Supplementary Materials

for

Estimating the order of mutations during tumorigenesis from tumor genome sequencing data

Table A: Estimates of $P_{k,i}$ for the simulated data when sample size N is 144 and 288

Gene	$P_{1,i}$			$P_{2,i}$			$P_{3,i}$			$P_{4,i}$			$P_{5,i}$		
	true	$N = 144$	$N = 288$	true	$N = 144$	$N = 288$	true	$N = 144$	$N = 288$	true	$N = 144$	$N = 288$	true	$N = 144$	$N = 288$
APC	0	0	0	0.04	0.02	0.05	0.06	0.06	0.05	0	0.03	0.01	0.04	0.01	0.04
ATM	0	0	0	0.11	0.11	0.11	0.04	0.03	0.04	0	0	0	0	0	0
CDKN2A	0.02	0.03	0.02	0	0	0	0.03	0.03	0.03	0	0.04	0	0.19	0.16	0.19
EGFR	0.2	0.21	0.21	0.05	0.03	0.04	0	0	0	0	0	0	0	0	0
EPHA3	0.02	0.02	0.02	0	0	0	0	0.01	0	0.15	0.16	0.15	0.07	0.06	0.04
EPHA5	0	0	0	0	0	0	0.07	0.07	0.07	0.09	0.03	0.07	0	0.02	0.01
EPHA7	0	0	0	0	0	0	0.04	0.05	0.04	0	0.01	0	0.07	0.08	0.06
ERBB4	0.03	0.02	0.03	0	0	0	0	0.01	0	0.08	0.04	0.08	0.04	0.05	0.03
FGFR4	0.01	0	0	0.03	0.02	0.03	0	0.01	0	0	0	0	0.08	0.05	0.09
INHBA	0	0	0	0	0	0	0.05	0.06	0.07	0	0	0	0	0	0
KDR	0.01	0.01	0.01	0	0	0	0	0	0	0.19	0.22	0.2	0	0	0.01
KRAS	0.41	0.4	0.4	0.01	0.01	0.01	0	0	0	0	0	0	0	0	0
LRP1B	0.01	0	0.01	0	0	0	0.06	0.05	0.07	0.28	0.27	0.28	0	0.07	0
LTK	0	0	0	0	0	0	0.03	0.02	0.04	0	0.02	0	0.14	0.08	0.14
MYO3B	0.02	0.01	0.01	0.01	0.01	0.01	0	0	0.01	0.08	0.04	0.07	0	0.02	0
NF1	0.05	0.05	0.05	0.01	0.01	0.01	0.04	0.05	0.03	0	0.01	0	0.12	0.09	0.11
NRAS	0.02	0.02	0.02	0	0	0	0	0	0	0	0.01	0	0	0	0
NTRK1	0	0.01	0.01	0.04	0.03	0.04	0.05	0.04	0.05	0	0.01	0	0	0.01	0
NTRK2	0	0	0	0	0	0	0.04	0.04	0.04	0	0.02	0	0.1	0.1	0.09
NTRK3	0	0	0	0.02	0	0.01	0.1	0.13	0.1	0	0.01	0	0	0	0
PAK3	0	0	0	0	0	0	0.05	0.04	0.06	0	0	0	0	0	0.01
PTEN	0.01	0.02	0.01	0	0	0	0.04	0.02	0.03	0	0	0	0	0.01	0
PTPRD	0.01	0.01	0.01	0.02	0.01	0.02	0.05	0.09	0.04	0.1	0.04	0.09	0	0.01	0.01
RB1	0.01	0	0.01	0	0	0	0.04	0.03	0.05	0	0.01	0	0.15	0.15	0.15
STK11	0.12	0.12	0.13	0.07	0.09	0.07	0.17	0.14	0.16	0	0	0	0	0	0
TFDP1	0.01	0.01	0	0	0	0	0.04	0.02	0.03	0	0.01	0	0	0.01	0
TP53	0.04	0.04	0.04	0.56	0.61	0.56	0	0	0	0.04	0.02	0.01	0	0	0
ZMYND10	0	0	0	0.03	0.03	0.03	0	0	0	0	0	0	0	0	0

Table B: Estimates of $P_{k,i}$ for non-small-cell lung tumors[1]

Gene	$P_{1,i}$		$P_{2,i}$		$P_{3,i}$		$P_{4,i}$		$P_{5,i}$	
	MLE	90% CI	MLE	90% CI	MLE	90% CI	MLE	90% CI	MLE	90% CI
APC	0	(0 , 0.009)	0.044	(0 , 0.102)	0.065	(0 , 0.191)	0	(0 , 0.108)	0.042	(0 , 0.279)
ATM	0	(0 , 0.067)	0.113	(0.04 , 0.195)	0.035	(0 , 0.186)	0	(0 , 0.136)	0	(0 , 0.152)
CDKN2A	0.017	(0 , 0.057)	0	(0 , 0.017)	0.028	(0 , 0.146)	0	(0 , 0.328)	0.192	(0.058 , 0.388)
EGFR	0.204	(0.157 , 0.269)	0.046	(0.012 , 0.091)	0	(0 , 0.041)	0	(0 , 0.036)	0	(0 , 0)
EPHA3	0.025	(0 , 0.068)	0	(0 , 0.083)	0	(0 , 0.083)	0.146	(0 , 0.354)	0.066	(0 , 0.331)
EPHA5	0	(0 , 0)	0	(0 , 0.026)	0.069	(0 , 0.201)	0.085	(0 , 0.381)	0	(0 , 0.177)
EPHA7	0	(0 , 0.019)	0	(0 , 0.03)	0.04	(0 , 0.141)	0	(0 , 0.105)	0.07	(0 , 0.229)
ERBB4	0.029	(0.007 , 0.06)	0	(0 , 0.037)	0	(0 , 0.091)	0.077	(0 , 0.27)	0.038	(0 , 0.178)
FGFR4	0.005	(0 , 0.021)	0.031	(0.011 , 0.085)	0	(0 , 0.015)	0	(0 , 0.185)	0.083	(0 , 0.237)
INHBA	0	(0 , 0.034)	0	(0 , 0.04)	0.055	(0.019 , 0.153)	0	(0 , 0.114)	0	(0 , 0.15)
KDR	0.007	(0 , 0.025)	0	(0 , 0)	0	(0 , 0.113)	0.192	(0 , 0.399)	0	(0 , 0.203)
KRAS	0.411	(0.351 , 0.476)	0.015	(0 , 0.048)	0	(0 , 0)	0	(0 , 0.077)	0	(0 , 0)
LRP1B	0.008	(0 , 0.032)	0	(0 , 0)	0.065	(0 , 0.254)	0.281	(0 , 0.582)	0	(0 , 0.32)
LTK	0	(0 , 0)	0	(0 , 0.04)	0.027	(0 , 0.158)	0	(0 , 0.225)	0.135	(0 , 0.413)
MYO3B	0.015	(0 , 0.043)	0.007	(0 , 0.035)	0	(0 , 0.058)	0.082	(0 , 0.304)	0	(0 , 0.199)
NF1	0.052	(0.023 , 0.091)	0.013	(0 , 0.054)	0.044	(0 , 0.206)	0	(0 , 0.188)	0.115	(0 , 0.302)
NRAS	0.021	(0.007 , 0.057)	0	(0 , 0.021)	0	(0 , 0.088)	0	(0 , 0.143)	0	(0 , 0.169)
NTRK1	0.005	(0 , 0.02)	0.036	(0.011 , 0.088)	0.051	(0.011 , 0.143)	0	(0 , 0.154)	0	(0 , 0.11)
NTRK2	0	(0 , 0.003)	0	(0 , 0.041)	0.042	(0 , 0.136)	0	(0 , 0.152)	0.103	(0 , 0.295)
NTRK3	0	(0 , 0.053)	0.016	(0 , 0.076)	0.099	(0 , 0.196)	0	(0 , 0.172)	0	(0 , 0.124)
PAK3	0	(0 , 0.023)	0	(0 , 0.037)	0.055	(0 , 0.148)	0	(0 , 0.162)	0	(0 , 0.131)
PTEN	0.014	(0 , 0.036)	0	(0 , 0)	0.036	(0 , 0.1)	0	(0 , 0.091)	0	(0 , 0.173)
PTPRD	0.009	(0 , 0.04)	0.017	(0 , 0.059)	0.048	(0 , 0.207)	0.097	(0 , 0.304)	0	(0 , 0.199)
RB1	0.007	(0 , 0.024)	0	(0 , 0.03)	0.036	(0 , 0.127)	0	(0 , 0.217)	0.154	(0 , 0.308)
STK11	0.124	(0.05 , 0.174)	0.074	(0.02 , 0.217)	0.17	(0.057 , 0.364)	0	(0 , 0.226)	0	(0 , 0.304)
TFDP1	0.007	(0 , 0.028)	0	(0 , 0.051)	0.036	(0.016 , 0.128)	0	(0 , 0.094)	0	(0 , 0.125)
TP53	0.038	(0.018 , 0.198)	0.558	(0.46 , 0.64)	0	(0 , 0.021)	0.04	(0 , 0.171)	0	(0 , 0.068)
ZMYND10	0	(0 , 0.007)	0.031	(0 , 0.074)	0	(0 , 0.059)	0	(0 , 0.08)	0	(0 , 0.043)

Table C: Estimates of $P_{k,i}$ for colorectal tumors[2]

Gene	$P_{1,i}$		$P_{2,i}$		$P_{3,i}$		$P_{4,i}$		$P_{5,i}$	
	MLE	90% CI	MLE	90% CI	MLE	90% CI	MLE	90% CI	MLE	90% CI
ADAMTS18	0.038	(0.008 , 0.075)	0	(0 , 0.036)	0	(0 , 0.048)	0	(0 , 0.066)	0	(0 , 0)
ADAMTSL3	0	(0 , 0.036)	0.026	(0 , 0.086)	0.03	(0 , 0.087)	0	(0 , 0.066)	0.024	(0 , 0.068)
APC	0.511	(0.151 , 0.81)	0.131	(0 , 0.815)	0.824	(0.259 , 0.967)	0	(0 , 0.62)	0.1	(0 , 0.256)
C10orf137	0	(0 , 0.024)	0	(0 , 0.008)	0	(0 , 0.029)	0	(0 , 0.083)	0.077	(0 , 0.183)
EPHA3	0.032	(0.022 , 0.074)	0	(0 , 0.021)	0.011	(0 , 0.057)	0.03	(0 , 0.172)	0.071	(0 , 0.213)
EPHB6	0.016	(0 , 0.06)	0.016	(0 , 0.061)	0	(0 , 0.046)	0	(0 , 0.076)	0	(0 , 0.1)
FBXW7	0	(0 , 0.035)	0.029	(0.014 , 0.082)	0	(0 , 0.033)	0.053	(0 , 0.221)	0.183	(0.047 , 0.305)
GNAS	0	(0 , 0.054)	0.034	(0.008 , 0.095)	0	(0 , 0.046)	0	(0 , 0.09)	0	(0 , 0.093)
GUCY1A2	0	(0 , 0.045)	0	(0 , 0.083)	0.071	(0.007 , 0.135)	0	(0 , 0.116)	0	(0 , 0.139)
KRAS	0.225	(0.061 , 0.676)	0	(0 , 0.146)	0	(0 , 0.659)	0.736	(0 , 0.898)	0	(0 , 0)
MAP2K7	0.019	(0 , 0.059)	0.005	(0 , 0.037)	0.03	(0.011 , 0.065)	0	(0 , 0.072)	0	(0 , 0.068)
MMP2	0	(0 , 0.03)	0.008	(0 , 0.041)	0.011	(0 , 0.051)	0	(0 , 0.074)	0.075	(0.017 , 0.21)
NAV3	0.025	(0 , 0.066)	0	(0 , 0.032)	0	(0 , 0.029)	0.041	(0 , 0.21)	0	(0 , 0.114)
OR51E1	0	(0 , 0.023)	0	(0 , 0.034)	0	(0 , 0.052)	0.047	(0.018 , 0.11)	0	(0 , 0.108)
PIK3CA	0	(0 , 0.132)	0.145	(0.081 , 0.231)	0.024	(0 , 0.197)	0	(0 , 0.211)	0.215	(0.124 , 0.508)
PTEN	0.03	(0.015 , 0.059)	0	(0 , 0.037)	0	(0 , 0.036)	0	(0 , 0.051)	0	(0 , 0.067)
RET	0	(0 , 0.05)	0.011	(0 , 0.049)	0	(0 , 0.042)	0	(0 , 0.078)	0.094	(0 , 0.197)
SEC8L1	0	(0 , 0.033)	0	(0 , 0.027)	0	(0 , 0.03)	0.016	(0 , 0.105)	0.076	(0 , 0.206)
TCF7L2	0.031	(0.02 , 0.063)	0.001	(0 , 0.067)	0	(0 , 0.032)	0.023	(0 , 0.127)	0.085	(0 , 0.222)
TNN	0.008	(0 , 0.027)	0.017	(0.001 , 0.058)	0	(0 , 0.031)	0.031	(0.012 , 0.116)	0	(0 , 0.054)
TP53	0.067	(0 , 0.249)	0.577	(0.031 , 0.676)	0	(0 , 0.495)	0.024	(0 , 0.151)	0	(0 , 0)

The effect of driver gene selection

Preselecting driver genes is sensible since only driver genes mutate in certain orders. Passenger mutations mutate in a random order. Adding all the passenger mutations introduces noise in the estimates and make the estimation computationally difficult. For a sample having n mutations, to calculate the likelihood of the sample, we sum up the probabilities of all possible orders, which is the sum of $n!$ probabilities. If we include all mutations, n will be quite large for some tumor samples. For example, in the lung cancer data, in which 623 genes were sequenced, the number of mutations in one sample is 54, which makes the computation difficult.

We will show how the result is affected as the cutoff for selecting driver genes changes. In our paper, we used the method of Youn and Simon (2011) [3] to select 28 driver genes for lung cancer data with the false discovery rate (FDR) controlled at 5%. If we loosen the cutoff to FDR level 0.2, 40 driver genes are selected. Table D presents the estimates $P_{k,i}^{40}$ obtained from using 40 driver genes and the estimates $P_{k,i}^{28}$ obtained from 28 driver genes. It shows that including more driver genes does not affect the result much. The values of $P_{k,i}^{40}$ and $P_{k,i}^{28}$ are very similar for $k = 1, 2$. Differences between those values increase for $k \geqq 3$. This is mainly caused by somewhat large values of $P_{k,i}^{40}$ for $k \geqq 3$ for some of the extra 12 genes added under FDR 0.2. Especially, the gene GNAS, PIK3C3, PDGFRA, INSRR have large values of $P_{k,i}^{40}$ for some k. Of the extra 12 genes, these are the ones that mutate most frequently and whose p values are smallest. These genes all mutate in samples with at least three mutations, thus have large values of $P_{k,i}^{40}$ for some of $k \geqq 3$. Other genes have very small values of $P_{k,i}^{40}$ for most of k, that they do not affect the result much. This result implies that adding some background genes which mutate infrequently does not affect the estimates of important driver genes.

Glioblastoma multiforme sequencing data analysis

We analyzed glioblastoma multiforme cancer sequencing data which is available from the TCGA data portal. (http://tcga-data.nci.nih.gov/tcga/) There were a total 148 samples in which 601 genes were sequenced in common.

We identified nine tumor driver genes using the method of Youn and Simon (2011)[3] with the false discovery rate controlled at 5%. Since there

Table D: Estimates of $P^{40}_{k,i}$ and $P^{28}_{k,i}$ for lung cancer data

Gene	$P^{40}_{1,i}$	$P^{28}_{1,i}$	$P^{40}_{2,i}$	$P^{28}_{2,i}$	$P^{40}_{3,i}$	$P^{28}_{3,i}$	$P^{40}_{4,i}$	$P^{28}_{4,i}$	$P^{40}_{5,i}$	$P^{28}_{5,i}$
EGFR	0.2	0.2	0.04	0.05	0	0	0	0	0	0
EPHA5	0	0	0	0	0.09	0.07	0	0.09	0	0
CDKN2A	0.01	0.02	0	0	0.04	0.03	0	0	0.09	0.19
FGFR4	0.01	0.01	0.03	0.03	0	0	0	0	0.05	0.08
EPHA3	0	0.02	0	0	0.07	0	0.11	0.15	0	0.07
MYO3B	0.02	0.02	0.01	0.01	0	0	0.06	0.08	0	0
ERBB4	0.02	0.03	0	0	0	0	0.11	0.08	0	0.04
EPHA7	0	0	0	0	0.02	0.04	0	0	0.05	0.07
KRAS	0.41	0.41	0.01	0.01	0	0	0	0	0	0
KDR	0.01	0.01	0	0	0	0	0.14	0.19	0	0
INHBA	0	0	0	0	0.02	0.05	0.05	0	0	0
LTK	0	0	0	0	0.04	0.03	0	0	0.05	0.14
APC	0	0	0.04	0.04	0.05	0.06	0.05	0	0	0.04
NTRK3	0	0	0	0.02	0.1	0.1	0	0	0	0
PTPRD	0.01	0.01	0.01	0.02	0.02	0.05	0	0.1	0.09	0
LRP1B	0.02	0.01	0	0	0.1	0.06	0	0.28	0.06	0
NTRK2	0	0	0	0	0.03	0.04	0	0	0.06	0.1
PTEN	0.01	0.01	0	0	0.03	0.04	0	0	0	0
PAK3	0	0	0	0	0	0.05	0	0	0.06	0
ZMYND10	0	0	0.03	0.03	0	0	0	0	0	0
NF1	0.05	0.05	0.01	0.01	0.02	0.04	0	0	0.08	0.12
RB1	0.01	0.01	0	0	0.03	0.04	0	0	0.08	0.15
NRAS	0.02	0.02	0	0	0	0	0	0	0	0
TFDP1	0.01	0.01	0	0	0	0.04	0.06	0	0	0
ATM	0	0	0.08	0.11	0.06	0.04	0	0	0	0
STK11	0.11	0.12	0.12	0.07	0	0.17	0.13	0	0.02	0
TP53	0.03	0.04	0.54	0.56	0	0	0	0.04	0.03	0
NTRK1	0	0	0.04	0.04	0	0.05	0.03	0	0.04	0
MERTK	0		0.01		0.03		0		0.03	
FYN	0.01		0		0		0		0	
GNAS	0		0		0.09		0		0	
JAK2	0.01		0		0.02		0		0.03	
INSRR	0		0.01		0		0		0.08	
ERAS	0		0		0.01		0.03		0	
PRKACB	0		0		0.02		0		0.01	
ROR2	0		0		0.06		0		0	
PIK3C3	0		0		0		0.09		0	
PIK3CG	0.01		0		0		0		0.04	
PDGFRA	0		0		0		0.14		0	
PTPN11	0		0.01		0.03		0		0	
TP73L	0.02		0		0		0		0.03	

Table E: Estimates of $P_{k,i}$ for glioblastoma multiformes

Gene	$P_{1,i}$		$P_{2,i}$		$P_{3,i}$		$P_{4,i}$	
	MLE	90% CI	MLE	90% CI	MLE	90% CI	MLE	90% CI
CHEK2	0.08	(0.05 , 0.14)	0.09	(0.01 , 0.2)	0.04	(0 , 0.26)	0.09	(0 , 0.33)
EGFR	0.15	(0.1 , 0.23)	0.17	(0.09 , 0.33)	0	(0 , 0.15)	0	(0 , 0.21)
ERBB2	0.02	(0 , 0.08)	0.06	(0 , 0.14)	0.12	(0.01 , 0.32)	0.27	(0.03 , 0.64)
NF1	0.04	(0 , 0.13)	0.22	(0.04 , 0.35)	0.17	(0 , 0.37)	0.18	(0 , 0.68)
PIK3CA	0.06	(0.05 , 0.13)	0	(0 , 0)	0	(0 , 0.07)	0	(0 , 0.12)
PIK3R1	0.07	(0.03 , 0.13)	0	(0 , 0.18)	0.18	(0 , 0.4)	0.01	(0 , 0.23)
PTEN	0.32	(0.17 , 0.41)	0	(0 , 0)	0.2	(0 , 0.63)	0	(0 , 0.16)
RB1	0	(0 , 0.03)	0.13	(0.02 , 0.29)	0.14	(0 , 0.38)	0.04	(0 , 0.19)
TP53	0.24	(0.13 , 0.35)	0.33	(0.17 , 0.53)	0.14	(0 , 0.44)	0.4	(0 , 0.78)

are few samples which have more than four nonsilent mutations, we assumed $P_{n,i} = P4, i$ for $n > 4$. The MLE of $P_{1,i}, P_{2,i}, P_{3,i}, P_{4,i}$ are shown in Table E with 90% confidence interval. It shows that the first mutational event occurs mainly in EGFR, PTEN, TP53. The estimated probability that the first mutation occurs in EGFR is 0.15 with a 90% CI (0.1,0.23). The probability in PTEN is 0.32 with a 90% CI (0.17,0.41) and that in TP53 is 0.24 with a 90% CI (0.13,0.35). The second mutational event mainly occurs in EGFR, NF1,RB1,TP53 with the probability 0.17,0.22,0.13,0.33 respectively. The third mutational event occurs most frequently in PTEN although its value of $P_{3,i}$ is only 0.2 and its CI covers 0. There are many other genes with small values of $P_{3,i}$ and the CIs of all genes except ERBB2 cover 0. The fourth mutational event mainly occurs in ERBB2, NF1, TP53 with the probability 0.27, 0.18, 0.4.

We calculated the conditional probability that a gene mutates early (a gene mutates at the $k'th$ event for $k \leq 2$) or late (a gene mutates at the $k'th$ event for $k > 2$) given that the gene is mutated in the sample. Table F shows the conditional probabilities and their 90% CIs. It shows that EGFR mutates early for sure and this is consistent with the conclusion from the lung cancer data analysis. PTEN mutates early more frequently than late and RB1 mutates late more frequently than late. This is also consistent with the conclusion from the lung and colon cancer analysis. ERBB2 mutates late and PIK3CA mutates early. TP53 mutates late more frequently than early although its estimated value of $P_{k,i}$ is large for every $k = 1, 2, 3, 4$.

Table F: Probabilities that observed mutations occur early* or late* for glioblastoma multiformes

Gene	Early	90% CI	Late	90% CI
CHEK2	0.41	(0.08 , 1)	0.71	(0 , 0.98)
EGFR	1	(1 , 1)	0	(0 , 0)
ERBB2	0.12	(0 , 0.32)	0.96	(0.67 , 1)
NF1	0.39	(0.18 , 0.73)	0.82	(0.58 , 0.98)
PIK3CA	1	(1 , 1)	0	(0 , 0)
PIK3R1	0.27	(0.11 , 1)	0.79	(0 , 0.94)
PTEN	0.7	(0.44 , 1)	0.44	(0 , 0.75)
RB1	0.37	(0.06 , 1)	0.72	(0 , 0.97)
TP53	0.55	(0.38 , 0.67)	0.9	(0.76 , 0.99)

*early means 1_{st} or 2_{nd} event and late means later events

References

1. Ding, L., Getz, G., Wheeler, D. A., Mardis, E. R., Mclellan, M. D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D. M., Morgan, M. B., Fulton, L., Fulton, R. S., Zhang, Q., Wendl, M. C., Lawrence, M. S., Larson, D. E., Chen, K., Dooling, D. J., Sabo, A., Hawes, A. C., Shen, H., Jhangiani, S. N., Lewis, L. R., Hall, O., Zhu, Y., Mathew, T., Ren, Y., Yao, J., Scherer, S. E., Clerc, K., Metcalf, G. A., Ng, B., Milosavljevic, A., Gonzalez-Garay, M. L., Osborne, J. R., Meyer, R., Shi, X., Tang, Y., Koboldt, D. C., Lin, L., Abbott, R., Miner, T. L., Pohl, C., Fewell, G., Haipek, C., Schmidt, H., Dunford-Shore, B. H., Kraja, A., Crosby, S. D., Sawyer, C. S., Vickery, T., Sander, S., Robinson, J., Winckler, W., Baldwin, J., Chirieac, L. R., Dutt, A., Fennell, T., Hanna, M., Johnson, B. E., Onofrio, R. C., Thomas, R. K., Tonon, G., Weir, B. A., Zhao, X., Ziaugra, L., Zody, M. C., Giordano, T., Orringer, M. B., Roth, J. A., Spitz, M. R., Wistuba, I. I., Ozenberger, B., Good, P. J., Chang, A. C., Beer, D. G., Watson, M. A., Ladanyi, M., Broderick, S., Yoshizawa, A., Travis, W. D., Pao, W., Province, M. A., Weinstock, G. M., Varmus, H. E., Gabriel, S. B., Lander, E. S., Gibbs, R. A., Meyerson, M., and Wilson, R. K. *Nature* **455**(7216), 1069–1075 October (2008).

2. Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjoblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K. V., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V. K., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E., and Vogelstein, B. *Science* **318**(5853), 1108–1113 (2007).

3. Youn, A. and Simon, R. *Bioinformatics* **27**(2), 175–181 (2011).