Functional Annotation of Human Long Non-Coding RNAs via Molecular Phenotyping

Jordan Ramilowski[1#], Chi Wai Yip[1#], Saumya Agrawal[1], Jen-Chien Chang[1], Yari Ciani[2], Ivan V Kulakovskiy[3], Mickael Mendez[4], Jasmine Li Ching Ooi[1], John F Ouyang[5], Nick Parkinson[6], Andreas Petri[7], Leonie Roos[8], Jessica Severin[1], Kayoko Yasuzawa[1], Imad Abugessaisa[1], Altuna Akalin[9], Ivan Antonov[10], Erik Arner[1], Alessandro Bonetti[1], Hidemasa Bono[11], Beatrice Borsari[12], Frank Brombacher[13], Carlo Vittorio Cannistraci[14], Christopher JF Cameron[17], Ryan Cardenas[15], Melissa Cardon[1], Howard Chang[16], Josée Dostie[17], Luca Ducoli[18], Alexander Favorov[3], Alexandre Fort[1], Diego Garrido[12], Noa Gil[19], Juliette Gimenez[20], Reto Guler[13], Lusy Handoko[1], Jayson Harshbarger[1], Akira Hasegawa[1], Yuki Hasegawa[1], Kosuke Hashimoto[1], Norihito Hayatsu[1], Peter Heutink[22], Tetsuro Hirose[23], Eddie L Imada[24], Masayoshi Itoh[21], Bogumil Kaczkowski[1], Aditi Kanhere[15], Emily Kawabata[1], Hideya Kawaji[21], Tsugumi Kawashima[1], Tom Kelly[1], Miki Kojima[1], Naoto Kondo[1], Haruhiko Koseki[1], Tsukasa Kouno[1], Anton Kratz[1], Mariola Kurowska-Stolarska[25], Andrew Tae Jun Kwon[1], Jeffrey Leek[24], Andreas Lennartsson[26], Marina Lizio[1], Fernando Lopez[1], Joachim Luginbühl[1], Shiori Maeda[1], Vsevolod Makeev[3], Luigi Marchionni[24], Yulia A Medvedeva[10], Aki Minoda[1], Ferenc Müller[15], Manuel Munoz Aguirre[12], Mitsuyoshi Murata[1], Hiromi Nishiyori[1], Kazuhiro Nitta[1], Shuhei Noguchi[1], Yukihiko Noro[1], Ramil Nurtdinov[12], Yasushi Okazaki[1], Valerio Orlando[27], Denis Paquette[17], Callum Parr[1], Owen JL Rackham[5], Patrizia Rizzu[22], Diego Fernando Sanchez[24], Albin Sandelin[29], Pillay Sanjana[15], Colin AM Semple[6], Harshita Sharma[1], Youtaro Shibayama[1], Divya Sivaraman[1], Takahiro Suzuki[1], Suzannah Szumowski[1], Michihira Tagami[1], Martin S Taylor[6], Chikashi Terao[1], Malte Thodberg[29], Supat Thongjuea[1], Vidisha Tripathi[30], Igor Ulitsky[19], Roberto Verardo[2], Ilya Vorontsov[3], Chinatsu Yamamoto[1], Robert S Young[6], J Kenneth Baillie[6], Alistair RR Forrest[31], Roderic Guigó[12], Michael M Hoffman[32], Chung Chau Hon[1], Takeya Kasukawa[1], Sakari Kauppinen[7], Juha Kere[26], Boris Lenhard[8], Claudio Schneider[2], Harukazu Suzuki[1], Ken Yagi[1], FANTOM consortium, Michiel de Hoon[1*], Jay W Shin[1*], Piero Carninci[1*]

[*] Corresponding authors
michiel.dehoon@riken.jp (M.D.H.)
jay.shin@riken.jp (J.W.S.)
carninci@riken.jp (P.C.)

[#] these authors contributed equally to this work

Affiliations
1 RIKEN, Center for Integrative Medical Sciences, Japan
2 Consorzio Interuniversitario per le Biotecnologie, Italy
3 Vavilov Institute of General Genetics, Russian Federation
4 University of Toronto, Canada
5 Duke NUS Graduate Medical School, Singapore
6 MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, United Kingdom
7 Center for RNA Medicine, Aalborg University, Denmark
8 Imperial College London, United Kingdom
9 Max Delbrück Center for Molecular Medicine, Germany
10 Research Center of Biotechnology of the Russian Academy of Sciences, Russian Federation
11 Research Organization of Information and Systems, Japan
12 Center for Genomic Regulation and Universitat Pompeu Fabra, Catalonia, Spain.
13 University of Cape Town, SAMRC & ICGEB, South Africa
14 Technische Universität Dresden, Germany
15 University of Birmingham, United Kingdom
16 Stanford University, United States
17 McGill University, Canada

18 Swiss Federal Institute of Technology in Zurich, Switzerland
19 Weizmann Institute of Science, Israel
20 Fondazione Santa Lucia, Italy
21 RIKEN, Preventive Medicine & Diagnostic Innovation Program, Japan
22 German Center for Neurodegenerative Diseases, Germany
23 Hokkaido University, Japan
24 Johns Hopkins University, United States
25 University of Glasgow, United Kingdom
26 Karolinska Institute, Sweden
27 King Abdullah University of Science and Technology, Saudi Arabia
29 University of Copenhagen, Denmark
30 National Centre for Cell Science, India
31 Harry Perkins Institute of Medical Research, Australia
32 Princess Margaret Cancer Centre, Canada

Abstract

Long non-coding RNAs (lncRNAs) constitute the majority of transcripts in mammalian genomes and yet, their functions remain largely unknown. We systematically suppressed 285 lncRNAs in human dermal fibroblasts and quantified cellular growth, morphological changes, and transcriptomic responses using Capped Analysis of Gene Expression (CAGE). The resulting transcriptomic profiles recapitulated the observed cellular phenotypes, yielding specific roles for over 40% of analyzed lncRNAs in regulating distinct biological pathways, transcriptional machinery, alternative promoter activity and architecture usage. Overall, combining cellular and molecular profiling provided a powerful approach to unravel the distinct functions of lncRNAs, which we highlight with specific functional roles for *ZNF213-AS1* and *lnc-KHDC3L-2.*

Introduction

Over 50,000 loci in the human genome transcribe long non-coding RNA (lncRNA) (Hon et al. 2017; Iyer et al. 2015), which are defined as transcripts at least 200 nt long with low or no protein-coding potential. While lncRNA genes outnumber protein-coding genes in mammalian genomes, they are comparatively less conserved (Ulitsky 2016), lowly expressed, and more cell-type-specific (Hon et al. 2017). However, the evolutionary conservation of lncRNA promoters (Carninci et al. 2005) and their structural motifs (Derrien et al. 2012) suggest that lncRNAs are fundamental biological regulators. To date, only a few hundred human lncRNAs have been extensively characterized (Quek et al. 2015; Volders et al. 2015; de Hoon et al. 2015), revealing their roles in regulating transcription (Engreitz, Ollikainen, et al. 2016), translation (Carrieri et al. 2012), and chromatin state (Gupta et al. 2010; Guttman and Rinn 2012; Guttman et al. 2011); (Ransohoff et al. 2018; Quinn and Chang 2016).

Our recent FANTOM5 analysis showed that 19,175 (out of ~27,000) human lncRNA loci are functionally implicated (Hon et al. 2017). However, genomic screens are necessary to comprehensively characterize each lncRNA. One common approach of gene knockdown followed by a cellular phenotype assay typically characterizes a small percentage of lncRNAs for a single observable phenotype. For example, a recent large-scale screening using CRISPR interference (CRISPRi) found that approximately 500 lncRNA loci (~3.7% of targeted lncRNA loci) are essential for cell growth or viability in a cell-type specific manner (Liu et al. 2017). In addition, CRISPR-Cas9 experiments targeting splice sites identified ~2.1% of lncRNAs that affect growth

of K562 chronic myeloid leukemia cells (Liu et al. 2018). A CRISPR activation study has also revealed ~0.11% lncRNAs to be important for drug resistance in melanoma (Joung et al. 2017).

As a part of the FANTOM 6 pilot project, we established an automated high-throughput cell culture platform to suppress 285 lncRNAs expressed in human primary dermal fibroblasts (HDF) using antisense LNA-modified gapmer antisense oligonucleotide (ASO) technology (Roux et al. 2017). We then quantified the effect of each knockdown on cell growth and morphology using real-time imaging, followed by Cap Analysis Gene Expression (CAGE; Murata et al. 2014) deep sequencing to reveal molecular pathways associated with each lncRNA. In contrast to cellular phenotyping, molecular phenotyping provides a detailed assessment of the response to an lncRNA knockdown at the molecular level, allowing biological functions to be associated to lncRNAs even in the absence of an observable cellular phenotype. We further investigated changes in promoter usage and architecture modulated by the knockdowns, and integrated genome-wide chromosome conformation capture data (Hi-C) with the CAGE data to explore regulatory functions of the lncRNA targets. All data produced for this study are publicly available at http://fantom.gsc.riken.jp/6/datafiles (user: review; password: f6plj0206) and analysis results can be explored at http://fantom.gsc.riken.jp/zenbu/reports/#FANTOM6

Results

Selection and ASO-mediated knockdown of lncRNA targets

Human dermal fibroblasts (HDF) are non-transformed primary cells that are commonly used for investigating cellular reprogramming (Takahashi et al. 2007; Ambasudhan et al. 2011), wound-healing (Li and Wang 2011), fibrosis (Kendall R., et al 2014), and cancer (Kalluri 2016). Fibroblasts are easy to transfect, proliferate in culture, and exhibit a distinctive spindle-shaped morphology indicative of their mesenchymal origin, making it an ideal model for cellular phenotypic screenings. Here, an unbiased selection of lncRNAs expressed in HDF was performed to choose 285 lncRNAs for functional interrogation (Methods; Supplementary Table S1, Fig. 1a-c). Using RNA-seq profiling of fractionated RNA we annotated the lncRNA subcellular localization in the chromatin (35%), nucleus (27%), or cytoplasm (38%) (Fig. 1d). We then designed a minimum of five non-overlapping antisense oligonucleotides (ASOs) against each lncRNA (Supplementary Table S2; Fig. 1e) and transfected them individually using an automated cell culture platform to minimize experimental variability (Fig. 1f).

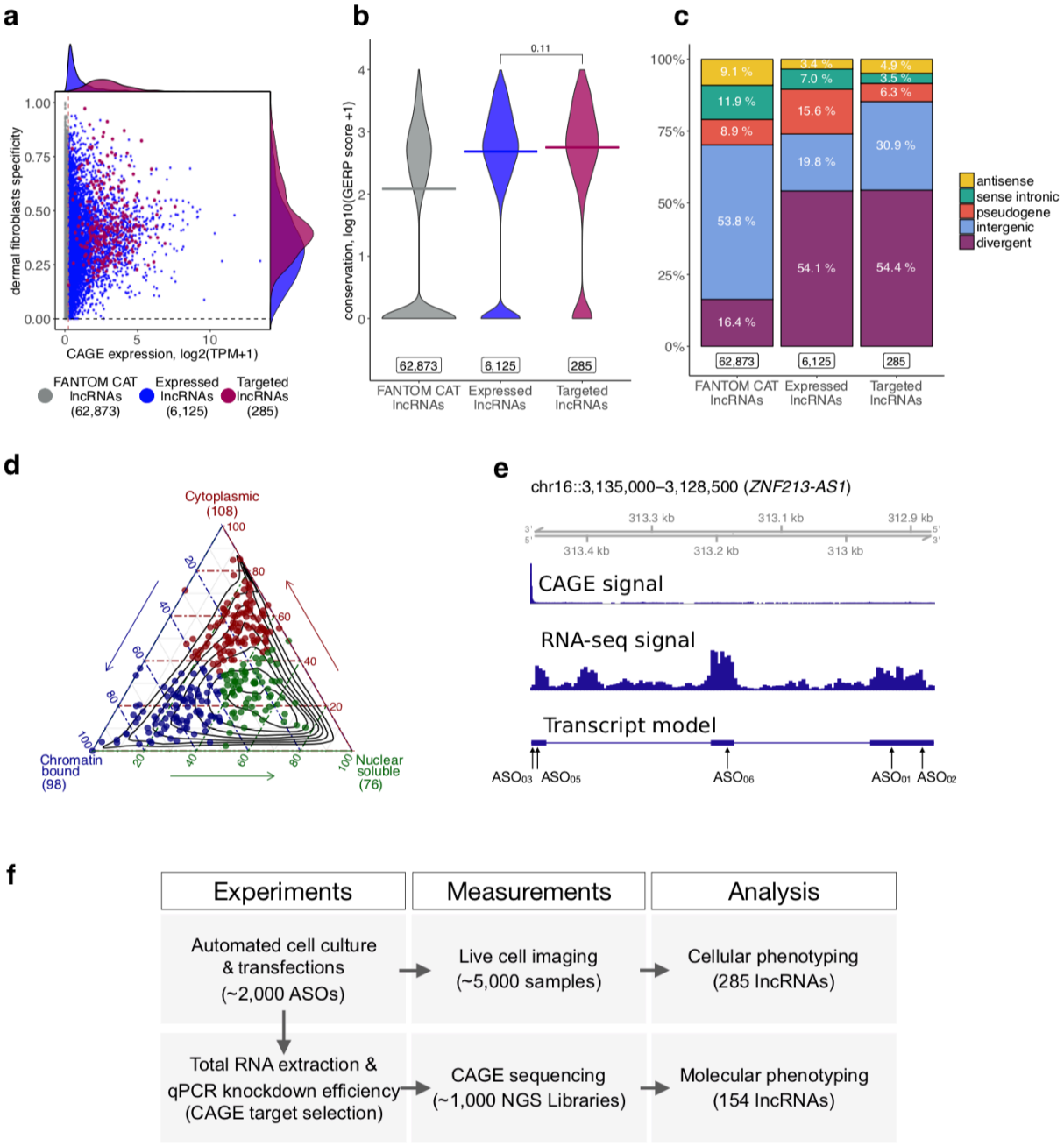

Fig. 1: Selection of lncRNA targets and the study overview. a, CAGE expression levels (TPM tags per million) and human dermal fibroblasts (HDF) specificity (Hon, et al., Nature 2017) of lncRNAs in the FANTOM CAT catalog (n=69,283; grey), lncRNAs expressed in HDFs (n=6,125; blue) and targeted lncRNAs (n=285; red). Dashed vertical line indicates most lowly expressed lncRNA target (~0.2 TPM). b, Gene conservation levels of lncRNAs in the FANTOM CAT catalog lncRNAs (grey), expressed in HDF (blue) and targeted lncRNAs (red). Crossbars indicate the median. No significant difference is observed when comparing targeted and expressed in HDF lncRNAs (Wilcoxon pvalue=0.11). c, Similar to that in b, but for genomic classes of lncRNAs. Most of the targeted lncRNAs and those expressed in HDF are expressed from divergent promoters (Hon, et al., Nature 2017). d, Subcellular localization (based on relative abundances from RNA-seq fractionation data) for targeted lncRNAs. Chromatin-bound (n=98; blue); Nuclear soluble

(n=76; green); Cytoplasmic (n=108; red). Black contours represent all lncRNA expressed in HDF. e, Example of *ZNF213-AS1* loci showing transcript model, CAGE and RNA-seq signal along with targeting ASOs. f, Schematics of the study.

Knockdown efficiencies had a median value of 34.6%, and for 27% of ASOs, the efficiencies were higher than 50% (Supplementary Table S2). Overall, ASOs targeting exons or introns were equally effective, and knockdown efficiencies were independent of the genomic class, expression level, and subcellular localization of the lncRNA (Supplementary Fig. 1).

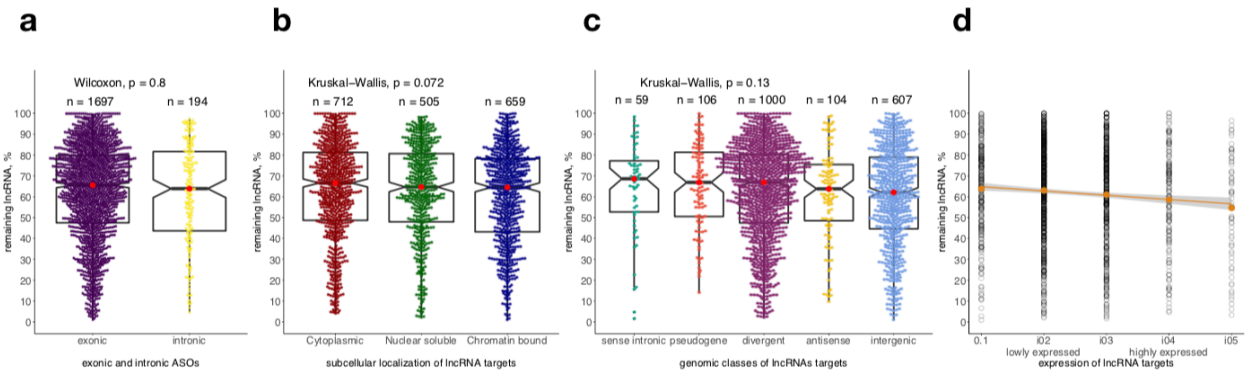

Fig.S1: Efficacy of ASOs. Comparison of knockdown efficiencies: a, for exonic and intronic ASOs. b, across different subcellular localizations of targeted lncRNAs. c, across different genomic classes of targeted lncRNAs. d, across expression levels of lncRNAs. In all panels, highest knockdown efficiency from the three primer pairs for a given ASO is shown. Red dot indicates the median value. Knockdown efficiencies with negative values were not considered.

A subset of lncRNAs is associated with cell growth and morphology changes

To evaluate the effect of each lncRNA knockdown on cell growth and morphology, we imaged ASO-transfected HDF in duplicates every 3 hours for a total of 48 hrs (Supplementary Table S3) and estimated their growth rate based on cell confluence measurements (Fig. 2a,b). 482 out of 2,021 (~23.8%) ASOs showed significant reduction and 20 (~1%) ASOs showed significant increase in the growth rate (Supplementary Table S3; *Student's* two-sided t-test FDR < 0.05) as compared to non-targeting controls (Fig. 2c). Different ASOs targeting the same lncRNA typically had different effects on growth, possibly due to variable knockdown efficiencies, differences in targeted lncRNA isoforms, as well as off-target effects. Requiring at least two independent ASOs to consistently impact cell growth, we identified 10/285 (~3.6%) lncRNAs (Supplementary Table S3; Z-score < -1.645, FDR < 0.05) involved in maintaining cell growth in HDF. Among lncRNAs with at least 25% reduced growth rate, we selected *A1BG-AS1*, which was previously implicated in cell growth (Bai et al. 2018), *CATG00000089639*, *RP11-195F19.9*, and *ZNF213-AS1* (Fig. 2d),

and validated their growth inhibition by measuring the Ki-67 proliferation protein marker upon knockdown with siRNAs and selected ASOs (Fig. 2e).

In addition to cell growth, we also explored changes in cell morphologies (Fig. 2f). Using a machine learning-assisted workflow (Methods; Fig. 2g), each cell was segmented and its morphological features representing various aspects of cell shapes and sizes were quantified (Carpenter et al. 2006) (Supplementary Table S3). As an example, knockdown of 12/285 lncRNAs affected the spindle-like morphology of fibroblasts, as indicated by consistent changes in their observed eccentricity (Z-score < -1.645, FDR < 0.05, at least two ASOs) without reducing the cell number, suggesting possible cellular transformation towards epithelial-like states. Collectively, we observed 44/285 lncRNAs (~15.5%) affecting cell growth and/or morphological parameters (Supplementary Table S3; Fig. 2h).

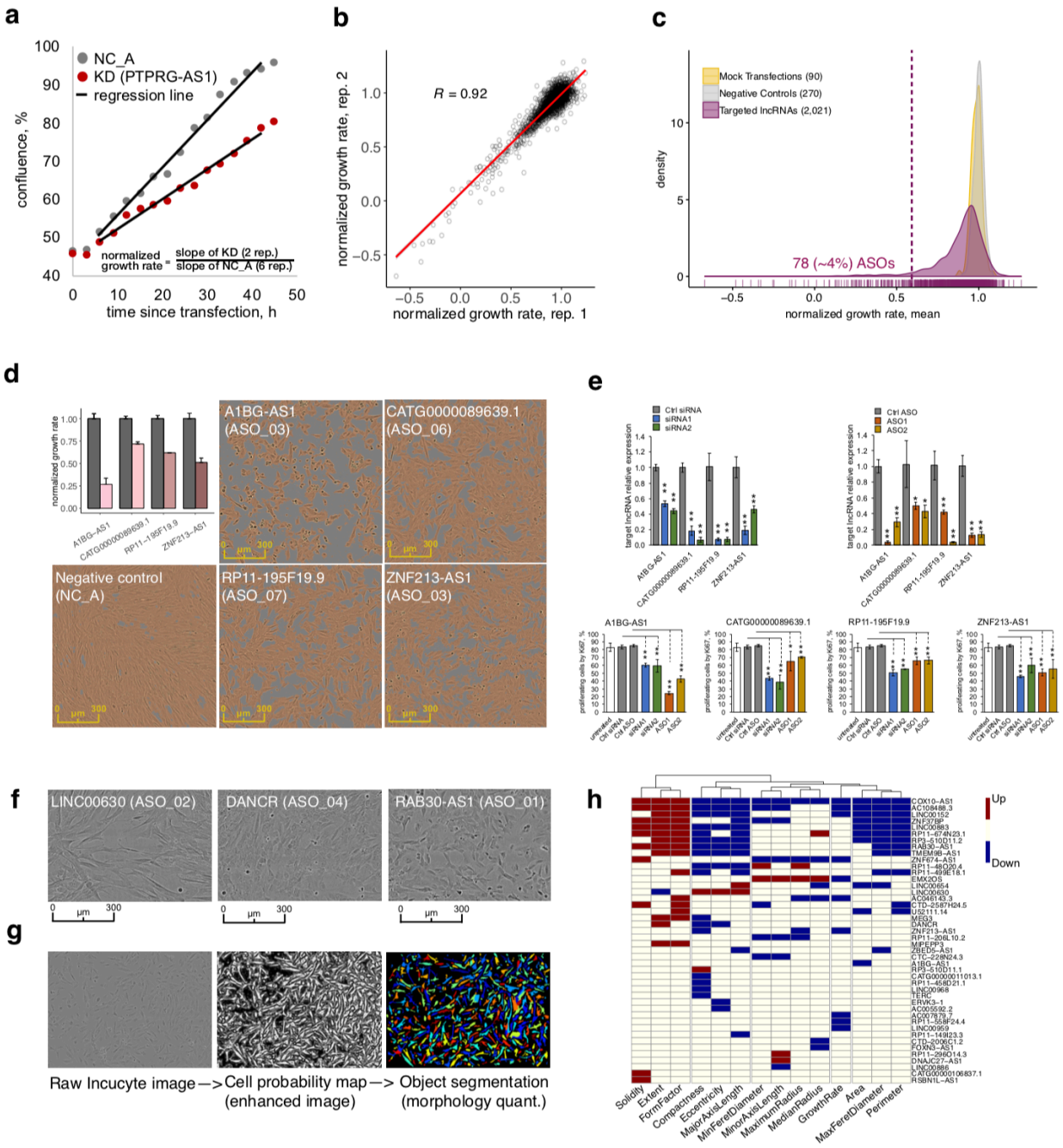

Fig. 2: Cell growth phenotype assessment. a, Selected example (PTPRG1-AS1) showing the normalized growth rate estimation using NC_A (negative control). b, Correlation of the normalized growth rate for technical duplicates across 2,456 IncuCyte® samples. c, Density distribution of normalized growth rates (technical replicates were averaged) for 2,021 ASO samples targeting lncRNA (red), 270 negative control (NC_A) samples (grey) and 90 mock-transfected cells (lipofectamine only) samples (yellow). Dashed vertical line indicates the 78 (~4%) ASOs with significantly reduced normalized growth rate as compared to other targeting ASOs (ZScore<-1.645; fdr<0.05). d, Growth inhibition for selected lncRNA targets (grey bars indicate matched negative control samples) and the corresponding Incucyte images at 48hrs after transfection. e, Ki67 staining (growth inhibition validation) for selected lncRNA targets after siRNA and ASOs suppression. f, IncuCyte® cell images of a few selected examples of distinct cell

morphologies upon the lncRNA knockdown, g, Overview of cell morphology image processing, h, lncRNAs (n=44) significantly (ZScore<-1.645; fdr<0.05) and consistently (at least two ASOs) reducing growth rate (n=10) or/and affecting cell morphologies (n=34).

Molecular phenotyping by CAGE recapitulates cellular phenotypes and highlights functions of lncRNAs

To evaluate transcriptome response upon the knockdowns and its agreement with the observed cellular phenotypes, we next selected 340 ASOs with high knockdown efficiency (mostly greater than 50%; median 71.4%) and generated and deep-sequenced 970 CAGE libraries to analyze 154 lncRNAs (Fig. 3a; Supplementary Table S4). Differential gene expression analysis revealed a median of 15 differentially regulated genes, where 61 (~18%) ASOs yielded zero and 108 (~32%) ASOs yielded 100 or more differentially regulated genes (Supplementary Table S5, Fig. 3b). lncRNAs can recruit a regulatory complex to mediate gene regulation in a site-specific manner (Long et al. 2017)). To test and identify focal regions of lncRNA-mediated regulation, we profiled differentially regulated CAGE promoters in genomic-clusters. We applied the Clustering of genomic REgions Analysis Method (CREAM) (Methods; Supplementary Fig2a; Tonekaboni et al. 2018) and identified linear clusters were significantly more compact than randomly generated clusters (Supplementary Fig. 2b; Mann-Whitney $p < 2.2 \times 10^{-16}$) thus supporting biological relevance of the transcriptome response detected by CAGE molecular phenotyping (Shin et al. 2018). Most identified clusters were distal to the genomic locus of the targeted lncRNA, suggesting a widespread usage of *trans* regulation mechanisms by lncRNAs (Supplementary Fig. 2c).

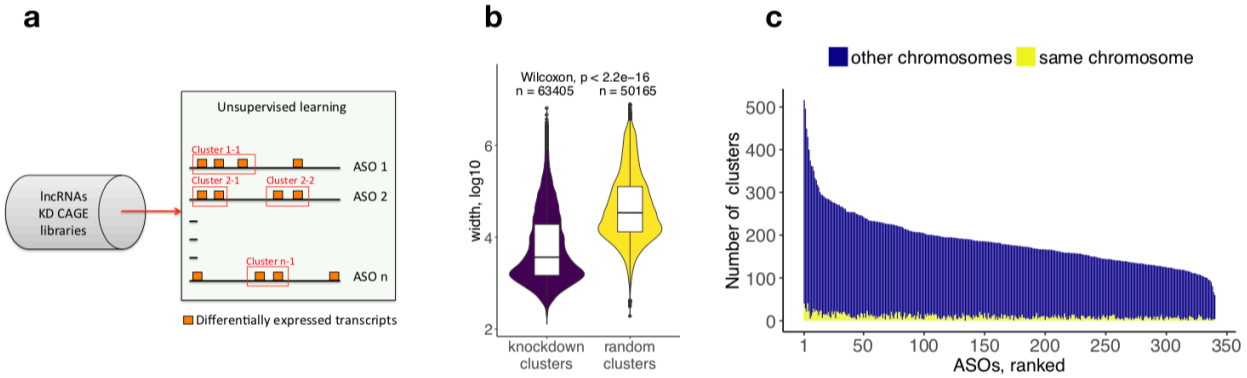

Fig. S2: Regulation of gene clusters by lncRNA knockdowns. a, Schematics of finding CORE clusters: differential transcripts are clustered with CREAM algorithm based on their genomic position. b, Distribution of the CORE clusters width in the knockdown experiment is narrower than that of clusters generated from randomly generated data. c,

Number of CORE clusters and their relative genomic position to the lncRNA target for each ASO knockdown, suggesting that lncRNA targets generally affect genes on the same as well as on other chromosomes.

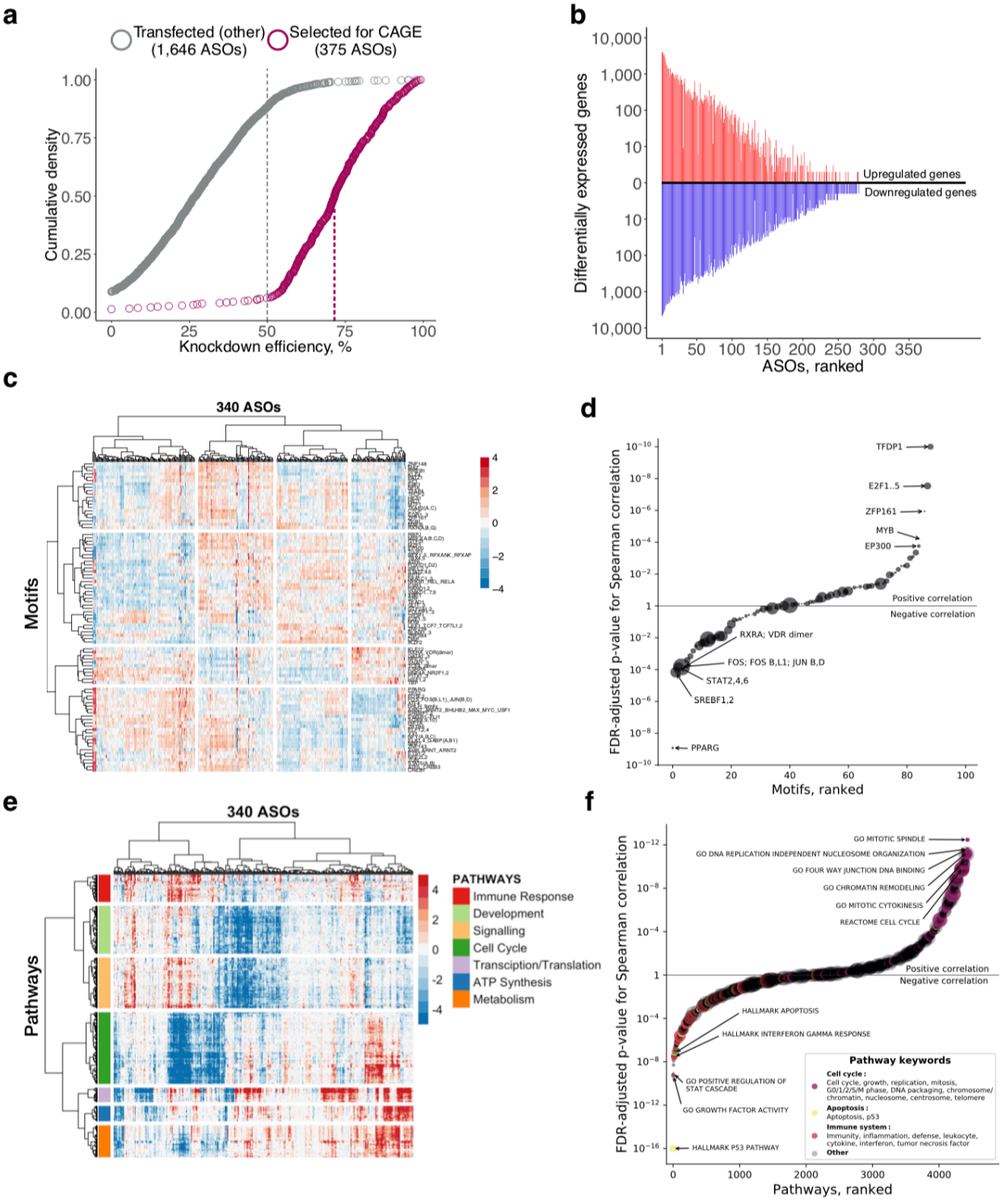

Fig. 3: CAGE molecular phenotype predicts cellular phenotype. a, q-PCR knockdown efficiency for 2,021 ASO-transfected samples (targeted lncRNAs). Grey dashed line indicates 50% knockdown efficiency generally required for CAGE selection. Purple dashed line indicates median knockdown efficiency (71.5%) of ASOs selected for CAGE

sequencing. b, Distribution of significantly differentially expressed genes (up-regulated: fdr<0.05, ZScore>1.645, log2FC>0.5 and down- regulated: fdr<0.05, ZScore<-1.645, log2FC<-0.5) across all 340 ASOs. c, Motif Response Activity Analysis (MARA) across 340 ASOs. Scale indicates ZScore of the relative motif activity, with abs(ZScore)>4=4. d, Correlation between normalized growth rate and motif activities across 340 ASOs targeting lncRNAs with highlighted examples Motifs are scaled based on associated TFs expression in HDF (1 to ~600TPM). e, Enriched biological pathways across 340 ASOs. Scale indicates GSEA enrichment value calculated as -log10(pvaule)*sign(NES). f, same as in d, but for selected GSEA pathways. Pathways are scaled based on the number of associated genes.

Next, to assess the specific molecular pathways affected by individual ASOs, we performed: i) transcription factor binding site motif activity response analysis (MARA; FANTOM Consortium et al. 2009) Fig. 3c, ii) gene set enrichment analysis (GSEA; Subramanian et al. 2005; Fig. 3e), and iii) analysis of FANTOM5 co-expression modules (FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al. 2014). Transcription factor motifs that promote cell growth (*e.g.* TFDP1, E2F1,2,3, and EP300) were positively correlated with the measured cell growth rate while transcription factor motifs known to inhibit growth or induce apoptosis (*e.g.* PPARG, SREBPF, and STAT2,4,6) were negatively correlated (Fig. 3d; Supplementary Fig. 3a; Supplementary Table S6). Moreover, correlations between GSEA pathways (Fig. 3f; Supplementary Fig. 3b, Supplementary Table S6) and FANTOM5 co-expression clusters (FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al. 2014); Supplementary Fig. 3c) showed that cell growth and replication related pathways were positively correlated with the measured growth rate, whereas those related to immunity, cell stress and cell death were negatively correlated. Additionally, morphological changes were reflected in the molecular phenotype assessed by CAGE (Supplementary Fig. 3d). Cell radius and axis length were associated with GSEA categories related to actin arrangement and cilia, while cell compactness was negatively correlated with apoptosis. The extensive molecular phenotyping analysis also revealed pathways not explicitly associated with cell growth and cell morphology, such as transcription, translation, metabolism, development and signaling (Fig. 3e).

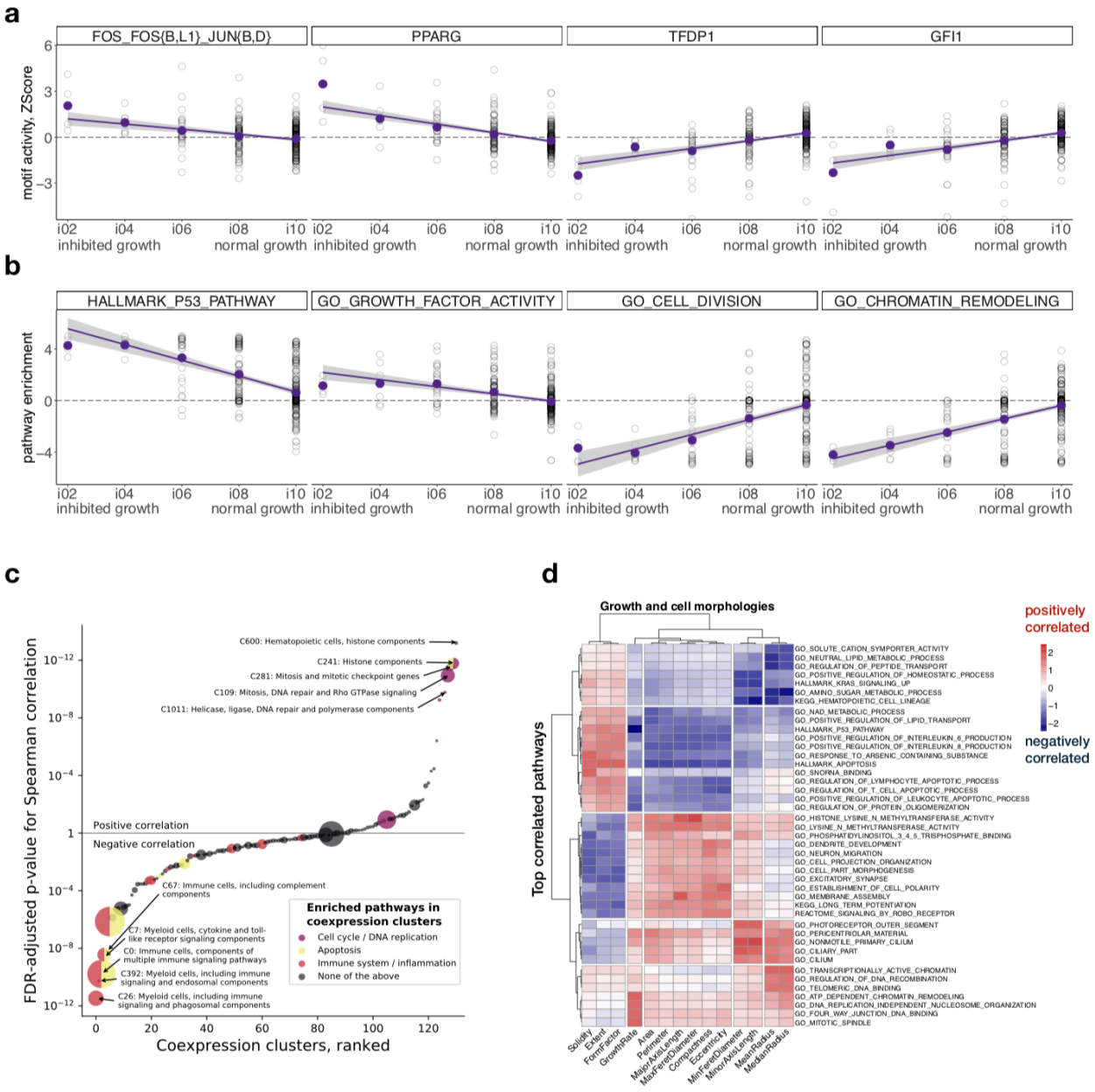

Fig. S3: Correlations with cell growth and morphologies. A, Correlations of selected transcription factor binding motifs activity with the normalized growth rate across 340 ASOs targeting lncRNAs. Each hollow circle represents a single ASO. b, Same as in a, but for selected GSEA pathways. c, Global correlations of FANTOM 5 co-expression clusters enrichment with growth phenotype. d, Top three significant GSEA pathways positively and negatively correlated with the normalized growth rate and 13 cell morphologies. Scale indicates FDR-adjusted p value for Spearman correlation re-scaled through each morphology.

Analyzing 119 lncRNAs targeted by at least two independent ASOs revealed that the transcriptome response was more concordant for distinct ASOs targeting the same lncRNA than for ASOs targeting different lncRNAs (Supplementary Fig. 4a). We found 20, 18, and 36 lncRNAs

significantly and consistently changed gene expression, motif activity and biological pathways, respectively (Methods; Supplementary Table S5; Supplementary Fig. 4b). In total, the molecular phenotype was affected upon knockdown of 52/119 (43.7%) lncRNAs in human dermal fibroblasts (Supplementary Fig. 4c).

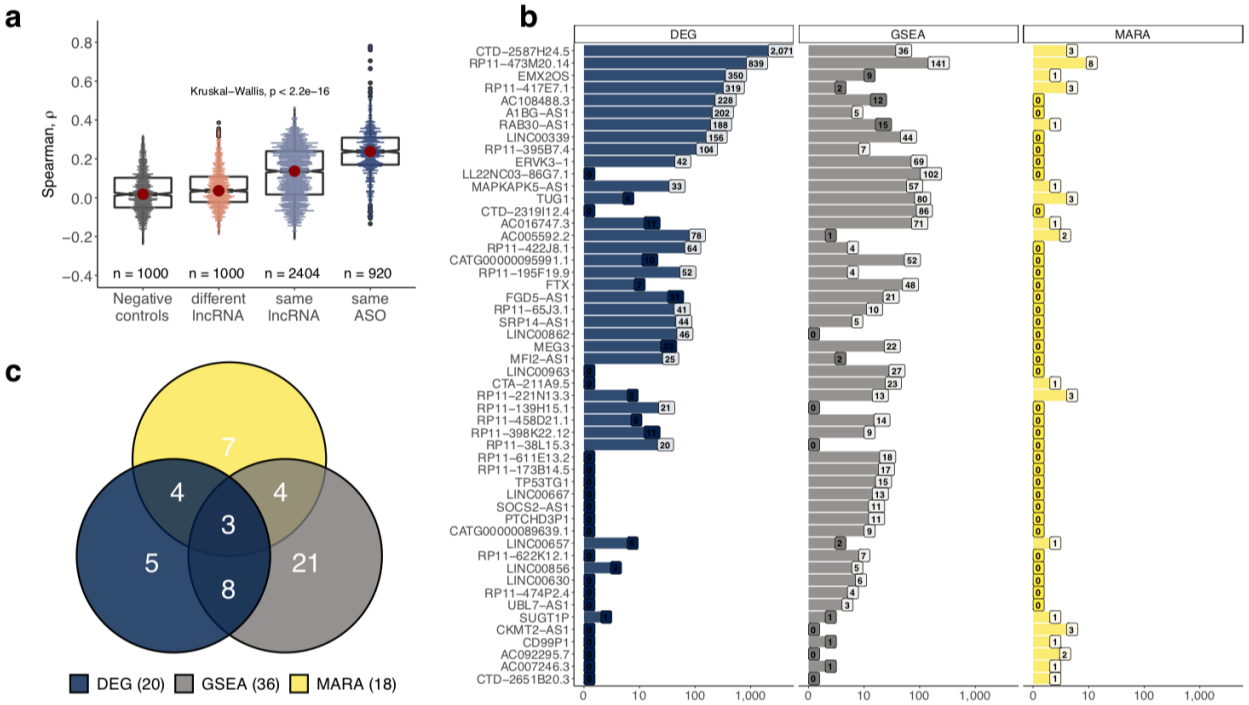

Fig. S4: Concordant functional response of lncRNAs. a, Pairwise correlations for CAGE libraries across: 1,000 randomly selected different lncRNAs comparisons, same lncRNAs (different ASOs), same ASOs (duplicates) and 1,000 randomly selected pairwise negative control comparisons. For each data point, the correlations are calculated taking all genes > 1TPM and dividing their expression by their average expression in all 143 negative controls. b, 52 lncRNA targets that in at least one functional category (DEG, GSEA, MARA) show a significant consistent response (in at least 2 ASOs) and above generated matched random background (Methods). The significant lncRNA targets in each category (light bar tip) are summarized in c.

Promoter usage and architecture modulated by lncRNA knockdown

Precise mapping of transcription start sites (TSS) with CAGE previously revealed usage of alternative promoters and architecture diversities mediated by either TATA or CpG transcriptional machineries (Carninci et al. 2006; Kim et al. 2005; Haberle et al. 2014; Garieri et al. 2017). Promoter switching (changes in the *relative* expression between a given pair of promoters of the same gene) suggest that lncRNAs might direct RNA polymerase II to alternative initiation sites (Garieri et al. 2017) resulting in different gene isoforms. To test whether we can define regulatory roles of lncRNAs based on the relative TSS activities in multi-promoter genes, we measured

knockdown-mediated promoter switching events, which occasionally lead to the change of the dominant promoter (Fig. 4a-c). For 28 out of 119 lncRNAs, we found 2,699 consistent switching events (Supplementary Table S7) in 17.8% (1,276 out of 7,161) actively transcribed multiple-promoter genes (Fig. 4d). The highest number of these events (n=1,159) was found for *CTD-2587H24.5* (ENSG00000267577) reported to be a "transcribed enhancer" and a hub in the chromatin interaction map in K562 cells (Thiel et al. 2018). Knockdown of five other lncRNAs: *AC013394.2, RP11-417E7.1, EMX2OS, ZNF213-AS1*, *RAB30-AS1* resulted in 100 or more promoter switching events (Fig. 4d).

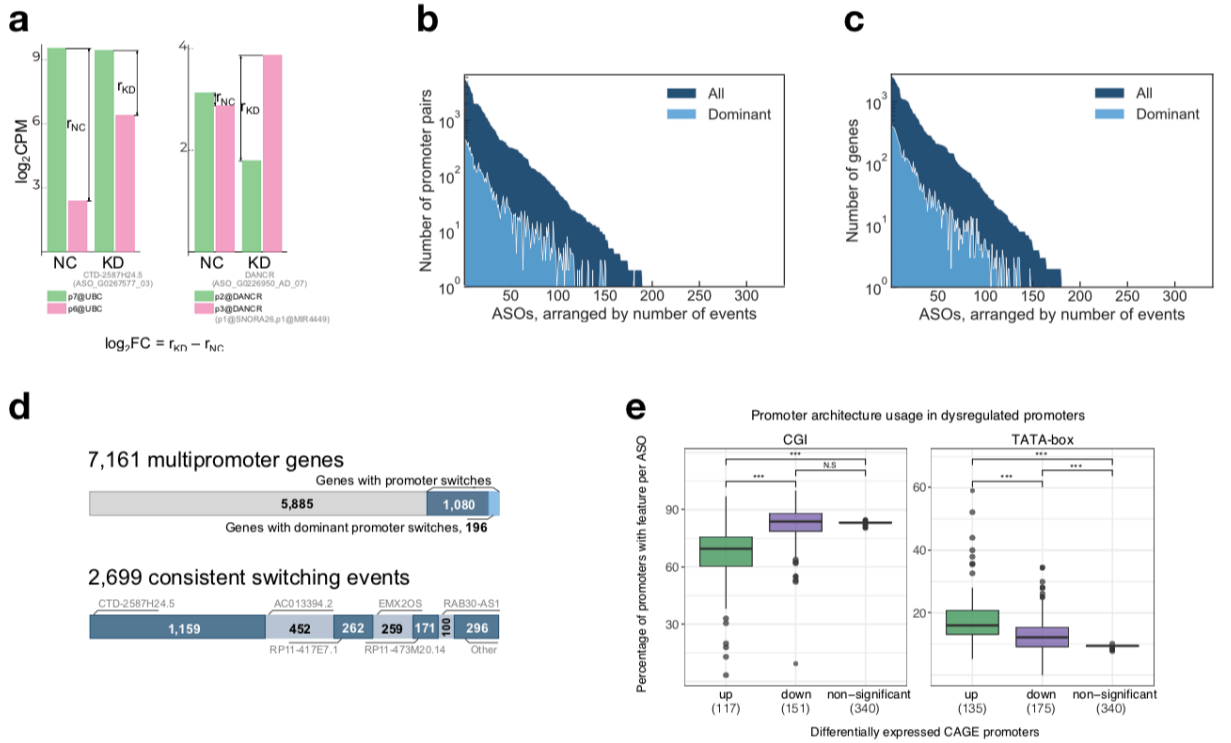

Fig. 4: Knockdown-mediated promoter usage and architecture. a, Selected example of a promoter switching event in the *UBC* gene upon the knockdown of *CTD-25877H24.5* lncRNA (left) and of the dominant promoter switching in *DANCR* upon its own knockdown (right). b, Number of significant promoter pair switching events (abs(log2FC)>0.5; FDR < 0.05; abs(ZScore) > 1.645) across all 340 ASOs, with switchings of the dominant promoter highlighted (light blue). c, 28 out of 119 lncRNA show 2,699 significant and consistent (in at least two ASOs) promoter switching events affecting 1,276 out of 7,161 expressed multipromoter genes. 196 of these resulted in switching of the dominant promoter (light blue). d, 2,699 consistent and significant switching events with most prominent lncRNA targets (out of 28 lncRNA targets showing consistent switching) highlighted. e, Up- and down-regulated CAGE promoters with CGI and TATA-box architectures compared to the background (all non-significantly deregulated promoters). Each data point represents a single ASO experiment with at least 20 promoters differentially expressed (abs(log2FC) > 0.5; FDR < 0.05; ZScore > 1.645) in a given category.

Promoter architecture is indicative of gene regulatory machineries with sharp transcription initiation peaks associated with TATA-box promoters, while broad distributions of transcription initiation are found in CpG island (CGI) regions. Also, TATA-boxes are often found in promoters of tissue-specific genes and they are absent in most housekeeping genes (Sandelin et al. 2007). Thus, the enrichment of TATA-box promoter usage upon knockdown suggest that lncRNAs may modulate regulatory machineries that govern cell- and pathway-specific promoters and elicit cell-specific function (Haberle et al. 2014). To assess which lncRNAs influence function in a cell-specific manner, we quantified the changes in TATA-box and CGI architecture at differentially regulated promoters and assessed significance for each ASO knockdown (Methods; Supplementary Table S8). Differentially expressed promoters were more likely to have a TATA-box both in case of upregulation (two sample Student t-test p-value < 2.2×10^{-16}) and downregulation (two sample Student t-test p-value = 1.3×10^{-16}; Fig. 4e). For example, 20-35% of promoters upregulated upon knockdown of *LINC01615*, *RNASEH1-AS1, RP11-221N13.3* contained a TATA-box, compared to the 10% of unchanged promoters (Supplementary Fig. 5a,b), and for *RP11-221N13.3,* additionally, 23-25% of the down-regulated promoters had a TATA-box architecture (Supplementary Fig. 5b). In contrast, CGI architecture was significantly underrepresented in upregulated promoters only (stratified sampling p value < 2.2×10^{-16}; Fig. 4e). Knockdown of *FGD5-AS1,* for example, showed ~60-70% depletion in two out of three ASOs as compared to the ~80% of CGI in the background promoters (Supplementary Fig. 5c).

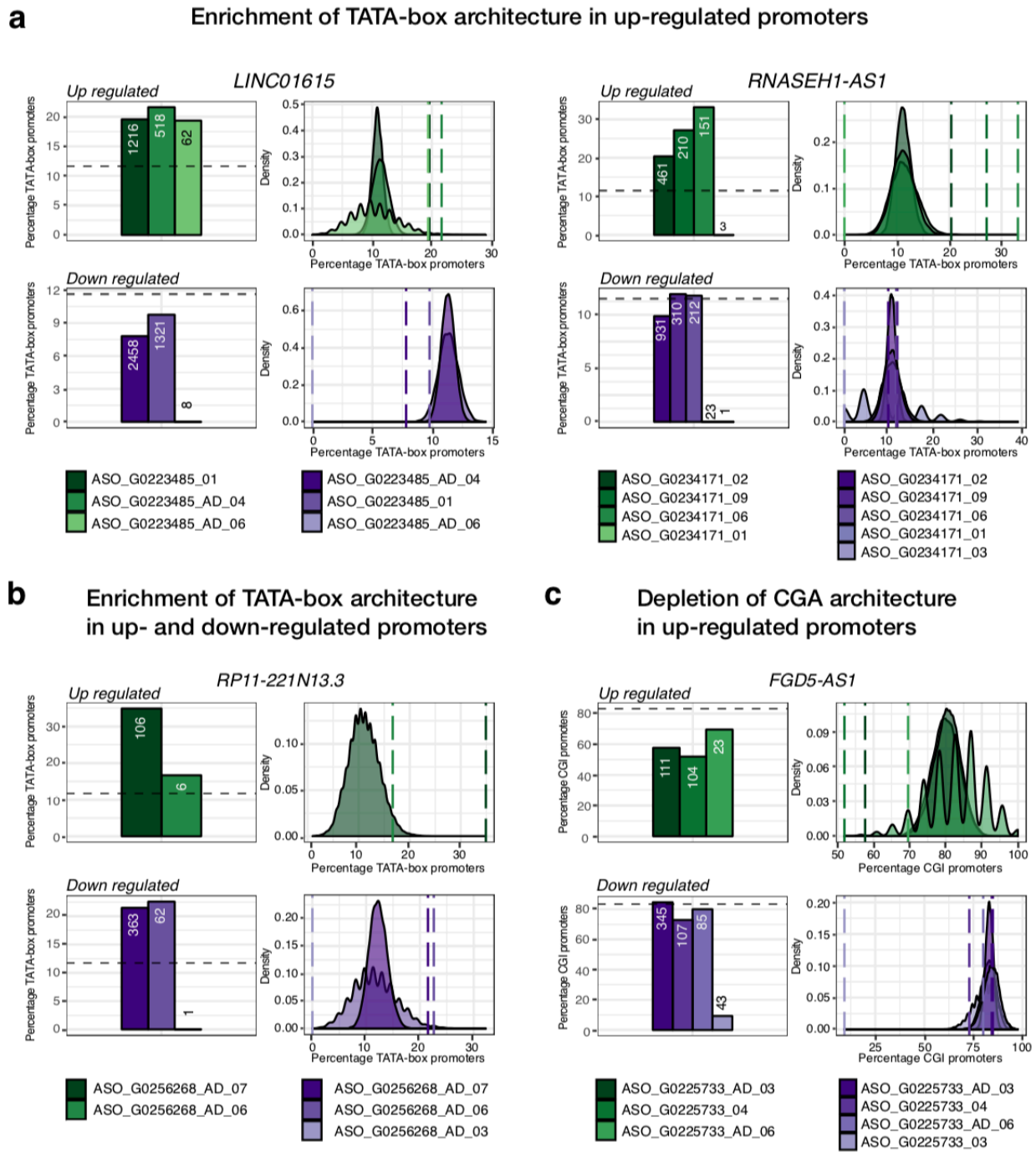

Fig. S5: Promoter architecture usage. a, Selected lncRNAs with significantly up regulated TATA-box promoters. b, Significantly up and down regulated TATA-box promoters. c, Significantly upregulated CGA promoters. In all panels, barplots (left) show the percentage of TATA-box or CGI promoters in a given ASO and the numbers represent the number of differentially expressed promoters. The density plots (right) show the distribution of the stratified random sampling (10,000 times) for the same number of promoters. The dotted lines indicate observed percentage for each ASO. The smaller the number of genes randomly sampled, the more discontinuous density distribution.

ZNF213-AS1 is associated with cell growth and migration.

As an example of an lncRNA associated with cell growth and morphology (Fig. 2h) we further analyzed *ZNF213-AS1* (*RP11-473M20.14*). This lncRNA is highly conserved in placental mammals, moderately expressed (~8 CAGE tags per million) in HDF and enriched in the chromatin. Four distinct ASOs (ASO_01, ASO_02, ASO_05, and ASO_06) strongly suppressed expression of *ZNF213-AS1*, while expression of the *ZNF213* sense gene was not significantly affected in any of the knockdowns. The four ASOs caused varying degrees of cell growth inhibition (Fig. 5a). ASO_01 and ASO_06 showed a reduction in cell number, as well as an upregulation of apoptosis, immune and defense pathways in GSEA suggesting cell death. While cell growth inhibition observed for ASO_02 and ASO_05 was confirmed by Ki-67 marker staining (Fig. 2e), the molecular phenotype revealed suppression of GSEA pathways related to cell growth, as well as to cell proliferation, motility, and extracellular structure organization (Fig. 5b), and consistent in two ASOs downregulation of related motifs, for example, EGR1, EP300, SMAD1..7,9 (Phan et al. 2004) (Fig. 5c). For data exploration, http://fantom.gsc.riken.jp/zenbu/reports/

As cell motility pathways were affected by the knockdown, we tested whether *ZNF213-AS1* could influence cell migration. Using wound closure (scratch) assays with transient cell growth inhibition (pre-incubation with mitomycin-C and serum starvation) on the transfected cells, we observed a substantial reduction of wound closure rate (~40% over a 24-hour period) in the *ZNF213-AS1* depleted HDFs (Fig. 5d). The reduced wound healing rate should thus mainly reflect reduced cell motility, further confirming affected motility pathways predicted by the molecular phenotype.

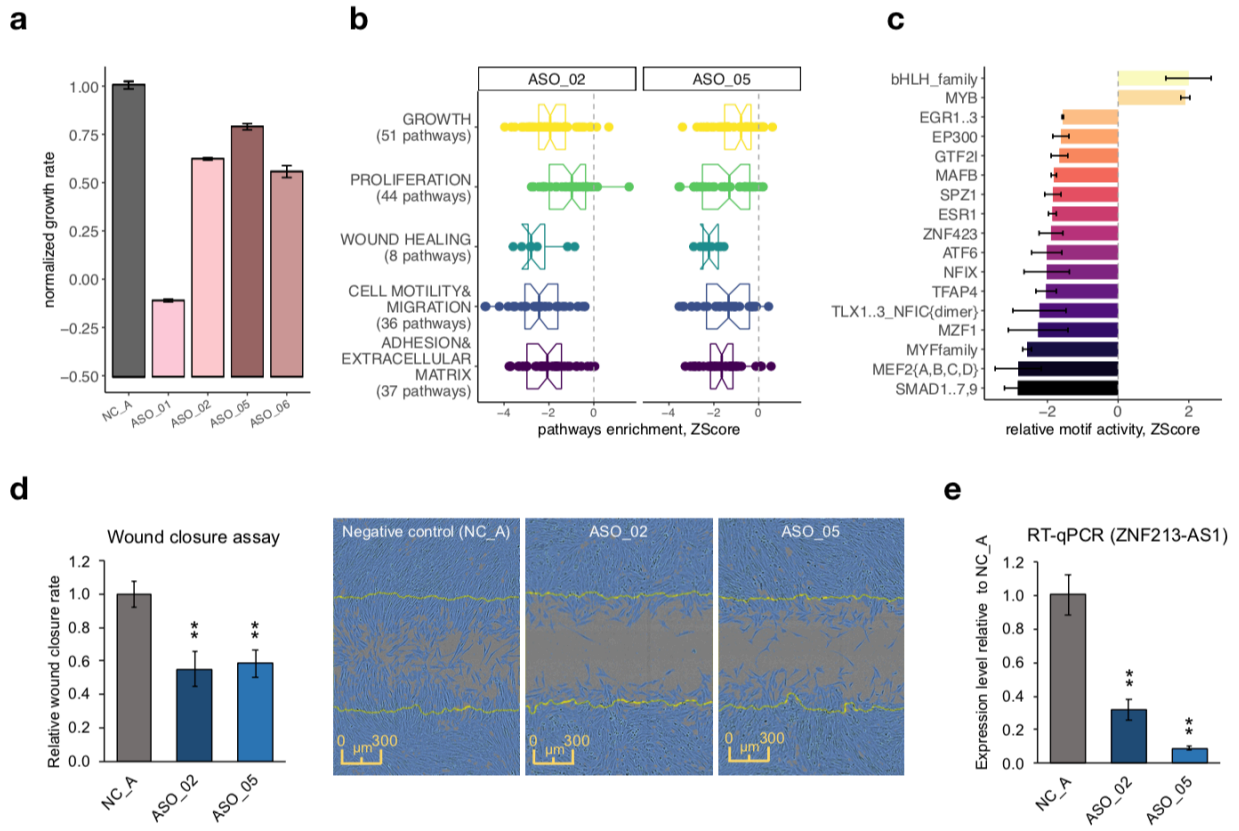

Fig. 5: *ZNF213-AS1* regulates cell growth, migration and proliferation. a, Normalized growth rate across four distinct ASOs (in duplicates) targeting ZNF213-AS1 as compared to six negative control samples (shown in grey). b, Enrichment of biological pathways associated with growth, proliferation, wound healing, migration and adhesion for ASO_02 and ASO_05. C, Down- and up-regulated binding motifs including examples of those for transcription factors known to modulate growth, migration and proliferation (EGR family, EP300, GTF2I). d, Relative wound closure rate calculated during the 20 hrs post scratching and the representative images of wound closure assay at 20hrs shows 40-45% reduction for the two targeting ASOs (ASO_02 and ASO_05; both in duplicates) as compared to six negative controls samples (shown in grey). e, Knockdown efficiency measured by RT-qPCR after wound closure assay (72 hrs post transfection) showing strong sustained suppression (65-90%) of ZNF213-AS1.

As these results indicated a potential role of *ZNF213-AS1* in cell growth and migration, we used FANTOM CAT Recount 2 atlas (co-submitted to *Genome Research*, GENOME/2019/254656, Imada et al. 2018), which incorporates the TCGA dataset (Collado-Torres et al. 2017), and found relatively higher expression of *ZNF213-AS1* in acute myeloid leukemia (LAML) and in low grade gliomas (LGG) as compared to other cancers (Supplementary Fig. 6a). In LAML, the highest expression levels were associated with mostly undifferentiated states, whereas in LGG, elevated expression levels were found in oligodendrogliomas, astrocytomas, and in IDH1 mutated tumors, suggesting that *ZNF213-AS1* is involved in modulating differentiation and proliferation of tumors

(Supplementary Fig. 6b-e). Further, univariate Cox proportional hazard analysis as well as Kaplan-Meier curves for LGG were significant and consistent with our findings (HR = 0.61, BH FDR = 0.0079). The same survival analysis on LAML showed weak association with poor prognostic outcome but the results were not significant (Supplementary Fig. 6f,g).

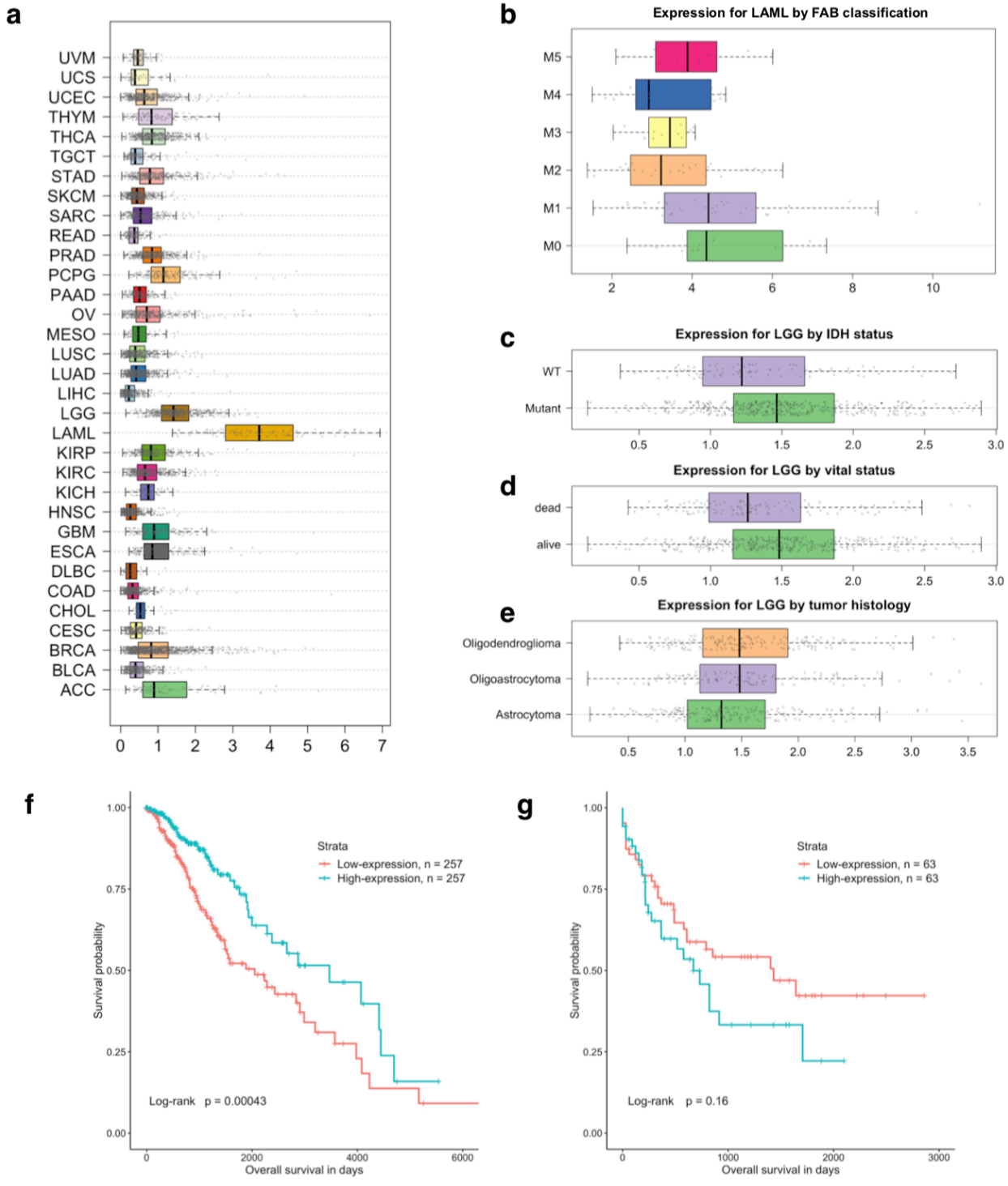

Fig. S6: Involvement of *ZNF213-AS1* in cancer. a, Expression of ZNF213-AS1 in TCGA b, Expression of ZNF213-AS1 for Acute Myeloid Leukemia (LAML) by French-American-British classification. The levels are higher in M0 and M1 subtypes. These subtypes are considered the most undifferentiated stages of LAML. Since ZNF213-AS1 shows the lowest expression level in blood cells (GTEx), the higher level of expression could be suggestive of dysregulation of its expression in undifferentiated stages or involvement in early precursor of blood cells. c, Expression of *ZNF213-AS1* in Low Grade Glioma (LGG) by tumor histology classification. Oligodendrocyte differentiation is related to good prognosis is this cancer type. d, Expression of *ZNF213-AS1* in LGG by tumor IDH (isocitrate dehydrogenase) mutational status. A total of 511 samples with mutational information were available, 417 of them were classified as "Mutant" and 94 as "wild type". IDH mutation has been repeatedly associated with better prognosis than wild type in gliomas. e, Expression of *ZNF213-AS1* by patients' vital status. The number of alive patients is 387 from the total of 513 available in the follow-up. Univariate Cox proportional hazard analysis as well as Kaplan-Meier curves for: f, LGG (HR: 0.61, FDR: 0.0079) g, LAML (HR: 1.32; FDR: 0.5455)

RP11-398K22.12 (KHDC3L-2) regulates *KCNQ5* in *cis*

Besides primarily observing in *trans* regulation in our knockdown data, several lncRNAs were previously shown to also regulate the expression in *cis* (Joung et al. 2017). Herein, we investigated in detail *RP11-398K22.12* (ENSG00000229852) where the knockdowns by two independent ASOs (ASO_03, ASO_05) successfully reduced the expression of the target lncRNA (67-82% knockdown efficiency, respectively) and further downregulated its neighboring genes, *KCNQ5* and its divergent partner novel lncRNA *CATG00000088862.1* (Fig. 6a). While the two genomic loci occupy chromosome 6 and are 650 kb away, Hi-C analysis showed that they are located within the same topologically associated domain (TAD) and spatially co-localized (Fig. 6b). Moreover, chromatin-enrichment and single molecule RNA-FISH of *RP11-398K22.12* (Fig. 6c) suggested its highly localized *cis*-regulatory role.

In FANTOM5 (Hon et al. 2017), expression of *RP11-398K22.12*, *KCNQ5* and *CATG00000088862.1* was enriched in brain and nervous system samples, while GTEx (GTEx Consortium 2015) showed highly specific expression in brain tissues, particularly in the cerebellum and the cerebellar hemisphere (Fig. 6d). GTEx data also showed that expression of *RP11-398K22.12* with *KCNQ5* and *CATG00000088862.1* was highly correlated across neuronal tissues (Fig. 6e,f), with the exception of cerebellum and cerebellar hemisphere, potentially due to relatively lower levels of *KCNQ5* and *CATG00000088862.1* while levels of *RP11-398K22.12* remained relatively higher. Additionally, we found an eQTL SNP (rs14526472) overlapping with *RP11-398K22.12* and regulating expression of *KCNQ5* in brain caudate ($p = 4.2 \times 10^{-6}$; normalized effect size -0.58). All these findings indicate that *RP11-398K22.12* is implicated in the nervous

system by maintaining the expression of *KCNQ5* and *CATG00000088862.1* in a *cis*-acting manner.

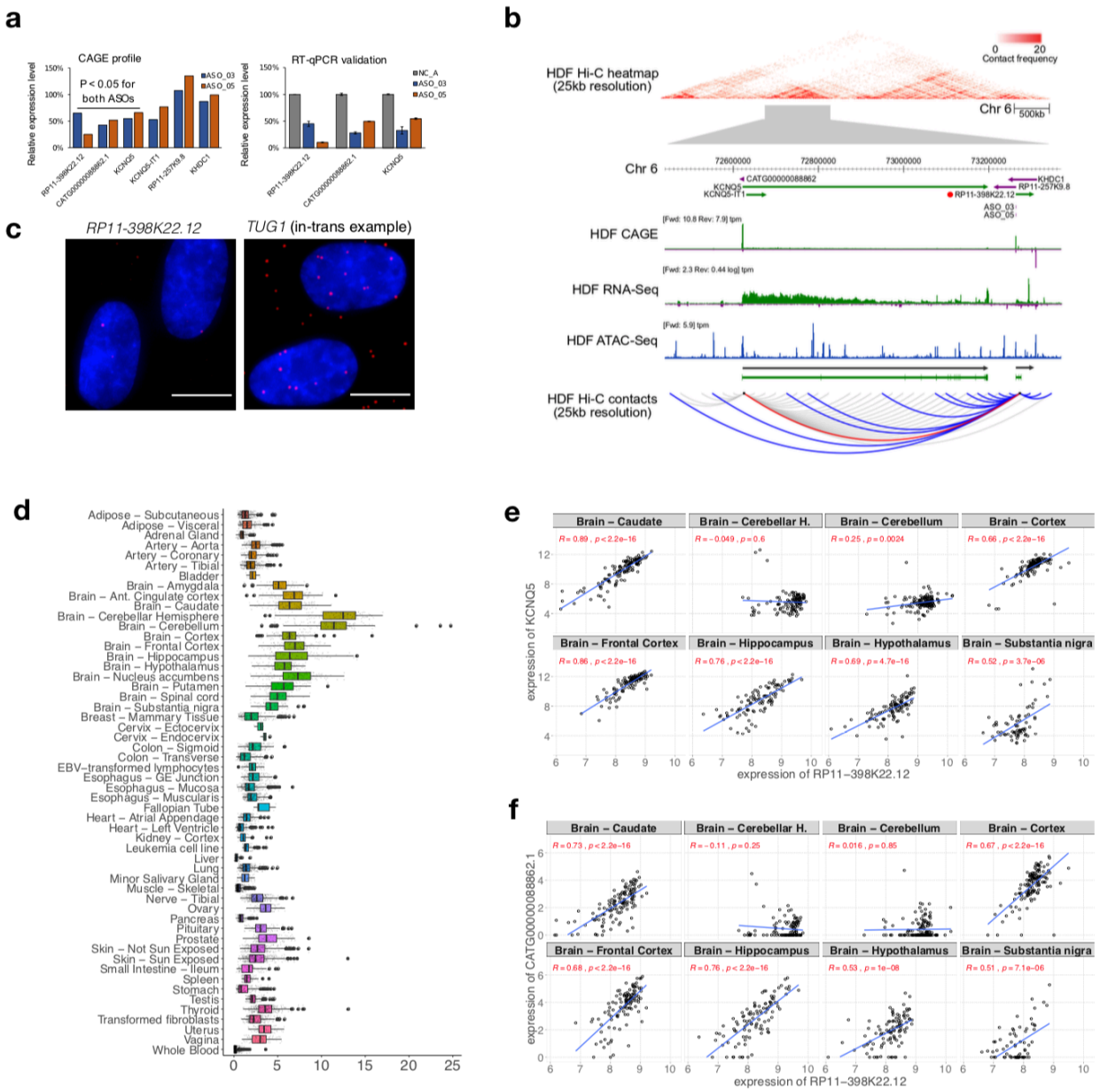

Fig. 6. RP11-398K22.12 mediates the down-regulation of KCNQ5 in cis. a, Changes in expression levels of genes shown by HiC analysis to be in the same topologically associated domain (TAD) as *RP11-398K22.12.* Both *KCNQ5* and *CATG00000088862.1* are down-regulated (pvalue < 0.05) upon the knockdown of *RP11-398K22.12* with two independent ASOs (left) as further confirmed with RT-qPCR (right). b, (top) Representation of the chromatin conformation in the 4Mb region proximal to the TAD containing *RP11-398K22.12,* followed by the locus gene annotation, CAGE, RNA-Seq and ATAC-Seq data for native HDFs. (bottom) Schematic diagram showing Hi-C predicted contacts of *RP11-398K22.12* (blue) and *KCNQ5* (grey) (25Kb resolution, frequency >= 5) in HDF cells. Red line indicates *RP11-398K22.12* and *KCNQ5* contact. c, FISH image for *RP11-398K22.12* suggesting proximal regulation. *TUG1* FISH image (suggesting trans regulation) is included for comparison; (bar = 10 μm). d, GTEx atlas

shows relatively high expression of *RP11-398K22.12* in brain samples. e, Correlation of *RP11-398K22.12* and *KCNQ5* in eight distinct brain regions. F, Correlation of *RP11-398K22.12* and *CATG00000088862.1* in eight distinct brain regions.

Discussion

This study systematically annotates lncRNAs through molecular and cellular phenotyping by selecting 285 lncRNAs from human dermal fibroblasts across a wide spectrum of expression, conservation levels and subcellular localization enrichments. Importantly, using ASO technology allowed observed phenotypes to be associated to the lncRNA transcripts, while in contrast CRISPR-based approaches may synchronically influence the transcription machinery at the site of the divergent promoter or affect regulatory elements of the targeted DNA site. Knockdown efficiencies obtained with ASOs were observed to be independent of lncRNA expression levels, subcellular localization, and of their genomic annotation, allowing us to apply the same knockdown technology to various classes of lncRNAs.

We investigated the *cis* regulation of nearby divergent promoters, which has been reported as one of the functional roles of lncRNA (Luo et al. 2016). However, in agreement with previous studies (Guttman et al. 2011) we did not observe general patterns in the expression response of divergent promoters. Recent studies suggest that transcription of lncRNA loci that do not overlap with other transcription unit may influence RNA polymerase II occupancy on neighboring promoters and gene bodies (Engreitz, Haines, et al. 2016, Cho et al. 2018). Thus, it is plausible that transcription of targeted lncRNA was maintained, despite suppression of mature or nascent transcripts using ASOs. This further suggests that the functional responses described in this study are due to interference of processed transcripts present either in the nucleus, the cytoplasm or both. While it is arguable that ASOs may interfere with general transcription by targeting the 5'-end of nascent transcripts and thus releasing RNA polymerase II followed by exonuclease-mediated decay and transcription termination (aka "torpedo model"; (Proudfoot 2016)), most of the ASOs were designed across the entire length of the transcript. Since we did not broadly observe dysregulation in nearby genes, interference of transcription or splicing activity is less likely to occur.

We observed a reduction in cell growth for ~3.6% of our target lncRNA genes, which is in-line with previous experiments using CRISPRi-pooled screening, which reported 5.9% (in iPS cells) of lncRNAs exhibiting a cell growth phenotype (Liu et al. 2017). While these rates are much lower

than for protein-coding genes (Sokolova et al. 2017), recurrent observations of cell growth (including cell death) phenotypes strongly suggest that a substantial fraction of lncRNAs play an essential role in cellular physiology and viability.

Several lncRNAs such as *MALAT1*, *NEAT1*, and *FIRRE* have been reported to orchestrate transcription, RNA processing, and gene expression (Kopp and Mendell 2018), but are not essential for mouse development or viability. These observations advocate for assays that can comprehensively profile the molecular changes inside perturbed cells. Therefore, in contrast to cell-based assays, functional elucidation via molecular phenotyping provides comprehensive information that cannot be captured by a single phenotypic assay. Herein, differential gene expression analysis, GSEA (affected pathways) and MARA (changes in transcription factors binding activity) indicated that 43.7% of lncRNAs exert a regulatory function in HDF. However, lncRNA targets that did not exhibit a molecular phenotype may be biologically relevant in other cell types or cell states (Li and Chang 2014; Liu et al. 2017). At the same time, our results showed that particular lncRNAs expressed broadly in other tissues (e.g., in the human brain) were functional in HDF (in case of *RP11-398K22.12*), suggesting that lncRNAs may be functionally relevant across multiple tissues in spite of the cell-type-specific expression of lncRNAs.

An increasing number of studies suggests that lncRNAs regulate gene expression in *trans* via interaction with chromatin-modifying complexes such as PRC2 (Rinn et al. 2007) or independent of PRC2 (Portoso et al. 2017) by directly binding to DNA (triplex; (Mondal et al. 2015) or other RNA binding proteins (Tichon et al. 2016). Analysis of cellular localization by fractionation followed by RNA-seq and *in situ* hybridization can indicate whether an lncRNA may act *in trans* by quantifying its abundance in the nuclear soluble fraction as compared to cytoplasm. While most lncRNAs in nuclear soluble fraction affected pathways associated with chromatin modification, additional experiments to globally understand their interaction partners will elucidate the molecular mechanism behind *trans*-acting lncRNAs (Li et al. 2017; Sridhar et al. 2017; companion paper).

In summary, our study demonstrates that lncRNAs elicit gene regulation largely *in trans* by regulating the transcriptional machinery for gene activation and repression. These results highlight the functional importance of lncRNAs regardless of their expression, localization and conservation levels. Molecular phenotyping is a powerful platform to reveal the functional relevance of lncRNAs that cannot be observed based on the cellular phenotypes alone. With

488 additional molecular profiling techniques, such as RNA duplex maps in living cells to decode
489 common structural motifs (Lu et al. 2016) and Oxford Nanopore Technology (ONT) to annotate
490 the full-length variant isoforms of lncRNA (Hardwick et al. 2019), structure-to-functional
491 relationship of lncRNAs may be elucidated further in the future. Data and analysis for this study
492 is made available at http://fantom.gsc.riken.jp/zenbu/reports/#FANTOM6

Online Methods

Gene Models and lncRNA targets selections

The gene models used in this study were primarily based on the FANTOM CAGE-associated transcriptome (CAT) at permissive level as defined previously (Hon et al. 2017), with additional *de novo* transcript models constructed from human dermal fibroblasts (HDFs) and induced pluripotent stem cells (iPSCs) RNA sequencing data. In brief, CAGE sequencing was performed on the total RNA, and RNA-Seq was performed on ribosomal-RNA depleted RNA, from HDFs and iPSCs (as described elsewhere). CAGE and RNA-Seq reads were mapped onto hg19 using Tophat2 with default parameters. RNA-Seq reads were *de novo* assembled for each cell line using Cufflinks as described previously (Hon et al. 2017) and the transcript models with their 5'ends supported by CAGE reads were retained. LncRNA genes were identified from these retained transcript models as previously described (Hon et al. 2017). The novel lncRNA genes (i.e. loci non-overlapping with FANTOM CAT) were merged with the permissive FANTOM CAT, and the merged assembly were lifted over (Hinrichs et al. 2006) from hg19 to hg38.

From this merged assembly, we selected lncRNA knockdown targets in an unbiased manner to broadly cover various types of lncRNAs. Briefly, we first identified a list of the lncRNA genes expressed in HDF, with RNA-Seq expression at least 0.5 fragments per kilobase per million and CAGE expression at least 1 tag per millions. Then we manually inspected each lncRNA locus in ZENBU genome browser for 1) its independence from neighboring genes on the same strand (if any), 2) support from RNA-Seq (for exons and splicing junctions) and CAGE data (for TSS) of its transcript models and 3) support from histone marks at TSS for transcription initiation (H3K27ac) and along gene body for elongation (H3K36me3), from Roadmap Epigenomics Consortium (Roadmap Epigenomics Consortium et al. 2015). A representative transcript model, which best represents the RNA-Seq signal, was manually chosen from each loci for design of antisense oligonucleotides (ASOs). In total, 285 lncRNA loci were chosen for ASO suppression. Additional controls (NEAT1, protein coding genes Supplementary Table S1) were added including MALAT1 as an experimental control.

ASO design

ASOs were designed as RNase H-recruiting locked nucleic acid (LNA) phosphorothioate gapmers with a central DNA gap flanked by 2-4 LNA nucleotides at the 5' and 3' ends of the ASOs. For each lncRNA target, we used the unspliced transcript sequence from FANTOM CAT as template for designing a minimum of 5 ASOs per lncRNA. A total of 2,055 ASOs targeting 285 lncRNAs were designed for the study (Supplementary Table S1). The gapmer ASOs were between 16 and

19 nucleotides in length (median of 17 nucleotides), had no predicted perfect match off-targets and 0-1 one-mismatched off-targets, and were designed with a predicted melting temperature (T_m) in the range of 50-56 °C (Pedersen et al. 2014). The ASOs were synthesized by Exiqon (~1,500 ASOs) and GeneDesign (~500 ASOs) Inc. and subsequently classified as exonic or intronic based on overlap with exons in FANTOM CAT gene models.

Cell culture

Human dermal fibroblasts (HDFs) were derived from the dermis of normal human neonatal foreskin cells (Lonza, catalog number: C2509). The cells were cultured in Dulbecco's Modified Eagle's medium (high glucose with L-glutamine) supplemented with 10% fetal bovine serum at 37 °C in a 5% CO2 incubator. The passage number of the cells for transfection was maintained at six or seven.

Automated cell culturing, ASO transfection and cell harvesting

Robotic automation (Hamilton®) was established to provide stable environment and accurate procedural timing control for cell culturing and transfection. In brief, trypsin-EDTA detachment, cell number and viability quantification, cell seeding, transfection and cell harvesting were performed with automation. All transfections were divided into 28 runs at weekly basis. ASO transfection was performed with duplication. In each run, there were 16 independent transfections with ASO negative control A (NC_A, Exiqon) and 16 wells transfected with an ASO targeting *MALAT-1* (Exiqon).

The HDF cells were seeded in 12-well plates with 80,000 cells in each well 24 hrs prior to the transfection. A final concentration of 20 nM ASO and 2 µl lipofectamine RNAiMAX (Thermo Fisher Scientific) were mixed in 200 µl OptiMEM (Thermo Fisher Scientific). The mixture was incubated at room temperature for 5 min and added to the cells, which were maintained in 1 ml complete medium. The cells were harvested 48 hrs post-transfection by adding 200 µl RLT buffer from the RNeasy 96 Kit (Qiagen) after PBS washing. The harvested lysates were kept at -80 °C.

RNA purification

The harvested lysates were subjected to purification using RNeasy 96 Kit (Qiagen) and epMotion automated liquid handling systems (Eppendorf) according to the manufacturer's instructions, except an additional washing step by RPE buffer. The eluted RNA was qualified and quantified by the Dropsense spectrophotometry platform (Trinean). The RNA was stored at -80 °C.

Real-time quantitative RT-PCR

Real-time quantitative RT-PCR was performed by One Step SYBR PrimeScript™ RT-PCR Kit II (Takara) using the epMotion automated liquid handling system (Eppendorf). For each sample, three primer pairs against the specific lncRNA target were used. The expression level was normalized by GAPDH while the knockdown efficiency was calculated from the fold-change between each sample and NC_A. The knockdown efficiency of *MALAT-1* was monitored along each run and across all the runs where all the samples have >90% knockdown. The primers' sequences are listed in (Supplementary Table S2). In general, the knockdown RNA samples having >50% knockdown efficiency shown consistently by one of the primer pairs were subjected to CAGE. Exceptions included insufficient amount of RNA due to cell death and great knockdown variation between replicates.

ASO transfection for real-time imaging

The HDF cells were transfected manually in 96-well plate to facilitate high-throughput real time imaging. The cells were seeded 24 hrs before transfection at a density of 5,200 cells per well. A final concentration of 20 nM ASO and 2 µl lipofectamine RNAiMAX (Thermo Fisher Scientific) were mixed in 200 µl OptiMEM (Thermo Fisher Scientific). After incubating at room temperature for 5 min, 18 µl of the transfection mix was added to 90 µl complete medium in each well. The ASOs were divided in 14 runs and transfected in duplicates. Each plate accommodated 6 wells of NC_A control, 2 wells of *MALAT1* ASO control and 2 wells of mock-transfection (lipofectamine alone) control.

Phase-contrast images of transfected cells were captured every 3 hrs for 2 days with 3 fields per well by the IncuCyte® live-cell imaging system (Essen Bioscience). The confluence in each field was analyzed by the IncuCyte® software. The mean confluence of each well was taken along the timeline until the mean confluence of the NC_A control in the same plate reached 90%. The growth rate in each well was calculated as the slope of a linear regression. A normalized growth rate of each replicate was calculated as the growth rate divided by the mean growth rate of the 6 NC_A controls from the same plate. Student's t-test was performed between the growth rate of the duplicated samples and the 6 NC_A controls, assuming equal variance.

Cell fractionation for RNA-sequencing

A previously described method (Conrad and Ørom 2017) was adopted for the isolation of cytoplasmic, nucleoplasmic and chromatin-associated RNA. Approximately 10 million cells were used per fractionation experiment. Briefly, trypsinized cells were washed and lysed using cold

lysis buffer containing 0.15% Igepal CA-630, 10 mM Tris pH 7.5, 150 mM NaCl. The lysate was centrifuged in a sucrose cushion, after which the supernatant was taken as the cytoplasmic fraction. The nuclear pellet was washed once in buffer containing 20mM HEPES pH 7.5, 50% glycerol, 75 mM NaCl, 1 mM DTT, 0.5 mM EDTA and suspended again in the same buffer. An equal volume of nuclear lysis buffer containing 20mM HEPES pH 7.5, 300mM NaCl, 1M Urea, 1% Igepal CA-630, 10mM $MgCl_2$, 1mM DTT, 0.2mM EDTA was added and incubated on ice for 5 min. After centrifugation, the supernatant was considered as the nucleoplasmic fraction and the pellet as the chromatin fraction. The chromatin pellet was washed once in buffer containing 10 mM HEPES pH 7.5, 10 mM KCl, 10% glycerol, 340 mM sucrose, 4 mM $MgCl_2$, 1 mM DTT and suspended in the same buffer. RNA from each fraction was isolated using Trizol LS (Invitrogen) according to manufacturer's instructions. To ensure RNA purity, DNase I treatment followed by phenol-chloroform extraction was conducted. RNA isolated from each fraction was subjected to RNA-sequencing.

Ki-67 staining upon lncRNA knockdown

For the selected four lncRNA targets showing >25% growth inhibition, we used two siRNAs and ASOs with independent sequences. The *Silencer® Select siRNAs* were obtained from Invitrogen and ASOs were from geneDesign (Supplementary Table S9). The HDF cells were transfected with 20 nM siRNA or ASO in 12-well plates by lipofection. The cells were seeded 24 hours before transfection at a density of 60,000 cells per well. At 48 hrs post-transfection, cells were washed by PBS and harvested by trypsin-EDTA. Cells from the two wells with the same transfection were collected into one tube. After PBS washing, the cells were fixed by adding pre-chilled 70% ethanol and incubated in -20 °C for at least 2 hours. Ethanol was then removed by centrifugation and the cells were washed by FACS buffer (2% FBS in PBS, 0.05% NaN3) twice. FITC-conjugated Ki-67 (20Raj1, eBioscience) was applied to cells at a ratio of 8 µl per 150,000 cells. Same concentration of FITC-conjugated mouse IgG1 kappa antibody (eBioscience) was used as isotypic control. After 1 hr incubation at 4 °C, cells were washed by FACS buffer and subjected to flow cytometric analysis. Knockdown efficiency by siRNA was determined by real-time quantitative RT-PCR using the same 3 primer pairs as for ASO knockdown efficiency.

Wound closure assay

The HDF cells were transfected by 20nM ASO as described earlier in 12-well plates. The cells were re-plated at 24 hrs post-transfection into a 96-well ImageLock plate (Essen BioScience) at a density of 20,000 cells per well. At 24 hrs after seeding, cells form a spatially uniform monolayer

with 95-100% cell confluence. The cells were incubated with 5 μg/mL mitomycin-C for 2 hrs to inhibit cell division. Then, medium was refreshed and a uniform scratch was created in each well by the WoundMaker™(Essen BioScience). After changing the medium twice, the cells were maintained in medium with 0.5% FBS. The condition of mitomycin C and serum concentrations was tested with HDF and showed complete growth inhibition without severe morphological change (unpublished data). The closure of the wound was monitored by IncuCyte® live-cell imaging system (Essen Bioscience) every 2 hrs for 24 hrs. At each time point, the relative wound density was calculated by the cell confluence within the wound area normalized by the cell confluence of the non-wound area in the same image. The relative wound closure rate was calculated as the slope of the linear regression of the relative wound density against time, followed by normalization with that of NC_A. The RNA was harvested after the assay for real-time quantitative RT-PCR.

Cell morphology quantification

For each transfection, representative phase-contrast image at a single time point was exported from the Incucyte time-series, when the confluence of NC_A-transfected cells in each batch was around 80% (at 30 – 36 hrs post-transfection). These raw images were first transformed to probability maps of cells by pixel classification using ilastik (v1.3.2) (https://github.com/ilastik/ilastik.github.io/blob/master/publications.html). Three-pixel categories including cell, cell boundary, and background were manually labeled in a set of randomly selected images. The trained model was then applied to all images in the batch mode. The predicted probability maps of cells (grey scale, 16 bits tiff format) were subsequently used for morphology quantification in CellProfiler (v3.1.5) (Carpenter et al. 2006). In brief, binarized segmented cell images were obtained using the module IdentifyPrimaryObjects (thresholding by Global Otsu method, followed by intensity-based declumping) and the morphology measurements were performed by the module MeasureObjectSizeShape (http://cellprofiler-manual.s3.amazonaws.com/CellProfiler-3.0.0/modules/measurement.html#measureobjectsizeshape). All the values (medians) were further normalized by those of NC_A from the matching transfection plate, identically to normalizing the growth rate.

Cap analysis of gene expression (CAGE)

Four micrograms of purified RNA were used to generate libraries according to the nAnT-iCAGE protocol (Murata et al. 2014). Briefly, random primer with anchor was used for cDNA synthesis,

followed by biotinylation and RNaseI digestion. After cap trapping of the 5′ end complete cDNA and linker ligation, second-strand was synthesized for dsDNA.Libraries were combined in 8-plex using different barcodes and were subjected to 50-base single-end sequencing using an Illumina HiSeq 2500 instrument. Tag were de-multiplexed and mapped to human genome assembly hg38 using TopHat 2.0.12. The average mapping rate was 68.9% with around 10 million mapped counts obtained on average across all samples.

The samples with mapped counts lower than 500,000 were excluded from further analysis. Several samples were flagged samples as “questionable” if their mapped counts were less than 1,000,000 or their A260/A230 ratio were less than 1.0. Additionally, we manually flagged as “questionable” if any exceptional QC metrics were detected after sequencing, which includes lower amounts of library volumes and possible errors in experiments.

CAGE promoter and gene expression and batch correction

Expression for CAGE promoters was estimated by counting the numbers of mapped tags falling using 379,953 promoter regions of gene models as described in ‘*Gene Models and lncRNA targets selections*’. From there, expression of each of 124,245 genes was estimated by summing up the expression values of all promoters assigned to a given gene.

Batch correction was performed on the log-transformed ‘cpm’ values with the prior.count set to 0.25 and normalized for the library sizes using ‘removeBatchEffect’ function from the ‘limma’ R package where the ‘batch’ was attributed to the CAGE sequencing runs and ‘batch2’, ‘design’ and ‘covariates’ parameters not used.

logCPM <- cpm(dge, log=TRUE, prior.count=0.25, normalized.lib.sizes = T)

removeBatchEffect(logCPM, batch = ‘CAGE_lsid’)

Differential promoter and gene expression

Differential promoter/gene expression was carried separately for each ASO knockdown (mainly duplicated CAGE libraries) against matching (the same CAGE sequencing run) negative controls using DESeq 2.16 with default parameters. For a few ASO knockdowns where CAGE replicates were in distinct sequencing runs, a generalized linear model (*glm*) with appropriate design was used. Only promoters/genes with a mean count >=1TPM in either knockdown or negative control libraries were tested. Promoter/gene was differentially expressed if |log2FC|>0.5; fdr<0.05 and |ZScore|>1.645; where ZScore was obtained by scaling relative expression change (log2FC) of each tested promoter/gene was across all experiments.

Motif Activity Response Analysis (MARA)

MARA was performed using batch corrected promoter expression for all the knock-down (KD) and control (both NC_A and NC_B) libraries (970 CAGE libraries). All promoters with expression >=1TPM at least in 70% CAGE libraries (24,014 promoters) were used for the analysis. Transcription factor binding sites (TFBS) for hg38 were predicted as described previously (Arner et al. 2015) using MotEvo (Arnold et al. 2012) for the set of 190 position-weight matrix motifs in SwissRegulon (released on 13 July 2015) (Pachkov et al. 2013) on a multiple alignment of genome assemblies hg38 (human), rheMac3 (macaque), mm10 (mouse), rn6 (rat), bosTau8 (cow), equCab2 (horse), canFam3 (dog), monDom5 (opossum), and galGal4 (chicken). The number of predicted TFBS were counted for each motif in the -300 to +100 base pair from the midpoint of the FANTOM CAT promoters. Next MARA was performed to decompose CAGE expression profiles of the promoters in terms of their associated motifs, yielding the activity profile of all the motifs with at least 150 TFBS associated with the expressed promoters across the HDF KD+controls samples.

Gene Set Enrichment Analysis

Gene set enrichment (GO Biological Process, GO Molecular Function, GO Cellular Component, KEGG, Hallmark and Reactome) analysis was carried separately for each pathway and each ASO knockdown using fgsea (1.8.0) R-package with the following parameters: set.seed(42), minSize=15, maxSize=1000, nperm=100,000, nproc=1. Each pathway was tested based on log2FC preranked genes values from DESEq2 analysis. Pathway was significantly enriched if |NES|>1; fdr<0.05 and |ZScore|>1.645.; where ZScore was obtained by scaling GSEA significance of enrichment: -log10(pvalue)*sign(NES) of each tested pathway across all knockdowns.

Fantom5 coexpression clusters

Enrichment of Fantom5 coexpression clusters was calculated using fgsea (1.8.0) R-package with the following parameters: set.seed(42), minSize=15, maxSize=5000, nperm=100,000, nproc=1. Coexpression clusters were tested based on log2FC preranked genes values from DESEq2 analysis.

Unsupervised clustering of genomic regions

The unsupervised clustering was performed with CREAM algorithm using all promotes with log2FC<-1 or log2FC>1 and based on genomic coordinates of dysregulated transcripts.

Cellular and molecular phenotype correlations

Growth and morphology values were correlated for each ASO against 1) each GSEA pathway using enrichment score, 2) motif activity of each MARA motif (>=1TPM) and 3) each Fantom5 coexpression clusters using enrichment score.

Significant lncRNA targets

The background was estimated 1,000 times for each functional feature (DEG, MARA, GSEA) and for each lncRNA separately by choosing n random ASOs from n distictint lncRNAs (where n is the number of all ASOs targeting a given lncRNA of) and, similarly, requiring that at least two of these ASOs show concordant significance. A target response was considered significant if the number of DEG or MARA or GSEA was above that estimated from the matching random background. So for example if a given lncRNA targeted by 3 ASOs has shown 100 DE genes in any of the two ASOs then it would be significant if 3 randomly chosen ASOs (repeated 1000 times) had median of less than 100 significant DE genes.

Promoter switching analysis

To identify promoter switching events we used normalized and batch-corrected promoter counts, considering only genes reaching 4 cpm in at least two of CAGE libraries and promoters reaching at least 2 cpm in any of CAGE libraries. From this pool, we only considered the ASOs and promoters that were included into the promoter-level differential expression data. Promoter pairs passing these expression thresholds corresponded to 7151 genes in total.

For each KD, for expression values of pairs of promoters of the same gene $<p_A, p_B>$, we calculated their log-transformed ratios $r = \log_2 p_A - \log_2 p_B$, and estimated statistical significance of the difference $abs(r_{NC} - r_{KD})$ between the data of a target knockdown (KD) induced by a given ASO and the negative control (NC) libraries. Student's T-test assuming the equal variance in NC and KD samples was used to evaluated significance. The p-values were FDR-corrected for multiple testing for the number of tested promoter pairs. Z-score transformation of the log2 fold changes between r_{KD} and r_{NC} across all ASOs was used to additionally filter by Z-score.

Promoter architecture usage

To identify a CGI overlap and TATA-box presence at ~30 bp upstream of the TSS, CAGE data were analyzed to identify the precise location of the most frequently used TSS. First, the mapped reads were imported using CAGEr package (Haberle et al. 2015), and CAGE supported TSSs (CTSS) were clustered within 20 bp distance, with single CTSS only allowed when > 5 TPM. These clusters were annotated using FANTOM CAT promoters (CAGE supported promoters). These were cross referenced with the differential expression gene list per experiment. A negative control per ASO experiment was chosen randomly to represent core promoter architecture and each CAGE supported promoter was annotated for a CGI overlapping the promoter and the presence of a TATA-box (PWM > 80 %) at -35 to -25 bp upstream of the most frequently used TSS. To avoid an effect of expression levels, we then divided, for each ASO, the CAGE supported promoters into 5 bins of expression level (low to high) and used stratified sampling (10,000 times) of the number of promoters in each class of differentially expressed genes separately (up and down) to assess significance. The background consisted of all promoters per ASO experiment in the negative control that were not significantly up- or downregulated.

Chromosome conformation capture (Hi-C)

Hi-C libraries were prepared essentially as described previously (Fraser, Ferrai, et al. 2015; Lieberman-Aiden et al. 2009) with minor changes to improve the DNA yield of Hi-C products (Fraser, Williamson, et al. 2015). The procedure followed is outlined as a flowchart in Supplementary Fig. 7a, and each step of the protocol is briefly described below.

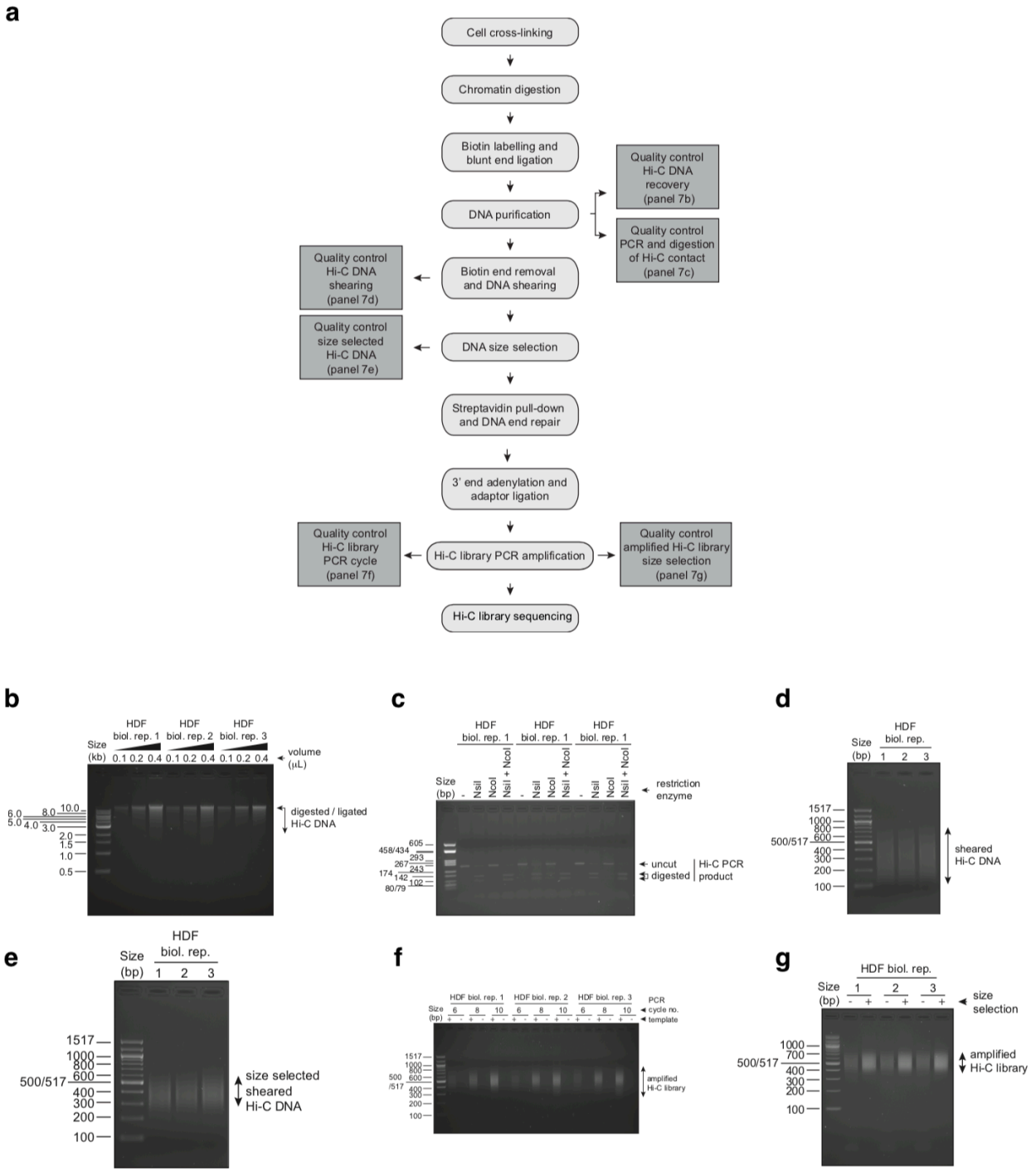

Fig. S7: Hi-C protocol overview. a, Flowchart of the Hi-C protocol with related quality control (QC) steps (b-g). b, cDNA recovery quality control. c, PCR and digestion QC. d, DNA shearing QC. e, DNA size-selection QC. f, PCR cycle QC. g, DNA size-selection QC.

Cell cross-linking

Three biological replicate samples of HDFs (~1 X 10^7 cells) grown as described above were fixed at room temperature in media containing 1% formaldehyde (Sigma, catalog number: F8775) for 10 minutes with gentle rocking every 2 minutes. Cross-linking was stopped by quenching the formaldehyde with glycine (Sigma, catalog number: G8898) at a final concentration of 125 mM for 5 minutes at room temperature, followed by 15 minutes on ice. The cells were then scraped off the plates and pelleted by gentle centrifugation at 400 *g* for 10 minutes. The quenched media was removed, and the cell pellets were quick-frozen on dry ice before storage at -80°C.

Chromatin digestion

Cell pellets were resuspended into 440 µl of 'Cold Lysis Buffer' containing protease inhibitors (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% Igepal; protease inhibitor cocktail as instructed by the manufacturer (Sigma, catalog number: P8340). The cells were incubated on ice for 15 minutes and lysed by twice 20 strokes with a Dounce homogenizer (Pestle B). Cell lysates were centrifuged at 5,000 rpm for 5 minutes at room temperature in a microcentrifuge. The pellets were washed twice with 400 µl of 1X NEBuffer™ 3.1, centrifuged 5 minutes at room temperature at 5,000 rpm, and resuspended in 200 µl of 1X NEBuffer™ 3.1. Each sample was divided equally into 4 tubes before adding 312 µl of 1X NEBuffer™ 3.1 to each tube. 1% SDS (38 µl) was added to each sample and incubated 10 minutes at 65°C. 10% Triton X-100 (44 µl) was mixed into each tube before adding 40 µl of NcoI (10 U/µl; 400 Units; NEB, catalog number: R0193L) and incubating a 37°C overnight.

Biotin labeling and blunt end ligation

Digested fragment ends were labeled with biotin by filling-in NcoI 5' overhangs with Klenow using biotinylated dCTP. Deoxynucleotides (1.5 µl each of 10 mM dATP, dGTP, dTTP, and 37.5 µl of 0.4 mM biotin-14-dCTP (Life Technologies, catalog number: 19518018)) were added to each tube, along with 10 µl of Klenow (5 U/µl; 50 Units; NEB, catalog number: M0210S), and incubated at 37°C for 45 minutes. 10% SDS (86 µl) was mixed into each sample before incubating 30 minutes at 65°C. The samples were transferred to 15 mL conical tubes containing 5.96 mL of water and 1.66 mL of 'Ligation mix' (10% Triton X-100 (750 µl), 10X Ligation Buffer (750 µl; 0.5 M Tris pH 7.5, 0.1 M $MgCl_2$, 0.1 M DTT), 10 mg/mL BSA (80 µl), 100 mM ATP (80 µl)). T4 DNA ligase (50 µl) was added to each tube (1 U/µl; 50 Units, Invitrogen, catalog number: 15224-025), and incubated 4 hours at 16°C.

DNA purification

The DNA was purified by proteinase K digestion followed with phenol/chloroform extraction and ethanol precipitation. Proteinase K (50 µl of 10 mg/mL) was added to each tube and incubated overnight at 65°C. Proteinase K was added again on the next day (50 µl of 10 mg/mL) and incubated for an additional 2 hours. The samples were transferred to 50 mL conical tubes and extracted with 10 mL of phenol by vortexing 2 minutes and centrifugation at 2,465 *g* for 15 minutes at room temperature. The samples were extracted again as above but with phenol/chloroform before precipitating with ethanol. To this end, samples were transferred to 35 mL centrifuge bottles, and their volumes adjusted to 10 mL with 1X TE before 1 mL of 3 M sodium acetate (pH 5.2) was mixed in and 25 mL of ice-cold ethanol was added. Each sample was gently mixed by inversion and incubated 1 hour at -80°C. The DNA was pelleted by centrifugation at 23,281 *g* (13,000 rpm if using SS34 rotor) for 25 minutes at 4°C, washed by resuspending in 10 mL of ice-cold 70% ethanol, and centrifuged again as above. Pellets were each dissolved into 450 µl of 1X TE pH 8.0 and extracted twice with 500 µl of phenol-chloroform by vortexing 1 minute and centrifuging at maximum speed for 5 minutes at room temperature in a microcentrifuge. The samples were precipitated with 3 M sodium acetate pH 5.2 (40 µl) and absolute ethanol (1 mL), incubated at -80°C overnight, and centrifuged at maximum speed for 25 minutes at 4°C. The DNA was washed five times by resuspending in ice-cold 70% ethanol (1 mL), each time centrifuging at maximum speed for 20 minutes at 4°C. The resulting pellets were each dissolved in 25 µl of 1X TE pH 8.0 before the 4 samples from original cell pellets were pooled. 10 mg/mL RNAse A (1 µl; Fermentas, catalog number: EN0531) was added and incubated for 30 minutes at 37°C. The Hi-C DNA was resolved on 0.8% agarose gel containing ethidium bromide (0.5 µg/mL) for qualitative assessment of digestion, ligation, and yield based on calibrated molecular weight markers (Supplementary Fig. 7b).

Quality control

The quality of Hi-C DNA was quantitatively assessed by PCR titration of an expected Hi-C contact and digestion of the PCR product as previously outlined (Lieberman-Aiden et al. 2009). The Hi-C DNA was serially diluted in water from 1 to ~0.0005 µL (2-fold; 11 dilutions), and 4 µL aliquots were PCR-amplified (35 cycles) with primers against a gene desert sequence (GD09 NcoI 5'-GCAATTAGTGCTATGCCCATGTTTCCTTTGTTCC-3', GD10 NcoI 5'-CAGTCTTCTACCGCTCTTGTAATGGGGTT-3'). Half of each PCR reaction was resolved on 1.5% agarose gels containing ethidium bromide (0.5 µg/mL) to verify the presence of amplification products. The remaining products of the first 5 PCR reactions from each titration were pooled,

and divided equally into 4 tubes to digest with either NcoI, NsiI, NcoI/NsiI, or buffer control as described previously (Lieberman-Aiden et al. 2009). Digestion reactions were incubated 1 hour at 37°C, and resolved on 2% agarose gels containing ethidium bromide (0.5 µg/mL) to verify greater digestion efficiency with NsiI. In all 3 HDF biological replicates, over 75% of this PCR product digested specifically with NsiI (Supplementary Fig. 7c), pointing to efficient Hi-C product formation in the samples.

Removal of biotin from unligated ends and DNA shearing

Biotin at unligated restriction fragment ends was removed with T4 DNA polymerase as we described previously (Fraser, Ferrai, et al. 2015). Briefly, 10 samples each of 5 µg DNA from individual HDF replicates were mixed with 10µl of 10X NEBuffer™ 2.1, 1 µl of 10 mM dCTP, and 5 Units of T4 DNA polymerase (NEB, catalog number: M0203L) in a final volume of 100 µl in MAXYMum Recovery™ PCR tubes (200 µl; Axygen, part number: PCR-02-L-C). The reactions were incubated in a thermocycler for 2 hours at 12°C, after which the enzyme was inactivated at 75°C for 20 minutes. The samples were transferred to 1.7 mL tubes (Corning™ Costar™ Low Binding Plastic Microcentrifuge Tubes, part # C3207), and the DNA was precipitated by mixing 10 µl of 3 M sodium acetate (pH 5.2) in each sample, followed by 275 µl of ice-cold ethanol, and incubating at -80°C overnight. The DNA was pelleted by centrifugation at 15,000 rpm in a microcentrifuge at 4°C for 25 minutes, washed twice with 500 µl ice-cold 70% ethanol each time centrifuging as above, and the resulting pellets were each dissolved in 130 µl of 10 mM Tris pH 8.5. The DNA was sheared to ~350 bp fragments by sonication with a 'Covaris M220 Focused-utrasonicator' using preset settings (DNA_0300_bp_130µl_Snap_Cap_Micro_TUBE). Samples were individually transferred to 1.7 mL tubes and shearing to appropriate size range was verified by resolving 400 ng of each sample onto 1.5% agarose gel containing ethidium bromide (0.5 µg/mL; Supplementary Fig. 7d).

DNA size selection

The sheared DNA was size-selected using AMPure XP Beads (Agencourt XP) with ratios of 0.6X and 0.85X according to the manufacturer's instructions (Beckman Coulter, catalog number: A63880). Briefly, 78 µl of beads was added to each 130 µl sample and incubated with rotation for 30 minutes at room temperature. The samples were placed on a DynaMag™ Magnet (Thermo Fisher Scientific) for 2 minutes, and the supernatant was re-extracted with 32.5 µl of beads as above to capture the appropriate size fragments. The beads were washed twice with 500 µl of ice-cold ethanol and air-dried 5 minutes. The DNA was eluted with 60 µl of 10 mM Tris pH 8.5 by

resuspending ten times with a pipette, and the 10 supernatants from each biological replicate were pooled to a 1.7 mL tube after placing samples on the magnetic stand for 2 minutes. The sample volumes were then reduced to 300 µl with a SpeedVac concentrator, and 5 µl of each sample was used to measure DNA concentrations with the Quant-iT™ PicoGreen™ dsDNA Assay Kit as per the manufacturer's instructions (Thermo Fisher Scientific, catalog number: P11496). This measure will be used below to calculate the amount of sequencing adaptor required. DNA recovery was verified by resolving 5 µl of each sample on a 1.5% agarose gel containing ethidium bromide (0.5 µg/mL; Supplementary Fig. 7e).

Streptavidin pull-down and DNA end repair

The biotinylated Hi-C products were pull-down on Dynabeads MyOne™ Streptavidin C1 Beads (Thermo Fisher Scientific, catalog number: 65001) to enhance DNA end repair, 3' end adenylation, and ligation of sequencing adaptors. The beads (60 µl for each Hi-C biological replicate) were washed with 400 µl 'Tween Wash Buffer' (TWB; 5 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween 20) twice for 3 minutes at room temperature with rotation, and pelleted on a DynaMag™ Magnet for 2 minutes between washes. Beads were next washed once with 300 µl of '2X Binding Buffer' (2XBB; 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 2 M NaCl), and resuspended in 300 µl of 2XBB. Each 300 µl-Hi-C library was added to a tube containing beads and incubated at room temperature for 15 minutes with rotation to bind the DNA. The supernatant was removed, and the beads were washed twice with 400µl of 1XBB (5 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 M NaCl), and once with 100 µl of 1X Ligation Buffer (diluted from 10X, NEB, catalog number: B0202S). The supernatant was removed and 100 µl of 'DNA End Repair Mix' was added to each sample (10X Ligation Buffer (10 µl), 10 mM dNTPs (4 µl), T4 DNA Polymerase (5µl of 3 U/µl, NEB), T4 Polynucleotide Kinase (5 µl of 10 U/µl, NEB), Klenow (1 µl of 5 U/µl, NEB), water (75 µl). The samples were transferred to MAXYMum Recovery™ 200 µl PCR tubes and placed in a thermocycler for 30 minutes at 20°C. Samples were then transferred to 1.7 mL tubes, and the beads were washed twice with 200 µl of TWB, and twice with 200 µl Elution Buffer (EB; 10 mM Tris-HCl pH 8.5), each time pelleting beads on a magnetic stand for 2 minutes.

3' end adenylation and sequencing adapter ligation

To prevent concatenation of Hi-C products, DNA fragments were 3' adenylated and ligated to Illumina Paired-End (PE) sequencing adaptors with 3'-T overhangs. To this end, the supernatant was first removed from the beads and 50 µl of 3' Adenylation Mix' was added to each tube (10X NEBuffer™ 2 (5 µl), 10 mM dATP (1 µl), Klenow Fragment (3'→5' exo-; 3 µl of 5 U/µl, NEB, catalog

number: M0212S), water (41 µl). The samples were transferred to MAXYMum Recovery™ 200 µl PCR tubes and incubated 60 minutes at 37°C in a thermocycler. The samples were transferred to 1.7 mL tubes, and the beads were washed twice with 200 µl of TWB, and twice with 200 µl EB, each time pelleting beads on a magnetic stand for 2 minutes.

The amount of Illumina PE Adaptor needed for ligation depends on the amount of DNA in Hi-C samples. As a rule, 6 pmol of Illumina PE Adaptor (TruSeq™ DNA PCR-Free LT Library Prep Kit-Set A, Illumina, catalog number: FC-121-3001) was used for every 1 µg of DNA measured with the Quant-iT™ PicoGreen™ dsDNA Assay Kit after size selection. The EB was removed from the beads and 45 µl of 'Adaptor Ligation Mix' (5X Invitrogen Ligation buffer (10 µl), Illumina PE Adaptors, water) was added. Sample were mixed by pipetting, and 5 µl of T4 DNA Ligase (1 Weiss U/µl, Invitrogen, catalog number: 15224-025) was added to each tube. The samples were transferred to MAXYMum Recovery™ 200 µl PCR tubes and incubated 60 minutes at 20°C in a thermocycler. The samples were transferred to 1.7 mL tubes, and the beads were washed twice for 6 minutes with 400 µl of TWB, and twice with 400 µl EB, each time pelleting beads on a magnetic stand for 1 minute. Each sample was finally resuspended in 50 µl of EB.

PCR amplification of Hi-C libraries

As a rule, the lowest possible PCR cycle number should be used to amplify enough Hi-C libraries to reduce amplification biases. To identify the optimal number of PCR cycles, 3 PCR reactions (25 µl) were prepared for each Hi-C library and amplified through either 6, 8, or 10 cycles. Individual PCR reactions were composed of 1 µl Hi-C library, 1.25 µl each of 10 µM Illumina PE 1.0 and PE 2.0 primers, 12.5 µl of Phusion High-Fidelity 2X Master Mix (NEB, catalog number: M0531S), and 9 µl of water. The reactions were conducted in MAXYMum Recovery™ 200 µl PCR tubes using the following PCR program: 1 cycle of 30 seconds at 98°C, either 6, 8, or 10 cycles of 10 seconds at 98°C / 30 seconds at 65°C / 30 seconds at 72°C, and a final cycle of 7 minutes at 72°C. The reactions were resolved on a 2.0% agarose gel containing ethidium bromide (0.5 µg/mL) to select the lowest possible cycle number when a product is detected (Supplementary Fig. 7f).

Eight PCR cycles was used for large-scale amplification of the libraries for which one set of 10 PCR reactions for each replicate were prepared as described above except that 1.5 µl of Hi-C library was used in each reaction. The 10 PCRs from each replicate were pooled and their combined volume was adjusted to 215 µl with water. The amplified DNA was then purified using AMPure XP Beads with a ratio of 0.8X by adding 172 µl of washed beads to each pooled reaction tube and incubating for 10 minutes at room temperature while mixing. The captured DNA was

pelleted on a magnetic stand 2 minutes and washed twice on the beads with 1 mL of ice-cold 70% ethanol. The beads were air-dried for no more than 5 minutes and the DNA was eluted from the beads by adding 33 µl of 10 mM Tris pH 8.0, 0.1 mM EDTA and pipetting to resuspend the beads. Sample concentrations were measured using 2 µl of the library and the Quant-iT™ PicoGreen™ dsDNA Assay Kit as per the manufacturer's instructions (Thermo Fisher Scientific). DNA recovery and quality was verified by resolving 5 µL of each sample on a 2.5% agarose gel containing ethidium bromide (0.5 µg/mL; Supplementary Fig. 7g).

Sequencing and processing

The libraries were sequenced on a HiSeq 2500 Illumina platform, and paired-end reads were mapped to hg38 and processed using HiCUP pipeline ver. 0.5.10 (Wingett et al. 2015). Ditags were mapped against the human genome assembly hg38. Experimental artifacts such as circularized, re-ligated, continuous and incorrect size fragments were filtered out. PCR duplicates were removed from the aligned data. The sequencing and mapping metrics are in (Supplementary Table S10). Processed mapped ditags from all the three replicates were merged using Samtools ver. 1.3.1 (Li et al. 2009) for the further analysis. Significant co-localized regions at 10kb resolution were identified using BioConductor package GOTHiC (Mifsud B., PMID: 28379994). All the identified co-localized regions with p-value =< 0.01 and fdr =< 0.05 were used for the downstream analysis. TADs were identified using Arrowhead from Juicer pipeline (Durand et al. 2016). For the downstream analysis FANTOM CAT promoters were mapped to the colocalized regions and TADs.

RNA Fluorescence in situ hybridization (FISH)

Oligonucleotide probes against target RNA were designed using the Stellaris Probe Designer version 4.2 (Biosearch Tech). Probes were flanked on both ends with overhang arms serving as annealing sites for secondary probes labeled with a fluorescent dye (Chen et al 2015). Overhang sequences were identical on both ends. Secondary probe sequences have been previously described (Moffit et al 2016) and were labeled on the 3' end with Atto 647N. All probe sequences used in this study can be found in (Supplementary Table S11). Two-step hybridization was performed using a previously-described procedure (Kouno et al. 2019). Briefly, fibroblasts were seeded onto coverslips overnight and fixed in 4% formaldehyde in PBS for 10 min at room temperature. Following fixation, coverslips were treated twice with ice-cold 0.1% sodium borohydride for 5 min at 4 °C. Coverslips were washed three times in PBS, followed by cell permeabilization in 0.5% Triton X-100 in PBS for 10 min at room temperature. Coverslips were

again washed three times in PBS and treated with 70% formamide in 2X SSC for 10 min at room temperature. Coverslips were washed twice in ice-cold PBS and once in ice-cold 2X SSC. Coverslips were either used immediately for hybridization or stored in 70% ethanol for no longer than a week, in which case were washed in PBS twice and once in 2X SSC at room temperature prior to hybridization. For hybridization, coverslips were incubated at 37 °C overnight in hybridization buffer containing 252 mM primary probe inside a humid chamber. Hybridization buffer contained 10% dextran sulfate, 1ug/ul yeast tRNA, 2mM vanadyl ribonucleoside complex, 0.02% BSA, 10% formamide, 2X SSC. Excess probe was removed by two washes for 30 min at room temperature in wash buffer containing 30% formamide, 2X SSC, 0.1% Triton X-100, followed by a rinse in 2X SSC. For second hybridization, coverslips were incubated at 37 °C for 1.5 hrs in hybridization buffer containing 30 nM secondary probe. Excess probe was washed twice for 20 min at room temperature. Coverslips were stained for 5 min in 2 mg/mL Hoescht in PBS, washed three times in PBS, and mounted on a glass slide with SlowFade Gold Antifade Mountant (Invitrogen). Imaging was done on a DeltaVision microscope (GE Healthcare) equipped with a sCMOS sensor. Image processing was done using FIJI (Chen et al. 2015; Moffitt et al. 2016; Kouno et al. 2019)

Acknowledgments

The work was supported by Ministry of Education, Science, Sports and Technology (MEXT) fund to RIKEN Center for Life Science Technologies (CLST) and Integrative Medical Sciences (IMS). The authors wish to acknowledge RIKEN GeNAS for generation and sequencing of the CAGE libraries using Illumina HiSeq 2500, and subsequent data processing. IVK and IEV were supported by RFBR 18-34-20024, AleFa was supported by NIH P30 CA006973 and RFBR 17-00-00208, VsMa was supported by the Russian Academy of Sciences Project 0112-2019-0001, NIH-NCI award R01CA200859 and DOD award W81XWH-16-1-0739, YAM was supported by RSF 18-14-00240, Wellcome Trust Investigator Award to FM and ZENCODE-ITN (H2020) programme of the European Commission to FM and AK. HB was supported by National Bioscience Database Center (NBDC), Japan Science and Technology Agency (JST). AF is supported by an Australian National Health and Medical Research Council Fellowship APP1154524 and project grant APP1146321. CAMS, MST and RSY were funded by a UK Medical Research Council (MRC) core grant to the MRC Human Genetics Unit. JD and DP were supported by the Canadian Institutes of Health Research (CIHR MOP-142451). Work from CIB was supported by grants MIUR n.974,CMPT177780 and MIUR n.257/Ric, CTN01_00177_817708, Y.Ciani was supported by a FIRC fellowship. RG acknowledges the support of the “Salvador de Madariaga” program from the Spanish Ministry of Education, Culture and Sports during his sabbatical at the Riken; MMH was supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-3948). We also thank Hiroto Atsui for research agreements and legal advice, Ri-ichiro Manabe for ethics, and Saito Tsutomu for finances. We also thank Dr. Minh Doan (Broad Institute, Boston, MA, USA) for discussion and advice on cell morphology quantification of Incucyte images.

Author Information

Competing interests - all authors declare no competing interest

Bibliography

Ambasudhan, R., Talantova, M., Coleman, R., Yuan, X., Zhu, S., Lipton, S.A. and Ding, S. 2011. Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. *Cell Stem Cell* 9(2), pp. 113–118.

Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., Freeman, T.C., Alhendi, A.M.N., Arner, P., Axton, R., Baillie, J.K., Beckhouse, A., Bodega, B., Briggs, J., Brombacher, F., Davis, M. and Hayashizaki, Y. 2015. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* 347(6225), pp. 1010–1014.

Arnold, P., Erb, I., Pachkov, M., Molina, N. and van Nimwegen, E. 2012. MotEvo: integrated Bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of DNA sequences. *Bioinformatics* 28(4), pp. 487–494.

Bai, J., Yao, B., Wang, L., Sun, L., Chen, T., Liu, R., Yin, G., Xu, Q. and Yang, W. 2018. lncRNA A1BG-AS1 suppresses proliferation and invasion of hepatocellular carcinoma cells by targeting miR-216a-5p. *Journal of Cellular Biochemistry*.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V. and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) 2005. The transcriptional landscape of the mammalian genome. *Science* 309(5740), pp. 1559–1563.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C., Forrest, A.R.R., Alkema, W.B., Tan, S.L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y. and Hayashizaki, Y. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics* 38(6), pp. 626–635.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Golland, P. and Sabatini, D.M. 2006. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology* 7(10), p. R100.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I.,

1070 Collavin, L., Santoro, C., Forrest, A.R.R., Carninci, P., Biffo, S., Stupka, E. and Gustincich, S.
1071 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded
1072 SINEB2 repeat. *Nature* 491(7424), pp. 454–457.

1073 Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S. and Zhuang, X. 2015. RNA imaging.
1074 Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348(6233), p.
1075 aaa6090.

1076 Cho, S.W., Xu, J., Sun, R., Mumbach, M.R., Carter, A.C., Chen, Y.G., Yost, K.E., Kim, J., He,
1077 J., Nevins, S.A., Chin, S.-F., Caldas, C., Liu, S.J., Horlbeck, M.A., Lim, D.A., Weissman, J.S.,
1078 Curtis, C. and Chang, H.Y. 2018. Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA
1079 Boundary Element. *Cell* 173(6), p. 1398–1412.e22.

1080 Collado-Torres, L., Nellore, A., Kammers, K., Ellis, S.E., Taub, M.A., Hansen, K.D., Jaffe, A.E.,
1081 Langmead, B. and Leek, J.T. 2017. Reproducible RNA-seq analysis using recount2. *Nature*
1082 *Biotechnology* 35(4), pp. 319–321.

1083 Conrad, T. and Ørom, U.A. 2017. Cellular Fractionation and Isolation of Chromatin-Associated
1084 RNA. *Methods in Molecular Biology* 1468, pp. 1–9.

1085 Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin,
1086 D., Merkel, A., Knowles, D.G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T.,
1087 Carninci, P., Brown, J.B., Lipovich, L., Gonzalez, J.M., Thomas, M. and Guigó, R. 2012. The
1088 GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure,
1089 evolution, and expression. *Genome Research* 22(9), pp. 1775–1789.

1090 Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S. and Aiden,
1091 E.L. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C
1092 Experiments. *Cell systems* 3(1), pp. 95–98.

1093 Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E.,
1094 Guttman, M. and Lander, E.S. 2016. Local regulation of gene expression by lncRNA promoters,
1095 transcription and splicing. *Nature* 539(7629), pp. 452–455.

1096 Engreitz, J.M., Ollikainen, N. and Guttman, M. 2016. Long non-coding RNAs: spatial amplifiers
1097 that control nuclear structure and gene expression. *Nature Reviews. Molecular Cell Biology*
1098 17(12), pp. 756–770.

1099 FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R.R., Kawaji, H., Rehli,
1100 M., Baillie, J.K., de Hoon, M.J.L., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh,

M., Andersson, R., Mungall, C.J., Meehan, T.F., Schmeier, S., Bertin, N., Jørgensen, M., Dimont, E., Arner, E., Schmidl, C. and et al. 2014. A promoter-level mammalian expression atlas. *Nature* 507(7493), pp. 462–470.

FANTOM Consortium, Suzuki, H., Forrest, A.R.R., van Nimwegen, E., Daub, C.O., Balwierz, P.J., Irvine, K.M., Lassmann, T., Ravasi, T., Hasegawa, Y., de Hoon, M.J.L., Katayama, S., Schroder, K., Carninci, P., Tomaru, Y., Kanamori-Katayama, M., Kubosaki, A., Akalin, A., Ando, Y., Arner, E. and Hayashizaki, Y. 2009. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature Genetics* 41(5), pp. 553–562.

Fraser, J., Ferrai, C., Chiariello, A.M., Schueler, M., Rito, T., Laudanno, G., Barbieri, M., Moore, B.L., Kraemer, D.C.A., Aitken, S., Xie, S.Q., Morris, K.J., Itoh, M., Kawaji, H., Jaeger, I., Hayashizaki, Y., Carninci, P., Forrest, A.R.R., FANTOM Consortium, Semple, C.A. and Nicodemi, M. 2015. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Molecular Systems Biology* 11(12), p. 852.

Fraser, J., Williamson, I., Bickmore, W.A. and Dostie, J. 2015. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiology and Molecular Biology Reviews* 79(3), pp. 347–372.

Garieri, M., Delaneau, O., Santoni, F., Fish, R.J., Mull, D., Carninci, P., Dermitzakis, E.T., Antonarakis, S.E. and Fort, A. 2017. The effect of genetic variation on promoter usage and enhancer activity. *Nature Communications* 8(1), p. 1358.

GTEx Consortium 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348(6235), pp. 648–660.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., Wang, Y., Brzoska, P., Kong, B., Li, R., West, R.B., van de Vijver, M.J., Sukumar, S. and Chang, H.Y. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464(7291), pp. 1071–1076.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., Yang, X., Amit, I., Meissner, A., Regev, A., Rinn, J.L., Root, D.E. and Lander, E.S. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477(7364), pp. 295–300.

Guttman, M. and Rinn, J.L. 2012. Modular regulatory principles of large non-coding RNAs. *Nature* 482(7385), pp. 339–346.

1132 Haberle, V., Forrest, A.R.R., Hayashizaki, Y., Carninci, P. and Lenhard, B. 2015. CAGEr:
1133 precise TSS data retrieval and high-resolution promoterome mining for integrative analyses.
1134 *Nucleic Acids Research* 43(8), p. e51.

1135 Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin,
1136 A., Suzuki, A.M., van IJcken, W.F.J., Armant, O., Ferg, M., Strähle, U., Carninci, P., Müller, F.
1137 and Lenhard, B. 2014. Two independent transcription initiation codes overlap on vertebrate core
1138 promoters. *Nature* 507(7492), pp. 381–385.

1139 Hardwick, S.A., Bassett, S.D., Kaczorowski, D., Blackburn, J., Barton, K., Bartonicek, N.,
1140 Carswell, S.L., Tilgner, H.U., Loy, C., Halliday, G., Mercer, T.R., Smith, M.A. and Mattick, J.S.
1141 2019. Targeted, High-Resolution RNA Sequencing Of Non-Coding Genomic Regions
1142 Associated With Neuropsychiatric Functions. *Biorxiv*, p. 539882.

1143 Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans,
1144 M., Furey, T.S., Harte, R.A., Hsu, F., Hillman-Jackson, J., Kuhn, R.M., Pedersen, J.S., Pohl, A.,
1145 Raney, B.J., Rosenbloom, K.R., Siepel, A., Smith, K.E., Sugnet, C.W., Sultan-Qurraie, A. and
1146 Kent, W.J. 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research*
1147 34(Database issue), pp. D590-8.

1148 Hon, C.-C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J.L., Gough, J.,
1149 Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J., Lizio, M., Kawaji, H., Kasukawa, T.,
1150 Itoh, M., Burroughs, A.M., Noma, S., Djebali, S., Alam, T., Medvedeva, Y.A., Testa, A.C. and
1151 Forrest, A.R.R. 2017. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*
1152 543(7644), pp. 199–204.

1153 de Hoon, M., Shin, J.W. and Carninci, P. 2015. Paradigm shifts in genomics through the
1154 FANTOM projects. *Mammalian Genome* 26(9–10), pp. 391–402.

1155 Imada, E.L., Sanchez, D.F., Collado-Torres, L., Wilks, C., Matam, T., Dinalankara, W.,
1156 Stupnikov, A., Lobo-Pereira, F., Yip, C.W., Yasuzawa, K., Kondo, N., Itoh, M., Suzuki, H.,
1157 Kasukawa, T., Hon, C.C., de Hoon, M., Shin, J.W., Carninci, P., FANTOM consortium, Jaffe,
1158 A.E., Leek, J.T., Favorov, A., Franco, G.R., Langmead, B., Marchionni L. Recounting the
1159 FANTOM Cage Associated Transcriptome bioRxiv 659490; doi: https://doi.org/10.1101/659490
1160
1161 Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner,
1162 J.R., Evans, J.R., Zhao, S., Poliakov, A., Cao, X., Dhanasekaran, S.M., Wu, Y.-M., Robinson,
1163 D.R., Beer, D.G., Feng, F.Y., Iyer, H.K. and Chinnaiyan, A.M. 2015. The landscape of long
1164 noncoding RNAs in the human transcriptome. *Nature Genetics* 47(3), pp. 199–208.

1165 Joung, J., Engreitz, J.M., Konermann, S., Abudayyeh, O.O., Verdine, V.K., Aguet, F.,
1166 Gootenberg, J.S., Sanjana, N.E., Wright, J.B., Fulco, C.P., Tseng, Y.-Y., Yoon, C.H., Boehm,
1167 J.S., Lander, E.S. and Zhang, F. 2017. Genome-scale activation screen identifies a lncRNA
1168 locus regulating a gene neighbourhood. *Nature* 548(7667), pp. 343–346.

1169 Kalluri, R. 2016. The biology and function of fibroblasts in cancer. *Nature Reviews. Cancer*
1170 16(9), pp. 582–598.

1171 Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D.
1172 and Ren, B. 2005. A high-resolution map of active promoters in the human genome. *Nature*
1173 436(7052), pp. 876–880.

1174 Kopp, F. and Mendell, J.T. 2018. Functional classification and experimental dissection of long
1175 noncoding rnas. *Cell* 172(3), pp. 393–407.

1176 Kouno, T., Moody, J., Kwon, A.T.-J., Shibayama, Y., Kato, S., Huang, Y., Böttcher, M., Motakis,
1177 E., Mendez, M., Severin, J., Luginbühl, J., Abugessaisa, I., Hasegawa, A., Takizawa, S.,
1178 Arakawa, T., Furuno, M., Ramalingam, N., West, J., Suzuki, H., Kasukawa, T. and Shin, J.W.
1179 2019. C1 CAGE detects transcription start sites and enhancer activity at single-cell resolution.
1180 *Nature Communications* 10(1), p. 360.

1181 Kuo, C.-C., Hänzelmann, S., Sentürk Cetin, N., Frank, S., Zajzon, B., Derks, J.-P., Akhade,
1182 V.S., Ahuja, G., Kanduri, C., Grummt, I., Kurian, L. and Costa, I.G. 2019. Detection of RNA-DNA
1183 binding sites in long noncoding RNAs. *Nucleic Acids Research*.

1184 Li, B. and Wang, J.H.-C. 2011. Fibroblasts and myofibroblasts in wound healing: force
1185 generation and measurement. *Journal of tissue viability* 20(4), pp. 108–120.

1186 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
1187 Durbin, R. and 1000 Genome Project Data Processing Subgroup 2009. The Sequence
1188 Alignment/Map format and SAMtools. *Bioinformatics* 25(16), pp. 2078–2079.

1189 Li, L. and Chang, H.Y. 2014. Physiological roles of long noncoding RNAs: insight from knockout
1190 mice. *Trends in Cell Biology* 24(10), pp. 594–602.

1191 Li, X., Zhou, B., Chen, L., Gou, L.-T., Li, H. and Fu, X.-D. 2017. GRID-seq reveals the global
1192 RNA-chromatin interactome. *Nature Biotechnology* 35(10), pp. 940–950.

1193 Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A.,
1194 Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A.,
1195 Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S. and Dekker, J.

2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950), pp. 289–293.

Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., Mandegar, M.A., Olvera, M.P., Gilbert, L.A., Conklin, B.R., Chang, H.Y., Weissman, J.S. and Lim, D.A. 2017. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 355(6320).

Liu, Y., Cao, Z., Wang, Y., Guo, Y., Xu, P., Yuan, P., Liu, Z., He, Y. and Wei, W. 2018. Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nature Biotechnology*.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., Mesirov, J.P., Cech, T.R. and Chang, H.Y. 2016. RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* 165(5), pp. 1267–1279.

Luo, S., Lu, J.Y., Liu, L., Yin, Y., Chen, C., Han, X., Wu, B., Xu, R., Liu, W., Yan, P., Shao, W., Lu, Z., Li, H., Na, J., Tang, F., Wang, J., Zhang, Y.E. and Shen, X. 2016. Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells. *Cell Stem Cell* 18(5), pp. 637–652.

Moffitt, J.R., Hao, J., Wang, G., Chen, K.H., Babcock, H.P. and Zhuang, X. 2016. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proceedings of the National Academy of Sciences of the United States of America* 113(39), pp. 11046–11051.

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A.R., Hoberg, E., Moustakas, A., Gyllensten, U., Jones, S.J.M., Gustafsson, C.M., Sims, A.H., Westerlund, F., Gorab, E. and Kanduri, C. 2015. MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. *Nature Communications* 6, p. 7743.

Murata, M., Nishiyori-Sueki, H., Kojima-Ishiyama, M., Carninci, P., Hayashizaki, Y. and Itoh, M. 2014. Detecting expressed genes using CAGE. *Methods in Molecular Biology* 1164, pp. 67–85.

Pachkov, M., Balwierz, P.J., Arnold, P., Ozonov, E. and van Nimwegen, E. 2013. SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. *Nucleic Acids Research* 41(Database issue), pp. D214-20.

1227 Pedersen, L., Hagedorn, P.H., Lindholm, M.W. and Lindow, M. 2014. A Kinetic Model Explains
1228 Why Shorter and Less Affine Enzyme-recruiting Oligonucleotides Can Be More Potent.
1229 *Molecular therapy. Nucleic acids* 3, p. e149.

1230 Portoso, M., Ragazzini, R., Brenčič, Ž., Moiani, A., Michaud, A., Vassilev, I., Wassef, M.,
1231 Servant, N., Sargueil, B. and Margueron, R. 2017. PRC2 is dispensable for HOTAIR-mediated
1232 transcriptional repression. *The EMBO Journal* 36(8), pp. 981–994.

1233 Proudfoot, N.J. 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II
1234 juggernaut. *Science* 352(6291), p. aad9926.

1235 Quek, X.C., Thomson, D.W., Maag, J.L.V., Bartonicek, N., Signal, B., Clark, M.B., Gloss, B.S.
1236 and Dinger, M.E. 2015. lncRNAdb v2.0: expanding the reference database for functional long
1237 noncoding RNAs. *Nucleic Acids Research* 43(Database issue), pp. D168-73.

1238 Quinn, J.J. and Chang, H.Y. 2016. Unique features of long non-coding RNA biogenesis and
1239 function. *Nature Reviews. Genetics* 17(1), pp. 47–62.

1240 Ransohoff, J.D., Wei, Y. and Khavari, P.A. 2018. The functions and unique features of long
1241 intergenic non-coding RNA. *Nature Reviews. Molecular Cell Biology* 19(3), pp. 143–157.

1242 Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H.,
1243 Helms, J.A., Farnham, P.J., Segal, E. and Chang, H.Y. 2007. Functional demarcation of active
1244 and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129(7), pp. 1311–
1245 1323.

1246 Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A.,
1247 Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W.,
1248 Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C. and
1249 Kellis, M. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539),
1250 pp. 317–330.

1251 Roux, B.T., Lindsay, M.A. and Heward, J.A. 2017. Knockdown of Nuclear-Located Enhancer
1252 RNAs and Long ncRNAs Using Locked Nucleic Acid GapmeRs. *Methods in Molecular Biology*
1253 1468, pp. 11–18.

1254 Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. and Hume, D.A. 2007.
1255 Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature*
1256 *Reviews. Genetics* 8(6), pp. 424–436.

1257 Shin, Y., Chang, Y.-C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S.,

1258 Haataja, M. and Brangwynne, C.P. 2018. Liquid nuclear condensates mechanically sense and
1259 restructure the genome. *Cell* 175(6), p. 1481–1491.e13.

1260 Sridhar, B., Rivas-Astroza, M., Nguyen, T.C., Chen, W., Yan, Z., Cao, X., Hebert, L. and Zhong,
1261 S. 2017. Systematic Mapping of RNA-Chromatin Interactions In Vivo. *Current Biology* 27(4), pp.
1262 602–609.

1263 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A.,
1264 Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P. 2005. Gene set
1265 enrichment analysis: a knowledge-based approach for interpreting genome-wide expression
1266 profiles. *Proceedings of the National Academy of Sciences of the United States of America*
1267 102(43), pp. 15545–15550.

1268 Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S.
1269 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*
1270 131(5), pp. 861–872.

1271 Thiel, D., Djurdjevac Conrad, N., Peschutter, R.X., Siebert, H. and Marsico, A. 2018. Identifying
1272 lncRNA-mediated regulatory modules via ChIA-PET network analysis. *Biorxiv*, p. 331256.
1273 Available at: https://doi.org/10.1101/331256 [Accessed: 22 April 2019].

1274 Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., Stern-Ginossar,
1275 N. and Ulitsky, I. 2016. A conserved abundant cytoplasmic long noncoding RNA modulates
1276 repression by Pumilio proteins in human cells. *Nature Communications* 7, p. 12209.

1277 Tonekaboni, S.A.M., Mazrooei, P., Kofia, V., Haibe-Kains, B. and Lupien, M. 2018. CREAM:
1278 Clustering of genomic REgions Analysis Method. *Biorxiv*, p. 222562.

1279 Ulitsky, I. 2016. Evolution to the rescue: using comparative genomics to understand long non-
1280 coding RNAs. *Nature Reviews. Genetics* 17(10), pp. 601–614.

1281 Volders, P.-J., Verheggen, K., Menschaert, G., Vandepoele, K., Martens, L., Vandesompele, J.
1282 and Mestdagh, P. 2015. An update on LNCipedia: a database for annotated human lncRNA
1283 sequences. *Nucleic Acids Research* 43(Database issue), pp. D174-80.

1284 Wiedenheft, B., Sternberg, S.H. and Doudna, J.A. 2012. RNA-guided genetic silencing systems
1285 in bacteria and archaea. *Nature* 482(7385), pp. 331–338.

1286 Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P. and
1287 Andrews, S. 2015. HiCUP: pipeline for mapping and processing Hi-C data. [version 1; peer
1288 review: 2 approved, 1 approved with reservations]. *F1000Research* 4, p. 1310.