




Article

Integrated Evolutionary and Multi-Omic Analysis of STAT Family Activation Across Solid Tumors

Dunja Lukic ¹, Pietro Hiram Guzzi ^{2,3} and Federico Manuel Giorgi ^{1,*}¹ Department of Pharmacy and Biotechnology, University of Bologna, Via Gobetti 85, 40129 Bologna, Italy² Department of Surgical and Medical Sciences, Magna Graecia University of Catanzaro, 88100 Catanzaro, Italy; hguzzi@unicz.it³ Data Analytics Research Center, University of Catanzaro, 88100 Catanzaro, Italy

* Correspondence: federico.giorgi@unibo.it; Tel.: +39-051-209-4521

Abstract

Background/Objectives: The STAT (Signal Transducer and Activator of Transcription) family of seven transcription factors mediates cytokine and growth-factor signaling, regulating proliferation, differentiation, and immunity. While STAT3/STAT5 are established oncogenes and STAT1/STAT2 are classically viewed as tumor suppressors, emerging evidence indicates context-dependent roles in tumorigenesis. This study aimed to integrate evolutionary analysis with bulk transcriptomic, regulon, single-cell, and exploratory chromatin-binding analyses of the STAT family in human solid tumors. **Methods:** Orthologs and paralogs of human STAT genes (81 sequences total) were retrieved across vertebrates and invertebrates; a phylogenetic tree was constructed using MUSCLE alignment and Neighbor-Joining in MEGA12. Differential expression was assessed in TCGA solid tumors versus GTEx normal tissues. Master-regulator activity was inferred using the *corto* algorithm. Single-cell RNA-seq datasets were used to compare malignant and non-malignant cell populations. STAT1 chromatin binding was examined via ChIP-seq in interferon-stimulated HeLa and K562 cells. **Results:** Phylogeny resolved seven conserved vertebrate clades, with endocrine-responsive STAT3/STAT5 showing higher conservation and immune-associated STAT1/STAT2/STAT4/STAT6 exhibiting faster divergence. The majority of STAT genes were frequently upregulated across multiple solid tumors, with activated regulons confirming functional transcriptional engagement. Single-cell analysis demonstrated tumor-cell-autonomous upregulation of STAT1 and STAT2 in the HNSCC dataset. STAT1 ChIP-seq revealed asymmetric forward/reverse-strand read density around peak summits, supporting non-canonical DNA recognition. **Conclusions:** The STAT family operates as an evolutionarily conserved, broadly activated transcriptional module in human solid cancers, combining quantitative upregulation with qualitative shifts in DNA-binding dynamics. These findings refine our understanding of JAK/STAT signaling in oncology and highlight opportunities for network-targeted therapies.



Academic Editor: Alexander Damalas

Received: 6 March 2026

Revised: 3 April 2026

Accepted: 21 April 2026

Published: 3 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).**Keywords:** STAT; pan-cancer; transcriptomics; master regulator analysis

1. Introduction

The Signal Transducer and Activator of Transcription (STAT) family comprises seven distinct genes in the human genome: STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6 [1]. These transcription factors are integral to the transmission of signals from extracellular cytokines and growth factors to the nucleus. They play an essential role in

regulating critical physiological processes, including cellular proliferation, differentiation, apoptosis, and immune system modulation [2]. The canonical activation of STAT proteins is primarily mediated by Janus kinases (JAKs) [3,4]. Upon ligand binding to cell-surface receptors, JAKs auto-phosphorylate and subsequently phosphorylate specific tyrosine residues on the recruited STAT monomers [5,6]. This phosphorylation event triggers STAT dimerization via their highly conserved Src homology 2 (SH2) domains, followed by rapid nuclear translocation and DNA binding to specific regulatory elements [7].

In the context of oncology, the dysregulation of STAT signaling contributes directly to tumorigenesis across multiple cancer types [8]. The molecular mechanisms driving this dysregulation, however, show a stark dichotomy between hematological malignancies and solid tumors. In blood cancers, particularly T-cell large granular lymphocytic (T-LGL) leukemia and natural killer/T-cell lymphomas, somatic gain-of-function mutations in the SH2 domains of STAT3 and STAT5B act as primary oncogenic drivers, occurring in up to 30% to 50% of patients [9,10]. These mutations directly induce constitutive phosphorylation, dimerization, and unconstrained proliferation. Conversely, somatic mutations in STAT genes are exceptionally rare in solid tumors [2]. Instead, solid tumors primarily exploit the constitutive activation and quantitative transcriptomic upregulation of wild-type STAT proteins [11]. STAT3 and STAT5 (comprising STAT5A and STAT5B) are classically defined as oncogenes [12,13]. Elevated expression and hyperactivation of STAT3 have been documented in a diverse array of solid tumors, including gastric cancer, glioblastoma, prostate cancer, and hepatic cancer, where they consistently correlate with poor patient prognosis and decreased overall survival [14]. In breast cancer, STAT3 activity is frequently associated with highly aggressive and dedifferentiated tumor subtypes, while STAT5 activity is more often linked to lower-grade tumors [15,16].

Conversely, STAT1 and STAT2 have traditionally been characterized as tumor suppressors, primarily due to their roles as the main mediators of interferon (IFN) signaling, which typically elicits anti-proliferative and pro-apoptotic responses. However, recent data suggest a context-dependent role. For instance, in advanced prostate cancer, elevated STAT1 expression has been implicated in mediating cellular resistance to both docetaxel chemotherapy and radiation therapy [17]. STAT1 can also act as an oncogene and tumor promoter in specific contexts, such as leukemia and breast cancer, where its upregulation drives cancer cell proliferation, metastasis, and immune evasion [18,19]. This duality highlights the need to evaluate the entire STAT family network simultaneously, rather than studying individual members in isolation.

The oncogenic potential of STAT proteins is largely driven by their capacity to transcriptionally regulate networks of genes essential for cell cycle progression, survival, and immunosuppression. STAT3 directly activates the transcription of Cyclin D1, c-Myc, and anti-apoptotic members of the Bcl-2 family (such as Bcl-xL and Mcl-1), while concomitantly repressing pro-apoptotic factors [20]. In glioblastoma, the aberrant activation of the IL-6/JAK/STAT3 axis not only drives rapid cellular proliferation but also heavily alters the tumor microenvironment, rendering it immunosuppressive and facilitating immune evasion [21]. Furthermore, STAT3 can interact with other oncogenic transcription factors, such as the Nuclear factor erythroid 2-related factor 2 (NRF2), to synergistically promote tumor progression and chemoresistance [22]. In addition to classical transcription activation, unphosphorylated STAT3 (u-STAT3) has also been shown to drive the expression of critical oncogenic and inflammatory mediators, such as IL-6, IL-8, CCL5, and the MET oncogene, further diversifying its pro-tumorigenic repertoire [23].

The complexity of STAT signaling in tumorigenesis has been extensively studied utilizing animal models, spanning from *Drosophila melanogaster* to transgenic mice, highlighting the deep evolutionary roots and conserved nature of these pathways [24]. Indeed, under-

standing the conserved oncogenic functions of STAT genes also requires an evolutionary perspective. The STAT gene family emerged early in metazoan evolution, with functional orthologs identified in invertebrates such as the nematode *Caenorhabditis elegans* [25]. The *Caenorhabditis elegans* STAT ortholog STA-1 lacks the N-terminal oligomerization domain present in vertebrate STATs, suggesting that this structural feature was acquired later in evolution to provide additional regulatory control over spatial arrangement and oligomerization [25]. During early vertebrate evolution, the STAT repertoire expanded rapidly from a single ancestral gene to six members, likely driven by two rounds of whole-genome duplication. The subsequent tandem duplication of STAT5 in early Eutherian mammals gave rise to the modern seven-member family found in humans. Phylogenetic analyses reveal a distinct evolutionary dichotomy: STAT proteins operating downstream of endocrine hormones (STAT3, STAT5A, STAT5B) exhibit higher sequence conservation, whereas those mediating immune signals (STAT1, STAT2, STAT4, STAT6) display a significantly faster evolutionary divergence rate [26].

The functional output of STAT proteins is ultimately governed by their DNA-binding specificities. Activated STAT dimers recognize and bind to conserved DNA consensus sequences, such as Gamma-activated sequence (GAS) elements and Interferon-stimulated response elements (ISRE) [7,27]. The advent of chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) has provided high-resolution, genome-wide maps of STAT binding [28,29]. For example, ChIP-seq profiling of STAT1 in interferon-stimulated HeLa cells identified tens of thousands of putative binding regions, revealing a massive scale of genomic interaction [30]. Surprisingly, transcriptome profiling coupled with these ChIP-seq datasets demonstrated that only a fraction of the genes harboring STAT1 binding sites are actually transcriptionally modulated, hinting at the complexity of STAT-mediated gene regulatory networks [31]. Furthermore, structural and biochemical studies indicate that STAT proteins can bind non-canonical or half-site elements under some conditions [27]. These observations suggest additional complexity in STAT DNA recognition beyond canonical GAS binding [32].

Despite extensive work on individual STAT members [1,4,33], an integrated family-level analysis spanning evolution, bulk transcriptomics, regulon activity, and single-cell expression in solid tumors remains limited. Here, we combine phylogenetic reconstruction, bulk differential expression, master-regulator analysis, and single-cell RNA-seq to characterize STAT family behavior across human solid tumors. We also include an exploratory STAT1 ChIP-seq analysis to examine whether peak-centered read distributions show reproducible asymmetry.

2. Materials and Methods

2.1. Sequence Retrieval and Phylogenetic Analysis

We retrieved orthologs and paralogs of human STAT genes as follows. All human gene symbols in the manuscript and in the pipeline are defined by HGNC (HUGO Gene Nomenclature Committee) [34], while non-human ortholog names follow the nomenclature adopted by the NCBI (National Center for Biotechnology Information) database. First, we included in the list of human STATs the following genes: STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, STAT6. Then, we defined orthologs using the OrthoDB database queried from the HGNC HCOP Orthology Prediction web interface [35]. We searched for orthologs, co-orthologs, and paralogs in the following species: Human, Chimp, Macaque, Mouse, Rat, Dog, Cat, Horse, Cattle, Pig, Opossum, Platypus, Chicken, Anole lizard, *Xenopus tropicalis*, Zebrafish, Fruit fly, and *Caenorhabditis elegans*. We found 74 genes that were in an orthologous or co-orthologous relationship with the human STAT genes, so together with the 7 human STATs, our analysis defined 81 genes. For each gene, we retrieved a

protein isoform as follows: when available, a curated RefSeq was selected (with “NP” identifier), and when multiple sequences were available, the longest was chosen. All sequence accession numbers are reported in the analysis. The retrieval was performed using R scripts and the *rentrez* package within the R statistical environment [36].

Protein sequences were aligned using the MUSCLE algorithm [37] with default parameters, including a gap open penalty of -2.90 , a gap extend penalty of 0.00 , and a hydrophobicity multiplier of 1.20 . Sequence clustering during the alignment phase was performed using the UPGMA method. Phylogenetic reconstruction was subsequently carried out in MEGA12 utilizing the Neighbor-Joining (NJ) method [38]. Evolutionary distances were computed using the Poisson substitution model, assuming uniform rates among sites and applying pairwise deletion for gaps and missing data. Topological support was evaluated using a bootstrap test with 1000 replicates.

2.2. RNA-Seq Analysis

To investigate the transcriptomic landscape of STAT genes across solid tumors, we utilized harmonized bulk RNA sequencing data from The Cancer Genome Atlas (TCGA) [39] and the Genotype-Tissue Expression (GTEx) project [40]. To account for study-specific biases and batch effects between these two repositories, the dataset was preprocessed using the Wang normalization pipeline [41]: this harmonization reduces cross-cohort technical bias but does not eliminate differences related to tissue composition, tumor purity, or residual batch effects. We restricted our analysis to 15 tumor types that possessed matched normal tissue data from GTEx: BLCA (Bladder Urothelial Carcinoma), BRCA (Breast Invasive Carcinoma), COAD (Colon Adenocarcinoma), ESCA (Esophageal Carcinoma), KIRC (Kidney Renal Clear Cell Carcinoma), KIRP (Kidney Renal Papillary Cell Carcinoma), LIHC (Liver Hepatocellular Carcinoma), LUAD (Lung Adenocarcinoma), LUSC (Lung Squamous Cell Carcinoma), PRAD (Prostate Adenocarcinoma), READ (Rectum Adenocarcinoma), STAD (Stomach Adenocarcinoma), THCA (Thyroid Carcinoma), UCEC (Uterine Corpus Endometrial Carcinoma), and UCS (Uterine Carcinosarcoma). Statistical analyses were conducted within the R environment (version 4.4.2). Differential gene expression contrasting tumor versus normal tissues was evaluated using the DESeq2 package (version 1.48.1) [42]. To assess the functional activation of STAT-driven transcriptional networks, we applied the *corto* algorithm (version 1.2.4) [43]. Regulatory networks were inferred using 200 bootstraps. Edge significance was set to a p -value of 10^{-10} for most datasets, with the exception of the UCS cohort; due to its substantially lower sample size ($n = 47$), a threshold of $p = 0.01$ was utilized to ensure the resulting networks were comparable in scale to the other cohorts as in [44]. All p -values were corrected for multiple testing using the Benjamini–Hochberg procedure. Only STAT genes with a network of at least 10 inferred targets are shown in the figures.

2.3. Single-Cell RNA-Seq Analysis

To examine whether bulk-expression patterns were also detectable at single-cell resolution, we interrogated the head and neck squamous cell carcinoma (HNSCC) dataset GSE103322 [45]. This specific dataset was selected due to its high-resolution profiling of the tumor microenvironment (TME) and robust pre-existing annotations, which reliably distinguish malignant epithelial cells from surrounding stromal and immune populations. The data, comprising 21,883 genes across 5902 individual cells, was queried and retrieved using the TMEexplorer R package v1.18.0 [46], chosen specifically for its efficiency in accessing and formatting standardized single-cell TME datasets. Downstream processing was performed utilizing the Seurat package v5.4.0 [47]. Cells were grouped into “Cancer” and “Non-Cancer” compartments according to the original dataset annotation. The expression

matrix underwent log-normalization, followed by the identification of 2000 highly variable features. Data scaling and Principal Component Analysis (PCA) were performed, and the first 15 principal components were utilized to generate Uniform Manifold Approximation and Projection (UMAP) embeddings for visualization. Statistical comparisons of STAT expression between the Cancer and Non-Cancer cell populations were conducted using a one-sided Wilcoxon rank-sum test, with p -values subsequently adjusted via the Benjamini–Hochberg method. To validate our findings across diverse tumor microenvironments, three additional single-cell RNA-seq datasets were acquired as Gene Expression Omnibus Series (GSE): GSE72056 (melanoma) [48], GSE75688 (breast cancer) [49], and GSE150430 (nasopharyngeal carcinoma) [50]. These datasets were processed, normalized, and analyzed utilizing the same Seurat-based pipeline described above to quantify STAT family expression in malignant versus non-malignant single cells.

2.4. ChIP-Seq Analysis

To assess the genomic binding dynamics of STAT1, we analyzed chromatin immunoprecipitation sequencing (ChIP-seq) data derived from interferon-stimulated HeLa S3 cells (GEO accession: GSE12782 [51]). To assess whether similar read-distribution patterns were also visible in a second dataset from a different cellular context, we additionally processed STAT1 ChIP-seq data from the K562 human chronic myelogenous leukemia cell line (GEO accession: GSE31477; sample GSM935472 [52]). Both ChIP-seq experiments utilized a total STAT1 antibody. Raw FASTQ files were aligned to the human reference genome using HISAT2 [53], and the resulting alignments were sorted and indexed utilizing SAMtools [54]. Peak calling was executed with MACS3 [55] using a q -value threshold of 0.05, with summit identification enabled, the building model bypassed, and an extension size of 200 bp. Genomic loci of interest, such as the IRF9, STAT2, and ICAM1 promoters, were visualized using the Integrative Genomics Viewer (IGV) [56], with the “color by read strand” option. To evaluate the strand-specific read distribution and investigate potential spatial binding asymmetries of the STAT1 dimer at target sites, we extracted sequence alignment information directly from the BAM files within the R environment. We calculated and plotted the density of forward- and reverse-mapped reads relative to the center of the identified MACS3 peak summits, generating an aggregated profile globally across the entire experiment.

3. Results

3.1. Evolutionary Expansion and Lineage-Specific Diversification of the STAT Family

The phylogenetic reconstruction of 81 STAT-related protein sequences derived via OrthoDB across vertebrate and invertebrate species resolved seven major clades corresponding to human STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6 (Figure 1). Invertebrate STAT orthologs from *Drosophila melanogaster* (Stat92E) and *Caenorhabditis elegans* (sta-1, sta-2) clearly form an external branch, supporting their role as ancestral outgroups and confirming the monophyletic expansion of vertebrate STATs from a single ancestral gene [24,25,57].

Within vertebrates, each human STAT gene clustered with its corresponding orthologs across mammals, birds, reptiles, amphibians, and fish, supported by robust bootstrap values at the principal nodes. STAT3, STAT5A, and STAT5B form deeply conserved branches with short interspecies branch lengths, consistent with strong purifying selection acting on endocrine and growth-factor-responsive STATs [26,58].



Figure 1. Phylogenetic tree of STAT proteins inferred from 81 protein sequences (derived via OrthoDB) spanning vertebrates and selected invertebrates. The tree was constructed using Neighbor-Joining with Poisson-corrected distances and 1000 bootstrap replicates. Distinct clades corresponding to STAT1 (red), STAT2 (dark blue), STAT3 (green), STAT4 (orange), STAT5A (cyan), STAT5B (purple), STAT6 (magenta). Invertebrate sequences from *Drosophila melanogaster* (Stat92E) and *Caenorhabditis elegans* (sta-1 and sta-2), highlighted in brown and tan color, respectively, belong to distinct outgroups, supporting a single ancestral STAT origin followed by vertebrate expansion. Bootstrap support values are indicated at internal nodes. The scale bar represents the number of amino acid substitutions per site. Lineage-specific duplications (e.g., teleost stat1a/stat1b and marsupial STAT5A-like sequences) are visible within respective clades.

In contrast, immune-associated STATs (STAT1, STAT2, STAT4, STAT6) display longer branch lengths and greater divergence across vertebrate clades, suggesting accelerated evolutionary dynamics, likely reflecting adaptation to pathogen-driven immune pressures [24,26]. This dichotomy between endocrine and immune STAT subclasses is clearly visible in the tree topology (Figure 1).

The tree reveals several lineage-specific duplication and loss events. Teleost *Danio rerio* contains two STAT1 paralogs (stat1a and stat1b), clustering within the STAT1 clade but forming distinct sub-branches, consistent with teleost-specific whole-genome duplication events [59]. Similarly, the marsupial *Monodelphis domestica* exhibits two STAT5A-like sequences (canonical STAT5A and LOC100011470), indicative of a lineage-specific duplication. It is important to note that while we deliberately retained the official NCBI nomenclature for all analyzed sequences to ensure database consistency, our phylogenetic topology strongly supports the biological reclassification of several uncharacterized loci. For instance, the clustering clearly identifies LOC100566752 in *Anolis carolinensis* and LOC100011470 in *Monodelphis domestica* as STAT5 genes. An evaluation of the genomic synteny of these loci on Chromosome 2 (NC_077228.1) of *Monodelphis domestica* confirms that the canonical STAT5A and LOC100011470 sequences are positioned in direct tandem proximity to each other (Supplementary Figure S1). This finding reinforces that this is a lineage-specific localized duplication event distinct from the broader eutherian genomic expansion.

Together, these findings reinforce a model in which early vertebrate whole-genome duplications expanded the STAT repertoire from a single ancestral gene to six members, followed by tandem duplication of STAT5 in eutherian mammals and additional lineage-restricted events. The preserved orthology across mammals, combined with selective paralog expansions in fish and marsupials, underscores both functional conservation and adaptive diversification within the STAT family.

3.2. Pan-Cancer Upregulation of STAT Transcription Factors

We next examined whether the evolutionary conservation of STAT genes is paralleled by recurrent expression changes in solid tumors. Using harmonized bulk RNA-seq datasets from TCGA tumors and GTEx normal tissues [39–41], we performed differential expression analysis across multiple solid cancer types (Figures 2 and 3).

Across the majority of tumor cohorts analyzed—including BLCA, BRCA, COAD, ESCA, KIRC, LUAD, LUSC, STAD, and others—STAT family members exhibited consistent upregulation in tumors relative to corresponding normal tissues. While STAT3 and STAT5 genes have historically been considered canonical oncogenes [12], our pan-cancer analysis demonstrates that the upregulation phenomenon extends broadly to STAT1, STAT2, STAT4, and STAT6.

STAT1 and STAT2, traditionally associated with interferon-mediated tumor suppression [1], were frequently elevated in tumor tissues. This pattern supports a context-dependent functional shift, where chronic interferon signaling may contribute to tumor adaptation, immune modulation, or therapy resistance rather than pure tumor suppression [18]. STAT6 and STAT4, typically associated with Th2 and immune polarization pathways [60], also showed elevated expression in multiple solid tumors, suggesting a broader rewiring of cytokine-responsive transcriptional programs in cancer.

The directionality of differential expression was consistent across tumor types, pointing to a systemic activation of STAT-mediated transcription rather than isolated tumor-specific events.

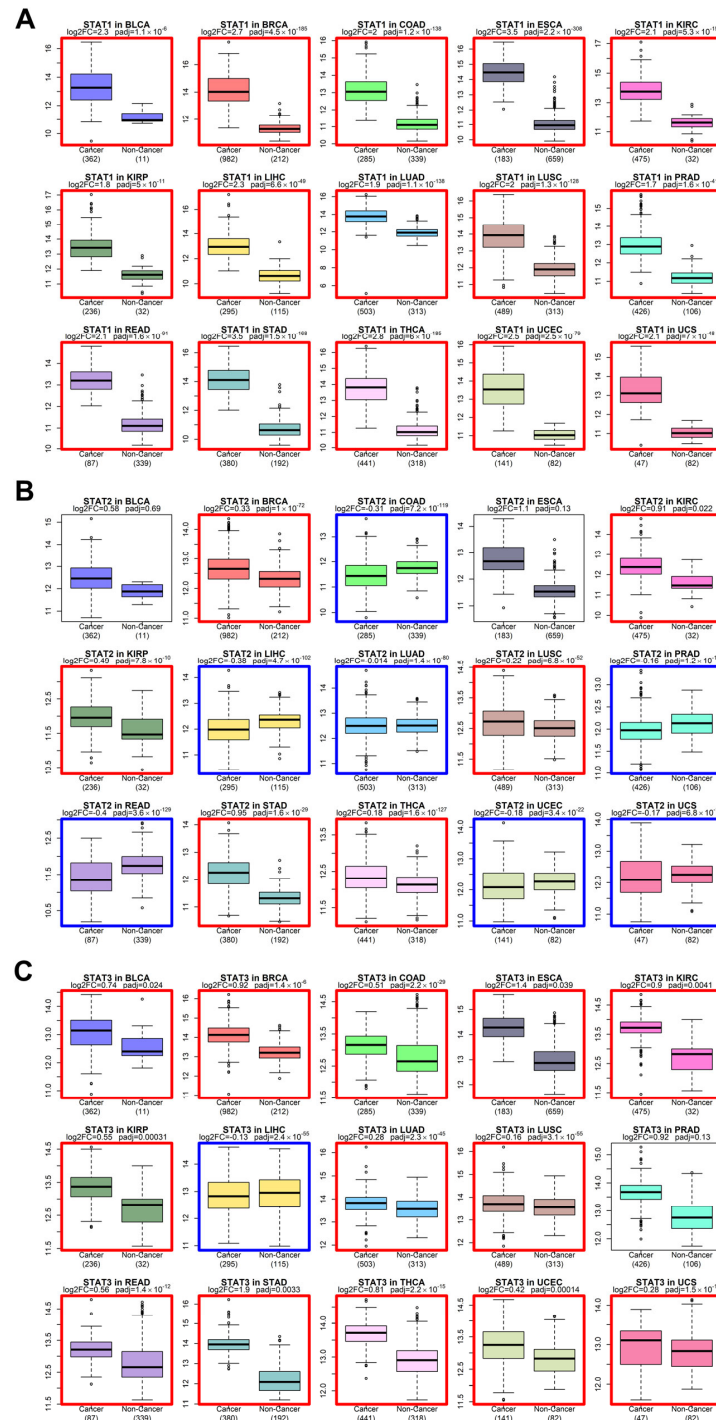


Figure 2. Differential expression analysis of STAT genes comparing tumor (TCGA) and corresponding normal tissue (GTEx) samples across multiple cancer types. Expression values represent normalized log-transformed RNA-seq counts. (A) STAT1 expression across tumor types. (B) STAT2 expression. (C) STAT3 expression. Each panel contains multiple boxplots corresponding to individual cancer types (BLCA, BRCA, COAD, ESCA, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, READ, STAD, THCA, UCEC, UCS). For each cancer type, the left box represents tumor samples (Cancer) and the right box represents non-cancer tissues (Non-Cancer). Sample sizes are indicated below each comparison. Center lines represent medians; boxes indicate interquartile ranges; whiskers denote distribution spread; outlier dots represent individual samples. Adjusted *p*-values are displayed above each comparison. Across most tumor types, STAT family members exhibit consistent upregulation in cancer relative to normal tissues. A colored box surrounds each comparison when a significant change (adjusted *p*-value < 0.05) occurs: red when the gene is upregulated in cancer, and blue when it is downregulated.

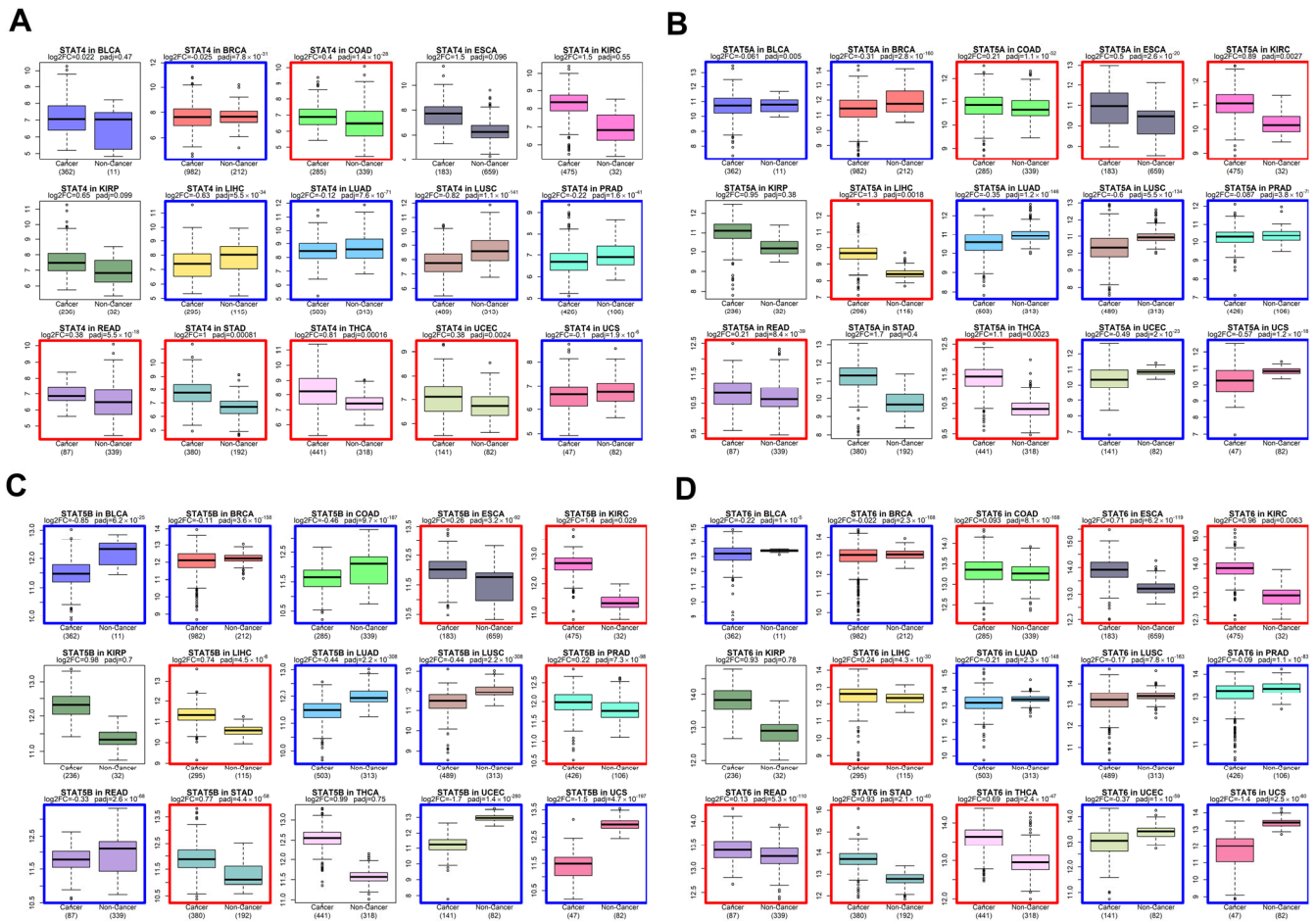


Figure 3. Differential expression analysis of STAT genes comparing tumor (TCGA) and corresponding normal tissue (GTEx) samples across multiple cancer types. Expression values represent normalized log-transformed RNA-seq counts. (A) STAT4 expression. (B) STAT5A expression. (C) STAT5B expression. (D) STAT6 expression. Each panel contains multiple boxplots corresponding to individual cancer types (BLCA, BRCA, COAD, ESCA, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, READ, STAD, THCA, UCEC, UCS). For each cancer type, the left box represents tumor samples (Cancer) and the right box represents non-cancer tissues (Non-Cancer). Sample sizes are indicated below each comparison. Center lines represent medians; boxes indicate interquartile ranges; whiskers denote distribution spread; outlier dots represent individual samples. Adjusted *p*-values are displayed above each comparison. Across most tumor types, STAT family members exhibit consistent upregulation in cancer relative to normal tissues. A colored box surrounds each comparison when a significant change (adjusted *p*-value < 0.05) occurs: red when the gene is upregulated in cancer, and blue when it is downregulated.

3.3. STAT Transcriptional Networks Are Functionally Activated in Cancer

Gene expression upregulation does not necessarily equate to transcriptional activity. Therefore, we evaluated STAT functional activity using a master regulator analysis implemented with the *corto* algorithm [43]. *Corto* infers regulatory networks based on co-expression and data processing inequality principles, enabling estimation of transcription factor activity independently of its mRNA abundance.

Across solid tumor types, STAT-driven regulons showed significant positive enrichment scores (NES), indicating coordinated upregulation of downstream targets (Figures 4 and 5). This observation demonstrates that STATs are not merely transcriptionally elevated but are acting as active master regulators within oncogenic gene networks.

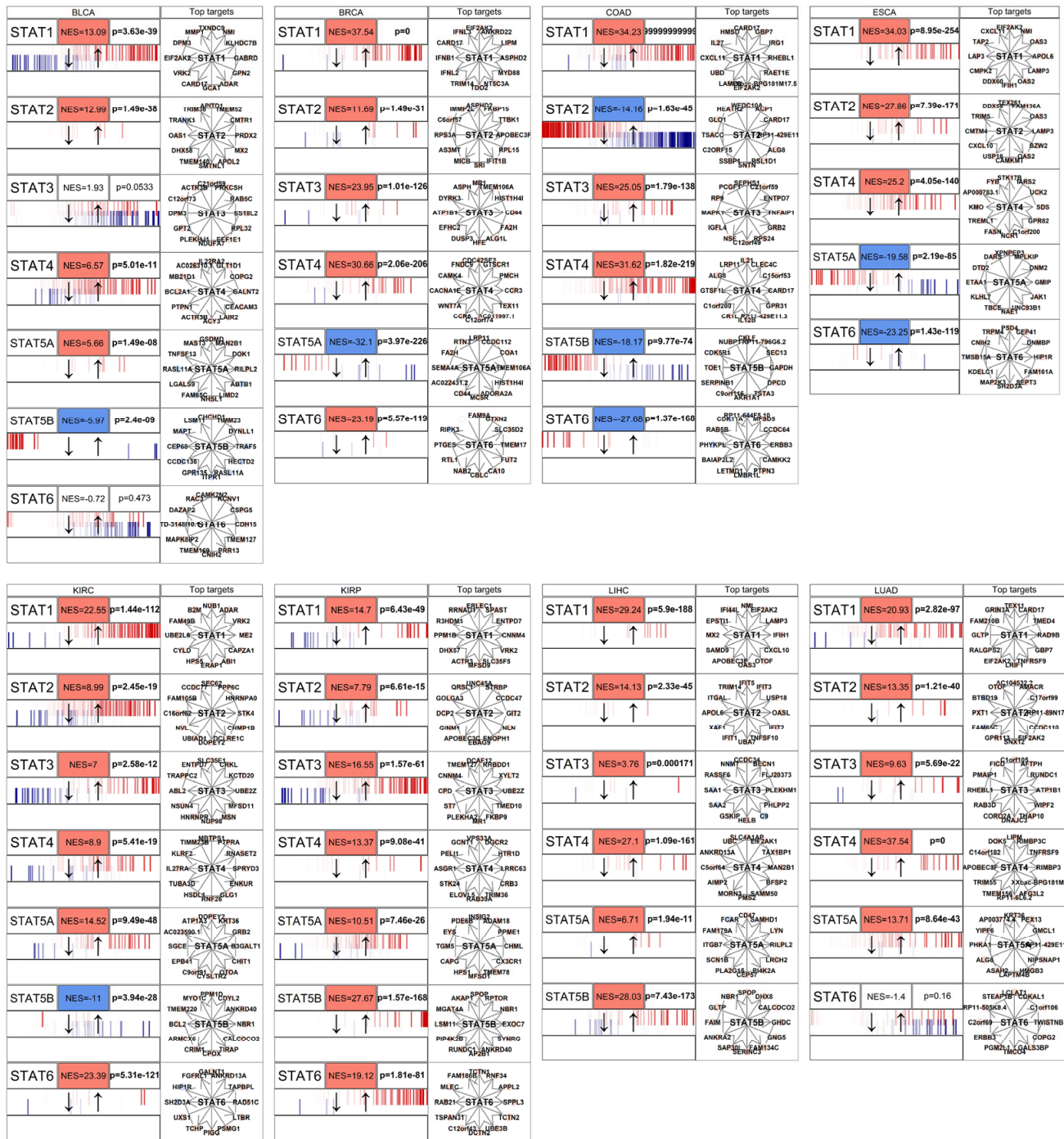


Figure 4. Corto-based master regulator analysis comparing STAT regulon activity between cancer and non-cancer samples across multiple tumor cohorts. For each tumor type (organized in grid layout), normalized enrichment scores (NES) and corresponding *p*-values are shown for individual STAT family members. The black arrows simply indicate the direction of the horizontal box (left genes are more downregulated, right genes are more upregulated). Red color gradients indicate positive enrichment (regulon activation in cancer), whereas blue gradients indicate negative enrichment (reduced activity in cancer). Small barcodes represent distribution of target gene differential expression across the ranked gene list. Network diagrams illustrate inferred STAT-centered regulatory networks, with edges connecting STAT factors to predicted target genes. Plots are organized alphabetically and show results for 8 tumors (BLCA, BRCA, COAD, ESCA, KIRC, KIRP, LIHC, LUAD). Across many tumor types, STAT regulons show coordinated activation, confirming functional transcriptional engagement beyond simple mRNA upregulation.

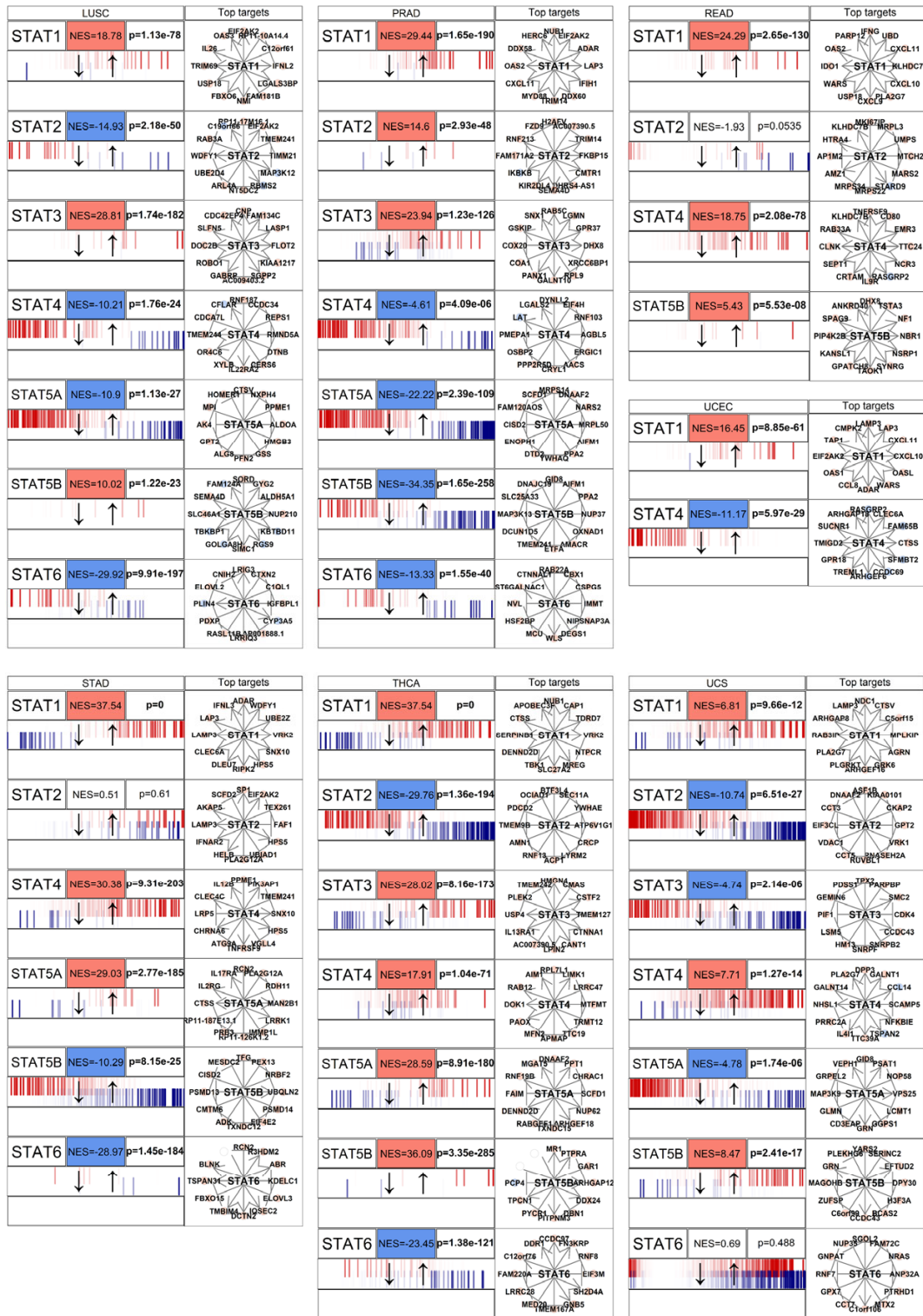


Figure 5. Corto-based master regulator analysis comparing STAT regulon activity between cancer and non-cancer samples across multiple tumor cohorts. For each tumor type (organized in grid layout), normalized enrichment scores (NES) and corresponding p-values are shown for individual STAT family members. The black arrows simply indicate the direction of the horizontal box (left genes are more downregulated, right genes are more upregulated). Red color gradients indicate positive enrichment (regulon activation in cancer), whereas blue gradients indicate negative enrichment (reduced activity in cancer). Small barcodes represent distribution of target gene differential expression across the ranked gene list. Network diagrams illustrate inferred STAT-centered regulatory networks, with edges connecting STAT factors to predicted target genes. Plots show results for 7 tumor types (LUSC, PRAD, READ, STAD, THCA, UCEC, UCS).

In multiple tumor contexts, STAT1, STAT3, and STAT6 exhibited particularly strong regulon activation, consistent with their known roles in regulating genes involved in cell cycle progression, inflammatory signaling, and immune evasion [33]. The concordance between mRNA upregulation and regulon activation supports a model of sustained pathway activation rather than passive overexpression.

Notably, STAT network activation was often broader than individual gene upregulation, suggesting cooperative or compensatory behavior among STAT paralogs. This reinforces the concept that STATs function as an interconnected transcriptional module rather than independent actors.

3.4. Single-Cell Validation of STAT1, STAT2, and STAT6 Upregulation in Malignant Cells

To determine whether the STAT upregulation observed in bulk RNA-seq data reflects tumor-intrinsic transcriptional activation rather than stromal or immune infiltration, we analyzed single-cell RNA-seq data from head and neck squamous cell carcinoma [45]. Cells were classified into “Cancer” and “Non-Cancer” groups based on the original annotation (see Methods), allowing us to compare malignant epithelial cells against all remaining cell populations (Figure 6).

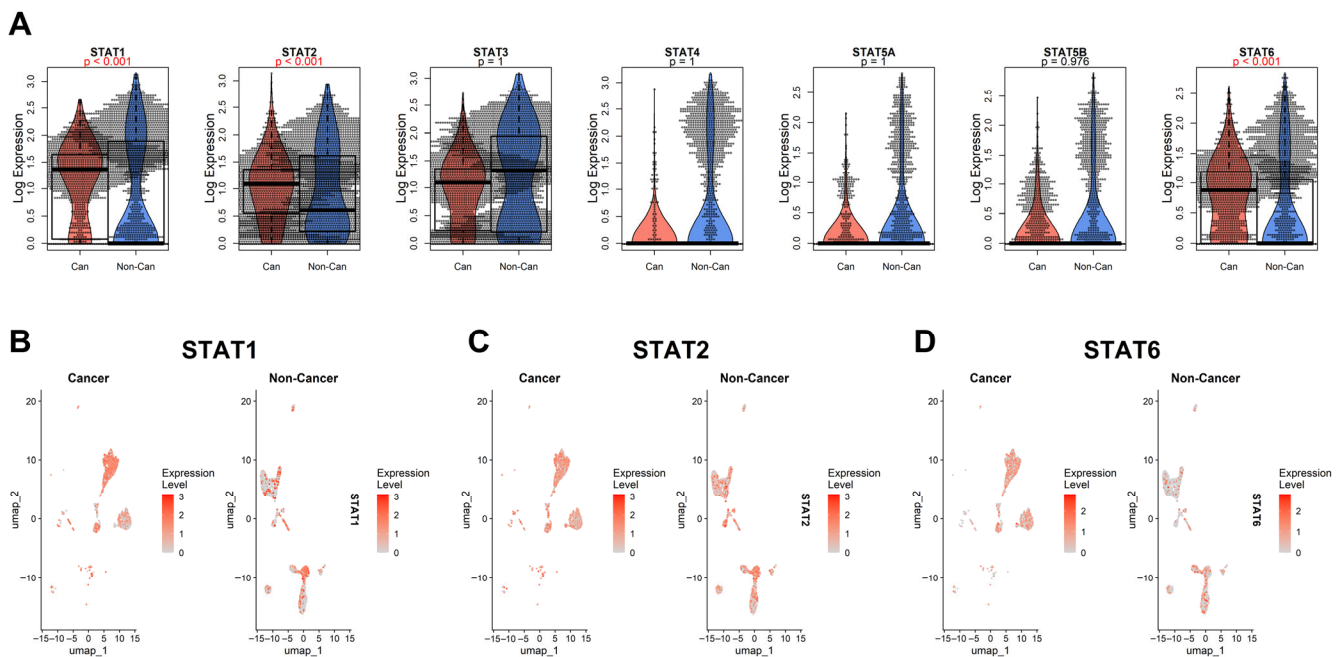


Figure 6. Single-cell RNA-seq analysis derived from the Puram dataset [45], comparing malignant epithelial cells (“Cancer”) and remaining cell populations (“Non-Cancer”). (A) Beeswarm/violin/boxplot representation of log-transformed expression levels for STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6. For each gene, cancer cells (red) are shown on the left and non-cancer cells (blue) on the right. Violin shapes depict distribution density; overlaid boxplots indicate median and interquartile range; dots represent individual cells. Adjusted *p*-values are displayed above each comparison (in red if significant). STAT1, STAT2, and STAT6 show significant global upregulation in cancer. (B) UMAP projection of single cells colored by STAT1 expression level in Cancer (left) and Non-Cancer (right) compartments. (C) UMAP projection colored by STAT2 expression. (D) UMAP projection colored by STAT6 expression. Color intensity reflects normalized expression levels. STAT1 and STAT2 show enrichment within malignant epithelial clusters, whereas STAT6 displays a more heterogeneous pattern with high expression in a minor non-cancer subpopulation, suggesting microenvironmental contribution.

At single-cell resolution, STAT1 and STAT2 showed clear and consistent upregulation within malignant cell clusters compared with the aggregated non-cancer compart-

ment. UMAP projections confirmed that their expression is enriched directly within tumor epithelial populations rather than being confined to infiltrating immune cells. These findings reinforce the notion that STAT1 and STAT2 activation is, at least in part, tumor-cell autonomous.

STAT6 displayed a more complex pattern. While global comparisons indicate significant upregulation in the cancer group, the beeswarm/violin/boxplot representation reveals that STAT6 expression is particularly elevated in a distinct minor subpopulation of non-cancer cells. This suggests that STAT6 activity may be enriched in specific stromal or immune subsets within the tumor microenvironment, potentially reflecting polarized immune states or cytokine-responsive niches.

It is worth mentioning that the “Non-Cancer” category in this dataset represents a heterogeneous mixture of cell types, including immune, stromal, endothelial, and possibly residual epithelial populations. It does not necessarily correspond to the precise normal cell of origin of the tumor, which remains undefined in this context. Therefore, while STAT6 appears significantly upregulated overall, its cellular specificity cannot be conclusively attributed solely to malignant epithelial cells in this dataset. Moreover, other foundational single-cell cancer datasets [48–50] did not reproduce this pattern; across those datasets, most STAT genes showed no consistent difference between malignant and non-malignant cells, with the main exception of STAT3 in nasopharyngeal carcinoma (Supplementary Figure S2), highlighting the need for better annotation and standardization for future single-cell datasets.

Future studies employing refined cell-type annotations and lineage-resolved comparisons (ideally contrasting tumor cells with their exact normal cellular counterpart) will be required to more precisely delineate STAT6’s tumor-intrinsic versus microenvironment-driven expression.

In the HNSCC [45] dataset, STAT1 and STAT2 were higher in malignant cells, whereas STAT6 showed a mixed pattern influenced by the non-malignant compartment.

3.5. Exploratory Analysis of Strand-Skewed STAT1 ChIP-Seq Read Distributions

Finally, we investigated STAT1 chromatin binding behavior by analyzing genome-wide ChIP-seq data derived from both interferon-stimulated HeLa cells [51] and K562 cells [52]. Visual inspection of canonical STAT1 target loci, including IRF9 and STAT2, which drive a positive feedback loop to amplify interferon signaling [61], and ICAM1, a key mediator of cell adhesion and immune infiltration [62], revealed a pronounced asymmetry between forward and reverse strand read distributions (Figure 7A–C). Rather than forming a perfectly centered pattern around the peak summit, reads showed strand-skewed distributions at these loci, indicating directional occupancy for STAT1. To evaluate this distribution globally, we calculated aggregate read density plots centered on all identified peak summits across both datasets. The resulting combined profile (Figure 7D) displays a clear spatial asymmetry. While this read distribution fundamentally reflects the standard bimodal strand shift inherent to single-end ChIP-seq fragment sequencing, it may also concurrently represent an asymmetric spatial occupancy of the transcription factor. STAT1 requires dimerization to effectively engage DNA [63]. Our data leaves open the possibility that the STAT1 dimer interacts with the Watson and Crick DNA fragments in an orientation-biased manner, further supporting prior structural evidence of half-site or non-palindromic binding [64,65]. Such asymmetric DNA recognition may expand the STAT1 regulatory repertoire, enabling interaction with composite regulatory elements or cooperative binding with additional transcription factors. This mechanistic flexibility may help explain why only a subset of STAT-bound genes undergo transcriptional modulation despite widespread chromatin occupancy [66–68]. Further studies should provide a clearer complement of

epigenomic and transcriptomic data, asking whether oncogenic STAT activity may involve both quantitative upregulation and qualitative shifts in DNA-binding dynamics.

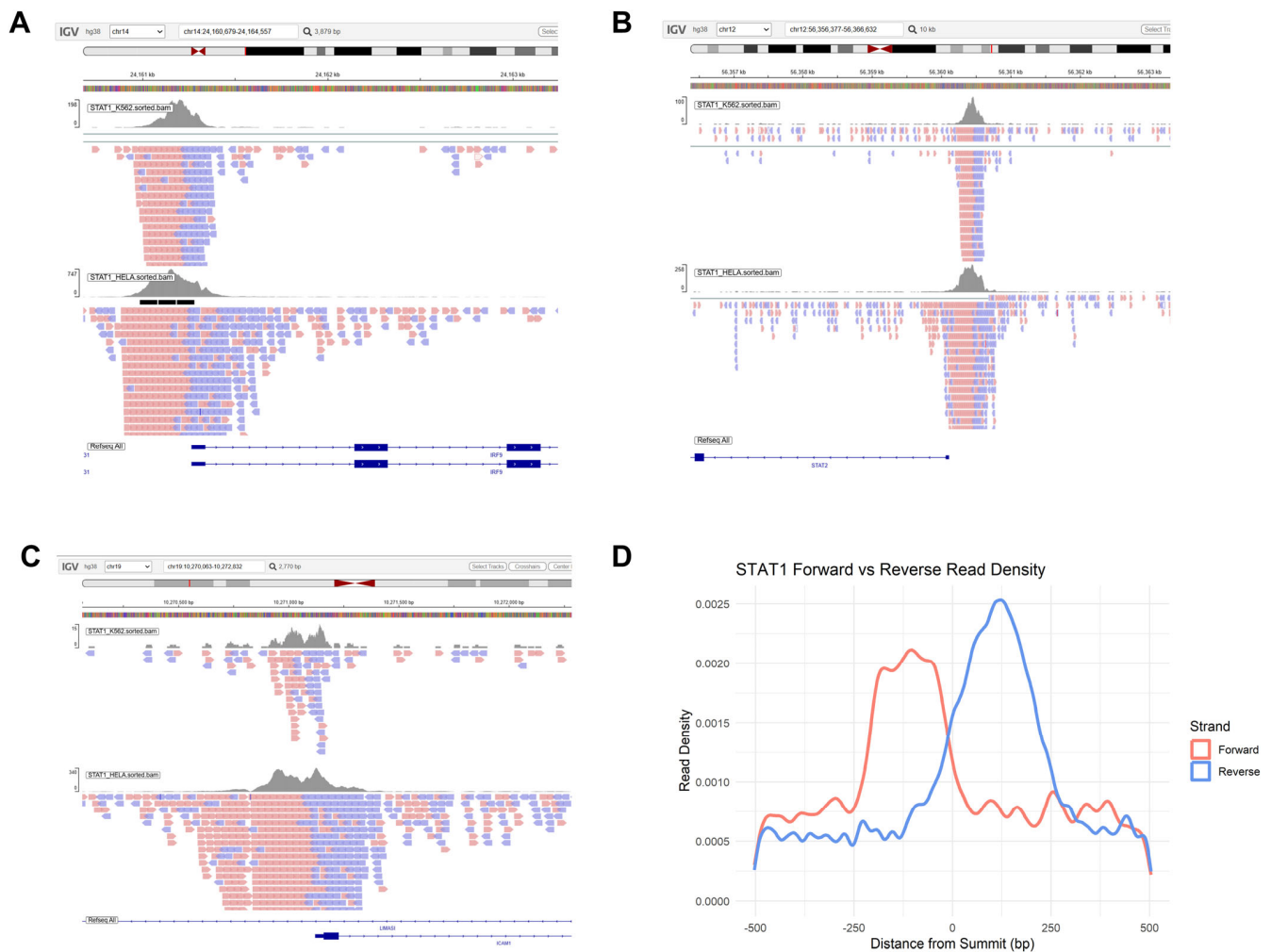


Figure 7. Genome-wide STAT1 binding profiles derived from interferon-stimulated HeLa cells [51] and K562 cells [52]. (A) IGV snapshot of a representative STAT1 binding locus at the IRF9 promoter. The upper tracks show read coverage density for K562 and HeLa cells. Individual reads are displayed below, with forward-strand reads in red and reverse-strand reads in blue, demonstrating strand-specific asymmetry relative to the peak summit. (B) IGV snapshot of a second representative locus at the STAT2 promoter, again demonstrating strand-biased distribution of reads around the binding peak. (C) IGV snapshot at the ICAM1 promoter locus, confirming identical non-canonical binding dynamics. (D) Aggregate metaprofile of STAT1 ChIP-seq signal centered on peak summits (± 500 bp), combining both datasets. Forward (red) and reverse (blue) strand read densities are plotted separately. The curves display asymmetry, with offset enrichment between strands rather than symmetric accumulation around the summit. This pattern leaves open the potential for non-canonical, orientation-biased DNA recognition by STAT1.

4. Discussion

In this study, we provide an integrated evolutionary, transcriptomic, network-level, single-cell, and chromatin-binding analysis of the STAT gene family across vertebrates and human solid tumors. Rather than focusing on individual STAT members, our results support a model in which the STAT family operates as an evolutionarily conserved but functionally plastic transcriptional module broadly activated in cancer.

Our phylogenetic reconstruction confirms the monophyletic expansion of vertebrate STATs, and the resulting topology structurally reinforces a functional dichotomy:

STAT3 and STAT5 paralogs display high conservation reflecting strict endocrine pathway constraints, whereas immune-associated STATs show greater evolutionary divergence [24,26,57]. These evolutionary dynamics mirror the functional specialization observed in mammals. Specifically, the deep conservation of STAT3 and STAT5 highlights their fundamental roles in basic cellular proliferation and survival pathways that are universally hijacked across the 15 solid tumor types analyzed here (such as breast, prostate, and lung carcinomas) to drive autonomous oncogenic growth [12,16]. Conversely, the accelerated evolutionary divergence of the immune-associated STATs (STAT1, STAT2, STAT4, STAT6) generated a highly plastic signaling network originally shaped by host-pathogen arms races [69]. In cancers characterized by complex inflammatory microenvironments, such as head and neck squamous cell carcinoma (HNSCC) and lung adenocarcinoma, this evolutionary plasticity is likely co-opted by malignant cells. Rather than acting strictly as static tumor suppressors, the rapid adaptive potential of these immune STATs enables tumors to dynamically rewire their cytokine responses, adapt to chronic inflammation, and facilitate immune evasion [70,71].

At the transcriptomic level, our pan-cancer analysis reveals widespread upregulation of STAT genes across solid tumors. While STAT3 and STAT5 are established oncogenic drivers [13], our data extend this pattern to STAT1, STAT2, STAT4, and STAT6. The elevation of STAT1 and STAT2 is particularly notable, given their classical characterization as interferon-mediated tumor suppressors [70,72]. Increasing evidence indicates that chronic interferon signaling may promote tumor adaptation, therapy resistance, and immune modulation in advanced cancers [71]. This immune modulation is further exemplified by the role of STAT1 and STAT3 as key transcriptional regulators of CD274 (PD-L1), a primary driver of adaptive immune evasion and a critical target for current checkpoint inhibitor therapies [73]. The consistency of STAT upregulation across tumor types suggests that persistent cytokine signaling and inflammatory circuits constitute a systemic feature of malignant transcriptional reprogramming [14].

Network inference using *corto* [43] demonstrates that STAT regulons are broadly activated in tumors, indicating functional transcriptional engagement rather than passive overexpression. Because transcription factor activity depends on post-translational regulation and chromatin context [74,75], the coordinated enrichment of downstream targets confirms that STAT proteins act as active master regulators in cancer. In several tumor types, regulon activation exceeded the magnitude of individual gene upregulation, supporting a cooperative model in which STAT paralogs function as an interconnected module. Given their ability to form heterodimers and bind overlapping motifs [76,77], oncogenic STAT signaling likely emerges from combinatorial network behavior rather than single-gene dominance. The pharmacological relevance of this network is underscored by the clinical success of Janus kinase (JAK) inhibitors, as well as the advanced clinical development of direct small-molecule inhibitors and targeted degraders (PROTACs) specifically directed against STAT3 and STAT5 [78].

Recent comprehensive pan-cancer analyses have mapped the broad clinical and genomic landscape of the STAT family. For instance, a multi-omic profiling detailed the widespread genomic alterations, epigenetic modifications, and immunological correlates of the JAK-STAT pathway across numerous cancer types [79]. Similarly, Zhou et al. analyzed bulk transcriptomic data to establish the prognostic and immunological value of individual STAT genes, correlating their expression profiles with patient survival and tumor microenvironment infiltration [80]. While these foundational studies firmly established the clinical and biomarker relevance of STATs, they primarily relied on bulk expression correlations and broad genomic mapping. Our study builds upon these macroscopic findings by addressing how this widespread upregulation is functionally executed at the molecular level

and separates cancer cells from paired physiological tissue. By applying master regulator analysis, we show that elevated STAT expression is accompanied by the activation of STAT downstream gene networks rather than existing as inert overexpression.

Single-cell RNA-seq analysis further refines this interpretation. STAT1 and STAT2 show clear enrichment within malignant epithelial populations, supporting a tumor-cell-intrinsic component of interferon pathway activation. STAT6, in contrast, displays a more complex pattern: although globally elevated in cancer, it is highly expressed in a minor subset of non-cancer cells, likely representing specific immune or stromal populations. This highlights both tumor-intrinsic and microenvironmental contributions to STAT signaling. The binary “Cancer vs. Non-Cancer” annotation available in current datasets aggregates heterogeneous cell types and does not necessarily represent the exact normal cell of origin. Future lineage-resolved single-cell analyses comparing tumor cells with their precise physiological counterparts will be necessary to fully disentangle tumor-driven from reactive STAT activation. More broadly, these results show the value of integrating pan-cancer cohorts with high-resolution single-cell data to achieve mechanistic clarity.

Finally, our ChIP-seq analysis reveals asymmetric STAT1 read distributions around peak summits, deviating from a perfectly centered model. This profile is likely driven by a combination of the standard technical bimodal shift of single-end sequencing and potential biological asymmetry. STAT proteins obligately dimerize to engage DNA [63], and structural and biochemical studies have shown that these dimers can bind half-sites or non-palindromic elements, particularly in alternative conformational states [4,65]. Thus, the observed asymmetry may partly stem from orientation-biased binding of the STAT dimer to the Watson and Crick strands. Asymmetric or context-dependent DNA recognition may help explain why only a fraction of STAT-bound loci undergo transcriptional modulation [28,81]. In cancer, such binding flexibility could expand the regulatory repertoire of STAT proteins beyond canonical interferon-stimulated genes, contributing to transcriptional plasticity within altered chromatin landscapes.

Our findings suggest that the STAT family represents an evolutionarily conserved transcriptional module that becomes broadly and functionally activated in cancer. Rather than acting solely as isolated oncogenes or tumor suppressors, STAT proteins operate as a coordinated regulatory network shaped by deep evolutionary constraints and tumor-specific signaling environments. Despite these findings, significant gaps in our understanding of STAT biology remain. While our ChIP-seq analysis suggests non-canonical, asymmetric DNA recognition dynamics, it is still unclear how this structural plasticity dictates definitive transcriptional outcomes. Future studies must investigate whether these asymmetric binding events represent half-site occupancy that facilitates cooperative binding with uncharacterized cofactors or composite regulatory elements. Furthermore, while our network analysis confirms widespread target engagement, the transition from pan-cancer transcriptomic profiling to patient-specific precision medicine remains a critical hurdle. Evaluating how these integrated STAT networks dynamically rewire in response to the emerging clinical therapies discussed above (such as targeted STAT degraders or immune checkpoint inhibitors) represents a vital next step for the field. Expanding integrative, multi-omics, and lineage-resolved studies will be essential to fully understand STAT-driven transcriptional architecture and to refine therapeutic strategies targeting JAK/STAT signaling in oncology.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/genes17050547/s1>. Supplementary Figure S1: Genomic Synteny of the STAT5 Locus in *Monodelphis domestica*. Visual representation from the NCBI Genome Data Viewer showing the tandem arrangement of the two STAT5A-like loci (canonical STAT5A and LOC100011470) on Chromosome 2 (NC_077228.1) of the *Monodelphis domestica* reference genome. Supplementary Figure S2: Single-cell transcriptomic profiling of STAT family expression

(STAT1-STAT6) in malignant versus non-malignant cells across three independent cancer datasets. Violin plots display log-normalized expression values, with internal boxplots indicating median and interquartile ranges. (A) Melanoma microenvironment (GSE72056). (B) Breast cancer microenvironment (GSE75688). (C) Nasopharyngeal carcinoma microenvironment (GSE150430), highlighting the significant upregulation of STAT3 in cancer cells ($p < 0.001$).

Author Contributions: D.L., validation, formal analysis, writing—review and editing, investigation; P.H.G., conceptualization, methodology, writing—review and editing; F.M.G., conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft, writing—review and editing, visualization, supervision, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Italian Ministry of University and Research, with the following programs: PON “Ricerca e Innovazione” 2014–2020; PRIN project 2022CEHEX8; PNRR program for HPC, Big Data, and Quantum Computing.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in this study is publicly available, with sources stated in the text.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. The Potential and Controversy of Targeting STAT Family Members in Cancer. *Semin. Cancer Biol.* **2020**, *60*, 41–56. [[CrossRef](#)] [[PubMed](#)]
2. Shahmarvand, N.; Nagy, A.; Shahryari, J.; Ohgami, R.S. Mutations in the Signal Transducer and Activator of Transcription Family of Genes in Cancer. *Cancer Sci.* **2018**, *109*, 926–933. [[CrossRef](#)]
3. Yu, H.; Lee, H.; Herrmann, A.; Buettner, R.; Jove, R. Revisiting STAT3 Signalling in Cancer: New and Unexpected Biological Functions. *Nat. Rev. Cancer* **2014**, *14*, 736–746. [[CrossRef](#)]
4. Philips, R.L.; Wang, Y.; Cheon, H.; Kanno, Y.; Gadina, M.; Sartorelli, V.; Horvath, C.M.; Darnell, J.E.; Stark, G.R.; O’Shea, J.J. The JAK-STAT Pathway at 30: Much Learned, Much More to Do. *Cell* **2022**, *185*, 3857–3876. [[CrossRef](#)]
5. Rah, B.; Rather, R.A.; Bhat, G.R.; Baba, A.B.; Mushtaq, I.; Farooq, M.; Yousuf, T.; Dar, S.B.; Parveen, S.; Hassan, R.; et al. JAK/STAT Signaling: Molecular Targets, Therapeutic Opportunities, and Limitations of Targeted Inhibitions in Solid Malignancies. *Front. Pharmacol.* **2022**, *13*, 821344. [[CrossRef](#)] [[PubMed](#)]
6. Samra, S.; Bergerson, J.R.E.; Freeman, A.F.; Turvey, S.E. JAK-STAT Signaling Pathway, Immunodeficiency, Inflammation, Immune Dysregulation, and Inborn Errors of Immunity. *J. Allergy Clin. Immunol.* **2025**, *155*, 357–367. [[CrossRef](#)] [[PubMed](#)]
7. Levy, D.E.; Darnell, J.E. Stats: Transcriptional Control and Biological Impact. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 651–662. [[CrossRef](#)]
8. Li, P.; Huang, D. Targeting the JAK-STAT Pathway in Colorectal Cancer: Mechanisms, Clinical Implications, and Therapeutic Potential. *Front. Cell Dev. Biol.* **2024**, *12*, 1507621. [[CrossRef](#)]
9. Muñoz-García, N.; Jara-Acevedo, M.; Caldas, C.; Bárcena, P.; López, A.; Puig, N.; Alcoceba, M.; Fernández, P.; Villamor, N.; Flores-Montero, J.A.; et al. STAT3 and STAT5B Mutations in T/NK-Cell Chronic Lymphoproliferative Disorders of Large Granular Lymphocytes (LGL): Association with Disease Features. *Cancers* **2020**, *12*, 3508. [[CrossRef](#)]
10. Küçük, C.; Jiang, B.; Hu, X.; Zhang, W.; Chan, J.K.C.; Xiao, W.; Lack, N.; Alkan, C.; Williams, J.C.; Avery, K.N.; et al. Activating Mutations of STAT5B and STAT3 in Lymphomas Derived from $\Gamma\delta$ -T or NK Cells. *Nat. Commun.* **2015**, *6*, 6025. [[CrossRef](#)]
11. Wang, W.; Lopez McDonald, M.C.; Hariprasad, R.; Hamilton, T.; Frank, D.A. Oncogenic STAT Transcription Factors as Targets for Cancer Therapy: Innovative Strategies and Clinical Translation. *Cancers* **2024**, *16*, 1387. [[CrossRef](#)]
12. Igelmann, S.; Neubauer, H.A.; Ferbeyre, G. STAT3 and STAT5 Activation in Solid Cancers. *Cancers* **2019**, *11*, 1428. [[CrossRef](#)]
13. de Araujo, E.D.; Keserü, G.M.; Gunning, P.T.; Moriggl, R. Targeting STAT3 and STAT5 in Cancer. *Cancers* **2020**, *12*, 2002. [[CrossRef](#)] [[PubMed](#)]
14. Wang, H.-Q.; Man, Q.-W.; Huo, F.-Y.; Gao, X.; Lin, H.; Li, S.-R.; Wang, J.; Su, F.-C.; Cai, L.; Shi, Y.; et al. STAT3 Pathway in Cancers: Past, Present, and Future. *MedComm* **2022**, *3*, e124. [[CrossRef](#)]
15. Maninang, C.; Li, J.; Li, W.X. The Prognostic Role of STAT5B Across Cancer Types and Comparative Analysis with STAT5A: A Systematic Review. *Biomolecules* **2025**, *15*, 1503. [[CrossRef](#)]
16. Temple, A.E.; Walker, S.R. The Roles of STAT3 and STAT5 in Breast Cancer. *Cancers* **2025**, *17*, 1781. [[CrossRef](#)]

17. Ebersbach, C.; Beier, A.-M.K.; Thomas, C.; Erb, H.H.H. Impact of STAT Proteins in Tumor Progress and Therapy Resistance in Advanced and Metastasized Prostate Cancer. *Cancers* **2021**, *13*, 4854. [[CrossRef](#)] [[PubMed](#)]
18. Kovacic, B.; Stoiber, D.; Moriggl, R.; Weisz, E.; Ott, R.G.; Kreibich, R.; Levy, D.E.; Beug, H.; Freissmuth, M.; Sexl, V. STAT1 Acts as a Tumor Promoter for Leukemia Development. *Cancer Cell* **2006**, *10*, 77–87. [[CrossRef](#)]
19. Zhang, Y.; Liu, Z. STAT1 in Cancer: Friend or Foe? *Discov. Med.* **2017**, *24*, 19–29. [[PubMed](#)]
20. Carpenter, R.L.; Lo, H.-W. STAT3 Target Genes Relevant to Human Cancers. *Cancers* **2014**, *6*, 897–925. [[CrossRef](#)]
21. McFarland, B.C.; Hong, S.W.; Rajbhandari, R.; Twitty, G.B.; Gray, G.K.; Yu, H.; Benveniste, E.N.; Nozell, S.E. NF- κ B-Induced IL-6 Ensures STAT3 Activation and Tumor Aggressiveness in Glioblastoma. *PLoS ONE* **2013**, *8*, e78728. [[CrossRef](#)]
22. Tian, Y.; Liu, H.; Wang, M.; Wang, R.; Yi, G.; Zhang, M.; Chen, R. Role of STAT3 and NRF2 in Tumors: Potential Targets for Antitumor Therapy. *Molecules* **2022**, *27*, 8768. [[CrossRef](#)]
23. Yang, J.; Liao, X.; Agarwal, M.K.; Barnes, L.; Auron, P.E.; Stark, G.R. Unphosphorylated STAT3 Accumulates in Response to IL-6 and Activates Transcription by Binding to NF κ B. *Genes Dev.* **2007**, *21*, 1396–1408. [[CrossRef](#)]
24. Liongue, C.; Ward, A.C. Evolution of the JAK-STAT Pathway. *Jak-Stat* **2013**, *2*, e22756. [[CrossRef](#)]
25. Wang, Y.; Levy, D.E.C. Elegans STAT: Evolution of a Regulatory Switch. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2006**, *20*, 1641–1652. [[CrossRef](#)]
26. Gorissen, M.; de Vrieze, E.; Flik, G.; Huising, M.O. STAT Genes Display Differential Evolutionary Rates That Correlate with Their Roles in the Endocrine and Immune System. *J. Endocrinol.* **2011**, *209*, 175–184. [[CrossRef](#)] [[PubMed](#)]
27. Babadei, O.; Strobl, B.; Müller, M.; Decker, T. Transcriptional Control of Interferon-Stimulated Genes. *J. Biol. Chem.* **2024**, *300*, 107771. [[CrossRef](#)] [[PubMed](#)]
28. Satoh, J.; Tabunoki, H. A Comprehensive Profile of ChIP-Seq-Based STAT1 Target Genes Suggests the Complexity of STAT1-Mediated Gene Regulatory Mechanisms. *Gene Regul. Syst. Biol.* **2013**, *7*, GRSB.S11433. [[CrossRef](#)]
29. Wu, J.; Jin, Y.-Y.; Gong, R.-L.; Yang, F.; Su, X.-Y.; Chen, T.-X. Genome-Wide ChIP-Seq and RNA-Seq Analyses of STAT3 Target Genes in TLRs Activated Human Peripheral Blood B Cells. *Front. Immunol.* **2022**, *13*, 821457. [[CrossRef](#)]
30. Robertson, G.; Hirst, M.; Bainbridge, M.; Bilenky, M.; Zhao, Y.; Zeng, T.; Euskirchen, G.; Bernier, B.; Varhol, R.; Delaney, A.; et al. Genome-Wide Profiles of STAT1 DNA Association Using Chromatin Immunoprecipitation and Massively Parallel Sequencing. *Nat. Methods* **2007**, *4*, 651–657. [[CrossRef](#)] [[PubMed](#)]
31. Giovannozzi, S.; Demeulemeester, J.; Schrijvers, R.; Gijsbers, R. Transcriptional Profiling of STAT1 Gain-of-Function Reveals Common and Mutation-Specific Fingerprints. *Front. Immunol.* **2021**, *12*, 632997. [[CrossRef](#)] [[PubMed](#)]
32. Karim, A.; Garg, R.; Saikia, B.; Tiwari, A.; Sahu, S.; Malhotra, M.; Minz, R.W.; Rawat, A.; Singh, S.; Suri, D. Unraveling the Unphosphorylated STAT3-Unphosphorylated NF- κ B Pathway in Loss of Function STAT3 Hyper IgE Syndrome. *Front. Immunol.* **2024**, *15*, 1332817. [[CrossRef](#)]
33. Cheng, M.; Liu, Y.; Guo, Y.; Li, M.; Xian, S.; Qin, H.; Yang, Y.; Qian, W.; Tang, J.; Lu, Y.; et al. Pan-Cancer Analysis Reveals Signal Transducer and Activator of Transcription (STAT) Gene Family as Biomarkers for Prognostic Prediction and Therapeutic Guidance. *Front. Genet.* **2023**, *14*, 1120500. [[CrossRef](#)]
34. Seal, R.L.; Braschi, B.; Gray, K.; Jones, T.E.M.; Tweedie, S.; Haim-Vilmovsky, L.; Bruford, E.A. Genenames.Org: The HGNC Resources in 2023. *Nucleic Acids Res.* **2023**, *51*, D1003–D1009. [[CrossRef](#)]
35. Eyre, T.A.; Wright, M.W.; Lush, M.J.; Bruford, E.A. HCOP: A Searchable Database of Human Orthology Predictions. *Brief. Bioinform.* **2007**, *8*, 2–5. [[CrossRef](#)]
36. Giorgi, F.M.; Ceraolo, C.; Mercatelli, D. The R Language: An Engine for Bioinformatics and Data Science. *Life* **2022**, *12*, 648. [[CrossRef](#)] [[PubMed](#)]
37. Edgar, R.C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
38. Kumar, S.; Stecher, G.; Suleski, M.; Sanderford, M.; Sharma, S.; Tamura, K. MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. *Mol. Biol. Evol.* **2024**, *41*, msae263. [[CrossRef](#)] [[PubMed](#)]
39. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An Immeasurable Source of Knowledge. *Contemp. Oncol.* **2015**, *19*, A68–A77. [[CrossRef](#)]
40. GTEx Consortium. The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues. *Science* **2020**, *369*, 1318–1330. [[CrossRef](#)]
41. Wang, Q.; Armenia, J.; Zhang, C.; Penson, A.V.; Reznik, E.; Zhang, L.; Minet, T.; Ochoa, A.; Gross, B.E.; Iacobuzio-Donahue, C.A.; et al. Unifying Cancer and Normal RNA Sequencing Data from Different Sources. *Sci. Data* **2018**, *5*, 180061. [[CrossRef](#)]
42. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
43. Mercatelli, D.; Lopez-Garcia, G.; Giorgi, F.M. Corto: A Lightweight R Package for Gene Network Inference and Master Regulator Analysis. *Bioinformatics* **2020**, *36*, 3916–3917. [[CrossRef](#)]

44. Pozzobon, D.; Bellezza, A.; Giorgi, F.M. Pan-Cancer Upregulation of the FOXM1 Transcription Factor. *Genes* **2025**, *16*, 56. [[CrossRef](#)]
45. Puram, S.V.; Tirosh, I.; Parikh, A.S.; Patel, A.P.; Yizhak, K.; Gillespie, S.; Rodman, C.; Luo, C.L.; Mroz, E.A.; Emerick, K.S.; et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **2017**, *171*, 1611–1624.e24. [[CrossRef](#)] [[PubMed](#)]
46. Christensen, E.; Naidas, A.; Chen, D.; Husic, M.; Shooshtari, P. TMExplorer: A Tumour Microenvironment Single-Cell RNAseq Database and Search Tool. *PLoS ONE* **2022**, *17*, e0272302. [[CrossRef](#)]
47. Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W.M.; Zheng, S.; Butler, A.; Lee, M.J.; Wilk, A.J.; Darby, C.; Zager, M.; et al. Integrated Analysis of Multimodal Single-Cell Data. *Cell* **2021**, *184*, 3573–3587.e29. [[CrossRef](#)]
48. Tirosh, I.; Izar, B.; Prakadan, S.M.; Wadsworth, M.H.; Treacy, D.; Trombetta, J.J.; Rotem, A.; Rodman, C.; Lian, C.; Murphy, G.; et al. Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-Seq. *Science* **2016**, *352*, 189–196. [[CrossRef](#)] [[PubMed](#)]
49. Chung, W.; Eum, H.H.; Lee, H.-O.; Lee, K.-M.; Lee, H.-B.; Kim, K.-T.; Ryu, H.S.; Kim, S.; Lee, J.E.; Park, Y.H.; et al. Single-Cell RNA-Seq Enables Comprehensive Tumour and Immune Cell Profiling in Primary Breast Cancer. *Nat. Commun.* **2017**, *8*, 15081. [[CrossRef](#)]
50. Chen, Y.-P.; Yin, J.-H.; Li, W.-F.; Li, H.-J.; Chen, D.-P.; Zhang, C.-J.; Lv, J.-W.; Wang, Y.-Q.; Li, X.-M.; Li, J.-Y.; et al. Single-Cell Transcriptomics Reveals Regulators Underlying Immune Cell Diversity and Immune Subtypes Associated with Prognosis in Nasopharyngeal Carcinoma. *Cell Res.* **2020**, *30*, 1024–1042. [[CrossRef](#)] [[PubMed](#)]
51. Rozowsky, J.; Euskirchen, G.; Auerbach, R.K.; Zhang, Z.D.; Gibson, T.; Bjornson, R.; Carriero, N.; Snyder, M.; Gerstein, M.B. PeakSeq Enables Systematic Scoring of ChIP-Seq Experiments Relative to Controls. *Nat. Biotechnol.* **2009**, *27*, 66–75. [[CrossRef](#)] [[PubMed](#)]
52. ENCODE Project Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **2012**, *489*, 57–74. [[CrossRef](#)]
53. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A Fast Spliced Aligner with Low Memory Requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)]
54. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
55. Zhang, Y.; Liu, T.; Meyer, C.A.; Eeckhoute, J.; Johnson, D.S.; Bernstein, B.E.; Nusbaum, C.; Myers, R.M.; Brown, M.; Li, W.; et al. Model-Based Analysis of ChIP-Seq (MACS). *Genome Biol.* **2008**, *9*, R137. [[CrossRef](#)]
56. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative Genomics Viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [[CrossRef](#)]
57. Rao, S.S.; Nelson, P.A.; Lunde, H.S.; Haugland, G.T. Evolutionary, Comparative, and Functional Analyses of STATs and Regulation of the JAK-STAT Pathway in Lumpfish upon Bacterial and Poly(I:C) Exposure. *Front. Cell. Infect. Microbiol.* **2023**, *13*, 1252744. [[CrossRef](#)]
58. Wang, Y.; Levy, D.E. Comparative Evolutionary Genomics of the STAT Family of Transcription Factors. *JAK-STAT* **2012**, *1*, 23–33. [[CrossRef](#)]
59. Davesne, D.; Friedman, M.; Schmitt, A.D.; Fernandez, V.; Carnevale, G.; Ahlberg, P.E.; Sanchez, S.; Benson, R.B.J. Fossilized Cell Structures Identify an Ancient Origin for the Teleost Whole-Genome Duplication. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2101780118. [[CrossRef](#)] [[PubMed](#)]
60. Tolomeo, M.; Cascio, A. STAT4 and STAT6, Their Role in Cellular and Humoral Immunity and in Diverse Human Diseases. *Int. Rev. Immunol.* **2024**, *43*, 394–418. [[CrossRef](#)] [[PubMed](#)]
61. Platanitis, E.; Decker, T. Regulatory Networks Involving STATs, IRFs, and NFκB in Inflammation. *Front. Immunol.* **2018**, *9*, 2542. [[CrossRef](#)]
62. Sikorski, K.; Chmielewski, S.; Przybyl, L.; Heemann, U.; Wesoly, J.; Baumann, M.; Bluysen, H.A.R. STAT1-Mediated Signal Integration between IFNγ and LPS Leads to Increased EC and SMC Activation and Monocyte Adhesion. *Am. J. Physiol. Cell Physiol.* **2011**, *300*, C1337–C1344. [[CrossRef](#)]
63. Chen, X.; Vinkemeier, U.; Zhao, Y.; Jeruzalmi, D.; Darnell, J.E.; Kuriyan, J. Crystal Structure of a Tyrosine Phosphorylated STAT-1 Dimer Bound to DNA. *Cell* **1998**, *93*, 827–839. [[CrossRef](#)]
64. Menon, P.R.; Staab, J.; Gregus, A.; Wirths, O.; Meyer, T. An Inhibitory Effect on the Nuclear Accumulation of Phospho-STAT1 by Its Unphosphorylated Form. *Cell Commun. Signal. CCS* **2022**, *20*, 42. [[CrossRef](#)]
65. Mao, X.; Ren, Z.; Parker, G.N.; Sondermann, H.; Pastorello, M.A.; Wang, W.; McMurray, J.S.; Demeler, B.; Darnell, J.E.; Chen, X. Structural Bases of Unphosphorylated STAT1 Association and Receptor Binding. *Mol. Cell* **2005**, *17*, 761–771. [[CrossRef](#)]
66. Begitt, A.; Droscher, M.; Meyer, T.; Schmid, C.D.; Baker, M.; Antunes, F.; Knobloch, K.-P.; Owen, M.R.; Naumann, R.; Decker, T.; et al. STAT1-Cooperative DNA Binding Distinguishes Type 1 from Type 2 Interferon Signaling. *Nat. Immunol.* **2014**, *15*, 168–176. [[CrossRef](#)] [[PubMed](#)]

67. Bhinge, A.A.; Kim, J.; Euskirchen, G.M.; Snyder, M.; Iyer, V.R. Mapping the Chromosomal Targets of STAT1 by Sequence Tag Analysis of Genomic Enrichment (STAGE). *Genome Res.* **2007**, *17*, 910–916. [[CrossRef](#)]
68. Bonham, A.J.; Wenta, N.; Osslund, L.M.; Prussin, A.J.; Vinkemeier, U.; Reich, N.O. STAT1:DNA Sequence-Dependent Binding Modulation by Phosphorylation, Protein:Protein Interactions and Small-Molecule Inhibition. *Nucleic Acids Res.* **2013**, *41*, 754–763. [[CrossRef](#)] [[PubMed](#)]
69. Veit, E.C.; Salim, M.S.; Jung, M.J.; Richardson, R.B.; Boys, I.N.; Quinlan, M.; Barrall, E.A.; Bednarski, E.; Hamilton, R.E.; Kikawa, C.; et al. Evolution of STAT2 Resistance to Flavivirus NS5 Occurred Multiple Times despite Genetic Constraints. *Nat. Commun.* **2024**, *15*, 5426. [[CrossRef](#)] [[PubMed](#)]
70. Meissl, K.; Macho-Maschler, S.; Müller, M.; Strobl, B. The Good and the Bad Faces of STAT1 in Solid Tumours. *Cytokine* **2017**, *89*, 12–20. [[CrossRef](#)]
71. Braverman, E.L.; Mognol, G.P.; Minn, A.J.; Vignali, D.A.A.; Varner, J.A. One Step Ahead: Preventing Tumor Adaptation to Immune Therapy. *Am. Soc. Clin. Oncol. Educ. Book* **2025**, *45*, e481556. [[CrossRef](#)] [[PubMed](#)]
72. Cheon, H.; Borden, E.C.; Stark, G.R. Interferons and Their Stimulated Genes in the Tumor Microenvironment. *Semin. Oncol.* **2014**, *41*, 156–173. [[CrossRef](#)]
73. Garcia-Diaz, A.; Shin, D.S.; Moreno, B.H.; Saco, J.; Escuin-Ordinas, H.; Rodriguez, G.A.; Zaretsky, J.M.; Sun, L.; Hugo, W.; Wang, X.; et al. Interferon Receptor Signaling Pathways Regulating PD-L1 and PD-L2 Expression. *Cell Rep.* **2017**, *19*, 1189–1201. [[CrossRef](#)]
74. Yang, X.; Liu, Y.; Cao, J.; Wu, C.; Tang, L.; Bian, W.; Chen, Y.; Yu, L.; Wu, Y.; Li, S.; et al. Targeting Epigenetic and Post-Translational Modifications of NRF2: Key Regulatory Factors in Disease Treatment. *Cell Death Discov.* **2025**, *11*, 189. [[CrossRef](#)] [[PubMed](#)]
75. Hyun, K.; Ahn, J.; Kim, H.; Kim, J.; Kim, Y.-I.; Park, H.-S.; Roeder, R.G.; Lee, J.E.; Kim, J. The BAF Complex Enhances Transcription through Interaction with H3K56ac in the Histone Globular Domain. *Nat. Commun.* **2024**, *15*, 9614. [[CrossRef](#)]
76. Hu, Y.; Shen, Y.; Zhao, Y.; Tang, Y.; Liu, C.; Gu, Y.; Yang, T.; Shen, Y. Genomic Distribution of Signal Transducer and Activator of Transcription (STAT) Family in Colorectal Cancer. *Hum. Cell* **2023**, *36*, 286–295. [[CrossRef](#)]
77. Begitt, A.; Krause, S.; Cavey, J.R.; Vinkemeier, D.E.; Vinkemeier, U. A Family-Wide Assessment of Latent STAT Transcription Factor Interactions Reveals Divergent Dimer Repertoires. *J. Biol. Chem.* **2023**, *299*, 104703. [[CrossRef](#)]
78. Shawky, A.M.; Almalki, F.A.; Abdalla, A.N.; Abdelazeem, A.H.; Gouda, A.M. A Comprehensive Overview of Globally Approved JAK Inhibitors. *Pharmaceutics* **2022**, *14*, 1001. [[CrossRef](#)]
79. Chen, R.J.; Lu, M.Y.; Williamson, D.F.K.; Chen, T.Y.; Lipkova, J.; Noor, Z.; Shaban, M.; Shady, M.; Williams, M.; Joo, B.; et al. Pan-Cancer Integrative Histology-Genomic Analysis via Multimodal Deep Learning. *Cancer Cell* **2022**, *40*, 865–878.e6. [[CrossRef](#)]
80. Zhou, M.; Zhang, P.; Da, M.; Yang, R.; Ma, Y.; Zhao, J.; Ma, T.; Xia, J.; Shen, G.; Chen, Y.; et al. A Pan-Cancer Analysis of the Expression of STAT Family Genes in Tumors and Their Relationship to the Tumor Microenvironment. *Front. Oncol.* **2022**, *12*, 925537. [[CrossRef](#)] [[PubMed](#)]
81. Muckenhuber, M.; Seufert, I.; Müller-Ott, K.; Mallm, J.-P.; Klett, L.C.; Knotz, C.; Hechler, J.; Kepper, N.; Erdel, F.; Rippe, K. Epigenetic Signals That Direct Cell Type-Specific Interferon Beta Response in Mouse Cells. *Life Sci. Alliance* **2023**, *6*, e202201823. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.