Article | Open Access | Published: 20 January 2021

Short H2A histone variants are expressed in cancer

Guo-Liang Chew, Marie Bleakley, Robert K. Bradley, Harmit S. Malik, Steven Henikoff, Antoine Molaro & Jay Sarthy

Nature Communications **12**, Article number: 490 (2021) | Cite this article

https://doi.org/10.1038/s41467-020-20707-x

short H2A (HistoneDB)

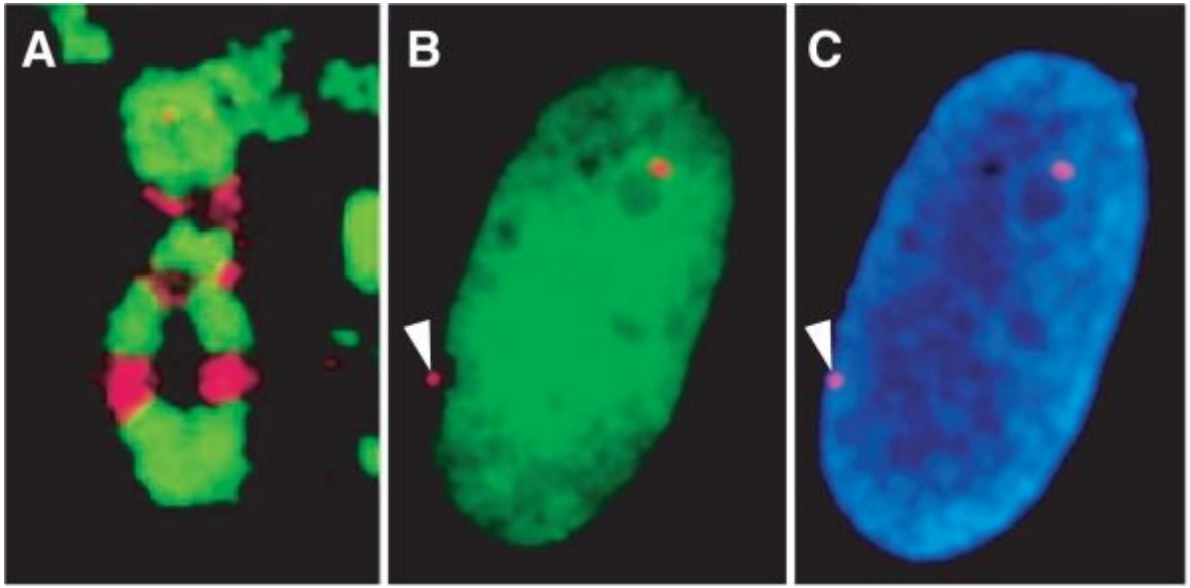
1. short_H2A is a class encompassing several histone H2A variants in placental (eutherian) mammals with shortened C-terminus expressed mainly during mammalian male germ cell development before the nearly complete replacement of histones by protamines in sperm nuclei.

2. The repertoires of short histone H2A variants vary extensively among eutherian mammals due to lineage-specific gains and losses. **Short H2A variants include H2A.B, H2A.L, H2A.P, H2A.Q, their genes are usually located on X chromosome and are intronless.** These four clades of eutherian mammal short H2A variants emerged from a single, well-supported monophyletic clade, confirming their common ancestry

3. Due to shortened docking domain and changes within the acidic patch nucleosomes incorporating **short H2As wrap less DNA (120-130 bp)** and form loosely packed chromatin.

4. There are few conserved residues in the histone fold domain of sH2As that distinguish them from each other, instead much of their specialization may stem from changes in the N- and C-terminal tails of these variants

H2A.B, previously known as "Barr body deficient" (H2A.Bbd)

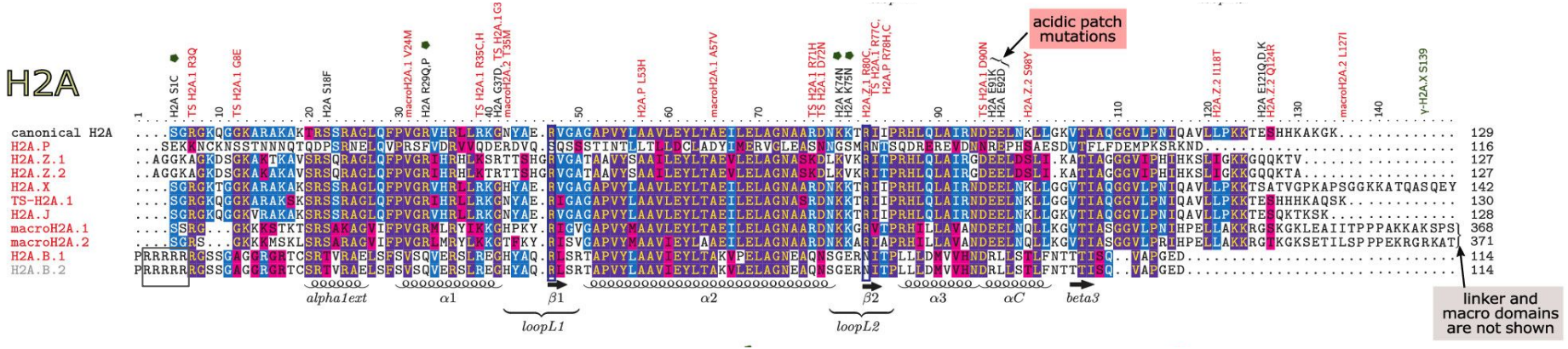


H2A variants and the inactive X chromosome of human females. (*A*) macroH2A (red) stains discrete regions of the inactive X chromosome that alternate with a marker for heterochromatin (histone H3K9me3). (*B*) H2A.B (green) is excluded from the inactive X chromosome (red dot with arrowhead pointing to it). (*C*) Same nucleus as in *B*, but stained with DAPI to show chromatin.

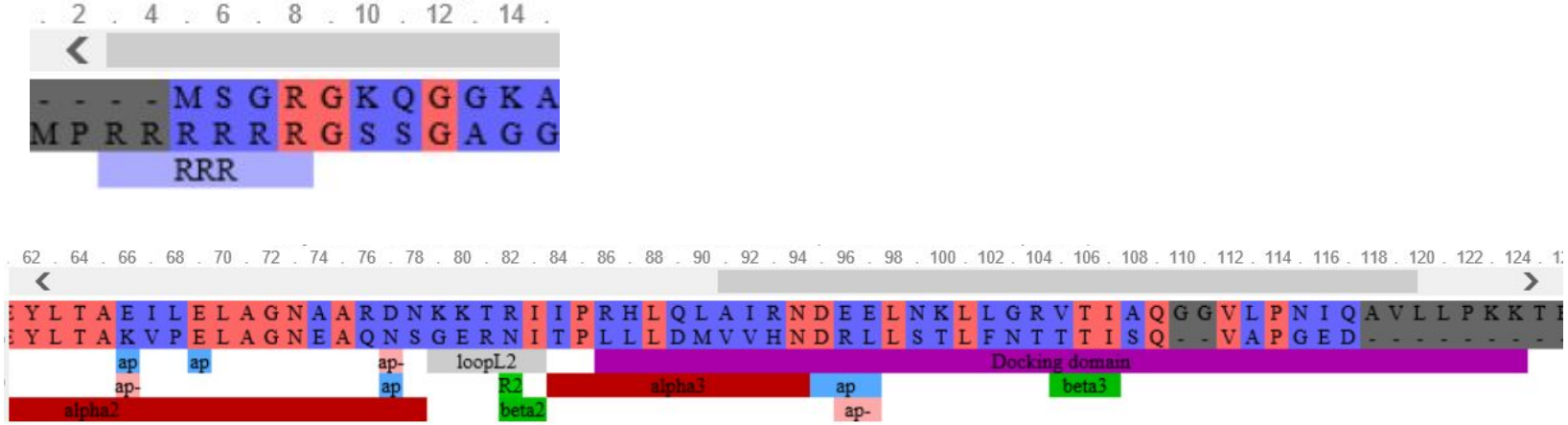
Henikoff S, Smith MM. Histone variants and epigenetics. *Cold Spring Harb Perspect Biol*. 2015;7(1):a019364. Published 2015 Jan 5. doi:10.1101/cshperspect.a019364

H2A.B sequence

- Sequence: Around 50% identity with the canonical H2A, has truncated docking domain, divergent histone fold domain, altered acidic patch, arginine rich N-terminus



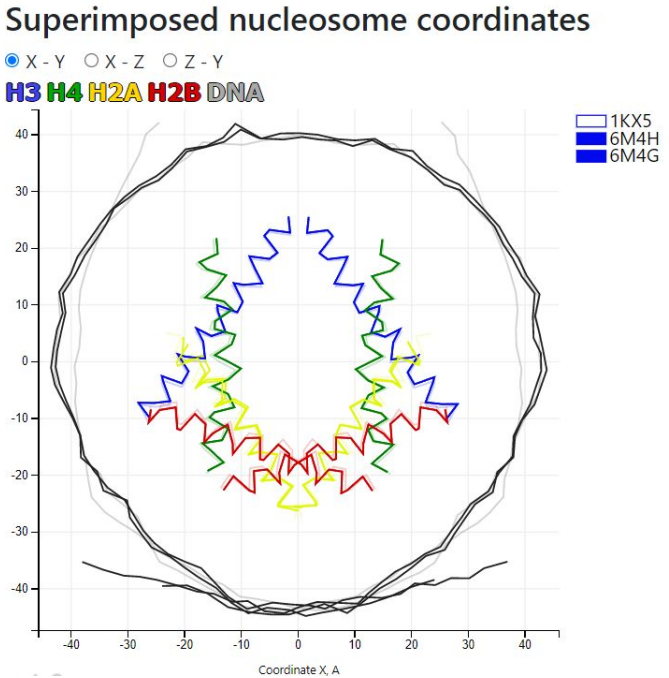
homo canonical_H2A vs H2A.B

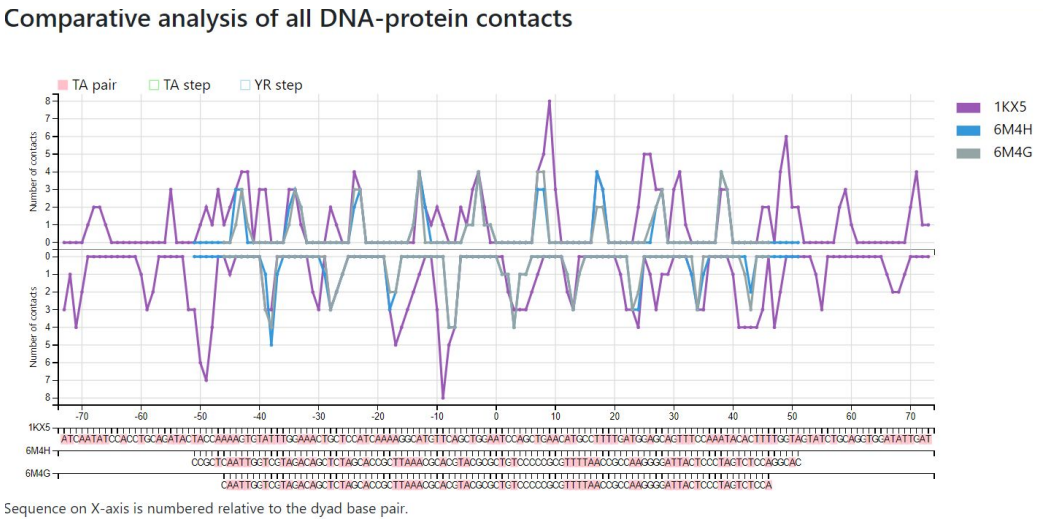


RRR	Stretch of arginines characteristic of H2A.B, at least in human
ap-	Loss of acidic patch residues
Docking domain	Docking domain locking H2A-H2B dimer on H3-H4 tetramer surface

Nucleosomes with H2A.B

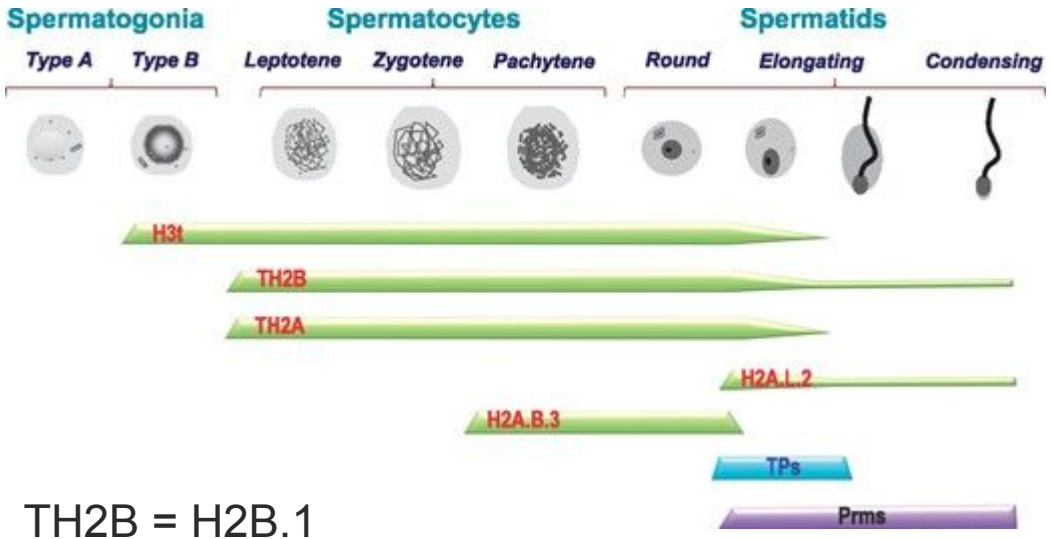
- Structural effects: H2A.B containing nucleosomes wrap less DNA (~120-130 bp instead of ~150 bp), form loosely packed chromatin.

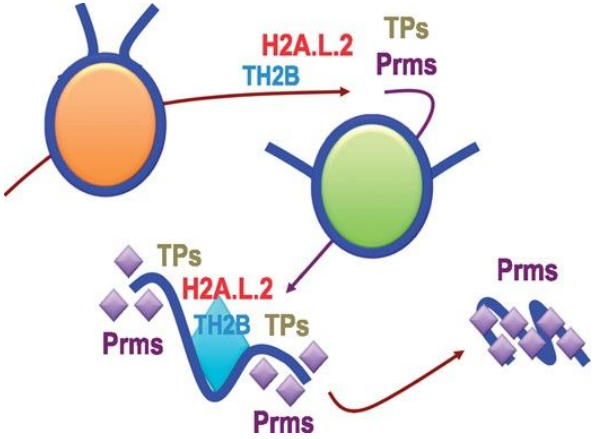




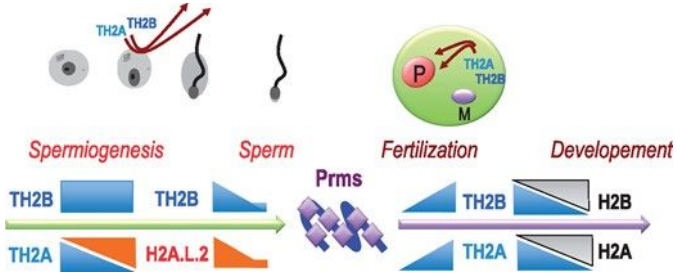
1. Knock-out: H2A.B knock-out mice are viable, subfertile and **display changes in splicing events**
2. Localization: H2A.B is expressed **during mammalian male germ cell development and in the brain**. Originally, H2A.B was characterized by its exclusion from the inactive X chromosome if overexpressed in female somatic cells. However, experiments in mouse testis revealed that H2A.B is in fact present on the inactive X chromosome.
3. **H2A.B can bind to RNA directly** in vitro and in vivo, and associates with mRNA at intron—exon boundaries.
4. Due to rapid evolution H2A.B function in different species may vary. For example, human H2A.B is retained during spermiogenesis, while is mouse it disappears and H2A.L is retained instead. Mouse H2A.B has additional negative residue in acidic patch, which is thought to increase its propensity to compact nucleosomal arrays relative to human H2A.B.

Spermatogenesis





TH2B = H2B.1
TH2A = H2A.1
H2A.L
H2A.B.3



Hoghoughi et al. Histone variants: essential actors in male genome programming, *The Journal of Biochemistry*, 2018, https://doi.org/10.1093/jb/mvx079

H2AB genes

There are five *X*-linked sH2A genes in humans: H2A.B.1.1 (*H2AFB2*), H2A.B.1.2 (*H2AFB3*), H2A.B.2 (*H2AFB1*), H2A.P (*HYPM*), and H2A.Q (unannotated)

All three H2AB genes are highly similar in sequence and encode a protein that is identical in the case of H2AB2 and H2AB3, with only one amino acid difference in the protein encoded by H2AB1.

In the literature these two proteins have sometimes been referred to as the variants H2A.B.1

https://histonedb.bioeng.ru/human/

H2A	H2A.B.1	H2AB1	474382	ENSG00000274183	ENST00000620016	NM_001017990	NP_001017990	115	29549088
H2A	H2A.B.2	H2AB2	474381	ENSG00000277858	ENST00000354514	NM_001017991	NP_001017991	115	29549088
H2A	H2A.B.2	H2AB3	83740	ENSG00000277745	ENST00000615853	NM_080720	NP_542451	115	29549088
H2A	H2A.P	H2AP	25763	ENSG00000187516	ENST00000341016	NM_012274	NP_036406	117	29549088
H2A		H2AQ1P	115482715	ENSG00000285989					
H2A	H2A.L	H2AL1Q	115482714	ENSG00000249467					29549088
H2A	H2A.L	H2AL3	115482686	ENSG00000229674					29549088
H2A		H2AL1MP	115482687	ENSG00000285944					

Background

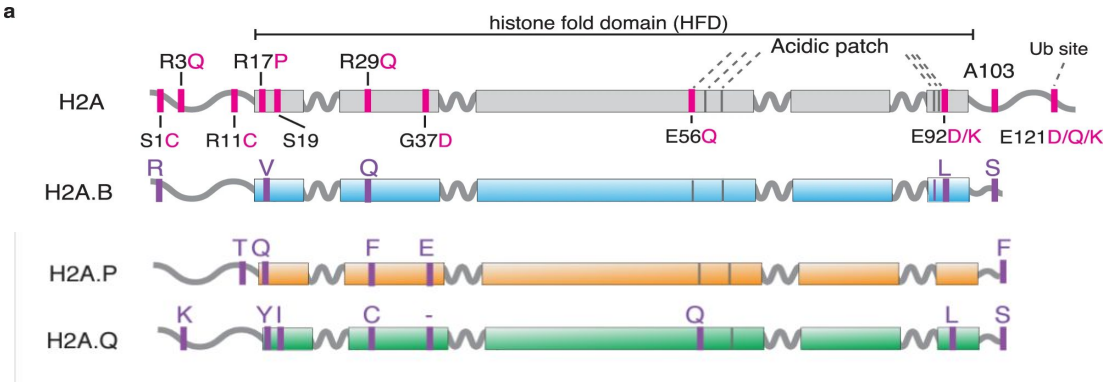
Additional evidence for a role for H2A.B in cancer comes from Hodgkin's lymphoma (HL), where H2A.B transcripts have been detected[23] and HL cells expressing H2A.B grow faster than H2A.B-negative cells[22].

https://link.springer.com/article/10.1007/s00262-012-1239-z

sH2As have evolved oncohistone features

many of the most common cancer-associated mutations in canonical H2A are already present in all wild-type sH2A sequences

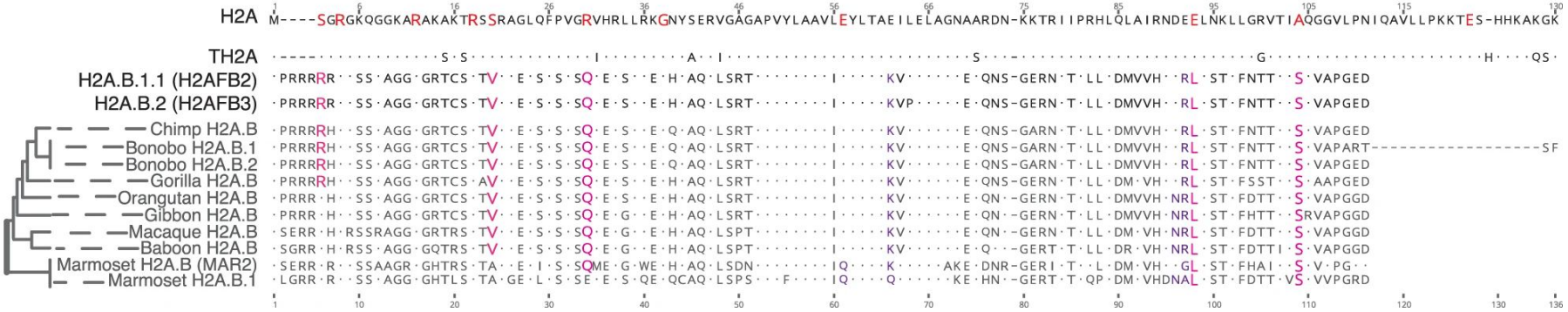
- R29Q/F substitutions that correspond to the second most frequent mutation in canonical H2A
- all wild-type sH2As have a C-terminal truncation that removes E121, the most common mutation in canonical H2A



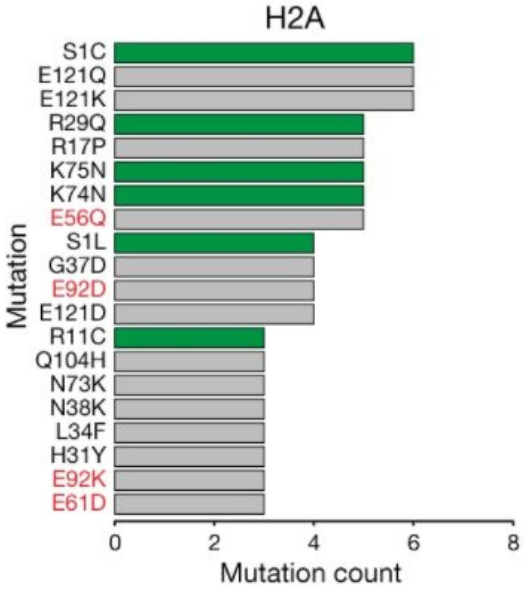
a Schematic of common oncomutations found in human core H2A and their status in H2A.B. Marked sites on core H2A show WT amino acid position followed by its most common cancer-specific substitution in TCGA (pink). Associated sites found in WT short H2As are shown in purple.

sH2As have evolved oncohistone features

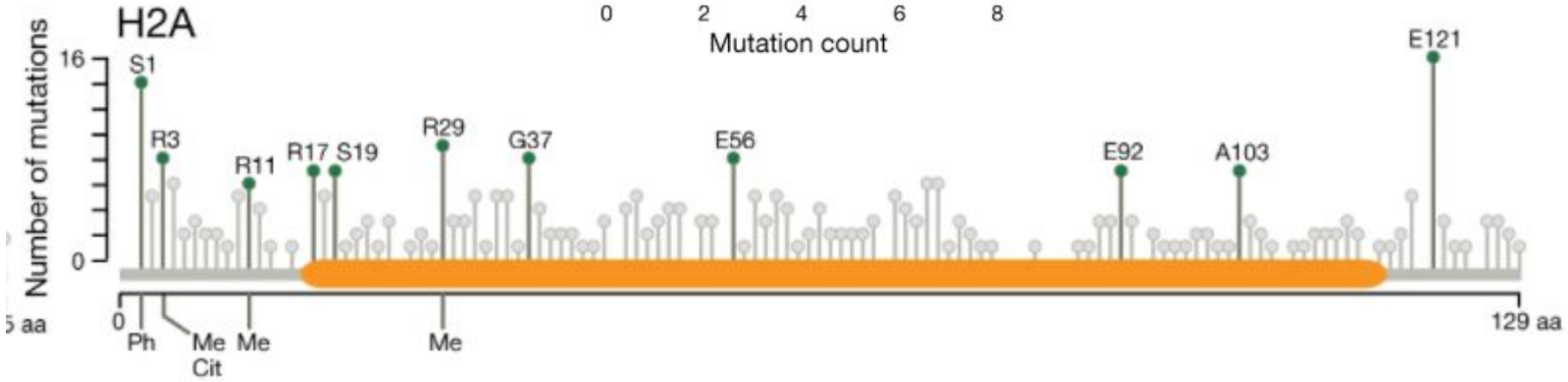
Phylogenetic analyses in primates showed that despite their rapid evolution, these oncohistone-like changes are highly conserved. This conservation implies functional consequences as many of these residues are critical contact points for histone-DNA or histone-histone interactions. These data show that sH2As contain oncohistone features similar to canonical H2A mutations in cancers.



b Protein alignment of core H2A, testis-specific H2A (TH2A), and H2A.B paralogs from Human and representative primates. Substitutions corresponding to oncohistone mutations in H2A (see Fig. 1.) are shown in pink.



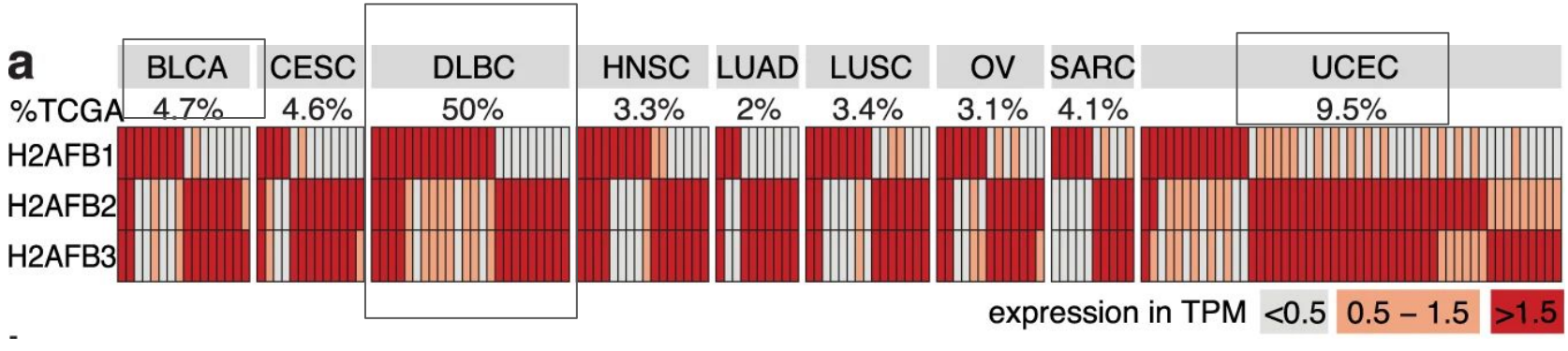
- tumour mutation burden (TMB) of ≤10 mutations per megabase (Mb)



Nacev, *et al.* The expanding landscape of 'oncohistone' mutations in human cancers. *Nature* 567, 473–478 (2019). https://doi.org/10.1038/s41586-019-1038-1

H2A.Bs are reactivated in a broad array of cancers

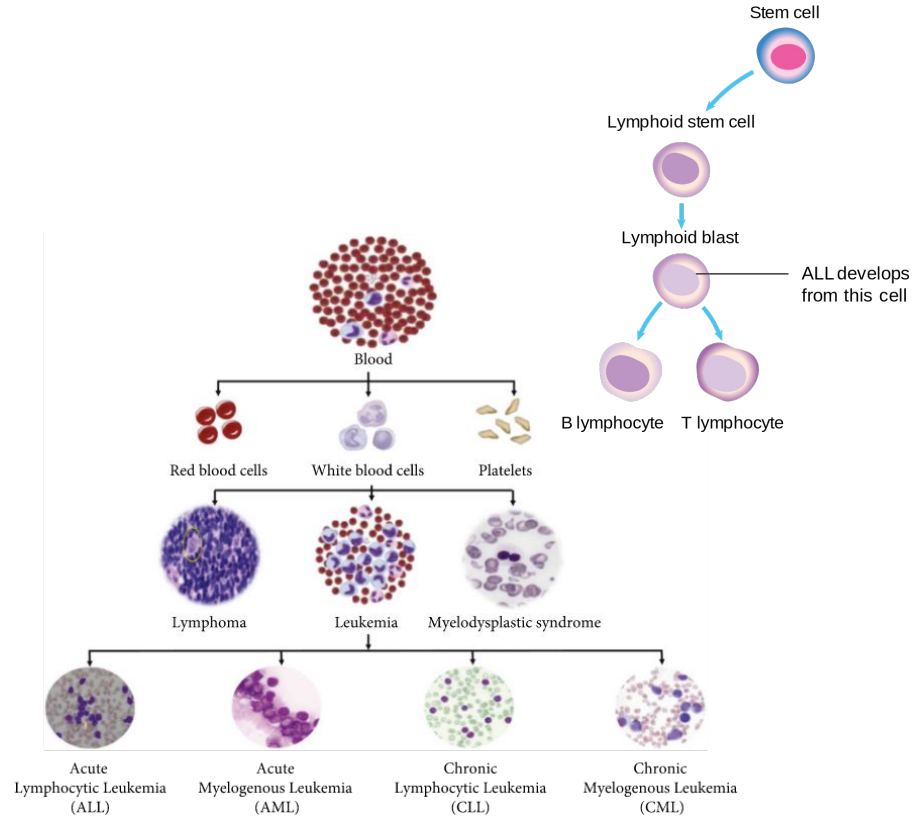
- TCGA: H2A.B paralogs are activated (at a threshold of >1.5 transcripts per million (TPM)) in numerous individual tumors across cancer types, but never in adjacent normal tissue, and very rarely (<1.5%) in non-testes tissue samples from the Genotype-Tissue Expression database

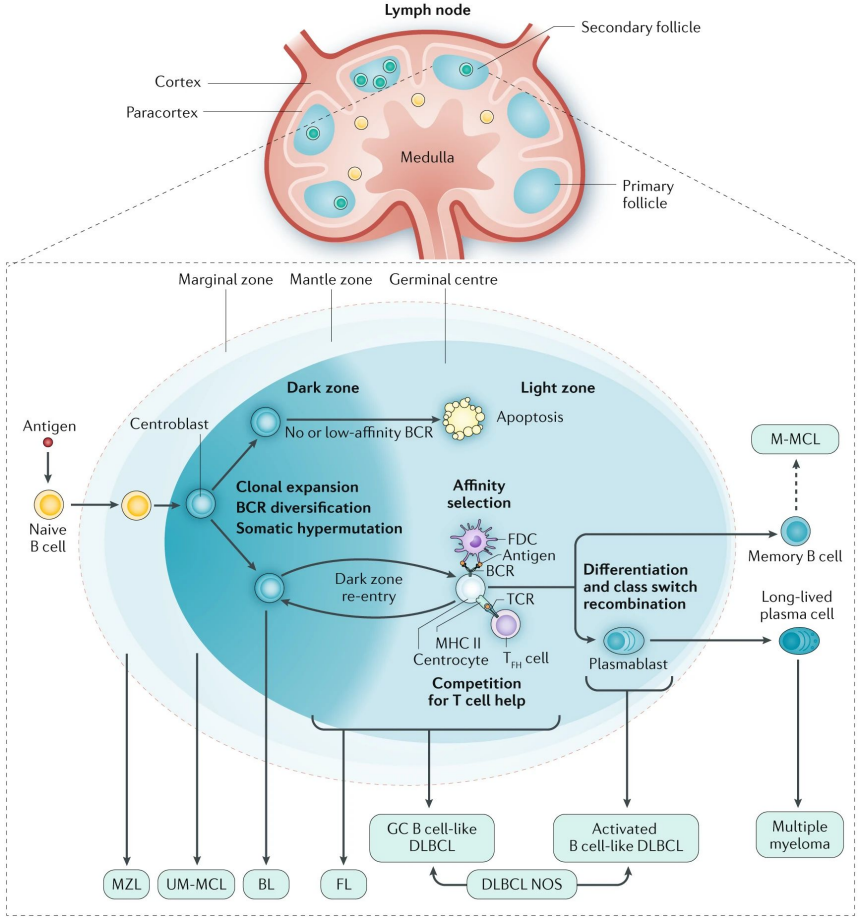


https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations

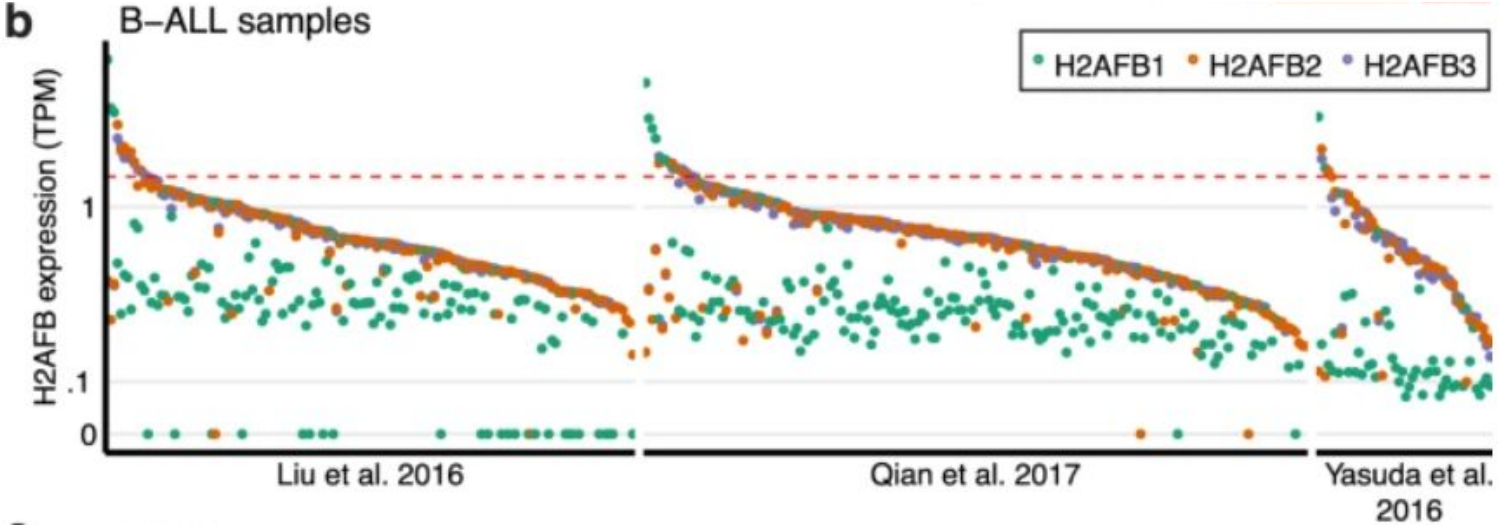
ALL - Острый лимфобластный лейкоз (пролиферация лимфобластов)

DLBCL - Диффузная В-крупноклеточная лимфома





- после DLBCL решили проанализировать и другие data sets from other lymphoid lineage-derived, low mutation cancers for aberrant H2A.B expression
-
- We queried four separate B-acute lymphoblastic leukemia (B-ALL) data sets and found 6–7% of specimens with H2A.B-encoding transcripts at >1.5 TPM in three of the data sets and 13% in the fourth



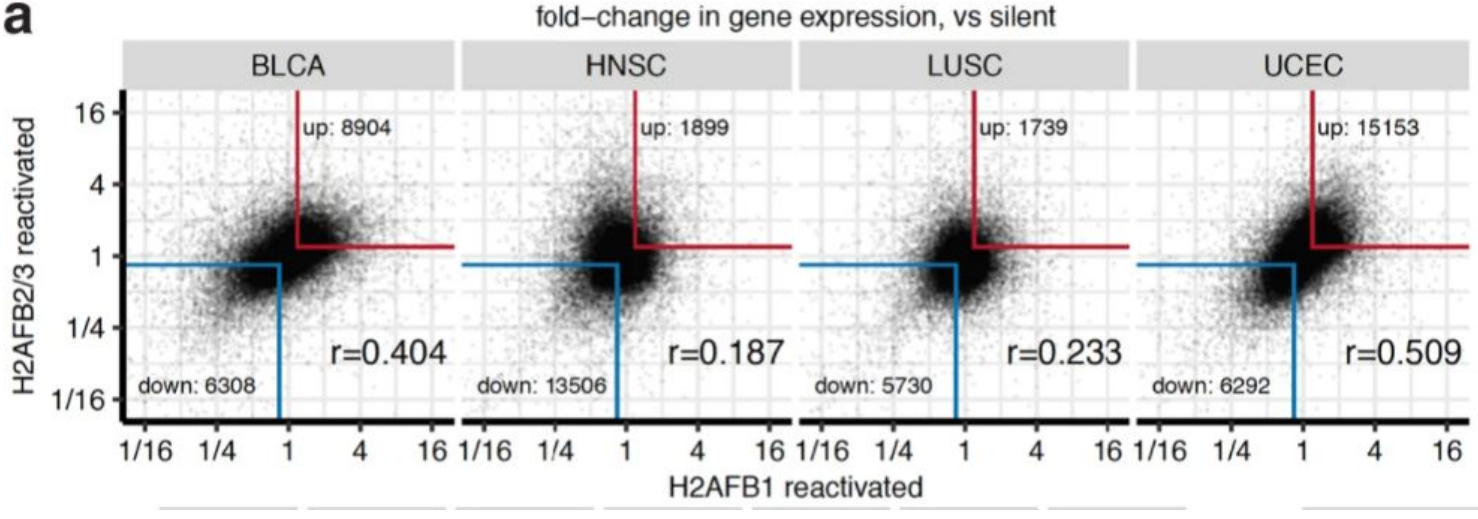
18, 5, and 1 samples

Although many tumors reactivate H2AFB1 alone, most tumors that express H2AFB2 also express H2AFB3 This finding may result from transcriptional co-regulation due to their genomic proximity or inability to distinguish these near-identical paralogs by short-read mapping

These results are consistent with our findings in the TCGA data set, where median H2A.B expression for the 232 H2A.B-positive samples is ~3 TPM, corresponding to 49th percentile of all expressed genes. **This level of expression is more likely the result of local, specific activation of individual H2AFB paralogues** than recurrent amplifications or broader X-chromosome dysfunction.

H2A.Bs are associated with cancer-specific, rather than pan-cancer gene expression programs

- сравнили экспрессию других генов между H2AB1 и H2AB2/H2AB3 - схожий паттерн ап и даун регуляции
- We found 146 genes were upregulated and 90 downregulated across H2A.B-positive cancers



Cancer-Testis Antigens

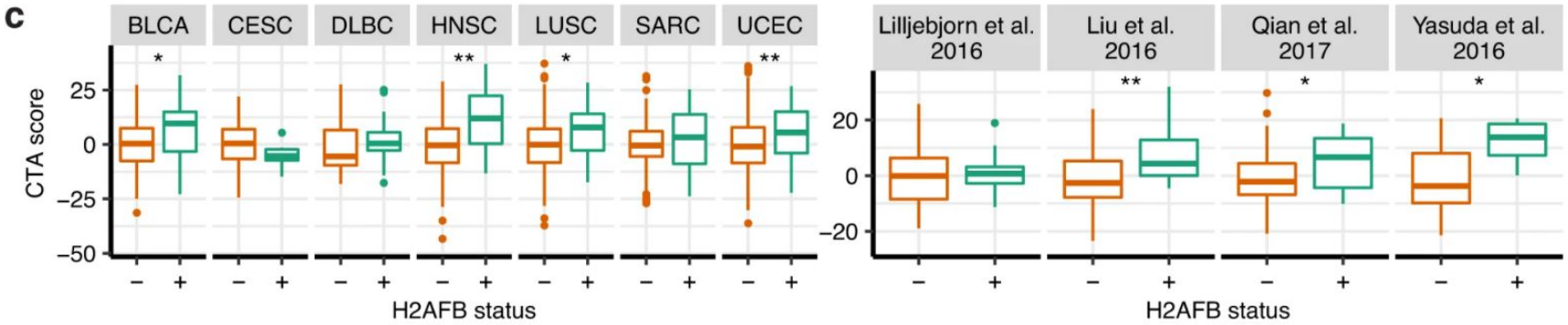
Cancer testis antigens (CTA) are a large family of tumor‐associated antigens expressed in human tumors of different histological origin, but not in normal tissues except for testis and placenta.

This tumor‐restricted pattern of expression, together with their strong in vivo immunogenicity, identified CTA as ideal targets for tumor‐specific immunotherapeutic approaches, and prompted the development of several clinical trials of CTA‐based vaccine therapy. Driven by this practical clinical interest, a more detailed characterization of CTA biology has been recently undertaken. So far, at least 70 families of CTA, globally accounting for about 140 members, have been identified. **Most of these CTA are expressed during spermatogenesis**, but their function is still largely unknown.

CTA can be divided in those that are encoded on the X chromosome, the X‐CTA genes, and those that are not, the non‐X‐CTA genes . It has been estimated that **10% of genes on the X chromosome belong to X‐CTA families**. The X‐CTA genes represent more than half of all CTA and often constitute multigene families organized in well‐defined clusters along the X chromosome, where the different members are arranged into complex direct and inverted repeats.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5528287/

- We noted that 12/146 of the commonly upregulated genes are Cancer-Testis Antigens. As *H2AFB1* was previously shown to be co-expressed with a subset of CTAs in HL[23], we determined whether H2A.B-reactivated cancers are generally associated with CTA upregulation. We summarized the expression of individual CTAs into a composite "CTA score" for each tumor and compared scores between H2A.B-reactivated and silent samples
- These data indicate that H2A.B expression is associated with CTA expression in several cancer types.



Supplementary Table 2: Number of H2AFB1/2/3 reactivated and silent samples in TCGA and B-ALL datasets.

Numbers of cancer samples in TCGA and B-ALL datasets used for gene expression and cancer testes antigen analyses (Fig 3a, b).

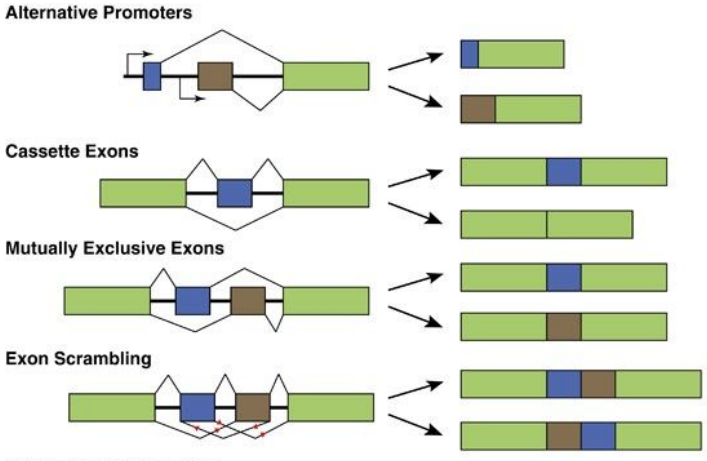
Cancer dataset	Reactivated	Silent
BLCA	16	280
CESC	13	231
DLBC	24	17
HNSC	16	442
LUAD	10	476
LUSC	15	409
SARC	10	199
UCEC	51	280
Lilljebjorn et al. 2016	26	69
Liu et al. 2016	11	74
Qian et al. 2017	13	77
Yasuda et al. 2016	4	24

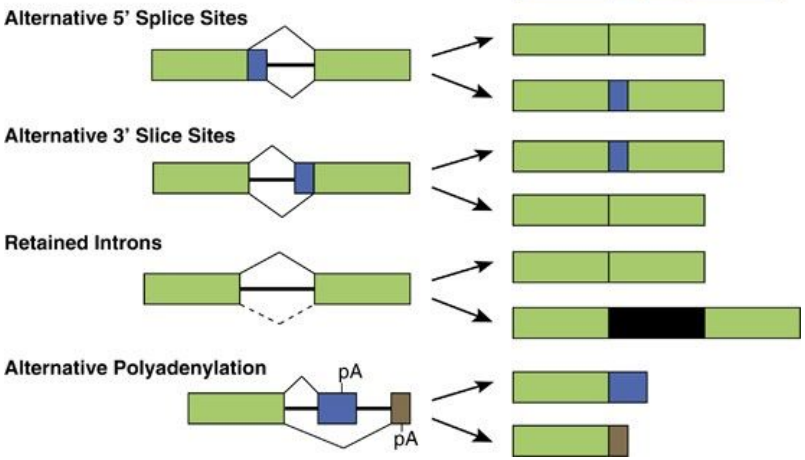
Supplementary Table 1. Expression of H2AFB1/2/3 in GTEx normal tissue

Numbers and percentages of samples with H2AFB1, H2AFB2, or H2AFB3 reactivation (expression > 1.5 TPM) from the GTEx dataset of normal tissues, by various tissue types.

GTEx tissue	Number with H2AFB1/2/3 reactivated	Total number	Percentage reactivated
Other	0	7900	0.00
Brain	4	2541	0.16
Colon	1	571	0.18
Thyroid	1	508	0.20
Blood Vessel	3	1320	0.23
Esophagus	6	1364	0.44
Prostate	1	160	0.63
Uterus	1	128	0.78
Spleen	3	202	1.49
Blood	14	929	1.51
Testis	152	252	60.32

Common mechanisms of alternative splicing



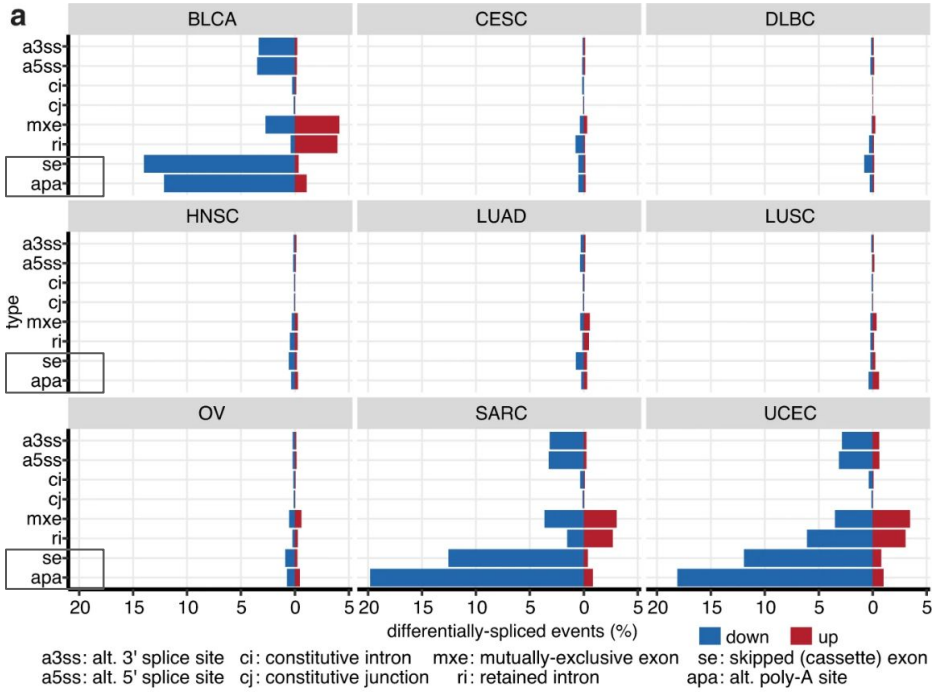


a3ss: alt. 3' splice site ci: constitutive intron mxe: mutually-exclusive exon se: skipped (cassette) exon
a5ss: alt. 5' splice site cj: constitutive junction ri: retained intron apa: alt. poly-A site

https://doi.org/10.1038/onc.2013.570

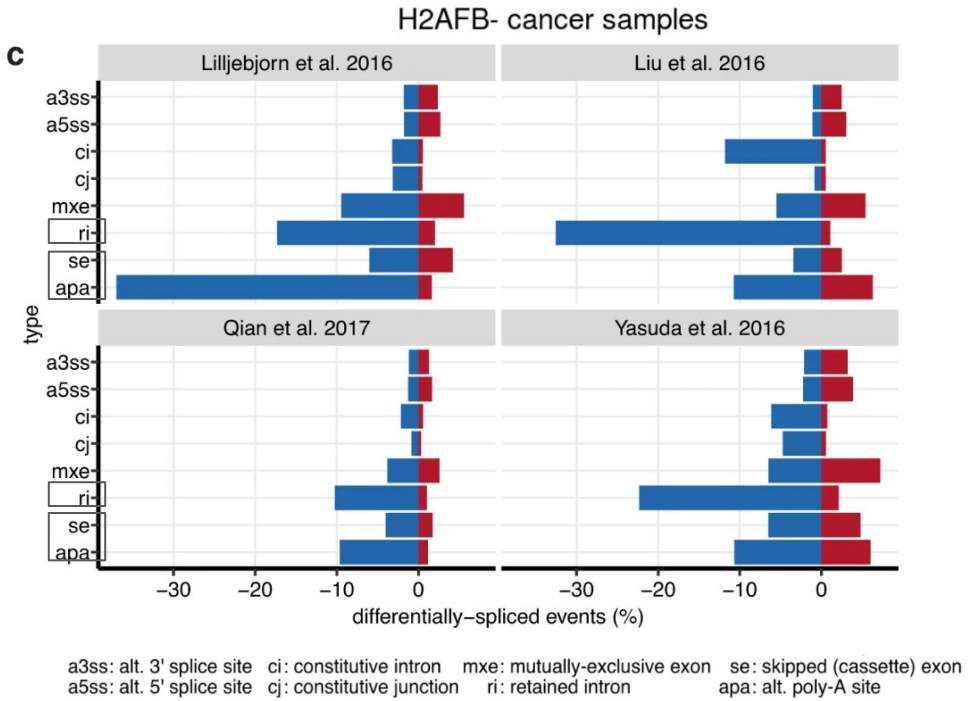
H2A.B-expressing cancers have distinct splicing patterns

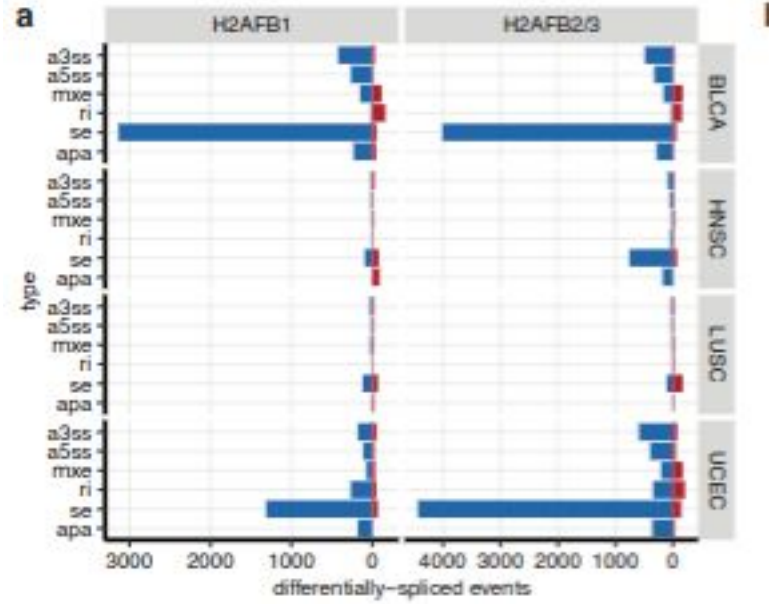
- H2A.B has been shown to directly bind RNA and interacts with splicing factors and H2A.B expression impacts alternative splicing patterns
- We found that H2A.B expression is associated with **reduced utilization of alternative "cassette exons" (se) and proximal alternative 3′ polyadenylation (APA) sites**
- These patterns are not H2A.B paralogue-specific, as similar patterns were observed in specimens expressing either H2AFB1 or H2AFB2/3



Bladder Urothelial Carcinoma, Sarcoma, Uterine Corpus Endometrial Carcinoma

- B-ALLs are not associated with mutations in splicing factors and global splicing dysregulation is not thought to be a major driver of these leukemias.
- When we compared splicing patterns in the H2A.B-reactivated and silent samples within each data set, we observed **aberrant splicing at a scale similar to that seen in H2A.B-positive TCGA cancers, with reductions in alternative exon and APA (**proximal alternative 3′ polyadenylation**) usage**.
- However, the most notable feature is a consistent decrease **in retained introns "ri"** in all four data sets.
- We conclude that H2A.B expression is associated with splicing dysfunction, with some features common among many cancers while others occur in a context-specific manner.




a
H2AFB1
H2AFB2/3
a3ss
a5ss
mxe
ri
se
apa
BLCA
HNSC
LUSC
UCEC
type
3000
2000
1000
0
4000
3000
2000
1000
0
differentially−spliced events

Discussion

- Nucleosome-destabilizing features are important for sH2As' roles in normal testis physiology but result in oncohistone properties when expressed out of context.
- H2A.B expression occurs in many common cancers. The diversity of H2A.B-expressing cancer types suggests that pathological histone dynamics play a more significant role in neoplasia than previously appreciated.
- H2A.B impacts different genes in different cancers.
- As nucleosomes protect DNA from inappropriate transcription factor binding,
    - nucleosome instability may allow oncogenic TFs access to different regulatory elements depending on cancer type[2,39].
    - Nucleosome destabilization also hastens RNA pol II elongation, which in turn reduces transcription-coupled splicing efficiency[40]. Alternative exons and proximal polyadenylation sequences are preferentially impacted by inefficient splicing owing to their weaker splice signals, resulting in a splicing phenotype similar to those observed in several H2A.B-positive cancers[40]. As some alternative exons promote mRNA degradation by targeting them for nonsense-mediated decay, even modest reductions in alternative splicing can increase oncogene expression[41]. H2A.B may operate at the nexus of several processes that cooperate to drive oncogenesis.

Whether potential similarities between histone mutant cancers and H2A.B-expressing cancers extend to prognoses and vulnerabilities merits further investigation, particularly in the context of DLBCL where larger data sets are needed to dissect these relationships.

Several cell lines show sensitivity to *H2AFB1*-gRNAs in the Sanger Cancer Dependency Map, with lymphoma-derived cell lines SU-DHL-8 and IM9 being among the most sensitive to *H2AFB1* disruption[45].

Better characterization of histone mutations and H2A.B expression across cancer cell lines is also needed in order to probe for similarities between H2A.B-expressing cancers and histone mutant cancers.

Finally, sH2A-derived short peptides that bind HLA molecules (Supplementary Data 6) may be useful immunotherapy targets, and global splicing dysregulation can also generate highly immunogenic neoantigens[46]. Thus, **our discovery of sH2A-expressing cancers may open new avenues of study and treatment for hundreds of thousands of cancer cases worldwide.**