CSCI 5465

Lab#7

Quincy Gu

11/19/2019

Part II – Loading and manipulating expression data in R

c). Print the data for the BRCA1 gene for all samples

GSM519722 GSM519723 GSM519724 GSM519725 GSM519726 GSM519727 GSM519728 GSM519729 GSM519730 GSM519731 GSM519732 GSM519733

1 7.274115 6.030057 5.857742 6.229241 5.732221 6.135794 5.784877 6.652859 5.872847 6.188124 6.454896 5.50438

GSM519734 GSM519735 GSM519736 GSM519737 GSM519738 GSM519739 GSM519740 GSM519741 GSM519742 GSM519743 GSM519744 GSM519745

1 4.807938 5.350852 5.622551 6.02982 5.535521 5.963843 5.903684 6.0243 6.065315 5.742128 5.654142 5.942336

GSM519746 GSM519747 GSM519748 GSM519749 GSM519750 GSM519751 GSM519752 GSM519753 GSM519754 GSM519755 GSM519756 GSM519757

1 6.128072 5.993853 5.460781 6.072657 9.892019 5.63162 5.901062 6.826039 6.004497 6.03452 6.015383 5.960794

GSM519758 GSM519759 GSM519760 GSM519761 GSM519762 GSM519763 GSM519764 GSM519765 GSM519767 GSM519768 GSM519769 GSM519770

1 6.265993 6.125837 7.433959 5.834553 5.643395 6.021743 5.457704 6.074645 6.369989 5.36483 6.878876 6.317075

GSM519771 GSM519772 GSM519773 GSM519774 GSM519775 GSM519776 GSM519777 GSM519778 GSM519779 GSM519780 GSM519781 GSM519782

1 7.669003 5.766564 6.318015 6.656691 6.023808 7.455667 6.201401 5.635156 5.561357 5.922497 5.678163 5.770562

GSM519783 GSM519784 GSM519785 GSM519786 GSM519787 GSM519788 GSM519789 GSM519790 GSM519791 GSM519792 GSM519793 GSM519794

1 7.107285 6.151594 5.831414 6.08856 6.146585 5.10978 5.812606 5.878297 5.711583 6.75326 5.991833 6.473865

GSM519795 GSM519796 GSM519797 GSM519798 GSM519799 GSM519800 GSM519801 GSM519802 GSM519803 GSM519804 GSM519805 GSM519806

1 5.243779 6.608707 6.031945 5.717139 5.644369 5.912723 6.089751 6.977763 5.553024 5.749142 5.277986 5.803057

GSM519807 GSM519808 GSM519809 GSM519810 GSM519811 GSM519812

1 5.917982 5.950098 5.003061 5.650207 5.579163 5.11519

NOTE: See results in my code (line 22)

d). [1] 10.029662 6.249125 7.654095 4.066509 3.517488 7.442540 4.720347 4.945013 5.822789 4.191514

See line 25.

e). See line 28-29.

f). See line 33-34.

g). [1] "AFFX-hum\_alu\_at" "AFFX-r2-P1-cre-3\_at" "ND3" "RPL41" "AFFX-r2-P1-cre-5\_at"

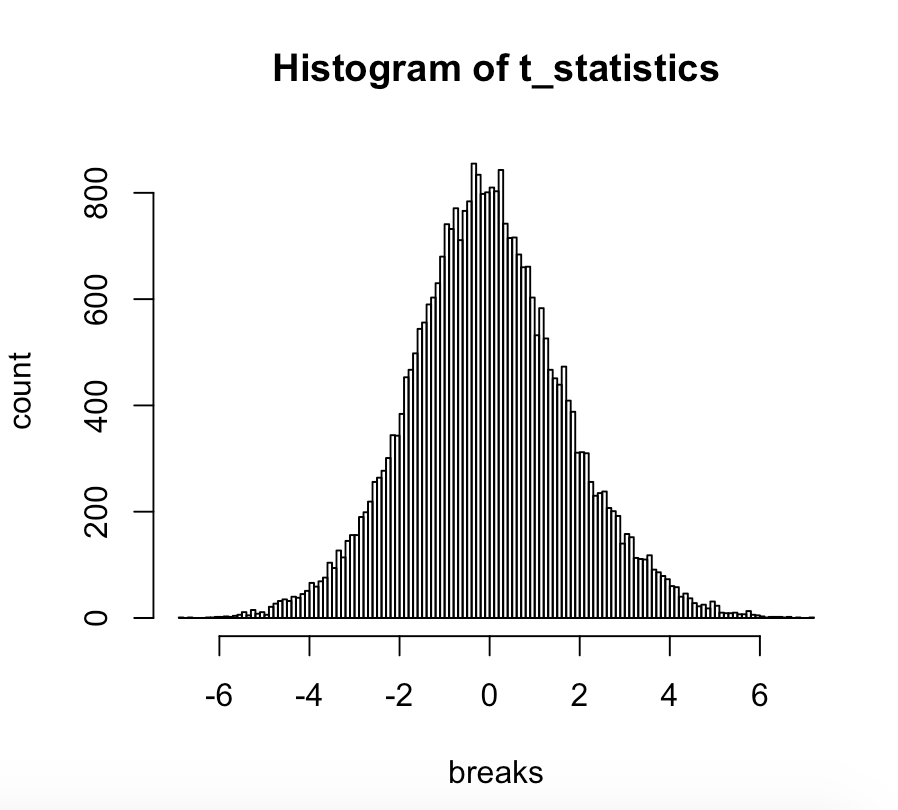
[6] "COX2" "AFFX-CreX-3\_at" "COX1" "221798\_x\_at" "AFFX-CreX-5\_at"

See line 39.

Note that all the other steps listed in Part II instruction sheet all could be found in my code with the corresponding comment there.

Part III – Statistical analysis of microarray gene expression data

b).



d).

pval = pt(-abs(tstat), N-2) only could consider as the integral areas under the t-distribution with the range from negative infinity to the negative of the absolutely value of tstat, since it missing the other part from the absolute value of tstat to the positive infinity, so this only could compute for a one-sided test.

Pval = 2\*pt(-abs(tstat), N-2) is different than the above one, since the range from negative infinity to the negative of the absolutely value of tstat is the same as the range from the absolute value of tstat to the positive infinity, so the integral areas under the t-distribution in those two ranges should be exactly the same, since the above one compute the one-sided test, 2 time the value in the above should return the values which computed for a two-sided test.

e).

i).

Gene\_Name P\_Values T-Statistics

1 NME5 2.983701e-10 7.126161

2 PPP1R14C 1.122778e-09 -6.834366

3 TBC1D9 1.208185e-09 6.818135

4 TARS 2.150219e-09 -6.690151

5 NME3 2.456984e-09 6.660447

6 ESR1 2.760007e-09 6.634514

7 ZNF516 4.823994e-09 6.509595

8 ANXA9 5.306539e-09 6.488196

9 DYNLRB2 7.598497e-09 6.407420

10 TTC8 8.109485e-09 6.392743

11 C6orf97 8.174488e-09 6.390941

12 FAM105B 1.377634e-08 -6.272838

13 FAM174A 1.549825e-08 6.246089

14 SCUBE2 1.604720e-08 6.238176

15 SPOPL 2.066088e-08 6.180627

16 MTMR12 2.441429e-08 -6.142513

17 1555817\_s\_at 3.371852e-08 6.068558

18 CBX2 3.391459e-08 -6.067227

19 NUP155 3.529497e-08 -6.058067

20 230451\_at 3.845827e-08 6.038341

21 SKP1 4.082736e-08 6.024590

22 MYO10 4.668876e-08 -5.993689

23 MYB 5.159998e-08 5.970613

24 232944\_at 5.411148e-08 5.959637

25 C9orf116 6.128828e-08 5.930837

26 LMNB2 6.769179e-08 -5.907820

27 CA12 6.917976e-08 5.902779

28 GFRA1 6.998706e-08 5.900089

29 233388\_at 7.031039e-08 5.899020

30 CCDC74A 7.132216e-08 5.895706

31 SUSD3 7.460364e-08 5.885269

32 AGR3 7.860856e-08 5.873126

33 TRIP13 8.006377e-08 -5.868865

34 MAP3K12 8.334072e-08 5.859541

35 DEGS2 8.414781e-08 5.857300

36 KCNN4 8.539009e-08 -5.853892

37 TTK 9.804879e-08 -5.821706

38 NAT1 1.142836e-07 5.785953

39 232855\_at 1.159929e-07 5.782484

40 ACADSB 1.185402e-07 5.777407

41 MAST4 1.185808e-07 5.777327

42 234222\_at 1.189781e-07 5.776545

43 TCTN1 1.189972e-07 5.776507

44 C6orf211 1.224670e-07 5.769786

45 ERBB4 1.274624e-07 5.760433

46 GREB1 1.304570e-07 5.754997

47 LOC100505614 1.377893e-07 5.742189

48 MICAL3 1.406737e-07 -5.737334

49 COMMD10 1.448668e-07 5.730447

50 THSD4 1.456026e-07 5.729259

51 PAPD7 1.613717e-07 -5.705122

52 215304\_at 1.643082e-07 5.700885

53 SLC25A13 1.671926e-07 -5.696795

54 CCNE1 1.715723e-07 -5.690715

55 PVR 1.748896e-07 -5.686211

56 CDC20 1.791358e-07 -5.680567

57 PRRT3 1.808609e-07 5.678311

58 LONRF2 1.869284e-07 5.670542

59 ABAT 1.899650e-07 5.666747

60 1567997\_x\_at 1.999426e-07 5.654683

61 SLC39A6 2.112444e-07 5.641713

62 TBCK 2.181927e-07 5.634074

63 1557315\_a\_at 2.422222e-07 5.609383

64 LRBA 2.540806e-07 5.598073

65 MELK 2.628070e-07 -5.590077

66 MOAP1 2.675671e-07 5.585824

67 YBX1 2.841679e-07 -5.571553

68 RERG 3.017845e-07 5.557279

69 QSOX2 3.045918e-07 -5.555080

70 NUS1 3.203879e-07 -5.543067

71 HDGFRP3 3.329604e-07 5.533914

72 BAIAP2L1 3.378894e-07 -5.530418

73 DACH1 3.606460e-07 5.514901

74 1562608\_at 3.723806e-07 5.507271

75 SCCPDH 3.731529e-07 5.506777

76 BRIX1 3.750094e-07 -5.505594

77 APOBEC3B 3.853335e-07 -5.499119

78 JRKL 3.950803e-07 -5.493159

79 ANKRA2 3.961156e-07 5.492535

80 VEGFA 4.046927e-07 -5.487422

81 DNALI1 4.057448e-07 5.486802

82 BRI3 4.180941e-07 -5.479643

83 LRP8 4.246633e-07 -5.475918

84 243736\_at 4.369622e-07 5.469095

85 ZNF131 4.485048e-07 -5.462861

86 CASC1 4.559639e-07 5.458915

87 KLHL7 4.831499e-07 -5.445052

88 GDAP2 4.906435e-07 -5.441365

89 KIF2C 5.114359e-07 -5.431417

90 MRFAP1L1 5.115270e-07 5.431374

91 236369\_at 5.255967e-07 5.424867

92 ANKRD30A 5.413309e-07 5.417789

93 1558102\_at 5.500010e-07 5.413974

94 RSPH1 5.552386e-07 5.411698

95 DNAJC12 5.620260e-07 5.408780

96 MICALL1 5.700684e-07 -5.405367

97 GTSE1 5.711248e-07 -5.404922

98 EVL 6.037367e-07 5.391572

99 MKI67 6.164242e-07 -5.386569

100 GAL 6.190053e-07 -5.385563

101 GATA3 6.225992e-07 5.384170

102 1566880\_at 6.565719e-07 5.371376

103 PSAT1 6.626615e-07 -5.369151

104 LOC100286909 7.260793e-07 5.347107

105 232614\_at 7.436897e-07 5.341321

106 228955\_at 7.679115e-07 -5.333577

107 CIRBP 7.718832e-07 5.332330

108 ANLN 7.775124e-07 -5.330574

109 METT5D1 8.083509e-07 5.321167

110 HHAT 8.361387e-07 5.312987

111 215397\_x\_at 8.789776e-07 5.300884

112 LMTK2 8.921429e-07 -5.297281

113 NANOG 9.034129e-07 5.294237

114 VLDLR 9.218261e-07 -5.289344

115 C11orf75 9.241857e-07 -5.288723

116 225197\_at 9.353552e-07 5.285809

117 C5orf34 9.410785e-07 -5.284328

118 ABTB2 9.919784e-07 -5.271538

119 LOC100505730 1.005039e-06 -5.268359

120 XPO5 1.017774e-06 -5.265299

121 TRPV6 1.044470e-06 -5.259003

122 236934\_at 1.055833e-06 5.256371

123 TFF1 1.087435e-06 5.249195

124 HDAC2 1.088515e-06 -5.248953

125 AMD1 1.103408e-06 -5.245645

126 244375\_at 1.138990e-06 5.237914

127 1569167\_at 1.176555e-06 5.230005

128 CYP2B6 1.181785e-06 5.228923

129 SRD5A1 1.193815e-06 -5.226453

130 CCNYL1 1.206486e-06 -5.223878

131 MCM5 1.211466e-06 -5.222873

132 ME2 1.250719e-06 -5.215091

133 GGH 1.254695e-06 -5.214316

134 OSCP1 1.274911e-06 5.210412

135 CCDC96 1.301640e-06 5.205343

136 RDH10 1.499860e-06 -5.170654

137 EME2 1.501251e-06 5.170427

138 C4A 1.648111e-06 5.147526

139 PPP2CA 1.658084e-06 5.146044

140 SKP2 1.708207e-06 -5.138726

141 ZFR 1.719197e-06 -5.137150

142 KIAA0141 1.748118e-06 5.133048

143 KCNK15 1.763206e-06 5.130934

144 235434\_at 1.767912e-06 -5.130278

145 SLC7A8 1.817923e-06 5.123413

146 PPARA 1.830562e-06 -5.121708

147 CENPA 1.845336e-06 -5.119729

148 TMEM144 1.868999e-06 5.116591

149 MAPKSP1 1.875210e-06 5.115773

150 SOX11 1.886524e-06 -5.114292

151 1558103\_a\_at 1.928560e-06 5.108861

152 NEK11 1.974353e-06 5.103075

153 CHD1L 1.988252e-06 -5.101345

154 TOMM40 2.008294e-06 -5.098871

155 RCOR2 2.025438e-06 -5.096773

156 UGCG 2.035306e-06 5.095574

157 DBF4 2.066140e-06 -5.091863

158 C16orf57 2.075622e-06 -5.090732

159 PREP 2.077592e-06 -5.090498

160 242657\_at 2.082481e-06 5.089918

161 NDC80 2.089272e-06 -5.089114

162 PALM 2.089362e-06 5.089103

163 TPRG1 2.099152e-06 5.087949

164 MTMR2 2.101236e-06 -5.087703

165 229425\_at 2.107592e-06 5.086958

166 DNAL1 2.207301e-06 5.075535

167 CMYA5 2.237869e-06 5.072134

168 241689\_at 2.285154e-06 5.066961

169 EIF2C2 2.304866e-06 -5.064835

170 FAM179B 2.376321e-06 5.057276

171 DNAJC16 2.414606e-06 5.053317

172 231171\_at 2.447814e-06 5.049932

173 1554007\_at 2.511925e-06 5.043522

174 C9orf64 2.511993e-06 5.043516

175 238243\_at 2.574013e-06 5.037465

176 MCM7 2.593364e-06 -5.035607

177 232307\_at 2.602574e-06 5.034727

178 METTL14 2.607078e-06 5.034298

179 NEK9 2.619285e-06 5.033138

180 ASPM 2.656382e-06 -5.029646

181 KIAA1737 2.660204e-06 5.029289

182 TCEAL1 2.666968e-06 5.028659

183 1567101\_at 2.677042e-06 5.027722

184 1564358\_at 2.723926e-06 5.023409

185 VAV3 2.831359e-06 5.013793

186 IL6ST 2.873228e-06 5.010142

187 CDC45 2.966695e-06 -5.002174

188 ADAMTS15 3.033896e-06 4.996596

189 ARNT2 3.058832e-06 4.994557

190 TCEAL3 3.071567e-06 4.993521

191 237365\_at 3.073889e-06 4.993333

192 MARCO 3.160112e-06 -4.986437

193 240485\_at 3.166516e-06 4.985932

194 238883\_at 3.205379e-06 4.982890

195 232210\_at 3.207669e-06 4.982712

196 KLHDC1 3.280589e-06 4.977103

197 1555893\_at 3.291057e-06 4.976308

198 PREX1 3.341310e-06 4.972525

199 241577\_at 3.380327e-06 4.969625

200 C14orf179 3.387572e-06 4.969091

201 IGF1R 3.478933e-06 4.962441

202 LOC692247 3.494293e-06 4.961340

203 OR7E37P 3.496098e-06 -4.961211

204 237301\_at 3.542312e-06 4.957928

205 ELP2 3.548700e-06 4.957477

206 FAHD1 3.569218e-06 4.956035

207 215571\_at 3.622563e-06 4.952324

208 B3GNT5 3.771495e-06 -4.942239

209 214202\_at 3.780304e-06 4.941655

210 238044\_at 3.893460e-06 4.934265

211 SLC22A5 4.005488e-06 4.927153

212 237881\_at 4.104171e-06 4.921046

213 FAM63A 4.106178e-06 4.920923

214 244696\_at 4.106487e-06 4.920905

215 CXXC5 4.151777e-06 4.918151

216 MLPH 4.210825e-06 4.914605

217 C16orf45 4.221157e-06 4.913990

218 FAM100B 4.290823e-06 -4.909878

219 IFRD1 4.310286e-06 -4.908741

220 AURKB 4.334446e-06 -4.907336

221 TTC30B 4.334841e-06 4.907313

222 TPBG 4.390394e-06 4.904112

223 SALL2 4.412697e-06 4.902838

224 240723\_at 4.414261e-06 4.902749

225 C19orf28 4.469716e-06 -4.899610

226 SYNJ2BP 4.651176e-06 4.889596

227 C1orf64 4.658923e-06 4.889177

228 ZNF434 4.698906e-06 4.887025

229 MCM10 4.725178e-06 -4.885621

230 DNMT3B 4.841181e-06 -4.879510

231 CST3 4.975613e-06 4.872605

232 NUDT12 4.995662e-06 4.871590

233 C9orf30 5.004505e-06 -4.871144

234 MED13L 5.016272e-06 4.870552

235 MYBL2 5.126577e-06 -4.865063

236 SUV39H2 5.212683e-06 -4.860858

237 PCBD2 5.220173e-06 4.860496

238 PSME4 5.233462e-06 -4.859854

239 EXO1 5.301073e-06 -4.856612

240 NAMPT 5.310579e-06 -4.856159

241 FAM114A2 5.340296e-06 4.854749

242 GAMT 5.340595e-06 4.854735

243 238544\_at 5.377657e-06 4.852988

244 EPHB4 5.533960e-06 -4.845744

245 FLJ36031 5.540839e-06 -4.845430

246 242142\_at 5.591953e-06 4.843108

247 UQCRH 5.620226e-06 -4.841832

248 233214\_at 5.628490e-06 4.841460

249 PHGDH 5.856091e-06 -4.831425

250 TMEM30B 5.908144e-06 4.829183

251 C6orf106 5.917310e-06 -4.828791

252 235938\_at 6.081667e-06 4.821847

253 CDCA2 6.083907e-06 -4.821754

254 CLCN5 6.113315e-06 -4.820531

255 CDC42EP1 6.156415e-06 -4.818750

256 CEP55 6.239543e-06 -4.815347

257 IL8 6.287318e-06 -4.813412

258 SLC44A4 6.367799e-06 4.810184

259 CASC5 6.376895e-06 -4.809822

260 227533\_at 6.392386e-06 4.809206

261 237377\_at 6.486519e-06 4.805494

262 ADCY9 6.535081e-06 4.803599

263 TAF5 6.594516e-06 -4.801299

264 227052\_at 6.673786e-06 4.798263

265 SRPK1 6.710099e-06 -4.796883

266 224989\_at 6.741831e-06 4.795684

267 ERCC6L 6.771549e-06 -4.794566

268 PAK1IP1 6.785207e-06 -4.794054

269 236139\_at 6.853819e-06 -4.791495

270 UPP1 6.872122e-06 -4.790817

271 RAB17 6.881933e-06 4.790454

272 THOC5 6.966498e-06 -4.787347

273 TAF1D 6.997124e-06 -4.786230

274 CELSR1 7.007040e-06 4.785870

275 PTP4A2 7.032023e-06 4.784964

276 FAM13B 7.049187e-06 4.784344

277 C10orf32 7.052792e-06 4.784213

278 VGLL1 7.169373e-06 -4.780039

279 STIL 7.170408e-06 -4.780003

280 E2F3 7.176921e-06 -4.779772

281 PRKAA1 7.507188e-06 -4.768307

282 ASCC2 7.535106e-06 -4.767361

283 242506\_at 7.607082e-06 -4.764936

284 CCDC150 7.639717e-06 -4.763844

285 CPSF3 7.843825e-06 -4.757116

286 YTHDC2 7.939730e-06 4.754013

287 HGF 7.972679e-06 4.752956

288 KIF15 8.020348e-06 -4.751433

289 TFRC 8.074030e-06 -4.749729

290 NFIL3 8.084993e-06 -4.749382

291 GNPDA2 8.112122e-06 4.748527

292 ORAI3 8.194494e-06 4.745945

293 222348\_at 8.245743e-06 4.744351

294 KCTD3 8.309772e-06 4.742374

295 229602\_at 8.397848e-06 4.739678

296 HSD17B2 8.401612e-06 -4.739563

297 CRIP1 8.444501e-06 4.738261

298 LOC375295 8.515036e-06 4.736133

299 230431\_at 8.634127e-06 4.732578

300 ASNS 8.729948e-06 -4.729753

301 KIAA0556 8.890402e-06 4.725089

302 MFHAS1 8.909321e-06 -4.724544

303 MEIS3P1 9.003020e-06 4.721864

304 BEND3 9.064815e-06 -4.720111

305 CDCA5 9.080996e-06 -4.719653

306 GPR77 9.091018e-06 4.719371

307 244625\_at 9.105000e-06 4.718977

308 DSC2 9.178915e-06 -4.716904

309 MUC15 9.218256e-06 -4.715807

310 TMEM178 9.259948e-06 4.714650

311 GARS 9.319647e-06 -4.713002

312 243671\_at 9.386465e-06 4.711169

313 JOSD1 9.559540e-06 -4.706481

314 243634\_at 9.741438e-06 4.701641

315 232286\_at 9.764481e-06 4.701034

316 AKR7A2 9.855556e-06 4.698650

317 PDAP1 9.908536e-06 -4.697272

318 FAM83D 9.942299e-06 -4.696398

319 HMGA1 9.976491e-06 -4.695516

320 BBS4 9.990017e-06 4.695168

321 FBXL5 1.002389e-05 4.694298

322 NT5DC2 1.007971e-05 -4.692870

323 HJURP 1.016182e-05 -4.690784

324 KIAA1804 1.016728e-05 -4.690646

325 232134\_at 1.025632e-05 -4.688404

326 A2ML1 1.025643e-05 -4.688401

327 C12orf76 1.026254e-05 4.688248

328 CDK19 1.053135e-05 -4.681595

329 TTC33 1.073012e-05 -4.676781

330 ZNF747 1.074290e-05 4.676475

331 C9orf150 1.081380e-05 4.674781

332 NCRNA00081 1.120819e-05 4.665549

333 230757\_at 1.131898e-05 4.663012

[ reached 'max' / getOption("max.print") -- omitted 7603 rows ]

ii).

7936 (line 129), and 137 (line 137).

iii).

3930 (line166)

iv).

4006 (line 167)

v).

Picked-up Gene name is TTC8.

Function:

“This gene encodes a protein that has been directly linked to Bardet-Biedi syndrome, which has the features that include retinal dystrophy, obesity, polydactyly, renal abnormalities and learning disabilities”, dieases associated with TTC8 include “Retinitis Pigmentosa 51 and Bardet-Biedi Syndrome 8”.

Note that all the other steps listed in Part III instruction sheet all could be found in my code with the corresponding comment there.

Part IV – Using R Packages to Process RNA-Seq Data

f).

[1] "KIAA1324.1" "MAPT.17" "C6orf97.6" "POPDC3.6" "TFF3.21" "APOBEC3C.22" "TFF1.21" "CADM1.11"

[9] "DEGS2.14" "PDZRN3.3" "ROR1.1" "ERBB4.2" "GREB1.2" "PREX1.20" "LIPG.18" "GRIK3.1"

[17] "PLCB1.20" "TLR6.4" "GBP1.1" "GNAO1.16" "DNAJA4.15" "CRIP1.14" "RPS6KA6.x" "SLC9A2.2"

[25] "F3.1" "PGR.11" "NPY1R.4" "CXCL12.10" "CACNA1H.16" "AGR3.7" "RASGRP3.2" "AR.x"

[33] "SCUBE2.11" "AMZ1.7" "UPP1.7" "CGREF1.2" "GPR77.19" "IGF2BP3.7" "TGFB3.14" "PUNC.15"

[41] "TOX3.16" "GLUD2.x" "KCNK15.20" "ADRA2A.10" "EGFR.7" "ACSL4.x" "TM7SF2.11" "CEACAM5.19"

[49] "RDHE2.8" "BDKRB2.14" "HS6ST3.13" "IL8.4" "MYRIP.3" "STAC.3" "ETS1.11" "ICAM1.19"

[57] "IGSF9B.11" "SYT10.12" "MAML2.11" "SH3TC2.5" "SOX2.3" "SAMD9L.7" "KCNF1.2" "TARP.7"

[65] "ABCC11.16" "FLJ14213.11" "NKD2.5" "ADAMTS19.5" "CDC42EP3.2" "UBASH3B.11" "NLRC5.16" "SULF1.8"

[73] "COTL1.16" "IL7R.5" "LEPREL1.3" "NOVA1.14" "F7.13" "DOCK3.3" "MYO7B.2" "KCNJ11.11"

[81] "LIN7A.12" "SEPT5.22" "MT2A.16" "KIAA2022.x" "DLG2.11" "AFAP1L1.5" "NAP1L2.x" "C20orf134.20"

[89] "STARD10.11" "MMP17.12" "LOC196993.15" "ARX.x" "COL13A1.10" "GSTT1.22" "CHI3L2.1" "CARD6.5"

[97] "UGT3A2.5" "CD44.11" "ACN9.7" "UBA7.3"