

SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

**Genome-wide investigation of COUP-TFII orphan nuclear receptor binding
events in cancer cells**

by Edina Erdős

Supervisor: Dr. Bálint László Bálint



UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLEKULAR CELL AND IMMUNE BIOLOGY

DEBRECEN, 2020

Genome-wide investigation of COUP-TFII orphan nuclear receptor binding events in cancer cells

By **Edina Erdős**, MSc

Supervisor: Dr. Bálint László Bálint, MD, PhD

Doctoral School of Molecular Cell and Immune Biology, University of Debrecen

Head of the **Examination Committee**: Prof. Dr. László Fésüs, PhD, DSc, MHAS

Members of the Examination Committee: Dr. Tibor Pankotai, PhD
Dr. Béla Nagy, PhD

The Examination will be held at 11 a.m., 15th of July, 2020.

Head of the **Defense Committee**: Prof. Dr. László Fésüs, PhD, DSc, MHAS

Reviewers: Dr. Attila Patócs, PhD, DSc
Dr. András Penyige, PhD

Members of the Defense Committee: Dr. Tibor Pankotai, PhD
Dr. Béla Nagy, PhD

The online PhD Defense will be held at 2 p.m., 15th of July, 2020.

Live online access will be provided. If you wish to take part in the discussion, please send an e-mail to edina.erdos@med.unideb.hu not later than 4 p.m. on the day before the discussion (14th of July, 2020). After the deadline, for technical reasons, it is no longer possible to join in to the defense.

1. INTRODUCTION AND THEORETICAL BACKGROUND

1.1. Introduction

There are many messenger molecules in our body that have hydrophobic properties, meaning they are poorly soluble in water. Carried proteins (such as globulins) are required for the distribution of these molecules, but due to their lipid soluble properties, they readily cross the plasma membrane and bind to specific intracellular receptors within the cell. Some of these receptors are DNA binding proteins and act as transcription factors. Their main function is to respond to environmental stimuli at the level of gene expression. These receptors are collectively referred to as nuclear receptors. Nuclear receptors are involved in several biological processes such as growth and embryonic regulation, phenotype maintenance, and regulation of metabolic processes (cholesterol, bile acid and fatty acid metabolism). Dysfunction in these processes can lead to infertility, obesity, diabetes and even the tumor development.

After cardiovascular disease, cancer is the second leading cause of death worldwide (WHO, 2018). According to the Central Statistical Office in Hungary, the number of deaths from malignant tumors and cardiovascular diseases is nearly the same (~ 33,000 deaths per year). Lung cancer was the most common cause of cancer in men, while breast cancer in women (WHO, 2016). The number of these deaths is increasing every year, despite of the increasingly effective therapeutic interventions.

Our work aimed to investigate the role of the COUP-TFII orphan nuclear receptor by functional genomic methods in different tumor cells, especially focusing on breast cancer.

1.2. Nuclear receptors

Nuclear receptors are ligand-activated transcription factors. Their ligands are hydrophobic messenger molecules and also present in our body, such as thyroid hormones, retinoids,

steroids, vitamin D, fatty acids, bile acids and xenobiotics. Nuclear receptors are highly conserved in their structure and they consist of similar functional domains. These domains are the highly variable N-terminal domain (A/B), the DNA binding domain (DBD) with two zinc finger motifs, the D domain and the C-terminal domain (E/F), which includes the ligand-binding domain (LBD).

Nuclear receptors can be divided into two groups based on their function. Receptors of the first group, are predominantly steroid receptors which move between the cytoplasm and the nucleus. The other group receptors are permanently located in the nucleus. Steroid receptors dimerize with heat shock proteins and bind their ligand in the cytoplasm. Ligand binding causes conformational changes in the nuclear receptor bridge and LBD, leading to facilitate their translocation to the nucleus. Nuclear receptors are detached from the carrier proteins to bind to specific DNA response elements in the nucleus and regulate transcription as homodimers. After binding to DNA, many proteins will bind to the nuclear receptors, collectively known as co-regulators. Another group is the nuclear receptors (RAR, PPAR), which often bind to DNA response elements in the absence of ligand, together with co-repressors and chromatin-modifying enzymes to maintain DNA in a less accessible state. Ligand binding results in a conformational change within the nuclear receptor that replaces the co-repressors with co-activators, creating a transcriptionally active state.

1.3. Estrogen receptor alpha (ER α) nuclear receptor

Estrogen receptor alpha (ER α) is a steroid hormone nuclear receptor whose natural ligand is estrogen. Estrogen is composed of estrone (E1), estradiol (E2) and estriol (E3), of which 17 β -estradiol is the most prominent in the circulation. 17 β -estradiol binds with the highest affinity to the estrogen receptor.

Estradiol (E2) acts through a mechanism specific to steroid receptors, it binds directly to the genome via the ERE motif. Besides, it is known that it is regulated via other transcription

factors; bound to the FOS/JUN dimer via the AP-1 motif. Non-genomic regulation of ER α is also known through several ways. For instance, a cell membrane associated with G protein-coupled ER (GPER) binds E2 and activates other signaling pathways to trigger a transcriptional response. Alternatively, when a ligand-independent activation occurs in the absence of E2 and the signal is derived from other signal transduction pathways, such as EGF or IGF1, in the form of phosphorylation.

Hisham Mohammed et al. in 2013 identified ER α -associated cofactors using rapid immunoprecipitation following mass-spectrometry of endogenous proteins (RIME). Among the 108 cofactors identified well known co-factors such as FoxA1, TLE1, AP2- γ , GATA3, p300, NCOA3, NRIP1 and RAR. Besides, less known transcriptional association with ER α was found like GREB1 and COT2 (COUP-TFII). Since there is limited knowledge about the transcriptional association between ER α and COUP-TFII, one of the aims of our work is to study this transcriptional association.

1.4. COUP-TFII orphan nuclear receptor

The chicken ovalbumin upstream promoter transcription factor II (COUP-TFII, also known as NR2F2, ARP-1, COT2) is a member of the steroid/thyroid nuclear receptor family. COUP-TF proteins were first isolated from HeLa nuclear extract in the 1980s. The human and mouse COUP-TFII have 100% amino acid identity, and there is 94% identity between *Drosophila* type 1 *svp* protein and human COUP-TF in DBD and 93% in LBD. This may suggest that COUP-TF proteins are one of the oldest members of the nuclear receptor family. COUP-TFII is generally considered as an orphan nuclear receptor because its natural ligand is not yet known, but retinoic acid induces a structural activation of COUP-TFII in LBD *in vitro*, but the used concentration of retinoic acid is well above the physiological levels. COUP-TFs could recognize the binding sites of other members of the nuclear receptor family, such as TR, RAR, VDR, HNF4 or ER α , thereby competing for binding site availability. Several studies on

the COUP-TFII orphan nuclear receptor have been reported that is involved in regulation of many biological processes, such as organogenesis, neural development, cardiovascular processes, reproduction, metabolism, and diseases (congenital heart failure (CHD) and cancer). COUP-TFII^{-/-} homozygous knockout mice died around 10 days of embryonic development due to cardiovascular disorders and defects in angiogenesis. Its role in angiogenesis has also been reported in other studies where COUP-TFII affects the expression of angiogenic processes such as vascular endothelial growth factors (VEGF) and their receptors (VEGFR), Angiotensin-1/Tie2 and Notch genes. COUP-TFII is expressed in all human tissues; it shows high levels of expression in endocrine, metabolic, reproductive and cardiovascular tissues.

1.5. Nuclear receptors in cancer

Nuclear receptors play a key role in many biological processes by providing a complex gene regulatory network that can lead to cancerous events. Expression of some nuclear receptors is associated with improved survival rate and could be therapeutic targets. One of the best examples is the presence of estrogen and progesterone receptors (ER and PR) in breast cancer, which determines the classification of breast cancers. ER and PR-positive breast cancer have better prognosis and better survival than those types of breast tumors where these receptors are not expressed. Selective estrogen receptor modulators (SERMs) brought about a breakthrough in the treatment of ER-positive breast cancer patients. One of the most well-known SERMs, tamoxifen, is still used therapeutic agent in ER-positive breast cancer patients; however newer SERMs, selective estrogen receptor downregulators (SERDs), aromatase and sulphatase inhibitors have been developed to decrease side effects of these drugs. Nevertheless, ~40% of patients develop endocrine resistance, which leads to cancer recurrence. Many factors can contribute to the development of endocrine resistance such as ER α and its co-regulators. According to a study, COUP-TFII may also play an important role

in maintaining endocrine sensitivity of breast cancer cells. COUP-TFII shows low expression in tamoxifen-resistant human breast cancer cells, however, re-expression of COUP-TFII in these cells restores tamoxifen sensitivity. Another study shows that COUP-TFII has a cell and drug type-dependent effect on chemoresistance in breast cancer cells. There is no specific response to the significance of COUP-TFII expression in breast cancer patients, as different studies have shown different results. The role of COUP-TFII expression in other cancers, such as prostate cancer, gastric cancer, colorectal cancer, pancreatic adenocarcinoma, and ovarian cancer, has also been investigated. So far, the prognostic significance of COUP-TFII differs depending on the study and the type of tumor and its exact role is not yet known.

1.6. Functional genomics tools for investigation of regulation by nuclear receptors

In recent decades, genomic technologies have evolved to allow a better understanding of gene regulation by nuclear receptors at the genomic level. Functional genomic methods made it possible to follow changes in gene expression, binding site, DNA methylation and chromatin availability. Mapping of binding sites of nuclear receptors, co-regulators, and other DNA binding proteins is possible by chromatin immunoprecipitation sequencing (ChIP-Seq). Chromatin immunoprecipitation (ChIP) is based on the interaction between protein and DNA. In this method, we use the antibody raised against the interest protein to extract the DNA fragments that our protein binds. Several steps in ChIP are critical: the starting cell number (10-20 million cells), fragment size, selection of the appropriate "ChIP-grade" antibody, selection of appropriate controls, and depth of sequencing. Because of all these critical steps, ChIP-Seq is less suitable for direct examination of patient samples.

The combination of ChIP-Seq data with RNA sequencing contributes to the study of nuclear receptor binding sites and its transcriptional consequence at the genome level. Combined, these can be used to examine how certain hormone or drug treatments affect nuclear receptor binding and gene expression, thereby predicting response to therapeutic

options. Since several nuclear receptors are present at the same time in cells and some of their regulatory regions overlap, studying the relationship between members of the nuclear receptor family may contribute to understand new roles in tumor progression and treatment.

2. AIMS OF OUR STUDY

The importance of the COUP-TFII orphan nuclear receptor in tumorigenesis has been reported in numerous studies, but genome-wide studies are not yet well-known. Our goal was to understand the mechanisms of COUP-TFII regulation at genomic level in various tumor cells. In our study, we aimed to

- perform characterizations of chromatin immunoprecipitation with ER α phage control
- map COUP-TFII cistrome in ER-positive breast cancer cells
- profile transcriptome of COUP-TFII-depleted ER-positive breast cancer cells
- investigate the level of COUP-TFII gene expression in patients with different breast cancer subtypes and its effect on survival of patients
- map of COUP-TFII cistrome in different cancer cells (breast, liver cancer and leukemia)
- examine the regulation of genes involved in angiogenesis by COUP-TFII
- investigate the significance of COUP-TFII gene expression level in the survival of patients with different cancer

3. MATERIALS AND METHODS

3.1. Cell culture and treatments

MCF-7 and HEK293T cells were cultured in DMEM media supplemented with 10% FBS, 1% L-glutamine and 1% penicillin-streptomycin. T47D cells were cultured in RPMI-1640 media supplemented with 10% FBS, 1% L-glutamine and 1% penicillin-streptomycin. COUP-TFII depleted MCF-7 and T47D cells were generated using the MISSION® Lentiviral Transduction Particle. The lentivirus containing shRNA sequences (NM_000125, TRCN0000003300 Clone) for silencing of COUP-TFII was used. The lentivirus containing shRNA sequence that has no target in human was used as a control (MISSION® pLKO.1-puro Non-Target shRNA Control Lentiviral Transduction Particles, SHC016V-1EA, Sigma). The experiment was performed as recommended by the manufacturer.

3.2. Phage production for mimicking epitope of polyclonal ER α antibody

A prepared random heptapeptide library of New England Biolabs (Ph.D.™ -7 Phage Display Peptide Library Kit) was used to generate phages that bind with high affinity to the hypervariable region of the widely used ER α antibody. About the experiment: 5 μ l (1011 pfu) of the phage peptide library was pre-blocked with 1 ml of TBST buffer containing 1% BSA and incubated for 10-60 minutes. To the blocked phage, 10 μ g of antibody was added and then 50 μ l of pre-blocked magnetic bead was added to the antibody-phage complex. The reaction was incubated for 20 minutes at room temperature with gentle rotation. The beads were washed ten times with TBST/BSA buffer and eluted twice with acidic elution solution (0.2 M glycine-HCl pH 2.2), 1 mg/ml BSA for 20 min at room temperature. To neutralize the acidic elution solution, 150 μ l of 1 M Tris-HCl buffer (pH 9.1) was added to 1 ml of eluate.

Phage amplification. Phages were grown in bacterial strain F + ER2738 by inoculating 25 ml of bacterial culture in the early log phase and shaking vigorously for 4.5-5 hours in non-selective LB media and centrifuging the samples. The phages within supernatant were

precipitated overnight at 4 °C in 1/6 volume of 20% (v/w) PEG-8000/2.5M NaCl, and the centrifuged pellet was suspended in TBS. The phages were again precipitated with the same solution. The centrifuged pellet was suspended in 200 µl TBS buffer. Absorbance was measured at 260 nm with a spectrophotometer to measure the concentration of phage particles. For prolonged storage, glycerol was added to a final concentration of 50% and stored at -20 °C.

Production of monoclonal phages. For this, we used the phage titration and plaque amplification protocol of the New England BioLabs with some modifications. The amplified polyclonal phages were diluted 10^9 - 10^{13} -fold in LB medium, then infected with 200 µl of bacterial strain ER2738 with 10 µl of diluted phage in the middle log phase, and the infected cells were plated on LB/IPTG/X plates and incubated overnight at 37 °C. The following day, an overnight incubation of the ER2738 bacterial culture was diluted 1: 100 and grown to an early log phase (OD₆₀₀ of 0.3-0.5) and a single colony from the plate was inoculated into 2 ml of bacterial culture. Next, the steps of "*Phage amplification*" were followed.

3.3. Chromatin immunoprecipitation (ChIP)

Phage ChIP. HEK293T fragmented chromatin was used for the experiment. A sufficient volume and amount of phage (1, 10, and 100 million) was added to HEK293T chromatin and the chromatin-phage mix precipitated with antibody (ER α : sc-543X) conjugated magnetic beads. Subsequently, the magnetic bead-binding complex was washed with various saline-containing ChIP wash buffer. The antibody-chromatin complex was eluted with elution buffer and the cross-links were dissolved by adding NaCl. Samples were digested with RNase and Proteinase K. Immunoprecipitated DNA was purified using the High Pure PCR Template Preparation Kit according to the manufacturer's instructions.

ChIP from breast cancer cells. For chromatin immunoprecipitation, we started from 15-20 million breast cancer cell lines (MCF-7 and T47D). Cells were fixed with 1% methanol-free

formaldehyde for 10 minutes at room temperature and stopped with 0.125 M glycine. Cells were washed with ice-cold 1xPBS buffer. The cells were scraped in 1 ml of ChIP lysis buffer and centrifuged at maximum speed. During nuclear isolation, the pellet was resuspended three times with ChIP lysis buffer. Fragmentation of chromatin was performed with a Bioruptor Plus Sonicator. For sequencing, the amount of chromatin corresponding to 15-20 million cells and for quantitative PCR 5 million cells was diluted ten-fold. Immunoprecipitation was performed overnight at 4 ° C using the following antibodies: ER α (sc-543X), COUP-TFII (sc-271265X) and IgG (sc-2027X). The amount of antibody for sequencing is 8 μ g, and the amount for quantitative PCR is 4 μ g. Following incubation, pre-blocked Protein A and Protein G conjugated magnetic beads were added. Samples were incubated again for a minimum of 6 hours at 4 ° C. The magnetic bead binding complex was then washed with different ChIP wash buffer and TE buffer. After eluting the antibody-chromatin complex and crosslinking, the samples were digested with RNase and Proteinase K. Immunoprecipitated genomic DNA was purified using the Qiagen MinElute PCR Purification Kit according to the manufacturer's instructions.

3.4. RNS isolation and reverse transcriptase PCR (RT-PCR)

Cells were grown on 6-well plates for RNA isolation, and then washed with PBS in 0.5 ml Trizolate reagent (UD-Genomed) following the manufacturer's protocol. RNA concentration was measured with NanoDrop.

For each cDNA synthesis, 1 μ g of total RNA was transcribed using SuperScript III Reverse Transcriptase (Thermofisher 18064071). Unlike to the manufacturer's instructions, the following amounts of reagents were calculated per reaction specifically for our laboratory: 4 μ l SSII buffer (5x), 2 μ l DTT (100 mM), 4 μ l dNTP (2.5 mM), 0.08 μ l random hexamer (3 μ g / μ l) and 0.08 μ l SSII MnIv (200 U / μ l), and 10 μ l RNA sample. The temperature profile

used for the reaction was 10 minutes at 25 ° C, 2 hours at 42 ° C and 15 minutes at 70 ° C. Following PCR, samples were diluted 5x with nuclease-free water.

3.5. qPCR

For quantitative polymerase chain reaction, 2x SYBR Green Master Mix supplemented with primers (final concentration of 2.5 µM), ROX reference dye and nuclease-free water was used. The final volume of the mix was 7.5 µl. To mix, 2.5 µl of sample was added. There were three types of samples: single-stranded phage DNA, genomic DNA from ChIP and cDNA. The temperature profile of the PCR was set at 5 min at 95°C, 30 sec at 60°C and 30 sec at 72°C with 40 cycles and melting point analysis on an ABI QuantStudio 12K Flex Real-Time PCR instrument. For the analysis, the Δ CT method was used. Normalisation was performed with input DNA in case of ChIP, or with housekeeping gene (ACTB) in case of profiling gene expression.

3.6. Next generation sequencing

Library preparation and next generation sequencing was performed by Genomic Medicine and Bioinformatic Core Facility, University of Debrecen, Hungary. The ChIP DNA libraries were constructed according to the Illumina's TruSeq ChIP Sample Preparation protocol, starting from 10 ng of ChIP DNA, and then sequenced in 50 bp read length (single-end) on Illumina NextSeq 500 system. RNA sequencing was performed using Illumina's TruSeq RNA Sample Preparation version 2 protocol, starting from 1 µg total RNA. The libraries were sequenced on the Illumina NextSeq 500 system in 50 bp read length (single-end).

3.7. Analysis of new generation sequencing data

ChIP-Seq data. Our ChIP-Seq and RNA-Seq data can be found under NCBI BioProject PRJNA602619, other ChIP-Seq data are from publicly available databases. ChIP-seq data

were analyzed by following in-house computer analysis steps. These steps were as follows: Sequencing reads were aligned to the hg19 (GRCh37) genome using Burrows-Wheeler Alignment v0.7.17 (BWA). Then the peaks were predicted using Model-based Analysis for ChIP-Seq (MACS2) v2.1.1 and artifacts based on ENCODE were extracted using Bedtools v2.27.1 (intersectBed). For ChIP-Seq data available from ENCODE, we ranked the peaks by coverage (RPKM) and used the peak at 25% for further analysis. Overlaps, subtractions, and merges between predicted peaks were performed with BedTools v2.27.1. Overlaps were examined using the diffBind R package. The motif analysis was performed using the findMotifsGenome.pl command of HOMER v4.9.1. Sequence enrichment distributions were calculated using the annotatePeaks.pl command of the HOMER program in 1000-2000 bp regions, and plotted on a heatmap or histogram. The annotation to genes was also done by the annotatePeaks.pl command of the HOMER program. Gene ontology analysis was performed using Genomic Regions Enrichment Annotations Tools (GREAT) v3.0.0.

RNA-Seq data. The raw data were mapped to the human hg19 reference genome using TopHat v2.1.1. The number of reads for each gene was determined using featureCounts v1.6.2. The differently expressed genes among the samples were identified using the edgeR v3.22.3 program running R v3.5.1 by setting the FDR 0.01 threshold. Genetic enrichment analysis was performed with GSEA v4.0.0.

3.8. Analysis of data derived from patients with cancer

Breast cancer patients' data derived from "The Cancer Genome Atlas", and downloaded using cBioPortal. Z-score data from processed RNA-Seq were used for gene expression analysis. Survival data were used from KMPlotter database of a large group of patients.

4. RESULTS

4.1. Characterisation of chromatin immunoprecipitation with ER α phage control

Chromatin immunoprecipitation is a complex and multi-step process that requires multiple levels of control and normalization. One of the most important is to know the proper antibody of our choice in the method used. Our aim was to set up a standardized ChIP experiment with a well-known antibody (ER α sc-543x), so we have developed a phage expressing ER α antibody epitope generated by phage display in our laboratory. The phage is well suited for studying the interaction between protein and DNA, since the phage envelope protein can mimic a particular portion of the protein of interest, and the known single-stranded DNA sequence in the phage can measure DNA by qPCR. Individual monoclonal phage clones were generated after several cycles using the NEB Ph.D.™ -7 random heptapeptide library. First, ten monoclonal phages were tested for which clones could be used for the ChIP reaction. We examined how many phages can be recovered from the total ChIP elution and how many of them remain as a loss in the IP buffer. We found two clones that we could regain after the whole process. By further examining these clones, we examined how much we can recover with our antibody by adding 100, 10,000, and 1,000,000 phage clones to ChIP. Our results showed that using 100 phages, ~60%, while using more phages, ~80% could be measured in the elution. In further experiments, we examined the reproducibility and stability of phage ER α . Reproducibility was tested in two separate ChIP experiments, which had in the same results obtained in two experiments despite the presence of different amounts of phages. Examining the effect of freezing, we found that the IP efficiency was reduced by ~10%, whereas in the case of freshly grown phages, the IP efficiency was similar to the previous results. The advantage of a phage-based system is that it is easily reproducible in bacterial strains, so it can be reached at any time within a short time. The phage system allows us to investigate epitope losses during the ChIP reaction. Before starting the ChIP experiments, the

production batch variability that is common to antibodies can be well screened. Taken together, these results suggest that phages can be used as spike-in controls during chromatin immunoprecipitation.

4.2. COUP-TFII as ER α coregulator in ER-positive breast cancer cells

Many research groups have reported on COUP-TF orphan nuclear receptor interaction with the ER α nuclear receptor, but less genome-wide studies are available. We investigated the role of COUP-TFII in ER α -mediated regulation at the genomic level in breast cancer cells by ER α and COUP-TFII ChIP-Seq from ER-positive MCF-7 and T47D cells. It has been found that ~90% of COUP-TFII binding sites overlap with ER α binding sites in MCF-7 cells, indicating that a significant portion of COUP-TFII is present in the ER α regulatory program. Shared binding sites have a higher signal intensity of ER α and COUP-TFII ChIP-seq than nuclear receptors alone. The same results could be confirmed in T47D cells. Examining the presence of active histone modifications at the shared COUP-TFII and ER α binding sites and only ER α binding sites, we found that the shared COUP-TFII and ER α regions show higher signal intensities for histone modifications associated with the presence of an active enhancer. To investigate which other transcription factors may be present at the shared COUP-TFII and ER α binding sites, we performed a motif enrichment analysis. Our results demonstrated that the NR half-sites, FOXA1 and GATA3 motifs showed a significant enrichment at the shared ER α and COUP-TFII binding sites, which could be confirmed at the binding level. These results suggested that COUP-TFII is present in an ER α -mediated transcription program bound by FOXA1 and GATA3 cofactors.

4.3. Role of COUP-TFII in gene expression in breast cancer cells

To examine gene expression regulation by COUP-TFII in breast cancer cells, COUP-TFII depleted ER α -positive breast cancer cell lines were generated using lentiviral-based shRNA

gene silencing. We found that our upregulated genes in COUP-TFII depleted cells are associated with genes that have decreased expression in tamoxifen-resistant cells, while increased expression after estrogen treatment. All this suggested that our differently expressed genes are also regulated by ER α . When we compared our downregulated gene set, we found genes that are also DREAM complex target genes playing an important role in cell cycle regulation.

4.4. Significance of COUP-TFII in patients with breast cancer

To examine COUP-TFII gene expression, different subgroups of breast cancer patients were used to gain a closer understanding of the importance of COUP-TFII in each subtype. The HER2-positive and basal subgroups are ER-negative breast tumors, thus, as expected, the level of ER α was low and the same is observed with COUP-TFII. COUP-TFII expression is also significant in ER-positive subgroups along with ER α . The results also showed that COUP-TFII expression in invasive lobular carcinoma (ILC) luminal A group is significantly ($p < 0.0001$) higher than in invasive ductal carcinoma (IDC) luminal A group. Subsequently, the disease-free survival rate (DFS) was examined based on the high or low expression level of COUP-TFII in the ER-positive and ER-negative breast tumors. In the ER-positive group, high expression of COUP-TFII significantly (logrank $P < 0.0001$, Mantel-Cox test) contributes to better survival compared to the low expression of COUP-TFII. Subsequently, the ER-positive group was further subdivided into luminal A and B subgroups for survival. It was found that improved survival rate associated with high expression of COUP-TFII can be observed only in luminal A, while no difference in survival in COUP-TFII expression level was observed in luminal B.

4.5. Map of COUP-TFII binding sites in different cancer cells

Little is known about the genomic distribution of COUP-TFII binding sites in various tumor cells. A total of 64 149 COUP-TFII binding sites were predicted in the three cell lines. 2.35% of which (1 509/64 149) overlapped between all three cell lines (shared COUP-TFII), whereas the majority of binding sites were present only in a particular cell type (cell type-specific COUP-TFII). Concerning to the distribution of the binding sites for the entire genome, ~40% of the shared COUP-TFII binding sites, while only ~10% of the cell-type COUP-TFII binding sites was located in the promoter region. We used the GREAT program to investigate the biological genes associated with COUP-TFII binding sites in certain cell types. As a result, we found that COUP-TFII binding sites are located near genes that are linked to cell specific biological processes.

4.6. Identification of cofactors at COUP-TFII binding sites

There is limited knowledge about the cofactors at COUP-TFII binding sites. Therefore, we performed motif analysis on the shared and cell type-specific COUP-TFII binding sites. Enrichment of the nuclear receptor-specific AGGTCA motif (NR) was observed in shared and individual cell lines. In addition, the FOXA1 motif was significantly enriched in MCF-7 and HepG2 cells, whereas the GATA motif was enriched in K562 cells. At the shared COUP-TFII binding sites, enrichment of the CTCF motif can be found. Based on transcription factor binding intensities, the cell-type specific COUP-TFII binding sites were determined by the master transcription factors, while the shared COUP-TFII binding sites were determined by the presence of CTCF.

4.7. COUP-TFII binding sites in cell-type specific regulation

To examine the chromatin environment of COUP-TFII, we first examined the proportion of COUP-TFII present in different cells besides the master transcription factor and CTCF. It was found that more than half of the master transcription factors overlap with COUP-TFII binding sites within a cell line and CTCF overlaps with COUP-TFII over master transcription factors. Cell type-specific COUP-TFII binding sites were predominantly with master transcription factors, while shared COUP-TFII binding sites showed greater overlap with CTCF. This confirms our previous findings that the cell type-specific COUP-TFII binding events are specific for the master transcription factor of the given cell, and the shared COUP-TFII binding events to the three cells are mainly related to CTCF. DNA enhancers and promoter regions are key elements of cell-type specific regulation, which are well characterized by the presence of various histone modifications. It was found that the COUP-TFII with its master transcription factor is preferentially located at active enhancer elements while alone or with CTCF located at active promoter regions.

4.8. Regulation of *VEGFA* gene by COUP-TFII

Earlier studies have reported that COUP-TFII plays an important role in angiogenesis. Therefore we investigated whether genes involved in angiogenesis can bind to a shared COUP-TFII binding site or not. Eleven genes (*ANG*, *ANGPTLA*, *CTSB*, *MDK*, *NRAS*, *PECAMI1*, *PTAFR*, *PTGS1*, *TGFBR3*, *VEGFA*, and *ZNF444*) were found to be associated with angiogenesis and shared COUP-TFII binding sites. ChIP-Seq data showed that COUP-TFII binding sites are located on the promoter and distal regulatory elements of the *VEGFA* gene in MCF-7, HepG2 and K562 cells. The pattern of COUP-TFII binding in the three cells is different, but there were several regions where these binding sites overlap with each other and also with CTCF. CTCF ChIA-PET data showed CTCF bounded chromatin loops. From these data, we could see that COUP-TFII and CTCF binding regions up to +45, -69, -183 and -305

kilobases from the *VEGFA* gene promoter are involved in chromatin interaction with the *VEGFA* promoter in at least two cell lines.

In MCF-7 cells, expression of the *VEGFA* gene was examined after the depletion of COUP-TFII. Gene silencing resulted in increased *VEGFA* gene expression. This suggested that COUP-TFII inhibits *VEGFA* gene expression. These results also suggested that COUP-TFII protects cells against a worse prognostic tumor phenotype.

4.9. Effect of COUP-TFII expression on survival of patients with different cancers

We previously described that high levels of COUP-TFII correlate with improved survival in ER-positive breast cancer patients. In this regard, we examined how COUP-TFII expression influences survival in patients with liver cancer and leukemia. Using the KMplotter database, no data are available for leukemia, but liver cancer and other types of cancer (lung, stomach, and ovarian) data are available. From these data, we could see that liver, lung, ovarian, and gastric cancers have the same effects as breast cancer, although high expression of COUP-TFII correlates with significantly better disease-free survival compare to the low expression of COUP-TFII.

5. DISCUSSION

COUP-TFII, a member of the steroid/thyroid nuclear receptor family plays an important role in various developmental processes. Several studies have reported that it plays a role in tumorigenesis, but there is limited knowledge about COUP-TFII-mediated transcriptional regulation at the level of the genome in tumor cells.

Advances in technologies for investigating binding sites and gene expression at the genomic level contribute to a better understanding of the molecular mechanisms of nuclear receptors. One of these technological advances is chromatin immunoprecipitation sequencing (ChIP-Seq), which is now a widely used method for studying transcription factors and

chromatin modifications. Because of the complexity of chromatin immunoprecipitation (ChIP) and the lack of standardization. In our work, we have developed a phage-based system that can be used as a control to monitor the specificity of the ER α antibody during the ChIP process. This ER α antibody is widely used in hundreds of publications, but unfortunately it is no longer commercially available. The outer envelope of the phage mimics the epitope (mimitope) of the ER α protein, whereas the single-stranded DNA inside it allows it to be detected in the PCR reaction. For this, polyclonal phage ER α was produced by phage display and, after multiple round selection, monoclonal individuals were generated. Monoclonal ER α phages were then tested during the ChIP process. We investigated epitope losses at different stages of ChIP, the stability and reproducibility of phage ER α were tested, too. Our results provided an easy-to-use, reproducible control system for antibody characterization during the ChIP process.

ER α is a key nuclear receptor in breast cancer as its presence correlates with a better prognosis and it is an important target for endocrine therapy. Nonetheless, patients often become resistant to this treatment, leading to recurrence of the disease. To understand these events, we need to know the mechanisms of regulation by ER α , including the function of ER α co-regulators. These co-regulators are proteins that act as a bridge-forming or helper to form a large transcriptional complex, influencing target genes activity. In our study, we identified the COUP-TFII orphan nuclear receptor as an ER α co-regulator. Based on ER α and COUP-TFII ChIP-Seq data from mammary origin MCF-7 and T47D cells, we found that nearly 90% of COUP-TFII overlaps with ER α binding sites and ChIP-Seq signal are greater at shared COUP-TFII and ER α binding sites than individual binding sites. Furthermore, the shared COUP-TFII and ER α binding sites correlate with specific histone modifications of the active enhancers (H3K27ac and H3K4me1). COUP-TFII is part of a transcriptionally active complex in which FOXA1 and GATA3 are present in addition to ER α . This suggested that

COUP-TFII binds to ER α transcriptionally active sites. In response to the question of which genes are affected by COUP-TFII, we profiled the transcriptome in COUP-TFII depleted ER-positive breast cancer. Our results demonstrated that after depletion; mainly genes, playing a role in endocrine resistance were upregulated, while expression of these genes was also increased by E2 treatment in other studies. In the next step, we examined how COUP-TFII expression changes in breast cancer patients. COUP-TFII expression was found to be higher in ER α -positive breast cancer patients than in ER-negative patients. In the survival rate of ER-positive patients subdivided into luminal A and B subgroups, we found that in the HER2-negative luminal A group, higher COUP-TFII expression levels correlate with better survival than low COUP-TFII expression levels.

In the next step, COUP-TFII cistrome was examined in different tumor cells using publicly available ChIP-Seq data. We found that the COUP-TFII shows a cell type-specific binding site distribution in the genome, with only a small fraction shared between the three cells. In addition to shared COUP-TFII between the three cells, CTCF binding was observed mainly in the active promoter region. No direct interaction between COUP-TFII and CTCF was reported yet. However, some other results suggested that they may be present in a complex. The cell-type specific COUP-TFII overlaps with nearly half of the cell-specific master transcription factor and these sites correlate with specific histone modifications (H3K27ac and H3K4me1) of active enhancers. However, it is not only our study suggesting that COUP-TFII binds to the master transcription factor. Rada-Iglesias et al. found that besides AP2 α (TFAP2A), the main regulator of neural crest cells, COUP-TFII is also present in active enhancers. In another study, COUP-TFII with TBX5 cardio-specific transcription factors plays a role in the regulation of cardio-specific genes. Our results also showed that COUP-TFII overlaps with active enhancer regions with master transcription factors, and these regions are determinants of cell type-specific regulation. Taken together, these results

demonstrate a novel role for the COUP-TFII nuclear receptor, which implies the presence in the cell type-specific regulation.

Since several studies have reported the role of the COUP-TFII nuclear receptor in angiogenesis, we investigated whether a cell-type-independent COUP-TFII binding site could be observed near the genes involved in angiogenesis. Eleven genes related to angiogenesis were found to which three cells overlap the COUP-TFII binding site. Examination of this gene found that each of the three cells of different origins displayed COUP-TFII and CTCF-bound promoter and distant enhancer regions in the presence of H3K4me3 histone modification, which is involved in chromatin loops. To determine how COUP-TFII influences the VEGFA gene, we examined the expression of VEGFA in COUP-TFII depleted MCF-7 cells. We found that expression of the VEGFA gene was increased in COUP-TFII depleted cells, suggesting that COUP-TFII inhibits VEGFA expression. This suggested that the COUP-TFII nuclear receptor has an anti-tumor role.

We have also examined the effect of COUP-TFII expression on disease-free survival (DFS) in cancer types other than breast cancer, such as liver cancer, lung cancer, ovarian cancer, and gastric cancer. In all cases, it was found that high expression of COUP-TFII correlates with better survival than low expression of COUP-TFII. Taken together, these results further enhanced the beneficial effects of the presence of COUP-TFII in tumor cells, patients, and may even serve as an important therapeutic target for the treatment of cancer.

6. SUMMARY

Nuclear receptors, as ligand-activated transcription factors, can influence directly the gene expression. Changes in gene expression are a hallmark of cancer. Genome-wide studies make it possible to study the transcriptional program by nuclear receptors, which brings us closer to understanding the disease. The most widely used method for the regulation of nuclear receptor is chromatin immunoprecipitation (ChIP). In our study, we have established a phage display control system through the example of estrogen receptor alpha (ER α) antibody that enables the characterization of ChIP antibodies in ChIP-qPCR experiments. ER α is a key nuclear receptor in the treatment and outcome of breast cancer. We investigated the role of COUP-TFII orphan nuclear receptor in ER-positive breast cancer cells and patients using functional genomics approaches. Our results indicate that COUP-TFII, as a co-regulator, is present in the ER α -mediated transcriptional complex and affects the expression of ER α target genes. Furthermore, high expression levels of COUP-TFII correlate with improved survival in patients with ER-positive luminal A phenotype. Examining the cistrome of COUP-TFII in other tumor cells, we found that COUP-TFII co-localizes with the cell-type-specific master transcription factor. Shared COUP-TFII binding sites in cells of different origins were identified in the regulatory regions of the *VEGFA* gene. In breast cancer cells, COUP-TFII inhibits the expression of this angiogenic gene. Investigation of survivals in various cancer patients was found that high expression levels of COUP-TFII correlate with improved survival compared to low expression levels of COUP-TFII. In summary, the presence of the COUP-TFII nuclear receptor has a tumor suppressor effect with master transcription factors in various cancers.



Registry number: DEENK/90/2020.PL
Subject: PhD Publikációs Lista

Candidate: Edina Erdős
Neptun ID: S6ZPAR
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology
MTMT ID: 10037512

List of publications related to the dissertation

1. **Erdős, E.**, Bálint, B. L.: NR2F2 Orphan Nuclear Receptor is Involved in Estrogen Receptor Alpha-Mediated Transcriptional Regulation in Luminal-A Breast Cancer Cells.
Int. J. Mol. Sci. "Accepted by Publisher", 2020.
IF: 4.183 (2018)
2. **Erdős, E.**, Bálint, B. L.: COUP-TFII is a modulator of cell-type-specific genetic programs based on genomic localization maps.
J. Biotechnol. 301, 11-17, 2019.
DOI: <http://dx.doi.org/10.1016/j.jbiotec.2019.05.305>
IF: 3.163 (2018)
3. Keresztessy, Z., **Erdős, E.**, Ozgyin, L., Kádas, J., Horváth, J., Zahuczky, G., Bálint, B. L.:
Development of an antibody control system using phage display.
J. Biotechnol. 300, 63-69, 2019.
DOI: <http://dx.doi.org/10.1016/j.jbiotec.2019.05.009>
IF: 3.163 (2018)





List of other publications

4. Patten, D. K., Corleone, G., Györfly, B., Perone, Y., Slaven, N., Barozzi, I., **Erdős, E.**, Saiakhova, A., Goddard, K., Vingiani, A., Shousha, S., Pongor, L. S., Hadjiminias, D. J., Schiavon, G., Barry, P., Palmieri, C., Coombes, R. C., Scacheri, P., Pruneri, G., Magnani, L.: Enhancer mapping uncovers phenotypic heterogeneity and evolution in patients with luminal breast cancer.
Nat. Med. 24 (9), 1469-1480, 2018.
DOI: <http://dx.doi.org/10.1038/s41591-018-0091-x>
IF: 30.641
5. Ozgyin, L., **Erdős, E.**, Bojcsuk, D., Bálint, B. L.: Nuclear receptors in transgenerational epigenetic inheritance.
Prog. Biophys. Mol. Biol. 118 (1-2), 34-43, 2015.
DOI: <http://dx.doi.org/10.1016/j.pbiomolbio.2015.02.012>
IF: 2.581
6. Bojcsuk, D., **Erdős, E.**, Bálint, B. L.: Az ösztrogénreceptor működése a legújabb genomikai kutatások tükrében.
LAM KID. 4 (2), 79-84, 2014.
7. **Erdős, E.**, Bálint, B. L.: Miért különböznek az ikerpárok?: az ikerkutatásról.
LAM KID. 2 (3), 33-38, 2012.

Total IF of journals (all publications): 43,731

Total IF of journals (publications related to the dissertation): 10,509

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

10 March, 2020

