Data in brief 29 (2020) 105139

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Gene expression dataset of prostate cells upon *MIR205HG*/*LEADR* modulation

Stefano Percio [a], Federica Rotundo [a], Paolo Gandellini [a, b, *]

[a] *Department of Applied Research and Technological Development, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, 20133, Italy*
[b] *Department of Biosciences, University of Milan, Milan, 20133, Italy*

ARTICLE INFO

Article history:
Received 22 October 2019
Received in revised form 7 January 2020
Accepted 10 January 2020
Available online 16 January 2020

Keywords:
Long noncoding RNA
Prostate
Differentiation
miR-205
MIR205HG
Alu

ABSTRACT

Although the role of *miR*-205 has been widely elucidated, the function of its host gene (*MIR205HG*) is yet to be clarified. We have recently investigated whether this gene is a simple endorsement for miRNA production or it may act independently, demonstrating its action as nuclear long noncoding RNA able to control basal-luminal differentiation in the human prostate context, thus deserving the reannotation as *LEADR*, Long Epithelial *Alu*-interacting Differentiation-related RNA. Here, we describe the loss and gain of function approaches experimentally used to modulate *LEADR* expression, and the bioinformatic procedures employed to analyze microarray data in our published article "LEADeR role of miR-205 host gene as long noncoding RNA in prostate basal cell differentiation" [1].

The high reproducibility of replicates, the strong concordance with a validation technique, and the coherent behavior observed for differentially expression features, both in terms of single genes and deregulated pathways, not only support the quality of the data, but also endorse their potential reuse. Very relevant are the diverse silencing and overexpression strategies employed (all of which analyzed in multiple biologically independent replicates), which should allow other scientists to analyze our dataset for the specific purpose of their research, may it be the study of *MIR205HG* function as miRNA host gene, the investigation of its miRNA-independent biological role or again the dissection of *Alu*

* Corresponding author. Department of Biosciences, University of Milan, Via Celoria 26, 20133, Milano, Italy.
E-mail address: paolo.gandellini@unimi.it (P. Gandellini).

https://doi.org/10.1016/j.dib.2020.105139

sequence involvement in the mechanism of action of long non-coding RNAs, which is a hot topic in the field.

Specifications Table

Subject	Biochemistry, Genetics and Molecular Biology (General)
Specific subject area	Transcriptomics of *MIR205HG*-modulated cells
Type of data	Table Image Chart Graph
How data were acquired	Illumina HumanHT-12 v4 chip microscope 7900HT Fast Real-Time PCR System SDS relative quantity (RQ) Manager 1.2.1 software ChemiDoc MP Imaging System R 3.5.3 Graphpad Prism 5
Data format	Raw Analyzed Filtered
Parameters for data collection	Prostate cells modulated for *MIR205HG* (*LEADR*)
Description of data collection	RWPE-1 cells were silenced for *MIR205HG* expression using different oligomers, while *MIR205HG* expression was reconstituted in DU145 cells with plasmid vectors. Total RNA was hybridized to the Illumina HumanHT-12 chip and scanned according to the standard Illumina protocol.
Data source location	Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy
Data accessibility	Repository name: Gene Expression Omnibus (GEO) Data identification number: GSE104003 Direct URL to data: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104003
Related research article	Valentina Profumo, Barbara Forte, Stefano Percio, Federica Rotundo, Valentina Doldi, Elena Ferrari, Nicola Fenderico, Matteo Dugo, Dario Romagnoli, Matteo Benelli, Riccardo Valdagni, Diletta Dolfini, Nadia Zaffaroni, and Paolo Gandellini. LEADeR role of miR-205 host gene as long noncoding RNA in prostate basal cell differentiation. Nature Communications DOI: 10.1038/s41467-018-08153-2

Value of the Data

- To our knowledge this dataset is the first-in-class transcriptomic profile obtained on human prostate cells modulated for *MIR205HG* and therefore it is a unique resource.
- Our dataset may be useful to all researchers interested in investigating *MIR205HG* function, which is still a largely unexplored field.
- The different silencing and overexpression strategies used could stimulate other scientists to consider our dataset for their specific research purpose concerning the study of *MIR205HG* itself, of the hosted *miR-205* and their interaction or of the *Alu* contribution in the *MIR205HG* function.

1. Data description

1.1. Experimental plan

The precursor sequence of human *miR-205* is located on the positive strand of the chr1q32 region and spans the last intron/exon junction of a gene originally named as *NPCA-5* (alias *LOC642587*). Given that no protein product has been experimentally observed, this gene was re-named as *miR-205* Host Gene (*MIR205HG*) in the assumption that it merely hosts *miR-205*. Nonetheless, latest researches have started to shed light on a possible regulatory role of *MIR205HG* as competing endogenous RNA, sponging *miR-590* and *miR-122* in head and neck and cervical tumors respectively [2,3].

In this regard, we have investigated, for the first time, the function of *MIR205HG* in human prostate, thus discovering that it is mostly expressed in the basal layer of epithelium, and progressively lost during luminal differentiation and tumorigenesis [1]. *MIR205HG* processed transcripts behave as nuclear intergenic long noncoding RNAs (lincRNAs), able *per se* to regulate basal-luminal differentiation, through the repression of the interferon pathway [1]. Mechanistically, we hypothesized that, by means of an *Alu*-like element present in the second exon, *MIR205HG* directly binds to the promoter regions of target genes, which are characterized themselves by an *Alu* element in proximity of an interferon-regulatory factor (IRF) binding site. In addition, we experimentally observed that this lincRNA interacts with IRF1 protein, thus modulating its occupancy on target genes. In line with this evidence, we proposed *MIR205HG* as a buffer of IRF1 transcription factor activity [1].

Since *MIR205HG* operates autonomously from *miR-205* and all its isoforms could be considered functionally active due to the presence of the *Alu* element, we decided to re-annotate it as *LEADR* [1].

In Profumo, Forte, Percio, Rotundo et al. [1], our main goal was to discover *LEADR* biological role in human prostate cells, and here we describe, in detail, the employed experimental procedures. Basically, we adopted gain and loss of function approaches to modify *LEADR* levels in different prostate cell lines and analyzed the consequent gene expression changes. Aside the techniques used to modulate *LEADR*-specific transcripts only, we also considered both the modification of its entire locus (including *miR-205*) and the overexpression of transcripts deleted for functional elements (i.e. *Alu*) (Fig. 1).

Transcriptome changes were measured through hybridization on a microarray platform comparing the different perturbations of *LEADR* expression operated in normal and tumor prostate cell lines (RWPE-1 and DU145, respectively).

Gene expression profile analysis led us to unearth *LEADR* function in prostate cells and speculate about the mechanism through which it regulates the transcription of its target genes. Asides this, our experiments could have a value for potential reuse in taking researchers to the discovery of new biological functions of *LEADR*. Therefore, the dataset described here paves the way for a better characterization of *LEADR* gene and a clarification of its transcriptional/functional interplay with the hosted *miR-205*.

1.2. Preprocessing pipeline and biological replicate management

Transcriptomic quantification from prostate cells modulated for *LEADR* expression was generated in our laboratory as three independent microarray experiments. For each dataset, the same preprocessing pipeline was applied (Fig. 2a), consisting of robust spline normalization (rsn) and filtering of low quality data. As illustrated in the box plots (Fig. 2b), this method confers a more stable inference even if the number of arrays is small, and it is able to remove any fluctuation due to systematic bias. Around 25% of probes in each microarray experiment showed sufficient quality to pass the threshold (Fig. 2a).

Several biological replicates for each modulation were introduced in the experimental design to enhance the statistical power and improve robustness of the biological variability estimation in the subsequent bioinformatic analyses. Specifically, all experiments were conducted as independent biological triplicates, except for the overexpression of *LEADR* RefSeq transcript and gene, which were performed as quadruplicates.

Principal component analysis (PCA) was performed to evaluate the quality of biological replicates in terms of explained variance. Upon plotting all samples in the space of principal components, a clear

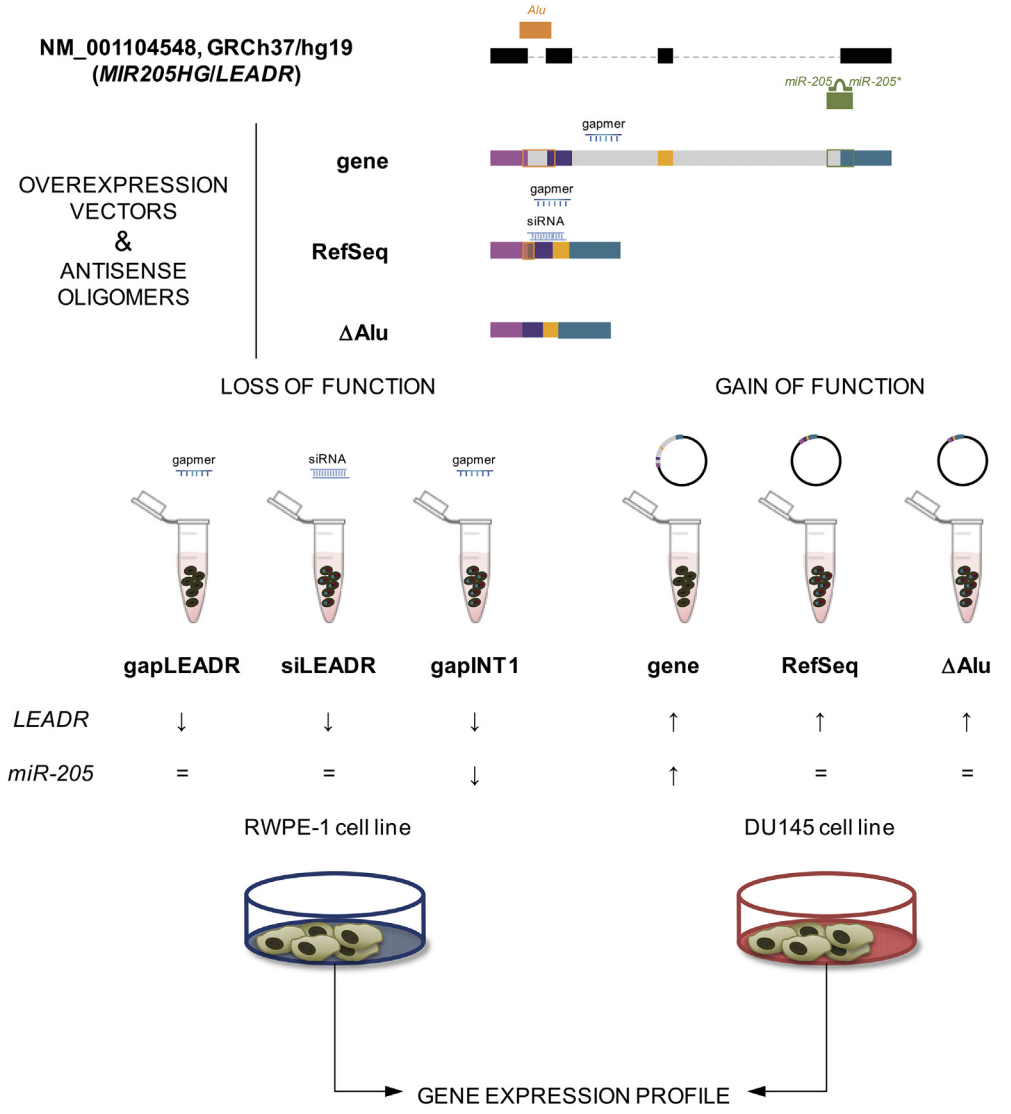

Fig. 1. LEADR modulation pipeline. The scheme describes, from top to bottom, i) NCBI RefSeq of *MIR205HG* with indication of the *Alu* element and pre-*miR-205*, ii) the constructs used for overexpression of whole *MIR205HG* gene, RefSeq, and *Alu*-deleted RefSeq (DAlu) transcripts with indication of target sequences of gapmers and siRNA used for silencing, iii) the loss and gain of function approaches with a summary of the effects induced by the different *MIR205HG* locus modulations, and iv) gene expression profile of cells perturbed for *MIR205HG* expression achieved through hybridization on a microarray platform.

segregation between replicates and other samples emerged in each dataset (Fig. 2c). The sample distribution showed a good congruency among the biological replicates confirming data consistency and reproducibility.

1.3. Evaluation of transfection efficacy

Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) was used to validate *LEADR* silencing in RWPE-1 cells upon transfection with siRNA or gapmers (Fig. 3a).

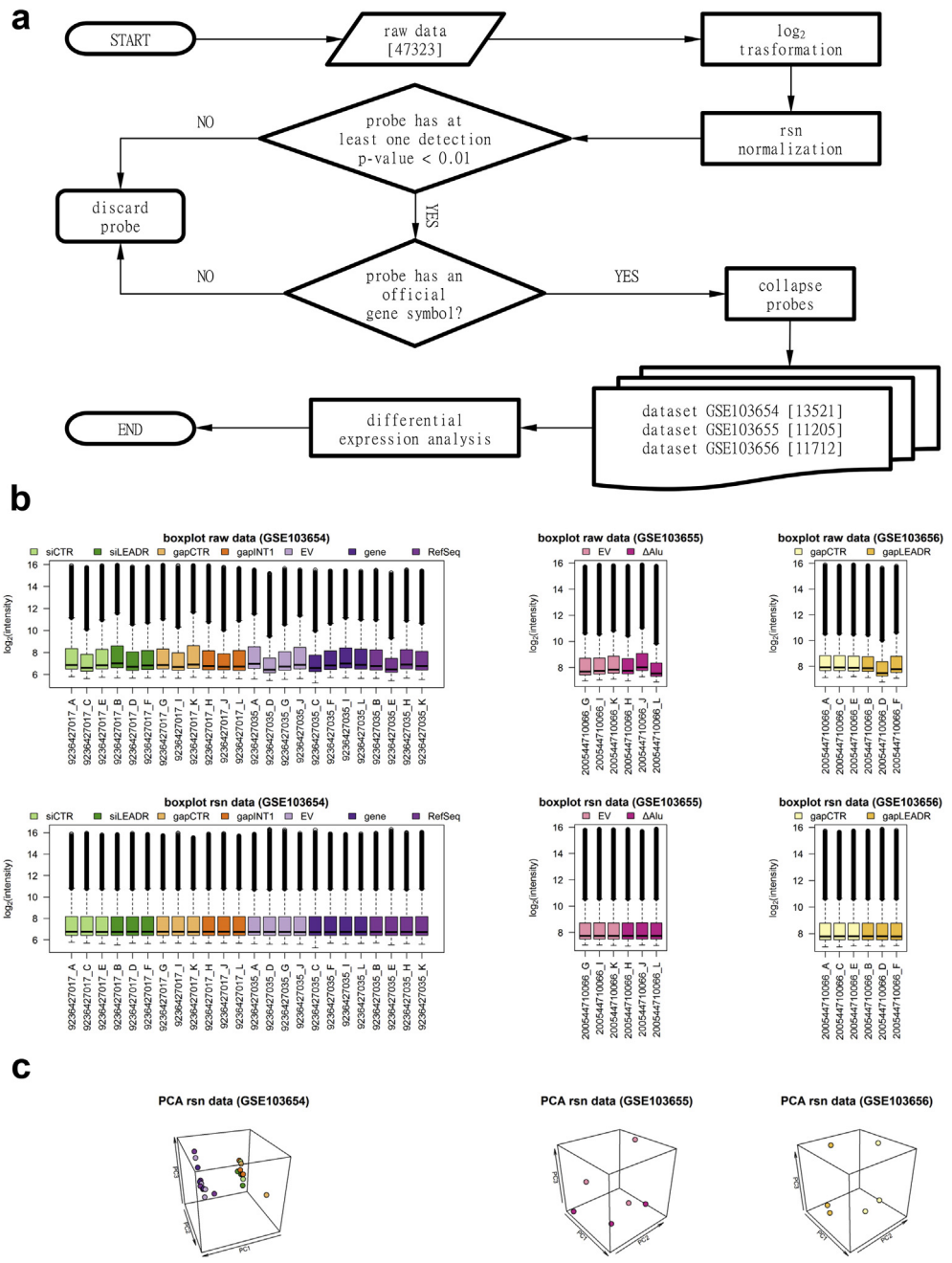

Fig. 2. Preprocessing pipeline and sample consistency. Flow chart describes the pipeline applied to analyze gene expression profiles. Raw data (total number of probes is indicated in brackets) obtained by the hybridization on Illumina HumanHT-12 v4 chips were log_2-transformed and normalized according to the rsn method. Probes with both at least one significant detection p-value (threshold of 0.01) and an associated official gene symbol were considered. Among filtered probes, for those with the same gene symbol, the probe with the highest variance was selected. For each of the three datasets, the final normalized matrix of expression profile was obtained (the number of retained probes is indicated in brackets) and used for differential expression analysis (a). Box plots for each GEO dataset with raw expression (*top*) and rsn normalized values (*bottom*), respectively (b). Scatter plot of sample distribution in the space of principal components (PCs) for each GEO dataset. Replicates affected by the same modulation are indicated with the same color (c).

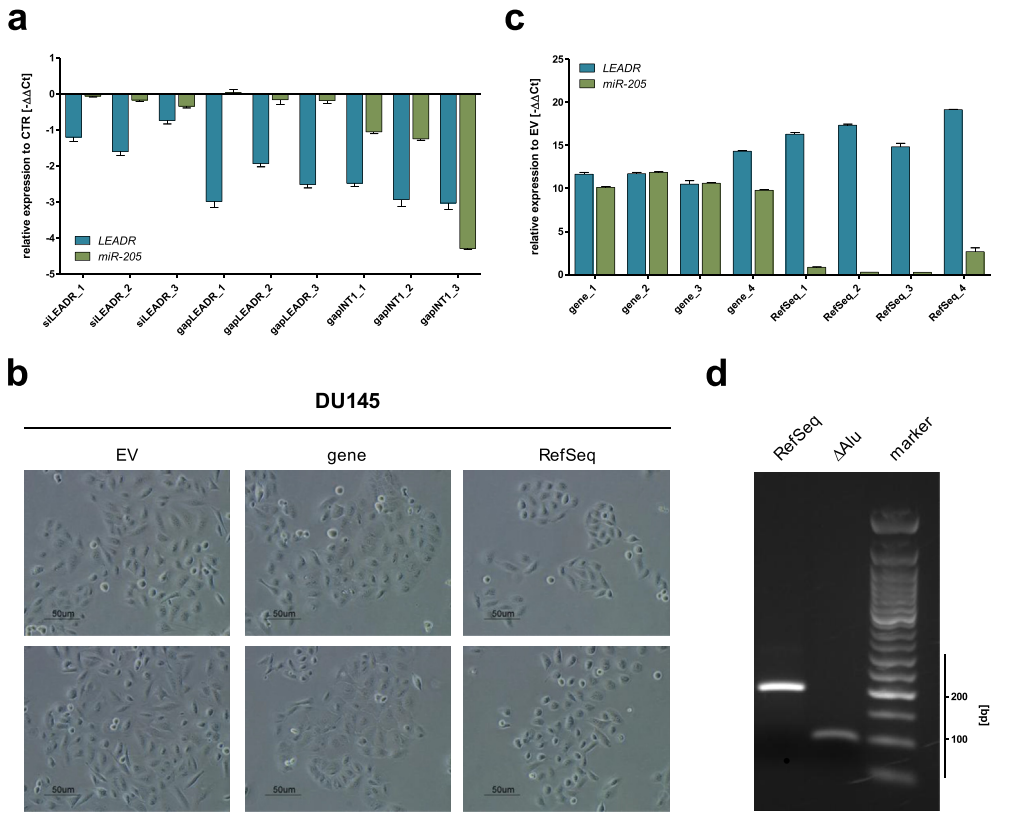

Fig. 3. Evaluation of transfection efficacy. Bar plots of qRT-PCR show the ability of both siLEADR and gapLEADR oligomers to abrogate *LEADR* expression and of gapINT1 oligomer to inhibit both *LEADR* and *miR-205*. Mean + s.d. (n = 3 qRT-PCR measurements) plotted (a). Full-size bright-field images show morphological changes occurring in DU145 cells transfected with empty vector (EV), RefSeq, and gene vectors respectively. Scale bar, 50 μm (b). Bar plots of qRT-PCR show the efficacy in restoring the expression of *LEADR* only or of *LEADR* and *miR-205*, upon transfection with RefSeq or gene vector, respectively. Mean + s.d. (n = 3 qRT-PCR measurements) plotted (c). End point RT-PCR shows the efficient transfection and transcription of the vector carrying RefSeq transcript depleted for a portion of the *Alu* element (ΔAlu). The product has an evidently lower molecular weight as compared to the wild type (RefSeq) (d).

Graphs confirmed a general abrogation of *LEADR* expression. All oligomers showed similar performance in terms of silencing efficacy across biological replicates, though the reduction of *LEADR* was overall more pronounced with gapmers as compared to siRNA. Specifically, the gapmer designed to target an intron within *LEADR* primary transcript (gapINT1) revealed to be able to reduce both *LEADR* and *miR-205* expression levels (Fig. 3a).

In addition, the efficacy of transfection with siRNA and gapmers in RWPE-1 cells was supported by changes in morphology, as assessed through bright-field imaging in Profumo, Forte, Percio, Rotundo et al. [1].

Morphological changes were appreciable also in DU145 cells transfected with either the entire *LEADR* gene or the RefSeq transcript, as compared to cells transfected with the empty vector (EV) (Fig. 3b). Consistent with this change, *LEADR* resulted to be overexpressed in both cases (Fig. 3c). Notably, transfection of the entire locus was reflected by a higher level of both *LEADR* and *miR-205*, while that of the RefSeq sequence showed an increase of *LEADR* only (Fig. 3c). Changes were consistent across biological replicates.

In overexpression experiments conducted with the RefSeq transcript deleted for the portion of the *Alu* element, *LEADR* expression levels increased more than ten times with respect to the control. This result corroborates the efficacy of the transfection albeit *LEADR* transcript was deprived of the *Alu* element, which is extremely important for its function as described in Profumo, Forte, Percio, Rotundo et al. [1]. Moreover, end point RT-PCR of RNA extracted from DU145 cells transfected with either *LEADR* RefSeq or correspondent *Alu*-deleted transcript showed amplicons of different molecular weight, confirming transcription of a shorter *LEADR* sequence in ΔAlu-cells (Fig. 3d).

1.4. Bioinformatic analyses

Transcriptomic changes arising from *LEADR* modulation were analyzed to infer *LEADR* function in prostate cells. We expected to observe an opposite trend according to the alterations induced in the two different cell lines (silencing *vs*. overexpression). Prevalently, *LEADR* revealed to have a repressive attitude since most of the genes were up-regulated upon its silencing (Fig. 4a). In line with this evidence, the up-regulated genes showed more overlap in all the three silencing experiments than the down-regulated ones, as reported in the Venn diagrams (Fig. 4a).

As a complementary approach, we ectopically expressed the RefSeq or gene sequences in DU145 cells and observed a high overlap in terms of both up- and down-regulated genes (Fig. 4b). Moreover, we wondered whether the genes overlapping across silencing experiments showed an opposite trend upon *LEADR* overexpression. This resulted to be the case, as major overlap was observed between genes up-regulated upon silencing and those down-regulated upon overexpression, as clearly illustrated in Profumo, Forte, Percio, Rotundo et al. [1]. Such evidence prompted us to focus on these genes as *bona fide LEADR* direct targets [1].

In order to corroborate our hypothesis, we applied a holist approach evaluating the concordance in terms of pathway modulation on genes commonly up-regulated upon silencing (65 genes according to the Venn diagram in Fig. 4a, *right*) and genes commonly down-regulated upon overexpression (97 genes according to the Venn diagram in Fig. 4b, *left*). On these selections, we performed an over-representation analysis investigating a possible enrichment of general molecular mechanisms belonging to the Hallmark collection of the MSigDB database [4]. Interestingly, we observed a high overlap among gene sets: the two lists of hallmarks shared seven pathways, though different for rank and size, with interferon pathways at the top of both rankings (Fig. 4c).

Overall, bioinformatic analyses revealed a strong concordance among *LEADR* modulations and a good coherence (albeit the sign of expression) between *LEADR* silencing and overexpression, thus highlighting the efficacy and the correctness of the experimental procedure.

1.5. Technical validation of microarray measurement

To confirm the validity of high-throughput microarray measures, we randomly selected 28 genes commonly modulated across *LEADR* silencing experiments in RWPE-1 cells, and we evaluated their expression levels using an independent technique, namely TaqMan-based qRT-PCR, in RWPE-1 cells transfected with gapLEADR and gapCTR. Plotting fold change (FC) values of both qRT-PCR and microarray analyses, a strong linear relationship clearly emerged (Fig. 4d), suggesting a considerable concordance in terms of trend and magnitude. Overall, this observation corroborates the quality and the informativeness of our high-throughput data, thus stimulating their potential reuse by other authors for the discovery of new *LEADR* functions.

1.6. Data records

Raw and normalized gene expression data in the study have been deposited at Gene Expression Omnibus as a superseries with the accession number GSE104003. This superseries is composed of three different datasets (GEO103654, GEO103655, and GEO103656), each one including LEADR modulation experiments and relative controls, as previously described (Table 1).

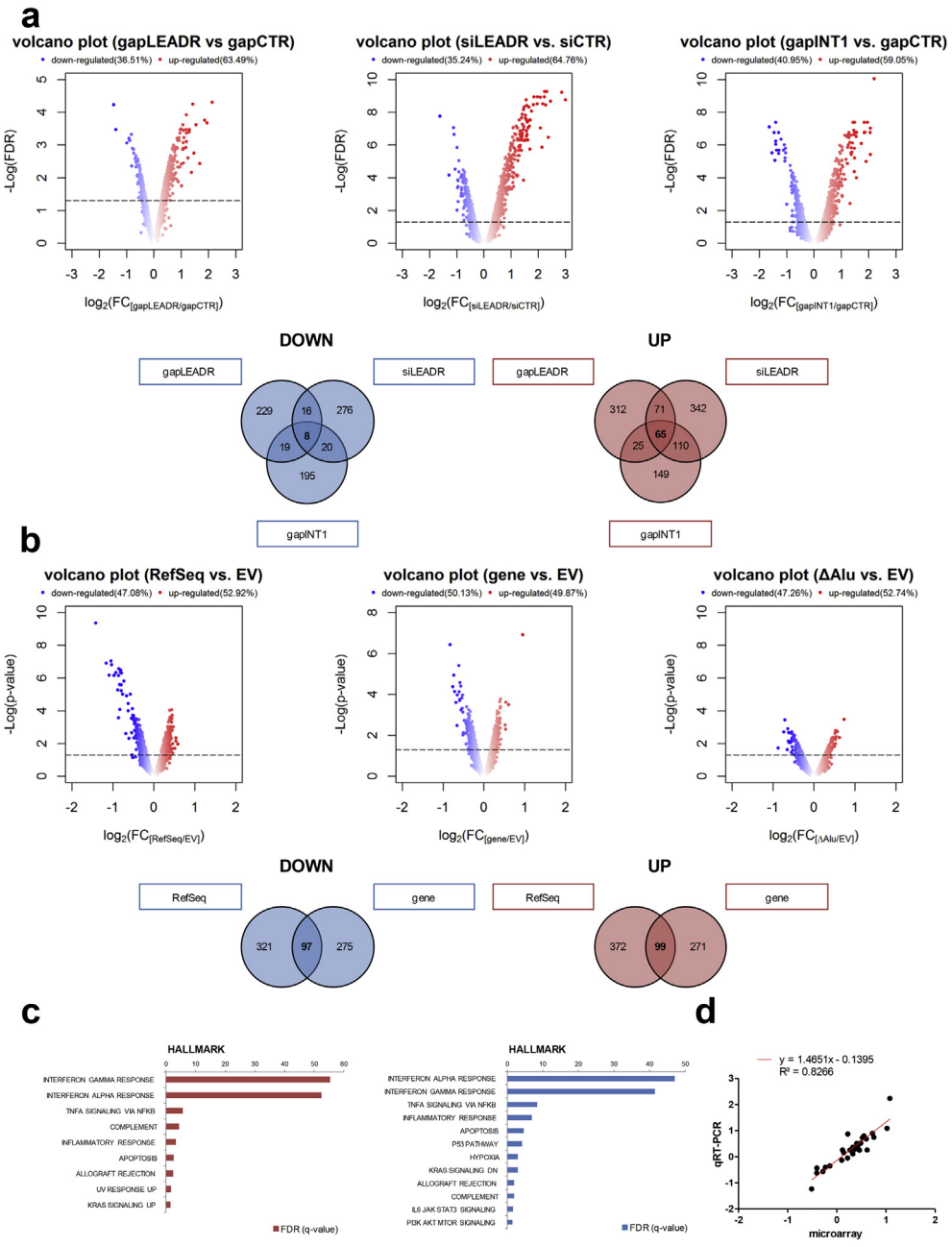

Fig. 4. Differential expression upon LEADR modulations. Differential expression analysis between gapLEADR, siLEADR, and gapINT1 and their respective controls (CTRs) was visualized in the volcano plots; blue dots represent down-regulated genes while red dots the up-regulated ones (*top*). Venn diagrams show the overlap among all three silencing methods in terms of down- and up-regulated genes (*bottom*) (a). Differential expression analysis between gene, RefSeq, and DAlu, and their respective control (EV) was visualized in the volcano plots; blue dots represent down-regulated genes while red dots the up-regulated ones (*top*). Venn diagrams show the overlap between gene and RefSeq overexpression methods in terms of down- and up-regulated genes (*bottom*) (b). Bar plots of overrepresentation analysis performed on genes commonly down-regulated upon overexpression (97 according to the left Venn diagram of Figure 4b), and those commonly up-regulated upon silencing (65 genes according to the right Venn diagram of

2. Experimental design, materials, and methods

2.1. LEADR modulation

We silenced *LEADR* in RWPE-1 cell line, a human normal prostate model, where it is endogenously expressed at high levels. To solely abrogate *LEADR* expression, a single-stranded LNA/DNA/LNA antisense gapmer (gapLEADR) or a conventional small interfering RNA (siLEADR), both complementary to an exon/exon junction in RefSeq sequence, were employed (Fig. 1). Gapmer nuclear translocation and RNAse H-mediated cleavage prompted us to adopt this approach to improve the efficacy of silencing in this compartment, which we and others demonstrated to be the main *LEADR* subcellular localization [1,2]. Another gapmer (gapINT1) was designed to target an intron within *LEADR* primary transcript (Fig. 1), and we showed that it was able to reduce the expression of both *LEADR* and *miR-205* [1]. This strategy was employed to study a possible transcriptional and functional relationship between *LEADR* and its hosted miRNA. In addition, a control siRNA (siCTR) and gapmer (gapCTR) with no homology to any known human mRNA were used for comparative analyses.

We reconstituted *LEADR* expression by stably transfecting plasmid vectors containing either the full length (hereafter gene) or the RefSeq transcript sequences in DU145 cell line (Fig. 1), a human prostate cancer model, where both *LEADR* and *miR-205* are normally not expressed.

In addition, in the same cell line, a vector containing the RefSeq sequence deleted for a portion of the *Alu* element (ΔAlu) was introduced to study the impact of this depletion on *LEADR* regulatory functions (Fig. 1). An empty vector was included in all experiments as control (EV).

2.2. RNA isolation and microarray hybridization

Total RNA was isolated from stable clones of DU145 overexpressing *LEADR* transcripts (gene, RefSeq or ΔAlu) or from RWPE-1 cells transiently silenced for *LEADR* expression (gapLEADR, siLEADR or gapINT1) 72 hours upon transfection. RNA was extracted with Qiazol reagent and purified with QIAGEN RNeasy mini kit (QIAGEN, Hilden, Germany) in accordance with the prescribed protocol. DNase I treatment was performed, and the quality of extracted RNA was evaluated through Agilent Bioanalyzer. Eight hundred ng of total RNA was reverse transcribed, labelled with biotin, and amplified overnight using the Illumina RNA TotalPrep Amplification kit (Thermo Fisher) according to manufacturer's protocol. One μg of the biotinylated cRNA was mixed with the Hyb E1 hybridization buffer and hybridized to the Illumina HumanHT-12 v4 chips (47,323 probes) at 58 °C overnight. Array chips were washed with manufacturer's E1BC solution, stained with 1 μg/mL Cy3-streptavidine, and scanned with the Illumina BeadArray Reader according to the standard Illumina protocols.

2.3. Preprocessing methods

Raw data were $\log_2$-transformed and normalized by using the rsn method implemented in the *lumi* package to deal with intensity-dependent and spatial bias [5]. Furthermore, normalized data were filtered by retaining probes with both at least one detection p-value < 0.01 across samples and an associated official gene symbol. For probes mapping on the same gene symbol, the one with the highest variance was selected.

2.4. Differential expression analyses

Differential expression was measured in terms of fold change (FC) using the *limma* package of R environment [6]. In general, significance was assessed in terms of False Discovery Rate (FDR) to manage

Figure 4a). Hallmark collection of MSigDB database is sorted according to the FDR associated to the enrichment score (c). Scatter plot showing the linear trend between fold-changes of selected genes upon gapLEADR silencing, as measured by microarray platform and qRT-PCR as technical validation. Equation of linear regression as well as r-squared (R^2) are plotted (d).

Table 1
GEO repository. Description of the three deposited datasets; the cell line (Source), the type of modulation (Protocol 1), the sample manipulation (Protocol 2), the hybridization (Protocol 3), the description of the sample (Samples), and the correspondent GEO accession number (Data) are indicated. Different colors are used to distinguish the three different datasets.

Source	Protocol 1	Protocol 2	Protocol 3	Samples	Data
RWPE-1 cells	control	RNA extraction	microarray hybridization	control siRNA	GEO103654
RWPE-1 cells	silencing	RNA extraction	microarray hybridization	LEADR siRNA	GEO103654
RWPE-1 cells	control	RNA extraction	microarray hybridization	control siRNA	GEO103654
RWPE-1 cells	silencing	RNA extraction	microarray hybridization	LEADR siRNA	GEO103654
RWPE-1 cells	control	RNA extraction	microarray hybridization	control siRNA	GEO103654
RWPE-1 cells	silencing	RNA extraction	microarray hybridization	LEADR siRNA	GEO103654
RWPE-1 cells	control	RNA extraction	microarray hybridization	control gapmer	GEO103654
RWPE-1 cells	silencing	RNA extraction	microarray hybridization	LEADR intronic gapmer	GEO103654
RWPE-1 cells	control	RNA extraction	microarray hybridization	control gapmer	GEO103654
RWPE-1 cells	silencing	RNA extraction	microarray hybridization	LEADR intronic gapmer	GEO103654
RWPE-1 cells	control	RNA extraction	microarray hybridization	control gapmer	GEO103654
RWPE-1 cells	silencing	RNA extraction	microarray hybridization	LEADR intronic gapmer	GEO103654
DU145 cells	control	RNA extraction	microarray hybridization	empty vector	GEO103654
DU145 cells	overexpression	RNA extraction	microarray hybridization	LEADR RefSeq	GEO103654
DU145 cells	overexpression	RNA extraction	microarray hybridization	LEADR gene	GEO103654
DU145 cells	control	RNA extraction	microarray hybridization	empty vector	GEO103654
DU145 cells	overexpression	RNA extraction	microarray hybridization	LEADR RefSeq	GEO103654
DU145 cells	overexpression	RNA extraction	microarray hybridization	LEADR gene	GEO103654
DU145 cells	control	RNA extraction	microarray hybridization	empty vector	GEO103654
DU145 cells	overexpression	RNA extraction	microarray hybridization	LEADR RefSeq	GEO103654
DU145 cells	overexpression	RNA extraction	microarray hybridization	LEADR gene	GEO103654
DU145 cells	control	RNA extraction	microarray hybridization	empty vector	GEO103654
DU145 cells	overexpression	RNA extraction	microarray hybridization	LEADR RefSeq	GEO103654

DU145 cells	overexpression	RNA extraction	microarray hybridization	LEADR gene	GEO103654
DU145 cells	control	RNA extraction	microarray hybridization	empty vector	GEO103654
DU145 cells	overexpression	RNA extraction	microarray hybridization	Alu-deleted LEADR	GEO103655
DU145 cells	control	RNA extraction	microarray hybridization	empty vector	GEO103655
DU145 cells	overexpression	RNA extraction	microarray hybridization	Alu-deleted LEADR	GEO103655
DU145 cells	control	RNA extraction	microarray hybridization	empty vector	GEO103655
DU145 cells	overexpression	RNA extraction	microarray hybridization	Alu-deleted LEADR	GEO103655
DU145 cells	control	RNA extraction	microarray hybridization	empty vector	GEO103655
RWPE-1 cells	control	RNA extraction	microarray hybridization	control gapmer	GEO103656
RWPE-1 cells	control	RNA extraction	microarray hybridization	control gapmer	GEO103656
RWPE-1 cells	control	RNA extraction	microarray hybridization	control gapmer	GEO103656
RWPE-1 cells	silencing	RNA extraction	microarray hybridization	LEADR gapmer	GEO103656
RWPE-1 cells	silencing	RNA extraction	microarray hybridization	LEADR gapmer	GEO103656
RWPE-1 cells	silencing	RNA extraction	microarray hybridization	LEADR gapmer	GEO103656

multiple test comparisons (threshold of 0.05). However, only few genes resulted significantly differentially expressed in terms of FDR correction in overexpression experiments. In this regard, such experiments have an intrinsic higher variability due to biological and technical causes. For example, since the DU145 cell line model does not constitutively express basal features, the sole reconstitution of *LEADR* could not be sufficient to induce the acquisition of a frankly basal phenotype, thus resulting in a hybrid state with marginal gene expression changes. Moreover, transfected cells have had the possibility to settle into plasmid integration for a long time, and they might have therefore adapted to this condition (in our case, we collected transcriptomic profile of stable clones of DU145 cells, while for silencing experiments performed in RWPE-1 cells we evaluated gene expression acutely 3 days after transfection). In this regard, we reasoned that FDR is a too restrictive correction for overexpression experiments, and we considered as significantly differentially expressed all genes with at least a nominal p-value < 0.05.

The overrepresentation analysis was conducted retaining only the significant enrichment scores (FDR threshold of 0.05).

2.5. Analysis of qRT-PCR data

In TaqMan-based qRT-PCR, we calculated FCs as $2^{-\Delta\Delta Ct}$ and averaged them among biological replicates. Specifically, for each *i*-measurement, FC values were estimated as the differential expression between gapLEADR and gapCTR, considering *GAPDH* as endogenous control as follows:

$$\Delta Ct_i = Ct_{LEADR_i} - Ct_{GAPDH_i}$$
$$-\Delta\Delta Ct = -\left(\Delta Ct_{i_{gapLEADR}} - \Delta Ct_{i_{gapCTR}}\right)$$
$$FC = 2^{-\Delta\Delta Ct}$$

Authors' contribution

Stefano Percio: Methodology, Validation, Formal Analysis, Writing – original draft.
Federica Rotundo: Methodology, Validation, Investigation, Writing – original draft.
Paolo Gandellini: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

We thank the Platform of Integrated Biology of Fondazione IRCCS Istituto Nazionale dei Tumori for microarray experiments. This work was supported by grants from: Italian Ministry of Health (GR-2013-02355625 to P.G.) and Cariplo Foundation (2015-0866 to P.G.).

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2020.105139.

References

[1] V. Profumo, B. Forte, S. Percio, F. Rotundo, V. Doldi, E. Ferrari, N. Fenderico, M. Dugo, D. Romagnoli, M. Benelli, R. Valdagni, D. Dolfini, N. Zaffaroni, P. Gandellini, LEADeR role of miR-205 host gene as long noncoding RNA in prostate basal cell differentiation, Nat. Commun. 10 (2019) 307, https://doi.org/10.1038/s41467-018-08153-2.
[2] S. Di Agostino, F. Valenti, A. Sacconi, G. Fontemaggi, M. Pallocca, C. Pulito, F. Ganci, P. Muti, S. Strano, G. Blandino, Long noncoding MIR205HG depletes Hsa-miR-590-3p leading to unrestrained proliferation in head and neck squamous cell carcinoma, Theranostics 8 (2018) 1850–1868, https://doi.org/10.7150/thno.22167.
[3] Y. Li, H. Wang, H. Huang, Long non-coding RNA MIR205HG function as a ceRNA to accelerate tumor growth and progression via sponging miR-122–5p in cervical cancer, Biochem. Biophys. Res. Commun. 514 (2019) 78–85, https://doi.org/10.1016/j.bbrc.2019.04.102.
[4] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. 102 (2005) 15545–15550, https://doi.org/10.1073/pnas.0506580102.
[5] P. Du, W.A. Kibbe, S.M. Lin, Lumi: a pipeline for processing Illumina microarray, Bioinformatics 24 (2008) 1547–1548, https://doi.org/10.1093/bioinformatics/btn224.
[6] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (2015), e47, https://doi.org/10.1093/nar/gkv007.