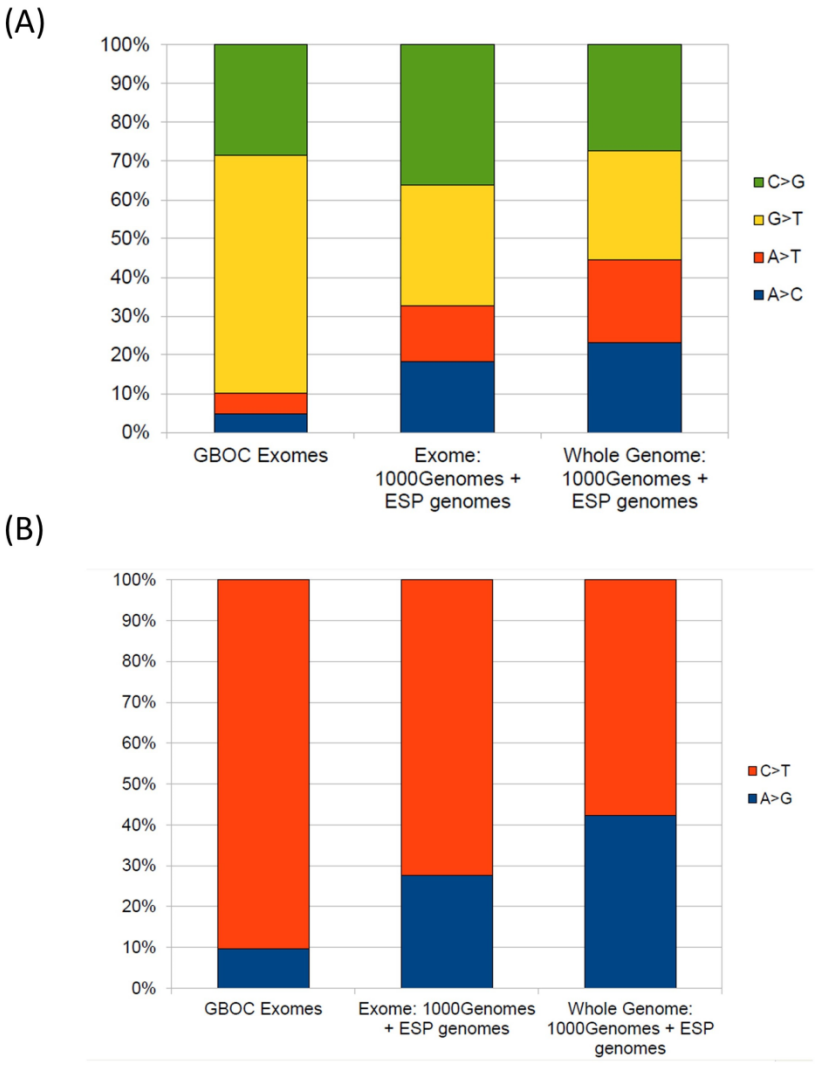
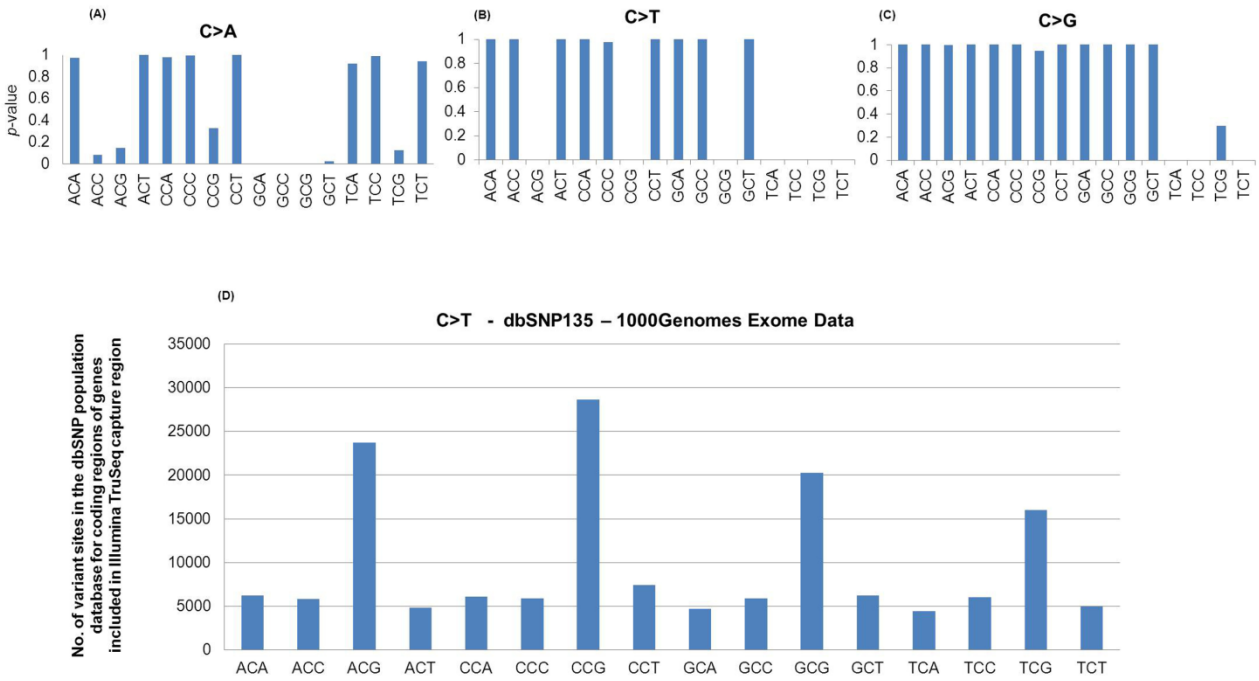
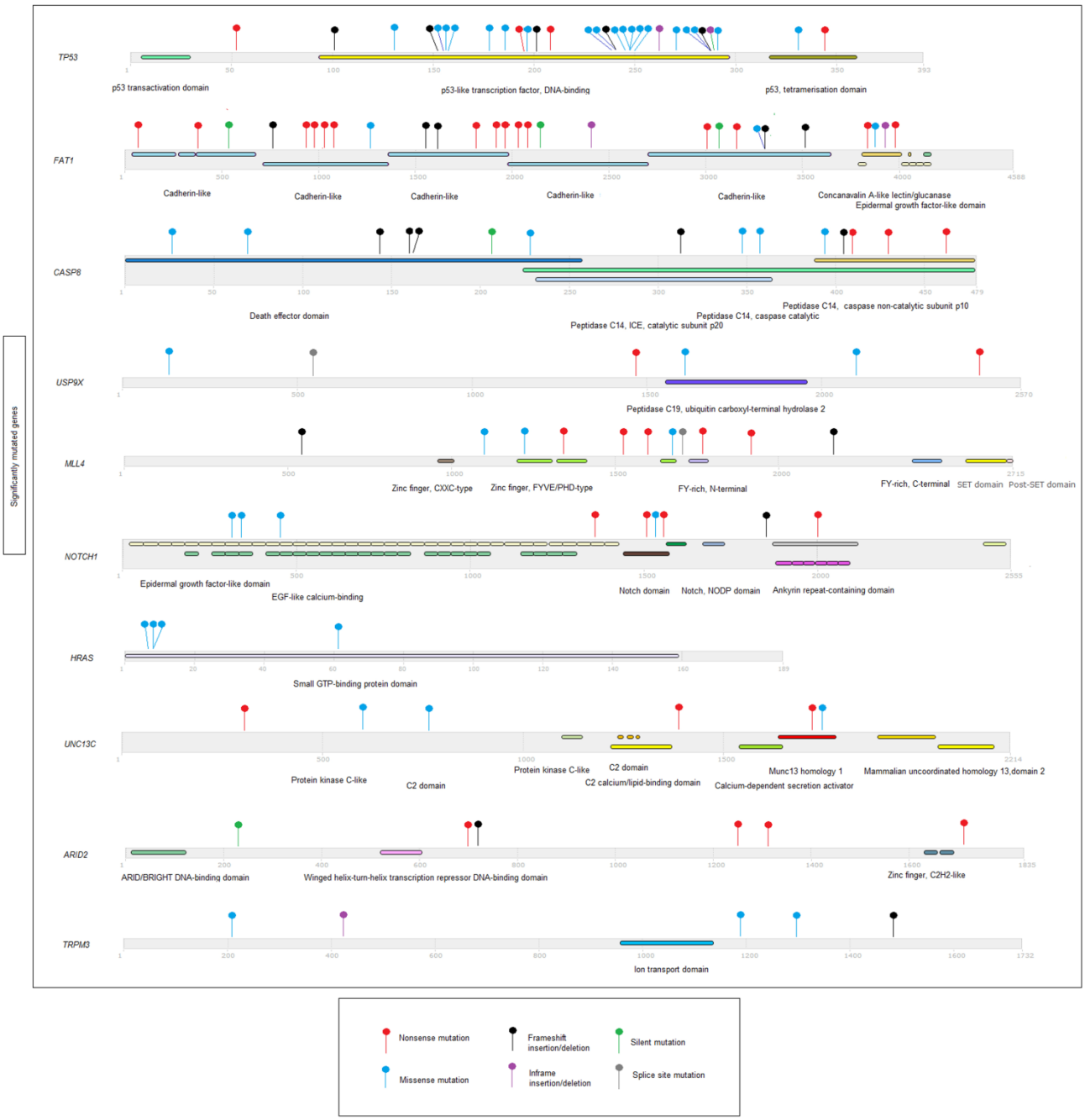


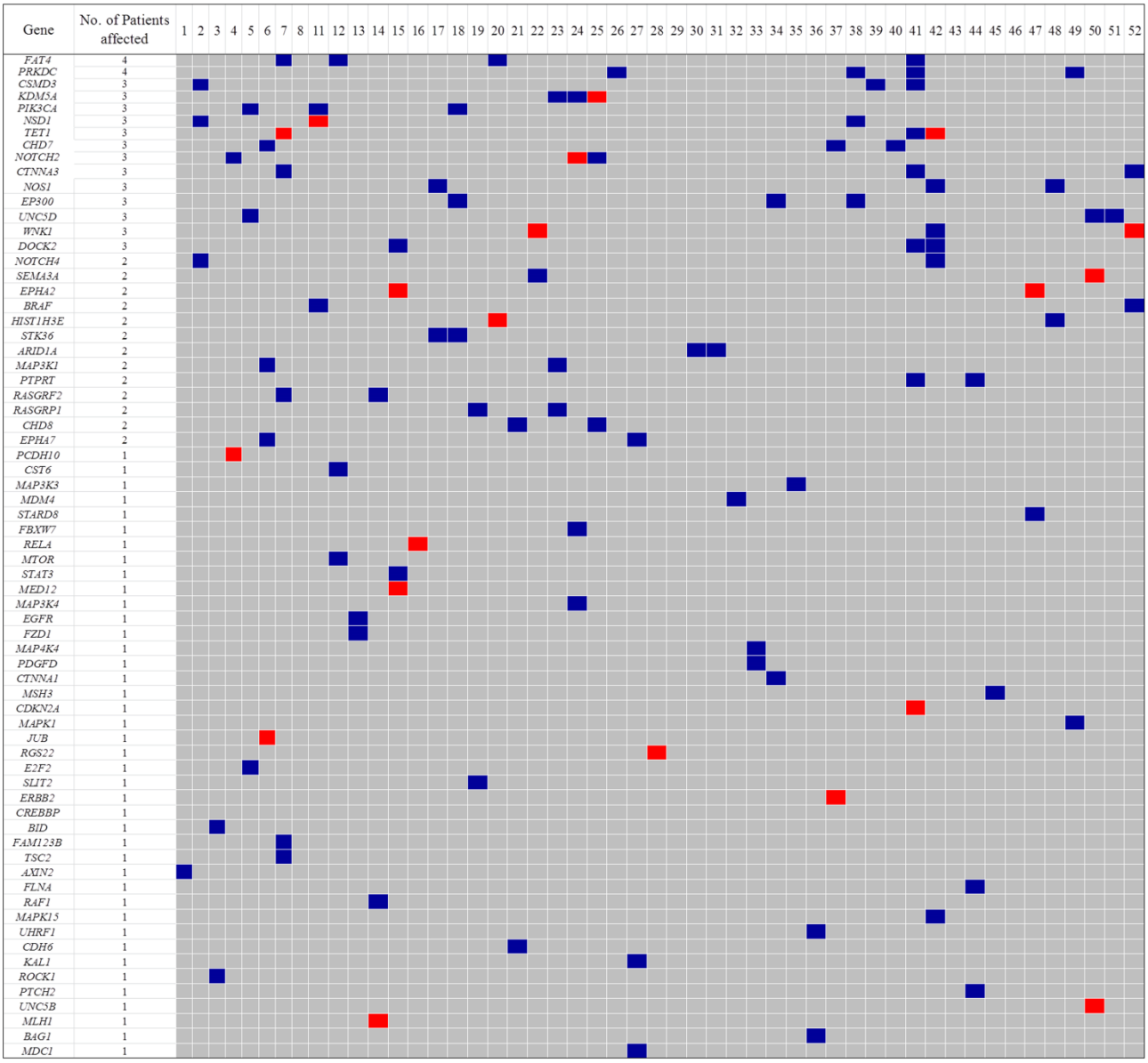
Supplementary Figure S1: Types and frequencies of mutations in gingivo-buccal oral cancer patients and in general populations. Frequencies of (A) Transversion and (B) Transition mutations in exomes of gingivo-buccal oral squamous cell tumors and in general populations (exome and whole genome)



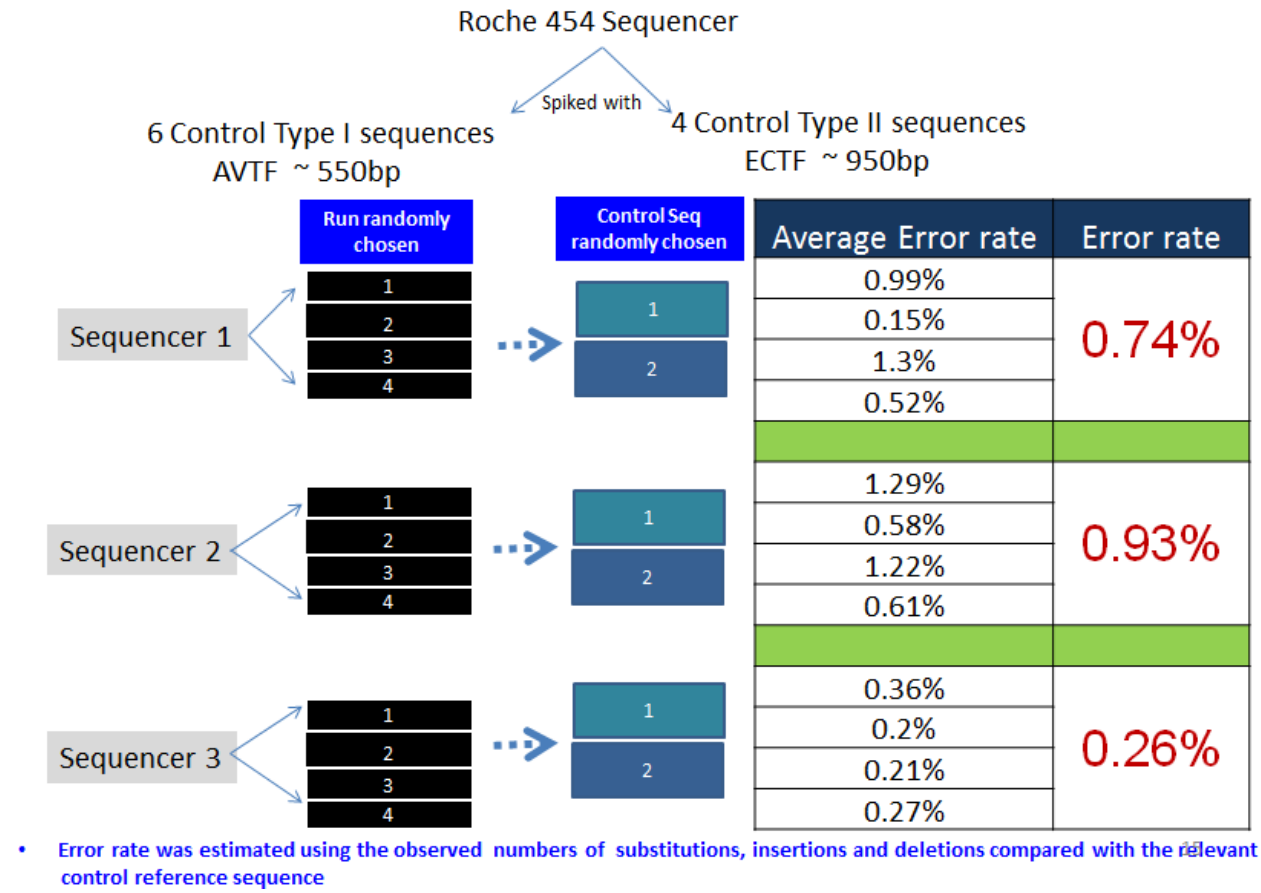
Supplementary Figure S2: *p*-Values pertaining to tests of equality of observed vs. expected frequencies of different C>X mutations. [(A) C>A+G>T, (B) C>T+G>A, (C) C>G+G>C] occurring at various possible sequence contexts. Expected frequencies were calculated by randomly sampling 10,000 triplets from the human exome captured by the Illumina TruSeq protocol. (D) Counts of C>T (+ G>A) changes in 1000Genomes and ESP data archived in dbSNP135 for different motifs.



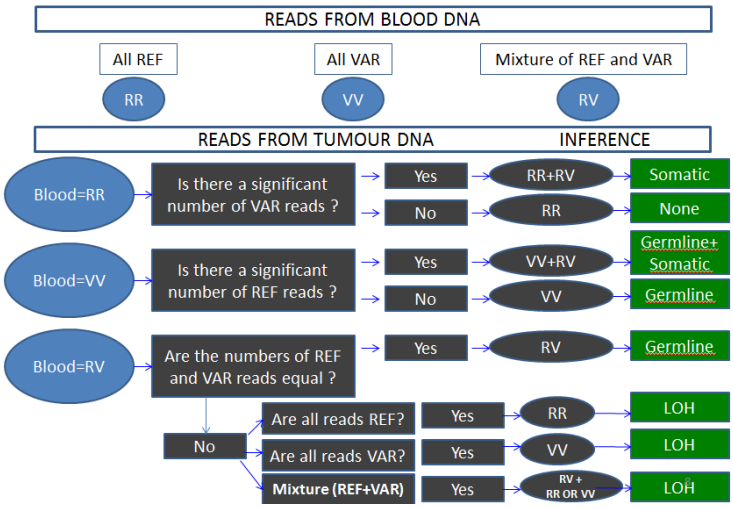
Supplementary Figure S3: Distributions of various types of mutations in the 10 genes found to be frequently mutated in gingivo-buccal oral squamous cell tumors. The backbone of each gene was copied from UniProt database (*www.uniprot.org).*



Supplementary Figure S4: Distributions of mutations observed in gingivo-buccal oral cancer patients. Distributions of all mutations in each OSCC-GB patient in genes other than those that are frequently mutated and displayed in Fig. 1.



Supplementary Figure S5: Strategy used for estimating sequencing error rate. Average and estimated error rates are shown. See Supplementary Methods for more details.



Supplementary Figure S6: Inference on genomic alterations. These data are based on genotypes inferred in tumor DNA, conditional on genotype inferred in blood DNA (Supplementary Methods)

Supplementary Table S1: Demographic, risk-exposure and clinical characteristics of 50 patients with gingivo-buccal oral squamous cell carcinoma

	Number	%
Age-group (yrs.) [Age Range: 26-70; Mean: 48.02±1.4]		
<40	7	14
40-45	14	28
46-50	10	20
51-55	9	18
56-60	3	6
>60	7	14
Gender		
Male	44	88
Female	6	12
Risk-habit		
Exposure to tobacco, but no exposure to alcohol	19	38
Exposure to both tobacco and alcohol	29	58
No exposure to tobacco or alcohol or areca nut	2	4
Tumor stage at first presentation*		
I-II	3	6
III	4	8
IV	43	86
Tumor stage (TNM)*		
T2 N0	2	4
T2 N1	1	2
T3 N0	3	6
T3 N2b	1	2
T4a N0	20	40
T4a N1	7	14
T4a N2b	11	22
T4a N2c	4	8
T4a N3	1	2
HPV infection		
Present	13**	26
Absent	37	74
HSV 1/2 infection		
Present	1	2
Absent	49	98
* All patients were M0 at first presentation when tissue was collected for analysis		
** 11 tumors with high-risk subtypes 16 and 18; 2 mixed subtypes		

Supplementary Table S2. Summary of sequence data and characteristics of single nucleotide variants and insertions/deletions identified in 50 gingivo-buccal oral squamous cell carcinoma patients using Illumina Hiseq-2000 and Roche GS-FLX platforms

Patient ID	Sample ID (T=Tumor; B=Blood)	No. of bases sequenced		No. of bases mapped to genome		Coverage (X) after QC and alignment		Proportion of targeted bases with at least 10 reads in blood and tumor		No. of germline mutations identified in target region [CDS+ncRNA]		No. of somatic mutations identified in target region [CDS+ncRNA]		
		HiSeq-2000	GS-FLX	HiSeq-2000	GS-FLX	HiSeq-2000	GS-FLX	HiSeq-2000	GS-FLX	HiSeq-2000	GS-FLX	HiSeq-2000		GS-FLX (CDS only)
												CDS	ncRNA	
1	1T	5852094400	4841250158	5504929900	4734448151	37.33	20.54	0.85	0.72	9256	3682	43	5	15
	1B	4684292400	2953713325	4441296700	2874450974	31.94	23.08	0.81	0.79					
2	2T	7326241800	5699352476	6892411300	5586503794	46.2	48.45	0.87	0.96	9728	5591	315	17	219
	2B	7528778600	3580371222	7119081200	3507624614	48.95	28.65	0.86	0.91					
3	3T	5315757200	4209665984	4997253800	4130940119	38.2	36.37	0.84	0.81	9308	4852	71	5	36
	3B	5000240000	3084646281	4688843300	2978007330	35.94	23.4	0.82	0.76					
4	4T	6185276000	4033359327	5838276400	3953734208	40.88	36.08	0.85	0.83	9766	4642	66	3	20
	4B	6611650800	2948160133	6232532700	2889418335	42.52	22.98	0.86	0.69					
5	5T	5553902400	8006918845	5211369200	7706692904	36.36	40.12	0.83	0.86	9538	4938	79	7	30
	5B	7079454200	4948228142	6650258900	4721612815	45.44	24.88	0.87	0.75					
6	6T	7173269000	4231021319	6750617900	4127588244	41	37.2	0.84	0.85	9290	4699	81	4	38
	6B	7660425800	3676692194	7163214400	3537639892	48.65	22.4	0.88	0.75					
7	7T	7826834800	4313995532	7370792700	4204169419	50.43	38.13	0.87	0.86	9563	5122	391	18	260
	7B	5907254000	4335896744	5552081900	4145177164	38.47	30.72	0.86	0.85					
8	8T	6773775600	4392387594	6385444300	4297651373	41.73	40.76	0.84	0.85	9292	4999	71	5	28
	8B	7109609000	3070499236	6722216600	2987740773	46.58	27.8	0.85	0.8					
11	11T	5268740200	4104290865	4949656300	4019043921	36.99	37.53	0.83	0.85	9497	4586	107	3	35
	11B	5459518800	2927219274	5128962800	2855953608	38.24	24.3	0.84	0.7					
12	12T	6022601400	4299697290	5621411800	4219200737	41.04	36.05	0.85	0.84	9699	5114	152	5	85
	12B	5777361400	2800803471	5424302600	2727365585	39.22	23.61	0.85	0.77					
13	13T	6065731000	5119963935	5700559500	5003178799	37.98	22.41	0.84	0.72	9710	3891	117	11	28
	13B	7980666200	4351610508	7511303500	4240993468	50.95	22.05	0.87	0.74					
14	14T	6119726800	3658738847	5736042400	3549607207	40.92	35.76	0.85	0.81	9592	4681	214	13	78
	14B	6502918200	2693794108	6133575800	2625411963	42.49	23.62	0.86	0.71					
15	15T	5786750600	3906308669	5416281200	3777747230	36.34	35.18	0.83	0.8	9945	4858	113	9	36
	15B	6675640600	3039782452	6241482000	2968103641	42.69	24.96	0.87	0.7					
16	16T	7412396000	3794398026	6876909000	3705496679	45.07	32.3	0.87	0.78	9186	4384	173	9	54
	16B	6368216800	2457984486	5997051300	2386944731	42.13	25.39	0.86	0.75					
17	17T	8341803800	5396766797	7877066300	5269667444	24.87	23.13	0.79	0.74	9130	3792	73	10	3
	17B	8323235600	3526182254	7899979000	3399821423	42.1	24.1	0.83	0.66					
18	18T	7447417600	4784626576	7044948200	4672236613	20.47	21.48	0.74	0.7	8696	3510	116	8	21
	18B	7482850400	2736134489	7100781500	2670328097	34.21	24.01	0.81	0.71					
19	19T	8567538200	4662411892	8114527600	4498611528	29.57	38.14	0.81	0.8	8931	4475	76	10	14
	19B	4128820400	2409459366	3911302000	2342656081	22.9	24.88	0.75	0.7					

20	20T	7092579200	3824067154	6681060100	3743106757	29.09	31.23	0.8	0.74	7878	3467	140	16	38
	20B	7650955200	2480071915	7252731400	2432301252	42.92	19.39	0.84	0.65					
21	21T	6489023800	3895695288	6180318300	3817761575	30.2	35.96	0.78	0.79	9062	4503	90	1	42
	21B	6771808600	2536728297	6440560200	2485538813	38.73	23.52	0.84	0.71					
22	22T	6507445600	3904419841	6174457400	3823230427	32.38	36.31	0.8	0.8	9333	4686	91	8	29
	22B	7458092200	2599845447	7084316000	2547652895	42.85	23.04	0.84	0.7					
23	23T	5703121200	4139601088	5319751300	4023389019	13.19	28.92	0.58	0.77	7578	3651	51	7	11
	23B	7855216800	4221454053	7379481500	4115234610	25.99	22.47	0.74	0.76					
24	24T	8749243600	3496097260	8233118600	3402918571	32.38	26.78	0.81	0.75	9244	3812	155	8	46
	24B	7956352600	2858292464	7536069700	2803690059	31.79	19	0.79	0.64					
25	25T	7593911800	3534768695	7144678200	3460209103	24.73	35.59	0.78	0.78	8801	4550	103	11	20
	25B	7573600000	2840967554	7175910000	2745660407	31.54	23.68	0.81	0.71					
26	26T	6431414400	4198912475	6122001000	4077893230	21.15	37.27	0.72	0.79	7531	4007	68	10	25
	26B	6212654000	2448090883	5916902800	2395842693	21.83	24.54	0.7	0.72					
27	27T	6859494800	4768616737	6512679400	4659176520	19.78	21.53	0.71	0.67	8112	3316	95	13	26
	27B	6486097200	2782351315	6178496700	2718658117	31.28	22.59	0.79	0.67					
28	28T	10794991400	4455991708	10150489700	4371924178	59.79	27.53	0.88	0.72	9037	4157	31	6	5
	28B	6482156400	2563353973	6178161300	2497391052	30.26	23.89	0.79	0.71					
29	29T	5999494200	3837979157	5620968300	3745299088	38.52	36.03	0.84	0.79	9790	4976	98	4	44
	29B	6510205400	2665164854	6096895500	2599967232	42.43	25.95	0.86	0.73					
30	30T	5910229400	4288549101	5565884600	4188265818	38.84	36.4	0.83	0.77	9316	4702	36	6	15
	30B	5015189000	2365956827	4718189200	2309701224	33.24	24.34	0.82	0.72					
31	31T	7246268400	4316635969	6812092800	4207650461	47.88	37.19	0.87	0.79	10025	5030	51	6	12
	31B	6261350600	2503689710	5879305200	2446566582	42.45	25.91	0.85	0.74					
32	32T	6185784200	4069755726	5784779500	3976587079	40.44	37.98	0.84	0.81	9351	4342	91	4	36
	32B	5404305000	2361856690	5046247700	1813654950	33.95	18.88	0.84	0.65					
33	33T	5870605800	3931962202	5507090300	3830628469	37.19	37.47	0.84	0.79	9702	4775	147	12	72
	33B	5499670800	2984900934	5186688700	2908102545	34.97	24.91	0.83	0.71					
34	34T	6117311200	4351799916	5679997300	4210890919	38.77	42.2	0.85	0.84	9822	4744	76	5	29
	34B	6247621000	3417107303	5804077000	3262737409	40.39	24.25	0.86	0.71					
35	35T	5202206800	4268442925	4877376400	4169875666	36.59	38.95	0.83	0.78	9568	4729	77	3	28
	35B	5044837200	2590782564	4733673200	2524621794	32.85	23.93	0.83	0.71					
36	36T	6254391800	4199230299	5807905300	4101503727	41	39.25	0.86	0.83	9389	4982	16	2	1
	36B	4147093000	2987004927	3887696400	2923577405	29.13	26.73	0.81	0.75					
37	37T	5797229400	3958489825	5432844000	3860676318	39.59	39.04	0.84	0.82	8543	4485	98	3	42
	37B	3843379000	2932739145	3595437400	2854027943	27.23	28.46	0.8	0.76					
38	38T	5625031600	4160505680	5227151400	4071721709	38.33	38.66	0.85	0.82	9516	4880	101	8	46
	38B	4855726200	2683553495	4545073700	2624451644	33.71	26.28	0.82	0.75					
39	39T	4941651800	4027848534	4663652300	3946087045	34.41	38.69	0.82	0.8	9611	4926	72	6	29
	39B	4919710400	3097972257	4630390500	2963489477	33.43	23.3	0.82	0.72					
40	40T	6795237000	4081636252	6299661500	3977907400	44.57	38.58	0.87	0.85	9834	5633	13	2	1
	40B	5803256000	3301506490	5369766600	3216375688	36.36	28.39	0.86	0.79					
41	41T	4522271800	5271032183	4187699800	4963080792	29.71	39.75	0.81	0.81	9316	4771	939	39	592
	41B	5397637200	3204881220	5023582700	3115954727	35.5	25.27	0.83	0.72					

42	42T	6264672800	5790658794	5808005500	5683373141	41.01	40.04	0.85	0.78	9276	4532	96	9	41
	42B	5112781000	3659704229	4725826500	3482890860	34.06	24.46	0.84	0.73					
43	43T	5593548000	4654067898	5227652100	4535332271	37.29	37.66	0.85	0.8	9826	5087	62	4	25
	43B	5718461200	3335507840	5327832400	3250013626	36.4	27.84	0.86	0.75					
44	44T	5776306200	3746407718	5422457800	3646790920	39.39	40.39	0.86	0.86	9918	5549	80	2	37
	44B	5369173200	2765853352	5044359300	2684356911	36.38	28.79	0.84	0.8					
45	45T	5444021800	3859800961	5134209800	3752201451	37.64	37.86	0.85	0.86	10218	5844	60	5	26
	45B	5614669800	2937575134	5293616000	2849240000	38.09	28.97	0.86	0.82					
46	46T	7518061000	3818484089	7094690200	3725240457	40.36	38.16	0.84	0.83	9669	5351	73	8	24
	46B	6195963200	2603890507	5869107400	2537504335	42.5	27.59	0.84	0.79					
47	47T	7279737400	3908046027	6863597700	3794467085	51	38.16	0.87	0.84	10209	5625	57	5	27
	47B	6805539600	2600974992	6398239400	2533768464	47.09	26.71	0.87	0.79					
48	48T	4032024600	4329464511	3803776000	4199081178	27.91	35.76	0.79	0.83	8967	4935	51	5	26
	48B	7391912400	3606974540	6986211500	3468249563	51.47	27.89	0.87	0.79					
49	49T	5793915000	4265717625	5444643500	4152937900	36.41	36.98	0.82	0.83	9392	5222	56	6	34
	49B	6143393800	4307857650	5787068300	4103774965	39.89	27.16	0.83	0.79					
50	50T	6554887000	4050510810	6165415000	3949917954	40.52	35.83	0.84	0.84	9172	4997	62	4	33
	50B	5412112200	2849644330	5106942000	2776484260	33.23	24.73	0.81	0.76					
51	51T	5403627200	3929436426	5126305400	3825186150	29.23	34.84	0.8	0.82	9255	5153	68	3	38
	51B	6074660400	2787616660	5757011600	2713088111	36.11	25.47	0.83	0.77					
52	52T	6400000000	3949914170	6028882800	3837057617	38.96	35.35	0.84	0.82	9315	5172	84	5	43
	52B	6400000000	3179204686	6045680800	3079245432	37.89	27.2	0.84	0.78					

Supplementary Table S3. Descriptive statistics of various parameters pertaining to mutations

Statistic	Minimum	Maximum	Median	Mean±SE
Total No. of mutations	13	939	79	112.78±19.25
Total no. of mutations, excluding synonymous mutations	12	637	60	85.82±13.28
Ratio of non-synonymous to synonymous mutations	2.11	12	3.5	3.70±0.21
Mutation rate per Mb	0.39	29.29	2.51	3.52±0.59
Mutation rate per Mb, excluding synonymous mutations	0.36	19.87	2.02	2.65±0.41

Supplementary Table S4. Numbers and types of mutations in each of the 50 gingivo-buccal oral squamous cell carcinoma patients

Patient no.	Coding regions and essential splice sites												Total	Mutations per mb of DNA	Ratio of non-synonymous to synonymous mutations	ncRNA	
	SNV							Indel									
	Synonymous	Missense	Nonsense	Splice site	Nonstop	Translation start site		Frame-shift deletion	Frame-shift insertion	Splice site	In-frame insertion	In-frame deletion				SNV	Indel
1	11	24	6	1				1					43	1.39	2.91	4	1
2	87	195	17	5		4		4	2			1	315	9.50	2.62	15	2
3	16	50	2	1					1			1	71	2.21	3.44	3	2
4	20	37	7	1					1				66	2.00	2.30	3	0
5	21	52	2	2				1				1	79	2.37	2.76	7	0
6	24	49	2	2				2	1			1	81	2.41	2.38	3	1
7	84	287	13	3				3			1		391	11.78	3.65	14	4
8	12	46	8	3				1	1				71	2.19	4.92	5	0
11	24	66	7	3				5	1			1	107	3.31	3.46	3	0
12	30	108	7	2				3	1			1	152	4.67	4.07	5	0
13	27	79	6	1		1		2				1	117	3.52	3.33	9	2
14	58	115	12	6		1		20	2				214	6.47	2.69	9	4
15	28	67	7	5		1		5					113	3.43	3.04	8	1
16	47	108	10	3		1		1	2			1	173	5.21	2.68	9	0
17	20	48	2	2				1					73	2.31	2.65	10	0
18	23	84	3	4					1			1	116	3.76	4.04	8	0
19	13	50	5	3		2		1	1			1	76	2.64	4.85	8	2
20	36	92	8	1	1			1	1				140	4.35	2.89	16	0
21	18	66	4					1				1	90	2.84	4.00	1	0
22	20	57	6	1		2		5					91	2.84	3.55	7	1
23	11	34	4	2									51	1.85	3.64	6	1
24	40	98	11	3				1	1		1		155	5.12	2.88	8	0
25	15	68	11	3		2		1	1			2	103	3.36	5.87	8	3
26	12	52	1	1				2					68	2.60	4.67	10	0
27	21	63	5	2				1	3				95	3.17	3.52	12	1
28	6	21	1	1		1		1					31	1.03	4.17	5	1
29	28	54	11	2		1		1	1				98	2.98	2.50	4	0
30	7	23	3					2				1	36	1.14	4.14	4	2
31	10	36		3								2	51	1.57	4.10	4	2
32	18	53	13	3				1	2			1	91	2.82	4.06	3	1
33	33	101	9	1				1	1		1		147	4.59	3.45	8	4
34	18	51	3	1		1		1				1	76	2.31	3.22	4	1
35	16	52	3	1				3	1	1			77	2.42	3.81	3	0
36	2	12		1				1					16	0.48	7.00	2	0
37	28	56	5	5	1	1		1				1	98	3.16	2.50	3	0
38	31	62	4					3	1				101	3.19	2.26	6	2
39	12	48	2	3				3	2	1		1	72	2.27	5.00	5	1

40	1	8		1						1		2	13	0.39	12.00	1	1
41	302	578	46	9		2		1	1				939	29.29	2.11	39	0
42	22	63	6	2				1	1			1	96	2.93	3.36	8	1
43	17	37	3	2				1	1		1		62	1.88	2.65	3	1
44	19	49	4			3		3				2	80	2.43	3.21	2	0
45	12	42	2						2			2	60	1.82	4.00	4	1
46	16	48	4					3		1		1	73	2.26	3.56	8	0
47	12	36	4			1		3	1				57	1.70	3.75	5	0
48	11	34	3	1		1			1				51	1.54	3.64	5	0
49	17	33	2			4							56	1.76	2.29	4	2
50	13	41	6					2					62	1.96	3.77	3	1
51	15	42	5					2	1		1	2	68	2.15	3.53	3	0
52	14	54	6	4		1		3	2				84	2.62	5	5	0
Total	1398	3629	311	104	2	30		99	38		5	30	5646	2.62	3.04	5	0

Supplementary Table S5. Summary of results of MutSigCV analysis

Gene*	p-value
CASP8	$<10^{-13}$
TP53	$<10^{-13}$
FAT1	4.10×10^{-13}
HRAS	3.42×10^{-05}
ARID2	1.85×10^{-04}
TRPM3	1.84×10^{-03}
UNC13C	1.90×10^{-03}
USP9X	2.56×10^{-03}
MLL4	1.22×10^{-02}
NOTCH1	1.43×10^{-02}

* This set of genes that turned out to be significant is identical with that identified to be significant by Genome MuSiC analysis (see text for details). P-values were generated using a Z-test.

Supplementary Table S6. List of mutations in genes significantly mutated in gingivo-buccal oral squamous cell carcinoma

Gene	Sl. no. of mutation in gene	ID of patient with mutation	Chromosome no.	Nucleotide position (hg 19)		Reference allele	Variant Allele	Transcript_ID	Exon no.	Amino-acid position	Variant-class	Amino-acid	
				Start	Stop							Reference	Variant
TP53	1	7	17	7574003	7574003	G	A	NM_001126112	10	1218	Nonsense_Mutation	R	*
	2	41	17	7574026	7574026	C	A	NM_001126112	10	1195	Missense_Mutation	G	V
	3	43	17	7577082	7577082	C	T	NM_001126112	8	1050	Missense_Mutation	E	K
	4	33	17	7577085	7577086	-	TGTGCGCCG	NM_001126112	8	1046_1047	In_Frame_Ins	-	RRT
	5	52	17	7577092	7577092	C	-	NM_001126112	8	1040	Frame_Shift_Del	R	fs
	6	34	17	7577093	7577093	C	T	NM_001126112	8	1039	Missense_Mutation	R	Q
	7	16	17	7577094	7577094	G	A	NM_001126112	8	1038	Missense_Mutation	R	W
		17	17	7577094	7577094	G	A	NM_001126112	8	1038	Missense_Mutation	R	W
		45	17	7577094	7577094	G	A	NM_001126112	8	1038	Missense_Mutation	R	W
		5	17	7577094	7577094	G	A	NM_001126112	8	1038	Missense_Mutation	R	W
	8	31	17	7577121	7577121	G	A	NM_001126112	8	1011	Missense_Mutation	R	C
	9	25	17	7577144	7577146	AGT	-	NM_001126112	8	986_988	In_Frame_Del	LL	L
	10	26	17	7577532	7577532	G	A	NM_001126112	7	943	Missense_Mutation	P	L
	11	13	17	7577538	7577538	C	T	NM_001126112	7	937	Missense_Mutation	R	Q
	12	49	17	7577539	7577539	G	A	NM_001126112	7	936	Missense_Mutation	R	W
		50	17	7577539	7577539	G	A	NM_001126112	7	936	Missense_Mutation	R	W
	13	27	17	7577548	7577548	C	T	NM_001126112	7	927	Missense_Mutation	G	S
	14	6	17	7577558	7577558	G	-	NM_001126112	7	917	Frame_Shift_Del	S	fs
	15	41	17	7577559	7577559	G	T	NM_001126112	7	916	Missense_Mutation	S	Y
	16	45	17	7577565	7577565	T	C	NM_001126112	7	910	Missense_Mutation	N	S
	17	28	17	7578212	7578212	G	A	NM_001126112	6	831	Nonsense_Mutation	R	*
	18	38	17	7578247	7578247	A	-	NM_001126112	6	796	Frame_Shift_Del	L	fs
	19	2	17	7578260	7578260	C	T	NM_001126112	6	783	Missense_Mutation	V	M
	20	19	17	7578263	7578263	G	A	NM_001126112	6	780	Nonsense_Mutation	R	*
		8	17	7578263	7578263	G	A	NM_001126112	6	780	Nonsense_Mutation	R	*
	21	34	17	7578389	7578389	G	A	NM_001126112	5	735	Missense_Mutation	R	C
	22	37	17	7578406	7578406	C	T	NM_001126112	5	718	Missense_Mutation	R	H
		33	17	7578406	7578406	C	T	NM_001126112	5	718	Missense_Mutation	R	H
	23	20	17	7578457	7578457	C	T	NM_001126112	5	667	Missense_Mutation	R	H
	24	48	17	7578461	7578461	C	A	NM_001126112	5	663	Missense_Mutation	V	F
	25	42	17	7578463	7578463	C	G	NM_001126112	5	661	Missense_Mutation	R	P
	26	3	17	7578475	7578476	-	G	NM_001126112	5	648_649	Frame_Shift_Ins	P	fs
	27	36	17	7578532	7578532	A	G	NM_001126112	5	592	Missense_Mutation	M	T
	28	44	17	7579377	7579381	GGTAG	-	NM_001126112	4	500_504	Frame_Shift_Del	T	fs
	29	51	17	7579528	7579528	C	T	NM_001126112	4	353	Nonsense_Mutation	W	*

FAT1	1	13	4	187521253	187521253	G	A	NM_005245	22	3968	Nonsense_Mutation	Q	*
	2	25	4	187521388	187521396	CATAGTTTC	-	NM_005245	22	3920_3922	In_Frame_Del	GNY	DEL
	3	39	4	187522472	187522472	G	A	NM_005245	21	3864	Missense_Mutation	T	M
	4	51	4	187522574	187522574	G	T	NM_005245	21	3830	Nonsense_Mutation	S	*
	5	29	4	187525001	187525002	-	A	NM_005245	19	3560	Frame_Shift_Ins	S	fs
	6	47	4	187530468	187530469	-	A	NM_005245	16	3358_3359	Frame_Shift_Ins	M	fs
	7	47	4	187530469	187530469	C	T	NM_005245	16	3358	Missense_Mutation	M	I
	8	37	4	187532783	187532783	G	A	NM_005245	14	3204	Nonsense_Mutation	Q	*
	9	3	4	187534324	187534324	C	A	NM_005245	13	3134	Silent	V	V
	10	16	4	187538220	187538220	G	C	NM_005245	11	3005	Nonsense_Mutation	S	*
	11	16	4	187540425	187540445	CAAAATGTTTATGATCATTGC	-	NM_005245	10	2432_2438	In_Frame_Del	GNDHKHF	DEL
	12	14	4	187540906	187540906	A	G	NM_005245	10	2278	Silent	D	D
	13	32	4	187541367	187541367	C	A	NM_005245	10	2125	Nonsense_Mutation	E	*
	14	30	4	187541586	187541586	C	A	NM_005245	10	2052	Nonsense_Mutation	E	*
	15	11	4	187541913	187541913	G	A	NM_005245	10	1943	Nonsense_Mutation	Q	*
	16	22	4	187542017	187542017	G	T	NM_005245	10	1908	Nonsense_Mutation	S	*
	17	19	4	187542357	187542357	G	A	NM_005245	10	1795	Nonsense_Mutation	R	*
	18	12	4	187542800	187542801	AT	-	NM_005245	10	1647	Frame_Shift_Del	M	fs
	19	52	4	187549326	187549327	-	CA	NM_005245	9	1597_1598	Frame_Shift_Ins	V	fs
	20	24	4	187557823	187557823	C	A	NM_005245	5	1296	Missense_Mutation	E	D
	21	24	4	187584594	187584594	C	A	NM_005245	3	1147	Nonsense_Mutation	E	*
	22	2	4	187627867	187627867	C	A	NM_005245	2	1039	Nonsense_Mutation	E	*
	23	4	4	187628023	187628023	G	A	NM_005245	2	987	Nonsense_Mutation	Q	*
	24	15	4	187628173	187628173	G	A	NM_005245	2	937	Nonsense_Mutation	R	*
	25	11	4	187628697	187628704	TCCTCATT	-	NM_005245	2	760_762	Frame_Shift_Del	N	fs
	26	1	4	187629388	187629388	G	T	NM_005245	2	532	Silent	R	R
	27	32	4	187629967	187629967	T	A	NM_005245	2	339	Nonsense_Mutation	K	*
	28	35	4	187630778	187630778	C	T	NM_005245	2	68	Nonsense_Mutation	W	*
CASP8	1	31	2	202131292	202131292	A	G	NM_033355	3	28	Missense_Mutation	D	G
	2	12	2	202131410	202131410	C	A	NM_033355	3	67	Missense_Mutation	F	L
	3	15	2	202137381	202137382	AG	-	NM_033355	5	144_145	Frame_Shift_Del	I	fs
	4	22	2	202137430	202137431	AA	-	NM_033355	5	161	Frame_Shift_Del	K	fs
	5	51	2	202137433	202137434	AG	-	NM_033355	5	162	Frame_Shift_Del	R	fs
	6	23	2	202139643	202139643	G	A	NM_033355	7	209	Silent	S	S
	7	13	2	202141586	202141586	C	T	NM_033355	8	233	Missense_Mutation	R	W
		33	2	202141586	202141586	C	T	NM_033355	8	233	Missense_Mutation	R	W
	8	35	2	202149683	202149683	C	-	NM_033355	9	316	Frame_Shift_Del	S	fs
	9	29	2	202149778	202149778	C	T	NM_033355	9	348	Missense_Mutation	L	F
	10	2	2	202149808	202149808	C	G	NM_033355	9	358	Missense_Mutation	Q	E
	11	24	2	202149917	202149917	C	T	NM_033355	9	394	Missense_Mutation	P	L
	12	11	2	202149961	202149962	TG	-	NM_033355	9	409	Frame_Shift_Del	C	fs
	13	32	2	202149973	202149973	C	T	NM_033355	9	413	Nonsense_Mutation	R	*
	14	4	2	202150039	202150039	C	T	NM_033355	9	435	Nonsense_Mutation	R	*
		8	2	202150039	202150039	C	T	NM_033355	9	435	Nonsense_Mutation	R	*

	15	38	2	202151270	202151270	C	T	NM_033355	10	465	Nonsense_Mutation	Q	*
		47	2	202151270	202151270	C	T	NM_033355	10	465	Nonsense_Mutation	Q	*
MLL4	1	47	19	36211899	36211899	C	-	NM_014727	3	550	Frame_Shift_Del	D	fs
	2	24	19	36215953	36215953	C	T	NM_014727	11	1165	Missense_Mutation	P	S
	3	5	19	36216421	36216421	C	G	NM_014727	13	1228	Missense_Mutation	F	L
	4	24	19	36218142	36218142	G	A	NM_014727	16	1363	Nonsense_Mutation	W	*
	5	52	19	36219731	36219731	G	A	NM_014727	21	1543	Nonsense_Mutation	W	*
	6	25	19	36220138	36220138	G	T	NM_014727	23	1620	Nonsense_Mutation	E	*
	7	24	19	36221004	36221004	A	G	NM_014727	24	1685	Missense_Mutation	H	R
	8	39	19	36221026	36221030	GGTGG	-	NM_014727	24	-	Splice_Site	K	Splice
	9	12	19	36221656	36221656	G	A	NM_014727	27	1775	Nonsense_Mutation	W	*
	10	16	19	36222886	36222886	C	T	NM_014727	28	1839	Nonsense_Mutation	Q	*
	11	16	19	36224521	36224521	C	-	NM_014727	30	2328	Frame_Shift_Del	T	fs
NOTCH1	1	29	9	139393624	139393624	C	A	NM_017617	32	2008	Nonsense_Mutation	E	*
	2	11	9	139396524	139396524	A	-	NM_017617	29	1801	Frame_Shift_Del	S	fs
	3	29	9	139399504	139399504	G	A	NM_017617	26	1547	Nonsense_Mutation	Q	*
	4	2	9	139399786	139399786	C	T	NM_017617	25	1521	Missense_Mutation	C	Y
	5	47	9	139399821	139399821	G	T	NM_017617	25	1509	Nonsense_Mutation	C	*
	6	32	9	139400154	139400154	G	T	NM_017617	25	1398	Nonsense_Mutation	Y	*
	7	35	9	139412297	139412297	C	T	NM_017617	8	450	Missense_Mutation	E	K
	8	46	9	139412744	139412744	C	G	NM_017617	7	367	Missense_Mutation	G	A
	9	16	9	139413147	139413147	C	T	NM_017617	6	332	Missense_Mutation	C	Y
HRAS	1	18	11	533875	533875	G	T	NM_005343	3	61	Missense_Mutation	Q	K
	2	41	11	534285	534285	C	T	NM_005343	2	13	Missense_Mutation	G	D
	3	12	11	534286	534286	C	G	NM_005343	2	13	Missense_Mutation	G	R
		15	11	534286	534286	C	G	NM_005343	2	13	Missense_Mutation	G	R
		39	11	534286	534286	C	G	NM_005343	2	13	Missense_Mutation	G	R
	4	22	11	534288	534288	C	T	NM_005343	2	12	Missense_Mutation	G	D
UNC13C	1	24	15	54305953	54305953	C	T	NM_001080534	1	285	Nonsense_Mutation	Q	*
	2	25	15	54307071	54307071	G	T	NM_001080534	1	657	Missense_Mutation	W	C
	3	22	15	54307337	54307337	C	G	NM_001080534	1	746	Missense_Mutation	S	C
	4	8	15	54592485	54592485	G	A	NM_001080534	11	1394	Nonsense_Mutation	W	*
	5	7	15	54792317	54792317	G	T	NM_001080534	19	1701	Nonsense_Mutation	E	*
	6	33	15	54793039	54793039	C	A	NM_001080534	20	1722	Missense_Mutation	Q	K
USP9X	1	2	X	40996157	40996157	A	G	NM_001039590	6	179	Missense_Mutation	N	S
	2	22	X	41010311	41010311	G	T	NM_001039590	13	-	Splice_Site	S	Splice
	3	25	X	41056704	41056704	C	T	NM_001039590	29	1441	Nonsense_Mutation	Q	*
	4	24	X	41069808	41069808	G	A	NM_001039590	33	1688	Missense_Mutation	E	K
	5	1	X	41077699	41077699	A	G	NM_001039590	37	2095	Missense_Mutation	N	S
	6	46	X	41088859	41088859	G	T	NM_001039590	43	2420	Nonsense_Mutation	E	*

ARID2	1	41	12	46230389	46230389	T	C	NM_152641	7	241	Silent	V	V
	2	29	12	46244016	46244016	C	T	NM_152641	15	704	Nonsense_Mutation	Q	*
	3	39	12	46244075	46244076	-	C	NM_152641	15	723_724	Frame_Shift_Ins	I	fs
	4	27	12	46245723	46245723	C	T	NM_152641	15	1273	Nonsense_Mutation	R	*
	5	24	12	46245843	46245843	C	T	NM_152641	15	1313	Nonsense_Mutation	Q	*
	6	52	12	46287315	46287315	C	T	NM_152641	19	1754	Nonsense_Mutation	R	*
TRPM3	1	30	9	73151542	73151542	T	-	NM_001007471	25	1509	Frame_Shift_Del	N	fs
	2	18	9	73152119	73152119	C	T	NM_001007471	25	1317	Missense_Mutation	A	T
	3	2	9	73164552	73164552	C	A	NM_001007471	24	1218	Missense_Mutation	D	Y
	4	3	9	73376523	73376525	CTT	-	NM_001007471	8	447	In_Frame_Del	K	DEL
	5	7	9	73461346	73461346	G	C	NM_001007471	4	208	Missense_Mutation	I	M

Supplementary Table S7. Mutations observed among the 50 gingivo-buccal oral squamous cell carcinoma patients at known mutational hotspots in *TP53

Codon no.	Amino acid change
175	R>H
	R>H
196	R>*
	R>*
213	R>*
245	G>S
248	R>Q
	R>W
	R>W
273	R>C
282	R>Frameshift
	R>Q
	R>W
	R>W
	R>W
	R>W

* Each mutation occurred in a separate patient

Supplementary Table S8. Frequencies (%) of recurrence in an independent validation sample set of 60 gingivo-buccal oral squamous cell carcinoma patients. The results are shown for the 10 significantly and recurrently mutated genes discovered and a comparative analysis.

Gene	TP53	FAT1	CASP8	MLL4	NOTCH1	USP9X	HRAS	UNC13C	ARID2	TRPM3
Frequencies in confirmation sample set (n=60)	71.7	18.3	45.0	5.0	21.7	11.7	11.7	6.7	8.3	5.0
Frequencies in discovery sample set (n=50)	62	40	34	16	16	12	12	12	10	10
p-value of test of equality of proportions between confirmation and discovery sets	0.28	0.01	0.24	0.11*	0.45	0.95	0.95	0.51*	0.76	0.46*

* *p*-value based on Fisher's exact test

Supplementary Table S9. Results of quantitative PCR assays to confirm deletions and amplifications. Fold change observed in sample from tumor tissue compared to blood.

Gene	Patient No.																			
	11	**12**	**14**	**16**	**17**	**18**	**20**	**21**	**22**	**26**	**28**	**33**	**36**	**37**	**38**	**42**	**43**	**44**	**49**	**50**
GSTT1 - Deletion	0.06												0.65		0.29					
CDKN2A-Deletion						0.42			0.33			0.33		0.41	0.66					
USP9X - Deletion		0.72	0.89	0.44			0.67											0.21		
CCND1 - Amplification			8.97	4.00	3.50			6.08		4.17		1.27				3.06	3.50		4.34	2.28
MECOM - Amplification	4.89		1.09							1.58										
MMP - Amplification											9.37	12.70					6.61	19.10		

Supplementary Table S10. Classification of non-synonymous mutations in significantly mutated genes associated with gingivo-buccal oral squamous cell carcinoma using bioinformatics tools, PROVEAN* and SIFT**

Gene	Status	Total no. of mutations	Mutations identified as 'deleterious' by PROVEAN		Mutations identified as 'damaging' by SIFT	
			No.	%	No.	%
TP53	Previously associated with HNSCC	84	81	96.43	83	98.81
CASP8		44	42	95.45	43	97.73
FAT1		39	36	92.31	36	92.31
NOTCH1		23	22	95.65	22	95.65
HRAS		13	13	100.00	13	100.00
MLL4	New associations detected in OSCC-GB	14	13	92.86	13	92.86
ARID2		11	10	90.91	11	100.00
USP9X		13	12	92.31	12	92.31
TRPM3		8	4	50.00	7	87.50
UNC13C		11	5	45.45	10	90.91
Total		260	238	91.54	250	96.15

*http://provean.jcvi.org/index.php

** http://sift.jcvi.org/

Supplementary Table S11: Inferences on nature of genomic alteration based on genotypes in blood and tumor*

Blood Genotype	Tumor Genotype		
	RR	**RV**	**VV**
RR	**No mutation (G0S0)**	**Somatic mutation (G0S1)**	**Somatic mutation (G0S2)**
RV	**LOH (G1S1)**	**Germline mutation (G1S0)**	**LOH (G1S1)**
VV**	**No inference (G2S2)**	**Somatic mutation (G2S1)**	**Germline mutation / No mutation (G2S0)**

* R denotes 'Reference' allele; that is, allele present in the Human Reference Sequence, hg19; V denotes 'Variant' allele. GnSm denotes *n* Germline mutations and *m* Somatic mutations.
** In some instances it was observed that the genotype in the blood DNA was VV, which is unlikely. In all such instances, it was found that there was an 'allelic swap'. The allele present in the human genome reference sequence was not the more common allele at that site in most populations (as ascertained from the HapMap3 and the 1000Genomes databases). See Supplementary Methods.

Supplementary Table S12: Primers for representative somatic CNVs located in recurrently mutated genes. Primers were designed using Primer Express software. See Supplementary Methods for details.

Sl. No.	Oligo ID	Sequence (5'-3')	Gene Name	Nature of Mutation
1	CNV-001	ACCCACCTCAACCGGAAAC	ADAM3A	Homozygous Deletion
2	CNV-002	TTGGACCTTCTGCTGCCTCTT	ADAM3A	Homozygous Deletion
3	CNV-003	AGCTGTTTTCCAAAGCGGTTAT	GSTT1	Homozygous Deletion
4	CNV-004	GTTGAGGATGTGTGAAGAAATTGAA	GSTT1	Homozygous Deletion
5	CNV-005	AAGATGGTGCCTCCGACTGT	LCE3B	Homozygous Deletion
6	CNV-006	CTCTAAAGTCGCTTGTCTCAGCAT	LCE3B	Homozygous Deletion
7	CNV-007	AGACGAGGAAAGAGGCTCCTAGA	SIGLEC14	Homozygous Deletion
8	CNV-008	CCAAAAGATCTGAGCCTGCTTCT	SIGLEC14	Homozygous Deletion
9	CNV-009	GAAAGTAAACTTGAGCTGGTACAATGG	SPINK14	Homozygous Deletion
10	CNV-010	TAAAATTGGTGCCGCAGATG	SPINK14	Homozygous Deletion
11	CNV-011	TCAAATCCCAGTTCTCTCTCACTTAC	CCND1	Amplification
12	CNV-012	GAGGAATTGGCACAGAGAGGTT	CCND1	Amplification
13	CNV-013	ACCACTGCCCTCAGCTCCTA	CDKN2A	Amplification
14	CNV-014	ACGTCTCCACAGTGAAACCAACT	CDKN2A	Amplification
15	CNV-015	TGCAACAACAGGCATGCATT	FADD	Amplification
16	CNV-016	GGCGTGGGCCAAATCA	FADD	Amplification
17	CNV-017	AGCTGCGGCTTTAGGATGAA	HRAS	Amplification
18	CNV-018	GCAGACGGGAAGCTCTACGA	HRAS	Amplification
19	CNV-019	TTTGATAAGGATTGGGATGAATCTG	IKBKG	Amplification
20	CNV-020	CATGGATTGGAAGATTTAATGTTGTT	IKBKG	Amplification
21	CNV-021	TATCATCTAGGATAGTTAACGCCATCTC	MECOM	Amplification
22	CNV-022	CGCAACCAGTTGAGCAGGTA	MECOM	Amplification
23	CNV-023	CAGCCATGGCCAATTCGTA	MED12	Amplification
24	CNV-024	CGTCATGATCTCTGGCTTCGT	MED12	Amplification
25	CNV-025	TTATTCCATGCAAGGTTGGATTC	MMP	Amplification
26	CNV-026	GCTGCTGGATTCCTATCCTCAT	MMP	Amplification
27	CNV-027	AACCCAACATCAGTGCAACCT	PRDM9	Amplification
28	CNV-028	GCTGTGGTCGATGGCATTATT	PRDM9	Amplification
29	CNV-029	ACAGCTGGTACAGATGAGGCAAT	SEMA5A	Amplification
30	CNV-030	ACGTGGTGTAGGACCTGACTTGT	SEMA5A	Amplification
31	CNV-031	CAAACGTTAGGCTAGACGAGCAA	UNC13C	Amplification
32	CNV-032	GGAGTGCAACCAACATGGTATG	UNC13C	Amplification
33	CNV-033	TCATGTTGTGTCACTCTAAATTGTACCTT	CCNC	Hemizygous Deletion
34	CNV-034	CCAGGTAGGCCCAGGGTAGT	CCNC	Hemizygous Deletion
35	CNV-035	TGAAATTAGTCCCTCAGAATAAACCA	CDKN2A	Hemizygous Deletion
36	CNV-036	ATTGCCTCTGAGCTTAGATTTTGAC	CDKN2A	Hemizygous Deletion
37	CNV-037	ACTTGGTCCAGCCACATGGT	CSMD1	Hemizygous Deletion
38	CNV-038	CATGCTGTCCACCCACATTG	CSMD1	Hemizygous Deletion

39	CNV-039	TCTAAGCAACCCTGCCACAAG	FHIT	Hemizygous Deletion
40	CNV-040	TGGCACACAGAAATTCCTTAACA	FHIT	Hemizygous Deletion
41	CNV-041	GGGCAGAGGAGGCATGAA	PTPRD	Hemizygous Deletion
42	CNV-042	TAGCCCAGGGTCCATTCCTA	PTPRD	Hemizygous Deletion
43	CNV-043	TTAAGAGCGTCTGCCTCACAAGT	USP9X	Hemizygous Deletion
44	CNV-044	CTCTATTGCACCAAGAGTGGCTAA	USP9X	Hemizygous Deletion

Supplementary Methods

Data Generated Using HiSeq-2000

For HiSeq-2000 data, sequence data were demultiplexed and FASTQ files generated by CASAVA v1.8.2 (Illumina). The sequence data generated per run were evaluated by FASTQC Ver. 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), using default parameters. Sequence reads were mapped to 1000 Genomes phase 2 reference sequence (hg19 with decoy sequence; ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/) using BWA[64] v0.6.1 with –q 15, -l 32 as arguments. Following alignment, the generated .sai files were used to create .sam (sequence alignment map) files by converting suffix array coordinates to chromosomal coordinates. Reads which mapped to multiple regions of the human genome were identified and removed. Reads which mapped to decoy segments of the reference genome were identified and removed. Finally, all unmapped reads were also removed from the alignment files. Duplicate reads were filtered out using the PICARD 1.60 (http://picard.sourceforge.net) software. Base quality recalibration, local realignment around indels using known SNPs was performed with GATK[67] v. 1.4. Alignment SAM files were converted to BAM files using Samtools[66] 0.1.12a. BAM files were converted to pileup format by SAMtools[66] (http://samtools.sourceforge.net/) for variant calling.

Somatic base-pair substitutions were called using Base-by-Base (BbB) caller developed by us. The BbB caller essentially identifies, for each base position with quality value $\geq$15, the total number of reads and the number of reads containing the (a) reference allele and (b) variant allele. These data are then statistically analysed [described in detail below] to draw inferences on

the nature of genomic alteration induced by the mutation. Indels were identified using GATK and BbB caller. The following criteria were used to select variants for further analysis: (1) The variant site should have a minimum depth of 10X in blood DNA; (2) The variant base should have a base quality value $\geq$15; (3) For germline and LOH variants, at least 20% of the reads mapping to such sites in the blood DNA should harbor the variant; (4) For somatic variants, at least 10% of the reads mapping to such sites in the tumor DNA should harbor the variant; (5) The variant site should have at least one read harboring the variant in each direction (to avoid strand bias). In every patient, clustered somatic mutations (two mutations in a 100bp window) were removed, as these mutations were possibly generated as part of read mis-alignment error. Somatic variants which are catalogued in the 1000 Genomes, ESP genome database (https://esp.gs.washington.edu/) with an allele frequency >0.01 were removed from further analysis. A mutation that was identified as a somatic mutation in a patient was removed if it was also identified as a germline mutation in more than one of the 50 patients included in this study. Somatic mutations thus identified were annotated with respect to gene structure, UCSC database information, COSMIC database, effect of mutation on the protein product using Oncotator module (http://www.broadinstitute.org/ cancer/cga/oncotator). Variants present in regions outside of the exons and UTRs were removed. All indels and recurrent substitutions were manually visualized and verified in IGV[70] v2.1.25.

Data Generated Using GS-FLX

For data generated using GS-FLX Genome Sequencers, sequence reads in FASTQ format were obtained from SFF files by sffinfo (GS-FLX) and a Linux script developed by us. The sequence data generated per run were evaluated by the same method described in the previous section. The

FASTQ files were mapped to human reference sequence hg19 (GRCh37) by SSAHA2 v2.5.5 . PICARD v1.60 (http://picard.sourceforge.net) was used to remove PCR duplicates and combine SAM files from multiple runs. Pileup files were generated by SAMtools[70] . The pileup files of tumor and blood DNA from each patient were used to identify putative variant sites using the BbB variant caller followed by statistical analysis [Methods section] to draw inferences on the nature of genomic alteration induced by the mutation. Variant positions were visualized in IGV[71] v2.1.25.

Verification of Somatic Mutations using Custom Ampliseq and Ion Torrent Platform

All somatic mutations – 109 mutations – belonging to genes which were found to be recurrently mutated (in at least 10% of the tumor samples), discovered using HiSeq-2000 and its analysis pipeline, Custom AmpliSeq Panel was used for amplification of these mutations and the products were sequenced using Ion Torrent PGM. These mutations could not be verified using GS-FLX, since these were located outside of the regions captured by the Roche-NimbleGen Exome Capture Probe Library. Highly multiplexed PCR assays were designed from about 100 bp of genomic sequences spanning each of these variants using the Ion Ampliseq Designer (Life Technologies). Amplification and library preparation for each variant was performed with about 10 ng of genomic DNA of each tumor sample and Ion Ampliseq Library kit v2 (Life Technologies) with barcoded adapters. The thermal cycling conditions used for PCR were 99 ºC for 2 minutes, 18 cycles of 99 ºC for 15 seconds and 60 ºC for 4 minutes, followed by hold at 10ºC. These libraries were pooled, purified by Ampure XP (Beckman Coulter) and the quality assessment and quantitation performed by High Sensitivity DNA kit (Agilent) and TaqMan assay based Ion Library Quantification kit in ABI 7900HT Real Time PCR system (Life Technologies)

respectively. The pooled DNA libraries were sequenced on Ion Torrent PGM (Life Technologies) using Ion 318 Chip and Ion PGM 200 Sequencing kit (Life Technologies) at an average depth of about 200X. Sequence reads obtained were mapped to HG19 sequence by TMAP (Homer, N., Cawley, S., and Merriman, B. Ion Torrent Community - TMAP: the Torrent Mapping Alignment Program for Ion Torrent. Available: http://ioncommunity.iontorrent.com/). Variant calling was performed with software tools developed in-house. Out of these 109 mutations, 107 (98%) could be verified from the sequence data obtained.

Sanger Sequencing of *TP53*:

Sanger sequencing of *TP53* was carried out for both blood and tumor DNA of each patient. Primers were synthesized to amplify exons 1-11 of *TP53* as described by Liu and Bodmer (2006)[72]. The amplified products were treated with Exonuclease I and Shrimp Alkaline phosphatase - Exo SAP IT (Affymetrix USB) at 37 ºC to remove the excess primers and dNTPs. The cleaned amplicon was used as a template for sequencing, at an appropriate concentration based on the size of the amplicon. Sequencing reactions were set up with BigDye® Terminator v3.1 Cycle Sequencing Kit as per the instructions of the manufacturer (Life Technologies). Post sequencing reaction clean-up was carried out using BigDyeX Terminator kit (Life Technologies). The samples for sequencing were electrophoresed through POP7 polymer on an automated DNA sequencer, Genetic Analyzer 3500 (Applied Biosystems).The data obtained were analyzed manually or with SeqScape software. Variants in the *TP53* sequence were confirmed by bidirectional sequencing.

Cross-Identification of Variants between the Two Platforms and Verification

The variants identified by the Illumina and the Roche pipelines were compared. Those variants that were identified by both the pipelines were categorized as 'confirmed variants'. Variant sites which were identified from Illumina-TruSeq capture region (38Mb: gene CDS and ncRNA CDS) and fell outside of the Roche-Nimblegen capture region (26Mb: CDS only) were retained for downstream analysis, provided the sites passed all of the above QC steps, and for further verification using Ion Torrent. Variant sites were classified into the following mutually exclusive and collectively exhaustive categories: Somatic, Somatic + LOH, LOH, Somatic+Germline, Germline+LOH and Somatic+LOH+Germline. It may be mentioned that (i) sites heterozygous for the variant allele in blood DNA were removed if homozygosity for the reference allele was noted in tumour DNA, and (ii) only those variants that were found to be somatic and not as germline in more than one patient were considered for further statistical analysis. Variant sites of acceptable quality that were present in HiSeq 2000 data, but were outside of the Nimblegen capture region and hence not present in GS-FLX data, and found to be recurrently mutated in $\geq$10% of the patients, were verified using the Ion Torrent platform.

Genome wide Scan

200 ng of each tumor and blood DNA sample were genotyped for 1.14 million SNP markers using the Omni-Quad chips and scanned on iScan (Illumina). Illumina provided GC content-polynomial fit script v1.4.6 was used to adjust LRR ratio for genomic waves associated with GC content[73]. For each normal - tumor pair, log-R ratio and B allele frequency[74] for individual SNPs were exported individually in separate files.

Analysis for CNV Detection

Regions of gains and losses were identified using ASCAT 2.0 (Alelle-Specific Copy number Analysis of Tumors) using exported data from Genome Studio software. ASCAT (http://heim.ifi.uio.no/bioinf/Projects/ASCAT/) uses SNP array data to predict regions of amplification, deletion and LOH while taking into account normal cell admixture and tumor aneuploidy from cancer samples. ASCAT segments and smoothes the normal-tumor matched genome-wide SNP array intensities and B-allele frequency for copy number prediction. Windows of SNPs with mean LogR > 0.15 was used to detect amplification segment and mean LogR < - 0.2 was used to detect deletion segments. Amplification and deletion CNVs shared across multiple samples were identified using in-house programs. Hi-confidence CNV regions were annotated using in-house programs with NCBI refgene transcript database (hg19).

Verification of Somatic CNVs by Real Time PCR

Following Real Time PCR, primers were designed for representative somatic CNVs which were found to be located in recurrently mutated genes using Primer Express software (Life Technologies). SYBR Green assays were set up for each blood and tumor DNA samples in ABI 7900HT Real Time PCR system using Power SYBR Green Master Mix (Life Technologies) with *GAPDH* gene assay (GAPDH-For: ATGCTGAGTGTACAAGCGTTTTCT, GAPDH-Rev: CACTATGCCACCCCAGGAAT) as endogenous control. The thermal cycling conditions used for all assays were 95^0C for 10 minutes, 40 cycles at 95^0C for 15 seconds and 60^0C for 1 minute, followed by dissociation curve analysis. Fold change of somatic CNVs in tumor with respect to blood was calculated by a published relative quantitation method [75].

Supplementary References

70. Ning, Z., Cox, A.J. & Mullikin, J.C. SSAHA: a fast search method for large DNA databases. *Genome research* **11**, 1725-1729 (2001).

71. Robinson, J.T. *et al*. Integrative genomics viewer. *Nat Biotechnol*. **29**, 24-26 (2011).

72. Liu, Y. & Bodmer, W.F. Analysis of P53 mutations and their expression in 56 colorectal cancer cell lines. *Proc. Natl. Acad. Sci. USA.* **103**, 976-981 (2006).

73. Diskin, S.J. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res*. **36** (19):e126 (2008).

74. Peiffer, D.A. *et al.* High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res*. **16** (9), 1136-1148. (2006)

75. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* **25**, 402-408 (2001).