Functional burdening analysis of cancer genomes

PCAWG Steering committee presentation

09/12/2016

Aim and deliverables for the functional impact paper

Decipher overall functional burdening in cancer genomes in the PCAWG project.

- Avg cancer has ~10 drivers & ~5000 mutations. What is the overall burdening of the many passengers in different cancers ?

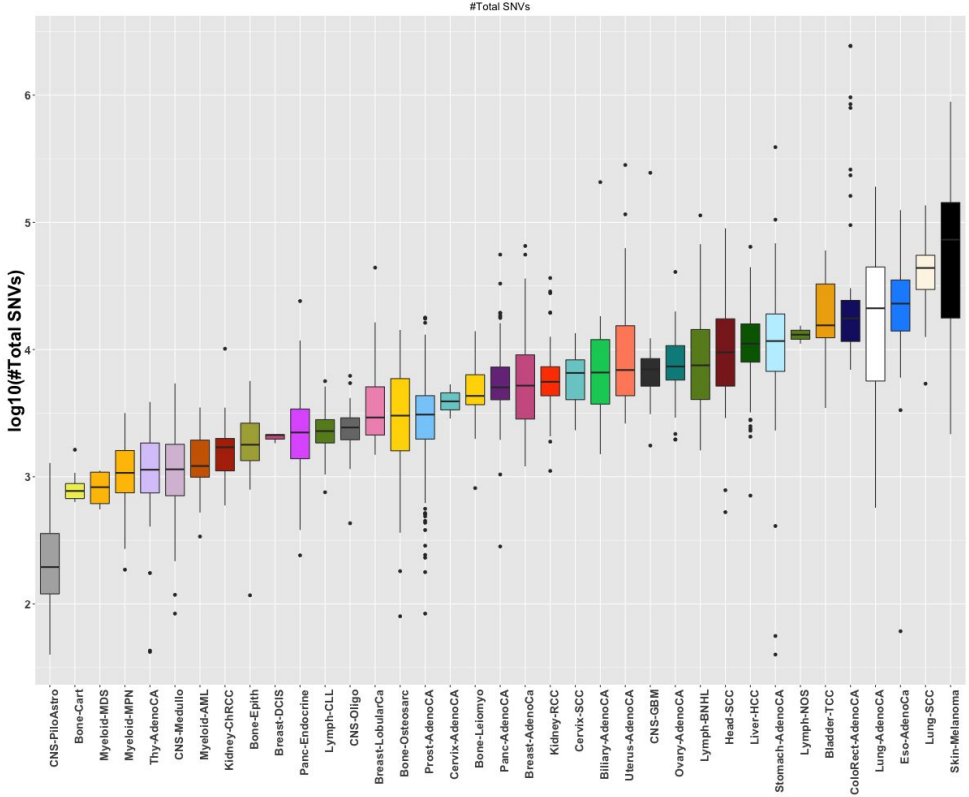
- Look at Overall variation burden observed in various genomic elements (coding & noncoding) in different PCAWG cohorts.
  - ➢ Comparison between real and simulated data to highlight genomic elements with significant burden from passengers in different cohorts
  - ➢ This work will provide **comprehensive functional annotations across all of pcawg** (FunSeq & aloft score)

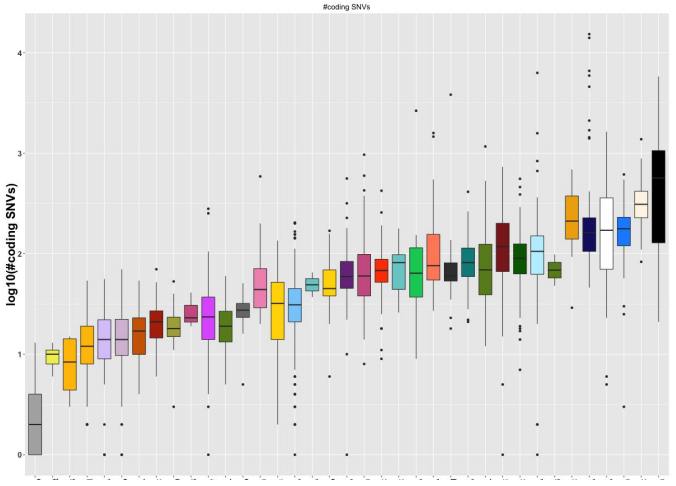
- Coding and noncoding functional impact score distribution across pan-cancer cohorts.
  - ➢ Enrichment/depletion of high impact passengers (other than drivers) in gene block/neighborhood
  - ➢ Correlation of passenger burdening with downstream gene expression changes
  - ➢ Framework to evaluate structural variation impact score

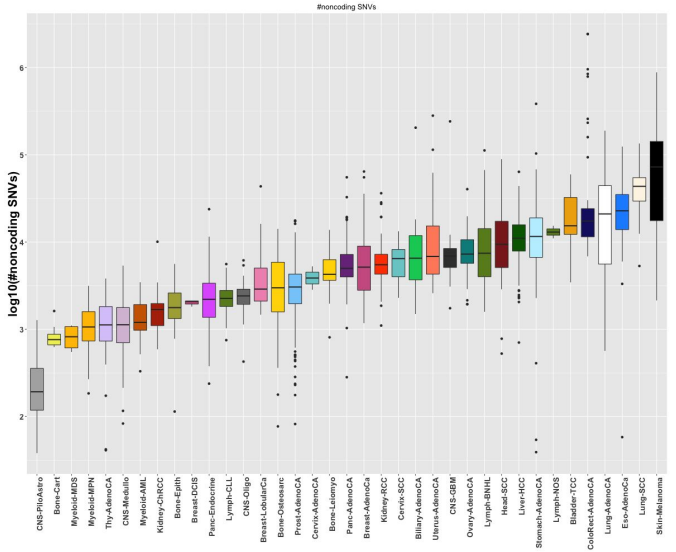
- Comparison between somatic and germline variation burdening
  - ➢ Investigate influence of germline mutational burden on the somatic genome variation profile

- Decipher the the differential passenger burdening in various cohorts (how it relates to mechanism)
  - ➢ Relate to different Signature, Ageing, sub-clonality & other clinical information
  - ➢ Any other suggestions ?

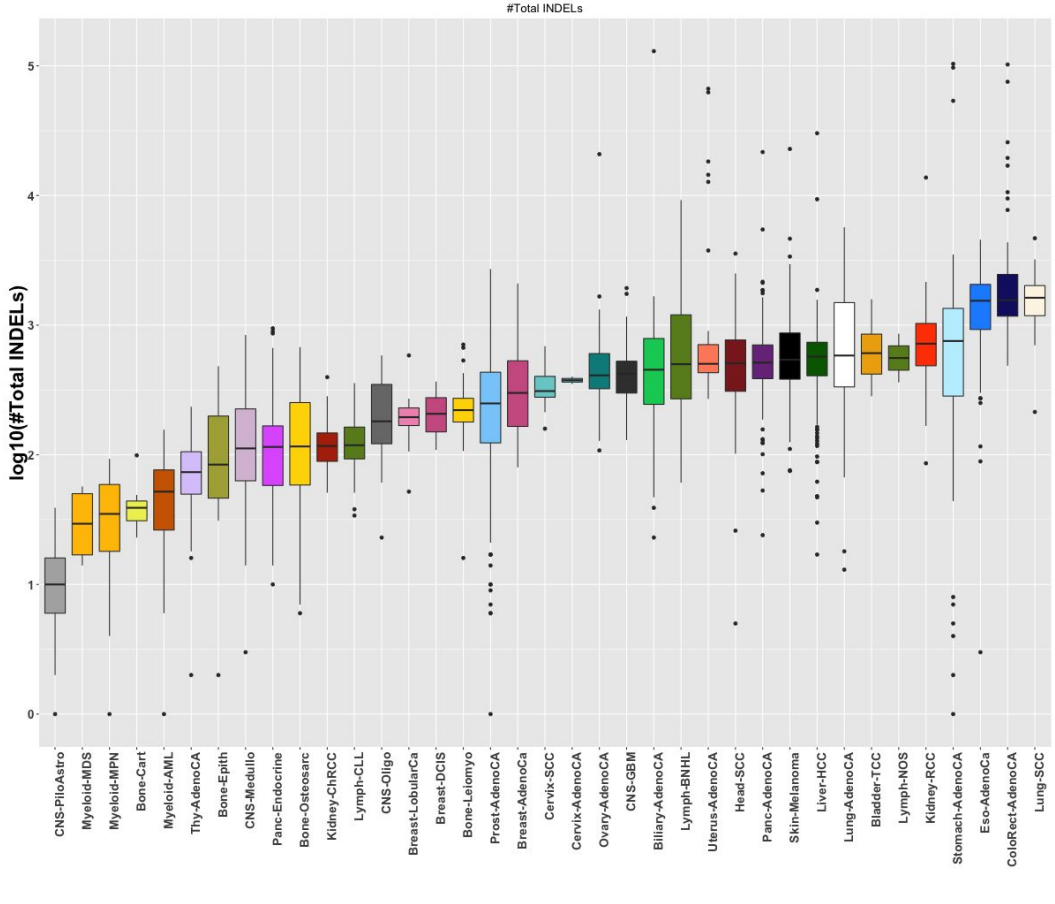
A backdrop
PCAWG SNV annotation overview

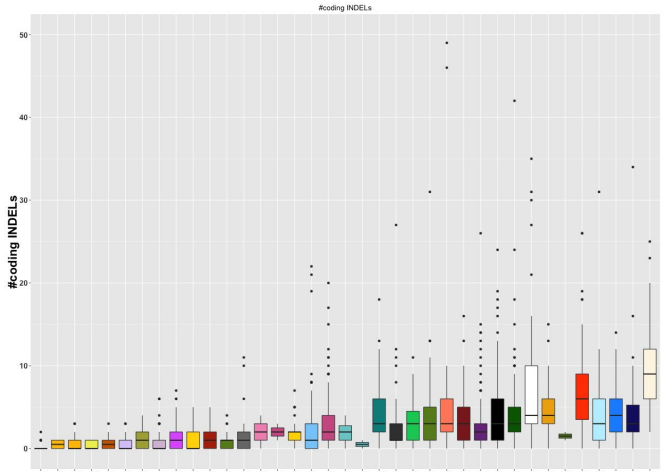


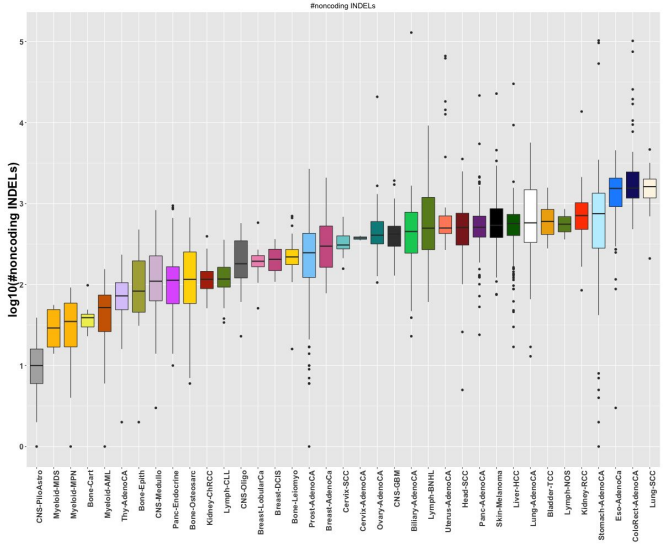




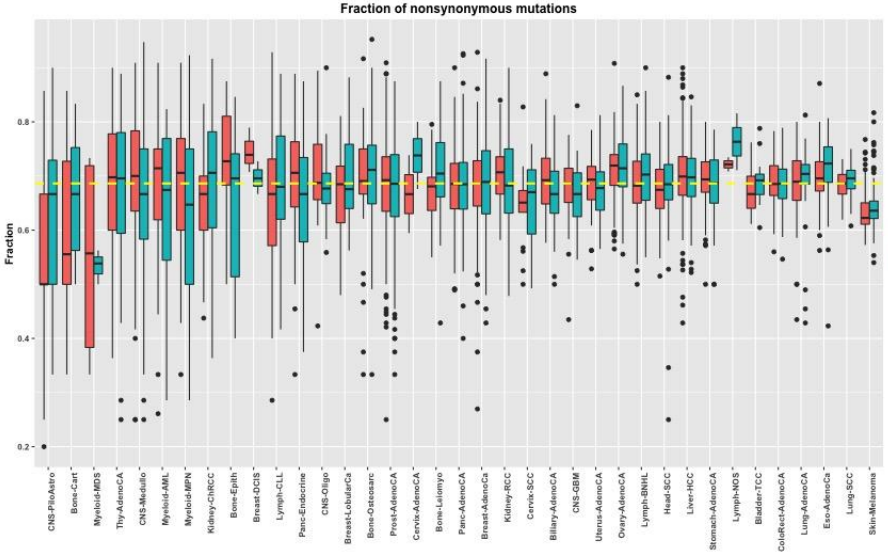
A backdrop PCAWG INDEL annotation overview

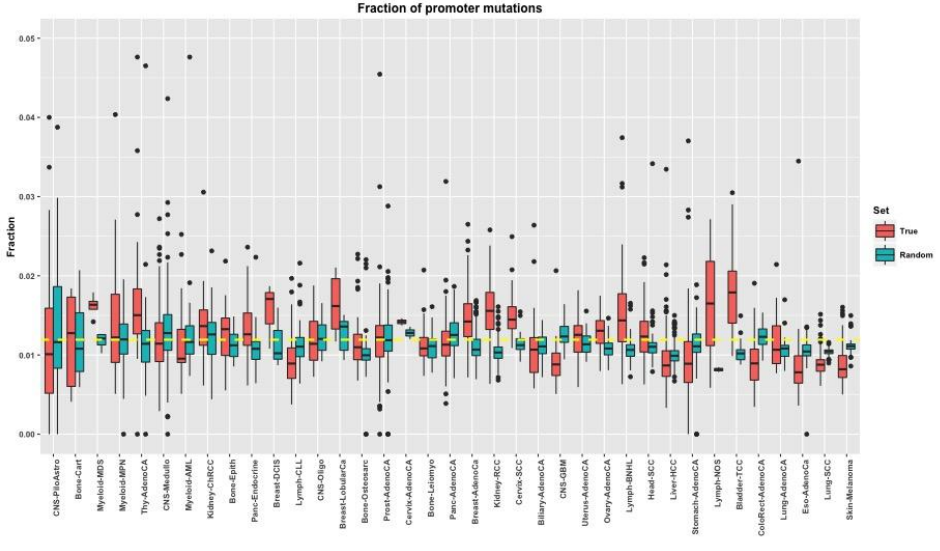




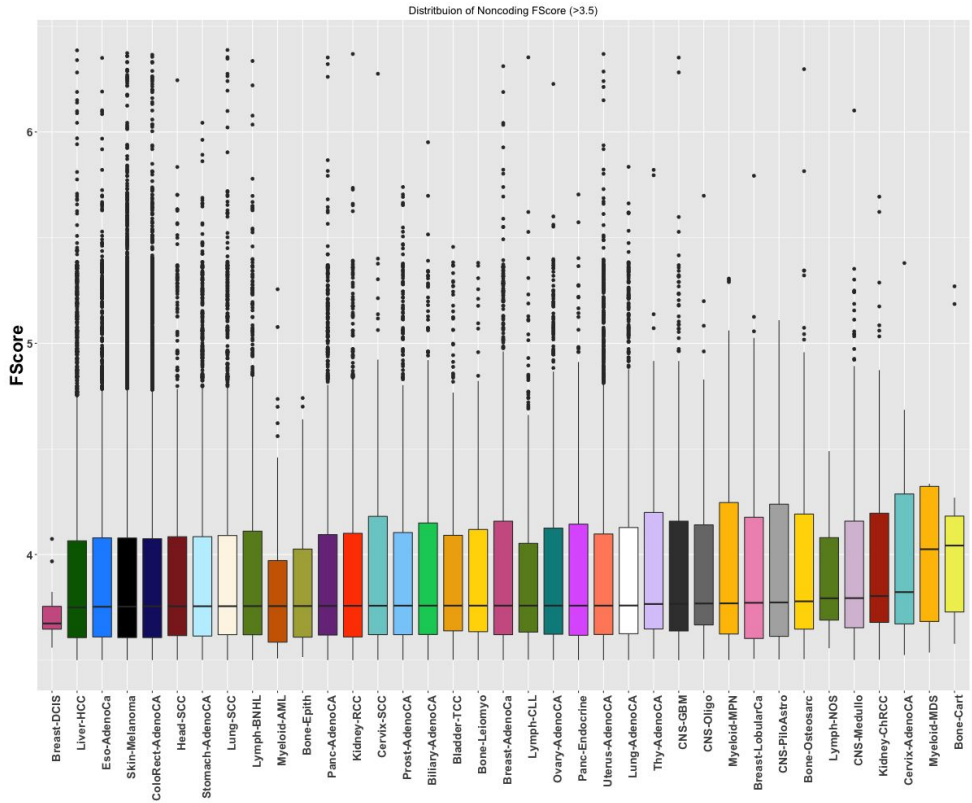


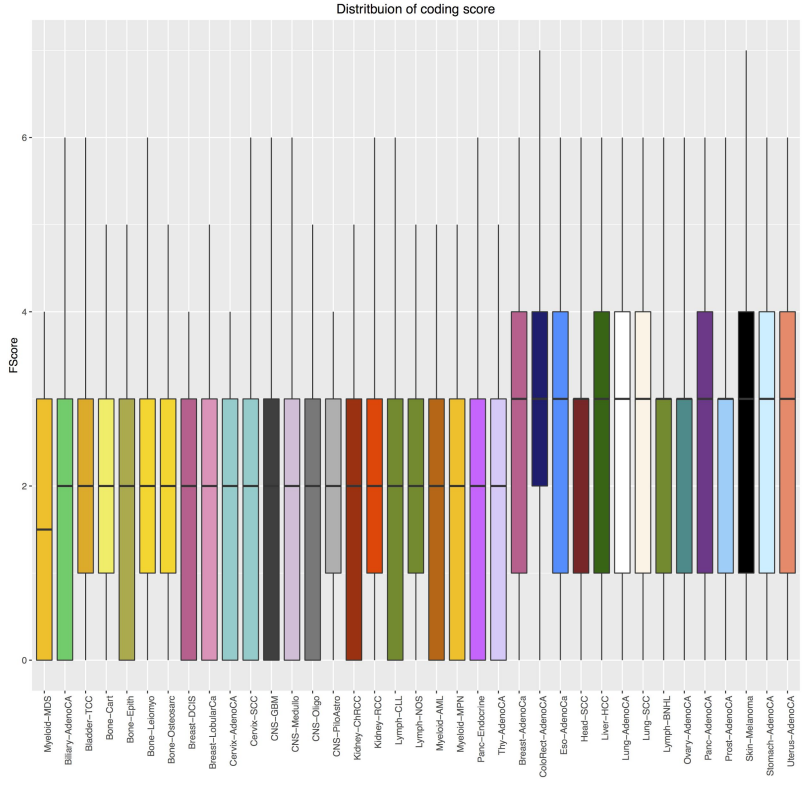
Comparison between original and randomized data set



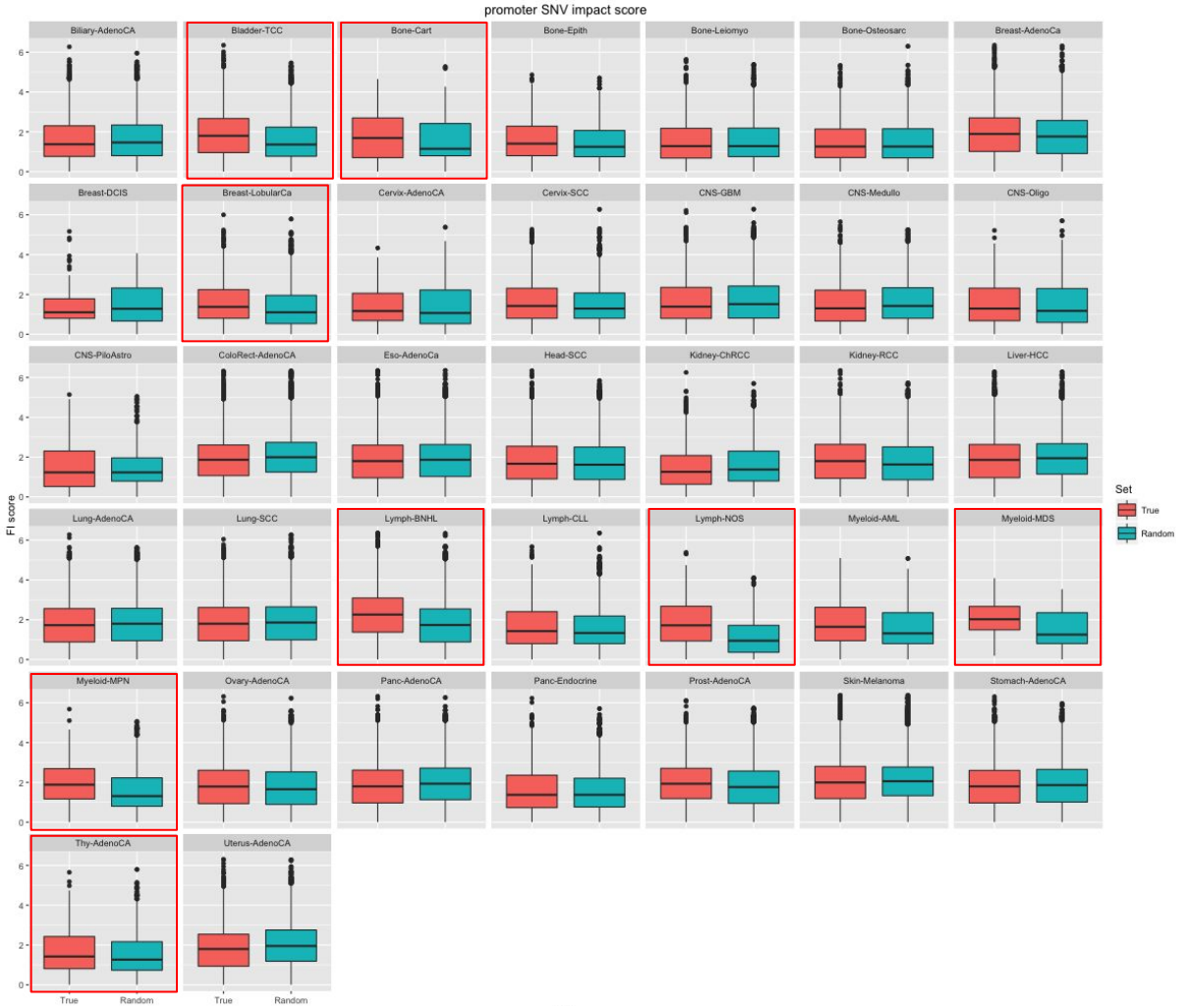


Functional impact score distribution of noncoding and coding SNVs





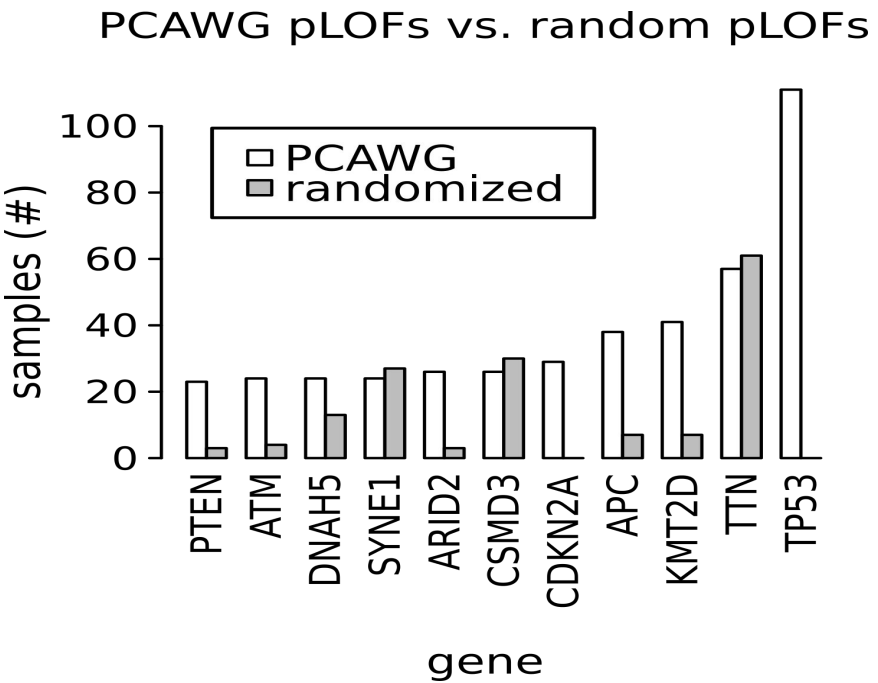
Functional impact score distribution of promoter SNVs



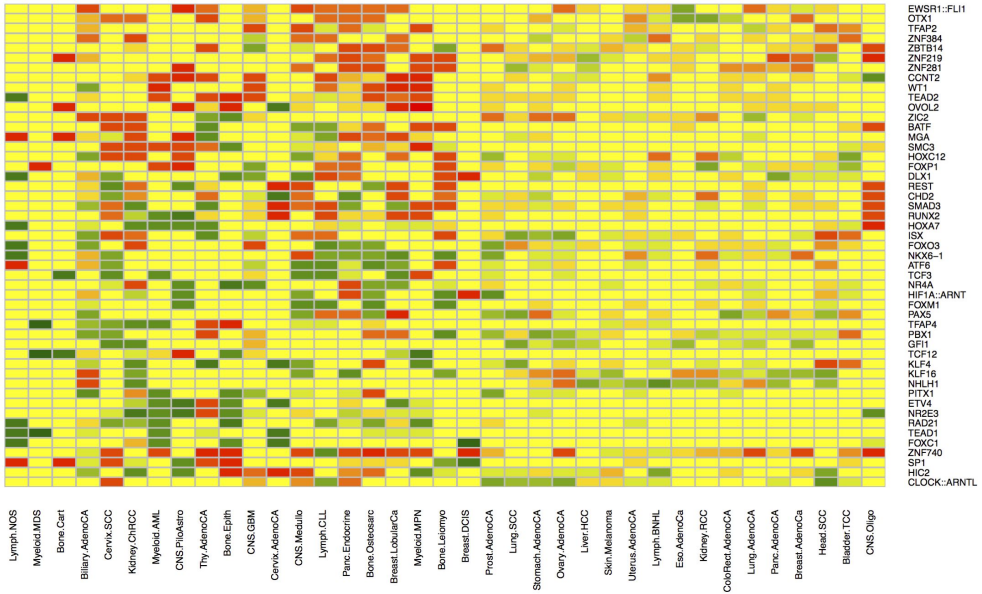
Functional impact score distribution of nonsynonymous SNVs

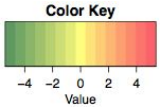
Loss of Function inducing SNVs in the PCAWG data

predicted loss of function (pLOF) mutation
Total LOF events is 28426
PCAWG pLOF mutations: 15435
#samples with at least 1 pLOF event = 2270



Functional burden of SNVs influencing TF motifs

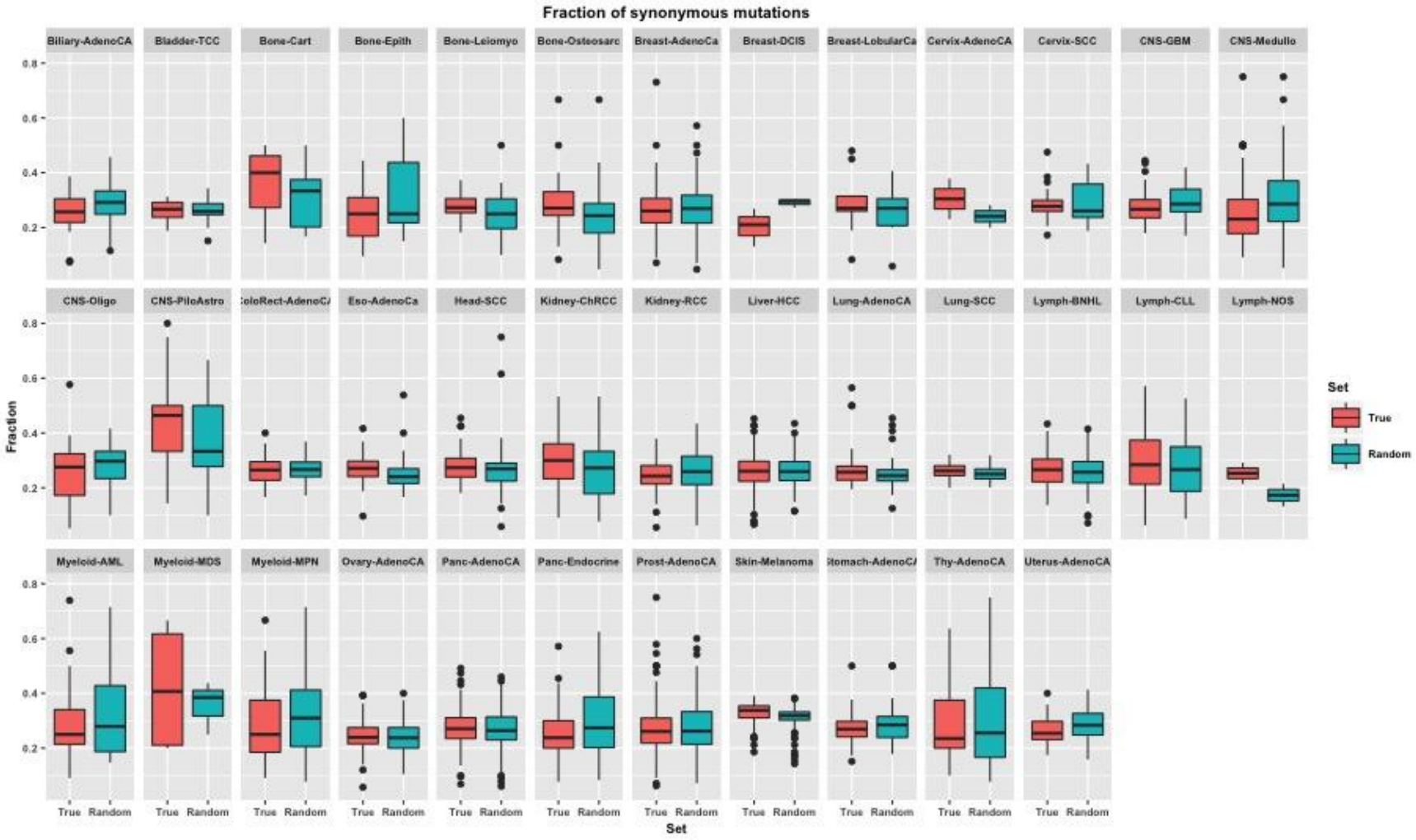


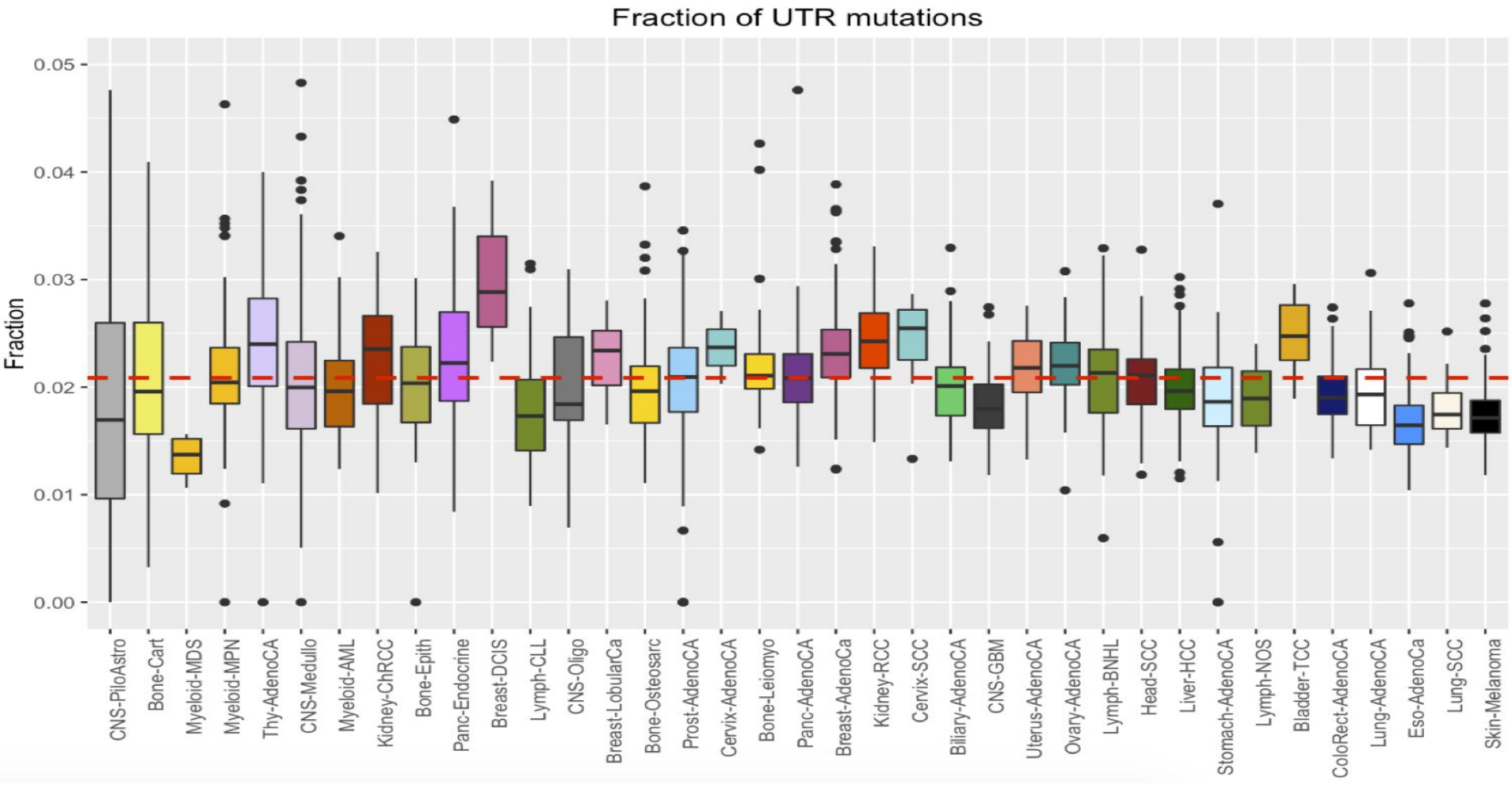


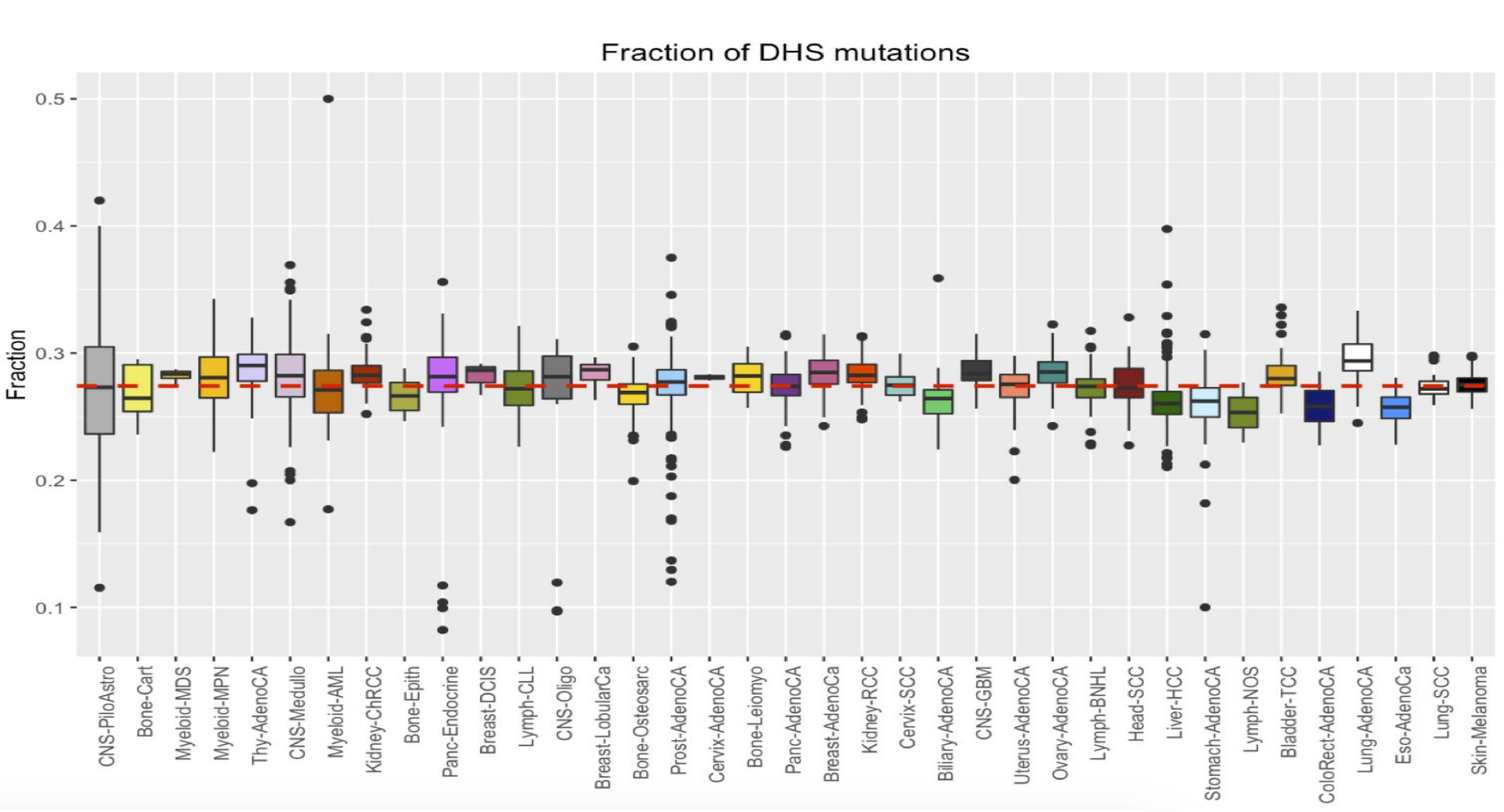
Motif Breaking

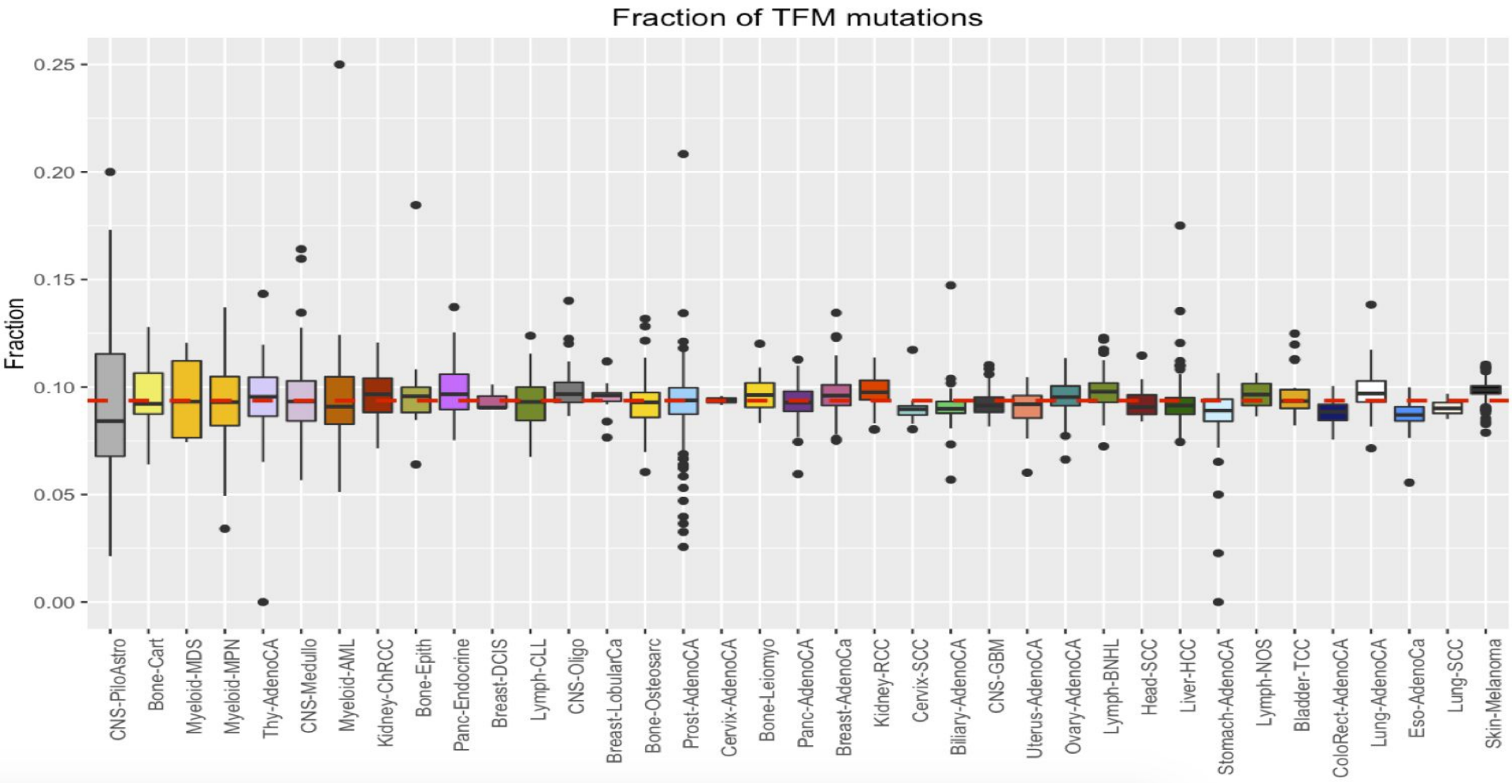
Motif Gaining

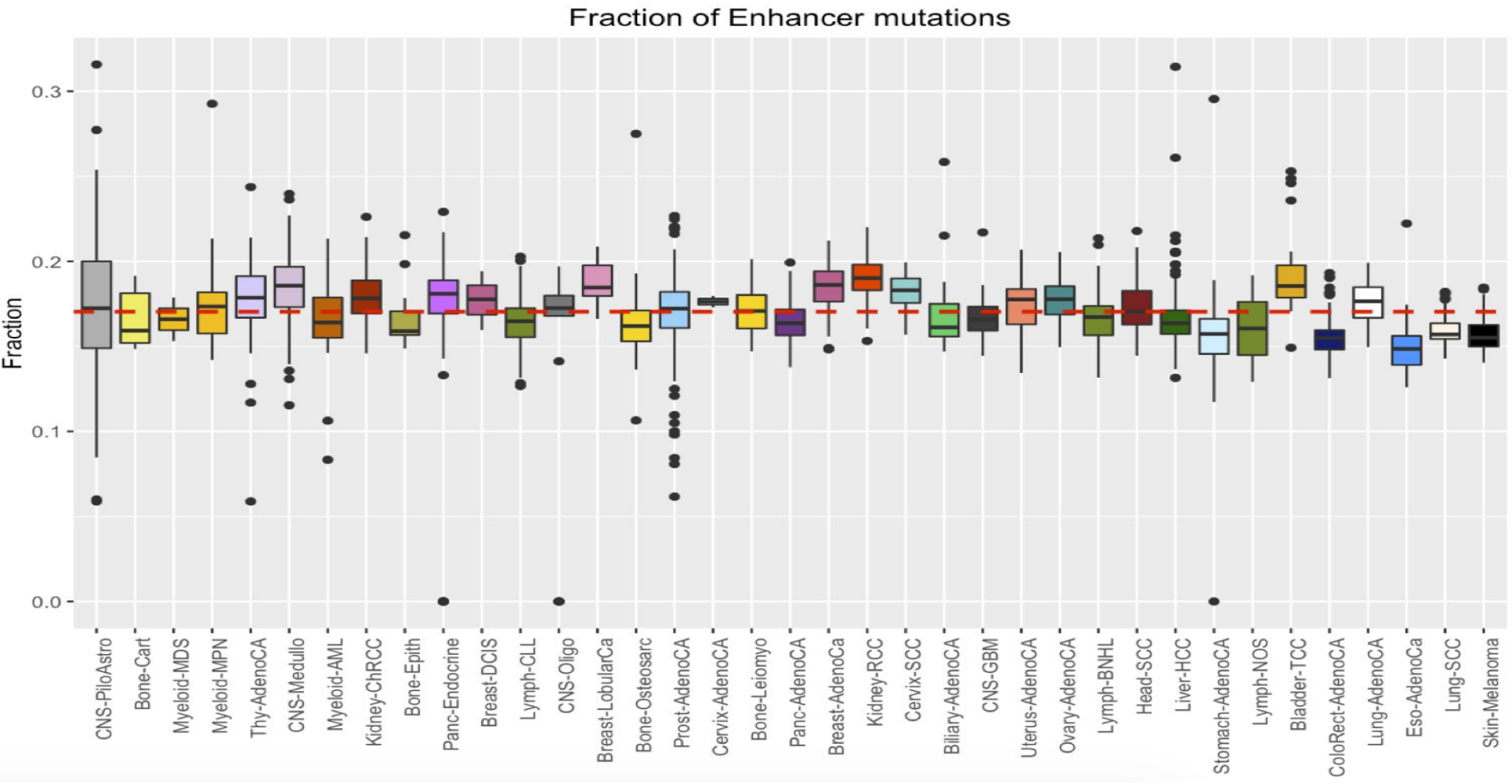
Extra Slides

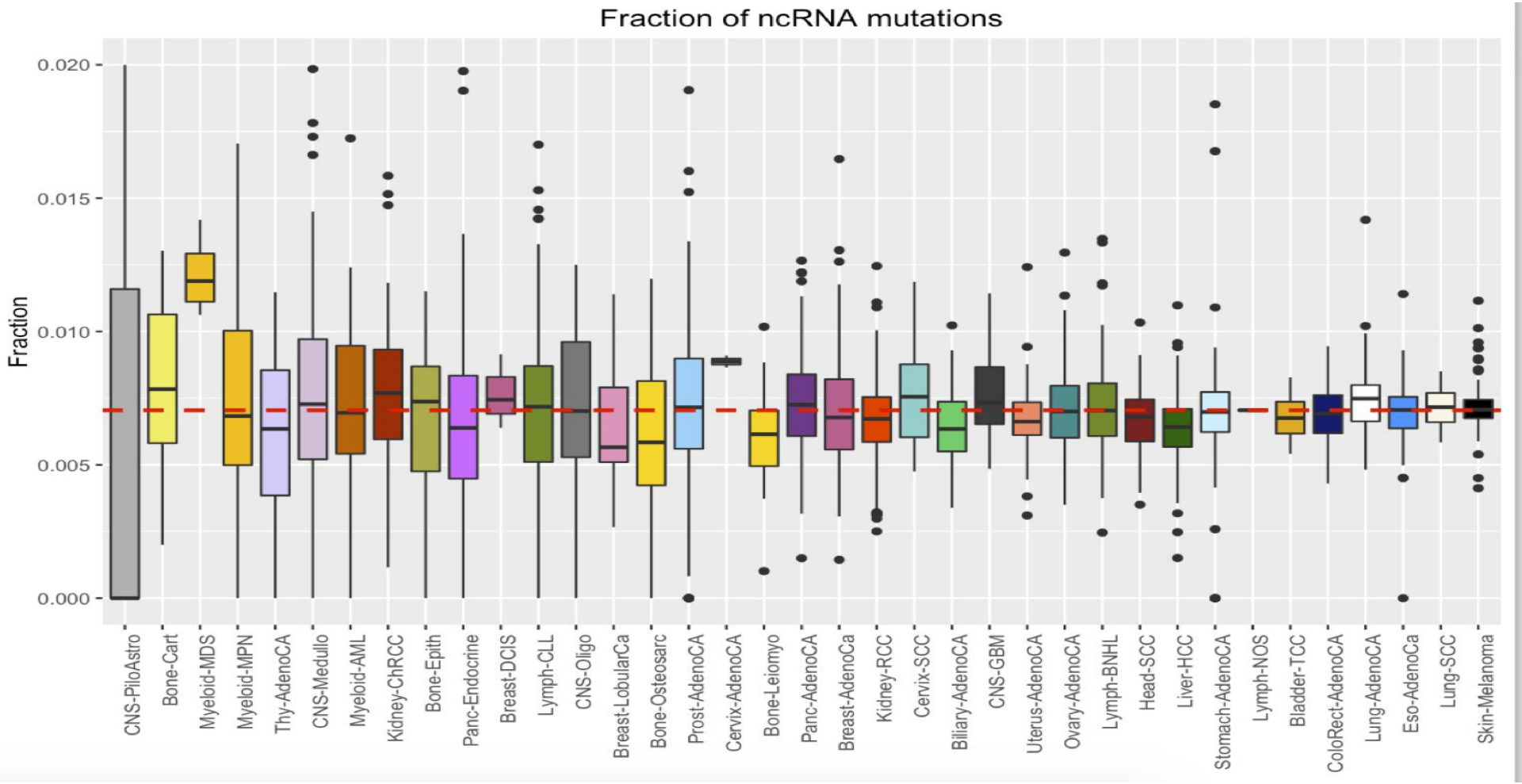

Fraction of synonymous mutations
Biliary-AdenoCA
Bladder-TCC
Bone-Cart
Bone-Epith
Bone-Leiomyo
Bone-Osteosarc
Breast-AdenoCa
Breast-DCIS
Breast-LobularCa
Cervix-AdenoCA
Cervix-SCC
CNS-GBM
CNS-Medullo
CNS-Oligo
CNS-PiloAstro
ColoRect-AdenoCA
Eso-AdenoCa
Head-SCC
Kidney-ChRCC
Kidney-RCC
Liver-HCC
Lung-AdenoCA
Lung-SCC
Lymph-BNHL
Lymph-CLL
Lymph-NOS
Myeloid-AML
Myeloid-MDS
Myeloid-MPN
Ovary-AdenoCA
Panc-AdenoCA
Panc-Endocrine
Prost-AdenoCA
Skin-Melanoma
Stomach-AdenoCA
Thy-AdenoCA
Uterus-AdenoCA
Fraction
0.8
0.6
0.4
0.2
True
Random
Set
Set
True
Random

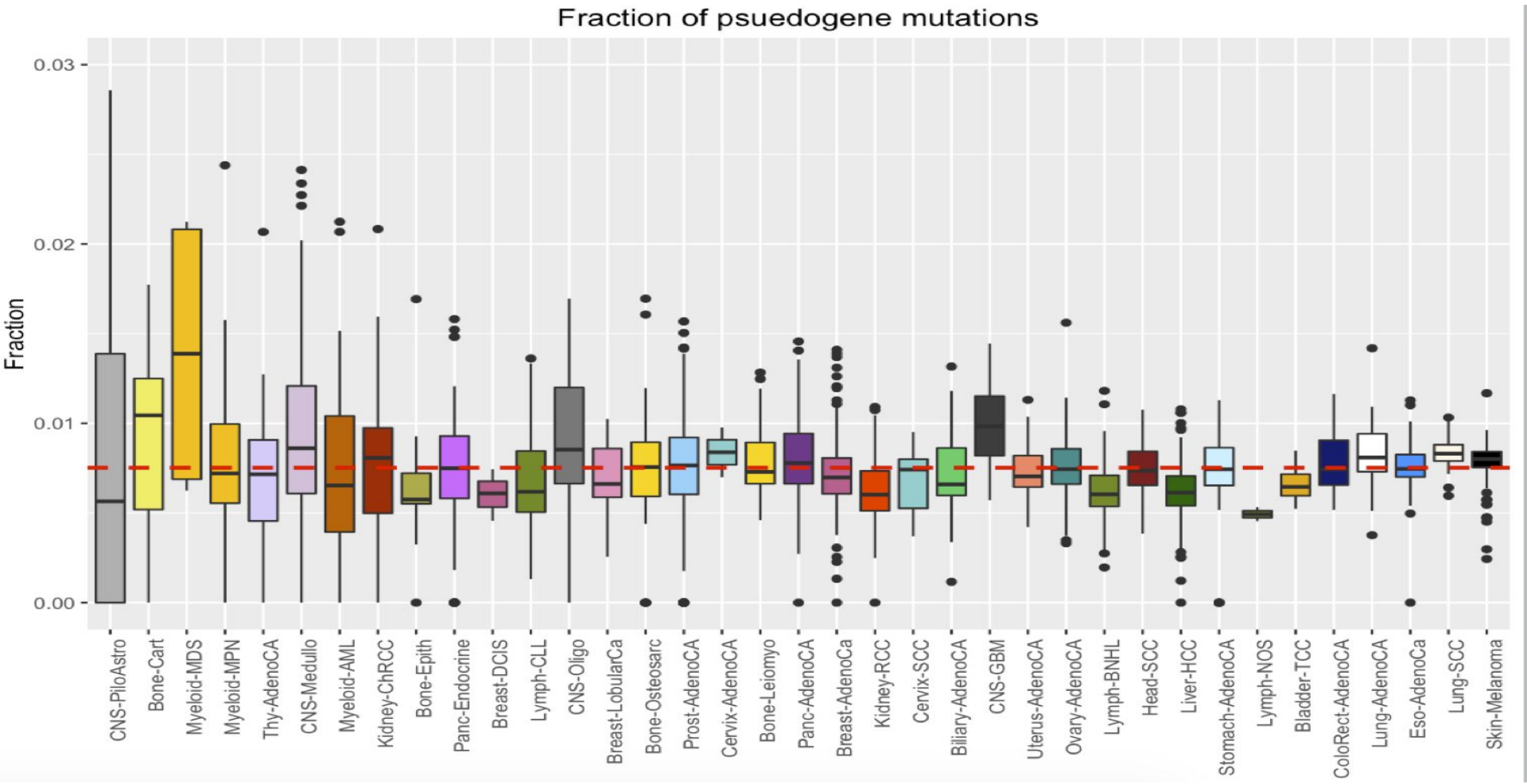

Fraction of UTR mutations
Fraction
0.05
0.04
0.03
0.02
0.01
0.00
CNS-PiloAstro
Bone-Cart
Myeloid-MDS
Myeloid-MPN
Thy-AdenoCA
CNS-Medullo
Myeloid-AML
Kidney-ChRCC
Bone-Epith
Panc-Endocrine
Breast-DCIS
Lymph-CLL
CNS-Oligo
Breast-LobularCa
Bone-Osteosarc
Prost-AdenoCA
Cervix-AdenoCA
Bone-Leiomyo
Panc-AdenoCA
Breast-AdenoCa
Kidney-RCC
Cervix-SCC
Biliary-AdenoCA
CNS-GBM
Uterus-AdenoCA
Ovary-AdenoCA
Lymph-BNHL
Head-SCC
Liver-HCC
Stomach-AdenoCA
Lymph-NOS
Bladder-TCC
ColoRect-AdenoCA
Lung-AdenoCA
Eso-AdenoCa
Lung-SCC
Skin-Melanoma


Fraction of DHS mutations
Fraction
0.5
0.4
0.3
0.2
0.1
CNS-PiloAstro
Bone-Cart
Myeloid-MDS
Myeloid-MPN
Thy-AdenoCA
CNS-Medullo
Myeloid-AML
Kidney-ChRCC
Bone-Epith
Panc-Endocrine
Breast-DCIS
Lymph-CLL
CNS-Oligo
Breast-LobularCa
Bone-Osteosarc
Prost-AdenoCA
Cervix-AdenoCA
Bone-Leiomyo
Panc-AdenoCA
Breast-AdenoCa
Kidney-RCC
Cervix-SCC
Biliary-AdenoCA
CNS-GBM
Uterus-AdenoCA
Ovary-AdenoCA
Lymph-BNHL
Head-SCC
Liver-HCC
Stomach-AdenoCA
Lymph-NOS
Bladder-TCC
ColoRect-AdenoCA
Lung-AdenoCA
Eso-AdenoCa
Lung-SCC
Skin-Melanoma


Fraction of TFM mutations
Fraction
0.25
0.20
0.15
0.10
0.05
0.00
CNS-PiloAstro
Bone-Cart
Myeloid-MDS
Myeloid-MPN
Thy-AdenoCA
CNS-Medullo
Myeloid-AML
Kidney-ChRCC
Bone-Epith
Panc-Endocrine
Breast-DCIS
Lymph-CLL
CNS-Oligo
Breast-LobularCa
Bone-Osteosarc
Prost-AdenoCA
Cervix-AdenoCA
Bone-Leiomyo
Panc-AdenoCA
Breast-AdenoCa
Kidney-RCC
Cervix-SCC
Biliary-AdenoCA
CNS-GBM
Uterus-AdenoCA
Ovary-AdenoCA
Lymph-BNHL
Head-SCC
Liver-HCC
Stomach-AdenoCA
Lymph-NOS
Bladder-TCC
ColoRect-AdenoCA
Lung-AdenoCA
Eso-AdenoCa
Lung-SCC
Skin-Melanoma

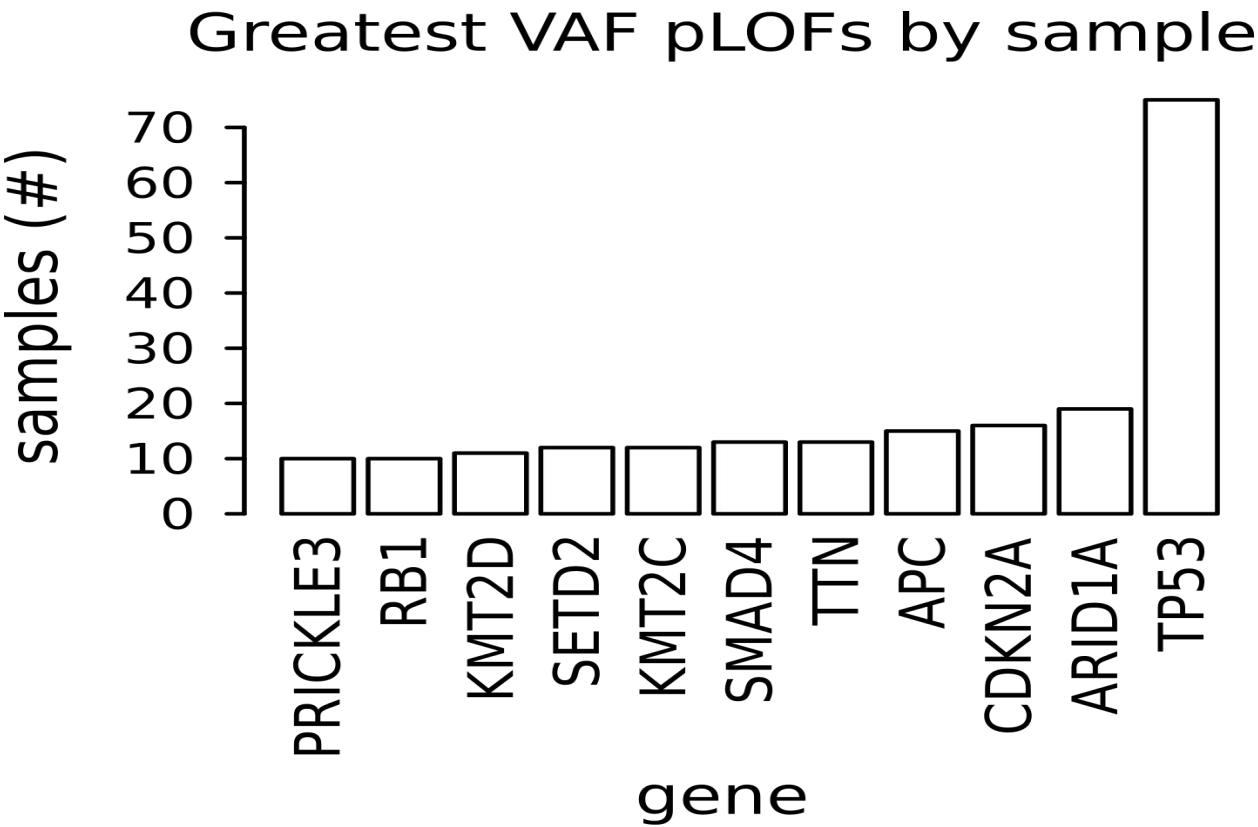

Fraction of Enhancer mutations
Fraction
0.3
0.2
0.1
0.0
CNS-PilloAstro
Bone-Cart
Myeloid-MDS
Myeloid-MPN
Thy-AdenoCA
CNS-Medullo
Myeloid-AML
Kidney-ChRCC
Bone-Epith
Panc-Endocrine
Breast-DCIS
Lymph-CLL
CNS-Oligo
Breast-LobularCa
Bone-Osteosarc
Prost-AdenoCA
Cervix-AdenoCA
Bone-Leiomyo
Panc-AdenoCA
Breast-AdenoCa
Kidney-RCC
Cervix-SCC
Biliary-AdenoCA
CNS-GBM
Uterus-AdenoCA
Ovary-AdenoCA
Lymph-BNHL
Head-SCC
Liver-HCC
Stomach-AdenoCA
Lymph-NOS
Bladder-TCC
ColoRect-AdenoCA
Lung-AdenoCA
Eso-AdenoCa
Lung-SCC
Skin-Melanoma


Fraction of ncRNA mutations
Fraction
0.020
0.015
0.010
0.005
0.000
CNS-PiloAstro
Bone-Cart
Myeloid-MDS
Myeloid-MPN
Thy-AdenoCA
CNS-Medullo
Myeloid-AML
Kidney-ChRCC
Bone-Epith
Panc-Endocrine
Breast-DCIS
Lymph-CLL
CNS-Oligo
Breast-LobularCa
Bone-Osteosarc
Prost-AdenoCA
Cervix-AdenoCA
Bone-Leiomyo
Panc-AdenoCA
Breast-AdenoCa
Kidney-RCC
Cervix-SCC
Biliary-AdenoCA
CNS-GBM
Uterus-AdenoCA
Ovary-AdenoCA
Lymph-BNHL
Head-SCC
Liver-HCC
Stomach-AdenoCA
Lymph-NOS
Bladder-TCC
ColoRect-AdenoCA
Lung-AdenoCA
Eso-AdenoCa
Lung-SCC
Skin-Melanoma


Fraction of psuedogene mutations
Fraction
0.03
0.02
0.01
0.00
CNS-PiloAstro
Bone-Cart
Myeloid-MDS
Myeloid-MPN
Thy-AdenoCA
CNS-Medullo
Myeloid-AML
Kidney-ChRCC
Bone-Epith
Panc-Endocrine
Breast-DCIS
Lymph-CLL
CNS-Oligo
Breast-LobularCa
Bone-Osteosarc
Prost-AdenoCA
Cervix-AdenoCA
Bone-Leiomyo
Panc-AdenoCA
Breast-AdenoCa
Kidney-RCC
Cervix-SCC
Biliary-AdenoCA
CNS-GBM
Uterus-AdenoCA
Ovary-AdenoCA
Lymph-BNHL
Head-SCC
Liver-HCC
Stomach-AdenoCA
Lymph-NOS
Bladder-TCC
ColoRect-AdenoCA
Lung-AdenoCA
Eso-AdenoCa
Lung-SCC
Skin-Melanoma


Gain of motif impact score
Biliary−AdenoCA
Bladder−TCC
Bone−Cart
Bone−Epith
Bone−Leiomyo
Bone−Osteosarc
Breast−AdenoCa
Breast−DCIS
Breast−LobularCa
Cervix−AdenoCA
Cervix−SCC
CNS−GBM
CNS−Medullo
CNS−Oligo
CNS−PiloAstro
ColoRect−AdenoCA
Eso−AdenoCa
Head−SCC
Kidney−ChRCC
Kidney−RCC
Liver−HCC
Lung−AdenoCA
Lung−SCC
Lymph−BNHL
Lymph−CLL
Lymph−NOS
Myeloid−AML
Myeloid−MDS
Myeloid−MPN
Ovary−AdenoCA
Panc−AdenoCA
Panc−Endocrine
Prost−AdenoCA
Skin−Melanoma
Stomach−AdenoCA
Thy−AdenoCA
Uterus−AdenoCA
FI score
0
1
2
3
4
True
Random
Set
True
Random

Top 10 anticorrelated pLOFs
correlation coefficient
0.000
-0.005
-0.010
-0.015
-0.020
-0.025
SETD2--TP53
CDKN2A--APC
ARID2--KMT2D
SMAD4--KMT2D
KMT2C--CDKN2A
ARID2--CDKN2A
SETD2--KMT2C
CREBBP--KMT2D
SETD2--PTEN
SYNE1--CDKN2A
gene pair


Greatest VAF pLOFs by sample
samples (#)
0
10
20
30
40
50
60
70
PRICKLE3
RB1
KMT2D
SETD2
KMT2C
SMAD4
TTN
APC
CDKN2A
ARID1A
TP53
gene