DISCOVERY AND CHARACTERIZATION OF ISOFORM LEVEL COMPLEXITY IN ESTROGEN-INDUCED MCF7 CELLS USING LONG-READ RNA-SEQUENCING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

İBRAHİM ÖZGÜL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
MOLECULER BIOLOGY AND GENETICS

AUGUST 2025

Approval of the thesis:

DISCOVERY AND CHARACTERIZATION OF ISOFORM LEVEL COMPLEXITY IN ESTROGEN-INDUCED MCF7 CELLS USING LONG-READ RNA SEQUENCING

submitted by **İBRAHİM ÖZGÜL** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** i**n Molecular Biology and Genetics, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, **Graduate School of Natural and Applied Sciences** _____

Prof. Dr. Meryem Beklioğlu
Head of the Department, **Department of Biological Sciences** _____

Prof. Dr. Ayşe Elif Erson Bensan
Supervisor, **Department of Biological Sciences, METU** _____

Examining Committee Members:

Prof. Dr. Sreeparna Banerjee
Department of Biological Sciences, METU _____

Prof. Dr. Ayşe Elif Erson Bensan
Department of Biological Sciences, METU _____

Prof. Dr. Petek Ballar Kırmızıbayrak
Basic Pharmaceutical Sciences, Ege University _____

Assoc. Prof. Dr. Ahmet Acar
Department of Biological Sciences, METU _____

Prof. Dr. Tolga Can
Computer Science, Colorado School of Mines, USA _____

Date: 26.08.2025

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : İbrahim Özgül

Signature :

ABSTRACT

DISCOVERY AND CHARACTERIZATION OF ISOFORM LEVEL COMPLEXITY IN ESTROGEN-INDUCED MCF7 CELLS USING LONG-READ RNA SEQUENCING

Özgül, İbrahim
Doctor of Philosophy, Molecular Biology and Genetics
Supervisor: Prof. Dr. Ayşe Elif Erson Bensan

August 2025, 194 pages

Estrogen (E2) signaling is a central driver of estrogen receptor-positive (ER+) breast cancer, yet our understanding of its transcriptomic impact is constrained by the limitations of short-read sequencing, which masks the complexity of transcript isoform regulation. This thesis aimed to overcome this gap by employing long-read direct RNA sequencing (DRS) to create a isoform-resolved map of the E2 response in the MCF7 cells. Our analysis revealed that E2 stimulation triggers widespread post-transcriptional remodeling of both the coding and non-coding transcriptome. For protein-coding genes, we detected E2 induced upregulation of isoforms generated through alternative splicing (AS) and alternative polyadenylation (APA). Notably, E2-induced intronic polyadenylation (IPA) generated truncated isoforms of ER pathway regulators, exemplified by TLE1 and GREB1, lacking critical C-terminus functional domains.

The non-coding transcriptome analysis also identified a variety of isoform structures. We focused on several long non-coding RNAs (lncRNAs) and confirmed the DRS

results, exemplified by the E2-regulated SNHG family. DRS analysis revealed that the *SNHG9* isoform in MCF7 cells is a spliced transcript. Based on this isoform, we proposed a sponge function using reporter assays. Functionally, overexpression of this *SNHG9* isoform, inhibited E2-driven proliferation, and clinically, high *SNHG9* expression correlates with improved survival in ER+ patients, suggesting a role in a feedback loop that limits proliferation.

In conclusion, we deciphered a complex architecture of isoforms regulated by E2. Hence, the use of long read sequencing techniques is likely to provide a more accurate understanding of gene expression dynamics, uncovering isoform structures that are overlooked using conventional techniques.

Keywords: Breast cancer, Estradiol, Direct RNA sequencing, Transcript isoform.

ÖZ

ÖSTROJENLE İNDÜKLENMİŞ MCF7 HÜCRELERİNDE İZOFORM DÜZEYİ KARMAŞIKLIĞININ UZUN OKUMALI RNA DİZİLEME KULLANILARAK KEŞFİ VE KARAKTERİZASYONU

Özgül, İbrahim
Doktora, Moleküler Biyoloji ve Genetik
Tez Yöneticisi: Prof. Dr. Ayşe Elif Erson Bensan

Ağustos 2025, 194 sayfa

Östrojen (E2) sinyali, östrojen reseptörü-pozitif (ER+) meme kanserinin temel bir itici gücü olmasına rağmen, bu sinyalin transkriptom üzerindeki etkisine dair mevcut anlayışımız, transkript izoform düzenlemesinin karmaşıklığını maskeleyen kısa okumalı dizilemenin getirdiği kısıtlamalarla sınırlıdır. Bu tez çalışması, uzun okumalı doğrudan RNA dizileme yöntemini kullanarak (DRS) MCF7 hücre hattındaki E2 yanıtının izoform düzeyinde bir haritasını oluşturmayı ve bu sayede mevcut bilgi boşluğunu gidermeyi amaçlamıştır. Analizlerimiz, E2 uyarımının hem kodlayan hem de kodlamayan transkriptomda yaygın bir transkripsiyon sonrası yeniden şekillenmeyi tetiklediğini ortaya koymuştur. Protein kodlayan genler özelinde, E2'nin alternatif uçbirleştirme (AS) ve alternatif poliadenilasyon (APA) yoluyla üretilen izoformların ekspresyonunu artırdığı saptanmıştır. Özellikle, E2'nin tetiklediği intronik poliadenilasyon (IPA) olaylarının, TLE1 ve GREB1 ile örneklendirilen ERα yolağı düzenleyicilerinin, kritik C-ucu işlevsel alanlarından yoksun, kısaltılmış izoformlarını ürettiği gözlemlenmiştir.

Kodlamayan transkriptomun incelenmesiyle de çeşitli izoform yapıları tanımlanmıştır. E2 tarafından düzenlenen SNHG ailesi ile örneklendirilen çeşitli uzun kodlamayan RNA'lara (lncRNA) odaklanılmış ve DRS sonuçları doğrulanmıştır. DRS analizi, MCF7 hücrelerindeki *SNHG9* izoformunun uçbirleştirilmiş bir transkript olduğunu ortaya çıkarmıştır. Bu izoforma dayanarak, raportör deneyleri kullanılarak bir "sünger" işlevi önerilmiştir. İşlevsel olarak, bu *SNHG9* izoformunun aşırı ifadesi E2 kaynaklı proliferasyonu engellemiş; klinik olarak ise yüksek *SNHG9* ifadesinin ER+ hastalarda artan sağkalım ile ilişkili olması, proliferasyonu sınırlayan bir geri bildirim döngüsünde rol oynadığına işaret etmektedir.

Sonuç olarak, bu çalışma ile E2 tarafından düzenlenen karmaşık bir izoform mimarisi çözümlenmiştir. Bu nedenle, uzun okumalı dizileme tekniklerinin kullanımı, geleneksel yöntemlerle gözden kaçırılan izoform yapılarını ortaya çıkararak gen ifadesi dinamiklerinin daha doğru bir şekilde anlaşılmasını sağlayacaktır.

Anahtar Kelimeler: Meme kanseri, Estradiol, Doğrudan RNA dizileme, Transkript izoform.

to all curious minds, and the journey that never ends

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to my supervisor, Prof. Dr. Ayşe Elif Erson Bensan. Her guidance, mentorship, and critical feedback were essential to the development and completion of this work. I am also incredibly grateful for the intellectual freedom she provided, which allowed me to grow as an independent scientist. This thesis would not have been possible without her constant support and dedication.

I extend my thanks to my thesis committee members for their time and dedication. I am sincerely grateful to Prof. Dr. Sreeparna Banerjee, Assoc. Prof. Dr. Ahmet Acar, Prof. Dr. Petek Ballar Kırmızıbayrak, and Prof. Dr. Tolga Can for their critical feedback, and insightful questions, which were instrumental in improving this dissertation.

A special note of gratitude is reserved for Prof. Dr. Tolga Can for his invaluable support and expert guidance with the sequencing data analysis, which significantly strengthened my findings. I would also like to express my appreciation to my former committee member, Prof. Dr. Mesut Muyan, for his valuable input and support during my graduate studies. My gratitude is also extended to Dr. Oğuzhan Beğik for his support in establishing the long-read sequencing protocol and for the insightful and helpful discussions that followed.

I must reserve a special and heartfelt acknowledgment for Didem Naz Dioken. Her significant participation in the data analysis and interpretation was crucial for this project, and her dedicated help truly shaped the final data presented in this thesis. Beyond her scientific talents, I am profoundly grateful for her unwavering friendship and emotional support. Her encouragement, humor, and willingness to listen made the most challenging moments of this PhD journey manageable. Thank you for being a wonderful colleague and an even better friend.

I would like to thank all past and present members of the Erson Lab who created and supported a collaborative work environment. Among them, I would like to extend a special thanks to Dr. Murat Erdem. I appreciate his willingness to share his knowledge, the many insightful scientific discussions we had, and the valuable lessons he taught me, which have been fundamental to my development as a scientist.

Finally, I would like to express my deepest gratitude to my family, Durdu Özgül, Zeynep Özgül, and Pervin Çakal, for their unwavering love and support throughout this journey. This achievement would not have been possible without their constant support and encouragement.

This work was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK), through grants 120Z464 and 124Z046.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xii

LIST OF TABLES xvi

LIST OF FIGURES xvii

LIST OF ABBREVIATIONS xix

LIST OF SYMBOLS xxiii

CHAPTERS

1 INTRODUCTION 1

1.1 Breast Cancer: Epidemiology and the Central Role of Estrogen 1

1.2 Molecular Mechanisms of Estrogen Action 2

1.2.1 Classical Genomic Pathway 4

1.2.2 Non-Classical Genomic and Non-Genomic Pathways 6

1.2.3 Temporal Dynamics of Estrogen Response 9

1.3 Transcriptomic Landscape in Estrogen Signaling 11

1.3.1 Limitations of Traditional Short-Read Sequencing 11

1.3.2 The Advantages of Long-Read Sequencing 14

1.3.3 Transcript Isoform Diversity 18

1.4 Long Noncoding RNAs (lncRNAs) and Their Roles in Breast Cancer and Estrogen Signaling 21

1.4.1 Definition and Biogenesis of lncRNAs 21

1.4.2 Diverse Mechanisms of lncRNA Function 22

1.4.3 LncRNAs in Breast Cancer Pathogenesis 26
1.4.4 LncRNAs as Targets and Modulators of Estrogen Signaling 29
1.4.5 Small Nucleolar RNA Host Gene (SNHG) Family of lncRNAs 35
1.5 Rationale and aim of the Study 38
2 MATERIALS AND METHODS 41
2.1 Cell Culture and Experimental Treatments 41
2.1.1 2.1.1 Cell Lines and Culture Conditions 41
2.1.2 Hormone and Drug Treatments 41
2.2 RNA Isolation and Direct RNA Sequencing 42
2.2.1 RNA Isolation and Quality Control 42
2.2.2 Library Preparation and Nanopore Sequencing 42
2.3 Bioinformatics Analysis 43
2.3.1 Processing and Quality Assessment of Nanopore Sequencing Data 43
2.3.2 Transcriptome Reconstruction with FLAIR 43
2.3.3 Quantification and Differential Isoform Analysis 44
2.3.4 Alternative Splicing and Polyadenylation Analysis 45
2.3.5 Functional and Pathway Enrichment Analysis 45
2.4 Analysis of Public Gene Expression Data 46
2.5 Gene Expression and Isoform Validation by RT-qPCR 46
2.6 Plasmid Construction and Dual-Luciferase Reporter Assays 48
2.6.1 Plasmid Construction 48
2.6.2 Dual-Luciferase Reporter Assays 50
2.7 Functional Assays 50
2.7.1 Subcellular Fractionation 50

2.7.2 Generation of Stable Overexpression Cell Lines 51
2.7.3 Cell Proliferation (MTT) Assay 52
2.7.4 Colony Formation Assay 52
2.8 Statistical Analysis 53
3 RESULTS 55
3.1 Quality Controls 55
3.2 Differentially Expressed Genes (DEG) 61
3.2.1 Gene Set Enrichment Analysis (GSEA) 65
3.2.2 Transcriptome Landscape 67
3.3 Changes in Protein-Coding Genes 68
3.3.1 E2 induced isoform expression 68
3.3.2 Detection of Read-Through Transcripts 80
3.4 Noncoding RNAs 82
3.4.1 Differentially Expressed ncRNAs 82
3.4.2 Pathway Analysis of E2-Regulated lncRNAs 86
3.4.3 AS events in lncRNAs 88
3.4.4 Experimental Confirmations 89
4 DISCUSSION 115
5 CONCLUSION 121
REFERENCES 123
APPENDICES 155
A. Supplementary Tables for Chapter 1 155
B. Protocol for Direct RNA Sequencing 163
C. Protocol for the Isolation of Nuclear and Cytoplasmic RNA 176

D. Supplementary Figures and Tables for Chapter 3 179

CURRICULUM VITAE 193

LIST OF TABLES

TABLES

Table 1.1. Overview of Key lncRNAs in General Breast Cancer Pathogenesis. 27
Table 1.2. LncRNAs Regulated by Estrogen Signaling 31
Table 2.1. List of RT-qPCR Primers. 47
Table 2.2. Oligonucleotides Used for Plasmid Construction. 48
Table 3.1. Summary of DRS Run Statistics and Read Quality by ToulligQC Report. 58

LIST OF FIGURES

FIGURES

Figure 1.1. Erα Signaling Pathway. 3
Figure 1.2. Comparison of short-read and long-read sequencing for transcriptome analysis. 15
Figure 1.3. Nanopore direct RNA sequencing from library preparation to data analysis. 17
Figure 1.4. Diverse Molecular Functions of lncRNAs. This figure illustrates the various mechanisms by which lncRNAs exert their regulatory effects. 22
Figure 1.5. Key Functions of lncRNA SNHGs in Breast Cancer Progression. 37
Figure 3.1. Time-dependent induction of *TFF1* gene expression by E2 in MCF7 cells. 56
Figure 3.2. Quality assessment of poly(A)-enriched RNA. 57
Figure 3.3. Read-length distribution of direct RNA sequencing reads. 60
Figure 3.4. E2-induced expression of *TFF1* and *GREB1* and their transcript isoforms 62
Figure 3.5. Differentially Expressed genes and MA plots. 65
Figure 3.6. GSEA results for DEGs and Enrichment scores in each time point. 66
Figure 3.7. Summary of differentially expressed transcripts upon E2 induction. .. 68
Figure 3.8. Quantification and characterization of alternative splicing events across the E2 treatment time course. 69
Figure 3.9. Splicing events *CRB3* and *HAGH*. 71
Figure 3.10. Enrichment Analysis of ES and IR events in DEGs using g:Profiler. 72
Figure 3.11. Differential expression of 3’UTR isoforms for selected genes. 74
Figure 3.12. Annotation of Regulatory Elements in Differentially Expressed 3' UTR Sequences. 75
Figure 3.13. Differential expression of transcripts terminating in Early Introns (EI) and Middle Introns (MI) after E2 treatment. 77
Figure 3.14. IPA events for *GREB1* and *TLE1*. 79

Figure 3.15. Read-through transcripts detected by DRS. 81
Figure 3.16. Summary of differentially expressed ncRNAs upon E2 induction. 83
Figure 3.17. Heatmap of DE-lncRNAs upon E2 induction. 84
Figure 3.18. Expression profiles of known ER-regulated lncRNAs from ONT sequencing data. 85
Figure 3.19. lncRNA enrichment and pathway analysis using RNAenrich and TLSEA. 87
Figure 3.20. Alternative splicing events in lncRNAs upon E2 treatment. 88
Figure 3.21. Characterization and Expression Analysis of the *lnc-GATA3-7*. 91
Figure 3.22. Characterization of the *VMP1*/miR-21 Locus. 93
Figure 3.23. Heatmap of relative expression for detected SNHG family members. 94
Figure 3.24. Expression analysis and validation of candidate SNHG lncRNAs. 97
Figure 3.25. *SNHG9* isoforms. 98
Figure 3.26. Subcellular localization of three candidate gene transcripts. 101
Figure 3.27. Stability of candidate SNHGs. 102
Figure 3.28. *SNHG9* functions as a miRNA sponge for miR-1277-3p. 104
Figure 3.29. *SNHG9* cell line expression and survival analysis. 106
Figure 3.30. *SNHG9* overexpression, proliferation and colony formation in breast cancer cell lines. 108
Figure 3.31. Overexpression of *SNHG9* inhibits E2-induced cell proliferation. 110
Figure 3.32. The Kaplan-Meier plots for *SNHG9*. 111
Figure 3.33. Cloned promoter regions for candidate SNHGs. 113

LIST OF ABBREVIATIONS

ABBREVIATIONS

ActD: Actinomycin D

Alt. Acceptor: Alternative 3' Acceptor site usage

Alt. Donor: Alternative 5' Donor site usage

AP-1: Activator Protein 1

APA: Alternative Polyadenylation

AS: Alternative Splicing

bp: base pairs

CAGE: Cap Analysis of Gene Expression

cAMP: cyclic Adenosine Monophosphate

cDNA: complementary DNA

ceRNA: competing endogenous RNA

DE-lncRNA: Differentially Expressed lncRNA

DEG: Differentially Expressed Gene

DMEM: Dulbecco's Modified Eagle's Medium

DRS: Direct RNA Sequencing

E2: 17β-estradiol

EGFR: Epidermal Growth Factor Receptor

EI: Early Intron

EMT: Epithelial-Mesenchymal Transition

ER: Estrogen Receptor

ER+: Estrogen Receptor positive

ER-: Estrogen Receptor negative

ERAR: Estrogen Receptor Agitation-Related

ERE: Estrogen Response Element

ES: Exon Skipping

EV: Empty Vector

FBS: Fetal Bovine Serum

FDR: False Discovery Rate

FLAIR: Full-Length Alternative Isoform analysis of RNA

gDNA: genomic DNA

GF: Growth Factor

GO: Gene Ontology

GPER1: G protein-coupled estrogen receptor 1

GSEA: Gene Set Enrichment Analysis

HD: Hormone Deprived

HER2: Human Epidermal growth factor Receptor 2

HSD: Honestly Significant Difference

HSP: Heat Shock Protein

IGV: Integrative Genomics Viewer

IHC: Immunohistochemistry

IPA: Intronic Polyadenylation

IR: Intron Retention

kb: kilobases

lncRNA: long noncoding RNA

M: Molar

MAPK: Mitogen-Activated Protein Kinase

MI: Mid-Intron

miRNA: microRNA

mRNA: messenger RNA

MTT: 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide

ncRNA: noncoding RNA

NES: Normalized Enrichment Score

ONT: Oxford Nanopore Technologies

ORF: Open Reading Frame

P/S: Penicillin/Streptomycin

PacBio: Pacific Biosciences

PAS: Polyadenylation Signal

PBS: Phosphate-Buffered Saline

PI3K: Phosphoinositide 3-kinase

PR: Progesterone Receptor

PRC2: Polycomb Repressive Complex 2

PSI: Percent Spliced In

qPCR: quantitative Polymerase Chain Reaction

RBP: RNA-Binding Protein

RNA-seq: RNA sequencing

rRNA: ribosomal RNA

RT-qPCR: Reverse Transcription quantitative Polymerase Chain Reaction

SD: Standard Deviation

SEM: Standard Error of the Mean

shRNA: short hairpin RNA

siRNA: small interfering RNA

SMRT: Single-Molecule Real-Time

SNHG: Small Nucleolar RNA Host Gene

snoRNA: small nucleolar RNA

Sp1: Specificity Protein 1

SR: Serine/Arginine-rich

SRE: Serum Responsive Elements

TCGA: The Cancer Genome Atlas

TF: Transcription Factor

TLSEA: Tissue-specific LncRNA Set Enrichment Analysis

TNBC: Triple-Negative Breast Cancer

TPM: Transcripts Per Million

TSS: Transcription Start Site

UTR: Untranslated Region

WDR: Tryptophan-Aspartic acid Repeat

LIST OF SYMBOLS

SYMBOLS

ΔΔCt: Delta-delta Cycle threshold

µg: microgram

µL: microliter

°C: degrees Celsius

CO_2: Carbon Dioxide

Ct: Cycle threshold

ERα/β: Estrogen Receptor alpha/beta

H3K27ac: Histone H3 Lysine 27 acetylation

H3K4me1/me3: Histone H3 Lysine 4 monomethylation/ trimethylation

HCl: Hydrochloric acid

HR: Hazard Ratio

KCl: Potassium chloride

mg: milligram

mL: milliliter

mM: millimolar

NaOH: Sodium hydroxide

ng: nanogram

nM: nanomolar

p: p-value

CHAPTER 1

INTRODUCTION

1.1 Breast Cancer: Epidemiology and the Central Role of Estrogen

Breast cancer represents a formidable global health challenge, standing as the most frequently diagnosed malignancy among women worldwide and a principal cause of cancer-related mortality [1]. In 2022 alone, an estimated 2.3 million new cases were diagnosed globally, surpassing lung cancer as the most common cancer overall, and approximately 670,000 women succumbed to the disease [2]. This translates to breast cancer accounting for nearly one in four female cancer diagnoses and one in six female cancer deaths globally [3]. The burden of this disease is substantial, with projections indicating a continued rise; estimates suggest that by 2050, annual incidence is projected to reach 3.2 million new cases and mortality to exceed 1.1 million deaths, driven largely by demographic shifts such as population growth and aging [4].

Breast cancer is not a single entity but rather a complex and heterogeneous collection of diseases [5]. Clinically and biologically, breast cancers are broadly classified based on the expression status of three key predictive and prognostic biomarkers: the estrogen receptor (ER), the progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2) [6]. These markers, typically assessed via immunohistochemistry (IHC), guide therapeutic decision-making and provide insights into tumor behavior [6]. Gene expression profiling studies have further refined this classification into intrinsic molecular subtypes, including Luminal A (typically ER+/PR+, HER2-, low proliferation), Luminal B (ER+/PR+, HER2- or HER2+, higher proliferation), HER2-enriched (ER-/PR-, HER2+), and Basal-like

(often triple-negative: ER-/PR-/HER2-) [7,8]. Each subtype exhibits distinct clinical features, metastatic patterns, prognoses, and responses to therapy [9].

Among these subtypes, hormone-dependent breast cancer, characterized by the expression of ER (ER+), represents the majority of diagnosed cases, accounting for approximately 70–80% of all invasive breast carcinomas [10]. The growth, proliferation, and survival of these ER+ cancer cells are critically dependent on signaling pathways activated by the steroid hormone estrogen [11]. This dependency positions estrogen signaling as a central mechanism for understanding breast cancer pathogenesis and highlights it as a primary target for therapeutic intervention through endocrine therapies [11]. A detailed understanding of these mechanisms is essential for improving therapeutic interventions for patients [12]. The primary form of estrogen in humans is 17β-estradiol (E2) [13]. E2 exerts its effects by binding to its nuclear receptors, estrogen receptor alpha (ERα) and estrogen receptor beta (ERβ) [14]. Both receptors are members of the nuclear receptor superfamily of transcription factors and share structural similarities, but are encoded by distinct genes (ESR1 for ERα on chromosome 6 and ESR2 for ERβ on chromosome 14) [11]. While both ERα and ERβ are expressed in normal breast tissue and can be present in tumors, ERα is the predominant subtype found in the majority of ER+ breast cancers [15]. Well-characterized ER+ breast cancer cell lines, such as MCF7, serve as essential models for studying the molecular mechanisms of estrogen action. Extensive experimental and clinical evidence identifies ERα as the primary mediator of estrogen-driven proliferation and tumorigenesis in the breast, making it the key target for endocrine therapies such as tamoxifen and aromatase inhibitors [16].

1.2 Molecular Mechanisms of Estrogen Action

Estrogen, primarily E2, orchestrates complex cellular responses via multiple distinct signaling pathways. These mechanisms are broadly categorized into genomic pathways, initiated within the nucleus and directly impacting gene transcription, and non-genomic pathways, often initiated at the cell membrane and involving rapid

activation of intracellular signaling cascades [14] (Figure 1.1). Understanding these diverse pathways is crucial for comprehending the complex role of estrogen in both normal physiology and breast cancer development (reviewed in [11]).

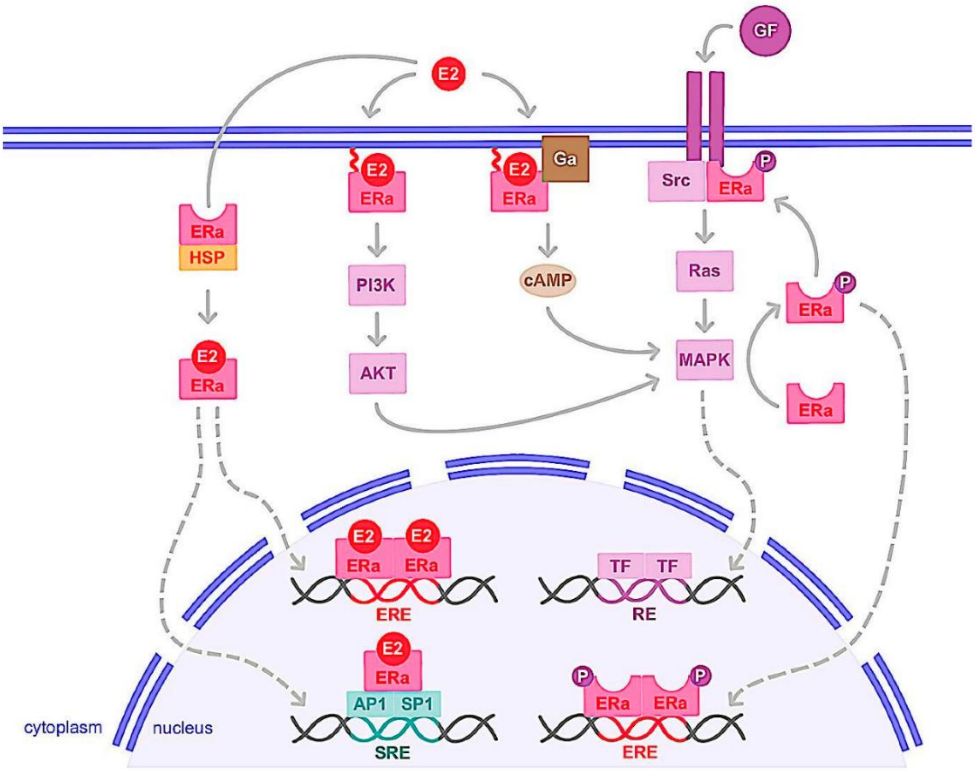


Figure 1.1. Erα Signaling Pathway. This figure illustrates the dual signaling mechanisms of Estrogen Receptor alpha (ERα), encompassing both genomic and non-genomic pathways. In the genomic pathway, ERα is activated by estrogen (E2) binding or after its phosphorylation (P) by cellular kinases following growth factor (GF) receptor stimulation, leading to its translocation into the nucleus. There, the ERα homodimer binds DNA to regulate gene transcription, either directly via estrogen responsive elements (EREs) or indirectly by tethering to other transcription factors like AP1 or SP1 that bind to serum responsive elements (SREs). Concurrently, in the non-genomic pathway, ERα can be anchored to the cell membrane and interact with G alpha proteins (Ga) or GF receptors. This activation

stimulates second messenger production, such as cyclic adenosine monophosphate (cAMP), and activates signaling pathways including phosphoinositide 3-kinase (PI3K)/AKT or Ras/mitogen-activated protein kinase (MAPK). This membrane-initiated activity ultimately leads to the activation of transcription factors (TFs) involved in regulating cell proliferation and survival. The figure was reproduced from [14].

1.2.1 Classical Genomic Pathway

The canonical or classical genomic pathway represents the most well-established mechanism of estrogen action mediated by nuclear estrogen receptors (ERs), predominantly ERα in the context of breast cancer [14]. This pathway involves a series of tightly regulated events [17]:

E2, being lipophilic, diffuses across the cell membrane and binds to ERα, which may reside in the cytoplasm complexed with chaperone proteins like heat shock protein 90 (HSP90) or within the nucleus. Ligand binding induces a critical conformational change in the ERα protein, leading to the dissociation of chaperone proteins.

The activated ERα monomers then form dimers, typically homodimers (ERα/ERα) but potentially heterodimers (ERα/ERβ) if ERβ is present [18]. These dimers translocate within the nucleus [18].

The ERα dimer recognizes and binds with high affinity to specific DNA sequences known as Estrogen Response Elements (EREs) located within the regulatory regions (promoters or enhancers) of target genes [19]. The consensus ERE is a palindromic inverted repeat, typically 13 base pairs long, with the sequence 5′-GGTCAnnnTGACC-3′, where 'n' denotes any nucleotide [20]. Each ERα monomer within the dimer contacts one half-site of the palindrome primarily within the DNA major groove, mediated by specific amino acids in the receptor's DNA-binding domain (DBD) [21]. While this consensus sequence confers high-affinity binding,

ERα can also bind to imperfect EREs or even half-sites, often influenced by flanking DNA sequences and the presence of co-regulators [22].

Upon binding to the ERE, the ligand-bound ERα dimer serves as a docking platform for a large multiprotein complex of co-regulators [17]. These co-regulators include co-activators (e.g., members of the p160/SRC family like SRC-1, SRC-2, SRC-3/AIB1; histone acetyltransferases like p300/CBP) and co-repressors (e.g., NCoR, SMRT) [23,24]. Co-activators typically possess enzymatic activities (like histone acetyltransferase activity) or recruit enzymes that modify chromatin structure (e.g., opening chromatin via histone acetylation) and facilitate the assembly of the basal transcription machinery, including RNA Polymerase II (Pol II), leading to gene activation [23]. Conversely, co-repressors often recruit histone deacetylases (HDACs) or other repressive complexes, leading to chromatin compaction and transcriptional repression [24]. The specific set of co-regulators recruited is context-dependent, influenced by the specific ERE sequence, the bound ligand (agonist vs. antagonist), post-translational modifications of ERα and co-regulators, and the cellular environment, ultimately determining the transcriptional outcome (activation or repression) of the target gene [25]. The intricate series of molecular interactions, from ligand binding to co-regulator recruitment, underscores that the classical genomic pathway is not a simple linear process but a highly tunable system. The diversity of ERE sequences, the array of available co-regulators, and the influence of the other cellular processes allow for a finely calibrated response to E2. These factors contribute to the tissue-specific effects of the hormone and the heterogeneity observed in breast cancer.

The efficiency and specificity of ERα binding and transcriptional regulation are significantly influenced by interactions with other transcription factors [26]. Notably, the pioneer factors Forkhead box protein A1 (FOXA1) and GATA binding protein 3 (GATA3) play critical roles. FOXA1 can bind to condensed chromatin, opening it up to allow subsequent ERα access to its binding sites [27]. Both FOXA1 and GATA3 frequently co-occupy genomic sites with ERα, forming functional complexes sometimes referred to as "enhanceosomes" [28]. These enhanceosomes,

often located at enhancer regions marked by high p300 and Pol II occupancy, appear necessary for regulating a core set of ERα target genes associated with breast cancer phenotypes and estrogen-dependent proliferation [28]. The dependence of ERα on pioneer factors like FOXA1 and GATA3 for accessing many of its chromatin targets highlights a hierarchical regulatory network. ERα does not act in isolation; its ability to engage with the genome and drive transcription is often follows or align with the actions of these collaborating factors. This model suggests that the levels and activity of these factors are as crucial as ERα status itself in determining estrogen responsiveness. The interplay is complex, with evidence suggesting ERα can also facilitate FOXA1 binding in some contexts [29], although some studies emphasize FOXA1's primary role as an upstream pioneer factor [30]. Other transcription factors, such as Activator Protein 1 (AP-1, a dimer typically composed of Fos and Jun family proteins) and Specificity Protein 1 (Sp1), also interact functionally and physically with ERα, often mediating estrogen responses at promoters lacking canonical EREs [31]. The requirement for these specific factors underscores the context-dependency of estrogen action; the cellular levels and activity states of FOXA1, GATA3, and various co-regulators likely shape the specific gene expression program elicited by E2 in breast cancer cells, potentially contributing to subtype differences and therapeutic response rates [28,32].

1.2.2 Non-Classical Genomic and Non-Genomic Pathways

Beyond the classical ERE-mediated pathway, estrogen may exert its influence through alternative genomic and non-genomic mechanisms, adding layers of complexity to its signaling network.

Non-Classical Genomic Signaling (Tethering): A significant portion of estrogen-responsive genes lack ERE sequences in their regulatory regions [31]. Estrogen regulates these genes through non-classical genomic pathways, primarily via a mechanism known as "tethering" [33]. In this mode, the ligand-activated ERα does not bind directly to DNA but instead interacts through protein-protein interactions

with other sequence-specific transcription factors that are already bound to their respective DNA response elements [31]. Prominent examples of transcription factors that tether ERα include AP-1 (binding to TREs - TPA Response Elements) and Sp1 (binding to GC-rich boxes) [31]. ERα can also interact with factors like NF-κB, STAT5, and C/EBPβ at their respective sites [34]. Through these interactions, ERα can function as a co-activator or co-repressor for the DNA-bound factor, modulating the transcription of the target gene without requiring an ERE [35]. This tethering mechanism significantly expands the repertoire of genes under estrogenic control and allows for complex crosstalk between estrogen signaling and other pathways (e.g., growth factor signaling often mediated via AP-1) [36]. The specific outcome (activation or repression) can depend on the ER subtype (ERα vs. ERβ), the specific ligand bound to ER, the identity of the tethering transcription factor, and the cellular context [37,38]. The existence of such tethering mechanisms means that estrogen's regulatory reach extends far beyond genes containing perfect EREs, allowing it to interface with a multitude of other signaling networks, such as those activated by growth factors, thereby contributing to its diverse physiological and pathological roles.

Non-Genomic Signaling (Membrane-Initiated): Estrogen can also elicit rapid cellular responses (occurring within minutes) that are too fast to be explained by changes in gene transcription and subsequent protein synthesis [39]. These non-genomic effects are typically initiated by E2 binding to receptors located at or near the plasma membrane [40]. These membrane receptors include a subpopulation of the classical nuclear receptors (ERα and ERβ) that are targeted to the membrane, often associated with caveolae, as well as the G protein-coupled estrogen receptor 1 (GPER1, formerly GPR30), a distinct seven-transmembrane receptor [41]. Binding of E2 to these membrane-associated receptors triggers the activation of various intracellular signaling cascades, most notably the mitogen-activated protein kinase (MAPK/ERK) pathway and the phosphoinositide 3-kinase (PI3K)/Akt pathway [42]. Activation often involves interactions with scaffold proteins and upstream kinases like Proto-oncogene tyrosine-protein kinase Src [43]. GPER1 activation,

specifically, can lead to the transactivation of the epidermal growth factor receptor (EGFR), further amplifying MAPK and PI3K/Akt signaling, as well as stimulating adenylyl cyclase activity, cyclic AMP (cAMP) production, and phospholipase C (PLC)-mediated increases in intracellular calcium [44]. These rapid, non-genomic signals can directly modulate cellular functions like ion channel activity, cell migration, and survival, and can also indirectly influence gene expression by phosphorylating and activating transcription factors (e.g., CREB, Elk-1, or even nuclear ERα itself) or by modifying chromatin structure [45]. The rapid activation of kinase cascades by membrane-initiated estrogen signaling provides a mechanism for immediate cellular responses and can also create feedback loops that modulate the slower, nuclear-initiated genomic effects, for instance, by phosphorylating nuclear ERα or its co-regulators, thereby priming or fine-tuning the transcriptional landscape.

The existence of these multiple signaling highlights the complexity of estrogen action [34]. E2 does not simply act as a switch for ERE-containing genes but rather integrates with numerous other cellular pathways and regulatory mechanisms [11]. This intricate network likely underlies the diverse biological effects of estrogen and contributes to the challenges observed in endocrine therapy, as resistance mechanisms may involve the upregulation or bypass via non-classical or non-genomic pathways when the classical ERE-dependent route is blocked [46]. This multi-level signaling capacity explains estrogen's strong effects and offers multiple avenues for cancer cells to evade therapeutic interventions, emphasizing the need for strategies that can address this signaling plasticity.

1.2.3 Temporal Dynamics of Estrogen Response

The cellular response to estrogen stimulation is not an instantaneous or static event but rather a highly dynamic process that unfolds over time, involving coordinated waves of gene activation and repression [47]. Early studies focusing on individual genes, and more recent genome-wide analyses, have revealed distinct temporal patterns in the estrogen-induced transcriptomic landscape [48].

Transcriptomic studies employing time-course analyses following E2 treatment have consistently identified sets of "early" and "late" response genes, characterized by different kinetics of induction or repression [49]. Early-response genes typically show changes in expression within minutes to a few hours of E2 exposure [49]. These genes are often direct targets of ERα, with ERα binding sites (EREs or tethering sites) located in their regulatory regions [50]. Pathway analyses indicate enrichment for canonical ER signaling and transcriptional regulation processes among these early targets [51]. Factors like ERα itself, along with its key collaborators FOXA1 and GATA3, are predicted central regulators of this early phase, which appears to occur in chromatin regions that are already relatively accessible prior to stimulation [27]. The hierarchical nature of this response, where early direct targets likely include regulators of subsequent events, suggests that interventions at this initial stage could have broad downstream consequences.

In contrast, late-response genes exhibit altered expression at later time points (e.g., >6-24 hours) [52]. The regulation of these genes may be indirect, potentially resulting from secondary effects mediated by proteins encoded by early-response genes, or may involve slower processes like extensive chromatin remodeling or progression through the cell cycle [53]. Pathway analyses of late-response genes often show enrichment for processes such as cell cycle progression, DNA replication, and metabolic pathways [54]. Interestingly, single-cell studies suggest that early responses might be more prominent in cells with high ER expression, while later responses dominate in cells with lower ERα levels, hinting at heterogeneity even within a population [55]. Furthermore, large-scale meta-analyses integrating

numerous E2 time-course datasets have revealed that while many genes show consistent upregulation or downregulation, a subset of genes can exhibit bidirectional regulation – being induced in some contexts and repressed in others [54,56]. This bidirectional behavior appears linked to distinct regulatory factors and pathways, including immune and MAPK signaling, adding another layer to the temporal complexity [57,58]. This integration with other major cellular pathways like immune and MAPK signaling further emphasizes that estrogen action is not isolated but is dynamically modulated by the broader cellular state.

Collectively, these findings paint a picture of the estrogen response as a multi-phasic process [34]. An initial wave of direct transcriptional activation and repression is followed by subsequent waves involving secondary effects, feedback mechanisms, and potentially more extensive chromatin alterations [11]. The precise kinetics and the balance between activation and repression evolve over time, highlighting the necessity of time-resolved studies to fully capture the complexity of estrogen signaling. Understanding these temporal dynamics is crucial for accurately interpreting experimental data and for potentially informing the timing and strategies of therapeutic interventions targeting the ER pathway.

Estrogen Receptor signaling pathway is integrated with the activation of key intracellular signaling cascades, including the PI3K/Akt and MAPK/ERK pathways [42,59]. This crosstalk allows estrogen to regulate a wide array of genes that lack traditional EREs, thereby integrating hormonal signals with major cellular growth and oncogenic pathways [60,61]. The activation of these pathways can also be ligand-independent, triggered by growth factors that phosphorylate the estrogen receptors, representing a critical mechanism in therapeutic resistance [46,62,63].

The activation of these pathways directly regulates well-characterized estrogen-responsive genes, including the Progesterone Receptor (*PGR*), Trefoil Factor 1 (*TFF1*), and Growth Regulating Estrogen Receptor Binding 1 (*GREB1*), which are critical mediators of estrogen-driven cell proliferation [64,65]. The expression of these and other target genes is subject to complex, multi-level control. This includes

deep integration with the cell's epigenetic machinery, where estrogen receptors recruit enzymes that direct DNA methylation and histone modifications to either activate or silence genes [66,67]. The regulation of these estrogen-responsive genes is a multi-layered process, further controlled by the epigenetic machinery, where ERs direct DNA methylation and histone modifications, and by post-transcriptional modulation via non-coding RNAs, such as miRNAs and lncRNAs, which can fine-tune the ultimate biological response [66,68–70].

This intricate network of signaling pathways ultimately converges on the transcriptome to orchestrate the cellular response. However, fully appreciating this response requires looking beyond simple gene expression changes and examining the entire landscape of RNA isoforms.

1.3 Transcriptomic Landscape in Estrogen Signaling

Understanding the ways estrogen signaling reshapes the cellular environment necessitates a deep dive into the transcriptome. While traditional methods have provided valuable insights, their inherent limitations have obscured the full complexity of RNA processing events modulated by estrogen [71]. The invention of long-read sequencing technologies offers a powerful new lens through which to view these processes [72].

1.3.1 Limitations of Traditional Short-Read Sequencing

For over a decade, RNA sequencing (RNA-seq) based on short-read technologies (predominantly Illumina) has been the cornerstone of transcriptomic analysis [73]. These methods generate vast quantities of sequence data (tens to hundreds of millions of reads per sample) at relatively low cost and high base-level accuracy (>99.9%), enabling robust quantification of overall gene expression levels and the detection of differential expression between conditions [73].

However, the fundamental principle of short-read sequencing, generating reads typically 50-150 base pairs (bp) long from fragmented RNA or cDNA molecules, imposes significant limitations on resolving the full complexity of the transcriptome [74]. Eukaryotic genes, particularly in mammals, frequently produce multiple distinct transcript isoforms through mechanisms like alternative transcription start sites (TSSs), AS, and APA [75]**.** These isoforms, while originating from the same gene locus, can encode different proteins or possess different regulatory elements within their untranslated regions (UTRs), leading to functional diversity [76]. The inability of short-read methods to fully capture this isoform diversity means that our understanding of gene regulation and cellular responses (stimuli like estrogen) may be biased if distinct isoforms are missed or their abundances miscalculated.

This challenge arises because short reads often fail to catch multiple variable regions (e.g., distant exons involved in an AS event) or the full length of the transcript [74]. Consequently, a single short read cannot be assigned to the specific transcript isoform from which it originated, especially when isoforms share significant sequence identity [72]. This ambiguity needs complex computational algorithms to infer or assemble full-length transcript structures and estimate the relative abundance of each isoform [77]. These reconstructions are difficult and prone to errors, particularly for genes with multiple, complex isoforms or those expressed at low levels [74].

Specifically, short-read RNA-seq faces difficulties in resolving isoform diversity, detecting complex AS events, APA isoforms and identifying novel isoforms [78,79].

First of all, short-read RNA-seq cannot reliably determine which distant exons are connected within the same transcript molecule, making the accurate identification and quantification of non-canonical isoforms challenging [78,80]. This issue is a direct consequence of the library preparation protocol. In short, short-read RNA-seq struggles to identify full-length isoforms because the sequencing reads (50-150 bp) are much shorter than the transcripts they originate from. Due to this fragmentation,

a single read cannot physically span and link distant exons. While short-read RNA-seq can quantify local events like the inclusion/exclusion of a single cassette exon (Percent Spliced In, PSI), resolving more complex splicing patterns involving multiple exons or alternative splice sites remains problematic [81].

Accurately mapping transcript 3' ends to identify APA sites is difficult due to biases in library preparation and sequencing that often lead to underrepresentation of transcript ends [82,83]. This hinders the analysis of 3' UTR length variation and the detection of intronic polyadenylation (IPA) events [84]. The underestimation of 3' UTR diversity is particularly concerning as these regions are critical hubs for post-transcriptional control via miRNAs and RBPs, meaning a vast layer of regulatory information is potentially missed.

Lastly, assembling novel transcript structures not present in reference annotations is particularly error-prone with short reads [85]. Therefore, while indispensable for measuring overall gene expression, traditional short-read RNA-seq provides an incomplete and potentially inaccurate view of the isoform landscape [86]. The functional diversity generated by E2-induced changes in AS and APA may be significantly underestimated or entirely missed by these approaches, hindering crucial layers of post-transcriptional regulation. This limitation means that many previous transcriptomic studies of estrogen signaling likely captured only a fraction of the molecular events occurring, highlighting a critical gap in our understanding [87,88].

1.3.2 The Advantages of Long-Read Sequencing

The emergence of third-generation sequencing technologies, primarily Oxford Nanopore Technologies (ONT) nanopore sequencing and Pacific Biosciences (PacBio) Single-Molecule Real-Time (SMRT) sequencing, has revolutionized genomics and transcriptomics by overcoming the fundamental read-length limitation of previous platforms [89]. These technologies can generate reads routinely spanning tens of kilobases (kb), and potentially exceeding a megabase (Mb) in the case of ONT, enabling the sequencing of entire DNA or RNA molecules in single reads [90,91].

Applying these technologies to transcriptomics (often termed long-read RNA-seq or Iso-Seq for PacBio) offers significant advantages. The major advantage is the ability to sequence complete, end-to-end cDNA copies of RNA transcripts, or even native RNA molecules directly (with ONT), in a single read [92]. Full-Length Transcript Sequencing eliminates the need for computational assembly of transcripts from short fragments [79]. Also, by capturing the entire structure of a transcript in one read, long-read sequencing allows for the direct and unambiguous identification of expressed isoforms, including the precise combination of exons used and the exact transcription start and end sites [81] (Figure 1.2). This drastically improves the accuracy of isoform-level quantification compared to the inferential methods required for short reads [79].

Long reads are particularly powerful for discovering novel transcript isoforms, including those with complex structures or arising from poorly annotated genes, that are often missed or incorrectly assembled by short-read approaches [74]. Comparative studies suggest long-read methods detect longer and lower-abundance transcripts more effectively [72].

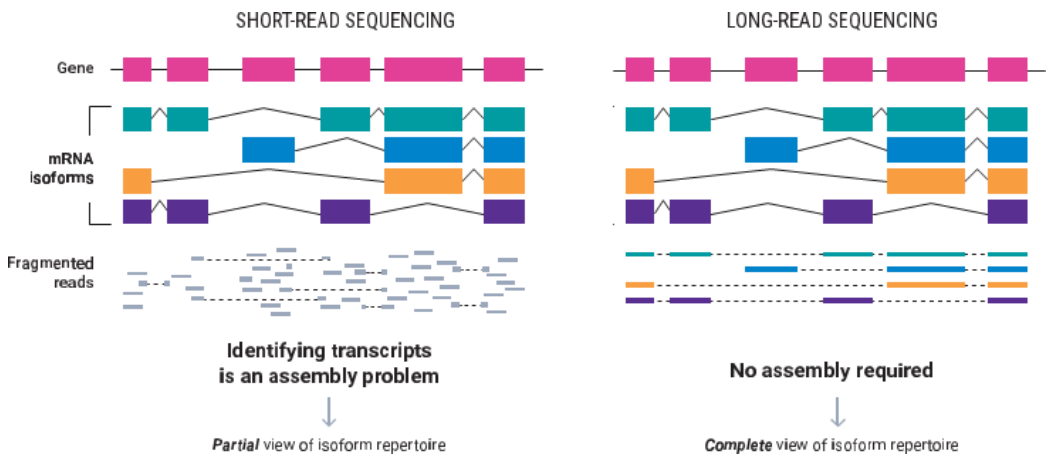


Figure 1.2. Comparison of short-read and long-read sequencing for transcriptome analysis. Short-read sequencing (left) generates fragmented reads that must be computationally assembled to infer mRNA isoforms, often resulting in a partial or incomplete view. In contrast, long-read sequencing (right) captures full-length transcripts in single reads, eliminating the need for assembly and providing a direct, complete view of the isoform repertoire. The figure was reproduced from Pacific Biosciences [93].

The ability to sequence full-length transcripts provides a much clearer picture of AS patterns, allowing for the characterization of complex events involving multiple exons [77]. Similarly, it enables precise mapping of polyadenylation sites, facilitating the study of 3' UTR diversity generated by APA and the identification of intronic polyadenylation (IPA) events [94]. Studies indicate better capture of true transcript ends compared to some short-read methods or even other long-read platforms under certain conditions [72].

With long reads, the detection of fusion transcripts and read-through transcripts is also relatively easier. Long reads spanning the breakpoint of fusion transcripts provide direct evidence for these events and allow for the characterization of specific fusion isoforms, which can be drivers and can be used as biomarkers in many cancers [95].

High-accuracy long reads (like PacBio HiFi or improving ONT reads) can also phase polymorphisms across a transcript, enabling the study of allele-specific isoform expression [96].

A unique capability of ONT is direct RNA sequencing (DRS), which analyzes native RNA molecules as they pass through a nanopore [97] (Figure 1.3.A). This approach avoids potential biases introduced during reverse transcription (RT) and PCR amplification steps inherent in cDNA-based methods (including standard short-read and PacBio/ONT cDNA sequencing) [98]. Crucially, DRS preserves endogenous RNA modifications (e.g., m6A, pseudouridine), allowing for the simultaneous profiling of the transcriptome and epitranscriptome from the same dataset [99]. This simultaneous detection of sequence, isoform structure, and RNA modifications in a single experiment represents an extensive view of RNA biology.

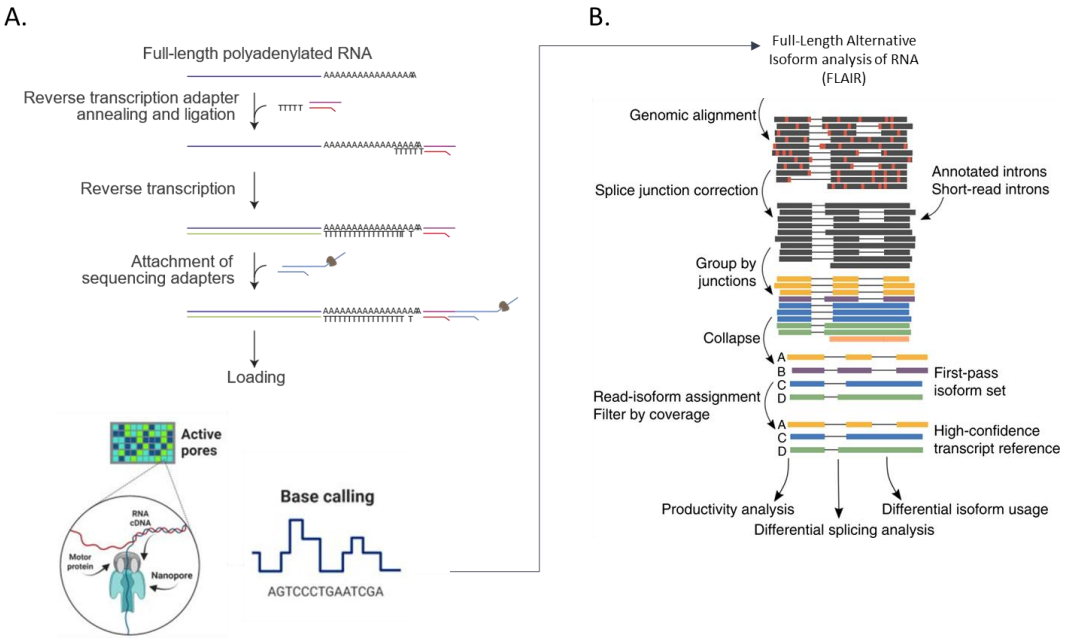


Figure 1.3. Nanopore direct RNA sequencing from library preparation to data analysis. (A) Nanopore direct RNA Sequencing workflow. Full-length RNA is converted to an RNA:cDNA hybrid and sequencing adapters are attached. The library is run on a nanopore sequencer, which reads the electrical signals generated by the molecule passing through the pore and converts them into sequence data. (B) FLAIR Analysis Workflow. Raw reads are aligned to a genome to identify exon-intron structures. Reads are then grouped and collapsed to define a set of transcript isoforms. Low-coverage isoforms are filtered out to create a high-confidence reference, which enables downstream functional studies like differential isoform usage and splicing analysis. The figure was reproduced from [77,100].

1.3.2.1 Transcript Isoform Discovery

To translate the raw long-read data into biologically meaningful insights, specialized computational workflows are required. One such tool is FLAIR (Full-Length Alternative Isoform analysis of RNA), a pipeline designed specifically for the analysis of long-read sequencing data [77]. FLAIR's workflow first aligns long reads to a reference genome and then performs splice site correction to account for mapping inaccuracies. Subsequently, it collapses the corrected reads into a set of high-confidence, full-length transcript isoforms, even enabling the discovery of novel splice variants not present in existing annotations (Figure 1.3.B). It also performs quantitative analysis by identifying differential isoform expression and differential transcript usage between conditions, providing a framework to characterize complex transcription events.

While initially hampered by lower throughput and higher error rates compared to short-read platforms, continuous improvements in long-read technologies are rapidly increasing throughput, accuracy and cost-effectiveness [98,101].

The application of long-read sequencing to cancer transcriptomics is yielding significant new insights into tumor biology [102]. By revealing the full spectrum of transcript isoforms, including tumor-specific variants previously hidden from view, these technologies provide a more complete understanding of the molecular alterations driving cancer and potential new targets for diagnosis and therapy [103]. This technological advancement enables a shift from gene-centric analyses to an isoform-resolved view of the transcriptome, which is essential for fully understanding the impact of signaling pathways like that of estrogen [104].

1.3.3 Transcript Isoform Diversity

Estrogen signaling exerts its effects on the transcriptome at multiple levels beyond simply altering the overall abundance of target genes [105]. Accumulating evidence indicates that E2 can regulate post-transcriptional RNA processing events, thereby

dynamically shaping the landscape of transcript isoform diversity in response to hormonal cues [106–108]. This regulation adds a critical layer of complexity to the estrogen response, allowing a single gene locus to produce multiple functionally distinct RNA and protein variants.

AS enables the differential inclusion or exclusion of exons or parts of exons from the final mature mRNA. Common types of AS events include exon skipping (cassette exons), the use of alternative 5' or 3' splice sites, mutually exclusive exons, and intron retention [109]. Estrogen signaling can influence these choices, leading to the production of different protein isoforms from the same gene [110]. These protein variants may possess altered functional domains, enzymatic activity, protein-protein interaction capabilities, subcellular localization, or stability, thereby contributing to the diverse cellular outcomes observed upon E2 stimulation [110,111]. For instance, AS of the BCL2 Like 1 (*BCL−X*) gene generates the anti-apoptotic Bcl-xL and the pro-apoptotic Bcl-xS isoforms; shifts in the ratio of these isoforms, potentially influenced by signaling pathways intertwined with estrogen, can dramatically impact cell survival [112]. Similarly, AS of cell surface receptors like CD44 or FGFR2 can generate variants with different ligand binding properties or signaling outputs, impacting processes like cell adhesion, migration, and proliferation [113,114].

APA affects the 3' end processing of pre-mRNAs, primarily by allowing the selection of different polyadenylation signals (PAS) within the transcript [115]. The most common form of APA occurs within the 3' untranslated region (3' UTR), leading to mRNA isoforms with the same coding sequence but different 3' UTR lengths (UTR-APA) [116]. Shorter 3' UTRs may arise from the use of proximal PAS, while longer 3' UTRs result from using distal PAS [115]. Since 3' UTRs harbor binding sites for microRNAs (miRNAs) and RNA-binding proteins (RBPs) that regulate mRNA stability, translation efficiency, and localization. APA-mediated changes in 3' UTR length can impact post-transcriptional gene regulation [117]. A general trend observed in proliferating cells and cancer is a global shift towards shorter 3' UTRs, which can allow mRNAs (including those encoding proto-oncogenes) to escape miRNA-mediated repression, leading to increased protein production [118]. Our lab

has identified a short 3'UTR isoform generated by APA upon E2 stimulation. The 3' UTR shortening of the DNA replication licensing factor *CDC6* in ER+ breast cancer cells, was linked to increased CDC6 protein , contributing to increased DNA replication [119].

Another form of APA, intronic polyadenylation (IPA), occurs when a PAS within an intron is recognized, leading to premature cleavage and polyadenylation [120]. IPA may result in truncated mRNA transcripts that may encode truncated proteins with altered or dominant-negative functions, or may lead to non-functional transcripts targeted for degradation [120,121]. The potential regulation of IPA by estrogen signaling in breast cancer remains an area for investigation.

Additionally, estrogen signaling can indirectly modulate RNA processing by altering the expression levels or activity (e.g., via phosphorylation) of key splicing factors (SFs) and RBPs [122]. Families of SFs like serine/arginine-rich (SR) proteins (e.g., SRSF6/SRp55, SRSF7) and heterogeneous nuclear ribonucleoproteins (hnRNPs) are known to be involved in splice site selection [122]. Several studies have reported estrogen-dependent regulation of specific SFs [123]. For example, E2 has been shown to downregulate SRp55 (SRSF6) in breast cancer cells, which in turn alters the splicing of the *CRHR1* gene [123]**.** Similarly, E2 can induce SRSF7 expression, potentially mediating downstream splicing changes [124]**.** Changes in the abundance or activity of core polyadenylation factors (e.g., components of CPSF, CstF, CFIm) can also shift the balance of PAS usage [125].

Therefore, estrogen utilizes both transcriptional and post-transcriptional mechanisms to orchestrate its biological effects [14]. Understanding how E2 impacts transcript isoform diversity through these mechanisms is essential for a complete picture of its role in breast cancer development, progression, and response to therapy [52].

1.4 Long Noncoding RNAs (lncRNAs) and Their Roles in Breast Cancer and Estrogen Signaling

The vast landscape of noncoding RNAs (ncRNAs) is increasingly recognized as a crucial layer of cellular regulation [126]. Among these, long noncoding RNAs (lncRNAs) have emerged as versatile and influential players in numerous biological processes, including cancer pathogenesis and hormone signaling pathways [127–129].

1.4.1 Definition and Biogenesis of lncRNAs

Long noncoding RNAs (lncRNAs) are broadly defined as RNA transcripts exceeding 200 nucleotides in length that lack significant protein-coding potential [130]. This size threshold distinguishes them from small regulatory ncRNAs like microRNAs (miRNAs) and small interfering RNAs (siRNAs), while their lack of functional open reading frames (ORFs) differentiates them from messenger RNAs (mRNAs) [131]. The human genome encodes tens of thousands of putative lncRNA genes, potentially outnumbering protein-coding genes, although only a fraction have been functionally characterized [130].

LncRNAs exhibit considerable diversity in their genomic origins, structure, and processing [132]. They can be transcribed from intergenic regions, intronic regions, or overlap with protein-coding genes in either the sense or antisense orientation [130]. Similar to mRNAs, the majority of lncRNAs are transcribed by RNA Polymerase II and undergo co-transcriptional and post-transcriptional modifications, including 5' capping, splicing, and 3' polyadenylation, although exceptions exist (e.g., some lncRNAs are non-polyadenylated or processed differently) [133–135]. Many lncRNAs undergo AS, further contributing to their diversity [136]. Compared to mRNAs, lncRNAs are generally expressed at lower levels and often exhibit more tissue-specific or cell-type-specific expression patterns, suggesting tightly regulated roles in development and cellular identity [137,138]. Their sequences are typically

less conserved across species than protein-coding genes, although conserved structural elements or short sequence motifs may be critical for their function [139].

1.4.2 Diverse Mechanisms of lncRNA Function

LncRNAs exert their biological functions through a remarkable variety of mechanisms, primarily by interacting with other macromolecules, including DNA, RNA (both mRNA and miRNA), and proteins (such as transcription factors, chromatin modifiers, and RNA-binding proteins) [140] (Figure 1.4). Their functional repertoire is often closely linked to their subcellular localization [141].

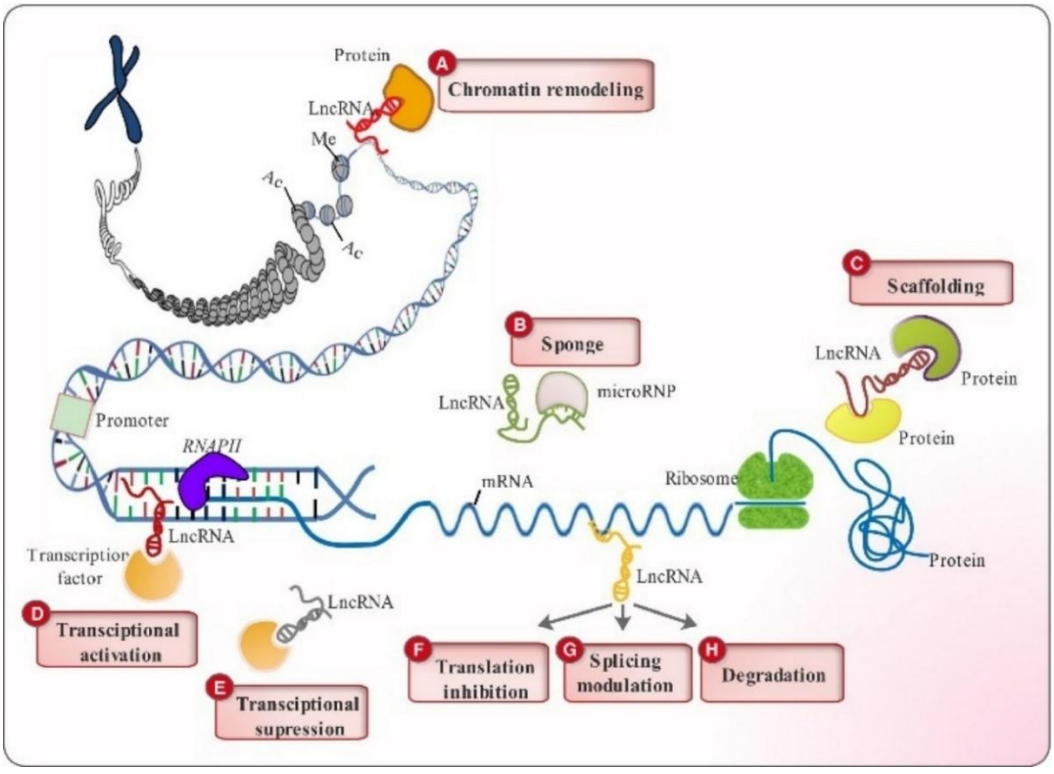


Figure 1.4. Diverse Molecular Functions of lncRNAs. This figure illustrates the various mechanisms by which lncRNAs exert their regulatory effects. (A) LncRNAs can facilitate changes in chromatin organization by recruiting chromatin-modifying

complexes to specific genomic locations. (B) They can function as competing endogenous RNAs (ceRNAs) or 'miRNA sponges,' binding to complementary microRNAs (miRNAs) and thereby preventing them from regulating their target mRNAs. (C) By acting as molecular scaffolds, lncRNAs provide a platform for the assembly of multiple proteins into functional complexes. (D) They can promote gene expression by guiding transcription factors to the promoter regions of target genes. (E) Conversely, lncRNAs can suppress transcription by sequestering transcription factors, preventing their binding to promoters. LncRNAs also modulate post-transcriptional events by directly binding to mRNA molecules, which can lead to (F) inhibition of protein translation, (G) modulation of AS patterns, or (H) induction of mRNA degradation. The figure was reproduced from [142].

1.4.2.1 Nuclear lncRNAs

Nuclear lncRNAs predominantly participate in the regulation of gene expression at the transcriptional and epigenetic levels:

They can act as guides, recruiting chromatin-modifying complexes (e.g., Polycomb Repressive Complex 2 (PRC2) or histone demethylases) to specific genomic loci to induce epigenetic silencing or activation For instance, *HOTAIR*-guided PRC2 activity can silence tumor suppressor genes like *HOXD10* and *PCDH10* in breast cancer [143], and *WIF-1* in esophageal squamous cell carcinoma by inducing H3K27 trimethylation at their promoter regions [144]. lncRNAs can function as scaffolds, bringing together multiple proteins to form functional regulatory complexes on chromatin. *NEAT1* functions as a structural scaffold by interacting with a multitude of over 50 paraspeckle proteins (PSPs), including SFPQ, NONO, FUS, RBM14, and HNRNPK, organizing them into coherent and functional paraspeckle bodies [145]. *NEAT1*-structured paraspeckles then contribute to gene expression regulation, often by sequestering specific RNAs and proteins, thereby modulating their availability or activity [146].

Some lncRNAs act as decoys, preventing transcription factors or other regulatory proteins from their intended targets. For example, *GAS5* (Growth Arrest-Specific 5) can function as a molecular decoy for the glucocorticoid receptor (GR) by binding to the DNA-binding domain of the GR, effectively mimicking a glucocorticoid response element (GRE). Consequently, *GAS5* suppresses the upregulation of GR-responsive genes, acting as a buffer to modulate the intensity of glucocorticoid signaling [147]. LncRNAs can also directly interact with DNA, forming RNA-DNA structures to modulate promoter activity, or participate in the organization of nuclear domains and chromatin architecture. *TERRA* (Telomeric Repeat-containing RNA) lncRNAs are transcribed from the subtelomeric regions of chromosomes and can localize back to telomeres and form stable RNA:DNA structures known as R-loops. These loops play active and crucial roles in telomere maintenance, integrity, and the regulation of telomerase activity [148].

Nuclear lncRNAs can also influence the processing of other RNAs, including pre-mRNA splicing, by interacting with splicing factors or modulating spliceosome assembly. *MALAT1*, also known as *NEAT2*, is a nuclear long non-coding RNA that regulates AS by interacting with and modulating the activity of SR splicing factors. This interaction influences the phosphorylation and distribution of SR proteins within nuclear speckles, thereby affecting the splicing patterns of various pre-mRNAs and impacting cellular processes like apoptosis [149].

1.4.2.2 Cytoplasmic lncRNAs

Cytoplasmic lncRNAs are primarily involved in post-transcriptional regulation. A widely studied mechanism involves lncRNAs acting as competing endogenous RNAs (ceRNAs) or "miRNA sponges" [150]. By harboring binding sites for specific miRNAs, these lncRNAs can sequester miRNAs, thereby preventing them from binding to their target mRNAs and leading to the de-repression (upregulation) of those mRNA targets. The lncRNA *H19* is a well-documented miRNA sponge, particularly for the let-7 family of miRNAs. *H19* possesses both canonical and non-

canonical binding sites for let-7 microRNAs. By binding and sequestering let-7, *H19* effectively reduces the available pool of these miRNAs, leading to the de-repression of their mRNA targets, such as *DICER* and *HMGA2* [151].

LncRNAs can interact with mRNAs directly through base-pairing or indirectly via RNA-binding proteins (RBPs) to influence mRNA stability (promoting decay or stabilization) or translation efficiency (enhancing or inhibiting protein synthesis) [152,153].

For example, *BACE1-AS* (BACE1 antisense RNA) stabilizes BACE1 (beta-secretase 1) mRNA by forming a direct duplex, which masks binding sites for microRNAs [154]. *LincRNA-p21* (also known as CDKN1A antisense RNA 1) binds to the mRNAs of *CTNNB1* (Catenin beta 1) and *JUNB* (JunB proto-oncogene, AP-1 transcription factor subunit), along with translational repressors, to inhibit their translation [155]. Also, *GAS5* (Growth arrest-specific 5) selectively blocks the translation of *MYC* (MYC proto-oncogene) by directly binding to the initiation factor eIF4E (Eukaryotic translation initiation factor 4E) [156].

Cytoplasmic lncRNAs can act as scaffolds for signaling complexes or regulate protein activity or localization [157]. In the cytoplasm, *HOTAIR* (HOX transcript antisense RNA) serves as a scaffold for DZIP3 (DAZ interacting zinc finger protein 3) and MEX3B (mex-3 RNA binding family member B), which facilitates the ubiquitination and subsequent degradation of SNUPN (snurportin 1) and ATXN1 (ataxin 1), respectively [158]. Another example, *LINK-A* (long intergenic non-protein coding RNA, kinase activator), scaffolds BRK (BRK proto-oncogene, Src family tyrosine kinase) and LRRK2 (leucine rich repeat kinase 2) to the EGFR (epidermal growth factor receptor):GPNMB (glycoprotein nmb) complex. This leads to the phosphorylation of HIF-1α and the upregulation of its target genes [159].

This functional versatility, enabled by their ability to fold into complex structures and interact with diverse partners, positions lncRNAs as important regulators of cellular processes [130].

1.4.3 LncRNAs in Breast Cancer Pathogenesis

Given their extensive regulatory roles, it is unsurprising that dysregulation of lncRNA expression and function is increasingly implicated in the pathogenesis of numerous diseases, including cancers [127,160]. In breast cancer, a growing body of evidence highlights the critical involvement of lncRNAs in virtually all aspects of tumor biology, including uncontrolled proliferation, evasion of apoptosis, angiogenesis, invasion and metastasis, metabolic reprogramming, and resistance to therapy [161,162]. LncRNAs can function as either oncogenes, promoting tumor growth and progression, or as tumor suppressors, inhibiting these processes [161,163].

For example, the lncRNA *BANCR* acts as an oncogene in breast cancer, where its high expression promotes tumor growth, metastasis, and resistance to therapies like trastuzumab. It exerts these effects primarily by acting as a molecular sponge for miR-34a-5p to activate the Wnt/β-catenin pathway, and it can also be transferred between cells in vesicles to spread drug resistance [164].

On the contrary, the tumor suppressor lncRNA *MEG3* is downregulated in breast cancer, and its presence is linked to better patient survival by inhibiting cancer cell growth and promoting cell death. It functions by sponging miR-330 to regulate the CNN1 axis. [165]. *MEG3* also counteracts estrogen-driven proliferation through a negative correlation with the *ESR1* [158].

Numerous specific lncRNAs have been identified as having altered expression and functional significance in breast cancer (Table 1.1). Among the most well-studied examples are:

HOTAIR (HOX Transcript Antisense RNA) is often upregulated in breast cancer, particularly metastatic disease, *HOTAIR* typically functions as an oncogene [166]. It acts primarily in the nucleus, recruiting chromatin-modifying complexes like PRC2 to silence tumor suppressor genes, thereby promoting proliferation, invasion, and

metastasis [167]. Its expression has been linked to poor prognosis and endocrine resistance [168,169].

MALAT1 (Metastasis Associated Lung Adenocarcinoma Transcript 1) is also frequently upregulated in breast cancer, *MALAT1* is associated with enhanced proliferation, migration, invasion, and metastasis [170,171]. It functions through various mechanisms, including regulating AS by modulating SR protein activity and acting as a miRNA sponge in the cytoplasm [172].

H19, an imprinted lncRNA with context-dependent roles, can act as either an oncogene or a tumor suppressor in breast cancer [173–175]. It has been implicated in regulating proliferation, apoptosis, EMT, and angiogenesis, often through mechanisms involving miRNA sponging [176,177].

Table 1.1. Overview of Key lncRNAs in General Breast Cancer Pathogenesis.

lncRNA	**Role in BC**	**Mechanism**	**Functional Impacts**	**Ref**
PVT1	Oncogene	Interacts with PRC2-ERα network; sponges miR-128-3p	Promotes proliferation, migration,; inhibits apoptosis; downregulates tumor suppressors (e.g., *BTG2*)	[178]
NEAT1	Oncogene	Paraspeckle component; repressed by BRCA1; upregulated by hypoxia; potential miRNA sponging	Promotes proliferation; inhibits apoptosis; regulates cell cycle (cyclin E1/D1)	[179, 180]

Table 1.1 (cont'd)

GAS5	Tumor Suppressor	Induces apoptosis; glucocorticoid receptor mimic	Inhibits tumor growth; low expression linked to poor prognosis	[181]
MEG3	Tumor Suppressor	Sponges miR-330 to regulate *CNN1*; negatively correlated with *ESR1*	Inhibits proliferation, migration, invasion; promotes apoptosis	[158, 165]
XIST	Tumor Suppressor-like	Recruits chromatin-modifying complexes for X-inactivation; may inhibit AKT signaling	Negatively regulates cell viability	[182]
BANCR	Oncogene	Sponges miR-34a-5p to upregulate *HMGA1*; activates Wnt/β-catenin	Promotes proliferation,, confers trastuzumab resistance	[164]
HOTAIR	Oncogene	Recruits PRC2 to silence tumor suppressor genes	Promotes proliferation, invasion, metastasis; linked to poor prognosis and endocrine resistance	[143, 167]

Table 1.1 (cont'd)

MALAT1	Oncogene	Regulates AS (SR proteins); miRNA sponge	Promotes proliferation, migration, invasion, metastasis	[149]
H19	Context-dependent	miRNA sponging; epigenetic regulation	Regulates proliferation, apoptosis, EMT, angiogenesis; involved in chemoresistance and endocrine resistance	[173]

These examples illustrate the diverse ways in which lncRNAs contribute to breast cancer pathogenesis, acting through transcriptional, post-transcriptional, and epigenetic mechanisms to influence key cancer hallmarks [127,183].

1.4.4 LncRNAs as Targets and Modulators of Estrogen Signaling

Given the centrality of estrogen signaling in ER+ breast cancer, the interplay between E2/ERα and lncRNAs is rapidly advancing area of research [184]. This relationship is bidirectional: estrogen signaling regulates the expression of numerous lncRNAs, and conversely, lncRNAs can modulate the activity and downstream effects of the estrogen signaling pathway [183].

Many lncRNAs have been identified as direct transcriptional targets of ERα [185]. These E2-regulated lncRNAs often possess functional EREs or other ERα binding sites in their promoter or enhancer regions, and their expression levels change significantly upon E2 stimulation or anti-estrogen treatment [185]. For example, the lncRNA *ERLC1* was identified through cross-analysis of ERα binding data and

expression datasets as being transcriptionally activated by ERα in breast cancer cells [183]. The lncRNA *ERINA* (Estrogen Inducible lncRNA) also responds directly to estrogen. Its expression is induced by E2 in ER+ breast cancer cells (T47D and MCF7) and is diminished by treatment with the ER antagonist Fulvestrant or by siRNA-mediated knockdown of ERα. A functional intronic ERE, acting as an enhancer, mediates the transactivation of *ERINA* by ERα, with ERα recruitment to this site observed upon E2 treatment [186].

In contrast, *HOTAIR* expression is repressed by estrogen. Its promoter and distant enhancer contain functional EREs, and and E2 stimulation enhances ERα binding to these sites, leading to transcriptional repression of *HOTAIR*. Consequently, tamoxifen treatment or hormone deprivation results in increased *HOTAIR* levels [187]. Finally, *DSCAM-AS1* expression is controlled by ERα, notably even in hormone-deprived conditions, indicating a ligand-independent mode of regulation, and it is highly specific for luminal (ER+) breast cancers [188].

The location of functional EREs is also not limited to proximal promoters; intronic enhancers (*ERINA*) or distant enhancers (*HOTAIR*) highlight the complex genomic architecture underlying ERα's regulation of lncRNA expression. Such complexity underscores the need to identify both proximal and distal regulatory elements for a complete map of ERα-lncRNA interactions.

These E2-regulated lncRNAs may then act downstream to mediate specific aspects of the estrogen response [184,185]. Table 1.2 summarizes key lncRNAs whose expression is modulated by the estrogen signaling pathway.

Table 1.2. LncRNAs Regulated by Estrogen Signaling.

LncRNA	**Cancer**	**Relationship with ER**	**Mechanism**	**Ref**
DSCAM-AS1	Breast Cancer	ERalpha interacts with promoter; Also estrogen-responsive	Highly specific for ERalpha+ luminal cells; Deletion mimics ERalpha deletion effects	[188–190]
H19	Breast Cancer, Endometrial context (ecESCs)	Upregulated by E2/ERalpha in ERalpha+ breast cancer cells; E2 promotes expression in ecESCs	Oncogenic in breast cancer; Knockdown decreases cell survival and blocks estrogen-induced cell growth	[151,175, 177,191]
HOTAIR	Breast Cancer, Endometrial Cancer	Expression regulated by E2/ERalpha (BC); E2 increases levels (EC)	Promoter contains EREs; ERalpha recruits co-regulatorsTranscriptionally suppressed by ER.	[166,169, 187]
ERINA	Breast Cancer	Estrogen-responsive oncogenic factor	ER-binding element identified within its intronic site; H3K27Ac enrichment suggests enhancer function	[131,186]

Table 1.2 (cont'd)

MIAT	Breast Cancer	Regulated via estrogenic signaling	E2 upregulates *MIAT* in MCF-7 cells via an ERalpha-dependent mechanism	[192]
LINC00472	Breast Cancer	Estrogen-responsive; Upregulated by ERalpha	ERalpha-binding site within its promoter region predicted and confirmed; Over-expression clinicopathological features	[193]
BNAT1	Breast Cancer	Estrogen-inducible antisense RNA	Expression enhanced by E2 in MCF7 cells	[194]
AER-lncRNAs (class)	Breast Cancer	Regulated by Apo-ERalpha	Largely restricted to luminal cells; Maintain luminal epithelial phenotype	[188]
RP11-48B3.4	Breast Cancer	Estrogen Receptor Agitation-Related (ERAR) lncRNA	Low expression associated with favorable prognosis in ER+ breast cancer patients	[195]

Table 1.2 (cont'd)

LOC100288637	Breast Cancer	Estrogen Receptor Agitation-Related (ERAR) lncRNA; correlation with HER2	Low expression associated with favorable prognosis in ER+ breast cancer patients	[195]
TC0101441	Ovarian Cancer	Upregulated by estrogen	Correlated with ERalpha+ ovarian cancer tissues; OE associated with advanced FIGO stage and/or high histological grade	[196]
TC0101686	Ovarian Cancer	Downregulated by estrogen	Correlated with ERalpha+ ovarian cancer tissues; Low expression associated with high histological grade	[196]
NEAT1	Prostate Cancer, Breast Cancer	ERalpha-regulated (PCa); Upregulated by E2 in ERalpha+ breast cancer cells (MCF7)	ERalpha recruited to *NEAT1* promoter; Most significantly overexpressed ERalpha-regulated lncRNA in prostate cancer	[197]

Conversely, lncRNAs can act as crucial modulators of the estrogen signaling pathway itself, creating regulatory feedback loops [183]. Mechanisms include regulation of *ESR1* expression, modulation of Erα activity and mediating endocrine resistance. A list of lncRNAs known to be involved in these interactions is provided in Appendix Table A.1.

Some lncRNAs can influence the expression of ERα by affecting the transcription or stability of the *ESR1* mRNA [183]. For instance, *ERLC1* was shown to stabilize the *ESR1* transcript, potentially by sequestering miRNAs that target *ESR1* or by interacting with RBPs like FXR1, thereby creating a positive feedback loop that enhances ERα signaling [183].

Other lncRNAs might repress *ESR1* expression. The lncRNA *LINP1* represses the estrogen response by downregulating ERα protein levels through a negative feedback loop in which reduced ER activity increases *LINP1*, leading to further ERα suppression [198] .

LncRNAs can interact directly with the ERα protein, acting as co-activators or co-repressors to modulate its transcriptional activity at target gene promoters. For instance, *BNAT1* lncRNA physically associates with the ERα protein. Knockdown of *BNAT1* leads to decreased ERα expression and represses ERα transactivation, suggesting it functions to modulate ERα-dependent transcriptional regulation, likely as a co-activator [194].

Other lncRNAs, such as *FGD5-AS1* and *EPB41L4A-AS1*, also interact with ERα in MCF-7 cell nucleus. While silencing of *FGD5-AS1* reduces cell proliferation and ERα protein levels, the tumor suppressor lncRNA *EPB41L4A-AS1,* which is often downregulated in tumors, shows a different modulatory effect [199]. Silencing *EPB41L4A-AS1* increases ERα trans-activating efficiency, suggesting it might normally act as a co-repressor or a modulator that moderates ERα activity [199].

A growing number of lncRNAs, such as *BNAT1*, *PVT1*, *FGD5-AS1*, *EPB41L4A-AS1*, and *MALAT1* and *HOTAIR*, have been found to physically bind to the ERα

protein [187,194,199,200]. These interactions can modulate ERα's conformation, its ability to bind DNA, or its interaction with co-regulators.

Several lncRNAs, such as *HOTAIR*, *H19*, *LINP1*, *DILA1*, *LINC-ROR*, *DSCAM-AS1*, *UCA1*, *MAFG-AS1*, *ERLC1*, and *BNAT1*, have been implicated in the development of resistance to endocrine therapies like tamoxifen, often by promoting ligand-independent ERα signaling or activating bypass pathways [187,190,194,201–204].

Identifying and functionally characterizing these estrogen-regulated and estrogen-modulating lncRNAs is crucial for a comprehensive understanding of the estrogen signaling network [205]. These lncRNAs represent potential biomarkers for predicting response to endocrine therapy and novel therapeutic targets for overcoming resistance in ER+ breast cancer [68]. The intricate feedback loops involving lncRNAs suggest that targeting these noncoding regulators might offer alternative strategies to disrupt aberrant estrogen signaling in cancer [183].

1.4.5 Small Nucleolar RNA Host Gene (SNHG) Family of lncRNAs

Among the diverse classes of lncRNAs implicated in cancer, the small nucleolar RNA host gene (SNHG) family has gathered increasing attention [206]. This family comprises a group of lncRNAs, currently numbered *SNHG1* through *SNHG22* and beyond, characterized by their genomic structure: they typically host genes encoding small nucleolar RNAs (snoRNAs) within their intronic regions [207]. SnoRNAs are well-known noncoding RNAs primarily involved in the post-transcriptional modification (e.g., methylation, pseudouridylation) and processing of ribosomal RNAs (rRNAs) and sometimes other RNAs like spliceosomal RNAs, playing essential roles in ribosome biogenesis [208].

While the hosting of snoRNAs is a defining feature, it is now clear that many SNHG lncRNAs possess biological functions independent of the snoRNAs they harbor [209]. The SNHG transcripts themselves, after splicing, can act as functional lncRNAs, participating in various cellular processes, particularly in the context of

cancer [210]. Numerous studies have reported the dysregulation (mostly upregulation) of various SNHG family members in a wide range of human malignancies, including endocrine-related cancers such as thyroid, pancreatic, ovarian, prostate, and notably, breast cancer [206].

Specific SNHG lncRNAs have been implicated in promoting key aspects of cancer progression, including enhanced cell proliferation, survival, migration, invasion, angiogenesis, stemness, and resistance to chemotherapy or radiotherapy [211,211] (Figure 1.5). Examples relevant to breast cancer include;

SNHG1, *SNHG6*, *SNHG15*, *SNHG16*, and *DANCR* (*SNHG13*) are generally reported to act as oncogenes. Their upregulation is often associated with poor prognosis. Their primary functions include promoting cell proliferation, migration, invasion, and inhibiting apoptosis [210,212–220]. A common mechanism for these lncRNAs is acting as a microRNA (miRNA) sponge, which prevents miRNAs from regulating their target genes. *SNHG1* is also suggested to be regulated by estrogen signaling and may modulate histone methylation [221]. *SNHG5* has a context-dependent role, sometimes acting as an oncogene and sometimes as a tumor suppressor, while *SNHG17* is linked to proliferation and metastasis in breast cancer [210,214]. In contrast, *GAS5* (SNHG2) typically functions as a tumor suppressor and is often downregulated in breast cancer. Its loss promotes proliferation and inhibits apoptosis [222]. *GAS5* has diverse mechanisms, including interacting with miRNAs (like miR-21 and miR-222), proteins, and DNA, thereby influencing major signaling pathways such as PI3K/Akt [223].

The mechanisms underlying SNHG functions are diverse but frequently involve acting as miRNA sponges (ceRNAs), thereby shielding the target mRNAs of those miRNAs [209]. Other reported mechanisms include interacting with RNA-binding proteins (RBPs) to affect mRNA stability or translation, interacting directly with proteins (e.g., transcription factors, signaling molecules) to modulate their activity, and participating in epigenetic regulation by guiding chromatin modifiers or influencing DNA methylation [209] (Appendix Table A.2).

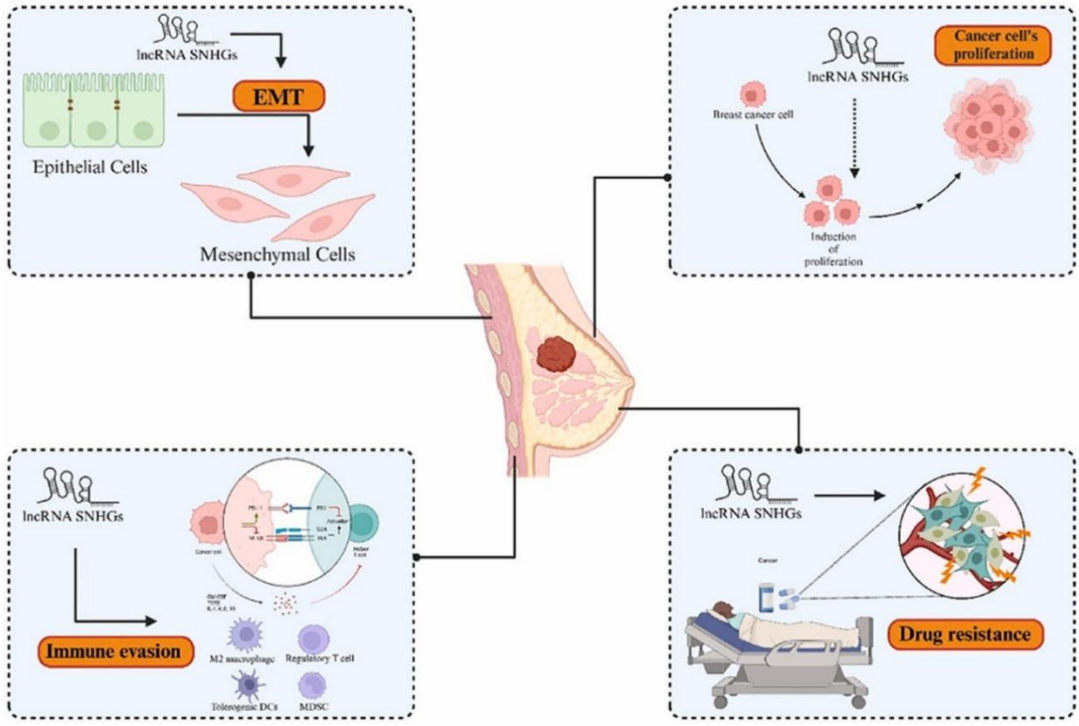


Figure 1.5. Key Functions of lncRNA SNHGs in Breast Cancer Progression. SNHG family contribute to breast cancer malignancy through several key pathways. Epithelial-Mesenchymal Transition (EMT): Promoting cancer cell invasion and metastasis by inducing the shift from an epithelial to a mesenchymal phenotype. Cancer Cell Proliferation: Directly stimulating the cell cycle to accelerate tumor growth. Immune Evasion: Fostering an immunosuppressive tumor microenvironment to help cancer cells avoid detection and destruction by the immune system. Drug Resistance: Enabling cancer cells to survive chemotherapeutic treatments, leading to therapeutic failure. The figure was reproduced from [209].

The frequent dysregulation of multiple SNHG members across various cancers, including breast cancer, raises the possibility of their coordinated regulation by central signaling pathways, such as the estrogen signaling pathway in ER+ breast cancer [224,225]. Understanding how estrogen might regulate the expression and

function of this lncRNA family could provide significant insights into estrogen-driven tumorigenesis [226].

The SNHG family represents a functionally important and potentially interconnected group of lncRNAs whose roles in estrogen signaling and ER+ breast cancer needs deeper investigation. Their diverse mechanisms, particularly as miRNA sponges, suggest they could collectively exert significant control over post-transcriptional gene regulation networks relevant to cancer.

1.5 Rationale and aim of the Study

Despite extensive research, our understanding of estrogen signaling in breast cancer may be incomplete due to the transcriptome's inherent complexity and the limitations of traditional research methodologies. Several critical knowledge gaps persist.

Primarily, most studies on E2's transcriptomic effects in ER+ breast cancer have utilized short-read sequencing ([55,227,228]**,** Section 1.3.1). These methods struggle to resolve the full diversity of transcript isoforms arising from AS and APA, thereby hindering a comprehensive, time-resolved view of how E2 dynamically modulates the isoform landscape for both coding and noncoding RNAs [51,226]. Consequently, the full extent and functional significance of E2-induced AS and APA events are likely underestimated.

The relationship between lncRNAs and estrogen signaling also needs further investigation. While numerous lncRNAs are known to be dysregulated in breast cancer and some are established E2 targets or modulators, many E2-regulated lncRNAs and their specific isoforms likely remain unidentified [68]. This is especially true for lncRNAs with complex splicing patterns or those expressed at lower levels, which are challenging to detect and characterize accurately using short reads [77,81]. The functional roles of most E2-associated lncRNAs in the context of ER+ breast cancer biology is poorly understood [185,226].

Despite its power for full-length transcript characterization (Section 1.3.2), long-read direct RNA sequencing has not been used to study the temporal dynamics of the E2 response in breast cancer.

Therefore, a time-resolved understanding of how E2 stimulation impacts isoform usage, AS, and APA is currently lacking. These gaps in knowledge highlight a blind spot. Focusing predominantly on gene-level expression changes measured by short reads neglects the potentially crucial regulatory information encoded at the isoform level and within the noncoding transcriptome. This missing information may be key to fully understanding E2's diverse biological effects, the heterogeneity of ER+ breast cancer, and the mechanisms underlying endocrine resistance.

Therefore, the aim of this thesis is to characterize the dynamic impact of E2 on the transcriptome of the ER+ breast cancer cell line MCF7, leveraging the unique capabilities of Oxford Nanopore Technology (ONT) long-read sequencing. The research will place particular emphasis on the changes in transcript isoform diversity, AS, and APA, and on identifying and functionally exploring E2-regulated long noncoding RNAs (lncRNAs), especially within the Small Nucleolar RNA Host Gene (SNHG) family.

We hypothesized that estrogen stimulation triggers a coordinated remodeling of the breast cancer transcriptome that extends beyond simple changes in gene expression for both coding and non-coding genes. We proposed that this reprogramming, which can only be resolved with long-read RNA sequencing, would manifest as widespread and dynamic shifts in AS and APA. We further posited that these isoform-level changes act in concert with differential gene expression to create a functionally distinct transcriptomic landscape that collectively drives the estrogen-induced proliferative state.

We exploited the advantages of long-read sequencing to obtain a high-resolution, dynamic, and isoform-aware view of the estrogen-regulated transcriptome. So that we can directly address the limitations of previous approaches and fill gaps in our understanding of E2 action.

CHAPTER 2

MATERIALS AND METHODS

2.1 Cell Culture and Experimental Treatments

2.1.1 2.1.1 Cell Lines and Culture Conditions

The human breast cancer cell lines MCF7, MDA-MB-231, and T47D were used (ATCC; HTB-22, HTB-26, HTB-133, respectively). All cell lines were maintained in Dulbecco's Modified Eagle's Medium (DMEM; Biological Industries, BI01-052-1A) supplemented with 10% Fetal Bovine Serum (FBS; Biological Industries, BI04-007-1A), 1% Penicillin/Streptomycin (P/S; Biological Industries, BI03-031-1B), 1% Sodium Pyruvate (Na/Pyr; Biological Industries, BI03-042-1B), and 1% L-glutamine (Biological Industries, BI03-020-1B). Cells were cultured in a humidified incubator at 37°C in an atmosphere of 5% CO_2.

2.1.2 Hormone and Drug Treatments

For estrogen (E2) stimulation experiments, MCF7 cells were cultured in phenol red-free DMEM with 10% dextran-coated charcoal-stripped FBS for 72 hours for hormone deprivation (HD). Subsequently, cells were treated with either 100 nM 17β-estradiol (E2; Sigma-Aldrich, E2257) or ethanol as a vehicle control. Samples were collected at 45 minutes, 3 hours, and 12 hours post-treatment. An HD control sample was collected at 72 hours, before treatment. For transcript stability assessment, transcription was inhibited with Actinomycin D (ActD; Tocris Biosciences, 1229/10) at 2 µg/mL for MCF7 and 10 µg/mL for MDA-MB-231 [229]. Total RNA

was isolated at 0, 2, 4, 6, 8, and 12 hours post-treatment. Transcript half-lives were calculated from RT-qPCR data. *MYC* was used as a low-stability control and *RPLP0* as a high-stability control.

2.2 RNA Isolation and Direct RNA Sequencing

2.2.1 RNA Isolation and Quality Control

Total RNA was extracted from cell lines using either TRIzol Reagent (Invitrogen, 15596026) or the High Pure RNA Isolation Kit (Roche, 11828665001). For general RNA analysis, samples were treated with DNase I, RNase-free (1 U/μL) (Thermo Fisher Scientific, EN0521) to remove genomic DNA, and RNA concentration was measured using a MaestroNano Micro-volume spectrophotometer (MaestroGen, MN-913). For the samples used in Nanopore sequencing, genomic DNA was removed with TURBO DNase (Invitrogen, AM2238), and RNA quantity, integrity, and quality were assessed using a Qubit 4 fluorometer (Invitrogen, Q33238) with its RNA Integrity and Quality (IQ) Assay (Invitrogen, Q33221).

2.2.2 Library Preparation and Nanopore Sequencing

Polyadenylated RNA was enriched from total RNA using oligo(dT)-conjugated magnetic beads (Dynabeads™ Oligo(dT)25, Thermo, 61002). The efficacy of rRNA depletion was confirmed using an Agilent 2100 Bioanalyzer system. Libraries were prepared from poly(A)-enriched RNA from each of the four experimental conditions (HD, E2 45 min, 3 hr, 12 hr) using the ONT Direct RNA Sequencing Kit (SQK-RNA002) . The detailed DRS protocol is in Appendix B. Sequencing was performed on a MinION Mk1C instrument using R9.4.1 flow cells (FLO-MIN106D). Each sample was sequenced on a separate flow cell.

2.3 Bioinformatics Analysis

2.3.1 Processing and Quality Assessment of Nanopore Sequencing Data

Raw electrical signal data in FAST5 format was basecalled into FASTQ format using the ONT's Guppy basecaller software [230] with default parameter sets for flowcell FLO-MIN106 and kit SQK-RNA002. Quality control of the basecalled reads was conducted using NanoPlot and further analyzed with the ToulligQC tool to ensure high-quality reads for downstream analysis. ToulligQC generated comprehensive run metrics, including total read counts, the percentage of reads passing the default quality filter, N50 read length, total data yield, and mean PHRED quality scores [231,232]. Nanoplot was subsequently used to generate visualizations of these metrics, including read length distribution plots [233]. The sequencing runs were of high quality, producing 1–2 million reads per sample, with over 91% of reads passing quality filters and N50 values in the range of 1377–1503 bp, confirming the successful capture of long transcripts suitable for full-length isoform analysis.

2.3.2 Transcriptome Reconstruction with FLAIR

The core transcriptomic analysis was conducted using the FLAIR (Full-Length Alternative Isoform analysis of RNA) computational pipeline [77]. The FLAIR workflow proceeded in several stages, beginning with the alignment of basecalled reads. High-quality raw nanopore FASTQ reads were aligned to the human reference genome GRCh37.75 using Minimap2 with the parameters -a -u f -k 14 -x splice. These parameters output alignments in SAM format, use a minimizer k-mer length of 14, find canonical splice sites on the transcript strand, and use preset options for splice-aware alignment, respectively. The resulting SAM alignments were then sorted to produce sorted BAM files using SAMtools with default parameters, and BAM indices were created using BAMtools with default parameters. To enhance the accuracy of exon-intron boundary identification, the aligned reads were then

corrected using a comprehensive set of known splice junctions derived from a reference genome annotation (Ensembl). Finally, the corrected reads were collapsed into a non-redundant, high-confidence set of full-length transcript isoforms. This process groups reads that represent the same transcript variant, effectively defining the complete repertoire of isoforms expressed in the samples, including both previously annotated and potentially novel variants. Aligned reads were visualized using the Integrative Genomics Viewer (IGV) [234].

2.3.3 Quantification and Differential Isoform Analysis

Following the definition of the isoform set, FLAIR was used to quantify the abundance of each isoform across all four samples. Expression levels were normalized and reported in Transcripts Per Million (TPM), a unit that accounts for both sequencing depth and transcript length, enabling direct comparison of transcript abundance across samples and conditions. Differential expression analysis was performed using the FLAIR diffIso module, which compared each E2-treated time point (45 min, 3 hr, 12 hr) against the hormone-deprived (HD) control [77].

A multi-level filtering strategy was applied to enhance the signal-to-noise ratio and focus on biologically meaningful changes. First, for a transcript to be included in the downstream analysis, it was required to have a minimum expression level of 10 TPM in at least one of the E2 treatment time points. For differential expression calling, a gene or transcript was considered significant if it met both a fold-change threshold (>1.5 for upregulation or <0.60 for downregulation) and an expression threshold. This expression threshold was tailored to the specific RNA type, reflecting their known biological properties: a minimum of 10 TPM for protein-coding genes and a lower threshold of 2 TPM for non-coding genes as lncRNAs are often expressed at lower levels than mRNAs [130]. Additionally, for a transcript to be included, its abundance had to be greater than or equal to 10% of the total transcript expression from the same gene.

2.3.4 Alternative Splicing and Polyadenylation Analysis

The FLAIR Diffsplice module was used to identify and quantify AS events, categorized as Exon Skipping (ES), Alternative 3' Acceptor site usage (Alt. Acceptor), Alternative 5' Donor site usage (Alt. Donor), and Intron Retention (IR) [77].

Changes in transcript 3' termini were analyzed to identify APA events, including differential usage of polyadenylation sites in the 3' UTR and intronic polyadenylation (IPA). The RegRNA 3.0 database was used to predict regulatory elements in differential 3'UTR sequences [235]. Identified intronic poly(A) sites were cross-referenced with the PolyA_DB v3 database [236]. The detection, analysis, and visualization of the Early Intronic and Middle Intron isoforms were performed by Didem Naz Dioken. In summary, after defining isoforms with FLAIR, the transcript boundaries were obtained. The transcript end locations of the isoforms were mapped to the NCBI Reference Sequence Database (RefSeq) to determine whether the genomic location was intronic or exonic. Regions not annotated in RefSeq were mapped to regions annotated in the Ensembl database, extracted using the BioMart data mining tool. Transcripts were then categorized into early intron, mid-intron, and exonic transcripts according to their transcript end location.

2.3.5 Functional and Pathway Enrichment Analysis

Multiple tools were used for functional analysis. Gene Set Enrichment Analysis (GSEA) “MSigDB” was performed on the ranked list of all expressed genes using the Hallmark gene sets collection (FDR q-value < 0.05) [237]. For more targeted analysis, g:Profiler was used for Gene Ontology (GO) enrichment analysis on gene lists exhibiting significant Exon Skipping or Intron Retention [238]. To analyze non-coding RNAs, two lncRNA-specific tools were employed: RNAenrich for GO and pathway analysis [239], and TLSEA [240] to identify associations between regulated lncRNAs and cancer types. Additionally, the snoRNA Database (snoDB 2.0) was

used to identify snoRNAs located within the intronic regions of E2-regulated SNHG loci [241].

2.4 Analysis of Public Gene Expression Data

Public gene expression datasets were used for in silico analysis. For bulk RNA-seq, data from TCGA and GTEx (via Xena Toil RNA-Seq Recompute Compendium) [242] were used to compare the expression of target genes in breast tumor versus normal tissue. For single-cell RNA-seq, public data (GSE75688) was analyzed using a custom script to screen for reads corresponding to intronic transcripts of *TLE1* and *GREB1*, with events visualized using the Integrative Genomics Viewer (IGV) [243]. Finally, for survival analysis, the Kaplan-Meier plotter was used to assess the prognostic value of *SNHG9* expression in ER-positive and ER-negative breast cancer patient cohorts [244]. Survival curves were compared using the log-rank test, and Hazard Ratios (HR) were calculated.

2.5 Gene Expression and Isoform Validation by RT-qPCR

RT-qPCR experiments were performed following MIQE guidelines [245]. 1 µg of total RNA was used for cDNA synthesis with the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, K1622), using a mix of oligo(dT) and random hexamer primers. qPCR was performed using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, 1725271) or NEB, LunaScript® RT SuperMix Kit, E3010L on a QIAGEN, Rotor-Gene Q device (9242698). For validation studies, a panel of ZR-75-1, BT474, CAL-85-1, HCC1143, and SKBR3 cDNAs was also used. Reactions were run in triplicate. Relative gene expression was calculated using the comparative Ct ($2^{-\Delta\Delta Ct}$) method, with normalization to the *RPLP0* or *ACTB* housekeeping genes [245]. The primers used for RT-qPCR are listed in Table 2.1.

Mature miR-21 was quantified using Applied Biosystems' TaqMan Universal Master Mix II (4440040), TaqMan MicroRNA Reverse Transcription Kit

(4366596),TaqMan MicroRNA Assay (Applied Biosystems, 4427975 for hsa-miR-21-5p and with *RNU43* as the endogenous control.).

Table 2.1. List of RT-qPCR Primers.

Gene	**Forward Primer (5'-3')**	**Reverse Primer (5'-3')**
ACTB	TGACGTGGACATCCGCAAAG	CTGGAAGGTGGACAGCGAGG
DANCR	GCTGGAATGTCCTGTTGCTG	TGTCACTGCTCTAGCTCCTGT
GAPDH	GGGAGCCAAAAGGGTCATCA	TTTCTAGACGGCAGGTCAGGT
lnc-GATA3–7	AGCTGACTTGTGCAGTGGTTA	CGTTCTCCTCACAGAATACAGGT
MYC	CAGCTGCTTAGACGCTGGATT	GTAGAAATACGGCTGCACCGA
NEAT1	ACTGGTATGTTGCTCTGTATGGT	AGGCAATGTGATAGGGGTCG
pri-miR-21	TCGTGACATCTCCATGGCTG	CCAGACAGAAGGACCAGAGT
RPLP0	GGAGAAACTGCTGCCTCATA	GGAAAAAGGAGGTCTTCTCG
SNHG15	GTTGCCTGACCATTCCTGAG	AGTAAGCTCTTCCACTTTGAGACC
SNHG17	CCGTGAATCTCTTGGTGGTGT	CATGTCTGGACCCCGAATCTT
SNHG9	GAGTGGCTATAAACGTCCG	AAGACGTGGGACAGCCAAGT
VMP1	ATCGGTACAGCAATCGGAGA	TCTGCATGTTCAGCATCTCT

2.6 Plasmid Construction and Dual-Luciferase Reporter Assays

2.6.1 Plasmid Construction

Two types of reporter vectors were constructed. First, for promoter activity assays, putative promoter regions of *SNHG9* (fragments P1, P2, P3), *SNHG15*, and *SNHG17* were PCR-amplified from MCF7 genomic DNA. The resulting fragments were digested with restriction enzymes and ligated into the pGL3-Basic vector (Promega, E1751) using T4 DNA ligase (Thermo Fisher Scientific, EL0011). Second, for miRNA sponge assays, the full-length *SNHG9* transcript, its exons (R1, R2), and specific sub-regions were cloned into the 3' UTR of the firefly luciferase gene in the pMIR-REPORT vector (Invitrogen, AM5795). This was achieved by digesting the vector and inserts with SacI (Thermo Scientific, FD1133) and MluI (Thermo Scientific, FD0564) restriction enzymes, followed by ligation with T4 DNA ligase (Thermo Fisher Scientific, EL0011). Deletion mutants for predicted miRNA binding sites were generated using synthetic oligonucleotides. The oligonucleotides used for PCR amplification and generating deletion mutants are listed in Table 2.2. All plasmids were verified by Sanger sequencing.

Table 2.2. Oligonucleotides Used for Plasmid Construction.

Oligonucleotide Name	Forward Sequence (5'-3')	Reverse Sequence (5'-3')
SNHG9 Overexpression	TTTCTAGCTAGCGCGGCCCGGGAATCTA	TTTCCCAAGCTTTTAGACAAAACAGCTT
Oligonucleotides for Promoter Assay		
SNHG15-promoter-650bp	GGTCGGGGTACCCTTTCCCACATGCTCCCACA	GGTGGAAGATCTAGTGAAGGCTTGGTCTCCA

Table 2.2 (cont'd)

SNHG17-promoter-460bp	GGTCGGGGTACCCAAACAGAAAAGCAAAGCAAGTGG	GGTGGAAGATCTACTGCACCAAGGACAGGG
SNHG9-P1-promoter-73bp	GGTCGGGGTACCGGGAATCTACGTCACCCGAA	GGTGGAAGATCTGTGGATTCCCGAGCCGC
SNHG9-P2-promoter-202bp	GGTCGGGGTACCATCTCCGCAGCCCCG	GGTGGAAGATCTGGGTGGATTCCCGAGCC
SNHG9-P3-promoter-473bp	GGTCGGGGTACCGCTGCACCGGCGTCAT	CTACCCAAGCTTGTGGATTCCCGAGCCGC
TFF1 promoter	A kind gift from Muyan Lab, previously described in [246].	
CMV promoter	A kind gift from Muyan Lab, previously described in [247,248].	

Oligonucleotides for Sponge Assay

SNHG9-Full Region	TTTCCGCTCGAGCCGGGCTCTGAGTGCTCTTG	TTTCCCAAGCTTTTAGACAAAACAGCTTTATT
SNHG9-R1	TTTCGCGAGCTCGCGGCCCGGGAATCTA	TTTCCGACGCGTCTAAGTGAAGAGGAGAGCGCA
SNHG9-R2	TTTCGCGAGCTCGACACTGGTCCTCCCA	TTTCCGACGCGTTTAGACAAAACAGCTT
R1.1-oligo	CGCGGCCCGGGAATCTACGTCACCCGAAAAGCGACTATAAACGCCGGCGCCTCCGTCCCCAGCCGCGGCTCGA	CGCGTCGAGCCGCGGCTGGGGACGGAGGCGCCGGCGTTTATAGTCGCTTTTCGGGTGACGTAGATTCCCGGGCCGCGAGCT
R1.2-oligo	CGGAATCCACCCGAAGAGTGGCTATAAACGTCCGCGCCTCCATTGCGCTCTCCTCTTCACTTAGA	CGCGTCTAAGTGAAGAGGAGAGCGCAATGGAGGCGCGGACGTTTATAGCCACTCTTCGGGTGGATTCCGAGCT

Table 2.2 (cont'd)

R1.1-mir-1277-seed-del-oligo	CGCGGCCCGGGAACCCGAAAAGCGACTATAAACGCCGGCGCCTCCGTCCCCAGCCGCGGCTCGA	CGCGTCGAGCCGCGGCTGGGGACGGAGGCGCCGGCGTTTATAGTCGCTTTTCGGGTTCCCGGGCCGCGAGCT
R1.2-multiple-seeds-del-oligo	CGGAATCCACCCGAAGAGTGGCTATAAACGTCCGCGCCTCCTCCTCTTCACTTAGA	CGCGTCTAAGTGAAGAGGAGGAGGCGCGGACGTTTATAGCCACTCTTCGGGTGGATTCCGAGCT

2.6.2 Dual-Luciferase Reporter Assays

MCF7 cells in 48-well plates were co-transfected with 225 ng of an experimental firefly luciferase construct and 25 ng of a control plasmid expressing Renilla luciferase (pRL-TK; Promega, E2241) using TurboFect Transfection Reagent (Thermo Fisher Scientific, R0531). Twenty-four hours post-transfection, cells were lysed with Passive Lysis Buffer. Firefly and Renilla luciferase activities were measured sequentially using the Dual-Luciferase Reporter Assay System (Promega, E1910) on a Molecular Devices, SpectraMax iD3s device (735-0390). Data were expressed as the ratio of firefly to renilla luciferase activity.

2.7 Functional Assays

2.7.1 Subcellular Fractionation

A lncRNA's function is often dictated by its subcellular localization. To determine whether candidate lncRNAs (*SNHG9*, *SNHG15*, *SNHG17*) are localized to the nucleus or cytoplasm, MCF7 cells were subjected to biochemical fractionation using the adaptation of "Frankenstein protocol" [249] by Novoa Lab (CRG, Spain). The

detailed protocol is in Appendix C. Briefly, cells were first lysed with a mild detergent buffer that selectively disrupts the plasma membrane while leaving the nuclear envelope intact. The cytoplasmic fraction was collected as the supernatant following a low-speed centrifugation step. The remaining nuclear pellet was then washed and lysed with a second, stronger buffer to release the nuclear contents.

RNA was isolated independently from both the nuclear and cytoplasmic fractions. To validate the purity of the fractions, PCR was performed for control lncRNAs with well-established localizations. *NEAT1*, a lncRNA exclusively retained in nuclear paraspeckles, served as the nuclear marker, while *DANCR*, which is predominantly cytoplasmic, served as the cytoplasmic marker [145,219]. The localization of target lncRNAs were detected by PCR followed by agarose gel electrophoresis.

2.7.2 Generation of Stable Overexpression Cell Lines

To investigate the long-term consequences of elevated *SNHG9* expression, stable overexpression cell lines were generated in both MCF7 (ER-positive) and MDA-MB-231 (ER-negative) cells. The full-length cDNA of the spliced *SNHG9* isoform was cloned into a mammalian expression vector (pcDNA3.1+; Invitrogen, V79020) containing a neomycin (G418) resistance gene. Cells were transfected with either the *SNHG9* overexpression plasmid or an empty vector (EV) control. Twenty-four hours post-transfection, the culture medium was replaced with a selection medium containing G418 (Roche, 04727878001) at a concentration empirically determined to be lethal to untransfected cells (800 µg/mL for MCF7; 1000 µg/mL for MDA-MB-231) [229].

Cells were cultured under continuous selection pressure for 2–3 weeks until resistant colonies emerged and untransfected cells were eliminated. These colonies were then pooled to create stable polyclonal populations, a strategy that averages out potential clonal artifacts. For maintenance, the G418 concentration was halved. The successful and stable overexpression of *SNHG9* in these cell lines was confirmed by RT-qPCR.

It is noted that attempts to establish stable *SNHG9* knockdown cell lines were unsuccessful due to significant cell death observed during the selection process, suggesting a potential essential role for basal *SNHG9* levels in cell viability.

2.7.3 Cell Proliferation (MTT) Assay

The effect of *SNHG9* overexpression on cell proliferation was assessed using the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Carl ROTH, 4022) assay [250]. *SNHG9*-overexpressing (*SNHG9*-OE) and empty vector (EV) control cells were seeded at a low density (5,000 cells/well) into 96-well plates. For experiments investigating E2-dependence, cells were hormone-deprived prior to the assay and then cultured with or without 100 nM E2. At specified time points (0, 24, 48, 72, and 96 hours), 10 µL of MTT stock solution (5 mg/mL) was added to each well. During a subsequent incubation, viable cells with active mitochondrial reductases convert the yellow MTT tetrazolium salt into purple formazan crystals. The culture medium was then removed, and the formazan crystals were dissolved by adding 100 µL of a HCl-SDS solution. The absorbance of the resulting solution, which is proportional to the number of viable cells, was measured at 570 nm to generate growth curves. At least three biological and five technical replicates were measured for each condition.

2.7.4 Colony Formation Assay

To evaluate the impact of *SNHG9* on long-term clonogenic survival and anchorage-dependent growth, a colony formation assay was performed [251]. *SNHG9*-OE and EV control cells were seeded at a very low density (1,000 cells per well) into 6-well plates. This low density ensures that individual cells form distinct colonies derived from a single progenitor. The cells were cultured for an extended period of 10 to 12 days, with the growth medium replaced every 3–4 days. At the end of the incubation, when visible colonies had formed, they were washed with phosphate-buffered saline

(PBS), fixed with cold 100% methanol for 20 minutes, and stained with a 0.5% Crystal Violet solution (Sigma-Aldrich, C0775) (in 25% methanol) for 5 minutes. After washing away excess stain, the plates were air-dried, photographed, and the number of stained colonies in each well was counted. Three biological and two technical replicates were measured for each condition.

2.8 Statistical Analysis

All quantitative data from experimental assays are presented as the mean ± standard deviation (SD) or mean ± standard error of the mean (SEM), derived from a minimum of two or three independent biological replicates as specified in the respective figure legends. Statistical analyses were performed to determine significance using appropriate tests based on the experimental design. For comparisons between two experimental groups, a two-tailed Welch's t-test was used. For comparisons involving three or more groups, a One-Way Analysis of Variance (ANOVA) was performed, followed by a Tukey's Honestly Significant Difference (HSD) post-hoc test to identify significant pairwise differences.

For survival analyses based on public data, Kaplan-Meier curves were generated using the KMplot tool, with 95% confidence intervals displayed [244]. The statistical significance of differences between survival curves was determined using the log-rank test. In the Gene Set Enrichment Analysis (GSEA), the significance of pathway enrichment was determined based on the False Discovery Rate (FDR) q-value. Across all analyses, statistical significance was denoted using a standard convention: *$p<0.05$, **$p<0.01$, ***$p<0.001$, and ****$p<0.0001$. A p-value of ≥ 0.05 was considered not significant (ns).

CHAPTER 3

RESULTS

To begin deciphering the isoform-level complexity of the estrogen (E2)-regulated transcriptome, we performed Direct RNA Sequencing using the Oxford Nanopore Technologies (ONT) platform on MCF7 cells, an ER-positive breast cancer cell line. Cells were treated with 100 nM 17β-estradiol (E2) in a time-course experiment to capture dynamic transcriptional changes. Prior to data analysis, rigorous quality control measures were implemented to ensure the reliability and integrity of the sequencing data.

3.1 Quality Controls

To confirm the effectiveness of the E2 treatment in our experimental system, the expression of the known E2-responsive gene, Trefoil Factor 1 (*TFF1*) [65], was checked by RT-qPCR using the same total RNA prepared for ONT sequencing. As expected, *TFF1* mRNA levels showed a time-dependent increase following E2 exposure compared to hormone-deprived (HD) control, indicating activation of the estrogen signaling pathway (Figure 3.1).

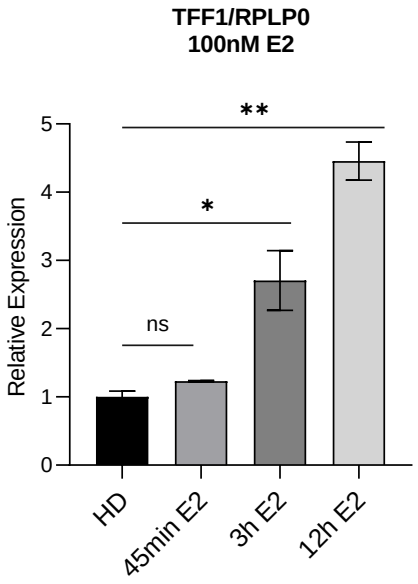


Figure 3.1. Time-dependent induction of *TFF1* gene expression by E2 in MCF7 cells. The bar chart displays the relative mRNA expression of *TFF1* in MCF7 breast cancer cells. Cells were treated with 100 nM of E2 for 45 minutes, 3 hours, and 12 hours. The 'HD' (Hormone Deprived) group serves as the untreated control. *TFF1* expression levels were normalized to the housekeeping gene *RPLP0*. The data are presented as the mean ± standard error. Statistical significance is indicated by asterisks (* $p < 0.05$, ** $p < 0.01$), and 'ns' denotes a non-significant difference.

After confirming the effectiveness of E2 treatment, we enriched mRNA from the RNA samples to deplete rRNA, minimizing its prevalence during sequencing. The poly(A) selection successfully depleted the rRNAs (Figure 3.2). Following poly(A) selection and sample preparation, the quality and characteristics of the RNA input were assessed using the Agilent Bioanalyzer “Eukaryote Total RNA Nano” chips. The electropherogram for poly(A) enriched RNA (HD sample) shows a reduction in the 18S and 28S rRNA peaks, which would typically dominate the total RNA pool and detected by sharp peaks (as indicated in the Figure 3.2 and Appendix Figure D.1). The total fraction of 18S and 28S rRNA is around 2.5% in our poly(A) enriched samples (0.4% for 18s and 2.1% for 28S rRNA), while typically they cover around

80-90 % of the total RNAs in eukaryotic cells. In addition, the samples display a broad distribution of RNA sizes, consistent with the heterogeneous nature of the enriched mRNA population with various lengths and abundances. These results indicated that we successfully enriched poly(A) RNAs to proceed with the library preparation.

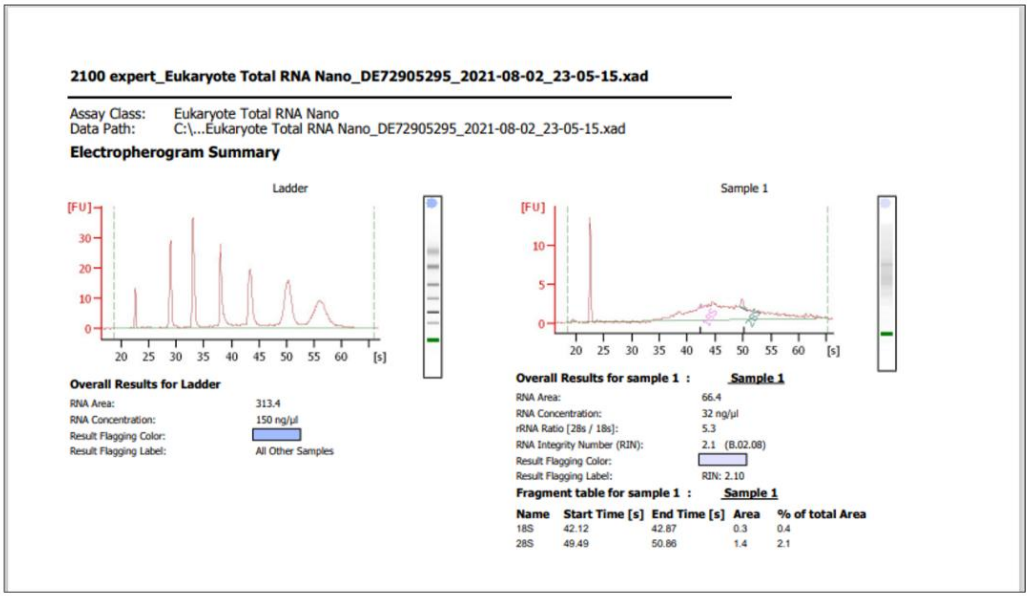


Figure 3.2. Quality assessment of poly(A)-enriched RNA. An Agilent Bioanalyzer electropherogram of the Hormone Deprived (HD) RNA sample after poly(A) selection to enrich for messenger RNA (mRNA). The trace demonstrates successful rRNA depletion, characterized by the absence of dominant 18S and 28S rRNA peaks and the presence of a broad distribution of fragment sizes, consistent with a heterogeneous mRNA population. The analysis table confirms that residual 18S and 28S rRNA constitutes only ~2.5% of the total RNA, indicating the sample quality is suitable for downstream library preparation and sequencing.

Direct RNA Sequencing (DRS) using Oxford Nanopore Technology was subsequently performed on poly(A)-selected RNA from MCF7 cells treated with E2 (45 min, 3 h, 12 h) and the control sample. The sequencing runs generated between 1-2 million total reads per sample, with around 92% of these reads passing standard quality filters, aligning well with good quality DRS runs (often >80-90%) [98].

Key sequencing metrics, including N50 read lengths reflecting the contribution of long reads to the dataset, and total data yield per sample are summarized in Table 3.1.

The "N50 (bp)" is a measure of read length continuity, signifying that 50% of the total sequenced bases are found in reads of this length or longer; the values here (1377 bp to 1503 bp) are consistent with successful DRS of eukaryotic full-length transcripts, which typically range from several hundred to a few thousand bases. "L50" value represents the number of the longest reads that together constitute half of the total data yield; for instance, for the HD sample, 935,655 of its longest reads make up 50% of its 1.42 GB yield.

The mean PHRED quality scores (Q) for these pass reads (9.33 to 10.3) correspond to a per-base accuracy of approximately 90%. These scores are usable for many transcriptomic applications like isoform detection and quantification [98]. While Oxford Nanopore has subsequently upgraded its sequencing chips, basecalling and buffers, enabling higher accuracies (often Q12-Q14+ for pass reads with the latest direct RNA chemistries), it's important to note that at the time our DRS was performed (2021), a PHRED score around 10, as observed here, was within the expected range and considered good for downstream applications [98].

Table 3.1. Summary of DRS Run Statistics and Read Quality by ToulligQC Report. This table summarizes key performance metrics for the sequencing of the four experimental samples: Hormone Deprivation (control), 45 min E2, 3 hours E2, and 12 hours E2. It details the run statistics (total reads, pass reads percentage, N50, yield), mean read lengths, and mean PHRED quality scores. The results demonstrate

high data output and quality across all samples, with yields exceeding 1 GB and over 91% of reads passing quality filters, confirming the dataset's robustness for subsequent comparative analysis.

Toullig Report Summary		HD	45 min E2	3 h E2	12 h E2
Run Statistics	Read count	1232807	973122	1762260	2174223
	Pass reads count (%)	1150375 (93.31%)	895051 (91.98%)	1640749 (93.1%)	2023918 (93.09%)
	N50 (bp)	1503	1499	1464	1377
	L50	935655	743314	1339206	1651978
	Yield (GB)	1.42	1.09	1.95	2.28
Read lengths (bp)	All reads (mean)	1153.5	1119.08	1109.12	1047.55
	Pass reads (mean)	1192.68	1165.75	1146.96	1083.74
PHRED score	All reads (mean)	9.98	9.62	9.95	9
	Pass reads (mean)	10.3	9.97	10.27	10.33

The sequencing reads exhibited high quality, facilitating subsequent isoform analysis. DRS successfully captured long reads, with lengths exceeding 10 kilobases (Figure 3.3).

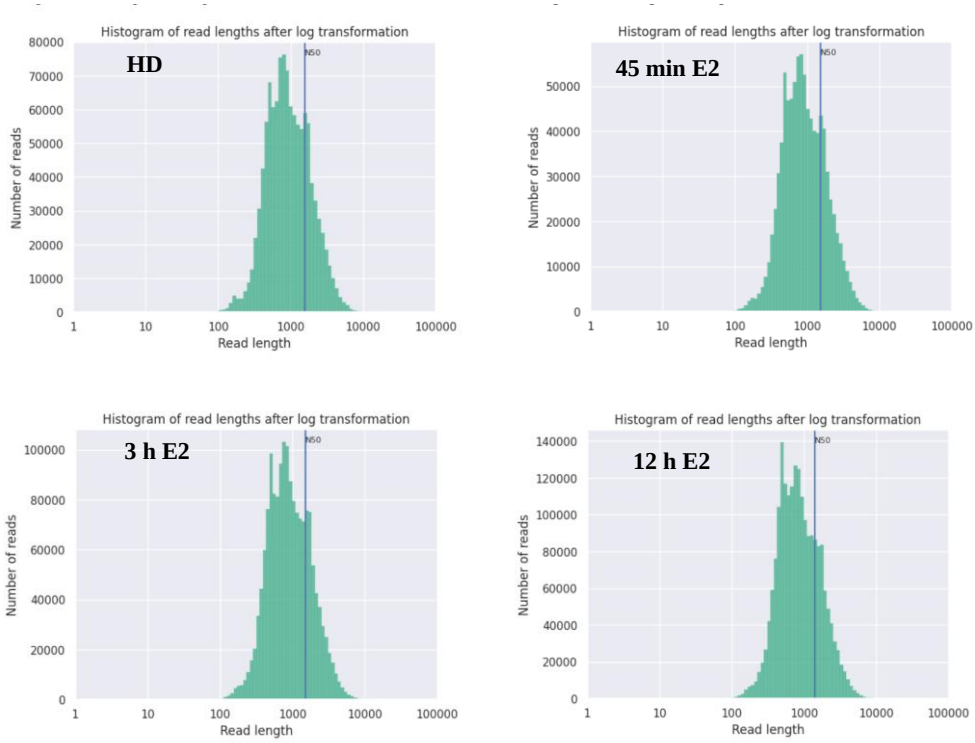


Figure 3.3. Read-length distribution of direct RNA sequencing reads. These Nanoplot histograms illustrate the read length distributions for each of the four experimental conditions HD, 45min E2, 3h E2, and 12h E2. Each plot displays the number of reads (y-axis) corresponding to a given read length (x-axis, log scale). The vertical blue line marks the N50 read length for each sample. The consistent shape and peak of the distributions across all four panels indicate a similar read length profile, confirming the uniformity of the sequencing data and its suitability for comparative analysis.

We obtained a high-quality, long-read transcriptome dataset, which was validated for E2 responsiveness and sequencing quality. This dataset formed the basis for a detailed analysis of E2-induced isoform expression and AS changes using the FLAIR (Full-Length Alternative Isoform analysis of RNA) pipeline.

FLAIR is a computational pipeline specifically developed for long-read RNA sequencing data, like ONT [77]. It is designed to improve the accuracy of transcript isoform identification and quantification by first correcting sequencing errors in reads, often using genome annotations or splice site information, and then collapsing redundant corrected reads to define a high-confidence set of full-length isoforms. FLAIR can then quantify the expression of these defined isoforms, enabling the detection of novel isoforms and the analysis of differential isoform usage and AS events between different conditions, such as the E2-induced changes in our samples.

3.2 Differentially Expressed Genes (DEG)

Having established the quality of the DRS data and confirmed the effectiveness of the E2 treatment, we next investigated the impact of E2 on the overall transcriptomic landscape by performing differential gene expression analysis using FLAIR's “Diff_iso” module.

Unlike short-read methods, where isoforms often need to be computationally pieced together from smaller fragments, DRS long reads allow many isoforms to be sequenced completely, or 'in one piece.' FLAIR is specifically made to accurately find and count these complete isoforms from the DRS data [77]. This precise isoform annotation is important because it provides a more comprehensive and accurate basis for the differential analysis that follows. By first identifying all the different expressed isoforms, the quantification used in differential expression calculations (whether at the isoform or gene level) is more solid, leading to a more reliable understanding of the actual biological effect of E2 treatment on the transcriptome.

As an initial validation, the expression of well-characterized E2-induced genes *TFF1* and *GREB1* was examined [64,65]. Consistent with the literature, both *TFF1* and *GREB1* mRNA levels were significantly upregulated in response to E2 stimulation (Figure 3.4).

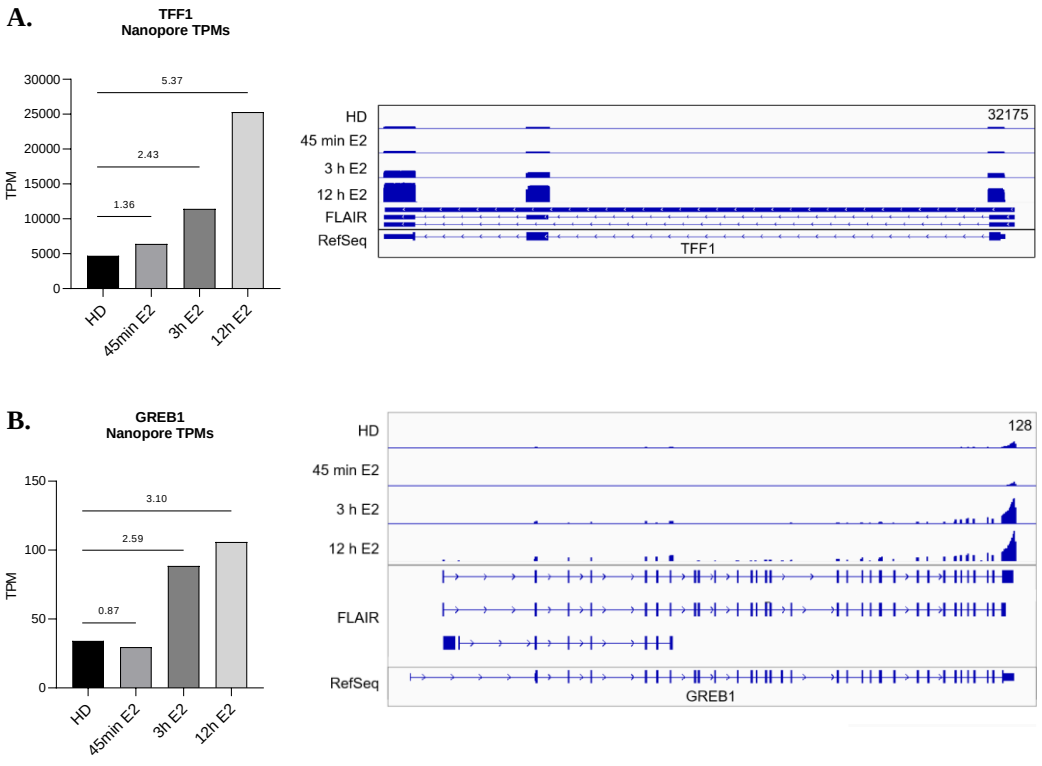


Figure 3.4. E2-induced expression of *TFF1* and *GREB1* and their transcript isoforms. The bar charts on the left show the time-dependent increase in expression for the E2-responsive genes *TFF1* (A) and *GREB1* (B) in Transcripts Per Million (TPM), based on DRS data. TPM fold change ratios (versus HD) are given for each time point. The corresponding panels on the right are Integrative Genomics Viewer (IGV) snapshots showing the detected transcript isoforms for each gene identified by the FLAIR pipeline, with the annotated RefSeq gene structure for reference. The number in the top-right of each track indicates the data range for the read coverage signal, indicating the maximum read coverage height displayed in that view.

For a more stringent analysis of DEGs, particularly because we had single biological samples for each time point, filters were applied requiring a minimum expression level of 10 TPM (Transcripts Per Million) in at least one sample and a fold change of >1.5 (upregulation) or <0.60 (downregulation) for coding genes. We used a lower

TPM threshold of 2 for noncoding genes due to their generally lower expression levels [130].

A summary of the number of DEGs meeting these criteria at each time point is presented in Figure 3.5.A. Notably, MCF7 cells exhibited differential expression of genes at 45 minutes (1943 upregulated vs. 1139 downregulated) and 3 hours (1849 upregulated vs. 1210 downregulated) post-E2 induction. However, at the 12-hour time point, this trend reversed, with a larger number of genes being downregulated (1399 upregulated vs. 1890 downregulated) (Figure 3.5.B). This pattern aligns with previous reports suggesting an initial transcriptional burst induced by E2, followed by broader repressive effects or feedback mechanisms at later stages [252,253].

A.

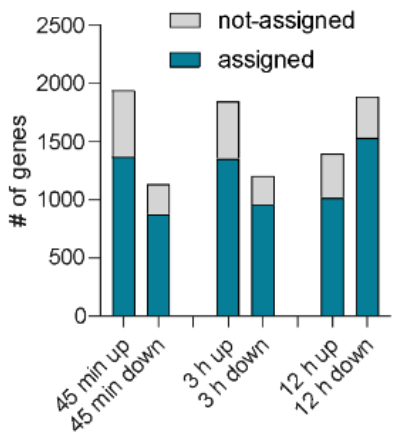


Figure 3.5 (cont'd)

B.

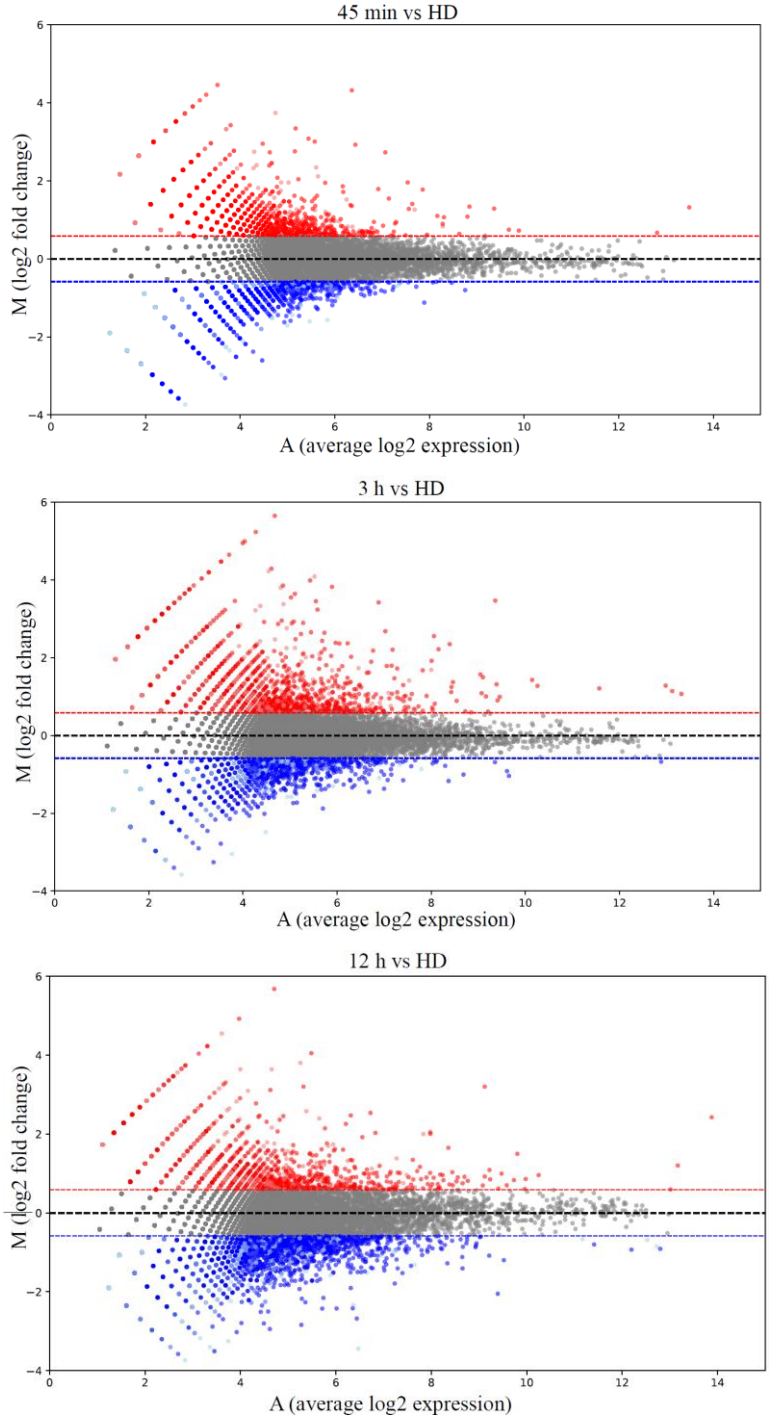
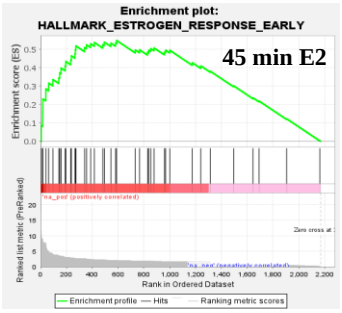


Figure 3.5. Differentially Expressed genes and MA plots. (A) Number of differentially expressed genes over time. The chart shows the count of upregulated (up) and downregulated (down) genes at 45 minutes, 3 hours, and 12 hours post-stimulus. Genes are grouped as “assigned" if they are annotated (teal) and as "not-assigned” if they are not annotated (gray). (B) For each time-point, an MA plot is provided to visualize DEGs. In an MA plot, the Y-axis (M-value) represents the log2 fold change in gene expression between conditions, while the X-axis (A-value) represents the average expression level of each gene. Genes shown in red are upregulated, and genes shown in blue are downregulated at that specific time-point.

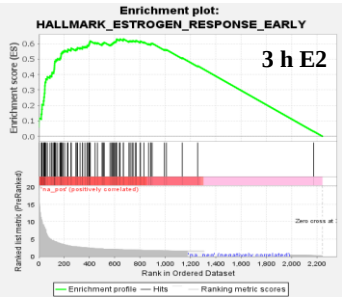
3.2.1 Gene Set Enrichment Analysis (GSEA)

To understand the broader biological pathways and functions associated with these differentially expressed genes, Gene Set Enrichment Analysis (GSEA) was performed using “GSEA MSigDB” [254,255] on the lists of upregulated and downregulated genes identified at each time point. E2 responsive gene signatures are well established [51] and we aimed to confirm that our data reflected similar patterns of DEGs, while also providing the resolution to detect isoform-level differences.

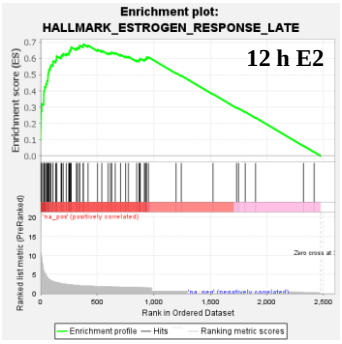
Across all durations, GSEA revealed significant enrichment (FDR q-value < 0.05) for hallmark gene sets associated with 'Estrogen Response Early' and 'Estrogen Response Late', further confirming the expected biological response (Figure 3.6).



Geneset	# of genes	n. Enrichment Score	n. p-value
HALLMARK_TNFA_SIGNALING_VIA_NFKB	55	1.84	0
HALLMARK_HYPOXIA	28	1.83	0
HALLMARK_MYOGENESIS	20	1.75	0
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	21	1.71	0.006
HALLMARK_ESTROGEN_RESPONSE_LATE	31	1.56	0.008
HALLMARK_ESTROGEN_RESPONSE_EARLY	42	1.53	0.001
HALLMARK_INFLAMMATORY_RESPONSE	25	1.41	0.031
HALLMARK_P53_PATHWAY	33	1.28	0.09
HALLMARK_INTERFERON_GAMMA_RESPONSE	25	1.22	0.166
HALLMARK_HEME_METABOLISM	21	1.18	0.185
HALLMARK_IL2_STAT5_SIGNALING	26	1.08	0.348
HALLMARK_UV_RESPONSE_UP	20	1.04	0.393
HALLMARK_KRAS_SIGNALING_UP	16	1.03	0.417



Geneset	# of genes	n. Enrichment Score	n. p-value
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	35	1.78	0.001
HALLMARK_ALLOGRAFT_REJECTION	16	1.77	0
HALLMARK_TNFA_SIGNALING_VIA_NFKB	74	1.73	0
HALLMARK_ESTROGEN_RESPONSE_EARLY	87	1.73	0
HALLMARK_INTERFERON_GAMMA_RESPONSE	35	1.71	0
HALLMARK_INFLAMMATORY_RESPONSE	28	1.63	0.001
HALLMARK_HYPOXIA	53	1.61	0
HALLMARK_ESTROGEN_RESPONSE_LATE	61	1.58	0
HALLMARK_IL6_JAK_STAT3_SIGNALING	15	1.52	0.02
HALLMARK_KRAS_SIGNALING_UP	19	1.42	0.056
HALLMARK_XENOBIOTIC_METABOLISM	24	1.41	0.042
HALLMARK_UV_RESPONSE_UP	39	1.4	0.018
HALLMARK_UV_RESPONSE_DN	40	1.39	0.02



Geneset	# of genes	n. Enrichment Score	n. p-value
HALLMARK_ESTROGEN_RESPONSE_LATE	65	1.72	0
HALLMARK_ESTROGEN_RESPONSE_EARLY	83	1.67	0
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	36	1.58	0
HALLMARK_MYC_TARGETS_V2	25	1.44	0.013
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	22	1.35	0.06
HALLMARK_INTERFERON_GAMMA_RESPONSE	34	1.33	0.037
HALLMARK_TNFA_SIGNALING_VIA_NFKB	46	1.31	0.011
HALLMARK_UV_RESPONSE_UP	29	1.28	0.062
HALLMARK_MYC_TARGETS_V1	19	1.27	0.118
HALLMARK_INFLAMMATORY_RESPONSE	16	1.26	0.123
HALLMARK_GLYCOLYSIS	38	1.25	0.064
HALLMARK_IL2_STAT5_SIGNALING	33	1.23	0.091
HALLMARK_COMPLEMENT	17	1.23	0.15

Figure 3.6. GSEA results for DEGs and Enrichment scores in each time point. This figure presents Gene Set Enrichment Analysis (GSEA) results, performed using "GSEA MSigDB" [254,255] for each E2 treatment timepoint compared to HD. Left panels: Representative enrichment plots illustrate the distribution of member genes; an accumulation of 'hits' towards the top of the ranked list (indicating positive correlation with the E2 response, such as upregulation) results in a positive Enrichment Score. Right panels list top enriched gene sets, their Normalized Enrichment Score (NES), which reflects the degree and direction of enrichment, and the normalized p-values, which typically considered significant when p-value < 0.05 (indicated as green backgrounds).

Additionally, pathways related to 'TNFα Signaling via NF-κB, 'Epithelial Mesenchymal Transition (EMT)', and 'Inflammatory Response' were consistently enriched among the differentially regulated genes across the time course, highlighting the pleiotropic effects of sustained estrogen signaling in this cell model. These findings also correlated with the well-established E2-regulated pathways [256].

3.2.2 Transcriptome Landscape

The results presented thus far confirm that the E2 induction produced the expected changes in the transcriptome. Leveraging the long-read capability of DRS, we next focused on E2's impact at the transcript isoform level using the quantification output from the FLAIR pipeline. Analysis of differentially expressed transcripts (using similar fold-change and TPM criteria as for genes) revealed that the majority were isoforms of protein-coding genes (~94%), with long noncoding RNAs (lncRNAs) comprising the next largest category (~4%), followed by other noncoding transcript types (Figure 3.7).

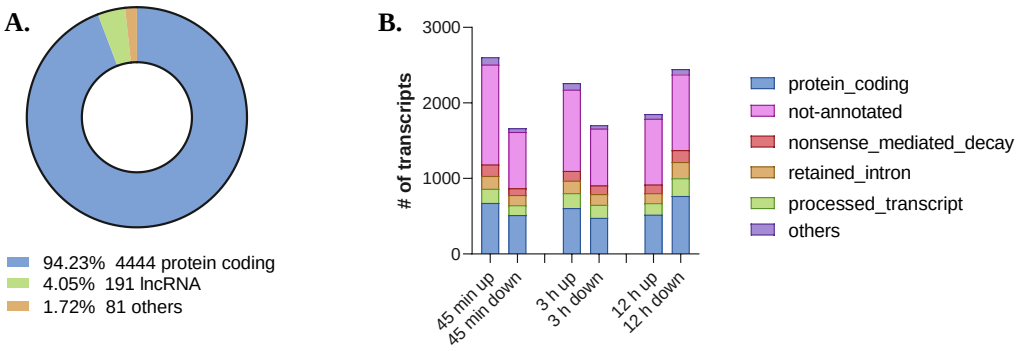


Figure 3.7. Summary of differentially expressed transcripts upon E2 induction. The donut chart illustrates the overall composition of all unique transcripts that were differentially expressed across the time course. The bar chart shows the number of upregulated (fold change > 1.5) and downregulated (fold change < 0.6) transcripts at each specific time point (45 min, 3h, and 12h). The bars are segmented by the standard "biotype" transcript labels of Ensembl database.

Notably, a large fraction of transcripts did not precisely match existing annotations; these are grouped as 'not-annotated' transcripts (summarized in Figure 3.7.B). While these not-annotated isoforms do not necessarily represent entirely new gene loci, they likely signify novel transcriptional variants of known genes, such as those with previously uncharacterized transcription start/end sites or AS patterns. Therefore, to begin exploring this not-annotated transcript group, we first looked into whether these isoforms were differentially expressed in both coding and non-coding genes.

3.3 Changes in Protein-Coding Genes

3.3.1 E2 induced isoform expression

E2 signaling is known to regulate not only transcription initiation but also AS and APA events, leading to the expression of distinct transcript isoforms with potentially different functions [76,109]. We investigated the isoform architecture of transcripts regulated by E2 in MCF7 cells using FLAIR's output.

3.3.1.1 E2-Induced Alternative Splicing Dynamics

Beyond changes in overall transcript abundance, we investigated how E2 treatment affects AS in coding genes. A dynamic regulation of alternatively spliced transcripts, including both previously annotated and potentially novel (non-annotated) isoforms, was observed across the E2 treatment time course (45 min, 3 h, 12 h) (Figure 3.8).

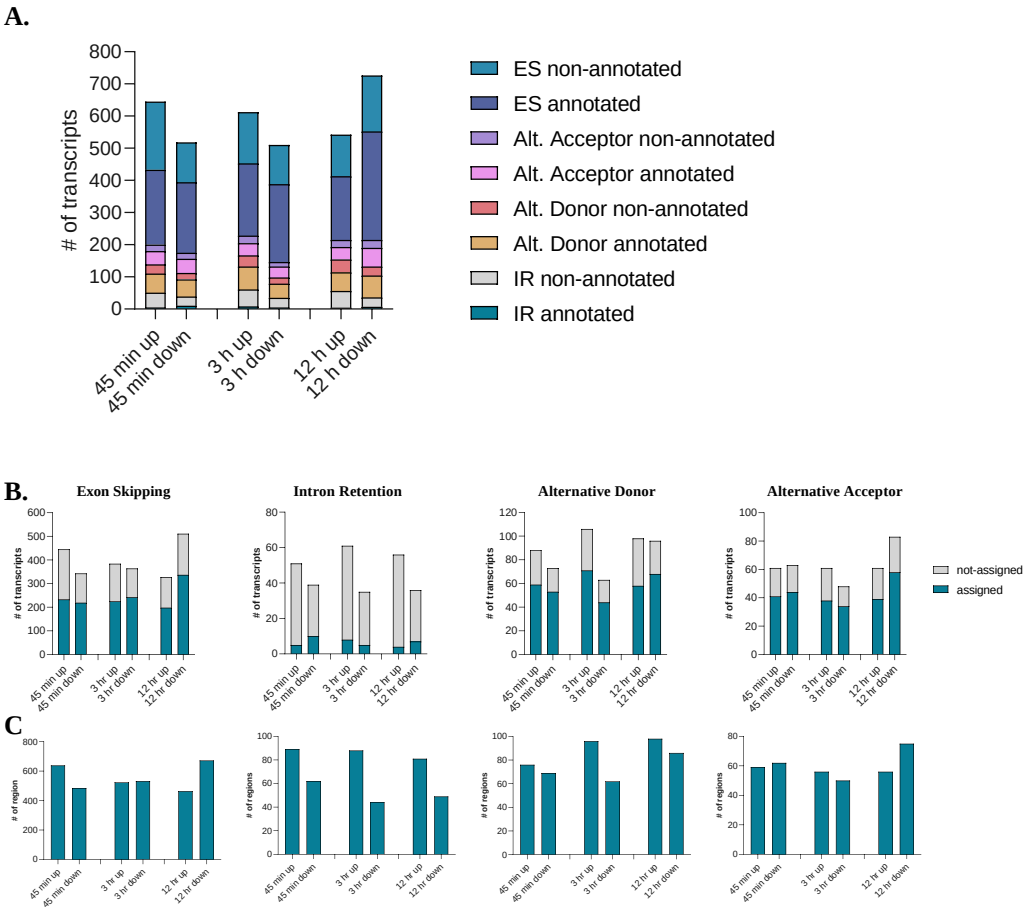


Figure 3.8. Quantification and characterization of AS events across the E2 treatment time course. This figure summarizes the dynamic changes in AS at different time points (45 min, 3 h, 12 h) following E2 treatment. Upper panel, (A) The stacked bar

chart shows the number of transcripts with different AS events, categorized by up- or down-regulation. The events include Exon Skipping (ES), Alternative Acceptor, Alternative Donor, and Intron Retention (IR), further classified as previously annotated or novel (non-annotated). Lower panel, these bar charts provide a more detailed breakdown for each major AS type, showing the number of affected transcripts (B) and genomic regions(C). Transcripts are grouped as “assigned" (teal) if they are annotated and as "not-assigned” (gray) if they are not annotated.

These E2-induced AS events were categorized based on the type of splicing change: alternative donor site usage (Alt. Donor; affecting 5’ splice sites), alternative acceptor site usage (Alt. Acceptor; affecting 3’ splice sites), intron retention (IR), or exon skipping (ES). Across all time points, exon skipping was the most frequently observed AS event type for differentially regulated transcripts derived from protein-coding genes, affecting both annotated and non-annotated isoforms (Figure 3.8.A). Intron retention events were notably more prevalent among the non-annotated isoforms compared to annotated ones, suggesting the identification of previously uncharacterized splicing patterns facilitated by long-read sequencing (Figure 3.8.B-C).

We detected multiple splicing isoforms across time points. These included alternative transcription start site (TSS) usage in *KEAP1* (Kelch-like ECH-associated protein 1) and *TPM1* (Tropomyosin 1), alternative 5' exon boundaries in *MRM2* (Mitochondrial RRNA Methyltransferase 2) (Appendix Figure D.2), intron retention leading to altered 3' ends in *CRB3* (Crumbs cell polarity complex component 3), and exon skipping in *HAGH* (Hydroxyacylglutathione hydrolase) (Figure 3.9). Notably, *CRB3* has been implicated in breast cancer stemness and tamoxifen resistance [257], while *HAGH* is involved in detoxification pathways [258]. However, caution is warranted, as ONT reads transcripts from their 3′ end to their 5′ end, which may result in incomplete or inaccurate capture of the 5′ ends of isoforms.

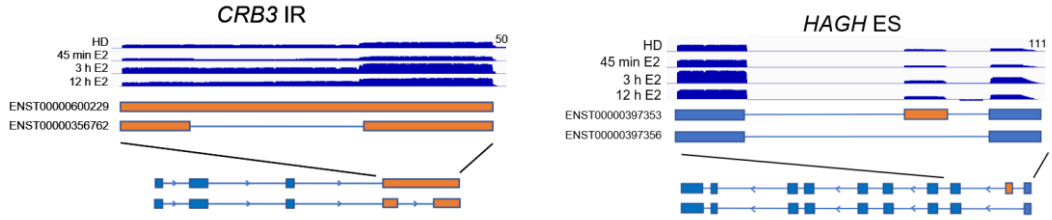


Figure 3.9. Splicing events *CRB3* and *HAGH*. For each gene, the upper tracks show the DRS read coverage, while the diagrams below show the corresponding transcript isoform structures. Left Panel: An example of Intron Retention (IR) is shown in the *CRB3* gene. Right Panel: An example of regulated Exon Skipping (ES) is shown in the *HAGH* gene. Orange rectangular indicate the corresponding exons in both images. The value in the top-right of each track indicates the maximum read coverage.

3.3.1.1.1 Functional Enrichment of AS transcripts

To explore the potential functional consequences of the observed splicing changes, Gene Set Enrichment Analysis (GSEA) was performed on the set of DEGs exhibiting significant exon skipping or intron retention events upon E2 treatment, using g:Profiler [238]. This analysis revealed enrichment for genes involved in biological processes such as 'mRNA processing', 'nucleocytoplasmic transport', and 'tRNA processing' (Figure 3.10). These findings suggest that E2-induced splicing changes may impact fundamental aspects of RNA metabolism and protein function through splicing changes.

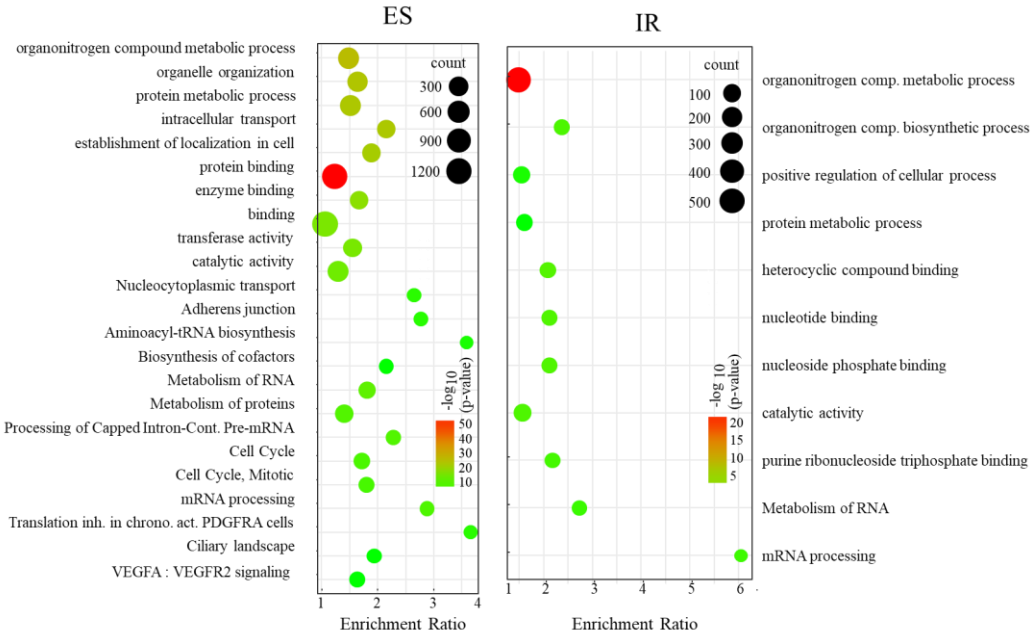


Figure 3.10. Enrichment Analysis of ES and IR events in DEGs using g:Profiler. This figure presents the functional enrichment analysis of mRNAs with Exon Skipping (ES, left panel) and Intron Retention (IR, right panel) events. In each plot, enriched Gene Ontology (GO) terms are listed on the y-axis, while the x-axis shows their statistical significance as the -log10(adjusted p-value). The dot color represents the significance level, with yellow being the most significant. The dot size represents the total number of isoforms.

In summary, the splicing analysis revealed that E2 treatment dynamically affects AS in MCF7 cells. Exon skipping emerged as the predominant splicing event, and long-read sequencing uncovered potentially novel splicing patterns. Functional enrichment analysis suggests these E2-induced splicing changes primarily impact genes involved in crucial RNA processing and transport pathways, highlighting AS as a significant layer of post-transcriptional regulation in estrogen response.

3.3.1.2 E2-Induced APA in Protein-Coding Genes

In addition to AS, E2 was found to influence the choice of polyadenylation sites in protein-coding genes, another layer of post-transcriptional gene regulation. This regulation can lead to transcript isoforms with different 3' untranslated regions (3'UTRs) or, in cases of intronic polyadenylation, to the production of truncated proteins.

3.3.1.2.1 3’UTR APA events

We detected differential levels of 3’UTR isoforms even at the early 45-minute time point. Notable examples included 3'UTR lengthening in genes such as *SOS1* (SOS Ras/Rac Guanine Nucleotide Exchange Factor 1) and *ARIH1* (Ariadne RBR E3 Ubiquitin Protein Ligase 1), and conversely, 3'UTR shortening in genes like *GSPT1* (G1 To S Phase Transition 1) and *CDC73* (Cell Division Cycle 73, Paf1/RNA Polymerase II Complex Component) 1 (Figure 3.11).

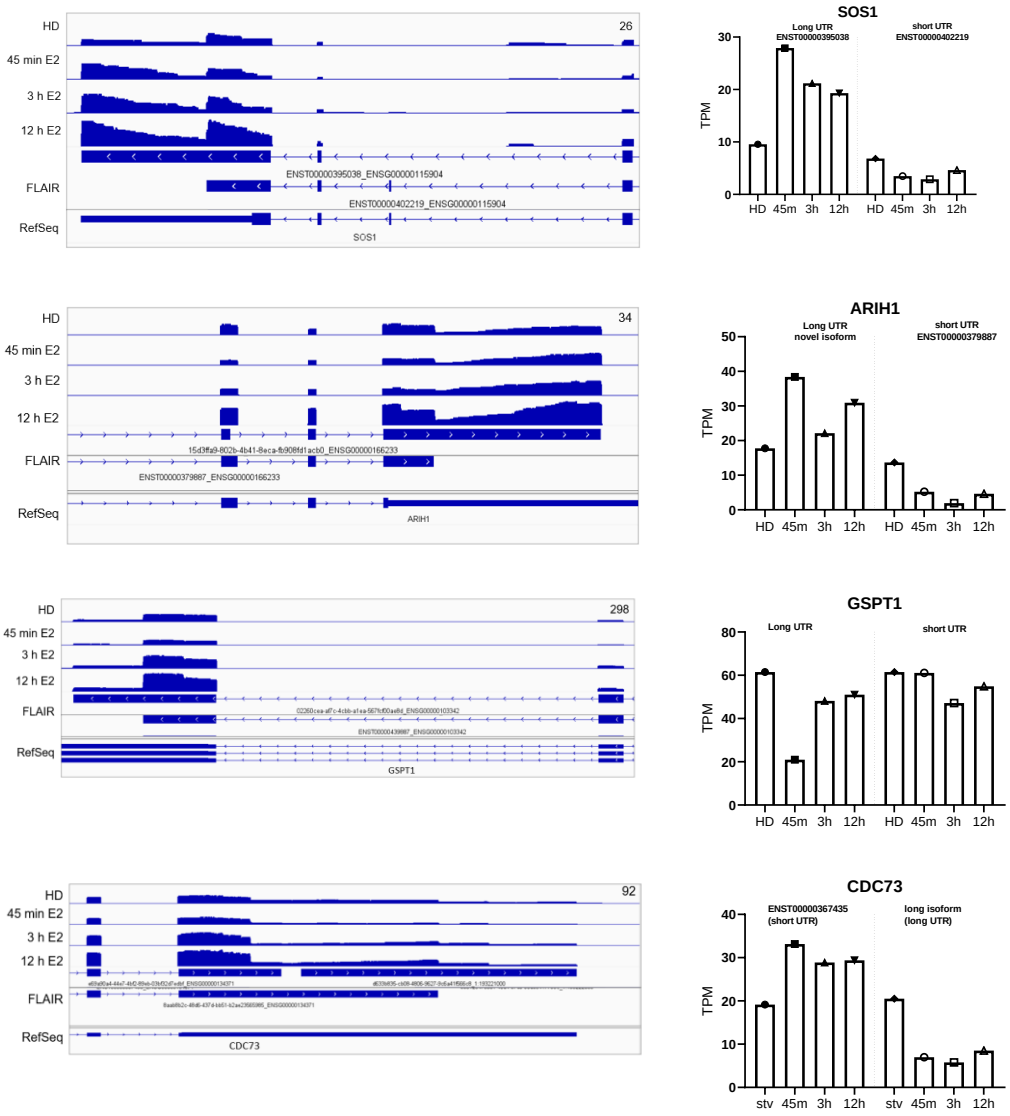


Figure 3.11. Differential expression of 3'UTR isoforms for selected genes. This figure displays data for four example cases (*SOS1*, *ARIH1*, *GSPT1*, and *CDC73*), illustrating changes in isoform expression. For each gene, the left panel shows a genome browser view of the isoforms for each gene identified by the FLAIR pipeline with the annotated RefSeq gene structure for reference. The right panel quantifies the expression levels (in TPM) of isoforms with different 3' UTRs across several time points. The value in the top-right of each track indicates the maximum read coverage.

Changes in 3'UTR length can significantly impact mRNA regulation by altering the availability of binding sites for microRNAs (miRNAs) and RNA-binding proteins, thereby affecting transcript stability, localization, and translation efficiency, without changing the primary protein sequence. Our analysis using the "RegRNA 3.0" indicated that 3' UTR sequences that are included in the longer 3'UTR isoforms of *SOS1* and *ARIH1* added numerous binding sites for RNA-binding proteins and miRNAs. On the other hand, shorter 3' UTR isoforms of *GSPT1* and *CDC73* may allow these transcripts to escape repression by specific miRNAs or RNA-binding proteins (Figure 3.12).

Figure 3.12. Annotation of Regulatory Elements in Differentially Expressed 3' UTR Sequences. This figure displays the analysis of the functional regulatory elements contained within the specific 3' UTR sequences that are gained or lost due to APA events in four selected genes. The tracks were generated using the RegRNA 3.0 database [235] and map various annotated features, including Untranslated Region (UTR) Motifs, RNA Ligand and Protein Binding Sites, and miRNA Target Sites.

The top two panels analyze the extended 3' UTR sequences that are included in the longer isoforms of *SOS1* and *ARIH1*. The presence of numerous potential binding sites suggests that these longer isoforms are subject to more complex post-transcriptional regulation. Lost 3'UTR sequence: The bottom two panels analyze the distal 3' UTR regions that are excluded in the shorter isoforms of *GSPT1* and *CDC73*. The loss of these regulatory element-rich sequences may allow the shorter transcripts to escape repression by specific miRNAs or RNA-binding proteins.

3.3.1.2.2 Intronic APA

While DRS identified transcript 3'-ends mapping to the expected terminal exons, it also detected a significant number of reads mapping to intronic regions. These findings pointed to widespread Intronic Polyadenylation (IPA) events. Overall, DRS captured 413 intronically polyadenylated and upregulated transcripts, a substantial portion of which, 246 transcripts, were not previously annotated. To explore these further, we organized the transcript ends into two main categories; Early Intron (EI) for transcripts terminating in the first or second intron, and Mid-Intron (MI) for those terminating in subsequent introns.

Analysis of these IPA transcripts revealed several key characteristics. A significant number of transcripts in both the EI (n=110) and MI (n=240) groups were differentially expressed following 45 minutes of E2 treatment (Figure 3.13). The vast majority of these intronic 3’-ends contained conserved polyadenylation signals (e.g., AAUAAA), suggesting they are generated through a regulated biological mechanism rather than random events.

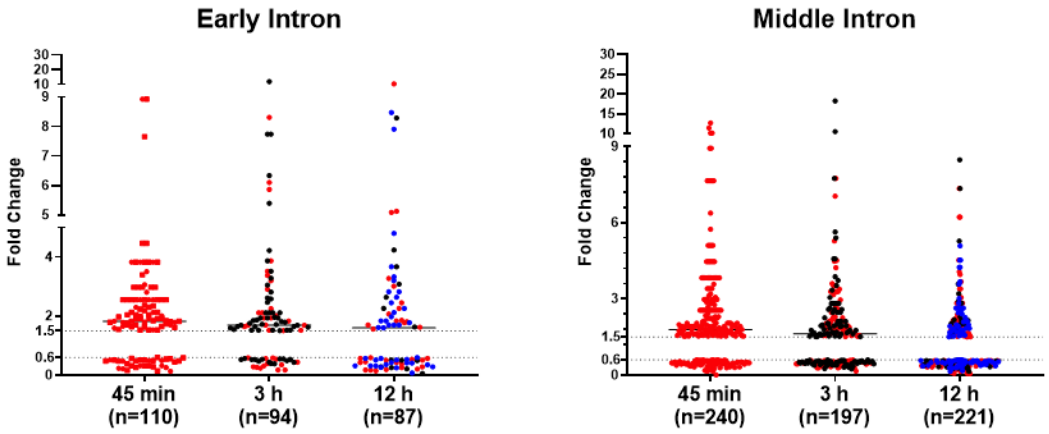


Figure 3.13. Differential expression of transcripts terminating in Early Introns (EI) and Middle Introns (MI) after E2 treatment. The scatter plots display the fold change for individual transcripts, identified by DRS, at 45 minutes (red), 3 hours (black), and 12 hours (blue). Dotted lines indicate the thresholds for upregulation (>1.5-fold change) and downregulation (<0.6-fold change). The total number of altered transcripts (n) is noted below each time point.

Within the detected IPA isoforms, we focused on the MI group, as these events are more likely to generate stable mRNAs that can be translated into C-terminally truncated proteins with potentially altered functions. We selected candidates from this group that are known to be involved in the E2 response for further investigation.

We investigated IPA events leading to truncated transcripts for *TLE1* (TLE family member 1) and the well-known E2-regulated gene *GREB1* (Growth Regulating Estrogen Receptor Binding 1) (as well as others detailed in Appendix Figure D.3) (Figure 3.14.A). GREB1 is an E2-responsive gene [259] and TLE1 mediates estrogen receptor binding and transcriptional activity in breast cancer cells [260].

For the *TLE1*, this IPA event occurs due to the utilization of a cryptic polyadenylation site located within the seventh intron of the pre-mRNA at chr9:84248387 [-] (Figure 3.14.A). The usage of this intronic poly(A) site results in

a prematurely terminated transcript. Similarly, a truncated mRNA isoform for *GREB1* is generated via an IPA event within the seventh intron at chr2:11721391[+] (Figure 3.14.A). While these poly(A) sites are not part of the canonical transcripts, they are experimentally proven polyadenylation sites listed in the PolyA_DB database [236].

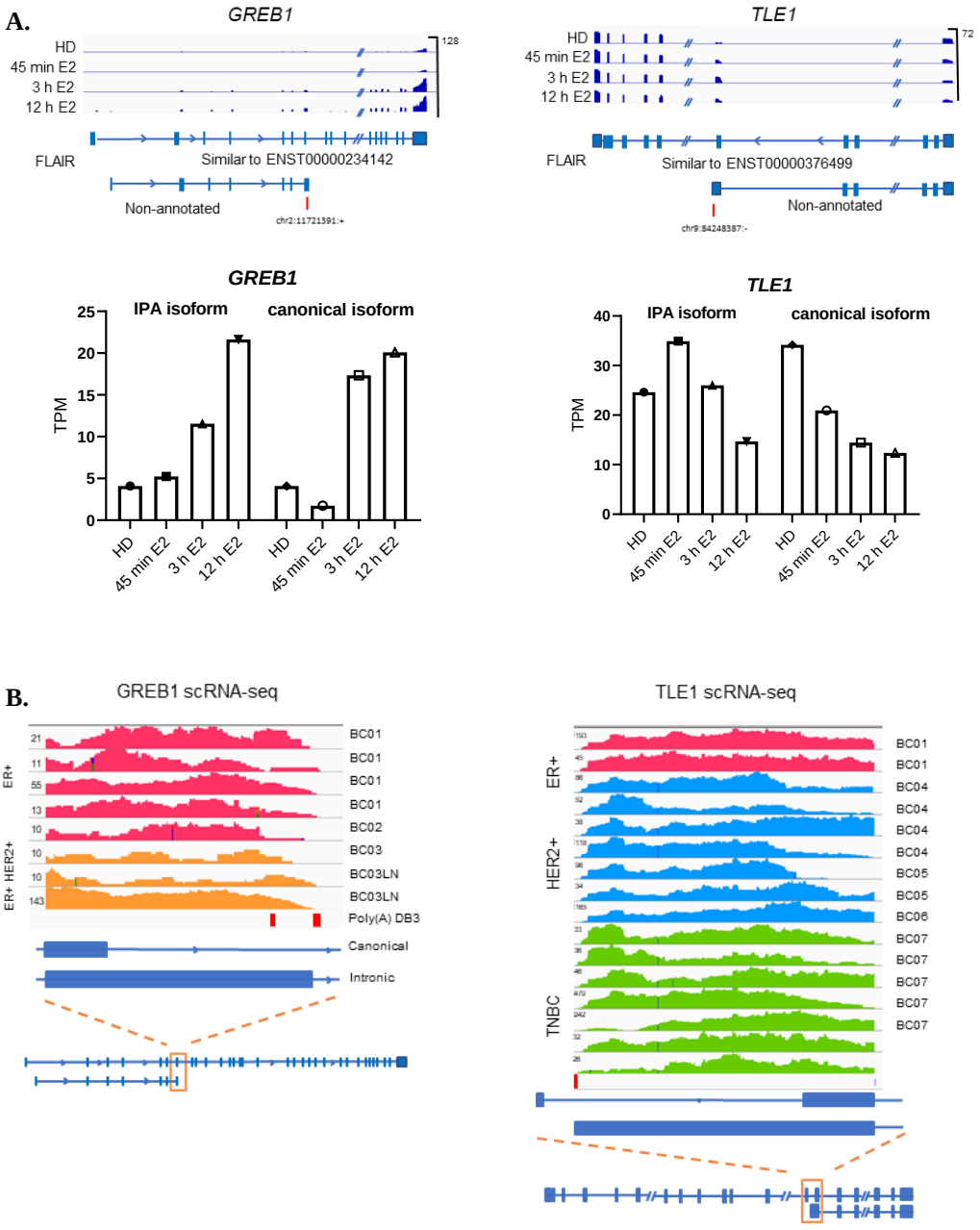


Figure 3.14. IPA events for *GREB1* and *TLE1*. (A) Gene structure and isoforms detected by RNA long-read sequencing (FLAIR pipeline). Red squares indicate the poly(A) site coordinates near the 3' ends of the intronic isoforms, detected by PolyA_DB V3.2 [236] The value in the top-right of each track indicates the maximum read coverage. (B) Supporting evidence from single-cell RNA-seq data from breast cancer patient samples with different subtypes (GSE75688) [243]. The value in the top-left of each track indicates the maximum read coverage.

The full-length TLE1 protein consists of 770 amino acids and contains an N-terminal glutamine-enriched (Q) domain, a Gly-Pro-enriched (GP) domain, a CcN cysteine-rich domain, a Ser-Pro-enriched (SP) domain, and a tryptophan-aspartic acid repeat (WDR) region. A truncated isoform, TLE1-IPA, however, codes for a much shorter protein of 200 amino acids. This isoform retains the N-terminal Q-domain and the GP domain but loses all other domains, including the C-terminal WDR domain, which is crucial for protein-protein interactions. The truncation occurs after the 192nd amino acid, with the addition of eight new amino acids before a stop codon. The loss of the WDR domain prevents it from engaging in its usual interactions, potentially altering transcriptional regulation.

The GREB1 protein has an N-terminal Zn-binding domain, a helicase domain, and a C-terminal Fringe-like GT domain. While the canonical isoform is 1949 amino acids long, the IPA isoform of GREB1 is only 364 amino acids. This shorter isoform possesses the Zn-binding domain but completely lacks the GT domain at the C-terminus. This GT domain is significant as it is responsible for the O-GlcNAcylation of the estrogen receptor alpha (ERα), a process that stabilizes the ERα protein. The loss of this domain in the truncated GREB1 isoform may indicate that it cannot perform this stabilizing function.

Notably, evidence supporting the specific IPA events in *TLE1* and *GREB1* was also found upon examining single-cell RNA-seq data (GSE75688) from breast cancer

patient samples [243] (Figure 3.14.B), and these IPA isoforms were subsequently validated experimentally by other members of our lab.

3.3.2 Detection of Read-Through Transcripts

We also detected 107 instances of read-through transcripts for protein-coding genes (FC>,1.5 and TPM>10 at least in one time point), where RNA Polymerase II fails to terminate at the canonical polyadenylation site, resulting in fusion transcripts incorporating downstream genes or genomic regions. This phenomenon can arise from various mechanisms and is sometimes associated with specific cellular states like stress or cancer [261]. Known and novel read-through events detected in our data included *ST20-MTHFS*, *USF1-F11R*, *TSTD1-F11R*, *DMWD-DMPK*, and *CTSD-IFITM10* (Figure 3.15). The *CTSD-IFITM10* read-through transcript, for instance, has previously been linked to proliferation in MCF7 cells [262]. Identifying these read-through transcripts adds another layer to the complexity of the transcriptome, highlighting mechanisms beyond canonical gene expression that may influence cellular phenotypes.

A.

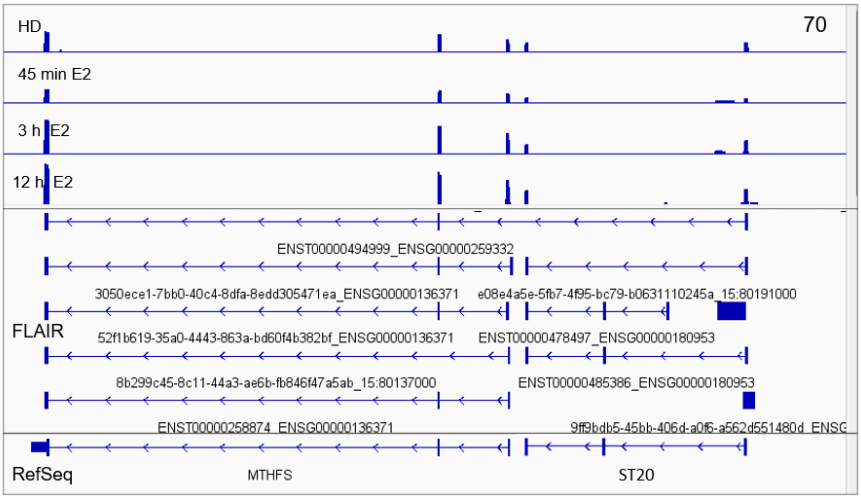
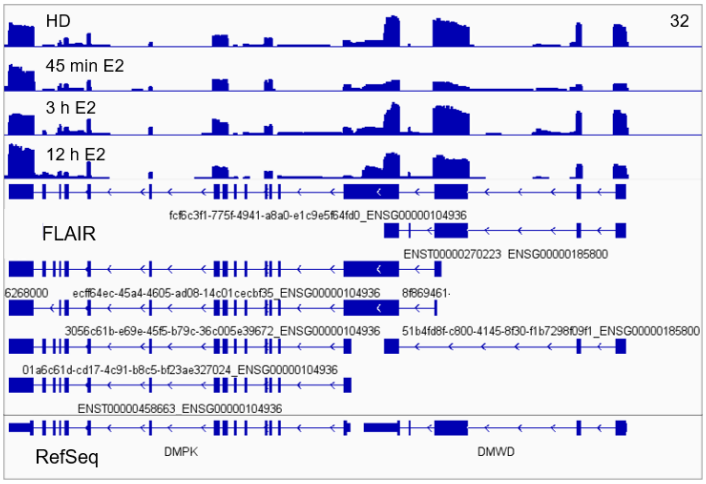
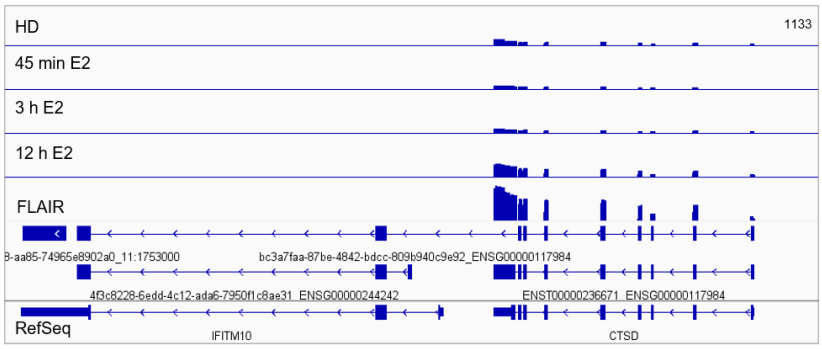


Figure 3.15 (cont'd)

B.



C.



D.

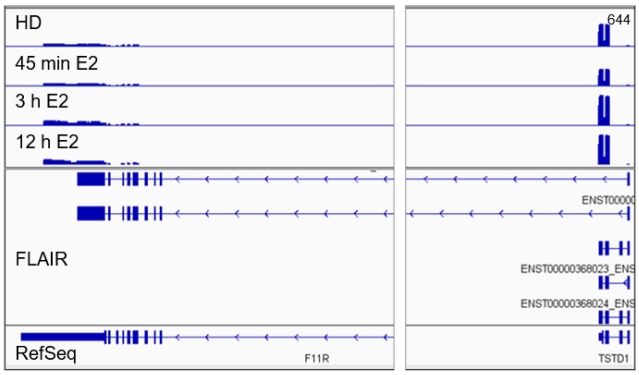


Figure 3.15. Read-through transcripts detected by DRS. This figure provides genome browser views of read-through transcription events identified using long-read ONT sequencing (FLAIR) along with the annotated RefSeq gene structure for reference. (A) *ST20-MTHFS*, (B) *DMWD-DMPK*, (C) *CTSD-IFITM10*, (D) *TSTD1-F11R* read-

through transcripts are illustrated. In all examples, the FLAIR outputs confirm the existence of single RNA molecules that physically span the intergenic region between the two genes. The value in the top-right of each track indicates the maximum read coverage.

We attempted to confirm these transcripts experimentally. However, we were unable to validate the selected transcripts, likely because such exceptionally long transcripts may not be reverse transcribed using standard cDNA synthesis protocols. Tailored techniques may be used to experimentally confirm these read-through transcripts.

3.4 Noncoding RNAs

Following the analysis of protein-coding transcript regulation, we then focused on noncoding RNAs (ncRNA). Estrogen signaling is known to interact with several ncRNAs, a part of the transcriptome now recognized as functionally significant alongside the protein-coding transcriptome [226]. To further uncover the complexity induced by E2, we focused on noncoding RNAs differentially expressed upon E2 induction.

3.4.1 Differentially Expressed ncRNAs

Among the differentially expressed ncRNAs identified using our filtering criteria (Fold change >1.5 or <0.6; TPM > 2), approximately 72% of ncRNAs were classified as lncRNAs, 20% as pseudogenes (processed), and the remaining 8% comprised other noncoding RNA types (Figure 3.16). LncRNAs are established key regulators in various cellular processes, including transcription, chromatin remodeling, and post-transcriptional gene regulation, with emerging roles in health and disease [163].

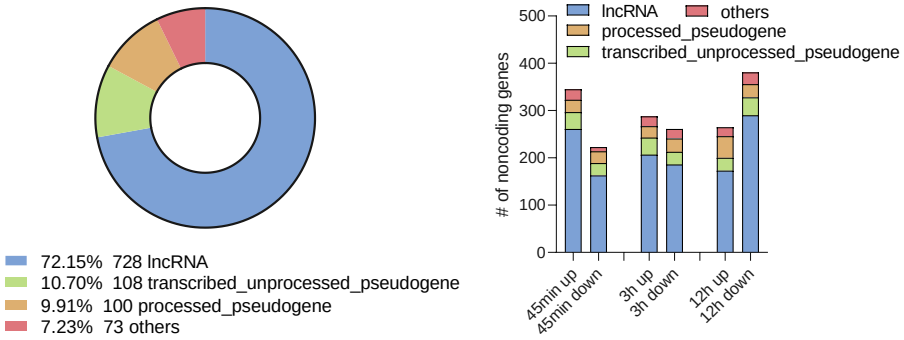


Figure 3.16. Summary of differentially expressed ncRNAs upon E2 induction. The donut chart illustrates the overall composition of the differentially expressed noncoding genes. LncRNAs represent the largest category, accounting for 72 % of the total, followed by various types of pseudogenes. The stacked bar chart shows the number of up-regulated and down-regulated noncoding genes at specific time points (45 min, 3 h, and 12 h) after E2 induction. The color-coded stacks correspond to the noncoding RNA types in the donut chart.

Similar to the trend observed for protein-coding genes, the overall pattern for differentially expressed lncRNAs (DE-lncRNA) showed an initial upregulation at 45 minutes and 3 hours post-E2 treatment, followed by an increase in the proportion of downregulated lncRNAs at the 12-hour time point (Figure 3.17).

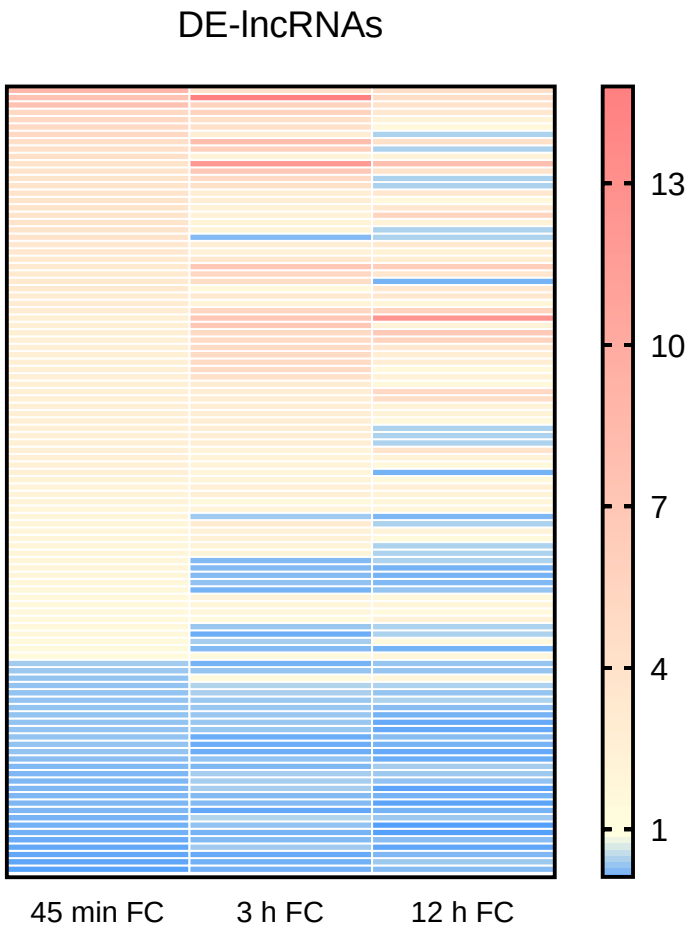


Figure 3.17. Heatmap of DE-lncRNAs upon E2 induction. This heatmap visualizes the expression patterns of lncRNAs that were differentially expressed at various time points following treatment with E2 hormone. Each horizontal row represents a single lncRNA, while the columns represent the experimental time points (e.g., 45 minutes, 3 hours, and 12 hours). The expression level is represented by the color scale on the right, where red indicates up-regulation (higher expression) and blue indicates down-regulation (lower expression) relative to a baseline control. The lncRNAs are clustered (supervised) by their expression profiles. Only the DE-lncRNAs expressed in all time-points (>2 TPM, FC <0.6 or >1.5 in all time points) (N=107).

LncRNAs are known targets of hormonal signaling pathways [188,226,263], including the E2 signaling pathway in breast cancer, where lncRNAs can both be regulated by E2 and participate in regulating the pathway itself. We observed differential expression of several lncRNAs previously linked to estrogen signaling, such as *DSCAM-AS1* and *HOTAIR* and [189,200] (Figure 3.18).

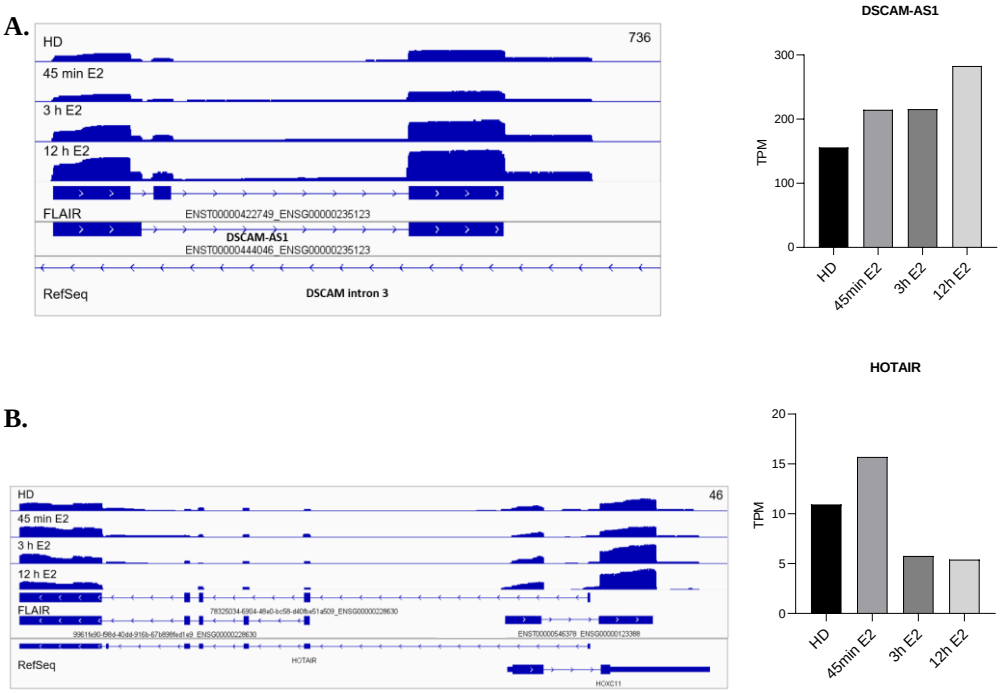


Figure 3.18. Expression profiles of known ER-regulated lncRNAs from ONT sequencing data. This figure presents the expression patterns of two well-characterized, ER-regulated lncRNAs, *DSCAM-AS1* (A) and *HOTAIR* (B). The expression levels, quantified as TPM from ONT data, were measured in a Hormone Deprived (HD) state and at 45 minutes, 3 hours, and 12 hours after treatment with (E2). Isoforms for each gene identified by the FLAIR pipeline, with the annotated RefSeq gene structure for reference. The value in the top-right of each track indicates the maximum read coverage.

3.4.2 Pathway Analysis of E2-Regulated lncRNAs

To infer the potential biological roles of the E2-regulated lncRNAs identified, functional enrichment and pathway analysis was performed using RNAenrich and TLSEA [239,240]. Across all time points, the DE-lncRNAs were associated with genes involved in biological processes such as 'regulation of apoptotic signaling pathway', 'gland development', TNF signaling, MAPK signaling, and 'epithelial cell proliferation'. A high number of These lncRNAs were linked to breast, gastric, colorectal and non-small cell lung cancers. These findings suggest that E2-regulated lncRNAs are intertwined with cell metabolism, key signaling pathways, and hormone-related developmental processes in MCF7 cells (Figure 3.19).

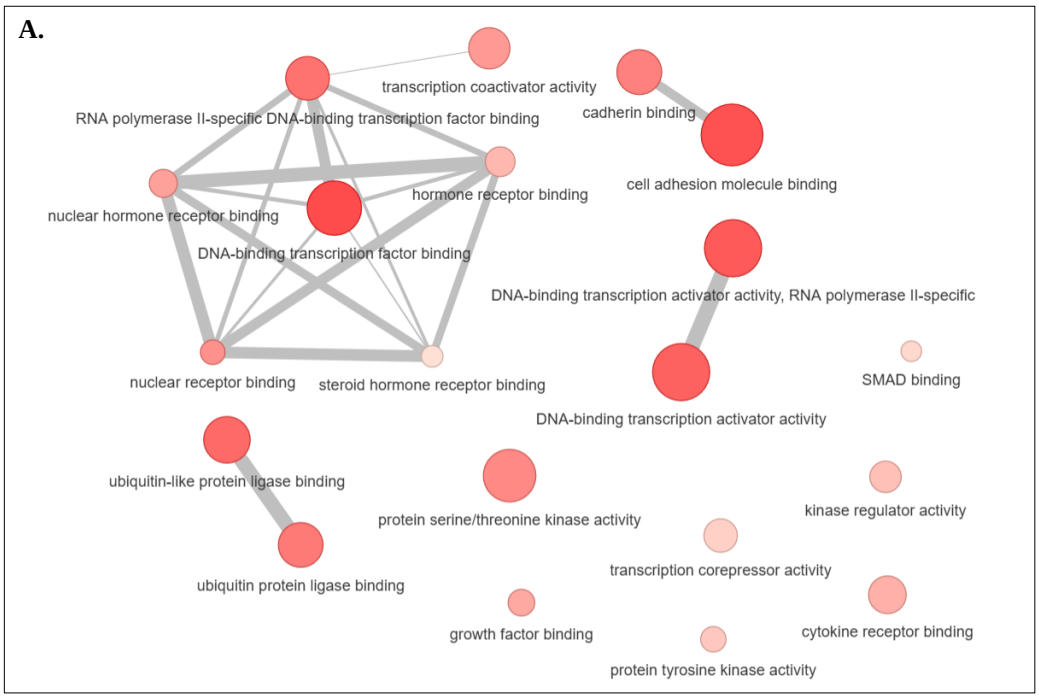


Figure 3.19 (cont'd)

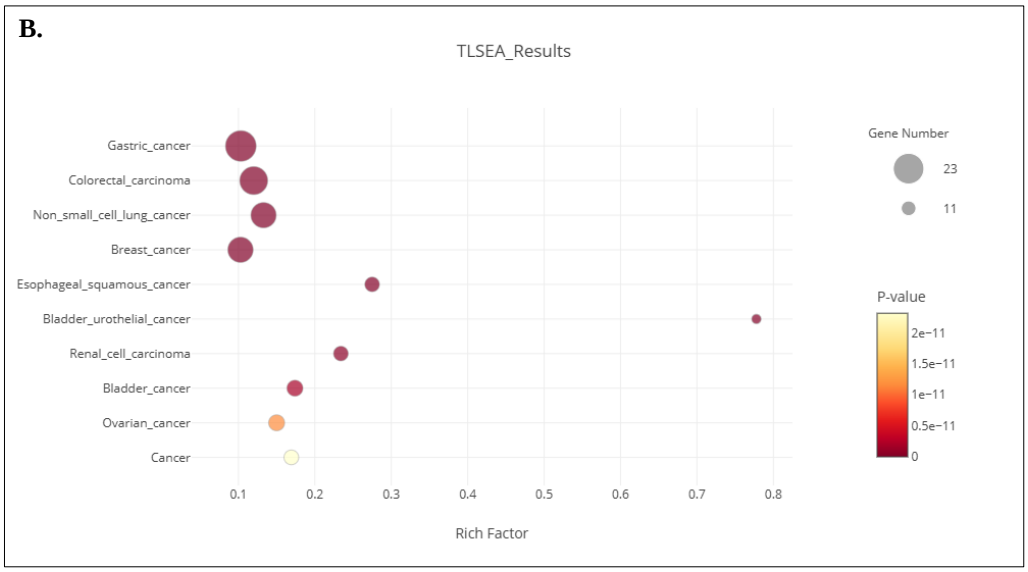


Figure 3.19. lncRNA enrichment and pathway analysis using RNAenrich and TLSEA. This figure presents two types of enrichment analyses to determine the potential functions and disease associations of a set of DE-lncRNAs. (A) Upper panel: Functional Enrichment Network generated by RNAenrich, visualizes the relationships between enriched Gene Ontology (GO) terms for molecular function. Each node represents a biological function. The size and color intensity of the node correlate with the number of genes involved and the enrichment significance. The lines (edges) connecting the nodes represent the degree of gene overlap between functions. (B) Lower Panel: Tissue-specific LncRNA Set Enrichment Analysis (TLSEA). This bubble plot shows the cancer types that are significantly associated with the input lncRNA list. Each bubble represents a specific cancer. Its position on the x-axis indicates the Rich Factor (enrichment level), its size is proportional to the number of genes from the list associated with that cancer, and its color corresponds to the statistical significance (P-value).

3.4.3 AS events in lncRNAs

Given that AS also generates lncRNA isoforms with potentially distinct functions, we specifically examined AS events within DE-lncRNAs. Similar to protein-coding genes, exon skipping was the predominant AS event type identified in lncRNAs, accounting for approximately 60% of the observed alternative processing events (Figure 3.20). This pattern agreed with recent reports on the widespread splice variants for lncRNAs [135].

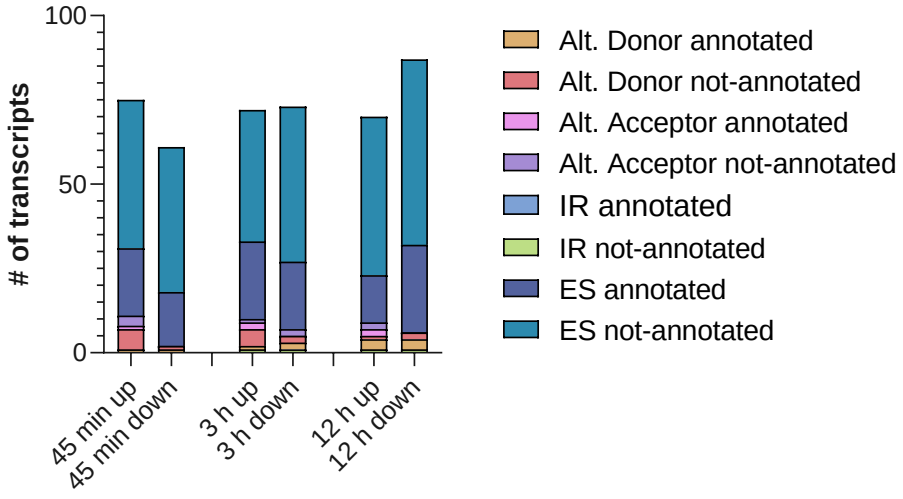


Figure 3.20. AS events in lncRNAs upon E2 treatment. This stacked bar chart quantifies the various types of AS events identified in lncRNAs. The analysis includes lncRNAs that were either up- or down-regulated at 45 minutes, 3 hours, and 12 hours following treatment with Estradiol (E2). Each colored segment within the bars corresponds to a distinct type of AS event (e.g., Exon Skipping, Intron Retention, Alternative Donor/Acceptor Sites).

Interestingly, when classifying these lncRNA isoforms using FLAIR, a substantial proportion of the detected exon skipping events (~70%) were categorized as 'unannotated', meaning they did not perfectly match existing reference annotations. While FLAIR's stringent classification considers potential sequencing errors and transcript boundaries, this high proportion of unannotated events may also reflect cell-type-specific lncRNA splicing or novel isoforms induced by E2. Further experimental validations would be needed to confirm the novelty of specific unannotated isoforms.

Overall, these findings indicate that E2 signaling impacts a wide array of lncRNAs in MCF7 cells. The diverse functions associated with these lncRNAs suggest that E2 may orchestrate complex cellular responses partly through the modulation of this noncoding layer of the transcriptome. Further studies are needed to elucidate the direct mechanisms of E2 regulation for specific lncRNAs and their precise functional roles in mediating E2 response.

3.4.4 Experimental Confirmations

3.4.4.1 Analysis of the *lnc-GATA3-7*

Examining individual lncRNA expression dynamics revealed diverse patterns, with some peaking at specific time points while others showed sustained increases or decreases (Figure 3.17). A notable example of downregulation was *lnc-GATA3-7* (LNCipedia: *lnc-GATA3-7*; ENST00000456526). Its expression continuously decreased over the time course, reaching approximately 80% reduction from its initial level by 12 hours (Figure 3.21.A). This downregulation was further confirmed by RT-qPCR in both E2-treated MCF7 cells (Figure 3.21.B). We observed *lnc-GATA3-7* expression specifically in ER-positive breast cancer cell lines (e.g., MCF7, T47D, ZR-75-1) compared to ER-negative cell lines (e.g., MDA-MB-231 and others) (Figure 3.21.C). Furthermore, analysis of public datasets confirmed its higher

expression in normal mammary tissue and Luminal A breast tumors compared to adjacent normal tissue (Figure 3.21.D). This highlights how E2 signaling can significantly alter the expression of tissue-specific lncRNAs, whose distinct isoforms may possess different functions, having need for further investigation.

Interestingly, an isoform of *lnc-GATA3-7*, previously reported as luminal-specific lncRNA (*LOL*), was described as a 1414 bp transcript in MCF7 cells [264]. Our analysis, utilizing DRS, has revealed a longer, 1740 bp isoform in MCF7 cells. This newly identified isoform shares a similar structure with *LOL* but extends an additional 56 bp at the 5' end and 271 bp at the 3' end (Appendix Figure D.4).

Based on DRS we have detected the complete, full-length isoform expressed in MCF7 cells. This was possible because direct RNA sequencing analyzes entire, full-length RNA molecules in single reads. This approach bypasses the transcript fragmentation step inherent to standard short-read sequencing, a process that severs the physical link between the 5' and 3' ends of a transcript [71]. This finding demonstrates the unique strength of direct RNA sequencing in detecting the accurate nature of transcripts.

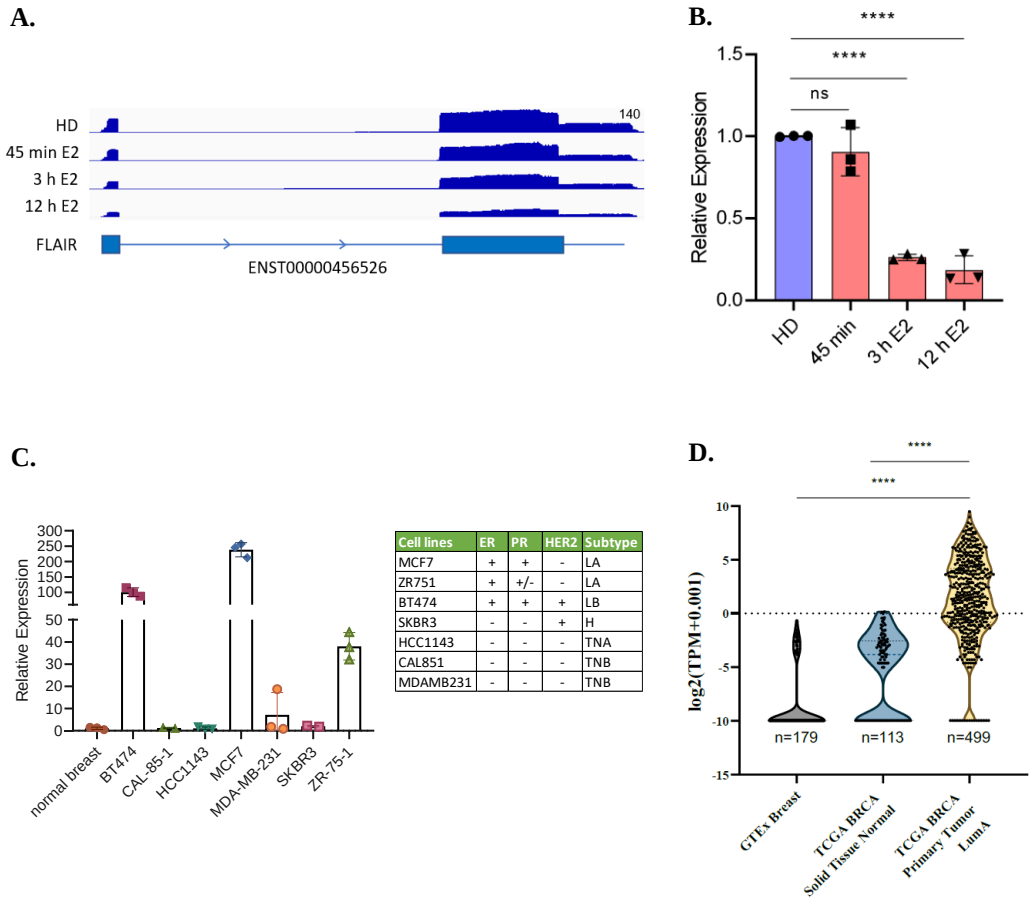


Figure 3.21. Characterization and Expression Analysis of the *lnc-GATA3-7*. (A) A genome browser view displaying DRS read coverage for the *lnc-GATA3-7* transcript following treatment with Estradiol (E2). The value in the top-right of each track indicates the maximum read coverage. (B) RT-qPCR results validating the expression pattern observed in the sequencing data. The bar chart confirms a significant down-regulation of *lnc-GATA3-7* at 3 and 12 hours after E2 treatment (**** $p < 0.0001$; ns, not significant). Experiment was repeated three times. (C) A dot plot showing the relative expression of *lnc-GATA3-7* across a panel of breast cancer cell lines compared to normal breast tissue. The adjacent table, detailing the receptor status of each cell line, shows that *lnc-GATA3-7* expression is highest in ER-positive luminal subtypes (MCF7, BT474, ZR-75-1). Experiment was repeated two times. (D) A violin plot comparing *lnc-GATA3-7* expression using data from the

GTEx (normal tissue) and The Cancer Genome Atlas (TCGA) databases. The plot demonstrates that *lnc-GATA3-7* is significantly overexpressed in the Luminal A subtype of breast cancer, compared to normal breast tissues.

3.4.4.2 Analysis of the *VMP1*/miR-21 Locus

The long-read data also allowed for the specific analysis of complex loci, such as the *VMP1* gene which hosts the microRNA miR-21 within one of its introns. While miR-21 biogenesis is often associated with the *VMP1* transcription, independent promoters for miR-21 have also been described within the intronic region of *VMP1*(Figure 3.22.A-C). Our DRS data enabled differentiation between full-length *VMP1* transcripts and distinct transcripts originating from the miR-21 locus (Figure 3.22.B). Interestingly, we observed a clear increase in level of miR-21-specific transcripts upon E2 treatment, while the host gene *VMP1* transcripts remained largely unchanged (Figure 3.22.B). Measurement of mature miR-21 levels using TaqMan assays suggested a decrease upon E2 stimulation over time (n=1) (See Figure 3.22.D).

However, attempts to validate the upregulation of the primary miR-21 transcript (*pri-miR-21*) using RT-qPCR across a detailed time course (5 min to 720 min) yielded inconsistent results between biological replicates (Figure 3.22.E). These inconsistencies could indicate a more complex regulatory network, where *pri-miR-21* expression is subject to modulation by additional factors or fluctuates in response to the dynamics of E2 treatment. Due to these conflicting results, this specific observation was not pursued further.

Nevertheless, the DRS approach successfully distinguished transcripts originating specifically from the miR-21 locus from those of its host gene, *VMP1*, demonstrating the utility of long-read data for resolving transcription at complex genomic regions.

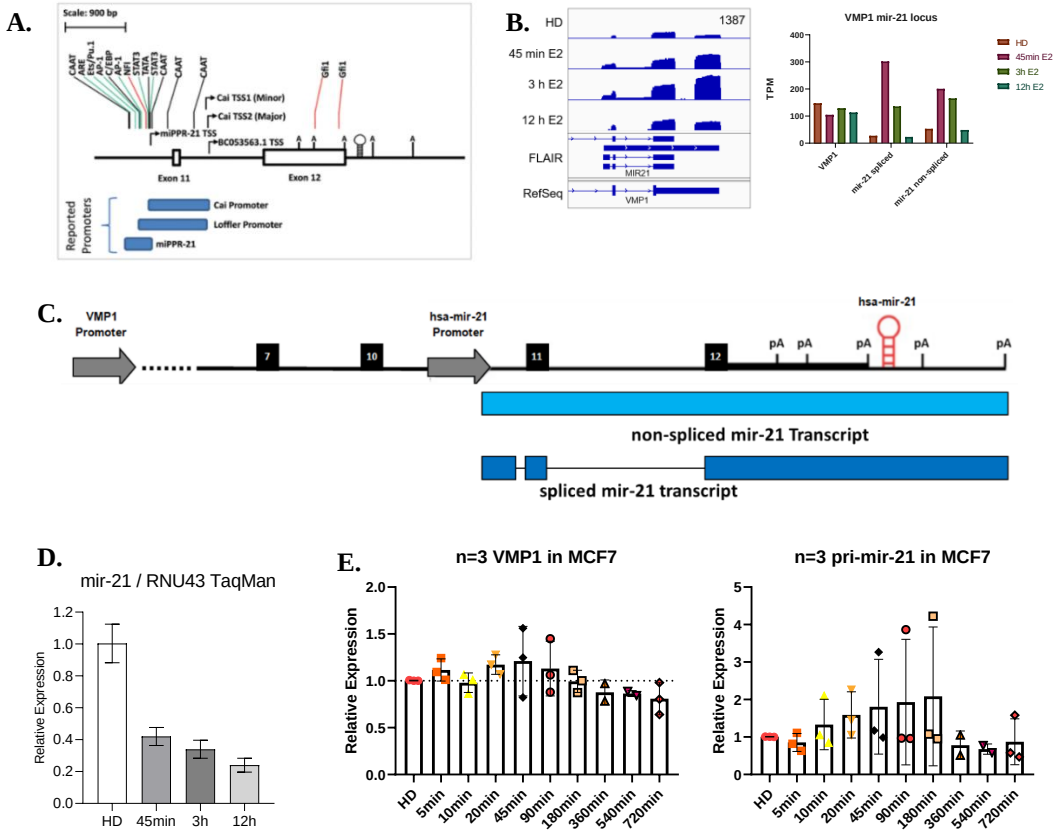


Figure 3.22. Characterization of the *VMP1*/miR-21 Locus. (A) Promoter Region Map: A schematic of the *VMP1* gene locus showing the location of the miR-21 promoters. The map highlights key transcription factor binding sites and transcription start sites (TSS) that regulate miR-21 expression [265]. (B) ONT Sequencing Results: IGV screenshot for the locus and the bar chart displays expression data in Transcripts Per Million (TPM) from DRS. It quantifies the levels of the *VMP1* gene and the different miR-21 transcript forms across a time course (HD, 45 min, 3 h, 12 h) of E2 treatment. (C) This Transcript Isoform Diagram illustrates how miR-21 is transcribed within the *VMP1*, showing both a "non-spliced" transcript and a "spliced" transcript for miR-21. (D) Mature miR-21 TaqMan qPCR assay validating the expression of mature miR-21. *RNU43* was used as reference gene (n=1). (E) *VMP1* and *pri-miR-21* qPCR: Time course analysis in MCF7 breast

cancer cells (n=3) showing the relative expression of the *VMP1* host gene (left) and the primary miR-21 transcript (*pri-miR-21*) (n=3).

3.4.4.3 Analysis of the of the SNHG Family of lncRNAs

Among the diverse lncRNAs regulated by E2, the Small Nucleolar Host Gene (SNHG) family was particularly prominent, with 20 members showing differential expression. Given the known roles of SNHGs in cancer and RNA processing, we chose to investigate this family in detail. (Figure 3.23).

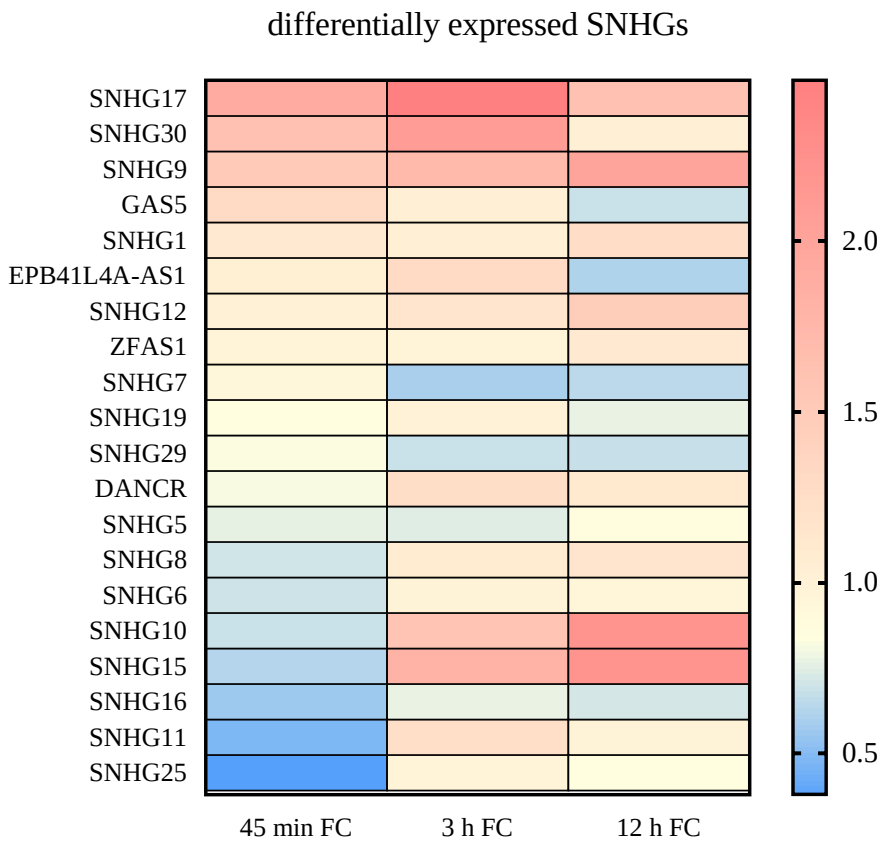


Figure 3.23. Heatmap of relative expression for detected SNHG family members. Each row corresponds to an individual SNHG, and each column represents the fold-change at the indicated time points of E2 treatment compared to HD sample. The color scale indicates the row-wise Z-score, with red representing higher expression

levels (upregulation) and blue representing lower expression levels (downregulation) relative to the HD sample. The SNHGs are clustered (supervised) by their expression profiles.

The SNHG family encompasses a diverse group of lncRNAs, initially named for their function in hosting snoRNAs involved in RNA maturation. However, subsequent research has revealed broader roles for SNHG lncRNAs in various cellular processes and cancers, acting through mechanisms such as DNA binding, epigenetic regulation, modulation of RBP activity, miRNA sponging, and direct mRNA interactions [209]. Numerous SNHG members (e.g., *SNHG1*, *SNHG5*, *SNHG6*, SNHG16, *SNHG15*, *SNHG17*, *GAS5*, *DANCR*, others) have been implicated in both tumor suppression and oncogenesis across different cancer types, including breast cancer (See Appendix Table A.2 for detailed table of SNHG functions in cancers).

The vast majority of snoRNAs are located within the introns of host genes, predominantly Small Nucleolar RNA Host Genes (SNHGs). Their biogenesis is intrinsically linked to the splicing of the host gene's pre-mRNA; after an intron is spliced out, it is processed by exonucleases to release the mature snoRNA [208]. This critical dependence on splicing prompted us to investigate whether E2 signaling could regulate snoRNA availability by modulating the splicing efficiency of their host genes. We observed that snoRNA-containing introns were consistently spliced out from all differentially expressed SNHG isoforms following E2 induction, with no evidence of intron retention. This finding suggested that E2-mediated regulation does not occur by modulating splicing efficiency. Therefore, we shifted our focus to the primary regulatory step, the transcription of the host genes themselves.

Our analysis revealed that E2 signaling alters the transcription of 19 different SNHGs. According to the snoRNA Database (snoDB 2.0) [241], these host genes harbor at least 145 unique snoRNAs, suggesting that E2 stimulation has the potential to broadly alter the cellular snoRNA landscape (Appendix Table D.1). Notably,

several of these E2-regulated SNHGs host snoRNAs with well-documented roles in breast cancer. For example, among the snoRNAs that already associated with breast cancer, there are snoRNAs that are hosted by differentially expressed SNHGs. For instance, *SNORD50A/B* hosted by the gene *SNHG5*, acts as a tumor suppressor in breast and other cancers by directly binding to and inhibiting the K-Ras oncoprotein, which in turn suppresses the oncogenic Ras-ERK1/2 signaling pathway [266]. *SNORA71*, hosted by the gene *SNHG17*, functions as an oncogene and it is. *SNORA71* promotes breast cancer metastasis and is linked to the metabolic stress response through its activation of the protein kinase RNA-activated (PKR) [267]. Lastly, *SNORD44* acts as a tumor suppressor in breast cancer and other malignancies. It is associated with its host gene, *GAS5*, and influences the mTOR signaling pathway to promote apoptosis [268].

Collectively, these findings indicate that E2-mediated transcriptional regulation of SNHGs might be a mechanism for controlling the availability of numerous cancer-associated snoRNAs, thereby potentially influencing key oncogenic and tumor-suppressive pathways.

3.4.4.3.1 E2 regulated SNHGs: *SNHG9, SNHG15, SNHG17*

Based on a combined analysis integrating the differential expression patterns upon E2 treatment (Figure 3.23), the presence of predicted EREs or ESR1 partner binding sites near their putative promoters (Appendix Figure D.5), and existing literature linking them to cancer biology [210,217,269–277]. We selected three candidates for further investigation: *SNHG9*, *SNHG15*, and *SNHG17*.

All three selected candidates, *SNHG9*, *SNHG17*, and *SNHG15*, exhibited a general trend of increased expression following sustained E2 induction (Figure 3.24.A). This overall upregulation for the candidate genes was subsequently confirmed by RT-qPCR (Figure 3.24.B).

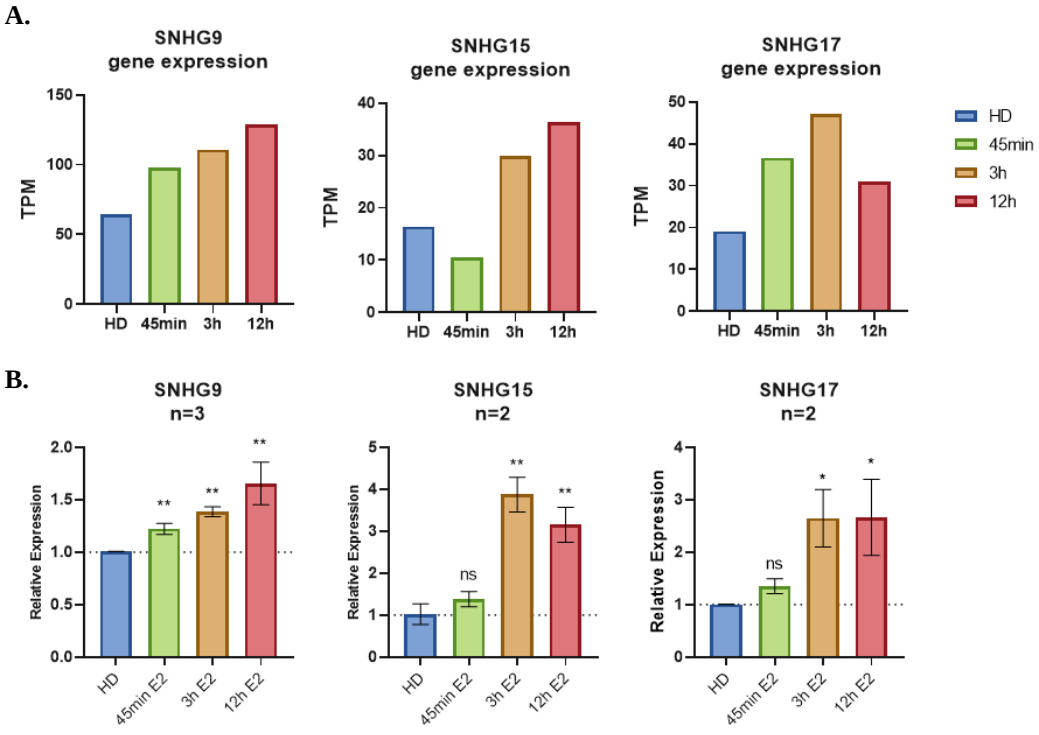


Figure 3.24. Expression analysis and validation of candidate SNHG lncRNAs. This figure shows the expression patterns of three candidate SNHG lncRNAs in response to E2 treatment. (A) Top Row: The bar charts display gene expression levels in Transcripts Per Million (TPM) as quantified by ONT. (B) Bottom Row: These charts show the validation of the expression trends using RT-qPCR. The data represents the relative gene expression (mean ± SD) from n=2 or n=3 biological replicates, confirming the patterns observed in the sequencing data. One-Way ANOVA followed by a Tukey's HSD post-hoc test to identify which specific groups differ. (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, "ns" denotes a non-significant).

3.4.4.3.1.1 *SNHG9*

While *SNHG15* and *SNHG17* have been previously studied in the context of breast cancer, implicated in various functions from RBP interactions to miRNA sponging [217,269], their regulation by E2 is not reported in the literature. *SNHG9*, conversely, has been primarily studied in other cancers (functioning as a miRNA sponge) and its role in breast cancer remained largely unexplored. The consistent upregulation of *SNHG9* upon E2 induction, its relatively simple gene structure compared to other SNHG members (Figure 3.25) and the limited existing literature on its function in breast cancer made it a particularly interesting candidate for investigating novel E2-lncRNA regulatory axes and functions in ER+ breast cancer cells.

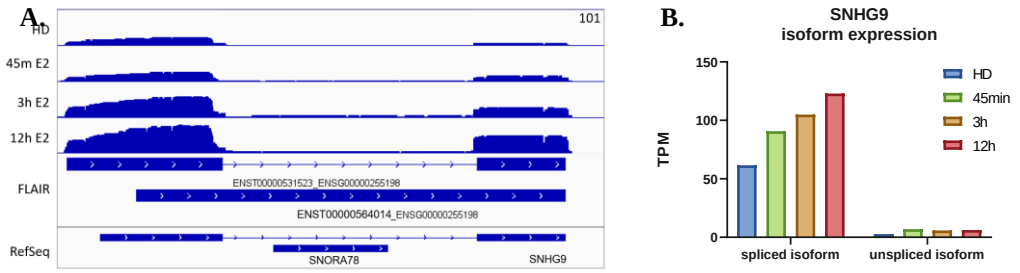


Figure 3.25. *SNHG9* isoforms. (A) IGV depiction of DRS reads and the *SNHG9* isoforms identified by FLAIR. ENST00000531523 is the spliced isoform and ENST00000564014 is the unspliced isoform. The height of the coverage tracks indicates the abundance of reads mapped to these regions. The value in the top-right of each track indicates the maximum read coverage. (B) Bar chart showing the expression levels of spliced and unspliced isoforms of *SNHG9* in Transcripts Per Million (TPM) for a control condition (HD) and at different time points (45 minutes, 3 h, and 12 h) following E2 stimulus.

As for the transcript isoforms of the *SNHG9*, there are only two isoforms (Figure 3.25). One transcript (ENST00000564014) is not spliced and the other one is a spliced isoform (ENST00000531523). Our analysis revealed that the dominant isoform in MCF7 cells is the spliced isoform (Figure 3.25.B). This distinction is

crucial, as the sequence and structural differences between isoforms can lead to distinct molecular functions, such as giving rise to snoRNAs, scaffolding for proteins, sponging miRNAs [161,225]. Despite this, many publications have studied the unspliced version as the canonical and functional isoform of the *SNHG9* gene [272,276].

SNHG9's role as a competing endogenous RNA (ceRNA) that sponges mir-214-5p was investigated in osteosarcoma, solely focusing on the unspliced isoform where the miRNA binding site is located within the exonized intron [277]. However, this sequence is absent in the spliced version we detected, meaning the spliced isoform cannot participate in this specific regulatory network in MCF7 cells. On top of that, many miRNA target prediction algorithms and genomic browsers default to the unspliced lncRNA transcript for the "*SNHG9*" locus, potentially leading to inaccurate functional predictions. This highlights the critical importance of identifying and studying the correct, cell-type-specific isoforms to accurately characterize gene function.

This finding highlights the critical importance of studying the correct, cell-type-specific isoform to accurately characterize lncRNA function. Therefore, to begin exploring the functional consequences of E2-mediated regulation of these lncRNAs, we proceeded with initial characterization experiments on the selected SNHG candidates.

3.4.4.3.2 Investigation of the candidate SNHGs

Following the identification and validation of *SNHG9*, *SNHG15*, and *SNHG17* as E2-regulated lncRNAs, we performed initial functional characterization experiments to gain insights into their potential roles in ER+ breast cancer cells.

3.4.4.3.2.1 Subcellular Localization of Candidate SNHGs

LncRNA function is often linked to subcellular localization [141]. Therefore, the subcellular localization and stability of these candidate SNHGs were next investigated to gain further insights into their potential mechanisms of action within ER+ breast cancer cells.

To predict the localization of our candidates, we utilized the DIANA-LncBase v3 localization tool [278], which suggested that *SNHG9*, *SNHG15*, and *SNHG17* could reside in both the nucleus and the cytoplasm (Appendix Table D.2). To experimentally determine their localization in MCF7 cells, we performed subcellular fractionation followed by PCR. Optimization of the fractionation protocol was confirmed by assessing the enrichment of known nuclear (*NEAT1*) [145] and mainly cytoplasmic (*DANCR*) [279] lncRNAs in the respective fractions (Figure 3.26). RT-PCR using cDNAs synthesized from cytoplasmic and nuclear RNAs revealed that *SNHG9*, *SNHG15*, and *SNHG17* transcripts were detectable in both the nuclear and cytoplasmic fractions of MCF7 cells (Figure 3.26). This dual localization suggests these lncRNAs might participate in regulatory mechanisms within both cellular compartments.

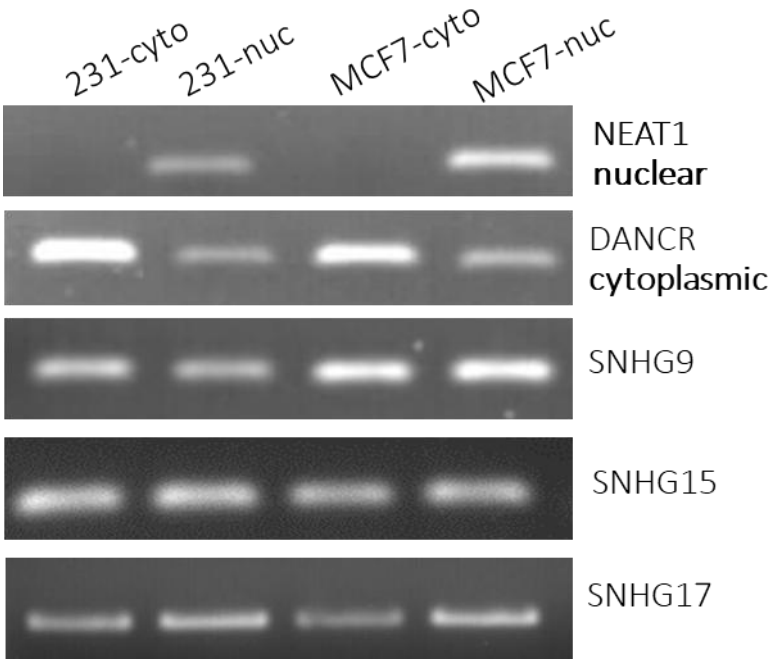


Figure 3.26. Subcellular localization of three candidate gene transcripts. Representative gel electrophoresis of PCR products from cytoplasmic and nuclear cell fractions for MDA-MB-231 and MCF7 cells. The lanes for each panel are loaded as follows: cytoplasmic fraction (cyto), nuclear fraction (nuc). *NEAT1* lncRNA was used as nuclear RNA control and *DANCR* lncRNA was used as predominantly cytoplasmic RNA control.

3.4.4.3.2.2 Transcript Stability of Candidate SNHGs

To understand the potential functional implications of these transcripts, we first determined their stability in the cell. MCF7 cells were treated with 2 µg/mL Actinomycin D (actD) to inhibit transcription [280], and RNA levels were measured over time by RT-qPCR. *SNHG9*, *SNHG15*, and *SNHG17* were found to be relatively

stable transcripts. *SNHG9* was stable in both cells. (Figure 3.27). Their stability might be relevant to their potential functionality.

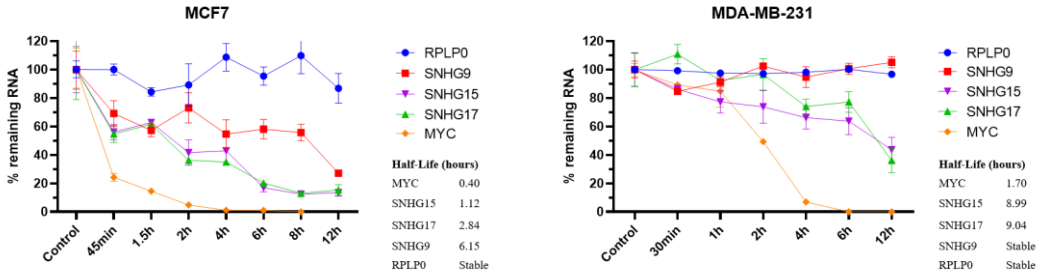


Figure 3.27. Stability of candidate SNHGs. Following transcriptional inhibition by actD (2 µg/mL for MCF7, 10 µg/mL for MDA-MB-231 [229]), the relative RNA levels of candidate SNHGs were quantified by RT-qPCR up to 12 hours. The graph displays the percentage of RNA remaining at each time point compared to the initial levels (0 hours) for each SNHG for both MDA-MB-231 and MCF7 cells. *RPLP0* was used as a stable RNA and *MYC* was used as a positive control. Half-lives (hours) were indicated next to the gene names. Data points represent the mean, and error bars indicate the standard error of the mean (SEM) derived from two independent experiments.

The stability of *SNHG9* transcript could be an indication of functionality. A stable lncRNA may function as a structural scaffold for proteins or miRNAs, a guide for epigenetic modifiers, or a decoy that needs to accumulate to be effective [281].

In contrast, an unstable lncRNA is more likely to act as a rapid signaling molecule, a quick-response element to cellular stress, or a switch for a specific phase of the cell cycle [281,282]. Consequently, its regulatory role is rapid and transient. Our positive control, mRNA of proto-oncogene *MYC*, has a very short half-life [283]. This instability is a critical feature of cellular regulation. Given that the *MYC* protein is a

powerful driver of cell proliferation, its production must be tightly controlled. By ensuring the rapid degradation of its mRNA, cells can quickly shut down MYC protein synthesis, preventing uncontrolled growth that could lead to cancer [284].

It is well known that (SNHG) family frequently functions as competing endogenous RNAs (ceRNAs), or "sponges," for various microRNAs and/or RBPs [209]. For a lncRNA to sponge miRNAs and thereby regulate the expression of their target genes, it should be abundant and stable enough to act as a regulator [285,286]. Therefore, based on both the established function of the SNHG family and the high stability we observed for *SNHG9*, we hypothesized that it could exert its function by acting as a sponge for trans-factors such as miRNAs.

3.4.4.3.2.3 Assessing the Sponge Potential of *SNHG9*

We next explored the potential for *SNHG9* to act as a sponge for trans-acting factors. *SNHG9* consists of two exons (134 [+37] bp) and (99 bp), and our FLAIR analysis indicated that the spliced isoform (lacking the intron) is the predominant form expressed in MCF7 cells (Figure 3.25). Using multiple prediction tools (e.g., miRNet, TargetScan, ENCORI), I identified potential binding sites for various miRNAs across the spliced *SNHG9* transcript (Figure 3.28.A).

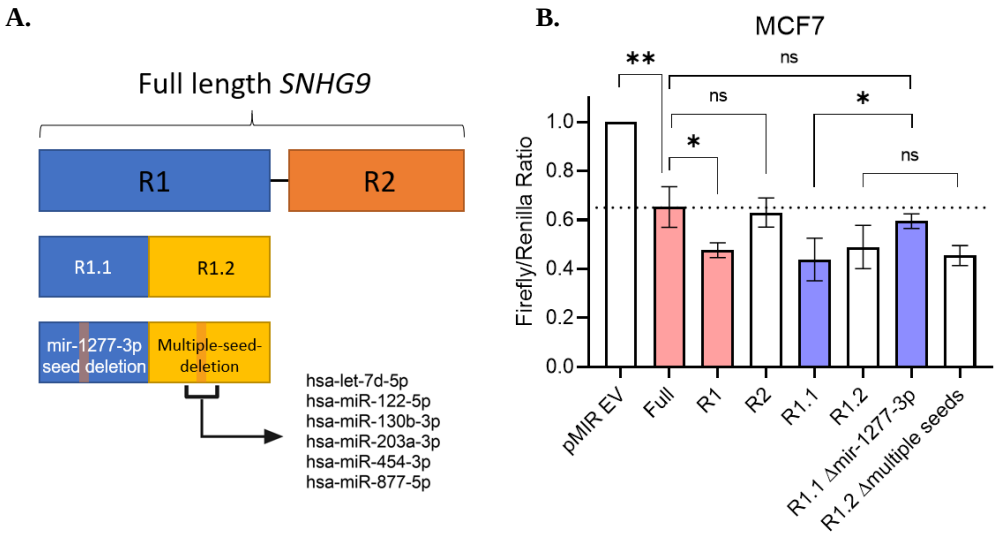
To experimentally test these interactions, I cloned the full-length *SNHG9* sequence (R1+R2), the first exon (R1), and the second exon (R2) separately into the 3' UTR of a luciferase reporter construct (Figure 3.28.A).

Transfection of these constructs into MCF7 cells resulted in decreased luciferase activity compared to the control vector, particularly for the R1 construct, suggesting potential binding of endogenous repressive molecules, likely miRNAs (Figure 3.28.B). R2 region showed the same luciferase levels as the full *SNHG9* sequence.

Previous studies in other cancers have examined interactions between *SNHG9* and various miRNAs. However, because lncRNAs like *SNHG9* exist as multiple

transcript isoforms with different sequences, validating the precise location of these binding sites is crucial. For example, mir-214-5p [276] and mir-23a-5p [272] are known to interact with the unspliced *SNHG9* transcript. In contrast, other interacting miRNAs, such as mir-199a-5p [277] and mir-150-5p [287], target the R2 region of *SNHG9*, which is common for both known *SNHG9* transcripts.

We further divided the R1 exon into two halves (R1-1, R1-2) based on miRNA prediction clusters (Figure 3.28.A). Luciferase assays showed comparable activity for R1-1 and R1-2 relative to the R1 (Figure 3.28.B).



C.

```
5' CCGCGGCCCGGGAAUCUACGUC 3' SNHG9
          |       ||||||||
3' UUUUAUGUAUAUAUAGAUGCAU 5' hsa-mir-1277-3p
```

Figure 3.28. *SNHG9* functions as a miRNA sponge for miR-1277-3p. (A) A schematic map of the *SNHG9* transcript illustrates the full-length construct ("Full") composed of regions R1 and R2, along with various deletion mutants created for the sponge assay. A predicted binding site for hsa-miR-1277-3p was identified in the first half of the R1 (R1.1) region. (B) Results from a dual-luciferase reporter assay in MCF7 cells (n=5). *SNHG9* fragments were cloned downstream of the luciferase

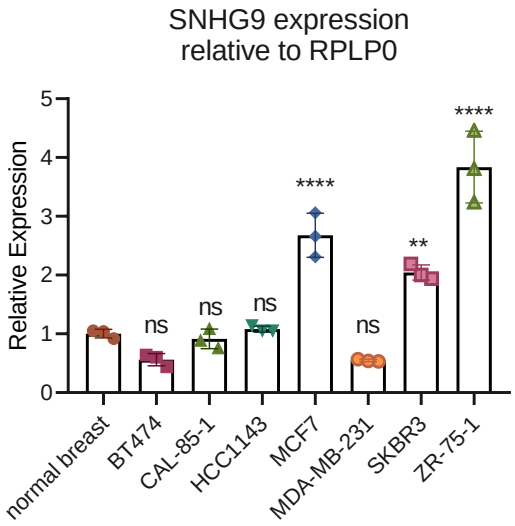
gene. A decrease in the Firefly/Renilla ratio relative to the empty vector (pMIR EV) indicates targeting by endogenous miRNAs. Transfection with the construct containing the first half of R1 (R1.1) resulted in significant repression of luciferase activity. This repression was significantly reversed upon deletion of the miR-1277-3p binding site R1.1 Δmir-1277-3p), indicating that *SNHG9* might acts as a sponge for endogenous miR-1277-3p. (C) The sequence alignment shows the predicted base-pairing between the hsa-miR-1277-3p seed region and its complementary binding site on the *SNHG9* transcript. Data are shown as mean ± SEM. One-Way ANOVA followed by a Tukey's HSD post-hoc test to identify which specific groups differ. (*$p < 0.05$, ** $p < 0.01$; ns, not significant).

To pinpoint specific interactions, we generated constructs with deletions in predicted miRNA seed regions within R1-1 (targeting the predicted miR-1277-3p site (Figure 3.28.C) and R1-2 (targeting multiple predicted sites). Interestingly, deletion of the miR-1277-3p site (R1.1 Δmir-1277-3p) in R1-first significantly increased luciferase activity compared to the intact R1-1 fragment, restoring it to levels similar to the control vector or full R1, while deletions in R1-2 had minimal effect (Figure 3.28.B). This suggests that the repressive effect observed with the R1.1 fragment was mediated, at least in part, through this specific site recognized by trans-acting factors, such as by miR-1277-3p (or other miRNAs sharing its seed sequence) (Figure 3.28.C). Verifying this potential sequestration would require more direct methods, such as the use of miRNA mimics/inhibitors or RNA pull-down assays.

Taken together, these findings present *SNHG9* as a stable, E2-induced lncRNA in MCF7 breast cancer cells with a potential mechanism of action as a sponge, possibly for mir-1277-3p. While future work is needed to elucidate the exact sponge mechanism, we turned to investigate the functional role of *SNHG9*. To address this, we sought to experimentally alter *SNHG9* expression in breast cancer cell line models.

3.4.4.3.2.4 *SNHG9* Overexpression

To investigate whether *SNHG9* expression correlates with Estrogen Receptor (ER) status, we first examined its expression in different breast cancer cell lines relative to normal breast tissue using RT-qPCR (Figure 3.29). *SNHG9* expression did not show a simple, direct correlation solely with ER status across this panel of cell lines. While high expression is seen in ER-positive Luminal A lines, it is also high in an ER-negative/HER2-positive line (SKBR3) and low in an ER-positive Luminal B line (BT474). TNBC cell lines like MDA-MB-231 did not show any upregulation.



Cell lines	ER	PR	HER2	Subtype
MCF7	+	+	-	LA
ZR751	+	+/-	-	LA
BT474	+	+	+	LB
SKBR3	-	-	+	H
HCC1143	-	-	-	TNA
CAL851	-	-	-	TNB
MDAMB231	-	-	-	TNB

Figure 3.29. *SNHG9* cell line expression and survival analysis. The bar chart displays the relative expression of the lncRNA *SNHG9* across various breast cancer cell lines (BT474, CAL-851, HCC1143, MCF7, MDA-MB-231, SKBR3, ZR-75-1) compared to normal breast tissue, with expression normalized to the reference gene *RPLP0*. The table provides details on the receptor status (ER, PR, HER2), and molecular subtype for each cell line. One-Way ANOVA followed by a Tukey's HSD post-hoc test to compare each cell line with normal breast. (**$p \leq 0.01$, ****$p \leq 0.0001$, "ns" denotes a non-significant).

Hence, to gain further insight into whether *SNHG9* levels may contribute to cancer phenotypes, we designed silencing and overexpression models. Multiple attempts to establish stable *SNHG9* knockdown using shRNA were unsuccessful due to cell death, suggesting a potential essential role for *SNHG9*. However, we successfully generated stable cell lines overexpressing *SNHG9* (*SNHG9*-OE) in two biological replicates (set-1 and set-2) of polyclonal MCF7 (LumA) cells and also in MDA-MB-231 cells (TNBC), which normally have low *SNHG9* expression. We confirmed the overexpression via RT-qPCR (Figure 3.30.A).

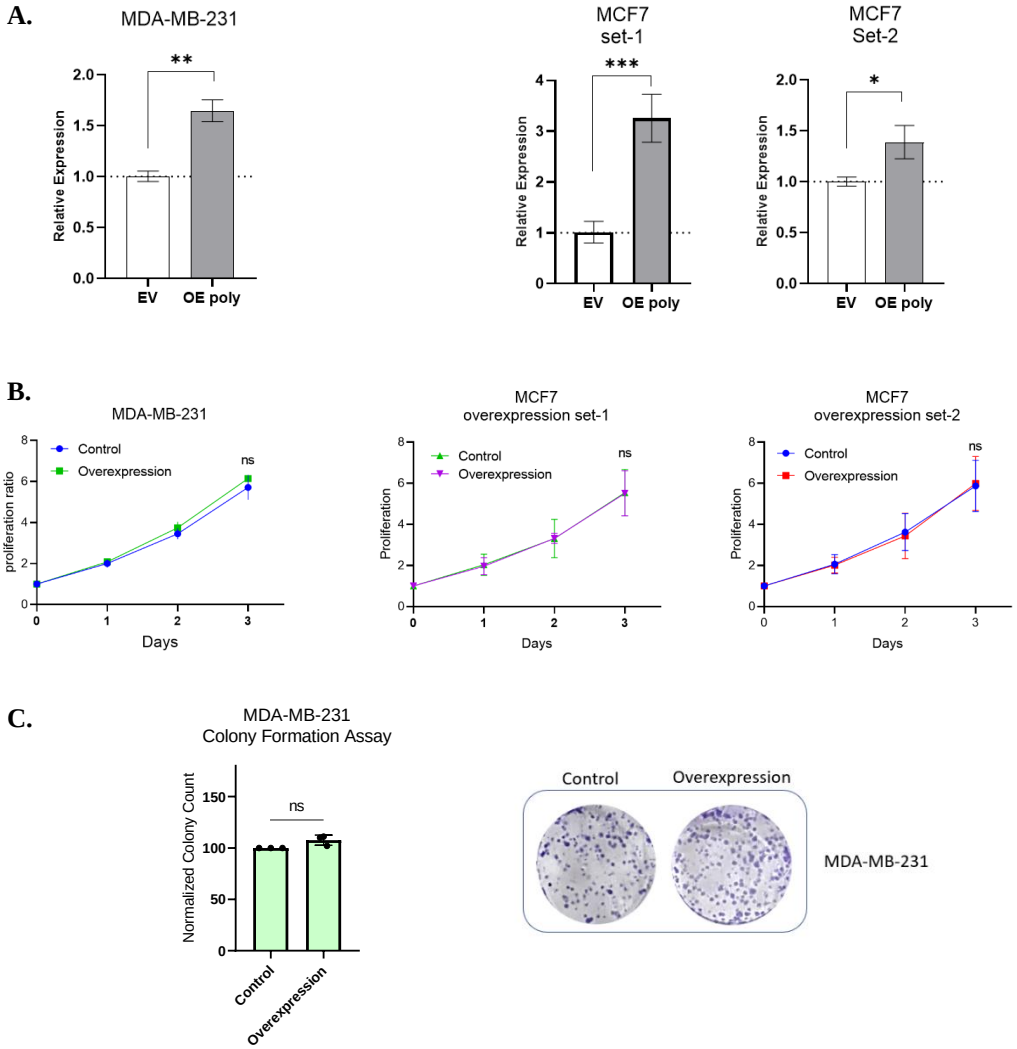


Figure 3.30 (cont'd)

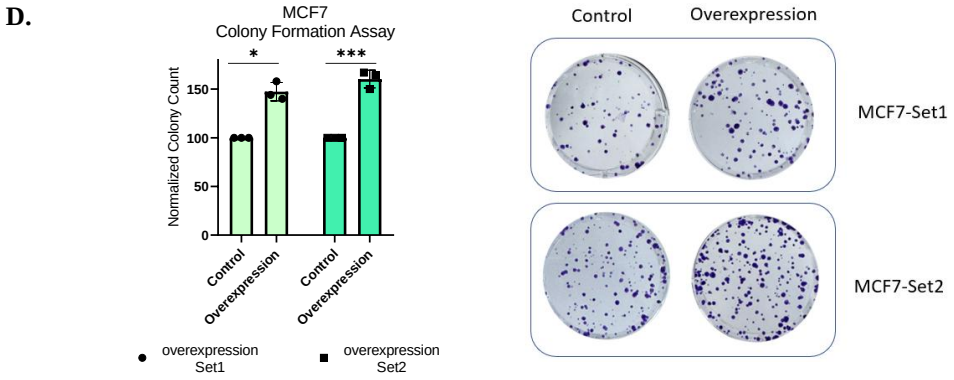


Figure 3.30. *SNHG9* overexpression, proliferation and colony formation in breast cancer cell lines. (A) Overexpression Confirmation: Relative expression of *SNHG9* was quantified by qRT-PCR in MDA-MB-231 and two independent sets of MCF7 cells (MCF7 Set-1 and Set-2). The results confirm successful *SNHG9* overexpression compared to the empty vector (EV) control in all cell lines. Data is normalized to an internal control gene, *RPLP0* (n=2). (B) Cell Proliferation Assay: The effect of *SNHG9* overexpression on cell proliferation was assessed using an MTT assay over a period of three days. Proliferation rates are compared between control cells (empty vector) and *SNHG9*-overexpressing cells (n=3). Long-term clonogenic potential was evaluated using a colony formation assay in both MDA-MB-231 (C) and MCF7 (D) cells for 10 days. The bar graphs show the normalized colony counts. The images on the right are representative pictures of the stained colonies (n=3). Data in all graphs are presented as mean ± SD from three biological replicates (n=3). Statistical significance was determined by a two-tailed Welch's t-test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, 'ns' not significant.

Under standard culture conditions, stable *SNHG9* overexpression did not significantly alter cell proliferation rates as measured by MTT assay compared to control cells (transfected with empty vector) in either MCF7 or MDA-MB-231 lines (Figure 3.30.B). While colony formation assays showed no significant difference in

MDA-MB-231 cells (Figure 3.30.C), we observed increased colony numbers in *SNHG9*-OE MCF7 cells (Figure 3.30.D).

Given that *SNHG9* expression was induced by E2, we considered its function might be context-dependent. We performed proliferation assays on *SNHG9*-OE and control MCF7 cells under hormone-deprived conditions followed by sustained E2 stimulation (up to 72 hours). As expected, E2 treatment increased proliferation in control cells. However, in the presence of E2, *SNHG9*-OE cells exhibited a significantly slower proliferation rate compared to E2-treated control cells (Figure 3.31).

This suggests that while *SNHG9* itself may not strongly drive proliferation under basal conditions, its E2-induced upregulation might participate in a negative feedback loop or pathway that dampens the overall proliferative response to sustained estrogen signaling. This could involve *SNHG9* modulating critical downstream effectors of the E2 signaling pathway or interacting with proteins involved in cell cycle checkpoint control, ultimately putting a brake on excessive proliferation [288–290].

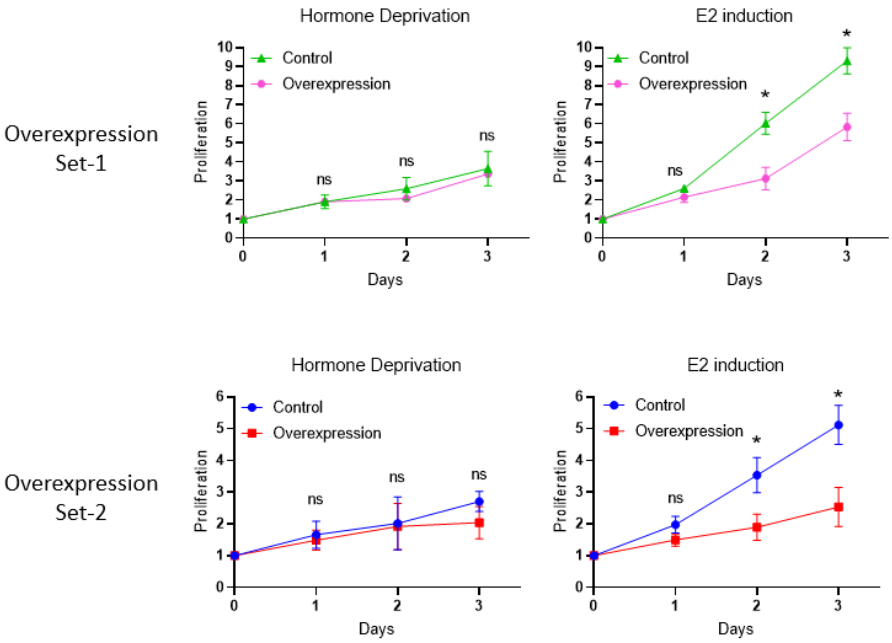


Figure 3.31. Overexpression of *SNHG9* inhibits E2-induced cell proliferation. The effect of *SNHG9* overexpression on cell proliferation was assessed using an MTT assay over a period of three days. The experiment was conducted in two independent *SNHG9* overexpression replicates (Set-1 and Set-2) in MCF7 cells. For each set, cells were tested under two conditions: hormone deprivation (left panels) and with estrogen (E2) induction (right panels). Data are presented as mean ± SEM. An asterisk (*) denotes a statistically significant difference ($p < 0.05$), while "ns" indicates a non-significant difference (n=3).

In summary, the functional exploration of the E2-induced *SNHG9* revealed that it has a context-dependent role in ER+ breast cancer cells. The phenotypic assays demonstrated that *SNHG9* overexpression exerted an inhibitory effect on E2-induced proliferation in MCF7 cells, suggesting it may function within a complex negative feedback mechanism controlling E2 responsiveness. To evaluate the clinical relevance of this hypothesis, we examined patient survival data. Our analysis using the KM plotter revealed that high *SNHG9* expression significantly correlates with

better survival rates specifically in ER-positive patients (Figure 3.32). In contrast, no significant survival difference was observed in ER-negative patients. This clinical data strongly supports our in vitro findings, suggesting that the E2-induced expression of *SNHG9*, by acting as a brake on proliferation, is a feature of less aggressive ER-positive tumors and confers a better prognosis.

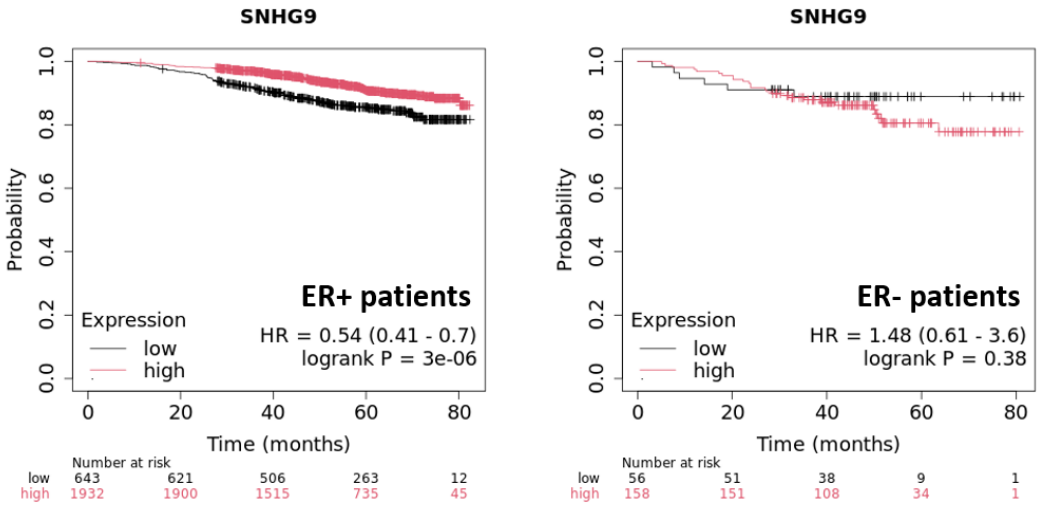


Figure 3.32. The Kaplan-Meier plots for *SNHG9*. These plots show Kaplan-Meier survival curves for breast cancer patients, stratified by high (red) and low (black) *SNHG9* expression. The left plot illustrates survival outcomes for ER-positive patients (n=2575), where high *SNHG9* expression is associated with better prognosis (HR = 0.54, p = 3e-06). The right plot shows survival outcomes for ER-negative patients (n=214), where *SNHG9* expression does not show a statistically significant association with survival (HR = 1.48, p = 0.38). Hazard ratios (HR) and log-rank p-values are presented for each cohort.

These functional results, which link *SNHG9* to the E2 signaling pathway, prompted us to investigate the transcriptional mechanism behind its E2-mediated induction.

3.4.4.3.2.5 Identification and Basal Activity of Putative SNHG Promoter Regions

E2-mediated differential expression of multiple SNHG family members suggested the possibility of a shared transcriptional control mechanism. To investigate this, we aimed to identify the uncharacterized promoter regions for our key candidate genes: *SNHG9*, *SNHG15*, and *SNHG17*.

Due to the limited annotation for lncRNA promoters in literature, we employed an integrative genomics approach to predict their locations. We identified putative promoter regions by searching for the convergence of multiple lines of evidence, including proximity to known transcription start sites (TSS), enrichment of active promoter histone marks (H3K4me3, H3K27ac), RNA Polymerase II occupancy, open chromatin signatures (DNase I hypersensitivity), and TSS mapping data from CAGE and FANTOM5 projects (See detailed images in Appendix Figure D.5). Furthermore, bioinformatic analysis of these predicted promoter regions identified potential binding sites for the estrogen receptor (EREs) and its key transcriptional partners, further supporting a potential link to E2 signaling.

Based on these predictions, we cloned the putative promoter regions for *SNHG9*, *SNHG15*, and *SNHG17* into the pGL3-Basic luciferase reporter vector. To validate their function, we performed dual-luciferase assays in MCF7 cells. The results demonstrated that all three cloned sequences exhibited significant promoter activity, driving luciferase expression at levels substantially higher than the promoterless empty vector (Figure 3.33.A). In the case of *SNHG9*, we tested several fragments and found that a broader region (P1) showed the highest activity, suggesting the presence of key regulatory elements within this sequence (Figure 3.33).

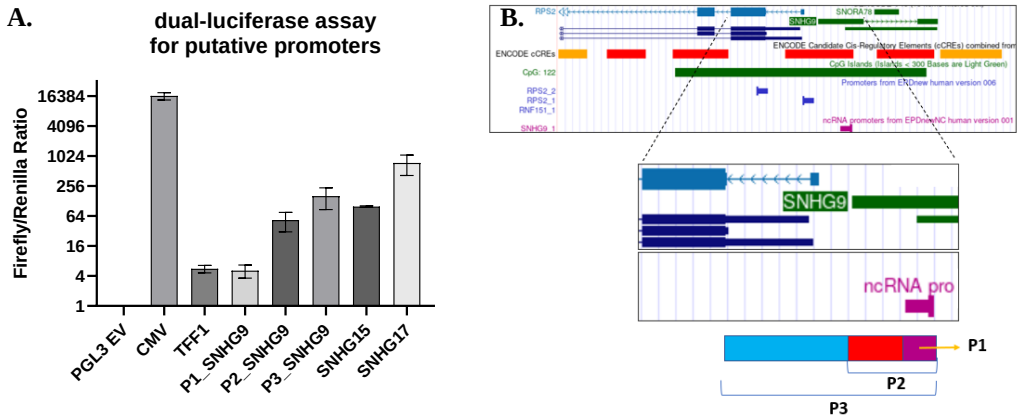


Figure 3.33. Cloned promoter regions for candidate SNHGs. To identify and characterize the promoter for *SNHG15*, SNG17 and several putative promoter regions of varying lengths ("P1", "P2", "P3") for *SNHG9* were cloned upstream of a luciferase reporter gene. (A) The bar chart (log2 scaled) shows the results of a dual-luciferase assay in MCF7 cells, quantifying the promoter activity of these constructs relative to a promoterless empty vector (PGL3 EV), a strong viral promoter (*CMV*), and an E2-responsive promoter (*TFF1*) (N=3 biological replicates for controls, N=5 for *SNHG9* promoter regions, N=2 for others). The schematic (B) illustrates the genomic context of the cloned fragments relative to the *SNHG9* locus.

These results successfully confirm that the regions we identified through bioinformatic analysis do indeed function as active promoters.

While our initial goal included testing the E2-responsiveness of these constructs, we encountered technical challenges with the luciferase assay under E2 stimulation conditions that prevented us from obtaining conclusive results. Therefore, while we have successfully identified and validated the basal promoters for these key SNHGs, determining whether E2 acts directly on these proximal regions or through more complex mechanisms will require further investigation. This work, however, provides the starting point for such future studies.

Overall, this work underscores the power of using long-read sequencing for the accurate characterization of the transcriptome. Our use of DRS allowed for the identification of the dominant, spliced isoform of *SNHG9* in MCF7 cells, a critical finding that corrected potential misconceptions from studies focused on the unspliced transcript. This accurate isoform-level resolution was fundamental to the subsequent functional assays, which suggested a novel role for *SNHG9* in modulating the E2 response pathway. In addition, the use of long-read direct RNA sequencing enabled accurate resolution of full-length transcript isoforms, capturing alternative transcription start sites, splicing events, and polyadenylation patterns. This approach provides a comprehensive view of transcript architecture and is a powerful strategy for elucidating the regulatory landscape of gene expression at isoform resolution. These findings enhance our understanding of isoform-specific regulation and their potential functional consequences at both the transcriptomic and proteomic level.

CHAPTER 4

DISCUSSION

This thesis characterized the dynamic impact of E2 on the transcriptome of the ER+ breast cancer cell line MCF7, with a specific focus on resolving transcript isoforms using DRS. Our central hypothesis was that E2 stimulation remodels the transcriptome beyond changes in gene expression. The findings presented here indicate that E2 drives a complex and dynamic post-transcriptional response affecting both coding and non-coding genes, including AS and APA. This work presents an isoform-resolved view that suggests potential regulatory mechanisms and highlights transcript isoforms that may be functionally distinct, contributing to a deeper understanding of the E2 response in breast cancer.

To ensure the reliability of our dataset, we first confirmed that our experimental system recapitulated the known global transcriptomic response to E2. DRS data was of high quality and accurately reflected established patterns of estrogen action, including the regulation of canonical target genes and the enrichment of hallmark E2-responsive gene sets. This alignment with established biology provides a solid foundation for interpreting the novel, isoform-specific findings of this study.

With the global response validated, we first narrowed our focus to the post-transcriptional regulation of protein-coding genes. A primary finding of our work is that E2 stimulation appears to induce a rapid remodeling of the splicing landscape in MCF7 cells. Our analysis of AS events reveals that exon skipping is a predominant event across all time points. These E2-induced splicing events may have notable functional implications. The observed IR event in the *CRB3* transcript is of particular interest, as *CRB3* has been linked to the suppression of cancer stem cell characteristics and the restoration of sensitivity to tamoxifen [257,291]. Similarly, the regulated ES event in the *HAGH* mRNA connects estrogen signaling directly to

cancer metabolism. *HAGH* encodes Glyoxalase 2, a critical enzyme in the detoxification of cytotoxic byproducts of glycolysis, a pathway often upregulated in cancer [292]. The functional enrichment analysis of all genes undergoing E2-regulated AS showed enrichment for terms such as 'mRNA processing' and 'nucleocytoplasmic transport', suggesting that E2 may modulate the splicing of genes whose protein products are themselves core components of the RNA processing machinery.

In addition to these AS events, our data also revealed another layer of post-transcriptional control through the differential use of polyadenylation sites. The detection of dynamically regulated IPA in response to E2 is one of the findings of this thesis. IPA can be viewed as a biological strategy for generating functionally distinct protein isoforms from a single gene locus [293,294]. We see this illustrated by the IPA events in *TLE1* and *GREB1*. In the context of ER-positive breast cancer, TLE1 has a role as a necessary co-activator of ERα [65]. Our data reveals that E2 stimulation induces an IPA event within the *TLE1* mRNA, resulting in a truncated protein lacking the C-terminal WDR domain, which is crucial for protein-protein interactions [295,296]. This truncated isoform could potentially function as a dominant-negative inhibitor of the full-length protein, suggesting a possible negative feedback loop.

A second potential feedback mechanism is suggested by our analysis of *GREB1*, a canonical E2-responsive gene whose protein product stabilizes ERα via its glycosyltransferase (GT) domain [64]. The long-read sequencing data shows that E2 also induces a GREB1 IPA isoform that completely lacks the GT domain. The loss of this ERα-stabilizing GT domain in the truncated GREB1 isoform could therefore dampen the cell's proliferative response to estrogen. However, the precise implications of these IPA events on protein function (whether they result in truncated proteins with dominant-negative effects, loss-of-function, or novel activities) require detailed further work for each individual gene and protein.

The analysis of 3'UTR-APA events further underscores the potential depth of post-transcriptional regulation. The finding that E2 induces a lengthening of the *SOS1* 3'UTR suggests a mechanism for fine-tuning this critical oncogenic pathway, while the E2-induced shortening of the *GSPT1* 3'UTR fits a well-established view where shorter 3'UTRs allow key transcripts to escape miRNA-mediated repression [118].

While these isoform-level changes in protein-coding genes highlight a significant layer of regulation, the E2-responsive landscape extends further into the non-coding genome. Indeed, the non-coding transcriptome also appears to be responsive to E2. The functional versatility of lncRNAs, enabled by their ability to fold into complex structures and interact with diverse partners, positions them as important regulators of cellular processes [161,282]. Our use of DRS allowed us to correct the annotation of known lncRNAs, such as *lnc-GATA3-7*, for which we identified a longer, more complete isoform than previously reported [264].

Our investigation of the SNHG family led us to focus on one member in particular, the lncRNA *SNHG9*. A crucial finding regarding *SNHG9* was the identification of the dominantly expressed and spliced isoform in MCF7 cells. This observation is important, as a significant portion of the existing literature on *SNHG9* has focused on the unspliced transcript, attributing its function to sponging miRNAs whose binding sites are located within the intron [272,276]. Our data suggest this previously proposed mechanism is likely not relevant in MCF7 ER-positive breast cancer cell line, highlighting the importance of identifying the correct, cell-type-specific isoform for functional studies. Functionally, we identified *SNHG9* as a stable, E2-upregulated lncRNA, and our luciferase reporter assays pointed to a potential miRNA sponge mechanism involving miR-1277-3p.

Notably, we found that stable overexpression of *SNHG9* in MCF7 cells significantly inhibited E2-induced proliferation. This key result supports a model where *SNHG9* is part of a negative feedback loop that prevents excessive proliferation in response to E2, a common regulatory strategy in biological systems [297–299]. This model is supported by clinical data, which reveals that high expression of *SNHG9* is

significantly correlated with better survival outcomes, specifically in ER-positive breast cancer patients.

The detailed findings for *SNHG9*, particularly the importance of identifying the correct isoform, exemplify the broader implications of our work. As a whole, this thesis supports a shift from a gene-centric to an isoform-centric view in the transcriptomic analysis of cancer. The data repeatedly suggests that gene-level expression measurements can be insufficient when different isoforms of the same gene have different/opposing functions [300–306].

This work suggests that the ratio of specific, functionally distinct isoforms could serve as a more nuanced prognostic biomarker than the overall expression level of a single gene. For example, the ratio of progesterone receptor isoforms A and B has been shown to be a prognostic and predictive factor in breast cancer [307]. Likewise, high ratio of the p53 isoform Δ40p53 to p53α in breast cancer cells leads to resistance to DNA-damaging chemotherapy like doxorubicin by reducing apoptosis and promoting DNA repair [308]. Similarly, the ratio of the full-length (co-activator) *TLE1* transcript to the truncated (dominant-negative) *TLE1-IPA* transcript could serve as a molecular indicator of the net transcriptional activity of the ERα complex.

It is also important that we acknowledge the limitations of this study. First, our findings are based on a single cell line, MCF7, which, while a standard model, cannot represent the full heterogeneity of ER+ breast cancer. Second, the initial DRS experiment was conducted using a single biological replicate for each time point. Future nanopore sequencing studies should aim for higher sequencing depth and coverage and include multiple cell lines and biological replicates to generate more data. The use of poly(A) enrichment for library preparation is another important consideration, as this method inherently excludes non-polyadenylated non-coding RNAs that may also play a role in the estrogen response. Consequently, the experiments presented here are suggestive, and our findings require further experimental support to fully understand the core mechanisms at play. For instance, while our data on the *VMP1*/miR-21 locus were intriguing, the inconsistent

validation results highlight the need for more detailed dynamic investigations, perhaps using more sensitive methods like digital PCR, to clarify the regulatory dynamics. Similarly, while we identified many differentially expressed lncRNAs, their direct regulation by E2 or its partners needs further work. For *SNHG9*, the potential sponge function requires more direct validation through experiments such as RNA pull-down assays to identify interacting RNA-binding proteins (RBPs), the use of miRNA mimics and inhibitors, and RBP-specific luciferase constructs. Furthermore, while our data on *SNHG9* and proliferation are compelling, additional experiments are needed to confirm the proposed negative feedback loop and investigate the exact molecular role of *SNHG9* in the E2 signaling pathway. For these findings to be translated into useful diagnostic or therapeutic tools, they must be rigorously validated and compared with data from patient-derived samples.

CHAPTER 5

CONCLUSION

In conclusion, this thesis has employed long-read direct RNA sequencing to reveal that the estrogen response in breast cancer is orchestrated through a complex and dynamic layer of post-transcriptional regulation. Our work synthesizes two major contributions into a more cohesive model of E2 action.

First, we demonstrate that E2 signaling drives widespread remodeling of the coding transcriptome through isoform-mediated mechanisms. We show that AS and APA are used to generate functionally distinct protein isoforms from key cancer-related genes, including *CRB3*, *HAGH*, *GREB1*, and *TLE1*, suggesting this is a primary strategy for fine-tuning the cellular response.

Second, we establish that this regulation extends deeply into the non-coding transcriptome. By leveraging the power of long reads, we were able to clearly detect the accurate sequence of lncRNAs.

Perhaps the most significant of these discoveries is the identification and characterization of *SNHG9* as an E2-induced lncRNA, which serves as a good example for our isoform-resolved approach. We have provided evidence that it can act like a sponge for trans-acting factors and it might function within a negative feedback loop to restrain estrogen-driven proliferation, a finding supported by clinical data showing that high *SNHG9* expression correlates with improved survival in ER+ breast cancer patients.

Ultimately, this research synthesizes these findings to underscore the necessity of an isoform-resolved perspective to fully comprehend hormone action in cancer. The regulatory circuits, functionally distinct isoforms, and corrected transcript annotations we have uncovered provide a rich resource for the new discoveries. They

generate numerous testable hypotheses that hold promise for developing new diagnostic and therapeutic strategies for ER+ breast cancer.

REFERENCES

1. Global cancer burden growing, amidst mounting need for services [Internet]. World Health Organization. 2024 [cited 2025 Jun 22]. Available from: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services

2. Kim J, Harper A, McCormack V, Sung H, Houssami N, Morgan E, et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nat Med. 2025 Apr;31(4):1154–62.

3. Arzanova E, Mayrovitz HN. The Epidemiology of Breast Cancer. In: Mayrovitz HN, editor. Breast Cancer [Internet]. Brisbane (AU): Exon Publications; 2022 [cited 2025 Jun 22]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK583819/

4. New report on global cancer burden in 2022 by world region and human development level [Internet]. 2024 [cited 2025 Jun 22]. Available from: https://www.iarc.who.int/news-events/new-report-on-global-cancer-burden-in-2022-by-world-region-and-human-development-level

5. Yara D, Oroszi T. Understanding Breast Cancer: A Comprehensive Review of Epidemiology, Risk Factors, and Treatment Strategies. Advances in Breast Cancer Research. 2025 Jan 3;14(1):1–15.

6. Breast Cancer Hormone Receptor Status | Estrogen Receptor [Internet]. [cited 2025 Jun 22]. Available from: https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-hormone-receptor-status.html

7. Schlamadinger D. Understanding the Molecular Subtypes of Breast Cancer [Internet]. Breast Cancer Research Foundation. 2025 [cited 2025 Jun 22]. Available from: https://www.bcrf.org/about-breast-cancer/molecular-subtypes-breast-cancer/

8. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000 Aug 17;406(6797):747–52.

9. Orrantia-Borunda E, Anchondo-Nuñez P, Acuña-Aguilar LE, Gómez-Valles FO, Ramírez-Valdespino CA. Subtypes of Breast Cancer. In: Mayrovitz HN, editor. Breast Cancer [Internet]. Brisbane (AU): Exon Publications; 2022 [cited 2025 Jun 22]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK583808/

10. Özdemir BC, Sflomos G, Brisken C. The challenges of modeling hormone receptor-positive breast cancer in mice. Endocr Relat Cancer. 2018 May;25(5):R319–30.

11. Yaşar P, Ayaz G, User SD, Güpür G, Muyan M. Molecular mechanism of estrogen-estrogen receptor signaling. Reprod Med Biol. 2017 Jan;16(1):4–20.

12. Berry DA, Cirrincione C, Henderson IC, Citron ML, Budman DR, Goldstein LJ, et al. Estrogen-receptor status and outcomes of modern chemotherapy for patients with node-positive breast cancer. JAMA. 2006 Apr 12;295(14):1658–67.

13. Jensen EV, Jacobson HI, Walf AA, Frye CA. Estrogen action: a historic perspective on the implications of considering alternative approaches. Physiol Behav. 2010 Feb 9;99(2):151–62.

14. Clusan L, Ferrière F, Flouriot G, Pakdel F. A Basic Review on Estrogen Receptor Signaling Pathways in Breast Cancer. Int J Mol Sci. 2023 Apr 6;24(7):6834.

15. Balfe P, McCann A, McGoldrick A, McAllister K, Kennedy M, Dervan P, et al. Estrogen receptor alpha and beta profiling in human breast cancer. Eur J Surg Oncol. 2004 Jun;30(5):469–74.

16. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Aromatase inhibitors versus tamoxifen in premenopausal women with oestrogen receptor-positive early-stage breast cancer treated with ovarian suppression: a patient-level meta-analysis of 7030 women from four randomised trials. Lancet Oncol. 2022 Mar;23(3):382–92.

17. Khan MZI, Uzair M, Nazli A, Chen JZ. An overview on Estrogen receptors signaling and its ligands in breast cancer. European Journal of Medicinal Chemistry. 2022 Nov 5;241:114658.

18. Chakraborty S, Willett H, Biswas PK. Insight into Estrogen Receptor Beta-Beta and Alpha-Beta Homo and Heterodimerization: A Combined Molecular Dynamics and Sequence Analysis Study. Biophys Chem. 2012;170:42–50.

19. Klinge CM. Estrogen receptor interaction with estrogen response elements. Nucleic Acids Res. 2001 Jul 15;29(14):2905–19.

20. Song HY, Sun W, Prabhakar S, Aung KMM, Su X. Study sequence rules of estrogen receptor α-DNA interactions using dual polarization interferometry and computational modeling. Anal Biochem. 2013 Feb 15;433(2):121–8.

21. Arao Y, Korach KS. The physiological role of estrogen receptor functional domains. Essays Biochem. 2021 Dec 17;65(6):867–75.

22. Petz LN, Nardulli AM. Sp1 binding sites and an estrogen response element half-site are involved in regulation of the human progesterone receptor A promoter. Mol Endocrinol. 2000 Jul;14(7):972–85.

23. Szwarc MM, Kommagani R, Lessey BA, Lydon JP. The p160/Steroid Receptor Coactivator Family: Potent Arbiters of Uterine Physiology and Dysfunction. Biol Reprod. 2014 Nov;91(5):122.

24. Varlakhanova N, Snyder C, Jose S, Hahm JB, Privalsky ML. Estrogen receptors recruit SMRT and N-CoR corepressors through newly recognized contacts between the corepressor N terminus and the receptor DNA binding domain. Mol Cell Biol. 2010 Mar;30(6):1434–45.

25. Altwegg KA, Vadlamudi RK. Role of estrogen receptor coregulators in endocrine resistant breast cancer. Explor Target Antitumor Ther. 2021;2(4):385–400.

26. Fiorito E, Katika MR, Hurtado A. Cooperating transcription factors mediate the function of estrogen receptor. Chromosoma. 2013;122(1):1–12.

27. Martin EM, Orlando KA, Yokobori K, Wade PA. The estrogen receptor/GATA3/FOXA1 transcriptional network: lessons learned from breast cancer. Curr Opin Struct Biol. 2021 Dec;71:65–70.

28. Kong SL, Li G, Loh SL, Sung WK, Liu ET. Cellular reprogramming by the conjoint action of ERα, FOXA1, and GATA3 to a ligand-inducible growth state. Mol Syst Biol. 2011 Aug 30;7:526.

29. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a critical determinant of Estrogen Receptor function and endocrine response. Nat Genet. 2011 Jan;43(1):27–33.

30. Pavithran H, Kumavath R. Emerging role of pioneer transcription factors in targeted ERα positive breast cancer. Explor Target Antitumor Ther. 2021;2(1):26–35.

31. Safe S, Kim K. NONCLASSICAL GENOMIC ER/Sp AND ER/AP-1 SIGNALING PATHWAYS. J Mol Endocrinol. 2008 Nov;41(5):263–75.

32. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. World J Clin Oncol. 2014 Aug 10;5(3):412–24.

33. Bourdeau V, Deschênes J, Laperrière D, Aid M, White JH, Mader S. Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. Nucleic Acids Res. 2008 Jan;36(1):76–93.

34. Fuentes N, Silveyra P. Estrogen receptor signaling mechanisms. Adv Protein Chem Struct Biol. 2019;116:135–70.

35. Jakacka M, Ito M, Weiss J, Chien PY, Gehm BD, Jameson JL. Estrogen receptor binding to DNA is not required for its activity through the nonclassical AP1 pathway. J Biol Chem. 2001 Apr 27;276(17):13615–21.

36. Heldring N, Isaacs GD, Diehl AG, Sun M, Cheung E, Ranish JA, et al. Multiple sequence-specific DNA-binding proteins mediate estrogen receptor signaling through a tethering pathway. Mol Endocrinol. 2011 Apr;25(4):564–74.

37. Paterni I, Granchi C, Katzenellenbogen JA, Minutolo F. Estrogen Receptors Alpha (ERα) and Beta (ERβ): Subtype-Selective Ligands and Clinical Potential. Steroids. 2014 Nov 15;0:13–29.

38. Maggi A. Liganded and unliganded activation of estrogen receptor and hormone replacement therapies. Biochim Biophys Acta. 2011 Aug;1812(8):1054–60.

39. Moriarty K, Kim KH, Bender JR. Minireview: estrogen receptor-mediated rapid signaling. Endocrinology. 2006 Dec;147(12):5557–63.

40. Rainville J, Pollard K, Vasudevan N. Membrane-Initiated Non-Genomic Signaling by Estrogens in the Hypothalamus: Cross-Talk with Glucocorticoids with Implications for Behavior. Front Endocrinol [Internet]. 2015 Feb 16 [cited 2025 Jun 22];6. Available from: https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2015.00018/full

41. Prossnitz ER, Barton M. The G protein-coupled oestrogen receptor GPER in health and disease: an update. Nat Rev Endocrinol. 2023 Jul;19(7):407–24.

42. Khatpe AS, Adebayo AK, Herodotou CA, Kumar B, Nakshatri H. Nexus between PI3K/AKT and Estrogen Receptor Signaling in Breast Cancer. Cancers (Basel). 2021 Jan 20;13(3):369.

43. Chu I, Arnaout A, Loiseau S, Sun J, Seth A, McMahon C, et al. Src promotes estrogen-dependent estrogen receptor α proteolysis in human breast cancer. J Clin Invest. 2007 Aug 1;117(8):2205–15.

44. Filardo EJ. Epidermal growth factor receptor (EGFR) transactivation by estrogen via the G-protein-coupled receptor, GPR30: a novel signaling pathway with potential significance for breast cancer. J Steroid Biochem Mol Biol. 2002 Feb;80(2):231–8.

45. Azuma K, Inoue S. Genomic and non-genomic actions of estrogen: recent developments. BioMolecular Concepts. 2012 Aug 1;3(4):365–70.

46. Belachew EB, Sewasew DT. Molecular Mechanisms of Endocrine Resistance in Estrogen-Receptor-Positive Breast Cancer. Front Endocrinol [Internet]. 2021 Mar 25 [cited 2025 Jun 22];12. Available from: https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2021.599586/full

47. Bernelot Moens SJ, Schnitzler GR, Nickerson M, Guo H, Ueda K, Lu Q, et al. Rapid Estrogen Receptor Signaling is Essential for the Protective Effects of Estrogen Against Vascular Injury. Circulation. 2012 Oct 16;126(16):1993–2004.

48. Li Z, Chen F, Chen L, Liu J, Tseng D, Hadi F, et al. The EstroGene2.0 database for endocrine therapy response and resistance in breast cancer. npj Breast Cancer. 2024 Dec 19;10(1):106.

49. Varju P, Chang KC, Hrabovszky E, Merchenthaler I, Liposits Z. Temporal profile of estrogen-dependent gene expression in LHRH-producing GT1-7 cells. Neurochem Int. 2009 Feb;54(2):119–34.

50. Welboren WJ, Stunnenberg HG, Sweep FCGJ, Span PN. Identifying estrogen receptor target genes. Mol Oncol. 2007 Sep;1(2):138–43.

51. Ikeda K, Horie-Inoue K, Inoue S. Identification of estrogen-responsive genes based on the DNA binding properties of estrogen receptors using high-throughput sequencing technology. Acta Pharmacol Sin. 2015 Jan;36(1):24–31.

52. Lin M, Du T, Tang X, Liao Y, Cao L, Zhang Y, et al. An estrogen response-related signature predicts response to immunotherapy in melanoma. Front Immunol. 2023;14:1109300.

53. Ovaska K, Matarese F, Grote K, Charapitsa I, Cervera A, Liu C, et al. Integrative Analysis of Deep Sequencing Data Identifies Estrogen Receptor Early Response Genes and Links ATAD3B to Poor Survival in Breast Cancer. PLOS Computational Biology. 2013 Jun 20;9(6):e1003100.

54. Frasor J, Danes JM, Komm B, Chang KCN, Lyttle CR, Katzenellenbogen BS. Profiling of Estrogen Up- and Down-Regulated Gene Expression in Human Breast Cancer Cells: Insights into Gene Networks and Pathways Underlying Estrogenic Control of Proliferation and Cell Phenotype. Endocrinology. 2003 Oct 1;144(10):4562–74.

55. Single-Cell Transcriptome Analysis Reveals Estrogen Signaling Coordinately Augments One-Carbon, Polyamine, and Purine Synthesis in Breast Cancer - ScienceDirect [Internet]. [cited 2025 Jun 23]. Available from: https://www.sciencedirect.com/science/article/pii/S2211124718317145

56. Garcia S a. B, Nagai MA. Transcriptional regulation of bidirectional gene pairs by 17-β-estradiol in MCF-7 breast cancer cells. Braz J Med Biol Res. 2011 Feb;44(2):112–22.

57. Kovats S. Estrogen receptors regulate innate immune cells and signaling pathways. Cellular Immunology. 2015 Apr 1;294(2):63–9.

58. Lu Y, Jiang S, Zhang J, Song H, Li L. [Estradiol activates MAPK signaling pathway by estrogen induced VEGF and bFGF in endometrial cancer cells]. Zhonghua Fu Chan Ke Za Zhi. 2014 Dec 1;49(12):925–31.

59. Vasan N, Toska E, Scaltriti M. Overview of the relevance of PI3K pathway in HR-positive breast cancer. Ann Oncol. 2019 Dec;30(Suppl 10):x3–11.

60. Barone I, Brusco L, Fuqua SAW. Estrogen Receptor Mutations and Changes in Downstream Gene Expression and Signaling. Clin Cancer Res. 2010 May 15;16(10):2702–8.

61. Marino M, Galluzzo P, Ascenzi P. Estrogen Signaling Multiple Pathways to Impact Gene Transcription. Curr Genomics. 2006 Nov;7(8):497–508.

62. Zhao M, Ramaswamy B. Mechanisms and therapeutic advances in the management of endocrine-resistant breast cancer. World J Clin Oncol. 2014 Aug 10;5(3):248–62.

63. Hanker AB, Sudhan DR, Arteaga CL. Overcoming Endocrine Resistance in Breast Cancer. Cancer Cell. 2020 Apr 13;37(4):496–513.

64. Shin EM, Huynh VT, Neja SA, Liu CY, Raju A, Tan K, et al. GREB1: An evolutionarily conserved protein with a glycosyltransferase domain links ERα glycosylation and stability to cancer. Sci Adv. 2021 Mar;7(12):eabe2470.

65. Prest SJ, May FEB, Westley BR. The estrogen-regulated protein, *TFF1*, stimulates migration of human breast cancer cells. FASEB J. 2002 Apr;16(6):592–4.

66. Hervouet E, Cartron PF, Jouvenot M, Delage-Mourroux R. Epigenetic regulation of estrogen signaling in breast cancer. Epigenetics. 2013 Mar 1;8(3):237–45.

67. Mann M, Cortez V, Vadlamudi RK. Epigenetics of Estrogen Receptor Signaling: Role in Hormonal Cancer Progression and Therapy. Cancers (Basel). 2011 Mar 29;3(2):1691–707.

68. Treeck O, Haerteis S, Ortmann O. Non-Coding RNAs Modulating Estrogen Signaling and Response to Endocrine Therapy in Breast Cancer. Cancers. 2023 Jan;15(6):1632.

69. Ratti M, Lampis A, Ghidini M, Salati M, Mirchev MB, Valeri N, et al. MicroRNAs (miRNAs) and Long Non-Coding RNAs (lncRNAs) as New Tools for Cancer Therapy: First Steps from Bench to Bedside. Target Oncol. 2020;15(3):261–78.

70. Klinge CM. miRNAs regulated by estrogens, tamoxifen, and endocrine disruptors and their downstream gene targets. Mol Cell Endocrinol. 2015 Dec 15;418(0 3):273–97.

71. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet. 2019 Nov;20(11):631–56.

72. Ament IH, DeBruyne N, Wang F, Lin L. Long-read RNA sequencing: A transformative technology for exploring transcriptome complexity in human diseases. Mol Ther. 2025 Mar 5;33(3):883–94.

73. Hu R, Islam MN, Varghese RS, Ressom HW. Short-Read RNA-Seq. Methods Mol Biol. 2024;2822:245–62.

74. Han SW, Jewell S, Thomas-Tikhonenko A, Barash Y. Contrasting and combining transcriptome complexity captured by short and long RNA sequencing reads. Genome Res. 2024 Oct 29;34(10):1624–35.

75. Zhang Y, Qian J, Gu C, Yang Y. Alternative splicing and cancer: a systematic review. Signal Transduct Target Ther. 2021 Feb 24;6(1):78.

76. Liu Q, Fang L, Wu C. Alternative Splicing and Isoforms: From Mechanisms to Diseases. Genes (Basel). 2022 Feb 24;13(3):401.

77. Tang AD, Soulette CM, van Baren MJ, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. Nat Commun. 2020 Mar 18;11(1):1438.

78. Leshkowitz D, Kedmi M, Fried Y, Pilzer D, Keren-Shaul H, Ainbinder E, et al. Exploring differential exon usage via short- and long-read RNA sequencing strategies. Open Biol. 12(9):220206.

79. Dong X, Tian L, Gouil Q, Kariyawasam H, Su S, De Paoli-Iseppi R, et al. The long and the short of it: unlocking nanopore long-read RNA sequencing data with short-read differential expression analysis tools. NAR Genom Bioinform. 2021 Jun;3(2):lqab028.

80. De Paoli-Iseppi R, Gleeson J, Clark MB. Isoform Age - Splice Isoform Profiling Using Long-Read Technologies. Front Mol Biosci [Internet]. 2021 Aug 2 [cited 2025 Jun 23];8. Available from:

https://www.frontiersin.org/journals/molecular-biosciences/articles/10.3389/fmolb.2021.711733/full

81. Wright DJ, Hall NAL, Irish N, Man AL, Glynn W, Mould A, et al. Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes. BMC Genomics. 2022 Jan 10;23(1):42.

82. Fahmi NA, Ahmed KT, Chang JW, Nassereddeen H, Fan D, Yong J, et al. APA-Scan: detection and visualization of 3′-UTR alternative polyadenylation with RNA-seq and 3′-end-seq data. BMC Bioinformatics. 2022 Sep 28;23(3):396.

83. van Dijk EL, Jaszczyszyn Y, Thermes C. Library preparation methods for next-generation sequencing: tone down the bias. Exp Cell Res. 2014 Mar 10;322(1):12–20.

84. Liu X, Jin W, Bao D, He T, Wang W, Li Z, et al. DIPAN: Detecting personalized intronic polyadenylation derived neoantigens from RNA sequencing data. Comput Struct Biotechnol J. 2024 Dec;23:2057–66.

85. Belchikov N, Hsu J, Li XJ, Jarroux J, Hu W, Joglekar A, et al. Understanding isoform expression by pairing long-read sequencing with single-cell and spatial transcriptomics. Genome Res. 2024 Nov;34(11):1735–46.

86. Veiga DFT, Nesta A, Zhao Y, Deslattes Mays A, Huynh R, Rossi R, et al. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. Sci Adv. 2022 Jan 21;8(3):eabg6711.

87. Lipovka Y, Konhilas JP. The complex nature of oestrogen signalling in breast cancer: enemy or ally? Biosci Rep. 2016 Jul;36(3):e00352.

88. Faltas CL, LeBron KA, Holz MK. Unconventional Estrogen Signaling in Health and Disease. Endocrinology. 2020 Mar 4;161(4):bqaa030.

89. Oehler JB, Wright H, Stark Z, Mallett AJ, Schmitz U. The application of long-read sequencing in clinical settings. Human Genomics. 2023 Aug 8;17(1):73.

90. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, et al. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nat Commun. 2019 Feb 14;10(1):754.

91. Grünberger F, Ferreira-Cerca S, Grohmann D. Nanopore sequencing of RNA and cDNA molecules in Escherichia coli. RNA. 2022 Mar;28(3):400–17.

92. Parker MT, Knop K, Sherwood AV, Schurch NJ, Mackinnon K, Gould PD, et al. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification [Internet]. eLife. eLife Sciences

Publications Limited; 2020 [cited 2025 Jun 22]. Available from: https://elifesciences.org/articles/49658/peer-reviews

93. RNA sequencing [Internet]. PacBio. [cited 2025 Jun 25]. Available from: https://www.pacb.com/products-and-services/applications/rna-sequencing/

94. Polenkowski M, Allister AB, Burbano de Lara S, Soltau M, Kendre G, Tran DDH. Mapping alternative polyadenylation in human cells using direct RNA sequencing technology. STAR Protoc. 2023 Sep 15;4(3):102420.

95. Davidson NM, Chen Y, Sadras T, Ryland GL, Blombery P, Ekert PG, et al. JAFFAL: detecting fusion genes with long-read transcriptome sequencing. Genome Biology. 2022 Jan 6;23(1):10.

96. Frontiers | Allele-specific quantification of human leukocyte antigen transcript isoforms by nanopore sequencing [Internet]. [cited 2025 Jun 23]. Available from: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2023.1199618/full

97. Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. Nat Biotechnol. 2016 May 6;34(5):518–24.

98. Liu-Wei W, van der Toorn W, Bohn P, Hölzer M, Smyth RP, von Kleist M. Sequencing accuracy and systematic errors of nanopore direct RNA sequencing. BMC Genomics. 2024 May 28;25(1):528.

99. Acera Mateos P, J Sethi A, Ravindran A, Srivastava A, Woodward K, Mahmud S, et al. Prediction of m6A and m5C at single-molecule resolution reveals a transcriptome-wide co-occurrence of RNA modifications. Nat Commun. 2024 May 9;15(1):3899.

100. Direct RNA sequencing (SQK-RNA004) [Internet]. Oxford Nanopore Technologies. 2023 [cited 2025 Jun 25]. Available from: https://nanoporetech.com/document/direct-rna-sequencing-sqk-rna004

101. Chiou CS, Chen BH, Wang YW, Kuo NT, Chang CH, Huang YT. Correcting modification-mediated errors in nanopore sequencing by nucleotide demodification and reference-based correction. Commun Biol. 2023 Nov 29;6:1215.

102. Ermini L, Driguez P. The Application of Long-Read Sequencing to Cancer. Cancers. 2024 Jan;16(7):1275.

103. O'Neill K, Pleasance E, Fan J, Akbari V, Chang G, Dixon K, et al. Long-read sequencing of an advanced cancer cohort resolves rearrangements, unravels

haplotypes, and reveals methylation landscapes. Cell Genomics. 2024 Nov 13;4(11):100674.

104. Sakamoto Y, Sereewattanawoot S, Suzuki A. A new era of long-read sequencing for cancer genomics. J Hum Genet. 2020 Jan;65(1):3–10.

105. Vrtačnik P, Ostanek B, Mencej-Bedrač S, Marc J. The many faces of estrogen signaling. Biochem Med (Zagreb). 2014 Oct 15;24(3):329–42.

106. Sengupta S, Schiff R, Katzenellenbogen BS. Post-Transcriptional Regulation of Chemokine Receptor CXCR4 by Estrogen in HER2 Overexpressing, Estrogen Receptor-Positive Breast Cancer Cells. Breast Cancer Res Treat. 2009 Sep;117(2):243–51.

107. Ors A, Chitsazan AD, Doe AR, Mulqueen RM, Ak C, Wen Y, et al. Estrogen regulates divergent transcriptional and epigenetic cell states in breast cancer. Nucleic Acids Research. 2022 Nov 11;50(20):11492–508.

108. Shults CL, Dingwall CB, Kim CK, Pinceti E, Rao YS, Pak TR. 17β-estradiol regulates the RNA-binding protein Nova1, which then regulates the alternative splicing of estrogen receptor β in the aging female rat brain. Neurobiology of Aging. 2018 Jan 1;61:13–22.

109. Wright CJ, Smith CWJ, Jiggins CD. Alternative splicing as a source of phenotypic diversity. Nat Rev Genet. 2022 Nov;23(11):697–710.

110. Elhasnaoui J, Ferrero G, Miano V, Cutrupi S, De Bortoli M. The Estrogen Receptor α Signaling Pathway Controls Alternative Splicing in the Absence of Ligands in Breast Cancer Cells. Cancers (Basel). 2021 Dec 13;13(24):6261.

111. Marasco LE, Kornblihtt AR. The physiology of alternative splicing. Nat Rev Mol Cell Biol. 2023 Apr;24(4):242–54.

112. Dou Z, Zhao D, Chen X, Xu C, Jin X, Zhang X, et al. Aberrant Bcl-x splicing in cancer: from molecular mechanism to therapeutic modulation. J Exp Clin Cancer Res. 2021 Jun 12;40:194.

113. Dougherty GJ, Cooper DL, Memory JF, Chiu RK. Ligand binding specificity of alternatively spliced CD44 isoforms. Recognition and binding of hyaluronan by CD44R1. J Biol Chem. 1994 Mar 25;269(12):9074–8.

114. Epstein RJ, Tian LJ, Gu YF. 2b or Not 2b: How Opposing FGF Receptor Splice Variants Are Blocking Progress in Precision Oncology. J Oncol. 2021;2021:9955456.

115. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol. 2017 Jan;18(1):18–30.

116. Zhang Q, Tian B. The emerging theme of 3'UTR mRNA isoform regulation in reprogramming of cell metabolism. Biochem Soc Trans. 2023 Jun 28;51(3):1111–9.

117. Zhang Y, Huang Z, Lu W, Liu Z. Alternative polyadenylation in cancer: Molecular mechanisms and clinical application. Crit Rev Oncol Hematol. 2025 Feb;206:104599.

118. Mayr C, Bartel DP. Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. Cell. 2009 Aug 21;138(4):673–84.

119. Akman BH, Can T, Erson-Bensan AE. Estrogen-induced upregulation and 3'-UTR shortening of CDC6. Nucleic Acids Res. 2012 Nov;40(21):10679–88.

120. Tian B, Pan Z, Lee JY. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res. 2007 Feb;17(2):156–65.

121. Cheng X, Jiang G, Zhou X, Wang J, Zhao Z, Zhang J, et al. The landscape and clinical relevance of intronic polyadenylation in human cancers. J Genet Genomics. 2024 Oct;51(10):1030–9.

122. Koedoot E, Wolters L, van de Water B, Le Dévédec SE. Splicing regulatory factors in breast cancer hallmarks and disease progression. Oncotarget. 2019 Oct 15;10(57):6021–37.

123. Lal S, Allan A, Markovic D, Walker R, Macartney J, Europe-Finner N, et al. Estrogen alters the splicing of type 1 corticotropin-releasing hormone receptor in breast cancer cells. Sci Signal. 2013 Jul 2;6(282):ra53.

124. Grammatopoulos DK. CRH-R splicing in estrogen-sensitive breast cancer. Cell Cycle. 2014 Mar 1;13(5):687–8.

125. Rehfeld A, Plass M, Krogh A, Friis-Hansen L. Alterations in Polyadenylation and Its Implications for Endocrine Disease. Front Endocrinol [Internet]. 2013 May 8 [cited 2025 Jun 23];4. Available from: https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2013.00053/full

126. Decoding noncoding RNAs. Nat Methods. 2022 Oct;19(10):1147–8.

127. Chi Y, Wang D, Wang J, Yu W, Yang J. Long Non-Coding RNA in the Pathogenesis of Cancers. Cells. 2019 Sep 1;8(9):1015.

128. Nadhan R, Isidoro C, Song YS, Dhanasekaran DN. Signaling by LncRNAs: Structure, Cellular Homeostasis, and Disease Pathology. Cells. 2022 Aug 13;11(16):2517.

129. Knoll M, Lodish HF, Sun L. Long non-coding RNAs as regulators of the endocrine system. Nat Rev Endocrinol. 2015 Mar;11(3):151–60.

130. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol. 2023 Jun;24(6):430–47.

131. Tsagakis I, Douka K, Birds I, Aspden JL. Long non-coding RNAs in development and disease: conservation to mechanisms. J Pathol. 2020 Apr;250(5):480–95.

132. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. RNA Biol. 2013 Jun 1;10(6):924–33.

133. Wilusz JE. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. Biochim Biophys Acta. 2016 Jan;1859(1):128–38.

134. Xu SM, Curry-Hyde A, Sytnyk V, Janitz M. RNA polyadenylation patterns in the human transcriptome. Gene. 2022 Mar 30;816:146133.

135. Khan MR, Wellinger RJ, Laurent B. Exploring the Alternative Splicing of Long Noncoding RNAs. Trends in Genetics. 2021 Aug 1;37(8):695–8.

136. Chen J, Liu Y, Min J, Wang H, Li F, Xu C, et al. Alternative splicing of lncRNAs in human diseases. Am J Cancer Res. 2021 Mar 1;11(3):624–39.

137. Mihaljevic A, Rubin PD, Chouvardas P, Esposito R. Cell type specific long non-coding RNA targets identified by integrative analysis of single-cell and bulk colorectal cancer transcriptomes. Sci Rep. 2024 May 13;14(1):10939.

138. Bjørklund SS, Aure MR, Häkkinen J, Vallon-Christersson J, Kumar S, Evensen KB, et al. Subtype and cell type specific expression of lncRNAs provide insight into breast cancer. Commun Biol. 2022 Aug 18;5(1):834.

139. Johnsson P, Lipovich L, Grandér D, Morris KV. Evolutionary conservation of long noncoding RNAs; sequence, structure, function. Biochim Biophys Acta. 2014 Mar;1840(3):1063–71.

140. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021 Feb;22(2):96–118.

141. Bridges MC, Daulagala AC, Kourtidis A. LNCcation: lncRNA localization and function. Journal of Cell Biology. 2021 Jan 19;220(2):e202009045.

142. Salehi S, Taheri MN, Azarpira N, Zare A, Behzad-Behbahani A. State of the art technologies to explore long non-coding RNAs in cancer. J Cell Mol Med. 2017 Dec;21(12):3120–40.

143. Bhan A, Mandal SS. LncRNA *HOTAIR*: a master regulator of chromatin dynamics and cancer. Biochim Biophys Acta. 2015 Aug;1856(1):151–64.

144. Shen WJ, Zhang F, Zhao X, Xu J. LncRNAs and Esophageal Squamous Cell Carcinoma - Implications for Pathogenesis and Drug Development. J Cancer. 2016 Jun 23;7(10):1258–64.

145. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. An Architectural Role for a Nuclear Non-coding RNA: *NEAT1* RNA is Essential for the Structure of Paraspeckles. Mol Cell. 2009 Mar 27;33(6):717–26.

146. An H, Williams NG, Shelkovnikova TA. *NEAT1* and paraspeckles in neurodegenerative diseases: A missing lnc found? Noncoding RNA Res. 2018 Nov 15;3(4):243–52.

147. Curci D, Stankovic B, Kotur N, Pugnetti L, Gasic V, Romano M, et al. The long non-coding RNA *GAS5* contributes to the suppression of inflammatory responses by inhibiting NF-κB activity. Front Pharmacol. 2024 Oct 9;15:1448136.

148. Sze S, Bhardwaj A, Fnu P, Azarm K, Mund R, Ring K, et al. TERRA R-loops connect and protect sister telomeres in mitosis. Cell Rep. 2023 Oct 31;42(10):113235.

149. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, et al. The Nuclear-Retained Noncoding RNA *MALAT1* Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. Mol Cell. 2010 Sep 24;39(6):925–38.

150. Xu J, Xu J, Liu X, Jiang J. The role of lncRNA-mediated ceRNA regulatory networks in pancreatic cancer. Cell Death Discov. 2022 Jun 14;8(1):287.

151. Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, et al. The Imprinted *H19* LncRNA Antagonizes Let-7 MicroRNAs. Mol Cell. 2013 Oct 10;52(1):10.1016/j.molcel.2013.08.027.

152. Zhang N, Wen K. The role of lncRNA binding to RNA-binding proteins to regulate mRNA stability in cancer progression and drug resistance mechanisms (Review). Oncology Reports. 2024 Nov 1;52(5):1–12.

153. Sebastian-delaCruz M, Gonzalez-Moro I, Olazagoitia-Garmendia A, Castellanos-Rubio A, Santin I. The Role of lncRNAs in Gene Expression Regulation through mRNA Stabilization. Noncoding RNA. 2021 Jan 5;7(1):3.

154. Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. Genome Biol. 2010;11(5):R56.

155. Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, et al. LincRNA-p21 suppresses target mRNA translation. Mol Cell. 2012 Aug 24;47(4):648–55.

156. Hu G, Lou Z, Gupta M. The long non-coding RNA *GAS5* cooperates with the eukaryotic translation initiation factor 4E to regulate c-*MYC* translation. PLoS One. 2014;9(9):e107016.

157. Noh JH, Kim KM, McClusky WG, Abdelmohsen K, Gorospe M. Cytoplasmic functions of long noncoding RNAs. Wiley Interdiscip Rev RNA. 2018 May;9(3):e1471.

158. Yang S, Manna C, Manna PR. Harnessing the Role of ESR1 in Breast Cancer: Correlation with microRNA, lncRNA, and Methylation. Int J Mol Sci. 2025 Mar 27;26(7):3101.

159. Lin A, Li C, Xing Z, Hu Q, Liang K, Han L, et al. The LINK-A lncRNA activates normoxic HIF1α signalling in triple-negative breast cancer. Nat Cell Biol. 2016 Feb;18(2):213–24.

160. Yin Q, Ma H, Bamunuarachchi G, Zheng X, Ma Y. Long Non-Coding RNAs, Cell Cycle, and Human Breast Cancer. Hum Gene Ther. 2023 Jun 1;34(11–12):481–94.

161. Ahadi A. Functional roles of lncRNAs in the pathogenesis and progression of cancer. Genes Dis. 2020 Apr 21;8(4):424–37.

162. Wang Y, Bu N, Luan X fei, Song Q qian, Ma BF, Hao W, et al. Harnessing the potential of long non-coding RNAs in breast cancer: from etiology to treatment resistance and clinical applications. Front Oncol [Internet]. 2024 Mar 5 [cited 2025 Jun 23];14. Available from: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2024.1337579/full

163. Jin H, Du W, Huang W, Yan J, Tang Q, Chen Y, et al. lncRNA and breast cancer: Progress from identifying mechanisms to challenges and opportunities of clinical treatment. Mol Ther Nucleic Acids. 2021 Aug 19;25:613–37.

164. Song X, Liu S, Zeng Y, Cai Y, Luo H. BANCR-containing extracellular vesicles enhance breast cancer resistance. J Biol Chem. 2025 Feb 11;301(4):108304.

165. Yi D, Wang Z, Yang H, Wang R, Shi X, Liu Z, et al. Long non-coding RNA *MEG3* acts as a suppressor in breast cancer by regulating miR-330/CNN1. Aging (Albany NY). 2024 Jan 17;16(2):1318–35.

166. Shi Y, Huang Q, Kong X, Zhao R, Chen X, Zhai Y, et al. Current Knowledge of Long Non-Coding RNA *HOTAIR* in Breast Cancer Progression and Its Application. Life. 2021 Jun;11(6):483.

167. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long noncoding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. Nature. 2010 Apr 15;464(7291):1071–6.

168. Zhang M, Wu K, Zhang P, Qiu Y, Bai F, Chen H. *HOTAIR* Facilitates Endocrine Resistance in Breast Cancer Through ESR1/miR-130b-3p Axis: Comprehensive Analysis of mRNA-miRNA-lncRNA Network. Int J Gen Med. 2021 Aug 18;14:4653–63.

169. Ökmen D, Pavlopoulou A, Toy Hİ, Ökmen D, Pavlopoulou A, Toy Hİ. *HOTAIR* as a Prognostic Predictor for Multiple Human Malignant Diseases: A Systematic Review and Meta-Analysis. Thoracic Research and Practice [Internet]. 2019 Aug 5 [cited 2025 Jun 23]; Available from: https://www.thoracrespract.org/articles/*HOTAIR*-as-a-prognostic-predictor-for-multiple-human-malignant-diseases-a-systematic-review-and-meta-analysis/doi/TurkThoracJ.2019.190

170. Wang N, Cao S, Wang X, Zhang L, Yuan H, Ma X. lncRNA *MALAT1*/miR-26a/26b/ST8SIA4 axis mediates cell invasion and migration in breast cancer cell lines. Oncology Reports. 2021 Aug 1;46(2):1–12.

171. Shao J, Zhang Q, Wang P, Wang Z. LncRNA *MALAT1* promotes breast cancer progression by sponging miR101-3p to mediate mTOR/PKM2 signal transmission. Am J Transl Res. 2021 Sep 15;13(9):10262–75.

172. Zhou Q, Liu L, Zhou J, Chen Y, Xie D, Yao Y, et al. Novel Insights Into *MALAT1* Function as a MicroRNA Sponge in NSCLC. Front Oncol. 2021 Oct 27;11:758653.

173. Zhang X, Luo M, Zhang J, Guo B, Singh S, Lin X, et al. The role of lncRNA *H19* in tumorigenesis and drug resistance of human Cancers. Front Genet [Internet]. 2022 Sep 27 [cited 2025 Jun 24];13. Available from: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.1005522/full

174. Matouk IJ, Halle D, Raveh E, Gilon M, Sorin V, Hochberg A. The role of the oncofetal *H19* lncRNA in tumor metastasis: orchestrating the EMT-MET decision. Oncotarget. 2015 Nov 25;7(4):3748–65.

175. Lecerf C, Le Bourhis X, Adriaenssens E. The long non-coding RNA *H19*: an active player with multiple facets to sustain the hallmarks of cancer. Cell Mol Life Sci. 2019 Jul 23;76(23):4673–87.

176. Peperstraete E, Lecerf C, Collette J, Vennin C, Raby L, Völkel P, et al. Enhancement of Breast Cancer Cell Aggressiveness by lncRNA *H19* and its Mir-675 Derivative: Insight into Shared and Different Actions. Cancers. 2020 Jul;12(7):1730.

177. Yang J, Qi M, Fei X, Wang X, Wang K. LncRNA *H19*: A novel oncogene in multiple cancers. International Journal of Biological Sciences. 2021 Jul 25;17(12):3188–208.

178. Liu S, Chen W, Hu H, Zhang T, Wu T, Li X, et al. Long noncoding RNA PVT1 promotes breast cancer proliferation and metastasis by binding miR-128-3p and UPF1. Breast Cancer Res. 2021 Dec 18;23(1):115.

179. Lo PK, Zhang Y, Wolfson B, Gernapudi R, Yao Y, Duru N, et al. Dysregulation of the BRCA1/long non-coding RNA *NEAT1* signaling axis contributes to breast tumorigenesis. Oncotarget. 2016 Oct 4;7(40):65067–89.

180. Choudhry H, Albukhari A, Morotti M, Haider S, Moralli D, Smythies J, et al. Tumor hypoxia induces nuclear paraspeckle formation through HIF-2α dependent transcriptional activation of *NEAT1* leading to cancer cell survival. Oncogene. 2015 Aug 20;34(34):4482–90.

181. Pickard MR, Williams GT. The hormone response element mimic sequence of *GAS5* lncRNA is sufficient to induce apoptosis in breast cancer cells. Oncotarget. 2016 Feb 3;7(9):10104–16.

182. Huang YS, Chang CC, Lee SS, Jou YS, Shih HM. Xist reduction in breast cancer upregulates AKT phosphorylation via HDAC3-mediated repression of PHLPP1 expression. Oncotarget. 2016 May 27;7(28):43256–66.

183. Yuan H, Yan L, Wu M, Shang Y, Guo Q, Ma X, et al. Analysis of the Estrogen Receptor-Associated LncRNA Landscape Identifies a Role for ERLC1 in Breast Cancer Progression. Cancer Res. 2022 Feb 1;82(3):391–405.

184. Lin CY, Kleinbrink EL, Dachet F, Cai J, Ju D, Goldstone A, et al. Primate-specific oestrogen-responsive long non-coding RNAs regulate proliferation and viability of human breast cancer cells. Open Biology. 2016 Dec;6(12):150262.

185. Zhang Y, Wang DL, Yan HY, Liao JY, He JH, Hu KS, et al. Genome-wide study of ER-regulated lncRNAs shows AP000439.3 may function as a key regulator of cell cycle in breast cancer. Oncology Reports. 2017 Nov 1;38(5):3227–37.

186. Fang Z, Wang Y, Wang Z, Xu M, Ren S, Yang D, et al. *ERINA* Is an Estrogen-Responsive LncRNA That Drives Breast Cancer through the E2F1/RB1 Pathway. Cancer Res. 2020 Oct 15;80(20):4399–413.

187. Xue X, Yang YA, Zhang A, Fong KW, Kim J, Song B, et al. LncRNA *HOTAIR* enhances ER signaling and confers tamoxifen resistance in breast cancer. Oncogene. 2016 May;35(21):2746–55.

188. Miano V, Ferrero G, Reineri S, Caizzi L, Annaratone L, Ricci L, et al. Luminal long non-coding RNAs regulated by estrogen receptor alpha in a ligand-independent manner show functional roles in breast cancer. Oncotarget. 2015 Nov 28;7(3):3201–16.

189. Ma Y, Bu D, Long J, Chai W, Dong J. LncRNA *DSCAM-AS1* acts as a sponge of miR-137 to enhance Tamoxifen resistance in breast cancer. J Cell Physiol. 2019 Mar;234(3):2880–94.

190. Niknafs YS, Han S, Ma T, Speers C, Zhang C, Wilder-Romans K, et al. The lncRNA landscape of breast cancer reveals a role for *DSCAM-AS1* in breast cancer progression. Nat Commun. 2016 Sep 26;7(1):12791.

191. Liao J, Chen B, Zhu Z, Du C, Gao S, Zhao G, et al. Long noncoding RNA (lncRNA) *H19*: An essential developmental regulator with expanding roles in cancer, stem cell differentiation, and metabolic diseases. Genes Dis. 2023 Jul;10(4):1351–66.

192. Rostom MM, Rashwan AA, Sotiropoulou CD, Hozayen SZ, Abdelhamid AM, Abdelhalim MM, et al. MIAT: A pivotal oncogenic long noncoding RNA tunning the hallmarks of solid malignancies. Transl Oncol. 2025 Feb 26;54:102329.

193. Wang Z, Katsaros D, Biglia N, Shen Y, Loo L, Yu X, et al. ERα upregulates the expression of long non-coding RNA LINC00472 which suppresses the phosphorylation of NF-κB in breast cancer. Breast Cancer Res Treat. 2019 Jun;175(2):353–68.

194. Horie K, Takagi K, Takeiwa T, Mitobe Y, Kawabata H, Suzuki T, et al. Estrogen-Inducible LncRNA *BNAT1* Functions as a Modulator for Estrogen Receptor Signaling in Endocrine-Resistant Breast Cancer Cells. Cells. 2022 Nov 15;11(22):3610.

195. Ilbeigi S, Naeimzadeh Y, Davoodabadi Farahani M, Rafi Monjezi M, Dastsooz H, Daraei A, et al. Clinical values of two estrogen receptor signaling targeted lncRNAs in invasive ductal breast carcinoma. Klin Onkol. 2021;34(5):382–91.

196. Qiu JJ, Ye LC, Ding JX, Feng WW, Jin HY, Zhang Y, et al. Expression and clinical significance of estrogen-regulated long non-coding RNAs in estrogen receptor α-positive ovarian cancer progression. Oncology Reports. 2014 Apr 1;31(4):1613–22.

197. Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li R, Hennig S, et al. The oestrogen receptor alpha-regulated lncRNA *NEAT1* is a critical modulator of prostate cancer. Nat Commun. 2014 Nov 21;5(1):5383.

198. Ma T, Liang Y, Li Y, Song X, Zhang N, Li X, et al. LncRNA LINP1 confers tamoxifen resistance and negatively regulated by ER signaling in breast cancer. Cell Signal. 2020 Apr;68:109536.

199. Melone V, Palumbo D, Palo L, Brusco N, Salvati A, Tarallo A, et al. LncRNA PVT1 links estrogen receptor alpha and the polycomb repressive complex 2 in suppression of pro-apoptotic genes in hormone-responsive breast cancer. Cell Death Dis. 2025 Feb 8;16(1):80.

200. Aiello A, Bacci L, Re A, Ripoli C, Pierconti F, Pinto F, et al. *MALAT1* and *HOTAIR* Long Non-Coding RNAs Play Opposite Role in Estrogen-Mediated Transcriptional Regulation in Prostate Cancer Cells. Sci Rep. 2016 Dec 6;6:38414.

201. Du T, Shi Y, Xu S, Wan X, Sun H, Liu B. Long Non-Coding RNAs in Drug Resistance of Breast Cancer . OTT. 2020 Jul 21;13:7075–87.

202. Shi Q, Li Y, Li S, Jin L, Lai H, Wu Y, et al. LncRNA DILA1 inhibits Cyclin D1 degradation and contributes to tamoxifen resistance in breast cancer. Nat Commun. 2020 Nov 2;11(1):5513.

203. Peng W xin, Huang J guo, Yang L, Gong A hua, Mo YY. Linc-RoR promotes MAPK/ERK signaling and confers estrogen-independent growth of breast cancer. Mol Cancer. 2017 Oct 17;16:161.

204. Feng J, Wen T, Li Z, Feng L, Zhou L, Yang Z, et al. Cross-talk between the ER pathway and the lncRNA MAFG-AS1/miR-339-5p/ CDK2 axis promotes progression of ER+ breast cancer and confers tamoxifen resistance. Aging. 2020 Oct 24;12(20):20658–83.

205. Sedano MJ, Harrison AL, Zilaie M, Das C, Choudhari R, Ramos E, et al. Emerging Roles of Estrogen-Regulated Enhancer and Long Non-Coding RNAs. Int J Mol Sci. 2020 May 25;21(10):3711.

206. Qin Y, Sun W, Wang Z, Dong W, He L, Zhang T, et al. Long Non-Coding Small Nucleolar RNA Host Genes (SNHGs) in Endocrine-Related Cancers. Onco Targets Ther. 2020;13:7699–717.

207. Xiao H, Feng X, Liu M, Gong H, Zhou X. SnoRNA and lncSNHG: Advances of nucleolar small RNA host gene transcripts in anti-tumor immunity. Front Immunol. 2023 Mar 15;14:1143980.

208. Huang ZH, Du YP, Wen JT, Lu BF, Zhao Y. snoRNAs: functions and mechanisms in biological processes, and roles in tumor pathophysiology. Cell Death Discov. 2022 May 12;8(1):259.

209. Saadh MJ, Hamid JA, Malathi H, Kazmi SW, Omar TM, Sharma A, et al. SNHG family lncRNAs: Key players in the breast cancer progression and immune cell's modulation. Experimental Cell Research. 2025 Apr 15;447(2):114531.

210. Zimta AA, Tigu AB, Braicu C, Stefan C, Ionescu C, Berindan-Neagoe I. An Emerging Class of Long Non-coding RNA With Oncogenic Role Arises From the snoRNA Host Genes. Front Oncol. 2020 Apr 7;10:389.

211. Zheng H, Wang G, Wang Y, Liu J, Ma G, Du J. Systematic analysis reveals a pan-cancer SNHG family signature predicting prognosis and immunotherapy response. iScience. 2023 Sep 25;26(10):108055.

212. Zhang M, Yang L, Hou L, Tang X. LncRNA *SNHG1* promotes tumor progression and cisplatin resistance through epigenetically silencing miR-381 in breast cancer. Bioengineered. 2021 Dec;12(2):9239–50.

213. Sun L, Chu H, Li H, Liu Y. LncRNA *SNHG1* correlates with higher T stage and worse overall survival, and promotes cell proliferation while reduces cell apoptosis in breast cancer. Transl Cancer Res. 2019 Apr;8(2):603–13.

214. Zeng H, Hou Y, Zhou X, Lang L, Luo H, Sun Y, et al. Cancer-associated fibroblasts facilitate premetastatic niche formation through lncRNA *SNHG5*-mediated angiogenesis and vascular permeability in breast cancer. Theranostics. 2022 Oct 17;12(17):7351–70.

215. Li K, Ma YB, Tian YH, Xu XL, Gao Y, He YQ, et al. Silencing lncRNA *SNHG6* suppresses proliferation and invasion of breast cancer cells through miR-26a/VASP axis. Pathol Res Pract. 2019 Oct;215(10):152575.

216. Khan MI, Ahmad A. LncRNA *SNHG6* sponges miR-101 and induces tamoxifen resistance in breast cancer cells through induction of EMT. Front Oncol. 2022 Sep 23;12:1015428.

217. Kong Q, Qiu M. Long noncoding RNA *SNHG15* promotes human breast cancer proliferation, migration and invasion by sponging miR-211-3p. Biochem Biophys Res Commun. 2018 Jan 8;495(2):1594–600.

218. Yang M, Wei W. SNHG16: A Novel Long-Non Coding RNA in Human Cancers. Onco Targets Ther. 2019 Dec 31;12:11679–90.

219. Zhang K, Tan X, Guo L. The long non-coding RNA *DANCR* regulates the inflammatory phenotype of breast cancer cells and promotes breast cancer progression via EZH2-dependent suppression of SOCS3 transcription. Mol Oncol. 2020 Feb;14(2):309–28.

220. Yan S, Teng L, Du J, Ji L, Xu P, Zhao W, et al. Long non-coding RNA *DANCR* aggravates breast cancer through the miR-34c/E2F1 feedback loop. Molecular Medicine Reports. 2024 Jun 1;29(6):1–13.

221. Xiong X, Feng Y, Li L, Yao J, Zhou M, Zhao P, et al. Long non-coding RNA *SNHG1* promotes breast cancer progression by regulation of LMO4. Oncol Rep. 2020 May;43(5):1503–15.

222. Kaur J, Salehen N, Norazit A, Rahman AA, Murad NAA, Rahman NMANAbd, et al. Tumor Suppressive Effects of *GAS5* in Cancer Cells. Noncoding RNA. 2022 May 28;8(3):39.

223. Liu J, Chen M, Ma L, Dang X, Du G. LncRNA *GAS5* Suppresses the Proliferation and Invasion of Osteosarcoma Cells via the miR-23a-3p/PTEN/PI3K/AKT Pathway. Cell Transplant. 2020 Oct 30;29:0963689720953093.

224. Fan Y, Gao Z, Xu J, Wang H, Guo Q, Xue H, et al. Identification and validation of SNHG gene signature to predict malignant behaviors and therapeutic responses in glioblastoma. Front Immunol [Internet]. 2022 Sep 8 [cited 2025 Mar 20];13. Available from: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2022.986615/full

225. Zhen N, Zhu J, Mao S, Zhang Q, Gu S, Ma J, et al. Alternative Splicing of lncRNAs From SNHG Family Alters snoRNA Expression and Induces Chemoresistance in Hepatoblastoma. Cellular and Molecular Gastroenterology and Hepatology. 2023 Jan 1;16(5):735–55.

226. Choo SW, Zhong Y, Sendler E, Goustin AS, Cai J, Ju D, et al. Estrogen distinctly regulates transcription and translation of lncRNAs and pseudogenes in breast cancer cells. Genomics. 2022 Jul;114(4):110421.

227. Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, et al. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell. 2011 May 13;145(4):622–34.

228. Lafront C, Germain L, Audet-Walsh É. Bulk mRNA-sequencing data of the estrogen and androgen responses in the human prostate cancer cell line VCaP. Data in Brief. 2024 Dec 1;57:111053.

229. Erdem M, Ozgul İ, Dioken DN, Gurcuoglu I, Guntekin Ergun S, Cetin-Atalay R, et al. Identification of an mRNA isoform switch for HNRNPA1 in breast cancers. Sci Rep. 2021 Dec 27;11(1):24444.

230. Cozzuto L, Liu H, Pryszcz LP, Pulido TH, Delgado-Tejedor A, Ponomarenko J, et al. MasterOfPores: A Workflow for the Analysis of Oxford Nanopore Direct RNA Sequencing Datasets. Front Genet. 2020 Mar 17;11:211.

231. Dias K, Lemoine S, Thomas-Chollier M, Le Crom S, Jourdren L. ToulligQC 2: fast and comprehensive quality control for Oxford Nanoporesequencing data [Internet]. SFBI; 2022 Jul [cited 2025 Jul 29]. Available from: https://hal.science/hal-05144856

232. Kahlke T. timkahlke/LongRead_tutorials [Internet]. 2025 [cited 2025 Jul 29]. Available from: https://github.com/timkahlke/LongRead_tutorials

233. Coster WD. wdecoster/NanoPlot [Internet]. 2025 [cited 2025 Jul 29]. Available from: https://github.com/wdecoster/NanoPlot

234. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.

235. Huang Y, Zhang Z, Zou Z, Zhang L, Chen Y, Wan J, et al. RegRNA 3.0: expanding regulatory RNA analysis with new features for motif, interaction, and annotation. Nucleic Acids Res. 2025 May 21;gkaf405.

236. Wang R, Nambiar R, Zheng D, Tian B. PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. Nucleic Acids Res. 2018 Jan 4;46(D1):D315–9.

237. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417–25.

238. Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. 2023 Jul 5;51(W1):W207–12.

239. Zhang S, Amahong K, Zhang Y, Hu X, Huang S, Lu M, et al. RNAenrich: a web server for non-coding RNA enrichment. Bioinformatics. 2023 Jul 1;39(7):btad421.

240. Li J, Li Z, Wang Y, Lin H, Wu B. TLSEA: a tool for lncRNA set enrichment analysis based on multi-source heterogeneous information fusion. Front Genet. 2023 May 2;14:1181391.

241. Bergeron D, Paraqindes H, Fafard-Couture É, Deschamps-Francoeur G, Faucher-Giguère L, Bouchard-Bourelle P, et al. snoDB 2.0: an enhanced interactive database, specializing in human snoRNAs. Nucleic Acids Res. 2023 Jan 6;51(D1):D291–6.

242. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun;45(6):580–5.

243. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017 May 5;8:15081.

244. Győrffy B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. Computational and Structural Biotechnology Journal. 2021 Jan 1;19:4101–9.

245. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009 Apr;55(4):611–22.

246. Huang J, Li X, Yi P, Hilf R, Bambara RA, Muyan M. Targeting estrogen responsive elements (EREs): design of potent transactivators for ERE-containing genes. Molecular and Cellular Endocrinology. 2004 Apr 15;218(1):65–78.

247. Muyan M, Yi P, Sathya G, Willmert LJ, Driscoll MD, Hilf R, et al. Fusion estrogen receptor proteins: toward the development of receptor-based agonists and antagonists. Molecular and Cellular Endocrinology. 2001 Sep 20;182(2):249–63.

248. Yi P, Bhagat S, Hilf R, Bambara RA, Muyan M. Differences in the Abilities of Estrogen Receptors to Integrate Activation Functions Are Critical for Subtype-Specific Transcriptional Responses. Molecular Endocrinology. 2002 Aug 1;16(8):1810–27.

249. Martelotto LG, Martelotto L. 'Frankenstein' protocol for nuclei isolation from fresh and frozen tissue for snRNAseq. 2020 Dec 21 [cited 2025 Jul 29];

Available from: https://www.protocols.io/view/frankenstein-protocol-for-nuclei-isolation-from-f-bqxymxpw

250. Mosmann T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. J Immunol Methods. 1983 Dec 16;65(1–2):55–63.

251. Crowley LC, Christensen ME, Waterhouse NJ. Measuring Survival of Adherent Cells with the Colony-Forming Assay. Cold Spring Harb Protoc. 2016 Aug 1;2016(8).

252. Jagannathan V, Robinson-Rechavi M. Meta-analysis of estrogen response in MCF-7 distinguishes early target genes involved in signaling and cell proliferation from later target genes involved in cell cycle and DNA repair. BMC Syst Biol. 2011 Aug 30;5:138.

253. Li Z, Li T, Yates ME, Wu Y, Ferber A, Chen L, et al. The EstroGene database reveals diverse temporal, context-dependent, and bidirectional estrogen receptor regulomes in breast cancer. Cancer Res. 2023 Aug 15;83(16):2656–74.

254. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005 Oct 25;102(43):15545–50.

255. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes | Nature Genetics [Internet]. [cited 2025 Jun 25]. Available from: https://www.nature.com/articles/ng1180

256. Kulkoyluoglu-Cotul E, Arca A, Madak-Erdogan Z. Crosstalk between Estrogen Signaling and Breast Cancer Metabolism. Trends in Endocrinology & Metabolism. 2019 Jan 1;30(1):25–38.

257. Li P, Feng C, Chen H, Jiang Y, Cao F, Liu J, et al. Elevated CRB3 expression suppresses breast cancer stemness by inhibiting β-catenin signalling to restore tamoxifen sensitivity. J Cell Mol Med. 2018 Jul;22(7):3423–33.

258. Scirè A, Cianfruglia L, Minnelli C, Romaldi B, Laudadio E, Galeazzi R, et al. Glyoxalase 2: Towards a Broader View of the Second Player of the Glyoxalase System. Antioxidants (Basel). 2022 Oct 28;11(11):2131.

259. Rae JM, Johnson MD, Scheys JO, Cordero KE, Larios JM, Lippman ME. GREB 1 is a critical regulator of hormone dependent breast cancer growth. Breast Cancer Res Treat. 2005 Jul;92(2):141–9.

260. Holmes KA, Hurtado A, Brown GD, Launchbury R, Ross-Innes CS, Hadfield J, et al. Transducin-like enhancer protein 1 mediates estrogen receptor binding and transcriptional activity in breast cancer cells. Proc Natl Acad Sci U S A. 2012 Feb 21;109(8):2748–53.

261. Caldas P, Luz M, Baseggio S, Andrade R, Sobral D, Grosso AR. Transcription readthrough is prevalent in healthy human tissues and associated with inherent genomic features. Commun Biol. 2024 Jan 15;7(1):100.

262. Varley KE, Gertz J, Roberts BS, Davis NS, Bowling KM, Kirby MK, et al. Recurrent read-through fusion transcripts in breast cancer. Breast Cancer Res Treat. 2014 Jul;146(2):287–97.

263. Hah N, Kraus WL. Hormone-regulated transcriptomes: lessons learned from estrogen signaling pathways in breast cancer cells. Mol Cell Endocrinol. 2014 Jan 25;382(1):652–64.

264. Sun W, Xu X, Jiang Y, Jin X, Zhou P, Liu Y, et al. Transcriptome analysis of luminal breast cancer reveals a role for LOL in tumor progression and tamoxifen resistance. Int J Cancer. 2019 Aug 1;145(3):842–56.

265. Ribas J, Lupold SE. The transcriptional regulation of miR-21, its multiple transcripts, and their implication in prostate cancer. Cell Cycle. 2010 Mar 1;9(5):923–9.

266. Siprashvili Z, Webster DE, Johnston D, Shenoy RM, Ungewickell AJ, Bhaduri A, et al. The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer. Nat Genet. 2016 Jan;48(1):53–8.

267. Youssef OA, Safran SA, Nakamura T, Nix DA, Hotamisligil GS, Bass BL. Potential role for snoRNAs in PKR activation during metabolic stress. Proc Natl Acad Sci U S A. 2015 Apr 21;112(16):5023–8.

268. Yuan S, Wu Y, Wang Y, Chen J, Chu L. An Oncolytic Adenovirus Expressing SNORD44 and *GAS5* Exhibits Antitumor Effect in Colorectal Cancer Cells. Hum Gene Ther. 2017 Aug;28(8):690–700.

269. Lei L, Huang Y, Shi L, Ye W, Lv X, Ying L, et al. Palbociclib sensitizes ER-positive breast cancer cells to fulvestrant by promoting the ubiquitin-mediated degradation of ER-α via *SNHG17*/Hippo-YAP axis. Breast Cancer Res Treat. 2024 Feb;203(3):613–25.

270. Tong J, Ma X, Yu H, Yang J. *SNHG15*: a promising cancer-related long noncoding RNA. Cancer Manag Res. 2019;11:5961–9.

271. Ma L, Gao J, Zhang N, Wang J, Xu T, Lei T, et al. Long noncoding RNA *SNHG17*: a novel molecule in human cancers. Cancer Cell Int. 2022 Mar 5;22(1):104.

272. Feng SG, Bhandari R, Ya L, Zhixuan B, Qiuhui P, Jiabei Z, et al. *SNHG9* promotes Hepatoblastoma Tumorigenesis via miR-23a-5p/Wnt3a Axis. J Cancer. 2021;12(20):6031–49.

273. Chen GY, Zhang ZS, Chen Y, Li Y. Long non-coding RNA *SNHG9* inhibits ovarian cancer progression by sponging microRNA-214-5p. Oncol Lett. 2021 Feb;21(2):80.

274. Wang D, Cao X, Han Y, Yu D. LncRNA *SNHG9* is Downregulated in Non-Small Cell Lung Cancer and Suppressed miR-21 Through Methylation to Promote Cell Proliferation. Cancer Manag Res. 2020;12:7941–8.

275. Li C, Hu J, Hu X, Zhao C, Mo M, Zu X, et al. LncRNA *SNHG9* is a prognostic biomarker and correlated with immune infiltrates in prostate cancer. Transl Androl Urol. 2021 Jan;10(1):215–26.

276. Bai CJ, Gao T, Liu JY, Li S, Wang XY, Fan ZF. *SNHG9*/miR-214-5p/SOX4 feedback loop regulates osteosarcoma progression. Neoplasma. 2022 Sep;69(5):1175–84.

277. Zhang H, Qin D, Jiang Z, Zhang J. *SNHG9*/miR-199a-5p/Wnt2 Axis Regulates Cell Growth and Aerobic Glycolysis in Glioblastoma. J Neuropathol Exp Neurol. 2019 Oct 1;78(10):939–48.

278. Karagkouni D, Paraskevopoulou MD, Tastsoglou S, Skoufos G, Karavangeli A, Pierros V, et al. DIANA-LncBase v3: indexing experimentally supported miRNA targets on non-coding transcripts. Nucleic Acids Res. 2020 Jan 8;48(D1):D101–10.

279. Ghafouri-Fard S, Khoshbakht T, Hussen BM, Baniahmad A, Taheri M, Samadian M. A review on the role of *DANCR* in the carcinogenesis. Cancer Cell International. 2022 May 19;22(1):194.

280. Ratnadiwakara M, Änkö ML. mRNA Stability Assay Using transcription inhibition by Actinomycin D in Mouse Pluripotent Stem Cells. Bio Protoc. 2018 Nov 5;8(21):e3072.

281. Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, et al. Genome-wide analysis of long noncoding RNA stability. Genome Res. 2012 May;22(5):885–98.

282. Aliperti V, Skonieczna J, Cerase A. Long Non-Coding RNA (lncRNA) Roles in Cell Biology, Neurodevelopment and Neurological Disorders. Noncoding RNA. 2021 Jun 17;7(2):36.

283. Miller DM, Thomas SD, Islam A, Muench D, Sedoris K. c-*MYC* and Cancer Metabolism. Clin Cancer Res. 2012 Oct 15;18(20):5546–53.

284. Dang CV. *MYC* on the Path to Cancer. Cell. 2012 Mar 30;149(1):22–35.

285. Yuan Y, Liu B, Xie P, Zhang MQ, Li Y, Xie Z, et al. Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. Proc Natl Acad Sci U S A. 2015 Mar 10;112(10):3158–63.

286. Bosson AD, Zamudio JR, Sharp PA. Endogenous miRNA and Target Concentrations Determine Susceptibility to Potential ceRNA Competition. Mol Cell. 2014 Nov 6;56(3):347–59.

287. Fu H, Si J, Xu L, Tang X, He Y, Lu N, et al. Long non-coding RNA *SNHG9* regulates viral replication in rhabdomyosarcoma cells infected with enterovirus D68 via miR-150-5p/c-Fos axis. Front Microbiol. 2022;13:1081237.

288. Pack LR, Daigh LH, Meyer T. Putting the brakes on the cell cycle: mechanisms of cellular growth arrest. Current Opinion in Cell Biology. 2019 Oct 1;60:106–13.

289. Fox EM, Andrade J, Shupnik MA. Novel Actions of Estrogen to Promote Proliferation: Integration of Cytoplasmic and Nuclear Pathways. Steroids. 2009 Jul;74(7):622–7.

290. Miziak P, Baran M, Błaszczak E, Przybyszewska-Podstawka A, Kałafut J, Smok-Kalwat J, et al. Estrogen Receptor Signaling in Breast Cancer. Cancers (Basel). 2023 Sep 23;15(19):4689.

291. Li P, Wang Y, Mao X, Jiang Y, Liu J, Li J, et al. CRB3 downregulation confers breast cancer stem cell traits through TAZ/β-catenin. Oncogenesis. 2017 Apr 24;6(4):e322.

292. Cordell PA, Futers TS, Grant PJ, Pease RJ. The Human hydroxyacylglutathione hydrolase (HAGH) gene encodes both cytosolic and mitochondrial forms of glyoxalase II. J Biol Chem. 2004 Jul 2;279(27):28653–61.

293. Mitschka S, Mayr C. Context-specific regulation and function of mRNA alternative polyadenylation. Nat Rev Mol Cell Biol. 2022 Dec;23(12):779–96.

294. Singh I, Lee SH, Sperling AS, Samur MK, Tai YT, Fulciniti M, et al. Widespread intronic polyadenylation diversifies immune cell transcriptomes. Nat Commun. 2018 Apr 30;9(1):1716.

295. Yu G, Chen Y, Hu Y, Zhou Y, Ding X, Zhou X. Roles of transducin-like enhancer of split (TLE) family proteins in tumorigenesis and immune regulation. Frontiers in cell and developmental biology [Internet]. 2022 Nov 11 [cited 2025 Jul 26];10. Available from: https://pubmed.ncbi.nlm.nih.gov/36438567/

296. Li D, Roberts R. WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cellular and molecular life sciences : CMLS [Internet]. 2001 Dec [cited 2025 Jul 26];58(14). Available from: https://pubmed.ncbi.nlm.nih.gov/11814058/

297. Carroll J, Liu X, Brodsky A, Li W, Meyer C, Aj S, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell [Internet]. 2005 Jul 15 [cited 2025 Jul 26];122(1). Available from: https://pubmed.ncbi.nlm.nih.gov/16009131/

298. Xiao T, Li W, Wang X, Xu H, Yang J, Q W, et al. Estrogen-regulated feedback loop limits the efficacy of estrogen receptor-targeted breast cancer therapy. Proceedings of the National Academy of Sciences of the United States of America [Internet]. 2018 Jul 31 [cited 2025 Jul 26];115(31). Available from: https://pubmed.ncbi.nlm.nih.gov/29987050/

299. Dizin E, Irminger-Finger I. Negative feedback loop of BRCA1–BARD1 ubiquitin ligase on estrogen receptor alpha stability and activity antagonized by cancer-associated isoform of BARD1. The International Journal of Biochemistry & Cell Biology. 2010 May 1;42(5):693–700.

300. Wu S, Lee C, Lai H, Yu C, Lee J, Zuo H, et al. Opposing Functions of BRD4 Isoforms in Breast Cancer. Molecular cell [Internet]. 2020 Jun 18 [cited 2025 Jul 26];78(6). Available from: https://pubmed.ncbi.nlm.nih.gov/32446320/

301. Stevenson B, Gorman M, Koach J, Cheung B, Marshall G, Parker M, et al. A structural view of PA2G4 isoforms with opposing functions in cancer. The Journal of biological chemistry [Internet]. 2020 Nov 20 [cited 2025 Jul 26];295(47). Available from: https://pubmed.ncbi.nlm.nih.gov/32952126/

302. Ko H, Chang Y, Park W, Ahn J. Opposing roles of the two isoforms of ErbB3 binding protein 1 in human cancer cells. International journal of cancer [Internet]. 2016 Sep 15 [cited 2025 Jul 26];139(6). Available from: https://pubmed.ncbi.nlm.nih.gov/27130196/

303. Ryvkin V, Rashel M, Gaddapara T, Ghazizadeh S. Opposing growth regulatory roles of protein kinase D isoforms in human keratinocytes. The Journal of biological chemistry [Internet]. 2015 Apr 24 [cited 2025 Jul 26];290(17). Available from: https://pubmed.ncbi.nlm.nih.gov/25802335/

304. Boise L, González-García M, Postema C, Ding L, Lindsten T, Turka L, et al. bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. Cell [Internet]. 1993 Aug 27 [cited 2025 Jul 26];74(4). Available from: https://pubmed.ncbi.nlm.nih.gov/8358789/

305. Bates D, Cui T, Doughty J, Winkler M, Sugiono M, Shields J, et al. VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, is down-regulated in renal cell carcinoma. Cancer research [Internet]. 2002 Jul 15 [cited 2025 Jul 26];62(14). Available from: https://pubmed.ncbi.nlm.nih.gov/12124351/

306. Bourdon J, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas D, et al. p53 isoforms can regulate p53 transcriptional activity. Genes & development [Internet]. 2005 Sep 15 [cited 2025 Jul 26];19(18). Available from: https://pubmed.ncbi.nlm.nih.gov/16131611/

307. Rojas PA, May M, Sequeira GR, Elia A, Alvarez M, Martínez P, et al. Progesterone Receptor Isoform Ratio: A Breast Cancer Prognostic and Predictive Factor for Antiprogestin Responsiveness. J Natl Cancer Inst. 2017 Jul 1;109(7):djw317.

308. Steffens Reinhardt L, Zhang X, Groen K, Morten BC, De Iuliis GN, Braithwaite AW, et al. Alterations in the p53 isoform ratio govern breast cancer cell fate in response to DNA damage. Cell Death Dis. 2022 Oct 28;13(10):907.

309. Wang M, Li Y, Xiao G, Zheng X, Wang J, Xu C, et al. *H19* regulation of oestrogen induction of symmetric division is achieved by antagonizing Let-7c in breast cancer stem-like cells. Cell Prolif. 2018 Oct 18;52(1):e12534.

310. Mitobe Y, Ikeda K, Suzuki T, Takagi K, Kawabata H, Horie-Inoue K, et al. ESR1-Stabilizing Long Noncoding RNA TMPO-AS1 Promotes Hormone-Refractory Breast Cancer Progression. Mol Cell Biol. 2019 Nov 12;39(23):e00261-19.

311. Hou P, Zhao Y, Li Z, Yao R, Ma M, Gao Y, et al. LincRNA-ROR induces epithelial-to-mesenchymal transition and contributes to breast cancer tumorigenesis and metastasis. Cell Death Dis. 2014 Jun;5(6):e1287.

312. Fang K, Xu ZJ, Jiang SX, Tang DS, Yan CS, Deng YY, et al. lncRNA FGD5-AS1 promotes breast cancer progression by regulating the

hsa-miR-195-5p/NUAK2 axis Retraction in /10.3892/mmr.2025.13457. Molecular Medicine Reports. 2021 Jun 1;23(6):1–8.

313. Shang Y, Cao T, Ma X, Huang L, Wu M, Xu J, et al. Estrogen-induced FXR1 promotes endocrine resistance and bone metastasis in breast cancer via BCL2 and GPX4. Front Cell Dev Biol [Internet]. 2025 Mar 24 [cited 2025 Jul 30];13. Available from: https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/10.3389/fcell.2025.1563353/full

314. Arun G, Diermeier S, Akerman M, Chang KC, Wilkinson JE, Hearn S, et al. Differentiation of mammary tumors and reduction in metastasis upon *MALAT1* lncRNA loss. Genes Dev. 2016 Jan 1;30(1):34–51.

315. Xiang Y, Liu H, Hu H, Li LW, Zong QB, Wu TW, et al. LINC00094/miR-19a-3p/CYP19A1 axis affects the sensitivity of ER positive breast cancer cells to Letrozole through EMT pathway. Aging (Albany NY). 2022 Jun 2;14(11):4755–68.

316. Li X, Wu Y, Liu A, Tang X. Long non-coding RNA UCA1 enhances tamoxifen resistance in breast cancer cells through a miR-18a-HIF1α feedback regulatory loop. Tumour Biol. 2016 Nov;37(11):14733–43.

317. Li W, Zhang Z, Liu X, Cheng X, Zhang Y, Han X, et al. The FOXN3-*NEAT1*-SIN3A repressor complex promotes progression of hormonally responsive breast cancer. J Clin Invest. 127(9):3421–40.

318. Zhang XF, Ye Y, Zhao SJ. LncRNA *GAS5* acts as a ceRNA to regulate PTEN expression by sponging miR-222-3p in papillary thyroid carcinoma. Oncotarget. 2017 Dec 16;9(3):3519–30.

319. Lambrou GI, Hatziagapiou K, Zaravinos A. The Non-Coding RNA *GAS5* and Its Role in Tumor Therapy-Induced Resistance. International Journal of Molecular Sciences. 2020 Jan;21(20):7633.

320. Liu Y, Li M, Yu H, Piao H. lncRNA CYTOR promotes tamoxifen resistance in breast cancer cells via sponging miR-125a-5p. International Journal of Molecular Medicine. 2020 Feb 1;45(2):497–509.

321. Kim CY, Oh JH, Lee JY, Kim MH. The LncRNA *HOTAIR*M1 Promotes Tamoxifen Resistance by Mediating HOXA1 Expression in ER+ Breast Cancer Cells. J Cancer. 2020 Mar 13;11(12):3416–23.

322. Deng L, Wang J, Song J, Wu Q, Gong Z, Song J, et al. Long noncoding RNA *SNHG1* promotes breast cancer progression by regulating the miR-641/RRS1 axis. Sci Rep. 2024 Feb 8;14(1):3265.

323. Chen Z, Pan T, Jiang D, Jin L, Geng Y, Feng X, et al. The lncRNA-*GAS5*/miR-221-3p/DKK2 Axis Modulates ABCB1-Mediated Adriamycin Resistance of Breast Cancer via the Wnt/β-Catenin Signaling Pathway. Mol Ther Nucleic Acids. 2020 Mar 6;19:1434–48.

324. Wang L, Su K, Wu H, Li J, Song D. LncRNA SNHG3 regulates laryngeal carcinoma proliferation and migration by modulating the miR-384/WEE1 axis. Life Sci. 2019 Sep 1;232:116597.

325. Jiang H, Li X, Wang W, Dong H. Long non-coding RNA SNHG3 promotes breast cancer cell proliferation and metastasis by binding to microRNA-154-3p and activating the notch signaling pathway. BMC Cancer. 2020 Sep 3;20(1):838.

326. Wang F, Quan Q. The long non-coding RNA *SNHG4*/microRNA-let-7e/KDM3A/p21 pathway is involved in the development of non-small cell lung cancer. Mol Ther Oncolytics. 2021 Mar 26;20:634–45.

327. Chi JR, Yu ZH, Liu BW, Zhang D, Ge J, Yu Y, et al. *SNHG5* Promotes Breast Cancer Proliferation by Sponging the miR-154-5p/PCNA Axis. Mol Ther Nucleic Acids. 2019 Sep 6;17:138–49.

328. Huang SL, Huang ZC, Zhang CJ, Xie J, Lei SS, Wu YQ, et al. LncRNA *SNHG5* promotes the glycolysis and proliferation of breast cancer cell through regulating BACH1 via targeting miR-299. Breast Cancer. 2022 Jan;29(1):65–76.

329. Lv Y, Lv X, Yang H, Qi X, Wang X, Li C, et al. LncRNA *SNHG6*/miR-125b-5p/BMPR1B Axis: A New Therapeutic Target for Triple-Negative Breast Cancer. Front Oncol. 2021 May 7;11:678474.

330. Wang YQ, Huang G, Chen J, Cao H, Xu WT. LncRNA *SNHG6* promotes breast cancer progression and epithelial-mesenchymal transition via miR-543/LAMC1 axis. Breast Cancer Res Treat. 2021 Jul;188(1):1–14.

331. Boone DN, Warburton A, Som S, Lee AV. *SNHG7* is a lncRNA oncogene controlled by Insulin-like Growth Factor signaling through a negative feedback loop to tightly regulate proliferation. Sci Rep. 2020 May 22;10(1):8583.

332. Zhang H, Zhang XY, Kang XN, Jin LJ, Wang ZY. LncRNA-*SNHG7* Enhances Chemotherapy Resistance and Cell Viability of Breast Cancer Cells by Regulating miR-186. Cancer Manag Res. 2020;12:10163–72.

333. Qian J, Lei X, Sun Y, Zheng L, Li J, Zhang S, et al. Long non-coding RNA *SNHG8* enhances triple-negative breast cancer cell proliferation and migration by regulating the miR-335-5p/PYGO2 axis. Biol Direct. 2021 Aug 6;16(1):13.

334. Li RH, Tian T, Ge QW, He XY, Shi CY, Li JH, et al. A phosphatidic acid-binding lncRNA *SNHG9* facilitates LATS1 liquid-liquid phase separation to promote oncogenic YAP signaling. Cell Res. 2021 Oct;31(10):1088–105.

335. Wang M, Liu K, Bao W, Hang B, Chen X, Zhu X, et al. Gut microbiota protect against colorectal tumorigenesis through lncRNA *SNHG9*. Dev Cell. 2025 Apr 7;60(7):1008-1017.e7.

336. Al-Hazani TMI, Al-Qahtani WS, Alwaili MA, Domiaty DM, Alshehri E, Al-Shamrani SM, et al. The function of long non-coding RNA *SNHG11* and its working mechanism in triple-negative breast cancer. Pathol Res Pract. 2023 Aug;248:154578.

337. Dong Y, Wang G. Knockdown of lncRNA *SNHG12* suppresses cell proliferation, migration and invasion in breast cancer by sponging miR-451a. Int J Clin Exp Pathol. 2020 Mar 1;13(3):393–402.

338. Dong H, Wang W, Mo S, Liu Q, Chen X, Chen R, et al. Long non-coding RNA *SNHG14* induces trastuzumab resistance of breast cancer via regulating PABPC1 expression through H3K27 acetylation. J Cell Mol Med. 2018 Oct;22(10):4935–47.

339. Zhang D, Ding X, Peng M. LncRNA *SNHG14* accelerates breast cancer progression through sponging miR-543 and regulating KLF7 expression. Arch Gynecol Obstet. 2022 Jun 1;305(6):1507–16.

340. Wang B, Xing AY, Li GX, Liu L, Xing C. *SNHG14* promotes triple-negative breast cancer cell proliferation, invasion, and chemoresistance by regulating the ERK/MAPK signaling pathway. IUBMB Life. 2024;76(12):1295–308.

341. Cai C, Huo Q, Wang X, Chen B, Yang Q. SNHG16 contributes to breast cancer cell migration by competitively binding miR-98 with E2F5. Biochem Biophys Res Commun. 2017 Apr 1;485(2):272–8.

342. Guan YX, Zhang MZ, Chen XZ, Zhang Q, Liu SZ, Zhang YL. Lnc RNA *SNHG20* participated in proliferation, invasion, and migration of breast cancer cells via miR-495. J Cell Biochem. 2018 Nov;119(10):7971–81.

343. Fang X, Zhang J, Li C, Liu J, Shi Z, Zhou P. Long non-coding RNA *SNHG22* facilitates the malignant phenotypes in triple-negative breast cancer via sponging miR-324-3p and upregulating SUDS3. Cancer Cell Int. 2020 Jun 17;20:252.

344. Broad/Illumina Genome Analyzer Boot Camp [Internet]. @broadinstitute. 2011 [cited 2025 Jul 29]. Available from: https://www.broadinstitute.org/genome-sequencing/broadillumina-genome-analyzer-boot-camp

APPENDICES

A. Supplementary Tables for Chapter 1

Table A.1. Overview of Key LncRNAs that regulate or are part of Estrogen Signaling.

LncRNA	Cancer	Relationship with ER	Mechanism	Ref
PVT1	Breast Cancer	Interacts with ERalpha; Key component of estrogenic signaling; Modulates ERalpha transcriptional modulation	Acts as a bridge factor between PRC2 (EZH2) and ERalpha, forming a transcriptional repressive complex; Downregulates tumor suppressor genes (e.g., *BTG2*)	[199]
BNAT1	Breast Cancer	Physically binds ERalpha protein; Modulates ERalpha-dependent transcription; Role in endocrine-resistant breast cancer	Knockdown decreases ERalpha expression and represses ERalpha transactivation	[194]
HOTAIR	Breast Cancer, Prostate Cancer	Modulates ERalpha activity/chromatin occupancy (BC); Promotes ligand-independent ER activity (BC); Interacts with ERalpha/ERbeta (PCa)	Elevates ERalpha occupancy on chromatin (BC); May block miR-130b-3p to increase *ESR1* expression (BC); Opposite regulation to *MALAT1* in PCa response to estrogen	[168]
H19	Breast	Upregulates ERalpha	Acts as ceRNA	[191,

	Cancer	expression; Protects ERalpha from degradation; Confers tamoxifen/fulvestrant resistance	for Let-7c, affecting estrogen-activated Wnt pathway	309]
LINP1	Breast Cancer	Downregulates ERalpha expression; Confers tamoxifen resistance	Expression increases upon tamoxifen treatment (ER-mediated transcriptional repression lifted)	[198]
DSCAM-AS1	Breast Cancer	Silencing decreases ERalpha and FOXA1 expression; Promotes tamoxifen resistance	Acts as a sponge for miR-137 (which prompts cell cycle arrest)	[189, 189]
TMPO-AS1	Breast Cancer	Positively regulates *ESR1* expression; Promotes hormone-refractory breast cancer	Stabilizes *ESR1* mRNA	[310]
LINC-ROR	Breast Cancer	Involved in tamoxifen resistance; Promotes estrogen-independent ERalpha signaling	Affects estrogen receptor activation through MAPK/ERK pathway (facilitates DUSP7 degradation); Acts as a sponge for miR-205-5p	[203, 311]
FGD5-AS1	Breast Cancer	Identified as an ERalpha-interacting RNA in MCF-7 cells	General oncogenic role promoting glycolysis and tumor progression via the miR-195-5p/NUAK2 axis; Direct ERalpha interaction mechanism on	[312]

			signaling not fully detailed	
ERLC1	Breast Cancer	Counteracts ERalpha downregulation; Potentiates ERalpha signaling; Leads to tamoxifen/fulvestrant resistance	Counteracts ERalpha downregulation by miR-129; Tethers FXR1, creating a positive feedback loop enhancing ERalpha signaling	[183, 313]
DILA1	Breast Cancer	Contributes to tamoxifen resistance by increasing stability of cyclin D1	Directly binds cyclin D1 protein, inhibiting its phosphorylation at Thr286, reducing ubiquitination and degradation	[202]
LOL	Breast Cancer	Contributes to tamoxifen resistance	Acts as a sponge for let-7 miRNA (which targets and downregulates *ESR1* mRNA)	[264]
MALAT1	Breast Cancer, Prostate Cancer	Reduces response to tamoxifen (BC); Interacts with ERalpha/ERbeta (PCa)	Regulates *ESR1* mRNA splicing and gene expression (BC); Opposite regulation to *HOTAIR* in PCa response to estrogen	[314]
SRA1	Breast Cancer	RNA coactivator, increasing transcriptional activity of ERalpha; Enhances agonist activity of tamoxifen-bound ERalpha	Binds to NCOA1, acting as RNA coactivator; Functions as scaffold interacting with miRNA processing	[68]

			proteins and TFs	
LINC00094	Breast Cancer	Increases sensitivity of ERalpha+ breast cancer cells to letrozole	Increases and maintains expression of aromatase (CYP19A1) by sponging miR-19a-3p	[315]
UCA1	Breast Cancer	Confers tamoxifen resistance	Regulates EZH2/p21 axis; Activates PI3K/AKT/mTOR signaling; Acts as ceRNA for miR-18a/HIF1α	[316]
NEAT1	Prostate Cancer, Breast Cancer	Modulates ERalpha target genes (PCa & BCa); Drives oncogenic growth (PCa)	Alters epigenetic landscape of target gene promoters (PCa); Part of a transcriptional repressor complex with SIN3A and FOXN3 in ERalpha+ breast cancer cells (BCa)	[317]
GAS5	Breast Cancer, Ovarian Cancer	Regulates endocrine resistance in BC	Acts as ceRNA to sponge miR-222/PTEN (BC)	[318, 319]
CYTOR	Breast Cancer	Regulates endocrine resistance in BC	Acts as ceRNA to sponge miR125a5p/SRF, Hippo, MAPK pathways	[320]
HOTAIRM1	Breast Cancer	Promotes acquired tamoxifen resistance	Interacts with EZH2, preventing PRC2-mediated H3K27me3 of *HOXA1* promoter	[321]

Table A.2. A Summary of SNHGs implicated in cancer.

SNHG	Cancer	Role in Cancer	Mechanism	Ref
SNHG1	Breast Cancer	Oncogene; Promotes proliferation, cell cycle progression, inhibits migration. Contributes to cisplatin resistance.	Sponges miR-573 to upregulate LMO4. Regulates miR-641/RRS1 axis. Epigenetically silences miR-381.	[212,221,322]
GAS5 (SNHG2)	Breast Cancer (Tamoxifen-Resistant, ERα+)	Tumor Suppressor; Sensitizes to tamoxifen, modulates Adriamycin resistance.	Sponges miR-222 to upregulate PTEN. Targets miR-221-3p/DKK2/β-catenin pathway.	[318,323]
SNHG3	Breast Cancer	Oncogene; Promotes proliferation, migration, invasion, metabolic reprogramming. Associated with ERα status.	Sponges miR-384 → ↑HDGF; Sponges miR-154-3p → activates Notch pathway.	[324,325]
SNHG4	Non-Small Cell Lung Cancer,	Oncogene; Promotes proliferation, invasion	NSCLC: Sponges miR-let-7e → KDM3A.	[326]
SNHG5	Breast Cancer (Luminal A: MCF-7, T47D)	Oncogene; Upregulated in BC cell lines.Promotes proliferation, cell cycle progression, glycolysis; inhibits	Sponges miR-154-5p → ↑PCNA. Regulates BACH1 via targeting miR-299.	[327,328]

		apoptosis.		
SNHG6	Breast Cancer (TNBC)	Oncogene; Promotes proliferation, migration; inhibits apoptosis.	Sponges miR-125b-5p to upregulate BMPR1B.	[329]
SNHG6	Breast Cancer (Luminal A: MCF-7)	Oncogene; Upregulated; Promotes migration, EMT. Overexpression in MCF-7 led to G1 progression and reduced migration	Sponges miR-543 to upregulate LAMC1/PI3K/AKT pathway.	[330]
SNHG7	Breast Cancer	Oncogene; Promotes proliferation (G0/G1 arrest on silencing). Enhances trastuzumab resistance.	Controlled by Insulin-like Growth Factor (IGF) signaling. Regulates miR-186 (trastuzumab resistance).	[331,332]
SNHG8	Breast Cancer (TNBC)	Oncogene; Upregulated; Promotes proliferation, migration, EMT.	Sponges miR-335-5p to upregulate PYGO2.	[333]
SNHG9	Breast Cancer	Oncogene; Upregulated; Promotes tumor growth, YAP activity.	Directly interacts with LATS1, inhibits LATS1 kinase activity, leading to YAP activation.	[334]
SNHG9	Colorectal Cancer	Tumor Suppressor	Suppressed by microbiota; inhibits p53 activity by	[335]

			dissociating SIRT1 from CCAR2.	
SNHG11	Breast Cancer (TNBC)	Oncogene; Upregulated; Promotes malignant behaviors.	Negatively regulates miR-7-5p, positively regulates SP2, which binds MUC-1 promoter.	[336]
SNHG12	Breast Cancer (Luminal A: MCF-7)	Oncogene; Upregulated in BC tissues and cells (MCF-7). Knockdown inhibits proliferation, invasion.	Regulates miR-451a to suppress AKT/mTOR pathway.	[337]
SNHG14	Breast Cancer	Oncogene; Increased expression; Promotes proliferation, migration, invasion. trastuzumab resistance.	Regulates ERK/MAPK signaling pathway. Sponges miR-543 to regulate KLF7.	[338–340]
SNHG14	Breast Cancer	Oncogene; Accelerates progression, promotes tumorigenesis,		
SNHG15	Breast Cancer	Oncogene; Upregulated; Promotes proliferation, migration, invasion; inhibits apoptosis.	Sponges miR-211-3p.	[217]
SNHG16	Breast Cancer	Oncogene; Upregulated; Promotes cell migration.	Competitively binds miR-98 with E2F5.	[341]

SNHG17	Breast Cancer (ERα+)	Oncogene; Involved in Palbociclib/fulvestrant sensitivity.	Modulates ERα degradation via Hippo-YAP/LATS1	[269]
SNHG20	Breast Cancer	Oncogene; Upregulated; Promotes proliferation, invasion, migration.	Regulates HER2 via miR-495.	[342]
SNHG22	Breast Cancer (TNBC)	Oncogene; Enhances malignant phenotypes.	Sponges miR-324-3p to upregulate SUDS3.	[343]

B. Protocol for Direct RNA Sequencing

This is the detailed protocol of the methods used for the enrichment of polyadenylated RNA, followed by library construction using the Oxford Nanopore Technologies (ONT) Direct RNA Sequencing Kit (SQK-RNA002) and subsequent sequencing on a MinION device. The procedure is an adaptation of the Novoa Lab (CRG, Barcelona) protocol and the standard ONT protocol, optimized for maximal RNA yield and library quality from the starting material.

All solutions were prepared using nuclease-free water and were handled with strict RNase-free techniques to prevent RNA degradation. Stock solutions of Tris, LiCl, and EDTA were autoclaved prior to use. Final prepared buffers were also autoclaved, with the exception of any buffer containing SDS, as noted below.

B.1 General RNase-Free Handling Precautions

The integrity of RNA is essential for direct RNA sequencing. All buffers were prepared to be RNase-free. In-house preparation of buffers allows for complete control over all components and their quality. A critical procedural note is that the Lysis/Binding Buffer, which contains a detergent (LiDS or SDS), was not autoclaved. High temperatures can cause SDS to precipitate upon cooling and may lead to chemical degradation, compromising its function in cell lysis and protein denaturation. Therefore, this buffer was made sterile by passing it through a 0.2 µm filter.

B.2 Buffer Recipes

The following buffers were prepared in-house according to the compositions detailed below.

B.2.1 Binding Buffer

A 10 mL solution was prepared containing a final concentration of 20 mM Tris-HCl (pH 7.5), 1 M LiCl, and 2 mM EDTA. This was achieved by combining 200 µL of a 1 M Tris-HCl (pH 7.5) stock, 1250 µL of an 8 M LiCl stock, 40 µL of a 0.5 M EDTA (pH 8.0) stock, and 8510 µL of nuclease-free water.

B.2.2 Lysis/Binding Buffer

A 10 mL solution was prepared containing a final concentration of 100 mM Tris-HCl (pH 7.5), 500 mM LiCl, 10 mM EDTA (pH 8.0), 1% (w/v) LiDS or SDS, and 5 mM DTT. This was prepared by combining 1000 µL of 1 M Tris-HCl (pH 7.5), 625 µL of 8 M LiCl, 200 µL of 0.5 M EDTA (pH 8.0), 100 mg of LiDS or SDS, 500 µL of 100 mM DTT, and 7675 µL of nuclease-free water. The buffer was warmed to room temperature and mixed until all components were fully dissolved before use.

B.2.3 Washing Buffer B

A 10 mL solution was prepared containing a final concentration of 10 mM Tris-HCl (pH 7.5), 150 mM LiCl, and 1 mM EDTA (pH 8.0). This was achieved by combining 100 µL of 1 M Tris-HCl (pH 7.5), 187.5 µL of 8 M LiCl, 20 µL of 0.5 M EDTA (pH 8.0), and 9692.5 µL of nuclease-free water.

B.2.4 Oligo(dT)25 Storage Buffer

A 10 mL solution was prepared containing a final concentration of 250 mM Tris-HCl (pH 7.5), 20 mM EDTA, 0.1% (v/v) Tween-20, and 0.02% (w/v) NaN3. This was prepared by combining 2500 µL of 1 M Tris-HCl (pH 7.5), 400 µL of 0.5 M EDTA, 10 µL of Tween-20, 2 mg of NaN3, and adjusting the final volume to 10 mL with nuclease-free water.

B.2.5 Reconditioning Solution

A solution of 0.1 M NaOH was prepared in nuclease-free water.

B.2.6 10 mM Tris-HCl (pH 7.5)

A solution of 10 mM Tris-HCl (pH 7.5) was prepared by diluting 100 µL of a 1 M Tris-HCl (pH 7.5) stock in 9.9 mL of nuclease-free water. This solution was chilled and maintained at 4 degrees Celsius for use in elution steps.

B.3 Experimental Procedures

The following sections describe the complete experimental workflow from total RNA input to the final sequencing-ready library, presented in chronological order. All steps were performed in a designated sterile area with appropriate RNase-free handling techniques.

B.3.1 Initial RNA Quality Control and DNase Treatment

Prior to library preparation, total RNA samples, isolated using a standard method such as TRIzol extraction, were optionally assessed for integrity using an Agilent Bioanalyzer with an RNA 6000 Nano Kit. A critical preparatory step was the removal of contaminating genomic DNA (gDNA), which can interfere with quantification and downstream reactions.

A DNase treatment was performed using TURBO™ DNase. The treatment strategy was deliberately optimized to prioritize RNA integrity, which is essential for long-read sequencing applications that aim to characterize full-length transcripts. Standard protocols often involve lengthy incubations that increase the risk of RNA

degradation by co-purified RNases. To prevent this, a high-concentration, short-incubation method was employed. For a given volume of RNA sample (up to 200 ng/µL), TURBO™ DNase enzyme was added to a final concentration of 10% (v/v), along with 10% (v/v) of 10X TURBO DNase Buffer and 2% (v/v) of an RNase inhibitor. The reaction was incubated at 37°C for a brief period of 10–20 minutes. This approach minimizes the time the RNA is exposed to potentially degradative conditions while ensuring efficient removal of gDNA. Following incubation, the reaction was immediately subjected to purification using RNAClean XP beads.

B.3.2 Poly(A)+ RNA Enrichment using Oligo(dT) Magnetic Beads

A highly rigorous, multi-pass enrichment strategy was employed to maximize the recovery of poly(A)+ RNA from the total RNA pool. This custom procedure, which deviates significantly from standard single-pass protocols, was designed to enhance capture efficiency, particularly for low-abundance transcripts, and to concentrate the final product. All buffers were brought to room temperature before use, except for the 10 mM Tris-HCl elution buffer, which was kept at 4°C.

B.3.2.1 Preparation of Dynabeads® Oligo(dT)25

For each sample, 200 µL (1 mg) of Dynabeads® Oligo(dT)25 were transferred to a 1.5 mL tube. The beads were washed by placing the tube on a magnetic rack for 1–2 minutes, removing the supernatant, and resuspending the beads in 100 µL of Binding Buffer. This wash step was repeated once, and the beads were finally resuspended in 100 µL of fresh Binding Buffer.

B.3.2.2 First-Pass RNA Capture

A total of 75 µg of total RNA was adjusted to a volume of 100 µL with nuclease-free water. An equal volume (100 µL) of Binding Buffer was added, and the sample

was heated at 65°C for 2 minutes to denature RNA secondary structures. The sample was immediately placed on ice. A volume of 3 µL of SUPERase•In™ or Murine RNase Inhibitor was added. These specific inhibitors were selected for their documented stability at the 65°C denaturation temperature. The 200 µL RNA sample was then added to the 100 µL of washed beads. The mixture was annealed by continuous rotation on a tabletop rotator for 5–10 minutes at room temperature. The tube was then placed on the magnet for 2 minutes, and the supernatant was carefully collected and stored on ice for a subsequent capture round. The beads were washed twice with 200 µL of Washing Buffer B. Finally, the RNA was eluted by resuspending the beads in 10–20 µL of cold 10 mM Tris-HCl, incubating at 75–80°C for 2 minutes, and immediately placing the tube on the magnet to collect the eluate (Elution A).

B.3.2.3 Second-Pass RNA Capture from Supernatant

The supernatant retained from the first pass was denatured at 65°C for 3 minutes and placed on ice. The same beads used for the first elution were washed with 300 µL of Lysis/Binding Buffer, and the denatured supernatant was added back to these beads. The mixture was annealed for 5–10 minutes at room temperature with rotation. Following magnetic capture, the supernatant was discarded, and the beads were washed twice with 200 µL of Washing Buffer B. A second elution was performed as before, yielding Elution B, which was pooled with Elution A.

B.3.2.4 Final Re-binding and Elution

This multi-stage process was designed to maximize the yield of captured RNA molecules. The first pass captures the majority of poly(A)+ RNA. The second pass, by reprocessing the initial supernatant, recovers transcripts that may have failed to bind in the first round due to stochastic effects or incomplete denaturation. To further purify and concentrate the sample, the pooled eluates (A and B) were subjected to a

third binding cycle. The volume was adjusted to 100 µL, mixed with 100 µL of Binding Buffer, and denatured at 65°C for 2 minutes. This sample was then applied to the same set of beads (which had been washed and re-equilibrated in 100 µL of Binding Buffer). After a 5–10 minute binding incubation and two washes with Washing Buffer B, the final, purified RNA was eluted in 10–20 µL of cold 10 mM Tris-HCl into a clean LoBind Eppendorf tube. This final product (Elution C) represented the concentrated, high-purity RNA sample ready for library construction.

B.3.3 Direct RNA Sequencing Library Preparation (ONT SQK-RNA002)

The poly(A)+ RNA was used as input for library preparation following the ONT Direct RNA Sequencing protocol. This process involves the sequential ligation of adapters and a reverse transcription step, with intermediate purifications designed to remove contaminants and select for appropriate fragment sizes.

B.3.3.1 Reverse Transcription and First Adapter (RTA) Ligation

In a 0.2 mL PCR tube, the following reagents were combined in order: 9 µL of the purified RNA sample (mixed with water if necessary), 1 µL of the Reverse Transcription Adapter (RTA) oligo, 0.5 µL of RNA CS (ONT-supplied control strand), 3 µL of NEBNext Quick Ligation Reaction Buffer (5X), 0.5 µL of SUPERase•In™ RNase Inhibitor, and 1.5 µL of T4 DNA Ligase. The reaction was mixed by pipetting, briefly centrifuged, and incubated for 15 minutes at room temperature.

Following ligation, a reverse transcription master mix was prepared by combining 8.5 µL of nuclease-free water, 2 µL of 10 mM dNTPs, 8 µL of 5X First Strand Buffer (from the SuperScript III kit), and 4 µL of 0.1 M DTT. This 22.5 µL master mix was added to the ligation reaction. Finally, 2 µL of SuperScript III Reverse Transcriptase and 0.5 µL of RNasin® Promega RNase Inhibitor were added. The reaction was

incubated in a thermal cycler under the following conditions: 50°C for 50 minutes, followed by heat inactivation at 70°C for 10 minutes, and holding at 4°C.

B.3.3.2 First Purification (Post-RT)

The 40 µL reverse transcription product was transferred to a 1.5 mL LoBind tube. A volume of 72 µL (a 1.8X ratio) of resuspended Agencourt RNAClean XP beads was added, and the mixture was incubated for 5–10 minutes at room temperature. This specific bead ratio provides a bulk purification, effectively removing small molecular weight contaminants from the reaction, such as unincorporated dNTPs, salts, and residual adapter oligos, while retaining the larger cDNA:RNA hybrid molecules. The beads were captured on a magnet, and the pellet was washed twice with 200 µL of freshly prepared 70% ethanol without resuspension. After removing all residual ethanol, the bead pellet was air-dried for approximately 1 minute and the nucleic acids were eluted in 20 µL of nuclease-free water.

B.3.3.3 Second Adapter (RMX) Ligation

The 20 µL eluate from the previous step was used for the second ligation. To this, the following were added: 8 µL of NEBNext Quick Ligation Reaction Buffer (5X), 6 µL of RNA Adapter Mix (RMX), 3 µL of nuclease-free water, and 3 µL of T4 DNA Ligase. The reaction was mixed and incubated for 15 minutes at room temperature.

B.3.3.4 Final Library Purification

The final library was purified using a 1.0X volume of RNAClean XP beads. To the 40 µL ligation reaction, 40 µL of resuspended beads were added and incubated on a rotator for 5 minutes. This more stringent 1.0X ratio serves a critical size-selection function. It is specifically calibrated to remove small DNA fragments, most

importantly the adapter-dimer byproducts that form when RMX adapters ligate to each other. These short, non-library molecules are highly detrimental as they compete efficiently for nanopores during sequencing, thereby reducing useful data yield. This step represents a calculated trade-off, enriching for full-length library molecules at the cost of removing the very shortest native transcripts.

After magnetic capture, the bead pellet was washed twice with 150 µL of Wash Buffer (WSB, supplied in the ONT kit). The pellet was air-dried for 1 minute, and the final sequencing library was eluted in 21 µL of Elution Buffer (EB, supplied in the ONT kit). The concentration of the final library was quantified using both the Qubit RNA HS and dsDNA HS assays.

B.3.4 Flow Cell Priming and Library Loading

The prepared library was loaded onto a MinION R9.4.1 flow cell for sequencing. The protocol includes several experiential steps beyond the standard manufacturer's instructions, which are critical for preventing common failure modes such as the introduction of air bubbles.

First, the flow cell was removed from 4°C storage and subjected to a quality control check using the MinKNOW software to determine the number of active pores. The priming port was opened, and a P1000 pipette set to 200 µL was used to draw back a small volume (~20 µL) to remove any air from the port entrance. A priming mix, prepared by combining one tube of Flush Buffer (FB) with 30 µL of Flush Tether (FLT), was then loaded. A volume of 800 µL of this mix was introduced slowly into the priming port. To prevent bubble introduction, the first drop was expelled outside the port before the tip was fully inserted. After a 5-minute incubation, the SpotON sample port was opened, and an additional 200 µL of the priming mix was loaded into the priming port, allowing observation of buffer emerging from the sample port, confirming a complete flush.

The final library (21 µL) was diluted with 17.5 µL of nuclease-free water and mixed with 37.5 µL of Running Buffer (RRB). This 76 µL final mix was loaded "drop-by-drop" into the sample port to avoid fluidic shear stress on the library complexes and to prevent bubble formation. The ports were sealed, and the sequencing run was initiated in the MinKNOW software with the appropriate kit (RNA002) selected and live basecalling disabled.

B.4. Supporting Information: Rationale for RNase Inhibitor Selection

The selection of an appropriate RNase inhibitor is critical and depends on the specific enzymatic and thermal conditions of each step in the protocol. The choices made in this protocol were informed by the known properties of commercially available inhibitors, as summarized in the tables below. This demonstrates a reasoned approach to reagent selection, ensuring inhibitor activity and compatibility throughout the workflow.

For instance, the poly(A)+ selection step involves a 65°C denaturation. As shown in Table B.1, SUPERase•In™ and RNase Inhibitor, Murine, are active up to 65°C or higher, making them suitable choices, whereas other inhibitors may be denatured. Similarly, SUPERase•In™ and the Murine inhibitor do not require DTT for activity, making them versatile for various downstream applications, as detailed in Table B.2.

Table B.1: Comparative Analysis of RNase Inhibitor Specificity and Reaction Conditions

Compound	Brand	RNase A	RNase B	RNase C	RNase 1	RNase T1	Max Temp (°C)	pH Range	Requires DTT
RNaseOUT™	Thermo	✓	✓	✓	×	×	N/A	Acidic	✓ (1 mM)
RNasin®	Promega	✓	✓	✓	×	×	70	5.0–8.0	✓ (1 mM)
SUPERase•In™	Thermo	✓	✓	✓	✓	✓	65	5.5–8.5	×
RNase Inhibitor, Murine	NEB	✓	✓	✓	×	×	65	N/A	×

Table B.2: Compatibility of Common RNase Inhibitors with Downstream Enzymatic Applications

Compound	DNase	Ligation	IVT	cDNA Synthesis	Poly(A) Tailing	Ribodepletion	Poly(A) Selection
RNaseOUT™	✓	✓	×	×	×	×	×
RNasin® (Promega)	✓	✓	✓	✓	×	×	×
SUPERase•In™	✓	✓	✓	✓	✓	✓	✓
RNase Inhibitor, Murine	✓	✓	✓	✓	✓	✓	✓

B.4.1 Supporting Figure: Bead-based Size Selection

Different bead-to-sample volume ratios selectively precipitate nucleic acids above a certain size threshold. We used the bead size selection criteria from Broad/Illumina Genome Analyzer Boot Camp [344] to select our RNAClean XP bead ratio.

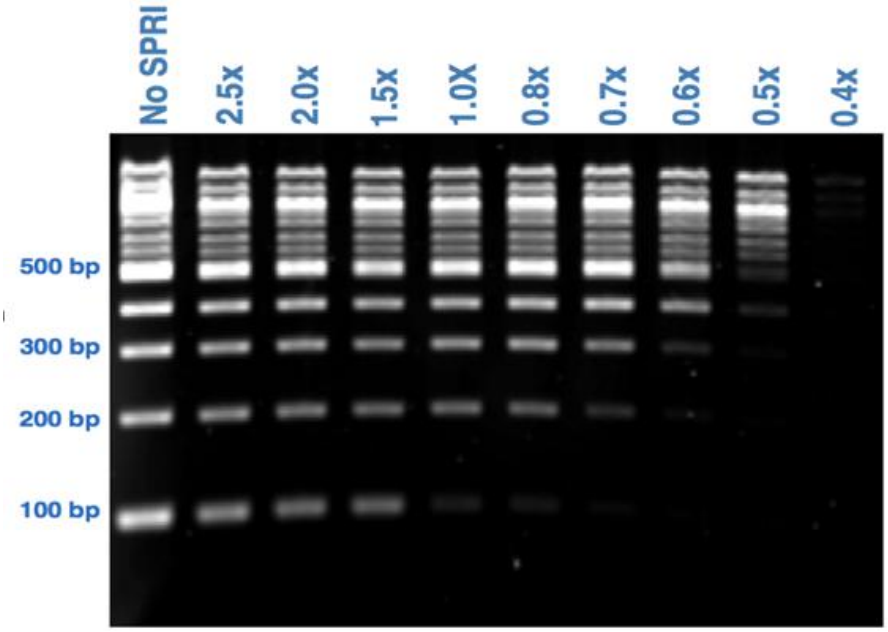


Figure B.1. This figure illustrates the principle of size selection using varying concentrations of Agencourt RNAClean XP beads [344]. The gel image compares the library profile after cleanup with a different bead ratio against a ladder (first lane). The 1.8X ratio retains a broad range of fragment sizes, making it suitable for the initial bulk purification step (see Section 2.3.2). In contrast, the more stringent 1.0X ratio effectively removes smaller fragments, which is critical for eliminating adapter-dimer byproducts during the final library purification (see Section 2.3.4). This demonstrates the importance of selecting the appropriate bead concentration to optimize the library for sequencing by enriching full-length transcripts and removing contaminants that would otherwise compete for nanopores.

B.5. Supporting Information: Materials and Equipment

This section details the commercial kits, chemical reagents, and laboratory equipment required for the successful execution of the protocol.

B.5.1 Kits and Reagents

- Direct RNA Sequencing Kit (Oxford Nanopore Technologies, cat. # SQK-RNA002)
- SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific, cat. # 18080044)
- 10 mM dNTP Solution (New England Biolabs, cat. # N0447)
- Concentrated T4 DNA Ligase, 2,000,000 U/mL (New England Biolabs, cat. # M0202)
- NEBNext® Quick Ligation Reaction Buffer (5X) (New England Biolabs, cat. # B6058)
- Agencourt RNAClean XP beads (Beckman Coulter)
- Dynabeads® Oligo (dT)25 (Thermo Fisher Scientific)
- TURBO™ DNase (Thermo Fisher Scientific)
- Qubit™ RNA HS Assay Kit (Thermo Fisher Scientific, cat. # Q32852)
- Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. # Q32851)
- RNaseOUT™ Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific)
- RNasin® Ribonuclease Inhibitor (Promega)
- SUPERase•In™ RNase Inhibitor (Thermo Fisher Scientific)
- RNase Inhibitor, Murine (New England Biolabs)
- Lithium Chloride (LiCl)
- Tris Base
- Ethylenediaminetetraacetic acid (EDTA)
- Lithium Dodecyl Sulfate (LiDS) or Sodium Dodecyl Sulfate (SDS)
- Dithiothreitol (DTT)
- Tween-20
- Sodium Azide (NaN3)
- Nuclease-free water (e.g., Thermo Fisher Scientific, cat. # AM9937)
- Ethanol (100%, molecular biology grade)
- 1.5 mL Eppendorf DNA LoBind tubes
- 0.2 mL thin-walled PCR tubes

Furthermore, the mandatory use of Eppendorf DNA LoBind tubes was a critical measure to maximize sample recovery. Standard polypropylene tubes can adsorb nucleic acids, leading to significant material loss, which is particularly detrimental during procedures involving multiple purification steps and low-volume elutions (e.g., 20 µL) as employed herein. The use of LoBind tubes mitigates this risk, ensuring the preservation of sample complexity and yield.

B.5.2 Laboratory Equipment

- Hula mixer (or equivalent gentle rotator mixer)
- Magnetic separation rack
- Microfuge (e.g., Eppendorf 5424 or equivalent)
- Vortex mixer
- Thermal cycler
- Calibrated micropipettes and tips (P1000, P200, P100, P20, P10, P2)
- Ice bucket and ice
- Laboratory timer
- Qubit Fluorometer (or equivalent for nucleic acid quantification)
- Agilent Bioanalyzer (or equivalent for RNA integrity analysis, optional)

C. Protocol for the Isolation of Nuclear and Cytoplasmic RNA

All buffers were prepared RNase-free to ensure the integrity of the RNA samples.

C.1. Buffers

C.1.1 Nuclei EZ Lysis Buffer

To prepare 10 mL of Nuclei EZ Lysis Buffer, 100 µL of 1 M Tris-HCl (pH 7.5), 300 µL of 2 M KCl, 20 µL of 0.5 M EDTA, 100 µL of 100X Protease Inhibitor, 50 µL of 100% NP-40, and 9,430 µL of nuclease-free water were combined. The final concentrations were 10 mM Tris-HCl, 60 mM KCl, 1 mM EDTA, 1X Protease Inhibitor, and 0.5% NP-40.

C.1.2 Nuclei Wash and Resuspension Buffer

To prepare approximately 5 mL of Nuclei Wash and Resuspension Buffer, 0.05 g (50 mg) of BSA powder was dissolved in 5,000 µL of 1X PBS to achieve a 1% BSA concentration. This buffer was supplemented with 25 µL of SUPERase•In™ or RNasin® Murine RNase inhibitor for a final concentration of 0.2 U/µL.

C.2. Experimental Procedure

All samples and reagents were kept on ice or at 4°C throughout the procedure. The cell pellet was lysed with 1500 µL of chilled Nuclei EZ Lysis Buffer, gently mixed with a wide bore tip, and incubated on ice for 5 minutes, with gentle mixing repeated during the incubation. The sample was centrifuged at 500g for 5 minutes at 4°C. The supernatant was removed, leaving approximately 50 µL, and the nuclei pellet was gently resuspended in another 1.5 mL of EZ Lysis Buffer for a second 5-minute incubation on ice. The nuclei were pelleted again at 500g for 5 minutes at 4°C. The

supernatant, which was the cytoplasmic fraction, was carefully removed and saved, leaving about 50 µL if the pellet was loose. The pellet was washed by adding 500 µL of Nuclei Wash and Resuspension Buffer and allowing it to sit for 5 minutes without resuspension, followed by the addition of 1 mL of wash buffer and gentle resuspension. The sample was centrifuged at 500g for 5 minutes at 4°C, the supernatant was removed, and the pellet was resuspended in 1.4 mL of Nuclei Wash and Resuspension Buffer, transferred to a 1.5 mL tube, and mixed by inversion. This wash step was repeated once more. The cleaned nuclei pellet was finally resuspended in 300 µL of supplemented Nuclei Wash and Resuspension Buffer. For RNA extraction, 500 µL of Trizol was added to the resuspended nuclei, shaken vigorously, and incubated at room temperature for 2 minutes. 200 µL of chloroform was added, the mixture was shaken vigorously, and centrifuged at 4°C for 15 minutes to separate the phases. The upper aqueous phase was carefully transferred to a new tube and mixed with an equal volume of 1X Isopropanol and Pellet Paint NF to precipitate the RNA. The mixture was incubated at room temperature for 15 minutes and the RNA was pelleted by centrifugation at maximum speed for 20 minutes at 4°C. The supernatant was discarded, and the pellet was washed with 70% ethanol with brief vortexing. The pellet was collected by centrifugation at maximum speed for 5 minutes at 4°C, all residual ethanol was carefully removed, and the pellet was air-dried for 2-3 minutes before resuspending the final, purified RNA in 50 µL of water.

D. Supplementary Figures and Tables for Chapter 3

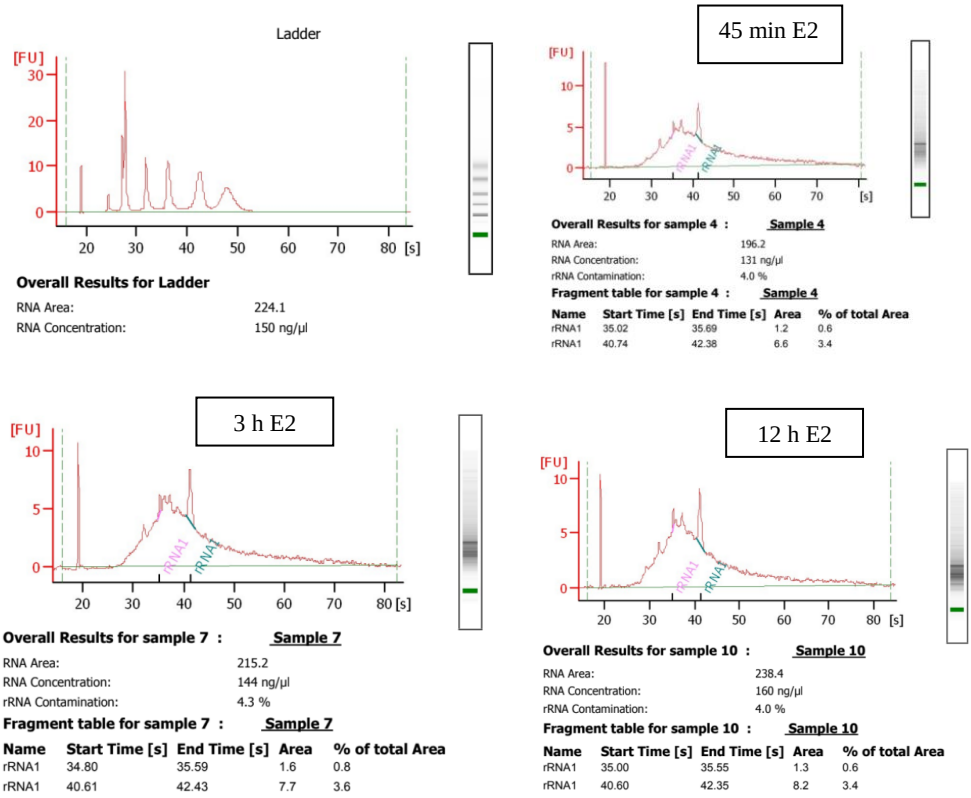


Figure D.1. Quality assessment of poly(A)-enriched RNA. An Agilent Bioanalyzer electropherogram of the 45 min E2, 3 h E2 and 12 h E2 RNA sample after poly(A) selection to enrich for messenger RNA (mRNA). The trace demonstrates successful rRNA depletion, characterized by the absence of dominant 18S and 28S rRNA peaks and the presence of a broad distribution of fragment sizes, consistent with a heterogeneous mRNA population. The analysis table confirms that residual 18S and 28S rRNA constitutes only ~2.5% of the total RNA, indicating the sample quality is suitable for downstream library preparation and sequencing.

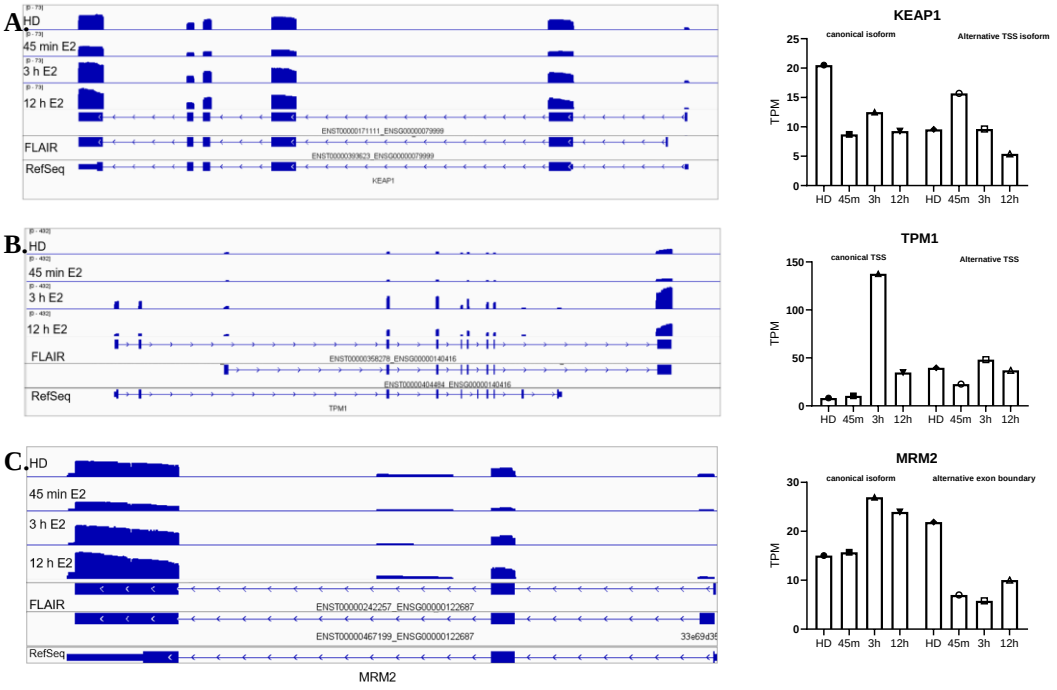


Figure D.2. Differential expression spliced isoforms for selected genes. This figure displays data for three example cases (*KEAP1*(A), *TPM1*(B), and *MRM2*(C)), illustrating changes in spliced isoform expression. For each gene, the left panel shows a genome browser view of the isoforms for each gene identified by the FLAIR pipeline with the annotated RefSeq gene structure for reference. The right panel quantifies the expression levels (in TPM) of spliced isoforms with different 3' UTRs across several time points.

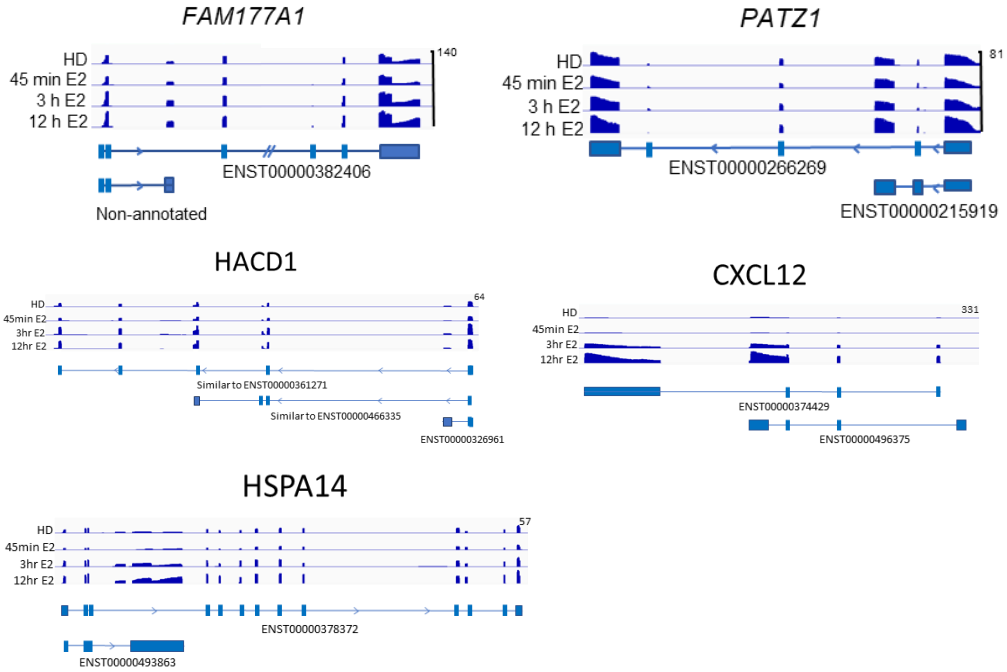


Figure D.3. IPA events for *FAM177A1*, *PATZ1*, *HACD1 CXCL12*, and *HSPA14*. Gene structure and isoforms detected by RNA long-read sequencing (FLAIR pipeline). The value in the top-right of each track indicates the maximum read coverage.

5’TCTTTTCCAGAGGCTAGATATATAGGAGGTTGTGGCTGCTCCAGCAGCT
TGGATTC**AGAGTGAGAAGGCATAAAGGAGAATCCCCAGCTGACTTGT**
GCAGTGGTTAATTGAAATTATTCAGGCAAGAGATGATGGTGTCTTGG
ACCAGGGGATGAGGAAGGCTACAAAATGTGTCTACCTGTATTCTGT
GAGGAGAACGTGTTCCCTGGTTTTAGATACTGTGAAGATGGATCAG
GAGAGAGTTTATCTAGACTGTTGGGGAAAGGTGTTGCGATTCCTTC
AGCTACACAGGATTGAAAGGAGACATTTCTGAAGGGGAAAAAGGAA
ATGAAAGAAAAGATGTTTCAGATTGAGGATATGCTGTGTGGTGAACT
TGTTCTTCACTCTGTAGGGTTCACAAATGACTCTTCACTGCCCTCTT
GGATGAAATAAACTGGTTCCCATAGAAATGGACCGTCTCTGATTTCA
CAGTCTAGAGCACACAAATAACTCAACTTTCTGGAAGCAAGGGTGA
TGGCTCGTGAGAGCTCAAGGACAAGGAGGAGTCTGTAGAAGCACA
CCTGGACTCGGGGAAATCATTGACCTCTCATCCTGAATTAAGTAGCG
CTTGAAAAGACTCTCTTTGAATCAGAAAAATAATGCGAGCTCCTATC
AAACTACCTTACACAGCCAACTGAGTGATCATAGGCTTGGGTGGAA
AAAATGTAATTCGTAGAACCTTTTAAGAGCTAAGATTCCTTAATCAAC
ATGGATACCGTCTTGTTGAATGACACCGAACTGTGCTGCGCAGGAT
TCGGTTGCATGGTGGCTGGAGCTGCCTCTCTGGGACCCAAATCTCA
GACCTGCTGCAAAGTGACTATATGATCTGCAACTTGCTAACCCTTCT
AGGCTTTTGTTTTATTATTTACATAATGAAGTGAGGGGAATATGTTCC
TTGTAAATTTGTTAAGAGGATTAAATTAGTTCATGTCAAATGATTTAG
AACACTGTCTGCCACATATTAAGGACTCAACAAATACTACCTATTATT
AACAATGTGCCCTAAATCTGTAAATCACTCAATAATACATGGCATTAG
GCTAAATATGGAAGGTAGAAGGAAATAAATTGTCCAGAAGCTAAAAA
TATGAAAGACTCTTCATATATTTAAAATATTTTCTTTAATTCTTTTGTT
TTTTTCTTTTTTGAGACAGAGTCTCGCTCCGTCACCCAGGCTGGAG
TGCAGTGGCATGTTCTCGGCTCGCTGCAGCCTCTGCTTCCTGGATT
CAAGTGATTCTCCTTCCTCAGCCTCCCAAGTAGCTGGAATTACAGGC
ATGTGCCAACATGCCGAGCTAATTTTTTTATTTTTTGTAGAGATGGG
GTTTCACCATGTTGCCCAGGCTGGTTTCAAACTCCTGACCTCAAGT
GAGCCACCCACCTTGGCCTTCCAAAGTGCTGAGATTACAGGCTTGA
GTCACTGTGCCCAGCCTGAAATATCTTCTTCACATCTTTTCATATGCATA
TTTGTTTTTACAAGTACATAAATATACATTTTTGATATAACTTGGAATCTTG
CATTTTCTAACTCAATGTAAGTATTTTTCTTCTCTTTTACCTATATAAATATA
AATATATATATATATACTTAGATGATATATTTACATACTTATTTTACTTAATAA
GTATTTTTCTATTATCAATAGTTAAAATTCCACTATGGTTTATGCCATTTATT
AAATGTCACTTTCCCTTCTCAAAAAA **3’**

Figure D.4. Sequence difference between luminal-specific lncRNA (*LOL*) and the sequence we obtained with DRS for *lnc-GATA3-7*. *LOL* was described as a 1414 bp transcript in MCF7 cells (indicated as bold). Our analysis DRS, has revealed a

longer, 1740 bp isoform. This newly identified isoform shares a similar structure with LOL but extends an additional 56 bp at the 5' end and 271 bp at the 3' end.

Table D.1. snoRNA Database (snoDB 2.0) results of DEG SNHGs.

snoDB ID	**Symbol**	**Box Type**	**Host Gene**	**Target Count**	**Target Types**
snoDB0297	*SNORD48*	C/D	SNHG32	30	rRNA;snRNA;Others
snoDB0484	*SNORD22*	C/D	SNHG1	25	rRNA;snRNA;Others
snoDB0679	*SNORD116-9*	C/D	*SNHG14*	23	rRNA;Others
snoDB0488	*SNORD27*	C/D	SNHG1	20	rRNA;snRNA;Others
snoDB0987	*SNORD47*	C/D	*GAS5*	20	rRNA;snRNA;Others
snoDB0257	*SNORA74A*	H/ACA	*SNHG4*	18	rRNA;snRNA;Others
snoDB0490	*SNORD25*	C/D	SNHG1	17	rRNA;snRNA;Others
snoDB1021	*SNORD50A*	C/D	*SNHG5*	17	rRNA;Others
snoDB0070	*SNORA73B*	H/ACA	SNHG3	15	Others
snoDB0900	*SNORD12C*	C/D	ZFAS1	13	rRNA;snRNA;Others
snoDB0822	*SNORD7*	C/D	SNHG30	13	rRNA;snRNA;Others
snoDB0069	*SNORA73A*	H/ACA	SNHG3	12	rRNA;snRNA;Others
snoDB0721	*SNORD115-11*	C/D	*SNHG14*	12	rRNA;Others
snoDB0988	*SNORD80*	C/D	*GAS5*	12	rRNA;snRNA;Others
snoDB0810	*SNORD49A*	C/D	SNHG29	11	rRNA;snRNA;Others
snoDB0669	*SNORD64*	C/D	*SNHG14*	10	snRNA;Others
snoDB0671	*SNORD109A*	C/D	*SNHG14*	9	snRNA;Others
snoDB0855	*SNORD104*	C/D	SNHG25	8	rRNA;snRNA;Others
snoDB04	*SNORD28*	C/D	SNHG1	8	rRNA;Others

87					
snoDB0298	*SNORD52*	C/D	SNHG32	8	rRNA;Others
snoDB0319	*SNORA9*	H/ACA	*SNHG15*	7	rRNA;snRNA;Others
snoDB0670	*SNORD108*	C/D	*SNHG14*	7	Others
snoDB0902	*SNORD12*	C/D	ZFAS1	7	rRNA;Others
snoDB0901	*SNORD12B*	C/D	ZFAS1	7	rRNA;Others
snoDB0890	*SNORA60*	H/ACA	*SNHG11*	6	rRNA;snRNA;Others
snoDB0892	*SNORA71A*	H/ACA	*SNHG17*	6	rRNA;Others
snoDB0256	*SNORA74D*	H/ACA	*SNHG4*	6	rRNA;snRNA;Others
snoDB0778	*SNORA78*	H/ACA	*SNHG9*	6	rRNA;Others
snoDB0994	*SNORD76*	C/D	*GAS5*	6	rRNA;Others
snoDB0598	*SCARNA13*	scaRNA	*SNHG10*	5	snRNA
snoDB0864	*SCARNA17*	scaRNA	*SNHG22*	5	snRNA
snoDB0206	*SNORA24*	H/ACA	*SNHG8*	5	rRNA
snoDB0073	*SNORA44*	H/ACA	*SNHG12*	5	rRNA;Others
snoDB0877	*SNORA71B*	H/ACA	*SNHG17*	5	rRNA;Others
snoDB0246	*SNORA13*	H/ACA	EPB…-AS1	4	rRNA;Others
snoDB0968	*SNORA17A*	H/ACA	*SNHG7*	4	rRNA;Others
snoDB0958	*SNORA50C*	H/ACA	SNHG25	4	rRNA;Others
snoDB0072	*SNORA61*	H/ACA	*SNHG12*	4	rRNA;Others
snoDB0893	*SNORA71C*	H/ACA	*SNHG17*	4	rRNA;snRNA
snoDB0894	*SNORA71D*	H/ACA	*SNHG17*	4	rRNA;Others
snoDB0304	*SNORD50B*	C/D	*SNHG5*	4	rRNA;Others
snoDB0216	*SNORA26*	H/ACA	*DANCR*	3	rRNA
snoDB01	*SNORA70I*	H/ACA	SNHG31	3	rRNA

13					
snoDB0889	*SNORA71E*	H/ACA	*SNHG11*	3	rRNA;snRNA
snoDB0711	*SNORD115-1*	C/D	*SNHG14*	3	rRNA;Others
snoDB0720	*SNORD115-10*	C/D	*SNHG14*	3	rRNA;Others
snoDB0722	*SNORD115-12*	C/D	*SNHG14*	3	rRNA;Others
snoDB0723	*SNORD115-13*	C/D	*SNHG14*	3	rRNA;Others
snoDB0724	*SNORD115-14*	C/D	*SNHG14*	3	rRNA;Others
snoDB0725	*SNORD115-15*	C/D	*SNHG14*	3	rRNA;Others
snoDB0726	*SNORD115-16*	C/D	*SNHG14*	3	rRNA;Others
snoDB0731	*SNORD115-21*	C/D	*SNHG14*	3	rRNA;Others
snoDB0732	*SNORD115-22*	C/D	*SNHG14*	3	rRNA;Others
snoDB0736	*SNORD115-26*	C/D	*SNHG14*	3	rRNA;Others
snoDB0738	*SNORD115-28*	C/D	*SNHG14*	3	rRNA;Others
snoDB0739	*SNORD115-29*	C/D	*SNHG14*	3	rRNA;Others
snoDB0740	*SNORD115-30*	C/D	*SNHG14*	3	rRNA;Others
snoDB0741	*SNORD115-31*	C/D	*SNHG14*	3	rRNA;Others
snoDB0743	*SNORD115-33*	C/D	*SNHG14*	3	rRNA;Others
snoDB0744	*SNORD115-34*	C/D	*SNHG14*	3	rRNA;Others
snoDB0745	*SNORD115-35*	C/D	*SNHG14*	3	rRNA;Others
snoDB0746	*SNORD115-36*	C/D	*SNHG14*	3	rRNA;Others
snoDB0748	*SNORD115-38*	C/D	*SNHG14*	3	rRNA;Others
snoDB0749	*SNORD115-39*	C/D	*SNHG14*	3	rRNA;Others
snoDB0714	*SNORD115-4*	C/D	*SNHG14*	3	rRNA;Others
snoDB0750	*SNORD115-40*	C/D	*SNHG14*	3	rRNA;Others
snoDB07	*SNORD115-*	C/D	*SNHG14*	3	rRNA;Others

51	*41*				
snoDB0752	*SNORD115-42*	C/D	*SNHG14*	3	rRNA;Others
snoDB0753	*SNORD115-43*	C/D	*SNHG14*	3	rRNA;Others
snoDB0754	*SNORD115-44*	C/D	*SNHG14*	3	rRNA;Others
snoDB0715	*SNORD115-5*	C/D	*SNHG14*	3	rRNA;Others
snoDB0716	*SNORD115-6*	C/D	*SNHG14*	3	rRNA;Others
snoDB0718	*SNORD115-8*	C/D	*SNHG14*	3	rRNA;Others
snoDB0719	*SNORD115-9*	C/D	*SNHG14*	3	rRNA;Others
snoDB0844	*SNORD1A*	C/D	SNHG16	3	rRNA;Others
snoDB0843	*SNORD1B*	C/D	SNHG16	3	rRNA;Others
snoDB0811	*SNORD65*	C/D	SNHG29	3	rRNA;Others
snoDB0996	*SNORD74*	C/D	*GAS5*	3	rRNA
snoDB0342	*SNORD93*	C/D	SNHG26	3	rRNA;Others
snoDB0071	*SNORD99*	C/D	*SNHG12*	3	rRNA;Others
snoDB0771	*SCARNA15*	scaRNA	*SNHG21*	2	snRNA
snoDB0845	*SCARNA16*	scaRNA	*SNHG20*	2	snRNA
snoDB0966	*SNORA16A*	H/ACA	*SNHG12*	2	rRNA
snoDB0971	*SNORA17B*	H/ACA	*SNHG7*	2	rRNA
snoDB0712	*SNORD115-2*	C/D	*SNHG14*	2	rRNA
snoDB0730	*SNORD115-20*	C/D	*SNHG14*	2	rRNA
snoDB0734	*SNORD115-24*	C/D	*SNHG14*	2	rRNA
snoDB0735	*SNORD115-25*	C/D	*SNHG14*	2	rRNA
snoDB0737	*SNORD115-27*	C/D	*SNHG14*	2	rRNA
snoDB0713	*SNORD115-3*	C/D	*SNHG14*	2	rRNA
snoDB07	*SNORD115-*	C/D	*SNHG14*	2	rRNA

42	*32*				
snoDB0756	*SNORD115-46*	C/D	*SNHG14*	2	rRNA
snoDB0717	*SNORD115-7*	C/D	*SNHG14*	2	rRNA
snoDB0226	*SNORD123*	C/D	*SNHG18*	2	rRNA
snoDB0842	*SNORD1C*	C/D	SNHG16	2	rRNA
snoDB0489	*SNORD26*	C/D	SNHG1	2	rRNA
snoDB0809	*SNORD49B*	C/D	SNHG29	2	rRNA
snoDB0962	*SNORD60*	C/D	*SNHG19*	2	rRNA;Others
snoDB0419	*SNORD87*	C/D	*SNHG6*	2	rRNA
snoDB0759	*SNORD109B*	C/D	*SNHG14*	1	snRNA
snoDB0757	*SNORD115-47*	C/D	*SNHG14*	1	rRNA
snoDB0664	*SNORD116-10*	C/D	*SNHG14*	1	rRNA
snoDB0665	*SNORD116-11*	C/D	*SNHG14*	1	rRNA
snoDB0680	*SNORD116-12*	C/D	*SNHG14*	1	rRNA
snoDB0681	*SNORD116-13*	C/D	*SNHG14*	1	rRNA
snoDB0682	*SNORD116-14*	C/D	*SNHG14*	1	rRNA
snoDB0683	*SNORD116-15*	C/D	*SNHG14*	1	rRNA
snoDB0684	*SNORD116-16*	C/D	*SNHG14*	1	rRNA
snoDB0666	*SNORD116-17*	C/D	*SNHG14*	1	rRNA
snoDB0685	*SNORD116-18*	C/D	*SNHG14*	1	rRNA
snoDB0686	*SNORD116-19*	C/D	*SNHG14*	1	rRNA
snoDB0687	*SNORD116-20*	C/D	*SNHG14*	1	rRNA
snoDB0688	*SNORD116-21*	C/D	*SNHG14*	1	rRNA
snoDB0689	*SNORD116-22*	C/D	*SNHG14*	1	rRNA
snoDB06	*SNORD116-*	C/D	*SNHG14*	1	rRNA

90	*23*				
snoDB0691	*SNORD116-24*	C/D	*SNHG14*	1	rRNA
snoDB0708	*SNORD116-29*	C/D	*SNHG14*	1	rRNA
snoDB0676	*SNORD116-6*	C/D	*SNHG14*	1	Others
snoDB1049	*SNORD31*	C/D	SNHG1	1	rRNA
snoDB0991	*SNORD44*	C/D	*GAS5*	1	rRNA
snoDB0995	*SNORD75*	C/D	*GAS5*	1	rRNA
snoDB0993	*SNORD77*	C/D	*GAS5*	1	rRNA
snoDB0990	*SNORD78*	C/D	*GAS5*	1	rRNA
snoDB0992	*SNORA103*	H/ACA	*GAS5*		
snoDB0668	*SNORD107*	C/D	*SNHG14*		
snoDB0727	*SNORD115-17*	C/D	*SNHG14*		
snoDB0728	*SNORD115-18*	C/D	*SNHG14*		
snoDB0729	*SNORD115-19*	C/D	*SNHG14*		
snoDB0733	*SNORD115-23*	C/D	*SNHG14*		
snoDB0747	*SNORD115-37*	C/D	*SNHG14*		
snoDB0755	*SNORD115-45*	C/D	*SNHG14*		
snoDB0758	*SNORD115-48*	C/D	*SNHG14*		
snoDB0672	*SNORD116-1*	C/D	*SNHG14*		
snoDB0673	*SNORD116-2*	C/D	*SNHG14*		
snoDB0692	*SNORD116-25*	C/D	*SNHG14*		
snoDB0693	*SNORD116-26*	C/D	*SNHG14*		
snoDB0694	*SNORD116-27*	C/D	*SNHG14*		
snoDB0695	*SNORD116-28*	C/D	*SNHG14*		
snoDB06	*SNORD116-*	C/D	*SNHG14*		

74	*3*				
snoDB0709	*SNORD116-30*	C/D	*SNHG14*		
snoDB0675	*SNORD116-4*	C/D	*SNHG14*		
snoDB0663	*SNORD116-5*	C/D	*SNHG14*		
snoDB0677	*SNORD116-7*	C/D	*SNHG14*		
snoDB0678	*SNORD116-8*	C/D	*SNHG14*		
snoDB0043	*SNORD64L1*	C/D	SNHG28		

A.

Observed Sequence	Predicted TF	Consensus Sequence	DEG SNHG Count	Source SNHGs
GGGATTTCC	NF-κB	GGGRNYYCC	18	*SNHG9*, *SNHG15*, *SNHG17*
TGATAA	GATA	WGATAR	3	*SNHG15*
TATTTAT	FOXA1	TRTTKRY	13	*SNHG17* (12)
TGAGTCA	AP1	TGASTCA	16	*SNHG17* (4)
TGGGCGAAT	SP1	KGGGCGRRY	17	*SNHG15*, *SNHG17*
AGGTCATTTTGACCT	ERE	AGGTCANNNTGACCT	16	*SNHG9*, *SNHG17*

B.

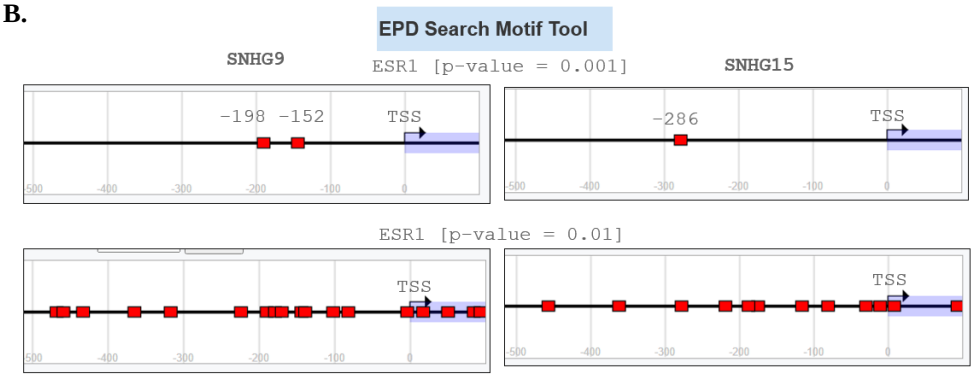
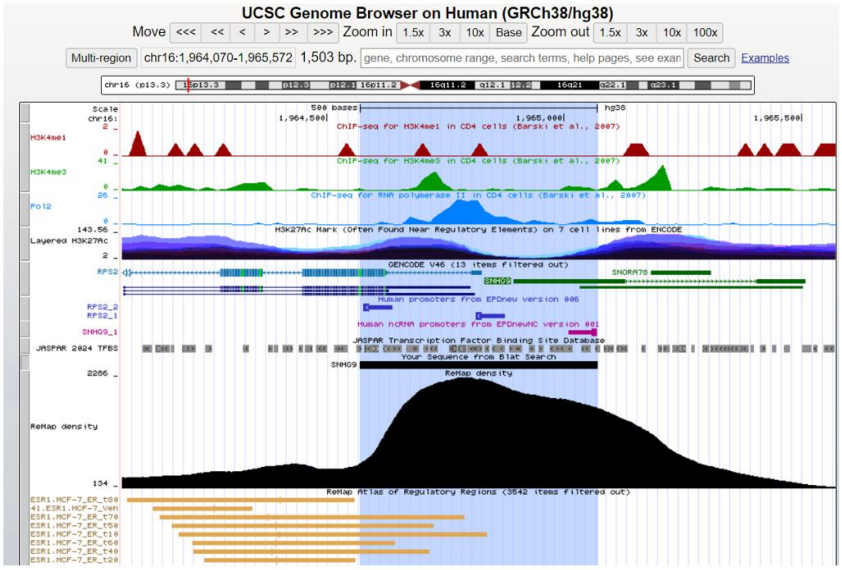


Figure D.5. (cont'd)

C.

SNHG9



SNHG15

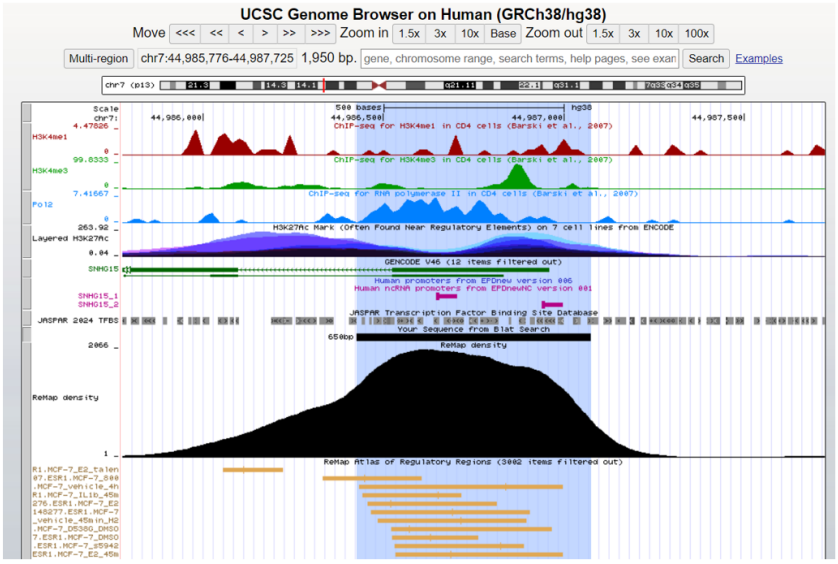


Figure D.5. (cont'd)

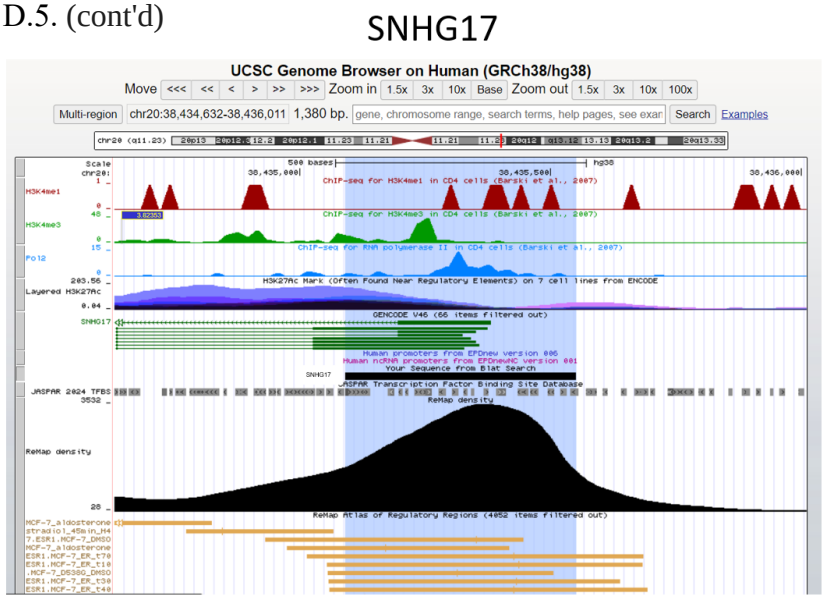


Figure D.5. Analysis of potential regulatory elements in the promoter regions of *SNHG9*, *SNHG15*, and *SNHG17*. (A) Results from a FIMO analysis showing the predicted motifs and counts for the Estrogen Response Element (ERE) and its partner transcription factors. (B) Predicted ESR1 binding sites upstream of the transcription start sites (TSS) for *SNHG9* and *SNHG15*, as identified by the EPD motif search tool. (C) This UCSC Genome Browser image displays the predicted promoter regions. The colocalization of multiple genomic features provides evidence that this region functions as an active promoter. There is an enrichment for active promoter histone modifications, including H3K4me3 (green peaks) and H3K27Ac (layered blue and purple peaks). The flanking H3K4me1 signal (red peaks) is typical of nearby enhancer elements. A peak from ChIP-seq for RNA Polymerase II (Pol2, blue) confirms the presence of the core transcription machinery. A DNase I hypersensitivity cluster (light blue bar) indicates the chromatin is in an open, accessible state, allowing regulatory factors to bind. Data from CAGE and FANTOM5 (blue vertical bars) map the transcription start sites to this location, providing direct evidence of gene expression initiation. Finally, ChIP-seq peaks for ESR1 (yellow lines) confirm the binding of this transcription factor in the selected regions.

Table D.2. DIANA-LncBase v3 Localization Predictions for Selected SNHG Genes.

Cell Type	Gene Name / Transcript	Nucleus TPM	Cytoplasm TPM	RCI
MCF7	*DANCR*/ENST00000411630	0.33035	1.01915	1.62531
MCF7	*SNHG17*/ENST00000414142	29.97125	2.78371	-3.4285
MCF7	*SNHG17*/ENST00000423536	40.54029	18.29609	-1.14782
MCF7	*SNHG17*/ENST00000424235	9.05088	0.75766	-3.57844
MCF7	*DANCR*/ENST00000425653	79.92041	126.24102	0.65954
MCF7	*DANCR*/ENST00000441504	7.73313	0.49716	-3.95926
MCF7	*SNHG17*/ENST00000449469	9.39501	0.44868	-4.38813
MCF7	*NEAT1*/ENST00000499732	136.87025	12.15469	-3.49322
MCF7	*NEAT1*/ENST00000501122	44.04444	0.72076	-5.93329
MCF7	*SNHG9*/ENST00000531523	7.88062	43.2087	2.45494
MCF7	*SNHG9*/ENST00000564014	13.29925	4.5679	-1.54174
MCF7	*SNHG15*/ENST00000577700	14.29234	3.57549	-1.99903
MCF7	*SNHG15*/ENST00000580458	9.19345	16.26382	0.82299
MCF7	*SNHG15*/ENST00000580528	1.0267	0.27645	-1.89291
MCF7	*SNHG15*/ENST00000582727	18.89122	0.97615	-4.27447
MCF7	*SNHG15*/ENST00000585030	1.64318	3.63247	1.14446
MCF7	*NEAT1*/ENST00000601801	19.80271	7.66554	-1.36924
MCF7	*NEAT1*/ENST00000612303	712.10216	45.84181	-3.95735
MCF7	*DANCR*/ENST00000623036	2.74164	0.19295	-3.82871
MCF7	*NEAT1*/ENST00000645023	723.09044	34.98175	-4.3695
MCF7	*NEAT1*/ENST00000646243	178.4568	4.69671	-5.24778
MCF7	*SNHG17*/ENST00000651115	3.99994	1.73708	-1.20332
MCF7	*SNHG17*/NR_015366.5	3.15639	2.87467	-0.13488
MCF7	*SNHG17*/NR_152753.1	2.83662	0.17806	-3.99376
MCF7	*SNHG17*/NR_152759.1	2.04931	0.13174	-3.95936
MCF7	*SNHG17*/NR_152763.1	3.29369	0.02293	-7.16617

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Özgül, İbrahim

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Molecular Biology and Genetics	2019
BS	METU Molecular Biology and Genetics	2016
High School	İzmir Bornova Anadolu High School	2011

WORK EXPERIENCE

Year	Place	Enrollment
2018-Present	METU Dept. of Molecular Biology and Genetics	Research Assistant

FOREIGN LANGUAGES

Advanced English, Elementary French

PUBLICATIONS

1. Dioken D.N.*, **Özgül İ.*** and Erson-Bensan A. E. "The 3′ end of the tale—neglected isoforms in cancer", FEBS Letters (2025) *** co-first authors**

2. Dioken D.N. *, **Özgül İ.***, Yilmazbilek I., Yakicier M.C., Karaca E. and Erson Bensan A. E. "An alternatively spliced PD-L1 isoform PD-L1Δ3, and PD-L2 expression in breast cancers: implications for eligibility scoring and immunotherapy response", Cancer Immunology, Immunotherapy, 72(12), 4065-4075 (2023)

3. Dioken D.N., **Özgül İ**., Köksal Bıçakcı G., Gol K., Can T. and Erson Bensan A. E. "Differential expression of mRNA 3′-end isoforms in cervical and ovarian cancers", Heliyon, 9(9) (2023) *** co-first authors**

4. Erdem M.*, **Özgül İ.***, Dioken D.N., Gurcuoglu I., Guntekin Ergun S., Cetin-Atalay R. et al. "Identification of an mRNA isoform switch for HNRNPA1 in breast cancers", Scientific Reports, 11(1) (2021) *** co-first authors**

5. Erdem M., **Özgül İ.** and Erson-Bensan A. E. "HNRNPA1 (Heterogeneous nuclear ribonucleoprotein A1)", Atlas of Genetics and Cytogenetics in Oncology and Haematology, 23(6), 137-142 (2019)