**Supplemental Table 1. Cohorts included for gene-outcome analysis\***

|  |  |  |
| --- | --- | --- |
|  | CHD | Stroke |
| EPIC-CVD | Included without EPIC-Norfolk |   |
| EPIC-NL  | Included | Included |
| UK biobank  | Included | Included  |
| MEGASTROKE |   | Included  |
| CARDIoGRAM GWAS | Included |   |

\* In case of participant overlap between the EPIC studies and consortium data, we excluded participants from the EPIC studies

**Supplemental Table 2. Description of baseline characteristics of 13,928 participants in the EPIC-CVD subcohort.**

|  |  |  |  |
| --- | --- | --- | --- |
|   | Total Subcohort |   | Missing\* (%) |
| N | 13114 |  |  |
| Age at recruitment | 52.3 ± 9.2 |  | 0.0 |
| Systolic blood pressure (mmHg) | 132.8 ± 19.5 |  | 24.2 |
| Diastolic blood pressure (mmHg) | 81.8 ± 10.8 |  | 24.2 |
| Body mass index (kg/m2) | 26.1 ± 4.2 |  | 0.6 |
| Waist to hip ratio  | 0.9 ± 0.1 |  | 8.1 |
| Total cholesterol (mmol/L) | 5.9 ± 1.1 |  | 4.5 |
| HDL cholesterol (mmol/L) | 1.5 ± 0.4 |  | 4.5 |
| Ratio of total to HDL cholesterol | 4.3 ± 1.6 |  | 4.5 |
| Non HDL cholesterol (mmol/L) | 4.5 ± 1.2 |  | 4.5 |
| Triglycerides (mmol/L) | 1.1 [0.8, 1.7] |  | 4.5 |
| Female sex (%)  | 60.7 |  | 0.0 |
| History of diabetes (%) | 3.1 |  | 8.7 |
| History of hypertension (%) | 37.1 |  | 0.9 |
| History of hypercholesterolemia (%) | 21.2 |  | 33.1 |
| Highly educated (%) | 44.6 |  | 1.5 |
| Physically active (%) | 20.8 |  | 1.3 |
| Never smoker (%) | 46.2 |   | 0.9 |

\* Expressed as percentage of total subcohort. Data are expressed as mean ± SD or median [p25,p75]

**Supplemental Table 3. Genotype frequencies\* and HWE test per country in the EPIC-CVD subcohort**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|   | CC | CT | TT |  | Total n | HWE p value |
| Italy | 69.9 | 26.6 | 3.5 |  | 1,596 | 0.06 |
| Spain | 29.8 | 48.6 | 21.6 |  | 2,860 | 0.27 |
| United Kingdom | 9.7 | 42.0 | 48.3 |  | 1,121 | 0.70 |
| The Netherlands | 11.6 | 42.8 | 45.6 |  | 1,206 | 0.28 |
| Germany | 19.3 | 49.6 | 31.1 |  | 1,821 | 0.84 |
| Sweden | 7.2 | 35.4 | 57.5 |  | 2,635 | 7.0\*10-3 |
| Denmark | 4.9 | 45.2 | 49.9 |  | 1,875 | 6.6\*10-9 |
| Total | 21,7 | 41,8 | 36,5 |  | 13,114 | 9.5\*10-62 |

HWE = Hardy Weinberg Equilibrium \* Expressed as percentage of total

**Supplemental Table 4. Description of baseline characteristics of 2,025 participants in the EPIC-NL subcohort.**

|  |  |  |  |
| --- | --- | --- | --- |
|   |  Total Subcohort |   | Missing\* (%) |
| N | 2,025 |  |  |
| Age at recruitment | 51 ± 11  |  | 0 |
| Systolic blood pressure (mmHg) | 127.0 ± 18.9  |  | 0.3 |
| Diastolic blood pressure (mmHg) | 78.1 ± 10.5  |  | 0.2 |
| Body mass index (kg/m2) | 25.6 ± 4.0  |  | 0 |
| Waist to hip ratio  | 0.8 ± 0.1  |  | 0.1 |
| Total cholesterol (mmol/L) | 5.6 ± 1.0  |  | 0 |
| HDL cholesterol (mmol/L) | 1.5 ± 0.4  |  | 0 |
| Ratio of total to HDL cholesterol | 4.0 ± 1.3  |  | 0 |
| Non HDL cholesterol (mmol/L) | 4.1 ± 1.1 |  | 0 |
| Female sex (%) | 80.1 |  | 0 |
| History of diabetes (%) | 1.5 |  | 0.1 |
| History of hypertension (%) | 38.4 |  | 0 |
| History of hypercholesterolemia (%) | 7.7 |  | 11.3 |
| Highly educated (%) | 21.5 |  | 0 |
| Physically active (%) | 42.1 |  | 0 |
| Never smoker (%) | 39.4 |  | 0 |
| Postmenopausal† (%) | 54.2 |   | 0 |

\* Expressed as percentage of total subcohort. † Among women. Data are expressed as mean ± SD

**Supplemental Table 5. Milk intake in the EPIC-CVD subcohort, stratified by country.**

|  |  |  |  |
| --- | --- | --- | --- |
| Country | N | Milk intake (g/day) | Non milk dairy intake (g/day)  |
| Italy | 1,596 | 120 [0, 192] | 81 [47, 138] |
| Spain | 2,860 | 200 [112, 307] | 49 [15, 105]  |
| United Kingdom | 1,121 | 294 [149, 440] | 67 [34, 112] |
| The Netherlands | 1,206 | 202 [82, 381] | 170 [105, 231]  |
| Germany | 1,821 | 32 [2, 97] | 138 [82, 222]  |
| Sweden | 2,635 | 196 [79, 402] | 123 [60, 228]  |
| Denmark | 1,875 | 182 [29, 509] | 99 [51, 212]  |

Data are expressed as median [p25, p75]

**Supplemental Table 6. Dietary intake of 13,928 participants in EPIC-CVD by rs4988235 genotype.**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Stratified by rs4988235 genotype |  |
|   |   | C/C | C/T | T/T |   |
| N |  | 2104 [1702, 2628] | 2036 [1645, 2509] | 2038 [1654, 2497] |  |
| Total energy (kcal/day)  |  | 143 [17, 224] | 170 [47, 307] | 195 [54, 387] |  |
| Milk (g/day) |  | 80 [39, 144] | 98 [45, 179] | 106 [53, 195] |  |
| Non milk dairy (g/day) |  | 0 [0, 0] | 0 [0, 3] | 0 [0, 3] |  |
|  Milk for coffee and creamers |  | 0 [0, 0] | 0 [0, 0] | 0 [0, 0] |  |
|  Dairy creams  |  | 0 [0, 2] | 1 [0, 3] | 2 [0, 4] |  |
|  Milk based puddings |  | 0 [0, 2] | 1 [0, 11] | 3 [0, 14] |  |
|  Curd  |  | 0 [0, 0] | 0 [0, 5] | 0 [0, 5] |  |
|  Yogurt, thick fermented milk  |  | 12 [0, 56] | 27 [0, 98] | 36 [2, 109] |  |
|  Ice cream  |  | 4 [0, 12] | 3 [0, 7] | 3 [1, 8] |  |
|  Cheese  |  | 36 [17, 64] | 26 [13, 48] | 25 [14, 44] |  |
| Vegetables (g/day)  |  | 160 [105, 246] | 157 [100, 239] | 142 [90, 218] |  |
| Fruit (g/day)  |  | 249 [139, 382] | 182 [99, 298] | 158 [88, 268] |  |
| Meat and meat products (g/day)  |  | 102 [71, 145] | 104 [69, 145] | 101 [68, 139] |  |
| Fish and shellfish (g/day) |  | 29 [16, 52] | 29 [15, 52] | 28 [13, 51] |  |
| Soft drinks (g/day) |  | 22 [17, 28] | 22 [17, 27] | 21 [17, 27] |  |
| Coffee (g/day)  |  | 0 [0, 44] | 4 [0, 77] | 16 [0, 97] |  |
| Tea (g/day)  |  | 120 [55, 261] | 293 [97, 580] | 450 [190, 700] |  |
| Alcohol intake (g/day)  |   | 0 [0, 52] | 3 [0, 190] | 12 [0, 250] |   |

\* Expressed as percentage of total subcohort. Data are expressed as median [p25, p75]

**Supplemental Table 7. Dietary intake of 2,025 participants in the EPIC-NL subcohort by best guess rs4988235 genotype.**

|  |  |  |
| --- | --- | --- |
|  |  | Stratified by best guess rs4988235 genotype |
|   |   | C/C | C/T | T/T |
| N |  | 196 | 909 | 920 |
| Total energy intake (kcal/day) |  | 1937 [1620, 2376] | 1928 [1652, 2325] | 1888 [1594, 2271] |
| Milk (g/day) |  | 193 [49, 292] | 215 [65, 407] | 226 [96, 419] |
|  Unfermented, unsweetened milk |  | 75 [20, 192] | 108 [25, 241] | 129 [30, 265] |
|  Buttermilk  |  | 0 [0, 34] | 3 [0, 86] | 2 [0, 121] |
|  Sweetened milk  |  | 9 [0, 26] | 13 [1, 31] | 15 [1, 31] |
| Non-milk dairy (g/day)  |  | 159 [92, 213] | 158 [90, 212] | 157 [98, 212] |
|  Curd  |  | 6 [2, 11] | 6 [2, 11] | 7 [2, 14] |
|  Yogurt  |  | 46 [14, 102] | 50 [13, 98] | 52 [17, 94] |
|  Cheese |  | 29 [18, 48] | 31 [20, 48] | 30 [20, 47] |
| Vegetables (g/day)  |  | 135 [102, 166] | 133 [105, 169] | 133 [104, 169] |
| Fruit (g/day)  |  | 240 [139, 359] | 248 [142, 364] | 251 [149, 368] |
| Meat (g/day)  |  | 114 [74, 147] | 104 [66, 136] | 100 [59, 133] |
| Fish (g/day) |  | 6 [2, 14] | 8 [3, 15] | 8 [3, 15] |
| Beverages (g/day)  |  | 1476 [1204, 1932] | 1472 [1161, 1843] | 1432 [1149, 1796] |
| Alcohol (g/day)  |  | 5 [0, 15] | 7 [1, 18] | 5 [1, 16] |

Data are expressed as median [p25, p75]

**Supplemental Table 8. Association between lactase persistence genotype and cardiovascular risk factors among the EPIC-CVD subcohort participants.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | estimate\*  | 95%CI | p-value†  | I2 (%)‡ | n |
| Systolic blood pressure (mmHg) |  0.35 |  -0.22 |  0.92 |  0.23 |  79.1 | 9,943 |
| Diastolic blood pressure (mmHg) |  0.11 |  -0.22 |  0.45 |  0.51 |  66.4 | 9,941 |
| Body mass index (kg/m2) |  0.15 |  0.04 |  0.27 |  0.01 |  7.8 | 13,035 |
| Waist to hip ratio  |  2.4\*10-3 |  7.4\*10-4 |  4.5\*10-3 |  0.01 |  0.0 | 12,053 |
| Total cholesterol (mmol/L) |  -0.03 |  -0.06 |  0.01 |  0.11 |  0.0 | 12,526 |
| HDL cholesterol (mmol/L) |  -0.02 |  -0.03 |  -0.01 |  3.0\*10-3 |  0.0 | 12,529 |
| Ratio of total to HDL cholesterol |  0.04 |  -0.01 |  0.08 |  0.09 |  0.0 | 12,526 |
| Non HDL cholesterol (mmol/L) |  -0.01 |  -0.04 |  0.02 |  0.59 |  0.0 | 12,526 |
| Triglycerides (log(mmol/L)) |  0.01 |  -0.01 |  0.02 |  0.23 |  18.0 | 12,527 |
| History of diabetes |  1.03 |  0.86 |  1.22 |  0.78 |  50.8 | 11,978 |
| History of hypertension |  1.05 |  0.99 |  1.12 |  0.13 |  0.0 | 12,999 |
| History of hypercholesterolemia |  0.90 |  0.83 |  0.99 |  0.02 |  0.0 | 8,773 |
| Highly educated  |  0.93 |  0.88 |  1.00 |  0.03 |  0.0 | 12,915 |
| Physically active |  1.03 |  0.96 |  1.11 |  0.40 |  0.0 | 12,948 |
| Never smoker  |  0.99 |  0.93 |  1.05 |  0.76 |  61.2 | 13,002 |

\* Estimate is OR for dichotomous and β for continuous variables, derived from linear or logistic regression model adjusting for sex, age, two genetic principal components, and study centre. † p-value of regression model. ‡ I2 for fixed effect meta-analysis of analyses stratified by participants with hard call and imputed rs4988235 data.

**Supplemental Table 9. Association between lactase persistence genotype and cardiovascular risk factors among the EPIC-NL subcohort.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|   | estimate\* | 95%CI | p-value | N |
| Systolic blood pressure (mmHg) | -0.04 | -1.23 | 1.15 | 0.95 | 2,018 |
| Diastolic blood pressure (mmHg) | -0.1 | -0.81 | 0.60 | 0.77 | 2,020 |
| Body mass index (kg/m2) | -0.06 | -0.32 | 0.21 | 0.69 | 2,025 |
| Waist to hip ratio  | 1.7\*10-3 | -0.01 | 2.9\*10-3 | 0.48 | 2,023 |
| Total cholesterol (mmol/L) | -0.01 | -0.08 | 0.05 | 0.72 | 2,024 |
| HDL cholesterol (mmol/L) | 0.02 | -0.01 | 0.04 | 0.25 | 2,024 |
| Ratio of total to HDL cholesterol | -0.04 | -0.13 | 0.05 | 0.37 | 2,024 |
| Non HDL cholesterol (mmol/L) | -0.03 | -0.1 | 0.04 | 0.43 | 2,024 |
| History of diabetes  | 1.06 | 0.61 | 1.94 | 0.83 | 2,022 |
| History of hypertension  | 1.00 | 0.86 | 1.16 | 0.96 | 2,025 |
| History of hypercholesterolemia | 0.83 | 0.63 | 1.11 | 0.20 | 1,797 |
| Highly educated  | 1.11 | 0.93 | 1.32 | 0.25 | 2,015 |
| Physically active  | 1.01 | 0.88 | 1.17 | 0.86 | 2,025 |
| Never smoker  | 1.19 | 1.03 | 1.37 | 0.02 | 2,022 |

\* Estimate is OR for dichotomous and β for continuous variables. Estimate derived from linear or logistic regression model adjusting for sex, age, two genetic principal components, and study centre.



**Supplemental Figure 1. Hazard ratio and 95%CI for each 25 g/day increase in genetically predicted milk intake and risk of ischaemic stroke in EPIC-CVD and EPIC-NL, assuming an additive effect of rs4988235.**

Estimates derived from Prentice-weighted Cox regression investigating effect of genetically predicted milk consumption (per 25g/day) on ischaemic stroke risk. Age was used as underlying timescale and estimates were adjusted for sex, the first two genetic principal components and study centre. In EPIC-CVD, the analysis was performed per country. Within countries, the analysis was stratified by participants with hard call versus imputed rs4988235 data and pooled using fixed effects. Country-specific results from EPIC-CVD and the additional EPIC-NL results were pooled with inverse variance weights in a random effect meta-analysis using restricted maximum likelihood estimation. Genetically predicted milk consumption was obtained from a linear regression model regressing rs4988235 on milk consumption in EPIC-CVD ad EPIC-NL separately, adjusted for age, sex, two genetic principal components and study centre.



**Supplemental Figure 2. Hazard ratio and 95%CI for each 25 g/day increase in genetically predicted milk intake and risk of total stroke in EPIC-CVD and EPIC-NL, assuming a dominant effect of rs4988235.**

Estimates derived from Prentice-weighted Cox regression investigating effect of genetically predicted milk consumption (per 25g/day) on CHD risk, assuming a dominant effect of rs4988235 T alleles. Age was used as underlying timescale and estimates were adjusted for sex, the first two genetic principal components and study centre. In EPIC-CVD, the analysis was performed per country. Within countries, the analysis was stratified by participants with hard call versus imputed rs4988235 data and pooled using fixed effects. Country-specific results from EPIC-CVD and the additional EPIC-NL results were pooled with inverse variance weights in a random effect meta-analysis using restricted maximum likelihood estimation. Genetically predicted milk consumption was obtained from a linear regression model regressing rs4988235 on milk consumption in EPIC-CVD ad EPIC-NL separately, adjusted for age, sex, two genetic principal components and study centre.



**Supplemental Figure 3. Hazard ratio and 95%CI for each 25 g/day increase in genetically predicted milk intake and risk of CHD in EPIC-CVD and EPIC-NL, assuming a dominant effect of rs4988235.**

Estimates derived from Prentice-weighted Cox regression investigating effect of genetically predicted milk consumption (per 25g/day) on CHD risk, assuming a dominant effect of rs4988235 T alleles. Age was used as underlying timescale and estimates were adjusted for sex, the first two genetic principal components and study centre. In EPIC-CVD, the analysis was performed per country. Within countries, the analysis was stratified by participants with hard call versus imputed rs4988235 data and pooled using fixed effects. Country-specific results from EPIC-CVD and the additional EPIC-NL results were pooled with inverse variance weights in a random effect meta-analysis using restricted maximum likelihood estimation. Genetically predicted milk consumption was obtained from a linear regression model regressing rs4988235 on milk consumption in EPIC-CVD ad EPIC-NL separately, adjusted for age, sex, two genetic principal components and study centre.