Cancer Variants in the Personal Genome Table: GENCODE v7

Now with fixed ncRNA numbers!

			Number of Genomic Elements			
			GENCODE v7			
			Genes (at least one transcript)	Pseudogenes	ncRNAs	ncRNAs (transcribed)
	Melanoma	SNVs	330	220	1588	129
	Melanoma	Indels	0	1	4	0
	Rest of ICGC	SNVs	897	409	2251	328
	Rest of ICGC	Indels	15	2	8	0
Rubin prostate cancer genomes	PR-0508	SNVs	18	75	292	21
	PR-0581	SNVs	93	117	701	42
	PR-1701	SNVs	26	37	313	12
	PR-1783	SNVs	34	35	344	12
	PR-2832	SNVs	17	34	268	13
	PR-3027	SNVs	50	34	415	17
	PR-3043	SNVs	17	19	229	7

Cancer Variants in the Personal Genome Table: Functional Elements

			Functional Elements - Binding Sites						
			Pol2	Ctcf	P300	cMyc	Max	Average (over all TFs)	TOTAL
	Melanoma	SNVs	142	127	22	22	4	59	31,035
	Melanoma	Indels	0	0	0	0	0	0.054	65
	Rest of ICGC	SNVs	280	254	43	39	3	116	68,711
	Rest of ICGC	Indels	1	1	0	0	0	0.189	76
Rubin prostate cancer genomes	PR-0508	SNVs	42	21	5	3	0	5.23	3,421
	PR-0581	SNVs	145	77	9	6	0	21.19	10,710
	PR-1701	SNVs	45	24	6	1	0	7.53	3,500
	PR-1783	SNVs	37	26	6	1	0	7.5	4,129
	PR-2832	SNVs	48	22	5	2	0	6.04	3,165
	PR-3027	SNVs	56	27	13	2	0	10.46	4,745
	PR-3043	SNVs	26	16	2	0	0	3.36	2,408

SNV Count Ratios: Normalized to Gene Count

		Genes	Pseudogenes	ncRNA	ncRNA transcribed	Pol2	Ctcf	P300	cMyc	Max	Average (over all TFs)
Counts	All variants NA12878 maternal SNPs	3748	3120	7171	3985	8417	9022	933	864	138	3077
	All variants NA12878 paternal SNPs	3770	3076	7168	3988	8448	9037	989	847	138	3095
	Melanoma SNVs	330	220	1588	129	142	127	22	22	4	59
	Rest of ICGC SNVs	897	409	2251	328	280	254	43	39	3	116
	PR-0508 SNVs	18	75	292	21	42	21	5	3	0	5.23
	PR-0581 SNVs	93	117	701	42	145	77	9	6	0	21.19
	PR-1701 SNVs	26	37	313	12	45	24	6	1	0	7.53
	PR-1783 SNVs	34	35	344	12	37	26	6	1	0	7.5
	PR-2832 SNVs	17	34	268	13	48	22	5	2	0	6.04
	PR-3027 SNVs	50	34	415	17	56	27	13	2	0	10.46
	PR-3043 SNVs	17	19	229	7	26	16	2	0	0	3.36
Ratio Normalized to Gene Count	All variants NA12878 maternal SNPs	1	0.83244397	1.91328709	1.063233725	2.24573106	2.40715048	0.24893276	0.23052295	0.03681964	0.820971185
	All variants NA12878 paternal SNPs	1	0.815915119	1.90132626	1.057824934	2.24084881	2.39708223	0.26233422	0.22466844	0.03660477	0.820954907
	Melanoma SNVs	1	0.666666667	4.81212121	0.390909091	0.43030303	0.38484848	0.06666667	0.06666667	0.01212121	0.178787879
	Rest of ICGC SNVs	1	0.455964326	2.50947603	0.365663322	0.31215162	0.28316611	0.04793757	0.04347826	0.00334448	0.129319955
	PR-0508 SNVs	1	4.166666667	16.2222222	1.166666667	2.33333333	1.16666667	0.27777778	0.16666667	0	0.290555556
	PR-0581 SNVs	1	1.258064516	7.53763441	0.451612903	1.55913978	0.82795699	0.09677419	0.06451613	0	0.227849462
	PR-1701 SNVs	1	1.423076923	12.0384615	0.461538462	1.73076923	0.92307692	0.23076923	0.03846154	0	0.289615385
	PR-1783 SNVs	1	1.029411765	10.1176471	0.352941176	1.08823529	0.76470588	0.17647059	0.02941176	0	0.220588235
	PR-2832 SNVs	1	2	15.7647059	0.764705882	2.82352941	1.29411765	0.29411765	0.11764706	0	0.355294118
	PR-3027 SNVs	1	0.68	8.3	0.34	1.12	0.54	0.26	0.04	0	0.2092
	PR-3043 SNVs	1	1.117647059	13.4705882	0.411764706	1.52941176	0.94117647	0.11764706	0	0	0.197647059

Indel Count Ratios: Normalized to Pseudogene Count

		Genes	Pseudogenes	ncRNA	ncRNA transcribed	Pol2	Ctcf	P300	cMyc	Max	Average (over all TFs)
Counts	All variants NA12878 maternal indels	110	166	654	4071	702	2044	1324	175	155	20
	All variants NA12878 paternal indels	89	182	655	4086	707	2073	1311	161	156	24
	Melanoma indels	0	1	4	0	0	0	0	0	0	0.054
	Rest of ICGC indels	15	2	8	0	1	1	0	0	0	0.189
Ratio Normalized to Gene Count	All variants NA12878 maternal indels	0.6626506	1	3.93975904	24.52409639	4.22891566	12.313253	7.97590361	1.05421687	0.93373494	0.120481928
	All variants NA12878 paternal indels	0.48901099	1	3.5989011	22.45054945	3.88461538	11.3901099	7.2032967	0.88461538	0.85714286	0.131868132
	Melanoma indels	0	1	4	0	0	0	0	0	0	0.054
	Rest of ICGC indels	7.5	1	4	0	0.5	0.5	0	0	0	0.0945

Comparison to Random Annotations

- Take all sets of annotations and "shift" them some number of bp forward on the chromosome
  - +1000
  - +2000
  - +5000
  - +10000
  - +20000
- Count # of annotations that contain at least one variant
- Compare to original counts

Melanoma SNVs

Shift	Genes	Pseudogenes	ncRNA	ncRNA transcribed	Pol2	Ctcf	P300	cMyc	Max	Average (over all TFs)
0	330	220	1588	129	142	127	22	22	4	59
+1000	291	221	1588	122	311	190	23	17	0	52.89
+2000	276	197	1590	143	300	209	25	13	0	54.89
+5000	280	212	1579	134	325	199	31	15	2	61
+10000	294	216	1556	141	340	225	42	29	2	65.97
+20000	324	213	1548	144	359	220	27	14	3	62.31

Rest of ICGC SNVs

Shift	Genes	Pseudogenes	ncRNA	ncRNA transcribed	Pol2	Ctcf	P300	cMyc	Max	Average (over all TFs)
0	897	409	2251	328	280	254	43	39	3	116
+1000	608	385	2263	277	645	399	72	37	4	119.22
+2000	606	385	2252	308	656	429	74	42	9	132.77
+5000	599	402	2250	342	693	412	75	57	3	130.89
+10000	606	408	2263	302	762	423	66	39	4	132.91
+20000	625	394	2257	329	742	431	68	56	7	132.42

PR-0508 SNVs

Shift	Genes	Pseudogenes	ncRNA	ncRNA transcribed	Pol2	Ctcf	P300	cMyc	Max	Average (over all TFs)
0	18	75	292	21	42	21	5	3	0	5.23
+1000	34	66	294	20	49	25	5	1	0	7.31
+2000	40	67	293	16	52	25	8	4	0	9.34
+5000	30	61	300	16	57	33	5	6	0	10.04
+10000	38	67	293	22	60	29	8	3	0	9.57
+20000	45	60	291	33	59	28	3	2	0	8.93

Melanoma Indels

Shift	Genes	Pseudogenes	ncRNA	ncRNA transcribed	Pol2	Ctcf	P300	cMyc	Max	Average (over all TFs)
0	0	1	4	0	0	0	0	0	0	0.054
+1000	1	1	4	0	0	0	1	0	0	0.108
+2000	1	1	5	0	0	1	1	0	0	0.149
+5000	0	1	5	0	0	0	0	0	0	0
+10000	0	2	5	0	0	0	0	0	0	0.135
+20000	1	1	9	0	1	0	0	0	0	0.027

Rest of ICGC Indels

Shift	Genes	Pseudogenes	ncRNA	ncRNA transcribed	Pol2	Ctcf	P300	cMyc	Max	Average (over all TFs)
0	15	2	8	0	1	1	0	0	0	0.189
+1000	2	1	8	0	1	2	0	0	0	0.257
+2000	3	1	7	0	2	1	0	0	0	0.473
+5000	2	1	6	1	1	0	0	0	0	0.243
+10000	2	0	7	1	3	0	1	1	0	0.649
+20000	1	1	5	0	1	1	0	0	0	0.135