Non-Coding analysis of TCGA KIRP WGS data

Jason Bedford, Ekta Khurana, Yao Fu,
Brian Shuch, and Mark Gerstein.

Yale University

10-06-2014

Can we identify which non-coding elements are under very strong, “coding-like”, selection?

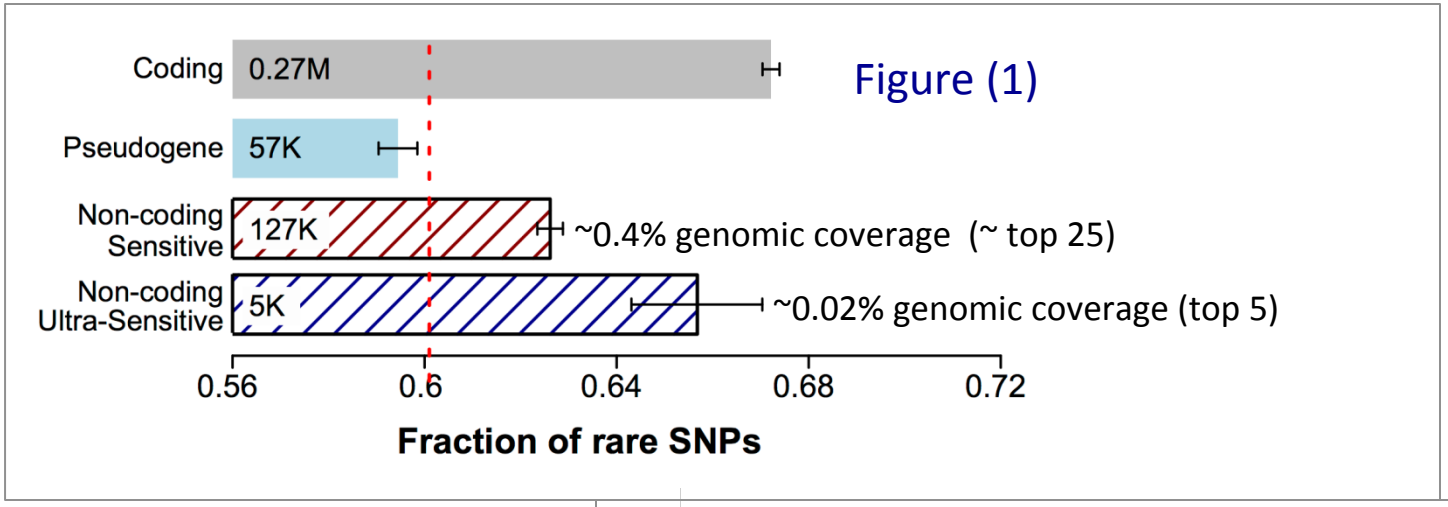


Figure (1)

- Start 677 high-resolution non-coding categories; Rank & find those under strongest selection
- Binding peaks of some general TFs (eg *FAM48A*)
- Core motifs of some TF families (eg *JUN*, *GATA*)
- DHS sites in spinal cord and connective tissue

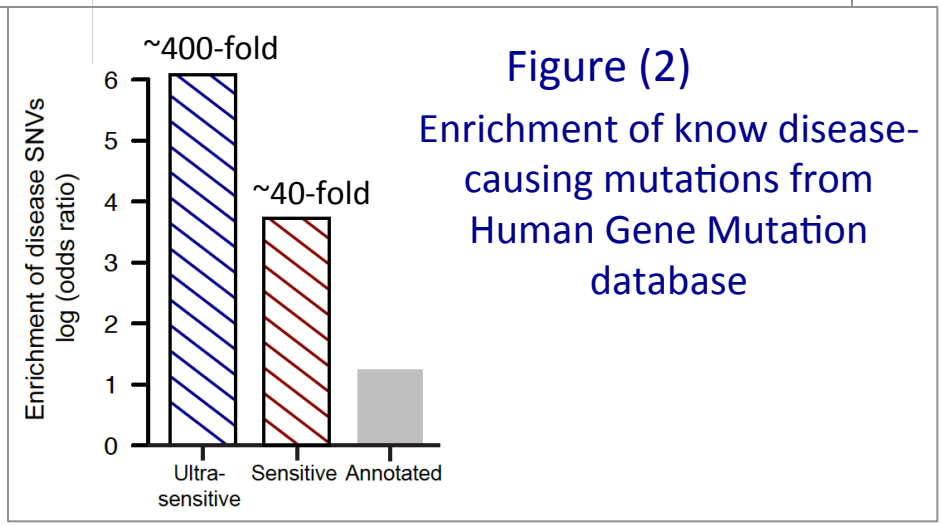
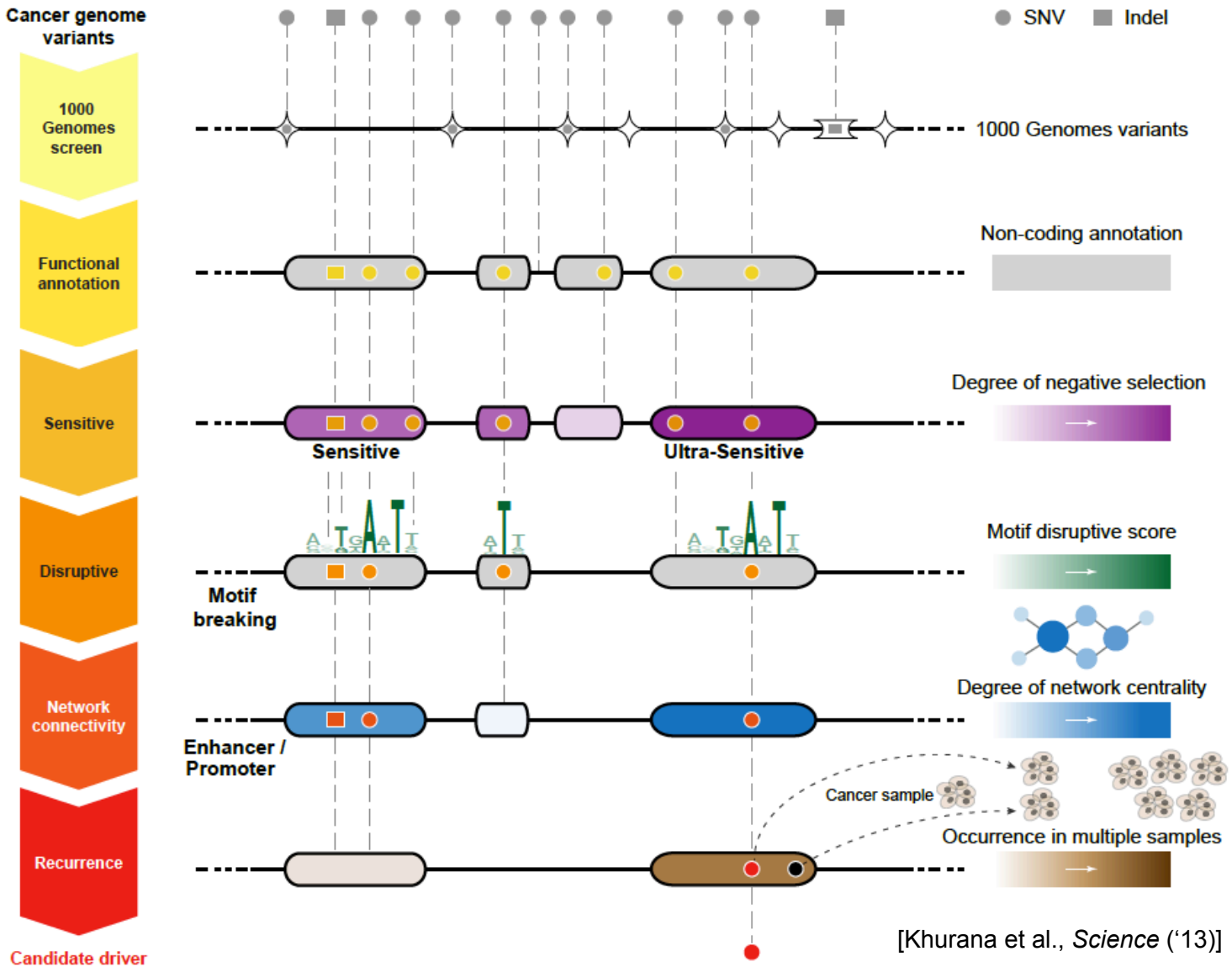


Figure (2)

Enrichment of know disease-causing mutations from Human Gene Mutation database

[Khurana et al., *Science* (‘13)]

Identification of non-coding candidate drivers amongst somatic variants: scheme



[Khurana et al., *Science* ('13)]

Overview of Non-Coding Variants in 32 TCGA KIRP WGS datasets

1000 Genomes screen

153,258 left
Avg: 4789
SD: 1804

Sensitive

Labeled Ulta-Sensitive : 43

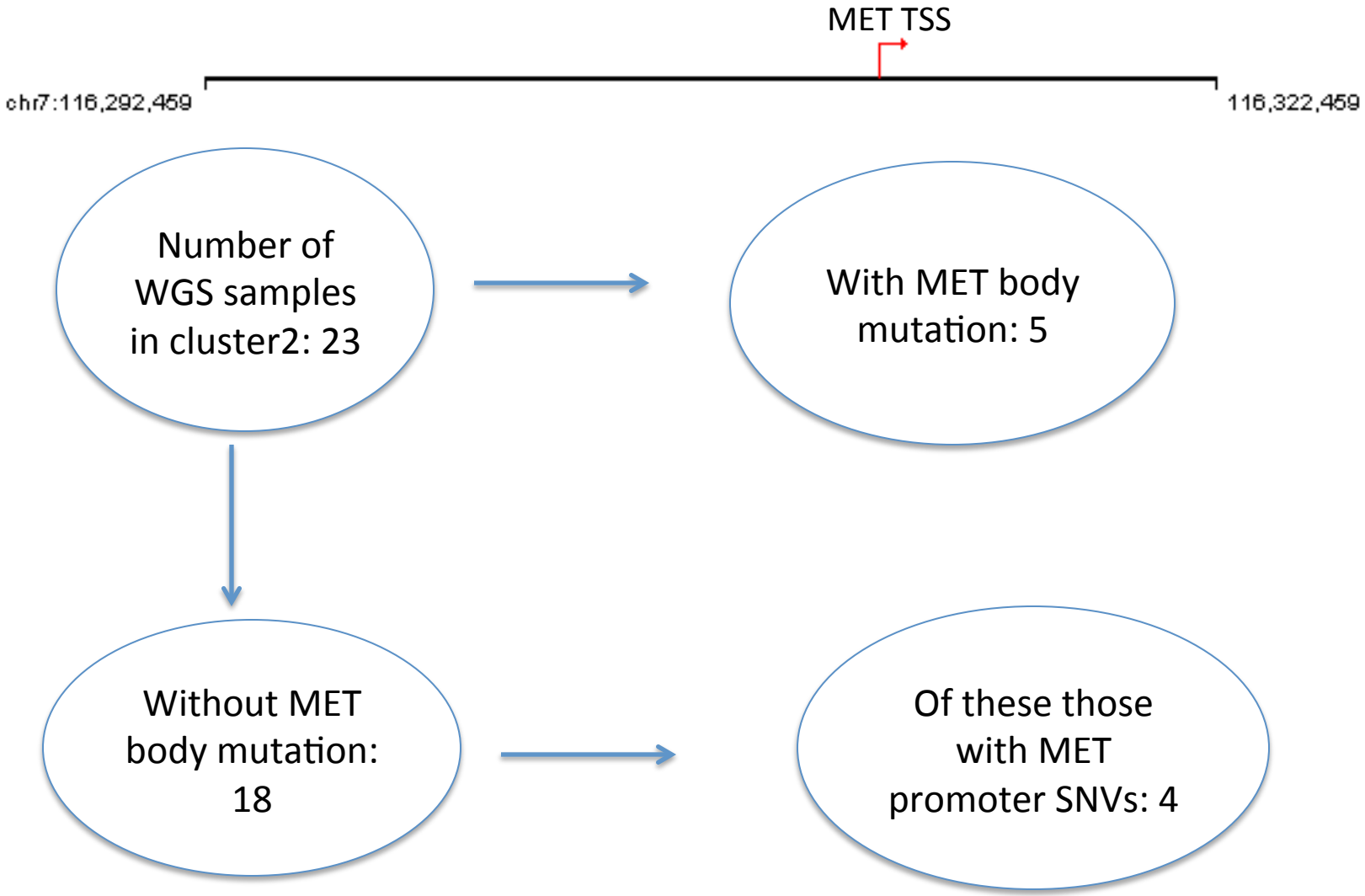
Functional annotation

90,145
Avg: 2817
SD: 1070

- Promoters: 4,629
  Avg: 144, SD: 64
- Enhancers: 4,237
  Avg: 133, SD: 56
- TF-Peaks: 20,654
  Avg: 645, SD: 261
- DHS: 23,354
  Avg: 730, SD: 297
- ncRNAs: 237
  Avg: 7, SD: 4
- Pseudogenes: 362
  Avg: 11, SD: 5
- Introns: 64,142
  Avg: 2006, SD: 824
- UTRs: 2030
  Avg: 63, SD: 30

[SNVs called by Suzi Fei]

Mutations in MET promoter for samples in cluster 2



FOXA2 and FoxA3 TF networks

- 12/32 samples are mutated
- FOXA1 acts as a pioneer factor for ERa

Recurrent SNVs in chr1:8064127-8071691

- TF-Peak of FOSL2 chr1:8064127-8071691 mutated in 8/32
- This region is also in H3K27ac peak

Recurrent SNVs in NEAT1

- ncRNA NEAT1 is mutated in 6/32
- Patents with mutated NEAT1 had worse outcome.
- NEAT1 is a downstream mediator of ERα signaling.

Recurrent SNVs in ETS domain

TCGA-B1-A47N.txt:chr5 36069063 36069064 A->C

TCGA-MH-A560.txt:chr5 36069063 36069064 A->C

GGA -> GGC change

All ETS-domain proteins bind to sequences that contain a central GGA motif.

Acknowledgments

- This work would not be possible with out the SNVs called by Suzi Fei from Paul Spellman's group at OHSU.
- Thank you for your time, I look forward to your feedback

- Backup slides and other stuff after this.

Ultra-Sensitive/Conserved ncSNVs

Ultra sensitive (depletion of common variants/an enrichment of rare variants)

CDC27(Intron), NRN1(Enhancer), CTDSP1(Promoter), SLC11A1(UTR), FANK1(Intron), VAC14(Intron), PARP16(Intron), PDE4DIP(Intron), HIST1H4H(Promoter), SMARCC2(Promoter), FANK1(Intron), CASC5(Promoter), ANXA8L1(Intron), MLL(Promoter), RP11-770J1.4(Promoter), CCDC107(Promoter), CUL2(Intron), FANK1(Intron), NRN1(Enhancer), PIWIL1(Intron&Promoter), ZNF544(UTR), ZBED5(Intron), NXA8L1(Intron), C6orf100(Promoter), GDPD5(Enhancer), RPS3(Enhancer), TIA1(Promoter), Pseudogene(ENST00000511272.1), Pseudogene(ENST00000511272.1), ncRNA(RNU5A) ncRNA(NEAT1), ncRNA(NEAT1).

Ultra conserved (defined by comparison across species)

POLA1(Intron), ZNF407(Intron), BTRC(Intron), SRSF6(Intron)

Disruptive SNVs

- Binding Motif Breaking: 532, Avg: 14, SD:6
- Binding Motif Making: 580, Avg: 17, SD: 8
- SNVs in HOT regions : 6,004, Avg: 172, SD: 67

Network connected SNVs

- Protein-protein interaction: 55,246, Avg: 1481, SD: 7
- Phosphorylation Network: 9,196, Avg: 263, SD: 98
- Regulatory Network: 32,209, Avg: 920, SD: 383

Intronic SNVs

- Total SNVs in Introns: 532, Avg: 14, SD:6

- SNVs in Introns and in H3K27ac peaks:

(ENCODE data)

FunSeq2 Annotation

- DNA Repair: 2303

Intron: 1709 Promoter: 68

Enhancer: 315 UTR: 63

- Altered in cancer: 2001

Intron: 1591 Promoter: 37

Enhancer: 247 UTR: 57