Appendix

Prostate cancer-specific gene interaction and MRA. Out of the 1,547 human TFs in AnimalTFDB 2.0, overlapping TFs with the 5,000 genes of high variance in the combined data between GSE67157 and GSE109505 were selected. A consensus gene network was generated from 100 times ARACNe bootstrapping. To run ARACNe bootstrapping, ARACNe C^{++} source file was downloaded from ARACNe homepage (http://califano.c2b2.columbia.edu/aracne/). From the prostate cancer-specific transcriptional interactome, the CCI signature became the candidate of regulons, which are regulated by MRs and FET inferred the master regulators.

Feature selection and identification of DEGs. In the GSE109505 data, SAM (1), a supervised feature selection method was performed to identify DEGs out of the 5,000 genes between CsA- and vehicle-treated PC-3 cells. Likewise, in the microarray data from patients with prostate cancer (GSE3325), out of 11,394 protein coding genes, the same 5,000 genes as GSE67157 were extracted and SAM performed to identify DEGs. 'Two class unpaired' in GSE109505 data and 'Multiclass' in GSE3325 data were used as a statistical problem type in SAM. A tuning parameter, delta of 0.4, optimized the cutoff for significance with the estimation of FDR *q*-value threshold 0.01. To find the genes that show reverse-correlation between CsA-treated PC-3 cells and patients with prostate cancer, the common DEGs from the SAM results of GSE109505 and GSE3325 datasets were summarized. The present study identified genes which are upregulated in patients with metastatic cancer against patients with benign and primary prostate cancer in GSE3325, but downregulated in CsA-treated PC-3 cells against vehicle-treated PC-3 cells in GSE109505. Likewise, the present study identified genes which are downregulated in patients with metastatic cancer against patients with benign and primary prostate cancer in GSE3325, but upregulated in CsA-treated PC-3 cells in GSE109505. These reverse-correlated genes were termed 'Clinically significant CsA-Induced gene expression (CCI) signature' and applied to further analyses. To obtain biological interpretation, GSEA for Hallmark (version 6.1 at MSigDB, http://software.broadinstitute.org/gsea) (2) was performed on GSE109505 data with the CCI signature.

FGES to discover E2F8 interacting genes. Bayesian networks (BNs) are an effective architecture for modeling causal relationships from observational biological and gene expression data (3,4). To construct BNs for prostate cancer gene expression profile data and identify E2F8-interacting genes, TCGA-PRAD gene expression data was downloaded from Xena Functional Genomics Explorer (https://tcga.xena-hubs.net/download/TCGA.PRAD.sampleMap/HiSeqV2.gz). From the TCGA-PRAD gene expression profile data, the E2F8 gene signature was extracted, which had been inferred in MRA-FET analysis and discretized the continuous value of expression levels in each gene to equal-width bins (8 bins) using the unsupervised discretization method available in the Infotheo R package (5). The FGES-discrete algorithm (http://www.ccd.pitt.edu/data-science/software/) (6) was ran on the TCGA-PRAD data to discover genes directly interacting to CCI signature master regulator. FGES-discrete algorithm uses the BDeu scoring to calculate the probability of data given model (6) and searches over possible models. The algorithm needs two basic parameters; 'structure-prior' and 'sample-prior'. 'structure-prior' provides prior probability for each node given a set of parents. 'sample-prior' indicates how confident that the prior probabilities in the BN are uniform for BDeu scoring. The default 'structure-prior=1' and 'sample-prior=1' were used. Prior 'knowledge' that indicates what is target node and what are predictor nodes in the discrete-BN was specified. This enables the algorithm to learn the BN faster and search for direct predictor nodes. CsA, cyclosporin A; SCAN, single channel array normalization; SAM, significance analysis of microarrays; DEGs, differentially expressed genes; FDR, false discovery rate; TFs, transcription factors; MRA, master regulator analysis; ARACNe, algorithm for the reconstruction of gene regulatory networks; MI, mutual information; FET, Fisher's exact test; MRs, master regulators; FGES, fast greedy equivalence search; BNs, Bayesian networks; TCGA, The Cancer Genome Atlas; PRAD, Prostate Adenocarcinoma; BDeu, Bayesian Dirichlet equivalent uniform.

References

1. Tusher VG, Tibshirani R and Chu G: Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98: 5116-21, 2001.
2. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, *et al*: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102: 15545-50, 2005.
3. Yu J, Smith VA, Wang PP, Hartemink AJ and Jarvis ED: Advances to Bayesian network inference for generating causal networks from observational biological data. Bioinformatics 20: 3594-603, 2004.
4. Wang YX and Huang H: Review on statistical methods for gene network reconstruction using expression data. J Theor Biol 362: 53-61, 2014.
5. Meyer PE, Kontos K, Lafitte F and Bontempi G: Information-theoretic inference of large transcriptional regulatory networks. EURASIP J Bioinform Syst Biol 79879, 2007.
6. Ramsey J, Glymour M, Sanchez-Romero R and Glymour C: A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. Int J Data Sci Anal 3: 121-129, 2017.