SUPPORTING INFORMATION

ASCC1 structures and bioinformatics reveal a novel Helix-Clasp-Helix RNA-binding motif linked to a two-histidine phosphodiesterase

Naga babu Chinnam, Roopa Thapar, Andrew S. Arvai, Altaf H. Sarker, Jennifer Soll, Tanmoy Paul, Aleem Syed, Daniel J. Rosenberg, Michal Hammel, Albino Bacolla, Panagiotis Katsonis, Abhishek Asthana, Miaw-Sheue Tsai, Ivaylo Ivanov, Olivier Lichtarge, Robert H. Silverman, Nima Mosammaparast, Susan E. Tsutakawa, John A. Tainer

ASCC1 Supplemental Table S1

ASCC1 germline mutations and phenotypes

Mutation cDNA (Protein)	*No cases*	*Phenotype*	*PMID*
c.157dupG (p.Glu53Glyfs19*) H	2	SMA and congenital bone fractures	26924529
c.157dupG (p.Glu53Glyfs19*) H	1	SMA and congenital bone fractures	28218388
c.710+1G>A (splicing, +1 FS) H	1	Fetal akinesia & intrauterine fetal death	28749478
c.871+1G>A splicing, +1 FS) H	1	Hydrops fetalis	28749478
c.157dupG (p.Glu53Glyfs19*) H	2	Muscle weakness with arthrogryposis & bone fractures	30327447
c.157dupG (p.Glu53Glyfs19*) and c.466C>T (p.Arg156*) CH	1	Muscle weakness with arthrogryposis & bone fractures	30327447
c.412C>T (p.Arg138*) H	3	Muscle weakness with arthrogryposis & bone fractures	30327447
64kb CNV - exons 7-10del and c.1027C>T (p.Arg343*) CH	1	SMABF2	31880396
Exon 5del and c.932C>G (p.Ser311*) CH	1	SMABF2	32160656
c.466C>T (p.Arg156*) and c.297-8T>G CH	1	SMABF2	33931933
c.626+1G>A (splicing) H	2	SMABF2	35338657
c.710+1G>A (p.Asp192*) H	1	SMABF2	35690317
c.897G>A (p.Trp299*) H	1	SMABF2	35690317
N290S	2	Barrett oesophagus/oesophageal adenocarcinoma	21791690

Supplemental Table S1. ASCC1 germline mutations and phenotypes

Mutations were taken from the original articles, which may have used different transcript versions. *SMA*, spinal muscular atrophy; *SMABF2*, spinal muscular atrophy with congenital bone fractures 2.

ASCC1 Supplemental Table S2

Supplemental Table S2. Kaplan-Meier analysis of TCGA patients expressing various levels of ASCC1 mRNA.

For each tumor type patients were divided in 3 groups, g_low expressing ASCC1 levels below mean-1SD, g_high expressing ASCC1 levels above mean+1SD and g_medium expressing ASCC1 levels between mean-1SD and mean+1SD. P-values were derived from log-rank tests and corrected with Holm-Sidak mehod for pairwise comparisons. Highlighted values show significant P-values lower than 0.05.

See excel Table S2

ASCC1 Supplemental Table S3

ASCC1					ZNF652				
dataset	N	mean	SD	p-value	dataset	N	mean	SD	p-value
ascc1_blca_patLow_mut_f1	192	2.4084	0.376135		znf652_blca_patLow_mut_f1	181	2.3429	0.406115	
ascc1_blca_patHigh_mut_f1	217	2.32541	0.364449	2.44E-02	znf652_blca_patHigh_mut_f1	228	2.38141	0.342167	3.08E-01
ascc1_brca_patLow_mut_f1	556	1.77152	0.39329		znf652_brca_patLow_mut_f1	543	1.8112	0.362993	
ascc1_brca_patHigh_mut_f1	521	1.81698	0.336817	4.15E-02	znf652_brca_patHigh_mut_f1	534	1.77553	0.371687	1.12E-01
ascc1_cesc_patLow_mut_f1	142	2.09357	0.451023		znf652_cesc_patLow_mut_f1	147	2.05387	0.451572	
ascc1_cesc_patHigh_mut_f1	157	2.06295	0.444509	5.55E-01	znf652_cesc_patHigh_mut_f1	152	2.10033	0.443069	3.70E-01
ascc1_coad_patLow_mut_f1	151	2.432	0.507098		znf652_coad_patLow_mut_f1	133	2.32515	0.513128	
ascc1_coad_patHigh_mut_f1	134	2.32322	0.490714	6.70E-02	znf652_coad_patHigh_mut_f1	151	2.43112	0.488915	7.69E-02
ascc1_dlbc_patLow_mut_f1	17	1.83681	0.427772		znf652_dlbc_patLow_mut_f1	17	1.9316	0.505366	
ascc1_dlbc_patHigh_mut_f1	20	2.05722	0.432315	1.29E-01	znf652_dlbc_patHigh_mut_f1	20	1.97664	0.385276	7.66E-01
ascc1_gbm_patLow_mut_f1	126	1.81366	0.186422		znf652_gbm_patLow_mut_f1	103	1.78211	0.321027	
ascc1_gbm_patHigh_mut_f1	112	1.75485	0.352909	1.16E-01	znf652_gbm_patHigh_mut_f1	136	1.79437	0.237379	7.45E-01
ascc1_hnsc_patLow_mut_f1	260	2.1163	0.333272		znf652_hnsc_patLow_mut_f1	259	2.10615	0.328231	
ascc1_hnsc_patHigh_mut_f1	241	2.14586	0.337201	3.25E-01	znf652_hnsc_patHigh_mut_f1	242	2.15661	0.341167	9.26E-02
ascc1_kich_patLow_mut_f1	33	1.21741	0.351535		znf652_kich_patLow_mut_f1	34	1.08788	0.27524	
ascc1_kich_patHigh_mut_f1	32	1.083	0.285058	9.51E-02	znf652_kich_patHigh_mut_f1	31	1.22073	0.36427	1.05E-01
ascc1_kirc_patLow_mut_f1	222	1.84695	0.255667		znf652_kirc_patLow_mut_f1	218	1.8669	0.26008	
ascc1_kirc_patHigh_mut_f1	268	1.84549	0.228622	9.47E-01	znf652_kirc_patHigh_mut_f1	270	1.83082	0.223438	1.06E-01
ascc1_kirp_patLow_mut_f1	124	1.85637	0.244743		znf652_kirp_patLow_mut_f1	129	1.86758	0.251976	
ascc1_kirp_patHigh_mut_f1	151	1.89094	0.287787	2.83E-01	znf652_kirp_patHigh_mut_f1	146	1.88222	0.284429	6.51E-01
ascc1_laml_patLow_mut_f1	85	1.24111	0.36883		znf652_laml_patLow_mut_f1	78	1.25155	0.413246	
ascc1_laml_patHigh_mut_f1	84	1.30022	0.514887	3.93E-01	znf652_laml_patHigh_mut_f1	91	1.28672	0.475829	6.08E-01
ascc1_lgg_patLow_mut_f1	216	1.5796	0.326751		znf652_lgg_patLow_mut_f1	257	1.55068	0.311016	
ascc1_lgg_patHigh_mut_f1	319	1.53852	0.238131	1.14E-01	znf652_lgg_patHigh_mut_f1	272	1.55754	0.240459	7.78E-01
ascc1_lihc_patLow_mut_f1	188	2.03258	0.235417		znf652_lihc_patLow_mut_f1	195	2.01768	0.298734	
ascc1_lihc_patHigh_mut_f1	181	2.02829	0.305588	8.80E-01	znf652_lihc_patHigh_mut_f1	174	2.04481	0.237933	3.33E-01
ascc1_luad_patLow_mut_f1	280	2.29025	0.509699		znf652_luad_patLow_mut_f1	230	2.32018	0.478214	
ascc1_luad_patHigh_mut_f1	236	2.41622	0.399014	1.75E-03	znf652_luad_patHigh_mut_f1	286	2.36765	0.454107	2.52E-01
ascc1_lusc_patLow_mut_f1	260	2.36634	0.400437		znf652_lusc_patLow_mut_f1	228	2.41985	0.448865	
ascc1_lusc_patHigh_mut_f1	230	2.44738	0.436356	3.35E-02	znf652_lusc_patHigh_mut_f1	261	2.3912	0.392681	4.56E-01
ascc1_ov_patLow_mut_f1	128	1.84896	0.333557		znf652_ov_patLow_mut_f1	136	1.85747	0.35735	
ascc1_ov_patHigh_mut_f1	145	1.84984	0.35194	9.83E-01	znf652_ov_patHigh_mut_f1	136	1.84321	0.329439	7.32E-01
ascc1_paad_patLow_mut_f1	87	1.57315	0.363197		znf652_paad_patLow_mut_f1	85	1.68256	0.215006	
ascc1_paad_patHigh_mut_f1	83	1.5314	0.432578	4.98E-01	znf652_paad_patHigh_mut_f1	85	1.42298	0.488145	1.72E-05
ascc1_prad_patLow_mut_f1	271	1.61054	0.262731		znf652_prad_patLow_mut_f1	267	1.55763	0.303817	
ascc1_prad_patHigh_mut_f1	226	1.60978	0.35076	9.79E-01	znf652_prad_patHigh_mut_f1	230	1.67121	0.296802	3.08E-05
ascc1_read_patLow_mut_f1	44	2.26098	0.3811		znf652_read_patLow_mut_f1	45	2.20698	0.324957	
ascc1_read_patHigh_mut_f1	47	2.18681	0.275281	2.93E-01	znf652_read_patHigh_mut_f1	48	2.22538	0.338108	7.90E-01
ascc1_sarc_patLow_mut_f1	121	1.70119	0.490081		znf652_sarc_patLow_mut_f1	117	1.7458	0.454256	
ascc1_sarc_patHigh_mut_f1	117	1.56548	0.407792	2.09E-02	znf652_sarc_patHigh_mut_f1	121	1.52683	0.432229	1.79E-04
ascc1_skcm_patLow_mut_f1	50	2.42776	0.615879		znf652_skcm_patLow_mut_f1	43	2.44586	0.588983	
ascc1_skcm_patHigh_mut_f1	54	2.56046	0.512794	2.37E-01	znf652_skcm_patHigh_mut_f1	61	2.53248	0.551073	4.50E-01
ascc1_stad_patLow_mut_f1	212	2.31044	0.543314		znf652_stad_patLow_mut_f1	191	2.30464	0.438264	
ascc1_stad_patHigh_mut_f1	200	2.21167	0.5027	5.60E-02	znf652_stad_patHigh_mut_f1	221	2.22606	0.589495	1.23E-01
ascc1_thca_patLow_mut_f1	203	1.20404	0.313955		znf652_thca_patLow_mut_f1	276	1.23004	0.301759	
ascc1_thca_patHigh_mut_f1	294	1.2398	0.291434	2.00E-01	znf652_thca_patHigh_mut_f1	221	1.21914	0.300714	6.89E-01
ascc1_ucec_patLow_mut_f1	82	2.36626	0.649737		znf652_ucec_patLow_mut_f1	69	2.14741	0.574131	
ascc1_ucec_patHigh_mut_f1	83	2.09564	0.568822	5.02E-03	znf652_ucec_patHigh_mut_f1	96	2.28959	0.653102	1.41E-01

Supplemental Table S3. For ASCC1 and ZNF652 we separated cancer patients into a higher (m_high, or patHigh) and lower expression (m_low or patLow) group for the respective gene based on RNA-seq data, approximately similar in population size as determined by the mean value. The dataset name is defined by the gene (ASCC1 or ZNF652), the TCGA cancer type, and the expression level (low vs. high).

We then compared the distribution of the log10 of the number of mutations that occurred for the low or high expression groups of the respective gene. The p-value from Student t-test is given for the significance of the difference.

ASCC1 Supplemental Table S4

Supplemental Table S4. Mutation loads in TCGA tumor samples expressing various levels of ASCC1 mRNA.

For each tumor type patients were divided in 3 groups, g_low (patLow) expressing ASCC1 levels below mean-1SD, g_high (patHigh) expressing ASCC1 levels above mean+1SD and g_medium (patMedium) expressing ASCC1 levels between mean-1SD and mean+1SD. For each tumor sample, the number of simple mutations exome-wide was obtained from COSMIC by intersecting TCGA patient codes with COSMIC patient codes and reported as a log10 transform. P-values were derived from pairwise t-tests. *Yellow highlights*, significant differences. *Most muts*, group with the higher mutational load.

See excel Table S4

ASCC1 Supplemental Table S5

Supplemental Table S5. Signature mutations in TCGA tumor samples expressing various levels of ASCC1 mRNA – 2 groups.

For each tumor type samples were divided in 2 groups, m_low expressing ASCC1 levels below the population mean, and m_high expressing ASCC1 levels above the population mean.

For each of the 30 mutational signatures (https://cancer.sanger.ac.uk/signatures/signatures_v2/) and for each of the 2 groups the number of samples containing mutations conforming to a given signature was recorded and compared using Fisher's tests. *Highlights*, significant p-values.

See excel Table S5

ASCC1 Supplemental Table S6

Supplemental Table S6. Signature mutations in TCGA tumor samples expressing various levels of ASCC1 mRNA- 3 groups.

For each tumor type patients were divided in 3 groups, g_low (patLow) expressing ASCC1 levels below mean-1SD, g_high (patHigh) expressing ASCC1 levels above mean+1SD and g_medium (patMedium) expressing ASCC1 levels between mean-1SD and mean+1SD.

For each of the 30 mutational signatures (https://cancer.sanger.ac.uk/signatures/signatures_v2/) and for each of the 3 groups the number of samples containing mutations conforming to a given signature was recorded and compared using Fisher's tests.

See excel Table S6

ASCC1 Supplemental Table S7

ApASCC1 full-length (SAXS)	SNAMDILRPPIICIDGFRFRRNPVQDKSPREYEEEEFGTVGSERVYCDDICDTLGEQIQETENGYKLELEIPSAY YKYIIGKKGETKKRLENETRTLIKIPGHGREGSVVISGHDRQGILSAKTRLDLLIESARRRQPFTHFISIPVNSQPI QDKFIEFKDDVVRFCSGDRGVDDTIFQNPHKLHLTIGTMPLLDKSEIDKAKAVLQQCKEELIAYDYIGHGGIT CQLRGLEYMNDDPGEVDVLYAKIQLQDNSDRLQCLADQLVNGFCESGLMNREHDRVKLHVTVMNTLMR KDPTGAGVTVSSGNKKPIKDRESFDANNILKLYGDYDFGPYQINTIHLSQRYSTSQDGYYACEDKIDF
ApASCC1 full-length with His Tag (Biochemical assay)	MHHHHHHENLYFQSNAMDILRPPIICIDGFRFRRNPVQDKSPREYEEEEFGTVGSERVYCDDICDTLGEQIQ ETENGYKLELEIPSAYYKYIIGKKGETKKRLENETRTLIKIPGHGREGSVVISGHDRQGILSAKTRLDLLIESARRR QPFTHFISIPVNSQPIQDKFIEFKDDVVRFCSGDRGVDDTIFQNPHKLHLTIGTMPLLDKSEIDKAKAVLQQC KEELIAYDYIGHGGITCQLRGLEYMNDDPGEVDVLYAKIQLQDNSDRLQCLADQLVNGFCESGLMNREHD RVKLHVTVMNTLMRKDPTGAGVTVSSGNKKPIKDRESFDANNILKLYGDYDFGPYQINTIHLSQRYSTSQD GYYACEDKIDF
ApASCC1Δ40 (crystallography and SAXS)	SNARVYCDDICDTLGEQIQETENGYKLELEIPSAYYKYIIGKKGETKKRLENETRTLIKIPGHGREGSVVISGHD RQGILSAKTRLDLLIESARRRQPFTHFISIPVNSQPIQDKFIEFKDDVVRFCSGDRGVDDTIFQNPHKLHLTIGT MPLLDKSEIDKAKAVLQQCKEELIAYDYIGHGGITCQLRGLEYMNDDPGEVDVLYAKIQLQDNSDRLQCLA DQLVNGFCESGLMNREHDRVKLHVTVMNTLMRKDPTGAGVTVSSGNKKPIKDRESFDANNILKLYGDYD FGPYQINTIHLSQRYSTSQDGYYACEDKIDF
HsASCC1 full-length with His Tag (SAXS and Biochemical assay)	MHHHHHHENLYFQSNAMEVLRPQLIRIDGRNYRKNPVQEQTYQHEEDEEDFYQGSMECADEPCDAYEV EQTPQGFRSTLRAPSLLYKHIVGKRGDTRKKIEMETKTSISIPKPGQDGEIVITGQHRNGVISARTRIDVLLDTF RRKQPFTHFLAFFLNEVEVQEGFLRFQEEVLAKCSMDHGVDSSIFQNPKKLHLTIGMLVLLSEEEIQQTCEM LQQCKEEFINDISGGKPLEVEMAGIEYMNDDPGMVDVLYAKVHMKDGSNRLQELVDRVLERFQASGLIVK EWNSVKLHATVMNTLFRKDPNAEGRYNLYTAEGKYIFKERESFDGRNILKLFENFYFGSLKLNSIHISQRFTV DSFGNYASCGQIDFS
HsASCC1Δ133 (crystallography)	MGSSHHHHHHSSGLVPRGSHMTHFLAFFLNEVEVQEGFLRFQEEVLAKCSMDHGVDSSIFQNPKKLHLTI GMLVLLSEEEIQQTCEMLQQCKEEFINDISGGKPLEVEMAGIEYMNDDPGMVDVLYAKVHMKDGSNRLQ ELVDRVLERFQASGLIVKEWNSVKLHATVMNTLFRKDPNAEGRYNLYTAEGKYIFKERESFD GRNILKLFENFYFGSLKLNSIHISQRFTVDSFGNYASCGQIDFS

Supplemental Table S7. Amino acid sequences for ASCC1 proteins used in this paper.

ASCC1 Supplemental Table S8

	ApASCC1Δ40	HsASCC1 Δ133
PDB Code	8TUK	8TLY
Wavelength	0.97946	0.97946
Resolution range	35.87 - 1.15 (1.22 - 1.15)	35.73 - 2.8 (3.0 - 2.8)
Space group	P 1 21 1	P 21 21 21
Unit cell	35.88 51.60 81.93 90 91.44 90	57.88 62.11 136.22 90 90 90
Total reflections	657320	82297
Unique reflections	101451 (13759)	12679 (1990)
Multiplicity	6.5 (5.0)	6.5 (6.5)
Completeness (%)	95.7 (80.8)	99.6 (99.6)
Mean I/sigma(I)	16.5 (2.0)	9.7 (1.72)
Wilson B-factor	15.76	44.11
R-meas	5.1 (70.1)	18.9 (97.6)
CC1/2	99.9 (82.9)	99.2 (79.6)
Reflections used in refinement	98954 (7189)	12158 (1115)
Reflections used for R-free	1929 (141)	599 (56)
R-work	0.15 (0.25)	0.21 (0.29)
R-free	0.17 (0.27)	0.26 (0.33)
Number of non-hydrogen atoms	2987	3693
-macromolecule atoms	2471	3536
-ligand atoms	70	0
-solvent atoms	486	157
Protein residues	290	437
RMS(bonds)	0.008	0.002
RMS(angles)	0.87	0.50
Ramachandran favored (%)	97.2	96.1
Ramachandran allowed (%)	2.8	3.9
Ramachandran outliers (%)	0.0	0.0
Rotamer outliers (%)	0.4	0.0
Clashscore	4.0	3.9
Average B-factor (all)	29.5	43.8
-macromolecules only	27.1	43.8
-ligands only	38.2	
-solvent only	41.0	43.9

Supplemental Table S8. Crystallographic statistics. Highest resolution shell is in parentheses.

ASCC1 Supplemental Table S9

RNA	uuuu**CACA**uuuu
RNA	uuuu**AACA**uuuu
RNA	uuuu**CGCA**uuuu
RNA	uuuu**AGCA**uuuu
RNA	uuuu**CACG**uuuu
RNA	uuuu**AACG**uuuu
RNA	uuuu**CGCG**uuuu
RNA	uuuu**AGCG**uuuu
DNA	tttt**CGCG**tttt

Supplemental Table S10. RNA oligonucleotides used in this study.

ASCC1 Supplemental Table S10

SAXS Data	ApASCC1 Δ40	ApASCC1 Full-length	Hs ASCC1 Full-length
Simple Scattering Dataset #	XSVL1FQL	XSA5YQYE	XSWQ5ILA
Beamline	ALS SIBYLS 12.3.1	ALS SIBYLS 12.3.1	ALS SIBYLS 12.3.1
Data Collection Mode	MALS-SEC-SAXS	MALS-SEC-SAXS	MALS-SEC-SAXS
SEC-SAXS Data reduction	SCÅTTER	RAW	SCÅTTER
Porod Debye (Px)	4.0	4.0	3.0
Low q ($Å^{-1}$)	0.030	0.025	0.026
High q ($Å^{-1}$)	0.33	0.36	0.35
Reciprocal Rg (Å)	25.33	30.7	28.62
Reciprocal I(0 (detector units)	18	4336	74.25
Real space Rg (Å)	25.85	32.33	30.18
Real space I(0 (detector units)	18.21	4422	75.73
Dmax (Å)	90.38	119.23	118.55
Energy (keV)	12	12	12
Temp (°C)	20	20	20
SAXS MW (kD)	35	46	44
MALS MW (kD)	37.7	49	44.5
Theoretical MW(kD)	36.5	41	44

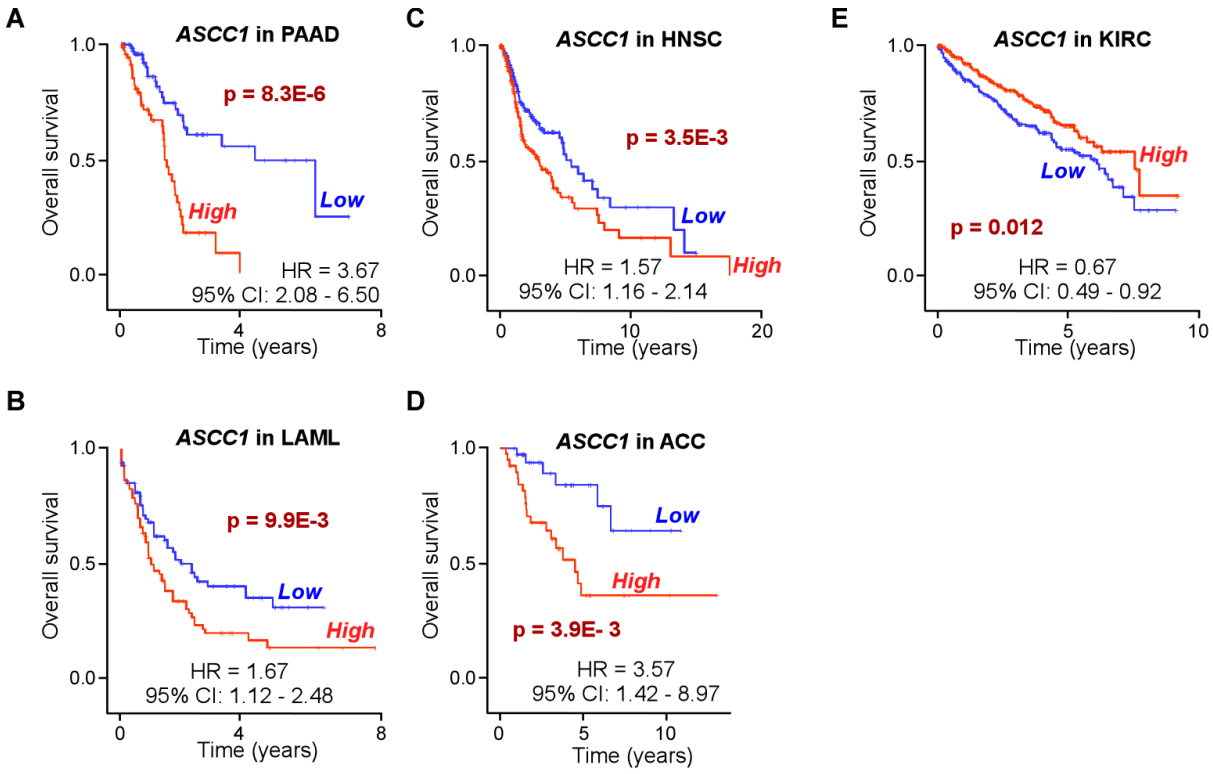
Supplemental Table S9. SAXS details and statistics.

ASCC1 Supplemental Table S11

Human Iso1	EA	Human Iso2	EA	Alvi	EA
Q201K	99.87	Q173	78.47	Q176	78.69
H207L	96.67	H179	96.9	H182	96.62
C231G	94.74	C203	91.98	C206	92.85
I345S	91.58	I317	90.93	I319	91.94
G58R	91.16	G58	87.78	G61	88.31
D283Y	90.46	D255	91.67	D259	83.72
P259H	90.26	P231	92.47	P235	91.28
C231S	88.99	c203	84.15	C206	86
G106E	88.45	G78	89.69	G81	92.4
R156Q	82.96	R128	82.56	R131	82.55
G194W	82.06	G166	79.3	G169	75.25
S67R	81.68	S67	75.33	S70	67.23
E113K	81.55	E85	82.7	E88	83.77
P203L	81.17	P175	81.58	P178	81.16
H207Y	80.62	H179	79.33	H182	79.39
H163Y	80.6	H135	79.31	H138	79.37
K205N	72.67	K177	68.79	K180	68.96
G211R	71.4	G183	74.37	G186	76.19
F181L	69.63	F153	67.46	F156	66.2
E298G	68.1	E270	61.77	E274	56.39
C189Y	67.57	C161	65.81	C164	57.68
L165V	63.36	L137	65.79	I140	44.63
Y70C	60.94	Y70	59.78	Y73	59.76
G241W	58.53	G213	43.17	G217	60.37
D31G	53.99	D31	33.34	E31	54.27
T134A	53.89	T106	44.21	S109	34.67
M309I	53.81	M281	46.51	M285	51.17
A267T	52.6	A239	50.63	A243	50.07
T116S	50.68	T88	39.86	T91	40.71
S67C	50.63	S67C	46.59	S70	42.54
L169I	49.89	L141	55.29	V144	43.99
V52M	49.55	V52	51.74	I55	38.61
L151F	49.27	L123	40.11	L126	42.29
E226V	47.06	E198	45.37	A201	23.36
V185I	46.49	V157	43.1	V160	42.26
I295T	45.37	I267	40.87	M271	28.97
R138Q	45.19	R110	45.78	R113	48.99
E218K	43.6	E190	47.32	K193	silent
R145Q	43.23	R117	35.8	K120	35.89
G260D	42.44	G232	30.17	G236	28.32
I210L	41.28	I182	45.41	I185	42.74
G241E	41.18	G214	43.79	G217	46.98
K331E	41.12	K303	31.73	K305	33.07
R288C	40.26	R260	40.52	G264	36.02
G177R	39.61	G149	19.15	K152	6.45
G177E	36.98	G149	15.19	K152	16.45
P160S	34.89	P132	44.31	P135	30.73
D263N	32.18	D235	29.84	D239	30.63
K188E	31.61	K160	36.83	F163	44.92
I131 V	31.27	I103	30.58	v106	silent
R343Q	31.2	R315	29.99	N317	34.2
R5H	30.43	R5	32.84	R5	31.85
F167L	29.04	F139	26.69	I142	21.39
R60Q	28.58	R60	11.6	K63	20.56
P317H	28.17	P289	23.41	P293	19.03
V195I	25.96	V167	28.52	V170	21.88
K335N	22.74	K307	14.64	K309	12.32
S39Y	21.15	S39	7.12	S39	25.63
K117R	20.25	K89	16.32	R92	silent
S39C	19.9	S39	8.99	S39	22.42
M261I	19.27	M233	17.78	E237	52.33
D44N	18.5	D44	14.08	D45	25.6
A328V	15.09	A300	18.79	S302	14.09
Q222K	12.07	Q194	15.17	D197	21.22
S198N	12.04	S170	7.15	T173	11.4
R10K	11.05	R10	6.07	C10	28.82
D128E	10.16	D100	5.88	E103	silent
Q229H	9.82	Q201	9.57	Q204	8.98
A250S	9.4	A222	7.71	R226	9.91
S121R	8.49	S93	2.7	K96	0.44
Q57K	7.49	Q57	2.51	N60	4.22
L179M	4.74	L151	4.64	I154	5.71
S121N	3.49	S93	0.24	K96	1.63
P20Q	2.86	P20	1.15	P20	1.06

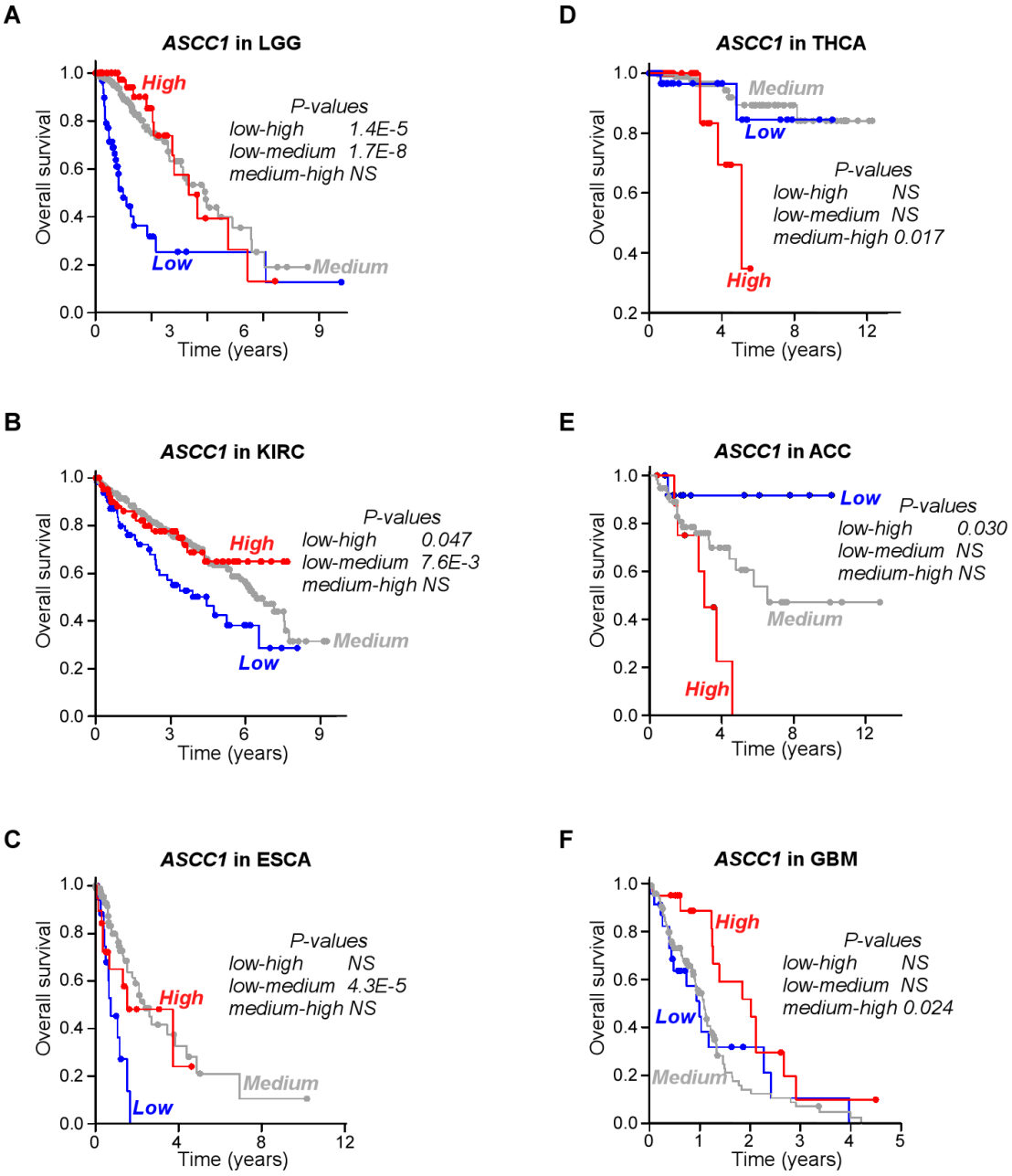
ASCC1 Supplemental Table S11. Table of Variants of Unknown Significance (VUS). VUS are listed based on human and *Alvinella* sequences, with their respective Evolutionary Action (EA) scores, a prediction on severity level. Red to green coloring is based on predicted severity to no impact, respectively.

ASCC1 Supplemental Fig S1



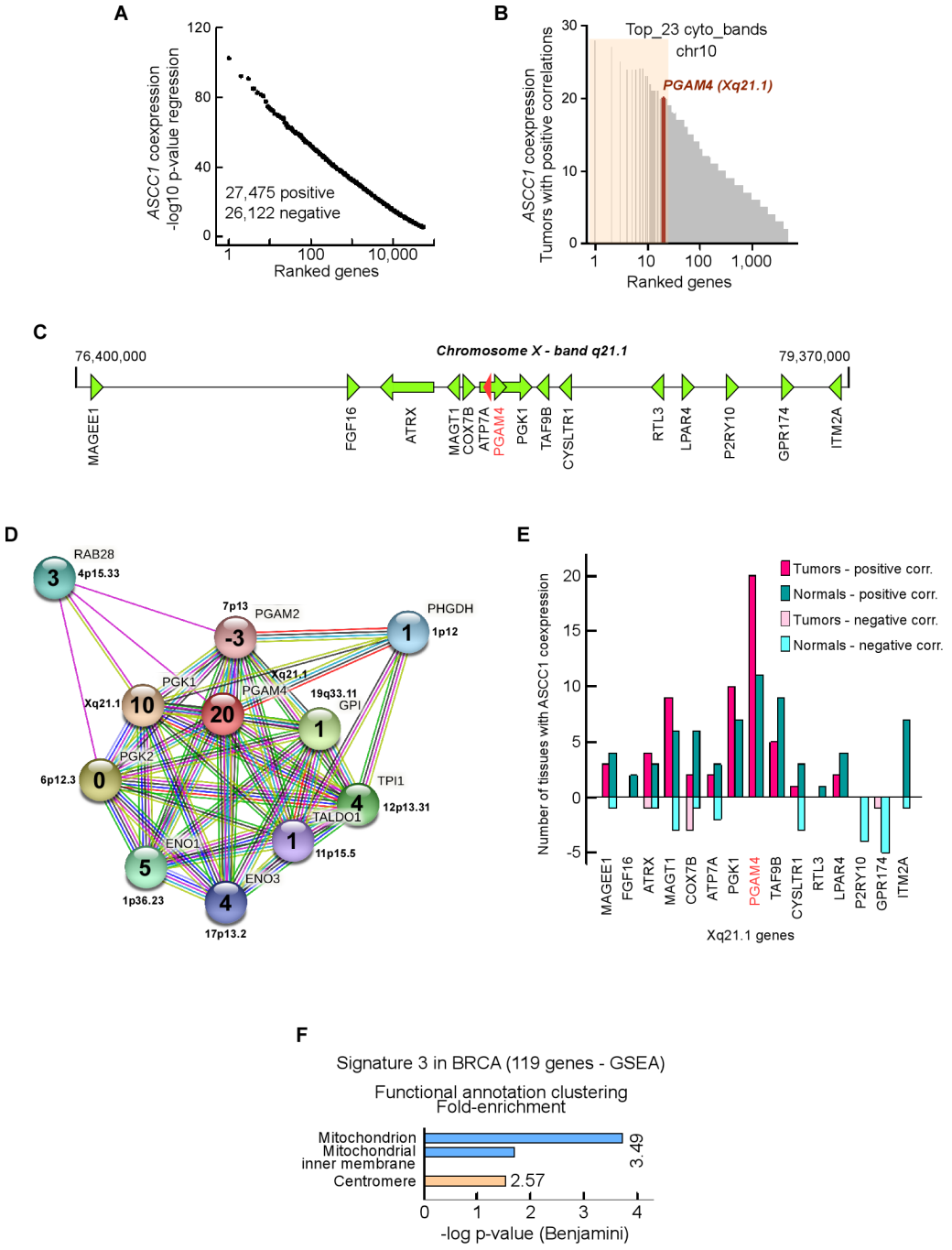
Supplemental Figure S1 . Kaplan-Meier survival curves in TCGA tumor types in which patients with High (above mean) versus Low (below mean) *ASCC1* mRNA levels displayed significant differences in survival. P-values from log-rank tests; HR, hazard ratios; CI, confidence intervals for HR. A) PAAD, pancreatic adenocarcinoma. B) LAML, acute myeloid leukemia. C) HNSC, head-neck squamous cell carcinoma. D) ACC, adrenocortical carcinoma. E) KIRC, kidney renal clear cell carcinoma.

ASCC1 Supplemental Fig S2



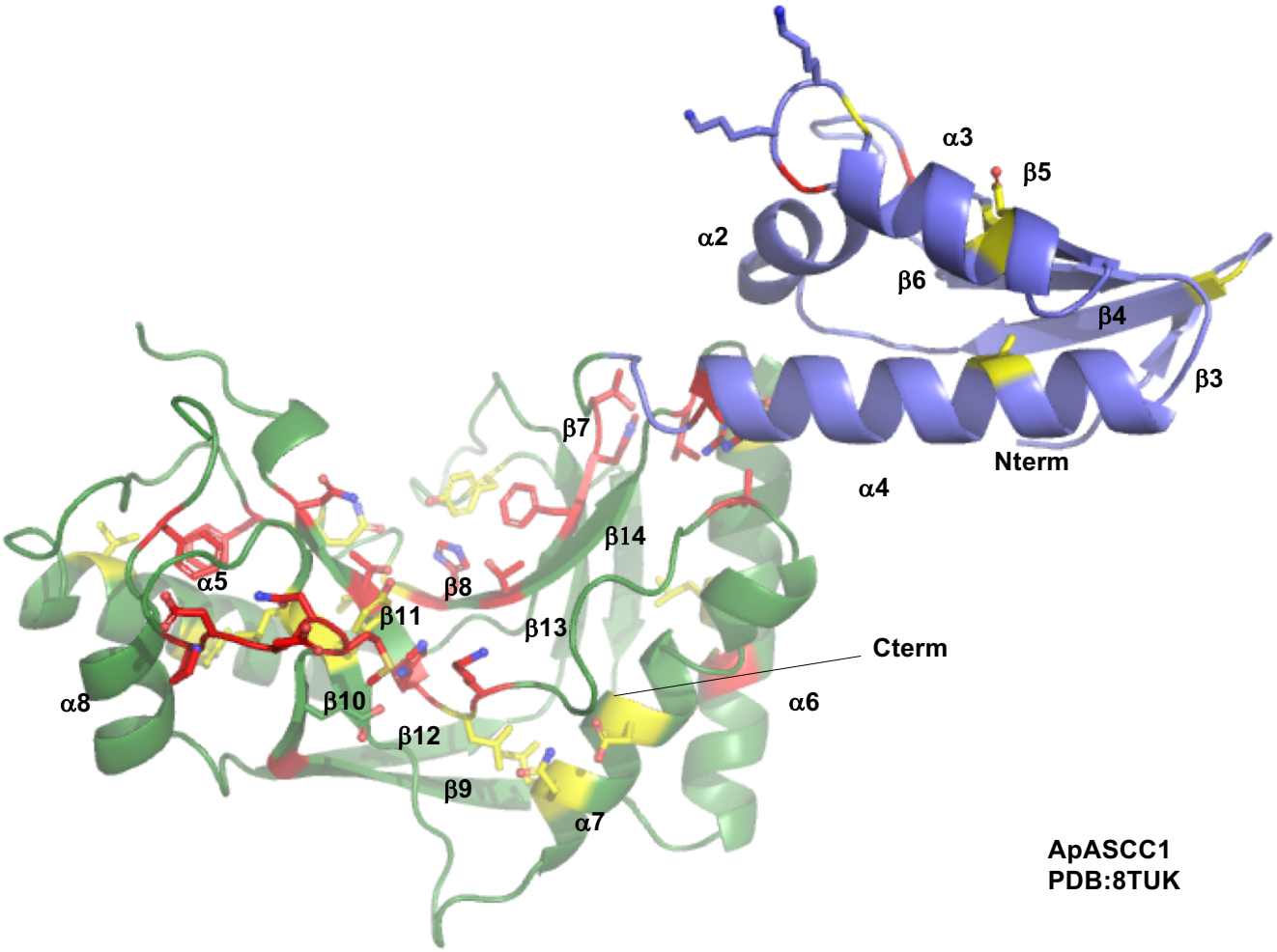
Supplemental Figure S2 . Kaplan-Meier survival curves in TCGA tumor types in which patients with High (above mean+1SD), versus Low (below mean-1SD), versus Medium (between mean-1SD and mean+1SD) *ASCC1* mRNA levels displayed significant differences in survival. P-values from Holm-Sidak corrected log-rank tests. A) LGG, low grade glioma .B) KIRC, kidney renal clear cell carcinoma. C) ESCA, esophageal carcinoma. D) THCA, thyroid cancer. E) ACC, adrenocortical carcinoma. F) GBM, glioblastoma.

ASCC1 Supplemental Fig S3



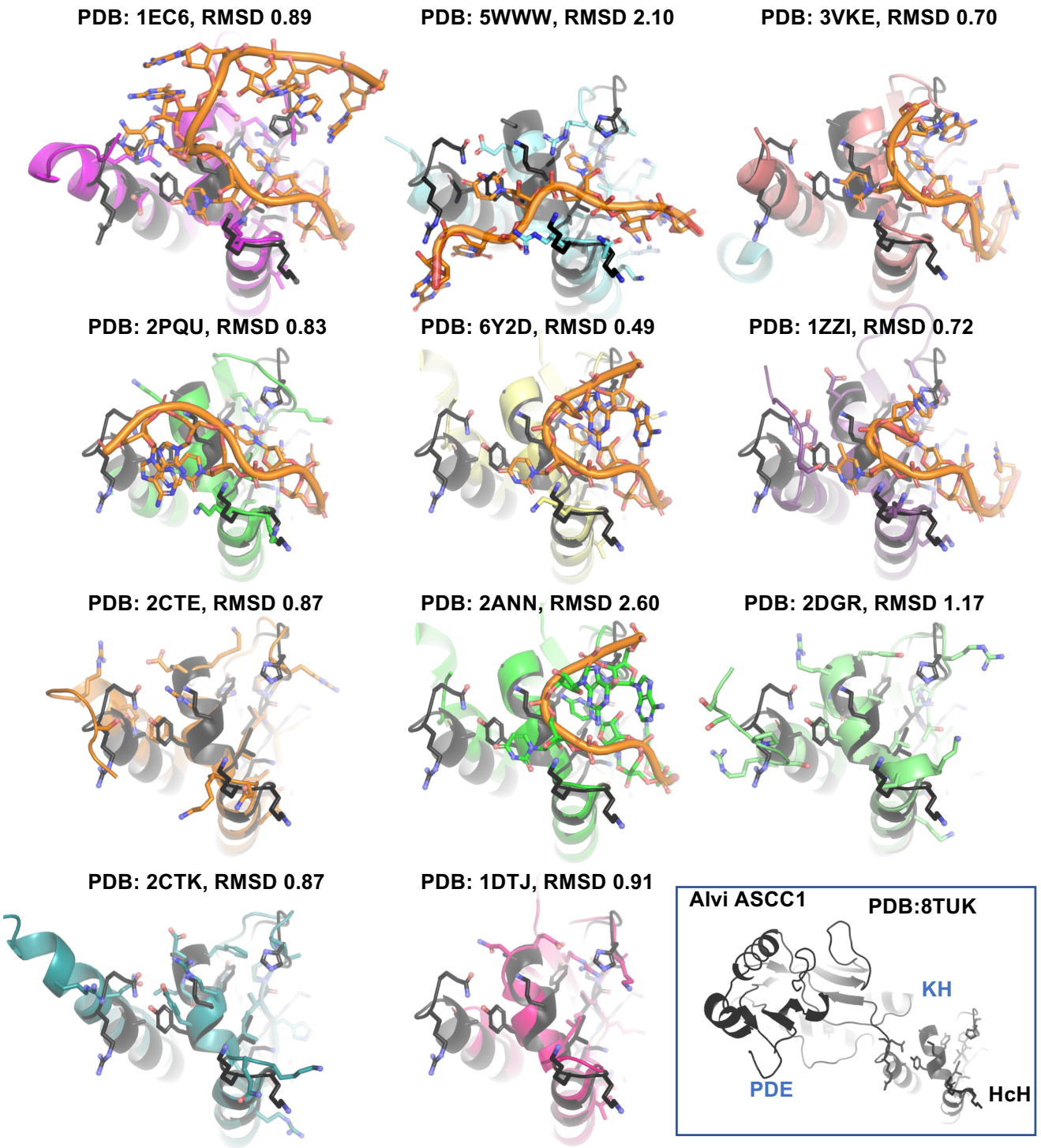
Supplemental Figure S3 . *ASCC1* is coexpressed with genes on chromosome Xq21.1. A-B) *ASCC1* coexpression. A, Ranked p-values (threshold set at a Bonferroni-corrected value of 2.5 x 10^{-6}) for regression coefficients between *ASCC1* and all other genes in 33 tumor types (>660,000 regression total). B, Ranked number of tumor types in which positive correlations in tumors were observed. *Orange highlight*, the top 23 genes were located on chromosome 10, where *ASCC1* resides, expect *PGAM4* (*maroon*). C) Cartoon of transcriptional units on chromosome Xq21.1 flanking *PGAM4*. D) High-confidence protein-protein interaction (PPI) network for PGAM4. Numbers indicate the number of tumors with positive correlation between *ASCC1* and the test protein-coding gene was found. *Lines*, number of algorithms or experimental methods supporting a PPI. E) Number of tissues where significant co-expression was observed between *ASCC1* and genes flanking *PGAM4* for both tumors and matched controls. Negative values indicate negative correlation. F) GSEA (gene set enrichment analysis) of all genes found to be significantly ($p<0.002$; Fisher's exact test) associated with Signature 3 (but not Signatures 3 + 13) in BRCA when expressed at high levels (i.e. above the mean). *x-axis*, p-values from Benjamini-corrected Fisher's exact tests; *y-axis*, gene ontology terms. *Fold-enrichment*, fold-enrichment relative to random occurrence for clusters of related gene ontology terms.

Supplemental Fig S4



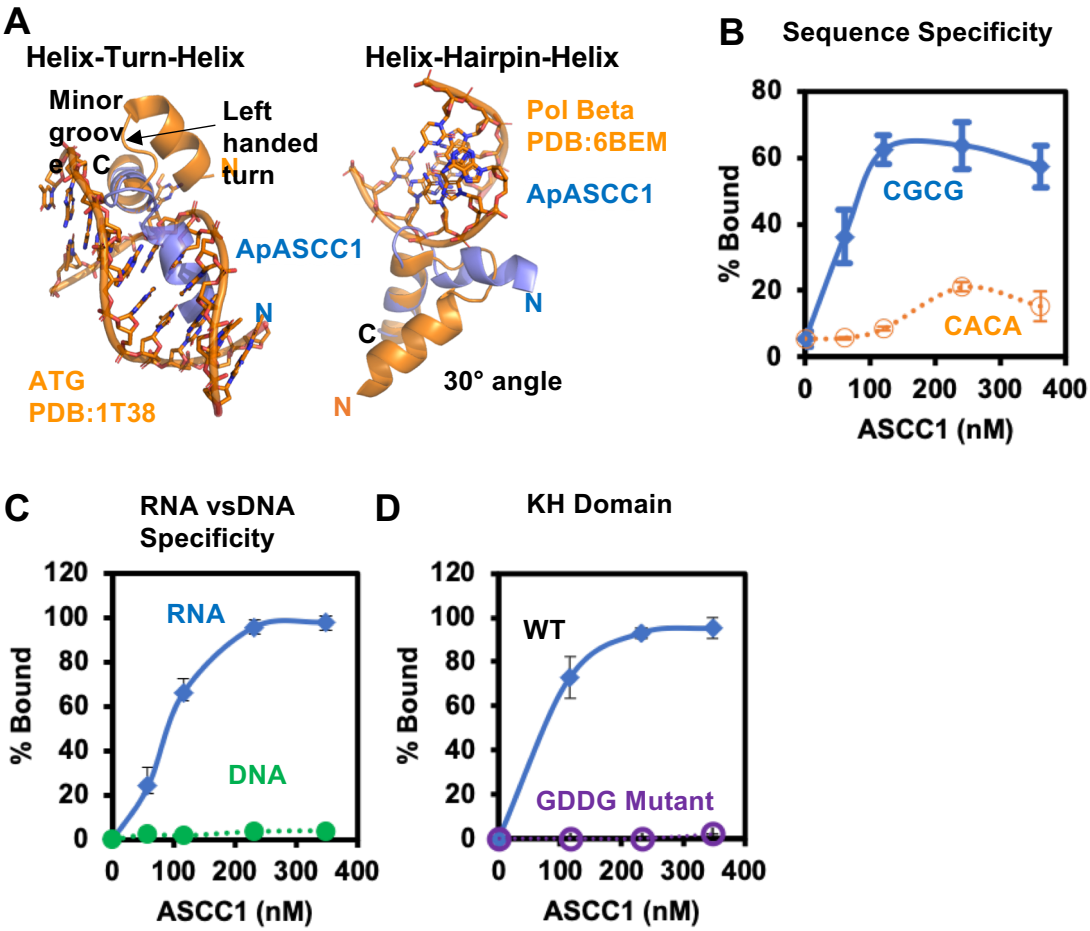
Supplementary Figure S4. Secondary structure numbering of Δ40 ApASCC1 corresponding to Fig. 3.

Supplemental Fig S5



Supplemental Figure S5. Overlay of top DALI and Foldseek KH domains onto ApASCC1 KH domain (black). Orientation is such that the highly conserved HcH is at the bottom and the PDE domain would be on the left side, as shown by inset of a zoomed out ASCC1. KH domains are known as RNA-binding domains. With nucleic acids colored orange, the nucleic acid path can vary from KH domain to KH domain. Varying steric clash between Alvi ASCC1 KH domain and nucleic acid bases suggest ASCC1 may bind RNA distinctly.

Supplemental Figure S6



Supplemental Figure S6. KH Domain Structure and ASCC1 RNA binding specificity.

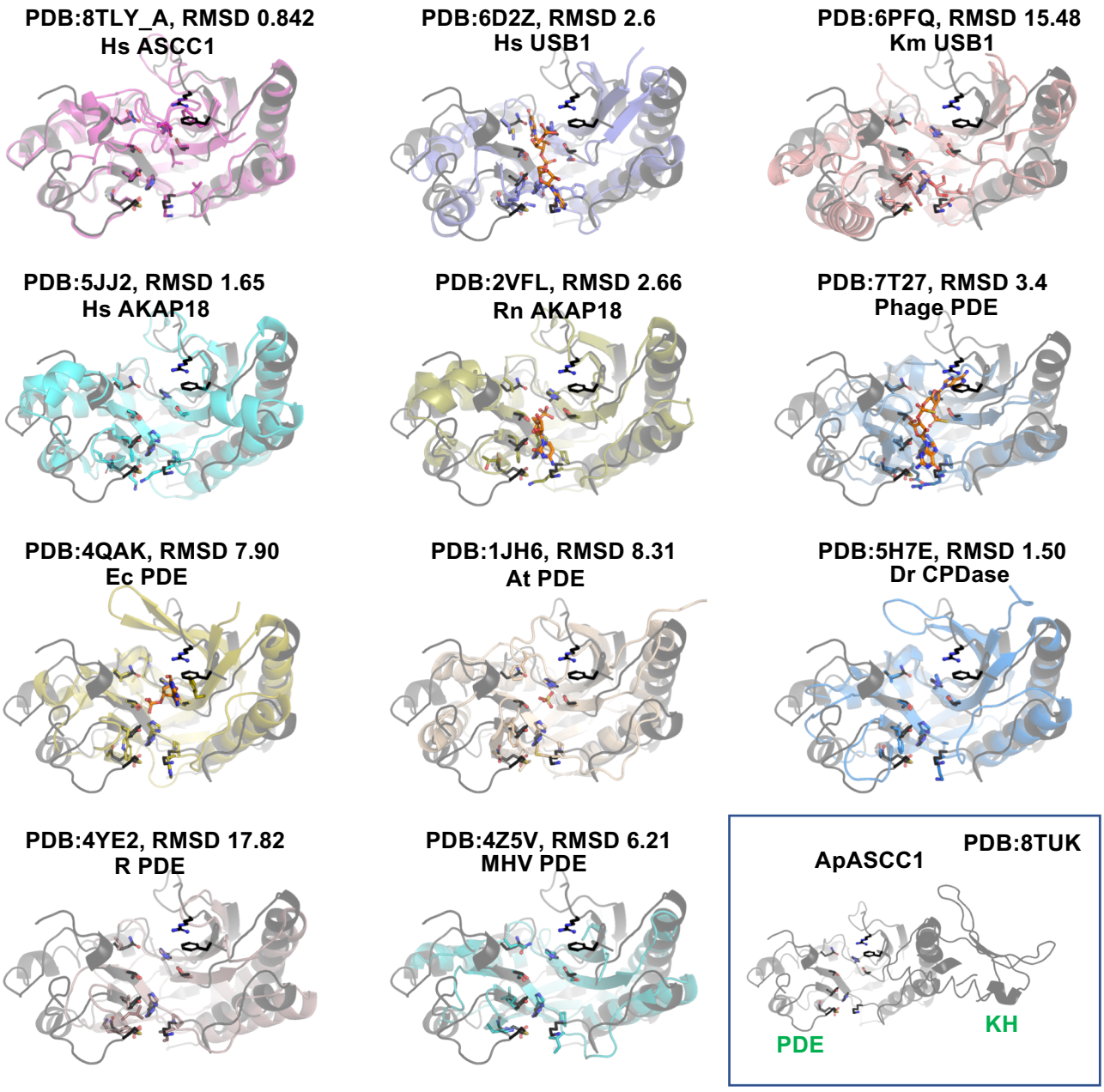
A) Representative Helix-Turn-Helix motif in ATG or a representative Helix-Hairpin-Helix motif in Pol Beta overlaid onto the C-terminal helices of ApASCC1 highlights their structural and recognition differences from the Helix-Clasp-Helix motif

B) Quantitation of EMSA showing HsASCC1 binding of RNA CGCG-containing sequence vs CACA-containing sequence, in Figure 4C. Dose-response curve were generated by plotting the percentage of bound RNA against total radiolabeled RNA. Error bars show ± standard deviation from the mean; n = 3 independent experiments.

C) Quantitation of EMSA showing HsASCC1 binding of RNA versus DNA, in Figure 4D. Dose-response curve were generated by plotting the percentage of bound RNA/DNA against total radiolabeled RNA/DNA. Error bars show ± standard deviation from the mean; n = 3 independent experiments.

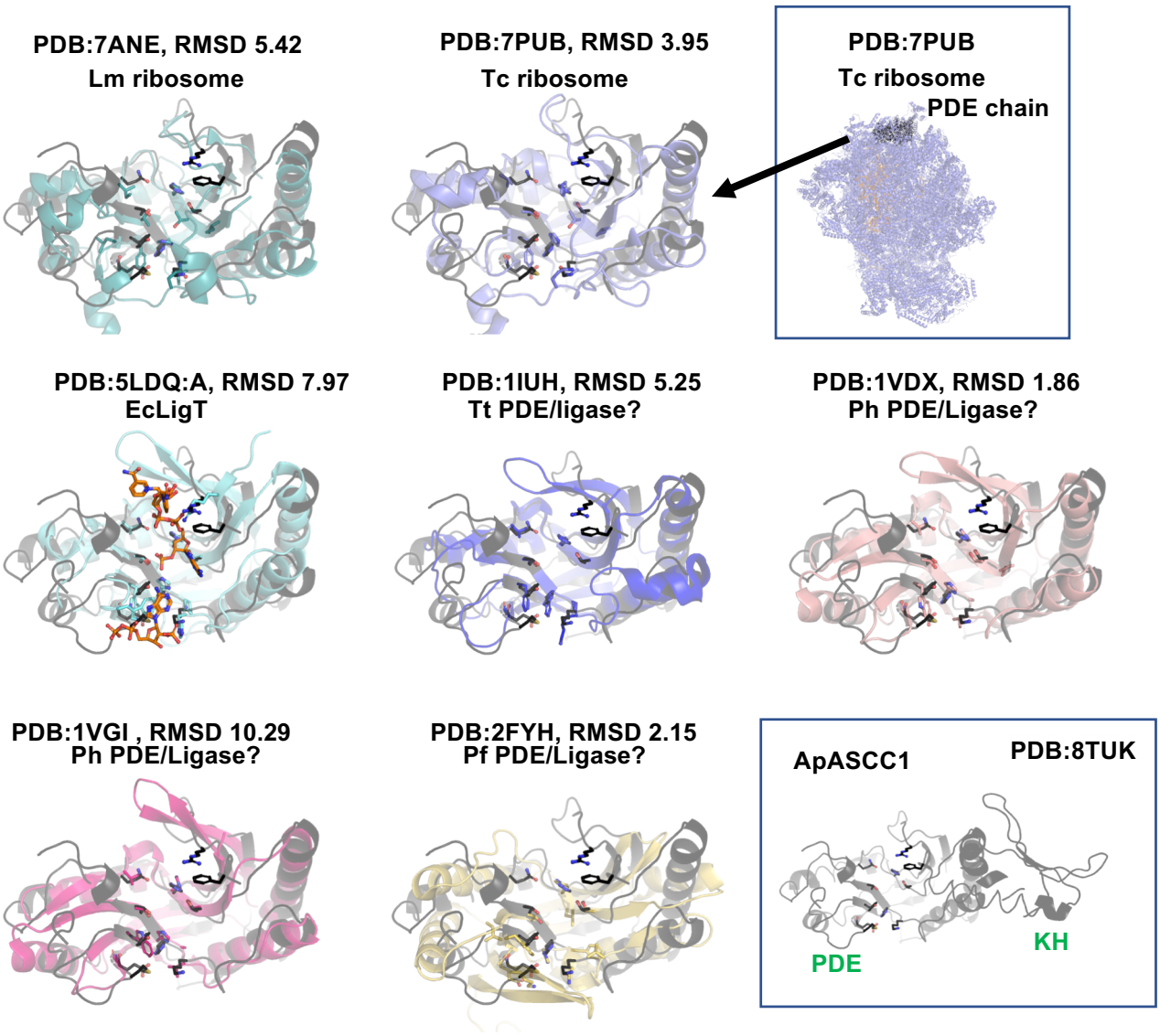
D) Quantitation of EMSA showing wild-type (WT) vs mutant GDDG HsASCC1 binding of CGCG-containing RNA, in Figure 4E. Dose-response curve were generated by plotting the percentage of bound RNA against total radiolabeled RNA. Error bars show ± standard deviation from the mean; n = 3 independent experiments.

ASCC1 Supplemental Fig S7



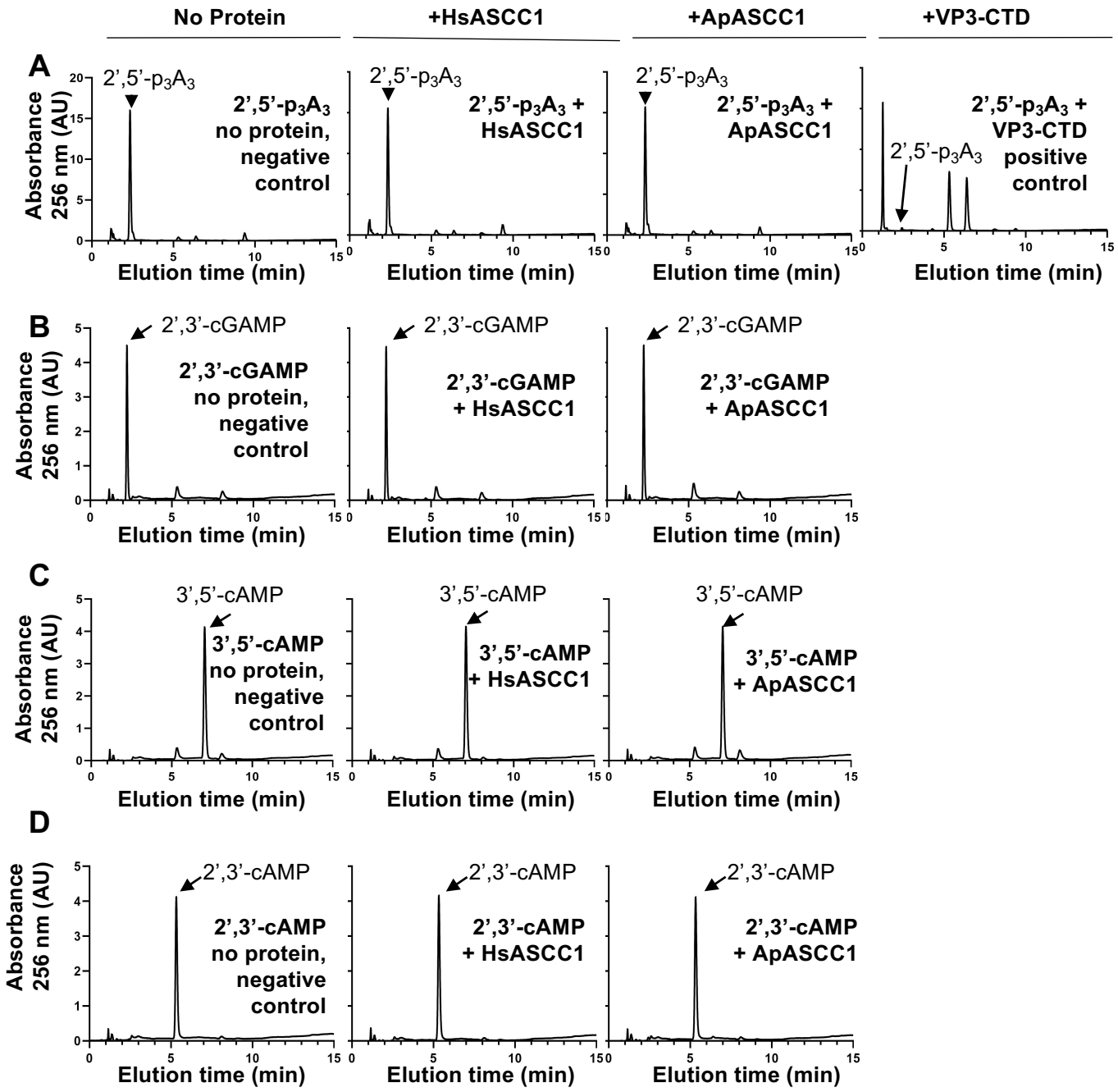
Supplemental Figure S7. Overlay of 2HXT domains onto ApASCC1 PDE domain (black). Orientation is such that the HXT motifs are facing up and KH domain would be on the right side, as shown by inset of a zoomed out ASCC1.

ASCC1 Supplemental Fig S8



Supplemental Figure S8. Overlay of ligase-associated 2HXT domains onto ApASCC1 PDE domain (black). Orientation is such that the HXT motifs are facing up and KH domain would be on the right side, as shown by inset of a zoomed out ASCC1.

ASCC1 Supplemental Fig S9

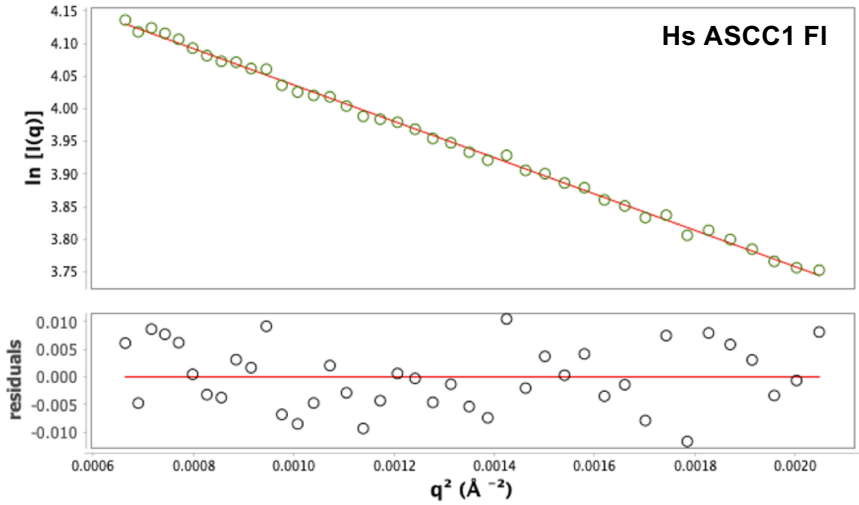


Supplemental Fig S9. ASCC1 had no detectable activity against 2',5'-p_3A_3 , 2',3'-cGAMP, 3',5'-cAMP, or 2',3'-cAMP

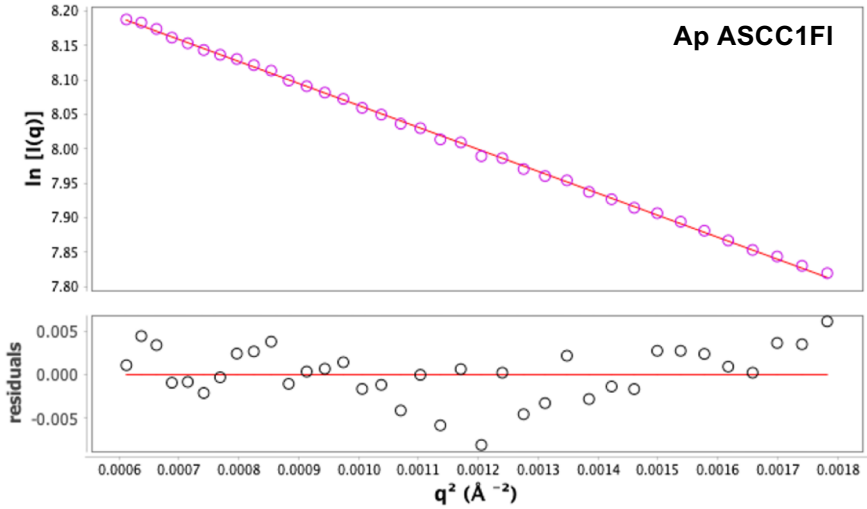
Ten μM of A) 2',5'-p_3A_3 , B) 2',3'-cGAMP, C) 3',5'-cAMP, or D) 2',3'-cAMP was incubated with 1 μM of human- or alvi-ASCC1 for 30 min at 30°C. Substrate without enzyme incubated under similar condition were used as un-degraded, negative control. Percent substrate remaining was calculated by measuring the area under the peaks in the HPLC chromatograms. As a positive control, RVA VP3-CTD degraded 2',5'-p_3A_3 (<2 % of substrate remaining) under the similar conditions. Shown are representative HPLC chromatograms. These experiments were performed in triplicate with similar results.

ASCC1 Supplemental Fig S10

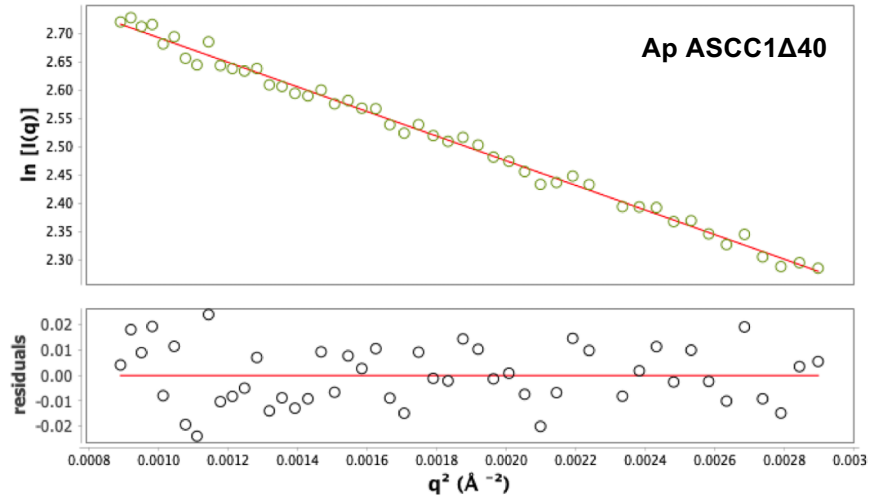
A



B

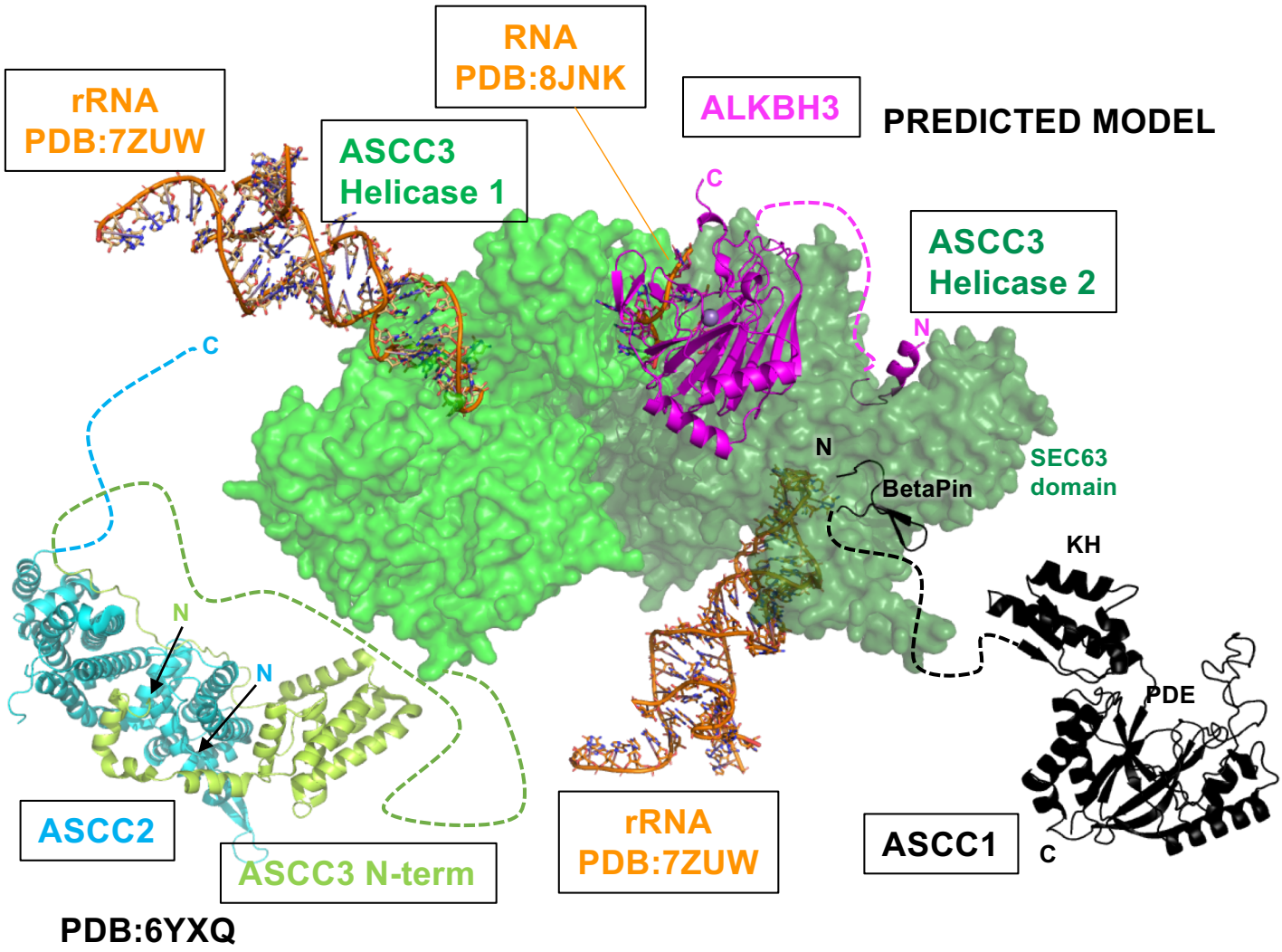


C



Supplemental Figure S10. Guinier analysis of SAXS data A. Hs ASCC1 FI, B. Ap ASCC1 FI and C. Ap ASCC1Δ 40 shows that the Guinier regions are linear, indicating that the protein samples had no significant amounts of large aggregation.

ASCC1 Supplemental Fig S11



Supplemental Figure S11. Model of Alkylation Response complex with ALKBH3, ASCC1, ASCC2, and ASCC3 suggests that ASCC1 and ASCC2 are tethered to ASCC3, while ALKBH3 is predicted to have two ASCC3 interfaces, a helix at the N-terminus and the catalytic domain. Suggestive to co-regulation, the ASCC1 and ALKBH3 tether on opposite sides of the SEC63 domain of the ASCC3 second helicase.

Atomic model was assembled based on pairwise AlphaFold-Multimer runs, HsASCC2/ASCC3 structure (PDB:6YXQ), RNA in complex with ALKBH3 (PDB:8JNK) and rRNA (PDB:7ZUW) from the yeast stalled ribosome structure with RQT (ASCC2, ASCC3, and TRIP4). Predicted interfaces modeled as cartoon are those that reproducibly occurred in all 5 predictions. rRNA position on ASCC3 helicase 1 was based on the yeast structure. Second rRNA position came from overlay of helicase 1 on helicase 2. Dotted lines represent regions with low confidence in the prediction. The two helicase domains were shown for clarity as semi-transparent surface, to highlight interactions with ALKBH3 and ASCC1. Pairwise runs between ALKBH3, ASCC1, and ASCC2 did not produce high confidence interfaces. Surprisingly, human ASCC2 and yeast ASCC2 are not well-conserved, and the human proteins are not predicted in the same interface as shown for the yeast structure (PDB:7ZUW).