|  |
| --- |
| **Supplemental Table 1. GO analysis for the 25 Boruta method selected gene transcripts** |
| GO biological process complete | Reference Gene | Transcripts | Fold Enrichment | FDR | Boruta Identified Gene |
| defense response to bacterium (GO:0042742) | 300 | 05 | 13.73 | 4.4e-2 | *BPI, MPO, FCGR1A, LCN2, and CTSG* |
| defense response to other organism (GO:0098542) | 1008 | 10 | 8.17 | 2.1e-3 | *IFI6, DDX58, MPO, LCN2, IL4R, IFITM1, BPI, IFI44L, FCGR1A, and CTSG* |
| response to other organism (GO:0051707) | 1373 | 11 | 6.6 | 1.7e-3 | *IFI6, DDX58, MPO, LCN2, IL4R, IFITM1, ODC1, BPI, IFI44L, FCGR1A, and CTSG* |
| response to external biotic stimulus (GO:0043207) | 1376 | 11 | 6.58 | 1.2e-3 | *IFI6, DDX58, MPO, LCN2, IL4R, IFITM1, ODC1, BPI, IFI44L, FCGR1A, and CTSG* |
| response to biotic stimulus (GO:0009607) | 1421 | 11 | 6.38 | 1.2e-3 | *IFI6, DDX58, MPO, LCN2, IL4R, IFITM1, ODC1, BPI, IFI44L, FCGR1A, and CTSG* |
| biological process involved in interspecies interaction between organisms (GO:0044419) | 1553 | 11 | 5.83 | 2.4e-3 | *IFI6, DDX58, MPO, LCN2, IL4R, IFITM1, ODC1, BPI, IFI44L, FCGR1A, and CTSG* |
| defense response (GO:0006952) | 1431 | 10 | 5.76 | 7.4e-3 | *IFI6, DDX58, MPO, LCN2, IL4R, IFITM1, BPI, IFI44L, FCGR1A, and CTSG* |
| immune response (GO:0006955) | 1649 | 11 | 5.49 | 3.6e-3 | *IFI6, DDX58, LCN2, IL4R, IFITM1, BPI, IFI44L, FCGR1A, CTSG, CEACAM8, and TNFSF13B* |
| immune system process (GO:0002376) | 2277 | 12 | 4.34 | 8.6e-3 | *IFI6, DDX58, MPO, LCN2, IL4R, IFITM1, ODC1, BPI, IFI44L, FCGR1A, CTSG, and CEACAM8* |
| response to external stimulus (GO:0009605) | 2422 | 12 | 4.08 | 1.5e-2 | *IFI6, DDX58, MPO, LCN2, IL4R, IFITM1, ODC1, BPI, IFI44L, FCGR1A, CTSG, and DOCK4* |

|  |
| --- |
| **Supplemental Table 2. OOB error rate and AUC for different predictors**  |
|  | AUC based on the lowest OOB error rate | The largest AUC and corresponding OOB error rate |
|  | Nondrinkers vs. Heavy drinkers | Nondrinkers vs. Moderate drinkers | Moderate drinkers vs. Heavy drinkers | Nondrinkers vs. Heavy drinkers | Nondrinkers vs. moderate drinkers | Moderate drinkers vs. Heavy drinkers |
|  | OOB | AUC | OOB | AUC | OOB | AUC | OOB | AUC | OOB | AUC | OOB | AUC |
| Boruta method 1 | 0.31 | 0.73(0.68, 0.79) | 0.33 | 0.69(0.66, 0.71) | 0.33 | 0.66(0.61, 0.72) | 0.31 | 0.73(0.70, 0.80) | 0.35 | 0.69(0.66, 0.72) | 0.60 | 0.67(0.61, 0.71) |
| 1,958 transcripts 2 | 0.37 | 0.74(0.68, 0.79) | 0.41 | 0.66(0.63, 0.69) | 0.36 | 0.65(0.59, 0.70) | 0.38 | 0.74(0.70, 0.79) | 0.41 | 0.66(0.63, 0.69) | 0.46 | 0.66(0.61, 0.71) |
| 25 transcripts 3 | 0.23 | 0.73(0.68, 0.79) | 0.33 | 0.68(0.65, 0.71) | 0.41 | 0.68(0.62, 0.73) | 0.23 | 0.73(0.69, 0.79) | 0.38 | 0.68(0.65, 0.71) | 0.47 | 0.68(0.63, 0.73) |
| 144 CpGs 4 | 0.18 | 0.77(0.72, 0.82) | 0.32 | 0.70(0.66, 0.73) | 0.25 | 0.68(0.61, 0.75) | 0.19 | 0.78(0.73, 0.83) | 0.32 | 0.70(0.66, 0.73) | 0.26 | 0.69(0.63, 0.76) |
| Combined 5 | 0.24 | 0.77(0.71, 0.82) | 0.33 | 0.69(0.66, 0.73) | 0.22 | 0.70(0.63, 0.77) | 0.24 | 0.77(0.71, 0.82) | 0.35 | 0.70(0.66, 0.73) | 0.22 | 0.70(0.63, 0.77) |
| Set 1 was based on the 25 Boruta method selected transcripts. Sets 2 and 3 were from alcohol-gene expression analyses using conventional linear regression (reference 9). Set 4 was from meta-analysis of alcohol associated DNA methylation markers (reference 21). Set 5 combined predictors from sets 1, 3, and 4. 95% CI for AUC was listed in parentheses.  |

|  |
| --- |
| **Supplemental Table 3. *P* derived from Delong tests comparing prediction capability (AUC) of the 25 Boruta method selected transcripts with predictors based on previous studies** |
|  | Nondrinkers vs. Heavy drinkers | Nondrinkers vs. Moderate drinkers | Moderate drinkers vs. Heavy drinkers |
| 1,958 transcripts 1 | 0.79 | 0.03 | 0.66 |
| 25 transcripts 2 | 0.99 | 0.37 | 0.54 |
| 144 CpGs 3 | 0.40 | 0.65 | 0.67 |
| Combined 4 | 0.35 | 0.68 | 0.41 |
| Sets 1 and 2 were from alcohol-gene expression analyses using conventional linear regression (reference 9). Set 3 was from meta-analysis of alcohol associated DNA methylation markers (reference 21). Set 4 combined the Boruta method selected transcripts with sets 2 and 3 |

|  |
| --- |
| **Supplemental Table 4. Comparisons of the largest AUC derived from the primary analyses and sensitivity analyses** |
|  | Primary analyses | Sensitivity analyses |
|  | Nondrinkers vs. Heavy drinkers | Nondrinkers vs. Moderate drinkers | Moderate drinkers vs. Heavy drinkers | Nondrinkers vs. Light drinkers | Light drinkers vs. Heavy drinkers |
| Boruta method 1 | 0.73(0.68, 0.79) | 0.69(0.66, 0.71) | 0.66(0.61, 0.72) | 0.72(0.69, 0.75) | 0.64(0.60, 0.70) |
| 1,958 transcripts 2 | 0.74(0.68, 0.79) | 0.66(0.63, 0.67) | 0.65(0.59, 0.70) | 0.71(0.68, 0.74) | 0.64(0.59 0.70) |
| 25 transcripts 3 | 0.73(0.68, 0.79) | 0.68(0.65, 0.71) | 0.68(0.62, 0.73) | 0.71(0.68, 0.75) | 0.65(0.61, 0.70) |
| 144 CpGs 4 | 0.77(0.72, 0.82) | 0.70(0.66, 0.73) | 0.68(0.61, 0.75) | 0.72(0.69, 0.76) | 0.69(0.62, 0.76) |
| Combined 5 | 0.77(0.71, 0.82) | 0.69(0.66, 0.73) | 0.70(0.63, 0.77) | 0.72(0.69, 0.75) | 0.68(0.61, 0.75) |
| Set 1 was based on the 25 Boruta method selected transcripts. Sets 2 and 3 were from alcohol-gene expression analyses using conventional linear regression (reference 9). Set 4 was from meta-analysis of alcohol associated DNA methylation markers (reference 21). Set 5 combined predictors from sets 1, 3, and 4. 95% CI for AUC was listed in parentheses. |

**Supplemental Tables 5. Association of new**

**defined hypertension and transcripts**

|  |  |  |  |
| --- | --- | --- | --- |
| Gene | OR | 95%CI | *P* |
| *IL4R* | 0.47 | 0.37, 0.60 | 1.4e-9 |
| *RBM38* | 1.36 | 1.16, 1.60 | 1.7e-4 |
| *DOCK4* | 0.70 | 0.58, 0.84 | 1.7e-4 |

|  |
| --- |
| **Supplemental Tables 6. Association of alcohol and transcripts with consistent direction of association with expectation** |
| Gene | beta | 95% CI | *P* |
| *DOCK4* | -0.0017 | -0.0024, -0.0011 | 1.8e-7 |
| *FCGR1A* | -0.0009 | -0.0018, 0.0000 | 5.8e-2 |
| *IFITM1* | -0.0010 | -0.0016, -0.0004 | 2.4e-3 |
| *IL4R* | -0.0016 | -0.0021, -0.0011 | 1.3e-10 |
| *SORT1* | -0.0007 | -0.0011, -0.0003 | 3.4e-4 |
| Values are derived from fixed-effect meta-analysis based on conventional linear regression models (reference 9) |

|  |
| --- |
| **Supplemental Table 7. Interaction analyses for the association between alcohol and gene transcripts** |
| with sex interaction | with age interaction |
| Gene | *β*\_alcohol | *P*\_alcohol | *β*\_interaction | *P*\_interaction | *β*\_alcohol | *P*\_alcohol | *β*\_interaction | *P*\_interaction |
| *IFI44L* | 3.5e-4 | 0.90 | -1.2e-4 | 0.95 | -2.1e-3 | 0.61 | 4.2e-5 | 0.56 |
| *FCGR1A* | 2.0e-3 | 0.17 | -2.3e-3 | 0.03 | -4.6e-3 | 0.03 | 6.7e-5 | 0.08 |
| *IFI6* | 6.7e-4 | 0.58 | -2.9e-4 | 0.75 | 7.3e-4 | 0.69 | -7.7e-6 | 0.81 |
| *SORT1* | -2.4e-4 | 0.69 | -3.7e-4 | 0.42 | 1.3e-3 | 0.15 | -3.7e-5 | 0.02 |
| *MEIS1* | -3.3e-4 | 0.64 | 1.6e-3 | 2.6e-3 | 3.4e-3 | 1.8e-3 | -3.1e-5 | 0.11 |
| *ODC1* | 1.0e-3 | 0.27 | 8.1e-4 | 0.25 | 1.3e-4 | 0.93 | 3.5e-5 | 0.17 |
| *P2RY14* | -6.9e-5 | 0.95 | -1.0e-3 | 0.23 | 6.0e-4 | 0.73 | -3.5e-5 | 0.25 |
| *PLAGL1* | -1.6e-4 | 0.79 | -6.1e-4 | 0.17 | 1.6e-3 | 0.08 | -4.6e-5 | 4.3e-3 |
| *ABCA13* | 1.5e-3 | 0.01 | -2.1e-4 | 0.63 | 1.0e-3 | 0.24 | 2.5e-6 | 0.87 |
| *DOCK4* | 6.9e-4 | 0.49 | -1.9e-3 | 0.01 | -7.0e-4 | 0.65 | -1.9e-5 | 0.49 |
| *GAPVD1* | 5.7e-4 | 0.35 | -4.0e-4 | 0.38 | 1.3e-3 | 0.17 | -2.2e-5 | 0.19 |
| *LCN2* | 2.3e-3 | 0.14 | 5.7e-4 | 0.62 | 1.7e-3 | 0.47 | 2.4e-5 | 0.56 |
| *DDX58* | 8.5e-4 | 0.42 | -8.4e-4 | 0.29 | -7.5e-4 | 0.64 | 9.8e-6 | 0.73 |
| *IFITM1* | 3.6e-4 | 0.71 | -1.0e-3 | 0.16 | -1.7e-3 | 0.27 | 1.3e-5 | 0.62 |
| *UTP20* | 4.2e-4 | 0.34 | 2.7e-4 | 0.42 | 1.9e-3 | 4.5e-3 | -2.1e-5 | 0.08 |
| *OLFM4* | 3.5e-3 | 3.7e-4 | -4.2e-4 | 0.65 | 1.6e-3 | 0.38 | 2.4e-5 | 0.47 |
| *TNFSF13B* | 3.6e-4 | 0.65 | -1.2e-3 | 0.04 | -3.9e-4 | 0.74 | -1.4e-5 | 0.50 |
| *CTSG* | 4.7e-3 | 0.00 | -9.2e-4 | 0.36 | 5.5e-3 | 0.01 | -3.5e-5 | 0.32 |
| *IL4R* | -6.1e-4 | 0.41 | -7.8e-4 | 0.16 | -2.3e-3 | 0.04 | 1.3e-5 | 0.51 |
| *MPO* | 1.5e-3 | 0.03 | -5.3e-6 | 0.99 | 3.0e-3 | 3.3e-3 | -2.7e-5 | 0.12 |
| *CYTH1* | 5.8e-5 | 0.92 | -3.5e-4 | 0.40 | 8.2e-4 | 0.34 | -2.2e-5 | 0.15 |
| *ATP5F1D* | -1.6e-4 | 0.72 | -4.6e-5 | 0.89 | -6.3e-4 | 0.35 | 7.6e-6 | 0.52 |
| *CEACAM8* | 3.6e-3 | 0.02 | -5.6e-4 | 0.62 | 4.1e-3 | 0.07 | -2.1e-5 | 0.60 |
| *BPI* | 3.9e-3 | 4.1e-3 | -8.5e-4 | 0.40 | 2.1e-3 | 0.31 | 1.3e-5 | 0.73 |
| *RBM38* | 1.7e-4 | 0.88 | 1.1e-3 | 0.21 | -1.3e-3 | 0.46 | 5.1e-5 | 0.10 |
|  |
| **Supplemental Tables 8. Interaction analyses for the association between transcripts and CVD risk factors**  |
|  | *P* for interaction with sex | *P* for interaction with age |
| Gene | Obesity | Hypertension | Type 2 diabetes | Obesity | Hypertension | Type 2 diabetes |
| *IFI44L* | 0.42 | 0.85 | 0.23 | 0.45 | 0.09 | 0.52 |
| *FCGR1A* | 0.01 | 0.72 | 0.18 | 0.07 | 0.29 | 0.36 |
| *IFI6* | 0.97 | 0.94 | 0.28 | 0.69 | 0.68 | 0.96 |
| *SORT1* | 0.05 | 0.69 | 0.78 | 0.22 | 0.90 | 0.47 |
| *MeIS1* | 0.17 | 0.09 | 0.93 | 0.09 | 0.32 | 0.79 |
| *ODC1* | 0.03 | 0.73 | 0.12 | 0.03 | 0.10 | 0.93 |
| *P2RY14* | 0.03 | 0.11 | 0.09 | 0.60 | 0.13 | 0.67 |
| *PLAGL1* | 0.02 | 0.13 | 0.14 | 0.56 | 0.76 | 0.04 |
| *ABCA13* | 0.16 | 0.28 | 0.39 | 0.74 | 0.92 | 0.16 |
| *DOCK4* | 5.5e-5 | 0.47 | 0.33 | 0.19 | 0.46 | 0.28 |
| *GAPVD1* | 0.12 | 0.14 | 0.66 | 0.28 | 0.43 | 0.18 |
| *LCN2* | 0.91 | 0.62 | 0.95 | 0.22 | 0.10 | 0.44 |
| *DDX58* | 0.25 | 0.94 | 0.76 | 0.69 | 0.07 | 0.50 |
| *IFITM1* | 0.05 | 0.38 | 0.26 | 0.54 | 0.01 | 0.62 |
| *UTP20* | 0.56 | 0.44 | 0.20 | 0.93 | 0.47 | 0.48 |
| *OLFM4* | 0.29 | 0.06 | 0.12 | 0.59 | 0.67 | 0.63 |
| *TNFSF13B* | 0.01 | 0.38 | 0.12 | 0.07 | 0.08 | 0.06 |
| *CTSG* | 1.6e-3 | 0.39 | 0.28 | 0.84 | 0.66 | 0.70 |
| *IL4R* | 0.02 | 0.92 | 0.03 | 0.01 | 0.34 | 0.28 |
| *MPO* | 2.9e-5 | 0.18 | 0.16 | 0.30 | 0.50 | 0.77 |
| *CYTH1* | 0.48 | 0.50 | 0.60 | 0.13 | 0.47 | 0.09 |
| *ATP5F1D* | 0.97 | 0.43 | 0.93 | 0.09 | 0.01 | 0.52 |
| *CEACAM8* | 0.36 | 0.69 | 0.38 | 0.34 | 0.23 | 0.74 |
| *BPI* | 0.07 | 0.13 | 0.74 | 0.11 | 0.60 | 0.93 |
| *RBM38* | 2.9e-4 | 0.14 | 0.96 | 0.48 | 0.68 | 0.55 |

|  |
| --- |
| **Supplemental Tables 9. Association of transcripts and CVD risk factors Stratified in Male** |
|   |  Obesity |   | Hypertension | Type 2 diabetes |
| Gene | OR |  95% CI | *P* | OR | 95% CI | *P* | OR | 95% CI | *P* |
| *IFI44L* | 0.94 | 0.86, 1.03 | 0.19 | 1.06 | 0.96, 1.17 | 0.23 | 1.10 | 0.97, 1.25 | 0.14 |
| *FCGR1A* | 1.33 | 1.14, 1.56 | 3.8e-4 | 0.95 | 0.79, 1.15 | 0.59 | 1.18 | 0.89, 1.57 | 0.26 |
| *IFI6* | 1.10 | 0.91, 1.33 | 0.31 | 1.13 | 0.91, 1.40  | 0.27 | 1.28 | 0.96, 1.70 | 0.09 |
| *SORT1* | 2.01 | 1.39, 2.91 | 2.0e-4 | 0.96 | 0.64, 1.43 | 0.83 | 1.20 | 0.65, 2.20 | 0.56 |
| *MeIS1* | 0.99 | 0.71, 1.38 | 0.97 | 1.18 | 0.81, 1.73 | 0.39 | 0.90 | 0.52, 1.54 | 0.70 |
| *ODC1* | 1.77 | 1.41, 2.22 | 1.1e-6 | 1.27 | 0.98, 1.65 | 0.07 | 2.10 | 1.41, 3.13 | 2.8e-4 |
| *P2RY14* | 1.04 | 0.85, 1.28 | 0.68 | 1.25 | 1.00, 1.58  | 0.05 | 1.29 | 0.94, 1.77 | 0.12 |
| *PLAGL1* | 0.83 | 0.57, 1.20 | 0.32 | 0.70 | 0.45, 1.08 | 0.11 | 0.35 | 0.18, 0.70 | 3.1e-3 |
| *ABCA13* | 2.49 | 1.77, 3.49 | 1.4e-7 | 1.01 | 0.70, 1.45 | 0.97 | 1.50 | 0.97, 2.30 | 0.07 |
| *DOCK4* | 1.30 | 1.03, 1.64 | 0.03 | 0.85 | 0.65, 1.11 | 0.24 | 0.78 | 0.52, 1.17 | 0.22 |
| *GAPVD1* | 2.49 | 1.73, 3.58 | 8.4e-7 | 1.31 | 0.88, 1.95 | 0.18 | 0.99 | 0.51, 1.91 | 0.98 |
| *LCN2* | 1.69 | 1.48, 1.94 | 6.4e-15 | 1.04 | 0.89, 1.21 | 0.62 | 1.27 | 1.06, 1.52 | 0.01 |
| *DDX58* | 1.04 | 0.83, 1.29 | 0.74 | 1.08 | 0.84, 1.39 | 0.56 | 0.86 | 0.59, 1.26 | 0.45 |
| *IFITM1* | 1.10 | 0.87, 1.38 | 0.43 | 1.22 | 0.95, 1.57 | 0.12 | 1.32 | 0.91, 1.92 | 0.15 |
| *UTP20* | 2.44 | 1.49, 4.02 | 4.4e-4 | 1.55 | 0.90, 2.68 | 0.11 | 0.67 | 0.29, 1.53 | 0.34 |
| *OLFM4* | 1.55 | 1.32, 1.81 | 5.7e-8 | 1.08 | 0.90, 1.29  | 0.44 | 1.15 | 1.45, 0.92 | 0.22 |
| *TNFSF13B* | 0.73 | 0.54, 0.98 | 0.04 | 1.05 | 0.75, 1.48 | 0.78 | 1.23 | 0.77, 1.98 | 0.39 |
| *CTSG* | 1.33 | 1.15, 1.53 | 1.0e-4 | 1.11 | 0.94, 1.31 | 0.22 | 1.14 | 0.92, 1.40 | 0.23 |
| *IL4R* | 0.92 | 0.69, 1.22 | 0.55 | 0.50 | 0.36, 0.71 | 8.0e-5 | 0.96 | 0.60, 1.55 | 0.88 |
| *MPO* | 1.46 | 1.09, 1.96 | 0.01 | 1.02 | 0.73, 1.43 | 0.89 | 1.16 | 0.76, 1.77 | 0.50 |
| *CYTH1* | 1.04 | 0.71, 1.51 | 0.85 | 1.32 | 0.86, 2.05 | 0.21 | 0.82 | 0.42, 1.63 | 0.57 |
| *ATP5F1D* | 0.96 | 0.57, 1.60 | 0.86 | 1.06 | 0.61, 1.87 | 0.83 | 1.48 | 0.63, 3.45 | 0.37 |
| *CEACAM8* | 1.60 | 1.40, 1.83 | 7.1e-12 | 1.02 | 0.88, 1.19 | 0.78 | 1.16 | 0.96, 1.42 | 0.13 |
| *BPI* | 1.39 | 1.20, 1.60 | 7.6e-6 | 0.99 | 0.84, 1.17 | 0.93 | 1.17 | 0.95, 1.44 | 0.15 |
| *RBM38* | 1.66 | 1.38, 2.00 | 7.9e-8 | 1.14 | 0.92, 1.41 | 0.22 | 1.68 | 1.22, 2.32 | 1.6e-3 |

|  |
| --- |
| **Supplemental Tables 10. Association of transcripts and CVD risk factors Stratified in Female** |
|  |  | Obesity |  |  | Hypertension |  | Type 2 diabetes |
| Gene | OR | 95% CI | *P* | OR | 95% CI | *P* | OR | 95% CI | *P* |
| *IFI44L* | 1.00 | 0.92, 1.08 | 0.99 | 1.08 | 0.99, 1.18 | 0.07 | 1.01 | 0.88, 1.17 | 0.84 |
| *FCGR1A* | 1.72 | 1.48, 2.00 | 1.9e-12 | 0.87 | 0.73, 1.04 | 0.13 | 0.93 | 0.69, 1.26 | 0.64 |
| *IFI6* | 1.11 | 0.93, 1.33 | 0.26 | 1.17 | 0.96, 1.43 | 0.12 | 1.11 | 0.81, 1.51 | 0.53 |
| *SORT1* | 3.10 | 2.12, 4.53 | 4.6e-9 | 0.90 | 0.59, 1.38 | 0.63 | 1.12 | 0.52, 2.39 | 0.77 |
| *MeIS1* | 1.35 | 1.00, 1.81 | 0.05 | 1.86 | 1.32, 2.62 | 3.7e-4 | 0.95 | 0.52, 1.72 | 0.86 |
| *ODC1* | 2.59 | 2.01, 3.32 | 1.3e-13 | 1.47 | 1.11, 1.96 | 0.01 | 1.36 | 0.84, 2.20 | 0.21 |
| *P2RY14* | 1.38 | 1.15, 1.65 | 4.9e-4 | 0.92 | 0.74, 1.14 | 0.44 | 0.84 | 0.59, 1.20 | 0.34 |
| *PLAGL1* | 1.48 | 1.04, 2.10 | 0.03 | 0.95 | 0.62, 1.45 | 0.81 | 0.70 | 0.37, 1.30 | 0.26 |
| *ABCA13* | 2.15 | 1.35, 3.43 | 1.3e-3 | 0.9 | 0.53, 1.52 | 0.69 | 2.42 | 1.09, 5.36 | 0.03 |
| *DOCK4* | 2.48 | 1.98, 3.11 | 2.0e-15 | 0.76 | 0.58, 0.99 | 0.04 | 1.08 | 0.70, 1.66 | 0.74 |
| *GAPVD1* | 3.36 | 2.28, 4.93 | 6.9e-10 | 0.82 | 0.52, 1.28 | 0.37 | 1.24 | 0.59, 2.61 | 0.58 |
| *LCN2* | 1.82 | 1.55, 2.15 | 8.0e-13 | 1.09 | 0.91, 1.32 | 0.36 | 1.37 | 1.02, 1.85 | 0.04 |
| *DDX58* | 1.29 | 1.06, 1.57 | 1.2e-2 | 1.09 | 0.87, 1.37 | 0.43 | 0.88 | 0.60, 1.29 | 0.53 |
| *IFITM1* | 1.57 | 1.26, 1.94 | 4.4e-5 | 1.10 | 0.85, 1.42 | 0.48 | 1.07 | 0.69, 1.66 | 0.76 |
| *UTP20* | 1.87 | 1.13, 3.08 | 0.01 | 1.43 | 0.79, 2.58 | 0.23 | 1.71 | 0.62, 4.73 | 0.30 |
| *OLFM4* | 1.48 | 1.21, 1.82 | 1.6e-4 | 0.90 | 0.72, 1.11 | 0.32 | 1.61 | 1.19, 2.18 | 2.3e-3 |
| *TNFSF13B* | 1.22 | 0.92, 1.63 | 0.16 | 0.82 | 0.59, 1.15 | 0.24 | 0.68 | 0.36, 1.26 | 0.22 |
| *CTSG* | 1.02 | 0.83, 1.24 | 0.88 | 1.05 | 0.84, 1.32 | 0.67 | 1.41 | 1.00, 1.99 | 0.05 |
| *IL4R* | 1.57 | 1.17, 2.10 | 2.5e-3 | 0.47 | 0.34, 0.66 | 1.4e-5 | 0.43 | 0.24, 0.79 | 0.01 |
| *MPO* | 0.65 | 0.42, 1.00 | 0.05 | 0.80 | 0.52, 1.25 | 0.33 | 2.08 | 1.04, 4.17 | 0.04 |
| *CYTH1* | 1.35 | 0.90, 2.02 | 0.14 | 0.94 | 0.59, 1.50 | 0.8 | 1.05 | 0.46, 2.35 | 0.91 |
| *ATP5F1D* | 0.95 | 0.59, 1.54 | 0.85 | 1.67 | 0.94, 2.99 | 0.08 | 1.68 | 0.62, 4.60 | 0.31 |
| *CEACAM8* | 1.62 | 1.37, 1.91 | 1.7e-8 | 1.06 | 0.88, 1.28 | 0.52 | 1.44 | 1.07, 1.93 | 0.02 |
| *BPI* | 1.24 | 1.02, 1.51 | 0.03 | 0.87 | 0.70, 1.08 | 0.21 | 1.32 | 0.92, 1.90 | 0.13 |
| *RBM38* | 2.65 | 2.17, 3.23 | 7.9e-22 | 1.56 | 1.23, 1.97 | 2.1e-4 | 1.76 | 1.20, 2.58 | 4.1e-3 |

|  |
| --- |
| **Supplemental Tables 11. Association of transcripts and CVD risk factors Stratified in younger people** |
|  |  | Obesity |  |  | Hypertension |  | Type 2 diabetes |
| Gene | OR | 95% CI | *P* | OR | 95% CI | *P* | OR | 95% CI | *P* |
| *IFI44L* | 0.99 | 0.91, 1.08 | 0.83 | 1.02 | 0.92, 1.14 | 0.68 | 1.04 | 0.84, 1.27 | 0.74 |
| *FCGR1A* | 1.57 | 1.34, 1.84 | 2.0e-8 | 0.85 | 0.68, 1.07 | 0.16 | 0.82 | 0.49, 1.37 | 0.45 |
| *IFI6* | 1.15 | 0.96, 1.37 | 0.13 | 1.08 | 0.85, 1.37 | 0.55 | 1.19 | 0.75, 1.90 | 0.46 |
| *SORT1* | 2.77 | 1.89, 4.06 | 1.6e-7 | 1.19 | 0.74, 1.91 | 0.48 | 0.90 | 0.34, 2.36 | 0.82 |
| *MeIS1* | 1.24 | 0.92, 1.68 | 0.16 | 1.73 | 1.17, 2.56 | 0.01 | 1.09 | 0.42, 2.86 | 0.86 |
| *ODC1* | 2.32 | 1.80, 3.00 | 1.3e-10 | 1.51 | 1.09, 2.09 | 0.01 | 1.15 | 0.60, 2.23 | 0.67 |
| *P2RY14* | 1.28 | 1.05, 1.56 | 0.01 | 0.94 | 0.72, 1.23 | 0.67 | 0.71 | 0.43, 1.18 | 0.19 |
| *PLAGL1* | 1.29 | 0.89, 1.85 | 0.18 | 0.98 | 0.61, 1.60 | 0.95 | 0.18 | 0.07, 0.46 | 3.9e-4 |
| *ABCA13* | 2.22 | 1.51, 3.27 | 4.5e-5 | 0.93 | 0.60, 1.44 | 0.76 | 2.19 | 1.13, 4.25 | 0.02 |
| *DOCK4* | 1.96 | 1.55, 2.48 | 2.5e-8 | 0.87 | 0.64, 1.18 | 0.36 | 0.58 | 0.31, 1.10 | 0.10 |
| *GAPVD1* | 3.17 | 2.20, 4.57 | 5.3e-10 | 1.33 | 0.84, 2.11 | 0.23 | 0.61 | 0.25, 1.47 | 0.27 |
| *LCN2* | 1.76 | 1.52, 2.04 | 9.4e-14 | 0.91 | 0.77, 1.09 | 0.32 | 1.21 | 0.90, 1.65 | 0.21 |
| *DDX58* | 1.13 | 0.92, 1.38 | 0.24 | 1.00 | 0.75, 1.32 | 0.98 | 0.68 | 0.38, 1.23 | 0.20 |
| *IFITM1* | 1.25 | 1.01, 1.55 | 0.04 | 0.83 | 0.62, 1.11 | 0.21 | 1.11 | 0.61, 2.02 | 0.73 |
| *UTP20* | 2.40 | 1.45, 3.97 | 6.3e-4 | 2.44 | 1.26, 4.72 | 0.01 | 1.76 | 0.51, 6.10 | 0.38 |
| *OLFM4* | 1.52 | 1.27, 1.82 | 5.9e-6 | 0.89 | 0.72, 1.10 | 0.26 | 1.41 | 1.00, 1.99 | 0.05 |
| *TNFSF13B* | 1.22 | 0.92, 1.60 | 0.17 | 0.79 | 0.53, 1.16 | 0.23 | 0.47 | 0.22, 1.03 | 0.06 |
| *CTSG* | 1.12 | 0.94, 1.33 | 0.22 | 1.02 | 0.83, 1.25 | 0.87 | 1.30 | 0.89, 1.89 | 0.18 |
| *IL4R* | 1.82 | 1.33, 2.50 | 1.9e-4 | 0.48 | 0.32, 0.72 | 4.0e-4 | 0.60 | 0.24, 1.54 | 0.29 |
| *MPO* | 1.24 | 0.87, 1.78 | 0.24 | 0.83 | 0.55, 1.26 | 0.39 | 1.63 | 0.78, 3.40 | 0.20 |
| *CYTH1* | 1.32 | 0.88, 1.96 | 0.18 | 1.14 | 0.69, 1.88 | 0.62 | 0.51 | 0.19, 1.35 | 0.18 |
| *ATP5F1D* | 0.96 | 0.58, 1.56 | 0.86 | 0.78 | 0.42, 1.46 | 0.44 | 2.90 | 0.78, 10.83 | 0.11 |
| *CEACAM8* | 1.61 | 1.38, 1.87 | 7.2e-10 | 0.91 | 0.76, 1.09 | 0.29 | 1.21 | 0.88, 1.66 | 0.24 |
| *BPI* | 1.40 | 1.18, 1.65 | 8.0e-5 | 0.88 | 0.72, 1.08 | 0.21 | 1.26 | 0.89, 1.78 | 0.19 |
| *RBM38* | 2.12 | 1.74, 2.59 | 1.4e-13 | 1.35 | 1.05, 1.73 | 0.02 | 1.55 | 0.90, 2.66 | 0.11 |

|  |
| --- |
| **Supplemental Tables 12. Association of transcripts and CVD risk factors Stratified in older people** |
|  |  | Obesity |  |  | Hypertension |  | Type 2 diabetes |
| Gene | OR | 95% CI | *P* | OR | 95% CI | *P* | OR | 95% CI | *P* |
| *IFI44L* | 0.97 | 0.90, 1.06 | 0.54 | 1.10 | 1.01, 1.19 | 0.03 | 1.05 | 0.94, 1.17 | 0.40 |
| *FCGR1A* | 1.49 | 1.28, 1.74 | 2.9e-7 | 0.94 | 0.81, 1.10 | 0.47 | 1.08 | 0.86, 1.36 | 0.51 |
| *IFI6* | 1.12 | 0.93, 1.34 | 0.24 | 1.17 | 0.97, 1.42 | 0.09 | 1.14 | 0.90, 1.45 | 0.28 |
| *SORT1* | 2.71 | 1.87, 3.92 | 1.5e-7 | 0.80 | 0.56, 1.16 | 0.25 | 1.06 | 0.62, 1.80 | 0.84 |
| *MeIS1* | 1.03 | 0.75, 1.41 | 0.85 | 1.33 | 0.96, 1.84 | 0.09 | 0.82 | 0.54, 1.26 | 0.37 |
| *ODC1* | 1.75 | 1.39, 2.20 | 2.0e-6 | 1.25 | 0.98, 1.58 | 0.07 | 2.08 | 1.48, 2.93 | 2.8e-5 |
| *P2RY14* | 1.20 | 0.99, 1.45 | 0.06 | 1.10 | 0.91, 1.34 | 0.34 | 1.09 | 0.83, 1.43 | 0.52 |
| *PLAGL1* | 1.14 | 0.78, 1.66 | 0.51 | 0.70 | 0.48, 1.03 | 0.07 | 0.6 | 0.35, 1.01 | 0.05 |
| *ABCA13* | 2.36 | 1.58, 3.50 | 2.3e-5 | 0.98 | 0.64, 1.48 | 0.91 | 1.57 | 1.00, 2.46 | 0.05 |
| *DOCK4* | 1.71 | 1.37, 2.14 | 2.4e-6 | 0.76 | 0.60, 0.96 | 0.02 | 0.96 | 0.68, 1.34 | 0.8 |
| *GAPVD1* | 2.93 | 1.97, 4.34 | 9.3e-8 | 0.92 | 0.63, 1.36 | 0.68 | 1.31 | 0.74, 2.32 | 0.35 |
| *LCN2* | 1.66 | 1.43, 1.92 | 2.0e-11 | 1.18 | 1.00, 1.38 | 0.05 | 1.35 | 1.13, 1.61 | 1.0e-3 |
| *DDX58* | 1.30 | 1.05, 1.60 | 0.01 | 1.14 | 0.92, 1.42 | 0.23 | 0.86 | 0.63, 1.18 | 0.34 |
| *IFITM1* | 1.38 | 1.10, 1.72 | 4.9e-3 | 1.40 | 1.11, 1.77 | 4.6e-3 | 1.20 | 0.87, 1.66 | 0.26 |
| *UTP20* | 1.88 | 1.14, 3.10 | 0.01 | 1.14 | 0.69, 1.88 | 0.61 | 0.81 | 0.39, 1.71 | 0.59 |
| *OLFM4* | 1.48 | 1.24, 1.76 | 1.0e-5 | 1.09 | 0.90, 1.32 | 0.39 | 1.27 | 1.02, 1.58 | 0.03 |
| *TNFSF13B* | 0.78 | 0.58, 1.04 | 0.09 | 1.01 | 0.76, 1.35 | 0.92 | 1.18 | 0.78, 1.80 | 0.43 |
| *CTSG* | 1.26 | 1.08, 1.48 | 3.5e-3 | 1.15 | 0.96, 1.38 | 0.12 | 1.22 | 1.00, 1.50 | 0.05 |
| *IL4R* | 0.84 | 0.64, 1.11 | 0.23 | 0.48 | 0.36, 0.64 | 4.3e-7 | 0.66 | 0.43, 0.99 | 0.05 |
| *MPO* | 0.97 | 0.70, 1.34 | 0.85 | 1.03 | 0.73, 1.46 | 0.86 | 1.38 | 0.92, 2.05 | 0.12 |
| *CYTH1* | 1.18 | 0.80, 1.75 | 0.40 | 1.19 | 0.79, 1.80 | 0.41 | 1.08 | 0.59, 1.99 | 0.80 |
| *ATP5F1D* | 0.95 | 0.56, 1.59 | 0.83 | 1.80 | 1.06, 3.05 | 0.03 | 1.37 | 0.67, 2.81 | 0.38 |
| *CEACAM8* | 1.55 | 1.34, 1.80 | 5.4e-9 | 1.15 | 0.98, 1.34 | 0.10 | 1.28 | 1.06, 1.55 | 0.01 |
| *BPI* | 1.25 | 1.06, 1.47 | 0.01 | 1.01 | 0.85, 1.20 | 0.89 | 1.21 | 0.98, 1.49 | 0.07 |
| *RBM38* | 1.94 | 1.60, 2.35 | 1.8e-11 | 1.24 | 1.02, 1.52 | 0.03 | 1.80 | 1.37, 2.37 | 3.1e-5 |

**Supplemental Figure 1. Gene expression level by drinking categories.** Nondrinker=1, moderate drinker=2, heavy drinker=3. The Y axis is beta regression coefficients for association analyses with gene expression levels in nondrinkers as the reference.

**Supplemental Figure 2. Pearson correlation coefficients of the 25 gene transcripts selected by Boruta method**

**Supplemental Figure 3. Pearson correlation coefficients of the 13 unique gene transcripts selected by Boruta method and 101 gene transcripts selected by linear regression**



**Supplemental Figure 4. Ancestor chart of the significant GO terms.** Figure downloaded from Quick GO for GO:0042742 and GO:0006955 (https://www.ebi.ac.uk/QuickGO/).

**Supplemental code**

To demonstrating the analytical R scripts, we generated the toydata set based on GSE21653 (https://www.ncbi.nlm.nih.gov/geo/). We transposed the data and randomly created two outcome variables (one continuous and one binary), as well as a set of covariates such as sex, age.

library(Boruta)

library(tidyverse)

library(randomForest)

library(pROC)

library(geneExpressionFromGEO)

library(caret)

geneExpressionDF1 <- getGeneExpressionFromGEO("GSE21653", FALSE, FALSE)

toydata<-as.data.frame(t(geneExpressionDF1))

colnames(toydata)<-paste0("X", colnames(toydata))

toydata$categorical\_outcome <-rbinom(n=266,size=1,prob=0.3)

toydata$continuous\_outcome <-runif(266, 0, 50)

toydata$sex<-rbinom(n=266,size=1,prob=0.45)

toydata$age<-sample(x = 35:70, size = 266, replace = TRUE)

set.seed(1234)

bor<-Boruta(continuous\_outcome~.,data= toydata ,doTrace=2)

important<-getSelectedAttributes(bor, withTentative = F)

toydata <- toydata[,colnames(toydata) %in% c("categorical\_outcome", important)]

#Split the dataset to traning dataset and testing dataset

set.seed(1234)

dataRF\_train\_all <- sample(266, 266\*0.7)

dataRF\_all\_train <- toydata[dataRF\_train\_all, ]

dataRF\_all\_test <- toydata[-dataRF\_train\_all, ]

set.seed(1234)

dataRFwalltrain.forest <- randomForest(as.factor(categorical\_outcome) ~ ., data = dataRF\_all\_train, importance = TRUE, na.action=na.omit)

test.predictions <- predict(dataRFwalltrain.forest, dataRF\_all\_test, type = "prob")

roc <- roc(dataRF\_all\_test$categorical\_outcome, test.predictions[,2])

auc\_d<-as.numeric(auc(roc))

thre\_pre<-function(x){

 set.seed(1234)

 dataRFwalltrain.forest <- randomForest(as.factor(categorical\_outcome) ~ ., data = dataRF\_all\_train, importance = TRUE, na.action=na.omit,cutoff = c(1-x, x))

 test.predictions <- predict(dataRFwalltrain.forest, dataRF\_all\_test, type = "prob")

 roc <- roc(dataRF\_all\_test$categorical\_outcome, test.predictions[,2])

 thres<-coords(roc, "best", ret = "threshold")

 return(as.numeric(thres))

}

thre\_list <- function(x){

 if (length(x)==10)

 {return(x)}

 else {x <- c(x, thre\_pre(x[length(x)]))

 ; thre\_list(x)

 }

}

threshold\_list <- as.data.frame(thre\_list(auc\_d))

table <- as.data.frame(matrix(NA, 10, 5))

#######auc list

for (i in 1:10){

 a<-threshold\_list[i,]

 set.seed(1234)

 dataRFwalltrain.forest <- randomForest(as.factor(categorical\_outcome) ~ ., data = dataRF\_all\_train, importance = TRUE, na.action=na.omit,cutoff = c(1-a, a))

 OOB <- dataRFwalltrain.forest$err.rate[500]

 test.predictions <- predict(dataRFwalltrain.forest, dataRF\_all\_test, type = "prob")

 roc <- roc(dataRF\_all\_test$categorical\_outcome, test.predictions[,2])

 auc <- as.numeric(auc(roc))

 data\_pre <- predict(dataRFwalltrain.forest,dataRF\_all\_test )

 conMa <- confusionMatrix(data\_pre, as.factor(dataRF\_all\_test$categorical\_outcome))

 output <- c(a, OOB, auc, conMa$byClass[1],conMa$byClass[2] )

 table[i,] <- output

}

names(table) = c("threshold","OOB","AUC","sensitivity","specificity")