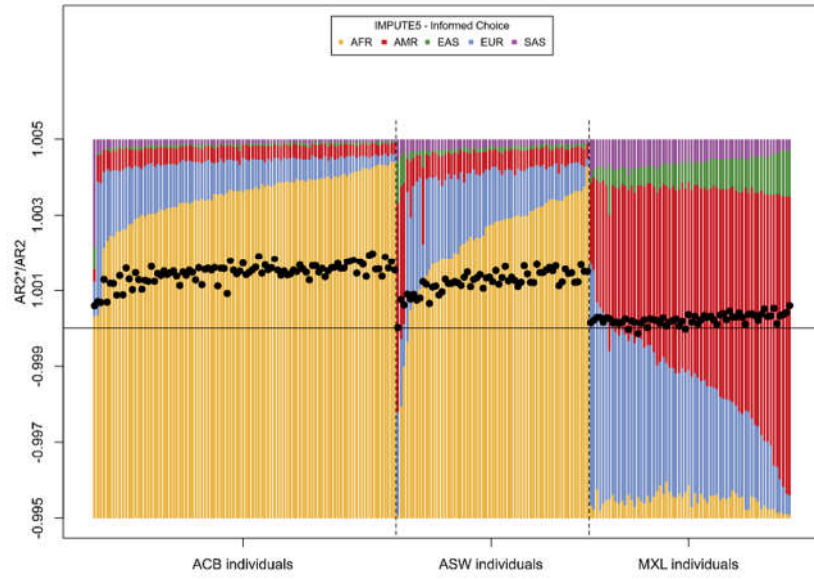* Corresponding Author

Figure S1: Individual AR2 (IMPUTE5 with all populations) and AR2* (IMPUTE5 with SOURCEFIND populations) statistics that correspond to those presented in Figure 2 in the main test. Individuals are coloured by their 1000 Genomes group, ACB, ASw, or MXL.
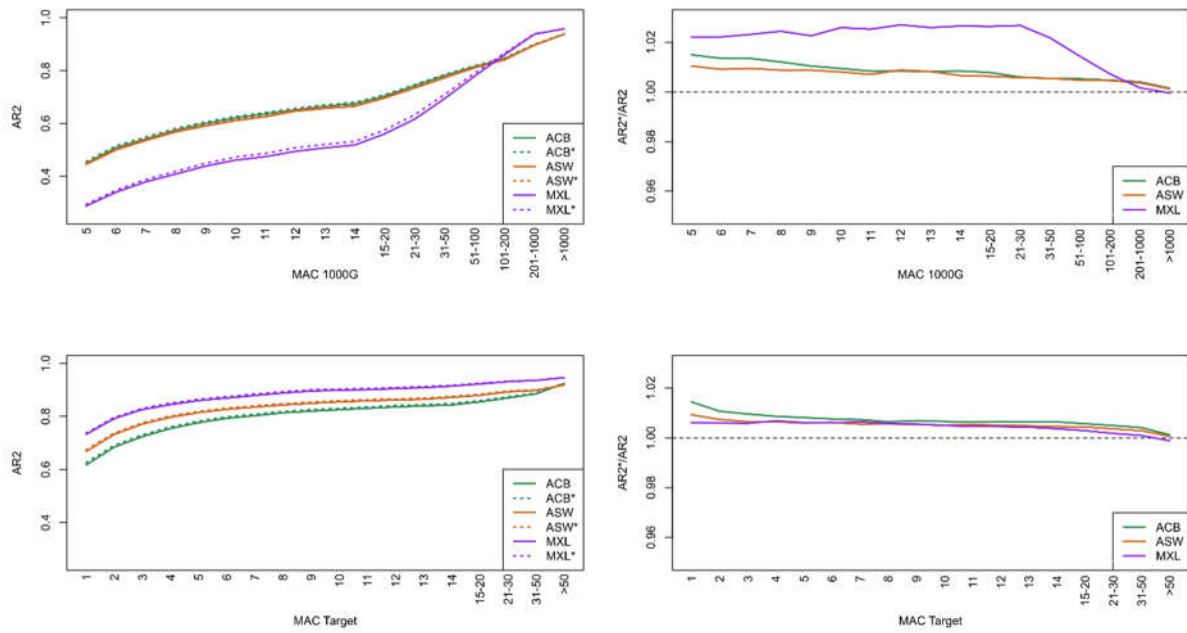
Figure S2. This figure echoes Figure 3 in the main text. The admixture plot in (a) shows the cumulative imputation barcodes used by IMPUTE5 under the informed choice approach; exactly as in Figure 3. The AR2 (aggregate R^2) overlaid in (a) and given in more detail in (b) however correspond to imputation using MINIMAC4.
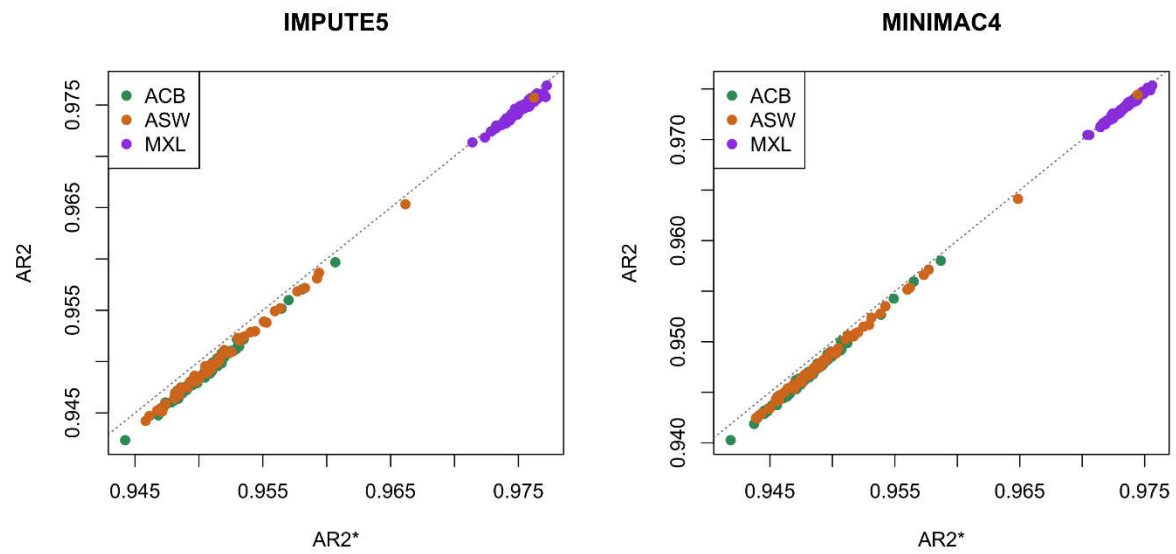
Figure S3: As Figure S1 but pertaining to the 'informed choice' strategy and showing the individual aggregate R^2 statistics from Figure 3 in the main text for IMPUTE5 and Figure S2 for MINIMAC4.
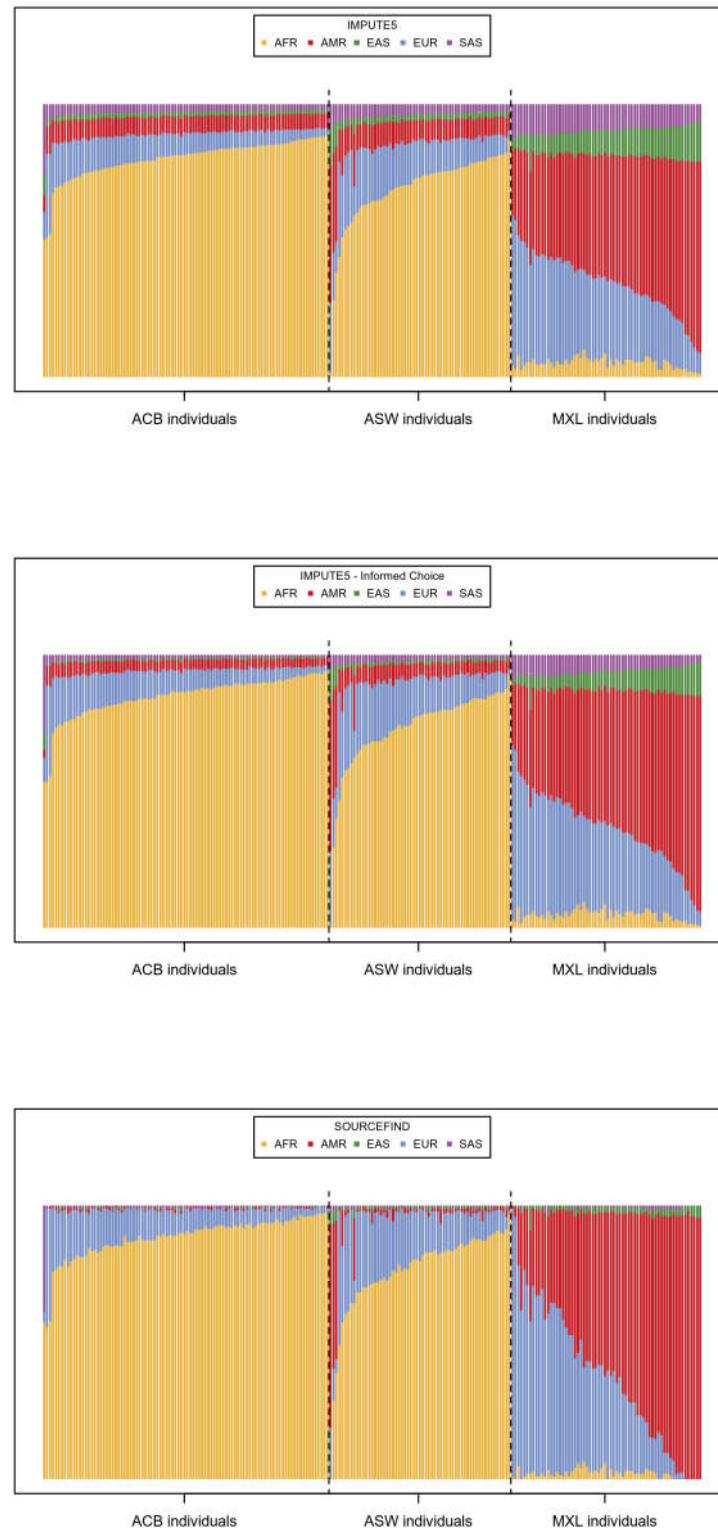
Figure S4: Proportions of the genome imputed with different populations when using IMPUTE5 with the complete 23 population reference panel (top), with the informed choice strategy (middle), or the proportions as were estimated by SOURCEFIND (bottom). The top and bottom panels appear in Figure 1 in the main text, and the middle panel appears in Figure 3 in the main text, they are put together here to facilitate their comparison.

	Hard-called dosage →	MINIMAC4			MINIMAC4, informed choice		
Truth ↓		AA	Aa	aa	AA	Aa	aa
MAC 0	AA	1438349128	570847	417	1438383014 + <0.01%	536997 -5.93%	381 -8.63%
	Aa	-	-	-	-	-	-
	aa	-	-	-	-	-	-
MAC 1	AA	626859413	511378	349	626879224 + <0.01%	491580 -3.87%	336 -3.72%
	Aa	1060454	1789942	1291	1056408 -0.38%	1794103 +0.23%	1176 -8.91%
	aa	-	-	-	-	-	-
MAC 2	AA	466653082	477660	331	466671271 + <0.01%	459495 -3.80%	307 -7.25%
	Aa	1400943	2841075	2196	1394398 -0.47%	2847745 +0.23%	2071 -5.69%
	aa	2260	3754	4848	2213 -2.08%	3883 3.44%	4766 -1.69%
MAC 3	AA	343451450	426792	326	343467622 + <0.01%	410666 -3.78%	280 -14.1%
	Aa	1412081	3275216	2998	1407515 -0.32%	3279980 +0.15%	2800 -6.60%
	aa	3533	7084	10229	3444 -2.52%	7227 2.02%	10175 -0.53%

Table S1. Again, this table echoes Table 2 in the main text but the results here correspond to MINIMAC4 and not IMPUTE5.