Transcriptional determinants of ERRα activities during cancer progression

Jingru Shi

Numéro National de Thèse :

THESE DE DOCTORAT DE L'UNIVERSITE DE LYON

Ecole Normale Supérieure de Lyon
en co-tutelle avec
East China Normal University
Ecole Doctorale N° 340 Biologie Moléculaire Intégrative et Cellulaire
Discipline : Biologie
Soutenue publiquement le 17/12/2024, par :

Madame Jingru SHI

Transcriptional determinants of ERRα activities during cancer progression

Déterminants transcriptionnels des activités de ERRα au cours de la progression cancéreuse

---

devant le jury composé de :

Claude BEAUDOIN, Pr. Univ. Clermont Ferrand	Rapporteur
Qi LIU, Pr. Univ. Tongji, Shanghai	Rapporteur
Jean-Marc VANACKER, DR CNRS, ENS de Lyon/CRCL	Directeur de thèse
Tieliu SHI, Pr. ECNU, Shanghai	Co-tuteur de thèse
Caroline MOYRET-LALLE, Pr. Univ. Lyon	Examinatrice
Julie CARAMEL, CR Inserm	Examinatrice

Acknowledgements

First of all, I want to thank all reviewers, supervisors and jury members. Thank you for taking the time to evaluate and give advice to my thesis. I will carefully accept your suggestions on my thesis.

When you read this, it means I have spent three meaningful years alone in France. I never thought I could live and study alone abroad for such a long time, and I want to thank all the people who have appeared in my world and helped me.

The first person to thank is my supervisor, Jean-Marc. It's difficult to describe in words all the help he has provided me throughout my doctoral studies. When the IGFL laboratory in ENS de Lyon closed, he didn't give up on me and persisted in seeking for a new laboratory that would be helpful for my research. When discussing my project, it is often difficult for me to understand some analysis results due to a lack of biological knowledge, but he always explained biological knowledge to me and helped me quickly understand. At the same time, he also guided me how to think about the next research direction. When I encountered a bottleneck in the project, he worked with me to find a solution, and even contacted relevant experts in the field. In addition, for the predicted results of my project, he performed extensive experimental verification. Without his help, I would not have been able to quickly finish my doctoral thesis within 3 years. After the completion of the journal paper, we planned to give a speech at the CRCL research institute and even the international conference. In order to present a good speech, he helped me to organize the content of the speech, told me how to add a short and concise title and summary to each slide, how to display results with clear and simple images, how to mark the key point of each slide, and even reminded me to add animation effects in appropriate places. Not only that, he also suggested that I should practice my speech a lot before the formal presentation. So, in the meeting room, I could always be seen practicing my speech in front of the big screen, and he sat listening carefully and taking notes on my problems. With his help, I have gained many skills in public presentation, which are my precious treasures that I will keep in mind in the future. How time flies, I am approaching the end of my doctoral learning. During the writing of my thesis, he helped me to build the framework, to organize the logical issues of the content and carefully reviewed every part. His seriousness and responsibility impressed me greatly. During this period, I was often depressed when modifying and correcting my thesis. He always kept encouraging me, saying, "This is what every PhD student goes through. You wrote great. We'll finish it soon!" With his constantly encouragement and guidance, I was able to persist in finishing my doctoral thesis and present it to all reviewers and researchers who is reading it. In addition to studying, he has also provided me with a lot of life assistance. When I rent a house, he said, "If you have any difficulties, I can provide financial assistance for you." He also introduced the beautiful views to me and gave advice for my vacation. I am very lucky to have met such a responsible and knowledgeable supervisor during my doctoral studying. I benefited a lot from him, and I wish him all the best in the future.

I also want to thank my Chinese supervisor in ECNU, Professor Tieliu SHI. He is a serious scholar. In each scientific discussion, he was always able to point out the potential problems and early mistakes of the project with his extensive research background. He always answered my questions at any time even though he was in the non-work time. I am impressed by his spirit in science. Thanks to his years of guidance. In my future work and life, I will keep his advice in mind and wish him all the best in the future.

The first person I met in France was Catherine. She is my teacher who guides me in the field of bioinformatics. I still remember the first day when I arrived in France, she picked me up at the airport. For the whole week, she helped me with my accommodation, the bank card, and school registration procedures. The day I arrived was close to Christmas. In order to help me quickly understand the project, she still took a lot of time to answer my questions. Although she had retired later, whenever I had problems of bioinformatics, she was willing to discuss with me and help to find a solution. I will remember all the help she has given me, and I hope her retirement life colorful.

I would also thank Christelle and Violaine in IGFL. Especially Christelle, she often arranged every "eating together" lunch for us. When I left the IGFL lab, I remember it was the end of 2022, she came in that day and asked me if I was ready to go to the new lab. Before leaving, she left a card with a wish to me for Merry Christmas. I will always remember that time and wish her everything is well.

I would like to thank the two leaders, Muriel and Olivier, of the new CRCL lab. I am appreciated for their opinions and encouragement on my project in each lab meeting. They always expressed their affirmation on my results which motivates me a lot. Thanks to Coralie who validated the partial results of my project. Thanks to Marine, Zélie, Ludivine and Julien. Thanks to the "Bonjour Jingru" every morning from Marine and Zélie. In particularly, I remember the day my first seminar presentation at CRCL, Zélie smiled and encouraged me. She told me, "Jingru, you did a great job and made huge progress compared to the first presentation!" I remember the overwhelming excitement I felt at that time. They are all so warm. Thanks to everyone at CRCL lab. I hope them everything goes well in the future.

Seven years ago, I successfully became a member of "PROGRAM OF SINO-FRENCH EDUCATION FOR RESEARCH (PROSFER)" program, which links me with France and doctoral studying. Therefore, I want to thank East China Normal University, Ecole Normale Supérieure de Lyon and all members of the program. I am especially grateful to Xiaoling LIU and Xiaoyan LIU, they have provided me with a lot of help and care throughout my master's and doctoral studies, and have accompanied me for 7 years of growth. My thanks for PROSFER are not only to all the organizers, but also to my classmates and friends who also belong to PROSFER. Thanks to Shijia WU, Wenqi XU, Wenqing JIAO, Yingquan SHAN, Wan XIANG..., we often cooked at home by ourselves and shared delicious food with each other in every Chinese festival. On countless nights of talking, we encouraged each other to persist. I also want to thank

my classmate in the ECNU lab, Jiajia ZHOU. She is an outgoing girl who has been helping me with various ECNU administrative procedures. Wishing everyone a bright future.

In addition to my classmates, I also want to thank my good friend Yumeng XIA who is far away in China. When I first arrived in France, I felt helpless and afraid, and she accompanied me all the time. Countless lonely or interesting days, we shared and grew together. For over ten years, including undergraduate, master's, and doctoral studies, we have supported, encouraged and comforted each other in our daily lives. We witness each other's growth and maturity. Wishing you the best and happiness!

Most importantly, I want to thank my parents. The road of studying alone is difficult, but it is my parents' phone care every day and night that drives me to keep moving forward. My parents not only guide me in the direction of my life and educate me, but also give me numerous care and love. Only because I said "I miss Chinese snacks", they immediately sent me a big package of snacks from China. My parents love and protect me like the strongest armor in my life. I love them so much! Not only that, I also want to thank my relatives, my cousins, and most importantly, my boyfriend Yang LUO. Due to the time difference, he stayed up late every day and waited for me to return home, because he wanted to accompany me for a while and help me to relieve the pressure of study and the loneliness of life. He also taught me to be happy and optimistic every day, to always treat everything with a positive attitude, not to be anxious, and to be calm always. He is a person as warm as the sun, and I want to be like him.

At last, I want to thank the China Scholarship Council (CSC) for providing financial support to my doctoral study in France.

After 23 years of studying, life is going to enter into a new stage. During the period of my doctoral studying, I seemed to constantly doubt, accept, and re-doubt myself. Similarly, my studying environment was also always changing, from ECNU to IGFL, and then to CRCL. It is a hard time. However, I gradually realized that these changes have made me more mature and calmer. Life is constantly changing, and we must face the fast-changing world at any time. At this moment, our greatest weapon is courage. The experience during my doctoral studying may have been my first weapon and the source of my courage. Therefore, I would like to thank myself for making the decision to start my PhD studying four years ago, for persisting all the way and for still being warm and bright. I am not perfect, but at least I have been running towards it.

If there's anything else I want to say, I hope the world can always be beautiful, and I hope all of us can always smile and get better!

Contents

Abstract **8**

List of Figures **11**

Abbreviations **16**

Introduction **23**

Chapter I - Breast cancer 23
1.1 General introduction 23
1.2 Breast cancer subtype classification 23
1.3 Breast tumor heterogeneity 24
1.4 Tumor microenvironment 25
1.5 Significance of tumor heterogeneity 26
1.6 Breast cancer treatment 26
1.7 Conclusion 27

Chapter II – Nuclear receptors and co-regulators 28
2.1 General introduction of nuclear receptors 28
2.2 Basic structure of nuclear receptors 29
2.3 Molecular functions of nuclear receptors 29
2.4 Relations between nuclear receptors and disease 30
2.5 Conclusion 31

Chapter III – ERRα 32
3.1 Basic structure of estrogen related receptors 32
3.2 Relation between ERRα and ERα 33
3.3 Biological functions of ERRα 35
3.4 Molecular mechanisms of ERRα actions 38
3.5 Conclusion 45

Chapter IV– ZEB1 46

Chapter V – Co-regulators prediction in cancer 47
5.1 Introduction of co-regulators prediction methods 47
5.2 PLS method introduction 52
5.3 Sparse PLS method introduction 53
5.4 Adaptive sparse PLS method introduction 54
5.5. Conclusion 55

Aim of my thesis **57**

General introduction of thesis **58**

Results **61**

Section I. Genes regulated by ERRα in cell migration pathway 61

Section II. ERRα co-activators prediction and common targets exploration 64
2.1 Determination of pre-selected TRs 68
2.2 ERRα's co-activators prediction 69
2.3 Identification of DEGs related to both ERRα and ZEB1 70
Section III. Stability evaluation of dataset size in sPLS modeling 72
Section IV. Transcriptional relationships between ERRα, ZEB1 and their targets in breast cancers. 78
4.1 Transcription activity in breast cancer single cells 79
4.2 Transcription activity in breast cancer cell lines 80
4.3 Transcription activity in breast tumors 81
Section V. Experimental validation of the transcriptional regulation exerted by ERRα and ZEB1 ... 84
5.1 Transcriptional regulation of ERRα and ZEB1 on 8 DEGs in TNBC cells 84
5.2 Transcriptional regulation of ERRα and ZEB1 on 8 DEGs in non-TNBC cells 87
5.3 Specificity of the 8 DEGs 88
Section VI. Co-operation of ERRα and ZEB1 on expression of the 8 DEGs 92
6.1 ZEB1 ChIP-Seq results 92
6.2 Cooperation mechanism through the binding of ERRα and ZEB1 on 8 DEGs 94
Section VII. Expression of 8 DEGs in tumor microenvironment 98
7.1 Expression of genes in the microenvironment of TNBC tumors 99
7.2 Expression of genes in the microenvironment of ER+ and Her2+ tumors 101
Section VIII. EMT in breast cancer 104
8.1 Correlation of the expression of the indicated genes and EMT status. 105
8.2 Comparison of EMT status between breast cancer subtypes 110
Section IX. Survival in breast cancer 112
9.1 Influence of ERRα, ZEB1 and the 8 DEGs on overall survival 113
9.2 Influence of ERRα, ZEB1 and the 8 DEGs on the relapse free survival and distant metastasis free survival 115
Discussion and perspectives 117
Relationship of ERRα and ZEB1 in TNBC 117
Using public data to explore the potential of other TRs to be ERRα co-regulators in breast cancer 118
Features of the TNBC MES subtype 120
Characteristic of the 8 DEGs in breast cancer 121
Characteristic of CAFs, endothelial cells and pericytes in breast cancer 126
Prognostic function of ERRα and ZEB1 in breast cancer 127
What's more? 130
General conclusion 132
References 133

ANNEXE 1 152

Publication list 152

ANNEXE 2 153

Articles published 153

Abstract

Estrogen-related receptor alpha (ERRα), an orphan nuclear receptor, participates in metabolism, cell migration, cell invasion, metastasis and progression of breast cancer in a coregulator-dependent manner. My thesis aims to explore the transcription activity of ERRα in breast cancer progression as well as the potential pathological and physiological pathways it may be involved in. To achieve this:

(1) We firstly predicted co-regulators for ERRα, using a mathematical algorithm, namely adaptive sparse partial least squares (sPLS) regression algorithm. Our results showed that ZEB1 is the most robust potential co-activator of ERRα. We validated that these two factors co-regulate the expression of eight migration-related targets (8 DEGs), specifically in triple negative breast cancer (TNBC) cells.

(2) We then investigated the mechanisms through which these two factors co-regulate gene expression. Our results showed that regulation of the expression of the 8 DEGs by ERRα depends on ZEB1, but not on other transcriptional regulators. ChIP analysis showed that ERRα directly (i.e., in a ZEB1-independent manner) binds to the promoters of the DEGs, whereas ZEB1 requires ERRα to bind to the same promoter elements. This suggests a physical interaction between these factors, that was demonstrated by proximity ligation assays.

(3) We next explored the pathological consequences of these interactions in breast tumors. We observed a correlation between the state of epithelial to mesenchymal transition (EMT) of the tumors and expression of ZEB1 and the 8 DEGs. Since EMT is correlated with breast cancer metastasis, we next investigated the prognosis ability of the 8 DEGS. Our results show that a high joint expression of the 8 DEGs predicts overall survival in TNBC patients.

In conclusion, we identified and experimentally validated a novel co-activator (ZEB1) of ERRα involved in breast cancer progression. These two factors act together on the transcription regulation of genes that are highly involved in cancer metastasis and their expression predicts the clinical outcome of TNBC patients.

Key words: transcription, breast cancer, co-regulators, ERRα, ZEB1

Résumé

Le récepteur Estrogen-Related Receptor (ERRα), un récepteur nucléaire orphelin intervient dans le métabolisme, la migration et l'invasion cellulaire, l'établissement de métastases et la progression des cancers du sein d'une manière dépendante de corégulateurs. Ma thèse vise à explorer l'activité transcriptionnelle de ERRα au cours de la progression cancéreuse ainsi que les voies pathologiques et physiologiques mises en œuvre. A ces fins:

(1) Nous avons prédit les corégulateurs de ERRα en utilisant un algorithme mathématique, l'algorithme de régression adaptive sparse partial least squares (sPLS). Nos résultats montrent que ZEB1 est le plus robuste coactivateur potentiel de ERRα. Nous avons validé que ces deux facteurs corégulent l'expression de huit cibles transcriptionnelles (differentially expressed genes; DEGs), spécifiquement dans les cellules de cancer du sein triple négatives (TNBC).

(2) Nous avons ensuite analysé les mécanismes par lesquels ces deux facteurs corégulent l'expression génique. Nos résultats montrent que la régulation de l'expression des 8 DEGs par ERRα dépend de ZEB1, mais pas d'autres corégulateurs transcriptionnels. No analyses par ChIP montrent que ERRα se fixe directement (i.e., indépendamment de ZEB1) sur les promoteurs des DEGs, alors que ZEB1 requiert la présence de ERRα pour se fixer sur ces mêmes éléments promoteurs. Ceci suggère une interaction physique entre les deux facteurs, que nous avons confirmé par proximity ligation assay.

(3) Nous avons ensuite exploré les conséquences pathologiques de ces interactions dans les tumeurs du sein. Nous avons observé une corrélation entre l'état de transition épithélium-mésenchyme (EMT) des tumeurs et l'expression de ZEB1 et des 8 DEGs. Comme l'EMT est corrélée aux métastases des cancers du sein, nous avons ensuite analysé la capacité pronostique de l'expression des 8 DEGs. Nos résultats montrent que la forte expression des 8 DEGs prédit la survie des patients TNBC.

En conclusion, nous avons identifié et validé expérimentalement un nouveau co-activateur (ZEB1) de ERRα impliqué dans la progression des cancers du sein. Ces deux facteurs agissent ensemble sur la régulation transcriptionnelle de gènes impliqués dans l'établissement de métastases et dont l'expression prédit le devenir clinique des patients TNBC.

Mot Clés: transcription, cancer du sein, corégulateurs, ERRα, ZEB1

摘要

雌激素相关受体 alpha（estrogen-related receptor α; ERRα）是一种孤儿核受体，它以依赖共调节因子（co-regulators; CoRs）的方式参与乳腺癌的代谢、细胞迁移、细胞侵袭、转移和癌症进展等疾病过程。本论文旨在探讨 ERRα 在乳腺癌进展中的转录调控功能及其可能参与的潜在病理和生理途径。本论文主要包括以下内容:

（1）首先利用自适应稀疏偏最小二乘（the adaptive sparse partial least squares; adaptive sPLS）回归算法预测 ERRα 的共调节因子 CoRs。我们的结果表明 ZEB1 是最显著且稳定的潜在 ERRα 共激活因子（co-activator）。不仅如此，实验还证明，尤其是在三阴性乳腺癌（triple negative breast cancer; TNBC）细胞中，这两个因子共同调节 8 个与癌细胞迁移相关的靶基因（8 DEGs）的表达。

（2）接下来我们研究了 ERRα 和 ZEB1 共同调控基因表达的分子机制。我们的研究结果表明，ERRα 对 8 DEGs 转录表达的调控需要 ZEB1 的参与，而不是依赖于其他的转录共调控因子。ChIP 实验分析表明，ERRα 能够以与不依赖 ZEB1 的方式直接结合在 8 DEGs 的启动子上，而 ZEB1 则需要 ERRα 结合到相同的启动子元件上。此外，通过 Proximity ligation assays（PLA）实验，我们还证明 ERRα 和 ZEB1 之间存在直接相互作用。

（3）接下来，我们探讨了这些相互作用在乳腺肿瘤中的病理影响。我们观察到肿瘤上皮向间质转化（epithelial to mesenchymal transition; EMT）状态与 ZEB1 和 8 DEGs 的表达之间存在相关性。由于 EMT 与乳腺癌转移密切相关，我们接下来探索了 8 DEGs 对患者生存的预后能力。我们的研究结果表明，8 个 DEGs 的联合高表达可以预测 TNBC 患者的总生存情况（overall survival; OS）。

总之，我们为 ERRα 预测并实验证实了一个新共激活因子 ZEB1，他们共同参与并影响乳腺癌的进展。此外，这两个因子还共同调控与乳腺癌转移高度关联的基因的转录水平，我们的结果表明这些下游靶基因的联合高表达可预测 TNBC 患者的生存情况。

关键词: 转录，乳腺癌，共调节因子，ERRα，ZEB1

List of Figures

Figure 1. Introduction of definition, pathogenesis, case percentage, prognosis and treatment for the three breast cancer subtypes from (Waks and Winer, 2019).

Figure 2. Pie chart from (Nolan et al., 2023) presenting the percentage of each subtype of breast cancers characterized by molecular profiling.

Figure 3. Distribution of stromal cells within tumors from (Anderson and Simon, 2020).

Figure 4. Introduction of 48 human nuclear receptors (NRs) from (Tao et al., 2020).

Figure 5. Structure of nuclear receptors (NRs).

Figure 6. Mechanism of regulation of gene expression from (Badia-I-Mompel et al., 2023).

Figure 7. Expression of ESRRA (encoding ERRα) in human normal tissues from (GTEx Consortium, 2020).

Figure 8. Structure of the ERR receptors from (Tang et al., 2021).

Figure 9. Structure of protein sequence of ERRα and ERα from (Tang et al., 2021).

Figure 10. Activities of ERα and ERRα.

Figure 11. General description of epithelial-mesenchymal plasticity in tumor metastasis from (Mittal, 2018).

Figure 12. Several molecular mechanisms of EMT are affected by ERRα in cancers.

Figure 13. Schematic model presenting the regulation of ERRα on target genes in a distant way from (Cerutti et al., 2023).

Figure 14. Role of ERRα and PGC-1α in the pathogenesis of Alzheimer's disease (AD) from (Sato et al., 2023).

Figure 15. Antagonistic regulation of NCoR1 and PGC-1α on ERRα in oxidative metabolism from (Pérez-Schindler et al., 2012).

Figure 16. The integrated binding motif profile of ESRRA from JASPAR database (matrix ID: MA0592.1) (Castro-Mondragon et al., 2022).

Figure 17. Examples describing the mode of regulation for the TF-target regulatory interactions.

Figure 18. A transcriptional regulatory network (TRN) from (Han et al., 2015) including identified TF-target interactions in TRRUST.

Figure 19. An example of the fitted models from (Lever et al., 2016).

Figure 20. A simple representation of the linear relationship between the new latent

component (comp1) with covariates (i.e., TFs).

Figure 21. The general principle of the application of sparse partial least square (sPLS) regression algorithm to explore the regulation of transcriptional regulators (TRs) on the target gene from (Cerutti et al., 2022).

Figure 22. Workflow used for the prediction of ERRα co-regulators.

Figure 23. Identification of ERRα targets involved in cell migration.

Figure 24. Human breast tumors datasets preparation used in sPLS modeling.

Figure 25. The core of sparse partial least square (sPLS) model computation.

Figure 26. Human TRs preparation used in sPLS modeling.

Figure 27. Identification of potential ERRα co-regulators in different breast cancers subtypes.

Figure 28. ERRα-ZEB1 DEGs.

Figure 29. Box plot shows the number of genes with high-quality sPLS model (nGenesModelOk).

Figure 30. Box plot shows the proportion of genes associated to ZEB1 in all genes with high-quality model (propGenesRelatedToZEB1).

Figure 31. Dot plot shows the frequency of the observed correlation between ZEB1 and ZEB1-relevant genes (genes in Table 5 in yellow: 19) in computed models of each sample set size (corrZEB1DEGs).

Figure 32. Box plot shows the significance of ZEB1 compared to top 2 TR (sigZEB1).

Figure 33. Atlas of breast cancer cell lines analyzed by single cell profiling in Gambardella et al. (2022).

Figure 34. Expression of the 8 DEGs in the clustered breast cancer cell lines.

Figure 35. Heatmap showing the expression levels of the indicated genes in 56 CCLE breast cancer cell lines.

Figure 36. Heatmap showing the expression levels of the indicated genes in human breast cancer tumors (GSE18229).

Figure 37. Heatmap showing the expression levels of the indicated genes in FUSCC triple negative breast cancer tumors.

Figure 38. Expression of the 8 DEGs and ZEB1 in TNBC.

Figure 39. Protein expression of ERRα and ZEB1 in transfected cells.

Figure 40. Common targets of ERRα and ZEB1.

Figure 41. ERRα and ZEB1 shared targets in several TNBC cells.

Figure 42. ERRα and ZEB1 shared targets in non-TNBC breast cancer cells.

Figure 43. ZEB1-responsive genes in TNBC (mRNA).

Figure 44. ERRα-responsive genes in TNBC (mRNA).

Figure 45. ERRα responsive genes in TNBC (ChIP).

Figure 46. Expression of 8 DEGs in TNBC.

Figure 47. Genomic features of ZEB1 binding sites in the whole genome.

Figure 48. Binding intensity of ZEB1 from TSS.

Figure 49. Distribution of ZEB1-binding site relative to TSS.

Figure 50. Distance of the significant peaks (IDR < 0.05) of ERRα and ZEB1 to the TSS on the 8 DEGs.

Figure 51. ERRα and ZEB1 bind to common DNA sites on DEG promoters.

Figure 52. ERRα and ZEB1 interact with each other and bind to common DNA sites on DEG promoters.

Figure 53. ERRα and ZEB1 interact with each other.

Figure 54. Schematic representation summarizing the ERRα-ZEB1 interactions on activated target genes.

Figure 55. Expression profile for nearly 430,000 single cells from 55 patients from (Pal et al., 2021).

Figure 56. Expression of ESRRA, ZEB1 and the 8 DEGs in single cells from TNBC tumors.

Figure 57. Dot plot showing the expression of indicated genes across non-epithelial cells in the microenvironment of TNBC tumors.

Figure 58. Expression of ESRRA, ZEB1 and the 8 DEGs in single cells from ER+ tumors.

Figure 59. Dot plot showing the expression of indicated genes across non-epithelial cells in the microenvironment of ER+ tumors.

Figure 60. Expression of ESRRA, ZEB1 and the 8 DEGs in single cells from Her2+ tumors.

Figure 61. Dot plot showing the expression of indicated genes across non-epithelial cells in the microenvironment of Her2+ tumors.

Figure 62. Scatter plots linking the expression of the indicated genes to the EMT score in TCGA all breast cancer tumors (KS method).

Figure 63. Correlogram showing the relation between expression (log-transformed) of the indicated genes and the EMT score in four independent breast cancer patient

datasets (KS method).

Figure 64. Correlogram showing the relation between expression (log-transformed) of the indicated genes and the EMT score in different subtypes of four independent breast cancer patient datasets (KS method).

Figure 65. Box plot showing the comparison of EMT score between breast cancer subtypes from TCGA (KS method).

Figure 66. Correlogram showing the relation between expression (log-transformed) of the indicated genes and the EMT score in different subtypes of four independent breast cancer patient datasets (76GS method).

Figure 67. Box plot showing the comparison of EMT score between breast cancer subtypes from TCGA (76GS method).

Figure 68. Box plots showing the comparison of EMT score between TNBC subtypes.

Figure 69. Kaplan-Meier curves from GSE96058 breast cancer RNA-Seq dataset showing the overall survival (OS) of tumor subtypes from KM plotter database, based on the expression of the indicated genes (single or in combination).

Figure 70. Kaplan-Meier curves from TCGA breast cancers RNA-Seq dataset showing the overall survival (OS) of tumor subtypes, based on the expression of the 8 DEGs.

Figure 71. Kaplan-Meier curves from chip datasets showing the overall survival (OS) of tumor subtypes from KM plotter database, based on the expression of the 8 DEGs.

Figure 72. Kaplan-Meier curves showing the impact of the combined expression of the 8 DEGs on the relapse free survival (RFS) of tumor subtypes

Figure 73. Kaplan-Meier curves from chip datasets showing the impact of the combined expression of the 8 DEGs on the distant metastasis free survival (DMFS) of tumor subtypes from KM plotter database.

Figure 74. Bar plot from Ma et al. (2019) showing the expression change of EMT features in the treatment of siERRα in breast cancer cells.

Figure 75. Kaplan-Meier curves from TCGA TNBC breast cancers RNA-Seq dataset showing the overall survival (OS), based on the expression of ADAMTS12.

Figure 76. Kaplan-Meier curves from GSE96058 breast cancer RNA-Seq dataset showing the overall survival (OS) from KM plotter database, based on the expression of MCU.

Figure 77. Kaplan-Meier curves from TCGA TNBC breast cancers RNA-Seq dataset showing the overall survival (OS), based on the expression of SPARC.

Figure 78. Kaplan-Meier curves showing the overall survival (OS) of breast tumor subtypes based on the expression of ESRRA and ZEB1 respectively.

Table 1. Number of variants of ESRRA from nine breast cancer projects in cBioPortal.

Table 2. Summary of copy number amplifications of ESRRA from nine breast cancer projects in cBioPortal.

Table 3. Summary of copy number homozygous deletions of ESRRA from nine breast cancer projects in cBioPortal.

Table 4. Summary of mutations of ESRRA from nine breast cancer projects in cBioPortal.

Table 5. The table displays genes with high-quality model (GenesModelOk) upon the modeling of each sample set size.

Table 6. Expression of ERRα and ZEB1 in breast cancer cells.

Table 7. Colocalized proteins of ERRα based on public ChIP-seq data of breast cancer cells in ChIP-Atlas database.

Table 8. 14 TFs share targets with ERRα in human in TRRUST v2 database.

Abbreviations

ACADM/MCAD	Acyl-CoA dehydrogenase medium chain
ACTB	Actin beta
AD	Alzheimer's disease
ADA	Adenosine deaminase
ADAMTS12	ADAM metallopeptidase with thrombospondin type 1 motif 12
ADGRG1	Adhesion G protein-coupled receptor G1
AEBP1	AE binding protein 1
AF-1	Activating function-1
AF-2	Activating function-2
Ang2	Angiogenin, ribonuclease A family, member 2
AR	Androgen receptor
ARID4A	AT-rich interaction domain 4A
ARTN	Artemin
ATP	Adenosine triphosphate
ATP5F1B/ATP5B	ATP synthase F1 subunit beta
ATPsynβ	ATP synthase, beta subunit
AXL	AXL receptor tyrosine kinase
Aβ	Amyloid beta/β
BACE1	Beta-site amyloid precursor protein cleaving enzyme 1
CCN1/CYR61	Cellular communication network factor 1
CCN2/CTGF	Cellular communication network factor 2
CDH1	Cadherin 1
CDH5	Cadherin 5
CEACAM1	CEA cell adhesion molecule 1
CETN3	Centrin 3
CHCHD10	Coiled-coil-helix-coiled-coil-helix domain containing 10
ChIP	Chromatin immunoprecipitation
ChIP-exo	Chromatin immunoprecipitation combines with lambda exonuclease digestion followed by high-throughput sequencing
ChIP-nexus	Chromatin immunoprecipitation experiments with nucleotide resolution through exonuclease, unique barcode and single ligation
ChIP-qPCR	Chromatin immunoprecipitation–Quantitative polymerase chain reaction
ChIP-seq	Chromatin immunoprecipitation sequencing
CLDN3	Claudin 3

CLDN7	Claudin 7
CNA	Copy number alteration
CNOT1	CCR4-NOT transcription complex subunit 1
CNV	Copy number variant
COL10A1	Collagen type X alpha 1 chain
COL11A1	Collagen type XI alpha 1 chain
CoR	Co-regulator
CRE	Cis-regulatory element
CREB1	cAMP responsive element binding protein 1
CSPG4	Chondroitin sulfate proteoglycan 4
CtBP	C-terminal Binding Protein
CTHRC1	Collagen triple helix repeat containing 1
CXCL11	C-X-C motif chemokine ligand 11
CXCL12	C-X-C motif chemokine ligand 12
CXCR6	C-X-C motif chemokine receptor 6
Cyt c	Cytochrome c, somatic
DAB2	DAB adaptor protein 2
DBD	DNA binding domain
DDR2	Discoidin domain receptor tyrosine kinase 2
DEGs	Ddifferentially expressed genes
DNA	DeoxyriboNucleic Acid
DUSP1	Dual specificity phosphatase 1
DYSF	Dysferlin
E2	Estrogen
EBAG9/RCAS1	Estrogen receptor binding site associated antigen 9
E-cadherin	Epithelial cadherin
ECM1	Extracellular matrix protein 1
EDN1	Endothelin 1
EDNRA	Endothelin receptor type A
ELF1	E74 like ETS transcription factor 1
ELK4	ETS transcription factor ELK4
EMT	Epithelial mesenchymal transition
ENO1	Enolase 1
EP300	E1A binding protein p300
EPAS1	Endothelial PAS domain protein 1
EPB41L4B	Erythrocyte membrane protein band 4.1 like 4B
EpCAM/CD326	Epithelial cell adhesion molecule
EPPK1	Epiplakin 1
ER	Estrogen receptor
ERE	Estrogen response element
ERRE	ERR response element

ERRα	Estrogen-related receptor alpha
ERRβ	Estrogen-related receptor β
ERRγ	Estrogen-related receptor gamma
ERα	Estrogen receptor alpha
ESR1	Estrogen receptor 1
ESRRA	Estrogen related receptor alpha
ESRRB	Estrogen related receptor beta
ETS1	ETS proto-oncogene 1, transcription factor
FDA	US Food and Drug Administration
FFPE	Formalin fixation and paraffin embedding
FGF1	Fibroblast growth factor 1
FOSL1	FOS like 1, AP-1 transcription factor subunit
fpkm	Fragments per kilo base of transcript per million mapped fragments
FSTL1	Follistatin like 1
FUSCC	Fudan University Shanghai Cancer Center
FUT8	Fucosyltransferase 8
FYN	FYN proto-oncogene, Src family tyrosine kinase
GADD45A	Growth arrest and DNA damage inducible alpha
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GOT2	Glutamic-oxaloacetic transaminase 2
GPS2	G protein pathway suppressor 2
GR	Glucocorticoid receptor
GRN	Gene regulatory network
GSK-3β	Glycogen synthase kinase-3 betat/β
GTEx	Genotype-Tissue Expression Project
HCFC2	Host cell factor C2
HDAC1	Histone deacetylase 1
HDAC3	Histone deacetylase 3
HDAC8	Histone deacetylase 8
HER2	Human epidermal growth factor receptor 2
HIPK3	Homeodomain interacting protein kinase 3
HMOX1	Heme oxygenase 1
IDH3A	Isocitrate dehydrogenase (NAD(+)) 3 catalytic subunit alpha
IGF	insulin-like growth factor
IGFBP3	Insulin like growth factor binding protein 3
IL12A	Interleukin 12A
IL6	Interleukin 6
IL6R	Interleukin 6 receptor
ITGA6/CD49f	Integrin subunit alpha 6

ITGAX	Integrin subunit alpha X
ITGB3	Integrin subunit beta 3
JAK1	Janus kinase 1
JUN	Jun proto-oncogene, AP-1 transcription factor subunit
KAT2B	K(lysine acetyltransferase 2B
KDM5B	Lysine demethylase 5B
KDR/VEGFR2	Kinase insert domain receptor
KO	Knock-out
LBD	Ligand binding domain
LCP1	Lymphocyte cytosolic protein 1
LEF1	lymphoid enhancer binding factor 1
LHFPL6/LHFP	LHFPL tetraspan subfamily member 6
LIMCH1	LIM and calponin homology domains 1
LLGL2	LLGL scribble cell polarity complex component 2
LMNB1	Lamin B1
LPAR1	Lysophosphatidic acid receptor 1
LSD1/KDM1A	Lysine demethylase 1A
LXR	Liver X receptors
MACIR	Macrophage immunometabolism regulator
MATN2	Matrilin 2
MCAM/CD146	Melanoma cell adhesion molecule
MCU	Mitochondrial calcium uniporter
MED1/PPARBP	Mediator complex subunit 1
MGAT5	Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase
MR	Mineralocorticoid receptors
MRGBP	MRG domain binding protein
mRNA	Messenger ribonucleic acid
MTA1	Metastasis associated 1
MYBL2	MYB proto-oncogene like 2
MYLK	Myosin light chain kinase
NANOS1	Nanos C2HC-type zinc finger 1
N-cadherin	Neural cadherin
NCOA2/SRC-2	Nuclear receptor coactivator 2
NCOA3/AIB1SRC-3	Nuclear receptor coactivator 3
NCOA4	Nuclear receptor coactivator 4
NCoR1	Nuclear receptor corepressor 1
NFATC3	Nuclear factor of activated T cells 3
NFKB1	Nuclear factor kappa B subunit 1
NFTs	Neurofibrillary tangles
NOP2	NOP2 nucleolar protein

NR	Nuclear receptor
NR3C1	Nuclear receptor subfamily 3 group C member 1
NR5A1	Nuclear receptor subfamily 5 group A member 1
NRIP1/RIP140	Nuclear receptor interacting protein 1
Nrp1	Neuropilin-1
NSD1/ARA267	Nuclear receptor binding SET domain protein 1
NUDT19	Nudix hydrolase 19
P300/CBP	CREB-binding protein and E1A binding protein p300
p63	Tumor protein p63
PBM	Protein binding-microarrays
PCR	Polymerase chain reaction
PDGFB	Platelet derived growth factor subunit B
PDLIM1	PDZ and LIM domain 1
PECAM1	Platelet and endothelial cell adhesion molecule 1
PGC-1	Peroxisome proliferator-activated receptor gamma coactivator 1
PGC-1α	Peroxisome proliferator-activated receptor-gamma coactivator - 1alpha
PIAS1	Protein inhibitor of activated STAT 1
PLA	Proximity ligation assay
PLXNC1	Plexin C1
POU2F1	POU class 2 homeobox 1
PPARGC1A	PPARG coactivator 1 alpha
PPARα/PPARβ/PPARδ/PPARγ	Peroxisome proliferator-activated receptor alpha/beta/delta/gamma
PPI	Protein-protein interaction
PR	Progesterone receptor
PREX1	Phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 1
PROX1	Prospero homeobox 1
qPCR	Quantitative polymerase chain reaction
R2	R-squared determination coefficient
RAR	Retinoic acid receptor
RARα/RARβ	Retinoic acid receptor alpha/beta
RBPJ	Recombination signal binding protein for immunoglobulin kappa J region
RE	Response element
RELN	Reelin
RHOBTB3	Rho related BTB domain containing 3
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing

RT-qPCR	Quantitative reverse transcription polymerase chain reaction
RXR	Retinoid X receptors
SELEX	Systematic evolution of ligands by exponential enrichment
SEMA6D	Semaphorin 6D
SEMA7A	Semaphoring 7A
SERPINE1	Serpin family E member 1
SETD7	SET domain containing 7, histone lysine methyltransferase
SINHCAF	SIN3-HDAC complex associated factor
SIRT1	Sirtuin 1
SLIT2	Slit guidance ligand 2
SNAIL/SNAI1/SNAIL1	Snail family transcriptional repressor 1
SNAIL2/SNAI2	Snail family transcriptional repressor 2
SP1	Sp1 transcription factor
SPARC	Secreted protein acidic and cysteine rich
SRC	SRC proto-oncogene, non-receptor tyrosine kinase
SRC-1	Nuclear receptor coactivator 1
STAT3	Signal transducer and activator of transcription 3
TBL1	Transducin β-like 1
TBLR1	TBL1X/Y related 1
TCA	Tricarboxylic acid cycle
TCF12	Transcription factor 12
TCGA	The Cancer Genome Atlas
TFAM/mtTFA	Transcription factor A, mitochondrial
TFBS	TF binding site
TFF1/pS2	Trefoil factor 1
TFs	Transcription factors
TGF-β	Transforming growth factor beta
THBS1	Thrombospondin 1
Tim22	Translocase of inner mitochondrial membrane 22
TMEM120B	Transmembrane protein 120B
TNM	TNM classification of malignant tumors
TOP2B	DNA topoisomerase II beta
tpm	Transcripts per kilobase million
TRERF1/TReP-132	Transcriptional regulating factor 1
TRN	Transcription regulatory network
TRs	Transcription regulators
TRα /TRβ	Thyroid hormone receptor alpha/beta
TSS	Transcriptional start site
UBP1	Upstream binding protein 1
Ucp1	Uncoupling protein 1
VCAM-1	Vascular cell adhesion protein 1

VDR	Vitamin D receptor
VEGF	Vascular endothelial growth factor
WNT7B	Wnt family member 7B
WT1	WT1 transcription factor
YAP/YAP1	Yes1 associated transcriptional regulator
YBX1	Y-box binding protein 1
YWHAB	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein beta
ZEB	Zinc-finger E-box-binding
ZEB1	Zinc finger E-box binding homeobox 1
ZNF619	Zinc finger protein 619
ZNF792	Zinc finger protein 792
α-SMA	Alpha smooth muscle actin

Introduction

Chapter I - Breast cancer

1.1 General introduction

Breast cancer (BC) is the most common cancer affecting women in 157 countries out of 185 (World Health Organization, 2024). Approximately 0.5–1% BC are diagnosed in men. Globally there are an estimated 2.3 million new cases and 670,000 deaths recorded in 2022. BC occurs in women at any stage of age after puberty and increases with age. The survival rate of BC has improved but the incidence of BC increases yearly.

The risk factors of BC include genetic and non-genetic factors. The genetic factors comprise DNA alterations and epigenetics. The non-genetic factors such as age, reproductive risk factors, lifestyle, radiation exposure and high mammographic density also have an impact on BC along with genetic alteration (Nolan et al., 2023).

1.2 Breast cancer subtype classification

BC is characterized by a high degree of tumor heterogeneity. There are two main classification methods of breast tumors. On the one hand, based on the presence or absence of immunohistochemical markers, estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (Her2), BCs are classified in three subtypes with different clinical features: ER positive/PR positive/Her2 negative (70% of patients), ER any status/PR any status/Her2 positive (15%-20%), and ER negative/PR negative/Her2 negative (triple negative breast cancer; TNBC: 15%) (**Figure 1**) (Waks and Winer, 2019). On the other hand, the RNA-based profiling method according to the PAM50 classification was also used to divide patients into four clinical molecular subtypes (also known as intrinsic subtypes: luminal A, luminal B, Her2-enriched, basal-like) and a normal-like subtype (**Figure 2**) (Pommier et al., 2020). An additional molecular subtype, claudin-low subtype, which has some common features with basal-like subtype, was discovered more recently in an integrated analysis on human and mouse tumors (Pommier et al., 2020). Claudin-low tumors present mesenchymal features, low expression of adhesion proteins and immune cell infiltration (Nolan et al., 2023). Basal-like and claudin-low tumors are grouped with the majority of TNBC to form the most aggressive subtype (Pommier et al., 2020).

	Hormone Receptor (HR) +/*ERBB2*−	*ERBB2*+ (HR+ or HR−)	Triple-Negative
Pathological definition	≥1% Of tumor cells stain positive for estrogen receptor or progesterone receptor proteins	Tumor cells stain strongly (3+) for ERBB2 protein or *ERBB2* gene is amplified in tumor cells. Approximately half of *ERBB2*+ tumors are also HR+	Tumor does not meet any pathologic criteria for positivity of estrogen receptor, progesterone receptor, or *ERBB2*
Molecular pathogenesis	Estrogen receptor α (a steroid hormone receptor) activates oncogenic growth pathways	The oncogene *ERBB2*, encoding ERBB2 receptor tyrosine kinase from the epidermal growth factor receptor family, is overactive	Unknown (likely various)
Percentage of breast cancer cases, %[12]	70	15-20	15
Prognosis			
Stage I (5-y breast cancer-specific survival), %[13,a]	≥99	≥94	≥85
Metastatic (median overall survival)[14-16,b]	4-5 y	5 y	10-13 mo
Typical systemic therapies for nonmetastatic disease (agents, route, and duration)	• Endocrine therapy (all patients): • Tamoxifen, letrozole, anastrozole, or exemestane • Oral therapy • 5-10 y • Chemotherapy (some patients): • Adriamycin/cyclophosphamide (AC) • Adriamycin/cyclophosphamide/ paclitaxel (AC-T) • Docetaxel/cyclophosphamide (TC) • Intravenous therapy • 12-20 wk	• Chemotherapy plus ERBB2-targeted therapy (all patients): • Paclitaxel/trastuzumab (TH) • Adriamycin/cyclophosphamide/ paclitaxel/trastuzumab ± pertuzumab (AC-TH±P) • Docetaxel/carboplatin/trastuzumab ± pertuzumab (TCH±P) • Intravenous therapy • 12-20 wk of chemotherapy; 1 y of ERBB2-targeted therapy • Endocrine therapy (if also hormone receptor positive) • Tamoxifen, letrozole, anastrozole, or exemestane • Oral therapy • 5-10 y	• Chemotherapy (all patients): • AC • AC-T • TC • Intravenous therapy • 12-20 wk

[a] Prognosis estimated from nearly 44 000 patients with breast cancer enrolled in the California Cancer Registry (2005-2008); stage I breast cancer defined by the American Joint Committee on Cancer staging manual anatomic staging table, 8th edition. 95% CIs for stage I 5-year breast cancer–specific survival estimates are as follows: 98%-100% for HR+/*ERBB2*−, 83%-100% for *ERBB2*+, and 75%-98% for triple-negative.

[b] Prognosis listed is from time of diagnosis of metastatic breast cancer.

Figure 1. Introduction of definition, pathogenesis, case percentage, prognosis and treatment for the three breast cancer subtypes from (Waks and Winer, 2019). ERBB2 is also known as human epidermal growth factor 2 (Her2). Hormone receptors (HR) include estrogen receptor (ER) and progesterone receptor (PR).

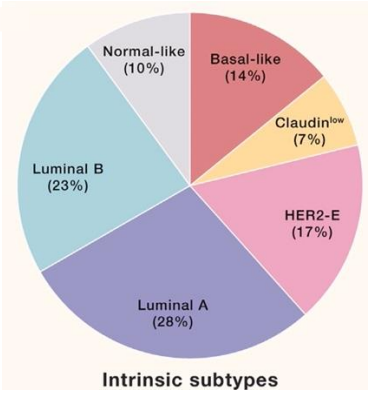


Figure 2. Pie chart from (Nolan et al., 2023) presenting the percentage of each subtype of breast cancers characterized by molecular profiling.

1.3 Breast tumor heterogeneity

Over the past decade, genetic, epigenetic and phenomenological data demonstrated a heterogeneity in BC which influences cancer progression and clinical therapeutic

strategies (Martelotto et al., 2014). Tumor heterogeneity includes intertumoral and intratumoral heterogeneity (Koren and Bentires-Alj, 2015). Intertumoral heterogeneity is caused by the different etiological and environmental factors between tumors (Liu et al., 2018), usually leading to distinct molecular intrinsic subtypes of cancer (**Figure 1-2**). Even though current studies mostly focus on exploring the underlying distinct features characterizing cancer subtypes, a growing number of studies gradually found it is also necessary to take intratumoral heterogeneity into account (Nolan et al., 2023). It should be noted that intratumoral heterogeneity exists within a tumor, or between a primary tumor and its metastatic implantation (Martelotto et al., 2014). Cells in one tumor show diverse genomic and biological variations based on the tumor cell evolution associated to genomic aberrations and tumor microenvironment (TME) (Martelotto et al., 2014). For instance, a study showed that the intratumoral heterogeneity had an impact on known driver genetic variants of TP53 and PIK3CA in breast cancer (Shah et al., 2012; Martelotto et al., 2014). Therefore, considering both of intertumoral and intratumoral heterogeneity in the investigation of cancer mechanism has become popular in genetics.

1.4 Tumor microenvironment

TME is composed of infiltrating immune cells, stromal cells (e.g. Cancer-Associated Fibroblasts, CAFs), blood vessels, and extracellular matrix (ECM) (Tan et al., 2022). Stromal cells have been gradually noticed and recognized as important components in cancer development. The subtype of stromal cells varies between cancer types, however, it commonly includes vascular endothelial cells, fibroblasts, adipocytes and stellate cells (Anderson and Simon, 2020). In TME, stromal cells interact with cancer cells and secret many factors which play a critical role in developing tumor heterogeneity, proliferation, invasion, metastasis and chemoresistance (**Figure 3**) (Anderson and Simon, 2020; Mehraj et al., 2021). A deeper dig in the tumor microenvironment will therefore be helpful to explore tumor heterogeneity.

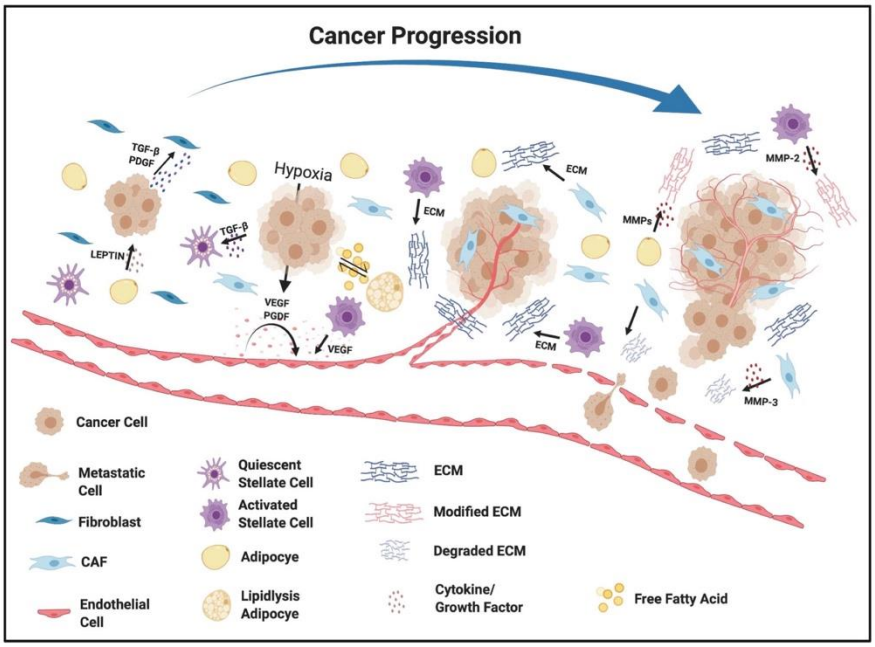


Figure 3. Distribution of stromal cells within tumors from (Anderson and Simon, 2020). Cancer cells attract and interact with diverse stromal cells that secret many factors influencing tumor features.

1.5 Significance of tumor heterogeneity

To better understand and investigate tumor heterogeneity, several technologies have been developed and used. For example, tumor bulk sequencing, detection of ultra-rare mutations, single-molecule sequencing and single-cell sequencing (Martelotto et al., 2014). Single-cell sequencing (scRNA-Seq), elucidating the expression of each gene in every individual cell, such as cells of TME, is helpful to dissect tumor heterogeneity between diverse tumor cell types (Nolan et al., 2023). Moreover, it can also distinguish the most potential molecules which contribute to tumor development and metastasis at a higher level of resolution (Ding et al., 2020). For instance, Wu et al. (2020) detected two subsets of CAFs and two subtypes of perivascular-like (PVL) cells based on the scRNA-Seq data of tumor tissues from five TNBC patients: myofibroblast-like CAFs (myCAFs), inflammatory-CAFs (iCAFs), differentiated-PVL (dPVL) cells and immature-PVL (imPVL) cells. They found these four subpopulations of stromal cells had distinct spatial features and functions. A significant association between iCAFs and dPVL cells with immune evasion has been identified in multiple TNBC cohorts. These specific subtypes of stromal cells provide candidates to be used to develop TME-directed therapy (Wu et al., 2020b). Altogether, high-throughput technologies such as scRNA-Seq are widely used to further investigate the cellular heterogeneity of tumor microenvironment, to reveal the underlying pathological mechanism and tumor heterogeneity. Moreover, they also give a new insight into the clinical treatment of cancers, including TNBC.

1.6 Breast cancer treatment

Early-stage BCs are mostly treated by surgery and radiotherapy along with systemic therapy if necessary. Primary endocrine therapy is usually used to treat patients with ER positive BCs. Neoadjuvant therapies that include chemotherapy, targeted drugs or immune checkpoint inhibitors have been a standard care in particular for early-stage Her2 positive and TNBC patients (Loibl et al., 2021; Hong and Xu, 2022). However, the response of TNBC to effective treatment strategies is very limited.

On the one hand, the lack of ER, PR, and Her2 expression in TNBC tumors causes an insensitivity to endocrine therapy or Her2-targeting treatment (Yin et al., 2020). The main treatment for TNBC is chemotherapy, whereas the efficacy of conventional adjuvant chemoradiotherapy after surgery is still limited (Yin et al., 2020). For example, Bevacizumab, which has been added to chemotherapy for the treatment of TNBC patients, does not show a significant benefit in survival time (Collignon et al., 2016; Yin et al., 2020). TNBC is still lethal even though some therapeutic agents were approved by the US Food and Drug Administration (FDA) for these patients (Derakhshan and Reis-Filho, 2022). On the other hand, TNBC also displays a

significant level of heterogeneity. Multiple TNBC subtypes have been characterized. For instance, Lehmann et al. (2011) initially identified six distinct subgroups of TNBC, based on the gene expression data from 587 TNBC cases, namely basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem–like (MSL) subtype, and luminal androgen receptor (LAR) subtypes (Lehmann et al., 2011).

In addition, a recent publication based on transcriptomic data, also classified a large Chinese cohort of 465 primary TNBC into four distinct subtypes in which specific potential targets for precision treatment have been suggested. These subtypes are basal-like immune-suppressed (BLIS) subtype, immunomodulatory (IM) subtype, luminal androgen receptor (LAR) subtype and mesenchymal-like (MES) subtype (Jiang et al., 2019).

Altogether, the huge heterogeneity of TNBC is an important factor affecting the therapeutic effects in TNBC patients. Taking tumor heterogeneity of breast cancer into account also guides scientific and clinical researchers to investigate novel pathological mechanisms and to explore potential therapeutic agents.

1.7 Conclusion

In summary, tumor heterogeneity, including intertumoral and intratumoral heterogeneity, is crucial to investigate the potential mechanisms of development of breast cancer. In this case, intertumoral heterogeneity mainly reflects the molecular subtype of breast tumors (**Figure 1-2**), while intratumoral heterogeneity relates to genomic aberrations and tumor microenvironment (TME).

In my thesis, taking into account the heterogeneity of breast cancers, in particular, of TNBC, we have explored new pathogenic mechanisms of breast cancer development to provide new prospects on survival prediction as well as on new possible therapeutic strategies.

Chapter II – Nuclear receptors and co-regulators

2.1 General introduction of nuclear receptors

Nuclear receptors (NRs) form a superfamily of transcription factors (TFs) displaying similar domain structure and functions (Ishigami-Yuasa and Kagechika, 2020). They regulate various biological phenomena such as proliferation, differentiation, metabolism, survival and reproduction in all animals (Tecalco-Cruz, 2018). 48 genes encoding nuclear receptors have been identified in human and are considered as potential important targets for the development of clinical therapeutic drugs (**Figure 4**) (Jin and Li, 2010; Tao et al., 2020). NRs are mainly composed of two groups: ligand-dependent receptors and orphan nuclear receptors (Olefsky, 2001; Lonard and O'malley, 2007; Chen, 2008).

Ligand-regulated NRs include receptors for hormonal ligands, comprising steroids and non-steroids. The steroid receptors include ER, PR, androgen receptor (AR), glucocorticoid receptor (GR) and mineralocorticoid receptor (MR) (Rochette-Egly, 2003; Chen, 2008; Treviño and Gorelick, 2021). The non-steroids receptors include retinoic acid receptors (RAR), thyroid hormone receptors, retinoid X receptors (RXR), vitamin D receptors (VDR) and peroxisome proliferator activated receptors (PPAR) (Chen, 2008). In contrast, orphan NRs, which have no identified ligand to date, are defined as functional receptors modulating the expression of their targets in a ligand-independent manner.

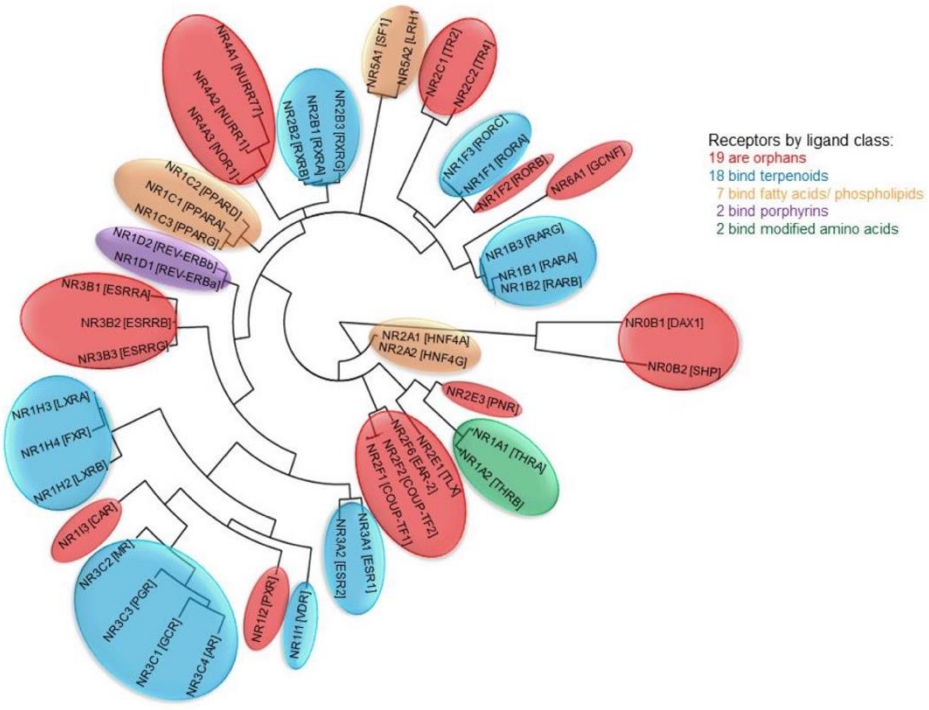


Figure 4. Introduction of 48 human nuclear receptors (NRs) from (Tao et al., 2020).

NRs are grouped based on their homology to the steroid receptors.

2.2 Basic structure of nuclear receptors

NRs have similar structural features. They comprise a variable A/B domain at the N-terminal end which includes the transcription activation domain AF-1. The DNA binding domain (DBD) is the central region which directly contacts a specific DNA sequence, known as response element (RE), in the promoter of the target genes. The ligand binding domain (LBD) is conserved and located at the C-terminal end of the protein. The LBD mediates the downstream transcription activity by specifically interacting with co-regulators through the conserved AF-2 helix located at the end of LBD. This domain is also involved in receptor homodimerization or heterodimerization. A hinge region of variable length links the DBD and LBD regions, and contributes to NR dimerization and DNA binding (**Figure 5**) (Olefsky, 2001; Chen, 2008; Huang et al., 2010).

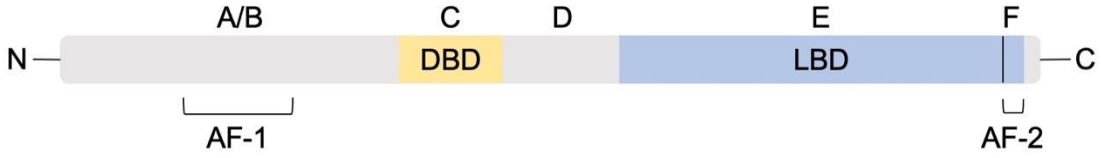


Figure 5. Structure of nuclear receptors (NRs). In their N-terminal part, NRs comprise the A/B domain (region A/B), containing the AF-1 transcription activation function. The DNA binding domain (DBD) (region C) is linked to ligand binding domain (LBD) (region E) by the non-conserved hinge region (region D). The LBD contains the ligand-inducible activation function domain AF-2 (region F) and is responsible for interacting with specific co-regulators (Olefsky, 2001; Chen, 2008).

2.3 Molecular functions of nuclear receptors

To better understand the effect of NRs in biological activities, it is necessary to briefly describe their molecular functions.

NRs play a central role in the physiological state through the regulation of the expression of their target genes. This regulation by NRs occurs via recognition of specific DNA sequences on the promoter region of genes by the DBD region, followed by an interaction with RNA polymerase II (pol II) multiprotein complex which promotes the initiation of transcription (**Figure 6**). For instance, upon estrogen (E2) binding, ER binds DNA directly through estrogen response elements (EREs) (Clusan et al., 2023). In addition, the regulation of the activity of NRs is highly controlled by co-regulators (CoRs) in a ligand-dependent or ligand-independent way. For instance, ER and co-activator p300 form a complex and are co-recruited to activate the expression of target genes in an estrogen-dependent manner (Guertin et al., 2014).

Currently, nearly 300 CoRs have been identified including co-activators which enhance the transcriptional regulation exerted by NRs and co-repressors which, in contrast, repress the expression of target genes (Lonard and O'malley, 2007). Noteworthy, there is only a small set of NRs but several of them participate in the pathophysiological responses in various cancers and in different tissues. This has revealed the contribution of tissue-specific CoRs, indicating that their combination with common NRs in different tissues led to particular tissue-specific biological responses (Chen, 2008).

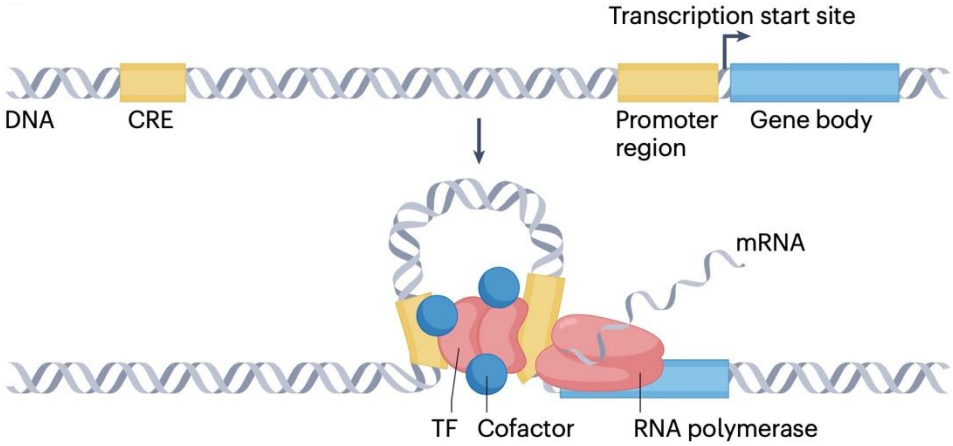


Figure 6. Mechanism of regulation of gene expression from (Badia-I-Mompel et al., 2023). Transcription factors (TFs) cooperate with cofactors and bind to response element (cis-regulatory elements; CREs) and to the promoter region, inducing the recruitment of RNA polymerase and the mRNA synthesis from their target genes.

2.4 Relations between nuclear receptors and disease

NRs have been demonstrated as associated with many human diseases, including cancers.

On the one hand, the cooperation of NRs and CoRs in cancers has gained much attention since a crucial role of CoRs has been demonstrated in the regulation of transcription exerted by NRs. For instance, Mediator subunit 1 (MED1) has been identified as an ER co-activator that affects BC metastasis and resistance to treatment (Leonard and Zhang, 2019). Other ER co-activators have also been identified in BC, such as AIB1 (Osborne et al., 2003), PPARBP (McCafferty et al., 2009) and SRC-1 (McDonnell et al., 2000; Shang et al., 2000) . In addition, ARA267 was found to be a co-regulator of AR and disrupting their interaction may serve as a potential therapeutic target in prostate cancer (Sampson et al., 2001). TReP-132 serves as a co-activator of PR to mediate cell growth inhibition and differentiation-enhancing activities of progesterone in breast cancer cells (Gizard et al., 2006). Together, NRs develop diverse pathological functions in a CoR-dependent manner in various cancers, including BC.

On the other hand, genetic defects of NRs were also demonstrated to be associated

with various clinical phenotypes and abnormal physiological states. Up to now, 20 out of the 48 human NRs have been shown to exhibit variations correlated to human disorders, thereby showing a potential as targets for drug development and therapeutic treatment (Achermann et al., 2017). A pathogenic homozygous variant (p.Arg157*) of ESR1 (encoding ERα) was reported in a young man (Smith et al., 1994). This person had abnormal bone growth, increased level of follicle-stimulating hormone and luteinizing hormone and impaired glucose tolerance. Homozygous mutations in ESRRB (encoding ERRβ) have been detected in patients with hearing loss (Collin et al., 2008). Globally, variations of NRs can result in important effects on human diseases, which suggests that they could be used as therapeutic targets in treatment.

Altogether, NRs have an impact on human diseases, including cancers, in a mutation-dependent manner and in a CoR-dependent manner.

2.5 Conclusion

In summary, NRs are transcription factors that are involved in many pathological and physiological activities. Their involvement in human diseases can occur through various mechanisms, including mutation-dependent or CoRs-dependent manners. The transcriptional regulation exerted by NRs on downstream pathways in cooperation with diverse CoRs may affect various pathologies, including cancers. Therefore, NRs and their recruited CoRs have the potential to be a research direction of human disease pathology, and even potential therapeutic targets of cancers, such as breast cancer.

Chapter III – ERRα

Estrogen-related receptor alpha (ERRα), an orphan nuclear receptor, is expressed in many tissues, in particular in high energy-demanding ones, such as muscle, liver, kidney, heart and colon (**Figure 7**) (Dwyer et al., 2010; Tsushida et al., 2018; Tran et al., 2021). ERRα participates in diverse processes, such as bone homeostasis, cell energy metabolism and cancer development (Fradet et al., 2011; Bianco et al., 2012). In cancer, ERRα is involved in angiogenesis, proliferation, cell adhesion, cell migration and invasion as well as metastasis (Dwyer et al., 2010; Deblois and Giguère, 2013; Sailland et al., 2014; Tribollet et al., 2022). No natural ligand has been identified for ERRα although some synthetic compounds have been proved to modulate the regulatory functions exerted by ERRα (Crevet and Vanacker, 2020). However, the recruitment of co-activators or co-repressors to ERRα is a crucial step to control the transcriptional output on target genes (Cerutti et al., 2023).

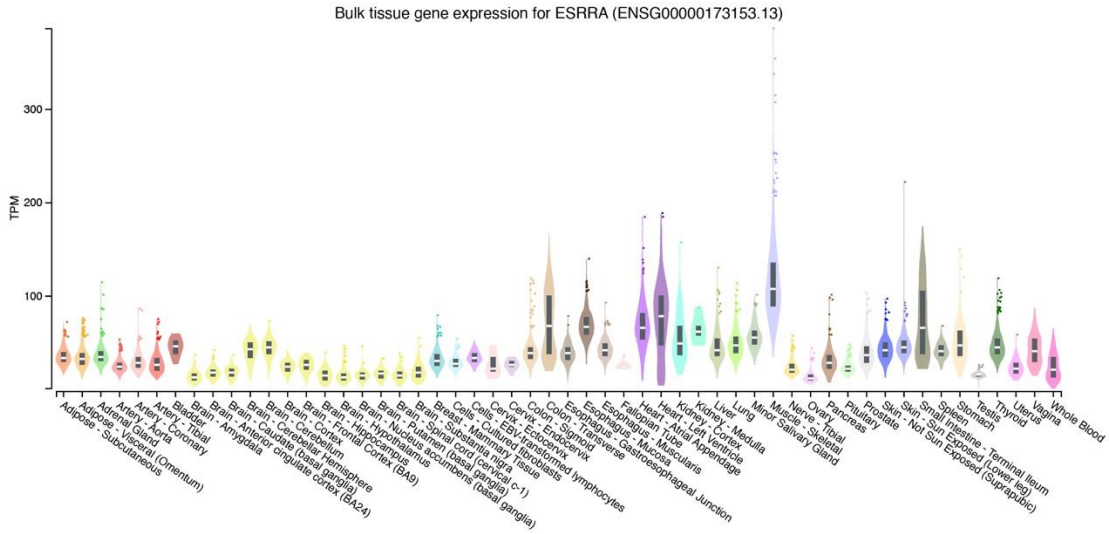


Figure 7. Expression of ESRRA (encoding ERRα) in human normal tissues from (GTEx Consortium, 2020). The data used for the analyses were obtained from the GTEx Portal on 08/28/2024.

3.1 Basic structure of estrogen related receptors

In addition to ERRα, the ERR sub-family also includes other two members: ERRβ and ERRγ (Horard and Vanacker, 2003). All ERR receptors have a similar structural organization and comprise the six conserved domains that have been identified in nuclear receptors (**Figure 8**) (Tang et al., 2021).

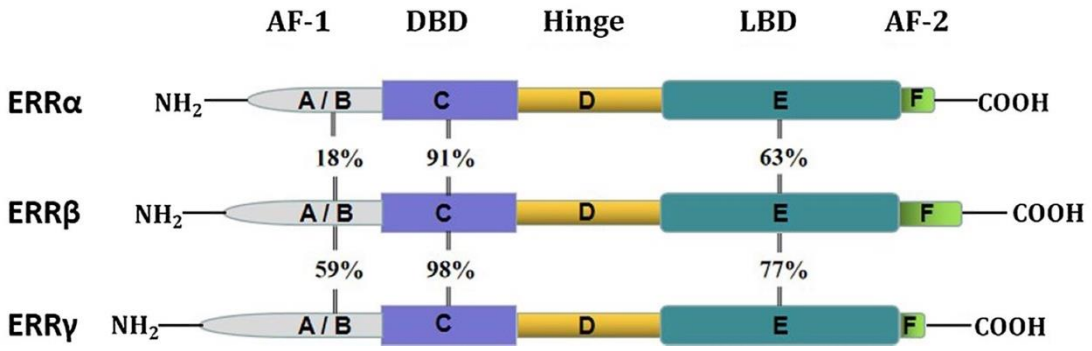


Figure 8. Structure of the ERR receptors from (Tang et al., 2021). Percentage homology of each domain of paired isoforms is displayed. AF-1: activating function-1; DBD: DNA binding domain; LBD: ligand-binding domain; AF-2: activating function-2.

Evidence demonstrated the role of ERRs in pathological and physiological processes, including in cancer, bone and metabolic development (Tang et al., 2021). The conserved DBD of ERRs family implies these three ERR receptors may share common targets. For instance, similar biological functions related to bone homeostasis were identified for ERRα and ERRγ (Tang et al., 2021). Inhibition of ERRα or ERRγ resulted in the promotion of bone formation and a compensation for bone loss resulting from ageing or estrogen-deficiency (Carnesecchi and Vanacker, 2016). However, ERR members also preserve specific features. ERRα has been suggested as a prognostic predictor of BCs, particularly in TNBC, while ERRβ has comparable low expression in BCs, as well as a significant lower expression in TNBC, compared to other subtypes (Ariazi et al., 2002; Fernandez et al., 2020). Overexpression of ERRγ has been shown in 75% of BC samples but not significantly observed in TNBC subtype (Treeck et al., 2020). These three ERRs family members have both similar and different physiological functions although they possess high sequence similarities.

3.2 Relation between ERRα and ERα

In addition to ERRβ and ERRγ, ERRα also shows a high sequence homology with ERα in the DBD (70%) and 36% homology in the LBD (**Figure 9**) (Tang et al., 2021). ERα is involved in numerous biological processes of BC, such as cell proliferation, tumor growth, cancer progression and treatment resistance (Shi et al., 2009; Tecalco-Cruz et al., 2019; Chen et al., 2022). However, ERα's activities (including the recruitment of CoRs) are ligand-dependent, contrasting with ligand-independent activities of ERRα (**Figure 10**) (Vrtačnik et al., 2014; Tang et al., 2021).

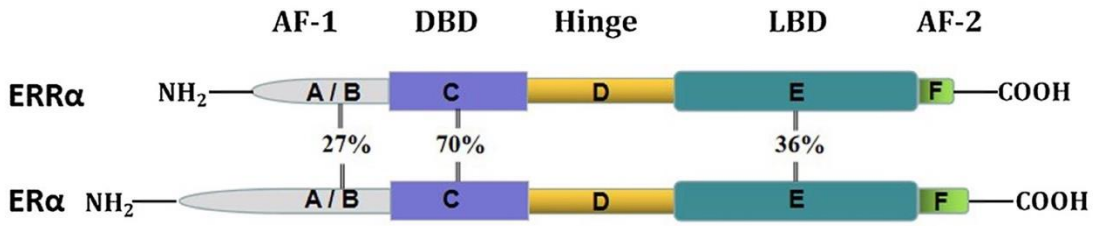


Figure 9. Structure of protein sequence of ERRα and ERα from (Tang et al., 2021). The percentage homology of each domain is shown.

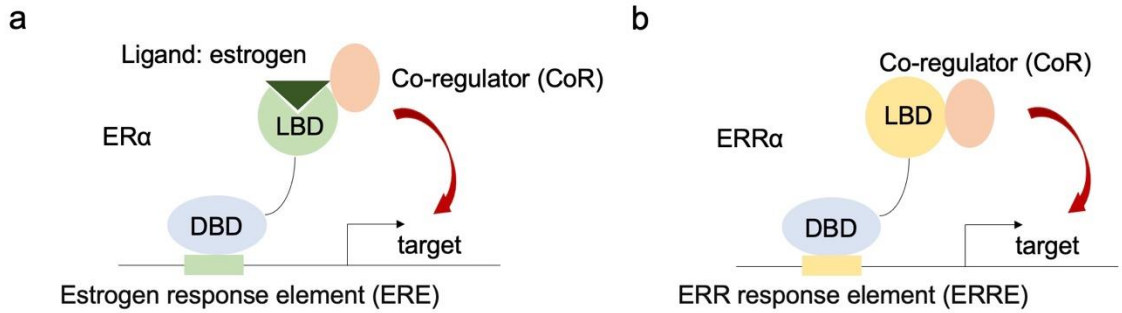


Figure 10. Activities of ERα and ERRα. a. ERα binds to the estrogen response element (ERE) on the promoter of its target genes through the DBD. Estrogen is necessary for ERα to recruit CoRs via its LBD and to induce the expression of its target genes. b. ERRα binds to ERR response element (ERRE) on the promoter of its target genes. ERRα interacts with co-regulators to modulate the expression of targets.

Although ERRα and ERα function in a different manner, they also associate to some common target genes and co-regulators (Horard and Vanacker, 2003; Bianco et al., 2012; Thewes et al., 2015). For instance, Heck et al. (2009) found that AIB1 not only regulates the transcription activity of ERα as a co-activator (Kiliti et al., 2023), but also modulates the activity of ERRα by acting as a substitute "protein ligand" (Heck et al., 2009). A study has shown that the association between ERα and its responsive genes (such as pS2 and EBAG9/RCAS1) depends on the status of ERRα in breast cancer patients, indicating a potential impact of ERRα on the transcriptional activity of ERα (Suzuki et al., 2004). A high involvement of ERRα and ERα in similar physiological pathways, such as bone formation and development was identified based on their co-expression in osteoblasts in vivo and in vitro (Bonnelye et al., 2002). In addition, the overlapped regulatory function of ERRα and ERα was also identified in promoting proliferation of breast cancer tamoxifen-resistant cells even though they target on a smaller proportion of common target genes (Thewes et al., 2015). In summary, ERRα and ERα share co-regulators and targets in BC, and ERRα may functionally affect ERα, suggesting a significant role of ERRα in BC (**Figure 10**).

3.3 Biological functions of ERRα

Literature reviewing indicates that ERRα is expressed in various tissues and functions in different processes such as bone homeostasis, metabolism and cancer development.

(1) Bone

ERRα knock-out (KO) mice show a resistance to the bone loss induced by age or ovariectomy (Delhon et al., 2009; Teyssier et al., 2009; Zhang et al., 2016). Compared to wild-type counterparts, mutant animals display an increase in bone formation rate and osteoblast (cells secreting bone matrix proteins) differentiation and activity (Zhang et al., 2016; Veis and O'Brien, 2023). In addition, Wei et al. (2016) also observed that the cholesterol-induced bone loss disappeared in ERRαKO mice (Wei et al., 2016). Moreover, an enhanced osteoblastic differentiation characterized by an increase of bone sialoprotein and a decrease of osteopontin secretion was detected upon treatment with ERRα inhibitor (Delhon et al., 2009). In summary, ERRα plays an important role in bone homeostasis, reducing osteoblast differentiation and inducing bone loss.

(2) Metabolism

ERRα is expressed in many human tissues, in particular in high energy-demanding tissues such as muscle, liver, kidney, heart and colon (**Figure 7**), in line with its demonstrated participation to cellular metabolism. Indeed, ERRα regulates the expression of genes related to lipid metabolism, mitochondrial biogenesis, oxidative phosphorylation, ATP synthesis, pyruvate metabolism, tricarboxylic acid cycle (TCA) and glycolysis (Ranhotra, 2022; Vanacker and Forcet, 2024).

ERRα was reported to regulate the expression of metabolic genes such as MCAD, IDH3A and ATP5B through the binding to the promoter regions of these targets and the recruitment of members of the PGC-1 family of co-activators (Giguère, 2008; Villena and Kralli, 2008). PGC-1α is considered as a surrogate protein ligands of ERRα because PGC-1α has a highly specific nuclear receptor-interacting surface for ERRα and is responsible for many cellular functions of ERRα (Deblois and Giguère, 2013). For instance, a synergistic co-regulation by ERRα and PGC-1α was shown to act on the regulation of Ucp1 which is a cold-induced gene in brown adipose tissue, suggesting their roles in response to cold (Angueira et al., 2020). In addition, inhibition of ERRα impairs the regulatory ability of PGC-1α on the expression of genes related to mitochondrial biogenesis such as mtTFA (involved in mitochondrial DNA replication and transcription), Tim22 (involved in protein import into mitochondria), IDH3A (involved in the TCA cycle), carnitine/acylcarnitine translocase (involved in fatty acid oxidation) or Cyt c and ATPsynβ (involved in oxidative phosphorylation) (Schreiber et al., 2004). Importantly, all these metabolic activities of ERRα are strictly PGC-1 dependent.

(3) Cancer

① Relationship between ERRα and EMT

Epithelial mesenchymal transition (EMT) is a process in which epithelial cells lose their cell polarity and cell-cell junctions, and gain invasive and migratory features to transdifferentiate to mesenchymal cells (Lamouille et al., 2014; Mittal, 2018). EMT process is characterized by the loss of epithelial markers such as epithelial cadherin (E-cadherin), increased expression of mesenchymal markers such as vimentin and neural cadherin (N-cadherin) (**Figure 11**) (Eastham et al., 2007; Mittal, 2018). This transition is involved in cancer metastasis and progression, and is usually mediated by transcription factors such as SNAIL, zinc-finger E-box-binding (ZEB) and basic helix-loop-helix transcription factors (Lamouille et al., 2014).

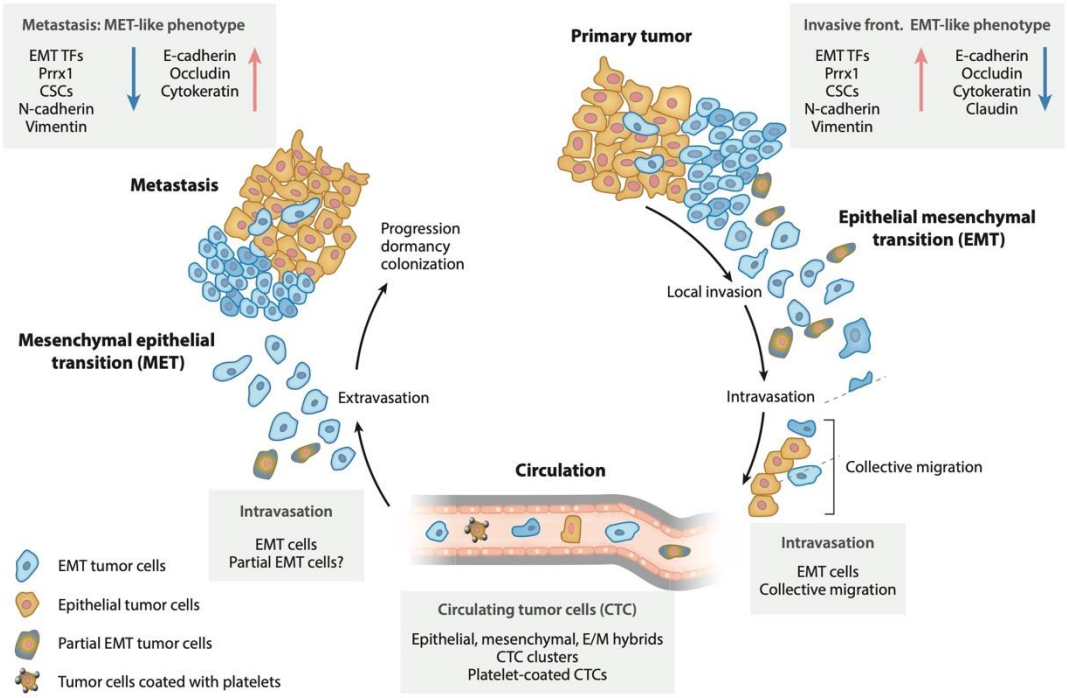


Figure 11. General description of epithelial-mesenchymal plasticity in tumor metastasis from (Mittal, 2018). Cancer cells experiencing EMT (EMT cells) in primary tumors have certain molecular and cellular changes, leading to loss of cell-cell junctions and adhesion that then promotes cell migration and cell invasion. Next, EMT cells enter in blood vessels, migrate and colonize to the distant specific tissues, and are capable of reversing back into an epithelial status through MET, thus achieving the goal of distant metastasis.

ERRα was reported in the induction of EMT in cancer cells and the receptor contributes to cell migration and invasion (Sailland et al., 2014; Carnesecchi et al., 2017; Chen et al., 2017, 2018; Yoriki et al., 2019). Several mechanisms have been summarized and described in the activation of ERRα-induced EMT process (**Figure 12**) (Vanacker and Forcet, 2024). For instance, in some cancer cells, ERRα triggers EMT by regulating the expression of factors such as SNAIL1, SNAIL2, ZEB1 and TGF-β.

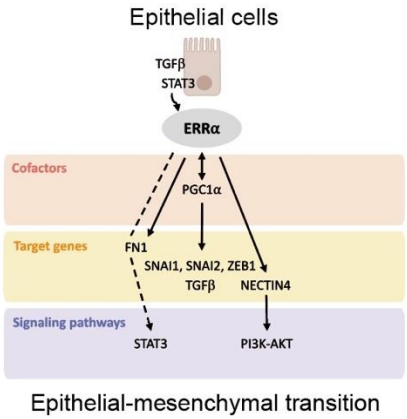


Figure 12. Several molecular mechanisms of EMT are affected by ERRα in cancers. The figure is from (Vanacker and Forcet, 2024).

In TNBC cells, ERRα plays an important role in the promotion of metastasis. For instance, inhibition of ERRα through a treatment with XCT-790 (a potent ERRα inverse agonist) or transfection with siRNA in MDA-MB231 cells (a TNBC subtype) revealed an increased expression of E-Cadherin and a reduced expression of mesenchymal markers (such as fibronectin and vimentin), demonstrating an inhibition of EMT (Wu et al., 2015). In contrast, overexpression of ERRα significantly promotes EMT in these cells. Another case is that MDA-MB231 cells transfected with siERRα showed the inhibited wound closure and in vitro motility, while overexpression of ERRα stimulated the migration and invasion of TNBC cells (Sailland et al., 2014; Wu et al., 2015). In summary, the effect of ERRα on BC cell migration and invasion may depend on its regulation of EMT.

② Cancer metastasis

ERRα also plays a role in cancer metastasis. In BC, ERRα favors bone metastasis which is common in almost 70% of BC patients (Fradet et al., 2011). For instance, overexpression of ERRα has been found to increase BC cell growth in mammary gland as well as to promote the primary tumor invasion in bone (Fradet et al., 2011; Misawa and Inoue, 2015; Tang et al., 2021).

③ Clinical influence of ERRα

Considering the crucial role of ERRα in BC cell migration and metastasis, the clinical relevance of ERRα in these tumors has been evaluated. Previous studies have established ERRα as an unfavorable prognostic predictor in BC and the expression of ERRα was significantly associated with high grade tumors and lymph node metastasis (Ariazi et al., 2002; Wu et al., 2015). In addition, target genes of ERRα were also proved to have a significant role in the prognosis of BC patients. For instance, Deblois et al. (2009) reported 86 ERRα targets as significantly associated with recurrence- or metastasis-free survival at least in two independent datasets (Deblois et al., 2009).

Besides, another study also identified a subset of ERRα targets as a prognostic predictor of BC patients (Chang et al., 2011). These two studies showed that the clinical prognostic influence of ERRα depends on the expression of its transcriptional targets. In other terms, the prognosis value of ERRα in BC may more depend on its activity than on its expression.

The clinical performance of ERRα in other cancers was also investigated. A significant association between increased expression of ERRα with short cancer-specific survival was observed in patients with prostate cancer (Fujimura et al., 2007). In human colorectal tumor, ERRα was found in all patients, as well as a special elevated expression level in tumors than in normal mucosa. Particularly, the expression level of ERRα increased from TNM stages II to IV of colorectal tumors, demonstrating the role of ERRα expression in the colorectal cancer development (Cavallini et al., 2005).

In summary, ERRα has a clinical influence on many cancers through its abnormal expression or a changed expression of its transcriptional targets.

3.4 Molecular mechanisms of ERRα actions

Various molecular mechanisms of ERRα action have been identified. The activities of ERRα on diverse pathological and physiological states can be classified in four manners:

(1) ERRα variations in breast cancer

Gene copy number variants (CNVs), characterized by the duplication or deletion of DNA segment, are a type of structural variants (> 50 bp). CNVs contribute to diverse diseases, including cancers (Zhang et al., 2021). Huang et al. (2021) observed a significant association between CNVs of ESRRA with the histological grade in patients with ovarian cancer (Huang et al., 2021). However, the contribution of ESRRA CNVs in BC patients is less characterized.

To gain insight into these possible relationships, we downloaded copy number alteration (CNA) and mutations from nine BC projects in cBioPortal (https://www.cbioportal.org; download time: 06/November11/2023). Copy number amplification, homozygous deletion and mutation of ESRRA in each BC project is shown in the table below **(Table 1)**. Totally, we observed a number of ESRRA amplifications much higher than homozygous deletion and mutation in BC patients.

Project (9)	Copy number alteration (CNA) (GRCh37)		Mutation (GRCh37)
	Amplification	homozygous deletion	
Breast Invasive Carcinoma (TCGA, Cell 2015, 818 samples)	7	1	/
Breast Invasive Carcinoma (TCGA, Firehose Legacy, 1108 samples)	9	1	/
Breast Invasive Carcinoma (TCGA, Nature 2012, 825 samples)	6	/	/
Breast Invasive Carcinoma (TCGA, PanCancer Atlas, 1084 samples)	5	1	2
Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016, 2509 samples)	20	/	/
Metastatic Breast Cancer (INSERM, PLoS Med 2016, 216 samples)	7	1	1
Proteogenomic landscape of breast cancer (CPTAC, Cell 2020, 122 samples)	3	/	/
The Metastatic Breast Cancer Project (Archived, 2020, 237 samples)	8	2	2
The Metastatic Breast Cancer Project (Provisional, December 2021, 379 samples)	85	25	2

Table 1. Number of variants of ESRRA from nine breast cancer projects in cBioPortal. TCGA and The Metastatic Breast Cancer Project data are used to research in different studies, and thus obtain different CNA annotation results.

We then summarized and compared the distribution of these variants across BC subtypes and ethnic groups. We found that almost all copy number amplifications in ESRRA were abnormal alteration of the entire gene, indicating that ESRRA may had no sensitive amplified gene region **(Table 2)**. We also found that most of the patients with copy number amplification were characterized by ER+ (green in **Table 2**). Only five TNBC samples showed copy number amplification, while they were all collected from metastatic tumor/patients (pink in **Table 2**). This implies that TNBC patients with copy number amplification may be more likely to develop metastasis, but there is not enough data for statistical supporting. Similarly, we found that copy number homozygous deletions of ESRRA were also presented in the entire gene region and most of which appeared in ER+ patients **(**green in **Table 3)**. Two copy number homozygous deletion were found in TNBC samples from metastatic tumor/patients (pink in **Table 3**). Besides, only five ESRRA mutations all from ER+ breast cancer patients (ER+/PR+/ Her2-) were observed and their mutation type, mutation location, alle change, located exon, protein position and domain were significantly distinct **(**green in **Table 4)**. We also noticed that most of the genetic variants were found in white population, but this may be caused by the population bias in data collection.

Project (9)	Copy number alteration (CNA) (GRCh37)										
	Amplification										
	CNA number	Sample								Sample type	CNA region
		Number	ER +	PR +	Her2 +	TNBC	white	black	Asian		
Breast Invasive Carcinoma (TCGA, Cell 2015, 818 samples)	7	7	6	5	4	0	5	1	1	Primary	Whole
Breast Invasive Carcinoma (TCGA, Firehose Legacy, 1108 samples)	9	9	8	7	6	0	6	2	1	Primary	Whole
Breast Invasive Carcinoma (TCGA, Nature 2012, 825 samples)	6	6	4	3	3	0	2	0	2	Primary	Whole
Breast Invasive Carcinoma (TCGA, PanCancer Atlas, 1084 samples)	5	5	4	3	4	0	3	1	1	Primary	Whole
Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016, 2509 samples)	20	20	13	5	7	0	UK and Canada			Primary	unknown
Proteogenomic landscape of breast cancer (CPTAC, Cell 2020, 122 samples)	3	3	2	2	2	0	2	0	1	Primary	unknown
Metastatic Breast Cancer (INSERM, PLoS Med 2016, 216 samples)	7	7	4	4	1	2	France			Metastasis	unknown
The Metastatic Breast Cancer Project (Archived, 2020, 237 samples)	8	8	7	6	2	0	8	0	0	Sample unknown; All patients develop metastatic breast cancer	7 whole, 1 Part
The Metastatic Breast Cancer Project (Provisional, December 2021, 379 samples)	85	78	51	42	13	3	82	1	0	Sample unknown; All patients develop metastatic breast cancer	Whole

*Primary: primary tumors
Whole: Whole ESRRA sequence
Part: part of ESRRA sequence

Table 2. Summary of copy number amplifications of ESRRA from nine breast cancer projects in cBioPortal.

Project (9)	Copy number alteration (CNA) (GRCh37)										
	Homozygous deletion										
	CNA number	Sample								Sample type	CNA region
		Number	ER +	PR +	Her2 +	TNBC	white	black	Asian		
Breast Invasive Carcinoma (TCGA, Cell 2015, 818 samples)	1	1	0	0	1	0	1	0	0	Primary	Whole
Breast Invasive Carcinoma (TCGA, Firehose Legacy, 1108 samples)	1	1	0	0	1	0	1	0	0	Primary	Whole
Breast Invasive Carcinoma (TCGA, Nature 2012, 825 samples)	/										
Breast Invasive Carcinoma (TCGA, PanCancer Atlas, 1084 samples)	1	1	1	0	0	0	1	0	0	Primary	Whole
Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016, 2509 samples)	/										
Proteogenomic landscape of breast cancer (CPTAC, Cell 2020, 122 samples)	/										
Metastatic Breast Cancer (INSERM, PLoS Med 2016, 216 samples)	1	1	0	0	0	1	France			Metastasis	unknown
The Metastatic Breast Cancer Project (Archived, 2020, 237 samples)	2	2	2	2	0	0	2	0	0	Sample unknown; All patients develop metastatic breast cancer	Whole
The Metastatic Breast Cancer Project (Provisional, December 2021, 379 samples)	25	22	20	18	4	1	22	1	0	Sample unknown; All patients develop metastatic breast cancer	24 Whole, 1 Part

*Primary: primary tumors
Whole: Whole ESRRA sequence
Part: part of ESRRA sequence

Table 3. Summary of copy number homozygous deletions of ESRRA from nine breast cancer projects in cBioPortal.

Project (9)	Mutation										
	Mutation description (GRCh37)						Protein		Patient		
	Number	Sample type	Location	Mutation type	Alle change	Exon	Protein position	Domain: DBD/LBD (UniProt)	ER	PR	Her2
Breast Invasive Carcinoma (TCGA, PanCancer Atlas, 1084 samples)	2	Primary	64082382	splice_region_variant,synonymous_variant	A: A/T	5	247	LBD	+	+	-
		Primary	64074726	Frame_Shift_Del	T: T/-	2	25	/	+	+	-
Metastatic Breast Cancer (INSERM, PLoS Med 2016, 216 samples)	1	Metastasis	64074712	missense_variant	G: G/A	2	21	/	+	+	-
The Metastatic Breast Cancer Project (Provisional, December 2021, 379 samples)	2	Sample unknown; All patients develop metastatic breast cancer	64081491	missense_variant	C: C/G	3	132	DBD	+	+	-
		Sample unknown; All patients develop metastatic breast cancer	64082342	missense_variant	G: G/A	5	234	LBD	NA	NA	NA

*Primary: primary tumors
Whole: Whole ESRRA sequence
Part: part of ESRRA sequence

Table 4. Summary of mutations of ESRRA from nine breast cancer projects in cBioPortal.

In summary, most of ESRRA variations occur in ER+ BC patients. They are often abnormal alterations of the entire gene region which indicate that ESRRA has no specific altered gene region. However, few ESRRA variations were observed in TNBC patients. We hypothesize that this may be because the number of TNBC samples is limited and that ERRα may function mainly in the form of non-self variation in TNBC, such as through a cooperation with co-regulators on transcriptional effects. Therefore, it is an important reason for us to explore the transcriptional partners of ERRα in breast cancer, in particular in TNBC.

(2) Interactions with co-activators

As an orphan nuclear receptor, ERRα performs regulatory function in a ligand-independent and CoR-dependent way. The binding motifs of ERRα, namely ERR-response elements (ERRE), usually reside in a distance of > 1kb upstream of the transcriptional start site (TSS) of targets (Cerutti et al., 2023) (**Figure 13**). Other TFs which bind to the region surrounding the TSS can be linked to the molecular complex formed by ERRα and the recruited cofactors in order to promote the regulatory effect of ERRα. Therefore, CoRs are helpful for ERRα to understand its molecular mechanism in human diseases.

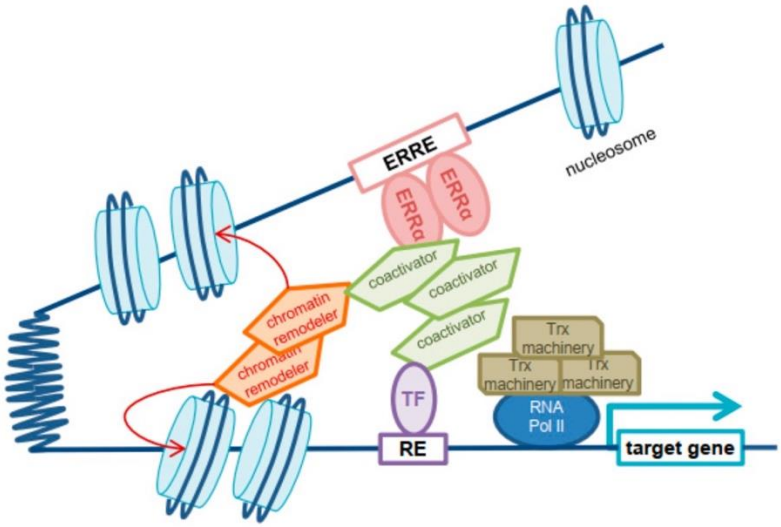


Figure 13. Schematic model presenting the regulation of ERRα on target genes in a distant way from (Cerutti et al., 2023). ERRα binds to the open chromatin region and recruits cofactors. The interaction between other TFs recruited to the TSS and the ERRα-cofactors complex promotes the transcription of target genes. TF: transcription factor; RE: response element; Trx: transcription.

Previous studies have proven the co-activating role of PGC-1α on ERRα, performing the cooperation function in energy metabolism (Ariazi and Jordan, 2006; Stein and McDonnell, 2006). In addition, an important role of the ERRα-PGC-1α complex has also been described in human Alzheimer's disease (AD) (Sato et al., 2023). AD is a progressive neurodegenerative disease affecting memory and cognitive ability and is characterized by senile plaques that result from insoluble deposits of amyloid β peptide (Aβ) and neurofibrillary tangles NFTs which are aggregates of hyperphosphorylated tau protein in neurons (Markesbery, 1997; Sato et al., 2023). Both of them cause neuronal cell death in AD. ERRα may works with PGC-1α to repress the expression of β-site amyloid precursor protein cleaving enzyme 1 (BACE1) which produces Aβ in neurons, as well as reduces the phosphorylation of glycogen synthase kinase 3β (GSK-3β) and phosphorylated tau levels (**Figure 14**) (Cai et al., 2001; Tang et al., 2018; Sato et al., 2023). Both RNA and protein expression of PGC-1α are decreased in AD brain (Qin et al., 2009), which may influence the accumulation of Aβ and hyperphosphorylated tau protein, indicating that ERRα and PGC-1α play a role in AD progression. In summary, PGC-1α acts as co-activator of ERRα in various pathological processes, including energy metabolism and human Alzheimer's disease.

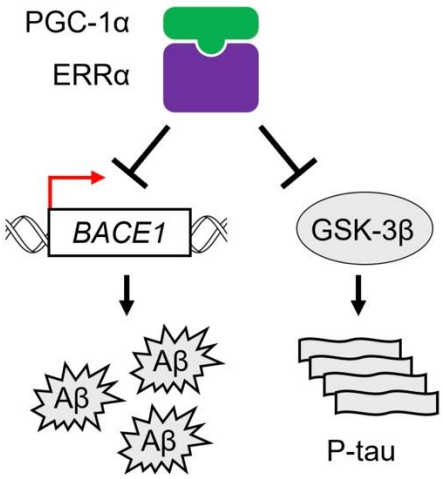


Figure 14. Role of ERRα and PGC-1α in the pathogenesis of Alzheimer's disease (AD) from (Sato et al., 2023). ERRα may work with PGC-1α to inhibit the expression of BACE1 and the kinase activity of GSK-3β, thus controlling the amount of Aβ and hyperphosphorylated tau protein which are pathological features of AD.

The three members of the p160 family of steroid receptor co-activators (SRC-1, SRC-2, and SRC-3) are crucial in human BC response to endocrine therapy (Saatci et al., 2021). They have been identified as cooperating with many NRs, including ERRα (Hong et al., 1999; Xie et al., 1999; Xu and Li, 2003; van der Laan et al., 2014). For instance, the simultaneously binding of ERRα and AIB1 (also known as SRC-3) to the promoter region on ERRα-target pS2 was detected by ChIPs and a positive correlation between ERRα and AIB1 was verified in both mRNA and protein level in breast tumors, suggesting an important role of AIB1 acting as an ERRα co-activator in BC (Heck et al., 2009).

(3) Interactions with co-repressors

The transcriptional activity of ERRα is also negatively regulated by co-repressors, such as the nuclear receptor interacting protein 1 (NRIP1, also known as RIP140) and the nuclear receptor corepressor 1 (NCoR1) (Castet et al., 2006; Misawa and Inoue, 2015; Cerutti et al., 2023).

NRIP1 is a nuclear protein which is highly expressed in adipose tissue, liver, heart and muscle and involved in energy expenditure, glucose metabolism and lipid metabolism (Christian et al., 2006; Fritah et al., 2010; Yi et al., 2017). NRIP1 was originally found in BC cell lines and specifically interacted with the AF-2 domain of nuclear receptors. It was identified as a factor which mediates down-regulation of nuclear receptors activity as well as acts a competition factor with co-activators such as SRC-1 (Treuter et al., 1998; Cerutti et al., 2023). Up to now, NRIP1 has been proven to repress a set of nuclear receptors such as the PPARs (α, β/δ and γ), TRα and TRβ, and ERRα and β mostly in metabolic tissues (Fritah et al., 2010). The transcription activity of ERRα is inhibited by NRIP1 based on their physical interaction. However, activation

was also observed in an ERR isoform- and DNA context-specific manners (Sanyal et al., 2004; Giguère, 2008). Altogether, this hypothesizes that NRIP1 could act as co-activator or co-repressor of ERRs.

The nuclear receptor corepressor 1 (NCoR1) has been reported as a co-repressor of thyroid hormone receptors and retinoic acid receptors (Hörlein et al., 1995). NCoR1 usually regulates gene expression in a form of complex constructed with some core subunits including G protein pathway suppressor 2 (GPS2), transducin β-like 1 (TBL1), TBL-related 1 (TBLR1) and histone deacetylase 3 (HDAC3) (Pérez-Schindler et al., 2012). NCoR1 is expressed in various tissues. Particularly in skeletal muscle, experiments showed NCoR1 muscle-specific knockout mice exhibited higher oxygen consumption, decreased maximal isometric force and increased fatigue resistance, demonstrating the critical role of NCoR1 in metabolic pathways (Pérez-Schindler et al., 2012). PGC-1α is also a crucial metabolic factor as well as a co-activator of ERRα in metabolism, as pointed above. Interestingly, Pérez-Schindler et al. (2012) identified ERRα as a common transcription factor for both NCoR1 and PGC-1α (Pérez-Schindler et al., 2012). NCoR1 and PGC-1α have opposing effects on ERRα transcriptional activity in the regulation of skeletal muscle function and oxidative metabolism (**Figure 15**). Altogether, NCoR1 acts as a transcription repressor of ERRα in metabolic phenomena.

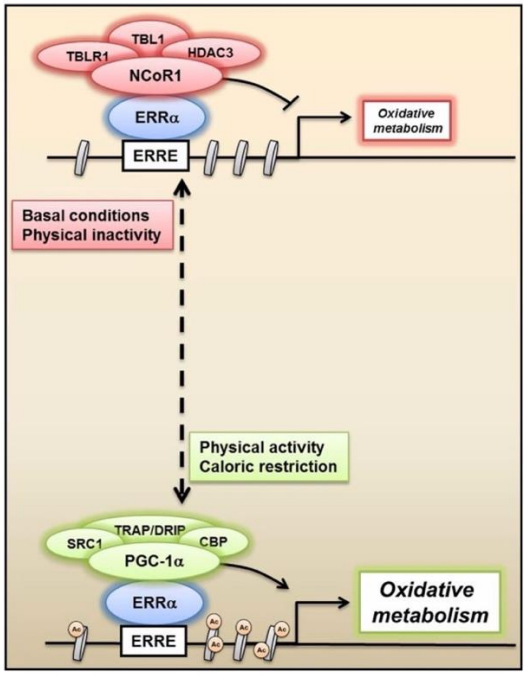


Figure 15. Antagonistic regulation of NCoR1 and PGC-1α on ERRα in oxidative metabolism from (Pérez-Schindler et al., 2012). NCoR1 represses the activities of ERRα on metabolic genes under basal condition. Under caloric restriction, PGC-1α acts as an ERRα co-activator on metabolic genes. Ac: histone acetylation.

(4) Post-translational modifications

Post-translational modifications also affect the transcriptional activity of ERRα. P300/CBP-associated factor (PCAF), also known as K(lysine) acetyltransferase 2B (KAT2B), acts as a co-regulator for various TFs (Ogryzko et al., 1996; Shikama et al., 2000). Wilson et al. (2010) observed that PCAF interacts and acetylates ERRα in vitro and in mouse liver (Wilson et al., 2010). Four conserved lysines within the DNA-binding domain (DBD) of ERRα are targeted by this acetylation, leading to a repressed activity of the receptor. In contrast, histone deacetylase 8 (HDAC8) and sirtuin 1 (SIRT1), both acting as co-activators, deacetylate ERRα and enhance its DNA binding activity. These results suggested the acetylation status of ERRα also has an impact on DNA binding and transcription activity.

3.5 Conclusion

ERRα gradually gained attention in the search of novel pathological mechanism and treatment strategies in BC. This is particularly the case in TNBC where the lack of markers ER, PR and Her2 complicates the access to conventional treatments. ERRα acts in cooperation with co-regulators in high energy-demanding tissues, such as muscle and liver (**Figure 7**). These activities appear dependent on the PGC-1α co-activator in mitochondrial biogenesis (Schreiber et al., 2004) and regulation of the expression of metabolic genes such as MCAD, IDH3A and ATP5B (Giguère, 2008; Villena and Kralli, 2008).

In addition, ERRα participates in other biological processes, such as Alzheimer's disease, cancer progression, including in BC through diverse molecular mechanisms. Importantly, CoRs-dependent regulations appear critical in these regulations. Moreover, the clinical significance of ERRα and, in particular, of its transcriptional targets in the prognosis of BC patients were also reported (Ariazi et al., 2002; Deblois et al., 2009; Chang et al., 2011). This highly suggests an important role of ERRα in BC. However, we found that ERRα does not regulate metabolic genes such as MCAD, IDH3A and ATP5B in BC cells (Sailland et al., 2014; Cerutti et al., 2022). We also noticed ERRα regulate BC cell progression in a PGC-1α-independent manner, suggesting the recruitment of other CoRs of ERRα in BC development. Therefore, the identification of novel CoRs of ERRα related to BC progression as well as the transcriptional targets of the ERRα-CoRs complex will be important to extend the understanding of ERRα and the therapy of BC, particularly TNBC.

Chapter IV– ZEB1

Zinc finger E-box binding homeobox 1 (ZEB1) is a member of zinc finger-homeodomain transcription factor family (Wu et al., 2020a) and is a crucial transcription factor in EMT and cancer progression. Abnormal expression of ZEB1 is associated with mesenchymal traits in tumor cells, as well as metastasis, multidrug resistance, proliferation and poor prognosis in various cancer types, including pancreatic cancer, liver disease, lung cancer, colon cancer and BC (Lehmann et al., 2016; Wu et al., 2020a).

ZEB1 has been shown as a transcriptional repressor interacting with the co-repressor CtBP to repress transcription (Postigo and Dean, 1999). It acts as a repressor of E-cadherin expression, a process which promotes EMT (Caramel et al., 2018). Besides, ZEB1 can also repress the expression of microRNA-200 family members (miR-200a, miR-200b, miR-200c, and miR-141) which, in turn, downregulate ZEB1 expression. This microRNA-mediated feedforward loop triggered by ZEB1 is able to stabilize EMT and promote cancer cell invasion (Burk et al., 2008).In addition to its role as a transcriptional repressor, ZEB1 also acts as a transcriptional activator. For instance, Lehmann et al. (2016) found an interaction between ZEB1 and the Hippo pathway factor YAP, which switches ZEB1 to an activator of a common ZEB1/YAP target gene set (AXL, LHFP, SERPINE1, SLIT2, DAB2, FSTL1, THBS1, CTGF) (Lehmann et al., 2016). This target gene set is a predictor of poor survival, treatment resistance and increased metastatic risk in BC, showing the impact of ZEB1 in cancer-promoting process. In 2020, Feldker et al. identified the AP-1 factors FOSL1 and JUN as new interactors of ZEB1, leading to activate the expression of tumor promoting genes. Especially, high expression of ZEB1, YAP, FOSL1 and JUN marks the most aggressive subtype of BC (claudin-low subtype). Altogether, this suggests a dual function of ZEB1 that depends on its cooperation with different transcription factor complexes and is involved in cancer progression, including in BC.

In our following analysis, ZEB1 is identified and validated as a co-activator of ERRα specifically in TNBC subtype, as well as targeting 8 common cell migration related ERRα/ZEB1 genes (DEGs). Our results show ZEB1 and the 8 DEGs have an association with EMT process, even poor overall survival in TNBC. This is in line with the hypothesis above, implying a co-activator role of ZEB1 in BC progression.

Chapter V – Co-regulators prediction in cancer

5.1 Introduction of co-regulators prediction methods

CoRs include DNA-binding TFs and non-DNA binding transcription regulators (TRs). Several high-throughput biochemical technologies exist that detect TF-DNA (TF-targets) interactions and then explore potential CoRs of TFs.

Biological experiments have been the most accurate methods to explore CoR of TFs. High-throughput in vitro technologies include systematic evolution of ligands by exponential enrichment (SELEX) (Jolma et al., 2010) and protein binding-microarrays (PBM) (Berger et al., 2006). High-throughput in vivo methods include chromatin immunoprecipitation-based methods such as ChIP-seq (Johnson et al., 2007), ChIP-exo (Franklin Pugh, 2012) and ChIP-nexus (He et al., 2015), as well as cleavage-based methods (Skene and Henikoff, 2017). However, many possible binding combinations of TFs increase the research complexity and these methods cannot perform on all TFs with every known CoRs in every type of cells or diverse biological conditions (Knock down/out, overexpressed, cancers…) (Rauluseviciute et al., 2024). Some approaches are limited to the investigation of pairwise interactions of 2 TRs, such as yeast two hybrid method (Francois et al., 2020). However, more than 2 proteins are possible to be simultaneously recruited and the relationship between them may be hard to detect by using this method. Therefore, with the development of computation technology and source, multiple CoRs can to be identified in the large-scale screening on numerous published data.

The regulation of TFs activities by CoRs on target genes depends on the direct binding on the genomic region near the TSS (referred to as “pre-defined region). Previous studies have identified regulator pairs for the queried gene through identifying the direct binding sites of TFs on pre-defined region from the TSS. For instance, the approach used in JASPAR web tool predicts TFs and the TF binding sites (TFBSs) for the queried target gene based on position weight matrix (PWM) (Castro-Mondragon et al., 2022). PWM reflects the occurrences of each type of nucleotide at each position which was summarized from experimentally determined DNA sequences bound by an individual TF (**Figure 16**) (Castro-Mondragon et al., 2022). Computational scanning on the DNA sequence of the queried gene (target) based on PWM is performed and scores the potential TFBSs (Stormo, 2013). A set of TFs with TFBSs predicted on the pre-defined region of the queried gene can be thought as or be used to predict CoRs (Castro-Mondragon et al., 2022). However, PWM is only a probabilistic matrix of each binding region, thus it may influence the predicted result because it is not the exact binding region (Stormo, 2013).

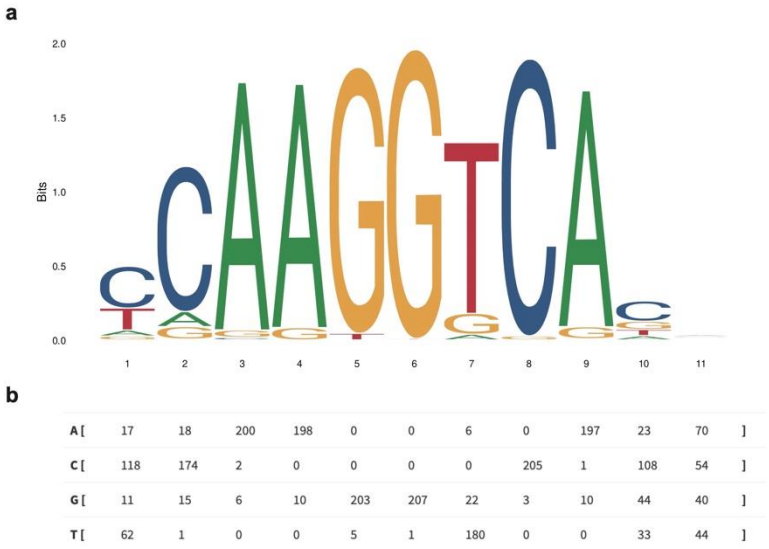


Figure 16. The integrated binding motif profile of ESRRA from JASPAR database (matrix ID: MA0592.1) (Castro-Mondragon et al., 2022). a. The sequence logo of the binding profile of ESRRA showing the distribution of nucleotides at each position based on b. the frequency matrix of ESRRA motif showing the frequency of each nucleotide at each position of the motif. The sequence logo and the frequency matrix used were downloaded from the JASPAR database on 09/2024 (https://jaspar.elixir.no/matrix/MA0592.1/).

Another study avoids this problem. The approach used in hTFtarget database predicts the candidate co-regulation of TFs for the queried genes based on the experimentally identified binding sites (Zhang et al., 2020). They first define a core region for the queried gene as 50kb region upstream from the TSS. TFs with high-confident peaks (from curated ChIP-seq datasets in hTFtarget) and each peak meets the screening criteria (beta-model score of each peak ≥ 0.517) within this core region will be identified as the potential co-regulators for this gene.

These two methods are based on TFBSs of TFs (referred to as "TFBSs-based approach"). However, they have some shortcomings, such as (1) the combination of TFs and CoRs usually regulate a list of targets, thus it is hard to combine and compare the predicted TFs from TFBSs-based approaches among all target genes; (2) TFBSs-based approaches are mainly used to find TF-TF co-binding pairs, do not include the relations between TF and non-DNA binding TRs, leading to miss CoRs prediction; (3) the length of pre-defined region on DNA sequence of target genes may influence the predicted result, leading to more unrelated TFs based on a longer length whereas correlated TFs based on a shorter length may be missed; (4) if a TR has never been or less studied in the ChIP-seq or in ChIP-seq under a specific condition (i.e., cancer type), then it will not be predicted by TFBSs-based approaches; (5) the binding motif for a TF sometimes cannot reveal a regulatory function on target genes enough because the regulation can also be influenced by other properties of the motif such as location,

orientation and methylation (Gardner and Faith, 2005). Therefore, TFBSs-based approaches are limited to predict novel potential co-regulation of TFs with CoRs.

From above, TFBSs-based approaches only collect and integrate large amounts of ChIP-seq datasets, thus it is hard to ensure the collection of all information. Text-mining methods have been developed to solve the information integration problem. They extract information from public texts such as biomedical articles, that do not only originate from ChIP-seq experiments. Some approaches collected and predicted TF-target interactions and TF cooperativity within a TF module (TFs grouped based on defined rules) based on the information of public literatures, such as the approach used in TRRUST. For instance, TRRUST identified human TF-target interactions based on a sentence-based text-mining method from 20 million abstracts of biomedical and life science articles (Han et al., 2015, 2018a). Sentences were extracted based on a TF name and an additional gene name and then manually curated to obtain reliable TF-targets interactions with the mode of regulation (i.e., activate or repress) (**Figure 17**). Then they integrated these TF-targets interactions and TF-TF interactions (Protein-protein interactions (PPI) collected from other databases) in order to measure TF cooperativity by applying a method inspired from the genome-wide functional network approach (Lee et al., 2011; Han et al., 2015). In the network (**Figure 18**), a TF module for a single target gene includes TFs which commonly regulate this gene. TFs tightly connecting with each other in a module based on PPI are assumed to present cooperation, suggesting the role as partners for each other on the transcription of the same target gene. In summary, TRRUST integrates and curates TF-targets relations from literature using sentence-based text-mining method. Based on TF-target interactions and PPI, they assume TF cooperativity among TFs of the target gene, predicting potential co-regulators.

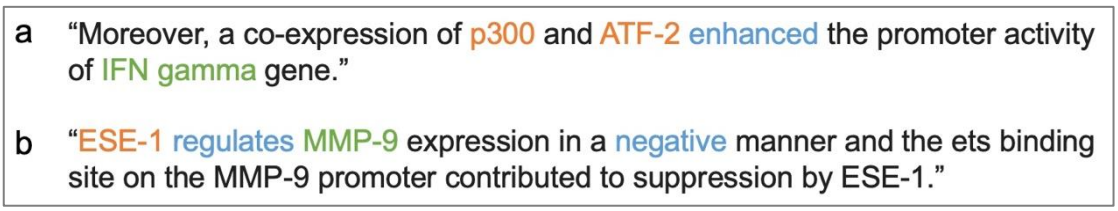


Figure 17. Examples describing the mode of regulation for the TF-target regulatory interactions. a. An example sentence describes the activation mode (blue word) for a TF (orange word) - target (green word) interaction. b. An example sentence describes the repression mode (blue word) for a TF (orange word) - target (green word) interaction (Han et al., 2015).

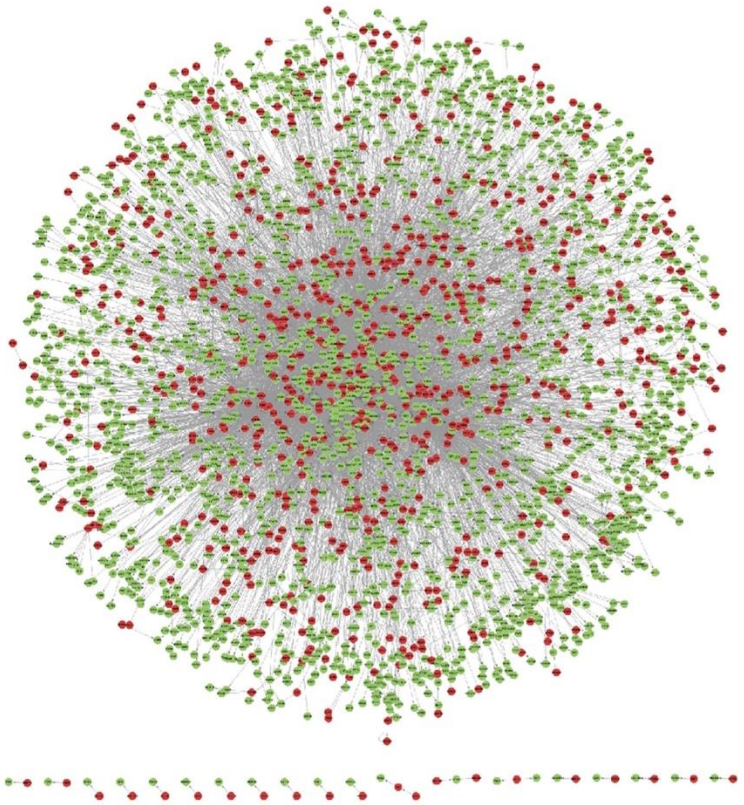


Figure 18. A transcriptional regulatory network (TRN) from (Han et al., 2015) including identified TF-target interactions in TRRUST. TFs (red nodes) and non-TF genes (green nodes) based on the regulatory interactions are shown.

However, there also show some disadvantages of text-mining method on the CoRs prediction, such as (1) TF-target relations extracted from literatures are only based on the co-occurrence of the name of TF and name of other genes in one sentence (**Figure 17**) and manual curation. However, the co-occurrence of two names in one sentence does not always reflect a precise relationship between them, such as the regulation mode because people do not always use the exact word (activate/repress) to describe the regulation mode. If the relationship was controversial or even incorrect since they were curated by researchers themselves, then the predicted results may also be unreliable. In addition, the regulation may be clearly described in several sentences or in a paragraph, not only in one sentence, thus the focus on the co-occurrence in one sentence may miss some information; (2) since these data were from published literature, it is not applicable to predict CoRs for less-studied TFs, leading to incomplete or unreliable prediction results.

Noticeably, the idea of network mentioned in the approach of TRRUST database is useful to connect factors and has also been widely used to explore the transcription control relationship between TFs, CoRs and targets, such as gene regulatory networks (GRNs) (Guan et al., 2022) and transcription regulatory network (TRN) (He and Tan, 2016) (**Figure 18**). For instance, TRN usually includes regulatory interactions between

TFs and targets in which direct interactions were represented by the edges. Considering the importance of gene expression data in the investigation of transcription activity, several computational methods are used to construct TRN models using genome-scale data, such as linear regression, statistical correlation, and Bayesian network (call as "regression-based approaches"). These regression-based approaches assume that TFs directly regulating the expression of the target are the most important and informative factors among all TFs, suggesting their expression level are responsible for predicting the expression of the target. Regression-based approaches aim to explore the statistical dependencies between the TFs and the target gene based on the constructed regression model which demonstrates the expression of the target is regressed on the expression of TFs. However, an important point needs to be considered when using regression-based approaches. A large number of TFs are often considered and used to construct the regression model for the target. However, the high number of parameters (e.g., the TFs) may bring much noise and problems of overfitting. Indeed, a model not only fits the potential relationship between variables (Here the TFs and the targets), but also fits the noise of each sample from biological or technical problems, and finally influence the future prediction results (**Figure 19**) (Lever et al., 2016). Thus, feature selection and dimensionality reduction should be considered before regression modeling.

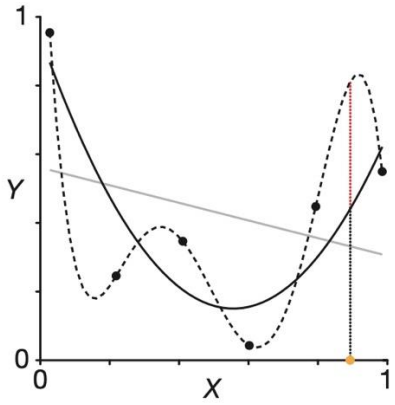


Figure 19. An example of the fitted models from (Lever et al., 2016). Exemplify underfitting (gray diagonal line), reasonable fitting (black curve) and overfitting (dashed curve) are presented on the same data. A huge difference (red dotted line) in Y prediction when X = 0.9 between the reasonable model and the overfitted model was observed.

In summary, diverse approaches have been applied to investigate the transcriptional regulation exerted by TFs on the target genes and next to predict CoRs for TFs. These include TFBSs-based approaches, text-mining method and regression-based approaches. However, all of them have both of advantages and disadvantages. Choosing the suitable method is important for answering a specific biological question.

5.2 PLS method introduction

Different approaches have been used to predict co-regulators for TFs. In particular, some studies also established a database to display the collected data, TFs-targets relations and co-regulation of TFs, based on the large scale of screening on published data. These studies include TRRUST (Han et al., 2018a), ChEA3 (Keenan et al., 2019), Cistrome DB (Zheng et al., 2019), TRANSFAC (Wingender et al., 1996), Harmonizonme (Rouillard et al., 2016), and hTFtarget (Zhang et al., 2020). However, in these databases, information for ERRα is limited, leading to few predicted co-regulated partners. More importantly, they cannot provide information under specific conditions we are interested in, such as CoRs of ERRα in cell migration pathway of breast cancer. Thus, a single study for the exploration of CoRs of ERRα and the further investigation of their specific pathological and physiological functions in breast cancer should be conducted using specific genomic data, instead of integration of published data.

From above, regression-based approaches can be used to better understand and explore the regulation of TFs and CoRs on the target genes based on condition-specific genomic data. They construct a regression model for the expression of target through the expression level of a list of TFs, and a non-zero regression coefficient will indicate the statistical dependency representing the regulation of TFs to the targets (He and Tan, 2016).

However, the high-dimensionality of genomic data renders computation for model construction complex. In addition, model overfitting may also raise up if trying to use a large number of TFs. To solve the problem of high-dimensionality and avoid the model overfitting, partial least squares (PLS) regression method has recently gained much attention (Chun and Keleş, 2010).

PLS aims at handling multicollinearity problems, i.e. the influence of highly correlated covariates (also called regressors or predictors) (i.e., TFs) on the response (also called observation or observed outcome) (i.e., target genes) (Durif, 2016). PLS is a compression method (dimension reduction of data) which assumes the regression relationship can be explained by a few numbers of latent variables presented in a form of n-dimensional vectors. The principle is that the latent variable (component) is a linear combination of covariates (i.e., TFs) (**Figure 20**) and its covariance with the response (i.e., the target gene) is maximum. The covariance of two variables reflects their joint variability, or the degree of association. If the association between variable X (x_1, x_2, …) and Y (x_1, y_2, …) is positive, that will denote that when variable X is larger than its mean, then Y tends to be larger than its mean. Covariance will thus be positive. If the association between variable X (x_1, x_2, …) and Y (x_1, y_2, …) is negative, that will denote that when variable X is larger than its mean, Y tends to be smaller than its mean. Their covariance will thus be negative. Here the covariance between the new latent variable with the response can be simply explained as their association (Rice, 2007).

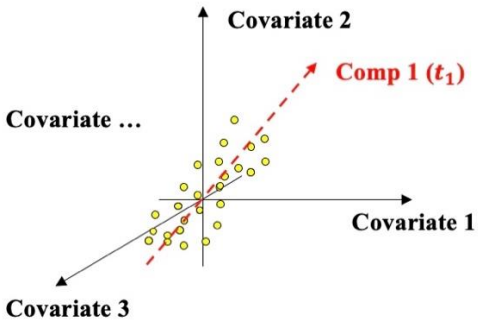


Figure 20. A simple representation of the linear relationship between the new latent component (comp1) with covariates (i.e., TFs).

Thus, the core of PLS is to establish new components (latent variables). The new component matrix $T\,(t_1,\ t_2,\ \ldots,\ t_n)$ can be explained as:

$$T = XW$$

where X is the predictor matrix, and $W\,(w_1,\ w_2,\ \ldots,\ w_n)$ is the weight vector matrix in which each weight vector w_n is defined to maximize the covariance between the new component t_n and the response (Chun and Keleş, 2010; Durif, 2016). The weight vectors also highlight the directions that explain the response (Durif, 2016).

After constructing the new components, the regression coefficients of covariates are calculated based on the matrix W and the covariates are able to regress to the explanation of the response (Boulesteix and Strimmer, 2007), which can be biologically applied to find the regulation of TFs on the expression of the target gene. The regression relation can be simplified described by the equation:

$$\begin{aligned}\text{expression}_{\text{Gene A}} = {} & \text{coefficient}_1 * \text{expression}_{\text{TR1}} + \\ & \text{coefficient}_2 * \text{expression}_{\text{TR2}} + \\ & \text{coefficient}_3 * \text{expression}_{\text{TR3}} + \\ & \ldots\end{aligned}$$

5.3 Sparse PLS method introduction

However, PLS uses all covariates to generate latent variables and to explain the response. For instance, in gene transcription, not all TFs have the ability to contribute to the expression of the target gene. In most situations, only a limited number of TFs are important factors to influence the transcription activity of the target regarding the different conditions, such as tissues. Particularly, in high-dimensional data like gene expression data, the contribution of accumulative noise from the large number of unrelated covariates to the response may exceed the contribution of a small set of

important covariates (Durif, 2016). Therefore, the selection of the most informative and important factors at the beginning may reduce the noise in high-dimensional data and improve the regression accuracy. In addition, modeling in genomic data faces the problem that the number of sample size is much smaller than the number of variables (Chun and Keleş, 2010). Therefore, sparse PLS (sPLS) method has been developed by introducing a selection step based on Lasso (least absolute shrinkage and selection operator) principle to PLS, thus combining both of compression and variable selection functions (Lê Cao et al., 2008).

sPLS works on similar modeling steps to PLS, whereas in the component construction progress, the l_1 -norm penalty (Lasso) was used in covariance maximization problem (Durif et al., 2018). This type of penalization allows the coefficient of less relevant covariates shrunk to zero and discard these irrelevant covariates. Therefore, the selected relevant covariates next generate the new components which act as the sparse linear combination of covariates, mapping the high-dimensional data into a lower dimension space (Durif et al., 2018). sPLS has been proved as showing a high prediction and accuracy in obtaining the relevant variables (Chun and Keleş, 2010). Moreover, to improve the selection accuracy, an adaptive version of sPLS (adaptive sPLS) was developed based on the optimization on sPLS inspiring from the adaptive Lasso (Durif et al., 2018).

5.4 Adaptive sparse PLS method introduction

The adaptive sPLS has been shown to display a stability of variable selection and an increased prediction accuracy. Durif et al. (2018) established the adaptive sPLS framework (Durif et al., 2018). In this work, they first evaluate the cross-validation stability of the adaptive sPLS method by comparing to other sPLS-based methods such as SPLS-log or SGPLS, because model quality always relies on the choice of hyper-parameters tuning by cross-validation and the chosen value should be stable under multiple running times on the same sample. The results in the adaptive sPLS are more stable than other two methods, showing the adaptive selection improves the cross-validation stability. Next, they also apply the adaptive sPLS method for logistic regression to construct a classification framework, called "logit-SPLS-ad" (the adaptive version of logit-SPLS). Then they test the performance of logit-SPLS-ad method in the classification of relapse outcome of 294 breast tumors. Results show a highest stability of the sparsity parameter selection and the highest precision of this parameter value in logit-SPLS-ad, by comparing to other methods: GLMNET, logit-PLS, logit-SPLS (not adaptive), SGPLS (sPLS-based) and SPLS-log (sPLS-based). Logit-SPLS-ad also gives better prediction results on the classification of breast tumors compared to other approaches. Altogether, this suggests that the adaptive sPLS has a good performance on hyper-parameter tuning and the adaptive selection improves the prediction accuracy. The adaptive sPLS can thus be used in breast tumor data (Durif et al., 2018). Recently, it has also been reported in CoR prediction in breast cancer cellular data (Cerutti et al., 2022) (**Figure 21**).

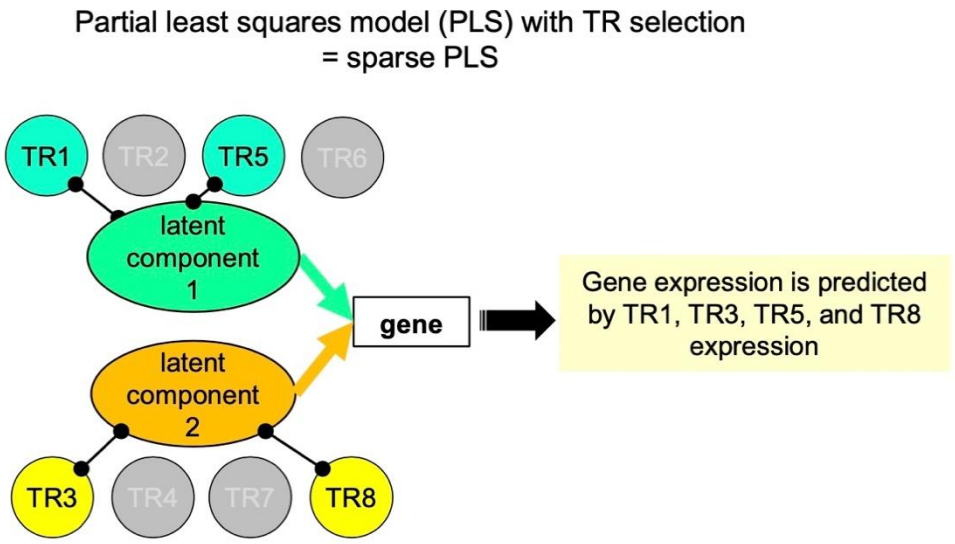


Figure 21. The general principle of the application of sparse partial least square (sPLS) regression algorithm to explore the regulation of transcriptional regulators (TRs) on the target gene from (Cerutti et al., 2022). By sPLS modeling algorithm, new latent components are generated as the linear combination of selected TRs and then are used to explain the expression of the target gene.

5.5. Conclusion

In summary, many biological technologies and computation approaches have been developed and validated in CoRs investigation. Genomic analyses, such as gene expression data, are helpful for exploring the relations between TFs (such as ERRα) and CoRs. In biology, regression-based approaches aim to find the statistical dependencies between the expression of TFs/CoRs with the expression of the target genes, suggesting the contribution of TFs/CoRs to the expression of the target and then predict CoRs. However, (1) high-dimensional expression data always cause huge amounts of noise in results. (2) In addition, it is hard to figure out the most contributing CoRs to the expression of the target gene from hundreds of CoRs since some unrelated TRs/CoRs may develop much false positive relationship in regression. (3) Besides, there also exist interactions between TFs/CoRs, which may further influence the regression model of the gene. Therefore, dimension reduction for data and feature selection for large number of TRs/CoRs at the beginning step are necessary to improve the precision and accuracy of the results.

sPLS regression method has a high performance on dealing with these problems as well as predicting CoRs based on expression data. The adaptive sPLS, which is the optimized version of sPLS improves the prediction accuracy in a real-word case of BC cohort (Durif et al., 2018). Altogether, the adaptive version of sPLS can be well used to investigate the statistical dependencies between covariates and the response, which can be applied to solve the biological problem: explore the regulation of TFs/CoRs on the

expression of the target gene and then predict CoRs of ERRα in BC.

Aim of my thesis

BC is currently the most common cancer of women globally. Many studies have explored the pathology of BC and developed proper treatments, such as drugs, chemotherapy, radiotherapy and neoadjuvant therapy. However, due to the tumor intertumoral and intratumoral heterogeneity, it is difficult to provide an accurate therapeutic advice to many patients. TNBC, the most aggressive BC subtype often occurring in young women, usually develop worse outcome. TNBC still lack efficient strategies to treat and prolong the survival of patients given the deficiency of ER, PR and Her2 markers. Therefore, novel biological molecules are needed to be identified and researched to explore new pathological mechanisms of BC, as well as to identify potential therapeutic targets.

Nuclear receptors (NRs) form a family of transcriptional factors (TFs) that directly regulate the expression of target genes by binding to specific response element on the DNA sequence. NRs mostly function in various pathological and physiological activities, such as BC, through the recruitment of specific co-regulators (CoRs). Thus, the cooperation of NRs and CoRs can help to investigate in BC. However, TNBC subtype lacks ER, PR and Her2 expressions, suggesting that the exploration of other NRs with their CoRs may be a highly promising approach. Interestingly, ERRα has gained attention in BC because its high sequence homology to ER, suggesting similar functions in BC (Misawa and Inoue, 2015). Moreover, ERRα can be identified in each BC subtype, including TNBC. Altogether, this suggests that ERRα could be used as a target in BC.

Several factors have already been identified as interacting with ERRα and modulating its cancer-promoting functions, such as NCOA3 (AIB1) (Heck et al., 2009), NCOA2 (SRC-2) (Liu et al., 2009), KDM1A (LSD1) (Carnesecchi et al., 2017) or NRIP1 (RIP140) (Castet et al., 2006). However, there is a lack of a systematic approach to screen and select potential CoRs of ERRα from a large amount of CoRs. To fill this research gap, a recent work by our lab has used the adaptive spare partial least squares (sPLS) regression method with improved prediction accuracy to explore the co-regulation of ERRα and CoRs on the expression of targets genes in genomic data of BC cell lines (Cerutti et al., 2022). However, this study only focused on BC cells, whereas potential CoRs of ERRα in human tumors are still unknown.

My thesis aims to identify novel co-regulators of ERRα, as well as to investigate their transcription activity on the commonly regulated target genes during BC progression. My work will be crucial to reinforce the understanding of the biological functions of ERRα in BC and could help to set a novel therapeutic strategy in BC, in particular in TNBC.

General introduction of thesis

ERRα, an orphan nuclear receptor, participates in many cancers' progression in a coregulator-dependent manner. My thesis aimed to explore the transcriptional activity of ERRα during breast cancer progression as well as the potential pathological and physiological pathways it may be involved in (**Figure 22**).

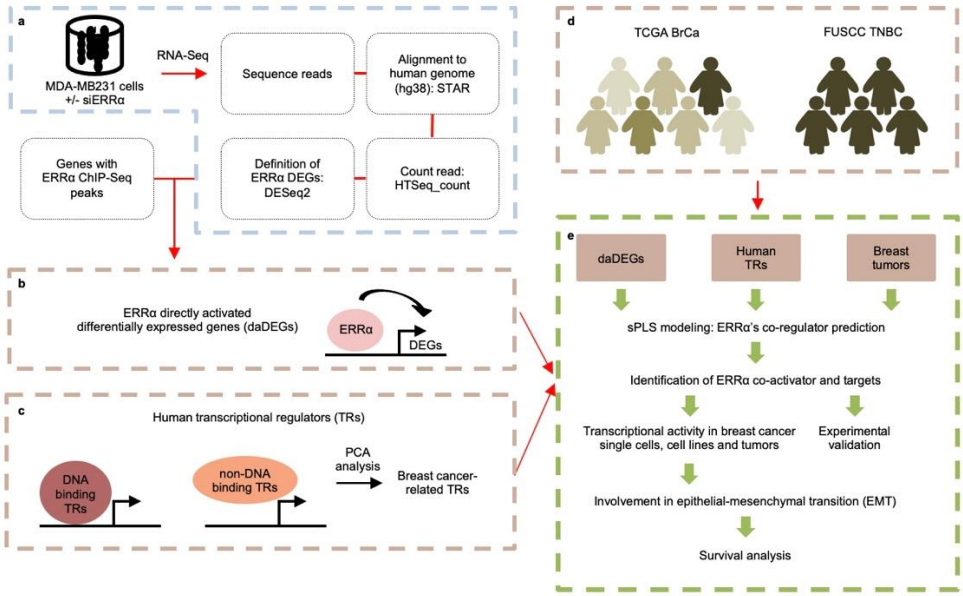


Figure 22. Workflow used for the prediction of ERRα co-regulators. a. ERRα target genes were identified by RNA-Seq in MDA-MB231 cells after transfection with siRNAs targeting ERRα. b. Establishment of directly activated differentially expressed genes (daDEGs) by identification of the DEGs that display at least one ERRα peak in ChIP-Seq experiment. c. Human (DNA-binding or non-DNA binding) transcriptional regulators were collected from open access databases. Principal component analysis (PCA) was used to determine their contribution to expression variability in all TCGA breast cancers tumors. d. RNA-Seq datasets from the TCGA breast cancer samples and from the Fudan University Shanghai Cancer Center (FUSCC) Triple Negative Breast Cancers (TNBC) were used. e. Directly activated DEGs, human TRs and breast tumor data were used to generate sPLS models to identify potential ERRα co-regulators in breast cancers. Bench experiments were conducted to validate the sPLS model's prediction results.

Here we used an adaptive sparse partial least squares (sPLS) regression algorithm to predict co-regulators for ERRα in breast cancer. 45 ERRα directly activated targets related to cell migration pathway and 493 breast cancer relevant transcriptional factors (TRs) were first selected and then were used as the input of computation. Based on the modeling results, ZEB1 was predicted and verified as the most potential and significant

co-activator of ERRα regardless of the size of the expression dataset, subtypes, ages and ethnics. 8 out of 45 ERRα directly activated differentially expressed genes (8 DEGs) were identified as ZEB1-related DEGs from the modeling results (i.e., that may be co-regulated by both ERRα and ZEB1 in breast cancer).

Next, we investigated the expression relationships between ERRα, ZEB1 and the 8 DEGs in breast cancer single cells, cell lines and tumors. The results showed a higher expression of ZEB1 and the 8 DEGs in TNBC subtype, especially in the MES (mesenchymal-like) subtype which is related to epithelial mesenchymal transition (EMT) process. In contrast, the expression of ESRRA (encoding ERRα) is scattered across all subtypes. RT-qPCR experiments were then performed on breast cancer cells transfected with siRNA targeting ERRα and/or ZEB1 to validate this relation pattern. Results verified the positive effect of ERRα and ZEB1 on the 8 DEGs specifically in TNBC cells, but not in other breast cancer subtypes.

Then, we investigated the cooperation of ERRα and ZEB1 in TNBC cells. Published ChIP-Seq experiments of ERRα and ZEB1 were analyzed separately, that indicate a direct binding of ERRα and suggested an indirect binding of ZEB1. The same binding regions of ERRα and ZEB1 on the 8 DEGs were detected in ChIP-qPCR experiments which suggested that they participate in the same molecular pathway. Proximity ligation assay (PLA) experiments were performed and identified a physical interaction between ERRα and ZEB1, as well as the observation of the decreased interaction upon depletion of each factor. Moreover, ChIP-qPCR after siRNA-mediated ERRα depletion or ZEB1 depletion respectively further demonstrated the regulation of ZEB1 on the expression of 8 DEGs depends on ERRα. All of this is consistent with the hypothesis that ZEB1 is a co-activator of ERRα to regulate the expression of the 8 identified DEGs in TNBC cells.

Tumors are a complicated mixture of cancer cells and microenvironment. After the experimental validation of the transcriptional activity of ERRα and ZEB1 on the 8 DEGs in TNBC cancer cells, we investigated these activities in non-cancer cells present in the tumor microenvironment. Published scRNA-Seq data of single cells from TNBC tumors, ER+ tumors and Her2+ tumors collected from breast cancer patients were analyzed. Interestingly, we observed an obvious enriched expression of ZEB1 and the 8 DEGs in cancer-associated fibroblasts (CAFs), endothelial and pericytic cells among the microenvironment of all types of tumors, whereas again the expression of ESRRA (encoding ERRα) was distributed among all types of cells. The high correlation between these three types of non-cancer cells and the EMT process suggested that the interaction between ERRα, ZEB1 and the 8 DEGs may be correlated to EMT and cancer metastasis through a function in the microenvironment.

Therefore, we next investigated whether the 8 DEGs are also involved in EMT since ZEB1 is a key factor of the EMT process. Results showed a positive correlation between the expression of ZEB1 and the 8 DEGs with the EMT status of the tumors. This suggests that tumors with higher expression of ZEB1 and the 8 DEGs are more likely to present a mesenchymal phenotype and may be more prone to establish

metastasis. Based on the observation in EMT analysis, we hypothesized that ZEB1 and these 8 ERRα-activated DEGs could be a potential target set used for the prognosis prediction of breast cancer patients. Survival analysis in three independent datasets all showed that the high combined expression of the 8 DEGs has a significant association with worse overall survival only in TNBC patients, but not in other breast cancer subtypes. In addition, we also observed a trend of tumor recurrence in TNBC MES subtype patients as well as an increased risk of distant tumor metastasis in five years in TNBC patients with higher joint expression of the 8 DEGs.

In summary, we predicted and validated ZEB1 as a co-activator of ERRα in breast cancer. In addition, 8 DEGs specifically co-regulated by ERRα and ZEB1 in TNBC subtype were identified and explored in EMT and predicting the prognosis of breast cancer patients. Altogether, these highly suggested that the 8 DEGs have significant associations with cancer metastasis and could be united to a new potential therapeutic target in TNBC patients.

Detailed results are presented in the following “Results” part.

Results

Section I. Genes regulated by ERRα in cell migration pathway

To investigate the activity of ERRα in breast cancer, we first identified its transcriptional targets. This section includes the introduction to RNA-Sequencing on breast cancer cells, the selection of differentially expressed genes (DEGs) and the identification of DEGs related to cell migration.

Materials and Methods

1. RNA-Seq of breast cancer cells

For RNA-sequencing, RNAs were extracted from MDA-MB231 breast cancer cells cultured in suspension after siRNA treatment (two different siRNA: siERRα#1, siERRα#2), using Qiagen extraction kit (Courtaboeuf, France). RNA-seq was performed on triplicate samples and libraries built using the mRNA-Seq Library Prep kit of Lexogen following the instructions for 5500 SOLiD and including a conversion step to SOLiD 5500W. Sequencing was performed with SOLiD 5500W System (Life Technologies, Strasbourg, France).

2. Detection of differentially expressed genes of ERRα

DEGs were selected as follows: (1) Fastq data from sequencing were preprocessed by STAR and HTSeq_count tools and raw counts gene profiles were obtained. (2) Normalization and differential expression analysis were performed on raw counts by using R package “DEseq2”. (3) After filtering out genes with expression level lower than 10 in control samples, up-regulated DEGs were defined as genes with significant higher expression (padj < 0.05, FoldChange > 1.5) and down-regulated DEGs were defined as genes with significant lower expression (padj < 0.05, FoldChange < 0.75) by comparing to control samples. (4) Final up/down-regulated DEGs were selected out based on the overlap of results in both siERRα#1 and siERRα#2 conditions.

Next, DEGs were then intersected with the 4,846 genes with ERRα peaks from our previous anti-ERRα ChIP-seq experiments (Cerutti et al., 2022) in order to obtain ERRα-directly or -indirectly regulated DEGs.

3. GO analysis on differentially expression genes

Gene ontology (GO) enrichment analysis was performed on ERRα-directly activated DEGs by R package “goseq”. Many redundant similar GO terms resulted from the tree-like structure of Gene Ontology make analysis results unclear. We then only preserved 44 GO terms located at level 4 in the GO tree, where the only level 1 GO term is “biological process” (GO:0008150). 12 out of 44 GO terms showed significant correlation (adj-pvalue < 0.05) with these ERRα directly activated DEGs. DEGs annotated to GO term “cell migration” were preserved and used into the subsequent

model computation.

Heatmap was generated using pheatmap package. Bar plot was generated using ggplot2 package in R software.

Results and discussion

We first investigated the transcriptomic function of ERRα in a cell model cultured under 3D condition. To remove the bias resulting from the quality of siRNA, we cultured cells transfected with two different types of siRNA (siERRα#1, siERRα#2) independently. For the RNA-Sequencing, we first extracted RNA from MDA-MB231 breast cancer cells, following with sequencing, aligning to human genome and reads counting analysis. We totally obtained 917 DEGs regulated by ERRα based on results from two conditions, in which 288 DEGs showed increased expression after silencing ERRα (ERRα-repressed DEGs) as well as the expression of 629 DEGs decreased in the absence of ERRα (ERRα-activated DEGs) (**Figure 23a**). Comparison with our lab's previous results of genes with ERRα binding peaks obtained in ChIP-seq experiments (Cerutti et al., 2022), we obtained 73 ERRα directly-repressed DEGs, 215 ERRα indirectly-repressed DEGs, 190 ERRα directly-activated DEGs and 439 ERRα indirectly-activated DEGs (**Figure 23a**). Since we focus on the positive transcriptional activity of ERRα, these 190 ERRα- activated DEGs with ERRα enrichment peaks were used to explore the pathways with the active participation of ERRα.

GO analysis was then performed on these 190 DEGs, demonstrating 12 significantly enriched biological pathways (adj-pvalue < 0.05) (**Figure 23b**). We noticed that ERRα directly-activated DEGs mainly play an important role in cell movement, with the term "cell migration" being the highest enriched pathway (adj-pvalue = 5.52e-10). Only 45 out of 190 ERRα directly-activated DEGs were found to be associated to cell migration process which is known to be highly involved in cancer cell invasion (**Figure 23b**).

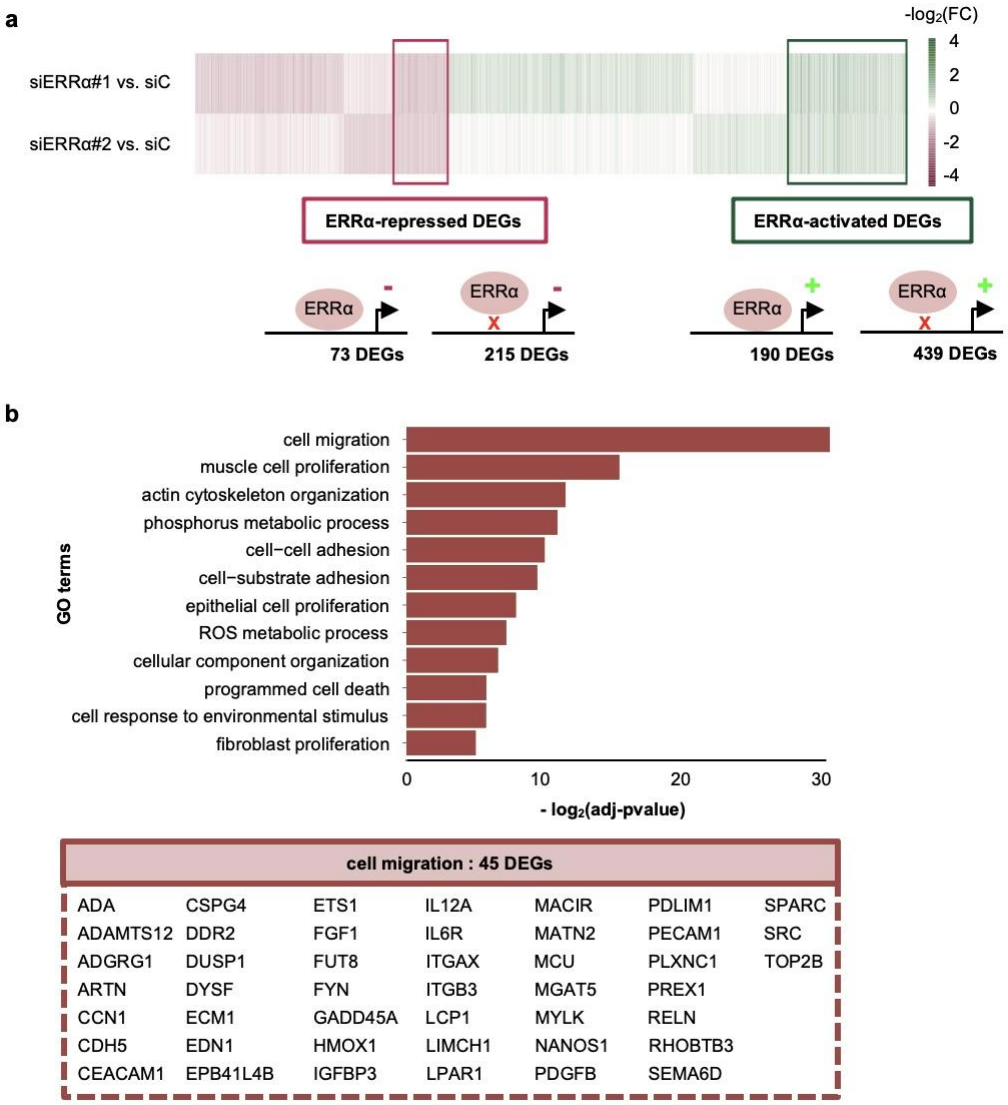


Figure 23. Identification of ERRα targets involved in cell migration. a. Heatmap of ERRα-regulated genes (differentially expressed genes; DEGs), determined by RNA-Seq after treatment with the indicated siRNAs (above). Number of genes with ERRα-binding peak identified by ChIP-Seq is indicated (below). b. Gene Ontology analysis. Main GO terms identified as significant are shown with their –log2(adj-pvalue). List of the 45 genes found in "cell migration" category is shown.

Section II. ERRα co-activators prediction and common targets exploration

The aim of this section is to predict co-regulators for ERRα using bioinformatic methods. Here we used a mathematical method called adaptive sparse partial least squares (sPLS) to build a regression model interpreted by a list of transcriptional regulators (TRs) for each gene (DEG). For model construction, we prepared three types of data from biological experiments and bioinformatic analysis as input. This section includes the preparation of three input data, co-activator prediction and determination of common targets.

Materials and Methods

1. Public expression datasets

TCGA

We downloaded the complete TCGA Breast Cancer RNA-Seq data and clinical records from Genomic Data Commons (GDC) data portal (https://portal.gdc.cancer.gov/, accessed on January, 28th 2022). Totally, we collected 1091 breast cancer primary tumor samples without FFPE (formalin fixation and paraffin embedding) cases and made the link between the RNA-Seq data and the clinical data. Breast cancer subtypes were clinically stratified based on the status of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (Her2) from IHC experiments recorded in TCGA (Nolan et al., 2023). Here "Luminal-like" subtype is defined as ER+/PR+/Her2-; "Her2+" subtype is defined as ER any status/PR any status/Her2 +; "TNBC" subtype is defined as ER-/PR-/Her2-; "Others" is defined as the rest of tumors (**Figure 24a**). In addition, "pre-menopause" group comprises patients aged ≤ 45 and "post-menopause" group comprises patients aged ≥55 according to the criteria of the WHO (https://www.who.int/news-room/fact-sheets/detail/menopause) (Davis and Baber, 2022).

FUSCC TNBC

We downloaded the Fudan University Shanghai Cancer Center Triple Negative Breast Cancer (FUSCC TNBC) RNA-Seq data and clinical features from (Jiang et al., 2019). The file includes the normalized expression data (fpkm) of 360 FUSCC TNBC samples and was linked to clinical data. FUSCC TNBC molecular subtyping information was derived from (Jiang et al., 2019) (**Figure 24b**).

Expression data was pre-processed by removing samples with $Expression_{ESRRA} = 0$ and removing genes not expressed in all of the samples.

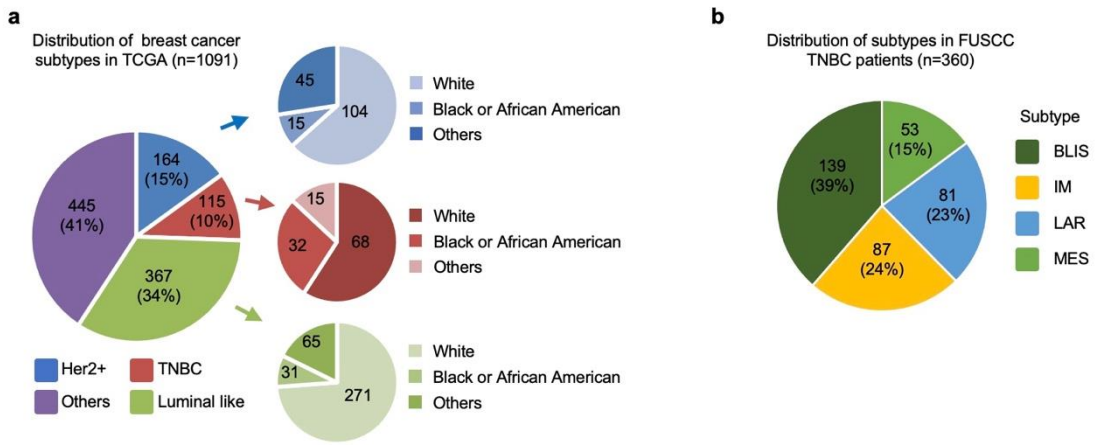


Figure 24. Human breast tumors datasets preparation used in sPLS modeling. Pie charts presenting breast cancer subtypes distribution in TCGA (a) and FUSCC TNBC (b) datasets. a. 1091 breast cancer primary tumor samples from TCGA divide into four subtypes based on the expression of ER, PR and Her2 determined by IHC. Luminal-like tumors are ER+/PR+/Her2-. Her2+ tumors are ER any status/PR any status/Her2+. TNBC are ER-/PR-/Her2-. Each TCGA subtype is classified in three different ethnic groups: White population, Black or African American population and Others. b. FUSCC TNBC tumors are divided into four different subtypes on the basis of multi-omics profiling. BLIS: basal-like and immune-suppressed; IM: immunomodulatory; LAR: luminal androgen receptor; MES: mesenchymal-like.

2. Determination of pre-selected TRs

TRs display distinct functions in different cancers. To eliminate the influence of the specific activity of TRs in different cancer types, we tried to construct a list of breast cancer-related TRs for the subsequent sPLS modeling. First, we collected known 2175 human TRs from public databases (Cerutti et al., 2022). The list of 2175 TRs were pre-processed in TCGA all breast primary tumors as follow:

(1) remove TRs with mean count < 10;

(2) remove TRs with mean expression < 0.1 or standard deviation < 0.1 in upper-quartile normalized data;

(3) perform log2-transformation on upper-quartile normalized expression of TRs.

Principal component analysis (PCA) analysis was next performed on log2-transformed normalized expression data of TRs, preserving the most breast cancer-correlated TRs according to the evaluated parameters: "|PCA Component 1| or |PCA Component 2| > 0.6". Particularly, we added 38 co-regulators of ERRα which are predicted from cells cultured in 2D condition from Cerutti et al. (2022). We removed TRs with mean count of three control samples in siERRα experiments under 3D condition < 100. Finally, we obtained 493 breast cancer-related TRs as the pre-selected TRs.

3. ERRα's co-activators prediction

Process for 45 ERRα-DEGs

Previously, 45 ERRα directly-activated DEGs (45 DEGs) related to cell migration pathway were detected in **section I**. Some additional steps on expression data were needed to ensure the quality of genes by removing genes which had insufficient expression in tumors before model computation:

(1) upper-quartile normalization;

(2) remove genes with mean expression < 0.01 or SD (standard deviation) < 0.01;

(3) remove genes including NA (missing data);

(4) remove genes whose expression is 0 in over half of samples;

(5) log2-transformation.

Process for 493 pre-selected breast cancer-related TRs

Additional steps on expression data were needed to ensure the quality of TRs by removing TRs which had insufficient expression in tumors before model computation:

(1) upper-quartile normalization;

(2) remove TRs with mean expression < 0.01 or SD < 0.01;

(3) remove TRs including NA (missing data);

(4) remove TRs whose expression is 0 in over half of samples;

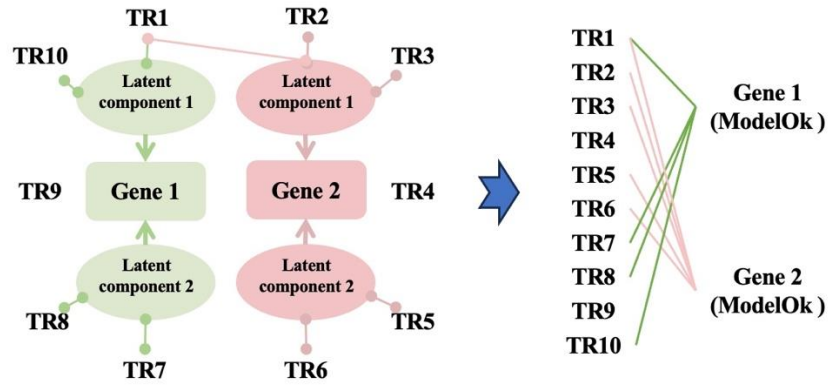
(5) remove TRs existing in 45 DEGs;

(6) log2-transformation.

Parameters training and sPLS modeling

We used an adaptive sPLS regression method in plsgenomics R package to generate a regression model for each gene in order to explain gene transcription activity through transcription factors. The complete modeling workflow was described in (Cerutti et al., 2022) (**Figure 25**). In short, for each computation, the value of latent components and lambda parameter were first tuned and determined by ten-fold cross-validation. Next, after modeling based on these optimized parameters, R-squared determination coefficient (R^2) was used to evaluate each gene model's quality. High-quality gene models ($R^2 > 0.6$) described by a set of TRs were selected and kept for TR prediction (**Figure 25a**).

a



b

	TR1	TR2	TR3	TR4	TR5	TR6	TR7	TR8	TR9	TR10
Frequency (number of related genes)	2	1	1	0	1	1	1	1	0	1
mean-prop (contribution to 45 DEGs)	1	0.5	0.5	0	0.5	0.5	0.5	0.5	0	0.5

Figure 25. The core of sparse partial least square (sPLS) model computation. a. General description of the sPLS modeling. Several latent components are first generated from the linear combination of a list of transcriptional regulators (TRs) determined by modeling algorithm, then used to build a model to interpret the expression of one gene. b. Models computed for each gene based on the same set of TRs. The regression relationship between genes and TRs are detected in a. The value "mean-prop" representing the contribution of one TR to all genes (with a high-quality computed model (ModelOK)) will be given to each TR in order to evaluate, rank and detect the most correlated and frequently detected TR (Cerutti et al., 2022).

ERRα's co-activators prediction

For each gene (DEG), the computation was performed in 10 replicates (one time). The contribution of each TR to 45 ERRα-DEGs' expression in one time (here called "mean-prop") was counted as the number of genes including this TR in their high-quality computation models (**Figure 25b**). The TR with the highest "mean-prop" value will be the best predicted co-factor (top 1 CoR) for ERRα.

Additionally, we modeled three times (30 replicates altogether) for each gene in order to check the variability of sPLS modeling and to obtain more robust results.

4. Identification of DEGs related to both ERRα and ZEB1

To investigate the transcriptional activity of ERRα and its predicted co-activators, we then identified DEGs which are related to both ERRα and top1 CoR based on sPLS modeling results. In each modeling replicate, one sPLS model was built for one gene (DEG) which reflects the contribution of TRs to this gene's expression and in the basis of these gene models, we summarized DEGs including top 1 CoR in their high-quality

sPLS model (**Figure 25b**). Finally, we counted a frequency number for each DEG (gene model includes top 1 CoR: 1, does not include: 0) from 30 modeling replicates. For example, frequency number of one DEG is 30 means the top 1 CoR always contributes to its transcription expression and this relationship is very robust.

Circular bar plot was generated using ggplot2 package. Heatmap displaying frequency rate (= frequency number / 30) was generated using pheatmap package in R software.

Results and discussion

2.1 Determination of pre-selected TRs

2175 known human transcription factors (TRs) were collected from three open access databases based on Cerutti et al. (2022). To ensure the participation of ERRα with our final predicted co-regulators (CoRs) to the biological activity of breast cancer, 500 TRs were selected first by using PCA analysis in TCGA breast tumors. We specifically removed TRs with an average expression count less than 100 in the three control cell models (**Figure 23a**). In addition, similar predictions for ERRα were conducted before, based on the cell model cultured under 2D condition (Cerutti et al., 2022) . Thus, 38 predicted and verified TRs obtained from 2D cell model were added. In total, 493 pre-selected TRs were finally considered as human breast cancer relevant TRs which were used to perform model computation (**Figure 26**).

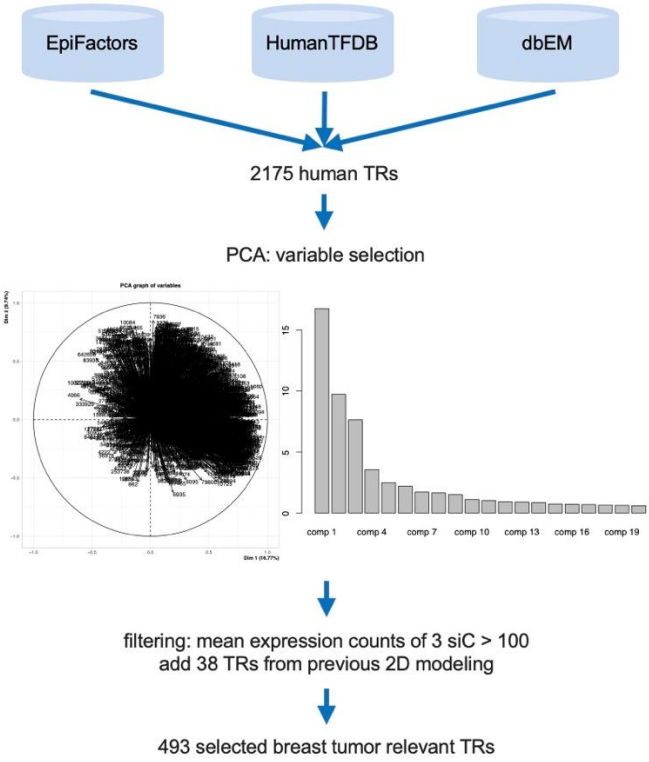


Figure 26. Human TRs preparation used in sPLS modeling. Flow chart presenting

known human breast cancer related TRs filtering steps. 2175 human TRs were collected from three open access databases (EpiFactors, HumanTFDB and dbEM) based on Cerutti et al. (2022). 500 breast cancer related TRs were selected out by PCA. Filtering and TR addition were used leading to selection of 493 breast tumor relevant TRs.

2.2 ERRα's co-activators prediction

We then used sPLS to predict CoRs for ERRα. Based on three types of input data including 45 ERRα directly-activated DEGs (45 DEGs), 493 pre-selected human breast tumor relevant TRs (493 TRs) and tumor expression data, we first performed sPLS model computation on TCGA all breast tumors. Three times computations each including ten replicates were conducted to reduce the variability of sample size. Top 5 predicted TRs ranking based on their contribution to the expression of 45 DEGs (call "mean-prop" hereafter) were displayed in **Figure 27a**. Robust results were observed indicating that ZEB1 ($p < 0.05$) was the most frequent TR (bottom: top 1) followed by PROX1, ACTB, EPAS1 and YBX1. As there is a huge heterogeneity in breast cancer, we also conducted model computation on each subtype distributed as Luminal-like, Her2 positive (Her2+), triple negative breast cancer (TNBC) and others. Strikingly, ZEB1 consistently ranked as the remarkable top 1 CoR for ERRα among all computed models (**Figure 27a above**). We also noticed that a certain level of diversity of TRs and a small fluctuation of the rankings of TRs were detected among results in subtypes, especially TNBC which included the lowest number of tumors (115) and the highest level of heterogeneity. Consequently, a larger TNBC sample set including 360 Chinese tumors from Fudan University Shanghai Cancer Center (FUSCC) was collected and used to conduct model computation again. Same computation procedures were performed on all FUSCC TNBC tumors as well as four molecular TNBC subtypes distributed as BLIS (basal-like and immune-suppressed), IM (immunomodulatory), LAR (luminal androgen receptor) and MES (mesenchymal-like) subtype. Similar results were observed in FUSCC TNBC tumors indicating ZEB1 as the most potential co-activator for ERRα (**Figure 27a below**).

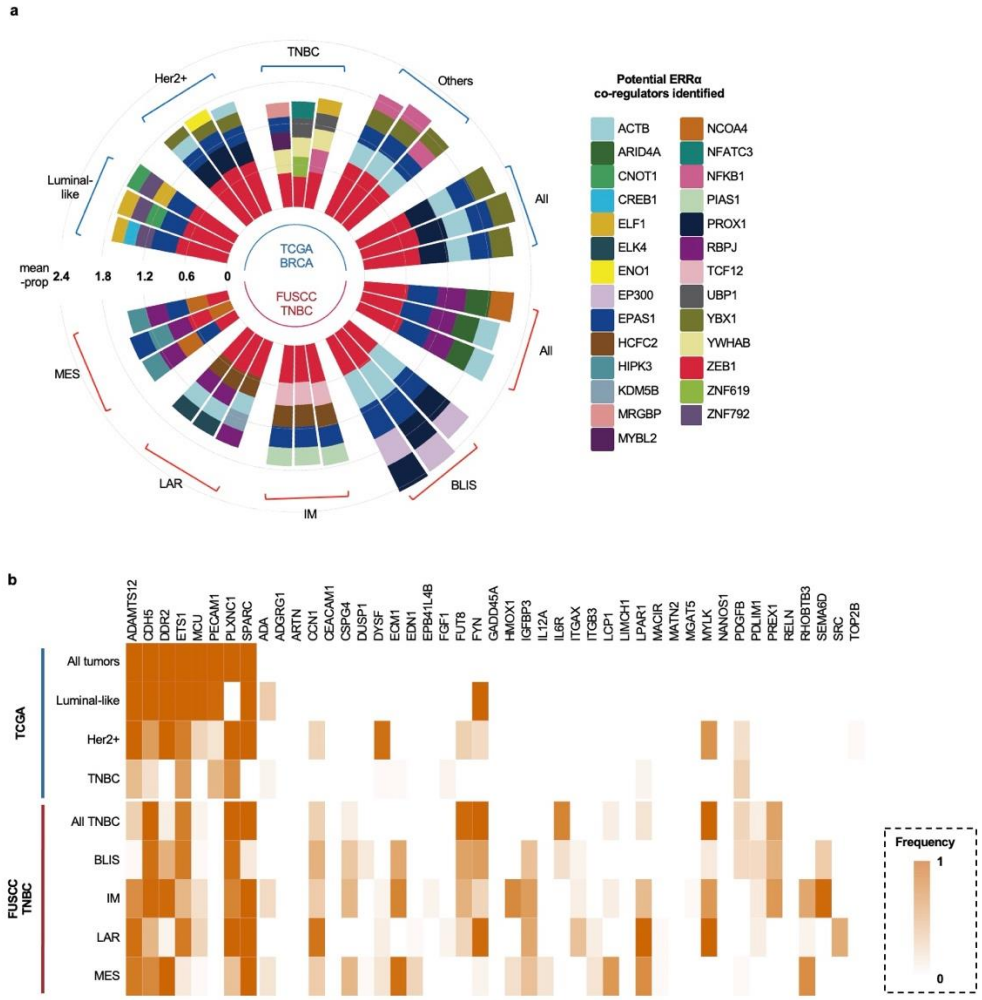


Figure 27. Identification of potential ERRα co-regulators in different breast cancers subtypes. a. Circular bar plot displaying top five candidate ERRα co-regulators obtained with sPLS modeling in breast cancer RNA-Seq datasets. Analysis was performed in TCGA (above) and FUSCC TNBC (below) datasets, as well as in each of their reported subtypes as indicated. In each category, modeling was performed three times, each including ten replicates. The height of each box represents one TR's mean proportion of 10 sPLS modeling replicates. Mean proportion of TRs is sorted from the largest (bottom) to the smallest (top). b. Heatmap showing the frequency of ZEB1 appearance as a TF explaining the expression of individual DEGs in subtypes of breast tumor datasets. BLIS: basal-like and immune-suppressed; IM: immunomodulatory; LAR: luminal androgen receptor; MES: mesenchymal-like.

2.3 Identification of DEGs related to both ERRα and ZEB1

As ZEB1 was predicted as a significant potential co-activator of ERRα, we wanted to explore the molecular mechanism of their common transcription activity. Among 45 cell migration relevant DEGs, we observed that ZEB1 always only keeps strong ability on modeling the expression of the 8 ERRα directly activated DEGs (ADAMTS12,

CDH5, DDR2, ETS1, MCU, PECAM1, PLXNC1, SPARC) in TGCA all breast tumors, while not on any other ERRα's DEGs. Next, in order to verify this observation and explore the subtype heterogeneity, we also explored each breast cancer subtype dataset of TCGA and FUSCC TNBC dataset. We observed that ZEB1 could still strongly influence the expression of most of these 8 DEGs in all subtypes despite it also impacted other DEGs in a certain subtype **(Figure 27b)**.

Moreover, we modeled on breast tumor dataset from (1) Black or African American and White patients of each subtype of TGCA; (2) pre-menopause and post-menopause patients of TGCA and FUSCC TNBC. Results showed the regulation of ZEB1 on most of the 8 DEGs are similar among all age group **(Figure 28a)** and different ethnic patients **(Figure 28b).**

Altogether, we suggested that ZEB1 is the most potential co-activator of ERRα regardless of breast cancer subtypes. We additionally observed that the expression of the 8 ERRα directly-activated DEGs can be influenced by ZEB1 across all tumor groups regardless of subtypes, ages and ethnics, which strongly suggests that the 8 DEGs are ERRα-ZEB1 common targets in human breast cancers.

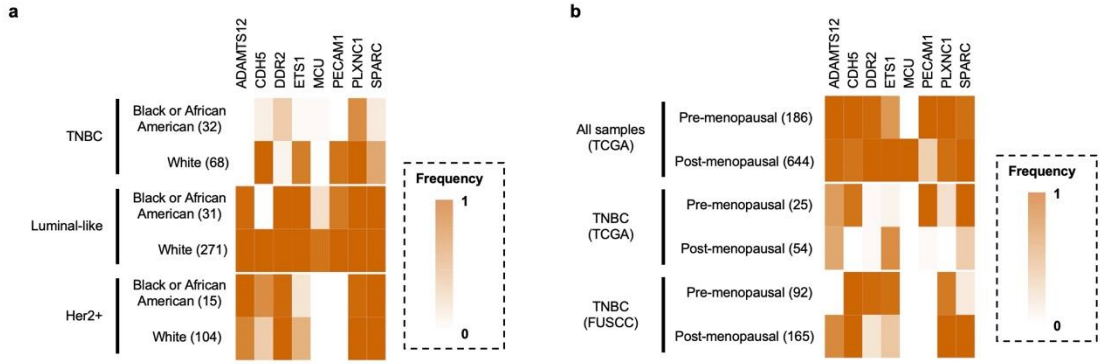


Figure 28. ERRα-ZEB1 DEGs. a-b. Heatmaps showing frequency of ZEB1 appearance as a TF explaining the expression of individual DEGs in subtypes of breast tumor datasets. a. Tumors were sorted according to ethnicity. b. Tumors were sorted according to age, and predicted pre- (≤ 45 y.o.) or post (≥55 y.o.)-menopausal status. Number of tumors per category is indicated.

Section III. Stability evaluation of dataset size in sPLS modeling

The size of the sample set required for mathematical computation modeling has a certain impact on the result. In order to better investigate and compare the modeling results from different breast cancer subtypes, age groups or ethnic groups, we explored the influence of sample set size from random samples. This section includes the description of the preparation of random sample sets, the modeling on random sample sets and the evaluation of the influence of sample size.

Materials and Methods

1. **Modeling on random sample sets**

Model computations based on 100 random samples were first performed. We randomly selected 100 samples to generate a new data set (random set) by using random sampling method on all breast tumors of TCGA (1091 samples) and then started sPLS modeling. The final computed models were obtained through ten-fold cross-validation method. In order to reduce deviation of random sampling, we built a total of ten random sets in which each random set contains 100 samples from random sampling and finally obtained ten sets of sPLS modeling result. Additionally, we also explored the influence of other sample set size (200, 300, 400, 500, 600, 700, 800, 900, 1000, 1091 samples). It is worth noting that when the sample set size becomes 1091, we used all samples to compute ten times without any random sampling.

2. **The evaluation of the influence of sample size**

We interpreted and evaluated the modeling results of each sample set size from 4 aspects:

(1) Number of genes with high-quality model (nGenesModelOk). sPLS modeling aims to build a fitted linear regression model for the dependent variable, which means to build an explanatory model in high-quality for each ERRα-activated gene based on the expressions of a set of TRs. So, we first checked the number of genes with high-quality models.

(2) Proportion of genes associated to ZEB1 in all genes with high-quality model (propGenesRelatedToZEB1). From our previous results **(Section II)** suggesting that ZEB1 is the most potential co-regulator (top 1) which may contribute to the transcription of genes directly activated by ERRα, we next compared the proportion of ZEB1-associated genes between different sample set size.

(3) Frequency of the observed correlation between ZEB1 and genes in modeling results (corrZEB19DEGs). We evaluated the association between ZEB1 and genes through the frequency of association in the ten sets of sPLS modeling result of each sample set size.

(4) Significance of ZEB1 compared to top 2 TR (sigZEB1). From above, we found that ZEB1 is not only the most stable but also the most influential co-regulator, while

other TRs rankings vary. Therefore, we also calculated the difference ratio between the influence of ZEB1 and that of other top 2 TR in order to confirm the dominance of ZEB1:

$$\text{sigZEB1} = \frac{\text{mean} - \text{prop of ZEB1}}{\text{mean} - \text{prop of top 2 TR}}$$

Box plots and the dot plot were generated using ggplot2 package in R software.

Results and discussion

The results showed that the median of the number of genes with high-quality models (nGenesModelOk) among ten modeling results is 15 when we used 100 random samples, which is larger than the results of other size of sample set. That is, nearly 15 genes closely associated with breast cancer TRs were identified when we modeled on 100 samples. When the sample set size is > 100, nGenesModelOk decreases as well as the number and the list of GenesModelOk gradually becomes stable (**Figure 29**) (**Table 5**). This suggested that in breast cancer, when the size of tumor sample set is > 100, a list of genes highly correlated to breast cancer relevant TRs can always be identified and preserved, while some of genes selected out at the beginning are gradually proved to be false positive results and dropped away.

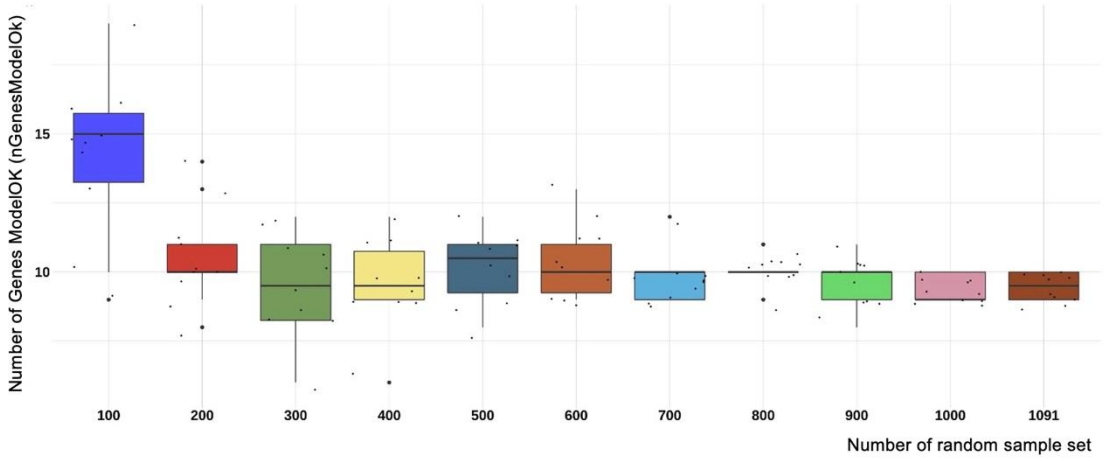


Figure 29. Box plot shows the number of genes with high-quality sPLS model (nGenesModelOk). For every sample size, model computation on ten random sets were performed in which each of them produced linear regression models for genes. High quality model is evaluated based on R-squared determination coefficient ($R^2 > 0.6$).

100 Random Samples	200 Random Samples	300 Random Samples	400 Random Samples	500 Random Samples	600 Random Samples	700 Random Samples	800 Random Samples	900 Random Samples	1000 Random Samples	1091 Random Samples
ADA	ADA	ADA	ADA	ADA	ADA	ADA	ADA			
ADAMTS12	ADAMTS12	ADAMTS12	ADAMTS12	ADAMTS12	ADAMTS12	ADAMTS12	ADAMTS12	ADAMTS12	ADAMTS12	ADAMTS12
ADGRG1	ADGRG1	ADGRG1	ADGRG1	ADGRG1	ADGRG1	ADGRG1				
CCN1	CCN1									
CDH5	CDH5	CDH5	CDH5	CDH5	CDH5	CDH5	CDH5	CDH5	CDH5	CDH5
DDR2	DDR2	DDR2	DDR2	DDR2	DDR2	DDR2	DDR2	DDR2	DDR2	DDR2
DYSF	DYSF		DYSF							
ECM1										
EDN1										
ETS1	ETS1	ETS1	ETS1	ETS1	ETS1	ETS1	ETS1	ETS1	ETS1	ETS1
FGF1	FGF1									
FUT8	FUT8	FUT8	FUT8	FUT8	FUT8	FUT8	FUT8			
FYN	FYN	FYN	FYN	FYN	FYN	FYN		FYN		
GADD45A										
IGFBP3										
IL6R	IL6R	IL6R								
LCP1										
LIMCH1										
LPAR1										
MACIR	MACIR	MACIR	MACIR	MACIR	MACIR	MACIR	MACIR	MACIR	MACIR	MACIR
MATN2										
MCU	MCU	MCU	MCU	MCU	MCU	MCU	MCU	MCU	MCU	MCU
MGAT5	MGAT5									
MYLK	MYLK									
PDGFB	PDGFB									
PECAM1	PECAM1	PECAM1	PECAM1	PECAM1	PECAM1	PECAM1	PECAM1	PECAM1	PECAM1	PECAM1
PLXNC1	PLXNC1	PLXNC1	PLXNC1	PLXNC1	PLXNC1	PLXNC1	PLXNC1	PLXNC1	PLXNC1	PLXNC1
PREX1	PREX1			PREX1	PREX1	PREX1				
RHOBTB3										
SPARC	SPARC	SPARC	SPARC	SPARC	SPARC	SPARC	SPARC	SPARC	SPARC	SPARC
SRC	SRC									
TOP2B	TOP2B	TOP2B	TOP2B	TOP2B	TOP2B	TOP2B	TOP2B	TOP2B	TOP2B	TOP2B

Table 5. The table displays genes with high-quality model (GenesModelOk) upon the modeling of each sample set size. Genes in yellow are ZEB1-related genes. Genes in blank are related to other TRs, not ZEB1. Altogether, 19 ZEB1-related genes were identified. Only 8 DEGs (ADAMTS12, CDH5, DDR2, ETS1, MCU, PECAM1, PLXNC1, SPARC) always show an association with ZEB1 regardless of the sample size.

Our results above show that ZEB1 was always identified as the most potential co-regulator (top1) (**Figure 27a**). We then checked the evolution of the list of GenesModelOk associated with ZEB1 (propGenesRelatedToZEB1) among all size of sample set. We first observed that when the sample size = 100, half of all GenesModelOk (17/32) are associated to ZEB1 (**Table 5, genes in yellow**) despite the proportion of ZEB1-related genes was the lowest (highest to 8/10 = 80 % when sample size = 1091) (**Figure 30**). When we used 200 or 300 samples, the proportion of genes associated with ZEB1 obviously increases; when the sample size ≥ 400, propGenesRelatedToZEB1 gradually becomes stable and fluctuates around 85% (**Figure 30**). This further indicates that when the sample size is small, some genes will be accidentally identified but they are not stable enough. With the increase of the sample size, the regression relationship is continuously corrected, some unstable genes are removed, and the list of ZEB1-relevant genes is gradually stabilized (**Table 5, genes in yellow**). It is worth noting that genes associated with ZEB1 are always dominant regardless of the sample size as well as nearly half of them were always retained (8/19) (**Table 5, genes in yellow**), suggesting that the direct regulation of ERRα on these cell migration-related target genes (45 DEGs) may be mostly affected by ZEB1. In particular for the 8 stable DEGs, ZEB1 always showed high stability and strong association in the contribution to the transcriptional expression of the 8 DEGs regardless of the size of sample set (**Figure 31**).

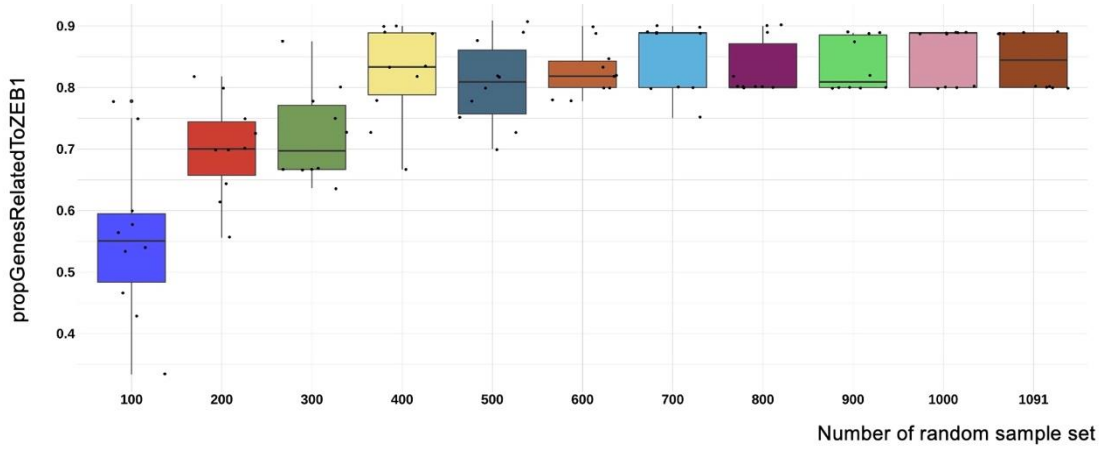


Figure 30. Box plot shows the proportion of genes associated to ZEB1 in all genes with high-quality model (propGenesRelatedToZEB1). High-quality model is evaluated based on R-squared determination coefficient ($R^2 > 0.6$).

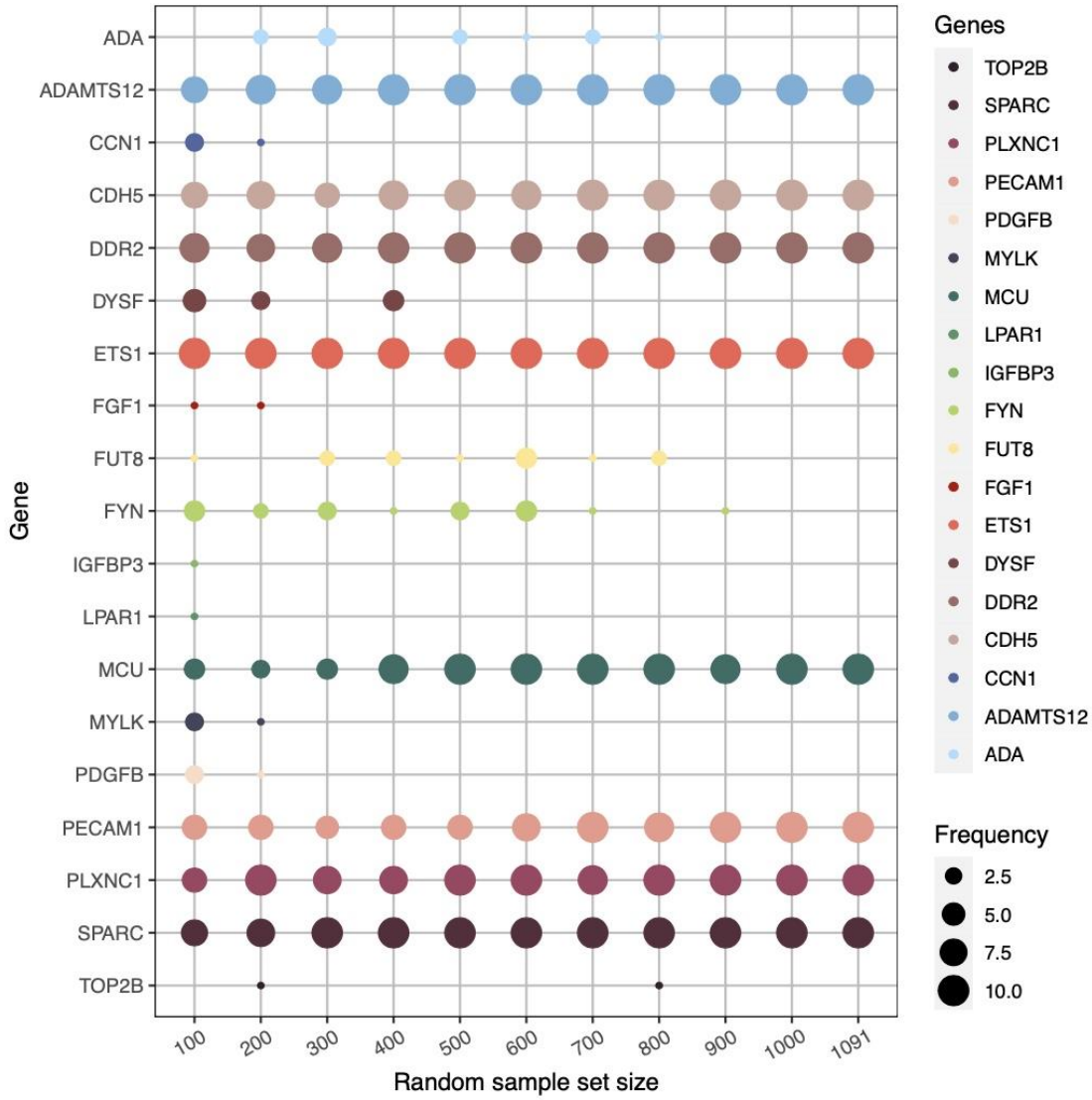


Figure 31. Dot plot shows the frequency of the observed correlation between ZEB1 and ZEB1-relevant genes (genes in **Table 5** in yellow**: 19**) in computed models of each sample set size (corrZEB1DEGs). The most frequency is 10 as computations on ten random sample sets for each sample size.

In order to further demonstrate the uniqueness of ZEB1 as ERRα's co-regulator, we also compared and evaluated the difference in the ability of explaining targets' transcription levels between ZEB1 (top 1) and top 2 TR (sigZEB1). To do this, we calculated a "mean-prop" for each TR which represents the contribution and explanatory ability of the individual TR to gene expression. The results showed that the contribution of ZEB1 was always greater than that of top 2 TR (sigZEB1 > 1) (**Figure 32**). The median of the contribution difference > 2 when the sample size > 100, indicating that ZEB1 could interpret the transcriptional expression of more ERRα target genes (nearly double) than top 2 TR. This suggested that the expression of most ERRα targets is mainly affected by ZEB1 and that the influence of other TRs is relatively limited. This further supports the fact that ZEB1 acts as the most unique and important co-regulator among all predicted TRs regardless of the sample set size.

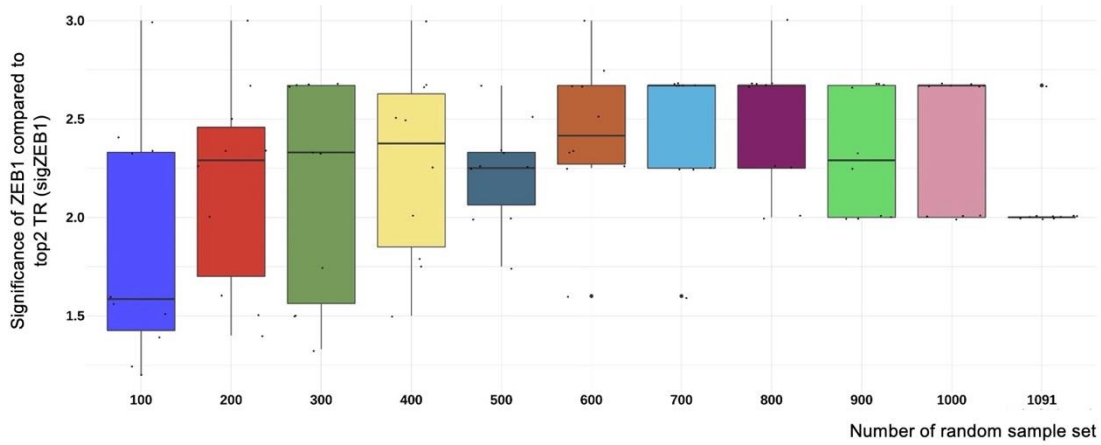


Figure 32. Box plot shows the significance of ZEB1 compared to top 2 TR (sigZEB1). The difference ratio is determined by the value of "mean-prop" of top 1 divided by "mean-prop" of top 2 TR.

From above, we concluded that when computing models on a small number of samples, the modeling results include a certain number of chaotic contents. The number of genes with high-quality model and the number of interpreted TRs was large, which leads to a low proportion of genes related to ZEB1 (17/32), but always more than 50% (**Figure 30**). With the increase of the sample size, computation bias was gradually eliminated as well as genes for which the relationship was not robust. Finally, 8 genes always showed a stably strong correlation with ZEB1 indicating they are robustly co-regulated by ERRα and ZEB1 regardless of sample size (**Table 5, Figure 31**). It is worth

noting that although the identification of irrelevant genes when the sample size is small can affect the modeling results, the genes associated with ZEB1 could be stably identified and always to be dominant (**Table 5**). In particular, the relationship between target genes and ZEB1 is much more significant than the ones with other TRs (**Figure 32**). In summary, we concluded that ZEB1 could consistently be identified as the most unique and influential co-regulator regardless of the sample set size, and could well and stably explain the transcriptional expression of the 8 ERRα directly activated genes which are active in cell migration.

Section IV. Transcriptional relationships between ERRα, ZEB1 and their targets in breast cancers.

In the conclusion of **section II** and the evaluation of **section III**, computed models highly suggested ZEB1 was the most stably potential co-activator of ERRα in breast cancer as well as the expression of the 8 ERRα-directly activated DEGs were also robustly modulated by ZEB1. Thus, the aim of this section is to investigate the transcription relations of ERRα, ZEB1 and their 8 common targets in breast cancer single cells, cell lines and tumors. This section includes the preparation of expression data of breast cancer and the comparison of gene expression.

Materials and Methods

1. Public expression datasets

Breast cancer single cell RNA-Seq data

scRNA-Seq data of breast cancer single cells were collected from (Gambardella et al., 2022). 35,276 individual cells from 32 human breast cell lines (31 cancer cell lines and 1 non-cancer cell line) were sequenced and recorded by cancer subtypes: LA (luminal A), LB (luminal B), H (Her2-enriched), TNA (triple-negative type A) and TNB (triple-negative type B).

Breast cancer cell lines RNA-Seq data

Breast cancer cell lines RNA-Seq data (log(tpm+1)) from the Cancer Cell Line Encyclopedia (CCLE) was downloaded from DepMap Public 22Q4 (https://depmap.org/portal/download/all/, accessed on April 14th 2023). We also downloaded the molecular subtype annotation from (Feldker et al., 2020). We collected 56 breast cancer cell lines data and made the link between the expression data and molecular subtypes. Molecular subtyping information was derived from (Feldker et al., 2020).

GSE18229 breast cancer tumors

We downloaded human breast cancer expression data, gene annotation and clinical features from Gene Expression Omnibus (GEO) database (accession number: GSE18229 only in microarray platform GPL13090). Average expression of several probes aligning to the same gene were calculated and kept. Totally, we obtained the expression profile of 180 human breast primary tumors with clinical annotations. Breast cancer molecular subtyping information was derived from (Feldker et al., 2020).

FUSCC TNBC & TCGA TNBC subtype information

Subtype classification of FUSCC TNBC and TCGA TNBC breast tumors are from (Jiang et al., 2019).

2. Gene expression comparison between breast cancer subtypes

Gene expression comparison in the breast cancer single cell atlas were obtained from the web server (http://bcatlas.tigem.it) built by the research team (Gambardella et al., 2022).

Gene expression comparison in human breast cancer cell lines and breast tumors were based on normalized expression data in log-transformed. For each gene, all samples' expression were normalized to [-1,1] by:

$$X_{norm} = \frac{2 * (X - X_{min})}{X_{max} - X_{min}} - 1$$

and then were compared between molecular subtypes.

Expression heatmaps were generated using pheatmap package. Box plots were generated using ggplot2 package in R software.

Results and discussion

4.1 Transcription activity in breast cancer single cells

To explore the transcription regulation ability of ERRα and ZEB1 in breast cancer, we investigated the expression level of ESRRA, ZEB1 and the 8 common targets in human breast cancer single cells, cell lines and tumors. Human breast cancer single cell RNA-Seq data (scRNA-Seq) of 35,276 cells from 32 human breast cell lines were collected, processed and analyzed by (Gambardella et al., 2022). Six cell clusters (luminal A, luminal B, Her2-enriched, triple-negative type A, triple-negative type B and basal-like) annotated with breast cancer subtypes were identified and showed in an atlas **(Figure 33)**. In this atlas, the expression of ESRRA was scattered across all subtypes whereas ZEB1's expression enriched in triple-negative subtypes. Interestingly, we observed that the expression of half of the 8 DEGs (ADAMTS12, DDR2, ETS1, SPARC) also have an enrichment in TNBC subtypes which follows that of ZEB1 **(Figure 34)**.

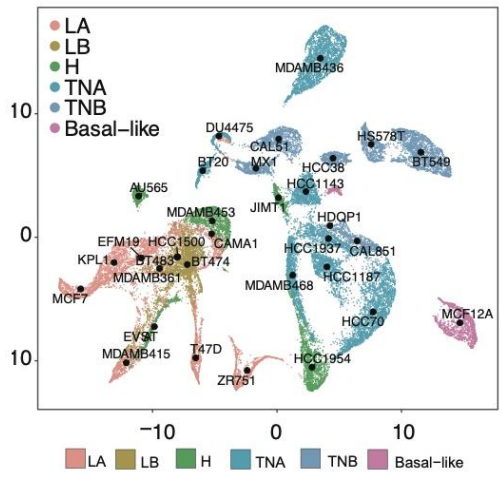


Figure 33. Atlas of breast cancer cell lines analyzed by single cell profiling in Gambardella et al. (2022). LA: luminal A; LB: luminal B; H: Her2-enriched; TNA: triple-negative type A; TNB: triple-negative type B.

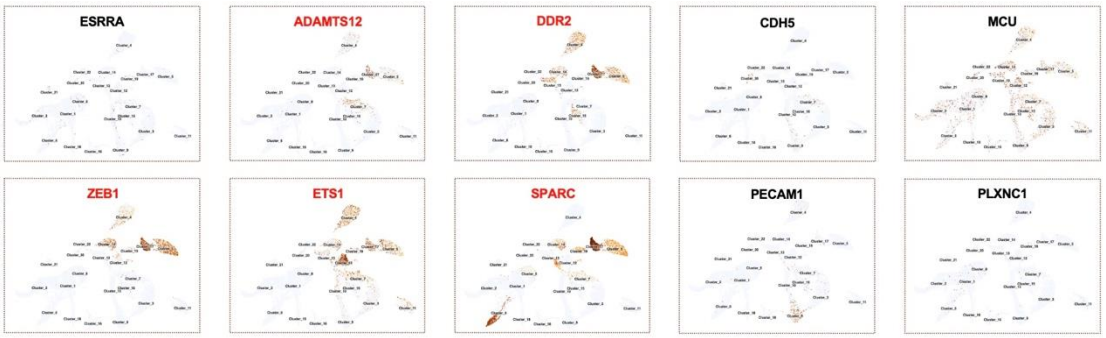


Figure 34. Expression of the 8 DEGs in the clustered breast cancer cell lines. The expression of genes in red enriches in TNBC cells.

4.2 Transcription activity in breast cancer cell lines

Next, we investigated the RNA expression data of 56 breast cancer cell lines with molecular subtype classification from Cancer Cell Line Encyclopedia (CCLE) (Feldker et al., 2020) . Similar to our results obtained in single cells, we observed that ZEB1 and most of the 8 DEGs (ADAMTS12, DDR2, ETS1, MCU, SPARC) displayed a higher normalized expression in the aggressive claudin-low subtype, which is a subfraction of TNBC (Nolan et al., 2023), while ESRRA displayed no obvious expression preference. Four ZEB1-repressed targets (CDH1, LLGL2, CLDN7, CLDN3) were set as a control set and showed an opposite expression pattern to that of ZEB1 (**Figure 35**).

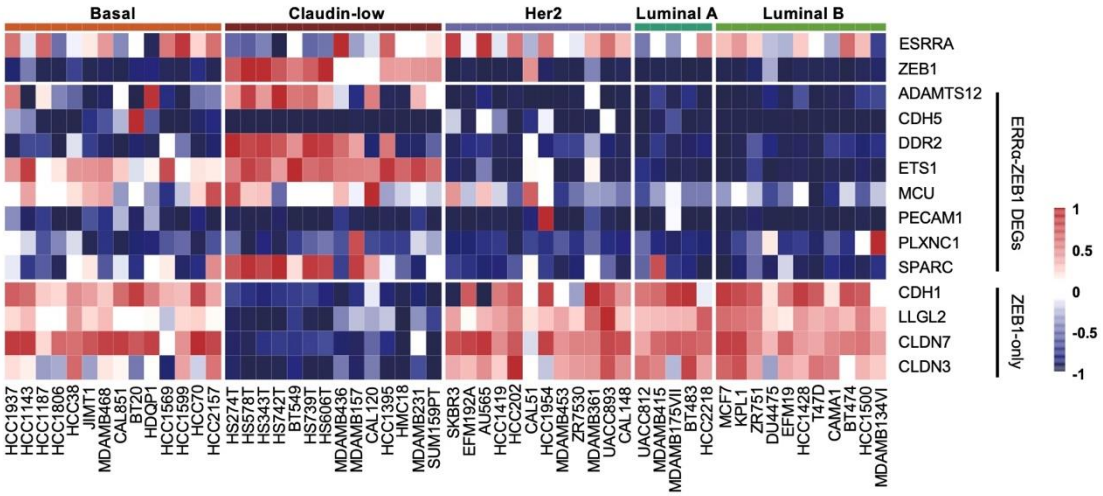


Figure 35. Heatmap showing the expression levels of the indicated genes in 56 CCLE breast cancer cell lines. The expression of each gene is normalized to [-1,1].

4.3 Transcription activity in breast tumors

In addition, we checked the mRNA expression of the above-mentioned genes in human breast tumors with identical molecular subtype annotation as cell lines (Feldker et al., 2020). Claudin-low subtype had an elevated expression of ZEB1 and of most of the 8 DEGs (CDH5, DDR2, ETS1, PECAM1, PLXNC1, SPARC) except MCU which is not included in this dataset. As controls, four ZEB1-repressed targets displayed a lower expression in claudin-low subtypes (**Figure 36**).

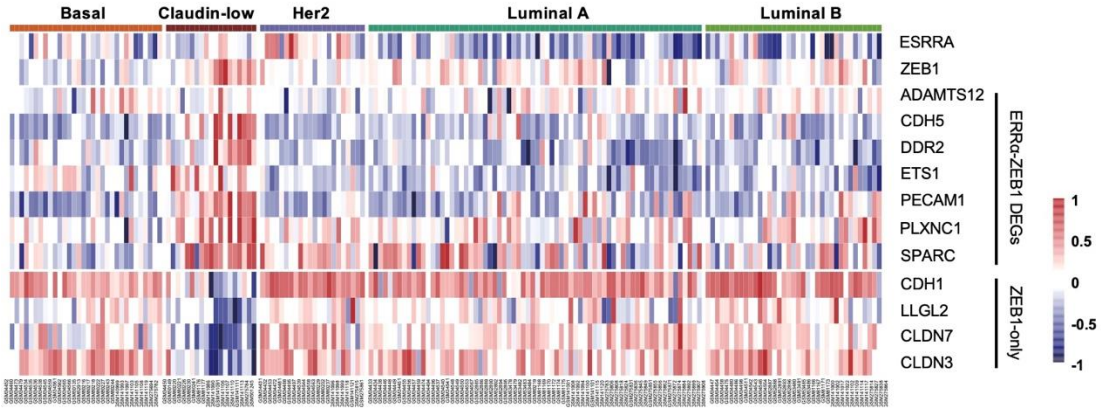


Figure 36. Heatmap showing the expression levels of the indicated genes in human breast cancer tumors (GSE18229). The expression of each gene is normalized to [-1,1].

So far, we knew that ZEB1 and the 8 DEGs displayed a higher expression in TNBC compared to other breast cancer subtypes. TNBC is a type of heterogeneous breast cancers. TNBC can be divided into four different molecular subtypes (BLIS, IM, LAR and MES) based on genetic characteristics (Jiang et al., 2019). Each of these subtypes has a different sensitivity to treatment. Thus, we further explored the expression differences of ESRRA, ZEB1 and the 8 DEGs in the four molecular subtypes of TNBC.

We observed that ESRRA has a higher expression in LAR subtypes of FUSCC TNBC patients. ZEB1 and almost all 8 DEGs (except MCU) are highly expressed in MES subtype and less expressed in BLIS subtypes. In addition, ZEB1, ETS1 and PLXNC1 genes also have higher expression in IM subtypes. Again, the four control genes still displayed an expression opposite to ZEB1 (**Figure 37**). We then compared the combined expression of the 8 DEGs between the four molecular subtypes of FUSCC TNBC as they are all regulated by both ERRα and ZEB1. We found that the average expression of the 8 DEGs in the BLIS subtypes was much lower than that of the other

three subtypes as well as the median of the combined expression of MES subtypes was the highest (**Figure 38a**). We also collected the data of TCGA TNBC annotating with the same subtype information based on the same classification method (Jiang et al., 2019). ZEB1, as expected, was significantly expressed in a higher manner in MES subtypes, as we observed in FUSCC TNBC (**Figure 38c, 38d**). Again, the results also showed that joint expression of the 8 DEGs is lower in BLIS, while MES has a relatively higher expression (**Figure 38b**).

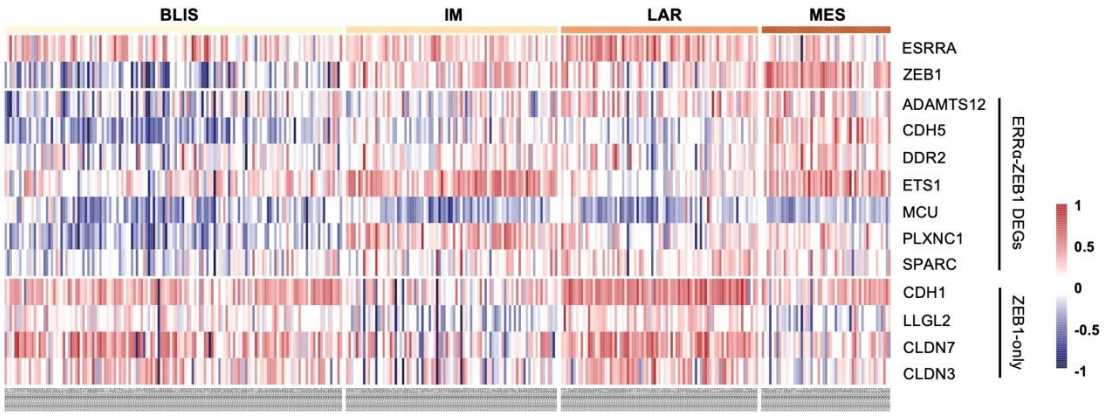


Figure 37. Heatmap showing the expression levels of the indicated genes in FUSCC triple negative breast cancer tumors. BLIS: basal-like and immune-suppressed; IM: immunomodulatory; LAR: luminal androgen receptor; MES: mesenchymal-like. The expression of each gene is normalized to [-1,1].

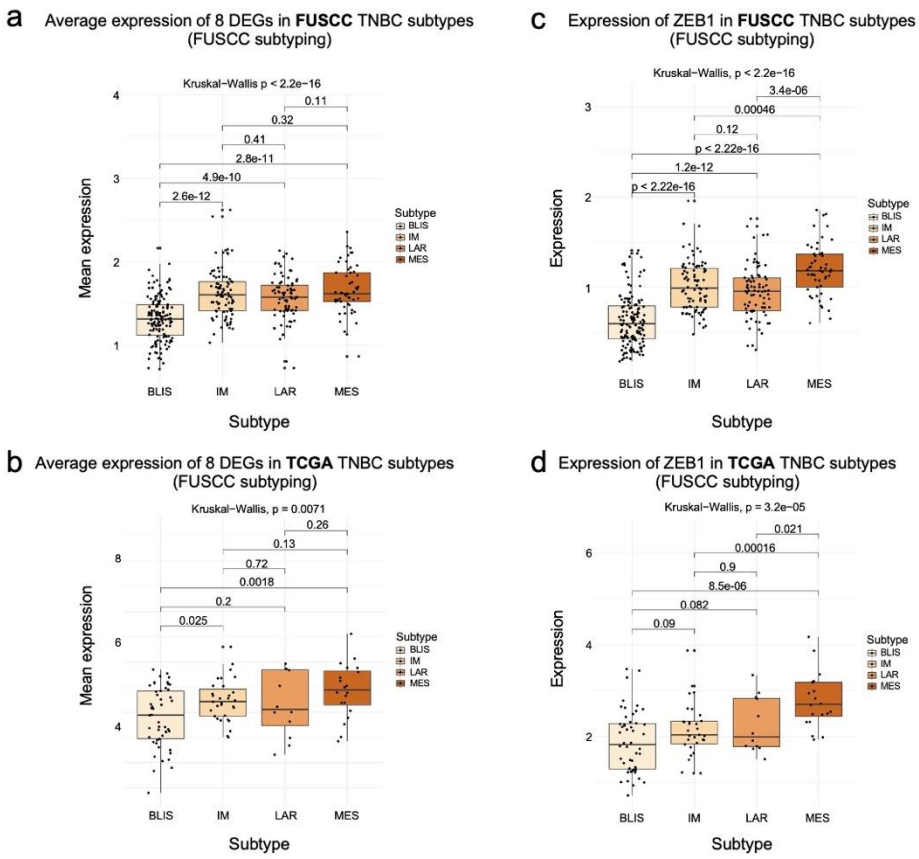


Figure 38. Expression of the 8 DEGs and ZEB1 in TNBC. a. Box plots showing the average expression of the 8 DEGs in FUSCC TNBC. b. Box plots showing the average expression of the 8 DEGs in TCGA TNBC. c. Box plots showing the expression of ZEB1 in FUSCC TNBC. d. Box plots showing the expression of ZEB1 in TCGA TNBC. Expression difference between each pair of groups was evaluated by Wilcox test. Expression difference between multiple groups was evaluated by Kruskal-Wallis test. p-value is presented on the top of the line. BLIS: basal-like and immune-suppressed; IM: immunomodulatory; LAR: luminal androgen receptor; MES: mesenchymal-like.

In summary, expression profiles demonstrated a higher expression of ZEB1 and most of the 8 DEGs in the aggressive claudin-low subtype (a part of TNBC), specifically in MES subtypes (mesenchymal-like) which is characterized by EMT features and correlated to breast cancer metastasis (discussed latter in discussion section), which strongly suggested ZEB1 could be a co-activator of ERRα on the regulation of the 8 DEGs and actively functions in the invasive TNBC.

Section V. Experimental validation of the transcriptional regulation exerted by ERRα and ZEB1

RT-qPCR and ChIP experiments were then performed in the lab to verify the relationships between ERRα, ZEB1 and the 8 DEGs. This section includes the preparation of cells transfected by siRNA targeting each factor and the validation of the transcriptional activity of ERRα and ZEB1 on the 8 DEGs in breast cancer cells.

Materials and Method

1. siRNA transfection

MCF7 (ER+/PR+/Her2-), BT474 (ER+/PR+/Her2+), SKBr3 (ER-/PR-/Her2+) and MDA-MB231 (TNBC; mesenchymal-like subtype) cells were cultured in DMEM supplemented with 10% fetal calf serum (FCS), 10 U/ml penicillin and 10 µg/ml streptomycin. BT549 (TNBC; mesenchymal subtype), HCC38 (TNBC; basal-like 1 subtype) and MDA-MB468 (TNBC; basal-like 1 subtype) cells were cultured in RPMI supplemented with 10% FCS, 10 U/ml penicillin and 10 µg/ml streptomycin. For siRNA transfection, 3.10^{-5} cells per ml were seeded in 6-well plate and 25 pmol/ml of total siRNA were transfected with INTERFERin (Polyplus Transfection; Illkirch-Graffenstaden, France) according to the manufacturer's recommendations. siRNAs were from Invitrogen (Strasbourg, France).

2. ChIP

10^7 cells were cross-linked with 1% formaldehyde and quenched with 1 M glycine. After centrifugation, chromatin was extracted from cell pellets and immunoprecipitated using the iDEAL ChIP kit (Diagenode, Liège Belgium) according to the manufacturer's instructions. Quantitative PCRs were performed using 2 µl of DNA in duplicate and enrichment was calculated related to input.

Results and discussion

5.1 Transcriptional regulation of ERRα and ZEB1 on 8 DEGs in TNBC cells

To experimentally investigate the transcriptional regulation exerted by ERRα and ZEB1 on the 8 DEGs, MDA-MB231 breast cancer cells were transfected with siRNA targeting ERRα or ZEB1 separately and analyzed by Western blots, showing the high efficiency of the siRNAs (**Figure 39**). The effects of siRNA-mediated silencing of ERRα or ZEB1 on the expression of the 8 DEGs was next analyzed by qPCR. We noted a reduction of the expression of all the 8 DEGs except DDR2 (**Figure 40a, 40b**). As ZEB1 is a potential co-activator of ERRα, we next explored the co-regulation of ERRα and ZEB1 in cell models transfected with siRNA targeting both of ERRα and ZEB1. No additional decrease in the expression of the DEGs (again except DDR2) upon depletion of both ERRα and ZEB1 was observed as compared to the result of the depletion of each single factor (**Figure 40c**). This suggests that both of ERRα and ZEB1

can regulate the expression of the 8 DEGs through the same molecular pathways.

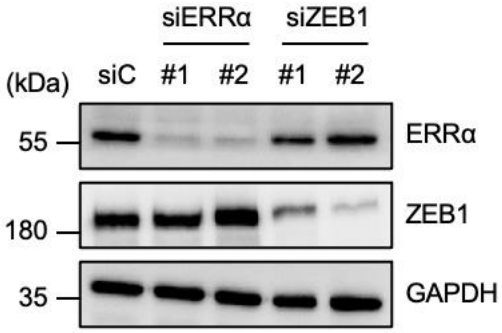


Figure 39. Protein expression of ERRα and ZEB1 in transfected cells. MDA-MB231 cells transfected with the indicated siRNAs were analyzed for ERRα and ZEB1 protein expression with GAPDH used as a loading control. Marker size is indicated.

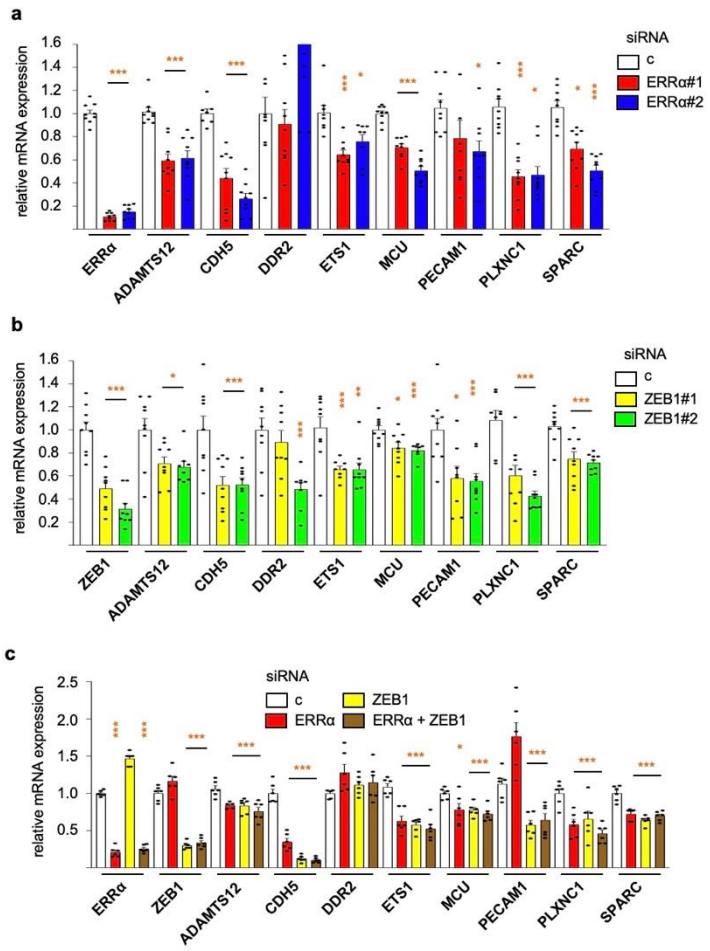


Figure 40. Common targets of ERRα and ZEB1. a. MDA-MB231 cells transfected with siRNAs directed against ERRα were analyzed by RT-qPCR for the expression of the

indicated genes. b. Same with MDA-MB231 cells transfected with siRNAs directed against ZEB1. c. Same with MDA-MB231 cells transfected with siRNA directed against ERRα and/or ZEB1 as indicated. Significance relative to control was evaluated by t-test with ***: p<0.005, **: p<0.01, *: p<0.05.

We then searched to validate this transcription regulation in other types of TNBC cells. Similar transfections with siRNA against ERRα or ZEB1 were performed in MDA-MB468 cells (BL1: basal-like 1 subtype), BT549 cells (M: mesenchymal subtype) and HCC38 cells (BL1: basal-like 1 subtype). Consistent results were observed in these three types of TNBC cells, showing a significant reduction of the expression of most of the 8 DEGs upon the depletion of ERRα or ZEB1 although some of the DEGs were not detected in some of the cell types (**Figure 41**).

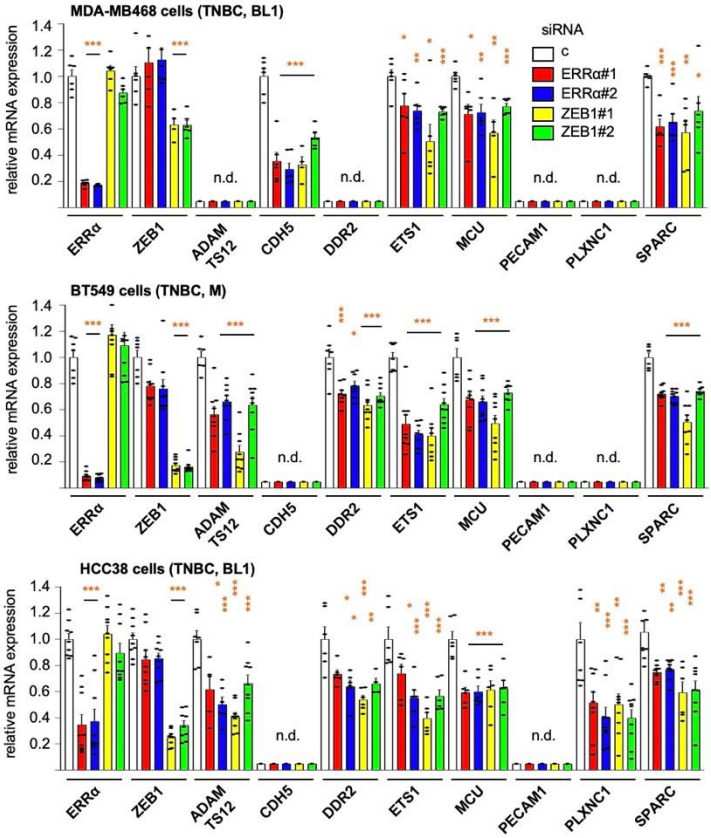


Figure 41. ERRα and ZEB1 shared targets in several TNBC cells. MDA-MB468, BT549 or HCC38 cells (all TNBC cells) transfected with the indicated siRNAs and analyzed for the expression of the indicated genes by RT-qPCR. BL1: basal-like 1 subtype; M: mesenchymal subtype according to (Lehmann et al., 2011). Significance relative to control was evaluated by t-test with ***: p<0.005, **: p<0.01, *: p<0.05. n.d.: not detected.

5.2 Transcriptional regulation of ERRα and ZEB1 on 8 DEGs in non-TNBC cells

In addition, we also investigated the cooperation of ERRα and ZEB1 on the 8 DEGs in other types of breast cancer cells. A preliminary RT-qPCR screen showed that ERRα is expressed in all types of breast cancer cells, including in TNBC cells. In contrast, ZEB1 is weakly or not expressed in non TNBC cells (MCF7, SKBr3 and BT474) but strongly in TNBC ones (MDA-MB231, BT-549, HCC38 and MDA-MB468) (**Table 6**).

	expression (Ct)		
	36b4	ZEB1	ERRα
MCF7	18.0 +/- 1.0	32.5 +/- 0.6	23.0 +/- 0.3
SKBr3	18.7 +/- 0.2	36.1 +/- 2.2	21.6 +/- 0.3
BT-474	17.0 +/- 0.6	30.9 +/- 1.5	23.1 +/- 0.2
MDA-MB231	17.5 +/- 0.1	22.4 +/- 0.2	23.8 +/- 0.2
BT-549	20.3 +/- 0.5	24.8 +/- 0.5	24.8 +/- 0.4
HCC38	19.9 +/- 0.2	28.0 +/- 0.3	26.5 +/- 0.1
MDA-MB468	19.5 +/- 0.4	28.5 +/- 0.3	23.2 +/- 0.6

Table 6. Expression of ERRα and ZEB1 in breast cancer cells. Expression of 36b4 (used as an endogenous control), ZEB1 and ERRα was evaluated by RT-qPCR in the indicated cell lines. Results are presented as mean of three independent experiments.

Interestingly, we observed the expression silencing of ERRα did not globally impact on the expression of the 8 DEGs in non TNBC cells while significantly influence the expression of other ERRα targets LMNB1 and NUDT19 in these three types of cells (**Figure 42**). Altogether, this demonstrates that the expression of the 8 DEGs is specifically modulated by ERRα and ZEB1 in TNBC cells, but not in non-TNBC cells.

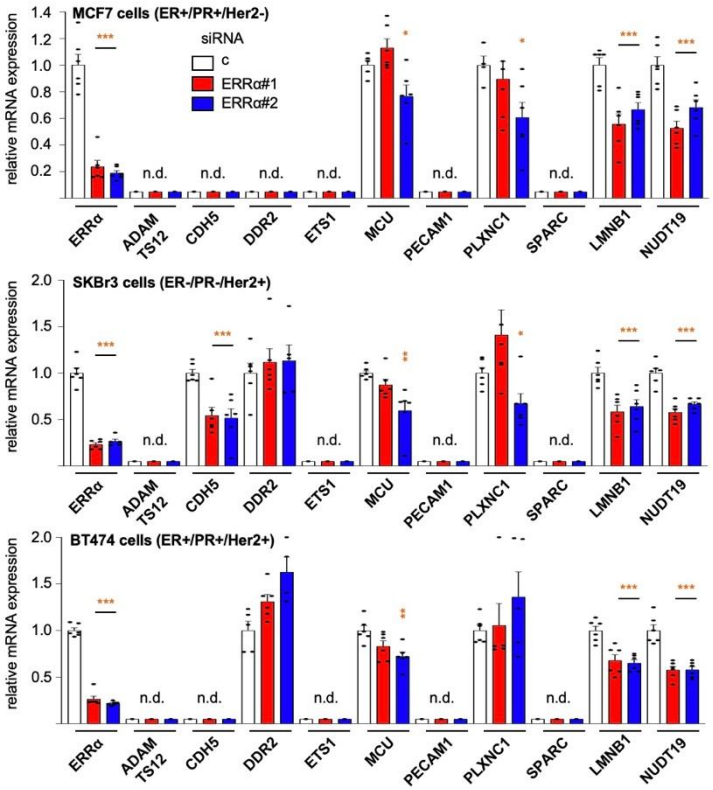


Figure 42. ERRα and ZEB1 shared targets in non-TNBC breast cancer cells. MCF7, SKBr3 or BT474 cells were transfected with siRNAs and analyzed for the expression of the indicated genes by RT-qPCR. Significance relative to control was evaluated by t-test with ***: p<0.005, **: p<0.01, *: p<0.05. n.d.: not detected.

5.3 Specificity of the 8 DEGs

Based on the validation of the cooperation of ERRα and ZEB1 in TNBC cells, we next evaluated the extent of their cooperative effect in these cells. We first evaluated the expression of four ZEB1 targets (Feldker et al., 2020). Expression reduction of CCN2 (CTGF), CCN1 (CYR61) and FOSL1, but not YAP1, upon the depletion of ERRα or ZEB1 was observed in three types of TNBC cells (MDA-MB231, MDA-MB468 and HCC38 cells) (**Figure 43**), indicating that ERRα could also regulate these four ZEB1-targest in some TNBC cells. In contrast, ZEB1 did not regulate the expression of four ERRα-only targets (GOT2, NOP2, SINHCAF or TMEM120B) (Cerutti et al., 2022) upon both of RT-qPCR and ChIP-qPCR experiments (**Figure 44, Figure 45**). This suggested the cooperation of ERRα and ZEB1 was limited to a set of genes which does not include all ERRα targets but may include additional ZEB1 targets.

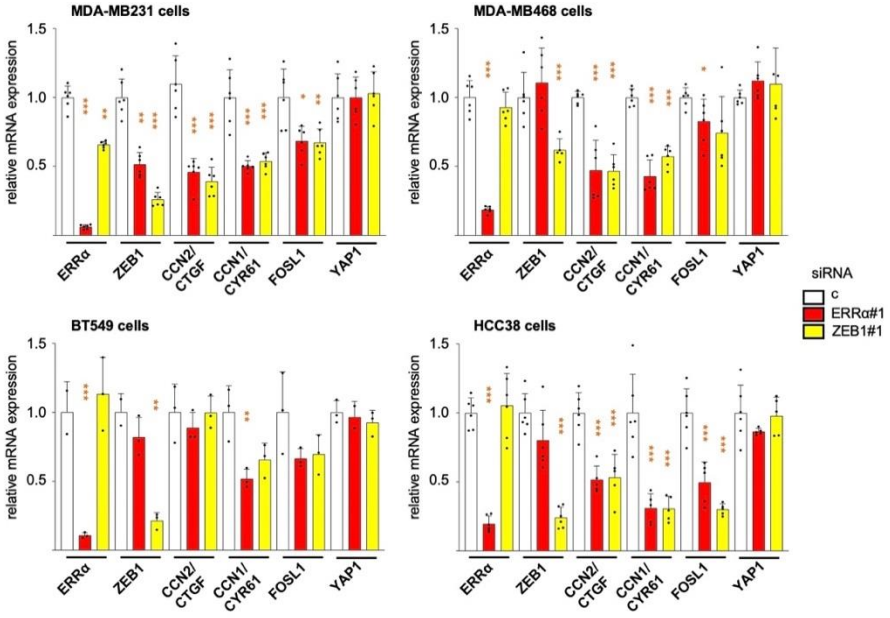


Figure 43. ZEB1-responsive genes in TNBC (mRNA). TNBC cells were transfected with the indicated siRNAs and analyzed for the expression of the indicated genes by RT-qPCR. Significance relative to control was evaluated by t-test with ***: p<0.005, **: p<0.01, *: p<0.05.

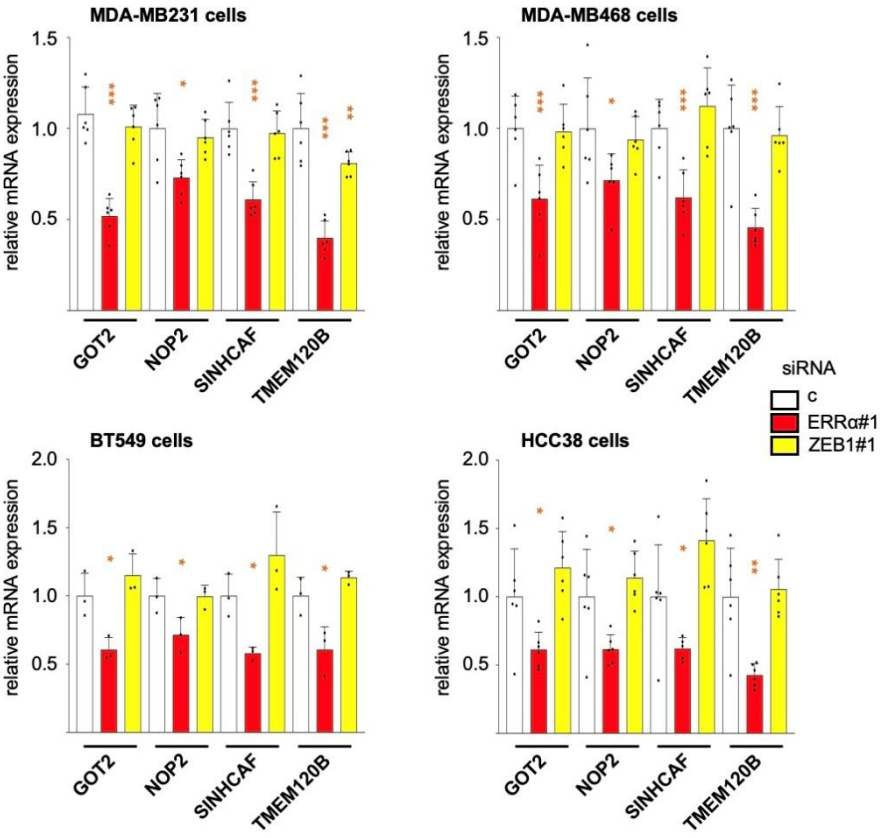


Figure 44. ERRα-responsive genes in TNBC (mRNA). TNBC cells were transfected with the indicated siRNAs and analyzed for the expression of the indicated genes by RT-qPCR. Significance relative to control was evaluated by t-test with ***: p<0.005, **: p<0.01, *: p<0.05.

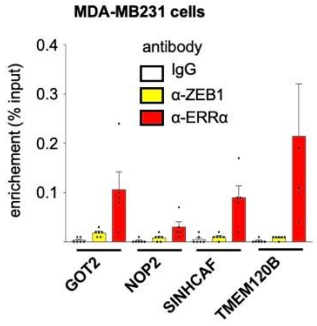


Figure 45. ERRα responsive genes in TNBC (ChIP). Binding of ERRα or ZEB1 to the indicated genomic sequences determined by ChIP-qPCR. Percent enrichments were determined amplifying a region comprising identified ERRα binding sites.

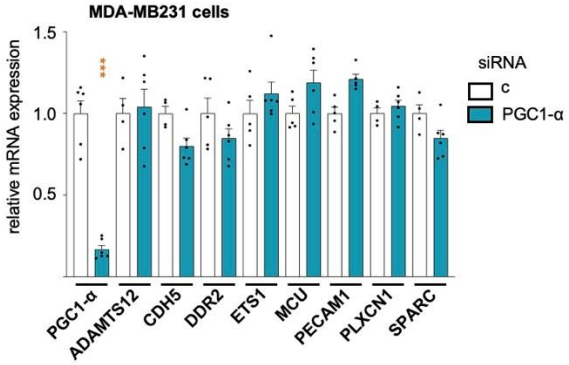


Figure 46. Expression of 8 DEGs in TNBC. MDA-MB231 cells were transfected with siRNA directed against PPARGC1A (encoding PGC-1α) or control and analyzed for the expression of the indicated genes by RT-qPCR. Significance relative to control was evaluated by t-test with ***: p<0.005, **: p<0.01, *: p<0.05.

In addition, PGC-1α, a well-known ERRα co-activator (Schreiber et al., 2003; Gaillard et al., 2007), had no impact on the expression of the 8 DEGs which also supported the observation that ZEB1 is a specific co-activator of ERRα on the regulation of the 8 DEGs (**Figure 46**).

In summary, ZEB1 is weakly expressed in non-TNBC breast cancer cells and ERRα does not regulate the expression of the 8 DEGs in these cells. In contrast, the mRNA expression level of the 8 DEGs was significantly regulated by ERRα and ZEB1 in all types of TNBC cells. The observation that ERRα and ZEB1 regulation is exerted within the same molecular pathway, strongly indicating their synergistic effect on the

expression of the 8 DEGs. Especially, the common targets of both ERRα and ZEB1 are restricted into a group of ERRα- and ZEB1-targets suggesting that the 8 DEGs are specific targets of ERRα in the cooperation of ZEB1, and not other co-regulators, in TNBC cells.

Section VI. Co-operation of ERRα and ZEB1 on expression of the 8 DEGs

We next explored the specific mechanisms used by ERRα and ZEB1 in the regulation of expression of the 8 DEGs. As indicated by our previously published ChIP-Seq data, ERRα can bind to the promoter regions of the 8 DEGs (Cerutti et al., 2022). This section includes the exploration of ZEB1-binding on the 8 DEGs as well as the cooperation of ERRα and ZEB1 on these genes.

Materials and Methods

Following the analysis pipeline in (Cerutti et al., 2022), ZEB1 published ChIP-seq data obtained in MDA-MB231 cells were re-analyzed from raw fastq sequencing files (50 or 100 bp paired-end sequencing using the Illumina HighSeq2500) (Feldker et al., 2020) including two anti-ZEB1 experiment replicates and two control inputs. Adapter trimming and quality filtered were performed by TrimGalore! through Cutadapt and FastQC tools. Trimmed Reads were mapped to the human reference genome (GRCh38) using Bowtie2 (paired-end). Sam files were converted to bam file format using Samtools and then sorted by Picard based on SortSam tool. Read duplicates were removed using Picard's MarkDuplicates function. Peaks were called, respectively, on each anti-ZEB1 replicate by MACS2 software using merged input files. The consistency of two anti-ZEB1 replicates was assessed using IDR tool. Annotations of identified peak were conducted by R package ChIPseeker based on searching peaks in the range of ± 100 kb to gene transcription start site (TSS) (keep both gene and transcript features). Together, 8,579 ZEB1-localized genomic regions annotated to 7,403 genes were found. Peak annotation profiles were generated using R package ChIPseeker. The location of conserved binding motifs (IDR < 0.05) of ERRα and ZEB1 on the 8 DEGs was determined and displayed based on the distance to gene TSS in R package ggplot2.

Results and discussion

6.1 ZEB1 ChIP-Seq results

We first reviewed the distribution of ZEB1 binding sites in different genomic features. 8,579 peaks annotated to 7,403 genes were distributed in all genome regions in which more than half (55.25%) in the promoter regions (**Figure 47**). We further observed that the highest TF-binding intensity was around the transcription start sites (TSS) (**Figure 47, 48**), and the binding sites mainly distributed in the region of ± 1 kb to TSS (**Figure 49**). This result indicates that ZEB1 is more likely to regulate gene transcription by acting on promoter regions close to the TSS.

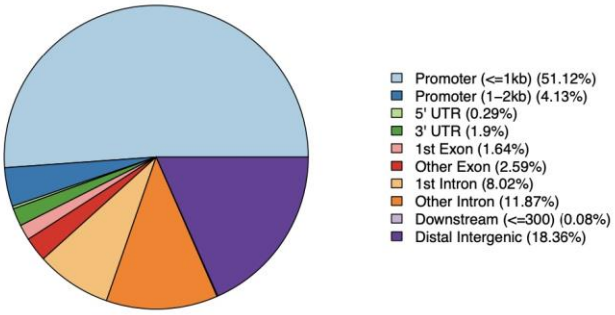


Figure 47. Genomic features of ZEB1 binding sites in the whole genome.

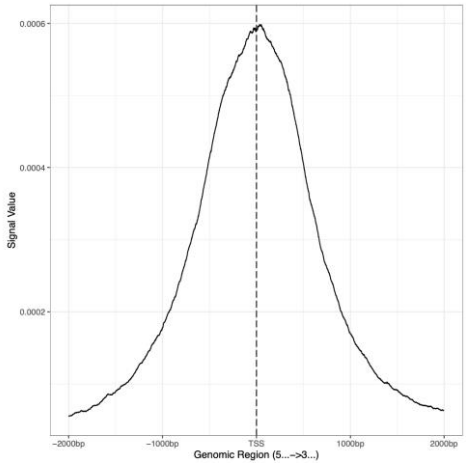


Figure 48. Binding intensity of ZEB1 from TSS.

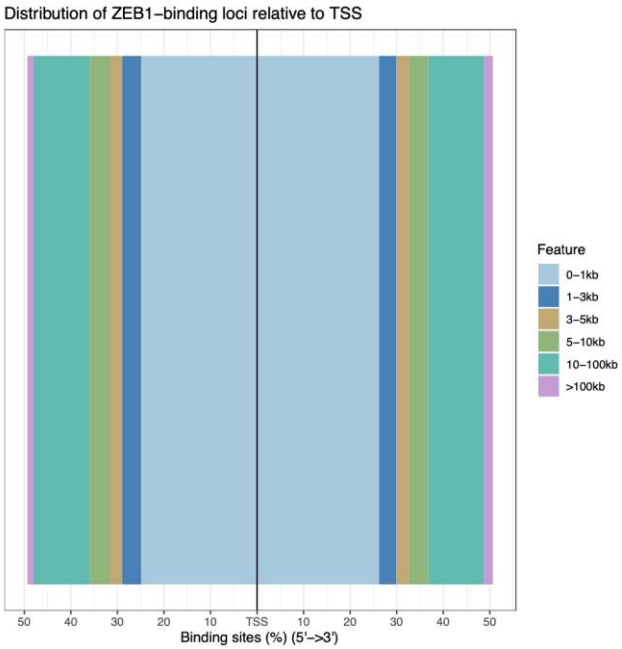


Figure 49. Distribution of ZEB1-binding site relative to TSS.

Next, we checked the distribution of binding sites of ERRα and ZEB1 on the 8 DEGs. We obtained the analysis results of the ERRα ChIP-Seq experiment from previous studies (Cerutti et al., 2022). We then selected out the peaks from the binding sites of ERRα and ZEB1 near the TSSs of the 8 DEGs. We found that the distribution of all ERRα binding sites on the 8 DEGs were far away from each other, whereas ZEB1 only binds to near 10kb upstream of TSS of ADAMTS12 gene (**Figure 50**). This suggests that ZEB1 only directly transcriptionally regulates ADAMTS12, but not the other seven DEGs. This result also suggests that ZEB1 is dependent on ERRα to regulate these genes in breast cancer.

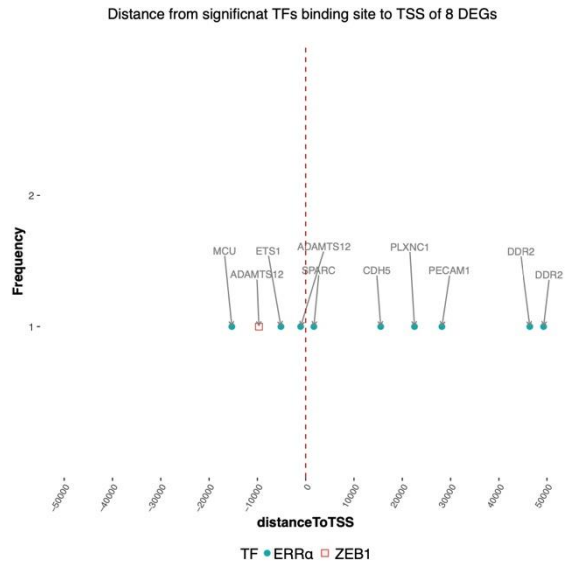


Figure 50. Distance of the significant peaks (IDR < 0.05) of ERRα and ZEB1 to the TSS on the 8 DEGs. Green circles represent the binding sites of ERRα to TSS on each DEG. Red box represents the binding site of ZEB1 to TSS on each DEG. Frequency represents the number of the binding sites.

6.2 Cooperation mechanism through the binding of ERRα and ZEB1 on 8 DEGs

Previous ChIP-Seq experiments confirmed the binding sites of ERRα through ERREs (ERRα response elements, 5'-TCAAGGTCA-3') on the promoter region of the 8 DEGs (Cerutti et al., 2022). ChIP-qPCR experiments were performed again to verify these binding sites of ERRα (**Figure 51a**). Surprisingly, ChIP-qPCR experiments also showed binding of ZEB1 on the same ERRα-binding region on the 8 DEGs (**Figure 51b**), where no conventional direct ZEB1 binding sites (5'-CAGGTG/A-3') were observed. All of this raised the hypothesis that the binding of ERRα and ZEB1 on the 8 DEGs may depend on each other.

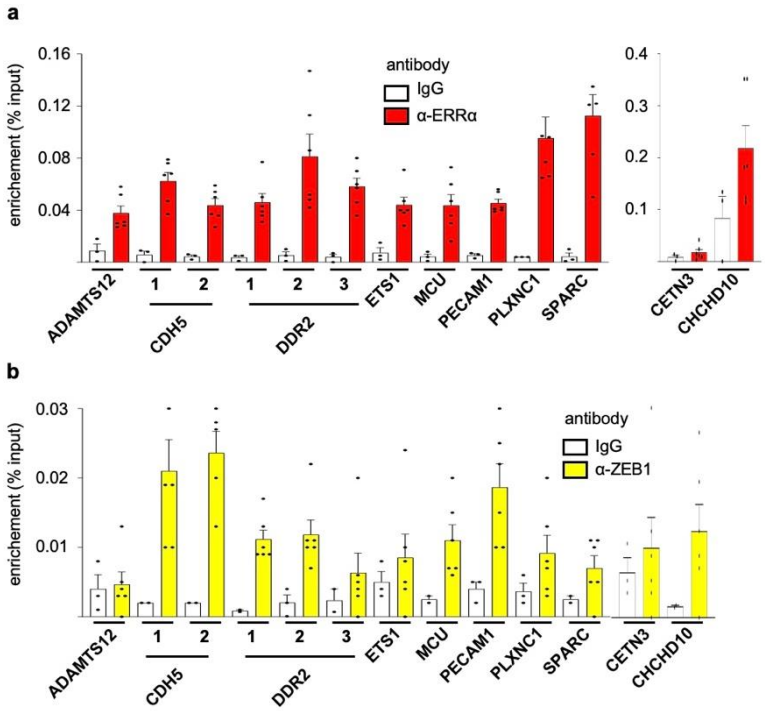


Figure 51. ERRα and ZEB1 bind to common DNA sites on DEG promoters. a. Binding of ERRα to genomic sequences identified by ChIP-Seq was confirmed by ChIP-qPCR experiments performed on independent samples. b. Binding of ZEB1 to the ERRα-binding sites was analyzed as in a. For ChIP experiments, percent enrichments were measured by qPCR amplifying a region encompassing the identified ERRα-binding sites. CETN3 and CHCHD10 represent negative and positive (respectively) controls for ERRα binding. IgG was used as a negative control.

Therefore, we next performed ChIP-qPCR experiments in the absence of ERRα or ZEB1 obtained by siRNA treatment. Interestingly, a decreased binding of ZEB1 on most of the 8 DEGs was detected upon the depletion of ERRα (**Figure 52a**). In contrast, silencing of ZEB1 did not reduce ERRα-binding on the 8 DEGs (**Figure 52b**), indicating ZEB1 can bind to the genomic region of 8 DEGs in the dependence of ERRα while for ERRα, ZEB1 is not necessary. This also suggests that ERRα and ZEB1 interact with each other. This hypothesis was confirmed by proximity ligation assays (PLA) which documented physical contacts between the two factors in the cell nuclei of MDA-MB231 breast cancer cells (**Figure 53**).

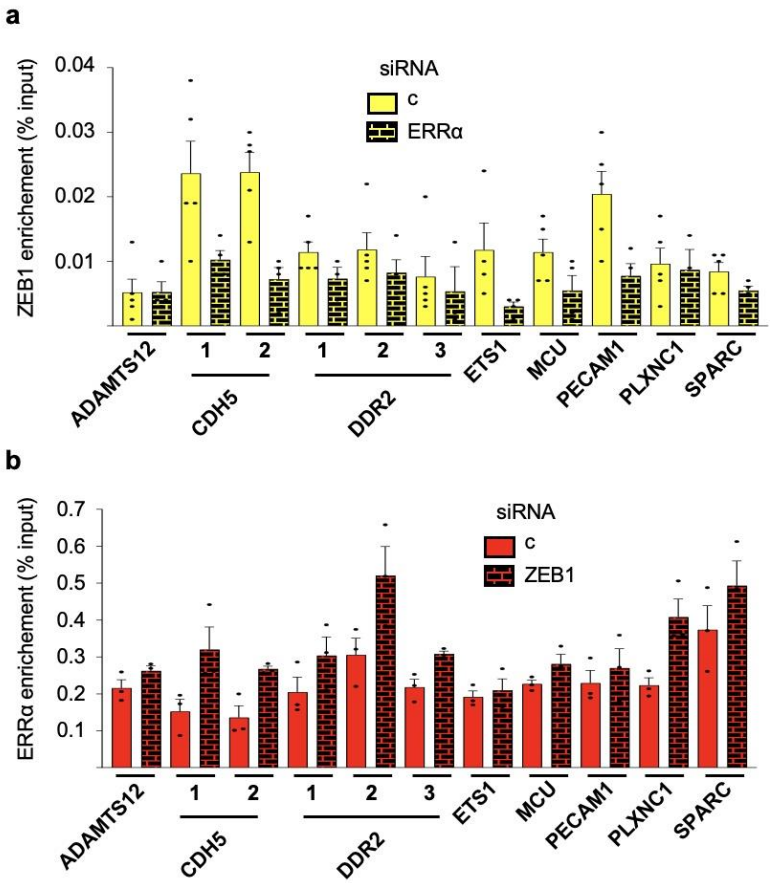


Figure 52. ERRα and ZEB1 interact with each other and bind to common DNA sites on DEG promoters. a. Binding of ZEB1 determined by ChIP-qPCR after siRNA-mediated ERRα depletion. b. Binding of ERRα determined by ChIP-qPCR after siRNA-mediated ZEB1 depletion.

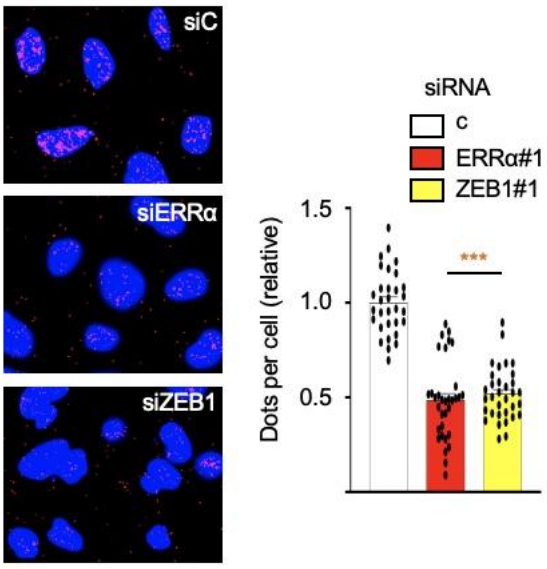


Figure 53. ERRα and ZEB1 interact with each other. Proximity ligation assays (PLA) used to detect interaction of the endogenous ERRα and ZEB1 in siRNA-transfected MDA-MB231 cells. Cells were counterstained with DAPI. Dot quantifications were performed on at least 8 fields from three independent experiments and are expressed relative to siC conditions. Error bars represent sem. Significance relative to control was evaluated by t-test with ***: $p<0.005$.

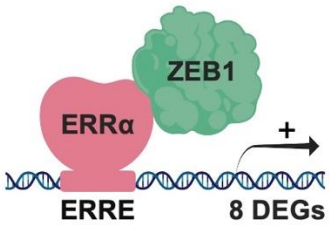


Figure 54. Schematic representation summarizing the ERRα-ZEB1 interactions on activated target genes. ERRα directly regulates the expression of the 8 DEGs through ERRE (ERRα response element) on the promoter regions. ZEB1 indirectly regulates the expression of the 8 DEGs in an ERRα-dependent manner. The factors display physical interactions.

In summary, the cooperation mechanism of ERRα and ZEB1 on the 8 DEGs was in the way of direct physical interaction. Specifically, the direct binding of ERRα on the 8 DEGs as well as the binding of ZEB1 in an indirect manner through the same binding genomic regions strongly suggested that the regulation of ZEB1 on the 8 DEGs depends on ERRα (**Figure 54**).

Section VII. Expression of 8 DEGs in tumor microenvironment

Our data above demonstrate a higher expression of ZEB1 and the 8 DEGs in TNBC as well as a specific regulation of the expression of these genes by ERRα and ZEB1 in TNBC cells. As tumors are composed of cancer cells and microenvironment (stromal cells, immune cells…), we next investigated the expression of indicated genes in non-cancer cells present in the tumor microenvironment in order to complete our understanding. This section includes the detection of the expression of indicated genes in microenvironment of TNBC, ER+ and Her2+ tumors.

Materials and Method

Pal et al. (2021) performed single-cell sequencing on 69 different tissue samples (normal tissue and tumors) from 55 patients and finally constructed a single-cell map displaying clustering results of nearly 430,000 single cells (**Figure 55**).

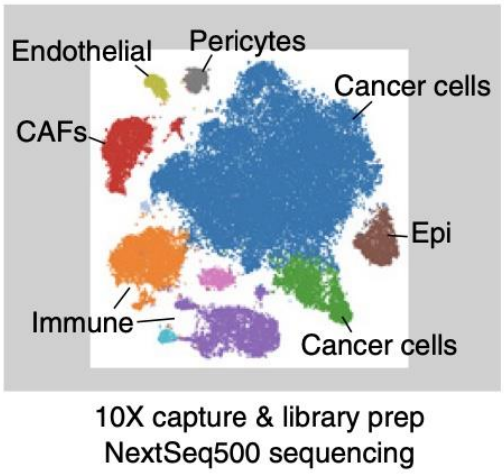


Figure 55. Expression profile for nearly 430,000 single cells from 55 patients from (Pal et al., 2021).

After quality control and filtration, nearly 342,000 cells of good quality were obtained and used for subsequent analysis. In order to explore the expression of ESRRA, ZEB1 and the 8 DEGs in the microenvironment of breast cancer tumors, we first selected all single cells of eight triple-negative breast cancer tumors (from patients who had not received treatment), 13 ER+ and six Her2+ tumors, and then performed clustering analysis respectively.

Based on the classification criteria of Pal et al. (2021), cells after clustering were divided into epithelial cells and non-epithelial cells (microenvironment) according to the expression levels of CD49f and CD326 (EpCAM). We next removed EpCAM+ and normal epithelial cells. The remaining stromal/immune cells (microenvironment) were

regrouped and re-clustered.

Cell type annotation comes from (Pal et al., 2021).

Expression maps and dot plots were generated using Seurat package in R software.

Results and discussion

7.1 Expression of genes in the microenvironment of TNBC tumors

There are about 12 major cell types in human normal breast tissue, including epithelial cells, lymphocytes, T cells, B cells, fat cells, fibroblasts, and perivascular cells (Kumar et al., 2023), hence, the development of breast cancer may also be influenced by these non-cancer cells to a certain extent (Chung et al., 2017; Pal et al., 2021).

The clustering results on all cells from different subtype tumors were shown in (**Figure 56a, 58a, 60a above left**). According to single-cell sequencing data from TNBC, ZEB1 had lower expression in cancer cells and higher expression in stromal/immune cells (non-epithelial cells). The expression of the 8 DEGs (except MCU) was similar to that of ZEB1 (i.e., enriched in stromal/immune cells), whereas the expression of ESRRA was scattered across all cell types (**Figure 56a**). Next, we removed epithelial cells and re-clustered the remaining stromal/immune cells. We found that ESRRA had no specific expression, but ZEB1 was mainly detected in cancer-associated fibroblasts (CAFs), endothelial and pericytes cells and highly expressed (**Figure 56b**). Although ZEB1 also had obvious expression in parts within B cells and T cells, the number of expressing cells is limited and most B cells and T cells do not express ZEB1. The percent of ZEB1-expressing cells in B cells and T cells is thus smaller than that in CAFs, endothelial and pericytes (**Figure 56b, 57**). Overall, the 8 DEGs were highly expressed in at least one cell type of CAFs, endothelial or pericytes, which was basically consistent with the expression pattern of ZEB1.

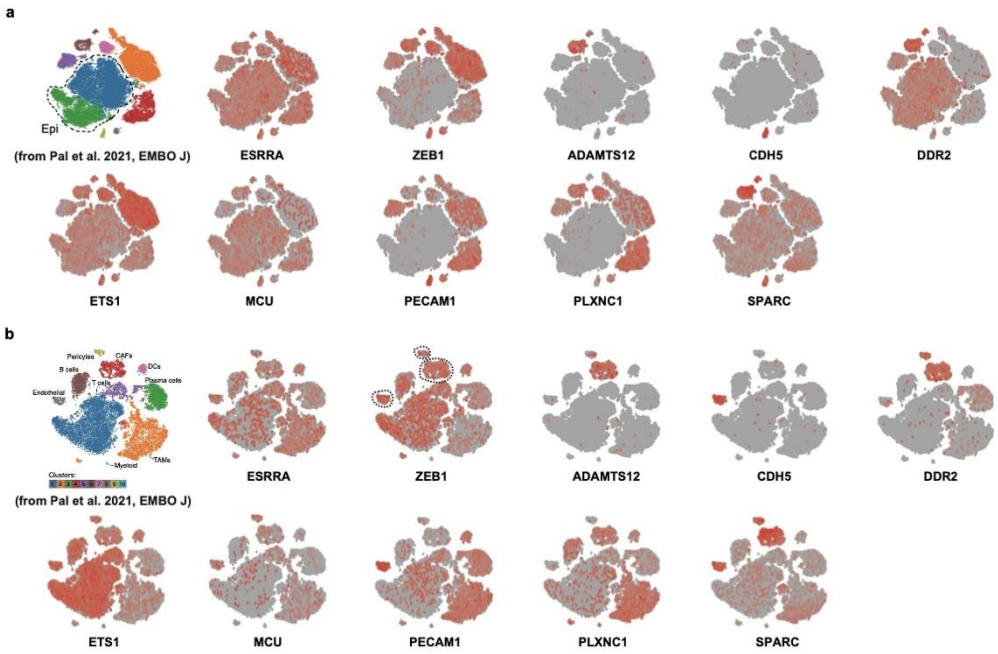


Figure 56. Expression of ESRRA, ZEB1 and the 8 DEGs in single cells from TNBC tumors. a. t-SNE map of scRNA-Seq profile of total cells from eight TNBC showing the expression of indicated genes of total cells. All cells colored by cluster (above left panel) in which epithelial cells are grouped with dotted lines. b. t-SNE map of the re-clustered non-epithelial cells identified in a, showing the expression of indicated genes of cells in microenvironment. All cells colored by cluster (above left panel). Cancer-associated fibroblasts (CAFs), endothelial and pericytes cells are indicated with dotted lines.

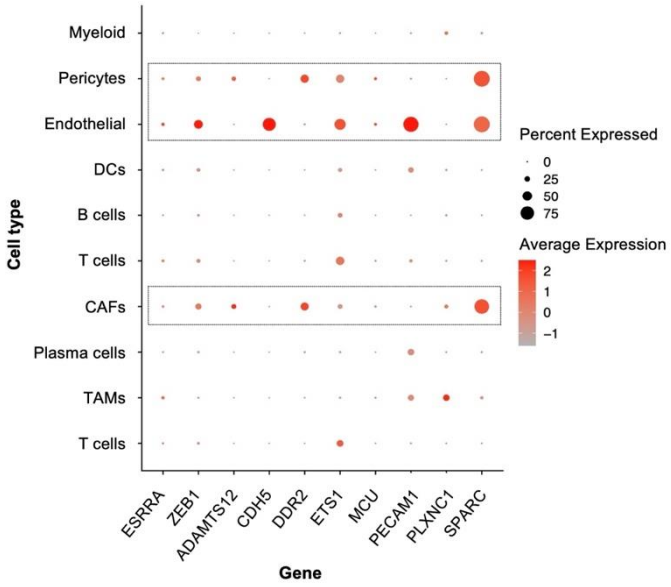


Figure 57. Dot plot showing the expression of indicated genes across non-epithelial cells in the microenvironment of TNBC tumors. Dot size represents the proportion of

cells expressing a specific gene. Color intensity represents the relative expression of specific genes.

7.2 Expression of genes in the microenvironment of ER+ and Her2+ tumors

In addition, we also analysed the single-cell data of ER+ and Her2+ tumors. We found that the expression pattern of ESRRA, ZEB1, and the 8 DEGs (except MCU) in the total cells and tumor microenvironment was highly similar to that observed in TNBC. ESRRA was expressed in all cell types, while ZEB1 and 8 DEGs were specifically highly expressed in at least one of CAFs, endothelial and pericytes cells (**Figure 58b, 59, 60b, 61**).

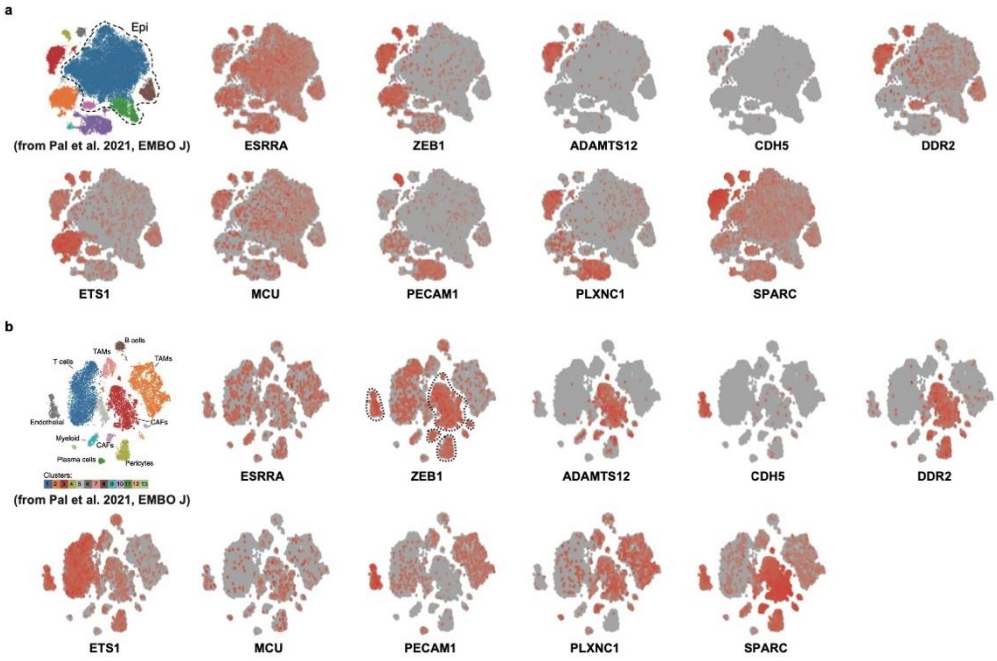


Figure 58. Expression of ESRRA, ZEB1 and the 8 DEGs in single cells from ER+ tumors. a. t-SNE map of scRNA-Seq profile of total cells from 13 ER+ tumors showing the expression of indicated genes of total cells. All cells colored by cluster (above left panel) in which epithelial cells are grouped with dotted lines. b. t-SNE map of the re-clustered non-epithelial cells identified in a, showing the expression of indicated genes of cells in microenvironment. All cells colored by cluster (above left panel). Cancer-associated fibroblasts (CAFs), endothelial and pericytes cells are indicated with dotted lines.

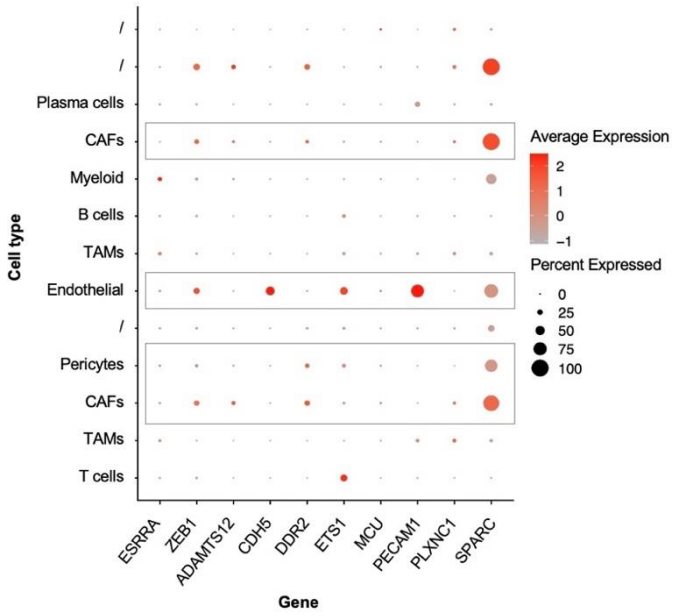


Figure 59. Dot plot showing the expression of indicated genes across non-epithelial cells in the microenvironment of ER+ tumors. Dot size represents the proportion of cells expressing a specific gene. Color intensity represents the relative expression of specific genes.

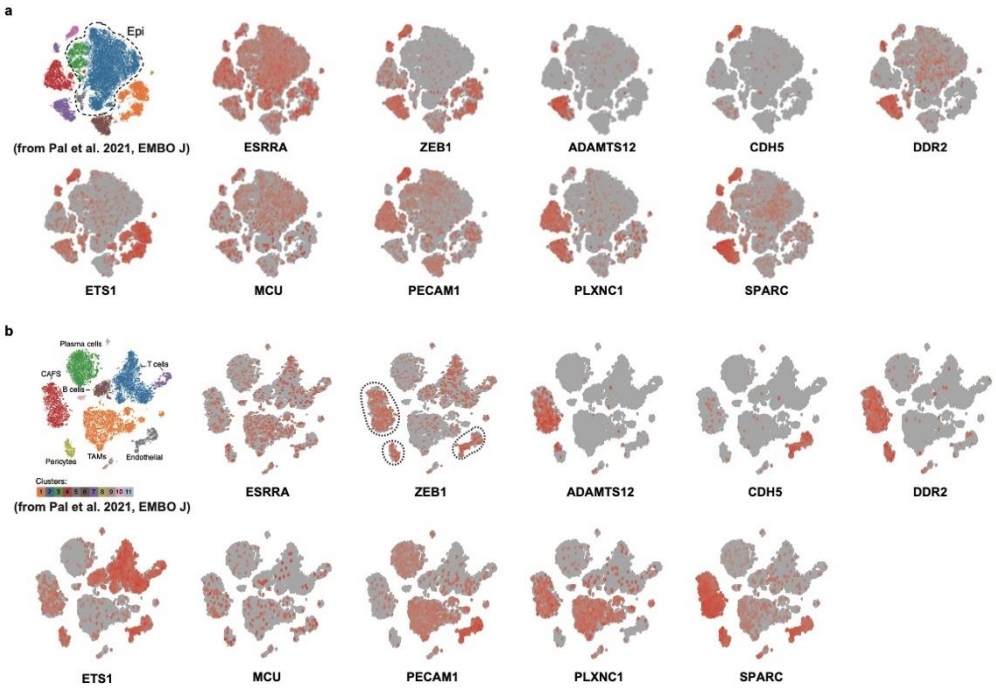


Figure 60. Expression of ESRRA, ZEB1 and the 8 DEGs in single cells from Her2+ tumors. a. t-SNE map of scRNA-Seq profile of total cells from six Her2+ tumors showing the expression of indicated genes of total cells. All cells colored by cluster (above left panel) in which epithelial cells are grouped with dotted lines. b. t-SNE map of the re-clustered non-epithelial cells identified in a, showing the expression of

indicated genes of cells in microenvironment. All cells colored by cluster (above left panel). Cancer-associated fibroblasts (CAFs), endothelial and pericytes cells are indicated with dotted lines.

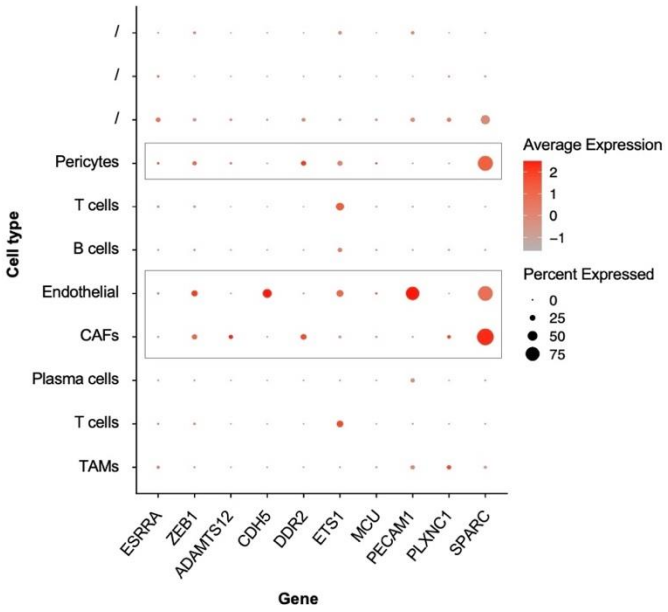


Figure 61. Dot plot showing the expression of indicated genes across non-epithelial cells in the microenvironment of Her2+ tumors. Dot size represents the proportion of cells expressing a specific gene. Color intensity represents the relative expression of specific genes.

Breast cancer types showed a considerable diversity when examining tumor cells and non-cancer cells. Despite the obvious diversity, we still found that ZEB1 and the 8 DEGs (except MCU) had specific expression enrichment in at least one of the CAFs, endothelial or pericytes cells, while the expression of ESRRA was scattered across all breast tumors. This suggests that they may be involved in the progression of breast tumors by showing a high physiological activity in the tumor microenvironment among all subtypes of breast cancer.

Based on the significant expression of ZEB1 and the 8 DEGs in TNBC cells (**Figure 40, Table 6**), on the fact that CAFs, endothelial and pericytes are highly involved in the EMT process (see the discussion section below), as well as the fact that ZEB1 is a key factor of EMT (Burk et al., 2008; Gregory et al., 2008; Caramel et al., 2018), we next explored whether the 8 DEGs are also related to the EMT process in breast tumors, in order to investigate the potential pathological and physiological pathways they may participate to.

Section VIII. EMT in breast cancer

Based on the potential function of ERRα, ZEB1 and the 8 DEGs in EMT process, we next quantified the EMT status of breast tumors. We also explored the correlations between these genes and the EMT process. This section includes the quantification of the EMT status of tumors, the correlation between the expression of indicated genes and EMT status and the comparison of EMT status between breast cancer subtypes.

Materials and Method

1. Public expression dataset

METABRIC

We downloaded human breast cancer mRNA expression z-scores (log microarray) and clinical features from cBioPortal (https://www.cbioportal.org) (Curtis et al., 2012). We obtained the expression profile of 2509 human breast primary tumors with clinical annotations.

FUSCC TNBC & TCGA TNBC

Same molecular subtype information of FUSCC TNBC and TCGA TNBC are from (Jiang et al., 2019).

2. Quantification of EMT status

To explore the correlation between gene expression and EMT progression, we used two unique methods to calculate an EMT score. With the first one called KS method, we quantified the EMT state for each breast cancer tumor sample (Chakraborty et al., 2020). First, 143 epithelia (Epi) signatures and 170 mesenchymal (Mes) signatures from tumors were collected and used to calculate their cumulative distribution function (CDF). Then, the EMT score was calculated and obtained based on the maximum distance between CDFs of Epi signatures and Mes signatures by using two-sample Kolmogorov-Smirnov test. Positive score for a tumor characterized a mesenchymal status and negative score was explained as epithelial status.

The second method is called the 76GS method. 76-gene expression signature reported by Byers et al. (2013) was used to calculate a weighted sum as EMT score for each tumor sample based on the method from (Guo et al., 2019; Chakraborty et al., 2020):

$$EMT\ score_j = \sum_{i=1}^{76} \omega_i * G_{ij}$$

where ω_i is the correlation coefficient between the expression of gene i (one of 76 signatures) and CDH1 (encoding E-cadherin). G_{ij} is the normalized expression of gene i in tumor sample j. Positive score for a tumor was explained as epithelial status and negative score was explained as mesenchymal status which is contrary to the explanation of the KS method.

Pearson correlation test was used to identify the correlation between EMT score and gene expression (log-transformed). Wilcox test was used to identify the EMT score difference between each pair of TCGA subtypes.

Scatter plots were generated using ggplot2 package. Correlograms were generated using ellipse package. Box plots were generated using ggplot2 package in R software.

Results and discussion

8.1 Correlation of the expression of the indicated genes and EMT status.

We first explored the relationship between gene expression and EMT score in TCGA all breast cancers using KS method. We found that tumors with high ZEB1 expression had a higher EMT score (positive correlation), demonstrating they were more likely to be in a mesenchymal state. We also found that the expression of the 8 DEGs (except MCU) individually had a positive correlation with EMT score, which was similar to that of ZEB1. However, there was a negative correlation between the expression of ESRRA and the EMT score (**Figure 62**). To verify this conclusion, we also checked the relationship in three additional independent breast cancer tumor datasets. The results were consistent with that in TCGA, showing that the expression of ZEB1 and the 8 DEGs (except MCU) positively correlated to EMT score (red) (**Figure 63**). We particularly collected four ZEB1 targets (Feldker et al., 2020) as the control group and we found that they always had the identical relationships similar to ZEB1. We noticed that ESRRA expression always showed a negative correlation with EMT score (blue). Therefore, we collected four additional ERRα targets (co-regulated by ERRα and other TFs, but not by ZEB1) as the second control set. The results showed these four ERRα targets also kept a negative correlation. In summary, tumors with high expression of ZEB1 and the 8 DEGs (except MCU) are more likely to display a mesenchymal phenotype as well as may participate to breast cancer metastasis and invasion. Particularly, the 8 ERRα-activated DEGs co-regulated by ZEB1 positively promote EMT process in breast tumors whereas four other ERRα targets co-regulated by other TFs, but not ZEB1 (Cerutti et al., 2022), showed an opposite correlation to EMT, suggesting that ZEB1 is specific for the expression of the 8 ERRα-targets to involved in EMT. It also should be noted that the expression of the 8 DEGs is directly regulated by ERRα although the EMT relationship of the 8 DEGs is similar to ZEB1. Based on our experiments above, we concluded that the transcriptional regulation exerted by ZEB1 on the 8 DEGs depends on ERRα. Both ERRα and ZEB1 are important for the 8 DEGs to participate to the EMT process and maybe to contribute to breast cancer progression.

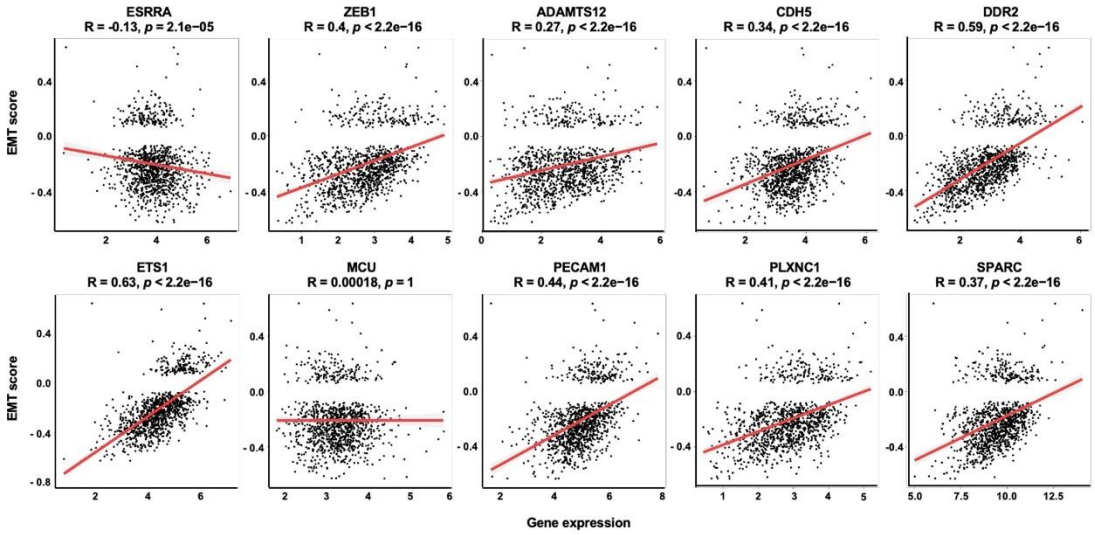


Figure 62. Scatter plots linking the expression of the indicated genes to the EMT score in TCGA all breast cancer tumors (KS method). Linear regression is estimated for each gene and plotted as a red line. Pearson's correlation coefficient and p-values are indicated on the top of each plot.

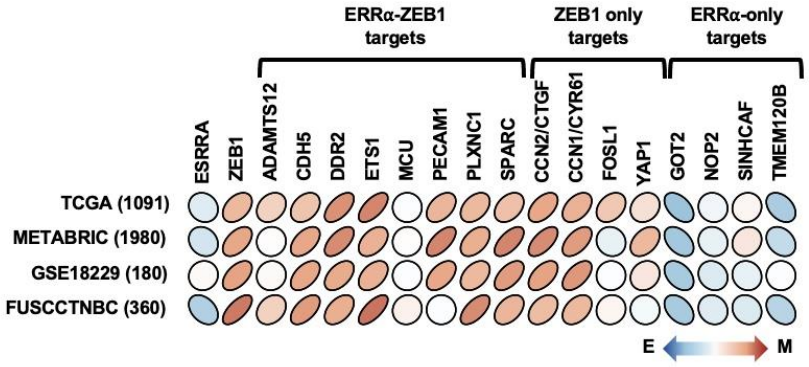


Figure 63. Correlogram showing the relation between expression (log-transformed) of the indicated genes and the EMT score in four independent breast cancer patient datasets (KS method). Positive score for a tumor was explained as mesenchymal status and negative score was explained as epithelial status. Pearson's correlation tests were used. Number of tumors in each dataset is shown. E: epithelial status; M: mesenchymal status.

Furthermore, we additionally explored the relationship in different breast cancer subtypes. We observed similar results in each subtype as before in all tumors, displaying samples with high expression of ZEB1 or the 8 DEGs tend to present a mesenchymal status regardless of breast cancer subtypes (**Figure 64**). In addition, we compared the differences of tumor EMT status between different breast cancer subtypes. We found that TNBC displayed significantly higher EMT scores than other subtypes, which is

consistent with the fact that TNBC is the most aggressive subtype of breast cancer (**Figure 65**).

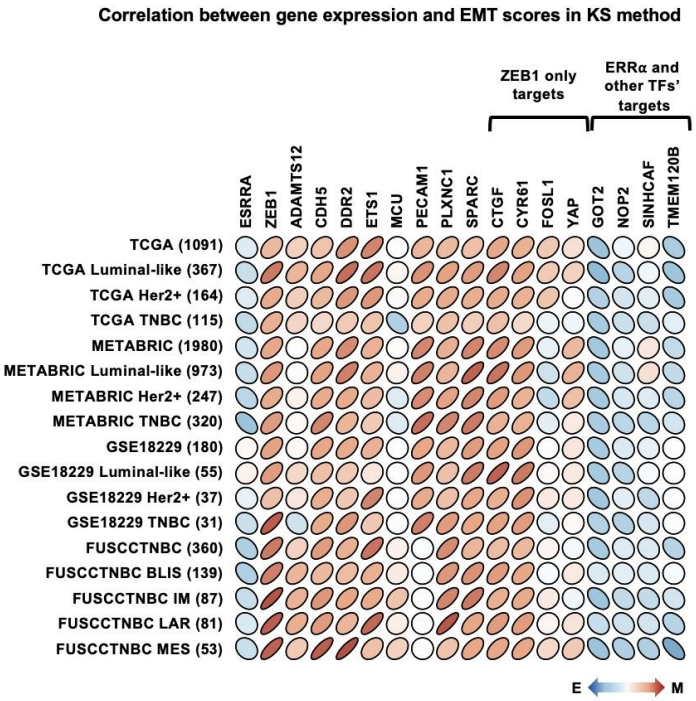


Figure 64. Correlogram showing the relation between expression (log-transformed) of the indicated genes and the EMT score in different subtypes of four independent breast cancer patient datasets (KS method). Positive score for a tumor indicates a mesenchymal status and negative score indicates an epithelial status. Pearson's correlation tests were used. Number of tumors in each dataset is shown. E: epithelial status; M: mesenchymal status.

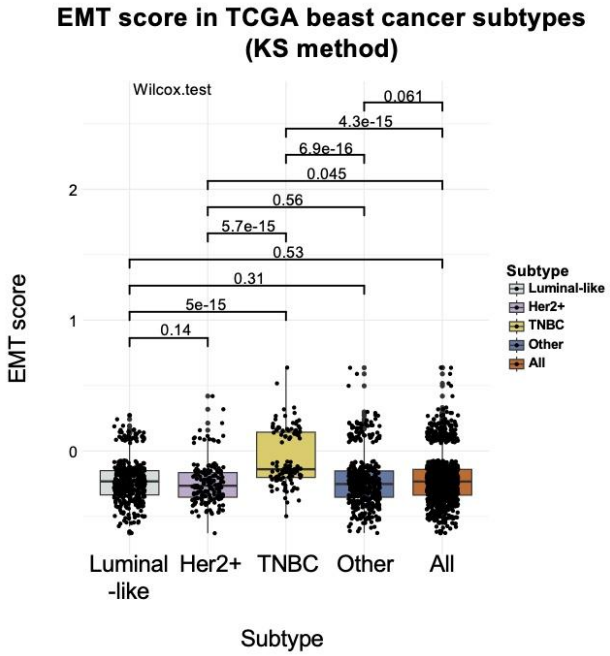


Figure 65. Box plot showing the comparison of EMT score between breast cancer subtypes from TCGA (KS method). Wilcox test was used to detect the difference between each pair of groups.

All these discoveries were also verified based on the 76GS method. 76GS method calculates the relationship between 76 signatures with CDH1 (encoding E-cadherin) based on the fact that the most important feature of EMT is the reduction of E-cadherin expression, promoting cell movement and metastasis. Therefore, tumors assigned a smaller EMT score by 76GS method are more likely to be in a mesenchymal status which is opposite to the results generated with the KS method. The results showed the negative correlation between the expression of ZEB1 and the 8 DEGs (except MCU) with EMT indicated that tumors with higher expression of these genes are characterized by a weakened epithelial phenotype but more inclined to present a mesenchymal phenotype (blue) (**Figure 66**), as expected with the conclusion of KS method. Same, TNBC kept a significantly more aggressive status (**Figure 67**).

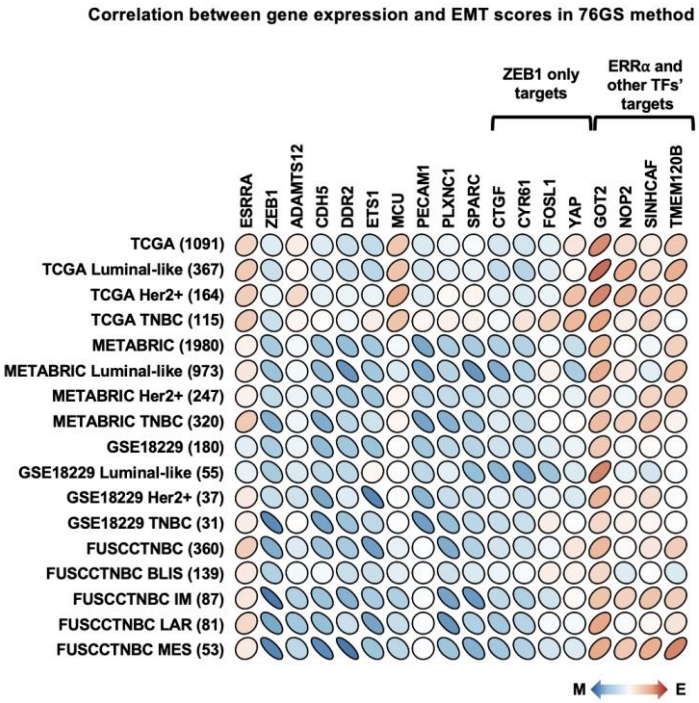


Figure 66. Correlogram showing the relation between expression (log-transformed) of the indicated genes and the EMT score in different subtypes of four independent breast cancer patient datasets (76GS method). Positive score for a tumor indicates an epithelial status and negative score indicates a mesenchymal status. Pearson's correlation tests were used. Number of tumors in each dataset is shown. E: epithelial status; M: mesenchymal status.

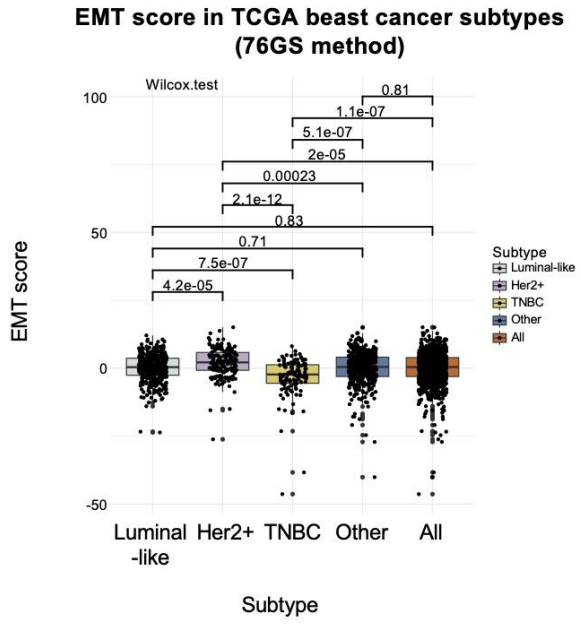


Figure 67. Box plot showing the comparison of EMT score between breast cancer subtypes from TCGA (76GS method). Wilcox test was used to detect the difference

between each pair of groups.

8.2 Comparison of EMT status between breast cancer subtypes

Combining the results of the two methods, we concluded that in the presence of ERRα, breast tumors with high expression of ZEB1 and the 8 DEGs (except MCU) were more likely to display a mesenchymal phenotype and EMT metastatic characteristics. TNBC kept the most significant EMT characteristics (**Figure 65, 67**) which is consistent with the aggressiveness of TNBC.

TNBC show high tumor heterogeneity. TNBC can be classified into different subtypes according to the molecular characteristics of tumors (Lehmann and Pietenpol, 2015; Jiang et al., 2019). We next explored the EMT differences between TNBC subtypes of FUSCC TNBC and TCGA TNBC tumors. Results demonstrated that IM and MES subtypes had significantly higher EMT score compared to the BLIS and LAR subtypes (**Figure 68**). The MES subtype is characterized by overexpression of genes related to breast cancer stem cells and up-regulation of JAK/STAT3 signaling pathway with both of these factors play an important role in the EMT process and metastasis (see the discussion section below). As we showed, ZEB1 and the 8 DEGs were significantly overexpressed in MES subtypes (**Figure 37, 38c, 38d**). This further suggests that tumors with high expression of ZEB1 and the 8 DEGs are extremely aggressive. A significant EMT correlation was not only shown in the aggressive TNBC tumors, but also in the MES subtype. We also suggest that not only the known ZEB1, but also the 8 DEGs can be used as potential targets to predict the invasiveness of tumors and help to prevent the metastasis of breast tumors in advance.

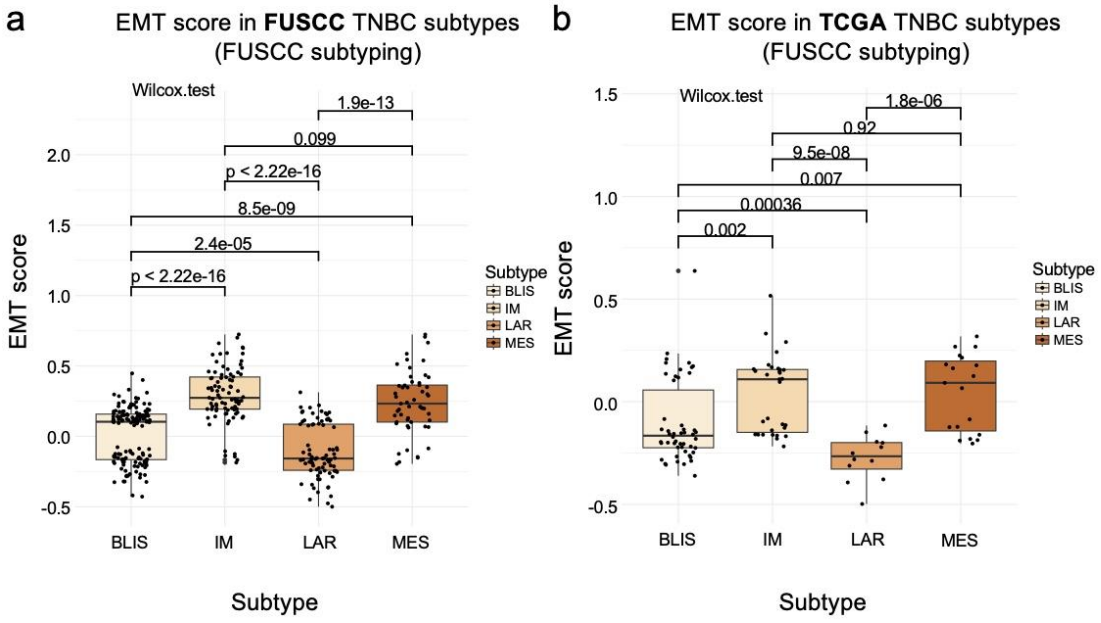


Figure 68. Box plots showing the comparison of EMT score between TNBC subtypes

of a. FUSCC and b. TCGA tumors (KS method). Difference between each pair of groups was evaluated by Wilcox test. P-value is presented on the top of the line. BLIS: basal-like and immune-suppressed; IM: immunomodulatory; LAR: luminal androgen receptor; MES: mesenchymal-like.

Section IX. Survival in breast cancer

A significant relation between the expression of ZEB1 and the 8 DEGs with the EMT status of breast tumors was detected. This suggested an involvement of these genes in cancer metastasis. We thus next investigated their predictive ability on the prognosis of breast cancer patients. This section includes the influence of indicated genes on the overall survival (OS), on the relapse free survival (RFS) and on the distant metastasis free survival (DMFS) in breast cancer patients.

Materials and Methods

1. Survival analysis in the GSE96058 dataset from KM Plotter database

Survival analysis of GSE96058 breast cancer RNA-seq dataset was performed using KM Plotter database (https://kmplot.com/analysis/) (Győrffy, 2024). For each gene, samples were divided into three terciles based on its expression. The top tercile group / bottom tercile group was assigned to "high" / "low" group separately. For the combination set of interested genes, mean expression of all genes was firstly calculated and then be used to classify samples to "high" / "low" group by using the same method in single gene above. Survival analysis in each breast cancer subtype was also conducted by restricting patients to different subtypes (Basal, LuminalA, LuminalB and Her2) in "PAM50 subtype" selection module.

2. Survival analysis in the TCGA dataset

Survival analysis of the TCGA breast cancer RNA-seq dataset was performed using the R package survival. For each gene, we labeled each sample with "high" or "low" based on the method used in previous research (Feldker et al., 2020) in which "high" sample is characterized by gene expression (upper-quartile normalized and then log-transformed) higher than the 60th percentile, while "low" sample is characterized by gene expression lower than the 40th percentile. For the combination set of interested genes, the "high" / "low" group comprises samples containing at least proper number (60% * the size of the combination set, for example, when we explore the survival outcome prediction probability of 8 DEGs together, the proper number is 60% * 8 = 5) of "high" / "low" labels except in two genes analysis in which proper number always to be 2. Kaplan-Meier curve was generated by statistical comparison between two groups based on log-rank test using survminer package in R software.

3. Survival analysis on the chip datasets from KM Plotter database

Survival analysis of a set of breast cancer chip datasets was also performed by KM Plotter database (https://kmplot.com/analysis/)(Győrffy, 2024). For the combination set of interested genes, mean expression of all gene probes was first calculated and then applied to classify samples to "high" / "low" group. The following microarray probes were used here: 221421_s_at (ADAMTS12), 204677_at (CDH5), 225442_at (DDR2), 214447_at (ETS1), 225320_at (MCU), 208983_s_at (PECAM1), 206470_at (PLXNC1) and 200665_s_at (SPARC).

Results and discussion

9.1 Influence of ERRα, ZEB1 and the 8 DEGs on overall survival

From **section VIII**, we observed that ZEB1 and the 8 DEGs are highly associated to the EMT process. We then hypothesized that they could also be used to predict survival in patients with breast cancer. In the human breast cancer RNA-Seq dataset (GSE96058) from the KM plotter database, we found that when ESRRA and ZEB1 were isolated or combined, they cannot evaluate overall survival (OS) of TNBC patients (named "Basal" in database). However, we surprisingly observed that the combined high expression of the 8 DEGs was associated with significantly worse prognosis (p = 0.048) in TNBC patients (**Figure 69a below**). This conclusion was obtained only in TNBC, and no OS differences were detected in all patients, luminal A, luminal B, and Her2+ subtypes (**Figure 69a above, 69b**). To validate this conclusion, we performed survival analysis in the TCGA RNA-Seq dataset. Similarly, the combined expression of the 8 DEGs showed the strongest correlation with worse OS only in TNBC group (**Figure 70**). Besides, we also surveyed chip datasets from the KM plotter database and obtained the consistent observation that TNBC patients with highly combined expression of the 8 DEGs were correlated to poor OS (**Figure 71**). Overall, TNBC patients with joint higher expression of the 8 DEGs significantly correlated with worse OS, as well as suggesting they can be used as a new potential set of targets to predict OS of TNBC patients.

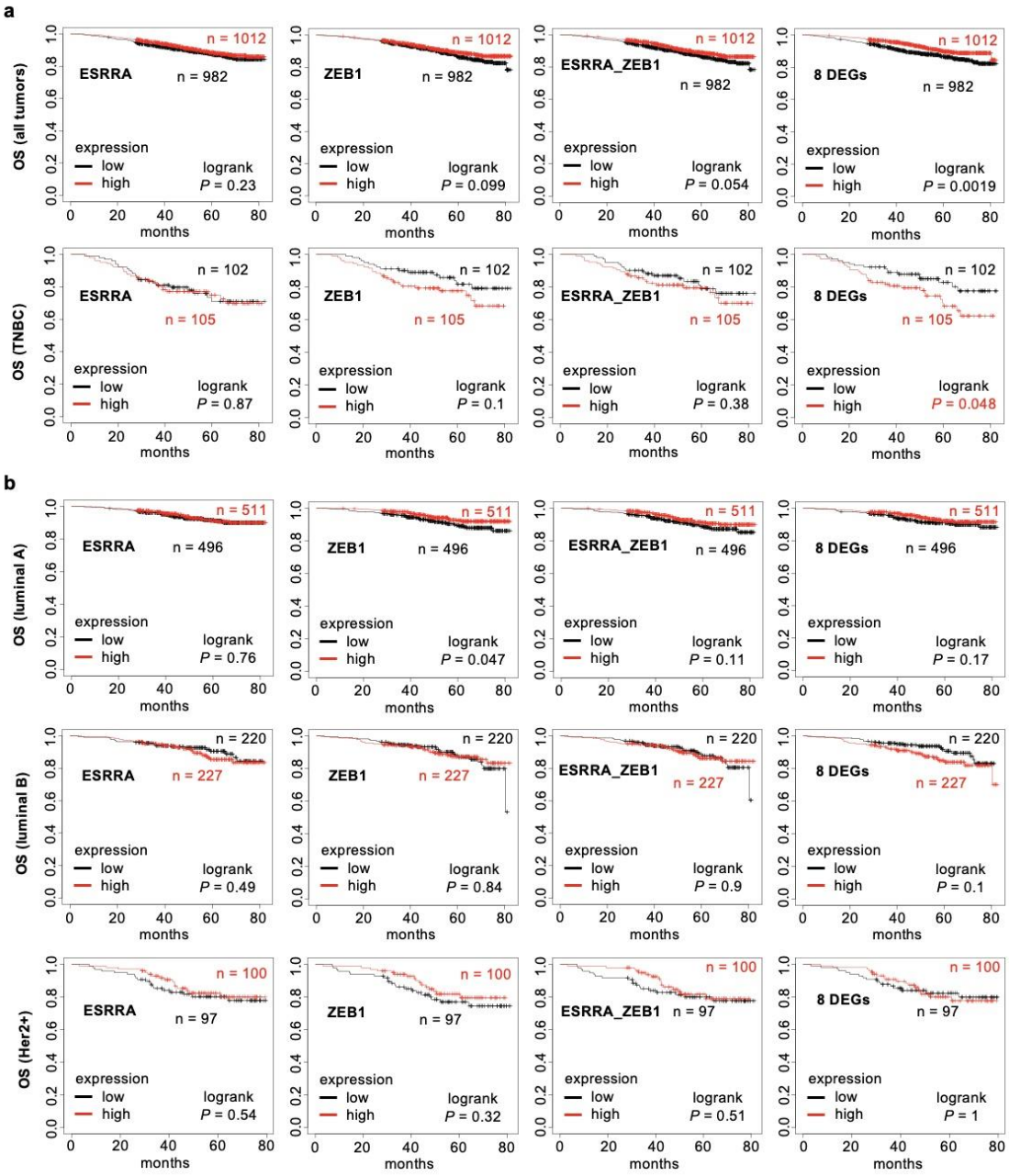


Figure 69. Kaplan-Meier curves from GSE96058 breast cancer RNA-Seq dataset showing the overall survival (OS) of tumor subtypes from KM plotter database, based on the expression of the indicated genes (single or in combination). a. Results in all breast tumors and TNBC (referred to as "basal" in the database). b. Results in other breast tumor subtypes. Number of tumors in each group and log-rank test p-values are indicated.

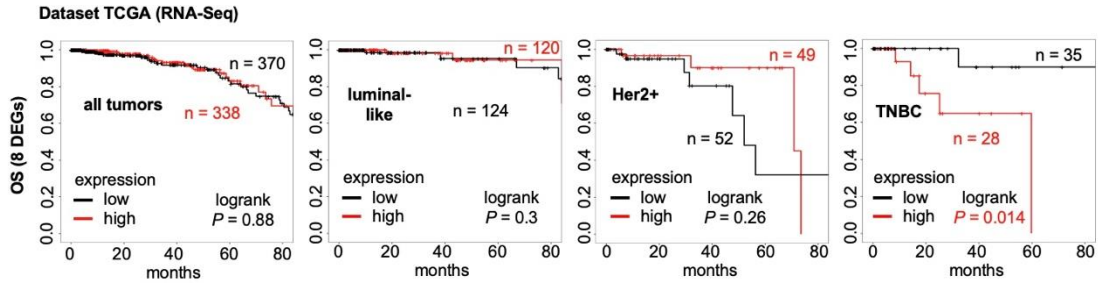


Figure 70. Kaplan-Meier curves from TCGA breast cancers RNA-Seq dataset showing the overall survival (OS) of tumor subtypes, based on the expression of the 8 DEGs. Number of tumors in each group and log-rank test p-values are indicated.

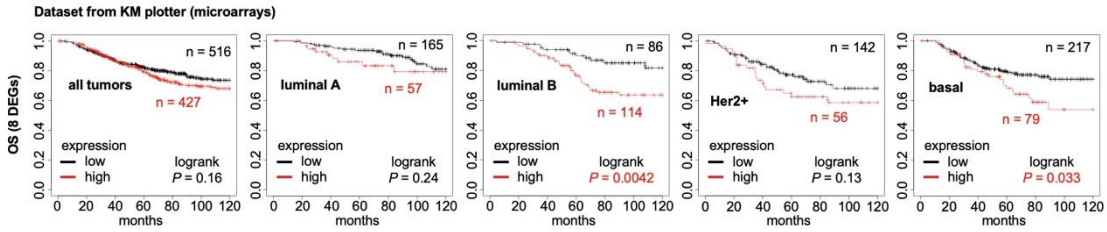


Figure 71. Kaplan-Meier curves from chip datasets showing the overall survival (OS) of tumor subtypes from KM plotter database, based on the expression of the 8 DEGs. Number of tumors in each group and log-rank test p-values are indicated.

9.2 Influence of ERRα, ZEB1 and the 8 DEGs on the relapse free survival and distant metastasis free survival

Moreover, we also explored the predicted ability of the 8 DEGs on relapse free survival (RFS) and distant metastasis free survival (DMFS) in the human breast cancer chip datasets from the KM plotter database. On the one hand, we did not detect a significant association between the combined expression of the 8 DEGs with RFS in all patients or each breast cancer subtype (**Figure 72a**). The same results were also observed in four TNBC molecular subtypes of FUSCC dataset, showing no significant association between the expression of the 8 DEGs with RFS in all Chinese TNBC population or each subtype (**Figure 72b**). However, in the MES subtype, we observed a trend of tumor recurrence in patients with highly combined expression of the 8 DEGs, in accordance with the higher EMT score in MES subtypes (**Figure 68**). On the other hand, the high combined expression of the 8 DEGs showed a significant correlation with worse DMFS in luminal B and Her2+ subtypes of chip datasets from KM plotter database. However, although it is not significant, we still observed that TNBC patients with highly combined expression of the 8 DEGs showed an increased risk of distant tumor metastasis in five years (**Figure 73**), which is consistent with the strong positive association of the 8 DEGs and EMT processes (**Figure 64, 66**).

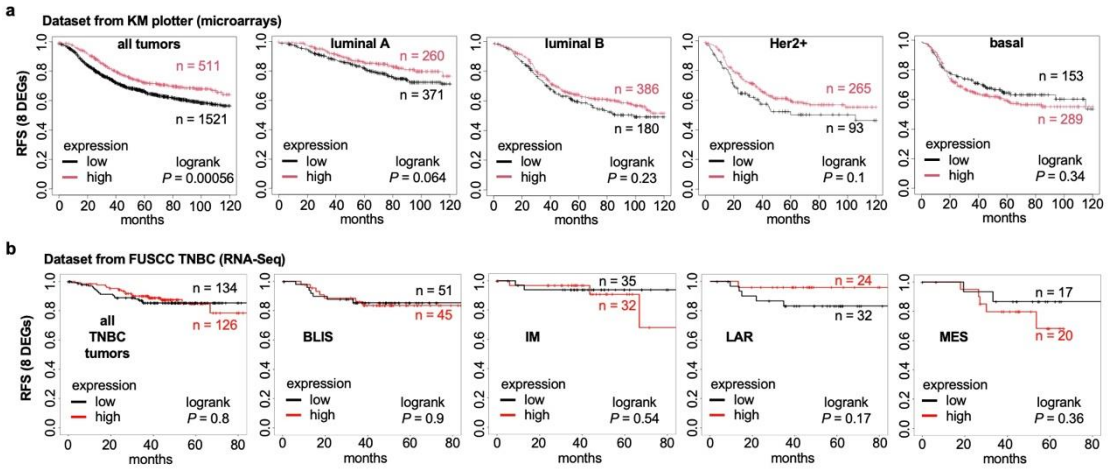


Figure 72. Kaplan-Meier curves showing the impact of the combined expression of the 8 DEGs on the relapse free survival (RFS) of tumor subtypes a. from chip datasets in KM plotter database, b. from FUSCC TNBC RNA-Seq dataset. Number of tumors in each group and log-rank test p-values are indicated.

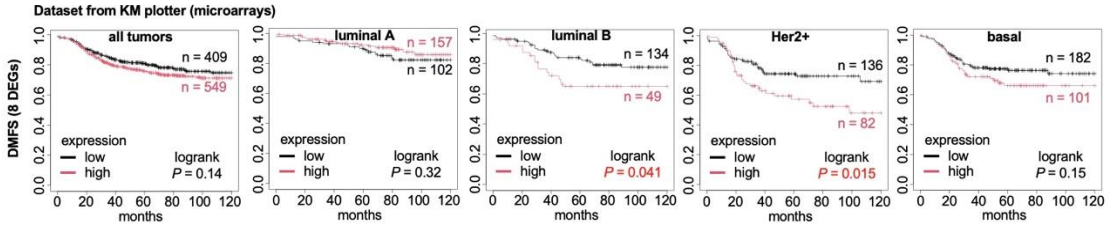


Figure 73. Kaplan-Meier curves from chip datasets showing the impact of the combined expression of the 8 DEGs on the distant metastasis free survival (DMFS) of tumor subtypes from KM plotter database. Number of tumors in each group and log-rank test p-values are indicated.

To summarize, the joint higher expression of the 8 DEGs was significantly and specifically associated with worse OS in TNBC patients while not in other breast cancer subtypes. The combined expression of the 8 DEGs showed no significant correlation with RFS and DMFS in TNBC patients, but we indeed observed a trend to relapse after treatment in TNBC MES subtype patients and a trend to distant tumor metastasis in five years in TNBC patients with jointly higher expression of 8 DEGs. We suggested that the 8 DEGs could be united to a new target set for predicting the prognosis of TNBC patients and a potential therapeutic strategy to prevent metastasis and recurrence after treatment.

Discussion and perspectives

ERRα, an orphan nuclear receptor, works with co-regulators to act on the expression of targets involved in tumorigenesis and cancer progression. To better understand the transcriptional activity of ERRα in breast cancer, the aim of my thesis is to predict co-regulators for ERRα using bioinformatic methods based on breast cancer expression data. For each gene that is directly activated by ERRα (differentially expressed genes; DEGs), we used a mathematical method called adaptive sparse partial least squares (sPLS). This allowed us to build regression models for each ERRα-DEG constructed by a list of human transcriptional regulators (TRs) that are relevant to breast cancer. The most frequent TR, contributing to the expression of these ERRα-DEGs, was suggested as a potential co-regulator of ERRα in breast cancer.

Relationship of ERRα and ZEB1 in TNBC

In our investigations, ZEB1 was always identified as the most potential co-factor of ERRα on the regulation of the expression of the 8 DEGs, questioning the molecular mechanisms of these regulations. Previous researches have independently established a participation of ERRα and ZEB1 in TNBC migration (Spaderna et al., 2008; Sailland et al., 2014; Wu et al., 2015; Carnesecchi et al., 2017). On another hand, data published by Ma et al. (2019) reported that ERRα activated by phosphorylated STAT3 could promote TNBC migration (Ma et al., 2019). In addition, they showed a decreased expression of ZEB1 upon ERRα depletion in MDA-MB231, a TNBC breast cancer cell line. Furthermore, a significant modification of EMT features (such as a reduction of vimentin and N-cadherin expression as well as an increase in E-cadherin) was observed upon ERRα knock-down. Altogether, this suggests that ERRα, activated by STAT3, could regulate ZEB1-mediated EMT progress. However, it should be noted that we did not observe a regulation of ZEB1 expression (at the mRNA or protein levels, **Figure 39-40**) upon ERRα knock down. This is not in favor of the hypothesis above although it remains possible that ERRα is involved in the EMT process in an STAT3-dependent manner.

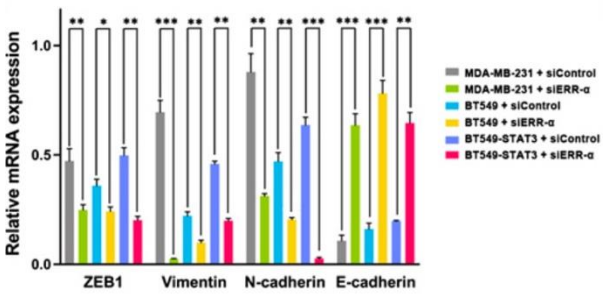


Figure 74. Bar plot from Ma et al. (2019) showing the expression change of EMT features in the treatment of siERRα in breast cancer cells. The depletion of ERRα decreases the mRNA expression of ZEB1, vimentin and N-cadherin while it increases

the mRNA expression of E-cadherin. Significance relative to control: ***: $p<0.001$, **: $p<0.01$, *: $p<0.05$.

Using public data to explore the potential of other TRs to be ERRα co-regulators in breast cancer

The significance of ZEB1 acting as a co-activator of ERRα in breast cancer has been experimentally explored in **section V** and **section VI** of "Results" above. The other identified top TRs (**Figure 27a**) still showed a relation, although unstable among subtypes, to ERRα targets. This suggests a potential to be a co-regulator of ERRα in a certain way. Therefore, we next checked if these other TRs have been verified or mentioned as co-regulators of ERRα in previous studies based on two open access databases. One of them (ChIP-Atlas) integrates publicly available experimental ChIP-Seq data and displays the colocalized proteins with ERRα in breast cancer on the web. Another database (TRRUST version2) provides the transcription regulators (TRs) sharing targets with ERRα in human and mouse, and all information in TRRUST is computationally extracted from published literature based on natural language learning methods.

We first explored ChIP-Atlas database which collects and analyzes a large amount of raw sequence ChIP-Seq data in the SRAs (Sequence Read Archives) database (Zou et al., 2024). These ChIP-Seq data come from multiple species, including *H. sapiens*, *M. musculus* or *D. melanogaster* and the results of the analyses are displayed on the web site (https://chip-atlas.org). We investigated the colocalization analysis results for ERRα in *H. sapiens* breast cancer. The online searched results displayed a number of TFs showing similar ChIP-seq profiles with ERRα on multiple genomic regions, suggesting a colocalization of these TFs with ERRα. (all searched results can be accessed from: https://chip-atlas.dbcls.jp/data/hg38/colo/ESRRA.Breast.html). In all the top 5 TRs predicted by our sPLS algorithm (**Figure 27a**), three TRs are also found in this database as proteins colocalized with ERRα in breast cancer (**Table 7**).

Cell types	ERRα's colocalization partners
ZR-75-1	EP300
MCF-7	ELF1
MCF-7	YBX1

Table 7. Colocalized proteins of ERRα based on public ChIP-seq data of breast cancer cells in ChIP-Atlas database. Three from our predicted TRs (**Figure 27a)** were identified in this database.

We then investigated the TRRUST v2 (transcriptional regulatory relationships unraveled by sentence-based text mining version 2) database which includes TF–target regulatory interactions in humans and mice as well as the interactions between TFs (Han et al., 2018b). All information in TRRUST were analyzed and obtained from published literature with sentence-based text mining method, followed by manual curation (https://www.grnpedia.org/trrust/). We searched other TFs sharing targets with ERRα in human on web (all searched results can be accessed from: https://www.grnpedia.org/trrust/result.php?gene=ESRRA&species=human&confirm=0). Results showed 14 TFs significantly sharing target genes with ERRα in human, in which two out of 14 TFs were also predicted as ERRα's potential co-regulators in our work (**Table 8, Figure 27a**).

TF	# of overlapped target genes	P value	FDR
CREB1	5	6.55E-09	5.78E-10
NR3C1	3	3.82E-06	8.96E-07
NR5A1	2	0.00011559	5.81E-05
MTA1	2	0.00015397	8.16E-05
LEF1	2	0.00027727	0.00016677
POU2F1	2	0.0002888	0.00017599
WT1	2	0.00056972	0.0003974
ESR1	2	0.00094225	0.00072446
HDAC1	2	0.00100453	0.00078173
ETS1	2	0.00129985	0.00105442
SP1	3	0.00211139	0.00182385
JUN	2	0.00320305	0.00289054
RELA	2	0.01157247	0.01130848
NFKB1	2	0.0116319	0.01137775

Table 8. 14 TFs share targets with ERRα in human in TRRUST v2 database. Two TRs (in yellow) from our predicted TRs by using sPLS method (**Figure 27a)** were identified in this database. P values are calculated with hypergeometric test.

However, we also noticed that ZEB1, the most robust and significant CoR in our

analysis, is not included in the results of these two databases (**Table 7, Table 8**). Data in ChIP-Atlas and TRRUST are integrated from the published studies or literature, so the first explanation may be an incomplete ChIP-Seq data collection for each TF, especially for ZEB1 in ChIP-Atlas database. Actually, we only found three sets of ZEB1 ChIP-Seq data (ChIP-Atlas ID: SRX13298923; SRX13298924; SRX13298925) in ChIP-Atlas (all searched results can be accessed from: https://chip-atlas.dbcls.jp/data/hg38/colo/ZEB1.Breast.html). These ZEB1 ChIP-Seq data are all from MCF-7, a non-TNBC cell line (ER+/PR+/Her2-). A ZEB1 ChIP-Seq study performed in MDA-MB231 (a TNBC cell line) can be publicly accessed since 2020, but it was unfortunately not included in ChIP-Atlas (Feldker et al., 2020). The TRRUST database provides information extracted from published literature based on text-mining methods. Thus, up to now, ZEB1 has not shown significant target sharing with ERRα in published literature. Altogether, this suggests that our data are the first to predict and validate a co-regulation between ERRα with ZEB1 in human breast cancer, especially TNBC.

Features of the TNBC MES subtype

To figure out the transcriptional activity of ERRα and its co-activator ZEB1, we first identified 8 ERRα-directly activated DEGs which are also correlated to ZEB1 as evidenced by modeling results (**Figure 27b**). Then we validated the variability and stability of ZEB1 and the 8 DEGs among different breast cancer subtypes, age groups, ethnics and datasets of various sample size, showing that ZEB1 was the most stable and significant factor for the transcription of the 8 ERRα-DEGs (**Figure 27b, Figure 28, Table 5, Figure 30-32**). We also studied ERRα, ZEB1 and these 8 common targets in expression profiles of breast cancer single cell, cell lines and tumors. Comparing to other breast cancer subtypes, we found that the expression of ZEB1 and the 8 DEGs (except MCU) was enriched in the most aggressive TNBC subtype, in particular in the TNBC MES subtype (mesenchymal-like) (**Figure 34-38**). Thus, we explored the specificity of the MES subtype to better understand the association between ZEB1 and the 8 DEGs with TNBC. From the conclusion of Jiang et al. (2019), the main features of MES subtype are an overexpression of genes related to breast cancer stem cells (CSC) and a higher expression of JAK1 and IL6 leading to up-regulation of JAK/STAT3 signaling pathway (Jiang et al., 2019). We thus performed literature reviewing on these two features to explore the association between each feature with breast cancer development.

CSC are cells with a low differentiation degree, which are characterized by self-renewal, proliferation and multidirectional differentiation capacity (Zeng et al., 2021). In addition, CSC tend to escape from the primary site of the tumor and invade lymphatic vessels or blood vessels (Guo, 2014; Mendoza-Almanza et al., 2020). Therefore, CSC play an important role in the progression, metastasis and recurrence of cancers, and are involved in the prognosis of cancer patients (Zeng et al., 2021; Bu et al., 2023).

Interleukin 6 (IL6) is associated with treatment resistance in breast cancer patients.

Activation of the IL-6/JAK/STAT3 signaling pathway not only promotes the proliferation and invasion of cancer cells, it can also inhibit their apoptosis (Tsoi et al., 2021; Manore et al., 2022). In addition, STAT3 can drive the transmission of IL-6 signal, thus promoting the cycle of malignant inflammation in tumors. Moreover, the JAK/STAT3 signaling pathway is also closely associated with the EMT process (Kim et al., 2015).

Based on these facts, it appears that MES subtypes are closely related to tumor migration, invasion, metastasis and EMT processes, suggesting a potential association between high expression of ZEB1 and the 8 DEGs with cancer development in MES subtype, along with abnormal up-regulation of genes related to CSC and JAK/STAT3 signaling pathway.

ZEB1 is a key factor of the EMT process (Feldker et al., 2020; Wu et al., 2020a). We observed that breast tumors with higher expression of ZEB1 and the 8 DEGs are more likely to present a mesenchymal status, again pointing to a strong correlation between ZEB1 and the 8 DEGs with the EMT process and cancer progression (**Figure 64, Figure 66**).

As mentioned above, we suggest that ERRα modulates the EMT process depending on STAT3. This should be taken together with (1) the fact that STAT3 promotes cancer cell migration by acting as an upstream mediator of EMT (Sadrkhanloo et al., 2022), (2) the up-regulation of JAK/STAT3 pathways as one of MES subtypes features, and (3) the high involvement of ERRα's co-activator (ZEB1) and ERRα's transcription targets (the 8 DEGs) in EMT process. Altogether, this suggests again an association of ERRα with the EMT process and breast cancer progression through the mediation of STAT3, especially in MES subtype, even though no obviously high ERRα expression was observed in the MES subtype.

In summary, we hypothesize that ERRα, ZEB1 and the 8 DEGs are all involved in EMT and tumor metastasis in MES subtype. It is also possible that the association of ERRα with the EMT process could involve upstream STAT3, its co-activator ZEB1 and the downstream transcription of the 8 DEGs, because of the specific high expression of ZEB1 and the 8 DEGs as well as up-regulation of STAT3 signaling pathway in the MES subtype.

Characteristic of the 8 DEGs in breast cancer

After exploring the expression pattern of ERRα, ZEB1 and the 8 DEGs in breast cancer expression datasets, further biological experiments were also performed to verify the specific co-regulatory activities of ERRα and ZEB1 on the 8 DEGs in TNBC subtypes (**Figure 39-54**). We have reviewed the function of ERRα and ZEB1 in introduction **Chapter III** and **Chapter IV**. Below, we summarize the functions of the 8 DEGs in cancer, based on the literature.

ADAMTS12 is a member of ADAMTS (a disintegrin and metalloproteinase with

thrombospondin motifs) family which are extracellular multifunctional enzymes (Mohamedi et al., 2021; He et al., 2023). ADAMTS12 participates to cell migration, to the EMT process and is a prognostic biomarker in pancreatic cancer (He et al., 2023). In our survival analysis of TCGA TNBC patients, we observed a significant association between high expression of ADAMTS12 with poor overall survival (**Figure 75**). Moreover, ADAMTS12 could also be used as a predicting marker along with eight other markers (COL10A1, COL11A1, EDNRA, CXCR6, WNT7B, CXCL11, AEBP1, EPPK1) to classify TNBC patients to CAF+ (cancer-associated fibroblasts) subtype, associated to worse outcome, and CAF- subtype (Wang et al., 2022). This is in line with our observation showing a high expression of ADAMTS12 in CAFs cells in TNBC tumor microenvironment (**Figure 56b, 57**). These two observations suggested the role of ADAMTS12 in breast cancer, particularly correlated to CAFs. Altogether, we suggest that ADAMTS12 is involved in tumor-promoting mechanisms, in particular in correlation to the microenvironment, as well as associated with survival of TNBC patients individually or cooperatively.

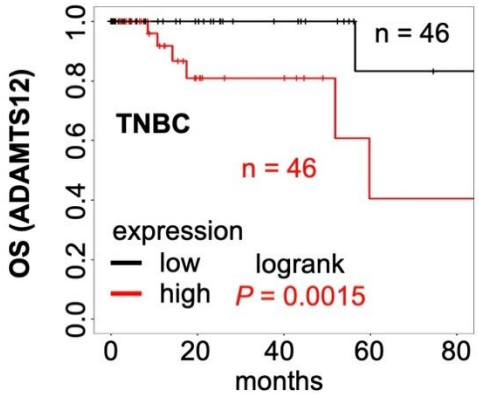


Figure 75. Kaplan-Meier curves from TCGA TNBC breast cancers RNA-Seq dataset showing the overall survival (OS), based on the expression of ADAMTS12. Number of tumors in each group and log-rank test p-values are indicated.

CDH5, or VE-cadherin (vascular endothelial cadherin), an endothelial junction protein (Usman et al., 2022), is a marker for metastatic breast cancers. As expected, we observed a specific expression enrichment of CDH5 in endothelial cells in the tumor microenvironment of all breast cancer subtypes (**Figure 56-61**). This is consistent with a nested case–control study showing an elevated expression of CDH5 in breast cancer metastatic patients (Fry et al., 2016). Milovanović et al. (2023) also found that there was a significant difference in mRNA expression of CDH5 between breast cancer patients with distant metastasis and patients without recurrences. Patients with the higher expression of CDH5 had a higher risk of worse outcome as well as a large incidence of distant recurrence raising to 26% compared with the 3% displayed by the group exhibiting the lower CDH5 expression (Milovanović et al., 2023).

Discoidin domain receptor tyrosine kinase 2 (DDR2) plays a critical role in both tumor cells and stromal cells in breast cancer metastasis (Zhang et al., 2013; Toy et al., 2015; Corsa et al., 2016). On the one hand, DDR2 facilitates the invasion and migration of tumor cells and stimulates EMT of breast cancer by regulating SNAIL1 stability (Zhang et al., 2013). On the other hand, Corsa et al. (2016) observed that DDR2 was important for the production of tumor ECM and the organization of collagen fibers. DDR2 in CAFs also enhances tumor cell collective invasion or migration (Corsa et al., 2016). As expected, we detected a significantly higher expression of DDR2 in the CAFs of human breast tumors (**Figure 56-61**). Interestingly, Klingen et al. (2024) found that the high expression of DDR2 was associated with significantly reduced recurrence-free survival (RFS), high tumor cell proliferation as well as more TNBC cases (Audun Klingen et al., 2024). Altogether, we concluded that DDR2 is highly involved in breast cancer metastasis through both tumor cells and stromal cells (Bayer et al., 2019).

E26 transcription factor-1 (ETS1) is a member of ETS domain family that highly participates in angiogenesis, EMT and cancer progression (Dittmer, 2015). ETS1 promotes angiogenesis in cancer cells through the overexpression of vascular endothelial growth factor (VEGF) receptor, Neuropilin-1 (Nrp1), and angiopoietin-2 (Ang2) (Kim et al., 2018). We observed a significant expression of ETS1 in endothelial cells (**Figure 56-61**). Additionally, ETS1 is also a prognostic predictor for survival of breast cancer patients (Span et al., 2002). Specifically in TNBC, the highly expression of ETS1 was associated with worse overall survival (Li et al., 2022b). Moreover, the overexpression of ETS1 gene was observed in breast cancer cells resistant to multidrug and may contribute to the development of the resistance of the cells (Kars et al., 2010).

Mitochondrial calcium uniporter (MCU) is a selective Ca^{+} channel transmembrane protein of the mitochondria. Previous studies demonstrated a role of MCU in cell motility, cell invasiveness, tumor metastasis in breast cancer, especially in TNBC (Tosatto et al., 2016). Overexpression of MCU was observed in breast cancer tissue compared to the paired adjacent normal tissues from four breast cancer patients. Increased protein expression of MCU led to enhanced migration capacities of MDA-MB231 breast cancer cells (Yuan et al., 2023). Moreover, MCU was used as a prognostic marker to evaluate the overall survival of breast cancer patients (Yuan et al., 2023; Li et al., 2024). In our study, we also identified MCU as a prognostic predictor in breast cancer. A high expression of MCU was associated with worse overall survival in breast cancer as well as in TNBC subtype (**Figure 76**).

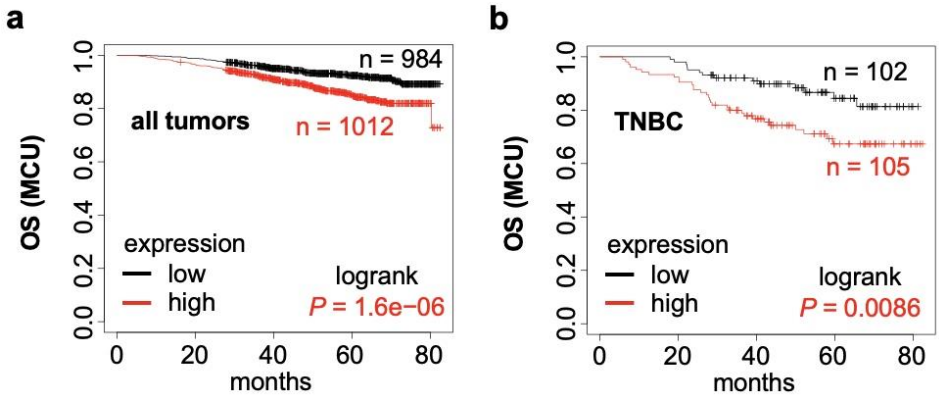


Figure 76. Kaplan-Meier curves from GSE96058 breast cancer RNA-Seq dataset showing the overall survival (OS) from KM plotter database, based on the expression of MCU. a. Result in all breast tumors. b. Result in TNBC (referred to as "basal" in the database). Number of tumors in each group and log-rank test p-values are indicated.

Platelet and endothelial adhesion molecules type 1 (PECAM1) is expressed in endothelial cells, monocytes, cancer-associated fibroblasts, immune and inflammatory cells and platelets (Abraham et al., 2018). We also observed a higher expression of PECAM1 in endothelial cells (**Figure 56-61**), supporting its role in transendothelial migration of cell and tumor metastasis (Pospelova et al., 2022). Gradually elevated expression of PECAM1 was detected with increasing grade of breast tumors. Furthermore, the expression of PECAM1 was significantly increased in patients who have died of breast cancer and in patients who have a poor outcome (Martin et al., 2005). In addition, Pospelova et al. (2022) found PECAM1 serum levels was significantly higher in female patients after breast cancer treatment compared to healthy individuals, as well as an association with cognitive dysfunction, depression, and vestibulocerebellar ataxia. This suggested PECAM1 may also act as a predictor of the damage evaluation in the central nervous system in post-treatment breast cancer patients. Altogether, PECAM1 is associated with a worse status of breast tumor and particularly associated with complications in breast cancer survivors.

Plexin-c1 (PLXNC1) is a receptor for semaphoring 7A (SEMA7A) which regulates cell motility, cell migration and immune reactions, as well as participate in cancer process. PLXNC1 was studied mostly in melanoma and gastric cancer (Balakrishnan et al., 2009; Wang et al., 2018; Chen et al., 2020; Ni et al., 2021), and less in breast cancer. Semaphorins, a super family consisting of 21 proteins, function as ligands for plexin receptors or function as receptors for plexin ligands. Thus, the diverse molecular regulation mechanism helps them activate and stimulate multiple signaling pathways, for example, tumor angiogenesis, metastasis and progression of breast cancer. Semaphorin 7A (SEMA7A), the ligand of PLXNC1, has also an impact on tumor progression and drug resistance in breast cancer. The expression of SEMA7A was associated with worse outcome and early metastasis. Especially, the depletion of SEMA7A led to an increased expression of E-cadherin and a decreased expression of

mesenchymal features such as p63 and vimentin, suggesting a strong correlation between SEMA7A and EMT process (Aiyappa-Maudsley et al., 2023).

Secreted protein acidic and rich in cysteine (SPARC), an albumin-binding glycoprotein, plays an important role in the formation and development of tumors. In our work, a higher expression of SPARC was found in stromal cells, especially CAFs, endothelial cells and pericytes, than that in cancer cells (**Figure 56-61**) (Shi et al., 2022). Alcaraz et al. (2023) found TNBC patients with SPARC-expressing CAFs had an association with worse RFS, indicating an important function of SPARC in stromal cells of breast cancer. Besides, additional increased expression of N-cadherin, vimentin, and β-catenin which are EMT-related markers was observed in SPARC overexpressing breast cancer cells while E-cadherin displayed a decreased expression (Li et al., 2022a). Altogether, this demonstrates a strong correlation of SPARC with the EMT process in breast cancer. The high expression of SPARC was associated with a significant poor outcome in TNBC patients which further suggested that SPARC may participate in the EMT process and thus influence the prognosis of breast cancer patients (**Figure 77**).

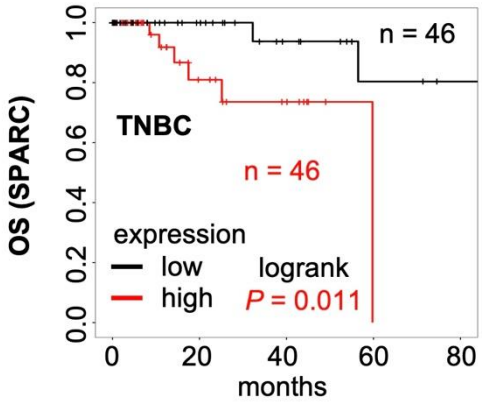


Figure 77. Kaplan-Meier curves from TCGA TNBC breast cancers RNA-Seq dataset showing the overall survival (OS), based on the expression of SPARC. Number of tumors in each group and log-rank test p-values are indicated.

In summary, the 8 DEGs are highly involved in tumor metastasis and associated to poor survival of breast cancer patients in published literatures. In our analysis, we observed that three of them (ADAMTS12, MCU, SPARC) have a significant association with overall survival in breast cancer (**Figure 75-77**), whereas the other five DEGs have no significant results. These results may depend on the expression datasets analyzed and need to be further validated.

Interestingly, we also noticed that ADAMTS12, DDR2 and SPARC have a particular function in stromal cells, such as CAFs (Corsa et al., 2016; Wang et al., 2022; Alcaraz et al., 2023), suggesting a potential role of these DEGs in the tumor microenvironment composed of non-cancer cells.

Characteristic of CAFs, endothelial cells and pericytes in breast cancer

Tumors not only comprise cancer cells and ECM, but also diverse cells such as CAFs, endothelial cells and pericytes (Turley et al., 2015; Zhao et al., 2023). This cellular microenvironment influences cancer cells and contributes to the development of tumors. We have investigated and validated the specific expression of ERRα, ZEB1 and the 8 DEGs in TNBC cells above (**Figure 34-35, Figure 40-42, Table 6**). Moreover, we observed a potential association between ADAMTS12, DDR2 and SPARC with stromal cells above (**Figure 75-77**). We thus next explored the expression of ERRα, ZEB1 and the 8 DEGs in the tumor cellular microenvironment. Interestingly, we observed that the expression distribution of these genes is similar among TNBC, ER+ and Her2+ tumor microenvironment. Results also showed a specific enrichment of ZEB1 and the 8 DEGs in at least one of three non-cancer cells (CAFs, endothelial cells and pericytes), while ERRα is expressed everywhere in all cells (**Figure 56-61**). Therefore, we thus investigated the biological function of CAFs, endothelial cells and pericytes in breast cancer by literature reviewing in order to explore the association between ZEB1 and the 8 DEGs with breast cancer microenvironment.

As a major component of the tumor microenvironment, CAFs can affect treatment resistance, recurrence and metastasis of breast cancer tumors (Hu et al., 2022). The promotion of metastasis is exerted by the modulation of the specific expression of molecular markers associated with EMT process. Indeed, the results by Soon et al. (2013) showed that MCF7 breast cancer cells expressed higher levels of EMT hallmark traits, such as CXCL12 and α-smooth actin (α-SMA) when co-cultured with CAFs as compared to co-culture with normal breast fibroblasts. There was also a higher expression of vimentin and lower expression of cell-cell contact protein E-cadherin in MCF7 cells co-cultured with CAFs. This suggests that CAFs can induce the progression of EMT in MCF7 cells and may further promote cancer cell metastasis and invasion (Soon et al., 2013). In addition, CAFs can promote EMT and influence breast cancer progression also by secreting proteins. For example, CAFs can regulate the expression of TGF-β and then promote cancer cell metastasis (Yu et al., 2014; Takai et al., 2016; Pelon et al., 2020; Hu et al., 2022); Furthermore, the deletion of Tiam1 in CAFs increases the expression of osteopontin and drives an increase of EMT features in co-cultured breast cancer cells, thus affecting the metastasis of these cells (Xu et al., 2016). CAFs from breast cancer tissues can activate the Wnt/β-catenin signaling pathway by secreting a large amount of collagen triple helix repeat containing-1 (CTHRC1). Abnormal activation of this pathway can reduce the expression of E-cadherin, thereby promoting EMT process and the invasion of cancer cells (Li et al., 2021). Altogether, these data suggest a strong association of CAFs with the EMT process, cancer cell metastasis and invasion.

Endothelial cells generally exist on the inner surface of the cardiovascular system. They regulate the formation of blood vessels through influencing angiogenesis. The binding of vascular endothelial growth factor (VEGF), a main factor of angiogenesis,

to VEGFR2, a receptor on endothelial cells (Hicklin and Ellis, 2005), activates downstream signaling pathways and promotes angiogenesis. It also provides oxygen and nutrients to tumor cells and promotes the EMT process in breast cancer cells as well as spread to other parts of the body (Ghiabi et al., 2014; Gao et al., 2023). In addition, vascular cell adhesion molecule-1 (VCAM-1) on the endothelial cells induces the contact and adhesion of tumor cells to endothelial cells, regulating the metastasis process of breast cancer tumors (Koning et al., 2023). Ferraro et al. (2019) also found that endothelial cells can promote the migration and invasion of SK-BR-3 breast cancer cells by activating STAT3 signaling pathway, which plays an important role in the EMT process. Globally, these data suggest an involvement of endothelial cells in the regulation of tumor progression.

Molnár et al. (2020) found that brain pericytes can attract and directly interact with TNBC cells. Pericyte-secreted soluble factors can strengthen the adhesion to each other and drive angiogenesis near the tumor. Besides, pericyte-secreted factors can significantly reduce the expression of E-cadherin in breast cancer cells, enabling these cells to acquire EMT features. In addition, high expression of insulin-like growth factors (IGFs) in pericytes contributes to the proliferation of breast cancer cells (Molnár et al., 2020). Mayo et al. (2021) also found that pericyte-like cells cultured on CD146+ cells can promote the acquisition of invasion and metastasis of 3384T TNBC cells. Altogether, these data show important functions of pericytic cells in breast cancer invasion.

In summary, previous researches have documented the significant role of CAFs, endothelial and pericytes in tumor invasion, migration and metastasis of breast cancer, suggesting that the enrichment of ZEB1 and the 8 DEGs in these cells may also participate in breast cancer progression in a microenvironment-dependent manner.

It is worth noticing that we almost did not observe a specific enrichment of these factors in immune cells such as T cells and B cells (**Figure 57, Figure 59, Figure 61**). This may be because there is only a small fraction of cells that are highly expressing these factors among the large number of immune cells. Thus, we did not focus on immune cells here.

Prognostic function of ERRα and ZEB1 in breast cancer

There is a strong association of our factors with different breast cancer processes such as EMT, invasion, metastasis and even clinical outcome. Furthermore, this association is obvious in cancer cells but also in the tumor microenvironment. This suggests diverse functions of these genes in breast cancer progression. Moreover, we observed that tumors with a high expression of ZEB1 and the 8 DEGs are more likely to display a mesenchymal phenotype which may develop cancer metastasis (**Figure 64, Figure 66**). Therefore, we next checked the influence of these genes on the clinical outcomes of breast cancer patients.

The role of the 8 DEGs in patient survival was discussed in “Characteristic of the

8 DEGs in breast cancer" discussion part above. ERRα and ZEB1 also have individually been demonstrated as prognosis relevant factors in breast cancer (Chang et al., 2011; May, 2014; Xiang et al., 2015; Wu et al., 2020a; Zawati et al., 2022). In contrast, other studies have shown that the expression of ERRα mRNA by itself cannot evaluate the outcomes of breast cancer patients (Deblois et al., 2009). However, a set of 86 ERRα target genes can be used as a prognostic predictor for breast cancer. In addition, another study surveyed the predictive ability of the mRNA expression of ERRα on patients' survival in six clinical breast tumor microarray datasets (Chang et al., 2011). Surprisingly, the significant association between higher ERRα expression with worse outcome was only found in one dataset, while a cluster of ERRα-regulated genes was identified, correlated to unfavorable clinical outcomes in patients with breast cancer. This suggested a changeable predictive ability of ERRα itself in prognosis, and that a clinical impact of ERRα on survival may depend on its transcriptional targets.

Similarly, in our research, we identified 8 ERRα directly activated targets as a combined predictor of overall survival (OS) in TNBC, but not in other breast cancer subtypes (**Figure 69-71**). Noticeably, the expressions of ERRα or ZEB1 have no impact on the OS of breast cancer patients studied in GSE96058 RNA-Seq dataset (**Figure 69**), TCGA RNA-Seq dataset and chip datasets (**Figure 78**). However, a significant prognostic effect of ERRα expression was detected in all tumors and Her2+ subtype of chip datasets (**Figure 78b top**), but this needs to be explored and validated further as microarray technology may harbor limitations in detection and troubles with normalization (Wang et al., 2009). Altogether, we conclude that the individual expression of ERRα or ZEB1 cannot reliably act as prognosis factors, whereas the combined expression of the 8 DEGs (reflecting ERRα-ZEB1 activity) can predict OS in TNBC patients.

We next explored the potential reasons of the non-significant influence of ERRα or ZEB1 expression on the OS of breast cancer patients.

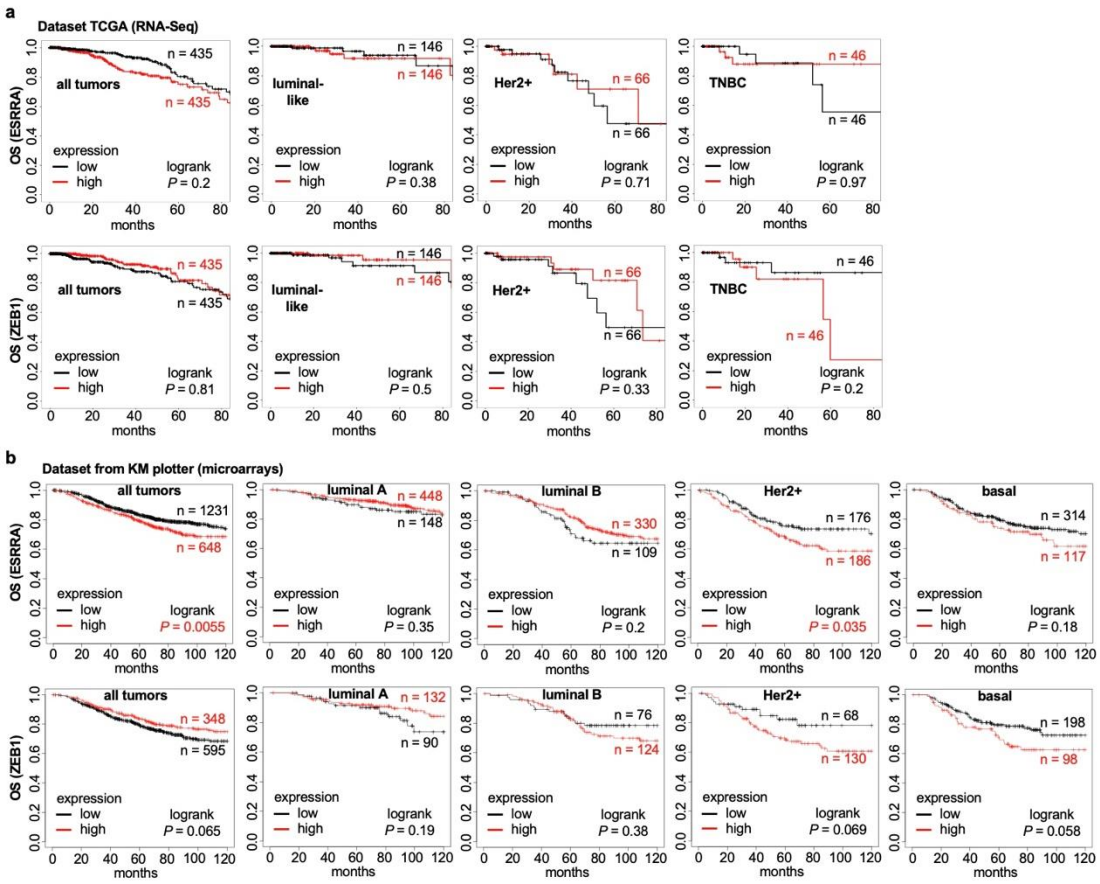


Figure 78. Kaplan-Meier curves showing the overall survival (OS) of breast tumor subtypes based on the expression of ESRRA and ZEB1 respectively. a. Results from TCGA breast cancer RNA-Seq dataset. Top: ESRRA. Below: ZEB1. b. Results from breast cancer chip datasets in KM plotter database. Top: ESRRA (the probe used in here: 1487_at). Below: ZEB1 (the probe used in here: 239952_at). Number of tumors in each group and log-rank test p-values are indicated.

We speculated that there may be an influence of distinct expression molecules (protein or RNA) for patient groups classification. For example, Ye et al. (2020) surveyed the association of ERRα expression with prognosis in a cohort of 199 cases of TNBC patients, including the OS, local recurrence-free survival (LRFS) and distant disease-free survival (DDFS). A significant decrease of these three parameters was observed in the ERRα-high TNBC patients compared to the ERRα-low group, indicating a strong association between ERRα with breast cancer patients' outcome. However, ERRα expression used in survival analysis was calculated based on two features (dyeing intensity and positive-staining scope) of tissue slides from the immunohistochemistry staining experiments on FFPE (formalin fixation and paraffin embedding) tumor tissues. Thus, they finally classified patients to ERRα-low and ERRα-high groups as well as performed survival analysis based on the protein expression of ERRα. However, this is in contrast to our survival analysis which was conducted on the RNA expression of ERRα or ZEB1 from RNA-Seq or chip datasets.

Therefore, we suggest that expression difference of RNA or protein level may influence the tumor groups classification and thus affect the correlation with patients' outcome.

In addition, the distribution of gene expression among tumor cell types may also have an impact on the results. Gene expression from RNA-Seq was the united expression of bulk RNA extracted from large cell populations of the tumor tissue, so higher or lower expression in some specific cell types may lead to the bias of transcriptomic analysis. An example is shown in **Figure 56-61**, where ZEB1 is more expressed in stromal cells than in tumor cells. Besides, Cortés et al. (2017) found that only the tumor-associated macrophages (TAMs) fully expressing Zeb1 were able to accelerate tumor growth in the mouse model of ovarian cancer. Moreover, they also identified ZEB1 in TAMs as a prognostic factor for worse outcome of human patients with ovarian carcinomas. These data suggest that ZEB1 affects patient's survival through a specific expression in immune cells such as TAMs. Therefore, our survival analysis based on the expression in the whole tumor, not a particular cell type in which the gene specifically enriched, may also lead to an unsignificant prognostic result.

In summary, we identified a combination of 8 DEGs as a prognostic predictor of overall survival in TNBC patients, but not in other breast cancer types. ERRα or ZEB1 mRNA expression individually have no impact on the overall survival at least as observed in GSE96058 and TCGA RNA-Seq datasets, although this needs to be investigated in more independent breast cancer datasets.

What's more?

We have identified ZEB1 as a co-activator of ERRα in TNBC and these factors specifically co-regulate the expression of 8 common targets involved in cell migration. We have discussed their potential role in cancer metastasis through cell migration and the EMT pathways. ZEB1 is also known as a factor participating to cancer metabolism by enhancing the Warburg effect (an adaptive change in metabolism in cancer cells that will use carbohydrates as an energy source) and then promotes cancer metastasis (Koppenol et al., 2011; Martinez-Outschoorn et al., 2014; Zhou et al., 2021). Since ERRα is also involved in the promotion of the Warburg effect (Deblois and Giguère, 2013), we hypothesize an underlying mechanism of the co-regulation of ERRα and ZEB1 on tumorigenesis through metabolic pathways. However, our RNA-Seq approaches showed that metabolic targets of ERRα are not regulated by this receptor in TNBC cells (Sailland et al., 2014; Cerutti et al., 2022). This suggests that the ERRα-ZEB1-metabolism connection is unlikely to happen in TNBC cells.

Mortality of African American women with TNBC is larger than that observed in other populations in the United States, even when considering the influence of treatment delays and social factors (Morris et al., 2007). Here we also observed the contribution difference of ZEB1 to the 8 DEGs between ethnics in TNBC cohort (**Figure 28a**) although this needs other independent cohorts to validate the ethnic heterogeneity of ZEB1 and the 8 DEGs in TNBC. We propose a similar suggestion regarding the

comparison of pre-menopause and post menopause age groups as TNBC mostly occurs in pre-menopause young women (**Figure 28b**) (Yin et al., 2020).

In addition, epigenetic phenomena such as DNA methylation and histone modifications also alter gene expression and then influence cancer development. The effect of DNA methylation or histone modification on ERRα and/or ZEB1 functions should be studied with respect to the activities of these factors on the expression of the 8 DEGs. Further investigations on these aspects will be needed to complete the understanding of the transcription activity of ERRα in breast cancer progression.

General conclusion

To summarize, my work identified and validated ZEB1 as an ERRα co-activator during breast cancer progression. For TNBC, we also suggest a novel therapeutic set of genes (8 DEGs), which are common targets activated by ERRα and ZEB1. The high combined expression of the 8 DEGs is specifically associated to the EMT status of breast tumors not only in cancer cells but also in the tumor microenvironment. Altogether, this results in a worse outcome for TNBC patients.

Our investigation enhances the comprehensive understanding of the transcription activity of ERRα in breast cancer. TNBC are the most aggressive subtype and have a high incidence in young women. However, TNBC lack effective treatments partly because of the deficient expression of ER, PR and Her2 in patients. Therefore, there is an urgent need to discover new molecules to target in TNBC patients. The new transcriptional mechanism of ERRα as well as the ZEB1 and the 8 DEGs may provide a novel approach to improve the clinical outcome of breast cancer patients, particularly in the case of TNBC.

References

Abraham V, Cao G, Parambath A, Lawal F, Handumrongkul C, Debs R, DeLisser HM (2018) Involvement of TIMP-1 in PECAM-1-mediated tumor dissemination. Int J Oncol 53(2): 488–502.

Achermann JC, Schwabe J, Fairall L, Chatterjee K (2017) Genetic disorders of nuclear receptors. J Clin Invest 127(4): 1181–1192.

Aiyappa-Maudsley R, McLoughlin LFV, Hughes TA (2023) Semaphorins and Their Roles in Breast Cancer: Implications for Therapy Resistance. Int J Mol Sci 24(17): 13093.

Alcaraz LB, Mallavialle A, Mollevi C, Boissière-Michot F, Mansouri H, Simony-Lafontaine J, Laurent-Matha V, Chardès T, Jacot W, Turtoi A, Roger P, Guiu S, et al. (2023) SPARC in cancer-associated fibroblasts is an independent poor prognostic factor in non-metastatic triple-negative breast cancer and exhibits pro-tumor activity. Int J Cancer 152(6): 1243–1258.

Anderson NM, Simon MC (2020) The tumor microenvironment. Curr Biol 30(16): R921–R925.

Angueira AR, Shapira SN, Ishibashi J, Sampat S, Sostre-Colón J, Emmett MJ, Titchenell PM, Lazar MA, Lim H-W, Seale P (2020) Early B Cell Factor Activity Controls Developmental and Adaptive Thermogenic Gene Programming in Adipocytes. Cell Rep 30(9): 2869-2878.e4.

Ariazi EA, Clark GM, Mertz JE (2002) Estrogen-related receptor alpha and estrogen-related receptor gamma associate with unfavorable and favorable biomarkers, respectively, in human breast cancer. Cancer Res 62(22): 6510–6518.

Ariazi EA, Jordan VC (2006) Estrogen-related receptors as emerging targets in cancer and metabolic disorders. Curr Top Med Chem 6(3): 203–215.

Audun Klingen T, Chen Y, Aas H, Akslen LA (2024) DDR2 expression in breast cancer is associated with blood vessel invasion, basal-like tumors, tumor associated macrophages, regulatory T cells, detection mode and prognosis. Hum Pathol 150: 29–35.

Badia-I-Mompel P, Wessels L, Müller-Dott S, Trimbour R, Ramirez Flores RO, Argelaguet R, Saez-Rodriguez J (2023) Gene regulatory network inference in the era of single-cell multi-omics. Nat Rev Genet 24(11): 739–754.

Balakrishnan A, Penachioni JY, Lamba S, Bleeker FE, Zanon C, Rodolfo M, Vallacchi V, Scarpa A, Felicioni L, Buck M, Marchetti A, Comoglio PM, et al. (2009) Molecular profiling of the “plexinome” in melanoma and pancreatic cancer. Hum Mutat 30(8):

1167–1174.

Bayer SV, Grither WR, Brenot A, Hwang PY, Barcus CE, Ernst M, Pence P, Walter C, Pathak A, Longmore GD (2019) DDR2 controls breast tumor stiffness and metastasis by regulating integrin mediated mechanotransduction in CAFs. Elife 8: e45508.

Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, Bulyk ML (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat Biotechnol 24(11): 1429–1435.

Bianco S, Sailland J, Vanacker J-M (2012) ERRs and cancers: effects on metabolism and on proliferation and migration capacities. J Steroid Biochem Mol Biol 130(3–5): 180–185.

Bonnelye E, Kung V, Laplace C, Galson DL, Aubin JE (2002) Estrogen receptor-related receptor alpha impinges on the estrogen axis in bone: potential function in osteoporosis. Endocrinology 143(9): 3658–3670.

Boulesteix A-L, Strimmer K (2007) Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Brief Bioinform 8(1): 32–44.

Bu J, Zhang Y, Wu S, Li H, Sun L, Liu Y, Zhu X, Qiao X, Ma Q, Liu C, Niu N, Xue J, et al. (2023) KK-LC-1 as a therapeutic target to eliminate ALDH+ stem cells in triple negative breast cancer. Nat Commun 14(1): 2602.

Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S, Brabletz T (2008) A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep 9(6): 582–589.

Cai H, Wang Y, McCarthy D, Wen H, Borchelt DR, Price DL, Wong PC (2001) BACE1 is the major beta-secretase for generation of Abeta peptides by neurons. Nat Neurosci 4(3): 233–234.

Caramel J, Ligier M, Puisieux A (2018) Pleiotropic Roles for ZEB1 in Cancer. Cancer Res 78(1): 30–35.

Carnesecchi J, Forcet C, Zhang L, Tribollet V, Barenton B, Boudra R, Cerutti C, Billas IML, Sérandour AA, Carroll JS, Beaudoin C, Vanacker J-M (2017) ERRα induces H3K9 demethylation by LSD1 to promote cell invasion. Proc Natl Acad Sci U S A 114(15): 3909–3914.

Carnesecchi J, Vanacker J-M (2016) Estrogen-Related Receptors and the control of bone cell fate. Mol Cell Endocrinol 432: 37–43.

Castet A, Herledan A, Bonnet S, Jalaguier S, Vanacker J-M, Cavaillès V (2006) Receptor-interacting protein 140 differentially regulates estrogen receptor-related receptor transactivation depending on target genes. Mol Endocrinol 20(5): 1035–1047.

Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Pérez N, Fornes O, Leung TY, et al. (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res 50(D1): D165–D173.

Cavallini A, Notarnicola M, Giannini R, Montemurro S, Lorusso D, Visconti A, Minervini F, Caruso MG (2005) Oestrogen receptor-related receptor alpha (ERRalpha) and oestrogen receptors (ERalpha and ERbeta) exhibit different gene expression in human colorectal tumour progression. Eur J Cancer 41(10): 1487–1494.

Cerutti C, Shi J-R, Vanacker J-M (2023) Multifaceted Transcriptional Network of Estrogen-Related Receptor Alpha in Health and Disease. Int J Mol Sci 24(5): 4265.

Cerutti C, Zhang L, Tribollet V, Shi J-R, Brillet R, Gillet B, Hughes S, Forcet C, Shi T-L, Vanacker J-M (2022) Computational identification of new potential transcriptional partners of ERRα in breast cancer cells: specific partners for specific targets. Sci Rep 12(1): 3826.

Chakraborty P, George JT, Tripathi S, Levine H, Jolly MK (2020) Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum. Front Bioeng Biotechnol 8: 220.

Chang C, Kazmin D, Jasper JS, Kunder R, Zuercher WJ, McDonnell DP (2011) The metabolic regulator ERRα, a downstream target of HER2/IGF-1R, as a therapeutic target in breast cancer. Cancer Cell 20(4): 500–510.

Chen H, Pan J, Zhang L, Chen L, Qi H, Zhong M, Shi X, Du J, Li Q (2018) Downregulation of estrogen-related receptor alpha inhibits human cutaneous squamous cell carcinoma cell proliferation and migration by regulating EMT via fibronectin and STAT3 signaling pathways. European Journal of Pharmacology 825: 133–142.

Chen J, Liu H, Chen J, Sun B, Wu J, Du C (2020) PLXNC1 Enhances Carcinogenesis Through Transcriptional Activation of IL6ST in Gastric Cancer. Front Oncol 10: 33.

Chen P, Li B, Ou-Yang L (2022) Role of estrogen receptors in health and disease. Front Endocrinol (Lausanne) 13: 839005.

Chen T (2008) Nuclear receptor drug discovery. Curr Opin Chem Biol 12(4): 418–426.

Chen Y, Zhang K, Li Y, He Q (2017) Estrogen-related receptor α participates transforming growth factor-β (TGF-β) induced epithelial-mesenchymal transition of osteosarcoma cells. Cell Adh Migr 11(4): 338–346.

Christian M, White R, Parker MG (2006) Metabolic regulation by the nuclear receptor corepressor RIP140. Trends Endocrinol Metab 17(6): 243–250.

Chun H, Keleş S (2010) Sparse partial least squares regression for simultaneous

dimension reduction and variable selection. J R Stat Soc Series B Stat Methodol 72(1): 3–25.

Chung W, Eum HH, Lee H-O, Lee K-M, Lee H-B, Kim K-T, Ryu HS, Kim S, Lee JE, Park YH, Kan Z, Han W, et al. (2017) Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun 8: 15081.

Clusan L, Ferrière F, Flouriot G, Pakdel F (2023) A Basic Review on Estrogen Receptor Signaling Pathways in Breast Cancer. Int J Mol Sci 24(7): 6834.

Collignon J, Lousberg L, Schroeder H, Jerusalem G (2016) Triple-negative breast cancer: treatment challenges and solutions. Breast Cancer (Dove Med Press) 8: 93–107.

Collin RWJ, Kalay E, Tariq M, Peters T, Zwaag B van der, Venselaar H, Oostrik J, Lee K, Ahmed ZM, Caylan R, Li Y, Spierenburg HA, et al. (2008) Mutations of ESRRB encoding estrogen-related receptor beta cause autosomal-recessive nonsyndromic hearing impairment DFNB35. Am J Hum Genet 82(1): 125–138.

Corsa CAS, Brenot A, Grither WR, Van Hove S, Loza AJ, Zhang K, Ponik SM, Liu Y, DeNardo DG, Eliceiri KW, Keely PJ, Longmore GD (2016) The Action of Discoidin Domain Receptor 2 in Basal Tumor Cells and Stromal Cancer-Associated Fibroblasts Is Critical for Breast Cancer Metastasis. Cell Rep 15(11): 2510–2523.

Crevet L, Vanacker J-M (2020) Regulation of the expression of the estrogen related receptors (ERRs). Cell Mol Life Sci 77(22): 4573–4579.

Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, et al. (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486(7403): 346–352.

Davis SR, Baber RJ (2022) Treating menopause - MHT and beyond. Nat Rev Endocrinol 18(8): 490–502.

Deblois G, Giguère V (2013) Oestrogen-related receptors in breast cancer: control of cellular metabolism and beyond. Nat Rev Cancer 13(1): 27–36.

Deblois G, Hall JA, Perry M-C, Laganière J, Ghahremani M, Park M, Hallett M, Giguère V (2009) Genome-wide identification of direct target genes implicates estrogen-related receptor alpha as a determinant of breast cancer heterogeneity. Cancer Res 69(15): 6149–6157.

Delhon I, Gutzwiller S, Morvan F, Rangwala S, Wyder L, Evans G, Studer A, Kneissel M, Fournier B (2009) Absence of estrogen receptor-related-alpha increases osteoblastic differentiation and cancellous bone mineral density. Endocrinology 150(10): 4463–4472.

Derakhshan F, Reis-Filho JS (2022) Pathogenesis of Triple-Negative Breast Cancer. Annu Rev Pathol 17: 181–204.

Ding S, Chen X, Shen K (2020) Single-cell RNA sequencing in breast cancer: Understanding tumor heterogeneity and paving roads to individualized therapy. Cancer Commun (Lond) 40(8): 329–344.

Dittmer J (2015) The role of the transcription factor Ets1 in carcinoma. Semin Cancer Biol 35: 20–38.

Durif G (2016) Multivariate analysis of high-throughput sequencing data. Statistics [math.ST]. Université de Lyon. English. ⟨NNT : 2016LYSE1334⟩. ⟨tel-01581175⟩.

Durif G, Modolo L, Michaelsson J, Mold JE, Lambert-Lacroix S, Picard F (2018) High dimensional classification with combined adaptive sparse PLS and logistic regression. Bioinformatics 34(3): 485–493.

Dwyer MA, Joseph JD, Wade HE, Eaton ML, Kunder RS, Kazmin D, Chang C, McDonnell DP (2010) WNT11 expression is induced by estrogen-related receptor alpha and beta-catenin and acts in an autocrine manner to increase cancer cell migration. Cancer Res 70(22): 9298–9308.

Eastham AM, Spencer H, Soncin F, Ritson S, Merry CLR, Stern PL, Ward CM (2007) Epithelial-mesenchymal transition events during human embryonic stem cell differentiation. Cancer Res 67(23): 11254–11262.

Feldker N, Ferrazzi F, Schuhwerk H, Widholz SA, Guenther K, Frisch I, Jakob K, Kleemann J, Riegel D, Bönisch U, Lukassen S, Eccles RL, et al. (2020) Genome-wide cooperation of EMT transcription factor ZEB1 with YAP and AP-1 in breast cancer. EMBO J 39(17): e103209.

Fernandez AI, Geng X, Chaldekas K, Harris B, Duttargi A, Berry VL, Berry DL, Mahajan A, Cavalli LR, Győrffy B, Tan M, Riggins RB (2020) The orphan nuclear receptor estrogen-related receptor beta (ERRβ) in triple-negative breast cancer. Breast Cancer Res Treat 179(3): 585–604.

Fradet A, Sorel H, Bouazza L, Goehrig D, Dépalle B, Bellahcène A, Castronovo V, Follet H, Descotes F, Aubin JE, Clézardin P, Bonnelye E (2011) Dual function of ERRα in breast cancer and bone metastasis formation: implication of VEGF and osteoprotegerin. Cancer Res 71(17): 5728–5738.

Francois M, Donovan P, Fontaine F (2020) Modulating transcription factor activity: Interfering with protein-protein interaction networks. Semin Cell Dev Biol 99: 12–19.

Franklin Pugh B (2012) Ultra-high resolution mapping of protein-genome interactions using ChIP-exo. BMC Proc 6(S6): O27.

Fritah A, Christian M, Parker MG (2010) The metabolic coregulator RIP140: an update. Am J Physiol Endocrinol Metab 299(3): E335-340.

Fry SA, Robertson CE, Swann R, Dwek MV (2016) Cadherin-5: a biomarker for metastatic breast cancer with optimum efficacy in oestrogen receptor-positive breast cancers with vascular invasion. Br J Cancer 114(9): 1019–1026.

Fujimura T, Takahashi S, Urano T, Kumagai J, Ogushi T, Horie-Inoue K, Ouchi Y, Kitamura T, Muramatsu M, Inoue S (2007) Increased expression of estrogen-related receptor alpha (ERRalpha) is a negative prognostic predictor in human prostate cancer. Int J Cancer 120(11): 2325–2330.

Gaillard S, Dwyer MA, McDonnell DP (2007) Definition of the molecular basis for estrogen receptor-related receptor-alpha-cofactor interactions. Mol Endocrinol 21(1): 62–76.

Gambardella G, Viscido G, Tumaini B, Isacchi A, Bosotti R, Bernardo D di (2022) A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. Nat Commun 13(1): 1714.

Gao F-Y, Li X-T, Xu K, Wang R-T, Guan X-X (2023) c-MYC mediates the crosstalk between breast cancer cells and tumor microenvironment. Cell Commun Signal 21(1): 28.

Gardner TS, Faith JJ (2005) Reverse-engineering transcription control networks. Phys Life Rev 2(1): 65–88.

Ghiabi P, Jiang J, Pasquier J, Maleki M, Abu-Kaoud N, Rafii S, Rafii A (2014) Endothelial cells provide a notch-dependent pro-tumoral niche for enhancing breast cancer survival, stemness and pro-metastatic properties. PLoS One 9(11): e112424.

Giguère V (2008) Transcriptional control of energy homeostasis by the estrogen-related receptors. Endocr Rev 29(6): 677–696.

Gizard F, Robillard R, Gross B, Barbier O, Révillion F, Peyrat J-P, Torpier G, Hum DW, Staels B (2006) TReP-132 is a novel progesterone receptor coactivator required for the inhibition of breast cancer cell growth and enhancement of differentiation by progesterone. Mol Cell Biol 26(20): 7632–7644.

Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol 10(5): 593–601.

GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369(6509): 1318–1330.

Guan J, Wang Y, Wang Y, Zhuang Y, Ji G (2022) SRGS: sparse partial least squares-

based recursive gene selection for gene regulatory network inference. BMC Genomics 23(1): 782.

Guertin MJ, Zhang X, Coonrod SA, Hager GL (2014) Transient estrogen receptor binding and p300 redistribution support a squelching mechanism for estradiol-repressed genes. Mol Endocrinol 28(9): 1522–1533.

Guo CC, Majewski T, Zhang L, Yao H, Bondaruk J, Wang Y, Zhang S, Wang Z, Lee JG, Lee S, Cogdell D, Zhang M, et al. (2019) Dysregulation of EMT Drives the Progression to Clinically Aggressive Sarcomatoid Bladder Cancer. Cell Rep 27(6): 1781-1793.e4.

Guo W (2014) Concise review: breast cancer stem cells: regulatory networks, stem cell niches, and disease relevance. Stem Cells Transl Med 3(8): 942–948.

Győrffy B (2024) Integrated analysis of public datasets for the discovery and validation of survival-associated genes in solid tumors. Innovation (Camb) 5(3): 100625.

Han H, Cho J-W, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, et al. (2018a) TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res 46(D1): D380–D386.

Han H, Cho J-W, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, et al. (2018b) TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res 46(D1): D380–D386.

Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, Kim H, Cho A, Kim E, Lee T, Kim H, Kim K, et al. (2015) TRRUST: a reference database of human transcriptional regulatory interactions. Sci Rep 5: 11432.

He B, Tan K (2016) Understanding transcriptional regulatory networks using computational models. Curr Opin Genet Dev 37: 101–108.

He Q, Johnston J, Zeitlinger J (2015) ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. Nat Biotechnol 33(4): 395–401.

He R-Z, Zheng J-H, Yao H-F, Xu D-P, Yang M-W, Liu D-J, Sun Y-W, Huo Y-M (2023) ADAMTS12 promotes migration and epithelial-mesenchymal transition and predicts poor prognosis for pancreatic cancer. Hepatobiliary Pancreat Dis Int 22(2): 169–178.

Heck S, Rom J, Thewes V, Becker N, Blume B, Sinn HP, Deuschle U, Sohn C, Schneeweiss A, Lichter P (2009) Estrogen-related receptor alpha expression and function is associated with the transcriptional coregulator AIB1 in breast carcinoma. Cancer Res 69(12): 5186–5193.

Hicklin DJ, Ellis LM (2005) Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis. J Clin Oncol 23(5): 1011–1027.

Hong H, Yang L, Stallcup MR (1999) Hormone-independent transcriptional activation and coactivator binding by novel orphan nuclear receptor ERR3. J Biol Chem 274(32): 22618–22626.

Hong R, Xu B (2022) Breast cancer: an up-to-date review and future perspectives. Cancer Commun (Lond) 42(10): 913–936.

Horard B, Vanacker J-M (2003) Estrogen receptor-related receptors: orphan receptors desperately seeking a ligand. J Mol Endocrinol 31(3): 349–357.

Hörlein AJ, Näär AM, Heinzel T, Torchia J, Gloss B, Kurokawa R, Ryan A, Kamei Y, Söderström M, Glass CK (1995) Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. Nature 377(6548): 397–404.

Hu D, Li Z, Zheng B, Lin X, Pan Y, Gong P, Zhuo W, Hu Y, Chen C, Chen L, Zhou J, Wang L (2022) Cancer-associated fibroblasts in breast cancer: Challenges and opportunities. Cancer Commun (Lond) 42(5): 401–434.

Huang P, Chandra V, Rastinejad F (2010) Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. Annu Rev Physiol 72: 247–272.

Huang X, Ruan G, Sun P (2021) Estrogen-related receptor alpha copy number variation is associated with ovarian cancer histological grade. J Obstet Gynaecol Res 47(5): 1878–1883.

Ishigami-Yuasa M, Kagechika H (2020) Chemical Screening of Nuclear Receptor Modulators. Int J Mol Sci 21(15): 5512.

Jiang Y-Z, Ma D, Suo C, Shi J, Xue M, Hu X, Xiao Y, Yu K-D, Liu Y-R, Yu Y, Zheng Y, Li X, et al. (2019) Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. Cancer Cell 35(3): 428-440.e5.

Jin L, Li Y (2010) Structural and functional insights into nuclear receptor signaling. Adv Drug Deliv Rev 62(13): 1218–1226.

Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316(5830): 1497–1502.

Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpää MJ, Bonke M, Palin K, et al. (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res 20(6): 861–873.

Kars MD, Işeri OD, Gündüz U (2010) Drug resistant breast cancer cells overexpress ETS1 gene. Biomed Pharmacother 64(7): 458–462.

Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, Jagodnik KM, Kropiwnicki E, Wang Z, Ma'ayan A (2019) ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res 47(W1): W212–W224.

Kiliti AJ, Sharif GM, Martin MB, Wellstein A, Riegel AT (2023) AIB1/SRC-3/NCOA3 function in estrogen receptor alpha positive breast cancer. Front Endocrinol (Lausanne) 14: 1250218.

Kim G-C, Kwon H-K, Lee C-G, Verma R, Rudra D, Kim T, Kang K, Nam JH, Kim Y, Im S-H (2018) Upregulation of Ets1 expression by NFATc2 and NFKB1/RELA promotes breast cancer cell invasiveness. Oncogenesis 7(11): 91.

Kim MS, Lee WS, Jeong J, Kim S-J, Jin W (2015) Induction of metastatic potential by TrkB via activation of IL6/JAK2/STAT3 and PI3K/AKT signaling in breast cancer. Oncotarget 6(37): 40158–40171.

Koning T, Cordova F, Aguilar G, Sarmiento J, Mardones GA, Boric M, Varas-Godoy M, Lladser A, Duran WN, Ehrenfeld P, Sanchez FA (2023) S-Nitrosylation in endothelial cells contributes to tumor cell adhesion and extravasation during breast cancer metastasis. Biol Res 56(1): 51.

Koppenol WH, Bounds PL, Dang CV (2011) Otto Warburg's contributions to current concepts of cancer metabolism. Nat Rev Cancer 11(5): 325–337.

Koren S, Bentires-Alj M (2015) Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy. Mol Cell 60(4): 537–546.

Kumar T, Nee K, Wei R, He S, Nguyen QH, Bai S, Blake K, Pein M, Gong Y, Sei E, Hu M, Casasent AK, et al. (2023) A spatially resolved single-cell genomic atlas of the adult human breast. Nature 620(7972): 181–191.

Laan S van der, Golfetto E, Vanacker J-M, Maiorano D (2014) Cell cycle-dependent expression of Dub3, Nanog and the p160 family of nuclear receptor coactivators (NCoAs) in mouse embryonic stem cells. PLoS One 9(4): e93663.

Lamouille S, Xu J, Derynck R (2014) Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol 15(3): 178–196.

Lê Cao K-A, Rossouw D, Robert-Granié C, Besse P (2008) A sparse PLS for variable selection when integrating omics data. Stat Appl Genet Mol Biol 7(1): Article 35.

Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res 21(7): 1109–1121.

Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA

(2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 121(7): 2750–2767.

Lehmann BD, Pietenpol JA (2015) Clinical implications of molecular heterogeneity in triple negative breast cancer. Breast 24 Suppl 2(0 2): S36-40.

Lehmann W, Mossmann D, Kleemann J, Mock K, Meisinger C, Brummer T, Herr R, Brabletz S, Stemmler MP, Brabletz T (2016) ZEB1 turns into a transcriptional activator by interacting with YAP1 in aggressive cancer types. Nat Commun 7: 10498.

Leonard M, Zhang X (2019) Estrogen receptor coactivator Mediator Subunit 1 (MED1) as a tissue-specific therapeutic target in breast cancer. J Zhejiang Univ Sci B 20(5): 381–390.

Lever J, Krzywinski M, Altman N (2016) Model selection and overfitting. Nat Methods 13(9): 703–704.

Li C-J, Tzeng Y-DT, Hsiao J-H, Tseng L-M, Hsu T-S, Chu P-Y (2024) Spatial and single-cell explorations uncover prognostic significance and immunological functions of mitochondrial calcium uniporter in breast cancer. Cancer Cell Int 24(1): 140.

Li H, Liu W, Zhang X, Wang Y (2021) Cancer-associated fibroblast-secreted collagen triple helix repeat containing-1 promotes breast cancer cell migration, invasiveness and epithelial-mesenchymal transition by activating the Wnt/β-catenin pathway. Oncol Lett 22(6): 814.

Li X-L, Li J-L, Qiu D-J, Ma L (2022a) Methylation-mediated expression of SPARC is correlated with tumor progression and poor prognosis of breast cancer. Neoplasma 69(4): 794–806.

Li Y, Wu T, Peng Z, Tian X, Dai Q, Chen M, Zhu J, Xia S, Sun A, Yang W, Lin Q (2022b) ETS1 is a prognostic biomarker of triple-negative breast cancer and promotes the triple-negative breast cancer progression through the YAP signaling. Am J Cancer Res 12(11): 5074–5084.

Liu D, Benlhabib H, Mendelson CR (2009) cAMP enhances estrogen-related receptor alpha (ERRalpha) transcriptional activity at the SP-A promoter by increasing its interaction with protein kinase A and steroid receptor coactivator 2 (SRC-2). Mol Endocrinol 23(6): 772–783.

Liu J, Dang H, Wang XW (2018) The significance of intertumor and intratumor heterogeneity in liver cancer. Exp Mol Med 50(1): e416.

Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G (2021) Breast cancer. Lancet 397(10286): 1750–1769.

Lonard DM, O'malley BW (2007) Nuclear receptor coregulators: judges, juries, and

executioners of cellular regulation. Mol Cell 27(5): 691–700.

Ma J-H, Qi J, Lin S-Q, Zhang C-Y, Liu F-Y, Xie W-D, Li X (2019) STAT3 Targets ERR-α to Promote Epithelial-Mesenchymal Transition, Migration, and Invasion in Triple-Negative Breast Cancer Cells. Mol Cancer Res 17(11): 2184–2195.

Manore SG, Doheny DL, Wong GL, Lo H-W (2022) IL-6/JAK/STAT3 Signaling in Breast Cancer Metastasis: Biology and Treatment. Front Oncol 12: 866014.

Markesbery WR (1997) Neuropathological criteria for the diagnosis of Alzheimer's disease. Neurobiol Aging 18(4 Suppl): S13-19.

Martelotto LG, Ng CKY, Piscuoglio S, Weigelt B, Reis-Filho JS (2014) Breast cancer intra-tumor heterogeneity. Breast Cancer Res 16(3): 210.

Martin TA, Watkins G, Lane J, Jiang WG (2005) Assessing microvessels and angiogenesis in human breast cancer, using VE-cadherin. Histopathology 46(4): 422–430.

Martinez-Outschoorn U, Sotgia F, Lisanti MP (2014) Tumor microenvironment and metabolic synergy in breast cancers: critical importance of mitochondrial fuels and function. Semin Oncol 41(2): 195–216.

May FE (2014) Novel drugs that target the estrogen-related receptor alpha: their therapeutic potential in breast cancer. Cancer Manag Res 6: 225–252.

McCafferty MPJ, McNeill RE, Miller N, Kerin MJ (2009) Interactions between the estrogen receptor, its cofactors and microRNAs in breast cancer. Breast Cancer Res Treat 116(3): 425–432.

McDonnell DP, Chang CY, Norris JD (2000) Development of peptide antagonists that target estrogen receptor-cofactor interactions. J Steroid Biochem Mol Biol 74(5): 327–335.

Mehraj U, Dar AH, Wani NA, Mir MA (2021) Tumor microenvironment promotes breast cancer chemoresistance. Cancer Chemother Pharmacol 87(2): 147–158.

Mendoza-Almanza G, Burciaga-Hernández L, Maldonado V, Melendez-Zajgla J, Olmos J (2020) Role of platelets and breast cancer stem cells in metastasis. World J Stem Cells 12(11): 1237–1254.

Milovanović J, Vujasinović T, Todorović-Raković N, Greenman J, Hranisavljević J, Radulovic M (2023) Vascular endothelial growth factor (VEGF) -A, -C and VE-cadherin as potential biomarkers in early breast cancer patients. Pathol Res Pract 252: 154923.

Misawa A, Inoue S (2015) Estrogen-Related Receptors in Breast Cancer and Prostate

Cancer. Front Endocrinol (Lausanne) 6: 83.

Mittal V (2018) Epithelial Mesenchymal Transition in Tumor Metastasis. Annu Rev Pathol 13: 395–412.

Mohamedi Y, Fontanil T, Cal S, Cobo T, Obaya ÁJ (2021) ADAMTS-12: Functions and Challenges for a Complex Metalloprotease. Front Mol Biosci 8: 686763.

Molnár K, Mészáros Á, Fazakas C, Kozma M, Győri F, Reisz Z, Tiszlavicz L, Farkas AE, Nyúl-Tóth Á, Haskó J, Krizbai IA, Wilhelm I (2020) Pericyte-secreted IGF2 promotes breast cancer brain metastasis formation. Mol Oncol 14(9): 2040–2057.

Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P, Schwartz GF, Park PK, Rosenberg AL, Brill K, Mitchell EP (2007) Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer 110(4): 876–884.

Ni Z, Huang C, Zhao H, Zhou J, Hu M, Chen Q, Ge B, Huang Q (2021) PLXNC1: A Novel Potential Immune-Related Target for Stomach Adenocarcinoma. Front Cell Dev Biol 9: 662707.

Nolan E, Lindeman GJ, Visvader JE (2023) Deciphering breast cancer: from biology to the clinic. Cell 186(8): 1708–1728.

Ogryzko VV, Schiltz RL, Russanova V, Howard BH, Nakatani Y (1996) The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell 87(5): 953–959.

Olefsky JM (2001) Nuclear receptor minireview series. J Biol Chem 276(40): 36863–36864.

Osborne CK, Bardou V, Hopp TA, Chamness GC, Hilsenbeck SG, Fuqua SAW, Wong J, Allred DC, Clark GM, Schiff R (2003) Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. J Natl Cancer Inst 95(5): 353–361.

Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X, Bryant VL, Penington JS, Di Stefano L, Tubau Ribera N, Wilcox S, Mann GB, et al. (2021) A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. EMBO J 40(11): e107333.

Pelon F, Bourachot B, Kieffer Y, Magagna I, Mermet-Meillon F, Bonnet I, Costa A, Givel A-M, Attieh Y, Barbazan J, Bonneau C, Fuhrmann L, et al. (2020) Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. Nat Commun 11(1): 404.

Pérez-Schindler J, Summermatter S, Salatino S, Zorzato F, Beer M, Balwierz PJ, Nimwegen E van, Feige JN, Auwerx J, Handschin C (2012) The corepressor NCoR1 antagonizes PGC-1α and estrogen-related receptor α in the regulation of skeletal muscle function and oxidative metabolism. Mol Cell Biol 32(24): 4913–4924.

Pommier RM, Sanlaville A, Tonon L, Kielbassa J, Thomas E, Ferrari A, Sertier A-S, Hollande F, Martinez P, Tissier A, Morel A-P, Ouzounova M, et al. (2020) Comprehensive characterization of claudin-low breast tumors reflects the impact of the cell-of-origin on cancer evolution. Nat Commun 11(1): 3431.

Pospelova M, Krasnikova V, Fionik O, Alekseeva T, Samochernykh K, Ivanova N, Trofimov N, Vavilova T, Vasilieva E, Topuzova M, Chaykovskaya A, Makhanova A, et al. (2022) Adhesion Molecules ICAM-1 and PECAM-1 as Potential Biomarkers of Central Nervous System Damage in Women Breast Cancer Survivors. Pathophysiology 29(1): 52–65.

Postigo AA, Dean DC (1999) ZEB represses transcription through interaction with the corepressor CtBP. Proc Natl Acad Sci USA 96(12): 6683–6688.

Qin W, Haroutunian V, Katsel P, Cardozo CP, Ho L, Buxbaum JD, Pasinetti GM (2009) PGC-1alpha expression decreases in the Alzheimer disease brain as a function of dementia. Arch Neurol 66(3): 352–361.

Ranhotra HS (2022) Estrogen-related receptor alpha in select host functions and cancer: new frontiers. Mol Cell Biochem 477(5): 1349–1359.

Rauluseviciute I, Launay T, Barzaghi G, Nikumbh S, Lenhard B, Krebs AR, Castro-Mondragon JA, Mathelier A (2024) Identification of transcription factor co-binding patterns with non-negative matrix factorization. Nucleic Acids Res : gkae743.

Rice JA (2007) Mathematical Statistics and Data Analysis. Thomson Brooks/Cole, Belmont, Calif.

Rochette-Egly C (2003) Nuclear receptors: integration of multiple signalling pathways through phosphorylation. Cell Signal 15(4): 355–366.

Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford) 2016: baw100.

Saatci O, Huynh-Dam K-T, Sahin O (2021) Endocrine resistance in breast cancer: from molecular mechanisms to therapeutic strategies. J Mol Med (Berl) 99(12): 1691–1710.

Sadrkhanloo M, Entezari M, Orouei S, Ghollasi M, Fathi N, Rezaei S, Hejazi ES, Kakavand A, Saebfar H, Hashemi M, Goharrizi MASB, Salimimoghadam S, et al. (2022) STAT3-EMT axis in tumors: Modulation of cancer metastasis, stemness and therapy response. Pharmacol Res 182: 106311.

Sailland J, Tribollet V, Forcet C, Billon C, Barenton B, Carnesecchi J, Bachmann A, Gauthier KC, Yu S, Giguère V, Chan FL, Vanacker J-M (2014) Estrogen-related receptor α decreases RHOA stability to induce orientated cell migration. Proc Natl Acad Sci U S A 111(42): 15108–15113.

Sampson ER, Yeh SY, Chang HC, Tsai MY, Wang X, Ting HJ, Chang C (2001) Identification and characterization of androgen receptor associated coregulators in prostate cancer cells. J Biol Regul Homeost Agents 15(2): 123–129.

Sanyal S, Matthews J, Bouton D, Kim H-J, Choi H-S, Treuter E, Gustafsson J-A (2004) Deoxyribonucleic acid response element-dependent regulation of transcription by orphan nuclear receptor estrogen receptor-related receptor gamma. Mol Endocrinol 18(2): 312–325.

Sato K, Takayama K-I, Inoue S (2023) Expression and function of estrogen receptors and estrogen-related receptors in the brain and their association with Alzheimer's disease. Front Endocrinol (Lausanne) 14: 1220150.

Schreiber SN, Emter R, Hock MB, Knutti D, Cardenas J, Podvinec M, Oakeley EJ, Kralli A (2004) The estrogen-related receptor alpha (ERRalpha) functions in PPARgamma coactivator 1alpha (PGC-1alpha)-induced mitochondrial biogenesis. Proc Natl Acad Sci U S A 101(17): 6472–6477.

Schreiber SN, Knutti D, Brogli K, Uhlmann T, Kralli A (2003) The transcriptional coactivator PGC-1 regulates the expression and activity of the orphan nuclear receptor estrogen-related receptor alpha (ERRalpha). J Biol Chem 278(11): 9013–9018.

Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, et al. (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 486(7403): 395–399.

Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M (2000) Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. Cell 103(6): 843–852.

Shi L, Dong B, Li Z, Lu Y, Ouyang T, Li J, Wang T, Fan Z, Fan T, Lin B, Wang Z, Xie Y (2009) Expression of ER-{alpha}36, a novel variant of estrogen receptor {alpha}, and resistance to tamoxifen treatment in breast cancer. J Clin Oncol 27(21): 3423–3429.

Shi S, Ma H-Y, Han X-Y, Sang Y-Z, Yang M-Y, Zhang Z-G (2022) Prognostic Significance of SPARC Expression in Breast Cancer: A Meta-Analysis and Bioinformatics Analysis. Biomed Res Int 2022: 8600419.

Shikama N, Chan HM, Krstic-Demonacos M, Smith L, Lee CW, Cairns W, La Thangue NB (2000) Functional interaction between nucleosome assembly proteins and p300/CREB-binding protein family coactivators. Mol Cell Biol 20(23): 8933–8943.

Skene PJ, Henikoff S (2017) An efficient targeted nuclease strategy for high-resolution

mapping of DNA binding sites. Elife 6: e21856.

Smith EP, Boyd J, Frank GR, Takahashi H, Cohen RM, Specker B, Williams TC, Lubahn DB, Korach KS (1994) Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. N Engl J Med 331(16): 1056–1061.

Soon PSH, Kim E, Pon CK, Gill AJ, Moore K, Spillane AJ, Benn DE, Baxter RC (2013) Breast cancer-associated fibroblasts induce epithelial-to-mesenchymal transition in breast cancer cells. Endocr Relat Cancer 20(1): 1–12.

Spaderna S, Schmalhofer O, Wahlbuhl M, Dimmler A, Bauer K, Sultan A, Hlubek F, Jung A, Strand D, Eger A, Kirchner T, Behrens J, et al. (2008) The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. Cancer Res 68(2): 537–544.

Span PN, Manders P, Heuvel JJ, Thomas CMG, Bosch RR, Beex LVAM, Sweep CGJ (2002) Expression of the transcription factor Ets-1 is an independent prognostic marker for relapse-free survival in breast cancer. Oncogene 21(55): 8506–8509.

Stein RA, McDonnell DP (2006) Estrogen-related receptor alpha as a therapeutic target in cancer. Endocr Relat Cancer 13 Suppl 1: S25-32.

Stormo GD (2013) Modeling the specificity of protein-DNA interactions. Quant Biol 1(2): 115–130.

Suzuki T, Miki Y, Moriya T, Shimada N, Ishida T, Hirakawa H, Ohuchi N, Sasano H (2004) Estrogen-related receptor alpha in human breast carcinoma as a potent prognostic factor. Cancer Res 64(13): 4670–4676.

Takai K, Le A, Weaver VM, Werb Z (2016) Targeting the cancer-associated fibroblasts as a treatment in triple-negative breast cancer. Oncotarget 7(50): 82889–82901.

Tan Z, Kan C, Sun M, Yang F, Wong M, Wang S, Zheng H (2022) Mapping Breast Cancer Microenvironment Through Single-Cell Omics. Front Immunol 13: 868813.

Tang J, Liu T, Wen X, Zhou Z, Yan J, Gao J, Zuo J (2021) Estrogen-related receptors: novel potential regulators of osteoarthritis pathogenesis. Mol Med 27(1): 5.

Tang Y, Min Z, Xiang X-J, Liu L, Ma Y-L, Zhu B-L, Song L, Tang J, Deng X-J, Yan Z, Chen G-J (2018) Estrogen-related receptor alpha is involved in Alzheimer's disease-like pathology. Exp Neurol 305: 89–96.

Tao LJ, Seo DE, Jackson B, Ivanova NB, Santori FR (2020) Nuclear Hormone Receptors and Their Ligands: Metabolites in Control of Transcription. Cells 9(12): 2606.

Tecalco-Cruz AC (2018) Molecular pathways involved in the transport of nuclear

receptors from the nucleus to cytoplasm. J Steroid Biochem Mol Biol 178: 36–44.

Tecalco-Cruz AC, Ramírez-Jarquín JO, Cruz-Ramos E (2019) Estrogen Receptor Alpha and its Ubiquitination in Breast Cancer Cells. Curr Drug Targets 20(6): 690–704.

Teyssier C, Gallet M, Rabier B, Monfoulet L, Dine J, Macari C, Espallergues J, Horard B, Giguère V, Cohen-Solal M, Chassande O, Vanacker J-M (2009) Absence of ERRalpha in female mice confers resistance to bone loss induced by age or estrogen-deficiency. PLoS One 4(11): e7942.

Thewes V, Simon R, Schroeter P, Schlotter M, Anzeneder T, Büttner R, Benes V, Sauter G, Burwinkel B, Nicholson RI, Sinn H-P, Schneeweiss A, et al. (2015) Reprogramming of the ERRα and ERα target gene landscape triggers tamoxifen resistance in breast cancer. Cancer Res 75(4): 720–731.

Tosatto A, Sommaggio R, Kummerow C, Bentham RB, Blacker TS, Berecz T, Duchen MR, Rosato A, Bogeski I, Szabadkai G, Rizzuto R, Mammucari C (2016) The mitochondrial calcium uniporter regulates breast cancer progression via HIF-1α. EMBO Mol Med 8(5): 569–585.

Toy KA, Valiathan RR, Núñez F, Kidwell KM, Gonzalez ME, Fridman R, Kleer CG (2015) Tyrosine kinase discoidin domain receptors DDR1 and DDR2 are coordinately deregulated in triple-negative breast cancer. Breast Cancer Res Treat 150(1): 9–18.

Tran A, Scholtes C, Songane M, Champagne C, Galarneau L, Levasseur M-P, Fodil N, Dufour CR, Giguère V, Saleh M (2021) Estrogen-related receptor alpha (ERRα) is a key regulator of intestinal homeostasis and protects against colitis. Sci Rep 11(1): 15073.

Treeck O, Schüler-Toprak S, Ortmann O (2020) Estrogen Actions in Triple-Negative Breast Cancer. Cells 9(11): 2358.

Treuter E, Albrektsen T, Johansson L, Leers J, Gustafsson JA (1998) A regulatory role for RIP140 in nuclear receptor activation. Mol Endocrinol 12(6): 864–881.

Treviño LS, Gorelick DA (2021) The Interface of Nuclear and Membrane Steroid Signaling. Endocrinology 162(8): bqab107.

Tribollet V, Cerutti C, Géloën A, Berger E, De Mets R, Balland M, Courchet J, Vanacker J-M, Forcet C (2022) ERRα coordinates actin and focal adhesion dynamics. Cancer Gene Ther 29(10): 1429–1438.

Tsoi H, Man EPS, Chau KM, Khoo U-S (2021) Targeting the IL-6/STAT3 Signalling Cascade to Reverse Tamoxifen Resistance in Estrogen Receptor Positive Breast Cancer. Cancers (Basel) 13(7): 1511.

Tsushida K, Tanabe K, Masuda K, Tanimura S, Miyake H, Arata Y, Sugiyama H, Wada

J (2018) Estrogen-related receptor α is essential for maintaining mitochondrial integrity in cisplatin-induced acute kidney injury. Biochem Biophys Res Commun 498(4): 918–924.

Turley SJ, Cremasco V, Astarita JL (2015) Immunological hallmarks of stromal cells in the tumour microenvironment. Nat Rev Immunol 15(11): 669–682.

Usman S, Jamal A, Bushaala A, Waseem NH, Al-Dehlawi H, Yeudall WA, Teh M-T, Tummala H, Waseem A (2022) Transcriptome Analysis Reveals Vimentin-Induced Disruption of Cell-Cell Associations Augments Breast Cancer Cell Migration. Cells 11(24): 4035.

Vanacker J-M, Forcet C (2024) ERRα: unraveling its role as a key player in cell migration. Oncogene 43(6): 379–387.

Veis DJ, O'Brien CA (2023) Osteoclasts, Master Sculptors of Bone. Annu Rev Pathol 18: 257–281.

Villena JA, Kralli A (2008) ERRalpha: a metabolic function for the oldest orphan. Trends Endocrinol Metab 19(8): 269–276.

Vrtačnik P, Ostanek B, Mencej-Bedrač S, Marc J (2014) The many faces of estrogen signaling. Biochem Med (Zagreb) 24(3): 329–342.

Waks AG, Winer EP (2019) Breast Cancer Treatment: A Review. JAMA 321(3): 288–300.

Wang M, Feng R, Chen Z, Shi W, Li C, Liu H, Wu K, Li D, Li X (2022) Identification of Cancer-Associated Fibroblast Subtype of Triple-Negative Breast Cancer. J Oncol 2022: 6452636.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10(1): 57–63.

Wang Z, Wang X, Zhou H, Dan X, Jiang L, Wu Y (2018) Long non-coding RNA CASC2 inhibits tumorigenesis via the miR-181a/PLXNC1 axis in melanoma. Acta Biochim Biophys Sin (Shanghai) 50(3): 263–272.

Wei W, Schwaid AG, Wang X, Wang X, Chen S, Chu Q, Saghatelian A, Wan Y (2016) Ligand Activation of ERRα by Cholesterol Mediates Statin and Bisphosphonate Effects. Cell Metab 23(3): 479–491.

Wilson BJ, Tremblay AM, Deblois G, Sylvain-Drolet G, Giguère V (2010) An acetylation switch modulates the transcriptional activity of estrogen-related receptor alpha. Mol Endocrinol 24(7): 1349–1358.

Wingender E, Dietze P, Karas H, Knüppel R (1996) TRANSFAC: a database on

transcription factors and their DNA binding sites. Nucleic Acids Res 24(1): 238–241.

World Health Organization (13 March 2024) Breast cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

Wu H-T, Zhong H-T, Li G-W, Shen J-X, Ye Q-Q, Zhang M-L, Liu J (2020a) Oncogenic functions of the EMT-related transcription factor ZEB1 in breast cancer. J Transl Med 18(1): 51.

Wu SZ, Roden DL, Wang C, Holliday H, Harvey K, Cazet AS, Murphy KJ, Pereira B, Al-Eryani G, Bartonicek N, Hou R, Torpy JR, et al. (2020b) Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. EMBO J 39(19): e104063.

Wu Y-M, Chen Z-J, Liu H, Wei W-D, Lu L-L, Yang X-L, Liang W-T, Liu T, Liu H-L, Du J, Wang H-S (2015) Inhibition of ERRα suppresses epithelial mesenchymal transition of triple negative breast cancer cells by directly targeting fibronectin. Oncotarget 6(28): 25588–25601.

Xiang S, Liu Y-M, Chen X, Wang Y-W, Ma R-R, Wu X-J, Gao P (2015) ZEB1 expression is correlated with tumor metastasis and reduced prognosis of breast carcinoma in Asian patients. Cancer Invest 33(6): 225–231.

Xie W, Hong H, Yang NN, Lin RJ, Simon CM, Stallcup MR, Evans RM (1999) Constitutive activation of transcription and binding of coactivator by estrogen-related receptors 1 and 2. Mol Endocrinol 13(12): 2151–2162.

Xu J, Li Q (2003) Review of the in vivo functions of the p160 steroid receptor coactivator family. Mol Endocrinol 17(9): 1681–1692.

Xu K, Tian X, Oh SY, Movassaghi M, Naber SP, Kuperwasser C, Buchsbaum RJ (2016) The fibroblast Tiam1-osteopontin pathway modulates breast cancer invasion and metastasis. Breast Cancer Res 18(1): 14.

Yi Z-J, Gong J-P, Zhang W (2017) Transcriptional co-regulator RIP140: An important mediator of the inflammatory response and its associated diseases (Review). Mol Med Rep 16(2): 994–1000.

Yin L, Duan J-J, Bian X-W, Yu S-C (2020) Triple-negative breast cancer molecular subtyping and treatment progress. Breast Cancer Res 22(1): 61.

Yoriki K, Mori T, Kokabu T, Matsushima H, Umemura S, Tarumi Y, Kitawaki J (2019) Estrogen-related receptor alpha induces epithelial-mesenchymal transition through cancer-stromal interactions in endometrial cancer. Sci Rep 9(1): 6697.

Yu Y, Xiao C-H, Tan L-D, Wang Q-S, Li X-Q, Feng Y-M (2014) Cancer-associated fibroblasts induce epithelial-mesenchymal transition of breast cancer cells through

paracrine TGF-β signalling. Br J Cancer 110(3): 724–732.

Yuan L, Lin Q, Shen F, Li Y, Li J, Xu B (2023) Mitochondrial calcium uniporter activates TFEB-driven autophagy to promote migration of breast cancer cells. Iran J Basic Med Sci 26(11): 1342–1349.

Zawati I, Jlassi A, Adouni O, Nouira M, Manai M, Rahal K, Driss M, Manai M (2022) Association of ZEB1 and Vimentin with poor prognosis in metaplastic breast cancer. Ann Diagn Pathol 59: 151954.

Zeng X, Liu C, Yao J, Wan H, Wan G, Li Y, Chen N (2021) Breast cancer stem cells, heterogeneity, targeting therapies and therapeutic implications. Pharmacol Res 163: 105320.

Zhang K, Corsa CA, Ponik SM, Prior JL, Piwnica-Worms D, Eliceiri KW, Keely PJ, Longmore GD (2013) The collagen receptor discoidin domain receptor 2 stabilizes SNAIL1 to facilitate breast cancer metastasis. Nat Cell Biol 15(6): 677–687.

Zhang L, Shi J, Ouyang J, Zhang R, Tao Y, Yuan D, Lv C, Wang R, Ning B, Roberts R, Tong W, Liu Z, et al. (2021) X-CNV: genome-wide prediction of the pathogenicity of copy number variations. Genome Med 13(1): 132.

Zhang L, Wong J, Vanacker J-M (2016) The estrogen-related receptors (ERRs): potential targets against bone loss. Cell Mol Life Sci 73(20): 3781–3787.

Zhang Q, Liu W, Zhang H-M, Xie G-Y, Miao Y-R, Xia M, Guo A-Y (2020) hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. Genomics Proteomics Bioinformatics 18(2): 120–128.

Zhao Y, Shen M, Wu L, Yang H, Yao Y, Yang Q, Du J, Liu L, Li Y, Bai Y (2023) Stromal cells in the tumor microenvironment: accomplices of tumor progression? Cell Death Dis 14(9): 587.

Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, Chen C-H, Brown M, Zhang X, Meyer CA, Liu XS (2019) Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res 47(D1): D729–D735.

Zhou Y, Lin F, Wan T, Chen A, Wang H, Jiang B, Zhao W, Liao S, Wang S, Li G, Xu Z, Wang J, et al. (2021) ZEB1 enhances Warburg effect to facilitate tumorigenesis and metastasis of HCC by transcriptionally activating PFKM. Theranostics 11(12): 5926–5938.

Zou Z, Ohta T, Oki S (2024) ChIP-Atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data. Nucleic Acids Res 52(W1): W45–W53.

ANNEXE 1

Publication list

1. **Jing-Ru Shi**, Coralie Poulard, Catherine Cerutti, Olivier Trédan, Muriel Le Romancer, Tie-Liu Shi, Jean-Marc Vanacker. (2024) An ERRα-ZEB1 transcriptional signature predicts survival in triple-negative breast cancers. **Under review in Breast Cancer Research.**

2. Catherine Cerutti, **Jing-Ru Shi**, Jean-Marc Vanacker. (2023) Multifaceted Transcriptional Network of Estrogen-Related Receptor Alpha in Health and Disease. ***Int J Mol Sci*. 24(5):4265.**

3. Catherine Cerutti, Ling Zhang, Violaine Tribollet, **Jing-Ru Shi**, Riwan Brillet, Benjamin Gillet, Sandrine Hughes, Christelle Forcet, Tie-Liu Shi, Jean-Marc Vanacker. (2022) Computational identification of new potential transcriptional partners of ERRα in breast cancer cells: specific partners for specific targets. ***Sci Rep***. 12(1):3826.

4. Li Zhang[#], **Jingru Shi[#]**, Jian Ouyang[#], Riquan Zhang Yiran Tao, Dongsheng Yuan, Chengkai Lv, Ruiyuan Wang, Baitang Ning, Ruth Roberts, Weida Tong, Zhichao Liu, Tieliu Shi. (2021). X-CNV: genome-wide prediction of the pathogenicity of copy number variations. ***Genome Med.*** 13(1):132. [#]= ***co first author***

5. Meng Ren[#], **Jingru Shi[#]**, Jinmeng Jia, Yongli Guo, Xin Ni, Tieliu Shi. (2020) Genotype-phenotype correlations of Berardinelli-Seip congenital lipodystrophy and novel candidate genes prediction. ***Orphanet J Rare Dis.*** 15(1):108. [#]= ***co first author***

6. **Jingru Shi[#]**, Meng Ren[#], Jinmeng Jia, Muxue Tang, Yongli Guo, Xin Ni and Tieliu Shi. (2019). Genotype-Phenotype Association Analysis Reveals New Pathogenic Factors for Osteogenesis Imperfecta Disease. ***Front Pharmacol.*** 10:1200. [#]= ***co first author***

ANNEXE 2

Articles published

(Three first pages presented for Article 1
which is under review on Oct. 14th, 2024)

An ERRα-ZEB1 transcriptional signature predicts survival in triple-negative breast cancers.

Jing-Ru Shi [1, 2, 3], Coralie Poulard [3], Catherine Cerutti [2], Olivier Trédan [3], Muriel Le Romancer [3], Tie-Liu Shi [1], Jean-Marc Vanacker [2, 3, 4]

[1] The Center for Bioinformatics and Computational Biology, Shanghai Key Laboratory of Regulatory Biology, Institute for Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China.

[2] Institut de Génomique Fonctionnelle de Lyon, Université Claude Bernard Lyon 1, CNRS UMR5242, Ecole Normale Supérieure de Lyon, 69007 Lyon, France.

[3] Centre de Recherche en Cancérologie de Lyon, CNRS UMR5286, INSERM U1052, Université Claude Bernard Lyon I, Lyon, France

[4] to whom correspondence should be addressed: jean-marc.vanacker@lyon.unicancer.fr

Keywords: transcription, breast cancer, co-regulators, ERRα, ZEB1

Abstract

Transcription factors (TF) act together with co-regulators to modulate the expression of their target genes, which eventually dictates their pathophysiological effects. Depending on the co-regulator, TF can exert different activities. The Estrogen Related Receptor α (ERRα) acts as a transcription factor that regulates several pathophysiological phenomena. In particular, interactions with PGC1 co-activators are responsible for the metabolic activities of ERRα. In breast cancers, ERRα exerts several tumor-promoting, metabolism-unrelated activities that do not depend on PGC1, questioning the identity of the co-activators involved in these cancer-related effects. Using bio-computing methods, we here suggest ZEB1 as a major ERRα co-factor that could be responsible for the expression of direct ERRα targets in triple-negative breast cancers (TNBC). Experimental validations establish that ERRα and ZEB1 interact together and are bound to the promoters of their target genes that they transcriptionally regulate. Our further analyses show that the ERRα-ZEB1 downstream signature can predict the survival of the TNBC patients.

Competing interests.

This work was funded by Ligue contre le Cancer (comité Ardèche and Drôme) and JoRiss/ENS research programs. JRS is funded by the Chinese Scholarship Council (CSC). OT received honoraria for lectures, presentations, and advisory board from Roche, Pfizer, Novartis-Sandoz, Lilly, MSD, Astra-Zeneca, Pierre Fabre, Seagen, Daiichi-Sankyo, Gilead, Eisai, Menarini-Stemline, Veracyte and Exact Sciences. Funders had no role in the design of the study, collection and analysis of data and decision to publish.

Introduction

Transcription factors (TFs) can be defined as proteins that bind to discrete response elements on the promoter and/or enhancer of their target genes and regulate their expression. However, to achieve the latter function, TFs require the binding of additional co-factors. Directly or indirectly, these factors will induce modifications of the chromatin on the vicinity of the target transcriptional start site, allowing the recruitment of active or repressive transcription modulating complexes. Literature shows an increased interest in studying the functional interactions between transcription factors and their co-factors. For instance, members of the TEAD family can recruit the YAP/TAZ co-activators to enhance the expression of their targets [1]. In contrast, binding of the VGLL4 co-repressor to TEAD will reduce the expression of these targets [2]. Though the Glucocorticoid Receptor (GR) acts as a transcriptional activator on its corresponding binding sites, it will antagonize the activation exerted by AP1 or NFκB on genes involved in inflammation (reviewed in [3]). Co-factors thus appear as instrumental in determining the activities of individual TFs.

ERRα (encoded by *ESRRA*) is a member of the nuclear receptor superfamily that acts as a transcription factor involved in various pathophysiologic effects. For instance, this protein is highly active in the regulation of cellular metabolism that it regulates by binding to the promoters of its metabolic targets such as *MCAD*, *IDH3A* and *ATP5B* and recruitment of members of the PGC-1 family of co-activators (reviewed in [4, 5]). These functions are operated in tissues displaying a high energy demand such as liver, muscle or adipose. Remarkably, in breast cancer cells, ERRα does not regulate these metabolic targets although it is bound to their promoters [6, 7]. In contrast, ERRα displays an array of cancer-related functions in these cells through the regulation of distinct transcriptional targets, such as *WNT11*, *TNFAIP1* or *MMP1* (reviewed in [8]). The receptor indeed modulates proliferation, resistance to hypoxia, angiogenesis, cell adhesion and migration, as well as extracellular matrix invasion [6, 9-13]. It is thus possible that co-factors differing from PGC-1s and that remain to be identified contribute to these activities. These data also suggest that ERRα could be used as a pharmacological target to be inactivated in various types of cancers. However, inhibiting the receptor could also impact its other activities, and could therefore yield unwanted and detrimental side effects. As an example, mice depleted from ERRα do not resist to cold exposure [14]. Identifying ERRα co-factors that are selectively involved in its cancer functions could suggest new targets to be examined against cancers. Several factors have already been identified as interacting with ERRα and modulating its cancer-promoting functions. This is the case for NCOA3 (AIB1, ref. [15]), NCOA2 (SRC-2, ref. [16]), KDM1A (LSD1, ref. [12]), SETD7 [7] or NRIP1 (RIP140, ref. [17]). These factors have been identified

on a gene candidate basis and this illustrates the lack of a general unbiased method to identify ERRα coregulators outside metabolic cells.

Here we used direct ERRα transcriptional targets that we have identified by RNA-Seq and ChIP-Seq in triple-negative breast cancer (TNBC) cells. This gene list was submitted to bio-computing approaches to suggest transcription factors that, together with ERRα, could be involved in their high expression in breast cancer samples. These methods suggested ZEB1 as the main factor responsible for these activities. This hypothesis was experimentally verified. We focus on eight target genes that we show to be transcriptionally regulated by ZEB1 and ERRα in TNBC cells. These two factors interact together and are recruited to ERRα response elements on the promoter of their target genes. Importantly, high expression of their transcriptional signature confer a poor survival prognosis to TNBC.

Results

We investigated the mechanisms through which ERRα positively regulates the expression of its target genes in breast cancers (workflow presented on Supplementary Figure S1). We first performed a transcriptomic analysis of this factor in a cellular model (Figure 1a). To this end, MDA-MB231 cells were grown in culture medium droplets after transfection with two different siRNAs directed against ERRα and RNA sequencing was performed. We observed that the expression of 917 genes (referred to as DEGs; Differentially Expressed Genes) were strongly deregulated by both siRNAs (Supplementary Table S2). These included 629 genes that displayed a reduced expression in the absence of ERRα (i.e. stimulated by the receptor). Comparison with our previously published ChIP-Seq data [7] revealed that 190 of these genes presented an enrichment peak for ERRα with a consensus ERRE (ERRα response element) at peak summit. We thus focused on these directly activated DEGs (listed on Supplementary Table S2). Gene Ontology (GO) analysis of this gene list showed a massive enrichment in terms related to cell migration (Figure 1b), consistent with the demonstrated role of ERRα in this phenomenon [6, 10, 18]. This GO term comprised 45 DEGs on which further analyses were conducted. We next established a list of transcriptional regulators (TRs) expressed in breast cancer (Supplementary Figure S2a). To this end, a comprehensive set of 2175 TRs was first collected from different databases [7] and submitted to expression-based filtering and PCA analysis of RNA-Seq data from TCGA breast cancers. Altogether, this produced a list of 493 breast cancer relevant TRs (Supplementary Table S3) that were used for model computation.

International Journal of
Molecular Sciences

MDPI

Review

Multifaceted Transcriptional Network of Estrogen-Related Receptor Alpha in Health and Disease

Catherine Cerutti, Jing-Ru Shi † and Jean-Marc Vanacker *,†

Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, CNRS UMR5242,
Ecole Normale Supérieure de Lyon, 69342 Lyon, France
* Correspondence: jean-marc.vanacker@lyon.unicancer.fr; Tel.: +33-(0)478782828
† Present address: Centre de Recherche en Cancérologie de Lyon, CNRS UMR5286, Inserm U1052, Université de Lyon, 69008 Lyon, France.

Abstract: Estrogen-related receptors (ERRα, β and γ in mammals) are orphan members of the nuclear receptor superfamily acting as transcription factors. ERRs are expressed in several cell types and they display various functions in normal and pathological contexts. Amongst others, they are notably involved in bone homeostasis, energy metabolism and cancer progression. In contrast to other nuclear receptors, the activities of the ERRs are apparently not controlled by a natural ligand but they rely on other means such as the availability of transcriptional co-regulators. Here we focus on ERRα and review the variety of co-regulators that have been identified by various means for this receptor and their reported target genes. ERRα cooperates with distinct co-regulators to control the expression of distinct sets of target genes. This exemplifies the combinatorial specificity of transcriptional regulation that induces discrete cellular phenotypes depending on the selected coregulator. We finally propose an integrated view of the ERRα transcriptional network.

Keywords: nuclear receptor; transcriptional regulator; target gene; regulatory network; metabolism; cancer

Citation: Cerutti, C.; Shi, J.-R.; Vanacker, J.-M. Multifaceted Transcriptional Network of Estrogen-Related Receptor Alpha in Health and Disease. *Int. J. Mol. Sci.* **2023**, *24*, 4265. https://doi.org/10.3390/ijms24054265

Academic Editor: Catherine Teyssier

Received: 12 January 2023
Revised: 15 February 2023
Accepted: 18 February 2023
Published: 21 February 2023

1. Introduction

In eukaryotes, regulation of gene expression relies on a combinatorial interplay between DNA-binding transcription factors (TFs) and non-DNA binding coactivators or corepressors. Among non-DNA-binding co-regulators, those involved in histone modifications are of importance to control chromatin accessibility and the dynamics of the transcriptional process [1]. The coordinated activity of all these cooperating components results in specific spatiotemporal effects on target gene expression [2–5]. Pairwise interactions between TFs or between TF and non-DNA binding coactivators can be demonstrated at the protein level [6–8]. However, simultaneous cooperative recruitment of more than two transcriptional partners may occur and is currently difficult to demonstrate experimentally.

Nuclear receptors (NRs) form a family of transcription factors whose activities are generally controlled by the recruitment of specific, endogenous ligands. NRs are present in all animals and 21 of them have been identified in *D. melanogaster* vs 48 in *H. sapiens*. NR proteins are organized in a similar manner. They comprise an N-terminal domain that can contribute to ligand-independent transcriptional activities, a centrally located DNA-binding domain (DBD) containing two Zn fingers, a hinge domain and a C-terminally located Ligand Binding Domain (LBD). The LBD is also involved in receptor homo- or heterodimerization. Furthermore, ligand recruitment induces a conformational change in the LBD that allows interactions with transcriptional cofactors, leading to the modulation of target gene expression. The DBD and, to a lesser extent, the LBD are the most conserved domains of NRs across evolution. The transcriptional activities exerted by NRs also require a large set of proteins to modulate chromatin structure and to recruit the basal transcription machinery. As for most of the TFs, the involvement of cofactors is both dynamic and hierarchical. Primary cofactors have been proposed as those directly binding to NRs to

scientific reports

Check for updates

OPEN

Computational identification of new potential transcriptional partners of ERRα in breast cancer cells: specific partners for specific targets

Catherine Cerutti[1], Ling Zhang[1], Violaine Tribollet[1], Jing-Ru Shi[1,2], Riwan Brillet[1], Benjamin Gillet[1], Sandrine Hughes[1], Christelle Forcet[1], Tie-Liu Shi[2] & Jean-Marc Vanacker[1]✉

Estrogen related receptors are orphan members of the nuclear receptor superfamily acting as transcription factors (TFs). In contrast to classical nuclear receptors, the activities of the ERRs are not controlled by a natural ligand. Regulation of their activities thus relies on availability of transcriptional co-regulators. In this paper, we focus on ERRα, whose involvement in cancer progression has been broadly demonstrated. We propose a new approach to identify potential co-activators, starting from previously identified ERRα-activated genes in a breast cancer (BC) cell line. Considering mRNA gene expression from two sets of human BC cells as major endpoint, we used sparse partial least squares modeling to uncover new transcriptional regulators associated with ERRα. Among them, *DDX21*, *MYBBP1A*, *NFKB1*, and *SETD7* are functionally relevant in MDA-MB-231 cells, specifically activating the expression of subsets of ERRα-activated genes. We studied SET7 in more details and showed its co-localization with ERRα and its ERRα-dependent transcriptional and phenotypic effects. Our results thus demonstrate the ability of a modeling approach to identify new transcriptional partners from gene expression. Finally, experimental results show that ERRα cooperates with distinct co-regulators to control the expression of distinct sets of target genes, thus reinforcing the combinatorial specificity of transcription.

In eukaryotes, regulation of gene expression relies on a combinatorial interplay between transcriptional regulators (TRs) including DNA-binding transcription factors (TFs) and non-DNA binding co-activators or co-repressors. Among non-DNA-binding co-regulators, those involved in histone modifications are of importance to control chromatin accessibility and the dynamics of the transcriptional process[1]. The coordinated activity of all these cooperating factors results in specific spatio-temporal effects on target gene expression[2].

Searching potential TFs by identifying TF-binding sites in pre-defined regions from the transcription start site (TSS) has often been used to unveil cooperative binding of TFs[3–5]. However, presence of binding sites is not sufficient to predict actual binding of TFs necessary for cooperation. Chromatin immunoprecipitation sequencing (ChIP-seq) studies provide actual genomic locations of TF DNA-binding. It however still remains challenging to determine whether these binding events are functional or incidental and whether they function in conjunction with other TFs nearby or at a distance. Several experimental approaches can demonstrate pairwise interactions at the protein level between TFs or between TF and non-DNA binding co-activator[6–8]. However simultaneous cooperative recruitment of more than two transcriptional partners may occur and is difficult to demonstrate experimentally. Various in silico methods have been proposed to infer either gene regulatory networks using dynamic or static mRNA gene expression[9–12] or transcriptional regulatory networks using mRNA and protein data to suggest direct relationships between regulators and target genes[13–15]. Such methods were developed in various biological contexts including breast cancer (BC) for subtypes identification[16,17]. Moreover, uncovering combinations of regulators that could be simultaneously or sequentially recruited remains to be achieved.

[1]Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS UMR5242, Ecole Normale Supérieure de Lyon, 32-34 Avenue Tony Garnier, 69007 Lyon, France. [2]The Center for Bioinformatics and Computational Biology, Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China. ✉email: jean-marc.vanacker@ens-lyon.fr

Zhang et al. Genome Medicine (2021) 13:132
https://doi.org/10.1186/s13073-021-00945-4

Genome Medicine

SOFTWARE Open Access

X-CNV: genome-wide prediction of the pathogenicity of copy number variations

Li Zhang[1,2†], Jingru Shi[1†], Jian Ouyang[1†], Riquan Zhang[2], Yiran Tao[1], Dongsheng Yuan[1], Chengkai Lv[1], Ruiyuan Wang[1], Baitang Ning[3], Ruth Roberts[4,5], Weida Tong[3*], Zhichao Liu[3*] and Tieliu Shi[1,2,6*]

Abstract

Background: Gene copy number variations (CNVs) contribute to genetic diversity and disease prevalence across populations. Substantial efforts have been made to decipher the relationship between CNVs and pathogenesis but with limited success.

Results: We have developed a novel computational framework X-CNV (www.unimd.org/XCNV), to predict the pathogenicity of CNVs by integrating more than 30 informative features such as allele frequency (AF), CNV length, CNV type, and some deleterious scores. Notably, over 14 million CNVs across various ethnic groups, covering nearly 93% of the human genome, were unified to calculate the AF. X-CNV, which yielded area under curve (AUC) values of 0.96 and 0.94 in training and validation sets, was demonstrated to outperform other available tools in terms of CNV pathogenicity prediction. A meta-voting prediction (MVP) score was developed to quantitively measure the pathogenic effect, which is based on the probabilistic value generated from the XGBoost algorithm. The proposed MVP score demonstrated a high discriminative power in determining pathogenetic CNVs for inherited traits/diseases in different ethnic groups.

Conclusions: The ability of the X-CNV framework to quantitatively prioritize functional, deleterious, and disease-causing CNV on a genome-wide basis outperformed current CNV-annotation tools and will have broad utility in population genetics, disease-association studies, and diagnostic screening.

Keywords: XGBoost, Copy number variation, Pathogenicity, Next-generation sequencing, Machine learning

Background

Gene copy number variants (CNVs) are a type of structural variant (> 50 bp), characterized as duplications or deletions of genomic segments in specific DNA regions [1]. For humans, CNVs are more prevalent than single nucleotide variants (SNVs) in terms of base-pair length. On average, each individual carries approximately 1000 CNVs. On aggregate, CNVs cover ~ 4 million bp across the genome [2]. CNVs are believed to originate via diverse mutational mechanisms such as errors in replication, meiotic recombination, and repair of double-strand breaks [2]. Evidence has mounted that CNVs make a significant contribution to rare variants involved in rare diseases [3–6] and more common diseases such as cancers [7, 8] and neurodevelopmental disorders [9–11].

Rapid advancements in emerging genomics technologies provide unprecedented breadth and depth to detect single nucleotide variations [12–14] and complex structural variants such as CNVs [15–17]. Furthermore, global collaborations established by large consortium efforts have enhanced our understanding of the distribution and functionality of CNVs across different ethnic groups

* Correspondence: weida.tong@fda.hhs.gov; Zhichao.liu@fda.hhs.gov; tlshi@bio.ecnu.edu.cn
†Li Zhang, Jingru Shi and Jian Ouyang contributed equally to this work.
[3]National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR 72079, USA
[1]Center for Bioinformatics and Computational Biology, and the Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai 200241, China
Full list of author information is available at the end of the article

Ren et al. Orphanet Journal of Rare Diseases (2020) 15:108
https://doi.org/10.1186/s13023-020-01383-y

Orphanet Journal of Rare Diseases

RESEARCH Open Access

Genotype-phenotype correlations of Berardinelli-Seip congenital lipodystrophy and novel candidate genes prediction

Meng Ren[1†], Jingru Shi[1†], Jinmeng Jia[1], Yongli Guo[2,3,4*], Xin Ni[2,3,4*] and Tieliu Shi[1,5*]

Abstract

Background: Berardinelli-Seip congenital lipodystrophy (BSCL) is a heterogeneous autosomal recessive disorder characterized by an almost total lack of adipose tissue in the body. Mutations in the *AGPAT2*, *BSCL2*, *CAV1* and *PTRF* genes define I-IV subtype of BSLC respectively and clinical data indicate that new causative genes remain to be discovered. Here, we retrieved 341 cases from 60 BSCL-related studies worldwide and aimed to explore genotype-phenotype correlations based on mutations of AGPAT2 and BSCL2 genes from 251 cases. We also inferred new candidate genes for BSCL through protein-protein interaction and phenotype-similarity.

Results: Analysis results show that BSCL type II with earlier age of onset of diabetes mellitus, higher risk to suffer from premature death and mental retardation, is a more severe disorder than BSCL type I, but BSCL type I patients are more likely to have bone cysts. In BSCL type I, females are at higher risk of developing diabetes mellitus and acanthosis nigricans than males, while in BSCL type II, males suffer from diabetes mellitus earlier than females. In addition, some significant correlations among BSCL-related phenotypes were identified. New candidate genes prediction through protein-protein interaction and phenotype-similarity was conducted and we found that CAV3, EBP, SNAP29, HK1, CHRM3, OBSL1 and DNAJC13 genes could be the pathogenic factors for BSCL. Particularly, CAV3 and EBP could be high-priority candidate genes contributing to pathogenesis of BSCL.

Conclusions: Our study largely enhances the current knowledge of phenotypic and genotypic heterogeneity of BSCL and promotes the more comprehensive understanding of pathogenic mechanisms for BSCL.

Keywords: Berardinelli-Seip congenital lipodystrophy, Genotype, Phenotype, Gene prediction, Protein-protein interaction, Phenotype-similarity

* Correspondence: guoyongli@bch.com.cn; nixin@bch.com.cn; tlshi@bio.ecnu.edu.cn
†Meng Ren and Jingru Shi contributed equally to this work.
[2]Beijing Key Laboratory for Pediatric Diseases of Otolaryngology, Head and Neck Surgery, MOE Key Laboratory of Major Diseases in Children, Beijing Children's Hospital, National Center for Children's Health, Beijing Pediatric Research Institute, Capital Medical University, Beijing, China
[1]Center for Bioinformatics and Computational Biology, and the Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China
Full list of author information is available at the end of the article

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 15 October 2019
doi: 10.3389/fphar.2019.01200

Genotype–Phenotype Association Analysis Reveals New Pathogenic Factors for Osteogenesis Imperfecta Disease

Jingru Shi[1†], Meng Ren[1†], Jinmeng Jia[1], Muxue Tang[1], Yongli Guo[2,3,4], Xin Ni[2,3,4*] and Tieliu Shi[1,2*]*

[1] Center for Bioinformatics and Computational Biology, and the Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China, [2] Big Data and Engineering Research Center, Beijing Key Laboratory for Pediatric Diseases of Otolaryngology, Head and Neck Surgery, MOE Key Laboratory of Major Diseases in Children, Beijing Children's Hospital, National Center for Children's Health, Beijing Pediatric Research Institute, Capital Medical University, Beijing, China, [3] Biobank for Clinical Data and Samples in Pediatrics, Beijing Children's Hospital, National Center for Children's Health, Beijing Pediatric Research Institute, Capital Medical University, Beijing, China, [4] Department of Otolaryngology, Head and Neck Surgery, Beijing Children's Hospital, National Center for Children's Health, Capital Medical University, Beijing, China

OPEN ACCESS

Edited by:
Peter Vee Sin Lee,
The University of Melbourne,
Australia

Reviewed by:
Amit P. Bhavsar,
University of Alberta,
Canada
Qing Lyu,
University of Rochester,
United States

***Correspondence:**
Yongli Guo
guoyongli@bch.com.cn
Xin Ni
nixin@bch.com.cn
Tieliu Shi
tlshi@bio.ecnu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Translational Pharmacology, a section of the journal Frontiers in Pharmacology

Received: 16 November 2018
Accepted: 17 September 2019
Published: 15 October 2019

Citation:
Shi J, Ren M, Jia J, Tang M, Guo Y, Ni X and Shi T (2019) Genotype–Phenotype Association Analysis Reveals New Pathogenic Factors for Osteogenesis Imperfecta Disease. Front. Pharmacol. 10:1200. doi: 10.3389/fphar.2019.01200

Osteogenesis imperfecta (OI), mainly caused by structural abnormalities of type I collagen, is a hereditary rare disease characterized by increased bone fragility and reduced bone mass. Clinical manifestations of OI mostly include multiple repeated bone fractures, thin skin, blue sclera, hearing loss, cardiovascular and pulmonary system abnormalities, triangular face, dentinogenesis imperfecta (DI), and walking with assistance. Currently, 20 causative genes with 18 subtypes have been identified for OI, of them, variations in *COL1A1* and *COL1A2* have been demonstrated to be major causative factors to OI. However, the complexity of the bone formation process indicates that there are potential new pathogenic genes associated with OI. To comprehensively explore the underlying mechanism of OI, we conducted association analysis between genotypes and phenotypes of OI diseases and found that mutations in *COL1A1* and *COL1A2* contributed to a large proportion of the disease phenotypes. We categorized the clinical phenotypes and the genotypes based on the variation types for those 155 OI patients collected from literature, and association study revealed that three phenotypes (bone deformity, DI, walking with assistance) were enriched in two variation types (the Gly-substitution missense and groups of frameshift, nonsense, and splicing variations). We also identified four novel variations (c.G3290A (p.G1097D), c.G3289C (p.G1097R), c.G3289A (p.G1097S), c.G3281A (p.G1094D)) in gene *COL1A1* and two novel variations (c.G2332T (p.G778C), c.G2341T (p.G781C)) in gene *COL1A2*, which could potentially contribute to the disease. In addition, we identified several new potential pathogenic genes (*ADAMTS2*, *COL5A2*, *COL8A1*) based on the integration of protein–protein interaction and pathway enrichment analysis. Our study provides new insights into the association between genotypes and phenotypes of OI and novel information for dissecting the underlying mechanism of the disease.

Keywords: osteogenesis imperfecta, genotype, phenotype, novel candidate pathogenic genes, novel candidate pathogenic variations