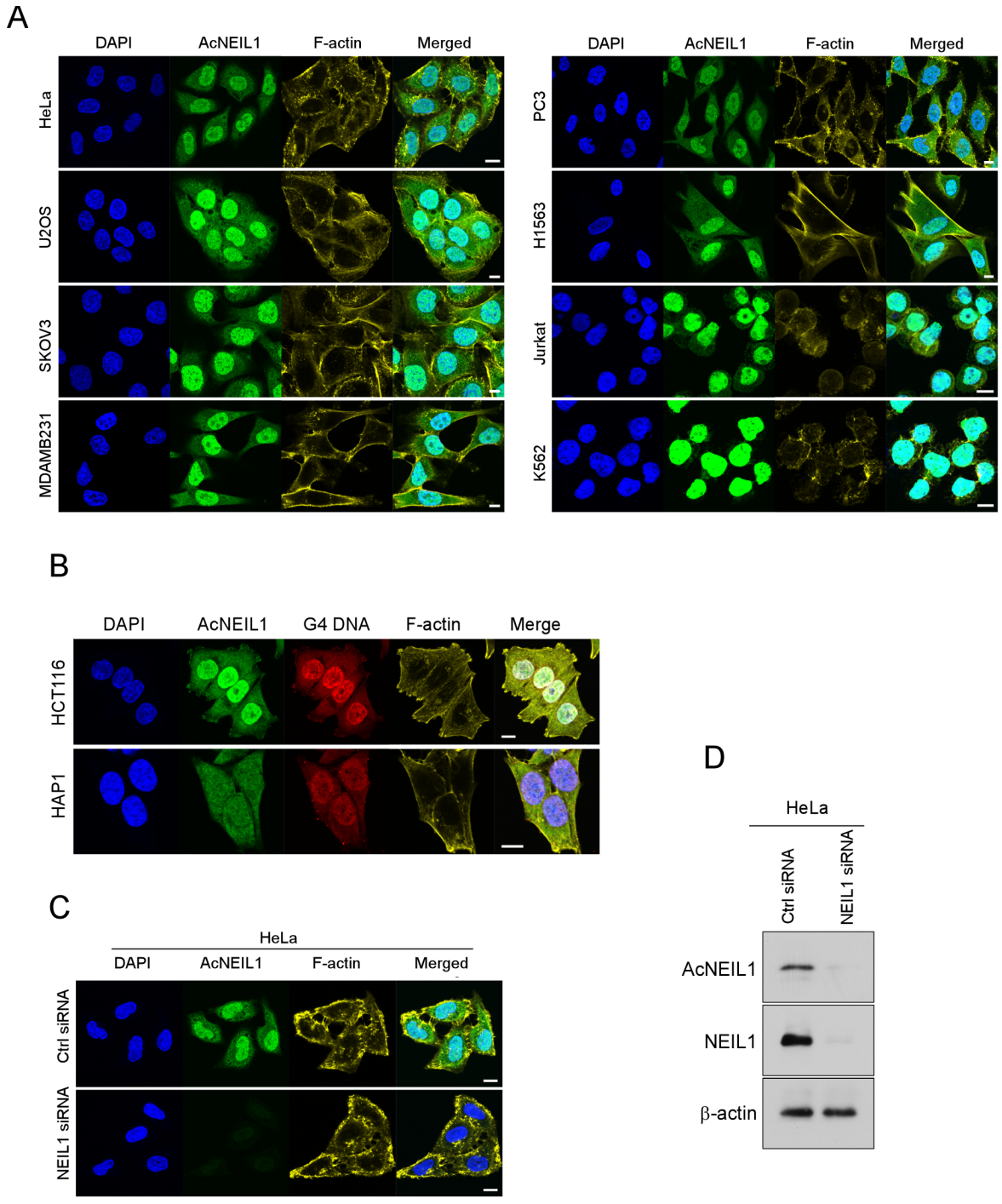
Heritable Pattern of Oxidized DNA Base Repair Coincides with Pre-Targeting of Repair Complexes to Open Chromatin

Albino Bacolla, Shiladitya Sengupta, Zu Ye, Chunying Yang, Joy Mitra, Ruth B. De-Paula, Muralidhar L. Hegde, Zamal Ahmed, Matthew Mort, David N. Cooper, Sankar Mitra, and John A. Tainer

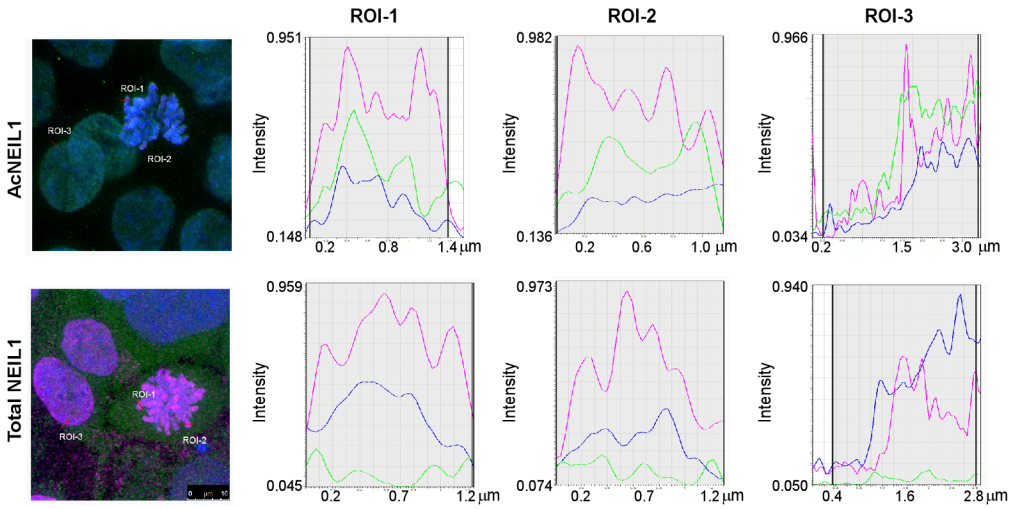
Supplementary Information



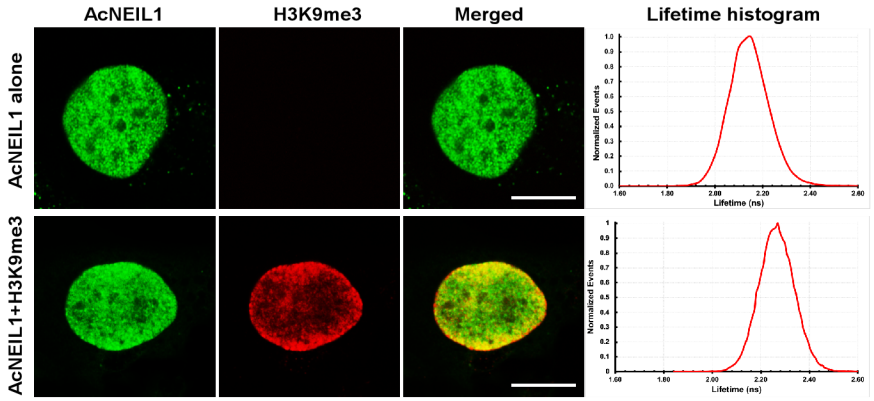
Supplementary Figure 1

Supplementary Figure 1. AcNEIL1 is predominantly localized in the nucleus. (**A**) Confocal microscopy images of eight cell lines stained with DAPI for nuclear DNA (blue), anti-AcNEIL1 specific antibody (green) and phalloidin for cytoplasmic cytoskeleton (yellow) and merged images, displaying nuclear accumulation of AcNEIL1; scale bar, 10 μm. (**B**) Confocal microscopy images of HCT116 and HAP1 cells stained with DAPI for nuclear DNA (blue), anti-AcNEIL1 specific antibody (green), a G4 DNA structure-specific antibody (red) and phalloidin for cytoplasmic cytoskeleton (yellow) displaying cell type-specific colocalization patterns of AcNEIL1. Scale bar, 10 μm. (**C**) Confocal microscopy images of HeLa cells stained with DAPI for nuclear DNA (blue), anti-AcNEIL1 specific antibody (green) and phalloidin for cytoplasmic cytoskeleton (yellow) plus merged images, showing the efficient knockdown of NEIL1 by siRNA; scale bar, 10 μm. (**D**) Immunoblotting of AcNEIL1 and total NEIL1 after siRNA treatment in HeLa cells.

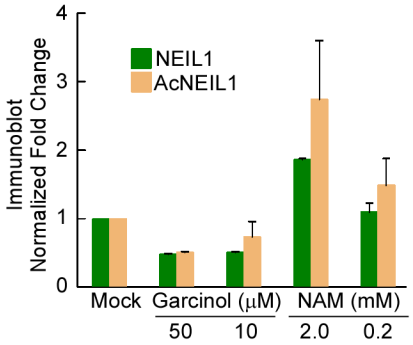
A



B

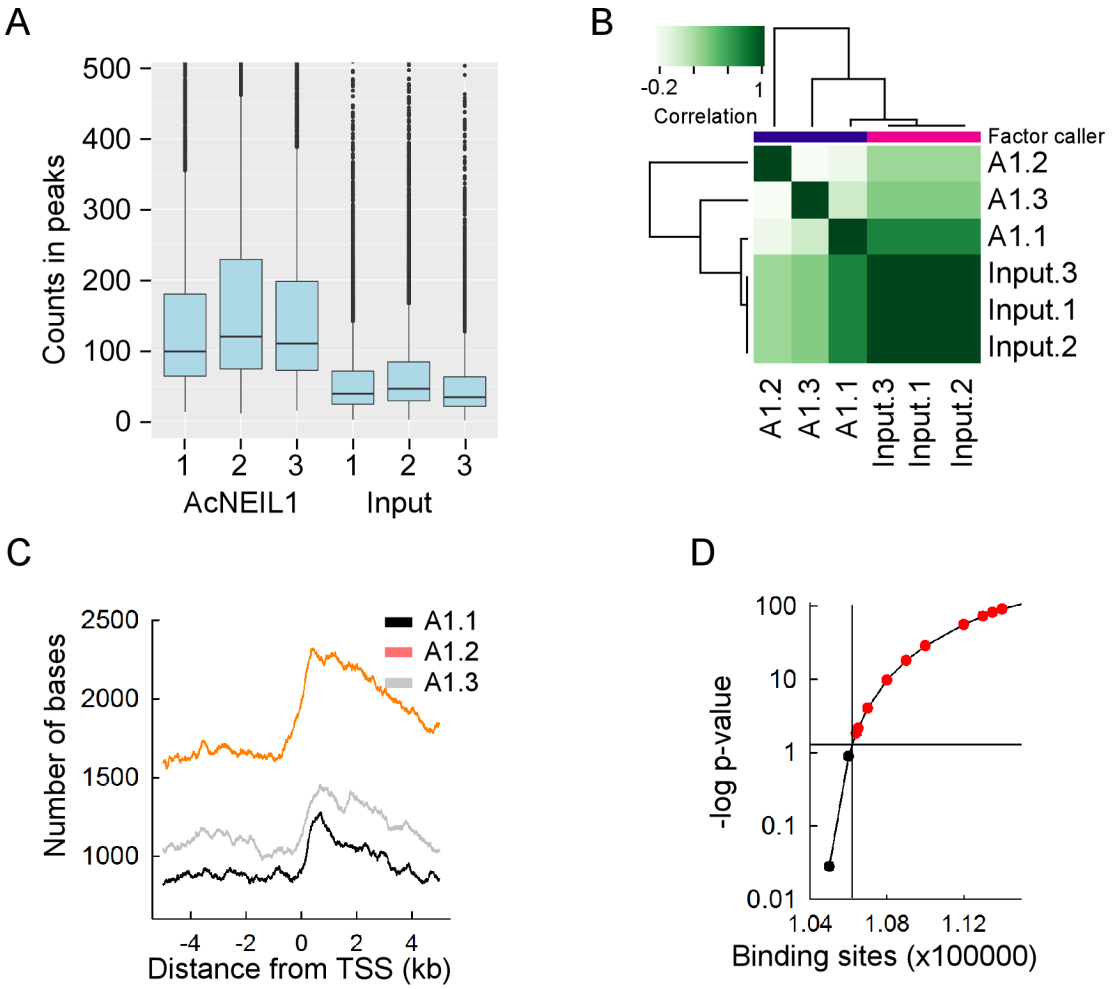


C



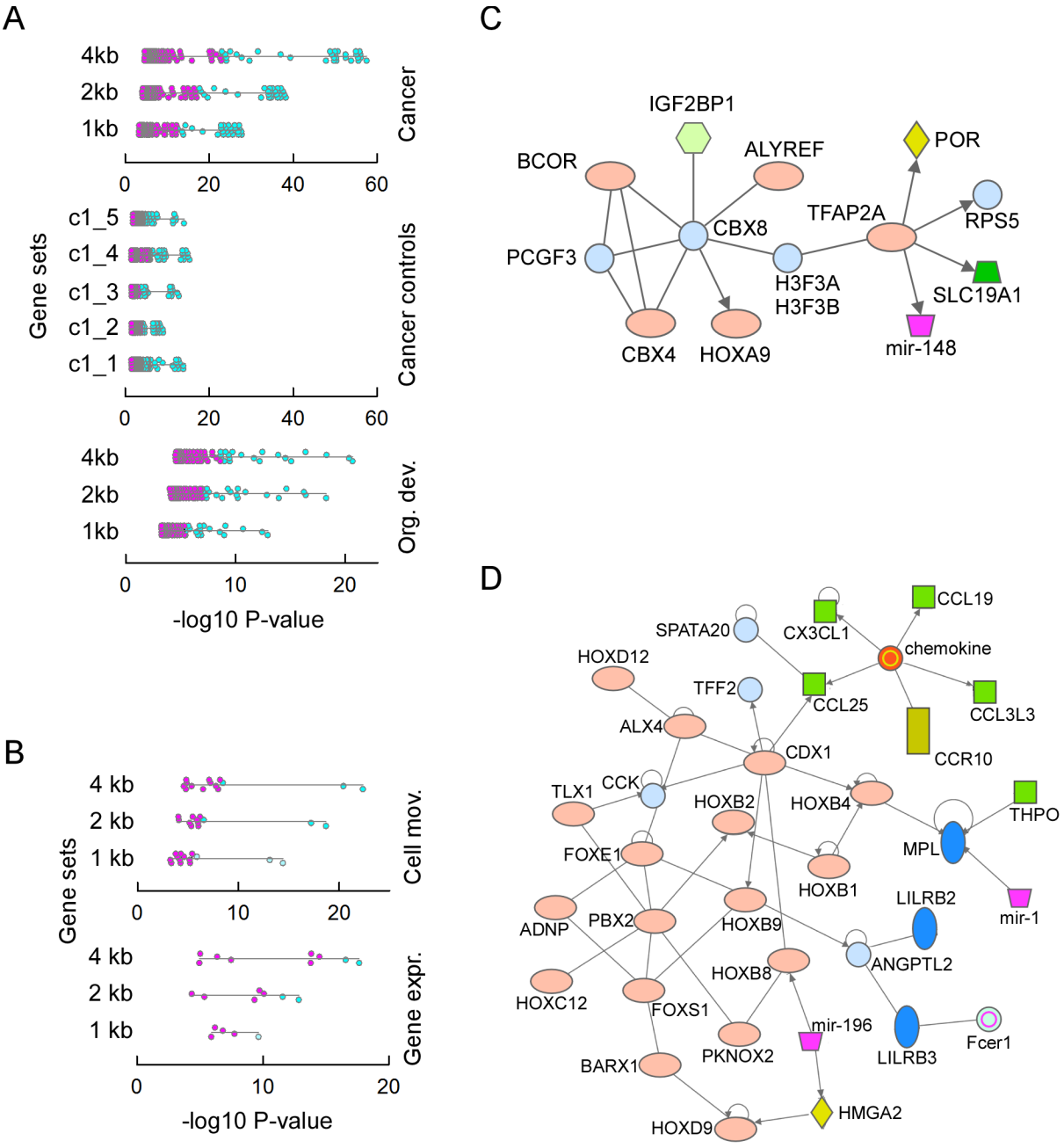
Supplementary Figure 2

Supplementary Figure 2. AcNEIL1 binds within active chromatin domains. (**A**) Co-localization measurements between AcNEIL1 and H3K27Ac using Leica Quantification tools on merged STED images. Three regions of interest (ROI) from respective images were selected and analyzed. Normalized intensities (y axis) for each ROI were plotted against distance (x axis). The results show that AcNEIL1 has a high degree of co-localization with H3K27Ac; green peak (AcNEIL1) superimposing on magenta (H3K27Ac) and DNA (blue). This is in contrast to total-NEIL1, which displays a diffuse distribution with ROI quantification showing limited co-localization with DNA and H3K27Ac. Note the relatively low intensities of total-NEIL1 signal as compared to H3K27Ac and DNA in the same focal point. (**B**) Left, Immunofluorescence staining of AcNEIL1 (top) and AcNEIL1 plus H3K9me3 (bottom) in HCT116 cells. Right, FLIM lifetime histogram of labelled AcNEIL1 (top) and labelled AcNEIL1 and H3K9me3 (bottom). FLIM samples were prepared as described in Materials and Methods; images were captured using a Leica SP8 FALCON and analyzed with a Leica FLIM analysis software. Scale bar, 10 μm. (**C**) Quantitation of immunoblotting (IB) of NEIL1, AcNEIL1 and histone H3 from the chromatin fractions of HCT116 cells. Intensities of IB bands were expressed as normalized fold-change and shown as mean and standard error from the mean from 2 independent experiments.



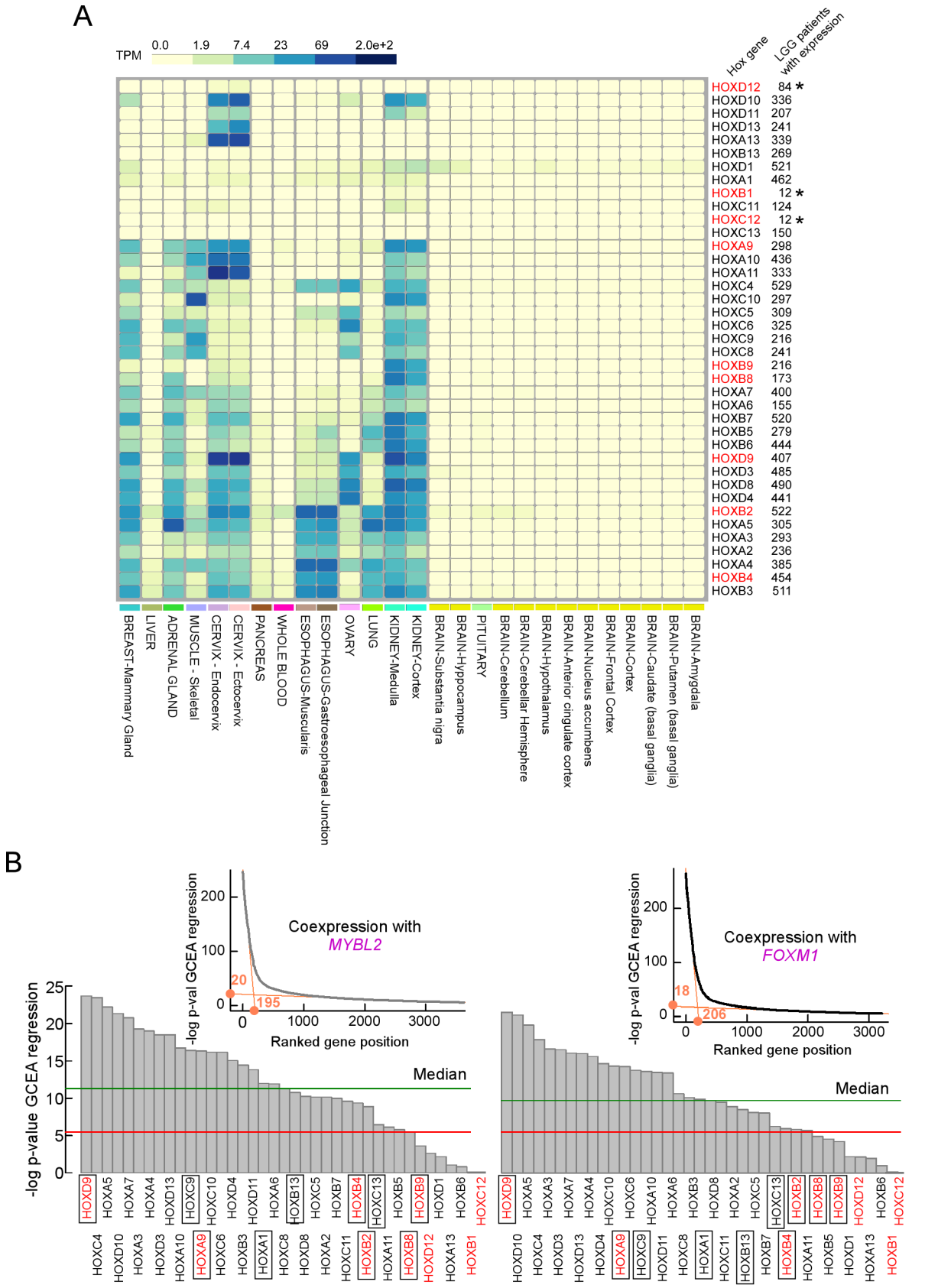
Supplementary Figure 3

Supplementary Figure 3. ChIP-seq quality-control metrics indicate specific and reproducible sequence amplification. (**A**) Bar plot of the number of reads overlapping with peaks for the AcNEIL1 and Input replicates from ChIPQC. (**B**) Peak correlation heatmap for the AcNEIL1 and control replicates based on the co-occurrence of peaks. (**C**) Bp-resolution of peak enrichment at TSSs. PIC summits were extended by ±100 positions and plotted as a function of distance from any TSS within ±5 kb. (**D**) p-values for the PIC overlap between replicates obtained with ChIPpeakAnno. The p-value for two overlapping peak sets is obtained from a hypergeometric test and depends on the total number of binding sites genome-wide. If this number is unknown, parameter "totalTest" will set this equal to: hg38_size * (2%(coding_DNA) + 1%(regulation_Region)) / (2 * average_peak_width), which for the weakest pair, A1.2 and A1.3, corresponded to 113900 and a p-value of 8.2×10^{-90}. The corresponding p-values for A1.1 with A1.3 was 0, and that of A1.1 with A1.2 was 9.7×10^{-114}. The plot displays the p-values for various number of binding sites. Red, significant p-values; black, non-significant p-values. Cross section, minimum number of binding sites for a significant p-value, 106400.



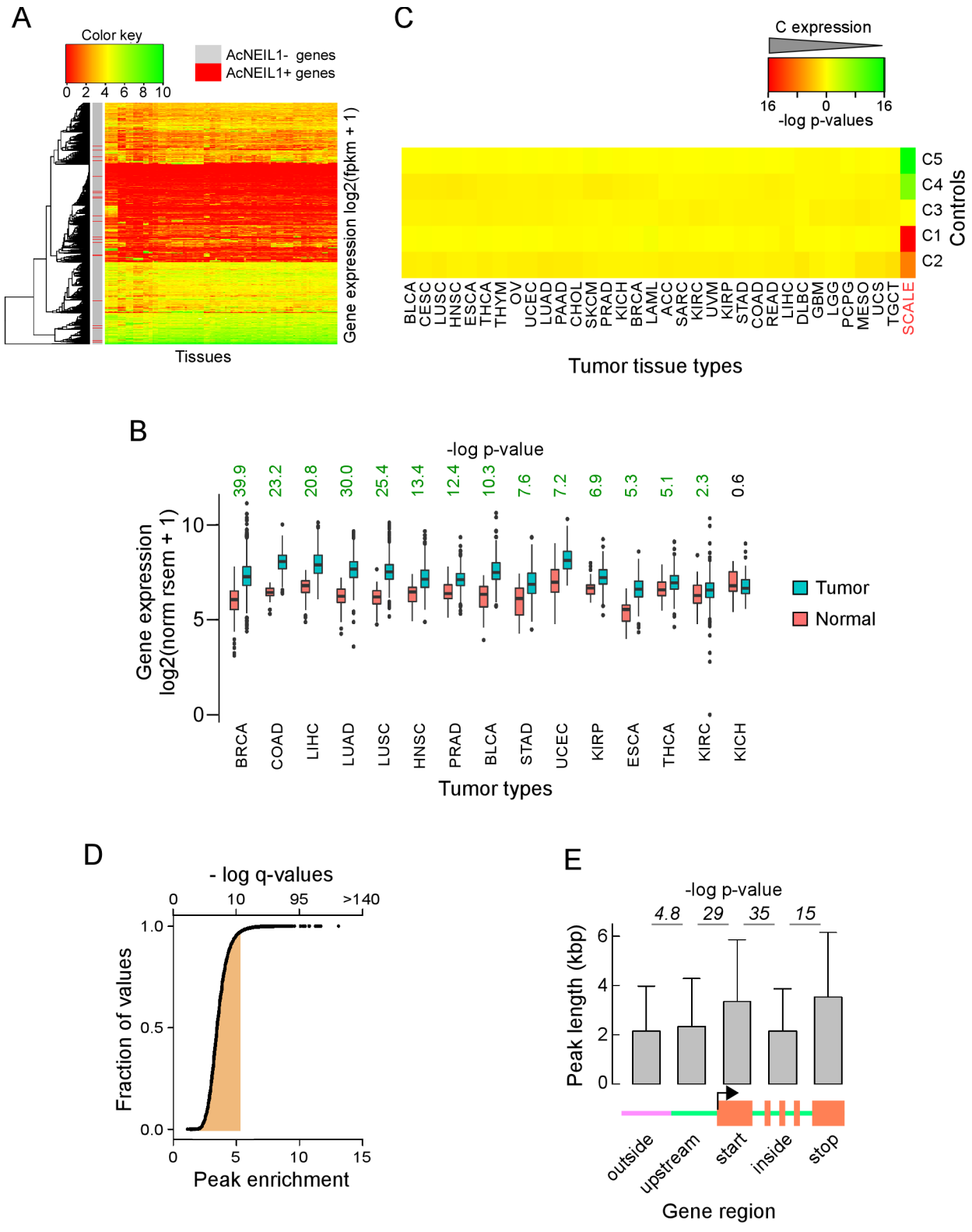
Supplementary Figure 4

Supplementary Figure 4. AcNEIL1 punctuates TSSs of genes involved in cancer and development. (**A**) Ranked p-values from IPA enrichment in gene sets associated with cell movement (top) and gene expression (bottom) for genes with PICs within 1, 2 and 4 kb of TSSs. Upper 25% (blue) p-values are distinguished from the lower 75% (magenta). (**B**) Ranked p-values from IPA enrichment in gene sets associated with cancer (top) or organismal development (bottom) for genes with PICs within 1, 2 and 4 kb of TSSs, and in gene sets associated with cancer for control genes (middle). Upper 25% (blue) p-values are distinguished from the lower 75% (magenta). (**C**) IPA pathway analysis for genes with PICs focused on *Hox*-genes interactions. (**D**) IPA pathway analysis for genes with PICs focused on the interactions of the *CBX8* cancer-related gene.



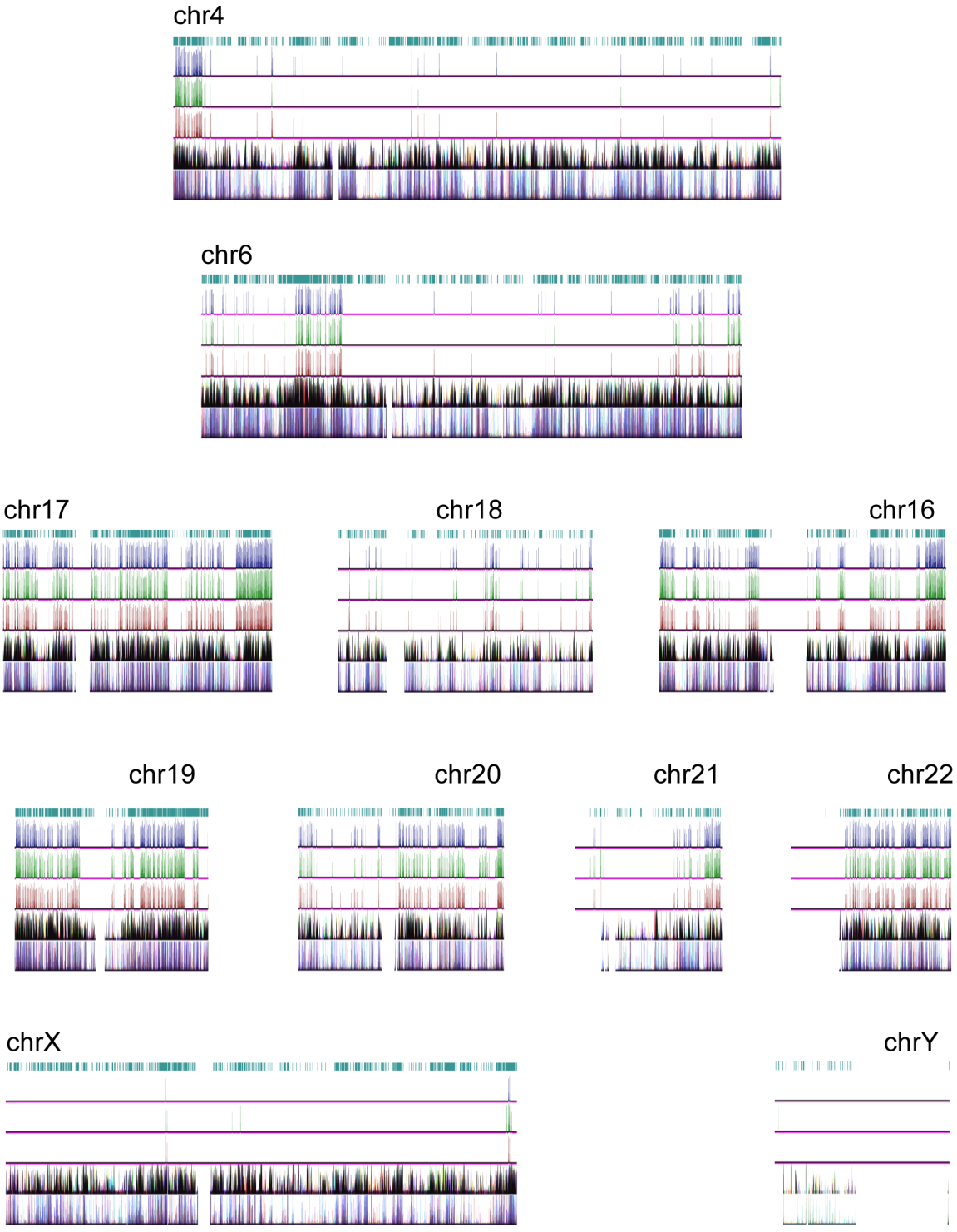
Supplementary Figure 5

Supplementary Figure 5. Overexpression of *Hox* genes in cancer is linked to decreased survival. (**A**) Heat map of *Hox* gene expression (https://gtexportal.org/home/) in adults for putative tissues of origin of tumors where overexpression was associated with poor survival (left) and number of patients with expressed *Hox* genes in LGG (right). *, *Hox* genes expressed in fewer than 100 patients out of 532. (**B**) Bar plots from GCEA of -log10 p-values for the coexpression of *Hox* genes with transcription factors *MYBL2* (left) and *FOXM1* (right). Red line, p-value threshold at 0.05; red, *Hox* genes with AcNEIL1 PICs; boxes, genes associated with poor survival in at least 3 cancer types.



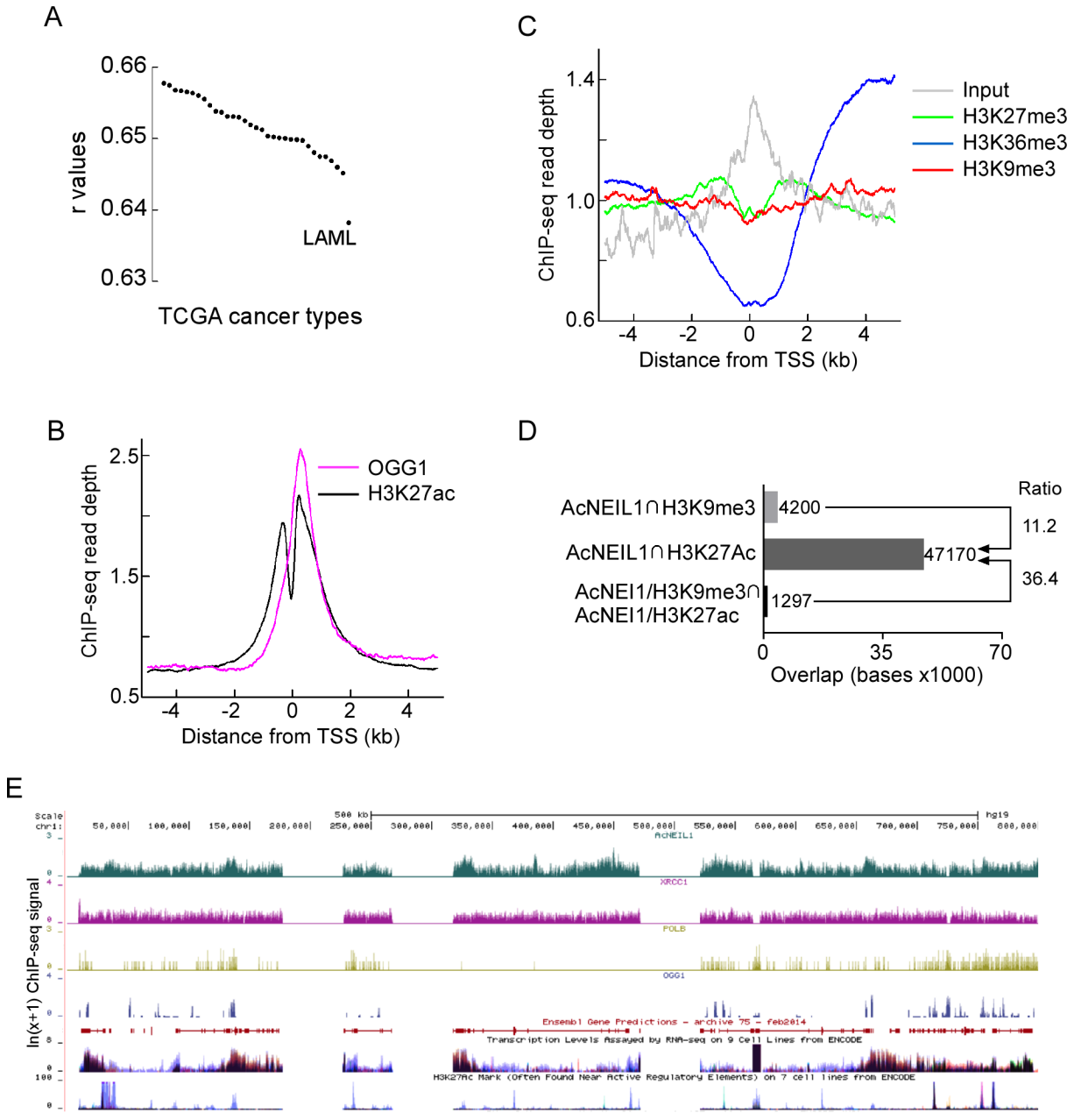
Supplementary Figure 6

Supplementary Figure 6. Transcription spreads AcNEIL1 along gene bodies. (**A**) Heat map and Euclidean hierarchical clustering of transcriptome profiling in 122 normal human tissue samples from [15]. Grey, clusters of genes without AcNEIL1 PICs within 1 kb of TSS (AcNEIL1– genes); red, clusters of genes with AcNEIL1 PICs within 1 kb of TSS (AcNEIL1+ genes). Most (21/30) discernable AcNEIL1+ clusters appear in the sector of weakly transcribed genes. (**B**) Box plots for *CBX8* gene expression in tumor and normal matched controls in TCGA. Only datasets with at least 10 controls were considered. P-values from Wilcoxon tests; green, above significant level (0.05). (**C**) 2-D hierarchical clustering of p-values from Wilcoxon tests for transcript levels in 5 control sets (each with1000 genes chosen at random) relative to all other genes in 33 TCGA tumor types. Scale, -log p-values. (**D**) Cumulative distribution of peak enrichment (bottom scale) and q-values (top scale) for the three AcNEIL1 replicates (from "macs2 callpeak"). Orange, range of PICs from ChIPpeakAnno. (**E**) Mean and SD of PICs lengths in different genomic features. Number of peaks were: outside, 6537; upstream, 2741; start, 1005; inside, 1216; stop, 275. P-values from Welch's t-tests between pairs (horizontal bars) of features.



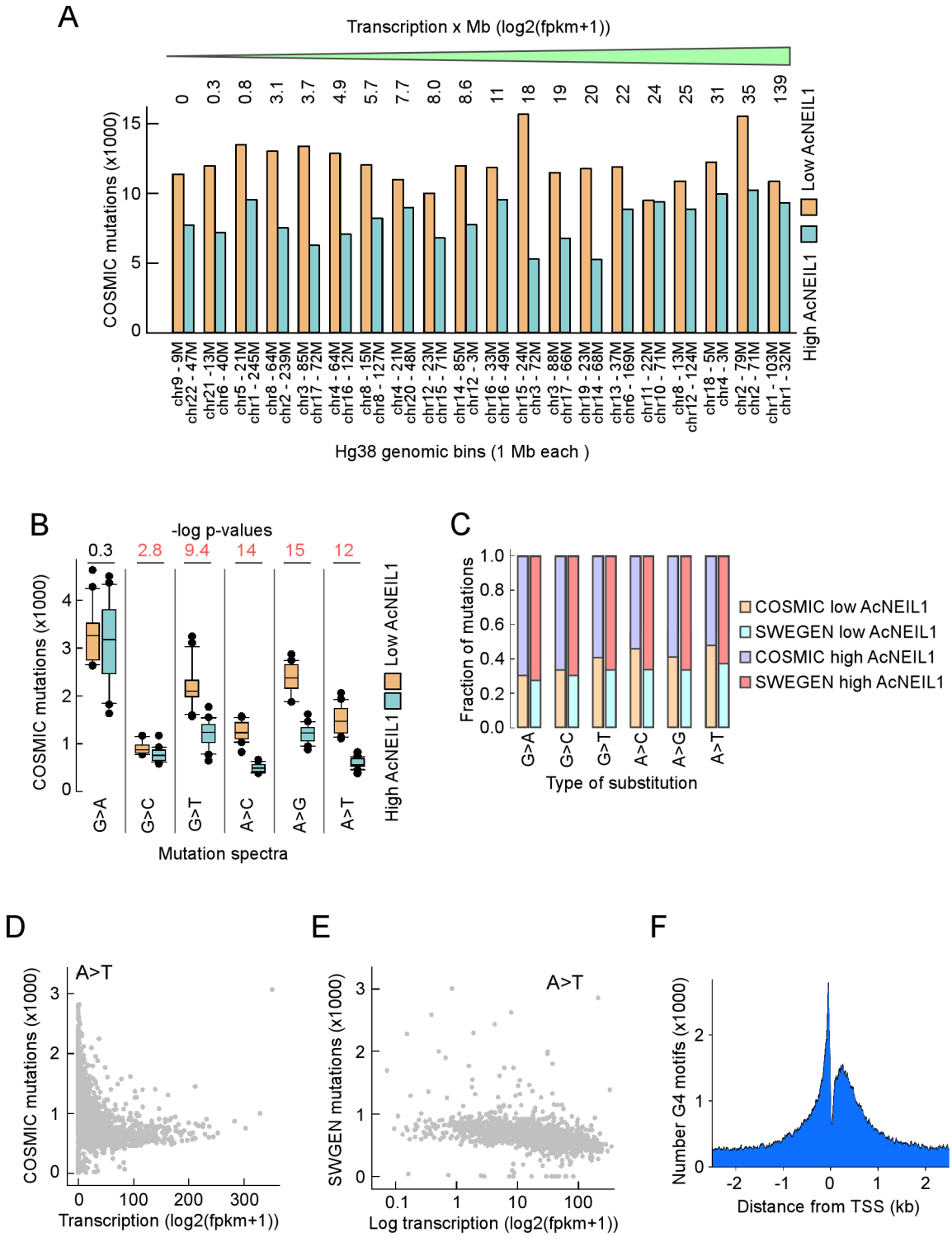
Supplementary Figure 7

Supplementary Figure 7. Gene-dense and highly transcribed domains are enriched in AcNEIL1. Visualization of the AcNEIL1 ChIP-seq peak replicates for selected chromosomes on the UCSC genome browser. Tracks are as in Figure 4F.



Supplementary Figure 8

Supplementary Figure 8. ChIP-seq shared peaks for BER components and modified histones. (**A**) Regression coefficients from correlations between ICS and transcription for the 33 TCGA cancer types. ICS as in Figure 4H; transcription is the value averaged over all patients within a given cancer type. LAML, result for the cohort of acute myeloid leukemia patients. (**B**) and (**C**) Profile of aggregate Chip-seq read depth near TSSs from the ENCODE project bigwig files in the MCF7 breast cancer cell line; Panel B, OGG1 and H3K27ac; Panel C, H3K27me3, H3K36me3, H3K9me3 and Input. (**D**) Number of overlapping bases between AcNEIL1 and H3K9me3, between AcNEIL1 and H3K27ac, and between the AcNEIL1/H3K9m3 and AcNEIL1/H3K27ac pairs from the ENCODE project narrow peak bed files in HCT116 cells. (**E**) Comparative ChIP-seq peak landscape of AcNEIL1, XRCC1, Pol β and OGG1 on chromosome 1 (hg19) visualized on the UCSC genome browser.



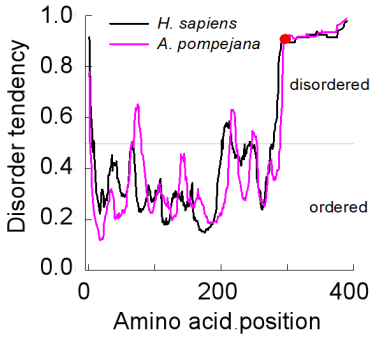
Supplementary Figure 9

Supplementary Figure 9. AcNEIL1 occupancy correlates with low mutation rates. (**A**) Genomic regions with low AcNEIL1 incur high mutation loads in cancer genomes. Twenty 1-Mb genomic intervals were chosen with either compound high (>950,000, turquoise) or compound low (<35,500, orange) AcNEIL1 ICS (sum of fold-enrichment signal at every bp over 1 Mb interval) for replicate 2 with matching transcriptional activity (sum of averaged log2(fpkm + 1) for all normal tissues for all transcripts with TSS within the 1 Mb interval, top) from (22). Number of mutations in cancer genomes were computed at each genomic interval and plotted; x-axis, starting position (M, million) of 1 Mb intervals. (**B**) Box plots of number of mutations for the 6 types of base substitution in the combined bins from **a**. Pairwise p-values from Welch's t-tests between low and high AcNEIL1-containing bins, n = 40. (**C**) Bar plot of relative fractions of base substitutions in cancer genomes and in the Swedish population. Number of base substitutions were computed in 1-Mb bins with low (<500,000) and high (>500,000) AcNEIL1, as per panel A, and plotted relative to the total. (**D**) Dot plot of number of A>T base transversions in cancer genomes as a function of transcription in 1 Mb bins. Transcription values as per panel A. (**E**) Dot plot of number of A>T base transversions in the Swedish population as a function of transcription in 1 Mb bins. Transcription values as per panel A; x-axis in log scale. (**F**) Distribution of G4 DNA-forming motifs around TSSs in hg38 annotated genes from UCSC refFlat.txt file.

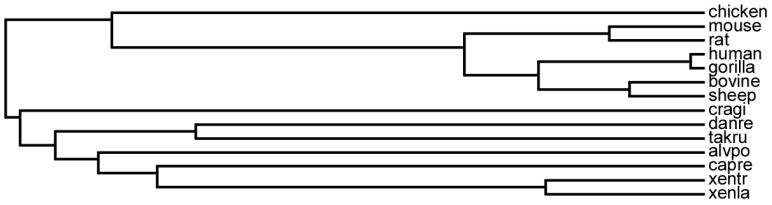
A

```
GEN   1   MPEGPELFIASRFVNAICKGRYFSGIVKKSDVSKCPSVMWDSDLYTIAAASRGKEIKLTL  60
          MPEGPELFIASRFVNAICKGRYFSGIVKKSDVSKCPSVMWDSDLYTIAAASRGKEIKLTL
EST   1   MPEGPELFIASRFVNAICKGRYFSGIVKKSDVSKCPSVMWDSDLYTIAAASRGKEIKLTL  60

GEN   61  TDVEANNTPGIKGNHKKQARRLSKEAKKLDIVFTFGMSGKFTFNPVNGTPKHAHLQFFTK  120
          TDVEAN TPGIKGNHKKQARRLSKEAKKLDIVFTFGMSGKFTFNPVNGTPKHAHLQFFTK
EST   61  TDVEAN-TPGIKGNHKKQARRLSKEAKKLDIVFTFGMSGKFTFNPVNGTPKHAHLQFFTK  119

GEN   121 DDNMSLCFVDTRRFGKWVPEGDWSPQRGPCVILEYEQFRENVLSSLKLSEFDKPICEVML  180
          DDNMSLCFVDTRRFGKWVPEGDWSPQRGPCVILEYEQFRENVLSSLKLSEFDKPICEVML
EST   120 DDNMSLCFVDTRRFGKWVPEGDWSPQRGPCVILEYEQFRENVLSSLKLSEFDKPICEVML  179

GEN   181 NQKYFNGVGNYLRAEVLYRAGVRPFEKARNVLEQLNTDGKMKSESPDILSLCHSVAREVV  240
          NQKYFNGVGNYLRAEVLYRAGVRPFEKARNVLEQLNTD KMKSESPDILSLCHSVAREVV
EST   180 NQKYFNGVGNYLRAEVLYRAGVRPFEKARNVLEQLNTDSKMKSESPDILSLCHSVAREVV  239

GEN   241 DLAKEHGSAYYNEDEDTTFMSWLQCYYNPNMKSISDHNGRTIWFAGDPGPLVPKDAKKMH  300
          DLAKEHGSAYYNEDEDTTFMSWLQCYYNPNMKSISDHNGRTIWFAGDPGPLVPKDAKKMH
EST   240 DLAKEHGSAYYNEDEDTTFMSWLQCYYNPNMKSISDHNGRTIWFAGDPGPLVPKDAKKMH  299

GEN   301 QGKSRKSVPKKSASESVAKVDLLSANSQISEDQDTLKEMKSRSNTKKARNDIVDDDVSSD  360
          QGKSRKSVPKKSASESVAKVDLLSANSQISEDQDTLKEMKSRSNTKKARNDIVDDDVSSD
EST   300 QGKSRKSVPKKSASESVAKVDLLSANSQISEDQDTLKEMKSRSNTKKARNDIVDDDVSSD  359

GEN   361 KKRLQLQLSSKQKVENVASRTRSKIKQKLT  390
          KKRLQLQLSSKQKVENVASRTRSKIKQKLT
EST   360 KKRLQLQLSSKQKVENVASRTRSKIKQKLT  389
```

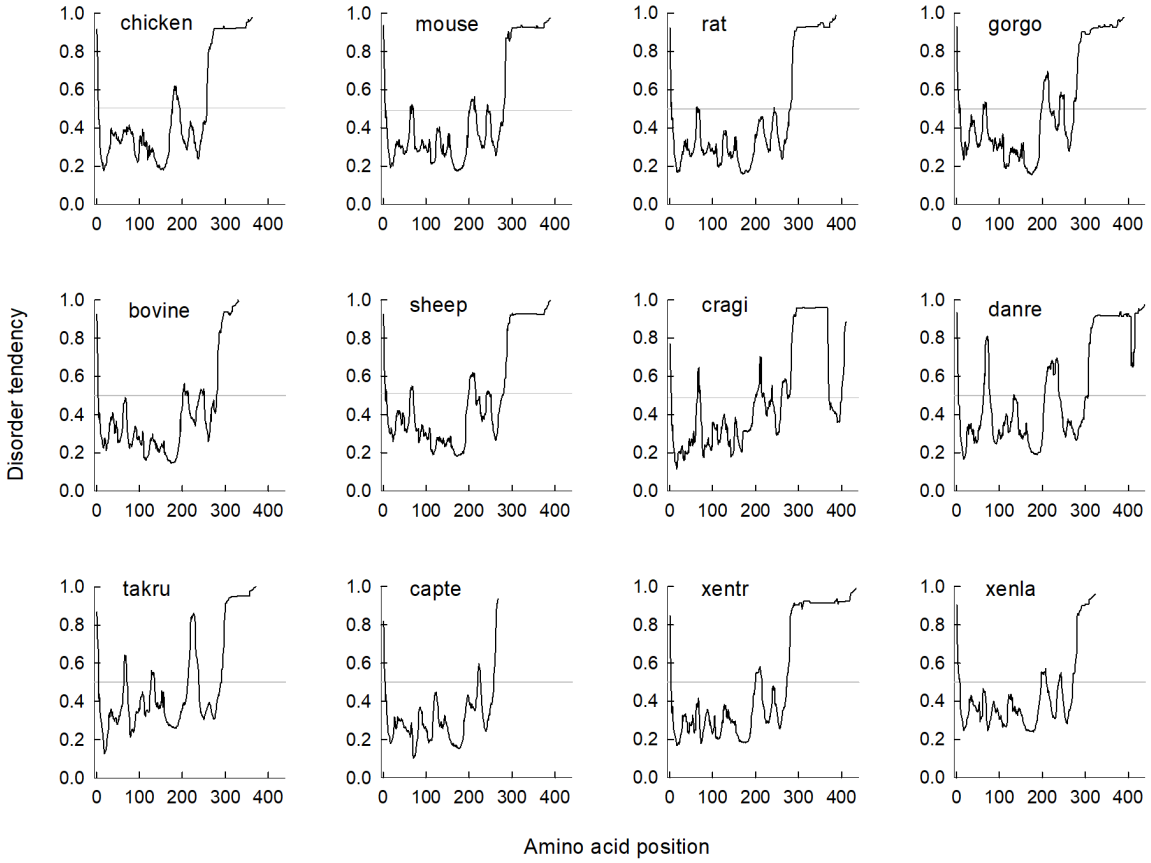
B



D



C



Supplementary Figure 10

Supplementary Figure 10. Evolutionary origin of AcNEIL1 acetylation center. (**A**) *Alvinella pompejana* Neil1 sequence from genomic DNA and EST. The protein sequences differ at two positions due to a tandem AACAAC motif encoding NN (amino acids 66-67) in genomic DNA vs. a single AAC in the EST, and a G→A transition resulting in a G→S change at amino acid 219. There are 7 additional base changes between the genomic and EST nucleotide sequence, which we suspect they may arise from RNA oxidation. (**B**) Intrinsic disorder prediction for human (black) and *Alvinella pompejana* (magenta) NEIL1. Red dot, acetylation center comprising Lys^{296}, Lys^{297}, and Lys^{298}. (**C**) intrinsic protein disorder scores (y axis <0.5, ordered; y axis >0.5, disordered) for Neil1 in representative species. (**D**) Phylogenetic tree based on Neil1 protein sequence alignment.

Supplementary Table 1. Summary of ChIPQC report

ID	Tissue	Factor	Condition	Replicate	Reads	Dup%	ReadL	FragL	RelCC	SSD	RiP%	RiBL%
AcNEIL1_1	HCT116	AcNEIL1	AcNEIL1	1	21894508	0	75	273	9.9	0.45	5.7	4.6e-06
AcNEIL1_2	HCT116	AcNEIL1	AcNEIL1	2	21601029	0	75	165	4.6	0.51	7.6	9.3e-06
AcNEIL1_3	HCT116	AcNEIL1	AcNEIL1	3	23848649	0	75	187	8.9	0.46	5.7	0
INP_1	HCT116	Control	AcNEIL1	c1	22559745	0	74	190	13	0.38	2	8.9e-06
INP_2	HCT116	Control	AcNEIL1	c2	26177874	0	74	199	14	0.41	2.1	0.000015
INP_3	HCT116	Control	AcNEIL1	c3	23191383	0	75	181	9.6	0.37	1.8	4.3e-06

Supplementary Table 2. Primers used in the study.

Gene symbol and application	Sequence (5´->3´)
MYC (ChIP qPCR)	F: GCGAGGATGTGTCCGATTCT R: CCCTTCGCACTCAATACGGA
CDKN1A (ChIP qPCR)	F: CAGGCTGTGGCTCTGATTGG R: TTCAGAGTAACAGGCTAAGG
HDAC1 (ChIP qPCR)	F: CAAACCCGCGTGTTGCTTTT R: CAGTCCCACTTCGTCGTGA
NEIL1 (ChIP qPCR)	F: CACTCCAGGATGGGAAACCC R: AAGGCAGCAAAGTCCTCCTC
MYL1 (ChIP qPCR)	F: GCTGGGGTTTCCCAGAACTT R: AGAGCACTCTAATGCTTCTTGCT
MAP2 (ChIP qPCR)	F: TCTATGTTCTCAAGCGCCCC R: AATCGGGGGCAACAGCTTTA
RARβ2 (ChIP qPCR)	F: CTCTGGCTGTCTGCTTTTGC R: CATGGGGGAATTCTGGTCCC
Non-specific region from chromosome 17 (ChIP qPCR)	F: TACTATCCCCGTGCTTCCCA R: CATTGAGGAGGGGGCAACAT
RARB (Reverse transcription Real Time PCR)	F: AAGTGAGCTGTTCAGAGGCA R: AATGCGTTCCGGATCCTACC
HPRT1 (Reverse transcription Real Time PCR)	F: TGACACTGGCAAAACAATGCA R: GGTCCTTTTCACCAGCAAGCT

Supplementary Table 3. Intersection between G4-DNA and gene mutations in 5´-UTRs in the HGMD dataset.

Case_01	HGMD Accession Number:	CR092368
Gene:		*IRF6*
Coordinate at mutation:		hg38:chr1:209806090 G>A
G4 DNA sequence:		CCCCGTCCCGCACCAGCCCTTACCTGCCCAGCCC
Distance from TSS in mRNA:		-219
Phenotype:		Van der Woude syndrome
HGMD entry tag:		DM
References:		19282774
Notes: truncation mutation due to out-of-frame start codon		

Case_02	HGMD Accession Number:	CR166119
Gene:		*TACR3*
Coordinate at mutation:		hg38:chr4:103719695 C>T
G4 DNA sequence:		CCCGGACCCTCCCACTCACCC
Distance from TSS in mRNA:		-20
Phenotype:		normosmic congenital hypogonatropic hypogonadism (nCHH)
HGMD entry tag:		DM?
References:		27094476
Notes: unknown significance - patient has mutation in the GNRHR protein		

Case_03	HGMD Accession Number:	CR149638
Gene:		*T*
Coordinate at mutation:		hg38:chr6:166167607 C>T
G4 DNA sequence:		CCCCTCCCCGCCGTCCCCGAAGCCC
Distance from TSS in mRNA:		-16
Phenotype:		Chordoma
HGMD entry tag:		DM?
References:		24990759
Notes: rs3734509 – SNP overlaps with transcription factor binding sites		

Case_04	HGMD Accession Number:	CR104634
Gene:		*FZD1*
Coordinate at mutation:		hg38:chr7:91264657 T>G
G4 DNA sequence:		CCCCGGCGCGCCCTAGCCACCCGGGTTCTCCCCGCCGCCC
Distance from TSS in mRNA:		-224
Phenotype:		Femoral neck geometry
HGMD entry tag:		FP
References:		20051274
Notes: rs2232157 – predicted altered transcription factor binding		

Case_05	HGMD Accession Number:	CR961721
Gene:		*SERPING1*
Coordinate at mutation:		hg38:chr11:57597645 C>G
G4 DNA sequence:		CCCCCCCCGCACCCCACCCTCCCTGACCC
Distance from TSS in mRNA:		-40
Phenotype:		Hereditary Angiodema
HGMD entry tag:		DM
References:		8755917
Notes: rs578018379 - unknown significance; patient has additional codon; deletion in *SERPING1*		

Case_06	HGMD Accession Number:	CR136167
Gene:		*PICALM*
Coordinate at mutation:		hg38:chr11:86068857 C>G
G4 DNA sequence:		CCCCACCCCCCACCGCACCCCCTACCCCC
Distance from TSS in mRNA:		-77
Phenotype:		Colorectal cancer
HGMD entry tag:		DM?
References:		23585368
Notes: suspected predisposition to colorectal cancer; additional frameshift mutation in *TGFBR2*		

Case_07	HGMD Accession Number:	CR091753
Gene:		*ZIC2*
Coordinate at mutation:		hg38:chr13:99982041 C>T
G4 DNA sequence:		GGGCTCGCAGGGGCGGGCGGG
Distance from TSS in mRNA:		-24
Phenotype:		Holoprosencephaly
HGMD entry tag:		DM

References: 19177455
Notes: unknown significance

Case_08 HGMD Accession Number: CR016062
Gene: *RAD51*
Coordinate at mutation: hg38:chr15:40695330 G>C
G4 DNA sequence: GGGGCGTTGGGGGCCGTGCGGGTCGGG
Distance from TSS in mRNA: -98
Phenotype: Myelodysplastic syndrome, cancer
HGMD entry tag: DFP
References: 25312513, 14724582, 26511493, 24930116, 11248061, 21708019
11535547, 16398215, 17999359, 19606696, 20396943, 20454923
20623332, 20640595, 21647442, 22611952, 24040396, 24859942
Notes: rs1801320 - suspected increased susceptibility to myelodysplastic; syndrome and cancer

Case_09 HGMD Accession Number: CR025892
Gene: *CDH1*
Coordinate at mutation: hg38:chr16:68737362 G>C
G4 DNA sequence: CCCGCTCCAGCCCGGCCCGACCCGACCGCACCC
Distance from TSS in mRNA: -54
Phenotype: Likely benign
HGMD entry tag: FP
References: Nakamura (2002) Mutat Res 502, 19
Notes: rs5030874 - reduced promoter activity

Case_10 HGMD Accession Number: CR132243
Gene: *CDH1*
Coordinate at mutation: hg38:chr16:68737367 G>T
G4 DNA sequence: CCCGCTCCAGCCCGGCCCGACCCGACCGCACCC
Distance from TSS in mRNA: -49
Phenotype: Likely benign
HGMD entry tag: FP
References: Chen (2013) World J Gastroenterol 19, 909
Notes: rs564350060 - increased promoter activity

Case_11 HGMD Accession Number: CR141061
Gene: *ELANE*
Coordinate at mutation: hg38:chr19:852326 A>T
G4 DNA sequence: CCCCGGAGCCCCAGCCCCACCATGACCC
Distance from TSS in mRNA: -3
Phenotype: Cyclic neutropenia
HGMD entry tag: DM
References: Tidwell (2014) Blood 123, 562
Notes: mutation activates translation from downstream initiation codons

Case_12 HGMD Accession Number: CR063419
Gene: *XRCC1*
Coordinate at mutation: hg38:chr19:43575535 G>A
G4 DNA sequence: GGGGTCCGAGGGGCAGGGAGAGTGGGAGGGGGCGGGGG
Distance from TSS in mRNA: -77
Phenotype: Increased risk of HNSSC, gastric and lung cancer
HGMD entry tag: DFP
References: 27372710, 24470137, 21427728, 16652158, 19116388, 19662459
20549339
Notes: rs3213245 - decreased promoter activity, stronger Sp1 binding site

Case_13 HGMD Accession Number: HR080003
Gene: *SRY*
Coordinate at mutation: hg38:chrY:2787733 C>G
G4 DNA sequence: CCCTCAACACCCCCTCAACCCCGCCC
Distance from TSS in mRNA: -130
Phenotype: XY sex reversal
HGMD entry tag: DM
References: 19694000, 9582429
Notes: mutation in Sp1 binding site

Supplementary Table 3. There are five different classes of variant listed in HGMD. Disease-causing mutations (**DM**) are entered into HGMD where the authors of the corresponding report(s) have established that the reported mutation(s) are involved (or very likely to be involved) in conferring the associated clinical phenotype upon the individuals concerned. The DM classification may, however, also appear with a question mark (**DM?**), denoting a probable/possible pathological mutation, reported as likely to be disease-causing in the corresponding report, but where (i) the author has indicated that there may be some degree of doubt or uncertainty; (ii) the HGMD curators believe greater interpretational caution is warranted, or (iii) subsequent evidence has appeared in the literature which has called the initial putatively deleterious nature of the variant into question (e.g. a negative functional, case–control or population-scale sequencing study). The DM and DM? variant classes may include mutations that are believed to contribute to disease susceptibility in a multi-factorial manner (e.g. autism or schizophrenia), exhibit complex polygenic inheritance or possess an environmental trigger component to their pathogenicity. Disease-associated polymorphisms (**DP**) are entered into HGMD where there is evidence for a significant association with a disease/clinical phenotype along with additional evidence that the polymorphism is itself likely to be of functional relevance (e.g. as a consequence of genic location, evolutionary conservation, transcription factor binding potential, etc.), although there may be no direct evidence (e.g. from an expression study) for a functional effect. The functional polymorphisms (**FP**) class includes those sequence changes for which a direct functional effect has been demonstrated (e.g. by means of an *in vitro* reporter gene assay or alternatively by protein structure, function or expression studies), but with no disease association reported as yet. Disease-associated polymorphisms with supporting functional evidence (**DFP**) must meet both of the above criteria in that the polymorphism should not only have been reported to be significantly associated with disease, but should also display direct evidence of being of functional relevance.