

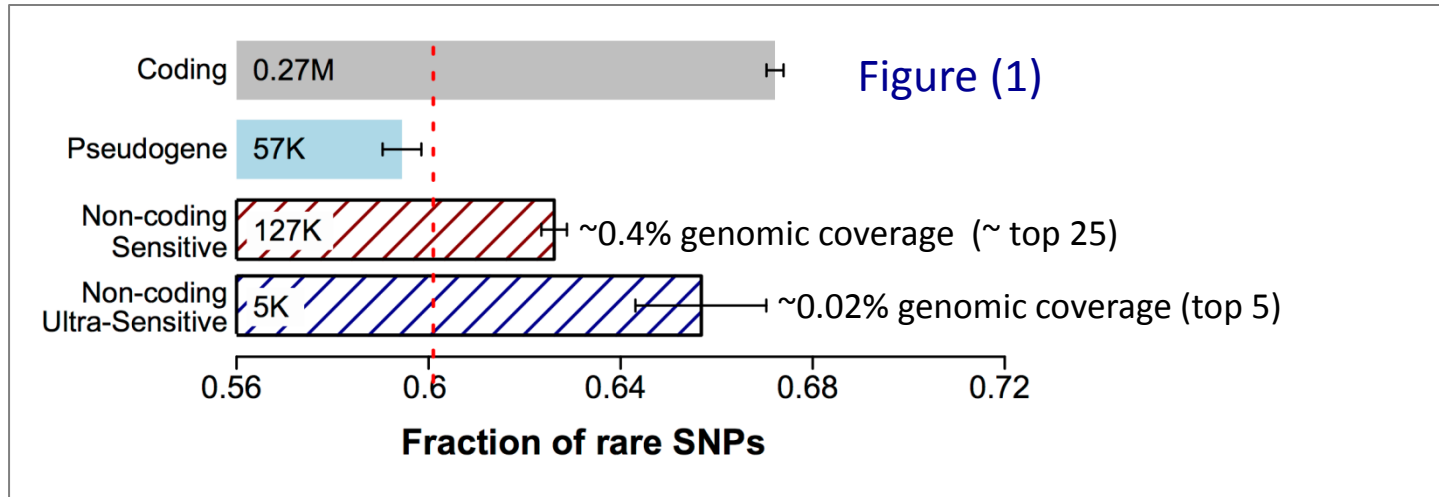
# Non-Coding analysis of TCGA KIRP WGS data

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

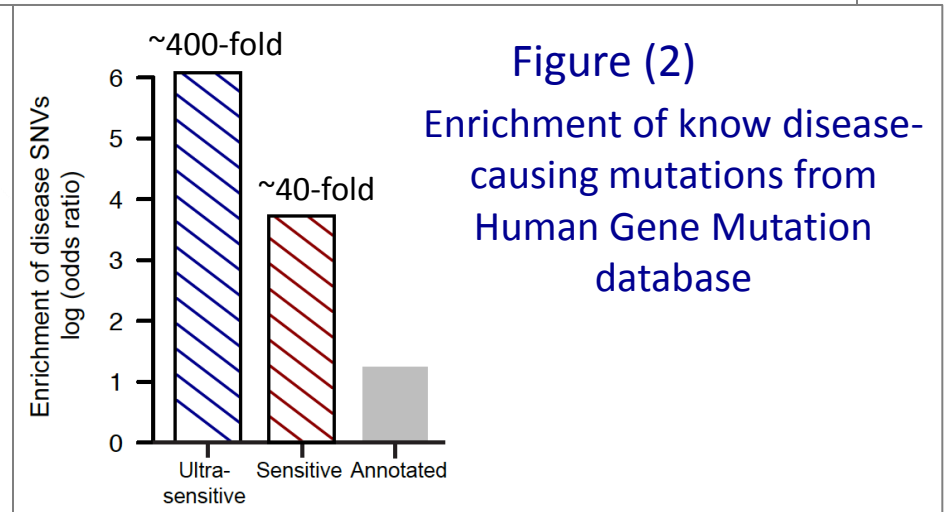
Yale University

06-02-2014

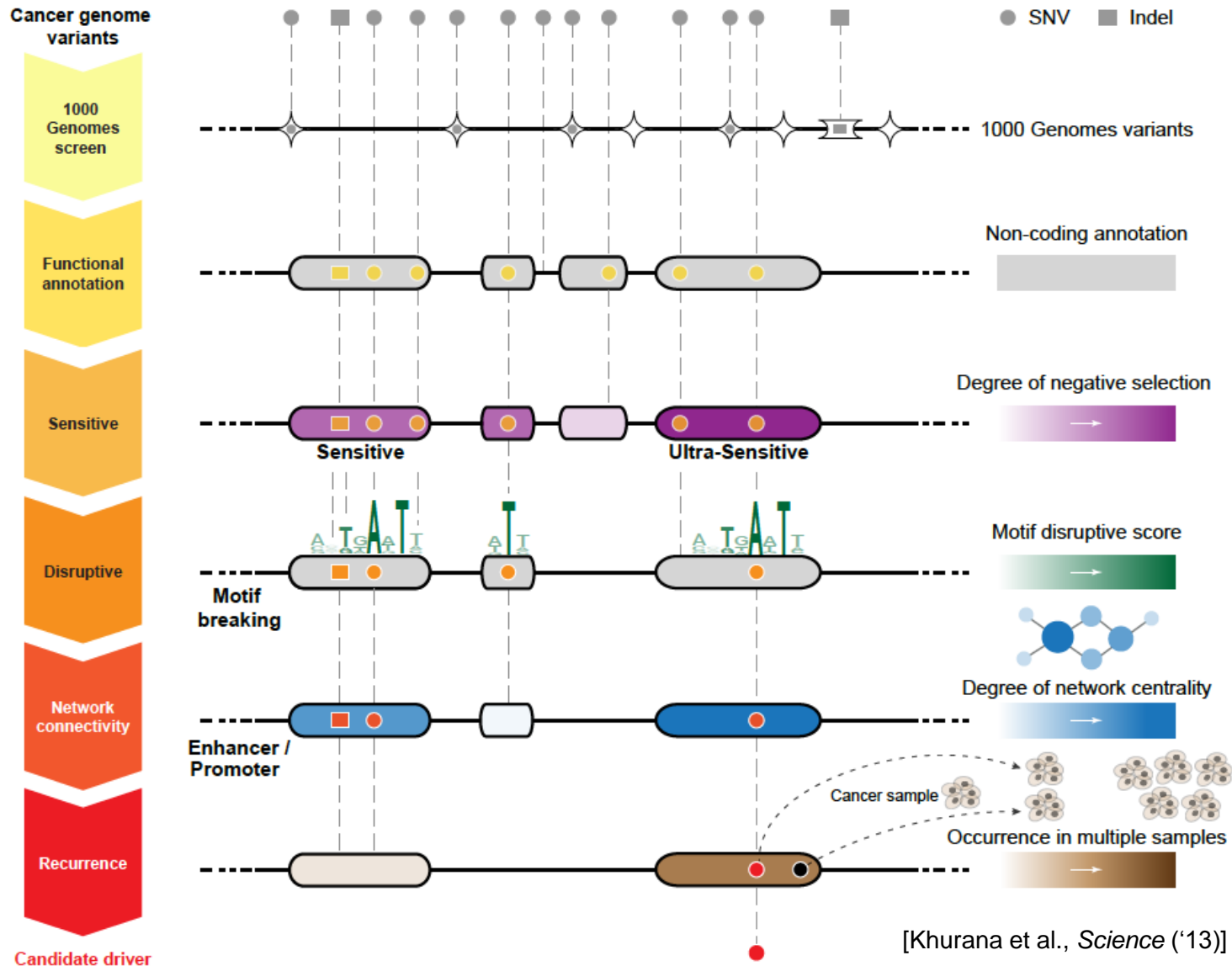
# Can we identify which non-coding elements are under very strong, “coding-like”, selection?



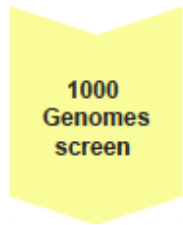
- ❑ Start 677 high-resolution non-coding categories; Rank & find those under strongest selection
- ❑ Binding peaks of some general TFs (eg *FAM48A*)
- ❑ Core motifs of some TF families (eg *JUN*, *GATA*)
- ❑ DHS sites in spinal cord and connective tissue



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



# Overview of Non-Coding Variants in 35 TCGA KIRP WGS datasets, part 1



162,840 left  
Avg: 4652  
SD: 1805



93,185  
Avg: 2662  
SD: 1070

- Promoters: 5,009  
Avg: 143, SD: 64
- Enhancers: 4,667  
Avg: 133, SD: 56
- TF-Peaks: 22,548  
Avg: 644, SD: 261
- DHS: 25,527  
Avg: 730, SD: 297
- ncRNAs: 249  
Avg: 7, SD: 4
- Pseudogenes: 391  
Avg: 11, SD: 5
- Introns: 70,433  
Avg: 2012, SD: 810
- UTRs: 2309  
Avg: 65, SD: 31

# Overview of Non-Coding Variants in 35 TCGA KIRP WGS datasets, part 2



- Binding Motif Breaking: 548 Avg: 16, SD: 6
- Binding Motif Making: 580 Avg: 17, SD: 8
- SNVs in HOT regions : 6,004 Avg: 172, SD: 67



- Protein-protein interaction: 55,350 Avg: 1581, SD: 641
- Phosphorylation Network: 9196 Avg: 263, SD: 110
- Regulatory Network: 32,209 Avg: 920, SD: 383

# Overview of Non-Coding Variants in 35 TCGA KIRP WGS datasets, part 3



- Labeled Sensitive: 597 Avg: 16, SD: 7
- Labeled Ultra-Sensitive : 46 Avg: 1, SD: 1
- Labeled Ultra-Conserved : 5,009 GERP Avg: 5.53, SD:0.2



- TF-Peak of FOSL2 chr1:8064127-8071691 mutated in 8/25
- ncRNA NEAT1 mutation in 6/35
- 10 different NC SNVs reoccurred in the exact same spot in 2 different samples
- 28 ultra-sensitive SNVs from 22 unique people are annotated (see last slide)

# Acknowledgments

---

- This work would not be possible without the SNVs called by Suzi Fei from Paul Spellman's group at OHSU.
- Thank you for your time, I look forward to your feedback

# FunSeq ultra-sensitive ENCODE annotated

TCGA-AL-3466 chr17 45213576 CDC27(Intron) TCGA-HE-A5NI chr6 6812161 NRN1(Enhancer)  
TCGA-AL-3468 chr10 127600320 FANK1(Intron) TCGA-HE-A5NI chr12 130826418 PIWIL1(Intron&Promoter)  
TCGA-AL-3473 chr16 70813081 VAC14(Intron) TCGA-HE-A5NJ chr19 58774355 ZNF544(UTR)  
TCGA-AL-3473 chr15 65588493 PARP16(Intron) TCGA-IA-A40Y chr11 10878453 ZBED5(Intron)  
TCGA-B1-A47O chr1 144928702 PDE4DIP(Intron) TCGA-MH-A55W chr10 47143311 ANXA8L1(Intron)  
TCGA-B3-3925 chr6 26286613 HIST1H4H(Promoter) TCGA-MH-A55W chr6 28911517 C6orf100(Promoter)  
TCGA-B9-4114 chr12 56584137 SMARCC2(Promoter) TCGA-MH-A560 chr11 75946623 GDPD5(Enhancer),RPS3(Enhancer)  
TCGA-B9-4115 chr10 127614341 FANK1(Intron) TCGA-MH-A561 chr2 70476430 TIA1(Promoter)  
TCGA-B9-4116 chr15 40885993 CASC5(Promoter) TCGA-B9-4114 chr4 26115035 Pseudogene(ENST00000511272.1)  
TCGA-B9-4116 chr10 47143212 ANXA8L1(Intron) TCGA-B9-4114 chr4 26115036 Pseudogene(ENST00000511272.1)  
TCGA-GL-A59R chr9 35658036 CCDC107(Promoter) TCGA-AL-3473 chr15 65588493 ncRNA(RNU5A)  
TCGA-HE-A5NF chr10 35310335 CUL2(Intron) TCGA-MH-A55Z chr11 65192208 ncRNA(NEAT1)  
TCGA-HE-A5NI chr10 127604185 FANK1(Intron) TCGA-B3-4103 chr11 65195871 ncRNA(NEAT1)  
TCGA-AL-3466 chr2 219260896 CTDSP1(Promoter),SLC11A1(UTR)  
TCGA-B9-A44B chr11 118306803 MLL(Promoter),RP11-770J1.4(Promoter)