university of groningen

umcg

Investigating the relation between copy number alterations and immune transcriptional footprints in the context of spatial transcriptomics in cancer

Research Project II

(WMBM902-30)

Intern: Argyro Anagnostopoulou (S5340411)

Daily supervisor(s): dr. A. (Arkajyoti) Bhattacharya, S. (Stefan) Loipfinger

Supervisor: Prof. dr. R.S.N. (Rudolf) Fehrmann

First examiner: Prof. dr. M. (Marco) de Bruyn

MSc Biomedical Sciences: Biology of Cancer and Immune System (2022-2024)

Translational Bioinformatics Group, Department of Medical Oncology, University Medical Center Groningen (UMCG), Netherlands

Start date: 15/01/2024 - **End date:** 16/06/2024

Abstract

Cancer is a leading cause of death worldwide, necessitating novel therapies and effective patient stratification. Copy Number Alterations (CNAs), characterized by chromosomal imbalances and genomic instability, are a hallmark of cancer. They contribute to tumorigenesis, treatment resistance, and poor prognosis, and together with immune evasion, CNAs shape the "cancer-immune set point," crucial for effective immunotherapy. Recent studies reveal CNAs impact the immune system, influencing immune evasion and therapy responses. However, the specific effects of CNAs on the cancer-immune set point remain unclear. This research project aimed to uncover new insights into the relationship between CNAs and immune transcriptional profiles using spatial transcriptomics data across various cancer types. We applied bioinformatics techniques using transcriptional components (TCs) derived from consensus independent component analysis (c-ICA) of 10,817 bulk transcriptomic profiles. Gene set enrichment analysis identified 235 TCs associated with immunological processes (immune-TCs), while Transcriptional Adaptation to CNA (TACNA) profiling identified 476 TCs capturing the transcriptional effects of CNAs (CNA-TCs). We used in-house tools for Projection and Colocalization analyses to obtain activity scores and correlation values for these TCs, evaluating the association between transcriptional processes affected by CNAs and immune processes. We identified the top 12 positively and inversely colocalized TCs in spatial transcriptomics samples and conducted literature search on the genomic regions where CNAs are present and the top genes. Most results aligned with existing literature, with some genes identified as cancer drivers and biomarkers in specific cancers, while others require further exploration. This project provides a comprehensive landscape of the association between CNAs and immune transcriptional footprints across various cancers using spatial transcriptomics. This approach could offer new insights into anti-cancer immune responses and guide the development of targeted cancer therapies.

Keywords: spatial transcriptomics, CNAs, immune evasion, cancer-immune set point, c-ICA, GSEA, TACNA profiling.

Abbreviations:

APCs	Antigen - presenting cells
cfDNA	Circulating cell-free DNA
c-ICA	Consensus independent component analysis
CIN	Chromosomal instability
CNAs	Copy number alterations
CNA-TCs	TCs capturing CNA-related processes
CRC	Colorectal cancer
CTLA-4	Cytotoxic T lymphocyte - associated protein 4
EDA	Exploratory data analysis
ESCC	Esophageal squamous cell carcinoma
GOBP	Gene ontology – biological processes
GSEA	Gene set enrichment analysis
H&E	Hematoxylin & eosin staining
HCC	Hepatocellular carcinoma
HLA	Human leukocyte antigen
HLA-DPA1	Major histocompatibility complex, class II, DP alpha 1
HLA-DRB5	Major histocompatibility complex, class II, DR beta 5
IFNs	Interferons
IL	Interleukin
Immune-TCs	TCs capturing immune-related processes
LOH	Loss of heterozygosity
MHC	Major histocompatibility complex
MM	Mixing matrix
MSigDB	Molecular signatures database
mtROS	Mitochondrial reactive oxygen species
NK	Natural killer cells
NSCLC	Non-small cell lung cancer
PC1	First principal component
PCA	Principal component analysis
PD-1	Programmed cell death protein 1
PI3K	1-phosphatidylinositol 3-kinase
PLEKHG7	Pleckstrin homology and RhoGEF domain containing G7
RSEM	RNA-seq with expectation maximization
SCNAs	Somatic copy number alterations
TACNA	Transcriptional Adaptation to CNA profiling
TC	Transcriptional component
TCGA	The cancer genome atlas
TLR	Toll-like receptors.
TMB	Tumor mutational burden
TME	Tumor microenvironment
TNBC	Triple-negative breast cancer
TNF	Tumor-necrosis factor
Top colocalized TCs	CNA- and/or immune-TCs exhibiting the highest/lowest correlation values from the colocalization analysis, including both the most positively and inversely colocalized TCs.
TP53	Tumor protein P53
Tregs	Regulatory T cells

Table of contents

1. Introduction

Cancer is a leading cause of death worldwide and novel treatments need to be developed. Traditional approaches such as radiation, chemotherapy, surgery and immunotherapy used in treatment schemes to reduce or completely eradicate cancer cell proliferation, have undoubtedly improved survival rates among cancer patients [1]. However, many still face recurrence and struggle to survive in long term, partially due to the inherent evolutionary nature of cancer, which allows it to develop resistance to chemotherapy [1] [2]. Thus, understanding the underlying molecular mechanisms is crucial for developing innovative therapeutics. In this research project, we delve into two critical cancer hallmarks which drive cancer progression: genomic instability and immune evasion.

1.1 Copy Number Alterations (CNAs): drivers of cancer progression

Genomic instability, a hallmark of cancer, drives random mutations, encompassing chromosomal rearrangements, facilitating cancer cell survival, growth and metastasis (Fig. 1) [3]. This instability, often characterized by chromosomal instability (CIN), is marked by frequent chromosomal missegregation during mitosis and significantly contributes to chromosomal alterations in cancer [2]. Aneuploidy, involving changes in the copy number of chromosomal segments, is prevalent in human tumors and is considered a primary cause of tumorigenesis [4]. It includes both broad somatic copy number alterations (CNAs) across large chromosomal regions and smaller focal CNAs [5], which are pervasive across all human malignancies, in approximately 50% of hematological tumors and 90% of solid cancers [6].

CNAs formation has been suggested to occur via processes connected to somatic alterations in DNA structure [7], such as replication stress causing double-strand breaks to accumulate in common genomic regions, thereby enhancing the occurrence of amplifications and deletions [8]. Accumulation of CNAs occurs in the majority of cancer types throughout disease progression [9] [10], with recurrent patterns linked to tumor prognosis and progression [11]. In translational research, analyzing CNA patterns across various cancer types is crucial for developing novel therapeutic strategies [12]. Large-scale studies have revealed recurrent CNA profiles spanning various cancer types [12], and utilizing CNAs in circulating cell-free DNA (cfDNA) as biomarkers has shown promise in classifying cancer types and their tissue origins [13] [14] [15].

Moreover, CNAs have been directly linked to the expression of driver genes [16] [17], where homozygous or heterozygous deletions can decrease tumor suppressor genes (e.g. CDKN2A/B, TP53, ATM loci) and copy number amplifications can boost the expression of oncogenes (e.g. MYC, MYCN, REL loci) [12] [18]. Recent research highlights the importance of chromosomal CNAs as a significant source of treatment resistance and poor prognosis for cancer patients [2]. For instance, the ovarian cancer cell line HeLa, when exposed to varying doses of methotrexate, developed resistance and recurrently gained extra copies of chromosome 5, which contains the DHFR gene targeted by methotrexate [19]. Despite significant advancements, the relationship between chromosomal abnormalities and tumor development remains incompletely understood

[6]. The effect of aneuploidy and CNAs varies greatly depending on the context [6] and factors such as tumor stage, cell type, genetic make- up, tumor microenvironment (TME) and immune system interactions [20]. While generally considered detrimental, under specific circumstances, these alterations can provide a fitness advantage, driving tumorigenesis [20].

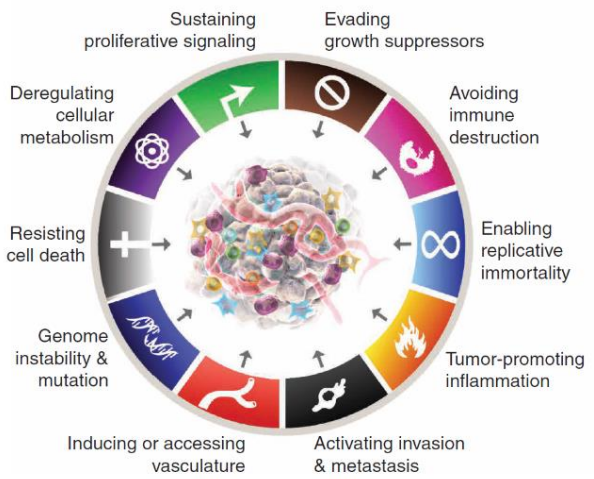

Figure 1: The hallmarks of cancer. The current definition of the hallmarks of cancer comprises eight hallmark capabilities and two enabling characteristics [21]. The two provisional emerging hallmarks introduced in 2011 [3], "reprogramming cellular metabolism" and "avoiding immune destruction", have been validated enough to be integrated into the fundamental set, alongside the original six acquired capabilities that were proposed in 2000 [22]. (Picture obtained from Hanahan [21])

1.2 Immune evasion in cancer

The immune system plays a crucial role in cancer immune destruction by preventing or eliminating the development and spread of early neoplasias, micrometastases, and advanced tumors [3]. This surveillance process is encapsulated in the cancer immunity cycle [23]. Immune evasion, a hallmark of cancer, involves the selection of tumor variants that develop resistance to immunological attacks, mainly by T cells and natural killer (NK) cells [24] [25]. Investigating the mechanisms through which the tumor somatic copy number alteration (SCNA) levels affect the immune evasion is crucial, as this could enhance the effectiveness of immune checkpoint blockade [26]. Immune evasion can be achieved by neoantigen editing, antigen presentation abnormalities, tumor infiltration suppression, and/or immune cell cytotoxicity [26]. Within the tumor site, tumor cells produce neoantigens, often resulting from non-synonymous somatic mutations [27]. These neoantigens are presented on the cell surface via the major histocompatibility complex (MHC) molecules to initiate the cancer immune cycle [27]. Antigen-presenting cells (APCs), like dendritic cells, harbor antigens in the body and present their fragments via MHC I or MHC II [23]. Upon capturing a cancer neoantigen and receiving signals from proinflammatory cytokines or cellular debris, APCs migrate to lymph nodes, where they stimulate the growth and maturation of naive T cells into cytotoxic effector T cells by receiving cues from proinflammatory cytokines or cell debris [23]. Subsequently, these activated effector T cells travel to the tumor location, where they identify cancer cell peptide-MHC I complexes via docking with their T cell receptors, leading to the eradication of

the tumor cells [23]. Cancer cells elimination further enhances the cancer immunity cycle by releasing more cancer neoantigens and immunogenic signals [23].

Nevertheless, disruptions to this cycle may occur if vital immune components are lacking or impaired in the microenvironment [23]. The genetic loss of the interferon (IFN) gene cluster and the epigenetic suppression of inflammatory mediators are direct mechanisms that impact the immune response within the TME [27]. Moreover, oncogenic changes, including RAS protein activation, can directly stimulate inflammation within the tumor, angiogenesis, and immune suppression [28]. Cancer cells may also evade immune surveillance by reducing tumor antigen presentation or completely destroying the antigen presentation system [27]. Strategies aimed at blocking inhibitory molecules that control immunological responses, forming the basis of immune checkpoint blockade treatments, are crucial for overcoming immune evasion [29].

1.3 The association between the immune-profile and CNAs in cancer

The immune profile of an individual is influenced by a multitude of extrinsic factors, like sunlight exposure, infections, the microbiome, medications, as well as intrinsic characteristics such as tumor genetic make-up, cytokine secretion, germline genetics, age or negative regulators (e.g. PD-L1/PD-1 axis) (Fig. 2) [30]. These factors establish the "cancer-immune set point," which is the balance between variables that either stimulate or inhibit anticancer immunity, denoting the threshold that must be crossed for immunotherapy to be effective for cancer patients [30].

An inflammatory microenvironment in tumors reflects changes in several cellular and external variables (Fig. 2b) [30]. The cellular composition of the tumor, such as immune cells in the parenchyma or at the invasive edge, can indicate the extent of inflammation [24] [31]. Additionally, proinflammatory cytokines such as type I and type II IFNs, interleukin (IL)-12, IL-23, IL-1β, tumor-necrosis factor (TNF)-α, and IL-2 are present in inflamed tumors and create a more favorable environment for T-cell activation and growth [30]. Nevertheless, it needs to be clarified if these cytokines cause or result from cellular influx [30]. On the other hand, non-inflamed tumors typically express cytokines linked to immune suppression or tolerance and may involve cells linked to immune suppression or tissue homeostasis, such as regulatory T cells (Tregs), myeloid-derived suppressor cells, and tumor-associated macrophages (M2 macrophages) [30]. A range of cytokines, chemokines, and tumor-associated cell types dictate the level of inflammation needed for an effective antitumor immune response post-immunotherapy [30]. Figure 2b illustrates a bell curve representing the population distribution of patients, where minor variations in these variables may shift the balance between tolerance and immunity, determining whether a patient or a specific tumor responds to immunotherapy [30].

Recent studies have investigated the relationship between CNAs and the immune system across various cancer types, revealing significant insights into how these genetic changes influence immune evasion and therapy responses. It has been demonstrated that broad and focal CNA loads exhibit varying associations with gene expression markers linked to immune evasion, cell proliferation, and other cancer hallmarks [26] [32], implicating CNAs in the carcinogenic

process through distinct pathways [5]. High levels of broad CNAs are strongly correlated with immune evasion markers in various cancer types and with a decreased immunotherapy response (e.g. in non-small cell lung cancer (NSCLC) and melanoma [26] [32] [33]). Conversely, elevated levels of focal events are associated with proliferation markers [26] and often have greater prognostic value in contrast with broad CNAs [34]. Therefore, gene dosage imbalance controlling specific gene changes may rule the interplay between the cancer genome and the immune system [5]. Furthermore, the strength and direction of the correlation between CNAs and tumor immunity may vary across different cancer types [33] [35] [36] [37], highlighting the necessity of examining the effects of CNAs in the context of each cancer tissue [5].

Davoli et al. [26] found that high SCNA levels are associated with immune evasion mechanisms, including decreased expression of cytotoxic immune cell markers and reduced leukocyte infiltration, leading to poorer outcomes in anti-CTLA-4 therapy for melanoma patients. Similarly, Thorsson et al. [32] identified six immune subtypes across 33 cancer types, linked to specific driver mutations and chromosomal alterations affecting immune cell infiltration, emphasizing the role of CNAs in shaping the tumor-immune microenvironment. In high-grade serous ovarian cancer, Jiménez-Sánchez et al. [36] demonstrated that immune cell exclusion is correlated with the Myc amplification and upregulated Wnt signaling, indicating a connection between CNAs and immune suppression. Additionally, Braun et al. [37] showed that in advanced clear cell renal cell carcinoma, chromosomal alterations, like the loss of 9p21.3, correlate with resistance to programmed cell death protein 1 (PD-1) blockade, further illustrating the impact of CNAs on the efficacy of immunotherapy. Bassaganyas et al. [5] focused on hepatocellular carcinoma (HCC) and showed that tumors with high broad CNA burdens exhibit immune exclusion traits and genetic dysfunctions, such as tumor protein P53 (TP53) mutations and DNA repair issues, while those with low CNA burdens present an immune-active profile with high inflammation and cytolytic activity. This study suggests that CNA scores could be predictive markers for immunotherapy response in HCC. Additionally, Budczies et al. [35] revealed that high CNA load is associated with low immune cell infiltrates across several cancer types, such as pancreatic adenocarcinoma, bladder urothelial carcinoma, stomach adenocarcinoma, esophageal carcinoma, breast invasive carcinoma, prostate adenocarcinoma, and papillary thyroid carcinoma. This negative correlation suggests that tumors with high CNA load tend to have a less active immune environment, which may impact the effectiveness of immunotherapies. Furthermore, a recent pan-cancer study by Bhattacharya et al. [38] introduced a novel computational method called Transcriptional Adaptation to CNA (TACNA) profiling. Through the analysis of over 34,000 gene expression profiles from cancer samples, they revealed how transcriptional adaptation to CNAs is linked to specific biological processes, and identified four genes that may play a role in tumor immune evasion when their expression is influenced by CNAs. Overall, these findings underscore the critical role of CNAs in shaping the cancer-immune microenvironment, often reducing immune cell activity and infiltration, contributing to immune evasion and therapy resistance.

Even though these studies highlight the significance of CNA burden in immunotherapy response, the specific effects of particular CNAs on the anti-cancer immune response remain unclear. Moreover, none of the abovementioned studies, nor any other published research, has employed spatial transcriptomics tools across multiple cancer types in combination with

consensus Independent Component Analysis (c-ICA) methodology, as explained in detail in the Materials & Methods section, to investigate this field. Current studies that conduct transcriptional analyses often face the challenge of mixed biological signals, where the more dominant signals tend to overshadow others, potentially obscuring critical subtle biological processes. This challenge could be addressed using c-ICA to desegregate these signals.

A recent, unpublished study of our research group explored the influence of CNAs on the cancer-immune set point in a pan-cancer setting, utilizing a novel spatial transcriptomics analysis, which included c-ICA on over 290,000 transcriptomic profiles. This analysis identified transcriptional patterns in the transcriptional components (TCs), associated with CNA burden and immune processes, which were further linked to specific biological pathways. The TCs explaining 90% variance in datasets were used in c-ICA. The study revealed a significant association between immune-TCs and high CNA burden across various cancer types was demonstrated, where immune-TCs were significantly inversely or positively associated with high CNA burden in more than 5 types of cancer. Of note, the majority (66%) of the immune-TCs exhibited an inverse correlation with high CNA burden.

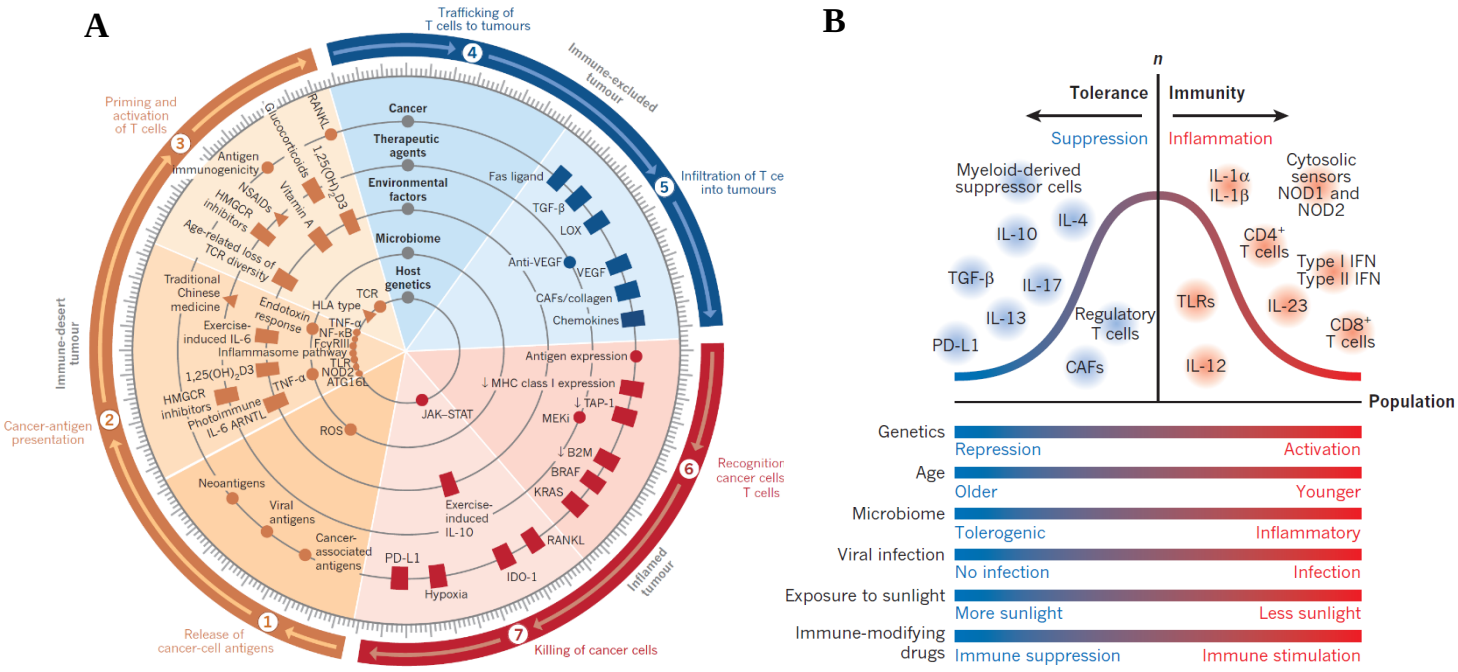

Figure 2: Factors of the cancer-immune set point. (A) Representation of cancer immunity and its variables influencing the cancer-immune set point, with each factor according to its primary action in the cancer-immunity cycle, indicated by rings [30]. **(B)** Multiple factors influence the balance between tolerance and immunity. Anticancer immunity is influenced by various factors that can result in the development of activated T-cell immunity or tolerance [30]. Factors such as genetics, age, microbiota, infections, sunlight exposure, and immune- suppressive medications, affecting the expression of cytokines and cell types that establish the cancer-immune set point [30]. The bell curve illustrates population differences in these factors, which affect whether a tumor will exhibit an inflamed or non-inflamed phenotype, establishing the cancer-immune set point [30]. Most people likely have a small imbalance between tolerogenic and immunogenic factors, indicating that random factors may influence whether a given patient, or tumor, responds to immunotherapy [30]. *n: number of people* (Pictures obtained from Chen & Mellman [30])

1.4 Research project aim

The aim of this research project was to uncover new insights into the relationship between the CNAs and the immune transcriptional profiles using spatial transcriptomics data across various cancer types. By employing a pan-cancer approach we aimed to comprehensively understand how CNAs influence the cancer-immune set point across a large number of different cancer samples, by investigating transcriptional patterns associated with CNAs and immune processes. To achieve this, we employed spatial transcriptomics tools together with the TACNA methodology, which is described in the Materials & Methods section. The main objective of this study was to assess the colocalization of the CNA-TCs and immune-TCs on the spatial transcriptomics data. To achieve this, the workflow presented in Figure 3 was followed.

First, bulk transcriptomic profiles derived from tumor biopsies from the public repository The Cancer Genome Atlas (TCGA) (10,817 profiles) were collected. To decompose the bulk transcriptome profiles into statistically independent TCs, c-ICA was applied, explaining 100% of the variance. This was previously conducted within the framework of a project in our research group. Following, Gene Set Enrichment Analysis (GSEA) was performed on the TCs to identify enriched biological pathways. Subsequently, Exploratory Data Analysis (EDA) on the spatial transcriptomics data was required, to obtain summary statistics and evaluate the data quality and distribution. Next, we identified the CNA-TCs and immune-TCs, which represented TCs reflecting the transcriptional effect of CNAs and immune-related processes respectively. For the Colocalization analysis of the identified TCs on the spatial transcriptomics cancer samples, the Projection method was required first to obtain the activity scores of the TCs. Finally, to address the main objective of this project, literature search of intriguing patterns in biological pathways was conducted, based on the findings from the Colocalization analysis.

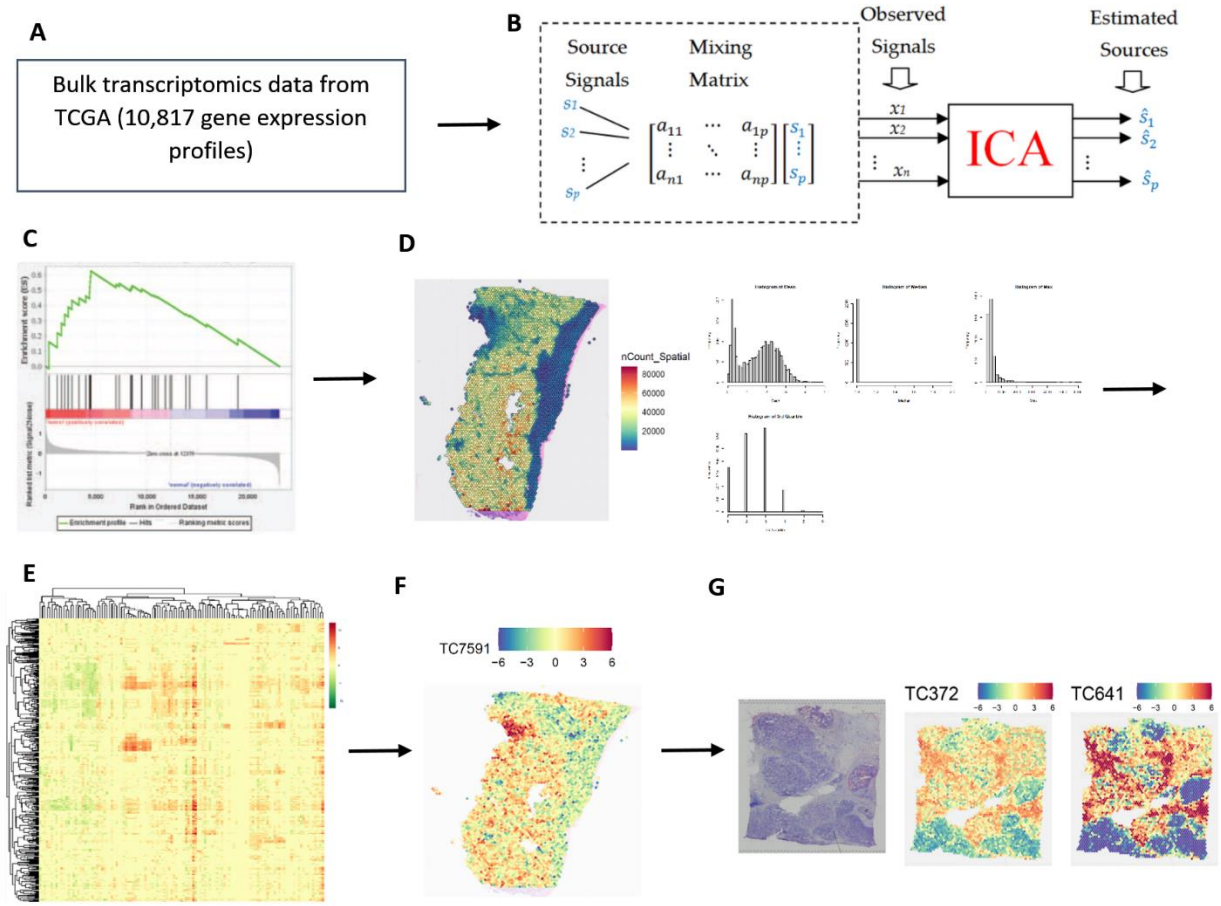

Figure 3: Workflow of the research project. (A) Bulk transcriptomics profiles, **(B)** consensus Independent Component Analysis (c-ICA), **(C)** Gene Set Enrichment Analysis (GSEA) on TCGA 100% explained variance TCs, **(D)** Exploratory Data Analysis (EDA) on the spatial transcriptomics samples, **(E)** Identification of CNA-TCs and immune-TCs. **(F)** Projection of CNA-TCs and immune-TCs on spatial transcriptomics samples, **(G)** Colocalization of the identified TCs on spatial transcriptomics samples and analysis of intriguing patterns in biological pathways. (Pictures for c-ICA and GSEA representations were obtained from [39] and [40] respectively.)

2. Materials and Methods

2.1 Data acquisition

Pre-processed and standardized level 3 RNA-Seq (version 2) data for 21 cancer samples from The Cancer Genome Atlas (TCGA) were collected through the Broad GDAC Firehose portal [41]. The TCGA data consisted of 10,817 profiles transcriptional profiles generated with RNA-Seq analysis, containing 20,392 genes [41]. For each sample, RNA-Seq with Expectation Maximization (RSEM) gene normalized data (identifier: illuminahiseq_rnaseqv2-RSEM_genes_normalized) were retrieved [42]. Additionally, RNA-Seq expression level read counts underwent further normalization in order to guarantee a constant normalization, using FPKMUQ (Fragments per Kilo-base of transcript per Million mapped reads upper quartile normalization). The most recent gene annotations, including HGNC, NCBI ID, and Ensembl ID, were retrieved by accessing the relevant data from the HUGO Gene Nomenclature Committee [43] in October 19, 2020. This step aimed to ensure that the gene annotations used in the analysis were comprehensive and up-to-date. Absent NCBI IDs and Ensembl IDs were supplemented with the latest non-curated data derived from external databases.

Spatial transcriptomics data were obtained from publicly available samples in the 10x Genomics Visium repository. The datasets comprised spatial transcriptomic profiles of cancer samples in h5 format, accompanied by stained images, scale factors, and tissue positions, where spots devoid of tissue filtered out. The selection of these samples was based on their availability within the platform at the time of study. Individual identifiers for all samples used in this research project are provided in Table 1. Prior to analysis, we performed quality control of the data. Genes that showed no expression in any spot within a spatial transcriptomic image were excluded. Additionally, we employed the SpatialFeaturePlot function on the raw data to visualize the total expression of all genes across the spots within each spatial transcriptomics sample. This allowed for a comprehensive examination of gene expression patterns across the entire image. Next, the "LogNormalize" function in Seurat v5.0.3 [44] was used to normalize the raw counts of every spot in the image, in order to account for variance in sequencing depth across data points. Through the normalization method the raw counts of each spot were divided by the total counts for that spot, multiplying the result by a scaling factor of 10,000, and then transformed using the log(x+1) function. To mitigate platform-specific effects, Principal Component Analysis (PCA) was performed on the correlation matrix of the log-normalized data of preprocessed transcriptomic profiles. The dominant features and patterns captured by the first principal component (PC1), constituting the maximum variance direction in the data - approximately 85% of the total variance- reflects platform-specific variability and exhibits consistent loadings across all samples. By excluding this component, we generated transcriptomic profiles devoid of platform-specific artifacts.

Table 1: Spatial transcriptomics datasets analyzed in this research project.

Sample	Cancer type	Additional information	Reference link in 10x Genomics
GB1	Brain cancer, glioblastoma		https://www.10xgenomics.com/datasets/human-brain-cancer-11-mm-capture-area-ffpe-2-standard

GB2	Brain cancer, glioblastoma		https://www.10xgenomics.com/datasets/human-glioblastoma-whole-transcriptome-analysis-1-standard-1-2-0
LC1	Lung cancer, neuroendocrine carcinoma	Stage (AJCC): IB, TNM system: T2aN0MX	https://www.10xgenomics.com/datasets/human-lung-cancer-11-mm-capture-area-ffpe-2-standard
LC2	Lung cancer, squamous cell carcinoma		https://www.10xgenomics.com/datasets/human-lung-cancer-ffpe-2-standard
CRC1	Colorectal cancer, adenocarcinoma	Stage (AJCC): II-A, TNM system: T3	https://www.10xgenomics.com/datasets/human-colorectal-cancer-11-mm-capture-area-ffpe-2-standard
CRC2	Large intestine colorectal cancer		https://www.10xgenomics.com/datasets/human-intestine-cancer-1-standard
CRC3	Invasive adenocarcinoma of the large intestine		https://www.10xgenomics.com/datasets/human-colorectal-cancer-whole-transcriptome-analysis-1-standard-1-2-0
OC1	Ovarian serous carcinoma	High grade	https://www.10xgenomics.com/datasets/human-ovarian-cancer-11-mm-capture-area-ffpe-2-standard
OC2	Ovarian cancer, serous papillary carcinoma		https://www.10xgenomics.com/datasets/human-ovarian-cancer-1-standard
OC3	Endometrial adenocarcinoma of the ovary		https://www.10xgenomics.com/datasets/human-ovarian-cancer-targeted-immunology-panel-stains-dapi-anti-pan-ck-anti-cd-45-1-standard-1-2-0
M1	Skin cancer, malignant melanoma		https://www.10xgenomics.com/datasets/human-melanoma-if-stained-ffpe-2-standard
PC1	Prostate cancer	Block 1E333_Tp11 Section 1, Stage II, Total Gleason score: 7, Sex: Male	https://www.10xgenomics.com/datasets/human-prostate-cancer-adjacent-normal-section-with-if-staining-ffpe-1-standard
PC2	Prostate cancer	Block 1D1061_Tp11 Section 1, Stage IV, Total Gleason score: 7, Sex: Male	https://www.10xgenomics.com/datasets/human-prostate-cancer-acinar-cell-carcinoma-ffpe-1-standard
PC3	Prostate cancer	Block 1E500_Tp12 Section 1, Stage III, Total Gleason score: 7, Sex: Male	https://www.10xgenomics.com/datasets/human-prostate-cancer-adenocarcinoma-with-invasive-carcinoma-ffpe-1-standard-1-3-0
BC1	Breast cancer	Block 738811QB Section 1. Grade II, Ethnicity: Asian, Age: 73	https://www.10xgenomics.com/datasets/human-breast-cancer-ductal-carcinoma-in-situ-invasive-carcinoma-ffpe-1-standard-1-3-0
BC2	Breast cancer	Stage AJCC/UICC - T2N0M0, ER positive, PR negative, Hercep Test 2+, Block ID: 1168993F, Tumor grade III, Age: 60s	https://www.10xgenomics.com/datasets/human-breast-cancer-visium-fresh-frozen-whole-transcriptome-1-standard
BC3	Breast cancer, invasive ductal carcinoma	Stage AJCC/UICC, Group IIA, ER positive, PR negative, Her2 positive and annotated with: ductal carcinoma in situ, lobular carcinoma in situ, invasive carcinoma	https://www.10xgenomics.com/datasets/human-breast-cancer-block-a-section-1-1-standard-1-1-0
BC4	Breast cancer	Section 2 of BC3	https://www.10xgenomics.com/datasets/human-breast-cancer-block-a-section-2-1-standard-1-1-0
BC5	Breast cancer, invasive lobular carcinoma	Stage AJCC/UICC, Group I, ER positive, PR positive, HER2 negative.	https://www.10xgenomics.com/datasets/human-breast-cancer-whole-transcriptome-analysis-1-standard-1-2-0
CC1	Invasive cervical cancer, squamous cell carcinoma	T1bN0M0 IB, Block C00084155.1a	https://www.10xgenomics.com/datasets/human-cervical-cancer-1-standard
IDC1	Infiltrating ductal carcinoma of the breast	Tumor grade - III, AJCC/UICC Stage - T2N0M0, AJCC/UICC Stage group - IIA, ER - Positive, PR - Negative, Hercep test - 2+, Menopausal age at excision: 50+	https://www.10xgenomics.com/datasets/invasive-ductal-carcinoma-stained-with-fluorescent-cd-3-antibody-1-standard-1-2-0

2.2 Exploratory Data Analysis (EDA) on the spatial transcriptomics samples

After the quality control on the spatial transcriptomics samples, the read-counts were examined to identify of missing values, and summary statistics, including Mean, Median, Min, Max, 1st quartile, and 3rd quartile, were computed for each gene and spot across all spatial transcriptomics samples. This was done to assess data quality, identify disparities, ensure quality control, detect outliers, and understand the overall data distribution.

2.3 Consensus Independent Component Analysis (c-ICA)

To separate the bulk transcriptome profiles into statistically independent TCs, the statistical technique c-ICA was performed in a prior study conducted by our research group, according to a previously outlined methodology [38]. The bulk transcriptional profiles contained measurements for p genes (mixed multivariate signals) and the resulting TCs (constituent source signals) exhibit non-Gaussian distributions. Briefly, i TCs of dimension 1×p are extracted when c-ICA is applied to a gene expression dataset consisting of p genes and n samples. These TCs capture distinct transcriptional patterns that correspond to underlying characteristics, such as biological processes, and each TC comprises p scalars which denote the direction and magnitude of the underlying biological process's effect on gene expression levels. The c-ICA method comprised the following steps: (a) whitening the bulk transcriptional profiles, (b) ICA on the whitened transcriptional profiles, (c) consensus approach, and (d) normalizing the consensus mixing matrix.

(a) Whitening the bulk transcriptional profiles

Prior to performing ICA, it was necessary to whiten the bulk transcriptional profiles. Therefore, matrix $\boldsymbol{X}$ (which contained measurements for p genes in columns and bulk transcriptional profiles in rows) was converted into a new matrix $\boldsymbol{X}_{whitened}$. This transformation enhanced the statistical features of the whitened transcriptional profiles (unit variance, zero covariance, orthogonality). The pre-processing step of whitening is essential to improve the effectiveness of ICA, reduce dimensionality, and minimize noise.

(b) Applying ICA on the whitened transcriptional profiles

FastICA is a fixed-point and iterative algorithm used in ICA, explained by Hyvarinen (1999) [45]. Briefly, it starts with a random initialization and converges to a fixed point where the output of the algorithm becomes stable and does not change notably through more iterations, and finds a solution. Each iteration comprises two steps aimed at maintaining the statistical independence of the estimated source transcriptional patterns: first, the unmixing matrix is calculated using a selected non-linear contrast function and followed by orthogonalization. The iterative procedure continues until a predefined convergence criterion, denoted by a threshold value (Epsilon), is met, which is indicated as a shift in the unmixing matrix between iterations. This criterion guarantees that the algorithm will stop when the unmixing matrix $\boldsymbol{W}$ stabilizes and source transcriptional patterns achieve a sufficient degree of statistical independence.

Consequently, the bulk transcriptional profiles have been decomposed by the FastICA algorithm into their statistically independent source TCs upon convergence.

(c) Consensus approach

The optimization method that fastICA utilizes may converge to local optima, especially in high-dimensional or noisy data spaces, resulting in suboptimal solutions. With various random initializations of the unmixing matrix $\boldsymbol{W}$, these local solutions could change, producing distinct sets of TCs $\boldsymbol{S}_{run}$ for every algorithm run. To address this issue, a consensus approach is employed, which gathers data from several fastICA algorithm runs -each with a unique random initialized unmixing matrix $\boldsymbol{W}$- and determines which TCs were recovered from the various runs. With this method, TCs that are potentially noise artifacts specific to a certain run or arising from convergence to suboptimal local solutions are filtered away. By prioritizing the stable and robust TCs that consistently emerge across multiple runs, the consensus approach enhances reliability and robustness. The resulting consensus mixing matrix, $\boldsymbol{MM}_{consensus}$, captures the weights that indicate the activity of each c-TC (consensus transcriptional component, $\boldsymbol{S}_{consensus}$) in a bulk transcriptional profile. The $\boldsymbol{MM}_{consensus}$ is derived by multiplying the bulk transcriptional profiles (matrix $\boldsymbol{X}$) by the pseudo-inverse of the c-TCs ($\boldsymbol{S}_{consensus}$), mathematically represented as $\boldsymbol{MM}_{consensus} = \boldsymbol{X}(\boldsymbol{S}_{consensus})^{-1}$.

(d) Normalizing the consensus mixing matrix

In the analysis of bulk transcriptional profiles, there is significant variability in gene expression levels. This is frequently due to technical factors like platform inconsistencies or batch effects, rather than biological differences. Moreover, the number of genes with high weights within each c-TC may vary, may inducing bias in the calculated $\boldsymbol{MM}_{consensus}$, which could impede direct comparisons. To overcome this, a normalization methodology for the $\boldsymbol{MM}_{consensus}$ was applied.

2.4 Gene Set Enrichment Analysis (GSEA)

To evaluate the effectiveness of the TCs in capturing biological characteristics, gene set enrichment analysis (GSEA) was conducted by employing the in-house tool, AnalyzerTool6. The enrichment analysis was performed on the TCs using 3 gene set databases from Molecular Signatures Database (MSigDB) (MSigDB 2023.2.Hs version) [46]. The data collections used were: Gene Ontology - Biological Processes (GOBP), Reactome, and Positional. Gene sets with less than 10 genes or exceeding 500 genes were omitted from the analysis. The selected threshold was >= 3.0. GSEA was performed using the 2x2 Fisher's Exact Test, which was applied after dividing the genes into two groups based on gene set membership (member of the gene set vs non-member) and the scalar value in the TC (absolute value of scalar >= 3 vs < 3). Fisher's exact p-values were Z-transformed to facilitate comparison of gene sets with different sizes.

2.5 Identification of CNA-TCs & Immune-TCs

To identify TCs capturing immune-related characteristics, we utilized the GOBP and Reactome databases in MSigDB. Initially, z-scores in the datasets that were generated from GSEA were transformed into p-values, followed by Bonferroni correction to perform multiple testing. Next, significance threshold of alpha=0.05 was applied to identify a subset of TCs exhibiting bio-enriched gene sets. Subsequently, we narrowed them down by selecting TCs associated with immune-related pathways. To achieve this, we utilized gene sets derived from a previous study conducted within our group, where gene sets of immune-related biological pathways were defined by filtering all nodes in the GOBP which contained terms related to the immune system process (GO:0002376) as well as the REACTOME immune system pathway (R-HSA-168256). Through this approach, we ensured that only TCs specifically related to immune processes were included for further analysis. Finally, immune-enriched TCs were prioritized by rank and filtered to retain TCs with a rank of 1, indicating the highest enrichment of immune-related gene sets. TCs meeting these criteria were designated as immune-TCs, representing TCs strongly linked with immune processes.

Likewise, to identify TCs associated with CNA processes, we utilized the Positional database following the same methodology outlined earlier. However, unlike the immune-related analysis, a specific gene set for subsetting was unavailable. Consequently, all TCs containing gene sets with p-values below the cutoff of 0.05 were selected for further analysis, representing TCs enriched for biological pathways. Next, TCs with statistically significant altered cytobands (CNAs) were retained. However, this approach resulted in excluding TCs that might capture CNA-related processes, as even one cytoband may contain CNAs in multiple genes. Therefore, we utilized the Transcriptional Adaptation to CNA profiling (TACNA profiling) technique [38] to identify the TCs that capture CNA-related pathways, which we termed CNA-TCs. Developed by Bhattacharya et al. [38], the TACNA method is a computational tool capable of assessing the influence of CNAs on gene expression patterns from gene expression profiles independently, without requiring paired CNA profiles. Thus, employing this approach, we identified CNA-TCs by selecting those containing more than 10 genes with CNAs.

2.6 Projection of CNA-TCs & Immune-TCs onto spatial transcriptomics samples

For the evaluation of the "activity" of CNA- and immune- TCs identified within the transcriptomic profiles of the spatial transcriptomics data, the in-house tool AnalyzerTool7 was employed. This is a Projection methodology through which the CNA- and immune- TC matrices were projected onto the spatial transcriptomics profiles of the spatial transcriptomics cancer samples (manuscript under preparation). Through this process, we acquired the projected mixing matrices (MM) from each dataset, containing the activity scores of each TC for all spots in every sample. In this method, the transcriptomic profiles of dataset A were projected onto the TCs matrix of dataset B, which derived from c-ICA analysis. Initially, the pseudo-inverse matrix after c-ICA analysis of dataset B was calculated (ref) and then we multiplied this matrix with the transcriptomic profiles of dataset A, focusing only on the shared genes between A and

B datasets, which resulted in the generation of the projected mixing matrix from dataset A. For every transcriptome profile in dataset A, the TC activity in this projected mixing matrix matched the TCs from dataset B.

Nevertheless, a significant portion of the observed variation in the expression levels may stem from platform or batch effects rather than biological differences, which also affects the mixing matrix in the projection process, making it challenging to compare samples across different platforms and experiments. Therefore, Johnson transformation method was employed to correct the mixing matrix values. Initially, the skewness of gene weights within the TCs was evaluated, and due to negative values, all signs of the gene weights and their corresponding activity in the mixing matrix were flipped. Next, the following process was applied to all profile-TC activities in the Mixing Matrix (MM):

1. Genes within each TC were randomly permuted for specific times, and the MM activity for the profile was calculated. These permutations generated the null distribution associated with the profile MM activity for that TC.
2. The Johnson transformation was used to change the null distribution into a normal distribution in case the Anderson-Darling test revealed that it did not follow a normal distribution. Using one of the three families of distributions (S, SU, or SL), this transformation determines the parameters that minimize the null distribution's deviation from a normal distribution. This transformation was also implemented on the initial mixing matrix.
3. The normally distributed null distribution was standardized to mean and standard deviation of zero and one, respectively. This standardization was also applied on the profile's activity value.
4. Finally, the corrected TC activity was derived by computing the z-value using the transformed null distribution and the computed mixing matrix activity, the final corrected TC activity was generated.

2.7 Colocalization of TC activities in spatial transcriptomics samples

The colocalization technique relies on principles from microscopy image analysis and is adapted from a previously published colocalization analysis [47]. Using the permutation-corrected mixing matrix, regions of low or high activity were identified by selecting locations where the TCs activity was above 2 for high activity TC processes and below -2 for low activity TC processes. Next, kernel density estimation was separately applied to regions of low and high TC activities for each image and TC using the R package ks 1.14.1. TCs with a minimum of 50 spots showing low or high activity were included. Using the least-squares cross-validation bandwidth selector with the "Hlscv.diag" method and an initial bandwidth matrix of [(9,0),(0,9)], the ideal kernel bandwidth was selected. Moreover, the size of the binned background grid (bgridsize) is set to match the dimensions of the spots in the spatial transcriptomic image. Then, the "kde" function was then used to fit the kernel, and the weights used were the absolute TC activity (above 2) for each spot. Subsequently, only regions where the fitted kernel density was higher than the 75th percentile of all densities for a TC within an

image were selected for further analysis. The similarity between kernel densities of all TCs was evaluated by computing Pearson correlation coefficients (ρ). If two TCs' up- or downregulated regions exhibited a ρ of 1, they displayed colocalization, whereas they were spatially exclusive when showing a ρ of -1.

In our analysis we took into account both positive and negative associations between CNA- and immune- TCs. For the quantification of these relationships, we used a colocalization score between TCs α and b and their regions with high activity and low activity, following the formula below:

$$\text{colocalization score}_{a,b} = (\rho_{a\_high\ activity,\ b\_\ high\ activity} - \rho_{a\_\ high\ activity,\ b\_\ low\ activity}) - (\rho_{a\_\ low\ activity,\ b\_\ high\ activity} - \rho_{a\_\ low\ activity,\ b\_\ low\ activity})$$

A colocalization score of 4 suggests that there is a strong spatial overlap between high CNA-TC and high immune-TC activity, while a score of -4 demonstrates spatial overlap, but with opposite activity patterns, meaning high immune-TC activity is found where CNA-TC activity is low, and vice versa. A score near 0 indicates that the activities of CNA- and immune-TCs are mostly spatially distinct and these TCs are not associated.

3. Results

3.1 Data acquisition

Bulk transcriptomic profiles derived from tumor biopsies from the public repository TCGA (10,817 profiles, 20,392 genes, generated with RNA-Seq) were collected [41]. The transcriptomic profiles assembly covered a wide range of experimental conditions, tissue types, disease statuses, and chemotherapy responses. This strategy increases the likelihood of identifying transcriptional patterns associated with immune-related activities and CNA footprints across different biological contexts and clinical settings. A bulk transcriptomic profile combines the transcriptional patterns of all cell types in a tissue, reflecting its biological processes. Therefore, subtle patterns can be masked by more prominent ones, making them more difficult to identify. To address this, our research group previously applied c-ICA to the TCGA dataset, separating bulk transcriptome profiles into 9,709 statistically independent TCs, where each gene in a TC had a specific weight and sign, indicating the degree to which the TC's underlying biological process affects the gene's expression level and in which direction. This method isolated both subtle and prominent transcriptional patterns.

3.2 Identification of 235 immune-TCs & 476 CNA-TCs

To identify the TCs that capture the transcriptional effects of immune-related processes, we conducted GSEA. Though this approach we identified 119 immune-TCs to be significantly enriched in 411 immune-related gene sets in the GOBP collection, as well as 158 immune-TCs and 120 gene sets in the Reactome collection. Among these, 41 immune-TCs were common to both Reactome and GOBP datasets, resulting in a total of 235 unique immune-TCs. Regarding the identification of CNA-TCs, we initially analyzed the GSEA results from the Positional database. However, this method was rejected because it excluded TCs related to CNA processes, as even one cytoband can contain CNAs in multiple genes. Instead, the TACNA method was employed, which successfully captured all TCs related to CNA processes. We identified 476 CNA-TCs, each containing more than 10 genes within a genomic region affected by CNAs. The total number of CNA-TCs and immune-TCs was 711, with 16 TCs being common to both categories, indicating that these TCs capture both transcriptional effects influenced by CNAs and immune system processes. This resulted in a total of 695 unique CNA- and immune-TCs. Subsequently, we generated heatmaps to visualize the identified TCs and their enrichment scores, represented by z-scores, in the gene sets (Fig. 4). Hierarchical

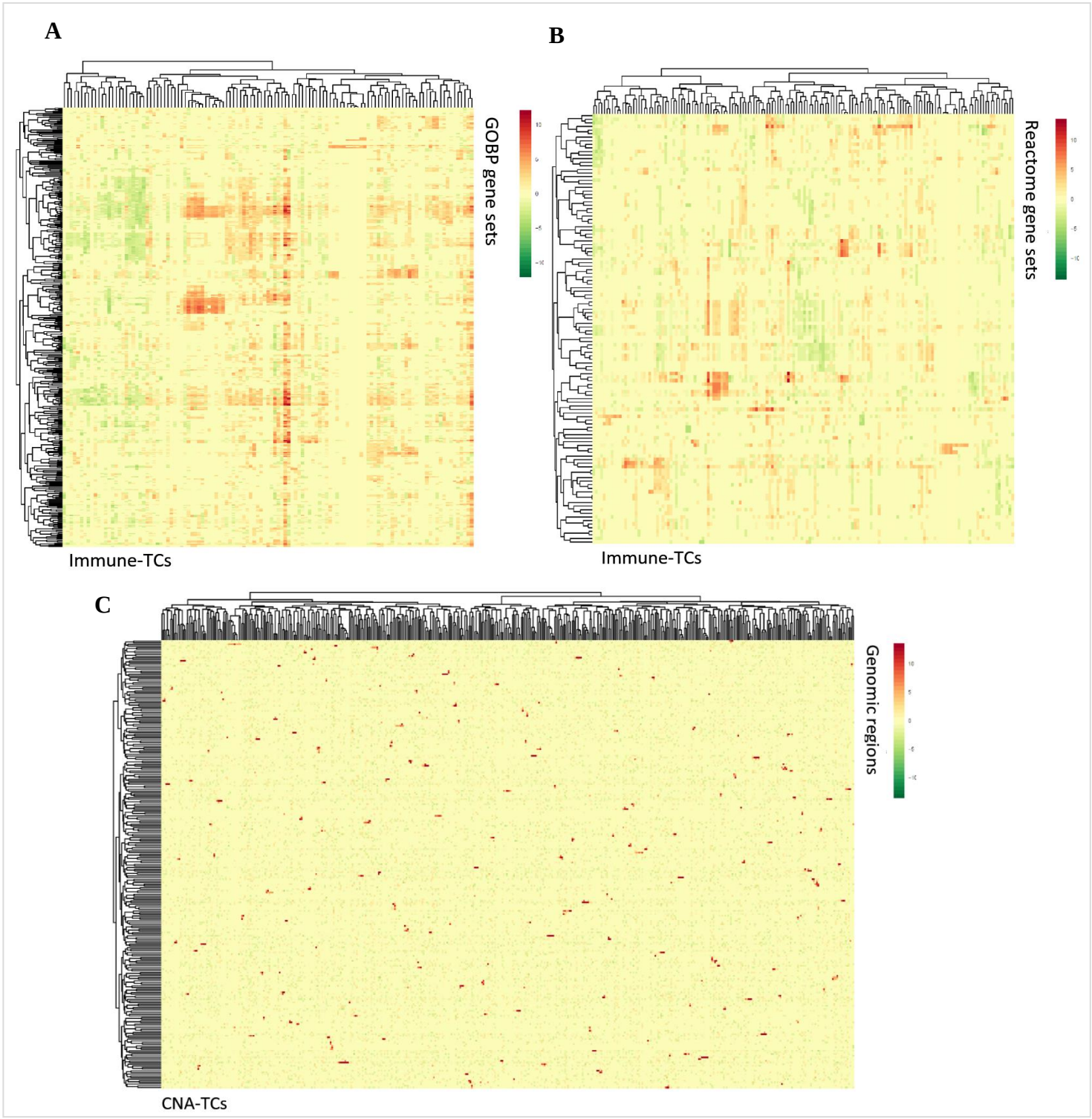

Figure 4: Identification of CNA-TCs and immune-TCs in an extensive dataset. (A) Heatmap of 119 TCs significantly enriched in 411 immune-related gene sets from the GOBP dataset via GSEA. **(B)** Heatmap of 158 TCs significantly enriched in 120 immune-related gene sets from the Reactome dataset via GSEA. **(C)** Heatmap of 476 CNA-TCs identified via the TACNA method. *TACNA: transcriptional adaptation to CNA profiling; GSEA: gene set enrichment analysis; The z-scores in the heatmaps represent the enrichment scores; Color bar: values above zero indicating positively enriched TCs, and values below zero showing negatively enriched TCs.*

clustering was performed using Pearson correlation (1-cor) as the distance measure and Ward's method (ward.D2) as the clustering method.

In Figures 4a and 4b illustrating the immune-TCs from the GOBP and Reactome collections, we observe a variety of enrichment patterns, indicating that the TCs capturing immune processes are highly positively enriched (red) in some gene sets, while others are negatively enriched (green) or show no significant enrichment (yellow). This distribution highlights the diverse roles of immune-TCs across different immune processes. The overlap of 41 immune-TCs between the Reactome and GOBP collections suggests a robust identification of immune-related TCs, indicating consistency in capturing significant immune processes across different gene set collections. The broad range of 235 unique immune-TCs indicates comprehensive coverage of immune responses, including both activated and suppressed pathways. Moreover, the negatively enriched TCs provides insights into repressed immune mechanisms. Figure 4c, represents TCs capturing transcriptional effects of CNA-related processes in different genomic regions. The relatively uniform distribution of both positively and negatively enriched TCs indicates that CNA processes influence a wide range of transcriptional activities. This suggests a widespread impact of CNAs on gene expression patterns across various chromosomal regions. This highlights the complexity and diversity of transcriptional responses to CNAs, resulting in both upregulation and downregulation of gene expression.

Additionally, we performed genomic mapping of the 476 CNA-TCs using gene annotation data to visualize the precise location of the copy number altered genomic region on the chromosomes (Appendix 2a). In Appendix 2b, the genomic plots of the 25 CNA-TCs containing at least 50 genes with CNAs are shown. These plots visualize the distribution of TC gene weights across the genome, which helps in understanding the genomic distribution and significance of various TCs. Overall, the results demonstrate the efficacy of the integrated strategy of combining c-ICA, GSEA, and TACNA on bulk transcriptomics to identify immune-TCs and CNA-TCs. This intermediate step was crucial for validating the results, ensuring the robustness of the identified TCs, and providing a foundation for further analysis.

3.3 EDA to assess the data quality of cancer samples

Moreover, we performed EDA on the spatial transcriptomics samples to gain insights into data quality and perform necessary preprocessing steps prior to further analysis. We visualized the read counts to display the total expression of all genes across the spots in the 21 spatial transcriptomics samples (Appendix 1). For instance, the breast cancer sample BC3 is represented in Figure 5a, where different regions of the tissue in the hematoxylin and eosin (H&E) staining show significant variance in read counts, potentially due to sequencing difficulties, technical limitations or inherent biological differences in certain areas. Moreover, Fig. 5b shows gene expression activity in various regions, with higher read counts indicating areas of increased transcriptional activity. Some regions display very low read counts compared to others with significantly higher expression (e.g. top right corner compared to bottom left corner). To account for this variance and mitigate platform-specific effects, we performed normalization and removed the PC1. This process generated transcriptomic profiles devoid of

platform-specific artifacts, allowing us to examine gene expression patterns across the entire sample.

Additionally, to assess any disparities in the data, ensure quality control, detect outliers, and understand the overall data distribution, we examined the read-counts for missing values and computed summary statistics - mean, median, min, max, 1st quartile, and 3rd quartile- for each gene and spot across all 21 spatial transcriptomics samples (Appendix 1). Subsequently, we generated two types of histograms to visualize these values: spot-wise and gene-wise. Values of zero were excluded from the histograms to focus on the distribution of actively expressed genes and to avoid skewing the data representation. For instance, the spot-wise histograms of the sample BC3 (Fig. 5c) show a concentration of low mean and maximum values, indicating that the majority of spots have relatively low gene expression levels, with only a few spots showing higher expression. This could indicate that certain regions of the tissue have higher transcriptional activity, possibly corresponding to areas with higher cellular density or specific tissue types. Gene-wise histograms (Fig. 5d) include a histogram of minimum values, which reveals a high bar at 1 and two lower bars at 2 and 3, suggesting that most genes are detected in at least one spot, with some genes detected in two or three spots. All histograms display right-skewed distributions with values close to zero and minimal variation. This indicates that the majority of genes are expressed at low levels across the sample, with only a few genes showing high average expression levels despite normalization.

Furthermore, we plotted the expression of genes of interest to examine potential patterns in a spatial context. In the sample BC3, we visualized the expression of the *ERBB2* (also known as *HER2* or *neu*) gene [48] due to its extensive documentation in cancer biology and its association with tumor progression. *ERBB2* is reported to be involved in a number of human cancers. About 30% of human breast cancers [49] and many other cancer types, such as ovarian [49], stomach [50], bladder [51], salivary [52], and lung carcinomas [53], show amplification or overexpression of *ERBB2* gene. Numerous studies have demonstrated that this amplification or overexpression impairs normal cellular regulatory mechanisms, leading to the development of aggressive tumor cells [54]. In Figure 5e, higher expression levels of *ERBB2* are observed in the top right region, where most of the read-counts show most decreased expression levels. However, these spots showing the highest expression are among the lowest ERBB2 expressed

spots. This suggests a localized overexpression of *ERBB2* in specific regions despite overall

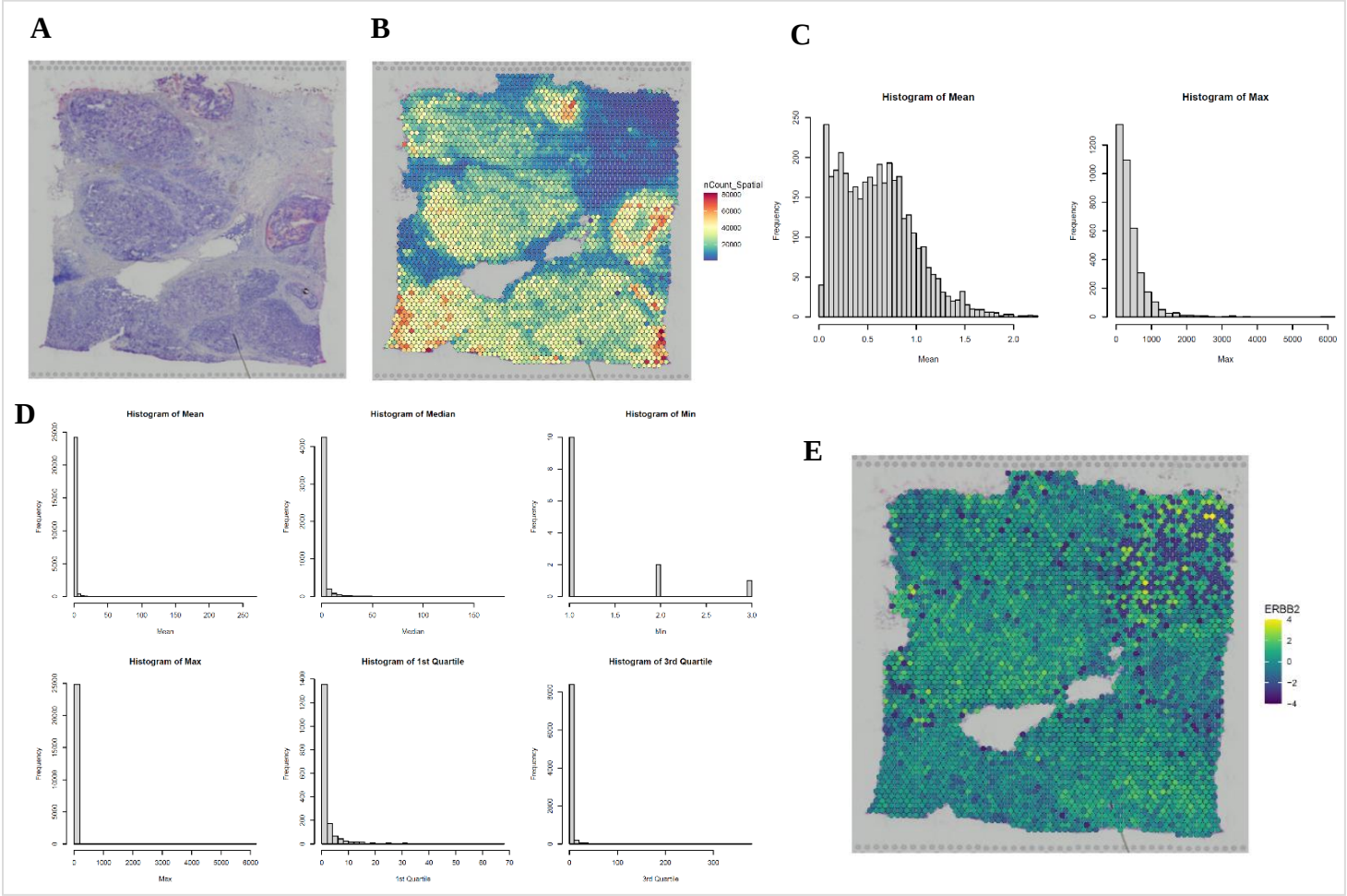

Figure 5: Exploratory data analysis on spatial transcriptomics sample BC3. (A) H&E staining of BC3. **(B)** Visualization of read counts across all spots in BC3, displaying the total expression of all genes. *Color bar: values represent raw expression counts, with red indicating high expression and blue indicating low expression.* **(C)** Summary statistics for spots excluding values of zero. **(D)** Summary statistics for genes excluding values of zero. **(E)** Visualization of ERBB2 gene expression levels in BC3. *Color bar: values above zero show high gene expression, while values lower than zero indicate low gene expression.*

low read-counts in the area. Additionally, in the bottom left and right corners, where the read counts show the highest expression, ERBB2 expression levels are close to zero or zero. Across the sample, *ERBB2* expression in most spots is consistently around a value of 2, indicating a relatively uniform expression level with exceptions in some regions. This pattern may reflect the heterogeneous nature of *ERBB2* expression within the tissue, however, further research is needed to understand the underlying biological pathways.

3.4 Colocalization of TC activities on spatial transcriptomics samples

The primary objective of this research project was to assess the relationship between CNAs and immune transcriptional footprints in 21 spatial transcriptomics cancer samples. To achieve this, we applied the Projection methodology on the previously identified TCs to acquire the mixing matrices of the TCs, which contained their activity scores for all spots in every spatial transcriptomics sample. Using the activity scores of the 25 most robust CNA-TCs, each

containing at least 50 genes with CNAs, we plotted them on spatial transcriptomics samples (Appendix 3a). This visualization aimed to elucidate the spatial distribution and activity of these significant TCs within different tissue samples. Understanding how these CNAs are spatially organized may help investigate their impact on tissue architecture and function. Mapping these components can provide insights into the spatial heterogeneity of CNAs and their roles in disease progression and tissue-specific gene expression patterns. Then, we conducted Colocalization analysis to obtain the correlation scores of all TCs with each other, evaluating whether specific patterns of high or low CNA-TC and immune-TC activities were reflected in the spatial distribution across different cancers. This analysis helped identify both positive and inverse colocalization between CNA-TCs and immune-TCs, providing insights into their spatial distribution within various cancer types.

3.4.1 Correlation analysis reveals patterns of CNA-TCs and immune-TCs in cancer samples

To investigate the patterns of CNA-TCs and immune-TCs in the spatial transcriptomics cancer samples, we generated two types of heatmaps per sample based on their correlation values (Appendices 3b and 3c). Hierarchical clustering was performed using Pearson correlation (1-cor) as the distance measure and Ward's method (ward.D2) as the clustering method. Correlation values of zero were removed to better visualize the clustering of TCs.

Figures 6 and 7 present examples of these heatmaps. In Figure 6a, all CNA-TCs are listed on both the x and y axes, with chromosome annotations distinguished by different colors. The CRC1 sample exhibits numerous and prominent clusters, indicated by regions with TCs that are either highly positively colocalized (red) or highly inversely colocalized (blue). This suggests that CRC1 may contain many clones characterized by TCs with CNAs in different chromosomes. For instance, a specific cluster in Fig. 6a shows TCs which are highly positively colocalized, capturing biological processes related to CNAs in chromosome 19. Some of the CNA-TCs identified in this cluster include TC8759, TC8693, TC7827, and TC7698 with their genomic plots displayed in Figure 6b, verifying their location in chromosome 19. The clustering of focal CNA-TCs on the same chromosome points toward a broader genomic region might be

implicated, possibly indicating a broader CNA affecting a significant portion of the

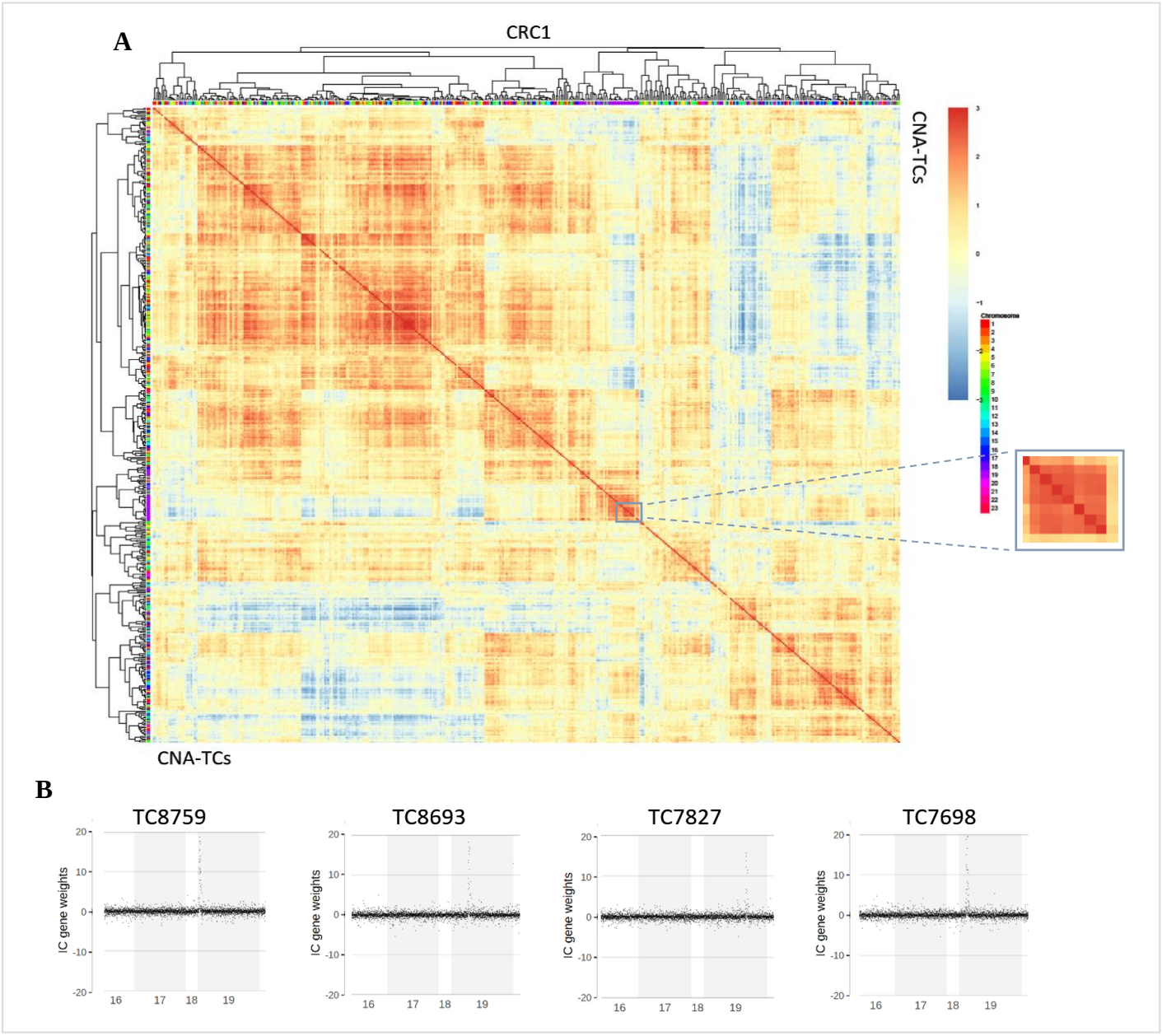

Figure 6: Clustering of CNA-TCs on the spatial transcriptomics sample CRC1. (A) Heatmap of CNA-TCs on the CRC1 sample with chromosome annotations. *Color bar: values above zero indicate positive colocalization (red), while values lower than zero indicate inverse colocalization (blue).* **(B)** Genomic plots of clustered CNA-TCs (TC8759, TC8693, TC7827, TC7698) on chromosome 19, indicating broader CNAs. *x axis: chromosomes, y axis: gene weights.*

chromosome (either deletion or amplification), such as an entire arm of chromosome 19. This highlights the importance of examining these clusters to understand the genomic alterations and their implications in cancer progression.

Figure 7 illustrates the clustering of immune-TCs in the OC2 sample. Notably, a small cluster of TCs, which are positively colocalized (dark red), includes the following TCs and their top gene sets, each characterized by a rank of 1, indicating that these TCs were most enriched in these pathways: (a) TC4962: antigen presentation, folding, assembly, and peptide loading of class I MHC (found in the Reactome dataset), (b) TC8928:

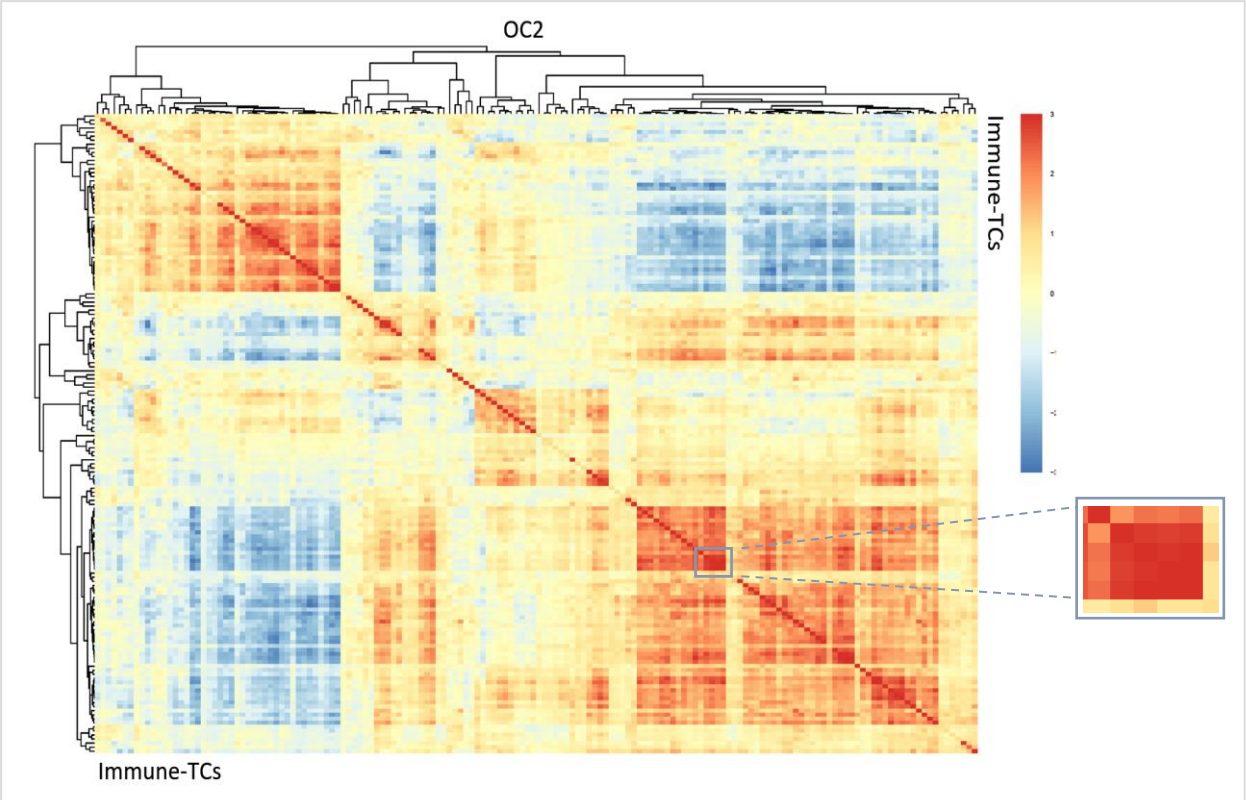

Figure 7: Clustering of immune-TCs on the spatial transcriptomics sample OC2. *Color bar: values above zero indicate positive colocalization (red), while values lower than zero indicate inverse colocalization (blue).*

negative regulation of activated T cell proliferation (GOBP), (c) TC2717: PD1 signaling (Reactome), and (d) TC641: MHC class II antigen presentation (Reactome). The clustering of these immune-TCs suggests a common biological theme centered on antigen presentation and immune regulation. The positive colocalization of these TCs highlights a coordinated expression pattern of genes involved in immune responses and regulatory pathways. By examining the top gene sets associated with these TCs, we can infer that they play crucial roles in antigen presentation and T cell regulation, contributing to the immune landscape of the OC2 cancer sample. This clustering approach may provide insights into the spatial organization of immune-related transcriptional patterns and their potential impact on the tumor microenvironment.

3.4.2 Identification of the top 12 colocalized TCs

To identify CNA- and immune-related transcriptional patterns that co-occur or spatially repel each other, we examined the highest and lowest correlation values between pairs of TCs. We termed "top colocalized TCs" as those CNA- and immune-TCs showing the highest positive or negative correlation values derived from the Colocalization analysis. This includes both the most positively colocalized TCs and the most inversely colocalized TCs, indicating strong spatial proximity or repulsion, for three distinct groups: CNA-TCs, immune-TCs, and CNA- and immune-TCs.

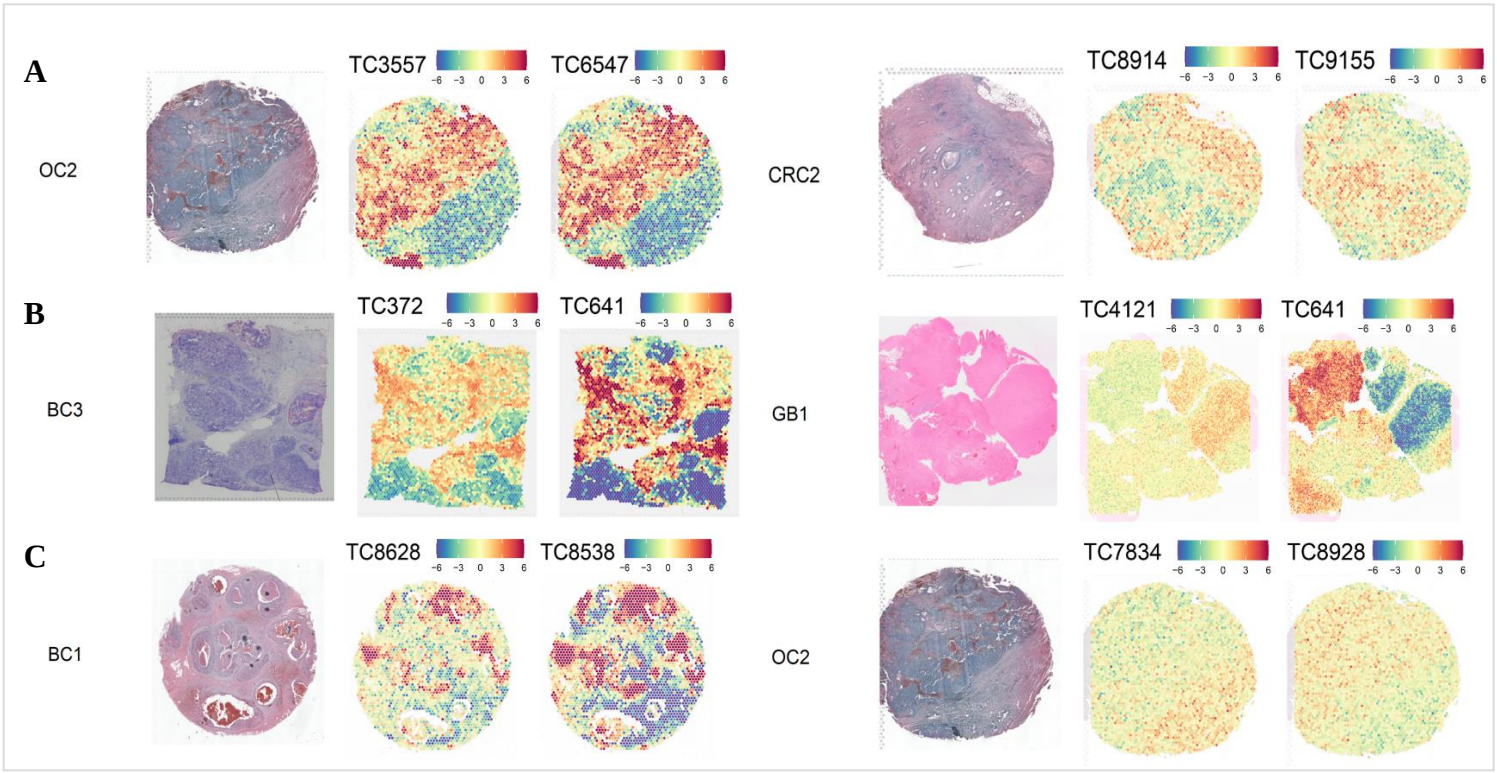

Figure 8: Identification of the top colocalized TCs. (A) Colocalization plots of the top colocalized CNA-TCs. Left: TC3557 and TC6547 in the spatial transcriptomics sample OC2, which show the highest positive colocalization among all cancer samples. Right: TC8914 and TC9155 in the spatial transcriptomics sample CRC2, showing the highest inverse colocalization among all cancer samples. **(B) Colocalization plots of the top colocalized immune-TCs.** Left: TC372 and TC641 in BC3, with the highest positive colocalization. Right: TC4121 and TC641 in GB1, with the highest inverse colocalization. **(C) Colocalization plots of the top colocalized CNA- & immune-TCs.** Left: TC8628 and TC8538 in BC1, with the highest positive colocalization. Right: TC7834 and TC8928 in OC2, with the highest inverse colocalization. *Color bar: an activity score above zero indicates high TC activity, while a score below zero indicates low TC activity.*

In Figure 8, the colocalization plots of the identified top colocalized TCs on the spatial transcriptomics samples are illustrated. The highest positive correlation score between CNA-TCs was found between TC3557 and TC6547 in the cancer sample OC2 (correlation value: +3.64, Fig. 8a left), where both TCs capture the transcriptional effect of CNAs in genomic regions on chromosome 19 (Fig. 9a). Conversely, the lowest correlation value between CNA-TCs was between TC8914 and TC9155 in CRC2 (correlation value: -3.185, Fig. 8a right), indicating an inverse colocalization, with CNAs located on different chromosomes, 8 and 20 respectively (Fig. 9b). Regarding the immune-TCs, the highest positive colocalization was between TC372 and TC641 in BC3 (correlation value: +3.394, Fig. 8b left), showing high

spatial proximity. The most significant inverse colocalization between immune-TCs was found for TC4121 and TC641 in GB1 (correlation value: -3.026, Fig. 8b right). The strongest positive colocalization between CNA-TCs and immune-TCs was observed in BC1, marked by TC8628 and TC8538 (correlation value: +3.203, Fig. 8c left). Notably, TC8628 functioned as both a CNA-TC and an immune-TC exhibiting a dual impact on genomic changed and immune system activity, while TC8538 represented the immune-TC. The highest inverse colocalization by

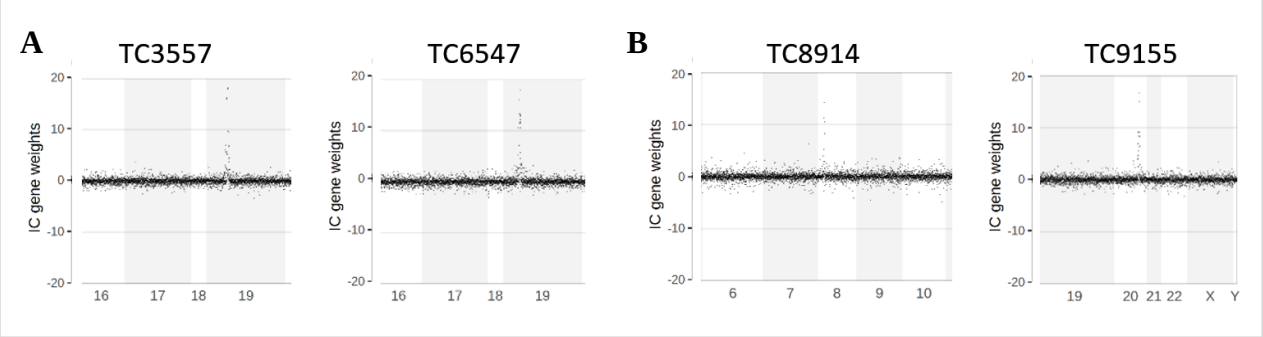

Figure 9: Genomic plots of the top colocalized CNA-TCs. (A) TC3557 and TC6547, which are positively colocalized TCs, show CNAs in genomic regions on chromosome 19. **(B)** TC8914 and TC9155, which are inversely colocalized TCs, exhibit CNAs on chromosome 8 and chromosome 20, respectively.

immune-TCs was observed between TC7834 and TC8928 in OC2 (correlation value: -3.176, Fig. 8c right), with TC7834 representing the CNA-TC and TC8928 the immune-TC. Lastly, Appendix 3d provides an example of TCs with a correlation value of zero, indicating no colocalization and spatial separation, specifically showing the CNA-TCs TC1039 and TC3753 in CRC3.

Table 2 presents the top colocalized immune-TCs and their top pathways, as identified by GSEA, with a rank of 1 indicating the highest level of enrichment in immune-related processes. Moreover, the top genes within each TC were identified based on the highest gene weight, highlighting the genes that most significantly contribute to the observed transcriptional patterns. Additionally, Table 3 lists the top colocalized CNA-TCs along with the chromosomes and cytobands where their CNAs are located. Based on these information we conducted literature search, which is further discussed in the report.

Table 2: Summary of top colocalized immune-TCs. This table lists the top colocalized immune-TCs, the database where they were found, their top associated pathways, corresponding genes, and the chromosomes in which they are located. *Entrez ID: unique, stable and tracked integers as gene identifiers by NCBI* [55].

Immune-TCs	Database	Top pathway	Entrez ID	Gene name	Symbol	Chromosome
TC372	Reactome	PD_1_SIGNALING	3127	major histocompatibility complex, class II, DR beta 5	HLA-DRB5	6
TC641	Reactome	MHC_CLASS_II_ANTIGEN_PRESENTATION	3113	major histocompatibility complex, class II, DP alpha 1	HLA-DPA1	6
TC4121	GOBP	T_CELL_PROLIFERATION	55799	calcium voltage-gated channel auxiliary subunit alpha2delta 3	CACNA2D3	3
TC8538	Reactome	INTERFERON_ALPHA_BETA_SIGNALING	2537	interferon alpha inducible protein 6	IFI6	1
TC8628	Reactome	INTERFERON_ALPHA_BETA_SIGNALING	3433	interferon induced protein with tetratricopeptide repeats 2	IFIT2	10
TC8928	GOBP	NEGATIVE_REGULATION_OF_ACTIVATED_T_CELL_PROLIFERATION	440107	pleckstrin homology and RhoGEF domain containing G7	PLEKHG7	12

Table 3: Summary of top colocalized CNA-TCs.

CNA-TCs	Chromosome	Cytoband	BP_Mapping
TC3557	19	p13.13	12949739.75
TC6547	19	p13.2	10365552
TC8914	8	p21.3	22249866.5
TC9155	20	q13.13	49909012
TC8628	10	q23.31	89305658
TC7834	20	q11.23	36246691.75

3.4.3 Investigation of top colocalized TCs across cancer samples

To investigate whether the top colocalized TCs are positively or inversely colocalized also in other cancer samples, we examined their correlation scores across all spatial transcriptomics cancer samples, as depicted in Table 4.

Table 4: Correlation scores of the top colocalized TCs from the Colocalization analysis of spatial transcriptomics samples. A high positive correlation value indicates positive colocalization of the TCs, whereas a high negative correlation value indicates inverse colocalization.

	BC1	BC2	BC3	BC4	BC5	CC1	CRC1	CRC2	CRC3	GB1	GB2	IDC1	LC1	LC2	M1	OC1	OC2	OC3	PC1	PC2	PC3
TC3557_ TC6547	0.696	0.928	0.603	1.614	1.823	0.922	2.217	0.696	2.184	2.026	2.220	0.667	1.911	-2.651	1.765	2.070	**3.640**	1.922	0.585	2.288	1.353
TC8914 _TC915	-1.297	0.352	-0.117	0.453	-0.642	0.378	-0.391	**-3.185**	-0.086	0.573	0.760	0.984	-1.057	-1.045	0.094	-1.994	-2.560	0.104	-0.449	0.072	-1.393
TC372_ TC641	0.075	3.328	**3.394**	3.348	3.098	0	1.060	0	2.440	0.108	0.892	2.910	-1.538	0.931	-0.177	0.756	2.315	3.285	1.941	0.919	0.7613
TC4121 _TC641	0	0	0	0	0	-0.359	-0.652	0	0	**-3.026**	0	0	-1.650	-1.699	0	0	0.392	-0.693	0.440	0.433	0.3405
TC8628 _TC853	**3.203**	2.553	3.1658	3.172	1.420	1.782	2.344	1.542	2.117	1.632	2.054	2.395	0.795	2.909	2.921	2.929	1.853	1.108	1.3620	1.6128	1.874
TC7834 _TC892	-1.078	-1.057	-1.008	-1.022	0.447	0.201	-0.147	-0.386	-0.481	-2.560	0.922	-0.606	1.089	-1.339	-0.677	1.782	**-3.176**	-0.386	0.327	-1.414	0.836

The top positively colocalized CNA-TCs (TC3557 and TC6547), found in OC2, also show high correlation values in OC1, OC3, GB1, CRC1, CRC3, LC2, and PC2. The top inversely colocalized CNA-TCs (TC8914 and TC9155), found in CRC2, exhibit high negative values in the sample OC2, as well. However, in the rest of the cancer samples, their correlation values are close to zero, indicating no significant spatial association. Additionally, the highest positively colocalized pair of immune-TCs (TC372 and TC641) identified in BC3 also shows high correlation values in BC2, BC4, BC5, CRC3, IDC1, OC2, and OC3. In the remaining samples, their values are close to zero, with zero in CC1 and CRC2 samples. The highest inversely colocalized pair of immune-TCs (TC4121 and TC641) found in GB1 shows very low values in other samples, approximating zero, indicating a lack of colocalization and spatial separation in almost all other cancer samples. Moreover, the top positively colocalized CNA- and immune-TCs (TC8628 and TC8538) identified in BC1 also exhibit high colocalization in BC2, BC3, BC4, CRC1, CRC3, GB2, IDC1, LC2, M1, and OC1. Finally, the top inversely colocalized CNA- and immune-TCs (TC7834 and TC8928) identified in OC2 also exhibit high negative correlation values in GB1. However, in the rest of the cancer samples, these values approach zero, suggesting spatial separation. The consistent high correlation values of top

colocalized TCs across multiple cancer samples suggest that certain transcriptional patterns may be characteristic of specific cancer types or subtypes.

Additionally, we investigated which pair of CNA- and immune-TCs with correlation score above 2 is present in the majority of the cancer samples. We discovered that it was TC8628 and TC8538 in 11 cancer samples: BC1, BC4, BC3, OC1, M1, LC2, BC2, CRC1, IDC1, CRC3, and GB2. These samples are listed in order from the highest to the lowest correlation value for these TCs. Of note, these TCs were also identified earlier as the top colocalized CNA- and immune-TCs (Fig. 8c). Similarly, we examined the pair of CNA- and immune-TCs with correlation score below -2, which was TC8914 and TC9516 in 6 cancer samples: OC2, CRC2, OC1, LC1, PC3, and CRC1, listed from higher to lower correlation scores. Notably, TC8914 was also a hit in the top inversely colocalized CNA-TCs in the sample CRC2 (Fig. 8a). Further research into these findings may provide insights into the spatial heterogeneity of cancer and its potential impact on disease progression and treatment response.

3.4.4 Variability of the correlation scores of TCs across the samples

To assess and compare the distribution of correlation values of TCs within each cancer sample, we generated histograms displaying these values across all samples. These histograms illustrate the frequency of TC colocalization, highlighting the number of TCs that are colocalized (e.g., +3 for most positively colocalized TCs and -3 for most inversely colocalized TCs) versus those that are not (i.e., have zero correlation values). This analysis provides insights into the extent of spatial colocalization of CNA-TCs and immune-TCs within each sample. Appendix 4 contains the histograms of the correlation scores between immune-TCs, and CNA-TCs with immune-TCs in the spatial transcriptomics samples, while Figure 10 illustrates the histograms of the correlation scores between CNA-TCs. In our analysis of 21 spatial transcriptomics samples, we observed notable deviations in the histograms of correlation scores between CNA-TCs, immune-TCs, and CNA-TCs with immune-TCs. Specifically, the GB1 sample exhibited a unique pattern, with significantly fewer non-colocalized TCs across all three histogram types. This indicates a higher number of significantly colocalized TCs in GB1, suggesting a strong association between CNAs and immune transcriptional footprints in this particular sample. However, this trend was not consistent in the GB2 sample, which showed a high number of TCs with zero correlation values, especially in the immune-immune and CNA-immune histograms, contradicting the observations in GB1. This observation might point to issues with the transcriptomics data quality in GB1 from the 10x Genomics platform. Additionally, CNA-TCs generally displayed a higher frequency of non-colocalized TCs compared to immune-TCs. However, CNA-TCs also exhibited a higher frequency of colocalized TCs, which may be attributed to their larger total number (476) compared to immune-TCs (235). The highest number of zero correlation values for TCs was observed in specific samples: BC1 and CRC3 for CNA-CNA correlations, CC1 and CRC3 for CNA-immune correlations, and CC1, CRC3, GB2, OC3, PC1, and PC2 for immune-immune correlations. The patterns observed, particularly in CRC3, suggest weaker associations between CNAs and immune footprints in these samples.

These findings highlight the variability in the association between CNAs and immune transcriptional footprints across different cancer types and samples.

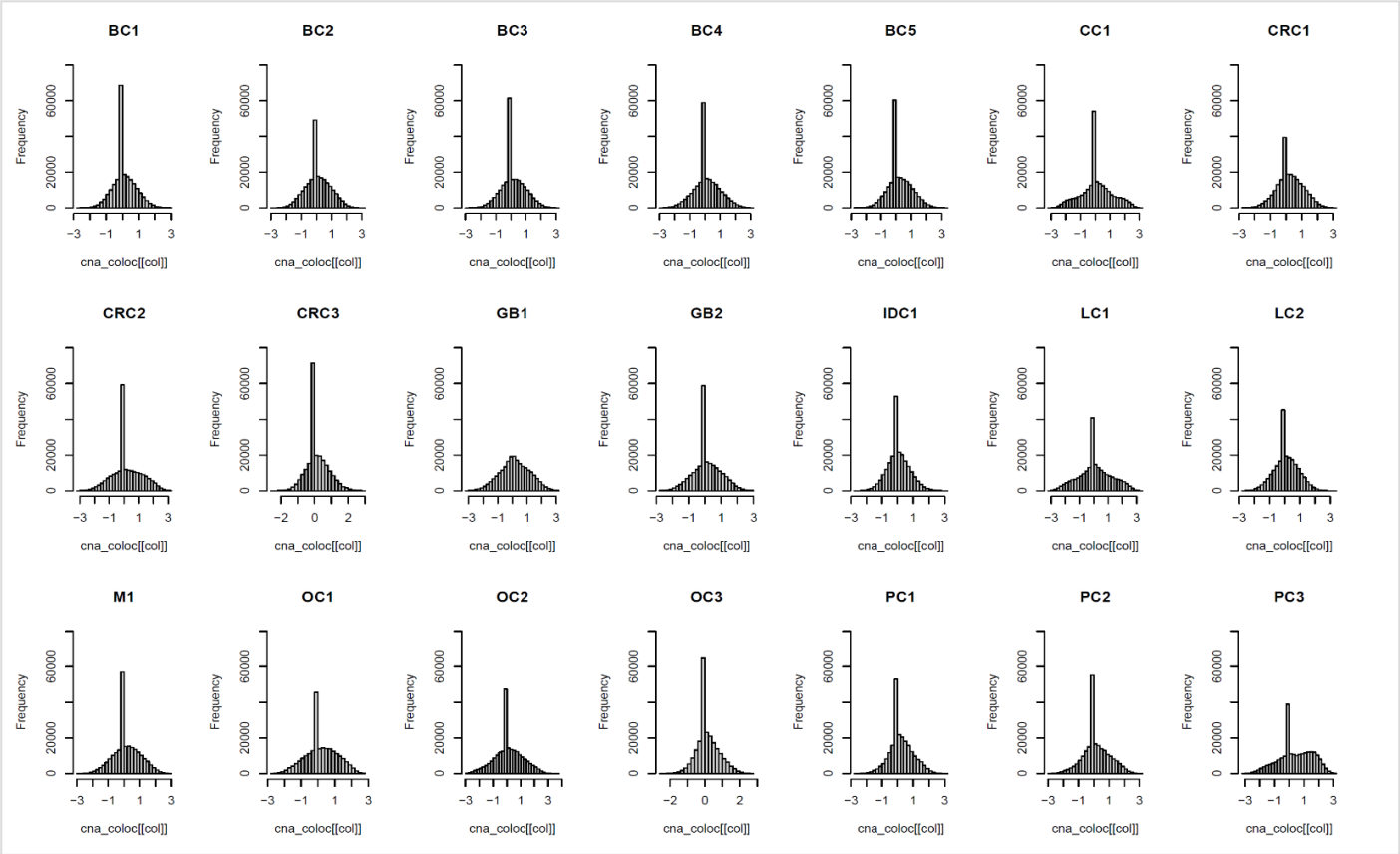

Figure 10: Histograms of the correlation scores between CNA-TCs in the spatial transcriptomics samples. *x axis: correlation values, y axis: total amount of TCs (frequency), +3: most positively colocalized TCs, -3: most inversely colocalized TCs, zero correlation value: not colocalized TCs.*

4. Discussion

In this research project, our main objective was to uncover new insights into the association between CNAs and the immune transcriptional footprints using spatial transcriptomics data in a pan-cancer setting. We addressed this by applying c-ICA on bulk transcriptomic profiles from the public repository TCGA to retrieve TCs capturing biological processes, which were then analyzed using GSEA. Through GSEA analysis and the TACNA methodology, we identified 476 CNA-TCs and 235 immune-TCs for further investigation. Using the Projection method, we obtained mixing matrices containing the activity scores of these TCs, which were subsequently used in the Colocalization analysis. By examining the correlation scores of the CNA- and immune-TCs we identified the top 6 colocalized pairs of TCs in the spatial transcriptomics samples. Finally, we conducted a literature search on these top 12 colocalized TCs to elucidate the underlying biological transcriptional patterns, and shed light on how CNAs influence the cancer-immune set point.

During this study, utilizing spatial transcriptomics tools, we successfully generated a comprehensive landscape encompassing the transcriptional effects of CNAs and immune processes in 21 spatial transcriptomics cancer samples. These data are available on a public website, which could serve as a valuable resource for further research. Using the correlation scores of the TCs, we generated two types of heatmaps illustrating the clustering of the CNA- and immune-TCs for each transcriptomic sample, revealing patterns between TCs (Appendices 3b and 3c). However, due to time constraints and the magnitude of the data during this project, these results could not be investigated in depth; nevertheless, they may provide valuable information for future research. Furthermore, a series of scripts has been developed for data analysis. These scripts facilitate the visualization of genes and TCs within these samples and provide tools for evaluating their colocalization, among other functionalities.

Top colocalized CNA-TCs

Literature search for the top positively colocalized CNA-TCs (TC3557 and TC6547, which contain CNAs located in the genomic regions 19p13.13 and 19p13.2 respectively) in the OC2 sample revealed consistency with previous data. Approximately 20% of ovarian cancers exhibit alterations in 19p13, often involving chromosomal translocations with material from chromosome 11 [56] [57] [58] [59] [60]. Specifically, in the study by L. Wang et al. [61], fusions between 11q13.2 and 19p13.2 were verified in 45% of primary ovarian cancer cell cultures. Moreover, in the study by Shih et al. [62], analysis of high-grade serous carcinomas from the TCGA database revealed elevated DNA copy numbers at the ch19p13.2 locus in 18% of samples. Of note, seven genes in this region showed significant DNA copy number and RNA expression correlations, ranking among the top 100 potential 'driver' genes [62]. Additionally, 54 of these top 100 genes were located in chromosome 19 subregions, indicating frequent structural rearrangements [62]. The positive colocalization of these TCs implies a potential interaction where specific genes within these genomic regions may mutually influence their expression patterns in a spatial context. The spatial proximity of these TCs may facilitate physical interactions between regulatory elements, such as enhancers or transcription factors,

within these genomic regions. This proximity could enhance the coordinated expression of genes involved in similar biological processes that are present especially in the phenotype of ovarian cancer.

Literature search was also conducted for the top inversely colocalized CNA-TCs, TC8914 and TC9155, which contain CNAs in the genomic regions 8p21.3 and 20q13.13, respectively, identified in the CRC2 sample. Studies have shown that chromosomal losses on the short arm of chromosome 8 are implicated in a wide range of colorectal cancers (CRC), particularly on the short arm with loss of heterozygosity (LOH) occurring at rates between 30% and 50%, particularly in regions 8p23.1-pter and 8p21, suggesting the presence of tumor suppressor genes [63] [64] [65] [66]. Moreover, LOH on 8p is associated with advanced and aggressive CRC, with minimal deletion regions at 8p22 and 8p21 linked to clinically aggressive disease markers [65] [67] [68]. Conversely, frequent amplifications on the chromosomal region 20q13 are observed in CRC [69]. An amplicon at 20q11–20q13 was detected in over 70% of CRC samples, including several elevated cancer-related genes, such as AHCY, POFUT1, and RPN2 [70]. TPX2 (20q11.21) and AURKA (20q13.2) were identified as oncogenes driving this amplicon [71]. Moreover, Aust et al. [72] found 20q13.2 amplification in 53% of CRC cases, correlating with worse overall survival, faster tumor progression, and association with left-sided colon tumors and lower histologic grade. Low-level 20q gains are common in breast, bladder, colon, and ovarian cancers, with high-level amplifications of 20q13.2 being less frequent [73]. Bui et al. [74] discovered a gain in 20q13.33 in 50.9% of CRC cases and 62.8% of colon polyps, suggesting its potential as an early detection biomarker and a therapy target. Overall, our findings align with previous studies, highlighting the relevance of these regions in CRC. The inverse correlation between deletions on 8p21.3 and amplifications on 20q13.13 suggests potential roles in cancer progression, possibly through spatially exclusive biological pathways. For example, tumor suppressor genes on 8p21.3 may counteract oncogenic processes driven by amplified genes on 20q13.13. This inverse relationship may also involve synthetic lethality, where the deletion of 8p21.3 genes creates a cellular environment requiring 20q13.13 amplification for survival. Targeting these regions therapeutically could exploit synthetic lethality to eliminate cancer cells effectively.

Top colocalized immune-TCs

TC372 and TC641 were found to be the top positively colocalized immune-TCs, found in the BC3 sample. In TC372, the *major histocompatibility complex, class II, DR beta 5 (HLA-DRB5)* was identified as the top gene within its respective pathway, while the top gene for TC641 was the *major histocompatibility complex DP alpha 1 (HLA-DPA1). HLA-DRB5* has been linked to predicting gliomas and skin cutaneous melanoma, but is not primarily associated with breast cancer [75] [76]. Interestingly, in the bioinformatics analysis by Wu et al. [77] on TCGA data, *HLA-DRB5* expression was not significantly different between breast cancer and normal samples. Likewise, *HLA-DPA1* has not been reported to associate with breast cancer in the literature. However, our analysis suggests a potential role for both these genes in antigen presentation processes and immune responses in breast cancer. Further research should be done to elucidate their specific associations in this type of cancer.

TC4121 and TC641 were identified as the top inversely colocalized immune-TCs in the GB1 sample. The top genes for TC4121 and TC641 were explored: *calcium channel regulatory subunit α2δ-3 (CACNA2D3*) and *HLA-DPA1*, respectively. In the study by Leone et al. [78], *CACNA2D3* was significantly associated with shorter survival in glioblastoma (GB) patients (median 200 days vs. 450 days, $P < 0.03$) and was found to be underexpressed. Previous research identified *CACNA2D3* as a potential tumor suppressor in esophageal squamous cell carcinoma (ESCC), lung and renal cell carcinomas, where it exhibited deletion, LOH, and reduced expression [79] [80] [81]. Functional analysis further supported *CACNA2D3*'s tumor suppressive role [82]. Jin et al. [83] showed that *CACNA2D3* is often downregulated in gliomas and linked to poor survival, while its overexpression increases intracellular calcium, inducing apoptosis and inhibiting cell proliferation, migration, invasion, and tumor development. Conversely, Yu & Fu [84] identified *CACNA2D3* as one of the six genes highly expressed in GB, conflicting with the previous findings. *CACNA2D3* methylation has been associated with shorter survival in gastric cancer [82], and proposed as a biomarker for metastasis risk in breast cancer [85]. These findings suggest *CACNA2D3* could serve as a potential marker in GB, given its known roles in other cancers, emphasizing its importance in cancer therapy through calcium channel regulation and apoptosis induction. Regarding the *HLA-DPA1* gene, literature search did not yield relevant results concerning its relation to GB. However, *HLA-DPA1* was identified in both the top positively and inversely colocalized pairs of immune-TCs in BC3 and GB1 in our analysis, suggesting its potential involvement in immune evasion mechanisms and the modulation of the tumor immune microenvironment. The inverse colocalization of TC4121 and TC641 suggests that their top genes are expressed in a mutually exclusive manner within the tumor microenvironment. This could indicate that the biological pathways regulated by *CACNA2D3* and *HLA-DPA1* may inhibit each other's processes. For instance, the tumor suppressor activity of *CACNA2D3* might counteract the immune evasion mechanisms facilitated by *HLA-DPA1*, or vice versa. Additionally, this inverse colocalization could imply synthetic lethality, where the loss of *CACNA2D3*'s tumor suppressive function necessitates the presence of *HLA-DPA1*-driven immune evasion for cell survival. Understanding this relationship could provide insights into potential therapeutic strategies, such as targeting these pathways to exploit their mutual exclusivity in cancer treatment.

Top colocalized CNA- and immune-TCs

TC8628 and TC8538 were identified as top positively colocalized CNA- and immune-TCs in the BC1 sample. Of note, this pair of TCs was also found to have a correlation score above 2 in the majority of the spatial transcriptomics samples. TC8628 captures both CNA- and immune-related processes. It contains CNAs in the genomic region 10q23.31 and its top identified gene is *interferon induced protein with tetratricopeptide repeats 2 (IFIT2)*. The top gene for TC8538 is *interferon alpha-inducible protein 6 (IFI6)*. In sporadic breast carcinomas, LOH is common in the 10q23 region, which includes the *PTEN* gene, with the incidence ranging from 10% to 40% [86] [87] [88]. *PTEN* regulates the 1-phosphatidylinositol 3-kinase (PI3K) pathway, which influences cell proliferation and survival [89], and it inhibits tumor growth by down-regulating this pathway, leading to G1 phase arrest and cell death [90] [91] [92]. Reduced PTEN levels are common in breast cancer and linked to worse outcomes such as disease-related death, lymph node metastasis, and loss of estrogen receptor staining, suggesting *PTEN*'s role as a tumor

suppressor [93]. This was further validated by the study of S. H. Chang et al. [94], which found that loss of *PTEN* expression in invasive ductal breast cancer samples was significantly associated with more aggressive cancer characteristics, including lymph node metastasis, and advanced stage of cancer.

IFIT2 is associated with various cancer types, including chronic myeloid leukemia, esophageal cancer, ESCC, oral cancer, and triple-negative breast cancer (TNBC) [95] [96] [97] [98] [99]. Decreased *IFIT2* expression correlates with more aggressive cancer traits, such as increased proliferation, invasion, migration, epithelial to mesenchymal transition, and cancer stem cell-like phenotypes, suggesting its role as a tumor suppressor [96] [98]. Moreover, Koh et al. [99], identified *IFIT2* as a significant factor in treatment resistance, including chemotherapy and radiation, among TNBC patients. Notably, higher *IFIT2* expression levels were associated with better prognosis and higher relapse-free survival. Additionally, this study found that baicalein downregulates *IFIT2* expression, thereby sensitizing resistant TNBC cells to treatment and promoting apoptosis, suggesting that targeting *IFIT2* with baicalein could be a promising strategy to overcome resistance in TNBC therapies.

IFI6 (also known as *G1P3*) plays a significant role in cancer progression, particularly in breast cancer and ESCC. In breast cancer, *IFI6* contributes to metastasis by increasing mitochondrial reactive oxygen species (mtROS), which enhances cell migration and invasion [100]. Suppressing mtROS or knocking down *IFI6* reduces these metastatic behaviors, suggesting that targeting IFI6's mitochondrial functions could improve clinical outcomes [100]. Moreover, Davenport et al. [101] found that IFI6 is localized in the endomembrane system and mitochondria of breast cancer cells, promoting apoptosis resistance and cancer cell survival. In ESCC, *IFI6* is highly expressed and associated with aggressive disease and poor prognosis [102]. Moreover, it plays a role in maintaining redox balance, preventing ROS accumulation and mitochondrial dysfunction; depleting *IFI6* increases ROS levels, leading to apoptosis, highlighting its therapeutic potential [102].

Overall, our findings align with previous data, validating the association of TC8628 and TC8538 with breast cancer. Their positive colocalization suggests a potential synergistic relationship between specific CNA and immune pathways. The dual role of TC8628, identified as both a CNA- and an immune-TC, indicates that the genomic alterations at 10q23.31 might not only drive tumor growth through genetic mechanisms, but also modulate the immune environment to enhance cancer progression. Moreover, the involvement of *PTEN* and *IFIT2* in TC8628 suggests that alterations in this region may impact tumor suppressor pathways and immune responses simultaneously. The loss of *PTEN*, a well-known tumor suppressor, coupled with the immune-modulating effects of *IFIT2*, underscore the multifaceted impact of CNAs in this genomic region. *IFI6's* role in enhancing mtROS and promoting metastasis underscores its contribution to a pro-tumorigenic microenvironment. Future research should focus on elucidating these mechanisms, as it could provide insights into new therapeutic targets. Interventions designed to disrupt these synergistic pathways could potentially lead to better clinical outcomes for breast cancer patients.

Finally, TC7834 and TC8928 were identified as the top inversely colocalized CNA- and immune-TCs in the OC2 sample. TC7834 contains CNAs located in the genomic region 20q11.23, and the top identified gene in TC8928 is *pleckstrin homology and RhoGEF domain*

containing G7 (PLEKHG7). Although the current literature does not provide evidence of their direct involvement in ovarian cancer progression, our analysis suggests a potential association and highlights the need for further investigation. The inverse colocalization of these TCs suggests that they may participate in biological pathways that are spatially exclusive within the tumor immune microenvironment. This implies that the cellular processes regulated by the CNAs in the 20q11.23 region and those influenced by *PLEKHG7* do not coexist in the same tumor regions. This could be due to distinct microenvironmental niches within the tumor, where different selective pressures drive the expression of these genes exclusively. For example, the CNAs in 20q11.23 might be associated with cellular proliferation and survival mechanisms that are not compatible with the immune-related functions regulated by *PLEKHG7*. Understanding these interactions could provide insights into ovarian cancer's heterogeneity and contribute to developing targeted therapies that exploit these biological processes.

Conclusion and future perspectives

This study provides a comprehensive landscape of the association between CNAs and immune transcriptional footprints across diverse cancer samples using spatial transcriptomics data. Our bioinformatics analysis identified numerous colocalized TCs, involving both CNA- and immune-related processes. Through computational methodologies, we gained new insights into the tumor immune microenvironment's complexity and its implications in cancer progression. By investigating the top genes and genomic regions of the top 12 colocalized TCs through literature search, we highlighted their significance in driving tumor-immune interactions and cancer characteristics. However, the precise underlying molecular mechanisms behind these associations between TCs reflecting the transcriptional effect of CNAs or immune processes in a spatial context are not yet known, and further research is needed.

Interestingly, the CNA- and immune-TCs, TC8628 and TC8538, showed strong positive colocalization in the majority of cancer samples, particularly in CRC and breast cancer. These TCs, sharing spatial proximity, with TC8628 being both a CNA- and immune-TC, suggest that CNAs in the genomic region 10q23.31 may modulate immune-related pathways involving *IFIT2* and *IFI6*, contributing to cancer progression and resistance. This association was present in multiple cancer phenotypes, making these TCs promising candidates for further investigation. Future research should focus on elucidating the molecular mechanisms underlying these interactions. Validation experiments, potentially in a wet-lab setting, are necessary to test this hypothesis and could reveal novel therapeutic targets for CRC and breast cancer.

An important limitation of this study was the limited number of spatial transcriptomics samples included. Future research should expand this analysis to incorporate data from additional public repositories that are currently available and include more cancer samples per cancer type. This would enhance our understanding of the transcriptional landscape by incorporating a broader range of tumor immune microenvironments. Notably, the GB1 sample showed discrepancies in the distribution of colocalized and non-colocalized TCs compared to GB2 and the rest of the cancer samples, indicating the need to ensure the data quality of GB1. Additionally, other highly

colocalized CNA- and immune-TCs could be explored, beyond the top 12 identified in this analysis, to uncover novel patterns and insights specific to each cancer type. Furthermore, integrating single-cell analysis could offer much-needed information about the specific cell types present in each spot across the spatial transcriptomics samples. This could enable a more holistic approach to unravel the heterogeneity of tumor-immune cell populations and elucidate their dynamic interactions within spatially resolved contexts.

Regarding computational challenges, the frequent updates to the Seurat package present a challenge, as its properties often change during the spatial transcriptomics analysis. Therefore, it must be installed and handled carefully to ensure accurate and reliable results. To address this, the development of harmonized versions would be beneficial, as it would streamline the workflow, reduce potential errors, and enhance the reproducibility and reliability of the results.

In conclusion, this research project demonstrates the significance of spatial transcriptomics and offers a comprehensive framework for investigating the underlying mechanisms of the transcriptional effects of CNAs and their associations with immune footprints in a big pool of cancer samples. This approach could provide new insights into their interactions, the spatial organization, the anti-cancer immune response in different cancer types, and inform the development of innovative, targeted cancer therapies.

Data availability

Pre-processed and standardized level 3 RNA-Seq (version 2) data for 21 cancer samples from the TCGA public data repository were collected through the Broad GDAC Firehose portal (https://gdac.broadinstitute.org/) [41]. The spatial transcriptomics data were collected from the 10x Genomics public database and they are available at https://www.10xgenomics.com/research-areas/cancer/visium-discovery-hub. Data generated during this research project are available in the website https://sites.google.com/student.rug.nl/research-project2-s5340411.

Code availability

R version 4.3.3 was used for the development of the code. The scripts required to reproduce the analysis in this research project are available on GitHub at https://github.com/IroAnagn/Research-project-II-in-MSc-Biomedical-Sciences.

5. Bibliography

[1] W. Yin, J. Wang, L. Jiang, and Y. James Kang, "Cancer and stem cells," *Exp. Biol. Med.*, vol. 246, no. 16, pp. 1791–1801, 2021.

[2] D. A. Lukow and J. M. Sheltzer, "Chromosomal instability and aneuploidy as causes of cancer drug resistance," *Trends in Cancer*, vol. 8, no. 1, pp. 43–53, 2022.

[3] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 144, no. 5, pp. 646–674, 2011.

[4] L. Sansregret, B. Vanhaesebroeck, and C. Swanton, "Determinants and clinical implications of chromosomal instability in cancer.," *Nat. Rev. Clin. Oncol.*, vol. 15, no. 3, pp. 139–150, Mar. 2018.

[5] L. Bassaganyas *et al.*, "Copy-number alteration burden differentially impacts immune profiles and molecular features of hepatocellular carcinoma," *Clin. Cancer Res.*, vol. 26, no. 23, pp. 6350–6361, 2020.

[6] A. Vasudevan, K. M. Schukken, E. L. Sausville, V. Girish, O. A. Adebambo, and J. M. Sheltzer, "Aneuploidy as a promoter and suppressor of malignant growth," *Nat. Rev. Cancer*, vol. 21, no. 2, pp. 89–103, 2021.

[7] E. S. Tan *et al.*, "Copy Number Alterations as Novel Biomarkers and Therapeutic Targets in Colorectal Cancer," *Cancers (Basel).*, vol. 14, no. 9, pp. 1–17, 2022.

[8] N. Crosetto *et al.*, "Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing.," *Nat. Methods*, vol. 10, no. 4, pp. 361–365, Apr. 2013.

[9] M. Baudis, "Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data," *BMC Cancer*, vol. 7, no. 1, p. 226, 2007.

[10] R. Beroukhim *et al.*, "The landscape of somatic copy-number alteration across human cancers.," *Nature*, vol. 463, no. 7283, pp. 899–905, Feb. 2010.

[11] T. I. Zack *et al.*, "Pan-cancer patterns of somatic copy number alteration.," *Nat. Genet.*, vol. 45, no. 10, pp. 1134–1140, Oct. 2013.

[12] B. Gao and M. Baudis, "Signatures of Discriminative Copy Number Aberrations in 31 Cancer Subtypes," *Front. Genet.*, vol. 12, no. May, 2021.

[13] R. J. Leary *et al.*, "Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing.," *Sci. Transl. Med.*, vol. 4, no. 162, p. 162ra154, Nov. 2012.

[14] S.-J. Dawson *et al.*, "Analysis of circulating tumor DNA to monitor metastatic breast cancer.," *N. Engl. J. Med.*, vol. 368, no. 13, pp. 1199–1209, Mar. 2013.

[15] E. Heitzer *et al.*, "Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing.," *Genome Med.*, vol. 5, no. 4, p. 30, 2013.

[16] M. B. Upender *et al.*, "Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells.," *Cancer Res.*, vol. 64, no. 19, pp. 6941–6949, Oct. 2004.

[17] C. Cox *et al.*, "A survey of homozygous deletions in human cancer genomes.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 12, pp. 4542–4547, Mar. 2005.

[18] N. Kumar, H. Cai, C. von Mering, and M. Baudis, "Specific genomic regions are differentially affected by copy number alterations across distinct cancer types, in aggregated cytogenetic data," *PLoS One*, vol. 7, no. 8, 2012.

[19] O. Shoshani *et al.*, "Chromothripsis drives the evolution of gene amplification in cancer.," *Nature*, vol. 591, no. 7848, pp. 137–141, Mar. 2021.

[20] U. Ben-David and A. Amon, "Context is everything: aneuploidy in cancer," *Nat. Rev. Genet.*, vol. 21, no. 1, pp. 44–62, 2020.

[21] D. Hanahan, "Hallmarks of Cancer: New Dimensions," *Cancer Discov.*, vol. 12, no. 1, pp. 31–46, 2022.

[22] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer.," *Cell*, vol. 100, no. 1, pp. 57–70, Jan. 2000.

[23] D. S. Chen and I. Mellman, "Oncology meets immunology: the cancer-immunity cycle.," *Immunity*, vol. 39, no. 1, pp. 1–10, Jul. 2013.

[24] T. F. Gajewski, H. Schreiber, and Y.-X. Fu, "Innate and adaptive immune cells in the tumor microenvironment.," *Nat. Immunol.*, vol. 14, no. 10, pp. 1014–1022, Oct. 2013.

[25] H. T. Khong and N. P. Restifo, "Natural selection of tumor variants in the generation of 'tumor escape' phenotypes.," *Nat. Immunol.*, vol. 3, no. 11, pp. 999–1005, Nov. 2002.

[26] T. Davoli, H. Uno, E. C. Wooten, and S. J. Elledge, "Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy.," *Science*, vol. 355, no. 6322, Jan. 2017.

[27] W. Liu *et al.*, "Dissecting the tumor microenvironment in response to immune checkpoint inhibitors via single-cell and spatial transcriptomics," *Clin. Exp. Metastasis*, no. 0123456789, 2023.

[28] A. van Weverwijk and K. E. de Visser, "Mechanisms driving the immunoregulatory function of cancer cells.," *Nat. Rev. Cancer*, vol. 23, no. 4, pp. 193–215, Apr. 2023.

[29] P. Sharma and J. P. Allison, "Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential.," *Cell*, vol. 161, no. 2, pp. 205–214, Apr. 2015.

[30] D. S. Chen and I. Mellman, "Elements of cancer immunity and the cancer-immune set point," *Nature*, vol. 541, no. 7637, pp. 321–330, 2017.

[31] S.-R. Woo, L. Corrales, and T. F. Gajewski, "Innate immune recognition of cancer.," *Annu. Rev. Immunol.*, vol. 33, pp. 445–474, 2015.

[32] V. Thorsson *et al.*, "The Immune Landscape of Cancer.," *Immunity*, vol. 48, no. 4, pp. 812-830.e14, Apr. 2018.

[33] A. M. Taylor *et al.*, "Genomic and Functional Approaches to Understanding Cancer Aneuploidy.," *Cancer Cell*, vol. 33, no. 4, pp. 676-689.e3, Apr. 2018.

[34] J. C. Smith and J. M. Sheltzer, "Systematic identification of mutations and copy number alterations associated with cancer patient prognosis.," *Elife*, vol. 7, Dec. 2018.

[35] J. Budczies *et al.*, "Integrated analysis of the immunological and genetic status in and across cancer types: impact of mutational signatures beyond tumor mutational burden," *Oncoimmunology*, vol. 7, no. 12, 2018.

[36] A. Jiménez-Sánchez *et al.*, "Unraveling tumor–immune heterogeneity in advanced ovarian cancer uncovers immunogenic effect of chemotherapy," *Nat. Genet.*, vol. 52, no. 6, pp. 582–593, 2020.

[37] D. A. Braun *et al.*, "Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma," *Nat. Med.*, vol. 26, no. 6, pp. 909–918, 2020.

[38] A. Bhattacharya, R. D. Bense, C. G. Urzúa-Traslaviña, E. G. E. de Vries, M. A. T. M. van Vugt, and R. S. N. Fehrmann, "Transcriptional effects of copy number alterations in a large set of human cancers," *Nat. Commun.*, vol. 11, no. 1, pp. 1–12, 2020.

[39] D. Mika, G. Budzik, and J. Józwik, "Single Channel Source Separation with ICA-Based Time-Frequency Decomposition," *Sensors*, vol. 20, p. 2019, 2020.

[40] D. Zhao *et al.*, "Identification of therapeutic targets and mechanisms of tumorigenesis in non-small cell lung cancer using multiple-microarray analysis," *Med. (United States)*, vol. 99, no. 44, p. E22815, 2020.

[41] J. N. Weinstein *et al.*, "The Cancer Genome Atlas Pan-Cancer analysis project.," *Nat. Genet.*, vol. 45, no. 10, pp. 1113–1120, Oct. 2013.

[42] B. Li and C. N. Dewey, "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.," *BMC Bioinformatics*, vol. 12, p. 323, Aug. 2011.

[43] S. Tweedie *et al.*, "Genenames.org: the HGNC and VGNC resources in 2021.," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D939–D946, Jan. 2021.

[44] Y. Hao *et al.*, "Integrated analysis of multimodal single-cell data.," *Cell*, vol. 184, no. 13, pp. 3573-3587.e29, Jun. 2021.

[45] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[46] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, and P. Tamayo, "The Molecular Signatures Database (MSigDB) hallmark gene set collection.," *Cell Syst.*, vol. 1, no. 6, pp. 417–425, Dec. 2015.

[47] F. J. Grisanti Canozo, Z. Zuo, J. F. Martin, and M. A. H. Samee, "Cell-type modeling in spatial transcriptomics data elucidates spatially variable colocalization and communication between cell-types in mouse brain," *Cell Syst.*, vol. 13, no. 1, pp. 58-70.e5, 2022.

[48] P. Nieto *et al.*, "A single-cell tumor immune atlas for precision oncology," *Genome Res.*, vol. 31, no. 10, pp. 1913–1926, 2021.

[49] D. J. Slamon *et al.*, "Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer.," *Science*, vol. 244, no. 4905, pp. 707–712, May 1989.

[50] N. R. Lemoine *et al.*, "Amplification and overexpression of the EGF receptor and c-erbB-2 proto-oncogenes in human stomach cancer.," *Br. J. Cancer*, vol. 64, no. 1, pp. 79–83, Jul. 1991.

[51] G. Sauter *et al.*, "Heterogeneity of erbB-2 gene amplification in bladder cancer.," *Cancer Res.*, vol. 53, no. 10 Suppl, pp. 2199–2203, May 1993.

[52] G. Stenman, J. Sandros, A. Nordkvist, J. Mark, and P. Sahlin, "Expression of the ERBB2 protein in benign and malignant salivary gland tumors.," *Genes. Chromosomes Cancer*, vol. 3, no. 2, pp. 128–135, Mar. 1991.

[53] M. Tateishi, T. Ishida, T. Mitsudomi, S. Kaneko, and K. Sugimachi, "Prognostic value of c-erbB-2 protein expression in human lung adenocarcinoma and squamous cell carcinoma.," *Eur. J. Cancer*, vol. 27, no. 11, pp. 1372–1375, 1991.

[54] M. C. Hung, A. L. Schechter, P. Y. Chevray, D. F. Stern, and R. A. Weinberg, "Molecular cloning of the neu gene: absence of gross structural alteration in oncogenic alleles.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 83, no. 2, pp. 261–264, Jan. 1986.

[55] D. Maglott, J. Ostell, K. D. Pruitt, and T. Tatusova, "Entrez Gene: gene-centered information at NCBI.," *Nucleic Acids Res.*, vol. 35, no. Database issue, pp. D26-31, Jan. 2007.

[56] W. Onkes *et al.*, "Breakpoint characterization of the der(19)t(11;19)(q13;p13) in the ovarian cancer cell line SKOV-3.," *Genes. Chromosomes Cancer*, vol. 52, no. 5, pp. 512–522, May 2013.

[57] R. B. Jenkins *et al.*, "Cytogenetic studies of epithelial ovarian carcinoma.," *Cancer Genet. Cytogenet.*, vol. 71, no. 1, pp. 76–86, Nov. 1993.

[58] F. H. Thompson *et al.*, "Simple numeric abnormalities as primary karyotype changes in ovarian carcinoma.," *Genes. Chromosomes Cancer*, vol. 10, no. 4, pp. 262–266, Aug. 1994.

[59] F. Micci *et al.*, "Reverse painting of microdissected chromosome 19 markers in ovarian carcinoma identifies a complex rearrangement map.," *Genes. Chromosomes Cancer*, vol. 48, no. 2, pp. 184–193, Feb. 2009.

[60] F. Micci *et al.*, "Array-CGH analysis of microdissected chromosome 19 markers in ovarian carcinoma identifies candidate target genes," *Genes, Chromosom. & cancer*, vol. 49, no. 11, p. 1046—1053, 2010.

[61] L. Wang *et al.*, "Frequent translocations of 11q13.2 and 19p13.2 in ovarian cancer.," *Genes. Chromosomes Cancer*, vol. 53, no. 6, pp. 447–453, Jun. 2014.

[62] I. M. Shih, K. Nakayama, G. Wu, N. Nakayama, J. Zhang, and T. L. Wang, "Amplification of the ch19p13.2 NACC1 locus in ovarian high-grade serous carcinoma," *Mod. Pathol.*, vol. 24, no. 5, pp. 638–645, 2011.

[63] M. L. Yaremko, M. L. Wasylyshyn, K. L. Paulus, F. Michelassi, and C. A. Westbrook, "Deletion mapping

reveals two regions of chromosome 8 allele loss in colorectal carcinomas.," *Genes. Chromosomes Cancer*, vol. 10, no. 1, pp. 1–6, May 1994.

[64] F. Lerebours *et al.*, "Deletion mapping of the tumor suppressor locus involved in colorectal cancer on chromosome band 8p21.," *Genes. Chromosomes Cancer*, vol. 25, no. 2, pp. 147–153, Jun. 1999.

[65] D. M. Takanishi *et al.*, "Chromosome 8 losses in colorectal carcinoma: Localization and correlation with invasive disease," *Mol. Diagnosis*, vol. 2, no. 1, pp. 3–10, 1997.

[66] C. Cunningham, M. G. Dunlop, A. H. Wyllie, and C. C. Bird, "Deletion mapping in colorectal cancer of a putative tumour suppressor gene in 8p22-p21.3.," *Oncogene*, vol. 8, no. 5, pp. 1391–1396, May 1993.

[67] M. Chang *et al.*, "Deletion mapping of chromosome 8p in colorectal carcinoma and dysplasia arising in ulcerative colitis, prostatic carcinoma, and malignant fibrous histiocytomas," *Am. J. Pathol.*, vol. 144, no. 1, pp. 1–6, 1994.

[68] T. Oyama *et al.*, "Isolation of a novel gene on 8p21.3-22 whose expression is reduced significantly in human colorectal cancers with liver metastasis.," *Genes. Chromosomes Cancer*, vol. 29, no. 1, pp. 9–15, Sep. 2000.

[69] Z. Fang *et al.*, "Coexistence of copy number increases of ZNF217 and CYP24A1 in colorectal cancers in a Chinese population.," *Oncol. Lett.*, vol. 1, no. 5, pp. 925–930, Sep. 2010.

[70] L. W. M. Loo *et al.*, "Integrated analysis of genome-wide copy number alterations and gene expression in microsatellite stable, CpG island methylator phenotype-negative colon cancer.," *Genes. Chromosomes Cancer*, vol. 52, no. 5, pp. 450–466, May 2013.

[71] A. H. Sillars-Hardebol, B. Carvalho, M. van Engeland, R. J. A. Fijneman, and G. A. Meijer, "The adenoma hunt in colorectal cancer screening: defining the target.," *J. Pathol.*, vol. 226, no. 1, pp. 1–6, Jan. 2012.

[72] D. E. Aust *et al.*, "Prognostic relevance of 20q13 gains in sporadic colorectal cancers: A FISH analysis," *Scand. J. Gastroenterol.*, vol. 39, no. 8, pp. 766–772, 2004.

[73] S. Knuutila *et al.*, "DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies.," *Am. J. Pathol.*, vol. 152, no. 5, pp. 1107–1123, May 1998.

[74] V. M. H. Bui, C. Mettling, J. Jou, and H. S. Sun, "Genomic amplification of chromosome 20q13.33 is the early biomarker for the development of sporadic colorectal carcinoma," *BMC Med. Genomics*, vol. 13, no. Suppl 10, pp. 1–11, 2020.

[75] X. Fan, J. Liang, Z. Wu, X. Shan, H. Qiao, and T. Jiang, "Expression of HLA-DR genes in gliomas: Correlation with clinicopathological features and prognosis," *Chinese Neurosurg. J.*, vol. 3, no. 1, pp. 1–9, 2017.

[76] Z.-H. Ji *et al.*, "Identification of immune-related biomarkers associated with tumorigenesis and prognosis in skin cutaneous melanoma.," *Am. J. Cancer Res.*, vol. 12, no. 4, pp. 1727–1739, 2022.

[77] G. Wu, G. Xiao, Y. Yan, C. Guo, N. Hu, and S. Shen, "Bioinformatics analysis of the clinical significance of HLA class II in breast cancer," *Med. (United States)*, vol. 101, no. 40, p. E31071, 2022.

[78] P. E. Leone *et al.*, "Integration of global spectral karyotyping, CGH arrays, and expression arrays reveals important genes in the pathogenesis of glioblastoma multiforme," *Ann. Surg. Oncol.*, vol. 19, no. 7, pp. 2367–2379, 2012.

[79] Y. R. Qin *et al.*, "Single-nucleotide polymorphism-mass array reveals commonly deleted regions at 3p22 and 3p14.2 associate with poor clinical outcome in esophageal squamous cell carcinoma.," *Int. J. cancer*, vol. 123, no. 4, pp. 826–830, Aug. 2008.

[80] A. L. S. Tai *et al.*, "High-throughput loss-of-heterozygosity study of chromosome 3p in lung cancer using single-nucleotide polymorphism markers.," *Cancer Res.*, vol. 66, no. 8, pp. 4133–4138, Apr. 2006.

[81] S. Hanke, P. Bugert, J. Chudek, and G. Kovacs, "Cloning a calcium channel alpha2delta-3 subunit gene from a putative tumor suppressor gene region at chromosome 3p21.1 in conventional renal cell carcinoma.," *Gene*, vol. 264, no. 1, pp. 69–75, Feb. 2001.

[82] A. Wanajo *et al.*, "Methylation of the calcium channel-related gene, CACNA2D3, is frequent and a poor

prognostic factor in gastric cancer.," *Gastroenterology*, vol. 135, no. 2, pp. 580–590, Aug. 2008.

[83] Y. Jin, D. Cui, J. Ren, K. Wang, T. Zeng, and L. Gao, "CACNA2D3 is downregulated in gliomas and functions as a tumor suppressor," *Mol. Carcinog.*, vol. 56, no. 3, pp. 945–959, 2017.

[84] F. Yu and W. M. Fu, "Identification of differential splicing genes in gliomas using exon expression profiling," *Mol. Med. Rep.*, vol. 11, no. 2, pp. 843–850, 2015.

[85] C. Palmieri *et al.*, "Methylation of the calcium channel regulatory subunit α2δ-3 (CACNA2D3) predicts site-specific relapse in oestrogen receptor-positive primary breast carcinomas.," *Br. J. Cancer*, vol. 107, no. 2, pp. 375–381, Jul. 2012.

[86] H. E. Feilotter *et al.*, "Analysis of the 10q23 chromosomal region and the PTEN gene in human sporadic breast carcinoma," *Br. J. Cancer*, vol. 79, no. 5–6, pp. 718–723, 1999.

[87] E. Rhei, L. Kang, F. Bogomolniy, M. G. Federici, P. I. Borgen, and J. Boyd, "Mutation analysis of the putative tumor suppressor gene PTEN/MMAC1 in primary breast carcinomas.," *Cancer Res.*, vol. 57, no. 17, pp. 3657–3659, Sep. 1997.

[88] S. Bose, S. I. Wang, M. B. Terry, H. Hibshoosh, and R. Parsons, "Allelic loss of chromosome 10q23 is associated with tumor progression in breast carcinomas," *Oncogene*, vol. 17, no. 1, pp. 123–127, 1998.

[89] A. Besson, S. M. Robbins, and V. W. Yong, "PTEN/MMAC1/TEP1 in signal transduction and tumorigenesis.," *Eur. J. Biochem.*, vol. 263, no. 3, pp. 605–611, Aug. 1999.

[90] L. P. Weng *et al.*, "PTEN suppresses breast cancer cell growth by phosphatase activity-dependent G1 arrest followed by cell death.," *Cancer Res.*, vol. 59, no. 22, pp. 5808–5814, Nov. 1999.

[91] J. Li *et al.*, "The PTEN/MMAC1 tumor suppressor induces cell death that is rescued by the AKT/protein kinase B oncogene.," *Cancer Res.*, vol. 58, no. 24, pp. 5667–5672, Dec. 1998.

[92] A. K. Ghosh, I. Grigorieva, R. Steele, R. G. Hoover, and R. B. Ray, "PTEN transcriptionally modulates c-myc gene expression in human breast carcinoma cells and is involved in cell growth regulation.," *Gene*, vol. 235, no. 1–2, pp. 85–91, Jul. 1999.

[93] P. L. Depowski, S. I. Rosenthal, and J. S. Ross, "Loss of expression of the PTEN gene protein product is associated with poor outcome in breast cancer," *Mod. Pathol.*, vol. 14, no. 7, pp. 672–676, 2001.

[94] S. H. Chang *et al.*, "Loss of PTEN Expression in Breast Cancers.," *Korean J. Pathol.*, vol. 39, no. 4, pp. 236–241, 2005.

[95] Z. Zhang *et al.*, "Overexpression of IFIT2 inhibits the proliferation of chronic myeloid leukemia cells by regulating the BCR-ABL/AKT/mTOR pathway," *Int. J. Mol. Med.*, vol. 45, no. 4, pp. 1187–1194, 2020.

[96] J. Chen *et al.*, "STAT1/IFIT2 signaling pathway is involved in PD-L1-mediated epithelial-to-mesenchymal transition in human esophageal cancer," *Clin. Transl. Oncol.*, vol. 24, no. 5, pp. 927–940, 2022.

[97] F. Ge, Z. Li, J. Hu, Y. Pu, F. Zhao, and L. Kong, "METTL3/m6A/IFIT2 regulates proliferation, invasion and immunity in esophageal squamous cell carcinoma," *Front. Pharmacol.*, vol. 13, no. October, pp. 1–17, 2022.

[98] K. C. Lai, P. Regmi, C. J. Liu, J. F. Lo, and T. C. Lee, "IFIT2 Depletion Promotes Cancer Stem Cell-like Phenotypes in Oral Cancer," *Biomedicines*, vol. 11, no. 3, 2023.

[99] S. Y. Koh, J. Y. Moon, T. Unno, and S. K. Cho, "Baicalein suppresses stem cell-like characteristics in radio-and chemoresistant MDA-MB-231 human breast cancer cells through up-regulation of IFIT2," *Nutrients*, vol. 11, no. 3, 2019.

[100] V. Cheriyath, J. Kaur, A. Davenport, A. Khalel, N. Chowdhury, and L. Gaddipati, "G1P3 (IFI6), a mitochondrial localised antiapoptotic protein, promotes metastatic potential of breast cancer cells through mtROS," *Br. J. Cancer*, vol. 119, no. 1, pp. 52–64, 2018.

[101] A. M. Davenport *et al.*, "G1P3/IFI6, an interferon stimulated protein, promotes the association of RAB5(+) endosomes with mitochondria in breast cancer cells.," *Cell Biol. Int.*, vol. 47, no. 11, pp. 1868–1879, Nov. 2023.

[102] Z. Liu *et al.*, "IFI6 depletion inhibits esophageal squamous cell carcinoma progression through reactive oxygen species accumulation via mitochondrial dysfunction and endoplasmic reticulum stress," *J. Exp. Clin. Cancer Res.*, vol. 39, no. 1, pp. 1–28, 2020.

Acknowledgements

In this report, ChatGPT was used to enhance grammar, syntax, and text-editing, thereby improving the overall quality of the English text. Retrieved from https://www.openai.com/chatgpt.

Appendix 1: EDA on the spatial transcriptomics samples

https://sites.google.com/student.rug.nl/research-project2-s5340411/appendix-1?authuser=1

Appendix 2: Genomic mapping of CNA-TCs

https://sites.google.com/student.rug.nl/research-project2-s5340411/appendix-2?authuser=1

Appendix 3: Colocalization of TCs on the spatial transcriptomics samples

https://sites.google.com/student.rug.nl/research-project2-s5340411/appendix-3?authuser=1

Appendix 4: Histograms of the correlation scores between the TCs

https://sites.google.com/student.rug.nl/research-project2-s5340411/appendix-4?authuser=1