

Supplementary data

Structural basis for transcription factor ZBTB7A recognition of DNA and effects of ZBTB7A somatic mutations that occur in human acute myeloid leukemia

Ren Ren¹, John R. Horton¹, Qin Chen¹, Jie Yang¹, Bin Liu¹, Yun Huang², Robert M. Blumenthal³, Xing Zhang^{1,*}, Xiaodong Cheng^{1,*}

¹Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

²Center for Epigenetics and Disease Prevention, Institute of Biosciences and Technology, College of Medicine, Texas A&M University, Houston, TX 77030, USA

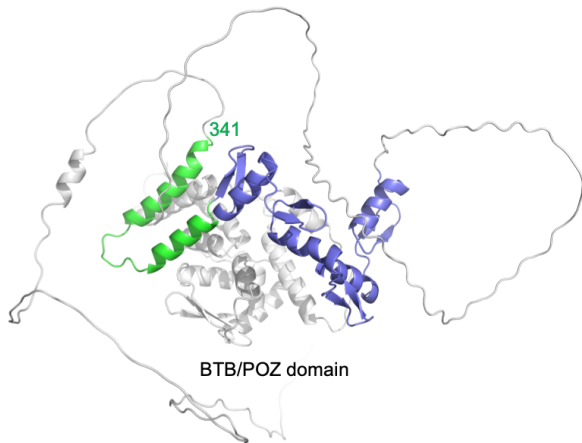
³Department of Medical Microbiology and Immunology, and Program in Bioinformatics, The University of Toledo College of Medicine and Life Sciences, Toledo, OH 43614, USA

*Correspondence: xzhang21@mdanderson.org (X.Z.); xcheng5@mdanderson.org (X.C.)

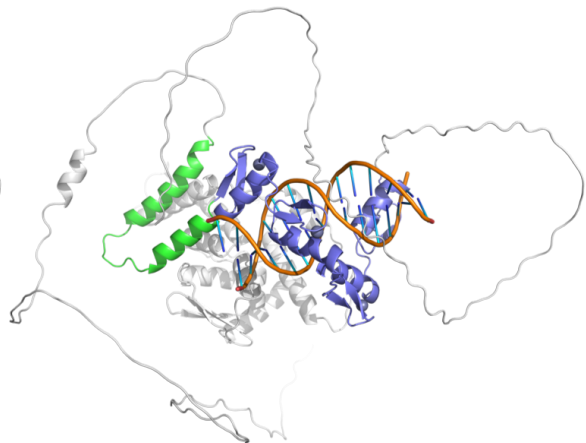
| A | Description | Query | E-Value | % | Residues | Accession |
|---|-------------------------------|-------|---------|-------|----------|----------------|
| | ZBTB7A protein [Homo sapiens] | 100% | 0.0 | 100.0 | 584 | AAH84568.1 |
| | ZBTB7B protein [Homo sapiens] | 88% | 2e-79 | 36.7 | 539 | AAH12070.1 |
| | ZBTB7C protein [Homo sapiens] | 51% | 2e-66 | 60.0 | 619 | NP_001358214.1 |

| | | | |
|--------|-----|---|-----|
| ZBTB7A | 1 | MAGGVDPGIPGIPFDHSSDILSGLNEQRTQGLLCDVVILVEGREFPTHRSLAACSQYFKKLFVTSGAVVDQON-----VYE | 76 |
| ZBTB7B | 1 | MGSPEDDLIGIPFDHSSSELLSCLNEQRLGHLCDLTIQTQGLEVRTHRAVLAACSHYFKKLFTEGGGGAVMGAGGGTATGGAGAGVCE | 90 |
| ZBTB7C | 1 | MANDIDELIGIPFPNHSSEVLCSLNEQRHDGLLCDVLLVVQQEYRTHRSLAACSQYFKKLFVTSGLASQPY-----VYE | 76 |
| ZBTB7A | 77 | IDFVSAEALTAALMDFAYTATLTVSTANVGDILSAARLLEIPAVSHVCADLLDRQILAADAGADAGQLDLVDQIDQRNLLRAKEYLEFPQS | 166 |
| ZBTB7B | 91 | LDFVGPALGALLEFAYTATLTTSSANMPAVLQAAARLLEIPCVIAACMEILQGSGLEASP----- | 151 |
| ZBTB7C | 77 | IDFVQPEALAAALIEFAYTSLTITAGNVKHLNAARMLLEIQCVNVVLEIMEPGDGGEE-----DDKEFQAGS | 146 |
| ZBTB7A | 167 | NPMNSLPPAAAAAASF PWSAFGASDDDL DATKEAVAAA VAAAGDCNGLDFYGP GPPAERPPTGDGEGDSNPGLWPERDEDAPTGGL | 256 |
| ZBTB7B | 152 | -----DEDDCERARQYLEAFATATASGVPNGEDSPQVPLPPPPPPPPRVARRSRKRPRKAFLOTK GARAN | 217 |
| ZBTB7C | 147 | PGHLGVIRD FSIESLLRENLD DDDDEDDDEDEDEEEEEEDDDDDT EDPADQENL PDPQDISCHQSPSKT DHLT EKAYS DTPRDF PDS | 236 |
| ZBTB7A | 257 | FPPPVAPPAATQNGHYGRGEEEEAAALSEAAPEPGD SPGLFSGAAEGDGDGDPVDGLAAS TLLQMMSSVGRAGAAA-GDSDEESR | 342 |
| ZBTB7B | 218 | HLVPEVPTVPAHPLTYEEEEVAGRVGSSGGSGPGDSSYPPTGTASPEPGQSYEPYEGEEEEELVYPAYGLAQGGPPLSPEELG | 304 |
| ZBTB7C | 237 | YFKANIPDRRSLSPFAPDFPHLWPGDFGAFALPEQPMDSGPLDLVIKNRKIKEEKEELPPPPPPPPNDFFKDMFPDLPGGPL | 323 |
| ZBTB7A | 343 | ADDKGVMDY--YLKYFSGAHDGDVYPAW-SQ-KVEKKIRAKAFQKCPICEKVIQGAGKLPFRHIRTHTGKPYECNICKVRFTRQDKLKVH | 428 |
| ZBTB7B | 305 | SDEDAIDPD--LMAYLSSLHQDNLAPGLDSQDKLVRRRSQMPQECVCHKLIHCAGKLPFRHIRTHTGKPYECNICKVRFTRQDKLKVH | 392 |
| ZBTB7C | 324 | GP IKAENDY GAYLNFLSATHLGLFPFW-PL-VEERKLPKASQQCPICHKVIHCAGKLPFRHIRTHTGKPYMCTICEVRFTRQDKLKH | 410 |
| ZBTB7A | 429 | MRKHTGKRPYLCOQCGAFAHNYDLKNHMRVHTGLRPYQDCSCCKTFVRSDDLHRLKDKGCGNVPSSRRGRKPRVR-----GGAPDP | 510 |
| ZBTB7B | 393 | MRKHTGERPYSCHPCARFLHSYDLKNHMLHTGDRPYECHLCHKAFKEDHLQRHLKGGQNCLEVRTRRRKDDAP-----PHYPPP | 474 |
| ZBTB7C | 411 | MRKHTGERPYLCTHCNAKVFVHNYDLKNHMRHTGVRPYQCFCKSFTVRSDDLHRLKROSCRMARPRGRKPAAWRAASLLFGPGGPAP | 500 |
| ZBTB7A | 511 | SPGATATPGAPQSSPDARRNGQEKHFKDEDEDEDVASPDGLCRLNVAGAGGGGDSGGGPGAATDGNFTAGLA* | 584 |
| ZBTB7B | 475 | STAAAAPAGLDLSNGHLDTFRLSLARFWEQSAPTGPVSTPGPPDDDEEGAPTTPQAE GAMESS* | 539 |
| ZBTB7C | 501 | DKAAFVMPALGEVGGHLGGAAVCLPGSPAKHFLAAPKGLSLQELERQFEETQMKLFGRAQLAERNAGGLLAFALAE NVAAARPYPF | 590 |
| ZBTB7C | 591 | LDPDWAAGLAGLPLAGLNHVASMSEANN* 619 | |

B Human ZBTB7A AlphaFold prediction



C Superimposed with bound DNA (PDB 7N5T)



| D | Class | Species | ZBTB7A Ortholog | Query cover | E-value | % identity | #AA |
|---|----------------|-----------------------------|-----------------|-------------|---------|------------|-----|
| | Mammalia | <i>Homo sapiens</i> | AAH84568.1 | 100% | 0.0 | 100 | 584 |
| | Reptilia | <i>Anolis carolinensis</i> | XP_008123039.1 | 90 | 0.0 | 66 | 563 |
| | Aves | <i>Gallus gallus</i> | XP_046760486.1 | 97 | 0.0 | 66 | 558 |
| | Chondrichthyes | <i>Rhinodon typus</i> | XP_048470032.1 | 95 | 1e-163 | 55 | 570 |
| | Osteichthyes | <i>Lepisosteus oculatus</i> | XP_015220868.1 | 85 | 1e-143 | 53 | 657 |
| | Amphibia | <i>Xenopustropicalis</i> | XP_004910395.1 | 85 | 1e-107 | 42 | 585 |

| | | ZF1 (5) | ZF2 ((5) | ZF3 (7) | ZF4 (8) |
|----------------|-----|---|----------|---------|---------|
| Mammalia | 381 | FQKCPICEKVIQGAGKLPFRHIRTHTGKPYECNICVRFTRQDKLKVHMRKHTGKPYLCQCGAFAHNYDLKNHMRVHTGLRPYQDCSCCKTFVRSDDLHRLKDKGCGN | (0) | | |
| Reptilia | 352 | FQKCPICEKVIQGAGKLPFRHIRTHTGKPYECNICVRFTRQDKLKVHMRKHTGKPYLCQCGAFAHNYDLKNHMRVHTGLRPYQDCSCCKTFVRSDDLHRLKDKGCGN | (2) | | |
| Aves | 370 | FQKCPICAKVIQGAGKLPFRHIRTHTGKPYECNICVRFTRQDKLKVHMRKHTGKPYLCQCGAFAHNYDLKNHMRVHTGLRPYQDCSCCKTFVRSDDLHRLKDKGCGN | (3) | | |
| Chondrichthyes | 370 | FQKCPICEKVIQGAGKLPFRHIRTHTGKPYECNICVRFTRQDKLKVHMRKHTGKPYLCQCGAFAHNYDLKNHMRVHTGLRPYQDCSCCKTFVRSDDLHRLKDKGCGN | (5) | | |
| Osteichthyes | 461 | FQKCPICSKVIQGAGKLPFRHIRTHTGKPYECNICVRFTRQDKLKVHMRKHTGKPYLCQCGAFAHNYDLKNHMRVHTGLRPYQDCSCCKTFVRSDDLHRLKDKGCGN | (6) | | |
| Amphibia | 344 | SQKCPICHKVIHCAGKLPFRHIRTHTGKPYMCTICEVRFTRQDKLKHRLKROSCRMARPRGRKPAAWRAASLLFGPGGPAP | (24) | | |

Figure S1. (A) Comparison of human ZBTB7A, B, and C. Identities in all three are highlighted in gray, while the four zinc finger motifs are in yellow, with cyan indicating substitutions relative to ZBTB7A. The DNA base interacting residues are highlighted in white letter with red background (ZF-3 makes only backbone contacts). (B) AlphaFold

prediction of ZBTB7A full-length protein, including two additional helices (colored in green) prior to the ZF DNA binding domain (colored in blue). The corresponding sequences for the green helices are not conserved and helices are not predicted to be present in ZBTB7B and 7C. **(C)** Superimposition of ZF-DNA binding complex onto the AlphaFold ZBTB7A full-length protein suggesting that the green helices might not be directly involved in DNA binding. **(D)**. ZF borders are as defined for the human ortholog, with cyan indicating differences from the human ortholog. Underlined residues make base-specific contacts to DNA, as defined for the human ortholog in this study, and are completely conserved. While ZF3 makes no direct DNA base contacts, it does not appear to have an elevated substitution rate compared to the other ZFs (5 positions in ZF1, 5 in ZF2, 7 in ZNF-3, and 8 in ZF4). The linkers contain conserved glycines (red font), that are proposed to be important for providing flexible orientation between ZFs.

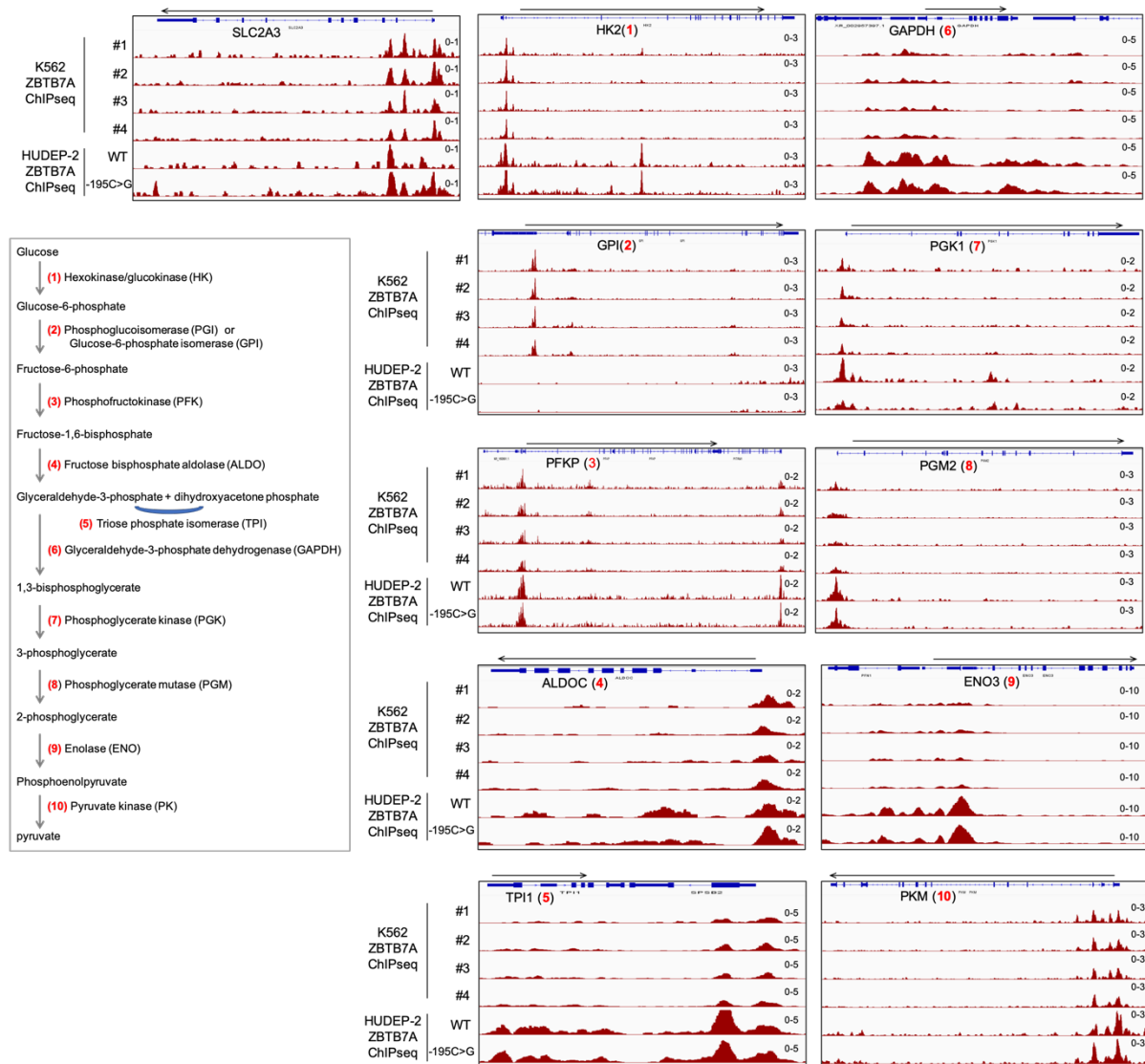


Figure S2. Representative tracks of ZBTB7A ChIP-seq in K562 cells (4 replicates) and HUDEP-2 (WT and -195C>G mutant), taken from Gene Expression Omnibus under accession GSE103445. Besides SLC2A3 (also known as GLUT3), we examined ten glycolytic genes shown in the boxed schematic. The results were normalized to track scale at each locus. We note that there are number of isozymes (such as HK1, HK2, ...) for each reaction step, and we only choose one representative.

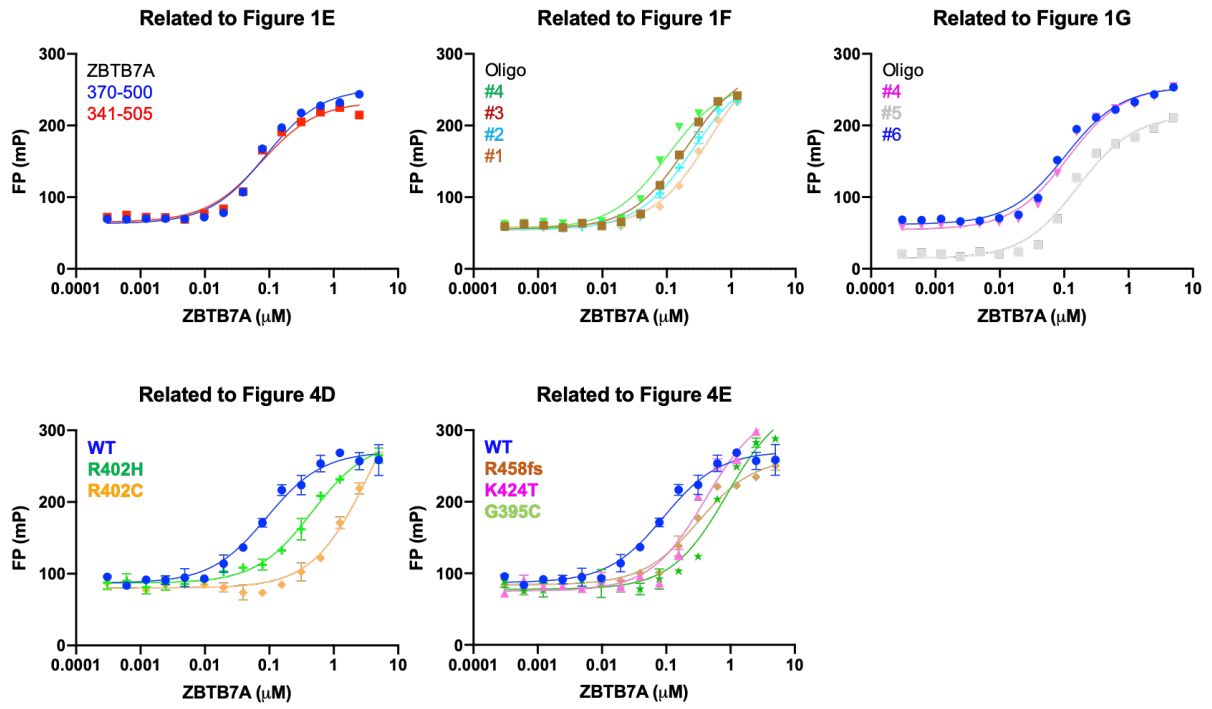


Figure S3. Raw data of millipolarization (mP) measured during fluorescence polarization using a Synergy 4 Microplate Reader (BioTek).

Table S1. Summary of X-ray data collection and refinement statistics (*) at SERCAT beamline 22ID

| Oligonucleotides | 5'-GGGACCCCTTGATGTTT-3' 3'-CCTGGGAACACTACAAAC-5' | 5'-GGGACCCCTTGATGTTT-3' 3'-CCCTGGGAACACTACAAA-5' | 5'-GGGACCCCTTCATGTTT-3' 3'-CCTGGGAAGTACAAAC-5' | 5'-GGTAAAAGACCCCTCCCAAT-3' 3'-CCAATTTCTGGGGAGGGTTTA-5' | 5'-GGTAAAAGACCCCTCCCAAA-3' 3'-CATTTTCTGGGGAGGGTTTC-5' |
|---|---|---|---|---|--|
| PDB code | 7N5U | 7N5V | 7N5W | 8E3D | 8E3E |
| Date collected | 2021-03-06 | 2021-03-06 | 2021-04-19 | 2022-06-09 | 2022-06-09 |
| Space group | $P2_12_12$ | $P2_1$ | $C222_1$ | $P6$ | $P2_1$ |
| Cell dimensions (Å) | 63.38, 103.61, 37.07 | 51.29, 60.07, 133.72 | 74.16, 103.43, 83.08 | 196.18, 196.18, 54.96 | 67.52, 147.18, 68.06 |
| α, β, γ (°) | 90, 90, 90 | 90, 98.5, 90 | 90, 90, 90 | 90, 90, 120 | 90, 96.5, 90 |
| Resolution (Å) | 40.11-2.85 (2.95-2.85) | 43.47-3.09 (3.20-3.09) | 37.08-2.25 (2.33-2.25) | 46.15-2.62 (2.71-2.62) | 49.80-2.99 (3.10-2.99) |
| ^a R_{merge} | 0.068 (1.06) | 0.142 (0.531) | 0.124 (0.904) | 0.182 (3.082) | 0.080 (0.420) |
| R_{pim} | 0.020 (0.377) | 0.067 (0.274) | 0.035 (0.470) | 0.039 (0.706) | 0.044 (0.331) |
| $CC_{1/2}, CC$ | (0.555, 0.845) | (0.992, 0.998) | (0.598, 0.865) | (0.398, 0.755) | (0.945, 0.986) |
| ^b $\langle I/\sigma I \rangle$ | 25.3 (0.9) | 7.9 (1.8) | 18.3 (1.2) | 21.7 (1.0) | 15.7 (2.1) |
| Completeness (%) | 79.5 (22.6) | 96.0 (86.7) | 89.5 (43.6) | 99.5 (100.0) | 98.7 (97.6) |
| Redundancy | 11.3 (6.6) | 5.4 (4.3) | 11.4 (2.5) | 22.1 (19.1) | 7.1 (6.4) |
| Observed reflections | 54,455 | 79,307 | 158,797 | 810,931 | 185,731 |
| Unique reflections | 4,830 (162) | 14,595 (1,338) | 13,984 (668) | 36,650 (3,610) | 26,193 (2,577) |
| | (3572 have both I+ and I-) | | | | |
| Wilson B-factor (Å ²) | 65.0 | 48.4 | 47.8 | 51.3 | 58.8 |
| Mean FOM (SAD) | 0.45 (at 5Å) | | | | |
| Density Modification Rfactor | 0.40 (at 5Å) | | | | |
| Refinement | | | | | |
| Resolution (Å) | 2.85 | 3.09 | 2.25 | 2.62 | 2.99 |
| No. reflections | 4,449 | 14,484 | 13,904 | 36,634 | 25,726 |
| ^c $R_{\text{work}} / ^d R_{\text{free}}$ | 0.248 / 0.300 | 0.264 / 0.314 | 0.191 / 0.229 | 0.206 / 0.237 | 0.203 / 0.235 |
| No. Atoms | (monomer) | (three molecules) | (monomer) | (three molecules) | (three molecules) |
| Protein | 631 | 1,731 | 670 | 1,777 (843) ^e | 2,703 |
| DNA | 649 | 1,947 | 650 | 1,800 (900) ^e | 2,565 |
| Solvent | 7 | 1 | 52 | 96 | 9 |
| Zn | 3 | 8 | 3 | 8 (4) ^e | 12 |
| B Factors (Å ²) | | | | | |
| Protein | 50.4 | 128.8 | 57.0 | 62.1 (170) ^e | 118.0 |
| DNA | 63.7 | 146.9 | 76.7 | 74.0 (191) ^e | 120.7 |
| Solvent | 24.6 | 80.5 | 56.2 | 59.8 | 85.7 |
| Zn | 72.6 | 147.0 | 58.9 | 61.4 (205) ^e | 140.9 |
| R.m.s. deviations | | | | | |
| Bond lengths (Å) | 0.003 | 0.003 | 0.006 | 0.004 | 0.003 |
| Bond angles (°) | 0.5 | 0.6 | 0.8 | 0.6 | 0.5 |

* Values in parenthesis correspond to highest resolution shell.

^a $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity and $\langle I \rangle$ is the averaged intensity from multiple observations.^b $\langle I/\sigma I \rangle$ = averaged ratio of the intensity (I) to the error of the intensity (σI).^c $R_{\text{work}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.^d R_{free} was calculated using a randomly chosen subset (5%) of the reflections not used in refinement.^e Three monomers are in asymmetric unit with density for one significantly worse than other two. Values in parenthesis are for the worse monomer.

Table S2. Expression plasmids used in this study

| Description | plasmid # |
|--|-----------|
| Human ZBTB7A residues 370-500 expressed in E. coli | pXC2222 |
| Human ZBTB7A residues 341-505 expressed in E. coli | pXC2311 |
| Human ZBTB7A mutant G395C (residues 370-500) expressed in E. coli | pXC2282 |
| Human ZBTB7A mutant R402H (residues 370-500) expressed in E. coli | pXC2280 |
| Human ZBTB7A mutant R402C (residues 370-500) expressed in E. coli | pXC2281 |
| Human ZBTB7A mutant K424T (residues 370-500) expressed in E. coli | pXC2279 |
| Human ZBTB7A mutant R458fs (residues 370-500) expressed in E. coli | pXC2283 |
| EGFP-ZBTB7A full length expression in PC3, U2OS and HeLa cells | pXC2135 |
| EGFP-ZBTB7A-R458fs expression in PC3, U2OS and HeLa cells | pXC2327 |

Table S3. Oligonucleotides used in this paper.

| Oligonucleotides | Application |
|--|----------------------------|
| FAM-5' -GTCGACCCGGCCTGGCG-3' 3' -CAGCTGGGCCGGACCGC-5' | FP binding |
| FAM-5' -TGGGACCCACGCACCGC-3' 3' -ACCCTGGGTGCGTGGCG-5' | FP binding |
| FAM-5' -GGTGACCCTCCGGATTC-3' 3' -CCACTGGGAGGCCTAAG-5' | FP binding |
| 5' -CGCCCCACCCCCACCA-3' 3' -GCGGGGGTGGGGGTGGT-5' -FAM | FP binding |
| 5' -GGCCCCTTCCCCACA-3' 3' -CCGGGGAAGGGGTGT-5' -FAM | FP binding |
| 5' -AAGACCCCTCCCCGAA-3' 3' -TTCTGGGGAGGGGCTT-5' -FAM | FP binding |
| 5' -GGGACCCTTCATGTTT -3' 3' - CCTGGGAAGTACAAAC-5' | Crystallization |
| 5' -GGGACCCTTGATGTTT -3' 3' - CCTGGGAAGTACAAAC-5' | Crystallization |
| 5' -GGGACCCTTGATGTTT-3' 3' -CCCTGGGAAGTACAAA-5' | Crystallization |
| 5' -GGTAAAAGACCCCTCCCCAAAT-3' 3' -CCATTTTCTGGGGAGGGGTTTA-5' | Crystallization |
| 5' -GGTAAAAGACCCCTCCCCAAA -3' 3' - CATTTTCTGGGGAGGGGTTTC-5' | Crystallization |
| 5' -CCC AAG CTT ATG GAC TAC AAA GAC GAT GAC GAC AAG ATG GTG AGC AAG GGC GAG GA-3' | Forward primer for pXC2315 |
| 5' -CCG GAA TTC TTC TTG TAC AGC TCG TCC ATG CCG AGA GTG-3' | Reverse primer for pXC2315 |