Histone H3F3A is encoded by different transcripts with distinct 5´-UTRs in both human and zebrafish

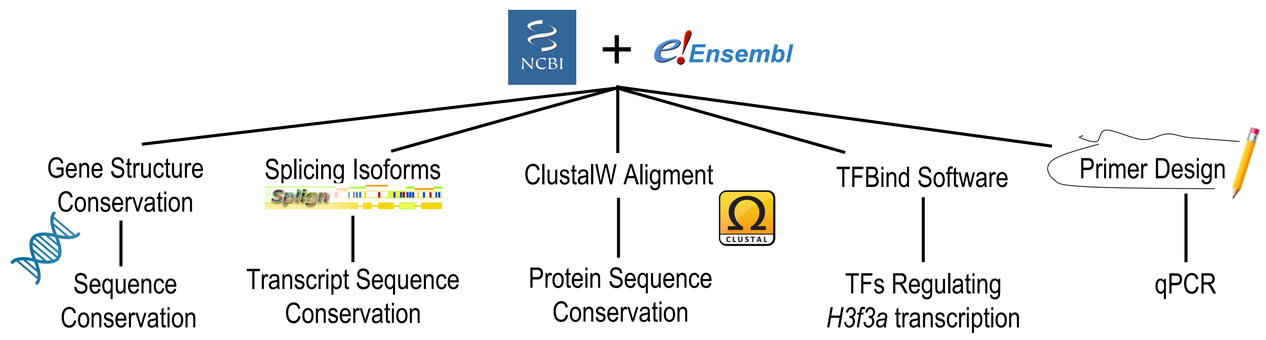
Komazec B. [1*], Cerqueira R. [1*], Horta S. [1*], Major SD. [1*], Rossetto V. [1*], Nunes A. [1*], Varela T. [3], Cancela ML. [2,3], Conceição N. [2,3]

[1]Master in Molecular and Microbial Biology, University of Algarve, Portugal
[2]Centre of Marine Sciences (CCMAR), University of Algarve, Portugal
[3]Dept of Biomedical Sciences and Medicine and Algarve Biomedical Centre, University of Algarve, Portugal
*All authors contributed equally to this work

Introduction

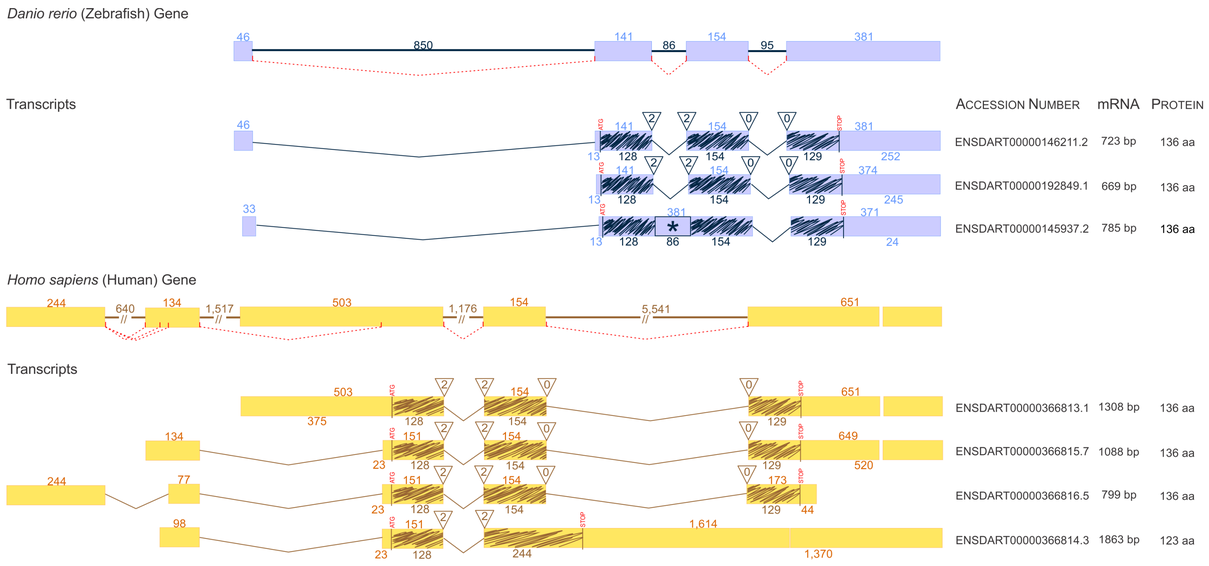
Histones are basic nuclear proteins that associate with DNA and are responsible for establishing the nucleosome structure. Five classes of histone genes were described: H1, H2A, H2B, H3, and H4. H3.3 is a replacement histone subtype expressed during cell cycle and overexpressed in transcriptionally active regions, promoter regions, intergenic or intragenic regulatory elements, and heterochromatin regions, helping to protect/maintain the integrity of the genome. The H3.3 histone is encoded by two genes called *H3.3A* (*H3F3A)* and *H3.3B* (*H3F3B*), which encode an identical protein despite differences in gene and promoter sequences. Somatic mutations in these genes were reported in human cancers such as chondroblastoma, osteosarcoma, giant cell tumor of bone, epithelial ovarian cancer, and pediatric gliomas. In both human and zebrafish, these are single copy genes. Because zebrafish has been validated as model to study human diseases, we investigated the genomic organization and the possible occurrence of alternative splicing of the zebrafish ortholog of human *H3F3A*.

Material & Methods



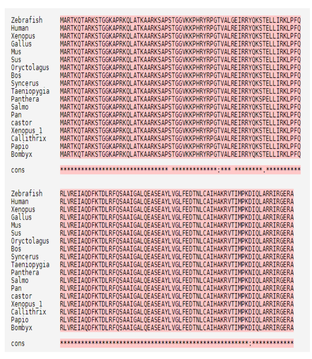
Results

Alternatively spliced variants



Representation and comparison between *Danio rerio* (ZEBRAFISH) and *Homo sapiens* (HUMAN) *H3F3A* gene and their respective transcripts. Genes are represented with all putative exons. Different variants of the alternative splicing are represented in both *D. rerio* and *H. sapiens* genes (red dashed lines) and each transcripts included the respective START (ATG) and STOP codon with the non-coding (solid colour) and coding (draw colour) sequence. Above exons and introns are represented the size (bp) at scale and the intron phase. Below exons are represented the size (bp) of the coding region. There are three zebrafish and four human transcripts of the *H3F3A* gene and each transcript information are disposed at the right side, such as, the ensemble number, transcript size (bp) and protein size (aa). The third zebrafish transcript include an intron retention (*).

Protein homology



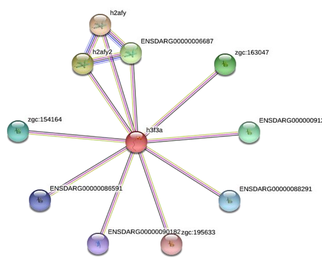
	Bmo	Pan	Cja	Xtro	Cca	Ptr	Ssa	Ptia	Tgu	Sca	Bta	Ocu	Ssc	Mmus	Gga	Hsa	Dre
Bmo	967																
Pan	100	411															
Cja	99,26	99,26	1394														
Xtro	100	100	99,26	543													
Cca	100	100	99,26	100	3384												
Ptr	100	100	99,26	100	100	1080											
Ssa	100	100	99,26	100	100	100	949										
Ptia	97,79	97,79	97,06	97,79	97,79	97,79	97,79	510									
Tgu	100	100	99,26	100	100	100	100	97,79	1108								
Sca	100	100	99,26	100	100	100	100	97,79	100	455							
Bta	100	100	99,26	100	100	100	100	97,79	100	100	1144						
Ocu	100	100	99,26	100	100	100	100	97,79	100	100	100	930					
Ssc	100	100	99,26	100	100	100	100	97,79	100	100	100	100	479				
Mmus	100	100	99,26	100	100	100	100	97,79	100	100	100	100	100	1086			
Gga	100	100	99,26	100	100	100	100	97,79	100	100	100	100	100	100	961		
Hsa	100	100	99,26	100	100	100	100	97,79	100	100	100	100	100	100	100	1077	
Dre	99,26	99,26	98,53	99,26	99,26	99,26	99,26	97,06	99,26	99,26	99,26	99,26	99,26	99,26	99,26	99,26	749

% similarity
97 - 98
98 - 99
99 - 100
Protein size (aa)

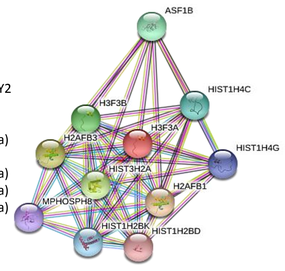
Abbreviation of species	
Bmo	*Bombyx mori*
Pan	*Papio anubis*
Cja	*Callithrix jacchus*
Xtro	*Xenopus tropicalis*
Cca	*castor canadensis*
Ptr	*Pan troglodytes*
Ssa	*Salmo salar*
Ptia	*Panthera tigris alta*
Tgu	*Taeniopygia guttata*
Sca	*Syncerus caffer*
Bta	*Bos taurus*
Ocu	*Oryctolagus cuniculu*
Ssc	*Sus scrofa*
Mmus	*Mus musculus*
Gga	*Gallus gallus*
Hsa	*Homo Sapiens*
Dre	*Danio rerio*

H3F3a protein conservation throughout different species and pairwise percent identities of the protein sequences. Protein sequences acquired from NCBI and identity values obtained from M-Coffee alignment. **Alignment of the protein sequence of the 17 different species studied resulted in a similarity of 99.6%.**

Protein connections



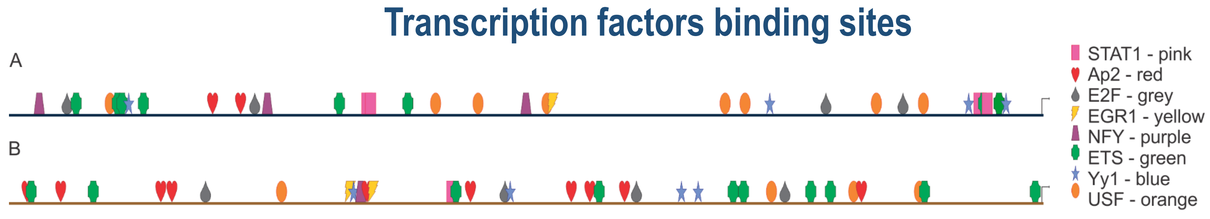
Danio rerio
- H2afy - H2A histone family, member Y;
- H2afy2 - H2A histone family, member Y2
- ENSDARG00000006687 - H2A histone family, member Y2 (367 aa)
- zgc:163047 - Histone cluster 2, H2ab (127 aa)
- ENSDARG00000091205 - Histone cluster 2, H2ab (128 aa)
- zgc:154164 - Histone cluster 2, H2ab (143 aa)
- ENSDARG00000088291 - Histone cluster 2, H2ab (135 aa)
- ENSDARG00000086591 - Histone cluster 2, H2ab (130 aa)
- ENSDARG00000090182 - Histone cluster 2, H2ab (128 aa)
- zgc:195633 - Histone cluster 2, H2ab (138 aa)



Homo sapiens
- H2AFB1 – H2A histone family, member B1
- H2AFB3 - H2A histone family, member B3
- HIST3H2A - Histone cluster 3, H2a
- H3F3B - H3 histone, family 3B (H3.3B)
- ASF1B - ASF1 anti-silencing function 1 homolog B (*S. cerevisiae*)
- HIST1H4C - - Histone cluster 1, H4c
- HIST1H2BK - - Histone cluster 1, H2bk
- HIST1H4G - Histone cluster 1, H4g
- MPHOSPH8 - M-phase phosphoprotein 8
- HIST1H2BD - Histone cluster 1, H2bd

***Homo sapiens* and *Danio rerio* H3F3A interaction network.** Both *Homo sapiens* and *Danio rerio* are linked to ten proteins none of which are identical. Data obtained from STRING CONSORTIUM 2018 ©.

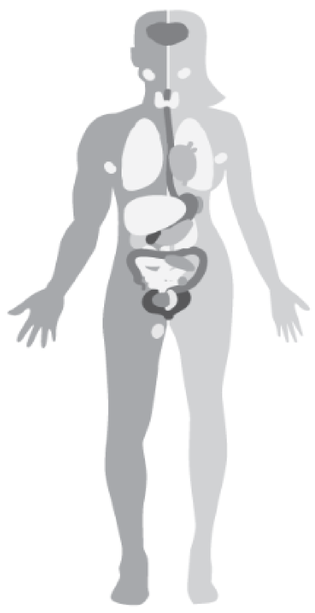
Transcription factors binding sites

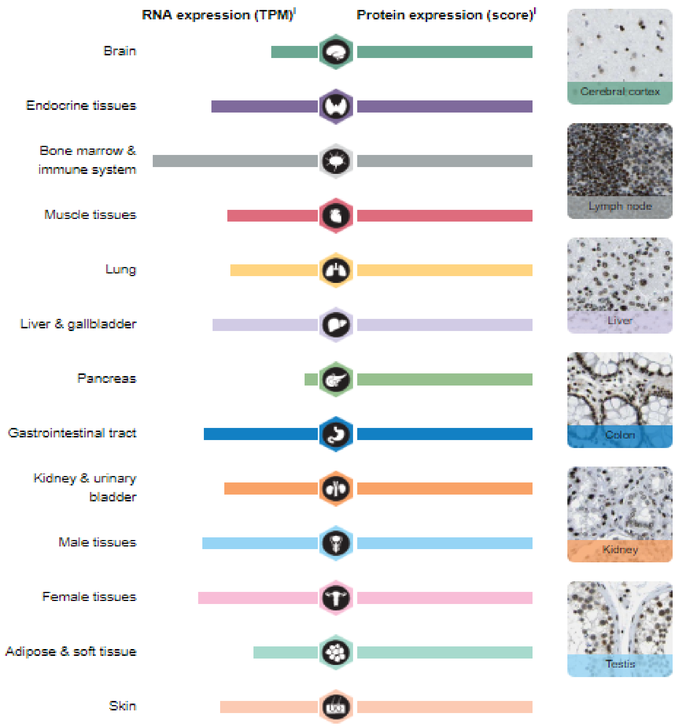


Comparative H3F3A promoter analysis between (A) *Danio rerio* and (B) *Homo sapiens*. Several putative transcription factors binding motifs found by Genome Browser Gateway and identified with TFBIND tool. The numbering is related to the translation start site. In zebrafish, minimal score for transcription factors AP2, E2F, EGR1, NFY, ETS, YY1 and USF was 0.85 and for STAT1 was 0.75. In human, minimal score for transcription factors E2F, EGR1, NFY, ETS, YY1 and USF was 0.85, for STAT1 was 0.80 and for AP2 as 0.90.

H3F3A expression

➢ ***Homo sapiens***





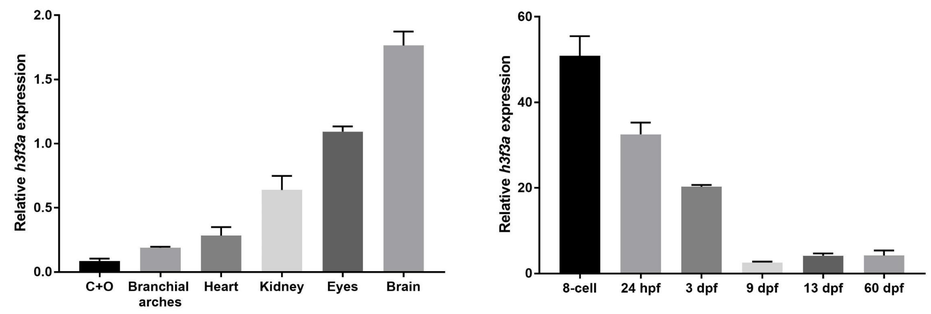
RNA and protein expression of *H3F3A* in all tissue examples analyzed from The Human Protein Atlas. RNA-seq results are reported as number of transcripts per million (TPM). Images of tissues provide a visual summary of the protein expression profile.



***H3F3A* expression in human tissues.** Forty- five human tissues were examined for *H3F3A* mRNA expression by mining human EST databases in NCBI-UniGene site. Relative mRNA expression (H3F3A/β-actin fraction) of the gene was obtained by normalizing gene transcripts per million (TPM) with that of β-actin. Tissue median adjusted mRNA expression levels (mREU/REU) were calculated. Confidence intervals of three reference genes (*GAPDH*, *ARHGDIA* and *RPS27A*) mRNA expression were established. Dashed line = upper limits of confidence intervals for reference genes. Left and right y axes = H3f3a/β-actin and mREU/REU fractions, respectively.

➢ ***Danio rerio***





***h3f3a* relative expression in zebrafish tissues (A) and in developmental stages (B).** Expression was measured by qPCR in zebrafish adult tissues and related to levels of β -actin gene. Data are expressed as mean ± SD. h3f3a is more expressed in the brain and in the initial developmental stages of zebrafish.

Conclusions

Through bioinformatic analysis we show that zebrafish *h3f3a* shares the same gene structure and protein-coding sequence as its human ortholog, the main differences corresponding to different 5´non-coding exons, resulting in different sized transcripts derived from distinct promoters. Interestingly, and despite the observed differences in promoter regions, we identified in zebrafish similar putative binding sites for transcription factors known to regulate human *H3F3A*. Since histones are crucial to maintain/regulate chromatin structure/function, their protein structure is transversal to all eukaryotic organisms. The differences identified in the 5' UTR regions of the different transcripts, both in human and in zebrafish, are likely to be associated with their tissue or cell-specific regulation, in both normal and pathological conditions, a possibility that should be further analyzed.

References

Ayoubi HA, Mahjoubi F, Mirzaei R. J Gastrointest Oncol. 2017 8:64-69.
Divisato G, Scotto di Carlo F, Pazzaglia L, Rizzo R, Coviello DA, Benassi MS, Picci P, Esposito T, Gianfrancesco F.. Oncotarget. 2017 8:63121-63131.
Scotto di Carlo F, Divisato G, Iacoangeli M, Esposito T, Gianfrancesco F. BMC Cancer. 2018 18:358.

Acknowledgements

This study received national funds from the Portuguese Foundation for Science and Technology (FCT) through the project UID/Multi/04326/2019 (CCMAR)