SUPPLEMENTARY METHODS

Assessment of healthy diet

Dietary intake was assessed using a validated food frequency questionnaire (FFQ), which included 100 foods and beverages items with specified fitted portion sizes. For each item, participants were asked to select their intake frequency (seven categories ranging from 'almost never' to 'twice or more per day' for foods and eight categories ranging from 'almost never' to 'four or more times per day' for beverages) in the last month. Using data from repeated measurements of the FFQ approximately 3 months apart and data from a 4-day weighed food record (WFR) in a randomly selected sample of 150 participants from our cohort, the reproducibility and validity of the FFQ were assessed. In summary, for reproducibility, the Spearman rank correlation coefficient for energy intake between the two measurements of FFQs was 0.68. The correlation coefficients for food and beverage items intake between the two measurements of the FFQ ranged from 0.62 to 0.79. For validity, the Spearman rank correlation coefficient for energy intake between WFR and the FFQ was 0.49. Correlation coefficients for nutrient intake between WFR and the FFQ ranged from 0.35 to 0.54 and 0.39 to 0.72 before and after adjustment of energy intake, respectively (all $P<0.05$).

Intake of food and beverage items was calculated by multiplying fitted portion sizes (according to sex, gram/time) by the frequency at which each item was consumed per day. The intake of each nutrient was calculated using the Chinese food composition tables [1] as the nutrient database, by first multiplying the quantity (grams) of each food and beverage item consumed by their nutrient content per gram; and then adding the nutrient contributions across all food items. Dietary patterns were derived using factor analysis by including the intake data of all 100 food and beverage items in grams. For greater interpretability of the dietary patterns, the varimax rotation procedure was applied. Major dietary patterns were identified with evaluation of eigenvalues (>1) and the scree test. Finally, three factors were determined and named descriptively according to the food and beverage items showing high factor loadings (absolute value >0.3). Three major dietary patterns (explaining 35.97% of the variance in food intake) were revealed using factor analysis and are presented in Supplementary Table 1. First, the fruit and sweet foods pattern, characterized by high intake of fruits, cakes, and ice cream; second, the vegetable pattern, characterized by high intake of vegetables, egg, soya bean products, and coarse cereals; third, the animal foods pattern, characterized by high intake of animal organs, animal blood, meat, processed meat products, and preserved egg.

Genotyping and quality control

In our cohort, the Illumina Asian Screening Array was used to genotype genomic DNA samples isolated from peripheral blood leukocytes in a randomly selected sample. Eventually, we identified the genotypes of 743,722 single nucleotide polymorphisms (SNPs) for 3,778 participants. After stringent quality control filtering [2], we excluded: (1) individual call rate <98%; (2) SNP genotype call rates with <98% completeness; (3) minor allele frequency <1%; (4) Hardy-Weinberg equilibrium tests with value of $P<1\times10^{-6}$; (5) a mismatch between genetic and reported sex; (6) duplication of genetic data; (7) extremely low or high heterozygosity (>3 standard deviations from the mean); and (8) SNPs in sex chromosomes or mitochondrial DNA. A total of 468,600 SNPs for 3,675 participants with an overall call rate of 99.8% were included. Further, after the exclusions in the present cohort study, 2,467 participants with genotype information were included in the analyses to study the effect modification of genetic risk on the associations between healthy lifestyle and the risk of metabolic dysfunction-associated fatty liver disease.

Genotype imputation was performed to capture information on unobserved SNPs and sporadically missing genotypes among the genotyped SNPs, using all haplotypes from the 1,000 Genomes Project Phase 3 reference panel (v2.3.2) (1,000 Genomes haplotypes; phase 3, October 2014). Pre-phasing was performed in SHAPEIT v2.12, and imputation was performed using IMPUTE2 [3]. Imputed SNPs with INFO scores <0.3 were excluded.

SUPPLEMENTARY REFERENCES

1. Yang Y, Wang G, Pan X. China food composition. 2nd ed. Beijing: Peking University Medical Press; 2009.
2. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc 2010;5:1564-73.
3. van Leeuwen EM, Kanterakis A, Deelen P, Kattenberg MV; Genome of the Netherlands Consortium; Slagboom PE, et al. Population-specific genotype imputations using minimac or IMPUTE2. Nat Protoc 2015;10:1285-96.