scientific reports

OPEN

An R package for generic modular response analysis and its application to estrogen and retinoic acid receptor crosstalk

Gabriel Jimenez-Dominguez[1,2,3], Patrice Ravel[1,2,3], Stéphan Jalaguier[1,2,3], Vincent Cavaillès[1,2,3✉] & Jacques Colinge[1,2,3✉]

Modular response analysis (MRA) is a widely used inference technique developed to uncover directions and strengths of connections in molecular networks under a steady-state condition by means of perturbation experiments. We devised several extensions of this methodology to search genomic data for new associations with a biological network inferred by MRA, to improve the predictive accuracy of MRA-inferred networks, and to estimate confidence intervals of MRA parameters from datasets with low numbers of replicates. The classical MRA computations and their extensions were implemented in a freely available R package called aiMeRA (https://github.com/bioinfo-ircm/aiMeRA/). We illustrated the application of our package by assessing the crosstalk between estrogen and retinoic acid receptors, two nuclear receptors implicated in several hormone-driven cancers, such as breast cancer. Based on new data generated for this study, our analysis revealed potential cross-inhibition mediated by the shared corepressors NRIP1 and LCoR. We designed aiMeRA for non-specialists and to allow biologists to perform their own analyses.

Modular response analysis (MRA) was introduced to infer the directions and strengths of connections between components of a biological system in a steady-state[1]. This approach can be applied to components at different levels of detail, *e.g.*, individual genes or subsystems, such as pathways or processes. MRA is based on the perturbation of individual components, the so-called modules. Various developments of MRA and related methods were recently reviewed[2]. In this report, we applied and extended MRA methodology to explore the crosstalk between two nuclear receptors (NRs). This resulted in the implementation of a new R[3] package called "aiMeRA".

Estrogen receptors (ERs) belong to the NR superfamily, which function as transcription factors, and are activated upon ligand binding. Both ER isoforms (ERα and ERβ) are involved in the control of cell proliferation and exhibit essential functions in tissue development and homeostasis, particularly in organs related to reproduction[4]. ERα overexpression is frequently observed in breast, ovarian, endometrial, and other hormone-driven tumors. The transcriptional activity of ERs is modulated by several coregulatory complexes, including coactivators and corepressors[4]. In the presence of estrogens or any agonist ligand, ERs interact preferentially with coactivators, or with a specific subclass of corepressors, including the nuclear receptor-interacting protein 1 (NRIP1, often named RIP140) and the ligand-dependent corepressor (LCoR). NRIP1 is a corepressor of particular interest because its expression is directly induced by estrogen, *i.e.*, NRIP1 installs a negative feedback loop to maintain ER signaling under control[5]. Indeed, abnormal NRIP1 expression is observed in ER-driven tumors[6,7]. LCoR represses the transcription of estrogen-induced gene expression[8], and NRIP1 expression has been shown necessary for LCoR inhibitory activity in breast cancer (BC) cells[9].

Interestingly, NRIP1 and LCoR function as corepressors for several ligand-dependent NRs. For instance, LCoR can repress the vitamin D receptor (VDR), retinoic acid receptors (RARs), but also RXR ligand-dependent transcription[8] in addition to ERs. Moreover, NRIP1 is a known direct target and negative regulator of RAR transcription[10].

There is experimental evidence of crosstalk between ER and RAR signaling[11]. For instance, ERα can suppress the basal expression of retinoic acid (RA)-responsive gene RARβ2, but also turns out to be necessary for its RA induction[12]. It was also found that ERα activates RARα1 expression in BC cells[13]. Other authors intersected

[1]Inserm U1194, Institut de Recherche en Cancérologie de Montpellier, Montpellier, France. [2]University of Montpellier, Montpellier, France. [3]ICM, Institut régional du Cancer de Montpellier, 208 avenue des Apothicaires, 34298 Montpellier cedex 5, France. ✉email: vincent.cavailles@inserm.fr; jacques.colinge@inserm.fr

1

RAR targets identified by chromatin immunoprecipitation-sequencing (ChIP-seq) with ER binding sites and discovered a significant overlap[14]. This study suggested a mechanism of space competition for ER and RA cross-talk in BC cells. A potential cooperative interaction between RARα and ER was also shown in BC cells[15]. Since NRIP1 and LCoR expression can both be regulated by RAR and ER transcription, we can further hypothesize that these molecules mediate part of ER-RAR crosstalk. The induced expression of NRIP1 and LCoR by one receptor produces molecules able to subsequently repress signaling of both receptors.

The first objective of this study was hence to explore the ER-RAR crosstalk at the transcriptional level by characterizing the ER-RAR-NRIP1-LCoR network by means of transcript abundance measurements and MRA. We considered a steady-state condition in a BC (MCF7-derived) cell line that would model BC cells under hormonal stimulation. We conducted perturbation experiments to generate quantitative PCR (qPCR) and mRNA sequencing (RNA-seq) data.

Given the nature of ER and RAR, *i.e.*, transcription factors, and the general ability of MRA to perform predictions[16], we introduced two extensions of MRA in a second part of this study. We exploited the RNA-seq data and tested whether the ER-NRIP1-LCoR network inferred by MRA could be used to search for novel genes under strong ER transcriptional control. Then, we used the same ER-NRIP1-LCoR network to predict the expression levels of E2-regulated and ER target genes that were not perturbed during the construction of the MRA-inferred network. These variations from classical MRA inference, but also a new procedure to estimate confidence intervals (CIs) around MRA parameter estimates from a very low number of replicates, are available in the R package developed in this study.

Based on the application of multivariate calculus, we also present a straightforward mathematical derivation of MRA.

To put our work in perspective, other authors have extended the classical MRA approach, *e.g.*, by introducing Bayesian variable selection to better infer pathway topologies[17], by pruning edges for a similar purpose[16], or by dealing with incomplete perturbation sets[18–20]. The Blüthgen Laboratory recently released another R package to perform MRA computations[21] with a specific focus on their edge-pruning and associated maximum likelihood extension of MRA[16].

Methods

Mathematical derivation. The seminal MRA publication by Kholodenko and colleagues introduced the concept of inferring interdependencies (connection coefficients or local response coefficients) modularly within a biological system[1]. That is, subsystems comprised of molecules and their relationship at a detailed level, which is outside the scope of this study, could be captured as a single module with one measurable quantity describing the overall module activity. For instance, in the case of ERα signaling, the complex process of ligand binding and transcriptional activity can be represented by a single module (Fig. 1A); the activity of this module is measured in MELN cells by the reporter gene luciferase. NRIP1 and LCoR activities were determined by their respective mRNA abundances. In this study, we assessed the dependence of each module activity with respect to that of the others; we computed connection coefficients to assign the directed edges of Fig. 1A. The solution was found in a steady-state after applying successive elementary perturbations on each module activity. Depending on the application, this framework can be applied to different molecular species and processes, *e.g.*, protein/metabolite concentrations or protein phosphorylation levels[2].

Now, in full generality, we assume that there are n modules whose activities are given by $x \in \mathbb{R}^n$. We further admit the existence of n intrinsic parameters, $p \in \mathbb{R}^n$, one per module, each of them perturbed by elementary perturbations. One can imagine mRNA abundance parameters for perturbations, such as siNRIP1 or siLCoR, and the number of available ERα-E2 bound complexes for the E2 perturbation. In other circumstances, perturbations may change affinity constants or other physical parameters. Lastly, we assume that there exists $S \subset \mathbb{R}^n \times \mathbb{R}^n$, an open subset, and $f : S \to \mathbb{R}^n$ of class $\mathcal{C}^1$, *i.e.*, continuously differentiable, such that.

$$\dot{x} = f(x, p). \tag{1}$$

We do not need to know $f(x, p) = (f_1(x, p), \ldots, f_n(x, p))^t$ explicitly, but we need one more hypothesis that is the existence of a time $T > 0$ such that all the solutions we consider for any p and initial conditions of x, have reached a steady-state, *i.e.*,

$$\dot{x} = 0, \forall t > T.$$

The unperturbed, basal state of the modules is denoted $x(p^0) \in \mathbb{R}^n$ and it has corresponding parameters $p^0 \in \mathbb{R}^n$. According to our hypotheses, $f(x(p^0), p^0) = 0 \Leftrightarrow f_i(x(p^0), p^0) = 0, \forall i \in$. By the implicit function theorem, $\forall i$, there exists open neighborhoods $V_i \times W_i \subset \mathbb{R}^{n-1} \times \times \mathbb{R}^n$ of $(x_1^0, \ldots, x_{i-1}^0, x_{i+1}^0, \ldots, x_n^0, p_1^0, \ldots, p_n^0)$ and $U_i \subset \mathbb{R}$ of x_i^0, as well as $g_i : V_i \times W_i \to U_i$ (also $\mathcal{C}^1$) with

$$f_i(x_1, \ldots, x_{i-1}, g_i(\cdots), x_{i+1}, \ldots, x_n, p_1, \ldots, p_n) = 0, \quad \forall (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n, p_1, \ldots, p_n) \in V_i \times W_i \tag{2}$$

We denote $x(p^0 + \Delta p)$, the steady-state corresponding to the changed parameters $p^0 + \Delta p$, *i.e.*, the solution of $\dot{x}(p^0 + \Delta p) = f(x(p^0 + \Delta p), p^0 + \Delta p^0)$. We introduce the notation $x_{j \neq i}$ to denote all the x_j but x_i. Now, if we assume that $(x_{j \neq i}(p^0 + \Delta p), p^0 + \Delta p)$ belong to $V_i \times W_i$ for all the perturbations considered experimentally, then by Taylor's Formula
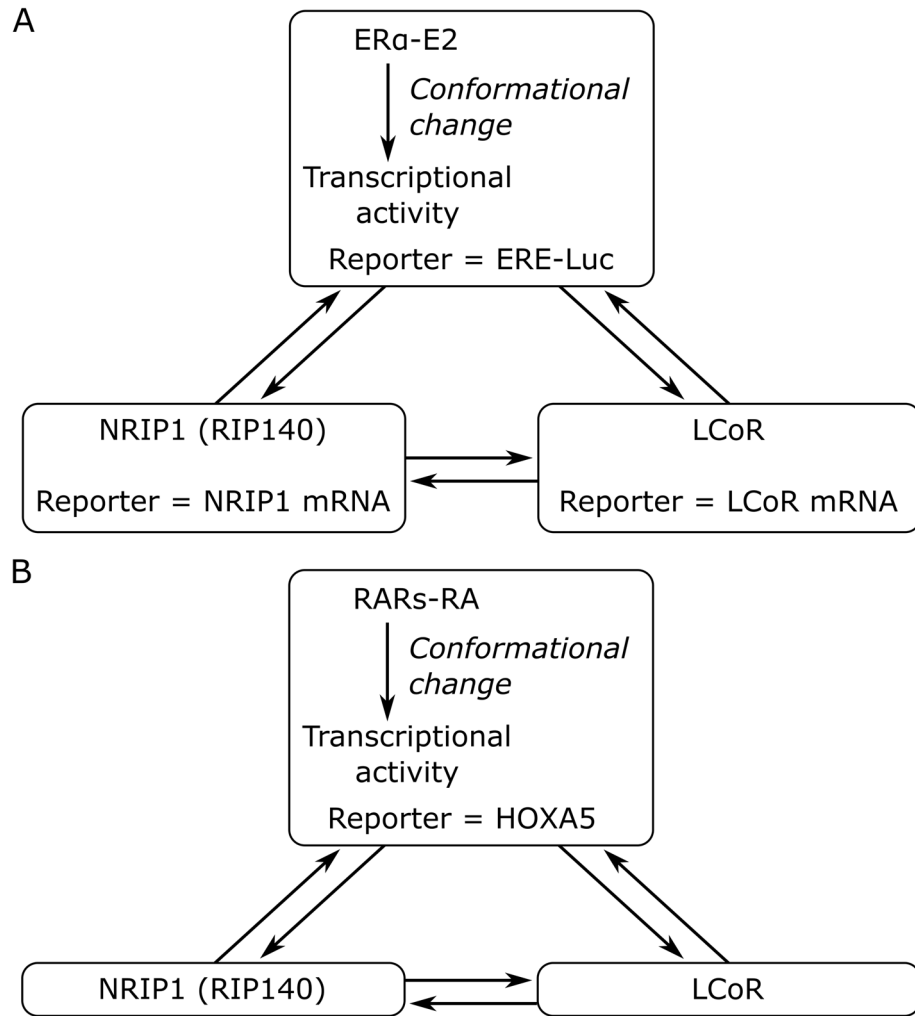
Figure 1. MRA general principle. (**A**) The ERα-NRIP1-LCoR transcriptional network. The activity level of each module is given by a measured reporter. Connection coefficients (edge weights) are determined from perturbation experiments. (**B**) The RARs-NRIP1-LCoR transcriptional network. Figure created with Inkscape 0.92 (www.inkscape.org).

$$x_i(p^0 + \Delta p) = x_i(p^0) + g_i'(p^0) \begin{pmatrix} x_1(p^0 + \Delta p) - x_1(p^0) \\ \vdots \\ x_{i-1}(p^0 + \Delta p) - x_{i-1}(p^0) \\ x_{i+1}(p^0 + \Delta p) - x_{i+1}(p^0) \\ \vdots \\ x_n(p^0 + \Delta p) - x_n(p^0) \\ \Delta p_1 \\ \vdots \\ \Delta p_n \end{pmatrix} + \sigma(\|\Delta p\|). \quad (3)$$

Dividing each side by $x_i(p^0)$, Eq. (3) can be rewritten

$$\frac{x_i(p^0 + \Delta p) - x_i(p^0)}{x_i(p^0)} =$$

$$\sum_{j=1, j \neq i}^{n} \frac{x_j(p^0)}{x_i(p^0)} \frac{\partial g_i}{\partial x_j}(p^0) \left(\frac{x_j(p^0 + \Delta p) - x_j(p^0)}{x_j(p^0)} \right) + \frac{1}{x_i(p^0)} \sum_{j=1}^{n} \frac{\partial g_i}{\partial p_j}(p^0)(\Delta p_j) + \sigma(\|\Delta p\|). \quad (4)$$

Since the parameter p_j influences the module j only, $\frac{\partial g_i}{\partial p_j} = 0$ if $j \neq i$. Moreover, $g_i(x_{j \neq i}, p) = x_i(x_{j \neq i}, p)$ in $V_i \times W_i$, and if we denote

$$\frac{\Delta x_i}{x_i} = \frac{x_i(p^0 + \Delta p) - x_i(p^0)}{x_i(p^0)},$$

and

$$r_{i,j} = \frac{x_j(p^0)}{x_i(p^0)} \frac{\partial x_i}{\partial x_j}(p^0), \quad j \neq i, \tag{5}$$

Then

$$\frac{\Delta x_i}{x_i} = \sum_{j \neq i} r_{i,j} \left(\frac{\Delta x_j}{x_j} \right) + \frac{1}{x_i(p^0)} \frac{\partial x_i}{\partial p_i}(p^0)(\Delta p_i) + \sigma(\|\Delta p\|). \tag{6}$$

We next consider elementary perturbations q_k, $k \in \{1, \ldots, n\}$, which only influence the module k, i.e., the parameter p_k. Neglecting the second-order term $\sigma(\|\Delta p\|)$ and writing

$$\left(\frac{\Delta x_i}{x_i} \right)_{q_k}$$

the relative difference in activity of module i upon Δp_k change induced by perturbation q_k, we find

$$\left(\frac{\Delta x_i}{x_i} \right)_{q_k} = \sum_{j \neq i} r_{i,j} \left(\frac{\Delta x_j}{x_j} \right)_{q_k}, \quad k \neq i, \tag{7}$$

$$\left(\frac{\Delta x_i}{x_i} \right)_{q_i} = \sum_{j \neq i} r_{i,j} \left(\frac{\Delta x_j}{x_j} \right)_{q_k} + \frac{\partial x_i}{\partial p_i}(p^0) \left(\frac{\Delta p_i}{x_i} \right). \tag{8}$$

By defining $r_{i,i} = -1$, we can write Eqs. (7) and (8) in matrix form:

$$rR = -P, \tag{9}$$

where R is the matrix of systems-level (relative) changes of modules to perturbations $R_{j,k} = \left(\frac{\Delta x_j}{x_j} \right)_{q_k}$, $j, k \in \{1, \ldots, n\}$. P is a diagonal matrix with $P_{i,i} = \frac{\partial x_i}{\partial p_i}(p^0) \left(\frac{\Delta p_i}{x_i} \right)$, $i \in \{1, \cdots, n\}$. The system (9) can be solved in two steps[1]. Firstly, $r = -PR^{-1}$ and since $r_{i,i} = -1$, we have $P_{i,i}(R^{-1})_{i,i} = 1$, thus $P_{i,i} = \frac{1}{(R^{-1})_{i,i}}$. Secondly,

$$r = -\left[\mathrm{diag}(R^{-1}) \right]^{-1} R^{-1}.$$

The elements of R are defined as $\left(\frac{x_i(p^0 + \Delta p_k) - x_i(p^0)}{x_i(p^0)} \right)_{q_k}$, but as suggested previously[1], we rather estimated this quantity by

$$R_{j,k} = 2 \left(\frac{x_i(p^0 + \Delta p_k) - x_i(p^0)}{x_i(p^0 + \Delta p_k) + x_i(p^0)} \right)_{q_k}, \tag{10}$$

which is numerically more stable and divisions by 0 are avoided.

Lastly, from Eq. (5), we see that $r_{i,j}$ contains the connection coefficients between MRA modules: the direct action of j on i normalized by the ratio x_j/x_i. Similarly, $P_{i,i}$ measures the relative effect of q_i on x_i. We call it q_i magnitude. The implicit function theorem provides analytical expressions for g_i' in terms of f partial derivatives, but since f is generally unknown we did not use them. To be rigorous, one should ultimately restrict the mathematical model to a neighborhood of $(x(p^0), p^0)$ included in all V_i's, W_i's, and U_i's.

MRA-inferred networks have been largely applied for their predictive capabilities [16]. Let us define a multiple perturbation q to be the linear combination of elementary perturbations q_k. For instance, a perturbation on modules i and j with the same individual magnitudes would be represented by a column vector c with 1's at positions i and j and 0's elsewhere. From Eq. (9), we compute

$$\left(\frac{\Delta x}{x} \right)_q = -r^{-1}Pc, \tag{11}$$

with $\left(\frac{\Delta x}{x} \right)_q$ the column vector containing the inferred relative changes on each module activity. Denoting Δp the parameter changes induced by q, the individual module activities are given by

$$\left(\frac{\Delta x_i}{x_i}\right)_q = -\left(r^{-1}Pc\right)_i = 2\frac{x_i\left(p^0 + \Delta p\right) - x_i\left(p^0\right)}{x_i\left(p^0 + \Delta p\right) + x_i\left(p^0\right)}$$

$$\Leftrightarrow x_i\left(p^0 + \Delta p\right) = \frac{x_i\left(p^0\right)}{1 + \frac{2}{\left(r^{-1}Pc\right)_i}}\left(\frac{2}{\left(r^{-1}Pc\right)_i} - 1\right). \tag{12}$$

If elementary perturbations contribute to different amounts to q, the vector c contains q_k's relative weights. In all cases, linearity between the perturbation strength and its impact on p is assumed.

Confidence interval estimations. CIs of the MRA parameters were estimated using a bootstrap procedure[22], an approach that was followed by other authors already[23,24]. In our implementation, we made special efforts to account for the experimental design with biological and technical replicates and to address the difficulty of working with a limited number of replicates. This was achieved by employing tools of statistical process control (SPC) theory[25,26]. SPC is a corpus of statistical methods developed after WW2 to control industrial production processes. One main application is the detection of fluctuations that are not caused by the intrinsic variability of a process. SPC provides estimators of the variance that were optimized for small series (randomly selected samples from a production line in the industrial setting) as well as a procedure to identify the samples outside the expected variability (outliers for our application). SPC can be used for normally distributed data, which will be the case here, but also in a nonparametric manner. In addition to SPC, our bootstrap procedure was also influenced by recent results discussing the propagation of noise in the data across MRA calculations[27]. These results showed that CI estimations on MRA parameters are best obtained by considering the noise on global response coefficients. Accordingly, a global response matrix R is computed for each of the 6 replicates of the complete set of qPCR measures, $i.e.$ a sample of 6 values is available for each $R_{j,k}$. SPC allows for the estimation of its intra-transfection standard deviation $s_{j,k}^{\text{intra}}$ considering the replicate structure: three biological replicates ($k_t = 3$) with two technical replicates ($n_r = 2$) each. According to SPC methodology, the standard deviation estimator is constructed by computing the mean of the 3 sample ranges, $i.e.$, the mean the absolute values of the differences between the two replicates of each transfection in our particular configuration, divided by Hartley's constant $d_2(n_r)$. Namely,

$$s_{j,k}^{\text{intra}} = \frac{\sum_{i=1}^{k_t} \text{Ra}_{j,k}^i}{k_t d_2(n_r)},$$

with $\text{Ra}_{j,k}^i = \max\limits_{0 \leq r \leq n_r} R_{j,k}^{i,r} - \min\limits_{0 \leq r \leq n_r} R_{j,k}^{i,r}$, where r is the technical replicate index and i the transfection index. Note that $d_2(2) = 1.128$[28] ($n_r = 2$ here). We did not use SPC to exclude samples outside the expected variability for qPCR data, but we did for the even more sparse RNA-seq data; see the specific Results subsection for the details. Lastly, one million R matrices were generated sampling normal distributions, one distribution for each $R_{j,k}$ with mean equal to the average of the 6 raw data replicates and specific variance obtained by SPC. Note that, in practice, different $R_{j,k}$ displayed very different variances in some cases on the contrary to Thomaseth et al. study[27]. The 95% CIs were obtained from the 2.5th and 97.5th percentiles. Departure from the assumed normal noise in the data was tested by Kolmogorov–Smirnov tests (data not shown), which detected no significant departure. This does not prove normality (no such test exists in statistics), especially a small number of replicates, but it clearly supports that the estimated 95% CIs are reasonable first approximation.

Cell culture and perturbation experiments. We used MELN cells, an MCF7-derived cell line stably transfected with the estrogen-responsive luciferase reporter gene ERE-βGlob-Luc-SV-Neo[29]. The cell line was authenticated by short tandem repeat profiling and tested for mycoplasma contamination.

MELN cells were cultured in phenol red-free Dulbecco's modified Eagle medium (Gibco) containing 5% dextran-charcoal treated FCS (Invitrogen) and antibiotics (Gibco). Perturbations at NRIP1 and LCoR were obtained by siRNAs that were transfected using Interferrin (Polyplus). Perturbations at ERα and RARs were induced by their respective natural ligands: the hormones estrogen (17β-estradiol or E2 for short) and all-trans retinoic acid (RA).

MELN cells were obtained in the following conditions: basal (untreated), E2, RA, E2 + RA, siNRIP1, siLCOR, siNRIP1 + siLCOR, E2 + siNRIP1, E2 + siLCOR, E2 + siNRIP1 + siLCOR, RA + siNRIP1, RA + siLCOR, RA + siN-RIP1 + siLCOR, E2 + RA + siNRIP1, E2 + RA + siLCOR, and E2 + RA + siNRIP1 + siLCOR. These experiments were realized in triplicates. We waited 24 h after siRNA transfection before treating with ligands. Cells were harvested after 18 h of culture under ligand stimulation. E2-treated cells received 100 nM E2, RA-treated cells 10 uM RA, and untreated cells ethanol. qPCR measures were realized in duplicates for each of the 3 biological replicates. Validations of the response to E2 and siRNA interference are in Suppl. Fig. 1.

mRNA quantification. RNA was isolated using "Quick-RNA MiniPrep" (Zymo Research) and reverse transcription (RT)-qPCR assays were done using qScript (VWR) according to the manufacturer's protocol. Transcripts were quantified using SensiFAST SYBR (BioLine) on an LC480 instrument. The nucleotide sequences of the primers used for real-time PCR were:

RIP140-f (5′- AATGTGCACTTGAGCCATGATG -3′),
RIP140-r (5′- TCGGACACTGGTAAGGCAGG -3′),
LCoR-f (5′- GAACCTAGCGAACAAGACGGTG -3′),
LCoR-r (5′- TGGAGAGTGGCTCAGGGAAGT -3′),

Luciferase-f (5′- CTCACTGAGACTACATCAGC -3′),
Luciferase-r (5′- TCCAGATCCACAACCTTCGC -3′),
HOXA5-f (5′- GCGCAAGCTGCACATAAGTC -3′),
HOXA5-r (5′- GAACTCCTTCTCCAGCTCCA -3′),
ERα-f (5′- TGGAGATCTTCGACATGCTG -3′),
ERα-r (5′- TCCAGAGACTTCAGGGTGCT -3′),
RARα-f (5′- GGATATAGCACACCATCCCC -3′),
RARα-r (5′- TTGTAGATGCGGGGTAGAGG -3′),
PGR-f (5′- CGCGCTCTACCCTGCACTC-3′),
PGR-r (5′-TGAATCCGGCCTCAGGTAGTT-3′).
(RT)-qPCR data are available from the R package.

mRNA sequencing.   For two of the triplicates, in each condition, RNA was extracted as above described. Libraries were prepared with Illumina TruSeq kit and submitted to NextSeq500 sequencing (1×75 bp/40 M reads). The first 13 and last 7 bps were cut by an in-house Perl script to eliminate compositional bias. Cut reads were submitted to sickle to eliminate remaining low-quality regions. Alignments were performed against the human genome (hg38) with TopHat v2.10[30] and read counts extracted with HTSeq-Count[31]. The read count matrix was normalized with edgeR[32] TMM algorithm. Data are available from GEO under GSE143956.

aiMeRA library implementation.   We implemented the MRA method according to the mathematical formulation above as an R library. (RT)-qPCR data of this project were embedded in the R library for convenience and to provide an example. We also included the data used in the MRA original paper[1] such that users can check that our code gives the same results as those reported in the latter publication.

Results

Transcriptional data.   Given ERβ and RARβ expression could not be quantified in MELN cells, we opted for networks involving an ERα module with a reported transcriptional activity by the ERE/luciferase construct, *i.e.*, luciferase mRNA abundance measured ERα activity. ERα mRNA abundance would combine both ligand-bound and free receptors, but only the ligand-bound receptors are relevant for the MRA-inferred network. RARα *versus* RARγ was not distinguished; we estimated their combined transcriptional activity by *HOXA5* gene mRNA abundance, and the corresponding MRA module was named RARs (Fig. 1B). NRIP1 and LCoR activity was determined by their gene mRNA abundance. Given that MELN cells are BC cells, we assumed E2-, RA-, or E2 & RA-stimulated conditions to be basal, *i.e.*, perturbations of ERα and RARs were negative (switch to ethanol condition). Perturbations of NRIP1 and LCoR were achieved by siRNAs, *i.e.*, they were also negative.

The ERα-NRIP1-LCoR network.   In an unstimulated condition (E2 absence), it is well-known that NRIP1 expression induces LCoR expression[9]. We confirmed the presence of this crosstalk under the basal E2-stimulated condition (Fig. 2A); in addition, we also observed a negative connection coefficient of LCoR to NRIP1. This is in accordance with the former assignment of NRIP1 as a direct target of E2-bound ERα, and LCoR one of its corepressors. The global response coefficients in matrix *R* were obtained comparing NRIP1 and LCoR expression under E2 stimulation with the successive E2 + siLCoR and E2 + siNRIP1 perturbations. Perturbations on ERα were obtained comparing the E2 stimulated condition with the ethanol condition. Next, we tested the ERα-NRIP1-LCoR network under the same E2-stimulated condition (Fig. 2B). We recapitulated the known induction of NRIP1 by ERα with a negative feedback[5]; we also reconstituted the known inhibition of ERα by LCoR[8]. Interestingly, the induction of LCoR upon NRIP1 expression observed in Fig. 2A became an inhibition. Indeed, this apparent contradiction is correct: in the inferred network including ERα (Fig. 2B), a strong double inhibition via ERα acts positively on LCoR and dominates the, now weak, direct negative connection coefficient. In other words, the 2-module NRIP1-LCoR only reflects the global responses, confounded by the missing ERα, whereas the 3-module ERα-NRIP1-LCoR network disentangles the global to local responses. Moreover, this makes sense biologically since there is no direct transcriptional control of LCoR by NRIP1; NRIP1 can only modulate ERα activity. The same reasoning applies to LCoR inhibition of NRIP1 that became an activation, as shown in Fig. 2A. The strong inhibition of ERα by LCoR, which in turn activates NRIP1 in Fig. 2B, results in a strong indirect inhibitory action counterbalancing the moderate direct positive connection coefficient. Perturbation magnitudes are reported in Fig. 2C. Finally, we assessed the validity of the inferred network by checking its predictive power in a validation experiment where the activity of each module was measured under double siNRIP1 & siLCoR perturbation. In Fig. 2D, reasonable fidelity of the predictions was noted, with relative errors proportional to the CI sizes, *i.e.*, with data variability. Connection coefficients whose 95% CI excluded 0 were marked by an asterisk (Fig. 2A,B,E,F); the sign of such coefficients was known at 5% significance. One may consider the removal of non-marked edges because they could be regarded as negligible (although in most cases where 0 lies inside the CI it is rather close to a boundary); such a procedure is outside the scope of this study, but has been investigated by others[16,21].

The RARs-NRIP1-LCoR network.   Subsequently, we built a RARs-NRIP1-LCoR network under RA stimulation (Fig. 1B). In the absence of RARs in the MRA-inferred network, NRIP1 and LCoR connection coefficients remained similar to that under E2 stimulation (Fig. 2E). This was expected since these two corepressors have the same function. Similar to above, the global response matrix *R* coefficients comparing gene expression under RA stimulation with the successive RA + siLCoR and RA + siNRIP1 perturbations. In Fig. 2F, including a RARs module in the inferred networks, we reconstituted the induction of NRIP1 expression by RARs, but also
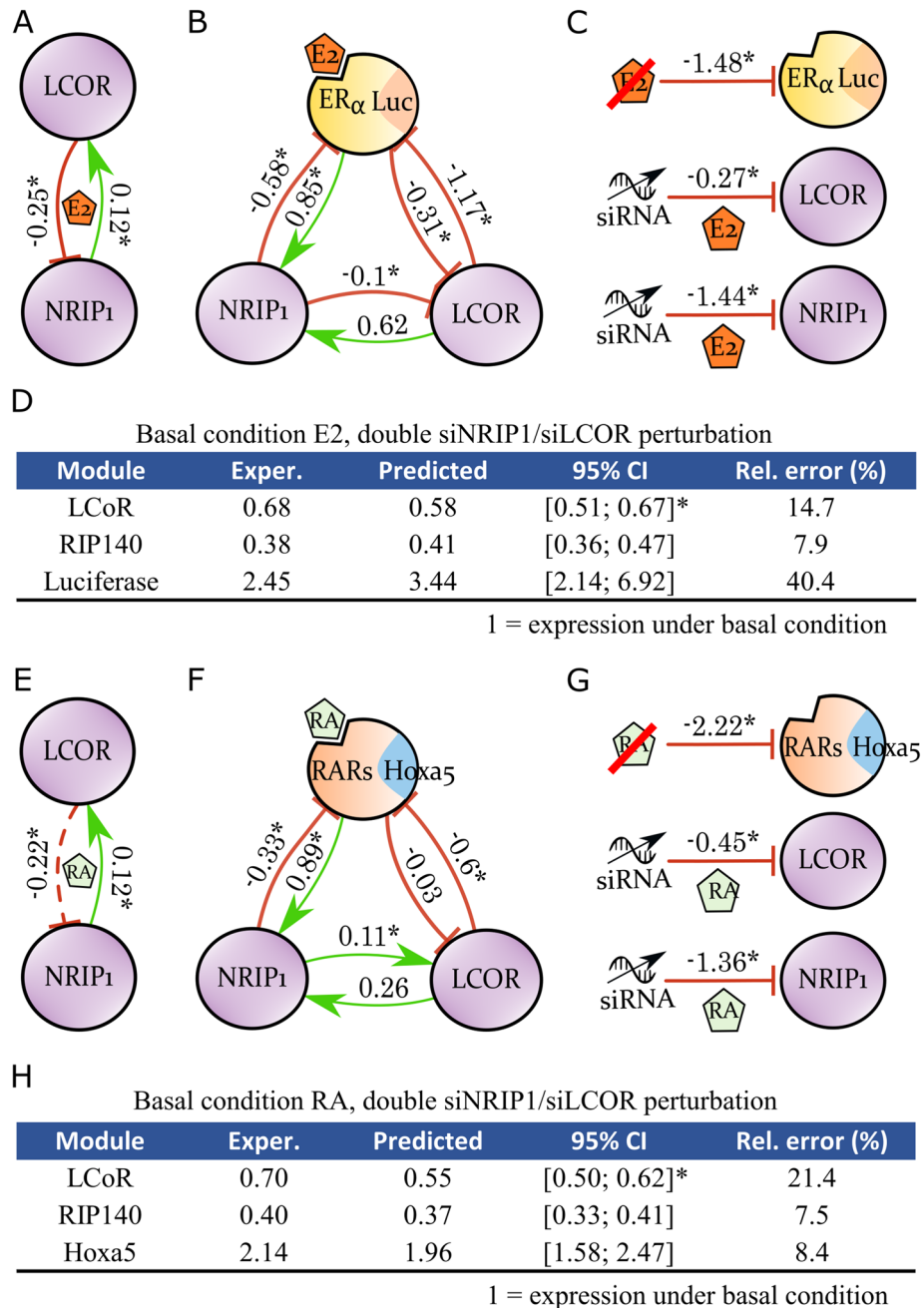
Figure 2. ER and RAR separated networks. (**A**) Connection coefficients between the two corepressors under the E2-stimulated condition. A 95% CI for each MRA parameter was estimated and the parameter denoted by an asterisk provided 0 was not included (sign known with 5% significance). (**B**) The ERα-NRIP1-LCoR transcriptional network. (**C**) ERα, NRIP1, and LCoR perturbation magnitudes. (**D**) Prediction of gene expression under dual siNRIP1 and siLCoR perturbation and E2 stimulation. *Note that LCoR experimental measure lies outside the 95% CI around the predicted value due to higher experimental data variability. A 96% CI equal to [0.45; 0.69] included LCoR experimental value. (**E**) Connection coefficients between the two corepressors under the RA-stimulated condition. (**F**) The RARs-NRIP1-LCoR transcriptional network. (**G**) RARs, NRIP1, and LCoR perturbation magnitudes. (**H**) Prediction of gene expression under dual siNRIP1 and siLCoR perturbation and RA stimulation. *LCoR 99% CI around the predicted value is equal to [0.50; 0.70], it includes LCoR experimental value. Figure created with Inkscape 0.92 (www.inkscape.org) and R 3.6 (r-project. org).

the inhibition of RARs expression by NRIP1[10]. The inhibition of RARs by LCoR has already been established[8]. The LCoR-to-NRIP1 connection coefficient changed of sign between Fig. 2E,F. This connection coefficient was found to be weak in both networks, and the change can be explained using the same arguments as for the ERα-
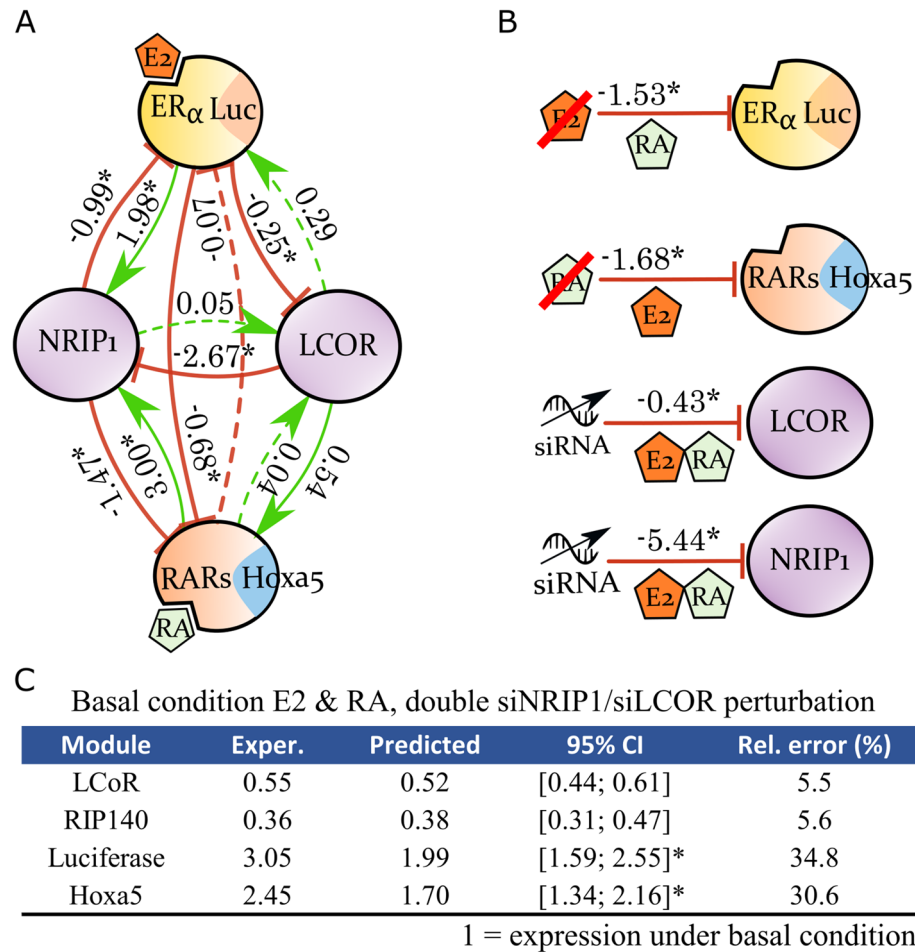
Figure 3. The ERα-RARs-NRIP1-LCoR network. (**A**) MRA-inferred network under dual E2 & RA stimulation. (**B**) Perturbation magnitudes with respect to the basal state E2 & RA stimulation (perturbations on ligands were determined by suppressing one of the two stimulations). (**C**) Predicted activity of the modules upon double siRNA inhibition of NRIP1 and LCoR. *Note that luciferase lies outside its 95% CI due to higher experimental data variability. A 98% CI equal to [1.29; 2.46] included luciferase experimental value. Figure created with Inkscape 0.92 (www.inkscape.org) and R 3.6 (r-project.org).

NRIP1-LCoR network reported above (stronger negative LCoR-to-RARs and positive RARs-to-NRIP1 connection coefficients counterbalance modest and direct positive connection coefficients to equate the weak and negative direct effect in Fig. 2E). We also recognized that the connection coefficient between LCoR and NRIP1 under RA (Fig. 2F) was similar to that under E2 stimulation (Fig. 2B), except for the weaker NRIP1-to-LCoR connection coefficient. This illustrated again the expected and similar functioning of both shared corepressors. NRIP1 perturbation magnitude remained similar, but a twofold decrease was found upon LCoR perturbation although the same siRNAs were used (Fig. 2G). Predictions supported the accuracy of the MRA-derived network in a second validation experiment (Fig. 2H).

The entire ERα-RARs-NRIP1-LCoR network.   We applied the same approach to infer a full network of ER-RARs crosstalk under dual E2 & RA stimulation (Fig. 3A). The global response coefficient matrix R was obtained comparing gene expression under dual RA + E2 stimulation with perturbations induced by RA + E2 + siNRIP1, RA + E2 + siLCoR, E2 (suppression of RA), and RA (suppression of E2). Perturbation magnitudes were in the same range as those we found before, but they were greater for NRIP1 (Fig. 3B). To our knowledge, the interaction between NRIP1 and LCoR under this particular E2 & RA condition has never been investigated; only the crosstalk between RARs and ER has been reported, as mentioned in the introduction of this article[14,15]. Hence, we first challenged the inferred network by testing its predictive accuracy (Fig. 3C), which was again satisfactory in a third validation experiment.

Interestingly, the cross-inhibition of ER and RARs signaling acted down two pathways. The MRA-inferred network showed a direct inhibition of ER transcriptional activity by the RARs module, which has already been described in the literature[14,15]. Reciprocal inhibition was suggested but not significant in our data. In agreement with our hypothesis, a parallel cross-talk mechanism was found through the induction of NRIP1 expression, which could subsequently repress both RARs and ERα. MRA inference thus supported the coexistence of both

phenomena. LCoR reversed its action on NRIP1 compared to the E2 and RA independent conditions. This reversed role could counterbalance the cross-inhibition of both NRs, but the connection coefficients and the much-attenuated induction of LCoR by NRIP1 suggest that it was not the case here.

MRA-inferred networks from RNA-seq data. Since MRA relies on relative changes of module activity (Eqs. (7, 8)), absolute quantitation is not necessary. Therefore, we inferred a RARs-NRIP1-LCoR network based on the HOXA5, NRIP1, and LCoR mRNA abundances that were available in our RNA-seq data (Fig. 4A). Compared with the qPCR-based network shown in Fig. 2F, we observed close values for almost all of the connection coefficients. The only change was the weak LCoR-to-RARs connection coefficient (-0.03) that became slightly positive (0.13) in the RNA-seq data.

CI estimation based on RNA-seq biological replicates under each condition is challenging. In the bootstrap procedure detailed above, we used a variance estimator adapted to limited numbers of replicates (six qPCR replicates) to conduct point estimations. Given we had only duplicates here, this was no longer an option. Therefore, we applied a common approach used in differential gene analysis to overcome this difficulty, *i.e.*, to learn variance collectively across multiple genes. Practically, our bootstrap method simulates replicates of the R matrix; hence we estimated relative changes of gene expression under the different perturbations investigated (RA-removed, siLCoR, and siNRIP1) *versus* the basal condition (RA stimulation). To have a relevant and sufficient sample of genes for this estimation, we used the 1,092 RA-regulated genes based on edgeR analysis (RA *vs.* ethanol conditions, Suppl. Table 1). Let G be an index set for the 1092 genes. R elements for a given perturbation (one column) are assumed to contain an additive Gaussian noise, which for all practical purposes is essentially correct here. The principle of the algorithm is to use SPC to estimate $\sigma_{.,j}$ from all genes under the perturbation q_j, excluding those with replicates that would not be under control according to SPC theory[26,28]. To achieve this, we eliminated such genes iteratively until G no longer changed, as follows:

```
Input: G, an index set for the genes to use in the estimation
```
$G' \leftarrow G$
```
Repeat
```
    $G \leftarrow G'$

    ```
    Compute one R matrix from each duplicate, R¹ and R², and define
    ``` $E_j \leftarrow \left\{ \left| R_{i,j}^1 - R_{i,j}^2 \right| ; i \in G \right\}.$

    ```
    Estimate the standard deviation after SPC:
    ``` $\sigma_{.,j} \leftarrow \frac{\overline{E_J}}{d_2(2)},$ ```with``` $d_2(2) = 1.128^\star$

    ```
    Compute the upper specified limit (USL) for α = 4.6% (corresponding to 2 standard
    deviations in SPC):
    ``` USL $\leftarrow D'_{\alpha/2}(2)\sigma_{.,j},$ ```with``` $D'_{\alpha/2}(2) = 2.834$

    ```
    Eliminate genes with replicates not under control:
    ```
    ```
    For i ∈ G do
    ```
           ```if``` $\left| R_{i,j}^1 - R_{i,j}^2 \right| >$ USL ```then``` $G' \leftarrow G' \backslash \{i\}$

    ```
    End-for
```
```
Until G' = G
```
```
Output: the estimate
``` $\sigma_{.,j}$ ```and the set of genes G that were kept in the estimation```

```
*Note that here we consider the 2 biological replicates as drawn from a single normal
distribution, i.e. n = 2 again for d₂ Hartley constant but for a different reason than in the qPCR
data case above..
```

This estimation combined with our bootstrap procedures yielded CIs reported in Suppl. Table 2, and was the basis for marking connection coefficients as significantly different from zero (see Fig. 4A).

Screening for new genes associated with the ERα-NRIP1-LCoR network. The results above indicated that MRA could be applied to RNA-seq data. We therefore decided to exploit this ability by performing a new type of investigation: we used MRA to identify genes under tight control via the ERα-NRIP1-LCoR network. We reasoned that the MRA-inferred ERα-NRIP1-LCoR network represents a core transcriptional regulator and, in MELN cells, its activity is accurately reported by the ERE-Luc construct. Accordingly, any gene which would behave almost identically to ERE-Luc in terms of expression, would be a natural candidate for tight ERα-NRIP1-LCoR network transcriptional control. We decided to test proximity of genes with ERE-Luc by their ability to yield an MRA-inferred network from RNA-seq data, together with NRIP1 and LCoR, as close as possible to the qPCR data-derived ERα(ERE-Luc)-NRIP1-LCoR network.

We performed intersection of existing ChIP-seq data[33] with E2-regulated genes from our RNA-seq data (E2 *vs.* ethanol conditions, Suppl. Table 3). This allowed us to identify 884 genes both targeted by ERα and E2-regulated (Suppl. Table 4). Hence, we computed 884 networks with siNRIP1 and siLCoR RNA-seq data, successively replacing ERE-Luc by each of these 884 genes. That is, global response coefficients were obtained comparing NRIP1 and LCoR expression under E2 stimulation with the successive E2 + siLCoR and E2 + siNRIP1 perturbations.
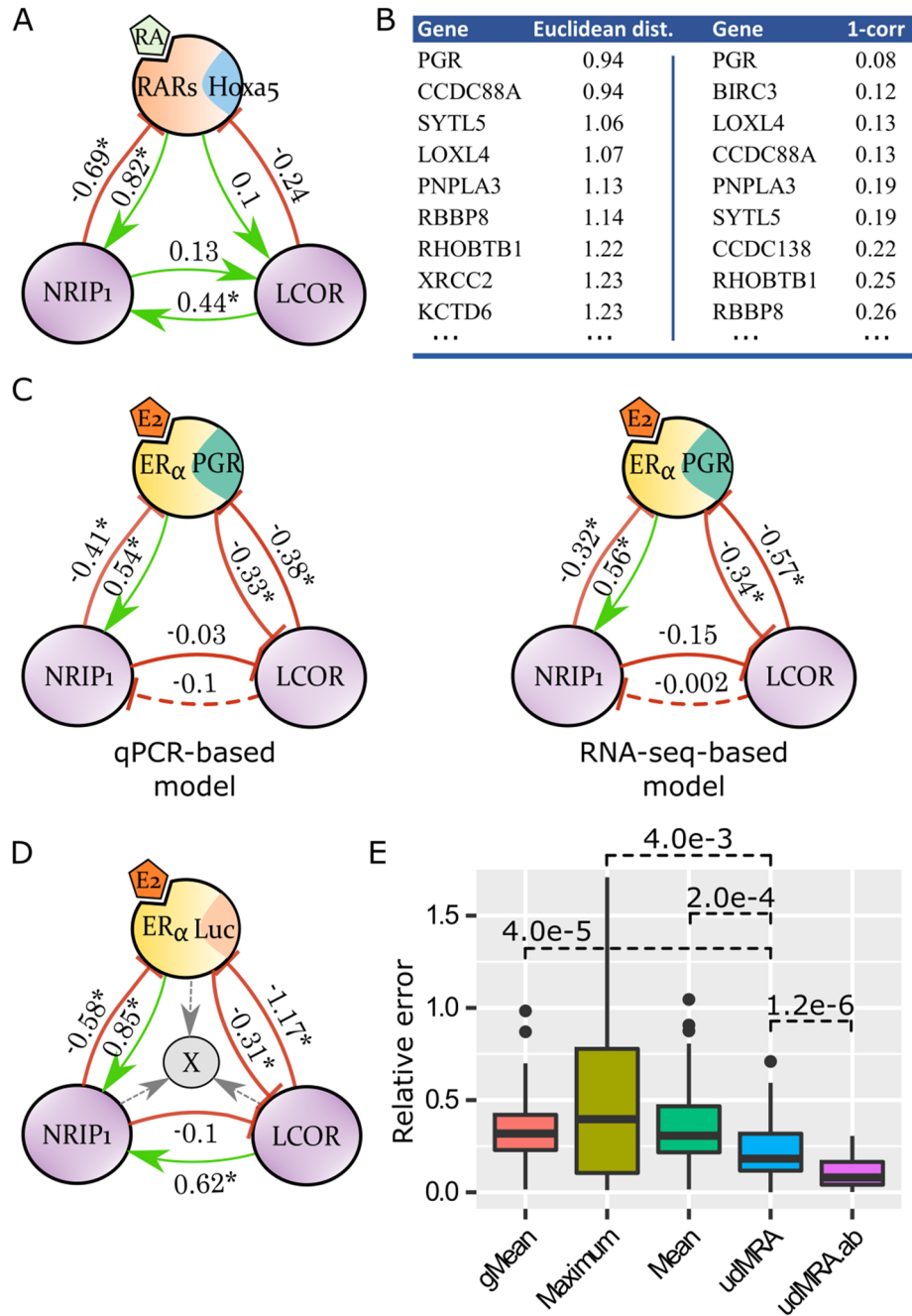
Figure 4. Application to RNA-seq data and genomic predictions. (**A**) RARs-NRIP1-LCoR network inferred from RNA-seq data. (**B**) Nine closest replacement genes for ERE-Luc in the ERα-NRIP1-LCoR network according to the Euclidean distance or 1-correlation. (**C**) ERα-NRIP1-LCoR networks with ERE-Luc replaced by PGR, trained from qPCR and RNA-seq data. (**D**) Principle of unidirectional MRA applications. (**E**) Accuracy of unidirectional MRA predictions (udMRA & udMRA.ab) under the E2 stimulation with double siNRIP1/siLCoR perturbation versus simple predictors (mean, geometric mean (gMean), and maximum of the two siRNAs). Wilcoxon test (n = 60). Figure created with Inkscape 0.92 (www.inkscape.org) and R 3.6 (r-project.org).

Perturbation on the candidate genes were obtained comparing their expression in the presence/absence of E2 stimulation. The genes with the shortest Euclidean distances or 1-correlation between their connection coefficients (the $r_{i,j}$ matrix) and those from the original qPCR-based connection coefficients (Fig. 2B) are listed in Fig. 4B. Given the progesterone receptor gene (PGR) ranked first with both measures, we performed a validation experiment by measuring *PGR* expression by qPCR. This enabled us to construct pure qPCR- and pure RNA-seq-based networks, as illustrated in Fig. 4C. These accurately reproduced the original network (Fig. 2B). CI estimations were obtained by applying the method introduced for the RARs-NRIP1-LCoR RNA-seq network; here the variance collected among the 884 E2-regulated/ERα-targeted genes was used.

Unidirectional MRA on genomic data. We next hypothesized that the ERα-NRIP1-LCoR MRA-derived network could provide means of predicting E2-regulated gene expression. To do this, we introduced a modified, hybrid version of the ERα-NRIP1-LCoR MRA network, including an additional module that cannot influence the other modules (Fig. 4D). The gray unidirectional arrows shown in Fig. 4D represent the connection coefficients between the NRIP1, LCoR, and ERα modules, and the added gene denoted by X. No connection coefficient from X to another module can be estimated due to the absence of perturbation data on X, i.e. we assume that X does not influence NRIP1, LCoR, or ERα transcriptional activity significantly. The connection coefficients from NRIP1, LCoR, or ERα towards X can be learned in the E2-stimulated basal condition by applying elementary perturbations, as described above. One separate network is constructed for each gene X considered. The mRNA abundance of gene X reports an $n + 1$th module activity and, by design, $r_{i,n+1} = 0, \forall i \in \{1, \ldots, n\}$, since no perturbation of X is available. From Eq. (7), we can compute $r_{n+1,j}, \forall j \in \{1, \ldots, n\}$, by solving the following system:

$$\left(\frac{\Delta x_{n+1}}{x_{n+1}}\right)_{q_k} = \sum_{j=1}^{n} r_{n+1,j}\left(\frac{\Delta x_j}{x_j}\right)_{q_k}, \quad k \in \{1, \ldots, n\}.$$

The performance of this type of MRA-inferred network that we call udMRA was assessed by its ability to predict the activity of the module $n + 1$ under the dual siNRIP1/siLCoR condition, as we did above for the conventional MRA-derived networks (Figs. 2D,H and 3C). To avoid trivially successful predictions on genes that would not vary, we limited the benchmark to the 884 E2-regulated/ERα-targeted genes that were significantly regulated upon siNRIP1 or siLCoR perturbation (under E2 stimulation); 60 genes fulfilled this condition (Suppl. Table 5). For comparison, the relative errors observed after applying udMRA or naïve predictions are depicted in Fig. 4E. udMRA yielded significantly better estimates of X expression, which is consistent with the complex dependency pattern between gene X expression and the modules NRIP1, LCoR, and ERα (Suppl. Fig. 2).

It is worth considering whether perturbation magnitudes during double siRNA interference on the same biological system remain identical; mathematically, this can be formulated as to whether filling the vector c in Eq. (11) with 1 values at the indices of the perturbed module, as conducted so far, is the best option. Equation (11) is written such that we can test different values. We examined the optimal set of coefficients a and b applied to the siNRIP1 and siLCoR perturbations (at the corresponding indices in vector c). This was based on the constraint that the prediction errors of luciferase, NRIP1, and LCoR expression (as in Fig. 2D) must be minimal; this yielded $a = 1$, and $b = 0.4$. Subsequently, these values were used in the udMRA network to more accurately predict the expression of the 60 benchmarked genes. Indeed, the relative error estimates confirmed that this modified inference procedure called udMRA.ab achieved greater accuracy (Fig. 4E).

aiMeRA usage. The R package was designed to be generally applicable; it relies on the formulae presented here, and is able to process any quantitative input, including biological and technical replicates. We included a functionality to facilitate the definition of MRA underlying network topologies (Fig. 5A). The design of a network can be specified through the execution of a few generic R functions and network plots can be generated within R directly (Fig. 5B). There is the possibility to export graphs in the graphML or gml formats for loading into Cytoscape[34] or yED graph editor. More details are provided in the package documentation and Supplementary Material. aiMeRA is available from GitHub and submission to Bioconductor is pending.

Discussion

MRA inference[1,2] is a widely used technique to uncover the directions and strengths of connections (connection coefficients or local response coefficients) between components of a biological system—the so-called modules—from systematic perturbation data (Fig. 1). We have developed an R package that implements classical MRA computations along with several extensions devoted to genomic data. The package was entirely implemented in R. The development of our MRA package came along with the opportunity to explore the crosstalk between ER and RAR, two important NRs involved in several tumors, such as BC. We generated unique qPCR and RNA-seq data using BC MELN cells, which are derived from MCF7 cells (see Methods) and a well-established BC model.

The transcriptional activity of ER and RAR is modulated by the shared corepressors NRIP1 and LCoR. Considering networks of growing complexity, *i.e.*, ER-NRIP1-LCoR, RAR-NRIP1-LCoR, and ER-RAR-NRIP1-LCoR, we showed that our MRA inference R package could recapitulate well-known interactions between these molecules by exploiting qPCR data (Figs. 2, 3). We successfully confirmed the known predictive power of MRA-inferred networks as a means to assess their validity in the validation experiments presented. Using original experimental data, the most complex network, ER-RAR-NRIP1-LCoR, enabled us to confront the hypothesis we stated in the Introduction section. Specifically, ER-RAR crosstalk has been described in previous studies as having potential mutual repressive consequences[11–15]. A spatial competition mechanism to bind DNA was proposed in one instance[14]. We reasoned that the sharing of corepressors (NRIP1 and LCoR) may contribute to cross-suppression. Indeed, NRIP1 expression is induced by both ER and RAR upon ligand binding, suggesting that NRIP1 induction by one NR could render NRIP1 available to dump the transcriptional activity of the other NR. The ER-RAR-NRIP1-LCoR network represented in Fig. 3A illustrates both the direct anti-estrogenic activity of RAR, which is in line with current literature, and indirect anti-estrogenic activity mediated by NRIP1, thus substantiating our hypothesis. Reciprocally, ER activity displayed a limited direct inhibition of RAR signaling, but a strong indirect repression via NRIP1 through the same mechanism. In addition, the inhibition of LCoR, which itself inhibits NRIP1, may amplify this phenomenon. This further supports our hypothesis. It is worth noting from Fig. 5 how easy the execution is of such computations in the aiMeRA package, even by a non-specialist.

A

```
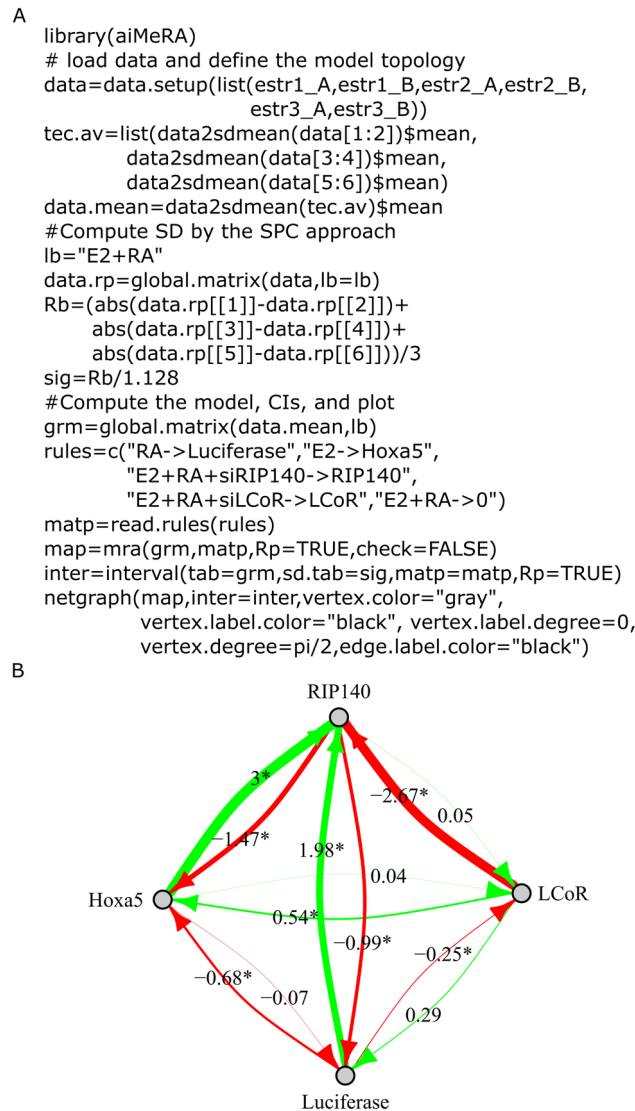library(aiMeRA)
# load data and define the model topology
data=data.setup(list(estr1_A,estr1_B,estr2_A,estr2_B,
                     estr3_A,estr3_B))
tec.av=list(data2sdmean(data[1:2])$mean,
            data2sdmean(data[3:4])$mean,
            data2sdmean(data[5:6])$mean)
data.mean=data2sdmean(tec.av)$mean
#Compute SD by the SPC approach
lb="E2+RA"
data.rp=global.matrix(data,lb=lb)
Rb=(abs(data.rp[[1]]-data.rp[[2]])+
    abs(data.rp[[3]]-data.rp[[4]])+
    abs(data.rp[[5]]-data.rp[[6]]))/3
sig=Rb/1.128
#Compute the model, CIs, and plot
grm=global.matrix(data.mean,lb)
rules=c("RA->Luciferase","E2->Hoxa5",
        "E2+RA+siRIP140->RIP140",
        "E2+RA+siLCoR->LCoR","E2+RA->0")
matp=read.rules(rules)
map=mra(grm,matp,Rp=TRUE,check=FALSE)
inter=interval(tab=grm,sd.tab=sig,matp=matp,Rp=TRUE)
netgraph(map,inter=inter,vertex.color="gray",
         vertex.label.color="black", vertex.label.degree=0,
         vertex.degree=pi/2,edge.label.color="black")
```

B

Figure 5. The aiMeRA R package. (**A**) Example R code for loading data, data preparation, and inference of a network. Note that NRIP1 was named by its common alternative name RIP140. Basal condition is E2 & RA stimulation (denoted "E2 + RA- > 0") and LCoR perturbation is defined as "E2 + RA + siLCoR- > LCoR". Same logic for RIP140 (= NRIP1). Perturbation on the HOXA5 module reporting RARs activity is defined as E2, *i.e.*, loss of RA stimulation compared to the basal condition was E2 & RA. Perturbation on the luciferase module is similarly defined as RA, *i.e.*, loss of E2. (**B**) Direct plot of an MRA-inferred network in R using the igraph library. Figure created with Inkscape 0.92 (www.inkscape.org) and R 3.6 (r-project.org).

Following this classical application of MRA, we showed that almost identical MRA-derived networks could be trained from qPCR and RNA-seq data obtained in the same conditions. By employing concepts of SPC theory, we introduced a new procedure that estimates data variance over a gene population when only a few experimental replicates are available. Such a method could be applied with proteomics or phospho-proteomics data when restricted to limited replicates, which is often the case.

Access to genomic data related to the ER-RAR crosstalk motivated an extension of MRA. Since ER acts as a transcription factor upon ligand binding, MRA enabled us to infer a core transcriptional regulatory network around it, *i.e.*, the ER-NRIP1-LCoR network. We thus explored the possibility to identify novel genes under tight control by this network by checking one by one their ability to replace the MELN cell ERE-Luc construct. This construct reports direct ER transcriptional activity. We constructed one MRA-inferred network *per* candidate gene and computed its similarity with the ERE-Luc-based reference network (Fig. 4B). The top candidate was the progesterone receptor gene (*PGR*) that is a widely used reporter of estrogen activity in BC in the clinic. A qPCR-based validation experiment confirmed our prediction (Fig. 4C). This observation supports the potential value of this use of MRA inference to explore unknown associations with a biological network of interest. It also suggests that *PGR* could be employed as surrogate direct reporter of ER transcriptional activity in cell lines devoid of specific constructs, such as ERE-Luc.

We further investigated the possibility of inferring the expression levels of 60 genes that were both E2-regulated and targets of ERα in published ChIP-seq data. This required the construction of hybrid networks, including unidirectional connections to add modules that were not perturbed in the training data set (Fig. 4D). We showed that such hybrid ERα-NRIP1-LCoR networks could outperform naïve predictors (Fig. 4E); this finding was not surprising because the dependencies between the 60 genes tested and the modules of the MRA-inferred network were complex (Suppl. Fig. 2). The accuracy of the predictions could be even further extended by introducing weights in MRA inferences. Other biological systems may be amenable to such modified inferences in the absence of strong feedback loops originating from the inferred genes. Weighted inferences also apply to standard MRA (see our mathematical derivation). Lastly, the inclusion in MRA inferences of molecules that are not perturbed directly and independently has been recently discussed in the special case where conservation laws can be invoked[19].

The aiMeRA package implements all the extensions discussed in this report.

References

1. Kholodenko, B. N. *et al.* Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12841–12846 (2002).
2. Santra, T., Rukhlenko, O., Zhernovkov, V. & Kholodenko, B. N. Reconstructing static and dynamic models of signaling pathways using modular response analysis. *Curr. Opin. Syst. Biol.* **9**, 11–21 (2018).
3. R Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, 2019).
4. Heldring, N. *et al.* Estrogen receptors: how do they signal and what are their targets. *Physiol. Rev.* **87**, 905–931 (2007).
5. Cavaillès, V. *et al.* Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. *EMBO J.* **14**, 3741–3751 (1995).
6. Docquier, A. *et al.* Negative regulation of estrogen signaling by ERβ and RIP140 in ovarian cancer cells. *Mol. Endocrinol.* **27**, 1429–1441 (2013).
7. Sixou, S. *et al.* Importance of RIP140 and LCoR Sub-Cellular localization for their association with breast cancer aggressiveness and patient survival. *Transl. Oncol.* **11**, 1090–1096 (2018).
8. Fernandes, I. *et al.* Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and -independent mechanisms. *Mol. Cell* **11**, 139–150 (2003).
9. Jalaguier, S. *et al.* Complex regulation of LCoR signaling in breast cancer cells. *Oncogene* **36**, 4790–4801 (2017).
10. White, K. A. *et al.* Negative feedback at the level of nuclear receptor coregulation self-limitation of retinoid signaling by RIP140. *J. Biol. Chem.* **278**, 43889–43892 (2003).
11. White, K. A., Yore, M. M., Deng, D. & Spinella, M. J. Limiting effects of RIP140 in estrogen signaling: potential mediation of anti-estrogenic effects of retinoic acid. *J. Biol. Chem.* **280**, 7829–7835 (2005).
12. Rousseau, C. *et al.* The N-terminal of the estrogen receptor (ERalpha) mediates transcriptional cross-talk with the retinoic acid receptor in human breast cancer cells. *J. Steroid Biochem. Mol. Biol.* **86**, 1–14 (2003).
13. Laganière, J., Deblois, G. & Giguère, V. Functional genomics identifies a mechanism for estrogen activation of the retinoic acid receptor alpha1 gene in breast cancer cells. *Mol. Endocrinol.* **19**, 1584–1592 (2005).
14. Hua, S., Kittler, R. & White, K. P. Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell* **137**, 1259–1271 (2009).
15. Ross-Innes, C. S. *et al.* Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. *Genes Dev.* **24**, 171–182 (2010).
16. Klinger, B. *et al.* Network quantification of EGFR signaling unveils potential for targeted combination therapy. *Mol. Syst. Biol.* **9**, 673 (2013).
17. Santra, T., Kolch, W. & Kholodenko, B. N. Integrating Bayesian variable selection with modular response analysis to infer biochemical network topology. *BMC Syst. Biol.* **7**, 57 (2013).
18. Bosdriesz, E. *et al.* Comparative network reconstruction using mixed integer programming. *Bioinformatics* **34**, i997–i1004 (2018).
19. Lill, D. *et al.* Mapping connections in signaling networks with ambiguous modularity. *NPJ Syst. Biol. Appl.* **5**, 1–11 (2019).
20. Stelniec-Klotz, I. *et al.* Reverse engineering a hierarchical regulatory network downstream of oncogenic KRAS. *Mol. Syst. Biol.* **8**, 601 (2012).
21. Dorel, M. *et al.* Modelling signalling networks from perturbation data. *Bioinformatics* **34**, 4079–4086 (2018).
22. Davison, A. C. & Hinkley, D. V. *Bootstrap Methods and their Applications.* (Cambridge University Press, 1997).
23. Santos, S. D. M., Verveer, P. J. & Bastiaens, P. I. H. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat. Cell Biol.* **9**, 324–330 (2007).
24. Andrec, M., Kholodenko, B. N., Levy, R. M. & Sontag, E. Inference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy. *J. Theor. Biol.* **232**, 427–441 (2005).
25. Oakland, J. S. *Statistical Process Control.* (Routledge, 2012).
26. Wheeler, D. J. & Chambers, D. S. *Understanding Statistical Process Control.* (SPC Press, Inc., 1992).
27. Thomaseth, C. *et al.* Impact of measurement noise, experimental design, and estimation methods on modular response analysis based network reconstruction. *Sci. Rep.* **8**, 16217 (2018).
28. Harter, H. L. Tables of range and studentized range. *Ann. Math. Stat.* **31**, 1122–1147 (1960).
29. Balaguer, P., Boussioux, A. M., Demirpence, E. & Nicolas, J. C. Reporter cell lines are useful tools for monitoring biological activity of nuclear receptor ligands. *Luminescence* **16**, 153–158 (2001).
30. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
31. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
32. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
33. Madak-Erdogan, Z. *et al.* Integrative genomics of gene and metabolic regulation by estrogen receptors α and β, and their coregulators. *Mol. Syst. Biol.* **9**, 676 (2013).
34. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).

Acknowledgements

Author contributions

P.R., V.C., and J.C. conceived the idea and supervised the study. GJD designed and implemented the R package, and performed data analysis. S.J. designed and realized the experimental work. P.R., G.J.D., S.J., V.C., and J.C. analyzed the results. J.C. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-86544-0.

Correspondence and requests for materials should be addressed to V.C. or J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.