SUPPLEMENTARY MATERIAL

EPIGEN-BRAZIL INITIATIVE RESOURCES: A LATIN AMERICAN IMPUTATION PANEL AND THE SCIENTIFIC WORKFLOW (A TOOL FOR TRANSPARENT AND REPRODUCIBLE BIOINFORMATICS ANALYSES)

The EPIGEN-Brazil Imputation Panel – Cohorts Section

1. TARGET DATASET

1.1. Salvador Cohort

The Salvador-SCAALA (Social Changes, Asthma and Allergy in Latin America Program) Project is a longitudinal study involving a sample of 1,445 children aged 4-11 years in 2005, living in Salvador, a city of 2.7 million inhabitants in Northeast Brazil. The population have been part of an earlier observational study that evaluated the impact of the sanitation program on diarrhoea in 24 small geographical areas selected to represent the population without sanitation in Salvador. From these study participants, 1,309 were successfully genotyped as part of the EPIGEN Project. Further details are available in (Barreto et al. 2006).

1.2. Bambuí Cohort

The Bambuí cohort study of ageing is in progress in Bambuí, a city in Minas Gerais State in Southeast Brazil, of approximately 15,000 inhabitants. The cohort population consisted of all residents aged 60 years and over on January 1997, who were identified from a complete census in the city. From 1,742 eligible residents, 1,606 constituted the original cohort, and 1,442 of these participants were successfully genotyped as part of the EPIGEN Project. Further details of the Bambuí study can be seen in (Lima-Costa et al. 2011).

1.3. Pelotas Cohort

The 1982 Pelotas birth cohort study was conducted in Pelotas, a city in Brazil extreme south, nears the Uruguay border, with 214,000 urban inhabitants in 1982. Throughout 1982, the

three maternity hospitals in the city were visited daily and births were recorded, corresponding to 99.2% of all births in the city. The 5,914 live born infants whose families lived in the urban area constituted the original cohort. We have genome-wide data for 3,736 individuals. Further details are available in (Victora and Barros 2006).

Summary of working target datasets:

1. EPIGEN 2.5M target dataset (2,235,109 SNPs for 6,487 samples)
2. EPIGEN 2.5M target dataset - Salvador Cohort (2,234,755 SNPs for 1,309 samples)
3. EPIGEN 2.5M target dataset - Bambui Cohort (2,233,665 SNPs for 1,442 samples)
4. EPIGEN 2.5M target dataset - Pelotas Cohort (2,234,985 SNPs for 3,736 samples)

2. RESULTS AND DISCUSSION

2.1. EPIGEN Reference Panels

All the cohort were imputed with the bioinformatic framework and panels proposed in according to section “1.5.1 EPIGEN Reference Panels” from the Supplementary Material of The EPIGEN-Brazil Scientific Workflow.

2.2. Target phasing with different reference panels

After haplotype phasing, the number of infered phased SNPs has been evaluated for each chromosome as shown in Table S1, S2 and S3. It confirms that the number of target SNPs phasing with EPIGEN-5M reference results is 159,124; 175,415 and 142,344 more SNPs than phasing with 1KGP for Salvador, Bambuí and Pelotas Cohorts, respectively.

2.2.1. EPIGEN 2.5M target dataset - Salvador Cohort

Table S1. Number of target SNPs before and after haplotype phase inference with 1KGP or EPIGEN-5M as reference.

Chr	Number of SNPs		
	Target	Imputation Basis Target phased with:	
	Study SNPs	1KGP	EPIGEN-5M
1	177,631	159,646	173,489
2	187,900	170,109	183,587
3	158,978	143,654	155,326
4	148,238	134,257	144,833
5	141,145	127,699	137,731
6	148,058	133,326	144,877
7	124,862	113,039	122,007
8	121,715	110,781	119,087
9	99,315	90,863	97,125
10	115,491	104710	112,932
11	112,252	101,493	109,566
12	108,975	98,221	106,504
13	80,932	73,605	79,150
14	74,143	67,237	72,643
15	69,858	63,523	68,336
16	73,445	66,970	71,789
17	63,433	57,453	61,895
18	66,449	61,000	65,117
19	45,034	40,535	44,076
20	54,510	50,183	53,409
21	30,939	28,209	30,310
22	31,452	29,029	30,877
Total	2,234,755	2,025,542	2,184,666

2.2.2. EPIGEN 2.5M target dataset - Bambui Cohort

Table S2. Number of target SNPs before and after haplotype phase inference with 1KGP or EPIGEN-5M as reference.

Chr	Number of SNPs		
	Target	Imputation Basis Target phased with:	
	Study SNPs	1KGP	EPIGEN-5M
1	177,524	158,324	173,428
2	187,794	168,624	183,716
3	158,881	142,508	155,401
4	148,174	133,283	144,885
5	141,083	126,514	137,753
6	147,973	132,344	144,821
7	124,803	112,134	121,971
8	121,650	109,984	119,051
9	99,276	90,201	97,223
10	115,431	103,863	112,953
11	112,209	100,656	109,610
12	108,933	97,324	106,456
13	80,909	73,057	79,236
14	74,117	66,707	72,579
15	69,828	63,018	68,357
16	73,391	66,381	71,751
17	63,392	56,948	61,910
18	66,425	60,591	65,084
19	45,016	40,262	44,078
20	54,490	49,807	53,343
21	30,930	27,994	30,326
22	31,436	28,841	30,848
Total	2,233,665	2,009,365	2,184,780

2.2.3. EPIGEN 2.5M target dataset - Pelotas Cohort

Table S3. Number of target SNPs before and after haplotype phase inference with 1KGP or EPIGEN-5M as reference.

Chr	Number of SNPs		
	Target	Imputation Basis Target phased with:	
	Study SNPs	1KGP	EPIGEN-5M
1	177,655	160,226	172,566
2	187,920	170,653	182,759
3	158,996	144,030	154,595
4	148,255	134,639	144,234
5	141,162	128,092	137,032
6	148,063	133,751	144,173
7	124,875	113,440	121,443
8	121,717	111,155	118,541
9	99,334	91,235	96,622
10	115,502	104,978	112,507
11	112,273	101,834	108,983
12	108,987	98,557	105,958
13	80,944	73,792	78,859
14	74,144	67,440	72,247
15	69,864	63,679	68,050
16	73,452	67,179	71,385
17	63,437	57,620	61,586
18	66,455	61,188	64,845
19	45,040	40,703	43,836
20	54,515	50,321	53,146
21	30,941	28,271	30,188
22	31,454	29,113	30,685
Total	2,234,985	2,031,896	2,174,240

2.3. Imputation Results

2.3.1. Number of SNPs throughout the imputation process

Tables S4 to S9 and Figures S1 to S3 show how the number of SNPs vary along the imputation process for chromosomes 1 to 22, in particular, the amount of target SNPs before and after haplotype phasing, total output SNPs, and number of SNPs after filtering for info ≥ 0.8 for different reference panels (1KGP and EPIGEN-5M+1KGP). As expected, we observe an increasing in the number of SNPs available for follow-up analyses after imputation for both panels, as described below for each cohort.

2.3.1.1. EPIGEN 2.5M target dataset - Salvador Cohort

EPIGEN 2.5M target dataset - Salvador Cohort imputed data using EPIGEN-5M+1KGP reference panel provides more output SNPs than the 1KGP reference panel (172,877 more SNPs in total and 600,870 more SNPs with info ≥ 0.8). Specifically, while the EPIGEN-5M+1KGP reference panel increased the number of SNPs imputed in approximately 17.02 times (35,809,959 more SNPs), with the 1KGP panel this increasing was 16.95 times (35,637,082 more SNPs). In addition, when comparing well imputed SNPs (info ≥ 0.8), EPIGEN-5M+1KGP reference panel increased the number of SNPs imputed in approximately 9.70 times (19,450,725 more SNPs) and 1KGP panel, 9.43 times (18,849,855 more SNPs) (Figure S1 and Tables S4 and S5).

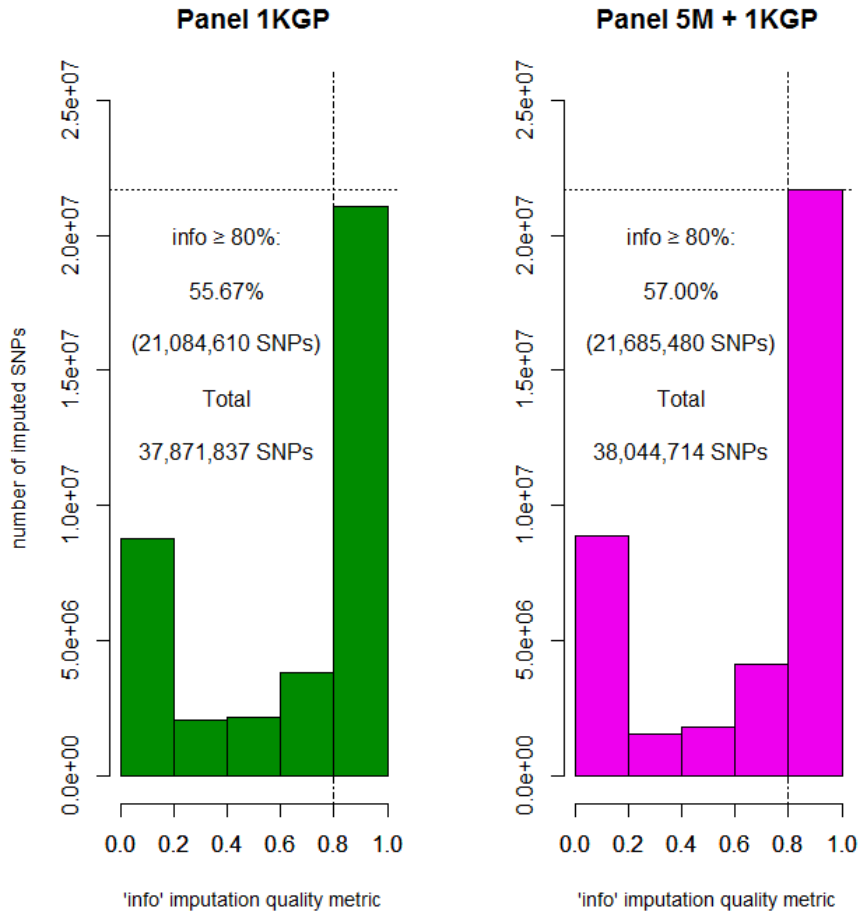

Fig S1: Comparison between 1KGP and EPIGEN-5M+1KGP reference panels for autosomal chromosomes showing the distribution of the info quality metric for imputation results. The dashed vertical line indicates the 0.8 threshold value and the horizontal one delimitates the highest number of SNPs info ≥ 0.8 achieved by a reference panel.

Table S4. Number of target SNPs before imputation, after phasing with 1KGP, the total output and after filtering for info ≥ 0.8 for 1KGP reference panel for chromosome 1 to 22.

Chr	Target Study SNPs	Imputation Basis: Phased 1KGP	Imputation Output:	
			Total Panel 1KGP	Filtered (info≥0.8) Panel 1KGP
1	177,631	159,646	2,979,836	1,646,301
2	187,900	170,109	3,277,621	1,812,056
3	158,978	143,654	2,739,450	1,533,257
4	148,238	134,257	2,712,882	1,533,461
5	141,145	127,699	2,508,712	1,409,011
6	148,058	133,326	2,404,586	1,371,594
7	124,862	113,039	2,196,064	1,234,911
8	121,715	110,781	2,164,610	1,209,012
9	99,315	90,863	1,638,139	905,800
10	115,491	104710	1,866,666	1,057,521
11	112,252	101,493	1,876,901	1,049,638
12	108,975	98,221	1,810,741	1,012,030
13	80,932	73,605	1,360,912	770,625
14	74,143	67,237	1,245,303	686,704
15	69,858	63,523	1,120,838	613,599
16	73,445	66,970	1,199,498	648,208
17	63,433	57,453	1,035,003	559,326
18	66,449	61,000	1,079,227	601,121
19	45,034	40,535	806,910	442,102
20	54,510	50,183	847,495	449,938
21	30,939	28,209	511,350	276,796
22	31,452	29,029	489,093	261,599
Total SNPs:	2,234,755	2,025,542	37,871,837	21,084,610

Table S5. Number of target SNPs before imputation, after phasing with EPIGEN-5M, the total output and after filtering for info ≥ 0.8 for EPIGEN-5M+1KGP reference panel for chromosome 1 to 22.

Chr	Target Study SNPs	Imputation Basis: Phased 5M	Imputation Output:	
			Total Panel 5M+1KGP	Filtered (info≥0.8) Panel 5M+1KGP
1	177,631	173,489	2,994,750	1,687,534
2	187,900	183,587	3,291,883	1,864,569
3	158,978	155,326	2,751,975	1,570,350
4	148,238	144,833	2,724,269	1,574,398
5	141,145	137,731	2,519,571	1,442,317
6	148,058	144,877	2,417,345	1,413,851
7	124,862	122,007	2,205,800	1,266,719
8	121,715	119,087	2,173,320	1,242,172
9	99,315	97,125	1,644,966	932,194
10	115,491	112,932	1,875,473	1,083,178
11	112,252	109,566	1,885,520	1,075,951
12	108,975	106,504	1,819,608	1,041,116
13	80,932	79,150	1,367,054	795,207
14	74,143	72,643	1,250,896	705,755
15	69,858	68,336	1,126,228	632,083
16	73,445	71,789	1,205,061	668,314
17	63,433	61,895	1,040,176	577,387
18	66,449	65,117	1,083,880	619,366
19	45,034	44,076	810,818	456,377
20	54,510	53,409	851,108	472,455
21	30,939	30,310	513,749	289,588
22	31,452	30,877	491,264	274,599
Total SNPs:	2,234,755	2,184,666	38,044,714	21,685,480

2.3.1.2. EPIGEN 2.5M target dataset - Bambui Cohort

EPIGEN 2.5M target dataset - Bambuí Cohort imputed data using EPIGEN-5M+1KGP reference panel provides more output SNPs than the 1KGP reference panel (173,470 more SNPs in total and 672,840 more SNPs with info ≥ 0.8). Specifically, while the EPIGEN-5M+1KGP reference panel increased the number of SNPs imputed in approximately 17.03 times (35,811,642 more SNPs), with the 1KGP panel this increasing was 16.96 times (35,638,172 more SNPs). In addition, when comparing well imputed SNPs (info ≥ 0.8), EPIGEN-5M+1KGP reference panel increased the number of SNPs imputed in approximately 9.15 times (18,210,274 more SNPs) and 1KGP panel, 8.85 times (17,537,434 more SNPs) (Figure S2 and Tables S6 and S7).

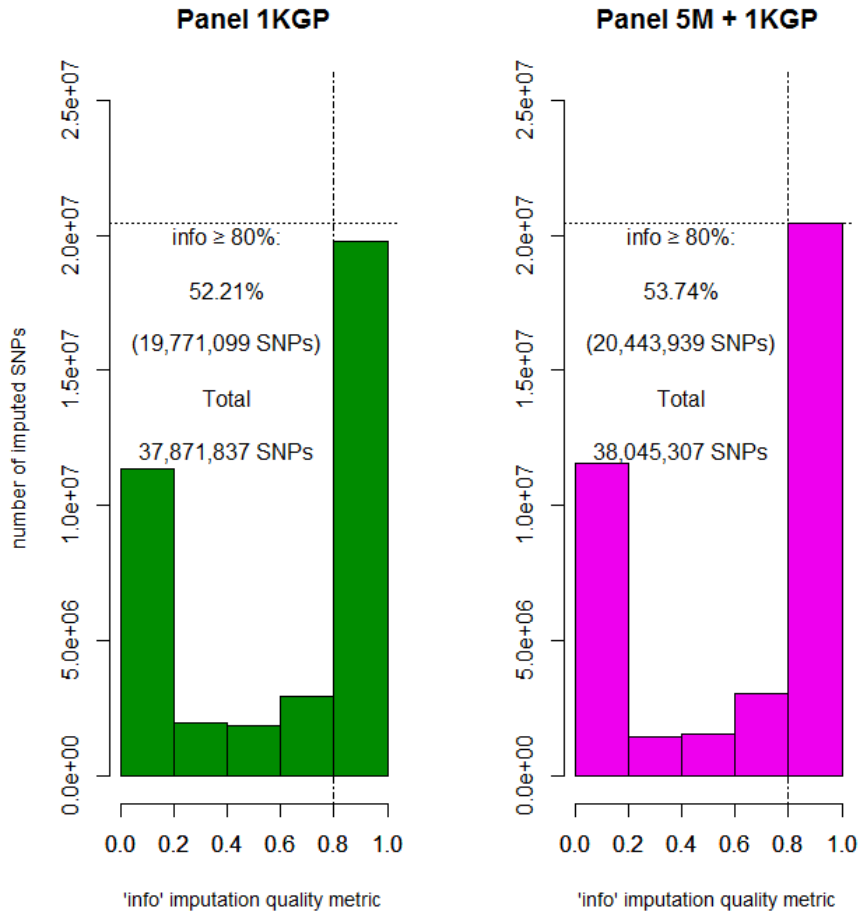

Fig S2: Comparison between 1KGP and EPIGEN-5M+1KGP reference panels for autosomal chromosomes showing the distribution of the info quality metric for imputation results. The dashed vertical line indicates the 0.8 threshold value and the horizontal one delimitates the highest number of SNPs info ≥ 0.8 achieved by a reference panel.

Table S6. Number of target SNPs before imputation, after phasing with 1KGP, the total output and after filtering for info ≥ 0.8 for 1KGP reference panel for chromosome 1 to 22.

Chr	Target Study SNPs	Imputation Basis: Phased 1KGP	Imputation Output:	
			Total Panel 1KGP	Filtered (info≥0.8) Panel 1KGP
1	177,524	158,324	2,979,836	1,542,057
2	187,794	168,624	3,277,621	1,695,066
3	158,881	142,508	2,739,450	1,440,207
4	148,174	133,283	2,712,882	1,443,903
5	141,083	126,514	2,508,712	1,319,810
6	147,973	132,344	2,404,586	1,293,877
7	124,803	112,134	2,196,064	1,156,067
8	121,650	109,984	2,164,610	1,133,423
9	99,276	90,201	1,638,139	852,343
10	115,431	103,863	1,866,666	991,313
11	112,209	100,656	1,876,901	982,116
12	108,933	97,324	1,810,741	948,423
13	80,909	73,057	1,360,912	726,528
14	74,117	66,707	1,245,303	644,582
15	69,828	63,018	1,120,838	573,042
16	73,391	66,381	1,199,498	605,884
17	63,392	56,948	1,035,003	525,472
18	66,425	60,591	1,079,227	562,880
19	45,016	40,262	806,910	418,188
20	54,490	49,807	847,495	410,909
21	30,930	27,994	511,350	258,327
22	31,436	28,841	489,093	246,682
Total SNPs:	2,233,665	2,009,365	37,871,837	19,771,099

Table S7. Number of target SNPs before imputation, after phasing with EPIGEN-5M, the total output and after filtering for info ≥ 0.8 for EPIGEN-5M+1KGP reference panel for chromosome 1 to 22.

Chr	Target Study SNPs	Imputation Basis: Phased 5M	Imputation Output:	
			Total Panel 5M+1KGP	Filtered (info≥0.8) Panel 5M+1KGP
1	177,524	173,428	2,994,759	1,585,046
2	187,794	183,716	3,291,952	1,750,352
3	158,881	155,401	2,752,048	1,480,816
4	148,174	144,885	2,724,301	1,486,478
5	141,083	137,753	2,519,608	1,355,547
6	147,973	144,821	2,417,359	1,340,749
7	124,803	121,971	2,205,832	1,187,378
8	121,650	119,051	2,173,367	1,170,159
9	99,276	97,223	1,645,007	883,112
10	115,431	112,953	1,875,481	1,023,024
11	112,209	109,610	1,885,543	1,013,587
12	108,933	106,456	1,819,645	979,336
13	80,909	79,236	1,367,077	754,073
14	74,117	72,579	1,250,931	667,144
15	69,828	68,357	1,126,263	594,181
16	73,391	71,751	1,205,071	632,080
17	63,392	61,910	1,040,200	548,339
18	66,425	65,084	1,083,897	586,022
19	45,016	44,078	810,833	436,555
20	54,490	53,343	851,114	434,664
21	30,930	30,326	513,760	273,691
22	31,436	30,848	491,259	261,606
Total SNPs:	2,233,665	2,184,780	38,045,307	20,443,939

2.3.1.3. EPIGEN 2.5M target dataset - Pelotas Cohort

EPIGEN 2.5M target dataset - Pelotas Cohort imputed data using EPIGEN-5M+1KGP reference panel provides more output SNPs than the 1KGP reference panel (168,058 more SNPs in total and 380,210 more SNPs with info ≥ 0.8). Specifically, while the EPIGEN-5M+1KGP reference panel increased the number of SNPs imputed in approximately 17.02 times (35,804,910 more SNPs), with the 1KGP panel this increasing was 16.95 times (35,636,852 more SNPs). In addition, when comparing well imputed SNPs (info ≥ 0.8), EPIGEN-5M+1KGP reference panel increased the number of SNPs imputed in approximately 9.42 times (18,829,912 more SNPs) and 1KGP panel, 9.25 times (18,449,702 more SNPs) (Figure S3 and Tables S8 and S9).

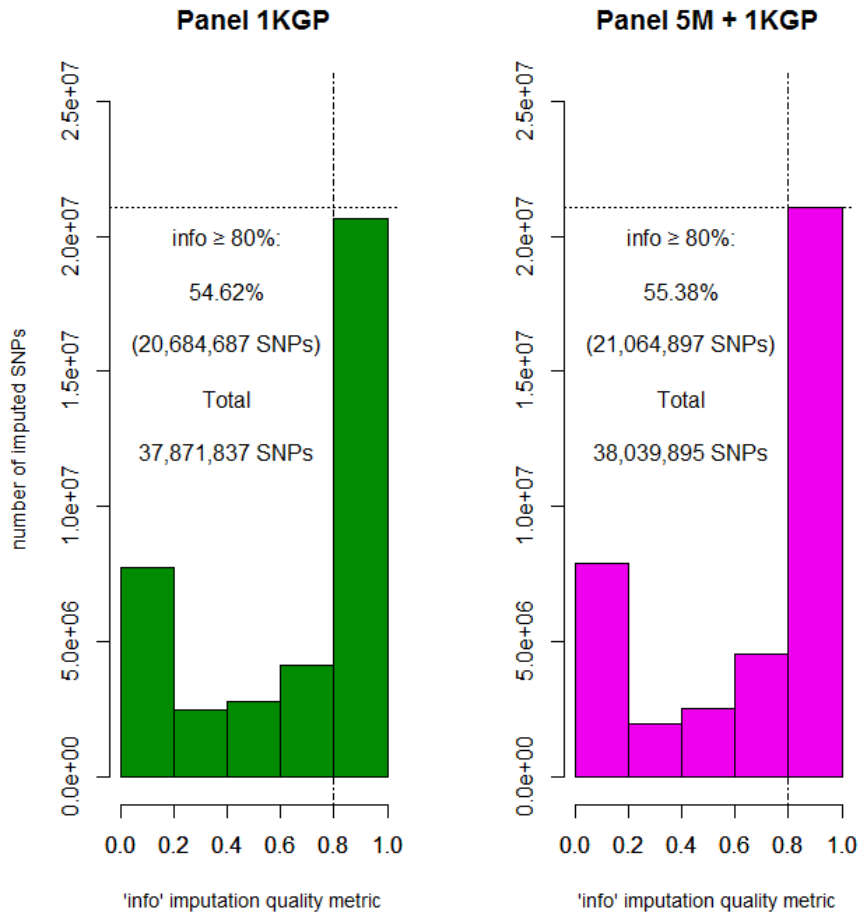

Fig S3: Comparison between 1KGP and EPIGEN-5M+1KGP reference panels for autosomal chromosomes showing the distribution of the info quality metric for imputation results. The dashed vertical line indicates the 0.8 threshold value and the horizontal one delimitates the highest number of SNPs info ≥ 0.8 achieved by a reference panel.

Table S8. Number of target SNPs before imputation, after phasing with 1KGP, the total output and after filtering for info ≥ 0.8 for 1KGP reference panel for chromosome 1 to 22.

Chr	Target Study SNPs	Imputation Basis: Phased 1KGP	Imputation Output:	
			Total Panel 1KGP	Filtered (info≥0.8) Panel 1KGP
1	177,655	160,226	2,979,836	1,616,279
2	187,920	170,653	3,277,621	1,779,391
3	158,996	144,030	2,739,450	1,504,329
4	148,255	134,639	2,712,882	1,504,690
5	141,162	128,092	2,508,712	1,382,961
6	148,063	133,751	2,404,586	1,348,003
7	124,875	113,440	2,196,064	1,210,903
8	121,717	111,155	2,164,610	1,186,631
9	99,334	91,235	1,638,139	889,047
10	115,502	104,978	1,866,666	1,037,169
11	112,273	101,834	1,876,901	1,030,198
12	108,987	98,557	1,810,741	991,891
13	80,944	73,792	1,360,912	754,566
14	74,144	67,440	1,245,303	674,406
15	69,864	63,679	1,120,838	600,358
16	73,452	67,179	1,199,498	635,431
17	63,437	57,620	1,035,003	547,706
18	66,455	61,188	1,079,227	590,691
19	45,040	40,703	806,910	434,980
20	54,515	50,321	847,495	438,940
21	30,941	28,271	511,350	270,836
22	31,454	29,113	489,093	255,281
Total SNPs:	2,234,985	2,031,896	37,871,837	20,684,687

Table S9. Number of target SNPs before imputation, after phasing with EPIGEN-5M, the total output and after filtering for info ≥ 0.8 for EPIGEN-5M+1KGP reference panel for chromosome 1 to 22.

Chr	Target Study SNPs	Imputation Basis: Phased 5M	Imputation Output:	
			Total Panel 5M+1KGP	Filtered (info≥0.8) Panel 5M+1KGP
1	177,655	172,566	2,994,278	1,636,495
2	187,920	182,759	3,291,502	1,810,417
3	158,996	154,595	2,751,642	1,527,773
4	148,255	144,234	2,723,953	1,528,575
5	141,162	137,032	2,519,262	1,400,444
6	148,063	144,173	2,417,031	1,376,474
7	124,875	121,443	2,205,553	1,227,517
8	121,717	118,541	2,173,078	1,208,562
9	99,334	96,622	1,644,764	907,522
10	115,502	112,507	1,875,218	1,054,731
11	112,273	108,983	1,885,249	1,044,387
12	108,987	105,958	1,819,384	1,010,125
13	80,944	78,859	1,366,924	772,395
14	74,144	72,247	1,250,742	686,123
15	69,864	68,050	1,126,098	613,240
16	73,452	71,385	1,204,872	649,405
17	63,437	61,586	1,040,055	560,181
18	66,455	64,845	1,083,741	604,223
19	45,040	43,836	810,696	443,669
20	54,515	53,146	851,001	455,394
21	30,941	30,188	513,698	280,852
22	31,454	30,685	491,154	266,393
Total SNPs:	2,234,985	2,174,240	38,039,895	21,064,897

2.3.2. Minor Allele Frequency (MAF) distribution throughout the process

We also compared the MAF distribution before and after imputation with the different panels without filtering for any info threshold (FigureS S4, S6, S8). For the three cohorts, the MAF distribution shows a considerable increase on the number of very rare SNPs (MAF ≤ 0.3%) when imputing the Target dataset (3 cohorts) with the 1KGP or the EPIGEN-5M+1KGP reference panel. Approximately 20% of the variants imputed with the 1KGP and the EPIGEN-5M+1KGP references are common (MAF > 5%). These findings are compatible with the MAF distribution through reference panels seen on Figure 1B in the Main Text.

This analysis was repeated for SNPs with info ≥ 0.8 cuttoff (Figure S5, S7, S9). After this filter, it shows a small increasing for very rare SNPs when imputing the Target dataset (3 cohorts) with the 1KGP and the EPIGEN-5M+1KGP reference panel. Finally, about 35-40% of the variants imputed with the 1KGP and the EPIGEN-5M+1KGP reference panels are common.

2.3.2.1. EPIGEN 2.5M target dataset - Salvador Cohort

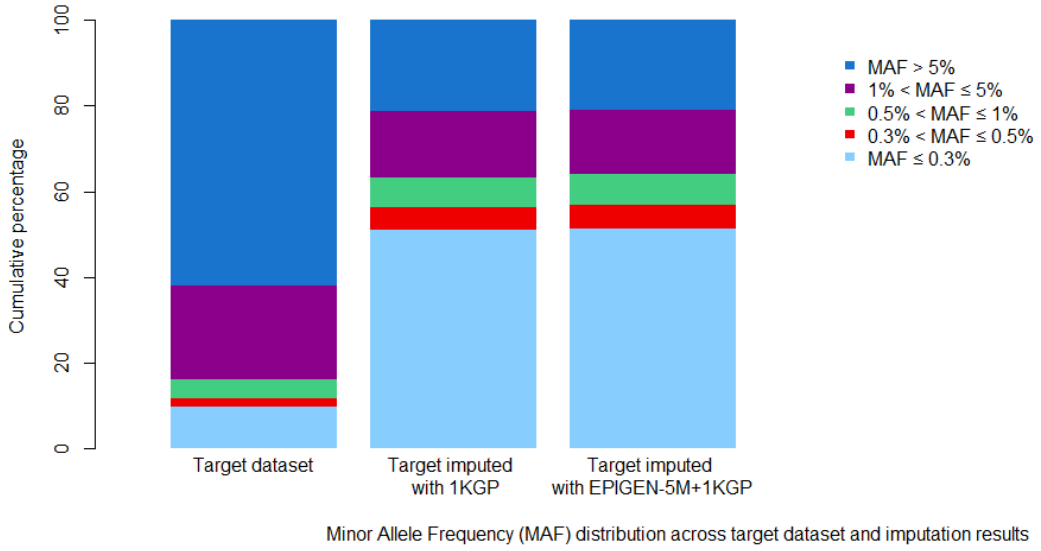

Figure S4. The cumulative percentage of variants by Minor Allele Frequency (MAF) of target dataset before and after imputation with distinct reference panels, without filtering for any info cutoff threshold.

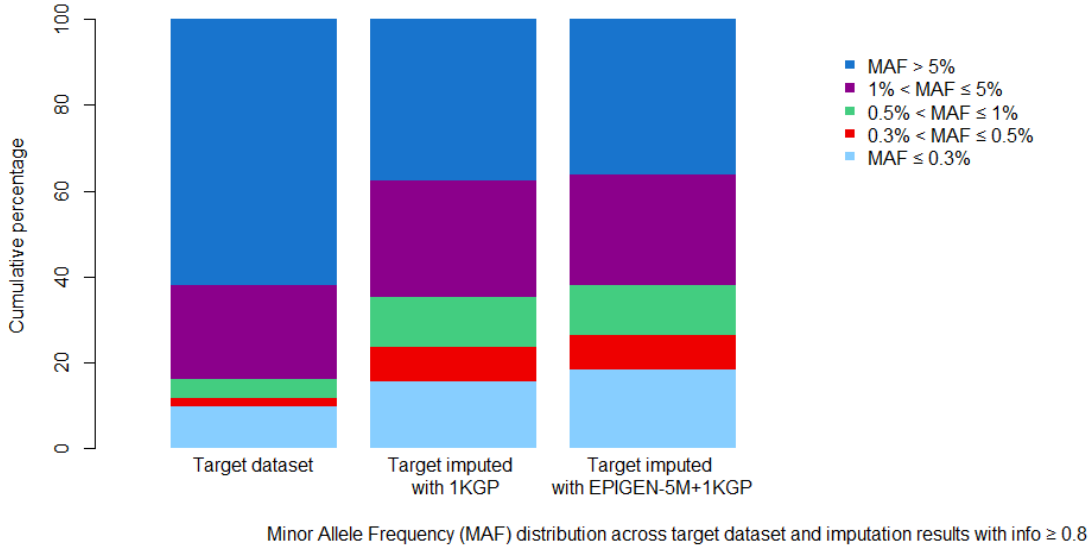

Figure S5. The cumulative percentage of variants by Minor Allele Frequency (MAF) of target dataset before and after imputation with distinct reference panels, using the cutoff of info ≥ 0.8.

2.3.2.2. EPIGEN 2.5M target dataset - Bambui Cohort

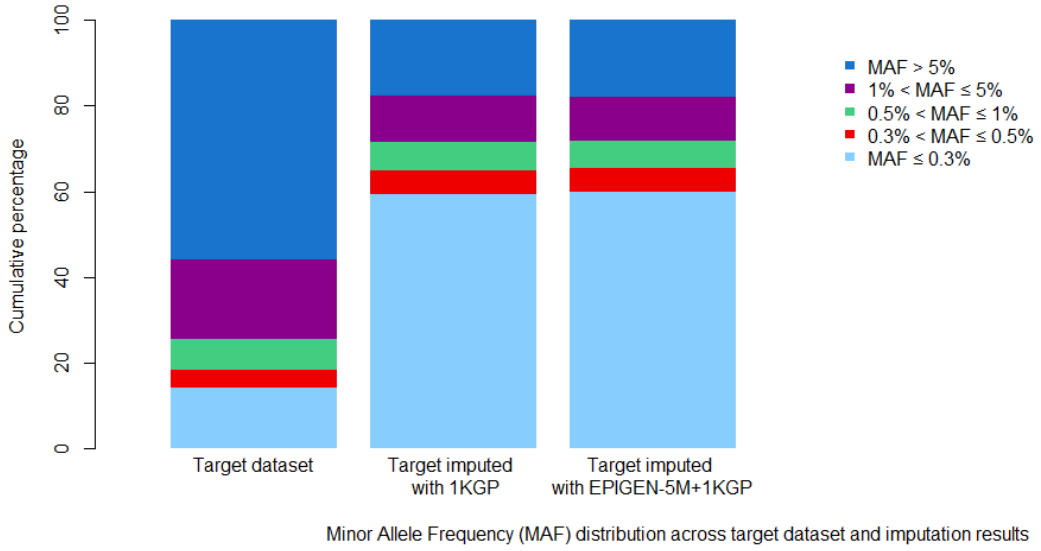

Figure S6. The cumulative percentage of variants by Minor Allele Frequency (MAF) of target dataset before and after imputation with distinct reference panels, without filtering for any info cutoff threshold.

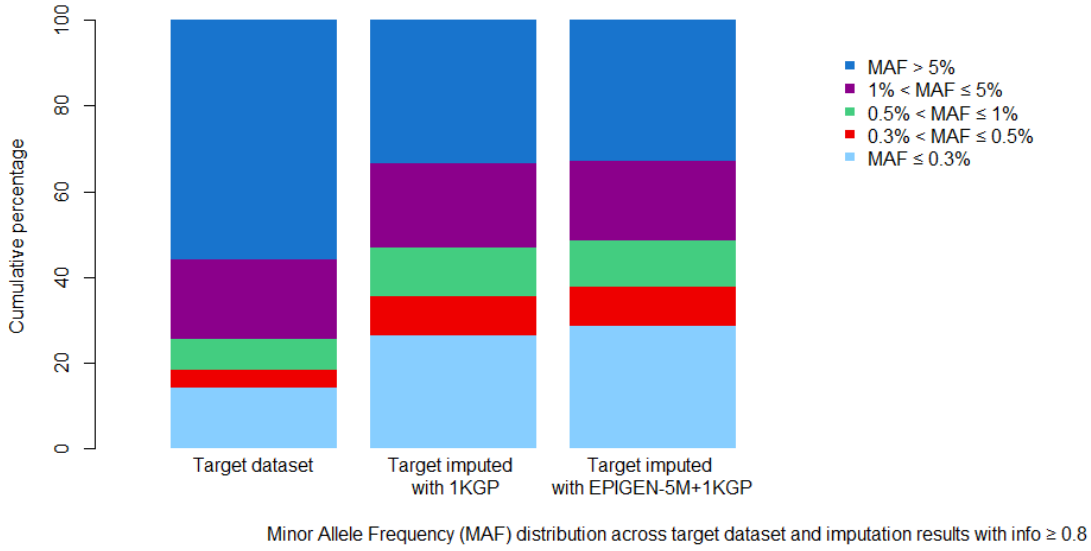

Figure S7. The cumulative percentage of variants by Minor Allele Frequency (MAF) of target dataset before and after imputation with distinct reference panels, using the cutoff of info ≥ 0.8.

2.3.2.3. EPIGEN 2.5M target dataset - Pelotas Cohort

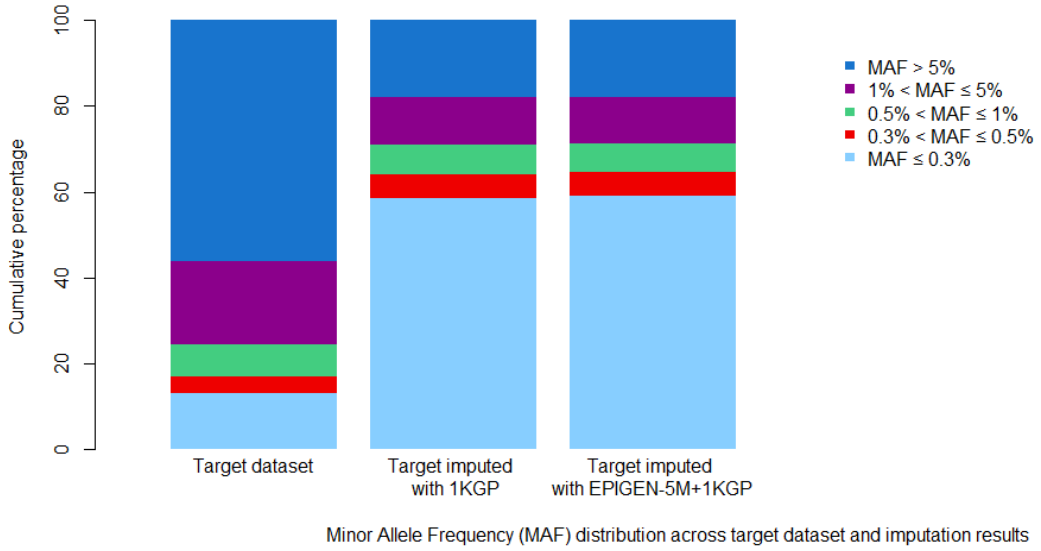

Figure S8. The cumulative percentage of variants by Minor Allele Frequency (MAF) of target dataset before and after imputation with distinct reference panels, without filtering for any info cutoff threshold.

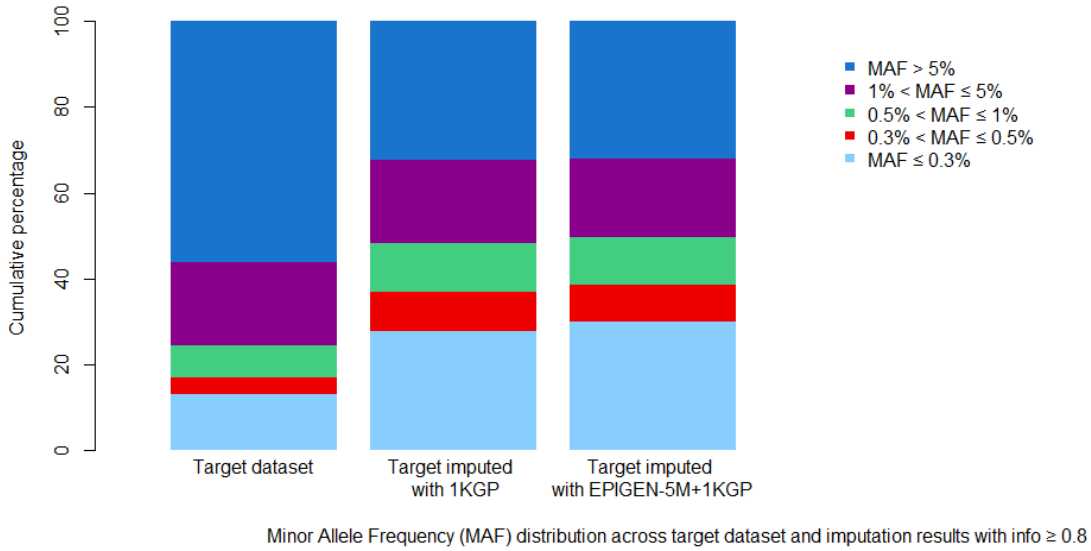

Figure S9. The cumulative percentage of variants by Minor Allele Frequency (MAF) of target dataset before and after imputation with distinct reference panels, using the cutoff of info ≥ 0.8.

Finally, we compared the MAF of each target dataset (EPIGEN 2.5M dataset and each cohort dataset) (Figure S10) trying to understand the differences between the cohorts imputation performance Figure S10 and Table S11. Further experiments and analysis are necessary for a better understanding of these results and will be performed.

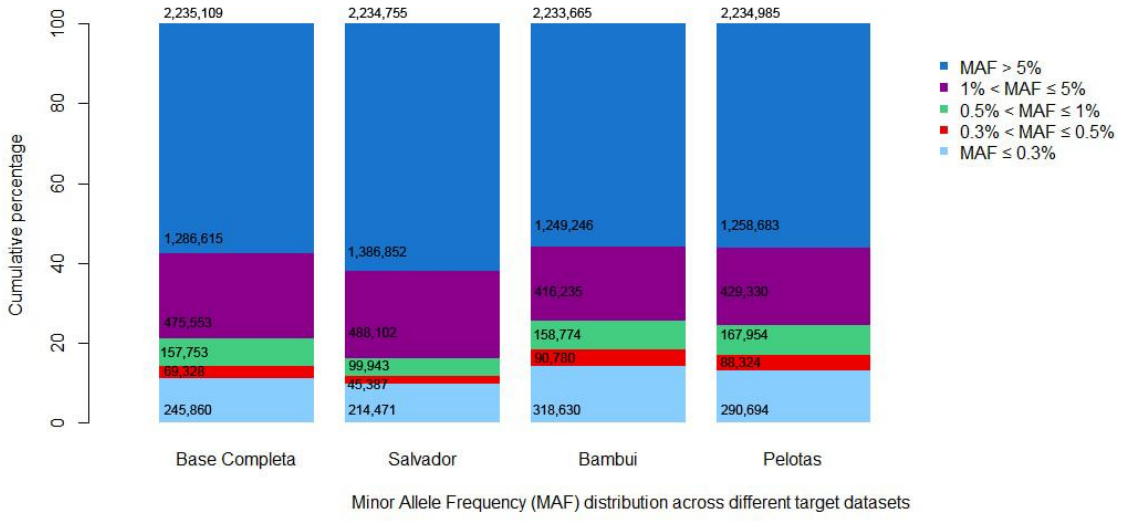

Figure S10. The proportion of variants by Minor Allele Frequency (MAF) of each EPIGEN 2.5M target dataset before imputation. The number of SNPs is indicated for each category and the total number of SNPs is above the graph.

Table S10. Summary of the number of target SNPs before and after haplotype phase inference with 1KGP or EPIGEN-5M as reference for the whole EPIGEN 2.5M dataset and for each cohort dataset.

Target and Imputation Basis

EPIGEN 2.5M

	Number of SNPs		
Chr	Target	Imputation Basis Target phased with:	
	Study SNPs	Phased 1KGP	Phased 5M
Total	2,235,109	2,038,898	2,167,159

Salvador Cohort

	Number of SNPs		
Chr	Target	Imputation Basis Target phased with:	
	Study SNPs	Phased 1KGP	Phased 5M
Total	2,234,755	2,025,542	2,184,666

Bambuí Cohort

	Number of SNPs		
Chr	Target	Imputation Basis Target phased with:	
	Study SNPs	Phased 1KGP	Phased 5M
Total	2,233,665	2,009,365	2,184,780

Pelotas Cohort

	Number of SNPs		
Chr	Target	Imputation Basis Target phased with:	
	Study SNPs	Phased 1KGP	Phased 5M
Total	2,234,985	2,031,896	2,174,240

Table S11. Summary of the number of target SNPs before and after imputation with 1KGP or EPIGEN-5M as reference for the whole EPIGEN 2.5M dataset and for each cohort dataset. The number of SNPs after quality control filtering for "info" ≥ 0.8 and their proportion in relation to the total number of SNPs are also described.

Target and Output			
EPIGEN 2.5M			
		Number of SNPs	
Chr	**Target**	**Imputation Output:**	
	Study SNPs	**Panel 1KGP**	**Panel 5M+1KGP**
Total SNPs:	2,235,109	37,871,837	38,035,721
SNPs info ≥ 80%:	-	20,968,285	21,562,106
Percentage info ≥ 80%:	-	55.37%	56.69%
Salvador Cohort			
		Number of SNPs	
Chr	**Target**	**Imputation Output:**	
	Study SNPs	**Panel 1KGP**	**Panel 5M+1KGP**
Total SNPs:	2,234,755	37,871,837	38,044,714
SNPs info ≥ 80%:	-	21,084,610	21,685,480
Percentage info ≥ 80%:	-	55.67%	57.00%
Bambuí Cohort			
		Number of SNPs	
Chr	**Target**	**Imputation Output:**	
	Study SNPs	**Panel 1KGP**	**Panel 5M+1KGP**
Total SNPs:	2,233,665	37,871,837	38,045,307
SNPs info ≥ 80%:	-	19,771,099	20,443,939
Percentage info ≥ 80%:	-	52.21%	53.74%
Pelotas Cohort			
		Number of SNPs	
Chr	**Target**	**Imputation Output:**	
	Study SNPs	**Panel 1KGP**	**Panel 5M+1KGP**
Total SNPs:	2,234,985	37,871,837	38,039,895
SNPs info ≥ 80%:	-	20,684,687	21,064,897
Percentage info ≥ 80%:	-	54.62%	55.38%

Supplementary References

Barreto ML, Cunha SS, Alcantara-Neves N, Carvalho LP, Cruz AA, Stein RT, Genser B, Cooper PJ, Rodrigues LC. 2006. Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in Northeastern Brazil (Salvador-SCAALA study). *BMC pulmonary medicine* **6**: 15.

Lima-Costa MF, Firmo JO, Uchoa E. 2011. Cohort profile: the Bambui (Brazil) Cohort Study of Ageing. *International journal of epidemiology* **40**: 862-867.

Victora CG, Barros FC. 2006. Cohort profile: the 1982 Pelotas (Brazil) birth cohort study. *International journal of epidemiology* **35**: 237-242.