Targeted Next Generation Sequencing of *CIC-DUX4* Soft Tissue Sarcomas Demonstrates Low Mutational Burden and Recurrent Chromosome 1p Loss

Lorena Lazo de la Vega B.S., Daniel H. Hovelson M.S., Andi K. Cani M.S., Chia-Jen Liu M.S., Jonathan B. McHugh M.D., David R. Lucas M.D., Dafydd G. Thomas M.D. PhD., Rajiv M. Patel* M.D., Scott A. Tomlins* M.D. PhD.

SUPPLEMENTARY INFORMATION

Supplementary Figure Legends

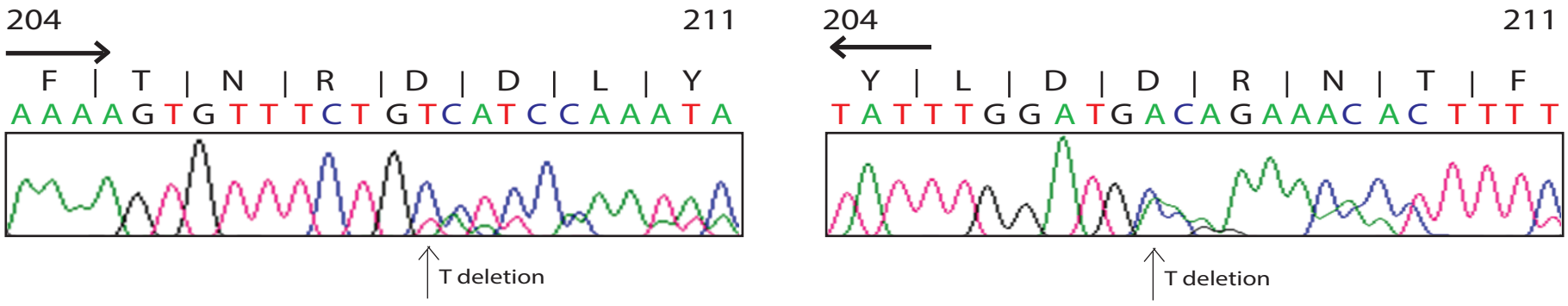
Fig. S1. Bidirectional Sanger Sequencing confirmation of prioritized mutations.

Sanger sequencing was performed for all prioritized mutations with variant allele frequencies >15% across all profiled cases. Sanger traces of cases with their respective mutation are shown in both directions with the indicated nucleotide and amino acid changes noted. These mutations were not observed in any other case (data not shown).

Figure S1

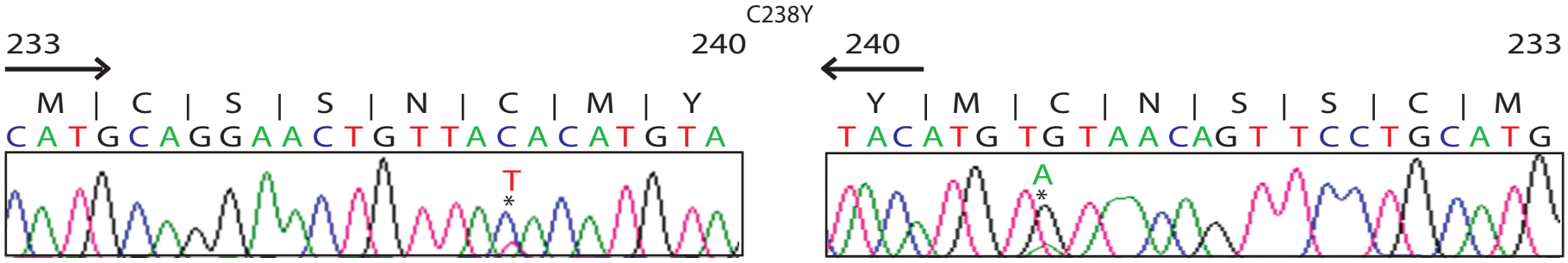
Case 1 *TP53*

D208fs



Case 1 *TP53*

C238Y



Case 7B *ARID1A*

R693X

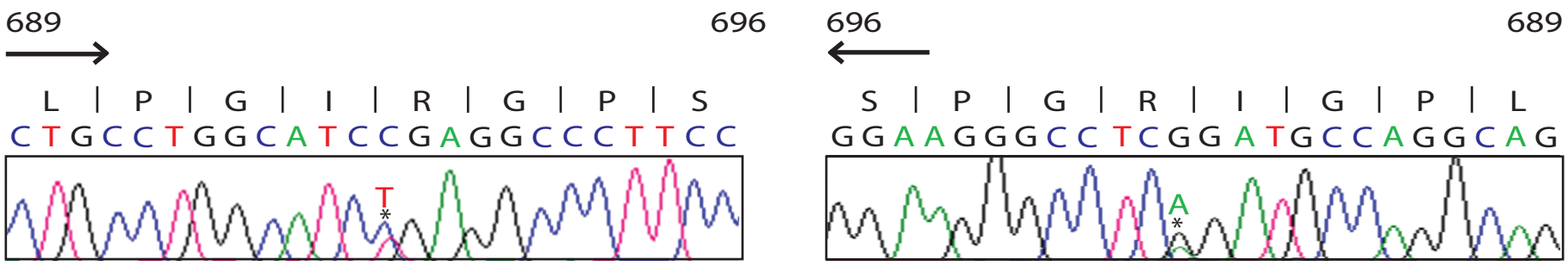


Table S1. Sequencing statistics and for profiled CIC-DUX4 samples

Case	Mapped Reads	Mean Depth	Uniformity	% reads on target	Total Called variants[1]	Variants passing filtering[2]	Prioritized variants[3]
1	9,465,769	566	95%	97%	1496	5	2;
2	10,854,865	670	86%	99%	1132	2	0
3	8,924,168	537	85%	98%	2434	3	0
4	8,846,125	548	87%	99%	1554	2	0
5A	9,586,657	592	90%	99%	1199	0	0
5B	8,363,000	514	88%	98%	1192	0	0
6A	14,130,758	883	88%	99%	1137	1	1;
6B	13,322,079	828	89%	99%	1173	0	0
7A	9,728,867	591	95%	98%	1177	2	0
7B	10,256,404	636	95%	99%	1134	2	1;
8	9,923,117	614	94%	99%	1189	4	0
Average	**10,309,255**	**634**	**90%**	**98%**	**1,347**	**2**	**<1**

Sequencing statisitics for 11 CIC-DUX4 sarcomas (from 8 cases) profiled by next generation sequencing (NGS) using the Ion Ampliseq Comprehensive Cancer Panel (CCP), which targets 1,688,650 bases in 15,992 amplicons representing 409 cancer genes. Given are the number of mapped reads per sample, the mean read coverage depth over targeted bases, uniformity of mapped reads and the % reads on target, as well as the average called variants prior to any filtering. [1]Variants called by automated low stringency variant calling. [2]Variants passing filtering of technical artifacts, poorly supported variants, germline SNPs and synonymous/non-coding variants. [3]Variants passing filtering described in [2] and prioritized as likely driving oncogenic or tumor suppressive mutations.

Table S2. Somatic nonsynonymous mutations in CIC-DUX4 sarcomas identified by next generation sequencing

Case	Location	Gene	REF	ALT	AA Change	Var. Allele Freq.	FAO	FDP	Transcript	Nuc change	ESP6500	KG_FEB2012	EXAC	FSAF	FSAR	HRUN
1	chr12:49444178	*KMT2D*	A	T	p.S1065T	0.3905	164	420	NM_003482	c.T3193A	NA	NA	0.00001653	51	113	1
	chr17:78320326	*RNF213*	C	A	p.L2731I	0.9515	255	268	NM_001256071	c.C8191A	NA	NA	0	152	103	1
	chr20:39708747	*TOP1*	C	T	p.P120S	0.4758	265	557	NM_003286	c.C358T	NA	NA	0	74	191	2
	chr17:7577568	***TP53***	**C**	**T**	**p.C238Y**	**0.3157**	**365**	**1156**	**NM_00546**	**c.G713A**	**NA**	**NA**	**0.00002471**	**209**	**156**	**1**
	chr17:7578226	***TP53***	**GT**	**G**	**p.D208fs**	**0.6758**	**1338**	**1980**	**NM_000546**	**c.623delA**	**NA**	**NA**	**0**	**682**	**656**	**1**
2	chr7:151859963	*KMT2C*	T	G	p.S3567R	0.54	96	178	NM_170606	c.A10699C	NA	NA	0	37	59	1
	chr12:49444845	*KMT2D*	C	G	p.C874S	0.35	123	354	NM_003482	c.G2621C	NA	NA	0	70	53	1
3	chr6:152712563	*SYNE1*	C	T	p.S2625N	0.48	529	1104	NM_033071	c.G7874A	NA	NA	0	307	222	1
	chr6:160500777	*IGF2R*	G	A	p.A1882T	0.11	12	111	NM_000876	c.G5644A	NA	NA	0	5	7	1
	chr9:139397766	*NOTCH1*	C	T	p.E1679K	0.14	24	169	NM_017617	c.G5035A	NA	NA	0	2	22	2
4	chr2:141459780	*LRP1B*	C	A	p.V2078L	0.41	37	90	NM_018557	c.G6232T	NA	NA	0	26	11	2
	chr8:41834733	*KAT6A*	G	A	p.R386W	0.48	967	1995	NM_006766	c.C1156T	NA	NA	0	417	550	1
6A	**chr3:41266163**	***CTNNB1***	**G**	**A**	**p.E54K**	**0.11**	**28**	**245**	**NM_001904**	**c.G160A**	**NA**	**NA**	**0**	**10**	**18**	**2**
7A	chr12:49431186	*KMT2D*	G	C	p.A3318G	0.5132	214	417	NM_003482	c.C9953G	NA	NA	0	106	108	1
7B	chr12:49431186	*KMT2D*	G	C	p.A3318G	0.5247	170	324	NM_003482	c.C9953G	NA	NA	0	82	88	1
	chr1:27087503	***ARID1A***	**C**	**T**	**p.R693X**	**0.3765**	**157**	**417**	**NM_006015**	**c.C2077T**	**NA**	**NA**	**0**	**84**	**73**	**2**
8	chr6:152201817	*ESR1*	C	G	p.T224S	0.1522	7	46	NM_000125	c.C671G	NA	NA	0	4	3	2
	chr10:76788660	*KAT6B*	GGAA	G	p.1360_1360del	0.1441	67	465	NM_012330	c.4078_4080del	NA	NA	0	15	52	2
	chr1:32741944	*LCK*	C	T	p.S213L	0.5197	158	304	NM_005356	c.C638T	NA	NA	0	87	71	1
	chr9:139396829	*NOTCH1*	C	T	p.R1760H	0.5262	241	458	NM_017617	c.G5279A	NA	NA	0	113	128	1

All high confidence somatic non-synonymous variants (see Methods) identified across 11 FFPE CIC-DUX4 sarcoma samples are shown. The location (hg19), reference (Ref.) and variant (Var.) alleles, amino acid (AA) change, variant allele frequency (Var Allele Freq. = FAO/FDP) are shown for each sample. Read level information (FAO= flow corrected variant allele containing reads, FDP=flow corrected read depth, FAO/FDP), function and transcript/nucleotide (Nuc) change are shown for each high confidence, somatic non-synonymous variant. Allele frequencies in normal populations (ESP6500, 1000 genomes project, and ExAC) are given. Variant allele support on both forward and reverse reads is given (FSAF=forward strand flow corrected variant allele calling reads; FSAR=reverse strand flow corrected variant allele calling reads). HRUN indicates the number of consecutive bases at the indicated variant allele position. Prioritized oncogenic/tumor suppressive mutations are bolded.