Driver Analysis

Bachisio Ziccheddu

12/27/2021

```
library(dndscv)
library(stringr)
library(plyr)
library(dplyr)
library(tidyr)
```

dNdScv

Input file

```
snv<-read.delim("~/Box/Bachisio/Documents/Project/Team/MauraF/HL/WXS_WGS_dnds_input_fil
e.txt",
                stringsAsFactors = F)
snv1<- snv[,c("sampleID","chr","pos","ref","alt")]
colnames(snv1)<-c("Tumor_Sample","Chrom","Pos","Ref","Alt")
```

dNdScv Algorithm

```
dn<- dndscv(unique(snv1), outmats=T)
#dn$globaldnds
WGS_driver_dnds<- dn$sel_cv
WGS_pos_dnds<- WGS_driver_dnds[WGS_driver_dnds$qglobal_cv<0.1,]
#WGS_pos_dnds$gene_name
```

```
WGS_pos_dnds[1:10,]
```

```
##       gene_name n_syn n_mis n_non n_spl n_ind    wmis_cv    wnon_cv    wspl_cv
## 1822        B2M     0    19     1     2     2 170.674065 293.607742 293.607742
## 8380      IGLL5    11    31     1     0     2  16.740406   6.821669   6.821669
## 7208      GNA13     1    11     5     0     5  59.471239 351.205187 351.205187
## 16421     SOCS1     2    20     2     0    40  37.686365 174.266310 174.266310
## 17946   TNFAIP3     1     4     3     2    16   8.973277 133.843699 133.843699
## 4367     CSF2RB     0     1     5     0     3   1.705294 113.324047 113.324047
## 1947      BCL7A     3    13     0     1     1  27.349173  18.398362  18.398362
## 3831       CISH     0     3     1     1     2  16.873312 193.755381 193.755381
## 8708      ITPKB     5    17     2     0     3   7.202024  14.623059  14.623059
## 16833     STAT6     1    17     0     0     0  37.616491   0.000000   0.000000
##         wind_cv      pmis_cv    ptrunc_cv  pallsubs_cv      pind_cv
## 1822   709.3870 0.000000e+00 4.588255e-07 0.000000e+00 3.671594e-05
## 8380   403.9565 0.000000e+00 1.503414e-01 0.000000e+00 1.104354e-04
## 7208   577.0807 7.367800e-09 1.197059e-09 4.518608e-13 4.872712e-08
## 16421 8231.5660 3.773093e-12 1.050684e-04 6.713519e-13 5.443054e-59
## 17946  882.4738 1.802083e-02 8.989011e-08 4.784390e-07 1.341162e-16
## 4367   144.7809 6.239944e-01 2.840619e-08 1.349829e-07 1.749479e-04
## 1947   188.0487 1.525455e-09 5.594304e-02 6.556133e-09 5.186864e-03
## 3831   316.1399 2.222685e-03 5.207242e-05 2.132118e-05 1.774664e-04
## 8708   138.2069 1.553749e-05 1.153716e-02 1.956747e-05 1.977671e-04
## 16833    0.0000 5.362321e-09 7.547564e-01 2.257534e-08 1.000000e+00
##            qmis_cv    qtrunc_cv  qallsubs_cv   pglobal_cv   qglobal_cv
## 1822  0.000000e+00 1.843653e-03 0.000000e+00 0.000000e+00 0.000000e+00
## 8380  0.000000e+00 9.759666e-01 0.000000e+00 0.000000e+00 0.000000e+00
## 7208  2.114664e-05 2.405011e-05 3.026112e-09 0.000000e+00 0.000000e+00
## 16421 2.526840e-08 2.110930e-01 3.372033e-09 0.000000e+00 0.000000e+00
## 17946 7.999232e-01 6.019941e-04 1.068035e-03 0.000000e+00 0.000000e+00
## 4367  7.999232e-01 2.853544e-04 3.389927e-04 6.014533e-10 2.013966e-06
## 1947  7.661977e-06 9.759666e-01 2.634385e-05 8.536977e-10 2.450234e-06
## 3831  7.999232e-01 1.494553e-01 2.677273e-02 7.716113e-08 1.759706e-04
## 8708  2.601365e-02 9.759666e-01 2.677273e-02 7.882812e-08 1.759706e-04
## 16833 2.114664e-05 9.759666e-01 7.559352e-05 4.200460e-07 8.439144e-04
```

Restricted hypothesis

```
cosmic<- read.delim("~/Box/Bachisio/Desktop/Tools/Reference/UCSC/COSMIC_census.csv", sep
=",")

Rossi<- c("STAT6","ITPKB","TNFAIP3","B2M","GNA13","HIST1H1E","CIITA","IRF8","ARID1A",
            "BTG1","IRF4","PCBP1","PIM1","STAT3","ATM","BCL6","BTK","CCND3","CD58","CXCR
4",
              "ID3","KMT2D","MYC","NFKBIE","NOTCH1","PRDM1","SPEN","TET2","TNFRSF14","TP
53","TRAF3","XPO1")

Shipp<- c("SOCS1","IKBKB","GNA13","ARID1A","DNAH12","NFKBIA","ACTB","HLA-B","STAT6","RBM
38",
          "XPO1","CSF2RB","TNFAIP3","NFKBIE","B2M", "HLA-B")


HL_drivers<- unique(c(cosmic$Gene.Symbol, Shipp,Rossi))
HL_drivers<- HL_drivers[!(HL_drivers %in% c("IGH","IGK","IGL","TRA","TRB","TRD","ZNF19
8","ZNF278","AFDN","DUX4L1","HMGN2P46","KNL1","LHFPL6","MALAT1",
                                            "MRTFA","NSD2","NSD3","SHTN1","TENT5C","WDC
P"))]
final_dnds<-snv1
dn_dri <- dndscv(unique(final_dnds), outmats=T, gene_list = HL_drivers)
#dn_dri$globaldnds
dn_dris_sel<- dn_dri$sel_cv
#dn_dris_sel[dn_dris_sel$qallsubs_cv<0.1,]

####estimate # drivers per sample - all nonsynonimous - not sig

dndsoutccfNE<- dn_dri
wmis = dndsoutccfNE$globaldnds["wall","mle"]
wmis_low  = dndsoutccfNE$globaldnds["wall","cilow"]
wmis_high = dndsoutccfNE$globaldnds["wall","cihigh"]
num_mis = nrow(unique(dndsoutccfNE$annotmuts[dndsoutccfNE$annotmuts$impact!="Synonymou
s",]))
num_missense_drivers = ((wmis-1)/wmis) * num_mis
num_missense_drivers_lowci  = ((wmis_low-1) /wmis) * num_mis
num_missense_drivers_hici   = ((wmis_high-1)/wmis) * num_mis
num_missense_drivers/nrow(as.data.frame(unique(final_dnds$Tumor_Sample)))
```

```
## [1] 1.764339
```

```
num_missense_drivers_hici/nrow(as.data.frame(unique(final_dnds$Tumor_Sample)))
```

```
## [1] 4.033966
```

```
num_missense_drivers_lowci/nrow(as.data.frame(unique(final_dnds$Tumor_Sample)))
```

```
## [1] -0.1331823
```

```
####estimate # drivers per sample - all missense - not sig
wmis = dndsoutccfNE$globaldnds["wmis","mle"]
wmis_low  = dndsoutccfNE$globaldnds["wmis","cilow"]
wmis_high = dndsoutccfNE$globaldnds["wmis","cihigh"]
num_mis = nrow(unique(dndsoutccfNE$annotmuts[dndsoutccfNE$annotmuts$impact=="Missens
e",]))
num_missense_drivers = ((wmis-1)/wmis) * num_mis
num_missense_drivers_lowci  = ((wmis_low-1) /wmis) * num_mis
num_missense_drivers_hici   = ((wmis_high-1)/wmis) * num_mis
num_missense_drivers/nrow(as.data.frame(unique(final_dnds$Tumor_Sample)))
```

```
## [1] 0.8974077
```

```
num_missense_drivers_hici/nrow(as.data.frame(unique(final_dnds$Tumor_Sample)))
```

```
## [1] 2.554394
```

```
num_missense_drivers_lowci/nrow(as.data.frame(unique(final_dnds$Tumor_Sample)))
```

```
## [1] -0.4832407
```

```
####estimate # drivers per sample - all trunc - not sig
wmis = dndsoutccfNE$globaldnds["wtru","mle"]
wmis_low  = dndsoutccfNE$globaldnds["wtru","cilow"]
wmis_high = dndsoutccfNE$globaldnds["wtru","cihigh"]
num_mis = nrow(unique(dndsoutccfNE$annotmuts[dndsoutccfNE$annotmuts$impact!="Missense" &
dndsoutccfNE$annotmuts$impact!="Synonymous",]))
num_missense_drivers = ((wmis-1)/wmis) * num_mis
num_missense_drivers_lowci  = ((wmis_low-1) /wmis) * num_mis
num_missense_drivers_hici   = ((wmis_high-1)/wmis) * num_mis
num_missense_drivers/nrow(as.data.frame(unique(final_dnds$Tumor_Sample)))
```

```
## [1] 1.330967
```

```
num_missense_drivers_hici/nrow(as.data.frame(unique(final_dnds$Tumor_Sample)))
```

```
## [1] 2.450127
```

```
num_missense_drivers_lowci/nrow(as.data.frame(unique(final_dnds$Tumor_Sample)))
```

```
## [1] 0.4955527
```

```
# plot
# Function to calculate confidence intervals for dN/dS values per gene under the dNdScv
model using profile likelihood.

dn_dri_driver_dnds<- geneci(dn_dri, gene_list = HL_drivers, level = 0.95)
#dn_dri_driver_dnds[dn_dri_driver_dnds$mis_low>1 | dn_dri_driver_dnds$tru_low>1,]

data("knownhotspots_hg19", package="dndscv") # Previously known hotspots
known_hotspots = known_hotspots[sapply(known_hotspots, function(x) strsplit(x, split
=":")[[1]][5]) %in%  HL_drivers] # Known hotspots in targetgenes
hotspots_siteRHT_81_sd = sitednds(dn_dri, site_list = known_hotspots, min_recurr = 1, me
thod = "LNP")
hotspot_HL<- hotspots_siteRHT_81_sd$recursites
#length(unique(final_dnds$Tumor_Sample))
#hotspot_HL

all_driver_HL<- unique(c(dn_dri_driver_dnds[dn_dri_driver_dnds$mis_low>1 | dn_dri_driver
_dnds$tru_low>1,]$gene,
                         dn_dris_sel[dn_dris_sel$qallsubs_cv<0.1,]$gene_name, WGS_pos_dn
ds$gene_name, hotspot_HL$gene))

all_driver_HL<- unique(c(dn_dris_sel[dn_dris_sel$qallsubs_cv<0.1,]$gene_name, WGS_pos_dn
ds$gene_name, hotspot_HL$gene))
all_driver_HL<- all_driver_HL[all_driver_HL!="FANCF"]


Shipp_u<- Shipp[!Shipp %in% all_driver_HL]
missed<- WGS_driver_dnds[WGS_driver_dnds$gene_name %in% Shipp_u,]

#hotspot_HL[order(hotspot_HL$gene),]
```

```
print(missed)
```

```
##       gene_name n_syn n_mis n_non n_spl n_ind  wmis_cv  wnon_cv  wspl_cv
## 11342    NFKBIA     1     0     0     2     1 0.000000 52.13431 52.13431
## 7949      HLA-B     2     0     2     0     1 0.000000 40.30555 40.30555
## 513         ACTB     2     8     0     0     1 8.013095  0.00000  0.00000
## 1465     ARID1A     0     2     0     0     3 1.824467  0.00000  0.00000
## 8398      IKBKB     0     2     0     0     0 5.671292  0.00000  0.00000
## 5159     DNAH12     0     0     0     0     0 0.000000  0.00000  0.00000
## 14291     RBM38     0     0     0     0     0 0.000000  0.00000  0.00000
##         wind_cv     pmis_cv   ptrunc_cv pallsubs_cv     pind_cv   qmis_cv
## 11342 137.19277 0.384900269 0.001749545 0.003105435 0.007045850 0.7999232
## 7949  120.18540 0.300512584 0.002237911 0.003141243 0.008005486 0.7999232
## 513   116.03005 0.002942151 0.749588790 0.010064717 0.008281072 0.7999232
## 1465   57.25367 0.439505772 0.688624949 0.684494150 0.001732905 0.7999232
## 8398    0.00000 0.062417928 0.752188333 0.164690899 1.000000000 0.7999232
## 5159    0.00000 0.103075403 0.568754469 0.225139118 1.000000000 0.7999232
## 14291   0.00000 0.483666056 0.865295645 0.771297351 1.000000000 0.7999232
##       qtrunc_cv qallsubs_cv   pglobal_cv qglobal_cv
## 11342 0.9759666   0.9979919 0.0002566557  0.1983257
## 7949  0.9759666   0.9979919 0.0002914751  0.2168898
## 513   0.9759666   0.9979919 0.0008661802  0.5438258
## 1465  0.9759666   0.9979919 0.0091773844  1.0000000
## 8398  0.9759666   0.9979919 0.4617413861  1.0000000
## 5159  0.9759666   0.9979919 0.5608298204  1.0000000
## 14291 0.9759666   0.9979919 0.9715888579  1.0000000
```

```
all_driver_HL
```

```
##  [1] "SOCS1"    "B2M"      "GNA13"    "TNFAIP3"  "CSF2RB"   "STAT6"
##  [7] "NONO"     "IGLL5"    "BCL7A"    "CISH"     "ITPKB"    "FAM230A"
## [13] "MSL2"     "HIST1H1E" "NFKBIE"   "PTPN1"    "SPDYE1"   "P2RY8"
## [19] "TMSB4X"   "MAL2"     "RAB19"    "MFHAS1"   "XPO1"     "TP53"
## [25] "EZH2"     "CCND3"
```

```
hotspot_HL
```

```
##    chr       pos ref mut  gene aachange   impact ref3_cod mut3_cod freq
## 1    2  61719472   C   T  XPO1    E571K Missense      CGA      CAA    5
## 2   15  45003779   T   C   B2M     L12P Missense      CTA      CCA    3
## 3   15  45003745   A   T   B2M      M1L Missense      GAT      GTT    2
## 4   15  45003746   T   A   B2M      M1K Missense      ATG      AAG    2
## 5   17   7577094   G   A  TP53    R282W Missense      CCG      CTG    2
## 6    7 148508727   T   A  EZH2    Y646F Missense      TAC      TTC    1
## 7    6  41903688   A   T CCND3    I290K Missense      ATA      AAA    1
## 8   17   7578235   T   A  TP53    Y205F Missense      TAT      TTT    1
## 9   17   7578406   C   T  TP53    R175H Missense      CGC      CAC    1
## 10  17   7577538   C   T  TP53    R248Q Missense      CGG      CAG    1
##              mu       dnds         pval         qval
## 1  8.813401e-04  5673.1791 9.536185e-15 1.089032e-11
## 2  1.438717e-03  2085.1915 4.958491e-09 2.831299e-06
## 3  4.293850e-04  4657.8247 1.989939e-07 7.575034e-05
## 4  1.031902e-03  1938.1694 1.147125e-06 3.275041e-04
## 5  2.420248e-03   826.3617 6.283328e-06 1.435112e-03
## 6  3.001536e-05 33316.2707 3.001421e-05 5.712704e-03
## 7  1.284294e-04  7786.3769 1.284108e-04 1.917261e-02
## 8  1.343293e-04  7444.3924 1.343090e-04 1.917261e-02
## 9  1.593795e-03   627.4332 1.591047e-03 2.018862e-01
## 10 2.207542e-03   452.9925 2.202281e-03 2.515005e-01
```

MutSigCV

Input file

```
annotated<-read.delim("~/Box/Bachisio/Documents/Project/Team/MauraF/HL/WXS_WGS_dnds_inpu
t_annot.hg19_multianno.txt",
                      stringsAsFactors = F)
annotated$ExonicFunc.refGene<-ifelse(annotated$ExonicFunc.refGene==".","Intron",annotate
d$ExonicFunc.refGene)
annotated.1<-separate(annotated,col = "Gene.refGene",into=c("Hugo_Symbol","rem"),sep
=";")
#head(annotated.1)
annotated<-annotated.1


annotated.1<-annotated[,c("Hugo_Symbol","Otherinfo10","Chr","Start","End","Ref","Alt","E
xonicFunc.refGene")]
colnames(annotated.1)<-c("Hugo_Symbol","Tumor_Sample_Barcode","Chromosome","Start_positi
on","End_position",
                         "Reference_Allele","Tumor_Seq_Allele2","Variant_Classificatio
n")
annotated.1$gene<-annotated.1$Hugo_Symbol
annotated.1$patient<-annotated.1$Tumor_Sample_Barcode
annotated.1$effect<-ifelse(annotated.1$Variant_Classification=="synonymous SNV","silen
t",NA)
annotated.1$effect<-ifelse(annotated.1$Variant_Classification%in%c("frameshift deletio
n","frameshift insertion","frameshift substitution",
                                                                   "nonframeshift deleti
on","nonframeshift insertion",
                                                                   "nonframeshift substi
tution","stopgain","stoploss"),"null",
                           annotated.1$effect)
annotated.1$effect<-ifelse(annotated.1$Variant_Classification=="nonsynonymous SNV","nons
ilent",annotated.1$effect)
annotated.1$effect<-ifelse(annotated.1$Variant_Classification=="Intron","noncoding",anno
tated.1$effect)
annotated.1$effect<-ifelse(annotated.1$Variant_Classification=="unknown",NA,annotated.1
$effect)
annotated.1$chr<-annotated.1$Chromosome
annotated.1$start<-annotated.1$Start_position
annotated.1$ref_allele<-annotated.1$Reference_Allele

annotated.2<-annotated.1[!is.na(annotated.1$effect),]

write.table(annotated.2,
            "~/Box/Bachisio/Documents/Project/Team/MauraF/HL/WXS_WGS_inputMutSigCV.maf",
            col.names = T,
            row.names = F,
            quote = F,
            sep="\t")
```

MutSigCV Algorithm

```
~/Desktop/Tools/MutSigCV/MutSigCV_1.41/run_MutSigCV.sh ~/Desktop/Tools/MutSigCV/mcr-8.3-
0/share/mcr-8.3-0/v83 \
~/Documents/Project/HL/WXS_WGS_inputMutSigCV.maf \
~/Desktop/Tools/MutSigCV/exome_full192.coverage.txt \
~/Desktop/Tools/MutSigCV/gene.covariates.txt \
~/Documents/Project/HL/Driver_Analysis/MutSigCV_output \
~/Desktop/Tools/MutSigCV/mutation_type_dictionary_file.txt \
~/Desktop/Tools/MutSigCV/chr_files_hg19
```

OncodriverFML

Input file

```
snv2<- snv[,c("chr","pos","ref","alt","sampleID")]
colnames(snv2)<-c("CHROMOSOME","POSITION","REF","ALT","SAMPLE")

write.table(snv2,
            "~/Box/Bachisio/Documents/Project/Team/MauraF/HL/WXS_WGS_oncodriverFML.txt",
            col.names = T,
            row.names = F,
            quote = F,
            sep="\t")
```

OncodriverFML algorithm

```
oncodrivefml \
--input ~Documents/Project/HL/WXS_WGS_oncodriverFML.txt \
--elements ~/Desktop/Tools/oncodrivefml/example/cds.tsv.gz \
--sequencing wgs \
--output  ~/Documents/Project/HL/Driver_Analysis/oncodriveFML_output
```