Supplementary Information

Multilayered control of mRNAs delivery to cell protrusions drives localized mini-cytoplasm establishment for stress-induced cell activation.

Carlotta Duval[1,2], Andrea Lauria[1,2,3], Alessandro Croce[1,2], Chiara Levra Levron[1,2], Osamu Ansai[1,2], Livia Caizzi[1,2], Francesca Anselmi[1,2,3], Gabriele Piacenti[1,2], Elisa Balma[1,2], Luca Elettrico[1,2], Daniela Donna[1,2], Francesco Neri[1,2], Mika Watanabe[4], Ken Natsuga[4], Valentina Proserpio[1,2,3], Salvatore Oliviero[1,2,3], Giacomo Donati[1,2,$].

Supplementary Fig.1
ReGenT-seq system setup: shRNA library specificity and potency, *Krt6b* promoter responsiveness, and epidermal cell validation.

Supplementary Fig.2
In vitro epidermal cell culture induces stress-associated gene expression programs characteristic of wound healing and SCC.

Supplementary Fig.3
ReGenT-seq results and validation.

Supplementary Fig.4
Viability screens and network analysis of hits shared with ReGenT-seq.

Supplementary Fig.5
Single-gene knockdown and RNA-seq analysis of chromatin factor hits identified by ReGenT-seq.

Supplementary Fig.6
Roles of *Cbx3* and *Kmt5c* in cell motility and proliferation.

Supplementary Fig.7
Histological stratification of wound healing regions.

Supplementary Fig.8
Histological stratification of SCC regions.

Supplementary Fig.9
Integration of transcriptomic profiling after CBX3 knockdown with ChIP-seq of endogenous CBX3.

Supplementary Fig.10
seCLIPs of endogenous CBX3.

Supplementary Fig.11
CBX3 RNA-binding profiles and multicopy RNA families bound by CBX3.

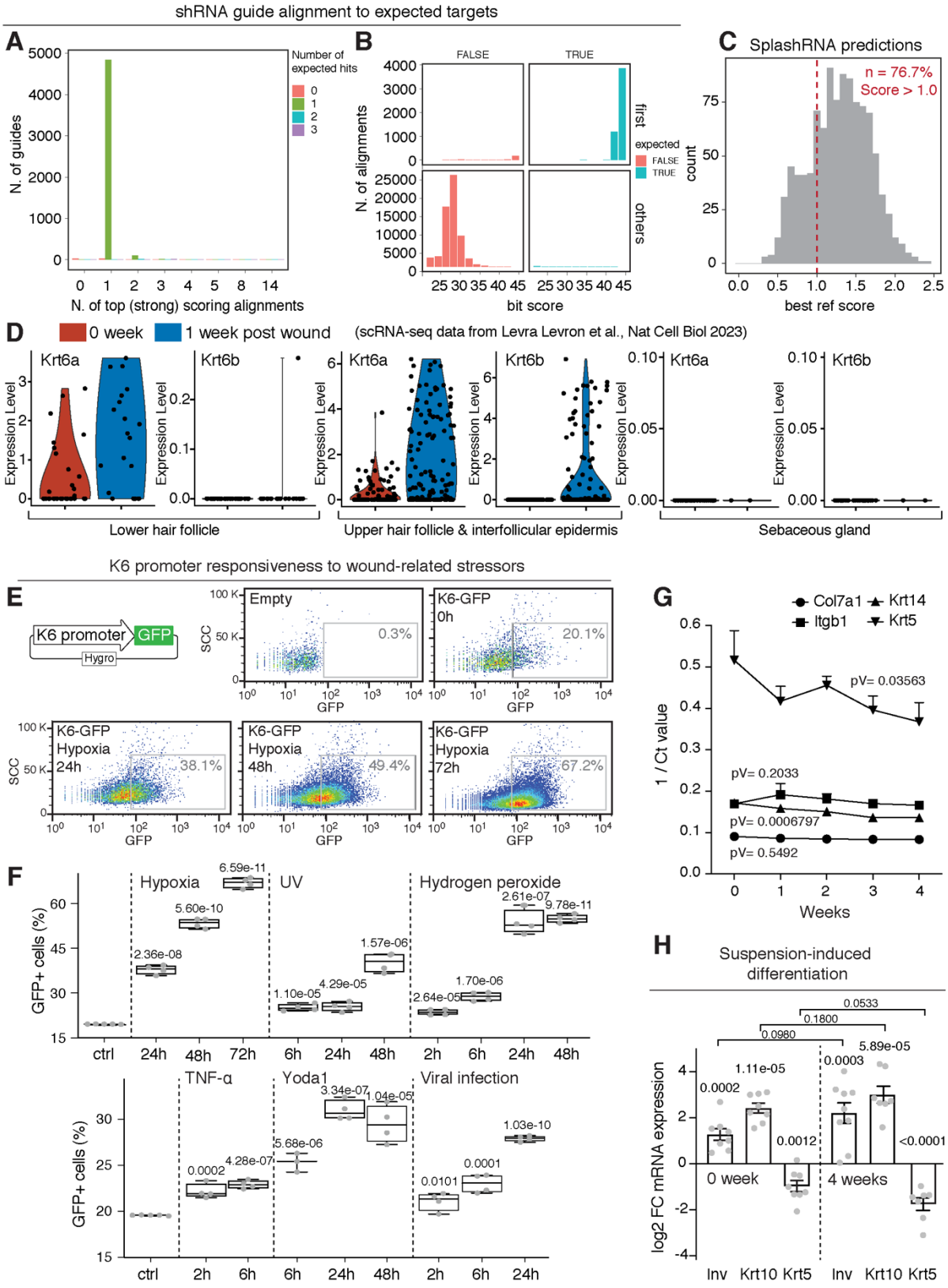
Supplementary Fig.12
Analysis of CBX3-dependent RNA splicing and nuclear export.

Supplementary Fig.13
Compartment-specific localization of RNAs in the nucleus, cytoplasm, and cell protrusions.

Supplementary Fig.14
Cbx3-dependent multilayered control of gene expression.

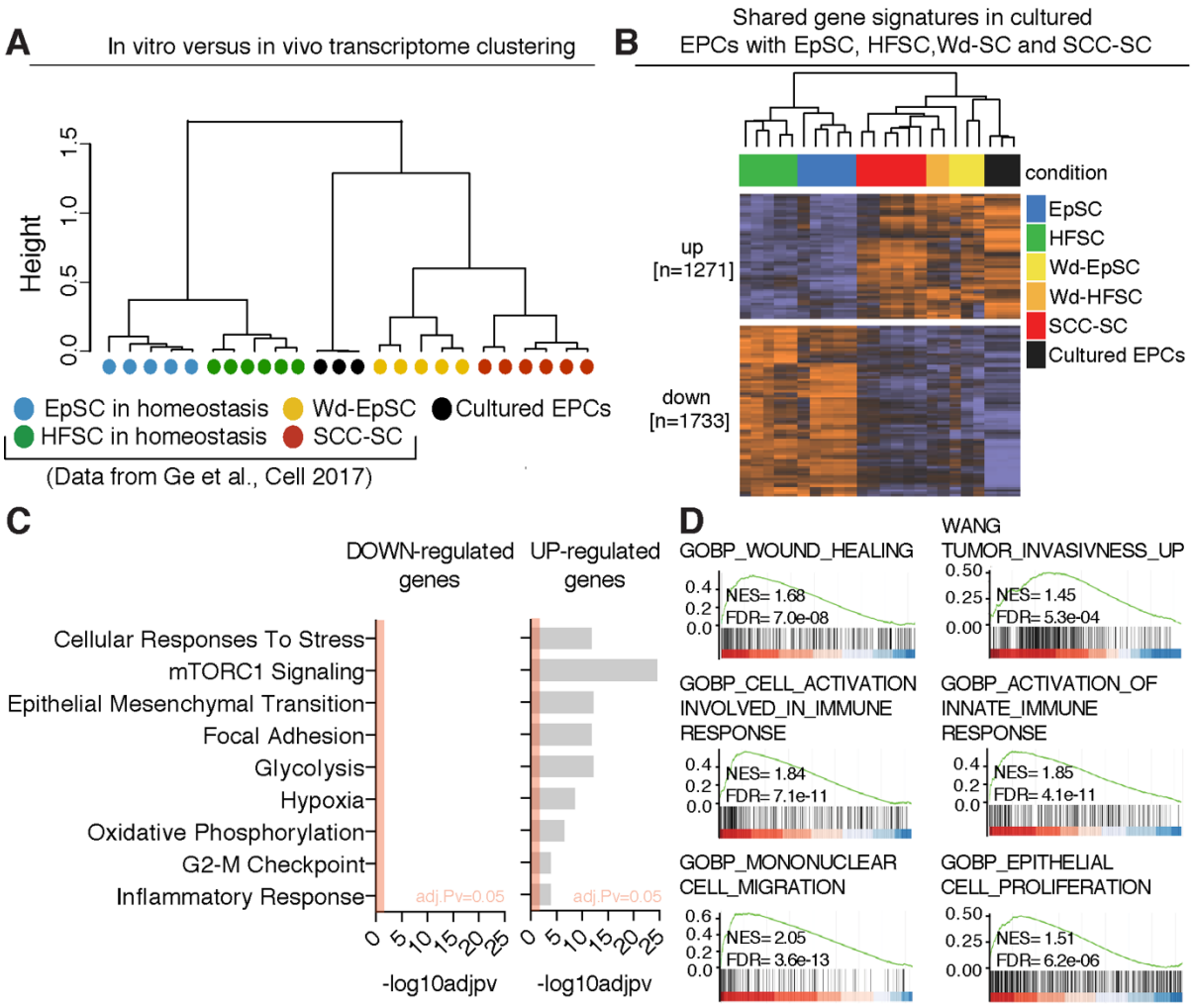
Supplementary Fig.15
CBX3 functional analysis in human epidermal cells.

Supplementary Fig.16
CBX3 controls 3D induced protrusions in human primary epidermal cells.



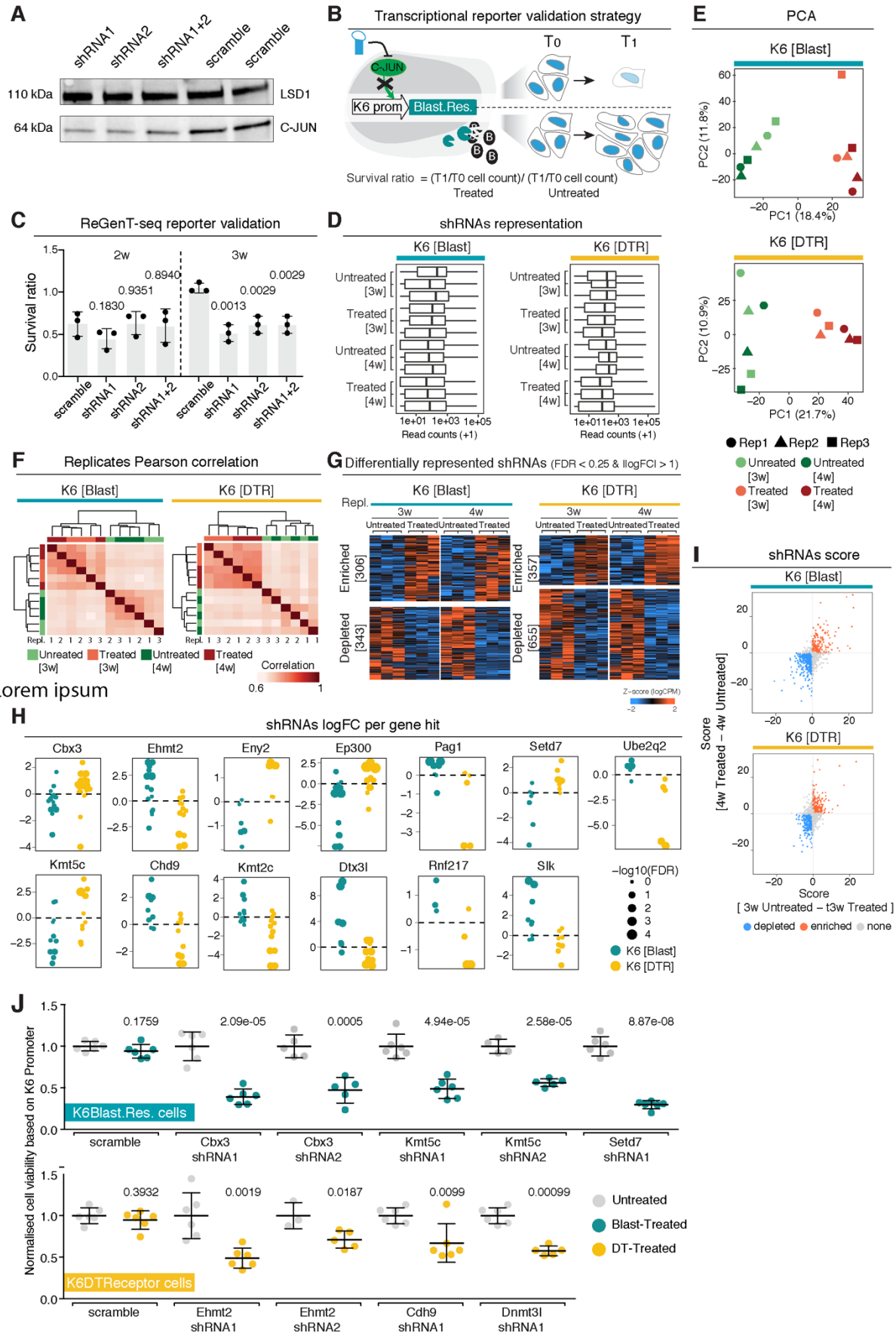
Supplementary Fig.1 | ReGenT-seq system setup: shRNA library specificity and potency, *Krt6b* promoter responsiveness, and epidermal cell validation. (A) Distribution of the number of top-scoring BLASTN alignment hits per shRNA guide, showing that most guides have a single strong

alignment. The y-axis reports the number of guides, the x-axis the number of strong alignments. Color code indicates the number of expected hits. **(B)** Distribution of BLASTN alignment bit scores, comparing the top-ranked (first) hit against all the remaining alignments (others) for each shRNA guide. Color and faceting separate the distributions according to whether the alignment corresponds to the expected target for each shRNA, showing that the strongest BLASTN alignments predominantly correspond to the expected shRNA targets (96.4%). **(C)** Prediction of shRNA library efficacy using SplashRNA. The histogram depicts the distribution of top SplashRNA prediction scores sampled from the shRNA library. Predicted potent shRNAs are defined as those with scores greater than 1. **(D)** Violin and dot plots showing mRNAs levels of *Krt6a* and *Krt6b*, measured by scRNA-seq, during homeostasis and one week after wounding in the indicated epidermal regions. **(E)** Representative flow cytometry plots of GFP expression in K6-GFP EPCs under hypoxia at the indicated time points, with percentages of GFP-positive cells shown. (**F**) Percentage quantification of GFP-positive EPCs following exposure to various stress conditions, including hypoxia (2% O_2), UV irradiation (254 nm, 400 mJ/cm²), hydrogen peroxide (0.5 nM H_2O_2), viral infection, mimic of mechanical stress simulation (10 μM Yoda-1), and inflammatory signaling (20 ng/μL TNF-α). Cells were analyzed at the indicated time points after treatment. GFP positivity is expressed relative to untreated control cells. n=4 independent replicates. The horizontal line indicates median, box indicates the interquartile range (IQR), and whiskers denote the 1.5 × IQR. **(G)** RT-qPCR analysis of epidermal stem/progenitor cell markers in epidermal cultures in vitro over the course of the 4-week ReGent-seq experiment. Gene expression levels are shown as relative expression normalized to *Actb* mRNA. Data are presented as mean ± SEM. from four technical replicates per condition. P-values from ANOVA test. **(H)** RT-qPCR analysis of epidermal differentiation markers in epidermal cultures under differentiation conditions, with suspension-induced differentiation applied for 36 h at the start and end of the ReGenT-seq experiment. Gene expression levels were normalized to *Actb* mRNA and are shown as relative expression. Data represent mean ± SEM. of four technical replicates per condition. In all graphs, statistical significance was assessed using an unpaired two-sided Student's t-test, unless otherwise stated.

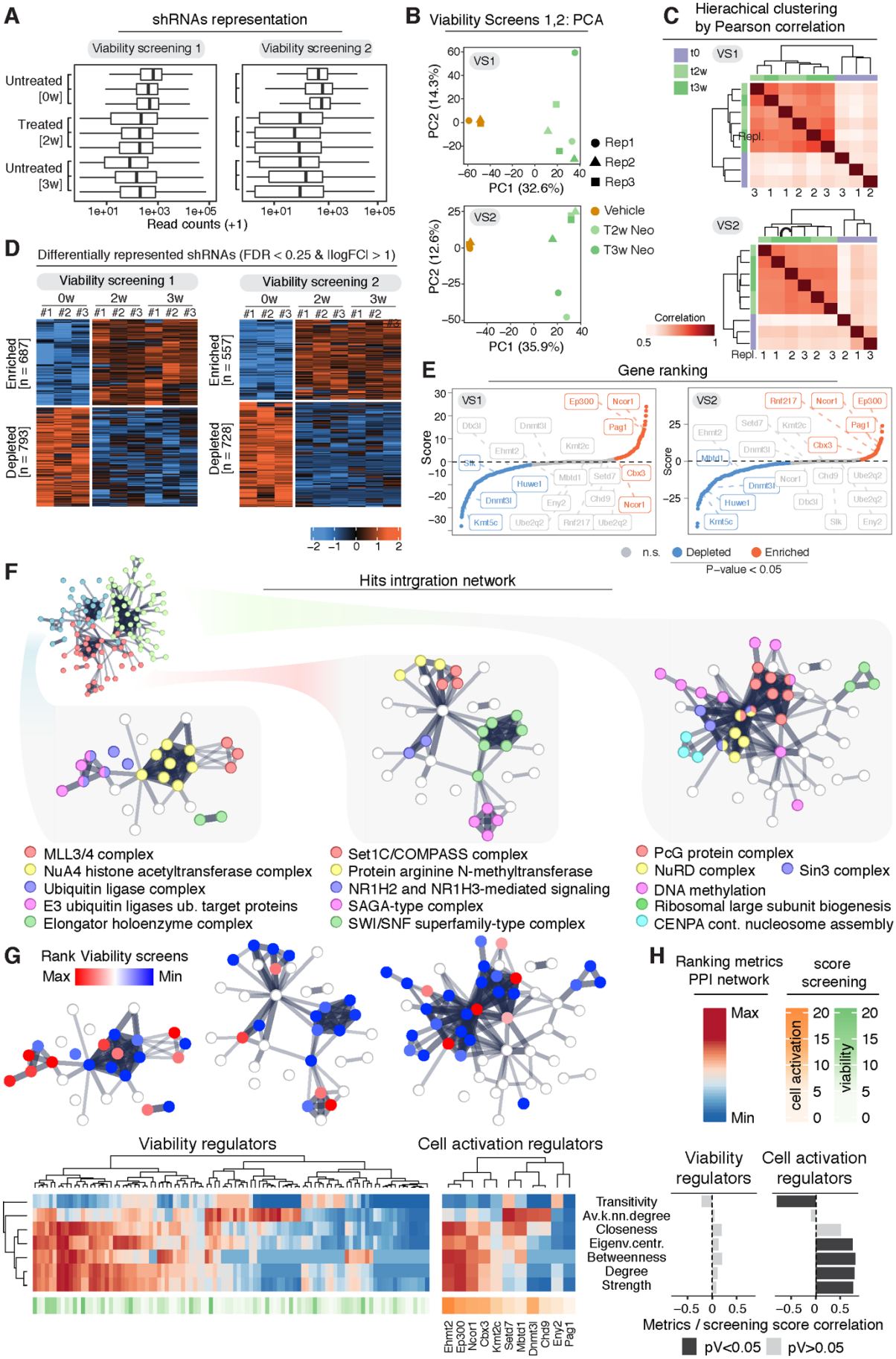


Supplementary Fig.2 | In vitro epidermal cell culture induces stress-associated gene expression programs characteristic of wound healing and SCC. (A) Hierarchical clustering of RNA-seq samples integrating in vitro epidermal progenitor cell (EPC) expression data from this study with in vivo datasets from Ge et al. (*Cell*, 2017), including wound-associated and tumor stem cells (Wd-EpSC and SCC-SC), as well as epidermal and hair follicle stem cells under homeostatic conditions (EpSC and HFSC). Cultured EPCs cluster more closely with Wd-EpSC and SCC-SC than with homeostatic EpSCs or HFSCs. **(B)** Heatmap showing expression levels of the gene signature shared by EPCs, SCC-SC and stem cells during wound healing (Wd-EpSC and Wd-HFSC), identified through differential expression analysis comparing them with EpSC and HFSC from homeostatic conditions. Heatmap clustering highlights up- and down-regulated genes. **(C, D)** GO term enrichment (C) and GSEA analysis (D) were performed on the gene signature shared by EPCs, Wd-SC, Wd-HF, and SCC-SC. GO term enrichment

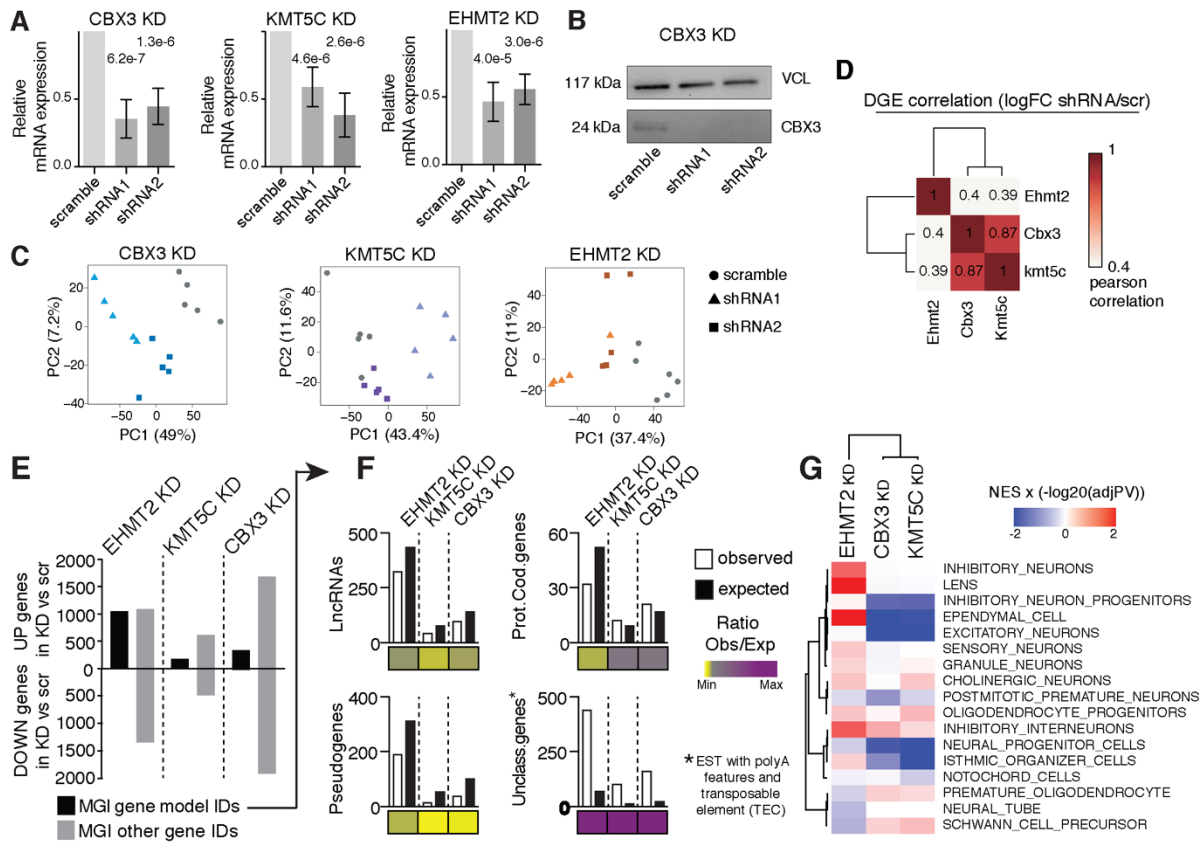
was conducted using Enrichr, and adjusted P values (Benjamini–Hochberg correction) are reported. GSEA was carried out using a running-sum statistic with permutation-based testing, with P values adjusted using the Benjamini–Hochberg method.


A
shRNA1
shRNA2
shRNA1+2
scramble
scramble
110 kDa
LSD1
64 kDa
C-JUN
B
Transcriptional reporter validation strategy
T0
T1
C-JUN
K6 prom
Blast.Res.
Survival ratio = (T1/T0 cell count)/ (T1/T0 cell count)
Treated
Untreated
C
ReGenT-seq reporter validation
Survival ratio
2w
3w
0.8940
0.9351
0.1830
0.0029
0.0013
0.0029
scramble
shRNA1
shRNA2
shRNA1+2
D
shRNAs representation
K6 [Blast]
K6 [DTR]
Untreated [3w]
Treated [3w]
Untreated [4w]
Treated [4w]
Read counts (+1)
E
PCA
K6 [Blast]
PC2 (11.8%)
PC1 (18.4%)
K6 [DTR]
PC2 (10.9%)
PC1 (21.7%)
Rep1 Rep2 Rep3
Unreated [3w]
Untreated [4w]
Treated [3w]
Treated [4w]
F
Replicates Pearson correlation
K6 [Blast]
K6 [DTR]
Repl.
Unreated [3w]
Treated [3w]
Untreated [4w]
Treated [4w]
Correlation
Lorem ipsum
G
Differentially represented shRNAs (FDR < 0.25 & llogFCl > 1)
K6 [Blast]
K6 [DTR]
Repl.
3w
4w
Untreated Treated Untreated Treated
Enriched [306]
Depleted [343]
Enriched [357]
Depleted [655]
Z-score (logCPM)
I
shRNAs score
K6 [Blast]
K6 [DTR]
Score [4w Treated – 4w Untreated]
Score [3w Untreated – t3w Treated]
depleted enriched none
H
shRNAs logFC per gene hit
Cbx3
Ehmt2
Eny2
Ep300
Pag1
Setd7
Ube2q2
Kmt5c
Chd9
Kmt2c
Dtx3l
Rnf217
Slk
−log10(FDR)
K6 [Blast]
K6 [DTR]
J
Normalised cell viability based on K6 Promoter
K6Blast.Res. cells
0.1759
2.09e-05
0.0005
4.94e-05
2.58e-05
8.87e-08
scramble
Cbx3 shRNA1
Cbx3 shRNA2
Kmt5c shRNA1
Kmt5c shRNA2
Setd7 shRNA1
K6DTReceptor cells
0.3932
0.0019
0.0187
0.0099
0.00099
scramble
Ehmt2 shRNA1
Ehmt2 shRNA2
Cdh9 shRNA1
Dnmt3l shRNA1
Untreated
Blast-Treated
DT-Treated

Supplementary Fig.3 | ReGenT-seq results and validation. (A) Western blot showing shRNA-induced downregulation (KD) of c-Jun. **(B)** Validation strategy for the cell activation (CA) reporter system. C-JUN KD reduces *Krt6b* promoter activity, resulting in an impaired cell viability upon Blasticidin treatment. **(C)** Bar plot validating the CA-reporter system. Survival ratio at 2 and 3 weeks (2w, 3w) were calculated as reported in (B). The horizontal bar represents the mean of three independent experiments per each condition, with error bar indicating ± SEM. **(D)** Number of detected shRNAs per sample (defined as shRNAs with at least one assigned read). **(E)** PCA of samples from the two individual ReGenT-seq sub-screens, K6 Blast (top) and K6 DTR (bottom) (Fig.1A-E,H). (**F**) Hierarchical clustering (Pearson correlation) of samples from the two individual ReGenT-seq sub-screens, K6 Blast (left) and K6 DTR (right). Clustering using Ward's method was performed based on Pearson's correlation distance of normalized read counts per shRNA, and the resulting pairwise correlation matrix is visualized as a heatmap.(**G**) Heatmaps showing differential shRNA representation between treated and untreated groups in the ReGenT-seq K6 Blast (left) and K6 DTR (right) sub-screens. Rows represent shRNAs and columns represent samples. Scaled hairpin abundance (Z-scores of TMM-normalized CPM) is plotted. Enriched and depleted shRNAs were identified using the edgeR package (FDR < 0.25 and log2FC > 1). n = 3 independent replicates per condition. **(H)** Dot plots representing all shRNAs targeting gene hits from ReGenT-seq, identified as pro- or anti- CA regulators. Log2FC is reported on the y-axis, and dot size represents the -log10(FDR). **(I)** Scatter plots visualizing the concordance of differential shRNA representation scores between the two time points (x-axis: 3 weeks; y-axis: 4 weeks) for the two ReGenT-seq sub-screens, K6 Blast (top) and K6 DTR (bottom). Red dots represent enriched hits, blue dots depict depleted hits, and non-significant hits are reported in gray. (**J)** Validation of gene hits affecting K6 promoter activity using a viability assay. Downregulation of gene hits in CA-reporter expressing cells reduces cell viability in a Blasticidin- or DT-dependent manner. For the Untreated (grey) and Treated (colored) groups, each dot represents the viability score calculated as the ratio of cells number at day 3 to day 0, normalized to maximum value observed. Horizontal bars represent the mean of 5–6 independent experiments per condition, with error bars indicating ± SEM. In all graphs, statistical significance was assessed using an unpaired two-sided Student's t-test, unless otherwise stated.

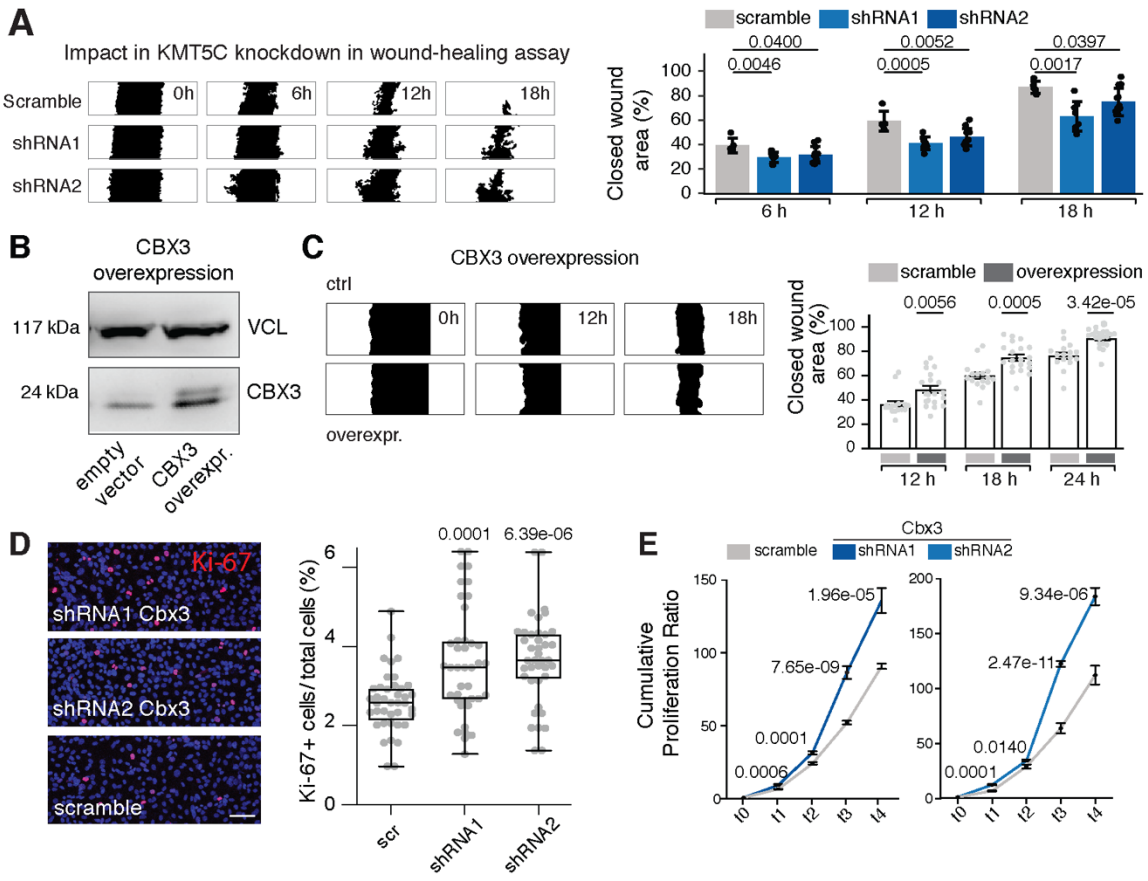

A
shRNAs representation
Viability screening 1
Viability screening 2
Untreated [0w]
Treated [2w]
Untreated [3w]
1e+01 1e+03 1e+05
Read counts (+1)
B
Viability Screens 1,2: PCA
VS1
PC2 (14.3%)
PC1 (32.6%)
Rep1
Rep2
Rep3
VS2
PC2 (12.6%)
PC1 (35.9%)
Vehicle
T2w Neo
T3w Neo
C
Hierachical clustering by Pearson correlation
t0
t2w
t3w
VS1
Repl.
VS2
Correlation
D
Differentially represented shRNAs (FDR < 0.25 & |logFC| > 1)
Viability screening 1
Viability screening 2
0w 2w 3w
Enriched [n = 687]
Depleted [n = 793]
Enriched [n = 557]
Depleted [n = 728]
E
Gene ranking
VS1
VS2
Score
n.s.
Depleted
Enriched
P−value < 0.05
F
Hits intrgration network
MLL3/4 complex
NuA4 histone acetyltransferase complex
Ubiquitin ligase complex
E3 ubiquitin ligases ub. target proteins
Elongator holoenzyme complex
Set1C/COMPASS complex
Protein arginine N-methyltransferase
NR1H2 and NR1H3-mediated signaling
SAGA-type complex
SWI/SNF superfamily-type complex
PcG protein complex
NuRD complex
Sin3 complex
DNA methylation
Ribosomal large subunit biogenesis
CENPA cont. nucleosome assembly
G
Rank Viability screens
Max
Min
Viability regulators
Cell activation regulators
H
Ranking metrics PPI network
score screening
Max
Min
cell activation
viability
Viability regulators
Cell activation regulators
Transitivity
Av.k.nn.degree
Closeness
Eigenv.centr.
Betweenness
Degree
Strength
Metrics / screening score correlation
pV<0.05
pV>0.05

Supplementary Fig.4 | Viability screens and network analysis of hits shared with ReGenT-seq. **(A)** Box plots illustrating shRNA representation (read counts per each shRNA) per sample. **(B)** PCA of samples from the two parallel viability screens (VS1 and VS2 respectively). **(C)** Heatmaps showing hierarchical clustering of samples from VS1 (left) and VS2 (right). Hierarchical clustering was performed using Ward's method based on Person's correlation distance of normalized shRNA read counts. The clustered pairwise correlation matrix is visualized as heatmap. **(D)** Heatmaps showing differential shRNA representation between time 0 and 2 or 3 weeks after sample collection for VS1 and VS2. Rows represent shRNA, columns represent samples. Scaled hairpin abundance (Z-scores of TMM-normalized CPM) is plotted. Enriched and depleted shRNAs were identified using the EdgeR package (Robinson et al., 2010) (FDR < 0.25 and log2FC > 1). n=3 independent replicates for each condition. **(E)** Gene-ranking plots for VS1 and VS2 screens. The effect of multiple shRNAs targeting the same gene was summarized by the ROAST gene set analysis method (Wu et al., 2010) (FDR<0.25). For both screens, gene hits in ReGenT-seq are labelled. Gene scores were calculated as -log(P-value) x (number of enriched - number of depleted shRNAs). **(F)** Protein-protein interaction network of all Viability and ReGenT-seq hits. Nodes represent the gene hits, edges indicate physical interactions, and edge thickness reflects interaction strength. K-means clustering of nodes is shown (top). Selected known chromatin-associated complexes are highlighted within each cluster (bottom). Viability hits without known interaction are not displayed but were included in the clustering analysis. All network analyses were performed using the STRING 11.5 database. **(G)** STRING-clusters coloured according to Viability screening score, highlighting that gene hits belonging to the same protein complex tend to exert coordinated effects on cell viability. **(H)** (left) Heatmap showing hierarchical clustering of Viability and ReGenT-seq gene hits (Viability regulators and CA regulators, respectively) based on major network metrics. Viability and ReGenT-seq scores are shown in green and orange, respectively (Top right). Bar plots showing Pearson's correlation between each network metric and the observed screening score for viability and CA regulators (right). This comparison reveals that the CA regulators exhibit higher network centrality than viability regulators, suggesting greater interaction promiscuity.



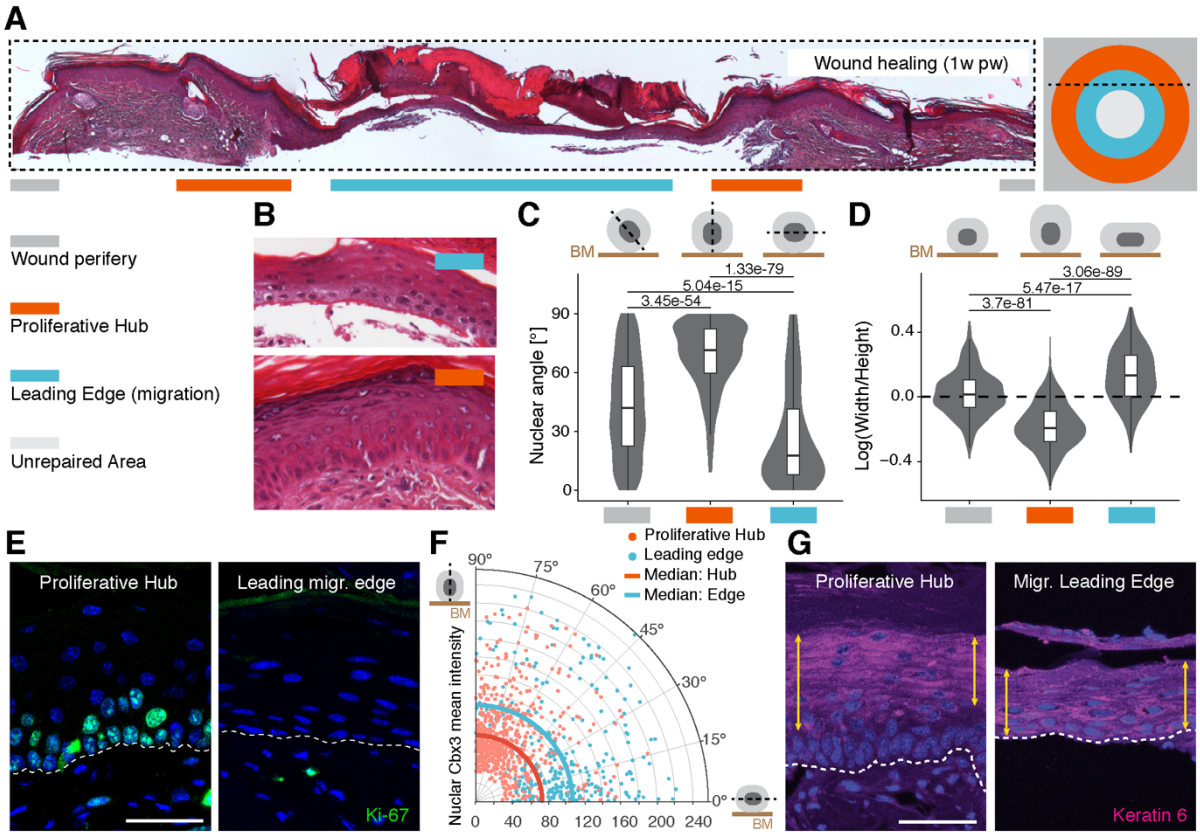
Supplementary Fig.5 | Single-gene knockdown and RNA-seq analysis of chromatin factor hits identified by ReGenT-seq. (A) RT-qPCR analyses showing shRNA-mediated knockdown (KD) of *Cbx3*, *Kmt5c* and *Ehmt2* in EPCs. mRNA levels are represented as relative expression using the 2-ΔΔCt method and normalized to control samples. n=3 independent replicates. Statistical significance was assessed using an unpaired two-sided Student's t-test. **(B)** Western blot showing CBX3 protein levels in EPCs transduced with shRNAs targeting *Cbx3* or scramble control. **(C)** PCA of RNA-seq data (5,000 most variable genes) comparing CBX3, KMT5C, and EHMT2 KD cells with control cells. n=5 independent replicates. (**D**) Heatmap showing pairwise differential gene expression (DGE) correlations among the selected CA regulators. The heatmap visualizes the clustered pairwise correlation matrix obtained comparing log2FC values between KD and control samples for each regulator. **(E)** Bar plot showing the number of DEGs between control and each individual KD condition. Genes are categorized by standard annotation (Grey; official gene symbols), or MGI annotation (black) when standard annotation is unavailable. **(F)** Bar plots showing observed and expected genes for each class in MGI

gene annotation (lncRNA, Pseudogene, Protein coding and Unclassified genes). The observed-to-expected ratio is colour-coded. **(G)** Heatmap showing, for each KD condition, hierarchical clustering based on Normalized Enrichment Score (NES) from GSEA performed on cell-type signature gene sets from the Descartes Atlas[37]. GSEA was carried out using a running-sum statistic with permutation-based testing, with P values adjusted using the Benjamini–Hochberg method.



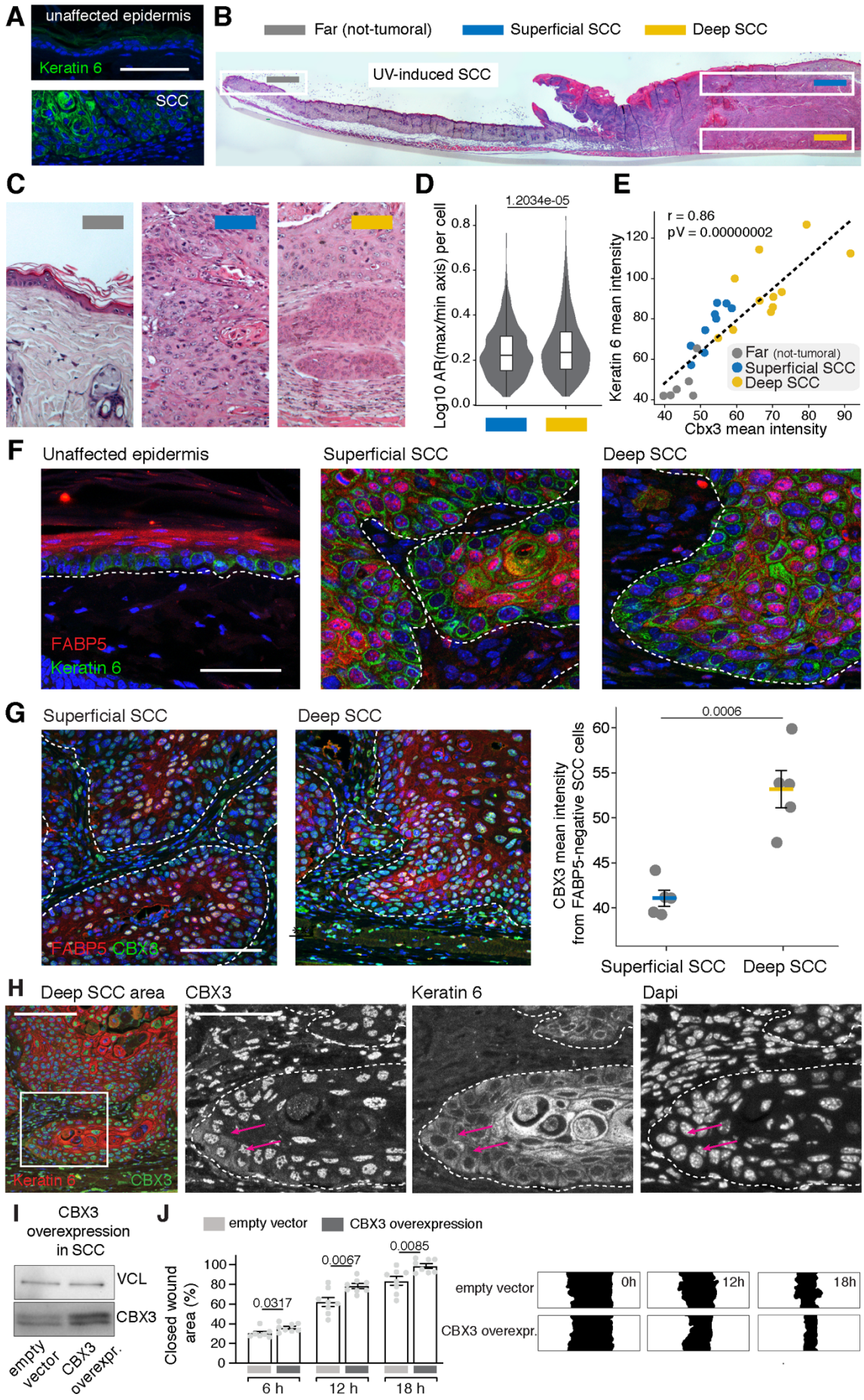
Supplementary Fig.6 | Roles of *Cbx3* and *Kmt5c* in cell motility and proliferation. (A) Scratch assay after Kmt5c knockdown (KD). Scratch masks and percentage of wound area closure at different time points are shown in the upper and lower panels, respectively. (**B)** Western blot showing CBX3 overexpression in EPCs. (**C)** Scratch assay performed with CBX3-overexpressing and control EPCs. Representative scratch masks (upper panel) and quantification of wound area closure over time (lower panel) are shown. n=16 independent replicates. **(D)** (left) Immunostaining for Ki-67 in CBX3 KD (shRNA1 and shRNA2) cells compared with control (scramble) cells. (Right) Box plot showing the percentage of Ki-67-positive cells relative to the total cell number. n=40 images. The horizontal line indicates the median, boxes represent the interquartile range (IQR), and whiskers denote the full data range. **(E)** Cumulative proliferation ratio in CBX3 KD cells. The proliferation ratio, defined as the number of cells at each time point divided by the number of cells at time 0, is shown over the time course (t0-t4). n=6 cell counts. Graphs shown in this figure represent data from 2 or 3 independent

experiments (2 for A and C, and 3 for E) independent experiments. Data are shown as means ± SEM. In all graphs, statistical significance was assessed using an unpaired two-sided Student's t-test, unless otherwise stated. Scale bar: 100μm (D).

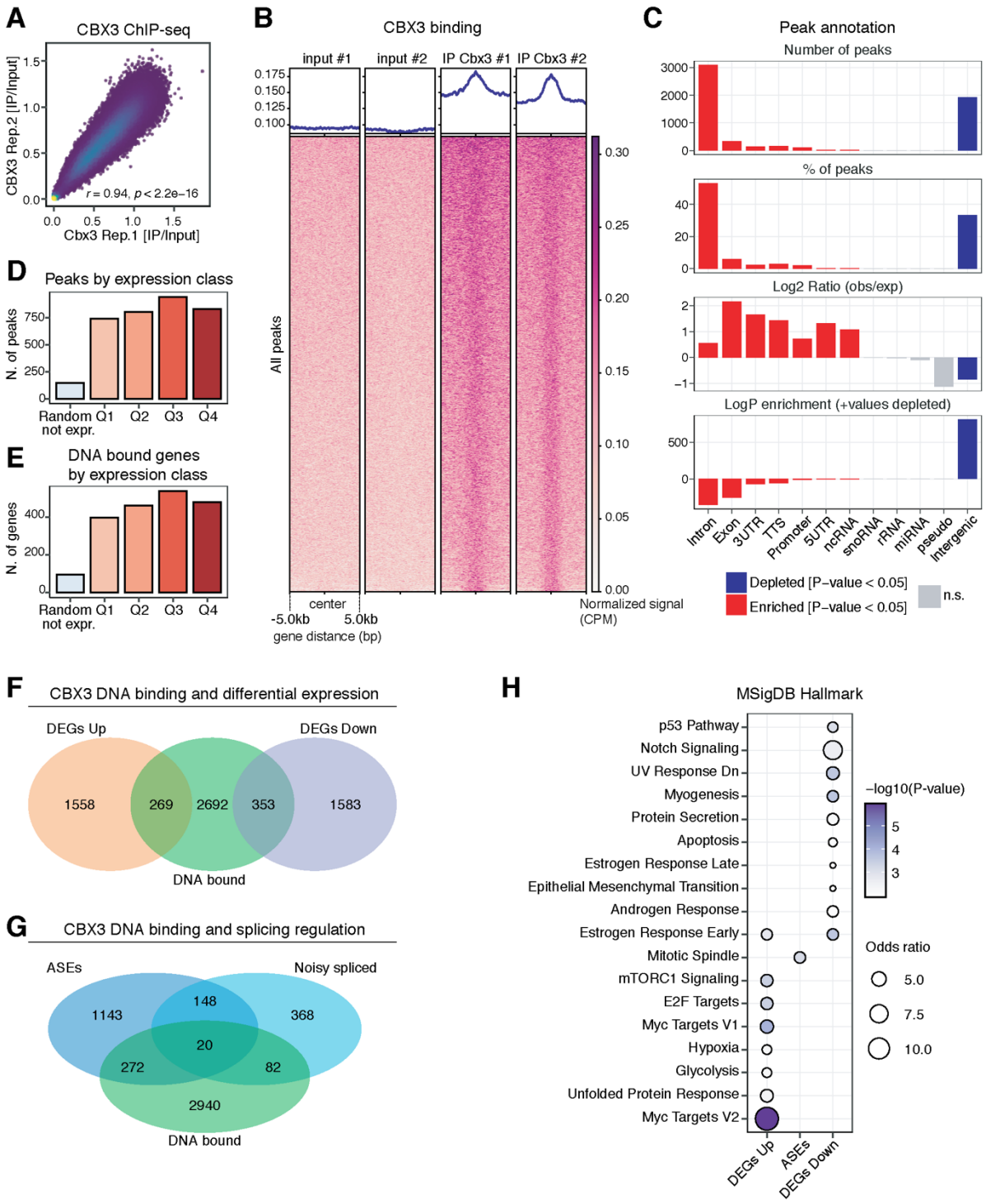


Supplementary Fig.7 | Histological stratification of wound healing regions. (A) H&E staining of skin sections at 1 week post-wounding (1w pw). Different zones relative to the wound center are indicated and schematically represented in their original position within the skin section. The dotted line indicates the sectioning site in the circular wound schematics (right panel). **(B)** H&E staining of the leading edge (upper panel) and the proliferative hub (lower panel), representing wound areas where EpSCs migrate and proliferate, respectively. **(C)** Basal cell orientation in skin sections was quantified as the angle between the nucleus and the basal membrane. Statistical significance was assessed using a t-test, with P-values adjusted using the Benjamini–Hochberg procedure. **(D)** Quantification of the rectangular cell shape descriptor (log(width/height)). Violin plots compare the wound periphery (grey), the proliferative hub (red), and the leading edge (light blue). Statistical significance was assessed using a t-test, with P-values adjusted using the Benjamini–Hochberg procedure. **(E)** Ki-67, a proliferation marker, in skin sections validates the proliferative hub and is absent in the leading edge. Representative

images of the proliferative hub and the leading edge are shown. **(F)** Correlation between nuclear CBX3 and basal cell position relative to the basal membrane, where each dot represents a single cell. The median CBX3 intensity in the proliferative hub and the leading edge is indicated. **(G)** Keratin 6 expression in the proliferative hub and the leading edge. Yellow double arrows indicate Keratin 6-positive zones in skin sections. For all graphs in the figure, n= 405 (wound periphery, ctrl), n=486 (proliferative hub), and n=263 (leading edge) images. Scale bar: 50μm (E, G).

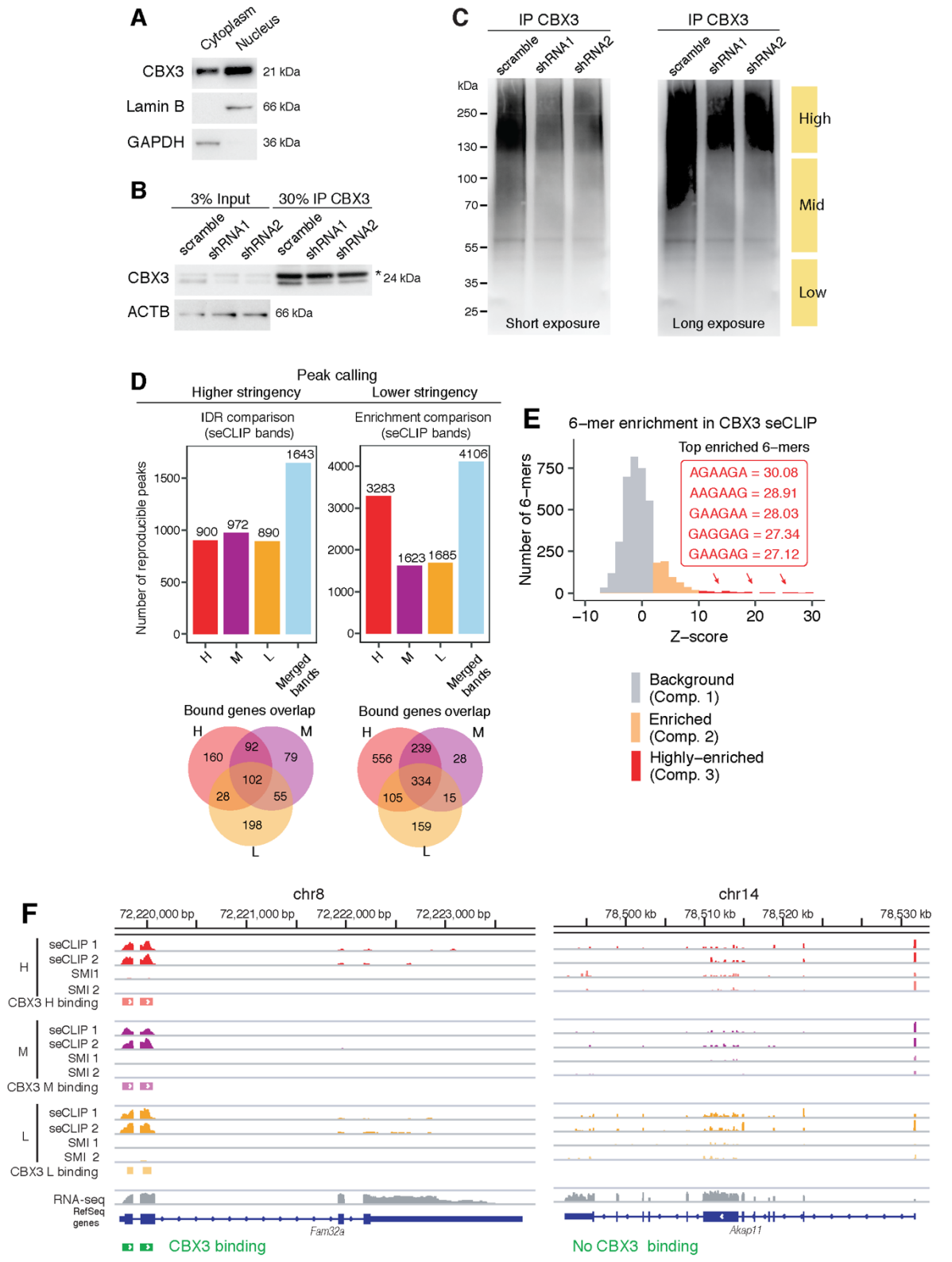

A
unaffected epidermis
Keratin 6
SCC
B
Far (not-tumoral)
Superficial SCC
Deep SCC
UV-induced SCC
C
D
1.2034e-05
Log10 AR(max/min axis) per cell
0.0
0.2
0.4
0.6
0.8
E
r = 0.86
pV = 0.00000002
Keratin 6 mean intensity
Cbx3 mean intensity
Far (not-tumoral)
Superficial SCC
Deep SCC
F
Unaffected epidermis
Superficial SCC
Deep SCC
FABP5
Keratin 6
G
Superficial SCC
Deep SCC
FABP5 CBX3
0.0006
CBX3 mean intensity from FABP5-negative SCC cells
Superficial SCC
Deep SCC
H
Deep SCC area
CBX3
Keratin 6
Dapi
Keratin 6
CBX3
I
CBX3 overexpression in SCC
VCL
CBX3
empty vector
CBX3 overexpr.
J
empty vector
CBX3 overexpression
Closed wound area (%)
0.0317
0.0067
0.0085
6 h
12 h
18 h
empty vector
CBX3 overexpr.
0h
12h
18h

Supplementary Fig.8 | Histological stratification of SCC regions. (A) Immunostaining of Keratin 6 in murine SCCs and unaffected epidermis, showing higher expression in tumors. **(B)** H&E staining of UV-induced SCCs. Far-, superficial-, and deep- SCC areas are distinguished by coloured bars (grey, blue, and yellow, respectively). **(C)** Higher magnification images of the far-, superficial- and deep- SCC regions shown in (B). **(D)** Cell shape descriptor (AR=major axis/minor axis of each nucleus) used to distinguish cells in superficial versus deep SCC areas. Violin plots show Welch's t-test results. n=1274-1802 cells form three samples. **(E)** Scatter plot showing the correlation between CBX3 and Keratin 6 levels in the three different zones defined from the wound. Pearson's correlation coefficient (r) and P-value are indicated. n=9 images form three samples. **(F)** Immunofluorescence images showing the epidermal stem/progenitor cell marker Keratin 14 and the early epidermal differentiation marker FABP5, across far-, superficial-, and deep- SCC regions, and in unaffected epidermis. (G) Immunostaining of CBX3 and FABP5 in superficial and deep SCC areas (left panels). Quantification of CBX3 in FABP5-negative epidermal cells is shown (right panels). Data are shown as means ± SEM. n=5 images from 5 independent regions. **(H)** Section of a deep SCC area stained with anti-CBX3 (green) and anti-Keratin 6 (red). Single channel images for CBX3, Keratin 6, and DAPI are shown. The dotted line to indicate the basal membrane. Magenta arrows highlight rare basal cells with nuclear and cytoplasmic staining. (**I**) Western blot showing CBX3 overexpression in SCC cells. **(L)** Scratch assay in CBX3-overexpressing and control SCC cells, showing representative masks (left) and wound closure quantification over time (right). Data are shown as means ± SEM. n=7-8. In all graphs, statistical significance was assessed using an unpaired two-sided Student's t-test, unless otherwise stated. Scale bar: 50μm (F, H-right), 100μm (B, G, H-left), and 1000 μm (A).



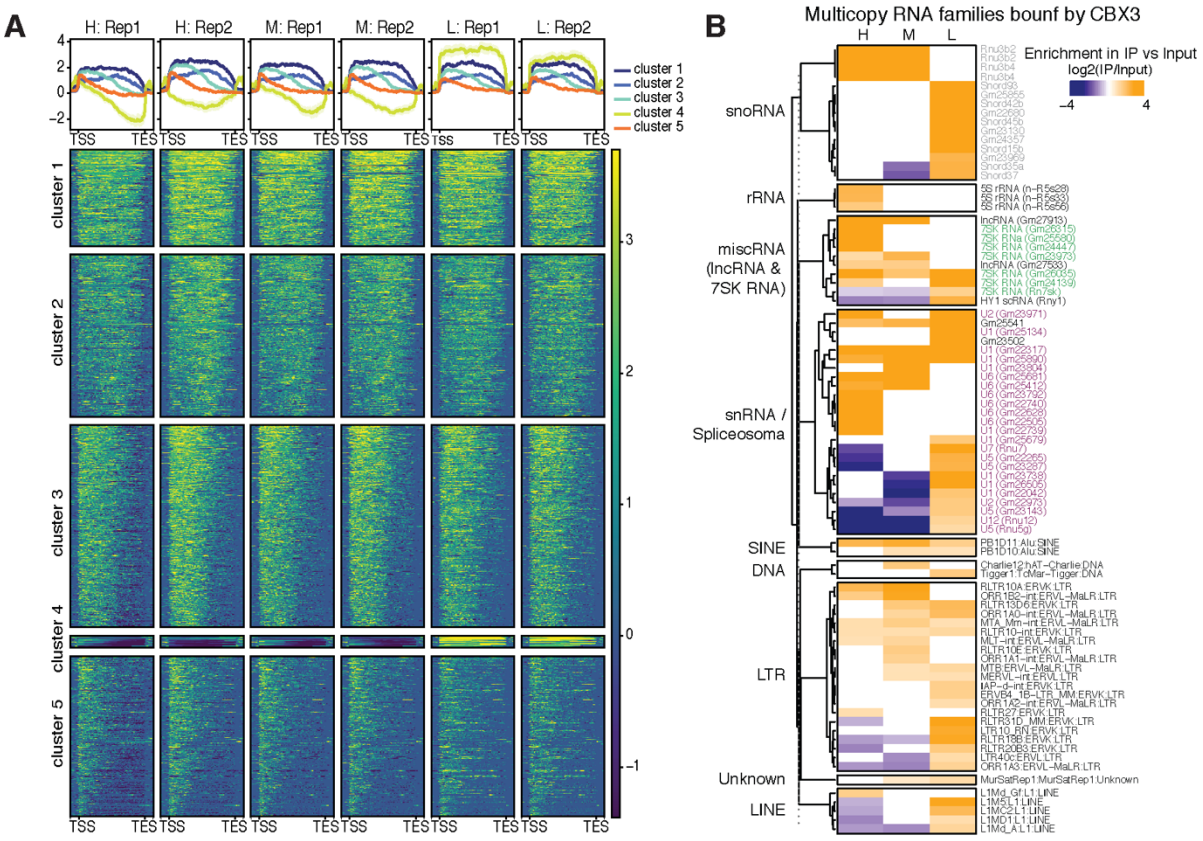
Supplementary Fig.9 | Integration of transcriptomic profiling after CBX3 knockdown with ChIP-seq of endogenous CBX3. (A) Scatter plot showing reproducibility between the two CBX3 ChIP-seq replicates, as measured by Pearson's correlation coefficient (r) of their average signal intensities

(IP/input) in 400-bp bins. **(B)** Heat map and profile plot showing ChIP-seq signal in CBX3 peaks. Signal for input and IP samples is reported in a 5 kb-window aligned on peak centers. **(C)** Annotation of CBX3 ChIP-seq peaks to the indicated genomic features, reporting number and percentage of peaks, observed over expected enrichment ratio, and its associated P-value. **(D, E)** Bar plots showing the distribution of CBX3 ChIP-seq peaks in gene bodies (d) and CBX3 bound genes by expression level (e) (Q1 = bottom quartile, Q4 = top quartile, including a random equally-sized set of not expressed genes as control). **(F)** Venn diagram showing the intersection of CBX3 DNA direct target genes (ChIP-seq data) with DEGs in CBX3 KD vs control EPCs (up- and down-regulated, RNA-seq data). **(G)** Venn diagram showing the intersection of CBX3 DNA direct target genes (ChIP-seq data) with genes affected by CBX3-mediated splicing regulation (ASEs and noisy spliced). **(H)** Gene set enrichment analysis of differentially expressed genes (DEGs) in CBX3 KD versus control EPCs (RNA-seq data), restricted to CBX3 direct DNA target genes identified by ChIP-seq, using the MSigDB Hallmark gene set collection. Enrichment analysis was performed using Enrichr. Dot size indicates the odds ratio, while dot color represents the −log10(P-value).

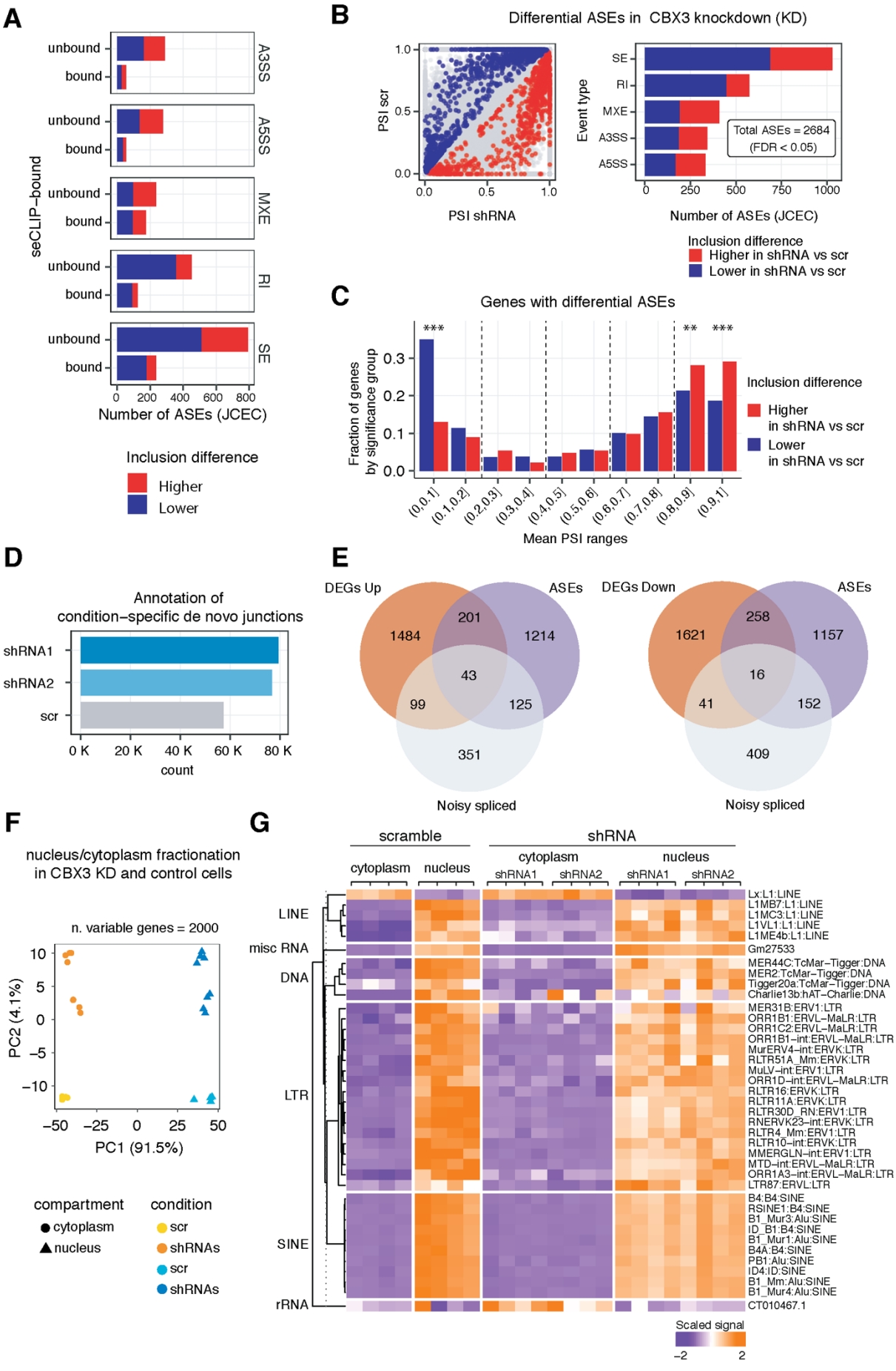


Supplementary Fig.10 | seCLIPs of endogenous CBX3. (A) Western blot analysis of cell fractionation. Cytoplasmic and nuclear proteins are shown. **(B)** CBX3 enrichment in seCLIP IP

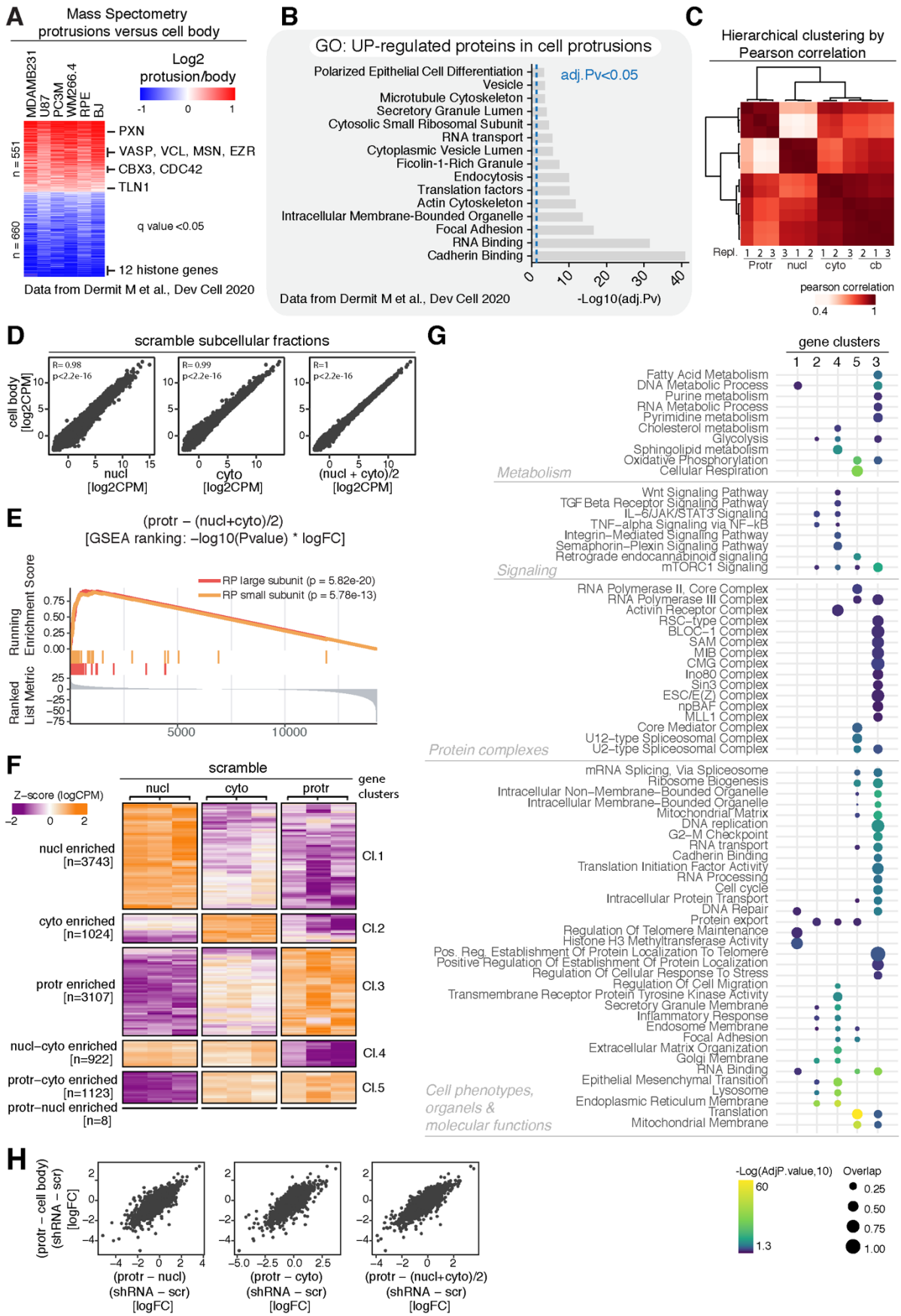
compared to INPUT is shown for scramble and shRNAs samples. (**C**) Western blot of immunoprecipitated endogenous CBX3 used in the seCLIP experiment. H, M, L refer to CBX3/RNA complexes migrating as a high weight (H), middle weight (M) and low weight protein complex (L). L corresponds to the weight of a single CBX3 protein. Scramble vs shRNAs is compared in a short- and long-time exposure. RNA material extracted for seCLIP library preparation is denoted by yellow bars. Asterisk marks antibody light chains. **(D)** Results of the higher stringency (left, IDR comparison analysis, P-value<0.001& FC>8) and lower stringency (right, overlapping enrichment analysis P-value<0.001& FC>1) seCLIP peak calling strategy are illustrated. Bar plots indicate the number of CBX3 seCLIP peaks in H, M, L, and all merged bands. Overlap of peaks and genes for the three bands is also shown as Venn diagrams. H, M, L refer to CBX3/RNA complexes migrating as a high weight (H), middle weight (M) and low weight protein complex (L), as illustrated in (**C**) L corresponds to the weight of a single CBX3 protein; M corresponds to the weight of two CBX3 proteins, potentially representing a CBX3 dimer[71]. **(E)** Z-score distribution relative to hexamers (6-mers) enrichment in CBX3 seCLIP peaks. Background (not enriched, grey) identified with a Gaussian mixture model, enriched (orange) and highly-enriched (red) 6-mers clusters are shown. Z-score values of top-5 enriched 6-mers are highlighted in red. **(F)** Representative IGV snapshot showing CBX3 seCLIP signal profile and peaks distribution in Fam32a (CBX3-bound) and an unbound control gene (Akap11). For detailed information about seCLIP analyses, see the Methods section.



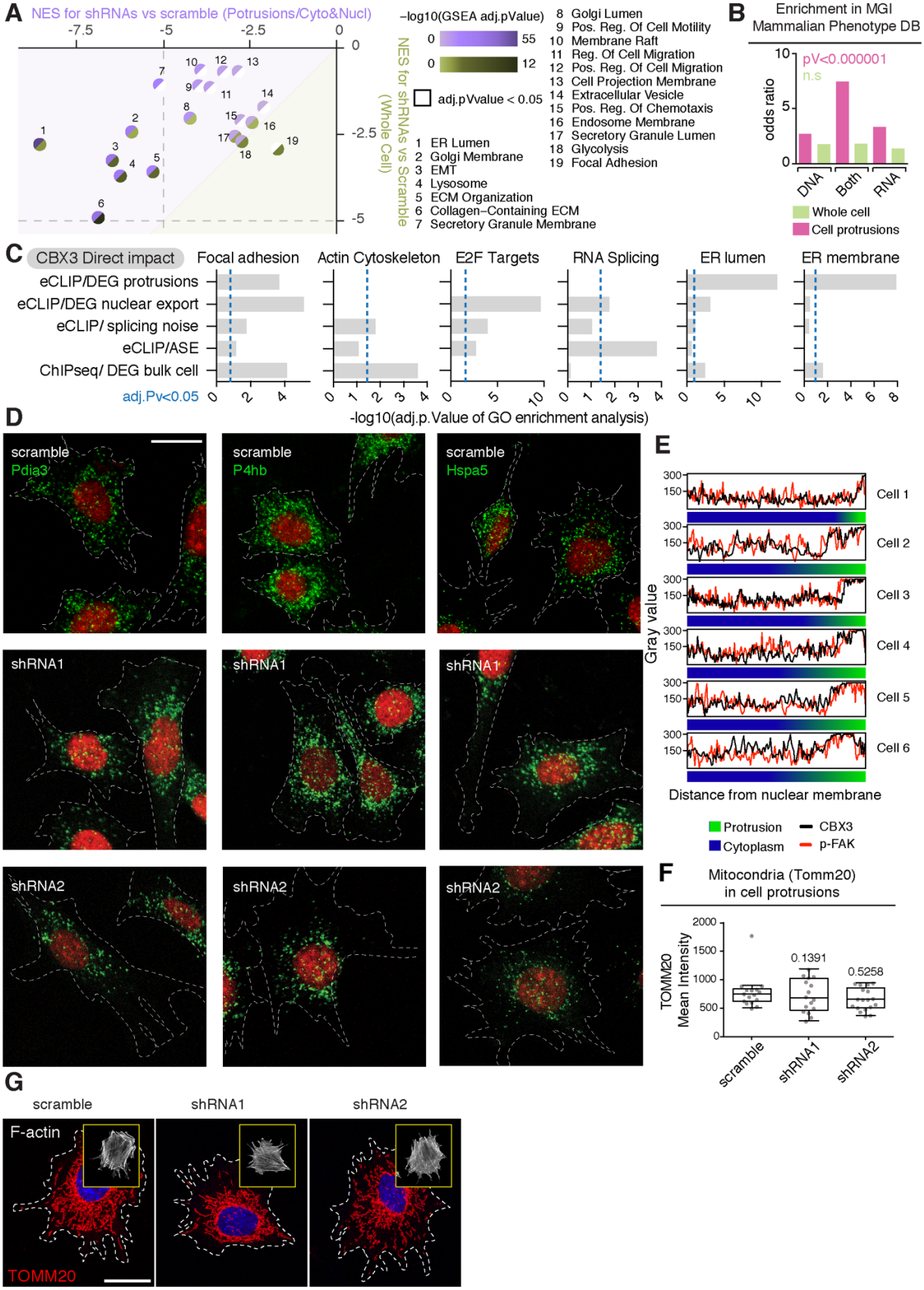
Supplementary Fig.11 | CBX3 RNA-binding profiles and multicopy RNA families bound by CBX3. (A) Heatmap of seCLIP-seq normalized read density profile at CBX3-bound gene bodies in two replicates for the three H, M, L bands (relative to Supplementary Fig.10C). The clustering analysis showed that CBX3 binding on spliced RNAs tends to be towards the 5'. In addition, most target RNAs are shared by the different size CBX3/RNA complexes, except for an RNA cluster 4 bound by CBX3/RNA complex characterized by the presence of a single CBX3 protein. **(B)** Multicopy RNAs families Enrichment analysis (IP vs Input) on CBX3-bound RNA in all the three bands. Rows are multicopy elements, columns are seCLIP bands, and the enrichment score (log2FC of IP vs size-matched input) is visualized. For detailed information about seCLIP analyses, see the Methods section.


A
unbound
bound
A3SS
A5SS
MXE
RI
SE
seCLIP−bound
0 200 400 600 800
Number of ASEs (JCEC)
Inclusion difference
Higher
Lower
B
Differential ASEs in CBX3 knockdown (KD)
PSI scr
PSI shRNA
Event type
Total ASEs = 2684
(FDR < 0.05)
Number of ASEs (JCEC)
Inclusion difference
Higher in shRNA vs scr
Lower in shRNA vs scr
C
Genes with differential ASEs
Fraction of genes by significance group
Mean PSI ranges
Inclusion difference
Higher in shRNA vs scr
Lower in shRNA vs scr
D
Annotation of condition−specific de novo junctions
shRNA1
shRNA2
scr
count
E
DEGs Up
ASEs
DEGs Down
Noisy spliced
F
nucleus/cytoplasm fractionation in CBX3 KD and control cells
n. variable genes = 2000
PC2 (4.1%)
PC1 (91.5%)
compartment
cytoplasm
nucleus
condition
scr
shRNAs
G
scramble
shRNA
cytoplasm
nucleus
shRNA1
shRNA2
LINE
misc RNA
DNA
LTR
SINE
rRNA
Scaled signal

Supplementary Fig.12 | Analysis of CBX3-dependent RNA splicing and nuclear export. (A) Differential Alternative Splicing Events (ASEs) in CBX3 knockdown (KD), grouped by CBX3-bound and unbound mRNAs (FDR < 0.05). n=3 replicates. Higher = higher percentage-spliced-in (Ψ) in KD vs scramble. Lower = lower Ψ in KD vs scramble. A3SS and A5SS: alternative 3'/5' splice sites, MXE: mutually exclusive exons, RI: retained, SE: skipped exons. **(B)** Differential alternative splicing event (ASE) analysis after CBX3 KD. Scatter plot showing exon-inclusion ratio difference between shRNA and scramble as percent spliced-in (ψ) (left). Number of ASEs as Junction and Exon Counts (JCEC) across event types. The number of total ASEs with FDR<0.05 is indicated within the graph (right). In both graphs red is used for higher exon inclusion in CBX3 KD vs control cells, while blue for lower inclusion with respect to scramble. **(C)** Higher and lower inclusion difference fraction of genes with differential ASEs, distributed over mean PSI ranges. **(D)** Read count of annotated *de novo* junctions specific to each condition. **(E)** Venn diagram with differentially expressed genes (DEGs) between CBX3 KD vs control cells (up and down, left, and right panel, respectively), CBX3-dependent-ASEs and CBX3-dependent Noisy spliced genes. The number of overlapping genes among these groups is reported. **(F)** Principal component analysis of RNA-seq samples from CBX3 KD and control cells after nucleus/cytoplasm fractionation. The top 2000 variable genes are used as input features. n=3 independent replicates **(G)** Heatmap showing enrichment of multicopy RNA families based on RNA-seq data from nuclear and cytoplasmic fractions of control and Cbx3 KD cells. Rows represent multicopy elements, columns represent samples, and scaled normalized concentration levels are shown, indicating that Cbx3 maintains these RNAs in the nucleus. For detailed information about all analyses shown in this figure, see the Methods section.


A
Mass Spectometry
protrusions versus cell body
MDAMB231
U87
PC3M
WM266.4
RPE
BJ
Log2
protusion/body
-1
0
1
n = 551
n = 660
PXN
VASP, VCL, MSN, EZR
CBX3, CDC42
TLN1
q value <0.05
12 histone genes
Data from Dermit M et al., Dev Cell 2020
B
GO: UP-regulated proteins in cell protrusions
adj.Pv<0.05
Polarized Epithelial Cell Differentiation
Vesicle
Microtubule Cytoskeleton
Secretory Granule Lumen
Cytosolic Small Ribosomal Subunit
RNA transport
Cytoplasmic Vesicle Lumen
Ficolin-1-Rich Granule
Endocytosis
Translation factors
Actin Cytoskeleton
Intracellular Membrane-Bounded Organelle
Focal Adhesion
RNA Binding
Cadherin Binding
0
10
20
30
40
Data from Dermit M et al., Dev Cell 2020
-Log10(adj.Pv)
C
Hierarchical clustering by
Pearson correlation
Repl. 1 2 3 3 1 2 1 2 3 2 1 3
Protr
nucl
cyto
cb
pearson correlation
0.4
1
D
scramble subcellular fractions
R= 0.98
p<2.2e-16
R= 0.99
p<2.2e-16
R=1
p<2.2e-16
cell body
[log2CPM]
nucl
[log2CPM]
cyto
[log2CPM]
(nucl + cyto)/2
[log2CPM]
E
(protr – (nucl+cyto)/2)
[GSEA ranking: –log10(Pvalue) * logFC]
RP large subunit (p = 5.82e-20)
RP small subunit (p = 5.78e-13)
Running
Enrichment Score
Ranked
List Metric
5000
10000
F
scramble
nucl
cyto
protr
gene
clusters
Z-score (logCPM)
–2
0
2
nucl enriched
[n=3743]
cyto enriched
[n=1024]
protr enriched
[n=3107]
nucl–cyto enriched
[n=922]
protr–cyto enriched
[n=1123]
protr–nucl enriched
[n=8]
Cl.1
Cl.2
Cl.3
Cl.4
Cl.5
G
gene clusters
1 2 4 5 3
Fatty Acid Metabolism
DNA Metabolic Process
Purine metabolism
RNA Metabolic Process
Pyrimidine metabolism
Cholesterol metabolism
Glycolysis
Sphingolipid metabolism
Oxidative Phosphorylation
Cellular Respiration
Metabolism
Wnt Signaling Pathway
TGF Beta Receptor Signaling Pathway
IL–6/JAK/STAT3 Signaling
TNF–alpha Signaling via NF–kB
Integrin–Mediated Signaling Pathway
Semaphorin–Plexin Signaling Pathway
Retrograde endocannabinoid signaling
mTORC1 Signaling
Signaling
RNA Polymerase II, Core Complex
RNA Polymerase III Complex
Activin Receptor Complex
RSC–type Complex
BLOC–1 Complex
SAM Complex
MIB Complex
CMG Complex
Ino80 Complex
Sin3 Complex
ESC/E(Z) Complex
npBAF Complex
MLL1 Complex
Core Mediator Complex
U12–type Spliceosomal Complex
U2–type Spliceosomal Complex
Protein complexes
mRNA Splicing, Via Spliceosome
Ribosome Biogenesis
Intracellular Non–Membrane–Bounded Organelle
Intracellular Membrane–Bounded Organelle
Mitochondrial Matrix
DNA replication
G2–M Checkpoint
RNA transport
Cadherin Binding
Translation Initiation Factor Activity
RNA Processing
Cell cycle
Intracellular Protein Transport
DNA Repair
Protein export
Regulation Of Telomere Maintenance
Histone H3 Methyltransferase Activity
Pos. Reg. Establishment Of Protein Localization To Telomere
Positive Regulation Of Establishment Of Protein Localization
Regulation Of Cellular Response To Stress
Regulation Of Cell Migration
Transmembrane Receptor Protein Tyrosine Kinase Activity
Secretory Granule Membrane
Inflammatory Response
Endosome Membrane
Focal Adhesion
Extracellular Matrix Organization
Golgi Membrane
RNA Binding
Epithelial Mesenchymal Transition
Lysosome
Endoplasmic Reticulum Membrane
Translation
Mitochondrial Membrane
Cell phenotypes,
organels &
molecular functions
H
(protr – cell body)
(shRNA – scr)
[logFC]
(protr – nucl)
(shRNA – scr)
[logFC]
(protr – cyto)
(shRNA – scr)
[logFC]
(protr – (nucl+cyto)/2)
(shRNA – scr)
[logFC]
-Log(AdjP.value,10)
60
1.3
Overlap
0.25
0.50
0.75
1.00

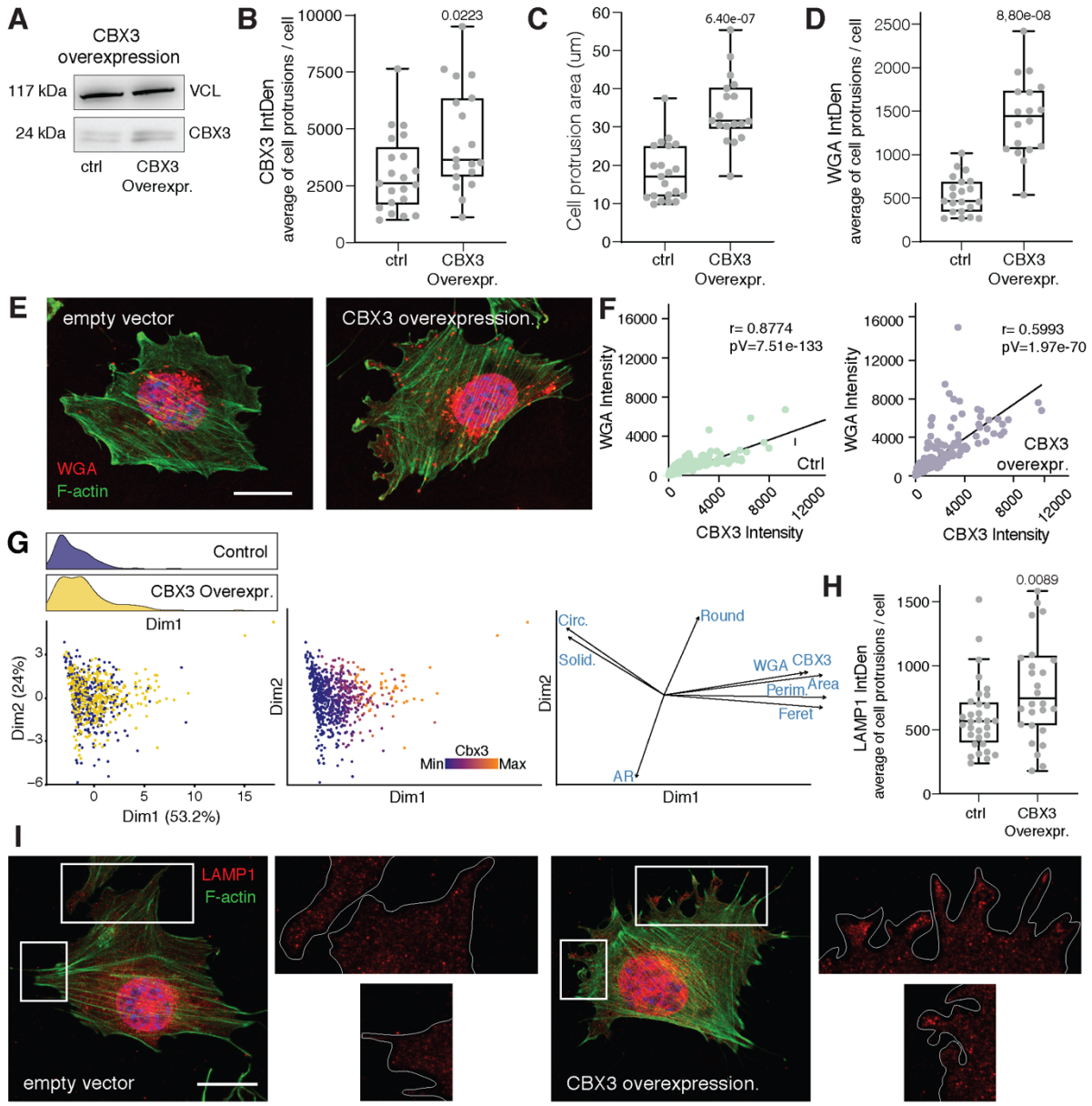
Supplementary Fig.13 | Compartment-specific localization of RNAs in the nucleus, cytoplasm, and cell protrusions. (A) Quantitative proteomics reanalysis showing protein levels in protrusions relative to cell bodies, across 6 cell lines (data and analysis from Dermit et al.[55]). Log_2 of protrusion/cell body protein ratio values from each cell line is shown. CBX3 and other control proteins (known to localize in the nucleus or in cell protrusions) are indicated. **(B)** GO analysis of upregulated proteins in cell protrusion (data from Dermit M. et al.[55]). Selected GO terms are ranked by -Log10(adj. P-value). **(C)** Hierarchical clustering of control RNA-seq samples from protrusion, nucleus, cytoplasm, and cell body compartments from EPCs. Hierarchical clustering by Ward's method was performed based on Person's correlation distance of normalized read counts per gene, and the clustered pairwise correlation matrix is visualized as heatmap. n=3 independent replicates per group. **(D)** Correlation plots of different subcellular fractions. Pearson's coefficient (r) and P-value are indicated within each graph. **(E)** GSEA enrichment plot for ribosome protein-large (red) and -small (orange) subunit gene sets in protrusion respect the sum of cytoplasm and nucleus fractions, in scramble cells. These represent positive controls for cell protrusion isolation[55]. P-value (p) is reported. **(F)** Heatmap of DEGs across subcellular fractions in control cells, clustered by compartmental enrichment. **(G)** GO analysis for each gene cluster from (F). GO term enrichment is shown by color code, while dot size represents the overlap between the number of DEGs and the gene universe in each list. **(H)** Scatter plot showing correlation of differences (reported as log2FC of ratios) between CBX3 knockdown (KD) and scramble EPCs in the protrusion to cell body ratio and the protrusion to nucleus, protrusion to cytoplasm, and protrusion to the sum of nucleus and cytoplasm ratios, respectively. Pearson's R and P-value from correlation test are reported.



Supplementary Fig.14 | *Cbx3*-dependent multilayered control of gene expression. (A) Correlation plot of negative normalized enrichment scores (NES) from GSEA, calculated using a running-sum

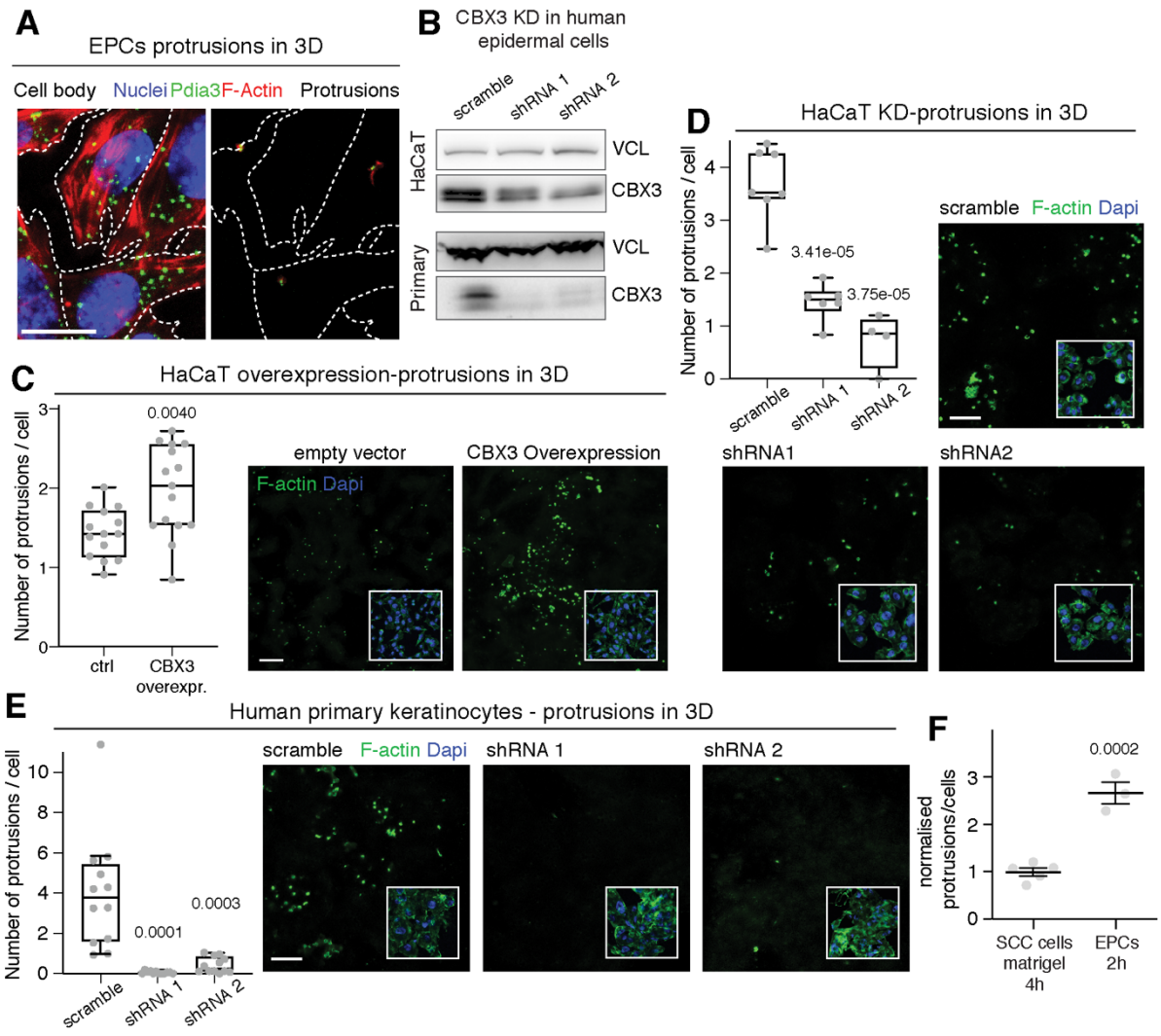
statistic with permutation-based testing. P-values were adjusted using the Benjamini–Hochberg procedure. GSEA was performed on transcriptomic data from subcellular compartments (protrusions relative to the cytoplasm and nucleus; violet) and from whole cells, comparing shRNA-treated cells with scramble controls (green). Enriched Gene Ontology (GO) terms, derived from Fig. 5H, are presented as a numbered list. The −log10(adjusted P-value) is represented by dot color, with white indicating non-significant results. **(B)** Bar plot showing enrichment analysis using the MGI Mammalian Phenotype database of DEGs in CBX3 knockdown (KD) versus control cells, from Fig. 5H. Genes are classified as direct chromatin targets based on ChIP-seq (DNA), RNA targets based on seCLIP (RNA), or targets identified by both methods. Enrichment analysis was performed using Enrichr. Enrichment is represented as an odds ratio with the associated P-value indicated in the plot. **(C)** Bar plots showing the direct impact of CBX3 on selected GO terms (from Fig. 5H) across distinct layers of gene expression regulation, expressed as −log10(adjusted P value). Direct regulatory layers were defined by the intersection of transcriptomic data from scramble versus CBX3 knockdown cells with genomic datasets (ChIP-seq or seCLIP). Specifically, these included: (i) seCLIP and RNA-seq comparing protrusions versus cytoplasm and nucleus (seCLIP/DEG protrusion); (ii) seCLIP and RNA-seq comparing cytoplasm versus nucleus (seCLIP/DEG nuclear export); (iii) seCLIP-seq and bulk RNA-seq for de novo junction analysis (seCLIP/splicing noise); (iv) seCLIP-seq and bulk RNA-seq for allele-specific expression analysis (seCLIP/ASE); and (v) ChIP-seq and bulk RNA-seq for whole-cell differential gene expression. This analysis reveals that gene sets such as focal adhesion are regulated at multiple layers of gene expression control, whereas others, including endoplasmic reticulum–associated genes, are regulated at a single layer, specifically within protrusions. **(D)** smFISH images of scramble, shRNA1 and 2 cells stained with probes against *Pdia3*, *P4hb* and *Hspa5* mRNAs (in Fig.6B the relative quantification is shown). For each image co-staining with DAPI (red) is also shown. Cell boundaries (white lines) were defined by F-actin. **(E)** Intensity line profile to show CBX3 (black) and p-FAK (red) distribution from nucleus (excluded) to protrusion through cytoplasm (from blue to green). **(F, G)** Immunofluorescence images (G) and corresponding quantification (F) of mitochondria in cell protrusions, detected using a TOM20 antibody. F-actin staining used for cell boundaries is shown in the yellow square. Dots in box plots represent the mean intensity value per cell. n=16-18 cells representative

of one of two independent experiments. Horizontal line indicates median, box indicates the interquartile range (IQR) and whiskers denote the 1.5 × IQR. Statistical significance was assessed using an unpaired two-sided Student's t-test.



Supplementary Fig.15 | CBX3 functional analysis in human epidermal cells. (**A**) Western blot showing CBX3 overexpression in HaCaT cells. **(B-D)** Quantification of CBX3 levels in protrusions, protrusion size, and wheat germ agglutinin (WGA) signal in control and CBX3-overexpressing HaCaT cells. Each dot represents the average of all protrusions per cell. n = 19–21 cells. **(E)** F-actin and WGA staining in control and CBX3-overexpressing HaCaT cells. **(F)** Scatter plot showing the correlation between WGA and CBX3 intensities in protrusions of control and CBX3-overexpressing HaCaT cells. Pearson coefficient (r) and P-value are indicated. n=413 protrusions from one of two independent

experiments. **(G)** Characterization of *Cbx3*-dependent features of protrusions in HaCaT cells. PCA of shape descriptors (AR=Aspect Ratio; Circ.=circularity; Solid.=Solidity; Feret=Feret's Diameter; Round=Roundness; Perim.=Perimeter; Area) and CBX3, and WGA intensities in CBX3-overexpressing (yellow) and control (blue) cells. Box plots highlight distributions along PCA dimension 1 (top panels). PCA data are colored by CBX3 intensity (middle panel) or represented by directional arrows for each descriptor included in the analysis (right panel). **(H, I)** LAMP1 immunostaining in HaCaT cells and quantification of lysosomal signal in cell protrusions of CBX3-overexpressing versus control cells. Each dot represents the average of all protrusions per cell. n=28-32 cells. For all box plots in the figure, dots represent data of one of two independent experiments performed. The horizontal line indicates the median, boxes represent the interquartile range (IQR), and whiskers denote the 1.5 × IQR. Statistical significance was assessed using unpaired two-sided Student's t-tests. Scale bars: 12.5 μm (E,I).



Supplementary Fig.16 | CBX3 controls 3D induced protrusions in human primary epidermal cells. (A) smFISH images of induced mature protrusions (3 um pores) and cell-body of EPCs. Cell body (left panel) and protrusions (right panel) were stained with probes against Pdia3 mRNA (green). Cell boundaries (dashed lines) were defined by F-acting staining. (**B**) Western blot showing CBX3 knockdown (KD) in HaCaT cells and in human primary keratinocytes. (**C, D**) F-actin and DAPI staining of HaCaT cells showing the induction of mature 3D protrusions in control, CBX3-overexpressing, scramble, and CBX3 KD conditions (right and lower panels). Insets highlight associated cell bodies (white squares). Quantification of mature protrusions per cell is shown in the left panels. n=14-15 and n=4-7 images for C and D respectively, from two independent replicates. (**E**) Representative F-actin

and DAPI staining of human primary keratinocytes illustrating mature 3D protrusion formation under scramble and CBX3 KD conditions (right). Cell bodies associated with protrusions are indicated in the insets (white squares). Protrusion numbers per cell are quantified in the left panels. n=12 images, from two independent replicates. **(F)**, Number of induced mature protrusions per cell in invading SCC cells (3um transwell+Matrigel, 4h) vs migrating EPCs (3um transwell, 2h). For all box plots in the figure, dots represent data of one of two independent experiments performed. The horizontal line indicates the median, boxes represent the interquartile range (IQR), and whiskers denote the 1.5 × IQR. Statistical significance was assessed using unpaired two-sided Student's t-tests. Scale bars: 12.5 μm (A) and 50 μm (C-E).