**Supplementary Material**

**A treelet transform analysis to relate nutrient patterns to the risk of hormonal receptor–defined breast cancer in the European Prospective Investigation into Cancer and Nutrition study.**

**Supplemental Table 1**: TT (cut-level 16) loadings of the third and fourth components.

|  |  |
| --- | --- |
| Variables \*  | TT 16 loadings |
| TC3 | TC4 |
| Calcium (Ca) |  |  |
| β-Carotene |  |  |
| Cholesterol | -0.178 |  0.448 |
| MUFA |  |  |
| PUFA |  |  |
| SFA |  |  |
| Iron (Fe) |  |  |
| Fibre |  |  |
| Potassium (K) |  |  |
| Magnesium (Mg) |  |  |
| Phosphorus (P) |  |  |
| Protein | -0.052 |  0.132 |
| Retinol | -0.410 | -0.609 |
| Riboflavin |  |  |
| Starch |  |  |
| Sugar |  |  |
| Thiamin |  |  |
| Vitamin B6 |  |  |
| Vitamin B12 | -0.254 | 0.641 |
| Vitamin C |  |   |
| Vitamin D | 0.856 |  |
| Vitamin E |  |  |
| Folate |  |  |
| Explained variance | 9% | 6% |

TC3, treelet component 3. TC4, treelet component 4.

\* log-transformed nutrient variables.

**Supplemental Table 2**: PCA loadings of the 4 derived components.

|  |  |
| --- | --- |
| Variables \*†  | PCA loadings |
| PC1 | PC2 | PC3 | PC4 |
| Calcium (Ca) | -0.024 | 0.12 | -0.136 | **0.314** |
| β-Carotene | **-0.275** | **0.601** | -0.121 | **-0.495** |
| Cholesterol | **0.276** | 0.07 | -0.172 | 0.064 |
| MUFA | 0.018 | -0.043 | -0.123 | -0.148 |
| PUFA | -0.006 | 0.102 | 0.131 | **-0.211** |
| SFA | 0.119 | -0.031 | -0.155 | -0.105 |
| Iron (Fe) | -0.054 | 0.102 | -0.019 | 0.048 |
| Fibre | -0.131 | 0.145 | 0.136 | 0.006 |
| Potassium (K) | -0.065 | 0.174 | 0.065 | 0.169 |
| Magnesium (Mg) | -0.045 | 0.142 | 0.042 | 0.115 |
| Phosphorus (P) | 0.003 | 0.108 | 0.01 | 0.19 |
| Protein | 0.042 | 0.077 | -0.003 | 0.159 |
| Retinol | **0.601** | **0.271** | **-0.295** | -0.275 |
| Riboflavin | 0.004 | **0.206** | -0.131 | **0.322** |
| Starch | -0.004 | -0.112 | 0.137 | -0.068 |
| Sugar | -0.098 | 0.073 | 0.01 | 0.175 |
| Thiamin | -0.076 | 0.174 | 0.133 | 0.183 |
| Vitamin B6 | -0.075 | 0.177 | 0.072 | 0.189 |
| Vitamin B12 | **0.362** | **0.254** | **-0.266** | **0.306** |
| Vitamin C | **-0.276** | **0.316** | -0.033 | 0.126 |
| Vitamin D | **0.431** | **0.25** | **0.796** | 0.006 |
| Vitamin E | -0.098 | 0.153 | 0.068 | **-0.256** |
| Folate | -0.141 | **0.249** | -0.014 | 0.105 |
| Explained variance | 28% | 22% | 10% | 8% |

PC1, principal component 1. PC2, principal component 2. PC3, principal component 3. PC4, principal component.

\* log-transformed nutrient variables
† In bold are PCA loadings >0.20

**Supplemental Table 3**: HRs (95%CI) for BC by quintiles of pattern scores (1st and 2nd components of TT cut-level 16) for PR positive and PR negative tumours in EPIC women.

|  |  |  |
| --- | --- | --- |
| Model\* | First component | Second component |
| PY | BC cases | HR (95% CI) | P-LRTa | P-trendb | PY | BC cases | HR (95% CI) | P-LRTa | P-trendb |
| PR Positive |  |  |  |  |  |  |  |  |  |  |
| Q1 | 723,730 | 611 | 1.00 (ref) | 0.31 | 0.28 | 738,063 | 801 | 1.00 (ref) | 0.17 | <0.01 |
| Q2 | 729,055 | 850 | 1.12 (1.01,1.25) | 727,815 | 823 | 0.96 (0.86,1.06) |
| Q3 | 726,226 | 805 | 1.10 (0.98,1.22) | 720,137 | 827 | 0.95 (0.85,1.05) |
| Q4 | 726,869 | 800 | 1.10 (0.98,1.23) | 716,542 | 766 | 0.90 (0.81,1.00) |
| Q5 | 717,755 | 812 | 1.10 (0.97,1.24) | 721,078 | 661 | 0.87 (0.77,0.98) |
| PR Negative |  |  |  |  |  |  |  |  |  |  |
| Q1 | 722,296 | 386 | 1.00 (ref) | 0.46 | 0.10 | 735,796 | 467 | 1.00 (ref) | 0.10 | 0.03 |
| Q2 | 726,449 | 468 | 0.98 (0.86,1.13) | 725,303 | 449 | 0.89 (0.78,1.02) |
| Q3 | 723,483 | 433 | 0.91 (0.79,1.06) | 717,455 | 434 | 0.84 (0.73,0.96) |
| Q4 | 724,668 | 468 | 0.99 (0.85,1.15) | 714,395 | 454 | 0.90 (0.78,1.03) |
| Q5 | 715,243 | 435 | 0.90 (0.77,1.06) | 719,189 | 386 | 0.84 (0.72,0.98) |
| *P- heterogeneity*c |  |  |  | 0.07 |  |  |  | 0.36 |

HR: hazard ratio. 95%CI, 95% confidence interval. BC, breast cancer. PR, progesterone receptor. PY, person-years.

a P-LRT, p-values for the likelihood ratio test (LRT), that was used to evaluate overall significance of a score variable in quintile categories compared with a chi-square distribution with 4 df.

b P-trend values were obtained by modelling score variables with quintile-specific medians as continuous variables.

c P-heterogeneity values for BC risks across PR status on 1 df were obtained using a data augmentation method.

\*Models were stratified by study centre and age in 1-y categories and adjusted for baseline menopausal status (premenopausal and perimenopausal [reference] or postmenopausal and women who underwent an ovariectomy), baseline alcohol intake (never drinkers [reference], former drinkers, drinkers only at recruitment, lifetime drinkers, unknown), height (continuous), BMI (below [reference] or above 25), schooling level (none, primary [reference], technical/professional/secondary, longer education, unknown /unspecified), age at first full-term pregnancy (nulliparous [reference], ≤ 21years, 21-30 years, > 30 years, unknown or missing), age at menarche (≤ 12 years [reference], 12-14 years, >14 years, missing), age at menopause (≤50 years [reference], > 50 years, pre-menopause or missing), use of hormones (never[reference], ever, unknown), levels of physical activity (inactive [reference], moderately inactive, moderately active, active, unknown) and alcohol-free energy(continuous).

**Supplemental Table 4**: HRs (95%CI) for BC by quintiles of pattern scores (1st and 2nd components of PCA) for overall, ER positive and ER negative tumours in EPIC women.

|  |  |  |
| --- | --- | --- |
| Model\* | First component | Second component |
| PY | BC cases | HR (95% CI) | P-LRTa | P-trendb | PY | BC cases | HR (95% CI) | P-LRTa | P-trendb |
| Overall |  |  |  |  |  |  |  |  |  |  |
| Q1 | 729,222 | 1,843 | 1.00 (ref) | 0.29 | 0.07 | 748,437 | 2,143 | 1.00 (ref) | 0.15 | 0.046 |
| Q2 | 736,877 | 2,292 | 1.03 (0.96,1.09) | 737,177 | 2,339 | 1.03 (0.97,1.10) |
| Q3 | 734,382 | 2,445 | 1.06 (1.00,1.13) | 732,009 | 2,280 | 0.98 (0.92,1.04) |
| Q4 | 735,659 | 2,478 | 1.06 (1.00,1.13) | 727,730 | 2,354 | 0.98 (0.99,1.05) |
|  Q5 | 734,300 | 2,509 | 1.07 (1.00,1.15) | 725,087 | 2,460 | 0.96 (0.89,1.02) |
| ER Positive |  |  |  |  |  |  |  |  |  |  |
| Q1 | 723,700 | 882 | 1.00 (ref) | 0.27 | 0.09 | 741,994 | 1,087 | 1.00 (ref) | 0.46 | 0.10 |
| Q2 | 730,480 | 1,201 | 1.07 (0.98,1.17) | 730,010 | 1,142 | 1.00 (0.92,1.09) |
| Q3 | 727,426 | 1,260 | 1.09 (0.99,1.19) | 725,034 | 1,113 | 0.94 (0.86,1.03) |
| Q4 | 728,361 | 1,286 | 1.11 (1.01,1.22) | 720,800 | 1,173 | 0.94 (0.86,1.03) |
| Q5 | 726,145 | 1,201 | 1.09 (0.99,1.21) | 718,273 | 1,315 | 0.95 (0.86,1.04) |
| ER Negative |  |  |  |  |  |  |  |  |  |  |
| Q1 | 719,177 | 215 | 1.00 (ref) | 0.56 | 0.91 | 736,399 | 280 | 1.00 (ref) | 0.02 | 0.11 |
| Q2 | 724,194 | 287 | 1.01 (0.85,1.22) | 724,298 | 312 | 1.10 (0.93,1.30) |
| Q3 | 720,958 | 333 | 1.13 (0.94,1.35) | 719,335 | 301 | 1.05 (0.88,1.25) |
| Q4 | 721,850 | 306 | 1.01 (0.83,1.22) | 714,609 | 245 | 0.83 (0.69,1.00) |
| Q5 | 720,190 | 297 | 1.04 (0.85,1.27) | 711,728 | 300 | 0.96 (0.80,1.16) |
| *P- heterogeneity*c |  |  |  | 0.80 |  |  |  |  0.13 |

HR: hazard ratio. 95%CI, 95% confidence interval. BC, breast cancer. ER, estrogen receptor. PY, person-years.

a P-LRT, p-values for the likelihood ratio test (LRT), that was used to evaluate overall significance of a score variable in quintile categories compared with a chi-square distribution with 4 df.

b P-trend values were obtained by modelling score variables with quintile-specific medians as continuous variables.

c P-heterogeneity values for BC risks across ER status on 1 df were obtained using a data augmentation method.

\*Models were stratified by study centre and age in 1-y categories and adjusted for baseline menopausal status (premenopausal and perimenopausal [reference] or postmenopausal and women who underwent an ovariectomy), baseline alcohol intake (never drinkers [reference], former drinkers, drinkers only at recruitment, lifetime drinkers, unknown), height (continuous), BMI (below [reference] or above 25), schooling level (none, primary [reference], technical/professional/secondary, longer education, unknown /unspecified), age at first full-term pregnancy (nulliparous [reference], ≤ 21years, 21-30 years, > 30 years, unknown or missing), age at menarche (≤ 12 years [reference], 12-14 years, >14 years, missing), age at menopause (≤50 years [reference], > 50 years, pre-menopause or missing), use of hormones (never[reference], ever, unknown), levels of physical activity (inactive [reference], moderately inactive, moderately active, active, unknown) and alcohol-free energy(continuous).

**Supplemental Table 5**: HRs (95%CI) for BC by quintiles of pattern scores (1st and 2nd components of PCA) for ER & PR positive and ER & PR negative tumours in EPIC women.

|  |  |  |
| --- | --- | --- |
| Model\* | First component | Second component |
| PY | BC cases | HR (95% CI) | P-LRTa | P-trendb | PY | BC cases | HR (95% CI) | P-LRTa | P-trendb |
| ER and PR Positive |  |  |  |  |  |  |  |  |  |  |
| Q1 | 721,384 | 525 | 1.00 (ref) | 0.07 | 0.04 | 718,901 | 161 | 1.00 (ref) | 0.77 | 0.65 |
| Q2 | 727,780 | 775 | 1.15 (1.03,1.29) | 723,803 | 211 | 1.00 (0.81,1.23) |
| Q3 | 724,554 | 805 | 1.16 (1.03,1.31) | 720,508 | 242 | 1.09 (0.89,1.35) |
| Q4 | 725,315 | 790 | 1.16 (1.03,1.31) | 721,445 | 224 | 0.98 (0.79,1.23) |
| Q5 | 723,543 | 758 | 1.17 (1.03,1.33) | 719,832 | 212 | 0.99 (0.78,1.25) |
| ER and PR Negative |  |  |  |  |  |  |  |  |  |  |
| Q1 | 739,692 | 743 | 1.00 (ref) | 0.38 | 0.09 | 736,067 | 215 | 1.00 (ref) | 0.06 | <0.05 |
| Q2 | 727,688 | 774 | 1.03 (0.93,1.14) | 723,975 | 241 | 1.10 (0.91,1.32) |
| Q3 | 722,601 | 720 | 0.96 (0.86,1.07) | 718,949 | 214 | 0.97 (0.80,1.19) |
| Q4 | 717,804 | 694 | 0.94 (0.84,1.05) | 714,277 | 180 | 0.82 (0.66,1.02) |
| Q5 | 714,791 | 722 | 0.94 (0.84,1.06) | 711,222 | 200 | 0.90 (0.72,1.12) |
| *P- heterogeneity*c |  |  |  | 0.45 |  |  |  | 0.12 |

HR: hazard ratio. 95%CI, 95% confidence interval. BC, breast cancer. ER, estrogen receptor. PR, progesterone receptor. PY, person-years.

a P-LRT, p-values for the likelihood ratio test (LRT), that was used to evaluate overall significance of a score variable in quintile categories compared with a chi-square distribution with 4 df.

b P-trend values were obtained by modelling score variables with quintile-specific medians as continuous variables.

c P-heterogeneity values for BC risks across ER\PR status on 1 df were obtained using a data augmentation method.

\*Models were stratified by study centre and age in 1-y categories and adjusted for baseline menopausal status (premenopausal and perimenopausal [reference] or postmenopausal and women who underwent an ovariectomy), baseline alcohol intake (never drinkers [reference], former drinkers, drinkers only at recruitment, lifetime drinkers, unknown), height (continuous), BMI (below [reference] or above 25), schooling level (none, primary [reference], technical/professional/secondary, longer education, unknown /unspecified), age at first full-term pregnancy (nulliparous [reference], ≤ 21years, 21-30 years, > 30 years, unknown or missing), age at menarche (≤ 12 years [reference], 12-14 years, >14 years, missing), age at menopause (≤50 years [reference], > 50 years, pre-menopause or missing), use of hormones (never[reference], ever, unknown), levels of physical activity (inactive [reference], moderately inactive, moderately active, active, unknown) and alcohol-free energy(continuous).

**Supplementaly Figure 1**: Relations between PCA nutrient patterns and BC risk (and associated 95%CI) obtained by using restrictive cubic splines with values of 1st and 99th percentile and medians of quintiles 1, 3 and 5 used as knots.

Models were stratified by study centre and age in 1-y categories and adjusted for baseline menopausal status (premenopausal and perimenopausal [reference] or postmenopausal and women who underwent an ovariectomy), baseline alcohol intake (never drinkers [reference], former drinkers, drinkers only at recruitment, lifetime drinkers, unknown), height (continuous), BMI (below [reference] or above 25), schooling level (none, primary [reference], technical/ professional/ secondary, longer education, unknown / unspecified), age at first full-term pregnancy (nulliparous [reference], ≤ 21years, 21-30 years, > 30 years, unknown or missing), age at menarche (≤ 12 years [reference], 12-14 years, >14 years, missing), age at menopause (≤50 years [reference], > 50 years, pre-menopause or missing), use of hormones (never[reference], ever, unknown), levels of physical activity (inactive [reference], moderately inactive, moderately active, active, unknown) and alcohol-free energy(continuous). P-linearity was obtained by evaluating the joint significance of variables other than the linear one in the model by using Wald’s test with 3 df.