



Article

Selective Concurrence of the Long Non-Coding RNA MALAT1 and the Polycomb Repressive Complex 2 to Promoter Regions of Active Genes in MCF7 Breast Cancer Cells

Felipe Arratia¹, Cristopher Fierro¹, Alejandro Blanco¹, Sebastian Fuentes¹, Daniela Nahuelquen¹, Martin Montecino¹ , Adriana Rojas² and Rodrigo Aguilar^{1,*}

¹ Institute of Biomedical Sciences (ICB), Faculty of Medicine and Faculty of Life Sciences, Universidad Andres Bello, Santiago 8370071, Chile

² Institute of Human Genetics, Faculty of Medicine, Pontificia Universidad Javeriana, Bogotá 110211, Colombia

* Correspondence: rodrigo.aguilar@unab.cl

Abstract: In cancer cells, the long non-coding RNA (lncRNA) MALAT1 has arisen as a key partner for the Polycomb Repressive Complex 2 (PRC2), an epigenetic modifier. However, it is unknown whether this partnership occurs genome-wide at the chromatin level, as most of the studies focus on single genes that are usually repressed. Due to the genomic binding properties of both macromolecules, we wondered whether there are binding sites shared by PRC2 and MALAT1. Using public genome-binding datasets for PRC2 and MALAT1 derived from independent ChIP- and CHART-seq experiments performed with the breast cancer cell line MCF7, we searched for regions containing PRC2 and MALAT1 overlapping peaks. Peak calls for each molecule were performed using MACS2 and then overlapping peaks were identified by bedtools intersect. Using this approach, we identified 1293 genomic sites where PRC2 and MALAT1 concur. Interestingly, 54.75% of those sites are within gene promoter regions (<3000 bases from the TSS). These analyses were also linked with the transcription profiles of MCF7 cells, obtained from public RNA-seq data. Hence, it is suggested that MALAT1 and PRC2 can concomitantly bind to promoters of actively-transcribed genes in MCF7 cells. Gene ontology analyses revealed an enrichment of genes related to categories including cancer malignancy and epigenetic regulation. Thus, by re-visiting occupancy and transcriptomic data, we identified a key gene subset controlled by the collaboration of MALAT1 and PRC2.

Keywords: long non-coding RNA; MALAT1; Polycomb Repressive Complex 2; chromatin



Citation: Arratia, F.; Fierro, C.; Blanco, A.; Fuentes, S.; Nahuelquen, D.; Montecino, M.; Rojas, A.; Aguilar, R. Selective Concurrence of the Long Non-Coding RNA MALAT1 and the Polycomb Repressive Complex 2 to Promoter Regions of Active Genes in MCF7 Breast Cancer Cells. *Curr. Issues Mol. Biol.* **2023**, *45*, 4735–4748. <https://doi.org/10.3390/cimb45060301>

Academic Editor: Yu Sun

Received: 17 March 2023

Revised: 23 May 2023

Accepted: 28 May 2023

Published: 30 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Polycomb Repressive Complex 2 (PRC2) is an epigenetic modifier that catalyzes the methylation of lysine 27 on histone H3 [1]. The presence of the complex and its mark in gene regulatory regions often leads to gene repression [2]. We and others have shown that such repression is required for successful cellular differentiation and development [1,3]. However, this complex has also been shown to bind to nascent transcripts in transcribing genes [4]. PRC2 is constituted of a core of at least three proteins: EZH2 (the catalytic subunit, which can be replaced by EZH1 in postmitotic cells such as neurons [5]), SUZ12 and EED, plus additional accessory subunits [1]. The misregulation of PRC2 is found in several types of cancer cells, including breast cancer, contributing to the nonmutational epigenetic reprogramming that facilitates the acquisition and maintenance of malignancy [6]. Currently, drugs targeting different sub-units of PRC2 are under development as cancer treatment alternatives [7].

In the epigenetics field, how PRC2 is targeted to specific genes is still a matter of active discussion [8]. Since the Polycomb response elements originally identified in *Drosophila* are not conserved in mammalian genomes, different research groups have unsuccessfully attempted to uncover DNA methylation- or protein-dependent mechanisms that apply

to the whole genome. Two seminal papers from the late 2000s reported that the long non-coding RNAs (lncRNAs) *Hotair* [9] and *Xist* [10] were sufficient for recruiting PRC2 to the *HOX* locus and the inactive X chromosome, respectively, portraying lncRNA-PRC2 interactions as a paradigm of molecular collaboration, at least at those specific genomic regions. Recently, the *Xist*-PRC2 interaction was specifically disrupted by novel small molecules designed to target the lncRNA [11].

Another PRC2-interacting lncRNA is MALAT1. This transcript has been under the spotlight of cancer biologists since its discovery in 2003 in small lung cancer cells [12]. In subsequent studies, MALAT1 was found in several additional types of tumors, including breast cancer, correlating with poor prognosis. Thus, MALAT1 has arisen as an archetype of lncRNA function in malignancies [13,14]. As for PRC2, MALAT1 overexpression has been proposed as a cancer biomarker [15], and it currently represents a promising therapeutic target to treat tumor progression [16].

There is increasing evidence indicating that MALAT1 partners with chromatin-modifying enzymes, including the histone deacetylase HDAC9, the chromatin remodeler BRG1, or the RNA modifier METTL16 [17,18]. PRC2 is, however, one of the most studied MALAT1 partners in tumors [19]: first, RNA immunoprecipitation (RIP) experiments performed on cancer cell extracts show that both molecules are part of a complex in the nucleoplasm [20–22]; second, genes such as *PCDH10* and *E-cadherin* are derepressed in cancer cells following MALAT1 or PRC2 down-regulation [21–23]. Additional works show that the MALAT1–PRC2 interaction seems to be dependent on the 3′-end of the lncRNA in cancer cells [23], but whether other complexes follow similar recruitment mechanisms is unclear.

One key question that remains to be addressed in cancer cells is whether the MALAT1–PRC2 partnership occurs in chromatin at a genome-wide level and whether this necessarily results in gene repression. Since the PRC2 subunit *EZH2* can bind and regulate target genes as a single protein by histone methylation-independent mechanisms [24], experiments focusing on the chromatin-binding profile of the core subunit *SUZ12* are better suited for studying the interaction sites of the full PRC2 [25]. Unlike chromatin immunoprecipitation (ChIP), which is the gold standard for the study of protein–DNA binding in vivo [26], examples of lncRNA-capture experiments (CHART, ChIRP or RAP) to define genomic binding sites for lncRNAs are scarce in the literature and databases. This is due to the challenge of capturing specific RNA molecules and performing the deep sequencing of the DNA associated with them [27–30]. Hence, up to the present date, reports claiming that a MALAT1–PRC2 complex modulates the expression of key target genes in cancer cells are based on the assumption that MALAT1 binds to that particular DNA sequence (e.g., [22]). Moreover, only a few authors have performed the experiments required to show the global MALAT1 binding profile in humans [28] or mice [27,29,30].

Here, we revisit two genome-wide sequencing studies performed by independent laboratories providing a novel perspective: we unified a previous report that used CHART to identify MALAT1 binding sites in MCF7 breast cancer cells [28] with an independent study performed in the same cell line evaluating the binding of the *SUZ12* subunit of PRC2 to the genome [31]. We searched in silico for genes associated with overlapped peaks of PRC2 and MALAT1, aiming to identify a set of genes and main Gene Ontology categories, where MALAT1-PRC2 may collaborate to regulate transcription in MCF7 cells. Our results indicate that a number of genes bound by MALAT1–PRC2 are enriched in the GO categories of cancer malignancy and epigenetic regulation, and they are mostly actively transcribed in MCF7 breast cancer cells.

2. Materials and Methods

2.1. CHIP- and CHART-Seq Datasets

The original MALAT1 CHART-seq experiments were performed by the Robert Kingston lab [28] in the MCF7 breast cancer cell line [32] and the raw sequencing data were deposited as Sequence Read Archives (SRA) that were recovered using the following run numbers:

Replica 1: SRR1386233 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1411209> (accessed on 1 July 2021) [33]); Replica 2: SRR1386234 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1411210> (accessed on 1 July 2021) [34]); Input: SRR1386236 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1411212> (accessed on 1 July 2021) [35]). The SUZ12 ChIP-seq experiments were performed by the Michael Snyder lab and the raw SRA were recovered from ENCODE [31] using the following run numbers: Replica 1: SRR6214195 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2828862> (accessed on 1 July 2021) [36]); replica 2: SRR6214196 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2828863> (accessed on 1 July 2021) [37]), input: SRR5111271 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2423179> (accessed on 1 July 2021) [38]).

2.2. Quality Control, Alignment and Peak Calling

Since MALAT1 and SUZ12 are broadly distributed throughout the genome [27,28], the fastq sequencing archives were subjected to the ENCODE 3 histone ChIP-seq pipeline in histone (non-transcription factor) mode (<https://github.com/ENCODE-DCC/chip-seq-pipeline2> (accessed on 1 July 2021)) [39]; see also [40]. Briefly, fastq sequence files derived from SRAs were cropped using Trimmomatic v0.39 [41]. Reads were aligned against the human hg38 genome using Bowtie2 v2.2.6 [42] with paired-end ChIP-seq parameters. Unmapped reads, multi-mapped reads, and duplicates were removed using MarkDuplicates in Picard v1.126 (<http://broadinstitute.github.io/picard/> (accessed on 1 July 2021)) [43]. Peak calling was performed using MACS2 [44] in paired-end mode, selecting the optimal peak set for subsequent experiments. Bed and subsequent Bigwig files were loaded into the Integrative Genomics Viewer (IGV) v2.8 or above to visualize peaks and signals.

2.3. Detection of Overlapped Peaks and Gene Ontology Analyses

To identify regions of the genome where MALAT1 and SUZ12 peaks overlap in at least 1 nucleotide, optimal peak bed files were analyzed with bedtools intersect v2.29.2 or higher (<https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html> (accessed on 1 July 2021)) [45] using the default parameters (“bedtools intersect -a malat1_peaks.bed -b suz12_peaks.bed >> intersect_malat1_vs_suz12.bed”). Peak annotation was performed using the R package ChIPseeker v3.10 or higher (<https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html> (accessed on 1 July 2021)) [46] using “flankDistance = 5000”. The nearest genes around peaks were analyzed using the gene ontology enrichment analysis tool from PANTHER Classification System v16.0 or higher (<http://pantherdb.org/> (accessed on 1 July 2021)) [47].

2.4. RNA-Seq Analyses

Two independent transcriptomic analyses of MCF7 cells were performed from raw sequencing data deposited as SRAs that were recovered using the following numbers: SRR2749729 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1915041> (accessed on 1 November 2022) [48]) [49] and SRR925723 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1172885> (accessed on 1 November 2022) [50]) [51]. Transcript level abundance was estimated using Salmon v1.9 (<https://salmon.readthedocs.io/en/latest/> (accessed on 1 November 2022)) [52] and expressed as Transcripts per Million (TPM). Transcripts were aggregated to the gene level using the R package tximport against the human transcriptome (<https://bioconductor.org/packages/release/bioc/html/tximport.html> (accessed on 1 November 2022)).

3. Results

Raw SRA sequence files from ChIP- [36–38] and CHART-seq [33–35] experiments were downloaded from the Gene Expression Omnibus and subjected to the ChIPseq processing pipeline recommended by ENCODE [39]. We confirmed that the deposited datasets were

suitable for this analysis after performing quality control checks showing sufficient mapped reads (>50 million for SUZ12 and MALAT1 after filtering) with acceptable reproducibility among replicates (key features shown in Supplementary Table S1).

We first analyzed the binding profile of each molecule separately. The number of optimal peaks (peaks consistently found in two replicates) was determined using an irreproducible discovery rate (IDR) value below 0.01 as filter. We identified 26,916 peaks for SUZ12 enrichment and 24,508 for MALAT1 (Figure 1a and Supplementary Table S1). These peaks were distributed throughout the human genome (Figure 1b) with no preference for promoter regions (Figure 1c). Instead, we found SUZ12 or MALAT1 preferentially located on introns and intergenic regions (Figure 1c). Additional metagene analyses revealed that, when focusing on a $-3000/+3000$ region (around TSSs), SUZ12 and MALAT1 showed a mutually exclusive enrichment pattern, with SUZ12 enriched over the TSS and MALAT1 voided from the same region (Figure 1d).

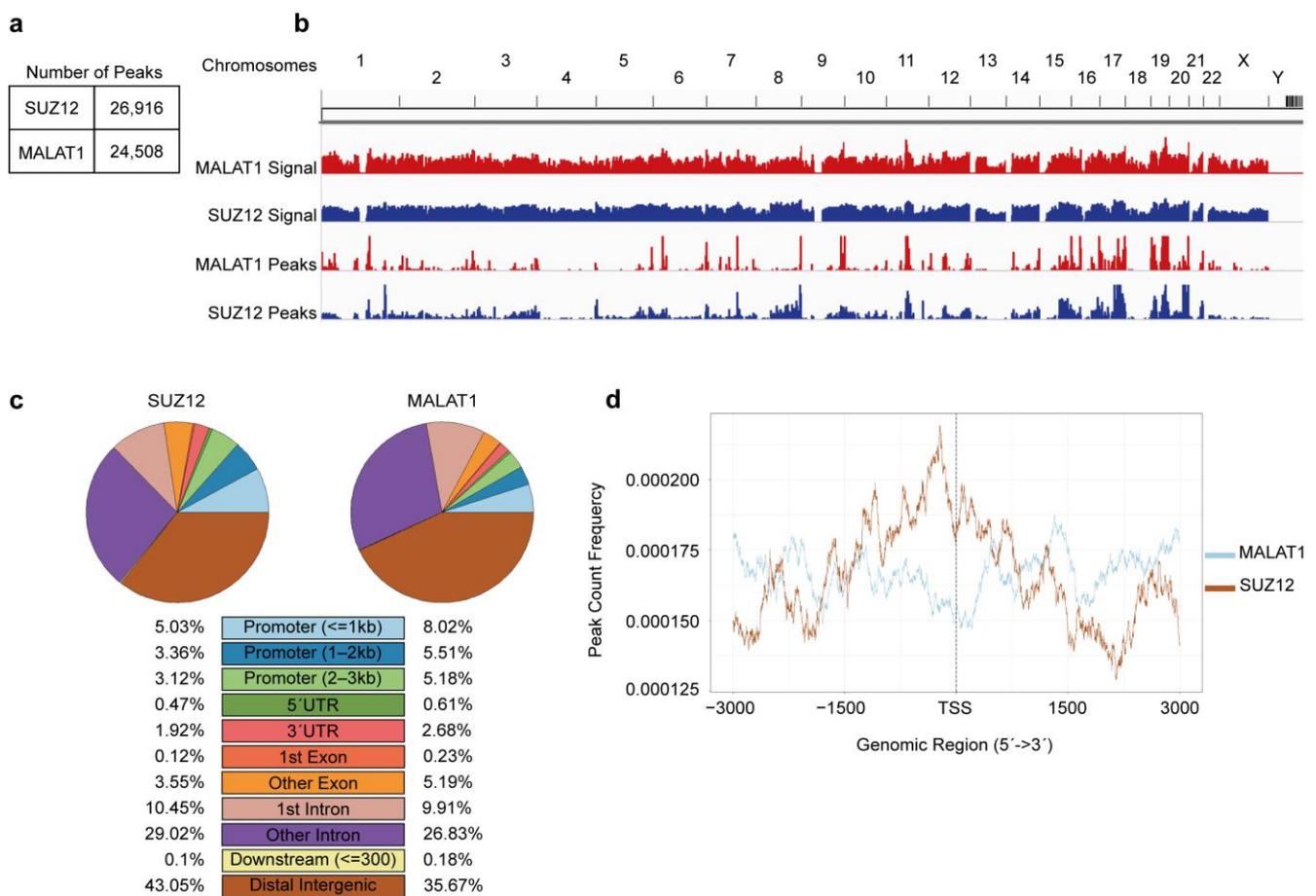


Figure 1. Genomic distribution of MALAT1 and SUZ12 in the MCF7 breast cancer cell line. (a) Raw sequencing data deposited in the Gene Expression Omnibus were subjected to the ChIPseq processing pipeline recommended by ENCODE. The number of reproducible peaks was determined for each molecule separately, as shown in the table. (b) Signal (bigwig) files and peak (bed) files were loaded into the Integrative Genomics Viewer. MALAT1 data are shown in red tracks and SUZ12 data in blue tracks. Chromosome numbers as shown on top of the graph. (c) Genomic annotations for optimal peaks were obtained using ChIPseeker. The percentage of peaks contained in each feature (colored rectangles) is shown for SUZ12 (left) and MALAT1 (right). (d) Metagene analyses were performed to evaluate the frequency of the peaks found around the transcription start site (TSS) for MALAT1 (blue line) and SUZ12 (brown line). Distance from the TSS in base pairs is shown in the x-axis.

We next assessed whether SUZ12 and MALAT1 present overlapped enrichment peaks throughout the genome. To identify those overlapping peaks, we applied bedtools intersect to the BED files containing the optimal peaks of SUZ12 and MALAT1. We found 1293 shared peaks between both molecules (Figure 2a,b), associated with 839 genes (Supplementary Table S2). Metagene analyses showed that the overlap of SUZ12 and MALAT1 occurs preferentially in genomic regions where a number of gene promoters are located, as 708 peaks (54.75% of the total number of peaks, which are associated with 531 genes) reside within the $-3000/+1$ region (Figure 2c, and Supplementary Table S2), with a notorious enrichment around the TSS (Figure 2d).

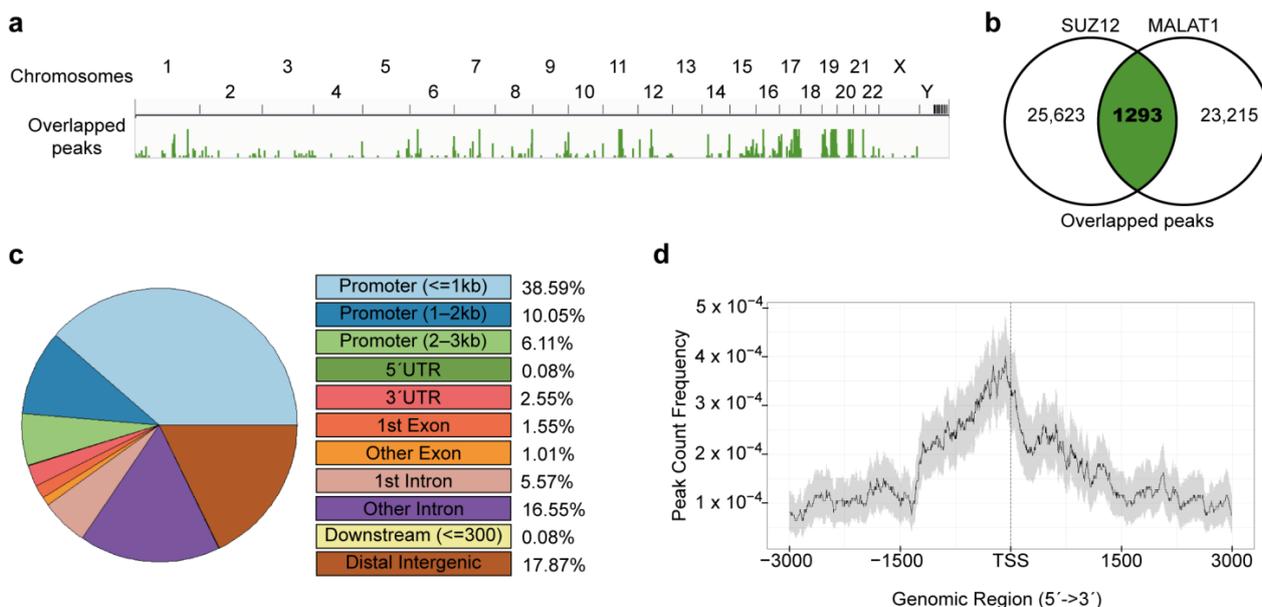


Figure 2. Genomic distribution of MALAT1-SUZ12 overlapped peaks in the MCF7 breast cancer cell line. (a) Reproducible peaks from SUZ12 and MALAT1 experiments were subjected to bedtools intersect to detect sites of overlap. The resulting bed file was loaded into the Integrative Genomics Viewer and is shown in green bars. Chromosome numbers as shown on top of the graph. (b) The number of peaks shared by SUZ12 and MALAT1 is shown with green background in the Venn diagram. (c) Genomic annotations for overlapped peaks were obtained using ChIPseeker. The percentage of peaks contained in each feature (colored rectangles) is shown. (d) Metagene analyses were performed to evaluate the frequency of the overlapped peaks (black line) found around the transcription start site (TSS) within a confidence interval of 95% (grey background). Distance from the TSS in base pairs is shown on the x-axis.

We then focused our analysis on a group of 403 genes where MALAT1 and PRC2 overlap immediately upstream of the TSS ($-1000/+1$ region, Supplementary Table S2), as this co-localization has a higher potential for establishing mechanisms that regulate the expression of those target gene promoters [1,4]. Thus, a gene ontology (GO) analysis was carried out by loading the IDs of these genes into the PANTHER 17.0 platform. As depicted in Figure 3a–c, we determined an over-representation of genes associated with cancer-relevant functions, including biological adhesion, locomotion (Figure 3a), transcriptional regulation (Figure 3b), and chromatin binding (Figure 3c). In Table 1, we highlight 10 genes that, among other functions, have been found participating in signaling (*MEN1*, *YWHAB*, *SEN3*, *MAPK15*, and *CADM4*), oxidative stress (*TXNIP*), acting as transcription factors (*ETS2*, *RARG*, and *VEZF1*) or chromatin binders (*MEN1*, *KDM6B*) in cancer cells. The enrichment of MALAT1 (Figure 3d, red tracks) and SUZ12 (Figure 3d, blue tracks) was visualized using the Integrative Genomics Viewer software. In addition, the position of overlapped peaks around the TSS (Figure 3d, green rectangle) on each of the genes listed in Table 1, was marked.

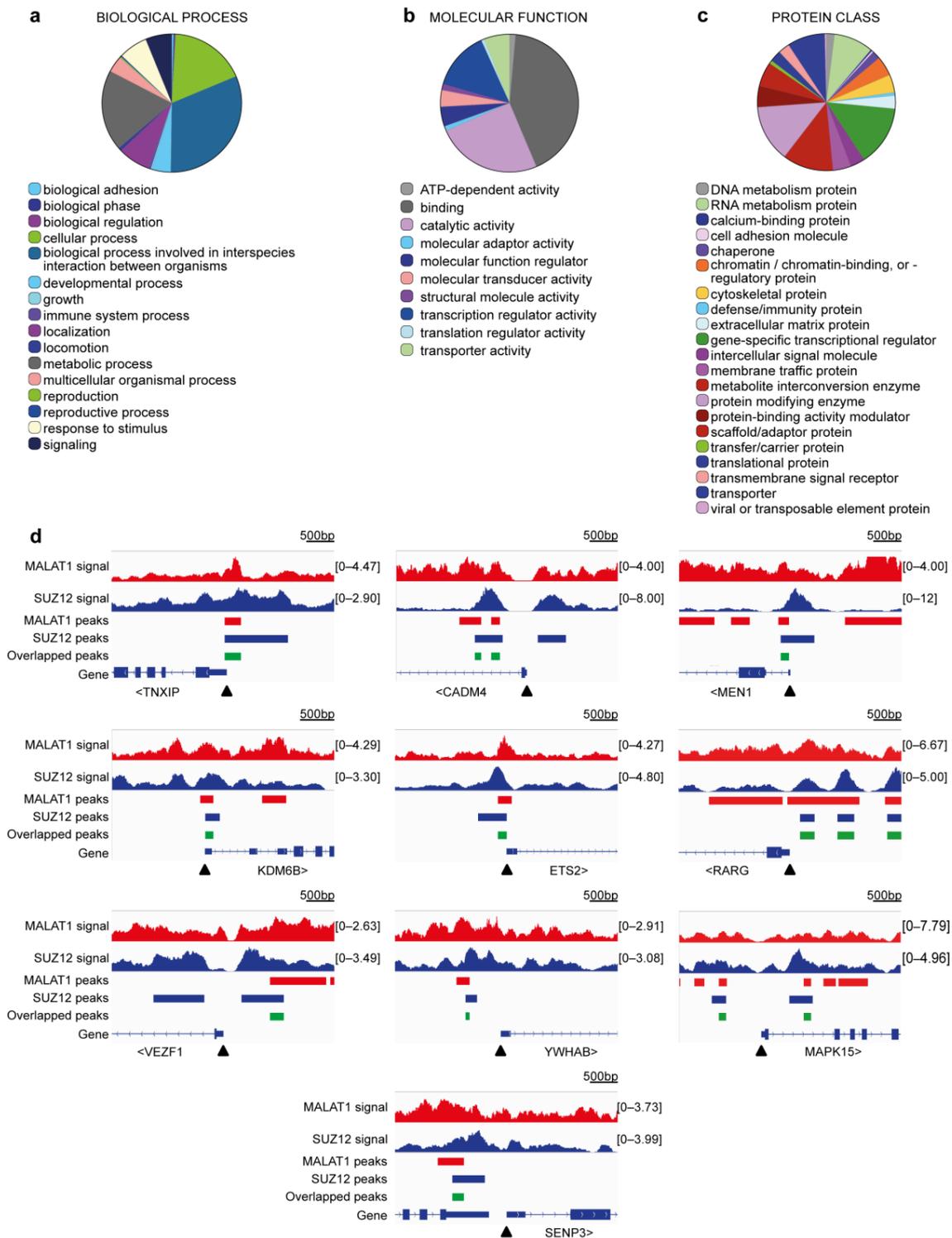


Figure 3. MALAT1 and SUZ12 concur to gene promoter regions of cancer-related genes. (a–c) Gene ontology analyses were performed in genes presenting MALAT1 and SUZ12 overlap within the 3000/+1 region. Categories for biological process (a), molecular function (b), and protein class (c) are shown. In (d) signal (bigwig) files and peak (bed) files were loaded into the Integrative Genomics Viewer. MALAT1 data are shown in red tracks, SUZ12 data in blue tracks, and overlap sites in green. Genes positions are shown, with arrowheads indicating the transcription start site. The fragments per million mapped fragments (FPM) scale is shown in brackets. Only one splice variant is shown for each gene. Scale bar: 500 bp.

Table 1. A subset of genes that present an overlap of MALAT1 and SUZ12 in the gene promoter region (see also Figure 3d).

Gene Symbol	Full Gene Name	Related Functions	Ref.
<i>MEN1</i>	Menin	Part of CCKR signaling. Chromatin binding (component of a MLL/SET1 histone methyltransferase (HMT) complex). Negative regulator of cell cycle. Negative regulation of proliferation. Inactivation favors tumorigenesis of mammary cells.	[53]
<i>YWHAB</i>	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein beta	Part of EGF, FGF, and CCKR signaling. Stimulates cell growth.	[54]
<i>SEN3</i>	SUMO specific peptidase 3	Involved in processing of sumoylated proteins. High levels are associated with poor survival in breast cancer.	[55]
<i>MAPK15</i>	Mitogen-activated protein kinase 15	Part of multiple signaling pathways, including: PDGF, TGF-beta, EGF, FGF, IFG. Associated with apoptosis. Regulates autophagy, ciliogenesis, protein trafficking/secretion and genome integrity. Downregulation activates cell motility in breast cancer.	[56]
<i>CADM4</i>	Cell adhesion molecule 4	Cell-cell adhesion. Associated to wound healing. Associated with cell growth. Low expression associated with advanced breast cancer stages.	[57]
<i>TXNIP</i>	Thioredoxin interacting protein	Oxidative stress mediator. Transcriptional repressor. Tumor suppressor. Low expression in breast cancer.	[58]
<i>ETS2</i>	ETS proto-oncogene 2, transcription factor	Transcription factor. Regulator of telomerase for breast cancer cell survival. Promotes hTERT expression.	[59]
<i>RARG</i>	Retinoic acid receptor gamma	Receptor for retinoic acid. Downregulated in tumors. Silencing causes tamoxifen resistance in breast cancer.	[60]
<i>VEZF1</i>	Vascular endothelial zinc finger 1	Regulator of angiogenesis in cancer.	[61]
<i>KDM6B</i>	Lysine demethylase 6B	Demethylates lysine 27 of histone H3. Regulates HOX expression. Inhibits metastasis in breast cancer cells	[62]

Finally, we evaluated the transcriptional status of the genes where MALAT1 and PRC2 overlap. Raw SRA sequencing files from RNA-seq experiments were downloaded from the Gene Expression Omnibus and transcript abundances were obtained using Salmon and aggregated to the gene level. When 25,935 gene transcripts expressed in MCF7 cells were considered [48] (after filtering out genes with TPM values = 0), their abundances in one RNA-seq replicate [48] were distributed as shown in Figure 4a, left (1st quartile = 183,340 TPM; median = 1,016,602 TPM; 3rd quartile = 6,692,790 TPM; average = 24,485,358 TPM). Another independent RNA-seq study [50] showed similar distribution (Supplementary Table S3). We then selected the genes where MALAT1 and SUZ12 overlap within or near the gene promoter region (<3000 bp from the TSS), where we found that 487 of the 531 genes showing overlapping also exhibited significant expression levels (>0 TPM). Transcript abundances in one replicate are distributed as shown in Figure 4a, right (1st quartile = 1,930,667 TPM; median = 10,344,484 TPM; 3rd quartile = 30,084,933 TPM; average = 66,526,895 TPM). This distribution was reproducible in the two independent RNA-seq experiments that we analyzed (Supplementary Table S3). Together, these data indicate that MALAT1 and PRC2 can be preferentially co-localized at promoters of actively-transcribed genes.

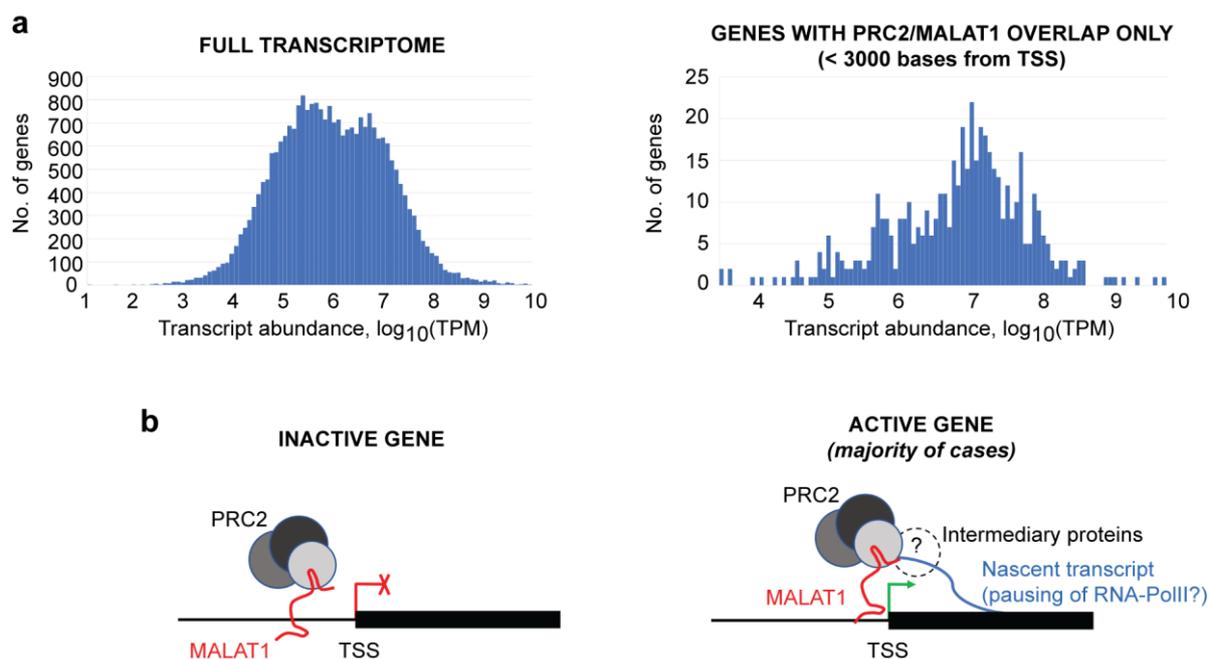


Figure 4. MALAT1 and SUZ12 concur to promoters of actively-transcribed genes. (a) transcriptomic abundance was determined from raw RNA sequencing files using Salmon and aggregated to the gene level. Left: frequency histogram showing all genes expressed in MCF7 cells with TPM (transcripts per kilobase million) > 0 derived from GSE74232/GSM1915041/SRR2749729 (total = 25,935 genes). Right: frequency histogram showing only the genes with MALAT1 and SUZ12 concurring to <3000 bp from TSS with TPM > 0 (total = 531 genes). (b) Scheme depicting MALAT1 and the PRC2 complex concurring to inactive (left) and active (right) genes. Our data indicate that, in most cases, MALAT1–PRC2 localize to active genes, where they may be interacting with nascent transcripts through unknown intermediary proteins (marked with “?”) (see Section 4).

4. Discussion

We present here an *in silico* study to identify genomic regions co-occupied by the lncRNA MALAT1 and the repressive epigenetic complex PRC2 in the MCF7 breast cancer cell line. We encountered a preferential concurrence of both macromolecules to genomic sequences near transcription start sites, often associated with actively-transcribed genes. A significant number of this set of target genes are related to signaling and gene expression control in cancer.

Since the discovery of MALAT1 in lung cancer cells [12], the epigenetic complex PRC2 has been proposed as a key regulatory partner [19], particularly after loss-of-function studies that measure the effect on specific target genes, where an independent knock-down of MALAT1 or of some of the PRC2 subunits results in gene de-repression [21–23]. The interaction of PRC2 and lncRNAs has been a matter of intense debate in the field, as some authors still consider the PRC2 catalytic subunit EZH2 as a promiscuous molecule, prone to bind to any RNA [63]. However, this idea has been challenged by several teams, proposing that PRC2–RNA complexes can collaborate in relevant regulatory mechanisms through specific interactions [64–66]. Therefore, there is a necessity for enlarging the number of studies that address, using multiple approaches, whether there is a binding of RNAs such as MALAT1 to DNA, narrowing the pool of genes where a potential PRC2–MALAT1 interaction may be mechanistically relevant.

We report that, despite the ubiquitous localization throughout the genome of both MALAT1 and the PRC2 subunit SUZ12, a clear overlap between the molecules is found only in a subset of 839 genes in MCF7 cells. There could be multiple mechanistic implications for this finding, and here we offer a few hypotheses that need further testing in the laboratory. First, these data support a model where MALAT1 may be involved in the

recruitment of PRC2 to a subset of genes in breast cancer cells. This resembles the roles of Xist and Hotair [9,10] except that, unlike these, this is not confined to the X chromosome or to the *Hox* locus, but occurs in a widespread fashion. Importantly, the preferential concurrence of both molecules to regions upstream and near the transcription start site (TSS) of actively-transcribed genes captured our attention. As mentioned above, reports proposing a MALAT—PRC2 complex focus almost exclusively on gene silencing [21–23], although this is in contrast to the seminal work demonstrating that MALAT1 is enriched on transcribed genes [28]. Together, these results may appear initially contradictory, but instead they argue in favor of a versatile role for RNA in general and for MALAT1 in particular, where these molecules can play a significant role during both gene activation and repression. Indeed, PRC2 can be found around the TSS, mediating at least two functions: H3K27me3 deposition, which leads to gene silencing [1] or *tempering* RNA Polymerase II elongation in actively-transcribed genes by binding to nascent transcripts [4,67]. The latter is further supported by a report showing that MALAT1 interacts with pre-mRNAs through RNA-binding proteins [29] which may include, according to our findings, PRC2 subunits.

Among the 403 genes that contain overlapping MALAT1–SUZ12 peaks at their proximal promoter regions (<1 kb from the TSS), special attention was given to genes involved in cancer progression, transcriptional regulation, and epigenetic activity. While the list presented in Table 1 is not exhaustive (the full gene list is available in Supplementary Table S2), it highlights interesting targets. Remarkably, all these genes are actively transcribing, exhibiting expression levels in the range of 1 to 10 million TPM in RNA-seq studies. First, two tumor suppressors caught our attention: the thioredoxin-interacting protein TXNIP and the cell adhesion molecule CADM4. Reduced levels of these two proteins have been linked to increased tumorigenesis and, accordingly, patients in higher cancer stages show low expression of both proteins [57,58]. Therefore, we are tempted to speculate that PRC2 is contributing to tempering their transcription, favoring breast cancer development. In the case of TXNIP, decreased levels prevent the clearance of oxidative species within the cells [58], while reduced levels of CADM4 limit the interaction with the extracellular matrix in cancer cells [57]. Menin and the lysine-demethylase KDM6B were also identified among the genes recognized by MALAT1 and SUZ12, and both are well-established epigenetic regulators. Menin is an essential component of the MLL/SET1 H3K4-methyltransferase complex, which is necessary for global gene activation, while KDM6B is an H3K27-demethylase, also regulating global gene transcription. In patients, low expression of menin favors tumorigenesis [53] and reduced levels of KDM6B favor the metastasis of breast cancer cells [62]. Additionally, we found that three transcription factor genes are targets of MALAT1–SUZ12: ETS2, RARG and VEZF1. ETS2 promotes telomerase expression [59], RARG has been linked with tamoxifen resistance in breast cancer patients [60], and VEZF1 is a regulator of angiogenesis [61]. In addition, we have highlighted the SENP3 gene, involved in protein sumoylation, whose elevated levels have been linked to poor survival in breast cancer patients [55]. Finally, genes participating in mitogen-activated pathways such as EGF and FGF, which are linked to cell growth and motility [54,56], are other genes that are targets of MALAT1 and SUZ12 binding. These genes encode the 14-3-3 protein YHWAB and the MAPK15 kinase. While YHWAB has a role in cell transformation [54], downregulation of MAPK15 (a proposed biomarker of breast cancer) increases cell motility in breast cancer and decreases apoptosis [56]. The potential involvement of a MALAT1–PRC2 complex in the regulation of genes connected to intricate regulatory networks (kinases, transcription factors, epigenetic regulators) explains why many authors have shown that loss-of-function experiments of MALAT1 or PRC2 subunits are sufficient to trigger phenotypic changes in cells. Additionally, our findings are an invitation not only to study the MALAT1–PRC2 pair from the perspective of H3K27me3-mediated gene silencing (which can be a minority of case, as depicted in Figure 4b, left), but also to increase the efforts in the study of mechanisms involving early transcription termination or RNA-Polymerase II pausing, where MALAT1–PRC2 may be acting as a rheostat that moderates transcription without

suppressing it (Figure 4b, right) [4]. Biologists prone to taking a reductionist approach may find in these data and specific genes interesting models to conduct further experiments.

We are conscious that MCF7 is a human cell line representing a specific stage of a particular cancer. Nevertheless, it is also one of the most studied and characterized at the molecular level [68] and the only human cancer cell where RNA binding to the genome has been assayed using deep sequencing that is publicly available. Since PRC2-RNA-mediated regulation is tissue-specific [69], the patterns identified in this study may change in cells from different origins, including samples from healthy vs. diseased subjects. The patterns may also change among breast tumor cell lines representing different cancer stages where MALAT1 expression is modulated [70]. Recently, the Valerio Orlando team reported that MALAT1 functions as a key cofactor of PRC2 in mouse myotubes C2C12 under oxidative stress. Interestingly, chromatin isolation by RNA purification (ChIRP, an alternative assay to CHART [71]) performed with MALAT1 revealed the preferential binding of MALAT1 alone to intergenic and intronic regions ([27], see also Supplementary Figure S1a generated by the authors). Thus, we decided to perform our bioinformatic pipeline combining chromatin binding data from MALAT1 [27] and SUZ12 [72], generated by the Orlando group in this muscle cell model. Briefly, our pipeline shows 4758 peaks for MALAT1 and 3039 peaks for SUZ12, with 70 shared peaks between both molecules (Supplementary Table S4, Supplementary Figure S1b). The lower number of peaks compared to the breast cancer model may be explained by the lower expression of MALAT1 in non-cancer cells [13,14]. Interestingly, we see 37.14% of shared peaks localizing in gene promoters, plus 8.58% of signals belonging to the 5'-UTR (Supplementary Figure S1c). In sum, we confirmed that MALAT1 and SUZ12 selectively co-localize in regulatory regions in an independent cell type. We are currently conducting ChIP- and CHART-seq experiments in other human cancers where MALAT1 expression should be higher.

A caveat of our analysis is the strict approach of bedtools intersect (the tool used to identify the regions occupied by both molecules). Intersect delivers a peak where independent peaks overlap by one or more nucleotides [45]. This excludes from our analysis any gene where peaks may sit beside each other without overlapping, which is a possibility for two molecules binding to a gene. However, as our team is studying transcriptional regulation by a MALAT1-PRC2 complex, intersect's strict approach is appropriate, narrowing the set of genes to those that exhibit overlap with high confidence. We are aware that our approach may be missing long-distance genomic interactions, due to higher-order interactions within the nucleus that go beyond gene promoter regions [73]. In fact, we have not included in our analyses the approximately 30% of MALAT1-SUZ12 overlapped peaks associated with introns and intergenic regions, which are contributing to higher-order spatial contacts or enhancer activity, among others.

5. Conclusions

In summary, by analyzing key datasets of genome-wide and transcriptomic studies performed in MCF7 cells, we identified a novel set of genes where a potential MALAT1-PRC2 functional collaboration can be now further tested at the molecular level. As epigenetic regulation has become a well-established driver for cancer development and progression [6], understanding the role of the PRC2-MALAT1 pair during the transcription of key genes in cancer cells will reveal alternative paths in the understanding and treatment of this disease.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/cimb45060301/s1>, Supplementary Table S1: Quality control data of the raw ChIP-seq SRA files used in this study from MCF7 breast cancer cells; Supplementary Table S2: The complete list of genes containing overlapped peaks of MALAT1 and SUZ12 in MCF7 cells; Supplementary Table S3: List of genes and their expression levels from GSE74232/GSM1915041/SRR2749729 and GSE48216/GSM1172885/SRR925723. TPM (Transcripts Per Kilobase Million) values are shown. Sheet 1: All genes from SRR2749729; Sheet 2: All genes from SRR925723; Sheet 3: Genes from SRR2749729 with MALAT1-PRC2 overlap in gene promoter region; Sheet 4: Genes from SRR925723 with MALAT1-PRC2 overlap in gene promoter region. Supplementary Table S4: The complete list

of genes containing peaks of SUZ12, MALAT1 and intersected peaks in C2C12 mouse myotubes. Sequence Read Archives (SRA) for MALAT1 binding experiments were recovered using the following run numbers: Replica 1: SRR12587947 (<https://www.ncbi.nlm.nih.gov/sra/?term=SRR12587947> (accessed on 15 May 2023)); Input replica 1: SRR12587951 (<https://www.ncbi.nlm.nih.gov/sra/?term=SRR12587951> (accessed on 15 May 2023)); Replica 2: SRR12587946 (<https://www.ncbi.nlm.nih.gov/sra/?term=SRR12587946> (accessed on 15 May 2023)); Input replica 2: (<https://www.ncbi.nlm.nih.gov/sra/?term=SRR12587950> (accessed on 15 May 2023)). SRA files for SUZ12 experiments were: Replica 1: ERR1458695 (<https://www.ncbi.nlm.nih.gov/sra/?term=ERR1458695> (accessed on 15 May 2023)); Input replica 1: ERR1458693 (<https://www.ncbi.nlm.nih.gov/sra/?term=ERR1458693> (accessed on 15 May 2023)). Supplementary Figure S1. Genomic distribution of MALAT1-SUZ12 overlapped peaks in the C2C12 mouse myotube cell line. (a) Genomic annotations for peaks were obtained using ChIPseeker. The percentage of peaks contained in each feature (colored rectangles) is shown for SUZ12 (left) and MALAT1 (right). (b) The number of peaks shared by SUZ12 and MALAT1 is shown with green background in the Venn diagram. (c) Genomic annotations for overlapped peaks were obtained using ChIPseeker. The percentage of peaks contained in each feature (colored rectangles) is shown.

Author Contributions: Conceptualization, F.A. and R.A.; methodology, F.A., C.F., A.B. and S.F.; formal analysis, F.A., D.N., M.M., A.R. and R.A.; writing—original draft preparation, R.A.; writing—review and editing, A.R. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ANID FONDECYT Iniciacion 11200308, and UNAB, Nucleo DI-03-22/NUC (to RA).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The scripts used in this study are available in the following GitHub repository: https://github.com/CferroR/rna_prc2_chromatin (accessed on 23 May 2023). ChIP-seq datasets used in this study can be downloaded from the GEO database (GSM2828862: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2828862> (accessed on 1 July 2021); GSM2828863: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2828863> (accessed on 1 July 2021); GSM2423179: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2423179> (accessed on 1 July 2021)). CHART-seq datasets used in this study can be downloaded from the GEO database (GSM1411209: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1411209> (accessed on 1 July 2021); GSM1411210: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1411210> (accessed on 1 July 2021); GSM1411212: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1411212> (accessed on 1 July 2021)). RNA-seq datasets used in this study can be downloaded from the GEO database (GSM1915041: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1915041> (accessed on 1 November 2022); GSM1172885: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1172885> (accessed on 1 November 2022)). Additional datasets supporting the conclusions of this article are included within the article and its supplementary files.

Acknowledgments: We thank Veronica Burzio for her diligent proofreading of this article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Margueron, R.; Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **2011**, *469*, 343–349. [[CrossRef](#)] [[PubMed](#)]
2. Li, B.; Carey, M.; Workman, J.L. The role of chromatin during transcription. *Cell* **2007**, *128*, 707–719. [[CrossRef](#)] [[PubMed](#)]
3. Aguilar, R.; Bustos, F.J.; Saez, M.; Rojas, A.; Allende, M.L.; van Wijnen, A.J.; van Zundert, B.; Montecino, M. Polycomb PRC2 complex mediates epigenetic silencing of a critical osteogenic master regulator in the hippocampus. *Biochim. Biophys. Acta* **2016**, *1859*, 1043–1055. [[CrossRef](#)] [[PubMed](#)]
4. Rosenberg, M.; Blum, R.; Kesner, B.; Aeby, E.; Garant, J.M.; Szanto, A.; Lee, J.T. Motif-driven interactions between RNA and PRC2 are rheostats that regulate transcription elongation. *Nat. Struct. Mol. Biol.* **2021**, *28*, 103–117. [[CrossRef](#)]
5. Henriquez, B.; Bustos, F.J.; Aguilar, R.; Becerra, A.; Simon, F.; Montecino, M.; van Zundert, B. Ezh1 and Ezh2 differentially regulate PSD-95 gene transcription in developing hippocampal neurons. *Mol. Cell Neurosci.* **2013**, *57*, 130–143. [[CrossRef](#)]

6. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **2022**, *12*, 31–46. [[CrossRef](#)]
7. Adibfar, S.; Elveny, M.; Kashikova, H.S.; Mikhailova, M.V.; Farhangnia, P.; Vakili-Samiani, S.; Tarokhian, H.; Jadidi-Niaragh, F. The molecular mechanisms and therapeutic potential of EZH2 in breast cancer. *Life Sci.* **2021**, *286*, 120047. [[CrossRef](#)]
8. Laugesen, A.; Hojfeldt, J.W.; Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell* **2019**, *74*, 8–18. [[CrossRef](#)]
9. Rinn, J.L.; Kertesz, M.; Wang, J.K.; Squazzo, S.L.; Xu, X.; Bruggmann, S.A.; Goodnough, L.H.; Helms, J.A.; Farnham, P.J.; Segal, E.; et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **2007**, *129*, 1311–1323. [[CrossRef](#)]
10. Zhao, J.; Sun, B.K.; Erwin, J.A.; Song, J.J.; Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **2008**, *322*, 750–756. [[CrossRef](#)]
11. Aguilar, R.; Spencer, K.B.; Kesner, B.; Rizvi, N.F.; Badmalia, M.D.; Mrozowich, T.; Mortison, J.D.; Rivera, C.; Smith, G.F.; Burchard, J.; et al. Targeting Xist with compounds that disrupt RNA structure and X inactivation. *Nature* **2022**, *604*, 160–166. [[CrossRef](#)]
12. Ji, P.; Diederichs, S.; Wang, W.; Boing, S.; Metzger, R.; Schneider, P.M.; Tidow, N.; Brandt, B.; Buerger, H.; Bulk, E.; et al. MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **2003**, *22*, 8031–8041. [[CrossRef](#)]
13. Gutschner, T.; Hammerle, M.; Diederichs, S. MALAT1—A paradigm for long noncoding RNA function in cancer. *J. Mol. Med.* **2013**, *91*, 791–801. [[CrossRef](#)] [[PubMed](#)]
14. Goyal, B.; Yadav, S.R.M.; Awasthee, N.; Gupta, S.; Kunnumakkara, A.B.; Gupta, S.C. Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer. *Biochim. Biophys. Acta Rev. Cancer* **2021**, *1875*, 188502. [[CrossRef](#)] [[PubMed](#)]
15. Tao, S.; Bai, Z.; Liu, Y.; Gao, Y.; Zhou, J.; Zhang, Y.; Li, J. Exosomes Derived from Tumor Cells Initiate Breast Cancer Cell Metastasis and Chemoresistance through a MALAT1-Dependent Mechanism. *J. Oncol.* **2022**, *2022*, 5483523. [[CrossRef](#)] [[PubMed](#)]
16. Abulwerdi, F.A.; Xu, W.; Ageeli, A.A.; Yonkunas, M.J.; Arun, G.; Nam, H.; Schneekloth, J.S., Jr.; Dayie, T.K.; Spector, D.; Baird, N.; et al. Selective Small-Molecule Targeting of a Triple Helix Encoded by the Long Noncoding RNA, MALAT1. *ACS Chem. Biol.* **2019**, *14*, 223–235. [[CrossRef](#)] [[PubMed](#)]
17. Lino Cardenas, C.L.; Kessinger, C.W.; Cheng, Y.; MacDonald, C.; MacGillivray, T.; Ghoshhajra, B.; Huleihel, L.; Nuri, S.; Yeri, A.S.; Jaffer, F.A.; et al. An HDAC9-MALAT1-BRG1 complex mediates smooth muscle dysfunction in thoracic aortic aneurysm. *Nat. Commun.* **2018**, *9*, 1009. [[CrossRef](#)] [[PubMed](#)]
18. Warda, A.S.; Kretschmer, J.; Hackert, P.; Lenz, C.; Urlaub, H.; Hobartner, C.; Sloan, K.E.; Bohnsack, M.T. Human METTL16 is a N(6)-methyladenosine (m(6)A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO Rep.* **2017**, *18*, 2004–2014. [[CrossRef](#)] [[PubMed](#)]
19. Achour, C.; Aguilo, F. Long non-coding RNA and Polycomb: An intricate partnership in cancer biology. *Front. Biosci.* **2018**, *23*, 2106–2132. [[CrossRef](#)]
20. Ye, M.; Xie, L.; Zhang, J.; Liu, B.; Liu, X.; He, J.; Ma, D.; Dong, K. Determination of long non-coding RNAs associated with EZH2 in neuroblastoma by RIP-seq, RNA-seq and ChIP-seq. *Oncol. Lett.* **2020**, *20*, 1. [[CrossRef](#)]
21. Huang, J.; Fang, J.; Chen, Q.; Chen, J.; Shen, J. Epigenetic silencing of E-cadherin gene induced by lncRNA MALAT-1 in acute myeloid leukaemia. *J. Clin. Lab. Anal.* **2022**, *36*, e24556. [[CrossRef](#)] [[PubMed](#)]
22. Qi, Y.; Ooi, H.S.; Wu, J.; Chen, J.; Zhang, X.; Tan, S.; Yu, Q.; Li, Y.Y.; Kang, Y.; Li, H.; et al. MALAT1 long ncRNA promotes gastric cancer metastasis by suppressing PCDH10. *Oncotarget* **2016**, *7*, 12693–12703. [[CrossRef](#)] [[PubMed](#)]
23. Li, P.; Zhang, X.; Wang, H.; Wang, L.; Liu, T.; Du, L.; Yang, Y.; Wang, C. MALAT1 Is Associated with Poor Response to Oxaliplatin-Based Chemotherapy in Colorectal Cancer Patients and Promotes Chemoresistance through EZH2. *Mol. Cancer Ther* **2017**, *16*, 739–751. [[CrossRef](#)] [[PubMed](#)]
24. Zovoilis, A.; Cifuentes-Rojas, C.; Chu, H.P.; Hernandez, A.J.; Lee, J.T. Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell* **2016**, *167*, 1788–1802.e1713. [[CrossRef](#)]
25. Fan, Y.; Shen, B.; Tan, M.; Mu, X.; Qin, Y.; Zhang, F.; Liu, Y. TGF-beta-induced upregulation of malat1 promotes bladder cancer metastasis by associating with suz12. *Clin. Cancer Res.* **2014**, *20*, 1531–1541. [[CrossRef](#)]
26. Orlando, V.; Paro, R. Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. *Cell* **1993**, *75*, 1187–1198. [[CrossRef](#)]
27. El Said, N.H.; Della Valle, F.; Liu, P.; Paytuyi-Gallart, A.; Adroub, S.; Gimenez, J.; Orlando, V. Malat-1-PRC2-EZH1 interaction supports adaptive oxidative stress dependent epigenome remodeling in skeletal myotubes. *Cell Death Dis.* **2021**, *12*, 850. [[CrossRef](#)]
28. West, J.A.; Davis, C.P.; Sunwoo, H.; Simon, M.D.; Sadreyev, R.I.; Wang, P.I.; Tolstorukov, M.Y.; Kingston, R.E. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* **2014**, *55*, 791–802. [[CrossRef](#)]
29. Engreitz, J.M.; Sirokman, K.; McDonel, P.; Shishkin, A.A.; Surka, C.; Russell, P.; Grossman, S.R.; Chow, A.Y.; Guttman, M.; Lander, E.S. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **2014**, *159*, 188–199. [[CrossRef](#)]
30. Lu, J.Y.; Shao, W.; Chang, L.; Yin, Y.; Li, T.; Zhang, H.; Hong, Y.; Percharde, M.; Guo, L.; Wu, Z.; et al. Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep.* **2020**, *30*, 3296–3311.e5. [[CrossRef](#)]

31. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [[CrossRef](#)] [[PubMed](#)]
32. Soule, H.D.; Vazquez, J.; Long, A.; Albert, S.; Brennan, M. A human cell line from a pleural effusion derived from a breast carcinoma. *J. Natl. Cancer Inst.* **1973**, *51*, 1409–1416. [[CrossRef](#)]
33. GSM1411209: MCF7 MALAT1 CO1 CHART-seq. Available online: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1411209> (accessed on 1 July 2021).
34. GSM1411210: MCF7 MALAT1 CO2 CHART-seq. Available online: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1411210> (accessed on 1 July 2021).
35. GSM1411212: MCF7 Input for MALAT1 CHART-seq. Available online: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1411212> (accessed on 1 July 2021).
36. GSM2828862: ChIP-seq from MCF-7. Available online: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2828862> (accessed on 1 July 2021).
37. GSM2828863: ChIP-seq from MCF-7. Available online: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2828863> (accessed on 1 July 2021).
38. GSM2423179: ChIP-seq from MCF-7. Available online: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2423179> (accessed on 1 July 2021).
39. ENCODE Transcription Factor and Histone ChIP-Seq Processing Pipeline. Available online: <https://github.com/ENCODE-DCC/chip-seq-pipeline2> (accessed on 1 July 2021).
40. ENCODE Transcription Factor and Histone ChIP-Seq Processing Pipeline: Input JSON. Available online: <https://github.com/ENCODE-DCC/chip-seq-pipeline2/blob/master/docs/input.md> (accessed on 1 July 2021).
41. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
42. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
43. Picard. Available online: <http://broadinstitute.github.io/picard/> (accessed on 1 July 2021).
44. Feng, J.; Liu, T.; Qin, B.; Zhang, Y.; Liu, X.S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **2012**, *7*, 1728–1740. [[CrossRef](#)]
45. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinform.* **2014**, *47*, 11.12.1–11.12.34. [[CrossRef](#)]
46. Yu, G.; Wang, L.G.; He, Q.Y. ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **2015**, *31*, 2382–2383. [[CrossRef](#)]
47. Mi, H.; Muruganujan, A.; Huang, X.; Ebert, D.; Mills, C.; Guo, X.; Thomas, P.D. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* **2019**, *14*, 703–721. [[CrossRef](#)]
48. Dieudonné, F.X.; O'Connor, P.B.; Gubler-Jaquier, P.; Yasrebi, H.; Conne, B.; Nikolaev, S.; Antonarakis, S.; Baranov, P.V.; Curran, J. GSM1915041: MCF7 Total RNA-seq, from GSE74232: The Effect of Heterogeneous Transcription Start Sites (TSS) on the Translatome: Implications for the Mammalian Cellular Phenotype. Available online: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1915041> (accessed on 1 November 2022).
49. Dieudonne, F.X.; O'Connor, P.B.; Gubler-Jaquier, P.; Yasrebi, H.; Conne, B.; Nikolaev, S.; Antonarakis, S.; Baranov, P.V.; Curran, J. The effect of heterogeneous Transcription Start Sites (TSS) on the translatome: Implications for the mammalian cellular phenotype. *BMC Genom.* **2015**, *16*, 986. [[CrossRef](#)]
50. GSM1172885: MCF7, RNA-Seq, from GSE48216: Modeling Precision Treatment of Breast Cancer. Available online: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1172885> (accessed on 1 November 2022).
51. Costello, J.C.; Heiser, L.M.; Georgii, E.; Gonen, M.; Menden, M.P.; Wang, N.J.; Bansal, M.; Ammad-ud-din, M.; Hintsanen, P.; Khan, S.A.; et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **2014**, *32*, 1202–1212. [[CrossRef](#)]
52. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [[CrossRef](#)] [[PubMed](#)]
53. Teinturier, R.; Abou Ziki, R.; Kassem, L.; Luo, Y.; Malbeteau, L.; Gherardi, S.; Corbo, L.; Bertolino, P.; Bachelot, T.; Treilleux, I.; et al. Reduced menin expression leads to decreased ERalpha expression and is correlated with the occurrence of human luminal B-like and ER-negative breast cancer subtypes. *Breast Cancer Res. Treat.* **2021**, *190*, 389–401. [[CrossRef](#)] [[PubMed](#)]
54. Takihara, Y.; Matsuda, Y.; Hara, J. Role of the beta isoform of 14-3-3 proteins in cellular proliferation and oncogenic transformation. *Carcinogenesis* **2000**, *21*, 2073–2077. [[CrossRef](#)] [[PubMed](#)]
55. Graves, J.D.; Lee, Y.J.; Liu, K.; Li, G.; Lin, F.T.; Lin, W.C. E2F1 sumoylation as a protective cellular mechanism in oxidative stress response. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 14958–14969. [[CrossRef](#)] [[PubMed](#)]
56. Chia, J.; Tham, K.M.; Gill, D.J.; Bard-Chapeau, E.A.; Bard, F.A. ERK8 is a negative regulator of O-GalNAc glycosylation and cell migration. *eLife* **2014**, *3*, e01828. [[CrossRef](#)]
57. Saito, M.; Goto, A.; Abe, N.; Saito, K.; Maeda, D.; Ohtake, T.; Murakami, Y.; Takenoshita, S. Decreased expression of CADM1 and CADM4 are associated with advanced stage breast cancer. *Oncol. Lett.* **2018**, *15*, 2401–2406. [[CrossRef](#)]

58. Chen, Y.; Ning, J.; Cao, W.; Wang, S.; Du, T.; Jiang, J.; Feng, X.; Zhang, B. Research Progress of TXNIP as a Tumor Suppressor Gene Participating in the Metabolic Reprogramming and Oxidative Stress of Cancer Cells in Various Cancers. *Front. Oncol.* **2020**, *10*, 568574. [[CrossRef](#)]
59. Xu, D.; Dwyer, J.; Li, H.; Duan, W.; Liu, J.P. Ets2 maintains hTERT gene expression and breast cancer cell proliferation by interacting with c-Myc. *J. Biol. Chem.* **2008**, *283*, 23567–23580. [[CrossRef](#)]
60. Mendes-Pereira, A.M.; Sims, D.; Dexter, T.; Fenwick, K.; Assiotis, I.; Kozarewa, I.; Mitsopoulos, C.; Hakas, J.; Zvelebil, M.; Lord, C.J.; et al. Genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2730–2735. [[CrossRef](#)]
61. Yin, R.; Guo, L.; Gu, J.; Li, C.; Zhang, W. Over expressing miR-19b-1 suppress breast cancer growth by inhibiting tumor microenvironment induced angiogenesis. *Int. J. Biochem. Cell Biol.* **2018**, *97*, 43–51. [[CrossRef](#)]
62. Xun, J.; Gao, R.; Wang, B.; Li, Y.; Ma, Y.; Guan, J.; Zhang, Q. Histone demethylase KDM6B inhibits breast cancer metastasis by regulating Wnt/beta-catenin signaling. *FEBS Open Bio.* **2021**, *11*, 2273–2281. [[CrossRef](#)] [[PubMed](#)]
63. Davidovich, C.; Zheng, L.; Goodrich, K.J.; Cech, T.R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1250–1257. [[CrossRef](#)] [[PubMed](#)]
64. Cifuentes-Rojas, C.; Hernandez, A.J.; Sarma, K.; Lee, J.T. Regulatory interactions between RNA and polycomb repressive complex 2. *Mol. Cell* **2014**, *55*, 171–185. [[CrossRef](#)] [[PubMed](#)]
65. Long, Y.; Bolanos, B.; Gong, L.; Liu, W.; Goodrich, K.J.; Yang, X.; Chen, S.; Gooding, A.R.; Maegley, K.A.; Gajiwala, K.S.; et al. Conserved RNA-binding specificity of polycomb repressive complex 2 is achieved by dispersed amino acid patches in EZH2. *eLife* **2017**, *6*, e31558. [[CrossRef](#)]
66. Wang, X.; Goodrich, K.J.; Gooding, A.R.; Naeem, H.; Archer, S.; Paucek, R.D.; Youmans, D.T.; Cech, T.R.; Davidovich, C. Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Mol. Cell* **2017**, *65*, 1056–1067.E5. [[CrossRef](#)]
67. Mousavi, K.; Zare, H.; Wang, A.H.; Sartorelli, V. Polycomb protein Ezh1 promotes RNA polymerase II elongation. *Mol. Cell* **2012**, *45*, 255–262. [[CrossRef](#)]
68. Comsa, S.; Cimpean, A.M.; Raica, M. The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research. *Anticancer Res.* **2015**, *35*, 3147–3154.
69. Wang, Y.; Xie, Y.; Li, L.; He, Y.; Zheng, D.; Yu, P.; Yu, L.; Tang, L.; Wang, Y.; Wang, Z. EZH2 RIP-seq Identifies Tissue-specific Long Non-coding RNAs. *Curr. Gene Ther.* **2018**, *18*, 275–285. [[CrossRef](#)]
70. Wang, Z.; Katsaros, D.; Biglia, N.; Shen, Y.; Fu, Y.; Loo, L.W.M.; Jia, W.; Obata, Y.; Yu, H. High expression of long non-coding RNA MALAT1 in breast cancer is associated with poor relapse-free survival. *Breast Cancer Res. Treat.* **2018**, *171*, 261–271. [[CrossRef](#)]
71. Chu, C.; Qu, K.; Zhong, F.L.; Artandi, S.E.; Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **2011**, *44*, 667–678. [[CrossRef](#)]
72. Bodega, B.; Marasca, F.; Ranzani, V.; Cherubini, A.; Della Valle, F.; Neguembor, M.V.; Wassef, M.; Zippo, A.; Lanzuolo, C.; Pagani, M.; et al. A cytosolic Ezh1 isoform modulates a PRC2-Ezh1 epigenetic adaptive response in postmitotic cells. *Nat. Struct. Mol. Biol.* **2017**, *24*, 444–452. [[CrossRef](#)] [[PubMed](#)]
73. Quinodoz, S.A.; Bhat, P.; Chovanec, P.; Jachowicz, J.W.; Ollikainen, N.; Detmar, E.; Sohalim, E.; Guttman, M. SPRITE: A genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. *Nat. Protoc.* **2022**, *17*, 36–75. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.