Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Long-read full-length transcriptome sequencing reveals multilayered sex-associated isoform regulation in thyroid carcinoma

Dilinigaer Dilinigaer
The Affiliated Tumor Hospital of Xinjiang Medical University
Yiliyaer Rousuli
517081945@qq.com
The Affiliated Tumor Hospital of Xinjiang Medical University
Alibiyati Aini
The Affiliated Tumor Hospital of Xinjiang Medical University
Dai Ran
The Affiliated Tumor Hospital of Xinjiang Medical University

Research Article

Keywords: thyroid carcinoma, sex differences, Oxford Nanopore Technologies, alternative splicing, alternative polyadenylation, differential transcript usage

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11143834/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Long-read full-length transcriptome sequencing reveals multilayered sex-associated isoform regulation in thyroid carcinoma

Dilinigaer 1, Alibiyati Aini 1, Dai Ran 1, Yiliyaer Rousuli 1*
1 Department of Head and Neck Surgery, The Affiliated Tumor Hospital of Xinjiang Medical University, Urumqi 830000, China
* Corresponding author: Yiliyaer Rousuli (Email: 517081945@qq.com)
ORCID: Dilinigaer, 0009-0006-6819-718X (https://orcid.org/0009-0006-6819-718X)

Abstract

Background: Sex disparities in thyroid carcinoma are well recognized, yet their molecular basis at the transcript isoform level remains poorly understood. We aimed to determine whether sex-associated molecular differences extend beyond gene-level expression to coordinated remodeling of transcript isoforms.
Methods: Tumor tissues from ten thyroid carcinoma patients (four males, six females) were profiled by Oxford Nanopore Technologies (ONT) full-length transcriptome sequencing. Gene-level, transcript-level and post-transcriptional analyses, including differential expression, alternative splicing, alternative polyadenylation and differential transcript usage, were integrated. External validation was performed by exon-usage analysis of the TCGA-THCA cohort (505 primary tumors).
Results: We identified 14,420 genes and 32,269 transcripts, including 1,064 novel genes and 9,755 novel transcripts. Transcript-level analysis detected 2,052 differentially expressed transcripts, nearly twice the number of differentially expressed genes (1,144), with 627 alterations detectable only at the isoform level. Sex-associated analysis revealed 68 differential splicing events (predominantly exon skipping), 340 alternative polyadenylation events, 94 3′UTR remodeling events and 956 transcripts with altered isoform preference. Integration across layers yielded 56 consistently dysregulated candidates, represented by SNHG5 and PTCSC2. External TCGA-THCA exon-usage analysis reproduced sex-associated signal at JPX and SNHG5 that was undetectable at the gene-expression level. Using GTEx normal thyroid (v8, n = 578; v10, n = 587) and TCGA adjacent-normal tissue (n = 59) as references, only 28 genes (26 in v10) showed sex-biased expression in normal thyroid, almost all sex-chromosomal or immunoglobulin genes, and no autosomal gene passed correction in tumours; 1,917 of the 2,139 sex-associated event genes (89.6%) showed no sex-biased gene-level expression in either normal thyroid or tumour.
Conclusions: Sex-associated regulation in thyroid carcinoma is mediated through coordinated changes in transcript abundance and transcript identity, establishing isoform-resolved profiling as an essential complement to conventional gene-level analysis of endocrine tumor heterogeneity.
Keywords: thyroid carcinoma; sex differences; Oxford Nanopore Technologies; alternative splicing; alternative polyadenylation; differential transcript usage

Background

Thyroid carcinoma is the most common malignancy of the endocrine system and represents an increasing global health burden [1,2]. Although differentiated thyroid carcinomas generally exhibit favorable clinical outcomes, considerable heterogeneity exists in tumor initiation, progression, recurrence, and molecular characteristics [3]. Among the factors contributing to thyroid carcinoma heterogeneity, sex disparity represents a well-recognized but incompletely understood phenomenon. Epidemiological studies have consistently demonstrated a higher incidence of thyroid carcinoma in females than in males, while differences in clinical outcomes

and disease characteristics between sexes have also been reported [1,4]. These observations suggest that biological sex may influence thyroid carcinoma development and progression. However, the molecular mechanisms underlying sex-associated differences remain largely unclear. Previous studies have investigated multiple factors potentially contributing to sex disparities in thyroid carcinoma, including sex hormone signaling, sex chromosome regulation, genomic alterations, and immune-related mechanisms [5,6]. The Cancer Genome Atlas study further characterized the genomic landscape of papillary thyroid carcinoma and revealed major molecular alterations associated with thyroid tumorigenesis [7]. However, most previous investigations have focused primarily on DNA-level alterations or gene-level expression changes. Such approaches may not fully capture transcriptomic complexity because individual genes can generate multiple transcript isoforms with distinct structural characteristics and biological functions. Recent studies have demonstrated that transcript-level regulation, including alternative splicing (AS), alternative polyadenylation (APA), and differential transcript usage (DTU), represents an important regulatory layer involved in cancer development and progression [8,9]. Changes in transcript isoform composition can influence RNA stability, protein diversity, and cellular functions even without substantial alterations in total gene expression [10,11]. Therefore, transcript isoform remodeling may represent an overlooked mechanism contributing to sex-associated heterogeneity in thyroid carcinoma.

Conventional short-read RNA sequencing has been widely used for transcriptome profiling and has greatly contributed to the identification of differentially expressed genes in cancer [12]. However, because short-read sequencing relies on fragmented RNA molecules, accurate reconstruction of full-length transcripts often requires computational inference, limiting the resolution of complex transcript structures, highly similar isoforms, and transcript-specific regulation [13]. Long-read RNA sequencing technologies provide an effective strategy to overcome these limitations by enabling direct characterization of full-length transcript molecules [14]. Oxford Nanopore Technologies (ONT)-based sequencing can capture complete transcript structures and facilitate high-resolution identification of transcript isoforms, alternative splicing events, alternative polyadenylation patterns, and differential transcript usage [15]. By providing an isoform-resolved view of transcriptomic regulation, long-read sequencing has revealed previously unrecognized layers of molecular complexity in cancer and other diseases [16,17].

In this study, we sought to determine whether sex-associated molecular differences in thyroid carcinoma extend beyond conventional gene expression changes and involve transcript isoform remodeling. Using Oxford Nanopore Technologies-based full-length transcriptome sequencing, we generated an isoform-resolved transcriptomic landscape of thyroid carcinoma samples from male and female patients. We systematically characterized sex-associated alterations across multiple regulatory layers, including differential gene and transcript expression, alternative splicing, alternative polyadenylation, and differential transcript usage. Through integrated analysis of these transcriptomic features, we identified coordinated isoform remodeling events associated with sex differences in thyroid carcinoma. This study provides new insights into the molecular basis of sex disparities in thyroid carcinoma and highlights transcript isoform remodeling as an additional regulatory layer contributing to tumor heterogeneity. To our knowledge, this represents the first isoform-resolved, multilayer characterization of sex-associated transcript regulation in thyroid carcinoma using full-length long-read sequencing.

Methods

Clinical samples and study design

Ten thyroid carcinoma tissue samples were included in this study, consisting of four male thyroid carcinoma samples (M_TC_1–M_TC_4) and six female thyroid carcinoma samples (F_TC_1–F_TC_6). This study aimed to characterize sex-associated transcriptomic differences in thyroid carcinoma at both gene and isoform levels using Oxford Nanopore Technologies (ONT)-based full-length transcriptome sequencing. All samples were collected following institutional ethical approval and informed consent procedures. Clinical characteristics, including age, histological subtype, TNM stage, and tumor size, were recorded for all patients and are summarized in Supplementary Table S1. No significant differences in clinical characteristics were observed

between the male and female groups (Fisher exact test for categorical variables and Student t test for continuous variables).

RNA extraction and ONT full-length transcriptome sequencing

Total RNA was extracted from thyroid carcinoma tissues according to the manufacturer's instructions. RNA quality was assessed using NanoDrop spectrophotometry, Qubit fluorometry, and agarose gel electrophoresis. Full-length transcriptome libraries were constructed using poly(A)-selected RNA. mRNA molecules were enriched using Oligo-dT primers, followed by reverse transcription to generate full-length cDNA. Libraries were sequenced using Oxford Nanopore Technologies. Nanopore long-read RNA sequencing enables direct characterization of transcript structures and transcript isoform diversity [18].

ONT sequencing data processing and quality control

Raw nanopore sequencing signals were processed using Dorado v0.8.2 for base calling. Reads were filtered using Chopper v0.7.0, and full-length transcript candidates were identified using Pychopper v2.7.9. Sequencing quality was assessed using NanoComp v1.23.1 and NanoPlot v1.42.0 [19]. Sequence statistics were summarized using SeqKit v2.9.0.

Full-length transcript identification and annotation

High-quality full-length reads were aligned to the human reference genome GRCh38.p14 using minimap2 v2.27-r1193 with splice-aware parameters (-ax splice -uf -k14) [20]. Genome annotation was based on GENCODE v49 annotation (Ensembl release 115) [21]. Transcript reconstruction was performed using StringTie v2.1.6 [22], and assembled transcript structures were compared with reference annotation using gffcompare v0.12.6 [23]. SQANTI3 v5.2 was used for long-read transcriptome quality control, transcript classification, and isoform characterization [24].

Transcript quantification and differential expression analysis

Transcript and gene expression levels were quantified using Salmon v1.7.0 [25]. Raw read count matrices generated from Salmon were used as input for differential expression analysis. Differential expression analysis was performed using DESeq2 v1.38.3 [26]. P values were adjusted for multiple testing using the Benjamini–Hochberg (BH) procedure. Genes and transcripts with $|\log2(\text{fold change})| \geq 0.585$ (fold change ≥ 1.5 or ≤ 0.667) and an adjusted P value (FDR) <0.05 were considered statistically significant. TPM values were used only for visualization analyses, including principal component analysis, correlation analysis, and heatmap visualization. Differentially expressed genes (DEGs) and differentially expressed transcripts (DETs) were identified between male and female thyroid carcinoma samples. Sequencing depth, full-length read counts, and mapping rates were compared between the male and female groups to exclude systematic coverage bias (Supplementary Table S2).

Alternative splicing analysis

Alternative splicing events were analyzed using SUPPA2 v2.3 [27]. Six major alternative splicing patterns were evaluated, including skipped exon (SE), alternative 5′ splice site (A5), alternative 3′ splice site (A3), retained intron (RI), alternative first exon (AF), and alternative last exon (AL). Differential alternative splicing events were defined as events with $|\Delta PSI| \geq 0.05$ and a BH-

adjusted P value ≤0.05.

Alternative polyadenylation analysis

Alternative polyadenylation (APA) events were identified using TAPAS (2018.5.26) [28]. APA analysis was performed using RefSeq transcript annotation with parameter -l 76. Differential APA events were identified using fold change ≥2 and a BH-adjusted P value ≤0.05. NanopLen v0.9.0 was additionally applied for differential poly(A) length analysis.

Differential transcript usage analysis

Differential transcript usage (DTU) analysis was performed using DRIMSeq v1.30.0 [29]. Transcript isoforms were considered significantly differentially utilized when |log2(fold change)| ≥0.585 and the BH-adjusted P value was <0.05. DTU events were integrated with DEG, DET, alternative splicing, and APA analyses to characterize isoform-level transcriptomic remodeling associated with sex differences.

Functional annotation and enrichment analysis

Novel transcripts were functionally annotated against NR, Pfam, UniProt, Gene Ontology (GO), KEGG, KOG, and transcription factor databases. KEGG functional annotation was performed using KOBAS v3.0 [30]. GO and KEGG enrichment analyses were performed using hypergeometric tests with whole-genome annotation as background. Terms with P value <0.05 were considered nominally enriched, and BH correction was further applied to identify terms remaining significant after multiple-testing adjustment. Database annotations were based on GENCODE [21], GO [31], UniProt [32], and Pfam [33].

TCGA-THCA external validation

Publicly available TCGA-THCA data were used for external validation at two levels. Gene-level expression (UCSC Xena TOIL RSEM, log2(x+1)) was compared between male (n = 136) and female (n = 369) primary tumors for the 56 multi-layer candidate genes using two-sided Mann-Whitney U tests with Benjamini-Hochberg (BH) correction, and with a linear model adjusting for age and AJCC stage. For isoform-level validation, exon-level expression (Xena HiSeqV2_exon) was extracted for candidate and control loci; per-sample exon inclusion was computed as the log2 ratio of exon signal to the locus mean and compared between sexes with Mann-Whitney U tests (BH correction across all tested exons). Sex-chromosome genes (XIST, KDM5D, RPS4Y1, UTY) served as internal positive controls.

Normal thyroid and adjacent-normal tissue reference analysis

To distinguish tumour-specific sex effects from baseline sex differences intrinsic to normal thyroid tissue, two independent external references were introduced. Normal thyroid baseline: raw RNA-seq count matrices for thyroid tissue were downloaded from the GTEx Portal in two releases (v8: 653 samples, of which 578 passed quality control at RNA integrity number [RIN] ≥ 6; v10: 684 samples, 587 passing) [34]; sex and age were obtained from the GTEx subject phenotype files. Tumour and adjacent-normal reference: gene-level STAR-Counts quantification for TCGA-THCA was downloaded from the GDC data portal (GENCODE v36 annotation; 564 samples comprising 505 primary tumours [136 male, 369 female] and 59 adjacent-normal tissues [17 male, 42 female]). Because demographic fields are no longer exposed in the GDC open tier, sex and age were retrieved from cBioPortal (thpa_tcga_gdc cohort) [35], mapped to GDC case identifiers without missingness.

Count matrices were normalised to counts per million (CPM) and log2(CPM + 1) transformed,

retaining genes with CPM > 1 in at least 20% of samples. In GTEx, a linear model log2CPM ~ sex + age + RIN was fitted to test the sex effect within normal thyroid; in TCGA, log2CPM ~ sex + tissue + age + sex × tissue was fitted and the interaction term evaluated by nested model F test. Multiple testing was controlled by the Benjamini-Hochberg procedure, with significance defined as FDR < 0.05 and |log2 fold change| > 1. All sex-associated event genes identified in this study were classified as: class A, sex-differential at baseline in normal thyroid (significant in both GTEx v8 and v10); class B, tumour-specific sex-biased expression (significant in TCGA tumours or the interaction term, but not in normal thyroid); class C, expression-independent (no sex-biased gene-level expression in either external cohort); class D, not evaluable (absent from the external reference sets). XIST and the Y-linked genes KDM5D, RPS4Y1, DDX3Y, UTY, EIF1AY and USP9Y served as internal positive controls for the correctness of sex labels. Statistical implementation used Python 3.13 (NumPy, SciPy).

Results

3.1 ONT full-length transcriptome sequencing reveals an isoform-resolved landscape of thyroid carcinoma

To establish an isoform-resolved transcriptomic landscape of thyroid carcinoma, ONT full-length transcriptome sequencing was performed on tumor tissues from four male and six female patients. In total, 14,420 genes and 32,269 transcripts were identified, of which 1,064 genes and 9,755 transcripts were absent from the reference annotation and were therefore defined as novel (Figure 1A). Because ONT single-molecule reads span entire transcript molecules, these novel candidates represent full-length transcript structures captured directly, without the assembly biases inherent to short-read reconstruction.

Sequencing quality control confirmed high data quality across all ten samples. A total of 63.1 million reads (61.1 Gb raw bases) were generated, with 57.0 million full-length non-chimeric (FLNC) reads at an FLNC rate of 90.8%. Read N50 ranged from 891 to 1,064 bp (raw) and 674 to 859 bp (FLNC), with mean read quality of 16.3–17.1. Mapping rates to the GRCh38.p14 reference genome ranged from 98.95% to 99.78%. Female (n = 6) and male (n = 4) samples showed no significant differences in FLNC rate, mean read quality, or mapping rate (all $P > 0.05$), confirming comparable sequencing quality between sex groups (Table S2).

Most genes were associated with one or two transcript isoforms, whereas a smaller subset exhibited complex isoform composition with multiple transcript variants (Figure 1B), indicating that isoform diversity in thyroid carcinoma is widespread but gene-dependent. In addition, known and novel transcripts displayed distinct exon-number distributions (Figure 1C), consistent with the capacity of full-length sequencing to resolve transcript architectures that remain incompletely represented in current genome annotations.

Each sample yielded an average of 6.3 million total reads, of which 90.8% were full-length non-chimeric (FLNC) reads, with a read N50 of 968 bp and a mean mapping rate of 99.4% to the GRCh38 reference genome (Supplementary Table S2).

Transcript-level analysis reveals sex-associated alterations beyond gene-level resolution

Principal component analysis (PCA) based on gene expression profiles showed a separation tendency between male and female thyroid carcinoma samples (Figure 2A). Gene-level differential expression analysis identified 1,144 differentially expressed genes (DEGs) between the two groups (fold change [FC] ≥ 1.5 or ≤ 0.67, $P < 0.05$), of which 526 were upregulated and 618 downregulated in male tumors (Figure 2B). Hierarchical clustering based on DEG expression profiles further grouped the samples according to sex (Figure 2C), confirming robust sex-associated differences at the conventional gene-expression level.

Consistent with biological sex as the grouping variable, sex chromosome-linked genes, including XIST on chromosome X and the Y-chromosome genes DDX3Y, RPS4Y1, KDM5D, and UTY,

ranked among the most strongly differentiated transcripts between the two groups and served as internal positive controls for the differential analysis. After excluding sex chromosome genes, 1,092 autosomal DEGs (95.5% of 1,144 total; 491 up and 601 down in male) and 1,947 autosomal DETs (94.9% of 2,052 total; 875 up and 1,072 down in male) remained, and all downstream splicing, APA, and DTU analyses were likewise evaluated with and without sex chromosome loci to ensure that the reported isoform remodeling was not driven by sex chromosome dosage effects.
We next asked whether the same comparison at isoform resolution would capture additional alterations. Transcript-level differential expression analysis identified 2,052 differentially expressed transcripts (DETs), including 945 upregulated and 1,107 downregulated transcripts in male tumors (Figure 2D). The number of transcript-level events was nearly twice that detected at the gene level (2,052 DETs versus 1,144 DEGs; Figure 2E), indicating that a substantial fraction of sex-associated regulation in thyroid carcinoma occurs at the isoform level and is not fully resolved by gene-level quantification alone.

3.3 Integrated gene- and transcript-level analysis distinguishes shared from isoform-specific regulation

To determine whether transcript-level alterations reflect gene abundance changes or constitute isoform-specific regulation, DEG and DET profiles were integrated. A total of 1,040 genes were shared between the two datasets, whereas 627 differential events were detected only at the transcript level and 97 only at the gene level (Figure 3A). The predominance of shared events indicates that most sex-associated expression changes are concordantly captured at both levels, whereas the transcript-only events represent a regulatory layer that is invisible to conventional gene-level analysis.
Genes simultaneously identified as DEGs and DETs were significantly enriched in Gene Ontology (GO) biological processes related to protein transport, apoptotic processes, transcriptional regulation, angiogenesis, cell migration, vesicle-mediated transport, and RNA splicing (Figure 3B). KEGG pathway analysis of the same gene set showed nominal enrichment of endocytosis, RNA transport, lysosome function, metabolic pathways, the citrate cycle, and cancer-associated signaling pathways (Figure 3C); however, none of these pathways remained significant after multiple-testing correction, and they are therefore reported as enrichment trends. Together, these enrichment patterns indicate that sex-associated differences converge on tumor-relevant processes, including intracellular trafficking, cellular metabolism, and RNA processing. These findings suggest that sex-associated regulation in thyroid carcinoma is mediated through coordinated changes in both transcript abundance and transcript identity. In parallel, the 624 genes detectable only at the transcript level (Figure 3A) showed significant enrichment for protein synthesis and mitochondrial energy metabolism: Gene Ontology analysis identified cytoplasmic translation (GO:0002181, adjusted $P = 9.9 \times 10^{-22}$), translation (GO:0006412, adjusted $P = 6.5 \times 10^{-20}$), ribosome biogenesis (GO:0042254, adjusted $P = 2.7 \times 10^{-5}$), and oxidative phosphorylation (GO:0006119, adjusted $P = 7.4 \times 10^{-4}$) among the top BH-significant terms (Supplementary Table S5). KEGG pathway analysis further confirmed ribosome (adjusted $P = 7.9 \times 10^{-18}$) and oxidative phosphorylation (adjusted $P = 6.8 \times 10^{-5}$) as the dominant enriched pathways, with the ATP-synthesis, mitochondrial electron-transport and ribosomal-protein families contributing the bulk of the enriched signal. These patterns indicate that isoform-specific, gene-level-invisible sex-associated regulation in PTC converges on translational machinery and mitochondrial oxidative metabolism—two axes repeatedly implicated in thyroid tumor growth and aggressiveness.

3.4 Sex-associated alternative splicing reveals post-transcriptional regulation in thyroid carcinoma

Because transcript-level alterations could not be fully attributed to changes in gene abundance, we next investigated whether post-transcriptional processing contributes to sex-associated transcriptomic differences. Seven major alternative splicing (AS) event types were profiled, including skipped exon (SE), mutually exclusive exon (MXE), alternative 5′ and 3′ splice sites (A5SS, A3SS), retained intron (RI), and alternative first and last exons (AF, AL). AS events were

broadly detected across all samples, with SE being the most abundant type, followed by A5SS, A3SS, and RI (Figure 4A), demonstrating extensive alternative transcript processing in thyroid carcinoma.

Differential AS analysis between male and female tumors ($|\Delta PSI| \geq 0.05$, $P \leq 0.05$) identified 68 sex-associated events, among which SE was the predominant category (24 events), followed by A5SS (15), RI (14), A3SS (8), AF (5), and AL (2); no significant differential MXE events were detected (Figure 4B). GO enrichment analysis of the associated genes revealed biological processes related to spermatogenesis and translation, cellular components including mitochondria, extracellular exosomes, and ribonucleoprotein complexes, and molecular functions dominated by protein-, RNA-, nucleotide-, and ATP-binding activities (Figure 4C). KEGG analysis showed enrichment trends in metabolic pathways, amino acid biosynthesis, RNA degradation, the tricarboxylic acid cycle, oxidative phosphorylation, and spliceosome-related pathways (Figure 4D); however, none of these pathways remained significant after multiple-testing correction, and they are therefore reported as functional trends rather than definitive enrichments.

These results indicate that sex-associated transcriptomic differences in thyroid carcinoma involve alterations in alternative transcript processing, adding a post-transcriptional regulatory layer beyond differential expression.

Sex-associated alternative polyadenylation drives transcript 3′ end remodeling

Alternative polyadenylation (APA) generates transcripts with distinct 3′ ends and thereby modifies their regulatory potential. Given the extensive isoform-level alterations described above, we investigated whether sex differences in thyroid carcinoma also involve differential APA regulation. APA sites were broadly detected across samples, with most genes containing one dominant polyadenylation site and a subset exhibiting multiple sites (Figure 5A), indicating widespread 3′ end heterogeneity. Differential APA analysis identified 340 sex-associated APA events, comprising 121 events with increased and 219 with decreased APA usage in male tumors (Figure 5B).

GO enrichment analysis performed separately for the two directions showed that genes with increased APA usage were mainly associated with response to oxidative stress, cell differentiation, RNA splicing, and regulation of DNA-templated transcription, whereas genes with decreased APA usage were primarily related to cell adhesion, homophilic cell-cell adhesion, cell migration, and regulation of apoptotic processes (Figure 5C). Because APA frequently acts through modification of 3′ untranslated regions (3′UTRs), we further quantified APA-associated length changes and identified 94 remodeling events, including 52 3′UTR shortening and 42 3′UTR lengthening events (Figure 5D). Thus, sex-associated APA regulation contributes to bidirectional transcript 3′ end remodeling, representing an additional layer of post-transcriptional regulation in thyroid carcinoma.

Multi-layer integration reveals sex-associated isoform remodeling in thyroid carcinoma

Although the analyses above revealed multiple regulatory layers, whether these alterations converge on preferential usage of specific isoforms remained unclear. Differential transcript usage (DTU) analysis identified 956 transcripts with significantly altered usage between male and female tumors, including 477 with increased and 479 with decreased usage in males (Figure 6A), demonstrating that sex-associated regulation involves selective modulation of isoform preference rather than transcript abundance alone.

Integration of DEG, DET, and DTU profiles identified 56 candidates consistently altered across all three regulatory layers (Figure 6B), representing potential key molecular signatures of sex differences in thyroid carcinoma. The full list of these multi-layer candidates is provided in Supplementary Table S8. Notably, the X chromosome-linked long non-coding RNA JPX was among these multi-layer candidates, consistent with a contribution of X chromosome-mediated regulation to sex-associated isoform remodeling. Representative candidates exhibited distinct

isoform usage patterns between the sexes, including the DTU-associated transcripts of SNHG5 and PTCSC2 and the APA-associated transcripts of PBX1 and VPS72 (Figure 6C). Leveraging the single-molecule full-length resolution of ONT sequencing, we further visualized the complete isoform architectures of representative loci: the SNHG5 locus displayed distinct full-length transcript configurations underlying its differential usage, whereas PBX1 exhibited structural differences involving alternative polyadenylation sites and 3′UTR remodeling (Figure 6D).
Collectively, these findings demonstrate that sex differences in thyroid carcinoma are associated with extensive isoform remodeling involving differential transcript usage and transcript structural variation. The integration of gene-level, transcript-level, and isoform-level analyses highlights the unique capability of ONT full-length transcriptome sequencing to characterize regulatory complexity that is difficult to resolve using conventional short-read approaches.

External validation in TCGA-THCA

To assess the generalizability of the sex-associated signatures, we performed two complementary validations in the independent TCGA-THCA cohort (505 primary tumors; 136 male, 369 female). At the gene-expression level, none of the 48 matchable candidates reached significance after multiple-testing correction (FDR < 0.05), whereas sex-chromosome controls (XIST, KDM5D, RPS4Y1, UTY) were overwhelmingly significant (e.g., XIST P = 2.8e-257), confirming adequate power and correct pipeline behavior.
We then tested the central premise of this study, that sex-associated regulation acts on transcript identity rather than transcript abundance, using exon-usage analysis as a short-read-accessible proxy for isoform remodeling. Among 157 exons across nine testable loci, all 13 KDM5D exons were strongly sex-differential (minimum FDR = 1.4e-78), validating the metric. Critically, exon-usage differences were detected at two multi-layer candidates: three exons of the X-linked lncRNA JPX (FDR = 0.043-0.047) and one exon of SNHG5 (FDR = 0.047; Figure 7B). PTCSC2 showed a sub-threshold trend (FDR = 0.053), and nominal-significance exons were enriched across candidate loci (18/134 autosomal exons at P < 0.05, binomial P = 0.003). Thus, in a cohort more than 50-fold larger than our discovery set, sex-associated signal is undetectable at the gene level but reproducible at the exon-usage level, a dissociation that directly supports the need for full-length long-read transcriptomics to resolve sex-biased isoform regulation in thyroid carcinoma.

Comparison with normal thyroid and adjacent-normal references

To determine whether the sex-associated events detected here merely reflect baseline sex differences intrinsic to normal thyroid, we compared our findings against GTEx normal thyroid and TCGA adjacent-normal tissue. In GTEx v8 normal thyroid (578 samples; 385 male, 193 female), only 28 genes showed significant sex-biased expression after adjusting for age and RNA integrity (FDR < 0.05, |log2FC| > 1; 18 male-biased, 10 female-biased); independent replication in GTEx v10 (587 samples) yielded 26 genes, with concordant conclusions. Nearly all of these genes were sex-chromosomal, namely XIST and TSIX with higher expression in females and 15 Y-linked genes (RPS4Y1, KDM5D, DDX3Y, UTY, EIF1AY, USP9Y, TXLNGY, TBL1Y, NLGN4Y, TMSB4Y and others), or immunoglobulin genes (IGHG1, IGKC, IGLC1/2/3, JCHAIN and others), the latter reflecting sex differences in immune cell infiltration composition rather than a sex programme of thyroid epithelium. Effect sizes at positive controls were as expected (XIST log2FC = −8.79; RPS4Y1 = +7.33; KDM5D = +6.90), confirming reliable sex labels (Figure 7A). Thus, apart from sex-chromosome dosage and immune infiltration composition, normal thyroid shows no significant autosomal sex-biased expression.
In TCGA-THCA, only 18 genes were sex-biased among 505 tumours, all sex-chromosomal or Y-linked, with no autosomal gene passing FDR < 0.05; findings in the 59 adjacent-normal tissues were consistent. The sex-by-tissue interaction test identified only four significant genes (TBL1Y, TMSB4Y, ZFY and AC011297.1), all Y-linked loci.
We next mapped all 2,139 sex-associated event genes detected in this study (1,137 DEG, 1,667 DET, 367 DTU, 53 alternative splicing, 329 APA) onto these references and classified them into four groups: class A (n = 11, 0.5%), sex-differential at baseline in normal thyroid, all sex-

chromosomal except CALCA, with effect sizes in our cohort correlating closely with those in GTEx normal thyroid (Pearson r = 0.50, n = 896), indicating that our cohort detected no gene-level sex bias beyond the normal baseline; class B (tumour-specific sex-biased expression), n = 0; class C (n = 1,917, 89.6%), showing no sex-biased gene-level expression in either normal thyroid or thyroid tumours despite significant isoform usage, alternative splicing or alternative polyadenylation events; and class D (n = 211, 9.9%), not evaluable because the gene was absent from the external reference sets (Figure 8). The predominance of class C indicates that the sex-associated events reported here cannot, in the great majority of cases, be explained by baseline sex differences or by differences in gene expression abundance, and instead point to sex-associated regulation operating at the post-transcriptional level, independent of expression level.
One class A gene warrants specific mention: CALCA, which encodes the calcitonin precursor secreted by thyroid parafollicular C cells, was sex-biased in normal thyroid as well and should therefore be regarded as a baseline feature of normal tissue, to be distinguished from tumour-associated alterations.

Discussion

Sex disparities represent a clinically recognized feature of thyroid carcinoma, with differences observed in disease incidence, clinical characteristics, and tumor behavior between male and female patients; however, the molecular mechanisms underlying these differences remain incompletely understood, particularly at the transcript isoform level [36,37]. In this study, we applied Oxford Nanopore Technologies (ONT)-based full-length transcriptome sequencing to systematically characterize sex-associated transcriptomic alterations in thyroid carcinoma. Long-read RNA sequencing enables direct characterization of full-length transcripts and provides improved resolution for transcript isoform diversity, alternative splicing, and transcript-level regulatory events that are difficult to fully resolve using conventional short-read approaches [38,39]. By integrating gene expression, transcript-level regulation, alternative splicing, alternative polyadenylation, and differential transcript usage analyses, we demonstrate that sex-associated molecular differences in thyroid carcinoma extend beyond gene abundance changes and involve extensive remodeling of transcript isoform landscapes. These findings establish transcript isoform remodeling as an additional molecular dimension underlying sex-associated heterogeneity in thyroid carcinoma.
Sex-associated heterogeneity in cancer is a complex biological phenomenon regulated by multiple molecular layers, including sex hormone signaling, sex chromosome-linked mechanisms, genetic alterations, and immune microenvironmental differences [40,41]. In thyroid carcinoma, previous studies have primarily focused on genomic alterations, hormonal regulation, and immune-related mechanisms to explain sex-associated differences. However, these approaches mainly capture molecular changes at the DNA or gene-expression level and may overlook additional regulatory processes occurring after transcription initiation. Our findings extend the current understanding of sex disparities in thyroid carcinoma by demonstrating that transcript-level regulation represents an important but previously underexplored layer of molecular heterogeneity.
Long-read RNA sequencing provides a unique opportunity to investigate this regulatory layer by directly capturing full-length transcript molecules. Unlike short-read sequencing approaches that require computational reconstruction of fragmented RNA molecules, long-read sequencing enables more accurate characterization of complete transcript structures and complex isoforms [39]. Previous studies have demonstrated that cancer transcriptomes contain extensive isoform diversity and that transcript usage alterations may not always parallel changes in overall gene expression [16]. Consistent with these observations, our study revealed that transcript-level alterations substantially expanded the molecular landscape associated with sex differences in thyroid carcinoma. These findings highlight the importance of moving beyond gene-centric analysis toward isoform-resolved characterization of tumor heterogeneity.
Importantly, this central premise received orthogonal support from external validation in the independent TCGA-THCA cohort (505 primary tumors). Although none of the matchable candidates reached gene-level significance after multiple-testing correction, sex-associated signal was reproduced at the exon-usage level at candidate loci, including the X-linked lncRNA JPX and SNHG5, and nominal-significance exons were enriched across candidate loci. This dissociation,

observed in a cohort more than 50-fold larger than our discovery set, indicates that sex-biased isoform regulation is reproducible across independent cohorts yet remains largely invisible to conventional gene-centric analysis, directly underscoring the necessity of full-length long-read transcriptomics. Together with recent long-read characterization of thyroid neoplasia [38], these findings position isoform-resolved profiling as an essential complement to conventional short-read transcriptomics in endocrine tumors.

Beyond gene expression: transcript identity as a new regulatory layer

Gene-level differential expression analysis remains a widely used strategy for identifying cancer-associated molecular changes; however, transcriptional regulation is inherently multilayered and cannot be fully represented by changes in total gene abundance alone. In our study, a proportion of differentially expressed transcripts (DETs) overlapped with differentially expressed genes (DEGs), indicating that changes in gene expression remain an important component of sex-associated regulation. Nevertheless, a considerable number of transcript-level alterations occurred without corresponding gene-level changes, suggesting that transcript-specific regulation provides additional biological information beyond conventional expression profiling. Previous studies have shown that isoform switching and differential transcript usage are widespread across cancers and may alter functional outputs without substantial changes in gene expression levels [42]. Therefore, our findings suggest that sex-associated heterogeneity in thyroid carcinoma involves coordinated regulation at both transcriptional abundance and transcript identity levels.

Long-read sequencing reveals hidden layers of tumor heterogeneity

Alternative splicing represents another critical layer of post-transcriptional regulation that expands transcriptomic diversity by generating multiple isoforms with distinct regulatory and functional properties from a single gene locus. Dysregulation of alternative splicing is increasingly recognized as a hallmark of cancer, as aberrant splicing can generate isoforms with altered functional properties that influence tumor cell proliferation, apoptosis, and migration [43,8]. Alterations in splicing regulators and the RNA-processing machinery can further reshape transcript composition and contribute to tumor heterogeneity through isoform-specific regulation [43]. In the present study, sex-associated differential splicing events, together with the enrichment of DEG/DET-associated genes in RNA splicing-related biological processes, suggest that altered RNA processing constitutes an important component of sex-associated transcriptomic remodeling in thyroid carcinoma. These findings indicate that sex-associated molecular differences involve not only differential expression but also distinct regulation of how transcripts are processed and utilized.

Alternative polyadenylation represents another important layer of transcript regulation that influences transcript stability, localization, and translational potential. Although APA has received less attention than gene expression alterations in cancer studies, changes in transcript 3′ end structures can generate functionally distinct transcript isoforms and contribute to tumor heterogeneity. In our study, ONT-based full-length sequencing enabled direct characterization of APA-associated transcript end remodeling, revealing sex-associated APA alterations and differential 3′UTR remodeling patterns. These observations indicate that sex-associated transcript regulation extends beyond transcript abundance and splicing patterns to include regulation of transcript terminal structures, further supporting a multilayer model of isoform remodeling in thyroid carcinoma.

Integration of DEG, DET, and DTU analyses further revealed that sex-associated transcript regulation involves coordinated changes in transcript usage preferences. Differential transcript usage reflects selective regulation of transcript isoforms generated from the same gene locus and may produce distinct molecular outputs without requiring changes in total gene expression. Representative isoform remodeling events identified in our study, involving SNHG5, PTCSC2, PBX1, and VPS72, provide potential molecular candidates for understanding sex-associated differences in thyroid carcinoma biology. Although the functional consequences of these transcript alterations require further experimental validation, our findings suggest that isoform switching may represent an important mechanism contributing to tumor heterogeneity and sex-associated molecular diversity [44].

Limitations, clinical implications, and future directions

Several limitations should be acknowledged. First, as an exploratory, first-in-field investigation, the discovery cohort was limited to ten samples. Although external exon-usage analysis in TCGA-THCA (n = 505) independently reproduced sex-associated signal at candidate loci, the short-read nature of that cohort precluded validation at full-length isoform resolution, and larger multicenter cohorts with long-read profiling will be required to confirm and extend these findings. Second, the discovery cohort included tumour tissue only, without matched adjacent-normal samples. We therefore introduced GTEx normal thyroid (578-587 samples) and TCGA-THCA adjacent-normal tissue (n = 59) as external references: neither normal thyroid nor tumour tissue showed autosomal sex-biased gene-level expression, and 89.6% of the sex-associated event genes showed no sex-biased expression in either tissue type, so these isoform-level events are unlikely to be attributable to pre-existing baseline sex differences. Two caveats apply. The public reference data are short-read gene-level quantification and cannot provide an equivalent control at isoform resolution, so our normal-tissue control for isoform events was implemented as the binary criterion of whether the gene shows sex-biased expression; and the GTEx and TCGA cohorts differ from our study in population origin, sample handling and annotation version, and are not matched adjacent-normal tissue from the same patients. Long-read sequencing of population-matched paired adjacent-normal tissue therefore remains the definitive route to confirming the tumour specificity of these isoform events. In addition, although no significant differences in clinical characteristics were observed between the male and female groups (Supplementary Table S1), residual confounding by age, tumor stage, or histology cannot be completely excluded given the limited cohort size. Third, although this study identified candidate isoform remodeling events and their potential biological associations, functional consequences require further investigation using isoform-specific manipulation approaches and protein-level validation. Fourth, bulk transcriptomic sequencing cannot fully resolve the cellular origin and spatial distribution of sex-associated isoform regulation. Future integration of single-cell long-read sequencing and spatial transcriptomic approaches may enable cell-type-specific and spatially resolved characterization of transcript isoform regulation [45].

Collectively, this study establishes an isoform-resolved framework for investigating sex-associated molecular heterogeneity in thyroid carcinoma. Our findings demonstrate that sex differences are not solely reflected by changes in gene expression abundance but also involve coordinated remodeling of transcript structures, transcript processing, and isoform usage. These results highlight transcript isoform remodeling as an additional molecular dimension of tumor biology and suggest that long-read transcriptomics may provide complementary insights beyond conventional gene-centric approaches for understanding cancer heterogeneity.

Conclusions

This study provides an isoform-resolved transcriptomic landscape of sex-associated heterogeneity in thyroid carcinoma based on ONT full-length transcriptome sequencing. We demonstrate that sex-associated molecular differences extend substantially beyond gene-level expression changes and involve multilayered isoform regulation, encompassing differential transcript expression, alternative splicing, alternative polyadenylation, transcript 3′ end remodeling, and differential transcript usage. The identification of 56 candidates consistently altered across gene, transcript, and transcript usage layers, together with directly resolved full-length isoform structures of representative loci such as SNHG5 (a multi-layer candidate) and PBX1 (an APA-associated locus), offers concrete entry points for future functional investigation. These findings underscore the importance of incorporating transcript-level regulation into studies of cancer biology and establish a framework for dissecting sex-associated tumor heterogeneity with long-read transcriptomics. External reference analysis further indicated that the great majority of these isoform-level sex-associated events cannot be explained by baseline sex differences in normal thyroid or by differences in gene expression abundance. Validation in larger independent cohorts and functional characterization of the identified isoforms will be essential next steps toward translating these molecular insights into clinical relevance.

List of abbreviations

A3SS: alternative 3′ splice site

A5SS: alternative 5′ splice site

AF: alternative first exon

AL: alternative last exon

APA: alternative polyadenylation

AS: alternative splicing

BH: Benjamini–Hochberg

CPM: counts per million

DEG: differentially expressed gene

DET: differentially expressed transcript

DTU: differential transcript usage

FDR: false discovery rate

GDC: Genomic Data Commons

GO: Gene Ontology

GSA: Genome Sequence Archive

GTEx: Genotype-Tissue Expression

KEGG: Kyoto Encyclopedia of Genes and Genomes

lncRNA: long non-coding RNA

MXE: mutually exclusive exon

NGDC: National Genomics Data Center

ONT: Oxford Nanopore Technologies

PCA: principal component analysis

PSI: percent-spliced-in

PTC: papillary thyroid carcinoma

QC: quality control

RI: retained intron

RIN: RNA integrity number

SE: skipped exon

TCGA: The Cancer Genome Atlas

THCA: thyroid carcinoma (TCGA project code)

TPM: transcripts per million

UTR: untranslated region

Declarations

Ethics approval and consent to participate: This study was approved by the Ethics Committee of Xinjiang Medical University (Approval No. K-2026259; approved on 30 September 2026) and was performed in accordance with the principles of the Declaration of Helsinki and its later amendments. Written informed consent was obtained from all individual participants included in the study.

Consent for publication: Not applicable.

Availability of data and materials: The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive for Human (GSA-Human) (Genomics, Proteomics & Bioinformatics 2025) at the National Genomics Data Center (NGDC) (Nucleic Acids Res 2025), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA020897; BioProject: PRJCA073698), and are publicly accessible at https://ngdc.cncb.ac.cn/gsa-human/browse/HRA020897. The dataset comprises Oxford Nanopore Technologies (ONT) PromethION full-length transcriptome data from 10 papillary thyroid carcinoma surgical specimens (4 male and 6 female). All remaining data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests: The authors have no relevant financial or non-financial interests to disclose.

Funding: This work was supported by the Natural Science Youth Research Program of Xinjiang Medical University (Grant No. 2024XYZR33).

Authors' contributions: Dilinigaer: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. Alibiyati Aini: Methodology, Formal analysis, Validation, Writing – review & editing. Dai Ran: Resources, Data curation, Investigation, Writing – review & editing. Yiliyaer Rousuli: Conceptualization,

Supervision, Funding acquisition, Project administration, Writing – review & editing. All authors read and approved the final manuscript.

Acknowledgements: The authors thank the Department of Head and Neck Surgery of The Affiliated Tumor Hospital of Xinjiang Medical University for assistance with clinical sample collection and pathological diagnosis, and RuiXing Biotechnology for technical support in ONT full-length transcriptome sequencing. We also thank all patients who participated in this study.

References

[1] Miranda-Filho A, Lortet-Tieulent J, Bray F, et al. Thyroid cancer incidence trends by histology in 25 countries: a population-based study. Lancet Diabetes Endocrinol. 2021;9(4):225-234. https://doi.org/10.1016/s2213-8587(21)00027-9

[2] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263. https://doi.org/10.3322/caac.21834

[3] Póvoa AA, Teixeira E, Bella-Cueto MR, et al. Clinicopathological features as prognostic predictors of poor outcome in papillary thyroid carcinoma. Cancers (Basel). 2020;12(11):3186. https://doi.org/10.3390/cancers12113186

[4] Heinrich M, Blickle E, Jedamzik T, et al. Sex-specific differences in patients with differentiated thyroid carcinoma and their possible impact on survival. Eur J Nucl Med Mol Imaging. 2026;53(8):5042-5049. https://doi.org/10.1007/s00259-026-07845-y

[5] Park SH, Sugier PE, Asgari Y, et al. Reproductive factors, sex hormone levels, and differentiated thyroid cancer risk: a Mendelian randomization study. Thyroid. 2025;35(4):433-443. https://doi.org/10.1089/thy.2024.0548

[6] Shobab L, Simpson J, McCoy M, et al. Differentiated thyroid cancer is associated with sex-specific immune response. J Endocr Soc. 2026;10(1):bvaf174. https://doi.org/10.1210/jendso/bvaf174

[7] Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. Cell. 2014;159(3):676-690. https://doi.org/10.1016/j.cell.2014.09.050

[8] Bradley RK, Anczuków O. RNA splicing dysregulation and the hallmarks of cancer. Nat Rev Cancer. 2023;23(3):135-155. https://doi.org/10.1038/s41568-022-00541-7

[9] Zhang Y, Huang Z, Lu W, et al. Alternative cleavage and polyadenylation: key regulatory mechanisms in health and disease. RNA Biol. 2025;22(1):1-33. https://doi.org/10.1080/15476286.2025.2529033

[10] Mandal SD, Mukherjee S. Pan-cancer analysis of cancer-specific transcript isoforms reveals the regulatory impact of isoform switching on the alteration of the interplay between RBPs and miRNAs in cancers. J Biosci. 2025;50. https://doi.org/10.1007/s12038-025-00509-3

[11] Pan T, Lu L, Gao R. The splice of life: an isoform-centric view of disease, technology, and therapeutics. J Clin Invest. 2026;136(14):e207476. https://doi.org/10.1172/jci207476

[12] Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet. 2019;20(11):631-656. https://doi.org/10.1038/s41576-019-0150-2

[13] Dhar R, Karmakar A, Badal AK, et al. From fragments to full pictures: how long-read sequencing platforms are revolutionizing oncology research. Mol Biol Rep. 2026;53(1). https://doi.org/10.1007/s11033-026-12076-3

[14] Chen Z, He X. Application of third-generation sequencing in cancer research. Med Rev. 2021;1(2):150-171. https://doi.org/10.1515/mr-2021-0013

[15] Tan X, Wang P, Li Y, et al. Long-read sequencing reveals RNA splicing complexity in human diseases. Comput Struct Biotechnol J. 2026;35(1):0052. https://doi.org/10.34133/csbj.0052

[16] Wijeratne S, Gonzalez MEH, Roach K, et al. Full-length isoform concatenation sequencing to resolve cancer transcriptome complexity. BMC Genomics. 2024;25(1):122. https://doi.org/10.1186/s12864-024-10021-x

[17] Byrne A, Felton C, Stephenson W. Beyond counting: how single-cell long-read sequencing turns transcriptome complexity into precision targets. Front Oncol. 2026;16:1800370. https://doi.org/10.3389/fonc.2026.1800370

[18] Workman RE, Tang AD, Tang PS, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. Nature Methods. 2019;16:1297-1305. https://doi.org/10.1038/s41592-019-

0617-2
[19] De Coster W, D'Hert S, Schultz DT, et al. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666-2669. https://doi.org/10.1093/bioinformatics/bty149
[20] Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094-3100. https://doi.org/10.1093/bioinformatics/bty191
[21] Frankish A, Carbonell-Sala S, Diekhans S, et al. GENCODE: reference annotation for the human and mouse genomes in 2023. Nucleic Acids Research. 2023;51(D1):D942-D949. https://doi.org/10.1093/nar/gkac1071
[22] Kovaka S, Zimin AV, Pertea GM, et al. Transcriptome assembly from long-read RNA sequencing. Genome Biology. 2019;20:278. https://doi.org/10.1186/s13059-019-1910-1
[23] Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Research. 2020;9:304. https://doi.org/10.12688/f1000research.23297.2
[24] Pardo-Palacios FJ, Arzalluz-Luque A, Kondratova L, et al. SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms. Nature Methods. 2024;21:793-797. https://doi.org/10.1038/s41592-024-02229-2
[25] Patro R, Duggal G, Love MI, et al. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. 2017;14:417-419. https://doi.org/10.1038/nmeth.4197
[26] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8
[27] Trincado JL, Entizne JC, Hysenaj G, et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biology. 2018;19:40. https://doi.org/10.1186/s13059-018-1417-1
[28] Arefeen A, Liu J, Xiao X, Jiang T. TAPAS: tool for alternative polyadenylation site analysis. Bioinformatics. 2018;34:2521-2529. https://doi.org/10.1093/bioinformatics/bty110
[29] Nowicka M, Robinson MD. DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. F1000Research. 2016;5:1356. https://doi.org/10.12688/f1000research.8900.2
[30] Bu D, Luo H, Huo P, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Research. 2021;49(W1):W317-W325. https://doi.org/10.1093/nar/gkab447
[31] Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Research. 2021;49:D325-D334. https://doi.org/10.1093/nar/gkaa1113
[32] UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research. 2023;51:D523-D531. https://doi.org/10.1093/nar/gkac1052
[33] Mistry J, Chuguransky S, Williams L, et al. Pfam: The protein families database in 2021. Nucleic Acids Research. 2021;49:D412-D419. https://doi.org/10.1093/nar/gkaa913
[34] GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369:1318-1330. https://doi.org/10.1126/science.aaz1776
[35] de Bruijn I, Kundra R, Mastrogiacomo B, et al. Analysis and visualization of longitudinal genomic and clinical data from the AACR Project GENIE biopharma collaborative in cBioPortal. Cancer Res. 2023;83:3861-3867. https://doi.org/10.1158/0008-5472.CAN-23-0816
[36] Shobab L, Burman KD, Wartofsky L. Sex differences in differentiated thyroid cancer. Thyroid. 2022;32:224-235. https://doi.org/10.1089/thy.2021.0361
[37] Morand GB, Tessler I, Krasner J, et al. Investigation of genetic sex-specific molecular profile in well-differentiated thyroid cancer: is there a difference between females and males? Clin Otolaryngol. 2023;48:748-755. https://doi.org/10.1111/coa.14075
[38] Wang Y, Zou B, Zhang Y, et al. Comprehensive long-read sequencing analysis discloses the transcriptome features of papillary thyroid microcarcinoma. J Clin Endocrinol Metab. 2024;109:1263-1274. https://doi.org/10.1210/clinem/dgad695
[39] Ament IH, DeBruyne N, Wang F, et al. Long-read RNA sequencing: a transformative technology for exploring transcriptome complexity in human diseases. Mol Ther. 2025;33:883-894. https://doi.org/10.1016/j.ymthe.2024.11.025
[40] Haupt S, Caramia F, Klein SL, et al. Sex disparities matter in cancer development and therapy. Nat Rev Cancer. 2021;21:393-407. https://doi.org/10.1038/s41568-021-00348-y
[41] Bernardez B, Higuera O, Martinez-Callejo V, et al. Sex and gender differences in cancer

pathogenesis and pharmacology. Clin Transl Oncol. 2025;27(10):3837-3848. https://doi.org/10.1007/s12094-025-03894-1
[42] Dolgalev G, Poverennaya EV. Quantitative analysis of isoform switching in cancer. Int J Mol Sci. 2023;24:10065. https://doi.org/10.3390/ijms241210065
[43] Stanley RF, Abdel-Wahab O. Dysregulation and therapeutic targeting of RNA splicing in cancer. Nat Cancer. 2022;3:536-546. https://doi.org/10.1038/s43018-022-00384-z
[44] Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12:31-46. https://doi.org/10.1158/2159-8290.cd-21-1059
[45] Belchikov N, Ivanova E, Evsyukov I, et al. Understanding isoform expression by pairing long-read sequencing with single-cell and spatial technologies. Genome Res. 2024;34:1735-1748. https://doi.org/10.1101/gr.279640.124

Figure Legends

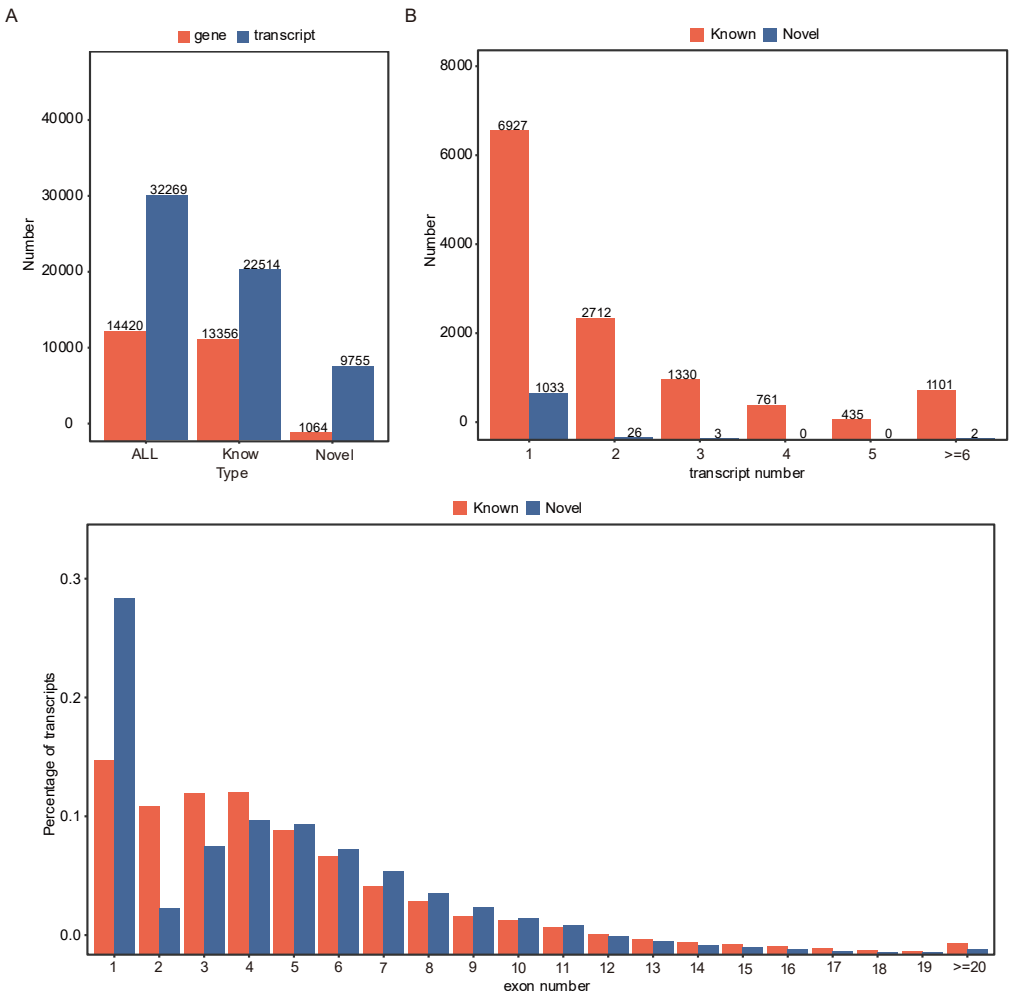


Figure 1. Comprehensive characterization of genes and transcripts identified by ONT full-length transcriptome sequencing.
(A) Classification of identified genes and transcripts into known and novel categories, with the numbers of total, known, and newly identified genes and transcripts.
(B) Distribution of genes containing different numbers of transcript isoforms.

(C) Distribution of exon numbers per transcript for known and novel transcripts.

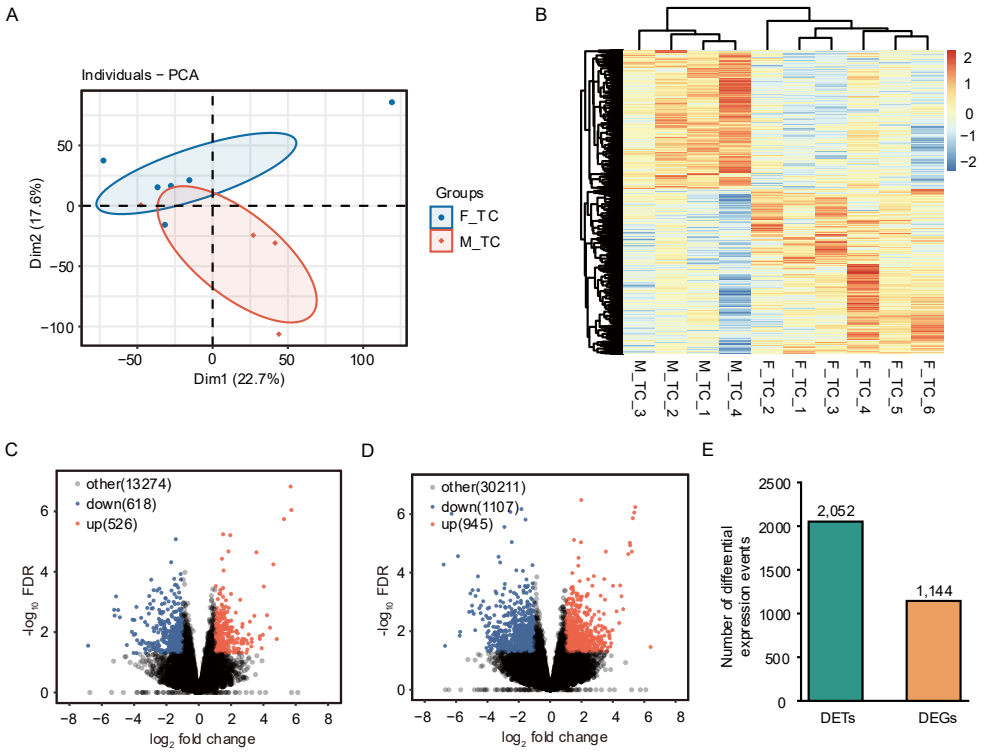


Figure 2. Gene-level and transcript-level differential expression reveals sex-associated transcriptional alterations in thyroid carcinoma.
(A) PCA of samples based on gene expression profiles. (B) Volcano plot of DEGs between male and female tumors. (C) Hierarchical clustering of samples based on DEG expression. (D) Volcano plot of DETs between male and female tumors. (E) Comparison of the numbers of DETs and DEGs identified between male and female thyroid carcinoma samples.

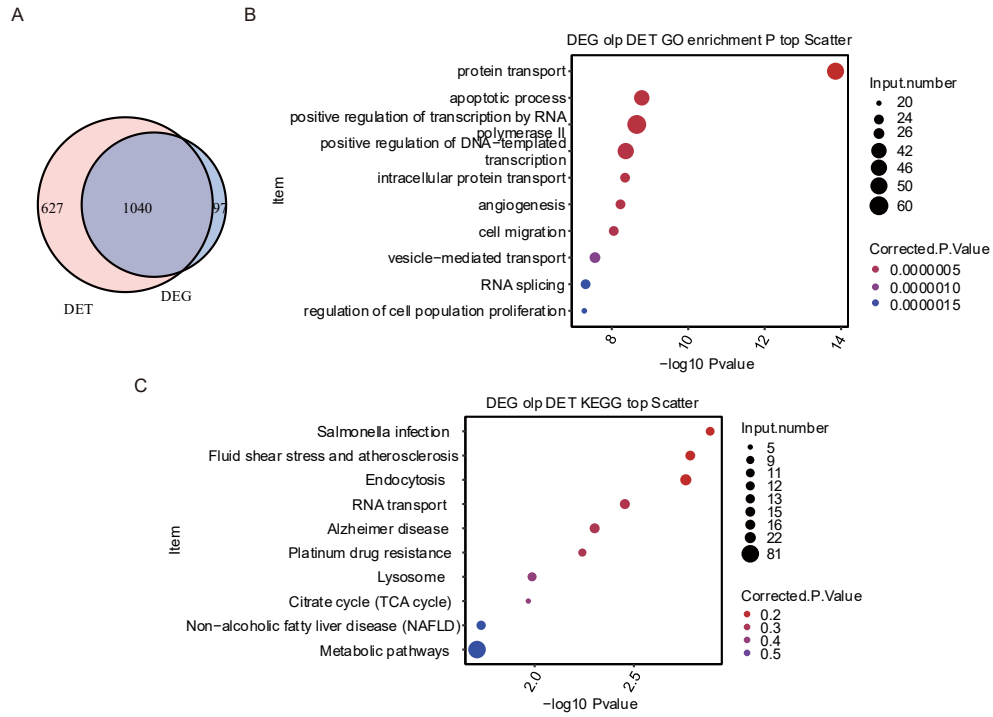


Figure 3. Integration of gene- and transcript-level alterations reveals transcriptomic complexity associated with sex differences in thyroid carcinoma.
(A) Venn diagram showing the overlap between DEGs and DETs. (B) GO enrichment analysis of genes simultaneously identified as DEGs and DETs. (C) KEGG pathway enrichment analysis of the same gene set. No pathway remained statistically significant after multiple-testing correction.

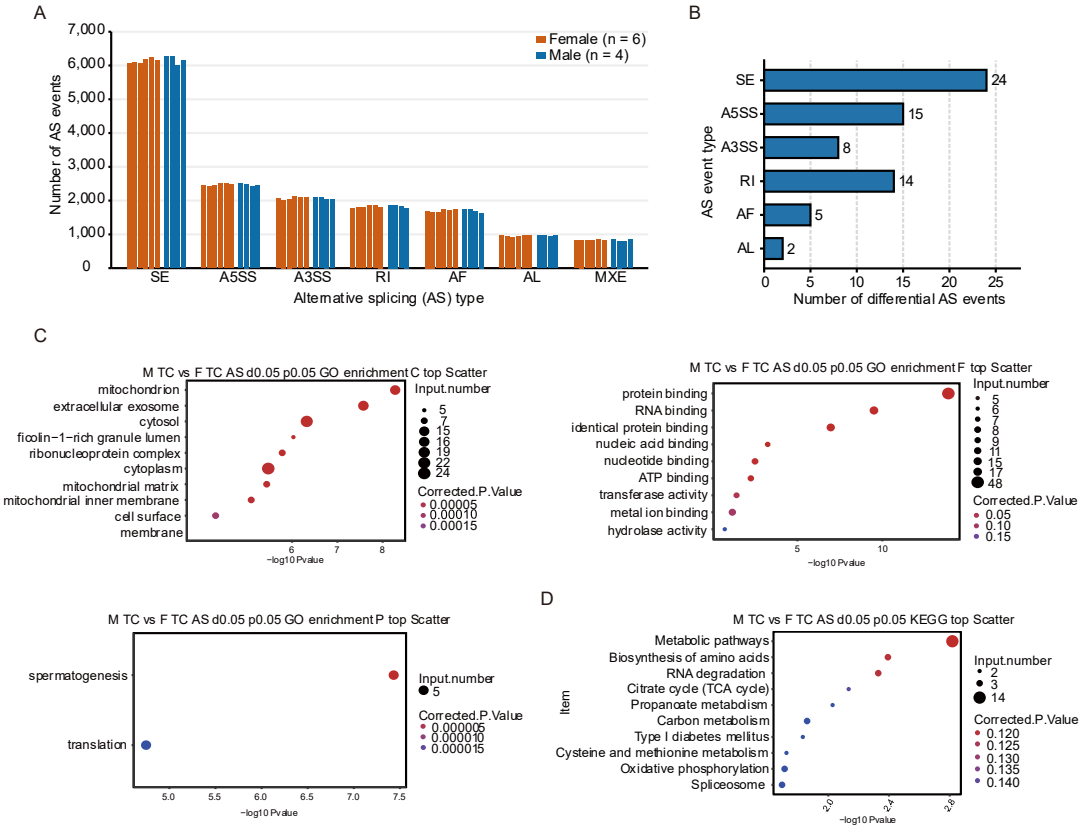


Figure 4. Sex-associated alternative splicing landscape reveals post-transcriptional regulation in thyroid carcinoma.
(A) Overview of AS events identified across male and female samples; seven event types were analyzed. (B) Distribution of the 68 sex-associated differential AS events by event type. (C) GO enrichment analysis of genes associated with differential AS events, shown for biological process, cellular component, and molecular function categories. (D) KEGG pathway analysis of the same genes; no pathway remained significant after multiple-testing correction.

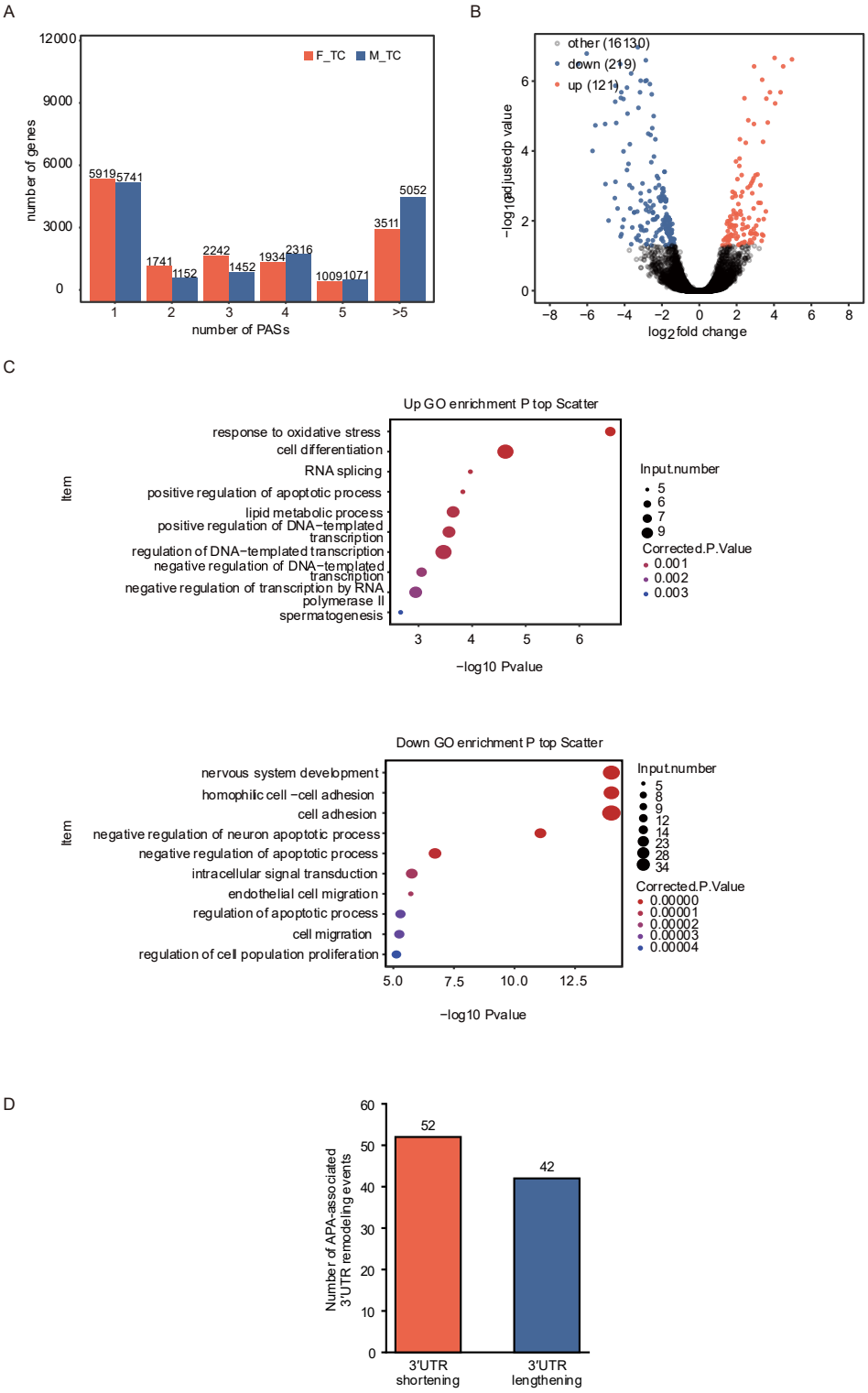


Figure 5. Sex-associated alternative polyadenylation landscape reveals transcript 3′ end remodeling in thyroid carcinoma.

(A) Distribution of APA site numbers per gene across samples. (B) Volcano plot of differential APA events between male and female tumors. (C) GO enrichment analysis of genes with increased and decreased APA usage. (D) Numbers of APA-associated 3′UTR remodeling events, classified as shortening (n = 52) or lengthening (n = 42).

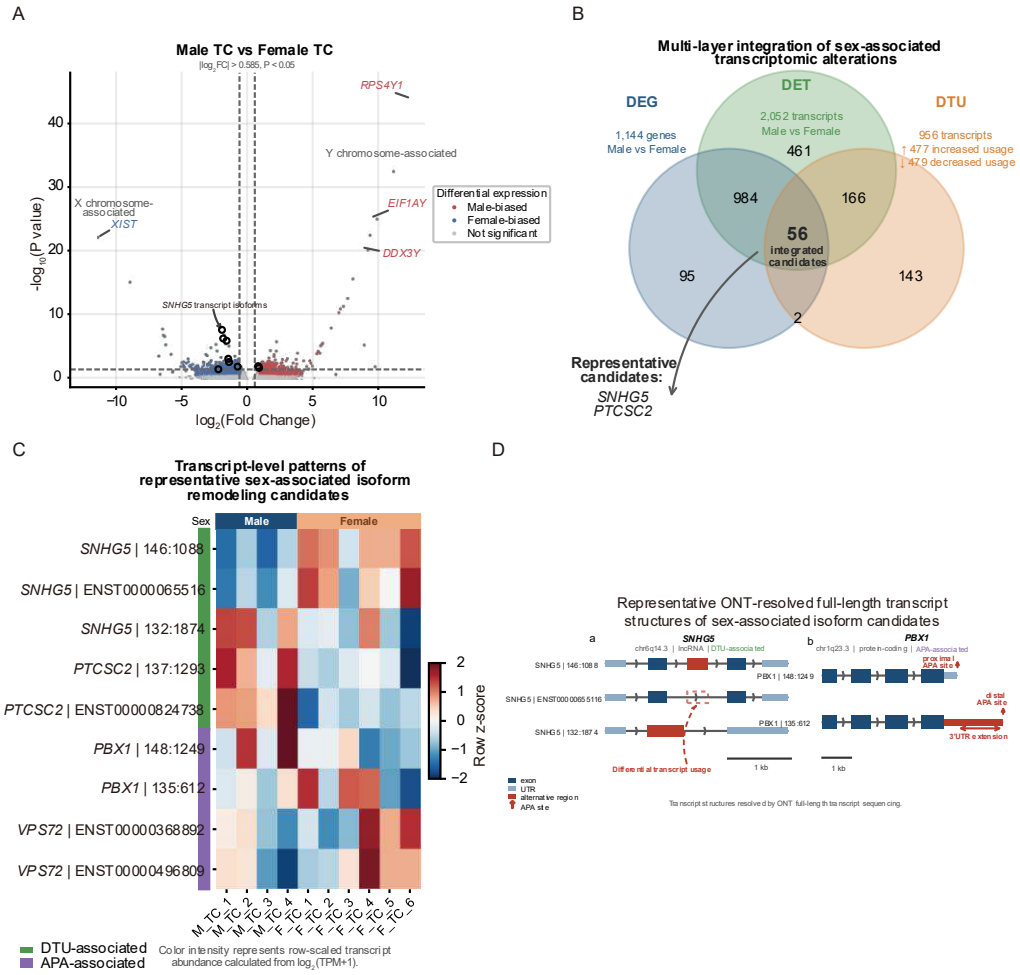


Figure 6. Multi-layer transcriptomic integration reveals sex-associated isoform remodeling in thyroid carcinoma.
(A) Volcano plot of DTU events between male and female tumors (477 increased, 479 decreased usage). (B) Venn diagram of the overlap among DEGs, DETs, and DTUs. For cross-layer integration, DETs and DTU transcripts were mapped to their parent genes, and numbers in the diagram represent gene-level counts; 56 candidates were shared across all three layers (see Supplementary Table S8). (C) Transcript usage patterns of representative candidates (SNHG5, PTCSC2, PBX1, VPS72). (D) ONT-resolved full-length transcript structures of SNHG5 and PBX1.

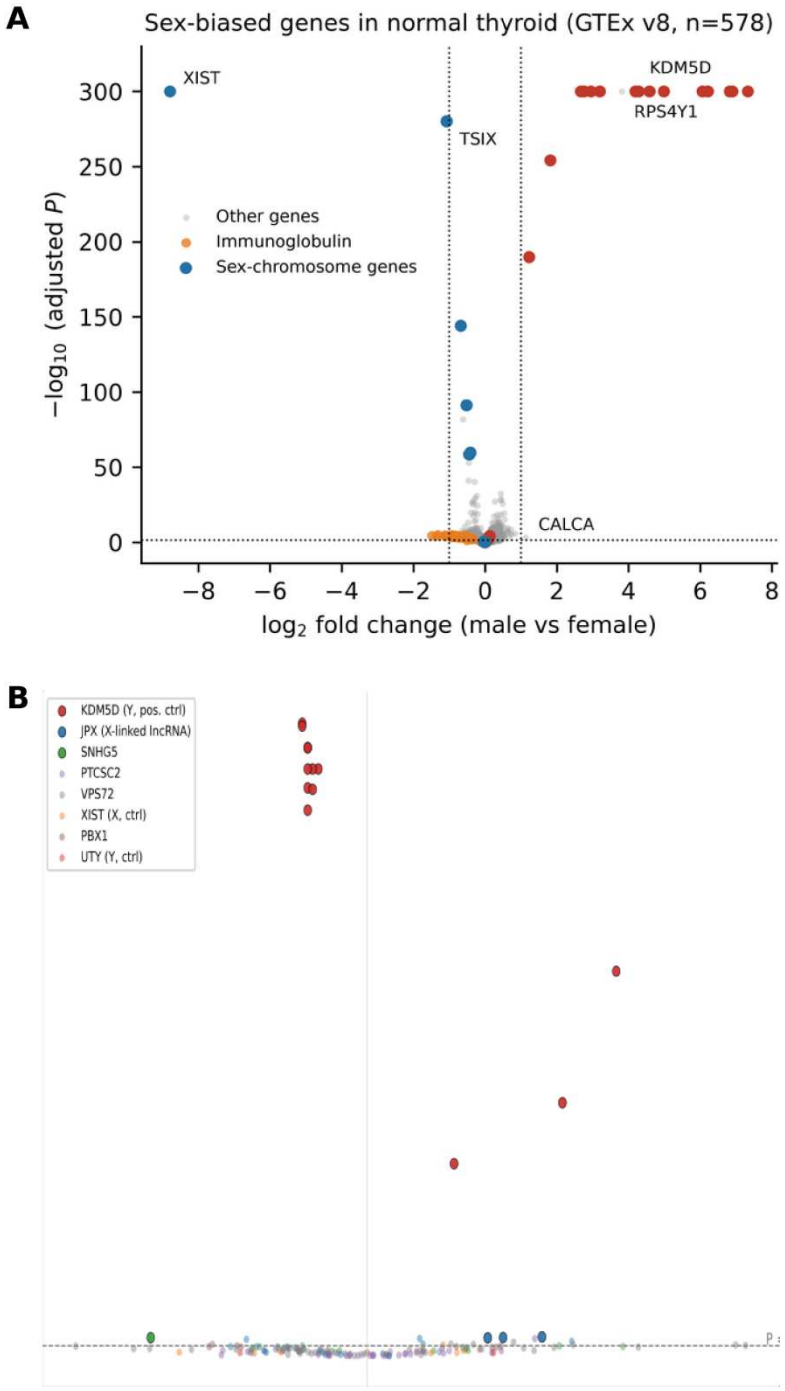


Figure 7. Sex-associated expression is absent from normal thyroid but detectable at the exon-usage level in tumours.

(A) Near-absence of autosomal sex-biased expression in normal thyroid tissue. Volcano plot of sex-differential genes in normal thyroid from GTEx v8 (578 samples; 385 males, 193 females). The x-axis shows log2 fold change (male minus female) and the y-axis shows −log10(FDR). A linear model of log2CPM ~ sex + age + RNA integrity number was fitted with Benjamini-Hochberg correction. Significance was defined as FDR < 0.05 and |log2FC| > 1, yielding 28 genes. Sex-chromosome genes are shown in red (male-biased) or blue (female-biased) and immunoglobulin genes in orange. Positive controls including XIST (log2FC = −8.79) and RPS4Y1 (+7.33) showed expected effect sizes. (B) External validation of sex-associated exon usage in TCGA-THCA. Volcano plot of exon-level sex differences in 505 TCGA-THCA samples (136 males vs. 369 females). Each point represents one exon from the nine candidate loci tested

(157 exons in total). The x-axis shows the difference in exon inclusion ratio (log2 exon-to-locus mean expression ratio, male minus female); the y-axis shows −log10(FDR) from Mann-Whitney U tests with Benjamini-Hochberg correction. The Y-chromosomal gene KDM5D (13 exons; minimum FDR = 1.4×10^{-78}) serves as the positive control. Among the candidate genes, JPX (3 exons) and SNHG5 (1 exon) show significant sex-biased exon usage (FDR < 0.05), whereas PTCSC2 shows a consistent sub-threshold trend (FDR = 0.053). Autosomal exons with nominally significant sex bias (18/134, binomial P = 0.003) support a broader, low-effect-size pattern of sex-associated exon usage.

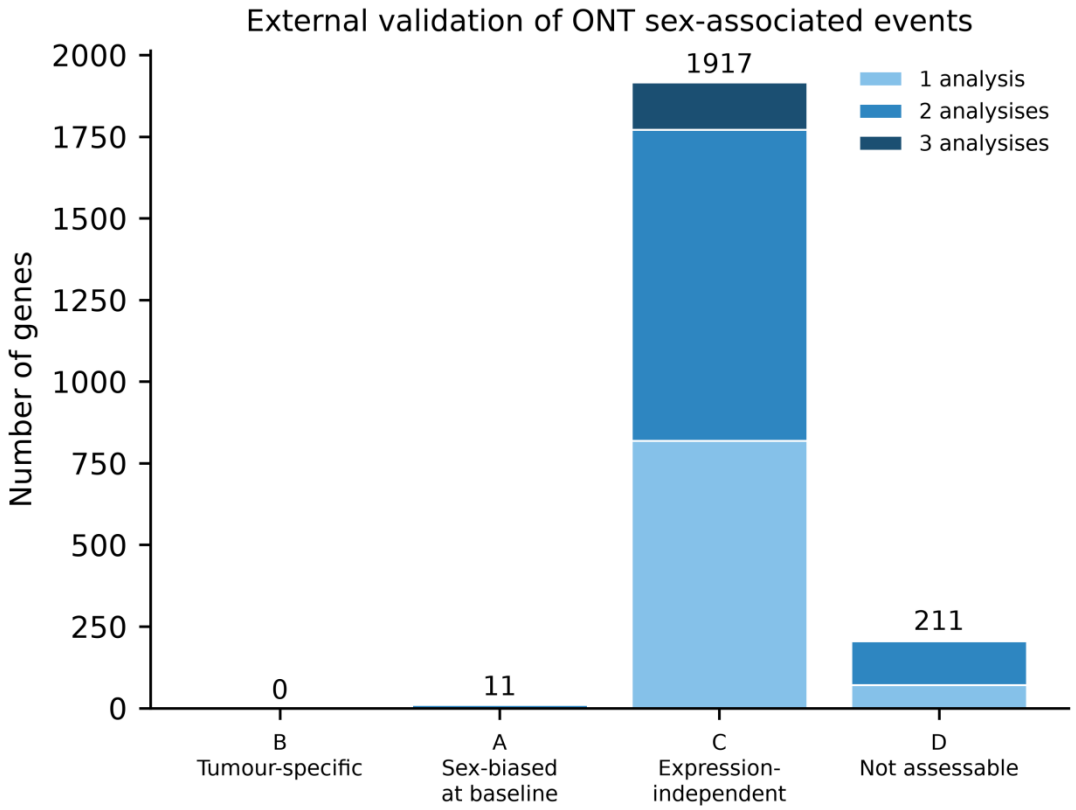


Figure 8. Classification of sex-associated event genes against external normal-tissue references. The 2,139 sex-associated event genes identified in this study were classified against external references into four groups: class A (n = 11, 0.5%), sex-differential in both GTEx v8 and v10 normal thyroid, all sex-chromosome genes except CALCA; class B (n = 0), tumour-specific sex-biased expression; class C (n = 1,917, 89.6%), no sex-biased gene-level expression in either normal thyroid or thyroid tumours, yet showing isoform usage, alternative splicing or alternative polyadenylation events; class D (n = 211, 9.9%), not evaluable because the gene was absent from the external reference sets. Shading within bars indicates the number of analyses in which each gene was detected.

Supplementary Tables

Supplementary Table S1. Clinical characteristics of the ten thyroid carcinoma patients.

Sample	Sex	Age (years)	Histology	TNM stage	Tumor size (cm)	Driver mutation
M_TC_1	Male	58	PTC, classical	pT1bN0M0	1.3	Not tested
M_TC_2	Male	34	PTC, classical + FV	pT1bN1aMx	1.3	Not tested
M_TC_3	Male	59	PTC,	pT3N0Mx	0.8	Not tested

Sample	Sex	Age (years)	Histology	TNM stage	Tumor size (cm)	Driver mutation
			classical			
M_TC_4	Male	35	PTC, classical	pT1aN0M0	0.8	Not tested
F_TC_1	Female	44	PTC, classical	pT1bN0M0	1.5	Not tested
F_TC_2	Female	31	PTC, classical	pT1aN0Mx	0.6	Not tested
F_TC_3	Female	48	PTC, follicular variant	pT1bN0Mx	1.4	Not tested
F_TC_4	Female	40	PTC, classical	pT1aN0Mx	0.6	Not tested
F_TC_5	Female	42	PTC, classical + FV	pT1aN0Mx	1.0	Not tested
F_TC_6	Female	68	PTC, tall cell	pT3bN1bMx	4.0	Not tested
P value (male vs female)	—	0.908	0.571	1.000	0.500	NA

Note: Categorical variables were compared between groups using Fisher exact test; continuous variables were compared using Student t test (or Mann–Whitney U test if non-normally distributed).

Supplementary Table S2. ONT full-length transcriptome sequencing quality metrics per sample.

Sample	Sex	Total reads	Full-length reads (FLNC)	FLNC (%)	Read N50 (bp)	Mean read quality	Mapping rate (%)
M_TC_1	Male	6,102,157	5,564,747	91.2	977	16.6	99.25
M_TC_2	Male	5,821,029	5,268,070	90.5	997	16.7	99.27
M_TC_3	Male	5,991,892	5,441,386	90.8	1,006	16.7	99.37
M_TC_4	Male	5,585,424	5,065,151	90.7	1,064	17.1	99.78
F_TC_1	Female	6,782,029	6,192,293	91.3	926	16.8	99.63
F_TC_2	Female	6,681,781	5,994,344	89.7	891	16.3	98.95
F_TC_3	Female	6,254,016	5,692,029	91.0	892	16.4	99.39
F_TC_4	Female	6,645,799	5,959,475	89.7	1,001	16.9	99.42
F_TC_5	Female	6,528,262	5,981,253	91.6	946	16.7	99.44
F_TC_6	Female	6,702,449	6,129,574	91.5	976	16.9	99.70
Male mean ± SD	—	5,875,126 ± 225,116	5,334,839 ± 217,102	90.8 ± 0.3	1,011 ± 37	16.8 ± 0.2	99.4 ± 0.2
Female mean ± SD	—	6,599,056 ± 188,288	5,991,495 ± 173,323	90.8 ± 0.9	939 ± 45	16.7 ± 0.3	99.4 ± 0.3
P value (male vs female)	—	0.010	0.010	0.610	0.038	0.914	0.610

Note: Metrics were summarized from NanoPlot/NanoComp and minimap2 alignment statistics. Group comparisons were performed using Mann–Whitney U test.

Supplementary Table S3. Complete list of differentially expressed genes (DEGs) between male and female thyroid carcinoma samples.(See Supplementary Excel file, Tables S1 – S8.)

Supplementary Table S4. Complete list of differentially expressed transcripts (DETs).(See Supplementary Excel file, Tables S1 – S8.)

Supplementary Table S5. Complete list of sex-associated differential alternative splicing

events.(See Supplementary Excel file, Tables S1 – S8.)

Supplementary Table S6. Complete list of differential alternative polyadenylation events and 3′ UTR remodeling events.(See Supplementary Excel file, Tables S1 – S8.)

Supplementary Table S7. Complete list of transcripts showing differential transcript usage (DTU) between male and female thyroid carcinoma samples.(See Supplementary Excel file, Tables S1 – S8.)

Supplementary Table S8. Multi-layer candidate genes consistently altered across DEG, DET, and DTU analyses.(See Supplementary Excel file, Tables S1 – S8.)

Supplementary Table S9. Classification of the sex-associated event genes against external normal-tissue references, with effect sizes and FDR values from GTEx v8/v10 and TCGA-THCA.(See Supplementary Excel file.)

Supplementary Table S10. Complete lists of sex-differential genes in GTEx normal thyroid and in TCGA-THCA adjacent-normal tissue.(See Supplementary Excel file.)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablesS1S10.xlsx