Summary for the grant K- 129166

The grant submitted in 2017 and launched in October 2018 contributed significantly to deepen our understanding on the components of cellular identity on chromatin level. Still, we can not provide strong answers on how stochastic events on chromatin level are influencing gene expression changes. We worked with several model systems: MCF-7 cells and LCL cells. We used extensive genomic technologies and a wide variety of bioinformatic tools. During this period three of the students from our research group competed successfully their PhD studies, and several undergraduate students worked on these project. Two of them received first price on the National Student Scientific Conferences (OTDK). The PI completed his habilitation process and became associate professor. A total of 10 publications were published with the support of the grant with a total impact factor of 44,352 (one of these was published within a collaboration in a new journal that has no IF yet). The dramatic changes in the research environment between 2018-2022 caused challenges that needed to shift our research activities during the grant period. Nevertheless we could stay on track, develop new tools, novel approaches and contribute significantly to the overall goals of the scientific community.

In December 2022 we published in Scientific Data the multiomic data of our comparison of the five theoretically isogenic LCL cell lines. In our previous paper we have shown that on chromatin level approx. a quarter of the enhancers are different between the five cell lines while only less than 5% of the genes are expressed differentially. By whole genome sequencing of these cell lines, we could prove that while the originate from the same person, still they are not isogenic. Slight genetic variations could be identified even at low passage numbers.

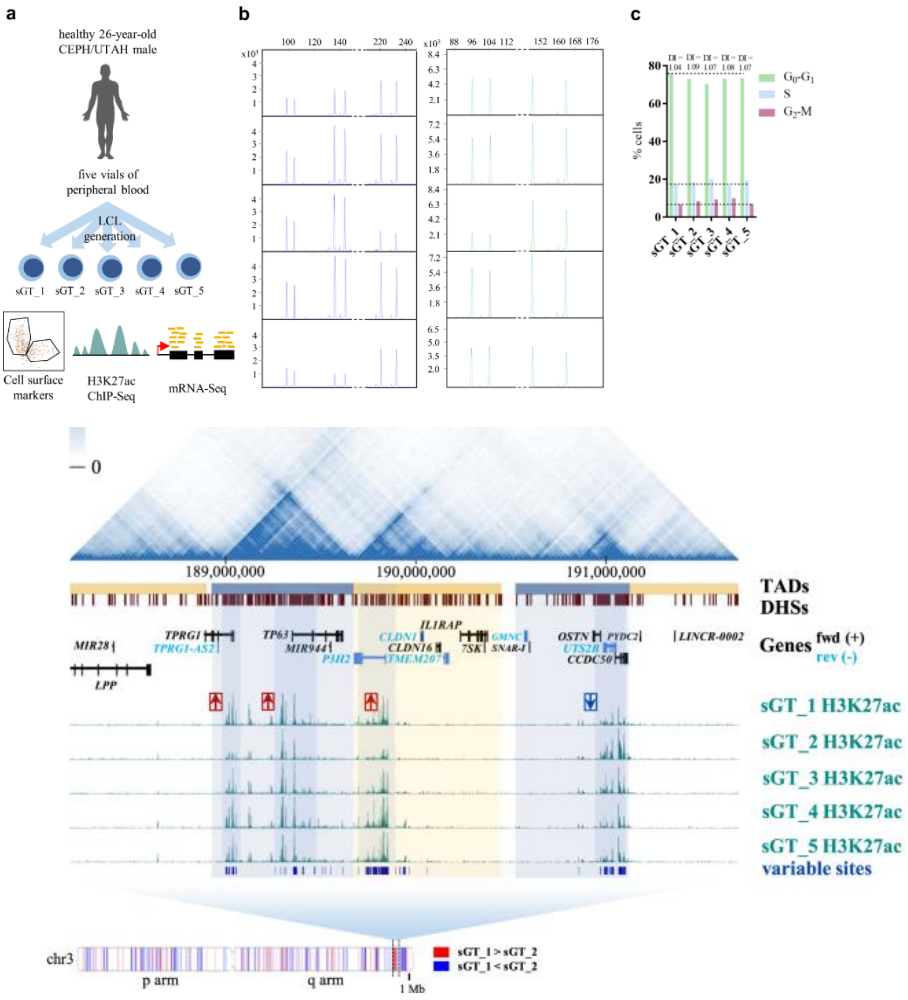
While our results are not sufficient to draw extensive conclusions, we speculate that human cells of an organism are not isogenic and the concept of isogenicity is a pure abstraction and we need to move past this outdated concept. We propose that the genome is in a continuous change and is in a plastic adaptation to the ever-changing environment.

We sought to understand different aspects of this concept by developing genetic and bioinformatic methods and studying different cellular systems.

Please find below some of the summaries of some of our published papers.

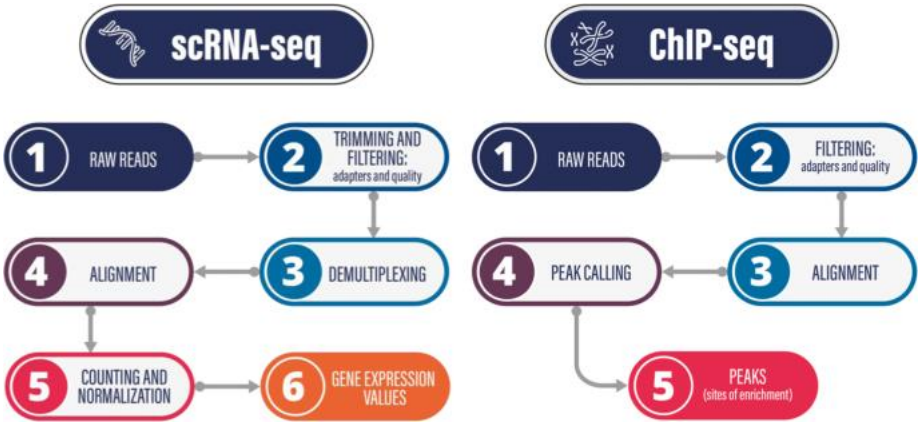
Genotyped human B-lymphoblastoid cell lines (LCLs) are widely used models in mapping quantitative trait loci for chromatin features, gene expression, and drug response. The extent of genotype-independent functional genomic variability of the LCL model, although largely overlooked, may inform association study design. In this study, we use flow cytometry, chromatin immunoprecipitation sequencing and mRNA sequencing to study surface marker patterns, quantify genome-wide chromatin changes (H3K27ac) and transcriptome variability, respectively, among five isogenic LCLs derived from the same individual. Most of the studied LCLs were non-monoclonal and had mature B cell phenotypes. Strikingly, nearly one-fourth of active gene regulatory regions showed significantly variable H3K27ac levels, especially enhancers, among which several were classified as clustered enhancers. Large, contiguous genomic regions showed signs of coordinated activity change. Regulatory differences were mirrored by mRNA expression changes, preferentially affecting hundreds of genes involved in specialized cellular processes including immune and drug response pathways. Differential expression of DPYD, an enzyme involved in 5-fluorouracil (5-FU) catabolism, was associated with variable LCL growth inhibition mediated by 5-FU. The extent of genotype-independent functional genomic variability might highlight the need to revisit study design strategies for LCLs in pharmacogenomics.

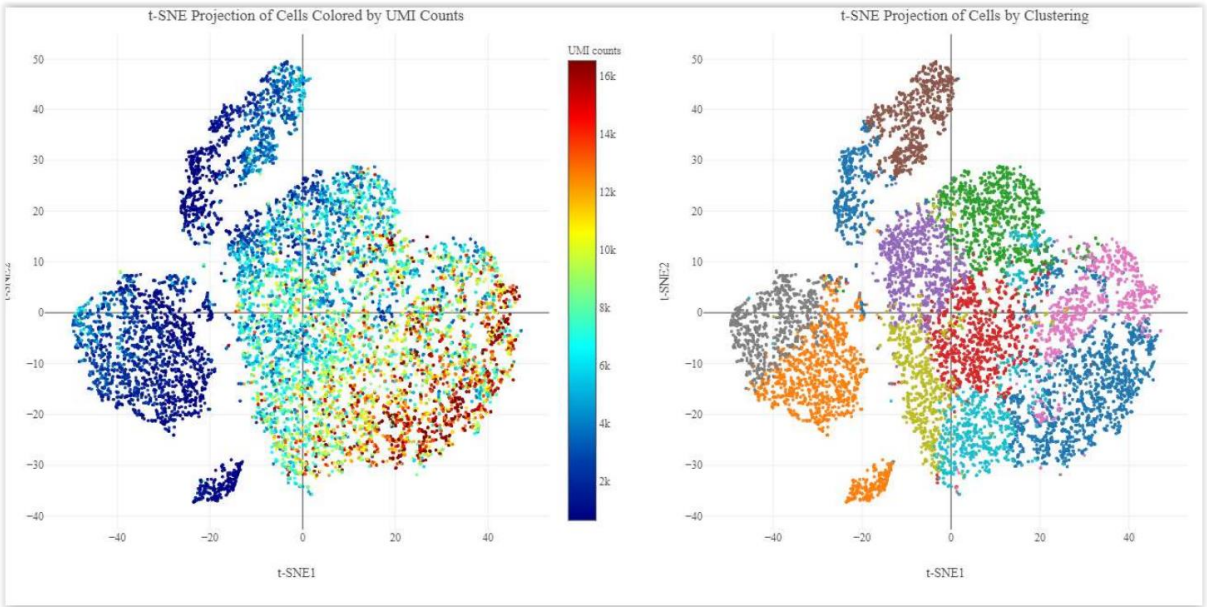
https://doi.org/10.1038/s41598-019-40897-9



In life-science research isogenic B-lymphoblastoid cell lines (LCLs) are widely known and preferred for their genetic stability - they are often used for studying mutations for example, where genetic stability is crucial. We have shown previously that phenotypic variability can be observed in isogenic B-lymphoblastoid cell lines. Isogenic LCLs present well-defined phenotypic differences on various levels, for example on the gene expression level or the chromatin level. Based on our investigations, the phenotypic variability of the isogenic LCLs is accompanied by certain genetic variation too. We have developed a compendium of LCL datasets that present the phenotypic and genetic variability of five isogenic LCLs from a multiomic perspective. In this paper, we present additional datasets generated with Next Generation Sequencing techniques to provide genomic and transcriptomic profiles (WGS, RNA-seq, single cell RNA-seq), protein-DNA interactions (ChIP-seq), together with mass spectrometry and flow cytometry datasets to monitor the changes in the proteome. We are sharing these datasets with the scientific community according to the FAIR principles for further investigations.

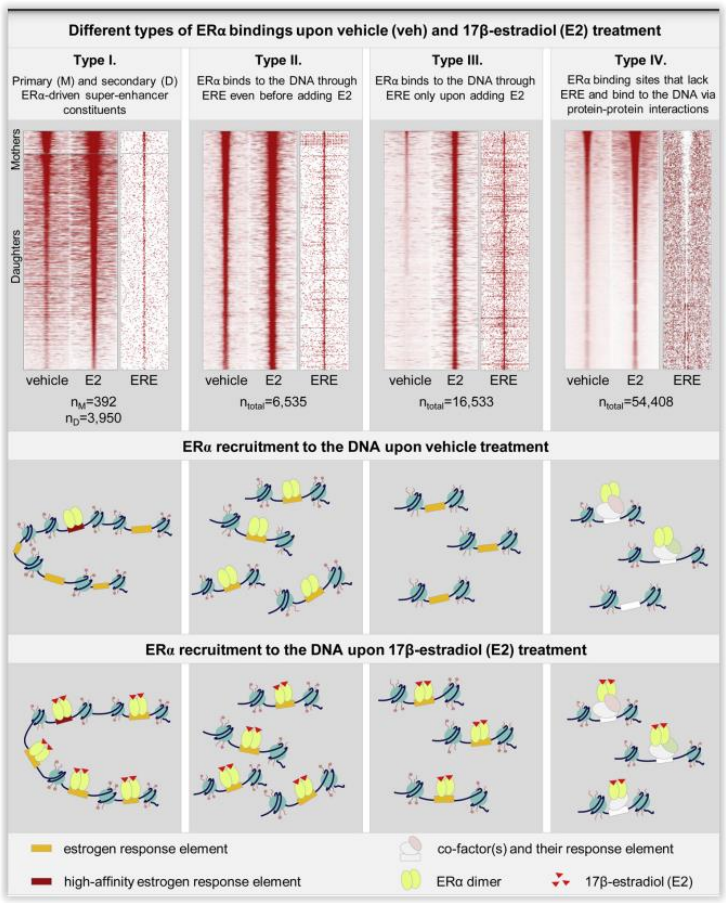
https://doi.org/10.1038/s41597-022-01871-9





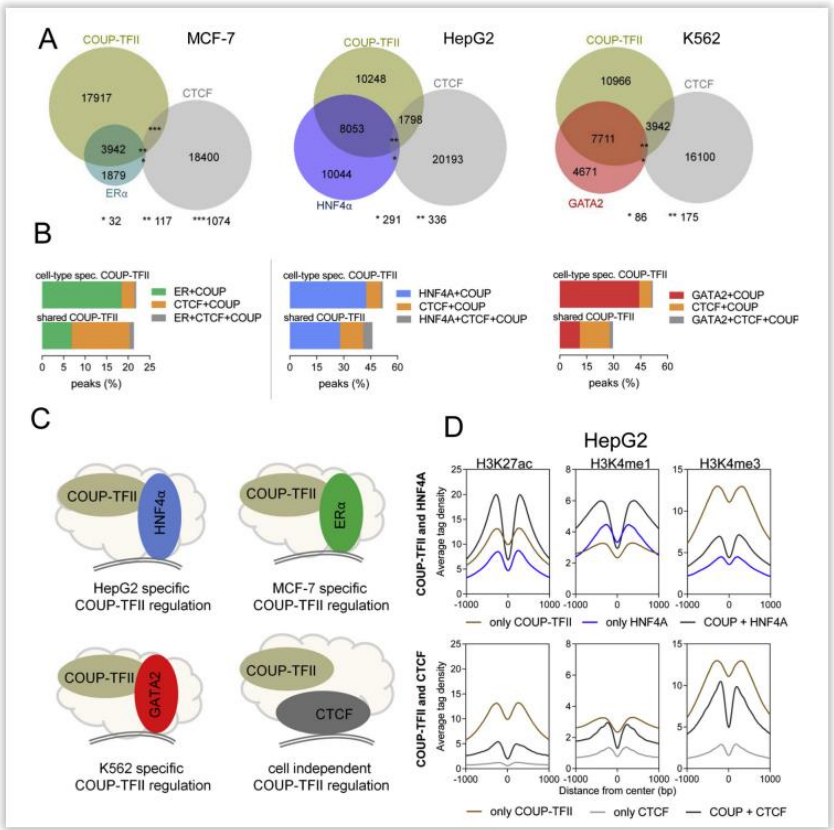
Estrogen Receptor alpha (ERα) is a ligand-activated transcription factor and it has a prominent role in both physiological and pathological processes of the reproductive system. ERα has been investigated extensively in breast cancer and the MCF-7 breast-cancer-derived cell line is a widely used model for the study of its behavior. In this paper we provide a systematic catalog of the possible scenarios of binding to more than 80,000 ERα transcription factor binding sites based on the mechanism of ERα binding to DNA (upon both vehicle and estradiol (E2) treatment). A key feature of the estrogen-driven genetic programs is the presence or absence of the specific response element referred to as the estrogen response element (ERE). While ERα-driven super-enhancers are key components of estrogen-dependent genetic programs, three additional classes of enhancers could be identified: one with the presence of ERE where the ERα bound to the DNA prior of E2-treatment, one where the E2 was required for ERα binding even in the presence of ERE, and one where the ERα binding is established through the response elements of the collaborating factors. Our results suggest that different scenarios of ERα binding result in different genetic programs.

10.1016/j.jbiotec.2019.04.016



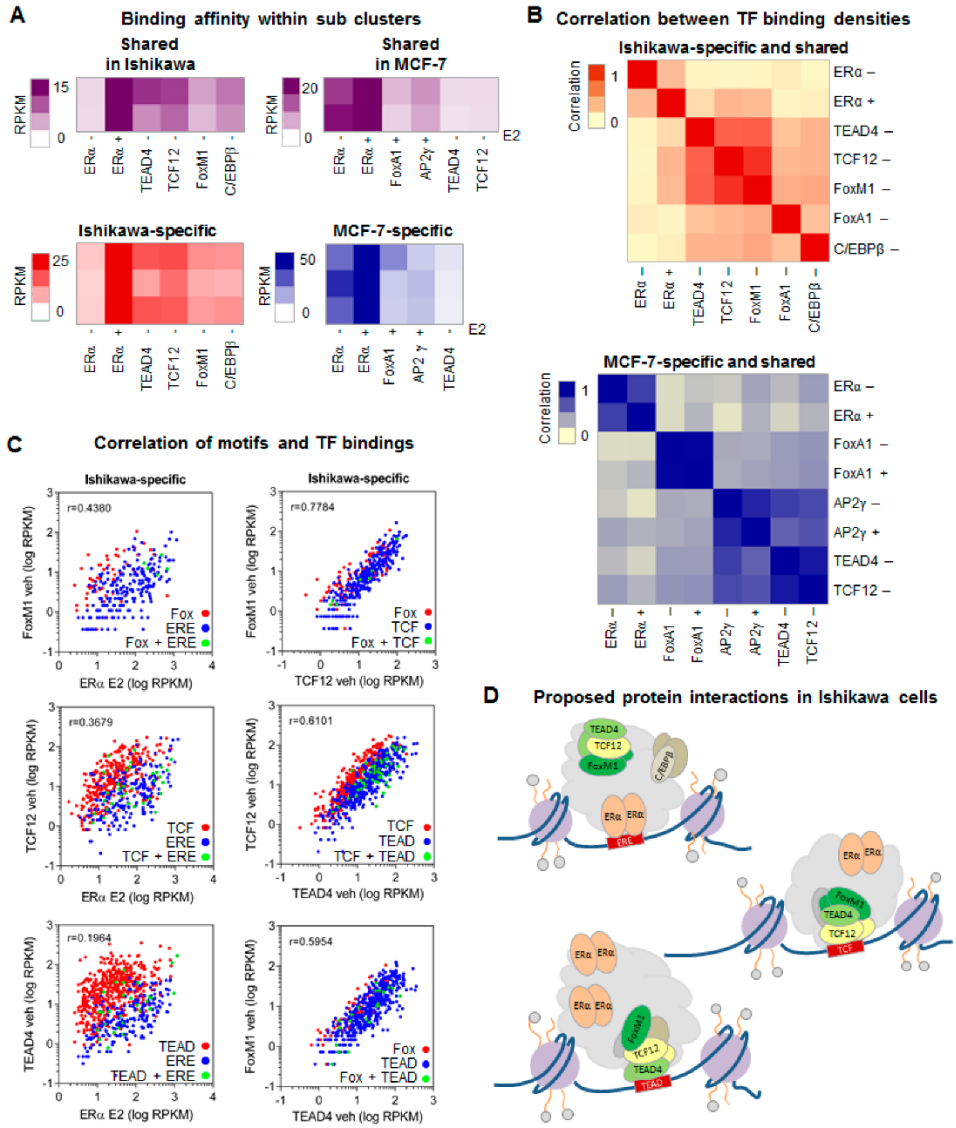
Chicken ovalbumin upstream promoter transcription factor II (COUP-TFII) is a member of the steroid/thyroid hormone receptor superfamily, but its ligand has not yet been identified. Little is known about the role of the COUP-TFII nuclear receptor in cancer cells. In this study, we mapped the cistrome of COUP-TFII in three different cancer cells, namely breast cancer cells (MCF-7), myelogenous leukaemia cells (K562) and liver cancer cells (HepG2) using publicly available ChIP-seq data. Our results show that COUP-TFII co-localises with master transcription factors (TFs) in a cell-specific manner such as estrogen receptor alpha in MCF-7, hepatocyte nuclear factor alpha in HepG2, and GATA-binding factor in K562, while the shared, non-specific COUP-TFII binding sites are co-occupied by CTCF. We identified chromatin environments for these COUP-TFII and master TF co-bound sites together with COUP-TFII and CTCF co-bound sites. Our results show that COUP-TFII and master TF co-bound sites are marked with active enhancer specific histone modifications (H3K27ac and H3K4me1), while COUP-TFII and CTCF co-bound sites reveal active promoter specific histone marks (H3K27ac and H3K4me3). These results describe the genomic context and role of COUP-TFII in the cell-type specific transcriptional programs. Furthermore, we report that the VEGFA gene regulated by shared COUP-TFII and CTCF co-bound regulatory elements is involved in long-range looping in a cell-type-independent manner. These findings provide a genomic insight into the regulation and angiogenic role of COUP-TFII.

10.1016/j.jbiotec.2019.05.305



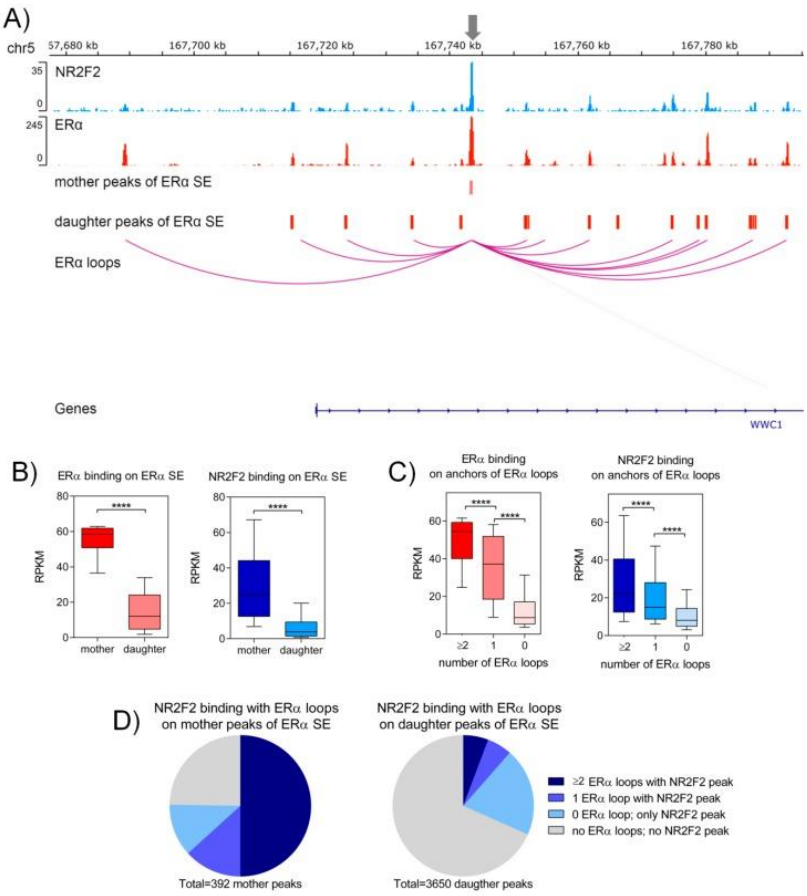
Super-enhancers (SEs) are clusters of highly active enhancers, regulating cell type-specific and disease-related genes, including oncogenes. The individual regulatory regions within SEs might be simultaneously bound by different transcription factors (TFs) and co-regulators, which together establish a chromatin environment conducting to effective transcription. While cells with distinct TF profiles can have different functions, how different cells control overlapping genetic programs remains a question. In this paper, we show that the construction of estrogen receptor alpha-driven SEs is tissue-specific, both collaborating TFs and the active SE components greatly differ between human breast cancer-derived MCF-7 and endometrial cancer-derived Ishikawa cells; nonetheless, SEs common to both cell lines have similar transcriptional outputs. These results delineate that despite the existence of a combinatorial code allowing alternative SE construction, a single master regulator might be able to determine the overall activity of SEs.

doi.org/10.3390/ijms21051630



Nuclear Receptor Subfamily 2 Group F Member 2 (NR2F2) is a member of the steroid/thyroid hormone receptor superfamily with a crucial role in organogenesis, angiogenesis, cardiovascular development and tumorigenesis. However, there is limited knowledge about the cistrome and transcriptome of NR2F2 in breast cancer. In this study, we mapped the regulatory mechanism by NR2F2 using functional genomic methods. To investigate the clinical significance of NR2F2 in breast cancer, The Cancer Genome Atlas (TCGA) data were used. These results show that a high *NR2F2* is associated with better survival of a specific subset of patients, namely those with luminal A breast cancer. Therefore, genome-wide NR2F2 and estrogen receptor alpha (ERα) binding sites were mapped in luminal A breast cancer cells using chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq), revealing that most NR2F2 overlap with ERα that are co-occupied by forkhead box A1 (FOXA1) and GATA binding protein 3 (GATA3) in active enhancer regions. NR2F2 overlaps with highly frequent ERα chromatin interactions, which are essential for the formation of ERα-bound super-enhancers. In the process of the transcriptome profiling of NR2F2-depleted breast cancer cells such differentially expressed genes have been identified that are involved in endocrine therapy resistance and are also ERα target genes. Overall, these findings demonstrate that the NR2F2 nuclear receptor has a key role in ERα-mediated transcription and it can offer a potential therapeutic target in patients with luminal A breast cancer.

10.3390/ijms21061910



The tumor grade of endometrioid endometrial cancer is used as an independent marker of prognosis and a key component in clinical decision making. It is reported that between grades 1 and 3, however, the intermediate grade 2 carries limited information; thus, patients with grade 2 tumors are at risk of both under- and overtreatment. We used RNA-sequencing data from the TCGA project and machine learning to develop a model which can correctly classify grade 1 and grade 3 samples. We used the trained model on grade 2 patients to subdivide them into low-risk and high-risk groups. With iterative retraining, we selected the most relevant 12 transcripts to build a simplified model without losing accuracy. Both models had a high AUC of 0.93. In both cases, there was a significant difference in the relapse-free survivals of the newly identified grade 2 subgroups. Both models could identify grade 2 patients that have a higher risk of relapse. Our approach overcomes the subjective components of the histological evaluation. The developed method can be automated to perform a prescreening of the samples before a final decision is made by pathologists. Our translational approach based on machine learning methods could allow for better therapeutic planning for grade 2 endometrial cancer patients.

https://doi.org/10.3390/cancers13174348