SI Appendix for

Mutational landscape of uterine and ovarian carcinosarcomas implicates histone genes in epithelial-mesenchymal transition

Siming Zhao[a,b], Stefania Bellone[c], Salvatore Lopez[c], Durga Thakral[a,b], Carlton Schwab[c], Diana P. English[c], Jonathan Black[c], Emiliano Cocco[c], Jungmin Choi[a,b], Luca Zammataro[c], Federica Predolini[c], Elena Bonazzoli[c], Mark Bi[a,b], Natalia Buza[d], Pei Hui[d], Serena Wong[d], Maysa Abu-Khalaf[e], Antonella Ravaggi[f], Eliana Bignotti[f], Elisabetta Bandiera[f], Chiara Romani[f], Paola Todeschini[f], Renata Tassi[f], Laura Zanotti[f], Franco Odicino[f], Sergio Pecorelli[f], Carla Donzelli[g], Laura Ardighieri[g], Fabio Facchetti[g], Marcella Falchetti[g], Dan-Arin Silasi[c], Elena Ratner[c], Masoud Azodi[c], Peter E. Schwartz[c], Shrikant Mane[a,b], Roberto Angioli[h], Corrado Terranova[h], Charles Matthew Quick[i], Babak Edraki[j], Kaya Bilgüvar[a,b], Moses Lee[k], Murim Choi[k], Amy L. Stiegler[l], Titus J. Boggon[l], Joseph Schlessinger[l], Richard P. Lifton[a,b,m,1], and Alessandro D. Santin[c]

[a]Department of Genetics, Yale University School of Medicine, New Haven, CT 06510; [b]Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, CT 06510; [c]Department of Obstetrics, Gynecology & Reproductive Sciences, Yale University School of Medicine, New Haven, CT 06510; [d]Department of Pathology, Yale University School of Medicine, New Haven, CT 06510; [e]Internal Medicine & Oncology, Yale University School of Medicine, New Haven, CT 06510; [f]"Angelo Nocivelli" Institute of Molecular Medicine, Department of Obstetrics & Gynecology, University of Brescia, 25100 Brescia, Italy; [g]Department of Pathology, University of Brescia, Brescia 25100, Italy; [h]Universita' Campus Bio-Medico di Roma, 00128 Rome, Italy, [i]Department of Pathology, University of Arkansas for Medical Sciences, Little Rock, AR 72205; [j]John Muir Health Clinical Research Center, Concord, CA 94598; [k]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 110-799, Korea; and [l]Department of Pharmacology, Yale University School of Medicine, New Haven, CT 06510; [m]Laboratory of Human Genetics and Genomics, The Rockefeller University, New York, NY 10065

[1]To whom correspondence should be addressed: richard.lifton@yale.edu

Materials and Methods

Patients and specimens

The study protocol was approved by the Yale Human Investigation Committee. Tumor DNA was extracted from biospecimens (n=68) collected from newly diagnosed patients with uterine (n=44) and ovarian (24) carcinosarcomas (CSs) who underwent surgical staging for the treatment for their disease at Yale University, New Haven, CT, the University of Arkansas for Medical Sciences, Little Rock, AR or the University of Brescia School of Medicine, Brescia, Italy. All cases were staged according to the 1988 FIGO staging system and each case was reviewed by board certified pathologists to confirm that the frozen section was histologically consistent with a uterine or ovarian CS. Main criteria for selection were: 1) the diagnosis of homologous or heterologous CS of ovarian or uterine origin confirmed by the review of a minimum of 2 Board certified surgical pathologists with specific expertise in Gynecologic oncology, 2) the availability for light microscopic evaluation of a hematoxylin and eosin stained section of each frozen tumor specimen submitted to sequencing for assessment of percent tumor nuclei and percent necrosis in addition to other pathology annotations and 3) the availability of sufficient fresh frozen tumor tissue for DNA extraction for NGS. Genomic DNA was prepared from venous blood, primary fibroblast cultures or frozen myometrium by standard procedures. Light microscopic evaluation was performed on a hematoxylin and eosin stained section of each frozen tumor specimen submitted to sequencing for assessment of percent tumor nuclei and percent necrosis in addition to other pathology annotations. We confirmed that each section contains a high purity of tumor (>60%) and minimal necrosis and used those sections for DNA extraction. Primary CS cell lines with tumor purity above 90% and less than 3 weeks of culture in vitro were used for sequencing. We also performed whole-exome multi-region sequencing on DNA extracted from four to five spatially separated areas obtained through macro-dissection from 6 randomly selected formalin fixed-paraffin embedded (FFPE) CS specimens harboring well-defined areas of sarcomatous and epithelial neoplastic tissue within the same tumor. All samples were identified and macro-dissected by a Board Certified surgical pathologist to determine tumor characteristics and cellularity and to remove contaminating normal tissue from the selected sarcomatous and epithelial matched areas used to perform multi-region whole exome sequencing.

Whole exome sequencing

For fresh frozen and cell line samples, genomic DNA was isolated by Allprep DNA kit (Qiagen # 80204). For FFPE samples, DNA was extracted by BiOstic® FFPE Tissue DNA Isolation Kit (MO BIO Laboratories #12250-50) with a modified protocol.

Genomic DNA was captured on the NimbleGen 2.1M human exome array and subjected to 74 base paired-end reads on the Illumina HiSeq 2000 instrument as described[1]. Sequence reads were mapped to the reference genome (hg19) using the ELAND program. Reads outside the targeted sequences were discarded and statistics on coverage were collected from the remaining reads using in-house Perl scripts.

Somatic point mutation and indel calling

For matched normal-tumor pairs, somatic point mutations were called by MuTect. The output from MuTect was further filtered to remove false positives caused by low base/read quality or mismapping. We also required minimum number of reads with non-reference alleles based on total coverage of the position and sequencing error of the platform. Somatic indels were called by an in-house pipeline detecting regions with multiple nearby loci showing differences of reference and non-reference reads coverage between tumor and normal samples. All indels have been manually curated. For unmatched tumors, SAMtools was used to call variant bases appended with quality scores. Among these, variants with frequency more than 2 $x10^{-5}$ in ExAC (Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org)) were excluded and the rest were considered as potential somatic variants. Identified variants were annotated based on novelty, impact on the encoded protein, conservation, and expression using an automated pipeline. For uterine carcinomas from TCGA, we only used samples without *POLE* mutations or MSI phenotypes to ensure similar mutation rates, which ended up with 46 uterine serous carcinomas and 104 endometrioid carcinomas. We processed 46 uterine serous carcinomas, 30 from Zhao et al., 2013[2] and 16 following the same protocol (see Table S7 for sequencing quality metrics). For ovarian serous carcinomas, all 316 samples were extracted from the TCGA project[3]. In the multi-region WES part, to ensure high specificity after combining mutations from multiple samples in the same tumor, we have only considered mutations with B-allele frequency (BAF) >10% to account for the slightly higher false positive rate at the low BAF spectrum. For each of these mutations, we extracted read coverage for reference and alternate allele information

across all samples from the same tumor. The presence of the mutation in one sample was defined by at least 2 independent reads with alternate allele for positions covered less than 200x and at least 3 reads for positions covered more than 200x. These settings ensured low false positive rate under the current sequencing error rate of the platform.

Somatic copy number mutation calling

The ratio of normalized coverage depth between tumor and normal samples was calculated for each exome capture probe. The distribution of coverage depth ratio was evaluated for each individual tumor-normal pairs and only samples showing strong clustering at discrete values were included in CNV analysis. This quality control procedure was detailed in Zhao et al., 2013[2] and we ended up with 29 CSs that passed the filter out of 41 matched tumors. The genome was then de-noised and segmented by circular binary segmentation (DNAcopy package from R), Copy number was assigned to each segment based on both coverage depth ratio and deviation of the B allele frequency from the genome-wide average using in-house scripts.

For a fair comparison, samples from previous USC cohort[2] have been re-analyzed using the same method (same set of parameters). GISTIC 2.0 was used to infer significantly altered regions. The genome-wide significance cut off for *q* value was set to 0.25 for both broad and focal CNV events. Cancer genes listed in either Pan-Cancer or Vogelstein cancer gene lists were examined within each of the significant intervals. LOH calling and purity estimation were performed as previously described[2].

Construction of phylogenetic trees

Mutations identified in at least one sample from the tumor with B-allele frequency more than 10% were used to construct phylogenetic trees. B-allele frequencies were adjusted by sample purity first and then used for hierarchical clustering (hclust function from R). This method not only took into account the presence or absence of the mutation in one sample but also the prevalence of the mutation when inferring the relationship between macro-dissected samples.

Real-time reverse transcription-PCR (qRT-PCR)

RNA isolation from carcinosarcomas cell lines and fresh frozen tissues was performed using AllPrep DNA/RNA/Protein Mini Kit (Qiagen, Valencia, CA, USA), according to the manufacturer's instructions. Quantitative PCR was carried out with a 7500 RealTime PCR System using the manufacturer's recommended protocol (Applied Biosystems, Foster City, CA, USA) to evaluate the expression of PIK3CA, PPP2R1A, FBXW7, KRAS, CHD4, TP53, MYC, CCNE1, ERBB2, BCOR, HIST1H2AB, HIST1H2AC, HIST1H2AA, HIST1H2AE, HIST1H2BK, HIST1H2BD, HIST1H2BC. The primers and probes were obtained from Applied Biosystems (i.e., FBXW7, Assay ID: Hs00217794_m1; PIK3CA, Assay ID: Hs00180679_m1; PPP2R1A, Assay ID: Hs00204426_m1; KRAS, Assay ID: Hs00270666_m1; CHD4, Assay ID: Hs00172349_m1; TP53, Assay ID: Hs01034249_m1; MYC, Assay ID: Hs00153408_m1; CCNE1, Assay ID: Hs01026536_m1; ERBB2, Assay ID: Hs00170433_m1; BCOR, Assay ID: Hs00372378_m1; HIST1H2AB, Assay ID: Hs01001083_s1; HIST1H2AC, Assay ID: Hs00185909_m1; HIST1H2AA, Assay ID: Hs00601656_s1; HIST1H2AE, Assay ID: Hs01027931_s1; HIST1H2BK, Assay ID: Hs00955067_g1; HIST1H2BD, Assay ID: Hs00371070_m1; HIST1H2BC, Assay ID: Hs00979148_g1). The comparative threshold cycle method was used to determine gene expression in each sample, relative to the value observed in non-malignant endometrial epithelial cell samples collected from similar age women, using glyceraldehyde-3-phosphate dehydrogenase (Assay ID Hs99999905_m1) RNA as an internal control.

RNA isolation from stably transfected cell lines expressing histone variants was performed using AllPrep DNA/RNA/Protein Mini Kit (Qiagen, Valencia, CA, USA), according to the manufacturer's instructions. Quantitative PCR was carried out with a 7500 RealTime PCR System using the manufacturer's recommended protocol (Applied Biosystems, Foster City, CA, USA) to evaluate the expression of CLDN3, CLDN4, CDH1, CDH2 and SNAI2/Slug. The primers and probes were obtained from Applied Biosystems (i.e., CLDN3, Assay ID: Hs00265816_s1; CLDN4, Assay ID: Hs00533616_s1; CDH1, Assay ID: Hs01023894_m1; CDH2, Assay ID: Hs00983056_m1; SNAI2/Slug, Assay ID:Hs00950344_m1). The comparative threshold cycle method was used to determine gene expression in each sample, relative to the value observed in ARK2 wild type (WT), using glyceraldehyde-3-phosphate dehydrogenase (Assay ID Hs99999905_m1) mRNA as an internal control.

Cell culture and stable transfection of USC-ARK2 cell line

USC-ARK2 cells were used to generate cell lines expressing histone variants by stable transfection. USC-ARK2 cells were grown in RPMI 1640 (Gibco® #11875-085) with 10% Hi-FBS (Gibco® #16140-071) and 1% Antibiotic- Antimycotic (Gibco® #15240-062). A G418 (Bio-Rad Laboratories, #170-3220) dilution curve was set up to determine the minimum killing concentration before stable transfection. Cells were transfected with Lipofectamin® 2000 following manufacture's protocol. 48 hours post- transfection, cells were re-plated in medium with 800ug/ml G418. After 12 days of selection, the concentration of G418 was lowered to 400ug/ml for maintenance.

Cell migration and Invasion assays.

Cell migration and invasion experiments were performed in 24-transwell chambers with polycarbonate membrane (6.5mm diameter, 8.0 μm pore size; Corning Incorporated, Corning, NY) as previously described[4]. Briefly, cells were kept in starvation for 24 hours before performing the assay. Cells (100,000/200μL) were added to the upper chambers in serum-free RPMI 1640. 800μL of RPMI 1640 with 10% FBS was added to the lower chambers as a chemoattractant. The plates were incubated at 37°C for 48 hours. Migration was terminated by fixing the membranes with 10% buffered formalin phosphate and staining with crystal violet (0.1% w/v in ddH2O) for 25 minutes at room temperature. After rinsing the filters in ddH2O, the interior of each insert was scrubbed with a cotton swab to remove cells that did not migrate. The stained cells were lysed with Triton X-100 (0.5% in ddH2O) overnight at room temperature. The resulting colored mixture was transferred to a 96-well plate and optical densities were read at 595nm in a VersaMax™ microtiter plate reader.

References

1. Choi, M. et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc. Natl. Acad. Sci. 106, 19096–19101 (2009).

2. Zhao, S. et al. Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. Proc. Natl. Acad. Sci. U. S. A. 110, 2916–2921 (2013).

3. TCGA. Integrated genomic analyses of ovarian carcinoma. Nature 474, 609–15 (2011).

4. Bellone M, et al. Expression of αV-Integrins in Uterine Serous Papillary Carcinomas; Implications for Targeted Therapy With Intetumumab (CNTO 95), a Fully Human Antagonist Anti–αV-Integrin Antibody. *Int J Gynecol Cancer* 21(6):1084–1090 (2011).

5. Dedes, K. J., Wetterskog, D., Ashworth, A., Kaye, S. B. & Reis-Filho, J. S. Emerging therapeutic targets in endometrial cancer. Nat. Rev. Clin. Oncol. 8, 261–71 (2011).

6. Shern, J. F. et al. Comprehensive Genomic Analysis of Rhabdomyosarcoma Reveals a Landscape of Alterations Affecting a Common Genetic Axis in Fusion-Positive and Fusion-Negative Tumors. Cancer Discov. (2014).

7. McWilliam, H. et al. Analysis tool web services from the EMBL-EBI. Nucleic Acids Res. 41, W597–W600 (2013).

8. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189-1191 (2009)

ID	Sample Type	CARC Type / %	SARC Type / %	Origin	Status	Age	Stage	Grade	Chemo
SARB1	fresh frozen	high-grade serous carcinoma 5%	heterologous rhabdomyosarcoma, chondrosarcoma 95%	uterus	matched	63	IIIA	G3	NO
SARB2	fresh frozen	endometrioid carcinoma 90%	heterologous osteosarcoma 10%	uterus	matched	70	IB	G3	NO
SARB3	fresh frozen	high-grade serous carcinoma 50%	heterologous chondrosarcoma 50%	ovary	matched	72	IIIC	G3	YES
SARB4	fresh frozen	high-grade serous carcinoma 30%	homologous,heterologous endometrial stromal sarcoma, minor rhabdomyosarcoma component 70%	uterus	matched	73	II	G3	NO
SARB5	fresh frozen	high-grade serous carcinoma, endometrioid carcinoma 85%	homologous undifferentiated 15%	uterus	matched	79	IIIC	G3	NO
SARB6	fresh frozen	high-grade serous carcinoma, endometrioid carcinoma 85%	heterologous rhabdomyosarcoma, chondrosarcoma 15%	ovary	matched	52	IIIC	G3	NO
SARB7	fresh frozen	endometrioid carcinoma 95%	homologous endometrial stromal sarcoma 5%	ovary	matched	59	IIIB	G3	NO
SARB8	fresh frozen	high-grade serous carcinoma, clear cell carcinoma 10%	heterologous rhabdomyosarcoma, chondrosarcoma 90%	ovary	matched	59	IIIA	G3	NO
SARB9	fresh frozen	high-grade serous carcinoma 10%	heterologous rhabdomyosarcoma 90%	ovary	matched	77	IIC	G3	NO
SARB10	fresh frozen	endometrioid carcinoma 95%	heterologous rhabdomyosarcoma, angiosarcoma 5%	ovary	matched	86	IIC	G3	NO
SARB11	fresh frozen	high-grade serous carcinoma, endometrioid carcinoma 95%	homologous endometrial stromal sarcoma 5%	uterus	matched	50	IVB	G3	NO
SARB12	fresh frozen	high-grade serous carcinoma 10%	heterologous rhabdomyosarcoma 90%	uterus	matched	68	IA	G3	NO
SARB13	fresh frozen	endometrioid carcinoma 60%	heterologous rhabdomyosarcoma 40%	uterus	matched	56	IIIC	G3	NO
SARB14	fresh frozen	high-grade serous carcinoma, endometrioid carcinoma 10%	homologous undifferentiated 90%	ovary	matched	69	IIIC	G3	NO
SARB15	fresh frozen	clear cell carcinoma 95%	homologous undifferentiated 5%	ovary	matched	61	IIIC	G3	NO
SARB16	fresh frozen	endometrioid carcinoma 75%	heterologous chondrosarcoma 25%	uterus	matched	56	II	G3	NO
SARB17	fresh frozen	endometrioid carcinoma 5%	heterologous rhabdomyosarcoma, liposarcoma 95%	uterus	matched	59	II	G3	NO
SARB18	fresh frozen	endometrioid carcinoma 5%	homologous endometrial stromal sarcoma 95%	uterus	matched	46	II	G3	NO
SARB19	fresh frozen	high-grade serous carcinoma, endometrioid carcinoma 40%	homologous undifferentiated 60%	ovary	unmatched	62	IV	G3	NO
SARB20	fresh frozen	endometrioid carcinoma 25%	homologous endometrial stromal sarcoma 75%	uterus	matched	85	IA	G3	NO
SARB21	fresh frozen	endometrioid carcinoma, undifferentiated 90%	homologous endometrial stromal sarcoma 10%	uterus	matched	66	IB	G3	NO

SARB22	fresh frozen	high-grade serous carcinoma, endometrioid carcinoma 5%	homologous endometrial stromal sarcoma 95%	uterus	matched	79	II	G3	NO
SARB23	fresh frozen	high-grade serous carcinoma, endometrioid carcinoma 5%	homologous undifferentiated 95%	uterus	unmatched	67	IB	G3	NO
SARB24	fresh frozen	high-grade serous carcinoma 5%	heterologous rhabdomyosarcoma, chondrosarcoma 95%	uterus	unmatched	67	IA	G3	NO
SARB25	fresh frozen	high-grade serous carcinoma, clear cell carcinoma 95%	heterologous rhabdomyosarcoma 5%	ovary	unmatched	61	IIIB	G3	YES
SARB26	fresh frozen	endometrioid carcinoma, undifferentiated 50%	heterologous rhabdomyosarcoma 50%	ovary	unmatched	65	IIIC	G3	NO
SARB27	fresh frozen	high-grade serous carcinoma 90%	heterologous rhabdomyosarcoma 10%	uterus	unmatched	81	IIIC	G3	NO
SARB28	fresh frozen	high-grade serous carcinoma, undifferentiated 10%	heterologous chondrosarcoma 90%	uterus	unmatched	75	IB	G3	NO
SARCL1	cell line	low-grade endometrioid 5%, clear cell carcinoma80%	homologous undifferentiated sarcoma 15%	uterus	matched	70	IC	G3	NO
SARCL4	cell line	high-grade serous carcinoma 50%	heterologous chrondrosarcoma 50%	ovary	matched	77	IIIC	G3	YES
SARCL5	cell line	high-grade serous carcinoma 95%	heterologous chrondrosarcoma 5%	ovary	matched	70	IIIB	G3	NO
SARCL6	cell line	high-grade serous carcinoma 60%	heterologous chrondrosarcoma 40%	ovary	unmatched	78	IIB	G3	YES
SARCL7	cell line	clear cell and serous carcinoma 90%	heterologous chondrosarcoma 10%	ovary	unmatched	55	IV	G3	YES
SAR10	fresh frozen	endometrioid carcinoma 30%	homologous 70%	uterus	matched	64	IIIA	G3	NO
SAR11	fresh frozen	endometrioid and serous carcinoma 45%	heterologous rhabdomyosarcoma, chondrosarcoma, undifferentiated sarcoma 55%	ovary	unmatched	60	IIIC	G3	NO
SAR12	fresh frozen	endometrioid carcinoma 98%	homologous undifferentiated sarcoma 2%	uterus	matched	66	IV	G3	NO
SAR13	fresh frozen	high-grade serous carcinoma 95%	homologous 5%	uterus	matched	72	IVB	G3	NO
SAR14	fresh frozen	endometrioid carcinoma with squamous differentiation 15%	heterologous chrondrosarcoma 50%, undifferentiated sarcoma 35%	uterus	matched	90	IB	G3	NO
SAR15	fresh frozen	high-grade serous carcinoma 98%	homologous undifferentiated sarcoma 2%	uterus	matched	62	IC	G3	NO
SAR16	fresh frozen	high-grade serous carcinoma 25%	homologous undifferentiated and heterologous rhabdomyosacoma 75%	uterus	unmatched	63	IB	G3	NO
SAR17	fresh frozen	high-grade endometrioid and serous adenocarcinoma 95%	homologous undifferentiated sarcoma,heterologous focal chondrosarcoma 5%	ovary	matched	73	IV	G3	NO
SAR18	fresh frozen	clear cell and endometrioid carcinoma 60%	homologous undifferentiated sarcoma 40%	ovary	unmatched	50	IIIA	G3	NO
SAR2	fresh frozen	high-grade serous carcinoma 30%	homologous 70%	uterus	unmatched	63	IB	G3	NO
SAR20	fresh frozen	high-grade serous carcinoma 50%	homologous 50%	uterus	matched	85	IIIC	G3	NO

SAR21	fresh frozen	endometrioid carcinoma 10%	homologous undifferentiated sarcoma and heterologous rhabdomyosarcoma 90%	uterus	unmatched	66	IB	G3	NO
SAR22	fresh frozen	high-grade serous carcinoma 40%	homologous 60%	uterus	unmatched	64	IB	G3	NO
SAR23	fresh frozen	high-grade serous and endometrioid carcinoma 50%	homologous undifferentiated sarcoma 50%	uterus	unmatched	82	IIIA	G3	NO
SAR24	fresh frozen	high-grade serous carcinoma 70%	heterologous elements rhabdomyosarcoma 30%	uterus	matched	69	IA	G3	NO
SAR25	fresh frozen	high-grade endometrioid carcinoma with squamous diff. 90%	homologous undifferentiated sarcoma 10%	uterus	matched	60	IB	G3	NO
SAR27	fresh frozen	high-grade serous and endometrioid carcinoma 70%	homologous 30%	ovary	matched	65	IIIB	G3	NO
SAR28	fresh frozen	high-grade endometrioid carcinoma95% serous carcinoma 3%	homologous undifferentiated sarcoma 2%	ovary	unmatched	72	IIIC	G3	NO
SAR29	fresh frozen	low and high-grade endometrioid carcinoma 15%	heterologous rhabdomyosarcoma 10%, homologous undifferentiated sarcoma 75%	uterus	matched	78	IIIA	G3	NO
SAR3	fresh frozen	high-grade serous and clear cell carcinoma 85%	homologous undifferentiated sarcoma, heterologous liposarcoma 15%	uterus	unmatched	76	IA	G3	NO
SAR30	fresh frozen	endometrioid and undifferentiated carcinoma 80%	homologous 20%	uterus	unmatched	57	IA	G3	NO
SAR31	fresh frozen	high-grade Mullerian (favor serous carcinoma) 40%	heterologous chondrosarcoma 10%, homologous undifferentiated sarcoma 50%	uterus	matched	76	IV	G3	NO
SAR32	fresh frozen	high-grade endometrioid carcinoma75%	homologous undifferentiated sarcoma 25%	uterus	matched	55	IIIC	G3	NO
SAR33	fresh frozen	high-grade serous carcinoma 50%	heterologous chrondrosarcoma 50%	ovary	matched	58	IIIC	G3	NO
SAR34	fresh frozen	adenoid basal carcinoma and squamous cell carcinoma 30%	homologous 70%	uterus	matched	73	IIB	G3	NO
SAR35	fresh frozen	high-grade serous and endometrioid carcinoma 80%	heterologous chrondrosarcoma 20%	uterus	matched	65	IA	G3	NO
SAR36	fresh frozen	high-grade serous carcinoma 50%	heterologous rhabdomyosarcoma, chondrosarcoma 50%	ovary	unmatched	46	IV	G3	NO
SAR38	fresh frozen	low-grade endometrioid carcinoma 50%	homologous undifferentiated sarcoma 50%	uterus	unmatched	81	IIIA	G3	NO
SAR39	fresh frozen	high-grade serous carcinoma15%	homologous undifferentiated sarcoma 85%	uterus	unmatched	68	IV	G3	NO
SAR41	fresh frozen	high-grade serous and endometrioid carcinoma 95%	homologous 5%	ovary	matched	64	IIIC	G3	NO
SAR42	fresh frozen	high-grade serous carcinoma 40%	homologous 60%	uterus	matched	70	IA	G3	NO
SAR6	fresh frozen	high-grade serous and endometrioid carcinoma 40%	heterologous rhabdomyoblastic and cartilagineous elements 60%	ovary	unmatched	63	IC	G3	NO

SAR7	fresh frozen	high-grade serous carcinoma 90%	homologous undifferentiated sarcoma 9%, heterologous rhabdomyosarcoma 1%	uterus	unmatched	73	IA	G3	NO
SAR8	fresh frozen	high-grade serous carcinoma 50%	homologous 50%	uterus	unmatched	72	IA	G3	NO
SAR9	fresh frozen	high-grade serous carcinoma 70%	homologous 30%	uterus	matched	65	IC	G3	NO

Table S1 Clinical features of the CS samples (n=68) used for whole exome sequencing.

[1] The six tumors used for multi-region sequencing were SAR24, SAR27, SAR33, SARB14, SARB18 and SARB22. They were later fixed by formalin and subsequently macro-dissected.

[2] CARC type: carcinomatous component histology type; SARC type: sarcomatous component histology type. The percentage was an estimation by the pathology department at Yale based on microscopic examination.

	Normal	Tumor
Read length (bp)	74	74
Number of reads (Million)	97	192
Median coverage (X)	83	156
Mean coverage (X)	101	197
% on genome	91.97%	92.32%
% on target	67.58%	66.47%
% of bases covered at least 8x	95.95%	97.25%
% of bases covered at least 20x	90.18%	94.45%
Mean error rate	0.42%	0.47%
% of PCR duplicate	5.26%	6.78%

Table S2 Sequencing metrics for tumor samples and matched normal samples.

	CS	Uterine serous carcinoma from Yale and TCGA	Uterine endometrioid carcinoma from TCGA	Ovarian serous carcinoma from TCGA
# H2A/H2B nonsilent mutations	10	11	6	13
# H2A/H2B genes affected	5	10	6	10
Mutation clustering *p* value	0.0006	0.8	1	0.4

Table S3 Distribution of nonsilent mutations found in 37 core H2A/H2B genes.

The clustering *p* value estimates how likely it is to get equal or less genes being hit given the observed number of genes by chance.

		10 CSs (uterine origin, serous histology)				25 USCs			
Arm	# Genes	Amp frequency	Amp q-value	Del Frequency	Del q-value	Amp frequency	Amp q-value	Del Frequency	Del q-value
1p	2121	30%	0.895	0%	0.943	12%	0.99	5%	0.979
1q	1955	**50%**	**0.167**	0%	0.943	20%	0.99	0%	0.979
2p	924	22%	0.908	12%	0.943	22%	0.99	10%	0.979
2q	1556	33%	0.786	14%	0.943	13%	0.99	9%	0.979
3p	1062	14%	0.936	33%	0.526	14%	0.99	14%	0.979
3q	1139	12%	0.936	22%	0.846	36%	0.586	18%	0.979
4p	489	0%	0.936	20%	0.846	0%	0.99	32%	0.372
4q	1049	0%	0.936	40%	0.283	0%	0.99	**48%**	**0.003**
5p	270	**50%**	**0.167**	0%	0.943	17%	0.99	10%	0.979
5q	1427	10%	0.936	0%	0.943	9%	0.99	9%	0.979
6p	1173	**50%**	**0.167**	0%	0.943	27%	0.99	16%	0.979
6q	839	30%	0.786	0%	0.943	19%	0.99	19%	0.979
7p	641	40%	0.359	0%	0.943	18%	0.99	14%	0.979
7q	1277	22%	0.908	12%	0.943	14%	0.99	18%	0.979
8p	580	25%	0.908	25%	0.785	32%	0.99	32%	0.438
8q	859	**50%**	**0.167**	33%	0.598	**50%**	**0.0329**	33%	0.372
9p	422	0%	0.936	40%	0.261	0%	0.99	28%	0.599
9q	1113	0%	0.936	**60%**	**0.0133**	0%	0.99	**40%**	**0.0315**
10p	409	22%	0.908	12%	0.943	11%	0.99	26%	0.715
10q	1268	22%	0.908	12%	0.943	11%	0.99	26%	0.52
11p	862	50%	0.26	**57%**	**0.0595**	19%	0.99	**41%**	**0.0476**
11q	1515	33%	0.786	14%	0.943	19%	0.99	**41%**	**0.0226**
12p	575	12%	0.936	22%	0.846	23%	0.99	15%	0.979
12q	1447	12%	0.936	22%	0.846	18%	0.99	14%	0.979
13q	654	0%	0.936	10%	0.943	26%	0.99	30%	0.483
14q	1341	0%	0.936	30%	0.605	17%	0.99	5%	0.979
15q	1355	0%	0.936	40%	0.283	0%	0.99	**44%**	**0.007**
16p	872	0%	0.936	**70%**	**0.001**	12%	0.99	**39%**	**0.06**
16q	702	0%	0.936	**70%**	**0.001**	15%	0.99	**52%**	**0.003**
17p	683	25%	0.908	**67%**	**0.004**	25%	0.99	**59%**	**0.0002**
17q	1592	0%	0.936	**60%**	**0.014**	12%	0.99	**39%**	**0.03**
18p	143	0%	0.936	40%	0.261	27%	0.99	16%	0.979
18q	446	0%	0.936	20%	0.846	18%	0.99	14%	0.979
19p	995	14%	0.936	33%	0.526	17%	0.99	32%	0.331
19q	1709	29%	0.908	38%	0.488	19%	0.99	**41%**	**0.0179**
20p	355	22%	0.908	12%	0.943	32%	0.99	17%	0.979
20q	753	30%	0.786	0%	0.943	29%	0.99	6%	0.979
21q	509	38%	0.564	29%	0.65	11%	0.99	26%	0.715
22q	921	0%	0.936	30%	0.598	9%	0.99	**58%**	**0.0001**
Xq	1312	**57%**	**0.167**	50%	0.254	10%	0.99	17%	0.979

Table S4. GISTIC2.0 identified significant broad CNV events in CSs with uterine serous histology and USCs.

Significance of broad CNV events (length > 80% of chromosomal arm) were assessed by GISTIC software for both CSs and USCs. Entries highlighted in bold format were events that passed the genome wide significance cut off (*q* value <0.25). For a fair comparison, we used 10 CSs from uterus with serous histology to compare with USC.

	Normal	Sarcomatous samples	Carcinomatous samples
Number of samples	6	12	12
Read length (bp)	74	74	74
# of reads (Million)	114	290	283
Median coverage (X)	95	190	204
Mean coverage (X)	118	256	266
% on genome	90.88%	87.10%	87.45%
% on target	67.92%	61.34%	64.50%
% of bases covered at least 4x	97.77%	98.41%	98.43%
% of bases covered at least 8x	96.46%	97.65%	97.70%
% of bases covered at least 20x	91.81%	95.30%	95.62%
Mean error rate	0.50%	1.00%	1.28%
% of PCR duplicate	5.62%	12.70%	13.80%

Table S5 Sequencing metrics for macro-dissected tumor samples and matched normal samples.

Driver genes	Uterine carcinoma	Sarcoma
TP53	20%-90%	0%
KRAS	2%-26%	0%
PIK3CA	20%-30%	5%
PIK3R1	10%-26%	0%
FBXW7	7%-20%	0%
CHD4	9%-20%	0%
TAF1	11%-14%	0%
PTEN	5%-80%	0%

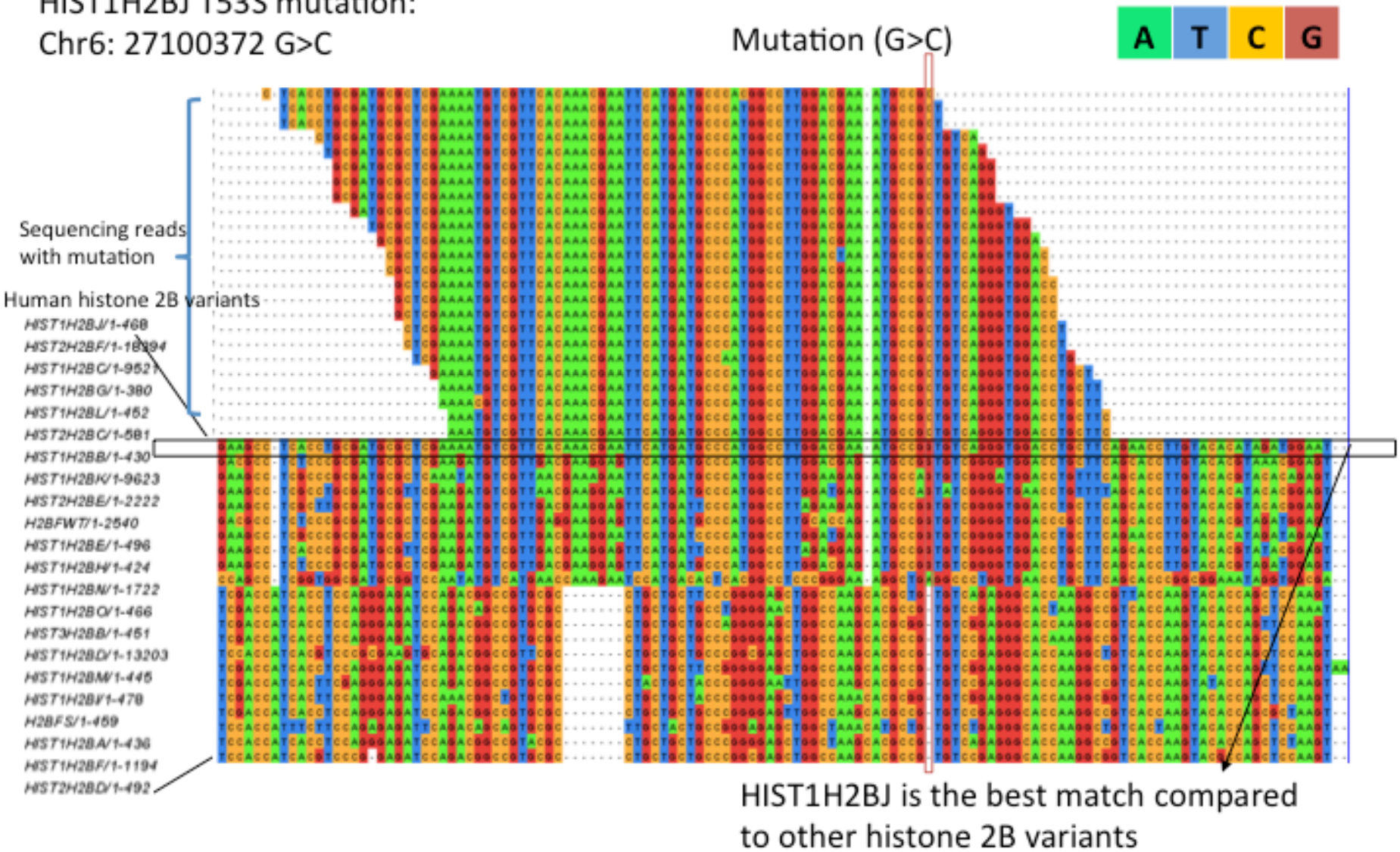
Table S6. Fraction of tumors affected by selected driver genes in uterine carcinomas and sarcomas.

8 driver genes were found having root mutations in the 6 CSs used for multi-region sequencing. The fraction of tumors affected by mutations in these genes are shown in uterine carcinoma[5] and 58 alveolar rhabdomyosarcomas[6].

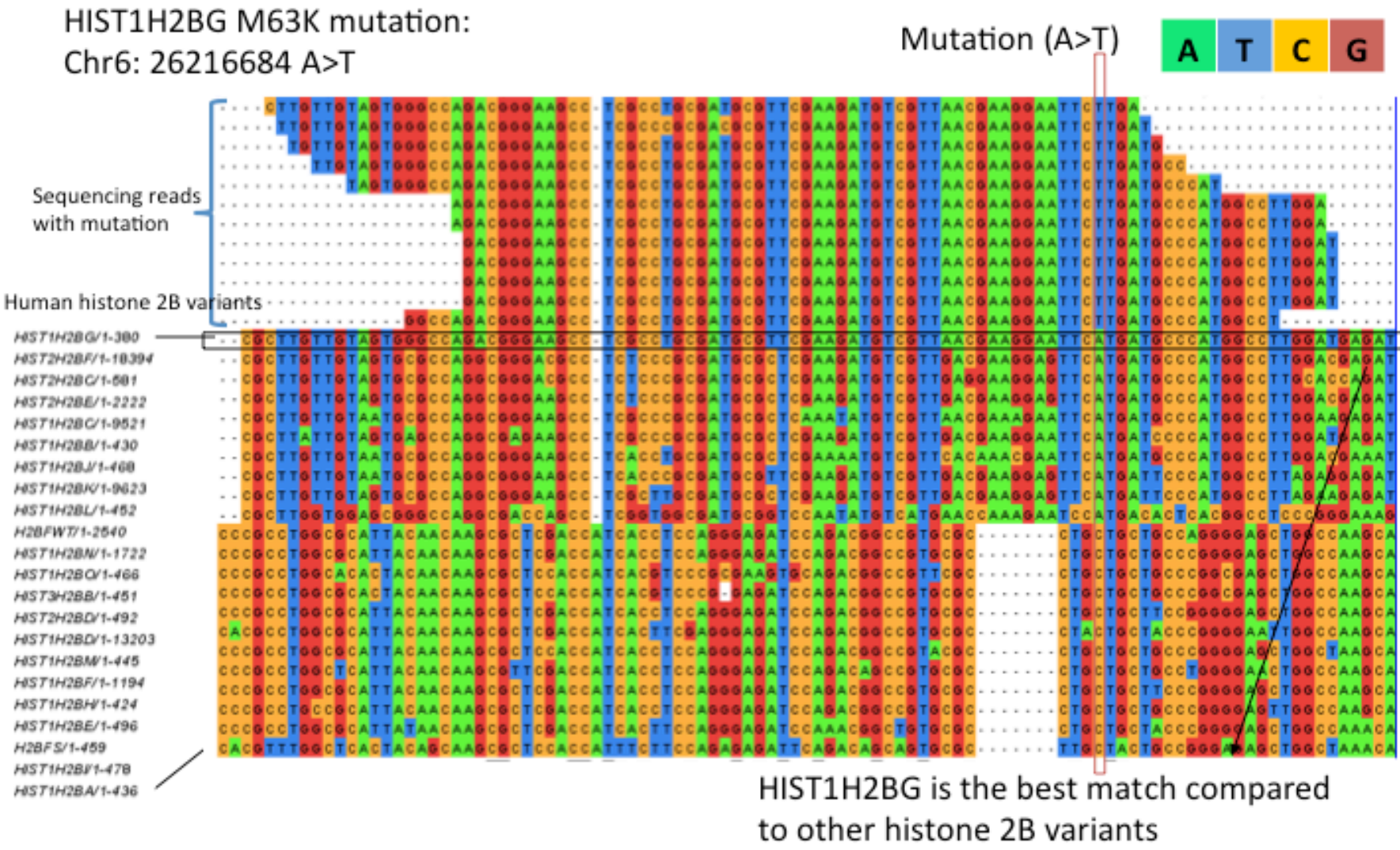
	Normal	Tumor
Number of samples	46	46
Read length (bp)	74	74
# of reads (Million)	92	191
Median coverage (X)	83	159
Mean coverage (X)	99	196
% on genome	92.00%	92.06%
% on target	70.40%	67.12%
% of bases covered at least 8x	95.76%	97.17%
% of bases covered at least 20x	90.07%	94.58%
Mean error rate	0.47%	0.50%
% of PCR duplicate	4.58%	8.19%

Table S7 Sequencing metrics for Yale processed uterine serous carcinomas which were used as comparators.

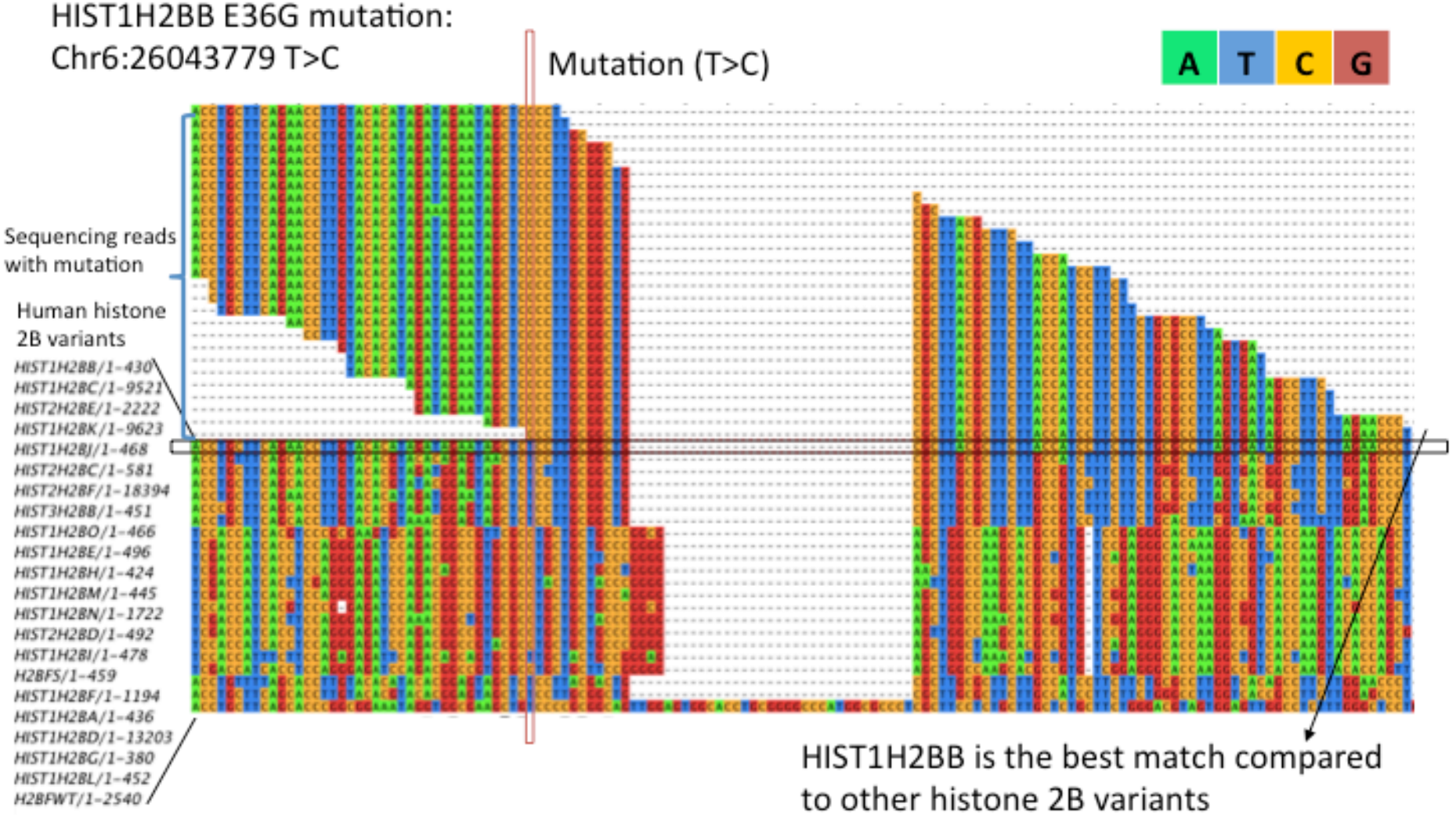
HIST1H2BJ T53S mutation:
Chr6: 27100372 G>C

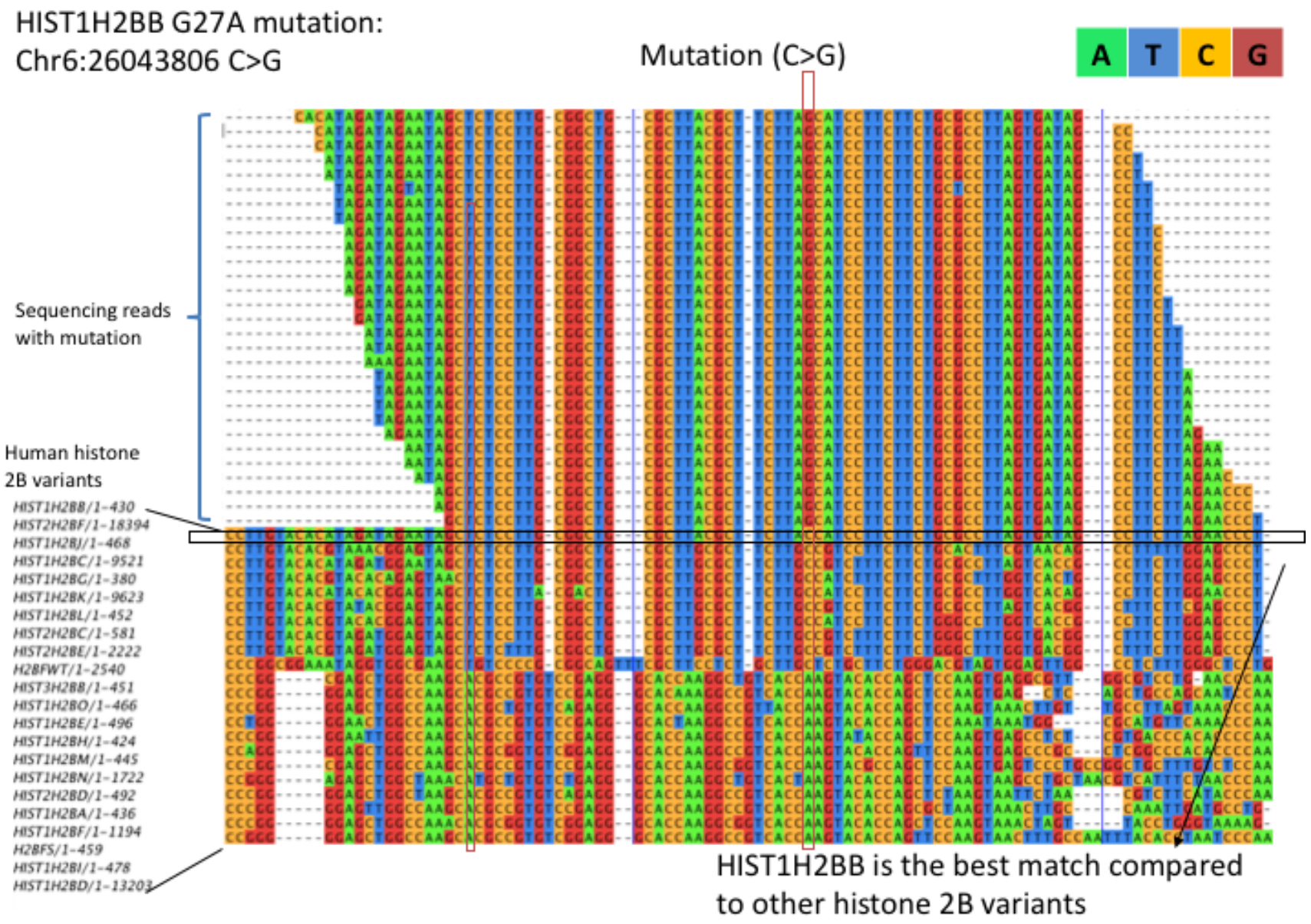


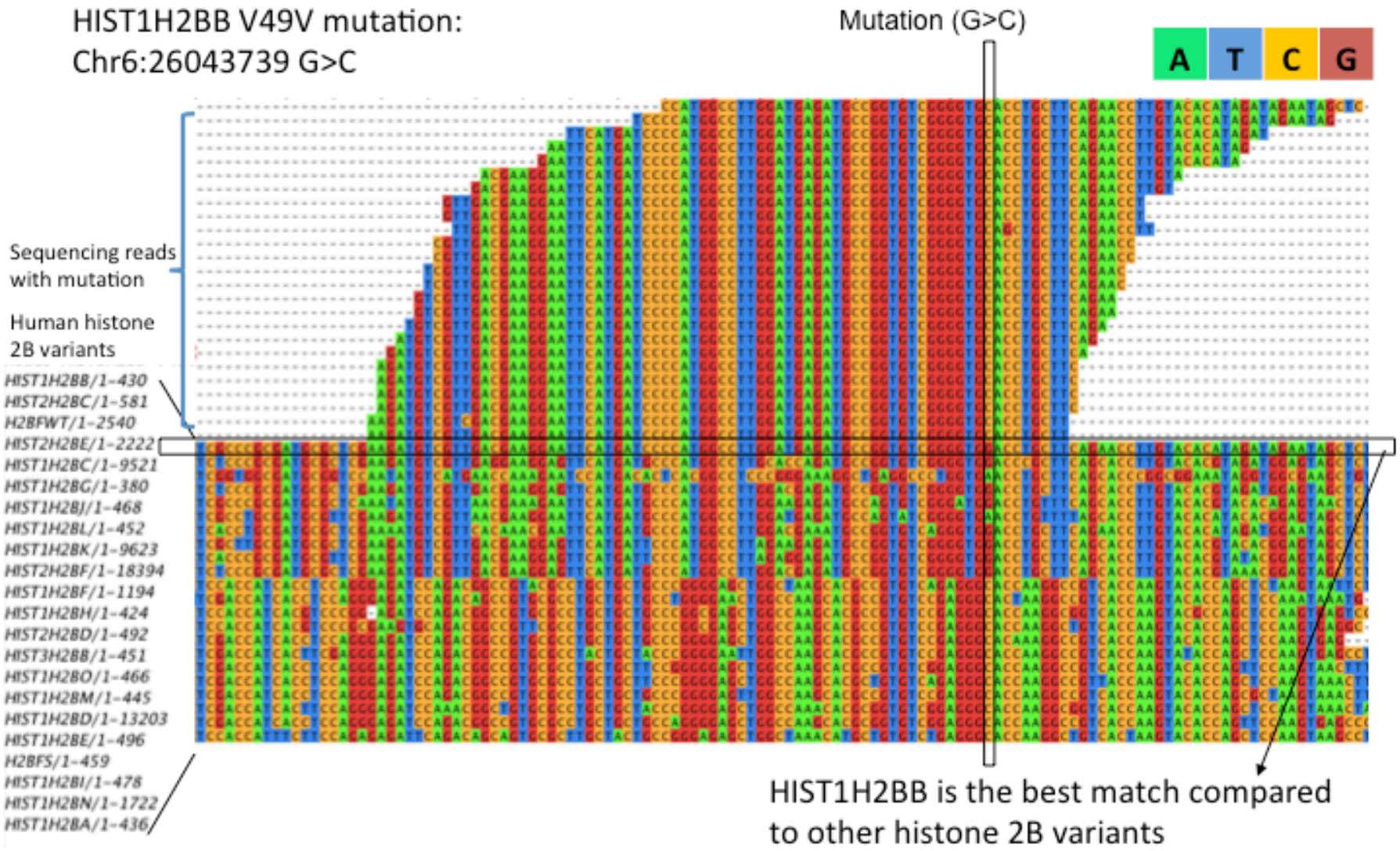

HIST1H2BJ K24X mutation:
Chr6: 27100460 TC>AT
A
T
C
G
Mutation (TC>AT)
Sequencing reads with mutation
Human histone 2B variants
HIST1H2BJ/1-468
HIST2H2BF/1-1839
HIST2H2BE/1-2222
HIST2H2BC/1-581
HIST1H2BL/1-452
HIST1H2BK/1-9623
HIST1H2BG/1-380
HIST1H2BC/1-9521
HIST1H2BB/1-430
H2BFWT/1-2540
HIST3H2BB/1-451
HIST2H2BD/1-492
HIST1H2BO/1-466
HIST1H2BN/1-1722
HIST1H2BM/1-445
HIST1H2BI/1-478
HIST1H2BH/1-424
HIST1H2BE/1-496
HIST1H2BD/1-1320
HIST1H2BA/1-436
H2BFS/1-459
HIST1H2BF/1-1194
HIST1H2BJ is the best match compared to other histone 2B variants

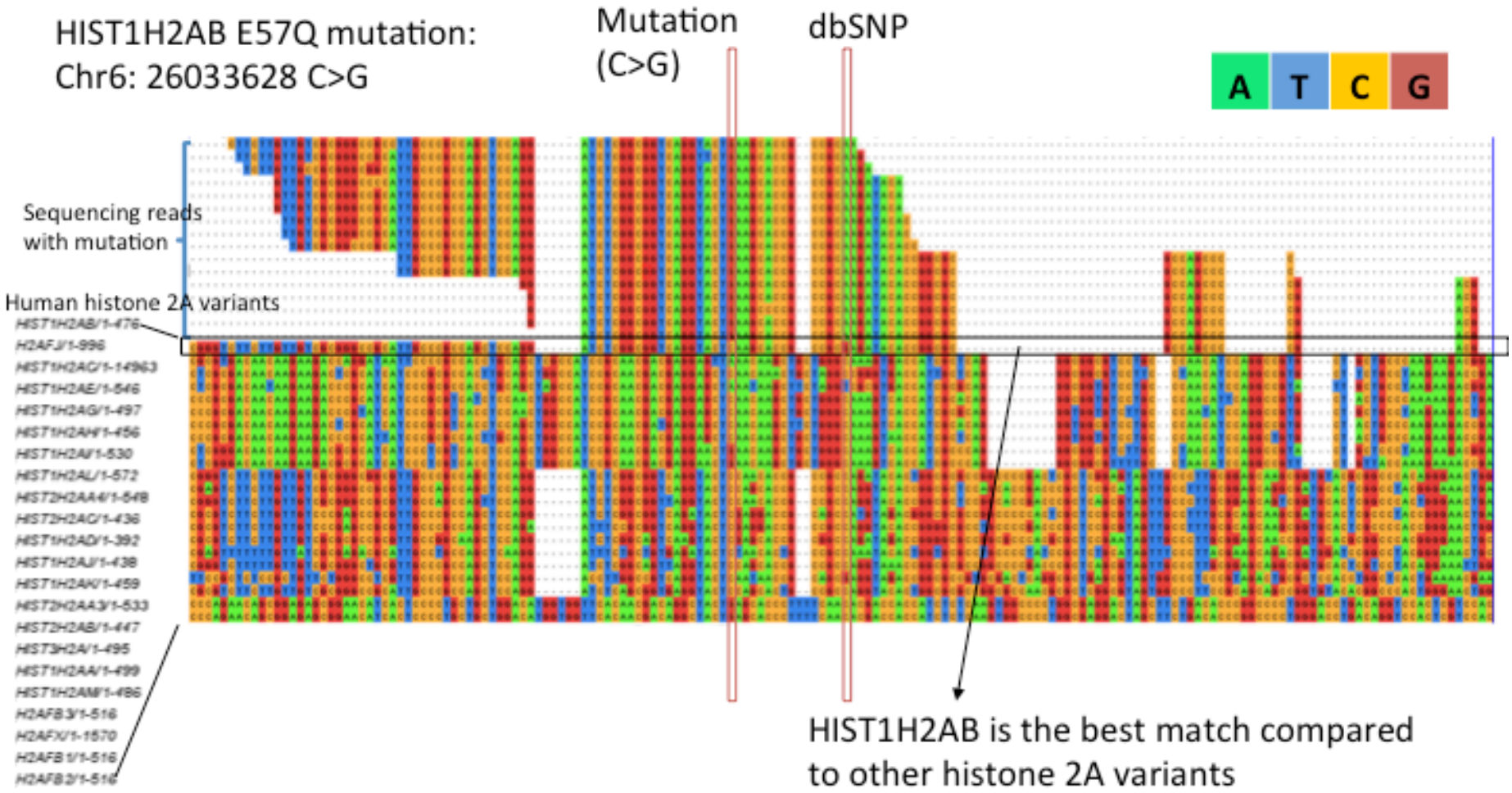

HIST1H2BG M63K mutation:
Chr6: 26216684 A>T
Mutation (A>T)
A
T
C
G
Sequencing reads with mutation
Human histone 2B variants
HIST1H2BG/1-380
HIST2H2BF/1-16394
HIST2H2BC/1-581
HIST2H2BE/1-2222
HIST1H2BC/1-9521
HIST1H2BB/1-430
HIST1H2BJ/1-468
HIST1H2BK/1-9623
HIST1H2BL/1-452
H2BFWT/1-2540
HIST1H2BN/1-1722
HIST1H2BO/1-466
HIST3H2BB/1-451
HIST2H2BD/1-492
HIST1H2BD/1-13203
HIST1H2BM/1-445
HIST1H2BF/1-1194
HIST1H2BH/1-424
HIST1H2BE/1-496
H2BFS/1-459
HIST1H2BI/1-478
HIST1H2BA/1-436
HIST1H2BG is the best match compared to other histone 2B variants


HIST1H2BG A18G mutation:
Chr6: 26216819 G>C
Mutation (G>C)
A
T
C
G
Sequencing reads with mutation
Human histone 2B variants
HIST1H2BG/1-380
HIST2H2BF/1-18394
HIST2H2BE/1-2222
HIST2H2BC/1-581
HIST1H2BL/1-452
HIST1H2BK/1-9623
HIST1H2BJ/1-468
HIST1H2BC/1-9521
HIST1H2BB/1-430
HIST3H2BB/1-451
HIST2H2BD/1-492
HIST1H2BO/1-466
HIST1H2BN/1-1722
HIST1H2BM/1-445
HIST1H2BI/1-478
HIST1H2BH/1-424
HIST1H2BF/1-1194
HIST1H2BE/1-496
HIST1H2BD/1-13203
HIST1H2BA/1-436
H2BFWT/1-2640
H2BFS/1-459
HIST1H2BG is the best match compared to other histone 2B variants


HIST1H2BB E36G mutation:
Chr6:26043779 T>C
Mutation (T>C)
A
T
C
G
Sequencing reads with mutation
Human histone 2B variants
HIST1H2BB/1-430
HIST1H2BC/1-9521
HIST2H2BE/1-2222
HIST1H2BK/1-9623
HIST1H2BJ/1-468
HIST2H2BC/1-581
HIST2H2BF/1-18394
HIST3H2BB/1-451
HIST1H2BO/1-466
HIST1H2BE/1-496
HIST1H2BH/1-424
HIST1H2BM/1-445
HIST1H2BN/1-1722
HIST2H2BD/1-492
HIST1H2BI/1-478
H2BFS/1-459
HIST1H2BF/1-1194
HIST1H2BA/1-436
HIST1H2BD/1-13203
HIST1H2BG/1-380
HIST1H2BL/1-452
H2BFWT/1-2540
HIST1H2BB is the best match compared to other histone 2B variants


HIST1H2BB G27A mutation:
Chr6:26043806 C>G
Mutation (C>G)
A T C G
Sequencing reads with mutation
Human histone 2B variants
HIST1H2BB/1-430
HIST2H2BF/1-18394
HIST1H2BJ/1-468
HIST1H2BC/1-9521
HIST1H2BG/1-380
HIST1H2BK/1-9623
HIST1H2BL/1-452
HIST2H2BC/1-581
HIST2H2BE/1-2222
H2BFWT/1-2540
HIST3H2BB/1-451
HIST1H2BO/1-466
HIST1H2BE/1-496
HIST1H2BH/1-424
HIST1H2BM/1-445
HIST1H2BN/1-1722
HIST2H2BD/1-492
HIST1H2BA/1-436
HIST1H2BF/1-1194
H2BFS/1-459
HIST1H2BI/1-478
HIST1H2BD/1-13203
HIST1H2BB is the best match compared to other histone 2B variants


HIST1H2BB V49V mutation:
Chr6:26043739 G>C
Mutation (G>C)
A
T
C
G
Sequencing reads with mutation
Human histone 2B variants
HIST1H2BB/1-430
HIST2H2BC/1-581
H2BFWT/1-2540
HIST2H2BE/1-2222
HIST1H2BC/1-9521
HIST1H2BG/1-380
HIST1H2BJ/1-468
HIST1H2BL/1-452
HIST1H2BK/1-9623
HIST2H2BF/1-18394
HIST1H2BF/1-1194
HIST1H2BH/1-424
HIST2H2BD/1-492
HIST3H2BB/1-451
HIST1H2BO/1-466
HIST1H2BM/1-445
HIST1H2BD/1-13203
HIST1H2BE/1-496
H2BFS/1-459
HIST1H2BI/1-478
HIST1H2BN/1-1722
HIST1H2BA/1-436
HIST1H2BB is the best match compared to other histone 2B variants


HIST1H2AB E57Q mutation:
Chr6: 26033628 C>G
Mutation
(C>G)
dbSNP
A
T
C
G
Sequencing reads
with mutation
Human histone 2A variants
HIST1H2AB/1-476
H2AFJ/1-996
HIST1H2AG/1-14963
HIST1H2AE/1-546
HIST1H2AG/1-497
HIST1H2AH/1-456
HIST1H2AI/1-530
HIST1H2AL/1-572
HIST2H2AA4/1-648
HIST2H2AC/1-436
HIST1H2AD/1-392
HIST1H2AJ/1-438
HIST1H2AK/1-459
HIST2H2AA3/1-533
HIST2H2AB/1-447
HIST3H2A/1-495
HIST1H2AA/1-499
HIST1H2AM/1-486
H2AFB3/1-516
H2AFX/1-1570
H2AFB1/1-516
H2AFB2/1-516
HIST1H2AB is the best match compared
to other histone 2A variants

HIST1H2AC R4H mutation:
Chr6: 2614471 G>A

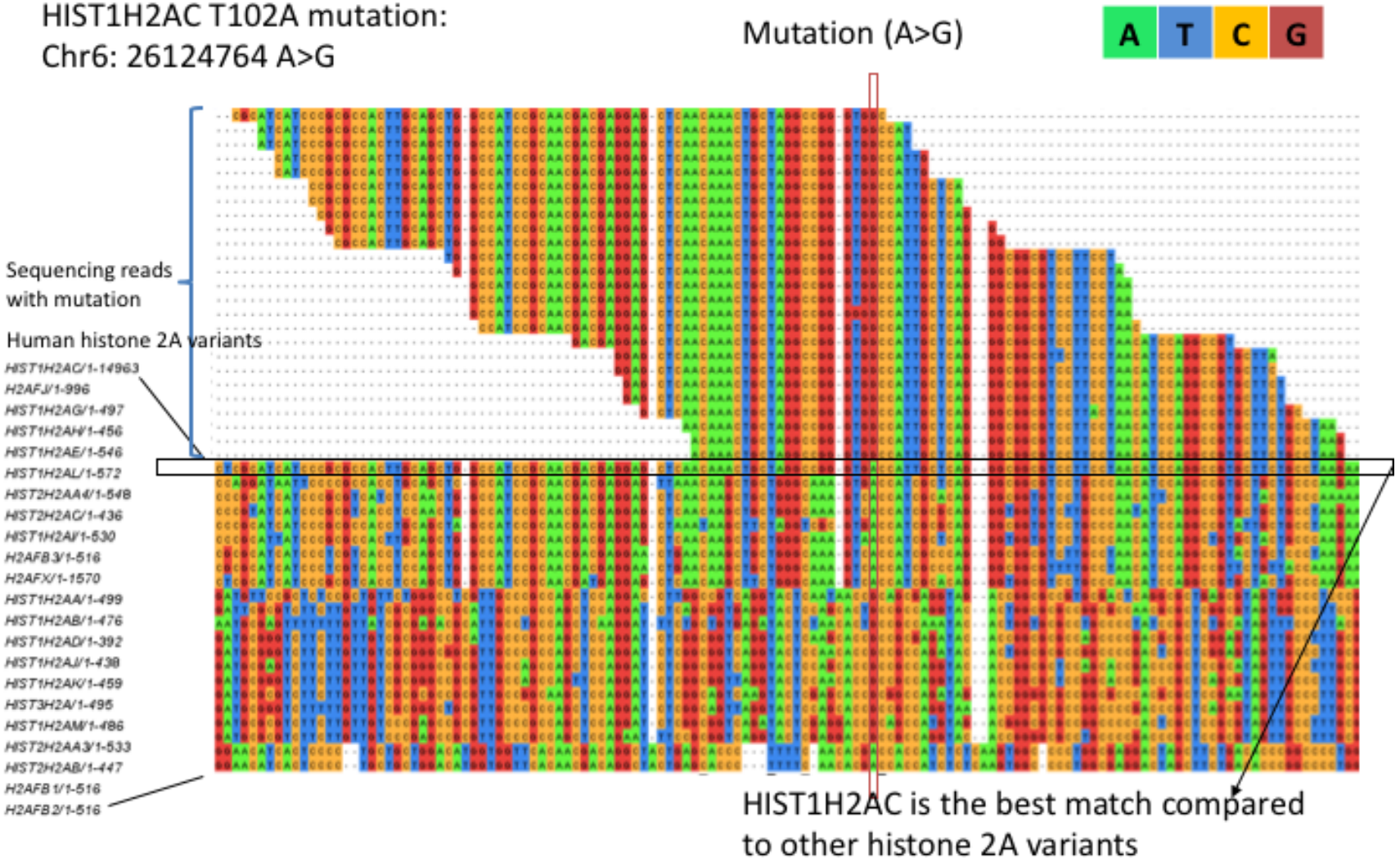

HIST1H2AC T102A mutation:
Chr6: 26124764 A>G
Mutation (A>G)
A
T
C
G
Sequencing reads with mutation
Human histone 2A variants
HIST1H2AC/1-14963
H2AFJ/1-996
HIST1H2AG/1-497
HIST1H2AH/1-456
HIST1H2AE/1-546
HIST1H2AL/1-572
HIST2H2AA4/1-548
HIST2H2AC/1-436
HIST1H2AI/1-530
H2AFB3/1-516
H2AFX/1-1570
HIST1H2AA/1-499
HIST1H2AB/1-476
HIST1H2AD/1-392
HIST1H2AJ/1-438
HIST1H2AK/1-459
HIST3H2A/1-495
HIST1H2AM/1-486
HIST2H2AA3/1-533
HIST2H2AB/1-447
H2AFB1/1-516
H2AFB2/1-516
HIST1H2AC is the best match compared to other histone 2A variants

Figure S1. Reads alignment for H2A/H2B mutations identified in this study.

For each mutation, we performed multiple sequence alignment by ClustalW2[7] for reads with the mutation and sequences of all histone variants. The results of alignments were displayed by Jalview[8] and colored based on nucleotide type. The gene that we identified always aligned the best with sequencing reads and only differed at the site of the mutation, the rest histone variants have multiple discordance sites and could not be sources for reads with the observed mutation.

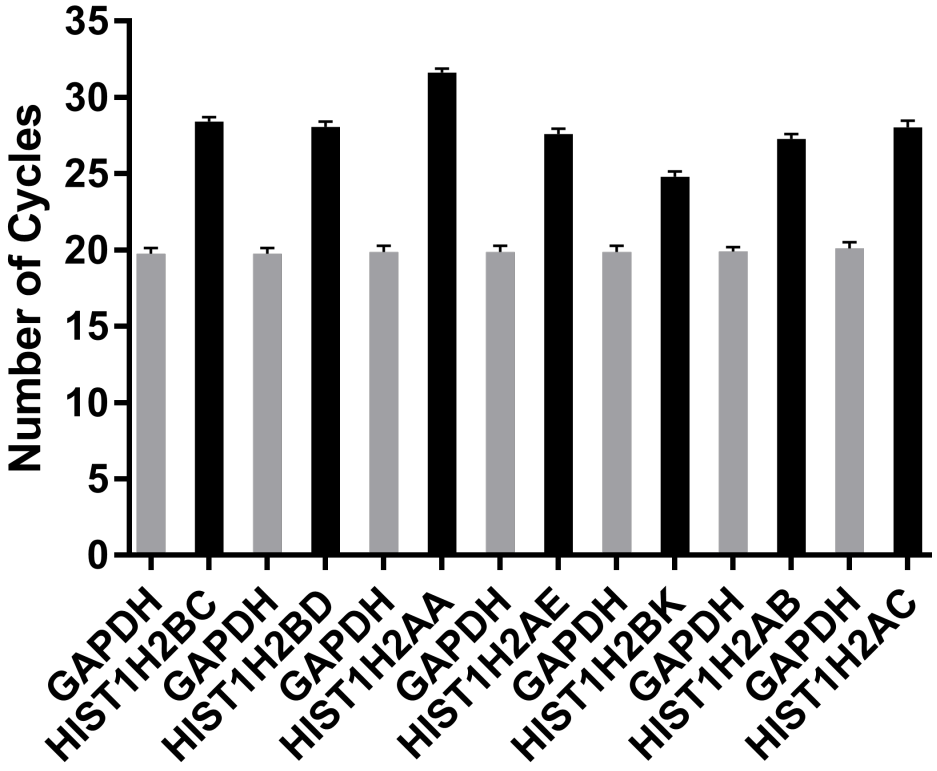


Figure S2. Representative histone core gene expression levels vs GAPDH in primary CS cell lines and fresh-frozen CS tissues.

Expression of mutated H2A/H2B genes and representative non-mutated histone core genes vs GAPDH in CSs by qRT-PCR. Bars describe the mean histone core gene expression ± SD in 10-25 representative CS samples.

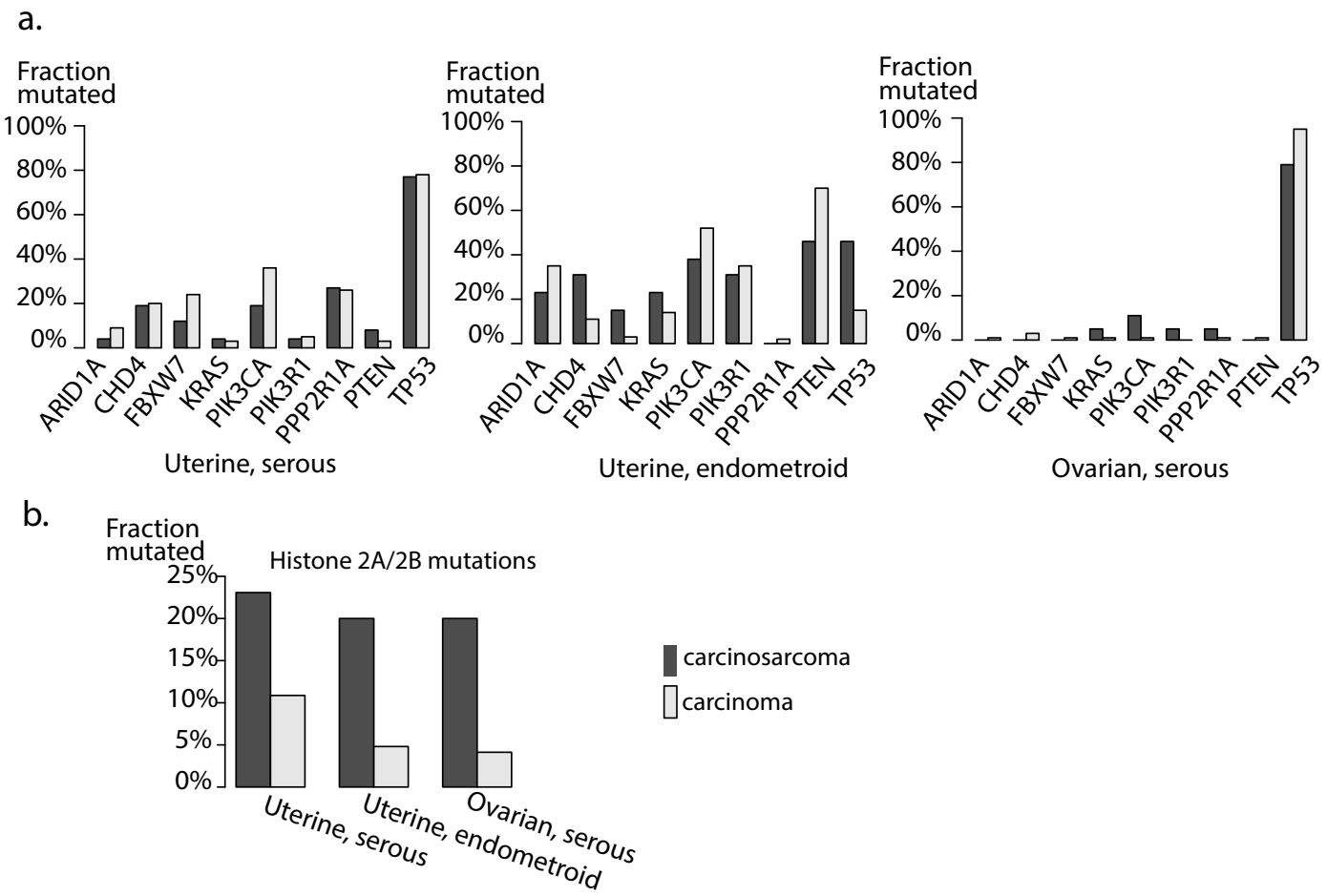


Figure S3. Comparison of driver mutation frequencies between CS and carcinomas.

a. Fraction of tumors altered by mutations found in driver genes in carcinosarcomas and carcinomas side by side for each of the three groups: uterine origin with serous histology, uterine origin with endometrioid histology and ovarian origin with serous histology. b. Fraction of tumors altered by mutations found in histone2A/2B genes in carcinosarcomas and carcinomas side by side for each of the three groups as described in a.

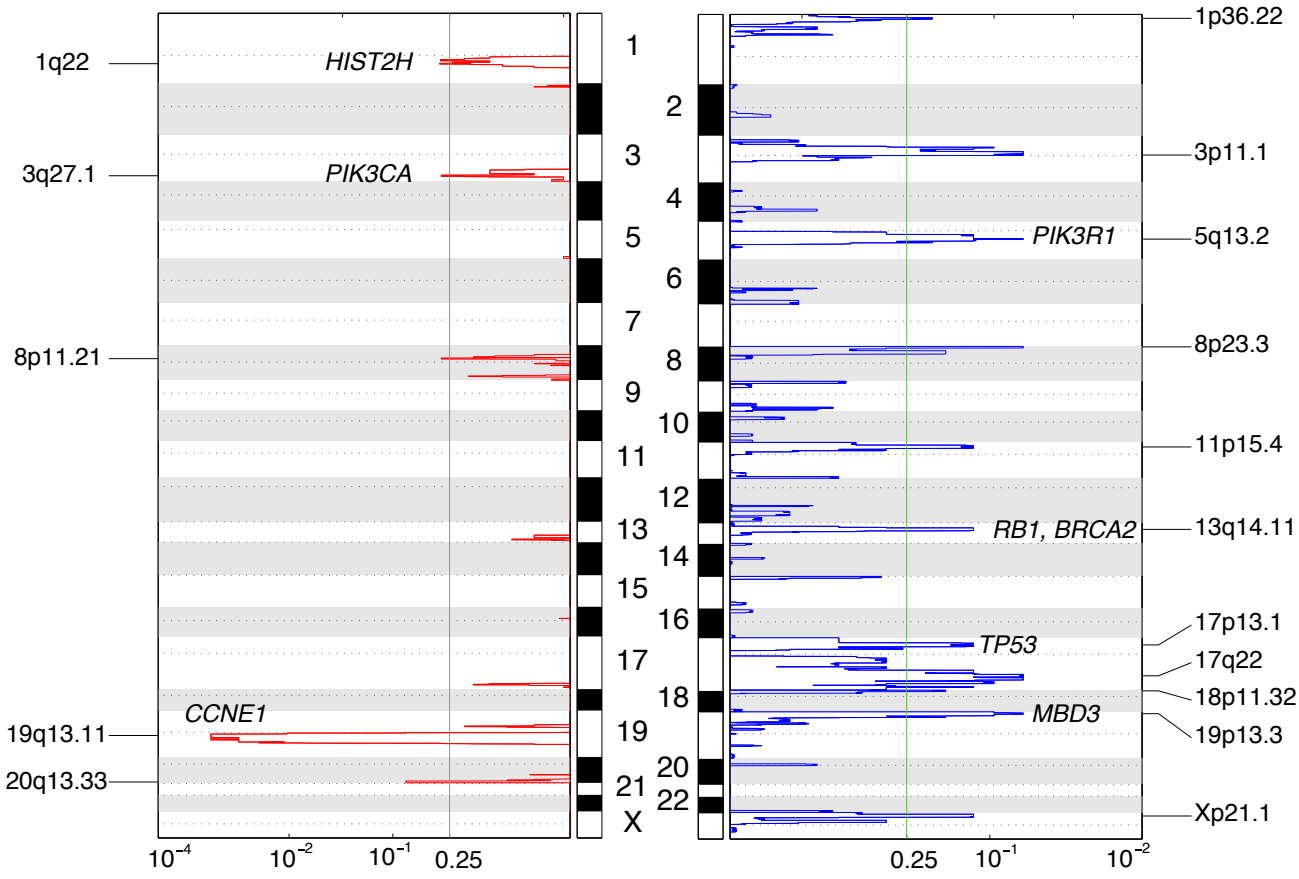


Figure S4. GISTIC 2.0 results for focal CNVs in 29 CS tumors.

GISTIC 2.0 *q* values (horizontal axis) are plotted across the genome (vertical axis). Left, result for CNV gain; right, result for CNV loss. Known gynecological cancer genes in significant gain or loss intervals are marked.

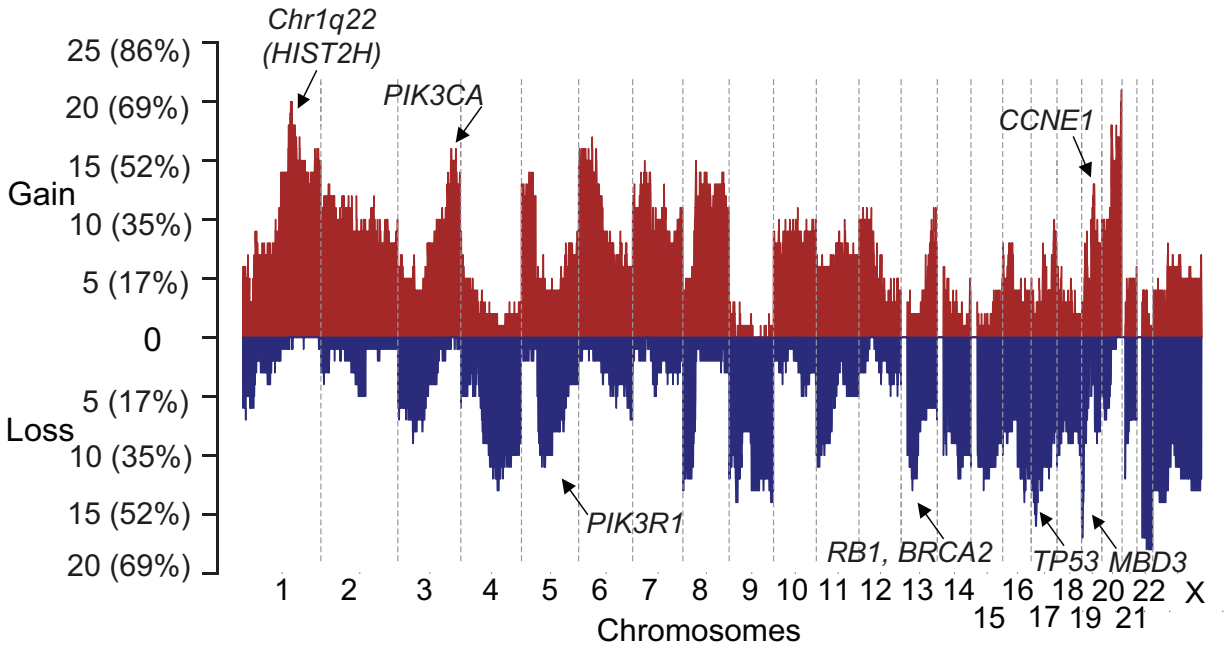


Figure S5. Copy-number profile of 29 CS tumors.

Frequency of copy-number gain (red) and copy-number loss (blue) are plotted along the genome. Known gynecological cancer genes in significant focal gain or loss intervals identified by GISTIC 2.0 are marked.

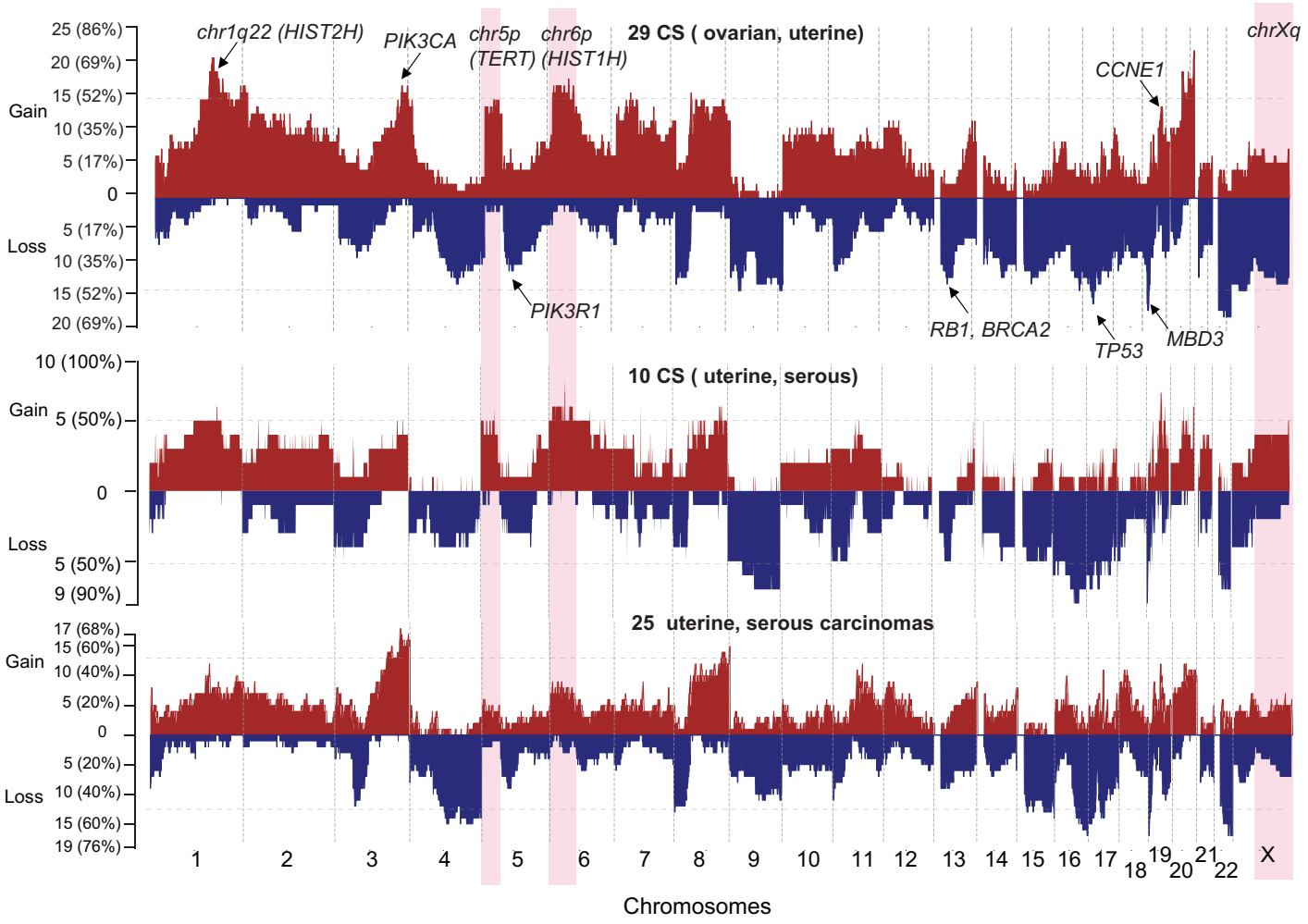


Figure S6. Comparison of copy number profile in uterine carcinosarcoma samples with serous histology and uterine serous carcinoma.

Frequency of copy number gain (red) and copy number loss (blue) are plotted along the genome for both cohorts. Significant focal CNV in CS are labeled with arrows. Significant broad CNV regions only present in CS with serous histology but not in USC are indicated by pink rectangles. The Y-axis labels also mark the percentage corresponding to each frequency value. Genes of interests are highlighted by arrows. "*HIST1H*" indicates cluster 1 of core histone genes on chr6p; "HIST2H" indicates cluster 2 of core histone genes on chr1q.

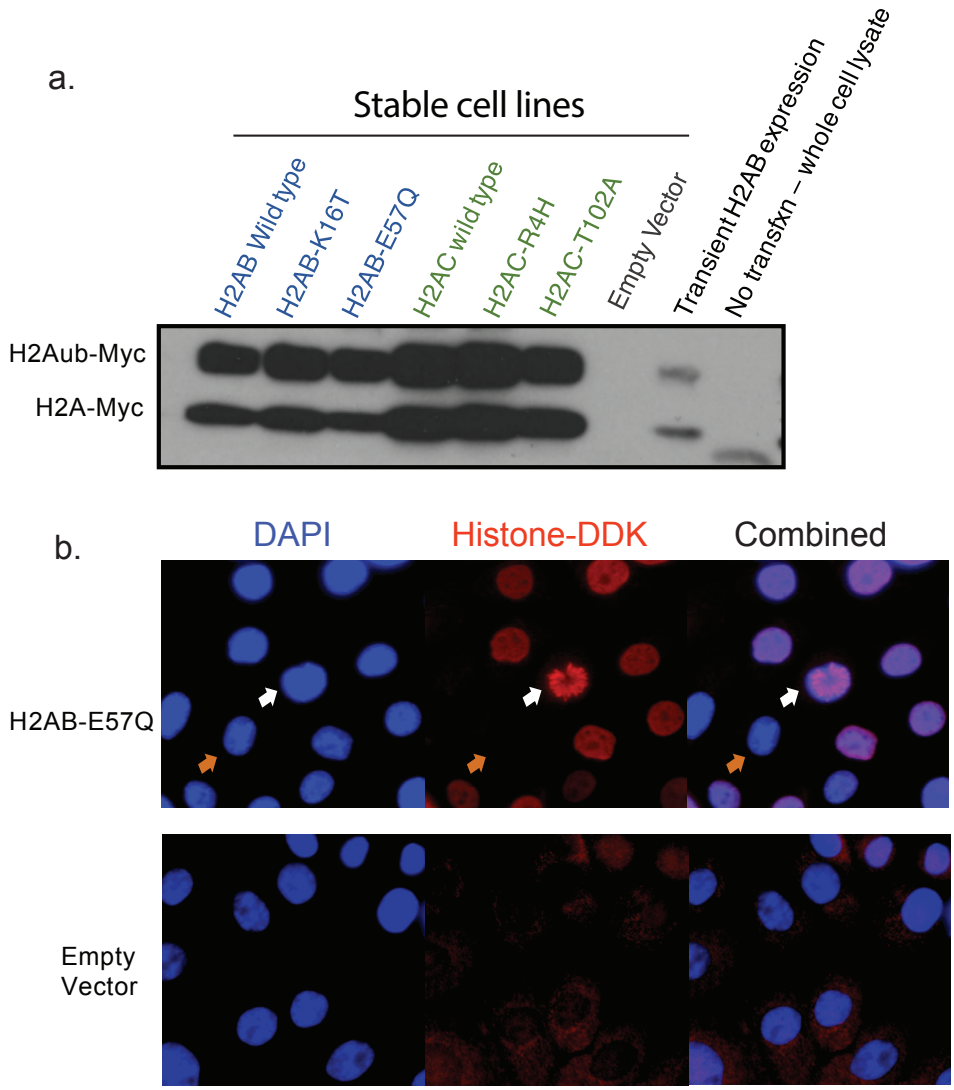


Figure S7. Establishing stable uterine carcinoma cell lines overexpressing histone 2A wild type or histone core gene mutations.

a. Western blot against Myc tag showing expression of transfected Myc-DDK- c terminal tagged wild type or mutated histones. The first 7 lanes are stably transfected cell lines. H2AB: HIST1H2AB; H2AC: HIST1H2AC. We also show histone expression level after 48 hours from transfection (Transient H2AB expression) and a control lane where no transfection was performed.

b. Representative immunofluorescent images of a cell line with stable transfection of mutated histone gene is shown at the top and a control cell line stably transfected with empty vector at the bottom. Nuclei are in blue (DAPI staining, first channel). DDK tag are in red (antibody against DDK, second channel). Most cells are expressing histone-DDK, while a few cells are not (e.g. the cell pointed by orange arrow). We mark one nucleus (white arrow) that clearly shows that exogenous histone has been incorporated into the chromosomal structure.

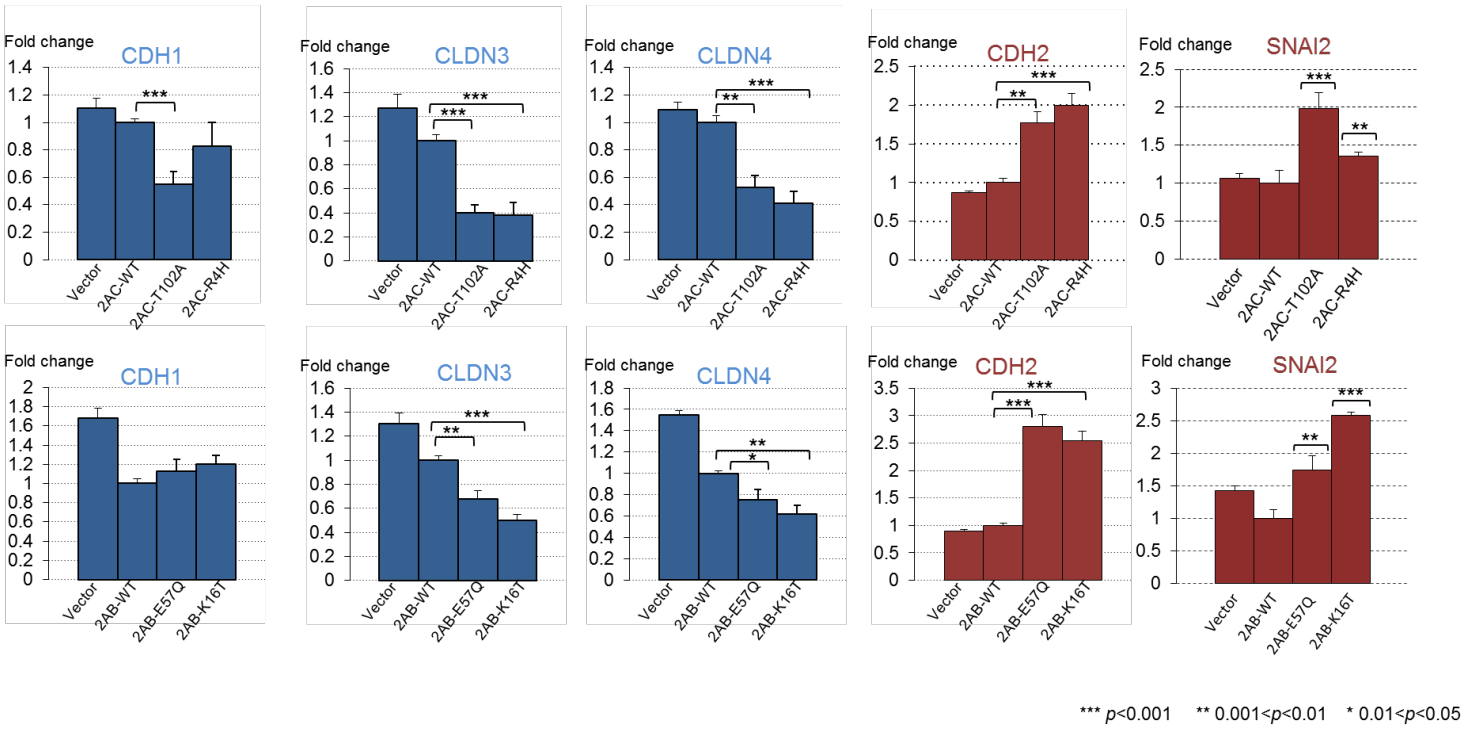


Figure S8. Down-regulation of epithelial markers and up-regulation of mesenchymal markers in stable cell lines expressing mutant histones.

A primary USC cell line (i.e., USC-ARK2, a cell line fully characterized by whole-exome sequencing and harboring no mutations in histone genes) was transfected with vectors expressing wild type HIST1H2AB (H2AB-WT) or E57Q (recurrent mutation) and K16T (found in one of the four outliers) mutants as well as wild type HIST1H2AC (H2AC-WT) or R4H and T102A mutants. Besides these 6 cell lines, we generated a seventh cell line expressing empty vector following the same procedures. qRT-PCR were performed following RNA extraction from these cell lines. qRT-PCR results for three epithelial markers (CDH1/E-cadherin, CLDN3/claudin-3, CLDN4/claudin-4) are shown in blue; results for mesenchymal markers CDH2/N-cadherin and SNAI2/Slug are shown in red. In the first row two HIST1H2AC mutations, 2AC R4H and 2AC T102A are compared with empty vector and H2AC wild type (2AC WT). In the second row two HIST1H2AB mutations, 2AB E57Q and 2AB K16T are compared with empty vector and with wild type H2AB (H2AB WT). Each experiments have three biological replicates. ***, $p<0.001$; ** , $0.001<p<0.01$; *, $0.05<p<0.01$.

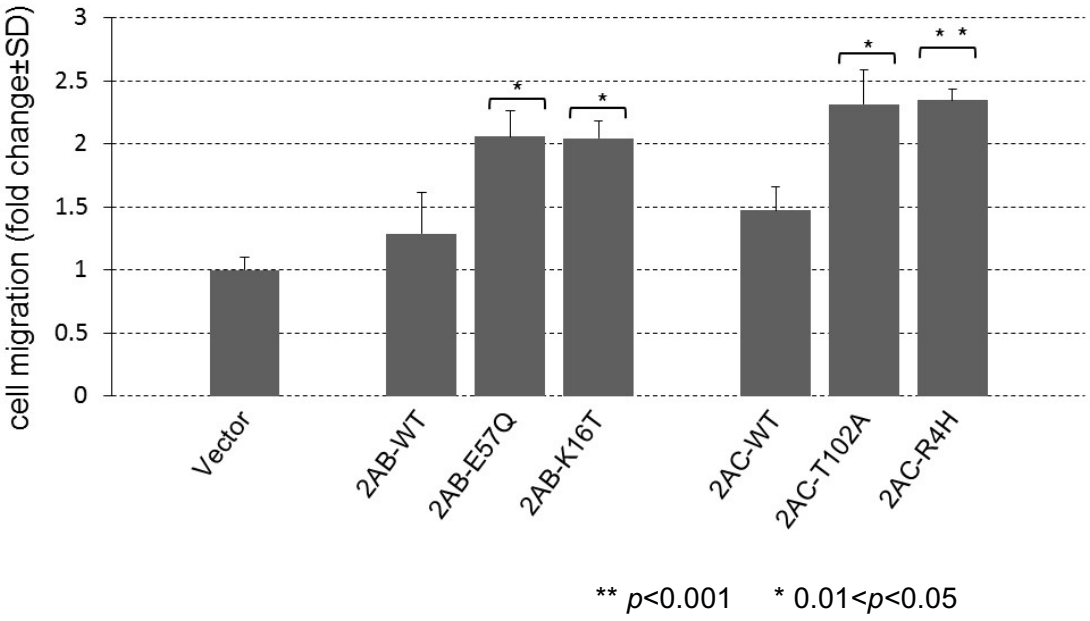


Figure S9. Migration and invasion assay of ARK-2 cells expressing E57Q, K16T, R4H and T102A mutations versus ARK-2 cells expressing empty vector or wild type HIST1H2AC or HIST1H2AB genes.

Primary USC cells (i.e., USC-ARK2), were transfected with vectors expressing wild type HIST1H2AB (H2AB-WT) or E57Q and K16T mutants as well as wild type HIST1H2AC (H2AC-WT) or R4H and T102A mutants. Migration and invasion assay results for stable cell lines expressing mutant histones are compared with empty vector control. A significant increase in cell migration and invasion of ARK-2 cells expressing E57Q, K16T, R4H and T102A mutations versus ARK-2 cells expressing empty vector or wild type HIST1H2AC or HIST1H2AB genes is shown (over 2-fold increase). Data shown are mean (±SD) of 3 experiments.

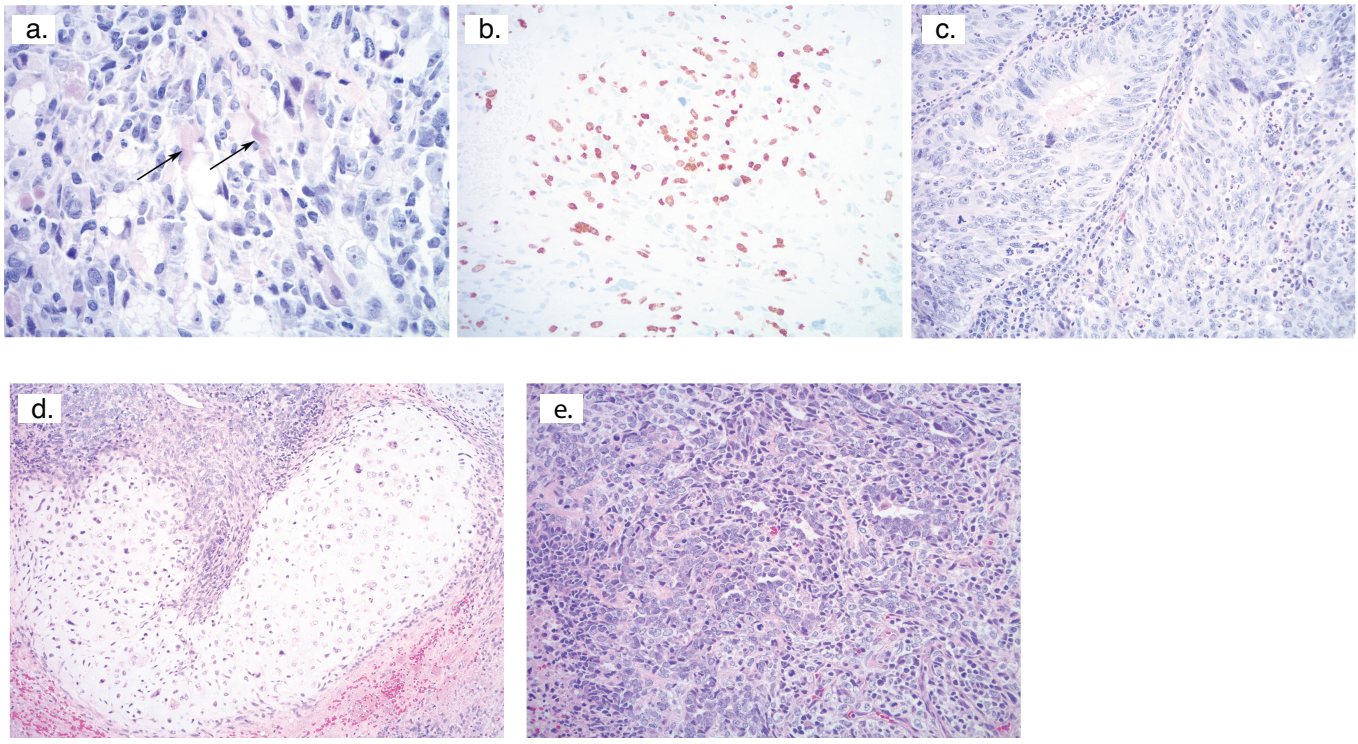
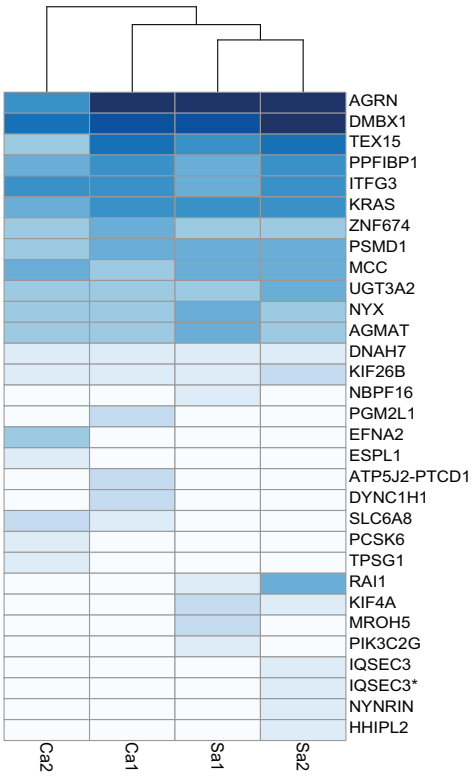
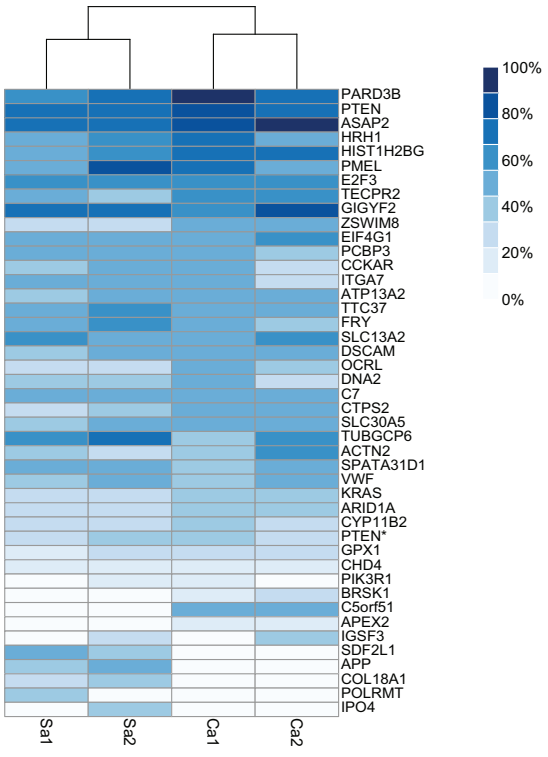


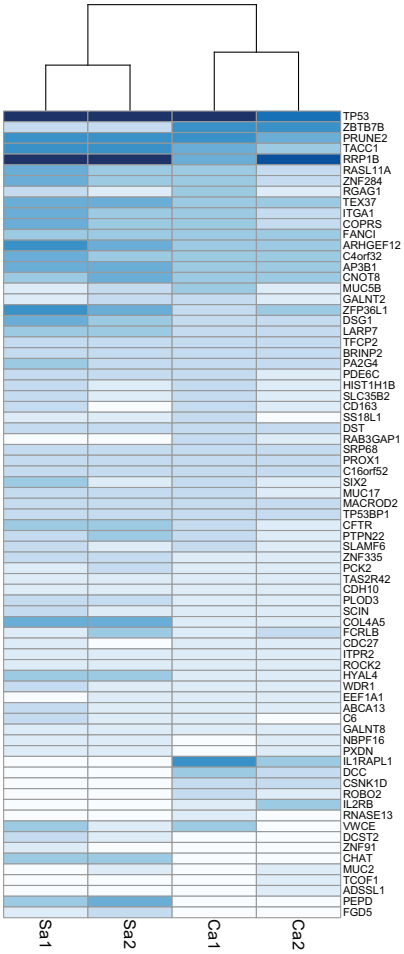
Figure S10. Photomicrographs of representative pathologic sections showing the histology of sarcomatous and carcinomatous areas chosen for WES and lineage analysis. Upper panel. Representative carcinosarcoma with heterologous rhabdomyosarcomatous and serous carcinomatous components. a: rhabdomyoblasts with cross striations (arrows) (400x original magnification). b: myogenin immunostain shows nuclear immunoreactivity in rhabdomyoblasts (200x original magnification). c: serous carcinoma component with glandular lumen formation and marked nuclear atypia (200x original magnification). Lower panel. Representative carcinosarcoma with heterologous chondrosarcomatous and serous carcinomatous components used for macro-dissection and lineage analysis. d: chondrosarcoma component (100x original magnification). e: serous carcinoma component with glandular lumen formation and marked nuclear atypia (200x original magnification).



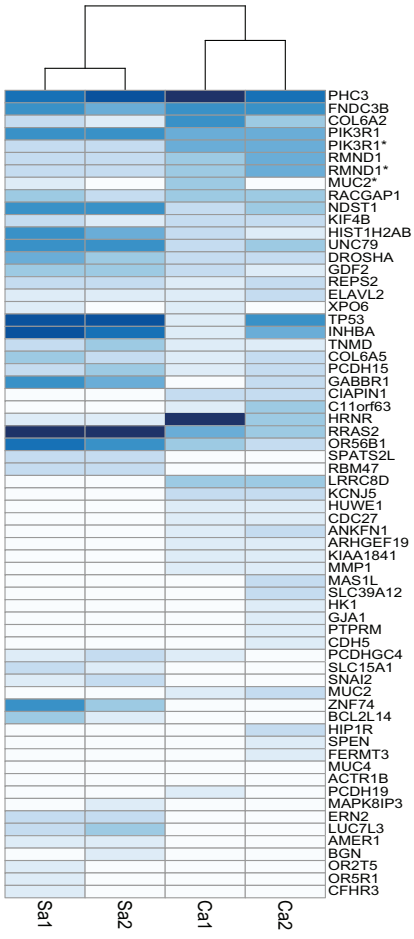
SAR27



SARB18



SAR33



SAR24

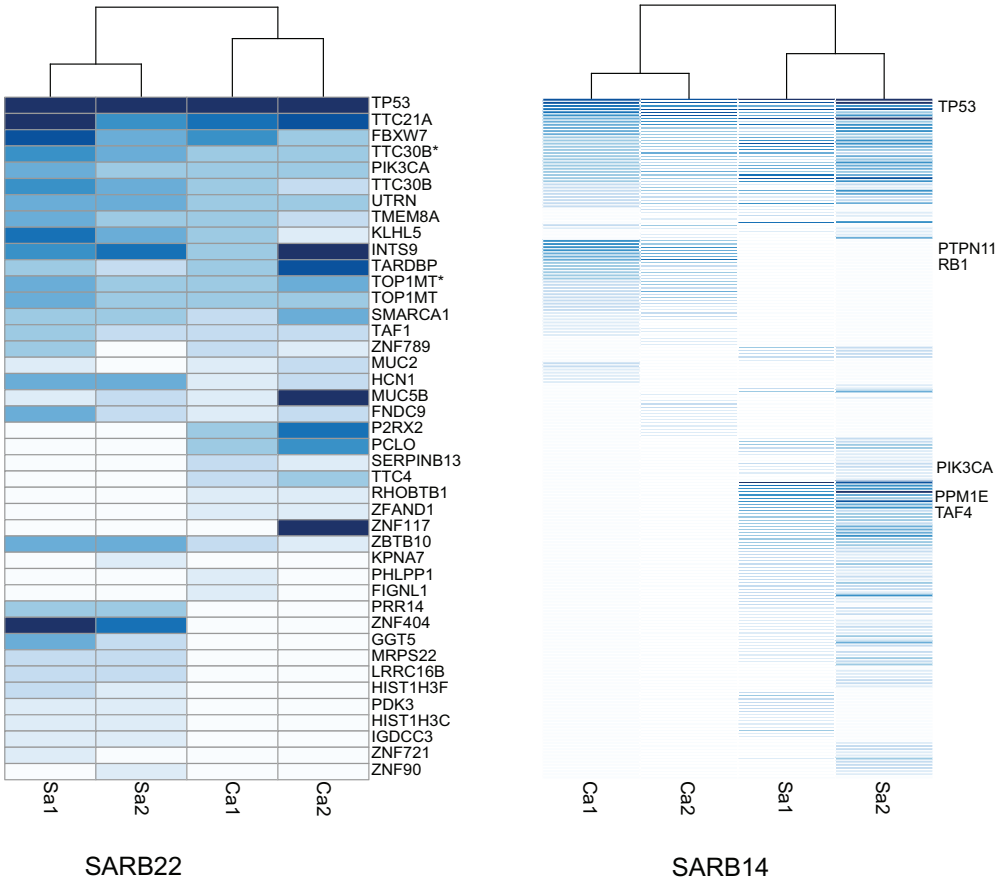


Figure S11. Hierarchical clustering of macro-dissected samples based on variant frequencies.

Each row in the heatmap represents one nonsynonymous mutations identified. Variant frequencies were somatic BAFs adjusted by purity and the values are indicated by the color of the cell. Ca1, Ca2: samples from the carcinoma areas; Sa1, Sa2: samples from the sarcoma areas. For sample SARB14, only mutations from cancer driver genes are labeled due to the limit of space.