**PRACTICAL Meta-analysis summary data files**

**This document provides information about:**

**A/ Description of the imputation**

**B/ Description of the summary statistics (overall analysis)**

**C/ Acknowledgements**

1. **Imputation**

Genotypes for ~70M SNPs were imputed for all samples using the October 2014 (Phase 3) release of the 1000 Genomes Project data as the reference panel. The OncoArray and GWAS datasets were imputed using a two-stage imputation approach, using SHAPEIT20 for phasing and IMPUTEv221 for imputation. The imputation was performed in 5Mb non-overlapping intervals. All subjects were split into subsets of ~10,000 samples, with subjects from the same grouped in the subset. We imputed genotypes for all SNPs that were polymorphic (MAF>0.1%) in either European or Asian samples. We excluded data for all monomorphic SNPs and those with an imputation r2<0.3 leaving a total of 21,465,239 SNP across chromosomes 1-22 and chromosome X.

1. **Statistical analysis**

Per-allele odds ratios and standard errors were generated for the OncoArray and each GWAS, adjusting for principal components and study relevant covariates using logistic regression. The OncoArray and iCOGS analyses were additionally stratified by country and study, respectively. We used the first seven principal components in European and first four principal components for Asian, as additional components did not further reduce inflation in the test statistics.

Odds ratio (OR) estimates were derived using either SNPTEST (https://mathgen.stats.ox.ac.uk/genetics\_software/snptest/snptest.html) or an in-house C++ program. OR estimates and standard errors were combined by a fixed effects inverse variance meta-analysis using METAL22. All statistical tests conducted were two-sided.

For the **overall analysis** no other covariates than the PCs were included in the model. A fixed effect model was used. The data was analysed stratified by country. Those countries with less than 50 samples (cases or controls) were excluded from analysis. The country variable is based on the country of each Study in each GWAS. The samples included in the analysis are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | PrCa Cases | Non-PrCa Controls | Total |
| OncoArray European |  | 44825 | 27904 | 72729 |
| iCOGS |  | 20219 | 20440 | 40659 |
| UK1 |  | 1854 | 1894 | 3748 |
| UK2 |  | 3650 | 3940 | 7590 |
| CAPS1 |  | 474 | 482 | 956 |
| CAPS2 |  | 1458 | 512 | 1970 |
| BPC3 |  | 2068 | 2993 | 5061 |
| Pegasus |  | 4600 | 2941 | 7541 |
| **Total** |  | **79148** | **61106** | **140254** |

Please refer to the **guidelines for PRACTICAL OncoArray QC document** (at <http://practical.icr.ac.uk/blog/?page_id=6297>) for a description of the imputation in the OncoArray set.

The **fields** included in the summary statistics file are:

|  |  |
| --- | --- |
| **Variable** | **Description** |
| MarkerName | Unique identifier for variant (Composed of chr\_position\_build37\_a0\_a1) |
| rs\_id |  |
| SNP | Name the variant was imputed as – either a 1000 Genomes identifier or the Oncoarray SNP Name. |
| Chr | Chromosome number |
| Position | Position (B37) |
| Allele1 | the first allele for this marker in the first file where it occurs |
| Allele2 | the second allele for this marker in the first file where it occurs |
| Freq1 | weighted average of frequency for allele 1 across all studies (for controls) |
| FreqSE | corresponding standard error for allele frequency estimate |
| MinFreq | minimum frequency for allele 1 across all studies |
| MaxFreq | maximum frequency for allele 1 across all studies |
| Effect | overall estimated effect size for allele1 \* |
| StdErr | overall standard error for effect size estimate |
| P-value | meta-analysis p-value |
| Direction | summary of effect direction for each study, with one '+' or '-' per study |
| OncoArray\_imputation\_r2 | QC score for the variant on OncoArray |

\* The effect in meta-analysis files are “beta”. To get OR = exp(beta).

The METAL program has been used for the meta-analysis..

1. **Acknowledgements**

When using these data, please cite: “Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci”. Schumacher, Al Olama, Berndt, et al. Nature Genetics 2018, acknowledge "The PRACTICAL consortium, CRUK, BPC3, CAPS, PEGASUS", and include the acknowledgement detailed at <http://practical.icr.ac.uk/blog/?page_id=8164>.

Kindly send a copy of any publication that results from these data to PRACTICAL@icr.ac.uk