**Supplementary Information A**

All seven arrays were updated to the GRCh37 genome build. This section contains a summary of the processes required for the update of each array. All strand files can be downloaded from Rayner, n.d-a, n.d-b, and n.d-c.

**Illumina 1.2M**

The Chipendium, a specifically designed online tool, was used to identify the correct strand for this array (Robertson & Rayner, 2018). The Illumina 1.2M array showed a 100% chance of a correct match with the Human1-2M-DuoCustom\_v1\_A ILMN strand. The GRCh37 version of the corresponding strand file was used to update this array to GRCh37, using a series of Unix and Plink commands (Robertson, 2012).

**Illumina 15k Custom Chip**

The same procedure as above was followed to update this array. The Chipendium was not able to find a match for the array, and so three custom strand files were generated, one each for GRCh35, GRCh36, and GRCh37, as the current build was unknown. The array build was updated using each of these three files separately, and the number of SNPs changed in the output for each was compared (Robertson, 2012). GRGCh36 showed the greatest similarity with the array data, and so the current data was assumed to be GRCh36. This was then updated a second time, this time to GRCh37.

**Illumina Human 660-Quad**

The array data was uploaded to the Chipendium, and showed a likelihood of 91.33% that the array matches to the Human610-Quadv1\_B SRC strand. The same procedure was followed as in Illumina 1.2M to update the array to build GRCh37.

**Infinium 550k v1.1**

Array data was uploaded to the Chipendium, and a 100% possibility of a correct match was found with the BDCHP-1X10-HUMANHAP550\_11218540\_C SRC strand. The same procedure as above was followed to update to GRCh37.

**Infinium 550k v3**

The same procedure was used. The Chipendium identified a 100% chance of match with the HumanHap550-2v3\_B ILMN strand, and the HumanHap550v3\_A ILMN strand. The HumanHap550-2v3\_B ILMN strand was used to update the build to GRCh37.

**Affymetrix 500k**

The Chipendium was unable to identify a match for this array, and so a series of custom strand files were generated in the form of GRCh35, GRCh36, and CHRCh37. The SNP identifiers in this array were labelled using the SNP\_A naming convention, which was not consistent with the remaining six files, which used rsIDs. Before updating the genome build, the identifiers were updated to rsIDs for consistency in meta-analysis. The file showed a 100% match to GRCh35, and was directly updated to GRCh37 file using the same process.

**Affymetrix v6**

The Chipendium was unable to identify a match for this array. The build was updated using custom-designed files to GRCh35, GRCh36, and GRCh37. The original file showed the most similarity with the GRCh36 updated version, and so the current build was assumed to be GRCh36. This file was updated a second time to GRCh37.

**References**

Rayner, W. (n.d.-a). *Illmn Strand Data*. Genotyping Chips Strand and Build Files. Retrieved 15 April 2022, from https://www.well.ox.ac.uk/%7Ewrayner/strand/ilmnStrand/index.html

Rayner, W. (n.d.-b). *Source Strand Data*. Genotyping Chips Strand and Build Files. Retrieved 15 April 2022, from https://www.well.ox.ac.uk/%7Ewrayner/strand/sourceStrand/index.html

Rayner, W. (n.d.-c). *Strand Home*. Genotyping Chips Strand and Build Files. Retrieved 15 April 2022, from [https://www.well.ox.ac.uk/%7Ewrayner/strand/index.html](https://www.well.ox.ac.uk/~wrayner/strand/index.html)

Robertson, N. (2012, January 17). *Update Build*. Genotyping Chips Strand and Build Files. Retrieved 18 April 2022, from https://www.well.ox.ac.uk/%7Ewrayner/strand/update\_build.sh

Robertson, N., & Rayner, W. (2018, October 18). *Chipendium*. Chipendium. Retrieved 18 April 2022, from http://mccarthy.well.ox.ac.uk/static/software/chipendium/

**Supplementary Information B**

The imputed output of all seven arrays was checked to produce tables of R2 imputation quality values and the frequency of alternate alleles, with reference and alternate alleles as defined by the reference panel. In most cases, the reference allele is the major allele, and the alternate allele is the minor allele. A post-imputation script from Rayner (2016) was used.

**Supplementary Table S1**

Frequencies of alternate allele frequencies, as designated by the reference panel, across the imputed genome from the Illumina 1.2M array.

|  |  |  |
| --- | --- | --- |
| **Alternate Allele Frequency** | **Count** | **%** |
| **0.0** | 35607197 | 88.13 |
| **.1** | 1224652 | 3.03 |
| **.2** | 812611 | 2.01 |
| **.3** | 618200 | 1.53 |
| **.4** | 497274 | 1.23 |
| **.5** | 405821 | 1.00 |
| **.6** | 351193 | 0.87 |
| **.7** | 293504 | 0.73 |
| **.8** | 245350 | 0.61 |
| **.9** | 348842 | 0.86 |
| **1.0** | 861 | 0.00 |
| **Total** | **40405505** | **100.00** |

**Supplementary Table S2**

Frequencies of genome-wide imputation scores for the Illumina 1.2M array.

|  |  |  |
| --- | --- | --- |
| **R2** | **Count** | **%** |
| **0.0** | 16662209 | 41.24 |
| **.1** | 1521038 | 3.76 |
| **.2** | 959126 | 2.37 |
| **.3** | 829163 | 2.05 |
| **.4** | 882371 | 2.18 |
| **.5** | 1130194 | 2.80 |
| **.6** | 1671531 | 4.14 |
| **.7** | 2473215 | 6.12 |
| **.8** | 3277542 | 8.12 |
| **.9** | 10765510 | 26.64 |
| **1.0** | 233606 | 0.58 |
| **Total** | **40405505** | **100.00** |

**Supplementary Table S3**

Frequencies of alternate allele frequencies, as designated by the reference panel, across the imputed genome from the Illumina 15k Custom Chip.

|  |  |  |
| --- | --- | --- |
| **Alternate Allele Frequency** | **Count** | **%** |
| **0.0** | 36476629 | 90.45 |
| **.1** | 1377926 | 3.42 |
| **.2** | 856060 | 2.12 |
| **.3** | 602297 | 1.49 |
| **.4** | 428755 | 1.06 |
| **.5** | 289753 | 0.72 |
| **.6** | 176674 | 0.44 |
| **.7** | 82769 | 0.21 |
| **.8** | 31028 | 0.08 |
| **.9** | 6532 | 0.02 |
| **Total** | **40328423** | **100.00** |

**Supplementary Table S4**

Frequencies of genome-wide imputation scores for the Illumina 15k Custom Chip.

|  |  |  |
| --- | --- | --- |
| **R2** | **Count** | **%** |
| **0.0** | 39134601 | 97.04 |
| **.1** | 446694 | 1.11 |
| **.2** | 219016 | 0.54 |
| **.3** | 145787 | 0.36 |
| **.4** | 104434 | 0.26 |
| **.5** | 82639 | 0.21 |
| **.6** | 66216 | 0.16 |
| **.7** | 46334 | 0.12 |
| **.8** | 26135 | 0.07 |
| **.9** | 56549 | 0.14 |
| **1.0** | 18 | 0.00 |
| **Total** | **40328423** | **100.00** |

**Supplementary Table S5**

Frequencies of alternate allele frequencies, as designated by the reference panel, across the imputed genome from the Illumina Human 660-Quad array.

|  |  |  |
| --- | --- | --- |
| **Alternate Allele Frequency** | **Count** | **%** |
| **0.0** | 35610360 | 88.15 |
| **.1** | 1218616 | 3.02 |
| **.2** | 817789 | 2.02 |
| **.3** | 617701 | 1.53 |
| **.4** | 495089 | 1.23 |
| **.5** | 407996 | 1.01 |
| **.6** | 350894 | 0.87 |
| **.7** | 293073 | 0.73 |
| **.8** | 244866 | 0.61 |
| **.9** | 341611 | 0.85 |
| **1.0** | 15 | 0.00 |
| **Total** | **40398010** | **100.00** |

**Supplementary Table S6**

Frequencies of genome-wide imputation scores for the Illumina Human 600-Quad array.

|  |  |  |
| --- | --- | --- |
| **R2** | **Count** | **%** |
| **0.0** | 22248791 | 55.07 |
| **.1** | 1131556 | 2.80 |
| **.2** | 644425 | 1.60 |
| **.3** | 526371 | 1.30 |
| **.4** | 549228 | 1.36 |
| **.5** | 654427 | 1.62 |
| **.6** | 996279 | 2.46 |
| **.7** | 1620281 | 4.01 |
| **.8** | 2463864 | 6.10 |
| **.9** | 9462511 | 23.42 |
| **1.0** | 100277 | 0.25 |
| **Total** | **40398010** | **100.00** |

**Supplementary Table S7**

Frequencies of alternate allele frequencies, as designated by the reference panel, across the imputed genome from the Infinium HumanHap 550K v1.1 array.

|  |  |  |
| --- | --- | --- |
| **Alternate Allele Frequency** | **Count** | **%** |
| **0.0** | 35600383 | 88.14 |
| **.1** | 1224388 | 3.03 |
| **.2** | 813285 | 2.01 |
| **.3** | 620164 | 1.54 |
| **.4** | 496316 | 1.23 |
| **.5** | 407235 | 1.01 |
| **.6** | 351583 | 0.87 |
| **.7** | 293797 | 0.73 |
| **.8** | 244873 | 0.61 |
| **.9** | 338990 | 0.84 |
| **1.0** | 18 | 0.00 |
| **Total** | **40391032** | **100.00** |

**Supplementary Table S8**

Frequencies of genome-wide imputation scores for the Infinium HumanHap 550K v1.1 array.

|  |  |  |
| --- | --- | --- |
| **R2** | **Count** | **%** |
| **0.0** | 20184982 | 49.97 |
| **.1** | 1322103 | 3.27 |
| **.2** | 779610 | 1.93 |
| **.3** | 655122 | 1.62 |
| **.4** | 684623 | 1.70 |
| **.5** | 841121 | 2.08 |
| **.6** | 1277979 | 3.16 |
| **.7** | 2010000 | 4.98 |
| **.8** | 2856586 | 7.07 |
| **.9** | 9698278 | 24.01 |
| **1.0** | 80628 | 0.20 |
| **Total** | **40391032** | **100.00** |

**Supplementary Table S9**

Frequencies of alternate allele frequencies, as designated by the reference panel, across the imputed genome from the Infinium HumanHap 550K v3 array.

|  |  |  |
| --- | --- | --- |
| **Alternate Allele Frequency** | **Count** | **%** |
| **0.0** | 35609256 | 88.14 |
| **.1** | 1224469 | 3.03 |
| **.2** | 814185 | 2.02 |
| **.3** | 619242 | 1.53 |
| **.4** | 496766 | 1.23 |
| **.5** | 408492 | 1.01 |
| **.6** | 350785 | 0.87 |
| **.7** | 292303 | 0.72 |
| **.8** | 246364 | 0.61 |
| **.9** | 339706 | 0.84 |
| **1.0** | 37 | 0.00 |
| **Total** | **40401605** | **100.00** |

**Supplementary Table S10**

Frequencies of genome-wide imputation scores for the Infinium HumanHap 550K v3 array.

|  |  |  |
| --- | --- | --- |
| **R2** | **Count** | **%** |
| **0.0** | 17377018 | 43.01 |
| **.1** | 1536894 | 3.80 |
| **.2** | 959979 | 2.38 |
| **.3** | 833098 | 2.06 |
| **.4** | 896083 | 2.22 |
| **.5** | 1153825 | 2.86 |
| **.6** | 1718481 | 4.25 |
| **.7** | 2539695 | 6.29 |
| **.8** | 3326762 | 8.23 |
| **.9** | 9983897 | 24.71 |
| **1.0** | 75873 | 0.18 |
| **Total** | **40401605** | **100.00** |

**Supplementary Table S11**

Frequencies of alternate allele frequencies, as designated by the reference panel, across the imputed genome from the Affymetrix 500k array.

|  |  |  |
| --- | --- | --- |
| **Alternate Allele Frequency** | **Count** | **%** |
| **0.0** | 34484690 | 88.16 |
| **.1** | 1189690 | 3.04 |
| **.2** | 787309 | 2.01 |
| **.3** | 602581 | 1.54 |
| **.4** | 482536 | 1.23 |
| **.5** | 394624 | 1.01 |
| **.6** | 339351 | 0.87 |
| **.7** | 282230 | 0.72 |
| **.8** | 238691 | 0.61 |
| **.9** | 315160 | 0.81 |
| **1.0** | 243 | 0.00 |
| **Total** | **39117105** | **100.00** |

**Supplementary Table S12**

Frequencies of genome-wide imputation scores for the Affymetrix 500k array.

|  |  |  |
| --- | --- | --- |
| **R2** | **Count** | **%** |
| **0.0** | 19628491 | 50.18 |
| **.1** | 1494248 | 3.82 |
| **.2** | 919689 | 2.35 |
| **.3** | 818866 | 2.09 |
| **.4** | 914504 | 2.34 |
| **.5** | 1179943 | 3.02 |
| **.6** | 1640345 | 4.19 |
| **.7** | 2194447 | 5.61 |
| **.8** | 2721818 | 6.96 |
| **.9** | 7534658 | 19.26 |
| **1.0** | 70096 | 0.18 |
| **Total** | **39117105** | **100.00** |

**Supplementary Table S13**

Frequencies of alternate allele frequencies, as designated by the reference panel, across the imputed genome from the Affymetrix v6 array.

|  |  |  |
| --- | --- | --- |
| **Alternate Allele Frequency** | **Count** | **%** |
| **0.0** | 35608646 | 88.14 |
| **.1** | 1225065 | 3.03 |
| **.2** | 812632 | 2.01 |
| **.3** | 619136 | 1.53 |
| **.4** | 497194 | 1.23 |
| **.5** | 406225 | 1.01 |
| **.6** | 350468 | 0.87 |
| **.7** | 292833 | 0.73 |
| **.8** | 245699 | 0.61 |
| **.9** | 342518 | 0.85 |
| **1.0** | 1189 | 0.00 |
| **Total** | **40401605** | **100.00** |

**Supplementary Table S14**

Frequencies of genome-wide imputation scores for the Affymetrix v6 array.

|  |  |  |
| --- | --- | --- |
| **R2** | **Count** | **%** |
| **0.0** | 16706440 | 41.32 |
| **.1** | 1710923 | 4.24 |
| **.2** | 1092506 | 2.70 |
| **.3** | 970571 | 2.40 |
| **.4** | 1064146 | 2.63 |
| **.5** | 1389556 | 3.44 |
| **.6** | 1966246 | 4.87 |
| **.7** | 2643136 | 6.54 |
| **.8** | 3155044 | 7.81 |
| **.9** | 9565112 | 23.68 |
| **1.0** | 137925 | 0.34 |
| **Total** | **40401605** | **100.00** |

**References**

Rayner, W. (2016). *Post-Imputation Checking*. McCarthy Group Tools. Retrieved 15 April 2022, from https://www.well.ox.ac.uk/%7Ewrayner/tools/Post-Imputation.html

**Supplementary Information C**

**Supplementary Figure S1**

Teacher ratings of reading skill of the study participant at age 7 in the full sample (A, n = 14,989 and subsample (B, n = 5,829).



**Supplementary Figure S2**

Level reached in a progressive reading scheme by the participants at age 7 in the full sample (A, n = 14,785) and subsample (B, n = 5,780).



*Note*: Levels are listed in order of difficulty, from least to most difficult.

**Supplementary Figure S3**

Teacher rating of participant's use of books at age 11 in the full sample (A, n = 14,097) and subsample (B, n = 5,578).

**

**Supplementary Figure S4**

Self-reported reading frequency of book reading of participants at age 11 in the full sample (A, n = 13,603) and subsample (B, n = 5,450)

******

**Supplementary Figure S5**

Self-reported reading frequency of newspapers, magazines or comics of participants at age 11in the full sample (A, n = 13,542) and subsample (B, n = 5,438)

****

**Supplementary Figure S6**

Self-reported reading frequency of book reading of participants at age 16 in the full sample (A, n = 10,974) and subsample (B, n = 4,619)



**Supplementary Figure S7**

Proportion of participants who can read well enough to cope, according to their teacher, at age 16 in the full sample (A, n = 12,412) and subsample (B, n = 5,087)

**

**Supplementary Figure S8**

Teacher-rated ability in the subject of English for students aged 16 in the full sample (A, n = 12,233) and the subsample (B, n = 5,029)



**Supplementary Figure S9**

Self-rated participant ability in the subject of English at age 16 compared to other people their age in the full sample (A, n = 11,873) and subsample (B, n = 4,973)



**Supplementary Figure S10**

Proportion of participants who have struggled with reading problems since finishing school in the full sample (A, n = 12,496 & C, n = 11,357) and subsample (B, n = 5,574, & D, n = 5,721).



**Supplementary Table S15**

Response rates and total sample size captured for each reading variable in the full NCDS sample, full sample size from Power & Elliot, 2006.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Response Rate (%)** | **Sample Size** |
| *Age 7* | 15425 |
| Southgate Test | 96.78 |  |
| Teacher Rating | 97.17 |  |
| Book Level | 96.81 |  |
| *Age 11* | 15337 |
| Comprehension Test | 89.70 |  |
| Book Use | 91.91 |  |
| Reads Books | 88.69 |  |
| Reads Other | 88.30 |  |
| *Age 16* | 14647 |
| Comprehension Test | 81.07 |  |
| Reads Books | 77.57 |  |
| Can Cope | 84.88 |  |
| English Ability | 83.74 |  |
| English Rating | 81.48 |  |
| *Age 23* | 12537 |
| Reading Problems | 99.89 |  |
| *Age 33* | 11407 |
| Reading Problems | 99.56 |  |

**References**

Power, C., & Elliott, J. (2005). Cohort profile: 1958 British birth cohort (National Child Development Study). *International Journal of Epidemiology*, *35*(1), 34–41. https://doi.org/10.1093/ije/dyi183

**Supplementary Information D**

**Supplementary Table S16**

Number of unique samples available on each array after ordering by array quality and removing duplicates.

|  |  |
| --- | --- |
| **Array** | **N** |
| **Illumina 1.2m** | 2774 |
| **Infinium HumanHap 550K v3** | 2553 |
| **Affymetrix v6** | 179 |
| **Infinium HumanHap 550K v1.1** | 64 |
| **Illumina Human 660-Quad** | 840 |
| **Affymetrix 500k**  | 5 |
| **Illumina 15K Custom Chip** | 16 |

**References**

Al‐Soufi, L., Martorell, L., Moltó, M., González‐Peñas, J., García‐Portilla, M. P., Arrojo, M., Rivero, O., Gutiérrez‐Zotes, A., Nácher, J., Muntané, G., Paz, E., Páramo, M., Bobes, J., Arango, C., Sanjuan, J., Vilella, E., & Costas, J. (2021). A polygenic approach to the association between smoking and schizophrenia. Addiction Biology, 27(1). https://doi.org/10.1111/adb.13104