

1 **Title:** Transcriptomic and Metabolomic analyses in Monozygotic and Dizygotic twins.

2 **Running Title:** Transcriptomics and metabolomics in MZ and DZ twins

3

4 **Nikki Hubers^{1,2,3}, Gabin Drouard⁴, Rick Jansen^{3,5}, René Pool^{1, 3}, Jouke Jan Hottenga¹,**
5 **Miina Ollikainen^{4,6}, Xiaoling Wang⁷, Gonneke Willemsen⁸, Jaakko Kaprio⁴, Dorret I.**
6 **Boomsma^{2,3,9} & Jenny van Dongen^{1,2,3}**

7 ¹Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the
8 Netherlands

9 ²Amsterdam Reproduction & Development (AR&D) research institute, Amsterdam, the
10 Netherlands

11 ³Amsterdam Public Health research institute, Amsterdam, the Netherlands

12 ⁴Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki,
13 Finland

14 ⁵ Amsterdam UMC location Vrije Universiteit Amsterdam, Department of Psychiatry &
15 Amsterdam Neuroscience - Complex Trait Genetics (VUmc) and Mood, Anxiety, Psychosis,
16 Stress & Sleep

17 ⁶ Minerva Foundation Institute for Medical Research, Helsinki, Finland.

18 ⁷Georgia Prevention Institute, Medical College of Georgia, Augusta University, Augusta, GA,
19 USA

20 ⁸Faculty of Health, Sport and Wellbeing, Inholland University of Applied Sciences, Haarlem,
21 the Netherlands

22 ⁹Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive
23 Research, Amsterdam, Vrije Universiteit Amsterdam

24

25 **Corresponding Author:**

26 Nikki Hubers

27 PhD student, Department of Biological Psychology, Vrije Universiteit Amsterdam

28 Van der Boechorststraat 7-9, Room MF-H-557, 1081 BT, Amsterdam, The Netherlands

29 Phone: +31625508825

30 Email: n.hubers@vu.nl

31

32 **Abstract**

33 Monozygotic (MZ) and dizygotic (DZ) twins are often studied to determine genetic and
34 environmental influences on complex traits, however, the biological mechanisms behind MZ
35 and DZ twinning are not completely understood. Genomic and epigenomic studies have
36 identified SNPs associated with DZ twinning and DNA methylation sites associated with MZ
37 twinning. To enhance the discovery of molecular biomarkers of twinning, we compare
38 transcriptomics and metabolomics data from MZ with those from DZ twins. We compared
39 42,663 RNA transcripts in 1,453 MZ twins and 1,294 DZ twins from the Netherlands Twin
40 Register (NTR), followed by sex-stratified analyses. The 5% transcripts with lowest p-values
41 were selected for replication analysis in 217 MZ and 158 DZ twins from the older Finnish Twin
42 cohort (FTC). In the NTR sample, we observed upregulations of the protein coding *PURG* gene
43 in MZ twins. The female-only analyses confirmed the *PURG* gene and identified four other
44 genes, while the male-only analyses indicated three other genes associated with either MZ and
45 DZ twinning. Replication results in the FTC, did not confirm *PURG*, but revealed seven
46 differentially expressed genes with nominal significant p-values in both cohorts, none of which
47 have been implicated for twinning before. Pathway enrichment showed differences in
48 expression in the WNT-pathway and cell adhesion processes, which were previously indicated
49 with MZ twinning though an epigenetic study, and the TGF-B pathway known to be associated
50 with DZ twinning through genetic studies. We additionally meta-analyzed 169 serum
51 metabolites from a NMR platform in 2,797 MZ and 2,040 DZ twins from the NTR, FTC and
52 the FinnTwin12 (FT12), and show no metabolomic differences between the MZ and DZ twins.
53 Overall, we identified novel transcriptomics biomarkers of twinning in peripheral blood and
54 provide partial converging evidence for multiple pathways previously identified in the GWAS
55 of DZ and EWAS of MZ twinning.

56

57 **Keywords:** Twinning, Twins, Transcriptomics, Metabolomics, Multi-omics

58

59 **Introduction**

60 Twinning is defined as the process that gives rise to either dizygotic (DZ) or monozygotic (MZ)
61 twins, and in rarer occasions to triplets and other higher-order multiples (van Dongen et al.,
62 2023). MZ and DZ twins are often studied to provide insight into the genetic and environmental
63 influences on complex traits through the classical twin design or other twin-family designs (D.
64 Boomsma et al., 2002; Hagenbeek et al., 2023). However, much remains unclear about the
65 etiology of MZ and DZ twinning. DZ twinning is the result of a spontaneous double ovulation,
66 runs in families and is often studied as a model for super fertility (Beck et al., 2021; D. I.
67 Boomsma, 2020). MZ twins arise after a fertilized egg cell splits, but the mechanisms behind
68 this are still largely a mystery (van Dongen et al., 2021, 2023).

69 Genome-wide association studies (GWAS) of DZ twinning highlighted multiple genes such as
70 *FSHB*, *FSHR* and *GNRH*, with obvious roles in female reproduction (Mbarek et al., 2016,
71 2024). An epigenome-wide association study (EWAS) highlighted DNA methylation
72 differences at over 800 sites in blood from MZ twins when compared to DZ twins and singletons
73 (van Dongen et al., 2021). The methylation differences in blood also replicated in buccal cells.
74 Still, mechanisms leading to the formation of twins, in particular MZ twins, are not fully
75 understood. Furthermore, it is unknown whether molecular markers of twinning might be found
76 in other omics layers that have not yet been studied in connection to twinning, including the
77 transcriptome and metabolome. Studying these layers may provide additional insights into the
78 etiology and can lead to the identification of biomarkers for MZ and DZ twinning (Johnson et
79 al., 2016; Soininen et al., 2015; Whipp et al., 2022).

80 Transcriptomics is the study of all RNA molecules from protein coding to noncoding RNA
81 (Thompson et al., 2016). Data from protein coding RNAs are sometimes integrated with GWAS
82 results to identify novel gene-trait associations (Yin et al., 2022). These so-called
83 transcriptome-wide association studies (TWASs) are usually performed as post-GWAS
84 analyses using databases such as GTEx and tools like SMR/HEIDI (Gamazon et al., 2015; Zhu
85 et al., 2016). In the largest DZ twinning GWAS, eight genes were identified using a TWAS
86 analysis in several female fertility related tissues: *ARL14EP*, *CAPRIN2*, *ZFPM1*, *SMAD3*,
87 *MPPED2-AS*, *GOLGA8T*, *PCBP2*, and *FAM66D*. Twins have also been used to study the
88 heritability of RNA sequence levels (Ouwens et al., 2019), but RNA transcripts have not been
89 used further to investigate the twinning etiologies.

90 Metabolomics is the study of the small molecules involved in cellular metabolism (Patti et al.,
91 2012). Unlike other omics, metabolomics provides tools to measure biochemical activity
92 directly, by monitoring the substrates and products involved in cellular metabolism. Studies
93 focusing on metabolomics are widely performed in search of biomarkers for physical and
94 mental health conditions, but have never been performed before for twinning (Guijas et al.,
95 2018).

96 In this study, we compare serum metabolomics and blood transcriptomics profiles between MZ
97 and DZ twins to enhance the discovery of molecular biomarkers of twinning. We include adult
98 participants from the Netherlands Twin Register (NTR) and from two Finnish twin cohorts,
99 FinnTwin12 (FT12) and the older Finnish Twin Cohort (FTC) of whom many have been
100 involved in one of the two previous omics studies into twinning (Kaprio et al., 2019; Ligthart
101 et al., 2019; Mbarek et al., 2024; Rose et al., 2019; van Dongen et al., 2021; Willemsen et al.,
102 2013). For the transcriptomics, we analyzed gene expression levels of 42,663 RNA transcripts
103 in 2,747 participants of the NTR and repeated our analysis in the 5% transcripts with the lowest
104 p-values in 375 participants of the FTC. For the metabolomics, the three cohorts had
105 measurements from the same platform allowing us to meta-analyze 169 metabolite levels from
106 4,837 DZ and MZ twins.

107

108 **Methods**

109 **Participants**

110 *NTR*

111 The NTR is a population based cohort that has been collecting data from twins and their families
112 since the 1980's (Ligthart et al., 2019). In 2004 the NTR started a large-scale biological sample
113 collection of nearly 10,000 participants to create a resource for future omics and biomarker
114 studies (Willemsen et al., 2010). During a home visit in the morning, eight tubes of fasting
115 blood and a morning urine sample were collected along with phenotypic information on health,
116 medication use, body composition and smoking. For fertile women samples were obtained, as
117 much as possible, on days 3–5 of their menstrual cycle or in the pill-free week if on oral
118 contraception. RNA data were generated with the Affymetrix U219 array for 3,362 participants
119 of the NTR, of whom 2,828 were twins. From these twins we excluded 81 individuals who were
120 pregnant or did not have complete covariate data leading to a total of 2,747 twins who we
121 included in our study (Table 1). Metabolomics data were collected for 4,227 participants of the
122 NTR of whom 3,638 were twins with complete phenotypic information that could be included
123 in this study. Zygosity of the participants was determined using genotype information, blood
124 group information or multiple survey items (Ligthart et al., 2019).

125 *FTC*

126 One subset of the Finnish Twin Cohort is the older Finnish Twin Cohort (FTC), which includes
127 twins born before 1958 (Wright et al., 2014). Four waves of questionnaires have been sent to
128 unselected members of the cohort in 1975, 1981, 1990 and 2011, respectively (Kaprio et al.,
129 2019). In 2015, a selected subset of these twins came in for measurement of their blood pressure,
130 completed interviews and questionnaires and provided a fasting blood sample for biochemical
131 measures, and samples for omics (Huang et al., 2018). Metabolomic data for 435 participants
132 were included in the study. Out of the 402 peripheral blood samples, high quality RNA was
133 obtained for 391 subjects and 375 twins of them had complete phenotypic data and could be
134 included in the current study (Table 1). Zygosity of the twins was determined by the first
135 questionnaires for twins and confirmed by genotyping.

136 *FT12*

137 FT12 is a population-based longitudinal cohort including twins born in Finland between 1983
138 and 1987 (Rose et al., 2019). FT12 was designed for baseline assessments at an early age

139 preceding onset of regular exposure to alcohol, tobacco or other substances (Rose et al., 2019).
140 Subsequent waves of follow-up in FT12 at ages 11-12, 14, 17, and 22 years, have created a rich
141 dataset of behavioral assessments from multiple sources across three stages of adolescence and
142 into early adulthood. The “age 22” assessment wave involved 1,347 twin individuals, with 779
143 individuals attending in-person assessments and thus venous blood plasma samples could be
144 collected (Table 1; (Whipp et al., 2022)). We removed 15 participants from the data due to
145 pregnancy or lipid lowering medication, resulting in a total sample of 764 participants from
146 FT12. Zygosity was determined using questionnaires at baseline and confirmed later by
147 genotyping. No transcriptomics data were available in the FT12.

148 **Omics data**

149 *Transcriptomics*

150 NTR

151 The generation of the gene expression arrays and data in the NTR was described before (Jansen
152 et al., 2014; Wright et al., 2014). Briefly, peripheral venous blood samples were drawn in the
153 morning (7:00—11:00 a.m.) after an overnight fast. Within 20 minutes of sampling, heparinized
154 whole blood was transferred into PAXgene Blood RNA tubes (Qiagen) and stored at -20°C.
155 The frozen PAXgene tubes were shipped to the Rutgers University and DNA Repository
156 (RUCDR, <http://www.rucdr.org>). RNA was extracted by the Qiagen Universal liquid handling
157 system, as per the manufacturer's protocol. RNA quality and quantity was assessed by Caliper
158 AMS90 with HT DNA5K/RNA LabChips. Samples were hybridized to Affymetrix U219 array
159 plates (GeneTitan) which contains 530,467 probes for 49,293 transcripts. Array hybridization,
160 washing, staining, and scanning were carried out in an Affymetrix GeneTitan System per the
161 manufacturer's protocol. Twin pairs were randomized over the sample plates. Expression data
162 were required to pass standard Affymetrix Expression Console quality metrics before further
163 undergoing quality control. Samples were excluded if they showed sex inconsistency. Log2
164 transformation and quantile normalization were applied to the gene expression data. In the main
165 analyses we excluded 1,578 transcripts located on the X and Y chromosome, leading to a total
166 of 42,663 RNA transcripts included in main analysis, 44,182 in the female-only analysis and
167 44,241 in the male-only analysis.

168 FTC

169 An in-depth description of the transcriptomics measurements of the transcriptomics data
170 collection can be found in Huang et al (2018) (Huang et al., 2018). RNA samples were extracted
171 from the peripheral leukocytes stored in the RNA cell protection reagents (QIAGEN Inc.
172 Valencia, CA, USA) using the miRNeasy Mini Kit (QIAGEN Inc. Valencia, CA, USA).
173 Transcriptome-wide gene expression data were obtained using the Illumina HumanHT-12 v4
174 Expression BeadChip (Illumina Inc) which targets more than 48,000 transcripts that provide
175 transcriptome-wide coverage of well-characterized genes, gene candidates and splice variants.
176 A block design was used to keep the distributions of sex and zygosity similar across chips (12
177 samples/chip) with the co-twins assigned to the same chip. The twin pairs were randomly
178 assigned as pairs to the 12 positions on each chip. Quantification was performed employing the
179 Genome-Studio Gene Expression Module and the lumi R package for data preprocessing and
180 quality control. The quality controls consisted of two key steps: 1; Transcripts with detection
181 P-value <0.05 in more than 50% of the samples were defined as “present”; 2; Log2
182 transformation and quantile normalization were applied to the gene expression data. There were
183 19,530 transcripts that passed the quality control steps and they were used as indices of gene
184 expression levels in further analyses. For the replication analyses, we selected the top 2,133
185 (5% lowest p-value fraction) RNA transcripts from the analyses in NTR and selected the same
186 transcript in the FTC data. We selected the probes based on genetic location and/or gene name.
187 A total of 1,325 transcripts shared overlap between the two platforms and were included in the
188 replication analyses.

189 The analysis of the transcriptomics data of the FTC cohort showed a high inflation ($\lambda =$
190 1.3) of the p-values (Supplementary figure 1). This is potentially due to the decision to retain
191 MZ and DZ twin pairs on the same RNA chips, as we observe a high correlation of 0.75 between
192 the RNA chips and age (Supplement Figure 2), which can introduce multi-collinearity (*10.4 -*
193 *Multicollinearity | STAT 462*, n.d.). To address this, we repeated our analyses of zygosity in
194 1,325 randomly selected RNA (excluding the initial 1,325 probes) transcripts in the FTC data.
195 When repeating the analyses to random RNA transcriptomics data, we do observe several
196 transcripts surpassing our significance threshold after multiple testing correction burden, but
197 when looking at the enrichment analyses we do not observe any pathways associated with MZ
198 or DZ twinning before (Supplementary Figures 3-4).

199 *Metabolomics*

200 In all three cohorts, metabolites were measured on the metabolomics platform from Nightingale
201 (Soininen et al., 2015). The panel included amino acids (alanine, glutamine, histidine,

202 isoleucine, leucine, phenylalanine, tyrosine, and valine) and ketone bodies (acetate,
203 acetoacetate, and 3-hydroxybutyrate), lipids and fatty acids. All metabolite data were available
204 in units mmol/L, nm, g/L or $\mu\text{mol/L}$ (Supplementary Table 1). The metabolomics data were
205 pre-processed for each of the three cohorts similarly, but separately.

206 In the NTR, the metabolomics data were collected in different batches, which were all pre-
207 processed separately (Hagenbeek et al., 2020; Pool et al., 2020). Metabolites were excluded
208 from the analysis when the mean coefficient of variation exceeded 25% and the missing rate
209 exceeded 5%. Metabolite measurements were set to missing if they were below the lower limit
210 of detection or quantification or could be classified as an outlier (five standard deviations
211 greater or smaller than the mean). If missing, the metabolite was imputed with half of the value
212 of this limit, or when this limit was unknown with half of the lowest observed level for this
213 metabolite.

214 In the FT12 and FTC, the metabolomics data were measured in one batch each. Metabolite
215 values below the limit of detection were reported as missing, as were metabolite values that
216 deviated from the mean by more than 5 standard deviations (SD). Metabolites with more than
217 10% missing values were discarded. Missing values were imputed in the FT12 and using the
218 sample minimum of each metabolite for which imputation was to be performed. Outlier
219 presence was assessed by verifying that no participant had at least one of its first three principal
220 components greater than or less than 5 SD from the mean.

221 In all three cohorts, metabolites were normalized by inverse normal rank transformation so that
222 the distribution of metabolites are of mean 0 and variance 1 (Demirkan et al., 2015; Kettunen
223 et al., 2016). In FT12, 119 metabolites could be included and in NTR and FTC all 169
224 metabolites were included (Supplementary Table 1 & 2).

225 **Statistical analyses**

226 *Transcriptomics*

227 We performed logistic regression for each of the 42,663 RNA transcripts in the NTR twin status
228 (MZ twins coded as= 1 and DZ twins coded as= 2) as a predictor. All analyses were performed
229 in R version 4.3.0. The models were run by using Generalized estimating equations (GEE) to
230 account for family structure (package = *geepack*; function = *geelm*). We included sex,
231 biological age at blood sampling, body mass index (BMI) at time of blood sampling, plate, well
232 and cell counts of blood cells including neutrophils, lymphocytes, monocytes, eosinophils and

233 basophiles, as covariates. Given that transcriptomics data are correlated (Jansen et al., 2014),
234 adjusting p-values to correct for multiple testing was based on the number of independent
235 dimensions in the data. We performed principal component analyses of the transcripts (Jolliffe
236 & Cadima, 2016), and corrected the transcriptomic analyses of twins for the number of effective
237 tests performed. 2,814 principal components captured 95% of the total variation in the RNA
238 transcript levels (Supplementary Table 3), leading to a correction where significance is reached
239 if a p-value is below $0.05/2,814 = 1.78E-05$. In addition, we ran sex stratified analyses without
240 the sex covariate and we performed sensitivity analyses of the female-only analyses in 669
241 females after excluding all participants taking contraceptives and post-menopausal women
242 (N=1,128). In the analyses of the females we additionally included the 1,519 transcripts on the
243 X chromosome and in the analyses of the males we included the X chromosome transcripts and
244 59 transcripts of the Y chromosome.

245 We took forward the 5% fraction of transcripts with the lowest p-value in NTR in the main
246 analyses (2,213 transcripts in 1,876 genes) and searched for matching transcripts in the FTC
247 data. We identified 1,325 transcripts from the Illumina chip for replication, based on the
248 location of the transcripts and the gene in which the transcript is located and included in both
249 platforms. Given that these transcripts have already been preselected, a more stringent
250 Bonferroni correction was used for replication in the FTC, leading to a p-value to be lower than
251 $0.05/1,325=3.77E-05$ to be significant.

252 *Metabolomics*

253 We performed logistic regression for each metabolite with twin zygosity as a predictor (MZ
254 and DZ twin, coded 1 and 2), for each cohort separately. The covariates for each cohort
255 comprise age, sex, smoking status, BMI, batch, and additionally only for NTR fasting status
256 (yes/no) and use of lipid lowering medication. All variables were measured at time of blood
257 sampling. All analyses were performed in R version 4.3.0. The models were run by using GEE
258 to account for family structure (package = *geepack*; function = *geelm*). For the meta-analyses
259 of NTR, FT12 and FTC, we applied fixed effect meta-analyses (package = *metafor*; function =
260 *rma*). We estimated the number of effective tests using the method of Matrix Spectral
261 Decomposition (Chen et al., 2022). In total, we included 169 metabolites which resulted in a
262 total of 27 effective tests, leading to a p-value to be lower than $0.05/27 = 1.85E-03$ to be
263 significant, based on Bonferroni correction.

264 **Enrichments analyses**

265 *Enrichr*

266 We performed enrichment analysis using Enrichr on the 5% most significant transcripts in NTR
267 and the 19 genes of the 21 significant transcripts in FTC. Enrichr is a gene set search engine
268 that enables the querying of hundreds of thousands of annotated gene sets (Xie et al., 2021).
269 Enrichr uniquely integrates knowledge from many high-profile projects to provide synthesized
270 information about mammalian genes and gene sets. We employed the *enrichR* R package as an
271 R interface to the Enrichr database (Kuleshov et al., 2016). We used the databases Gene
272 Ontology 2023 molecular functions database (GO2023) and KEGG pathway database (KEGG)
273 to look at the enrichment (Consortium et al., 2023; Kanehisa et al., 2023).

274 *BIOS consortium database*

275 We performed a look-up of the 5% most significant transcripts in the NTR data and the
276 significantly different expressed transcripts in FTC in the methylation–expression association
277 (expression quantitative trait methylation; eQTM) database of the BIOS consortium (Bonder et
278 al., 2017; Zhernakova et al., 2016). The eQTM database did not include participants from NTR.
279 The database contains information on associations between DNA methylation levels and gene
280 expression, based on RNA sequencing data from ~2000 whole blood samples. We checked
281 whether the identified RNA transcripts identified in the current study were located in genes that
282 showed overlap with the eQTM of the CpGs associated with MZ twinning from our earlier
283 EWAS study (van Dongen et al., 2021).

284

285 **Results**

286 *Transcriptomics*

287 We compared 42,663 RNA transcripts between 1,453 MZ and 1,291 DZ twins of the NTR
288 (Table 1). One RNA transcript, located at 11727318_at (chr8:30853321:30854213), was
289 significantly differentially expressed between MZ and DZ twins (higher expression in MZ
290 twins; beta = 1.78E-02; p value = 6.19E-06). The transcript maps to chromosome 8 located in
291 the *PURG* gene (Figure 1; Supplement Figure 5; Supplementary table 4). We performed
292 enrichment analyses in the GO2023 and the KEGG dataset on the genes in which 5% transcripts
293 with the lowest p-values are located (2,133 transcripts in 1,876; Figure 2; Supplementary table
294 5-6). In the GO2023 database, 56 pathways were significantly enriched in the genes and the
295 most strongly enriched pathway was "molecular function of the estrogen response element
296 binding". The KEGG database highlighted 24 significant enrichment of pathways involved in
297 breast cancer and cell adhesion elements, as well as the WNT-signaling pathway. We performed
298 a look-up of the 5% most significant transcripts in the BIOS consortium eQTM database and
299 identified that two transcripts (located in the genes *HTT* and *FANCC*) that were significantly
300 associated with DNA methylation of CpG sites in the genes previously found to be differentially
301 methylated in MZ twins ($P < 3e-08$; Supplementary table 7). Three methylation sites in the *HTT*
302 gene (cg01807241, cg02566259 and cg06421590) showed lower methylation in MZ twins
303 while in our result the *HTT* is higher expressed in the MZ twins. In line with this, in the BIOS
304 dataset, lower methylation correlated with higher expression. The methylation site in *FANCC*,
305 cg14127626, also showed less methylation in the BIOS dataset and more gene expression in
306 the MZ twins.

307 *Sex-stratified analyses*

308 We reran the analyses in females and males separately, including the sex chromosomes. In the
309 female-only analyses we included 1,008 MZ twins and 789 DZ twins and 44,182 RNA
310 transcripts. In these analyses the same transcript in *PURG*, 11727318_at
311 (chr8:30853321:30854213), showed significantly higher expression in MZ twins (beta = 2.50E-
312 02; p value = 2.65E-07). In addition, we found four more transcripts, 11745032_a_at
313 (ch4:114821440:114821440), 11736249_x_at (chr15:64657193:64658274), 11732371_at
314 (chr19:51479729:51480947) and 11759996_at (chr5:35960858:35963053) which significantly
315 associated with zygoty in the female-only analyses (Supplementary table 8). These transcripts
316 are of the genes *ARSL* (beta = -2.30E-02; p value = 1.04E-05), *KIAA0101* (beta = -3.19E-02;

317 p value = 1.26E-05), *KLK7* (beta = -2.43E-02; p value = 1.42E-05) and *UGT3A1* (beta = -3.81E-
318 02; p value = 1.73E-05), respectively. We further performed enrichment analyses in the
319 GO2023 and the KEGG dataset on the genes in which the 5% transcripts with the lowest p-
320 value are located (2,009 transcripts; Figure 3; Supplementary tables 9-10). In total 44 pathways
321 were enriched in the KEGG dataset and 58 function in GO2023. The strongest enriched
322 pathways were the DNA binding Transcription Activator Activity, RNA Polymerase II-specific
323 and pathways in cancer.

324 We performed a sensitivity analysis in the NTR female-only sample after removing all females
325 that took contraceptives or indicated to be past menopause (N=1,128) at the time of the
326 biological sampling (new N=669) In this analysis, six genes in the SMAD pathway (*SMAD2*,
327 *SMAD4*, *TGFB1*, *SMAD5*, *TGFBRI*, *SMAD7*), previously indicated for DZ twinning, were
328 downregulated in the female DZ twins based on a nominal significance (p<0.05). In addition,
329 the R-SMAD pathway was the most significantly enriched (Supplementary tables 16-19),
330 which was also among the enriched pathways in the female-only analyses (Supplementary table
331 10).

332 In the male-only analyses we included 445 MZ twins and 505 DZ twins and 44,241 RNA
333 transcripts, including both sex chromosomes. In this analysis the transcript in *PURG* was not
334 significantly different in MZ and DZ twins (beta = 7.00E-03; p value = 0.260). However, we
335 observed three other transcripts, 11739295_a_at (chr21:45651163: 45655445), 11738638_a_at
336 (chr20:2397878:2413399) and 11726545_x_at (chr12: 100550135: 100550836), that were
337 significantly associated with zygosity in the male-only analyses (Supplementary table 11).
338 These transcripts are located respectively in the *ICOSLG* (beta = 3.93E-02; p value = 4.08E-
339 07), *TGM6* (beta = 3.22E-02; p value = 7.29E-06) and *GOLGA2P5* (beta = 3.43E-02; p value
340 = 1.50E-05) genes. We further performed enrichment analyses in the GO2023 and the KEGG
341 dataset on the genes in which the 5% most significant transcripts are located (2,112 transcripts;
342 Figure 4; Supplementary Tables 12-13). In total 21 pathways were enriched in the KEGG
343 dataset and 55 functions in GO2023. The most strongly enriched pathways were the Ankyrin
344 Binding pathway and the Thyroid hormone signaling pathway.

345 *Replication analyses*

346 Replication analyses were carried out for 1,325 transcripts of the 2,133 (5%) most significant
347 ones in the results in NTR in 375 twins of the FTC cohort (Table 1). We observed that 21
348 transcripts, from 19 genes, were differentially expressed in the FTC (Supplementary table 14).

349 Eight of these genes showed lower expression levels in the MZ twins while the other 11 showed
350 higher expression in the MZ twins. We performed enrichment analyses in the GO2023 and
351 KEGG databases on the 19 genes (Figure 5). The GO2023 indicated 10 significant enrichments
352 including the cAMP regulatory function and the R-SMAD pathway (Supplementary table 15).
353 The TGF-B pathway was also enriched in the KEGG database, together with the cell cycle, the
354 WNT-pathway and 7 other pathways (Supplementary table 16). A look-up of the 19 genes of
355 the 21 significant transcripts in the BIOS consortium eQTM database did not reveal associations
356 with CpGs that were previously associated with MZ twinning. The NTR findings for *PURG*
357 were not replicated. For the FTC cohort the most significant transcript (ILMN_7746) lies in the
358 *JUN* gene (beta = -1.165; $p < 1E-10$). Of the 19 genes, 7 (37%) showed a replication of the
359 same direction of effect in the NTR (Table 2).

360 ***Metabolomics***

361 We meta-analyzed 169 metabolites in 2,797 MZ twins and 2,040 DZ from the NTR, FT12 and
362 the FTC (Table 1) with a fixed effect model. The results of the individual analyses of each
363 cohort and the meta-analyses are summarized in supplementary table 2. The results show that
364 none of the metabolite levels differed significantly between the MZ and DZ twins after multiple
365 testing correction ($\alpha=2.96E-04$). Five metabolites showed differences between the MZ and
366 DZ twins, on the nominal p-value of 0.05. The levels of glycoprotein acetyls, linoleic acid,
367 omega-6 fatty acids, albumin and phospholipids in medium LDL were lower in the MZ twins.
368 In three out of these five metabolites the effects of the NTR and the FT12 show the same
369 direction, while the FTC shows the opposite effect (Supplementary figure 6). We found that the
370 overall correlations in the betas between the three cohorts were weak with a range from -0.29
371 to 0.24 (Supplementary Figure 6).

372

373

374 Discussion

375 In this study we compared transcriptomic and metabolomic data between MZ and DZ twins of
376 the NTR, FT12 and the FTC. We show differences in the RNA transcript levels between MZ
377 and DZ twins in genes whose relations to twinning have partially been established through
378 other, previous, omics studies (GWAS, and EWAS) and show that the differences between the
379 transcriptome in MZ and DZ twins depend on sex. We find some overlap in the differentially
380 expressed gene transcripts with the monozygotic twinning epigenetic signature and the
381 enrichment analyses point to pathways, such as cell adhesion, the WNT-pathways and the TGF-
382 B pathway, that have been indicated before with either MZ or DZ twinning (Mbarek et al.,
383 2024; van Dongen et al., 2021).

384 We found one transcript located in the *PURG* gene to be significantly associated with zygosity
385 in the total NTR sample. When dividing into males and females, we identified an additional
386 seven transcripts (three in males and four in females). None of these eight genes were replicated
387 in the Finnish samples (which contained both men and women), but eight other transcripts
388 located in these seven genes showed nominal significant association with zygosity in both
389 cohorts. These seven genes could potentially serve as biomarkers of DZ or MZ twinning. In the
390 females-only analyses, but not in the males-only analyses, we confirm the finding of *PURG*
391 indicating a sex-specific twinning maker. *PURG* is a gene that codes for Purine Rich Element
392 Binding Protein G, and the direct function of the gene is unknown. It is however highly similar
393 to purine-rich element binding protein A, which is a DNA-binding protein which has been
394 implicated in the control of both DNA replication and transcription (Liu & Johnson, 2002).
395 *PURG* is highly expressed in the brain, endometrium and testis and, in all analyses, showed
396 higher expression in the MZ twins (*PURG*; Gene - NCBI, n.d.). In the FinnGen GWAS study,
397 *PURG* was also associated with hypertension, BMI and pregnancy related conditions including
398 poor fetal growth (Kurki et al., 2023). Of note, previous studies have shown that mothers of
399 dizygotic twins have a higher BMI compared to mothers of monozygotic twins (Hubers et al
400 TRHG in press).

401 The four extra genes identified in the females-only analyses all showed increased expression in
402 the DZ twins and all show some relation to female reproduction. *ARSJ* and *UGT3A1* have a
403 direct link to female fertility as both are involved in the synthesis of hormones (Mackenzie et
404 al., 2008). *ARSJ* is also associated with height in the FinnGen study, possible reflecting the
405 increased body composition of mothers of DZ twins (Kurki et al., 2023; van Dongen et al.,
406 2023). *KIAAA0101* has been associated before with ovarian cancer and *KLK7* has been

407 implicated with litter size in pigs and encodes for a member of the kallikrein subfamily of serine
408 proteases, which have diverse physiological functions and many kallikrein genes are
409 biomarkers for cancer. (Dan & Yonggang, 2020; Jin et al., 2018; Kurki et al., 2023). DZ
410 twinning is often seen as super fertility and these findings further highlight the relation between
411 DZ twinning and fertility. It would be interesting to investigate if these genes are also more
412 upregulated in mothers of DZ twins and if the differences in gene expression are also present
413 around ovulation.

414 When comparing the results with the TWAS results from the latest GWAS we find *PCBP2* to
415 be higher expressed in MZ twins in the 5% transcripts with lowest p-values of the NTR results
416 (Mbarek et al., 2024). The TWAS analyses were performed in eight tissues mostly relevant to
417 female fertility; breast, hypothalamus, ovary, pituitary gland, testis, uterus, vagina, and whole
418 blood tissue, while we are only focusing on RNA transcript levels in blood (Mbarek et al.,
419 2024). In the TWAS, the whole blood data showed the lowest amount of genes with an
420 association with DZ twinning, hinting that blood probably is not the most relevant tissue to
421 study. Obtaining RNA transcript data from any other tissue at such a large scale is quite
422 challenging and the previously detected DNA methylation signature was observed in both blood
423 and buccal samples from MZ twins (van Dongen et al., 2021). In the current study, we show
424 that differential DNA methylation in monozygotic twins is accompanied by a difference in
425 transcript levels at two genes in blood. Still, investigating the RNA transcript levels of other
426 tissues has potential in understanding the twinning traits.

427 *HTT* and *FANCC*, are indicated both in the 5% transcripts with lowest p-values RNA transcripts
428 and in the epigenetic MZ twinning signature (Bonder et al., 2017; van Dongen et al., 2021;
429 Zhernakova et al., 2016). The former gene has been associated with Huntington's diseases and
430 the latter with a process that is activated when DNA replication is blocked called the Fanconi
431 anemia pathway (Gordon & Buchwald, 2013; Jain & Roy, 2023). The directions of effects
432 appear reversed in the two studies, implying that the increased methylation of these genes found
433 by van Dongen et al. (2021) correlated with a lower expression of these genes in the MZ twins
434 as shown in our study (van Dongen et al., 2021).

435 Additionally, we observe enrichment of the associated RNA transcripts in cell adhesion
436 processes and the WNT-signaling pathways. Both were also enriched in the epigenetic signature
437 and our results further underline the hypothesis that these processes may therefore be important
438 in the etiology of MZ twinning (van Dongen et al., 2021) The WNT-signaling pathway may
439 also be important for DZ twinning as *SMAD3* has been hypothesized to influence the WNT-

440 signaling through an interaction with β -catenin (Funa et al., 2015; Hernandez Gifford, 2015).
441 *SMAD3* is the second top hit of the DZ twinning GWAS and also plays a central role in the
442 TGF-B pathway as it is needed to translocate transcription factors to the cell nucleus (Hata &
443 Chen, 2016; Mbarek et al., 2024). In our transcriptomics analysis, two genes in the TGF-B
444 pathways pathway, *CDKN2B* and *SKP1*, were respectively lower and higher expressed in DZ
445 twins of the FTC. These genes have not been previously indicated for any type of twinning, but
446 the TGF-B pathway has been implicated for DZ twinning before (Mbarek et al., 2024).

447 As for metabolites, we do not observe differences between MZ and DZ twins after correcting
448 for multiple testing. We show that the three cohorts did not show strong correlations between
449 coefficients, and even showed a negative correlation between the NTR and FTC, based on
450 analyses in the total sample of males and females combined. The negative correlation could
451 have been caused by menopause or contraceptive effects, which are known to influence
452 metabolite levels (Bot et al., 2020; Hagenbeek et al., 2020) or the small sample of the FTC
453 cohort. Furthermore, since dizygotic twinning depends on a double ovulation, a molecular
454 signature of dizygotic twinning may disappear when ovulation stops beyond the female
455 reproductive age or through the use of contraceptives. In the transcriptomics data, we performed
456 a sensitivity analysis in the NTR female-only sample by removing all females that took
457 contraceptives or indicated to be in menopause at the time of the biological sampling, and saw
458 that the signal became stronger for genes in the SMAD pathway implicated in dizygotic
459 twinning.

460 A final consideration of our study is that we do not have a non-twin control group. Therefore,
461 for the transcriptome differences we observe between MZ and DZ twins, it is not necessarily
462 evident whether they are connected to MZ or DZ twinning. However, some pathways clearly
463 show overlap with previous DZ twinning GWAS results and thus are more likely driven by
464 differential expression in DZ twins, whereas other transcriptomic results converge with the
465 DNA methylation signature for MZ twinning (Mbarek et al., 2024; van Dongen et al., 2021).

466 To conclude, our study identifies novel transcriptomic biomarkers for twinning and provides
467 converging evidence for multiple pathways previously identified in the GWAS of DZ and
468 EWAS of MZ twinning. More functional studies on twinning are to be encouraged, and
469 pursuing efforts to include other omics (e.g. proteomics) and other tissues (e.g. ovary, and pre-
470 and perinatal tissues) may pave the way to a better understanding of how these biological
471 pathways are involved in twinning.

472 **Declarations**

473 *Ethics approval*

474 The study protocol was approved for the NTR by the Ethical Review Board of the VU
475 University Medical Center and informed consent was obtained from all participants.

476 Ethical approval for all data collection waves from FT12 was obtained from the ethical
477 committee of the Helsinki and Uusimaa University Hospital District and the Institutional
478 Review Board of Indiana University. All data collection and sampling protocols were
479 performed in compliance with the ethical guidelines. Parents provided consent for the twins
480 aged 12 and 14 years old, while twins aged 17 and 22 years old provided written consent
481 themselves for sample collection.

482 The study protocol for the FTC was approved by the Institutional Ethics Board of the Hospital
483 District of Helsinki and Uusimaa, Finland (ID 154/13/03/00/11) and the Institutional Review
484 Board of Augusta University.

485 *Availability of data and materials*

486 The data of the Netherlands Twin Register (NTR) may be requested through the NTR data
487 access committee (https://tweelingenregister.vu.nl/information_for_researchers/working-with-ntr-data).
488 ntr-data).

489 The Finnish Twin Cohort data used in the analysis is deposited in the Biobank of the Finnish
490 Institute for Health and Welfare (<https://thl.fi/en/web/thl-biobank/forresearchers>). It is available
491 to researchers after written application and following the relevant Finnish legislation.

492 FinnTwin12 data analyzed in this study is not publicly available due to the restrictions of
493 informed consent. Requests to access these datasets should be directed to the Institute for
494 Molecular Medicine Finland (FIMM) Data Access Committee (DAC) (fimmdac@helsinki.fi)
495 for authorized researchers who have IRB/ethics approval and an institutionally approved study
496 plan. To ensure the protection of privacy and compliance with national data protection
497 legislation, a data use/transfer agreement is needed, the content and specific clauses of which
498 will depend on the nature of the requested data.

499 *Competing interest*

500 None of the authors have competing interests.

501 *Funding*

502 NH is supported by the Royal Netherlands Academy of Science Professor Award (PAH/6635)
503 to DIB and the 2023 talent travel grand award by the faculty of behaviour and movement
504 sciences at the Vrije Universiteit Amsterdam. JvD is supported by NWO Large Scale
505 infrastructures, X-omics (184.034.019). The NTR is supported by multiple grants from the
506 Netherlands Organizations for Scientific Research (NOW; 480-15-001/674; 480-04-004; 400-
507 05-717); and Medical Research (ZonMW; 912-10-020); the European Science Council (ERC)
508 Genetics of Mental Illness (ERC Advanced, 230374); Developmental trajectories of
509 psychopathology (NIMH 1RC2 MH089995). The metabolomics data of the NTR was collected
510 through BBMRI: BBMRI.1 (184.021.007) and BBMRI.2 (184.033.111).

511 The FTC has been supported by the Academy of Finland (Grants 265240, 263278, 308248,
512 312073, 336832 to Jaakko Kaprio and 297908 to Miina Ollikainen) and the Sigrid Juselius
513 Foundation (to Miina Ollikainen). The transcriptome study in FTC was supported by
514 NIH/NHLBI grant HL104125. Phenotype and genotype data collection in FinnTwin12 cohort
515 has been supported by, ENGAGE – European Network for Genetic and Genomic
516 Epidemiology, FP7-HEALTH-F4-2007, grant agreement number 201413, National Institute of
517 Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to R J Rose;
518 AA15416 and K02AA018755 to D M Dick; R01AA015416 to Jessica Salvatore) and the
519 Academy of Finland (grants 100,499, 205,585, 118,555, 141,054, 264,146, 308,248 to JK, and
520 the Centre of Excellence in Complex Disease Genetics (grants 312,073, 336,823, and 352,792
521 to JKaprio).

522

523 *Authors contributions*

524 NH, GD, RP and JvD have contributed to the conceptualization of the study. NH, GD, RJ and
525 RP prepared the data for inclusion. NH has performed the analyses and was responsible for the
526 first draft of the manuscript. RP, JvD, JJH, JK, GW and DIB provided supervision of the
527 analyses and writing. GW and DIB were involved in the initial data collection of the NTR and
528 MO and XW in the data collection of the FTC and FT12. All authors provided feedback and
529 revision on the manuscript and supported the submitted version of the manuscript.

530 *Acknowledgement*

531 This work largely builds upon the biobanking efforts performed by the BIOS and the BBMRI
532 consortia and we thank all participants and staff for their contributions. Additionally we thank
533 Dr. Fiona Hagenbeek for providing expertise and feedback throughout the project.

534

535 **References**

- 536 *10.4 - Multicollinearity | STAT 462*. (n.d.). Retrieved January 29, 2024, from
537 <https://online.stat.psu.edu/stat462/node/177/>
- 538 Beck, J. J., Bruins, S., Mbarek, H., Davies, G. E., & Boomsma, D. I. (2021). Biology and Genetics of
539 Dizygotic and Monozygotic Twinning. *Twin and Higher-Order Pregnancies*, 31–50.
540 https://doi.org/10.1007/978-3-030-47652-6_3/FIGURES/6
- 541 Bonder, M. J., Luijk, R., Zhernakova, D. V., Moed, M., Deelen, P., Vermaat, M., Van Itersson, M., Van
542 Dijk, F., Van Galen, M., Bot, J., Slieker, R. C., Jhamai, P. M., Verbiest, M., Suchiman, H. E. D.,
543 Verkerk, M., Van Der Breggen, R., Van Rooij, J., Lakenberg, N., Arindrarto, W., ... Heijmans, B. T.
544 (2017). Disease variants alter transcription factor levels and methylation of their binding sites.
545 *Nature Genetics*, 49(1), 131–138. <https://doi.org/10.1038/NG.3721>
- 546 Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews*.
547 *Genetics*, 3(11), 872–882. <https://doi.org/10.1038/NRG932>
- 548 Boomsma, D. I. (2020). The Genetics of Human DZ Twinning. *Twin Research and Human Genetics*,
549 23(2), 74–76. <https://doi.org/10.1017/THG.2020.15>
- 550 Bot, M., Milaneschi, Y., Al-Shehri, T., Amin, N., Garmaeva, S., Onderwater, G. L. J., Pool, R., Thesing, C.
551 S., Vijfhuizen, L. S., Vogelzangs, N., Arts, I. C. W., Demirkan, A., van Duijn, C., van Greevenbroek,
552 M., van der Kallen, C. J. H., Köhler, S., Ligthart, L., van den Maagdenberg, A. M. J. M., Mook-
553 Kanamori, D. O., ... Sattar, N. (2020). Metabolomics Profile in Depression: A Pooled Analysis of
554 230 Metabolic Markers in 5283 Cases With Depression and 10,145 Controls. *Biological*
555 *Psychiatry*, 87(5), 409–418. <https://doi.org/10.1016/J.BIOPSYCH.2019.08.016>
- 556 Chen, Y., Li, E. M., & Xu, L. Y. (2022). Guide to Metabolomics Analysis: A Bioinformatics Workflow.
557 *Metabolites*, 12(4). <https://doi.org/10.3390/METABO12040357>
- 558 Consortium, T. G. O., Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D.,
559 Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J.,
560 Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., ... Westerfield, M. (2023).
561 The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1).
562 <https://doi.org/10.1093/GENETICS/IYAD031>
- 563 Dan, L., & Yonggang, L. (2020). Molecular characteristics and association analysis with litter size trait
564 for porcine KLK7 gene. *Animal Biotechnology*, 31(5), 377–381.
565 <https://doi.org/10.1080/10495398.2019.1604379>
- 566 Demirkan, A., Henneman, P., Verhoeven, A., Dharuri, H., Amin, N., van Klinken, J. B., Karszen, L. C., de
567 Vries, B., Meissner, A., Göraler, S., van den Maagdenberg, A. M. J. M., Deelder, A. M., C't Hoen,
568 P. A., van Duijn, C. M., & van Dijk, K. W. (2015). Insight in Genome-Wide Association of
569 Metabolite Quantitative Traits by Exome Sequence Analyses. *PLoS Genetics*, 11(1), 1004835.
570 <https://doi.org/10.1371/JOURNAL.PGEN.1004835>
- 571 Funa, N. S., Schachter, K. A., Lerdrup, M., Ekberg, J., Hess, K., Dietrich, N., Honoré, C., Hansen, K., &
572 Semb, H. (2015). β -Catenin Regulates Primitive Streak Induction through Collaborative
573 Interactions with SMAD2/SMAD3 and OCT4. *Cell Stem Cell*, 16(6), 639–652.
574 <https://doi.org/10.1016/J.STEM.2015.03.008>

- 575 Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., Eyler,
576 A. E., Denny, J. C., Nicolae, D. L., Cox, N. J., & Im, H. K. (2015). A gene-based association method
577 for mapping traits using reference transcriptome data. *Nature Genetics*, *47*(9), 1091–1098.
578 <https://doi.org/10.1038/NG.3367>
- 579 Gordon, S. M., & Buchwald, M. (2013). *The FANCC Gene and Its Products*.
580 <https://www.ncbi.nlm.nih.gov/books/NBK6419/>
- 581 Guijas, C., Montenegro-Burke, J. R., Warth, B., Spilker, M. E., & Siuzdak, G. (2018). Metabolomics
582 activity screening for identifying metabolites that modulate phenotype. *Nature Biotechnology*,
583 *36*(4), 316. <https://doi.org/10.1038/NBT.4101>
- 584 Hagenbeek, F. A., Hirzinger, J. S., Breunig, S., Bruins, S., Kuznetsov, D. V., Schut, K., Odintsova, V. V., &
585 Boomsma, D. I. (2023). Maximizing the value of twin studies in health and behaviour. *Nature*
586 *Human Behaviour*, *7*(6), 849–860. <https://doi.org/10.1038/S41562-023-01609-6>
- 587 Hagenbeek, F. A., Pool, R., van Dongen, J., Draisma, H. H. M., Jan Hottenga, J., Willemsen, G.,
588 Abdellaoui, A., Fedko, I. O., den Braber, A., Visser, P. J., de Geus, E. J. C. N., Willems van Dijk, K.,
589 Verhoeven, A., Suchiman, H. E., Beekman, M., Slagboom, P. E., van Duijn, C. M., Barkey Wolf, J.
590 J. H., Cats, D., ... Boomsma, D. I. (2020). Heritability estimates for 361 blood metabolites across
591 40 genome-wide association studies. *Nature Communications*, *11*(1).
592 <https://doi.org/10.1038/S41467-019-13770-6>
- 593 Hata, A., & Chen, Y. G. (2016). TGF- β Signaling from Receptors to Smads. *Cold Spring Harbor*
594 *Perspectives in Biology*, *8*(9). <https://doi.org/10.1101/CSHPERSPECT.A022061>
- 595 Hernandez Gifford, J. A. (2015). The role of WNT signaling in adult ovarian folliculogenesis.
596 *Reproduction (Cambridge, England)*, *150*(4), R137. <https://doi.org/10.1530/REP-14-0685>
- 597 Huang, Y., Ollikainen, M., Sipil, P., Mustelin, L., Wang, X., Su, S., Huan, T., Levy, D., Wilson, J., Snieder,
598 H., Kaprio, J., & Wang, X. (2018). Genetic and environmental effects on gene expression
599 signatures of blood pressure: a transcriptome-wide twin study. *Hypertension (Dallas, Tex. :*
600 *1979)*, *71*(3), 457. <https://doi.org/10.1161/HYPERTENSIONAHA.117.10527>
- 601 Jain, S., & Roy, I. (2023). Aptamer Reduces Aggregation of Mutant Huntingtin and Rescues
602 Proteostasis Network in Non-Neuronal and Neuronal Cells. *ACS Chemical Neuroscience*, *14*(12),
603 2385–2395. <https://doi.org/10.1021/ACSCHEMNEURO.3C00226>
- 604 Jansen, R., Batista, S., Brooks, A. I., Tischfield, J. A., Willemsen, G., Van Grootheest, G., Hottenga, J. J.,
605 Milaneschi, Y., Mbarek, H., Madar, V., Peyrot, W., Vink, J. M., Verweij, C. L., de Geus, E. J. C.,
606 Smit, J. H., Wright, F. A., Sullivan, P. F., Boomsma, D. I., & Penninx, B. W. J. H. (2014). Sex
607 differences in the human peripheral blood transcriptome. *BMC Genomics*, *15*(1), 1–12.
608 <https://doi.org/10.1186/1471-2164-15-33/FIGURES/3>
- 609 Jin, C., Liu, Z., Li, Y., Bu, H., Wang, Y., Xu, Y., Qiu, C., Yan, S., Yuan, C., Li, R., Diao, N., Zhang, Z., Wang,
610 X., Liu, L., & Kong, B. (2018). PCNA-associated factor P15PAF, targeted by FOXM1, predicts poor
611 prognosis in high-grade serous ovarian cancer patients. *International Journal of Cancer*, *143*(11),
612 2973–2984. <https://doi.org/10.1002/IJC.31800>
- 613 Johnson, C. H., Ivanisevic, J., & Siuzdak, G. (2016). Metabolomics: beyond biomarkers and towards
614 mechanisms. *Nature Reviews. Molecular Cell Biology*, *17*(7), 451.
615 <https://doi.org/10.1038/NRM.2016.25>

- 616 Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments.
617 *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*,
618 374(2065). <https://doi.org/10.1098/RSTA.2015.0202>
- 619 Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for
620 taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, 51(D1), D587–
621 D592. <https://doi.org/10.1093/NAR/GKAC963>
- 622 Kaprio, J., Bollepalli, S., Buchwald, J., Iso-Markku, P., Korhonen, T., Kovanen, V., Kujala, U.,
623 Laakkonen, E. K., Latvala, A., Leskinen, T., Lindgren, N., Ollikainen, M., Piirtola, M., Rantanen, T.,
624 Rinne, J., Rose, R. J., Sillanpää, E., Silventoinen, K., Sipilä, S., ... Waller, K. (2019). The Older
625 Finnish Twin Cohort - 45 Years of Follow-up. *Twin Research and Human Genetics : The Official
626 Journal of the International Society for Twin Studies*, 22(4), 240–254.
627 <https://doi.org/10.1017/THG.2019.54>
- 628 Kettunen, J., Demirkan, A., Würtz, P., Draisma, H. H. M., Haller, T., Rawal, R., Vaarhorst, A., Kangas, A.
629 J., Lyytikäinen, L. P., Pirinen, M., Pool, R., Sarin, A. P., Soininen, P., Tukiainen, T., Wang, Q.,
630 Tiainen, M., Tynkkynen, T., Amin, N., Zeller, T., ... Ala-Korpela, M. (2016). Genome-wide study
631 for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nature
632 Communications*, 7. <https://doi.org/10.1038/NCOMMS11122>
- 633 Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S.,
634 Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G.
635 W., & Maayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server
636 2016 update. *Nucleic Acids Research*, 44(W1), W90–W97.
637 <https://doi.org/10.1093/NAR/GKW377>
- 638 Kurki, M. I., Karjalainen, J., Palta, P., Sipilä, T. P., Kristiansson, K., Donner, K. M., Reeve, M. P.,
639 Laivuori, H., Aavikko, M., Kaunisto, M. A., Loukola, A., Lahtela, E., Mattsson, H., Laiho, P., Della
640 Briotta Parolo, P., Lehisto, A. A., Kanai, M., Mars, N., Rämö, J., ... Palotie, A. (2023). FinnGen
641 provides genetic insights from a well-phenotyped isolated population. *Nature* 2023 613:7944,
642 613(7944), 508–518. <https://doi.org/10.1038/s41586-022-05473-8>
- 643 Ligthart, L., Van Beijsterveldt, C. E. M., Kevenaar, S. T., De Zeeuw, E., Van Bergen, E., Bruins, S., Pool,
644 R., Helmer, Q., Van Dongen, J., Hottenga, J. J., Van'T Ent, D., Dolan, C. V., Davies, G. E., Ehli, E.
645 A., Bartels, M., Willemsen, G., De Geus, E. J. C., & Boomsma, D. I. (2019). The Netherlands Twin
646 Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Research and
647 Human Genetics : The Official Journal of the International Society for Twin Studies*, 22(6), 623–
648 636. <https://doi.org/10.1017/THG.2019.93>
- 649 Liu, H., & Johnson, E. M. (2002). Distinct proteins encoded by alternative transcripts of the PURG
650 gene, located contrapodal to WRN on chromosome 8, determined by differential
651 termination/polyadenylation. *Nucleic Acids Research*, 30(11), 2417–2426.
652 <https://doi.org/10.1093/NAR/30.11.2417>
- 653 Mackenzie, P. I., Rogers, A., Treloar, J., Jorgensen, B. R., Miners, J. O., & Meech, R. (2008).
654 Identification of UDP Glycosyltransferase 3A1 as a UDP N-Acetylglucosaminyltransferase.
655 *Journal of Biological Chemistry*, 283(52), 36205–36210.
656 <https://doi.org/10.1074/JBC.M807961200>
- 657 Mbarek, H., Gordon, S. D., Duffy, D. L., Hubers, N., Mortlock, S., Beck, J. J., Hottenga, J.-J., Pool, R.,
658 Dolan, C. V., Actkins, K. V., Gerring, Z. F., Van Dongen, J., Ehli, E. A., Iacono, W. G., McGue, M.,

- 659 Chasman, D. I., Gallagher, C. S., Schilit, S. L. P., Morton, C. C., ... Martin, N. G. (2024). Genome-
660 wide association study meta-analysis of dizygotic twinning illuminates genetic regulation of
661 female fecundity. *Human Reproduction*, *39*(1), 240–257.
662 <https://doi.org/10.1093/HUMREP/DEAD247>
- 663 Mbarek, H., Steinberg, S., Nyholt, D. R., Gordon, S. D., Miller, M. B., McRae, A. F., Hottenga, J. J., Day,
664 F. R., Willemsen, G., De Geus, E. J., Davies, G. E., Martin, H. C., Penninx, B. W., Jansen, R.,
665 McAloney, K., Vink, J. M., Kaprio, J., Plomin, R., Spector, T. D., ... Boomsma, D. I. (2016).
666 Identification of Common Genetic Variants Influencing Spontaneous Dizygotic Twinning and
667 Female Fertility. *American Journal of Human Genetics*, *98*(5), 898–908.
668 <https://doi.org/10.1016/J.AJHG.2016.03.008>
- 669 Ouwens, K. G., Jansen, R., Nivard, M. G., van Dongen, J., Frieser, M. J., Hottenga, J. J., Arindrarto, W.,
670 Claringbould, A., van Ijzerman, M., Mei, H., Franke, L., Heijmans, B. T., A. C. 't Hoen, P., van
671 Meurs, J., Brooks, A. I., Heijmans, B. T., A. C. 't Hoen, P., van Meurs, J., Isaacs, A., ... Penninx, B.
672 W. J. H. (2019). A characterization of cis- and trans-heritability of RNA-Seq-based gene
673 expression. *European Journal of Human Genetics* *2019 28:2*, *28*(2), 253–263.
674 <https://doi.org/10.1038/s41431-019-0511-5>
- 675 Patti, G. J., Yanes, O., & Siuzdak, G. (2012). Metabolomics: the apogee of the omic trilogy. *Nature*
676 *Reviews. Molecular Cell Biology*, *13*(4), 263. <https://doi.org/10.1038/NRM3314>
- 677 Pool, R., Hagenbeek, F. A., Hendriks, A. M., van Dongen, J., Willemsen, G., de Geus, E., Metabolomics
678 Consortium, B., Willems van Dijk, K., Verhoeven, A., Eka Suchiman, H., Beekman, M., Eline
679 Slagboom, P., Harms, A. C., Hankemeier, T., Boomsma, D. I., D Suchiman, H. E., Amin, N.,
680 Beulens, J. W., van der Bom, J. A., ... Slagboom, P. (2020). Genetics and Not Shared Environment
681 Explains Familial Resemblance in Adult Metabolomics Data. *Twin Research and Human*
682 *Genetics*, *23*(3), 145–155. <https://doi.org/10.1017/THG.2020.53>
- 683 *PURG purine rich element binding protein G [Homo sapiens (human)] - Gene - NCBI*. (n.d.). Retrieved
684 May 15, 2024, from
685 <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=29942>
- 686 Rose, R. J., Salvatore, J. E., Aaltonen, S., Barr, P. B., Bogl, L. H., Byers, H. A., Heikkilä, K., Korhonen, T.,
687 Latvala, A., Palviainen, T., Ranjit, A., Whipp, A. M., Pulkkinen, L., Dick, D. M., & Kaprio, J. (2019).
688 FinnTwin12 Cohort: An Updated Review. *Twin Research and Human Genetics : The Official*
689 *Journal of the International Society for Twin Studies*, *22*(5), 302.
690 <https://doi.org/10.1017/THG.2019.83>
- 691 Soininen, P., Kangas, A. J., Würtz, P., Suna, T., & Ala-Korpela, M. (2015). Quantitative serum nuclear
692 magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circulation*.
693 *Cardiovascular Genetics*, *8*(1), 192–206. <https://doi.org/10.1161/CIRCGENETICS.114.000216>
- 694 Thompson, S. D., Prahalad, S., & Colbert, R. A. (2016). Integrative Genomics. *Textbook of Pediatric*
695 *Rheumatology*, 43-53.e3. <https://doi.org/10.1016/B978-0-323-24145-8.00005-3>
- 696 van Dongen, J., Gordon, S. D., McRae, A. F., Odintsova, V. V., Mbarek, H., Breeze, C. E., Sugden, K.,
697 Lundgren, S., Castillo-Fernandez, J. E., Hannon, E., Moffitt, T. E., Hagenbeek, F. A., van
698 Beijsterveldt, C. E. M., Jan Hottenga, J., Tsai, P. C., van Dongen, J., Hottenga, J. J., McRae, A. F.,
699 Sugden, K., ... Boomsma, D. I. (2021). Identical twins carry a persistent epigenetic signature of
700 early genome programming. *Nature Communications* *2021 12:1*, *12*(1), 1–14.
701 <https://doi.org/10.1038/s41467-021-25583-7>

- 702 van Dongen, J., Hubers, N., & Boomsma, D. I. (2023). New insights into the (epi)genetics of twinning.
703 *Human Reproduction, 0*, 1–8. <https://doi.org/10.1093/HUMREP/DEAD131>
- 704 Whipp, A. M., Heinonen-Guzejev, M., Pietiläinen, K. H., van Kamp, I., & Kaprio, J. (2022). Branched-
705 chain amino acids linked to depression in young adults. *Frontiers in Neuroscience, 16*, 935858.
706 <https://doi.org/10.3389/FNINS.2022.935858/BIBTEX>
- 707 Willemsen, G., De Geus, E. J. C., Bartels, M., Van Beijsterveldt, C. E. M. T., Brooks, A. I., Estourgie-van
708 Burk, G. F., Fugman, D. A., Hoekstra, C., Hottenga, J. J., Klufft, K., Meijer, P., Montgomery, G. W.,
709 Rizzu, P., Sondervan, D., Smit, A. B., Spijker, S., Suchiman, H. E. D., Tischfield, J. A., Lehner, T., ...
710 Boomsma, D. I. (2010). The Netherlands Twin Register biobank: a resource for genetic
711 epidemiological studies. *Twin Research and Human Genetics : The Official Journal of the*
712 *International Society for Twin Studies, 13*(3), 231–245. <https://doi.org/10.1375/TWIN.13.3.231>
- 713 Willemsen, G., Vink, J. M., Abdellaoui, A., Den Braber, A., Van Beek, J. H. D. A., Draisma, H. H. M., Van
714 Dongen, J., Van 'T Ent, D., Geels, L. M., Van Lien, R., Ligthart, L., Kattenberg, M., Mbarek, H., De
715 Moor, M. H. M., Neijts, M., Pool, R., Stroo, N., Klufft, C., Suchiman, H. E. D., ... Boomsma, D. I.
716 (2013). The Adult Netherlands Twin Register: Twenty-Five Years of Survey and Biological Data
717 Collection. *Twin Research and Human Genetics : The Official Journal of the International Society*
718 *for Twin Studies, 16*(1), 271. <https://doi.org/10.1017/THG.2012.140>
- 719 Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W.,
720 Zhou, Y. H., Abdellaoui, A., Batista, S., Butler, C., Chen, G., Chen, T. H., D'Ambrosio, D., Gallins,
721 P., Ha, M. J., Hottenga, J. J., ... Boomsma, D. I. (2014). Heritability and genomics of gene
722 expression in peripheral blood. *Nature Genetics 2014 46:5, 46*(5), 430–437.
723 <https://doi.org/10.1038/ng.2951>
- 724 Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A.,
725 Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., Jeon, M., & Ma'ayan, A. (2021). Gene Set
726 Knowledge Discovery with Enrichr. *Current Protocols, 1*(3), e90.
727 <https://doi.org/10.1002/CPZ1.90>
- 728 Yin, X., Bose, D., Kwon, A., Hanks, S. C., Jackson, A. U., Stringham, H. M., Welch, R., Oravilahti, A.,
729 Fernandes Silva, L., Locke, A. E., Fuchsberger, C., Service, S. K., Erdos, M. R., Bonnycastle, L. L.,
730 Kuusisto, J., Stitzel, N. O., Hall, I. M., Morrison, J., Ripatti, S., ... Wen, X. (2022). Integrating
731 transcriptomics, metabolomics, and GWAS helps reveal molecular mechanisms for metabolite
732 levels and disease risk. *American Journal of Human Genetics, 109*(10), 1727.
733 <https://doi.org/10.1016/J.AJHG.2022.08.007>
- 734 Zhernakova, D. V., Deelen, P., Vermaat, M., Van Iterson, M., Van Galen, M., Arindrarto, W., Van't Hof,
735 P., Mei, H., Van Dijk, F., Westra, H. J., Bonder, M. J., Van Rooij, J., Verkerk, M., Jhamai, P. M.,
736 Moed, M., Kielbasa, S. M., Bot, J., Nooren, I., Pool, R., ... Franke, L. (2016). Identification of
737 context-dependent expression quantitative trait loci in whole blood. *Nature Genetics 2016 49:1,*
738 *49*(1), 139–145. <https://doi.org/10.1038/ng.3737>
- 739 Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., Montgomery, G. W., Goddard, M.
740 E., Wray, N. R., Visscher, P. M., & Yang, J. (2016). Integration of summary data from GWAS and
741 eQTL studies predicts complex trait gene targets. *Nature Genetics 2016 48:5, 48*(5), 481–487.
742 <https://doi.org/10.1038/ng.3538>

743

Figure 1: Manhattan plot showing the results of the transcriptomics analyses in the Netherlands Twin Register (NTR) comparing monozygotic and dizygotic twins.

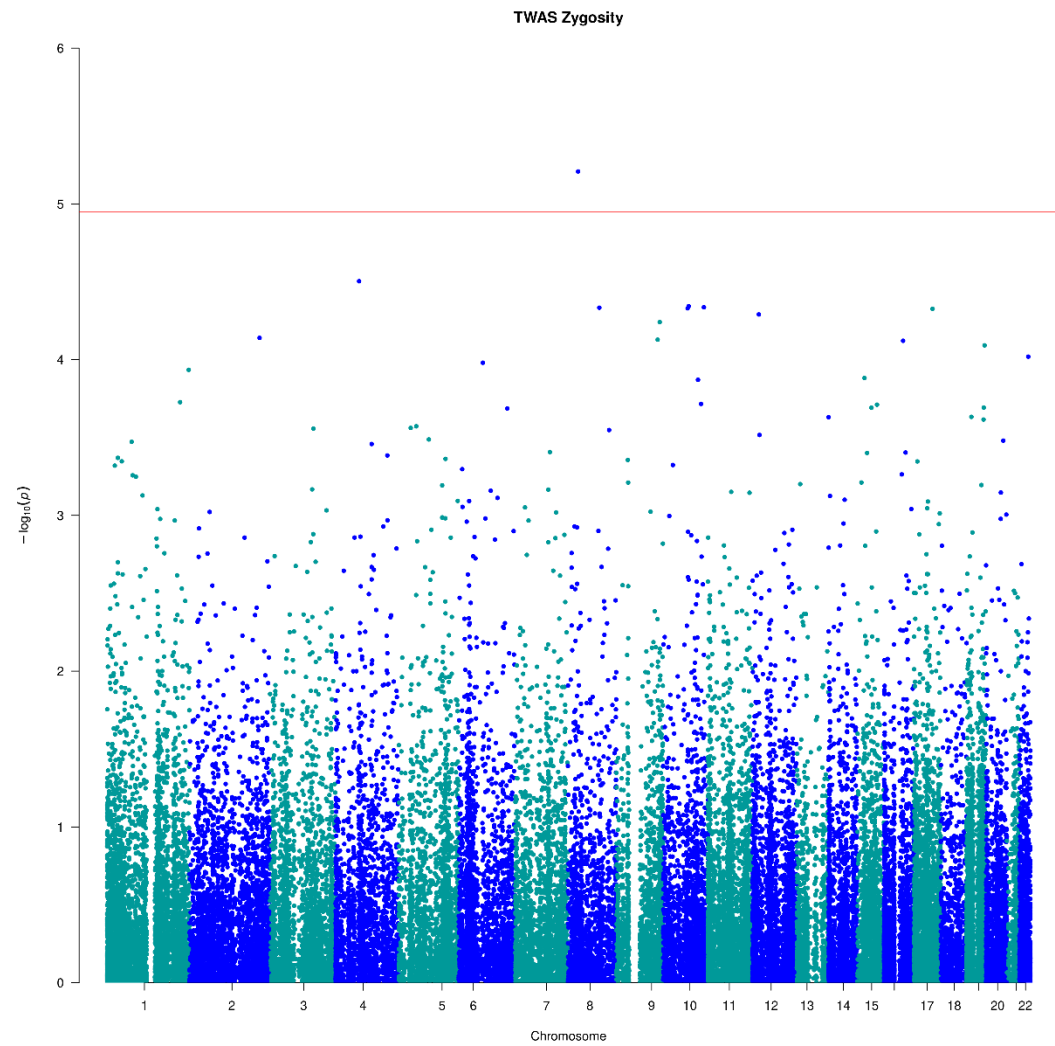


Figure 2: Enrichment analyses of the 5% transcripts with lowest p-values RNA transcript with the lowest p-value in the NTR. The figure only shows the ten most enriched pathways and molecular functions. The left figure shows the enrichment in the Gene Ontology 2023 molecular functions database and the right figure shows the enrichment in the KEGG human pathway database.

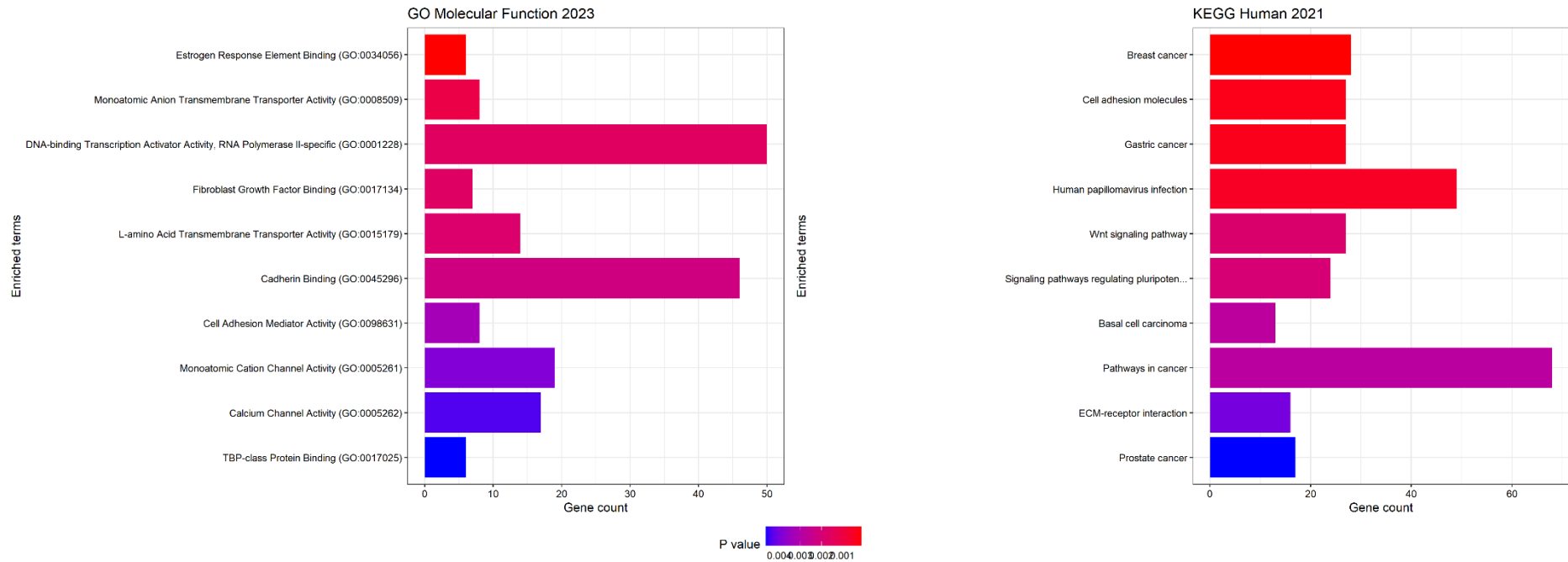


Figure 3: Enrichment analyses of the 5% transcripts with lowest p-values RNA transcript with the lowest p-value in the NTR females. The figure shows the ten most enriched pathways and molecular functions. The left figure shows the enrichment in the Gene Ontology 2023 molecular functions database and the right figure shows the enrichment in the KEGG human pathway database.

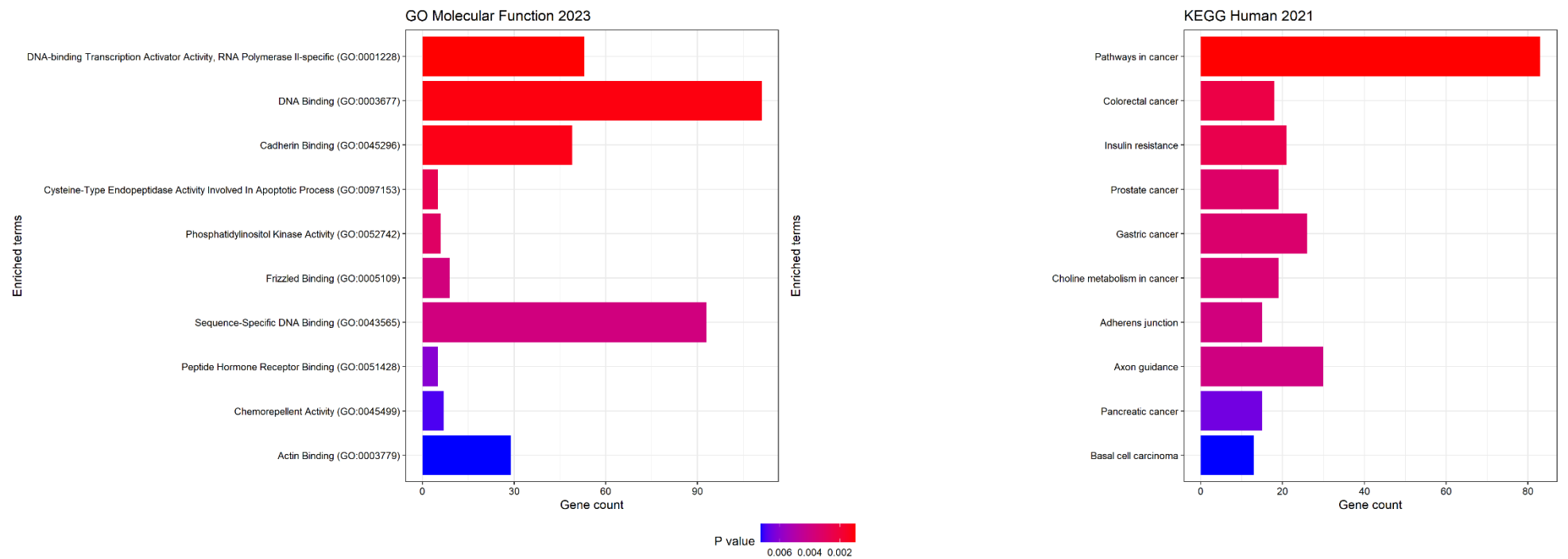


Figure 4: Enrichment analyses of the 5% transcripts with lowest p-values RNA transcript with the lowest p-value in the NTR males. The figure shows the ten most enriched pathways and molecular functions. The left figure shows the enrichment in the Gene Ontology 2023 molecular functions database and the right figure shows the enrichment in the KEGG human pathway database.

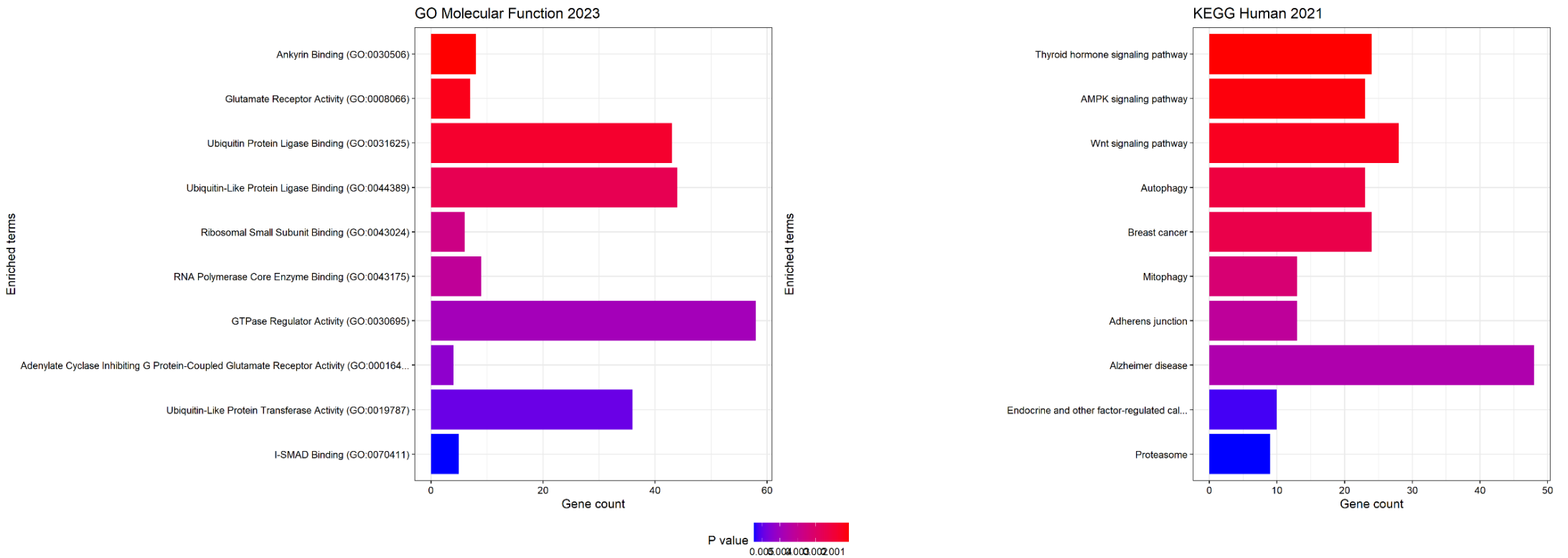


Figure 5: Enrichment analyses of the 19 genes from the 21 significant ($p < 3.77E-05$) RNA transcripts for the transcriptomics analyses in the FTC. The figure shows the ten most enriched pathways and molecular functions. The left figure shows the enrichment in the Gene Ontology 2023 molecular functions database and the right figure shows the enrichment in the KEGG human pathway database.

