CADTAD: CAncer Driver Topologically Associated Domains identify oncogenic and tumor suppressive lncRNAs

Ziyan Rao[1,2,#], Min Zhang[3,#,*], Shaodong Huang[1,2], Chenyang Wu[1,2], Yuheng Zhou[1], Weijie Zhang[4,5], Xia Lin[3], Dongyu Zhao[1,2,*]

1. Department of Biomedical Informatics, School of Basic Medical Sciences, Peking University, Beijing, 100191, China
2. State Key Laboratory of Vascular Homeostasis and Remodeling, Peking University, Beijing, 100191, China
3. Zhejiang Provincial Key Laboratory of Pancreatic Disease, the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310058, China
4. Zhejiang Provincial Key Laboratory of Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang 310058, China
5. Department of Orthopaedic Surgery, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang 310058, China

# These authors contributed equally to this work

*Corresponding to:
The Lead Contact: Dongyu Zhao: zhaodongyu@bjmu.edu.cn
Min Zhang: minzh@zju.edu.cn

Abstract

Cancer lncRNAs have been identified by both experimental and in silico methods. However, the current approaches for mining cancer lncRNAs are not sufficient and accurate. To deeply discover them, we focus on the core cancer driver lncRNAs (CDLs) which directly interact with cancer driver protein-coding genes. We investigated various aspects of cancer-related lncRNAs (CRLs), including their genomic locations, expression patterns, and their direct interactions with cancer driver protein-coding genes. We found that most CRLs located in cancer driver topologically associated domains (CDTs). Moreover, some CRLs showed a high tendency for co-expression and binding sites with cancer driver protein-coding genes. Utilizing these distinctive characteristics and integrating >4000 multi-omics data, we developed a pipeline CADTAD to unearth conserved candidate CDLs in pan-cancer, including 256 oncogenic lncRNAs, 177 tumor suppressive lncRNAs, and 75 dual-function lncRNAs, as well as some specific candidate CDLs in three individual cancer types and validated their cancer-related characteristics. Importantly, the function of 10 putative CDLs in prostate cancer was subsequently validated through cell studies. In light of these findings, our study offers a new perspective from the 3D genome to study the roles of lncRNAs in cancer. Furthermore, we provide a valuable set of potential lncRNAs that could deepen our understanding of the oncogenic mechanism of CDLs.

Introduction

Genetic[1] and epigenetic[2,3] features are commonly employed to discover cancer driver protein-coding genes, pivotal contributors to cancer progression[4]. However, the coding genes represent less than 2% of the entire genomic sequence, while most of the genome is transcribed into large number of noncoding RNAs, among which long noncoding RNAs (lncRNAs) stands out as a significant constitution. lncRNAs is a class of molecules exceeding 200 base pairs in length and typically has limited protein-coding potential[5]. They can engaged with DNA, RNA, or proteins to orchestrate the gene expression and chromatin architecture through various processes such as epigenetic regulation, transcription, or translation[6–11]. So it has been considered that lncRNAs may transcend the boundaries of genomics and epigenomics, assuming a supragenomic role[9], challenging the Central Dogma by exerting influence over DNA, RNA, and protein[11]. Numerous studies have reported the pivotal roles of the lncRNAs in cancer development[12–14]. Since subcellular localization of lncRNAs largely determines their function[15], most lncRNAs located in

55 the nucleus are strongly enriched within 3D proximity of their transcriptional loci[16] and can exert
56 cis-regulatory influence on neighboring genes. For example, lncRNA *CDKN2B-AS1* mediates the
57 epigenetic silencing of proximal genes *CDKN2A* and *CDKN2B* to increase cell proliferation[17,18].
58 Furthermore, certain lncRNAs in the nucleus can also affect distant genes expression, such as
59 *HOTAIR* repressing the expression of distant gene *HOXD* cluster by epigenetic silencing, resulting
60 in enhanced tumor metastasis[19,20]. On the other hand, some cytoplasmic lncRNAs can act as
61 microRNA sponges. For instance, *PTENP1* effectively sequesters specific microRNAs, thereby
62 preventing them from binding to *PTEN* and subsequently promoting its degradation[21]. Hence,
63 unraveling the function of lncRNAs is conducive to promoting our understanding of the initiation
64 and development of tumors and identifying the potential therapeutic targets for cancer.

65 The efficiency of experimental methodologies for identification cancer lncRNAs has
66 exhibited poor performance, while many bioinformatic approaches have predominantly yielded
67 cancer-related lncRNAs (CRLs) that have limited influence on tumorigenesis. The majority of
68 investigations about lncRNAs concentrated on individual molecules and their respective functions
69 in tumor[22,23]. Although this approach facilitates an in-depth exploration of specific functions of
70 lncRNA in tumors, it lacks the capacity for large-scale screening. Consequently, a limited subset
71 of lncRNAs can be examined in tumors due to insufficient detection, leaving the roles of numerous
72 other lncRNAs in tumors unexplored. In addition, large-scale experimental screening tools, such
73 as CRISPRi have been employed to silence all lncRNAs in order to assess their influence on tumor
74 development through cell proliferation[24]. While, this method is relatively reliable in identifying
75 lncRNAs with potential tumor-related effects, it is associated with substantial experimental costs.
76 Furthermore, the reproducibility of results may be compromised by variations in experimental
77 conditions and cellular environment. Moreover, a comprehensive understanding of the
78 mechanisms underlying the impact of lncRNAs on tumorigenesis remains unclear. One of the
79 bioinformatics methods entailed a meticulous manual curation process aimed at reviewing and
80 collating all lncRNAs associated with tumors. Subsequently, these curated lncRNAs were
81 integrated into CRL database[25,26]. It should be noted that this method relied exclusively on existing
82 literature and exhibited some leniency in the inclusion criteria for lncRNAs. The lncRNAs with
83 different expression patterns in tumor tissues compared with normal counterparts were considered
84 as CRLs. However, the only evidence of expression difference is not sufficient to prove the pivotal
85 roles of lncRNAs in tumorigenesis. An alternative approach was to utilize machine learning

86 techniques and combine multiple features to predict CRLs[27–32]. These predictive models only
87 identified CRLs reliant on training sets. Therefore, the current studies are not capable to
88 comprehensively and accurately discern cancer driver lncRNAs (CDLs), and not able to
89 distinguish between oncogenic and tumor suppressive lncRNA in silico.

90 In a wide range of eukaryotes, DNAs in the chromatin are folded into higher-ordered
91 structures. One of the fundamental structural entities of chromatin is the Topologically Associated
92 Domain (TAD), representing a discrete region of gene regulation. Within the confines of TAD,
93 DNA elements exhibited frequent interactions with one another, while inter-TAD chromatin
94 interactions tend to be relatively less pronounced. Perturbation of TAD boundaries can cause
95 profound consequences, including long-range transcriptional dysregulation and aberrant gene
96 expression[33,34]. Within the same TAD territory, lncRNA and protein-coding genes may frequently
97 engage in interactions and show similar functions.

98 Given the critical role of lncRNAs playing in cancer progression and the strong correlations
99 with the expression of proximate genes, acting as cis-regulatory elements [35], we link lncRNAs to
100 cancer driver protein-coding genes to deeply and accurately identify CDLs. CDLs are defined as
101 those lncRNAs capable of modulating the expression of cancer driver protein-coding genes,
102 thereby promoting tumorigenesis. These CDLs can be divided into oncogenic lncRNAs and tumor
103 suppressive lncRNAs according to their association with cancer driver protein-coding genes. In
104 this study, we first explored the interplay between CRLs and cancer driver protein-coding genes.
105 Notably, CRLs exhibited a pronounced propensity to localize within cancer driver topologically
106 associated domains (CDTs). Furthermore, our analysis of CRLs revealed enhanced co-expression
107 profile and exhibited an elevated degree of lncRNA-DNA interactions with cancer driver protein-
108 coding genes. We constructed a pipeline CADTAD and employed these insights to explore
109 potential CDL datasets across various cancer types. This endeavor yielded a subset of candidates
110 including 256 oncogenic lncRNAs, 177 tumor suppressive lncRNAs and 75 dual-function
111 lncRNAs for pan-cancer and putative CDLs in three individual cancer types (prostate cancer,
112 gastric cancer, and lung cancer). To validate the efficacy of identified CDLs, we utilized cancer
113 genomic, epigenomic, and phenotype data. Additionally, we conducted CRISPR-Cas9 knockout
114 experiments to confirm the function of 6 oncogenic lncRNAs and 4 tumor suppressive lncRNAs
115 specifically in prostate cancer.
116

Results

CDT is an important feature to identify CDLs.

We initially collected a set of 556 cancer driver protein-coding genes, which can be categorized as either oncogenes or tumor suppressor genes according to the Cancer Gene Census (CGC) database[36] v.94. Meanwhile, 426 and 822 CRLs were respectively obtained from Cancer LncRNA Census 2 (CLC2)[25] and Lnc2Cancer3.0[26]. We explored the relationship between these CRLs and cancer driver protein-coding genes within the context of cancer (See Methods).

It has been reported that lncRNA expression more extremely correlates with either neighboring (cis) or distal (trans) protein-coding genes compared to correlations of all-against-all protein-coding genes[5]. Therefore, our initial investigation focused on the associations between the expressions of CRLs and cancer-driver protein-coding genes. We found that CRLs obtained from both CLC2 and Lnc2Cancer exhibited significantly higher levels of co-expression with cancer driver protein-coding genes when compared to randomly selected lncRNAs or protein-coding genes (Fig. 1A). Furthermore, when considering the transcription start sites (TSSs) of cancer driver protein-coding genes or housekeeping genes as the center, we observed a notable predominance of CRLs in proximity to cancer driver protein-coding genes compared with their distribution around housekeeping genes across a series of extended genomic regions (Fig. 1B, Fig. S1A, see Methods). According to our findings, we proposed a scenario that lncRNAs may control their neighboring cancer driver protein-coding. Since TADs are characterized by high-density chromatin contacts and their alterations may cause aberrant gene expression[34], we consider TADs in cancer samples as functional units within which lncRNAs can potentially affect the expression of cancer driver protein-coding genes. TADs can be categorized into two groups: CDTs, which are defined as TADs that contain at least one cancer driver protein-coding gene, and the remaining TADs referred to as typical TADs. Our results revealed that the densities of CRLs were higher in CDTs than those in neighboring typical TADs or randomly selected typical TADs of similar length in prostate cancer cell line LNCaP (Fig. 1C-D, Fig. S1B-C, see Methods). In breast cancer cell line T47D, we observed the consistent results (Fig. 1E-F). These observations suggested that CRLs located close to cancer driver protein-coding genes, especially in the same TAD. Consistently, the known onco-lncRNA PVT1 is adjacent to the oncogene MYC in the same TAD in T47D cell line, and a previous study demonstrated that lncRNA PVT1 enhanced the stability of MYC, resulting in high protein level and promoted cancer progression, indicating a potent cis-regulatory

148 mechanism of PVT1 regulating MYC[23]. Also, lncRNA CCAT1 could regulate MYC
149 transcriptional expression through promoting loop interaction, and they are also found in the same
150 TAD in LNCaP cell line. However, lncRNA PCAT1, which overlapped with TAD boundary both
151 in LNCaP and T47D cell line, was co-amplified with MYC and induced MYC expression by a
152 post-transcriptional mechanism[12,13](Fig. 1G-H).
153 In addition, we observed almost half of the CDLs identified by a recent publication[37]
154 located in CDTs (Fig. 1I). It has been documented that the distance between lncRNA-mRNA
155 neighboring pairs could have an impact on coexpression[5]. Next, we divided CRLs into two
156 categories: I-CRLs, which were located in CDTs, and O-CRLs, which were discovered outside of
157 the CDT regions. When calculating the expression correlation between I/O-CRLs and cancer
158 driver protein-coding genes within CDTs, I-CRLs exhibited a higher degree of co-expression with
159 cancer driver protein-coding genes within the same TAD, as compared to O-CRLs (Fig. 1J, see
160 Methods). In addition, lncRNAs can play an important role in biological networks for
161 implementing specific biological functions, and there are many methods to infer lncRNA-mRNA
162 interaction networks based on their co-expression pattern[38]. So we employed the concept of
163 weighted gene co-expression network analysis (WGCNA) to construct a cancer-related lncRNA-
164 mRNA network[39,40] (Fig. S2A-B, see Methods). Then, we only focused on the interaction between
165 I/O-CRLs and the important genes (see Methods). While both I-CRLs and O-CRLs exhibited the
166 propensity to serve as hub-lncRNAs within the cancer-related lncRNA-mRNA network (Fig. 1K),
167 centrality score comparison and GSEA analysis revealed that I-CRLs showed a significantly
168 higher likelihood than O-CRLs to function as hub-lncRNAs (Fig. 1K-L). Thus, our findings
169 suggest that I-CRLs play more crucial roles in cancer-related processes.
170 Additionally, we examined the effectiveness of cancer lncRNAs identified by CDT and
171 expression correlation, and applying an expression correlation threshold of 0.5 was comparable to
172 mining lncRNAs with CDT (Fig. S3A-B).
173 All these results suggest that CDT is an important feature to discover CDLs, and lncRNAs
174 in CDTs may influence tumor suppressor or oncogene expression in the same CDT to enhance
175 cancer phenotypes.
176
177 **CDLs directly interact with cancer driver protein-coding genes.**

178 LncRNAs, which are located in proximity to protein-coding genes or associated regions,
179 play pivotal roles in the regulation of transcription and chromatin remodeling[5,8,9]. Given their
180 ability to interact with DNA, RNA, or protein to modulate many important cancer-related
181 processes[13], we are motivated to identify core driver lncRNAs in CDTs, which directly bind to
182 DNA sequences to influence the transcription of cancer driver protein-coding genes. So far,
183 lncRNAs have been discovered to directly bind DNAs in two ways: triplex and R-loops, which
184 may have different impacts on gene expression[41,42]. Triplex is a triple-helical structure in which
185 double-stranded DNA accommodates a single-stranded RNA in the major groove[43]. R-loop is a
186 three-stranded nucleic acid structure that forms RNA-DNA hybrid between a nascent guanine-rich
187 RNA transcript segment and a DNA template whilst leaving the non-template DNA strand in a
188 single-stranded conformation[44]. Here, we used Triplexator[43] and QmRLFS-finder[44] to predict
189 triplex and R-loops separately between lncRNAs and corresponding protein-coding genes.
190 Compared to randomly selected no matter protein-coding genes or lncRNAs, CRLs had more
191 potential triplex binding sites and R-loops with cancer driver protein-coding genes (Fig. 2A-B, see
192 Methods). The results suggested that CRLs have a higher affinity with cancer driver protein-coding
193 genes.

194 The expression levels of I-CRL were highly correlated with cancer driver protein-coding
195 genes as mentioned above. Here, we investigated whether I-CRLs also exhibited a higher affinity
196 for binding to the DNA sequences of cancer driver protein-coding genes. Indeed, our results
197 showed that I-CRLs contained obviously more binding sites in the context of both triplex and R-
198 loops structures compared with other groups (Fig. 2C-D). But there was a little difference of O-
199 CRLs between two structures: the triplex binding sites of O-CRLs were significantly less than I-
200 CRLs, while the R-loop number of O-CRLs was almost the same as I-CRLs. We infer triplex
201 primarily functions as cis-regulatory elements, whereas R-loops have the capacity to regulate gene
202 expression in both cis and trans contexts. In order to demonstrate the potential of these two RNA-
203 DNA interaction mechanisms to affect gene expression, we next categorized I-CRL into two
204 groups: one group comprised of I-CRLs that were predicted to bind to cancer driver protein-coding
205 genes in the same CDT and is known as 'bind' group; the remaining I-CRLs were 'no bind' group,
206 which were not predicted to bind to cancer driver protein-coding genes in the same CDT. We
207 examined the expression correlation values between the two groups and corresponding cancer
208 driver protein-coding gene pairs. No matter whether triplex or R-loop, potential binding pairs had

209 higher co-expression tendencies than no binding pairs (Fig. 2E-F). All these outcomes indicate
210 that these two RNA-DNA interaction mechanisms have the potential to function in cancer driver
211 protein-coding gene expression regulation by I-CRLs.
212
213 **Discovery of pan-cancer driver lncRNAs by constructing CADTAD pipeline.**
214 Given I-CRLs play important roles in cancers, we next identified CDLs through the
215 characteristics mentioned above in pan-cancer. First, we collected 139 TAD data from GEO and
216 did principal component analysis (PCA). We noticed that there was a difference in TAD among
217 normal solid tissue samples, solid tumors, normal blood samples, and hematological malignancies.
218 PC1 distinguished solid tissues and blood samples, and PC2 indicated the difference between
219 normal and cancer samples (Fig. 3A, see Methods). Taking into consideration the heterogeneity
220 between solid tumors and hematological malignancies and given the high prevalence of solid
221 tumor[45], we exclusively utilized TAD data of solid tumors to identify pan-cancer driver lncRNAs.
222 To accomplish this, we developed an integrated computational approach CADTAD to predict
223 CDLs potentially involved in regulating the expression of cancer driver protein-coding genes
224 within CDTs. As a result of our efforts, we identified a total of 256 oncogenic lncRNAs, 177 tumor
225 suppressive lncRNAs, and 75 dual-function lncRNAs (Fig. 3B, see Supplementary Data 5, see
226 Methods).
227 To ascertain the tumor-related attributes of the identified CDLs, we first examined the
228 overlap between CDLs and CRL datasets CLC2 and Lnc2Cancer. The analysis revealed a
229 significant enrichment of CDLs in both datasets, indicating that the mined CDLs retain certain
230 characteristics shared with CRLs, thus potentially shedding light on their deeper cancer-related
231 functions (Fig. 3C). Subsequently, we investigated the number of lncRNAs that were differentially
232 expressed in CCLE compared to HPA by using data from LncExpDB[46] including 1319 up-
233 regulated and 1755 down-regulated lncRNAs (Fig. S4A, see Methods). Potential oncogenic
234 lncRNAs exhibited enrichment in the up-regulated lncRNAs, while candidate tumor suppressive
235 lncRNAs were enriched among the down-regulated lncRNAs in cancer (Fig. 3D). As expected, we
236 observed no enrichment of potential oncogenic lncRNAs among down-regulated lncRNAs nor did
237 we find an enrichment of candidate tumor suppressive lncRNAs within up-regulated lncRNAs (Fig.
238 S4B-C). However, potential dual function lncRNAs were not enriched in either of the differentially
239 expressed lncRNA groups (Fig. S4D). Subcellular localization can provide valuable clues about

240 the function of lncRNAs[15]. To investigate this aspect, we leveraged the nucleus and cytosol-
241 featured lncRNAs from LncBook2.0[47]. Our analysis revealed that the identified CDLs displayed a
242 notable enrichment within the nucleus (Fig. 3E), whereas there was no significant enrichment in
243 the cytosol (Fig. S4E). These findings underscore the predominant role of our CDLs within the
244 nucleus.

245 Traditionally, researchers often rely on mutation data for the discovery of cancer driver
246 protein-coding genes. In our study, we sought to determine if candidate CDLs shared similar
247 characteristics. Remarkably, CDLs exhibited a higher number of SNPs classified as pathogenic
248 variation effects from COSMIC when compared to the remaining lncRNAs (Fig. 3F), suggesting
249 the pathogenic role of CDLs in cancer. Beyond mutation analysis, various other approaches are
250 used to identify cancer genes through epigenetic markers. For instance, broad H3K4me3 marks
251 are associated with tumor-suppressor genes[3], while broad H3K27me3 marks are indicative of
252 oncogenes[2]. Furthermore, our findings demonstrated that the mined oncogenic lncRNAs possessed
253 wider H3K27me3 marks, whereas candidate tumor suppressive lncRNAs exhibited broader
254 H3K4me3 marks (Fig. 3G-H). Collectively, these results provide compelling evidence for the
255 potential of our mined CDLs to drive cancer progression.

256

257 **CADTAD analysis for discovery of putative cancer driver lncRNAs in individual cancers.**

258 Given the heterogeneity inherent in tumors, CADTAD is applied to individual cancer types.
259 To decide which specific cancer to focus on, we considered the distribution of collected TAD
260 sample numbers across various tumors. Prostate cancer, which ranked at the top, was chosen for
261 identifying CDLs in an individual cancer context (Fig. 4A). We gathered cancer driver protein-
262 coding genes specific to prostate cancer from TUSON[1] and analyzed their distribution (Fig. 4B).
263 Next, we employed the pipeline for pan-cancer analysis to prostate cancer and identified a total of
264 97 candidate oncogenic lncRNAs, 29 tumor suppressive lncRNAs, and three dual-function
265 lncRNAs in prostate cancer (Fig. 4C, see Supplementary Data 6).

266 Leveraging functional data relevant to prostate cancer, like pan-cancer CDLs, we found
267 that the mined oncogenic lncRNAs in prostate cancer exhibited a pronounced enrichment in up-
268 regulated lncRNAs, while candidate tumor suppressive lncRNAs were enriched in down-regulated
269 lncRNAs (Fig. 4D, S5A). Meanwhile, the mined CDLs in prostate cancer displayed a higher
270 likelihood of being localized in the nucleus (Fig. 4E), and featured a greater number of SNPs

271 classified as pathogenic variations (Fig. 4F). Additionally, broader H3K27me3 and H3K4me3
272 marks have been respectively detected in putative oncogenic lncRNAs and tumor suppressive
273 lncRNAs in prostate cancer (Fig. 4G).

274 To further validate the tumor-specific characteristics of potential CDLs, we conducted
275 experimental evaluations. We randomly selected multiple candidate CDLs in prostate cancer and
276 experimentally assessed six prospective oncogenic lncRNAs and four tumor suppressive lncRNAs.
277 Each candidate lncRNA was disrupted by using two CRISPR-Cas9 guide RNAs. The cutting
278 efficiencies of each pair of guide RNAs were validated through T7 Endonuclease I assay (Fig. S6).
279 We also detected the expression level of each candidate lncRNA respectively by RT-qPCR.
280 Intriguingly, knockdown of each candidate oncogenic lncRNAs significantly inhibited the
281 proliferation of LNCaP cells as we expected (Fig. 5A). Conversely, we observed an obvious cell
282 growth increase in the LNCaPs which knockdown two out of four candidate tumor suppressive
283 lncRNAs, and one candidate tumor suppressive lncRNA had the same tendency (Fig. 5B).
284 Furthermore, knockdown of our candidate oncogenic lncRNAs slowed the wound closure in PC-
285 3 prostate cancer cells (Fig. 6A). However, cells with downregulation of tumor suppressive
286 lncRNAs closed the wound faster than the control groups (Fig. 6B). In addition, Kaplan–Meier
287 survival curve demonstrated that prostate adenocarcinoma (PRAD) patients with lower levels of
288 candidate oncogenic lncRNA expression showed extended survival compared to patients with high
289 levels of oncogenic lncRNA expression (Fig. S7A). Conversely, candidate tumor suppressive
290 lncRNAs exhibited the opposite effect in the terms of patient survival (Fig. S7B). Thus, these
291 survival outcomes underline the potential of these candidate CDLs as prognostic markers indicate
292 of prostate cancer progression.

293 Furthermore, we selected gastric cancer and lung cancer, which were ranked at the top
294 based on the distribution of TAD sample numbers to identify their respective CDLs (Fig. 4A).
295 Through the CADTAD approach, we successfully identified a total of 139 candidate oncogenic
296 lncRNAs, 193 tumor suppressive lncRNAs, and 18 dual-function lncRNAs for gastric cancer (Fig.
297 S8A-B, see Supplementary Data 7). We also identified 32 oncogenic lncRNAs, 254 tumor
298 suppressive lncRNAs, and 20 dual-function lncRNAs for lung cancer (Fig. S9A-B, see
299 Supplementary Data 8). The CDLs identified in these two cancer types exhibited consistent cancer-
300 related features, including enrichment of differentially expressed lncRNAs in cancer, nuclear

301 localization, a higher proportion of SNPs classified as pathogenic, and the presence of histone
302 markers associated with cancer genes (Fig. S5B-C, S8C-F, S9C-F).
303 All of these findings underscore the significance of the CDTs combined with expression
304 and interaction as a robust signature for the discovery of CDLs.
305

306 ## Discussion

307 LncRNAs can create an additional regulatory dimension superimposed over the genomic
308 and epigenomic programs by interacting with DNAs, RNAs, or proteins to orchestra gene
309 expression[9], defying the conventional Central Dogma of molecular biology[11]. In our investigation
310 into the relationship between CRLs and cancer driver protein-coding genes, we are the pioneer to
311 uncover that CRLs exhibit a proximity to cancer driver protein-coding genes within CDTs. This
312 finding represents a novel characteristic of cancer lncRNAs. Furthermore, the expressions of I-
313 CRLs demonstrated a notably correlation with the gene expression levels of cancer driver protein-
314 coding gene in the same TAD, indicating a potential regulation connection. Additionally, a
315 comprehensive comparison of lncRNA-DNA interactions further accentuated the likelihood that
316 I-CRLs play a role in regulating the expression of cancer driver protein-coding genes. By
317 establishing a direct link between lncRNAs and cancer driver protein-coding genes, we constructed
318 a pipeline CADTAD and identified potential core CDL datasets in pan-cancer analysis and across
319 three specific cancer types. To validate the effectiveness of these datasets, we used various
320 bioinformatics function datasets to unveil the cancer related characteristics associated with mined
321 CDLs. We also conducted in vitro assay using CRISPR-Cas9 system to validate the influence of
322 candidate CDLs on cancer phenotype. All these results proved that these CDL datasets indeed
323 exhibited features closely linked to cancer. Additionally, we classified CDLs into three distinct
324 categories: oncogenic lncRNA, tumor suppressive lncRNA, and dual-function lncRNA. Notably,
325 our method CADTAD represents the first in-silico approach to achieve this comprehensive
326 categorization.
327 In some studies, the identification of driver lncRNAs primarily revolves around a genomic
328 perspective. For instance, oncodriveFML[48] could identify both coding and non-coding cancer
329 driver mutations including mutations found in lncRNAs. This is achieved by estimating the
330 cumulative functional impact bias of somatic mutations within tumors, using signals of positive
331 selection as a guiding principle. Furthermore, ExInAtor1/2[37,49] provides a framework to identify

driver lncRNAs by discerning driver mutations within them. This identification is rooted in the detection of signals of positive selection that act on somatic mutations, thereby considering the mutational burden and functional impact. And mutations in these regions are presumed to affect lncRNA function.

Other studies delve into exploring lncRNA function in tumors by focusing on lncRNA-mediated regulation. For example, they may identify dysregulated lncRNA-TF-gene triplets in glioblastoma[50], or develop methods like lncMAP[51] to reveal distinct types of lncRNA regulatory molecules. Additionally, an effective approach to understanding lncRNA function in cancer involves associating these lncRNAs with pivotal mutational driver events. In such case, lncRNAs are often regarded as downstream effectors in oncogenic pathways[52]. DriverLncNet[53], on the other hand, offers a comprehensive strategy that combines both genomic and regulatory features. This method identifies cancer driver lncRNAs through the analysis of copy number alteration and the construction of regulatory networks. These networks aim to capture key functional effectors, thereby dissecting the functional role of driver lncRNAs. While numerous bioinformatics methods already exist for mining cancer driver or associated lncRNAs, there remains room for improvement, particularly in enhancing the precision of CDL prediction.

First, CDLs are typically classified into two categories based on their functions, namely, oncogenic lncRNAs and tumor suppressive lncRNAs. The current methods for discovering CDLs have primarily remained at the level of CRLs, lacking the specificity to differentiate their precise roles in either promoting or inhibiting cancer progression. Second, the identification of driver lncRNAs based on genomic characteristics has yielded relatively low numbers[37,48,49]. To address this, we advocate approaching CDLs from a distinct perspective, one that is different from the paradigm of cancer driver protein-coding genes. In this context, we define CDLs as lncRNAs that directly regulate the expression of cancer driver protein-coding genes, with the objective of uncovering the lncRNAs with the most significant potential impact on cancer. While our approach CADTAD does not consider alterations within the CDLs themselves, such as mutations or epigenetic modifications, it focuses on the crucial task of discerning CDLs that govern the core drivers of cancer. CADTAD also facilitates the distinction of lncRNA-specific functions in either promoting or inhibiting cancer progress.

Nonetheless, our exploration of CDLs was limited to just three cancer types due to insufficient Hi-C data. In the future, the expansion of Hi-C data across various cancer types will

363 enable the comprehensive mining of CDLs in individual cancers, thereby enlarging the CDL
364 datasets. It is essential to emphasize that the present CDL dataset exclusively focuses on lncRNAs
365 that directly regulate the expression of cancer driver protein-coding genes through lncRNA-DNA
366 interaction prediction. On one hand, many novel experimental techniques, such as GRID-seq,
367 RADICL-seq[10], have been developed to detect RNA-chromatin interactions. These high
368 throughput sequencing can be integrated to enhance the accuracy of lncRNA-DNA interaction. On
369 the other hand, lncRNAs have the capacity to influence coding gene expression through various
370 mechanisms, including liquid–liquid phase separation[54] and miRNA sponging[55]. This CDL dataset
371 represents the core driver dataset within the broader spectrum of all CDLs capable of directly
372 modulating cancer driver protein-coding gene expression. While acknowledging the intricacies of
373 lncRNA regulatory mechanisms, there is an inherent need for a more comprehensive approach to
374 mining CDLs that encompasses all conceivable lncRNA regulatory pathways. This comprehensive
375 approach will enable us to explore the multifaceted landscape of CDLs more comprehensively.
376 In summary, this study effectively utilizes 3D genome interaction information to achieve
377 highly precise mining of CDLs. This pipeline CADTAD holds the potential for broader
378 applications in the study of other diseases, offering a new perspective on cancer biology. Notably,
379 it represents an innovative approach to identify new biomarkers and prognostic indicators for
380 cancer treatment.
381

382 # Methods

383 **Source of datasets analyzed in this project.**

384 lncRNA gene and protein coding gene reference was from LncExpDB[46] and converted to hg19
385 using LiftOver downloaded from the UCSC Genome Browser website
386 (https://hgdownload.soe.ucsc.edu/downloads.html#utilities_downloads). Thus, we got 98410
387 lncRNAs and 19840 coding genes. The expression of lncRNA and protein coding genes also came
388 from LncExpDB with combined CCLE and HPA TPM data. Fasta file hg19.fa was from UCSC
389 (https://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/hg19.fa.gz) and used for later Hi-C
390 mapping and lncRNA-DNA interaction prediction. Cancer-related lncRNA (CRL) datasets were
391 collected from CLC2[25] and Lnc2Cancer3.0[26] and converted to LncBook lncRNA id according to
392 LncExpDB reference, and we obtained respectively 426 and 822 CRLs to do the further analysis.
393 Cancer driver protein-coding genes were from CGC database[36] v.94 and were classified as tumor
394 suppressor gene or oncogene according to column 'Role in Cancer' containing 'TSG' or
395 'oncogene'. If one gene's column 'Role in Cancer' contained both 'TSG' and 'oncogene', it was
396 classified as dual-function gene. We got a total of 556 cancer driver protein-coding genes
397 according to the LncExpDB reference. 3804 housekeeping genes were extracted from a
398 publication[56]. 1000 oncogenes and 1000 tumor suppressor genes were from a publication[1]. 531
399 genes in the PATHWAY IN CANCER (hsa05200) were downloaded from Kyoto Encyclopedia
400 of Genes and Genomes (KEGG) database[57] (see Supplementary Data 1).

401 Cancer driver protein-coding genes in individual cancer type were collected from a publication[1],
402 remained under p values < 0.5, and converted from alias to gene symbol, getting both tumor
403 suppressor genes, oncogenes and dual-function genes (Fig. 4B, S8A, S9A, see Supplementary Data
404 2). The statistic of individual cancer genes was calculated using python package upsetplot (v.0.8.0).
405 RefLnc[58] was used to supplement the TCGA (PRAD, STAD, LUSC, and LUAD) and GTEx
406 projects with combined FPKM data to obtain the expression of 26494 lncRNAs and 19490 coding
407 genes in a single cancer type. The 17 driver lncRNAs identified by ExInAtor2 were from a
408 publication[37].

409

410 **Bioinformatics analysis of Hi-C data.**

411 Hi-C raw data were collected from ENCODE and GEO, getting 139 samples (see Supplementary
412 Data 3). Raw data were mapped to hg19 using bwa-mem2 (v.2.0) with parameters E=50, L=0, and

413 M, and then transferred to bam files using samtools[59] (v.1.6). The downstream analysis was
414 performed using HiCExplorer[60] (version 3.7.2). Each enzyme's specific restriction sites were
415 located using hicFindRestSite. Then, to obtain a Hi-C matrix h5 file, we utilized hicBuildMatrix
416 with the following parameters: binSize=1000, inputBufferSize=200000, corresponding enzyme's
417 restriction sites, and dangling sequence. Next, we merged each h5 file under parameters nb=40 to
418 create an h5 file with bin size 40kb, normalized them solely using hicNormalize with parameters
419 normalize=smallest, and corrected these files using hicCorrectMatrix with default parameters.
420 Finally, normalize them all using hicNormalize with parameters normalize=smallest.
421 We then utilized the normalized matrix to calculate topologically associated domain (TAD) using
422 hicFindTADs with parameters correctForMultipleTesting=fdr. Domains.bed files were used for
423 downstream TAD analysis. To distinguish TADs differences among normal and tumor samples,
424 we then converted h5 files to cool files using hicConvertFormat and chose the file with the largest
425 TAD number as the basis to calculate the interval scores using chromosight[61] (v.1.6.1). Then we
426 analyzed these interval scores to plot Principal Component Analysis (PCA) in R v.4.2.0.
427 pyGenomeTracks was used to plot TAD.
428
429 **Method to define CDTs and calculation of lncRNA ratio.**
430 Using the intersect function in bedtools[62] (v.2.30.0) with parameter F=1, we identified TADs as
431 cancer driver TADs (CDTs) if they entirely included at least one cancer driver protein-coding gene.
432 The rest TAD completely containing non-cancer genes, which excluded cosmic genes, top1000
433 TUSON OGs/TSGs, and genes about the PATHWAY IN CANCER, were defined as typical TADs.
434 Then we randomly selected the same number of typical TADs as CDTs with almost the same
435 length, and determined the ratio by dividing the total number of CRLs by the summed length of
436 the TAD in each TAD type (Fig. S1B). To compare with neighbor TADs around CDTs, we also
437 calculated the ratio of up-/down-stream 5 non-CDTs. Once there was one CDT in up-/down-stream
438 5 TADs, we skipped it until finding a total of 5 non-CDTs. If the up-/down-stream of CDTs didn't
439 have 5 non-CDTs, we then assigned a ratio of 0 in the non-TAD region (Fig. S1C).
440
441 **Randomly select non-cancer genes and non-cancer lncRNAs.**
442 For TAD ratio calculation (Fig. 1C-F), we chose typical TAD as a control, and this type of TAD
443 had to be overlapped with one non-cancer genes. We defined non-cancer genes there excluding

444 cosmic genes, top1000 TUSON OGs/TSGs, and genes about PATHWAY IN CANCER in KEGG.
445 For comparison of co-expression and binding (Fig. 1A, 2A-B), non-cancer genes and non-cancer
446 lncRNAs were chosen at random. Non-cancer genes were randomly selected with the same number
447 as cancer driver protein-coding genes from the genes which were located in typical TAD, while
448 non-cancer lncRNAs were selected with the same number as corresponding CRL datasets from
449 lncRNAs excluding both CRL datasets. For comparison of co-expression with cancer driver
450 protein-coding genes between I-CRL and O-CRL (Fig. 1J), we first got the pair between I-CRL
451 and cancer driver protein-coding gene in the same TAD, and then randomly selected the same
452 number O-CRL excluding I-CRL to replace I-CRL of the pair. For comparison among 4 type
453 lncRNAs (Fig. 2D), there was no random selection for the reason that we considered all binding
454 between each type of lncRNA and cancer driver protein-coding genes.
455
456 **Expression correlation calculation and comparison of cancer lncRNA discovery effectiveness.**
457 Here, we calculated the Spearman correlation between the expression of paired lncRNAs and
458 protein coding genes. We got mined 30289 cancer lncRNAs by CDT feature in pan-cancer. In
459 addition, we acquired cancer lncRNAs by calculating the expression correlation and choosing
460 different thresholds. Next, in order to assess the cancer lncRNA finding effectiveness by CDT and
461 co-expression characteristics, we performed enrichment analysis and calculated the ratio between
462 these two types of mined cancer lncRNAs and cancer-related lncRNA datasets.
463
464 **lncRNA-DNA interaction analysis.**
465 We were aware of two mechanisms in which the lncRNA might interact with DNA: triplex and R-
466 loop. We reviewed all relevant prediction techniques and finally chose classic methods Triplextor[43]
467 and QmRLFS-finder[44] to predict lncRNA-DNA interaction. Before prediction, we got lncRNAs'
468 and genes' sequences from hg19.fa using getfasta function in bedtools[62] (v.2.30.0). For triplex
469 prediction, we predicted the number between lncRNA and genes using the default parameter. For
470 R-loop prediction, we first predicted the R-loop structure in each lncRNA and then counted all
471 matched numbers of paired genes whose sense or antisense strand had RIZ sequence[22], as RIZ
472 sequence is most important to form R-loop.
473

474 **Construction of a cancer-related lncRNA-mRNA network and definition of lncRNA**
475 **centrality.**
476 We calculated the Pearson correlation between the expression of all lncRNAs and all mRNAs
477 getting a similarity matrix, and then transferred to an adjacency matrix by raising the similarity to
478 a power[39,40]. We used the scale-free topology criterion for soft thresholding and the minimum
479 scale-free topology fitting index R^2 was set as 0.9, so we got 6 as the soft threshold (Fig. S2A-B).
480 To mine cancer-related functions, we defined important genes containing cancer genes from CGC
481 database v.94, the top 500 oncogenes and tumor suppressor genes from TUSON, and genes in
482 PATHWAY IN CANCER from KEGG. We used power 6 to calculate the network between all
483 lncRNAs and important genes and summed the values of one lncRNA with all important genes to
484 define lncRNA centrality. Then we only focused on the I-CRL's and O-CRL's rank in all lncRNA,
485 so we did GSEA to compare them using the Python package GSEApy[63] (v.1.0.5).
486
487 **Construct pipeline CADTAD to mine CDLs.**
488 We constructed a pipeline to mine CDLs in pan-cancer and individual cancer types (Fig. 3B, 4C,
489 S8B, S9B). First, we got all TAD completely covering cancer driver protein-coding genes, and
490 then we merged all intersected TAD to define CDT using intersect and merge function in
491 pybedtools[62,64] (v. 0.9.0). Then, we discovered lncRNAs located in these CDT with a reference of
492 all lncRNAs and matched these lncRNAs with cancer driver protein-coding genes in the same
493 CDT. Next, we predicted potential lncRNA-DNA interaction pairs of matched lncRNA-cancer
494 driver protein-coding gene pairs by Triplexator and QmRLF-finder as soon as one of the methods
495 predicted. Finally, we used tumor samples' RNA-seq expression data to calculate the expression
496 correlation of remaining lncRNA-cancer driver protein-coding gene pairs and obtained lncRNAs
497 with absolute correlation values greater than 0.5 as CDLs. We considered CDLs whose expression
498 correlation values were only positive with oncogenes and/or negative with tumor suppressor genes
499 as oncogenic lncRNAs, only positive with tumor suppressor genes and/or negative with oncogenes
500 as tumor suppressive lncRNAs, and had the above two cases as dual-function lncRNAs. And we
501 applied CADTAD to pan-cancer and individual cancers to identify CDLs (see Supplementary Data
502 5-8).
503
504 **Bioinformatics methods related to cancer characteristics verification.**

505 To obtain up-regulated and down-regulated lncRNAs in cancer, we compared cancer samples with
506 normal samples using the Wilcoxon test and calculating fold change, and lncRNAs whose p values
507 were less than 0.05 and |logFC| were more than 1 were considered as differential expressions.
508 Nucleus and cytosol-featured lncRNAs were taken from LncExpDB[46]
509 (ftp://download.big.ac.cn/lncexpdb/5-SubcellularLocalization/1-
510 SRP150867/deseq2_SRP150867.tar.gz) by using cutoff adjusted p-value < 0.05 and
511 log2FoldChange < -1 to obtain respectively 3770 nucleus and 38 cytosol-featured lncRNAs.
512 Variation data were extracted from LncBook2.0[47]
513 (https://ngdc.cncb.ac.cn/lncbook/files/variation_LncBook2.0.csv.gz) and we only used COSMIC
514 Variant Effect which was pathogenic to compare. H3K4me3 and H3K27me3 peak files were
515 downloaded from ENCODE (see Supplementary Data 4), and we calculated each lncRNA's peak
516 width by summing up the intersected peaks' length. Survival data were from UCSC Xena[65] and
517 survival analyses were conducted by Python package lifelines (v.0.27.7).
518
519 **CRISPR gRNA and lentiviral vector design.**
520 For each candidate lncRNA, two pairs of guide RNAs were designed to delete about 4 kb of the
521 promoter region defined as from 5 kb upstream to 1 kb downstream of transcription site (TSS).
522 The sequences of each guide RNAs were listed as follows:
523

Gene	pgRNA	gRNA1 sequence	gRNA2 sequence
L024	p1	AGGCTCCCTCTGAGCCGACA	GTGTGGCCACCGGAACACCA
	p2	AGGCTCCCTCTGAGCCGACA	CGTCCCCCTCCTAGTCTCCG
L873	p1	GCTCTGCTAGATAAATCTGG	AGGGCGCCGCTTAAAGTGTG
	p2	GCTCTGCTAGATAAATCTGG	CAGGGTGTGGAGCCGCACTG
L989	p1	GTCAGGACACAGATGACCGA	ACCTTGGCAAGTGAAGACGC
	p2	GTCAGGACACAGATGACCGA	AAACATTCAGGCCCCCACCG
L569	p1	GCTGTGTCAACCATGCCCCG	TTCCAACCAGAATAGGACCA
	p2	GCTGTGTCAACCATGCCCCG	CAGAGCTCGTAATTGCTGCG
L538	p1	GCTCTTCTGATACTGACCCG	TAGTAGGGCTCAAATCTCTG
	p2	CTAAGATGTCCAAGTCCTAG	TAGTAGGGCTCAAATCTCTG
L581	p1	GTCCACAGCAAGAGTGTCCG	CTTCCCTCCAAAAATCACCG

	p2	TGTTCCTGGGTGGCGACGTG	AGGTTTCTAGACGTGACCCA
L774	p1	GTTGCAGCCATCGTCCTGGG	TTTCGGCATGTGATCCGAGG
	p2	GTTGCAGCCATCGTCCTGGG	GCGGGACTTCCAGAGCCCGA
L614	p1	GGTGTGATAACTTTACTGAT	CTCGCCCGCTTCGCCCGGGT
	p2	ATGTTATCAGCCGCACACGG	CTCGCCCGCTTCGCCCGGGT
L011	p1	CACTTGGAAACAATTACTGT	AAGTCGCCGACAACTGCCTG
	p2	TTTCTGATATTAATCAGGCG	AAGTCGCCGACAACTGCCTG
L120	p1	GCTGAGGCTATTGTCCGGG	AGGACACTGTTAGACCACCT
	p2	TCACCCATCGACTACCCGGG	GTGAAACATTAACCCCCTCA

The oligos for each target site were annealed by heating to 95ºC for 5 minutes and cooling to 85 ºC at 2 ºC per second and then further down to 25 ºC at 0.1 ºC per second. The annealed oligos were cloned into lentiCRISPR v2 plasmid and sequenced to confirm successful ligation.

Cell culture and lentiviral transduction.

LNCaP, PC-3 and HEK293T were purchased from ATCC. All cells were maintained in DMEM medium supplemented with 10% FBS and cultured in 5% CO_2. These cell lines were mycoplasma negative during routine tests.

Lentiviral was packaged by co-transfection of second packaging vectors pMD2.G, psPAX2 into HEK293T cells. After 48 hours of transfection, the medium was harvest and centrifuged at 800 × g for 10 minutes to remove cell debris. Then the supernatant was used for infection when LNCaP or PC-3 were grown to 70% confluence. The medium was changed 24 hours after viral transduction and infected cells incubated for another 24 hours before puromycin selection. Positive cells were collected for cell proliferation assay, woundheal assay and total RNA and genomic DNA extraction.

PCR and RT-qPCR verification of CRISPR/Cas9 induced deletion at candidate lncRNAs promoter.

Mutagenesis on candidate lncRNAs was verified by PCR using primers flanking the deletion regions. RT-qPCR was used to detect the RNA levels of lncRNA candidates after CRISPR/Cas9 editing. Total RNA of edited cells was extracted and converted to cDNA. Then qPCR was

546 performed with ChamQ Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China). The
547 primers used for PCR and RT-qPCR were listed as follows.
548

Gene	PCR primers	Real-time PCR primers
L024	F: TGAGACTACCTGCAGGACCA	F: CCTTCCTCAGCCATTGTCCA
	R: GGAAGTGACACTCCGTCCTC	R: TTGCTTCACAGTGCCCCTAA
L873	F: GTTCTGTCCGGCCATTCTGT	F: CCTAGGACGTCTCCATTGCG
	R: ATCATCTTCCATCTCCGCCG	R: ACTCCGCCTAGTAGTGACCG
L989	F: GCAGACAGAGGGGTCAACAG	F: TGGGGGCCATTCTCAATGTT
	R: CTCAGGTGCCTCTGGAACTC	R: AGAGAGCAACGCTCTGTGAG
L569	F: CCAGGGTGGAACCCAACTATT	F: CCAGCCATGTGGTGTAGACG
	R: ACATCCTGATGATGCCTCCA	R: AGATGCAATCGCAGATGTGG
L538	F: TGCGGGAGGAGAGACTACAT	F: GGCCCAGCGATGCTTCTTAT
	R: GGTTAGCTTCAGAGCTGCCA	R: TAGCTTCAGAGCTGCCAAGG
L581	F: GGGGACACCACTCCTCCATC	F: GAGAATGGAGCGGAGGAACC
	R: TGGCTCAGGAATTTGCTCCC	R: TGTGTCATGGCTGACATGCT
L774	F: CGGCTCTGACCTGAGCTATG	F: TGAGAGGCTACGCTGTATGG
	R: CCCCCTGCACTCTGAATAGG	R: GAGACATGGTACAGGGTGGG
L614	F: AGCCATGTGAAACAACACACA	F: ACCTGATTCTTTACGTGATTCCCA
	R: AGTTTCTGAAGCGTCCCTCG	R: TCAGGACGCAGAGGAATTAGT
L011	F: CCCTACAAACCTAACCCCAACA	F: AGAGGCACGTGGTTCGAAAT
	R: CATTTTGGCTTCAGGGGCTG	R: AGTCATTGGCATGGAGTGGG
L120	F: AGCCTGGGAAGTTCGCTTAC	F: CAGCCCGGTTTGGGAGTAAA
	R: AAAGATTGAGGGTCGGGCTG	R: AGCCAGACAGAAACTACGCC

549

550 **Cell proliferation and wound healing assay.**

551 LNCaP cells after viral transduction and puromycin selection were seeded in 96-well plates at a
552 density of 6000 cells per well and allowed to attach for 24 hours. Cell viability was measured
553 utilizing the CCK8 kit across four days. CRISPR/Cas9 guide RNA edited PC-3 were inoculated
554 into 6-well plate and the wound was created until the cell confluence arrived at 90%. The photos
555 were taken at the indicated time by using microscope.

556

557 **Reference**

558 1. Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy
559 patterns and shape the cancer genome. *Cell* **155**, 948 (2013).
560 2. Zhao, D. *et al.* Broad genic repression domains signify enhanced silencing of oncogenes.
561 *Nat. Commun.* **11**, (2020).
562 3. Chen, K. *et al.* Broad H3K4me3 is associated with increased transcription elongation and
563 enhancer activity at tumor-suppressor genes. *Nat. Genet.* **47**, 1149–1157 (2015).
564 4. Martínez-Jiménez, F. *et al.* A compendium of mutational cancer driver genes. *Nat. Rev.*
565 *Cancer* **20**, 555–572 (2020).
566 5. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of
567 their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
568 6. Melé, M. & Rinn, J. L. 'Cat's Cradling' the 3D Genome by the Act of LncRNA
569 Transcription. *Mol. Cell* **62**, 657–664 (2016).
570 7. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs
571 and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
572 8. Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs.
573 *Nat. Rev. Genet.* **21**, 102–117 (2020).
574 9. Herman, A. B., Tsitsipatis, D. & Gorospe, M. Integrated lncRNA function upon genomic
575 and epigenomic regulation. *Mol. Cell* **82**, 2252–2266 (2022).
576 10. Mattick, J. S. *et al.* Long non-coding RNAs: definitions, functions, challenges and
577 recommendations. *Nat. Rev. Mol. Cell Biol.* (2023) doi:10.1038/s41580-022-00566-8.
578 11. Chang, H. Y. & Qi, L. S. Reversing the Central Dogma: RNA-guided control of DNA in
579 epigenetics and genome editing. *Mol. Cell* **83**, 442–451 (2023).
580 12. Huarte, M. The emerging role of lncRNAs in cancer. *Nat. Med.* **21**, 1253–1261 (2015).
581 13. Schmitt, A. M. & Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell*
582 **29**, 452–463 (2016).
583 14. Bhan, A., Soleimani, M. & Mandal, S. S. Long noncoding RNA and cancer: A new
584 paradigm. *Cancer Res.* **77**, 3965–3981 (2017).
585 15. Bridges, M. C., Daulagala, A. C. & Kourtidis, A. LNCcation: lncRNA localization and
586 function. *J. Cell Biol.* **220**, 1–17 (2021).
587 16. Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus.

588 *Cell* **184**, 5775-5790.e30 (2021).
589 17. Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA.
590 *Nature* **451**, 202–206 (2008).
591 18. Kotake, Y. *et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to
592 and silencing of p15 INK4B tumor suppressor gene. *Oncogene* **30**, 1956–1962 (2011).
593 19. Rinn, J. L. *et al.* Functional Demarcation of Active and Silent Chromatin Domains in
594 Human HOX Loci by Noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
595 20. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote
596 cancer metastasis. *Nature* **464**, 1071–1076 (2010).
597 21. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates
598 tumour biology. *Nature* **465**, 1033–1038 (2010).
599 22. Luo, H. *et al.* HOTTIP-dependent R-loop formation regulates CTCF boundary activity and
600 TAD integrity in leukemia. *Mol. Cell* **82**, 833-851.e11 (2022).
601 23. Tseng, Y. Y. *et al.* PVT1 dependence in cancer with MYC copy-number increase. *Nature*
602 **512**, 82–86 (2014).
603 24. Liu, S. J. *et al.* CRISPRi-based genome-scale identification of functional long noncoding
604 RNA loci in human cells. *Science (80-.).* **355**, (2017).
605 25. Vancura, A. *et al.* Cancer LncRNA Census 2 (CLC2): an enhanced resource reveals clinical
606 features of cancer lncRNAs. *NAR Cancer* **3**, 1–15 (2021).
607 26. Gao, Y. *et al.* Lnc2Cancer 3.0: An updated resource for experimentally supported
608 lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq
609 data. *Nucleic Acids Res.* **49**, D1251–D1258 (2021).
610 27. Zhao, T. *et al.* Identification of cancer-related lncRNAs through integrating genome,
611 regulome and transcriptome features. *Mol. Biosyst.* **11**, 126–136 (2015).
612 28. Zhang, X., Wang, J., Li, J., Chen, W. & Liu, C. CRlncRC: A machine learning-based
613 method for cancer-related long noncoding RNA identification using integrated features.
614 *BMC Med. Genomics* **11**, (2018).
615 29. Zhang, X. *et al.* Identification of cancer-related long non-coding RNAs using XGboost with
616 high accuracy. *Front. Genet.* **10**, 1–14 (2019).
617 30. Liu, Z. *et al.* Identifying Cancer-Related lncRNAs Based on a Convolutional Neural
618 Network. *Front. Cell Dev. Biol.* **8**, 1–7 (2020).

619 31. Zhang, J. *et al.* CLING: Candidate Cancer-Related lncRNA Prioritization via Integrating
620 Multiple Biological Networks. *Front. Bioeng. Biotechnol.* **8**, 1–12 (2020).
621 32. Yuan, L., Zhao, J., Sun, T. & Shen, Z. A machine learning framework that integrates multi-
622 omics data predicts cancer-related LncRNAs. *BMC Bioinformatics* **22**, 1–18 (2021).
623 33. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre.
624 *Nature* **485**, 381–385 (2012).
625 34. Valton, A. L. & Dekker, J. TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.*
626 **36**, 34–40 (2016).
627 35. Werner, M. S. *et al.* Chromatin-enriched lncRNAs can act as cell-type specific activators of
628 proximal gene transcription. *Nat. Struct. Mol. Biol.* **24**, 596–603 (2017).
629 36. Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction across
630 all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).
631 37. Esposito, R. *et al.* Tumour mutations in long noncoding RNAs enhance cell fitness. *Nat.*
632 *Commun.* **14**, 2021.11.06.467555 (2023).
633 38. Zhang, J., Le, T. D., Liu, L. & Li, J. Inferring and analyzing module-specific lncRNA-
634 mRNA causal regulatory networks in human cancer. *Brief. Bioinform.* **20**, 1403–1419
635 (2018).
636 39. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network
637 analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, (2005).
638 40. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network
639 analysis. *BMC Bioinformatics* **9**, (2008).
640 41. Niehrs, C. & Luke, B. Regulatory R-loops as facilitators of gene expression and genome
641 stability. *Nat. Rev. Mol. Cell Biol.* **21**, 167–178 (2020).
642 42. Li, Y., Syed, J. & Sugiyama, H. RNA-DNA Triplex Formation by Long Noncoding RNAs.
643 *Cell Chem. Biol.* **23**, 1325–1333 (2016).
644 43. Buske, F. A., Bauer, D. C., Mattick, J. S. & Bailey, T. L. Triplexator: Detecting nucleic acid
645 triple helices in genomic and transcriptomic data. *Genome Res.* **22**, 1372–1381 (2012).
646 44. Jenjaroenpun, P., Wongsurawat, T., Yenamandra, S. P. & Kuznetsov, V. A. QmRLFS-
647 finder: A model, web server and stand-alone tool for prediction and analysis of R-loop
648 forming sequences. *Nucleic Acids Res.* **43**, W527–W534 (2015).
649 45. Chizuka, A. *et al.* Difference between hematological malignancy and Solid tumor research

650 articles published in four major medical journals. *Leukemia* **20**, 1655–1657 (2006).
651 46. Li, Z. *et al.* LncExpDB: An expression database of human long non-coding RNAs. *Nucleic*
652 *Acids Res.* **49**, D962–D968 (2021).
653 47. Li, Z. *et al.* LncBook 2.0: integrating human long non-coding RNAs with multi-omics
654 annotations. *Nucleic Acids Res.* **51**, 186–191 (2022).
655 48. Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N.
656 OncodriveFML: A general framework to identify coding and non-coding regions with
657 cancer driver mutations. *Genome Biol.* **17**, 1–13 (2016).
658 49. Lanzós, A. *et al.* Discovery of Cancer Driver Long Noncoding RNAs across 1112 Tumour
659 Genomes: New Candidates and Distinguishing Features. *Sci. Rep.* **7**, 1–16 (2017).
660 50. Li, Y. *et al.* Identification and characterization of lncRNA mediated transcriptional
661 dysregulation dictates lncRNA roles in glioblastoma. *Oncotarget* **7**, 45027–45041 (2016).
662 51. Li, Y. *et al.* LncMAP: Pan-cancer Atlas of long noncoding RNA-mediated transcriptional
663 network perturbations. *Nucleic Acids Res.* **46**, 1113–1123 (2018).
664 52. Ashouri, A. *et al.* Pan-cancer transcriptomic analysis associates long non-coding RNAs with
665 key mutational driver events. *Nat. Commun.* **7**, 13197 (2016).
666 53. Zhang, Y. *et al.* Identifying Cancer Driver lncRNAs Bridged by Functional Effectors
667 through Integrating Multi-omics Data in Human Cancers. *Mol. Ther. - Nucleic Acids* **17**,
668 362–373 (2019).
669 54. Adnane, S., Marino, A. & Leucci, E. LncRNAs in human cancers: signal from noise. *Trends*
670 *Cell Biol.* **32**, 565–573 (2022).
671 55. Du, Z. *et al.* Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory
672 network in prostate cancer. *Nat. Commun.* **7**, (2016).
673 56. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29**,
674 569–574 (2013).
675 57. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-watanabe, M. KEGG
676 for taxonomy-based analysis of pathways and genomes. **51**, 587–592 (2023).
677 58. Jiang, S. *et al.* An expanded landscape of human long noncoding RNA. *Nucleic Acids Res.*
678 **47**, 7842–7856 (2019).
679 59. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
680 2079 (2009).

681 60. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome
682 organization in flies. *Nat. Commun.* **9**, (2018).
683 61. Matthey-Doret, C. *et al.* Computer vision for pattern detection in chromosome contact maps.
684 *Nat. Commun.* **11**, (2020).
685 62. Quinlan, A. R. & Hall, I. M. BEDTools : a flexible suite of utilities for comparing genomic
686 features. **26**, 841–842 (2010).
687 63. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set
688 enrichment analysis in Python. *Bioinformatics* **39**, 1–3 (2023).
689 64. Dale, R. K., Pedersen, B. S. & Quinlan, A. R. Pybedtools : a flexible Python library for
690 manipulating genomic datasets and annotations. **27**, 3423–3424 (2011).
691 65. Caicedo, H. H., Hashimoto, D. A., Caicedo, J. C., Pentland, A. & Pisano, G. P. Visualizing
692 and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 669–673
693 (2020).
694
695

Author contributions

D.Z. conceived the project. D.Z. and Z.R. designed the bioinformatics analyses and interpreted the results. Z.R. performed the bioinformatics analyses, with assistance from S.H., C.W., and Y.Z. M.Z. designed and performed the experiments, and interpreted the results, with assistance from W.Z, and X.L. D.Z., M.Z., and Z.R. wrote the manuscript.

Acknowledgements

We are grateful to Lina Ma, and Zhao Li for their helpful support of lncRNA data. This work was supported by the National Key R&D Program of China (2021YFF1201100), the National Natural Science Foundation of China (NSFC, 32270603) and the Fundamental Research Funds for the Central Universities (BMU2021YJ057). This work was supported by High-performance Computing Platform of Peking University.

Competing interests

The authors declare that they have no conflict of interest.

712 **Figure legends**

713 **Figure 1**

714 **CRLs are more co-expressed and located near cancer driver protein-coding genes, especially**
715 **in the same TAD, and play an important role in a cancer-related lncRNA-mRNA network.**

716 **A**. Boxplot to show the difference of absolute Spearman correlation between the experimental pair
717 and 2 control pairs. P values are determined by one tail Wilcoxon test.

718 **B**. Line plot of the statistic of the lncRNA number in CLC2/Lnc2Cancer in regions whose step
719 size is 100kb extending from 100kb to 50000kb near the TSS of cancer driver genes/housekeeping
720 genes. The red line indicates cancer driver genes, while the blue line indicates housekeeping genes.
721 P values are determined by one tail Wilcoxon test.

722 **C**. Boxplot to indicate the TAD length of CDT and typical TAD are almost the same in LNCaP. P
723 value is determined by two tail Wilcoxon test.

724 **D**. Bar plot to show the ratio that the number of CRLs divided by TAD length in each TAD of
725 LNCaP. The red bar represents CDT, while the blue bar represents typical TAD. The grey bar
726 represents flanking typical TAD. CDT: cancer driver TAD.

727 **E**. Boxplot to indicate the TAD length of CDT and typical TAD are almost the same in T47D. P
728 value is determined by two tail Wilcoxon test.

729 **F**. Bar plot to show the ratio that the number of CRLs divided by TAD length in each TAD of
730 T47D. The red bar represents CDT, while the blue bar represents typical TAD. The grey bar
731 represents flanking typical TAD. CDT: cancer driver TAD.

732 **G**. The two-dimensional heatmap shows the normalized Hi-C interaction frequencies in LNCaP.
733 Black line box out the calculated TAD. Blue rectangles represent genes.

734 **H**. The two-dimensional heatmap shows the normalized Hi-C interaction frequencies in T47D.
735 Black line box out the calculated TAD. Blue rectangles represent genes.

736 **I**. Heatmap to show the CDT characteristic of the driver lncRNAs identified by ExInAtor2. The
737 red box represents that lncRNA has this characteristic. The gray box represents that lncRNA
738 doesn't have this characteristic. Underlined lncRNAs represent known driver lncRNAs.

739 **J**. Boxplot to show the difference of absolute Spearman correlation between I-CRLs/O-CRLs and
740 cancer driver protein-coding genes in the same TAD. I-CRL: cancer-related lncRNAs in CDT. O-
741 CRL: cancer-related lncRNAs outside CDT. P values are determined by one tail Wilcoxon test.

742 **K**. GSEA analysis shows enrichment of the I-CRLs and O-CRLs in hub lncRNAs in a cancer-
743 related lncRNA-mRNA network.
744 **L**. Boxplot to indicate the difference of centrality score between I-CRLs and O-CRLs. P value is
745 determined by one tail Wilcoxon test.
746

747 **Figure 2**

748 **CRLs have a higher bind degree with cancer driver protein-coding genes in the same TAD.**
749 **A**. Boxplot to show the difference of triplex number between the experimental pair and 2 control
750 pairs. P values are determined by one tail Wilcoxon test.
751 **B**. Boxplot to show the difference of R-loop number between the experimental pair and 2 control
752 pairs. P values are determined by one tail Wilcoxon test.
753 **C**. Cartoon pattern to show the definition of the 4 lncRNA type.
754 **D**. Boxplot to show the difference of triplex number and R-loop number among I-CRL, I-TL, O-
755 CRL, and O-TL. I-CRL: cancer-related lncRNAs in CDT. I-TL: typical lncRNAs in CDT. O-CRL:
756 cancer-related lncRNAs outside CDT. O-TL: typical lncRNAs outside CDT. P values determined
757 by one tail Wilcoxon test.
758 **E**. Boxplot to show the absolution Spearman expression correlation between predicted binding
759 pairs and no binding pairs in the same CDT about triplex. P values are determined by one tail
760 Wilcoxon test.
761 **F**. Boxplot to show the absolution Spearman expression correlation between predicted binding
762 pairs and no binding pairs in the same CDT about R-loop. P values are determined by one tail
763 Wilcoxon test.
764

765 **Figure 3**

766 **Mine of pan-cancer driver lncRNAs and validate their cancer characteristics using genomics**
767 **and epigenomics data.**
768 **A**. PCA plot to show the difference of TAD among normal blood samples, blood cancers, normal
769 solid samples, and solid tumors. The blue dots represent normal blood samples. The yellow
770 triangles represent blood cancers. The gray rectangles represent normal solid samples. The red
771 crosses represent solid tumors.
772 **B**. The flowchart of mining pan-cancer driver lncRNAs.

773 **C**. Bar plot to show enrichment of the mined pan-cancer driver lncRNAs in CLC2 and Lnc2Cancer
774 CRL datasets. P values are determined by one tail Fisher's exact test.
775 **D**. Bar plot to show the enrichment of the mined oncogenic lncRNAs in up-regulated lncRNAs in
776 cancer, and mined tumor suppressive lncRNAs in down-regulated lncRNAs in cancer. P values
777 are determined by one tail Fisher's exact test.
778 **E**. Bar plot showing the enrichment of the mined pan-cancer driver lncRNAs in nucleus-featured
779 lncRNAs. P value is determined by one tail Fisher's exact test.
780 **F**. Empirical cumulative distribution plot to show the difference of pathogenic variations between
781 putative CDLs and the rest other lncRNAs. CDL: cancer driver lncRNAs. P value is determined
782 by one tail Wilcoxon test.
783 **G**. Empirical cumulative distribution plot to show the difference of H3K27me3 width between
784 candidate oncogenic lncRNAs and the rest other lncRNAs. P value is determined by one tail
785 Wilcoxon test.
786 **H**. Empirical cumulative distribution plot to show the difference of H3K4me3 width between
787 candidate tumor suppressive lncRNAs and the rest other lncRNAs. P value is determined by one
788 tail Wilcoxon test.
789

790 **Figure 4**

791 **Mine of prostate cancer driver lncRNAs and validate their cancer characteristics using**
792 **genomics and epigenomics data.**
793 **A**. Bar plot to show the statistics of TAD sample numbers about individual cancers.
794 **B.** Upset plot to show the statistic of tumor suppressor genes, and oncogenes in prostate cancer.
795 **C**. The flowchart of mining CDLs in prostate cancer.
796 **D**. Bar plot to show the enrichment of the mined oncogenic lncRNAs in up-regulated lncRNAs in
797 prostate cancer, and mined tumor suppressive lncRNAs in down-regulated lncRNAs in prostate
798 cancer. P values are determined by one tail Fisher's exact test.
799 **E**. Bar plot showing the enrichment of the mined CDLs about prostate cancer in nucleus-featured
800 lncRNAs. P value is determined by one tail Fisher's exact test.
801 **F**. Empirical cumulative distribution plot to show the difference of pathogenic variations between
802 putative CDLs and the rest other lncRNAs. CDL: cancer driver lncRNAs. P value is determined
803 by one tail Wilcoxon test.

804 **G**. Empirical cumulative distribution plot to show the difference of H3K27me3 width between
805 candidate oncogenic lncRNAs and the rest other lncRNAs, and H3K4me3 width between tumor
806 candidate tumor suppressive lncRNAs and the rest other lncRNAs. P values are determined by one
807 tail Wilcoxon test.
808

809 **Figure 5**

810 **Depletion of identified putative CDLs impaired or promoted cancer phenotypes of tumor**
811 **cells.**

812 LncRNA expression level and cell proliferation of prostate cancer cell line LNCaP with each
813 putative oncogenic lncRNA (**A**) or tumor suppressive lncRNA (**B**) disrupted by CRISPR-Cas9
814 guide RNA g1, g2, or under control condition. Bar plot to show lncRNA expression level and line
815 plot to show cell proliferation.
816

817 **Figure 6**

818 **Depletion of identified putative CDLs slowed or quickened wound closure of tumor cells.**

819 Wound healing assay of PC-3 prostate cancer cells with each putative oncogenic lncRNA (**A**) or
820 tumor suppressive lncRNA (**B**) disrupted by CRISPR-Cas9 guide RNA g1, g2, or under control
821 condition.

Fig 1

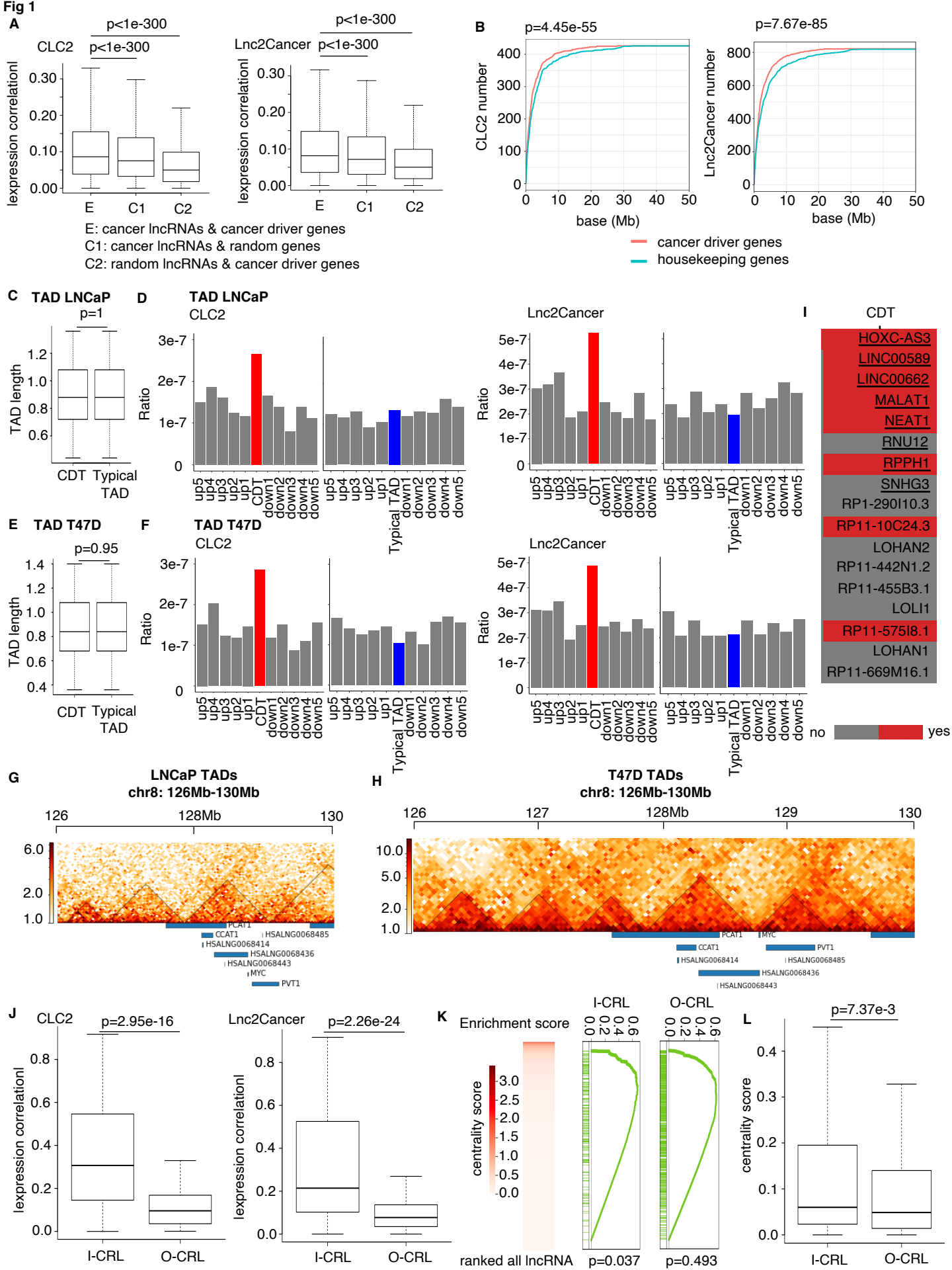

A
p<1e-300
CLC2
p<1e-300
lexpression correlationl
E
C1
C2
Lnc2Cancer
p<1e-300
E: cancer lncRNAs & cancer driver genes
C1: cancer lncRNAs & random genes
C2: random lncRNAs & cancer driver genes
B
p=4.45e-55
p=7.67e-85
CLC2 number
Lnc2Cancer number
base (Mb)
cancer driver genes
housekeeping genes
C
TAD LNCaP
p=1
TAD length
CDT
Typical TAD
D
TAD LNCaP
CLC2
Lnc2Cancer
Ratio
E
TAD T47D
p=0.95
F
TAD T47D
I
CDT
HOXC-AS3
LINC00589
LINC00662
MALAT1
NEAT1
RNU12
RPPH1
SNHG3
RP1-290I10.3
RP11-10C24.3
LOHAN2
RP11-442N1.2
RP11-455B3.1
LOLI1
RP11-575I8.1
LOHAN1
RP11-669M16.1
no
yes
G
LNCaP TADs
chr8: 126Mb-130Mb
126
128Mb
130
PCAT1
CCAT1
HSALNG0068485
HSALNG0068414
HSALNG0068436
HSALNG0068443
MYC
PVT1
H
T47D TADs
chr8: 126Mb-130Mb
127
129
J
CLC2
p=2.95e-16
Lnc2Cancer
p=2.26e-24
I-CRL
O-CRL
K
Enrichment score
centrality score
ranked all lncRNA
p=0.037
p=0.493
L
p=7.37e-3

Fig 2

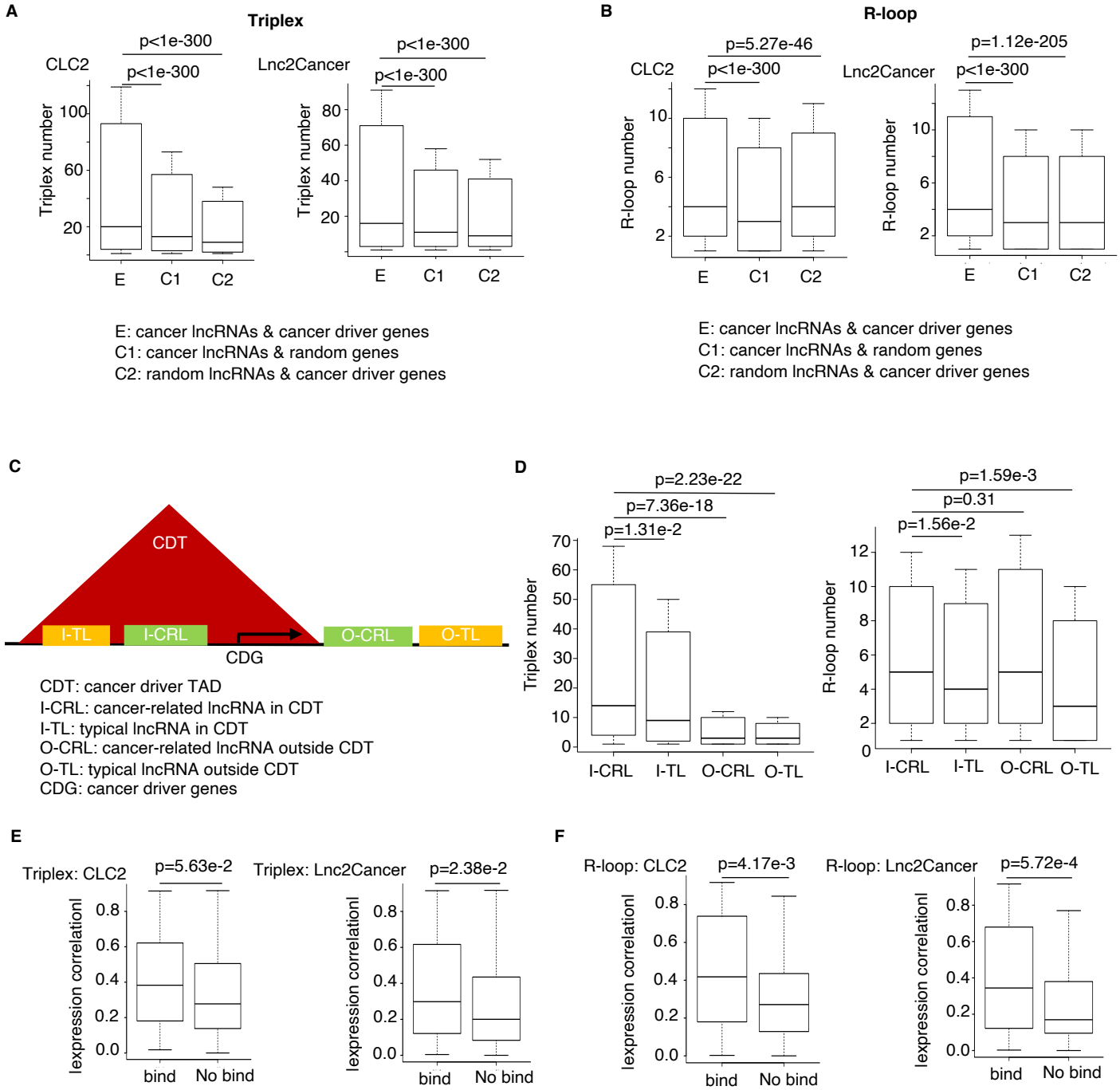

A
Triplex
CLC2
p<1e-300
p<1e-300
Triplex number
100
60
20
E
C1
C2
Lnc2Cancer
p<1e-300
p<1e-300
Triplex number
80
60
40
20
E
C1
C2
E: cancer lncRNAs & cancer driver genes
C1: cancer lncRNAs & random genes
C2: random lncRNAs & cancer driver genes
B
R-loop
CLC2
p=5.27e-46
p<1e-300
R-loop number
10
6
2
E
C1
C2
Lnc2Cancer
p=1.12e-205
p<1e-300
R-loop number
10
6
2
E
C1
C2
E: cancer lncRNAs & cancer driver genes
C1: cancer lncRNAs & random genes
C2: random lncRNAs & cancer driver genes
C
CDT
I-TL
I-CRL
CDG
O-CRL
O-TL
CDT: cancer driver TAD
I-CRL: cancer-related lncRNA in CDT
I-TL: typical lncRNA in CDT
O-CRL: cancer-related lncRNA outside CDT
O-TL: typical lncRNA outside CDT
CDG: cancer driver genes
D
p=2.23e-22
p=7.36e-18
p=1.31e-2
Triplex number
70
60
50
40
30
20
10
0
I-CRL
I-TL
O-CRL
O-TL
p=1.59e-3
p=0.31
p=1.56e-2
R-loop number
12
10
8
6
4
2
0
I-CRL
I-TL
O-CRL
O-TL
E
Triplex: CLC2
p=5.63e-2
lexpression correlationl
0.8
0.6
0.4
0.2
0.0
bind
No bind
Triplex: Lnc2Cancer
p=2.38e-2
lexpression correlationl
0.8
0.6
0.4
0.2
0.0
bind
No bind
F
R-loop: CLC2
p=4.17e-3
lexpression correlationl
0.8
0.6
0.4
0.2
0.0
bind
No bind
R-loop: Lnc2Cancer
p=5.72e-4
lexpression correlationl
0.8
0.6
0.4
0.2
0.0
bind
No bind

Fig 3

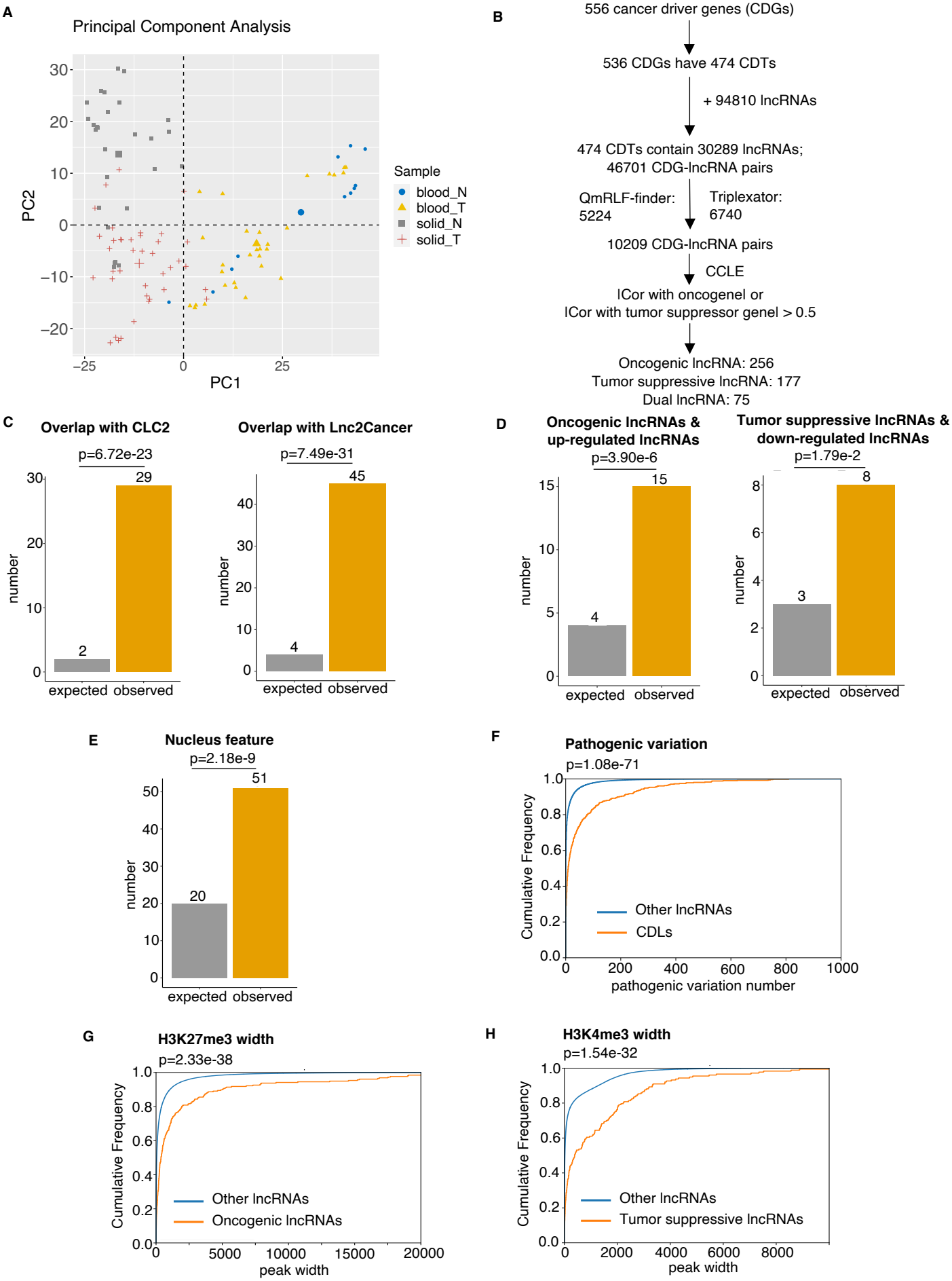

A
Principal Component Analysis
PC2
PC1
Sample
blood_N
blood_T
solid_N
solid_T
B
556 cancer driver genes (CDGs)
536 CDGs have 474 CDTs
+ 94810 lncRNAs
474 CDTs contain 30289 lncRNAs;
46701 CDG-lncRNA pairs
QmRLF-finder:
5224
Triplexator:
6740
10209 CDG-lncRNA pairs
CCLE
lCor with oncogenel or
lCor with tumor suppressor genel > 0.5
Oncogenic lncRNA: 256
Tumor suppressive lncRNA: 177
Dual lncRNA: 75
C
Overlap with CLC2
p=6.72e-23
Overlap with Lnc2Cancer
p=7.49e-31
number
expected
observed
D
Oncogenic lncRNAs &
up-regulated lncRNAs
p=3.90e-6
Tumor suppressive lncRNAs &
down-regulated lncRNAs
p=1.79e-2
E
Nucleus feature
p=2.18e-9
F
Pathogenic variation
p=1.08e-71
Cumulative Frequency
Other lncRNAs
CDLs
pathogenic variation number
G
H3K27me3 width
p=2.33e-38
Oncogenic lncRNAs
peak width
H
H3K4me3 width
p=1.54e-32
Tumor suppressive lncRNAs

Fig 4

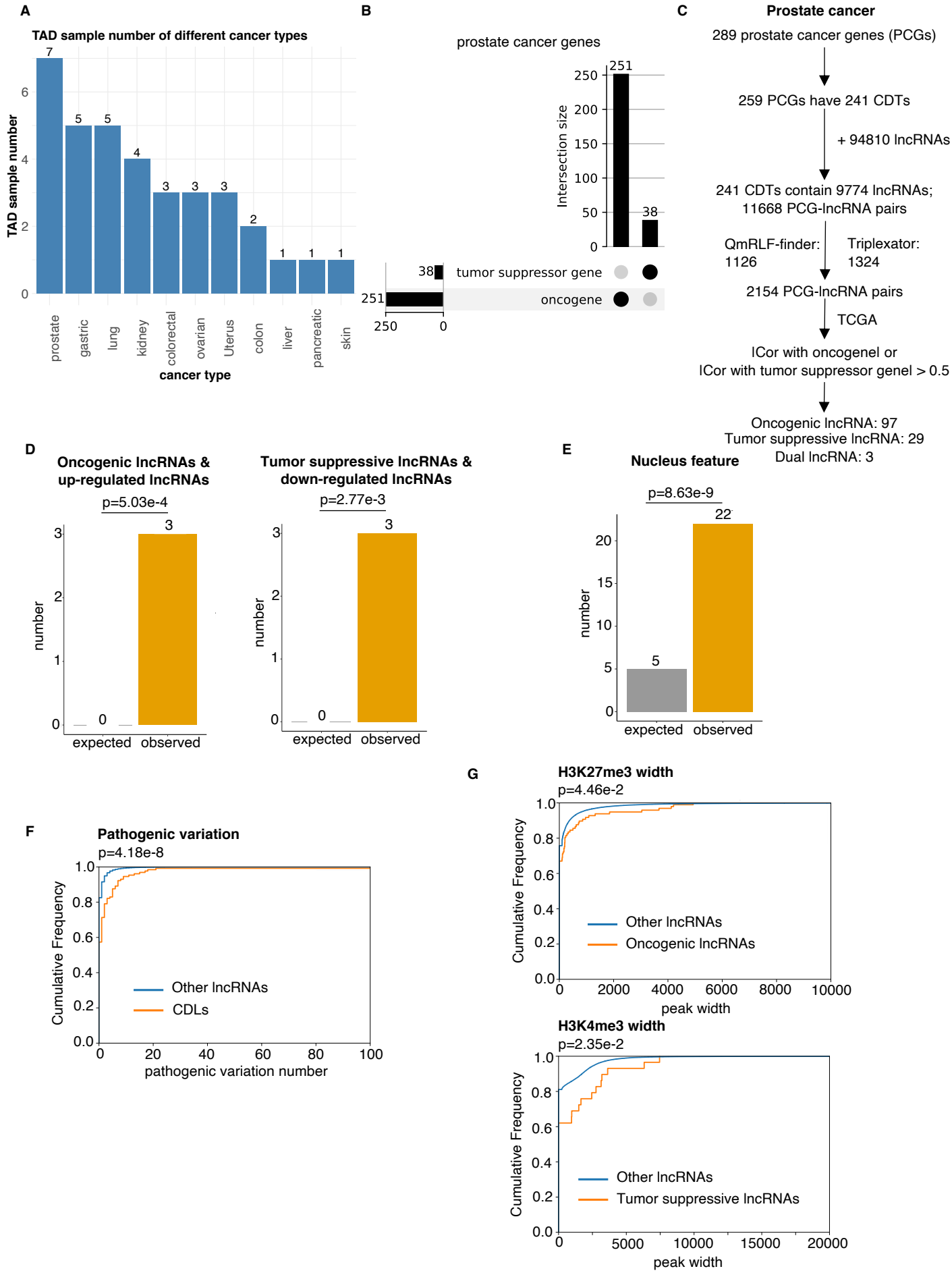

A
TAD sample number of different cancer types
TAD sample number
cancer type
prostate
gastric
lung
kidney
colorectal
ovarian
Uterus
colon
liver
pancreatic
skin
B
prostate cancer genes
Intersection size
tumor suppressor gene
oncogene
C
Prostate cancer
289 prostate cancer genes (PCGs)
259 PCGs have 241 CDTs
+ 94810 lncRNAs
241 CDTs contain 9774 lncRNAs;
11668 PCG-lncRNA pairs
QmRLF-finder: 1126
Triplexator: 1324
2154 PCG-lncRNA pairs
TCGA
lCor with oncogenel or
lCor with tumor suppressor genel > 0.5
Oncogenic lncRNA: 97
Tumor suppressive lncRNA: 29
Dual lncRNA: 3
D
Oncogenic lncRNAs & up-regulated lncRNAs
p=5.03e-4
Tumor suppressive lncRNAs & down-regulated lncRNAs
p=2.77e-3
number
expected
observed
E
Nucleus feature
p=8.63e-9
F
Pathogenic variation
p=4.18e-8
Cumulative Frequency
Other lncRNAs
CDLs
pathogenic variation number
G
H3K27me3 width
p=4.46e-2
Oncogenic lncRNAs
peak width
H3K4me3 width
p=2.35e-2
Tumor suppressive lncRNAs

Fig 5

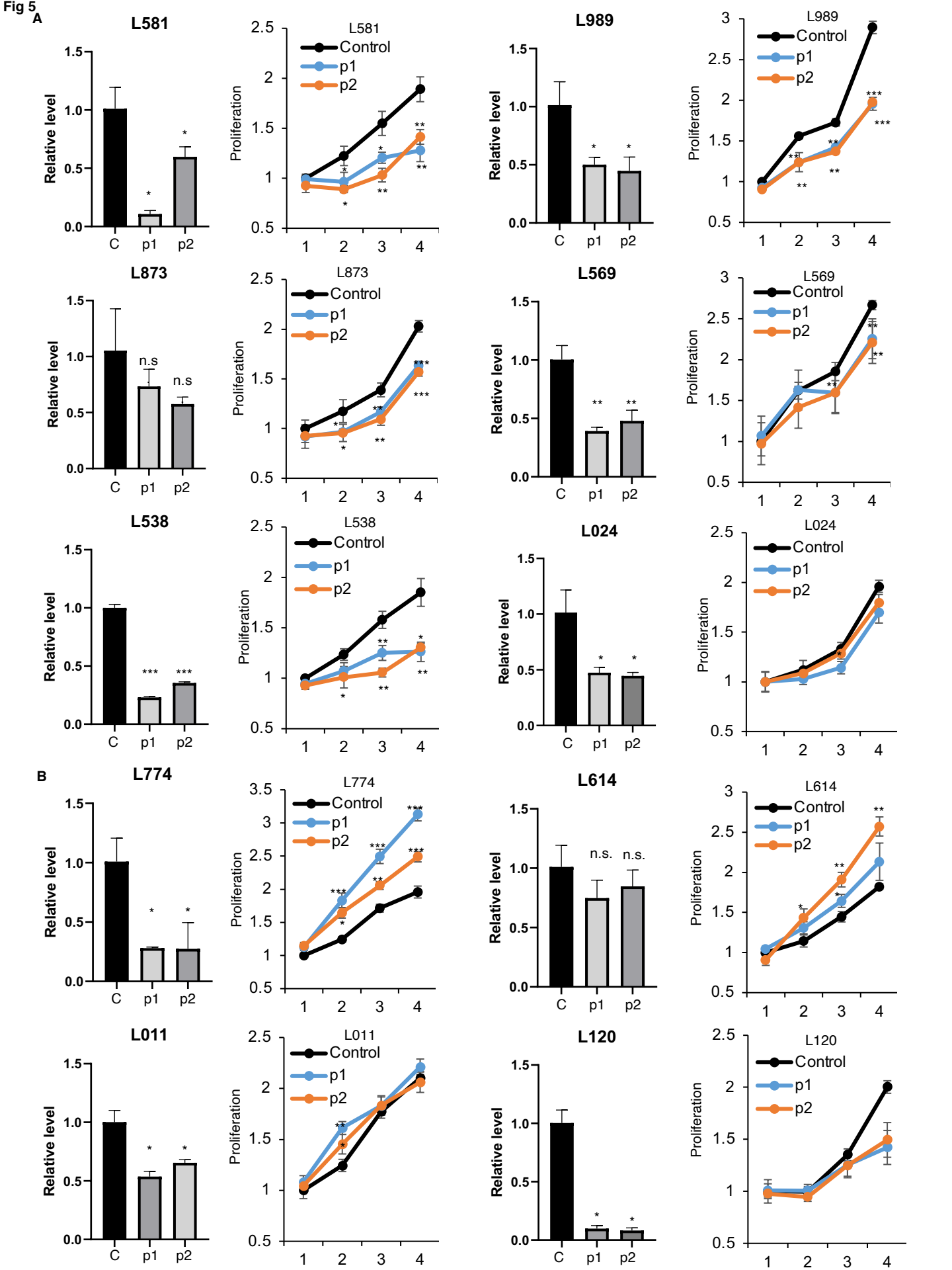
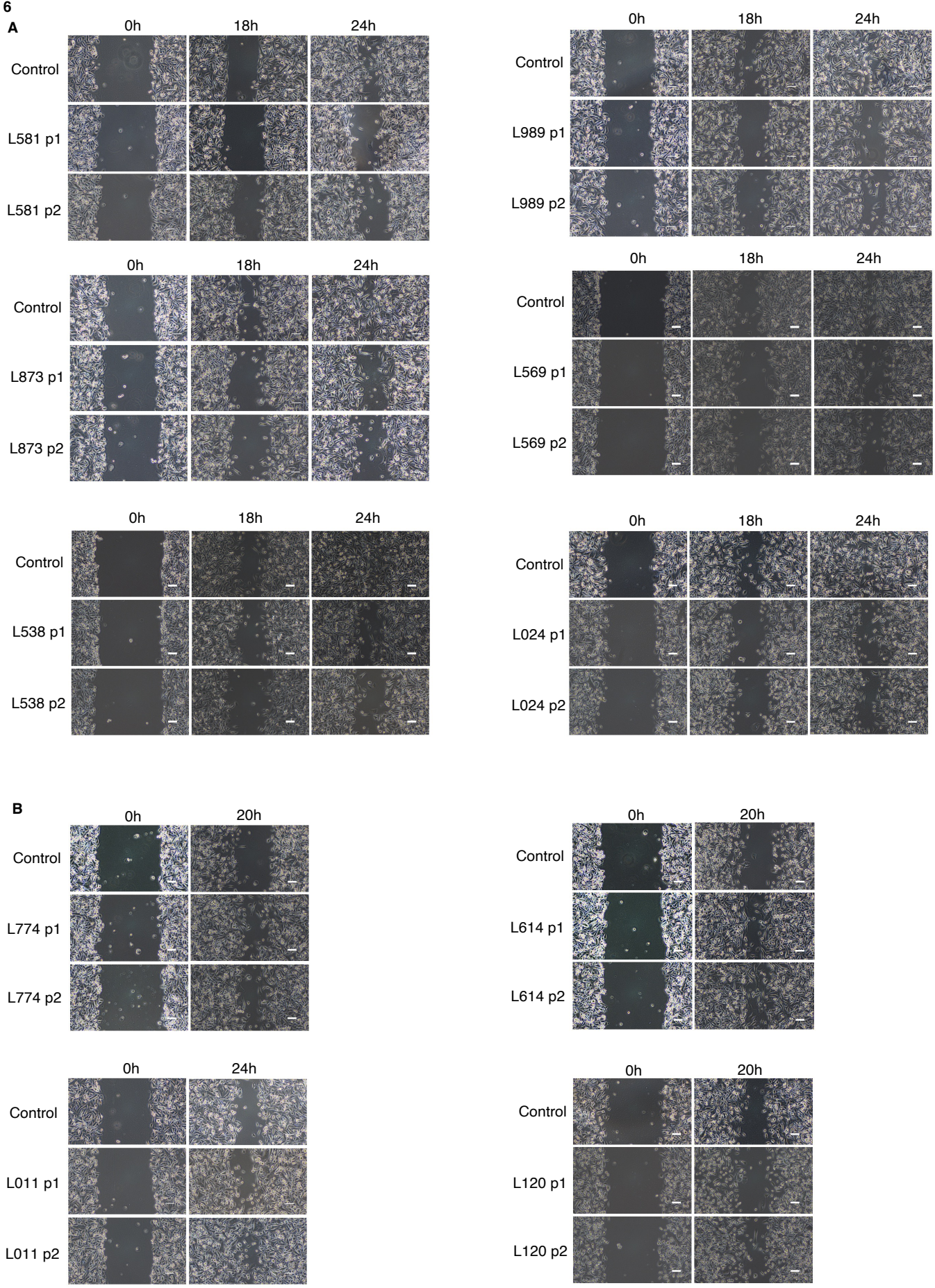

A
L581
Relative level
C
p1
p2
L581
Control
p1
p2
Proliferation
L989
L989
L873
L873
n.s
n.s
L569
L569
L538
L538
L024
L024
B
L774
L774
L614
L614
n.s.
n.s.
L011
L011
L120
L120

Fig 6


A
0h
18h
24h
Control
L581 p1
L581 p2
L989 p1
L989 p2
L873 p1
L873 p2
L569 p1
L569 p2
L538 p1
L538 p2
L024 p1
L024 p2
B
20h
L774 p1
L774 p2
L614 p1
L614 p2
L011 p1
L011 p2
L120 p1
L120 p2