Актуальные проблемы общей и клинической биохимии – 2025 : сборник материалов республиканской научно-практической конференции с международным участием, посвящённой 100-летию со дня рождения академика Ю. М. Островского, Гродно, 27 июня 2025 года. – Гродно: Гродненский государственный медицинский университет, 2025. – С. 223–226.

4. G-quadruplex structures in the human genome as novel therapeutic targets / J. Bidzinska [et al.]. // Molecules. – 2013. – Vol. 18, № 10. – P. 12368–12395.

5. National Center for Biotechnology Information // National Library of Medicine (US). – URL: https://www.ncbi.nlm.nih.gov/datasets/ (date of access: 19.03.2026).

6. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins / S. K. Mishra [et al.]. // Scientific reports. – 2016. – Vol. 6, № 1. – P. 38144.

УДК [577.213.3:577.117.2]:618.19-006.6

R. Vuggu

Scientific supervisor: Associate Professor, PhD, A. N. Koval

Educational Establishment
"Gomel State Medical University"
Gomel, Republic of Belarus

INTEGRATED IN SILICO ANALYSIS OF G-QUADRUPLEX FORMING SEQUENCES AND EXPRESSION PATTERNS OF THE *ESR1* GENE IN BREAST CANCER

Introduction

Breast Cancer (BC) is a highly heterogeneous disease, remaining a primary cause of cancer-related mortality worldwide. Its clinical management is dictated by the presence or absence of specific receptors, leading to the classification into four primary molecular subtypes: Luminal A (ER+, PR+, HER2–), Luminal B (ER+, PR+/–, HER2+/–), HER2–Enriched, and Basal-like (Triple-Negative Breast Cancer, TNBC) [1, 2].

The *ESR1* gene encodes the Estrogen Receptor alpha (ERα), a ligand-activated transcription factor. Approximately 80% of breast cancers are hormone receptor-dependent, meaning they rely on estrogen signaling to fuel tumor growth and survival. When estrogen binds to the receptor encoded by *ESR1*, it translocates to the nucleus, binding to Estrogen Response Elements (EREs) to trigger the transcription of genes involved in cell proliferation and anti-apoptosis. *ESR1* expression is inextricably linked to Progesterone Receptor (PR) status, as PR is a downstream target of ER signaling. The co-expression of *ESR1* and PR serves as a functional indicator of an intact estrogen signaling pathway and a critical biomarker for predicting the success of endocrine therapies like Tamoxifen or Aromatase Inhibitors [3].

G-quadruplexes (G4) are four-stranded DNA structures forming in guanine-rich regions, recognized as epigenetic regulators that influence gene expression by modulating transcription factor binding and RNA polymerase processivity [4]. Given the central role of *ESR1* in driving luminal BCs, identifying G4–forming sequences within this gene could uncover novel regulatory mechanisms explaining inter-individual variations in gene expression contributing to disease susceptibility or treatment response.

Goal

To perform an integrated in silico analysis of the human *ESR1* gene by: (1) identifying and characterizing putative G-quadruplex forming sequences (PQS) within the gene, and [2] analyzing its expression patterns across different breast cancer subtypes using public genomic data, to evaluate the potential regulatory significance of G4 structures in the context of gynecological oncology.

Material and methods of research

G-Quadruplex Prediction. The complete genomic sequence of *ESR1* (Gene ID: 2099, NC_000006.12) was retrieved from NCBI Datasets (https://www.ncbi.nlm.nih.gov/datasets/gene/) in FASTA format. The sequence was analyzed using G4IPDB, a database and tool for predicting putative G-quadruplex-forming sequences in both sense and antisense strands. Analysis was performed using default G4 motif detection parameters (G≥3NxG≥3NxG≥3NxG≥3, x≥1). The output includes strand, start position, sequence, length, and confidence scores (cG, cC, cG/cC). A cG/cC score>20 indicates a high-confidence G4; "INF" denotes sequences with exceptionally high guanine content (cC=0) [5].

Assessment of ESR1 Expression in Breast Cancer. Transcriptomic data from invasive breast carcinoma (BRCA) were retrieved from The Cancer Genome Atlas (TCGA) through the UCSC Xena platform. Specimens were divided into three categories: non-tumorous normal breast tissue (n=113), ER+/PR+ breast tumors (n=566), and other breast cancer subtypes (HER2+, TNBC; n=281). Relative ESR1 expression values (log2(norm_count+1)) were evaluated across these groups. Differences were analyzed statistically using Student's t-test, where $p < 0.001$ indicated statistical significance. Visualization employed box plots.

The results of the research and their discussion

G-Quadruplex Architecture within the ESR1 Locus. G4IPDB screening detected 1,331 putative quadruplex structures (PQS) in the ESR1 gene (~300 kb): 678 located on the forward strand (50.9%) and 653 on the reverse strand (49.1%), indicating symmetrical distribution. Larger intronic segments contain notably greater numbers of PQS, including several exhibiting extreme thermodynamic stabilities (cG/cC >100 or "INF"), specifically at coordinates 12694, 68728, and 208353 (Table 1). These intronic G-quadruplexes may influence alternative splicing and transcription elongation pauses, potentially modulating the generation of various ESR1 isoforms linked to clinical resistance against hormonal therapies [3].

Table 1 – High-confidence PQS in the ESR1 gene

Strand	Start (bp)	Putative G4 Sequence (5'–>3')	Length	cG/cC
Sense	12694	GGGGGAGTGGGGAGAGGGG	20	**INF**
Sense	68728	GGTTGTGGGGGGCAGGGGGTGG	23	100.0
Sense	123620	GGGGGGTTGGGGGTCAAGGGG	22	111.0
Sense	208353	GGGGTGGGGGGAGGGGGGAGGG	23	**INF**
Sense	427686	GGGGTTGGGGGGCTGGGGGAGGG	24	121.0
Anti	209882	GGGGGTGGGGGGCTAGGGAAGGG	23	111.0
Anti	352515	GGGGGTGGGGGAAGGGAGGG	20	**INF**

Note. Positions based on NCBI Datasets files. "INF" indicates infinite cG/cC score, signifying very high predicted stability.

ESR1 Expression Profile Across Breast Cancer Subtypes. Evaluation of TCGA BRCA data demonstrated clear differences in ESR1 expression among breast cancer molecular subtypes (Figure 1). ER+/PR+ cases displayed markedly elevated ESR1 levels ($p < 0.001$) relative to normal breast tissue and all other tumor subtypes, particularly hormone receptor-negative and triple-negative forms.

Conversely, other BC subtypes (HER2+ and TNBC) displayed significantly lower expression of *ESR1* ($p<0.001$) than that of non-cancerous normal breast tissue. These data confirm that *ESR1* expression, and consequently enhanced estrogen signaling, plays an important role specifically in luminal BC subtypes [6].

Integrating G4 Landscapes with ESR1 Expression. The differential expression of *ESR1* between BC subtypes may be partially explained by its distinctive G4 landscape. The high expression

of *ESR1* in luminal tumors suggests that its chromatin is in an "open" configuration, potentially facilitated by G4–resolving helicases that prevent the formation of repressive G4 structures. Conversely, the significantly lower expression in TNBC and other subtypes could be associated with the formation of stable G4 structures, particularly in the promoter or first intron, acting as transcriptional roadblocks. This aligns with previous work comparing G-quadruplex content in human genes regulating apoptosis, which demonstrated the regulatory potential of such structures [7]. The identification of several "INF"-score G4s within *ESR1* supports their potential to form stable, regulatory structures in vivo. In TNBC, where *ESR1* is silenced, these G4s could represent an additional layer of epigenetic repression, potentially interacting with DNA methylation or histone modifications to maintain the silenced state [8].

Conclusions

A total of 1,331 PQS were identified in the *ESR1* gene, with a balanced distribution between sense and antisense strands.

ESR1 contains multiple PQS with exceptionally high predicted stability (cG/cC>100 or "INF"), predominantly located in large intronic regions, suggesting potential roles in regulating alternative splicing and transcript variant expression.

Analysis of TCGA data confirmed significantly higher *ESR1* expression in ER+/PR+ breast cancer subtypes ($p<0.001$) compared to normal tissue and other BC subtypes, validating its central role in luminal tumors.

The distinctive G4 landscape of *ESR1* may contribute to its differential expression across breast cancer subtypes, with stable G4 structures potentially acting as repressive elements in hormone-independent and TNBC tumors.

Identified G4 motifs in the *ESR1* gene represent potential molecular targets for therapeutic intervention, particularly for strategies aimed at downregulating ESR1 expression in hormone-resistant or TNBC subtypes.

LITERATURE

1. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries / H. Sung [et al.]. // CA: A Cancer Journal for Clinicians. – 2021. – Vol. 71, № 3. – P. 209–249.

2. Cancer statistics, 2025 / R. L. Siegel [et al.]. // CA: A Cancer Journal for Clinicians. – 2025. – Vol. 75, № 1. – P. 10–45.

3. Jeselsohn, R. ESR1 mutations – a mechanism for acquired endocrine resistance in breast cancer / R. Jeselsohn, G. Buchwalter, C. De Angelis // Nature Reviews Clinical Oncology. – 2015. – Vol. 12, № 10. – P. 573–583.

4. Rhodes, D. G-quadruplexes and their regulatory roles in biology / D. Rhodes, H. J. Lipps // Nucleic Acids Research. – 2015. – Vol. 43, № 18. – P. 8627–8637.

5. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins / S. K. Mishra [et al.]. // Scientific Reports. – 2016. – Vol. 6. – P. 38144.

6. Comprehensive molecular portraits of human breast tumours / C. Lynda [et al.]. // Nature. – 2012. – Vol. 490, № 7418. – P. 61–70.

7. Бобровник, Д. А. Сравнение содержания гуаниновых квадруплексов в генах человека, регулирующих апоптоз / Д. А. Бобровник // Проблемы и перспективы развития современной медицины : Сборник научных статей XIX Республиканской научно-практической конференции с международным участием студентов и молодых ученых: в 6 т., Гомель, 5–6 мая 2022 года. – Гомель: Гомельский государственный медицинский университет, 2022. – С. 89–91.

7. Бенько, А. С. Мутации в гене C9orf72, митохондриальная дисфункция и роль гуаниновых квадруплексов в нейродегенерации / А. С. Бенько, А. М. Пашкевич // Молодые исследователи – регионам : материалы Международной научной конференции, Вологда, 22 апреля 2025 года. – Вологда : Вологодский государственный университет, 2025. – С. 347–349.