

## **Additional file 2: Supplementary methods**

### **Optimizing colorectal cancer screening through polygenic risk score-based risk stratification: evidence from a population-based cohort and screening trial**

**Authors:** Hongda Chen, Jiahui Si, Chenyu Luo, Jianbo Tian, Junyi Xin, Canqing Yu, Dianjianyi Sun, Pei Pei, Ling Yang, Iona Y Millwood, Robin G Walters, Yiping Chen, Huaidong Du, Zhengming Chen, Liming Li, Masaru Koido, Yoichiro Kamatani, Meilin Wang, Xiaoping Miao, Jun Lv, Min Dai

## Supplementary methods

### Study population

#### *Population-based screening trial for PRS validation and application*

The TARGET-C study is a large-scale, multicenter CRC screening randomized controlled trial (RCT) conducted in 6 cities of 5 provinces in China, which aiming to compare the feasibility, participation, yield and cost of colonoscopy, fecal immunochemical test (FIT), and risk-adapted strategies for CRC screening. This study was approved by the Ethics Committee of the National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, and registered in the Chinese Clinical Trial Registry (registration number: ChiCTR1800015506). All participants provided written informed consent.

In the present study, we included participants who underwent colonoscopy at the baseline screening phase and provided qualified plasma samples (N=3173). Eligible subjects were genotyped using the Infinium™ Chinese Genotyping Array Bead Chip. 1000 Genomes Project Phase 3 data was used as reference for imputation. We first pre-phased the genotypes with SHAPEIT2 and then performed imputation using IMPUTE2<sup>1</sup>. As for quality control (QC), SNPs were excluded based on the SNPs (i) on sex chromosomes, (ii) with call rates < 95%, (iii) with *P* value for Hardy Weinberg equilibrium (HWE) <  $1.0 \times 10^{-6}$ , (iv) with minor allele frequency < 0.01. For QC of samples, we excluded those (i) with sample call rates < 95%, (ii) deviate  $\pm 3$  standard

deviation (SD) from the samples' heterozygosity rate mean value, (iii) with relatedness above a  $\pi$ -hat threshold of 0.25, (iv)  $\pm 4$  SD outliers identified by principal component analysis. Finally, a total of 7,895,576 SNPs and 2821 samples were included.

### **Ascertainment of outcomes**

In the TARGET-C trial, all outcomes were classified based on the most advanced findings from colonoscopy and confirmed by histopathological review of biopsies. The colonoscopic findings were defined and categorized as (i) CRC (adenocarcinoma of the colon or rectum), (ii) Advanced adenoma (AA) (adenomas with high-grade dysplasia, villous or tubular-villous histologic features, or diameter  $\geq 10$  mm, or traditional or sessile serrated lesions with diameter  $\geq 10$  mm or dysplasia), (iii) Non-advanced adenoma (NAA) (diameter  $< 10$  mm with no advanced histology), and (iv) Health control (HC) (no findings of adenomas)<sup>2,3</sup>. Both CRC and AA were regarded as advanced neoplasia (AN)<sup>2</sup>.

### **Construction methods of PRS**

#### ***Approach 1: Well-established PRS based on GWAS-identified CRC risk SNPs***

To date, limited studies have thoroughly developed and validated PRSs for CRC risk prediction in East Asian populations. A previous study used GWAS data from the Asia Colorectal Cancer Consortium (ACCC), including 24,192 CRC cases and 214,186

controls, to develop and validate performance of CRC PRSs<sup>4</sup>. PRS<sub>115-EAS</sub>, a PRS of 115 GWAS-identified risk variants consistently associated with CRC risk across East Asian and European populations, with weights derived from East Asian data, validated significantly better than PRS<sub>115-EUR</sub> derived from European descendants or PRSs constructed using results from fine-mapping and genome-wide algorithms<sup>4</sup>. Xin et al systematically constructed candidate PRSs via different approaches by incorporating EAS and European (EUR) GWAS meta-analysis<sup>5</sup>. The PRS-CSx approach-derived PRS (defined as PRS<sub>CSx-EAS\_EUR</sub>) achieved the optimal performance. Therefore, we used PRS<sub>115-EAS</sub> and PRS<sub>CSx-EAS\_EUR</sub> as well-established EAS PRS for subsequent assessment. We calculated PRS based on the following formula:  $PRS = \sum_{i=1}^n \beta_i SNP_i$ , where n means the number of SNPs,  $SNP_i$  means the number of the risk allele for the i-th SNP (0, 1, or 2), and  $\beta_i$  means effect size carried by the risk allele.

### **Calculation of modified Asia-Pacific Colorectal Screening (APCS) score in TARGET-C**

A standardized epidemiological questionnaire was administered by trained interviewers to all TARGET-C participants. Information including sociodemographic factors, lifestyles, disease history and family history was collected. We used a modified Asia-Pacific Colorectal Screening (APCS) score, which enables risk stratification using 5 elementary clinical information, including age (0: 50–54 years; 1:

55–64 years; 2: 65–74 years), sex (0: female; 1: male), family history of CRC among first-degree relatives (0: absent; 1: present), smoking status (0: non-smoker; 1: current or past smoker), and body mass index (BMI) (0:  $<23 \text{ kg/m}^2$ ; 1:  $\geq 23 \text{ kg/m}^2$ )<sup>6</sup>.

Participants with a score  $\geq 4$  were defined as high risk and those with a score  $< 4$  were defined as low risk.

### **Risk-adapted screening strategy in the TARGET-C trial**

The risk-adapted screening strategy was defined as follows: participants identified as high-risk were recommended to undergo colonoscopy, while those classified as low-risk were offered FIT. Individuals with positive FIT results were subsequently referred for diagnostic colonoscopy.

All colonoscopies were performed by experienced endoscopists following standard procedures. Abnormal findings during colonoscopy were carefully checked and sent for pathology examination under standard clinical procedures. Any findings during colonoscopy were required to be documented photographically. Clinical information such as the examination duration, sedation status, completeness of colonoscopy, bowel preparation status, complications, polyp features, description of other abnormal findings, as well as pathology diagnosis were collected and documented in the web-based data management system. To ensure the consistency of pathology diagnosis among different hospitals, we conducted a central review of the pathology

sections for all CRCs, all advanced adenomas, and a random selection of 10% of the nonadvanced adenomas by an experienced gastrointestinal pathologist from the National Cancer Center of China. Any inconsistent diagnosis was resolved by consensus of at least 2 central pathologists. The FIT product used in our study was the FIT OC-SENSOR (Eiken Chemical Co, Tokyo, Japan). Following the manufacturer's recommendation, a cut-off value of  $\geq 100$  ng Hb/mL was used to define a positive result.

## References

1. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
2. Gupta S, Lieberman D, Anderson JC, et al. Recommendations for Follow-Up After Colonoscopy and Polypectomy: A Consensus Update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2020;158:1131-1153.e5.
3. Rex DK, Ahnen DJ, Baron JA, et al. Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am J Gastroenterol* 2012;107:1315–1329; quiz 1314, 1330.
4. Ping J, Yang Y, Wen W, et al. Developing and validating polygenic risk scores for colorectal cancer risk prediction in East Asians. *Intl Journal of Cancer* 2022;151:1726–1736.
5. Xin J, Du M, Gu D, et al. Risk assessment for colorectal cancer via polygenic risk score and lifestyle exposure: a large-scale association study of East Asian and European populations. *Genome Med* 2023;15:4.
6. Sung JJY, Wong MCS, Lam TYT, et al. A modified colorectal screening score for prediction of advanced neoplasia: A prospective study of 5744 subjects. *J Gastroenterol Hepatol* 2018;33:187–194.