Supplementary Information:

Camello, a novel family of Histone Acetyltransferases that acetylate histone H4 and is essential for zebrafish development

Krishanpal Karmodiya[1§], Krishanpal Anamika[1,2], Vijaykumar Muley[1], Saurabh J. Pradhan[1], Yoshita Bhide[1] and Sanjeev Galande[1]

[1] Center of Excellence in Epigenetics, Indian Institute of Science Education and Research, Pashan, Pune 411 021, India

[2]Present address: Persistent Systems Limited Pingala - Aryabhata, Erandwane, Pune 411004, India.

§Corresponding author

Running title: Genome-wide survey of HATs

Keywords: histone acetyltransferase/Camello proteins/histone acetylation/Histone 4/ evolution/zebrafish/.

Supplementary Information

Methods

Homologs of mouse CMLO3 protein and phylogenetic analysis

Homologs of CMLO3 protein of mouse were fished by searching against the database of protein sequences from 23 completely sequenced eukaryotic organisms (Supplementary Table 1). Splice isoforms were removed from ENSEMBL genome datasets prior to search for homologs. Jackhmmer program of HMMER package was used to search database for 3 iterations. The e-value for reporting homologs of CMLO3 protein sequence was set to the 1 and e-value significance for including sequences in next round of search was 0.01. This search yielded 450 proteins as probable homologs of CMLO3 protein after search of three iterations. The alignments generated for homologs of CMLO3 were retained and used for further analysis. These proteins were then subjected to the hmmscan program of HMMER package to identify domains and superfamilies reported in Pfam and Superfamily databases [1]

Phylogenetic analysis of CMLO3 homologs

Mulitple sequence alignments (MSAs) were constructed for Pfam and Superfamily detected acetyltransferase domains only and also for complete 450 sequences. These MSAs were subjected to the phyml software to reconstruct Maximum-likelihood Phylogenetic trees using default parameters [2].

Cell culture and transfection

HeLa cells were grown in DMEM medium supplemented with 10% fetal bovine serum at 37 ^{0}C with 5 % CO_2. For transient transfections, 0.5 x 10^6 cells were transfected with 2.5 and 5.0 µg of plasmid DNA by Lipofectamine 2000 (Invitrogen). Transfection was performed according to the manufacturer's instructions.

Immunostaining and microscopy

HeLa cells were stained with anti-calnexin (ab22595, Abcam) and anti-Lamin B1 (Ab16048, Abcam) antibodies after fixation and permeabilization. DNA counterstaining was performed using DAPI (4', 6'-diamidino-2-phenylindole). Z stacks of the images were collected as 0.2 µm

optical slices by confocal microscope (Carl Zeiss). Images were processed using ImageJ software.

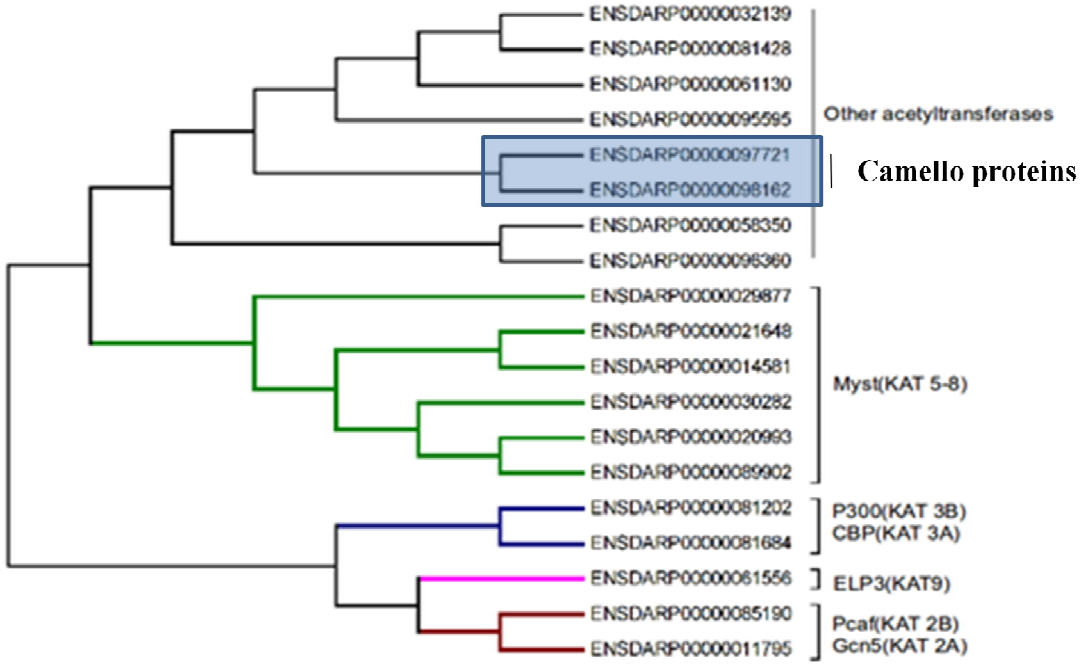
References:

1. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002) The Pfam protein families database. *Nucleic Acids Res*, **30:**276–280.

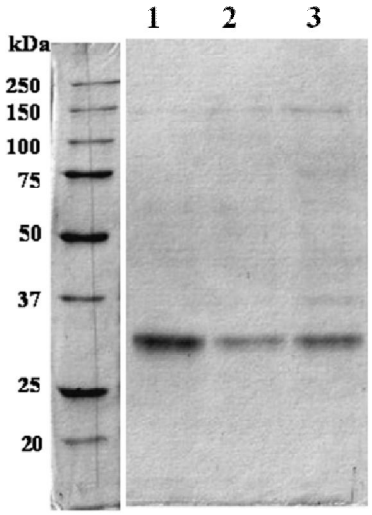
2. Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, **59(3):**307-321.

Supplementary Figure legends:

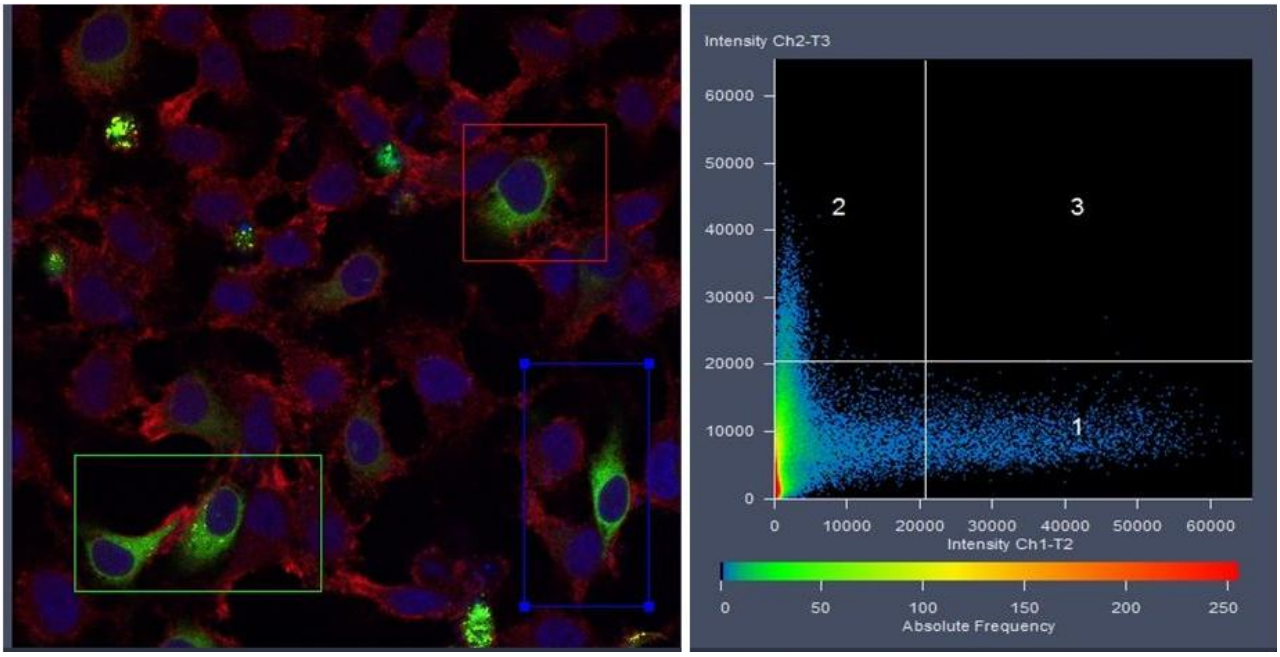
Additional file 1: Figure S1. Complement of KAT homologs in Zebrafish genome. Dendrogram of KATs in mouse classified into known and new family containing camello proteins. Homologs of different KAT families are represented in different colors. The new family of KAT 'Camello family' identified in this study is highlighted in blue color. Presence of Camello proteins in zebrafish suggest that this family of proteins is conserved across genomes.



Additional file2: Figure S2. Purification of camello proteins, CMLO3, E0CYR6 and CMLO2 by Ni-NTA chromatography. 12 % SDS ⁄ PAGE of recombinant camello proteins. Protein molecular weight markers is from MBI Fermentas. Lane 1: purified CMLO3. Lane 2: E0CYR6. Lane 3: CMLO2. CMLO3, E0CYR6 and CMLO2 (~28 kDa proteins) are purified to homogeneity.

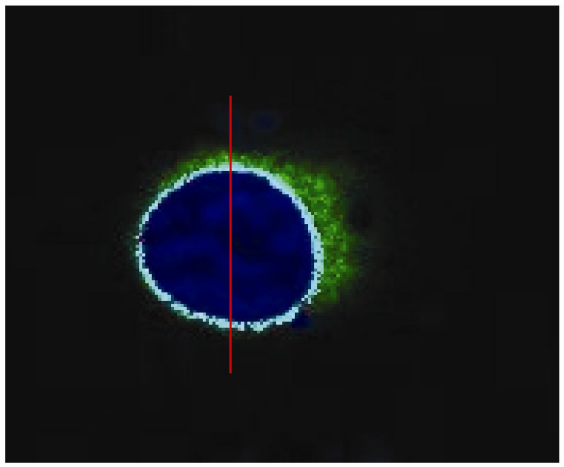


Additional file3: Figure S3. CMLO3 is expressed as GFP fusion protein in HeLa cells. Co-localization of CMLO3-GFP fusion protein (green, channel 2) was compared with endoplasmic reticulum marker, calnexin (red, channel 3). CMLO3 is not co-localized with endoplasmic reticulum marker.

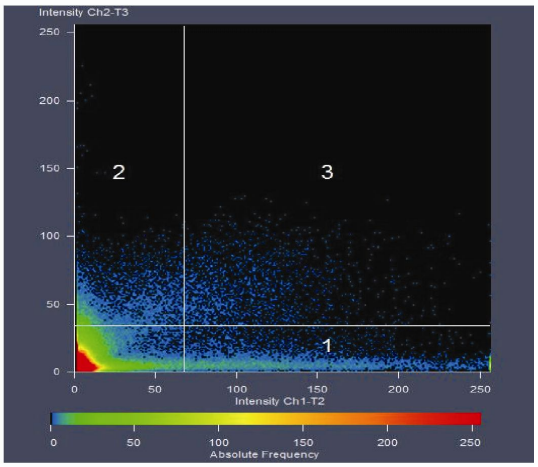


Additional file4: Figure S4. CMLO3 is expressed as GFP fusion protein in HeLa cells. Co-localization of CMLO3-GFP fusion protein (green, channel 1) was compared with marker of inner nuclear membrane, β-lamin (red, channel 2). CMLO3 is co-localized with marker of inner nuclear membrane lamin B1.

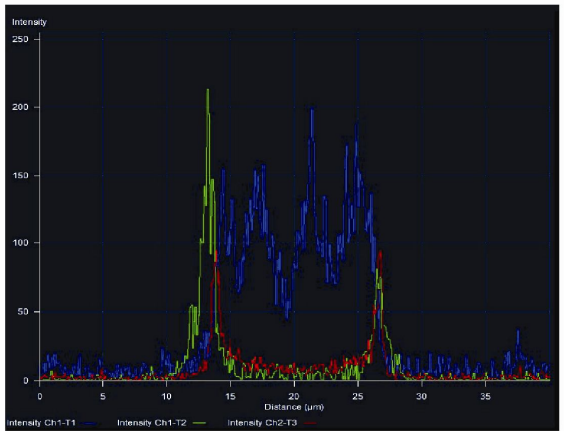
A



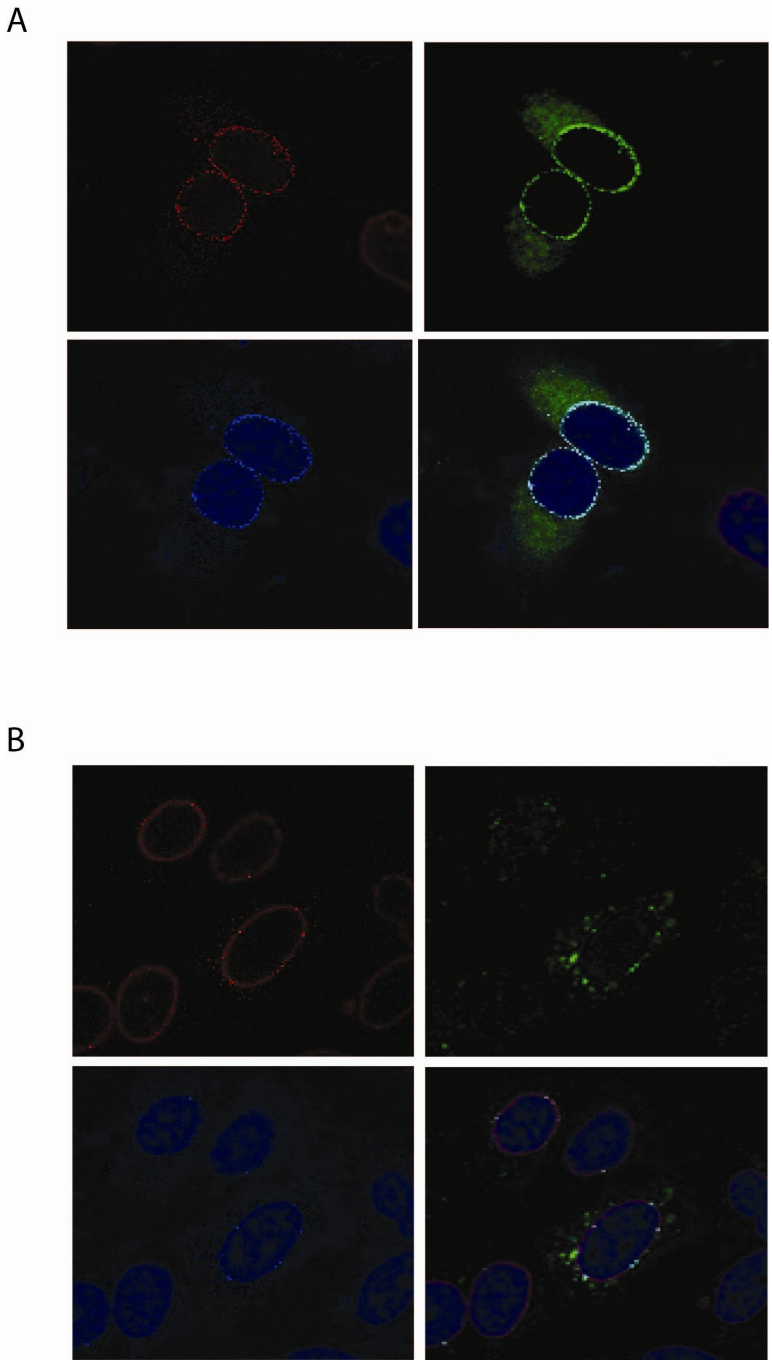
B



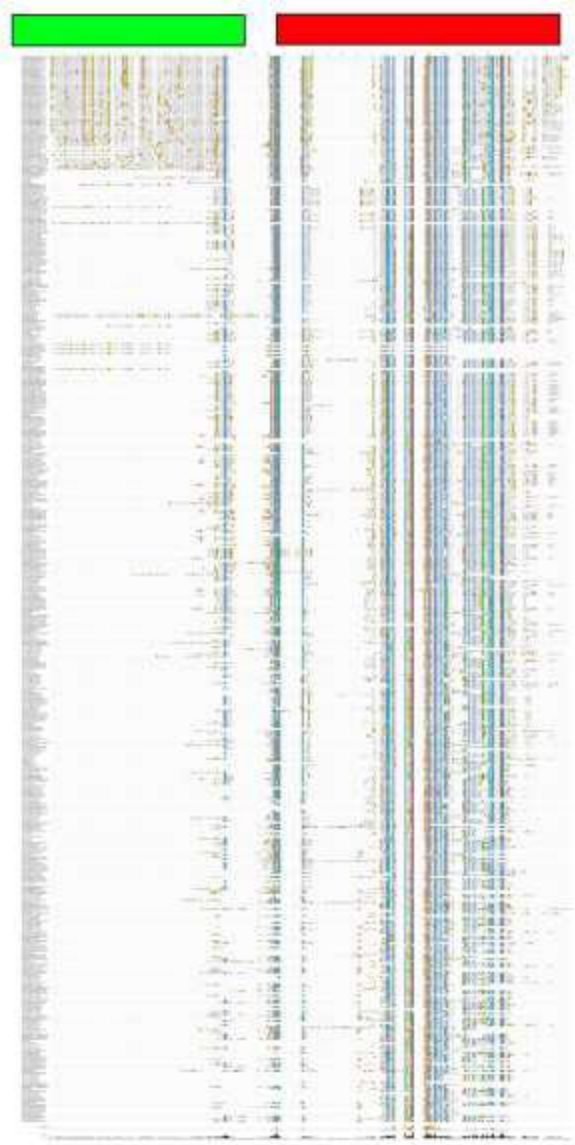
C



Additional file5: Figure S5. (A) C-terminal his-myc tagging of the CMLO3 recapitulate the localization profile of the CMLO3-GFP fusion protein and its overlap with lamin B1, marker of inner nuclear membrane. **(B)** N-terminal tagging of CMLO3 with his-myc affects the localization of the CMLO3 and was found mostly in the cytoplasm.



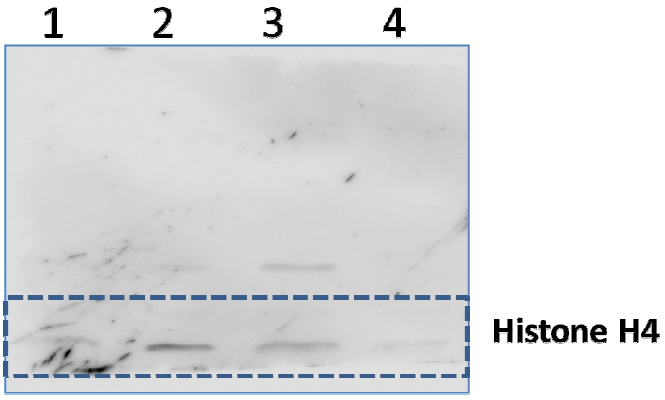
Additional file7: Figure S6. Multiple sequence alignment of CMLO3 homologs. 450 proteins detected by jackhmmer in 23 genomes aligned using clustalo program. Green color box shows sequence region unique to the CMLO3 proteins and red color box shows approximate location of acetyl transferase domain. I could detect homologs in the unicellular yeast, monosiga, and sponge for acetyl transferase domain but Cnidarian Nematostella encodes full length CMLO3-like protein suggesting that the N-terminus (green box) region of CMLO3 proteins is acquired along with the tool kit necessary for two germ layer formations. Cnidarians are the early branching multicellular organisms with two germ layers and body axis.



Additional file6: Table S1. Details of eukaryotic organisms used for CMLO3 protein homologs. Complete sets of protein sequences for all organisms were downloaded from Ensembl, NCBI, and JGI databases in January, 2013. Ensembl sequences were updated in April, 2013.

Name of the organism	Phylum	Data source
Saccharomyces cerevisiae	Ascomycota	Ensembl
Monosiga brevicollis	Craspedida	JGI
Amphimedon queenslandica	Porifera	Ensembl
Hydra magnipapillata	Cnidaria	NCBI
Nematostella vectensis	Cnidaria	Ensembl
Drosophila melanogaster	Arthropoda	Ensembl
Caenorhabditis elegans	Nematoda	Ensembl
Strongylocentrotus purpuratus	Echinodermata	Ensembl
Saccoglossus kowalevskii	Hemichordata	NCBI
Branchiostoma floridae	Chordata	NCBI
Ciona intestinalis	Chordata	Ensembl
Petromyzon marinus	Chordata	Ensembl
Danio rerio	Chordata	Ensembl
Xenopus tropicalis	Chordata	Ensembl
Gallus gallus	Chordata	Ensembl
Anolis carolinensis	Chordata	Ensembl
Ornithorhynchus anatinus	Chordata	Ensembl
Monodelphis domestica	Chordata	Ensembl
Bos Taurus	Chordata	Ensembl
Oryctolagus cuniculus	Chordata	Ensembl
Mus musculus	Chordata	Ensembl
Pan troglodytes	Chordata	Ensembl
Homo sapiens	Chordata	Ensembl

Additional file8: Figure S7: Camello family proteins CMLO3 and E0CYR6 are active histone acetyltransferases. (A) HAT assay using baculo produced recombinant histones, H3 and H4. E0CYR6 and CMLO3 are showing histone H4 acetylation (lanes 2 and 3, respectively).



Additional file9: Figure S8: Western blot analysis of CMLO3-GFP overexpressing HeLa cells after cytoplasmic and nuclear fractionation. Lamin B1 and tubulin were used as markers of nuclear and cytoplasmic fractions respectively. CMLO3-GFP was mainly observed in nuclear fraction. C - Cytoplasmic fraction; N- Nuclear fraction.