

A Pan Cancer Atlas

```

setwd("~/desktop/THCA")
ex <- read.delim("~/Desktop/THCA/ex_LUAD.tsv")
cn <- ex #backup

rownames(cn) <- cn$gene_id
cn <- cn[,-1:-2]

# Duplicated Counts from patients
cn <- cn[,-c(178,223,226,246,339,428,485,474)]

#sample types
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

sampleID <- read.delim("~/desktop/THCA/TCGA-LUAD/Data/gdc_sample_sheet.2022-06-11.tsv")
i1 <- sampleID$Data.Type == "Gene Expression Quantification"
sampleID <- sampleID[i1, , drop = F]
sampleID <- sampleID[,7:8]
sampleID <- distinct(sampleID, Sample.ID, .keep_all = T)
a <- colnames(cn)
for(i in 1:3){a <- sub('\\\\.', '-', a)}
b <- data.frame(IDs = a)
c <- merge(b, sampleID, by.x = "IDs", by.y = "Sample.ID", sort = F)
cc <- data.frame(sample = sub("Primary Tumor", "PRT", c$Sample.Type))
cc <- data.frame(sample = sub("Solid Tissue Normal", "STN", cc$sample))
for(i in 1:3){c$IDs <- sub('\\\\-', '.', c$IDs)}
colnames(cn) <- cc$sample
head(cn)

##
##          PRT  STN  STN   PRT  PRT  PRT  PRT  PRT  PRT  PRT  STN  PRT
## ENSG00000000003.15 7434  548  573 11426 1624 2567 2393 1456 3719 1990 2143
## ENSG00000000005.6    0    1    1     3    0    1    0    0    2     9    0
## ENSG000000000419.13 2968  709  619  2428 1588 1771 2238 1851 2154 1984 1452
## ENSG000000000457.14  555  399  370  1236  498  354  640  650 1319   608  737
## ENSG000000000460.17  643   92   68   888  114  488  468  378  563   161  430
## ENSG000000000938.13 2145 3718 2599  1375  743  343 1232 1686  357 12548 1304
##
##          PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  STN  PRT
## ENSG00000000003.15 3131 2778 3345 1430 3632 2018 4313 2405 7805 5100  950 1443
## ENSG00000000005.6    0    3    1    0    1    0    1   48   19    1    1    0

```

```

## ENSG00000000419.13 3193 1145 3374 1196 1796 1252 1337 1770 9358 2394 706 1275
## ENSG00000000457.14 1326 4417 1037 617 774 556 591 1076 1789 939 475 230
## ENSG00000000460.17 483 421 806 190 228 316 144 672 2743 673 87 104
## ENSG00000000938.13 580 390 3583 341 1171 420 166 1053 1775 2717 2201 930
##
## PRT PRT PRT PRT PRT STN PRT PRT PRT PRT PRT PRT
## ENSG00000000003.15 848 6145 2120 1919 3601 459 8297 2255 4383 669 3249 3715
## ENSG00000000005.6 1 2 0 0 2 0 1 0 0 0 0 0
## ENSG00000000419.13 797 2342 1161 2082 2721 725 1960 3896 2006 889 1969 2468
## ENSG00000000457.14 446 761 432 899 1287 372 2100 2317 1632 442 1345 249
## ENSG00000000460.17 95 451 126 262 1075 74 543 1258 482 662 427 395
## ENSG00000000938.13 370 1512 981 2432 281 5527 519 756 1209 247 2713 150
##
## PRT PRT PRT STN PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG00000000003.15 3010 2818 4519 985 2233 2477 3495 7090 3857 4092 1472 2325
## ENSG00000000005.6 0 1 0 2 0 0 3 7 0 0 0 0
## ENSG00000000419.13 1343 1328 1322 1278 2803 1325 1625 1599 4734 3568 545 2313
## ENSG00000000457.14 956 1262 911 413 940 587 788 1377 1142 622 558 776
## ENSG00000000460.17 175 270 1180 135 647 363 440 441 689 508 168 1012
## ENSG00000000938.13 675 1200 1796 12378 1136 942 480 1778 1489 302 702 296
##
## PRT PRT STN PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG00000000003.15 2519 1959 6760 4409 2906 9492 1459 3513 4242 1566 667 2165
## ENSG00000000005.6 0 0 3 0 0 1 0 1 16 0 10 1
## ENSG00000000419.13 3146 1104 2070 3799 2724 2524 1031 1436 1827 2666 554 2351
## ENSG00000000457.14 1741 837 1110 1998 909 741 861 492 661 701 716 776
## ENSG00000000460.17 1144 524 202 1683 760 225 187 363 405 642 619 960
## ENSG00000000938.13 1340 1645 4876 1088 286 110 6849 162 2584 1327 455 744
##
## PRT STN STN PRT PRT PRT PRT STN PRT PRT PRT PRT PRT
## ENSG00000000003.15 13245 510 1120 1218 3595 1329 945 622 4545 2295 1759 4597
## ENSG00000000005.6 77 2 7 10 1 1 0 1 0 1 0 0
## ENSG00000000419.13 3874 660 980 558 1163 1259 901 619 1557 1113 963 2710
## ENSG00000000457.14 2285 346 343 830 329 274 709 411 1185 430 227 1064
## ENSG00000000460.17 2616 65 72 463 186 190 127 79 322 222 229 326
## ENSG00000000938.13 657 1310 4442 305 1872 917 489 1896 424 654 532 1230
##
## PRT PRT PRT PRT PRT PRT STN PRT PRT PRT PRT PRT STN
## ENSG00000000003.15 5460 1583 1975 1188 9092 3739 1304 5042 735 5351 2786 881
## ENSG00000000005.6 1 2 4 0 4 0 1 4 0 0 0 17 1
## ENSG00000000419.13 1682 483 1197 1564 2861 2047 791 1966 705 1269 1443 941
## ENSG00000000457.14 855 597 1705 269 743 2240 375 1920 478 909 680 536
## ENSG00000000460.17 253 183 560 219 283 525 57 410 112 456 242 144
## ENSG00000000938.13 1256 598 1658 321 1023 355 1842 1404 1499 986 1458 2954
##
## PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT STN
## ENSG00000000003.15 1501 1768 6379 1569 626 1501 4413 1901 6067 2101 9732 1395
## ENSG00000000005.6 0 2 0 0 0 0 1 6 0 0 0 20 8
## ENSG00000000419.13 1769 1703 2002 1070 576 1160 1726 961 2126 1078 2166 1031
## ENSG00000000457.14 790 997 658 929 264 473 1895 225 677 750 1000 541
## ENSG00000000460.17 330 381 345 310 163 282 411 169 651 237 345 135
## ENSG00000000938.13 348 248 1287 1076 958 346 1527 447 362 484 855 3245
##
## PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG00000000003.15 8028 2504 1233 2611 1198 1567 1225 3434 1486 2474 6428 7709
## ENSG00000000005.6 3 3 0 0 0 1 1 0 2 188 4 4
## ENSG00000000419.13 2142 1060 411 828 815 1007 340 2676 1226 1754 1522 2754
## ENSG00000000457.14 1352 631 272 503 680 541 347 1269 465 278 594 2163
## ENSG00000000460.17 622 316 34 332 146 203 136 765 441 204 767 499
## ENSG00000000938.13 2641 657 285 2151 1094 2945 398 801 1223 305 533 1212
##
## PRT PRT STN PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT

```

```

## ENSG000000000003.15 2607 1329 780 1728 3661 4382 1841 6832 1501 2897 5316 6084
## ENSG000000000005.6 148 0 2 3 18 7 19 0 0 2 5 0
## ENSG000000000419.13 1051 1395 789 781 662 2336 1308 1608 1483 3171 2886 2553
## ENSG000000000457.14 565 297 406 490 436 1320 445 978 363 771 631 856
## ENSG000000000460.17 284 320 146 218 96 1540 414 409 286 249 716 300
## ENSG000000000938.13 874 387 10095 2347 471 1939 1800 846 560 532 1468 1005
## PRT PRT PRT PRT PRT STN PRT PRT PRT PRT PRT
## ENSG000000000003.15 2352 2291 13000 4537 4915 651 10120 1703 2261 2655 785
## ENSG000000000005.6 1 2 1 3 0 0 0 0 0 2 3
## ENSG000000000419.13 911 586 2073 1202 2023 927 6714 702 1792 574 969
## ENSG000000000457.14 760 384 426 1099 1156 462 1006 457 246 362 574
## ENSG000000000460.17 206 206 624 568 670 94 2088 178 131 142 136
## ENSG000000000938.13 136 373 654 972 1463 4866 1873 494 275 463 6791
## PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 9058 3190 1845 1678 1085 3932 3190 4673 1815 1428 3582 2783
## ENSG000000000005.6 3 2 9 7 0 0 0 2 4 1 1 1
## ENSG000000000419.13 4525 1083 1669 987 1382 980 496 1715 999 1011 1222 1131
## ENSG000000000457.14 977 1131 1203 755 610 407 555 1975 918 695 789 672
## ENSG000000000460.17 870 226 293 219 599 207 243 594 263 151 344 173
## ENSG000000000938.13 4621 535 1004 1202 543 554 994 421 1054 3410 1746 398
## STN PRT PRT PRT STN PRT PRT PRT STN PRT PRT PRT
## ENSG000000000003.15 1988 3169 3334 613 823 1194 8522 2547 618 2652 8038 2330
## ENSG000000000005.6 6 47 8 0 1 1 0 0 0 0 6 0
## ENSG000000000419.13 1520 2278 1317 789 923 1812 2548 1365 570 2261 1763 2728
## ENSG000000000457.14 815 943 644 889 230 489 2384 612 346 327 1005 965
## ENSG000000000460.17 170 585 120 577 76 617 1790 184 78 368 552 318
## ENSG000000000938.13 6303 2660 1241 1914 3373 951 491 1579 1476 455 510 10286
## PRT PRT PRT STN PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 2918 4165 5108 1229 2777 10056 2138 1413 1804 1998 2272 1807
## ENSG000000000005.6 0 0 1 1 11 1 4 86 0 3 0 41
## ENSG000000000419.13 2183 1534 1570 1703 1205 6290 2179 1452 1118 1969 851 620
## ENSG000000000457.14 1616 1162 641 345 611 672 830 375 554 983 565 1368
## ENSG000000000460.17 781 446 525 122 376 768 610 387 196 665 199 695
## ENSG000000000938.13 2165 2130 577 3772 763 633 2447 886 243 576 2245 227
## PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 5849 12949 1599 1196 5980 2982 1107 2063 1093 2508 870 8875
## ENSG000000000005.6 2 6 0 4 1 0 0 1 0 0 0 0
## ENSG000000000419.13 1227 2197 515 359 3549 657 776 1209 752 775 734 3908
## ENSG000000000457.14 757 1530 295 621 1297 341 495 717 1113 466 672 805
## ENSG000000000460.17 520 701 21 373 673 179 115 251 213 172 184 924
## ENSG000000000938.13 936 1531 246 328 1529 340 135 609 1106 303 1939 1227
## PRT STN PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 4241 2370 3373 5052 1260 1516 857 2730 3083 3347 3541 3103
## ENSG000000000005.6 0 1 0 7 2 1 0 0 1 0 0 0
## ENSG000000000419.13 2643 2770 1045 1288 1139 837 1000 1583 808 899 1982 1030
## ENSG000000000457.14 950 621 662 1049 334 293 863 804 767 824 854 489
## ENSG000000000460.17 1010 166 386 320 302 87 346 356 276 245 431 462
## ENSG000000000938.13 1615 17257 1067 1272 1037 695 236 2852 532 835 1763 1242
## PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 854 3540 1597 3866 908 5440 4558 3401 3500 2960 3641 2834
## ENSG000000000005.6 9 1 1 0 0 0 0 0 17 1 0 2
## ENSG000000000419.13 689 2662 1124 1516 722 2397 2089 1626 1525 1384 1804 1307
## ENSG000000000457.14 789 1091 538 779 412 756 1025 405 467 1039 687 715
## ENSG000000000460.17 777 819 292 218 113 530 315 474 420 204 338 285

```

```

## ENSG000000000938.13 134 3082 531 1468 398 870 1760 685 1345 935 1533 1045
## PRT STN PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 2015 743 2307 3269 1952 3341 4460 1086 1370 2905 4196 1338
## ENSG000000000005.6 15 0 1295 0 0 1 0 0 0 125 1 0
## ENSG000000000419.13 1215 739 2030 1540 956 1340 1109 687 614 1930 2406 2217
## ENSG000000000457.14 693 405 631 1171 573 603 411 304 408 1103 2073 1101
## ENSG000000000460.17 212 60 823 318 195 231 283 180 130 574 1106 620
## ENSG000000000938.13 417 3111 115 1915 175 1189 2004 270 510 701 1454 1683
## PRT PRT PRT PRT PRT PRT PRT PRT STN PRT PRT STN
## ENSG000000000003.15 3875 4288 15402 2627 3446 401 2752 1139 513 3440 2078 2497
## ENSG000000000005.6 8 1 20 1 2 73 1 8 0 333 5 7
## ENSG000000000419.13 1171 1335 3402 602 1469 154 1162 552 585 862 1004 1525
## ENSG000000000457.14 421 637 1474 452 706 221 1088 234 304 717 368 803
## ENSG000000000460.17 814 539 863 171 415 102 888 127 77 267 117 165
## ENSG000000000938.13 162 565 1484 283 1332 79 692 839 3081 1567 763 3215
## STN STN PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 835 764 6961 4813 2720 3760 7666 6709 1157 4118 2405 9590
## ENSG000000000005.6 0 1 0 0 1 3 1 6 2 0 1 3
## ENSG000000000419.13 572 1088 3794 2160 1963 632 1574 1668 1235 1117 1434 2299
## ENSG000000000457.14 274 265 1337 785 1690 351 948 997 805 632 590 1423
## ENSG000000000460.17 93 74 1038 305 307 106 424 541 208 389 335 419
## ENSG000000000938.13 1564 2501 1538 3024 4064 239 1210 1609 4548 472 816 2024
## STN PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 1453 1221 3528 5937 1584 4916 2190 788 7284 3841 6341 1384
## ENSG000000000005.6 2 0 0 0 2 2 2 0 2 18 31 1
## ENSG000000000419.13 937 1129 1092 1435 1594 873 1125 1386 1092 1625 2299 642
## ENSG000000000457.14 354 545 1081 1093 835 699 1122 1305 1218 1118 637 602
## ENSG000000000460.17 91 318 563 254 510 162 357 228 176 762 289 119
## ENSG000000000938.13 2558 196 1724 1487 114 636 560 492 534 2424 2541 184
## PRT STN PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 4337 2510 4768 3428 4144 1946 8559 1717 2170 1343 6656 3942
## ENSG000000000005.6 0 4 1 1 0 1 2 1 21 0 0 1
## ENSG000000000419.13 3201 881 1494 1453 2439 734 2325 755 1380 907 3399 3034
## ENSG000000000457.14 1367 506 719 497 1644 724 1365 608 655 823 1292 1280
## ENSG000000000460.17 625 96 426 753 670 450 671 456 367 189 735 913
## ENSG000000000938.13 307 1062 298 521 1364 776 775 3008 1714 1047 1325 602
## PRT PRT PRT STN PRT PRT PRT PRT PRT PRT PRT STN PRT
## ENSG000000000003.15 2723 1826 7372 1294 4407 4028 2366 3887 1475 4026 696 3014
## ENSG000000000005.6 0 1 4 3 2 1 1 0 0 0 0 7
## ENSG000000000419.13 1637 1309 1315 1108 513 3934 2760 2823 1275 2021 558 1924
## ENSG000000000457.14 805 521 1351 673 336 1783 807 1024 373 1398 290 1208
## ENSG000000000460.17 358 238 332 114 97 881 1011 724 172 863 54 506
## ENSG000000000938.13 1153 245 1245 3578 576 702 491 4651 81 1090 2471 596
## PRT PRT PRT STN STN PRT STN PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 5487 12232 9468 1281 1745 4199 2615 1073 2292 3424 746 5638
## ENSG000000000005.6 0 0 1 107 2 5 1 1 3 1 0 1
## ENSG000000000419.13 1804 4063 2420 1042 1381 2074 1664 892 654 1700 691 1208
## ENSG000000000457.14 1798 1599 1100 334 574 1952 858 351 460 2170 300 883
## ENSG000000000460.17 337 1161 417 94 162 666 172 307 165 457 142 211
## ENSG000000000938.13 3921 139 1515 1647 3844 878 9948 809 658 263 2646 736
## PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 2311 1887 2184 1806 2021 4735 1443 544 3893 3519 4410 6732
## ENSG000000000005.6 0 1 0 1 0 2 1 0 0 3 0 0
## ENSG000000000419.13 1331 855 1549 1066 833 1031 1716 1174 1818 1422 1630 2738

```

```

## ENSG000000000457.14 385 517 855 355 900 765 669 309 1487 490 940 1133
## ENSG000000000460.17 499 320 298 260 133 152 302 382 919 459 763 1258
## ENSG000000000938.13 540 1318 102 2381 237 783 2315 1052 1416 698 907 592
## PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 4422 5479 3418 7321 886 1853 1305 3513 2347 4662 3883 942
## ENSG000000000005.6 0 3 0 4 0 1 0 0 2 1 0 2
## ENSG000000000419.13 2578 1015 2937 1475 988 675 732 3093 603 1414 2952 878
## ENSG000000000457.14 794 549 1008 629 601 409 929 1291 469 543 1682 635
## ENSG000000000460.17 879 204 592 303 299 93 344 1794 301 161 1267 150
## ENSG000000000938.13 1896 196 1926 1770 1954 1364 160 1332 240 1236 1829 589
## PRT PRT STN PRT PRT PRT PRT STN PRT PRT PRT PRT
## ENSG000000000003.15 3666 2157 829 1403 1811 4887 4236 348 6839 2784 1575 4702
## ENSG000000000005.6 5 0 6 1 1 5 1321 0 150 4 0 0
## ENSG000000000419.13 1861 2330 829 662 865 1677 1799 752 2512 2321 1819 1571
## ENSG000000000457.14 647 631 334 256 788 982 1301 211 2889 1529 917 432
## ENSG000000000460.17 959 567 93 158 195 509 376 38 1865 637 173 488
## ENSG000000000938.13 1275 3184 2289 372 2406 2570 823 1506 2257 1087 38 738
## PRT PRT PRT STN PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 4577 1848 7867 1546 848 2134 1779 4477 2936 1877 843 1539
## ENSG000000000005.6 0 0 44 1 23 3 0 12 0 4 0 0
## ENSG000000000419.13 2329 1264 3030 1563 552 1890 911 2813 1447 920 739 1282
## ENSG000000000457.14 1257 590 1011 975 900 511 565 1014 827 417 653 445
## ENSG000000000460.17 823 172 203 145 562 163 204 525 241 279 148 198
## ENSG000000000938.13 3079 579 1878 4125 286 2963 263 742 343 448 312 852
## PRT PRT PRT PRT STN PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 4275 5948 1097 8534 1110 1179 2626 8462 2203 1638 1672 2546
## ENSG000000000005.6 0 0 1 1 0 0 0 3 0 0 1 108
## ENSG000000000419.13 1071 1174 1486 2291 1192 1243 865 3576 889 1563 729 3384
## ENSG000000000457.14 630 628 200 886 275 661 585 1871 535 655 476 1409
## ENSG000000000460.17 542 572 215 419 74 229 730 828 188 369 206 908
## ENSG000000000938.13 372 113 633 1134 3373 1866 266 788 632 1734 278 807
## PRT STN PRT STN PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 3071 689 2223 1353 8235 5001 3861 2473 4123 2559 8645 7254
## ENSG000000000005.6 0 2 2 2 129 0 0 8 0 1 8 2
## ENSG000000000419.13 1155 653 711 842 2969 1452 3149 1801 1390 1415 3076 2428
## ENSG000000000457.14 660 269 397 417 954 1308 419 839 568 561 1623 716
## ENSG000000000460.17 362 57 155 106 434 789 502 394 116 336 517 521
## ENSG000000000938.13 664 2924 329 2699 1646 1963 69 666 1770 1778 1273 735
## PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 1237 4884 2096 5137 10178 6072 2370 2165 5963 5315 8641 1536
## ENSG000000000005.6 0 0 0 0 10 0 8 0 0 9 3 0
## ENSG000000000419.13 1662 1736 1665 1429 2362 4411 1290 1587 1075 3541 2135 765
## ENSG000000000457.14 671 647 403 831 3230 593 713 747 979 1756 1175 510
## ENSG000000000460.17 435 543 270 244 505 748 496 370 222 918 549 172
## ENSG000000000938.13 611 245 565 156 1773 488 847 1634 856 969 1499 911
## STN PRT PRT PRT PRT PRT PRT PRT PRT PRT STN PRT PRT
## ENSG000000000003.15 2056 2351 1577 2585 1168 2296 1694 6038 3868 1500 3365 2436
## ENSG000000000005.6 5 3 0 0 0 2 2 0 3 3 0 2
## ENSG000000000419.13 1573 953 829 1542 1460 951 2119 1470 1718 869 2244 1918
## ENSG000000000457.14 774 699 352 664 681 1196 761 1446 1126 367 777 895
## ENSG000000000460.17 154 152 307 454 363 141 385 244 450 89 451 780
## ENSG000000000938.13 3092 270 113 895 6236 230 451 1059 604 2130 507 126
## PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT PRT
## ENSG000000000003.15 2326 801 3018 9134 1146 4006 4911 4238 3214 997 3342 3035

```

```

## ENSG000000000005.6      0 16   0   1   0   3   1   0  18   0   0   1
## ENSG000000000419.13 3111 469 1159 2215 1397 1433 1213 2382 1552 1227 2713 1344
## ENSG000000000457.14  969 877  444 1106  335  820 2786 1372  414  852 1920 1052
## ENSG000000000460.17 1208 377  323  770  343  258  753  363  316  527  953  157
## ENSG000000000938.13  382  86 1040 1592  509 2105  858  684 1211 1034 4108  382
##
##      PRT  PRT  STN  PRT  PRT  PRT  PRT  STN  PRT  PRT  PRT  PRT
## ENSG000000000003.15 2492 2235 1028 9980  943 3950 2721 1509 2319 7249 1588 2261
## ENSG000000000005.6    2  18   3   0   0  10   0   2   0   0   1   1
## ENSG000000000419.13 3880  918 1206 2705 1236  608 1993 1306 2237 3144  816  815
## ENSG000000000457.14 1100  488  486 1459 1215  375  824  361  805 1768  359 1190
## ENSG000000000460.17  204  187  105 1699  390  210  461   59  762 1359  193  285
## ENSG000000000938.13 1036 1107 4093  884 1161  595  658 1744 3807  798  783  599
##
##      PRT  PRT  PRT  PRT  PRT  PRT  STN  PRT  PRT  PRT  PRT  PRT
## ENSG000000000003.15 9079 3254 8815 2472 3257 2421 1200 2698 5010 4736 2119 2050
## ENSG000000000005.6    3   1   2   2   0   1   2   0  917   2   0   4
## ENSG000000000419.13 3519 1781 3824 2471 1092 1617  909  738  640 1460 1026 1410
## ENSG000000000457.14 1630  725 1011  755  688  814  373  611  522 1936  752  254
## ENSG000000000460.17 1097  556  751  209  262  415   70  203  275  486  189  203
## ENSG000000000938.13 1745 1467  920 1891  362  380 2123  862  649 1481  539 1652
##
##      PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT
## ENSG000000000003.15 1121 2687 3825 10673 4586 5009 3014 2604 2714  769 4548 2279
## ENSG000000000005.6    25   3   0   3   0  13   0   1   0   7   4   0
## ENSG000000000419.13 1434 1157 1177  2002 2629 2731 2149 1848 1785  725 1531 1583
## ENSG000000000457.14  809  443  400   605 1512  919  843  914  711  536  704  804
## ENSG000000000460.17  236  345  449   335  311  321  663  511  571  109  295  303
## ENSG000000000938.13 4004 1059  367  1185  266 2299 1084  421 1693 1204  376 2305
##
##      PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT
## ENSG000000000003.15 5471 3709 7006 3577 9713 2376 3580 1686 2966 1889 2336 1398
## ENSG000000000005.6    0   0   0  13   0   0   1   0   0   0   1   0
## ENSG000000000419.13  617 1283 2860 1206 1640 1027 2836 1064 1346 1941 1095 1287
## ENSG000000000457.14  390 1022  983  263  382  599  533 1167 1050  548  715  656
## ENSG000000000460.17  223  452  883  172  475  333  433  329  376  301  286  303
## ENSG000000000938.13 1079  830  968  758  648  719  533 1603  434  879  412  739
##
##      PRT  STN  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT
## ENSG000000000003.15 3623 2183 3091 5889 1676 1113 2257 3859 3647 3719 1622 1419
## ENSG000000000005.6    0   3   0   0   1   0   0   2   1   0   0   0
## ENSG000000000419.13 1388 1573 1757 2845  856 1180  644 1149 1123 2174 1300 1693
## ENSG000000000457.14  688  480  740  476  361  982  538  600  569  677  341  541
## ENSG000000000460.17  452   87  813  252  135  642  212  123  255  386  235  508
## ENSG000000000938.13 1312 1621  309  742  638  361  616 1524 1827  117  281  557
##
##      PRT  PRT  PRT  STN  STN  PRT  PRT  PRT  PRT  PRT  PRT  PRT
## ENSG000000000003.15 4820 4375 2650  776 1524 2085 2633 2121 2582 12460 2182 1040
## ENSG000000000005.6    3   2   5   2   2   2   5   2   0  8122   0   1
## ENSG000000000419.13 1405 4095 1322  992 1677 1867 1583 1609  878 1295  737  859
## ENSG000000000457.14  459  934  661  307  414  605  908 1181  571  575  376  597
## ENSG000000000460.17  398  673  249   67   99  227  246  580  424  309  142  151
## ENSG000000000938.13  408  425 1125 3583 1392  659 1402 2021  324  143  387  901
##
##      PRT  STN  PRT  PRT  PRT  PRT  PRT  PRT  STN  PRT  PRT  PRT
## ENSG000000000003.15 3444 1189 4156 1766 8673 5176 1381 2096  849 3440 1538 2904
## ENSG000000000005.6    3   7   0   0   1   8   5   0   1   0   2   2
## ENSG000000000419.13 2569  709 2481 1661 3153 1899  998 3780 1237 1478 2094 1071
## ENSG000000000457.14 1977  361 1271  475 1713  856  715  698  290  625  403  437
## ENSG000000000460.17  545   54 1147   95 1067  441  195  594   95  749  415  260
## ENSG000000000938.13  678 1815  547  368 1226 4358 1337  254 4712  227  433  713

```

```

##          PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  STN
## ENSG000000000003.15 1028 3630 9390 3399 7079 1213 4509 8260 3663 6122 1155 1275
## ENSG000000000005.6    1    2    5    6    1    0    3    6    3    0    0    0
## ENSG0000000000419.13 997 1122 1698 967 1899 1359 2182 3235 2084 2714 460 690
## ENSG0000000000457.14 216 399 1277 407 959 805 1523 1587 640 1042 156 440
## ENSG0000000000460.17 302 289 294 107 252 667 819 779 228 712 90 57
## ENSG0000000000938.13 991 338 1068 423 293 1226 3999 2943 1200 840 437 1439
##          PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT
## ENSG000000000003.15 4154 3729 936 4435 1523 3759 5817 3352 3334 6642 583 3659
## ENSG000000000005.6   20    0    0    1    1    3    1    0    1    1    1 188
## ENSG0000000000419.13 3191 1589 705 1059 388 2482 1664 2922 2504 2471 1436 981
## ENSG0000000000457.14 2074 2702 175 1243 329 1322 754 1147 657 806 482 456
## ENSG0000000000460.17 1090 631 164 209 69 229 605 1262 621 495 278 158
## ENSG0000000000938.13 1082 468 332 1434 432 3772 1174 1333 1294 1613 248 1645
##          PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT
## ENSG000000000003.15 1655 2475 1786 1945 1703 4279 3794 1849 8804 3136 923 1053
## ENSG000000000005.6    2    3    0    0    0    0    7    1    2    0    0    1
## ENSG0000000000419.13 895 1276 737 1268 1635 3773 1695 812 1445 1460 1785 1817
## ENSG0000000000457.14 1100 1158 435 441 380 757 1007 406 1075 950 392 598
## ENSG0000000000460.17 199 193 118 119 479 685 295 186 327 531 289 339
## ENSG0000000000938.13 814 1520 2878 1445 906 236 408 535 427 586 282 902
##          PRT  STN  PRT  PRT  PRT  PRT  STN  PRT  STN  PRT  PRT  PRT
## ENSG000000000003.15 3101 783 2624 1961 4473 3991 513 3943 2794 1548 7846 4513
## ENSG000000000005.6    0    0    0    0    0    2    1    0    0    0    0    0
## ENSG0000000000419.13 2763 783 1964 385 2253 2162 754 3162 1659 1239 2104 4348
## ENSG0000000000457.14 1255 380 431 355 1465 703 350 726 905 681 922 1080
## ENSG0000000000460.17 1518 72 198 261 348 492 73 603 204 184 268 1161
## ENSG0000000000938.13 1104 2418 634 315 1150 755 2077 1578 7083 522 1152 904
##          STN  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  PRT  STN
## ENSG000000000003.15 706 6616 4201 3352 1002 2860 10798 996 2438 3153 6052 577
## ENSG000000000005.6    1    0    1 11 10 0 36 0 0 3 3 1
## ENSG0000000000419.13 811 2332 1793 1447 284 1911 3551 737 593 2063 3726 941
## ENSG0000000000457.14 635 998 2799 1294 549 815 542 478 701 683 1336 275
## ENSG0000000000460.17 104 614 432 310 497 759 1115 275 297 375 1243 36
## ENSG0000000000938.13 2901 734 1042 1325 177 106 1698 827 450 2700 3469 2662
##          PRT  PRT  PRT  PRT  PRT  PRT  PRT
## ENSG000000000003.15 5209 1369 1267 2718 2020 3087 591
## ENSG000000000005.6    2    0    0    0    1    0 20
## ENSG0000000000419.13 1229 703 3955 1165 1419 2389 383
## ENSG0000000000457.14 666 479 362 639 891 1171 871
## ENSG0000000000460.17 358 140 210 293 813 282 462
## ENSG0000000000938.13 909 1004 492 525 322 240 204

```

```
# we reached to our tidy data set
```

```
library(DESeq2)
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
## Loading required package: BiocGenerics
```

```
##
```

```
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:dplyr':
```

```

##
##      combine, intersect, setdiff, union
## The following objects are masked from 'package:stats':
##
##      IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
##      anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##      dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##      grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##      order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##      rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##      union, unique, unsplit, which.max, which.min
##
## Attaching package: 'S4Vectors'
## The following objects are masked from 'package:dplyr':
##
##      first, rename
## The following objects are masked from 'package:base':
##
##      expand.grid, I, unname
## Loading required package: IRanges
##
## Attaching package: 'IRanges'
## The following objects are masked from 'package:dplyr':
##
##      collapse, desc, slice
## Loading required package: GenomicRanges
## Loading required package: GenomeInfoDb
## Loading required package: SummarizedExperiment
## Loading required package: MatrixGenerics
## Loading required package: matrixStats
##
## Attaching package: 'matrixStats'
## The following object is masked from 'package:dplyr':
##
##      count
##
## Attaching package: 'MatrixGenerics'
## The following objects are masked from 'package:matrixStats':
##
##      colAlls, colAnyNAs, colAnys, colAvgPerRowSet, colCollapse,
##      colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##      colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##      colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,

```



```

##      colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
##      colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##      colWeightedMeans, colWeightedMedians, colWeightedSds,
##      colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
##      rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##      rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##      rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##      rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##      rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
##      rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##      rowWeightedSds, rowWeightedVars

## Loading required package: Biobase

## Welcome to Bioconductor
##
##      Vignettes contain introductory material; view with
##      'browseVignettes()'. To cite Bioconductor, see
##      'citation("Biobase")', and for packages 'citation("pkgname)".

##
## Attaching package: 'Biobase'

## The following object is masked from 'package:MatrixGenerics':
##
##      rowMedians

## The following objects are masked from 'package:matrixStats':
##
##      anyMissing, rowMedians

con <- factor(paste0(c$Sample.Type, seq_along(c$Sample.Type)))
gr <- factor(cc$sample)
#colSums(cn)
colData <- data.frame(condition = con, group = gr , type = "paired-end")
cds <- DESeqDataSetFromMatrix(cn, colData, design = ~group)
cds <- DESeq(cds) #Constructing DESeq2 dataset

## estimating size factors
## estimating dispersions
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates
## fitting model and testing
## -- replacing outliers and refitting for 8106 genes
## -- DESeq argument 'minReplicatesForReplace' = 7
## -- original counts are preserved in counts(dds)
## estimating dispersions
## fitting model and testing
cnt <- log2(1+(counts(cds, normalize = T))) #getting normalized counts `count matrix`

```