

Comparative Effect of Herceptin and Its Biosimilar: Transcriptomic and Proteomic
Insights from HER2-Positive Cancer Cell Lines

by

Karsten Rinas

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Biology

Waterloo, Ontario, Canada, 2025

©Karsten Rinas 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Gary Bader Professor, The Donnelly Centre, University of Toronto
Supervisor(s)	Dr. Brendan J. McConkey Associate Professor, Department of Biology University of Waterloo
Internal Members	Dr. Marc Aucoin Professor, Department of Chemical Engineering University of Waterloo Dr. Brian Dixon Professor, Department of Biology University of Waterloo
Internal-external Member	Dr. Brian Ingalls Professor, Department of Applied Mathematics University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

I was the sole author of Chapters 1, 3, 4, and 5, which were written under the supervision of Dr. Brendan J. McConkey.

Herceptin (Genentech, San Francisco, USA) and its biosimilar, Aponte-Trastuzumab (ApoTras; Apobiologix, Toronto, Canada), were generously provided by Apobiologix. Apobiologix also provided the results for the potency, purity, and identity comparison of the used batches of Herceptin and biosimilar ApoTras summarized in **Table 1-1**.

For Chapter 2, the treatment and preparation of BT474 bulk RNA-sequencing samples were conducted by Steve Mooney at the University of Waterloo, while library preparation and sequencing of BT474 samples were performed at the Donnelly Centre for Cellular and Biomolecular Research (Toronto, Canada). I performed the data analysis and drafted the manuscript under the supervision of Dr. Brendan J. McConkey. Chapter 2 was prepared as a manuscript intended for publication and is pending submission.

For Chapter 3, the data analysis, algorithm development, and manuscript drafting were performed by me. Zihan Dong, an undergraduate researcher, assisted in implementing the power analysis algorithm under my supervision.

For Chapter 4, I designed the experiments and completed the data collection. Library preparation and RNA sequencing were performed by the Centre for Applied Genomics, SickKids Hospital (Toronto, Canada). I performed the data analysis and drafted the manuscript.

In Chapter 5, I conducted the proteomics experiments for treated BT474, SKBR3, and MCF7 cell lines. Tianchunlu Yu, also an undergraduate researcher, assisted with the treatment and preparation of BT474 proteomics samples for mass spectrometric analysis. The SPARC Biocentre, SickKids (Toronto, Canada) performed the mass spectrometric analysis. I performed the data analysis.

Abstract

The first patent for the therapeutic monoclonal antibody Herceptin expired in 2019, opening the door for the development of biosimilars. Biosimilars are copies of the originator product that must demonstrate biosimilarity for regulatory approval. This is a very difficult task. Antibodies are large molecules that undergo extensive post-translational modifications (PTMs), which affect their potency, stability, binding affinity, and immunogenicity. Batches of the originator products often vary in their PTM patterns, making structural similarity very difficult to prove. Despite these challenges, most biosimilar development workflows rely on targeted wet-lab assays to assess structural similarity, purity, and efficacy, often without addressing the full complexity of the task.

Herceptin targets the surface receptor HER2 (human epidermal growth factor receptor 2) on breast cancer cells, inducing global cellular changes including cell arrest, programmed cell death, reduced motility, and angiogenesis. The complete mode of action is under investigation, and the development of resistance to Herceptin is common and complex. Due to these gaps in knowledge and complexity of Herceptin's effect, untargeted omics techniques, such as transcriptomics and proteomics, are ideally suited for functionally comparing biosimilars to their originator product.

Therefore, the primary goal of this thesis is to evaluate whether transcriptomics and proteomics experiments are useful for investigating the impact of the biosimilar on target cells compared to the originator product. To this end, Herceptin's effect on the breast cancer cell lines BT474, SKBR3 and MCF7 will be compared to a biosimilar, called Apotex-Trastuzumab (ApoTras), provided by Apobiologix (Toronto, Canada). This analysis will allow us to investigate the sensitivity of the methods in detecting differences and determining how many of the known effects can be identified. This work will provide valuable insights into the use of untargeted omics approaches for investigating the functional similarity of a biosimilar to its originator product.

Acknowledgements

It's hard to know where to begin in expressing gratitude to those who have helped me along the way. The work described in this thesis represents the conclusion of years of research, growth, and collaboration, made possible by the encouragement, expertise, and support of many individuals.

First, I am very grateful to my supervisor, Dr. Brendan McConkey, for his constant support of my research and for providing me with everything I need to grow as a scientist. He created an environment where I could develop both personally and professionally. I would also like to thank my committee members, Dr. Brian Dixon and Dr. Marc Aucoin, for their feedback and mentorship, and Dr. Gary Bader for sharing his expertise and acting as my external examiner.

Thank you to my lab members, Monica, Mark, and Janet, for their scientific insights and emotional support during challenging times. You always made me laugh when I needed it.

On a personal level, I would like to extend my gratitude to my family and all my teachers and mentors for always believing in me and cheering me on along my way. Your love and encouragement provided the foundation for my success.

Finally, to everyone who contributed to this work, I offer my deepest thanks. This thesis is not just a reflection of my effort but a testament to the collective input of an amazing community of mentors, colleagues, friends, and family.

Dedication

This work is dedicated to my parents, Horst and Angelika Rinas, for their love and unwavering support since my first breaths on this earth.

Table of Contents

Examining Committee Membership	ii
Author’s Declaration	iii
Statement of Contributions	iv
Abstract	v
Acknowledgements.....	vi
Dedication	vii
List of Figures.....	xiii
List of Tables	xvi
List of Abbreviations.....	xvii
Chapter 1: Introduction	1
1.1 Breast Cancer	1
1.1.1 General Overview.....	1
1.1.2 Breast Cancer Risk Factors	1
1.1.3 Biological Mechanisms in Breast Cancer Progression.....	2
1.1.4 HER2-Positive Breast Cancer	4
1.2 General Cellular Functions of HER2	6
1.2.1 The HER Family and Their Ligands	6
1.2.2 Structure and Activation of HER2	7
1.2.3 Cellular Functions of the HER2 Receptor.....	8
1.2.4 Epidemiology and Clinical Characteristics	11
1.3 Cell Line Models for Breast Cancer Research.....	16
1.4 Anti-Her2 Monoclonal Antibody (trastuzumab).....	17
1.4.1 History of Monoclonal Antibodies	17
1.4.2 Monoclonal Antibody Structure	18
1.4.3 Production of mABs	18
1.4.4 Therapeutic mAB Advancements.....	20
1.4.5 Effects of Therapeutic mAbs.....	21
1.4.6 Types of Therapeutic Monoclonal Antibodies	23

1.4.7 Trastuzumab	25
1.5 Biosimilars	34
1.5.1 Introduction	34
1.5.2 Regulatory Framework for Biosimilars	36
1.5.3 Issues in Development.....	38
1.5.4 Current Best Practice to Assess Biosimilarity	40
1.5.5 Problems and Opportunities for Biosimilars	44
1.6 Thesis Relevant Information	46
1.6.1 Herceptin biosimilar characteristics (Apotex-trastuzumab)	46
1.6.2 BT474, SKBR3 and MCF7 sensitivity to Herceptin and Apotex-trastuzumab.....	48
1.7 Thesis Outline	49
Chapter 2: Assessing biosimilarity of trastuzumab monoclonal antibody	
therapeutics using RNA sequencing.....	51
2.1 Introduction	51
2.2 Materials and Methods	54
2.2.1 Cell Culture.....	54
2.2.2 Antibody Treatment.....	54
2.2.3 RNA Extraction and Sequencing.....	54
2.2.4 Data Analysis.....	54
2.2.5 Additional Files and Code Availability	57
2.3 Results	58
2.3.1 Quality control and data preprocessing	60
2.3.2 Determine the information level of the comparisons for different minFC.....	60
2.4 Discussion	70
Chapter 3: Type II error estimation for RNA-sequencing experiments using a	
bootstrap approach	74
3.1 Introduction	74
3.2 Materials and Methods	76
3.2.1 Data Set Preparation	76

3.2.2 Data Filtering.....	76
3.2.3 Algorithm	77
3.2.4 Visualization.....	79
3.2.5 Additional Files and Code Availability	79
3.3 Results	80
3.3.1 Real-life Data Introduction.....	80
3.3.2 Power Analysis Method Implementation	80
3.3.3 Overall Computed Power Results for Real Life Datasets	83
3.3.4 Expression Level Strongly Impacts the Calculated Power for Genes.....	84
3.3.5 Power Estimation on Gene Level	88
3.4 Discussion	91
Chapter 4: Molecular Characterization of HER2-Positive Breast Cancer Cell Lines	
Following Trastuzumab and Biosimilar Treatment	94
4.1 Introduction	94
4.2 Materials and Methods.....	97
4.2.1 Cell Line Culture and Treatment.....	97
4.2.2 RNA Extraction and Sequencing.....	97
4.2.3 Bioinformatic Analysis.....	98
4.2.4 Statistical Analysis	99
4.2.5 Pathway Activity Calculation.....	100
4.2.6 Gene Set Enrichment Analysis.....	101
4.2.7 Visualizations	102
4.2.8 Additional Files and Code Availability	102
4.3 Results	103
4.3.1 Validation of HER2-Positive Breast Cancer Cell Models	103
4.3.2 Molecular Treatment Efficacy is Highly Similar	103
4.3.3 PAM50 Gene Expression and Intrinsic Subtype Shifts After Treatment.....	104
4.3.4 Untargeted DEG Analysis	107
4.3.5 Adaptation of Cancer Hallmark Pathway Activity.....	109

4.3.6 Regulatory Network Insights	111
4.3.7 Distinct and Shared Pathway Changes in Response to Treatment	122
4.4 Discussion	145
4.4.1 Overview of Findings	145
4.4.2 Cell Line Characteristics and Suitability for Treatment Studies	146
4.4.3 Mechanisms of Action and Cellular Responses to trastuzumab.....	148
4.4.4 Biosimilarity Assessment	153
4.4.5 Research Implications.....	156
Chapter 5: Multi-omics data analysis to understand the phenotypic differences	
between Herceptin and biosimilar treatment.....	158
5.1 Introduction	158
5.2 Materials and Methods	161
5.2.1 Transcriptomics Data.....	161
5.2.2 Proteomics data.....	161
5.2.3 Data analysis.....	163
5.2.4 Visualization.....	165
5.2.5 Additional Files and Code Availability	166
5.3 Results	167
5.3.1 Proteomics data analysis.....	167
5.3.2 Multi-omics Profile of Targeted Response to trastuzumab	177
5.4 Discussion	186
5.4.1 Proteomics data.....	186
5.4.2 Cellular trastuzumab Response Profile.....	187
5.4.3 Assessing Biosimilarity with multi-modal data.....	190
5.4.4 Challenges and Opportunities for improvement.....	192
Chapter 6: Conclusion	194
6.1 Summary	194
6.2 Functional Insights and Follow-Up Experiments.....	196
6.3 Implications and Methodological Advances	198

6.4 Biological and Clinical Relevance	200
6.5 Limitations and Outlook	202
6.6 Final remarks.....	205
References	206
Appendix A	257
Appendix B	260

List of Figures

Figure 1-1 Sensitivity test of SKBR3, MCF7, and BT474 cell lines for Herceptin (Her) and its biosimilar, ApoTras.	48
Figure 2-1 Experimental workflow summary for the comparison of monoclonal antibodies. BT474 cells were exposed to biosimilar (ApoTras) or originator (Her) compounds, with no treatment (control) as a reference.	59
Figure 2-2 Sample clustering and detected differentially expressed genes. A) Principal components for sample clustering based on gene expression.	62
Figure 2-3 Histogram of the estimated power for all genes at a fold change of 1.5 (-33%/+50%) and 2.0 (-50%/+100%).	64
Figure 2-4 Functional profiling of differentially expressed genes. Functional profiling of DEG for Her&ApoTras vs. Con.	67
Figure 3-1 Bootstrap resampling implementation.	82
Figure 3-2 Power estimation stably increases for differing effect size, replicate number and dataset structure.	84
Figure 3-3 Per gene power is highly dependent on expression level.	87
Figure 3-4 Power curves for representative genes for each expression group and comparison. ...	90
Figure 4-1 General data structure for comparing the originator Herceptin with the biosimilar treated samples versus control.	104
Figure 4-2 Intrinsic subtype shift caused by trastuzumab treatment.	106
Figure 4-3 Comparative Analysis of Differential Gene Expression and Pathway Enrichment in BT474 and SKBR3 Cell Lines.	108
Figure 4-4 Scatter plot comparing log fold-changes (logFC) in gene expression between BT474 and SKBR3 cell lines for the comparison of ApoTras biosimilar against Herceptin (ApoTrasvsHer).	109
Figure 4-5 Cancer hallmark pathway (PROGENy) activities in BT474 and SKBR3 cell lines treated with Herceptin and/or its biosimilar.	111
Figure 4-6 Analysis of Transcription Factor Activity and Pathway Enrichment.	113

Figure 4-7 Network Analysis of Transcription Factors and Target Gene Overlap.....	115
Figure 4-8 Transcription Factor Activity in Response to Biosimilar vs. Herceptin Treatment in BT474 and SKBR3 Cell Lines.	119
Figure 4-9 Transcription Factor Networks and Volcano Plot Analysis of Biosimilar vs. Herceptin Treatment in BT474.	121
Figure 4-10 Overview of pathway clustering and similarity analysis of Reactome pathways comparing both treatments, Herceptin and ApoTras, against their respective controls for BT474 and SKBR3 cell lines.	123
Figure 4-11 Sankey plot illustrating the distribution and overlap of gene set enrichment clusters between BT474 and SKBR3 cell lines in response to Herceptin and ApoTras treatment compared to control.	124
Figure 4-12 Overview of pathway clustering and similarity analysis of Reactome pathways comparing Herceptin against the biosimilar ApoTras, for BT474 and SKBR3 cell lines.	136
Figure 4-13 Sankey plot illustrating the distribution and overlap of gene set enrichment clusters between BT474 and SKBR3 cell lines in response to Herceptin treatment compared to ApoTras treatment.	137
Figure 5-1 Technical quality assessment of quantified reporter intensities across all three proteomics datasets.	169
Figure 5-2 Biological quality assessment of proteomics data for treated BT474, SKBR3, and MCF7 cell lines.	171
Figure 5-3 Comparison of protein identifications across datasets.	173
Figure 5-4 Comparison of differentially expressed protein results for all three cell lines. Threshold for significance is 5% FDR.	176
Figure 5-5 Intersection visualization of differentially expressed gene (DEG) and proteins (DEP) for all three cell lines.	177
Figure 5-6 Combined trastuzumab effect profile of enriched gene sets in Transcriptomics and Proteomics data.	179
Figure 5-7 Hub genes for Cell cycle regulation gene set group.	181

Figure 5-8 Overlap of enriched gene sets from biological processes BP across all four datasets for the comparison ApoTrasvsHer.	183
Figure 5-9 Shared leading-edge genes for BT474 cell line modalities.....	185

List of Tables

Table 1-1 Comparative Analysis of Potency, Identity, and Purity Attributes of ApoTras Biosimilar (#FAZ1502) and Herceptin (US; #3014296).	47
Table 2-1 Numbers of significant DEG for different fold change thresholds (FCT) and an FDR of 0.05.	60
Table 2-2 Differentially expressed genes (Name and HNGC symbol) identified in ApoTras vs. Her.	69
Table 3-1 Summary of identified genes per dataset.....	80
Table 4-1 Gene set enrichment analysis of Reactome pathways comparing Herceptin and ApoTras treatments against the control samples in BT474 and SKBR3 cell lines.	128
Table 4-2 Gene set enrichment analysis of Reactome pathways comparing Herceptin and ApoTras treatments in BT474 and SKBR3 cell lines.	141
Table 5-1 Proteomics Quality Control Metrics for three 10xTMT-Labeled shotgun proteomics data sets of treated BT474, SKBR3, and MCF7 cell lines.	167
Table 5-2 Top 5 hub genes for each gene set group.	182

List of Abbreviations

ADC	antibody-drug conjugate
ADCC	antibody-dependent cellular cytotoxicity
AKT	serine/threonine kinase
ApoTras	trastuzumab biosimilar provided from Apobiologix
BP	Biological Processes
CDC	complement-dependent cytotoxicity
CDR	complementarity-determining region
CPM	counts per million
DEG	differentially expressed gene
DEP	differentially expressed proteins
DMSO	dimethyl sulfoxide
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EMA	European Medicines Agency
EMT	epithelial-mesenchymal transition
ER	estrogen receptor
FC	fold change
FCT	Fold change threshold
FDA	Food and Drug Administration
FDR	false discovery rate
G1	first gap phase (cell cycle)
G2	second gap phase (cell cycle)
GSEA	Gene set enrichment analysis
HAMA	human anti-mouse antibody
HER2	human epidermal growth factor receptor 2
HER2+	HER2-positive
HiSeq	High throughput sequencing platform

HPLC	high-performance liquid chromatography
HR-	hormone receptor-negative
HR+	hormone receptor-positive
iTRAQ	Isobaric Tags for Relative and Absolute Quantitation
LC-MS/MS	liquid chromatography-tandem mass spectrometry
lcpm	log2 counts per million
LFC	log2 fold change
M phase	mitosis (cell cycle)
mAb	monoclonal antibody
minFC	Minimum fold change
mRNA	messenger RNA
mTOR	mechanistic target of rapamycin
NGS	next-generation sequencing
NK	natural killer (cells)
PCA	principal component analysis
PR	progesterone receptor
PSM	Peptide-Spectrum Match
PTM	post-translational modification
RNA	ribonucleic acid
RNA-seq	RNA sequencing
S phase	synthesis phase (cell cycle)
SI	Similarity index
T-DM1	trastuzumab emtansine
TCGA	The Cancer Genome Atlas
TF	transcription factor
TMT	Tandem Mass Tags
TNBC	triple-negative breast cancer
TPM	transcripts per million
VEGF	vascular endothelial growth factor

VSN

Variance-Stabilizing Normalization

Chapter 1: Introduction

1.1 Breast Cancer

1.1.1 General Overview

Breast cancer continues to be a major health challenge worldwide. It is the most diagnosed cancer and the leading cause of cancer-related mortality in women. In 2023, the World Health Organization (WHO) estimated that there were around 2.3 million new breast cancer cases in women and 685,000 deaths in the same year globally (WHO, 2023). With the high incidence and mortality of the disease, solutions for prevention, diagnosis, and treatment are paramount. Breast cancer is a heterogeneous disease with multiple subtypes that have been defined by distinct molecular and clinical features in recent decades. Clinically, breast cancer *in situ* is categorized based on underlying molecular subtypes primarily by the presence or absence of hormone receptors and human epidermal growth factor receptor 2 (HER2).

The main subtypes include hormone receptor-positive (HR+), HER2-positive (Jørgensen, 2022b, 2022a), and triple-negative breast cancer (TNBC) (Lehmann et al., 2011; Liedtke et al., 2008). HR+ subtypes include Luminal A (Kennecke et al., 2010; Parker et al., 2009) and Luminal B (Cheang et al., 2009; Nielsen et al., 2010) which differ in growth rates and outcomes. For example, HER2-positive breast cancer, although aggressive, responds well to antibody therapies like trastuzumab, but TNBC, with negative ER, PR, and HER2 receptors, is the most challenging to treat with less effective treatment options. Management of breast cancer often includes combinations of antibodies, chemotherapy, and targeted therapies (Agostinetti et al., 2022; Arora et al., 2022; Jerusalem et al., 2015; J. Wang & Wu, 2023).

1.1.2 Breast Cancer Risk Factors

There are multiple genetic, environmental, and lifestyle factors that increase the risk for breast cancer. Age is one of the strongest risk factors, for which rates of diagnosis for breast cancer increase as women get older (Anderson et al., 2008). Genetic factors are also a part of this risk, such as carrying mutations in the BRCA1 and BRCA2 genes (Antoniou et al., 2003).

The effects of estrogen on breast cancer development are very well-documented in epidemiological studies. Late menopause, and use of hormone replacement therapy are both associated with increased risk due to prolonged exposure to estrogen. Reproductive history and lifestyle factors are known risk factors for breast cancer (Kelsey et al., 1993; Łukasiewicz et al., 2021). Obesity elevates risk by stimulating circulating hormone concentrations (Calle & Kaaks, 2004). Alcohol intake and poor diet are significantly associated with breast cancer risk (Friedenreich & Cust, 2008).

Environmental exposures can also modulate breast cancer risk. Examples include exposure to ionizing radiation and certain chemical pollutants (Boyd et al., 2011). High breast density may also be associated with increased risk (Hiatt & Brody, 2018). Further research will likely identify environmental risk factors supporting the development of more effective breast cancer prevention strategies.

1.1.3 Biological Mechanisms in Breast Cancer Progression

The development of breast cancer derives from the dysregulation of normal breast tissue, with the form and function of breast tissue impacting development of breast cancer. The breast is composed of four main tissue types, including epithelial tissues, stromal tissues that have a structural role, and adipose tissue that serves as energy store and hormone producer. The functional unit of the breast is the mammary gland. Mammary glands consist of lobules and their branching ducts surrounded by connective tissue (Macias & Hinck, 2012; Russo & Russo, 2004). Ducts are tubular structures that transports milk and are lined with epithelial cells. Lobules are clusters of milk-producing glands. The connective tissue surrounding the lobules and ducts is a mix of fibroblasts, adipocytes, and extracellular matrix, providing structural support and defines the microenvironment of the breast. Hormones, most notably estrogen and progesterone, regulate the growth and function of the breast tissue by promoting growth, differentiation, and production of milk (Macias & Hinck, 2012; Russo & Russo, 2004). These hormones are recognized by specific cell surface receptors on the breast cell. These hormones-receptor complexes in turn produce signals that influence gene expression and ultimately, cell function (Jensen & Jordan, 2003).

Estrogen plays a central role in regulating cell proliferation, differentiation, and apoptosis, and their associated complexities are on display during the menstrual cycles and pregnancy (Jensen & Jordan, 2003; Macias & Hinck, 2012). Proliferation is simply an increase in the number of cells mediated through cell division. Differentiation refers

to the acquisition of specialized functions, such as the capacity to produce milk. Apoptosis is the death of injured or unneeded cells, a critical regulator of tissue homeostasis. Defects in any of these processes, particularly those caused by hormone receptors, are known to cause carcinogenesis (S.-S. Cao & Lu, 2016). The cellular origins of most breast cancers are thought to reside in the duct or lobule epithelia. For example, basal-like breast cancer is thought to initiate in luminal progenitor cells (Lim et al., 2009), while more mature luminal cells could be the origination points for the luminal A and B subtypes (Shehata et al., 2012).

Dysregulation of the estrogen receptor (ER), progesterone receptor (PR), and HER2 pathways by genetic or epigenetic events is typically an early initiating event in the multistep progression from normal breast epithelium to breast cancer (Arteaga et al., 2012; Hanahan & Weinberg, 2011). These pathways control proliferative signalling, a basic hallmark in the progression to a cancer cell (Hanahan & Weinberg, 2011). A common pro-proliferative pathway downstream of growth factor signalling is the PI3K/AKT/mTOR signalling pathway, which serves as a central coordinator of cellular growth, survival, and metabolism (Fruman & Rommel, 2014; Gordon & Banerji, 2013).

Another common mechanism through which cancer cells may evade growth suppression is through mutations in growth suppressor genes, e.g., TP53 and BRCA1/2 that result in reversal of cell cycle control and DNA repair (Venkitaraman, 2002). Cancer progression is driven by genomic instability and mutations, such as HER2 amplification, and BRCA mutations, among many others (Arteaga et al., 2012; Negrini et al., 2010; Venkitaraman, 2002). Not only is HER2 overexpression an advantageous alteration of signalling in cancer cells, but it is also an important prognostic and predictive biomarker in breast cancer (Arteaga et al., 2012; Perou et al., 2000).

Another hallmark of breast cancer cells, related to their ability to evade cell death, is achieved through the dysregulation of both intrinsic and extrinsic apoptotic pathways. ER-positive tumours, for example, often overexpress anti-apoptotic proteins such as BCL-2, which also endow the cells with chemoresistance (Vander Heiden et al., 2009). Through repeated cell divisions, the increased action of telomerase contributes to breast cancer cells replicative immortality (Blasco, 2005; Vander Heiden et al., 2009). Collectively, these changes to a cell are associated with the stem-like characteristic of self-renewal, giving rise to the invasive and de novo metastatic potential of breast cancers (Blasco, 2005; Vander Heiden et al., 2009).

To survive and thrive, the tumour must secure sufficient supply of oxygen and nutrients. A common method is to stimulate the growth of blood vessels via angiogenesis, which is predominantly driven by VEGF signalling in endothelial cells (Carmeliet & Jain, 2000). Another essential feature of tumour progression is immune evasion. This helps the cells avoid immune surveillance by modifying the tumour microenvironment. To this end, tumour cells suppress immune cells, communicate with stromal cells, and secrete cytokines that mediate inflammation (Coussens & Werb, 2002; Vinay et al., 2015). A common next stage in metastasis is for a cell to undergo epithelial-mesenchymal transition (EMT). EMT is the process in which epithelial cells lose cell-cell adhesion and polarity and gain migratory mesenchymal-like characteristics. EMT is thought to be involved in the metastasis of breast cancer cells to the bones, lungs, and other critical organs that often follow invasion (Chaffer & Weinberg, 2011; Negrini et al., 2010).

Triple-negative breast cancers (TNBCs) lack the expression of the ER, PR, and HER2, and are aggressive and difficult to treat (Foulkes et al., 2010). TNBCs can occur at different times during tumorigenesis, can occur independently on different subtypes, and can occur in different combinations on different subtypes. This complexity of disease processes also suggests many possible targets, which presents opportunities for tumour-specific targeted and effective therapies. For example, trastuzumab and Pertuzumab (HER2 inhibitors), CDK4/6 inhibitors, and immune checkpoint inhibitors, have each drastically improved patient survival for their specific subtype (Arteaga et al., 2012; S.-S. Cao & Lu, 2016; Foulkes et al., 2010). To develop effective, personalized, and transformative therapeutic strategies, it is imperative to understand the molecular features of sub-types of breast cancer.

1.1.4 HER2-Positive Breast Cancer

HER2-positive breast cancer accounts for approximately 15-20% of all subtypes. It is caused by the overexpression or amplification of the HER2 gene, which is a fundamental driver of aggressive tumour growth (Engel & Kaklamani, 2007; Loibl & Gianni, 2017). At the biologic level, HER2-positive cancers are aggressive; they proliferate quickly and are more likely to metastasize than other subtypes of breast cancer (Balduzzi et al., 2014). HER2 status is determined by immunohistochemistry (IHC), which detects protein expression on the cell membrane, or fluorescence in situ hybridization (FISH), which assesses HER2 amplification, for treatment planning (Wolff et al., 2007).

Additionally, as with TNBCs, HER2-positive tumours are larger at diagnosis and with nodal involvement.

The treatment of HER2-positive breast cancer has been revolutionized by targeted therapies, such as the therapeutic monoclonal antibodies trastuzumab (Herceptin), pertuzumab (Perjeta), and trastuzumab-DM1 (Kadcyla) (Swain et al., 2015). Targeted therapies are combined with chemotherapy and, with hormone-sensitive tumours, hormone therapy (von Minckwitz et al., 2017). The widespread use of HER2-targeted therapies has greatly increased survival rates for women with HER2-positive breast cancer. The five-year disease-free survival rates are as high as 92.4% under current treatment regimens (Cameron et al., 2017), although between 15-31% of all treated women will eventually relapse (Pinto et al., 2013).

Challenges in the treatment of HER2-positive cancers include the resistance to treatment, a high risk for central nervous system metastases (5%), and potential cardiotoxicity from the treatment (Ewer & O'Shaughnessy, 2007; Olson et al., 2013). This has driven research into resistance mechanisms, development of other therapies, and the exploration of more precise biomarkers, culminating in more personalized and subtype specific approaches (Fizman & Jasniss, 2011). Treatment of HER2-positive breast cancer became more streamlined due to the favourable toxicity profile of targeted therapies, but on the other hand, there remains much ground to cover in terms of optimizing overall survival and quality of life for patients.

1.2 General Cellular Functions of HER2

1.2.1 The HER Family and Their Ligands

HER2, a member of the HER family of receptor tyrosine kinases, regulates cell survival and proliferation. Accordingly, HER2 is one of the most important therapeutic targets in HER2-positive breast cancer. The HER family, also known as the ErbB family, consists of four closely related receptor tyrosine kinases: HER1 (EGFR or ErbB1), HER2 (ErbB2), HER3 (ErbB3) and HER4 (ErbB4) (Roskoski, 2014). Together, these receptors regulate many of the critical aspects of cell physiology, including growth, survival, adhesion, migration, and differentiation (Olayioye et al., 2000). Each of these receptors, except for HER2, has a specific ligand that activates it.

- HER1 (EGFR or ErbB1): Ligands for HER1 include EGF, TGF- α , and amphiregulin. This receptor is responsible for initiating signal-transduction cascades that result in cell growth and differentiation. Overexpression or activating mutations of this class of receptors have been associated with a variety of cancers, including lung and colorectal cancers. (Normanno et al. 2006; Yarden & Sliwkowski 2001).
- HER2 (ErbB2): HER2 is unique in that it does not have a known ligand and is constitutively active. It forms heterodimers with other HER family members, leading to potent signalling cascades (Hynes & Lane, 2005; Yarden & Sliwkowski, 2001).
- HER3 (ErbB3): HER3 binds heregulins (neuregulins) to activate its function, when dimerized with other HER family receptor members, primarily HER2. Although HER3 lacks an active kinase domain, this receptor plays a vital role, when bound to the other HER family receptor members, primarily HER2, in PI3K/AKT pathway activation (Roskoski, 2014; Yarden & Sliwkowski, 2001).
- HER4 (ErbB4): Heregulins, neuregulins, and betacellulin all bind to HER4 to facilitate cell differentiation and proliferation, which has been implicated in normal development as well as in the context of cancer (Yarden & Sliwkowski, 2001).

1.2.2 Structure and Activation of HER2

HER2 is a 185 kDa transmembrane glycoprotein involved in several cellular signalling pathways. The molecule is composed of three major domains: an extracellular ligand-binding domain, a transmembrane domain, and an intracellular tyrosine kinase domain (Gajria & Chandarlapaty, 2011).

With respect to the proposed mechanisms of transformation by HER2, this analysis reviews only the roles of the kinase domain and its truncated mutants, which are predominant in HER2-positive breast cancer cell lines (Pupa et al., 2021). The C-terminus of the full-length HER2 receptor and the juxta membrane region between the kinase and transmembrane domains are important for regulating receptor activity. Mutations in these regions also contribute to transformation (Gajria & Chandarlapaty, 2011). Moreover, although the extracellular domain of HER2 has recently been recognized as a ligand-independent dimerization arm, mutations in this region can also contribute to transformation (Arteaga & Engelman, 2023). The understanding of transformation by HER2 is likely incomplete without considering these regions. The extracellular ligand-binding domain of HER2 consists of four subdomains. In other family members, ligand binding has been localized to subdomains I and III. However, HER2 is unique in that its extracellular domain does not bind any known ligands, keeping HER2 in a constitutively active state. This feature gives HER2 strong signalling capacity, especially when it heterodimerizes with other members of the HER family (Clavero-Álvarez et al., 2018).

The transmembrane domain is responsible for anchoring the receptor to the cell membrane. This hydrophobic domain crosses the lipid bilayer and is critical for dimerization. Following ligand binding with another HER2 receptor or with other members of the HER family, this dimerization is crucial for receptor activation and the initiation of downstream signalling pathways (Zahavi & Weiner, 2018).

The intracellular active tyrosine kinase domain is the enzymatic domain of the receptor and contains multiple tyrosine residues that are phosphorylated upon receptor activation. The phosphorylated tyrosines serve as docking sites for various intracellular signalling proteins that initiate multiple downstream signalling cascades. HER2 activates the MAPK/ERK and PI3K/AKT pathways during signal transduction downstream of these phosphorylated tyrosine sites. These pathways directly affect cell proliferation,

survival, and differentiation, contributing to HER2's role in tumorigenesis (Weigelt et al., 2010).

1.2.3 Cellular Functions of the HER2 Receptor

The HER family of receptors, and HER2 in particular, has been implicated in a wide variety of cellular processes. One of the most important functions of HER2 signalling is proposed to be the regulation of cell proliferation and differentiation. In normal tissues, this role is critical to achieving and maintaining proper tissue development and ensuring the maintenance of tissue integrity and function (Moasser, 2007).

Deregulation of HER2 in these tissues has been documented to drive oncogenesis. In cardiac development, HER2 is required for the specification of cardiac muscle cells and proper heart function, establishing it as a critical cardiogenic mediator (Negro et al., 2004). HER2 signalling has been demonstrated to be essential for the proper development and differentiation of the nervous system, directing appropriate neural tissue patterning and preserving neural tissue function (Moasser, 2007).

Throughout organismal development, HER2 is indispensable for both embryogenesis and postnatal development. In embryogenesis specifically, HER2, in coordination with the other HER family receptors, is required for the progression of the embryo as well as the proper development of tissues and organs (Negro et al., 2004). HER2 also has unique and essential functions in early stem cells. During embryogenesis, early stem cells must divide and differentiate with high temporal and spatial precision. As a receptor involved in all these functions, HER2 is required to regulate the growth and differentiation of early embryonic stem cells (Weigelt et al., 2010). After birth, HER2 is still required to allow cells, tissues, and organs to maintain homeostasis and repair themselves in response to injury or other stresses. HER2 continues to play similar roles throughout the life of the organism. When HER2 is needed again to regulate cell growth and differentiation to replace and regenerate tissue, in some situations, HER2 regulates cell transformation and drives cancer development (Moasser, 2007).

Together, HER2 can modulate many core aspects of cellular function and development. These critical roles in normal and pathological states, such as cancer, highlight the central importance of HER2 in cellular biology and make it a very compelling therapeutic target (Weigelt et al., 2010).

1.2.3.1 Her2 Downstream Signalling Pathways

HER2 activates two fundamental signalling pathways, the MAPK/ERK and PI3K/AKT cascades, which regulate cell proliferation and survival. Both pathways play a regulatory role in many aspects of the cell's function, such as gene transcription, organization of the actin cytoskeleton, and inhibition of apoptosis. Alterations in the regulation of these pathways as a result of HER2 overexpression or amplification lead to increased proliferation of cells and tumorigenesis (L. Pan et al., 2024).

The MAPK/ERK pathway is a primary signal transduction pathway downstream of HER2. This cascade is initiated by HER2 forming heterodimers most commonly with HER3 or with EGFR. Heterodimerization results in autophosphorylation of tyrosine residues in the intracellular kinase domains of HER2. These phosphorylated tyrosines then serve as binding sites for adaptor proteins, such as GRB2 and SOS. SOS catalyses the exchange of GDP with GTP on RAS. After RAS is activated, it initiates a cascade of signalling through RAF, MEK, and ERK. ERK is then phosphorylated and relocates to the nucleus, where it can phosphorylate transcription factors or other nuclear effectors to alter gene transcription (Weigelt et al., 2010). Among these many potential target genes, active ERK is necessary and sufficient to promote transcription of cell cycle regulators including cyclin D1. This ultimately promotes progression from G1 to S (L. Pan et al., 2024).

A second pathway, the PI3K/AKT cascade is a critical regulator of cell survival, metabolism, and other processes. This cascade is also activated through HER2 dimerization and autophosphorylation. Phosphorylated tyrosines on HER2 serve as docking sites for the p85 subunit of PI3K, which subsequently activates its catalytic subunit, p110. Conversion of PIP2 to PIP3 at the plasma membrane ensues. PIP3 recruits proteins that contain pleckstrin homology domains, such as AKT, to the membrane, where AKT undergoes a conformational change, resulting in phosphorylation and activation by PDK1 and mTORC2 (He et al., 2021). Active AKT is a promiscuous kinase that targets multiple downstream effectors, inhibiting the pro-apoptotic signalling and promoting protein synthesis through mTOR activation. Moreover, the PI3K/AKT pathway supports cellular metabolism, promoting cellular growth through upregulated glucose uptake and glycolysis. This enhanced metabolism provides necessary resources for cell division. Tight control of this cascade, often disrupted through HER2 overexpression or mutations in PI3K or PTEN, ensures the survival of cells and is

increased even further through oncogenic signalling cascades (He et al., 2021; L. Pan et al., 2024).

1.2.3.2 Her2 in Cancer

HER2-positive tumours grow faster and metastasize more frequently than HER2-negative tumours, due to activation of downstream signalling pathways that promote cell survival and resistance to apoptosis (Clavero-Álvarez et al., 2018). HER2-positive breast cancer is highly proliferative and highly metastatic. This rapid growth is caused by HER2's ability to homodimerize or heterodimerize with another EGFR receptor in the absence of a ligand. The bound receptor pair forms a signalling dimer with kinase activity, which then auto phosphorylates its intracellular tyrosine residues (Weigelt et al., 2010). This activation proceeds down signalling cascades that exacerbate all aspects of the PI3K/AKT pathway. AKT, a central signalling molecule, is frequently overactivated in HER2-positive breast tumours. In normal cells, AKT signalling prevents apoptosis following exposure to growth factors and regulates cellular metabolism, thus allowing cells to enter the cell cycle to support cell division. The PI3K/AKT pathway is nearly always overactive in HER2-overexpressing cancer cells, and provides a major survival advantage, contributing to its malignancy (He et al., 2021).

The second pivotal signalling pathway that is overactive due to excessive HER2 signalling is the MAPK/ERK pathway. This pathway lies downstream of growth factor receptors and transmits pro-proliferative signals to the nucleus. When signalling through ERK occurs, non-transformed cells respond with proliferation, transformation, and eventually resistance to standard therapies (Clavero-Álvarez et al., 2018). As a result, HER2-overexpressing tumours are more vascularized; this requires the upregulation of the pro-angiogenic growth factor VEGF. VEGF has been confirmed by several labs to be a suppressed target of HER2, and repression of VEGF following effective inhibition of HER2 can be sufficient to stop tumour growth altogether (Weigelt et al., 2010).

Similarly, lymph nodes serve as supportive niches for tumour cell proliferation. Invasive breast cancer with positive HER2 status has a striking propensity to metastasize both earlier and to a larger extent to certain sites like the brain or liver, where a specialized microenvironment may support continued tumour cell growth despite standard treatments (L. Pan et al., 2024). As a result, HER2 breast cancer often does not respond to standard treatments, like chemotherapies or hormonal therapies, and novel

therapies should target HER2 or associated mechanisms to stop the secretion of VEGF (He et al., 2021).

1.2.4 Epidemiology and Clinical Characteristics

Compared to HER2-negative subtypes, HER2-positive breast cancers are more likely to present aggressively and to show higher rates of recurrence and metastasis (Zahavi & Weiner, 2018). As a function of patient population, HER2-positive breast cancer can occur in women of all ages but is diagnosed more commonly in younger women than other breast cancer subtypes (National Breast Cancer Foundation, 2024). Clinically, HER2-positive tumours are more likely to be high-grade, show increased proliferation, present larger at presentation and have positive lymph nodes (Zahavi & Weiner, 2018). In the past, HER2-positive breast cancer had an unfavourable prognosis based on its aggressive biology. However, with the advent of HER2-targeted therapies, outcomes have improved and now fall into a more favourable category with more effective management strategies and rates of recurrence (Pernas et al., 2019). HER2-positive breast cancer is characterized by specific genetic alterations, chromosomal abnormalities, and molecular subtypes that account for its aggressive nature and clinical behaviour (Zahavi & Weiner, 2018).

1.2.4.1 Molecular and Genetic Features

HER2-positive breast cancer is characterized by a series of genetic aberrations, chromosomal rearrangements, and distinct molecular phenotypes that are associated with its oncogenic behaviour and clinical progression (Nik-Zainal et al., 2016).

A hallmark of HER2-overexpressing breast cancer is the gene amplification or overexpression of HER2, also known as ERBB2. The HER2 gene amplification locus lies on chromosome 17q12, leading to multiple copies of the HER2 gene. Multiple copies of the HER2 gene result in overexpression of the HER2 protein at the cell surface, transducing exponentially increased signals for the transformation of HER2+ cells (Koboldt et al., 2012). Although point mutations within HER2 can lead to overactivity, such events are far less frequent than gene amplification (Bose et al., 2013).

HER2-positive breast cancers are often characterized by specific chromosomal abnormalities and mutations. In addition to amplification of the HER2 gene locus, recurrent gains of chromosome 17 at other sites, such as 17q21-q22, have also been reported (Orsetti et al., 2017). These regions are thought to harbour additional oncogenes

that cooperate with HER2 in tumorigenesis. The TP53 gene, a crucial tumour suppressor, is mutated at a high frequency in HER2-positive breast cancers. Mutations in TP53 drive genomic instability and provide a mechanism for resistance to apoptosis (Koboldt et al., 2012). Mutations that constitutively activate the PI3K/AKT signalling pathway have also been identified in a substantial number of HER2-positive cancers, with the PIK3CA gene, which encodes the p110 α catalytic subunit of PI3K, being the most frequently altered (Prat et al., 2014). Activation of this signalling pathway enhances cell survival and proliferation. In contrast, PTEN, a chromosomal tumour suppressor gene that functions upstream in the PI3K/AKT pathway to inhibit pathway activity, is often lost or mutated in HER2-positive cancers, leading to further pathway activation (Koboldt et al., 2012).

Molecular profiling has identified several biomarkers of HER2-positive breast cancer that are critical for prognosis and therapeutic decision-making (Z. Jiang et al., 2024). Among these, immunohistochemistry (IHC) is widely used to quantify HER2 protein overexpression within tumour tissue. A score of 3+ on a scale of 0–3 indicates high HER2 expression and makes the patient eligible for HER2-targeted therapies. As an alternative method, fluorescence in situ hybridization (FISH) is used to identify HER2 gene amplification. A HER2/CEP17 ratio greater than 2.0 denotes amplification (Kurozumi et al., 2019). A truncated isoform of HER2, known as P95HER2—which lacks the extracellular domain—has been linked to trastuzumab resistance. If the HER2 pathway is adequately suppressed, for example by Lapatinib, P95HER2 is sensitive to other HER2-targeted therapies (J. Wang & Xu, 2019). The analysis of circulating tumour DNA (ctDNA) offers real-time information about HER2-positive tumours, including the development of somatic mutations with potential clinical relevance, such as those in the PIK3CA gene (Z. Jiang et al., 2024).

However, the complexity of signalling pathways downstream of HER2, along with those cross-regulated by HER2, suggests that upstream activating mutations alone are insufficient for disease development. Rather, some downstream steps promote tumorigenesis while others may be protective. Indeed, mounting evidence indicates that HER2 signalling does not act in isolation but instead communicates with other pathways to amplify its oncogenic potential (J. Wang & Xu, 2019). For example, in HER2+/hormone receptor-positive tumours, HER2 signalling significantly interacts with estrogen receptor signalling. This cross-talk influences treatment response, as therapies targeting HER2 can eventually lead to upregulation or increased dependence on the

estrogen receptor for growth and survival, indicating coordinated regulation of tumour progression by both pathways (Prat et al., 2014). HER2 signalling also upregulates angiogenic proteins such as VEGF, leading to increased angiogenesis, tumour growth, and metastatic potential by enhancing blood flow to the tumour (J. Wang & Xu, 2019).

HER2-positive breast cancers differ from other breast cancer subtypes in their molecular and genomic features. While hormone receptor-positive (HR+) breast cancer relies on signalling through estrogen (ER) and progesterone receptors (PR), HER2-positive breast cancer does not. Instead, HER2-amplified tumours depend on signalling through the HER2 receptor. This fundamental biological difference has downstream implications for the biology and response to treatment of these cancers. Unlike HR+ breast cancer, which depends on the estrogen receptor, HER2-amplified breast cancer activates the HER2/PI3K/AKT and HER2/MAPK/ERK signalling pathways (Z. Jiang et al., 2024). Activation of these pathways promotes rapid and robust tumour cell growth and survival in aggressive contexts. Additionally, studies have shown that HER2-positive breast cancers exhibit higher levels of genomic instability compared to HR+ breast cancers, often due to the frequent loss of critical tumour suppressors such as TP53 and PTEN (Nik-Zainal et al., 2016).

HER2-positive tumours uniquely benefit from efficacious HER2-directed therapies, such as trastuzumab and pertuzumab, which are largely ineffective in HR+ and triple-negative breast cancers (Metzger-Filho et al., 2021). Disease heterogeneity is further illustrated in HER2-positive breast cancer subtypes with differing hormone receptor co-expression. Among these, HER2+/HR+ tumours may exhibit distinct biology and unique responses to therapy compared to HER2+/HR- tumours (Luo et al., 2021).

1.2.4.2 Role of HER2 in tumorigenesis

The biological role of HER2 in tumorigenesis is multifaceted, including driving oncogenesis, enhancing tumour growth, promoting survival and metastasis, crosstalk with other oncogenic pathways, and mediating the tumour microenvironment. Each of these processes likely contributes to the aggressive nature and poor prognosis observed in HER2-positive breast cancer (Weigelt et al., 2010).

Activation of downstream genes following HER2 receptor activation represents one of the central consequences of HER2 amplification in breast cancer. Increased HER2 signalling, either through overexpression or gene amplification, fosters overactivation of downstream signalling networks, supporting cellular proliferation and survival. HER2

signalling is mediated by the formation of intracellular docking sites following the autophosphorylation of tyrosine residues within the intracellular domain of HER2. This promotes the recruitment of diverse signalling proteins, ultimately activating multiple downstream pathways that regulate cell physiology (Prat et al., 2014).

The phosphorylated HER2 receptor triggers downstream signalling pathways that directly contribute to tumorigenesis, growth, survival, metastasis, and resistance to therapy. In breast cancer, four general functional themes have emerged that amplify oncogenic HER2 signalling outputs. Promoting cell proliferation, the MAPK/ERK pathway positively regulates the gene expression of cell cycle regulatory proteins, such as cyclin D1. Consequently, activated ERK1/2 kinases phosphorylate and modulate downstream effectors and transcription factors, driving cells from the G1 phase to the S phase and enabling rapid proliferation. Promoting cell survival, the PI3K/AKT pathway is the major survival pathway downstream of HER2. AKT inhibits pro-apoptotic proteins such as BAD and caspase-9 while conferring a metabolic advantage through enhanced glucose uptake and glycolysis. Promoting metastasis, HER2 signalling enhances the ability of cells to leave the primary tumour, invade the extracellular matrix, and intravasate into the bloodstream to colonize distant tissues. HER2 also positively regulates the expression and activation of matrix metalloproteinases, which degrade the extracellular matrix to facilitate cell invasion. Furthermore, HER2 signalling contributes to epithelial-to-mesenchymal transition (EMT), a phenotypic shift that enhances mesenchymal features and cellular motility (Weigelt et al., 2010).

HER2 signalling is not linear but interacts with multiple other oncogenic pathways. HER2 activates PI3K through phosphorylation, converting PIP2 to PIP3, which recruits AKT to the plasma membrane and enables its phosphorylation by PDK1 and mTOR. The PI3K/AKT pathway is critical for cellular survival, growth, and metabolism. Mutations in the PI3KCA gene are common in many malignancies and can drive breast cancer independently of HER2 overexpression. Alternatively, loss of the tumour suppressor PTEN, a negative regulator of the PI3K/AKT pathway, also leads to enhanced PI3K/AKT signalling. Together, these mechanisms suggest that HER2-overexpressing cells, through either PTEN loss or PI3KCA mutations, are more likely to resist apoptosis and undergo uncontrolled proliferation. HER2 also activates the MAPK/ERK pathway, a kinase cascade that promotes gene expression and cell cycle progression. Additionally, HER2 signalling crosstalk with estrogen and progesterone receptor pathways, influencing response and resistance to HER2-targeted therapies. As a result, combination

therapies targeting HER2 and hormone receptors may be necessary for HER2+/HR+ tumours (Mercogliano et al., 2023).

HER2 also plays a critical role in regulating the tumour microenvironment, essential for tumour expansion and metastasis. HER2 signalling upregulates VEGF, a pro-angiogenic protein, leading to the development of new blood vessels that provide the tumour with oxygen and nutrients for rapid growth and metastatic dissemination. HER2-positive tumours can also exploit the immune microenvironment to avoid detection. For example, HER2 signalling upregulates PD-L1, a protein expressed on immune and tumour cells, which suppresses cytotoxic T cell activity, allowing the tumour to evade immune surveillance. HER2 signalling can also activate stromal fibroblasts, transforming them into cancer-associated fibroblasts that support tumour growth by secreting growth factors, cytokines, chemokines, and extracellular matrix proteins (Weigelt et al., 2010).

Chemokines and their receptors provide a mechanism for breast cancer to adapt and evade the anti-cancer immune response. For example, CXCL12 secreted in the tumour microenvironment is a potent chemoattractant for T cells expressing its receptor, CXCR4. When Treg cells infiltrate breast cancer through a CXCR4/CXCL12 gradient, their presence is associated with poor prognosis, particularly in triple-negative breast cancer, where Treg populations inversely correlate with CD8+ T cell infiltration and patient outcomes. Tofacitinib, a clinically proven JAK inhibitor, reduces cancer cell CXCL12 secretion in a preclinical HER2-driven breast cancer model, retaining CXCR4+ CD8+ T lymphocytes in the tumour. JAK inhibition can, therefore, promote the recruitment of effector CD8+ T cells into the tumour despite the immunosuppressive cytokine milieu (Mercogliano et al., 2023).

1.3 Cell Line Models for Breast Cancer Research

Breast cancer cell lines are important for studying the molecular causes of breast cancer and identifying the weaknesses of tumour cells to therapeutics, such as trastuzumab. Although numerous breast cancer cell lines are being developed, only a subset, including BT474, SKBR3, and MCF7, have been widely used. MCF7 is often used as a non-tumorigenic, luminal A-like, ER-positive, HER2-negative control. HER2-amplified lines, including HCC1954, JIMT-1, and ZR-75-1, have been useful models for HER2-positive breast cancer (Sigma-Aldrich, 2024). These cell lines represent the major subtypes of breast cancer and have supported many important discoveries in breast cancer biology and therapeutics.

BT474 is a luminal B model, derived from a primary ductal carcinoma. This cell line has HER2 amplification and overexpression, as well as estrogen and progesterone receptors. It has been used to understand the mechanisms of HER2 signalling, crosstalk with hormone receptors, and the ability of trastuzumab to inhibit proliferation and induce apoptosis. SKBR3, derived from a pleural effusion, also features HER2 amplification but lacks detectable hormone receptor expression. This characteristic makes SKBR3 valuable for studies of HER2-targeted therapies. MCF7 is a model for luminal A breast cancer. It was derived from a metastatic adenocarcinoma and is negative for HER2. It is typically used as a receptor-positive, HER2-negative control to demonstrate the lack of trastuzumab efficacy (Y. Pan et al., 2017).

HCC1954 carries HER2 amplification, and PIK3CA mutations. It is valuable for examining mechanisms of resistance, as well as for identifying synergistic therapeutic combinations. JIMT-1 originates from a trastuzumab-resistant tumour and is a model to understand resistance and screen for novel therapeutic options. ZR-75-1 expresses low levels of HER2 and high levels of ER and PR and is useful for studying mechanisms of hormone receptor function as well as the crosstalk between HER2 and hormone receptor signalling pathways. Comparative studies between ZR-75-1 and other cell line models also allow researchers to investigate how drug response may be influenced by the baseline receptor status of the cells under study (Hendriks et al., 2013).

These cell lines have been instrumental in advancing the understanding of breast cancer progression for drug development, like trastuzumab.

1.4 Anti-Her2 Monoclonal Antibody (trastuzumab)

1.4.1 History of Monoclonal Antibodies

A historical perspective on antibody therapeutics can be tracked back to Paul Ehrlich's reference to the "magic bullet" more than a century ago, describing a molecule that "drags the therapeutic agent directly to the cellular location where it develops its maximum effect." Nearly 80 years later, in 1975, this prediction was realized by Georges Köhler and César Milstein when they developed hybridoma technology, allowing for the production of monoclonal antibodies (mAbs) (Köhler & Milstein, 1975). This pioneering work was recognized with the Nobel Prize in Physiology or Medicine in 1984.

The first monoclonal antibody (mAb) approved for use as a therapeutic was Orthoclone OKT3 (muromonab-CD3), which was granted FDA approval in 1986 to reduce the risk of rejection among kidney transplant patients. Although Orthoclone OKT3 was effective, its use was limited by immunogenicity, a general limitation of murine antibodies (Waldmann, 1984). To overcome these limitations, three significant advances in antibody engineering emerged:

1. Humanized antibodies were engineered by modifying murine sequences to reduce immunogenicity (RapidNovor, 2024).
2. Chimeric antibodies were produced by assembling DNA from mouse and human sources (Sino Biological, 2024).
3. Fully human antibodies were developed using transgenic mice or phage display systems (Lonberg, 2008).

A major milestone was reached in 1997 with the FDA approval of Rituxan (rituximab), a monoclonal antibody for cancer therapy. Since then, monoclonal antibodies have been approved for a variety of conditions, including cancers, autoimmune diseases, and inflammatory diseases. Worldwide, the monoclonal antibody market is now estimated to exceed \$20 billion (Ecker et al., 2015). Current areas of research include identifying new targets, increasing therapeutic efficacy, preventing resistance, and developing novel therapies, such as bispecific antibodies and antibody-drug conjugates.

1.4.2 Monoclonal Antibody Structure

Monoclonal antibodies (mAbs) are highly specific therapeutic proteins engineered to recognize predetermined antigens. MAbs are derived from a single ancestral immune cell that recognizes an antigen with defined specificity and binding properties. This immune cell is fused to immortalized myeloma cells to generate a clone that produces mAbs (Zahavi & Weiner, 2018). Structurally, mAbs are composed of two identical heavy chains connected by disulfide bonds to two identical light chains, forming a Y-shaped structure. The two arms of the Y are the variable regions responsible for recognizing the antigen of interest. The stem of the Y structure forms the constant region, which interacts with immune cell receptors to mediate responses and with Fc receptors on non-immune cells to facilitate targeted killing of antigen-expressing cells (Suzuki et al., 2015). Because of their precision in targeting antigens, mAbs have become essential agents in cancer therapy.

Monoclonal antibodies are classified according to their source and structure, both of which impact their therapeutic properties and immunogenicity (Scott et al., 2012):

- Murine antibodies (e.g., muromonab-CD3) contain no human sequences and elicit strong immune responses in humans.
- Chimeric antibodies (e.g., rituximab) contain mouse variable regions and human constant regions, reducing immunogenicity.
- Humanized antibodies (e.g., trastuzumab) are predominantly human but retain mouse antigen-binding sites, further lowering immunogenicity.
- Fully human antibodies (e.g., adalimumab) are derived entirely from human sources and exhibit minimal immunogenic potential.

1.4.3 Production of mABs

Producing therapeutic mAbs involves several critical steps:

1. Antigen Selection and Expression

The process begins with selecting an antigen associated with the disease of interest, such as HER2 in breast cancer. After selection, the antigen is isolated, purified, and expressed in high yields using recombinant DNA technology.

2. Hybridoma Technology

Laboratory animals are immunized with the antigen, initiating antibody production. Spleen cells from the immunized animal are harvested and fused with myeloma cells to create hybridomas capable of producing antigen-specific antibodies. These hybridomas are screened for the production of the desired antibody (Schofield, 2023).

3. Cloning and Expansion

Promising hybridomas are selected, subcloned to ensure clonality, and expanded in bioreactors under optimized conditions to produce large quantities of the monoclonal antibody (Schofield, 2023).

4. Antibody Purification

The antibody-containing culture medium is harvested, and antibodies are typically purified using Protein A affinity chromatography. Protein A binds with a high affinity to the Fc region of most IgG subtypes. In this method, Protein A is immobilized on a chromatography column to enable selective capture. This approach efficiently isolates IgG-class antibodies. To improve purity, charge variants can be removed using ion exchange chromatography and antibody aggregates can be eliminated with size exclusion chromatography.

5. Characterization and Quality Control

Purified antibodies undergo extensive characterization to confirm structure, binding affinity, and specificity. Analytical techniques such as mass spectrometry and X-ray crystallography are used to confirm the structural integrity of the antibody. Additionally, quality control tests are performed to ensure sterility, bioactivity, and safety.

6. Formulation and Stability Testing

Antibodies are formulated into stable pharmaceutical products by optimizing pH, buffers, and excipients. Stability studies are conducted to determine the effectiveness of the product under various storage conditions (Kelley, 2020).

7. Clinical Trials and Regulatory Approval

mAbs are tested in preclinical investigations and clinical trials (Phases I–III) to evaluate safety, efficacy, and dosage. Regulatory agencies, such as the FDA or EMA, review these data before granting approval (Kelley, 2020).

1.4.4 Therapeutic mAB Advancements

Substantial progress has been made in developing therapeutic monoclonal antibodies to address the limitations of the early murine antibodies. The primary goals of mAb therapeutics described here were to reduce immunogenicity, improve efficacy, and expand the range of potential applications. One of the challenges of early mAb therapeutics emerged from what was known as the human anti-mouse antibody response (HAMA). HAMA can have clinically meaningful impacts such as neutralizing the mAbs, making them less effective, or even causing an allergic or other serious adverse reaction. Researchers have primarily addressed this problem through the development of humanization techniques (Clavero-Álvarez et al., 2018).

The use of chimeric antibodies was an important advancement combining the variable regions derived from mouse antibodies with human constant regions. Chimeric antibodies could be produced that retained the murine specificity yet were substantially less immunogenic in clinical settings. The next variation on this technology was CDR grafting. As the name suggests, in CDR grafting one transfers the CDR's from a mouse antibody onto a human antibody framework. A second variation was framework shuffling in which researchers combine human antibody framework regions with CDRs from antibodies of a non-human species in an iterative fashion (Dall'Acqua et al., 2005). A third method was resurfacing, replacing non-human surface-exposed domains.

Addressing the challenge of completely human monoclonal antibodies required two technological breakthroughs: transgenic mice and phage display systems. One option for transgenic mice was to genetically manipulate mice to produce human antibodies. XenoMouse technology is an example of this approach, which was used to generate monoclonal antibodies, such as Panitumumab. This allowed for completely human antibodies to be generated using traditional hybridoma techniques with far less concern for immunogenicity. Phage display systems enabled the isolation of fully human antibodies from large libraries without requiring animal immunization (Lonberg, 2008).

Computational techniques have continued to advance and transform the field of therapeutic antibody development. For example, *in silico* humanization tools like HumAb and the BioPhi platform, make use of machine learning and transformer algorithms to improve antibody sequences, identify CDR residues that can be changed to retain or improve specificity, predict potential immunogenicity, and optimize overall antibody fitness. These tools support further antibody humanization by enabling *in silico*

prediction and optimization of sequence properties before experimental validation (Marks & Deane, 2024).

A final area of investigation to improve mAb effectiveness is the development of alternative antibody formats. Two emerging formats, bispecific antibodies and antibody-drug conjugates, are broadening the therapeutic horizons of mAbs. Bispecific antibodies with two unique antigen binding sites can target two different antigens at the same time, which allows for new targets in cancer immunotherapy, such as activating immune cells only near cancer cells. Conversely, antibody-drug conjugates can combine the specificity of mAbs with the cytotoxicity of small molecules, allowing for targeting of potent therapeutic agents to a specific cell (Mercogliano et al., 2023).

Furthermore, there is ongoing research to enhance the efficacy of antibody-mediated therapies and prevent the development of resistance. New targets are being identified, and new ways are being devised to maximize antibody-dependent cellular cytotoxicity ADCC. ADCC describes an immune effector mechanism in which immune cells, especially natural killer (NK) cells, recognize and eliminate target cells bound by antibodies. Understanding and overcoming resistance mechanisms are critical for the long-term success of mAb therapies. Improvements in pharmacokinetics remain a priority for the field, and technologies for half-life extension and methods to enhance tissue penetration are being developed that should lead to more durable treatments with greater efficacy, and potentially enabling less frequent dosing and, ultimately, a better patient outcome. Advances in these areas will be moving the field from where it stands today, toward the development of increasingly more potent, safe, and effective antibody-based therapeutics.

1.4.5 Effects of Therapeutic mAbs

Therapeutic monoclonal antibodies (mAbs) target specific antigens, which can cause a range of downstream effects depending on their mechanism of action. Broadly, these downstream effects derive from three major phenomenological categories: direct effects, immune mediated effects, and synergistic effects when combined in combination therapies (Zahavi & Weiner, 2018). MAb can also induce direct apoptosis, as in the case of Rituximab, used to treat B-cell lymphomas. Finally, mAbs can engage the immune system to lyse cancer cells through two main effector mechanisms—antibody-dependent cellular cytotoxicity and complement-dependent cytotoxicity. Antibody-dependent cellular cytotoxicity occurs when natural killer (NK) cells recognize the Fc region of an

mAb bound to a tumour cell, leading to release of cytotoxic granules that lyse the bound tumour cell. Trastuzumab and rituximab are two mAbs that utilize ADCC. Complement-dependent cytotoxicity on the other hand, occurs when the complement system is activated through binding of a molecule to a target cell. Activation of the complement cascade results in the formation of the membrane-attack complex, leading to lysis of the target cell. Of note, the ability of rituximab to induce complement is overridden by its greater ability to induce ADCC (Scott et al., 2012).

Monoclonal antibodies can also synergize with other anti-cancer treatments, including chemotherapy and radiation therapy, to enhance efficacy above that expected by summing their independent effects. A number of mechanisms may explain these synergistic effects, including:

- Cytotoxicity: mAbs can enhance the cytotoxicity of chemotherapy agents. Trastuzumab, for example, enhances the efficacy of docetaxel for HER2-positive breast cancer.
- Enhanced drug delivery: Antibodies such as bevacizumab (Avastin), which inhibit vascular endothelial growth factor (VEGF), may also normalize tumour vasculature. This is predicted to facilitate more efficient delivery of chemotherapeutic agents to the tumour.
- Radiosensitization: Cetuximab was also reported to potentiate the effects of radiation therapy in head and neck cancers, presumably by decreasing DNA repair in cancer cells.
- Immune checkpoint blockade: The combination of mAbs that target different elements of the immune system, e.g., PD-1 inhibitors (e.g., pembrolizumab) with CTLA-4 inhibitors (e.g., ipilimumab) can synergize to induce more potent antitumour immune responses (Weigelt et al., 2010).

Despite their high target specificity, mAbs can have off-target effects and adverse reactions. mAbs have the potential to cause a wide variety of adverse effects (AEs). At the extreme end are immune-related adverse events, as well as a milder set of infusion reactions. Skin reactions, including rashes, have been the most frequently reported early AEs to draw notice, and most cases can be managed with appropriate topical treatments.

Some of the more serious AEs associated with mAbs include cardiotoxicity, particularly with trastuzumab; severe infections associated with rituximab or any other mAb that depletes B cells; and immune-related adverse events, particularly with checkpoint inhibitors given alone or in combination therapies (Hansel et al., 2010). Some of these AEs, such as cardiotoxicity, have led to detailed guidance on monitoring patients on these agents and the intervention to perform when an AE does occur. Efforts to ensure proper clinical care of patients receiving mAbs will need to continue to rely on the quest for biomarkers that may predict the likelihood of an AE. In sum, the diverse mechanisms of action of therapeutic monoclonal antibodies, spanning from direct tumour cell inhibition to immune system modulation and synergistic effects in combination therapies, underlie the fundamental importance of these agents in modern cancer care. The challenges that remain, especially in managing their off-target effects, will be met in part through ongoing development of mAb molecules and an increasing understanding of how best to mitigate these complications in the future.

1.4.6 Types of Therapeutic Monoclonal Antibodies

Altogether, targeted protein antagonists account for only two of the nine possible categories in which targeted therapy options can be placed.

Unconjugated antibodies constitute the majority of approved mAbs and lack any conjugated molecules. This class is comprised of several subtypes on the basis of their structure and derivation. Humanized and fully human mAbs, e.g. Trastuzumab (Herceptin) and adalimumab (Humira), are engineered to reduce immunogenicity and improve tolerability in patients. Chimeric antibodies such as rituximab (Rituxan) are composed of both mouse- and human-derived antibody components. Unconjugated antibodies can function in diverse ways, including to block growth signals, to direct the immune system to cellular targets and to disrupt immune inhibitory checkpoints to facilitate anti-tumour responses (Zahavi & Weiner, 2018).

Bispecific antibodies are a category of novel naked mAb that are able to bind two separate antigens simultaneously. For example, blinatumomab is a bispecific antibody that links cancerous B cells expressing CD19 to CD3-expressing T cells, thereby promoting the T cells to the destruction of the tumour cells (Klein et al., 2024).

Although naked monoclonal antibodies have demonstrated some success, their potency can be substantially increased by conjugation. Conjugates of mAbs are used to link other molecules to the antibody and thereby enhance its effects, or to deliver potent

agents directly to cancer cells. The two primary classes of conjugated mAbs are Antibody-Drug Conjugates and radiolabelled antibodies (Suzuki et al., 2015).

ADCs comprise three components: a monoclonal antibody (mAb) that can recognize and bind to a tumour, a potent drug capable of killing cells, and a linker – the connection between the mAb and the drug. The mAb in an ADC is able to selectively bind to antigens expressed on the surface of tumours but not on most normal cells. This means that the mAb can be developed as if it were a guided missile that will deliver the toxin only to cancer cells, but not to healthy cells. Upon binding the antigen, the antibody and antigen complex is rapidly internalized by endocytosis. Within the cell, the linker is susceptible to cleavage, liberating the cytotoxic load, which ultimately causes the cells to die (J. Lu et al., 2020).

A number of ADCs have been approved by regulatory authorities and have shown clinical benefit. Two examples of ADCs include ado-trastuzumab emtansine (Kadcyla) approved for HER2-positive breast cancer, which is conjugate of trastuzumab and the microtubule inhibitor DM1. Adcetris, which has been approved for several subtypes of lymphomas, is a combination of an anti-CD30 antibody and an antimetabolic agent, monomethyl auristatin E. Most recently, another ADC fam-trastuzumab deruxtecan-nxki (Enhertu) has shown an especially profound effect in HER2-positive cancers, which is a combination of the antibody-trastuzumab and a topoisomerase I inhibitor (DXd) (Mercogliano et al., 2023).

Radioactive isotopes can be coupled to the antibody to create radiolabeled antibodies, or radioimmunoconjugates. When administered to patients, these antibodies can deliver targeted radiation selectively to the malignant cells. This radioimmunotherapy strategy allows for the delivery of high doses of radiation directly to the tumour while minimizing radiation exposure to surrounding normal tissues (Suzuki et al., 2015).

There are two radiolabeled antibodies that make up the majority of the current clinical experience. Ibritumomab tiuxetan (Zevalin), directed against CD20, is chelated to yttrium-90 or indium-111 and used in the management of specific B-cell non-Hodgkin lymphomas. The anti-CD20 antibody tositumomab was also radiolabeled to iodine-131 for the treatment of similar diseases; however, the product has been discontinued (Zahavi & Weiner, 2018).

The development of antibody-drug conjugates continues to be an active area of research, with many candidates advancing to the clinic. Advances in linker technology, payload selection, and antibody engineering will continue to drive progress in the field.

These advances are generally aimed to improve the therapeutic index of ADCs, increase their circulating half-lives, and allow for the treatment of a wider range of cancers and other diseases (J. Lu et al., 2020).

In summary, although the field is still dominated by unaltered full length naked antibodies, conjugates of antibodies with drugs or effector cells represent a growing and highly promising class of mAb drugs. By combining the high specificity of antibodies with the potent killing ability of cytotoxic agents or radiation, these conjugates promise to provide a direct route to targeted therapy, with the potential benefits of enhanced efficacy and reduced systemic toxicity (Mercogliano et al., 2023).

1.4.7 Trastuzumab

1.4.7.1 Development

The development of trastuzumab has been a critical milestone in the progress of targeted cancer therapy, especially for the management of breast cancer. Trastuzumab can be traced back to research on oncogenes that began in the mid-1980s (Maadi et al., 2021a). In 1984, the HER2 oncogene was discovered in chemically induced rat neuroblastomas, and the human homologue was cloned a year later by Lisa Coussens and colleagues. These studies led to the important discovery that the HER2 gene was amplified and/or overexpressed to the points of being pathological in as many as 20-30% of human breast tumours, making HER2 an excellent target for therapy (Sawyers, 2019).

The efforts to generate an effective antibody directed against HER2 began in 1989 when investigators produced a mouse monoclonal antibody, 4D5, that specifically bound HER2. However, using mouse antibodies in humans proved to be problematic because of concerns about immunogenicity. In 1992, in an effort to avoid this problem, Carter and colleagues humanized the 4D5 antibody, resulting in trastuzumab (Bradley et al., 2021).

Clinical development of trastuzumab moved very rapidly, consistent with the important clinical unmet need for effective HER2-targeted breast cancer therapies. The dose-escalating tracer trial of 15 women was initiated in 1992, and the success of this trial from 1996 prompted the initiation of larger Phase II and Phase III trials in over 900 women. In fact, results from these trials were so encouraging, that Genentech implemented a lottery to directly allocate drug to patients and oncologists on an open access basis, to meet heavy demand (Sawyers, 2019).

In recognition of its potential, the U.S. Food and Drug Administration granted accelerated approval for trastuzumab, and in September 1998 the FDA approved trastuzumab for the treatment of metastatic HER2+ breast cancer. In the wake of these findings, trastuzumab was approved in Europe in August 2000. Together, these approvals constituted global recognition of trastuzumab's antitumour effects (Zahavi & Weiner, 2018).

Thus far, the potential impact of trastuzumab on the treatment of breast cancer has been enormous. Trastuzumab has improved the survival of women with HER2-positive metastatic breast cancer and has significantly reduced the risk of recurrence in this population as well as in women with early disease. More fundamentally, trastuzumab altered the field of oncology by providing a high-profile example of the efficacy of biomarker-driven drug discovery and development, heralding the dawn of an era of personalized cancer care (Zahavi & Weiner, 2018).

The success of trastuzumab provided a springboard for further developments in HER2 targeted therapies. In 2008, trastuzumab emtansine was generated as an antibody-drug conjugate combining the targeting specificity of trastuzumab with a highly potent cytotoxic agent. This represented a whole new approach to targeted cancer therapy (Zahavi & Weiner, 2018).

As the patents for trastuzumab have expired – within Europe in 2014 and in the United States in 2019 – and biosimilars have been developed, potentially allowing for greater access to this life-saving treatment. These two events highlight the on-going importance of trastuzumab within the field of oncology (Bradley et al., 2021).

In summary, the story of trastuzumab is a powerful example of the triumph of translational research in oncology. Tracing its roots back to basic science research nearly 40 years ago, through to its current role in clinical practice, trastuzumab has changed, the treatment of HER2-positive breast cancer, and has created a new model for targeted cancer therapies more broadly. As such, the history of trastuzumab has had a profound and unambiguous effect on drug development strategies in oncology across the board (Bradley et al., 2021).

1.4.7.2 Indication and Clinical outcome

Trastuzumab is approved for HER2-positive breast cancer, for both metastatic and early-stage breast cancer. Accurate assessment of the HER2 status of the tumour is

critical for selection of appropriate therapy and is commonly done with immunohistochemistry and/or in situ hybridization techniques. This allows treatment to be targeted to patients most likely to benefit (Pinto et al., 2013).

In the metastatic setting, trastuzumab has been highly effective. Clinical trial data reveal that trastuzumab increases overall survival from 20.3 to 25.1 months in patients with metastatic HER2+ breast cancer. The combination of trastuzumab and chemotherapy increases overall survival and response rates compared to trastuzumab alone, suggesting the added benefit to this combinatorial strategy (Gajria & Chandarlapaty, 2011).

For early-stage breast cancer, trastuzumab is a highly effective adjuvant treatment. On average, the treatment leads to an estimated 9.5% absolute reduction in the risk of recurrence within three years. The number needed to treat with trastuzumab to prevent one recurrence within three years calculated from this rate of recurrence is 10, which is comparable to other adjuvant systemic therapies for early-stage breast cancer. Trastuzumab also has an effect on preventing death within three years of 3% in absolute terms. This effect, however, varies widely across studies. Nevertheless, even with this wide variation, trastuzumab has been statistically significantly associated with improved rates of long-term survival across multiple investigations (Tarantino et al., 2022).

Long-term outcome data continue to demonstrate the efficacy of trastuzumab. For example, the 9-week versus 1-year trastuzumab treatment arms of the SOLD trial, with a median follow-up of 8.1 years, revealed that a full 1-year of trastuzumab treatment provided significantly better results than just 9-weeks of therapy. This trial was seminal in determining the appropriate length of time that trastuzumab therapy should be administered to patients with early-stage breast cancer.

The ability of trastuzumab to be combined continues to extend beyond breast cancer. For example, combining trastuzumab with other targeted therapies, specifically lapatinib, has been effective in other HER2-positive cancers, such as colorectal cancer, thereby expanding potential utility in oncology (Bradley et al., 2021).

However, trastuzumab has associated risks. It raises the absolute risk of severe cardiac event by 2.1%, requiring regular cardiac monitoring for the early detection of this adverse event. Fortunately, most cardiac events are not life threatening and can be resolved with the discontinuation of the drug, making it possible for women to balance their risk against its benefits (Pinto et al., 2013).

Despite the compelling clinical benefits observed in adjuvant and metastatic populations, trastuzumab has been plagued by issues of treatment resistance. For instance, even though first-line trastuzumab makes up the most efficacious and widely used treatment for HER2-positive BC, approximately 70% of patients with HER2-positive disease remain unresponsive to the therapy. In addition, resistance typically develops rapidly in nearly all patients after a period, further emphasizing the importance of ongoing studies to elucidate the mechanisms of resistance and identify strategies to overcome it (Gajria & Chandarlapaty, 2011).

In sum, trastuzumab has provided a significant benefit to HER2-positive breast cancer patients in both the metastatic and adjuvant settings. The benefits of survival advantage, risk reductions for recurrence, and synergy of action with other treatments have established trastuzumab as a standard component of therapy for HER2-positive breast cancer. However, the continued challenges of cardiac toxicity and treatment resistance underscore the importance of thoughtful patient selection, monitoring, and ongoing research to optimize its use and develop new approaches to resistance (Maadi et al., 2021a).

1.4.7.3 Structure and Interaction with Her2

At its core mechanism of action, trastuzumab binds to domain IV of the extracellular segment of the HER2 receptor with a relatively high affinity ($K_d \approx 5$ nM). This binding involves domain IV in three separate regions, combining electrostatic and hydrophobic interactions (Maadi et al., 2021a). Despite the initial hopes, trastuzumab does not block HER2 homodimers, nor heterodimers with other HER family receptors as were initially expected. Indeed, several studies have suggested that trastuzumab may modestly promote HER2 homodimerization, particularly at higher doses and longer exposure times. Moreover, the binding of trastuzumab does not impede phosphorylation of HER2 at most tyrosine residues, and in some instances may even promote HER2 phosphorylation.

The specificity of trastuzumab's interaction with HER2 is striking. At saturating concentrations, trastuzumab does not bind to only EGFR or HER3, but also to other members of the HER family, despite their far more significant sequence similarities to the HER2 ectodomain itself. This specificity implies that the effects of trastuzumab on

cells that express more than one HER receptor are primarily those mediated by its effects on HER2 (Weigelt et al., 2010).

Of interest, the binding of trastuzumab does not significantly influence the internalization and recycling of HER2 receptors itself. Thus, HER2 receptors are still being internalized and recycled back to the plasma membrane, with trastuzumab on board during this process. Trastuzumab was detected in internalized vesicles, but these vesicles primarily derive from recycling endosomes, rather than degradative compartments (Pinto et al., 2013).

Moreover, the clinical synergy observed upon combining trastuzumab with pertuzumab, a second anti-HER2 antibody, does not result from enhanced binding affinity or global conformational rearrangements of HER2. Recent Cryo-EM studies have demonstrated that both antibodies can bind to HER2 simultaneously with very little conformational rearrangement, indicating that the combinatorial binding of the two sets of extracellular domains or the geometry of the overall complex is sterically possible and certainly does not explain the mechanism of synergy (Pinto et al., 2013). Despite the ability of trastuzumab to interact with HER2 specifically and with high affinity, its biological activity is mediated through a series of downstream effects on HER2 signalling rather than through direct inhibition of HER2 dimerization (Scott et al., 2012).

1.4.7.4 Effect of Trastuzumab

1.4.7.4.1 Direct Cellular Impact

Trastuzumab has a profound impact on the intracellular processes of HER2-positive breast cancer cells, affecting multiple pathways and machineries that are critical to proliferation, cell survival and signalling. One of the most prominent intracellular effects of trastuzumab is its effective and persistent inhibition of the PI3K/AKT signalling pathway, which is essential for cell survival and growth. This inhibition is achieved through multiple mechanisms, such as the inhibition of AKT phosphorylation. Trastuzumab can also induce the inhibition of HER3 phosphorylation, a preferred dimerization partner of HER2, or activation of PTEN, a known tumour suppressor that negatively regulates the PI3K pathway. This downregulation of this pathway results in a decreased signal for cell survival, and increased signals of susceptibility to apoptosis in these cells (Vu & Claret, 2012).

Another major consequence of trastuzumab treatment is the ability to stimulate cell cycle arrest. Specifically, trastuzumab treatment results in an increase in the percentage of cells in the G0/G1 phase of the cell cycle, largely because of trastuzumab's effect on the cyclin-dependent kinase inhibitor p27KIP1. Trastuzumab promotes an increase in both the expression and the subcellular localization of p27KIP1, allowing p27KIP1 to more effectively inhibit cyclin-CDK complexes and stop the cell cycle. Trastuzumab also acts directly on the cell cycle through the suppression of numerous genes governing the cell cycle and DNA replication. This directly slows DNA synthesis and cellular proliferation, promoting a response that is consistent with the observed accumulation of cells in G0/G1 (Baselga & Albanell, 2001).

Trastuzumab has also been shown to have an impact on the intracellular HER2 receptor dynamics. Herceptin does not significantly affect the total internalization and recycling of HER2, but it has been suggested to cause a small increase in HER2 internalization. More significantly, trastuzumab has been shown to induce HER2 degradation through interaction with an E3 ubiquitin ligase, c-Cbl. This leads to the ubiquitination and subsequent proteasomal degradation of the HER2 receptor. The role and significance of this mechanism in the overall anti-tumour activity of trastuzumab remain unclear and require further studies (Weigelt et al., 2010).

In addition to these direct effects, trastuzumab may also regulate the number of proteins contributing to apoptosis, indirectly favouring pro-apoptotic factors. Indeed, treatment with trastuzumab has been associated with increased protein levels of pro-apoptotic molecules, such as BAX, as well as decreased anti-apoptotic proteins, such as BCL-2. Consequently, this may tip cancerous cells more toward an apoptotic fate in response to subsequent stimuli (Pinto et al., 2013).

Trastuzumab also modulates cellular metabolism. The cells in HER2-positive cancer frequently have abnormal metabolic profiles such as increased glycolysis and lipid synthesis. Emerging evidence suggests that trastuzumab may in part correct some of these metabolic abnormalities, which could play a role in its anti-tumour efficacy. Trastuzumab has been shown to reduce glucose uptake and lactate formation, markers of glycolytic activity, and can also alter the expression of key enzymes involved in metabolism (Scott et al., 2012).

Lastly, trastuzumab has been demonstrated to have effects on intracellular signalling networks in addition to the PI3K/AKT pathway. For example, trastuzumab has been shown to have effects on other signalling cascades, such as the MAPK pathway. These effects seem to be highly context- and cell-type specific (Jeyakumar & Younis, 2012).

In summary, the intracellular consequences of trastuzumab are numerous and profound. Through its actions on signal transduction, cell cycle, receptor trafficking, apoptosis and metabolism, trastuzumab has a broad intracellular antitumour impact. The combination of these diverse actions accounts for the clinical success of the agent and emphasize the complexity of therapeutic strategies for cancer based on molecular targeting (Arnould et al., 2006; Gajria & Chandarlapaty, 2011; Mercogliano et al., 2023).

1.4.7.4.2 Extracellular Effects

The extracellular effects of trastuzumab are predominantly mediated through its ability to alter the immune system's response to HER2-positive cancer cells. These effects are integral to the efficacy of the drug and underscore the relevance of the immune system to cancer therapy. A central extracellular mechanism of trastuzumab is its ability to promote antibody-dependent cellular cytotoxicity. In this process, trastuzumab binds to HER2 receptors on the surface of cancer cells, and then the Fc region of trastuzumab recruits and activates immune effector cells, especially natural killer cells, through ligation to Fc gamma receptors on their surface. This engagement triggers a signal transduction cascade within NK cells that leads to activation of the SOS/RAS/ERK/p38/MAPK signalling pathway and ultimately to the generation of phosphatidylinositol 4,5-bisphosphate [PI(4,5)P₂]. Activation of NK cells results in the release of cytotoxic granules containing perforin and granzymes, which induce apoptosis of HER2+ cancer cells (Nuti et al., 2011).

Trastuzumab also has immunomodulatory effects that increase the adaptive immune response to HER2+ tumours, which may be more pronounced in some adjuvant study populations. Trastuzumab has been shown to cause internalization of HER2, promoting uptake of HER2 by dendritic cells in a manner that depends on the dendritic cell-FcγRI complexes with HER2. Trastuzumab binding to HER2 has been associated with increased uptake of the HER2 by dendritic cells, thus enhancing antigen presentation (Scott et al., 2012). This suggests that binding of trastuzumab to HER2 on the surface of the tumour cell may benefit immune surveillance by anti-tumour circulating

lymphocytes. Recent studies have also shown associations with trastuzumab treatment and reduced circulating regulatory T-Cells, and an increased ratio of T-regulatory cells to Th17 cells, both of which are known to have potentially suppressive effects on tumour immunity and may be expected to be found in a less favorable anti-tumour immune environment (Pinto et al., 2013).

Another important extracellular function of trastuzumab is its role in promoting a sustained immune response. The cancer cell apoptosis triggered by NK cells generates apoptotic bodies, which are highly phagocytosable by antigen presenting cells. In this process, the CTLs traveling in the blood specific for tumour-associated antigens will expand and acquire new epitopes. This process has been termed epitope spreading, which leads to the expansion of a host immune response to tumour-associated antigens beyond the original HER2 target. The new immune responses generated by this process may potentiate and enhance the therapy (Suzuki et al., 2015).

trastuzumab has additional effects on the tumour microenvironment through its modulation of angiogenesis. Indeed, trastuzumab has been associated with a decrease in vascular endothelial growth factor secretion, which results in the inhibition of the formation of new blood vessels that the tumour relies on for oxygen and nutrients. This anti-angiogenic effect of trastuzumab also adds to its overall anti-tumour effects (Weigelt et al., 2010).

Curiously, the extracellular effects of trastuzumab are most prominent in low-baseline immune infiltrate tumours. In these cases, the effects of trastuzumab treatment led to robust increases in immune-related gene expression, including MHC-II genes, interferon-related genes, and T-cell markers. These findings may suggest that trastuzumab can have the most impact in “cold” tumours, with the potential to convert them to “hot” tumours that are more susceptible to immune clearance (Zahavi & Weiner, 2018).

The extracellular effects of trastuzumab are also evident by changes in circulating immune cells. Patients that have a clinical response to trastuzumab-containing therapy also generally have a change in their circulating leukocyte counts, namely a decrease in circulating neutrophils has been commonly observed. This peripheral immune change may be a potential biomarker of trastuzumab response.

In summary, the extracellular actions of trastuzumab are diverse and impact its overall anti-tumour activity. Through activation of both the innate and adaptive immune

systems, effects on the tumour microenvironment, and the promotion of long-lasting immune responses, these extracellular functions of trastuzumab act in concert with the direct effects on HER2-positive cancer cells to make a significant contribution to the clinical effectiveness of trastuzumab (Gajria & Chandarlapaty, 2011; Vu & Claret, 2012).

1.4.7.4.3 Resistance Mechanisms

Despite the significant clinical benefits of trastuzumab for the treatment of HER2-positive breast cancer, the development of resistance is a significant problem. As many as two-thirds of patients either do not respond to the initial trastuzumab treatment or will eventually develop resistance after their initial response. Acquired resistance mainly emerges because of prior treatment with the drug and is reversible upon discontinuation of the drug. Several key mechanisms of trastuzumab resistance have been uncovered (Wong & Lee, 2012).

One way in which this might happen is through alterations to the HER2 receptor itself. Some breast cancer cells make a shorter, extracellular region-lacking version of HER2 termed p95HER2. Like full-length HER2, truncated p95HER2 is constitutively active and trastuzumab-resistant. Mucin4, which is overexpressed in a subset of breast tumours, can bind to HER2 and prevent the binding of trastuzumab to its target by sterically hindering where trastuzumab binds (Tolaney & Krop, 2009).

Resistance can also develop through the upregulating of alternative signalling pathways that circumvent the inhibition of HER2:

- **PI3K/AKT Pathway Activation:** Activating mutations of the PI3K catalytic subunit or loss of the tumour suppressor PTEN can result in constitutive activation of the PI3K/AKT pathway, a major downstream effector of HER2 signalling. Constitutive activation of this pathway can confer resistance of cells to HER2 inhibition by trastuzumab.
- **Enhanced IGF-1R Signalling:** The upregulation of insulin-like growth factor 1 receptor IGF-1R is one mechanism by which cells can acquire resistance to HER2 therapy. IGF-1R interacts with HER2 to form heterodimers, and signals through many of the same downstream pathways as HER2. This allows the cell to survive and proliferate despite the trastuzumab therapy.

- Activation of other HER Family Members: Increased signalling through EGFR (HER1) or HER3 may represent alternative proliferative and survival signals. HER2 heterodimers are powerful activators of the PI3K/AKT pathway and therefore may also contribute to trastuzumab resistance.

Trastuzumab presumably exerts some of its effect through induction of antibody-dependent cellular cytotoxicity (ADCC). Changes that affect this immune-mediated response, such as polymorphisms that block the activity of Fc gamma receptors on immune effector cells required for ADCC, could make ADCC less efficient, thereby potentially blunting the antitumour effects of trastuzumab (Gajria & Chandarlapaty, 2011).

Resistance can also arise from deregulation of cell cycle control and apoptosis pathways. Down-regulation of p27KIP1, a critical effector of trastuzumab-mediated cell cycle arrest, has been associated with resistance. Cells that overexpress anti-apoptotic proteins or underexpress pro-apoptotic proteins, such cells will fail to undergo programmed cell death despite exposure to trastuzumab (Wong & Lee, 2012).

Some resistance mechanisms may arise as adaptive responses to sustained HER2 inhibition. These mechanisms can include increased receptor tyrosine kinase expression, upregulation of HER2 ligands, or the activation of feedback loops that will re-establish HER2 signalling (Gajria & Chandarlapaty, 2011).

To promote strategies to overcome or prevent trastuzumab resistance, recognizing these mechanisms of resistance is essential. Proposed strategies under investigation to combat this resistance include combination therapy against multiple HER family members or downstream pathways, antibody-drug conjugates such as T-DM1, and approaches to enhance immune-mediated effects. Ongoing investigations to identify biomarkers of resistance and optimize tailored treatment strategies will aid in improving the outcomes of patients with HER2-amplified breast cancer (Wong & Lee, 2012).

1.5 Biosimilars

1.5.1 Introduction

Biosimilars constitute a major technological advance in biopharmaceuticals that could greatly increase access to, and lower the cost of, critical life-saving biological therapies (DiPaola, 2017). By definition, a biosimilar is highly similar to an authorized biological

medicine, which we will refer to as the originator product. The concept of biosimilars has attracted considerable attention, especially with respect to monoclonal antibodies, a class of therapeutic proteins that has been growing rapidly (C. Liu & Morrow, 2016).

The primary reason that biosimilar development is much more difficult than the development of small molecule generic drugs is that small molecules can be exactly reproduced because they have well-defined chemical structures (C. Liu & Morrow, 2016). In contrast, the size, molecular weight, and complexity of biologic molecules prevent them from being perfectly duplicated. Biologics are composed of a core protein structure surrounded by a heterogeneous mixture of post-translational modifications. A single IgG molecule, the molecular scaffold for a number of currently marketed therapeutic monoclonal antibodies, is a 150 kDa molecule composed of four peptide chains (C. Liu & Morrow, 2016).

The complexity of biosimilars arises from several factors:

1. **Structural Complexity:** Biosimilars are produced by living organisms, resulting in products that are not simple molecular entities, but collections of closely related species (C. Liu & Morrow, 2016).
2. **Manufacturing Sensitivity:** Biosimilar production is highly sensitive to the cell culture conditions, where even minor changes can lead to variability in the final product (C. Liu & Morrow, 2016).
3. **Post-Translational Modifications (PTMs):** The particular PTMs that proteins undergo, such as glycosylation, can also vary depending on how the protein is assembled, and those PTMs can in turn impact the biological activity, pharmacokinetics, and immunogenicity of that protein (Kang et al., 2021).
4. **Analytical Challenges:** The high degree of structural complexity inherent to biological products implies that biosimilars of these products must be subjected to numerous analyses using a variety of analytical techniques to confirm similarity to the reference product (C. Liu & Morrow, 2016).

The primary goal of biosimilarity development is to demonstrate there are no clinically meaningful differences between the biosimilar and the reference product with respect to quality, safety, and efficacy (Walsh, 2018). This is achieved through extensive comparability assessments, composed of a series of analytical, non-clinical, and clinical studies (C. Liu & Morrow, 2016).

The development of biosimilars has important implications for drug manufacturers, patients, and health care systems. First, by offering potential alternatives to established biologic therapies at a potentially lower price, biosimilars could increase patient access to these essential therapies that have remained inaccessible to many patients. Second, competition from biosimilars may generate spillover benefits leading to continued improvement in the drug development and manufacturing process for all biologics (DiPaola, 2017). As the field of biosimilars evolves, regulatory guidelines are developing, tools for analytics are improving, and with increasing experience, sponsors are better able to address the challenges of biosimilar development.

1.5.2 Regulatory Framework for Biosimilars

After their introduction, the regulatory environment for biosimilars has developed significantly, and regulatory agencies around the world have created comprehensive strategies to ensure the quality, safety, and efficacy of the complex biological products. Overall, the goal has been to create an appropriate regulatory pathway for approval of biosimilars, while maintaining high standards of all biologic products. Although individual regulations vary somewhat regionally, the core objectives and policies are consistent across all major regulatory agencies including the European Medicines Agency (EMA), the Food and Drug Administration (FDA), and the World Health Organization (WHO) (Krishnan et al., 2015).

The EMA was the first agency to develop a regulatory pathway for biosimilars, and guidelines were first introduced in 2005 (Wolff-Holz et al., 2019). These guidelines recommend a comparability exercise comprised of a series of sequential steps that address quality, non-clinical, and clinical information. The FDA published guidelines under the Biologics Price Competition and Innovation Act of 2009 that created an abbreviated approval pathway for biosimilars, requiring the demonstration of a totality of the evidence that includes an evaluation of analytical data, animal studies, and clinical data (U.S. Food & Administration, 2019). The WHO has also posted guidelines on evaluating similar biotherapeutic products, and they serve as a model of a biosimilar development path that several countries have adopted or modified (World Health Organization, 2009).

Regardless of the regulatory agency, elements in the evaluation of biosimilars typically include robust physicochemical characterization, biological activity assessment,

non-clinical studies, clinical studies, and post-marketing surveillance. The stage of physicochemical characterization involves a series of analytical studies designed to establish that the biosimilar's structure and function are similar to that of the reference product. These studies have traditionally included mass spectroscopy, chromatographic techniques, such as size-exclusion chromatography, reverse-phase high-performance liquid chromatography, and hydrophobic interaction chromatography, and a variety of spectroscopic methods, such as assays for protein-carbohydrate composition and proteolysis (Krishnan et al., 2015).

Non-clinical studies may include pharmacodynamic (PD) studies to show that the products have similar biological activity and toxicology studies to evaluate safety profiles. Clinical studies may include both pharmacokinetic (PK) and PD studies to establish bioequivalence, efficacy trials in one or more disease indications to demonstrate safety and efficacy equivalent to the reference product, and immunogenicity trials to compare any potential to elicit immune responses (Wolff-Holz et al., 2019). Regulatory guidance and well-accepted scientific principles may permit extrapolation of data from the clinical study program to certain non-studied indications of the reference product if substantial evidence supports this.

As the biosimilars field develops, the regulatory landscape continues to evolve. Some agencies, such as the FDA, have implemented further criteria for interchangeability in addition to biosimilarity (U.S. Food & Administration, 2019). Ongoing efforts to standardize regulations across regions will enable development programs to be implemented more broadly. Regulatory authorities increasingly appreciate the importance of using advanced analytical technologies, such as mass spectrometry-based multi-attribute methods and transcriptomics, to demonstrate biosimilarity (Wolff-Holz et al., 2019).

The regulatory framework for biosimilars is a compromise between ensuring product quality on the one hand and facilitating market entry on the other (World Health Organization, 2009). As scientific knowledge and analytical prowess improve, it is likely that these frameworks will progressively develop. The aim of future regulatory development would be to accelerate the path from concept through clinical evaluation to product approval, while still ensuring that standards are in place to guarantee the safety and efficacy of products in the target patient population (Krishnan et al., 2015). The ongoing evolution of biosimilar regulation highlights the fluidity of the field, and the

dedication of regulatory agencies to stay up-to-date on scientific knowledge and the evidence base for products that may be developed, to protect patient and public health.

1.5.3 Issues in Development

The development of biosimilars is fraught with a distinct set of challenges that make the development of generics as straightforward as that of small-molecule drugs. Due to the intrinsic complexity of biological molecules and the rigorous regulatory guidelines for demonstrating biosimilarity, these challenges affect nearly every step along the developmental pathway from manufacturing to clinical evaluation (Prasad, 2019).

Another challenge to biosimilar development is the complex post-translational modifications that proteins undergo during their expression. Small molecule drugs are chemically synthesized, ensuring consistent production, but proteins are manufactured by living cells, which makes their post-translational system highly unpredictable. Therefore, proteins produced by living cells are inherently heterogeneous mixtures of related sequences. The profile of a protein's post-translational modifications can be different even within the same experimental system depending on the conditions of cell culture and the conditions of protein purification downstream. Minor changes in culture conditions can lead to significant variation in a proteins' post-translational modifications. This is important because post-translational modifications can significantly influence the biological activity of the protein, such as drug pharmacokinetics, potency and even immunogenicity (Berkowitz et al., 2012).

Another critical issue is the effect of manufacturing processes on product attributes. Manufacturers of biosimilars will need to design their own manufacturing processes for production of their biosimilar, which is almost certainly distinct from the process of the reference product. For example, the choice of expression system may impact the glycosylation of the final product. Sullivan et al. have also described how specifics of purification procedures can alter product impurities and related protein concentrations. Hence, developers need to optimize these processes to create a product that imitates the reference biologic and establish a robust process able of producing a consistent product (Sullivan & Digrazia, 2017).

The need for rigorous and extensive analytical characterization of a biosimilar is unique to biosimilars. Leading analytical methods allow the molecular structure and function of the biosimilar to be critically assessed in numerous ways. However, there are no tools or procedures that can assure complete and exhaustive assessment of a molecule

as complex as a biosimilar. Developers must instead combine complementary techniques so that analytical characterization of the biosimilar is as complete as possible, using all relevant information to evaluate the biosimilar (Berkowitz et al., 2012). Yet even with the best available techniques, there will likely be aspects of the biosimilar molecular structure that cannot be fully elucidated. The batch-to-batch variability of multi-cell line biological materials, and the various vendors of raw materials, can also introduce batch-to-batch variability introducing differences from the reference product. When present, this variability generates random error and demonstrates that all datasets contain some measurement error. Establishing statistical similarity becomes more challenging when the data are inherently noisy due to biological and analytical variability (Chow, 2018).

Moreover, the design and conduct of clinical trials for biosimilars involves unique challenges. Whereas the purpose of trials for novel biologics is to establish the efficacy of the intervention, the objective of biosimilar clinical trials is to demonstrate that differences between the biosimilar and the reference product are not clinically relevant. This feature of biosimilar clinical trial affects decisions regarding patient population, endpoints, and statistical testing strategies. Additionally, the requirement that similarity between the products must be demonstrated using appropriate confidence intervals for clinical equivalence, rather than traditional superiority or noninferiority bounds used for most novel drug approvals, adds complexity to clinical trial design and interpretation (Prasad, 2019).

Immunogenicity assessment is another critical aspect of biosimilar development that presents significant challenges. The likelihood of an immunological response from a patient can be dependent upon small differences in product attributes, manufacturing processes, or packaging and storage conditions. Predicting this immunogenicity based on analytical or preclinical assessments can be hard to do, so the evaluation must also include comprehensive clinical evaluation, along with long-term post-marketing surveillance (Prasad, 2019).

Finally, the regulatory landscape is constantly evolving, and this in itself presents challenges to biosimilar developers. As regulatory agencies update their regulations and guidelines, or issue responses to manufacturers, the strategies employed by developers must be updated to align with these new expectations and requirements. This can be especially difficult in the context of a global development program, where regions may have discrepant expectations, and requirements, as well as differing interpretations about what it means for a molecule to be biosimilar.

In conclusion, although the development of biosimilars offers an important opportunity for increasing patient access to necessary biological therapies, it also poses several difficult scientific, technical, and regulatory issues. Addressing these challenges will require collaboration across many scientific disciplines, including protein chemistry, analytical sciences, manufacturing, clinical development, and regulatory affairs. As the field progresses, new scientific approaches and technologies are beginning to resolve some of these difficulties and will likely make the process of biosimilar development more efficient while maintaining the exacting standards of safety and efficacy demanded for these critical therapies (Berkowitz et al., 2012).

1.5.4 Current Best Practice to Assess Biosimilarity

The establishment of biosimilarity requires an exhaustive battery of analytical, functional, and clinical trials intended to ensure a total body of evidence supporting the similarity between the biosimilar and its reference product. These trials employ a broad range of methods, from physicochemical characterization to in vivo clinical trials, all yielding critical information for the overall biosimilarity confirmation (Pisupati et al., 2017).

Physicochemical characterization lays the foundational tests of biosimilarity. In order to recognize the composition of the biosimilar and the innovator product, high-resolution analytical methods are put to use to define the primary, secondary and tertiary structures. Mass spectrometry (MS) is a crucial tool for these comparisons, providing detailed structural information regarding the primary sequence, post-translational modifications and higher-order structure. Methods such as peptide mapping coupled with liquid chromatography-mass spectrometry (LC-MS) are used, which identify and quantify amino acid modifications and sequence variants. Fourier transform infrared spectroscopy (FTIR), and circular dichroism (CD) spectroscopy determine the composition of secondary structure, whereas higher-order structure can be elucidated via hydrogen-deuterium exchange mass spectrometry (HDX-MS) and nuclear magnetic resonance (NMR) spectroscopy (Fekete et al., 2013).

Chromatographic methods are widely employed for confirming product purity and heterogeneity. Size-exclusion chromatography (SEC) assesses the presence of aggregates and fragments, while ion-exchange chromatography (IEX) and capillary isoelectric focusing (cIEF) provide insights into charge variants. Hydrophobic interaction chromatography (HIC) is often used to separate and quantify different glycoforms. When

used with MS detection, each of these separation methods provide an additional level of characterization (Visser et al., 2013).

One of the most crucial aspects of characterization in this context is glycosylation profiling, because of its profound influence on biological activity and immunogenicity. Several methods to identify and analyze individual glycan species have been described, the most common of which include normal-phase high performance liquid chromatography (HPLC) for profiling and separation of N-linked glycans, and capillary electrophoresis for profiling of glycan populations, both of which may be coupled with MS for exact mass and structural elucidation. Following separation, N- or O-linked glycan pools are comprehensively profiled to reveal the identity and distribution of each glycan in the pool. Individual glycan structures associated with relevant effector functions or safety concerns are specifically quantified and compared between biosimilar and reference product samples (Reusch et al., 2015).

Functional assessment is of equal importance when establishing biosimilarity. Cell-based potency assays are frequently relied upon to compare the biological activity of the biosimilar and originator product. For monoclonal antibodies that bind to cell surface receptors, a standard potency assay is a cell-based, in vitro anti-proliferation assay of the relevant cell target. For example, a biosimilar directed against HER2 could use an assay assessing proliferation of breast cancer cells in culture. This assay provides a direct readout of the ability of the product to bring about the intended biological effect (Jung et al., 2014).

Typically, the binding of the immunogen to the appropriate antigen and FcIIIa receptor are evaluated using surface plasmon resonance or biolayer interferometry. These techniques allow for the assessment of on and off rates and equilibrium dissociation constants, permitting direct comparisons to be made about binding. Where the antibodies are also intended to have effector functions, cell-based assays can be done to assess antibody-dependent cellular cytotoxicity and complement-dependent cytotoxicity, thus evaluating the effector function profiles on a number of batches (Pisupati et al., 2017).

Forced degradation studies are designed to assess the stability profiles of the biosimilar and reference product when subjected to stress conditions. Such studies also provide insights into degradation pathways, and potential differences in susceptibility to various environmental factors (Janzen, 2019).

Though not an assay of biosimilarity per se, the assessment of process and host cell proteins (HCPs) and other process-related impurities is critical to ensure both safety and

consistent product potency and consistency. Enzyme-linked immunosorbent assays (ELISAs) and Liquid chromatography tandem mass spectrometry LC-MS/MS methods are in widespread use for both quantitative and qualitative assessment of HCPs (Auclair & Rathore, 2022).

Recently, advanced analytical approaches have been developed that offer a more thorough evaluation of biosimilarity. Among these are multi-attribute methods (MAM), which, when combined with LC-MS/MS, have become popular for their ability to monitor multiple critical quality attributes. The benefits of these methods are the ability to provide site-specific information on post-translational modifications and product variants simultaneously (Nupur et al., 2022).

RNA sequencing (RNA-seq), while not yet widely adopted in biosimilar assessments, may offer novel insights into biosimilarity at the cellular transcriptomic level. The data would provide a detailed perspective on the effect a biosimilar has on biological systems as compared to the effects from the reference product and have the potential to capture minor distinctions in cellular pathways that neither analytical nor functional tests might detect (Ozsolak & Milos, 2011).

Lastly, and not systematically analyzed as a test, careful design and conduct of appropriate clinical trials are also necessary to establish biosimilarity. These trials typically consist of pharmacokinetic (PK) and pharmacodynamic (PD) evaluations to show bioequivalence, as well as comparative effectiveness and safety trials in a sensitive patient population (European Medicines Agency, 2019).

In summary, the demonstration of biosimilarity requires a comprehensive approach that integrates state-of-the-art analytical methodologies with functional assays and clinical studies. The goal of this totality of the evidence strategy is to establish a substantial body of evidence supporting the similarity of the biosimilar and the reference product, such that any remaining uncertainties are reduced to a level consistent with safe and effective use of these essential drugs.

1.5.4.1 Transcriptomics approaches for assessing functional biosimilarity in cellular responses

Aside from structural comparison, demonstrating that the biosimilar compound induces the same cellular and molecular response is crucial in confirming biosimilarity. Transcriptomics analyses provide detailed insights into cellular processes at every stage of the experimental workflow.

The use of microarrays and RNA-sequencing in comparative biological studies has been reviewed previously (Kolitz et al., 2014). It includes issues such as the need of establishing control groups, data normalization, and testing power as well as the need for proper data analysis. The review concluded that transcriptomic profiling has the capability to identify subtle shifts in gene expression in cell-based assays. The use of transcriptomics has been extended to the analysis of several biosimilars and their originator products. Generic equivalent of Copaxone was tested against the brand Copaxone, using circulating immune cells taken from multiple sclerosis patients and analyzed with microarrays for transcriptomics studies (Towfic et al., 2014). They found that the biosimilar had a higher expression of genes related to inflammation and cell adhesion, which may suggest a later onset of side effects. The researchers characterized the drug-specific immunomodulatory signatures of anti-TNF- alpha biologics, etanercept and adalimumab, by means of performing microarray studies on gene expression profile from peripheral blood cells of treated patients (Meugnier et al., 2011). The above were the first studies using either RNA- seq or microarray to analyze complex drugs, although they did not include monoclonal antibodies.

More recently, transcriptome analyses have been applied to biosimilarity studies for several specific purposes. One such application involves exploring mAb expression in Chinese Hamster Ovary cell lines using RNA-seq (Schulze et al., 2022). Anti-TNF biologic therapy was compared to evaluate treatment effectiveness and the normalization of expression in disease states. The FDA's emphasis is on the application of transcriptomics in the identification of pharmacodynamic biomarkers that can show drug activity, be used for dosing, and detect resistance, that represent a more ancillary usage of transcriptomics techniques to prove biosimilarity.

The use of microarray profiling was also recently applied by Avila-Portillo et al. (2020) in order to compare the impact of Granulocyte Colony-Stimulating Factor (G-CSF) and its biosimilar on target cells (Avila-Portillo et al., 2020). This approach elicited a nearly identical, though not entirely identical, cellular response. The signature response found that genes that are differentially expressed are linked to the PI3K-Akt, NF-kB, TNF signaling, and anti-apoptotic signaling pathways, suggesting a phenotypic response in common. This research applied transcriptomic methodologies to demonstrate an equivalent cellular response *in vitro* to establish global biosimilarity prior to clinical trials involving recombinant human proteins, though engineered proteins such as therapeutic monoclonal antibodies. Furthermore, the integration of proteomics analyses

with transcriptomics data in the study of mAb biosimilarity remains unexplored (Haider & Pal, 2013; Nie et al., 2007).

1.5.5 Problems and Opportunities for Biosimilars

Accelerating the field of biosimilars can be facilitated by addressing key scientific, regulatory, and market-related challenges and opportunities. The ability to anticipate potential problems and to leverage the available resources to robustly address these challenges will shape the success or failure of biosimilars and their impact on healthcare going forward (Kos et al., 2018).

One of the key barriers to biosimilar development is the inherent complexity of biological molecules. Unlike small molecule generics, biosimilars cannot be identical to its originator product because the complexity of protein structure and post-translational modifications. This complexity requires extensive analytical characterization and functional testing, which can be time-consuming and expensive. In addition, the sensitivity of biological products to manufacturing conditions in production processes, can lead to differences in the final product. Finally, the lack of access to proprietary information regarding a reference product's manufacturing process requires that many details of production, like the reference product itself, must be reverse engineered (Berkowitz et al., 2012).

The present analytical strategy is thorough, but incomplete. The majority of assays used for biosimilarity testing—targeted approaches that assay a prior defined list of critical quality attributes— provide quantitative information that is indispensable for assessing biosimilarity between products. However, focusing on just a limited range of critical quality attributes may not uncover all possible sources of dissimilarity between the biosimilar and reference product. Most notably, there are no examples of untargeted approaches, such as RNA sequencing or shotgun proteomics, in use for routine biosimilarity assessments today. These omics technologies have the power to uncover subtle differences in eukaryotic cell responses or targeted protein expression profiles that would be invisible using conventional approaches.

Regulatory challenges are also a known limitation in biosimilar development. The regulatory environment has become significantly more robust and interpretable over time, but there is still some variability region-to-region. This can create challenges for global development plans and has the potential to produce unnecessary iterative testing requirements. In addition, the notion of interchangeability, which is important in the U.S.

environment, presents yet added layers of complexity around the development and approval of biosimilars (Kang et al., 2021).

But there are also unique barriers for biosimilars to being adopted in the market. Unlike generic small molecule drugs, which can often be automatically substituted, biosimilars typically require healthcare providers to actively prescribe, rather than passively allow switching. This necessitates a large number of educational efforts to build confidence among healthcare professionals and patients. Additionally, complex pricing and reimbursement structures, and strategies such as patent litigation by originator companies, can impact biosimilar market access and uptake (Edgar et al., 2021).

However, in the face of these hurdles, biosimilars also present considerable opportunities to increase the reach and reduce the cost of healthcare. The advent of biosimilars may lower spending on healthcare considerably, especially for biological therapies that are often high-priced medical products. These less expensive biosimilars may improve patient access to necessary treatments and enable earlier interventions for disease.

Additionally, the pursuit of biosimilar development has spurred innovations in analytical technologies and biopharmaceutical manufacturing, which require highly sensitive and robust analytical techniques including mass spectrometry and chromatography so that all aspects of these drugs are identified and examined. These advancements have broad implications for biopharmaceuticals (Berkowitz et al., 2012).

One potential new direction is the use of advanced omics technologies for biosimilar development and characterization. Although not specifically outlined in current guidelines, the use of RNA-seq or proteomics could be useful for understanding biosimilarity at the level of cellular response. These mechanisms are likely to be most beneficial for biosimilars that have complex or unknown mechanisms of action (Berkowitz et al., 2012).

A third opportunity is the potential of biosimilars to enhance personalized medicine. The greater availability of biological therapies that results from the increasing number of biosimilars on the market can improve and extend research on treatment responses in larger and more diverse patient populations. This research can in turn help identify the factors that predict response to treatment and toxicity, which can then be used to develop processes for more personalized treatment.

In sum, although biosimilars must overcome significant hurdles in development, regulation, and market penetration, they also offer significant opportunities to improve healthcare and drive scientific progress. Research in the field of biosimilars is ongoing, and technologies and analytic tools are emerging that may offer new solutions for establishing similarity and understanding biological effects. As the science and regulation of biosimilars continues to advance, they are likely to make greater and potentially transformative impact on the biotherapeutic landscape and improve patients' access to critical biological therapies (Kos et al., 2018).

1.6 Thesis Relevant Information

1.6.1 Herceptin biosimilar characteristics (Apotex-trastuzumab)

In this study, two therapeutic monoclonal antibodies were utilized with Apotex-trastuzumab (a biosimilar of trastuzumab (ApoTras) provided by Apobiologix (Canada)) and the originator product Herceptin (Genetech, US). Prior to performing in-depth functional analysis of transcriptomics and proteomics data, a typical structural and functional comparison was performed to ensure comparability. A summary of critical quality attributes (CQAs), including potency, identity, and purity, was provided by Apobiologix (Canada). The comparison of the two products revealed both similarities and minor differences (**Table 1-1**). For instance, relative potency assays demonstrated comparable anti-proliferative effects and HER2 binding affinity. The measured values were within the acceptable quality ranges, defined by the comparison of three different Herceptin batches. However, differences were observed in antibody-dependent cellular cytotoxicity (ADCC), with ApoTras exhibiting a higher relative potency (1.38) compared to Herceptin (0.62).

Analysis of deamidation, oxidation and lysine glycation for the light and heavy antibody chains revealed distinct variations, with most being slightly outside of the quality range. In general, deamidation levels were lower in the biosimilar ApoTras, while oxidation and lysine glycation were reduced for Herceptin. Glycan composition and site-specific glycosylation also showed differences between both antibodies, such as ApoTras having lower levels of G0F and higher levels of G1F glycans compared to Herceptin. Despite these differences, both products were generally within or close to the quality range. However, it remains unclear how the measured differences might impact the efficacy, limiting the usability for assessing biosimilarity.

Table 1-1 Comparative Analysis of Potency, Identity, and Purity Attributes of ApoTras Biosimilar (#FAZ1502) and Herceptin (US; #3014296). Results include potency, identity, and purity. Units are provided for each parameter, and the 'Quality Range' represents the acceptable variability based on three independent measurements of three different lots (N = 3). Glycosylation types: G0: Core glycan (mannose + GlcNAc), no fucose or galactose; G0F: Core glycan with fucose, no galactose; G1F: Core glycan with fucose and one galactose; G2F: Core glycan with fucose and two galactose residues. LC: Light chain; HC: Heavy chain.

Parameter	Method	Measurement	Units	ApoTras (#FAZ1502)	Herceptin (US; #3014296)	Quality Range (N = 3)
Potency						
Potency	Anti-Proliferation Assay	Relative potency	na	1.01	1.05	0.88-1.12
Potency	ADCC	Relative potency	na	1.38	0.62	0.34 – 1.18
Potency	HER2 binding	Relative potency	na	0.92	0.99	0.80 – 1.28
Identity						
Identity/Glycan Composition	Glycoprofiling by Rapifluor	% Afucosylation	%	10.14	7.38	1.87 – 17.17
Identity	Asn/Gln Deamidation analysis	Report Results (HC)	% Deamidation	14	19	16.9 – 19.3
Identity	Asn/Gln Deamidation analysis	Report Results (LC)	% Deamidation	16	28	26.1 – 29.1
Identity	Met/Trp Oxidation Analysis	Report Results (HC)	% Oxidation	33	12	10.7 – 15.5
Identity	Met/Trp Oxidation Analysis	Report Results (LC)	% Oxidation	2	1	0.6 – 1.2
Identity	Lysine Glycation	Report Results (HC)	% Lysine Glycation	7	4	2.7 – 4.5
Identity	Lysine Glycation	Report Results (LC)	% Glycation	5	3	2.8 – 2.8
Purity						
Glycan Composition	Sialic Acid Analysis	Report Results	pmol Neu5Ac/pmol protein	0.04	0.02	0 – 0.09
Glycan Composition	Sialic Acid Analysis	Report Results	pmol Neu5Gc/pmol protein	n.d.	n.d.	
Glycopeptide analysis	N-Glycosylation site determination	Report Results	Glycosylation level N300	100	100	
Glycopeptide analysis	Site Specific N-Glycosylation composition	Report Results (HC)	% G0F – HexNAc	2	5	5
Glycopeptide analysis	Site Specific N-Glycosylation composition	Report Results (HC)	% G0	4	4	2 -8
Glycopeptide analysis	Site Specific N-Glycosylation composition	Report Results (HC)	% G0F	37	63	51 – 69
Glycopeptide analysis	Site Specific N-Glycosylation composition	Report Results (HC)	% G1F	37	21	8 -32
Glycopeptide analysis	Site Specific N-Glycosylation composition	Report Results (HC)	% G2F	7	2	2
Glycopeptide analysis	Site Specific N-Glycosylation composition	Report Results (HC)	% Others (< 5%)	13	6	0 - 18

1.6.2 BT474, SKBR3 and MCF7 sensitivity to Herceptin and Apotex-trastuzumab

Three breast cell lines were selected for this thesis including BT474, SKBR3 and MCF7. Based on prior characterization, BT474 was expected to respond strongly to mAb treatment, SKBR3 moderately, and MCF7 not at all. The concentration of 4 $\mu\text{g/ml}$ was chosen based on a concentration gradient and in accordance with T. Ichikawa's study (Ichikawa et al., 2012). To confirm these expectations, a WST-1 cell viability test was performed to assess cell responses (**Figure 1-1**). The WST-1 colorimetric assay demonstrated that the order of responsiveness among the cell lines was BT474 > SKBR3 > MCF7. All treatments in this thesis were performed with a concentration of 4 $\mu\text{g/ml}$ for 6 days.

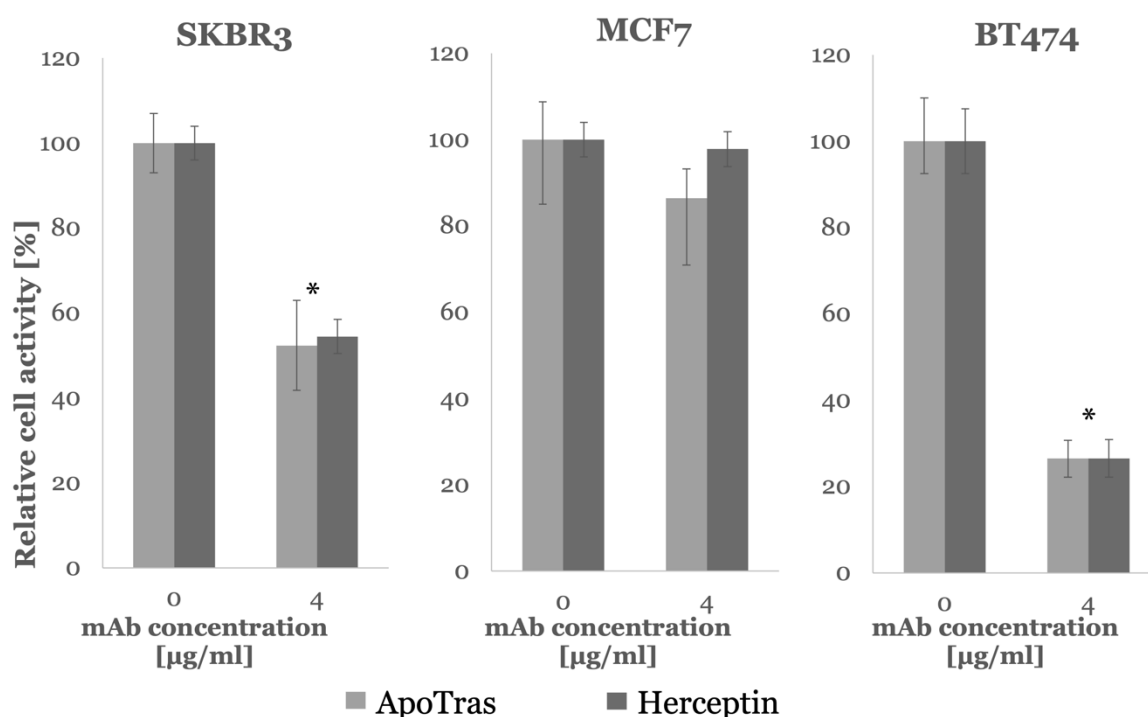


Figure 1-1 Sensitivity test of SKBR3, MCF7, and BT474 cell lines for Herceptin (Her) and its biosimilar, ApoTras. Each concentration was tested with six replicates ($N = 6$; mean $\pm$ SD). The cells were exposed to 4 $\mu\text{g/ml}$ of mAbs for 6 days. The media was replaced every 72 hours. The cell activity was measured with WST-1 colorimetric assay. The ratio of the absorbance to that of the non-treated cells represents the sensitivity of cells to Her or ApoTras. No significant differences were observed between the treatment with ApoTras and Her. * Indicates p-value < 0.05 for comparison of treated samples versus control samples using Student t-test.

1.7 Thesis Outline

Therapeutic monoclonal antibodies (mAbs) for targeted therapies have revolutionized cancer treatment in the last 30 years. One major drawback of trastuzumab treatment is the high costs, limiting the access for patients. However, the patents for therapeutic mAbs started to expire allowing the creation of more affordable biosimilars. Proving biosimilarity to needs compliance with extensive regulatory frameworks. They currently emphasize targeted wet-lab studies by focusing on comparing potency, purity, and efficacy. However, high-throughput approaches to assess molecular and phenotypic effects have been neglected so far. This is concerning because therapeutic mAbs like trastuzumab elicit strong cellular responses in target cells and the mode of action (MoA) and development of resistance are not fully understood. This thesis attempts to fill this gap and explore the application of proteomics and transcriptomics data for assessing the biosimilarity of Herceptin (Her) and a biosimilar candidate Apotex provided by Apobiologix (Toronto, Canada) (ApoTras).

Chapter 2 establishes proof-of-concept for using transcriptomics data from a mAb-treated BT474 cell line for confirming biosimilarity. The focus is on the bioinformatic analysis for providing simple comparisons, identifying differentially expressed genes (DEGs) and implementing a similarity index to confirm overall transcriptional similarity between the Her and ApoTras. The chapter also introduces the need for investigating the type two error rates for different effect sizes to evaluate our confidence in the identification of DEGs.

Chapter 3 introduces a novel bootstrap-based methodology to estimate Type II errors for different comparisons to detect DEGs. This chapter investigates the reliability of RNA-seq data analysis by evaluating the per-gene power and reproducibility across TCGA breast cancer dataset under varying experimental conditions. The algorithm was employed to identify Type II errors for data sets with different in-group and between-group variance, effect sizes and replicate numbers. This allowed for a robust exploration of statistical power in RNA-seq experiments.

Chapter 4 builds upon these analyses by delving into the molecular characterization of treatment responses including additional data from treated SKBR3 cell samples. The transcriptomics differences and similarities between cells treated with Her and ApoTras are deeply explored using the transcriptomic data. The PROGENy pathway activities were investigated to identify differences in cancer hallmark processes. Additionally,

changes in regulatory network activity are investigated to understand the driver transcription factors causing the observed changes in expression. These analyses were complemented by an untargeted gene set enrichment analysis providing the most complete view of cellular pathways to highly sensitive approach for identifying all functional changes. In summary, the chapter provides a comprehensive analysis of the MoA of Herceptin and potential functional differences between the originator product Her and its biosimilar ApoTras.

In Chapter 5, the transcriptomic data was extended with proteomic data from BT474, SKBR3 and MCF7 cell lines to confirm transcriptional changes as phenotypic changes. The MCF7 cell line served as a negative control because the cell line lacks expression of HER2 which is the target of trastuzumab. The datasets were integrated on pathway enrichment level to reduce the impact on study biases and arbitrarily chosen threshold for statistical or biological significance. The correlations between RNA-seq and proteomics pathway signatures highlight key biomarkers and functional pathways. This integrated analysis provides a comprehensive view of phenotypic changes distinct and shared therapeutic impacts of the two therapeutic mAb Her and ApoTras.

Together, this thesis provides a reproducible framework to perform an untargeted, high-throughput biosimilarity assessment. Additionally, valuable insights into the transcription dynamics of trastuzumab treatments extend our understanding of the MoA of trastuzumab.

Chapter 2: Assessing biosimilarity of trastuzumab monoclonal antibody therapeutics using RNA sequencing

2.1 Introduction

Therapeutic monoclonal antibodies (mAbs) represent a key class of biopharmaceuticals, with cancer and chronic inflammatory diseases being the most common targets. As of 2019, over 90 mAb based products were approved by the U.S. Food and Drug administration (FDA) for clinical practice (U.S. Food and Drug Administration, n.d.). However, the cost of immunotherapy can be prohibitive, limiting patients' access. The patents for eight mAbs expired in 2020 in the USA, allowing the production of more affordable biosimilar versions of mAbs (Generics and Biosimilars Initiative (GaBI), 2011).

Therapeutic mAbs are composed of a heterogeneous mixture of post-translational modifications (PTMs) around a common core protein structure. A typical structure for a mAb is a combination of four peptide chains (two heavy chains and two light chains) plus PTMs, resulting in a total molecular weight of approximately 150 kDa for IgG-based mAbs. Types of PTMs present include N-glycosylation (Hodoniczky et al., 2005; Jefferis, 2012; Mimura et al., 2001; Raju, 2010), disulfide bonds (Dillon et al., 2008; Wypych et al., 2008), deamidation (Vlasak et al., 2009), and methionine or tryptophan oxidation (Chumsae et al., 2007; Gaza-Bulsecu et al., 2008). The pattern of PTMs depends on the protein expression system (Frenzel et al., 2013; Wells & Robinson, 2017), cell culture conditions (Butler, 2005; F. Li et al., 2010), and downstream purification processes (Roque et al., 2004). PTMs can influence the biological properties of mAbs including pharmacokinetics, potency (Hmiel et al., 2015), and immunogenicity (Kuriakose et al., 2016). The development of a biosimilar with equivalent biological activity and very similar combination of PTMs as an approved mAb is challenging, but the regulatory approval process for a biosimilar still typically requires the demonstration of a highly similar product. The physicochemical profile is especially important, as the *in vitro* and *in vivo* biological activity is measured via binding affinity and potency. These tests help confirm that the biosimilar has no clinically meaningful differences or safety concerns in comparison to an approved reference product.

An established humanized mAb therapeutic is trastuzumab (Herceptin®), used in the treatment of HER2 positive breast cancers. Trastuzumab is listed on the World Health

Organization Model list of essential medicines, those considered to be the safest, most effective, and most essential to meet the needs of a modern Health Care system (World Health Organization, 2017). Trastuzumab targets the surface receptor tyrosine kinase erbB-2 (HER2), which is a member of the epidermal growth factor receptor family (EGFR). HER2 is overexpressed in approximately 20-30% of all breast cancers (King et al., 1985). The receptor is activated by homo- or hetero-dimerization leading to an activation of the intracellular tyrosine kinase domain. The autophosphorylated tyrosines in turn interact with various other proteins which control important cellular pathways. These include the PLC-gamma (Peles et al., 1991), Ras/Raf/MEK/ERK (Duru et al., 2012; J. Liu & Kern, 2002) and PI3K/Akt (Delord et al., 2005) pathways, plus additional cross-talk with other pathways (Y. Lu et al., 2001, 2004). These signalling pathways influence a variety of cellular processes, including migration, angiogenesis, apoptosis, and proliferation, by controlling numerous cellular targets or altering gene expression. The overexpression of HER2 results in the increased activation of the above signalling cascades and can promote cancer progression (Eccles, 2001; Faltus et al., 2004; Hudziak et al., 1987; Yakes et al., 2002; Zhou et al., 2004).

The binding of trastuzumab to HER2 has a complex effect on the target cells. Most notably, binding of trastuzumab to the HER2 receptor can markedly reduce cell growth and proliferation, slowing the progression of tumours (Valabrega et al., 2007). Additionally, the constant region (Fc) of trastuzumab can recruit immune effector cells such as natural killer cells resulting in antibody-dependent cell-mediated cytotoxicity (ADCC), (Clynes et al., 2000; Collins et al., 2012) or macrophages enabling antibody-dependent cellular phagocytosis (ADCP) (Petricevic et al., 2013). When used in combination with other therapeutics, it may additionally help induce complement-dependent cytotoxicity (CDC) (Mamidi et al., 2013). While HER2 protein signalling has been well studied the downstream transcriptional events after the binding of trastuzumab takes place remain less clear.

A common method for assessing analytical similarity of biosimilar mAbs is cell-based potency assays. An *in vitro* anti-proliferation assay with breast cancer cell line is often the primary assay to compare the bioactivity of a biosimilar product to Herceptin (J. Lee et al., 2018; Miranda-Hernández et al., 2015). This assay measures the change in cell numbers over time but does not provide a detailed characterization of pathway activities or regulation of pivotal genes which may be highly specific to the given mAb treatment. The use of transcriptomics as a measure of biosimilarity provides an

additional level of analytical detail about mAb function and can provide important insights into mechanisms of action and differences in response at the cellular level. Transcriptomics techniques can comprehensively assess expressed gene transcripts (mRNAs) within cell-based models and more specifically RNA-seq experiments assessing transcriptional responses to mAbs can provide important insights into the effects of mAb treatments (Merry et al., 2016; Von Heyde et al., 2015). To date, transcriptomics has not yet been applied as a tool to directly examine biosimilarity.

The current work presents results of gene expression analysis using RNA-seq to quantify the similarity of biological responses to marketed originator product Herceptin (Her) and a recently developed biosimilar version (ApoTras). We present a comparative analysis of the response of HER2 positive BT474 breast cancer cells to treatment with the mAb Her and the biosimilar ApoTras. The resulting transcriptomic changes, relative to a common untreated control group, were examined. The presented methodology directly compares the two treatments to identify potential differences in effect and secondly identifies changes common to both treatments relative to controls. This is formulated as two orthogonal contrasts, providing independent tests of differences between the treatments and of commonalities relative to the controls. We further introduce a measure to quantify similarity of treatments based on ratios of the number of statistically significant changes in the two comparisons. In addition, a type II error analysis for adjusted effect size was implemented to confirm that the statistical power was reasonable. In the present study, both originator and biosimilar compounds are expected to have highly similar effects on target cells.

2.2 Materials and Methods

2.2.1 Cell Culture

The HER2-positive human breast cancer cell line, BT474 (HTB-20; ATCC, American Type Culture Collection, Manassas, USA), was cultured in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific, San Jose, USA) containing 10% (v/v) fetal bovine serum (FBS; Gibco, Mississauga, Canada) and 1% (v/v) penicillin-streptomycin [10,000 U/ml penicillin, 10,000 µg/ml streptomycin] (Gibco, Mississauga, Canada) at 37°C and 5% CO₂. The cell line was confirmed to be free of mycoplasma contamination with the PCR Mycoplasma Test Kit (PromoKine, Heidelberg, Germany) according to the manufacturer's protocol.

2.2.2 Antibody Treatment

The experimental design included ten biological replicates per condition. On Day 0, approximately 1.2×10^6 cells were plated into 10 cm Petri dishes for the control and treatment groups. From Days 1 - 6 cells were incubated with either Herceptin (Genentech, San Francisco, USA) or the trastuzumab biosimilar, ApoTras (Apobiologix, Toronto, Canada) at a concentration of 4 µg/ml (Ichikawa et al., 2012) (Appendix Figure A 3), or with no treatment (control). The culture media was replaced on Day 1 and Day 4. Total RNA extraction took place on Day 7.

2.2.3 RNA Extraction and Sequencing

Total RNA was isolated from BT474 cells with Quick-RNA™ MiniPrep kit (Zymo Research, Irvine, USA) following the manufacturer's instructions. The cDNA library was constructed with *TruSeq Stranded mRNA* Library Prep Kit (Illumina, San Diego, USA) and sequenced by the Donnelly Centre for Cellular and Biomolecular Research (Toronto, Canada). 2 x 100 bp paired-end reads were produced with the HiSeq 2000 (Illumina) platform. The raw reads were received in FASTA file format.

2.2.4 Data Analysis

2.2.4.1 Sequence quality control and adaptor trimming

The quality of the sequence data was assessed with FastQC v3 (Andrews, 2010) before and after adapter trimming with the BBDuk method from the BBTools suite

(Bushnell, 2016) and summarized using MultiQC v1.25.2 (Ewels et al., 2016). A k-mer size of 23 was used for trimming the 3' ends of the reads.

2.2.4.2 Alignment and Quantification

The fast, short-read alignment tool Bowtie v1.2.1.1 (Langmead et al., 2009) was used to align the reads allowing a maximum of two mismatches in the first 25 bases of each read. Reads were aligned to the human transcriptome (version: GRCh38.p7) of the consensus coding sequence (CCDS) project (Farrell et al., 2014). The release contains 32,524 CCDS entries and 18,892 Gene IDs. Afterwards, RSEM v1.3.0 (B. Li & Dewey, 2011) was used to estimate the accurate abundance of genes with default parameters.

2.2.4.3 Pearson correlation test of housekeeping genes

A previously established list of 'housekeeping' genes was used for quality control (Eisenberg & Levanon, 2013). The authors analyzed expression data from the Human BodyMap (HBM) 2.0 Project. Housekeeping genes are expected to maintain consistent expression levels all within and across tissue groups and have low experimental variance relative to other genes. Samples with housekeeping genes deviating from group or overall means are an indicator of poor data quality. The Pearson correlation test was used to assess housekeeping genes and was done with the package `corrplot` in RStudio (RStudio Team, 2016). For the BT474 cell line, a mean correlation value of $R < 0.90$ between any sample and the rest of the data set was used as a threshold to identify sample outliers. Any identified outliers were removed from further analysis.

2.2.4.4 Testing for differentially expressed genes (DEG)

The expected count tables were imported into RStudio v1.1.442 (RStudio Team, 2016) and analyzed with the `edgeR` v3.18.1 package (Robinson et al., 2009). The `edgeR` package uses a negative binomial distribution to model the dispersion of read counts. Prior to differential expression analysis, excluding genes with low counts improved the detection power of differentially expressed and biologically relevant genes. The filter threshold was set to 1 count per million (CPM) per sample, present in at least 9 biological replicates. The library sizes were normalized with the weighted trimmed mean of M values (TMM) method. Dispersion was calculated with the Cox-Reid profile-adjusted likelihood method to fit a generalized linear model, which allows to test the expression levels for significant differences. A generalized linear model likelihood ratio

test is carried out with the `glmQLFit` function. For the three experimental conditions (Her, ApoTras, and controls) two comparisons were made. One comparison is between ApoTras and Her, identifying differences between the two treatments; the second is a comparison of both treatments with controls to identify common changes in gene expression. These two comparisons, termed ApoTras vs. Her and Her&ApoTras vs. Con form an orthogonal contrast, partitioning the data into two independent tests.

Identification of DEG was corrected for multiple hypothesis testing using false discovery rates within the `glmTreat()` function within `edgeR`. The default false discovery rate parameter was set to $q = 0.05$, representing a 5% false discovery rate (FDR) within each comparison. FDR were set independently for each comparison. Fold change thresholds (`minFC`) were used to set the minimum difference in gene expression to be considered of potential biological relevance, where the FC is calculated as the ratio between two comparison means with the larger value in the numerator. Thresholds for `minFC` of 1.25, 1.5 or 2.0 were used, with `minFC` = 1.25 representing a minimum increase of 25% in gene expression level for one of the comparison groups.

2.2.4.5 Type II error analysis

For the type II error analysis, a bootstrapping method was implemented for random sampling with replacement (1000 replicates). This step was integrated within the `edgeR` analysis pipeline after library size and TMM normalizations. Prior to resampling the $\log_2(\text{CPM})$ values were adjusted so that the mean effect size for each gene was set to a specific value (e.g. $\text{FC}_p = 1.5$ or 2.0). The adjusted $\log_2(\text{CPM})$ values were then transformed to read counts while keeping the library size of the samples fixed, and values were randomly resampled with replacement. The dispersion parameters of the genes were kept consistent with the original dataset so that the variance estimates are not impacted by the resampling. This workflow was implemented with an in-house R script and used to estimate the type II error of all genes and the overall power for the comparisons.

2.2.4.6 Similarity index

We implement an index based on direct comparison of statistical differences between A vs. B and changes common to both treatments, i.e. (A and B) vs. C. Equivalently, the similarity measure is calculated based on two independent contrasts as we have used in

this study. For treatment groups A and B with control group C, the SI is defined as follows:

$$SI_{minFC} = \frac{(\text{common changes} - \text{differences})}{\text{all observed changes}}$$

$$SI_{minFC} = \frac{DEG(A\&B \text{ vs. } C) - DEG(A \text{ vs. } B)}{DEG(A\&B \text{ vs. } C) \cup DEG(A \text{ vs. } B)}$$

The subscript to SI indicates a minFC, DEG (A&B vs. C) represents the number of DEG in the contrast of combined treatments (A and B) vs. C, and DEG (A vs. B) represents the number of DEG in the contrast of treatments A vs. B. The index will tend towards 1 with increasing similarity between treatments and should equal 1.0 if A and B are replicates of the same treatment. If A and B have significant differences between treatments but no common differentially expressed genes when compared to controls, the index will equal -1.0.

2.2.4.7 Pathway analysis of DEG

Identification of overrepresented KEGG pathways excluding disease related gene sets (261 pathways) was performed with the command `kegga` from the `r-studio` package `limma` 3.33.14 (Ritchie et al., 2015). This command implements a Wallenius non-central hypergeometric distribution to calculate *p*-values. Default parameters were used for the analysis.

2.2.4.8 Data visualization

Figures 2-7 were generated with the package `ggplot2` (Wickham, 2016) in RStudio.

2.2.5 Additional Files and Code Availability

Additional files and scripts for preprocessing, analysis, and figure creation are available in the GitHub repository at https://github.com/karstenrinass/Thesis/Chapter_2.

2.3 Results

RNA-seq was used to assess gene expression in BT474 cell lines exposed to Herceptin (Her) or Apotex-Trastuzumab (ApoTras), as well as untreated controls (Con), with the larger goal of assessing similarity between ApoTras and Her treatments (**Figure 2-1**). Two statistical contrasts of gene expression were performed. The first identified common changes due to both trastuzumab treatments relative to controls, Her&ApoTras vs. Con, and the second identified differences between the two trastuzumab treatments, ApoTras vs. Her.

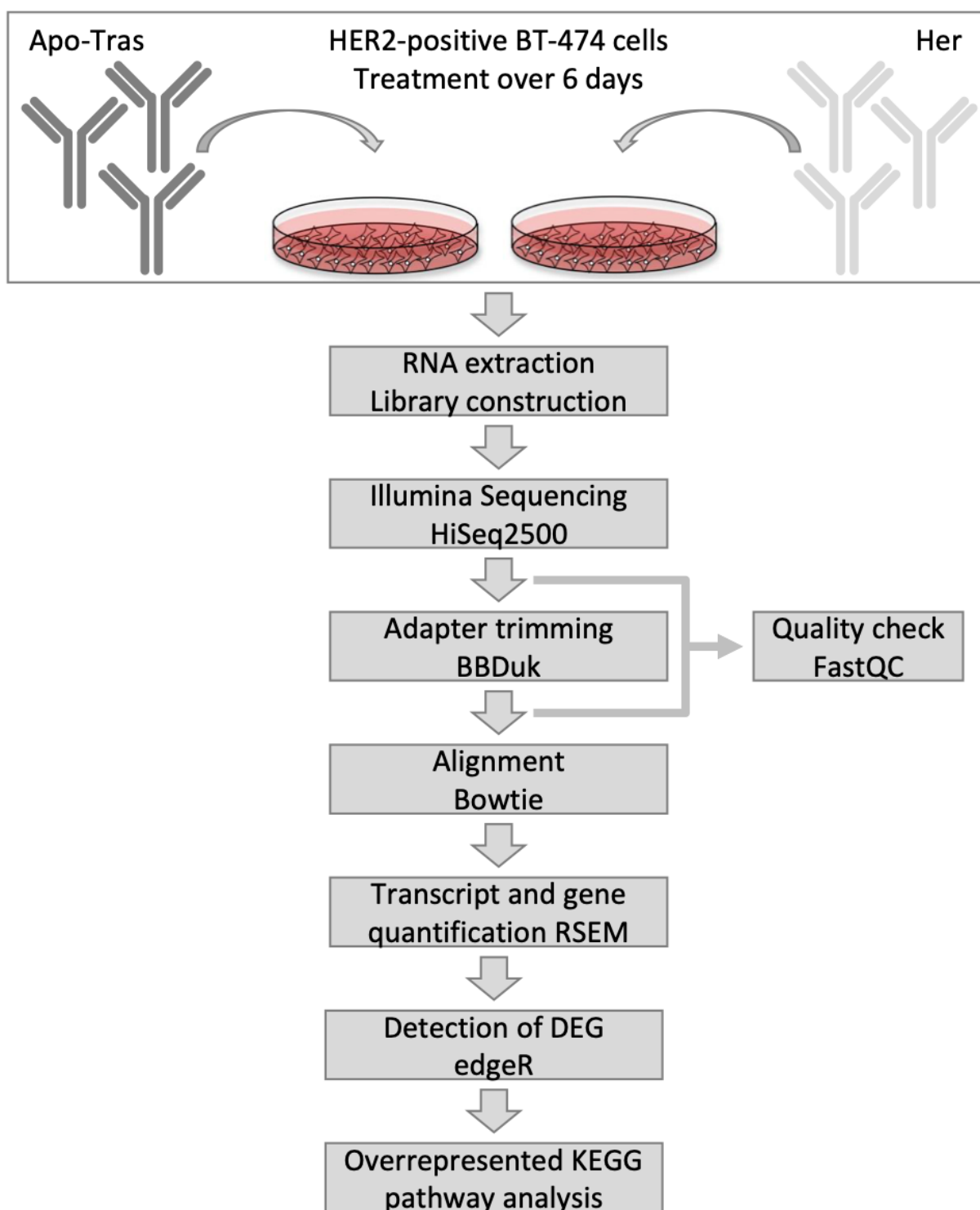


Figure 2-1 Experimental workflow summary for the comparison of monoclonal antibodies. BT474 cells were exposed to biosimilar (ApoTras) or originator (Her) compounds, with no treatment (control) as a reference. Total RNA was extracted after six days, and a cDNA library was prepared from the mRNA content. Samples were sequenced using the Illumina HiSeq 2500 platform. Resulting reads were adapter trimmed, and the quality was checked with FastQC. Processed reads were aligned to the human transcriptome (Version: GRCh38.p7) of the consensus coding sequence (CCDS) project. Subsequently, the program RSEM was used to quantify transcript and gene expression, and the R package edgeR v3.18.1 used for statistical analysis of DEGs. The resulting gene lists was used for functional profiling of DEG by identifying overrepresented KEGG pathways.

2.3.1 Quality control and data preprocessing

The MultiQC report summarizing the FastQC results is available in [Additional file 1](#). After alignment, filtering and normalization, 11,910 genes were identified and used for differential expression analysis ([Additional file 2 Table S2](#)). Sample data quality was further assessed using housekeeping genes, which are expected to maintain relatively consistent expression levels between all samples. Expression values for housekeeping genes were assessed using Pearson correlation analysis, including 3649 genes. Out of 30 samples, 29 exhibited highly similar housekeeping gene expression (mean correlation scores > 0.95) except for one ApoTras sample, which had a mean correlation coefficient of less than 0.75 when compared to all other samples (Appendix Figure A 1). This sample was identified as a low data quality sample and was excluded from further analyses.

Table 2-1 Numbers of significant DEG for different fold change thresholds (FCT) and an FDR of 0.05. See [Additional file 3 Tables S3-S8](#) for identified genes.

Contrast	Gene expression	FCT 1.25	FCT 1.5	FCT 2
ApoTras&Her vs. Con	Up	1311	484	165
	Down	1312	477	70
ApoTras vs. Her	Up	24	0	0
	Down	0	0	0

Prior to detailed analysis, the effect of the fold change threshold used within edgeR (Robinson et al., 2009) (termed minFC) on the number of DEG was assessed for sensitivity. The false discovery rate was set to 5%, and fold change thresholds of minFC = 1.25, 1.5 or 2 were compared (**Table 2-1**; See [Additional file 3 Tables S3-S8](#)), with minFC = 1.25 representing a minimum threshold of 25% difference in gene expression level between comparison groups. No significant changes were observed for the contrast ApoTras vs. Her for parameter values minFC = 1.5 or minFC = 2.0. Changes were only observed for minFC = 1.25, for which 24 genes were identified with a minimum change in gene expression of 25% for upregulation and ~20% for downregulation.

2.3.2 Determine the information level of the comparisons for different minFC

2.3.2.1 Overall sample similarity

Principal components analysis was used to visualize overall differences between samples (**Figure 2-2A**). The samples separated into two groups along principal

component 1 (PC1), with the control samples clustered closely together. In comparison, the treatment samples were clearly separated from the control samples along the PC1 axis. Compared to control samples, treated samples had greater spread along the principal component 2 (PC2) axis, exhibited greater within-group variability. The variance within the two treatment groups was primarily present along PC2 axis, indicating that a different combination of transcript expression changes is responsible for the variability within the treated samples as compared to the control samples. However, the two distributions were not clearly partitioned. Accordingly, the variation between the two treatments was highly comparable to the within-group variation in terms of the differential expression of transcripts.

The magnitude of effect on gene expression for each of the two treatments was compared by plotting the ratio of expression levels for ApoTras/Control and Her/Control (**Figure 2-2B**; [Additional file 3 Tables S3, S4](#)). At a minFC of 1.25 and FDR of 5%, 2623 genes were identified as differentially expressed when treated with trastuzumab (Her&ApoTras) vs. controls, and 24 genes are differentially expressed for ApoTras vs. Her ([Additional File 3 Tables S3, S4](#)). Here, $\log_2(\text{FC})$ values vary from approximately -3 (8-fold reduction in expression) to +5 (32-fold increase in expression). The data clustered around a 1:1 ratio line from lower left to upper right, indicating that common DEG had a similar magnitude and direction of response for both treatments. Genes with significant differences between trastuzumab treatments appeared in a band near the center of the figure, consistent with relatively small changes in expression for these genes. For the 24 genes showing differences between the two treatments, the fold change was low and overlaps with the background (non-differentially regulated) genes in terms of difference in expression.

At a minFC of 1.25 and FDR of 5%, 2623 genes were identified as differentially expressed when treated with trastuzumab (Her&ApoTras) vs. controls, and 24 genes are differentially expressed for ApoTras vs. Her ([Additional File 3 Tables S3, S4](#)). To compare the dependence of the fold change (FC) on gene expression level, the change in relative abundance was plotted as $\log_2$ transformed FC ($\log_2(\text{FC})$, y-axis) against the absolute abundance expressed as $\log_2$ counts per million reads ($\log_2(\text{CPM})$, x-axis) (**Figure 2-2C and D**). Data was presented for Her&ApoTras vs. Con (**Figure 2-2C**) and ApoTras vs. Her (**Figure 2-2D**), with changes much more abundant and with greater magnitude of change in the Her&ApoTras vs. Con comparison. For the ApoTras vs. Her comparison there were relatively few DEG, the magnitude of relative change $\log_2(\text{FC})$

was considerably smaller, and a reduced dependence on absolute gene abundance was apparent.

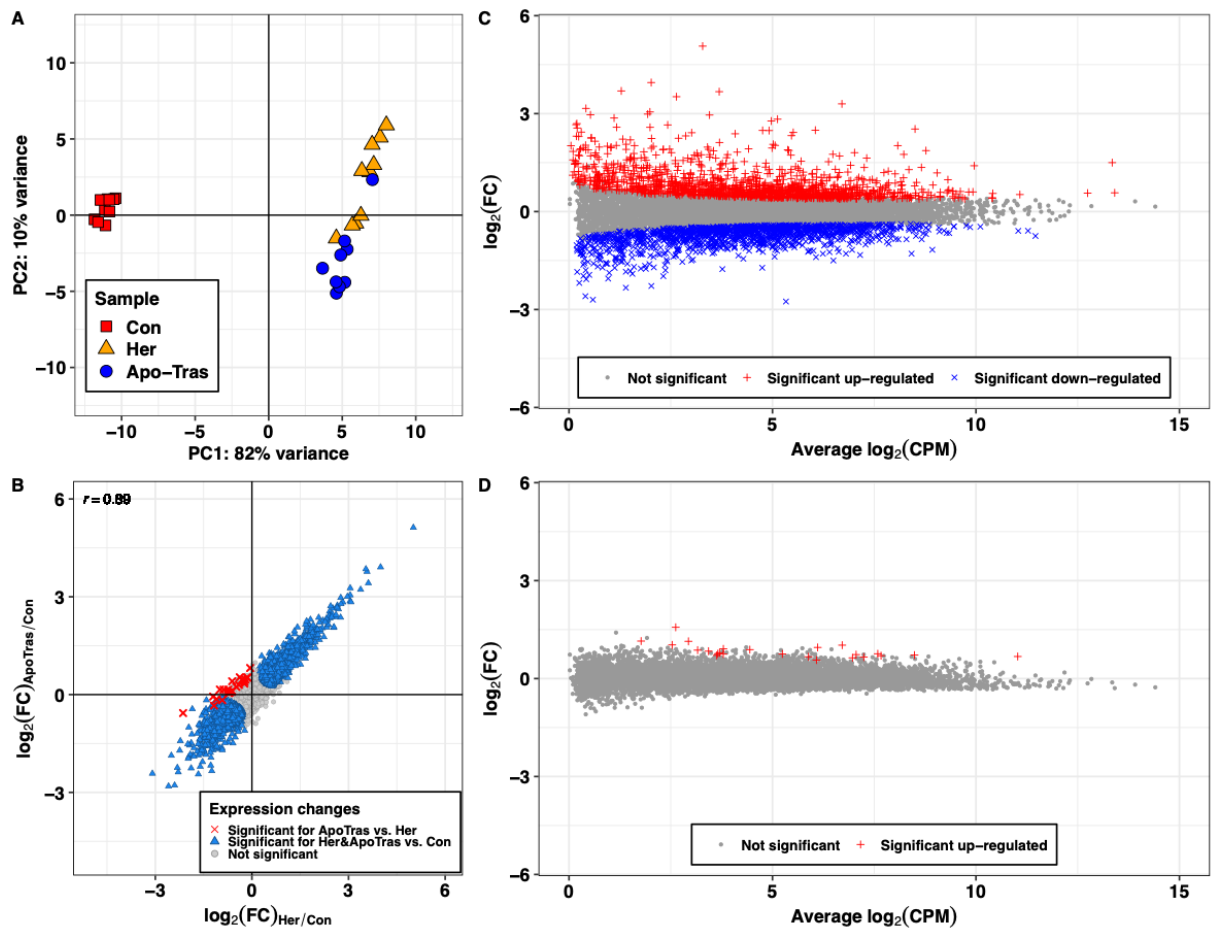


Figure 2-2 Sample clustering and detected differentially expressed genes. A) Principal components for sample clustering based on gene expression. BT474 cells in response to trastuzumab treatment for Her (n = 10), ApoTras (n = 9) and control groups (n = 10) are shown. Each symbol represents one sample. B) $\log_2(FC)$ gene expression ratios of Her/Control (x-axis) and ApoTras/Control (y-axis). Significant differentially expressed genes for the two contrasts ApoTras vs. Her and Her&ApoTras vs. Con are highlighted. Each symbol represents one gene of 11,910 total. Pearson's correlation R is shown in the upper left corner for gene expression ratios. C) and D) Differences in gene expression $\log_2(FC)$ as a function of gene expression level $\log(\text{CPM})$. Her&ApoTras vs. Con (C) and ApoTras vs. Her (D) contrasts are shown. They show the distribution for Her&ApoTras vs. Con ($q = 0.05$, $p = 0.011$) and ApoTras vs. Her ($q = 0.05$, $p = 0.000098$) respectively. Each symbol represents one gene of 11,910.

2.3.2.2 Type II error analysis for adjusted effect size

We implemented a resampling method to estimate type II error rate, in which expression differences were adjusted to a fixed fold change, then repeatedly resampled with replacement, and reanalysed for DEGs (Figure 2-3). The resampling method was applied to the complete data sets and for the same comparisons used for the detection of DEG. The examined contrasts were Her&ApoTras (n = 10 and 9 respectively) vs. Con (n = 10) and ApoTras (n = 9) vs. Her (n = 10), consistent with the sample replication in the

original dataset. FDR threshold α -values calculated in the original data analysis were held constant, while gene expression values were recalculated based on resampling data where the fold change was adjusted to 1.5 or 2 for all genes. A fold change of 2 was selected as it has been commonly used as a threshold to define biologically relevant changes and therefore can help to compare across studies. Additionally, we showed the results for the stricter fold change of 1.5 to emphasize that even with a minimum change in expression of 50% a high overall power was still achieved for the experiment.

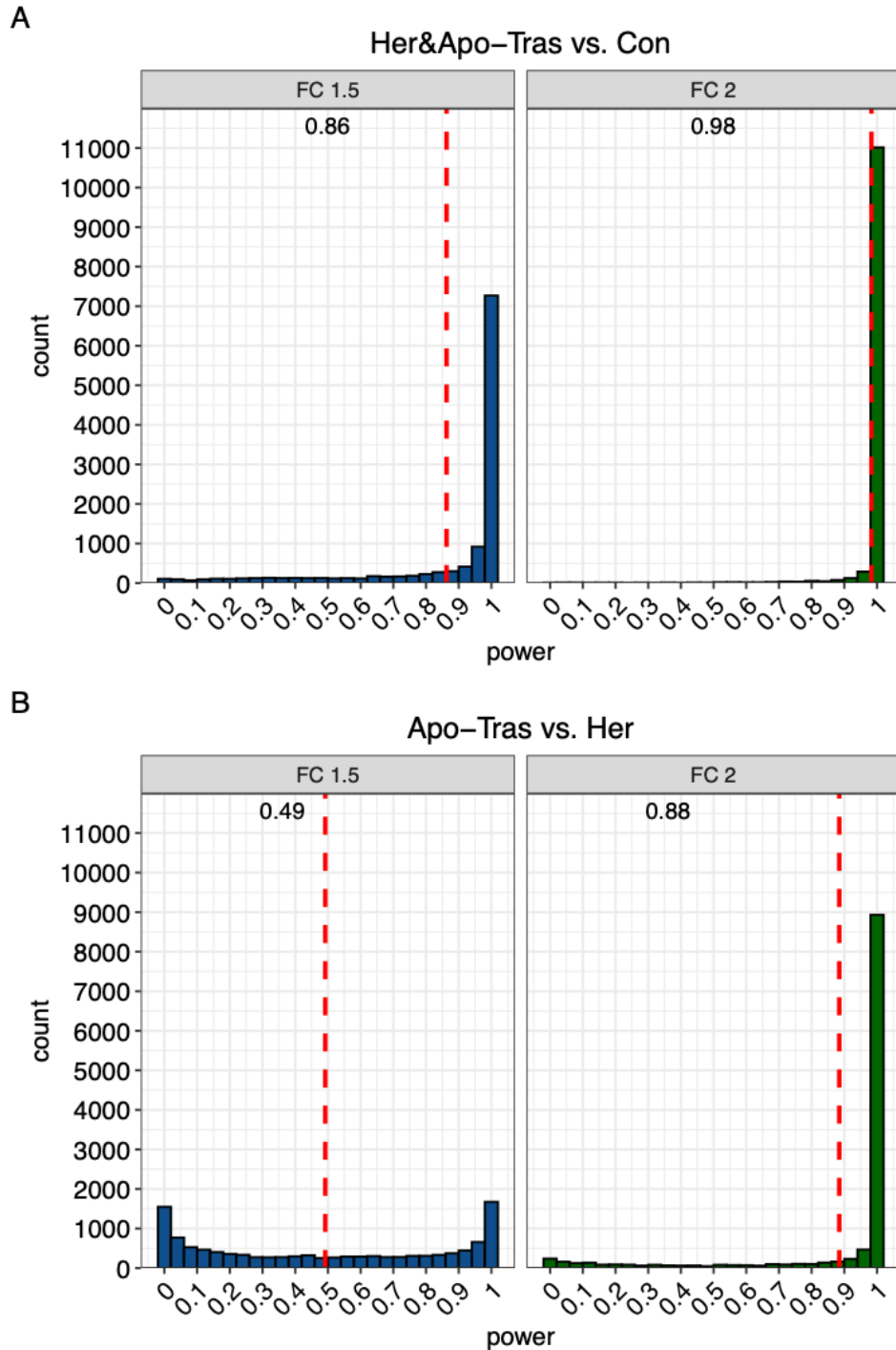


Figure 2-3 Histogram of the estimated power for all genes at a fold change of 1.5 (-33%/+50%) and 2.0 (-50%/+100%). A) Her&ApoTras vs. Con, FDR $q = 0.05$ and $\alpha = 0.011$; B) ApoTras vs. Her, FDR $q = 0.05$ and $\alpha = 0.000098$. The dashed red vertical line marks the average power for all genes. The corresponding overall power is noted at the top of the plot.

Within the current study a high number of sample replicates were used, and a relatively low variability is expected for cell line samples with minimal between-replicate genetic variation. Statistical significance for DEG was determined using a fixed $q = 0.05$, which results in an α -value of 0.011 for Her&ApoTras vs. Con, and an α -value

of 0.000098 for the ApoTras vs. Her. The considerably smaller α -value for ApoTras vs. Her was due to relatively few significant differences present. In **Figure 2-3A**, statistical power is presented for the Her&ApoTras vs. Con comparison at effect levels of $FC_P = 1.5$ and 2.0, and was calculated for all 11,910 genes (as $\min FC = 1.25$ was used as the threshold for detecting significant changes, expected power at $FC_P = 1.25$, approaches zero for the resampled data). The corresponding estimated power at $FC_P = 1.5$ and 2.0 is 0.86 and 0.98 respectively, so that we expected that about 86% of expression changes at $FC_P = 1.5$ would be detected, and about 98% of two-fold changes would have been detected. The ApoTras vs. Her comparison had less power, consistent with the lower number of statistically significant changes present and the corresponding lower value of α (**Figure 2-3B**). However, reasonable power was still achieved, with 0.49 and 0.88 for effect sizes of 1.5-fold and 2.0-fold respectively.

2.3.2.3 Similarity index for comparison of treatments

Our implemented similarity index focused on the significant differences between comparison of the treatments against the control and the comparison of the treatments against each other. The similarity index was calculated for the $\min FC$ of 1.25 and 2 for our dataset:

$$SI_{1.25} = \frac{DEG(\text{Her\&ApoTras vs. Con}) - DEG(\text{ApoTras vs. Her})}{DEG(\text{Her\&ApoTras vs. Con}) \cup DEG(\text{ApoTras vs. Her})} = \frac{2623 - 24}{2640} \\ = 0.984$$

$$SI_2 = \frac{235 - 0}{235} = 1$$

For the treatments within this study, with an SI of 0.984 and 1 for a $\min FC$ of 1.25 or 2 respectively, indicating a high degree of transcriptomic similarity between the originator (Herceptin) and the biosimilar (ApoTras) in the BT474 cell line.

2.3.2.4 Biological analysis of DEG for Her&ApoTras vs. Con and ApoTras vs. Her

2.3.2.4.1 Common changes in pathways identified in Herceptin and ApoTras treatments

DEG identified in Her&ApoTras vs. Con were classified by function using a gene set overrepresentation analysis of KEGG pathways, excluding the human disease pathway subset. The human disease pathways were excluded because many cancer pathways and viral infections impact cell cycle regulation genes, which trastuzumab also impacts due

to its mode of action. As a result, changes that correlate with specific disease pathways are less informative due the link through cell cycle genes and may hinder the detection of other differentially down-regulated pathways. KEGG pathways with the highest enrichment of DEG are shown in **Figure 2-4**, along with the number and ratio of DEG within each pathway. This summarizes the induced expression changes and identifies essential cellular functions linked to trastuzumab treatments. The annotation of differentially expressed genes to KEGG pathways are shown in [Additional file 4 Table S9](#).

Overall, 54 KEGG pathways were detected as overrepresented in the up- or down-regulated gene list with a threshold of 5% FDR. Cell cycle and DNA-replication related pathways were mainly overrepresented in the down-regulated genes whereas cellular processes, drug metabolism and signal processing related pathways were overrepresented in up-regulated genes. Pathways related to biosynthesis show a more complex pattern. Most of these changed pathways are related to lipid, amino acid, and nucleotide metabolism. Pathways which showed mainly upregulated gene expression included metabolism of caffeine, retinol, sugars, long-chain unsaturated fatty acids, glycerophospholipids and steroid hormones. In contrast, pathways which showed mainly downregulated genes can be summarized as folate, ketone bodies, amino acids, nucleotide related metabolism. This network of changed pathways depicts clear and common changes of Herceptin and the biosimilar against the control samples.

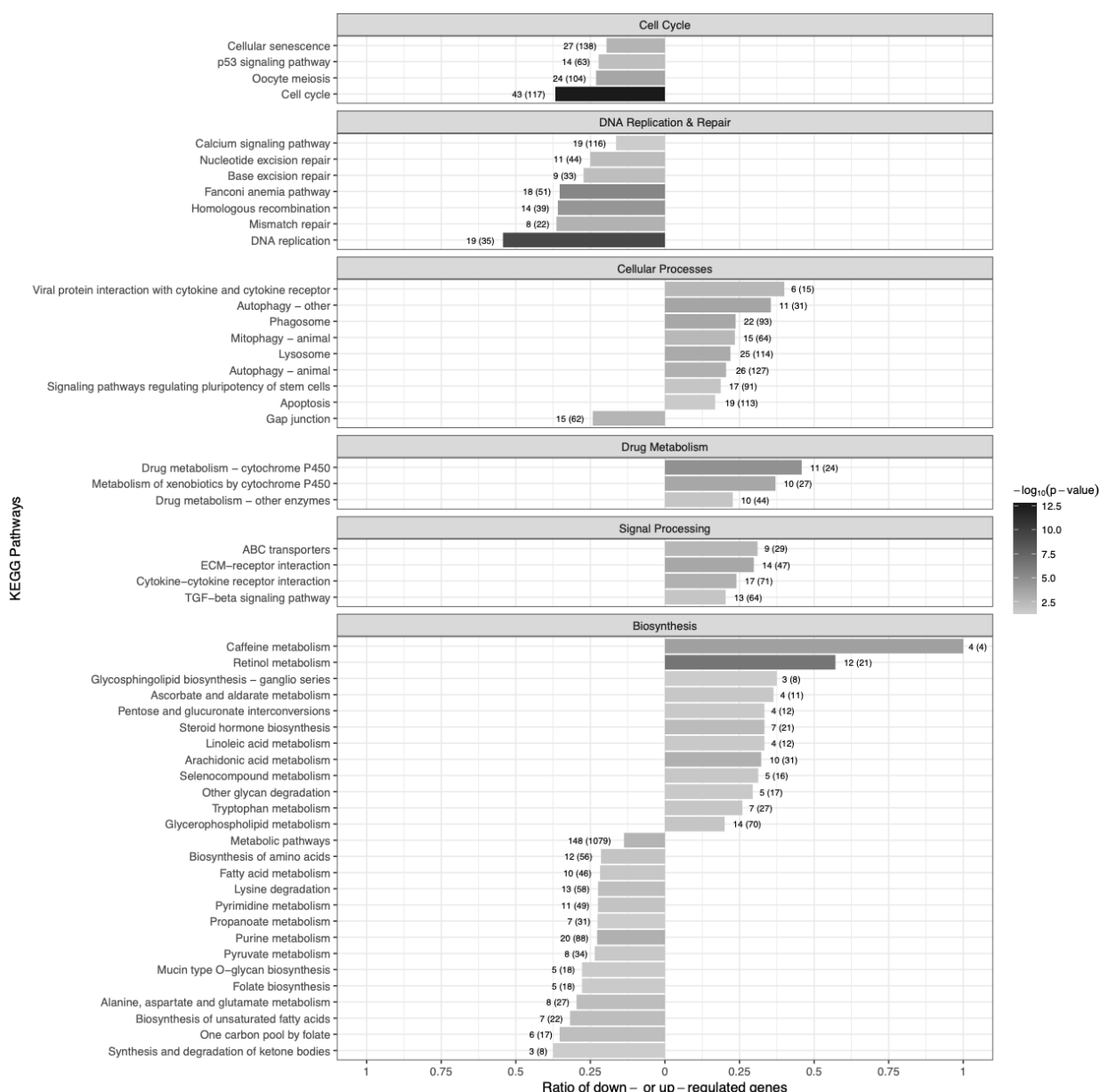


Figure 2-4 Functional profiling of differentially expressed genes. Functional profiling of DEG for Her&ApoTras vs. Con. Analysis of overrepresented KEGG pathways excluding disease related gene sets. The pathways are grouped according to their cellular function. The analysis was performed with the function `kegga` in the package `limma`. Bar length indicates the proportion of genes within the pathways that are down-regulated, and the numbers adjacent to bars indicate the number of genes identified. $\log_{10}(p\text{-values})$ are represented by the colour scheme. A p-value of 0.05 was used as the threshold. The ratio was calculated by dividing the number of identified genes by the total number of annotated genes in that pathway.

2.3.2.4.2 Differences between Herceptin and biosimilar treatments

The direct comparison of ApoTras vs. Her identifies only a small number of DEG. The 24 genes and the corresponding $\log_2(FC)$ values are shown in **Table 2-2**. All detected DEG have a higher expression in the ApoTras treated cells relative to Her samples. The limited number of DEGs identified within ApoTras vs. Her prohibits a statistically

significant gene set overrepresentation analysis. Relative to controls, most of the identified genes had reduced expression in Her, whereas the ApoTras samples had expression levels closer to control values, with some exceptions such as DST and AHNAK (Appendix Figure A 2). However, both treatments cause relatively small changes in gene expression, and the observed FC is less than two for 19 out of 24 differentially expressed genes. Five genes identified within the ApoTras vs. Her comparison, ANKRD36, KMT2A, NCR3LG1, POTEF, and ZNF445, are also decreased in abundance in the comparison of (Her&ApoTras) vs. Con. In total, 13 of the 24 DEG are known to be localized in the nucleus. Six of them contain DNA binding domains (REL, ATM, KMT2C, KMT2A, ZFHX4, ZNF445). Another functional group in the DEGs are membrane associated genes, including NC3LG1, PCLO, FAT1, MDN1, HERC1, FMN1, FREM2, AHNAK, DST. Six of these genes, PCLO, FAT1, FMN1, FREM2, AHNAK and DST, are directly involved in cell adhesions and junctions. In summary, the small number of DEG genes identified between Her and ApoTras are primarily membrane bound proteins and DNA binding domain containing proteins.

Table 2-2 Differentially expressed genes (Name and HNGC symbol) identified in ApoTras vs. Her. Genes are ordered according to their log2(FC). FDR value of 5% was applied as threshold.

HGNC symbol	Gene name	logFC	logCPM	P-Value	FDR
POTEF	POTE ankyrin domain family member F	1.57	2.62	0.000027	0.01978
REL	REL proto-oncogene, NF-kB subunit	1.15	1.77	0.000013	0.01978
NCR3LG1	neurobeachin like 1	1.14	2.93	0.000001	0.00714
ZFH4	zinc finger homeobox 4	1.03	2.53	0.000003	0.01588
FREM2	FRAS1 related extracellular matrix 2	1.02	6.71	0.000038	0.02375
KMT2A	lysine methyltransferase 2A	0.95	6.11	0.000023	0.01978
NBEAL1	natural killer cell cytotoxicity receptor 3 ligand 1	0.91	3.79	0.000024	0.01978
PCLO	piccolo presynaptic cytomatrix protein	0.89	4.44	0.000005	0.01953
MAP1B	microtubule associated protein 1B	0.87	3.16	0.000029	0.01978
ANKRD36	ankyrin repeat domain 36	0.84	3.43	0.000040	0.02375
FMN1	formin 1	0.79	3.70	0.000020	0.01978
EXPH5	exophilin 5	0.76	3.66	0.000023	0.01978
KMT2C	lysine methyltransferase 2C	0.76	7.59	0.000087	0.04502
ATM	ATM serine/threonine kinase	0.75	5.24	0.000013	0.01978
ZNF445	zinc finger protein 445	0.75	3.79	0.000025	0.01978
MDN1	midasin AAA ATPase 1	0.73	7.06	0.000029	0.01978
SYNE2	spectrin repeat containing nuclear envelope protein 2	0.72	8.49	0.000030	0.01978
ADGRV1	adhesion G protein-coupled receptor V1	0.71	3.63	0.000030	0.01978
FAT1	FAT atypical cadherin 1	0.70	7.67	0.000024	0.01978
AHNAK	AHNAK nucleoprotein	0.68	11.03	0.000074	0.04212
CEP350	centrosomal protein 350	0.67	5.89	0.000098	0.04862
DST	dystonin	0.66	7.24	0.000080	0.04313
HERC1	HECT and RLD domain containing E3 ubiquitin protein ligase family member 1	0.64	6.97	0.000008	0.01978
BRWD1	bromodomain and WD repeat domain containing 1	0.57	6.08	0.000021	0.01978

2.4 Discussion

The goal was to analyze RNA-seq to support the claim of biosimilarity between the therapeutic antibody Herceptin and its biosimilar. The impact of trastuzumab on the cell line's gene expression network goes far beyond reduced proliferation with potential off-target effects caused by slight difference in post-translational modification. Currently, the functional comparison of a biosimilar to its originator product is mostly limited to proliferation assays or target binding estimations. However, the mechanism of action for trastuzumab can impact and change multiple pathways regulating cell proliferation, apoptosis, angiogenesis. Furthermore, patients often acquire resistance during the treatment or are intrinsically resistant to Herceptin. Multiple probable resistance mechanisms have been described. Therefore, a narrowly targeted approach to establish biosimilarity cannot be sufficient to prove similarity of effect. Thus, this concise, untargeted RNA-seq analysis approach can provide a more comprehensive picture to compare and prove the same mode of action. The main difficulty here is that RNA-seq data mainly focused on hypothesis development but not for testing a hypothesis. Here, we showed a way of how RNA-seq can be interpreted to use its full informational content to establish biosimilarity with the example of Herceptin.

To push the boundaries of RNA-seq application, we established four key steps for the analysis after transcript quantification. The main comparisons for the analysis were not comparing each treatment separately to the control samples but by grouping both treatments and comparing that group to the control group (Her & ApoTras vs. Con) while also checking for significant changes between both treatments (ApoTras vs. Her). To our knowledge, this type of contrast has not been previously applied for treatment comparisons. Using these two contrasts, we established a similarity index (SI) that focuses on the detected significant changes. This similarity index avoids assumptions about non-significantly changed genes by focusing on the detected differentially expressed genes which reduces the assumptions about not significantly changed gene expressions. Next, studying the type II error for our fold change of interest, we can deduct whether genes were non-significant due to small effect size or insufficient group size. This is essential to differentiate between unaffected genes and genes with high type II error rate. And finally, the gene set overrepresentation analysis provides the biological interpretation of the data. This allows integration of the common changes caused by both treatments with known research to check for missing pathways or unexpected changes.

Similarly, for the contrast between both treatments we can confirm that there are no biologically concerning changes changed between the two treatments.

For our concrete comparison of the trastuzumab biosimilar produced by Apotex, we can confirm a strong biosimilarity based on the RNA-seq data. We estimated two SI indices for minFC of 2 and 1.25 with 0.984 and 1 showing a high similarity of detected expression changes. The advantage of the presented SI is that it penalizes differences between two treatments in proportion to the total number of DEG which the treatments have in common. Additionally, when DEG detected between both treatments are differentially regulated in the contrast of both treatments against the control as well, the penalizing effect is slightly smaller. This effect is preferable because if the genes are differentially expressed in both contrasts, the direction of change for both treatments against the control is likely to be identical. Therefore, the similarity index emphasizes the difference of the treatments by the actual significant detectable differences as DEG between both groups instead of the absence thereof. That is a key difference compared to other common binary asymmetric similarity indices like the Jaccard (Jaccard, 1912), Dice (Dice, 1945), Sokal (Sokal & Sneath, 1963) or Szymkiewicz-Simpson (Simpson, 1943) coefficient. Jaccard and Szymkiewicz-Simpson are often used to determine the overlap of gene sets (Merico et al., 2010). All of them would only consider the contrasts of either A or B vs C and would not consider DEG detected in the contrast of A vs. B. Accordingly, the introduced SI value presents a useful tool to compare the overall outcome of the measured transcriptomic changes induced by the two treatments.

Next, the type II error analysis reveals that the power for detecting genes as significantly changed was over the commonly used 80% for a minFC of 2. That further strengthens the result of the similarity index because the type II error rate is low for this specific effect size. In contrast, we detected that for the minFC of 1.25 both contrasts have a much lower power which could conceal significantly changed genes because of a potentially higher false negative rate. That means our sample size was big enough for a minFC of 2 but not for an minFC of 1.25 to detect significant changes with a common 80% power threshold. Therefore, the power analysis supports the calculated similarity index of 1 for an minFC of 2. In contrast, for the minFC1.25 a larger sample size would be needed to confirm the similarity index. However, an expression changes of 25%/20% represents relatively small change which might not be biologically relevant.

The third step was the biological analysis allows the confirmation of detected known changes of the drug. The biological interpretation of the significant changes for an

minFC of 1.25 allows interpretation the biological effect of the two treatments. For the contrast of both treatments against the control approximately 1311 genes were upregulated, and 1312 genes were downregulated. For this comparison, an overrepresentation analysis was performed to test for overrepresented KEGG pathways for either up- or down-regulated genes. As expected, trastuzumab treatments resulted in multiple pathways with reduced gene expression relative to controls, and cell proliferation being one of the most significantly impacted pathways. Several of the identified pathways with decreased gene expression are additionally connected to cell proliferation, such as cell cycle, nucleotide synthesis, and DNA replication. Reduced proliferation is one of the key cellular effects of Her treatment on target cancer cells (Yakes et al., 2002). Furthermore, mechanisms involving DNA repair are also decreased which is a common response to the treatment of breast cancer cells with Her treatment (Mayfield et al., 2001). These changes in key pathways of proliferation are detected in mAb treatments which support the expectation that the biosimilar ApoTras has a similar effect to the originator product Herceptin. Increased expression in pathways of drug metabolism is also observed and can be interpreted as a generalized response to a mAb treatment. Additionally, genes related to autophagy are upregulated. Increased autophagy was identified as common feature of Herceptin treatments and has been previously implicated in the development of Her resistance in breast cancer (Vazquez-Martin et al., 2009). In addition, Herceptin treatment increased expression of genes regulating the production of lipids involved in cell signalling and membrane composition. Changes in lipid metabolism are considered as a hallmark of cancer (Zhang, 2012), though the impact of Herceptin treatment on the lipid metabolism remains to be investigated. However, that effect was observed for both mAb treatments, and would have been missed in cell-based assays.

Next, we analyzed the short list of DEG between both treatments with 24 genes. Because of the limited number of genes, no overrepresentation analysis was performed, and the genes were mainly analyzed by their known impact on cancer progression. The resulting picture was inconclusive. All 24 genes were upregulated in ApoTras samples compared to Her samples. Currently, none of the genes has been studied in context of trastuzumab treatment of Her2⁺ breast cancer. However, REL, ATM, KMT2C, KMT2A, FAT1, and AHNAK have been associated with cancer progression, but in opposite directions. REL (Barkett & Gilmore, 1999; Rayet & G  linas, 1999), KMT2C (Dawkins et al., 2016) were previously described as tumour promoting factors whereas ATM (Vo

et al., 2004), KMT2A (W. Lin et al., 2016) and FAT1 (Morris et al., 2013) were associated with reduced cancer growth. In summary, several of DEG may influence cancer progression; however, their net effect is unclear as the potential effects are both positive and negative. In terms of demonstrating biosimilarity, this comparison of transcriptomic data reduces the number of potentially relevant genes to a short list which could be further investigated by targeted approaches. Overall, in comparison with the large changes in gene expression within the Her&ApoTras vs. Con comparison, the differences between ApoTras and Her are small and do not suggest a clear direction of possible effect. The net biological importance of the detected DEG appears quite limited.

Overall, in combination of the different analysis parts, we can state that the similarity index shows strong agreement in treatment effects for a minFC of 2, with robust support from the type II error analysis proving that the power was over 80% to detect significant changes for each contrast. Furthermore, the biological interpretation of the overrepresentation analysis confirms that the common changes by both treatments are mostly related to a trastuzumab treated cell response. The impact of the treatments on the lipid metabolism have not been studied in detail but might of interest considering the importance of lipids in cancer regulation. Also, the detailed analysis of the contrast ApoTras vs. Her suggests an inconclusive pattern among differentially regulated genes related to cancer progression, but by considering the power analysis, which showed a power of below 50%, indicates that multiple truly differentially expressed might not have been detected and would need further study. Therefore, the biosimilarity for minFC of 2 is high whereas the reliability of the results for a minFC of 1.25 is not as strong.

Even though the analysis was performed for one therapeutic monoclonal antibody, we are confident this type of analysis is easily transferable to other biologics and potential biosimilars to compare the mechanism of action. Therefore, this analysis can help to filter for promising biosimilars in an early stage of the development based on a comprehensive untargeted approach. This can be especially helpful for biologics whose mechanisms of action are complex or not yet fully understood.

Chapter 3: Type II error estimation for RNA-sequencing experiments using a bootstrap approach

3.1 Introduction

The power of a study indicates the likelihood that the null hypothesis is correctly rejected. Although RNA-seq has become a commonly used tool, it is very powerful for detecting global transcriptomic changes in organisms, however studies about the statistical power of RNA-sequencing (RNA-seq) experiments are very limited. RNA-seq experiments are effective at detecting novel transcripts and have very good dynamic range and sensitivity. This is especially helpful when a treatment has unknown effects or exerts global impacts by changing the activity of transcription factors. This has led to an extensive amount of research to optimize the analysis of RNA-seq experiments in terms of quality assessment, alignment quality, and statistical inference. However, a common conception is that RNA-seq experiments are underpowered (Cui et al., 2018; Y. Liu et al., 2014; Schurch et al., 2016). This assumption often limits the interpretation of RNA-seq experiments.

The amount of research concerned with the power of RNA-seq experiments is mainly focused on the determination of a minimum group size for future studies. Several developed programs like PROPER (H. Wu et al., 2015), RnaSeqSampleSize (Zhao et al., 2018), or RNAseqPS (Guo et al., 2014) predict the power of future RNA-seq experiments to estimate appropriate group sizes. They simulate data by sampling from statistical distributions to create expression data. For that, some constraints that can be set randomly are the number of detected genes, the proportion of differentially expressed genes or the expression level distribution of the genes. That implicates multiple biases by assuming properties as well as breaking the regulation network of gene expression. Furthermore, their estimations are intentionally conservative to avoid issues of too high predicted power. These tools are useful for planning experiments but have limited value for estimating power for effect sizes of interest in a conducted experiment.

The statistical power for a specific effect size can add valuable information to the analysis of RNA-seq experiments and provides a complement to identifying differentially expressed genes. By studying the power for a specific effect size, we can differentiate between the impact of small group size versus too small an effect size. The

latter would imply that the experiment may have enough power to detect differences, if the effect size is of a suitable magnitude. This analysis can widen the understanding of studies expecting small differences in expression, such as the comparison between similar drug treatments, e.g. therapeutic monoclonal antibodies and biosimilars. To assess drug treatment similarity, RNA-seq experiments can add considerable value because they can cover most of the transcriptome and thus are less likely to miss unexpected changes. In these instances, the comparison of a few biomarkers or the target-binding affinities is unlikely to be sufficient for establishing similar efficacy. Estimating the power for a specific effect size in such experiments can shed a light on the reproducibility of these studies and show how the sample size or the specified effect size were suitable for the experiment. As highly dimensional data, RNAseq power analysis is not a simple post hoc procedure, however.

One way to investigate complex datasets is to simulate data closely related to the original, using resampling or bootstrap approaches (Efron, 1979). Bootstrapping allows the simulation of datasets with a highly similar structure to the original dataset. In general, bootstrapping refers to sampling with replacement from the original dataset while maintaining the same group size (N). This simplicity of the bootstrap allows to test if the results are stable without assuming distribution characteristics. Bootstrap approaches are often used to estimate confidence intervals (DiCiccio & Efron, 1996), error rates (Jain et al., 1987) or statistical hypothesis testing (Hall & Wilson, 1991). By sampling with replacement from each experimental group, the in-sample data structure is not impacted and thus, less biases are introduced into the analysis. Also, the expression level will still be representing the specific gene as well as the dependency of transcription levels between genes. One limitation of bootstrapping is the dependency of the dataset being representative for the population distribution. A bootstrap approach represents a complementary tool to estimate the statistical power of RNA-seq data without disturbing the sample structure, but to date has not been implemented for power analysis of RNA-seq experiments.

We hypothesize that a bootstrap approach with effect size adjustment provides a method to accurately estimate the statistical power of RNA-seq experiments. Importantly, this can provide information on whether genes are not differentially expressed, or the experiment was underpowered and had a low ability to detect differences. Furthermore, this can inform the dependency of the power of a gene on the expression level.

3.2 Materials and Methods

3.2.1 Data Set Preparation

Two extant data sets were analyzed to demonstrate the utility of the developed method on real data. The datasets have very different data structures in terms of their sequencing depth, sample origin and replication number. The first dataset originates from the breast cell line BT474. The RSEM-expected count data for the study is accessible through the Gene Expression Omnibus (GEO) series accession number (GSE: 158969). Samples were untreated ($n = 10$, Con), treated with the anti-HER2 monoclonal antibody trastuzumab ($n = 10$, Treat #1) or treated with a trastuzumab biosimilar ($n = 9$, Treat #2). The methods are described in detail in Chapter 2: and biosimilar treated sample replicate 8 was excluded due to poor quality sample, as noted. In summary, the cells were treated with either Her or ApoTras antibodies twice over seven days in three-day intervals, before the total RNA content was extracted. The mRNA selection, library preparation and sequencing were performed in the Donnelly Centre, University of Toronto, Toronto, Canada. The generated FASTQ files were aligned to the human transcriptome release from the consensus coding sequence (CCDS)(Farrell et al., 2014) version GRCh38.p7 by using Bowtie v1.2.1.1 (Langmead et al., 2009). The transcripts were quantified with RSEM v1.3.0 (B. Li & Dewey, 2011) to produce the expected count table.

The second RNA-seq data set was a subset of the breast invasive carcinoma (BRCA) project from the Cancer Genome Atlas (TCGA) program (Weinstein et al., 2013). The dataset is intensively studied and commonly used for testing RNA-seq analysis related algorithms. The focus in this study was to classify cancer samples for stage IIB and stage IIIA tumours. In total, 88 biological replicates per group were inspected. The pipeline of the National Institutes of Health (NIH) was followed and reads were aligned to the human reference genome GRCh38.p0 with STAR v2.4.2a, followed by HTSeq v0.6.1p1(Anders et al., 2015) to quantify expression levels. For all 88 samples, the unnormalized counts were downloaded from Genomic Data Commons (GDC) (<https://portal.gdc.cancer.gov/>) of the National Institutes of Health (NIH).

3.2.2 Data Filtering

The expression data were transformed to count per million (CPM) values with the edgeR function `cpm()`. Then, the expression data for each experiment were filtered for a

CPM value equivalent to at least 10 counts in the sample with the smallest library size, present in at least 9 samples for the treated cell line data or at least 88 samples in the TCGA dataset. The number of genes passing this filtering step were 11,910 for the treated cell line study and 20,415 for the TCGA dataset. The filtered data were then sampled to the noted group size per conditions with function `sample()` part of the R base package without replacement. The output was then used for our power analysis algorithm.

3.2.3 Algorithm

3.2.3.1 FDR-threshold calculation

The first step of the algorithm is to calculate the FDR threshold for the original dataset with `edgeR` (Robinson et al., 2009). The library sizes were normalized with the weighted trimmed mean of M values (TMM) method from `edgeR`. The subsequent step was to obtain the posterior dispersion estimates with the local fit method in `estimateDisp()`. Then, the quasi-likelihood negative binomial generalized log-linear model was adjusted to the expression data using the function `glmQLFit()`. Next, the fold changes between the conditions were tested for significance with the quasi-likelihood F-test function `glmQLFTest()` with the desired FDR threshold. The FDR threshold was set to 10%, 5%, or 1% as noted. FDR-values were calculated with the Benjamini-Hochberg method. The p-values equalling the corresponding FDR-thresholds were retrieved from the results table and were later used to assess statistical significance in the power estimation of the mean-adjusted and bootstrapped datasets.

3.2.3.2 Mean adjustment

To determine power at a given effect size, mean expression values for the target gene were adjusted to the target effect size, where effect size is defined as the fold change between the conditions. Here, we investigated the effect size of 1.25, 1.5 or 2 as noted. The effect size adjustment between conditions was done for each gene separately. First, the data were transformed $\log_2$ CPM (lcpm) values, then, the means of the lcpm values were calculated for each gene and group. Following this, the mean difference was tested against the $\log_2$ value of the effect size.

To find effect size as a function of power, effect sizes were adjusted until the target power was attained. Both comparison groups were iteratively adjusted symmetrically around the overall mean to preserve in-group variance. In summary, the samples were

moved by the minimum amount to achieve the same effect size for all genes. As the last step the data were transformed back into expression data for the sampling procedure.

3.2.3.3 Smoothed bootstrap sampling

A smoothed bootstrapping approach was implemented for resampling, using random sampling with replacement. Here, it was used to estimate the power for detecting genes as differentially expressed. The smoothing eliminates the possibility of zero within-group variance after sampling with replacement by adding noise to the data. The approach involves fitting a kernel density distribution to the expression data and then sample from the smoothed distribution to add noise to the expression values. For the expression values $x_1, \dots, x_n$, the kernel density estimator is:

$$\hat{f}(x) = \frac{1}{n h} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right), \quad (1)$$

where K represents the kernel distribution and h is the bandwidth which defines the smoothing as the standard deviation. A Gaussian kernel distribution with the standard deviation of 1 and mean of 0 was implemented. The bandwidth is chosen by minimizing the mean integrated squared error and can be estimated with the Scott method for a Gaussian density function. The bandwidth was defined as follows:

$$h = \frac{1.06 \times \min(s, \frac{IQR}{1.34})}{n^{1/5}}, \quad (2)$$

in which s is the standard deviation of gene expression and interquartile range (IQR) represents the difference from the 25th percentile to the 75th percentile. N is the sample size of the group. The function is implemented in base R command `bw.nrd()`. From the estimated kernel density function, the noise is generated and then added to the `lcpm`. The smoothing was applied after the sampling step of n samples with replacement from the mean adjusted data set. This resampling and smoothing step was repeated for the indicated number of iterations in the range of 10 to 400 times as noted.

3.2.3.4 Power assessment

Statistical analysis was performed with `edgeR`. For that, the `lcpm` expression values were transformed back to count data using the original library size and normalization factor of the original dataset to prevent any changes in the library size and sample composition. As before, the `glmQLFit` and `glmQLFTest` function were used to apply the quasi-likelihood F-test in `edgeR`. Then the previous estimated p-values bordering the FDR threshold were used to define significant tests. The statistical power is defined as 1

– β in which β is the Type II Error of a test. Here, the power may also be described as the proportion of successful tests for each gene and iteration for a given comparison. From these estimates, we calculated the per-gene power and overall experiments power.

Error evaluation of adjusted effect changes

The final calculated fold changes from the power analysis section were used to estimate the resampling error of the bootstrapping approach. For that, we calculated root mean square errors (RMSE). The RMSE can be written as

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}}, \quad (3)$$

in which x_i represents the mean derived from the mean of all adjusted and resampled dataset, $\hat{x}_i$ being the mean of the mean-adjusted dataset without resampling, and n representing the number of calculated means which equals the number of sampling iterations. Therefore, RMSE can be used to estimate the standard deviation of the error distribution. Because the RMSE value is relative, it has to be set into context to the actual mean expression value. RMSE is sensitive to large errors, which are undesirable for the bootstrap approach.

3.2.4 Visualization

All figures were generated with the ggplot2 (Wickham, 2016) package for R.

3.2.5 Additional Files and Code Availability

Additional files and scripts for preprocessing, analysis, and figure creation are available in the GitHub repository at https://github.com/karstenrinas/Thesis/Chapter_3.

3.3 Results

3.3.1 Real-life Data Introduction

To evaluate our resampling approach, we examined two RNA-seq data sets. The datasets differ in replication number, sample origin, sequencing depth and study design. One dataset was derived from the breast cancer cell line BT-474 treated with no drug (Con), drug #1 (Treat #1) or a biosimilar version of drug #1 (Treat #2). Genes typically showed a very low within-group variance but a high between-group variances. The replicates per condition were derived from the same cell line and three contrasts were analyzed: Con vs Treat# 1 (treated cell line), Treat #1 vs Treat #2 (treated cell line), and Stage IIB vs Stage IIIA (TCGA).

The second RNA-seq data set was a subset of the invasive breast carcinoma (BRCA) study from the Cancer Genome Atlas project (TCGA dataset)(Ciriello et al., 2015). For cancer stage IIB and IIIA, 88 replicates were selected to have equal group size. These samples are derived from different individuals, and as a result the in-group variance is significantly higher, and the between group variance is small compared to the BT474 cell line dataset. The sample clustering can be seen in the PCA plots (Appendix Figure B 1) and the identified and quantified genes per dataset are summarized in Table 1. In the following, three contrasts were analyzed: Con vs Treat# 1 (treated cell line), Treat #1 vs Treat #2 (treated cell line), and Stage IIB vs Stage IIIA (TCGA).

Table 3-1 Summary of identified genes per dataset. TCGA: Dataset contig RNA-seq data derived from breast cancer samples. Treated cell line data: BT474 cell line with a control group and two treatment groups. LCPM: $\log_2$ (counts per million).

Dataset	# of identified genes	# of groups	# of replicates per group	Mean expression (lcpm)	Range of expression (lcpm)	Average expression variance
TCGA	20,415	2	88	2.76	-4.97 – 16.80	3.518
Treated cell line	11,910	3	10 (Con, Treat #1), 9 (Treat #2)	4.65	-1.57 – 14.82	0.043

3.3.2 Power Analysis Method Implementation

The implemented bootstrapping method **Figure 3-1** can estimate the power for two group comparisons within an RNA-seq study. First, the input data set was filtered for a minimum expression level and a subset for the number of replicates per group selected without replacement. The table with gene expression levels for each sample and condition was then used to compute the original FDR p-value threshold for a specific

group size. After that, raw counts were transformed to lcpm values and the fold change between groups was adjusted to a target effect size for power analysis. Next, the samples per condition were resampled with replacement with 10 to 400 iterations. For the smoothed bootstrap method, noise was sampled from a gaussian kernel distribution for each gene and sample. The bandwidth estimator controlling the standard deviation of the kernel was estimated with the Scott method. For each resampled dataset, the differentially expressed genes (DEG) were detected with edgeR, and the p-value estimated from the original dataset was applied. The statistical power was estimated for the complete contrast or for specific genes based on comparison of the p-value for each iteration vs. FDR threshold. The results estimate how likely a gene would have been detected as DEG for the specified effect size, group size and FDR threshold.

The stability of the resampling method was assessed for the estimated means from data used for error detection. The root-mean-square error (RMSE) was used to determine differences of the estimated mean values compared to the observed means. RMSE values are useful for detecting the variation in the estimator, with emphasis on large errors. The plotted results are shown Figure 3-1B and C for different group sizes and iteration numbers, respectively. The shown data are based on the Con samples from the treated cell line dataset. Overall, the RMSE values did not vary that much. The RMSE values are low even for small group sizes like 3 with a mean of 0.135 and a standard deviation of 0.114. With a group size of ten the RMSE mean value is the lowest with 0.0551 and a standard deviation of 0.0421 and over 89% of values are smaller than for a group size of 3. Also, over 95% of RMSE values of the 75th percentile is for genes with an average expression below 4 lcpm. **Figure 3-1C** depicts the effects of increased iteration numbers for a group size of 5. The mean varies less than 2% independent of iteration number. Clearly, most values scatter around the 45-degree line indicating only small deviations. However, the variance of RMSE values drops slightly from 0.0739 for 10 iterations to 0.0730 for 100 iterations with no further decrease when doubling or quadrupling the iteration number. Therefore, the iteration number was kept at 100 for the following power estimations.

Another important criterion to inspect is the effect of the resampling and empirical Bayes correction of the variance on the adjusted effect size. The LFCs were calculated for the Con samples against the Treat#1 samples with 100 iterations after detection of DEG with edgeR. The histograms were then plotted for the absolute LFCs (**Figure 3-1D**). The histogram shows high numbers of genes have an LFC close to the adjusted effect

size. The standard deviation for all three is smaller than 0.04% of the LFC. In comparison, the bottom plot shows the deviation of the LFCs from the adjusted effect size decreases tremendously with higher mean expression levels referring to more stable means. Overall, the LFC does not show any significant bias or unexpected outliers after the sampling and edgeR corrections. This small effect is expected because the variance was adjusted over all samples independent of group association.

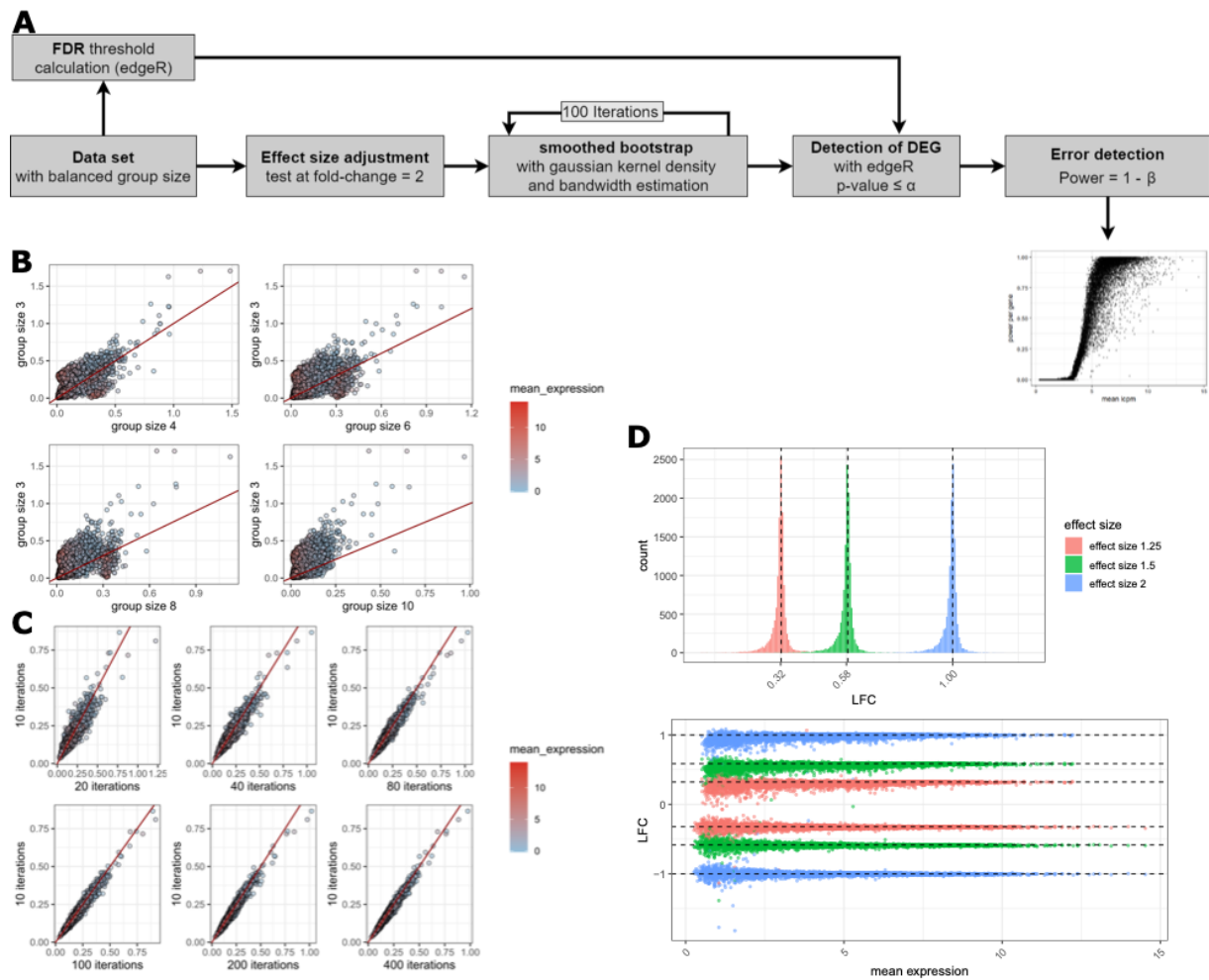


Figure 3-1 Bootstrap resampling implementation. A) Overview of the implemented power analysis algorithm. B) Increasing group size per condition reduces resulting error variation. Calculated root-mean-square error (RMSE) for resampled means from Con samples (Tras dataset, $n = 10$, 11910 genes) were plotted for group sizes of 3 against group sizes of 4, 6, 8 and 10. Each data point represents the RMSE value for one gene. C) Higher iteration numbers reduce the variance of RMSE. Like in B, Con samples (Tras dataset, $n = 10$) were used to calculate RMSE for different iteration numbers from 10 to 400. The RMSE values were plotted against the lowest iteration number of 10. Each data point represents the RMSE value for one gene. D) Log foldchange (LFC) for most genes cluster close together after effect size adjustment, resampling and edgeR analysis. The histogram (top) shows the frequencies of absolute LFCs, and the LFC against expression level are shown at the bottom. Foldchanges were calculated for Con samples ($n = 10$) against Treat #1 samples ($n = 10$) from treated cell line dataset for 11,910 genes with 100 iterations for three LFCs ($\log_2(1.25)$, $\log_2(1.5)$, $\log_2(2)$). The dashed black lines indicate the originally adjusted effect size. The mean expression level is shown as lcpm .

3.3.3 Overall Computed Power Results for Real Life Datasets

The first step was to analyze the overall estimated power based on the smoothed bootstrap method. This allows to analyze the trend and stability of power estimation for different group size, effect sizes, FDR thresholds and data structures. The overall power curves are shown in **Figure 3-2**. In general, the power is much higher for the treated cell line experiment than for the TCGA dataset which is expected because of the different between and in group variance (Appendix Figure B 1). The same is the case for the differences in the estimated power values between the two contrasts for the treated cell line. The contrast against the control shows in average a higher power than the contrast of both treatments against each other. The power increases for each comparison with increased FDR threshold and effect size. For an effect size of 2, the treated cell line contrasts show even for a small group size of three a power over 80% except for Treat #1 vs Treat #2 contrast with the stringent FDR threshold of 1%. For the TCGA data, the power surpasses 80% for a group size as small as 54 for an FDR threshold of 10% and 5%. Additionally, the impact by applying different FDR threshold shrinks with increased groups size and power for all contrasts. Another observation is the power decrease for the comparison of Treat #1 and Treat #2 despite the group size increase to 8. The effect is smaller with higher FDR thresholds but does not disappear completely. The next step to a group size of 10 replicates then follows the pattern of the other power curves. Power increased reliably across all contrasts as group size increased, with low variance.

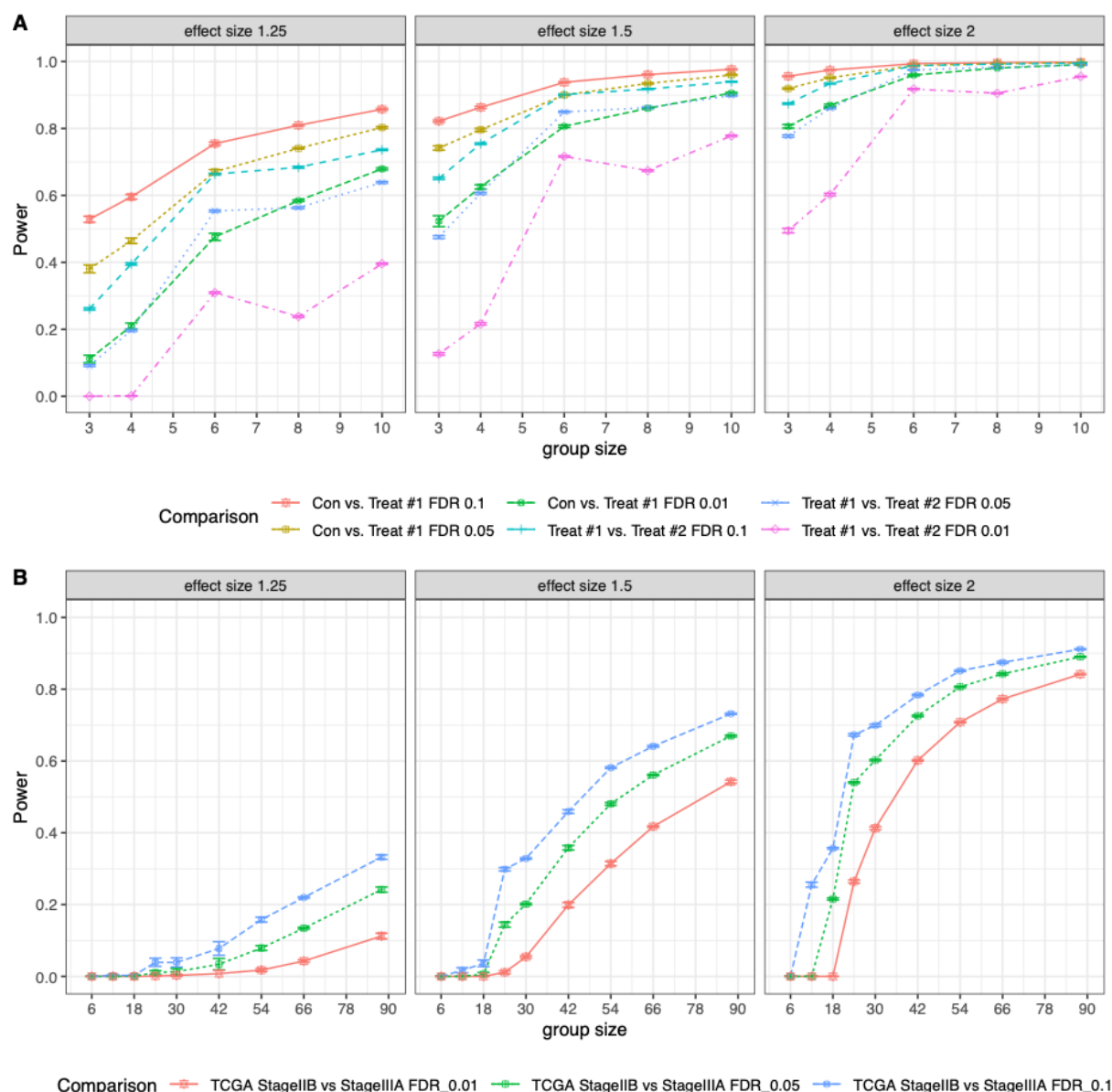


Figure 3-2 Power estimation stably increases for differing effect size, replicate number and dataset structure. A) and B) plotted power curves for range of group sizes against overall power with 100 iterations. Section A) shows the power for the treated cell line data with 11910 genes. The shown comparisons are the control group Treat #1 (Con vs. Treat 1#) and the group Treat #1 against the group Treat #2 (Treat #1 vs. Treat #2). Section B) shows the power for the TCGA dataset with 20,415 genes. The shown comparison is the group of cancer stage IIB samples against a group of cancer stage IIIA samples. For A and B, error bars show the variance for the power estimations over the 100 iterations of the bootstrap approach. The plots are split for the different effect sizes and the results are presented for three FDR thresholds (1%, 5%, 10%).

3.3.4 Expression Level Strongly Impacts the Calculated Power for Genes

The distribution of the power per gene reveals an internal dependency of the power on the expression level. This tendency is there even though the dataset was filtered for low expressed genes to improve the power which is a common procedure. To detect differences in power estimations based on the expression level, the genes were grouped

into three equally sized gene sets with data dependent expression ranges based on the calculated lcpm values. The ranges were named low, medium and high expression and were dataset dependent. For the treated cell line, the low expression group was below a lcpm of 4, the medium group was between 4 and 5.7 lcpm and the high expression group showed an expression level over 5.7 lcpm. In comparison, TCGA expression grouping showed lower expression thresholds with below 1.8 lcpm, 1.8 to 5.0 lcpm and above 5.0 lcpm for low, medium and high expression levels, respectively. The number of genes per group were 3970 for the treated cell line dataset and 6805 for the TCGA dataset. The histograms of the power distribution for all genes can be seen in **Figure 3-3**. **Figure 3-3** shows a subset of the parameters group size, FDR threshold and effect size as noted in the figure description to focus on key characteristics. The complete set of histograms can be seen in Appendix Figure B 2 and Appendix Figure B 3. The power distributions show a strong impact of the expression level of genes on the power. Genes with higher expression are highly likely to show a higher power.

For Con vs. Treat#1, 80% of the high expressed genes had a power over 80% for an FDR of 1% and effect size of 1.5. This stands in contrast to the overall power which was only slightly above 60% (**Figure 3-2A**). Indeed, only one half of the medium expressed genes have a power over 80%. In contrast, the low expressed group showed a power below 50% for most genes. A groups size increase to six showed the expected shift towards higher power, but differences between the expression groups follow the same pattern. Almost all high and medium expressed genes have a power over 90% while the low expressed genes show a more uniform distribution. Therefore, increases in group size mainly improves the power of low expressed genes with little to no effect on medium to high expressed genes. In the same way, increasing group size impacts mostly medium to high expression levels for the contrast of Treat#1 vs Treat#2, with lesser impact on the low expressed genes. With a group size of six the power distribution is highly similar for high expressed genes between both treated cell line contrasts, although the overall power difference between both contrasts is almost 10% (**Figure 3-2A**).

The impact of expression level was similar in the TCGA dataset. For a group size of 18, the power is low with most genes having a power below 10% for low or medium expressed genes and a uniform distribution for high expressed genes. However, even for this small group size the power is over 80% for over 7% of genes, mostly in the high expressed level. This is the case despite the overall estimated power is below 30%. The category of high expressed genes receives a strong power boost by a group size increase

to 30 with almost two thirds of the genes having a power of 80%. A similar effect but to a lesser degree is the increase in power for genes in medium expression group. The low expressed gene distribution becomes uniform with less than 1000 genes having a power above 80%. In summary, the power of individual genes is highly dependent on the expression level, with a higher expression level leading to a higher power for given effect size. For both the TCGA and trastuzumab datasets, an increase of group size mainly improves the power of low expressed genes.

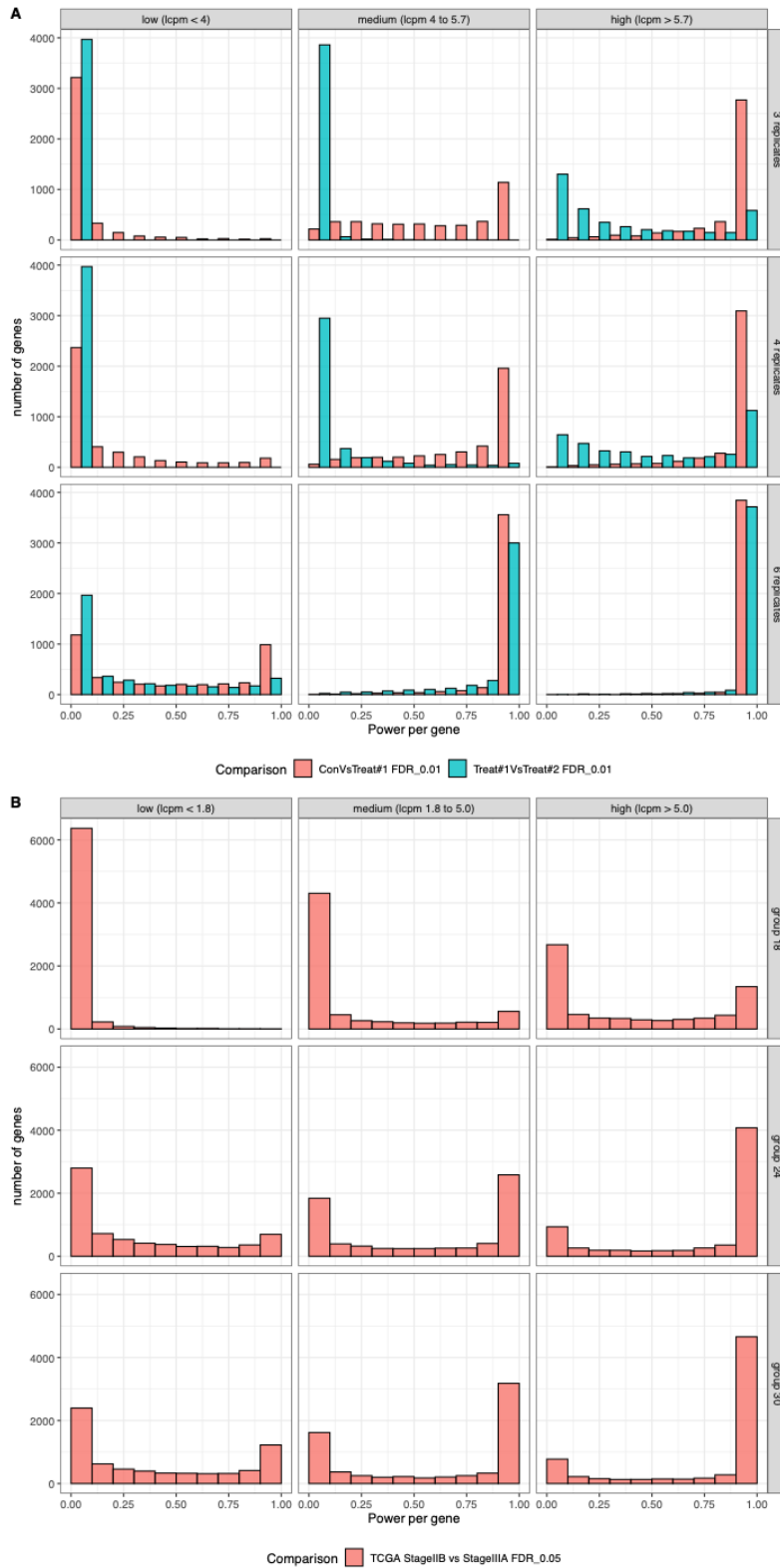


Figure 3-3 Per gene power is highly dependent on expression level. The histograms show the power distribution of genes for both datasets for a subset of parameters. The genes were divided into three groups of equal size based on the expression level (lcpm) with the expression ranges noted at the top. The number of genes per group is 3970 (treated cell line) and 6805 (TCGA dataset). A: histograms of power distribution for treated cell line dataset for an effect size of 1.5, FDR threshold of 1% and group size numbers of three, four, and six. B: histograms of power distribution for TCGA dataset for an effect size of 2, FDR threshold of 5% and group size of 18, 24 and 30.

3.3.5 Power Estimation on Gene Level

The power for single genes differed greatly from the overall power estimations and was further analyzed. The power curves of single genes were selected for different lcpm expression values (**Figure 3-4**). For the treated cell line (**Figure 3-4A, B**), 16 genes were selected for a range from 0.1 to 14.5 lcpm with steps of 1 lcpm. The plots were split by effect sizes of 1.25, 1.5 and 2. The comparison Con vs Treat#1 exhibits the same trend as observed in **Figure 3-3**. The estimated power for high expressed genes reached 80% at all group sizes for effect sizes of 1.5 and 2 and at group size bigger than 6 for an effect size as small as 1.25. The medium expressed genes followed closely. Only the power of low expressed genes showed a low power throughout all group sizes for an effect size of 1.25. The power crossed a level of 80% for a group size of 10 for an effect size of 1.5 or 2 except for one gene. The lowest expressed gene NPIPA3 did reach a power over 10% or any parameter combination.

The trend of the power curves is similar for the contrast of Treat#2 vs Treat#1, but the rise of power is shifted towards bigger group sizes. The strongest surge in power appears for the group size growth from four to six. After that, the power of some genes decreases when the group size is increased to eight. This observation is weaker for bigger effect sizes. That is comparable to the observation from the overall power curve for the contrast of Treat#2 vs Treat#1. Like for Con vs Treat#1, the two lowest expressed genes did not reach a power of 80% for any combination of parameters. In summary, the single gene power for both treated cell line contrasts have the same trend as the overall power curves in **Figure 3-2**.

The results differ for the TCGA dataset. The course of the power curves showed a similar tendency for an increased group size and effect size. However, the power of high, medium expressed genes did not exceed the power of the low expressed genes for 1.25 effect sizes as much as for the treated cell line. Furthermore, the power occasionally showed the opposite effect for a growing group size, for example, for an effect size of 1.5 in the group size range from 66 to 88 for the high expressed gene ENSG00000120868. Therefore, the effect of changes in group size and effect size for all eight selected genes are more similar compared to the treated cell line results. The increase of group size kept benefiting the power of high and medium expressed genes too and not only for the low expressed genes as for the treated cell line data. For the effect size of 2, the curve was more like the power curve of treated cell line contrast for

four of the five highest expressed genes (ENSG00000133112, ENSG00000083845, ENSG00000134954, ENSG00000120868). The power of these four genes achieved 80% power earlier than the other genes or the overall power curve (**Figure 3-2**) for a group size of 30. In summary, the single gene power estimation for the TCGA dataset were not as stable as the treated cell line data set with the higher between group variances.

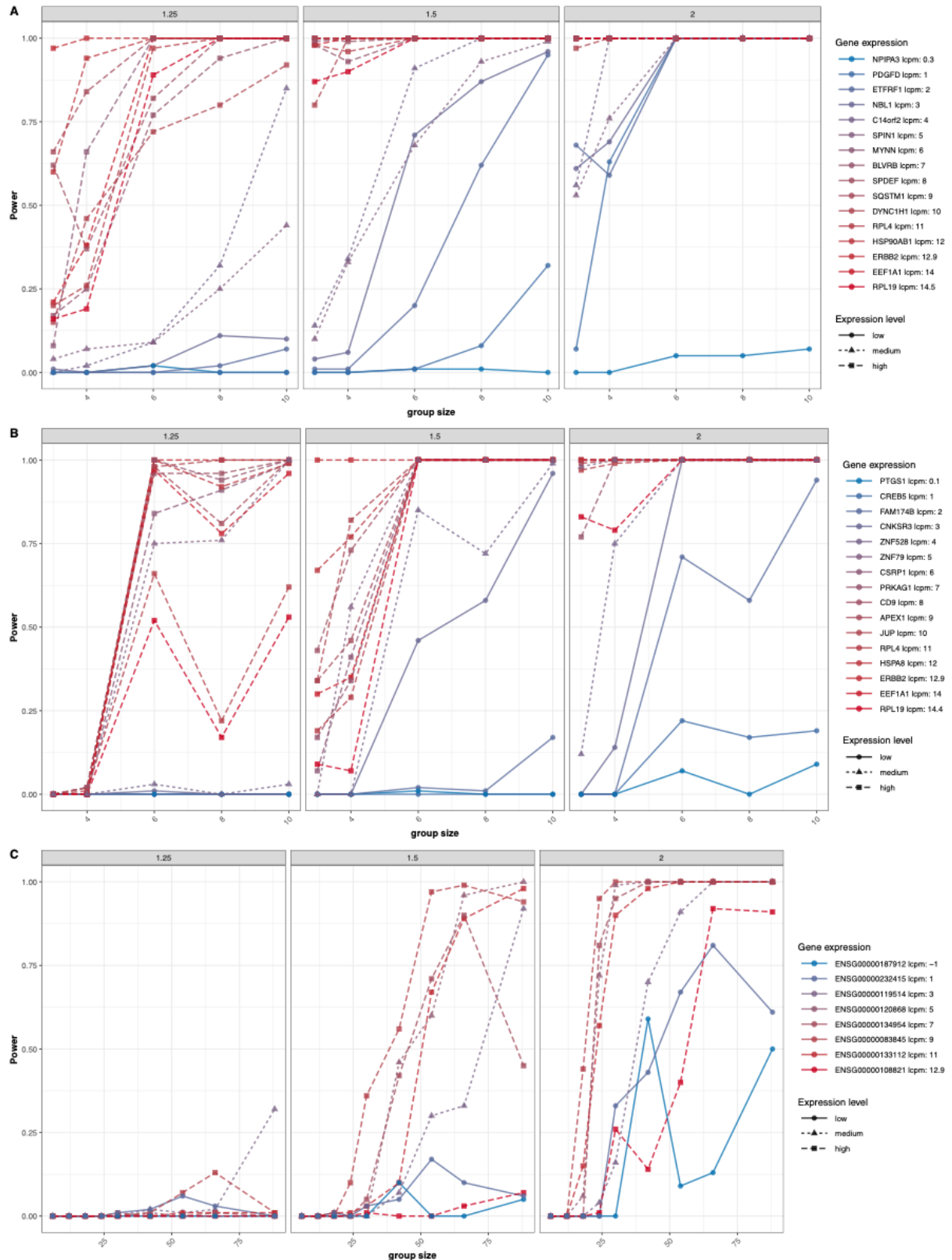


Figure 3-4 Power curves for representative genes for each expression group and comparison. The data are split for effect sizes of 1.25, 1.5 and 2, and the FDR threshold is 5% for all comparisons. The curves are split for the three comparisons Con vs Treat#1 (A), Treat#1 vs Treat#2 (B) both of which are for the treated cell line and StageIIB vs. StageIIIA (C) for the TCGA dataset. The genes were chosen based on lcpm expression values as labelled on the right side of the plots. The expression level groups were assigned as in Figure 3 and were labelled by different line types and symbols as shown on the right side of the plots.

3.4 Discussion

RNA-seq data are highly complex, typically involving over ten thousand genes and hypotheses. Expression data are impacted by many factors, such as sample preparation and the sequencing process. Therefore, a common assumption is that RNA-seq experiments are frequently underpowered when used in exploratory studies. This can lead to the conclusion that if no significant changes were detected in an experiment the sample sizes must have been too small, without further analysis of type II errors. All currently available tools to analyze the power of experiments are for a priori power estimation to predict sample sizes. For that purpose, the programs are based on read count simulations that assume various characteristics of the gene count distribution. Most of them intentionally underestimate the power to provide additional safety for the experimental planning. Nevertheless, there are no tools to control the actual post-hoc power of an experiment, leaving questions open about what magnitude of effect can be detected within a given experiment. A method that reliably estimates the power for a target effect size of interest will help answer the question of whether differentially expressed genes have been missed or are not present.

In this study, we implemented a bootstrap approach with effect size adjustment, while maintaining the library size, gene expression levels, and variance. The usefulness of bootstrapping to analyze RNA-seq data has been previously shown for estimating confidence in expression levels for hypothesis testing for RNA-seq data in the tools IsoEM2 or kallisto (Bray et al., 2016; Mandric et al., 2017). Bootstrap sampling differs from the simulation methods that sample from a specific distribution used in other power analysis tools (Guo et al., 2014; H. Wu et al., 2015; Zhao et al., 2018). One key difference is that bootstrapping does not assume any distribution characteristics by simulating data from the observed values, avoiding bias. Furthermore, the added smoothing factor prevents the case of a variance of zero for a group and adds some variance dependent noise to the data. The step of effect size adjustment before sampling allows analysis of the power for detecting a significant difference in expression for a target effect size. Also, using the FDR threshold derived from the original dataset controls whether calculated p-values would pass the threshold in the original dataset if that gene was the only effect size-adjusted gene. The combination of the bootstrap sampling, effect size adjustment and use of the observed FDR threshold result in a useful post-hoc power analysis and is not a simple function of the observed p-values as noted

by Ruxton (Colegrave & Ruxton, 2003) and Hoenig and Heisey (Hoenig & Heisey, 2001).

Our approach reveals that the power for the two analysed datasets is reasonably high and that acceptable power levels (>80%) can be achieved for smaller group sizes per condition. For the treated cell line data set, both comparisons exceed 80% power for a group size of 3 replicates, an effect size of 2, and an FDR threshold of 5%. Similarly, for the TCGA dataset 80% power was obtained for a group size of 66, 24 replicates less than the total available number of replicates per condition. This is surprising because RNA-seq experiments are considered frequently underpowered, especially those involving patient-derived cancer tissue (Cui et al., 2018; Y. Liu et al., 2014; Schurch et al., 2016). Furthermore, the analysis shows the two comparisons for the treated cell line data set have a similar power for effect sizes of interest and show a power of close to 95% considering the complete dataset with a standard effect size of 2 and a stringent FDR threshold of 1%. This is reliably above the often-acceptable power level of 80%. The variance of the estimated overall power has also been shown to be quite small. This analysis has a major implication for the treated cell line data, as no differentially expressed genes were identified between both treatments (Treat#1 and Treat#2), but the comparison of both treatments against the control identified over 2000 genes as differentially expressed in the initial study. Thus, the findings support a high similarity of both treatments.

We also studied the power dependency of single genes on the expression levels. It has been previously shown that filtering samples by expression level, for example using a minimum of 10 counts in the smallest samples increases the power (H. Wu et al., 2015). Dispersion values of genes are dependent on the mean expression of genes, with lower expressed genes generally having a higher dispersion. Therefore, our results suggest that this dependency is much stronger for detecting differentially expressed genes (**Figure 3-3**). For example, the contrasts for treated cell line data show a power level of over 80% for most of medium- or highly expressed genes for strict parameters (effect size 1.5, 1% FDR, group size of 6). In contrast, most low expressed genes show a power level of 10%. The same trend can be seen for the more variable TCGA gene set. Therefore, this dependency should be considered for the interpretation of RNA-seq data. For example, key genes of a pathway of interest are unlikely to be detected as differentially expressed if their expression level is too low, particularly for more variable datasets.

Additionally, we discovered that the power estimation per gene is stable for both datasets. Indeed, most genes pass consistently either pass or fail the test across all resampled data sets. That means the overall power stems mainly from specific genes. For treated cell line contrast of Treat#1 vs. Treat#2, 30% of genes pass the FDR threshold over 80% of the time when the overall power was below 50% at an FDR of 5%, effect size of 1.25 and group size of 4. Also, 30% of genes had an estimated power of less than 10% for an FDR of 5%, effect size of 1.25 and group size of 4. This suggests that the power estimation is highly stable even on a single gene level. These results challenge the assumption that RNA-seq studies are generally underpowered. The single gene curves are similar to the overall power curves for the treated cell line data. The single gene power curves also showed that especially low expressed genes are more likely to have low power and not be detectable as differentially expressed for any combination of parameters. Therefore, for these genes a future experiment would need a higher number of replicates. These results can influence the analysis of cases where specific genes are of interest, for example with cancer or disease biomarkers. A specific case for which this can be helpful is the use of TCGA data sets to confirm results of smaller studies. With our approach it is possible to control when genes of interest are not significantly changed results from either a low statistical power or small effect size.

Power estimation at the single-gene level appears less stable for the more variable TCGA dataset. To improve that, a weighted bootstrapping approach could further stabilize estimates by reducing the influence of outliers in large-sample datasets such as the TCGA dataset. Another addition could be the identification of genes which are commonly underpowered or show a low effect size, through a meta-analysis of multiple datasets. The estimation of group size for future RNA-seq experiments could also benefit from this approach as the number of publicly available RNA-seq datasets grows.

In conclusion, the bootstrap approach estimates overall and single gene power, and post-hoc power analysis may support a more comprehensive analysis of RNA-seq studies by adding information beyond the p-values of the original data set. Future research like RNA-seq experiments for drug comparison may benefit from this additional step to confirm statistical power and identify the potential type II error rates of experiments.

Chapter 4: Molecular Characterization of HER2-Positive Breast Cancer Cell Lines Following Trastuzumab and Biosimilar Treatment

4.1 Introduction

HER2-positive breast cancer accounts for 15–20% of all cases. It is characterized by overexpression of the transmembrane tyrosine kinase HER2, which promotes cancer progression by enhancing cell proliferation and survival signalling pathways (Slamon et al., 2001). This is a subtype that is associated with rapid growth and higher likelihood of metastasis and poorer survival rates than other subtypes (Arteaga et al., 2012). HER2-targeted therapies that inhibit changes in HER2 signalling have greatly increased the efficacy of treatment. The standard therapy is the therapeutic monoclonal antibody against HER2, trastuzumab (Herceptin), which binds to HER2 and blocks downstream signalling within the HER2 pathway (Hudis, 2007; Vu & Claret, 2012). Inhibited downstream signalling leads to inactivation of the hyperactive PI3K/AKT and MAPK pathways, which are otherwise sustained by feedback mechanisms (Yarden & Sliwkowski, 2001). Ultimately, HER2 inhibition enhances apoptosis and cytostasis, reduces cell survival, and halts cancer progression (Baselga & Albanell, 2001). However, resistance to trastuzumab evolves in a major fraction of patients treated with this drug, thereby reducing the choice of treatments and blunting long-term survival rates (Beckwith & Yee, 2014).

Several mechanisms of resistance have been proposed, most of which involve persistent downstream signalling activation despite HER2 inhibition. Those molecular mechanisms can be described depending on mutations in the HER2 receptor (Sun et al., 2015), activation of alternative growth factor pathways, e.g. IGF-1R (Beckwith & Yee, 2014), MET (Rivas et al., 2022; Shattuck et al., 2008), and increased signalling through parallel pathways like PI3K/AKT (Saal et al., 2005), loss of phosphatase and tensin homolog (PTEN) (Pandolfi, 2004; Saal, Holm, Maurer, Memeo, Su, Wang, Yu, Malmström, et al., 2005), or upregulation of EGFR (Gallardo et al., 2012), MUC4 (Mukhopadhyay et al., 2011; Mukohara, 2011) or HER3 (Dey et al., 2015; B. Liu et al., 2011; Pohlmann et al., 2009). However, while many mechanisms of resistance have been investigated and proposed, a comprehensive list of the mechanisms of trastuzumab resistance under long term treatment ultimately remains vague (Baselga & Albanell, 2001;

Beckwith & Yee, 2014; Carey et al., 2016; Perez et al., 2014; Pohlmann et al., 2009; Slamon et al., 2001; Vogel et al., 2002).

The expiration of the U.S. patent for Herceptin in 2019 has opened the door for the development of copies, so-called biosimilars, which will expand access to this critical drug. These findings highlight the importance of understanding trastuzumab's mechanisms of action and resistance to make meaningful comparisons among biosimilars and the reference product Herceptin (Tkaczuk & Jacobs, 2014). Biosimilars are biological products highly similar to an approved reference product, with no clinically meaningful differences in terms of safety, purity, and potency (Center for Drug Evaluation and Research, 2015). Nevertheless, the demonstration of biosimilarity for structurally complex molecules as therapeutic antibodies and with complex effects on target and immune cells, is extraordinarily challenging (European Medicines Agency, 2012; Schellekens, 2009). Although many biosimilars receive regulatory approval, over 20 have been withdrawn due to insufficient biosimilarity or delayed authorization. In the U.S. only six biosimilars have been accredited (European Medicines Agency, 2017; Rathore et al., 2020; U.S. Food and Drug Administration, 2017). This complex impact and incomplete understanding of mode of action of many biosimilars, such as trastuzumab, precludes detection of eventual differences between biosimilars and the originator product.

Currently, rigorous testing involves physicochemical characterization, biological activity assessment, and pharmacodynamic studies which are required by regulators for biosimilar approval (European Medicines Agency, 2012; FDA, 2015). However, these approaches often lack untargeted, global analyses of complex signalling responses in target cells, which could reveal subtle differences in efficacy or unexpected side effects (Schellekens, 2009).

This void can be complemented by RNA sequencing (RNA-seq), a powerful technique that allows for non-targeted global analysis of the transcriptome to generate hypothesis and study systems that can be followed up on. RNA-seq offers a complete picture of the cell's transcriptome and is capable of identifying small changes to the cell in response to perturbations (Z. Wang et al., 2009). This becomes especially important for complex biologics such as trastuzumab, where the originator product's batches have variations in glycosylation, oxidation, deamidation, amino acid sequence, and even at the three-dimensional level, the detailed protein model. Even these minor variations could influence the binding affinities and hence the downstream signalling and gene expression patterns (Al-Qasem et al., 2021; Conesa et al., 2016; Y. Jiang, Li, et al., 2021). However, currently RNA-

seq has not been reported in the scientific literature as a key component for demonstrating biosimilarity in regulatory submissions.

For trastuzumab, only limited RNA-seq data are available for treated breast cancer cell lines with sufficient sample numbers. This scarcity of available data to fully understand the complete transcriptomic changes caused by trastuzumab on target cells strongly limits our understanding.

This study treats two well-established HER2-positive breast cancer cell lines, BT474 and SKBR3 with trastuzumab and its biosimilar, Apotex-Trastuzumab, provided by ApoTex (Canada) to address this gap and assess the biosimilarity via RNA-seq. These cell lines are characterized by different HER2 levels, and they are also known to be differently responsive to trastuzumab, thus representing distinct biological models of HER2 positive breast cancer. While BT474 cells are estrogen receptor (ER) and progesterone receptor (PR) positive, SKBR3 cells are ER/PR negative, representing a more aggressive phenotype (Holliday & Speirs, 2011; Neve et al., 2006).

High-throughput RNA-seq with sufficient replicate numbers will generate detailed gene expression profiles of these cell lines following trastuzumab treatment. The examination will discuss transitions in intrinsic subtype as well as gains and losses in cancer hallmark pathway activation, regulon network formation, and gene set enrichment analysis. These analyses will permit the study of trastuzumab influence on target cells, the evaluation of biosimilarity, and subtype- specific variations in the cell lines response.

This holistic approach is crucial for future biosimilarity comparisons. This work provides a proof of concept for the use of RNA-seq to identify small disparities between biosimilars and their reference products and further contributes to the repertoire of methods available for biosimilarity analysis. Perhaps transcriptomic analysis should become part of the approval process for biosimilars by regulatory agencies. This could lower the barrier for biosimilars to go through the lengthy clinical trials and ultimately to the market.

4.2 Materials and Methods

4.2.1 Cell Line Culture and Treatment

The human breast cancer cell lines BT474 (HTB-20; ATCC, American Type Culture Collection, Manassas, USA) and SKBR3 (HTB-30; ATCC, American Type Culture Collection, Manassas, USA) were used for the experiments. Methods for culturing BT474 are presented in Chapter 2. SKBR3 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Thermo Fisher Scientific, San Jose, USA) supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco, Mississauga, Canada) and 1% (v/v) penicillin/streptomycin [10,000 U/ml penicillin, 10,000 µg/ml streptomycin] (Gibco, Mississauga, Canada). Cultures were maintained at 37°C in a humidified atmosphere with 5% CO₂. For all experiments, cells were used at a passage number of approximately 20 to ensure consistency across replicates.

4.2.1.1 Mycoplasma Contamination testing

Mycoplasma contamination of the cell line was tested using the PCR Mycoplasma Test Kit (PromoKine, Heidelberg, Germany) following manufacturer's instructions.

4.2.1.2 Cell line treatment

The SKBR3 cell line was used for the experiment; cells were plated in 6 well plates with a seeding density of 6500 cells/cm². At Day 0, in both the control and treatment groups, eight replicates were set up to ensure uniformity. Medium was refreshed on Day one and four. The media was changed, and cells were treated with 4 µg/mL of Herceptin (Genentech, San Francisco, CA) or ApoTras (Apobiologix, Toronto, CA)(Ichikawa et al., 2012) in 1X PBS. For the control group, the same amount of 1× PBS was added to the media as in the treatment groups. Here the effect of Herceptin and ApoTras on SKBR3 cell proliferation and viability is examined. Whole RNA extraction was performed on Day 7.

4.2.2 RNA Extraction and Sequencing

RNA was isolated from samples with Quick-RNA™ MiniPrep kit (Zymo Research, Irvine, USA) following manufacturer's instructions. RNA quality and quantity were determined with a Qubit fluorometer (Thermo Fisher Scientific) and samples below 100 ng/µl were excluded from downstream analyses. RNA quality was monitored using the

Agilent 2100 Bioanalyzer and RIN values higher than nine were considered for the analysis. This ensured high-quality RNA for sequencing and analysis.

RNA libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, USA). In the case of libraries sequenced on the Illumina HiSeq 2500, 2 x 126 bp paired-end sequencing was performed and an average of 15 million reads were generated per sample. RNA quality was assessed and the libraries were generated and sequenced at the Centre for Applied Genomics, Sick Kids Hospital Toronto, Canada.

4.2.3 Bioinformatic Analysis

Bioinformatic analyses were performed on SKBR3 and BT474 cell lines after treatment and sequencing.

4.2.3.1 Data availability

Molecular subtyping controls were obtained from the breast cancer data set (BRCA) of The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013) were accessed via the R package `curateTCGADData` v1.1.38. The data, denoted as “RNASeq2GeneNorm”, provided normalized log2 CPM values. The BT474 sequencing data and SKBR3 sequencing data are available in the NCBI Gene Expression Omnibus (GEO) under accession numbers GSE158969 and GSE284388, respectively.

4.2.3.2 Sequencing quality control and adapter trimming

Initial quality control was performed using FastQC v0.11.6 (Andrews, 2010) before and after adapter removal and trimming of low-quality bases with BBDuk method of BBTools suite (Bushnell, 2016). The 3' ends of the reads were trimmed using a k-mer size of 23. A summary of the FastQC reports was obtained with MultiQC v1.25.2 (Ewels et al., 2016).

4.2.3.3 Alignment

Reads were mapped against the human reference genome GRCh38.p12 using the ENSEMBL v95 gene annotation as a reference database. Alignment was performed using STAR aligner v2.5.0 (Dobin et al., 2013) with parameters recommended by G. Baruzzo et al. following a benchmarking study (Baruzzo et al., 2017). Metrics reported by the STAR aligner were summarized using MultiQC v1.25.2 (Ewels et al., 2016).

4.2.3.4 Quantification

Gene expression levels were quantified post-alignment using RSEM v1.3.0 with default parameters (B. Li & Dewey, 2011). Subsequent analyses were performed on both raw count data and normalized log2 counts per million (CPM) depending on the analytical requirements.

4.2.4 Statistical Analysis

All statistical analyses were performed with R v4.3.1 and RStudio (RStudio Team, 2016).

4.2.4.1 Testing for differentially expressed genes (DEG)

Low expressing genes were excluded from further analyses. Genes that had less than 1 CPM in more than 8 SKBR3 or 9 BT474 samples were filtered out. DEGs were detected using edgeR v3.18.1 (Robinson et al., 2009) with raw counts. The first step is normalization to account for differences in library sizes, variations in the composition of RNA and technical biases, which is achieved by the `calcNormFactors()` function and uses the Trimmed Mean of M-values (TMM) normalization. The `estimateDisp()` function was used for estimating the dispersion of biological variation. This step is essential because its output serves as input for the `glmQLFit()` function for fitting the generalized linear model with quasi-likelihood. After fitting the models, the significance of expression changes was tested on the fitted data with `glmTreat()`. Differences were calculated and limits to detect significant changes using FDR and LogFC for all of them are described in the results section.

4.2.4.2 Principal component analysis

The principal component analysis (PCA) was conducted with the `prcomp` function of the `stats` package v.3.6.2. LogCPM values from the expression data of both cell lines were used for the analysis.

4.2.4.3 Cell cycle regulated genes analysis

The Cyclebase 3.0 database was analyzed to identify genes associated with cell cycle expression. It is a database of cell cycle regulated genes. Specific genes that are known to change their expression levels during the cell cycle's phases G1, S, G2, and M have been investigated. The top 500 genes linked to a particular cell cycle phase were binned according to the cell cycle peak time for the cell cycle. Clusters were formed on the basis

of their time of peak expression in the cell cycle and associated frequencies of those genes across the clusters were used to define phase-specific patterns.

4.2.4.4 Cancer subtype assignment

The intrinsic cancer subtypes of the samples, such as Luminal A, Luminal B, Basal-like, HER2-enriched, and Normal-like were assigned based on the PAM50 gene signature (Bastien et al., 2012; Parker et al., 2009). PAM50 subtyping was conducted using the Genefu v2.32.0 package (Gendoo et al., 2016). The package includes centroid expression profiles for each subtype, based on expert-classified samples. The classification of samples was based on the highest Spearman correlation with centroids.

4.2.5 Pathway Activity Calculation

4.2.5.1 Pathway Activity score calculation

To calculate the activity score, first the DEG analysis results were transformed by calculating a combined score from LFC and FDR values.

$$\text{Combined Score} = \text{LFC} \times -\log_{10}(\text{FDR})$$

To obtain pathway activity scores, the consensus method was applied using the decoupleR package (Badia-I-Mompel et al., 2022). Activity scores were calculated for the Pathway Responsive Genes for activity inference (PROGENy) (Holland et al., 2020; Schubert et al., 2018) with a total of 14 sets and the Collection of Transcription Regulation Interactions (CollecTRI) (Müller-Dott et al., 2023) with a total of 564 transcription factors with at least 10 target genes.

4.2.5.2 Heatmap

Hierarchical clustering was used to enable visualization of clusters identified in our analysis. Data were plotted with the pheatmap function, of the pheatmap package v1.0.12 (Kolde, 2022), by the normalized activity scores. Euclidean distance was used as a metric to measure dissimilarity and complete linkage clustering was used to group the samples based on their scaled activity scores.

4.2.5.3 Network creation

To elucidate the commonalities between top-ranked transcription factors between the transcriptional datasets of CollecTRI, the overlapping genes detected between significantly altered transcription factor target gene sets were obtained. Pearson's

correlations were used to compute similarity. The network was plotted as an undirected graph using the igraph package with the Kamada-Kawai layout algorithm (Kamada & Kawai, 1989).

4.2.6 Gene Set Enrichment Analysis

To identify pathways impacted in their expression by the treatments, gene set enrichment analysis (GSEA) was performed with the package clusterProfiler v4.10.0 (T. Wu et al., 2021). Four distinct comparisons were created per cell line:

1. Herceptin against the control group (HervsCon)
2. ApoTras against the control group (ApoTrasvsCon)
3. Both treatments together against the control group (ApoTras&HervsCon)
4. Treatment groups against each other (ApoTrasvsHer)

Gene set enrichment analyses for Reactome pathways were performed using the gsePathway function (Fabregat et al., 2016). Gene sets were confined to those containing between 10 and 500 genes. The pathway enrichment significance was determined using the false discovery rate (FDR) threshold of 5%, correcting for multiple comparisons using with the Benjamini-Hochberg method.

4.2.6.1 Grouping Gene set enrichment results

4.2.6.1.1 Overlap estimation

To estimate the overlap between significantly enriched pathways, first the ratio of shared leading-edge genes between each pair of significant pathways was calculated. Specifically, the overlap ratio was calculated as the number of shared DEGs divided by the total number of leading-edge genes in the smaller of the two pathways. This approach prevents smaller pathways from being penalized for having fewer leading-edge genes. Additionally, the Jaccard index was calculated. Both scores were combined. The formula for calculating the overlap ratio O for two pathways, A and B, is:

Where:

$$O = \frac{|A \cap B|}{\min(|A|, |B|)}$$

$|A \cap B|$ is the number of shared DEGs between pathways

$\min(|A|, |B|)$ is the total number of DEGs in the smaller of the two pathways.

The overlap ratio was then combined with the Jaccard index:

$$J = \frac{|A \cap B|}{|A \cup B|}$$

$|A \cup B|$ is the total number of unique DEGs in both pathways.

Each measure, the overlap ratio and the Jaccard index, was weighted at 0.5 to create a final combined score S for the similarity between each pair of pathways:

$$S = 0.5 \times O + 0.5 \times J$$

This combined score reflects the overall similarity of the pathways.

4.2.6.1.2 Hierarchical clustering

Next, hierarchical clustering was performed on the pathways using R base functions. The Euclidean distance was used to compute the dissimilarity between pathways, while complete linkage was applied for agglomeration, ensuring that closely related pathways are grouped together. This clustering method resulted in the identification of 20 pathway groups. These groups represent clusters of Reactome pathways that share common leading-edge genes, indicating similar biological processes. Clustering results were visualized using a Sankey plot, highlighting similarities and differences between the two cell lines. Each group represents a cluster of pathways with similar DEGs, revealing shared biological mechanisms. The pathway groups were then analyzed to identify key biological processes.

4.2.7 Visualizations

Visualizations, unless otherwise noted were created using ggplot2 v3.4.2 (Ginestet, 2011) and when needed modified with Inkscape v1.2 (The Inkscape Team, 2024).

4.2.8 Additional Files and Code Availability

Additional files and scripts for preprocessing, analysis, and figure creation are available in the GitHub repository https://github.com/karstenrinas/Thesis/Chapter_4.

4.3 Results

The data set used in this study focused on HER2-positive subtype of breast cancer for which the BT474 and SKBR3 cell lines serve as models. The gene expression profiling was conducted by high-throughput RNA sequencing in cells treated with Herceptin (Her) and the Apotex-Trastuzumab (ApoTras) biosimilar. It allows concluding on the primary modifications of the signalling pathways with targeted therapy. More detailed quality control metrics can be found in the MultiQC reports of FastQC results in [Additional file 1 to 2](#).

SKBR3 and BT474 cells were treated with Her or ApoTras at 8 replicates each condition for SKBR3 and at 9 or 10 replicates of each condition for BT474. Alignment rates per sample exceeded 98% for BT474 and of 10-20 million reads, and 9-12 million reads per sample for SKBR3. Summary statistics for the alignment are also reported in the MultiQC reports found in [Additional file 3 to 4](#). The expression data tables for both cell lines are provided in [Additional file 5 Table S1-2](#).

4.3.1 Validation of HER2-Positive Breast Cancer Cell Models

HER2 expression levels were similar in both cell lines, exceeding a logTPM of 10.5 (X. Dai et al., 2017). These cells were characterized as Luminal B type with logTPM values of 6.6 and 3.8 for progesterone receptor (PR) and estrogen receptor (ER), respectively, meaning they were progesterone receptor and estrogen receptor positive. Conversely, SKBR3 displayed logTPM levels for PR and ER of 0 and 0.5, respectively, being classified as belonging to the more aggressive HER2-enriched subtype of breast cancer.

Only BT474 expressed CDH1 (logTPM 7.6; undetectable in SKBR3). Both cell lines expressed KRT8, KRT18, and EPCAM, but not the mesenchymal markers VIM, SNAI1, or ZEB1 and thus classified as epithelial cells (Y. Wang & Zhou, 2016). These data show that in terms of the studies regarding breast cancer both models have an epithelial phenotype. The fact that SKBR3 cells do not express CDH1 also suggests that there are minor differences in the epithelial phenotype of the two cell lines.

4.3.2 Molecular Treatment Efficacy is Highly Similar

Principal component analysis (PCA) of the 500 most variable genes showed clear separation between treated and control samples (**Figure 4-1**). The principal source of variation was due to the treatment, while PC2 represented variation inherent to either cell

line. For BT474 and SKBR3 their samples treated with Her and ApoTras clustered nearest to each other, indicating similar molecular response to treatments. Cultured SKBR3 samples were more closely grouped, relative to BT474.

To further examine downstream effects of the treatments, cell cycle gene expression data from Cyclebase 3.0 were analyzed (Santos et al., 2014) and rendered in (Figure 4-1). BT474 and SKBR3 exhibited a significant reduction in the expression of the same relevant genes in all cell cycle phases, relative to the untreated controls. Of note, BT474 cells, which rely to a larger extent on HER2 signalling, were affected to a greater extent, as expected. The strongest reductions in expression were observed for genes involved in G2/M and M phases, such as CCNB1, CDK1, BUB1, further supporting the efficiency of HER2 inhibition in blocking mitosis progression. In contrast, the downregulation in SKBR3 was less evident, indicating a poorer response to the treatments.

Only small differences in the modulation of cell cycle by Herceptin and ApoTras were identified. This is a testament to the extremely high molecular resemblance between trastuzumab and the biosimilar as the origin of variation was mostly among cell lines rather than treatments. These are clear examples of how RNA-seq can detect subtle differences and demonstrates biosimilarity.

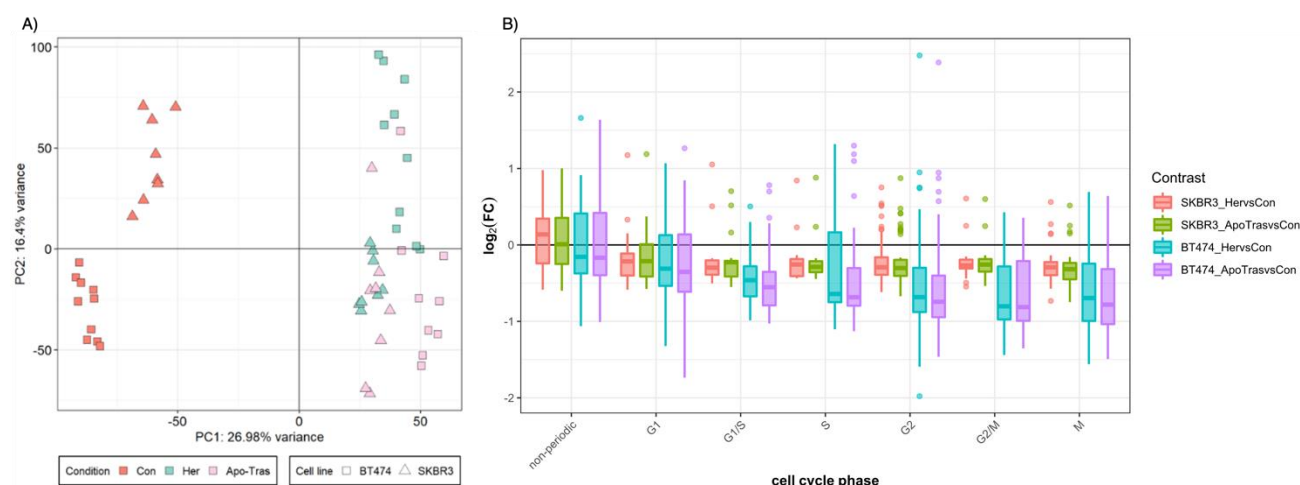


Figure 4-1 General data structure for comparing the originator Herceptin with the biosimilar treated samples versus control. A) PCA plot of the 500 most variable genes after normalizing data of all samples. Con: Control; Her: Herceptin as originator product; ApoTras. B) Boxplot of LogFC values for cell cycle phase related genes (Cyclebase 3.0; Santos et al., 2014). The comparisons show a clear downregulation for all comparisons. HerVsCon: Herceptin vs. Control; ApoTrasVsCon: ApoTras vs. Control.

4.3.3 PAM50 Gene Expression and Intrinsic Subtype Shifts After Treatment

PAM50 gene expression profiles were examined in order to evaluate the effects of Her and ApoTras on breast cancer subtypes. These profiles provide information on

treatment-induced molecular changes by categorizing breast cancer into intrinsic subtypes - Luminal A, Luminal B, HER2-enriched, Basal-like, and Normal-like - (Parker et al., 2009). Both BT474 and SKBR3 samples demonstrated a significant shift from the Basal subtype in untreated controls to the Luminal A subtype following treatment with either Her or ApoTras. In line with clinical findings of HER2-targeted treatments encouraging decreased tumour aggressiveness, this return to a less aggressive molecular phenotype is evident (**Figure 4-2**). The detailed results of the PAM50 signature scoring are provided in [Additional file 6 Table S3](#).

Cell cycle-associated genes, such as CCNE1, CDC20, and MKI67, which are essential for cell proliferation, were significantly downregulated in treated samples. This downregulation was consistent across both treatments and cell lines, reflecting the effectiveness of HER2 inhibition in halting tumour growth. In contrast, upregulated genes, such as ESR1 and GRB7, indicated compensatory activation of hormone receptor signalling and remaining HER2 pathway activity, which are related to Her resistance (**Figure 4-2**).

Significant differences were observed between the two cell lines. In BT474, the treatment-induced Luminal A subtype shift coincided with increased expression of luminal markers such as PGR and MAPT. In contrast, SKBR3 displayed increased expression of keratinization-related genes like KRT17 and KRT5, suggesting epithelial characteristics associated with a more aggressive phenotype.

The PAM50 analysis shows the efficacy of Her and ApoTras in inducing advantageous molecular changes in HER2-positive breast cancer subtypes. The comparable responses between the two treatments support the biosimilarity of ApoTras to Her, while cell line-specific differences highlight the importance of customising therapeutic approaches to individual tumour profiles.

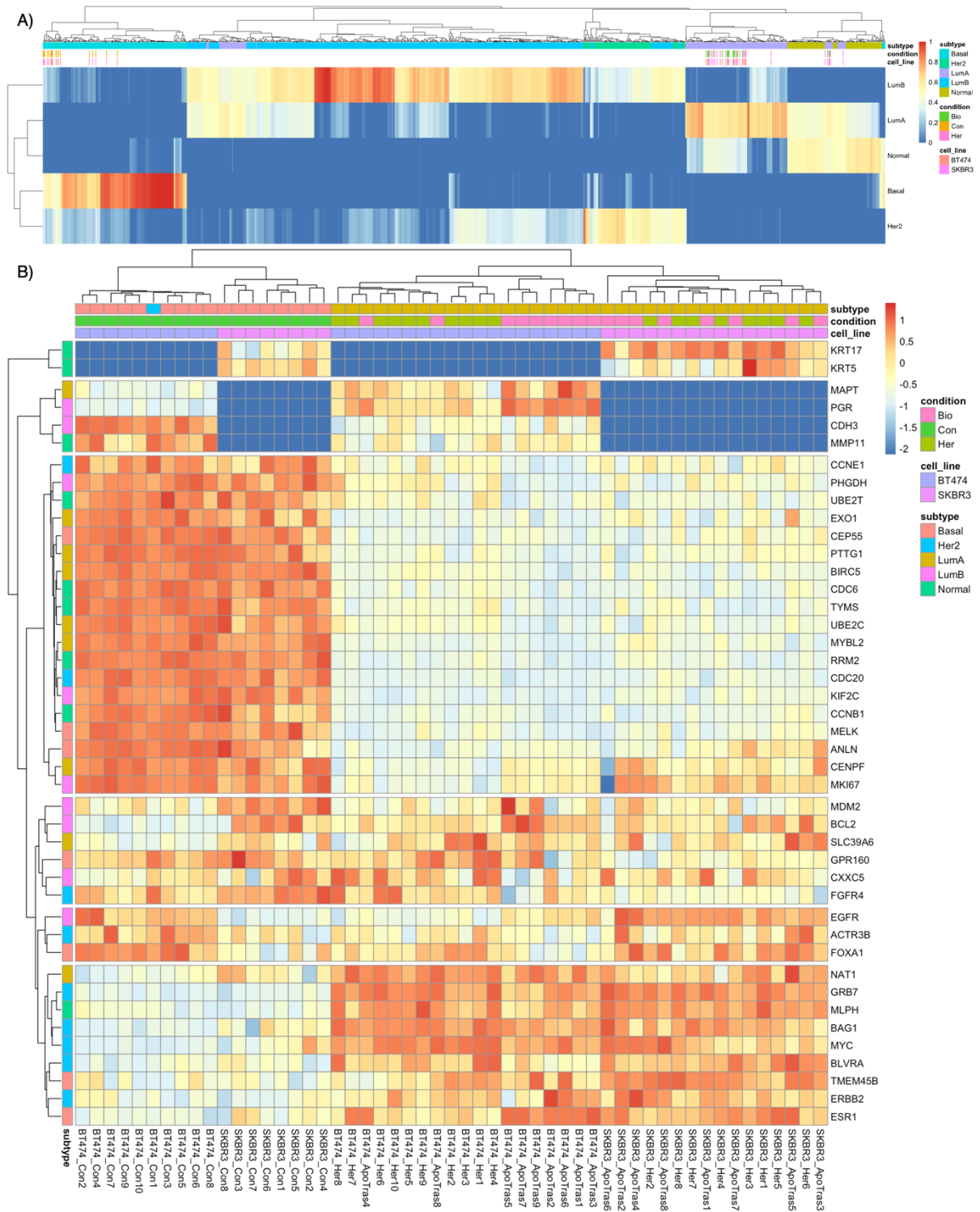


Figure 4-2 Intrinsic subtype shift caused by trastuzumab treatment. A) Heatmap of clustered PAM50 signature scores for the samples and TCGA breast cancer samples. B) Normalized expression levels for PAM50 subtype genes of all samples.

4.3.4 Untargeted DEG Analysis

To evaluate the difference between trastuzumab and its biosimilar ApoTras effects, two comparisons for each cell line, the combined treatments versus controls (ApoTras&Herceptin vs. Control; ApoTras&HervsCon) and treatments against each other (ApoTras vs. Herceptin; ApoTrasvsHer), were performed. This twofold methodology permitted the ability to find common gene expression signatures associated to HER2 blockade and discrete variations between the treatments. The tables of the differentially expressed genes for all comparisons per cell line can be accessed here [Additional file 7 Table S4-11](#).

BT474 also had a large transcriptomic response in the ApoTras & Herceptin vs. Control contrast: 1,659 genes up-regulated and 1,584 down-regulated (FDR < 5%, LogFC > ±0.32) (**Figure 4-3**). A similar but less pronounced effect was observed in SKBR3 cells, with 481 up and 269 down-regulated genes. The majority of differentially expressed genes (DEGs) were cell line-specific, and a small percentage of 180 upregulated genes and 118 downregulated genes were identified as differentially expressed in both cell lines. These shared DEGs were enriched in Reactome (Fabregat et al., 2016) pathways related to cell cycle regulation, particularly mitotic and G2/M checkpoints. Genes such as BIRC5, CDC20, and CDK1 were strongly downregulated, reflecting Her's and ApoTras' shared ability to inhibit cell proliferation by disrupting key cell cycle regulators.

Functional biases were found for differences in Reactome pathways for the overrepresentation analysis of DEGs specific to cell types. In SKBR3, downregulated genes related to mitochondrial translation included MRPL13, MRPL47, and MRPL52. BT474, on the other hand, showed repression of DNA repair, predominantly of homologous recombination molecules (e.g. BRCA1, BRCA2, RAD51, POLD1, POLD2) and polymerase activity.

In summary, the results indicate marked cell line-dependent responses to HER2 targeted therapy with BT474 cell line presenting more extensive transcriptome changes associated with the compromised DNA repair process, and SKBR3 developing more specific metabolic adaptation. These results provide the comparison to identify general and unique effects of the HER2 inhibitory activity at a molecular level and the findings highlight unique biological mechanisms of HER2 inhibition in both models and the need

of studying complementary cell lines to cover the heterogeneity of HER2 positive breast cancer.

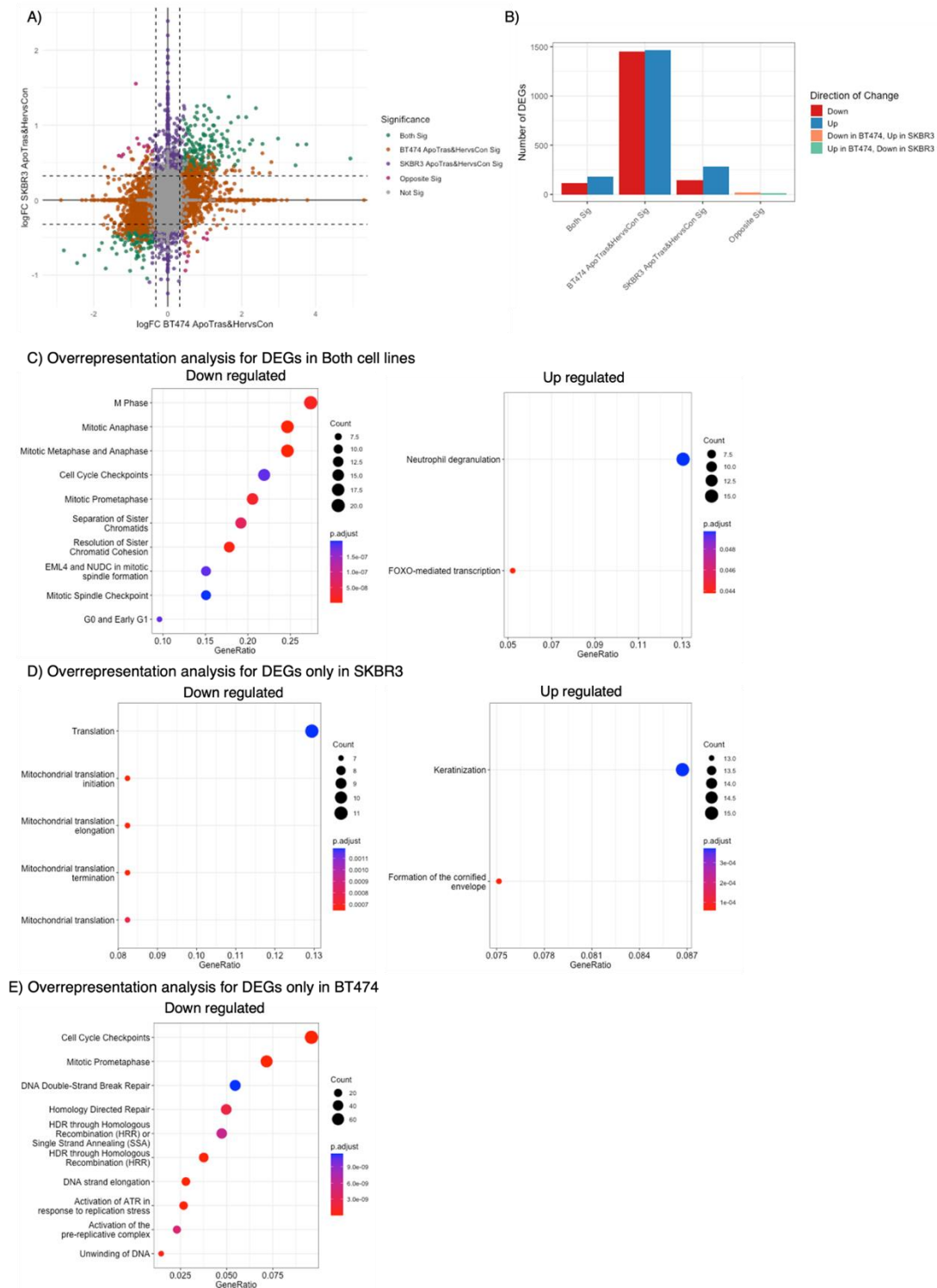


Figure 4-3 Comparative Analysis of Differential Gene Expression and Pathway Enrichment in BT474 and SKBR3 Cell Lines. A) Distribution of LogFC values for differentially expressed genes (DEGs) in BT474 and SKBR3 cell lines for the comparison. B) Barplot of overlapping and unique DEGs between BT474 and SKBR3. C-E) Pathway enrichment analysis highlighting shared and unique Reactome (Fabregat et al., 2016) pathways affected by Herceptin and biosimilar ApoTras treatments in BT474 and SKBR3.

Mean differences were very small for the ApoTrasvsHer comparison. In BT474, the expression of five genes was altered including the POTE family members POTEE, POTEF involved in cancer progression and ANKRD36C, ZFHX4 and an unnamed gene (ENSG00000269937) (**Figure 4-4**). These genes were related to protein interaction and tumour cell survival. In contrast, no significant DEGs were observed in SKBR3, reinforcing that at the molecular level the two treatments are very similar.

These data suggest that while Herceptin and ApoTras elicit similar, highly correlated, transcriptomic responses in HER2-positive breast cancer cell lines, the magnitude of the responses is different in BT474 versus SKBR3. While this gives more insight into the relative sensitivity to DEG in BT474, the changes in SKBR3 also indicate alterations in metabolic and survival response. These data confirmed that ApoTras can be considered biologically equivalent to Herceptin and highlighted the relevance of considering cell line-dependent responses in cellular modelling studies.

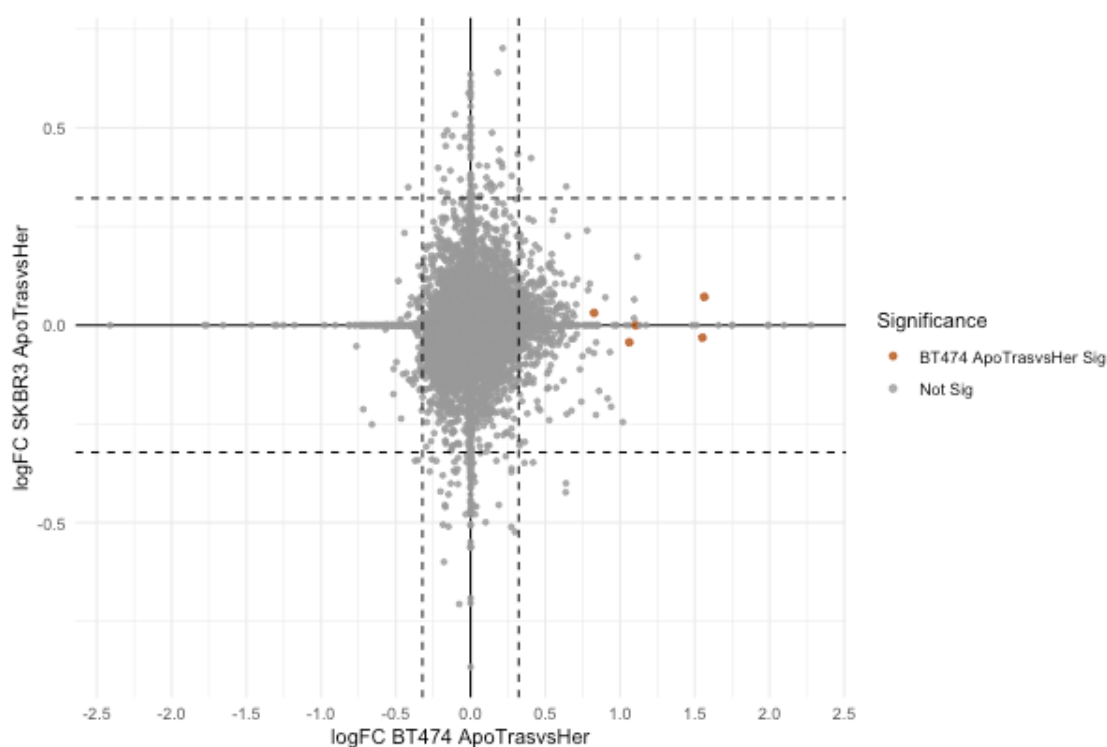


Figure 4-4 Scatter plot comparing log fold-changes (logFC) in gene expression between BT474 and SKBR3 cell lines for the comparison of ApoTras biosimilar against Herceptin (ApoTrasvsHer). Genes significantly differentially expressed (FDR < 0.05) are coloured in orange. Non-significant genes are shown in grey. Dashed lines represent thresholds for $\logFC \pm \log_2(1.25)$.

4.3.5 Adaptation of Cancer Hallmark Pathway Activity

The treatment effects on cancer hallmarks were investigated by performing a pathway analysis with the Pathway Responsive Genes for activity inference (PROGENy) gene

sets (Holland et al., 2020; Schubert et al., 2018). Relevant pathway enrichments in both BT474 and SKBR3 cells treated with ApoTras and Herc are also apparent in the analysis. Significantly, both cell lines have diminished PI3K and MAPK activity and increased p53 activity (**Figure 4-5**). The complete results of the PROGENy pathway activity analysis can be found [Additional file 8 Table S11-14](#).

For SKBR3 cells, a comparable decrease in PI3K and MAPK activity was detected, accompanied by a modest increase in p53 activity. Additionally, there was significant upregulation in pathways including NFkB, TNFA, JAK-STAT, and Hypoxia, accompanied by reduced activity in the Estrogen pathway in ApoTras-treated SKBR3 samples. The comparison of ApoTras to Her showed a reduced activity for the pathways TRAIL, NFkB, and TNFA signalling in SKBR3 cells. This suggests a specific modulation of pathways between the biosimilar and the original treatment.

These results suggest subtle functional differences between the biosimilar and original trastuzumab, with implications for the treatment response in HER2-positive breast cancer.

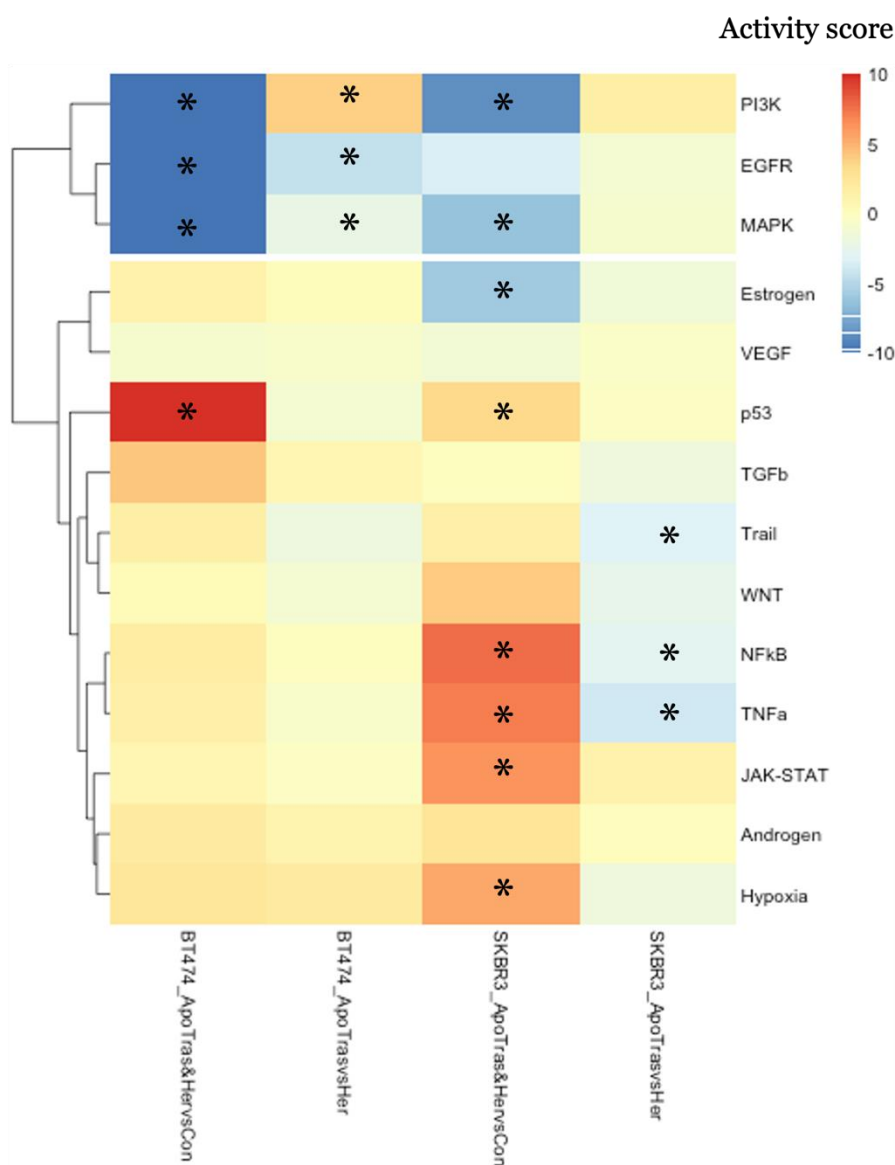


Figure 4-5 Cancer hallmark pathway (PROGENy) activities in BT474 and SKBR3 cell lines treated with Herceptin and/or its biosimilar. Significantly altered pathways are marked with asterisks (FDR < 0.05).

4.3.6 Regulatory Network Insights

4.3.6.1 Trastuzumab effect

The analysis of transcription factor (TF) activity in BT474 and SKBR3 breast cancer cell lines after treatment showed significant shifts in TF activity. The analysis was performed with decouplR (Badia-I-Mompel et al., 2022) and the TF database CollectTRI (Müller-Dott et al., 2023). When comparing treated samples against controls (ApoTras&HervsCon), a 5% false discovery rate (FDR) threshold was applied for defining significance. In the BT474 cell line, 48 TFs were upregulated and 26 downregulated, while in SKBR3, 101 TFs were upregulated and 24 downregulated. Of

these, 22 and 12 TFs were up- and down-regulated, respectively, in both cell lines, with no TF exerting opposing activity changes. This means the effect is consistent (**Figure 4-6**). Details of the calculated TF activities for each comparison can be found here [Additional file 8 Table S15-16](#).

Some of the relevant TFs, instead, have already been reported to be affected by trastuzumab. FOXO3 and FOXO1, involved in the control of (Eijkelenboom & Burgering, 2013; Furukawa-Hibi et al., 2005; Webb & Brunet, 2016), were significantly up-regulated in both BT474 and SKBR3 cell lines. AP-1 and TP53 were up-regulated in both cell lines. AP-1 represents an important family of proteins involved in cell proliferation and differentiation (Eferl & Wagner, 2003; Shaulian & Karin, 2001) and TP53 is a tumour suppressor gene essential in the cell response to DNA damage and regulation of cell cycle arrest (Meek, 2009; Vogelstein et al., 2000). These two TFs together contribute to trastuzumab's role in inducing cell death and activating tumour-suppression pathways. There were some cell line dependent responses for certain TFs. TNF- α stimulated NF- κ B was exclusively up-regulated in BT474. These results are in consonance with the involvement of NF- κ B in classically understood immune response (Hayden & Ghosh, 2011; T. Liu et al., 2017; Oeckinghaus & Ghosh, 2011). In contrast, SKBR3 was the only cell line to upregulate SP1, HIF1A, and STAT3. These TFs are associated with angiogenesis, immune modulation, and tumour resistance pathways (L. Dai et al., 2022; Kimura et al., 2023; McIntyre & Harris, 2016) indicating that such processes are increased in SKBR3 cells upon trastuzumab treatment. These two cell lines also consistently downregulated the activity of MYCN, E2F3, E2F4, E2F1, and MYC. These TF reductions are indicators that trastuzumab successfully inhibits the cell cycle progression and proliferation, which ultimately impact tumour growth (Albanell et al., 2003; Leone et al., 2001; Schwab, 2004).

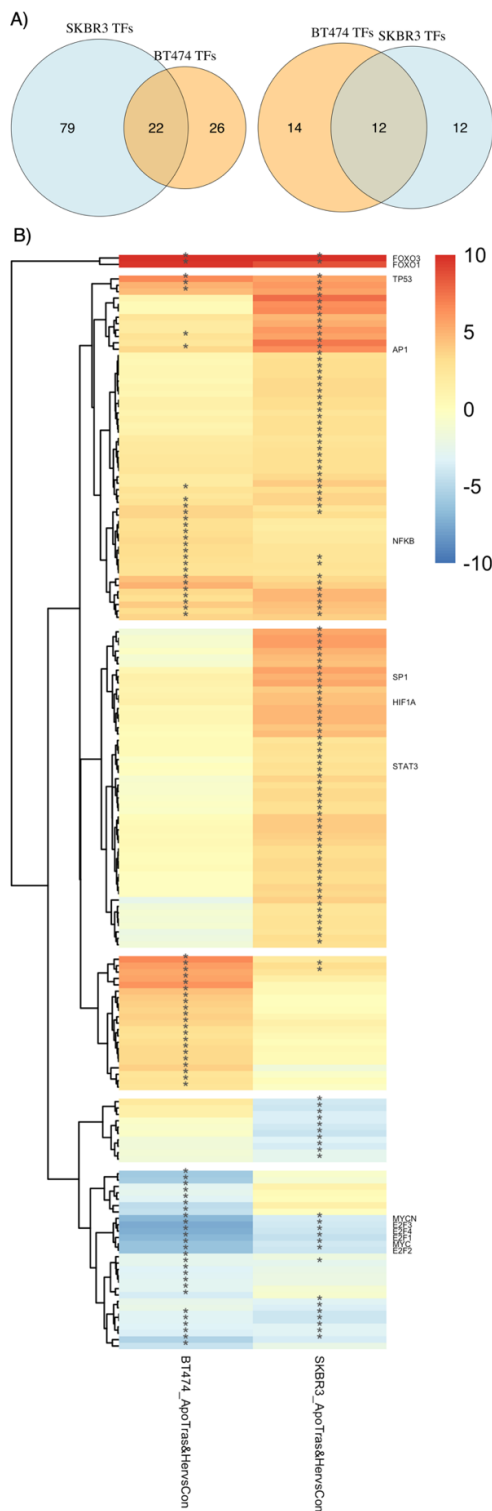


Figure 4-6 Analysis of Transcription Factor Activity and Pathway Enrichment. A) Venn diagram illustrating the number of transcription factors (TFs) with increased or decreased activity in BT474 and SKBR3 cell lines following trastuzumab treatment either biosimilar or Herceptin. The diagram highlights the overlap and unique TFs between the two cell lines. The left Venn diagram shows the TFs with increased activity for each cell line and the right one shows the TFs with decreased activity. B) Heatmap representing the activity scores of significant transcription factors across BT474 and SKBR3 cell lines. The heatmap visualizes clustering of the TFs based on their activity scores, with Euclidean distance used to calculate dissimilarities between pathways. Asterisks indicate an FDR value of < 1%.

Transcription factors (TFs) of interest were then subjected to cluster analysis based on their target genes by using the method cluster_walktrap (Pons & Latapy, 2006) focussing on their target genes. A total of 27 non-redundant DEGs were found among the significant transcription factors (TFs) in thirteen clusters of BT474, which did not share any significant TF with SKBR3 (**Figure 4-7**). Key clusters include Cluster 2 MYC family of transcription factors, Cluster 4 cell cycle regulation, and Cluster 7 FOXO family involved in apoptosis. These are the groups that highlight the role of MYC in proliferation, E2F in cell cycle progression, and FOXO in stress response.

SKBR3 showed the highest number of exclusive clusters, with 22 and 34 unique TFs. The most relevant clusters included Cluster 8, which comprised additional members of the MYC family, Cluster 6, which contained regulatory transcription factor SPI1, and Cluster 3, containing cell survival and differentiation signalling transcription factors. These clusters can be related to proliferation (MYC), immune modulation (SPI1), and tumour progression signalling.

The analysis also identified several overlapping clusters between BT474 and SKBR3, indicating shared regulatory mechanisms in response to trastuzumab treatment. For example, PPARG and PPARGC1A, transcription factors of lipid metabolism and inflammation, are found in Cluster 1 of BT474 and Cluster 17 of SKBR3. These transcription factors have critical functions in metabolic reprogramming and in the modulation of immune responses. Both of them are critical in tumour development as well as in the control of the tumour microenvironment (Desvergne & Wahli, 1999; Kersten, 2014; Palomer et al., 2018). This intersection indicates a common regulation network that trastuzumab has contributed to forming in the two cell lines.

Both SMAD2 and SMAD4 are part of the important intracellular mediators of the TGF- β signalling and are included in Cluster 2 in BT474 and Cluster 3 in SKBR3. These transcription factors control processes such as cell growth, differentiation and apoptosis. Their induction in both PD cell lines supports the hypothesis of an antitumor tumour inhibition by trastuzumab likely mediated by the TGF- β signalling activation following receptor blockade and thus favouring differentiation and apoptosis (Massagué et al., 2000; Shi & Massagué, 2003; Y. Yang et al., 2021).

Lastly, Cluster 5 in BT474 and Cluster 11 in SKBR3 also contain TP53, TP73 and TP63, the three members of the p53 family of transcription factors. They also play key roles in the DNA damage response, cell cycle control and apoptosis. Activation of these was nonetheless observed in both cell lines, indicating an efficient induction of p53 signalling by trastuzumab. Importantly, this highlights the role of the treatment in activating cellular (Ozaki & Nakagawara, 2005), thereby reinforcing the tumour-suppressive effects.

Other transcription factors shared by both cell lines play important roles in treatment resistance and tumour progression. AR (Androgen Receptor), detected in both cell lines, is involved in tumour cell survival and proliferation (Z. Chen et al., 2023). SMARCA1, which was shared, is involved in tumour suppression and DNA repair.

The known immune response and cell differentiation-regulator IRF1 and JUN are also acquired by both cell lines. Altered activity of these TFs may indicate roles in immune modulation, immune evasion, and tumour cell plasticity (Frontini et al., 2009).

Lastly, the analysis found MYC, MYCN, ZNF91, REST, and GLI2 in both cell lines. The oncogenes MYC and MYCN, which work to control proliferation and cell survival, have also been shown to be associated with the regulation of cell fate (R. et al. Huang, 2021; Ruiz-Pérez et al., 2017).

Several clusters were unique to BT474, indicating that the cell line has specific regulatory responses to trastuzumab treatment. More interesting is Cluster 10 that includes IRF 7 and KLF15, associated with immune regulation as well as transcriptional activation (Y. Cao et al., 2024; Liang et al., 2011) and that could relate trastuzumab to immune-modulatory outcomes in BT 474. Cluster 12, which includes NFATC4 and NFATC3, also indicates calcium signalling and immune-related pathways (Vihma et al., 2016). Similarly, Cluster 4, which includes DMTF1, FOXC2, ASCL2, FOXL2, FOXC1, FOXH1, ARX and IRX1, is interesting because of the presence of FOXC2, a transcription factor that plays a role in epithelial-to-mesenchymal transition (EMT) and metastasis. Its participation in tumour progression along with chemoresistance may indicate a role in drug resistance as well in BT474.

In SKBR3 a number of distinct clusters expose relevant cell line-specific functions or phenotypes that may account for treatment resistance or cancer pathogenesis. Cluster 6, which include NR3C2, TFAP2C, POU3F1, and HOXA7, is interesting because it contains HOXA7, one of the HOX gene family members, associated with cancer cell differentiation and drug resistance. In fact, HOXA7 is associated with the advancement and drug resistance in tumours (Dang et al., 2022; X. Duan et al., 2017; Y. Li et al., 2015). Cluster 9, comprising ELF1, PBX1, MBD2, and KDM2A is interesting for the central role of PBX1, a transcription factor involved in cell differentiation and growth (H. Chen et al., 2023; Magnani et al., 2011). Specifically, Cluster 8, containing CREB1, ATF2 and ATF1, is significant for CREB1, a transcription factor participating in cell survival, tumour progression, and chemoresistance (Kim, 2021). Lastly, Cluster 12, which includes ZNF263, CTNNB1, and TCF7, also contains CTNNB1 (or β -catenin), an important player in the Wnt signalling pathway. The Wnt/ β -catenin pathway is involved in processes of cell proliferation (Clevers & Nusse, 2012; MacDonald et al., 2009; Nusse, 2012). These particular SKBR3 clusters support a cell line-specific role of FOXC2, HOXA7, PBX1 and CREB1 in response to trastuzumab that also relates chemoresistance, tumour invasion, and metastasis. IRF7 was identified as uniquely upregulated TF in BT474, while ELF3, EHF and SP7 were identified in SKBR3. These TFs all had unique regulatory signatures in each of these cell lines.

Among other findings, the clustering analysis revealed some overlap as well as cell line-specific regulatory circuits in the two cell types. The intersecting clusters comprised TFs like MYC, E2F, FOXO1/3 and TP53. Plus, there were distinct clusters for each cell line like IRF7 in BT474 and ELF3 in SKBR3-specific clusters. In summary, it can be

concluded that trastuzumab activates both cell line-specific and shared clusters to define transcriptional regulatory networks in BT474 and SKBR3.

4.3.6.2 Assessment of Biosimilarity

To investigate the differential expression of transcription Factors (TF), BT474 and SKBR3 cell lines were treated with ApoTras, a biosimilar and Herceptin and an analysis for TF's activity was conducted. Of these, in BT474, 43 generated TFs were significantly differentially expressed; 36 were upregulated and 7 were downregulated (FDR < 0.05). In SKBR3, a total of thirty TFs were significantly regulated, of which seven were up- and twenty-three down-regulated (**Figure 4-8**). An overall greater number of up-regulated TFs can be observed in BT474 while more down-regulated TFs are present in SKBR3. Transcription factor activity calculation results for each comparison can be found here [Additional file 8 Table S17-18](#).

The two cell lines overlapped only for a very small portion of the transcription factors where USF2, and FOXN1 appeared to play a role and had increasing activity following the exposure. Growth factor signalling, immune function and hormonal mechanisms, which are regulated in part by USF2 and FOXN1, are other important targets of these genetic pathways (Gawronska-Kozak, 2020; Gawronska-Kozak et al., 2019; Kim, 2021; Romano et al., 2013; Webb & Brunet, 2016).

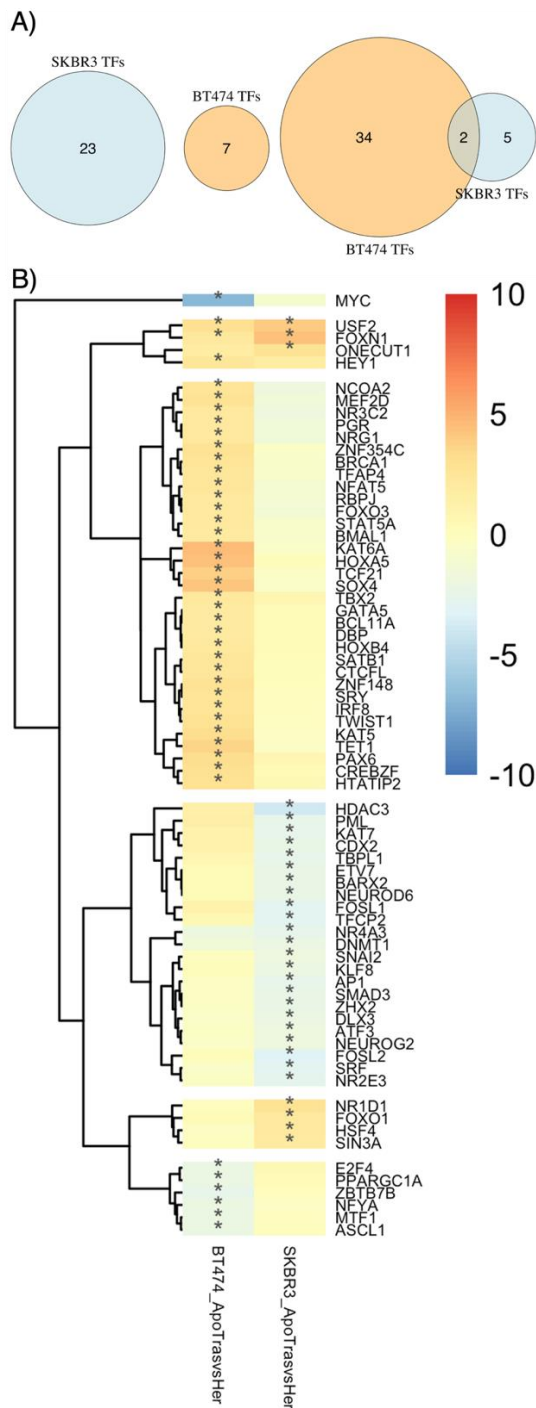


Figure 4-8 Transcription Factor Activity in Response to Biosimilar vs. Herceptin Treatment in BT474 and SKBR3 Cell Lines. A) Venn diagram showing the overlap of significantly regulated transcription factors (TFs) between BT474 and SKBR3 cell lines following treatment with Biosimilar vs. Herceptin. A total of 43 TFs were significantly regulated in BT474, with 36 upregulated and 7 downregulated. For SKBR3, 30 TFs were significantly regulated, with 7 upregulated and 23 downregulated. The overlap between the two cell lines consists of 2 shared TFs, USF2 and FOXN1, which show increased activity in both cell lines. B) Heatmap illustrating the activity of significantly regulated TFs in BT474 and SKBR3 following biosimilar and Herceptin treatment. The heatmap is based on scaled activity scores, with upregulated TFs shown in yellow/orange and downregulated TFs in blue. Statistical significance is indicated by asterisks (p-value < 0.05). The clustering of TFs reveals distinct patterns of regulation between the two cell lines, highlighting shared and unique responses to the treatments.

After treatment with biosimilar (ApoTras) or Herceptin across BT474 and SKBR3, breast cancer cells, a high-level view of differential transcription factor (TF) activity, as shown by network diagrams and volcano plots, is seen in (**Figure 4-9**).

In BT474, seven distinct clusters of TFs that are significantly regulated were identified using clustering analysis. Cluster 2 contains FOXO3 and FOXN1 and cluster 4 comprises STAT5A and GATA5. Two distinct clusters were identified in SKBR3. FOXN1, FOXO1 and NR4A3 were included in Clusters 1 and 2, demonstrating the distinctive regulation of these TFs in the two cell lines.

Cluster 2 in BT474 and 1 in SKBR3 presented a transcription factor, FOXN1 that is common to both cell lines. Specifically, in BT474 cluster 4, which include STAT5A and GATA5, include TFs linked to regulation of the cell while in SKBR3 clusters were associated with immune and stress response. As shown in the volcano plots in **Figure 4-9** the target genes related to USF2 and FOXN1 were significantly upregulated in BT474. USF2 overexpressed MYH9, IGF2R, AR, POU2F1, ACACA, THBS1, CREBBP, EGFR, APC, PPARA. Genes that had been upregulated are FOXN1 targets AR, EGFR, EP400, BAZ2A, ATP7A, KDM3A, CHD4 and ESR1. In BT474, EGFR and AR were common targets of USF2 and FOXN1. ESR1 was uniquely overexpressed among FOXN1 targets. All three genes have been associated to trastuzumab resistance.

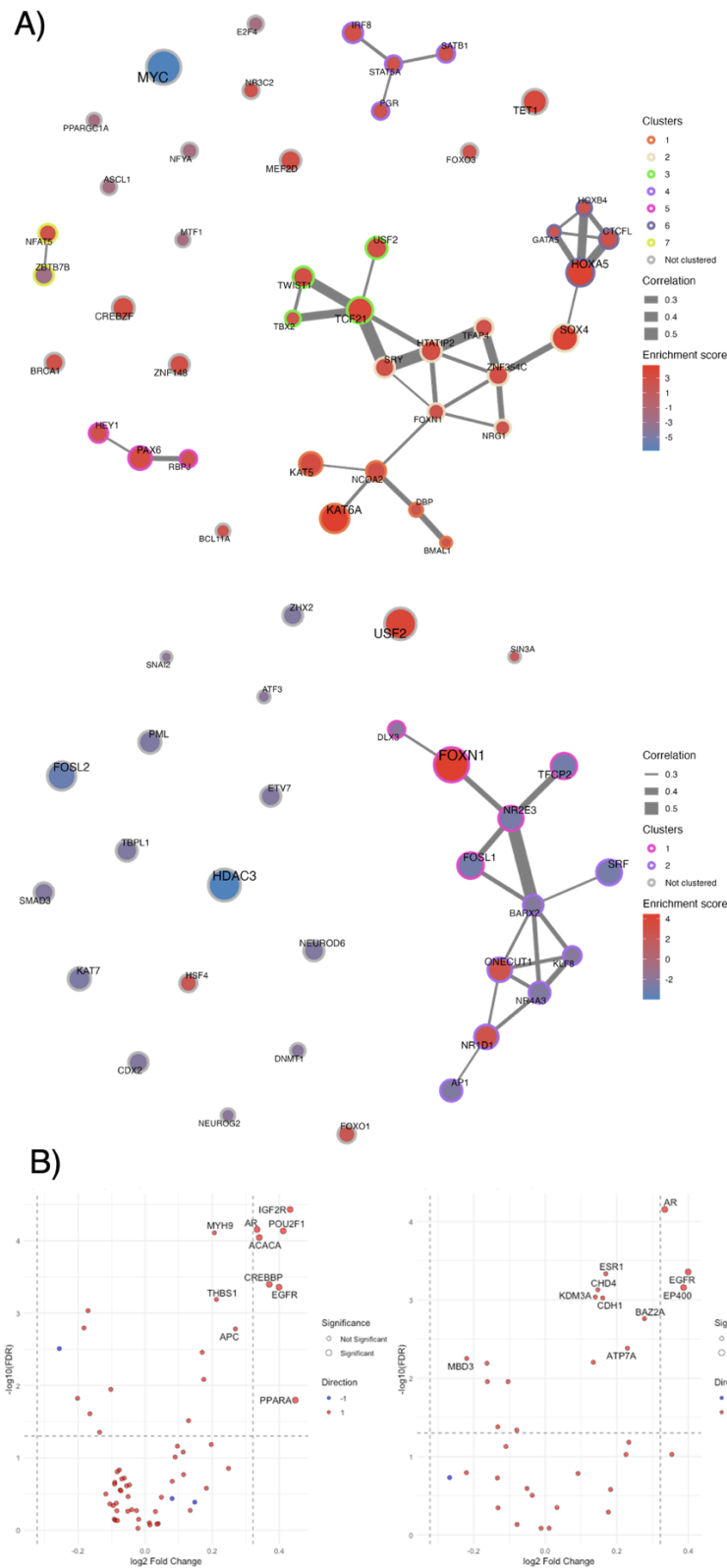


Figure 4-9 Transcription Factor Networks and Volcano Plot Analysis of Biosimilar vs. Herceptin Treatment in BT474. A) Network diagrams showing the relationships between significantly regulated transcription factors (TFs) in BT474 (top) and SKBR3 (bottom) cell lines following biosimilar and Herceptin treatment. The nodes represent individual TFs, coloured according to their activity (upregulated: red, downregulated: blue). The edges between TFs indicate shared target genes with thicker edges representing stronger correlations. B) Volcano plots of TF targets of USF2 (left) and FOXN1 (right) for BT474 cell line.

4.3.7 Distinct and Shared Pathway Changes in Response to Treatment

4.3.7.1 Trastuzumab effects

Reactome-based GSEA revealed a total of 457 gene sets that were significantly enriched ($p_{\text{adjust}} < 0.05$) between the BT474 and SKBR3 cell lines. Of these, 197 pathways showed common enrichment in the two cell lines, thus indicating a shared biological response and baseline characteristics of trastuzumab treatment. Interestingly, they found 52 unique gene sets significantly enriched in BT474 cell line and 208 in SKBR3, thus indicating independently activated pathways in the two cell lines (**Figure 4-10**). Complete GSEA analysis results for each comparison are provided here [Additional file 9 Table S18-19](#).

The significant Reactome gene sets were subjected to a hierarchical clustering analysis to remove redundancy and aid interpretation. The leading-edge method highlighted sets of DEGs involved in multiple enriched gene sets and the similarity between pathways was measured in terms of the overlapping DEGs between the different pathways. The overlap ratio was combined with the Jaccard index to compute the similarity index between pathways, with each contributing equally (i.e. 0.5 each) and 1 - Euclidean distance was used to cluster pathways.

Applied to our 457 gene sets, we obtained 40 clusters, with cluster size varying from 1 to 62 gene sets (mean of 11 gene sets per cluster; **Figure 4-10**). Cell cycle regulation and protein degradation formed the largest clusters, each containing 62 gene sets, highlighting their central role in cellular activity. Smaller clusters with single gene sets were represented by more specific biological processes, such as the Fanconi Anemia Pathway and the Metabolism of folate and pterins. Hierarchical clustering helped reduce pathway redundancy and identify biologically relevant patterns improving interpretation of the enrichment results.

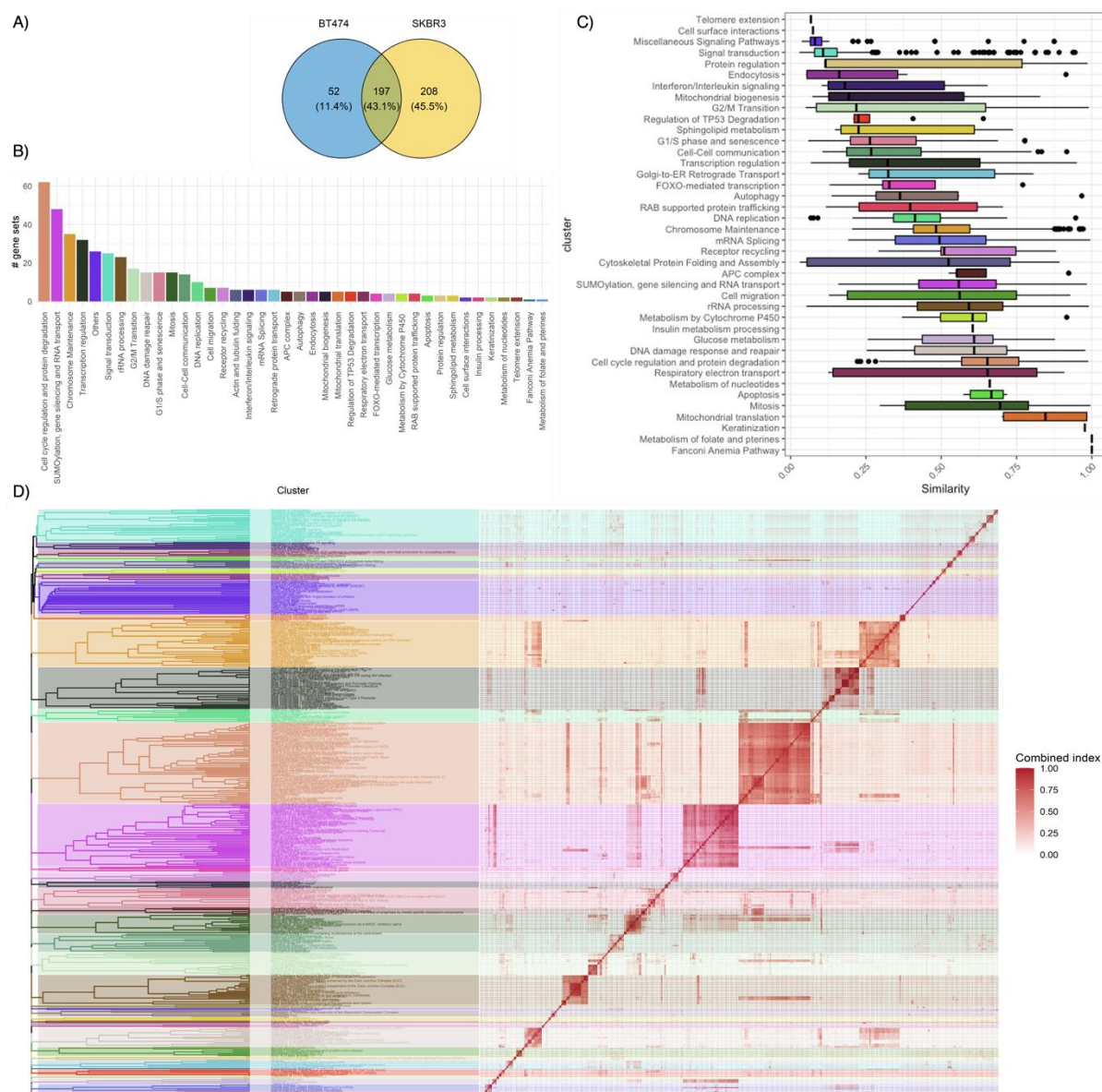


Figure 4-10 Overview of pathway clustering and similarity analysis of Reactome pathways comparing both treatments, Herceptin and ApoTras, against their respective controls for BT474 and SKBR3 cell lines. A) Venn diagram displaying the overlap and unique significant gene sets identified in BT474 and SKBR3 cell lines. B) Bar plot summarizing the number of gene sets per cluster after hierarchical clustering. Clusters were generated by cutting the dendrogram into 40 groups, with the number of gene sets per cluster ranging from 1 to 62 and an average of 11 gene sets per cluster. C) Boxplot of similarity values (based on shared differentially expressed genes) for each cluster. The median similarity value across clusters is 0.49. D) Dendrogram showing the hierarchical clustering of pathways on the left and the corresponding heatmap of similarity values on the right. The heatmap visualizes the combined Jaccard index and overlap values, with higher similarity represented by darker colours. Clusters are annotated with distinct colours.

Among the 40 identified clusters, 11 showed complete overlap between the BT474 and SKBR3 cell lines, indicating shared mechanisms of response to the treatments, independent of the specific HER2-positive cell line. Partial overlap was observed in 20 clusters, while 12 clusters were cell line-specific (**Figure 4-11, Table 4-1**). The complete

results of the clustering are available in [Additional file 9 Table S20](#). Three clusters were unique to BT474, reflecting distinct molecular responses in this cell line. In contrast, nine clusters were exclusive to SKBR3, demonstrating greater heterogeneity in pathway regulation and treatment responses in this cell line.

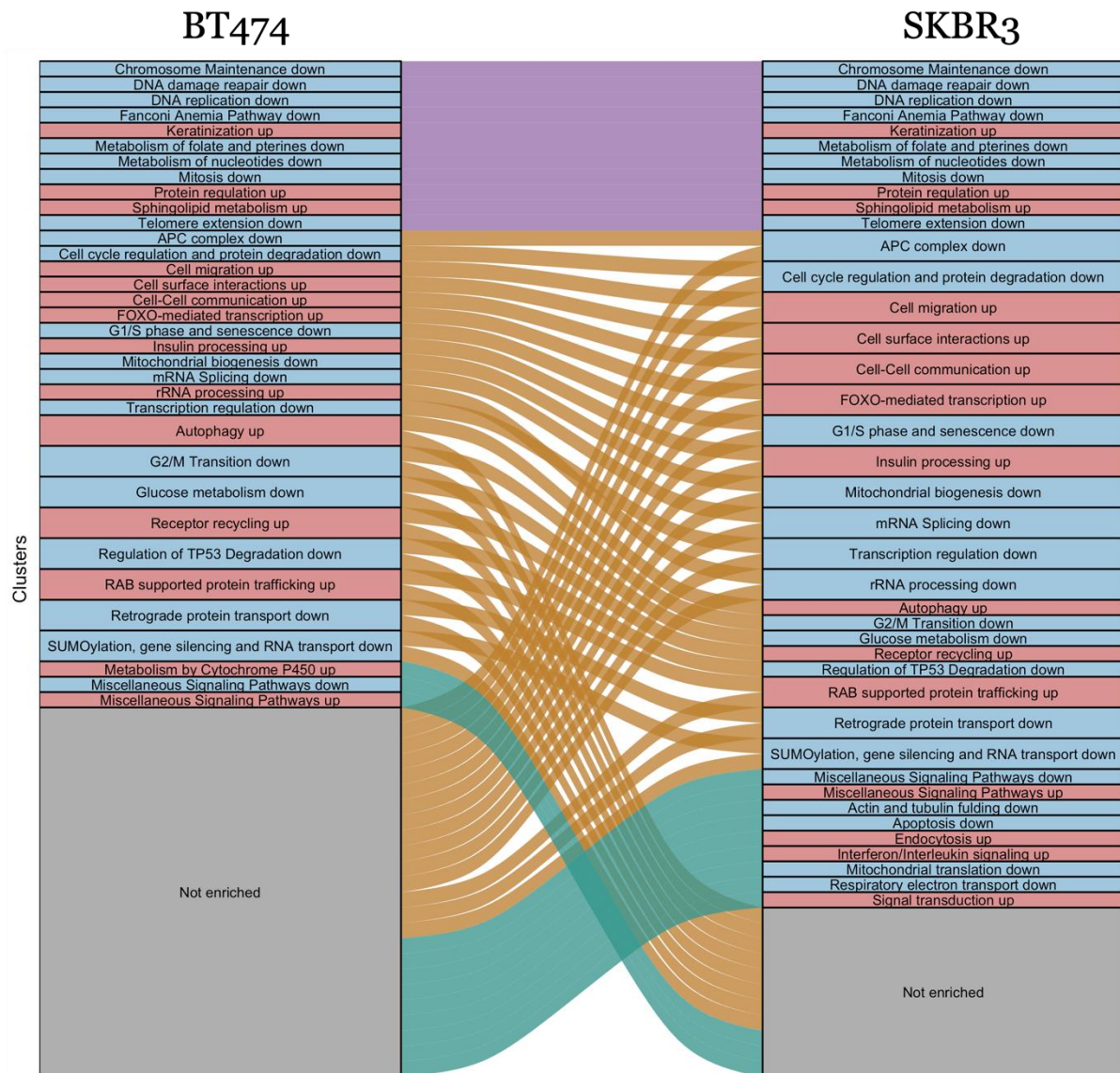


Figure 4-11 Sankey plot illustrating the distribution and overlap of gene set enrichment clusters between BT474 and SKBR3 cell lines in response to Herceptin and ApoTras treatment compared to control. The Sankey plot visualizes pathway enrichment patterns and their relationships between BT474 (left) and SKBR3 (right) cell lines. Clusters are color-coded to indicate their regulation direction: red for upregulated and blue for downregulated pathways. The connecting flows between clusters demonstrate three types of relationships: purple bands indicate complete overlap of pathways between cell lines, brown bands show partial overlap with connections to "Not enriched" categories, and green bands represent cell line-specific responses.

Overall, the results at the pathway level indicated that DNA maintenance and repair were impacted by trastuzumab and ApoTras treatments. Highly relevant clusters such as Chromosome Maintenance, DNA Damage Repair, DNA Replication, DNA Replication pathways, Fanconi Anemia Pathway and Telomere Extension all exhibit consistent downregulation in both BT474 and SKBR3 cell lines (Biertümpfel et al., 2010). These include PCNA, RPA1, POLD1, POLD2, and POLD3 genes, for which an attenuated expression was identified. In BT474 cells this included DNA2, TOP3A, BRIP1, RMI2 and BLM and in SKBR3 RMI1, BLM, EXO1, RBBP8 and BARD1 were among the DNA Damage Repair-related genes that were included in the DNA Damage Repair cluster that was also downregulated. Also in vitro samples, inhibition of cell division and therefore of DNA replication was shown by the decrease of the DNA Replication cluster and of MCM4, MCM8 and ORC1 (Johnson et al., 2003; S. Wang & Ding, 2019). On top of that, downregulation of RTEL1, CIAO1 and RUVBL2 in BT474 and FAM96B, NOP10, DKC1 in SKBR3 cells have been demonstrated to disrupt the Fanconi Anaemia Pathway, responsible for the DNA interstrand cross-link repair, and telomere maintenance. Altogether, these data indicate that the treatment induced common changes in replication and in DNA repair processes in both cell types.

Cell cycle regulation pathways were also affected, including the Cell Cycle Regulation and Protein Degradation cluster and those associated with Mitosis, G2/M Transition, Regulation of TP53 Degradation, FOXO-mediated Transcription, and the APC Complex. PSMB8, PSME2, PSMD11, PSMC4 were significantly downregulated among the genes associated to the Cell Cycle Regulation and Protein Degradation cluster, a finding that indicates a repression of proteasomal degradation which is critically involved in regulating essential cell cycle components in both cell lines (Bousquet et al., 2019; Livneh et al., 2016). In addition, genes of G2/M Transition cluster, including PLK1, CDK1, and HAUS6, were downregulated, indicating an impaired progression from G2 to M phase of mitosis (Hydbring et al., 2017). Analogous result was obtained in the Mitosis cluster, in which CENPI, CENPM, CENPL and all other genes were downregulated and showed evidence of problems in mitotic spindle formation and chromosome alignment (Pavin & Tolić, 2021). Also, CDK1, CDC25C, and CHEK2 expression levels were lower, indicating a reduction in pathways which degrade TP53. On the contrary, the CAV1, BTG1, and CCNG2 transcripts, mediated by FOXO, are upregulated, indicating that FOXO promotes cell cycle arrest in association with the apoptotic environment (Webb & Brunet, 2016). Also, genes associated with the

APC Complex cluster, responsible for the transition from metaphase into anaphase, were also down-regulated from genes such as CDC27, ANAPC1, UBE2S (Alfieri et al., 2016).

Treatment had, in fact, a great impact on cellular differentiation processes and metabolic pathways. Thus, alternate regulation patterns were present in clusters connected with sphingolipid metabolism, general protein regulation, keratinisation, metabolization of pterines, and metabolization of nucleotides. The repression of genes belonging to the clusters Metabolism of Nucleotides NME1, DTYMK and NME4 as well as Metabolism of Pterines SLC19A1, SHMT2, and MTHFD2 indicate reduced nucleotide and pterin metabolism, fundamental for DNA synthesis and repair processes (Fuhs & Hunter, 2017; Nazki et al., 2014). Conversely, genes involved in the regulation of proteins, such as CST3, ADAM10 and STC2, were up-regulated, pointing towards activation of pathways connected to protein homeostasis and stress responses. The Sphingolipid Metabolism cluster with members ASAH1, CTSA and PSAP and Keratinisation, with KRT15, KLK12 and PERP showed up-regulation. This suggests eventually the possible modulation of lipid signalling (Ogretmen, 2018) and differentiation pathways (Candi et al., 2005). A number of genes from the Mitochondrial Biogenesis cluster, e.g. ATP5B and HSPA9, as well as the Respiratory Electron Transport cluster, e.g. NDUFAB1 and NDUFB1, were downregulated, highlighted inhibition in mitochondrial metabolism (Scarpulla et al., 2011).

In addition to common pathway activation, BT474 and SKBR3 also showed unique clusters. In BT474 cells, the Cytochrome P450 metabolism cluster was induced, including upregulation of CYP1A1, CYP4F3 and CYP4B1. The genes are involved in the oxidative metabolism of various substances, suggesting a metabolic specificity for the cell line's activity profile (Jiang, Sun, et al., 2021). Among the most interesting responses in SKBR3 were the downregulation of the Apoptosis cluster, including CASP3, CDKN2A, MAPK3 (Pastena et al., 2024) and the Actin and Tubulin Folding cluster and the relevance of this downregulation is that proteins as TCP1, CCT6A, and CCT3 are key to cytoskeletal stability (and thus tubulin polymerization) (Jönsson et al., 2007). The Endocytosis cluster was upregulated and included genes like PIK3C2A, OCRL and ARRB1, which are indicative of increased receptor internalization. The Interferon/Interleukin Signalling cluster, which includes JAK1, IL18, and STAT1, upregulated indicates activation of specific immune signalling pathways. Among these, downregulation of the Mitochondrial Translation cluster and the Respiratory Electron Transport cluster was observed, with MRPL9, ERAL1 and NDUFAB1 expression being

reduced. On the other hand, the Signal Transduction cluster, comprising APP, EGF and THBS1, was upregulated.

Both BT474 and SKBR3 cells share common molecular pathways that are perturbed by trastuzumab treatment, some specific responses to trastuzumab treatment occur only in one or the other cell line. These results reveal the complexity of molecular effects of trastuzumab and the heterogeneity in response observed in HER2-positive breast cancer tumour models.

Table 4-1 Gene set enrichment analysis of Reactome pathways comparing Herceptin and ApoTras treatments against the control samples in BT474 and SKBR3 cell lines. Clusters were grouped based on shared leading-edge genes, and the overlap, direction of regulation, and significant gene sets in each cluster were identified for both cell lines. The table lists the top genes for each cluster, the number of significant gene sets in each cell line and overlapping pathways between the cell lines. Clusters are classified by their overlap type (Complete, BT474-specific, SKBR3-specific, or partial overlap) and the direction of regulation (up-regulated or down-regulated).

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Sets	SKBR3 Top 5 Genes	SKBR3 Significant Sets	Overlapping Pathways
Actin and tubulin folding	SKBR3-specific	Down-regulated	-	-	TCP1, CCT6A, CCT5, CCT2, CCT3	6	
APC complex	Partial	Down-regulated	CDC27, ANAPC1, UBE2S, UBE2C, MAD2L1	4	ANAPC5, CDC27, ANAPC10, UBE2E1, UBE2C	5	APC-Cdc20 mediated degradation of Nek2A, Inactivation of APC/C via direct inhibition of the APC/C complex, Inhibition of the proteolytic activity of APC/C required for the onset of anaphase by mitotic spindle checkpoint components, Phosphorylation of the APC/C
Apoptosis	SKBR3-specific	Down-regulated	-	-	CASP3, CDKN2A, CYCS, MAPK3, C1QBP	3	
Autophagy	Partial	Up-regulated	MAP1LC3B, PIK3C3, CHMP4C, UBC, BECN1	5	MAP1LC3B, EPAS1, CHMP4B, NBR1, CHMP2A	2	Pexophagy, Translation of Replicase and Assembly of the Replication Transcription Complex
Cell cycle regulation and protein degradation	Partial	Down-regulated	PSMB8, PSME2, PSMD11, PSMC4, PSMC3	17	PSMD10, PSMB1, PSMB3, PSME2, PSMD11	62	Activation of APC/C and APC/C:Cdc20 mediated degradation of mitotic proteins, APC:Cdc20 mediated degradation of cell cycle proteins prior to satisfaction of the cell cycle checkpoint, APC/C-mediated degradation of cell cycle proteins, APC/C:Cdc20 mediated degradation of mitotic proteins, APC/C:Cdh1 mediated degradation of Cdc20 and other APC/C:Cdh1 targeted proteins in late mitosis/early G1, Assembly of the pre-replicative complex, Cdc20:Phospho-APC/C mediated degradation of Cyclin A, CDK-mediated phosphorylation and removal of Cdc6, DNA Replication, DNA Replication Pre-Initiation, G1/S DNA Damage Checkpoints, Orc1 removal from chromatin, Regulation of APC/C activators between G1/S and early anaphase, Regulation of mitotic cell cycle, S Phase, Switching of origins to a post-replicative state, Synthesis of DNA

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Sets	SKBR3 Top 5 Genes	SKBR3 Significant Sets	Overlapping Pathways
Cell migration	Partial	Up-regulated	FXYD3, TPM4, CACNG4, ANXA1, VIM	1	MYL9, MYL12A, MYL12B, MYH14, MYH9	7	Muscle contraction
Cell surface interactions	Partial	Up-regulated	PSG4, CAV1, CD47, SDC4, CD44	1	CD44, SDC4, GPC1, CEACAM6, GUSB	2	Cell surface interactions at the vascular wall
Cell-Cell communication	Partial	Up-regulated	ID4, CAV1, IRS2, ID2, ID3	1	ITGB1, FN1, ITGA2, LAMB3, LAMC1	14	Signalling by Receptor Tyrosine Kinases
Chromosome Maintenance	Complete	Down-regulated	PCNA, RPA1, POLD2, POLD1, POLD3	35	PCNA, RPA3, POLD3, POLD1, POLD2	35	Base Excision Repair, Chromosome Maintenance, DNA Damage Bypass, DNA Repair, Dual Incision in GG-NER, Dual incision in TC-NER, Extension of Telomeres, Gap-filling DNA repair synthesis and ligation in GG-NER, Gap-filling DNA repair synthesis and ligation in TC-NER, Global Genome Nucleotide Excision Repair (GG-NER), Lagging Strand Synthesis, Leading Strand Synthesis, Mismatch Repair, Mismatch repair (MMR) directed by MSH2:MSH3 (MutSbeta), Mismatch repair (MMR) directed by MSH2:MSH6 (MutSalpha), Nucleotide Excision Repair, PCNA-Dependent Long Patch Base Excision Repair, Polymerase switching, Polymerase switching on the C-strand of the telomere, Processive synthesis on the C-strand of the telomere, Processive synthesis on the lagging strand, Recognition of DNA damage by PCNA-containing replication complex, Removal of the Flap Intermediate, Removal of the Flap Intermediate from the C-strand, Resolution of Abasic Sites (AP sites), Resolution of AP sites via the multiple-nucleotide patch replacement pathway, Telomere C-strand (Lagging Strand) Synthesis, Telomere Maintenance, Termination of translesion DNA synthesis, Transcription-Coupled Nucleotide Excision Repair (TC-NER), Translesion Synthesis by POLH, Translesion synthesis by POLI, Translesion synthesis by POLK, Translesion synthesis by REV1, Translesion synthesis by Y family DNA polymerases bypasses lesions on DNA template
DNA damage repair	Complete	Down-regulated	DNA2, TOP3A, BRIP1, RMI2, BLM	15	RMI1, BLM, EXO1, RBBP8, BARD1	15	DNA Double-Strand Break Repair, G2/M DNA damage checkpoint, HDR through Homologous Recombination (HRR), HDR through Homologous Recombination (HRR) or Single Strand Annealing (SSA), HDR through Single Strand Annealing (SSA), Homologous DNA Pairing and Strand Exchange, Homology Directed Repair, Presynaptic phase of homologous DNA pairing and strand exchange, Processing of DNA double-strand break ends, Regulation of TP53 Activity, Regulation of TP53 Activity through Phosphorylation, Resolution of D-Loop Structures, Resolution of D-

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Sets	SKBR3 Top 5 Genes	SKBR3 Significant Sets	Overlapping Pathways
							loop Structures through Holliday Junction Intermediates, Resolution of D-loop Structures through Synthesis-Dependent Strand Annealing (SDSA), Transcriptional Regulation by TP53
DNA replication	Complete	Down-regulated	POLA2, MCM8, MCM4, ORC1, MCM3	10	MCM8, ORC5, ORC6, MCM2, ORC1	10	Activation of ATR in response to replication stress, Activation of the pre-replicative complex, CDC6 association with the ORC:origin complex, Cell Cycle Checkpoints, DNA strand elongation, E2F mediated regulation of DNA replication, G1/S Transition, G2/M Checkpoints, Mitotic G1 phase and G1/S transition, Unwinding of DNA
Endocytosis	SKBR3-specific	Up-regulated	-	-	PIK3C2A, OCRL, ARRB1, DNAJC6, VAMP8	5	
Fanconi Anemia Pathway	Complete	Down-regulated	C17orf70, STRA13, FANCF, MUS81, DCLRE1A	1	C19orf40, FANCM, C17orf70, C1orf86, APITD1	1	Fanconi Anemia Pathway
FOXO-mediated transcription	Partial	Up-regulated	CAV1, BTG1, FOXO4, CCNG2, RBL2	2	FOXO3, SMAD3, SMAD2, CITED2, BTG1	4	FOXO-mediated transcription, FOXO-mediated transcription of cell cycle genes
G1/S phase and senescence	Partial	Down-regulated	E2F1, CCNE1, TFDP1, CCNE2, CDK2	7	E2F1, TFDP1, CCNE1, CDK2, CCNE2	15	Aberrant regulation of mitotic cell cycle due to RB1 defects, Cellular Senescence, Diseases of mitotic cell cycle, G0 and Early G1, G1/S-Specific Transcription, TP53 Regulates Transcription of Genes Involved in G1 Cell Cycle Arrest, Transcriptional Regulation by E2F6
G2/M Transition	Partial	Down-regulated	CDK1, PLK1, NEDD1, HAUS6, CEP41	17	CDK1, PLK1, HAUS6, CEP76, PLK4	15	Anchoring of the basal body to the plasma membrane, AURKA Activation by TPX2, Centrosome maturation, Condensation of Prometaphase Chromosomes, Condensation of Prophase Chromosomes, Cyclin A/B1/B2 associated events during G2/M transition, G2/M Transition, Initiation of Nuclear Envelope (NE) Reformation, Loss of Nlp from mitotic centrosomes, Loss of proteins required for interphase microtubule organization from the centrosome, Mitotic G2-G2/M phases, Polo-like kinase mediated events, Recruitment of mitotic centrosome proteins and complexes, Recruitment of NuMA to mitotic centrosomes, Regulation of PLK1 Activity at G2/M Transition
Glucose metabolism	Partial	Down-regulated	ALDOA, PGK1, ENO1,	4	PGK1, PGAM1, GPI,	3	Gluconeogenesis, Glucose metabolism, Glycolysis

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Sets	SKBR3 Top 5 Genes	SKBR3 Significant Sets	Overlapping Pathways
Insulin processing	Partial	Up-regulated	TPH1, GAPDH INHBB, KIF5C, CTSD, EXOC4, EXOC1	1	GAPDH, ALDOC ERO1L, CPE, EXOC3, EXOC1, EXOC4	2	Peptide hormone metabolism
Interferon/Interleukin signalling	SKBR3-specific	Up-regulated	-	-	JAK1, STAT1, ICAM1, CCL22, IL18	6	
Keratinization	Complete	Up-regulated	KRT15, KLK12, PERP, FURIN, KLK5	2	IVL, KRT23, JUP, KRT27, DSC2	2	Formation of the cornified envelope, Keratinization
Metabolism by Cytochrome P450	BT474-specific	Up-regulated	CYP1A1, CYP4F3, CYP4B1, CYP4F8, CYP4F11	4		-	
Metabolism of folate and pterines	Complete	Down-regulated	SLC19A1, SHMT2, SLC25A3 2, MTHFD1 L, MTHFD2	1	SHMT1, DHFR, FPGS, MTHFD2, SLC19A1	1	Metabolism of folate and pterines
Metabolism of nucleotides	Complete	Down-regulated	NME1, NME4, DUT, DCTPP1, DTYMK	2	GUK1, CTPS1, TXN, AK6, NME1	2	Interconversion of nucleotide di- and triphosphates, Metabolism of nucleotides
Mitochondrial biogenesis	Partial	Down-regulated	ATP5B, HSPA9, MARK4, NDUFB8, DYNLL1	2	CHCHD3, MTX2, HSPA9, ATP5G1, ATP5B	5	Mitochondrial protein import, Organelle biogenesis and maintenance

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Sets	SKBR3 Top 5 Genes	SKBR3 Significant Sets	Overlapping Pathways
Mitochondrial translation	SKBR3-specific	Down-regulated	-	-	MRPL9, MRPS18B, ERAL1, MRPL54, MRPL48	5	
Mitosis	Complete	Down-regulated	CENPI, CENPM, CENPL, CENPH, CENPU	15	CENPH, CENPO, CASC5, CENPU, CENPM	15	Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal, Amplification of signal from the kinetochores, Deposition of new CENPA-containing nucleosomes at the centromere, EML4 and NUDC in mitotic spindle formation, M Phase, Mitotic Anaphase, Mitotic Metaphase and Anaphase, Mitotic Prometaphase, Mitotic Spindle Checkpoint, Nucleosome assembly, Resolution of Sister Chromatid Cohesion, RHO GTPase Effectors, RHO GTPases Activate Formins, Separation of Sister Chromatids, Signalling by Rho GTPases
mRNA Splicing	Partial	Down-regulated	SNRPF, SRSF3, DHX38, LOC102724594, MAGOHB	4	SNRPF, SNRPE, SNRPG, SNRPB, FIP1L1	6	mRNA Splicing, mRNA Splicing - Major Pathway, Processing of Capped Intron-Containing Pre-mRNA, RNA Polymerase II Transcription Termination
Protein regulation	Complete	Up-regulated	CST3, ADAM10, STC2, PRSS23, IGFBP3	3	CST3, APP, ADAM10, MFGE8, GOLM1	3	Amyloid fiber formation, Post-translational protein phosphorylation, Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)
RAB supported protein trafficking	Partial	Up-regulated	RAB5B, RAB7A, RAB11A, RAB6A, RAB5A	3	RAB5B, RAB5A, ULK1, RAB7A, OPTN	3	Rab regulation of trafficking, TBC/RABGAPs
Receptor recycling	Partial	Up-regulated	ATP6V1D, ATP6V1H, ATP6V1E1, ATP6V1G1, ATP6V1C1	7	NEDD4L, ATP6V0A4, ATP9A, ATP8B1, TRPM4	1	Ion channel transport

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Sets	SKBR3 Top 5 Genes	SKBR3 Significant Sets	Overlapping Pathways
Regulation of TP53 Degradation	Partial	Down-regulated	CDK1, TP53, CDC25C, CHEK2, SFN	5	CDC25C, ZNF385A, RBL1, CCNE1, CDK2	1	TP53 Regulates Transcription of Cell Cycle Genes
Respiratory electron transport	SKBR3-specific	Down-regulated		-	NDUFAB1, NDUFB1, NDUFAF6, NDUFC1, NDUFB11	5	
Retrograde protein transport	Partial	Down-regulated	KIF22, TUBB3, KIF2A, CENPE, KIF20B	4	TUBA4A, TUBA1C, TUBB4B, TUBA1B, KIF23	4	COPI-dependent Golgi-to-ER retrograde traffic, Kinesins
rRNA processing	Partial	Up-regulated	RPS27L, RPL22, RPL23, RPS4X, RPL7	10	RPS7, RPS9, RPS2, RPL21, RPL7	23	Eukaryotic Translation Elongation, Eukaryotic Translation Termination, Formation of a pool of free 40S subunits, Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC), Peptide chain elongation, Response of EIF2AK4 (GCN2) to amino acid deficiency, Selenoamino acid metabolism, Selenocysteine synthesis, SRP-dependent cotranslational protein targeting to membrane, Viral mRNA Translation
Signal transduction	SKBR3-specific	Up-regulated		-	APP, PSAP, EGF, THBS1, DLL4	25	
Sphingolipid metabolism	Complete	Up-regulated	ASAHI, CTSA, PSAP, ARSD, HEXA	3	GM2A, CTSA, PSAP, ASAHI, HEXB	3	Glycosphingolipid metabolism, Neutrophil degranulation, Sphingolipid metabolism
SUMOylation, gene silencing and RNA transport	Partial	Down-regulated	SEH1L, NUPL1, NUP85, NDC1, NUP155	43	NUP160, NUP155, NUP43, NUP35, NUP37	45	Defective TPR may confer susceptibility towards thyroid papillary carcinoma (TPC), Export of Viral Ribonucleoproteins from Nucleus, HIV Infection, HIV Life Cycle, Host Interactions of HIV factors, Interactions of Rev with host cellular proteins, Interactions of Vpr with host cellular proteins, Metabolism of non-coding RNA, Mitotic Prophase, NEP/NS2 Interacts with the Cellular Export Machinery, NS1 Mediated Effects on Host Pathways, Nuclear Envelope (NE) Reassembly, Nuclear Envelope Breakdown, Nuclear import of Rev protein, Nuclear Pore Complex (NPC) Disassembly, Postmitotic nuclear pore complex (NPC) reformation, Regulation of Glucokinase by Glucokinase Regulatory Protein, Regulation of HSF1-mediated heat shock response, Rev-mediated nuclear export of HIV RNA, snRNP Assembly, SUMO

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Sets	SKBR3 Top 5 Genes	SKBR3 Significant Sets	Overlapping Pathways
							E3 ligases SUMOylate target proteins, SUMOylation, SUMOylation of chromatin organization proteins, SUMOylation of DNA damage response and repair proteins, SUMOylation of DNA replication proteins, SUMOylation of RNA binding proteins, SUMOylation of SUMOylation proteins, SUMOylation of ubiquitylation proteins, Transcriptional regulation by small RNAs, Transport of Mature mRNA derived from an Intron-Containing Transcript, Transport of Mature mRNA Derived from an Intronless Transcript, Transport of Mature mRNAs Derived from Intronless Transcripts, Transport of Mature Transcript to Cytoplasm, Transport of Ribonucleoproteins into the Host Nucleus, Transport of the SLBP Dependant Mature mRNA, Transport of the SLBP independent Mature mRNA, tRNA processing, tRNA processing in the nucleus, Viral Messenger RNA Synthesis, Vpr-mediated nuclear import of PICs
Telomere extension	Complete	Down-regulated	RTEL1, CIAO1, RUVBL2, ACD, NDOR1	2	FAM96B, NOP10, DKC1, ERCC2, CIAPIN1	2	Cytosolic iron-sulfur cluster assembly, Telomere Extension By Telomerase
Transcription regulation	Partial	Down-regulated	ERCC2, CDK9, SUDS3, CD3EAP, TCEA1	2	POLR2H, POLR2K, POLR2E, MNAT1, CCNH	32	Epigenetic regulation of gene expression, TP53 Regulates Transcription of DNA Repair Genes

4.3.7.2 Assessment of Biosimilarity

Gene set enrichment analysis (GSEA) using the Reactome database was performed to compare pathway regulation changes between Herceptin and ApoTras treatments at an FDR threshold of 0.05. In BT474 cells, 404 gene sets were significantly altered, while in SKBR3 cells, 46 gene sets were identified as altered (**Figure 4-12**). Among these, 42 gene sets were shared between the two treatments across both cell lines, and no pathways exhibited opposite regulation between Herceptin and ApoTras. The complete GSEA results are provided in [Additional file 9 Table S21-22](#).

Thirty clusters were formed from the significant gene sets, categorised by shared leading-edge genes, yielding an average similarity score of 0.57 among the clusters (**Figure 4-12**). Only one cluster, Mitochondrial Protein Synthesis, exhibited complete overlap between the treatments, signifying highly consistent regulation within this pathway. Six clusters showed partial overlap, whereas 23 clusters were exclusive to Herceptin treatment in BT474 cells, and three clusters were specific to ApoTras treatment in SKBR3 cells.

Cluster sizes exhibited significant variation, spanning from one to 116 pathways. The largest cluster, Cell Cycle and Mitotic Regulation, comprised 116 gene sets, underscoring its pivotal role in pathway regulation. The smallest cluster, Keratinisation, comprised two pathways linked to cornified envelope formation, indicating a more specialised biological function.

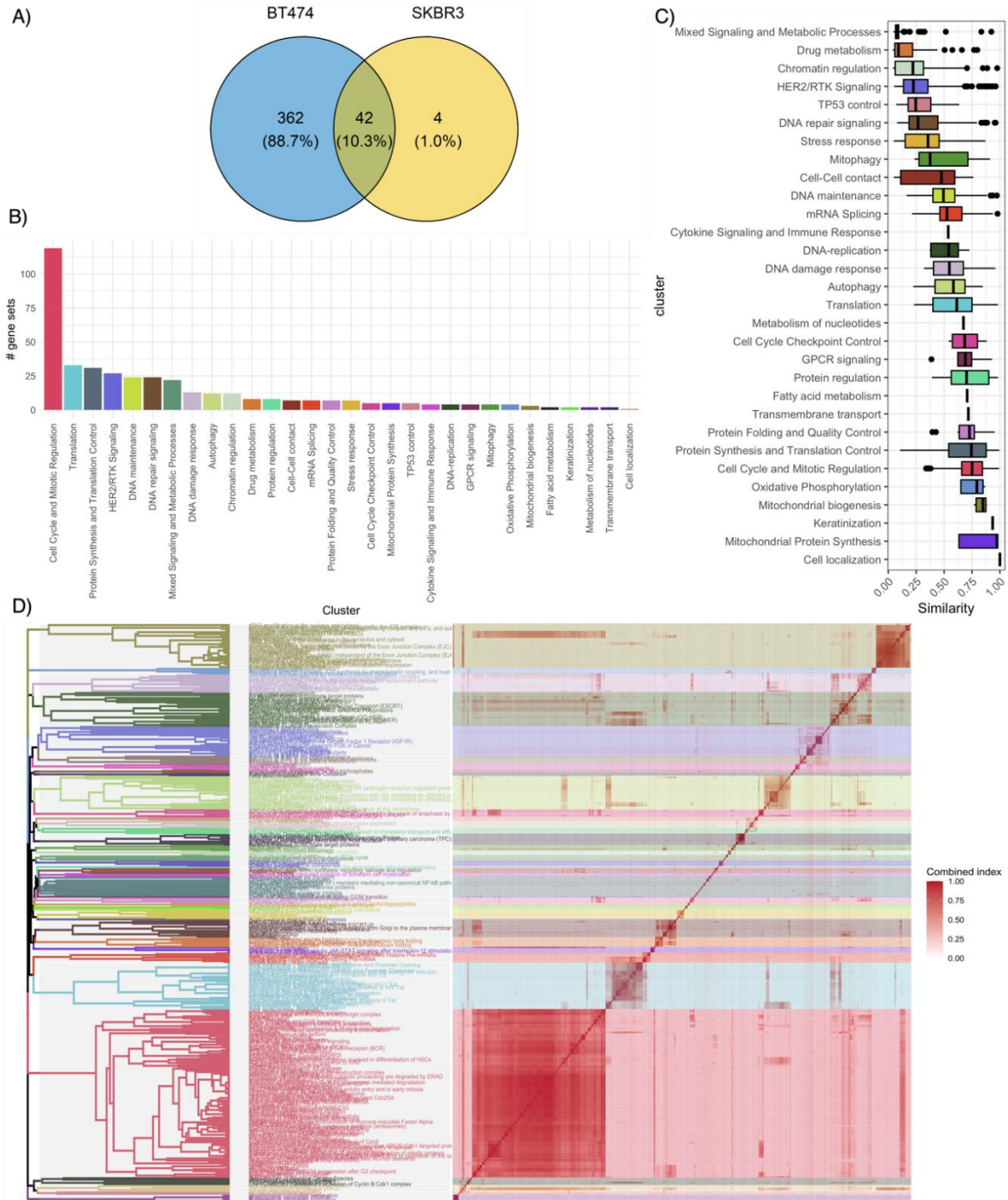


Figure 4-12 Overview of pathway clustering and similarity analysis of Reactome pathways comparing Herceptin against the biosimilar ApoTras, for BT474 and SKBR3 cell lines. A) Venn diagram displaying the overlap and unique significant gene sets identified in BT474 and SKBR3 cell lines. B) Bar plot summarizing the number of gene sets per cluster after hierarchical clustering. Clusters were generated by cutting the dendrogram into 30 groups. C) Boxplot of similarity values (based on shared differentially expressed genes) for each cluster. The median similarity value across clusters is 0.49. (D) Dendrogram showing the hierarchical clustering of pathways on the left and the corresponding heatmap of similarity values on the right. The heatmap visualizes the combined Jaccard index and overlap values, with higher similarity represented by darker colours. Clusters are annotated with distinct colours.

Intersection clusters analyses identified the most relevant biological processes affected by treatment with Herceptin and ApoTras in both BT474 and SKBR3 cell lines (**Figure 4-13, Table 4-2**). Clusters were mainly defined according to the biological function of genes in the clusters, having similar regulation in the two treatments with the two cell lines. The clustering information of pathways can be accessed in [Additional file 9 Table S23](#).

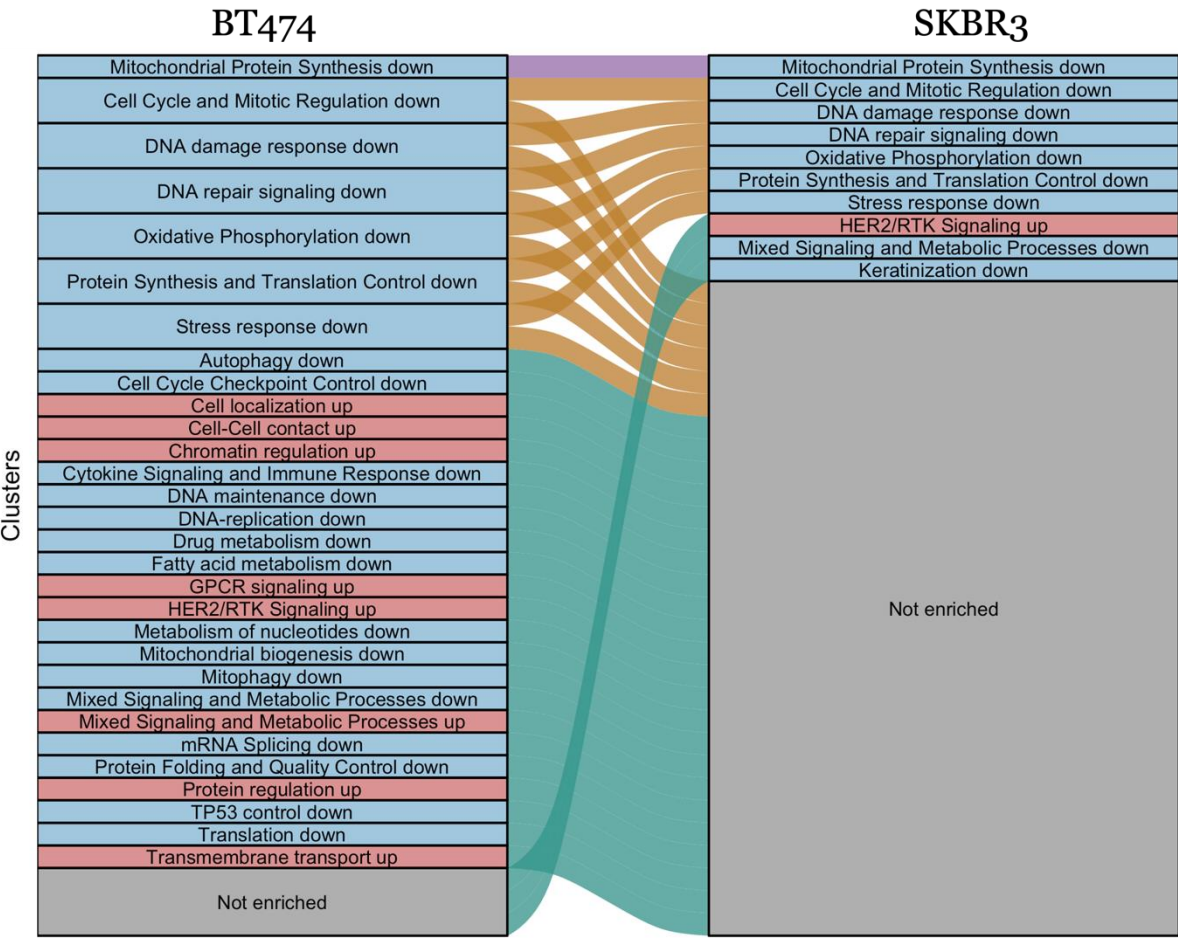


Figure 4-13 Sankey plot illustrating the distribution and overlap of gene set enrichment clusters between BT474 and SKBR3 cell lines in response to Herceptin treatment compared to ApoTras treatment. The Sankey plot visualizes pathway enrichment patterns and their relationships between BT474 (left) and SKBR3 (right) cell lines. Clusters are color-coded to indicate their regulation direction: red for upregulated and blue for downregulated pathways. The connecting flows between clusters demonstrate three types of relationships: purple bands indicate complete overlap of pathways between cell lines, brown bands show partial overlap with connections to "Not enriched" categories, and green bands represent cell line-specific responses.

The analysis of all cell cycle progression and DNA repair-related pathways shows great down regulation in BT474 and SKBR3 cells when treated with Herceptin or ApoTras. The genes in this cluster had reduced expression upon ApoTras treatment, are associated with

the cell cycle progression and named Cell Cycle and Mitotic Regulation cluster. Among annotated genes were PSMD8, PSMB8, PSME1, PSMD14 and PSMD3 in BT474, and PSMA6, PSMB10, CDT1, MCM10 and SKP2 in SKBR3. These clusters are linked by the commonality of the DNA replication pathway to processes necessary for tumour cell proliferation.

DNA Damage Response was downregulated in both cell lines. Among the altered genes downregulated were RPA2, RPA3, PCNA, POLD2 and RPA1 in BT474, and POLD3, POLD2, UBA52, RPA3 and RPS27A in SKBR3. These results are in agreement with defects in gap-filling DNA repair synthesis and ligation during the GG-NER pathway (Marteijn et al., 2014; Moser et al., 2007; Ogi et al., 2010). In particular, BT474 showed a downregulation of genes UBB, RPS27A, UBA52, UBC, and MNAT1, while POLR2H, POLD2, UBA52, RPA3, and RPS27A were observed in SKBR3 in the DNA Repair Signaling cluster. All these alterations mirrored some common processes and, in particular, all referred to Dual Incision in TC-NER and Transcription-Coupled Nucleotide Excision Repair (TC-NER), pointing at DNA repair mechanisms that are affected by ApoTras treatment (Marteijn et al., 2014; Reed & Gillette, 2007; Spivak, 2015). Clusters that were linked with mitochondrial maintenance, activity, and energy production were heavily downregulated in both cell lines. Among these are MRPS14, MRPS5, MRPL15, MRPS35 and MRPL44 in BT474 and MRPS33, MRPL28, MRPS31, MRPL3 and MRPS27 in SKBR3 within the Mitochondrial Protein Synthesis cluster. For example, mitochondrial translation, mitochondrial translation elongation, and mitochondrial translation initiation are all common pathways highlighting the central role of mitochondrial protein synthesis in maintaining cellular energy production (Ott et al., 2016). Cluster of Oxidative Phosphorylation included genes NDUFAF1, NDUFA9, NDUFS7, NDUFAF2, NDUFS3 in BT474; NDUFS6, NDUFB10, COX5A, UQCRB, NDUFS8 in SKBR3 downregulated (Faubert et al., 2020; A. Huang et al., 2025). Common pathways, including eIF and the cap-binding complex activating mRNA as well as eukaryotic translation initiation and eukaryotic translation elongation, highlight the targeting of multiple phases of protein translation. This leads to the inactivation of a complex required for the synthesis of proteins necessary for growth and division of the tumour (Vyas et al., 2016).

The Stress Response cluster exhibited consistent downregulation in both cell lines. In BT474, the genes ATP6V1D, ATP6V0C, ATP6V0B, ATP6V1G1, and ATP6V0E1 exhibited alterations, whereas in SKBR3, the genes HIST1H2AE, RPS15, HIST1H2BG, RPS11, and CCS were downregulated (Feng et al., 2017; G. Li et al., 2019). The common route Cellular responses to stress indicate a diminished capacity to handle treatment-induced stress. Both treatments diminish the survival capacity of tumour cells under therapeutic pressure by impairing mechanisms including autophagy, protein regulation, and mitochondrial function (Kroemer et al., 2010).

Regarding the HER2/RTK signalling cluster, the latter displayed cell line specific results, with no overlapping pathways between BT474 and SKBR3. In BT474, the cluster was upregulated and included pathways such as Constitutive Signaling by Aberrant PI3K in Cancer, FGFR1 mutant receptor activation, IGF1R signalling cascade, and PI3K Cascade that are known to be part of PI3K/AKT and ERBB2 signalling (H. Li et al., 2021). The major mutated genes in BT474 are PIK3R1, PIK3CA, SOS1, KRAS and PIK3C3. In SKBR3 cells, the only altered pathway that was found involved nitric oxide activating guanylate cyclase and included important genes PRKG1, PDE5A, ITPR1, PKP2 and KRT78 (Vasudevan et al., 2009).

In BT474, relevant higher-order cellular processes that were identified by the pertinent clusters were, Cell Cycle and DNA Replication Regulation, Metabolism and Drug Response, and Gene Expression and Protein Synthesis. Among the downregulated pathways that include genes such as ANAPC2, UBE2E1, MCM2, RPA2, those related to DNA replication, cell cycle checkpoint and DNA repair were reported as well (Bartek et al., 2005). Clusters based on metabolism, such as Fatty Acid Metabolism, Drug Metabolism, Mitochondrial Biogenesis, which included genes such as GGT6, MGST2, DECR1 downregulated (Carracedo et al., 2013). Both BT474 and SKBR3 cell lines, after treatment with Herceptin and ApoTras, showed a common molecular modification in the down-regulation of pathways related to cell cycle regulation, DNA damage repair, mitochondrial function, protein synthesis and stress response. This discontinuation experience highlights the possibility that the impact of the biosimilar, ApoTras could vary from that of Herceptin. Particularly interesting are responses of particular cell lines, such as increased activation of HER2/RTK signalling in BT474 and decreased activation of keratinization signals in

SKBR3, indicating that the molecular adaptations to treatment with the biosimilar ApoTras in the different HER2- positive breast cancer models may be unique.

Table 4-2 Gene set enrichment analysis of Reactome pathways comparing Herceptin and ApoTras treatments in BT474 and SKBR3 cell lines. Clusters were grouped based on shared leading-edge genes, and the overlap, direction of regulation, and significant gene sets in each cluster were identified for both cell lines. The table lists the top genes for each cluster, the number of significant gene sets in each cell line and overlapping pathways between the cell lines. Clusters are classified by their overlap type (Complete, BT474-specific, SKBR3-specific, or partial overlap) and the direction of regulation (up-regulated or down-regulated).

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Gene Sets	SKBR3 Top 5 Genes	SKBR3 Significant Gene Sets	Overlapping Pathways
Autophagy	BT474-specific	Down-regulated	TUBA1B, TUBA1C, TUBB2A, TUBB3, TUBB4B	12	-	-	
Cell Cycle and Mitotic Regulation	Partial	Down-regulated	PSMD8, PSMB8, PSME1, PSMD14, PSMD3	119	PSMA6, PSMB10, CDT1, MCM10, SKP2	1	DNA Replication
Cell Cycle Checkpoint Control	BT474-specific	Down-regulated	ANAPC2, UBE2E1, ANAPC15, CDC26, ANAPC7	5	-	-	
Cell localization	BT474-specific	Up-regulated	DYNC1H1, ALMS1, AKAP9, TTBK2, CLASP1	1	-	-	
Cell-Cell contact	BT474-specific	Up-regulated	ITGA2, LAMC1, LAMA3, FN1, ITGAV	7	-	-	
Chromatin regulation	BT474-specific	Up-regulated	TBL1XR1, NCOA1, CREBBP, EP300, NCOR1	12	-	-	
Cytokine Signalling and Immune Response	BT474-specific	Down-regulated	RALA, LCP1, HNRNPF, ANXA2, LMNB1	4	-	-	
DNA damage response	Partial	Down-regulated	RPA2, RPA3, PCNA, POLD2, RPA1	13	POLD3, POLD2, UBA52, RPA3, RPS27A	1	Gap-filling DNA repair synthesis and ligation in GG-NER
DNA maintenance	BT474-specific	Down-regulated	HIST1H2BK, HIST2H2AA3, HIST1H2BD, H2AFZ, H2AFX	24	-	-	
DNA repair signalling	Partial	Down-regulated	UBB, RPS27A, UBA52, UBC, MNAT1	24	POLR2H, POLD2, UBA52, RPA3, RPS27A	3	Dual incision in TC-NER, Gap-filling DNA repair synthesis and ligation in TC-NER, Transcription-Coupled Nucleotide Excision Repair (TC-NER)

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Gene Sets	SKBR3 Top 5 Genes	SKBR3 Significant Gene Sets	Overlapping Pathways
DNA-replication	BT474-specific	Down-regulated	MCM2, MCM7, MCM5, MCM3, CDC45	4	-	-	
Drug metabolism	BT474-specific	Down-regulated	GGT6, MGST2, GSTZ1, MGST3, MGST1	8	-	-	
Fatty acid metabolism	BT474-specific	Down-regulated	DECRI, ACSF2, ACOT13, HADHB, HADH	2	-	-	
GPCR signalling	BT474-specific	Up-regulated	ARHGEF12, AKAP13, SOS1, ITSN1, TRIO	4	-	-	
HER2/RTK Signalling	BT474-specific	Up-regulated	PIK3R1, PIK3CA, SOS1, KRAS, PIK3C3	26	-	-	
HER2/RTK Signalling	SKBR3-specific	Up-regulated	-	-	PRKG1, PDE5A, ITPR1	1	
Keratinization	SKBR3-specific	Down-regulated	-	-	PKP2, KRT8, KRT80, KRT78, KRT7	2	
Metabolism of nucleotides	BT474-specific	Down-regulated	DCTD, RRM1, DCTPP1, NME1-NME2, NME3	2	-	-	
Mitochondrial biogenesis	BT474-specific	Down-regulated	ATP5C1, ATP5L, ATP5A1, ATP5J2, ATP5E	3	-	-	
Mitochondrial Protein Synthesis	Complete	Down-regulated	MRPS14, MRPS5, MRPL15, MRPS35, MRPL44	5	MRPS33, MRPL28, MRPS31, MRPL3, MRPS27	5	Mitochondrial translation, Mitochondrial translation elongation, Mitochondrial translation initiation, Mitochondrial translation termination, Translation
Mitophagy	BT474-specific	Down-regulated	TOMM40, TOMM22, TOMM6, TOMM5, TOMM7	4	-	-	
mRNA Splicing	BT474-specific	Down-regulated	SNRPD3, SNRPG, SNRPB, SNRPF, SNRPE	7	-	-	
Oxidative Phosphorylation	Partial	Down-regulated	NDUFAF1, NDUFA9, NDUFS7, NDUFAF2, NDUFS3	4	NDUFS6, NDUFB10, COX5A, UQCRRB, NDUFS8	1	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.
Protein Folding and Quality Control	BT474-specific	Down-regulated	CCT4, CCT7, CCT6A, TCP1, CCT3	7	-	-	

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Gene Sets	SKBR3 Top 5 Genes	SKBR3 Significant Gene Sets	Overlapping Pathways
Protein regulation	BT474-specific	Up-regulated	NUP214, NUP98, RANBP2, TPR, NUP153	8	-	-	
Protein Synthesis and Translation Control	Partial	Down-regulated	RPS9, RPS6, RPS14, RPS2, RPS29	31	RPS27A, RPS18, RPS27, RPS15A, RPS9	30	Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S, Axon guidance, Cap-dependent Translation Initiation, Eukaryotic Translation Elongation, Eukaryotic Translation Initiation, Eukaryotic Translation Termination, Formation of a pool of free 40S subunits, Formation of the ternary complex, and subsequently, the 43S complex, GTP hydrolysis and joining of the 60S ribosomal subunit, Influenza Infection, Influenza Viral RNA Transcription and Replication, L13a-mediated translational silencing of Ceruloplasmin expression, Major pathway of rRNA processing in the nucleolus and cytosol, Metabolism of amino acids and derivatives, Nervous system development, Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC), Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC), Nonsense-Mediated Decay (NMD), Peptide chain elongation, Regulation of expression of SLITs and ROBOs, Response of EIF2AK4 (GCN2) to amino acid deficiency, Ribosomal scanning and start codon recognition, rRNA processing, rRNA processing in the nucleus and cytosol, Selenoamino acid metabolism, Selenocysteine synthesis, Signalling by ROBO receptors, SRP-dependent cotranslational protein targeting to membrane, Translation initiation complex formation, Viral mRNA Translation

Cluster	Overlap Type	Direction	BT474 Top 5 Genes	BT474 Significant Gene Sets	SKBR3 Top 5 Genes	SKBR3 Significant Gene Sets	Overlapping Pathways
Stress response	Partial	Down-regulated	ATP6V1D, ATP6V0C, ATP6V0B, ATP6V1G1, ATP6V0E1	7	HIST1H2AE, RPS15, HIST1H2BG, RPS11, CCS	1	Cellular responses to stress
TP53 control	BT474-specific	Down-regulated	YWHAЕ, PRDX2, PRDX1, YWHAQ, YWHAH	5	-	-	
Translation	BT474-specific	Down-regulated	POLR2F, POLR2K, POLR2E, POLR2L, POLR2H	33	-	-	
Transmembrane transport	BT474-specific	Up-regulated	SLC7A1, SLC7A2, SLC38A1, SLC7A6, SLC36A1	2	-	-	

4.4 Discussion

4.4.1 Overview of Findings

The analysis of our HER2-positive breast cancer cell lines BT474 and SKBR3 treated with trastuzumab (Herceptin) and its biosimilar (ApoTras) provides insights to treatment response mechanisms, cell line differences, and biosimilarity. It paves the way for the full understanding of cellular response to HER2 inhibition that may enable refined treatment strategies in HER2-positive breast cancers and provides a foundation to assess trastuzumab biosimilars.

BT474 and SKBR3 proved to be useful models, they exhibited high HER2 expression but were otherwise negative for opposite receptor status. BT474 is estrogen and progesterone receptor positive while SKBR3 is receptor negative. Thus, they weakly model different HER2-positive molecular subtypes. Both models were epithelial, and a baseline principal component analysis separated treated samples from untreated samples for both cell lines. BT474 showed a stronger treatment effect compared to SKBR3, while cell cycle-related genes were broadly downregulated in both models. This report sets the baseline parameters for these anticipated experiments and confirms trastuzumab responsiveness on which all these experiments depend.

The next section explores ApoTras vs Herceptin comparisons and examines differential gene expression, pathway activity, and transcription factor changes. BT474 shows stronger pathway activity changes compared to SKBR3, but both cell lines shared cell cycle checkpoint downregulation and FOXO-mediated transcription upregulation, plus SKBR3 unique NFkB, TNFA, and hypoxia pathway activation. Opposite of the differential expression of any factors, FOXO3, FOXO1, AP-1, and TP53 are upregulated in both cell lines, while SKBR3 features unique signals for SP1 and STAT3. The collective findings emphasize the complex panoply of signals provoked by HER2 inhibition, signalling a promising era of personalized treatments.

The biosimilarity assessment confirms high similarity between ApoTras and Herceptin. BT474 cells showed little gene expression difference supporting molecular similarity. Both

treatments broadly impacted hallmark pathway such as cell cycle regulation and apoptosis in similar profiles. ApoTras, comparatively, appeared to enforce stronger repression on proliferation in BT474 through a more severe proteasome and DNA repair pathway downregulation, potentially indicative of deeper cell cycle arrest, but proscribes questions of efficacy.

Cell line specific modifications to ApoTras involve enhanced PI3K/AKT signalling in BT474 and nitric oxide pathway alterations in SKBR3 that may have implications for resistance and metastasis. Indeed, subtle effects, but the translational evaluation confirms the therapeutic promise of ApoTras and now lends support for the use of ApoTras as a trastuzumab replacement.

4.4.2 Cell Line Characteristics and Suitability for Treatment Studies

HER2-positive BT474 and SKBR3 breast cancer cell lines have been extensively used as preclinical models of HER2-driven tumours to define the biology of the disease and assess the response to HER2-targeted therapy, like trastuzumab (Brasó-Maristany et al., 2020; Diermeier-Daucher et al., 2011). Both cell lines have high HER2 expression with logTPM = 10.5, confirming their classification as HER2-positive disease models. Nonetheless, direct comparison of the bulk transcriptomes of BT474 and SKBR3 cells reveals that there are a number of transcripts significantly more or less abundant in one cell line compared to the other (Kokot et al., 2024).

BT474 express both estrogen and progesterone receptors at logTPM of 6.6 and 3.8, respectively, categorizing it as a Luminal B subtype of breast cancer. In contrast, SKBR3 does not express estrogen or progesterone receptors, logTPM of 0 and 0.5, respectively and therefore is a HER2-enriched subtype of breast cancer. The inclusion of BT474 and SKBR3 allows for assessing treatment response in vitro across breast cancer subtypes that are relevant in the clinic, as well as the canonical model system for studying the crosstalk between hormonal signalling and HER2 signalling. Consistent with this classification, both BT474 and SKBR3 express defining epithelial markers, including KRT8, KRT18, KRT19, EPCAM, and OCLN, but not mesenchymal markers VIM, SNAIL, and ZEB1. One notable

exception to this distribution is CDH1, barely expressed in either line, but enriched in BT474 over SKBR3. These established properties of BT474 and SKBR3 together confirm that the lines are representative models of HER2-positive breast cancer but also establish a baseline for describing differences between these two HER2-positive breast cancer models (Sammut et al., 2022).

Furthermore, as has previously been reported both BT474 and SKBR3 cells all showed highly significant trastuzumab-induced down-regulation of cell cycle-related genes, a well-described effect of HER2-targeted therapy (Stanley et al., 2017). ER/PR status, genes involved in cell cycle regulation during the G1, G2/M, and M phases were significantly downregulated, indicating coupling towards cell cycle arrest. However, within each of these phases, BT474 cells also displayed a stronger response to trastuzumab than ER/PR- SKBR3 cells (Kar et al., 2017). Only one prior study has examined the differential effects of HER2-targeted therapy by ER/PR status in vitro, but this result is in line with existing clinical evidence that in cell lines and tumours, both ER/PR- and ER/PR+ HER2+ breast tumours, constitutive HER2 signalling in ER/PR-positive cells is sufficient to drive proliferation. The downstream effect of proliferation further includes predicted treatment-driven negative regulation of genes that are positively correlated with increased mitotic activity, including CCNB1, MKI67, and induction of Luminal-associated genes (e.g. ESR1, PGR) in this experiment's cell lines (Le et al., 2012). The results demonstrate that treatment induces a shift toward the Luminal A subtype, consistent with previous in vivo findings on HER2-targeted therapies, and underscore the clinical implications (Sahlberg et al., 2013).

The utility of BT474 and SKBR3 is further supported by their reproducibility in relation to clinical data such as that of the TCGA breast cancer cohort. Other investigators have demonstrated by comparing these two cell lines, that they exhibit the same molecular characteristics of their patient tumours, most of them being hormone receptor positive, Luminal A and Luminal B tumour subtypes (Witt & Tollefsbol, 2023). Clinically this reaffirms the rationale of using the cell lines for the evaluation of the efficacy of anti-HER2 approaches and personalized medicine. BT474 and SKBR3 are also interesting models for the study of HER2-positive breast cancers taken as a pair; these two lines present both

luminal and HER2-enriched features and, thus, a number of characteristics that can explain the heterogeneity in drug response.

4.4.3 Mechanisms of Action and Cellular Responses to trastuzumab

One important treatment for HER2-positive breast cancer is trastuzumab, which inhibits tumour growth by targeting the overexpressed HER2 receptor. Even though in certain cases it can be very beneficial, resistance primarily remains a challenge: 65-75% of HER2-positive metastatic cases are non-responsive. The mechanisms of action of trastuzumab are examined in this study, with particular attention paid to the BT474 and SKBR3 cell lines' cell cycle regulation, DNA repair, metabolism, differentiation, immune modulation, and cell-specific adaptations (Kokot et al., 2024).

Cell Cycle Regulation

Trastuzumab stops the growth of HER2-positive cancer cells by causing cell cycle arrest at the G1/S and G2/M checkpoints (Brasó-Maristany et al., 2020). It blocks cell cycle progression by negatively regulating CDK1, CDK2, CCNE1, and CCNB1 among others (Stanley et al., 2017). As BT474 cells are dependent on HER2 signalling and hormone receptor positive, they are more responsive than SKBR3 cells. One example of this is exhibited by more pronounced downregulation of proliferation markers including CCNB1 and MKI67 in BT474 cells (Sabbaghi et al., 2017).

Trastuzumab also disrupts mitosis by suppressing PLK1 and BUB1, leading to G2/M arrest and potentially mitotic catastrophe (Lewis Phillips et al., 2008). Centromere proteins CENP-A and CENP-E are also downregulated, abrogating chromosomal segregation and inhibition of spindle assembly (Upton et al., 2021). This multilevel inhibition underlines the ability of trastuzumab to abrogate proliferation and the possibility of associating trastuzumab with mitosis-targeted agents to increase effectiveness.

Mechanisms of DNA Repair

Since trastuzumab leads to the inhibition of DNA repair mechanisms, cancerous cells are particularly vulnerable to an increased level of genomic instability (Le et al., 2005). BRCA1,

BRCA2, and RAD51, key repair genes that would be repressed in all the cell lines, are not detectably repressed in BT474, despite strong repression of homologous recombination (Sahlberg et al., 2013). This is in line with findings that such patients are more sensitive to HER2-directed therapies. The fact that the repair genes are not as down regulated in SKBR3 cells suggest SKBR3 cells might be showing some adaptation. This is likely related to the activation of the DNA damage response, as inferred by high activity of p53 and increased expression of their targets, including CDKN1A (p21) in both cell lines (Zazo et al., 2016). However, prevention of the vital repair pathway will prevent DDR from stabilizing when necessary, leading to mitotic catastrophe or apoptosis.

Metabolic Changes

Both cell lines undergo metabolic reprogramming induced by trastuzumab with a reduction of mitochondrial oxidative phosphorylation (OXPHOS) and increase of glycolysis (Lv et al., 2022). ALDOA, PGK1, and GAPDH are among the glycolytic enzymes that are increased, especially in SKBR3, which also shows more pronounced mitochondrial downregulation. Increased expression of sphingolipid metabolism genes such as ASAH1 and CTSA indicates that lipid metabolism pathways also adapt, indicating a survival strategy under HER2 suppression (Das et al., 2024). Because BT474 cells are more sensitive to trastuzumab, they exhibit a greater activation of glycolytic pathways. Although the metabolic adaptations vary, indicating cell-specific responses, FOXO transcription factors, which are active in both cell lines, may be controlling these changes (Maadi et al., 2021).

Cellular Differentiation Changes

Among the effects of trastuzumab there is also the induction of keratinisation and of other differentiation-related functions that alter the aggressiveness of HER2- positive cancer cells (Rediti et al., 2024). Overexpression of epithelial markers like KRT15 and KLK5 suggest that differentiation is increased in BT474 cells. SKBR3 cells, on the other hand, exhibit both keratinisation and the expression of genes linked to EMT, namely KRT23, underscoring their mixed differentiation and potential for invasion (Khateb et al., 2024). The SKBR3 line maintains its adaptive and aggressive characteristics, while the BT474 line assumes a more

stable phenotype, a pattern which is consistent with the behavior of HER2-positive tumors in general (Kar et al., 2017).

Cellular Communication and Immune Modulation

In addition to the HER2 pathway itself, trastuzumab, as an antibody, modulates immune responses and the microenvironment influenced by the tumour. These are changes that may offer some advantage in promoting growth, evading the immune system, or developing resistance. The analysis also showed that cell migration and motility were the biological processes associated with the upregulation of genes found in trastuzumab-treated BT474 and SKBR3 cells. Such is the case with MYL12A, a functional gene involved in the regulation of the actin cytoskeleton and contractility, which is upregulated in SKBR3 cells, and FXYD3, a functional ion transport and epithelial integrity gene that is induced in BT474 cells. Considering that these genes are upregulated, they may contribute to enhanced motility and possibly even invasive capability as a sort of ‘compensatory’ mechanism in cancer cells exposed to therapeutic stress.

These migratory dynamics are associated with a cluster of upregulated integrin-related genes, including ITGB1, which is able to promote cell–ECM adhesion in SKBR3 cells. This is in line with increased integrin-mediated signaling, which has previously been implicated in migration, invasion, and survival in the context of anti-HER2 therapy (Arnould et al., 2006). These findings indicate that SKBR3 cells may be more reactive in their ability to reorganize the microenvironment to support proliferation and escape from targeted therapy.

More importantly, both cell lines have different immune signaling profiles, revealing the different ways in which cell type-specific responses can be modulated. The JAK-STAT pathway is not yet fully characterized but plays a role in cytokine responses, cell proliferation, and immune signaling, and might be active in SKBR3 cells. In contrast, BT474 cells exhibit activation of the NF- κ B signaling pathway, a master regulator of inflammatory signaling and anti-apoptotic responses that promote tumor-associated inflammation. Among the multiple immune and inflammatory pathways that, once activated during inflammation, may cause constitutive immune signaling and treatment resistance, NF- κ B activation is a well-known driver of resistance in breast cancer.

The second element common to both cell lines is represented by the upregulation of FOXO family transcription factors. Due to their ability to render cells resistant to chemo- and radiotherapy, FOXO proteins can be considered strong candidates to support tumor survival under stress stimuli exerted by therapies (Webb and Brunet, 2016). Both lines also exhibit an upregulation of CD44, the cell adhesion molecule that has been widely associated with roles in cancer stemness, immune evasion, and therapy resistance (Vivekanandhan and Knutson, 2022). CD44 also interacts with the ECM and can transmit anti-apoptotic signals to the cell, enabling anchorage-independent growth—a requirement for the metastatic process.

According to these findings, the effect of trastuzumab on the behavior of HER2+ cancer cells appears to be strongly linked to modifications in immune response, innate signaling, and the tumor microenvironment. An immunomodulatory balance between immune activation and agonist mechanisms versus evasion and antagonist factors (e.g., CD44, FOXO) may contribute to this complex behavior. These data support the plasticity of luminal HER2+ cells to adopt immune-evasive and motile phenotypes, as almost all TGF- β signature genes driving metastasis are upregulated in these cells treated in vitro with Trastuzumab-Emtansin, even under the influence of an antitumoral immune response. These findings suggest the necessity of multi-drug treatment strategies, including the inhibition not only of HER2 but also of compensatory pro-survival and immune-modulating mechanisms. Future strategies could involve combinations including JAK-STAT and NF- κ B inhibitors, anti-CD44 therapy, or FOXO-targeting agents to increase therapeutic efficacy and reduce the risk of resistance.

Cell Stress and Apoptosis

Trastuzumab promotes the activation of stress responses and regulates apoptotic cell death (Vu & Claret, 2012). Both cell lines show activation of autophagy associated genes such as MAP1LC3B, with SKBR3 cells boasting higher activation and suggesting adaptive resistance. Despite the downregulation of critical apoptosis regulators such as CASP3, SKBR3 cells have a strong apoptotic response while BT474 cells exhibit an even stronger apoptotic responses with activation of caspases, potentially enhancing treatment efficacy.

Activation of FOXO1 and FOXO3 in both cell lines elicits apoptosis and oxidative stress, but SKBR3 compensates with other pathways such as JAK-STAT and NF- κ B (Goutsouliak et al., 2020). These results suggest a potential for combined trastuzumab with apoptosis inducers or autophagy inhibitors for therapeutic purposes.

Protein Trafficking and Transport

Trastuzumab modifies protein trafficking and transport via increasing RAB GTPases, which improves HER2 internalisation and degradation (Pohlmann et al., 2009). Both cell lines activate caveolin mediated endocytosis, but SKBR3 cells also upregulate the recycling receptor genes NEDD4L, which perhaps helps to maintain the HER2 signalling. Since BT474 cells are more responsive to trastuzumab, they also present less receptor recycling. These findings suggest that targeting receptor trafficking pathways might enhance the efficacy of trastuzumab and counteract resistance mechanisms of SKBR3.

Angiogenesis

In both cell lines, trastuzumab acts through PI3K/MAPK and VEGF signalling to block pro-angiogenic pathways of action, which impact angiogenesis (Swain et al., 2023). In response to stress, SKBR3 upregulates hypoxia-related genes like HIF1A, enabling resistance to therapy in part by promoting angiogenesis. Regarding this, and related to its enhanced responsiveness, BT474 consistently down-regulates angiogenesis-related pathways. For breast tumours, for example trastuzumab in combination with anti-angiogenic agents or inhibitors of the hypoxia pathway might be able to target compensatory mechanisms in patients with a phenotype similar to SKBR3 and improve outcomes.

Response Specific to Cell Type

BT474 cells were shown to respond more effectively to trastuzumab treatment showing a clear cell cycle arrest, blocking of DNA repair pathways and promotion of differentiation into an epithelial like phenotype. In contrast, SKBR3 has evolved means allowing resistance and invasive capabilities, such as angiogenesis, EMT and JAK-STAT activation (Kokot et al., 2024). For example, anti-angiogenic therapy can be used for SKBR3-like tumours, while

trastuzumab in combination with DNA repair inhibitors could be used for BT474-like cancers. Such combinations could in principle increase the efficacy of trastuzumab and reduce the magnitude of non-responders.

Integration of Mechanisms

The actions of trastuzumab on cell cycle control, DNA repair, and modulation of metabolism and the immune system are interrelated. BT474 cells are also receptor dependent on HER2, and trastuzumab exerts a potent cytostatic and cytotoxic activity on these cells. SKBR3 cells, which adapt with pathways such as hypoxia signalling and EMT, show only partial resistance and so the need for combination therapies with additional agents (Maadi et al., 2021).

Clinical Implications

Dissimilar responses of BT474 and SKBR3 to trastuzumab emphasizes the need for personalized therapy. Other combination strategies that can potentially lead to better outcomes include the use of immunomodulators, anti-angiogenic agents, and DNA repair inhibitors. By characterizing signalling, metabolic, and immune adaptations amongst HER2 positive tumours, personalized approaches to circumvent resistance and enhance efficacy will be possible (Goutsouliak et al., 2020).

4.4.4 Biosimilarity Assessment

The comparison between ApoTras and Herceptin show that they are both molecularly and functionally very similar, with slight differences in mechanisms. The combination of distinct gene expressions, pathway regulation, transcription factor activity, and categories found in the Reactome gene set enrichment are compelling evidence of the biosimilarity among these drugs.

DEG Analysis

Gene expression signature analysis demonstrates almost complete molecular concordance with only five genes differentially expressed between ApoTras- treated and Her- treated BT474 cells. Among these are ankyrin repeat domain containing family members POTE,

known to play a role in cell signal transduction and cell proliferation. For instance, the up-regulation of POTEE and POTEF in ApoTras- treated cells could be consistent with subtle effects on the cell cycle control, and the increased expression of ANKRD36C in the ApoTras treated cells might be indicative of less obvious effects on signalling, but the actual involvement of the protein in HER2 positive cancer is not known as of yet. Further analysis of the roles played by transcriptional factors highlighted ZFHX4, responsible for cell differentiation and EMT, another factor whose expression was enhanced in ApoTras treated cells, whose possible role in cellular plasticity remained unknown. But, given the scarcity of the DEGs, the overall molecular similarity between treatments is noticeable, supporting the use of ApoTras as a good alternative to trastuzumab.

Impairment of Proliferation and DNA Repair

Both ApoTras and trastuzumab can effectively induce cell cycle arrest through the inhibition of CDKs and cyclins, including CDK1, CDK2, CCNB1, and CCND3, blocking progression at the G1/S and G2/M checkpoints (Kokot et al., 2024). The lower Ki-67 expression in BT474 and SKBR3 cells confirms this anti-proliferative effect, that in BT474 cells is more pronounced, being strictly related to HER2 signalling action. Also, both therapies downregulate homologous recombination genes such as BRCA1, BRCA2, and RAD51, which in turn further negatively affects the DNA repair capability of cancer cells and makes the cancer more susceptible to DNA damage by other mechanisms. Only minor disruption of the repair pathways is present in SKBR3 cells, with few γ -H2AX foci, while BT474 display high levels of DNA damage using an increased number of γ -H2AX. Finer differences such as the slightly elevated DNA damage response observed in ApoTras-treated BT474 cells, will have to be further explored to elucidate the real biological implications of ApoTras treatment.

Increased Risk for Resistance Development

In addition, both treatments have commonalities in constructs that are likely related to therapeutic resistance as well. First, the levels of EGFR, ESR1 and AR themselves can be up-regulated to allow signals through different pathways having pro-survival and pro-proliferative functions (Cortese et al., 2023). Activated EGFR forms heterodimers with

HER2 that may be capable of cellular signalling even in the existence of a HER2 blockade. Hormonal signalling also plays a role in the resistance to HER2 targeting (Mayer, 2015). For instance, ESR1 and AR are described to be involved in both survival and proliferation when HER2 signalling is lost by the cells.

ApoTras also had different transcriptional effects as compared to Her. As an example, PIK3CD was the most upregulated downstream molecule after ApoTras treatment of BT474 cells, and AKT, downstream of PI3K, was the only predicted activated upstream regulator (Duro-Sánchez et al., 2023). PIK3R3 was the one showing the highest increase in SKBR3 cells and was also predicted to further active the PI3K/AKT signalling cascade. Accordingly, the PI3K/AKT signal transduction was significantly higher with ApoTras than with Her in BT474 cells (Uifălean et al., 2018). These results imply that ApoTras required more activation of the PI3K/AKT pathway as a response to the treatment with both drugs in BT474 cells, and the inhibition of this pathway might have been the mechanism through which neratinib increased Apo cytotoxicity. Also, the SKBR3 cells treated with ApoTras exhibited activation of the Epithelial-to-Mesenchymal Transition pathway (in line with a partial EMT) and downregulation of keratinization signalling. These findings pointed out that ApoTras treated SKBR3 cells would be somehow primed to undergo phenotypic changes associated with enhanced invasiveness.

HER2/RTK Signalling Adaptations

ApoTras and Her both antagonize PI3K/AKT and RAS/MAPK signalling cascades downstream of HER2, and concomitantly increase HER2 turnover, by acting upstream of the receptor (Vu & Claret, 2012). The ApoTras and Her ultimately produce their clinical effect via these cross-coupling pathways. However, ApoTras also activates a distinctive repertoire of pathway adaptation responses, for instance hyper PI3K/AKT signalling in BT474 cells, that could eventually benefit the cell through an alternative escape route. Nitric oxide and calcium signalling are two additional pathways affected by the ApoTras, but not trastuzumab in SKBR3 cells. These latter two processes alter the tumour's ability to adhere and migrate. These data show subtle differences between ApoTras and Her treatment in their

effects on the HER2 pathway. While this argument does not challenge biosimilarity, it notes how ApoTras and Her likely exert slightly variable effects in given tumours.

Cell Type-Specific Responses

The responses of HER2-positive breast cancer cell lines BT474 and SKBR3 to ApoTras and trastuzumab reveal potential differences tied to their distinct molecular profiles (Dong et al., 2021). First, since BT474 cells are extremely HER2-dependent cell lines, the sensitivity to both treatments is higher in BT474 cells. Other examples associated to cell cycle such as CCNB1, and DNA repair such as BRCA1 are significantly lower in expression. On the contrary, the expression of nitric oxide signalling related genes is upregulated, and genes related to EMT are downregulated in response to ApoTras in SKBR cells. Loss of EMT genes was associated with higher mesenchymal phenotype on SKBR3 that could be activating this program as a compensatory resistance mechanism. Such differences show indeed the heterogeneity of HER2-positive breast cancer and, although the two treatments perform effectively on a global basis, it warrants further investigation into the specifics of the adaptive responses.

4.4.5 Research Implications

This study demonstrated the utility of RNA-sequencing data in the evaluation of biosimilarity. Tumour heterogeneity, such as the distinct cellular responses of SKBR3 and BT474 to trastuzumab, should be considered when developing individualized treatment strategies (Kokot et al., 2024). BT474 were extremely sensitive to trastuzumab at the transcriptional level, downregulating cell cycle regulators and DNA repair genes due to the high reliance on HER2 signalling (Uifălean et al., 2018). Likewise, tumours with molecular profiles similar to BT474 are expected to benefit from combination therapies with DNA-damaging agents or PARP inhibitors.

The biosimilarity comparison study of ApoTras and the reference product Herceptin indicated a high similarity in all analyses conducted, with slight variances (Niazi, 2023). ApoTras tended to enhance the inhibition of a few genes involved in the cell cycle and proteasome activity, which might act as stronger signals for cell cycle arrest than the

originator product. This stronger inhibition could also have unanticipated consequences in regard to downstream signalling. The findings support the use of molecular profiling as a robust approach for assessing biosimilarity (Center for Drug Evaluation and Research, 2015). This enables informed follow-up experiments to evaluate whether the identified differences have a meaningful biological impact (Xie et al., 2020). This approach is especially helpful for comparing drugs like trastuzumab with broad effects on target cells that cannot be captured by conventional experiments.

Chapter 5: Multi-omics data analysis to understand the phenotypic differences between Herceptin and biosimilar treatment

5.1 Introduction

The expiration of patents for biologics like trastuzumab (Herceptin®) has made possible the development of biosimilar monoclonal antibodies which are expected to increase accessibility and reduce costs for cancer treatment (Uifălean et al., 2018). Biosimilars must undergo rigorous analytical and clinical evaluations to demonstrate comparable efficacy and safety to their reference biologics, despite being structurally and functionally similar (Mascarenhas-Melo et al., 2024). This shift brings benefits to patients, cost and accessibility being chief among them. One such biosimilar, Apo-Trastuzumab (Apobiologix; ApoTras), was recently designed to be bioequivalent to Herceptin, and their mechanisms of action are compared here using advanced multi-omics approaches.

The monoclonal antibody against human epidermal growth factor receptor 2 (HER2) trastuzumab plays a major role in treating approximately 20-30% of all diagnosed breast cancers (Maadi et al., 2021). By binding the extracellular domain of HER2, trastuzumab sharply reduces cell growth and division through several mechanisms (Baselga & Albanell, 2001). Importantly, binding prevents receptor dimerization, which blocks multiple downstream signalling pathways including PI3K/AKT and MAPK, which are responsible for survival and proliferation signalling (DrugBank, 2023).

The effects of trastuzumab on cancer cells are varied and diverse, with cell cycle arrest as one of its primary mechanisms (Maadi et al., 2021). Trastuzumab triggers G1-phase arrest by downregulating p21/WAF1, and modulating p27Kip1 phosphorylation (Maadi et al., 2021). At the same time, trastuzumab induces apoptosis by reducing levels of survivin, a protein that protects cancer cells from programmed cell death. This double strike is a major impediment to tumour growth.

Furthermore, DNA integrity and angiogenesis are inhibited by trastuzumab. By promoting G1-phase exit, leading to DNA damage accumulation in cancer cells, compromising their genomic stability (Maadi et al., 2021). Trastuzumab also exhibits anti-

angiogenic activity through a reduction in VEGF secretion, which inhibits new blood vessel formation necessary for tumour growth and metastasis (Izumi et al., 2002).

At the molecular level, trastuzumab has been suggested to trigger the internalization and degradation of HER2 receptors, although the precise mechanism remains incompletely understood. Studies have also shown that trastuzumab blocks Src signalling, leading to increased levels and activity of PTEN (Phosphatase and tensin homolog), which in turn suppresses the PI3K/Akt signalling pathway, reinforcing the anti-tumour effects (Tortora, 2011).

These diverse mechanisms of action cumulatively contribute to the clinical efficacy of trastuzumab in treating HER2-positive breast cancers, but despite its success, the full scope of downstream transcriptional and proteomic changes inflicted by trastuzumab remains incompletely understood (Albanell et al., 2003). To fill this void and to fully characterize and comprehensively compare trastuzumab with its biosimilar counterpart, ApoTras, a multi-omics approach integrating of transcriptomics and proteomics is essential (Menyhárt & Györffy, 2021).

Transcriptomics and proteomics offer complementary insights into cellular processes, each with specific advantages and limitations (Hasin et al., 2017). RNA-sequencing (RNA-seq) provides high sensitivity for detecting differential gene expression and offers a comprehensive view of the transcriptome. RNA-seq benefits from established bioinformatics pipelines and is relatively cost-effective. However, RNA-seq data may not accurately reflect protein abundance due to post-transcriptional regulation and can be sensitive to technical noise (Menyhárt & Györffy, 2021).

Proteomics, on the other hand, directly measures protein levels, capturing post-translational modifications that are crucial for understanding cellular function (Wahjudi et al., 2021). Proteomics provides a more accurate representation of the functional state of cells and can detect changes in protein abundance that may not be detectable at the mRNA level (Hasin et al., 2017). However, proteomics typically has lower coverage compared to transcriptomics, especially for low-abundance proteins, and it is often more technically challenging and costly (Menyhárt & Györffy, 2021).

The integration of these two approaches can partially solve their individual limitations and provide a more complete understanding of cellular processes (Raufaste-Cazavieille et al., 2022). Multi-omics integration enables the identification of similar changes across both molecular levels. This will increase the confidence in the biological importance of the observed changes and discover mechanisms that might not be detectable from either dataset alone (Perna et al., 2017).

Transcriptomic and proteomic analyses have been used to study the effects of trastuzumab in HER2-positive cell lines, but most studies select one of the omics layers (Magnifico et al., 2009). That reduces the reliability of the data by which the therapeutic effects can be investigated. Integration of transcriptomic and proteomic data increases confidence in findings, identifying commonly affected pathways at gene and protein levels (Perna et al., 2017).

This study establishes a shared pathway profile from transcriptomic and proteomic datasets to establish a robust core mechanism of action (MOA) shared by Herceptin and biosimilar ApoTras for the cell lines SKBR3, BT474 and MCF7. The integration of the findings from these omics' layers will further the understanding of the therapeutic similarity between the reference product Herceptin and its biosimilar ApoTras. Additionally, a relevant set of key pathways can be defined for the response of HER2-positive breast cancer to trastuzumab. This method will investigate the molecular basis of trastuzumab's efficacy and confirm the biosimilarity of ApoTras.

5.2 Materials and Methods

5.2.1 Transcriptomics Data

The RNA-seq data for BT474 and SKBR3 are available in the NCBI Gene Expression Omnibus (GEO) under accession numbers GSE158969 and GSE284388, respectively. The cell culture methods and treatments are described in Chapter 4, section 4.2.1. The RNA-extraction and sequencing methods are described in Chapter 4, section 4.2.2.. In this Chapter, the results of the differential gene expression analysis from Chapter 4 section 4.2.4.1, for the comparisons of ApoTras vs Her and ApoTras & Her vs Con, were used for gene set enrichment analysis.

5.2.2 Proteomics data

5.2.2.1 Cell Culture and Treatment

The three cell lines (BT474, SKBR3 and MCF7) were cultured and treated as described in Chapter 4, section 4.2.1.. MCF7 cells do not express HER2 and served as a negative control. For each cell line, four control samples, three Apotex-trastuzumab (ApoTras) treated, and three Herceptin (Her) treated samples were collected on Day 7.

5.2.2.2 Protein extraction and TMT Labelling

Protein extraction was performed using Pierce™ Mass Spectrometry Sample preparation kits (Thermo Fisher Scientific, Catalog #84840) following to the manufacturer's protocols for cultured cells. Briefly, the cells were lysed with the provided cell lysis buffer, followed by sonication to ensure thorough disruption. Subsequently, 100 µg of lysate protein per sample was reduced with 1,5-Dithiothreitol (DTT) and alkylated with iodoacetamide (IAA) followed by acetone precipitation to isolate the proteins. The proteins were initially digested with Lys-C for 2 h, followed by overnight trypsin digestion for 10 h. The peptide concentration was determined in triplicate using the Pierce™ Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific, Catalog #23275). For each sample, 80 µg was transferred and dried with Thermo Scientific™ SpeedVac™ Vacuum Concentrator. The

peptides were resuspended in 100 μ L of 100 mM TEAB pH 8.5 for Tandem Mass Tags (TMT) labeling (Thermo Fisher Scientific, Catalog #90110). TMT labeling was performed according to the manufacturer's protocols to enable multiplexed proteomic analysis. The labeled samples were pooled and dried under vacuum. Excess TMT label was neutralized using 8% ammonium hydroxide, after which the TMT-labeled samples were pooled in a 1:1:1:1:1:1:1:1:1 ratio, vacuum-dried, and stored at -80°C . The samples were shipped on dry ice to SPARC Biocentre, SickKids (Toronto, Canada) for mass spectrometric analysis.

5.2.2.3 Liquid Chromatography Tandem Mass spectrometry Analysis

The proteomic analysis was conducted at the SPARC BioCentre using an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific, Rockford, IL). Briefly, peptides were lyophilized and resuspended in 0.1% TFA. The liquid chromatography (LC) gradient was run over a total duration of 60 minutes, with the following solvent system:

- Mobile Phase A: 0.1% formic acid in water.
- Mobile Phase B: 80% acetonitrile with 0.1% formic acid.

The gradient profile included:

1. 5%–30% solvent B over 45 minutes
2. 30%–50% solvent B over 5 minutes
3. 50%–95% solvent B for column washing over 5 minutes
4. 5-minute re-equilibration phase

Mass spectrometry was performed using the Orbitrap Fusion Lumos instrument in data dependent acquisition (DDA) mode. The MS1 spectra were acquired in the Orbitrap at a resolution of 120,000, covering a mass range of 375–1,500 m/z , with an AGC target of 4×10^5 and a maximum injection time of 50 milliseconds.

Fragmentation of precursor ions was performed using Higher-energy C-trap Dissociation (HCD), enabling the detection of reporter ions from Tandem Mass Tags (TMT) in the low m/z range. The MS2 spectra, used for TMT reporter quantification, were acquired at a resolution of 50,000, with a fixed first mass of 100 m/z , an AGC target of 1×10^5 , and a

maximum injection time of 100 milliseconds. The isolation window for MS3 was set to 0.7–1.0 Da, and the HCD collision energy was approximately 35%. The resulting RAW files were shared for analysis and can be accessed in [Additional file 1-3](#).

5.2.3 Data analysis

5.2.3.1 Proteomics Data

5.2.3.1.1 Preprocessing

Peptide and protein identification was performed with the open-source Trans-Proteomic Pipeline (TPP) (Deutsch et al., 2023) for MS data processing. ProteoWizard was used to convert the RAW data files to mzML. Protein database searches were conducted with Comet (Eng et al., 2015) and the TPP-specific X!Tandem (Craig & Beavis, 2004). The fragment lists were searched against the human UniProt Reference database (downloaded Aug 23, 2019) (Apweiler, 2008), including randomized decoy peptides created with the Trans-Proteomic Pipeline (TPP) (Deutsch et al., 2023) decoy generation script. Both search engines were configured with trypsin as the digest enzyme, allowing for one missed cleavage per peptide. Additionally, Methionine oxidation was set as a variable modification, while carbamidomethylation of cysteine and the TMT balance tag on N-terminal and lysine residues were set as static modifications. Searches were performed with a mass tolerance of 10 ppm for parent mass, and 0.6 Da for fragment mass. The complete search engine parameters can be accessed at [Additional file 4-5](#). PeptideProphet (Ma et al., 2012) and iProphet (Shteynberg et al., 2011) were used to assign confidence scores to peptide identifications in each experiment. ProteinProphet (Deutsch et al., 2023) and was used to assign protein identifications and groupings with confidence scores from the results of Comet (E(Craig & Beavis, 2004)X!Tandem (Craig & Beavis, 2004)). The protein identifications were filtered for an estimated false discovery rate (FDR) of 1% based on the detected decoy peptides.

5.2.3.1.2 Differentially Expressed Protein (DEP) Analysis

TMT reporter ion intensities for each peptide-spectrum match (PSM) were quantified using the TPP Libra module (Deutsch et al., 2023) during each mass spectrometry run. The complete set of PSMs, including confidence scores, raw reporter ion intensities, and the corresponding protein IDs, was exported from the TPP web interface as a text file. These files, along with the list of contaminants, can be found in [Additional file 6-15](#). Protein IDs from all protein groups were initially filtered to retain those with a false discovery rate (FDR) of 1% or lower. Proteins were then further filtered to exclude randomized decoy sequences, contaminants, and trastuzumab-related proteins (UniProt IDs: P01834 (Immunoglobulin kappa constant), P01857 (Immunoglobulin heavy constant gamma 1)).

For each mass spectrometry run, PSMs with fewer than two unique TMT reporter ions were removed, and proteins with fewer than two unique peptide sequence were excluded. Reporter ion intensities for all peptides assigned to the same protein were summed to calculate the total intensity per TMT tag and protein. The summed peptide intensities were then normalized using variance-stabilizing normalization with the `justvs` (Huber et al., 2002) function from the `limma` package (Ritchie et al., 2015; Smyth, 2005). Proteins without assigned gene symbols were excluded from downstream analyses. Differential expression analysis was performed using the `limma` (Ritchie et al., 2015; Smyth, 2005) package, with empirical Bayes methods applied to identify differentially expressed proteins. Proteins were considered significantly altered in expression at an FDR threshold of 5%.

5.2.3.1.3 Principal Component Analysis

Principal component analysis (PCA) was conducted as described in Chapter 4, section 4.2.4.2 using $\log_2$ VSN-normalized intensities as input.

5.2.3.1.4 Pearson Correlation Calculation of Housekeeping Genes

Pearson correlation analysis of housekeeping genes (Eisenberg & Levanon, 2013) was conducted following the methodology outlined in Chapter 2, section 2.2.4.3. Results were visualized using the `ComplexHeatmap v2` (Z. Gu et al., 2016) package in R.

5.2.3.1.5 Subcellular protein location

The information about subcellular locations of proteins was derived from Thul and colleagues (Thul et al., 2017).

5.2.3.2 Combined Proteomics and Transcriptomics Analysis

5.2.3.2.1 Gene Set Enrichment Analysis (GSEA)

The GSEA analysis was performed with the GSEA v4.0.3 (A. Subramanian et al., 2005) software. For this analysis, the genes were ranked according to the formula described in Chapter 4, section 4.2.5.1 using DEG and DEP analysis results as input. FDR thresholds are indicated in the result section as needed. The GSEA analysis was performed for Gene Ontology Biological Processes gene sets (Gene Ontology Consortium, 2004).

5.2.3.2.2 Shared Pathway Network

Cytoscape v3.1.3 (Shannon et al., 2003; Smoot et al., 2011) was used to identify the overlap in leading-edge genes of significant pathways to visualize the resulting network. The overlap was estimated using a combined score of overlap and Jaccard index, with both metrics given equal weight. The leading-edge genes were downloaded for each identified cluster.

5.2.3.2.3 Hub Gene Identification of Gene Set Groups

Hub gene identification was performed by selecting the leading-edge genes for each gene set group from GSEA analysis and using STRINGdb R package to determine network information. The degree of connectivity was then determined with igraph to find the top 5 most connected genes.

5.2.4 Visualization

Unless stated otherwise, visualizations were generated using ggplot2 v3.5.1 (Ginestet, 2011) in Rstudio (RStudio Team, 2016). Venn diagrams were produced using the VennDiagram package v1.7.3. UpSet plots were created with the package UpsetR v1.1.0 (Conway et al., 2017).

5.2.5 Additional Files and Code Availability

Additional files and scripts for preprocessing, analysis and figure creation are available in the GitHub repository at https://github.com/karstenrinas/Thesis/Chapter_5.

5.3 Results

5.3.1 Proteomics data analysis

5.3.1.1 Technical Quality Control

Three proteomics experiments were performed, one for each cell line: BT474, SKBR3, and MCF7. Each experiment included four control replicates, three samples treated with Apotex-trastuzumab (ApoTras), and three samples treated with Herceptin (Her). The database search for the three experiments identified 3914 proteins for the BT474 dataset, 3054 for the SKBR3 datasets, and 1140 proteins for MCF7 with at least one unique peptide and an estimated FDR of 1%. BT474 and SKBR3 datasets showed similar total Peptide-Spectrum Matches (PSM) numbers per run and identified peptides whereas MCF7 showed significantly fewer identified proteins, total PSMs and total peptides. However, the number of peptides per protein and protein coverage range is not significantly worse. The summarized data are presented in Table 5-1.

Table 5-1 Proteomics Quality Control Metrics for three 10xTMT-Labeled shotgun proteomics data sets of treated BT474, SKBR3, and MCF7 cell lines. PSMs (Peptide-Spectrum Matches): The number of times a peptide is identified by matching tandem mass spectra to the peptide sequence. Total Peptides: Number of uniquely identified peptides. Protein coverage: The percentage of a protein's amino acid sequence that is covered by identified peptides. Peptides: Number of identified peptides for a protein.

Cell line	Quantified Proteins	Total PSMs	PSMs Range; Median	total Peptides	Peptides per protein Range; Median	Protein Coverage Range; Median (%)
BT474	3914	38373	1 to 403; 4	27408	1 to 234; 4	0.2 to 90.9; 12.3
SKBR3	3054	34261	1 to 778; 4	25408	1 to 621; 4	0.1 to 87.4; 12.9
MCF7	1140	21768	1 to 185; 3	11693	1 to 195; 3	0.2 to 78.9; 10.9

The reporter tag intensities per sample were controlled (**Figure 5-1**). The total reporter tag intensity is somewhat variable for SKBR3 and very stable for BT474 and MCF7, though MCF7 showing a considerably lower total for reporter tags (**Figure 5-1A**). The distribution of the log₂ intensities before and after VSN normalization was plotted to control if the normalization was successful (**Figure 5-1B, C**). In **Figure 5-1C** the stabilized variance is visible for all samples across runs, except Con2 from MCF7, whose median was lower than for the others. In the following steps Con2 for MCF7 was significantly lower. As a results,

Con2 for MCF7 was excluded leaving three control replicates for further analysis. The normalized intensity matrices can be accessed at [Additional file 16 Table S1-3](#).

These results indicate that BT474 and SKBR3 provide robust proteomic data, while MCF7 exhibits a lower number of identifications, potentially reflecting differences in sample quality.

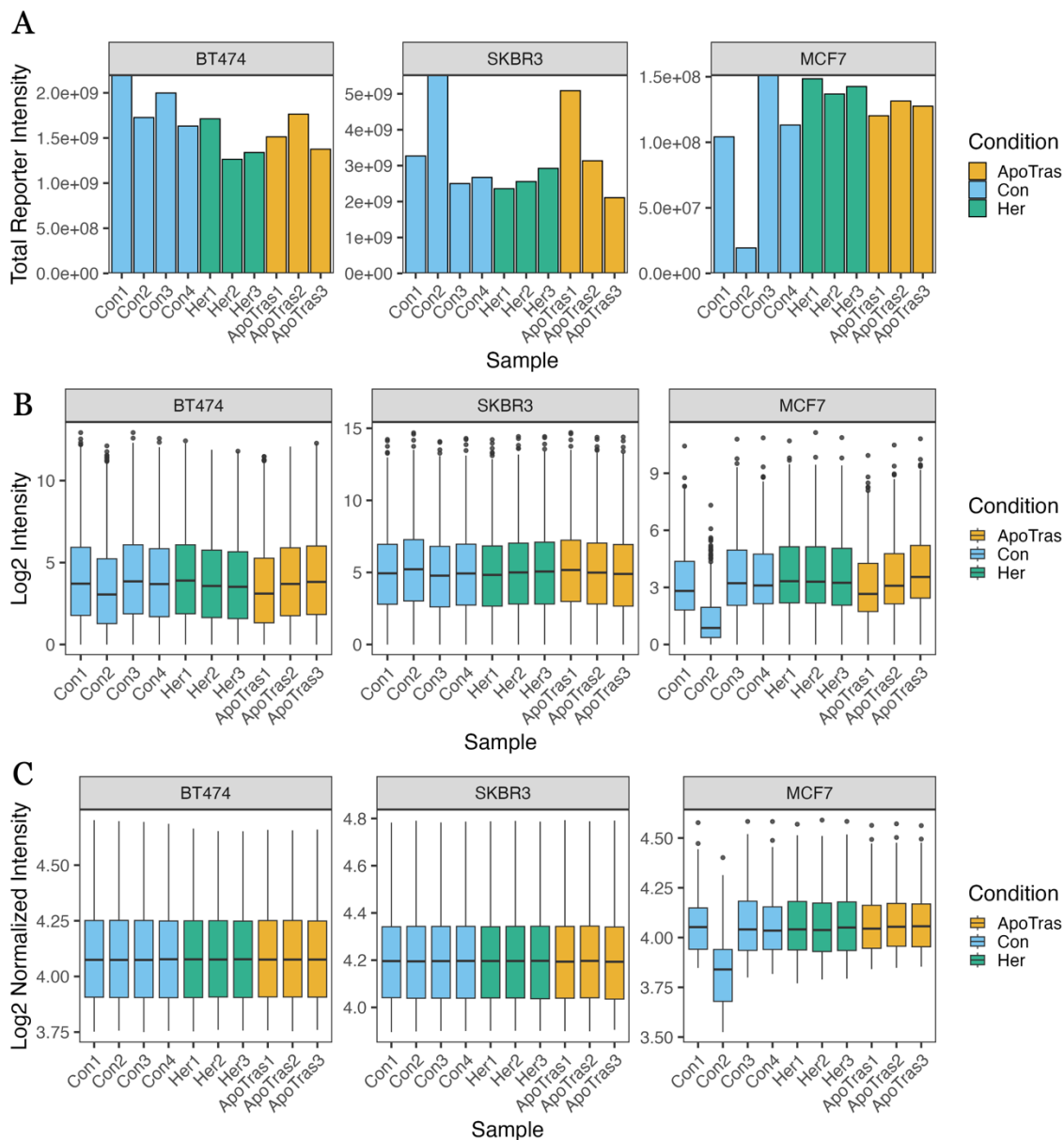


Figure 5-1 Technical quality assessment of quantified reporter intensities across all three proteomics datasets. A) Total reporter intensities per sample (barplot). B) Log2 intensities of quantified proteins per sample (boxplot). C) Log2 normalized intensities of quantified proteins per sample (boxplot).

5.3.1.2 Biological Quality Control

Prior to further analyses, quality control measures were calculated for the three experiments. First, the Pearson correlation of housekeeping genes was calculated (**Figure 5-2A**). For B474, the overall sample correlation was very high (> 0.94), but a separation

between control and treated sample is still visible potentially indicating the strong treatment impact. For SKBR3, the correlation between the samples is similar to BT474 with $R > 0.94$ and no visible separation between samples. For MCF7, the overall correlation is a little bit lower with values above 0.9. The grouping of the samples per experiment was investigated via Principal Components Analysis of the 500 most variable genes (**Figure 5-2B**). For BT474, a clear separation of control samples from treated samples is visible along principal component 1 (PC1) representing over 70% of the variance. For SKBR3 samples a similar separation is visible along PC1 but representing only 24% of the variability, indicating much less impact by the treatment. For MCF7, PC1 represented 19% of the variance with no visible separation by condition.

The protein extraction and TMT labelling was performed in order of replicate numbers to avoid preparation in the order of conditions. A batch effect based on sample preparation order cannot be observed and no batch correction was calculated in the detection of differentially expressed proteins.

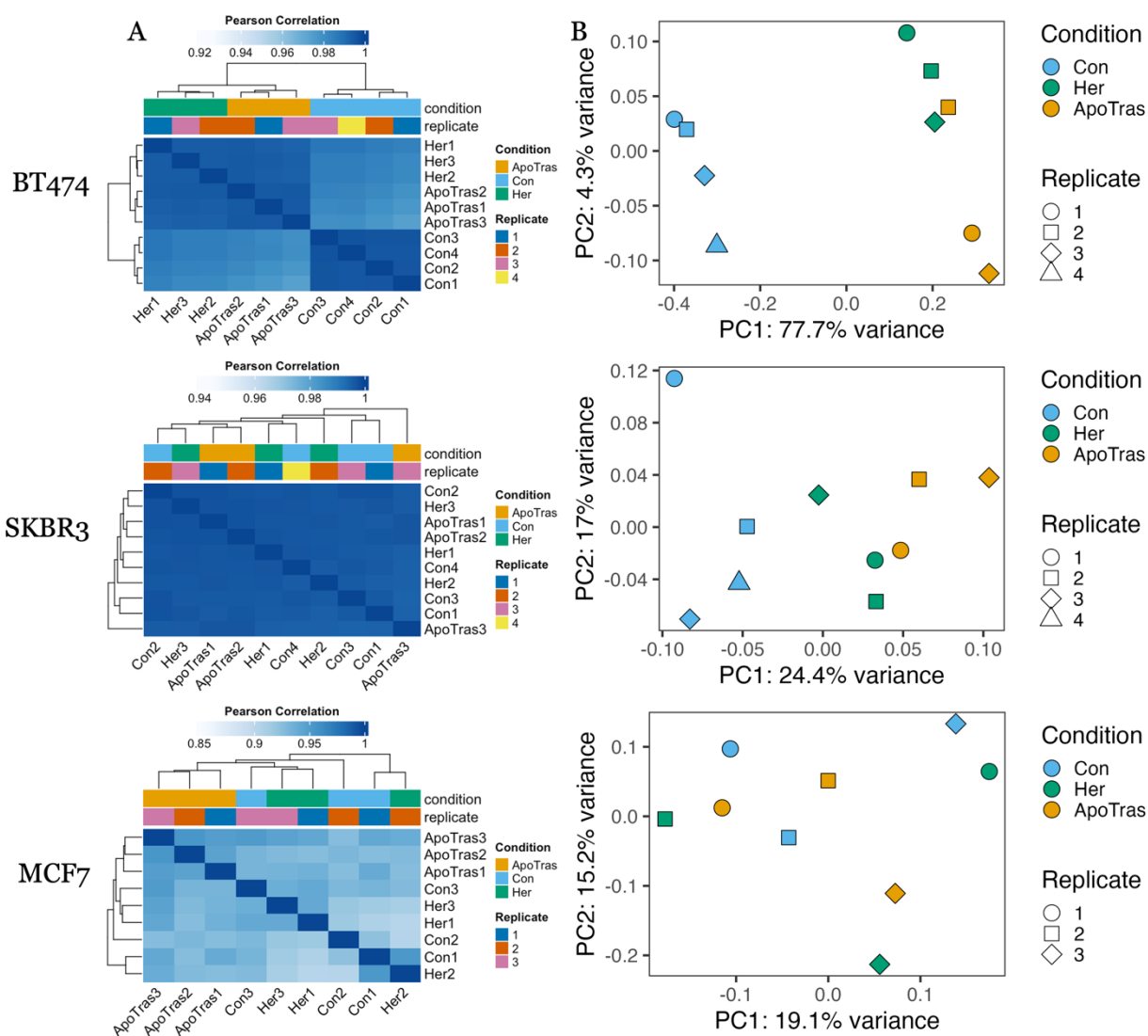


Figure 5-2 Biological quality assessment of proteomics data for treated BT474, SKBR3, and MCF7 cell lines. A) Heatmap of Pearson correlations of detected housekeeping genes in each experiment. B) Principal component analysis (PCA) of the top 500 most variable genes for experiment.

5.3.1.3 Comparing Proteomics experiments

Next, the proteomics experiments were compared. The three experiments share 962 proteins which accounts for 84% of all detected proteins of the MCF7 experiment supporting that the experiment provides usable data despite the smaller number of identified proteins. Only 3% of proteins were unique for the MCF7 cell line. For SKBR3, 81% of

identified proteins are shared with at least one of the runs and only 16% of identified proteins are unique for SKBR3. For BT474, the proportion of common proteins drops due to the significantly higher number of identified proteins. The BT474 samples share 65% of proteins with at least one other experiment and 34% of identified proteins are unique to BT474. Next, the variance was investigated via PCA plot of the 500 most variable genes of the combined scaled data set. It shows the already expected grouping. The BT474 samples separated the most according to treatment along PC1 followed by SKBR3 samples, which showed similar division but with much smaller separation along PC1, indicating that the protein levels are impacted by treatment but not as strongly as for BT474. The MCF7 sample group in the center of the PCA plot shows no clear separation along PC1 and only minor separation along PC2. Next, the proportions of locations with which the identified proteins were identified were calculated (**Figure 5-3C**). The proportion of protein locations are very similar across cell lines with the nuclear structure, cytosol, and mitochondria representing over 60% of the identified proteins. The next group of proteins are localized at plasma membrane and vesicles presenting a combined proportion of 10%. A small proportion of proteins are localized at the cytoskeleton, endoplasmic reticulum and Golgi apparatus.

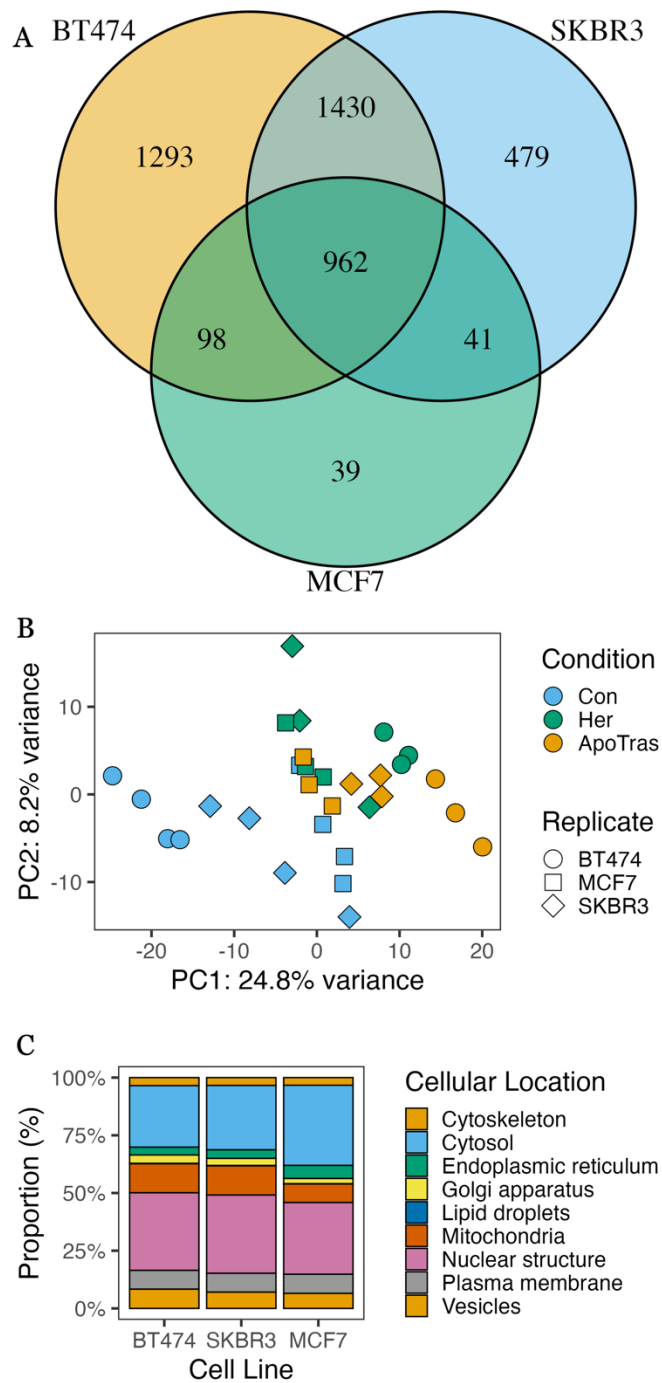


Figure 5-3 Comparison of protein identifications across datasets. A) Venn diagram showing overlap of quantified proteins between all three cell lines, BT474, SKBR3, and MCF7. B) Combined PCA plot of the shared 500 most variable genes. C) Mosaic plot illustrating the proportions of subcellular locations for associated proteins. The information about subcellular locations was derived from the study from (Thul et al., 2017).

5.3.1.4 Comparison of identified differentially expressed proteins

Differentially expressed proteins (DEPs) were identified using limma to evaluate the treatment effects across cell lines. The following comparison were calculated:

- Her vs Con (HervsCon)
- ApoTras vs Con (ApoTrasvsCon)
- ApoTras & Her vs Con (ApoTras&HervsCon)
- ApoTras vs Her (ApoTrasvsHer)

A 5% FDR threshold was applied for identifying significant DEPs, and the complete DEP analysis results are provided in [Additional file 17 Table S4-16](#). The overlap between ApoTrasvsCon and HervsCon comparisons was examined to assess the consistency of treatment effects within each cell line (**Figure 5-4**).

For BT474, the highest number of differentially expressed proteins (DEPs) were detected with 1846 DEPs for ApoTrasvsCon and 1608 for HervsCon (**Figure 5-4A**). The overlap between both was also the highest with 72%. Accordingly, the comparison for ApoTras&HervsCon also showed a similar number of DEPs with 1816 of which 99% overlap with one of the other comparisons. In agreement, the correlation coefficient was between the log fold changes of the two-treatment comparison was over 0.9 indicating very high similarity (**Figure 5-4B**). No DEPs were detected which changed in the opposite direction for the two treatment comparisons.

For SKBR3, significantly fewer DEPs were detected with much less overlap between the comparisons. For ApoTrasvsCon, 125 DEPs were detected while 124 DEPs were detected for HervsCon. The overlap in DEPs was 45% still indicating similarity but much lower (**Figure 5-4A**). As a result, the combined comparison (ApoTras&HervsCon) showed a lower overlap for SKBR3 then for BT474 with 45% and detecting 62 additional DEPs which are not shared with the other comparisons. When comparing the log fold changes (LFCs) for both treatment comparisons the calculated LFCs cover a much smaller range with the

majority of LFCs being below 1 for both comparisons indicating a smaller treatment effect. The LFCs significantly correlate with 0.35 which is smaller for than for BT474.

For MCF7, almost no DEPs were detected, with 4 for ApoTras&HervsCon, 5 for ApoTrasvsCon and 5 for HervsCon. Only three DEPs were shared across all comparisons, reflecting the limited response of MCF7 to treatment. This aligns with prior observations suggesting that MCF7 is less responsive to these treatments.

Comparing the ApoTras&HervsCon comparisons across SKBR3 and BT474 experiments, they share 59 DEPs from the total 125 for SKBR3 comparison indicating a similarity in the treatment effect with limited confidence level because of the difference in magnitude of detected DEPs. The comparison of both treatments (ApoTras&Her) revealed only 1 DEP for BT474 and 0 for SKBR3. This indicates minimal differential expression between the two treatments for these cell lines. In summary, BT474 and SKBR3 showed a clear effect of the treatment with BT474 showing significantly stronger induced changes by the treatment and the measured difference between treatments is rather small. As expected MCF7 cell line showed only a very limited effect to the treatment.

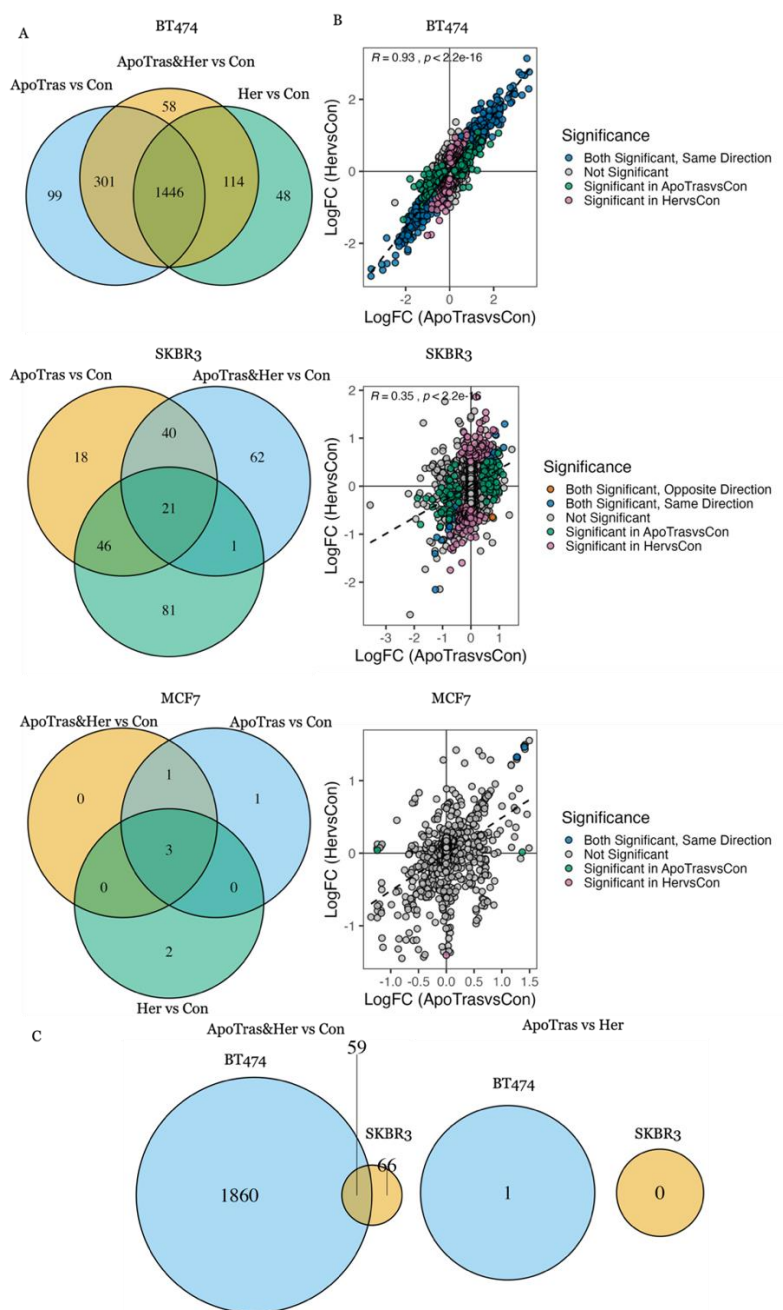


Figure 5-4 Comparison of differentially expressed protein results for all three cell lines. Threshold for significance is 5% FDR. A) Venn diagrams representing the overlap between the comparisons ApoTrasvsCon, HervsCon, and ApoTras&HervsCon. B) Scatterplot showing the similarity of LogFC values for the comparisons ApoTrasvsCon and HervsCon for each cell line. The black dashed line represents the linear regression trend line (Pearson correlation). The Pearson correlation coefficient (**r**) and p-value (**p**) are displayed in the corner of the plot. C) Venn diagrams presenting the overlap between BT474 and SKBR3 for ApoTras&HervsCon or ApoTrasvsHer.

5.3.2 Multi-omics Profile of Targeted Response to trastuzumab

To combine both modalities, the overlap in differentially expressed genes (DEGs) and differentially expressed proteins (DEPs) were evaluated. The applied thresholds were a log fold-change (LFC) of $\log_2(1.25)$ and a 5% false discovery rate (FDR) for DEGs, and a 5% FDR without an LFC threshold for DEPs Figure 5-5. The number of DEGs and DEPs varied significantly between comparisons, ranging from 4 DEPs in MCF7 to 4,532 DEGs in BT474. No feature was found to be differentially expressed across all modalities and responding cell lines. This divergence does not allow to find commonalities across the datasets on this level.

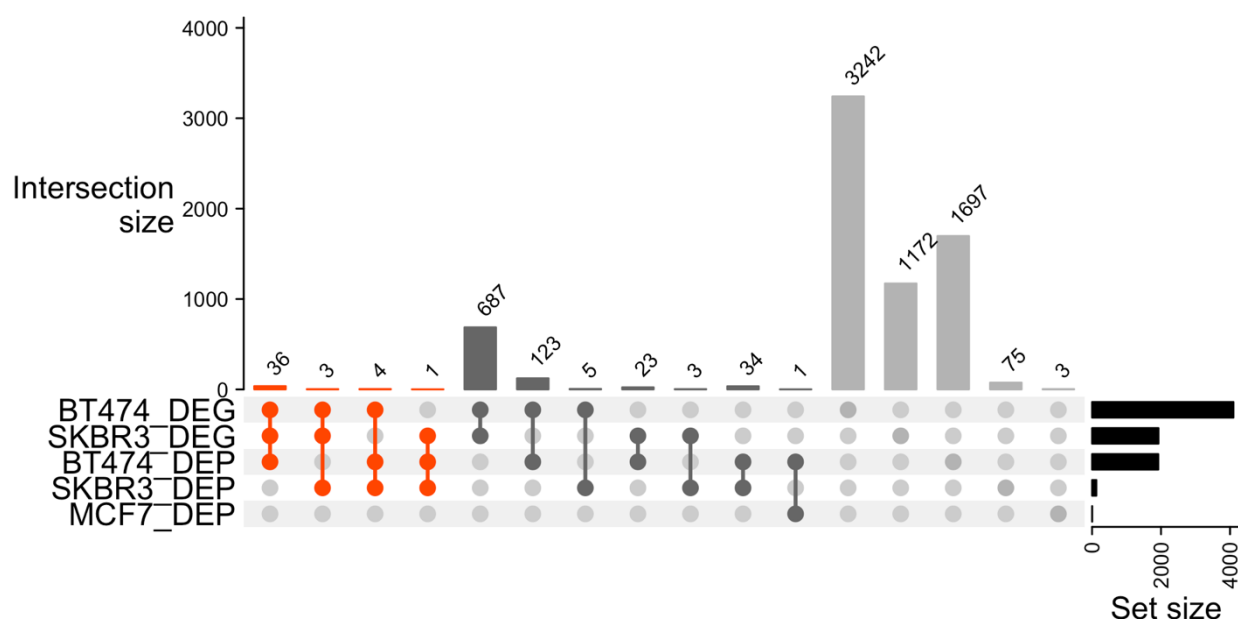


Figure 5-5 Intersection visualization of differentially expressed gene (DEG) and proteins (DEP) for all three cell lines. Thresholds for identifying DEGs were set to a fold-change (FC) of 1.25 and an FDR of 5%. For DEPs, the threshold was an FDR of 5%.

5.3.2.1 Shared enriched Gene sets for trastuzumab treatment

Combining the datasets after GSEA for biological processes shows a different picture. The applied threshold for statistical significance were 10% for the RNA modality and 20% for the protein modality. The complete GSEA results are available in [Additional file 18 Table S17-29](#). MCF7 protein data did not show any significantly enriched gene sets. The

range of enriched pathways for the other comparisons was more comparable than on DEG and DEP level, with 446 for BT474_GSEA_RNA and 881 for SKBR3_GSEA_RNA. A total of 40 enriched gene sets were shared across all comparisons **Figure 5-6A**. These 40 enriched genes sets were then visualized using Cytoscape (Shannon et al., 2003) to identify overlap between the genes sets in their leading-edge genes **Figure 5-6B** resulting in 14 functional groups of gene sets. Among these, seven groups were negatively enriched, and seven groups were positively enriched. The size of the gene set groups ranges from two to eight.

The negatively enriched gene set groups were dominated by cell cycle-related processes, including Cell Cycle Regulation, DNA Replication, Chromatin remodelling, and Cell Division. Additionally, groups related to RNA Splicing, DNA Repair, and mRNA Metabolism were also negatively enriched. These gene set groups likely represent a direct effect of treatment with the monoclonal antibodies Herceptin or ApoTras or a cellular response to the treatment.

The positively enriched gene set groups were more diverse, including metabolic-related processes such as Phospholipid Metabolism, Lipid Transport, and Lipid Oxidation. Other positively enriched groups included Leukocyte Activation, Endocytosis, Differentiation, and Protein Localization. Together, these gene set groups present a clear profile of the cellular response to trastuzumab, supported by multimodal evidence from two HER2+ cell lines. Additionally, the MCF7 protein data supports the specificity of this profile as a targeted cellular response to trastuzumab, rather than an untargeted reaction.

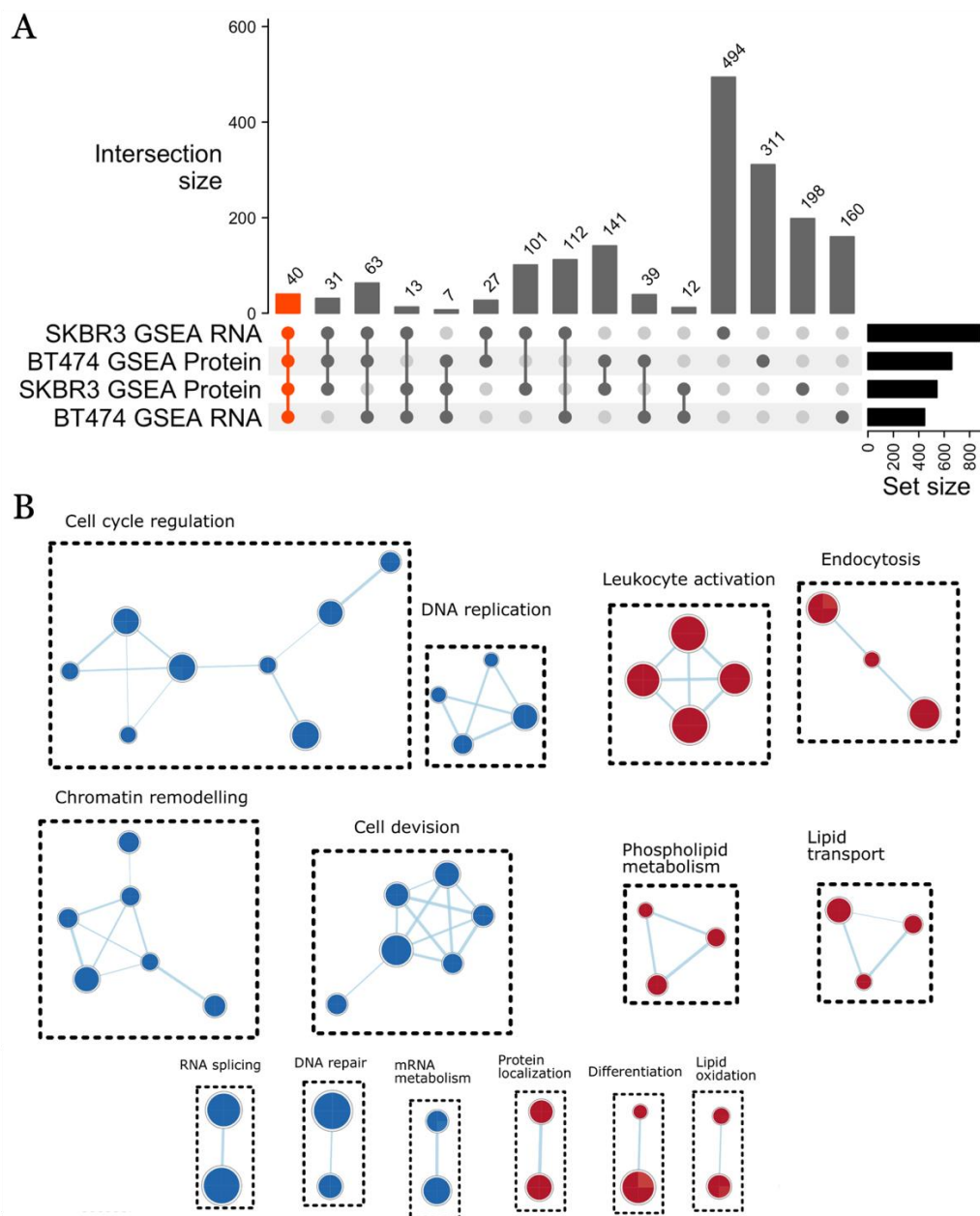


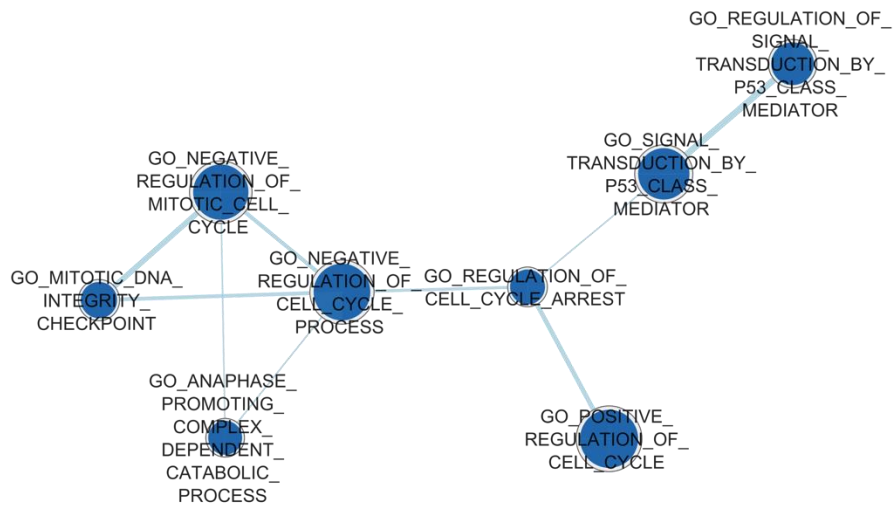
Figure 5-6 Combined trastuzumab effect profile of enriched gene sets in Transcriptomics and Proteomics data. The comparison ApoTras&HervsCon was used. A) Overlap of enriched gene sets across all four datasets and their combinations. A 10% threshold was applied to the GSEA results for RNA-seq data, and a 20% threshold was applied to the proteomics data. B) Gene set network showing shared enriched gene sets. Nodes represent the gene sets and are quartered and coloured by the direction of change of the gene set. from the GSEA analysis from the different data sets. The top-left quarter of each node represents SKBR3 RNA data, the bottom-left quarter represents SKBR3 protein data, the top-right quarter represents BT474 RNA data, and the bottom-right quarter represents BT474 protein data. Edge thickness indicates the strength of overlap between the gene sets.

5.3.2.2 Marker genes for trastuzumab response profile

To robustly connect trastuzumab response profile of gene to the protein/gene level, the shared leading-edge genes for each gene set group were clustered using STRINGdb (Szkarczyk et al., 2023) network information. The top 5 most connected genes were identified as hub genes and selected as representative markers for each gene set group.

Figure 5-7 illustrates this process for the Cell cycle regulation gene set group. The complete list of hub genes for each gene set group is provided in Table 5-2. A couple of genes appear multiple times, with TOP2A being the most frequent, appearing four times across Cell cycle regulation, Chromatin remodelling, DNA repair and DNA replication gene set groups. CDK1 is listed three times with Cell cycle regulation, Chromatin remodelling, DNA repair. VAMP8 also appears three times including Leukocyte activation, Endocytosis, and Protein localization. This highlights that these clusters are not independent and are often regulated through similar pathways.

A



B

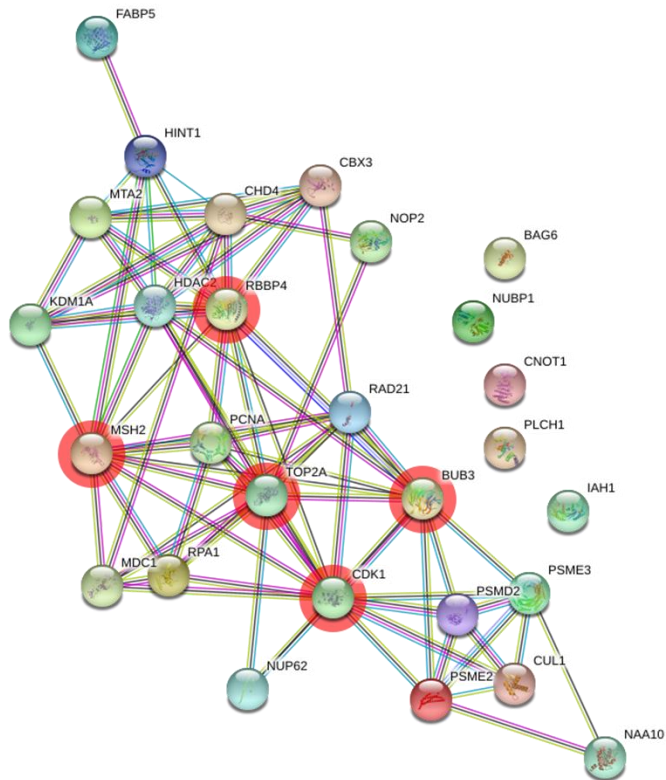


Figure 5-7 Hub genes for Cell cycle regulation gene set group. A) Subnetwork of Cell cycle regulation gene set group for gene ontology biological processes. B) Network of leading-edge genes shared by both cell lines and modalities. The identified hub genes are highlighted with red circles.

Table 5-2 Top 5 hub genes for each gene set group.

Gene set group	Genes	Direction
Cell cycle regulation	CDK1, MSH2, BUB3, RBBP4, TOP2A	Down regulated
DNA replication	TOP2A, RAD21, SMC4, BUB3, SMC2	Down regulated
Chromatin remodelling	MCM7, NASP, DNMT1, CDK1, TOP2A	Down regulated
Cell division	NUP62, RAN, NAA50, NAA10, NSUN2	Down regulated
RNA splicing	SRSF1, SRSF7, SRSF3, TRA2B, WBP1	Down regulated
DNA repair	TOP2A, MSH2, RPA1, CDK1, MCM4	Down regulated
mRNA metabolism	PUF60, MBNL1, SRPK1, PSME2, PSMD2	Down regulated
Leukocyte activation	CD63, CD81, VAMP8, RAB5B, ANXA2	Up regulated
Endocytosis	RAB5A, RAB5B, VAMP8, ARF6, CD63	Up regulated
Phospholipid metabolism	CERS2, SGPL1, TECR, ESYT2, ESYT1	Up regulated
Lipid transport	NPC2, NPC1, ANXA2, HEXB, ACSL3	Up regulated
Protein localization	RAB10, RAB3D, ARF6, EHD4, VAMP8	Up regulated
Differentiation	PKP2, DSC2, PRSS8, RAB25, ANXA1	Up regulated
Lipid oxidation	ECH1, ACOX1, ACOT8, ETFB, ABCD3	Up regulated

5.3.2.3 Biosimilarity Assessment

The GSEA analysis of GO biological processes for the comparison ApoTras vs Her was performed to identify differences between the biosimilar ApoTras and the originator product Her across modalities. The thresholds for statistical significance were set at 10% for the RNA level and 20% for the protein level. The results revealed a more cell-line-specific pattern **Figure 5-8**. For BT474 GSEA RNA, 689 gene sets were significantly enriched and BT474 GSEA protein showed 81 enriched gene sets. In contrast, SKBR3 GSEA protein showed no enriched gene sets and SKBR3 GSEA RNA identified only in 21 enriched gene sets. The only observed overlap of enriched gene sets across RNA and protein level appeared between BT474 GSEA RNA and BT474 GSEA protein, with 4 shared pathways. These gene sets included somite development, circadian regulation of gene expression chaperone-mediated protein folding and protein folding. The detected difference is limited and observed only in BT474 cell line.

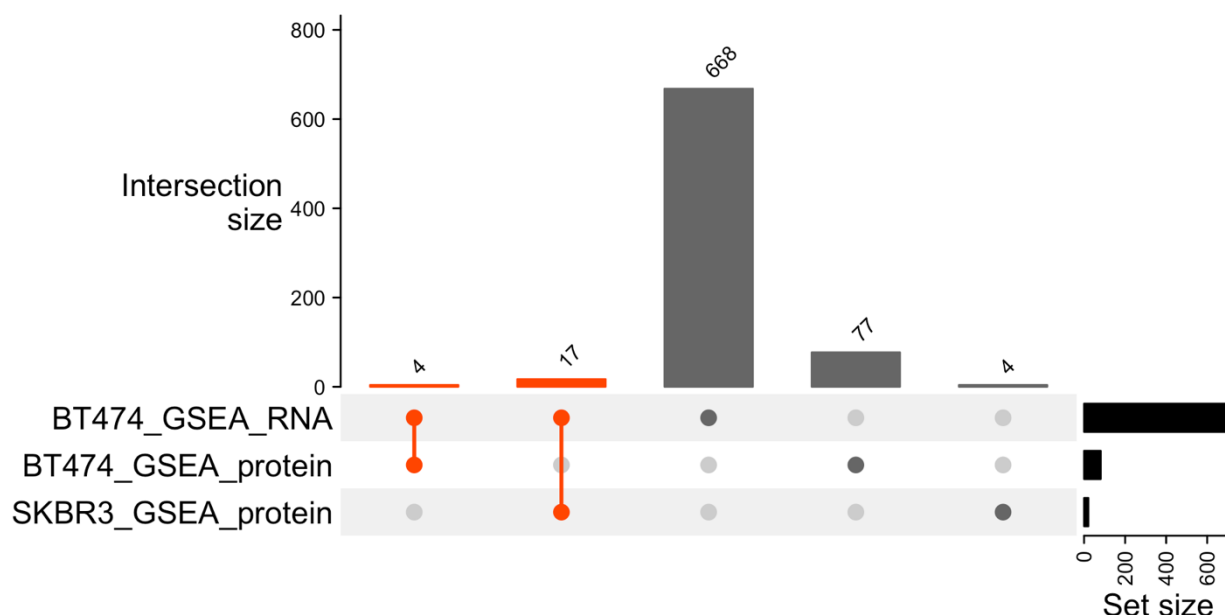


Figure 5-8 Overlap of enriched gene sets from biological processes BP across all four datasets for the comparison ApoTrasvsHer. A 10% threshold was applied to the GSEA results for RNA-seq data, and a 20% threshold was applied to the proteomics data.

The shared overlap in enriched gene sets for BT474 RNA and protein included two positively enriched gene sets, somite development and circadian regulation of gene expression, indicating increased expression in ApoTras treated samples. Additionally, two negatively enriched gene sets with Chaperone-mediated protein folding and protein folding showed reduced expression in ApoTras (**Figure 5-9A**).

When investigating the overlap in leading edge genes of the GSEA, it showed that Chaperone-mediated protein folding and protein folding were found to share multiple leading-edge genes, with total of nine genes across modalities (**Figure 5-9B**). When combining the leading-edge genes shared across modalities a total of 23 leading-edge genes is shared between BT474 RNA and BT474 protein.

Analysis of these 23 genes using STRING DB network information revealed a highly connected network, suggesting a shared function (**Figure 5-9C**). The top five genes with the highest degree of connectivity were primarily heat shock proteins (HSP), including HSPA8, HSPA9, HSPD1, HSPE1, and the HSP90-stabilizing PTGES3. Other noteworthy genes included Prefoldins (PFDN1, PFDN2, PFDN6), Clusterin (CLU), and Cell Division Cycle

37 (CDC37). The reduced expression of genes related to protein folding in ApoTras-treated samples suggests biologically meaningful differences.

Two leading-edge genes are shared between RNA and protein data for the circadian regulation of gene expression gene set: O-linked N-acetylglucosamine Transferase (OGT) and HECT, UBA, and WWE Domain Containing E3 Ubiquitin Protein Ligase 1 (HUWE1). Both genes were enriched in ApoTras-treated samples compared to Her-treated samples in both protein and RNA data. Although these differences are not shared between cell lines and are observed only in BT474, they are supported at both the protein and transcript levels.

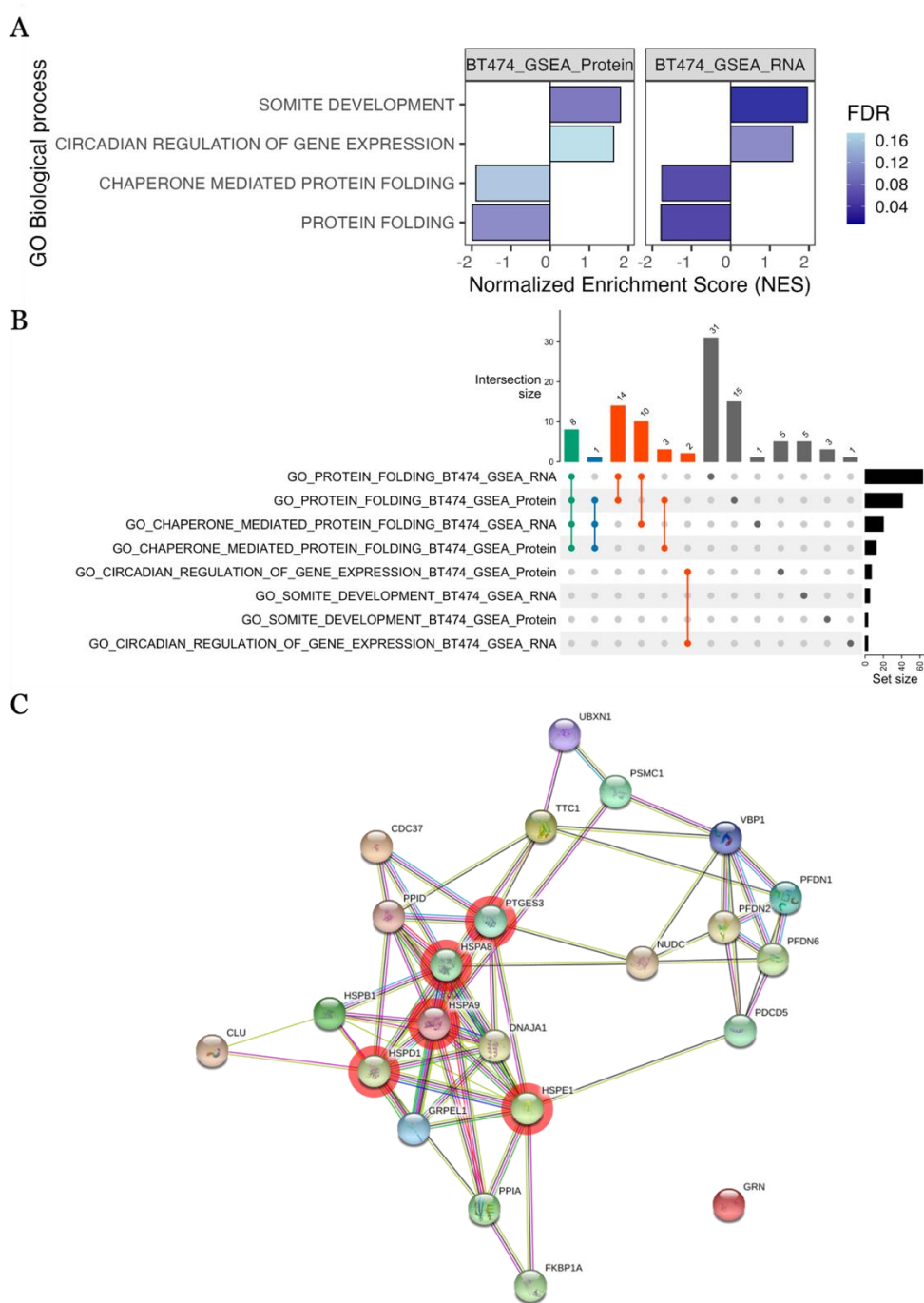


Figure 5-9 Shared leading-edge genes for BT474 cell line modalities. A) Normalized Enrichment Score for significantly enriched biological processes in RNA and protein data of BT474. B) Overlap in leading edge genes between the shared enriched gene sets. C) Network of the shared leading-edge genes for Protein folding and Chaperon mediated protein folding gene sets across RNA and protein level. The most connected genes are highlighted in red.

5.4 Discussion

Trastuzumab has changed the game for the treatment of HER-2 positive breast cancer with its targeted effect on HER2 signaling pathways, ability to down regulate the cell cycle and activate apoptotic signals. But, downstream molecular changes, particularly at the protein level, that occur following successful protein knockdown are still poorly understood.

These data collectively provided a comprehensive “systems biology” view of Herceptin and ApoTras’ mechanisms of action in this chapter. A deeper comprehension of the molecular mechanisms responsible for the action of trastuzumab can further clarify its multifarious action. These data sets will help to test for biosimilarity in relation to ApoTras. The preceding discussion provides the relevant considerations, including the explanation of proteomics and transcriptomics- defined cellular response profiles, the variability in biosimilarity, and the implications for the future of biosimilars and treatment of HER2-positive breast cancer.

5.4.1 Proteomics data

Important information on the cellular reaction to trastuzumab and its biosimilar, ApoTras, was revealed by the proteomics analysis. While MCF7, a HER2-negative control, showed little response, high-quality data from BT474 and SKBR3 cell lines identified high numbers protein and showed clear treatment effects. For BT474, principal component analysis (PCA) demonstrated a distinct separation between the treated and control samples. This effect was also noticable for SKBR3 but was less pronounced. Analysis of differentially expressed proteins (DEPs) showed that SKBR3 had moderate alterations, with fewer DEPs and smaller fold changes, whereas BT474 had significant changes, with over 1,800 DEPs for both treatments. The consistency of the proteomics studies was validated by the overlap of identified proteins across cell lines. All datasets showed a similar distribution according to functional evaluation of protein localisation, with nuclear, cytosolic, and mitochondrial proteins predominating.

5.4.2 Cellular trastuzumab Response Profile

In integrating transcriptomics with proteomics data, significant cellular pathways relevant for trastuzumab response are found. Enrichment of common pathways also revealed large changes in biological processes, lending additional multimodal evidence in support of the effects of trastuzumab. Functional annotations related to cell cycle control, DNA replication, RNA splicing and chromatin remodeling processes were shown to be significantly depleted in the presence of trastuzumab, indicating that this compound also had an impact in proliferating and genomic stability. Among the positively enriched gene set groups, leukocyte activation, protein localization, and lipid metabolism highlight adaptive cellular responses and putative immune-related effects. The finding of hub genes within the enriched groups of gene sets is informative and helps to unveil important information about the molecular processes underlying the therapeutic effect of trastuzumab. Also, many of the same hub genes, such as TOP2A, CDK1, VAMP8, were present across various functional groups, suggesting an important role of these genes in relevant trastuzumab responses (Y. Gu et al., 2020; H. Wang et al., 2018).

TOP2A (Topoisomerase II Alpha) is an established cell proliferation marker and therapeutic target of chemotherapy in various types of cancer, including breast cancer. This is in line with the reported enrichment of pathways correlated with inhibition of DNA replication and cell cycle arrest. While not being mechanistically related to the central action of trastuzumab on its target HER2, due to its role in replication stress and DNA damage response TOP2A could represent a bystander target in HER2-positive cells treated with trastuzumab (Y. Lin et al., 2020).

Cell cycle progression, particularly transition through the G2/M checkpoint, is highly dependent on CDK1 (Cyclin-Dependent Kinase 1). Of note, some of the findings reported in the present study would coincide with the role of trastuzumab in cell cycle arrest, particularly at G1-phase as also reviewed elsewhere (N. Jiang et al., 2021; Y. Li et al., 2023). Despite the low research of trastuzumab on CDK1, due to its important function in the progression of BC and in the response to target therapies, CDK1 can be explored as a

target for a possible involvement in the efficacy of the treatment (Y. Lin et al., 2020; Xue et al., 2020).

VAMP8 (Vesicle-Associated Membrane Protein 8) is thought to code for a protein that plays a central role in leukocyte activation as well as in endocytosis and in the localisation of proteins. The contribution of VAMP8 in vesicle trafficking as well as in recruitment of immune cells might be indirectly associated with trastuzumab treatment, whose benefits are known to be mediated by antibody- dependent cellular cytotoxicity (ADCC) even if not specifically mentioned here (Arnould et al., 2006).

Among the most important hub genes, RBBP4 (Retinoblastoma-Binding Protein 4) and MSH2 (MutS Homolog 2) are representative of functions in chromatin remodelling and DNA repair. Their decreased levels further suggest that trastuzumab can interfere with the mechanisms that control maintenance of the genome, making cancer cells more susceptible (Pattanayak et al., 2022). Although the tumour-specificity of the RBBP4 and MSH2 genes and their relevance to trastuzumab are not yet elucidated in the literature, both have been associated with progression of breast cancer and, so, might regulate functions that contribute to augmenting the effectiveness of this therapy.

Several of these hub genes present strong connections to endocytosis and intracellular trafficking pathways, such as RAB5A, RAB5B, ARF6. These genes are indirectly related to the hypothesized mechanism of trastuzumab through the internalization and degradation of HER2 receptors. Their participation suggests that trastuzumab might also have other actions rather than just blocking direct signalling in HER2-positive cells (Maldonado et al., 2024).

The hub genes identified can be related to metabolic processes, particularly of lipid oxidation, and lipid transport such as ECH1, ACOX1 and NPC1. Despite that this is not a well-documented route for trastuzumab, it is a well-known metabolic role for cancer metabolism (O'Neill et al., 2022). Cancer biology is marked by metabolic reprogramming, and the effects of trastuzumab on these pathways can tell us about metabolic vulnerabilities of HER2-positive tumours more generally (Gandhi & Das, 2019).

In the present literature, there is no clear relationship between trastuzumab treatment and

some of the hub genes and gene set groups that were identified. Among these genes, there are several that are well known contributing factors to breast cancer, also to other cancers. The leukocyte activation pathway was enriched in trastuzumab-treated BT474 and SKBR3 cells, despite the absence of immune cells in the experimental setup. The identified genes, CD63, CD81, VAMP8, RAB5B, and ANXA2, are traditionally associated with immune cell function, particularly in vesicle trafficking, membrane dynamics, and intercellular communication. However, these genes are not exclusive to leukocytes and are increasingly recognized as part of cancer cell-intrinsic programs related to stress adaptation, membrane repair, and exosome-mediated signaling. For instance, CD63 and CD81 are tetraspanins enriched on exosomes and endosomal membranes and are involved in receptor trafficking and cell motility (Kalluri & LeBleu, 2020; LeBleu & Kalluri, 2020). Similarly, VAMP8 and RAB5B regulate vesicle fusion and endocytic sorting and have been shown to promote invasive behavior and focal adhesion dynamics in breast cancer cells in vitro (Frittoli et al., 2014). ANXA2 plays a critical role in membrane stabilization, plasmin generation, and stress-response signaling, and is secreted via CD63+ vesicles upon interferon- γ stimulation in epithelial cells—even in the absence of immune components (Ecsédi et al., 2017; Zheng et al., 2011).

The expression of these genes in monocultured cancer cells suggests that their upregulation may be part of a cell-intrinsic response to trastuzumab-induced stress. Trastuzumab is known to promote internalization and degradation of HER2, leading to remodeling of membrane domains and increased trafficking demands, potentially triggering compensatory activation of vesicle-related genes. Furthermore, these gene programs may reflect a form of immune mimicry or preparatory signaling, in which cancer cells activate immune-associated machinery in anticipation of immune surveillance. This aligns with prior reports showing that cancer cells can express immune-modulatory genes even in vitro, as part of autocrine or paracrine loops or broader survival strategies (Tang et al., 2020). These findings suggest that trastuzumab-treated cells may adopt an immune-interactive phenotype, even in the absence of immune effectors, which could have implications for therapy resistance or response to combination regimens involving immune-based strategies.

Lipid metabolism genes such as genes involved in the synthesis of ceramide, CERS2 or in fatty acids activation ACSL3, have been associated with tumor progression and chemoresistance in breast cancer (Fan et al., 2015; S. Tan et al., 2024). These results validate the growing recognition of metabolic reprogramming as a hallmark of cancer and as a factor influencing tumour development, survival and sensitivity/ resistance to therapies. The fact that lipid-related pathways are over-represented among the trastuzumab response signature could point out to metabolic weaknesses of HER2-positive tumors. Combining the use of drugs that target the metabolism of trastuzumab may represent new approaches to revert resistance mechanisms and obtain better responses in patients with HER2-positive breast cancer (Gandhi & Das, 2019).

Novel or not investigated genes, in particular in metabolic and immune related pathways we identified indicate the potential of this multi-omics technique to unveil novel aspects of trastuzumab-related mechanisms. These data can be further explored to identify and develop new therapies and biomarkers for HER2 positive breast cancer and other more broadly applicable tumours that function via similar mechanisms.

5.4.3 Assessing Biosimilarity with multi-modal data

Incorporation of transcriptomics and proteomics analysis allowed a detailed comparison study of Herceptin and its bio-similar, ApoTras. If the treatments showed little differences candidalysin-treated BT474 cell lines display a slightly enhanced overall sensitivity to radiation compared with SKBR3. Gene set enrichment analysis (GSEA) indicated distinct profiles of enriched pathways, primarily differing at the level of the gene. These proteins belong to the protein folding and chaperone mediated protein folding pathways and suggest possible different cellular stress/responses or protein homeostasis mechanisms between ApoTras-treated and untreated BT474 samples (Yuan et al., 2022). These pathways involve heat shock proteins including HSPA8, HSPD1, HSP90, which stabilize PTGES3, thus emphasizing the role of proper protein folding and stabilization in cellular adaption to treatment (Kauraniemi et al., 2004).

Among enriched pathways positively enriched in circadian gene regulation in the ApoTras samples were also found OGT and HUWE1 as leading edge genes (S. Wang & Ding, 2019). Their biological relevance is further supported by the fact that these phenotypes are observed at the RNA and protein levels, suggesting that ApoTras also elicits functionally defined effects opposite to those of Herceptin (Lesurf et al., 2017). The differences observed were only seen in BT474 and not SKBR3, highlighting cell line-specific responses and also suggesting that they may not be applicable across all HER2-positive models (Y. Lin et al., 2020).

Variability observed between BT474 and SKBR3 is likely due to the heterogeneous nature of HER2-positive breast cancer subtypes (Kokot et al., 2024). In fact, HER2-positive breast cancers represent a heterogeneous group of diseases that include several subtypes defined by distinct molecular profiles and signaling pathway activation, and/or response to therapy (Baumann et al., 2018; D. Li et al., 2025). BT474 and SKBR3 are two HER2 positive cell lines with different intrinsic characteristics such as hormone receptor status, growth factor signaling, and metabolic activity, broadly representing preferences of the Luminal or Basal like molecular subtypes. BT474 cells present also higher level of HER2 amplification and are more sensitive to HER2-targeted therapies than SKBR3 cells (McClure et al., 2019). These intrinsic discrepancies could explain the distinct molecular responses of BT474 cell line to Herceptin and ApoTras, as BT474 might be able to more easily sense subtle changes in the mechanisms of action of the drugs, like protein folding pathways (Chang et al., 2022). Appreciating this heterogeneity is important in relation to the interpretation of biosimilarity studies and points to the need for diverse models in biosimilar assessments.

The fact that “Herceptin pathways” and “ApoTras pathways” overlapped in various functions such as cell cycle regulation, lipid metabolism, and immune response, confirm the evaluation of biosimilarity. It is possible that the observed variations in protein folding trajectory reflect limited variability in molecular interactions dynamics, rather than a meaningful difference in network pharmacology properties and therapeutic utility for these complex drugs (Yuan et al., 2022). The results indicate the power of multi-omic analyses in

detecting minute molecular differences that could evade detection by analyses that are either exclusively based in a single omic or conventional approaches for demonstrating biosimilarity (N. Duan et al., 2024).

It will be important to replicate these data in other HER2-positive models. It is not possible to determine from the current observations if these variations lead to any phenotypic variation capable of affecting clinical outcomes without further functional studies. These findings can be understood in the context of the limitation that, survival differences in an experimental context might have been due to non- drug-specific direct effects of the drug or to a beneficial adaptive response to a form of treatment. The double approach employed to define and understand the translational significance of Aporras and trastuzumab represents an example of the multiple multi-omics analyses that should be pursued to address the questions surrounding the safety and efficacy of biosimilars.

5.4.4 Challenges and Opportunities for improvement

This work illustrates the many challenges of a multi-omics approach to drug response and biosimilarity assessment. The heterogeneous nature of HER2-positive breast cancer, as revealed by the behaviour and response of BT474 and SKBR3 cells, represents one of the major limitations to the full understanding of trastuzumab MOA (Sammur et al., 2022). Similarly to studies merging two or more omics levels, further molecular insights are gained from multiomics approaches which, but cannot be generalized because of the multi-omics aspect of modelling. Additional HER2-positive cell lines of different histopathology will be required in order to definitively address this question and account for all possible responses.

Yet, the challenge of deciphering small molecular changes, such as those involved in protein folding pathways, also complicates this type of study. The biological relevance of these variations is currently unknown, although they may not directly lead to changes in the effectiveness of treatment (Zeng et al., 2018). This will obviously need to be resolved by validation in the wet-lab, assessing the effects of these variations on important studies including, but not limited to, key biological pathways, protein homeostasis, stress responses, or tumor cell survival. In addition to this, orthogonal data such as set across patient level

sequencing or proteomic analysis may help make sense of these observed differences (Rossi et al., 2022).

Despite these limitations this is a very fruitful area for advancing this type of research. The integration of transcriptomics and proteomics is a powerful approach to uncover slight molecular differences not detected in single-omics studies (Gómez-Cebrián et al., 2021). This can help us in the understanding of biosimilarity and in getting additional knowledge about pharmacological processes, that is, secondary and adaptive reactions. As an example, the identification of hub genes and common enriched pathways provide a basis for further investigations on their use as biomarkers for assessing the therapeutic effect and biosimilarity (Wei et al., 2021).

The implications of this for clinical work are considerable. The trastuzumab response profile generated in the present study may be used in the construction of predictive tools for patient stratification, helping to identify patients who are most likely to respond to trastuzumab, or its biosimilar (Rashid & Selvarajoo, 2024). In addition, multi-omics studies may aid in the evaluation of further HER2-targeted agents, facilitating the comparison between originator and biosimilar compounds. This process could become faster for the regulatory bodies to issue their final approval to products similar to biotherapeutics and can be utilized for more medications that have a more complex mechanism of action such as these biosimilar medicines (Z. Wang et al., 2023).

It is clear that, although there are still challenges associated with making sense of large amounts of multi-omics data, the potential for furthering precision medicine and biosimilar development is quite promising. The obstacles described in this review will require several changes in experimental approaches, larger models and more functional validations to fully take advantage of the potentials of multi-omics analysis for the development of biosimilars (F. Yang et al., 2024).

Chapter 6: Conclusion

6.1 Summary

The goal of this thesis was to create and apply a rigorous untargeted multi-omics methodology for the evaluation of functional biosimilarity of monoclonal antibodies and to test this methodology on the comparison of Herceptin and its biosimilar ApoTras. By combining transcriptomic and proteomic analysis, this pipeline provided comprehensive view of the molecular responses induced by the drugs in HER2-positive breast cancer cell lines. This methodology goes further than the traditional comparability tests, by describing common as well as unique RNA and protein expression and thus providing a system level, mechanistic view of drug action.

This was accomplished through a number of methodological innovations. To provide a measure of the similarity at the molecular level that correlates well with the differential gene and protein expression. To address this, the concept of a Similarity Index was introduced that provides measure of similarity at the level of molecular features relevant to the biologic activity of a biopharmaceutical. The unified pathway approach was then introduced to describe coordinated changes in expression across sets of molecules to provide a more detailed insights into the biological processes that are affected. In addition, a bootstrapping procedure was performed in order to estimate statistical power and type II error, and and to estimate confidence in detecting the minimum meaningful expression differences across treatments.

These results indicate that at the molecular or phenotypic level, ApoTras is very similar to Herceptin, as over 95% of the measured features are in common. Functional annotation showed consistent downregulation of proliferative and survival pathways and upregulation of apoptotic signaling pathways. However, this analysis also revealed cell-line specific as well as pathway-level differences which could potentially indicate possible compound-specific features, especially in secondary signaling or stress response pathways. In addition to the fact that these regulators corroborate the biosimilarity of ApoTras, the power of integrating multi-omics to the analysis of biosimilarity is clear from these findings. More

importantly, the framework has applicability across other biologics that are complex or not well understood illustrating the relevance of the present test case beyond its specific case.

6.2 Functional Insights and Follow-Up Experiments

The functional study described herein establishes that ApoTras functionally recapitulates the effects of Herceptin in breast cancer cell lines. The transcriptional and proteomic signatures were largely concordant between the two agents and demonstrated downregulation of E2F and MYC target genes, attenuation of G2/M checkpoint control genes, and activation of pro-apoptotic transcriptional programs, suggesting an anti-proliferative, pro-apoptotic stress response. These results are consistent with the mechanism of action of trastuzumab and support the application of ApoTras as a biosimilar.

However, some subtle distinctions were also highlighted by the study at the transcription factor and pathway level. FOXM1 was more efficiently downregulated by ApoTras in BT474 cells, whereas E2F1 was more efficiently suppressed by Herceptin in SKBR3 cells. Similarly, pathway enrichment analysis led to the identification of pathways relevant to the known mode of action of Herceptin against DNA repair and inflammation in BT474, and ApoTras against metabolism and signal transduction in SKBR3. Though the observed differences were small, the replication of effects suggests that they may have biological significance, and future research should pursue following up on these findings.

This was mechanistically supported by the multi-omics integrative studies. Overlap in the suppression of common pathways was observed in the transcriptomic and proteomic data. Herceptin down-regulated CDK1 and CCNB1 more than ApoTras, whereas ApoTras inhibited PLK1 and AURKA to a larger degree than the other compounds. Such discrepancies could be a consequence of divergent cellular responses by the drugs through feedback loops, post-transcriptional regulatory factors or stress response factors.

Several promising directions for future experiments are offered to better understand the implications of these dissimilarities. First, functional studies should validate the regulation of FOXM1, E2F1, PLK1, and AURKA, using protein-level assays like Western blotting or phosphoprotein profiling. Second, targeted perturbation studies such as siRNA knockdown or the use of specific pathway inhibitors could ascertain if these differences influence both treatment susceptibility and apoptotic thresholds, as well as cell cycle kinetics. Third, applying this approach to a larger panel of HER2-positive models and/or primary tumour

samples would allow to assess whether the observed divergence reflect generalizable biological variation or are specific to these cell lines. In general, this chapter illustrates the strength of omics-based profiling not just to verify biosimilarity, but also to identify possible differences worth pursuing in follow-up on mechanistic or preclinical investigations of the biosimilar. These subtle observations also underline the importance of considering the whole complexity of action of drugs when assessing biosimilars.

6.3 Implications and Methodological Advances

This thesis proposes a scalable and flexible omics-based approach for evaluating biosimilarity, offering a comprehensive methodology for making sense of transcriptomic and proteomic information in the context of understanding cellular responses to a drug. Simultaneous analysis of transcriptomic and proteomic data enables identification of upstream regulatory signals and downstream functional effects, providing a more holistic and more accurate view of the drug activity.

Another main methodological contribution of this work is the novel introduction of the Similarity Index, which quantifies the similarity of treatment in thousands of molecular features. This provides a quantitative measure that will enable researchers and regulators to assess “how similar” a biosimilar is to its originator. In addition, pathway-centric enrichment analyses illuminated the functional significance of omics changes by merging genes/proteins into biologically meaningful modules that provide insight into common or contrasting patterns of regulation.

This framework benefits greatly from the statistical rigor it incorporates. This allowed for confirmation of statistically significant results based on power, according to information obtained from the bootstrapping procedure. The latter is especially important for biosimilarity studies due to the need to interpret slight differences in data that is by definition sensitive to detect such differences. These results suggest that the present study had enough power to detect small, but true differences, providing credence not only to the observed similarities but also to the lack of significant differences.

Additionally, the datasets produced in this dissertation remain valuable for future use. The data can serve as a valuable, well-annotated reference source for biosimilar assessments in the future. In the future, when new biosimilars that target HER2 or its downstream pathways are generated, the trastuzumab-induced network documented in this study could be used as a comparison view to assess the functional profile of new biosimilars. The analysis pipeline is modular and can be updated with new pathway databases, statistical tools, or extended to include additional omics layers like phosphoproteomics or metabolomics.

More broadly, the developments outlined here could be applied to other novel biologics with complex unknown modes of resistance. In the case of these drugs, omics studies without specific targets represent a form of hypothesis generation to follow up hypotheses generated by traditional functional and clinical laboratory tests. It provides a way to empirically examine drug behaviour from a system-level perspective using both mechanistic and quantitative perspective.

6.4 Biological and Clinical Relevance

The findings of this dissertation have relevant implications for biosimilar development, as well as translational and clinical oncology research. The transcriptomic and proteomic signatures, representative of the profile of response to trastuzumab generated here, may be used as a potential “gold standard” to assess the response to therapy in HER2+ tumours. Such a signature could be applied in the clinic for biosimilarity testing, or as a predictive tool for patient stratification based on expected response to therapy, therefore allowing for more personalized treatment modalities in the context of anti-HER2 therapy.

A key insight from this study is the concordance between transcriptomic and proteomic data. Transcriptomics permits the detection of various types of transcripts that are often out of reach with proteomic analyses, including genes expressed at lower levels, un-spliced or non-coding RNAs, and alternative isoforms. This property renders RNA-seq an excellent means of profiling the regulatory landscape operating upstream of protein translation. Proteomics, on the other hand, allows for direct measurement of a cell’s functions; it measures the mature proteins and the post-translationally modifications and molecular interactions that select the cell phenotype. The duality of function, as potential (RNA) or action (protein), can provide insight into how drugs are capable of reprogramming cellular function.

Integrating both layers, this study detects concordant alterations in key oncogenic pathways, demonstrates reproducibility across platforms, and clarifies subtle differences in the treatment response. These findings show the promise of multi-omics in terms of elucidating not just whether a biosimilar is functional, but also the mechanisms by which it is, and how those mechanisms could differ by context. This is crucial in clinical settings where drugs are evaluated not only for efficacy but also on their safety, mechanisms of action and applicability to specific patient subgroups.

The framework developed in this thesis could be important as regulatory bodies have now begun to adopt a totality-of-evidence approach for assessing biosimilarity. It does this by offering a comprehensive understanding of mechanisms at scale, where traditional pharmacokinetic and immunogenicity analyses are unable to provide this level of focus, and

thus helps to translate preclinical efficacy to clinical applications. In addition, because the data produced here can be reanalysed, built upon, and modified, these data can gain value as a point of reference. In the end it becomes clear that the multi-omics and integrative approach described in this dissertation is poised for advancement not only in regulatory science, but also in clinical practice. This can at least pave the way for more intelligent decision-making in regard to the development of biosimilars and even provide this field's approach to the next phase of system biology-based approach for bioassays.

6.5 Limitations and Outlook

The study is limited in several directions including experimental scope, mechanistic and immunologic insights, as well as data-related challenges. However, many of these can be alleviated with additional follow-up studies. The study focused on two breast cancer cell lines with BT474 and SKBR3 and even though they present some major HER2+ subtypes, they do not fully reflect the diverse picture of HER2+ breast cancer. However, this can easily expand with additional data from additional cell lines, such as AU565, MDA-MB-361 or HCC1954 (Thennavan et al., 2021). Another option would be to leverage large genomic resources, such as TCGA BRCA (Koboldt et al., 2012) data containing breast cancer samples or DepMap (Tsherniak et al., 2017) which contains expression profiles for many drug-treated cell line studies. These data sets provide access to diverse HER2+ tumour profiles from patients (TCGA BRCA) and transcriptomic and proteomic drug response profiles across a wide range of cell lines (DepMap). This enables the comparison of biosimilar signatures within a wider biological context. In addition, the comparison may also provide insights into molecular cancer subtypes or genetic backgrounds associated with drug sensitivity to the biosimilar versus the originator, enabling more personalized predictions of treatment efficacy.

Furthermore, this study includes data from one time point and dose, which limits the ability to study resistance mechanisms, dose-dependent responses, or distinguishing between the direct impact of the drug on the cells and the adaptive response of the cell. Resistance mechanisms are commonly investigated via time-course experiments coupled with dose escalation treatments to evaluate the cell response over multiple months of treatment with trastuzumab. These can inform on the activation of compensating signalling pathways, including but not limited to PI3K/AKT or upregulation of other HER family receptors, that would permit cells to escape from HER2 inhibition. These types of experiments could add additional details to the response comparison for assessing biosimilarity. Similarly, 3D cell culture models or primary cancer samples can help address tumour complexity and interactions within the tumour microenvironment (Salinas-Vera et al., 2022). These

experiments in combination with spatial or single-cell method application will enable a more detailed and nuanced analysis (S. Z. Wu et al., 2021).

Furthermore, the study is missing immunological and mechanistic information. Trastuzumab is known to cause antibody-dependent cellular cytotoxicity (ADCC) as a central therapeutic efficacy (Nuti et al., 2011). Immune cell-mediated effects could be measured with a similar experimental setup including co-cultivation with immune cells like Natural Killer Cells or Macrophages to measure ADCC activation (Beano et al., 2008; Shi et al., 2015). An adverse effect of trastuzumab is cardiotoxicity which would need testing of cardiomyocytes to assess the risk for the biosimilar and compare it to the originator product (T. Y. T. Lee et al., 2024).

RNA-sequencing and proteomics experiments suffer from biases. Transcriptomics data analyses assume that transcript abundance correlates with protein abundance which is often not the case (Guimaraes et al., 2014; Rocca et al., 2015). However, proteomics has a low sensitivity and most often captures only a part of the total proteome of a cell. As we could see in this study with RNA-seq experiment measuring approximately 13000 expressed genes, while the proteomics studies detected fewer than 4000 proteins. Additionally, protein isolation protocols are normally biased due to the different solubility, stability and later ionization of protein and peptides (Glatter et al., 2015). Also, proteomics experiments struggle measuring post-translational modifications (PTMs) unless specifically enriched for phosphorylation or other PTMs. Particularly, phosphoproteomics studies would help investigating the direct downstream signalling of trastuzumab and identify differences because it is mainly dependent on kinases for protein activation (Bose et al., 2006). For both methods it is still common to quantify the data at the gene level rather than the isoform level, losing valuable information (Ezkurdia et al., 2012). In addition, during the analysis of these data set external knowledge will be applied at multiple stages like alignment to the genome and quantification by the annotation, identifying peptides by matching the fragmentation patterns to in-silico digested databases or assuming that pathway and gene set databases are unbiased (Schnoes et al., 2013). Furthermore, these untargeted omics experiments are

explorative and mainly help to build a hypothesis and always need validation by follow-up experiments to confirm findings.

Another limiting factor for scaling the approach developed in this thesis are the costs for RNA-sequencing and mass spectrometry measurement which are still easily over \$100 per sample which makes large scale experiments with multiple cell lines, time points and dosages very cost intensive (Genomics, 2024; Hou et al., 2015). However, the costs are constantly dropping, and novel more cost-effective techniques are in development (Choi et al., 2024; Teng et al., 2024; Veeranagouda et al., 2019). In summary, while the limitations outlined above are significant, they can be addressed through additional experiments and refinements in methodology. These limitations do not undermine the established framework and findings but instead highlight opportunities to enhance the robust and reliable biosimilarity evaluations.

6.6 Final remarks

This thesis highlights several key advances. Overall, untargeted omics methods, such as RNA-sequencing and proteomics, especially when integrated, are highly valuable tools for investigating biosimilarity and can be readily reproduced for similar research questions including downstream analysis. In this thesis, the test case of ApoTras, a biosimilar of Herceptin, showed high overall similarity in both its transcriptomic and proteomic profiles. Furthermore, the pathway profile can be reused for other Herceptin biosimilars in development, thereby reducing experimental redundancy. This dataset uniquely combines high numbers of replicates, deep sequencing, a clinically relevant biosimilar model, and matched proteomic data. This makes the dataset highly valuable for other biosimilar developments of Herceptin. The value of untargeted omics approaches becomes clear in Chapter 4 and 5, highlighting the complexity of trastuzumab's mode of action, that cannot be fully captured by targeted wet-lab experiments. It is anticipated that the regulatory agencies, such as FDA, consider requiring untargeted omics experiments for biosimilar applications. The preferred option would be that the regulatory agencies provide standardized transcriptomic and proteomic benchmarks for the development of biosimilars to reduce repetitive experiments. For Herceptin, extending the framework to include additional cell lines would enhance HER2+ subtype coverage. For example, AU565 represents a more aggressive HER2-enriched subtype, while MDA-MB-361 models high ER/PR expression providing insights into receptor cross-talk mechanisms. Overall, with more patents of monoclonal antibodies ending the development of biosimilars will become increasingly important, and this work may help streamline this process.

References

- Agostinetto, E., Gligorov, J., & Piccart, M. (2022). Systemic therapy for early-stage breast cancer: learning from the past to build the future. *Nature Reviews Clinical Oncology*, 19(12), 763–774.
- Albanell, J., Codony, J., Rovira, A., Mellado, B., & Gascón, P. (2003). Mechanisms of action of trastuzumab and scientific update. *Seminars in Oncology*, 30, 3–14.
- Alfieri, C., Chang, L., Zhang, Z., Yang, J., Maslen, S., Skehel, M., & Barford, D. (2016). Molecular basis of APC/C regulation by the spindle assembly checkpoint. *Nature*, 536(7617), 431–436.
- Al-Qasem, A. J., Alves, C. L., & Ditzel, H. J. (2021). Resistance mechanisms to combined CDK4/6 inhibitors and endocrine therapy in ER+/HER2- advanced breast cancer: Biomarkers and potential novel treatment strategies. *Cancers*, 13(21), 5397.
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*.
<https://doi.org/10.1093/bioinformatics/btu638>
- Anderson, W. F., Rosenberg, P. S., Menashe, I., Mitani, A., & Pfeiffer, R. M. (2008). Age-specific incidence of breast cancer subtypes: understanding the black-white crossover. *Journal of the National Cancer Institute*, 100(23), 1804–1814.
- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*.
<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Antoniou, A., Pharoah, P. D. P., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., Loman, N., Olsson, H., Johannsson, O., Borg, Å., Pasini, B., Radice, P., Manoukian, S., Eccles, D. M., Tang, N., Olah, E., Anton-Culver, H., Warner, E., Lubinski, J., ... Easton, D. F. (2003). Average Risks of Breast and Ovarian Cancer Associated with BRCA1 or BRCA2 Mutations Detected in Case Series Unselected for Family History: A Combined Analysis of 22 Studies. *The American Journal of Human Genetics*, 72(5), 1117–1130. <https://doi.org/10.1086/375033>

- Apweiler, R. (2008). The Universal Protein resource (UniProt). *Nucleic Acids Research*, 36(SUPPL. 1). <https://doi.org/10.1093/nar/gkm895>
- Arnould, L., Gelly, M., Penault-Llorca, F., Benoit, L., Bonnetain, F., Migeon, C., Cabaret, V., Fermeaux, V., Bertheau, P., Garnier, J., Jeannin, J.-F., & Coudert, B. (2006). Trastuzumab-based treatment of HER2-positive breast cancer: an antibody-dependent cellular cytotoxicity mechanism? *British Journal of Cancer*, 94(2), 259–267. <https://doi.org/10.1038/sj.bjc.6602930>
- Arora, S., Narayan, P., Osgood, C. L., Wedam, S., Prowell, T. M., Gao, J. J., Shah, M., Krol, D., Wahby, S., & Royce, M. (2022). US FDA drug approvals for breast cancer: a decade in review. *Clinical Cancer Research*, 28(6), 1072–1086.
- Arteaga, C. L., & Engelman, J. A. (2023). Molecular Pathways and Mechanisms of HER2 in Cancer Therapy. *Cancer Discovery*, 13(7), 1648–1671.
- Arteaga, C. L., Sliwkowski, M. X., Osborne, C. K., Perez, E. A., Puglisi, F., & Gianni, L. (2012). Treatment of HER2-positive breast cancer: current status and future perspectives. *Nature Reviews Clinical Oncology*, 9(1), 16–32.
- Auclair, J., & Rathore, A. (2022). *Analytical Characterization of Host Cell Proteins (HCPs)*.
- Avila-Portillo, L. M., Aristizabal, F., Perdomo, S., Riveros, A., Ospino, B., Avila, J. P., Butti, M., & Abba, M. C. (2020). Comparative analysis of the biosimilar and innovative G-CSF modulated pathways on umbilical cord blood–derived mononuclear cells. *Bioinformatics and Biology Insights*, 14, 1177932220913307.
- Badia-I-Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., Taus, P., Dugourd, A., Holland, C. H., Ramirez Flores, R. O., & Saez-Rodriguez, J. (2022). decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances*, 2(1). <https://doi.org/10.1093/bioadv/vbac016>
- Balduzzi, S., Mantarro, S., Guarneri, V., Tagliabue, L., Pistotti, V., Moja, L., & D’Amico, R. (2014). Trastuzumab in the treatment of HER2-positive metastatic breast cancer: a meta-analysis of randomized controlled trials. *Cancer Treatment Reviews*, 40(1), 123–132.

- Barkett, M., & Gilmore, T. D. (1999). Control of apoptosis by Rel/NF- κ B transcription factors. *Oncogene*, 18(49), 6910–6924. <https://doi.org/10.1038/sj.onc.1203238>
- Baruzzo, G., Hayer, K. E., Kim, E. J., DI Camillo, B., Fitzgerald, G. A., & Grant, G. R. (2017). Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nature Methods*, 14(2). <https://doi.org/10.1038/nmeth.4106>
- Baselga, J., & Albanell, J. (2001). Mechanism of action of trastuzumab and scientific update. *Seminars in Oncology*, 28, 4–11.
- Bastien, R. R., Rodríguez-Lescure, Á., Ebbert, M. T., Prat, A., Munárriz, B., Rowe, L., Miller, P., Ruiz-Borrego, M., Anderson, D., Lyons, B., Álvarez, I., Dowell, T., Wall, D., Seguí, M. Á., Barley, L., Boucher, K. M., Alba, E., Pappas, L., Davis, C. A., ... Martín, M. (2012). PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC Medical Genomics*, 5. <https://doi.org/10.1186/1755-8794-5-44>
- Baumann, J., Kokabee, M., Wong, J., Balasubramaniyam, R., Sun, Y., & Conklin, D. S. (2018). Global metabolite profiling analysis of lipotoxicity in HER2/neu-positive breast cancer cells. *Oncotarget*, 9(43). <https://doi.org/10.18632/oncotarget.25500>
- Beano, A., Signorino, E., Evangelista, A., Brusa, D., Mistrangelo, M., Polimeni, M. A., Spadi, R., Donadio, M., Ciuffreda, L., & Matera, L. (2008). Correlation between NK function and response to trastuzumab in metastatic breast cancer patients. *Journal of Translational Medicine*, 6(1), 1–9.
- Beckwith, H., & Yee, D. (2014). Insulin-like growth factors, insulin, and growth hormone signaling in breast cancer: implications for targeted therapy. *Endocrine Practice*, 20(11), 1214–1221.
- Berkowitz, S. A., Engen, J. R., Mazzeo, J. R., & Jones, G. B. (2012). Analytical strategies for the development of biosimilars. *Nature Reviews Drug Discovery*, 11(7), 527–540.
- Biertümpfel, C., Zhao, Y., Kondo, Y., Ramón-Maiques, S., Gregory, M., Lee, J. Y., Masutani, C., Lehmann, A. R., Hanaoka, F., & Yang, W. (2010). Structure and mechanism of human DNA polymerase η . *Nature*, 465(7301), 1044–1048.

- Blasco, M. A. (2005). Telomeres and human disease: ageing, cancer and beyond. *Nature Reviews Genetics*, 6(8), 611–622.
- Bose, R., Kavuri, S. M., Searleman, A. C., Shen, W., Shen, D., Koboldt, D. C., Monsey, J., Goel, N., Aronson, A. B., Li, S., Ma, M., Ding, J. M., Larson, J. M., Ding, L., Mardis, E. R., Ellis, M. J., Ley, T. J., Wilson, R. K., Cooper, M. A., & Ma, C. X. (2013). Activating HER2 Mutations in HER2 Gene Amplification Negative Breast Cancer. *Cancer Discovery*, 3(2), 224–237. <https://doi.org/10.1158/2159-8290.CD-12-0349>
- Bose, R., Molina, H., Scott Patterson, A., Bitok, J. K., Periaswamy, B., Bader, J., Pandey, A., & Cole, P. A. (2006). Phosphoproteomic analysis of Her2/neu signaling and inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(26). <https://doi.org/10.1073/pnas.0603948103>
- Bousquet, G., El Bouchtaoui, M., Leboeuf, C., Battistella, M., Varna, M., Ferreira, I., Plassa, L.-F., Hamdan, D., Bertheau, P., & Janin, A. (2019). Pan-cancer analysis of genomic and transcriptomic data reveals the heterogeneous landscape of proteasome genes. *Cancers*, 11(9), 1280. <https://doi.org/10.3390/cancers11091280>
- Boyd, N. F., Martin, L. J., Yaffe, M. J., & Minkin, S. (2011). Mammographic density and breast cancer risk: current understanding and future prospects. *Breast Cancer Research*, 13(6), 1–12.
- Bradley, R., Braybrooke, J., Gray, R., Hills, R., Liu, Z., Peto, R., Davies, L., Dodwell, D., McGale, P., Pan, H., Taylor, C., Anderson, S., Gelber, R., Gianni, L., Jacot, W., Joensuu, H., Moreno-Aspitia, A., Piccart, M., Press, M., ... Swain, S. M. (2021). Trastuzumab for early-stage, HER2-positive breast cancer: a meta-analysis of 13 864 women in seven randomised trials. *The Lancet Oncology*, 22(8), 1139–1150. [https://doi.org/https://doi.org/10.1016/S1470-2045\(21\)00288-6](https://doi.org/https://doi.org/10.1016/S1470-2045(21)00288-6)
- Brasó-Maristany, F., Griguolo, G., Pascual, T., Paré, L., Nuciforo, P., Llombart-Cussac, A., Bermejo, B., Oliveira, M., Morales, S., Martínez, N., Vidal, M., Adamo, B., Martínez, O., Pernas, S., López, R., Muñoz, M., Chic, N., Galván, P., Garau, I., ... Prat, A. (2020). Phenotypic changes of HER2-positive breast cancer during and after dual HER2 blockade. *Nature Communications*. <https://doi.org/10.1038/s41467-019-14111-3>

- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. <https://doi.org/10.1038/nbt.3519>
- Bushnell, B. (2016). *BBtools Software Package*.
<https://doi.org/10.1080/00498250701449431>
- Butler, M. (2005). Animal cell cultures: Recent achievements and perspectives in the production of biopharmaceuticals. *Applied Microbiology and Biotechnology*, 68(3), 283–291. <https://doi.org/10.1007/s00253-005-1980-8>
- Calle, E. E., & Kaaks, R. (2004). Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nature Reviews Cancer*, 4(8), 579–591.
- Cameron, D., Piccart-Gebhart, M. J., Gelber, R. D., Procter, M., Goldhirsch, A., de Azambuja, E., Castro Jr, G., Untch, M., Smith, I., & Gianni, L. (2017). 11-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: final analysis of the HERceptin Adjuvant (HERA) trial. *The Lancet*, 389(10075), 1195–1205.
- Candi, E., Schmidt, R., & Melino, G. (2005). The cornified envelope: a model of cell death in the skin. *Nature Reviews Molecular Cell Biology*, 6(4), 328–340.
- Cao, S.-S., & Lu, C.-T. (2016). Recent perspectives of breast cancer prognosis and predictive factors. *Oncology Letters*, 12(5), 3674–3678.
- Cao, Y., Li, J., Zhang, G., Fang, H., Du, Y., & Liang, Y. (2024). KLF15 transcriptionally activates LINC00689 to inhibit colorectal cancer development. *Communications Biology*, 7(1), 130.
- Carey, L. A., Ollila, D. W., Anders, C. K., Hoadley, K. A., Iglesia, M., Perou, C. M., Cirrincione, C. T., Pitcher, B. N., Berry, D. A., Barry, W. T., Krop, I. E., Winer, E. P., Harris, L. N., Henry, N. L., Weckstein, D. J., Singh, B., Hudis, C. A., & Cheang, M. C. U. (2016). Molecular heterogeneity and response to neoadjuvant human epidermal growth factor receptor 2 targeting in CALGB 40601, a randomized phase III trial of paclitaxel plus trastuzumab with or without lapatinib. *Journal of Clinical Oncology*, 34(6), 542–549. <https://doi.org/10.1200/JCO.2015.62.1268>

- Carmeliet, P., & Jain, R. K. (2000). Angiogenesis in cancer and other diseases. *Nature*, 407(6801), 249–257.
- Carracedo, A., Cantley, L. C., & Pandolfi, P. P. (2013). Cancer metabolism: fatty acid oxidation in the limelight. *Nature Reviews Cancer*, 13(4), 227–232.
- Center for Drug Evaluation and Research. (2015). Scientific Considerations in Demonstrating Biosimilarity to a Reference Product. *FDA*.
- Chaffer, C. L., & Weinberg, R. A. (2011). A perspective on cancer cell metastasis. *Science*, 331(6024), 1559–1564.
- Chang, C. A., Jen, J., Jiang, S., Sayad, A., Mer, A. S., Brown, K. R., Nixon, A. M. L., Dhabaria, A., Tang, K. H., Venet, D., Sotiriou, C., Deng, J., Wong, K. K., Adams, S., Meyn, P., Heguy, A., Skok, J. A., Tsirigos, A., Ueberheide, B., ... Neel, B. G. (2022). Ontogeny and Vulnerabilities of Drug-Tolerant Persisters in HER2+ Breast Cancer. *Cancer Discovery*, 12(4). <https://doi.org/10.1158/2159-8290.CD-20-1265>
- Cheang, M. C. U., Chia, S. K., Voduc, D., Gao, D., Leung, S., Snider, J., Watson, M., Davies, S., Bernard, P. S., Parker, J. S., Perou, C. M., Ellis, M. J., & Nielsen, T. O. (2009). Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer. *JNCI: Journal of the National Cancer Institute*, 101(10), 736–750. <https://doi.org/10.1093/jnci/djp082>
- Chen, H., Yu, Z., Niu, Y., Wang, L., Xu, K., & Liu, J. (2023). Research progress of PBX1 in developmental and regenerative medicine. *International Journal of Medical Sciences*, 20(2), 225.
- Chen, Z., Han, F., Du, Y., Shi, H., & Zhou, W. (2023). Hypoxic microenvironment in cancer: molecular mechanisms and therapeutic interventions. *Signal Transduction and Targeted Therapy*, 8(1), 70.
- Choi, J., Hyun, J., Hyun, J., Kim, J. H., Lee, J. H., & Bang, D. (2024). Cost and time-efficient construction of a 3'-end mRNA library from unpurified bulk RNA in a single tube. *Experimental and Molecular Medicine*, 56(2). <https://doi.org/10.1038/s12276-024-01164-8>

- Chow, S.-C. (2018). *Analytical similarity assessment in biosimilar product development*. Chapman and Hall/CRC.
- Chumsae, C., Gaza-Bulseco, G., Sun, J., & Liu, H. (2007). Comparison of methionine oxidation in thermal stability and chemically stressed samples of a fully human monoclonal antibody. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 850(1–2), 285–294.
<https://doi.org/10.1016/j.jchromb.2006.11.050>
- Ciriello, G., Gatza, M. L., Beck, A. H., Wilkerson, M. D., Rhie, S. K., Pastore, A., Zhang, H., McLellan, M., Yau, C., Kandoth, C., Bowlby, R., Shen, H., Hayat, S., Fieldhouse, R., Lester, S. C., Tse, G. M. K., Factor, R. E., Collins, L. C., Allison, K. H., ... Perou, C. M. (2015). Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*. <https://doi.org/10.1016/j.cell.2015.09.033>
- Clavero-Álvarez, A., Di Mambro, T., Perez-Gaviro, S., Magnani, M., & Bruscolini, P. (2018). Humanization of antibodies using a statistical inference approach. *Scientific Reports*, 8(1), 14820.
- Clynes, R. a, Towers, T. L., Presta, L. G., & Ravetch, J. V. (2000). Inhibitory Fc receptors modulate in vivo cytotoxicity against tumor targets. *Nature Medicine*, 6(4), 443–446.
<https://doi.org/10.1038/74704>
- Colegrave, N., & Ruxton, G. D. (2003). Confidence intervals are a more useful complement to nonsignificant tests than are power calculations. In *Behavioral Ecology*.
<https://doi.org/10.1093/beheco/14.3.446>
- Collins, D. M., O'donovan, N., McGowan, P. M., O'sullivan, F., Duffy, M. J., & Crown, J. (2012). Trastuzumab induces antibody-dependent cell-mediated cytotoxicity (ADCC) in HER-2-non-amplified breast cancer cell lines. *Annals of Oncology*, 23(7), 1788–1795. <https://doi.org/10.1093/annonc/mdr484>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., Mortazavi, A., Brazma, A., Gingeras, T. R., Marioni, J. C., Robinson, M. D., Love, M. I., Huber, W., Anders, S.,

- Morgan, J. T. A. T., ... Eisen, M. B. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33(18). <https://doi.org/10.1093/bioinformatics/btx364>
- Cortese, K., Ponassi, M., Profumo, A., Coronel Vargas, G., Iervasi, E., Gagliani, M. C., Bellese, G., Tavella, S., & Castagnola, P. (2023). Lipid Metabolism Reprogramming and Trastuzumab Resistance in Breast Cancer Cell Lines Overexpressing the ERBB2 Membrane Receptor. *Membranes*, 13(6), 540.
- Coussens, L. M., & Werb, Z. (2002). Inflammation and cancer. *Nature*, 420(6917), 860–867.
- Craig, R., & Beavis, R. C. (2004). TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics*, 20(9). <https://doi.org/10.1093/bioinformatics/bth092>
- Cui, P., Lin, Q., Ding, F., Xin, C., Gong, L., Zhang, L., Gao, F., Wang, J., Shen, N., Deng, X., Zhao, H., Fu, J., Wang, W., Xu, X., Yang, H., Wang, J., & Zhang, X. (2018). Comparative performance of four RNA-seq platforms for analysis of human whole blood samples. *Scientific Reports*, 8(1), 1–10. <https://doi.org/10.1038/s41598-018-23226-4>
- Dai, L., Mugaanyi, J., Zhang, T., Tong, J., Cai, X., Lu, C., & Lu, C. (2022). A pan-cancer bioinformatic analysis of the carcinogenic role of SMARCA1 in human carcinomas. *PLoS One*, 17(9), e0274823.
- Dai, X., Cheng, H., Bai, Z., & Li, J. (2017). Breast cancer cell line classification and Its relevance with breast tumor subtyping. In *Journal of Cancer*. <https://doi.org/10.7150/jca.18457>
- Dall’Acqua, W. F., Damschroder, M. M., Zhang, J., Woods, R. M., Widjaja, L., Yu, J., & Wu, H. (2005). Framework shuffling of antibodies to reduce immunogenicity and manipulate functional and biophysical properties. *Methods*, 36(1), 43–60.

- Dang, Y., Yu, J., Zhao, S., Cao, X., & Wang, Q. (2022). HOXA7 promotes the metastasis of KRAS mutant colorectal cancer by regulating myeloid-derived suppressor cells. *Cancer Cell International*, 22(1), 88.
- Das, C., Bhattacharya, A., Adhikari, S., Mondal, A., Mondal, P., Adhikary, S., Roy, S., Ramos, K., Yadav, K. K., & Tainer, J. A. (2024). A prismatic view of the epigenetic-metabolic regulatory axis in breast cancer therapy resistance. *Oncogene*, 1–15.
<https://doi.org/10.1038/s41388-024-03213-4>
- Dawkins, J. B. N., Wang, J., Maniati, E., Heward, J. A., Koniali, L., Kocher, H. M., Martin, S. A., Chelala, C., Balkwill, F. R., Fitzgibbon, J., & Grose, R. P. (2016). Reduced Expression of Histone Methyltransferases KMT2C and KMT2D Correlates with Improved Outcome in Pancreatic Ductal Adenocarcinoma. *Cancer Research*, 76(16), 4861–4871. <https://doi.org/10.1158/0008-5472.CAN-16-0481>
- Delord, J. P., Allal, C., Canal, M., Mery, E., Rochaix, P., Hennebelle, I., Pradines, A., Chatelut, E., Bugat, R., Guichard, S., & Canal, P. (2005). Selective inhibition of HER2 inhibits AKT signal transduction and prolongs disease-free survival in a micrometastasis model of ovarian carcinoma. *Annals of Oncology*, 16(12), 1889–1897.
<https://doi.org/10.1093/annonc/mdi405>
- Desvergne, B., & Wahli, W. (1999). Peroxisome proliferator-activated receptors: nuclear control of metabolism. *Endocrine Reviews*, 20(5), 649–688.
- Deutsch, E. W., Mendoza, L., Shteynberg, D. D., Hoopmann, M. R., Sun, Z., Eng, J. K., & Moritz, R. L. (2023). Trans-Proteomic Pipeline: Robust Mass Spectrometry-Based Proteomics Data Analysis Suite. *Journal of Proteome Research*, 22(2).
<https://doi.org/10.1021/acs.jproteome.2c00624>
- Dey, N., Williams, C., Leyland-Jones, B., & De, P. (2015). A critical role for HER3 in HER2-amplified and non-amplified breast cancers: function of a kinase-dead RTK. *American Journal of Translational Research*, 7(4), 733.
- Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*. <https://doi.org/10.2307/1932409>

- DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*.
<https://doi.org/10.1214/ss/1032280214>
- Diermeier-Daucher, S., Breindl, S., Buchholz, S., Ortmann, O., & Brockhoff, G. (2011). Modular anti-EGFR and anti-Her2 targeting of SK-BR-3 and BT474 breast cancer cell lines in the presence of ErbB receptor-specific growth factors. *Cytometry Part A*, 79(9), 684–693.
- Dillon, T. M., Ricci, M. S., Vezina, C., Flynn, G. C., Liu, Y. D., Rehder, D. S., Plant, M., Henkle, B., Li, Y., Deechongkit, S., Varnum, B., Wypych, J., Balland, A., & Bondarenko, P. V. (2008). Structural and functional characterization of disulfide isoforms of the human IgG2 subclass. *Journal of Biological Chemistry*, 283(23), 16206–16215. <https://doi.org/10.1074/jbc.M709988200>
- DiPaola, M. (2017). The impact of biosimilars on the biopharmaceuticals industry. *GEN-Genetic Engineering & Biotechnology News*, 37(1), 10–11.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dong, C., Wu, J., Chen, Y., Nie, J., & Chen, C. (2021). Activation of PI3K/AKT/mTOR Pathway Causes Drug Resistance in Breast Cancer. In *Frontiers in Pharmacology* (Vol. 12). <https://doi.org/10.3389/fphar.2021.628690>
- DrugBank. (2023). Trastuzumab: Uses, Interactions, Mechanism of Action. *DrugBank Online*. <https://go.drugbank.com/drugs/DB00072>
- Duan, N., Hua, Y., Yan, X., He, Y., Zeng, T., Gong, J., Fu, Z., Li, W., & Yin, Y. (2024). Unveiling Alterations of Epigenetic Modifications and Chromatin Architecture Leading to Lipid Metabolic Reprogramming during the Evolutionary Trastuzumab Adaptation of HER2-Positive Breast Cancer. *Advanced Science*, 11(18).
<https://doi.org/10.1002/advs.202309424>
- Duan, X., Chen, H., Ma, H., & Song, Y. (2017). The expression and significance of the HOXA7 gene in oral squamous cell carcinoma. *Journal of Oral Science*, 59(3), 329–335.

- Duro-Sánchez, S., Alonso, M. R., & Arribas, J. (2023). Immunotherapies against HER2-Positive Breast Cancer. In *Cancers* (Vol. 15, Issue 4).
<https://doi.org/10.3390/cancers15041069>
- Eccles, S. A. (2001). The role of c-erbB-2/HER2/neu in breast cancer progression and metastasis. *Journal of Mammary Gland Biology and Neoplasia*, 6(4), 393–406.
<https://doi.org/10.1023/A:1014730829872>
- Ecker, D. M., Jones, S. D., & Levine, H. L. (2015). The therapeutic monoclonal antibody market. *MAbs*, 7(1), 9–14.
- Ecsédi, P., Kiss, B., Gógl, G., Radnai, L., Buday, L., Koprivanacz, K., Liliom, K., Leveles, I., Vértessy, B., Jeszenői, N., Hetényi, C., Schlosser, G., Katona, G., & Nyitray, L. (2017). Regulation of the Equilibrium between Closed and Open Conformations of Annexin A2 by N-Terminal Phosphorylation and S100A4-Binding. *Structure*, 25(8), 1195–1207.e5. <https://doi.org/10.1016/j.str.2017.06.001>
- Edgar, B. S., Cheifetz, A. S., Helfgott, S. M., Smith, G. P., Bandekar, T., Hoye-Simek, A., Liu, Y., Singh, R. M., Fajardo, K. I., Carter, J., & others. (2021). Overcoming barriers to biosimilar adoption: real-world perspectives from a national payer and provider initiative. *Journal of Managed Care & Specialty Pharmacy*, 27(8), 1129–1135.
- Eferl, R., & Wagner, E. F. (2003). AP-1: a double-edged sword in tumorigenesis. *Nature Reviews Cancer*, 3(11), 859–868.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1–26. <https://doi.org/10.1214/aos/1176344552>
- Eijkelenboom, A., & Burgering, B. M. T. (2013). FOXO: linking new signaling pathways. *Nature Reviews Molecular Cell Biology*, 14(3), 180–191.
- Eisenberg, E., & Levanon, E. Y. (2013). Human housekeeping genes, revisited. *Trends in Genetics*, 29(10), 569–574. <https://doi.org/10.1016/j.tig.2013.05.010>
- Eng, J. K., Hoopmann, M. R., Jahan, T. A., Egertson, J. D., Noble, W. S., & MacCoss, M. J. (2015). A Deeper Look into Comet - Implementation and Features. *Journal of the American Society for Mass Spectrometry*, 26(11). <https://doi.org/10.1007/s13361-015-1179-x>

- Engel, R. H., & Kaklamani, V. G. (2007). HER2-positive breast cancer: current and future treatment strategies. *Drugs*, 67(9), 1329–1341.
- European Medicines Agency. (2012). Guideline on Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins As Active Substance: Quality Issues. In *E.M.A. Publication*. <https://doi.org/10.1136/bmj.333.7574.873-a>
- European Medicines Agency. (2017). *Ontruzant: Summary of Product Characteristics*. <https://www.ema.europa.eu>
- European Medicines Agency. (2019). *Biosimilars in the EU: Information guide for healthcare professionals*.
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Ewer, M. S., & O’Shaughnessy, J. A. (2007). Cardiac toxicity of trastuzumab-related regimens in HER2-overexpressing breast cancer. *Clinical Breast Cancer*, 7(8), 22–29.
- Ezkurdia, I., Del Pozo, A., Frankish, A., Rodriguez, J. M., Harrow, J., Ashman, K., Valencia, A., & Tress, M. L. (2012). Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. *Molecular Biology and Evolution*, 29(9). <https://doi.org/10.1093/molbev/mss100>
- Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., K€orninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., ... D’Eustachio, P. (2016). The reactome pathway knowledgebase. *Nucleic Acids Research*, 44(D1), D481–D487. <https://doi.org/10.1093/nar/gkv1351>
- Faltus, T., Yuan, J., Zimmer, B., Kr€amer, A., Loibl, S., Kaufmann, M., & Strebhardt, K. (2004). Silencing of the HER2/neu gene by siRNA inhibits proliferation and induces apoptosis in HER2/neu-overexpressing breast cancer cells. *Neoplasia*, 6(6), 786–795. <https://doi.org/10.1593/neo.04313>
- Fan, S. H., Wang, Y. Y., Lu, J., Zheng, Y. L., Wu, D. M., Zhang, Z. F., Shan, Q., Hu, B., Li, M. Q., & Cheng, W. (2015). CERS2 Suppresses Tumor Cell Invasion and is Associated

- with Decreased V-ATPase and MMP-2/MMP-9 Activities in Breast Cancer. *Journal of Cellular Biochemistry*, 116(4). <https://doi.org/10.1002/jcb.24978>
- Farrell, C. M., O’Leary, N. A., Harte, R. A., Loveland, J. E., Wilming, L. G., Wallin, C., Diekhans, M., Barrell, D., Searle, S. M. J., Aken, B., Hiatt, S. M., Frankish, A., Suner, M. M., Rajput, B., Steward, C. A., Brown, G. R., Bennett, R., Murphy, M., Wu, W., ... Pruitt, K. D. (2014). Current status and new features of the Consensus Coding Sequence database. In *Nucleic Acids Research* (Vol. 42, Issue D1, pp. 865–872). <https://doi.org/10.1093/nar/gkt1059>
- Faubert, B., Solmonson, A., & DeBerardinis, R. J. (2020). Metabolic reprogramming and cancer progression. *Science*, 368(6487), eaaw5473.
- FDA. (2015). Scientific considerations in demonstrating biosimilarity to a reference product. *Guidance for Industry*.
- Fekete, S., Gassner, A.-L., Rudaz, S., Schappler, J., & Guilleme, D. (2013). Analytical strategies for the characterization of therapeutic monoclonal antibodies. *TrAC Trends in Analytical Chemistry*, 42, 74–83.
- Feng, S., Deng, L., Chen, W., Shao, J., Xu, G., & Li, Y.-P. (2017). Osteoclast proton pump regulator Atp6v1c1 enhances breast cancer growth by activating the mTORC1 pathway and bone metastasis by increasing V-ATPase activity. *Oncotarget*, 8(29), 47675.
- Fiszman, G. L., & Jasniz, M. A. (2011). Molecular mechanisms of trastuzumab resistance in HER2 overexpressing breast cancer. *International Journal of Breast Cancer*, 2011(1), 352182.
- Foulkes, W. D., Smith, I. E., & Reis-Filho, J. S. (2010). Triple-negative breast cancer. *New England Journal of Medicine*, 363(20), 1938–1948.
- Frenzel, A., Hust, M., & Schirrmann, T. (2013). Expression of recombinant antibodies. *Frontiers in Immunology*, 4, 217. <https://doi.org/10.3389/fimmu.2013.00217>
- Friedenreich, C. M., & Cust, A. E. (2008). Physical activity and breast cancer risk: impact of timing, type and dose of activity and population subgroup effects. *British Journal of Sports Medicine*, 42(8), 636–647.

- Frittoli, E., Palamidessi, A., Marighetti, P., Confalonieri, S., Bianchi, F., Malinverno, C., Mazzarol, G., Viale, G., Martin-Padura, I., Garré, M., Parazzoli, D., Mattei, V., Cortellino, S., Bertalot, G., Di Fiore, P. P., & Scita, G. (2014). A RAB5/RAB4 recycling circuitry induces a proteolytic invasive program and promotes tumor dissemination. *Journal of Cell Biology*, 206(2), 307–328.
<https://doi.org/10.1083/jcb.201403127>
- Frontini, M., Vijayakumar, M., Garvin, A., & Clarke, N. (2009). A ChIP–chip approach reveals a novel role for transcription factor IRF1 in the DNA damage response. *Nucleic Acids Research*, 37(4), 1073–1085. <https://doi.org/10.1093/nar/gkn1051>
- Fruman, D. A., & Rommel, C. (2014). PI3K and cancer: lessons, challenges and opportunities. *Nature Reviews Drug Discovery*, 13(2), 140–156.
- Fuhs, S. R., & Hunter, T. (2017). The NM23/NDPK family of metastasis suppressor genes. *Trends in Molecular Medicine*, 23(10), 963–975.
- Furukawa-Hibi, Y., Kobayashi, Y., Chen, C., & Motoyama, N. (2005). FOXO transcription factors in cell-cycle regulation and the response to oxidative stress. *Antioxidants & Redox Signaling*, 7(5–6), 752–760.
- Gajria, D., & Chandarlapaty, S. (2011). HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies. *Expert Review of Anticancer Therapy*, 11(2), 263–275.
- Gallardo, A., Lerma, E., Escuin, D., Tibau, A., Muñoz, J., Ojeda, B., Barnadas, A., Adrover, E., Sánchez-Tejada, L., Giner, D., Ortiz-Martínez, F., & Peiró, G. (2012). Increased signalling of EGFR and IGF1R, and deregulation of PTEN/PI3K/Akt pathway are related with trastuzumab resistance in HER2 breast carcinomas. *British Journal of Cancer*, 106(8). <https://doi.org/10.1038/bjc.2012.85>
- Gandhi, N., & Das, G. M. (2019). Metabolic reprogramming in breast cancer and its therapeutic implications. In *Cells* (Vol. 8, Issue 2).
<https://doi.org/10.3390/cells8020089>
- Gawronska-Kozak, B. (2020). Foxn1 control of skin function. *Applied Sciences*, 10(16), 5685.

- Gawronska-Kozak, B., Bukowska, J., Kur-Piotrowska, A., & Kozak, W. (2019). FOXN1 transcription factor in epithelial growth and wound healing. *International Journal of Molecular Sciences*, 20(7), 1659.
- Gaza-Bulseco, G., Faldu, S., Hurkmans, K., Chumsae, C., & Liu, H. (2008). Effect of methionine oxidation of a recombinant monoclonal antibody on the binding affinity to protein A and protein G. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 870(1), 55–62.
<https://doi.org/10.1016/j.jchromb.2008.05.045>
- Gendoo, D. M. A., Ratanasirigulchai, N., Schröder, M. S., Paré, L., Parker, J. S., Prat, A., & Haibe-Kains, B. (2016). Genefu: An R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics*, 32(7).
<https://doi.org/10.1093/bioinformatics/btv693>
- Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(90001), 258D – 261.
<https://doi.org/10.1093/nar/gkh036>
- Generics and Biosimilars Initiative (GaBI). (2011). *US\$67 billion worth of biosimilar patents expiring before 2020*. <http://www.gabionline.net/Biosimilars/General/US-67-billion-worth-of-biosimilar-patents-expiring-before-2020>
- Genomics, A. (2024). *The cost of RNA-sequencing (RNA-seq) ranges from approximately 36.9 to 173 for a single sample in an mRNA-seq experiment*.
<https://www.linkedin.com/pulse/budgeting-mrna-seq-project-here-main-cost-drivers-keep-j15ve/>
- Ginestet, C. (2011). ggplot2: Elegant Graphics for Data Analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. https://doi.org/10.1111/j.1467-985x.2010.00676_9.x
- Glatter, T., Ahrné, E., & Schmidt, A. (2015). Comparison of different sample preparation protocols reveals lysis buffer-specific extraction biases in gram-negative bacteria and human cells. *Journal of Proteome Research*, 14(11).
<https://doi.org/10.1021/acs.jproteome.5b00654>

- Gómez-Cebrián, N., Domingo-Ortíz, I., Poveda, J. L., Vicent, M. J., Puchades-Carrasco, L., & Pineda-Lucena, A. (2021). Multi-omic approaches to breast cancer metabolic phenotyping: Applications in diagnosis, prognosis, and the development of novel treatments. *Cancers*, 13(18), 4544.
- Gordon, V., & Banerji, S. (2013). Molecular pathways: PI3K pathway targets in triple-negative breast cancers. *Clinical Cancer Research*, 19(14), 3738–3744.
- Goutsouliak, K., Veeraraghavan, J., Sethunath, V., De Angelis, C., Osborne, C. K., Rimawi, M. F., & Schiff, R. (2020). Towards personalized treatment for early stage HER2-positive breast cancer. In *Nature Reviews Clinical Oncology* (Vol. 17, Issue 4). <https://doi.org/10.1038/s41571-019-0299-9>
- Gu, Y., Li, J., Guo, D., Chen, B., Liu, P., Xiao, Y., Yang, K., Liu, Z., & Liu, Q. (2020). Identification of 13 key genes correlated with progression and prognosis in hepatocellular carcinoma by weighted gene co-expression network analysis. *Frontiers in Genetics*, 11, 153.
- Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18). <https://doi.org/10.1093/bioinformatics/btw313>
- Guimaraes, J. C., Rocha, M., & Arkin, A. P. (2014). Transcript level and sequence determinants of protein abundance and noise in Escherichia coli. *Nucleic Acids Research*, 42(8), 4791–4799.
- Guo, Y., Zhao, S., Li, C. I., Sheng, Q., & Shyr, Y. (2014). RNAseqPS: A web tool for estimating sample size and power for RNAseq experiment. *Cancer Informatics*. <https://doi.org/10.4137/CIN.S17688>
- Haider, S., & Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. *Current Genomics*, 14(2), 91–110.
- Hall, P., & Wilson, S. R. (1991). Two Guidelines for Bootstrap Hypothesis Testing. *Biometrics*. <https://doi.org/10.2307/2532163>
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5), 646–674.

- Hansel, T. T., Kropshofer, H., Singer, T., Mitchell, J. A., & George, A. J. T. (2010). The safety and side effects of monoclonal antibodies. *Nature Reviews Drug Discovery*, 9(4), 325–338.
- Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology*, 18(1), 1–15.
- He, Y., Sun, M. M., Zhang, G. G., Yang, J., Chen, K. S., Xu, W. W., & Li, B. (2021). Targeting PI3K/Akt signal transduction for cancer therapy. *Signal Transduction and Targeted Therapy*, 6(1), 425.
- Hendriks, B. S., Klinz, S. G., Reynolds, J. G., Espelin, C. W., Gaddy, D. F., & Wickham, T. J. (2013). Impact of tumor HER2/ERBB2 expression level on HER2-targeted liposomal doxorubicin-mediated drug delivery: multiple low-affinity interactions lead to a threshold effect. *Molecular Cancer Therapeutics*, 12(9), 1816–1828.
- Hiatt, R. A., & Brody, J. G. (2018). Environmental determinants of breast cancer. *Annual Review of Public Health*, 39(1), 113–133.
- Hmiel, L. K., Brorson, K. A., & Boyne, M. T. (2015). Post-translational structural modifications of immunoglobulin G and their effect on biological activity. *Analytical and Bioanalytical Chemistry*, 407(1), 79–94. <https://doi.org/10.1007/s00216-014-8108-x>
- Hodoniczky, J., Yuan, Z. Z., & James, D. C. (2005). Control of recombinant monoclonal antibody effector functions by Fc N-glycan remodeling in vitro. *Biotechnology Progress*, 21(6), 1644–1652. <https://doi.org/10.1021/bp050228w>
- Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *American Statistician*. <https://doi.org/10.1198/000313001300339897>
- Holland, C. H., Szalai, B., & Saez-Rodriguez, J. (2020). Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1863(6), 194431. <https://doi.org/https://doi.org/10.1016/j.bbagr.2019.194431>

- Holliday, D. L., & Speirs, V. (2011). Choosing the right cell line for breast cancer research. *Breast Cancer Research*, 13(4), 1–7.
- Hou, Z., Jiang, P., Swanson, S. A., Elwell, A. L., Nguyen, B. K. S., Bolin, J. M., Stewart, R., & Thomson, J. A. (2015). A cost-effective RNA sequencing protocol for large-scale gene expression studies. *Scientific Reports*, 5. <https://doi.org/10.1038/srep09570>
- Huang, A., Xue, H., Xie, T., Xiang, L., Chen, Z., Ma, A., Yan, H., & Yuan, J. (2025). A review of the pathogenesis of mitochondria in breast cancer and progress of targeting mitochondria for breast cancer treatment. *Journal of Translational Medicine*, 23(1), 70. <https://doi.org/10.1186/s12967-025-06077-2>
- Huang, R. et al. (2021). Biological Role of MYCN in Medulloblastoma: Novel Therapeutic Targets and Strategies. *Frontiers in Oncology*, 11, 694320.
- Huber, W., Von Heydebreck, A., Sültmann, H., Poustka, A., & Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18(SUPPL. 1). https://doi.org/10.1093/bioinformatics/18.suppl_1.S96
- Hudis, C. A. (2007). Trastuzumab—mechanism of action and use in clinical practice. *New England Journal of Medicine*, 357(1), 39–51.
- Hudziak, R. M., Schlessinger, J., & Ullrich, A. (1987). Increased expression of the putative growth factor receptor p185HER2 causes transformation and tumorigenesis of NIH 3T3 cells. *Proceedings of the National Academy of Sciences*, 84(20), 7159–7163. <https://doi.org/10.1073/pnas.84.20.7159>
- Hydbring, P., Wang, Y., Fassl, A., Li, X., Matia, V., Otto, T., Choi, Y. J., Sweeney, K. E., Suski, J. M., Yin, H., & others. (2017). Cell-cycle-targeting MicroRNAs as therapeutic tools against refractory cancers. *Cancer Cell*, 31(4), 576–590.
- Hynes, N. E., & Lane, H. A. (2005). ERBB receptors and cancer: The complexity of targeted inhibitors. In *Nature Reviews Cancer* (Vol. 5, Issue 5). <https://doi.org/10.1038/nrc1609>
- Ichikawa, T., Sato, F., Terasawa, K., Tsuchiya, S., Toi, M., Tsujimoto, G., & Shimizu, K. (2012). Trastuzumab produces therapeutic actions by upregulating miR-26a and miR-

- 30b in breast cancer cells. *PLoS ONE*, 7(2), e31422.
<https://doi.org/10.1371/journal.pone.0031422>
- Izumi, Y., Xu, L., di Tomaso, E., Fukumura, D., & Jain, R. K. (2002). Tumour biology: hereceptin acts as an anti-angiogenic cocktail. *Nature*, 416(6878), 279–280.
- Jaccard, P. (1912). THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE. *The New Phytologist*, 11, 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
- Jain, A. K., Dubes, R. C., & Chen, C. C. (1987). Bootstrap Techniques for Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
<https://doi.org/10.1109/TPAMI.1987.4767957>
- Janzen, H. (2019). Forced degradation studies—comparison between ICH, EMA, FDA and WHO guidelines and ANVISA’s resolution RDC 53/2015. *Master’s Thesis, Bonn-Rhein-Sieg University of Applied Sciences*.
- Jefferis, R. (2012). Isotype and glycoform selection for antibody therapeutics. *Archives of Biochemistry and Biophysics*, 526(2), 159–166.
<https://doi.org/10.1016/j.abb.2012.03.021>
- Jensen, E. V., & Jordan, V. C. (2003). Estrogen receptor action in breast cancer. *Endocrine Reviews*, 24(4), 523–538.
- Jerusalem, G., Bachelot, T., Barrios, C., Neven, P., Di Leo, A., Janni, W., & de Boer, R. (2015). A new era of improving progression-free survival with dual blockade in postmenopausal HR+, HER2- advanced breast cancer. *Cancer Treatment Reviews*, 41(2), 94–104.
- Jeyakumar, A., & Younis, T. (2012). Trastuzumab for HER2-positive metastatic breast cancer: clinical and economic considerations. *Clinical Medicine Insights: Oncology*, 6, CMO–S6460.
- Jiang, N., Zhang, X., Qin, D., Yang, J., Wu, A., Wang, L., Sun, Y., Li, H., Shen, X., Lin, J., Kantawong, F., & Wu, J. (2021). Identification of Core Genes Related to Progression and Prognosis of Hepatocellular Carcinoma and Small-Molecule Drug Predication. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.608017>

- Jiang, Y., Li, J., Zeng, X., Ren, Y., & Li, X. (2021). Comprehensive analysis of the molecular mechanism of a novel anti-HER2 ADC using RNA sequencing. *Molecular Therapy-Oncolytics*, 20, 47–56. <https://doi.org/10.1016/j.omto.2020.11.001>
- Jiang, Y., Sun, B., Zhu, X., Wang, Y., Jiao, R., & Zhao, J. (2021). Cytochrome P450: Implications for human breast cancer. *Frontiers in Oncology*, 11, 681320.
- Jiang, Z., Prat, A., Carey, L. A., Adamo, B., Vidal, M., Tabernero, J., Cort'es, J., Parker, J. S., Perou, C. M., & Baselga, J. (2024). Molecular Characterization and Classification of HER2-Positive Breast Cancer Subtypes. *Cancer Discovery*, 14(1), 68–85.
- Johnson, E. M., Kinoshita, Y., & Daniel, D. C. (2003). A new member of the MCM protein family encoded by the human MCM8 gene, located contrapodal to GCD10 at chromosome band 20p12. 3-13. *Nucleic Acids Research*, 31(11), 2915–2925.
- Jönsson, G., Staaf, J., Olsson, E., Heidenblad, M., Vallon-Christersson, J., Osoegawa, K., de Jong, P., Oredsson, S., Ringnér, M., Höglund, M., & Borg, Å. (2007). High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. *Genes, Chromosomes and Cancer*, 46(6), 543–558. <https://doi.org/10.1002/gcc.20438>
- Jung, S. K., Lee, K. H., Jeon, J. W., Lee, J. W., Kwon, B. O., Kim, Y. J., Bae, J. S., Kim, D.-H., Lee, S. Y., & Chang, S. J. (2014). Comparison of the physicochemical and biological properties of a biosimilar granulocyte colony-stimulating factor with its reference product. *Journal of Pharmaceutical Sciences*, 103(10), 3411–3419.
- Kalluri, R., & LeBleu, V. S. (2020). The biology, function, and biomedical applications of exosomes. *Science*, 367(6478), eaau6977. <https://doi.org/10.1126/science.aau6977>
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1). [https://doi.org/10.1016/0020-0190\(89\)90102-6](https://doi.org/10.1016/0020-0190(89)90102-6)
- Kang, H.-N., Thorpe, R., Knezevic, I., Casas Levano, M., Chilufya, M. B., Chirachanakul, P., Chua, H. M., Dalili, D., Foo, F., Gao, K., & others. (2021). Regulatory challenges with biosimilars: an update from 20 countries. *Annals of the New York Academy of Sciences*, 1491(1), 42–59.

- Kar, A., Liu, B., & Gutierrez-Hartmann, A. (2017). ESE-1 knockdown attenuates growth in trastuzumab-resistant HER2+ breast cancer cells. *Anticancer Research*, 37(12), 6583–6591.
- Kauraniemi, P., Hautaniemi, S., Autio, R., Astola, J., Monni, O., Elkahloun, A., & Kallioniemi, A. (2004). Effects of Herceptin treatment on global gene expression patterns in HER2-amplified and nonamplified breast cancer cell lines. *Oncogene*, 23(4). <https://doi.org/10.1038/sj.onc.1207200>
- Kelley, B. (2020). Developing therapeutic monoclonal antibodies at pandemic pace. *Nature Biotechnology*, 38(5), 540–545.
- Kelsey, J. L., Gammon, M. D., & John, E. M. (1993). Reproductive factors and breast cancer. *Epidemiologic Reviews*, 15(1), 36–47.
- Kennecke, H., Yerushalmi, R., Woods, R., Cheang, M. C. U., Voduc, D., Speers, C. H., Nielsen, T. O., & Gelmon, K. (2010). Metastatic behavior of breast cancer subtypes. *Journal of Clinical Oncology*, 28(20), 3271–3277. <https://doi.org/10.1200/JCO.2009.25.9820>
- Kersten, S. (2014). Integrated physiology and systems biology of PPARα. *Molecular Metabolism*, 3(4), 354–371.
- Khateb, H., Alshaer, W., Alqudah, A., & Zihlif, M. (2024). Unveiling the Molecular Mechanism of Trastuzumab Resistance in HER2-Positive Breast Cancer Cell Lines. *Cancers*, 16(7), 1180.
- Kim, M. (2021). CREB1 Identified as Potential Therapeutic Target for the Treatment of Pancreatic Cancer. *AACR Annual Meeting*.
- Kimura, K., Jackson, T. L. B., & Huang, R. C. C. (2023). Interaction and collaboration of SP1, HIF-1, and MYC in regulating the expression of cancer-related genes to further enhance anticancer drug development. *Current Issues in Molecular Biology*, 45(11), 9262–9283.
- King, C., Kraus, M., & Aaronson, S. (1985). Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science*, 229(4717), 974–976. <https://doi.org/10.1126/SCIENCE.2992089>

- Klein, C., Brinkmann, U., Reichert, J. M., & Kontermann, R. E. (2024). The present and future of bispecific antibodies for cancer therapy. *Nature Reviews Drug Discovery*, 23(4), 301–319.
- Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., McMichael, J. F., Fulton, L. L., Dooling, D. J., Ding, L., Mardis, E. R., Wilson, R. K., Ally, A., Balasundaram, M., Butterfield, Y. S. N., Carlsen, R., Carter, C., Chu, A., Chuah, E., Chun, H.-J. E., ... Center, M. D. A. C. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), 61–70. <https://doi.org/10.1038/nature11412>
- Köhler, G., & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, 256(5517), 495–497.
- Kokot, A., Gadakh, S., Saha, I., Gajda, E., Łaźniewski, M., Rakshit, S., Sengupta, K., Mollah, A. F., Denkiewicz, M., Górczak, K., Claesen, J., Burzykowski, T., & Plewczynski, D. (2024). Unveiling the Molecular Mechanism of Trastuzumab Resistance in SKBR3 and BT474 Cell Lines for HER2 Positive Breast Cancer. *Current Issues in Molecular Biology*, 46(3). <https://doi.org/10.3390/cimb46030171>
- Kolde, R. (2022). Package “pheatmap”: Pretty heatmaps. *R Package*.
- Kolitz, S. E., Towfic, F., Grossman, I., Hayden, M. R., & Zeskind, B. (2014). Use of genetic technologies to compare medicines. *Clinical Genetics*, 86(5), 419–427. <https://doi.org/10.1111/cge.12462>
- Kos, I. A., Azevedo, V. F., Neto, D. E., & Kowalski, S. C. (2018). The biosimilars journey: current status and ongoing challenges. *Drugs in Context*, 7, 212543.
- Krishnan, A., Mody, R., & Malhotra, H. (2015). Global regulatory landscape of biosimilars: emerging and established market perspectives. *Biosimilars*, 19–32.
- Kroemer, G., Mariño, G., & Levine, B. (2010). Autophagy and the integrated stress response. *Molecular Cell*, 40(2), 280–293.
- Kuriakose, A., Chirmule, N., & Nair, P. (2016). Immunogenicity of Biotherapeutics: Causes and Association with Posttranslational Modifications. *Journal of Immunology Research*, 2016, 1298473. <https://doi.org/10.1155/2016/1298473>

- Kurozumi, S., Alsaleem, M., Monteiro, C. P., Bhardwaj, K., Joosten, S. E. J., Fujii, T., Shirabe, K., Green, A. R., Ellis, I. O., & Rakha, E. A. (2019). HER2 Gene Amplification in ER-positive HER2 Immunohistochemistry 0 or 1+ Breast Cancer with Early Distant Recurrence. *Anticancer Research*, 39(3), 1485–1493.
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Le, X. F., Almeida, M. I., Mao, W., Spizzo, R., Rossi, S., Nicoloso, M. S., Zhang, S., Wu, Y., Calin, G. A., & Bast, R. C. (2012). Modulation of microrna-194 and cell migration by her2-targeting trastuzumab in breast cancer. *PLoS ONE*, 7(7). <https://doi.org/10.1371/journal.pone.0041170>
- Le, X. F., Lammayot, A., Gold, D., Lu, Y., Mao, W., Chang, T., Patel, A., Mills, G. B., & Bast, R. C. (2005). Genes affecting the cell cycle, growth, maintenance drug sensitivity are preferentially regulated by anti-HER2 antibody through phosphatidylinositol 3-kinase-AKT signaling. *Journal of Biological Chemistry*, 280(3). <https://doi.org/10.1074/jbc.M403080200>
- LeBleu, V. S., & Kalluri, R. (2020). Exosomes as a Multicomponent Biomarker Platform in Cancer. *Cancer Treatment Reviews*, 86, 102019. <https://doi.org/10.1016/j.ctrv.2020.102019>
- Lee, J., Kang, H. A., Bae, J. S., Kim, K. D., Lee, K. H., Lim, K. J., Choo, M. J., & Chang, S. J. (2018). Evaluation of analytical similarity between trastuzumab biosimilar CT-P6 and reference product using statistical analyses. *MAbs*, 10(4), 547–571. <https://doi.org/10.1080/19420862.2018.1440170>
- Lee, T. Y. T., Coles, J. G., & Maynes, J. T. (2024). iPSC-cardiomyocytes in the preclinical prediction of candidate pharmaceutical toxicity. *Frontiers in Pharmacology*, 15, 1308217.
- Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y., & Pietenpol, J. A. (2011). Identification of human triple-negative breast cancer subtypes

- and preclinical models for selection of targeted therapies. *The Journal of Clinical Investigation*, 121(7), 2750–2767.
- Lenferink, A. E. G., Busse, D., Flanagan, W. M., Yakes, F. M., & Arteaga, C. L. (2001). ErbB2/neu kinase modulates cellular p27Kip1 and cyclin D1 through multiple signaling pathways. *Cancer Research*, 61(17), 6583–6591.
- Leone, G., Sears, R., Huang, E., Rempel, R., Nuckolls, F., Park, C.-H., Giangrande, P., Wu, L., Saavedra, H. I., Field, S. J., Thompson, J. R., & Nevins, J. M. (2001). Myc requires distinct E2F activities to induce S phase and apoptosis. *Molecular Cell*, 8(1), 105–113. [https://doi.org/10.1016/S1097-2765\(01\)00287-4](https://doi.org/10.1016/S1097-2765(01)00287-4)
- Lesurf, R., Griffith, O. L., Griffith, M., Hundal, J., Trani, L., Watson, M. A., Aft, R., Ellis, M. J., Ota, D., Suman, V. J., Carey, L. A., Barry, W. T., Berry, D. A., & Perou, C. M. (2017). Genomic characterization of HER2-positive breast cancer and response to neoadjuvant trastuzumab and chemotherapy—results from the ACOSOG Z1041 (Alliance) trial. *Annals of Oncology*, 28(5), 1070–1077. <https://doi.org/10.1093/annonc/mdx064>
- Lewis Phillips, G. D., Li, G., Dugger, D. L., Crocker, L. M., Parsons, K. L., Mai, E., Blättler, W. A., Lambert, J. M., Chari, R. V. J., Lutz, R. J., Wong, W. L. T., Jacobson, F. S., Koeppen, H., Schwall, R. H., Kenkare-Mitra, S. R., Spencer, S. D., & Sliwkowski, M. X. (2008). Targeting HER2-positive breast cancer with trastuzumab-DM1, an antibody-cytotoxic drug conjugate. *Cancer Research*, 68(22). <https://doi.org/10.1158/0008-5472.CAN-08-1776>
- Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), 323. <https://doi.org/10.1186/1471-2105-12-323>
- Li, D., Jin, P., Cai, Y., Wu, S., Guo, X., Zhang, Z., Liu, K., Li, P., Hu, Y., & Zhou, Y. (2025). Clinical significance of lipid pathway-targeted therapy in breast cancer. *Frontiers in Pharmacology*, 15, 1514811.

- Li, F., Vijayasankaran, N., Shen, A., Kiss, R., & Amanullah, A. (2010). Cell culture processes for monoclonal antibody production. *MAbs*, 2(5), 466–479.
<https://doi.org/10.4161/mabs.2.5.12720>
- Li, G., Wang, X., Hibshoosh, H., Jin, C., & Halmos, B. (2019). Targeting of the HER2/HER3 signaling axis overcomes ligand-mediated resistance to trastuzumab in HER2-positive breast cancer. *Cancer Letters*, 473, 156–168.
- Li, H., Prever, L., Hirsch, E., & Gulluni, F. (2021). Targeting PI3K/AKT/mTOR signaling pathway in breast cancer. *Cancers*, 13(14), 3517.
- Li, Y., Yang, X. H., Fang, S. J., Qin, C. F., Sun, R. L., Liu, Z. Y., Jiang, B. Y., Wu, X., & Li, G. (2015). HOXA7 stimulates human hepatocellular carcinoma proliferation through cyclin E1/CDK2. *Oncology Reports*, 33(2), 990–996.
- Li, Y., Zhou, X., & Lyu, Z. (2023). Analysis of two-gene signatures and related drugs in small-cell lung cancer by bioinformatics. *Open Medicine (Poland)*, 18(1).
<https://doi.org/10.1515/med-2023-0806>
- Liang, Q., Deng, H., Sun, C.-W., Townes, T. M., & Zhu, F. (2011). Negative regulation of IRF7 activation by activating transcription factor 4 suggests a cross-regulation between the IFN responses and the cellular integrated stress responses. *The Journal of Immunology*, 186(2), 1001–1010.
- Liedtke, C., Mazouni, C., Hess, K. R., André, F., Tordai, A., Mejia, J. A., Symmans, W. F., Gonzalez-Angulo, A. M., Hennessey, B., Green, M., Cristofanilli, M., Hortobagyi, G. N., & Pusztai, L. (2008). Response to Neoadjuvant Therapy and Long-Term Survival in Patients With Triple-Negative Breast Cancer. *Journal of Clinical Oncology*, 26(8), 1275–1281. <https://doi.org/10.1200/JCO.2007.14.4147>
- Lim, E., Vaillant, F., Wu, D., Forrest, N. C., Pal, B., Hart, A. H., Asselin-Labat, M.-L., Gyorki, D. E., Ward, T., & Partanen, A. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nature Medicine*, 15(8), 907–913.
- Lin, W., Francis, J. M., Li, H., Gao, X., Pedamallu, C. S., Ernst, P., & Meyerson, M. (2016). Kmt2a cooperates with menin to suppress tumorigenesis in mouse pancreatic islets.

- Cancer Biology and Therapy*, 17(12), 1274–1281.
<https://doi.org/10.1080/15384047.2016.1250986>
- Lin, Y., Fu, F., Lv, J., Wang, M., Li, Y., Zhang, J., & Wang, C. (2020). Identification of potential key genes for HER-2 positive breast cancer based on bioinformatics analysis. *Medicine*, 99(1).
- Liu, B., Fan, Z., Edgerton, S. M., Yang, X., Lind, S. E., & Thor, A. D. (2011). Potent anti-proliferative effects of metformin on trastuzumab-resistant breast cancer cells via inhibition of erbB2/IGF-1 receptor interactions. *Cell Cycle*, 10(17), 2959–2966.
- Liu, C., & Morrow, K. J. (2016). Biosimilars of monoclonal antibodies: a practical guide to manufacturing, preclinical, and clinical development. *John Wiley & Sons*.
- Liu, Y., Zhou, J., & White, K. P. (2014). Statistics: The next generation. *Nature Reviews Genetics*, 15(12), 843–856. <https://doi.org/10.1038/nrg3835>
- Livneh, I., Cohen-Kaplan, V., Cohen-Rosenzweig, C., Avni, N., & Ciechanover, A. (2016). Expression of immunoproteasome genes is regulated by cell-intrinsic and –extrinsic factors in human cancers. *Scientific Reports*, 6(1), 34019.
- Loibl, S., & Gianni, L. (2017). Epidemiology and management of HER2-positive breast cancer. *The Lancet*, 389(10087), 2415–2429.
- Lonberg, N. (2008). Fully human antibodies from transgenic mouse and phage display platforms. *Current Opinion in Immunology*, 20(4), 450–459.
<https://doi.org/10.1016/j.coi.2008.06.003>
- Lu, J., Jiang, F., Lu, A., & Zhang, G. (2020). Antibody-drug conjugates: current status and future directions. *MAbs*, 12(1), 1704199.
- Lu, Y., Zi, X., & Pollak, M. (2004). Molecular mechanisms underlying IGF-I-induced attenuation of the growth-inhibitory activity of trastuzumab (herceptin) on SKBR3 breast cancer cells. *International Journal of Cancer*, 108(3), 334–341.
<https://doi.org/10.1002/ijc.11445>
- Lu, Y., Zi, X., Zhao, Y., Mascarenhas, D., & Pollak, M. (2001). Insulin-like growth factor-I receptor signaling and resistance to transtuzumab (Herceptin). *Journal of the National Cancer Institute*, 93(24), 1852–1857. <https://doi.org/10.1093/jnci/93.24.1852>

- Łukasiewicz, S., Czezelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021). Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—An updated review. In *Cancers* (Vol. 13, Issue 17). <https://doi.org/10.3390/cancers13174287>
- Luo, T., Zhong, X., Huang, H., Lv, F., Li, Y., Zhao, H., Wang, X., Xu, F., Zhang, L., & Wang, J. (2021). Clinical features of patients with HER2-positive breast cancer and brain metastases. *BMC Cancer*, 21(1), 1–9. <https://doi.org/10.1186/s12885-021-07890-4>
- Lv, L., Yang, S., Zhu, Y., Zhai, X., Li, S., Tao, X., & Dong, D. (2022). Relationship between metabolic reprogramming and drug resistance in breast cancer. In *Frontiers in Oncology* (Vol. 12). <https://doi.org/10.3389/fonc.2022.942064>
- Ma, K., Vitek, O., & Nesvizhskii, A. I. (2012). A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. *BMC Bioinformatics*, 13 Suppl 16. <https://doi.org/10.1186/1471-2105-13-S16-S1>
- Maadi, H., Soheilifar, M. H., Choi, W.-S., Moshtaghian, A., & Wang, Z. (2021). Trastuzumab mechanism of action; 20 years of research to unravel a dilemma. *Cancers*, 13(14), 3540.
- Macias, H., & Hinck, L. (2012). Mammary gland development. *Wiley Interdisciplinary Reviews: Developmental Biology*, 1(4), 533–557.
- Magnani, L., Ballantyne, E. B., Zhang, X., & Lupien, M. (2011). PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLoS Genetics*, 7(11), e1002368.
- Magnifico, A., Albano, L., Campaner, S., Delia, D., Castiglioni, F., Gasparini, P., Sozzi, G., Fontanella, E., Menard, S., & Tagliabue, E. (2009). Tumor-initiating cells of HER2-positive carcinoma cell lines express the highest oncoprotein levels and are sensitive to trastuzumab. *Clinical Cancer Research*. <https://doi.org/10.1158/1078-0432.CCR-08-1327>
- Maldonado, H., Dreger, M., Bedgood, L. D., Kyriakou, T., Wolanska, K. I., Rigby, M. E., Marotta, V. E., Webster, J. M., Wang, J., Rusilowicz-Jones, E. V., Murphy, D. J.,

- Townsend, P. A., Luzio, J. P., Parker, P. J., Metcalf, D. J., Gilmore, A. G., Paluch, E. K., Frame, M. C., & Caswell, P. T. (2024). A trafficking regulatory subnetwork governs $\alpha V\beta 6$ integrin-HER2 cross-talk to control breast cancer invasion and drug resistance. *Science Advances*, *10*(49), eadk9944.
<https://doi.org/10.1126/sciadv.adk9944>
- Mamidi, S., Cinci, M., Hasmann, M., Fehring, V., & Kirschfink, M. (2013). Lipoplex mediated silencing of membrane regulators (CD46, CD55 and CD59) enhances complement-dependent anti-tumor activity of trastuzumab and pertuzumab. *Molecular Oncology*, *7*(3), 580–594. <https://doi.org/10.1016/j.molonc.2013.02.011>
- Mandric, I., Temate-Tiagueu, Y., Shcheglova, T., Al Seesi, S., Zelikovsky, A., & Măndoiu, I. I. (2017). Fast bootstrapping-based estimation of confidence intervals of expression levels and differential expression from RNA-Seq data. *Bioinformatics*.
<https://doi.org/10.1093/bioinformatics/btx365>
- Marks, C., & Deane, C. M. (2024). Humanization of Antibodies using a Machine Learning Approach on Large-Scale Repertoire Data. *Bioinformatics*, *37*(22), 4041–4048.
- Marteijn, J. A., Lans, H., Vermeulen, W., & Hoeijmakers, J. H. J. (2014). Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews Molecular Cell Biology*, *15*(7), 465–481.
- Mascarenhas-Melo, F., Diaz, M., Gonçalves, M. B. S., Vieira, P., Bell, V., Viana, S., Nunes, S., Paiva-Santos, A. C., & Veiga, F. (2024). An overview of biosimilars—development, quality, regulatory issues, and management in healthcare. *Pharmaceuticals*, *17*(2), 235.
- Massagué, J., Blain, S. W., & Lo, R. S. (2000). TGF- β signaling in growth control, cancer, and heritable disorders. *Cell*, *103*(2), 295–309.
- Mayer, I. A. (2015). Clinical Implications of Mutations in the PI3K Pathway in HER2+ Breast Cancer: Prognostic or Predictive? In *Current Breast Cancer Reports* (Vol. 7, Issue 4). <https://doi.org/10.1007/s12609-015-0197-9>
- Mayfield, S., Vaughn, J. P., & Kute, T. E. (2001). DNA strand breaks and cell cycle perturbation in herceptin treated breast cancer cell lines. *Breast Cancer Research and Treatment*, *70*(2), 123–129. <https://doi.org/10.1023/A:1012999012192>

- McClure, R. S., Wendler, J. P., Adkins, J. N., Swanstrom, J., Baric, R., Kaiser, B. L. D., Oxford, K. L., Waters, K. M., & McDermott, J. E. (2019). Unified feature association networks through integration of transcriptomic and proteomic data. *PLoS Computational Biology*, 15(9). <https://doi.org/10.1371/journal.pcbi.1007241>
- McIntyre, A., & Harris, A. L. (2016). Overcoming Hypoxia-Mediated Tumor Progression: Combinatorial Approaches Targeting pH Regulation, Angiogenesis and Immune Dysfunction. *Frontiers in Cell and Developmental Biology*, 4, 27.
- Meek, D. W. (2009). Tumour suppression by p53: a role for the DNA damage response? *Nature Reviews Cancer*, 9(10), 714–723.
- Menyhárt, O., & Györffy, B. (2021). Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Computational and Structural Biotechnology Journal*, 19, 949–960.
- Mercogliano, M. F., Bruni, S., Mauro, F. L., & Schillaci, R. (2023). Emerging Targeted Therapies for HER2-Positive Breast Cancer. *Cancers*, 15(7), 1987. <https://doi.org/10.3390/cancers15071987>
- Merico, D., Isserlin, R., Stueker, O., Emili, A., & Bader, G. D. (2010). Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE*, 5(11). <https://doi.org/10.1371/journal.pone.0013984>
- Merry, C. R., McMahon, S., Forrest, M. E., Bartels, C. F., Saiakhova, A., Bartel, C. A., Scacheri, P. C., Thompson, C. L., Jackson, M. W., Harris, L. N., & Khalil, A. M. (2016). Transcriptome-wide identification of mRNAs and lincRNAs associated with trastuzumab-resistance in HER2-positive breast cancer. *Oncotarget*, 7(33), 53230–53244. <https://doi.org/10.18632/oncotarget.10637>
- Metzger-Filho, O., Viale, G., Trippa, L., Li, T., Yardley, D. A., Mayer, I. A., Abramson, V. G., Arteaga, C. L., Springer, L. M., Waks, A. G., Krop, I. E., Winer, E. P., Tolaney, S. M., Burstein, H. J., Barry, W. T., Berry, D. A., Piccart-Gebhart, M., Perou, C. M., & Carey, L. A. (2021). Impact of HER2 Heterogeneity on Treatment Response of Early-Stage HER2-Positive Breast Cancer: Phase II Neoadjuvant Clinical Trial of T-DM1

- Combined with Pertuzumab. *Cancer Discovery*, 11(10), 2474–2487.
<https://doi.org/10.1158/2159-8290.CD-21-0187>
- Meugnier, E., Coury, F., Tebib, J., Ferraro-Peyret, C., Rome, S., Bienvenu, J., Vidal, H., Sibia, J., & Fabien, N. (2011). Gene expression profiling in peripheral blood cells of patients with rheumatoid arthritis in response to anti-TNF- α treatments. *Physiological Genomics*, 43(7), 365–371.
- Mimura, Y., Church, S., Ghirlando, R., Ashton, P. R., Dong, S., Goodall, M., Lund, J., & Jefferis, R. (2001). The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: Properties of a series of truncated glycoforms. *Molecular Immunology*, 37(12–13), 697–706. [https://doi.org/10.1016/S0161-5890\(00\)00105-X](https://doi.org/10.1016/S0161-5890(00)00105-X)
- Miranda-Hernández, M. P., López-Morales, C. A., Piña-Lara, N., Perdomo-Abúndez, F. C., Pérez, N. O., Revilla-Beltri, J., Molina-Pérez, A., Estrada-Marín, L., Flores-Ortiz, L. F., Ruiz-Argüelles, A., & Medina-Rivero, E. (2015). Pharmacokinetic Comparability of a Biosimilar Trastuzumab Anticipated from Its Physicochemical and Biological Characterization. *BioMed Research International*, 2015.
<https://doi.org/10.1155/2015/874916>
- Moasser, M. M. (2007). The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene*, 26(45), 6469–6487.
- Morris, L. G. T., Kaufman, A. M., Gong, Y., Ramaswami, D., Walsh, L. A., Turcan, Ş., Eng, S., Kannan, K., Zou, Y., Peng, L., Banuchi, V. E., Paty, P., Zeng, Z., Vakiani, E., Solit, D., Singh, B., Ganly, I., Liao, L., Cloughesy, T. C., ... Chan, T. A. (2013). Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. *Nature Genetics*, 45(3), 253–261. <https://doi.org/10.1038/ng.2538>
- Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L. H. F., & Foulster, M. I. (2007). Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III α in a cell-cycle-specific manner. *Molecular Cell*, 27(2), 311–323.

- Mukhopadhyay, P., Chakraborty, S., Ponnusamy, M. P., Lakshmanan, I., Jain, M., & Batra, S. K. (2011). Mucins in the pathogenesis of breast cancer: implications in diagnosis, prognosis and therapy. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1815(2), 224–240.
- Mukohara, T. (2011). Mechanisms of resistance to anti-human epidermal growth factor receptor 2 agents in breast cancer. *Cancer Science*, 102(1), 1–8.
- Müller-Dott, S., Tsirvouli, E., Vazquez, M., Ramirez Flores, R. O., Badia-I-Mompel, P., Fallegger, R., Türei, D., Laegreid, A., & Saez-Rodriguez, J. (2023). Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic Acids Research*, 51(20), 10934–10949.
<https://doi.org/10.1093/NAR/GKAD841>
- National Breast Cancer Foundation. (2024). *HER2 positive breast cancer*.
- Nazki, F. H., Sameer, A. S., & Ganaie, B. A. (2014). Folate: metabolism, genes, polymorphisms and the associated diseases. *Gene*, 533(1), 11–20.
- Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability—an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology*, 11(3), 220–228.
- Negro, A., Brar, B. K., & Lee, K.-F. (2004). Essential Roles of Her2/erbB2 in Cardiac Development and Function. *Recent Progress in Hormone Research*, 59, 1–12.
- Neve, R. M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F. L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F., Speed, E., Spellman, T., Gray, J., Mills, D. A., Spellman, P., & Gray, J. W. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*, 10(6), 515–527.
<https://doi.org/10.1016/j.ccr.2006.10.008>
- Niazi, S. K. (2023). A Critical Analysis of the FDA’s Omics-Driven Pharmacodynamic Biomarkers to Establish Biosimilarity. In *Pharmaceuticals* (Vol. 16, Issue 11).
<https://doi.org/10.3390/ph16111556>
- Nie, L., Wu, G., Culley, D. E., Scholten, J. C. M., & Zhang, W. (2007). Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. *Critical Reviews in Biotechnology*, 27(2), 63–75.

- Nielsen, T. O., Parker, J. S., Leung, S., Voduc, D., Ebbert, M., Vickery, T., Davies, S. R., Snider, J., Stijleman, I. J., & Reed, J. (2010). A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clinical Cancer Research*, 16(21), 5222–5232.
- Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L. B., Martin, S., Wedge, D. C., Shlien, D. J. H., Morganella, D. J., Raine, K., Butler, A., Jones, D., Campbell, P. J., & Stratton, M. R. (2016). A whole-genome sequence and transcriptome perspective on HER2-positive breast cancers. *Nature Communications*, 7, 12276. <https://doi.org/10.1038/ncomms12276>
- Nupur, N., Joshi, S., Gulliarne, D., & Rathore, A. S. (2022). Analytical similarity assessment of biosimilars: global regulatory landscape, recent studies and major advancements in orthogonal platforms. *Frontiers in Bioengineering and Biotechnology*, 10, 832059.
- Nuti, M., Bellati, F., Visconti, V., Napoletano, C., Domenici, L., Caccetta, J., Zizzari, I. G., Ruscito, I., Rahimi, H., Benedetti-Panici, P., & Rughetti, A. (2011). Immune Effects of Trastuzumab. *Journal of Cancer*, 2, 317. <https://doi.org/10.7150/jca.2.317>
- Ogi, T., Limsirichaikul, S., Overmeer, R. M., Volker, M., Takenaka, K., Cloney, R., Nakazawa, Y., Niimi, A., Miki, Y., Jaspers, N. G., Mullenders, L. H. F., Tanaka, K., & Vermeulen, W. (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Molecular Cell*, 37(5), 714–727. <https://doi.org/10.1016/j.molcel.2010.02.021>
- Ogretmen, B. (2018). Sphingolipid metabolism in cancer signalling and therapy. *Nature Reviews Cancer*, 18(1), 33–50.
- Olayioye, M. A., Neve, R. M., Lane, H. A., & Hynes, N. E. (2000). The ErbB signalling network: receptor heterodimerization in development and cancer. *The EMBO Journal*, 19(13), 3159–3167. <https://doi.org/10.1093/emboj/19.13.3159>

- Olson, E. M., Abdel-Rasoul, M., Maly, J., Wu, C. S., Lin, N. U., & Shapiro, C. L. (2013). Incidence and risk factors for brain metastases in patients with HER2-positive breast cancer. *Annals of Oncology*, 24(12), 2999–3005.
- O'Neill, K. I., Kuo, L. W., Williams, M. M., Lind, H., Crump, L. S., Hammond, N. G., Spoelstra, N. S., Caino, M. C., & Richer, J. K. (2022). NPC1 Confers Metabolic Flexibility in Triple Negative Breast Cancer. *Cancers*, 14(14).
<https://doi.org/10.3390/cancers14143543>
- Orsetti, B., Nugoli, M., Cervera, N., Lasorsa, L., Chuchana, P., Ursule, L., Nguyen, C., Redon, R., du Manoir, S., Rodriguez, C., Bibeau, F., Bonnet, F., Rochaix, P., Sastre-Garau, X., Jacquemier, C., MacGrogan, G., Gestraud, P., de Reyniès, A., Bonnet, F., ... Finetti, P. (2017). Genome-wide copy number aberrations and HER2 and FGFR1 alterations in primary breast cancer by molecular inversion probe microarray. *Molecular Cancer Research*, 15(3), 305–316. <https://doi.org/10.1158/1541-7786.MCR-16-0302>
- Ott, M., Gogvadze, V., Orrenius, S., & Zhivotovsky, B. (2016). Mitochondrial proteins and complexes in breast and mammary cancer. *Seminars in Cancer Biology*, 47, 110–118.
- Ozaki, T., & Nakagawara, A. (2005). p73, a sophisticated p53 family member in the cancer world. *Cancer Science*, 96(11), 729–737.
- Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: Advances, challenges and opportunities. *Nature Reviews Genetics*, 12(2), 87–98. <https://doi.org/10.1038/nrg2934>
- Palomer, X., Barroso, E., Pizarro-Delgado, J., Peña, L., Botteri, G., Zarei, M., Aguilar, D., Montori-Grau, M., & Vázquez-Carrera, M. (2018). PPAR β/δ : a key therapeutic target in metabolic disorders. *International Journal of Molecular Sciences*, 19(3), 913.
- Pan, L., Li, J., Xu, Q., Gao, Z., Yang, M., Wu, X., & Li, X. (2024). HER2/PI3K/AKT pathway in HER2-positive breast cancer: A review. *Medicine*, 103(24), e38508.
- Pan, Y., Wang, X., Dai, S., Yang, Q., Li, M., & Ling, L. (2017). CD147 and HER2 expression in different HER2-positive cancer cell lines. *Oncology Letters*, 13(6), 4099–4106.

- Pandolfi, P. P. (2004). Breast cancer: loss of PTEN predicts resistance to treatment. *The New England Journal of Medicine*, 351(22), 2337–2338.
- Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J. F., Haibe-Kains, A., Bernard, D. S., Ellis, J. M., & Perou, C. M. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, 27(8), 1160–1167.
<https://doi.org/10.1200/JCO.2008.18.1370>
- Pastena, P., Perera, H., Martinino, A., Kartsonis, W., & Giovino, F. (2024). Unraveling biomarker signatures in triple-negative breast cancer: a systematic review for targeted approaches. *International Journal of Molecular Sciences*, 25(5), 2559.
- Pattanayak, B., Lameirinhas, A., Torres-Ruiz, S., Burgués, O., Rovira, A., Martínez, M. T., Tapia, M., Zazo, S., Albanell, J., Rojo, F., & Prat, A. (2022). Role of SALL4 in HER2+ breast cancer progression: regulating PI3K/AKT pathway. *International Journal of Molecular Sciences*, 23(21), 13292. <https://doi.org/10.3390/ijms232113292>
- Pavin, N., & Tolić, I. M. (2021). Mechanobiology of the mitotic spindle. *Developmental Cell*, 56(2), 192–201.
- Peles, E., Levy, R. B., Or, E., Ullrich, A., & Yarden, Y. (1991). Oncogenic forms of the neu/HER2 tyrosine kinase are permanently coupled to phospholipase C gamma. *The EMBO Journal*, 10(8), 2077–2086. <https://doi.org/10.1002/j.1460-2075.1991.tb07739.x>
- Perez, E. A., Romond, E. H., Suman, V. J., Jeong, J.-H., Sledge, G., Geyer, C. E. J., Martino, S., Rastogi, P., Gralow, J., Swain, S. M., Winer, E. P., Colon-Otero, G., Davidson, N. E., Mamounas, E., Zujewski, J. A., & Wolmark, N. (2014). Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2–positive breast cancer: planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. *Journal of Clinical Oncology*, 32(33), 3744–3752.
<https://doi.org/10.1200/JCO.2014.55.5730>
- Perna, F., Berman, S. H., Soni, R. K., Mansilla-Soto, J., Eyquem, J., Hamieh, M., Hendrickson, R. C., Brennan, C. W., & Sadelain, M. (2017). Integrating Proteomics

- and Transcriptomics for Systematic Combinatorial Chimeric Antigen Receptor Therapy of AML. *Cancer Cell*, 32(4). <https://doi.org/10.1016/j.ccell.2017.09.004>
- Pernas, S., Barroso-Sousa, R., & Tolaney, S. M. (2019). Overall Survival for HER2-Positive Breast Cancer Patients Treated With Pertuzumab and Trastuzumab (P+T) in Clinical Practice: A Real-World Data Analysis. *Journal of Clinical Oncology*, 37(15_suppl), 1030.
- Perou, C. M., Sørlie, T., Eisen, M. B., Van De Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., & Akslen, L. A. (2000). Molecular portraits of human breast tumours. *Nature*, 406(6797), 747–752.
- Petricevic, B., Laengle, J., Singer, J., Sachet, M., Fazekas, J., Steger, G., Bartsch, R., Jensen-Jarolim, E., & Bergmann, M. (2013). Trastuzumab mediates antibody-dependent cell-mediated cytotoxicity and phagocytosis to the same extent in both adjuvant and metastatic HER2/neu breast cancer patients. *Journal of Translational Medicine*, 11(1), 307. <https://doi.org/10.1186/1479-5876-11-307>
- Pinto, A. C., Ades, F., de Azambuja, E., & Piccart-Gebhart, M. (2013). Trastuzumab for patients with HER2 positive breast cancer: delivery, duration and combination therapies. *The Breast*, 22, S152–S155.
- Pisupati, K., Benet, A., Tian, Y., Okbazghi, S., Kang, J., Rand, M., Schrimsher, J., Carlson, M., Abadie, J., Cummings, T., Nupur, N., Chhabra, N., Dash, R., & Rathore, A. S. (2017). Assessment of structural and functional similarity of biosimilar products: Rituximab as a case study. *MAbs*, 9(1), 143–158. <https://doi.org/10.1080/19420862.2017.1402996>
- Pohlmann, P. R., Mayer, I. A., & Mernaugh, R. (2009). Resistance to trastuzumab in breast cancer. *Clinical Cancer Research*, 15(24), 7479–7491. <https://doi.org/10.1158/1078-0432.CCR-09-0636>
- Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2). <https://doi.org/10.7155/jgaa.00124>

- Prasad, M. (2019). Top 5 Challenges Faced By Biosimilars: Navigating the Complex Landscape. *DrugPatentWatch*. <https://www.drugpatentwatch.com/blog/top-5-challenges-faced-biosimilars/>
- Prat, A., Carey, L. A., Adamo, B., Vidal, M., Tabernero, J., Cort'es, J., Parker, J. S., Perou, C. M., & Baselga, J. (2014). Molecular Features and Survival Outcomes of the Intrinsic Subtypes Within HER2-Positive Breast Cancer. *Journal of the National Cancer Institute*, 106(8), dju152.
- Pupa, S. M., Ligorio, F., Cancila, V., Franceschini, A., Tripodo, C., Vernieri, C., & Castagnoli, L. (2021). HER2 signaling and breast cancer stem cells: the bridge behind HER2-positive breast cancer aggressiveness and therapy refractoriness. *Cancers*, 13(19), 4778.
- Raju, T. S. (2010). Impact of Fc Glycosylation on Monoclonal Antibody Effector Functions and Degradation by Proteases. In S. J. Shire, W. Gombotz, K. Bechtold-Peters, & J. Andya (Eds.), *Current Trends in Monoclonal Antibody Development and Manufacturing: Vol. XI* (pp. 249–269). Springer. <https://doi.org/10.1002/9781119116493.ch16>
- RapidNovor. (2024). Antibody Humanization Strategies, Challenges, and Innovations. *RapidNovor Blog*. <https://www.rapidnovor.com/antibody-humanization-strategies-challenges-innovations/>
- Rashid, M. M., & Selvarajoo, K. (2024). Advancing drug-response prediction using multi-modal and-omics machine learning integration (MOMLIN): a case study on breast cancer clinical data. *Briefings in Bioinformatics*, 25(4).
- Rathore, A. S., Chhabra, H., & Bhargava, A. (2020). Approval of Biosimilars: A Review of Unsuccessful Regulatory Filings. In *Expert Opinion on Biological Therapy*. <https://doi.org/10.1080/14712598.2020.1793954>
- Raufaste-Cazavieille, V., Santiago, R., & Droit, A. (2022). Multi-omics analysis: Paving the path toward achieving precision medicine in cancer treatment and immuno-oncology. In *Frontiers in Molecular Biosciences* (Vol. 9). <https://doi.org/10.3389/fmolb.2022.962743>

- Rayet, B., & G  linas, C. (1999). Aberrant rel/nfkb genes and activity in human cancer. *Oncogene*, 18(49), 6938–6947. <https://doi.org/10.1038/sj.onc.1203221>
- Rediti, M., Venet, D., Joaquin Garcia, A., Maetens, M., Vincent, D., Majjaj, S., El-Abed, S., Di Cosimo, S., Ueno, T., Izquierdo, M., Piccart, M., & Sotiriou, C. (2024). Identification of HER2-positive breast cancer molecular subtypes with potential clinical implications in the ALTTO clinical trial. *Nature Communications*, 15(1), 10402. <https://doi.org/10.1038/s41467-024-54621-3>
- Reed, S. H., & Gillette, T. G. (2007). Nucleotide excision repair and the ubiquitin proteasome pathway—do all roads lead to Rome? *DNA Repair*, 6(2), 149–156.
- Reusch, D., Habegger, M., Falck, D., Peter, B., Maier, B., Gassner, J., Hook, M., Wagner, K., Bonnington, L., Bulau, P., Wuhler, M., Deelder, A. M., & Spengler, M. (2015). High-throughput glycosylation analysis of therapeutic immunoglobulin G by capillary gel electrophoresis using a DNA analyzer. *MAbs*, 7(1), 180–191. <https://doi.org/10.4161/19420862.2015.989785>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47. <https://doi.org/10.1093/nar/gkv007>
- Rivas, S., Mar  n, A., Samtani, S., Gonz  lez-Feli  , E., & Armis  n, R. (2022). MET signaling pathways, resistance mechanisms, and opportunities for target therapies. *International Journal of Molecular Sciences*, 23(22), 13898.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rocca, J. D., Hall, E. K., Lennon, J. T., Evans, S. E., Waldrop, M. P., Cotner, J. B., Nemergut, D. R., Graham, E. B., & Wallenstein, M. D. (2015). Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *The ISME Journal*, 9(8), 1693–1699.

- Romano, R., Palamaro, L., Fusco, A., Giardino, G., Gallo, V., Del Vecchio, L., & Pignata, C. (2013). FOXP1: A Master Regulator Gene of Thymic Epithelial Development. *Frontiers in Immunology*, 4, 187.
- Roque, A. C. A., Lowe, C. R., & Taipa, M. Â. (2004). Antibodies and genetically engineered related molecules: Production and purification. *Biotechnology Progress*, 20(3), 639–654. <https://doi.org/10.1021/bp030070k>
- Roskoski, R. (2014). The ErbB/HER family of protein-tyrosine kinases and cancer. In *Pharmacological Research* (Vol. 79). <https://doi.org/10.1016/j.phrs.2013.11.002>
- Rossi, C., Cicalini, I., Cufaro, M. C., Consalvo, A., Upadhyaya, P., Sala, G., Antonucci, I., Del Boccio, P., Stuppia, L., & De Laurenzi, V. (2022). Breast cancer in the era of integrating “Omics” approaches. In *Oncogenesis* (Vol. 11, Issue 1). <https://doi.org/10.1038/s41389-022-00393-8>
- RStudio Team. (2016). *RStudio: Integrated development environment for R*. Rstudio, Inc. <http://www.rstudio.com/>
- Ruiz-Pérez, M. V., Henley, A. B., & Arsenian-Henriksson, M. (2017). The MYCN protein in health and disease. *Genes*, 8(4), 113.
- Russo, J., & Russo, I. H. (2004). Development of the human breast. *Maturitas*, 49(1), 2–15.
- Saal, L. H., Holm, K., Maurer, M., Memeo, L., Su, T., Wang, X., Yu, J. S., Malmström, P.-O., Mansukhani, M., Enoksson, J., Hibshoosh, H., Borg, Å., & Parsons, R. (2005). PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. *Cancer Research*, 65(7), 2554–2559. <https://doi.org/10.1158/0008-5472.CAN-04-3913>
- Sabbaghi, M. A., Gil-Gomez, G., Guardia, C., Servitja, S., Arpi, O., García-Alonso, S., Menendez, S., Arumi-Uria, M., Serrano, L., Salido, M., Muntasell, A., Martínez-García, M., Zazo, S., Chamizo, C., Gonzalez-Alonso, P., Madoz-Gurpide, J., Eroles, P., Arribas, J., Tusquets, I., ... Albanell, J. (2017). Defective cyclin B1 induction in trastuzumab-emtansine (T-DM1) acquired resistance in HER2-positive breast cancer. *Clinical Cancer Research*, 23(22). <https://doi.org/10.1158/1078-0432.CCR-17-0696>

- Sahlberg, K. K., Hongisto, V., Edgren, H., Mäkelä, R., Hellström, K., Due, E. U., Moen Vollan, H. K., Sahlberg, N., Wolf, M., Børresen-Dale, A. L., Perälä, M., & Kallioniemi, O. (2013). The HER2 amplicon includes several genes required for the growth and survival of HER2 positive breast cancer cells. *Molecular Oncology*, 7(3). <https://doi.org/10.1016/j.molonc.2012.10.012>
- Salinas-Vera, Y. M., Valdés, J., Pérez-Navarro, Y., Mandujano-Lazaro, G., Marchat, L. A., Ramos-Payán, R., Nuñez-Olvera, S. I., Pérez-Plascencia, C., & López-Camarillo, C. (2022). Three-Dimensional 3D Culture Models in Gynecological and Breast Cancer Research. In *Frontiers in Oncology* (Vol. 12). <https://doi.org/10.3389/fonc.2022.826113>
- Sammut, S. J., Crispin-Ortuzar, M., Chin, S. F., Provenzano, E., Bardwell, H. A., Ma, W., Cope, W., Dariush, A., Dawson, S. J., Abraham, J. E., Dunn, J., Hiller, L., Thomas, J., Cameron, D. A., Bartlett, J. M. S., Hayward, L., Pharoah, P. D., Markowitz, F., Rueda, O. M., ... Caldas, C. (2022). Multi-omic machine learning predictor of breast cancer therapy response. *Nature*, 601(7894). <https://doi.org/10.1038/s41586-021-04278-5>
- Santos, A., Wernersson, R., & Jensen, L. J. (2014). Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Research*, 43(D1), D1140–D1144. <https://doi.org/10.1093/nar/gku1092>
- Sawyers, C. L. (2019). Herceptin: a first assault on oncogenes that launched a revolution. *Cell*, 179(1), 8–12.
- Scarpulla, R. C., Vega, R. B., & Kelly, D. P. (2011). Metabolic control of mitochondrial biogenesis through the PGC-1 family regulatory network. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1813(7), 1269–1278.
- Schellekens, H. (2009). Biosimilars: the long and winding road to clinical equivalence. *Drugs*, 69(8), 967–982.
- Schnoes, A. M., Ream, D. C., Thorman, A. W., Babbitt, P. C., & Friedberg, I. (2013). Biases in the experimental annotations of protein function and their effect on our understanding of protein function space. *PLoS Computational Biology*, 9(5), e1003063.
- Schofield, D. (2023). *Monoclonal antibody production: Process, Technologies & Steps*.

- Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., Garnett, M. J., Blüthgen, N., & Saez-Rodriguez, J. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-017-02391-6>
- Schulze, M., Kumar, Y., Rattay, M., Niemann, J., Wijffels, R. H., & Martens, D. E. (2022). Transcriptomic analysis reveals mode of action of butyric acid supplementation in an intensified CHO cell fed-batch process. *Biotechnology and Bioengineering*, 119(9), 2359–2373.
- Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M., & Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, 22(6), 839–851. <https://doi.org/10.1261/rna.053959.115>
- Schwab, M. (2004). MYCN in neuronal tumours. *Cancer Letters*, 204(2), 179–187.
- Scott, A. M., Wolchok, J. D., & Old, L. J. (2012). Antibody therapy of cancer. *Nature Reviews Cancer*, 12(4), 278–287. <https://doi.org/10.1038/nrc3236>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Shattuck, D. L., Miller, J. K., Carraway Kermit L., I. I. I., & Sweeney, C. (2008). Met Receptor Contributes to Trastuzumab Resistance of Her2-Overexpressing Breast Cancer Cells. *Cancer Research*, 68(5), 1471–1477. <https://doi.org/10.1158/0008-5472.CAN-07-5962>
- Shaulian, E., & Karin, M. (2001). AP-1 in cell proliferation and survival. *Oncogene*, 20(19), 2390–2400.
- Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, I. A., Avril, S., Prater, M., Eirew, P., Caldas, C., & Watson, C. J. (2012). Phenotypic and functional

- characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Research*, 14, 1–19.
- Shi, Y., Fan, X., Deng, H., Brezski, R. J., Ryczyn, M., Jordan, R. E., Strohl, W. R., Zou, Q., Zhang, N., & An, Z. (2015). Trastuzumab Triggers Phagocytic Killing of High HER2 Cancer Cells In Vitro and In Vivo by Interaction with Fc γ Receptors on Macrophages. *The Journal of Immunology*, 194(9), 4379–4386.
<https://doi.org/10.4049/jimmunol.1402891>
- Shi, Y., & Massagu'e, J. (2003). Mechanisms of TGF- β signaling from cell membrane to the nucleus. *Cell*, 113(6), 685–700.
- Shteynberg, D., Deutsch, E. W., Lam, H., Eng, J. K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R. L., Aebersold, R., & Nesvizhskii, A. I. (2011). iProphet: Multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Molecular and Cellular Proteomics*, 10(12).
<https://doi.org/10.1074/mcp.M111.007690>
- Sigma-Aldrich. (2024). *Authenticated Breast Cancer Cell Lines for Cancer Research*.
- Simpson, G. G. (1943). Mammals and the nature of continents. *American Journal of Science*. <https://doi.org/10.2475/ajs.241.1.1>
- Sino Biological. (2024). *Chimeric Monoclonal Antibody Overview*.
- Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., & Norton, L. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *New England Journal of Medicine*, 344(11), 783–792.
<https://doi.org/10.1056/NEJM200103153441101>
- Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., & Ideker, T. (2011). Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics*, 27(3), 431–432. <https://doi.org/10.1093/bioinformatics/btq675>
- Smyth, G. K. (2005). limma: Linear Models for Microarray Data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (pp. 397–420).
https://doi.org/10.1007/0-387-29362-0_23

- Sokal, R. R., & Sneath, P. H. A. (1963). Principles of numerical taxonomy. In *W. H. Freeman and Company*. W.H.Freeman. <https://doi.org/10.2307/1377831>
- Spivak, G. (2015). Nucleotide excision repair in humans. *DNA Repair*, 36, 13–18.
- Stanley, A., Ashrafi, G. H., Seddon, A. M., & Modjtahedi, H. (2017). Synergistic effects of various Her inhibitors in combination with IGF-1R, C-MET and Src targeting agents in breast cancer cell lines. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-04301-8>
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545–15550. <https://doi.org/10.1073/PNAS.0506580102>
- Sullivan, P. M., & Digrazia, L. M. (2017). Analytic characterization of biosimilars. *American Journal of Health-System Pharmacy*, 74(8). <https://doi.org/10.2146/ajhp150971>
- Sun, Z., Shi, Y., Shen, Y., Cao, L., Zhang, W., & Guan, X. (2015). Analysis of different HER-2 mutations in breast cancer progression and drug resistance. *Journal of Cellular and Molecular Medicine*, 19(12), 2691–2701.
- Suzuki, M., Kato, C., & Kato, A. (2015). Therapeutic antibodies: their mechanisms of action and the pathological findings they induce in toxicity studies. *Journal of Toxicologic Pathology*, 28(3), 133–139.
- Swain, S. M., Baselga, J., Kim, S.-B., Ro, J., Semiglazov, V., Campone, M., Ciruelos, E., Ferrero, J.-M., Schneeweiss, A., & Heeson, S. (2015). Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *New England Journal of Medicine*, 372(8), 724–734.
- Swain, S. M., Shastry, M., & Hamilton, E. (2023). Targeting HER2-positive breast cancer: advances and future directions. In *Nature Reviews Drug Discovery* (Vol. 22, Issue 2). <https://doi.org/10.1038/s41573-022-00579-0>

- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & Von Mering, C. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(1 D). <https://doi.org/10.1093/nar/gkac1000>
- Tan, S., Sun, X., Dong, H., Wang, M., Yao, L., Wang, M., Xu, L., & Xu, Y. (2024). ACSL3 regulates breast cancer progression via lipid metabolism reprogramming and the YES1/YAP axis. *Cancer Biology & Medicine*, 21(7), 606.
- Tang, S., Ning, Q., Yang, L., Mo, Z., & Tang, S. (2020). Mechanisms of immune escape in the cancer immune cycle. *International Immunopharmacology*, 86, 106700.
- Tarantino, P., Hamilton, E., Tolaney, S. M., Cortes, J., Morganti, S., Lin, N. U., & Curigliano, G. (2022). Management of Early-Stage Human Epidermal Growth Factor Receptor 2–Positive Breast Cancer. *Journal of Clinical Oncology*, 40(16), 1832–1845.
- Teng, S., Wang, D., Qian, Y., Bahitwa, R., Shao, J., Suo, M., Xu, M., Yang, L., Li, T., & Yu, Q. (2024). QUIC-seq: Quick ultra-affordable high-throughput convenient RNA sequencing. *Plant Biotechnology Journal*, 23(1), 281. <https://doi.org/10.1111/pbi.14122>
- The Inkscape Team. (2024). *Inkscape*. <https://inkscape.org>
- Thennavan, A., Beca, F., Xia, Y., Recio, A. V., Allott, E. H., Geradts, J., Hoadley, K. A., & Perou, C. M. (2021). Molecular analysis of TCGA breast cancer histologic types. *Cell Reports*, 36(13), 109689.
- Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., ... Lundberg, E. (2017). A subcellular map of the human proteome. *Science*, 356(6340). <https://doi.org/10.1126/science.aal3321>
- Tkaczuk, K. H. R., & Jacobs, I. A. (2014). Biosimilars in oncology: from development to clinical practice. *Seminars in Oncology*, 41, S3–S12.

- Tolaney, S. M., & Krop, I. E. (2009). Mechanisms of trastuzumab resistance in breast cancer. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)*, 9(3), 348–355.
- Tortora, G. (2011). Mechanisms of resistance to HER2 target therapy. *Journal of the National Cancer Institute Monographs*, 2011(43), 95–98.
- Towfic, F., Funt, J. M., Fowler, K. D., Bakshi, S., Blaugrund, E., Artyomov, M. N., Hayden, M. R., Ladkani, D., Schwartz, R., & Zeskind, B. (2014). Comparing the biological impact of glatiramer acetate with the biological impact of a generic. *PloS One*, 9(1), e83757.
- Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., Gill, S., Harrington, W. F., Pantel, S., Krill-Burger, J. M., Meyers, G. J., Ali, J. M., Goodale, B., Lee, Y., Wang, L., Ghandi, M., Garraway, J. E., Sellers, W. R., Boehm, J. S., ... Boehm, J. D. (2017). Defining a cancer dependency map. *Cell*, 170(3), 564–576. <https://doi.org/10.1016/j.cell.2017.06.010>
- Uifălean, A., Ilieș, M., Nicoară, R., Rus, L. M., Hegheș, S. C., & Iuga, C.-A. (2018). Concepts and challenges of biosimilars in breast cancer: the emergence of trastuzumab biosimilars. *Pharmaceutics*, 10(4), 168.
- Upton, R., Banuelos, A., Feng, D., Biswas, T., Kao, K., McKenna, K., Willingham, S., Ho, P. Y., Rosental, B., Tal, M. C., Raveh, T., Volkmer, J. P., Pegram, M. D., & Weissman, I. L. (2021). Combining CD47 blockade with trastuzumab eliminates HER2-positive breast cancer cells and overcomes trastuzumab tolerance. *Proceedings of the National Academy of Sciences of the United States of America*, 118(29). <https://doi.org/10.1073/pnas.2026849118>
- U.S. Food, & Administration, D. (2019). *Biosimilars Action Plan: Balancing Innovation and Competition*.
- U.S. Food and Drug Administration. (n.d.). *Drugs@FDA: FDA Approved Drug Products*. Retrieved January 7, 2019, from <https://www.accessdata.fda.gov/scripts/cder/daf/>
- U.S. Food and Drug Administration. (2017). *Ogivri (trastuzumab-dkst) Approval*. <https://www.fda.gov>

- Valabrega, G., Montemurro, F., & Aglietta, M. (2007). Trastuzumab: Mechanism of action, resistance and future perspectives in HER2-overexpressing breast cancer. *Annals of Oncology*, 18(6), 977–984. <https://doi.org/10.1093/annonc/mdl475>
- Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*, 324(5930), 1029–1033.
- Vasudevan, K. M., Barbie, D. A., Davies, M. A., Rabinovsky, R., McNear, C. J., Kim, J. J., Hennessy, B. T., Tseng, H.-Y., Pochanard, P., Kim, S. Y., Mills, G. B., Settleman, J., & Cantley, L. C. (2009). AKT-independent signaling downstream of oncogenic PIK3CA mutations in human cancer. *Cancer Cell*, 16(1), 21–32. <https://doi.org/10.1016/j.ccr.2009.05.006>
- Vazquez-Martin, A., Oliveras-Ferraros, C., & Menendez, J. A. (2009). Autophagy facilitates the development of breast cancer resistance to the anti-HER2 monoclonal antibody trastuzumab. *PLoS ONE*, 4(7). <https://doi.org/10.1371/journal.pone.0006251>
- Veeranagouda, Y., Remaury, A., Guillemot, J. C., & Didier, M. (2019). RNA Fragmentation and Sequencing (RF-Seq): Cost-Effective, Time-Efficient, and High-Throughput 3' mRNA Sequencing Library Construction in a Single Tube. *Current Protocols in Molecular Biology*, 129(1). <https://doi.org/10.1002/cpmb.109>
- Venkitaraman, A. R. (2002). Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell*, 108(2), 171–182.
- Vihma, H., Luhakooder, M., Pruunsild, P., & Timmusk, T. (2016). Regulation of different human NFAT isoforms by neuronal activity. *Journal of Neurochemistry*, 137(3), 394–408.
- Vinay, D. S., Ryan, E. P., Pawelec, G., Talib, W. H., Stagg, J., Elkord, E., Lichtor, T., Decker, W. K., Whelan, R. L., Kumara, H. M. C. S., Signori, E., Honoki, K., Georgakilas, A. G., Amin, A., Helferich, W. G., Boosani, C. S., Guha, G., Ciriolo, M. R., Chen, S., ... Kwon, B. S. (2015). Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Seminars in Cancer Biology*, 35, S185–S198. <https://doi.org/https://doi.org/10.1016/j.semcancer.2015.03.004>

- Visser, J., Feuerstein, I., Stangler, T., Schmiederer, T., Fritsch, C., & Schiestl, M. (2013). Physicochemical and functional comparability between the proposed biosimilar rituximab GP2013 and originator rituximab. *BioDrugs*, 27(5), 495–507.
- Vlasak, J., Bussat, M. C., Wang, S., Wagner-Rousset, E., Schaefer, M., Klinguer-Hamour, C., Kirchmeier, M., Corvaia, N., Ionescu, R., & Beck, A. (2009). Identification and characterization of asparagine deamidation in the light chain CDR1 of a humanized IgG1 antibody. *Analytical Biochemistry*, 392(2), 145–154.
<https://doi.org/10.1016/j.ab.2009.05.043>
- Vo, Q. N., Kim, W.-J., Cvitanovic, L., Boudreau, D. A., Ginzinger, D. G., & Brown, K. D. (2004). The ATM gene is a target for epigenetic silencing in locally advanced breast cancer. *Oncogene*, 23(58), 9432–9437. <https://doi.org/10.1038/sj.onc.1208092>
- Vogel, C. L., Cobleigh, M. A., Tripathy, D., Gutheil, J. C., Harris, L. N., Fehrenbacher, L., Slamon, D. J., Murphy, M., Novotny, W. F., Burchmore, M., Shak, S., Stewart, S. J., & Press, M. (2002). Efficacy and Safety of Trastuzumab as a Single Agent in First-Line Treatment of HER2-Overexpressing Metastatic Breast Cancer. *Journal of Clinical Oncology*, 20(3), 719–726. <https://doi.org/10.1200/JCO.20.3.719>
- Vogelstein, B., Lane, D., & Levine, A. J. (2000). Surfing the p53 network. *Nature*, 408(6810), 307–310.
- Von Heyde, S. Der, Wagner, S., Czerny, A., Nietert, M., Ludewig, F., Salinas-Riester, G., Arlt, D., & Beißbarth, T. (2015). mRNA profiling reveals determinants of trastuzumab efficiency in HER2-positive breast cancer. *PLoS ONE*, 10(2), 1–27.
<https://doi.org/10.1371/journal.pone.0117818>
- von Minckwitz, G., Procter, M., de Azambuja, E., Zardavas, D., Benyunes, M., Viale, G., Suter, T., Arahmani, A., Rouchet, N., & Clark, E. (2017). Adjuvant pertuzumab and trastuzumab in early HER2-positive breast cancer. *New England Journal of Medicine*, 377(2), 122–131.
- Vu, T., & Claret, F. X. (2012). Trastuzumab: Updated Mechanisms of Action and Resistance in Breast Cancer. *Frontiers in Oncology*, 2(June), 1–6.
<https://doi.org/10.3389/fonc.2012.00062>

- Vyas, S., Zaganjor, E., & Haigis, M. C. (2016). Mitochondria and cancer. *Cell*, 166(3), 555–566.
- Wahjudi, L. W., Bernhardt, S., Abnaof, K., Horak, P., Kreutzfeldt, S., Heining, C., Borgoni, S., Becki, C., Berg, D., Richter, D., Hutter, B., Uhrig, S., Pfütze, K., Leichsenring, J., Glimm, H., Brors, B., von Kalle, C., Stenzinger, A., Korf, U., ... Wiemann, S. (2021). Integrating proteomics into precision oncology. *International Journal of Cancer*, 148(6). <https://doi.org/10.1002/ijc.33301>
- Waldmann, T. A. (1984). Monoclonal antibodies in diagnosis and therapy. *Science*, 226(4675), 1296–1303.
- Walsh, G. (2018). Biosimilars: pharmacovigilance considerations. *Clinical Therapeutics*, 40(7), 1152–1160.
- Wang, H., Deng, Y., Nong, J., Huang, Y., Wang, X., Liu, Y., Li, Y., & Zhang, Y. (2018). Identification of key pathways and genes in response to trastuzumab treatment in breast cancer using bioinformatics analysis. *Oncology Letters*, 15(4), 4683–4688.
- Wang, J., & Wu, S.-G. (2023). Breast cancer: an overview of current therapeutic strategies, challenge, and perspectives. *Breast Cancer: Targets and Therapy*, 721–730.
- Wang, J., & Xu, B. (2019). Targeted therapeutic options and future perspectives for HER2-positive breast cancer. *Signal Transduction and Targeted Therapy*, 4(1), 34.
- Wang, S., & Ding, Z. (2019). Analysis of HER2 genomic binding in breast cancer cells identifies a global role in direct gene regulation. *PloS One*, 14(11), e0225180.
- Wang, Y., & Zhou, B. P. (2016). Biomarkers for EMT and MET in breast cancer: An update. *Oncology Letters*, 12(6), 4869–4876.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57–63.
<https://doi.org/10.1038/nrg2484>
- Wang, Z., Li, Y., Zhao, W., Jiang, S., Huang, Y., Hou, J., Zhang, X., Zhai, Z., Yang, C., Wang, J., Zhu, J., Pan, J., Jiang, W., Li, Z., Ye, M., Tan, M., Jiang, H., & Dang, Y. (2023). Integrative multi-omics and drug-response characterization of patient-derived

- prostate cancer primary cells. *Signal Transduction and Targeted Therapy*, 8(1).
<https://doi.org/10.1038/s41392-023-01393-9>
- Webb, A. E., & Brunet, A. (2016). FOXO transcription factors: key regulators of cellular quality control. *Trends in Biochemical Sciences*, 41(6), quality control.
- Wei, L. M., Li, X. Y., Wang, Z. M., Wang, Y. K., Yao, G., Fan, J. H., & Wang, X. S. (2021). Identification of hub genes in triple-negative breast cancer by integrated bioinformatics analysis. *Gland Surgery*, 10(1). <https://doi.org/10.21037/gs-21-17>
- Weigelt, B., Lo, A. T., Park, C. C., Gray, J. W., & Bissell, M. J. (2010). HER2 signaling pathway activation and response of breast cancer cells to HER2-targeting agents is dependent strongly on the 3D microenvironment. *Breast Cancer Research and Treatment*, 122(1), 35–43.
- Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., & Stuart, J. M. (2013). The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, 45(10), 1113–1120.
- Wells, E., & Robinson, A. S. (2017). Cellular engineering for therapeutic protein production: product quality, host modification, and process improvement. In *Biotechnology Journal* (Vol. 12, Issue 1). <https://doi.org/10.1002/biot.201600105>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.
https://doi.org/10.1111/j.1467-985X.2010.00676_9.x
- Witt, B. L., & Tollefsbol, T. O. (2023). Molecular, Cellular, and Technical Aspects of Breast Cancer Cell Lines as a Foundational Tool in Cancer Research. In *Life* (Vol. 13, Issue 12). <https://doi.org/10.3390/life13122311>
- Wolff, A. C., Hammond, M. E. H., Schwartz, J. N., Hagerty, K. L., Allred, D. C., Cote, R. J., Dowsett, M., Fitzgibbons, P. L., Hanna, W. M., & Langer, A. (2007). American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Archives of Pathology & Laboratory Medicine*, 131(1), 18–43.
- Wolff-Holz, E., Tiitso, K., Vleminckx, C., & Weise, M. (2019). Evolution of the EU biosimilar framework: past and future. *BioDrugs*, 33(6), 621–634.

- Wong, A. L. A., & Lee, S.-C. (2012). Mechanisms of resistance to trastuzumab and novel therapeutic strategies in HER2-positive breast cancer. *International Journal of Breast Cancer*, 2012(1), 415170.
- World Health Organization. (2009). *Guidelines on evaluation of similar biotherapeutic products (SBPs)*.
- World Health Organization. (2017). *WHO Model List of Essential Medicines*.
[https://doi.org/10.1016/S1473-3099\(14\)70780-7](https://doi.org/10.1016/S1473-3099(14)70780-7)
- Wu, H., Wang, C., & Wu, Z. (2015). PROPER: Comprehensive power evaluation for differential expression using RNA-seq. *Bioinformatics*.
<https://doi.org/10.1093/bioinformatics/btu640>
- Wu, S. Z., Al-Eryani, G., Roden, D. L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J. R., Bartonicek, N., Wang, T., Larsson, L., Kaczorowski, D., Weisenfeld, N. I., Uytingco, C. R., Chew, J. G., Bent, Z. W., Chan, C.-L., Gnanasambandapillai, V., ... Swarbrick, A. (2021). A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics*, 53(9), 1334–1347.
<https://doi.org/10.1038/s41588-021-00911-1>
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation*, 2(3).
<https://doi.org/10.1016/j.xinn.2021.100141>
- Wypych, J., Li, M., Guo, A., Zhang, Z., Martinez, T., Allen, M. J., Fodor, S., Kelner, D. N., Flynn, G. C., Liu, Y. D., Bondarenko, P. V., Ricci, M. S., Dillon, T. M., & Balland, A. (2008). Human IgG2 antibodies display disulfide-mediated structural isoforms. *Journal of Biological Chemistry*, 283(23), 16194–16205.
<https://doi.org/10.1074/jbc.M709987200>
- Xie, L., Zhang, E., Xu, Y., Gao, W., Wang, L., Xie, M. H., Qin, P., Lu, L., Li, S., Shen, P., Jiang, W., & Liu, S. (2020). Demonstrating Analytical Similarity of Trastuzumab Biosimilar HLX02 to Herceptin® with a Panel of Sensitive and Orthogonal Methods

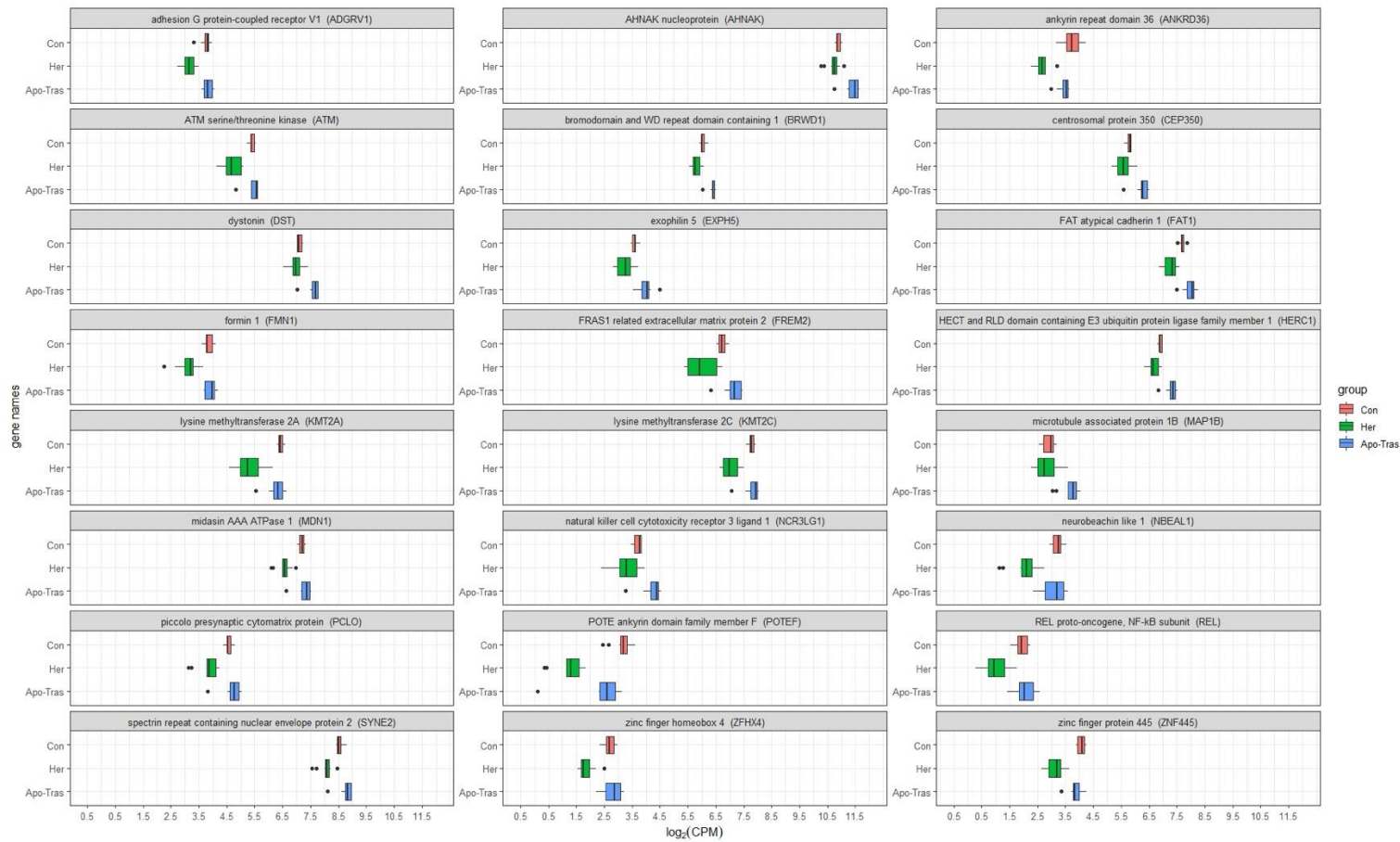
- Including a Novel FcγRIIIa. *BioDrugs*, 34(3), 363–379.
<https://doi.org/10.1007/s40259-020-00407-0>
- Xue, J., Liu, Y., Wan, L., & Zhu, Y. (2020). Comprehensive analysis of differential gene expression to identify common gene signatures in multiple cancers. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 26, e919953–1.
- Yakes, F. M., Chinratanalab, W., Ritter, C. A., King, W., Seelig, S., & Arteaga, C. L. (2002). Herceptin-induced inhibition of phosphatidylinositol-3 kinase and Akt is required for antibody-mediated effects on p27, cyclin D1, and antitumor action. *Cancer Research*, 62(14), 4132–4141. <https://doi.org/10.1200/JCO.2005.05.4692>
- Yang, F., Zhou, L., Shen, Y., Wang, X., Fan, X., & Yang, L. (2024). Multi-omics approaches for drug-response characterization in primary biliary cholangitis and autoimmune hepatitis variant syndrome. *Journal of Translational Medicine*, 22(1), 214.
- Yang, Y., Ye, W.-L., Zhang, R.-N., He, X.-S., Wang, J.-R., Liu, Y.-X., Wang, Y., Yang, X.-M., Zhang, Y.-J., & Gan, W.-J. (2021). The role of TGF-β signaling pathways in cancer and its potential as a therapeutic target. *Evidence-Based Complementary and Alternative Medicine*, 2021(1), 6675208.
- Yarden, Y., & Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nature Reviews Molecular Cell Biology*, 2(2), 127–137.
- Yuan, Q., Deng, D., Pan, C., Ren, J., Wei, T., Wu, Z., Zhang, B., Li, S., Yin, P., & Shang, D. (2022). Integration of transcriptomics, proteomics, and metabolomics data to reveal HER2-associated metabolic heterogeneity in gastric cancer with response to immunotherapy and neoadjuvant chemotherapy. *Frontiers in Immunology*, 13, 951137.
- Zahavi, D., & Weiner, L. (2018). Monoclonal Antibodies in Cancer Therapy. *Antibodies*, 7(4), 38. <https://doi.org/10.3390/antib7040038>
- Zazo, S., González-Alonso, P., Martín-Aparicio, E., Chamizo, C., Cristóbal, I., Arpí, O., Rovira, A., Albanell, J., Eroles, P., Lluch, A., Madoz-Gúrpide, J., & Rojo, F. (2016). Generation, characterization, and maintenance of trastuzumab-resistant HER2+ breast cancer cell lines. *American Journal of Cancer Research*, 6(11).

- Zeng, F., Fu, J., Hu, F., Tang, Y., Fang, X., Zeng, F., & Chu, Y. (2018). Identification of key pathways and genes in response to trastuzumab treatment in breast cancer using bioinformatics analysis. *Oncotarget*, 9(63), 32149.
- Zhang, F. (2012). Dysregulated lipid metabolism in cancer. *World Journal of Biological Chemistry*, 3(8), 167. <https://doi.org/10.4331/wjbc.v3.i8.167>
- Zhao, S., Li, C. I., Guo, Y., Sheng, Q., & Shyr, Y. (2018). RnaSeqSampleSize: Real data based sample size estimation for RNA sequencing. *BMC Bioinformatics*. <https://doi.org/10.1186/s12859-018-2191-5>
- Zheng, L., Foley, K., Huang, L., Leubner, A., Mo, G., Olino, K., Edil, B. H., Mizuma, M., Sharma, R., Le, D. T., Jaffee, E. M., Anders, R. A., Maitra, A., & Laheru, D. (2011). Tyrosine 23 phosphorylation of annexin A2 promotes proliferation, invasion, and STAT3 phosphorylation in the nucleus of human breast cancer cells. *Journal of Cell Science*, 124(Pt 4), 591–600. <https://doi.org/10.1242/jcs.075275>
- Zhou, X., Tan, M., Hawthorne, V. S., Klos, K. S., Lan, K. H., Yang, Y., Yang, W., Smith, T. L., Shi, D., & Yu, D. (2004). Activation of the Akt/mammalian target of rapamycin/4E-BP1 pathway by ErbB2 overexpression predicts tumor progression in breast cancers. *Clinical Cancer Research*, 10(20), 6779–6788. <https://doi.org/10.1158/1078-0432.CCR-04-0112>

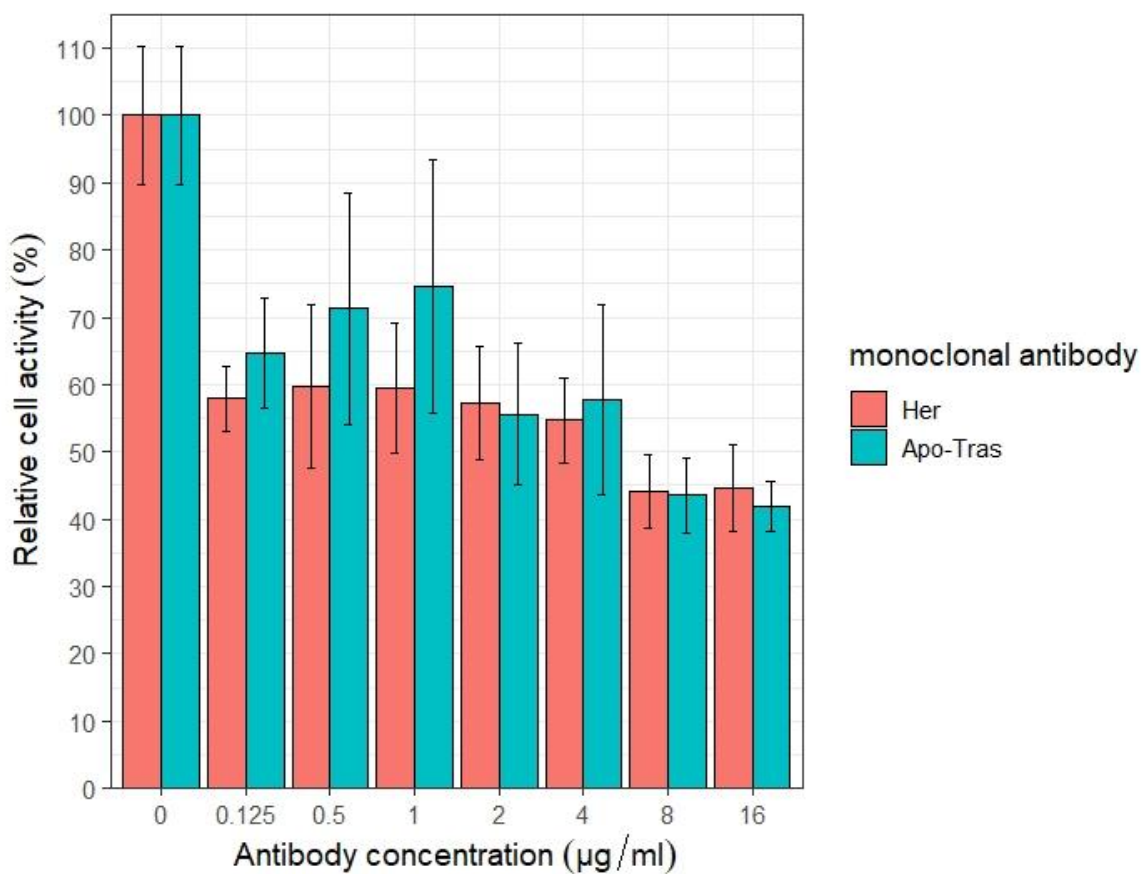
Appendix A

	Con1	Con2	Con3	Con4	Con5	Con6	Con7	Con8	Con9	Con10	Her1	Her2	Her3	Her4	Her5	Her6	Her7	Her8	Her9	Her10	Apo-Tras1	Apo-Tras2	Apo-Tras3	Apo-Tras4	Apo-Tras5	Apo-Tras6	Apo-Tras7	Apo-Tras8	Apo-Tras9	Apo-Tras10
Con1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,97	0,98	0,98	0,96	0,97	0,96	0,98	0,96	0,96	0,98	0,97	0,98	0,63	0,97	0,98	0,98	0,98	0,98	0,97	0,97
Con2		1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,96	0,98	0,98	0,95	0,97	0,95	0,97	0,95	0,95	0,97	0,97	0,98	0,61	0,97	0,97	0,98	0,98	0,97	0,96	0,97
Con3			1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,97	0,98	0,98	0,96	0,97	0,96	0,97	0,95	0,96	0,97	0,97	0,98	0,62	0,97	0,97	0,98	0,98	0,98	0,96	0,97
Con4				1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,96	0,98	0,98	0,96	0,97	0,95	0,97	0,95	0,95	0,97	0,97	0,98	0,61	0,97	0,97	0,98	0,98	0,97	0,96	0,97
Con5					1,00	1,00	1,00	1,00	1,00	1,00	0,96	0,97	0,98	0,95	0,96	0,95	0,97	0,95	0,95	0,97	0,97	0,97	0,60	0,97	0,97	0,97	0,97	0,97	0,96	0,97
Con6						1,00	1,00	1,00	1,00	1,00	0,97	0,98	0,98	0,96	0,97	0,96	0,97	0,96	0,96	0,98	0,97	0,98	0,63	0,97	0,97	0,98	0,98	0,98	0,97	0,97
Con7							1,00	1,00	1,00	1,00	0,96	0,97	0,98	0,95	0,96	0,95	0,97	0,95	0,95	0,97	0,97	0,98	0,60	0,97	0,97	0,98	0,98	0,97	0,96	0,97
Con8								1,00	1,00	1,00	0,96	0,98	0,98	0,96	0,97	0,96	0,97	0,95	0,96	0,97	0,97	0,98	0,62	0,97	0,97	0,98	0,98	0,97	0,96	0,97
Con9									1,00	1,00	0,96	0,98	0,98	0,96	0,97	0,96	0,97	0,95	0,96	0,97	0,97	0,98	0,62	0,97	0,97	0,98	0,98	0,97	0,96	0,97
Con10										1,00	0,96	0,97	0,98	0,95	0,96	0,95	0,97	0,94	0,95	0,97	0,97	0,98	0,60	0,97	0,97	0,98	0,98	0,97	0,96	0,97
Her1											1,00	0,99	0,99	1,00	1,00	1,00	0,99	1,00	1,00	1,00	0,99	0,98	0,73	0,98	0,99	0,98	0,98	0,99	1,00	0,99
Her2												1,00	1,00	0,99	1,00	0,99	1,00	0,98	0,99	1,00	1,00	1,00	0,67	1,00	1,00	1,00	1,00	1,00	0,99	1,00
Her3													1,00	0,99	0,99	0,99	1,00	0,98	0,99	1,00	1,00	1,00	0,67	1,00	1,00	1,00	1,00	1,00	0,99	1,00
Her4														1,00	1,00	1,00	0,99	1,00	1,00	0,99	0,98	0,97	0,74	0,98	0,99	0,98	0,98	0,98	1,00	0,99
Her5															1,00	1,00	1,00	1,00	1,00	0,99	0,99	0,72	0,99	1,00	0,99	0,99	0,99	0,99	1,00	0,99
Her6																1,00	0,99	1,00	1,00	1,00	0,99	0,98	0,74	0,98	0,99	0,98	0,98	0,99	1,00	0,99
Her7																	1,00	0,99	0,99	1,00	1,00	0,99	0,70	1,00	1,00	0,99	0,99	1,00	1,00	1,00
Her8																		1,00	1,00	0,99	0,98	0,97	0,75	0,98	0,99	0,97	0,97	0,98	1,00	0,98
Her9																			1,00	0,99	0,99	0,98	0,74	0,98	0,99	0,98	0,98	0,99	1,00	0,99
Her10																				1,00	1,00	0,99	0,70	0,99	1,00	0,99	0,99	1,00	1,00	1,00
Apo-Tras1																					1,00	1,00	0,68	1,00	1,00	1,00	1,00	1,00	0,99	1,00
Apo-Tras2																						1,00	0,65	1,00	1,00	1,00	1,00	1,00	0,98	0,99
Apo-Tras3																							1,00	0,67	0,69	0,66	0,65	0,67	0,73	0,68
Apo-Tras4																								1,00	1,00	1,00	1,00	1,00	0,99	1,00
Apo-Tras5																									1,00	1,00	1,00	1,00	1,00	1,00
Apo-Tras6																									1,00	1,00	1,00	0,99	1,00	
Apo-Tras7																										1,00	1,00	0,98	0,99	
Apo-Tras8																											1,00	0,99	1,00	
Apo-Tras9																												1,00	0,99	
Apo-Tras10																													1,00	

Appendix Figure A 1 Pearson correlation matrix for 3649 housekeeping genes for all 30 samples. The labels are Control (Con), Herceptin (Her), biosimilar (ApoTras). The correlation values are colour coded according to the bar on the right side. All samples are highly similar beside the sample ApoTras3 which was excluded. Values are rounded to two decimals.

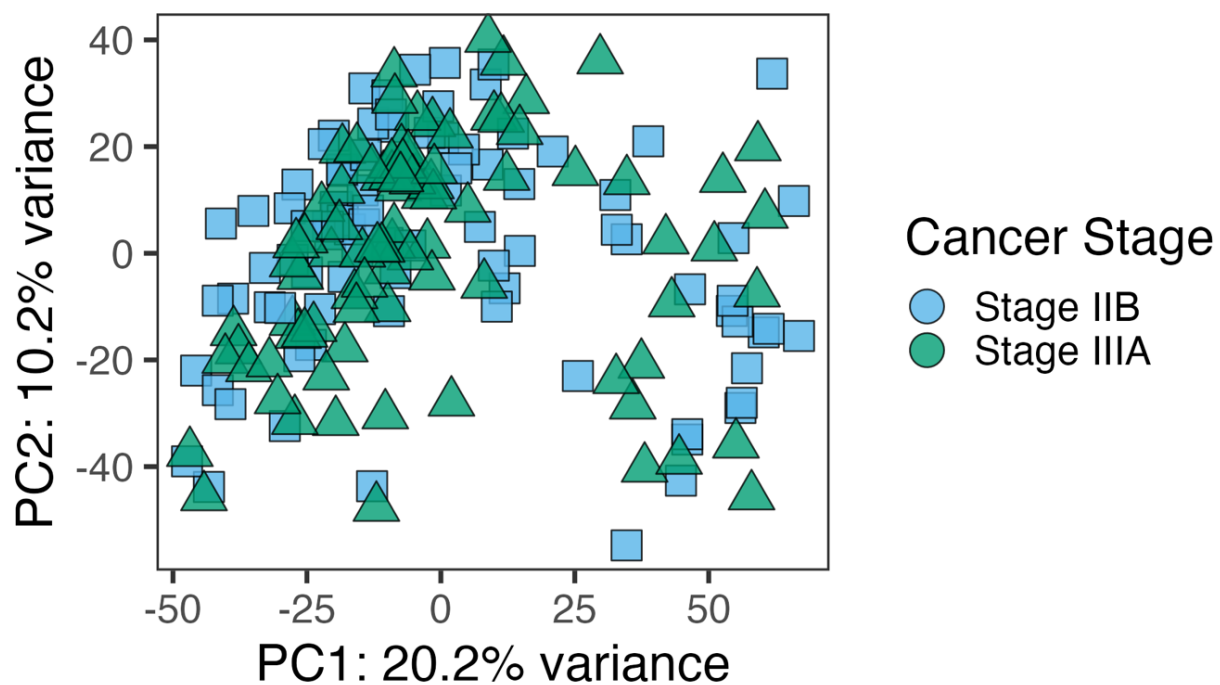


Appendix Figure A 2 Boxplot of expression values of 24 DEG from the ApoTras vs. Her. Numbers of replication are 10 for Con, 10 for Her and 9 for ApoTras. The variance of the expression values is relatively low, but the differences are also rather small.

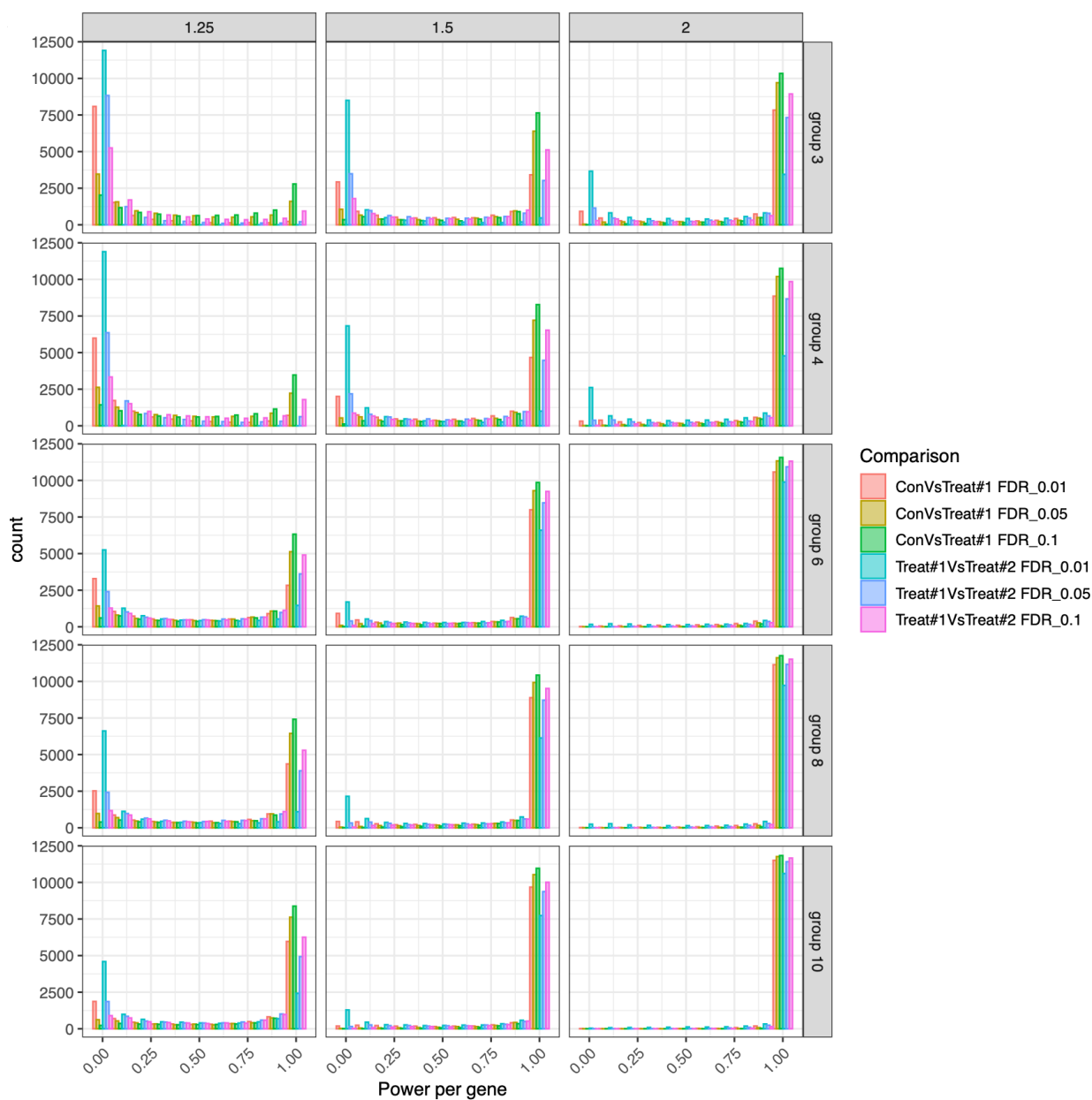


Appendix Figure A 3 Sensitivity test of BT474 cell line for Herceptin (Her) and biosimilar (ApoTras). The number of replicates is six per concentration. The cells were exposed to different concentrations of mAbs for 6 days. the media was replaced every 72 hours. The cell activity was measured with WST-1 colorimetric assay. The ratio of the absorbance to that of the non-treated cells represents the sensitivity of cells to Her or ApoTras. No significant differences were measured between the treatment with ApoTras and Her.

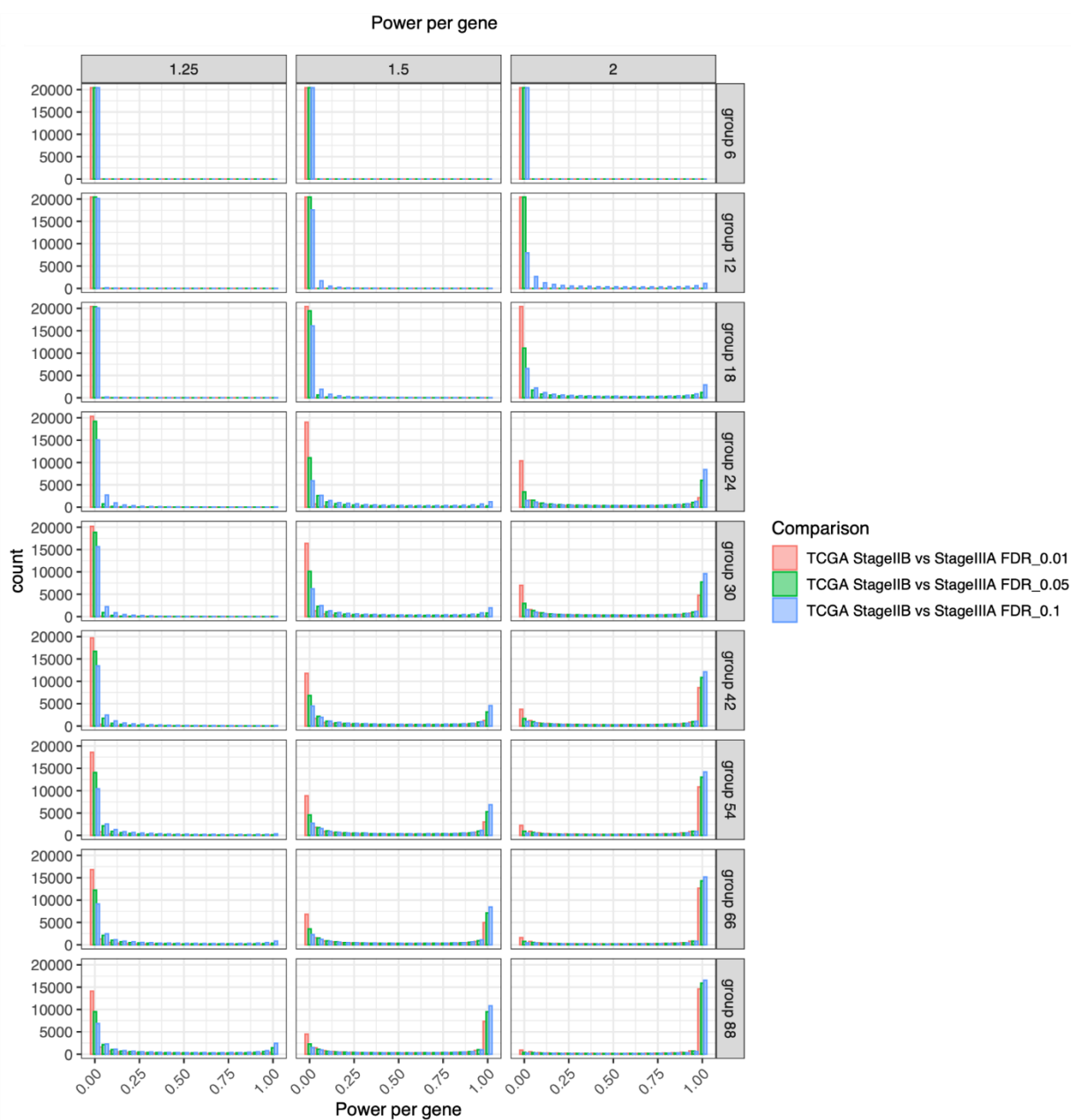
Appendix B



Appendix Figure B 1 PCA plot of TCGA BRCA data set for the 500 most variable genes. The axis labels show the represented variance of the principal component. 88 samples per cancer stage are represented.



Appendix Figure B 2 The histograms show the power distribution of genes for the treated cell line datasets for a subset of parameters. The genes were split into three groups of equal size based on the expression level (lcpm) with the expression ranges noted at the top. The number of genes per group is 3970 (treated cell line). Histograms of power distribution for treated cell line dataset for an effect size of 1.25, 1.5, and 2, and FDR thresholds of 10%, 5%, and 1%, and group size numbers of three, four, six, eight, and ten.



Appendix Figure B 3 The histograms show the power distribution of genes for the TCGA datasets for a subset of parameters. The genes were split into two groups of equal size based on the expression level (lcpm). The number of genes per group is 6805 (TCGA dataset). Histograms of power distribution for treated cell line dataset for an effect size of 1.25, 1.5, and 2, and FDR thresholds of 10%, 5%, and 1%, and group size numbers of three, four, six, eight, and ten.