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

Jordan A Ramilowski[1,2#], Chi Wai Yip[1,2#], Saumya Agrawal[1,2], Jen-Chien Chang[1,2], Yari Ciani[3], Ivan V Kulakovskiy[4], Mickaël Mendez[5], Jasmine Li Ching Ooi[2], John F Ouyang[6], Nick Parkinson[7], Andreas Petri[8], Leonie Roos[9], Jessica Severin[1,2], Kayoko Yasuzawa[1,2], Imad Abugessaisa[1,2], Altuna Akalin[10], Ivan V Antonov[11], Erik Arner[1,2], Alessandro Bonetti[2], Hidemasa Bono[12], Beatrice Borsari[13], Frank Brombacher[14], Chris JF Cameron[15], Carlo Vittorio Cannistraci[16], Ryan Cardenas[17], Melissa Cardon[1], Howard Chang[18], Josée Dostie[19], Luca Ducoli[20], Alexander Favorov[21], Alexandre Fort[2], Diego Garrido[13], Noa Gil[22], Juliette Gimenez[23], Reto Guler[14], Lusy Handoko[2], Jayson Harshbarger[2], Akira Hasegawa[1,2], Yuki Hasegawa[2], Kosuke Hashimoto[1,2], Norihito Hayatsu[1], Peter Heutink[24], Tetsuro Hirose[25], Eddie L Imada[26], Masayoshi Itoh[2,27], Bogumil Kaczkowski[1,2], Aditi Kanhere[17], Emily Kawabata[2], Hideya Kawaji[27], Tsugumi Kawashima[1,2], S. Thomas Kelly[1], Miki Kojima[1,2], Naoto Kondo[2], Haruhiko Koseki[1], Tsukasa Kouno[1,2], Anton Kratz[2], Mariola Kurowska-Stolarska[28], Andrew Tae Jun Kwon[1,2], Jeffrey Leek[26], Andreas Lennartsson[29], Marina Lizio[1,2], Fernando López-Redondo[1,2], Joachim Luginbühl[1,2], Shiori Maeda[1], Vsevolod J Makeev[21], Luigi Marchionni[26], Yulia A Medvedeva[11], Aki Minoda[1,2], Ferenc Müller[17], Manuel Muñoz-Aguirre[13], Mitsuyoshi Murata[1,2], Hiromi Nishiyori[1,2], Kazuhiro R Nitta[1,2], Shuhei Noguchi[1,2], Yukihiko Noro[2], Ramil Nurtdinov[13], Yasushi Okazaki[1,2], Valerio Orlando[30], Denis Paquette[19], Callum JC Parr[1], Owen JL Rackham[6], Patrizia Rizzu[24], Diego Fernando Sánchez Martinez[26], Albin Sandelin[31], Pillay Sanjana[17], Colin AM Semple[7], Youtaro Shibayama[1,2], Divya M Sivaraman[1,2], Takahiro Suzuki[1,2], Suzannah C Szumowski[2], Michihira Tagami[1,2], Martin S Taylor[7], Chikashi Terao[1], Malte Thodberg[31], Supat Thongjuea[2], Vidisha Tripathi[32], Igor Ulitsky[22], Roberto Verardo[3], Ilya Vorontsov[21], Chinatsu Yamamoto[2], Robert S Young[33], J Kenneth Baillie[7], Alistair RR Forrest[1,2,34], Roderic Guigó[13], Michael M Hoffman[35], Chung Chau Hon[1,2], Takeya Kasukawa[1,2], Sakari Kauppinen[8], Juha Kere[29], Boris Lenhard[9], Claudio Schneider[36], Harukazu Suzuki[1,2], Ken Yagi[1,2], FANTOM consortium, Michiel de Hoon[1,2*], Jay W Shin[1,2*], Piero Carninci[1,2*]

[*] 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, Yokohama, Kanagawa 230-0045 Japan
2. RIKEN Center for Life Science Technologies, Yokohama, Kanagawa 230-0045 Japan
3. Laboratorio Nazionale Consorzio Interuniversitario Biotecnologie (CIB), Trieste 34127 Italy
4. Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991 Russia
5. Department of Computer Science, University of Toronto, Toronto, ON M5S 1A1 Canada
6. Program in Cardiovascular and Metabolic Disorders, Duke NUS Graduate Medical School, Singapore 169857 Singapore
7. MRC Human Genetics Unit, University of Edinburgh, Edinburgh EH4 2XU UK

8. Center for RNA Medicine, Department of Clinical Medicine, Aalborg University, Copenhagen 9220 Denmark
9. Institute of Clinical Sciences MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, London W12 0NN UK
10. Berlin Institute for Medical Systems Biology, Max-Delbrük-Center for Molecular Medicine in the Helmholtz Association, Berlin 13125 Germany
11. Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Moscow 117312 Russia
12. Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima City 739-0046 Japan
13. Bioinformatics and Genomics, Centre for Genomic Regulation (CRG), Barcelona 8003 Catalonia, Spain
14. International Centre for Genetic Engineering and Biotechnology (ICGEB), University of Cape Town, Cape Town 7925 South Africa
15. School of Computer Science, McGill University, Montréal, Québec H3G 1Y6 Canada
16. Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Center for Molecular and Cellular Bioengineering (CMCB), Center for Systems Biology Dresden (CSBD), Cluster of Excellence Physics of Life (PoL), Department of Physics, Technische Universität Dresden, Dresden 01062 Germany
17. Institute of Cancer and Genomics Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT UK
18. Center for Personal Dynamic Regulome, Stanford University, Stanford, California 94305 USA
19. Department of Biochemistry and Rosalind & Morris Goodman Cancer Research Center, McGill University, Montréal, Québec H3G 1Y6 Canada
20. Institute of Pharmaceutical Sciences, Swiss Federal Institute of Technology, Zurich 8093 Switzerland
21. Computational Systems Biology, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 119991 Russia
22. Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100 Israel
23. Epigenetics and Genome Reprogramming Laboratory, IRCCS Fondazione Santa Lucia, Rome Italy
24. Genome Biology of Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), Tübingen 72076 Germany
25. Graduate School of Frontier Biosciences, Osaka University, Suita 565-0871 Japan
26. Department of Oncology, Johns Hopkins University, Baltimore, Maryland 21287 USA
27. RIKEN Preventive Medicine and Diagnosis Innovation Program (PMI), Saitama 351-0198 Japan
28. Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, Scotland UK
29. Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge 14157 Sweden
30. Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900 Kingdom of Saudi Arabia
31. Department of Biology and BRIC, University of Copenhagen, Denmark, Copenhagen N DK2200 Denmark
32. National Centre for Cell Science, Pune, Maharashtra 411007 India
33. Centre for Global Health Research, Usher Institute, University of Edinburgh, EH4 2XU UK
34. Harry Perkins Institute of Medical Research, QEII Medical Centre and Centre for Medical

Research, The University of Western Australia, Nedlands, Perth, WA 6009 Australia

35. Princess Margaret Cancer Centre, Toronto, ON M5G 1L7 Canada
36. Department of Medicine and Consorzio Interuniversitario Biotecnologie p.zle Kolbe 1 University of Udine, Udine 33100 Italy

36

Abstract

Long non-coding RNAs (lncRNAs) constitute the majority of transcripts in the mammalian genomes and yet, their functions remain largely unknown. We systematically knockdown 285 lncRNAs expression in human dermal fibroblasts and quantified cellular growth, morphological changes, and transcriptomic responses using Capped Analysis of Gene Expression (CAGE). Antisense oligonucleotides targeting the same lncRNA exhibited global concordance, and the molecular phenotype, measured by CAGE, recapitulated the observed cellular phenotypes while providing additional insights on the affected genes and pathways. Here, we disseminate the largest to-date lncRNA knockdown dataset with molecular phenotyping (over 1,000 CAGE deep-sequencing libraries) for further exploration and highlight 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 nucleotides 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 of lncRNAs (Xue et al. 2016), (Chu et al. 2015) 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; Ma et al. 2019), 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 FANTOM 5 computational analysis showed that 19,175 (out of 27,919) human lncRNA loci are functionally implicated (Hon et al. 2017). Yet, 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 ~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 (Liu et al. 2018) and a CRISPR activation study revealed ~0.11% lncRNAs to be important for drug resistance in melanoma (Joung et al. 2017). However, many of these studies target the genomic DNA, potentially perturbing the chromatin architecture, or focus on a single cellular assay, possibly missing other relevant functions and underlying molecular pathways.

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 pathways to be associated to lncRNAs even in the absence of an observable cellular phenotype. All data and analysis results are publicly available (see Data Access) and results can be interactively explored using our in-house portal https://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). Here, an unbiased selection of lncRNAs expressed in HDF was performed to choose 285 lncRNAs for functional interrogation (Methods; Supplemental Table S1, Fig. 1A-C). Using RNA-seq profiling of fractionated RNA, we annotated the lncRNA subcellular localization in the chromatin-bound (35%), nucleus-soluble (27%), or cytoplasm (38%) (Fig. 1D). We then designed a minimum of five non-overlapping antisense oligonucleotides (ASOs) against each lncRNA (Supplemental Methods; Supplemental Table S2; Fig. 1E,F) and transfected them individually using an

automated cell culture platform to minimize experimental variability (Fig. 1G). The overall knockdown efficiencies across 2,021 ASOs resulted in median value of 45.4%, and we could successfully knockdown 879 out of 2,021 (43.5%) ASOs (>40% knockdown efficiency in at least two primer pairs or >60% in one primer pair; Supplemental Table S2). 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 (Supplemental Fig. S1A–D).

A subset of lncRNAs 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 hours (Supplemental Table S3) and estimated their growth rate based on cell confluence measurements (Fig. 2A,B). First, we observed across all ASOs that changes in cell growth and morphological parameters were significantly correlated with knockdown efficiency (Supplemental Fig. S1E). Considering both successful knockdown and significant growth inhibition (Student's two-sided *t*-test FDR ≤ 0.05), 246 out of 879 ASOs (~28%) showed cellular phenotype (Fig. 2C, Table S3).

To assess globally whether the observed growth inhibition is lncRNA-specific, we used all 194 lncRNAs successfully targeted by at least two ASOs (Supplemental Fig. S2A) and found that ASOs targeting the same lncRNA were significantly more likely to have a concordant growth response than ASOs targeting different lncRNA (empirical p = 0.00037; Supplemental Methods; Supplemental Fig. S2B). However, different ASOs targeting the same lncRNA typically showed different effects on growth, possibly due to variable knockdown efficiencies, differences in targeted lncRNA isoforms, as well as off-target effects. To reliably identify target specific cellular phenotype, we applied conditional cutoffs based on the number of successful ASOs per each lncRNA (Supplemental Methods; Supplemental Fig. S2C) and identified 15/194 lncRNAs (7.7%) with growth phenotype (adjusted background less than 5%; Supplemental Fig. S2D). We validated *A1BG-AS1*, which was previously implicated in cell growth (Bai et al. 2019), *CATG00000089639*, *RP11-195F19.9*, and *ZNF213-AS1* by measuring the MKI67 proliferation protein marker upon knockdown with siRNAs and selected ASOs (Fig. 2D, Supplemental Fig. S2E).

In addition to cell growth, we also explored changes in cell morphology (Fig. 2E). Using a machine learning-assisted workflow (Methods), each cell was segmented and its morphological features representing various aspects of cell shapes and sizes were quantified (Carpenter et al. 2006) (Fig. 2F; Supplemental Table S3). As an example, knockdown of 14/194 lncRNAs (7.2%) affected the spindle-like morphology of fibroblasts, as indicated by a consistent decrease in their observed eccentricity without reducing the cell number, suggesting possible cellular transformation towards epithelial-like states. Collectively, we observed 59/194 lncRNAs (~30%) affecting cell growth and/or morphological parameters (Fig. 2G; Supplemental Table S3)

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

Next, we selected 340 ASOs with high knockdown efficiencies (mostly greater than 50%; median 71.4%) and sequenced 970 CAGE libraries to analyze 154 lncRNAs (Fig. 3A; Supplemental Table S4). To assess functional implications by individual ASOs, we performed differential gene expression, Motif Activity Response Analysis (MARA; (FANTOM Consortium et al. 2009), and Gene Set Enrichment Analysis (GSEA; (Subramanian et al. 2005); Fig3B-E), and compared them with cellular phenotype.

We globally observed significant knockdown-mediated transcriptomic changes (which generally correlated with KD efficiency; Supplemental Fig. S3A), with ~57% of ASOs showing at least 10 differentially expressed genes (FDR $\leq$ 0.05; abs($\log_2$FC) > 0.5). For 84 divergent-antisense lncRNAs (targeted by 186 independent ASOs) (Supplemental Methods), we found their partner gene to be generally unchanged (median abs($\log_2$FC) = ~0.13), with an exception of two significantly downregulated and three significantly upregulated genes (FDR $\leq$ 0.05; Supplemental Fig. S3B). We have, however, noticed common response in a large number of ASOs (~30-35% of all responding ASOs) such as down-regulation of cell-cycle related pathways, upregulated stress genes and pathways or altered cell metabolism and energetics (Supplemental Fig. S3C,D).

When comparing knockdown-mediated molecular and cellular response, we found that transcription factor motifs that promote cell growth, including 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; Supplemental Fig. S4A; Supplemental Table S6). Moreover, correlations between GSEA pathways (Fig. 3F; Supplemental Fig. S4B; Supplemental Table S6) and FANTOM5 co-expression clusters (Supplemental Fig. S4C) 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. We found that amongst 53 ASOs implicated in growth inhibition pathway based on the CAGE profiles, only 43% of them showed growth inhibition in the real-time imaging. This might suggest better sensitivity of transcriptomic profiling when detecting phenotypes as compared to live cell imaging methods, which are more prone to a delayed cellular response to the knockdown.

Additionally, morphological changes were reflected in the molecular phenotype assessed by CAGE (Supplemental Fig. S4D). 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).

Next, to globally assess whether individual ASO knockdowns lead to lncRNA-specific effects, we scaled the expression change of each gene across the whole experiment and compared differentially expressed genes (Fig. 3B) of all possible ASO pairs targeting the same lncRNA target versus different lncRNAs (Supplemental Methods; Supplemental Table S5). We found that the concordance of the same target group was significantly greater than that of the different target group (comparing the Jaccard indices across 10,000 permutations; Supplemental Fig. S5A), suggesting that ASO knockdowns are non-random and lead to more lncRNA specific effects than the non-targeting ASO pairs. Further, by requiring at least five common DEGs (FDR ≤ 0.05, abs($\log_2$FC) > 0.5, abs(Z-score) > 1.645) and ASO-pairs significantly above the non-targeting ASO pairs background ($p \leq 0.05$), we identified 16 ASO-pairs, targeting 13 lncRNAs,

exhibiting reproducible knockdown-mediated molecular responses in human dermal fibroblasts (Supplemental Fig. S5B). Corresponding GSEA pathways and MARA motifs of these 16 ASO-pairs are shown in Supplemental Figure S5C.

siRNA validation experiments

To evaluate whether the lncRNA-specific effects can be measured by other knockdown technologies, nine lncRNAs, with relatively mild growth phenotype, were subjected to siRNA knockdown. We noted that higher concordance was observed for ASO modality alone (Supplemental Fig. S5D). The observed discrepancies in the transcriptional response between ASO and siRNA-mediated knockdowns could be contributed by their mode of action and variable activities in different subcellular compartments. Next, a concordant response was found for (5/36) ASO-siRNA pairs targeting three lncRNAs (Supplemental Fig. S5E; Supplemental Table S5), enriched in the cytoplasm (*MAPKAPK5-AS1*), soluble nuclear fraction (*LINC02454*) and in the chromatin-bound fraction (*A1BG-AS1*). While we cannot completely exclude the technical artefacts of each technology, concordant cellular response exhibited by using ASOs alone suggests that lncRNA, in part, are essential regulatory elements in cells. Yet, our study generally warrants a careful assessment of specific findings from different knockdown technologies, including CRISPR-inhibition, and demonstrates a requirement of using multiple replicates in a given target per each modality.

***ZNF213-AS1* is associated with cell growth and migration.**

Extensive molecular and cellular phenotype data for each ASO knockdown can be explored using our portal https://fantom.gsc.riken.jp/zenbu/reports/#FANTOM6. As an example of an lncRNA associated with cell growth and morphology (Fig. 2G), we showcase *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-bound fraction. 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. 4A). 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 MKI67 marker staining (Fig. 2D; Supplemental Tables S7), the molecular phenotype revealed suppression of GSEA pathways related to cell growth, as well as to cell proliferation, motility, and extracellular structure organization (Fig. 4B), and consistent in two ASOs downregulation of related motifs, for example, EGR1, EP300, SMAD1..7,9 (Fig. 4C).

As cell motility pathways were affected by the knockdown, we tested whether *ZNF213-AS1* could influence cell migration. Based on the wound-closure assay after transient cell growth inhibition (mitomycin-C and serum starvation), we observed a substantial reduction of wound closure rate (~40% over a 24-hour period) in the *ZNF213-AS1* depleted HDFs (Fig. 4D,E). The reduced wound healing rate should thus mainly reflect reduced cell motility, further confirming affected motility pathways predicted by the molecular phenotype.

As these results indicated a potential role of *ZNF213-AS1* in cell growth and migration, we used FANTOM CAT Recount 2 atlas (Imada et al. 2020), 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 (Supplemental Fig. S6A). 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 (Supplemental Fig. S6B–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 a weak association with poor prognostic outcome but the results were not significant; (Supplemental Fig. S6F,G).

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

Next, 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.

5A). While the two genomic loci occupy Chromosome 6 and are 650kb away, Hi-C analysis (Supplemental Methods; Supplemental Fig. S7) showed that they are located within the same topologically associated domain (TAD) and spatially co-localized (Fig. 5B). Moreover, chromatin-enrichment and single molecule RNA-FISH of *RP11-398K22.12* (Fig. 5C) suggested its highly localized *cis*-regulatory role.

In FANTOM5 (Hon et al. 2017), expression levels of *RP11-398K22.12*, *KCNQ5* and *CATG00000088862.1* were enriched in brain and nervous system samples, while GTEx (GTEx Consortium 2015) showed their highly-specific expression in the brain, particularly in the cerebellum and the cerebellar hemisphere (Fig. 5D). GTEx data also showed that expression of *RP11-398K22.12* with *KCNQ5* and *CATG00000088862.1* was highly correlated across neuronal tissues (Fig. 5E,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.

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. 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 (Supplemental Fig. S3B). Recent studies suggest that transcription of lncRNA loci that do not overlap with other transcription units 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 ~7.7% 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. Further, when applying image-based analysis, we found that lncRNAs affect cell morphologies (Fig. 2G), which has not been so far thoroughly explored.

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, the number of

overlapping differentially expressed genes between 2 ASOs of the same lncRNA targets, indicated that 10.9% of lncRNAs exert a reproducible regulatory function in HDF.

Although the features of selected lncRNAs being generally similar to those of other lncRNAs expressed in HDF (Fig. 1B-D), the cell type specific nature of lncRNAs and the relatively small sampling size (119 lncRNAs with knockdown transcriptome profiles) used in our study may not fully represent the whole extent of lncRNA in other cell types. 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*). Although the exact molecular mechanisms of *RP11-398K22.*12 are not yet fully understood, its potential role in HDF suggests that lncRNAs may be functionally relevant across multiple tissues in spite of the cell-type-specific expression of lncRNAs.

Further, we used siRNA technology to knockdown lncRNA targets as a method for independent validation. When comparing the transcriptomes perturbed by ASOs and siRNAs, concordance was observed only for 3 out of 9 lncRNAs. This discrepancy is likely due to different modes of actions of the two technologies. While ASOs invoke RNase H-mediated cleavage, primarily active in the nucleus, the siRNAs use RNA-inducing silencing complex (RISC) mainly active in the cytoplasm. LncRNAs are known to function in specific subcellular compartments (Chen 2016) and their maturity, secondary structures, isoforms and functions could be vastly different across compartments (Johnsson et al. 2013). Since the majority functional lncRNA are reported to be inside the nucleus (Palazzo and Lee 2018), (Sun et al. 2018), ASO-mediated knockdowns, which mainly target nuclear RNAs, are generally more suitable for functional screenings of our lncRNA (62% found in the nuclear compartment). Besides, the dynamics of secondary effects mediated by different levels of knockdown from different technologies are likely to be observed as discordance when considering the whole transcriptome, where this kind of discordance has been reported previously (Stojic et al. 2018). In contrast, in the MKI67 assay where only a single feature such as growth phenotype is assayed, siRNA knockdown revealed higher reproducibility with ASO knockdown. This suggested that the growth phenotype might be triggered by different specific pathways in ASO- and siRNA- knockdowns.

Previous studies suggests that lncRNAs regulate gene expression in *trans* epigenetically, via direct or indirect interaction with regulators such as DNMT1 (Di Ruscio et al. 2013) or 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 a given 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 may affect 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).

In summary, our study highlights the functional importance of lncRNAs regardless of their expression, localization and conservation levels. Molecular phenotyping is a powerful and generally more sensitive to knockdown mediated changes platform to reveal the functional relevance of lncRNAs that cannot be observed based on the cellular phenotypes alone. With additional molecular profiling techniques, such as RNA duplex maps in living cells to decode common structural motifs (Lu et al. 2016), and Oxford Nanopore Technology (ONT) to annotate the full-length variant isoforms of lncRNA (Hardwick et al. 2019), structure-to-functional relationship of lncRNAs may be elucidated further in the future.

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). From this merged assembly, there were ~2,000 lncRNAs robustly expressed in HDF (TPM ≥ 1). However, we selected lncRNA knockdown targets in an unbiased manner to broadly cover various types of lncRNAs (TPM ≥ 0.2). 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 locus for design of antisense oligonucleotides (ASOs). In total, 285 lncRNA loci were chosen for ASO suppression. Additional controls (*NEAT1*, protein coding genes Supplemental Table S1) were added including *MALAT1* as an experimental control. For details please refer to the Supplemental Methods.

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 details please refer to the Supplemental Methods.

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 hours prior to the transfection. A final concentration of 20 nM ASO and 2 µl lipofectamine RNAiMAX (Thermo Fisher Scientific) were mixed in 200 µl Opti-MEM (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 hours 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 was extracted from the lysate for real time quantitative RT-PCR (Supplemental Methods).

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 hours 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 Opti-MEM (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 hours 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. Negative growth rate was derived when cells shrink and/or detach. As these rates of cell depletion could not be normalized by the rate of growth, negative values were maintained to indicate severe growth inhibition. Student's *t*-test was performed between the growth rate of the duplicated samples and the 6 NC_A controls, assuming equal variance.

426

427 **Cell morphology quantification**

428 For each transfection, representative phase-contrast image at a single time point was exported
429 from the Incucyte time-series. These raw images were first transformed to probability maps of
430 cells by pixel classification using ilastik (1.3.2) (Berg et al. 2019). The trained model was then
431 applied to all images where the predicted probability maps of cells (grey scale, 16 bits tiff
432 format) were subsequently used for morphology quantification in CellProfiler (3.1.5) (Carpenter
433 et al. 2006). For details please refer to the Supplemental Methods.

434

435 **MKI67 staining upon lncRNA knockdown**

436 For the selected four lncRNA targets showing >25% growth inhibition, we used two siRNAs and
437 ASOs with independent sequences. The transfected cells were fixed by adding pre-chilled 70%
438 ethanol and incubated in -20°C. The cells were washed by FACS buffer (2% FBS in PBS, 0.05%
439 NaN3) twice. FITC-conjugated MKI67 (20Raj1, eBioscience) was applied to the cells and
440 subjected to flow cytometric analysis. Knockdown efficiency by siRNA was determined by real-
441 time quantitative RT-PCR using the same 3 primer pairs as for ASO knockdown efficiency. For
442 details please refer to the Supplemental Methods.

443

444 **Wound closure assay**

445 The HDF cells were transfected by 20nM ASO as described earlier in 12-well plates. The cells
446 were re-plated at 24 hours post-transfection into a 96-well ImageLock plate (Essen BioScience)
447 at a density of 20,000 cells per well. At 24 hours after seeding, cells form a spatially uniform
448 monolayer with 95-100% cell confluence. The cells were incubated with 5 µg/mL mitomycin-C
449 for 2 hours to inhibit cell division. Then, medium was refreshed and a uniform scratch was
450 created in each well by the WoundMaker™(Essen BioScience). The closure of the wound was
451 monitored by IncuCyte® live-cell imaging system (Essen Bioscience) every 2 hours for 24
452 hours. The RNA was harvested after the assay for real-time quantitative RT-PCR. For details
453 please refer to the Supplemental Methods.

454

455 ***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). For details please refer to the Supplemental Methods.

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). For details please refer to the Supplemental Methods.

Data Access

All raw and processed sequencing data generated in this study have been submitted to the DNA Data Bank of Japan (DDBJ; https://www.ddbj.nig.ac.jp/) under accession numbers (DRA008311, DRA008312, DRA008436, DRA008511) or can be accessed through the FANTOM6 project portal https://fantom.gsc.riken.jp/6/datafiles. The analysis results can be downloaded from https://fantom.gsc.riken.jp/6/suppl/Ramilowski_et_al_2020/data/ and interactively explored using our in-house portal https://fantom.gsc.riken.jp/zenbu/reports/#FANTOM6/.

Acknowledgements

General: We would also like to thank Linda Kostrencic, Hiroto Atsui, Emi Ito, Nobuyuki Takeda, Tsutomu Saito, Teruaki Kitakura, Yumi Hara, Machiko Kashiwagi, Masaaki Furuno at RIKEN Yokohama for assistance in arranging collaboration agreements, ethics applications, computational infrastructure and the FANTOM6 meetings. The authors wish to acknowledge RIKEN GeNAS for generation and sequencing of the CAGE libraries and subsequent data processing.

Funding: FANTOM6 was made possible by a Research Grant for RIKEN Center for Life Science Technology, Division of Genomic Technologies (CLST DGT) and RIKEN Center for Integrative Medical Sciences (IMS) from MEXT, Japan. I.V.K. and I.E.V. were supported by RFBR 18-34-20024, B.B. is supported by the fellowship 2017FI_B00722 from the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement (Generalitat de Catalunya) and the European Social Fund (ESF), A.V.F. was supported by NIH P30 CA006973 and RFBR

17-00-00208, D.G.M. is supported by a "la Caixa"-Severo Ochoa pre-doctoral fellowship (LCF/BQ/SO15/52260001), E.L.I and L.M. were supported by NIH National Cancer Institute Grant R01CA200859 and DOD award W81XWH-16-1-0739, M.K.S. was supported by Versus Arthritis UK 20298, A.L. was supported by the Swedish Cancer society, The Swedish research council, the Swedish Childhood Cancer fund, Radiumhemmets forsknigsfonder, V.J.M. was supported by the Russian Academy of Sciences Project 0112-2019-0001, Y.A.M. was supported by RSF grant 18-14-00240, A.S. was supported by Novo Nordisk Foundation, Lundbeck Foundation, Danish Cancer Society, Carlsberg Foundation, Independent Research Fund Denmark, A.R.R.F. is currently supported by an Australian National Health and Medical Research Council Fellowship APP1154524, M.M.H. was supported by Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-3948), C.S. was supported by CIB from grants MIUR n.974,CMPT177780. J.L. was supported by Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Foreign Researchers. C.J.C.P. was supported by RIKEN Special Post-Doctoral Research (SPDR) fellowship.

Author Information

Competing interests - all authors declare no competing interest

References

Ambasudhan, R., Talantova, M., Coleman, R., et al. 2011. Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. *Cell Stem Cell* 9(2), pp. 113–118.

Bai, J., Yao, B., Wang, L., et al. 2019. lncRNA A1BG-AS1 suppresses proliferation and invasion of hepatocellular carcinoma cells by targeting miR-216a-5p. *Journal of Cellular Biochemistry* 120(6), pp. 10310–10322.

Berg, S., Kutra, D., Kroeger, T., et al. 2019. ilastik: interactive machine learning for (bio)image analysis. *Nature Methods* 16(12), pp. 1226–1232.

Carninci, P., Kasukawa, T., Katayama, S., et al. 2005. The transcriptional landscape of the mammalian genome. *Science* 309(5740), pp. 1559–1563.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., et al. 2006. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology* 7(10), p. R100.

Carrieri, C., Cimatti, L., Biagioli, M., et al. 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491(7424), pp. 454–457.

Chen, L.-L. 2016. Linking long noncoding RNA localization and function. *Trends in Biochemical Sciences* 41(9), pp. 761–772.

Cho, S.W., Xu, J., Sun, R., et al. 2018. Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* 173(6), pp. 1398-1412.e22.

Chu, C., Zhang, Q.C., da Rocha, S.T., et al. 2015. Systematic discovery of Xist RNA binding proteins. *Cell* 161(2), pp. 404–416.

Collado-Torres, L., Nellore, A., Kammers, K., et al. 2017. Reproducible RNA-seq analysis using recount2. *Nature Biotechnology* 35(4), pp. 319–321.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., et al. 2013. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503(7476), pp. 371–376.

Engreitz, J.M., Haines, J.E., Perez, E.M., et al. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539(7629), pp. 452–455.

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

FANTOM Consortium, Suzuki, H., Forrest, A.R.R., et al. 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., et al. 2015. Hierarchical folding and reorganization of

538 chromosomes are linked to transcriptional changes in cellular differentiation. *Molecular Systems*
539 *Biology* 11(12), p. 852.

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

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

545 Gupta, R.A., Shah, N., Wang, K.C., et al. 2010. Long non-coding RNA HOTAIR reprograms
546 chromatin state to promote cancer metastasis. *Nature* 464(7291), pp. 1071–1076.

547 Guttman, M., Donaghey, J., Carey, B.W., et al. 2011. lincRNAs act in the circuitry controlling
548 pluripotency and differentiation. *Nature* 477(7364), pp. 295–300.

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

551 Hardwick, S.A., Bassett, S.D., Kaczorowski, D., et al. 2019. Targeted, High-Resolution RNA
552 Sequencing Of Non-Coding Genomic Regions Associated With Neuropsychiatric Functions.
553 *Biorxiv*, p. 539882.

554 Hon, C.-C., Ramilowski, J.A., Harshbarger, J., et al. 2017. An atlas of human long non-coding
555 RNAs with accurate 5' ends. *Nature* 543(7644), pp. 199–204.

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

558 Imada, E.L., Sanchez, D.F., Collado-Torres, L., et al. 2020. Recounting the FANTOM
559 cageassociated transcriptome. *Genome Research*.

560 Iyer, M.K., Niknafs, Y.S., Malik, R., et al. 2015. The landscape of long noncoding RNAs in the
561 human transcriptome. *Nature Genetics* 47(3), pp. 199–208.

562 Johnsson, P., Ackley, A., Vidarsdottir, L., et al. 2013. A pseudogene long-noncoding-RNA
563 network regulates PTEN transcription and translation in human cells. *Nature Structural &*
564 *Molecular Biology* 20(4), pp. 440–446.

565 Joung, J., Engreitz, J.M., Konermann, S., et al. 2017. Genome-scale activation screen identifies a
566 lncRNA locus regulating a gene neighbourhood. *Nature* 548(7667), pp. 343–346.

567 Kalluri, R. 2016. The biology and function of fibroblasts in cancer. *Nature Reviews. Cancer*
568 16(9), pp. 582–598.

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

571 Lieberman-Aiden, E., van Berkum, N.L., Williams, L., et al. 2009. Comprehensive mapping of
572 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., et al. 2017. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 355(6320).

Liu, Y., Cao, Z., Wang, Y., et al. 2018. Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nature Biotechnology*.

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

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

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

Luo, S., Lu, J.Y., Liu, L., et al. 2016. Divergent lncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells. *Cell Stem Cell* 18(5), pp. 637–652.

Lu, Z., Zhang, Q.C., Lee, B., et al. 2016. RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* 165(5), pp. 1267–1279.

Ma, L., Cao, J., Liu, L., et al. 2019. LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Research* 47(D1), pp. D128–D134.

Mondal, T., Subhash, S., Vaid, R., et al. 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.

Palazzo, A.F. and Lee, E.S. 2018. Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs. *Frontiers in genetics* 9, p. 440.

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

Quek, X.C., Thomson, D.W., Maag, J.L.V., et al. 2015. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Research* 43(Database issue), pp. D168-73.

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

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

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., et al. 2015. Integrative

analysis of 111 reference human epigenomes. *Nature* 518(7539), pp. 317–330.

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

Sridhar, B., Rivas-Astroza, M., Nguyen, T.C., et al. 2017. Systematic Mapping of RNA-Chromatin Interactions In Vivo. *Current Biology* 27(4), pp. 602–609.

Stojic, L., Lun, A.T.L., Mangei, J., et al. 2018. Specificity of RNAi, LNA and CRISPRi as loss-of-function methods in transcriptional analysis. *Nucleic Acids Research* 46(12), pp. 5950–5966.

Subramanian, A., Tamayo, P., Mootha, V.K., et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102(43), pp. 15545–15550.

Sun, Q., Hao, Q. and Prasanth, K.V. 2018. Nuclear long noncoding rnas: key regulators of gene expression. *Trends in Genetics* 34(2), pp. 142–157.

Takahashi, K., Tanabe, K., Ohnuki, M., et al. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131(5), pp. 861–872.

Tichon, A., Gil, N., Lubelsky, Y., et al. 2016. A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nature Communications* 7, p. 12209.

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

Volders, P.-J., Verheggen, K., Menschaert, G., et al. 2015. An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Research* 43(Database issue), pp. D174-80.

Xue, Z., Hennelly, S., Doyle, B., et al. 2016. A G-Rich Motif in the lncRNA Braveheart Interacts with a Zinc-Finger Transcription Factor to Specify the Cardiovascular Lineage. *Molecular Cell* 64(1), pp. 37–50.

Figure legends

Figure 1. Selection of lncRNA targets, their properties and the study overview. (*A*) CAGE expression levels at $\log_2$TPM (tags per million) and human dermal fibroblasts (HDF) specificity of lncRNAs in the FANTOM CAT catalog (Hon, et al., Nature 2017; *N* = 62,873; grey), lncRNAs expressed in HDF (*N* = 6,125; blue) and targeted lncRNAs (*N* = 285; red). The dashed vertical line indicates most lowly expressed lncRNA target (~0.2 TPM). (*B*) Gene conservation levels of lncRNAs in the FANTOM CAT catalog (grey), lncRNAs 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 p = 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. (*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 the distribution of all lncRNAs expressed in HDF. (*E*) Example of *ZNF213-AS1* loci showing transcript model, CAGE and RNA-seq signal along with targeting ASOs. (*F*) Number of ASOs for target lncRNAs and controls used in the experiment. (*G*) Schematics of the study.

Figure 2. Cell growth and morphology assessment. (*A*) Selected example (*PTPRG1-AS1*) showing the normalized growth rate estimation using a matching 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 averaged) 252 ASOs targeting lncRNAs with successful knockdown (KD) and growth phenotype (blue) consistent in 2 replicates (FDR < 0.05 as compared to matching NC_A; 246 ASOs inhibited growth), 627 ASOs targeting lncRNAs with successful KD (purple), 270 negative control (NC_A) samples (grey) and 90 mock-transfected cells (lipofectamine only) samples (yellow). (*D*) MKI67 staining (growth inhibition validation) for four selected lncRNA targets after siRNA and ASOs suppression. (*E*) IncuCyte® cell images of selected distinct cell morphologies changes upon an lncRNA KD. (*F*) An overview of cell morphology imaging processing pipeline using a novel lncRNA target, *CATG000089639.1*, as an example. (*G*) lncRNAs (*N* = 59) significantly (FDR < 0.05) and consistently (after adjusting for the number of successfully targeting ASOs) affecting cell growth (*N* = 15) and cell morphologies (*N* = 44).

Figure 3. CAGE predicts cellular phenotypes. (*A*) RT-qPCR knockdown (KD) efficiency for 2,021 ASO-transfected samples (targeted lncRNAs only). Grey dashed line indicates 50% KD efficiency generally required for CAGE selection. Purple dashed lines indicate median KD efficiency (71.5%) for 375 ASOs selected for CAGE sequencing. After quality control, 340 ASOs targeting lncRNAs were included for further analysis. (*B*) Distribution of significantly differentially expressed genes (up-regulated: FDR < 0.05, Z-score > 1.645, log2FC > 0.5 and down-regulated: FDR < 0.05, Z-score < -1.645, log2FC < -0.5) across all 340 ASOs. (*C*) Motif

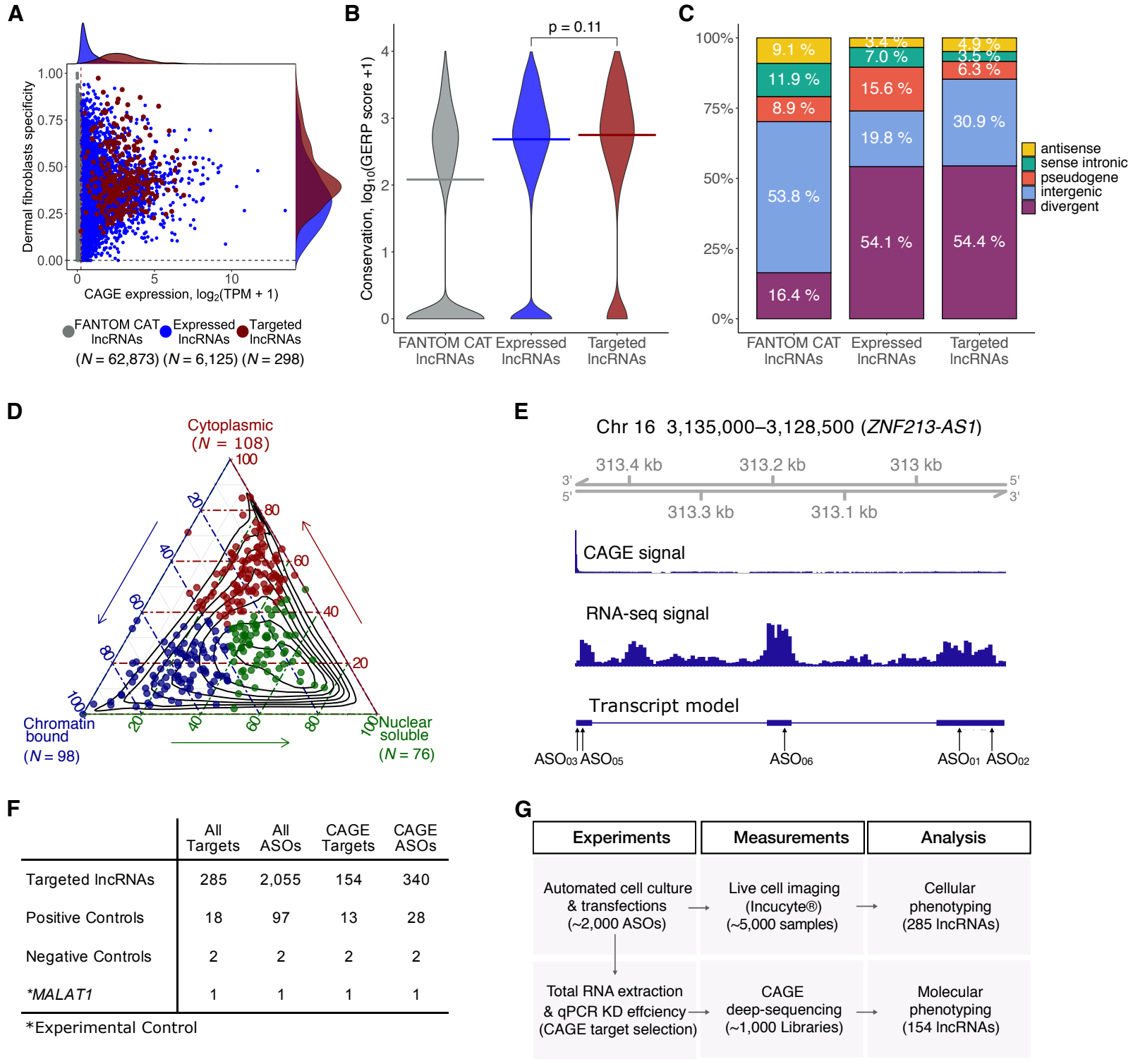
673 Response Activity Analysis (MARA) across 340 ASOs. Scale indicates Z-score of the relative
674 motif activity (the range was set to abs(Z-score) =< 5 for visualization purposes). (*D*) Correlation
675 between normalized growth rate and motif activities across 340 ASOs targeting lncRNAs with
676 highlighted examples. Motifs sizes shown are scaled based on the HDF expression of their
677 associated TFs (range 1 to ~600 TPM). (*E*) Enriched biological pathways across 340 ASOs.
678 Scale indicates GSEA enrichment value calculated as -log10(p) × sign(NES). (*F*) same as in D,
679 but for selected GSEA pathways. Pathways sizes are scaled based on the number of associated
680 genes.

681 **Figure 4.** *ZNF213-AS1* regulates cell growth, migration and proliferation. (*A*) Normalized
682 growth rate across four distinct ASOs (in duplicates) targeting *ZNF213-AS1* as compared to six
683 negative control samples (shown in grey). (*B*) Enrichment of biological pathways associated with
684 growth, proliferation, wound healing, migration and adhesion for ASO_02 and ASO_05. (*C*)
685 Most consistently down- and upregulated transcription factor binding motifs including those for
686 transcription factors known to modulate growth, migration and proliferation such as for example
687 EGR family, EP300, GTF2I. (*D*) Transfected, re-plated and mitomycin-C (5 µg/mL)-treated
688 HDF cells were scratched and monitored in the IncuCyte® imaging system. Relative wound
689 closure rate calculated during the 24 hours post-scratching shows 40-45% reduction for the two
690 targeting ASOs (ASO_02 (N = 10) and ASO_05 (N = 13)) as compared to NC_A transfection
691 controls (N = 33, shown in grey) and the representative images of wound closure assay 16 hours
692 post-scratching. (*E*) Knockdown efficiency measured by RT-qPCR after wound closure assay
693 (72 hours post-transfection) showing sustained suppression (65-90%) of *ZNF213-AS1*.

694 **Figure 5.** *RP11-398K22.12* down-regulates *KCNQ5* and *CATG00000088862.1* in cis. (*A*)
695 Changes in expression levels of detectable genes in the same topologically associated domain
696 (TAD) as *RP11-398K22.12* based on Hi-C analysis. Both *KCNQ5* and *CATG00000088862.1* are
697 down-regulated ($p < 0.05$) upon the knockdown of *RP11-398K22.12* by two independent ASOs
698 in CAGE analysis (left) as further confirmed with RT-qPCR (right). (*B*) (top) Representation of
699 the chromatin conformation in the 4Mb region proximal to the TAD containing *RP11-*
700 *398K22.12*, followed by the locus gene annotation, CAGE, RNA-seq and ATAC-seq data for
701 native HDFs. (bottom) Schematic diagram showing Hi-C predicted contacts of *RP11-398K22.12*
702 (blue) and *KCNQ5* (grey) (25Kb resolution, frequency >= 5) in HDF cells. Red line indicates
703 *RP11-398K22.12* and *KCNQ5* contact. (*C*) FISH image for *RP11-398K22.12* suggesting

proximal regulation. *TUG1* FISH image (suggesting trans regulation) is included as a comparison; (bar = 10 µm). (*D*) GTEx atlas across 54 tissues (*N* = 9,662 samples) shows relatively high expression levels of *RP11-398K22.12* in 13 distinct brain regions samples (highlighted). (*E*) Expression correlation for *RP11-398K22.12* and *KCNQ5* in 8 out of 13 distinct brain regions, as highlighted in D. (*F*) Expression correlation for *RP11-398K22.12* and *CATG00000088862.1* in 8 out of 13 distinct brain regions, as highlighted in D.

FIGURE 1



F

	All Targets	All ASOs	CAGE Targets	CAGE ASOs
Targeted lncRNAs	285	2,055	154	340
Positive Controls	18	97	13	28
Negative Controls	2	2	2	2
MALAT1	1	1	1	1

*Experimental Control

G

FIGURE 2

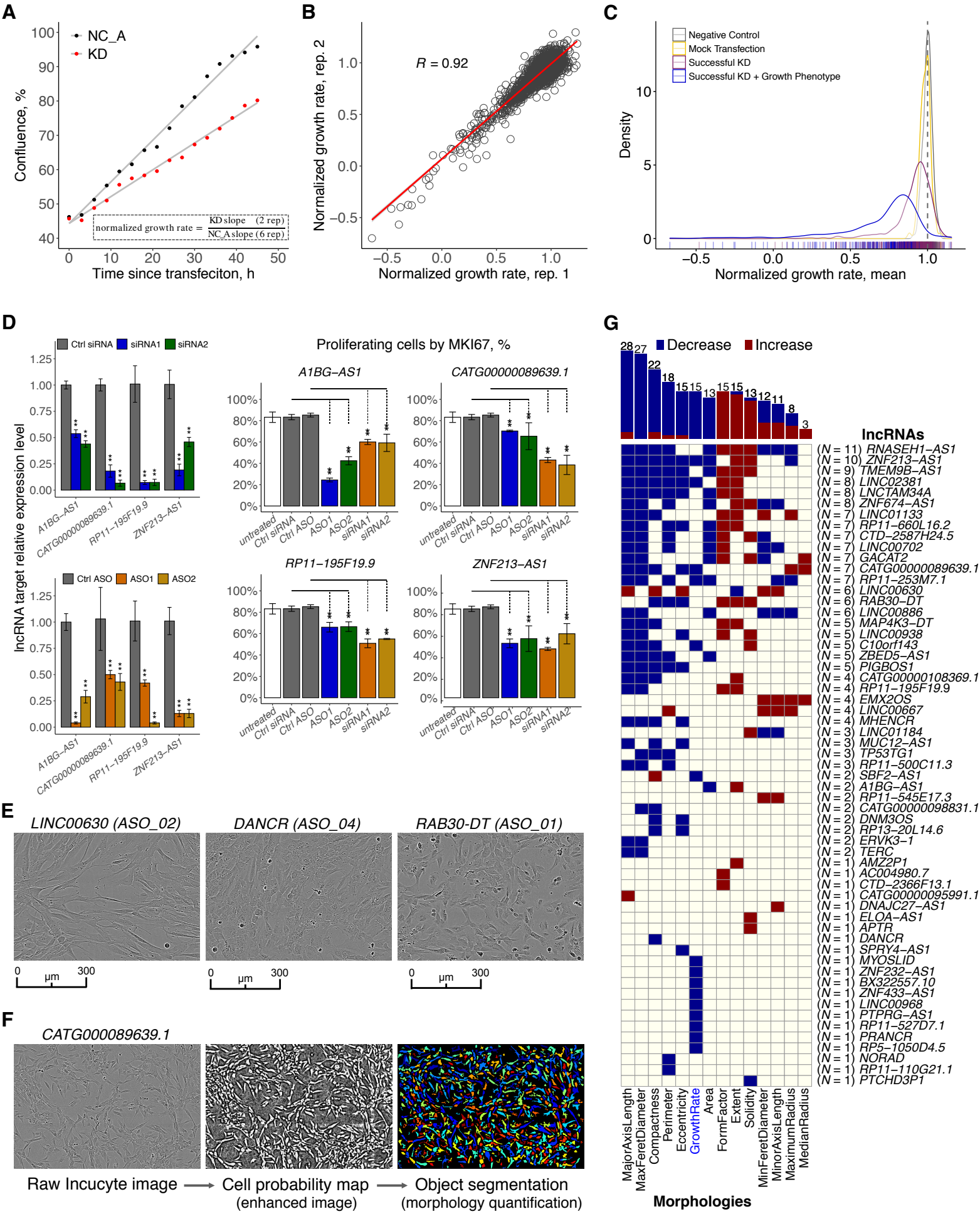

A
NC_A
KD
Confluence, %
Time since transfeciton, h
normalized growth rate = KD slope (2 rep) / NC_A slope (6 rep)
B
R = 0.92
Normalized growth rate, rep. 2
Normalized growth rate, rep. 1
C
Negative Control
Mock Transfection
Successful KD
Successful KD + Growth Phenotype
Density
Normalized growth rate, mean
D
Ctrl siRNA
siRNA1
siRNA2
lncRNA target relative expression level
Proliferating cells by MKI67, %
A1BG–AS1
CATG00000089639.1
RP11–195F19.9
ZNF213–AS1
Ctrl ASO
ASO1
ASO2
untreated
siRNA1
siRNA2
E
LINC00630 (ASO_02)
DANCR (ASO_04)
RAB30-DT (ASO_01)
0
µm
300
F
CATG000089639.1
Raw Incucyte image
Cell probability map
(enhanced image)
Object segmentation
(morphology quantification)
G
Decrease
Increase
lncRNAs
(N = 11) RNASEH1–AS1
(N = 10) ZNF213–AS1
(N = 9) TMEM9B–AS1
(N = 8) LINC02381
(N = 8) LNCTAM34A
(N = 8) ZNF674–AS1
(N = 7) LINC01133
(N = 7) RP11–660L16.2
(N = 7) CTD–2587H24.5
(N = 7) LINC00702
(N = 7) GACAT2
(N = 7) CATG00000089639.1
(N = 7) RP11–253M7.1
(N = 6) LINC00630
(N = 6) RAB30–DT
(N = 6) LINC00886
(N = 5) MAP4K3–DT
(N = 5) LINC00938
(N = 5) C10orf143
(N = 5) ZBED5–AS1
(N = 5) PIGBOS1
(N = 4) CATG00000108369.1
(N = 4) RP11–195F19.9
(N = 4) EMX2OS
(N = 4) LINC00667
(N = 4) MHENCR
(N = 3) LINC01184
(N = 3) MUC12–AS1
(N = 3) TP53TG1
(N = 3) RP11–500C11.3
(N = 2) SBF2–AS1
(N = 2) A1BG–AS1
(N = 2) RP11–545E17.3
(N = 2) CATG00000098831.1
(N = 2) DNM3OS
(N = 2) RP13–20L14.6
(N = 2) ERVK3–1
(N = 2) TERC
(N = 1) AMZ2P1
(N = 1) AC004980.7
(N = 1) CTD–2366F13.1
(N = 1) CATG00000095991.1
(N = 1) DNAJC27–AS1
(N = 1) ELOA–AS1
(N = 1) APTR
(N = 1) DANCR
(N = 1) SPRY4–AS1
(N = 1) MYOSLID
(N = 1) ZNF232–AS1
(N = 1) BX322557.10
(N = 1) ZNF433–AS1
(N = 1) LINC00968
(N = 1) PTPRG–AS1
(N = 1) RP11–527D7.1
(N = 1) PRANCR
(N = 1) RP5–1050D4.5
(N = 1) NORAD
(N = 1) RP11–110G21.1
(N = 1) PTCHD3P1
MajorAxisLength
MaxFeretDiameter
Compactness
Perimeter
Eccentricity
GrowthRate
Area
FormFactor
Extent
Solidity
MinFeretDiameter
MinorAxisLength
MaximumRadius
MedianRadius
Morphologies

FIGURE 3

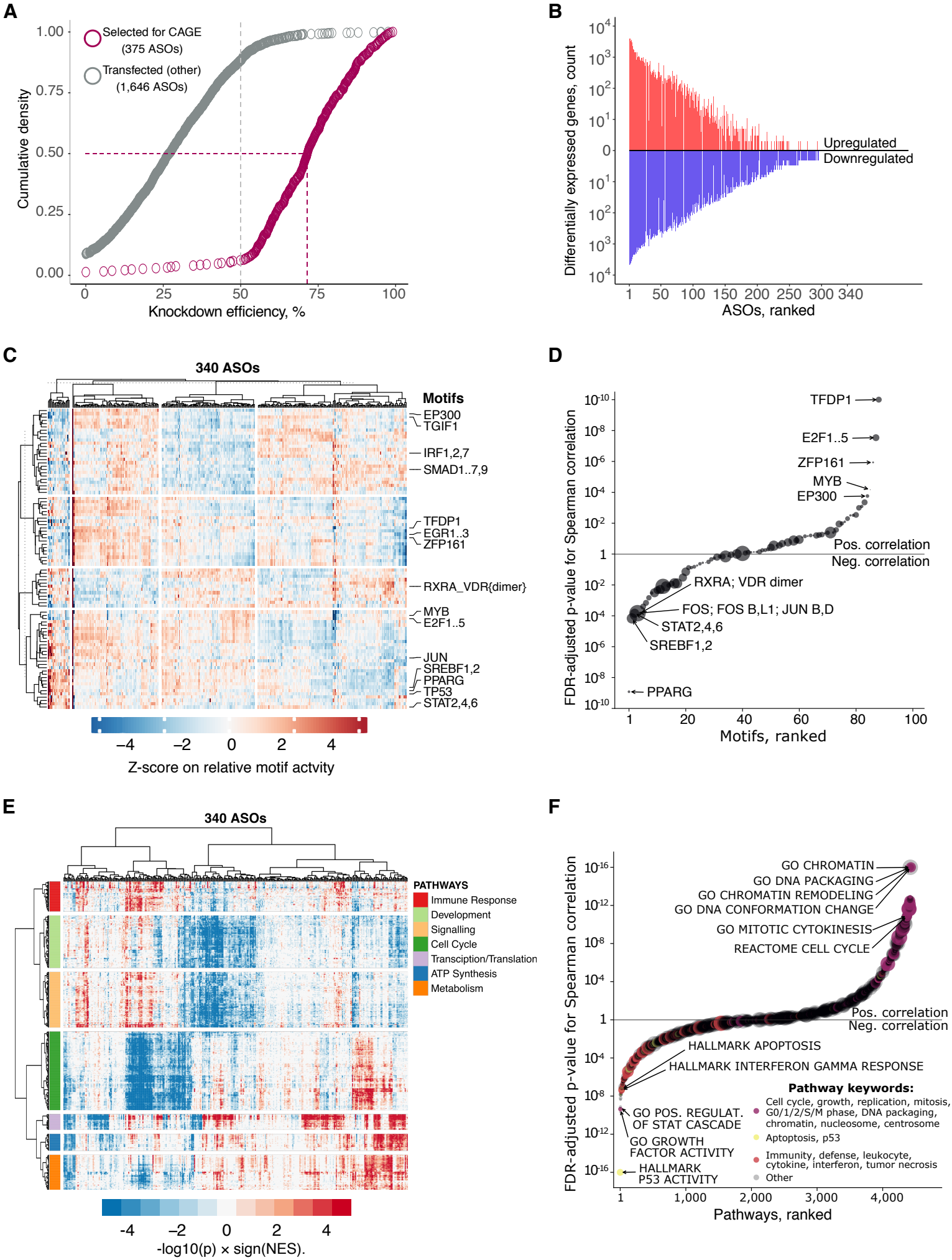


FIGURE 4

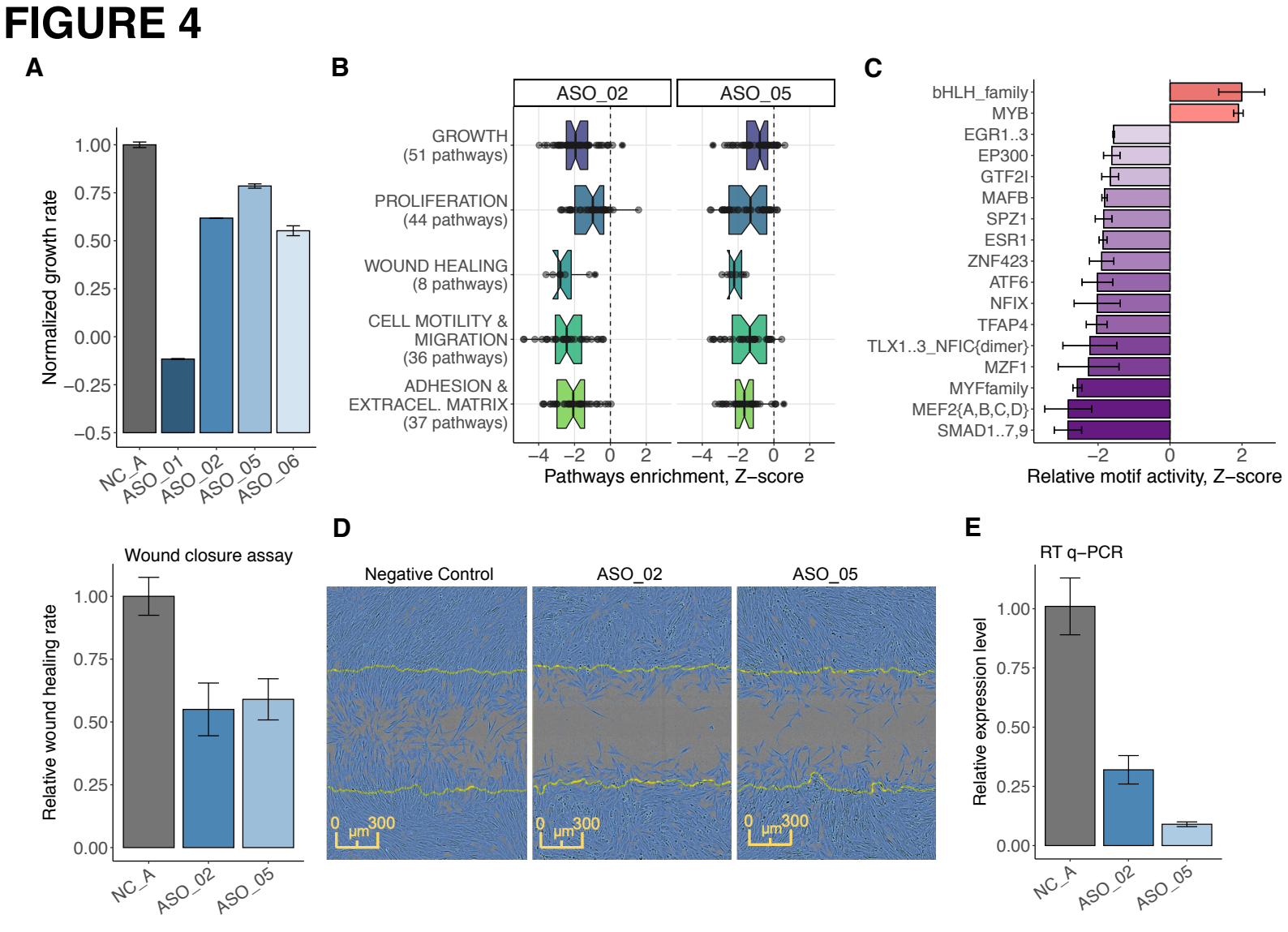
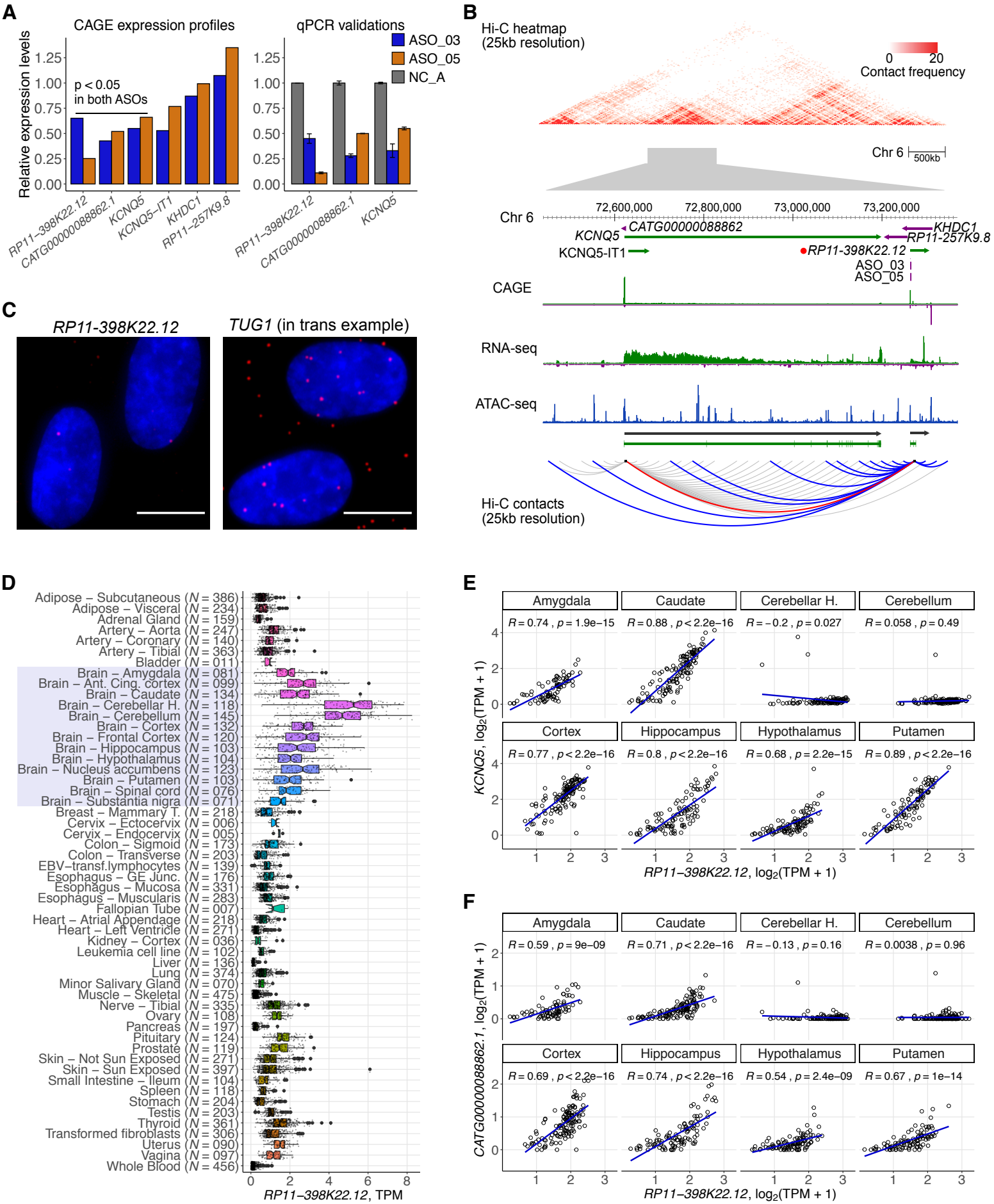

A
Normalized growth rate
1.00
0.75
0.50
0.25
0.00
−0.25
−0.5
NC_A
ASO_01
ASO_02
ASO_05
ASO_06
B
ASO_02
ASO_05
GROWTH
(51 pathways)
PROLIFERATION
(44 pathways)
WOUND HEALING
(8 pathways)
CELL MOTILITY &
MIGRATION
(36 pathways)
ADHESION &
EXTRACEL. MATRIX
(37 pathways)
−4 −2 0 2
Pathways enrichment, Z−score
C
bHLH_family
MYB
EGR1..3
EP300
GTF2I
MAFB
SPZ1
ESR1
ZNF423
ATF6
NFIX
TFAP4
TLX1..3_NFIC{dimer}
MZF1
MYFfamily
MEF2{A,B,C,D}
SMAD1..7,9
−2 0 2
Relative motif activity, Z−score
Wound closure assay
Relative wound healing rate
1.00
0.75
0.50
0.25
0.00
NC_A
ASO_02
ASO_05
D
Negative Control
ASO_02
ASO_05
0 300 μm
E
RT q−PCR
Relative expression level
1.00
0.75
0.50
0.25
0.00
NC_A
ASO_02
ASO_05

FIGURE 5


A
CAGE expression profiles
qPCR validations
ASO_03
ASO_05
NC_A
Relative expression levels
p < 0.05
in both ASOs
RP11–398K22.12
CATG00000088862.1
KCNQ5
KCNQ5–IT1
KHDC1
RP11–257K9.8
B
Hi-C heatmap
(25kb resolution)
Contact frequency
Chr 6
500kb
72,600,000
72,800,000
73,000,000
73,200,000
KCNQ5
CATG00000088862
KHDC1
RP11-257K9.8
KCNQ5-IT1
RP11-398K22.12
CAGE
RNA-seq
ATAC-seq
Hi-C contacts
(25kb resolution)
C
RP11-398K22.12
TUG1 (in trans example)
D
Adipose – Subcutaneous (N = 386)
Adipose – Visceral (N = 234)
Adrenal Gland (N = 159)
Artery – Aorta (N = 247)
Artery – Coronary (N = 140)
Artery – Tibial (N = 363)
Bladder (N = 011)
Brain – Amygdala (N = 081)
Brain – Ant. Cing. cortex (N = 099)
Brain – Caudate (N = 134)
Brain – Cerebellar H. (N = 118)
Brain – Cerebellum (N = 145)
Brain – Cortex (N = 132)
Brain – Frontal Cortex (N = 120)
Brain – Hippocampus (N = 103)
Brain – Hypothalamus (N = 104)
Brain – Nucleus accumbens (N = 123)
Brain – Putamen (N = 103)
Brain – Spinal cord (N = 076)
Brain – Substantia nigra (N = 071)
Breast – Mammary T. (N = 218)
Cervix – Ectocervix (N = 006)
Cervix – Endocervix (N = 005)
Colon – Sigmoid (N = 173)
Colon – Transverse (N = 203)
EBV–transf.lymphocytes (N = 139)
Esophagus – GE Junc. (N = 176)
Esophagus – Mucosa (N = 331)
Esophagus – Muscularis (N = 283)
Fallopian Tube (N = 007)
Heart – Atrial Appendage (N = 218)
Heart – Left Ventricle (N = 271)
Kidney – Cortex (N = 036)
Leukemia cell line (N = 102)
Liver (N = 136)
Lung (N = 374)
Minor Salivary Gland (N = 070)
Muscle – Skeletal (N = 475)
Nerve – Tibial (N = 335)
Ovary (N = 108)
Pancreas (N = 197)
Pituitary (N = 124)
Prostate (N = 119)
Skin – Not Sun Exposed (N = 271)
Skin – Sun Exposed (N = 397)
Small Intestine – Ileum (N = 104)
Spleen (N = 118)
Stomach (N = 204)
Testis (N = 203)
Thyroid (N = 361)
Transformed fibroblasts (N = 306)
Uterus (N = 090)
Vagina (N = 097)
Whole Blood (N = 456)
RP11–398K22.12, TPM
E
Amygdala
Caudate
Cerebellar H.
Cerebellum
Cortex
Hippocampus
Hypothalamus
Putamen
KCNQ5, log2(TPM + 1)
RP11–398K22.12, log2(TPM + 1)
F
CATG00000088862.1, log2(TPM + 1)
RP11–398K22.12, log2(TPM + 1)