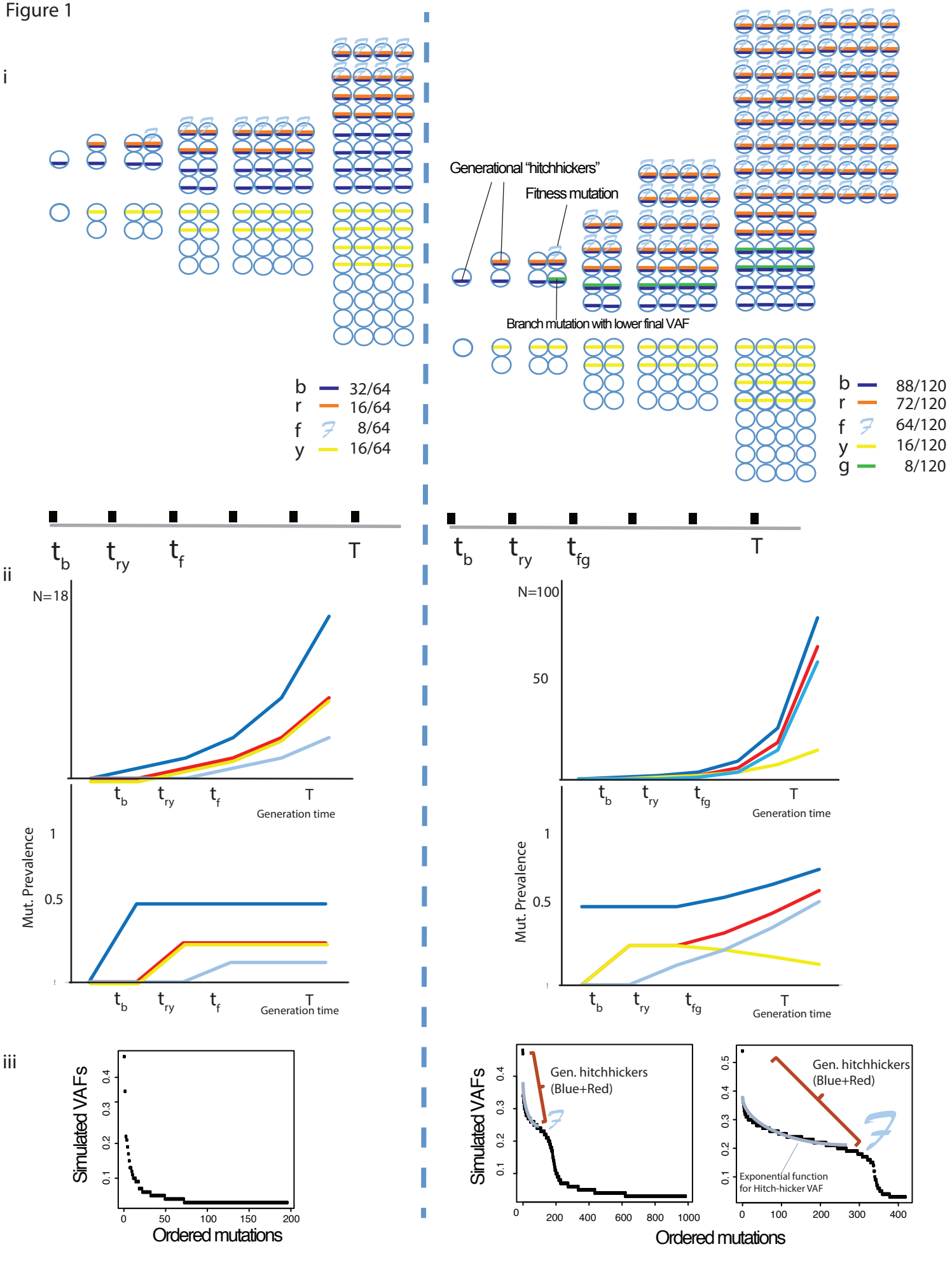
Figure 1


i
Generational "hitchhickers"
Fitness mutation
Branch mutation with lower final VAF
b 32/64
r 16/64
f 8/64
y 16/64
b 88/120
r 72/120
f 64/120
y 16/120
g 8/120
ii
N=18
N=100
50
Generation time
Mut. Prevalence
1
0.5
iii
Simulated VAFs
Ordered mutations
Gen. hitchhickers (Blue+Red)
Exponential function for Hitch-hicker VAF

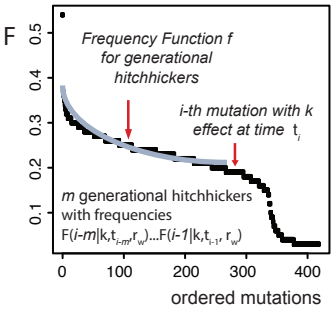
Using a single tumor

VCF file

Normalize VAF (based on ploidy, CNV and purity) to assign mutational frequency F

Use subclonal Information for each mutation to select non-overlapping subclones

Order mutations from earlier to last



For each mutation i > m using m generational hitchickers

A) estimate of growth 'r'

B) NLS optimization for effect 'k' as:

$$\mathbb{Q}(k, t_g \mid r, F, m) = \sum_m (F_m - (A * e^{-r(m+t_g)} + B))^2$$

C) Identify peaks 'K_p'

Output

$$\vec{r} = \{r_1, r_2, .. r_n\} \in \mathcal{R}^N_{+/-}$$
$$\vec{k} = \{k_1, k_2, .. k_n\} \in \mathcal{R}^N_{+}$$
$$\overrightarrow{Kp} = \{k_1, k_2, .. k_{p<n}\} \in \mathcal{R}^N_{+}, \quad where\ \overrightarrow{Kp} \subset \vec{k}$$

Glossary

Variant Allele Frequency (VAF): The fraction of sequencing reads overlapping a genomic coordinate that support the non-reference allele. This fraction can be further normalized based on the sample's ploidy and purity.

Variant Call Format (VCF) file: A text file format that includes sequencing information such as the position and frequency of every mutation in the sample.

Subclone: Cells that belong to a single lineage during population growth. Within the subclone, a higher mutational frequency is associated with an earlier time of occurence.

Linear subclones: A population growth where every subclone has at most one child subclone (e.g. subclone A -> Subclone B -> Subclone C).

Fitness mutation: A mutation that increases the growth of the population. Typically, a fitness mutation might lead to the formation of a subclone.

Generational hitchhiker (g-hitchiker): a hitchhike mutation that occurred before the fitness mutation. They have increased VAF (higher than their respective fitness mutation) and represent generational time as their respective branching mutations have typically low VAF (see figure 1).

Growth *r*: Before the **fitness mutation**, the population was growing with a rate *r*. In our model, we use the prevalence of **generational hitchhikers** to estimate **growth *r***.

Fitness effect k_i: After the **fitness mutation** *i* occurs, the population is growing with rate $k_i * r$.

Frequency F(i): The frequency for mutation *i* at time of sequencing.

Frequency Function $f(F, t_g, t_{i-m})$: The function that describes the frequency F(*i-m*) for m **g-hitchhikers** occurring before **fitness mutation** *i*.

Generational time t_g: A local re-optimized constant to nullify time for *m* respective **g-hitchhikers.**

Growth vector $\vec{r}$: For each mutation i>m in the tumor sample we estimate **growth** r_{i-1}.

Effect vector $\vec{k}$: For each mutation i>m in the tumor sample we estimate **fitness effect** k_i

Peak vector $\overrightarrow{Kp}$: Local peaks for **effect vector** $\vec{k}$ correspond to **fitness mutations** with **effect** k_i.

Optimizing function $\mathbb{Q}(k, t_g \mid r, F, m)$: Using the **Frequency F** of *m* **g-hitchhikers** occurring before mutation *i*, we use a nonlinear least square (NLS) fitting to calculate **effect** k_i and **generational time** t_g.

Positive Growth Enrichment(PGE): A type of mutation (eg. missense TP53) is assessed whether it occurs significantly more often than random during periods of positive **growth *r***.

Frequency function f() for m "hitchhickers" (with $F_m > F_i$)

$$F(T, t_{i-m}) = \frac{e^{-r_w * t_{i-m}} * [\, N_{tot} - F(t_i) * N_{tot} + \sqrt[k_i]{F(t_i) * N_{tot}}\,] + F(t_i) * N_{tot} - \sqrt[k_i]{F(t_i) * N_{tot}}}{N_{tot}}$$

r_w: growth r corresponding to window [*i-m*, *i*-1]

t_{i-m}: the time when the (*i-m*)-th mutation occured

k_i: the effect K of the i-th mutation

$F(t_i)$: the frequency of the hypothetical fitness mutation *i*

N_{tot}: the total number of mutations

Frequency function for m "hitchhickers" with local reoptimization

$$F(T, t_g, t_{i-m}) = \frac{e^{-r(t_g + t_{i-m})} * (N_{tot} - F(t_i) * N_{tot} + \sqrt[k_i]{F(t_i) * N_{tot}}) + F(t_i) * N_{tot} - \sqrt[k_i]{F(t_i) * N_{tot}}}{N_{tot}}$$

t_g: locally optimized generational time to adjust for local hitchhickers

Using M multiple tumors

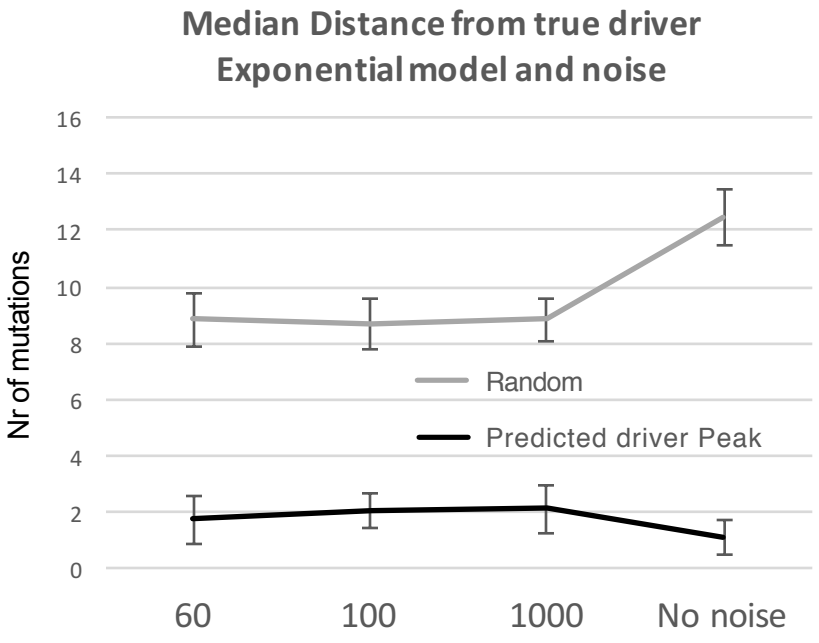
A) Estimate **Positive Growth Enrichment (PGE):**

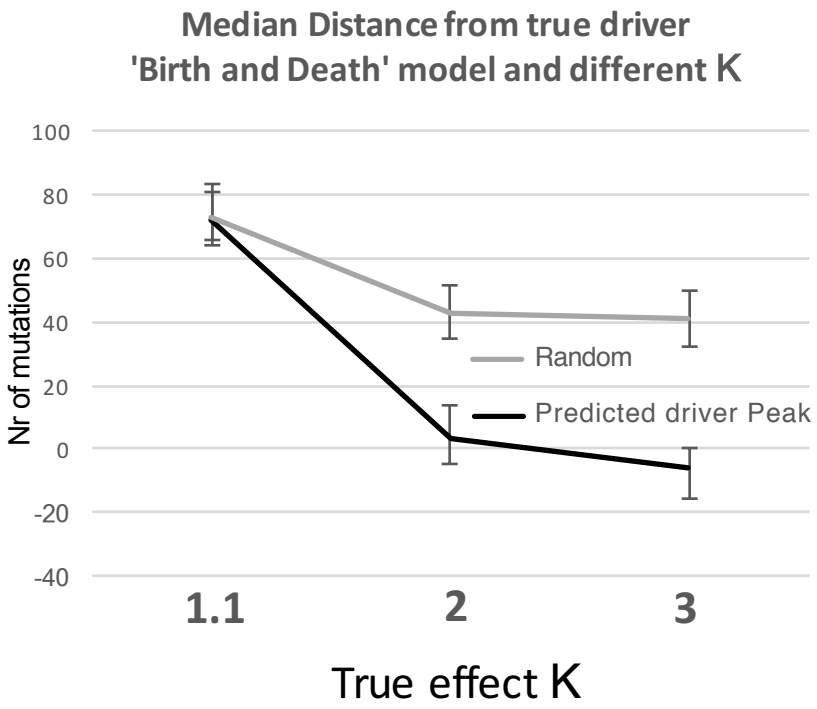
Across all M tumors,
A **type** of mutation (eg. all *missense TP53* mutations, or *all premature-stop* mutations *in Tumor Suppressor Genes*), found w times in M tumors,
is **enriched during positive growth** if:
mutational growth $\vec{r}_{mut} > 0$ more often than random

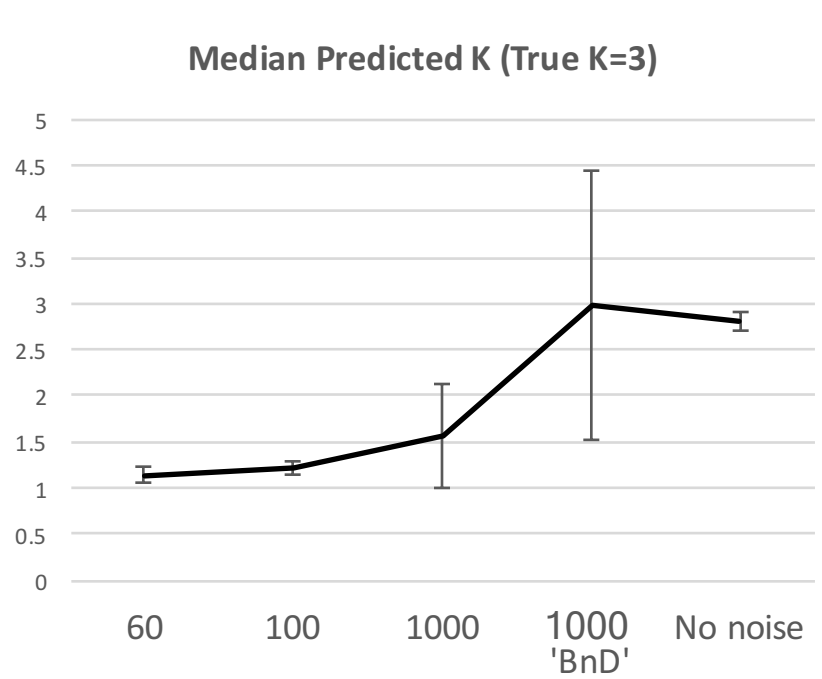
B) Estimate the range of effect k (eg. [1.2-1.4]) within a type of mutation based on enrichment

Figure 2

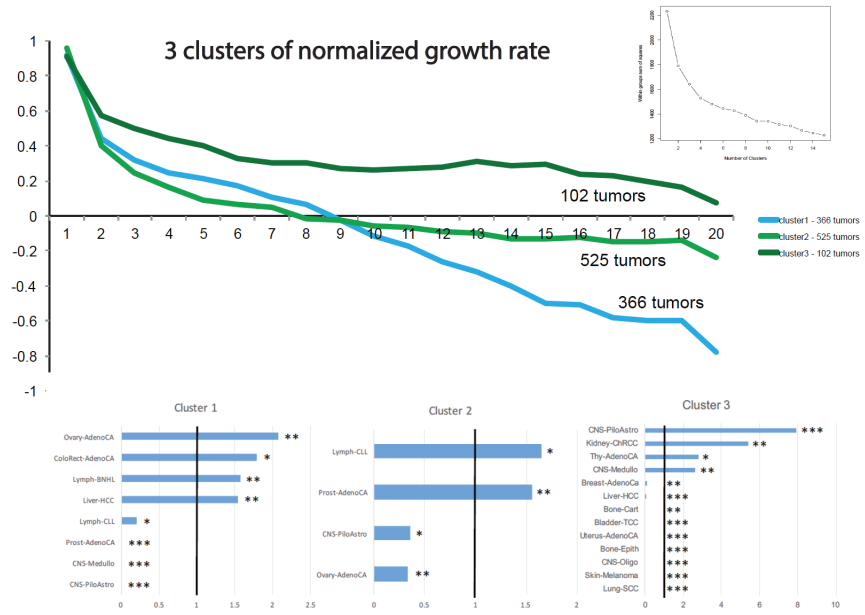
Simulations: Driver Prediction Under Exponential and 'BnD' Model With Noise (coverage)



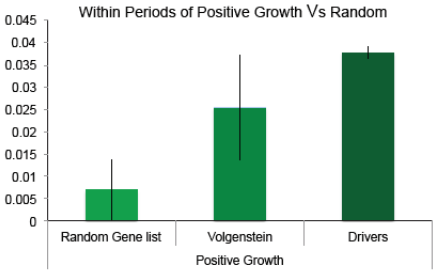


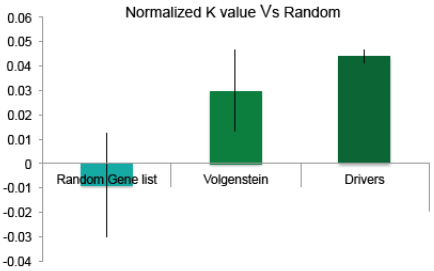


3i

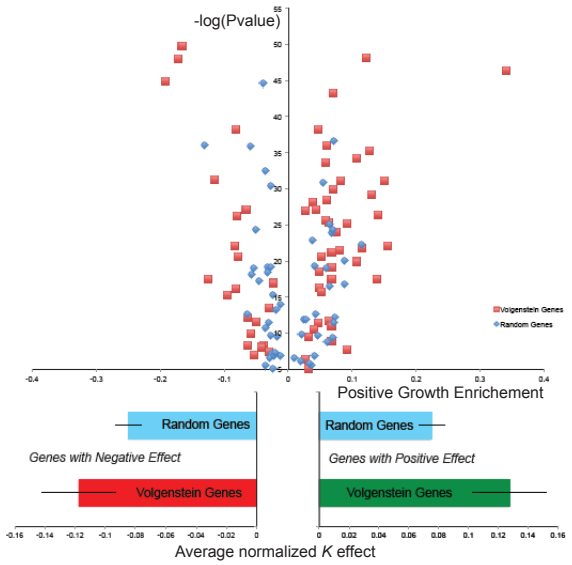


3ii



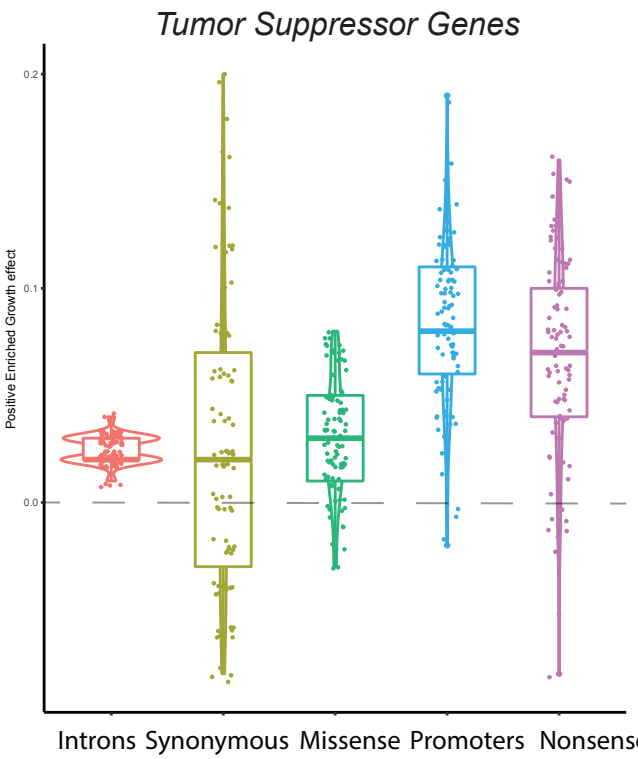


3iii

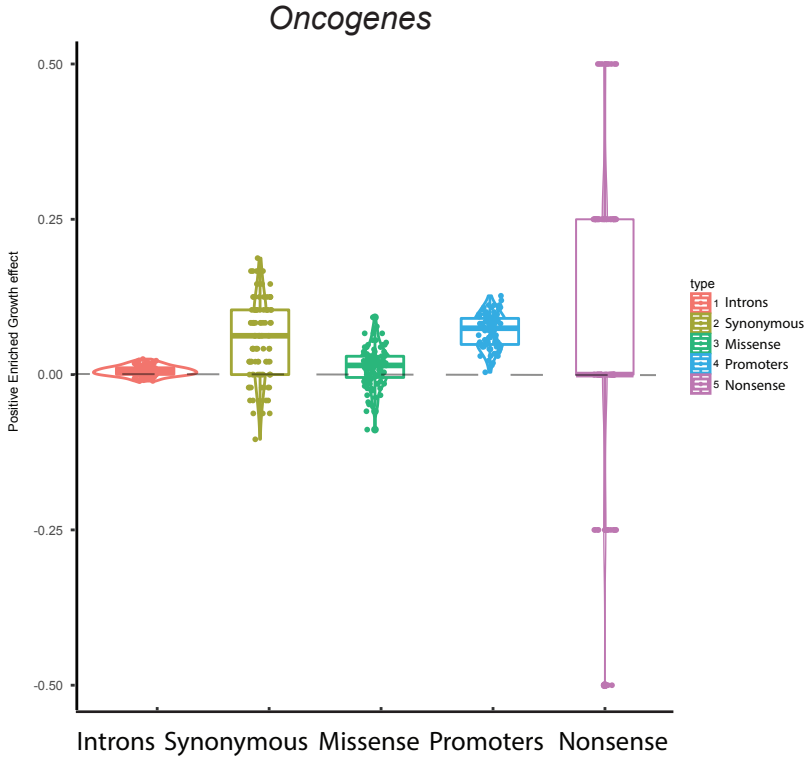


3iv

4i


Tumor Suppressor Genes
Positive Enriched Growth effect
0.2
0.1
0.0
Introns
Synonymous
Missense
Promoters
Nonsense

4ii


Oncogenes
Positive Enriched Growth effect
0.50
0.25
0.00
-0.25
-0.50
Introns
Synonymous
Missense
Promoters
Nonsense
type
1 Introns
2 Synonymous
3 Missense
4 Promoters
5 Nonsense

4iii

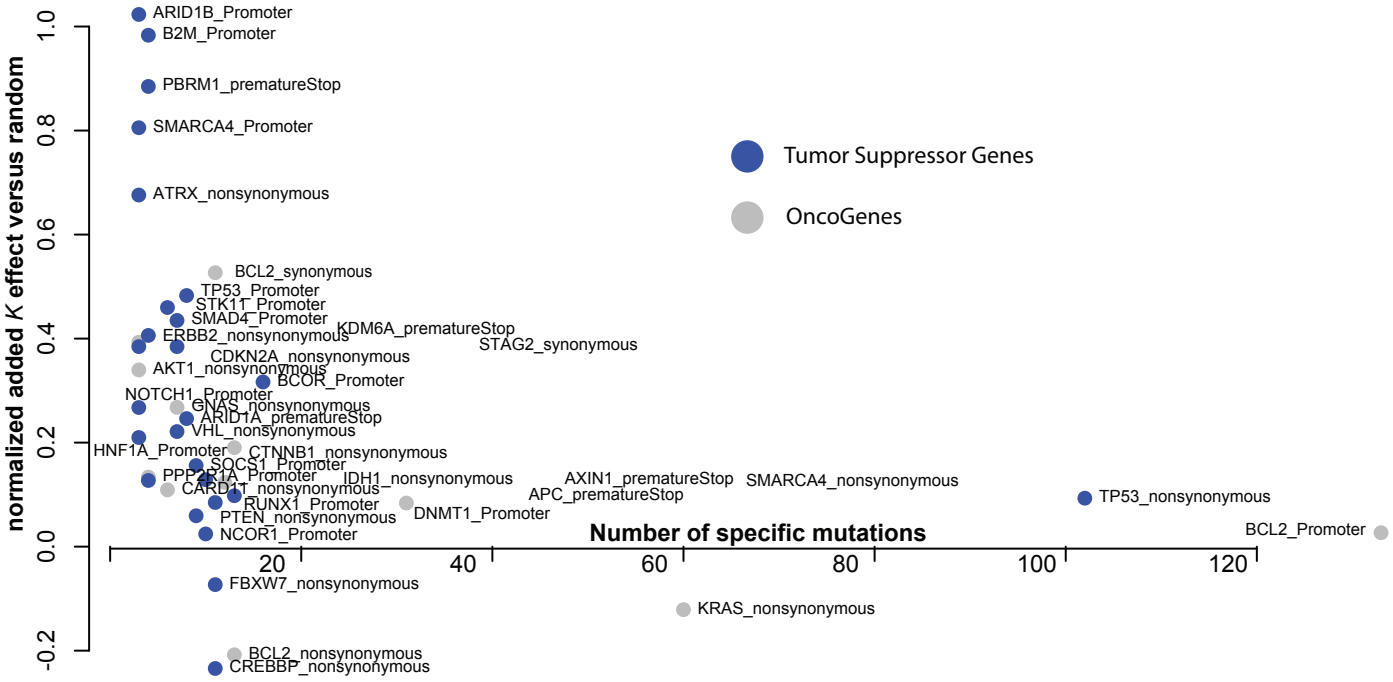
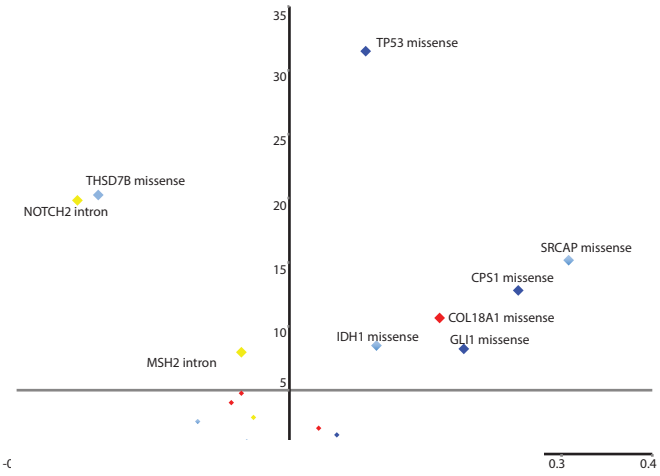
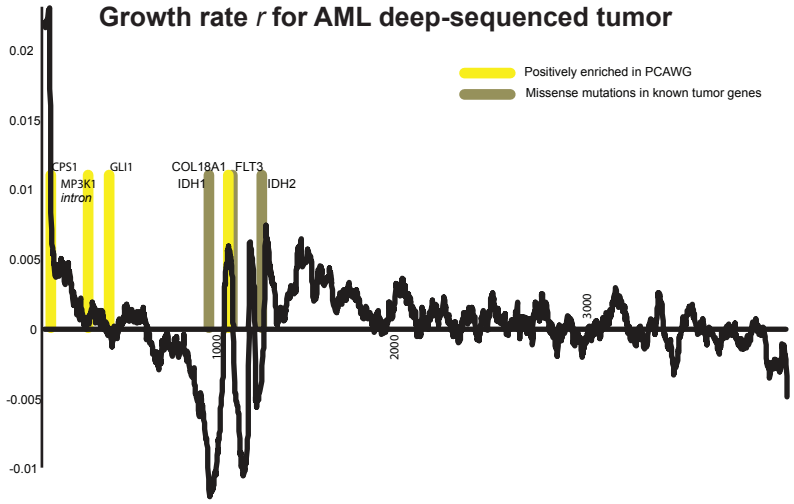

normalized added K effect versus random
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
Number of specific mutations
20
40
60
80
100
120
Tumor Suppressor Genes
OncoGenes
ARID1B_Promoter
B2M_Promoter
PBRM1_prematureStop
SMARCA4_Promoter
ATRX_nonsynonymous
BCL2_synonymous
TP53_Promoter
STK11_Promoter
SMAD4_Promoter
ERBB2_nonsynonymous
KDM6A_prematureStop
STAG2_synonymous
CDKN2A_nonsynonymous
AKT1_nonsynonymous
BCOR_Promoter
NOTCH1_Promoter
GNAS_nonsynonymous
ARID1A_prematureStop
VHL_nonsynonymous
HNF1A_Promoter
CTNNB1_nonsynonymous
SOCS1_Promoter
PPP2R1A_Promoter
IDH1_nonsynonymous
AXIN1_prematureStop
SMARCA4_nonsynonymous
CARD11_nonsynonymous
APC_prematureStop
TP53_nonsynonymous
RUNX1_Promoter
DNMT1_Promoter
PTEN_nonsynonymous
NCOR1_Promoter
BCL2_Promoter
FBXW7_nonsynonymous
KRAS_nonsynonymous
BCL2_nonsynonymous
CREBBP_nonsynonymous

Figure 5

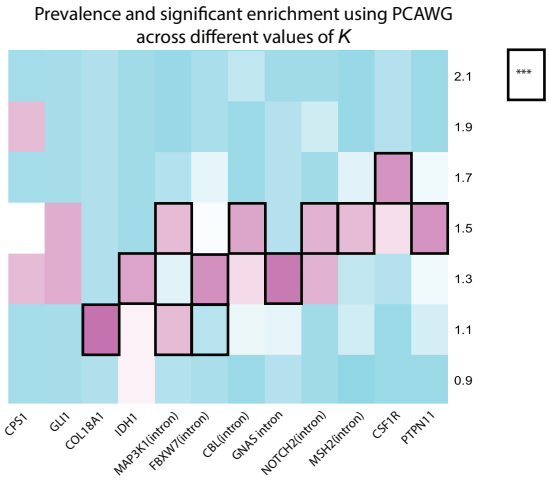
5i



5ii



5iii



5iv