Università degli Studi di Napoli Federico II
Ph.D. Program in
Computational and Quantitative Biology
XXXVI Cycle

Thesis for the Degree of Doctor of Philosophy

Single cell lineage tracing reveals clonal dynamics of anti-EGFR therapy resistance in Triple Negative Breast Cancer

by
Melania Franchini

Advisor: Prof. Diego di Bernardo
Co-advisor: Dr. Gennaro Gambardella

23/02/2024

Scuola Politecnica e delle Scienze di Base
Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione

"Dicette ó pappice á noce
damme ó tiempo ca te spertose"

Neapolitan proverb

Single cell lineage tracing reveals clonal dynamics of anti-EGFR therapy resistance in Triple Negative Breast Cancer

Ph.D. Thesis presented
for the fulfillment of the Degree of Doctor of Philosophy
in Computational and Quantitative Biology
by

Melania Franchini

January 2024

Approved as to style and content by

_____

Prof. Diego di Bernardo, Advisor

_____

Dr. Gennaro Gambardella, Co-advisor

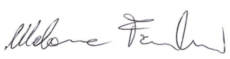
Università degli Studi di Napoli Federico II
Ph.D. Program in Computational and Quantitative Biology
XXXVI cycle - Chairman: Prof. Michele Ceccarelli

https://cqb.dieti.unina.it

Candidate's declaration

I hereby declare that this thesis submitted to obtain the academic degree of Philosophiæ Doctor (Ph.D.) in Computational and Quantitative Biology is my own unaided work, that I have not used other than the sources indicated, and that all direct and indirect sources are acknowledged as references.
Parts of this dissertation have been published in international journals and/or conference articles (see list of the author's publications at the end of the thesis).

Napoli, February 23, 2024



Melania Franchini

Abstract

Targeting the epidermal growth factor receptor (EGFR) in Triple Negative Breast Cancer (TNBC) resulted in varying and unpredictable clinical responses.

To understand the reasons behind this variability, we utilized cellular barcoding and single-cell transcriptomics to reconstruct the subclonal dynamics of EGFR-amplified TNBC cells when exposed to Afatinib, a tyrosine kinase inhibitor (TKI) that irreversibly inhibits EGFR.

Through integrated lineage tracing analysis, we identified a rare pre-existing subpopulation of cells exhibiting a distinctive biological signature, characterized by enhanced expression levels of IGFBP2 (Insulin-Like Growth Factor Binding Protein 2). Our findings demonstrate that overexpression of IGFBP2 is sufficient to confer tolerance to Afatinib treatment in TNBC cells by activating the compensatory IGF1-R signaling pathway.

Lastly, we employed deep learning techniques to create an algorithm that predicts the sensitivity of TNBC cells to Afatinib based on the reconstructed mechanisms of resistance.

In conclusion, our approach has proven successful in unraveling the intricate signaling network responsible for resistance to EGFR-targeted therapy, providing novel insights for tailoring individualized treatment strategies in TNBC.

Keywords: lineage tracing, scRNA-seq, TNBC, EGFR-TKIs.

Sintesi in lingua italiana

Il trattamento mirato al recettore del fattore di crescita epidermico (EGFR) nel cancro al seno Triplo Negativo (TNBC) ha portato a risposte cliniche variabili ed imprevedibili.
Per comprendere le ragioni di questa variabilità, abbiamo utilizzato il barcoding cellulare e la trascrittomica a singola cellula per ricostruire la dinamica subclonale delle cellule TNBC amplificate per EGFR quando esposte ad Afatinib, un inibitore della tirosin-chinasi (TKI) che inibisce irreversibilmente l'EGFR.
Attraverso un'analisi integrata di ricostruzione clonale delle linee cellulari, abbiamo identificato una rara sottopopolazione pre-esistente di cellule che mostra una distintività biologica, caratterizzata da elevati livelli di espressione di IGFBP2 (Insulin-Like Growth Factor Binding Protein 2).
I nostri risultati dimostrano che la sovraespressione di IGFBP2 è sufficiente a conferire tolleranza al trattamento con Afatinib nelle cellule TNBC attivando la via di segnalazione compensatoria dell'IGF1-R.
Grazie all'impiego di tecniche di deep learning abbiamo sviluppato un algoritmo che predice la sensibilità delle cellule TNBC ad Afatinib sulla base dei meccanismi di resistenza ricostruiti.
In conclusione, il nostro approccio si è dimostrato efficace nello svelare la complessa rete di segnalazione responsabile della resistenza alla terapia mirata contro l'EGFR, fornendo nuove informazioni per adattare strategie di trattamento personalizzate nei casi di TNBC.

Parole chiave: ricostruzione clonale, sequenziamento a singola cellula, tumore al seno triplo negativo, fattore epidermico della crescita.

Acknowledgements

A Ph.D. is defined in a thousand ways: "The worst experience of my life" or "Couldn't I have opened a bakery?"
For me, it was a wild, formative, emotionally challenging experience, full of tears, frustration and sense of inadequacy, but also smiles, victories, beautiful days, and wonderful people.
The first thanks goes to Gennaro. For transmitting me Mondays' traumas when, stepping into his office for weekly meetings, I knew I would face nervous discussions, but I almost never left without a new idea.
The next one goes to Diego, who offered me a chance to join a diverse and interdisciplinary group and for its quiet yet influential presence, notably in lab meetings when I just wanted to run away.

This is a thesis on one of the biggest women nightmare, written by a woman, thinking about the stories of many other women, so, the first thank goes to you Mamma, for being a determined, strong, free woman who, unlike the one from your favorite writer, Leopardi, always pushed me to kick beyond the hedge and bite into the infinite. Hey mom, you were right, it tastes good!
Thank you Babbo, for being a good, perhaps too good man, with that gruff and complicated character that fortunately you transmitted to me. If I've grown into a "pappice," it's because of you, but now, cracking the nut isn't sufficient. Shall we aim for new horizons together?
Last but not least, thank you Federico. It's devastating and scary to know that your heart has legs of its own, but I'm calm knowing that it's still awake, vigilant, and ironic about the world, guiding me with laughter and jazz, discovering corners I never imagined looking at.

It is said that a person is not defined only by themselves or their his-

tory. I don't know what is true, but I do know that without my friends, I wouldn't be who I am.
So, thanks to Grazia, Miss Ostetria, for the fabulous work she does, for the passion, dedication, calmness, irony, and for all the supportive chats over these years. In her eyes, I may appear a superhero, but for me, she's the true one.
Thanks to Lucia, Coco, for being the historical and stoic memory of who I was, I am, and, I hope, I will be. She is one of those people who, despite life making you take absurd turns, you always meet again. And thanks also for introducing me to Manfredi (Manfre), for all the musical chats I still remember. Thanks Silvia, la piccola italiana, for the talks about dreams, the cherishing memories of me, and for the unwavering determination. You and the serene Domenico yet occasionally contentious, deeply inspiring those who observe your dynamic relationship.
Thanks Veronica for your energetic way of acting and for being my biggest fan since the elementary school and thanks Candida for your quiet and gentle way of talking about your ideas.

Sometimes you grow with people by age while sometimes by experience, so thanks Fabiola, Dotti, for being a disillusioned person who still believes, for the constant laughs, silent support, alcoholic evenings, and that turbulent journey to rediscover the new me.
Thanks Luigi, the dictator, for the evenings talking about everything and nothing, the journey we've made together, and for being my conscience/talking brake in all the cases where I wanted to break everything.

A PdD and the scientific career are complex experiences and, sometimes, life could be even worse. So meet the right people at the right time could save your life, or, at least, your mind.
Thanks Luca, for that genetic network hand-drawn like a cartographer

that still reminds me that, as you used to say, "Power without control is useless".
Thanks Sergio, Seggi, for being the guiding figure, a bit of an older brother, especially in the first years of this journey. Your little Marie is lucky to have a dad like you. Thanks Antonella, Nannonella, for your constant energy and for making statistics plastic and understandable, and for all the times you let us laugh at your oddities. Thanks Alessandra, Parapallo, for your calm way of speaking and constant questions about what is right in both professional and personal life.
Thanks Stefania, my psycologist, for being the soft mattress to land on during that leap into the void that was the journey into my mind.

Even though no one wants to admit it, we spend the majority of the time at work, so some of its dynamics still remains in your attitude.
So, a huge thank goes to the guys from the Bioinformatics Core : la signora madre (Rossella), Diego, Gegge (Eugenio) e zio Spagna (Xabi) for not kicking me out when I went crazy, for all the chats and exchanges of opinions and the hilarious jokes.

No one is an island and most of the time is spent at work. But what if we could have time with wonderful people at work?
Thanks Adele (Stella) and Virginia (Cucciolo), I can't separate them, for being one the opposite of the other, one sensitive and the other tough, one reflective and the other impulsive. Thanks Virgi for teaching me how to walk into the storm with a smile and thanks Adele for crying the tears I don't have the courage to cry.
Thanks Riccardo, Arni, for patience, laughter, irony, days spent thinking, and, above all, for teaching me what it means to supervise someone. It was an honor to see you take your first steps in research, and I hope to be able to do it again.

Thanks Simona for working and screaming side by side for the results presented in this thesis. Hey, Simo, we didi it!
Thanks Gaetano, for being the carefree Gae, for your unwavering calm even in tense moments and your thoughtful advice in tranquility ones.
Thanks Stefania, Raggio di Sole, for your straightforward and gruff way, magnetism, irreverent questions, and laughs. Thanks Mario, zio Satana, for all the jokes on the edge of human, forget legality, and for always having a comforting word. Thanks Francesco, Franozzo, for your sometimes shouted and stinging, sometimes peaceful points of view, which enchanted me for the open-mindedness. Thanks Giuliano, Yogi, for all your unimaginable outfits and disorder, your stoicism in carrying forward your projects with a molecular biology that I didn't think could exist. Thanks Iacopo, TFEB, for all your "Teano-style" outings that made us bend over laughing and for your almost maniacal order that I have always admired. Thanks Lorena, Lori, for not making me feel alone and for the advice on the future. Thanks Alessio, the prince, for raising the level of fashion in tha lab and for the "dB per il sociale" that I still remember as lively and constructive.

While I write these final lines, Negrita's song continues to resonate in my head. I was listening to it when I found out I was admitted to the PhD program and was a bit of a premonition. I will use it to thank the last person, the teenager/young adult Melli who dreamed of, one day, giving cancer patients care opportunities that the important people in her life did not have. To them, I dedicate all the effort of these and the coming years, to her, instead, I dedicate this song.

"È un biglietto per le stelle, quello lì davanti a te
Cambierai la pelle, ma resta speciale, non ti buttare via
In questo inferno, di ombre piatte, in questo vecchio luna park
Resta ribelle, non ti buttare via"

Contents

List of Acronyms

The following acronyms are used throughout the thesis.

NGS Next-Generation Sequencing

scRNA-seq Single-cell RNA sequencing

ITH Intra-Tumor Heterogeneity

TNBC riple Negative Breast Cancer

EGFR Epidermal Growth Factor Receptor

EGFR-TKI Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor

IGFBP2 Insulin-like Growth Factor Binding Protein 2

List of Figures

List of Tables

Chapter 1

Introduction

Intra-tumor heterogeneity (ITH) refers to the coexistence of diverse cancer cell populations within a single tumor, presenting a substantial challenge in the development of effective treatment strategies in clinical practice.This is due to the high cellular plasticity and phenotypic diversity, reflecting the complex interplay of genetic and non-genetic factors that enable cancer cells to survive and proliferate despite anti-cancer treatments [9] [10] [1] [11]. Accumulating evidence from independent studies demonstrates that ITH gives rise to rare sub-populations of malignant cells characterized by distinct epigenetic and transcriptional profiles, rendering them refractory to a variety of anticancer drugs [12]. These drug-tolerant subpopulations play a pivotal role in cancer recurrence and treatment resistance, protecting the entire tumor from eradication.

Single-cell RNA sequencing (scRNA-seq) has emerged as a powerful tool to elucidate the intricate landscape of intra-tumor heterogeneity [13] [14] [15], identify rare subpopulations of cells with unique gene expression profiles, and pinpoint key genes that undergo dynamic changes in response to treatments [16] [17]. However, to effectively monitor the emergence of drug-resistant cell populations, it is also crucial to track how surviving clones adapt to the treatment and to compare their evolutionary trajectories. To this end, cellular barcoding [18] has emerged as a powerful DNA marking technique that enables tracking of individual cells within a complex population [16] [19] [20]. This technique utilizes synthetic DNA sequences, or

barcodes, to label individual cells, allowing researchers to trace their fate over time and across different conditions. Cellular barcoding has revolutionized our ability to track cancer progression, metastasis, and cell differentiation at the single-clone level, providing invaluable insights into the intricate dynamics of tumor evolution. Therefore, developing novel methods that integrate DNA marking techniques with techniques that measure cellular states at single-cell resolution is paramount to decipher the diverse adaptive pathways leading to drug resistance.

Triple-Negative Breast Cancer (TNBC) stands as a particularly challenging cancer type due to its inherent aggressiveness, high heterogeneity, and the absence of recurrent oncogenic alterations [13] [14] [15]. Indeed, current therapeutic landscape for TNBC primarily centers on Neoadjuvant Chemotherapies (NAC) even though its effectiveness is often hindered by the frequent development of drug resistance [13] [14] [21]. As a result, there's an urgent need for innovative approaches to identify biomarkers that predict drug response, enabling effective stratification of TNBC patients.
While EGFR-activating mutations and amplifications (≥ 5 copies) are uncommon in TNBC [22] [23] [24] [15], the majority of primary TNBCs exhibit enhanced expression of EGFR due to an increase in gene copy number (three to four copies) [25] [26] [27] [24] [17] [28], thus representing a valuable vulnerability for TNBC patients [29].
Unlike other tumor types, where inhibition of wild-type EGFR by monoclonal antibodies or tyrosine kinase inhibitors (TKIs) has been showed to be beneficial [30] [31] [32] [33], EGFR-targeted therapies in TNBCs have shown variable and unpredictable clinical responses (1.7% to 38.7%) [34] [35] [36] [37] [38]. However, genomic variants that have been found to be predictive of anti-EGFR resistance in other tumor types [39] are infrequent in TNBC patients [34] [23] [40]. This suggests that non-genomic mechanisms of resistance might play a significant role in TNBC. As a result, the lack of predictive biomarkers for response to anti-EGFR therapies in TNBC has hindered the translation of EGFR inhibitors in breast cancer.

In this study, we leverage single-cell transcriptomics, cellular barcoding, and time-resolved computational analyses to explore the non-genetic mech-

anisms underlying the cellular response to Afatinib, a highly potent EGFR and HER2 inhibitor used in the treatment of various cancers. We focus on MDA-MB-468 cells, a wild-type amplified triple-negative cell line that serves as a preclinical model of TNBC [41] [42] and exhibits high sensitivity to anti-EGFR treatment.
Our approach revealed both pre-existing and adaptive drug resistance mechanisms. Retrospective lineage tracing identified a distinct Afatinib-tolerant MDA-MB-468 subpopulation characterized by elevated IGFBP2 gene expression. We demonstrated through chemical and genetic manipulations that IGFBP2 overexpression is sufficient to render TNBC cells tolerant to afatinib treatment through activation of the compensatory IGF1-R signaling pathway. Prospective lineage tracing, on the other hand, highlighted additional adaptive mechanisms, including lysosome biogenesis, reactive oxygen species (ROS) homeostasis, and fatty acid metabolism.

Finally, we employed deep learning techniques to develop an algorithm predicting the Afatinib sensitivity of TNBC cells based on reconstructed mechanisms of resistance. In conclusion, our findings provide a new understanding of the intricate signaling network underlying EGFR-targeted therapy resistance in TNBC that can be helpful in devising novel strategies for TNBC patient stratification and therapeutic intervention.

Chapter 2 explores the fundamental biological aspects of drug resistance in cancer, with a specific focus on common molecular mechanisms and explanatory models. Furthermore, the chapter provides an in-depth characterization of Triple Negative Breast Cancer, covering its molecular features, clinical behavior, and the challenges encountered in its treatment.

Chapter 3 describes the different scRNA-seq platforms, with related advantages and disadvantages, along with common pipeline used for data analysis.

Chapter 4 explores the state-of-the art of lineage tracing methods with related application in drug response studies.

Chapter 5 contains the description of the approach we used to improve our knowledge on EGFR-TKIs resistance in a cellular model of TNBC (MDA-MB-468).

Chapter 6 summarizes the findings of the present work, emphasizing the unresolved questions that still persist

Chapter 2

Cancer drug resistance and Triple Negative Breast Cancer

2.1 Introduction

This chapter contains the biological basics of cancer drug resistance specifically addressing common molecular mechanisms in Section 2.2 and the models employed to elucidate this behavior, spanning both genetic (2.2.1) and non-genetic (2.2.2) mechanisms.
In Section 2.3, a detailed overview of Breast Cancer molecular and clinical classification is presented with a focus on the characterization of Triple Negative Breast Cancer (2.3.1), encompassing molecular features, clinical behavior, and treatment challenges.

2.2 Cancer drug resistance: Mechanisms and models

Drug resistance in cancer refers to the absence of clinical benefit that patients experience during treatment [43] [11] [44] thereby reducing their chances of survival and emerging as a significant challenge in anticancer therapies [45] [46].
It can be categorized as either intrinsic (primary) or acquired (secondary), depending on the timing at which the decline in therapeutic effectiveness

occurs [45] [47] [1] [44]. More precisely, intrinsic drug resistance is defined by the absence of clinical response to a treatment that has not been administered previously [45] [47] [48], while acquired drug resistance is characterized by the reappearance of malignancy following an initial positive response [45] [48].

The main causes of intrinsic drug resistance involve three main molecular mechanisms: *(i)* inherited genetic alterations in genes associated with proliferation and apoptosis [47] [49] [50] [11] [51]; *(ii)* the presence of unresponsive subpopulations, such as cancer stem cells, contributing to tumor heterogeneity and promoting further tumor relapse [9] [10] [52] [43] [11] and *(iii)* the activation of protective pathways reduces intracellular drug concentrations by overexpressing efflux channels on the plasma membrane and/or by directly detoxifying them [9] [45] [43] or diminishing their effects, as in the case of Reactive Oxygen Species (ROS) detoxification pathway [53] [19] [54].

Conversely, acquired drug resistance emerges through the activation of secondary proto-oncogenes (or the inactivation of tumor suppressors) and their associated signaling pathways. This compensates for the initial signaling blockade induced by the treatment, enabling cells to overcome the initial barrier imposed by the treatment [55] [56] [57].
This adaptation can also involve the modification of drug targets, thereby reducing drug binding [9] [43] [51], or altering the tumor microenvironment (TME), the complex environment surrounding cancer cells composed of various cellular types such as fibroblasts and immune cells [43] [58] [59]. Changes in the TME can lead to drug neutralization, the release of pro-survival factors, and the limitation of drug-induced cytotoxicity, particularly in hypoxic conditions [60] [59].
Extensive studies on individual drug resistance mechanisms revealed their inner nature which could be genetic, non-genetic or at the interplay of these two, as the current knowledge seems to report [61] [43] [1] [62] and it is shown in *Figure* 2.1.

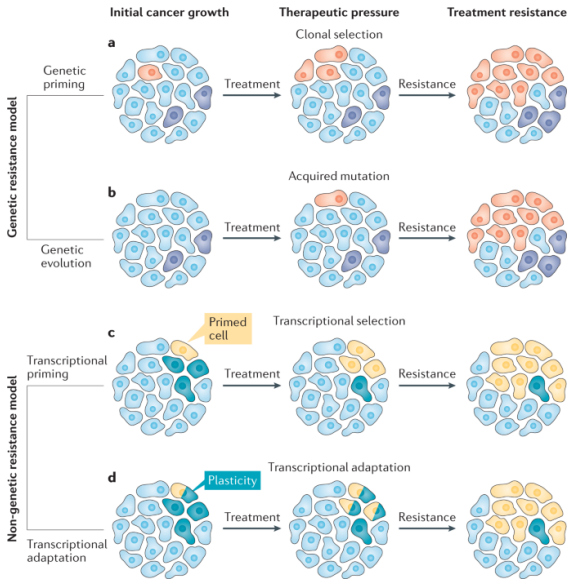


Figure 2.1. Cancer drug resistance models. A) Genetic mechanisms contributing to drug resistance involve pre-existing mutations within a subset of cells, known as the resistant sub-population, which inherently resist specific therapies. **B)** Conversely, cancer cells accumulate mutations during treatment, conferring them a proliferative or survival advantage that counteracts the therapeutic effects. **C)** Non-genetic mechanisms encompass the activation of distinct transcriptional programs within a subset of cells, leading to inherent resistance against specific drugs. **D)** Under therapeutic pressures, cancer cells rapidly adjust by reorganizing their gene expression, acquiring programs that provide a selective advantage within the treatment environment, thereby circumventing therapeutic interventions. *Figure* from [1]

2.2.1 Genetic models of drug resistance: Darwinian evolution

Genetic resistance involves the acquisition of specific genetic alterations, spanning SNV (Single Nucleotide Variations) [63] [64] [65], CNV (Copy Number Variations) [63] [64] and big chromosomal rearrangements [66] [67] [68], allowing cancer cells to evade therapeutic pressure. These changes may be pre-existing, shown in *Figure* 2.1*A*, suggesting intrinsic resistance [47] [49] [50] [11] [51], or they may occur during therapy, as re-

ported in *Figure* 2.1*B*, giving cancer cells a competitive advantage against treatment, thereby contributing to acquired resistance [69] [70].

Multiple models rooted in *Darwinian* theory [71] [72] [61] have been proposed over time to elucidate genetic drug resistance.
According to this theory, genetically independent entities, like cancer cells, display inherent differences. Some of these variances can enhance survival through adaptation, enabling growth under limited resources such as nutrient deficiencies, hypoxic conditions or adverse environments, such as drug treatments, while others may lack these adaptive capabilities [71] [73] [74]. The overall result is the formation of a cellular population fully composed by cells with these favorable genetic setting [75] [70] [72], as shown in *Figure* 2.1*A*, thanks to genetic heterogeneity and clonal selection phenomena.

Genetic heterogeneity involves mutational variability within the primary tumor, known as intra-tumor heterogeneity (ITH) [64] [73] [70] [76], or between primary tumor and related metastases (i.e., inter-tumor heterogeneity) [75] [70].
Intratumor heterogeneity leads to diverse subclones with variable behaviors and fitness levels due to the discordant inheritance of extrachromosomal DNA (ecDNA). This type of DNA often carries oncogenes like MYC, EGFR, or PDGFRA, among the others, [77] [78] [76] and, as it lacks centromeres, it cannot be balanced inheritance between daughter cells, thus amplifying cell-to-cell variability in oncogene copy number and transcriptional levels. This imbalance is further intensified by the accumulation of de novo mutations during metastatic process [79] [70], thus explaining differences in mutational loads between primary tumor and metastases.
Clonal evolution, instead, while directly influenced by ITH is driven by environmental selective pressures shaping selection and survival of the best-fitter subclones, as the drug-resistant ones. This process is usually explained by different models, including the *(i)* linear, *(ii)* neutral, *(iii)* branched and *(iv)* macroevolution models. Each model describes well aspects of the mechanisms while falling short in explaining others [73] [74].

The linear model indicates a stepwise accumulation of genetic alterations in progenitor cells, which offers a selective advantage over the other clones

resulting in enhanced proliferation, thus aligning with Darwinian theory [61] [73] [74] [80].
Regrettably, various studies have demonstrated that this model fails to account for the stability in clonal composition over time [81], as illustrated in the neutral model [82]. Moreover, it does not address the coexistence of multiple subclones [83] [74], a phenomenon elucidated by the branched model.
The neutral evolution model elucidates the diminished variation in clonal composition by considering that tumor cell evolution is characterized by a lack of strong selective pressure [83] [74] [80].
On the other hand, the simultaneous presence of several subclones, a direct consequence of intratumor heterogeneity, involves the acquisition of distinct genetic alterations in different subpopulations of cells as considered in the branched evolution model [83] [74].
All the mentioned models revolve around small genetic changes with effects observable over extended periods, a phenomenon termed microevolution. However, this stands in contrast with the existence of large chromosomal rearrangements such as chromosomal instability, chromoplexy, or chromothripsis, which introduce a multitude of variations in a brief time span. This phenomenon aligns with the macroevolution theory [74] [82].

The existence of different models deriving from different and partially discordant experimental evidence [82] [84], implies that none of them can fully explain the mechanism of cancer evolution and drug resistance, suggesting a potential additional involvement of non-genetic mechanisms.

2.2.2 Non genetic models of drug resistance: Lamarckian evolution

Non-genetic drug resistance involves transcriptional and epigenetic changes promoting intratumor heterogeneity (ITH) and maintaining a sub-cellular population with a pre-existing transcriptional program for survival against external stresses like cancer therapies [85] [84]. This leads to two different effects, namely *(i)* drug tolerance and *(ii)* drug adaptation, depicted in *Figure*2.1*C* and *Figure*2.1*D*, respectively.

Drug tolerance and adaptation share the reversibility property with respect

to genetic drug resistance [86] [87] [88], while differing in the duration and effectiveness of clonal advantage under stressful conditions.
In drug tolerance, cells survive treatment without proliferation, even though they re-acquire sensitivity after drug withdrawal [86] [87] [88].
In contrast, in drug adaptation, cells overcome therapeutic pressure by rapidly rewiring their gene expression programs, processes called transcriptional priming [1] [89] and transcriptional adaptation [1], thus reaching a different stable state [1]. This adaptation can persist even after drug withdrawal, conferring a long-term advantage. These alterations give rise to cellular plasticity, also known as phenotype switching [90] [47] [62], allowing cells to adopt a drug irresponsiveness identity reversibly and dynamically. This plasticity can manifest in various forms, including dedifferentiation [16] [91] [52] [19] [20] [92] and transdifferentiation [93] [90] [94] [95] [89].
Dedifferentiation involves transitioning to a stem-like state through processes like epithelial-mesenchymal transition (EMT), linked to increased drug resistance and aggressiveness [16] [91] [52] [19] [20] [92].
On the contrary, transdifferentiation induces cells to differentiate into a different cellular type with subsequent changes in morphology and function in response to specific triggers [93] [90] [94] [95] [89].
The need for adaptation to changing external conditions, in conjunction with cellular plasticity, mirrors the principles of *Lamarckian* theory [93] [96], where is stated that organisms can flexibly adapt to their environment during their lifetime and transmit these adaptations to their descendants [96] [62]. In the context of cancer, a small proportion of cells can undergo temporary transitions towards a drug-resistant phenotype, guided by epigenetic modifications, such as alterations in chromatin structure, DNA methylation, and histone modifications [97] [86] [98] [99], directly impacting their transcriptional profiles.

In conclusion, tackling drug resistance in anticancer therapies is a complex challenge, requiring a comprehensive understanding of overlapping genetic and non-genetic mechanisms.
Genetic models, such as Darwinian selection, propose pre-existing genetic alterations in drug-tolerant cells, while Lamarckian induction suggests epigenetic changes driving tolerance.
In a recent perspective, there's a focus on the fluctuating expression of

resistance markers, where rare cells manage to survive and develop stable resistance. Within this framework, the utilization of single-cell RNA-seq emerges as the most promising approach for enhancing the detection and characterization of these elusive drug-tolerant cells.

2.3 Breast Cancer: Molecular and clinical characteristics

Breast cancer is the most diagnosed cancer in women worldwide, reporting 2.26 million new cases in 2020 and causing nearly 626,679 deaths annually [100]. This highlights its substantial impact as a global health challenge.
Incidence rates, which show an annual increase of 3.1% [101] [102], are notably elevated in developed regions [103] [104] [102] like North America, Australia, New Zealand, and Europe. This trend is attributed to lifestyle factors such as obesity and declining parity [103] [102].
Mortality rates, instead, vary based on geographical location, with Asian countries demonstrating lower rates compared to the United States and Europe [105] [106] [103] [102].
Finally, socioeconomic factors, coupled with racial disparities [105] [102], exert a considerable influence, accentuating elevated mortality rates among women facing economic hardships. This is predominantly attributed to delayed diagnoses and suboptimal treatments [105] [103] [102].

Breast cancer is molecularly classified based on the immunohistochemical expression of hormone receptors, Progesterone (PR) and Estrogen receptor (ER), and the HER2 receptor. These markers play a crucial role in the management and treatment strategies for breast cancer [107] [108].
This approach defines four primary classes: *(i)* Luminal A (LA), *(ii)* Luminal B (LB), *(iii)* HER2-enriched (HER2+), and *(iv)* Triple Negative (TN) [107] [108], offering a detailed insight into the intricate molecular landscape of the disease (*Figure* 2.2).

The Luminal A subtype (*Figure* 2.2, *upper panel*), constituting nearly 50% of cases, is defined by the presence of ER and/or PR, absence of HER2, and low Ki-67 expression (<20%) [109] [110] [108].
Clinically, these tumors exhibit a low-grade, slow growth pattern, indica-

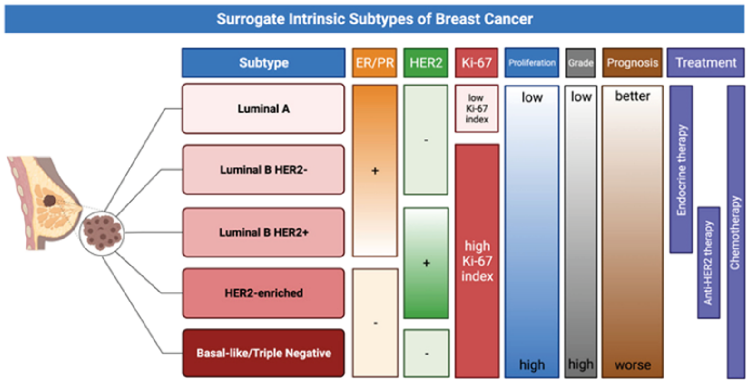


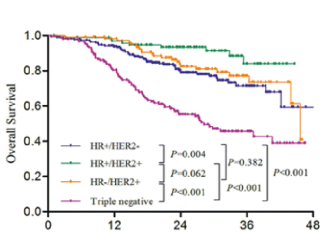
Figure 2.2. Breast Cancer (BC) molecular classification and clinical implications. The figure illustrates molecular subtypes of Breast Cancer (BC) characterized by molecular features (ER/PR status in orange, HER2 status in green and Ki67 status in red). Additionally, indications on proliferation potential (blue) and tumor grade (grey) are reported. These subtypes carry important implications for prognosis (brown bar) and the selection of treatment strategies (right bars, in violet). *Figure* from [2]

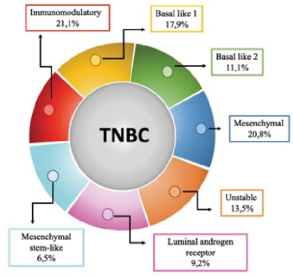
tive of the most favorable prognosis [111] [112]. They demonstrate a robust response to hormone therapy, particularly with agents like Tamoxifen or aromatase inhibitors, while minimal chemotherapy benefits are observed [113] [114]. The rare relapses are more commonly seen in bones, with lower rates in visceral and CNS sites and, remarkably cases of recurrence exhibit extended survival [115].

In contrast, Luminal B tumors, comprising 10–20% of cases, present a higher histologic grade and a less favorable prognosis compared to Luminal A [109] [116] [110] [114]. Their molecular profile is characterized by ER positivity, often PR negativity, HER2 negativity, and elevated Ki67 (>20%), indicating a faster growth rate [117] [108]. Despite the application of hormonal therapy combined with chemotherapy, these tumors pose a higher risk of visceral recurrence, leading to lower survival rates from diagnosis to relapse [117] [108].

Notably, both Luminal A and Luminal B subtypes frequently (40%) harbor PIK3CA mutations, resulting in the hyperactivation of the Phosphatidylinositol-3 kinase (PI3K) signaling pathway [118] [119].

The HER2-positive breast cancer, *Figure* 2.2 *mid panel*, constitutes 10–15% of diagnoses, exhibiting high HER2 expression with absence of ER and PR [109] [110] [120] [108]. This subgroup primarily metastasizes to bones, showing a higher incidence of visceral relapses compared to the luminal group.
It could be further categorized as luminal HER2 and HER2-enriched, which differ in hormonal receptor expression and Ki-67 rates. Luminal HER2 exhibits lower Ki-67 (15–30%) with hormonal receptors, while HER2-enriched has higher Ki-67 (>30%) without hormonal receptors [120] [108].
Despite heightened growth and aggressiveness compared to luminal subtypes, prognosis has improved with HER2-targeted therapies like Trastuzumab and Pertuzumab [121] [117] [122].
The most frequently identified genetic variants in HER2-positive breast cancer were found in the TP53, VHL and ATM genes [123] [122].
Triple-negative breast cancer (TNBC), accounting for 20% of breast cancers, lacks ER, PR, and HER2 expression[13] [14] [22] [15]. It is often associated with BRCA1/2 mutations (80%), along with TP53 (75%–80%) and PIK3CA (16-18%) mutations [124] [123] [51]. This results in DNA repair gene alterations, increased genomic instability, and p53 signaling loss [124] [125] [51]. While TP53 mutation alone did not predict a poor prognosis, TNBCs with PIK3CA mutation exhibited significantly worse outcomes [124] [22] [51]. Histologically, TNBC is poorly differentiated, highly proliferative, aggressive, prone to early relapse, and often diagnosed at an advanced stage [126] [117] [127].
The lack of identifiable biomarkers has impeded the incorporation of targeted therapies into the standard care for this type of breast cancer.
Currently, the standard approach relies on conventional chemotherapy, typically a combination of anthracyclines (DNA intercalators) and taxanes (mitotic spindle inhibitors). Unfortunately, only a minority, approximately 20% of patients, show positive responses to standard treatments [14]. Regrettably, a substantial proportion, ranging from 30% to 50% of patients, rapidly develops drug resistance [13] [14] [21], emphasizing TNBC as a compelling unmet clinical need (*Figure* 2.3, *left panel*).





Sub-types	Characteristics	Molecular Target
BL1	Proliferation drivers such as cell cycle, cell division and DNA replication	PARP1, RADS1, PLK1, TTK, CHEK1, AURKA/B
BL2	Growth factor and metabolic signaling with myoepithelial markers	EGFR, mTOR, MET, EPHA2
M	Epitelial-to-mesenchymal transition and differentiation	PI3K, mTOR, IGF1R, SRC, PDGFR, FGFR
UNS	DNA damage responses and cell proliferation	PARP1, RADS1, PLK1, TTK, CHEK1, AURKA/B
LAR	Hormonale-mediated signaling-androgen receptor	AR, Hsp90, PI3K, FGFR4
MSL	Epitelial-to-mesenchymal transition, differentiation, angiogenesis, stemness, growth factor	SRC, PI3K, MEK1/2, mTOR, PDGFR, NFkB, FGFR, IGFR, TGFBRII
IM	Immunemediated signaling	JAK ½, LYN, STATs, IRF1/7/8, BTK, NFkB

Figure 2.3. Triple Negative Breast Cancer survival rates and classification (left) Breast Cancer 2-years survival rates based on each molecular subtype defined by HR and HER2 statuses. *Figure* from [3] **(midddle)** Lehmann classification of TNBC which is categorized into six main subgroups: two basal-like classes (BL1 and BL2), an immunomodulatory (IM) class, a mesenchymal (M) class, a mesenchymal stem cell (MSL) class, and the luminal androgen receptor (LAR) class. **(right)** The table reports main molecular features of each TNBC subtype. *Figure* adapted from [4]

2.3.1 Triple Negative Breast Cancer subtypes: molecular characteristics and clinical implications

Triple Negative Breast Cancer (TNBC) is a heterogeneous neoplasm, predominant in premenopausal young women (under 40), especially African Americans [128] [129], which is further categorized into six subtypes (*Figure 2.3 middle panel*) including *(i-ii)* basal-like (BL1 and BL2), *(iii)* mesenchymal (MES), *(iv)* luminal androgen receptor positive (LAR), *(v)* immunomodulatory (IM) and *(vi)* unstable (UNS) each exhibiting distinctive clinical outcomes and pharmacological responses [22] [130] [131] [51] [15].

Constituting the majority (60-80%) of TNBC cases, Basal-like TNBCs are further categorized into two subtypes, BL1 and BL2 (*Figure* 2.3, *middle panel*), both exhibiting alterations in cellular proliferation regulation at distinct levels (*Figure* 2.3, *right panel*).
Specifically, the BL1 subtype exhibits amplification in proliferation genes like MYC, CDK6, and AKT2, while harboring deletions in DNA repair-related genes such as BRCA2, MDM2, RB1, and TP53 [51] [15]. This underscores the potential efficacy of PARP (poly (ADP-ribose) polymerase),

MYC or RB1 downstream effectors inhibitors and DNA intercalators, such as cisplatin, for which BL1 TNBCs show increased sensitivity [51] [15].
In contrast, BL2 types manifest AKT1 copy number gain and abnormal activation of proliferation-controlling signaling pathways, involving EGFR, MET, NGF, Wnt/-catenin, and IGF-1R pathways [14] [132] [131]. Therapeutic options for this subtype include mTOR, AKT1 and EGFR inhibitors, alongside antimitotic agents such as taxanes. Administering taxanes results in a notable four-fold increase in clinical remission rates for both BL1 and BL2 subtypes [13] [133] [15].
The immunomodulatory subtype (IM) shows overexpression of immune signaling genes with hyperactivation of the related pathways such as T cell receptor (TCR) signaling, interleukin (IL)-12 and IL-7 (*Figure* 2.3, *right panel*), so the most promising therapeutic approach relies on the administration of immune checkpoint inhibitors targeting PD1/PDL1 proteins [133] [127].
The peculiarity of LAR subtype, apart from its higher occurrence in Asian women and the overall survival rate (OS), is the hyperactivation of the Androgen Receptor (AR) pathway indicating a potential clinical benefit of this patients from the usage of anti-AR therapy [13] [15].
The mesenchymal group further divides into mesenchymal (M) and mesenchymal stem-like (MSL), each distinguished by their proliferation and differentiation status (*Figure* 2.3, *right panel*).
Specifically, the M subtype demonstrates enhanced activation of cell migration, extracellular matrix–receptor interaction, and differentiation pathways (Wnt and TGF-) [15]. This subtype is predisposed to chemotherapeutic resistance, prompting the consideration of mTOR inhibitors or drugs targeting epithelial–mesenchymal transition [134] [135].
In contrast, the MSL subtype exhibits diminished expression of cell proliferation-related genes and elevated levels of stemness-related genes. In this scenario, potential treatment options include PI3K inhibitors, Src antagonists, or antiangiogenic drugs [130] [15].

In conclusion, breast cancer features diverse subtypes with favorable treatment responses, unlike Triple-Negative Breast Cancer (TNBC), which displays notable heterogeneity. Various TNBC subtypes exhibit unique molecular profiles and treatment responses, underscoring the imperative for com-

prehensive investigations to refine therapeutic approaches for this intricate and challenging breast cancer subtype.

Chapter 3

Single-cell RNA sequencing

3.1 Introduction

This section provides an overview of the state-of-the-art in both single cell RNA sequencing with particular attention to standard workflow for data analysis advantages, disadvantages, and applications in research.
Section 3.2 reports a description of single-cell RNA sequencing experimental platforms, comparing it to bulk RNA sequencing, followed by a detailed overview of a standard workflow of scRNA-seq data analysis.
This encompasses *(i)* data pre-processing (paragraph 3.2.1), *(ii)* normalization and eventual batch effect correction (paragraph 3.2.2), *(iii)* feature selection and dimensionality reduction (paragraph 3.2.3), *(iv)* clustering (paragraph 3.2.4), *(v)* differential expression and enrichment analyses (paragraph 3.2.5) and, finally, *(vi)* pseudotime reordering (paragraph 3.2.6).

3.2 Exploring single-cell RNA sequencing: Platforms and data analysis

Single-cell RNA sequencing (scRNA-seq) has recently revolutionized the detection and quantification of messenger RNA (mRNA) at the individual cell level [136] [137] by assigning a unique cell barcode to each RNA molecule in a cell. In this way scRNA-seq allows for the exploration of individual cell transcriptional profiles and, compared to bulk RNA-seq, it

had facilitated the identification of rare cell subpopulations and enhanced cell type characterization [138] [139].

However, despite its potency in uncovering crucial cell types for developmental and disease insights, scRNA-Seq faces several challenges, including zero inflation due to both *(i)* limited mRNA in a single cell and *(ii)* technological constraints [140] [52] [141]. Additionally, it demands larger starting material than bulk RNA-seq, making it costlier and challenging for limited samples [140] [52] [141]. Finally, the substantial data volume presents storage issues, and its analysis complexity requires specialized software and bioinformatics expertise [140] [52] [141].

Single-cell RNA sequencing holds the advantage of isolating individual cells from a population using a variety of methods that have been continuously refined over time. Early approaches, such as Smart-seq2, were highly effective but limited to processing a small number of cells [142]. This limitation was addressed by microfluidic advancements, such as the Fluidigm C1 platform, which enabled the automated processing of 96 cells [143].
Recent advances in droplet microfluidics, including Drop-seq, inDrop, and 10X Chromium, have further expanded scRNA-seq throughput, making them the prevailing methods [144] [145] [5] [139]. Despite sharing the co-encapsulation technique, these platforms differ in terms of materials, bead manufacturing, protocols, and capture efficiency. Drop-seq stands out for its cost-effectiveness [5], while 10X Chromium excels in capture efficiency due to the use of soft hydrogel beads and synchronized encapsulation, a feature it shares with the InDrop platform [144] [145] [139].
Ultimately, platform selection depends on specific research needs, highlighting the dynamic nature of scRNA-seq methodologies.

In droplet microfluidics (*Figure* 2.1 *A*), the generation of nanoliter droplets involves the co-flow of oil and aqueous solutions, encapsulating single cells and barcoded beads [5]. This process, following a Poissonian distribution, results in numerous empty droplets and a small fraction containing multiple cells (doublets). The barcoded beads play a pivotal role in post-mRNA capture amplification on beads, known as STAMPs (Single-cell Transcriptomes Attached to Microparticles).

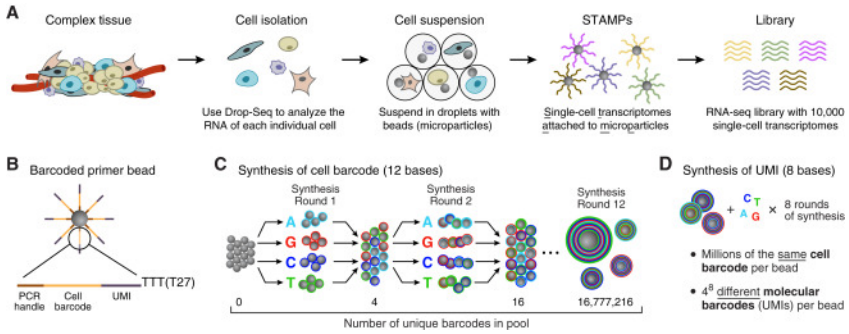


Figure 3.1. Microfluidics-based single-cell RNA-sequencing (scRNA-seq). (A) Overview of single-cell RNA-seq library preparation using Drop-seq. In microfluidic devices, cells and barcoded beads co-flow, entering aqueous droplets. Cell lysis, facilitated by the lysis buffer within the droplet, releases the transcriptome. Barcoded bead surfaces capture cell transcriptomes with primers. After emulsion breakage, uniquely barcoded STAMPs are recovered, representing the original cell's transcriptome. Subsequent bulk reverse transcription yields amplified STAMPs, prepared for next-generation sequencing. **(B)** Microparticle primers include a common PCR handle for amplification, with unique cell barcodes (BC) and molecular identifiers for mRNA counting (UMI), finally the oligo dT sequence captures mRNAs. **(C)** Cell barcodes are generated by splitting microparticle pools into oligonucleotide synthesis cycles, creating unique sequences reflecting synthesis paths. **(D)** UMIs are synthesized post split-and-pool cycles, with microparticles undergoing eight degenerate synthesis rounds. *Figure* from [5]

The beads carry a common PCR handle sequence, a unique molecular identifier (UMI, *Figure* 2.1 *D*), a cellular barcode (BC, *Figure* 2.1 *C*), and a 30 bp oligo dT sequence for mRNA capture [5].
The 12bp cellular barcode uniquely assigns each transcript to its cell of origin during bioinformatic analysis, facilitating efficient pooled library processing. UMIs offer distinctive tags to mRNA transcripts, ensuring accurate computational counting and preventing the double-counting of reads from the same mRNA transcript.
UMI length varies based on library preparation chemistry/platform, ranging from 8bp in Drop-seq [5] to 14 bp in the latest chemistry, v3, of the 10X Chromium platform [144].
Droplets serve as reaction chambers, facilitating cell lysis and poly(A)-

RNA capture on bead surfaces.
For subsequent Next Generation Sequencing (NGS), with Illumina technology predominant [146], captured mRNAs undergo reverse transcription, and Illumina adapters are incorporated via protocols like tagmentation [147] [148]. Resulting cDNA fragments, typically 200 to over 600 bp, undergo limited-cycle PCR to ensure complete adapter sequence incorporation [146]. Following ligation with adapters, these fragments adhere to a flow cell, initiating clonal amplification and generating clusters of identical copies. Sequencing employs the sequencing by synthesis (SBS) technique, utilizing a polymerase and chemically modified nucleotides.
The Illumina platform remains the primary choice for this NGS process. Sequencing throughput varies across Illumina sequencers (MiniSeq, MiSeq, NextSeq, NovaSeq, and HiSeq). Notably, the NovaSeq 6000 sequencer enhances performance, offering up to 500 Gb with a 2x150 bp kit for the S1 flow cell and up to 3,000 Gb with a 2x150 bp read length for the S4 flow cell [149].

Following sequencing, the data undergo analysis using specific bioinformatic workflows. The steps of these workflows are detailed in the subsequent paragraphs, adhering to the most recent best practices.

3.2.1 scRNA-seq data pre-processing: from sequencing files to expression matrix

In paired-end single cell sequencing, Read 1 (R1) usually provides cellular barcodes and UMIs, while Read 2 (R2) contains captured cDNA sequences.
After demultiplexing, associating reads with samples, low-quality barcodes and associated cDNA sequences are removed. Following quality control, 5' Illumina adapters and 3' poly-A tails are trimmed and Read 2 sequences align to the reference genome, with STAR aligner [150], optimized for scRNA-seq, i.e. in "solo" mode.
Finally, bead synthesis errors are identified and corrected, ensuring robust data processing in scRNA-seq. After pre-processing, single-cell transcriptomes are reconstructed into an expression matrix, where columns represent cells and rows correspond to genes. Quality control ensures analysis includes individual, undamaged cells, addressing challenges like low

sequencing depth, low detected gene count, high mitochondrial gene fraction, ambient RNA contamination, and doublet formation.
Tools like CellRanger [144] distinguish "true" cells from background barcodes by evaluating count depth distribution, identifying damaged cells with low gene counts, and dying cells with a high mitochondrial gene fraction.
SoupX [151] tackles cell-free RNA contamination, estimating fractions in potentially 'empty' droplets.
Doublets, detected by excessive genes and high-count depth, are addressed by DoubletFinder [152] by generating artificial doublets.
Quality control thresholds differ depending on tissue, protocols, and platforms, necessitating reference to similar experiments and analysis of various QC metric distributions.

Ultimately, the expression matrix may be influenced by read count variability arising from both technical and biological variations. To mitigate technical-derived variability, normalization and, occasionally, batch effect correction are frequently employed.

3.2.2 Takling scRNA-seq variability: normalization and batch effect correction

Traditional normalization methods, effective in bulk RNA-seq, are generally unsuitable for scRNA-seq due to zero inflation.
Dedicated scRNA-seq methods such as MAST [153] and Scran [154] address dropout events and scaling factor computation challenges to achieve accurate normalization.

Batch effect, a phenomenon where non-biological factors unrelated to biology lead to changes in the experimentally generated data, can be effectively removed in bulk RNA-seq [155] [156] [157]. However, in scRNA-seq, it poses a significant challenge due to intricate nonlinear effects.
Correction methods like ComBat [157], mnnCorrect [155], Harmony [158], CCA [159], LIGER [160], and scGen [161] aim to preserve biological signals while mitigating batch-specific biases.
Harmony offers speed for initial exploration, Seurat3/4-CCA demonstrates moderate batch mixing, and LIGER excels at batch mixing with poten-

tial impacts on cell type purity [162] [163]. Evaluation involves comparing clustering or visualization results and using metrics like kBET [163] to assess batch-effect correction effectiveness.
Biological effects like cell cycle, on the other hand, are corrected by scoring related biological features (e.g., cell cycle genes expression), followed by a simple linear regression against the calculated scores as implemented in Seurat [164] [159] [165] [166]. Caution is advised to avoid unintentional signal interference for specific analyses like cell differentiation or proliferation.
After normalization, values are typically pseudocounted, meaning that one count is added to avoid numerical issues in the subsequent log-transformation. Gene expression variance is stabilized by applying regularized negative binomial regression as in the case of SCTransform [167] [168].

At this stage, it is crucial to underscore the inherent biological information for characterizing cell heterogeneity or distinguishing between cell types/states in the data. This entails the need for simplifying high-dimensional data, given the abundance of profiled genes, through the application of feature selection and dimensionality reduction methods.

3.2.3 Improve characterization of cell heterogeneity: feature selection and dimensionality reduction

Feature selection simplifies data by strategically choosing around most biologically relevant genes from over 20,000 in the human genome. Typically unsupervised, it aims to identify genes with significant biological variability, using Highly Variable Genes (HVGs) to filter out technical noise. Tools like Seurat [164] [159] [165] [166] identify HVGs by assessing the coefficient of variation and expression means, prioritizing those with higher-than-expected zeros. While the theoretically selected gene number depends on sample complexity, studies suggest that downstream analyses can tolerate variations in the exact number of HVGs.

Following Highly Variable Gene (HVG) selection to emphasize relevant information, dimensionality reduction techniques project cells from a high-dimensional space to a lower-dimensional embedding while preserving cell-to-cell variability.

Commonly used methods for dimensionality reduction include principal component analysis (PCA) [169] [170] [171] [172], non-negative matrix factorization (NMF) [173], t-distributed stochastic neighbor embedding (t-SNE) [169] [171] [6] [174], and uniform manifold approximation and projection (UMAP) [169] [171] [175].
Principal Component Analysis (PCA) is a linear projection method that identifies dominant patterns, or principal components (PCs), by maximizing variance through linear combinations of original variables. PC1 captures the highest variance, and subsequent uncorrelated PCs, like PC2, account for the next largest variances [169] [170] [171] [172]. Additional PCs are often retained for downstream analyses such as clustering and trajectory inference and their number determined by dataset complexity.
However, PCA doesn't consider dropout events, leading to methods like Zero-Inflated Factor Analysis (ZIFA) [176], introducing an extra zero-inflation modulation layer to address dropouts. This enhanced projection capability, while valuable, comes at the expense of increased computational complexity.
Non-negative Matrix Factorization (NMF), analogous to PCA, is a linear projection method designed to decompose a nonnegative data matrix (V) with a positive rank (r) into basis (W) and coefficient (H) matrices. When applied to the expression matrix (V), NMF produces a basis matrix W of size (m, r), where m is the number of cells and r is the rank. The objective is to approximate V by expressing it as the product of nonnegative matrices, W and H, using the l2 norm (squared Euclidean distance) as an effective error function. NMF exhibits robustness to noise in scRNA-seq data, showing reliable performance in cell clustering [102] [177]. Moreover, the coefficient matrix derived from NMF aids in identifying genes specifically expressed in subgroups, offering an unbiased approach for selecting feature genes [102] [177].

For visualization, nonlinear methods like t-SNE and UMAP effectively preserve distances among cells in the original space offering improved scalability ideal for large-scale scRNA-seq. This method transforms high-dimensional Euclidean distances between data points into conditional probabilities, utilizing a Student's t distribution to calculate similarity in the low-dimensional space.

t-SNE faces runtime limitations with large datasets, a challenge overcome by the UMAP algorithm, recognised for its superior runtime performance and global structure retention. UMAP operates on three pivotal data assumptions: a uniform distribution on the Riemannian manifold, locally constant or approximable Riemannian metric, and local connectivity of the manifold. These assumptions enable modeling with fuzzy topology, determining the embedding by seeking the low-dimensional projection with the closest equivalent fuzzy topology. During model construction, UMAP begins by creating a weighted k-neighbor graph (kNN) through the nearest-neighbor descent algorithm [178], ultimately computing a low-dimensional representation that effectively preserves the desired characteristics of this graph, enhancing its suitability for large-scale datasets.

After simplifying the data, the identification of cellular subpopulations, one of the most potent capabilities of single-cell analysis, can be accomplished through the application of clustering methods.

3.2.4 Identifying cellular subpopulations: clustering methods

Clustering approaches aims at group data objects into clusters based on similarities or shared characteristics. In the context of scRNA-seq clustering must face multiple challenges like uncertainties in the number of clusters, variable cell types, and scalability issues.

Clustering methods, usually applied in scRNA-seq data analysis can be roughly classified into three groups: *(i)* K-means [179] [180] [181], *(ii)* hierarchical-based clustering [182] [183] and *(iii)* graph-clustering [181] [183].
K-means, an unsupervised machine learning algorithm, partitions data points into a predefined number of clusters (k) by iteratively assigning data points to the nearest cluster center based on Euclidean distances. The algorithm then updates the cluster centers by averaging the assigned data points, repeating the process until the cluster assignments stabilize or converge. Despite scalability for large datasets, it has limitations, potentially missing the global optimum and sensitivity to the choice of clusters (k). Its tendency to identify globular clusters and sensitivity to outliers

pose challenges, particularly in detecting rare cell types.
Applied in various tools like SC3 [184], SIMLR [185] and RaceID [186], k-means is adapted diversely.
SC3 integrates PCA/Laplacian graphs for improved performance on droplet-based datasets. SIMLR uses Gaussian kernel-based distance matrices with varied hyperparameters. RaceID employs k-means considering absolute transcript counts, utilizing a gap statistic for cluster determination, and includes an outlier detection step.

Hierarchical clustering encompasses agglomerative and divisive strategies [182] [183].
Agglomerative clustering merges cells progressively based on distance measures, while divisive clustering recursively splits clusters. This hierarchical approach aids in identifying rare cell types as small clusters. Unlike some methods, hierarchical clustering doesn't demand a predetermined cluster number or assumptions about data distributions, making it adaptable and widely incorporated in single-cell RNA-seq clustering algorithms at a high computational cost. It is implemented in CIDR [187], BackSPIN [188] and pcaReduce [189].
CIDR combines dimension reduction and hierarchical clustering, addressing dropout effects through implicit imputation.
BackSPIN employs divisive hierarchical clustering, thus with lower computational complexity respect to agglomerative counterpart, and sorting points into neighborhoods (SPIN) to cluster genes and cells simultaneously, while pcaReduce utilizes agglomerative hierarchical clustering with PCA, offering a hierarchical representation of clustering results.

Graph-based clustering has gained popularity in overcoming the limitations of k-means and hierarchical methods in large-scale datasets. This approach, applicable to single-cell RNA-seq data represented as graphs, involves nodes (cells) and edge weights (pairwise distances).
Common methods include the *(i)* clique algorithm [190] [191], *(ii)* spectral clustering [192], and *(iii)* the Louvain algorithm [193] [194].
The clique algorithm represents fully connected points in a graph, this defining a cluster.
SNN-Clip utilizes shared nearest neighbors for calculating cell similarity

to find quasi-cliques, thus reducing computational costs. Despite its ability to avoid manual cluster number specification, SNN-Clip suffers from scalability issues with resulting clusters lacking stability.
Spectral clustering, adapting to data distribution, relies on eigenvalues of the cell similarity matrix, but has high time complexity. It is implemented in SIMLR [185], tailored for scRNA-seq, which combines spectral clustering with kernel similarity measures, excelling in processing large-scale datasets with noise.
SinNLRR [195], instead, applies spectral clustering with a nonnegative, low-rank structure to detect cell types, though users must set the number of clusters.
The Louvain algorithm recursively merges communities and executes modularity clustering on condensed graphs. Its efficient time complexity sets it apart from other community-detection algorithms, making it a key component in popular tools like Scanpy (Python-based) [196] and Seurat (R-based).
In 2020, the Leiden algorithm enhanced Louvain clustering [197], overcoming limitations related to community division precision, node distribution density, and poorly connected communities. With faster and more accurate partitioning, Leiden provides explicit guarantees and bounds, and it has been successfully implemented in Seurat and Scanpy suites.

Identifying cellular subpopulations is crucial, requiring characterization involving distinctive biomarker identification and exploration of activated biological mechanisms. This step is essential for addressing questions in pathological and healthy conditions, employing differential expression and enrichment analysis.

3.2.5 Characterizing cellular subpopulations: differential expression and enrichment analysis

Differential Expression Analysis (DEA) [198] [199] examines variations in gene transcript expression levels across conditions, clusters, or cell states. The robustness of DE analysis relies in employing statistical methods that effectively handle the inherent variability among biological replicates [198]. Neglecting this variability leads to a plethora of false discoveries, significantly diminishing the robustness of scRNA-seq experiment results.

Seven statistical tests are usually used in DE analysis in scRNA-seq data, as demonstrated by their implementation in the dedicated function (FindMarkers) of Seurat package, including a t-test, Wilcoxon rank-sum test, logistic regression (LR), negative binomial (NB) and Poisson generalized linear models, a likelihood ratio test, and MAST's two-part hurdle model [167] [47] [165].
Two-tailored Student's t-test compares mean gene expression between the groups into consideration to discern treatment effects or group disparities in populations.
Wilcoxon rank-sum test, instead, is non-parametric thus it compares the ranks of genes in a gene set against those of genes outside of it.
Logistic regression constructs models predicting group membership for each gene while comparing it with a null model using a likelihood ratio test.
The negative binomial test relies on a distribution where the mean is less than the variance, making it apt for experiments with a modest number of replicates (typically 3-5).
Conversely, Poisson generalized linear models are favored in experiments with a higher number of replicates.
Likelihood ratio test compares the probability of given expression level for a gene in the target condition to the likelihood of the same expression level for a gene not in the target condition.
MAST employs a two-part generalized linear model, suitable for scRNA-seq data in complex experimental designs. It integrates an empirical Bayesian approach, enhancing inference for sparsely expressed genes, and utilizes the Cell Detection Rate (CDR) as a covariate for nuanced analyses [153].
Numerous benchmarking studies [200] [201] [202] [51] underscores substantial distinctions among statistical testes, showing the superior performance of the nonparametric Wilcoxon's rank-sum test compared to other methods. This holds particularly true for its notable advantages, such as low false positive rates, especially in extensive sample sizes.

In addition to dedicated scRNA-seq DE analysis tools, those adapted from bulk RNA-seq, involving pseudobulkerization (aggregating read counts

across biological replicates), commonly yield comparable results. Notable examples include DESeq2 [203], edgeR [204], and limma [205].
DESeq2 [203] utilizes the likelihood ratio test and the Wald test to assess gene model coefficients, incorporating shrinkage estimators and empirical Bayes priors for stable results across diverse data types.
edgeR [204] models count data using an overdispersed Poisson model, employing an empirical Bayes procedure to moderate overdispersion. Statistically, it employs likelihood ratio and quasi-likelihood F-tests for robust error rate control.
limma [205] employs linear models, adjusting for factors and utilizing parametric empirical Bayes for reliable inference, particularly in small-sample experiments. It can be used in two modes: limma-trend considers the mean-variance relationship at the gene level, while voom applies it at the level of individual observations.
Ultimately, tool selection hinges on the experimental context. Varied features in each tool provide diverse scenarios, emphasizing the need for methodological adaptation to specific study requirements.

Following the identification of biomarkers through differential expression analysis, the resulting gene list requires biological interpretation to unveil the underlying mechanisms. Gene Set Enrichment (GSE) analysis is a common approach, often employing Gene Set Enrichment Analysis (GSEA) [206] [207], which ranks genes based on expression changes and assesses the significant overrepresentation of gene sets or pathways. In the context of scRNA-seq analysis, various methods developed for bulk RNAseq, such as GSVA [208] and ssGSEA [209], are applicable.
Single sample gene set enrichment analysis (ssGSEA) and Gene Set Variation Analysis (GSVA) both extend the GSEA method by incorporating elements of its algorithm.
While ssGSEA employs the same gene expression ranking and Kolmogorov-Smirnov (K-S) like random walk statistic, it distinguishes itself by transforming gene expression into a Pathway Activity Score (PAS) profile for each cell-gene set pair, without requiring phenotype labels. On the other hand, GSVA retains the K-S like random walk statistic but estimates kernel-based cumulative density for each gene, using the classical maximum deviation method by default.

While conventional bulk RNA-seq tools can be adapted for scRNA-seq enrichment analysis, they often face challenges related to zero-inflation and cellular heterogeneity. Specialized tools such as Vision [210], Pagoda2 [157], AUCell [210] and scGSEA [211] are tailored for scRNA-seq, employing unique strategies to tackle these issues.

Vision [210], an annotation toolkit, leverages autocorrelation statistics to detect cellular biological variations. It identifies K-nearest neighbors (KNN) for each cell, computes pathway scores by averaging expressed genes, and corrects sample-level metrics using means and standard deviations.

In contrast, Pathway and Gene Set Overdispersion Analysis (Pagoda2 [157]) focuses on overdispersion in pathway activities to address cell heterogeneity in scRNA-seq data. It uses an error model for each cell, normalizes gene variance, and quantifies Pathway Activity Scores (PAS) through the first weighted principal component.

Area Under the ranked gene expression Curve (AUCell [210]) is a recovery-based method identifying cells with activity scores. This approach calculates PAS using the AUC score among all ranked genes in a particular cell, estimating the proportion of highly expressed genes in each gene set.

Finally, scGSEA [211] uses tf-idf normalized count matrix [212] [213] [214] which are decomposed through NMF linear projection method [173]. Then, GSEA is computed against the gene weights matrix (W) yielding matrix P were positive and significant scores (FDR < 0.05) were retained. Pathway activity in each cell was computed by the weighted sum of normalized enrichment scores across shared factors, achieved by the dot product of P and H (latent factors representing biological processes recurring throughout the set of sequenced cells) matrices.

Pagoda2 and Vision face challenges with the log-transform method for normalization, a concern shared by AUCell, which is notably affected by normalization methods. Despite these challenges, Pagoda2 exhibits higher stability with gene dropout and effectiveness in mitigating batch effects. However, new tools and advancements are required to enhance functional annotation in scRNA-seq, emphasizing the need for continued development in this field.

In evaluating cellular responses to stimuli from static snapshots, com-

prehending the dynamics of transcriptional states is essential. Pseudotime reordering methods are commonly employed for this purpose, but they encounter challenges in non-developmental processes. In this context, scGSEA has showed a strong capability to capture recurring patterns of pathway activity across numerous cells in diverse conditions, thus enhancing pseudotime reordering in time-based non-developmental processes such as drug response.

3.2.6 Reconstructing gene expression dynamics over time: pseudotime reordering

Pseudotime refers to the arrangement of cells along a trajectory, quantifying the relative activity or progression of the underlying biological process [215] [216] [217].

Pseudotime methods are essential in differentiation and time-based studies, constructing various trajectories such as linear, cycling, bifurcating, multifurcating, tree-shaped, or graph-based structures.
The classification of computational methods for pseudotime reordering is determined by the topology of the trajectories they create.
In this respect, Wanderlust [218] deals with linear trajectories, while ElPiGraph [219] and reCAT [220] are suitable for cyclical trajectories. Differences construction of trajectory topologies reflect in the type of algorithm they rely on. In this context, two are main algorithm applied: *(i)* minimum spanning tree (MST) and *(ii)* the graph-based ones.
The minimum spanning tree algorithm is utilized in tools like Monocle [221] [222], TSCAN [223] and Slingshot [224], making them suitable for complex developmental processes with branching points.
Monocle [221] [222] identifies pseudotime through the longest path in its first version, while the second version learns the cell trajectory using the MST algorithm and calculates pseudotime based on geodesic distances along the MST.
TSCAN [223] applies the MST algorithm on clusters, reducing computational workload, and Slingshot uses a simultaneous principal curves algorithm to create smooth curves from the MST during pseudotime calculation.
Graph-based methods, compute diffusion pseudotime (DPT) which com-

putes cell transition probabilities via random walks from a designated root cell, utilizing differences in these probabilities as pseudotime, thus accommodating loops or multiple separated trajectories, are applied in PAGA [**?**] and Monocle3 [221].
In this context, a weighted KNN approach is employed to construct the trajectory and compute pseudotime in the "diffusion map space". Partition-based graph abstraction (PAGA [225]) compresses and denoises data, creating a symmetrized KNN-like graph for community detection using the Louvain algorithm. Monocle3 utilizes the principal graph algorithm to generate the trajectory and assesses pseudotime as the geodesic distance from a user-selected root node.

Multiple benchmarking studies have evaluated the strengths and weaknesses of trajectory inference tools, recommending method selection based on anticipated trajectory topology [226] [183] and validate results with multiple trajectory inference (TI) methods to avoid bias and ensure robust predictions.
Given these limitations, particularly when studying non-developmental processes, emphasize the need for further improvements in this area.

Chapter 4

Lineage tracing

4.1 Introduction

This section provides an overview of the state-of-the-art in lineage tracing methods with particular attention to the advantages, disadvantages, and applications in research.
Section 4.2 describes lineage tracing principles and presents the current state of the art in related methods divided by experimental platforms in optical (paragraph 4.2.1), genetic (paragraph 4.2.2) and expression compatible (paragraph 4.2.3).

4.2 Lineage tracing: A state-of-the-art exploration of methods

Lineage tracing comprises a collection of methods enabling the tracking of individual cell fate and their descendants while minimally perturbing their physiological function during time-dependent processes, such as development or drug resistance [227] [228] [157].
This reconstruction spans both molecular and evolutionary perspectives, shedding light on the quantities and the molecular type of daughter cells generated throughout the process.
Lineage tracing encompasses continuous direct observation of cell divisions or cellular labeling for subsequent analysis. Figure 4.1 illustrates various experimental methods, showing advancements in optical, genetic,

and expression-compatible approaches.

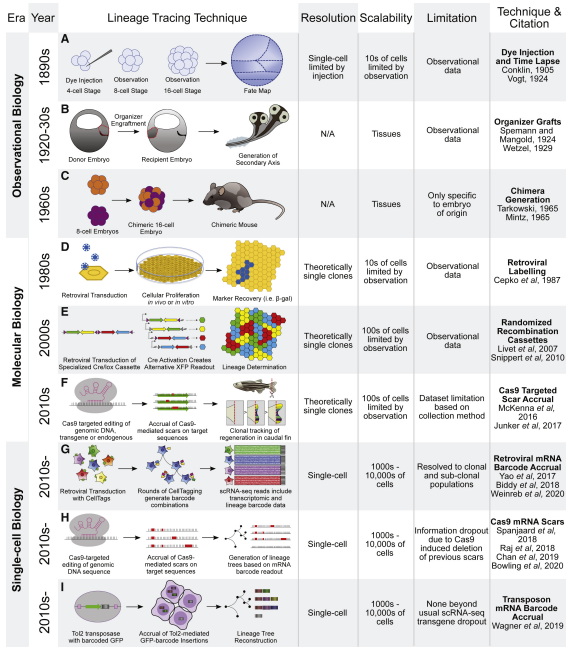


Figure 4.1. Overview of the lineage tracing methods. Visual representations illustrating **(A)** dye injection for fate mapping in the 1890s, **(B)** axis development and grafting, **(C)** creation of chimeric mouse embryos, **(D)** retroviral labeling of cells, **(A)** application of specialized Cre-loxP cassettes for clonal analysis, **(F)** accumulation of Cas9 scars in organisms, **(G)** viral barcoding methods for clonal and lineage analysis, **(H)** compatibility of Cas9 scars with scRNA-seq, and **(I)** barcode accrual through transposon-mediated processes. Figure from [6]

4.2.1 Optical lineage tracing: dyer and reporter genes barcoding

Since early fate mapping in the late 1800s, Figure 4.1*A,B*, lineage tracing primarily involves continuous observation of cell divisions or labeling cells for subsequent analysis, driven by advances in light microscopy and dye injection techniques [229] [230].

Notably applied in developmental studies on leeches and Styela partita (a tunicate model organism), these techniques allowed for monitoring progenitor cells and revealed the deterministic role of stereotypical cell divisions in fate determination [231]; [232]. Progress continued with time-lapse cinematography, culminating in the microscopy-based observation of live animals, exemplified by the complete lineage mapping of C. elegans development [233].
While these findings were pivotal, they were based on observations of transparent model organism embryos with limited cell numbers and invariant lineages. However, opaque embryos required additional techniques like dyeing or radiolabeling for fate mapping and lineage tracing, extending their application to mice and zebrafish in the 1920s [234]. Tracer dye injection in zebrafish revealed that the position of cells at the onset of gastrulation determines their future cell type [235]. While dye injection allows clonal analysis in wildtype organisms, it has limitations in time span and control over labeled cells.

To overcome these constraints, advancements in genetic manipulations, coupled with the advent of gene cloning technologies, have enabled the integration of reporter transgenes using viral transduction [236] [237]. This method, employing adenoviruses, adeno-associated viruses, and lentiviruses, is considered more practical and less cytotoxic than conventional transfection approaches. However, viral transduction faces challenges, including factors influencing its efficacy, such as the antiviral response [227].
Reporter genes encompass biochemical and fluorescent variants (Figure 4.1 *D*), with β-galactosidase as a prominent biochemical reporter. Its initial application in the late 80s involved transducing developing rat retinas with lentivirus to introduce β-galactosidase [238].
While aiding clonal analysis by inferring lineages from individual cells, these methods face limitations in constructing a comprehensive lineage hierarchy. Although capable of identifying progeny, they encounter challenges in resolving ancestor-progeny relationships crucial for a detailed lineage tree. Advances in fluorescent reporter usage have addressed this limitation.

Fluorescent reporters, particularly Green Fluorescent Protein (GFP), facil-

itate the direct observation of dividing cells and are widely recognized and utilized [157]. Its application has evolved with advancements in molecular biology, especially with the introduction of recombinase enzymes like Cre and FLP (Figure 4.1*E*). These enzymes facilitate controlled recombination between strategically placed loxP and FRT DNA target sites, resulting in predictable inversions or deletions. Inducing this process excises stop codons upstream of GFP, permanently marking the cell population. The precision is further enhanced by tissue-specific promoters regulating enzyme expression, ensuring accurate labeling in a minor proportion of cells, specifically target cells.

This approach has proven particularly valuable for studying adult tissue stem cells in organs like the intestine [239] or skin [240], significantly advancing our understanding of tissues development and maintenance.

Despite advancements, recombination-based lineage tracing encounters challenges in cell labeling frequency, impeding an accurate representation of overall population behavior due to complexities in discerning shared ancestry.

To address these challenges, "mosaic" labeling strategies, such as Mosaic Analysis with Double Markers (MADM) [241], integrate cell labeling and gene knockout through Cre-mediated interchromosomal recombination. This method employs two fluorescent markers (e.g., GFP and RFP) for wildtype and mutant cells, expanding the population captured. This enhances multiplexing, as in the case of multicolor lineage tracing techniques like "Brainbow," [242] which employs various spectral fluorophores, providing up to 90 distinct fluorescent signatures through stochastic Cre-mediated recombination that assigns unique colors to cells originating from a common ancestor.

A related method, Confetti [237], produced four distinguishable fluorescent signatures. In mouse models, it unveiled symmetric divisions of sparsely labeled intestinal stem cells, resulting in an expanded stem cell population. Long-term lineage tracing demonstrated neutral drift towards clonality in intestinal crypts, ultimately leading to one clone dominating the entire crypt. For instance, Split-Cre [243] utilizes two inactive fragments, each controlled by a distinct promoter. Here, recombination is coupled with the expression of multiple genes by simultaneously activating both promoters within a cell, resulting in Cre enzyme reconstitution and driving recombi-

nation.
Beyond reporter genes, genetic mosaicism studies offer an alternative for lineage tracing. Originating with Drosophila simulans gynandromorphs, a mosaic of male and female cells derived from separate cleavage nuclei, its application helped explore an alternative Drosophila mosaics model via mitotic recombination, revealing compartments that emphasize lineage segregation and influence functional subdivisions [244].
Mosaic organisms progressed to mammals, in the 1960s, with chimeric mouse embryos (Figure 4.1*C*), revealing the clonal origins of melanocytes, effects of mosaic genetic hermaphroditism, and developmental consequences of lethal mutations [6]. Chimerism advanced to interspecies aggregations using chick and quail embryos, enabling the tracking of neural crest cell migration, plasticity, and fate through phenotypic differences [245].

Observational lineage tracing employs phenotypic labels like morphology, dye marking, and genetic recombination for continuous cell observation. While informative, these methods may lack the precision necessary for constructing lineage hierarchies in specific organisms. Genetic recombination, induced by agents like Tamoxifen, can trigger apoptosis and disrupt tissue balance, posing challenges [6] [157]. Methods relying on a single gene for labeling may also lack specificity for the target cell population, limiting their effectiveness.

4.2.2 Genetic lineage tracing: Polylox and CRISPR/Cas9 barcoding and mutational patterns

As discussed earlier, optical techniques face challenges due to their limited color palettes, hindering the construction of comprehensive lineages and fate maps.

Addressing this limitation, methods like Polylox, DNA barcoding and CRISPR/Cas9-based scar generation have been introduced, generating DNA fingerprints in each cell but sacrificing imaging information, Figure 4.1*F*.
Polylox, a genetic DNA barcoding technique, utilizes random Cre-LoxP-mediated recombination events to generate approximately 1.9 million unique genetic barcodes [246]. This approach minimizes the risk of "label colli-

sion," where unrelated cells might acquire the same color combination, a common issue in optical techniques. Its improved robustness respect to optical methods was demonstrated in hematopoietic stem cell clones [246]. Polylox, with its increased label diversity, improves lineage tracing resolution, yet challenges persist. Silencing of reporter expression, evident in retrovirus-mediated labeling, hinders genuine lineage relationships. This could be overcome by integrating genetic labeling into non-functional genome, but it could induce differential labeling, inefficient capture, apoptosis-induced cell loss, and lineage termination. Despite confounders, genomic technologies uniquely potentiate lineage tracing, demanding careful consideration of associated factors for accurate interpretation and application [247] [248] [249].
Genomic technologies are pivotal for lineage tracing, especially in manipulable model organisms. In human samples, challenging direct manipulations necessitate lineage reconstruction based on naturally occurring somatic mutations, arising from DNA polymerase errors during cell division, analyzed through sequencing both at bulk and single-cell level, to retrospectively identify cells originating from a common ancestor [250] [228] unlike other genomic tracers. This flexible approach, unconstrained by specific labeling times, gains significance, notably in tumors with extensive genomic aberrations.

One challenge in this method is the high sequencing error rate, particularly for rare somatic mutations. A solution involves using copy-number variants (CNVs), microsatellites (MSs), and retrotransposition due to their detectability with low genome coverage [251] [252].
CNVs, abundant in cancer cells, have been employed to reconstruct breast cancer lineage trees. Single-cell whole-genome sequencing integrates CNV profiles with spatial information for cancer lineage tree reconstruction. Well-defined mutation sites in MSs have long delineated lineage trees. LINE1 retrotransposition serves as a lineage tracer in brain development [253].
These markers address challenges in exploring tumorigenesis and organ-specific development. However, the broad application is hindered by the high cost of sequencing of large numbers of single cells which can be solved by using targeted sequencing of mutation hotspots with high depth [251]

[252], DNA hydroxymethylation [254] and mitochondrial DNA (mtDNA) [255].

The limited occurrence of somatic mutations, particularly in healthy tissues with high turnover, hinders the resolution of somatic mutations-based clonal tree reconstruction. To overcome this limitation, innovative methods such as the CRISPR-Cas9 system have emerged Figure 4.1*H*.
This system introduces random indel mutations generated from DNA repair following Cas9-induced double-strand breaks (DSBs), uniquely labeling cells. Genetic barcodes within a multi-copy transgenic reporter, targeted by sgRNA, accumulate edits over time through Cas9 expression, facilitating lineage reconstruction by genomic DNA sequencing [256]; [257].
Early CRISPR-Cas9 methods faced challenges of editing saturation and potential false-positive lineage issues, especially in rapidly developing organisms. Mammalian lineage tracing lacked label diversity for prolonged periods, prompting self-targeting approaches using homing guide RNA (hgRNA) to create a genetic barcode, although with limited single-cell resolution. Recent genomic advancements have improved in-vivo modifications Figure 4.1 *I*, incorporating transposases like PiggyBac [258], often in conjunction with CRISPR-Cas9, or independently, as seen in the Sleeping Beauty transposon system [259].

The mentioned methods frequently depended on DNA-based sequencing, lacking the capability to capture the transcriptome and provide detailed information about cell identity and state. Consequently, there is a need for approaches compatible with single-cell RNA sequencing (scRNA-seq) to address these limitations.
This solution is given by DNA barcoding discussed in the next section.

4.2.3 Expression-based lineage tracing: DNA barcoding

DNA barcoding techniques have evolved for compatibility with scRNA-seq, expressing barcodes as RNA transcripts within the 3'UTR of a transgene, allowing simultaneous capture with the transcriptome Figure4.1 *G*.
The length of barcodes, measured in base pairs (bp), is directly linked to the number of unique labeling combinations. Therefore, the choice of library is influenced by this correlation.

Retro/lentiviruses secure the inheritance of DNA barcodes in progeny cells, aiding lineage relationship identification across numerous cells, with NGS revealing barcode-cell associations. Applied to hematopoietic cells, this approach unveils diverse clonal differentiation patterns in vivo and studies cancer cells, uncovering heterogeneity and clonal evolution in various cancer subtypes.
Lentiviral transduction introduces unique barcodes into a cell's genome, facilitating clonal analysis but is confined to identifying populations from a single cell.
To address this limitation, modified barcodes known as "CellTags" [260] enable labeling in successive rounds, providing a relative timescale for lineage tree construction. Lentiviral barcoding proves valuable in in vitro cell culture or regenerative systems, allowing clonal resampling and linking early cell states to eventual fates. This method, exemplified by CellTagging [260], reveals insights into lineage conversion trajectories and gene regulatory factors in reprogramming and hematopoiesis. Integrating additional information, such as chromatin accessibility, may unveil hidden heritable properties, offering deeper mechanistic insight into development and reprogramming.
CRISPR-Cas9 lineage tracing methods have evolved for enhanced compatibility with scRNA-seq [261] [256] [262].
Notably, the GESTALT (genome editing of synthetic target arrays for lineage tracing) method stands out in this regard [261]. Utilizing CRISPR-Cas9 editable cassettes integrated into genomes, GESTALT facilitates lineage tracing in intact, developing organisms like zebrafish models. The method has illustrated that the majority of cells in adult fish can be traced back to a limited number of embryonic progenitors. In the study of neural development, GESTALT has identified migratory neural progenitor populations, shedding light on their greater potency compared to previous understanding. However, limitations, such as large deletions erasing lineage records, have prompted the development of alternative approaches like ScarTrace [263] and LINNAEUS ([256]), which distribute Cas9 targets across the genome to mitigate such effects.

Chapter 5

Single cell lineage tracing reveals clonal dynamics of anti-EGFR therapy resistance in Triple Negative Breast Cancer

5.1 Introduction

While EGFR-activating mutations and amplifications (≥ 5 copies) are relatively rare in Triple-Negative Breast Cancer (TNBC) [22] [23] [24] [15], a majority of primary TNBCs exhibit increased expression of EGFR due to an increased gene copy number (three to four copies) [25] [26] [27] [24] [17] [28] (Figure 5.1 *A*-*G*), representing a potential vulnerability for TNBC patients [29].
However, EGFR-targeted therapies have demonstrated variable and unpredictable clinical responses in TNBC [34] [35] [36] [37] [38].
On the other hand, genomic variants found to be predictive of anti-EGFR resistance are relatively infrequent in TNBC patients [34] [23] [40] (Figure 5.1*H*), suggesting that non-genomic mechanisms of resistance must be at play in TNBCs.

Therefore, the lack of predictive biomarkers for response to anti-EGFR therapies in TNBC has hampered the translation of EGFR inhibitors in breast cancer.

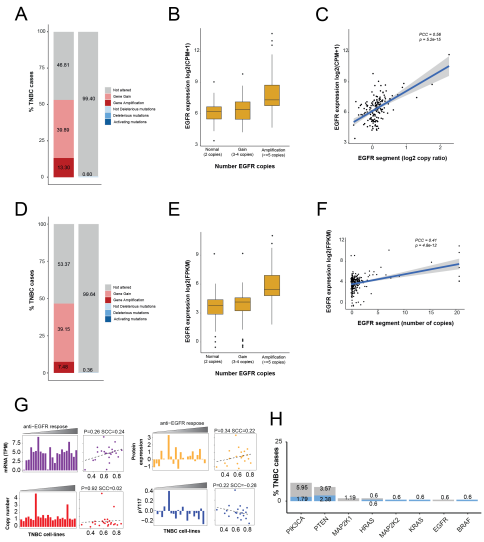


Figure 5.1. EGFR genetic alterations in patients' cohorts. (A) Percentage of TNBC patients from TCGA cohort with CNVs (left bar) or mutations (right bar) in EGFR gene. CNVs were divided in gain (3-4 copies, pink) or amplification (≥ 5 copies, red), while mutations were separated in activating (dark blue), deleterious (light blue) and not deleterious ones (cyan). **(B,C)** Distribution **(B)** and correlation **(C)** of EGFR gene expression relative to its copy number status in TNBC patients within the TCGA cohort. **(D)** EGFR genetic alterations at the copy number and mutational level (Asian cohort of patients). **(E, F)** EGFR gene expression distribution **(E)** and correlation **(F)** as a function of its copy number status in TNBC patients from Asian cohort. **(G)** Correlation between efficacy of anti-EGFR and EGFR status (mRNA, protein expression, CNV and EGFR phosphorylation status) in TNBC cell-lines. Data were downloaded from GDSC [7] portal and anti-EGFR efficacy was computed as average IC_{50} of all drugs targeting EGFR. **(H)** Percentage of TNBC patients harboring both deleterious (blue) and not deleterious (grey) mutations in known anti-EGFRs resistance genes reported on x-axis. Details on data analysis of patients cohort are reported in 6.2.8

To reveal these underlying mechanisms and identify drug resistance

signatures to potentially improve patients' stratification, we developed an experimental and computational platform to enable and analyze single cell lineage tracing in a human EGFR-dependent TNBC cell line (Section 5.2). Next, as described in Section 5.3, we took advantage of retrospective lineage tracing method to identify a pre-existing cellular population, characterized by high expression of IGFBP2 gene which involvement in drug resistance have been widely validated with orthogonal approaches.
The peculiar features identified thank to the retrospective lineage tracing have been employed to develop a contrastive learning approach to automatically predict Afatinib response in model cell lines, which is describef in Section 5.4.
Finally, in Section 5.5, prospective lineage tracing was applied to reconstruct the molecular dynamics of Afatinib resistance in MDA-MB-468.

5.2 Single-cell lineage tracing in EGFR-dependent Triple Negative Breast Cancer (TNBC)

We integrated lineage tracing with single-cell transcriptomics to discover novel predictive biomarkers of resistance to EGFR inhibitors in MDA-MB-468 cell line treated with Afatinib.

MDA-MB-468 is a human TNBC cell line characterized by wild-type EGFR gene amplification, which is highly sensitive to anti-EGFR therapies [41] [42] including Afatinib (Figure 5.2*B*). Additionally, it lacks mutations in anti-EGFR resistance-associated genes such as KRAS and HRAS [39].
Afatinib is a highly selective second-generation EGFR TKI that covalently binds to the intracellular tyrosine kinase domain of both EGFR and HER2 receptors, impeding the activation of EGFR signaling pathway [264] [265]. It exhibited substantial antitumor efficacy in a preclinical study assessing the effectiveness of six EGFR inhibitors across 14 distinct TNBC cell lines [266]. Its effectiveness was further demonstrated in TNBC PDX models, where it showed the highest synergistic cytotoxicity, among 1,363 drugs screened, when combined with YM155 (BIRC5 inhibitor) [267]. Finally, it produced a modest effect in TNBC patients when co-administered with neoadjuvant paclitaxel [265].

To allow single-cell lineage tracing, each cell in the initial MDA-MB-468 population was individually labeled through transduction with Cellecta's CloneTracker XP 10M Barcode-3' Libraries [268] at a low Multiplicity Of Infection (MOI, 0.05) to avoid for multiple infections into a single cell as described in 6.1.1 (Figure 5.2*D*). Cellecta library is a commercially available lentiviral barcodes library [268] in which each virion carries the plasmid depicted in Figure 5.2*C*, containing Venus, a red fluorescent reporter gene for infected cell detection, and a Puromycin resistance gene for their corrected selection. Importantly, the Venus gene incorporates a 48-nucleotide barcode cassette, referred as lineage barcode, positioned approximately 70 nucleotides upstream of the polyA sequence in its 3'-UTR region. This strategic placement makes the library compatible with oligo-dT cDNA synthesis, a crucial step in RNA-sequencing library preparation. In practice, lentivirus integrates the plasmid into the DNA of infected cells, facilitating barcode transcription. The resulting RNA fragments, carrying the barcode, are captured at the 3' end using indexed oligo-dT primers, ensuring the presence of the lineage barcode in sequencing files. This enables its computational identification in both single-cell and bulk RNA sequencing data, as described in 6.2.1.

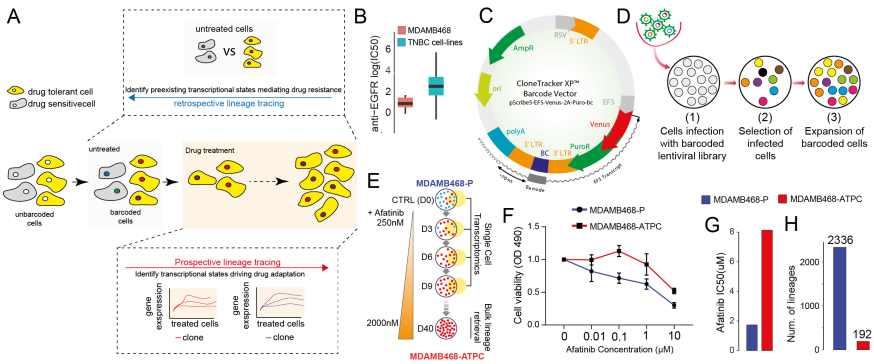


Figure 5.2. Platform for the identification of drug response biomarkers with single cell lineage tracing. **(A)** An overview of strategies of lineage tracing. Once cells are marked, they can be exposed to a selected drug concentration to drive the selection of resistant clones (yellow cells). In prospective analyses resistant clones (i.e. lineages) are followed over time to reconstruct driving mechanisms of drug adaptation. In retrospective analyses resistant lineages are retrospectively mapped on the untreated condition to reconstruct pre-existing mechanisms of drug resistance. **(B)** Average log IC_{50} distribution of MDA-MB-468 cells and of all the other available TNBC cell lines to anti-EGFRs. **(C)** The Cellecta's lentiviral construct. **(D)** Infection of MDA-MB-468 cells to enable lineage tracing. **(E)** Scheme of the time series experiment performed on the barcoded MDA-MB-468 cell line. D0 are untreated parental MDA-MB-468 cells (MDA-MB-468-P) a few hours before Afatinib addition, and D3 through D40 is the duration of the experiment where cells where subjected to incremental doses of Afatinib. Cells at D40 were considered as Afatinib tolerant persisted cells (ATPC). 10x Chromium sequencing was performed at D0, D3, D6 and D9 of 250 nM Afatinib treatment. Quantseq-Flex to retrieve Afatinib-tolerant lineages in bulk was performed at Day 40 with 2000 nM Afatinib treatment. **(F)** Dose-response curve showing cell viability of MDA-MB-468-P and MDA-MB-468-ATPC cells following treatment with the indicated concentrations of Afatinib. **(G)** Estimated IC_{50} for the dose response curve in **(F)**. **(H)** Number of unique lineages in the MDA-MB-468-P and MDA-MB-468-ATPC cell lines.

To gain a deeper understanding of clonal complexity of barcoded MDA-MB-468 cells, we improved lineage barcode detection by utilizing a custom bulk RNA-sequencing approach (see 6.1.1). Through this method, we identified 2,336 distinct clonal populations within the pool of barcoded MDA-

MB-468 cells (*Figure* 5.2*H*), representing groups of related cells descended from a common ancestor.

Next, we employed the lineage tracing experimental platform to investigate the mechanisms of drug resistance in TNBC. This exploration took into consideration its temporal dynamics via a time series experiment (refer to 6.1.2). We induced the selection of drug-tolerant cells in our cellular system by exposing barcoded MDA-MB-468 cells to escalating concentrations of Afatinib, ranging from 250 nM to 2000 nM. The chosen concentration range intentionally mimics a drug adaptation process, ranging from the minimal level at which a breast cancer cell line responds to an EGFR inhibitor from GDSC1 screening [7], as shown in *Figure* 5.3 *left*, to the highest concentration, which is twice the IC_{50} value of Afatinib in the MDA-MB-468 cell line (*Figure* 5.3 *right*).

Drug concentration was doubled every 9 days to allow cells to adapt to the drug treatment and sustain survival. This interval was chosen not only for its ability to facilitate cellular adjustment to the drug but also to ensure experiment reproducibility. Attempts to double the drug concentration at shorter intervals resulted in notable variability in cell viability and inconsistent outcomes across experiments.

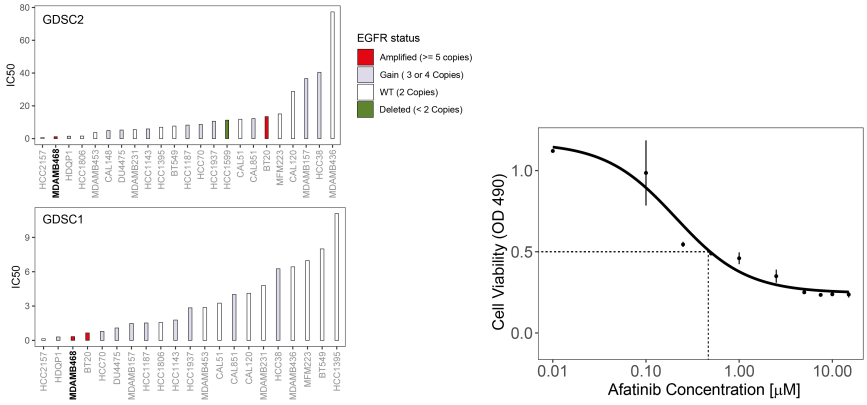


Figure 5.3. (left) Barplot showing IC_{50} values of anti-EGFRs (y-axis) in BC cell lines screened in GDSC2 (*top*) and GDSC1 (*bottom*) dataset and colored based on their CNV status (x-axis, ordered by median IC_{50}). **(right)** Dose-response curve in terms of cell viability (y-axis) after Afatinib treatment at the indicated concentrations (x-axis) on MDA-MB-468 cell line.

Given that the IC_{50} represents the drug concentration at which 50% of the treated cellular population dies, our decision ensures the identification of fully drug-tolerant cells. Indeed, after 40 days, we obtained MDA-MB-468 Afatinib Tolerant Persister Cells (MDA-MB-468 ATPC) which were able to grow stably in a medium with 2000 nM of Afatinib showing a 4-fold increase in Afatinib IC_{50} values with respect to the parental population (*Figure* 5.2 *F,G*).
Interestingly, this population comprises only the 8% of the initial clones (i.e., 192 out of the initial 2,336 ones), as we found from the retrieval of the lineage barcode using RNA bulk sequencing (*Figure* 5.2*E*, *lower part*). During the time-series experiment a subset of cells were collected and frozen in order to be later sequenced with single cell transcriptomics.

In summary, our time series experiment accurately recapitulates a typical drug adaptation process. The lineage tracing platform we developed, combining both experimental and computational approaches, facilitates the evaluation of the clonal composition of cancer cell populations.

5.3 Retrospective lineage tracing to elucidate pre-existing mechanisms of drug resistance

To unravel the dynamic non-genomic cellular states influencing the response to Afatinib, we systematically collected cells at three-day intervals for single-cell transcriptomics, providing close snapshots of their evolving gene expression patterns (refer to *Figure* 5.2*E*).

Specifically, we performed single-cell transcriptomics using the 10X Chromium platform (see 6.1.3) on 4,101 cells obtained at day 0, day 3, day 6, and day 9 post-Afatinib treatment (see *Figure* 5.4*A*). The Chromium 10X platform was chosen for its heightened sensitivity and depth, enabling the detection of both lowly expressed genes and those expressed in a small fraction of cells. This capability is particularly crucial for identifying rare cellular populations. Ultimately, the sequencing data underwent analysis following a standard workflow, as theoretically outlined in *Chapter* 3, and further detailed for the specific analysis in 6.2.2.

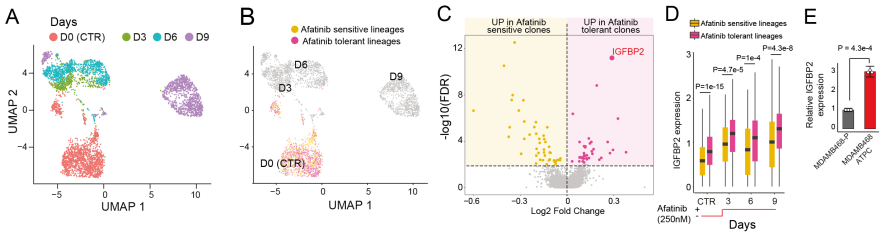


Figure 5.4. Retrospective lineage tracing analysis to reveal intrinsic drug resistance biomarkers. (A) UMAP representation of MDA-MB-468 cells colour-coded for the days of treatment. **(B)** Retrospective projection of the Afatinib tolerant persisted lineages on a UMAP representation of untreated MDA-MB-468 cells at Day 0. **(C)** Differential expression between Afatinib tolerant persisted lineages versus Afatinib sensitive lineages on untreated MDA-MB-468 cells (i.e. cells from Day 0). **(D)** Expression distribution of IGFBP2 in Afatinib tolerant persisted lineages and Afatinib sensitive lineages over time. Statistical differences were estimated using a two-tailed Wilcoxon test. **(E)** IGFBP2 expression in MDA-MB-468-P and MDA-MB-468-ATPC cells measured by real-time quantitative PCR. The statistical difference was estimated using a two tailed T-test.

The UMAP plot reported in *Figure* 5.4*A*, show that cellular populations within each time point are quite homogeneous, with extreme time points, i.e. control and day 9, with higher homogeneity respect to intermediate time points (day 3 and 6). This could indicate an initial drug-induced stress which is rapidly overcome at day 9 that show a different "stable" state, in line with the current knowledge of drug adaptation mechanism.
Next, the clonal composition of the cellular populations into each time point was evaluated by analyzing transcriptional barcodes in single cells, following the steps reported in 6.2.1, allowing us to capture the transcriptional state of 448 distinct clones at day 0, and 186, 206, and 228 distinct clones at day 3, day 6, and day 9, respectively.
To identify the rare drug-tolerant pre-existing cellular population, we conducted a retrospective analysis (i.e. retrospective lineage tracing). This involved mapping the barcodes of the 192 fully Afatinib-tolerant clones observed on day 40 onto the single-cell data of untreated MDA-MB-468 cells at day 0, as depicted in *Figure* 5.4*B*.

Subsequently, we explored the possibility of identifying genes with dis-

tinct expression patterns in untreated cells at day 0 which contribute to understanding the pre-existing mechanisms of resistance to Afatinib treatment. To this end, we compared cells originating from Afatinib-tolerant lineages with those from Afatinib-sensitive lineages, the latter being those disappeared during 40 days of continuous Afatinib exposure (see 6.2.2).
Remarkably, we observed differential expression in 374 genes between these two cell populations (refer to *Table* 6.1), with 221 genes exhibiting up-regulation in the Afatinib-tolerant lineages and 153 genes up-regulated in the Afatinib-sensitive lineages (see *Figure* 5.4*C*).

To unravel the mechanisms underlying Afatinib resistance associated with the genes in our signature, we conducted a Gene Ontology Enrichment Analysis (GOEA) on the 221 up-regulated genes in the Afatinib-tolerant cell population. This analysis unveiled involvement in biological processes associated with resistance to EGFR TKIs observed in other cancer types, such as oxidative phosphorylation (OXPHOS) and fatty acid metabolism [16] [269] [262] [270] (refer to *Table* 6.2).
These processes are implicated in reducing drug-induced stress by detoxifying secondary products derived from drug action on cellular components. Conversely, the Afatinib-sensitive cell population exhibited up-regulation of 153 genes, including EGFR and several genes associated with the estrogen signaling pathway (refer to *Table* 6.3).
This observation validates a diminished cellular reliance on these signaling pathways, potentially blocked by the treatment, serving as an initial indication of resistance.

To validate our approach, we concentrated on characterizing the most significantly upregulated gene within the untreated Afatinib-tolerant lineages. As shown in *Figure* 5.4*C*, it was Insulin-Like Growth Factor Binding Protein 2 (IGFBP2), recently linked to the progression and metastasis of various tumor types [271] [272].
This protein, a member of a six-gene family, exhibits high affinity for binding to Insulin-like Growth Factors I and II (IGF1 and IGF2), thereby amplifying the activation of Insulin-like Growth Factor Receptors (IGF1-R and IGF2-R). Notably, these receptors have been associated with resistance to EGFR TKIs in glioblastoma [273] [274] and NSCLC [275] [276].

Interestingly, the high expression of IGFBP2 observed in Afatinib-tolerant lineages, existing prior to drug treatment, persists throughout the treatment period, as illustrated in *Figure* 5.4*D*. This implies a potential involvement of this gene, among the others, in both the initiation and maintenance of drug resistance.
Indeed, the increased expression of the IGFBP2 gene in Afatinib-tolerant clones was subsequently validated through qRT-PCR in MDA-MB-468-ATPC (see 6.1.4) showing a 2-fold increase (*Figure* 5.4*E*).

Targeted drugs usually function by inhibiting a specific target situated upstream of a signaling pathway. Consequently, when cells undergo treatment and this target is suppressed, the entire pathway experiences blockade. Nevertheless, cells manage to overcome this blockade by activating compensatory pathways, effectively activating downstream components of the initially impeded one. In the case of EGFR-TKIs resistance, numerous studies have demonstrated the activation of IGF1-R, MET, and the PI3K-AKT-mTOR signaling signaling pathways [39] [277] [237] [278] [279].
This understanding, coupled with the knowledge that IGFBP2 is part of the IGF1-R pathway, prompted us to co-treat parental MDA-MB-468 cells with different concentrations of Afatinib and GSK-1904529A, a selective inhibitor of IGF1-R and Insulin Receptor (INSR). As shown in *Figure* 5.5 (see 6.1.5 for experimental details) co-treatment of parental MDA-MB-468 cells at different concentrations of Afatinib and GSK-1904529A had a strong synergistic effect on these cells.
The occurrence of synergism in drug combination studies is validated when the combined effect of two or more drugs surpasses the sum of their individual ones. This substantiates the effectiveness of simultaneously blocking two distinct pathways (i.e., EGFR and IGF1-R) in this experimental context.

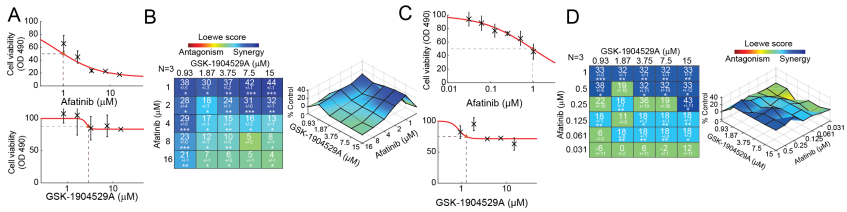


Figure 5.5. Compensatory activation of IGF1-R and INSR pathways to overcome EGFR inhibition. (A) Single agent dose-response data and fitted curve for Afatinib *(upper)* and GSK-1904529A *(lower)* on MDA-MB-468 parental cells. **(B)** *(left)*. Synergy scores calculated using Combenefit tool according to the Loewe additivity model (see 6.1.5) for different concentration of Afatinib and GSK-1904529A on MDA-MB-468 parental cells. The larger the value stronger is the synergism while negative values mean antagonism between the two drugs. The number below the synergy score is standard deviation. Asterisks below standard deviation indicate results that are statistically significant by the one-sample t-test. The degree of significance are as follows: * p < 5x10-2; ** p<10-3; *** p<10-4. The number of biological replicates (N) is indicated at the top left of the matrix. *(right)* Combination dose-response surface, expressed as a percentage of the control value. Overlain on the dose response surface are the Loewe synergy scores.**(C)** Same as **(A)** but where concentration of Afatinib vary from 31nM to 1μM. **(D)** Same of **(B)** but where concentration of Afatinib vary from 31nM to 1μM.

To gain a more thorough understanding of the dose-response relationship, we conducted two independent experiments in triplicate. One experiment involved varying Afatinib concentrations from 1 to 16 M (see *Figure* 5.5 *A,B*), while the other focused on lower Afatinib concentrations, ranging from 31 nM to 1 M (see *Figure* 5.5 *C,D*).
As depicted in *Figure* 5.5 *B*, the synergistic interaction between Afatinib and GSK-1904529A is noticeable across a broad spectrum of Afatinib concentrations, with particular potency observed around 1 M, which is approximately the IC_{50} value for MDA-MB-468 (*Figure* 5.3*B*).
These findings further validate the existence of a pre-existing subset of cells exhibiting heightened activity in the compensatory IGF1-R signaling pathway sustained by IGFBP2.

In conclusion, the integration of scRNA-seq and lineage tracing techniques enabled us to pinpoint a set of potential biomarkers for intrinsic drug re-

sistance, enhancing the stratification of patients. Initial characterization of the most up-regulated gene, IGFBP2, suggests promising prospect.

5.4 Contrastive learning to predict Afatinib response in TNBC cells.

Employing retrospective lineage tracing within our cellular model, we successfully delineated a unique signature encompassing 374 potential biomarkers linked to intrinsic drug resistance. This discovery prompted an exploration of the predictive potential of these genes as biomarkers for Afatinib therapy in other Triple Negative Breast Cancer cells.

To address this, we developed scASTRAL (single-cell Afatinib reSponse of TRiple negAtive ceLls), a novel deep doublet learning approach detailed in the 6.2.3.
Leveraging contrastive learning and Supported Vector Machine (SVM), scASTRAL predicts Afatinib sensitivity from single-cell data using the expression values of the 374 identified marker genes for Afatinib response shown in *Figure* 5.6.

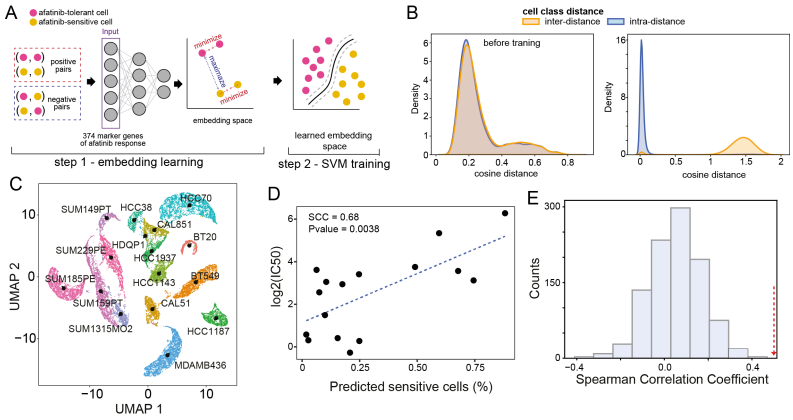


Figure 5.6. Single-cell Afatinib response prediction in TNBC cells. **(A)** Schematics of scASTRAL. Contrastive learning is used in the first step to build an embedding space where positive examples (two cells of the same class) are separated from negative examples (two cells of different classes). In the second step an SVM classifier is trained on the learned embedded space to predict the class (i.e. Afatinib-tolerant or Afatinib-sensitive) of novel cells. **(B)** Inter- and intra- cosine distance between sensitive and tolerant MDA-MB-468 cells before and after the training. Intra-distance is the cosine distance among cells of the same Afatinib response class, while inter-distance is the distance among cells belonging to two different Afatinib response classes **(C)** UMAP representation of 22,724 TN breast cancer cells from 16 cell lines. **(D)** Spearman Correlation Coefficient (SCC) between scASTRAL predicted sensitivity to Afatinib and experimentally estimated IC_{50}. **(E)** Histogram of SCC values computed between scASTRAL predicted sensitivity to Afatinib and experimentally estimated IC_{50} using 374 random genes. 1,000 simulations were performed. Red arrow indicates SCC value obtained using the 374 Afatinib response marker genes we identified with our retrospective lineage tracing approach.

Contrastive learning enhances low-dimensional data embeddings by bringing similar samples closer and pushing dissimilar ones farther apart [280] [281] [282].
In our specific context, it constructs an embedding space where positive examples (e.g., two cells with the same drug response label) are segregated from negative examples (e.g., two cells with different drug response labels). On the other hand, the SVM excels in binary classification, utilizing an op-

timal decision boundary, the hyperplane, to maximize the margin between data points of distinct classes., thus improving clarity in class differentiation [283] [284] [285].
As illustrated in *Figure* 5.6*A*, scASTRAL takes cells as input, representing their transcriptional status through the expression values of the 374 marker genes identified in the intrinsic Afatinib response.
The method involves two main steps. In the first step, scASTRAL constructs an embedding space, minimizing cosine distance within the same Afatinib-response class clones and maximizing distance between different clones (*Figure* 5.6*B*). This ensures robust and scale-independent comparisons, as cosine distance assesses similarity without considering object size. In the second step, scASTRAL employs an SVM classifier to distinguish between Afatinib-sensitive and Afatinib-tolerant cells within the learned space from the first step.

We trained scASTRAL using the drug response labeled 1,541 MDA-MB-468 control cells (see 6.2.3) and assessed its performance in predicting Afatinib sensitivity on 22,724 single-cell transcriptional profiles from 16 TNBC cell lines (refer to *Figure* 5.6*C*). Specifically, data for 11 of the 16 cell lines (comprising 15,022 cells) were sourced from the recently released breast cancer atlas [136], while the remaining 5 cell lines (SUM149PT, SUM185PE, SUM22PE, SUM159PT, and SUM1315MO2), totaling 7,702 cells, were de novo sequenced using the DROP-seq platform [5] (see 6.1.6 for experimental details and 6.2.4 for computational ones).
Subsequently, we utilized scASTRAL to project the 22,724 triple-negative cells into the acquired Afatinib-response space, classifying each cell as sensitive or resistant to Afatinib, following the procedure reported in 6.2.5.

To assess scASTRAL's predictive performance for Afatinib sensitivity, we translated single-cell predictions to the cell-line level. This involved computing the fraction of predicted Afatinib-sensitive cells in each cell line and correlating these values with experimentally estimated Afatinib IC_{50} values obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) database [7] or measured de novo when unavailable (refer to *Table* 6.4).
Figure 5.6 *D* illustrates that scASTRAL accurately predicted cell-line sensitivity values, exhibiting significant correlation (P=0.0038, SCC=0.68)

with the corresponding experimentally estimated IC_{50} values (see *Table* 6.5). To assess the predictive power of our 374-marker gene signature, we utilized Monte Carlo simulations which model the probability of various outcomes in processes influenced by random variables.
In each simulation, we randomly selected 347 genes, trained scASTRAL on MDA-MB-468 control cells, applied the model to predict Afatinib responses in the 16 TNBC cell lines mentioned earlier, and assessed performance by correlating predicted sensitivities with experimentally estimated IC_{50} values. Despite conducting 1,000 simulations, *Figure* 5.6*E* demonstrates that none yielded a set of 347 genes surpassing the predictive capability of those identified through our retrospective lineage tracing strategy.

In summary, these findings show the efficacy of our cell lineage tracing strategy, which retrospectively maps drug-tolerant clones onto untreated cells. Beyond facilitating the identification of predictive biomarker genes for drug response, this approach also exhibits promise in proficiently classifying cell lines, especially in situations where drug response data are unavailable.

5.5 Prospective lineage tracing to identify biomarkers of acquired drug resistance

In our earlier analyses (see *paragraph* 5.2), it was revealed that among the 2,336 clones (i.e., lineages) initially present in the parental MDA-MB-468 cellular population, only 192 eventually developed full tolerance to Afatinib treatment.

To understand the evolutionary trajectory of these tolerant clones over time, we computed, at each sequenced time point, the percentage of cells associated with each Afatinib-tolerant lineage.
This was done by estimating the number of cells linked to a specific lineage and dividing it by the total number of sequenced cells. The trends of these percentages over time are depicted in *Figure* 5.7*A* (and detailed in *Table* 6.6).

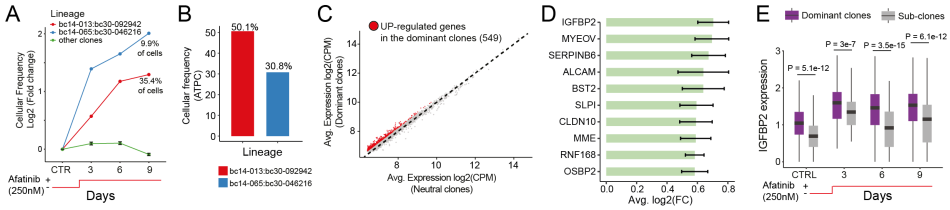


Figure 5.7. Prospective lineage tracing analysis to reveal acquired drug resistance biomarkers. (A) Cellular frequency of the Afatinib tolerant clones bc14-013:bc30-092942 and bc14-013:bc30-092942 and all other clones at the sequenced time points (CTRL, Day 0), Day 3, Day 6, Day 9). **(B)** Cellular frequency of clones bc14-013:bc30-092942 and bc14-013:bc30-092942 in the MDA-MB-468-ATPC cell line. **(C)** Up-regulated genes identified by comparing cells belonging to the dominant clones to cells belonging to the neutral clones. **(D)** Average log2 fold change (FC) of the top ten up-regulated genes from **(C)**. A 95% confident interval is reported. **(E)** Expression distribution of IGFBP2 in dominant and neutral clones over time. Statistical differences were estimated with a two-tailed Wilcoxon test.

This analysis revealed that 190 of the 192 lineages maintain a constant relative population size (i.e. the same percentage) over time, which we termed "neutral". The remaining 2 clones, however, showed significantly increased relative size frequency, constituting nearly half (45%) of the total population after nine days of Afatinib treatment (see *Figure* 5.7*A*), which we termed "dominant".
Subsequent analysis verified that these two clones emerged as the predominant ones, constituting 81% (see *Figure* 5.7*B*) of the drug-tolerant MDA-MB-468-ATPC cell line (i.e., cells after 40 days of Afatinib exposure).

To identify key genes driving the cellular expansion of these two dominant clones during the first nine days of Afatinib treatment, we categorized cells into two groups based on the behavior of their clone of origin, distinguishing them as either "dominant" or "neutral". Next, we applied a semi-supervised pseudotime reordering algorithm [286] to combine functional and time-dependent trajectories, providing a comprehensive depiction of the evolutionary pattern for both types of clones. Subsequently, we applied time-course differential expression analysis [287] to pinpoint

genes that displayed up-regulation specifically in the cells of the dominant clones (for additional details, refer to 6.2.6). This led to the discovery of a comprehensive set of 549 driver genes, all demonstrating significant up-regulation (FDR $< 5\%$) in the dominant clones (consult *Figure* 5.7*C* and *Table* 6.7 for details).
Particularly noteworthy is the consistent emergence of IGFBP2 as the most up-regulated gene (see *Figure* 5.7*D*). These findings, along with its constant upregulation through time points in dominant clones respect to neutral ones (*Figure* 5.7*E*) strongly suggest that IGFBP2 may be instrumental in both the initiation of resistance and the subsequent cellular expansion during Afatinib treatment.

To test for IGFBP2 implication in drug resistance, we generated two novel cell lines from the parental MDA-MB-468 cell line using CRISPR-based transcriptional modulation techniques to enhance/reduce the expression of IGFBP2 gene in physiological conditions (details in 6.1.7).
This comes in two variants: CRISPRi for repression and CRISPRa for activation. They both contains a nuclease deactivated Cas9 (dCas9), to the target genomic region thanks to a guide RNA (sgRNA), without generating a DSB, fused with a transcriptional repression domain, or a transcriptional activation one, respectively. In our case CRISPRa model was the MDA-MB-468-VPH-IGFBP2, stably overexpressing IGFBP2, and MDA-MB-468-KRAB-IGFBP2 stably downregulating IGFBP2 expression, as reported in *Figure* 5.8.

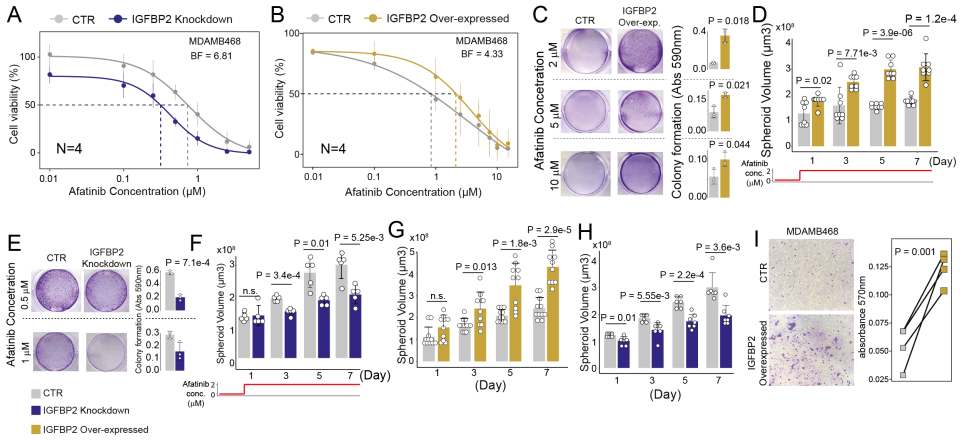


Figure 5.8. Identification of genes driving cellular expansion during Afatinib adaptation. **(A)** Dose-response curve in terms of cell viability following treatment of MDA-MB-468 control (CTR) and IGFBP2-knockdown cells with Afatinib at the indicated concentrations (see 6.2.7). A value of 6.81 corresponds to a very strong significant difference between the two dose-response curves [8]. **(B)** Dose-response curve in terms of cell viability of MDA-MB-468 control (CTR) and IGFBP2-overexpressing cells following treatment with Afatinib at the indicated concentrations. Significance was assessed as in **(A)**. A value of 4.3 corresponds to a strong significant difference between the two dose-response curves [8]. **(C)** Colony assay (representative images) for MDA-MB-468 control (CTR) and IGFBP2-overexpressing cells after 10 days of Afatinib exposure at 2, 5 and 10 μM *(left)*. Quantification is reported on the right. Experiments were performed in triplicate. **(D)** Spheroid volume growth of MDA-MB-468 control and IGFBP2-overexpressing cells with 2 μM Afatinib (see 6.1.9) over time. **(E)** Colony assay (representative images) for MDA-MB-468 control and IGFBP2-knockdown cells after 3 days of Afatinib exposure at 0.5 and 1 μM (left). Quantification is reported on the right. Experiments were performed in triplicate. **(F)** Spheroid volume growth of MDA-MB-468 control and IGFBP2-knockdown cells with 2 μM of Afatinib over time. **(G)** Spheroid volume growth over time of MDA-MB-468 IGFBP2-overexpressing cells. **(H)** Spheroid volume growth over time of MDA-MB-468 IGFBP2-knockdown cells. μ Transwell migration assay of MDA-MB-468 IGFBP2-overexpressing cells. Reported p values from panels C to I were estimated using two-tailed t-test.

As depicted in *Figure* 5.8*A*, exposing these cells to escalating concentrations of Afatinib revealed a substantial increase in cytotoxicity across

the concentration range (see 6.2.7), after IGFBP2 knockdown. This improvement is evident in *Figure* 5.8*A*, showing a notable reduction in the IC_{50} value for these cells.
The data additionally highlights almost complete cellular death in the population at 2 μM Afatinib, emphasizing the compromised capacity of these cells to endure adverse conditions. Conversely, IGFBP2 overexpression significantly decreased cytotoxicity, elevating both the IC_{50} value and the concentration at which the drug's cytotoxic effect was total.
Conversely, to demonstrate the potential involvement of IGFBP2 in cellular expansion under treatment, we conducted a cell colony growth assay on both cellular systems (see 6.1.8).
The system with IGFBP2 overexpression exhibited robust cellular growth even at high Afatinib dosages, ranging from 2 to 10 M, showing a significant enhancement compared to the control, as presented in *Figure* 5.8*B*. This behavior persisted in 3D cell culture (see 6.1.9), where IGFBP2 overexpression amplified MDA-MB-468 spheroid growth during a nine-day exposure to 2 μM Afatinib, mirroring the findings of our time series experiment (*Figure* 5.2*F*).
In contrast, IGFBP2 knockdown hindered colony formation after 5 days of exposure to 1 μM Afatinib (see *Figure* 5.8*C*) and significantly impeded MDA-MB-468 spheroid growth throughout the nine days of Afatinib treatment (see *Figure* 5.8*D*).

To assess whether IGFBP2 might serve as a potential generalized biomarker, among others, imparting resistance to Afatinib in other TNBC cell lines, namely HCC1806 and HDQ-P1, we conducted gene overexpression investigations.
The first cell line represents the most sensitive cell line to Afatinib, while the latter, despite harboring a heterozygous deletion of IGFBP2, stands as the second most sensitive. For the first cell line, a CRISPRa system was employed to generate cells stably overexpressing IGFBP2, whereas in the second cell line, plasmid transfection was utilized (6.1.10).
As shown in *Figure* 5.9, again overexpression decreased Afatinib cytotoxicity in both cell lines, thus indicating a possible involvement in Afatinib resistance for a broader range of TNBC cell lines.

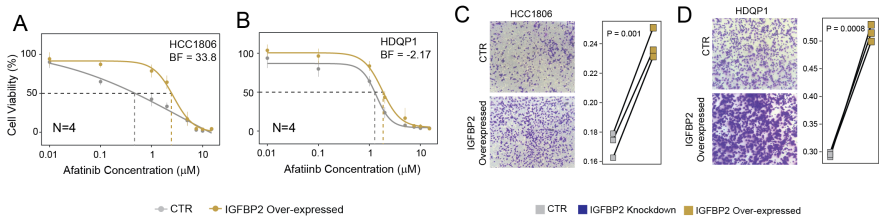


Figure 5.9. IGFBP2 implication in Afatinib resistance. (A) Dose-response curve in terms of cell viability following treatment with Afatinib at the indicated concentrations on HCC1806 IGFBP2 overexpressing cells and relative control cells. **(B)** Dose-response curve in terms of cell viability following treatment with Afatinib at the indicated concentrations on HDQ-P1 IGFBP2 overexpressing cells and relative control cells. **(C)** Transwell migration assay of HCC1806 IGFBP2 overexpressing cells. **(D)** Transwell migration assay of HDQ-P1 IGFBP2 overexpressing cells.

All together, these results demonstrate the capacity of our cell lineage tracing strategy to enable the reconstruction of cell expansion at single-clone resolution, providing insights into the dynamic nature of drug resistance mechanisms. Our findings also reinforce the role of IGFBP2 as one of the players influencing Afatinib adaptation in MDA-MB-468 cells.

5.6 Reconstruction of gene expression dynamics of acquired mechanisms of drug resistance

The signature related to acquired mechanisms that we identified allowed us to validate the involvement of IGFBP2 over time. However, it's crucial to recognize that other genes may play a role, and their activation might be induced in later stages of the drug response.

To pinpoint these potential targets and obtain a more comprehensive understanding of gene expression dynamics during Afatinib response, we analyzed the expression patterns of the 549 driver genes that were upregulated in drug-tolerant lineages along their reconstructed pseudotime trajectory (refer to 6.2.6).
After clustering analysis (see 6.2.6) four distinct patterns of transcriptional adaptation to Afatinib treatment were observed (*Figure* 5.10).

Cluster one encompassed genes with an "early" transcriptional response, showing a significant increase after three days of treatment, while cluster two comprised genes with a "delayed" transcriptional response, demonstrating a significant increase after six days of treatment. Cluster three included genes that consistently downregulated in expression during Afatinib treatment, whereas genes in cluster four exhibited a “waving” behavior with a transient decrease in expression followed by a return to baseline levels.

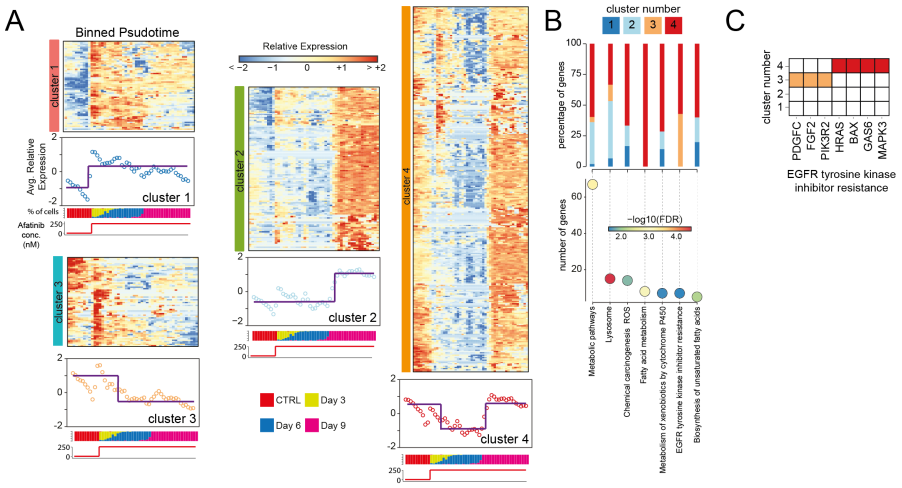


Figure 5.10. Gene expression dynamics in response to Afatinib adaptation. (A) 549 driver genes clustered according to their expression. Cells are binned according to their reconstructed pseudotime. The average expression profile of the genes in the cluster and the cell composition of each bin as a percentage of cells in each timepoint is reported below each cluster. **(B)** Gene ontology enrichment analysis of 549 genes (bottom) and for each enriched term the percentage of genes belonging to a specific cluster of **(A)**. **(C)** Genes in the enriched term EGFR tyrosine kinase inhibitor resistance and the cluster of **(A)** in which they fall.

Subsequently, we conducted time-resolved Gene Ontology Enrichment Analyses (GOEA) to unravel the sequential activation of transcriptional programs guiding the adaptation to Afatinib.
Initially, we performed GOEA on all 549 driver genes (see *Table* 6.8) and then determined the specific activation times of each significant GO term

by calculating the proportion of genes it included from each cluster.
The GOEA analysis of the 549 driver genes (refer to *Figure* 5.10*B*) unveiled numerous biological processes previously associated with drug resistance in other cancer types, encompassing lysosome biogenesis, Reactive Oxygen Species (ROS) homeostasis, and fatty acid metabolism [288] [289] [290] [291] [292] [293] [294].

Additionally, it highlighted specific genes linked to EGFR resistance, such as PDGF-C [295], FGF2 [296] [297], PIK3R2 [298], HRAS [299], MAPK3 [300] [301], GAS6 [279] and BAX [302] (refer to *Figure* 5.10*C*).
Notably, PDGF-C and FGF2 serve as ligands for PDGFR- and FGFR receptors, respectively, which related pathways are well-established compensatory ones activated to counteract EGFR inhibition [303] [92]. This activation is often linked to the acquisition of mesenchymal features [303] [92], so pointing out to dediferrentiation, one of the most known non genetic drug resistance mechanism, a trait frequently sustained by the AXL pathway [304] [305] where GAS6 acts as an activator, an additional EGFR resistance gene we identified (refer to *Figure* 5.10*C*).

Furthermore, lysosome biogenesis and lipogenesis have been extensively studied for their involvement in drug adaptation across various cancer types, encompassing both standard chemotherapeutics and targeted therapy [288] [289] [290] [291] [292] [293] [294].
Enhanced lysosomal biogenesis facilitates the sequestration of hydrophobic weak base compounds, thereby mitigating the cytotoxic potential of a drug by limiting its availability at the site of action [288] [289]. Conversely, enhanced lipogenesis has been reported to contribute to drug adaptation by reducing drug uptake and participating in antioxidant cell defense through the regulation of membrane fluidity [290] [291] [292] [293] [294].

As depicted in *Figure* 5.9*A*, a discernible temporal program emerges, although most pathways feature genes distributed across all four clusters.
Notably, lysosomal genes show significant enrichment in clusters one and two, indicating a potential early cellular response to drug-induced damage. In contrast, genes associated with fatty acid metabolism, Reactive Oxygen Species (ROS), and EGFR TKI resistance (refer to *Figure* 5.9*B*) display

expression profiles falling within clusters three and four. Here, gene expression is elevated from day 0 (i.e., untreated cells), suggesting that these pathways may represent inherent characteristics of drug-tolerant cells, either required in later stages of drug adaptation (cluster four) or not (cluster three).

In conclusion, these findings substantiate the hypothesis that, in response to Afatinib, tolerant clones not only induce xenobiotic detoxification mechanisms to limit drug availability but also activate compensatory pathways beyond IGF1-R to sustain pro-survival signals.

This capability allows them to diminish Reactive Oxygen Species (ROS) production [306], concurrently alleviating their detrimental effects by generating antioxidants and facilitating Epithelial to Mesenchymal Transition (EMT) [92] [303].

Chapter 6

Conclusions

Due to the intricate genetic makeup of triple-negative breast cancer (TNBC) and the absence of recurrent oncogenic alterations, treatment options for TNBC are currently limited to conventional chemotherapies [14] [13] [21].
Despite increased EGFR expression in the majority of primary TNBCs, EGFR-targeted therapies have shown inconsistent clinical responses, necessitating the urgent exploration of novel approaches to identify biomarker genes that can effectively guide tailored EGFR-targeted treatments [35] [34] [36] [38] [37]. Clinical studies of EGFR inhibitors in various tumor types have revealed variable response rates, emphasizing the need for robust biomarker identification strategies to predict patient outcomes accurately and optimize EGFR-targeted therapy in TNBC.

In this investigation, we employed single-cell transcriptomics, cellular barcoding, and time-resolved computational analyses to comprehensively characterize the transcriptional profile of MDA-MB-468 cells in response to Afatinib treatment, a potent tyrosine kinase inhibitor (TKI) targeting both EGFR and HER2 proteins. Retrospective lineage tracing analysis uncovered a pre-existing subpopulation of Afatinib-tolerant MDA-MB-468 cells with distinct features, including elevated IGFBP2 gene expression.
Experimental validation demonstrated that IGFBP2 overexpression renders TNBC cells tolerant to Afatinib by activating the compensatory IGF1-R signaling pathway in MDA-MB-468 cells. Prospective lineage tracing

analysis revealed several mechanisms contributing to drug adaptation, including lysosome biogenesis, reactive oxygen species homeostasis, and fatty acid metabolism [288] [289] [290] [291] [292] [293] [294].

Our approach not only shed light on the molecular mechanisms of Afatinib resistance in MDA-MB-468 cells but also led to the development of an algorithm, scASTRAL, through deep learning techniques.
This algorithm accurately predicts Afatinib sensitivity based on the expression levels of 374 genes identified through retrospective lineage tracing. The signature demonstrated generalizability by accurately predicting Afatinib response in both publicly available and de novo single-cell datasets of TNBC cell lines.
Further studies are required to explore whether the transcriptional programs linked to Afatinib resistance are conserved across different tumor types and shared by other tyrosine kinase inhibitors (TKIs).

Understanding the diverse responses of cancer cell clones to treatment is crucial for comprehending how intra-tumor heterogeneity affects drug efficacy and identifying druggable targets for personalized cancer therapy. Our findings underscore the promising potential of lineage tracing techniques to reconstruct single-clone behaviors under various stimuli, facilitating the identification of predictive biomarker genes that can inform personalized treatment strategies for cancer patients.

Appendix - Materials and Methods

6.1 Experimental Methods

6.1.1 Set up of the lineage tracing experimental platform

For single-cell lineage tracing, we engineered MDA-MB-468 cells with Cellecta's CloneTracker XP 10M Barcode-3' Libraries that is a commercially available lentiviral barcodes library [268].
Each virion carries a plasmid containing Venus, a red fluorescent reporter gene for infected cell detection, and a Puromycin resistance gene for correctly infected cells selection. Importantly, the 3'-UTR of the Venus gene incorporates a 48-nucleotide barcode sequence.
To establish the cellular barcoded pool, 50,000 cells were infected with the library using LentiTrans™ Polybrene Transduction Reagent at a low MOI (0.05), resulting in a 5% infection rate and generating a population with 2,500 unique barcodes. After 72 hours post-infection, 1 μg/mL of Puromycin (Thermo Fisher Scientific) was added to the medium to selectively retain transduced MDA-MB-468 cells.
Following 5 days, the selected cells were evaluated for Venus expression using the BD AccuriTM C6 flow cytometer (BD Biosciences, Franklin Lakes, NJ, USA), confirming 100% positivity. Subsequently, the cells were expanded in culture medium supplemented with 1 μg/mL of Puromycin.
To enhance barcode identification in the parental and fully tolerant cellular populations, we utilized QuantSeq-Flex technology for bulk RNA-seq. For this purpose, total RNA was extracted from three independent biological replicates using the Qiagen RNeasy Mini kit, following the manufacturer's instructions. Subsequently, libraries were prepared according to the protocol provided by Lexogene. To specifically capture Cellecta barcodes, target-specific reverse transcription primers were designed, each incorporating a partial Illumina P7 adapter extension followed by the target-specific sequence, in accordance with the guidelines outlined in the QuantSeq-Flex manual. These primers were utilized as a pool at a concentration of 50 nM, with lengths ranging from 44 to 50 nucleotides as required by the manufacturer's protocol.

6.1.2 Time series experiment

To explore the mechanisms of drug resistance in TNBC, we subjected MDA-MB-468 cells to incremental concentrations of Afatinib and performed single-cell sequencing at intermediate time points. This approach allowed us to reconstruct changes in transcriptomic profiles over the course of treatment.
To achieve this, 5 x 105 barcoded cells were plated in triplicate in a 6-well plate and allowed to adhere for 24 hours. Concomitantly, control cells were plated in triplicate, and their viability was monitored throughout the experiment. Subsequently, the cells were exposed to incremental concentrations of Afatinib (Selleckchem cat. Num. S1011), starting from 250 nM, with each treatment lasting 72 hours.
The Afatinib concentration was doubled every nine days, reaching up to 2 μM. At intervals of 3 days, cell viability was assessed for both treated and untreated cells, and 1 x 10^4 treated cells were collected and cryopreserved from each treated replicate. This was achieved by centrifuging the cells at 1,100 rpm for 3 minutes, resuspending them with resuspension buffer (RB) (100% FBS) to obtain a final concentration of 1,000 cells/μl. Subsequently, 10 μl of this cell suspension was combined with 90 μl of RB to achieve a final concentration of 100 cells/μl (total: 1 x 10^4 cells). Finally, 100 μl of cold freezing buffer (FB) (100% FBS and 5% DMSO) was added before storing the cells at -80°C.

6.1.3 Single-cell sequencing with 10X Chromium platform

To single cell profile the cryopreserved samples derived from intermediate picking during time series experiment underwent thawing in a 37°C water bath and were thoroughly washed with MACS® Separation Buffer (Miltenyi Biotech).
Single-cell suspension viability was assessed using Trypan Blue on a LUNA-II™ Automated Cell Counter (Logos Biosystems). After filtering through a 40-μm filter and centrifugation, cells were resuspended to a final concentration of 1 x 10^3 cells/ μl. Single-cell libraries were generated using a Chromium Single Cell Gene Expression 3 Library and Gel Bead kit v3.1 [144]. The cells, along with gel beads, were injected into microfluidic chips

of the 10x Chromium instrument, forming nanoliter-scale Gel Beads-in-Emulsion (GEMs). Reverse transcription within the GEMs was carried out, followed by sample purification, cDNA amplification, and quantification. Amplified cDNAs were then fragmented, end-repaired, A-tailed, and subjected to post-ligation amplification.
Sequencing-ready libraries were purified, and their quality and concentration were assessed before sequencing on a NovaSeq 6000 SP 100 cycles flow cell.

6.1.4 RNA extraction, reverse transcription, and quantitative RT-PCR (qPCR)

Cell line total RNA extraction employed the Qiagen RNeasy Mini kit (Qiagen, Hilden, Germany) in accordance with the manufacturer's instructions. Subsequently, 1 μg of total RNA per sample was utilized for double-strand cDNA synthesis through the QuantiTect Reverse Transcription Kit (Qiagen). Quantitative Real-Time PCR (qRT-PCR) involved a reaction mixture comprising 10 μL of 2× Sybr Green (LightCycler 96, Roche Molecular Systems, Inc.), 200 nM of each primer, and 20 ng of the previously generated cDNA.
Relative gene expression was determined using the comparative C(T) method, with RP18S serving as the housekeeping gene.

6.1.5 Drug combination assays

A total of 4 x 10^4 MDA-MB-468 cells were plated in a 96-well plate and incubated overnight at 37°C. Afatinib and GSK-1904529A drugs were prepared in various dilutions and then combined in all possible pairs to generate either a 5 x 5 or 5 x 6 drug combination matrix. Subsequently, cells were exposed to either single-agent drugs or to the drug pairs from the combination matrix, with negative controls treated with DMSO (each treatment performed in triplicate). After a 72-hour incubation at 37°C, cell viability was assessed using the CellTiter (Promega), and absorbance was read at 490 nm with the GloMax® Discover instrument plate reader. Interactions between Afatinib and GSK-1904529A drugs, along with expected drug responses, were calculated using the Combenefit tool [307] employing the Loewe additivity model.

6.1.6 Single-cell sequencing with Drop-Seq platform

The SUM149PT, SUM185PE, SUM22PE, SUM159PT, and SUM1315MO2 cell lines underwent single-cell transcriptomics using DROP-seq technology. Library preparation followed the procedure outlined in [136]. Subsequently, scRNA libraries were sequenced on a NovaSeq 6000 machine using an SP 100 cycles flow cell.

6.1.7 dCas9 plasmids cloning, lentiviral packaging and transduction and sgRNA cloning

The plasmids facilitating cellular integration of dCas9, namely pRDKCE2B-EFS-dCas9-KRAB-2A-Blast, pRDVCE2B-EFS-dCas9-VPH-2A-Blast, pRSG16N-U6-sg-UbiC-TagRFP-2A-Neo, and pRSG17H-U6-sg-UbiC-TagGFP2-2A-Hygro, were obtained from Cellecta (#9E2B-PS, #9E2B-PS, #616N-L, and #617H-L, respectively).
Subsequently, lentiviral packaging was conducted following the manufacturer's guidelines (Cellecta, Mountain View, CA). Briefly, 24 hours before transfection, 4 x 10^6 HEK293T cells were seeded in a 100-mm tissue-culture dish. On the day of transfection, 2μg of the plasmid of interest was combined with 10μg of the packaging plasmid mix (Cellecta, psPAX2: pMD2.G) in DMEM without serum or antibiotics, along with Plus Reagent™ and Lipofectamine™ (Life Technologies). Cells were incubated with the DNA/Plus Reagent™/Lipofectamine™ mix for 24hours.
Viral supernatant was collected 48hours post-transfection, filtered through a 0.45μm PES filter, and concentrated using LentiFuge™ Viral Concentration Reagent (Cellecta) following the manufacturer's protocol. The concentrated lentiviral particles were re-suspended in PBS and stored at 80°C. Next, 5 x 10^5 MDA-MB-468 and HCC1806 cells were seeded in a 6-well plate and transduced with lentivirus to stably express dCas9-KRAB or dCas9-VHP in the presence of 8 μg/mL of LentiTrans™ Polybrene Transduction Reagent (Cellecta, LTDR1) at a MOI of 0.5. After 72 hours of transduction, 2.5 μg/mL of Blasticidin (Thermo Fisher Scientific, Waltham, MA, USA) was added to MDA-MB-468 cells, while 1 μg/mL of Blasticidin was used for the HCC1806 cell line. Once dCas9-KRAB and dCas9-VHP stable cell lines were established, they were transduced with pRSU6-gRNA lentivirus to modulate IGFBP2 expression with a MOI of

0.5. After 5 days of selection, sgRNA RFP-positive or GFP-positive cells were further sorted using BD FACS ARIA III.
The sgRNAs used to overexpress or downregulate the IGFBP2 gene, instead, was designed using the CRISPick Broad Institute tool [308]. IGFBP2 protospacer sequences were synthesized, hybridized, phosphorylated, and inserted into pRSU6-gRNA plasmids using BbsI sites.

6.1.8 Clonogenic survival assay

Cells (5 x 10^4) were seeded in a 12-well plate and allowed to adhere for 24 hours. Subsequently, the cells were treated with varying concentrations of Afatinib and incubated for either 10 days or 5 days. Fresh media was replenished on the fifth day. On the tenth day, the media was aspirated, and the dishes were washed once with ice-cold PBS.
Colonies were fixed and stained with a crystal violet solution for 45 minutes on a rocking platform. Following three rinses with PBS, the dishes were air-dried. The stained crystal violet was then solubilized with PBS-0.1% SDS, and absorbance at 590 nm was measured. The experiments were conducted in triplicate, and the results are presented as mean ± SD.

6.1.9 Spheroids generation and image acquisition

Monolayer-cultured cells were detached to create a single-cell suspension, diluted to 2.5 × 10^4 cells per milliliter of ice-cold medium, and combined with Cell Basement Membrane (ATCC, ATCC-ACS-3035) at a final concentration of 2.5%. The resulting cell suspension (200 μl) was dispensed into each well of an Ultra-Low Attachment (ULA) 96-well plate with a round or conical bottom (PerkinElmer, #6055330). Spheroid formation commenced with centrifugation of the plates at 300g for 10 minutes, followed by incubation under standard cell culture conditions (37 °C, 7% CO2) in humidified incubators.
For Afatinib response studies, the drug was introduced as follows: after 24 hours, when spheroids had formed, 100 μl of medium was aspirated, and treatment was administered by adding 100 μl of medium containing 2 μM of Afatinib to each well. Treatments were renewed after 3 days in fresh complete medium. Images were captured before treatment and at days 3, 5, and 7 post-treatment using the High Content Analysis System Operetta

CLS (PerkinElmer) with controlled temperature (37 °C) and CO2 (5%). Acquisition involved multiple planes, and area analyses were conducted on maximum projections using Harmony software (PerkinElmer). Volumetric analyses were performed through stack processing with 3D analysis.

6.1.10 IGFBP2 plasmid transfection in HDQ-P1 cell line

The pCMV6-AC-IGFBP2-GFP expression vector, containing the human IGFBP2 (NM_00597) gene fused to GFP in the C-terminal region, was acquired from Origene Technologies (RG202573). This plasmid facilitated the generation of HDQ-P1 cells with stable overexpression of the IGFBP2 gene. Transfection of HDQ-P1 cells was accomplished using Lipofectamine 2000 (Life Technologies, Grand Island, NY, USA) following the manufacturer's instructions. The stable-transfected HDQ-P1 cells were subjected to selection in a medium containing 500 μg/mL G418 (Life Technologies). After 7 days of selection, cells were sorted using BD FACS ARIA III to isolate the transfected population.

6.1.11 Cell migration assay

Transwell chambers (8-μm pores) (Euroclone) were used to perform transwell migration assays. Briefly, breast cancer cells were plated in the upper side of the transwell chamber in a serum-deprived medium. Next, as chemoattractant, 300 μl of complete medium was added in the lower chamber. After 24 hours, cells migrated to the lower side of the chambers were stained with crystal violet solution (crystal violet 0.05%, methanol 20%). Then, crystal violet in the chamber was de-stained with PBS-0.1% SDS solution and read at 570 nm.

6.1.12 Cell culture

HEK293T, MDA-MB-468, HCC1806 and HDQ-P1 were obtained from ATCC biobank. HEK293T and HDQ-P1 cells were cultured with DMEM supplemented with 10% foetal bovine serum (FBS), 1% L-glutamine, 1% penicillin/streptomycin (Euroclone, Milan, Italy), whereas MDA-MB-468 and HCC1806 cells were cultured with RPMI supplemented with 10% foetal bovine serum (FBS), 1% L-glutamine, 1% penicillin/streptomycin

(Euroclone). Cells were kept at 37 °C under 5% CO2 atmosphere. SUM series cell lines were obtained from BioIVT biobank. SUM149PT, SUM185PE, SUM225CWN, SUM229PE and SUM159PT were cultured with Ham's F12 medium (Gibco) supplemented with 5% heat-inactivated FBS (Gibco), 10 mM HEPES (Sigma-Aldrich), 1 μg/ml hydrocortisone (Sigma-Aldrich), 5 μg/ml insulin (Sigma-Aldrich) and 1% L-glutamine (Euroclone). SUM1315MO2 were cultured with the Ham's F12 medium supplemented as described above with in addition 10 ng/mL EGF. No penicillin-streptomycin was added to the SUM cell line medium, as recommended by the suppliers.

6.2 Computational Methods

6.2.1 Lineage barcode identification from RNA sequencing data

The lineage barcode sequence introduced by lentivirus in Cellecta library [268] is strategically located 70nt upstream of polyA sequence making it compatible with oligo-dT cDNA synthesis in RNA-sequencing library preparation. Practically, lentivirus integrates the plasmid into infected cell DNA, facilitating barcode transcription. Resulting RNA fragments, carrying the barcode, are captured at the 3' end using indexed oligo-dT primers, initiating essential cDNA first strand synthesis. This method enables barcode identification from both single cell and bulk RNA sequencing data.
Next, we developed a computational barcode identification from FASTQ files takes advantage of the distinctive structure of the barcode. The 48-nucleotide sequence comprises two sub-sequences: one 14-nt long (bc14) and the other 30-nt long (bc30), separated by a 4-nt (TGGT) anchor. The first sub-sequence, bc14, can occur in 100 different ways, while bc30 can manifest in 100,000 different occurrences, resulting in a total of 10,000,000 possible combinations and, consequently, the same number of potential clones to be uniquely tagged. Notably, Cellecta provides a whitelist containing the 100 possible bc14, and 100,000 bc30 can appear on both sides of the anchor.
The initial step of barcode identification is the search of the TGGT anchor, allowing for one possible mismatch. Then, it extracts putative bc14 (14-nt bases before the anchor) and putative bc30 (30-nt bases after the

anchor) if a match is found. If both putative bc14 and bc30 are present in the Cellecta whitelist, the lineage barcode, represented by the combined string (bc14:bc30), is assigned to the corresponding cell from read one (for single-cell data).
If either putative lineage barcode is absent from the Cellecta whitelist, correction is made using the closest barcode sequence with a Hamming distance of one, but only if no other barcodes in the whitelist share this distance. In cases where both putative bc14 and bc30 are not in the Cellecta whitelist, the bc14 barcode undergoes correction first, following the described method.
The corresponding bc30 is corrected only if the bc14 correction is successful, reducing the computational demand for sequencing errors without sacrificing sensitivity. At the conclusion of this iterative process, each sequenced cell is associated with the most abundant pair of lineage barcodes (bc14:bc30) identified. Only reads where both bc14 and bc30, after correction, match those provided by Cellecta's whitelist are utilized, ensuring precise lineage barcode assignment.
This process was implemented in an R script using the Biostrings R package for string matching and a custom C++ function for computing Hamming distances between DNA sequences of the same length.

6.2.2 Analysis of single-cell RNA-seq data from 10X Chromium platform

The scRNA-seq data analysis initiated with demultiplexing and the alignment of raw sequencing reads to the Hg38 human reference genome using Cell Ranger tool version 6.1.2.
For alignment, we utilized GENCODE annotation v32 of Hg38, focusing on genes with biotypes protein coding, lincRNA, and antisense. Subsequently, cells with fewer than 5,000 Unique Molecular Identifiers (UMI) were excluded.
To further refine the dataset, putative cell doublets and cells expressing a high fraction of mitochondrial reads were removed using the filterCell function from the gficf version 2 R package, which is available at [309].
This function, employing loess regression, modeled the relationship between the total UMI count (log scale) and the ratio of UMIs in mitochondrial genes over the total UMI count. Cells with a significantly deviating

ratio, as indicated by an adjusted p-value (FDR) < 0.1, were discarded.
A similar strategy was employed to identify and remove putative cell doublets. Here, loess regression modeled the relationship between the total UMI count in a cell (log scale) and the ratio of detected genes over the total UMI count. Cells with a significantly deviating ratio, again indicated by an FDR < 0.1, were discarded.
Drawing on our extensive experience in single-cell data analysis, this approach not only effectively eliminated cell doublets but also addressed technical issues associated with cells exhibiting a high total UMI count concentrated in a small number of genes.
The resultant dataset, consisting of 4,101 highly covered cells post-filtering, formed the basis for all subsequent analyses. To address potential batch effects arising from sequencing day 0 cells using two distinct captures in different flow cells, we applied the CombatSeq function from the sva R package [157].
Furthermore, genes expressed in fewer than 50 cells and in less than 5% of sequenced cells were excluded, but only if their average expression was less than 1.12 UMI, following a methodology similar to that in [310].
Finally, this pre-processed and filtered expression matrix underwent analysis using the standard Seurat pipeline (v3) with default parameters. The pipeline includes normalization with the LogNormalize function, feature selection with FindVariableFeatures, and data scaling with the ScaleData function. Subsequently, dimensionality reduction was performed by computing PCA with RunPCA, followed by UMAP computation using the RunUMAP function. Differentially expressed genes in MDA-MB-468 control cells between tolerant and sensitive Afatinib clones were identified using the FindMarkers function with logfc.threshold and min.pct parameters set to 0.

6.2.3 scASTRAL architecture and training

The scASTRAL architecture is built upon a contrastive autoencoder model. The encoder includes an input layer with 374 neurons, representing the number of Afatinib response biomarker genes identified through retrospective lineage tracing.
Subsequent to the input layer, two hidden layers, comprising 64 and 32 neurons, respectively, utilize the Rectified Linear Unit (ReLU) activation

function. The decoder replicates the architecture of the encoder.
scASTRAL underwent training using the 1,541 MDA-MB-468 control cells categorized as afatinib-tolerant or afatinib-sensitive in the training set. Prior to training initiation, the matrix of normalized Counts Per Million (CPM) was trimmed to the 374 intrinsic Afatinib response biomarker genes, whose expression was then rescaled using tf-idf transformation [212]. This dataset was randomly divided into 5 batches of equal size, ensuring a balanced distribution of Afatinib-tolerant and Afatinib-sensitive cells across batches. Then, the contrastive model training proceeded as follows. First, for each cell in the batch, a positive and a negative example were randomly generated, therefore ensuring that the number of training examples for each batch equaled twice the number of cells in that batch. Then, the contrastive model was trained using a loss function composed of the weighted sum of three terms: *(i)* the cosine embedding loss, *(ii)* the reconstruction error, and *(iii)* the latent space regularization factor. The cosine embedding loss maximizes the cosine distance between pairs of cells labeled with -1 (negative example) while simultaneously maximizing the cosine similarity between pairs labeled with +1 (positive example). The reconstruction error ensures that the decoder could reconstruct the original data from the latent representation generated by the encoder, incorporating dependencies among the 374 input genes into the final embedding space.
Subsequently, an SVM classifier was trained using the reconstructed embedding space of the considered cell batch. Specifically, the SVM classifier employed a cosine kernel with a regularization hyperparameter set to 100. The classification accuracy of the SVM trained on a batch was assessed using the left-out cells. The Adam optimizer [311] was employed with a learning rate of 0.0001, and training utilized an early stopping criterion with a maximum of 250 epochs. Metrics for evaluating improvement across epochs included validation loss and SVM classification accuracy. Training stopped if no improvement was observed over the last 20 epochs. Hyperparameters for the contrastive model were determined through a grid search and after training the five contrastive models along with their associated SVM models, the model achieving the highest classification accuracy was chosen for predicting Afatinib sensitivity at the single-cell level, detailed below.

scASTRAL was implemented in Python using PyTorch version 1.13.1, and the code is accessible at [312].

6.2.4 Analysis of single-cell sequencing data from Drop-Seq platform

Raw reads pre-processing was performed as described in Macosko et al.. Only high depth cells with at least 5,000 UMI were retained and used to test the scATRAL tool.
The alignment pipeline can be found at [313].

6.2.5 Evaluation of scASTRAL performance on single-cell profiles from TNBC cell lines

To validate scASTRAL and assess its predictive capability for Afatinib sensitivity in TNBC cell lines, we utilized 22,724 single-cell transcriptional profiles derived from 16 TNBC cell lines of which 11 were sourced from [136], while the remaining 5 were newly sequenced using the Drop-seq platform [5], see experimental details in 6.1.6.
Prior to input into the scASTRAL method, the raw UMI count matrix for each cell line was normalized using sklearn to obtain Counts Per Million (CPM).
Subsequently, the matrix was trimmed to include only the 374 Afatinib response biomarker genes and their expression values were then rescaled using tf-idf transformation before applying scASTRAL.
For the classification of Afatinib sensitivity, a cell was considered sensitive if the SVM classification probability exceeded 0.75; otherwise, it was classified as tolerant. This threshold was determined during the SVM training phase to maximize classification accuracy.
To translate predictions from the single-cell level to the cell-line level, we calculated the fraction of predicted Afatinib-sensitive cells in each cell line, by dividing the number of cells predicted as sensitive by the total number of cells predicted as either sensitive or tolerant.

6.2.6 Pseudotime reordering and time-course differential expression analysis in dominant/neutral clones

The 2,836 single-cell expression profiles linked to Afatinib-tolerant clones were categorized based on the behavior of their clone of origin into two groups: "dominant" and "neutral." The matrix of dominant clones consisted of 1,293 cells, while the matrix of neutral clones comprised 1,543 cells.
Subsequently, we ordered cells over a time-course aware pseudotime trajectory by employing the psupertime function from the psupertime R package [286] with default parameters, utilizing the sequencing day (i.e., 0, 3, 6, 9 days) as cell labels.
For both matrices, the expression profile of each gene underwent a three-step process: *(i)* ordering based on the reconstructed pseudotime, (ii) smoothing using the moving median method implemented in the rollmedian function from the zoo R package which was employed with a rolling window of 51, and *(iii)* division into 50 expression bins according to the inferred pseudotime, so the expression profiles of each gene in both matrices were represented by 50 pseudotime points.
Finally, to identify upregulated genes in cells associated with dominant clones over the first 9 days of treatment, we utilized the splineTimeR package to compare the reconstructed time-course profile of a gene across cells of the dominant clones with its counterpart in the neutral clones. Upregulated genes were defined as those with a False Discovery Rate (FDR) $<$ 5% and a positive average fold change across the 50 binned pseudotime points.

6.2.7 Significance evaluation of dose-response curves trend using Gaussian processes

We employed Gaussian processes to assess the significance of variations in dose-response curve trends. Following the methodology in [314], we reconstructed the dose-response trend for each condition by interpolating response values at each tested concentration. Subsequently, we identified differences between the two trends (i.e. the two dose-response curves) by calculating the log-likelihood and the associated Bayesian factor (BF) which is reported in bits.

6.2.8 Estimation of EGFR copy number alterations, mutations, and expression in TNBC.

Two independent cohorts of TNBC patients were used to estimate gene copy number alterations, mutations and expression reported in *Figure* 5.1.
The first cohort composed of 192 TNBC patients was obtained from the Genomic Data Common (GDC) portal. GDC raw bulk expression, mutational data and copy number variation data along with clinical information were retrieved using TCGAbiolinks R package v.2.25.3.
The second cohort was composed of 465 TNBC Asiatic patients [22] for whom gene expression, gene copy number alteration and mutations were available was gathered from the NODE database after the author's authorization.
The raw expression counts downloaded from GDC were first normalized with edgeR package and transformed in log2(CPM+1), while those from the Asiatic cohort were already normalized as FPKM values, and thus we only transformed them into log2(FPKM + 1).
A sample in the GDC cohort was defined to have EGFR copy gain if the number of copies was 3 or 4, while EGFR amplification was defined when this number was greater than 5.
For the Asiatic cohort, EGFR gain and amplification were defined based on GISTIC scores, respectively, +1 and +2.
For both cohorts, mutations were considered only if they were in the coding region (i.e., missense, non-sense, in-frame and out-of-frame deletions). Next, the deleterious tag was assigned when at least two functional annotations among VEP, PolyPhen and SIFT reported a negative impact of the mutation on protein function.
Finally, EGFR activating mutations tag was conferred based on indications in [315].

Appendix - Supplementary Tables

Table 6.1. Differentially expressed genes in Afatinib-tolerant vs Afatinib-sensitive subpopulations in untreated MDA-MB-468 cells. **(A)** Ensembl gene ID **(B)** HGNC gene symbol **(C)** Log2 fold change in gene expression between the two subpopulations. A positive sign means that the gene is upregulated in Afatinib-tolerant cells, while a negative sign means that the gene is upregulated in Afatinib-sensitive cells **(D)** Percentage of cells expressing given gene in Afatinib-tolerant subpopulation **(E)** Percentage of cells expressing given gene in Afatinib-sensitive subpopulation **(F)** p-value of Student's t-test **(G)** False discovery rate (FDR) by Benjamini-Hochberg method from the p-value of Student's t-test.
Only significant (FDR < 0.1) are shown

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000108826	MRPL27	0.16	0.85	0.78	2.30E-07	1.144E-04
ENSG00000008018	PSMB1	-0.14	0.94	0.95	2.49E-07	1.149E-04
ENSG00000188001	TPRG1	0.08	0.19	0.10	2.46E-07	1.149E-04
ENSG00000179862	CITED4	0.26	0.98	0.98	2.59E-07	1.151E-04
ENSG00000100097	LGALS1	-0.18	0.99	1.00	3.20E-07	1.376E-04
ENSG00000134986	NREP	0.15	0.67	0.57	3.69E-07	1.532E-04
ENSG00000182196	ARL6IP4	0.11	0.97	0.96	6.69E-07	2.690E-04
ENSG00000167815	PRDX2	0.12	0.98	0.98	7.07E-07	2.753E-04
ENSG00000126709	IFI6	-0.12	0.09	0.18	7.89E-07	2.981E-04
ENSG00000143977	SNRPG	0.12	0.98	0.96	9.11E-07	3.339E-04
ENSG00000171208	NETO2	-0.20	0.78	0.85	1.25E-06	4.464E-04
ENSG00000112769	LAMA4	0.17	0.89	0.85	1.38E-06	4.791E-04
ENSG00000089280	FUS	-0.15	0.98	0.99	1.53E-06	5.153E-04
ENSG00000124610	H1-1	0.30	0.71	0.62	1.69E-06	5.536E-04
ENSG00000164104	HMGB2	0.16	0.94	0.89	2.30E-06	7.342E-04
ENSG00000197061	H4C3	0.38	0.99	0.98	2.82E-06	8.778E-04

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000204262	COL5A2	-0.11	0.04	0.10	2.99E-06	9.095E-04
ENSG00000135046	ANXA1	-0.15	0.99	0.98	3.10E-06	9.205E-04
ENSG00000135480	KRT7	-0.11	0.99	0.99	3.43E-06	9.930E-04
ENSG00000169908	TM4SF1	-0.18	0.84	0.90	3.55E-06	1.006E-03
ENSG00000111786	SRSF9	0.11	0.96	0.96	3.91E-06	1.058E-03
ENSG00000113389	NPR3	0.28	0.53	0.44	3.90E-06	1.058E-03
ENSG00000175315	CST6	-0.30	0.63	0.70	4.53E-06	1.201E-03
ENSG00000091436	MAP3K20	0.16	0.90	0.90	4.67E-06	1.212E-03
ENSG00000164485	IL22RA2	0.13	0.39	0.29	6.46E-06	1.642E-03
ENSG00000122641	INHBA	-0.14	0.45	0.55	6.79E-06	1.691E-03
ENSG00000146648	EGFR	-0.20	1.00	1.00	7.14E-06	1.721E-03
ENSG00000173156	RHOD	0.12	0.92	0.91	7.18E-06	1.721E-03
ENSG00000104979	C19orf53	0.11	0.96	0.95	7.49E-06	1.760E-03
ENSG00000135940	COX5B	0.09	0.99	0.99	1.02E-05	2.356E-03
ENSG00000198176	TFDP1	0.11	0.90	0.88	1.04E-05	2.356E-03
ENSG00000139714	MORN3	0.08	0.66	0.56	1.12E-05	2.490E-03
ENSG00000148773	MKI67	0.22	0.88	0.83	1.23E-05	2.689E-03
ENSG00000101639	CEP192	0.14	0.59	0.48	1.31E-05	2.771E-03
ENSG00000177700	POLR2L	0.10	0.98	0.97	1.29E-05	2.771E-03
ENSG00000065054	SLC9A3R2	0.12	0.81	0.78	1.48E-05	3.021E-03
ENSG00000198542	ITGBL1	-0.04	0.03	0.07	1.48E-05	3.021E-03
ENSG00000231826	LINC01819	0.13	0.87	0.84	1.50E-05	3.021E-03
ENSG00000006625	GGCT	0.11	0.96	0.96	1.62E-05	3.211E-03
ENSG00000131876	SNRPA1	0.13	0.90	0.86	1.85E-05	3.610E-03
ENSG00000071994	PDCD2	-0.11	0.71	0.77	1.91E-05	3.655E-03
ENSG00000118193	KIF14	0.21	0.55	0.45	2.05E-05	3.879E-03

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000178209	PLEC	-0.17	0.93	0.93	2.12E-05	3.946E-03
ENSG00000168243	GNG4	-0.06	0.08	0.14	2.19E-05	4.012E-03
ENSG00000026508	CD44	-0.16	1.00	1.00	2.41E-05	4.299E-03
ENSG00000109452	INPP4B	-0.05	0.04	0.09	2.42E-05	4.299E-03
ENSG00000125266	EFNB2	-0.12	0.28	0.37	2.51E-05	4.410E-03
ENSG00000116667	C1orf21	-0.08	0.37	0.46	2.61E-05	4.453E-03
ENSG00000134508	CABLES1	-0.07	0.19	0.27	2.58E-05	4.453E-03
ENSG00000102038	SMARCA1	0.03	0.15	0.08	3.29E-05	5.400E-03
ENSG00000169021	UQCRFS1	0.08	0.99	0.99	3.28E-05	5.400E-03
ENSG00000197409	H3C4	0.10	0.37	0.27	3.27E-05	5.400E-03
ENSG00000162496	DHRS3	-0.07	0.07	0.13	3.59E-05	5.806E-03
ENSG00000133112	TPT1	-0.08	0.99	1.00	3.68E-05	5.882E-03
ENSG00000196230	TUBB	0.11	1.00	1.00	4.22E-05	6.658E-03
ENSG00000135318	NT5E	-0.20	0.35	0.44	5.27E-05	8.208E-03
ENSG00000107738	VSIR	-0.07	0.20	0.28	6.13E-05	9.429E-03
ENSG00000138119	MYOF	-0.13	0.94	0.95	6.39E-05	9.710E-03
ENSG00000035115	SH3YL1	0.13	0.78	0.69	6.82E-05	1.017E-02
ENSG00000119333	DYNC2I2	0.12	0.96	0.95	6.85E-05	1.017E-02
ENSG00000136518	ACTL6A	0.13	0.79	0.72	7.15E-05	1.048E-02
ENSG00000135679	MDM2	0.11	0.46	0.37	7.51E-05	1.075E-02
ENSG00000142507	PSMB6	0.10	0.95	0.92	7.48E-05	1.075E-02
ENSG00000131747	TOP2A	0.23	0.84	0.79	7.70E-05	1.091E-02
ENSG00000188486	H2AX	0.13	0.84	0.77	8.62E-05	1.207E-02
ENSG00000139433	GLTP	0.08	0.86	0.83	8.82E-05	1.221E-02
ENSG00000186591	UBE2H	-0.11	0.73	0.78	9.48E-05	1.298E-02
ENSG00000111275	ALDH2	0.11	0.54	0.47	1.000E-04	1.342E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000153113	CAST	-0.13	0.95	0.96	1.002E-04	1.342E-02
ENSG00000075618	FSCN1	-0.09	0.26	0.34	1.030E-04	1.366E-02
ENSG00000053254	FOXN3	0.08	0.45	0.37	1.071E-04	1.386E-02
ENSG00000111057	KRT18	-0.15	0.99	0.99	1.079E-04	1.386E-02
ENSG00000140153	WDR20	0.08	0.44	0.35	1.073E-04	1.386E-02
ENSG00000159228	CBR1	0.10	0.91	0.89	1.102E-04	1.402E-02
ENSG00000172531	PPP1CA	0.08	0.97	0.98	1.123E-04	1.412E-02
ENSG00000234741	GAS5	-0.10	0.99	0.99	1.133E-04	1.412E-02
ENSG00000171863	RPS7	0.07	1.00	1.00	1.200E-04	1.481E-02
ENSG00000151914	DST	-0.05	0.66	0.73	1.218E-04	1.488E-02
ENSG00000172348	RCAN2	-0.07	0.17	0.25	1.271E-04	1.538E-02
ENSG00000146674	IGFBP3	-0.19	0.50	0.58	1.290E-04	1.546E-02
ENSG00000124098	FAM210B	0.11	0.85	0.81	1.307E-04	1.551E-02
ENSG00000154518	ATP5MC3	0.09	0.97	0.97	1.337E-04	1.571E-02
ENSG00000108244	KRT23	-0.14	0.92	0.92	1.414E-04	1.617E-02
ENSG00000147676	MAL2	-0.14	0.93	0.94	1.414E-04	1.617E-02
ENSG00000176890	TYMS	0.10	0.97	0.96	1.410E-04	1.617E-02
ENSG00000132530	XAF1	-0.02	0.01	0.04	1.531E-04	1.734E-02
ENSG00000107779	BMPR1A	0.07	0.41	0.32	1.575E-04	1.768E-02
ENSG00000141753	IGFBP4	0.11	0.79	0.72	1.621E-04	1.771E-02
ENSG00000225663	MCRIP1	0.10	0.71	0.67	1.632E-04	1.771E-02
ENSG00000232124		0.02	0.08	0.03	1.634E-04	1.771E-02
ENSG00000244509	APOBEC3C	-0.08	0.51	0.61	1.614E-04	1.771E-02
ENSG00000235823	OLMALINC	0.13	0.56	0.49	1.694E-04	1.820E-02
ENSG00000085063	CD59	-0.12	0.95	0.94	1.724E-04	1.836E-02
ENSG00000144063	MALL	-0.10	0.48	0.55	1.753E-04	1.851E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000180198	RCC1	0.08	0.65	0.56	1.857E-04	1.928E-02
ENSG00000198824	CHAMP1	0.08	0.68	0.61	1.842E-04	1.928E-02
ENSG00000130429	ARPC1B	-0.09	0.98	0.98	1.913E-04	1.944E-02
ENSG00000137700	SLC37A4	0.08	0.52	0.43	1.919E-04	1.944E-02
ENSG00000161203	AP2M1	0.11	0.96	0.94	1.892E-04	1.944E-02
ENSG00000077942	FBLN1	-0.16	0.75	0.79	2.092E-04	2.062E-02
ENSG00000133216	EPHB2	-0.03	0.04	0.08	2.102E-04	2.062E-02
ENSG00000186665	C17orf58	0.03	0.26	0.18	2.086E-04	2.062E-02
ENSG00000242125	SNHG3	0.08	0.86	0.81	2.056E-04	2.062E-02
ENSG00000123179	EBPL	0.10	0.86	0.82	2.154E-04	2.097E-02
ENSG00000111341	MGP	-0.37	0.98	0.98	2.196E-04	2.121E-02
ENSG00000074071	MRPS34	0.10	0.96	0.95	2.247E-04	2.154E-02
ENSG00000089220	PEBP1	0.08	0.98	0.98	2.330E-04	2.166E-02
ENSG00000144029	MRPS5	0.11	0.93	0.91	2.300E-04	2.166E-02
ENSG00000174547	MRPL11	0.09	0.92	0.91	2.316E-04	2.166E-02
ENSG00000253438	PCAT1	0.06	0.13	0.07	2.324E-04	2.166E-02
ENSG00000089009	RPL6	0.07	1.00	1.00	2.367E-04	2.185E-02
ENSG00000151498	ACAD8	0.06	0.32	0.23	2.450E-04	2.245E-02
ENSG00000099817	POLR2E	0.09	0.90	0.88	2.549E-04	2.302E-02
ENSG00000148908	RGS10	-0.12	0.90	0.90	2.568E-04	2.302E-02
ENSG00000272068	BCAN-AS1	-0.01	0.01	0.04	2.558E-04	2.302E-02
ENSG00000162552	WNT4	0.05	0.25	0.17	2.616E-04	2.328E-02
ENSG00000040275	SPDL1	0.12	0.75	0.68	2.707E-04	2.392E-02
ENSG00000143469	SYT14	-0.06	0.09	0.16	2.726E-04	2.392E-02
ENSG00000090520	DNAJB11	0.10	0.73	0.68	2.786E-04	2.425E-02
ENSG00000099624	ATP5F1D	0.09	0.96	0.96	2.802E-04	2.425E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000163041	H3-3A	0.08	1.00	1.00	2.946E-04	2.497E-02
ENSG00000167700	MFSD3	0.07	0.65	0.58	2.938E-04	2.497E-02
ENSG00000183648	NDUFB1	0.09	0.97	0.95	2.928E-04	2.497E-02
ENSG00000163993	S100P	-0.08	0.08	0.13	3.034E-04	2.542E-02
ENSG00000176102	CSTF3	0.09	0.94	0.94	3.039E-04	2.542E-02
ENSG00000176912	TYMSOS	0.10	0.77	0.74	3.106E-04	2.580E-02
ENSG00000178999	AURKB	0.14	0.63	0.58	3.159E-04	2.590E-02
ENSG00000130312	MRPL34	0.10	0.91	0.90	3.259E-04	2.654E-02
ENSG00000125445	MRPS7	0.07	0.94	0.93	3.336E-04	2.667E-02
ENSG00000147100	SLC16A2	-0.11	0.40	0.49	3.310E-04	2.667E-02
ENSG00000185950	IRS2	-0.13	0.65	0.73	3.339E-04	2.667E-02
ENSG00000231426	FILNC1	-0.04	0.01	0.03	3.370E-04	2.675E-02
ENSG00000103534	TMC5	-0.06	0.10	0.16	3.581E-04	2.820E-02
ENSG00000117713	ARID1A	0.07	0.74	0.67	3.599E-04	2.820E-02
ENSG00000013364	MVP	-0.09	0.24	0.30	3.655E-04	2.847E-02
ENSG00000231131	LNCAROD	-0.07	0.14	0.21	3.768E-04	2.916E-02
ENSG00000136810	TXN	-0.09	0.97	0.98	3.827E-04	2.926E-02
ENSG00000188042	ARL4C	-0.14	0.80	0.83	3.826E-04	2.926E-02
ENSG00000104852	SNRNP70	0.10	0.94	0.91	4.012E-04	3.049E-02
ENSG00000140451	PIF1	0.13	0.26	0.18	4.091E-04	3.090E-02
ENSG00000144366	GULP1	-0.01	0.00	0.02	4.117E-04	3.091E-02
ENSG00000198836	OPA1	0.08	0.51	0.42	4.320E-04	3.224E-02
ENSG00000170312	CDK1	0.17	0.83	0.77	4.421E-04	3.279E-02
ENSG00000185479	KRT6B	-0.21	0.28	0.35	4.514E-04	3.329E-02
ENSG00000073282	TP63	0.02	0.08	0.04	4.652E-04	3.334E-02
ENSG00000178053	MLF1	-0.09	0.40	0.48	4.596E-04	3.334E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000184924	PTRHD1	0.10	0.86	0.83	4.632E-04	3.334E-02
ENSG00000204592	HLA-E	-0.10	0.83	0.87	4.655E-04	3.334E-02
ENSG00000230124	ACBD6	0.10	0.89	0.86	4.625E-04	3.334E-02
ENSG00000090530	P3H2	-0.10	0.28	0.35	4.753E-04	3.365E-02
ENSG00000138668	HNRNPD	0.10	0.97	0.96	4.739E-04	3.365E-02
ENSG00000080986	NDC80	0.14	0.76	0.70	4.836E-04	3.405E-02
ENSG00000115241	PPM1G	0.09	0.98	0.97	4.874E-04	3.412E-02
ENSG00000111716	LDHB	0.08	0.97	0.97	4.981E-04	3.468E-02
ENSG00000277443	MARCKS	0.11	0.97	0.97	5.080E-04	3.517E-02
ENSG00000122644	ARL4A	0.12	0.75	0.72	5.235E-04	3.570E-02
ENSG00000177879	AP3S1	-0.08	0.41	0.48	5.294E-04	3.570E-02
ENSG00000188015	S100A3	0.05	0.52	0.44	5.301E-04	3.570E-02
ENSG00000204922	UQCC3	0.08	0.93	0.93	5.262E-04	3.570E-02
ENSG00000240541	TM4SF1-AS1	-0.05	0.11	0.17	5.228E-04	3.570E-02
ENSG00000156411	ATP5MJ	0.08	0.97	0.95	5.483E-04	3.673E-02
ENSG00000075856	SART3	0.05	0.64	0.55	5.653E-04	3.767E-02
ENSG00000167393	PPP2R3B	0.07	0.33	0.25	5.683E-04	3.767E-02
ENSG00000111665	CDCA3	0.12	0.51	0.43	5.754E-04	3.777E-02
ENSG00000176485	PLAAT3	-0.12	0.64	0.69	5.759E-04	3.777E-02
ENSG00000198771	RCSD1	0.07	0.12	0.07	5.814E-04	3.793E-02
ENSG00000102359	SRPX2	-0.07	0.19	0.26	5.896E-04	3.823E-02
ENSG00000171865	RNASEH1	0.10	0.92	0.89	5.922E-04	3.823E-02
ENSG00000086570	FAT2	0.05	0.11	0.06	6.200E-04	3.942E-02
ENSG00000145386	CCNA2	0.13	0.60	0.54	6.198E-04	3.942E-02
ENSG00000188910	GJB3	-0.06	0.22	0.30	6.201E-04	3.942E-02
ENSG00000087884	AAMDC	-0.10	0.79	0.82	6.293E-04	3.980E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000255690	TRIL	0.02	0.11	0.06	6.405E-04	4.031E-02
ENSG00000177030	DEAF1	0.07	0.48	0.40	6.494E-04	4.066E-02
ENSG00000124102	PI3	0.09	0.29	0.21	6.673E-04	4.076E-02
ENSG00000148671	ADIRF	-0.20	0.85	0.87	6.569E-04	4.076E-02
ENSG00000167754	KLK5	-0.11	0.97	0.97	6.673E-04	4.076E-02
ENSG00000183696	UPP1	-0.15	0.95	0.94	6.607E-04	4.076E-02
ENSG00000242265	PEG10	0.14	0.99	0.98	6.635E-04	4.076E-02
ENSG00000107438	PDLIM1	-0.09	0.60	0.66	6.741E-04	4.077E-02
ENSG00000180035	ZNF48	0.05	0.23	0.16	6.724E-04	4.077E-02
ENSG00000065882	TBC1D1	-0.12	0.64	0.68	6.873E-04	4.126E-02
ENSG00000126777	KTN1	0.09	0.99	1.00	6.887E-04	4.126E-02
ENSG00000172922	RNASEH2C	0.10	0.86	0.84	6.935E-04	4.135E-02
ENSG00000228716	DHFR	0.09	0.87	0.82	6.979E-04	4.141E-02
ENSG00000003096	KLHL13	0.06	0.29	0.21	7.029E-04	4.151E-02
ENSG00000283235		-0.01	0.00	0.01	7.142E-04	4.198E-02
ENSG00000165949	IFI27	-0.17	0.04	0.08	7.224E-04	4.226E-02
ENSG00000114115	RBP1	-0.15	0.84	0.84	7.383E-04	4.279E-02
ENSG00000197467	COL13A1	-0.07	0.16	0.23	7.359E-04	4.279E-02
ENSG00000126878	AIF1L	0.06	0.74	0.65	7.427E-04	4.284E-02
ENSG00000159055	MIS18A	0.08	0.60	0.53	7.472E-04	4.286E-02
ENSG00000224259	LINC01133	-0.17	0.45	0.52	7.498E-04	4.286E-02
ENSG00000277283		0.02	0.11	0.06	7.566E-04	4.305E-02
ENSG00000109654	TRIM2	-0.06	0.23	0.30	7.603E-04	4.306E-02
ENSG00000029993	HMGB3	0.11	0.97	0.94	7.779E-04	4.386E-02
ENSG00000126216	TUBGCP3	0.10	0.61	0.52	7.901E-04	4.422E-02
ENSG00000137965	IFI44	-0.03	0.04	0.08	7.913E-04	4.422E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000132746	ALDH3B2	0.09	0.65	0.59	8.017E-04	4.440E-02
ENSG00000261122	LINC02167	0.07	0.24	0.18	8.005E-04	4.440E-02
ENSG00000172893	DHCR7	0.09	0.87	0.82	8.138E-04	4.487E-02
ENSG00000124575	H1-3	0.14	0.47	0.38	8.227E-04	4.505E-02
ENSG00000176101	SSNA1	0.09	0.93	0.93	8.243E-04	4.505E-02
ENSG00000006125	AP2B1	-0.14	0.77	0.79	8.435E-04	4.522E-02
ENSG00000079931	MOXD1	-0.04	0.11	0.17	8.486E-04	4.522E-02
ENSG00000136824	SMC2	0.13	0.83	0.76	8.453E-04	4.522E-02
ENSG00000138735	PDE5A	-0.10	0.24	0.31	8.371E-04	4.522E-02
ENSG00000151748	SAV1	-0.11	0.65	0.69	8.477E-04	4.522E-02
ENSG00000170558	CDH2	-0.01	0.00	0.02	8.492E-04	4.522E-02
ENSG00000134690	CDCA8	0.15	0.68	0.63	8.542E-04	4.529E-02
ENSG00000143546	S100A8	0.12	0.14	0.09	8.650E-04	4.567E-02
ENSG00000166794	PPIB	0.08	0.97	0.97	8.709E-04	4.579E-02
ENSG00000143322	ABL2	-0.12	0.56	0.62	8.750E-04	4.581E-02
ENSG00000117115	PADI2	0.08	0.55	0.48	8.797E-04	4.587E-02
ENSG00000112242	E2F3	-0.12	0.75	0.80	9.173E-04	4.763E-02
ENSG00000120594	PLXDC2	0.07	0.79	0.71	9.241E-04	4.778E-02
ENSG00000089685	BIRC5	0.12	0.93	0.91	9.618E-04	4.936E-02
ENSG00000135390	ATP5MC2	0.09	0.95	0.94	9.625E-04	4.936E-02
ENSG00000213190	MLLT11	0.03	0.19	0.13	9.698E-04	4.941E-02
ENSG00000250802	ZBED3-AS1	0.02	0.06	0.02	9.715E-04	4.941E-02
ENSG00000134333	LDHA	-0.12	0.97	0.97	1.037E-03	5.255E-02
ENSG00000122952	ZWINT	0.10	0.78	0.72	1.042E-03	5.256E-02
ENSG00000012124	CD22	0.00	0.04	0.07	1.058E-03	5.314E-02
ENSG00000167985	SDHAF2	0.07	0.75	0.72	1.063E-03	5.321E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000169288	MRPL1	0.09	0.79	0.75	1.073E-03	5.347E-02
ENSG00000125868	DSTN	-0.07	0.98	0.98	1.078E-03	5.353E-02
ENSG00000171792	RHNO1	0.06	0.43	0.35	1.091E-03	5.394E-02
ENSG00000143870	PDIA6	0.07	0.98	0.98	1.099E-03	5.412E-02
ENSG00000258920	FOXN3-AS1	0.04	0.20	0.13	1.111E-03	5.452E-02
ENSG00000117525	F3	-0.06	0.24	0.31	1.120E-03	5.473E-02
ENSG00000101384	JAG1	-0.07	0.23	0.30	1.137E-03	5.491E-02
ENSG00000130816	DNMT1	0.12	1.00	1.00	1.133E-03	5.491E-02
ENSG00000156299	TIAM1	-0.05	0.37	0.45	1.135E-03	5.491E-02
ENSG00000124380	SNRNP27	0.11	0.88	0.84	1.148E-03	5.525E-02
ENSG00000149262	INTS4	0.05	0.35	0.28	1.159E-03	5.535E-02
ENSG00000171345	KRT19	-0.10	1.00	1.00	1.159E-03	5.535E-02
ENSG00000171848	RRM2	0.14	0.84	0.77	1.187E-03	5.646E-02
ENSG00000136522	MRPL47	0.08	0.84	0.80	1.225E-03	5.803E-02
ENSG00000168610	STAT3	-0.12	0.81	0.86	1.229E-03	5.803E-02
ENSG00000184990	SIVA1	0.08	0.94	0.94	1.261E-03	5.929E-02
ENSG00000116698	SMG7	-0.13	0.69	0.74	1.276E-03	5.973E-02
ENSG00000127419	TMEM175	0.04	0.20	0.14	1.280E-03	5.973E-02
ENSG00000089127	OAS1	-0.07	0.15	0.21	1.290E-03	5.999E-02
ENSG00000182197	EXT1	-0.05	0.66	0.71	1.306E-03	6.048E-02
ENSG00000156011	PSD3	0.04	0.15	0.09	1.323E-03	6.107E-02
ENSG00000160298	C21orf58	0.12	0.80	0.71	1.330E-03	6.117E-02
ENSG00000131981	LGALS3	-0.11	0.95	0.94	1.346E-03	6.164E-02
ENSG00000124795	DEK	0.08	0.99	1.00	1.379E-03	6.273E-02
ENSG00000173141	MRPL57	0.09	0.96	0.94	1.379E-03	6.273E-02
ENSG00000179271	GADD45GIP1	0.07	0.97	0.97	1.391E-03	6.304E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000174775	HRAS	0.09	0.93	0.90	1.405E-03	6.344E-02
ENSG00000102974	CTCF	0.13	0.76	0.67	1.423E-03	6.388E-02
ENSG00000142208	AKT1	0.05	0.89	0.85	1.426E-03	6.388E-02
ENSG00000205581	HMGN1	0.05	0.99	1.00	1.430E-03	6.388E-02
ENSG00000162599	NFIA	0.09	0.64	0.57	1.447E-03	6.416E-02
ENSG00000251602		0.04	0.18	0.12	1.447E-03	6.416E-02
ENSG00000136098	NEK3	0.08	0.29	0.22	1.453E-03	6.419E-02
ENSG00000149212	SESN3	0.07	0.24	0.17	1.462E-03	6.438E-02
ENSG00000130340	SNX9	-0.05	0.44	0.50	1.472E-03	6.459E-02
ENSG00000136108	CKAP2	0.14	0.70	0.64	1.489E-03	6.501E-02
ENSG00000171346	KRT15	-0.18	0.88	0.89	1.494E-03	6.501E-02
ENSG00000204264	PSMB8	-0.10	0.82	0.82	1.497E-03	6.501E-02
ENSG00000106153	CHCHD2	-0.06	0.99	1.00	1.525E-03	6.596E-02
ENSG00000186847	KRT14	-0.24	0.36	0.44	1.543E-03	6.653E-02
ENSG00000165507	DEPP1	-0.04	0.18	0.25	1.549E-03	6.657E-02
ENSG00000105767	CADM4	0.07	0.45	0.38	1.601E-03	6.834E-02
ENSG00000171861	MRM3	0.06	0.36	0.30	1.600E-03	6.834E-02
ENSG00000188229	TUBB4B	0.10	0.99	0.98	1.643E-03	6.987E-02
ENSG00000119655	NPC2	-0.10	0.86	0.88	1.676E-03	7.102E-02
ENSG00000189266	PNRC2	0.12	0.80	0.74	1.710E-03	7.225E-02
ENSG00000071073	MGAT4A	0.08	0.52	0.46	1.718E-03	7.225E-02
ENSG00000154723	ATP5PF	0.08	0.97	0.96	1.728E-03	7.225E-02
ENSG00000253293	HOXA10	-0.03	0.09	0.14	1.722E-03	7.225E-02
ENSG00000114416	FXR1	0.11	0.86	0.85	1.753E-03	7.294E-02
ENSG00000167397	VKORC1	0.09	0.88	0.84	1.756E-03	7.294E-02
ENSG00000163882	POLR2H	0.08	0.89	0.88	1.771E-03	7.333E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000133131	MORC4	-0.09	0.46	0.53	1.809E-03	7.415E-02
ENSG00000151849	CENPJ	0.08	0.47	0.40	1.804E-03	7.415E-02
ENSG00000187608	ISG15	-0.24	0.57	0.59	1.809E-03	7.415E-02
ENSG00000175592	FOSL1	-0.06	0.64	0.71	1.823E-03	7.448E-02
ENSG00000155368	DBI	0.08	0.97	0.97	1.852E-03	7.542E-02
ENSG00000162614	NEXN	-0.06	0.29	0.35	1.899E-03	7.708E-02
ENSG00000250722	SELENOP	0.05	0.19	0.13	1.942E-03	7.857E-02
ENSG00000152795	HNRNPDL	0.08	0.98	0.97	1.956E-03	7.888E-02
ENSG00000170471	RALGAPB	0.03	0.34	0.26	1.979E-03	7.955E-02
ENSG00000137267	TUBB2A	-0.09	0.57	0.63	1.998E-03	8.001E-02
ENSG00000175449	RFESD	0.02	0.09	0.05	2.003E-03	8.001E-02
ENSG00000137312	FLOT1	-0.07	0.79	0.81	2.025E-03	8.060E-02
ENSG00000013810	TACC3	0.09	0.74	0.66	2.040E-03	8.095E-02
ENSG00000121864	ZNF639	0.08	0.43	0.36	2.073E-03	8.175E-02
ENSG00000237512	UNC5B-AS1	-0.06	0.23	0.30	2.071E-03	8.175E-02
ENSG00000075234	TTC38	-0.04	0.22	0.28	2.082E-03	8.185E-02
ENSG00000099785	MARCHF2	-0.05	0.21	0.27	2.097E-03	8.216E-02
ENSG00000111335	OAS2	-0.01	0.01	0.02	2.114E-03	8.257E-02
ENSG00000167272	POP5	0.07	0.86	0.83	2.136E-03	8.302E-02
ENSG00000247746	USP51	0.04	0.11	0.07	2.139E-03	8.302E-02
ENSG00000137699	TRIM29	0.10	0.66	0.61	2.188E-03	8.416E-02
ENSG00000163659	TIPARP	-0.07	0.24	0.30	2.182E-03	8.416E-02
ENSG00000188313	PLSCR1	-0.14	0.76	0.78	2.181E-03	8.416E-02
ENSG00000141644	MBD1	-0.06	0.37	0.43	2.255E-03	8.645E-02
ENSG00000134222	PSRC1	0.10	0.49	0.43	2.285E-03	8.729E-02
ENSG00000171617	ENC1	0.09	0.88	0.85	2.291E-03	8.729E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000197958	RPL12	-0.05	0.99	0.99	2.308E-03	8.769E-02
ENSG00000129187	DCTD	0.04	0.39	0.31	2.345E-03	8.826E-02
ENSG00000165215	CLDN3	0.07	0.78	0.72	2.352E-03	8.826E-02
ENSG00000262814	MRPL12	0.07	0.94	0.93	2.347E-03	8.826E-02
ENSG00000286522	H3C2	0.12	0.49	0.43	2.342E-03	8.826E-02
ENSG00000134874	DZIP1	0.08	0.57	0.51	2.373E-03	8.880E-02
ENSG00000170906	NDUFA3	0.09	0.93	0.88	2.397E-03	8.914E-02
ENSG00000175193	PARL	0.08	0.67	0.61	2.404E-03	8.914E-02
ENSG00000197429	IPP	0.06	0.20	0.15	2.398E-03	8.914E-02
ENSG00000198931	APRT	0.07	0.97	0.97	2.420E-03	8.949E-02
ENSG00000183684	ALYREF	0.08	0.94	0.93	2.432E-03	8.966E-02
ENSG00000184260	H2AC20	0.11	0.60	0.53	2.446E-03	8.990E-02
ENSG00000183734	ASCL2	0.03	0.11	0.06	2.465E-03	9.033E-02
ENSG00000042753	AP2S1	0.06	0.96	0.96	2.503E-03	9.066E-02
ENSG00000100813	ACIN1	0.10	0.87	0.83	2.484E-03	9.066E-02
ENSG00000106615	RHEB	-0.10	0.94	0.94	2.500E-03	9.066E-02
ENSG00000183049	CAMK1D	-0.08	0.47	0.54	2.499E-03	9.066E-02
ENSG00000135870	RC3H1	0.06	0.53	0.45	2.517E-03	9.089E-02
ENSG00000035141	FAM136A	0.08	0.92	0.91	2.540E-03	9.098E-02
ENSG00000064651	SLC12A2	-0.05	0.70	0.73	2.566E-03	9.098E-02
ENSG00000084774	CAD	0.05	0.71	0.64	2.554E-03	9.098E-02
ENSG00000109881	CCDC34	0.10	0.94	0.92	2.577E-03	9.098E-02
ENSG00000134882	UBAC2	0.11	0.71	0.64	2.555E-03	9.098E-02
ENSG00000145390	USP53	-0.12	0.53	0.58	2.573E-03	9.098E-02
ENSG00000168298	H1-4	0.12	0.64	0.57	2.573E-03	9.098E-02
ENSG00000226197		0.02	0.08	0.04	2.552E-03	9.098E-02

Table 6.1 continued from previous page

ENSEMBL Gene ID	HGNC Gene Symbol	Fold change (log2)	Afatinib tolerant cells (%)	Afatinib sensitive cells (%)	Student's t - test p - value	FDR
ENSG00000107164	FUBP3	0.07	0.59	0.51	2.608E-03	9.130E-02
ENSG00000159450	TCHH	-0.06	0.18	0.25	2.603E-03	9.130E-02
ENSG00000169062	UPF3A	0.05	0.96	0.97	2.603E-03	9.130E-02
ENSG00000147614	ATP6V0D2	-0.08	0.58	0.64	2.647E-03	9.214E-02
ENSG00000163491	NEK10	0.05	0.05	0.02	2.642E-03	9.214E-02
ENSG00000174705	SH3PXD2B	-0.08	0.46	0.52	2.699E-03	9.343E-02
ENSG00000254024		-0.02	0.06	0.10	2.699E-03	9.343E-02
ENSG00000196924	FLNA	-0.14	1.00	1.00	2.743E-03	9.469E-02
ENSG00000146425	DYNLT1	0.08	0.93	0.89	2.764E-03	9.515E-02
ENSG00000029364	SLC39A9	-0.04	0.26	0.33	2.811E-03	9.648E-02
ENSG00000160789	LMNA	-0.08	1.00	1.00	2.850E-03	9.705E-02
ENSG00000173928	SWSAP1	0.03	0.39	0.31	2.844E-03	9.705E-02
ENSG00000184207	PGP	0.09	0.89	0.86	2.850E-03	9.705E-02
ENSG00000179059	ZFP42	0.00	0.00	0.01	2.860E-03	9.710E-02
ENSG00000077152	UBE2T	0.11	0.90	0.87	2.886E-03	9.772E-02
ENSG00000229989	MIR181A1HG	0.06	0.15	0.10	2.903E-03	9.802E-02
ENSG00000138430	OLA1	-0.10	0.94	0.96	2.920E-03	9.833E-02
ENSG00000171320	ESCO2	0.05	0.41	0.33	2.964E-03	9.903E-02
ENSG00000177508	IRX3	0.13	0.93	0.90	2.964E-03	9.903E-02
ENSG00000225383	SFTA1P	0.08	0.63	0.58	2.951E-03	9.903E-02
ENSG00000116991	SIPA1L2	0.08	0.70	0.65	2.974E-03	9.910E-02
ENSG00000109089	CDR2L	-0.07	0.30	0.36	2.999E-03	9.966E-02

Table 6.2. GOEA for genes up-regulated in tolerant subpopulation in the untreated condition. **(A)** Term name **(B)** Percentage of up-regulated genes in tolerant population contained in the given term **(C)** Total number of genes in the given term **(D)** Population Hits represents the number of background population genes associated with the gene ontology term **(E)** Population Total is the number of background population genes **(F)** Fold Enrichment is the percentage of up-regulated genes in tolerant population belonging to given term, divided by the corresponding percentage in the background **(G)** p-value of Fisher's Exact of the enrichment **(H)** False Discovery Rate (FDR) correction of enrichment p-value for the given term.
Only significant (FDR < 0.1) are shown.

Term Name	% genes inters.	N° genes term	Pop Hits	Pop Total	Fold Enrich	Fisher's Exact p - value	FDR
GO:0006412 translation	5.48	193	223	19308	5.38	1.510e-5	5.674e-3
GO:0006334 nucleosome assembly	4.57	193	148	19308	6.76	1.760e-5	5.674e-3
hsa03010 Ribosome	4.57	100	158	8164	5.17	1.190e-4	1.157e-2
GO:0030261 chromosome condensation	2.28	193	23	19308	21.75	7.240e-5	1.866e-2
hsa05208 Chemical carcinogenesis reactive oxygen species	5.02	100	223	8164	4.03	3.510e-4	2.284e-2
hsa04714 Thermogenesis	5.02	100	232	8164	3.87	4.810e-4	2.345e-2
hsa05010 Alzheimer disease	6.39	100	384	8164	2.98	7.010e-4	2.475e-2
hsa05415 Diabetic cardiomyopathy	4.57	100	203	8164	4.02	7.620e-4	2.475e-2
GO:0015986 ATP synthesis coupled proton transport	2.28	193	27	19308	18.53	1.390e-4	2.987e-2
GO:0090267 positive regulation of mitotic cell cycle spindle assembly checkpoint	1.83	193	12	19308	33.35	1.990e-4	3.212e-2

Table 6.2 continued from previous page

Term Name	% genes inters.	N° genes term	Pop Hits	Pop Total	Fold Enrich	Fisher's Exact p - value	FDR
GO:0034501 protein localization to kinetochore	1.83	193	12	19308	33.35	1.990e-4	3.212e-2
hsa00190 Oxidative phosphorylation	3.65	100	134	8164	4.87	1.165e-3	3.246e-2
hsa05012 Parkinson disease	5.02	100	266	8164	3.38	1.379e-3	3.362e-2
GO:0007052 mitotic spindle organization	2.74	193	57	19308	10.53	2.540e-4	3.532e-2
GO:0007094 mitotic spindle assembly checkpoint	2.28	193	32	19308	15.63	2.740e-4	3.532e-2
hsa05020 Prion disease	5.02	100	273	8164	3.29	1.675e-3	3.629e-2
GO:0016584 nucleosome positioning	1.83	193	16	19308	25.01	4.930e-4	5.774e-2
GO:1901970 positive regulation of mitotic sister chromatid separation	1.37	193	4	19308	75.03	5.830e-4	6.257e-2
GO:0000278 mitotic cell cycle	3.65	193	154	19308	5.20	8.970e-4	7.798e-2
GO:0007059 chromosome segregation	2.74	193	76	19308	7.90	9.630e-4	7.798e-2
GO:1903490 positive regulation of mitotic cytokinesis	1.37	193	5	19308	60.02	9.650e-4	7.798e-2
GO:0007049 cell cycle	5.48	193	359	19308	3.34	9.680e-4	7.798e-2
hsa05014 Amyotrophic lateral sclerosis	5.48	100	364	8164	2.69	4.433e-3	8.644e-2

Table 6.3. GOEA for genes up-regulated in sensitive subpopulation in the untreated condition. **(A)** Term name **(B)** Percentage of up-regulated genes in tolerant population contained in the given term **(C)** Total number of genes in the given term **(D)** Population Hits represents the number of background population genes associated with the gene ontology term **(E)** Population Total is the number of background population genes **(F)** Fold Enrichment is the percentage of up-regulated genes in tolerant population belonging to given term, divided by the corresponding percentage in the background **(G)** p-value of Fisher's Exact of the enrichment **(H)** False Discovery Rate (FDR) correction of enrichment p-value for the given term.
Only significant (FDR < 0.1) are shown.

Term	% genes inters.	N° genes inters	Pop Hits	Pop Total	Fold Enrich	Fisher's Exact p-value	FDR
GO:0045109 intermediate filament organization	7.28	11	66	19308	23.15	3.755E-11	4.972E-08
hsa05169: Epstein-Barr virus infection	7.95	12	202	8164	6.06	3.362E-06	5.950E-04
GO:0098609 cell-cell adhesion	6.62	10	190	19308	7.31	9.212E-06	5.534E-03
GO:0009615 response to virus	5.30	8	111	19308	10.01	1.496E-05	5.534E-03
GO:0002486 antigen processing presentation endogenous antigen MHCI ER pathway, TAP-independent	2.65	4	8	19308	69.45	1.949E-05	5.534E-03
GO:0019882 antigen processing presentation	3.97	6	47	19308	17.73	2.090E-05	5.534E-03

Table 6.3 continued from previous page

Term	% genes inters.	N° genes inters	Pop Hits	Pop Total	Fold Enrich	Fisher's Exact p-value	FDR
hsa05163:Human cytomegalovirus infection	6.62	10	225	8164	4.54	3.007E-04	2.661E-02
GO:0048870 cell motility	3.31	5	38	19308	18.28	1.527E-04	3.370E-02
GO:0045071 negative regulation of viral genome replication	3.31	5	46	19308	15.10	3.230E-04	6.107E-02
GO:0045087 innate immune response	9.27	14	599	19308	3.25	3.690E-04	6.107E-02
GO:0045104 intermediate filament cytoskeleton organization	2.65	4	21	19308	26.46	4.324E-04	6.361E-02
GO:0030855 epithelial cell differentiation	3.97	6	94	19308	8.87	5.719E-04	7.572E-02
GO:0042270 protection from NK cell mediated cytotoxicity	1.99	3	6	19308	69.45	7.466E-04	8.237E-02
GO:0031581 hemidesmosome assembly	1.99	3	6	19308	69.45	7.466E-04	8.237E-02
hsa04915:Estrogen signaling pathway	4.64	7	138	8164	5.18	2.115E-03	8.812E-02
hsa05150: Staphylococcus aureus infection	3.97	6	96	8164	6.38	2.297E-03	8.812E-02

Table 6.3 continued from previous page

Term	% genes inters.	N° genes inters	Pop Hits	Pop Total	Fold Enrich	Fisher's Exact p-value	FDR
hsa05416: Viral myocarditis	3.31	5	60	8164	8.50	2.627E-03	8.812E-02
hsa04145:Phagosome	4.64	7	152	8164	4.70	3.430E-03	8.812E-02
hsa04218: Cellular senescence	4.64	7	156	8164	4.58	3.898E-03	8.812E-02
hsa04514:Cell adhesion molecules	4.64	7	157	8164	4.55	4.023E-03	8.812E-02
hsa05165:Human papillomavirus infection	6.62	10	331	8164	3.08	4.481E-03	8.812E-02
hsa05166: Human T-cell leukemia virus 1 infection	5.30	8	222	8164	3.68	5.514E-03	9.280E-02
hsa05330: Allograft rejection	2.65	4	38	8164	10.74	5.767E-03	9.280E-02
GO:0001916∼positive regulation of T cell mediated cytotoxicity	2.65	4	27	19308	20.58	9.215E-04	9.385E-02
hsa04612:Antigen processing presentation	3.31	5	78	8164	6.54	6.741E-03	9.943E-02

Table 6.4. Dose response curves of SUM cell lines. (A) Cell line name **(B)** Biological replicate ID **(C)** Technical replicate ID **(D)** Condition of the assay i.e., only vehicle (DMSO), background (bg) or Afatinib treatment. **(E)** Treatment concentration (μM) for Afatinib or % of DMSO evaluated **(F)** Lumicescence value"

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM159PT	1	1	vehicle(DMSO)	0.15%	16740
SUM159PT	1	2	vehicle(DMSO)	0.15%	19710
SUM159PT	1	3	vehicle(DMSO)	0.15%	22630
SUM159PT	1	1	afatinib	15	16160
SUM159PT	1	2	afatinib	15	20450
SUM159PT	1	3	afatinib	15	20500
SUM159PT	1	1	afatinib	10	17400
SUM159PT	1	2	afatinib	10	19730
SUM159PT	1	3	afatinib	10	20540
SUM159PT	1	1	afatinib	7.5	11420
SUM159PT	1	2	afatinib	7.5	15590
SUM159PT	1	3	afatinib	7.5	13420
SUM159PT	1	1	afatinib	5	2790
SUM159PT	1	2	afatinib	5	1800
SUM159PT	1	3	afatinib	5	1720
SUM159PT	1	1	afatinib	2.5	2430
SUM159PT	1	2	afatinib	2.5	2800
SUM159PT	1	3	afatinib	2.5	1400
SUM159PT	1	1	afatinib	1	3140
SUM159PT	1	2	afatinib	1	1980
SUM159PT	1	3	afatinib	1	450
SUM159PT	1	1	afatinib	0.1	2990
SUM159PT	1	2	afatinib	0.1	3731

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM159PT	1	3	afatinib	0.1	4291
SUM159PT	1	1	afatinib	0.01	4531
SUM159PT	1	2	afatinib	0.01	4301
SUM159PT	1	3	afatinib	0.01	5431
SUM159PT	1	1	bg	no	25
SUM159PT	1	2	bg	no	20
SUM159PT	1	3	bg	no	20
SUM159PT	2	1	vehicle(DMSO)	0.15%	6450
SUM159PT	2	2	vehicle(DMSO)	0.15%	6121
SUM159PT	2	3	vehicle(DMSO)	0.15%	5611
SUM159PT	2	1	afatinib	15	5201
SUM159PT	2	2	afatinib	15	5571
SUM159PT	2	3	afatinib	15	5601
SUM159PT	2	1	afatinib	10	4711
SUM159PT	2	2	afatinib	10	5281
SUM159PT	2	3	afatinib	10	5281
SUM159PT	2	1	afatinib	7.5	4861
SUM159PT	2	2	afatinib	7.5	4621
SUM159PT	2	3	afatinib	7.5	4521
SUM159PT	2	1	afatinib	5	1680
SUM159PT	2	2	afatinib	5	1610
SUM159PT	2	3	afatinib	5	1330
SUM159PT	2	1	afatinib	2.5	1050
SUM159PT	2	2	afatinib	2.5	1030
SUM159PT	2	3	afatinib	2.5	880
SUM159PT	2	1	afatinib	1	770

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM159PT	2	2	afatinib	1	910
SUM159PT	2	3	afatinib	1	850
SUM159PT	2	1	afatinib	0.1	320
SUM159PT	2	2	afatinib	0.1	340
SUM159PT	2	3	afatinib	0.1	380
SUM159PT	2	1	afatinib	0.01	1570
SUM159PT	2	2	afatinib	0.01	270
SUM159PT	2	3	afatinib	0.01	580
SUM159PT	2	1	bg	no	10
SUM159PT	2	2	bg	no	30
SUM159PT	2	3	bg	no	50
SUM159PT	3	1	vehicle(DMSO)	0.15%	148700
SUM159PT	3	2	vehicle(DMSO)	0.15%	222400
SUM159PT	3	3	vehicle(DMSO)	0.15%	248300
SUM159PT	3	1	afatinib	15	200300
SUM159PT	3	2	afatinib	15	209200
SUM159PT	3	3	afatinib	15	209000
SUM159PT	3	1	afatinib	10	200600
SUM159PT	3	2	afatinib	10	202600
SUM159PT	3	3	afatinib	10	182800
SUM159PT	3	1	afatinib	7.5	143800
SUM159PT	3	2	afatinib	7.5	136200
SUM159PT	3	3	afatinib	7.5	122100
SUM159PT	3	1	afatinib	5	54910
SUM159PT	3	2	afatinib	5	57110
SUM159PT	3	3	afatinib	5	50130

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM159PT	3	1	afatinib	2.5	8703
SUM159PT	3	2	afatinib	2.5	11200
SUM159PT	3	3	afatinib	2.5	6962
SUM159PT	3	1	afatinib	1	25970
SUM159PT	3	2	afatinib	1	31430
SUM159PT	3	3	afatinib	1	36060
SUM159PT	3	1	afatinib	0.1	27050
SUM159PT	3	2	afatinib	0.1	31030
SUM159PT	3	3	afatinib	0.1	29040
SUM159PT	3	1	afatinib	0.01	22740
SUM159PT	3	2	afatinib	0.01	21740
SUM159PT	3	3	afatinib	0.01	17790
SUM159PT	3	1	bg	no	75
SUM159PT	3	2	bg	no	110
SUM159PT	3	3	bg	no	70
SUM159PT	4	1	vehicle(DMSO)	0.15%	169320
SUM159PT	4	2	vehicle(DMSO)	0.15%	122100
SUM159PT	4	3	vehicle(DMSO)	0.15%	121600
SUM159PT	4	1	afatinib	15	140800
SUM159PT	4	2	afatinib	15	139300
SUM159PT	4	3	afatinib	15	137900
SUM159PT	4	1	afatinib	10	116800
SUM159PT	4	2	afatinib	10	124300
SUM159PT	4	3	afatinib	10	117600
SUM159PT	4	1	afatinib	7.5	70120
SUM159PT	4	2	afatinib	7.5	76660

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM159PT	4	3	afatinib	7.5	65360
SUM159PT	4	1	afatinib	5	24030
SUM159PT	4	2	afatinib	5	9514
SUM159PT	4	3	afatinib	5	16620
SUM159PT	4	1	afatinib	2.5	15150
SUM159PT	4	2	afatinib	2.5	27670
SUM159PT	4	3	afatinib	2.5	31880
SUM159PT	4	1	afatinib	1	24290
SUM159PT	4	2	afatinib	1	21910
SUM159PT	4	3	afatinib	1	21780
SUM159PT	4	1	afatinib	0.1	11910
SUM159PT	4	2	afatinib	0.1	15802
SUM159PT	4	3	afatinib	0.1	16712
SUM159PT	4	1	afatinib	0.01	22140
SUM159PT	4	2	afatinib	0.01	47100
SUM159PT	4	3	afatinib	0.01	45100
SUM159PT	4	1	bg	no	100
SUM159PT	4	2	bg	no	90
SUM159PT	4	3	bg	no	70
SUM225CWN	1	1	vehicle(DMSO)	0.15%	59500
SUM225CWN	1	2	vehicle(DMSO)	0.15%	54990
SUM225CWN	1	3	vehicle(DMSO)	0.15%	48370
SUM225CWN	1	1	afatinib	15	40790
SUM225CWN	1	2	afatinib	15	48000
SUM225CWN	1	3	afatinib	15	49430
SUM225CWN	1	1	afatinib	10	36040

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM225CWN	1	2	afatinib	10	26530
SUM225CWN	1	3	afatinib	10	26490
SUM225CWN	1	1	afatinib	7.5	21680
SUM225CWN	1	2	afatinib	7.5	22670
SUM225CWN	1	3	afatinib	7.5	25300
SUM225CWN	1	1	afatinib	5	19760
SUM225CWN	1	2	afatinib	5	18360
SUM225CWN	1	3	afatinib	5	19590
SUM225CWN	1	1	afatinib	2.5	19980
SUM225CWN	1	2	afatinib	2.5	19780
SUM225CWN	1	3	afatinib	2.5	18520
SUM225CWN	1	1	afatinib	1	19890
SUM225CWN	1	2	afatinib	1	14010
SUM225CWN	1	3	afatinib	1	13030
SUM225CWN	1	1	afatinib	0.1	18820
SUM225CWN	1	2	afatinib	0.1	9390
SUM225CWN	1	3	afatinib	0.1	12730
SUM225CWN	1	1	afatinib	0.01	8700
SUM225CWN	1	2	afatinib	0.01	8840
SUM225CWN	1	3	afatinib	0.01	7060
SUM225CWN	1	1	bg	no	50
SUM225CWN	1	2	bg	no	55
SUM225CWN	1	3	bg	no	70
SUM225CWN	2	1	vehicle(DMSO)	0.15%	1800
SUM225CWN	2	2	vehicle(DMSO)	0.15%	1950
SUM225CWN	2	3	vehicle(DMSO)	0.15%	2430

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM225CWN	2	1	afatinib	15	1940
SUM225CWN	2	2	afatinib	15	1840
SUM225CWN	2	3	afatinib	15	1060
SUM225CWN	2	1	afatinib	10	990
SUM225CWN	2	2	afatinib	10	1010
SUM225CWN	2	3	afatinib	10	990
SUM225CWN	2	1	afatinib	7.5	1230
SUM225CWN	2	2	afatinib	7.5	850
SUM225CWN	2	3	afatinib	7.5	680
SUM225CWN	2	1	afatinib	5	430
SUM225CWN	2	2	afatinib	5	410
SUM225CWN	2	3	afatinib	5	390
SUM225CWN	2	1	afatinib	2.5	350
SUM225CWN	2	2	afatinib	2.5	550
SUM225CWN	2	3	afatinib	2.5	460
SUM225CWN	2	1	afatinib	1	530
SUM225CWN	2	2	afatinib	1	470
SUM225CWN	2	3	afatinib	1	490
SUM225CWN	2	1	afatinib	0.1	440
SUM225CWN	2	2	afatinib	0.1	380
SUM225CWN	2	3	afatinib	0.1	370
SUM225CWN	2	1	afatinib	0.01	470
SUM225CWN	2	2	afatinib	0.01	480
SUM225CWN	2	3	afatinib	0.01	610
SUM225CWN	2	1	bg	no	20
SUM225CWN	2	2	bg	no	10

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM225CWN	2	3	bg	no	40
SUM225CWN	3	1	vehicle(DMSO)	0.15%	90360
SUM225CWN	3	2	vehicle(DMSO)	0.15%	97560
SUM225CWN	3	3	vehicle(DMSO)	0.15%	93470
SUM225CWN	3	1	afatinib	15	67420
SUM225CWN	3	2	afatinib	15	60230
SUM225CWN	3	3	afatinib	15	62080
SUM225CWN	3	1	afatinib	10	39620
SUM225CWN	3	2	afatinib	10	40120
SUM225CWN	3	3	afatinib	10	40530
SUM225CWN	3	1	afatinib	7.5	39450
SUM225CWN	3	2	afatinib	7.5	37630
SUM225CWN	3	3	afatinib	7.5	35830
SUM225CWN	3	1	afatinib	5	30610
SUM225CWN	3	2	afatinib	5	27950
SUM225CWN	3	3	afatinib	5	23620
SUM225CWN	3	1	afatinib	2.5	23670
SUM225CWN	3	2	afatinib	2.5	23840
SUM225CWN	3	3	afatinib	2.5	23640
SUM225CWN	3	1	afatinib	1	18130
SUM225CWN	3	2	afatinib	1	16730
SUM225CWN	3	3	afatinib	1	19410
SUM225CWN	3	1	afatinib	0.1	19200
SUM225CWN	3	2	afatinib	0.1	15540
SUM225CWN	3	3	afatinib	0.1	10370
SUM225CWN	3	1	afatinib	0.01	13450

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM225CWN	3	2	afatinib	0.01	13260
SUM225CWN	3	3	afatinib	0.01	16880
SUM225CWN	3	1	bg	no	150
SUM225CWN	3	2	bg	no	110
SUM225CWN	3	3	bg	no	110
SUM225CWN	4	1	vehicle(DMSO)	0.15%	39300
SUM225CWN	4	2	vehicle(DMSO)	0.15%	35990
SUM225CWN	4	3	vehicle(DMSO)	0.15%	34950
SUM225CWN	4	1	afatinib	15	74450
SUM225CWN	4	2	afatinib	15	19850
SUM225CWN	4	3	afatinib	15	6732
SUM225CWN	4	1	afatinib	10	14370
SUM225CWN	4	2	afatinib	10	15060
SUM225CWN	4	3	afatinib	10	13890
SUM225CWN	4	1	afatinib	7.5	12320
SUM225CWN	4	2	afatinib	7.5	11940
SUM225CWN	4	3	afatinib	7.5	11380
SUM225CWN	4	1	afatinib	5	5391
SUM225CWN	4	2	afatinib	5	7432
SUM225CWN	4	3	afatinib	5	8373
SUM225CWN	4	1	afatinib	2.5	10610
SUM225CWN	4	2	afatinib	2.5	11250
SUM225CWN	4	3	afatinib	2.5	10050
SUM225CWN	4	1	afatinib	1	5331
SUM225CWN	4	2	afatinib	1	5121
SUM225CWN	4	3	afatinib	1	5441

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM225CWN	4	1	afatinib	0.1	3771
SUM225CWN	4	2	afatinib	0.1	5721
SUM225CWN	4	3	afatinib	0.1	5421
SUM225CWN	4	1	afatinib	0.01	3711
SUM225CWN	4	2	afatinib	0.01	4111
SUM225CWN	4	3	afatinib	0.01	4031
SUM225CWN	4	1	bg	no	100
SUM225CWN	4	2	bg	no	80
SUM225CWN	4	3	bg	no	85
SUM149PT	1	1	vehicle(DMSO)	0.15%	10660
SUM149PT	1	2	vehicle(DMSO)	0.15%	9864
SUM149PT	1	3	vehicle(DMSO)	0.15%	10320
SUM149PT	1	1	afatinib	15	9432
SUM149PT	1	2	afatinib	15	9572
SUM149PT	1	3	afatinib	15	9132
SUM149PT	1	1	afatinib	10	8862
SUM149PT	1	2	afatinib	10	8002
SUM149PT	1	3	afatinib	10	9442
SUM149PT	1	1	afatinib	7.5	9481
SUM149PT	1	2	afatinib	7.5	7341
SUM149PT	1	3	afatinib	7.5	3220
SUM149PT	1	1	afatinib	5	2470
SUM149PT	1	2	afatinib	5	2720
SUM149PT	1	3	afatinib	5	2750
SUM149PT	1	1	afatinib	2.5	3101
SUM149PT	1	2	afatinib	2.5	2591

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM149PT	1	3	afatinib	2.5	2451
SUM149PT	1	1	afatinib	1	2270
SUM149PT	1	2	afatinib	1	2130
SUM149PT	1	3	afatinib	1	1560
SUM149PT	1	1	afatinib	0.1	1950
SUM149PT	1	2	afatinib	0.1	1650
SUM149PT	1	3	afatinib	0.1	1430
SUM149PT	1	1	afatinib	0.01	930
SUM149PT	1	2	afatinib	0.01	120
SUM149PT	1	3	afatinib	0.01	740
SUM149PT	1	1	bg	no	50
SUM149PT	1	2	bg	no	40
SUM149PT	1	3	bg	no	50
SUM149PT	2	1	vehicle(DMSO)	0.15%	34710
SUM149PT	2	2	vehicle(DMSO)	0.15%	31300
SUM149PT	2	3	vehicle(DMSO)	0.15%	33090
SUM149PT	2	1	afatinib	15	29870
SUM149PT	2	2	afatinib	15	35570
SUM149PT	2	3	afatinib	15	33480
SUM149PT	2	1	afatinib	10	31110
SUM149PT	2	2	afatinib	10	28610
SUM149PT	2	3	afatinib	10	11730
SUM149PT	2	1	afatinib	7.5	12510
SUM149PT	2	2	afatinib	7.5	15150
SUM149PT	2	3	afatinib	7.5	19030
SUM149PT	2	1	afatinib	5	1170

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM149PT	2	2	afatinib	5	1470
SUM149PT	2	3	afatinib	5	1180
SUM149PT	2	1	afatinib	2.5	1310
SUM149PT	2	2	afatinib	2.5	990
SUM149PT	2	3	afatinib	2.5	1190
SUM149PT	2	1	afatinib	1	1810
SUM149PT	2	2	afatinib	1	1720
SUM149PT	2	3	afatinib	1	2250
SUM149PT	2	1	afatinib	0.1	2650
SUM149PT	2	2	afatinib	0.1	660
SUM149PT	2	3	afatinib	0.1	2140
SUM149PT	2	1	afatinib	0.01	650
SUM149PT	2	2	afatinib	0.01	710
SUM149PT	2	3	afatinib	0.01	1240
SUM149PT	2	1	bg	no	20
SUM149PT	2	2	bg	no	20
SUM149PT	2	3	bg	no	10
SUM149PT	3	1	vehicle(DMSO)	0.15%	33800
SUM149PT	3	2	vehicle(DMSO)	0.15%	39230
SUM149PT	3	3	vehicle(DMSO)	0.15%	35170
SUM149PT	3	1	afatinib	15	36860
SUM149PT	3	2	afatinib	15	34970
SUM149PT	3	3	afatinib	15	38870
SUM149PT	3	1	afatinib	10	30930
SUM149PT	3	2	afatinib	10	37560
SUM149PT	3	3	afatinib	10	47710

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM149PT	3	1	afatinib	7.5	32280
SUM149PT	3	2	afatinib	7.5	33210
SUM149PT	3	3	afatinib	7.5	17130
SUM149PT	3	1	afatinib	5	5701
SUM149PT	3	2	afatinib	5	4621
SUM149PT	3	3	afatinib	5	4041
SUM149PT	3	1	afatinib	2.5	3290
SUM149PT	3	2	afatinib	2.5	2410
SUM149PT	3	3	afatinib	2.5	950
SUM149PT	3	1	afatinib	1	5852
SUM149PT	3	2	afatinib	1	7342
SUM149PT	3	3	afatinib	1	7322
SUM149PT	3	1	afatinib	0.1	7122
SUM149PT	3	2	afatinib	0.1	5451
SUM149PT	3	3	afatinib	0.1	7483
SUM149PT	3	1	afatinib	0.01	5101
SUM149PT	3	2	afatinib	0.01	6182
SUM149PT	3	3	afatinib	0.01	12290
SUM149PT	3	1	bg	no	100
SUM149PT	3	2	bg	no	90
SUM149PT	3	3	bg	no	90
SUM149PT	4	1	vehicle(DMSO)	0.15%	10064
SUM149PT	4	2	vehicle(DMSO)	0.15%	10600
SUM149PT	4	3	vehicle(DMSO)	0.15%	10380
SUM149PT	4	1	afatinib	15	12040
SUM149PT	4	2	afatinib	15	9211

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM149PT	4	3	afatinib	15	9483
SUM149PT	4	1	afatinib	10	9053
SUM149PT	4	2	afatinib	10	8833
SUM149PT	4	3	afatinib	10	8912
SUM149PT	4	1	afatinib	7.5	5741
SUM149PT	4	2	afatinib	7.5	5931
SUM149PT	4	3	afatinib	7.5	11680
SUM149PT	4	1	afatinib	5	4201
SUM149PT	4	2	afatinib	5	3801
SUM149PT	4	3	afatinib	5	4081
SUM149PT	4	1	afatinib	2.5	2940
SUM149PT	4	2	afatinib	2.5	2400
SUM149PT	4	3	afatinib	2.5	2620
SUM149PT	4	1	afatinib	1	2020
SUM149PT	4	2	afatinib	1	2690
SUM149PT	4	3	afatinib	1	3911
SUM149PT	4	1	afatinib	0.1	3701
SUM149PT	4	2	afatinib	0.1	4441
SUM149PT	4	3	afatinib	0.1	4201
SUM149PT	4	1	afatinib	0.01	3580
SUM149PT	4	2	afatinib	0.01	3870
SUM149PT	4	3	afatinib	0.01	3650
SUM149PT	4	1	bg	no	40
SUM149PT	4	2	bg	no	50
SUM149PT	4	3	bg	no	20
SUM131MO2	1	1	vehicle(DMSO)	0.15%	2290

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM131MO2	1	2	vehicle(DMSO)	0.15%	2640
SUM131MO2	1	3	vehicle(DMSO)	0.15%	3441
SUM131MO2	1	1	afatinib	15	2810
SUM131MO2	1	2	afatinib	15	2430
SUM131MO2	1	3	afatinib	15	2300
SUM131MO2	1	1	afatinib	10	1100
SUM131MO2	1	2	afatinib	10	2190
SUM131MO2	1	3	afatinib	10	2550
SUM131MO2	1	1	afatinib	7.5	1510
SUM131MO2	1	2	afatinib	7.5	2170
SUM131MO2	1	3	afatinib	7.5	1880
SUM131MO2	1	1	afatinib	5	1440
SUM131MO2	1	2	afatinib	5	90
SUM131MO2	1	3	afatinib	5	60
SUM131MO2	1	1	afatinib	2.5	70
SUM131MO2	1	2	afatinib	2.5	80
SUM131MO2	1	3	afatinib	2.5	60
SUM131MO2	1	1	afatinib	1	50
SUM131MO2	1	2	afatinib	1	40
SUM131MO2	1	3	afatinib	1	30
SUM131MO2	1	1	afatinib	0.1	20
SUM131MO2	1	2	afatinib	0.1	20
SUM131MO2	1	3	afatinib	0.1	50
SUM131MO2	1	1	afatinib	0.01	20
SUM131MO2	1	2	afatinib	0.01	10
SUM131MO2	1	3	afatinib	0.01	50

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM131MO2	1	1	bg	no	10
SUM131MO2	1	2	bg	no	20
SUM131MO2	1	3	bg	no	10
SUM131MO2	2	1	vehicle(DMSO)	0.15%	300
SUM131MO2	2	2	vehicle(DMSO)	0.15%	280
SUM131MO2	2	3	vehicle(DMSO)	0.15%	270
SUM131MO2	2	1	afatinib	15	210
SUM131MO2	2	2	afatinib	15	250
SUM131MO2	2	3	afatinib	15	250
SUM131MO2	2	1	afatinib	10	150
SUM131MO2	2	2	afatinib	10	200
SUM131MO2	2	3	afatinib	10	250
SUM131MO2	2	1	afatinib	7.5	310
SUM131MO2	2	2	afatinib	7.5	150
SUM131MO2	2	3	afatinib	7.5	140
SUM131MO2	2	1	afatinib	5	60
SUM131MO2	2	2	afatinib	5	50
SUM131MO2	2	3	afatinib	5	50
SUM131MO2	2	1	afatinib	2.5	20
SUM131MO2	2	2	afatinib	2.5	60
SUM131MO2	2	3	afatinib	2.5	70
SUM131MO2	2	1	afatinib	1	40
SUM131MO2	2	2	afatinib	1	60
SUM131MO2	2	3	afatinib	1	50
SUM131MO2	2	1	afatinib	0.1	70
SUM131MO2	2	2	afatinib	0.1	60

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM131MO2	2	3	afatinib	0.1	80
SUM131MO2	2	1	afatinib	0.01	20
SUM131MO2	2	2	afatinib	0.01	10
SUM131MO2	2	3	afatinib	0.01	30
SUM131MO2	2	1	bg	no	20
SUM131MO2	2	2	bg	no	10
SUM131MO2	2	3	bg	no	10
SUM131MO2	3	1	vehicle(DMSO)	0.15%	25350
SUM131MO2	3	2	vehicle(DMSO)	0.15%	35980
SUM131MO2	3	3	vehicle(DMSO)	0.15%	28730
SUM131MO2	3	1	afatinib	15	30220
SUM131MO2	3	2	afatinib	15	30470
SUM131MO2	3	3	afatinib	15	34180
SUM131MO2	3	1	afatinib	10	34440
SUM131MO2	3	2	afatinib	10	37180
SUM131MO2	3	3	afatinib	10	27680
SUM131MO2	3	1	afatinib	7.5	32530
SUM131MO2	3	2	afatinib	7.5	25120
SUM131MO2	3	3	afatinib	7.5	15660
SUM131MO2	3	1	afatinib	5	4751
SUM131MO2	3	2	afatinib	5	3761
SUM131MO2	3	3	afatinib	5	4911
SUM131MO2	3	1	afatinib	2.5	13030
SUM131MO2	3	2	afatinib	2.5	3551
SUM131MO2	3	3	afatinib	2.5	780
SUM131MO2	3	1	afatinib	1	4371

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM131MO2	3	2	afatinib	1	7292
SUM131MO2	3	3	afatinib	1	5531
SUM131MO2	3	1	afatinib	0.1	4551
SUM131MO2	3	2	afatinib	0.1	4821
SUM131MO2	3	3	afatinib	0.1	5972
SUM131MO2	3	1	afatinib	0.01	4801
SUM131MO2	3	2	afatinib	0.01	5491
SUM131MO2	3	3	afatinib	0.01	6622
SUM131MO2	3	1	bg	no	25
SUM131MO2	3	2	bg	no	20
SUM131MO2	3	3	bg	no	10
SUM229PE	1	1	vehicle(DMSO)	0.15%	1470
SUM229PE	1	2	vehicle(DMSO)	0.15%	1950
SUM229PE	1	3	vehicle(DMSO)	0.15%	1910
SUM229PE	1	1	afatinib	15	1990
SUM229PE	1	2	afatinib	15	1390
SUM229PE	1	3	afatinib	15	1950
SUM229PE	1	1	afatinib	10	1450
SUM229PE	1	2	afatinib	10	1780
SUM229PE	1	3	afatinib	10	270
SUM229PE	1	1	afatinib	7.5	330
SUM229PE	1	2	afatinib	7.5	430
SUM229PE	1	3	afatinib	7.5	250
SUM229PE	1	1	afatinib	5	180
SUM229PE	1	2	afatinib	5	110
SUM229PE	1	3	afatinib	5	150

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM229PE	1	1	afatinib	2.5	130
SUM229PE	1	2	afatinib	2.5	110
SUM229PE	1	3	afatinib	2.5	120
SUM229PE	1	1	afatinib	1	190
SUM229PE	1	2	afatinib	1	160
SUM229PE	1	3	afatinib	1	140
SUM229PE	1	1	afatinib	0.1	220
SUM229PE	1	2	afatinib	0.1	180
SUM229PE	1	3	afatinib	0.1	190
SUM229PE	1	1	afatinib	0.01	190
SUM229PE	1	2	afatinib	0.01	150
SUM229PE	1	3	afatinib	0.01	160
SUM229PE	1	1	bg	no	10
SUM229PE	1	2	bg	no	30
SUM229PE	1	3	bg	no	10
SUM229PE	2	1	vehicle(DMSO)	0.15%	2101
SUM229PE	2	2	vehicle(DMSO)	0.15%	2450
SUM229PE	2	3	vehicle(DMSO)	0.15%	2040
SUM229PE	2	1	afatinib	15	2440
SUM229PE	2	2	afatinib	15	2780
SUM229PE	2	3	afatinib	15	2830
SUM229PE	2	1	afatinib	10	2750
SUM229PE	2	2	afatinib	10	2490
SUM229PE	2	3	afatinib	10	2260
SUM229PE	2	1	afatinib	7.5	330
SUM229PE	2	2	afatinib	7.5	590

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM229PE	2	3	afatinib	7.5	550
SUM229PE	2	1	afatinib	5	150
SUM229PE	2	2	afatinib	5	550
SUM229PE	2	3	afatinib	5	230
SUM229PE	2	1	afatinib	2.5	540
SUM229PE	2	2	afatinib	2.5	450
SUM229PE	2	3	afatinib	2.5	430
SUM229PE	2	1	afatinib	1	450
SUM229PE	2	2	afatinib	1	460
SUM229PE	2	3	afatinib	1	480
SUM229PE	2	1	afatinib	0.1	310
SUM229PE	2	2	afatinib	0.1	230
SUM229PE	2	3	afatinib	0.1	250
SUM229PE	2	1	afatinib	0.01	210
SUM229PE	2	2	afatinib	0.01	170
SUM229PE	2	3	afatinib	0.01	150
SUM229PE	2	1	bg	no	20
SUM229PE	2	2	bg	no	20
SUM229PE	2	3	bg	no	30
SUM229PE	3	1	vehicle(DMSO)	0.15%	4101
SUM229PE	3	2	vehicle(DMSO)	0.15%	4002
SUM229PE	3	3	vehicle(DMSO)	0.15%	4031
SUM229PE	3	1	afatinib	15	4431
SUM229PE	3	2	afatinib	15	3441
SUM229PE	3	3	afatinib	15	4391
SUM229PE	3	1	afatinib	10	4801

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM229PE	3	2	afatinib	10	4231
SUM229PE	3	3	afatinib	10	3621
SUM229PE	3	1	afatinib	7.5	3250
SUM229PE	3	2	afatinib	7.5	3260
SUM229PE	3	3	afatinib	7.5	3170
SUM229PE	3	1	afatinib	5	3170
SUM229PE	3	2	afatinib	5	3340
SUM229PE	3	3	afatinib	5	3010
SUM229PE	3	1	afatinib	2.5	1610
SUM229PE	3	2	afatinib	2.5	1260
SUM229PE	3	3	afatinib	2.5	1380
SUM229PE	3	1	afatinib	1	1410
SUM229PE	3	2	afatinib	1	1640
SUM229PE	3	3	afatinib	1	1410
SUM229PE	3	1	afatinib	0.1	1980
SUM229PE	3	2	afatinib	0.1	1160
SUM229PE	3	3	afatinib	0.1	2270
SUM229PE	3	1	afatinib	0.01	1090
SUM229PE	3	2	afatinib	0.01	1140
SUM229PE	3	3	afatinib	0.01	1230
SUM229PE	3	1	bg	no	30
SUM229PE	3	2	bg	no	10
SUM229PE	3	3	bg	no	50
SUM185PE	1	1	vehicle(DMSO)	0.15%	4710
SUM185PE	1	2	vehicle(DMSO)	0.15%	4550
SUM185PE	1	3	vehicle(DMSO)	0.15%	4854

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM185PE	1	1	afatinib	15	4651
SUM185PE	1	2	afatinib	15	4581
SUM185PE	1	3	afatinib	15	4761
SUM185PE	1	1	afatinib	10	4031
SUM185PE	1	2	afatinib	10	3951
SUM185PE	1	3	afatinib	10	3681
SUM185PE	1	1	afatinib	7.5	2980
SUM185PE	1	2	afatinib	7.5	2930
SUM185PE	1	3	afatinib	7.5	3070
SUM185PE	1	1	afatinib	5	2480
SUM185PE	1	2	afatinib	5	2490
SUM185PE	1	3	afatinib	5	2360
SUM185PE	1	1	afatinib	2.5	2550
SUM185PE	1	2	afatinib	2.5	2340
SUM185PE	1	3	afatinib	2.5	2180
SUM185PE	1	1	afatinib	1	1580
SUM185PE	1	2	afatinib	1	1340
SUM185PE	1	3	afatinib	1	1420
SUM185PE	1	1	afatinib	0.1	1640
SUM185PE	1	2	afatinib	0.1	1710
SUM185PE	1	3	afatinib	0.1	1260
SUM185PE	1	1	afatinib	0.01	1610
SUM185PE	1	2	afatinib	0.01	1440
SUM185PE	1	3	afatinib	0.01	1540
SUM185PE	1	1	bg	no	20
SUM185PE	1	2	bg	no	10

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM185PE	1	3	bg	no	50
SUM185PE	2	1	vehicle(DMSO)	0.15%	500
SUM185PE	2	2	vehicle(DMSO)	0.15%	580
SUM185PE	2	3	vehicle(DMSO)	0.15%	480
SUM185PE	2	1	afatinib	15	410
SUM185PE	2	2	afatinib	15	470
SUM185PE	2	3	afatinib	15	460
SUM185PE	2	1	afatinib	10	450
SUM185PE	2	2	afatinib	10	400
SUM185PE	2	3	afatinib	10	420
SUM185PE	2	1	afatinib	7.5	420
SUM185PE	2	2	afatinib	7.5	440
SUM185PE	2	3	afatinib	7.5	400
SUM185PE	2	1	afatinib	5	200
SUM185PE	2	2	afatinib	5	110
SUM185PE	2	3	afatinib	5	150
SUM185PE	2	1	afatinib	2.5	110
SUM185PE	2	2	afatinib	2.5	160
SUM185PE	2	3	afatinib	2.5	170
SUM185PE	2	1	afatinib	1	40
SUM185PE	2	2	afatinib	1	60
SUM185PE	2	3	afatinib	1	50
SUM185PE	2	1	afatinib	0.1	70
SUM185PE	2	2	afatinib	0.1	60
SUM185PE	2	3	afatinib	0.1	80
SUM185PE	2	1	afatinib	0.01	20

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM185PE	2	2	afatinib	0.01	10
SUM185PE	2	3	afatinib	0.01	30
SUM185PE	2	1	bg	no	10
SUM185PE	2	2	bg	no	10
SUM185PE	2	3	bg	no	20
SUM185PE	3	1	vehicle(DMSO)	0.15%	510
SUM185PE	3	2	vehicle(DMSO)	0.15%	490
SUM185PE	3	3	vehicle(DMSO)	0.15%	500
SUM185PE	3	1	afatinib	15	510
SUM185PE	3	2	afatinib	15	450
SUM185PE	3	3	afatinib	15	490
SUM185PE	3	1	afatinib	10	390
SUM185PE	3	2	afatinib	10	370
SUM185PE	3	3	afatinib	10	400
SUM185PE	3	1	afatinib	7.5	410
SUM185PE	3	2	afatinib	7.5	450
SUM185PE	3	3	afatinib	7.5	370
SUM185PE	3	1	afatinib	5	390
SUM185PE	3	2	afatinib	5	300
SUM185PE	3	3	afatinib	5	80
SUM185PE	3	1	afatinib	2.5	80
SUM185PE	3	2	afatinib	2.5	80
SUM185PE	3	3	afatinib	2.5	80
SUM185PE	3	1	afatinib	1	60
SUM185PE	3	2	afatinib	1	70
SUM185PE	3	3	afatinib	1	30

Table 6.4 continued from previous page

CCL	experiment	replicate	Condition	Afatinib[uM] or DMSO[%]	Raw Luminescence Value
SUM185PE	3	1	afatinib	0.1	20
SUM185PE	3	2	afatinib	0.1	30
SUM185PE	3	3	afatinib	0.1	40
SUM185PE	3	1	afatinib	0.01	20
SUM185PE	3	2	afatinib	0.01	20
SUM185PE	3	3	afatinib	0.01	20
SUM185PE	3	1	bg	no	10
SUM185PE	3	2	bg	no	20
SUM185PE	3	3	bg	no	10

Table 6.5. Afatinib sensitivity prediction results of scASTRAL. (A) Cell line name **(B)** Log2 IC_50 of the cell line **(C)** Percentage of cells predicted to be sensitive to Afatinib by scASTRAL **(D)** Source of the IC_50 value which could or GDSC [7] or the estimated one.

CCL	Log2 IC50	% sensitive cells	source IC50
SUM1315MO2	0.411	0.49	estimated
SUM149PT	0.275	0.55	estimated
SUM159PT	0.310	0.58	estimated
SUM185PE	1.491	0.40	estimated
SUM22PE	-0.272	0.31	estimated
BT20	3.752	0.86	GDSC
BT549	2.942	0.99	GDSC
CAL51	3.564	0.97	GDSC
CAL851	3.615	0.95	GDSC
HCC1143	2.562	0.80	GDSC

Table 6.5 continued from previous page

CCL	Log2 IC50	% sensitive cells	source IC50
HCC1187	3.051	1.00	GDSC
HCC1937	3.409	0.99	GDSC
HCC38	5.339	0.97	GDSC
HCC70	3.121	1.00	GDSC
HDQP1	0.579	0.73	GDSC
MDAMB436	6.276	0.99	GDSC

Table 6.6. Clones (i.e., lineage) frequencies over time. (A) Lineage (clone) ID **(B)** Frequency of the given lineage in the control condition (Day 0) b At 3 days **(D)** At 6 days **(E)** At 9 days **(F)** Afatinib response group of given lineage as Tolerant or Sensitive

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-021:bc30-021932	2.206	3.340	4.034	2.809	Tolerant
bc14-037:bc30-021892	4.348	1.461	2.017	2.019	Tolerant
bc14-012:bc30-099590	1.168	0.835	1.274	1.405	Tolerant
bc14-021:bc30-048793	2.401	2.296	2.123	1.405	Tolerant
bc14-054:bc30-099259	2.206	2.714	2.335	1.317	Tolerant
bc14-006:bc30-019687	0.844	0.626	0.531	0.966	Tolerant
bc14-061:bc30-092152	0.584	1.253	0.531	0.966	Tolerant
bc14-011:bc30-090251	0.973	0.209	1.380	0.878	Tolerant
bc14-060:bc30-072759	0.649	0.209	0.955	0.790	Tolerant
bc14-065:bc30-030058	0.714	3.132	0.637	0.702	Tolerant
bc14-043:bc30-034364	0.714	1.044	0.531	0.615	Tolerant
bc14-064:bc30-086398	0.584	0.626	0.849	0.615	Tolerant
bc14-065:bc30-029505	0.973	2.088	0.531	0.615	Tolerant

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-096:bc30-014468	0.779	0.209	0.849	0.615	Tolerant
bc14-052:bc30-005120	0.389	0.835	1.380	0.527	Tolerant
bc14-052:bc30-029064	0.714	1.044	0.106	0.527	Tolerant
bc14-087:bc30-061352	0.519	0.209	0.212	0.527	Tolerant
bc14-022:bc30-047533	0.324	0.209	0.425	0.439	Tolerant
bc14-059:bc30-072335	0.324	0.209	0.531	0.439	Tolerant
bc14-019:bc30-041430	0.324	#N/A	0.637	0.351	Tolerant
bc14-029:bc30-078291	0.324	0.209	0.212	0.351	Tolerant
bc14-039:bc30-051511	0.324	0.418	0.743	0.351	Tolerant
bc14-065:bc30-057163	0.195	0.418	0.531	0.351	Tolerant
bc14-020:bc30-010913	0.649	0.418	0.425	0.263	Tolerant
bc14-033:bc30-023121	0.195	0.418	0.212	0.263	Tolerant
bc14-055:bc30-057547	0.454	0.418	0.318	0.263	Tolerant
bc14-061:bc30-015052	0.260	#N/A	0.106	0.263	Tolerant
bc14-063:bc30-070885	0.519	1.044	0.318	0.263	Tolerant
bc14-003:bc30-090195	0.389	0.418	0.425	0.176	Tolerant
bc14-006:bc30-066015	0.454	#N/A	0.425	0.176	Tolerant
bc14-006:bc30-071388	0.260	#N/A	0.106	0.176	Tolerant
bc14-018:bc30-066885	0.195	#N/A	0.106	0.176	Tolerant
bc14-022:bc30-069303	0.454	0.209	#N/A	0.176	Tolerant
bc14-037:bc30-094922	0.195	#N/A	#N/A	0.176	Tolerant
bc14-061:bc30-007079	0.260	#N/A	0.212	0.176	Tolerant
bc14-065:bc30-033306	0.324	0.209	0.318	0.176	Tolerant
bc14-071:bc30-020429	0.584	0.209	0.425	0.176	Tolerant
bc14-086:bc30-044324	0.519	0.209	0.106	0.176	Tolerant
bc14-095:bc30-012872	0.260	0.209	0.212	0.176	Tolerant

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-001:bc30-037375	0.065	#N/A	#N/A	0.088	Tolerant
bc14-007:bc30-038130	0.130	#N/A	#N/A	0.088	Tolerant
bc14-015:bc30-078014	0.065	#N/A	#N/A	0.088	Tolerant
bc14-024:bc30-096023	0.065	0.209	0.425	0.088	Tolerant
bc14-029:bc30-044951	0.130	0.209	#N/A	0.088	Tolerant
bc14-037:bc30-069628	0.065	0.418	#N/A	0.088	Tolerant
bc14-045:bc30-024511	0.324	#N/A	#N/A	0.088	Tolerant
bc14-050:bc30-014578	0.065	0.209	#N/A	0.088	Tolerant
bc14-055:bc30-073573	0.260	0.418	0.425	0.088	Tolerant
bc14-058:bc30-071926	0.065	#N/A	0.106	0.088	Tolerant
bc14-059:bc30-086131	0.065	#N/A	0.106	0.088	Tolerant
bc14-061:bc30-079696	0.195	#N/A	0.106	0.088	Tolerant
bc14-065:bc30-098517	0.065	0.209	0.106	0.088	Tolerant
bc14-069:bc30-027116	0.260	#N/A	#N/A	0.088	Tolerant
bc14-083:bc30-009620	0.324	0.626	0.106	0.088	Tolerant
bc14-087:bc30-065831	0.130	0.418	0.106	0.088	Tolerant
bc14-095:bc30-033735	0.065	#N/A	#N/A	0.088	Tolerant
bc14-095:bc30-059525	0.195	#N/A	0.212	0.088	Tolerant
bc14-013:bc30-092942	14.471	21.503	32.696	35.470	Tolerant
bc14-065:bc30-046216	2.466	6.472	7.749	9.921	Tolerant
bc14-029:bc30-097278	0.649	1.670	0.849	2.019	Tolerant
bc14-057:bc30-088917	0.519	0.418	1.911	1.141	Tolerant
bc14-040:bc30-048917	0.324	0.209	0.849	1.054	Tolerant
bc14-002:bc30-074580	0.389	0.835	0.955	0.878	Tolerant
bc14-063:bc30-058638	0.260	0.209	0.531	0.615	Tolerant
bc14-040:bc30-018935	0.130	#N/A	0.425	0.439	Tolerant

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-051:bc30-000703	0.195	0.418	0.318	0.439	Tolerant
bc14-029:bc30-007942	0.065	#N/A	#N/A	0.263	Tolerant
bc14-038:bc30-044710	0.065	#N/A	#N/A	0.263	Tolerant
bc14-099:bc30-079659	0.065	#N/A	0.212	0.263	Tolerant
bc14-065:bc30-087150	0.065	#N/A	0.106	0.176	Tolerant
bc14-001:bc30-003651	0.065	0.209	#N/A	#N/A	Tolerant
bc14-001:bc30-050813	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-001:bc30-073725	0.195	#N/A	#N/A	#N/A	Sensitive
bc14-002:bc30-005546	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-002:bc30-018515	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-002:bc30-023232	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-002:bc30-037900	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-002:bc30-045329	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-002:bc30-051804	0.195	0.209	0.212	#N/A	Sensitive
bc14-002:bc30-059123	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-002:bc30-097262	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-003:bc30-006836	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-003:bc30-018637	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-003:bc30-023649	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-003:bc30-060369	#N/A	0.209	0.106	#N/A	Sensitive
bc14-004:bc30-003418	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-004:bc30-008394	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-004:bc30-027226	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-004:bc30-031423	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-004:bc30-071961	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-004:bc30-087344	0.065	#N/A	#N/A	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-005:bc30-009892	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-006:bc30-030371	0.195	#N/A	0.106	#N/A	Sensitive
bc14-006:bc30-067911	0.065	0.209	#N/A	#N/A	Sensitive
bc14-006:bc30-085689	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-007:bc30-000605	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-007:bc30-007531	0.065	#N/A	0.106	#N/A	Sensitive
bc14-008:bc30-004652	0.195	#N/A	#N/A	#N/A	Sensitive
bc14-008:bc30-027839	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-008:bc30-059667	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-008:bc30-062269	0.065	0.209	#N/A	#N/A	Sensitive
bc14-010:bc30-004435	0.065	0.209	#N/A	#N/A	Sensitive
bc14-010:bc30-007792	0.065	0.209	#N/A	#N/A	Sensitive
bc14-010:bc30-043637	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-010:bc30-099753	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-011:bc30-059389	0.065	0.209	#N/A	#N/A	Sensitive
bc14-011:bc30-061613	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-011:bc30-062136	0.130	0.209	0.106	#N/A	Tolerant
bc14-011:bc30-062576	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-011:bc30-069933	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-011:bc30-080773	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-011:bc30-092844	0.130	#N/A	#N/A	#N/A	Tolerant
bc14-011:bc30-096594	#N/A	0.209	0.106	#N/A	Sensitive
bc14-011:bc30-099377	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-012:bc30-026933	0.324	#N/A	0.212	#N/A	Sensitive
bc14-012:bc30-027008	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-012:bc30-035872	#N/A	#N/A	0.106	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-012:bc30-068431	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-012:bc30-080160	#N/A	#N/A	#N/A	#N/A	Tolerant
bc14-012:bc30-083558	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-013:bc30-009766	0.130	0.209	#N/A	#N/A	Tolerant
bc14-013:bc30-018603	0.065	0.209	#N/A	#N/A	Sensitive
bc14-013:bc30-028407	0.195	0.418	#N/A	#N/A	Sensitive
bc14-013:bc30-085462	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-013:bc30-093973	#N/A	#N/A	0.106	#N/A	Tolerant
bc14-014:bc30-000356	0.130	#N/A	0.212	#N/A	Sensitive
bc14-014:bc30-064469	0.844	#N/A	0.106	#N/A	Sensitive
bc14-014:bc30-095687	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-015:bc30-045353	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-015:bc30-060120	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-015:bc30-066453	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-015:bc30-077167	0.065	0.209	#N/A	#N/A	Sensitive
bc14-015:bc30-079242	0.065	0.209	0.106	#N/A	Sensitive
bc14-015:bc30-081240	0.130	0.209	0.106	#N/A	Sensitive
bc14-015:bc30-091592	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-016:bc30-007195	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-017:bc30-037177	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-018:bc30-015072	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-018:bc30-028041	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-018:bc30-029572	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-018:bc30-034063	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-018:bc30-054089	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-018:bc30-065781	0.065	#N/A	#N/A	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-019:bc30-065627	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-019:bc30-068533	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-020:bc30-079030	0.065	#N/A	0.212	#N/A	Sensitive
bc14-021:bc30-010896	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-021:bc30-024916	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-021:bc30-031240	0.130	#N/A	#N/A	#N/A	Tolerant
bc14-021:bc30-046704	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-021:bc30-053172	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-021:bc30-061643	0.065	0.209	#N/A	#N/A	Sensitive
bc14-021:bc30-076957	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-021:bc30-089309	0.130	#N/A	0.106	#N/A	Sensitive
bc14-021:bc30-099996	0.130	0.418	#N/A	#N/A	Sensitive
bc14-022:bc30-029811	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-022:bc30-032129	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-022:bc30-032693	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-022:bc30-034018	0.130	0.209	#N/A	#N/A	Sensitive
bc14-022:bc30-059089	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-022:bc30-059265	0.130	#N/A	0.212	#N/A	Sensitive
bc14-022:bc30-063862	#N/A	#N/A	#N/A	#N/A	Tolerant
bc14-022:bc30-067450	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-022:bc30-098837	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-029:bc30-001138	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-029:bc30-001504	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-029:bc30-002160	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-029:bc30-005813	0.195	0.209	0.212	#N/A	Sensitive
bc14-029:bc30-018914	0.130	#N/A	#N/A	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-029:bc30-022186	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-029:bc30-052996	0.195	#N/A	#N/A	#N/A	Tolerant
bc14-029:bc30-072575	0.065	0.418	#N/A	#N/A	Sensitive
bc14-029:bc30-076836	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-029:bc30-078747	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-029:bc30-091740	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-030:bc30-014189	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-030:bc30-079766	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-032:bc30-088851	0.130	0.209	#N/A	#N/A	Sensitive
bc14-033:bc30-023722	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-033:bc30-036686	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-033:bc30-088008	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-034:bc30-025803	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-034:bc30-065305	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-034:bc30-067943	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-034:bc30-088466	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-034:bc30-094912	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-036:bc30-011076	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-036:bc30-013232	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-036:bc30-033544	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-036:bc30-043462	0.260	#N/A	#N/A	#N/A	Sensitive
bc14-036:bc30-045775	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-036:bc30-055776	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-036:bc30-055997	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-036:bc30-072252	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-036:bc30-074269	0.130	#N/A	#N/A	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-037:bc30-006301	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-009106	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-009931	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-015631	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-018534	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-037:bc30-033005	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-034693	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-038326	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-039971	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-037:bc30-047211	0.130	#N/A	#N/A	#N/A	Tolerant
bc14-037:bc30-048607	0.065	#N/A	0.106	#N/A	Sensitive
bc14-037:bc30-064907	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-065837	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-078298	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-080081	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-037:bc30-081298	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-037:bc30-086504	0.065	#N/A	#N/A	#N/A	Tolerant
bc14-038:bc30-004062	#N/A	0.209	#N/A	#N/A	Tolerant
bc14-038:bc30-022777	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-038:bc30-025776	0.260	#N/A	#N/A	#N/A	Sensitive
bc14-038:bc30-043693	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-038:bc30-055830	0.195	#N/A	#N/A	#N/A	Sensitive
bc14-038:bc30-057819	0.065	0.209	#N/A	#N/A	Sensitive
bc14-038:bc30-059241	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-039:bc30-005606	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-039:bc30-067041	0.260	#N/A	#N/A	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-039:bc30-071070	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-039:bc30-071573	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-039:bc30-074798	0.324	#N/A	0.212	#N/A	Tolerant
bc14-039:bc30-085698	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-039:bc30-086156	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-039:bc30-097361	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-040:bc30-041540	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-040:bc30-052198	0.130	#N/A	#N/A	#N/A	Tolerant
bc14-040:bc30-077147	0.130	#N/A	#N/A	#N/A	Tolerant
bc14-040:bc30-077164	0.260	#N/A	#N/A	#N/A	Sensitive
bc14-040:bc30-086653	#N/A	0.209	0.106	#N/A	Sensitive
bc14-040:bc30-087074	0.324	0.209	#N/A	#N/A	Sensitive
bc14-041:bc30-019652	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-041:bc30-024244	0.065	0.209	#N/A	#N/A	Sensitive
bc14-041:bc30-028986	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-041:bc30-040419	0.195	#N/A	#N/A	#N/A	Sensitive
bc14-043:bc30-089487	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-044:bc30-024918	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-045:bc30-055277	0.130	#N/A	0.106	#N/A	Sensitive
bc14-045:bc30-082359	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-046:bc30-017426	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-046:bc30-047148	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-046:bc30-059550	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-046:bc30-085341	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-046:bc30-088001	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-047:bc30-092195	#N/A	#N/A	0.106	#N/A	Tolerant

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-049:bc30-005445	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-050:bc30-038125	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-050:bc30-069935	#N/A	#N/A	0.106	#N/A	Tolerant
bc14-051:bc30-024824	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-051:bc30-041241	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-051:bc30-045497	0.260	0.209	0.106	#N/A	Sensitive
bc14-051:bc30-059824	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-051:bc30-059884	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-051:bc30-087661	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-051:bc30-094255	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-051:bc30-097150	0.195	0.209	#N/A	#N/A	Sensitive
bc14-052:bc30-043614	0.065	#N/A	#N/A	#N/A	Tolerant
bc14-052:bc30-071932	0.065	#N/A	0.106	#N/A	Sensitive
bc14-052:bc30-072418	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-052:bc30-084859	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-053:bc30-014960	#N/A	#N/A	0.212	#N/A	Sensitive
bc14-053:bc30-047684	0.065	#N/A	#N/A	#N/A	Tolerant
bc14-054:bc30-019543	0.324	0.209	0.106	#N/A	Sensitive
bc14-054:bc30-040931	0.454	0.209	0.106	#N/A	Sensitive
bc14-054:bc30-051394	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-054:bc30-053680	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-054:bc30-056461	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-054:bc30-058015	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-054:bc30-063716	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-054:bc30-067588	0.065	#N/A	0.106	#N/A	Tolerant
bc14-054:bc30-067885	0.130	#N/A	#N/A	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-054:bc30-077234	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-054:bc30-083274	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-054:bc30-091290	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-054:bc30-095388	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-001342	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-055:bc30-003559	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-007437	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-055:bc30-017992	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-024585	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-029529	0.389	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-037305	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-041960	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-056904	0.065	#N/A	0.106	#N/A	Tolerant
bc14-055:bc30-056969	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-059348	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-059564	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-064685	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-066496	0.065	#N/A	0.106	#N/A	Tolerant
bc14-055:bc30-068091	0.130	#N/A	#N/A	#N/A	Tolerant
bc14-055:bc30-069491	0.260	#N/A	#N/A	#N/A	Sensitive
bc14-055:bc30-082544	0.065	0.209	#N/A	#N/A	Tolerant
bc14-055:bc30-083868	0.324	0.418	0.212	#N/A	Sensitive
bc14-055:bc30-091658	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-056:bc30-031480	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-056:bc30-054790	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-057:bc30-011210	0.065	0.209	0.106	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-057:bc30-072952	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-058:bc30-006803	0.065	#N/A	0.106	#N/A	Sensitive
bc14-058:bc30-010204	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-058:bc30-035786	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-058:bc30-037894	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-058:bc30-054741	0.065	0.209	#N/A	#N/A	Sensitive
bc14-058:bc30-057807	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-059:bc30-055555	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-061:bc30-003529	0.065	#N/A	0.106	#N/A	Sensitive
bc14-061:bc30-008845	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-061:bc30-013855	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-061:bc30-026269	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-061:bc30-028070	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-061:bc30-035108	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-061:bc30-050888	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-061:bc30-098212	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-062:bc30-048318	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-062:bc30-082235	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-063:bc30-041979	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-063:bc30-080432	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-064:bc30-053728	0.260	0.418	0.318	#N/A	Sensitive
bc14-065:bc30-026474	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-065:bc30-033004	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-065:bc30-051915	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-065:bc30-067976	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-066:bc30-020663	0.130	#N/A	#N/A	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-067:bc30-077257	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-067:bc30-081937	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-067:bc30-086683	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-069:bc30-016428	0.649	#N/A	#N/A	#N/A	Sensitive
bc14-069:bc30-036996	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-069:bc30-061227	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-069:bc30-062798	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-069:bc30-075416	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-069:bc30-082362	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-069:bc30-091431	0.065	#N/A	0.212	#N/A	Tolerant
bc14-069:bc30-098271	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-069:bc30-098454	0.065	0.209	0.106	#N/A	Sensitive
bc14-071:bc30-012516	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-071:bc30-014545	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-071:bc30-030902	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-071:bc30-042924	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-071:bc30-061562	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-071:bc30-062014	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-071:bc30-075980	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-071:bc30-080963	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-071:bc30-083356	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-071:bc30-086570	0.065	0.418	0.106	#N/A	Sensitive
bc14-073:bc30-098970	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-074:bc30-015189	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-074:bc30-029197	0.324	0.209	#N/A	#N/A	Sensitive
bc14-074:bc30-056463	#N/A	#N/A	0.106	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-074:bc30-084891	0.324	#N/A	#N/A	#N/A	Sensitive
bc14-075:bc30-012281	#N/A	0.418	#N/A	#N/A	Sensitive
bc14-075:bc30-017658	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-075:bc30-017829	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-075:bc30-078925	0.130	#N/A	0.106	#N/A	Sensitive
bc14-076:bc30-002264	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-076:bc30-076229	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-077:bc30-018113	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-077:bc30-027907	0.065	#N/A	0.212	#N/A	Sensitive
bc14-077:bc30-052322	0.065	0.209	#N/A	#N/A	Tolerant
bc14-077:bc30-090562	0.065	#N/A	#N/A	#N/A	Tolerant
bc14-079:bc30-029643	0.065	0.209	#N/A	#N/A	Tolerant
bc14-079:bc30-045148	0.260	#N/A	#N/A	#N/A	Sensitive
bc14-079:bc30-049731	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-079:bc30-085071	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-079:bc30-087174	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-080:bc30-006100	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-080:bc30-035943	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-080:bc30-087209	#N/A	#N/A	0.106	#N/A	Tolerant
bc14-083:bc30-035442	0.195	0.209	0.106	#N/A	Sensitive
bc14-084:bc30-034645	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-084:bc30-082356	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-085:bc30-049090	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-085:bc30-051927	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-085:bc30-052053	#N/A	#N/A	#N/A	#N/A	Tolerant
bc14-085:bc30-095436	0.065	0.209	#N/A	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-086:bc30-011266	0.065	0.209	0.106	#N/A	Sensitive
bc14-086:bc30-012804	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-086:bc30-014797	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-086:bc30-016545	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-086:bc30-026510	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-086:bc30-064451	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-086:bc30-094800	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-087:bc30-017654	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-087:bc30-023694	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-087:bc30-061156	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-087:bc30-085371	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-087:bc30-088025	0.195	#N/A	#N/A	#N/A	Sensitive
bc14-088:bc30-021201	0.130	0.209	#N/A	#N/A	Sensitive
bc14-088:bc30-042359	0.065	#N/A	#N/A	#N/A	Tolerant
bc14-088:bc30-077874	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-088:bc30-081283	0.260	#N/A	#N/A	#N/A	Tolerant
bc14-090:bc30-063687	#N/A	0.209	0.212	#N/A	Sensitive
bc14-090:bc30-079989	0.130	#N/A	#N/A	#N/A	Sensitive
bc14-090:bc30-091729	0.065	#N/A	0.106	#N/A	Sensitive
bc14-093:bc30-017643	0.130	#N/A	#N/A	#N/A	Tolerant
bc14-093:bc30-021312	0.260	#N/A	#N/A	#N/A	Tolerant
bc14-093:bc30-031743	#N/A	#N/A	#N/A	#N/A	Sensitive
bc14-093:bc30-046216	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-093:bc30-055999	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-095:bc30-010829	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-095:bc30-017886	0.324	#N/A	0.318	#N/A	Tolerant

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-095:bc30-018626	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-095:bc30-023975	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-095:bc30-026140	0.065	#N/A	0.318	#N/A	Sensitive
bc14-095:bc30-028359	0.130	#N/A	0.106	#N/A	Sensitive
bc14-095:bc30-030772	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-095:bc30-048456	0.195	#N/A	#N/A	#N/A	Sensitive
bc14-095:bc30-052882	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-095:bc30-053340	0.260	0.209	#N/A	#N/A	Sensitive
bc14-095:bc30-058330	#N/A	0.209	#N/A	#N/A	Sensitive
bc14-095:bc30-068182	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-095:bc30-070011	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-095:bc30-079714	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-095:bc30-090215	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-095:bc30-093295	0.065	#N/A	0.106	#N/A	Sensitive
bc14-096:bc30-017133	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-096:bc30-037139	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-096:bc30-053569	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-096:bc30-085993	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-097:bc30-004602	0.065	#N/A	#N/A	#N/A	Tolerant
bc14-097:bc30-007881	0.065	0.209	0.106	#N/A	Sensitive
bc14-097:bc30-027180	#N/A	#N/A	0.106	#N/A	Sensitive
bc14-097:bc30-030681	0.065	0.209	#N/A	#N/A	Sensitive
bc14-097:bc30-031611	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-097:bc30-090378	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-097:bc30-097790	0.454	0.209	0.106	#N/A	Tolerant
bc14-098:bc30-094274	0.065	#N/A	#N/A	#N/A	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-099:bc30-006725	0.389	0.209	0.212	#N/A	Tolerant
bc14-099:bc30-024293	#N/A	0.418	#N/A	#N/A	Sensitive
bc14-099:bc30-056416	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-100:bc30-095175	0.065	#N/A	#N/A	#N/A	Sensitive
bc14-087:bc30-018814	0.324	#N/A	0.106	0.615	Sensitive
bc14-036:bc30-057524	0.065	0.626	0.318	0.527	Sensitive
bc14-052:bc30-015400	0.130	0.418	0.425	0.527	Sensitive
bc14-061:bc30-056416	#N/A	1.253	0.531	0.527	Sensitive
bc14-062:bc30-023729	0.195	0.835	0.531	0.527	Sensitive
bc14-030:bc30-021720	0.195	0.209	0.318	0.439	Sensitive
bc14-054:bc30-009684	0.519	0.835	0.425	0.439	Sensitive
bc14-001:bc30-020424	#N/A	#N/A	#N/A	0.351	Tolerant
bc14-010:bc30-010957	0.649	0.209	0.106	0.351	Sensitive
bc14-032:bc30-098530	0.130	#N/A	0.318	0.351	Sensitive
bc14-057:bc30-073331	0.195	#N/A	0.106	0.351	Sensitive
bc14-097:bc30-049818	0.454	1.461	0.955	0.351	Sensitive
bc14-002:bc30-036494	0.130	#N/A	0.106	0.263	Sensitive
bc14-006:bc30-062663	0.454	1.253	0.531	0.263	Sensitive
bc14-016:bc30-017899	0.130	#N/A	#N/A	0.263	Sensitive
bc14-018:bc30-079210	0.065	#N/A	0.106	0.263	Sensitive
bc14-029:bc30-051262	0.195	0.209	#N/A	0.263	Sensitive
bc14-058:bc30-063078	0.195	0.209	0.106	0.263	Sensitive
bc14-071:bc30-014542	#N/A	#N/A	#N/A	0.263	Sensitive
bc14-002:bc30-005695	0.130	#N/A	#N/A	0.176	Sensitive
bc14-003:bc30-016505	0.389	#N/A	0.318	0.176	Sensitive
bc14-005:bc30-040269	0.324	0.209	0.318	0.176	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-006:bc30-076376	0.065	#N/A	#N/A	0.176	Sensitive
bc14-008:bc30-013525	0.324	0.626	0.106	0.176	Sensitive
bc14-013:bc30-079065	#N/A	#N/A	0.106	0.176	Tolerant
bc14-018:bc30-024519	0.324	0.209	0.212	0.176	Sensitive
bc14-018:bc30-097470	#N/A	#N/A	0.212	0.176	Sensitive
bc14-029:bc30-099942	#N/A	#N/A	#N/A	0.176	Sensitive
bc14-033:bc30-026128	0.065	#N/A	#N/A	0.176	Sensitive
bc14-036:bc30-047443	0.324	#N/A	0.106	0.176	Sensitive
bc14-037:bc30-017983	0.195	0.209	0.106	0.176	Sensitive
bc14-037:bc30-098606	#N/A	#N/A	#N/A	0.176	Sensitive
bc14-039:bc30-001884	#N/A	0.209	#N/A	0.176	Sensitive
bc14-039:bc30-004780	0.195	#N/A	0.212	0.176	Sensitive
bc14-039:bc30-093289	0.260	0.209	0.106	0.176	Sensitive
bc14-046:bc30-001122	#N/A	0.209	#N/A	0.176	Sensitive
bc14-051:bc30-007396	0.065	#N/A	#N/A	0.176	Sensitive
bc14-051:bc30-042898	#N/A	#N/A	#N/A	0.176	Tolerant
bc14-052:bc30-014687	0.130	0.209	#N/A	0.176	Sensitive
bc14-054:bc30-008363	#N/A	0.209	#N/A	0.176	Sensitive
bc14-054:bc30-043716	#N/A	#N/A	0.425	0.176	Sensitive
bc14-054:bc30-086176	1.103	0.418	0.318	0.176	Sensitive
bc14-055:bc30-009568	0.260	0.209	0.212	0.176	Sensitive
bc14-055:bc30-096249	#N/A	0.209	#N/A	0.176	Sensitive
bc14-056:bc30-081470	0.065	0.418	0.212	0.176	Sensitive
bc14-058:bc30-090278	#N/A	#N/A	#N/A	0.176	Sensitive
bc14-061:bc30-010643	#N/A	#N/A	#N/A	0.176	Sensitive
bc14-065:bc30-055225	0.324	#N/A	#N/A	0.176	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-076:bc30-038301	0.130	0.209	0.212	0.176	Sensitive
bc14-080:bc30-044858	0.065	#N/A	0.637	0.176	Sensitive
bc14-085:bc30-035729	#N/A	#N/A	#N/A	0.176	Sensitive
bc14-085:bc30-073908	0.065	#N/A	0.106	0.176	Sensitive
bc14-086:bc30-029282	#N/A	#N/A	#N/A	0.176	Tolerant
bc14-089:bc30-041178	0.065	#N/A	#N/A	0.176	Sensitive
bc14-092:bc30-045445	0.260	0.835	0.318	0.176	Sensitive
bc14-093:bc30-078550	0.065	#N/A	#N/A	0.176	Sensitive
bc14-001:bc30-030855	#N/A	#N/A	0.106	0.088	Sensitive
bc14-003:bc30-016579	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-005:bc30-016671	0.130	0.418	#N/A	0.088	Sensitive
bc14-005:bc30-022606	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-005:bc30-030128	0.130	#N/A	#N/A	0.088	Sensitive
bc14-005:bc30-066773	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-006:bc30-037654	0.389	#N/A	0.106	0.088	Sensitive
bc14-008:bc30-086830	0.195	0.209	0.212	0.088	Sensitive
bc14-010:bc30-012146	0.260	#N/A	#N/A	0.088	Sensitive
bc14-010:bc30-083133	0.065	#N/A	#N/A	0.088	Sensitive
bc14-011:bc30-034804	#N/A	#N/A	0.212	0.088	Tolerant
bc14-011:bc30-049090	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-012:bc30-006577	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-012:bc30-083329	0.065	#N/A	#N/A	0.088	Sensitive
bc14-013:bc30-027978	0.195	0.209	#N/A	0.088	Sensitive
bc14-015:bc30-070768	0.130	0.209	#N/A	0.088	Sensitive
bc14-015:bc30-082024	0.195	#N/A	#N/A	0.088	Sensitive
bc14-021:bc30-031111	0.065	#N/A	#N/A	0.088	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-021:bc30-071239	#N/A	#N/A	0.212	0.088	Sensitive
bc14-029:bc30-010283	0.065	#N/A	0.106	0.088	Sensitive
bc14-033:bc30-014112	#N/A	0.209	#N/A	0.088	Sensitive
bc14-033:bc30-045424	0.195	#N/A	#N/A	0.088	Sensitive
bc14-033:bc30-091477	#N/A	#N/A	0.106	0.088	Sensitive
bc14-034:bc30-035948	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-036:bc30-000431	0.065	#N/A	#N/A	0.088	Sensitive
bc14-036:bc30-044408	0.065	#N/A	#N/A	0.088	Sensitive
bc14-036:bc30-056689	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-037:bc30-015258	0.130	0.209	#N/A	0.088	Sensitive
bc14-037:bc30-092499	0.130	#N/A	#N/A	0.088	Sensitive
bc14-038:bc30-012487	0.324	0.209	0.106	0.088	Sensitive
bc14-039:bc30-001593	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-039:bc30-054153	0.389	#N/A	0.106	0.088	Sensitive
bc14-039:bc30-090948	0.065	#N/A	#N/A	0.088	Sensitive
bc14-039:bc30-094541	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-040:bc30-043189	0.130	0.209	#N/A	0.088	Sensitive
bc14-040:bc30-048938	0.195	0.209	0.212	0.088	Sensitive
bc14-040:bc30-059270	0.065	0.209	0.106	0.088	Sensitive
bc14-041:bc30-051143	#N/A	0.626	0.212	0.088	Sensitive
bc14-041:bc30-053766	0.130	#N/A	#N/A	0.088	Sensitive
bc14-041:bc30-092680	0.195	0.209	#N/A	0.088	Sensitive
bc14-045:bc30-059780	0.130	#N/A	#N/A	0.088	Sensitive
bc14-046:bc30-063899	0.065	0.209	0.212	0.088	Sensitive
bc14-047:bc30-080252	#N/A	#N/A	0.106	0.088	Tolerant
bc14-049:bc30-072859	0.195	#N/A	0.106	0.088	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-050:bc30-021964	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-051:bc30-033549	0.130	#N/A	#N/A	0.088	Sensitive
bc14-052:bc30-017825	#N/A	0.209	0.106	0.088	Sensitive
bc14-052:bc30-027804	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-053:bc30-001774	0.065	#N/A	0.106	0.088	Sensitive
bc14-053:bc30-020023	#N/A	#N/A	0.212	0.088	Sensitive
bc14-054:bc30-034945	#N/A	0.209	#N/A	0.088	Tolerant
bc14-054:bc30-047787	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-054:bc30-057477	#N/A	#N/A	0.106	0.088	Tolerant
bc14-054:bc30-092374	0.195	#N/A	0.106	0.088	Sensitive
bc14-055:bc30-045298	#N/A	#N/A	#N/A	0.088	Tolerant
bc14-055:bc30-058298	#N/A	0.209	#N/A	0.088	Tolerant
bc14-055:bc30-062697	#N/A	0.209	#N/A	0.088	Sensitive
bc14-055:bc30-082736	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-057:bc30-005958	0.130	#N/A	#N/A	0.088	Sensitive
bc14-057:bc30-033121	0.130	#N/A	#N/A	0.088	Sensitive
bc14-057:bc30-076524	0.649	0.209	#N/A	0.088	Sensitive
bc14-057:bc30-087557	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-058:bc30-060180	0.130	0.209	#N/A	0.088	Sensitive
bc14-058:bc30-095762	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-059:bc30-056389	0.195	#N/A	0.212	0.088	Sensitive
bc14-061:bc30-023425	0.260	#N/A	#N/A	0.088	Sensitive
bc14-065:bc30-029169	0.195	#N/A	0.106	0.088	Sensitive
bc14-065:bc30-064666	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-065:bc30-064882	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-065:bc30-072137	#N/A	0.209	#N/A	0.088	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-065:bc30-095370	0.065	0.209	#N/A	0.088	Sensitive
bc14-065:bc30-099357	#N/A	#N/A	#N/A	0.088	Tolerant
bc14-066:bc30-073305	0.454	#N/A	0.106	0.088	Sensitive
bc14-069:bc30-004604	0.065	#N/A	#N/A	0.088	Sensitive
bc14-069:bc30-046350	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-073:bc30-057065	0.844	0.418	0.212	0.088	Sensitive
bc14-076:bc30-045840	0.065	#N/A	#N/A	0.088	Sensitive
bc14-079:bc30-062970	0.130	0.209	#N/A	0.088	Sensitive
bc14-080:bc30-060261	0.195	0.418	0.106	0.088	Sensitive
bc14-080:bc30-094133	#N/A	0.209	#N/A	0.088	Sensitive
bc14-086:bc30-017902	0.130	#N/A	0.106	0.088	Sensitive
bc14-086:bc30-031919	0.389	#N/A	0.212	0.088	Sensitive
bc14-086:bc30-034547	0.260	#N/A	#N/A	0.088	Sensitive
bc14-086:bc30-041658	0.130	#N/A	#N/A	0.088	Sensitive
bc14-086:bc30-047583	#N/A	0.209	#N/A	0.088	Tolerant
bc14-086:bc30-073968	#N/A	#N/A	0.106	0.088	Sensitive
bc14-086:bc30-085498	0.130	#N/A	#N/A	0.088	Sensitive
bc14-087:bc30-004923	#N/A	0.209	#N/A	0.088	Sensitive
bc14-087:bc30-063176	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-088:bc30-067626	0.065	0.209	#N/A	0.088	Sensitive
bc14-089:bc30-007827	#N/A	0.209	#N/A	0.088	Sensitive
bc14-089:bc30-073923	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-090:bc30-065489	#N/A	0.209	#N/A	0.088	Tolerant
bc14-093:bc30-013754	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-095:bc30-005522	#N/A	#N/A	#N/A	0.088	Tolerant
bc14-095:bc30-006695	0.065	#N/A	#N/A	0.088	Sensitive

Table 6.6 continued from previous page

Lineage ID	Frequency Day 0	Frequency Day 3	Frequency Day 6	Frequency Day 9	Afatinib response group
bc14-095:bc30-045251	0.130	#N/A	#N/A	0.088	Sensitive
bc14-096:bc30-089923	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-097:bc30-011582	#N/A	#N/A	#N/A	0.088	Sensitive
bc14-097:bc30-045223	0.195	#N/A	0.106	0.088	Sensitive
bc14-097:bc30-075976	0.130	0.209	#N/A	0.088	Sensitive
bc14-097:bc30-093253	#N/A	#N/A	#N/A	0.088	Tolerant

Table 6.7. Differentially expressed genes using cells of dominant clones with clonals versus cells associated to neutral clones. (A) HGNC Gene Symbol **(B)** Average expression of the given gene in dominant clones **(C)** Average expression of the given gene in not dominant clones **(D)** Average fold change in the expression of given gene between dominant and not dominant clones **(E)** Mean standard deviation of fold change in the expression of given gene between dominant and not dominant clones **(F)** Cumulative fold change in the expression of given gene in both populations **(G)** p-value of a Student's t-test **(H)** False Discovery Rate (FDR) correction of p-value for the given gene. *Only significant (FDR < 0.1) are shown*

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
IGFBP2	8.58	7.87	0.70	0.358	35.188	3.25e-7	1.60e-6
MYEOV	7.56	6.87	0.70	0.39	34.746	2.10e-2	4.63e-2
SERPINB6	7.59	6.92	0.67	0.389	33.646	4.28e-5	1.49e-4
ALCAM	8.07	7.43	0.64	0.586	31.935	2.23e-14	5.02e-13
BST2	7.55	6.91	0.64	0.485	31.923	3.86e-8	2.19e-7
SLPI	8.46	7.86	0.59	0.384	29.68	3.15e-5	1.12e-4
CLDN10	7.50	6.91	0.59	0.377	29.507	8.96e-6	3.49e-5
MME	7.36	6.78	0.59	0.35	29.438	2.42e-3	6.38e-3
RNF168	7.02	6.43	0.58	0.216	29.116	3.10e-4	9.41e-4

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
OSBP2	7.17	6.59	0.58	0.303	29.103	1.24e-2	2.83e-2
ALDH2	7.66	7.09	0.57	0.377	28.719	2.24e-10	1.94e-9
GLYATL2	7.27	6.71	0.56	0.382	28.197	2.28e-9	1.63e-8
XAGE2	7.30	6.75	0.56	0.419	27.807	4.44e-9	2.98e-8
NPR3	6.99	6.44	0.55	0.435	27.68	4.50e-23	1.70e-20
AP2M1	8.66	8.12	0.54	0.278	27.217	1.93e-8	1.17e-7
LAMP1	8.60	8.07	0.53	0.287	26.496	1.51e-7	7.83e-7
SLC1A6	7.97	7.44	0.53	0.504	26.444	6.35e-10	5.02e-9
QPRT	8.37	7.84	0.52	0.307	26.203	2.21e-5	8.02e-5
NDUFB5	7.65	7.13	0.52	0.588	26.105	5.67e-15	1.49e-13
CD82	8.06	7.54	0.52	0.279	26.08	1.56e-6	6.81e-6
MDK	8.02	7.50	0.52	0.551	25.794	7.54e-10	5.86e-9
CLDN3	7.79	7.27	0.52	0.414	25.775	1.83e-5	6.75e-5
MAP3K20	8.27	7.76	0.51	0.201	25.379	2.66e-3	6.97e-3
FGFBP1	6.93	6.43	0.50	0.299	25.079	4.29e-8	2.42e-7
MAGEF1	7.54	7.05	0.50	0.19	24.863	6.58e-4	1.90e-3
OLMALINC	7.25	6.76	0.50	0.424	24.729	8.41e-11	7.92e-10
SELENOP	6.93	6.43	0.49	0.266	24.636	1.18e-9	8.83e-9
NUDT8	7.46	6.97	0.49	0.492	24.492	4.19e-7	2.02e-6
S100A9	7.08	6.59	0.48	0.425	24.114	9.31e-13	1.40e-11
NFIA	7.12	6.64	0.48	0.254	24.112	4.35e-5	1.51e-4
SUSD2	7.50	7.01	0.48	0.269	24.051	6.51e-8	3.58e-7
NCBP2AS2	7.62	7.14	0.48	0.439	24.025	8.33e-13	1.27e-11
PI3	6.99	6.52	0.48	0.345	23.84	3.17e-6	1.32e-5
DLG1	7.53	7.05	0.47	0.282	23.576	9.01e-5	2.97e-4
MVD	7.76	7.30	0.46	0.291	23.128	8.67e-4	2.46e-3

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
SEMA3E	6.93	6.48	0.46	0.27	22.955	5.33e-6	2.15e-5
COMT	8.54	8.09	0.46	0.322	22.898	1.29e-7	6.79e-7
LGMN	7.15	6.69	0.46	0.372	22.841	4.11e-4	1.23e-3
ITGB6	6.95	6.49	0.46	0.286	22.776	1.58e-6	6.91e-6
DHCR7	7.82	7.36	0.45	0.338	22.629	8.66e-8	4.66e-7
ARID5A	7.02	6.57	0.45	0.27	22.618	3.72e-4	1.12e-3
HES6	7.76	7.31	0.45	0.367	22.391	1.34e-4	4.31e-4
DYNLT2B	7.56	7.11	0.45	0.442	22.337	1.63e-9	1.19e-8
TCF7L1	7.78	7.33	0.45	0.293	22.335	5.88e-3	1.42e-2
POLR2H	8.09	7.64	0.44	0.352	22.105	1.03e-12	1.54e-11
CYP27A1	7.01	6.58	0.44	0.351	21.927	9.49e-5	3.11e-4
TRIM29	7.30	6.86	0.44	0.365	21.832	4.75e-9	3.16e-8
TFDP1	8.07	7.64	0.44	0.219	21.792	1.62e-4	5.15e-4
BCL11A	6.93	6.49	0.44	0.323	21.741	4.82e-7	2.30e-6
ATP1B1	8.31	7.88	0.43	0.31	21.657	6.48e-10	5.11e-9
ANK3	7.75	7.32	0.43	0.324	21.509	1.44e-5	5.38e-5
FBXO45	6.90	6.47	0.43	0.244	21.451	1.33e-2	3.03e-2
PIGX	7.16	6.74	0.43	0.308	21.434	3.42e-5	1.21e-4
RFC4	7.15	6.72	0.43	0.373	21.422	2.31e-7	1.16e-6
ADAMTS1	6.87	6.44	0.43	0.439	21.414	2.46e-27	3.07e-24
PTMS	8.02	7.60	0.43	0.257	21.244	6.04e-3	1.46e-2
CITED4	9.60	9.18	0.42	0.257	21.102	2.87e-6	1.20e-5
ATP13A3	6.99	6.57	0.42	0.321	21.04	3.47e-2	7.32e-2
C1GALT1C1	7.08	6.66	0.42	0.401	20.833	5.92e-7	2.78e-6
SCPEP1	7.11	6.70	0.41	0.415	20.628	7.64e-6	3.01e-5
ALG3	7.38	6.97	0.41	0.518	20.523	5.55e-19	5.28e-17

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
ENC1	8.21	7.80	0.41	0.304	20.436	1.03e-6	4.61e-6
IRX3	8.24	7.84	0.41	0.171	20.363	3.42e-4	1.03e-3
TEX30	7.47	7.07	0.41	0.451	20.304	1.61e-8	9.82e-8
PROM1	7.28	6.87	0.41	0.377	20.295	1.55e-8	9.51e-8
BLVRB	8.09	7.69	0.40	0.418	20.203	2.84e-8	1.66e-7
VGLL1	7.08	6.68	0.40	0.457	20.076	1.33e-19	1.61e-17
DCUN1D1	6.93	6.53	0.40	0.269	20.067	2.85e-9	2.00e-8
TMEM205	7.36	6.96	0.40	0.483	20.038	4.96e-13	7.94e-12
SLC2A11	7.01	6.61	0.40	0.347	20.033	2.95e-5	1.05e-4
MUC15	7.46	7.07	0.40	0.252	19.898	3.52e-4	1.06e-3
PDZK1IP1	6.85	6.45	0.40	0.288	19.884	1.06e-14	2.60e-13
DNAJC3	7.76	7.37	0.40	0.215	19.848	5.87e-5	1.99e-4
ACTL6A	7.39	6.99	0.40	0.351	19.848	6.50e-9	4.23e-8
UBAC2	7.33	6.94	0.40	0.294	19.809	3.83e-6	1.57e-5
TMED9	8.55	8.15	0.40	0.22	19.798	2.30e-4	7.12e-4
LINC00662	9.24	8.84	0.39	0.696	19.687	5.07e-16	1.79e-14
MPZL1	7.77	7.38	0.39	0.22	19.648	1.29e-2	2.95e-2
LPP	7.18	6.79	0.39	0.272	19.565	1.29e-2	2.94e-2
PPP1R2	7.13	6.73	0.39	0.294	19.551	1.38e-6	6.07e-6
CCDC88B	6.81	6.42	0.39	0.221	19.396	1.78e-7	9.06e-7
CLN3	7.47	7.08	0.39	0.275	19.365	5.86e-5	1.99e-4
ATP11A	8.18	7.79	0.39	0.394	19.327	2.95e-7	1.46e-6
ITM2C	7.57	7.18	0.39	0.263	19.303	4.19e-5	1.46e-4
SYNGR1	6.98	6.59	0.39	0.267	19.28	3.07e-8	1.78e-7
FBXO32	7.41	7.03	0.39	0.385	19.254	3.68e-5	1.29e-4
EIF2B5	7.10	6.72	0.38	0.288	19.223	9.53e-3	2.22e-2

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
ITGA6	7.91	7.53	0.38	0.212	19.123	1.64e-2	3.68e-2
UBE2E3	7.47	7.08	0.38	0.428	19.11	1.00e-5	3.87e-5
PMF1	7.82	7.44	0.38	0.454	19.092	6.63e-8	3.64e-7
COMMD4	7.56	7.18	0.38	0.464	19.025	1.39e-7	7.24e-7
CYP1B1	9.20	8.82	0.38	0.436	18.99	4.71e-25	4.20e-22
LGALS3BP	7.82	7.44	0.38	0.392	18.925	3.18e-5	1.13e-4
ID4	8.49	8.11	0.38	0.45	18.884	5.04e-9	3.34e-8
OPA1	6.87	6.50	0.38	0.244	18.841	9.39e-3	2.19e-2
CCN1	7.06	6.68	0.38	0.301	18.828	2.86e-2	6.13e-2
MESP1	7.61	7.24	0.38	0.382	18.797	3.98e-10	3.27e-9
BBOX1	6.91	6.54	0.38	0.281	18.794	3.97e-18	2.86e-16
PTPN14	7.81	7.44	0.37	0.309	18.687	2.30e-2	5.03e-2
NUPR1	9.07	8.70	0.37	0.292	18.559	1.22e-4	3.93e-4
SLC9A3R2	7.74	7.37	0.37	0.291	18.539	2.07e-3	5.53e-3
NXPH4	7.02	6.65	0.37	0.379	18.52	1.34e-7	7.01e-7
ALDH9A1	7.35	6.98	0.37	0.335	18.502	2.06e-8	1.24e-7
LGR4	7.38	7.02	0.37	0.34	18.403	1.10e-4	3.56e-4
C5orf38	7.51	7.15	0.37	0.367	18.351	6.60e-5	2.22e-4
CLN5	6.89	6.53	0.37	0.222	18.295	2.20e-4	6.84e-4
GRTP1	7.13	6.76	0.37	0.267	18.272	4.70e-3	1.16e-2
BOC	6.93	6.57	0.37	0.372	18.248	2.03e-5	7.43e-5
FDPS	9.09	8.73	0.36	0.243	18.202	3.72e-8	2.13e-7
ERG28	7.16	6.80	0.36	0.384	18.188	1.51e-10	1.35e-9
TRA2B	8.17	7.81	0.36	0.234	18.143	1.50e-2	3.38e-2
STOM	6.80	6.44	0.36	0.218	18.065	7.55e-12	9.02e-11
KLHDC3	7.41	7.05	0.36	0.411	18.051	4.73e-6	1.92e-5

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
EMP2	7.87	7.51	0.36	0.257	18.029	3.21e-10	2.71e-9
IL22RA2	6.84	6.48	0.36	0.259	18.027	1.99e-3	5.34e-3
COMMD9	7.30	6.94	0.36	0.352	18.014	1.30e-10	1.17e-9
NDFIP2	7.25	6.89	0.36	0.308	17.989	1.16e-7	6.14e-7
PARL	7.09	6.73	0.36	0.37	17.961	1.75e-11	1.94e-10
ARGLU1	8.52	8.16	0.36	0.285	17.874	6.42e-13	9.92e-12
GCHFR	6.81	6.46	0.36	0.25	17.79	2.47e-23	9.62e-21
ARL4A	7.71	7.36	0.36	0.309	17.786	3.58e-7	1.75e-6
AHCYL1	8.21	7.85	0.36	0.272	17.771	2.31e-7	1.16e-6
MRPL47	7.46	7.10	0.35	0.491	17.626	8.69e-11	8.18e-10
BCL7C	7.69	7.34	0.35	0.366	17.597	6.88e-16	2.31e-14
TXN2	7.98	7.63	0.35	0.367	17.518	1.18e-8	7.37e-8
RERG	6.92	6.57	0.35	0.313	17.507	8.20e-8	4.43e-7
TMCO3	7.02	6.67	0.35	0.34	17.472	3.45e-3	8.89e-3
CDC42BPG	6.91	6.57	0.35	0.344	17.375	3.62e-4	1.09e-3
EEF1AKMT4	6.91	6.57	0.35	0.267	17.369	1.07e-9	8.08e-9
GAS1	6.76	6.42	0.35	0.216	17.28	1.73e-3	4.69e-3
CSRP2	6.95	6.61	0.35	0.262	17.241	1.95e-11	2.15e-10
NOL12	7.41	7.06	0.34	0.475	17.218	6.42e-14	1.30e-12
ECHDC2	7.03	6.69	0.34	0.305	17.197	8.72e-6	3.40e-5
FRMD4A	6.81	6.46	0.34	0.221	17.194	2.28e-2	4.98e-2
NEAT1	9.26	8.91	0.34	0.632	17.143	5.87e-14	1.20e-12
POLR2J3	8.51	8.17	0.34	0.385	17.136	4.17e-10	3.42e-9
TMX4	7.24	6.90	0.34	0.382	17.075	2.49e-2	5.41e-2
SIPA1	6.84	6.50	0.34	0.267	17.016	1.32e-9	9.77e-9
FAM241B	7.10	6.76	0.34	0.37	17.015	4.55e-11	4.58e-10

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
PCDH9	6.77	6.43	0.34	0.259	17.01	7.82e-11	7.45e-10
TMED1	7.08	6.74	0.34	0.336	16.972	6.52e-6	2.60e-5
GGA1	7.22	6.88	0.34	0.223	16.936	3.40e-3	8.78e-3
EPHB3	6.76	6.43	0.34	0.174	16.934	2.30e-2	5.03e-2
BACE2	7.48	7.15	0.34	0.342	16.922	3.26e-8	1.88e-7
ANKRD10	7.29	6.95	0.34	0.297	16.921	9.38e-8	5.04e-7
CD99	7.73	7.39	0.34	0.414	16.899	2.79e-9	1.96e-8
PFN2	7.52	7.18	0.34	0.452	16.879	1.53e-7	7.93e-7
AGA	6.84	6.50	0.34	0.227	16.863	8.11e-6	3.18e-5
VTCN1	7.35	7.01	0.34	0.343	16.841	6.35e-6	2.53e-5
PLEKHH3	6.97	6.64	0.34	0.358	16.792	1.18e-5	4.48e-5
ATF6B	6.85	6.52	0.34	0.311	16.748	3.29e-3	8.52e-3
COA5	7.16	6.83	0.33	0.383	16.723	5.74e-13	9.07e-12
SEC62	8.22	7.89	0.33	0.298	16.63	3.19e-4	9.69e-4
CRABP2	9.72	9.39	0.33	0.536	16.565	1.06e-12	1.56e-11
ISYNA1	7.67	7.34	0.33	0.367	16.55	4.59e-9	3.07e-8
DIPK2A	6.84	6.51	0.33	0.263	16.543	5.37e-7	2.54e-6
FDX2	8.46	8.13	0.33	0.431	16.514	6.96e-15	1.80e-13
FOS	6.97	6.64	0.33	0.323	16.446	4.78e-4	1.41e-3
MGST2	7.07	6.74	0.33	0.345	16.43	1.99e-6	8.54e-6
SREBF1	6.88	6.55	0.33	0.273	16.4	1.50e-2	3.38e-2
KEAP1	7.83	7.51	0.33	0.322	16.363	1.32e-9	9.80e-9
TXNRD1	7.11	6.78	0.33	0.298	16.359	1.08e-2	2.49e-2
ALDH3B2	7.41	7.09	0.33	0.298	16.311	1.64e-3	4.46e-3
CCPG1	6.93	6.61	0.33	0.282	16.3	2.61e-6	1.10e-5
SCAMP3	7.46	7.13	0.33	0.455	16.287	2.08e-11	2.27e-10

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
CRELD2	7.20	6.88	0.33	0.248	16.249	1.46e-2	3.29e-2
ABHD8	7.13	6.80	0.33	0.304	16.237	3.66e-5	1.29e-4
EPC1	7.08	6.76	0.32	0.339	16.217	1.03e-5	3.96e-5
BCAM	6.89	6.57	0.32	0.287	16.185	4.37e-5	1.52e-4
TADA3	6.92	6.60	0.32	0.243	16.145	1.44e-8	8.88e-8
PRKCI	7.27	6.95	0.32	0.211	16.095	8.64e-3	2.03e-2
GCAT	7.00	6.68	0.32	0.252	16.089	7.00e-9	4.53e-8
CYB5R3	7.49	7.17	0.32	0.386	16.034	3.34e-7	1.64e-6
NFKBIA	7.23	6.91	0.32	0.347	16.029	3.48e-8	2.00e-7
VPS28	8.28	7.96	0.32	0.308	16.022	1.03e-4	3.36e-4
SCD	7.89	7.57	0.32	0.328	16.009	1.03e-5	3.95e-5
TSC22D3	6.76	6.44	0.32	0.225	15.972	2.54e-18	1.97e-16
MCAT	6.78	6.46	0.32	0.202	15.95	2.88e-10	2.45e-9
HERPUD1	7.92	7.60	0.32	0.604	15.938	2.90e-17	1.58e-15
TCEAL1	6.79	6.47	0.32	0.212	15.934	7.05e-11	6.81e-10
SLC25A1	7.57	7.25	0.32	0.479	15.886	1.06e-13	2.03e-12
PILRB	6.99	6.68	0.32	0.361	15.858	2.89e-8	1.69e-7
CC2D1A	7.10	6.79	0.32	0.261	15.803	6.89e-4	1.98e-3
IER5	7.91	7.59	0.32	0.34	15.802	4.24e-13	6.90e-12
ALKBH7	7.71	7.39	0.32	0.491	15.794	2.05e-8	1.23e-7
STK24	7.95	7.63	0.32	0.288	15.786	9.99e-3	2.33e-2
CLDN4	7.81	7.50	0.32	0.282	15.775	2.92e-2	6.23e-2
NQO2	7.79	7.47	0.32	0.38	15.769	3.88e-7	1.88e-6
ALYREF	8.28	7.96	0.32	0.402	15.753	2.02e-8	1.21e-7
IRX6	6.74	6.43	0.32	0.194	15.74	1.35e-3	3.71e-3
TM9SF2	7.93	7.62	0.31	0.248	15.721	7.75e-4	2.21e-3

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
LDLR	7.64	7.32	0.31	0.529	15.712	1.30e-12	1.88e-11
LMNB1	7.45	7.14	0.31	0.398	15.688	3.28e-4	9.94e-4
SLC27A3	6.79	6.48	0.31	0.261	15.686	2.55e-4	7.83e-4
MARCKSL1	9.33	9.02	0.31	0.224	15.675	1.77e-5	6.52e-5
GSN	8.13	7.81	0.31	0.295	15.646	6.46e-6	2.57e-5
CNIH2	6.89	6.58	0.31	0.3	15.644	5.81e-18	3.87e-16
CHD7	7.01	6.70	0.31	0.33	15.629	2.21e-2	4.84e-2
CREG1	7.72	7.40	0.31	0.338	15.61	4.35e-4	1.29e-3
HOOK3	6.95	6.64	0.31	0.189	15.558	4.26e-2	8.83e-2
RAB25	8.29	7.98	0.31	0.492	15.552	3.74e-11	3.85e-10
OSTF1	7.23	6.92	0.31	0.363	15.549	8.19e-6	3.21e-5
FAM210B	7.90	7.59	0.31	0.331	15.54	6.88e-6	2.73e-5
IRF2BPL	7.05	6.74	0.31	0.348	15.538	3.07e-9	2.13e-8
GLI4	6.82	6.51	0.31	0.255	15.536	1.60e-9	1.17e-8
RGL2	6.76	6.45	0.31	0.183	15.504	4.31e-2	8.93e-2
HEXA	6.80	6.49	0.31	0.231	15.5	8.48e-5	2.81e-4
TOB1	6.79	6.48	0.31	0.209	15.482	1.21e-6	5.39e-6
MED11	6.76	6.45	0.31	0.244	15.443	5.40e-11	5.34e-10
ACO2	7.42	7.12	0.31	0.232	15.442	1.19e-3	3.30e-3
NT5DC2	8.09	7.79	0.31	0.248	15.426	1.26e-2	2.88e-2
TK1	7.49	7.18	0.31	0.613	15.412	9.23e-6	3.58e-5
CALR	9.62	9.32	0.31	0.215	15.379	5.31e-7	2.51e-6
CARHSP1	7.80	7.49	0.31	0.488	15.379	6.38e-13	9.90e-12
PLEKHF1	6.98	6.68	0.31	0.283	15.36	3.71e-5	1.30e-4
TMEM219	7.39	7.08	0.31	0.471	15.359	1.63e-7	8.40e-7
VSNL1	6.77	6.46	0.31	0.251	15.353	1.25e-5	4.76e-5

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
PLXDC2	7.80	7.49	0.31	0.254	15.346	1.08e-5	4.16e-5
HSD17B4	6.96	6.65	0.31	0.335	15.319	4.62e-4	1.37e-3
MGST1	8.93	8.63	0.31	0.527	15.316	1.35e-17	8.06e-16
TTC14	6.75	6.44	0.31	0.193	15.307	1.88e-5	6.91e-5
NINJ1	6.88	6.57	0.31	0.248	15.3	6.45e-7	3.01e-6
OBI1	6.87	6.57	0.31	0.282	15.229	1.55e-9	1.14e-8
FIS1	8.02	7.72	0.31	0.447	15.225	2.75e-9	1.93e-8
SREBF2	7.10	6.79	0.30	0.295	15.225	7.87e-6	3.09e-5
TWF2	7.34	7.04	0.30	0.415	15.181	1.14e-10	1.04e-9
WDR83OS	8.63	8.33	0.30	0.406	15.149	5.48e-9	3.61e-8
MYCBP2	7.96	7.66	0.30	0.23	15.144	2.31e-6	9.80e-6
PDGFC	6.93	6.62	0.30	0.265	15.097	7.27e-9	4.70e-8
RNF13	6.95	6.64	0.30	0.284	15.093	7.10e-7	3.29e-6
HES1	8.29	7.99	0.30	0.502	15.074	1.17e-15	3.69e-14
ACAT2	7.86	7.56	0.30	0.343	15.061	2.02e-14	4.61e-13
IDI1	7.82	7.52	0.30	0.204	15.056	1.26e-7	6.62e-7
CYB5R1	6.77	6.47	0.30	0.202	15.015	2.14e-5	7.79e-5
NUCB1	6.82	6.52	0.30	0.241	15.004	1.88e-2	4.18e-2
GCA	6.98	6.68	0.30	0.284	14.996	7.24e-8	3.95e-7
FGF2	6.86	6.56	0.30	0.251	14.971	2.67e-5	9.56e-5
DDAH2	8.24	7.94	0.30	0.311	14.97	2.54e-6	1.08e-5
SCAND1	8.04	7.74	0.30	0.419	14.967	7.75e-10	6.00e-9
PDCD10	7.88	7.58	0.30	0.328	14.95	1.10e-13	2.09e-12
TTC32	6.94	6.64	0.30	0.324	14.931	2.78e-3	7.26e-3
MCM3	7.83	7.54	0.30	0.393	14.915	1.45e-8	8.90e-8
EFHD1	8.87	8.57	0.30	0.249	14.904	2.11e-7	1.06e-6

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
COPG1	6.85	6.55	0.30	0.237	14.89	2.35e-4	7.27e-4
MAP3K13	7.10	6.80	0.30	0.282	14.88	2.45e-2	5.33e-2
IFT27	6.80	6.50	0.30	0.204	14.869	1.11e-7	5.92e-7
TAF15	7.55	7.25	0.30	0.373	14.868	2.99e-7	1.48e-6
RBM26	7.74	7.45	0.30	0.277	14.846	1.14e-3	3.16e-3
HRAS	7.90	7.61	0.30	0.371	14.837	1.33e-12	1.91e-11
CCDC50	6.97	6.67	0.30	0.294	14.797	1.21e-4	3.92e-4
AARD	9.16	8.86	0.30	0.222	14.728	1.73e-5	6.42e-5
CBR3	7.30	7.01	0.29	0.425	14.67	4.15e-7	2.01e-6
FANCL	6.87	6.57	0.29	0.271	14.661	1.23e-7	6.47e-7
TP53I11	6.91	6.62	0.29	0.381	14.646	2.14e-3	5.71e-3
OS9	6.89	6.60	0.29	0.238	14.638	9.57e-6	3.71e-5
MGP	10.00	9.71	0.29	0.414	14.61	2.61e-2	5.64e-2
RWDD3	6.78	6.49	0.29	0.186	14.561	4.60e-6	1.87e-5
FAM189B	6.91	6.62	0.29	0.26	14.543	2.52e-3	6.63e-3
ZNF428	8.02	7.73	0.29	0.302	14.525	6.23e-6	2.49e-5
DNAJC19	7.26	6.97	0.29	0.378	14.521	2.95e-4	8.99e-4
CHID1	7.02	6.73	0.29	0.275	14.52	5.30e-7	2.51e-6
KCTD3	7.07	6.78	0.29	0.28	14.52	2.01e-4	6.28e-4
METTL13	7.03	6.74	0.29	0.242	14.47	3.01e-6	1.26e-5
TMPRSS3	6.84	6.55	0.29	0.209	14.464	7.23e-5	2.41e-4
SUPT7L	6.81	6.52	0.29	0.242	14.458	7.05e-5	2.36e-4
SNRNP48	6.99	6.70	0.29	0.307	14.457	9.16e-4	2.58e-3
PCID2	7.38	7.09	0.29	0.214	14.455	1.04e-3	2.91e-3
ENDOV	7.29	7.00	0.29	0.353	14.429	7.50e-4	2.15e-3
OSGEP	6.99	6.70	0.29	0.235	14.427	1.63e-9	1.19e-8

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
RHOV	6.98	6.70	0.29	0.31	14.426	3.05e-4	9.29e-4
UPF3A	8.70	8.41	0.29	0.166	14.42	1.13e-2	2.60e-2
RFXANK	7.58	7.30	0.29	0.47	14.418	3.55e-12	4.62e-11
IFNGR1	7.57	7.29	0.29	0.295	14.414	2.49e-6	1.05e-5
TMC6	6.79	6.50	0.29	0.234	14.408	7.53e-7	3.48e-6
UFC1	8.69	8.40	0.29	0.287	14.395	6.86e-7	3.19e-6
ATP6V0E2	6.81	6.52	0.29	0.26	14.392	3.03e-10	2.57e-9
TMEM106B	6.86	6.57	0.29	0.219	14.39	1.11e-7	5.90e-7
UGGT2	6.75	6.47	0.29	0.217	14.39	3.64e-2	7.65e-2
ASF1B	7.31	7.02	0.29	0.528	14.351	1.29e-4	4.14e-4
PDIA4	7.61	7.32	0.29	0.245	14.347	5.91e-3	1.43e-2
PITHD1	7.59	7.30	0.29	0.391	14.322	2.64e-6	1.11e-5
TMEM161A	7.10	6.81	0.29	0.317	14.318	1.89e-4	5.94e-4
PEX13	6.77	6.49	0.29	0.186	14.317	2.14e-2	4.71e-2
GPR108	6.75	6.47	0.29	0.232	14.308	1.61e-4	5.12e-4
SNAP47	6.83	6.54	0.29	0.232	14.298	2.45e-3	6.45e-3
FAM111B	7.46	7.18	0.29	0.475	14.268	7.52e-9	4.85e-8
GADD45B	7.06	6.78	0.29	0.318	14.265	1.90e-4	5.96e-4
HPS1	6.78	6.50	0.29	0.252	14.243	5.24e-10	4.20e-9
LAMA4	7.66	7.37	0.29	0.391	14.233	1.27e-10	1.15e-9
CYYR1	6.93	6.65	0.28	0.254	14.217	1.74e-3	4.70e-3
RPAP2	6.82	6.53	0.28	0.293	14.159	9.95e-7	4.49e-6
NPTN	6.96	6.68	0.28	0.315	14.152	1.76e-5	6.49e-5
PSAP	8.14	7.86	0.28	0.232	14.151	2.73e-6	1.15e-5
TCAF1	7.43	7.15	0.28	0.272	14.148	3.76e-3	9.42e-3
GSTM3	6.95	6.66	0.28	0.358	14.144	3.55e-17	1.82e-15

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
PRDX2	9.92	9.63	0.28	0.511	14.133	2.99e-16	1.15e-14
PSPC1	7.44	7.16	0.28	0.299	14.131	4.45e-4	1.32e-3
GRN	7.41	7.13	0.28	0.34	14.129	9.21e-6	3.57e-5
NCKAP5	6.71	6.43	0.28	0.202	14.128	1.03e-2	2.39e-2
EHD1	7.28	7.00	0.28	0.254	14.124	2.50e-2	5.43e-2
DNASE2	8.32	8.04	0.28	0.323	14.121	3.00e-4	9.14e-4
MGMT	7.46	7.18	0.28	0.559	14.11	4.85e-10	3.91e-9
ENSA	8.60	8.31	0.28	0.313	14.085	3.80e-8	2.17e-7
BAX	7.56	7.27	0.28	0.459	14.085	2.16e-9	1.55e-8
TRAPPC6A	7.18	6.90	0.28	0.507	14.071	2.20e-12	3.03e-11
OBSL1	7.19	6.91	0.28	0.305	14.059	7.83e-4	2.23e-3
DRG1	7.34	7.06	0.28	0.527	14.052	2.02e-12	2.80e-11
EAPP	6.74	6.46	0.28	0.177	14.052	2.61e-2	5.65e-2
LINC01315	6.83	6.55	0.28	0.236	14.03	6.39e-9	4.16e-8
DNPEP	7.48	7.20	0.28	0.334	14.03	7.56e-11	7.26e-10
VKORC1	7.73	7.45	0.28	0.578	14.014	8.89e-11	8.35e-10
BRD1	6.77	6.49	0.28	0.264	14.011	1.44e-3	3.96e-3
RBM4B	6.79	6.51	0.28	0.201	14.001	6.49e-5	2.19e-4
MICALL1	6.72	6.44	0.28	0.16	13.956	1.04e-5	4.02e-5
DNAJC4	7.33	7.05	0.28	0.516	13.905	7.51e-15	1.92e-13
AKR1A1	7.35	7.08	0.28	0.438	13.899	6.13e-9	4.01e-8
TMEM59	7.95	7.67	0.28	0.317	13.891	3.10e-11	3.24e-10
PTOV1	7.44	7.16	0.28	0.295	13.876	1.61e-3	4.40e-3
COMMD2	7.65	7.37	0.28	0.227	13.871	2.45e-3	6.45e-3
NUDT22	7.72	7.45	0.28	0.366	13.864	9.48e-16	3.06e-14
CAMK2G	7.13	6.85	0.28	0.309	13.862	2.95e-8	1.72e-7

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
TECR	9.18	8.91	0.28	0.383	13.814	1.19e-10	1.08e-9
ECT2	7.11	6.83	0.28	0.46	13.814	1.05e-4	3.42e-4
GAS6	6.81	6.54	0.28	0.259	13.807	8.94e-4	2.53e-3
LSG1	6.80	6.52	0.28	0.229	13.802	1.72e-4	5.42e-4
MTX1	7.31	7.03	0.28	0.482	13.781	7.38e-12	8.83e-11
YIPF3	7.52	7.24	0.28	0.308	13.779	5.37e-5	1.83e-4
RAB3IP	6.82	6.55	0.28	0.257	13.776	1.07e-3	2.98e-3
MTFP1	7.11	6.83	0.28	0.399	13.743	4.80e-8	2.69e-7
PRSS8	7.08	6.80	0.28	0.315	13.733	4.48e-7	2.15e-6
DYNC2I2	8.46	8.18	0.27	0.256	13.713	1.95e-3	5.24e-3
CD59	8.81	8.54	0.27	0.234	13.7	7.19e-12	8.64e-11
LINC01278	7.48	7.20	0.27	0.419	13.688	1.65e-7	8.50e-7
TMEM237	6.99	6.72	0.27	0.244	13.647	5.40e-3	1.32e-2
MVB12A	7.56	7.28	0.27	0.405	13.644	1.62e-6	7.05e-6
ZNF639	6.69	6.42	0.27	0.145	13.616	5.81e-4	1.69e-3
ERP29	8.97	8.70	0.27	0.298	13.615	4.57e-8	2.57e-7
SGCE	7.02	6.75	0.27	0.303	13.587	1.55e-3	4.24e-3
CCDC167	7.17	6.90	0.27	0.329	13.584	2.50e-3	6.58e-3
CKAP4	7.23	6.96	0.27	0.28	13.582	4.59e-2	9.46e-2
PRADC1	6.75	6.48	0.27	0.218	13.568	2.69e-5	9.63e-5
FKBP8	8.42	8.15	0.27	0.218	13.567	6.54e-7	3.06e-6
MORN3	6.89	6.61	0.27	0.3	13.564	4.43e-14	9.28e-13
PCNA	8.80	8.53	0.27	0.401	13.547	6.56e-10	5.17e-9
MCM3AP	6.71	6.44	0.27	0.196	13.546	1.14e-2	2.62e-2
TMEM223	7.35	7.08	0.27	0.515	13.542	1.10e-17	6.74e-16
ABHD5	7.03	6.76	0.27	0.36	13.526	4.78e-5	1.65e-4

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
TRIOBP	7.40	7.13	0.27	0.285	13.513	2.57e-2	5.57e-2
NUBP2	7.30	7.03	0.27	0.441	13.512	3.09e-9	2.14e-8
DDX17	8.39	8.12	0.27	0.249	13.512	2.32e-6	9.85e-6
MCF2L	6.69	6.42	0.27	0.142	13.496	3.26e-2	6.91e-2
NEU1	6.91	6.65	0.27	0.255	13.484	4.34e-5	1.50e-4
AP1B1	7.12	6.85	0.27	0.286	13.482	4.43e-2	9.16e-2
	6.72	6.45	0.27	0.279	13.473	8.69e-7	3.96e-6
CCDC34	8.53	8.26	0.27	0.268	13.459	3.54e-2	7.44e-2
SFT2D2	7.92	7.65	0.27	0.336	13.458	4.69e-7	2.24e-6
INTS6	6.81	6.54	0.27	0.305	13.445	1.87e-4	5.87e-4
TMED10	7.53	7.27	0.27	0.216	13.403	5.13e-4	1.51e-3
SPOP	6.81	6.54	0.27	0.261	13.395	1.29e-11	1.47e-10
HCG11	6.79	6.52	0.27	0.226	13.378	2.12e-2	4.67e-2
SPG21	7.20	6.93	0.27	0.313	13.367	2.76e-3	7.22e-3
CDC16	7.20	6.94	0.27	0.306	13.36	1.86e-2	4.14e-2
C1orf35	7.99	7.72	0.27	0.312	13.353	6.30e-8	3.47e-7
PNKD	8.14	7.87	0.27	0.339	13.351	4.64e-3	1.15e-2
LHPP	6.74	6.47	0.27	0.199	13.347	2.55e-5	9.17e-5
SSR2	8.90	8.63	0.27	0.262	13.342	6.18e-8	3.41e-7
ALG14	6.79	6.53	0.27	0.199	13.34	5.78e-8	3.20e-7
CORO1B	8.10	7.84	0.27	0.235	13.326	8.64e-4	2.45e-3
PLAAT4	6.68	6.42	0.27	0.145	13.313	1.23e-9	9.15e-9
MYDGF	7.67	7.40	0.27	0.271	13.296	7.76e-6	3.05e-5
TLCD4	6.70	6.44	0.27	0.151	13.279	1.60e-3	4.35e-3
TMEM106C	7.80	7.53	0.27	0.369	13.265	4.63e-9	3.09e-8
TMEM9B	6.87	6.60	0.27	0.269	13.258	2.92e-9	2.04e-8

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
DRG2	6.84	6.57	0.27	0.225	13.253	1.19e-5	4.53e-5
P4HB	8.65	8.39	0.27	0.214	13.251	3.13e-6	1.30e-5
GPR180	7.03	6.76	0.27	0.268	13.25	1.30e-3	3.59e-3
PARD6A	6.69	6.42	0.27	0.162	13.249	2.99e-6	1.25e-5
EXD3	6.75	6.48	0.27	0.204	13.238	3.23e-3	8.37e-3
MMS22L	6.87	6.60	0.27	0.26	13.234	7.98e-5	2.65e-4
CIRBP	8.05	7.78	0.27	0.36	13.232	3.85e-6	1.58e-5
SCCPDH	6.86	6.59	0.27	0.26	13.227	5.05e-10	4.05e-9
H1-10	8.22	7.96	0.26	0.358	13.223	1.53e-5	5.71e-5
MTX2	7.25	6.98	0.26	0.334	13.21	3.26e-5	1.15e-4
ZNF576	6.79	6.52	0.26	0.222	13.198	3.32e-5	1.17e-4
MRPL28	7.40	7.14	0.26	0.506	13.189	5.00e-5	1.71e-4
GPAA1	7.90	7.63	0.26	0.269	13.179	1.21e-4	3.92e-4
TOR3A	7.37	7.10	0.26	0.333	13.165	1.75e-5	6.46e-5
POLR2E	7.83	7.57	0.26	0.274	13.156	2.52e-9	1.79e-8
HIBCH	6.91	6.64	0.26	0.302	13.153	4.70e-10	3.80e-9
CYB5A	7.46	7.20	0.26	0.446	13.148	3.20e-7	1.57e-6
ITSN2	6.94	6.68	0.26	0.307	13.141	1.16e-3	3.21e-3
LY6K	7.97	7.70	0.26	0.427	13.122	3.63e-7	1.77e-6
THOC6	6.84	6.57	0.26	0.273	13.119	1.65e-2	3.69e-2
MANF	7.70	7.44	0.26	0.427	13.111	5.28e-3	1.29e-2
NFYB	7.07	6.81	0.26	0.295	13.109	5.59e-3	1.36e-2
SUCLG1	7.53	7.27	0.26	0.516	13.107	9.76e-14	1.88e-12
MGAT4A	7.11	6.85	0.26	0.303	13.091	3.42e-5	1.21e-4
ACOT13	7.45	7.19	0.26	0.428	13.091	2.42e-12	3.30e-11
EBP	8.29	8.03	0.26	0.334	13.084	1.10e-8	6.91e-8

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
SECISBP2	7.27	7.01	0.26	0.295	13.076	7.81e-3	1.85e-2
CCHCR1	6.99	6.73	0.26	0.317	13.067	5.92e-4	1.72e-3
TM7SF3	6.83	6.56	0.26	0.256	13.064	3.28e-3	8.50e-3
MAPK3	6.73	6.47	0.26	0.178	13.06	3.18e-3	8.23e-3
GMPPA	6.80	6.54	0.26	0.233	13.033	3.66e-7	1.78e-6
MYO9B	7.09	6.83	0.26	0.288	13.027	1.06e-5	4.09e-5
PHF3	7.20	6.94	0.26	0.286	13.02	1.66e-3	4.50e-3
TET1	7.18	6.92	0.26	0.302	13.02	7.35e-4	2.10e-3
PIK3R2	6.84	6.58	0.26	0.272	13.016	5.89e-3	1.43e-2
ATF7	6.90	6.64	0.26	0.258	13.004	4.59e-8	2.58e-7
PODXL2	7.34	7.08	0.26	0.45	12.994	2.36e-13	4.14e-12
LYNX1	6.68	6.42	0.26	0.16	12.985	8.03e-6	3.15e-5
RAB4A	7.43	7.17	0.26	0.441	12.983	9.88e-9	6.26e-8
HNMT	6.84	6.58	0.26	0.309	12.954	4.74e-13	7.61e-12
SEC22B	6.89	6.64	0.26	0.321	12.95	1.93e-8	1.17e-7
FJX1	7.57	7.32	0.26	0.257	12.942	7.32e-6	2.90e-5
SPDEF	6.74	6.48	0.26	0.209	12.941	4.75e-9	3.16e-8
PBX1	7.77	7.51	0.26	0.379	12.926	1.43e-5	5.35e-5
GMPR2	6.82	6.56	0.26	0.255	12.918	3.84e-5	1.35e-4
ZBTB1	6.70	6.44	0.26	0.159	12.877	2.68e-2	5.79e-2
UNC50	6.86	6.60	0.26	0.246	12.875	4.90e-6	1.99e-5
SUGP2	7.61	7.35	0.26	0.276	12.871	5.30e-4	1.55e-3
ENOSF1	7.85	7.59	0.26	0.258	12.853	2.75e-8	1.61e-7
TMEM208	7.42	7.16	0.26	0.447	12.852	8.84e-9	5.63e-8
PPIL2	6.68	6.42	0.26	0.159	12.846	2.99e-5	1.07e-4
ZNF518A	7.03	6.77	0.26	0.362	12.838	1.87e-6	8.08e-6

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
AAMP	7.59	7.33	0.26	0.434	12.822	3.57e-8	2.05e-7
PAFAH1B3	8.67	8.42	0.26	0.4	12.822	2.41e-8	1.42e-7
AMOT	6.68	6.43	0.26	0.154	12.814	2.14e-3	5.70e-3
RPN2	8.25	8.00	0.26	0.295	12.808	2.27e-3	6.02e-3
COPE	9.11	8.86	0.26	0.283	12.787	4.26e-7	2.05e-6
TANC1	6.75	6.50	0.26	0.202	12.787	4.45e-3	1.10e-2
CALD1	8.96	8.71	0.26	0.239	12.786	1.48e-6	6.47e-6
SNRNP25	7.21	6.96	0.26	0.503	12.767	1.46e-11	1.64e-10
NCBP2	7.25	6.99	0.26	0.294	12.765	7.50e-7	3.46e-6
ATP6AP1	7.85	7.60	0.26	0.285	12.758	1.40e-3	3.84e-3
EFR3A	7.16	6.90	0.26	0.363	12.754	5.81e-5	1.97e-4
MYL6B	7.61	7.35	0.26	0.447	12.754	1.47e-7	7.65e-7
MPHOSPH6	7.32	7.06	0.26	0.444	12.75	2.02e-10	1.76e-9
APP	8.89	8.64	0.26	0.401	12.75	2.76e-10	2.36e-9
ZFAND6	7.02	6.76	0.26	0.282	12.745	3.62e-4	1.09e-3
MFGE8	6.98	6.73	0.26	0.255	12.741	2.67e-3	7.00e-3
BPNT1	6.87	6.61	0.26	0.238	12.738	8.22e-9	5.27e-8
HM13	7.44	7.19	0.26	0.367	12.732	1.69e-8	1.03e-7
SLC39A7	7.67	7.41	0.26	0.282	12.732	2.67e-5	9.58e-5
HYAL2	6.77	6.52	0.25	0.247	12.722	4.38e-5	1.52e-4
STX17	6.84	6.59	0.25	0.334	12.718	1.25e-6	5.54e-6
SCP2	7.40	7.15	0.25	0.408	12.706	2.83e-7	1.40e-6
GTPBP8	6.86	6.60	0.25	0.236	12.693	2.69e-5	9.62e-5
MED13	6.71	6.46	0.25	0.182	12.686	2.33e-3	6.17e-3
SQLE	7.83	7.58	0.25	0.169	12.68	2.09e-4	6.51e-4
TMEM214	6.80	6.55	0.25	0.294	12.678	4.67e-7	2.24e-6

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
SDR39U1	6.78	6.53	0.25	0.267	12.675	6.93e-8	3.79e-7
SRA1	7.08	6.82	0.25	0.431	12.671	2.36e-15	6.90e-14
PMVK	7.99	7.74	0.25	0.365	12.663	3.69e-10	3.07e-9
VEZF1	6.73	6.47	0.25	0.186	12.657	2.19e-5	7.98e-5
MIA3	7.01	6.75	0.25	0.337	12.652	8.01e-4	2.28e-3
PGAP2	6.70	6.44	0.25	0.142	12.646	5.45e-3	1.33e-2
EIF4A2	8.12	7.87	0.25	0.148	12.639	3.79e-2	7.94e-2
HNRNPH2	7.02	6.76	0.25	0.298	12.629	1.07e-2	2.47e-2
ACTR6	7.19	6.93	0.25	0.306	12.627	2.70e-3	7.08e-3
HSPA14	7.15	6.90	0.25	0.325	12.606	9.36e-5	3.07e-4
PDIA6	9.21	8.96	0.25	0.26	12.605	1.80e-9	1.31e-8
PPP1R1A	6.67	6.42	0.25	0.152	12.599	1.23e-23	5.31e-21
RNF187	8.17	7.92	0.25	0.318	12.592	6.08e-6	2.43e-5
MPC2	8.30	8.05	0.25	0.463	12.559	2.18e-19	2.47e-17
SSH1	6.68	6.43	0.25	0.124	12.531	2.38e-2	5.19e-2
B4GALT3	6.85	6.60	0.25	0.232	12.524	6.77e-5	2.27e-4
NT5DC1	7.03	6.78	0.25	0.353	12.522	1.80e-4	5.67e-4
TRPV4	6.72	6.47	0.25	0.228	12.517	1.55e-7	8.01e-7
SC5D	6.80	6.55	0.25	0.254	12.517	1.56e-9	1.14e-8
HSPA5	8.56	8.31	0.25	0.41	12.509	1.61e-15	4.88e-14
CNOT8	7.02	6.77	0.25	0.286	12.503	9.32e-6	3.62e-5
FAM32A	7.43	7.18	0.25	0.385	12.46	1.44e-5	5.38e-5
PSMC2	7.55	7.30	0.25	0.391	12.46	5.81e-8	3.21e-7
SYNJ2BP	6.70	6.45	0.25	0.171	12.448	1.65e-3	4.48e-3
NELFE	7.47	7.22	0.25	0.542	12.444	1.69e-11	1.88e-10
MAN1B1	6.85	6.60	0.25	0.315	12.438	8.85e-5	2.92e-4

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
VDR	6.84	6.59	0.25	0.354	12.438	6.88e-5	2.31e-4
STX4	6.84	6.59	0.25	0.239	12.431	8.18e-5	2.71e-4
GTPBP6	6.88	6.63	0.25	0.293	12.43	3.30e-10	2.77e-9
MED4	6.93	6.68	0.25	0.304	12.414	1.91e-8	1.15e-7
ATP13A2	7.36	7.11	0.25	0.212	12.411	1.16e-4	3.74e-4
MED30	6.79	6.54	0.25	0.254	12.393	1.10e-2	2.54e-2
ADIPOR1	7.41	7.16	0.25	0.448	12.392	6.54e-9	4.25e-8
PLPP2	8.39	8.14	0.25	0.245	12.387	1.62e-6	7.06e-6
RRP7A	7.61	7.36	0.25	0.394	12.385	5.89e-23	2.10e-20
ZNF695	6.70	6.46	0.25	0.178	12.384	1.67e-5	6.20e-5
SNX7	6.92	6.67	0.25	0.262	12.382	1.52e-5	5.67e-5
GPR160	6.67	6.42	0.25	0.124	12.377	7.67e-4	2.19e-3
KIAA0319L	6.73	6.49	0.25	0.226	12.374	5.00e-4	1.47e-3
CBR1	7.92	7.68	0.25	0.399	12.37	2.97e-11	3.13e-10
TMEM38B	6.85	6.60	0.25	0.226	12.365	6.65e-7	3.10e-6
FN3KRP	7.02	6.77	0.25	0.336	12.363	1.16e-13	2.19e-12
DCPS	7.57	7.32	0.25	0.41	12.353	2.82e-5	1.01e-4
DNPH1	7.91	7.66	0.25	0.314	12.346	1.39e-6	6.10e-6
TRIM59	6.68	6.43	0.25	0.158	12.342	2.78e-2	5.98e-2
LAMTOR4	7.96	7.71	0.25	0.421	12.34	8.52e-3	2.00e-2
EFNA4	6.84	6.60	0.25	0.231	12.335	1.50e-4	4.80e-4
RPIA	6.78	6.54	0.25	0.253	12.322	5.32e-4	1.56e-3
STX5	6.89	6.65	0.25	0.296	12.321	9.62e-14	1.86e-12
SLC39A1	7.91	7.67	0.25	0.284	12.317	7.67e-11	7.35e-10
DYNLT1	8.18	7.93	0.25	0.324	12.313	1.72e-13	3.12e-12
DPY19L3	6.76	6.51	0.25	0.256	12.309	7.24e-3	1.72e-2

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
RIPK4	6.84	6.60	0.25	0.447	12.308	1.73e-13	3.14e-12
MMAB	7.47	7.23	0.25	0.472	12.293	3.97e-7	1.92e-6
LRRC42	6.96	6.71	0.25	0.283	12.289	2.35e-4	7.28e-4
BACE1	7.04	6.80	0.25	0.306	12.284	3.35e-3	8.65e-3
CAPN1	7.87	7.63	0.25	0.258	12.279	3.63e-7	1.77e-6
HMCES	7.11	6.86	0.25	0.318	12.271	1.19e-3	3.30e-3
HACD3	7.97	7.72	0.25	0.224	12.268	1.03e-5	3.97e-5
YPEL3	6.67	6.42	0.25	0.134	12.257	2.17e-2	4.77e-2
DPM2	7.37	7.12	0.25	0.57	12.255	9.13e-13	1.38e-11
ACADM	6.99	6.75	0.25	0.293	12.253	9.13e-6	3.55e-5
NUP58	7.09	6.85	0.25	0.266	12.237	1.96e-6	8.44e-6
HMGXB4	7.63	7.39	0.25	0.346	12.226	3.89e-5	1.36e-4
NANOS1	6.69	6.45	0.24	0.172	12.224	1.38e-2	3.13e-2
MCF2L-AS1	6.88	6.64	0.24	0.32	12.219	7.84e-17	3.54e-15
LAMTOR1	8.32	8.08	0.24	0.468	12.217	9.24e-16	2.99e-14
GRB14	6.67	6.43	0.24	0.131	12.21	8.20e-5	2.72e-4
CTSD	7.90	7.65	0.24	0.317	12.206	1.08e-9	8.11e-9
DHCR24	7.73	7.48	0.24	0.19	12.188	1.68e-3	4.56e-3
HDGFL2	7.02	6.78	0.24	0.301	12.181	3.66e-4	1.10e-3
GJC1	6.70	6.46	0.24	0.215	12.176	3.05e-4	9.29e-4
VDAC3	7.49	7.25	0.24	0.425	12.176	1.94e-12	2.71e-11
CCDC88A	7.15	6.91	0.24	0.351	12.163	7.63e-8	4.14e-7
ATAD1	6.92	6.68	0.24	0.296	12.157	2.21e-4	6.88e-4
TENT4A	6.92	6.68	0.24	0.298	12.155	8.14e-4	2.32e-3
TUSC3	7.27	7.03	0.24	0.385	12.154	2.70e-7	1.34e-6
RNF5	6.99	6.74	0.24	0.322	12.152	1.36e-9	1.01e-8

Table 6.7 continued from previous page

HGNC Gene Symbol	Average Exp T1	Average Exp T2	Average Fold Change	SD Fold Change	Cum Fold Change	p - value	FDR
TIMM29	7.05	6.81	0.24	0.276	12.147	1.36e-6	5.98e-6
UBA6	6.82	6.58	0.24	0.318	12.141	5.95e-4	1.73e-3
MRPL46	6.87	6.62	0.24	0.258	12.132	5.63e-10	4.48e-9
RPL39L	7.32	7.08	0.24	0.424	12.13	2.04e-9	1.47e-8
PALM2AKAP2	7.46	7.22	0.24	0.213	12.122	1.58e-2	3.55e-2
RBM3	7.69	7.45	0.24	0.337	12.12	2.18e-9	1.56e-8
TM2D3	6.66	6.42	0.24	0.129	12.114	2.36e-9	1.67e-8
LITAF	7.71	7.46	0.24	0.378	12.111	2.52e-9	1.78e-8
MIR99AHG	6.68	6.44	0.24	0.154	12.108	1.06e-3	2.96e-3
EPS15	6.82	6.58	0.24	0.263	12.107	2.47e-4	7.61e-4
HMGCS1	6.68	6.44	0.24	0.122	12.105	1.70e-4	5.37e-4
DDOST	7.63	7.38	0.24	0.302	12.087	1.72e-5	6.37e-5
SAC3D1	8.33	8.09	0.24	0.283	12.079	8.32e-4	2.36e-3
SLC25A26	6.71	6.46	0.24	0.182	12.075	6.81e-5	2.28e-4
GK5	6.84	6.60	0.24	0.287	12.064	3.04e-3	7.89e-3
TUBGCP3	6.84	6.60	0.24	0.221	12.05	3.77e-4	1.13e-3
RRAGA	7.39	7.15	0.24	0.329	12.043	4.86e-8	2.72e-7
GOLT1B	6.68	6.44	0.24	0.145	12.041	1.02e-2	2.38e-2
GABARAP	8.52	8.28	0.24	0.478	12.041	2.55e-12	3.45e-11
ZNF519	6.73	6.49	0.24	0.266	12.038	2.38e-5	8.60e-5

Table 6.8. Gene Ontology (GO) analysis results against KEGG pathways on genes up-regulated in tolerant clones with clonal fitness respect to tolerant clones without clonal fitness. **(A)** Category ID **(B)** Term name **(C)** Number of up-regulated genes in tolerant clones with clonal fitness contained in the given term; **(D)** Percentage of up-regulated genes in clones with clonal fitness in the intesection with the gene set **(E)** p-value of Fisher's Exact of the enrichment.
Only significant (FDR < 0.1) are shown.

Category ID	Term Name	N° genes inters	% genes inters	p-value
hsa01100	Metabolic pathways	67	12.27	5.160e-4
hsa04141	Protein processing in endoplasmic reticulum	21	3.85	1.120e-7
hsa04142	Lysosome	15	2.75	2.940e-5
hsa05208	Chemical carcinogenesis - reactive oxygen species	14	2.56	1.382e-2
hsa05166	Human T-cell leukemia virus 1 infection	13	2.38	2.975e-2
hsa05225	Hepatocellular carcinoma	12	2.20	1.024e-2
hsa04210	Apoptosis	10	1.83	1.818e-2
hsa05161	Hepatitis B	10	1.83	4.815e-2
hsa00513	Various types of N-glycan biosynthesis	9	1.65	1.350e-5
hsa00510	N-Glycan biosynthesis	9	1.65	8.890e-5
hsa01212	Fatty acid metabolism	8	1.47	1.281e-3
hsa05218	Melanoma	7	1.28	1.849e-2
hsa00980	Metabolism of xenobiotics by cytochrome P450	7	1.28	2.633e-2
hsa01521	EGFR tyrosine kinase inhibitor resistance	7	1.28	2.782e-2
hsa05210	Colorectal cancer	7	1.28	3.985e-2
hsa00900	Terpenoid backbone biosynthesis	6	1.10	4.500e-4
hsa00280	Valine, leucine and isoleucine degradation	6	1.10	1.267e-2
hsa05213	Endometrial cancer	6	1.10	2.683e-2
hsa00100	Steroid biosynthesis	5	0.92	2.398e-3
hsa01040	Biosynthesis of unsaturated fatty acids	5	0.92	7.399e-3
hsa05216	Thyroid cancer	5	0.92	2.219e-2

Table 6.8 continued from previous page

Category ID	Term Name	N° genes inters	% genes inters	p-value
hsa00340	Histidine metabolism	4	0.73	2.530e-2

Bibliography

[1] Jean-Christophe Marine, Sarah-Jane Dawson, and Mark A. Dawson. Non-genetic mechanisms of therapeutic resistance in cancer. *Nature Reviews Cancer*, 20:743–756, 12 2020.

[2] Michele Balma, Virginia Liberini, Manuela Racca, Riccardo Laudicella, Matteo Bauckneht, Ambra Buschiazzo, Daniele Giovanni Nicolotti, Simona Peano, Andrea Bianchi, Giovanni Albano, Natale Quartuccio, Ronan Abgral, Silvia Daniela Morbelli, Calogero D'Alessandria, Enzo Terreno, Martin William Huellner, Alberto Papaleo, and Désirée Deandreis. Non-conventional and investigational pet radiotracers for breast cancer: A systematic review. *Frontiers in Medicine*, 9, 4 2022.

[3] Jieqiong Liu, Kai Chen, Wen Jiang, Kai Mao, Shunrong Li, Min Ji Kim, Qiang Liu, and Lisa K. Jacobs. Chemotherapy response and survival of inflammatory breast cancer by hormone receptor- and her2-defined molecular subtypes approximation: an analysis from the national cancer database. *Journal of Cancer Research and Clinical Oncology*, 143:161–168, 1 2017.

[4] L. Gerratana, D. Basile, G. Buono, S. De Placido, M. Giuliano, S. Minichillo, A. Coinu, F. Martorana, I. De Santo, L. Del Mastro, M. De Laurentiis, F. Puglisi, and G. Arpino. Androgen receptor in triple negative breast cancer: A potential target for the targetless subtype. *Cancer Treatment Reviews*, 68:102–110, 7 2018.

[5] Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, John J Trombetta, David A Weitz, Joshua R Sanes, Alex K Shalek, Aviv Regev, and Steven A McCarroll. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161:1202–1214, 5 2015.

[6] Sadie VanHorn and Samantha A. Morris. Next-generation lineage tracing and fate mapping to interrogate development. *Developmental Cell*, 56:7–21, 1 2021.

[7] https://www.cancerrxgene.org.

[8] Alfredo A Kalaitzis and Neil D Lawrence. The theory of probability. *Oxford Classic Texts in the Physical Sciences*, 1939.

[9] Ian A. Cree and Peter Charlton. Molecular chess? hallmarks of anti-cancer drug resistance. *BMC Cancer*, 17:10, 12 2017.

[10] Caitriona Holohan, Sandra Van Schaeybroeck, Daniel B. Longley, and Patrick G. Johnston. Cancer drug resistance: an evolving paradigm. *Nature Reviews Cancer*, 13:714–726, 10 2013.

[11] Ruth Nussinov, Chung-Jung Tsai, and Hyunbum Jang. Anticancer drug resistance: An update and perspective. *Drug Resistance Updates*, 59:100796, 12 2021.

[12] Corbin E. Meacham and Sean J. Morrison. Tumour heterogeneity and cancer cell plasticity. *Nature*, 501:328–337, 9 2013.

[13] Giampaolo Bianchini, Justin M Balko, Ingrid A Mayer, Melinda E Sanders, and Luca Gianni. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nature reviews. Clinical oncology*, 13:674–690, 11 2016.

[14] William D. Foulkes, Ian E. Smith, and Jorge S. Reis-Filho. Triple-negative breast cancer. *New England Journal of Medicine*, 363:1938–1948, 11 2010.

[15] Li Yin, Jiang-Jie Duan, Xiu-Wu Bian, and Shi cang Yu. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research*, 22:61, 12 2020.

[16] Alexandre F. Aissa, Abul B. M. M. K. Islam, Majd M. Ariss, Cammille C. Go, Alexandra E. Rader, Ryan D. Conrardy, Alexa M. Gajda, Carlota Rubio-Perez, Klara Valyi-Nagy, Mary Pasquinelli, Lawrence E. Feldman, Stefan J. Green, Nuria Lopez-Bigas, Maxim V. Frolov, and Elizaveta V. Benevolenskaya. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. *Nature Communications*, 12:1628, 3 2021.

[17] Paul Savage, Alexis Blanchet-Cohen, Timothée Revil, Dunarel Badescu, Sadiq M I Saleh, Yu-Chang Wang, Dongmei Zuo, Leah Liu, Nicholas R Bertos, Valentina Munoz-Ramos, Mark Basik, Kevin Petrecca, Jamil Asselah, Sarkis Meterissian, Marie-Christine Guiot, Atilla Omeroglu, Claudia L

Kleinman, Morag Park, and Jiannis Ragoussis. A targetable egfr-dependent tumor-initiating program in breast cancer. *Cell reports*, 21:1140–1149, 10 2017.

[18] Vijay G Sankaran, Jonathan S Weissman, and Leonard I Zon. Cellular barcoding to decipher clonal dynamics in disease. *Science (New York, N.Y.)*, 378:eabm5874, 10 2022.

[19] Yaara Oren, Michael Tsabar, Michael S. Cuoco, Liat Amir-Zilberstein, Heidie F. Cabanos, Jan-Christian Hütter, Bomiao Hu, Pratiksha I. Thakore, Marcin Tabaka, Charles P. Fulco, William Colgan, Brandon M. Cuevas, Sara A. Hurvitz, Dennis J. Slamon, Amy Deik, Kerry A. Pierce, Clary Clish, Aaron N. Hata, Elma Zaganjor, Galit Lahav, Katerina Politi, Joan S. Brugge, and Aviv Regev. Cycling cancer persister cells arise from lineages with distinct programs. *Nature*, 596:576–582, 8 2021.

[20] Kamen P Simeonov, China N Byrns, Megan L Clark, Robert J Norgard, Beth Martin, Ben Z Stanger, Jay Shendure, Aaron McKenna, and Christopher J Lengner. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid emt states. *Cancer cell*, 39:1150–1162.e9, 8 2021.

[21] Marlous Hoogstraat, Esther H Lips, Isabel Mayayo-Peralta, Lennart Mulder, Petra Kristel, Ingrid van der Heijden, Stefano Annunziato, Maartje van Seijen, Petra M Nederlof, Gabe S Sonke, Wilbert Zwart, Jelle Wesseling, and Lodewyk F A Wessels. Comprehensive characterization of pre- and post-treatment samples of breast cancer reveal potential mechanisms of chemotherapy resistance. *NPJ breast cancer*, 8:60, 5 2022.

[22] Yi-Zhou Jiang, Ding Ma, Chen Suo, Jinxiu Shi, Mengzhu Xue, Xin Hu, Yi Xiao, Ke-Da Yu, Yi-Rong Liu, Ying Yu, Yuanting Zheng, Xiangnan Li, Chenhui Zhang, Pengchen Hu, Jing Zhang, Qi Hua, Jiyang Zhang, Wanwan Hou, Luyao Ren, Ding Bao, Bingying Li, Jingcheng Yang, Ling Yao, Wen-Jia Zuo, Shen Zhao, Yue Gong, Yi-Xing Ren, Ya-Xin Zhao, Yun-Song Yang, Zhenmin Niu, Zhi-Gang Cao, Daniel G Stover, Claire Verschraegen, Virginia Kaklamani, Anneleen Daemen, John R Benson, Kazuaki Takabe, Fan Bai, Da-Qiang Li, Peng Wang, Leming Shi, Wei Huang, and Zhi-Ming Shao. Genomic and transcriptomic landscape of triple-negative breast cancers: Subtypes and treatment strategies. *Cancer cell*, 35:428–440.e5, 3 2019.

[23] Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, 490:61–70, 10 2012.

[24] C M Perou, T Sørlie, M B Eisen, M van de Rijn, S S Jeffrey, C A Rees, J R Pollack, D T Ross, H Johnsen, L A Akslen, O Fluge, A Pergamenschikov,

C Williams, S X Zhu, P E Lønning, A L Børresen-Dale, P O Brown, and D Botstein. Molecular portraits of human breast tumours. *Nature*, 406:747–52, 8 2000.

[25] Remah Ali and Michael K Wendt. The paradoxical functions of egfr during breast cancer progression. *Signal transduction and targeted therapy*, 2:16042–, 2017.

[26] Rohit Bhargava, William L Gerald, Allan R Li, Qiulu Pan, Priti Lal, Marc Ladanyi, and Beiyun Chen. Egfr gene amplification in breast cancer: correlation with epidermal growth factor receptor mrna and protein expression and her-2 status and absence of egfr-activating mutations. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 18:1027–33, 8 2005.

[27] Heae Surng Park, Min Hye Jang, Eun Joo Kim, Hyun Jeong Kim, Hee Jin Lee, Yu Jung Kim, Jee Hyun Kim, Eunyoung Kang, Sung-Won Kim, In Ah Kim, and So Yeon Park. High egfr gene copy number predicts poor outcome in triple-negative breast cancer. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 27:1212–22, 9 2014.

[28] Marc Tischkowitz, Jean-Sébastien Brunet, Louis R Bégin, David G Huntsman, Maggie C U Cheang, Lars A Akslen, Torsten O Nielsen, and William D Foulkes. Use of immunohistochemical markers can refine prognosis in triple negative breast cancer. *BMC cancer*, 7:134, 7 2007.

[29] Ricardo Costa, Ami N Shah, Cesar A Santa-Maria, Marcelo R Cruz, Devalingam Mahalingam, Benedito A Carneiro, Young Kwang Chae, Massimo Cristofanilli, William J Gradishar, and Francis J Giles. Targeting epidermal growth factor receptor in triple negative breast cancer: New discoveries and practical insights for drug development. *Cancer treatment reviews*, 53:111–119, 2 2017.

[30] Federico Cappuzzo, Giovanna Finocchiaro, Francesco Grossi, Paolo Bidoli, Adolfo Favaretto, Antonio Marchetti, Maria Luisa Valente, Agnieszka Cseh, Laura Clementi, Dan Massey, and Armando Santoro. Phase ii study of afatinib, an irreversible erbb family blocker, in egfr fish-positive non-small-cell lung cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, 10:665–72, 4 2015.

[31] Malcolm J Moore, David Goldstein, John Hamm, Arie Figer, Joel R Hecht, Steven Gallinger, Heather J Au, Pawel Murawa, David Walde, Robert A Wolff, Daniel Campos, Robert Lim, Keyue Ding, Gary Clark, Theodora

Voskoglou-Nomikos, Mieke Ptasynski, Wendy Parulekar, and National Cancer Institute of Canada Clinical Trials Group. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase iii trial of the national cancer institute of canada clinical trials group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 25:1960–6, 5 2007.

[32] Eric Van Cutsem, Marc Peeters, Salvatore Siena, Yves Humblet, Alain Hendlisz, Bart Neyns, Jean-Luc Canon, Jean-Luc Van Laethem, Joan Maurel, Gary Richardson, Michael Wolf, and Rafael G Amado. Open-label phase iii trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 25:1658–64, 5 2007.

[33] Jan B Vermorken, Ricard Mesia, Fernando Rivera, Eva Remenar, Andrzej Kawecki, Sylvie Rottey, Jozsef Erfan, Dmytro Zabolotnyy, Heinz-Roland Kienzer, Didier Cupissol, Frederic Peyrade, Marco Benasso, Ihor Vynnychenko, Dominique De Raucourt, Carsten Bokemeyer, Armin Schueler, Nadia Amellal, and Ricardo Hitt. Platinum-based chemotherapy plus cetuximab in head and neck cancer. *The New England journal of medicine*, 359:1116–27, 9 2008.

[34] José Baselga and Carlos L Arteaga. Critical update and emerging trends in epidermal growth factor receptor targeting in cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 23:2445–59, 4 2005.

[35] Giampaolo Bianchini, Carmine De Angelis, Luca Licata, and Luca Gianni. Treatment landscape of triple-negative breast cancer — expanded options, evolving needs. *Nature Reviews Clinical Oncology*, 19:91–113, 2 2022.

[36] Lisa A Carey, Hope S Rugo, P Kelly Marcom, Erica L Mayer, Francisco J Esteva, Cynthia X Ma, Minetta C Liu, Anna Maria Storniolo, Mothaffar F Rimawi, Andres Forero-Torres, Antonio C Wolff, Timothy J Hobday, Anastasia Ivanova, Wing-Keung Chiu, Madlyn Ferraro, Emily Burrows, Philip S Bernard, Katherine A Hoadley, Charles M Perou, and Eric P Winer. Tbcrc 001: randomized phase ii study of cetuximab in combination with carboplatin in stage iv triple-negative breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 30:2615–23, 7 2012.

[37] Katsuya Nakai, Mien-Chie Hung, and Hirohito Yamaguchi. A perspective on anti-egfr therapies targeting triple-negative breast cancer. *American journal of cancer research*, 6:1609–23, 2016.

[38] Gunter von Minckwitz, Walter Jonat, Peter Fasching, Andreas du Bois, Ulrich Kleeberg, Hans-Joachim Lück, Erika Kettner, Jörn Hilfrich, Wolfgang Eiermann, Julie Torode, and Andreas Schneeweiss. A multicentre phase ii study on gefitinib in taxane- and anthracycline-pretreated metastatic breast cancer. *Breast cancer research and treatment*, 89:165–72, 1 2005.

[39] Jingyi He, Zhengrong Huang, Linzhi Han, Yan Gong, and Conghua Xie. Mechanisms and management of 3rd-generation egfr-tki resistance in advanced non-small cell lung cancer (review). *International journal of oncology*, 59, 11 2021.

[40] Sohrab P. Shah, Andrew Roth, Rodrigo Goya, Arusha Oloumi, Gavin Ha, Yongjun Zhao, Gulisa Turashvili, Jiarui Ding, Kane Tse, Gholamreza Haffari, Ali Bashashati, Leah M. Prentice, Jaswinder Khattra, Angela Burleigh, Damian Yap, Virginie Bernard, Andrew McPherson, Karey Shumansky, Anamaria Crisan, Ryan Giuliany, Alireza Heravi-Moussavi, Jamie Rosner, Daniel Lai, Inanc Birol, Richard Varhol, Angela Tam, Noreen Dhalla, Thomas Zeng, Kevin Ma, Simon K. Chan, Malachi Griffith, Annie Moradian, S.-W. Grace Cheng, Gregg B. Morin, Peter Watson, Karen Gelmon, Stephen Chia, Suet-Feung Chin, Christina Curtis, Oscar M. Rueda, Paul D. Pharoah, Sambasivarao Damaraju, John Mackey, Kelly Hoon, Timothy Harkins, Vasisht Tadigotla, Mahvash Sigaroudinia, Philippe Gascard, Thea Tlsty, Joseph F. Costello, Irmtraud M. Meyer, Connie J. Eaves, Wyeth W. Wasserman, Steven Jones, David Huntsman, Martin Hirst, Carlos Caldas, Marco A. Marra, and Samuel Aparicio. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*, 486:395–399, 6 2012.

[41] J Filmus, M N Pollak, R Cailleau, and R N Buick. Mda-468, a human breast cancer cell line with a high number of epidermal growth factor (egf) receptors, has an amplified egf receptor gene and is growth inhibited by egf. *Biochemical and biophysical research communications*, 128:898–905, 4 1985.

[42] T Thomas, S Balabhadrapathruni, C R Gardner, J Hong, C A Faaland, and T J Thomas. Effects of epidermal growth factor on mda-mb-468 breast cancer cells: alterations in polyamine biosynthesis and the expression of p21/cip1/waf1. *Journal of cellular physiology*, 179:257–66, 6 1999.

[43] Behzad Mansoori, Ali Mohammadi, Sadaf Davudian, Solmaz Shirjang, and Behzad Baradaran. The different mechanisms of cancer drug resistance: A brief review. *Advanced Pharmaceutical Bulletin*, 7:339–348, 9 2017.

[44] Hiba Zahreddine and Katherine L. B. Borden. Mechanisms and insights into drug resistance in cancer. *Frontiers in Pharmacology*, 4, 2013.

[45] Talha Bin Emran, Asif Shahriar, Aar Rafi Mahmud, Tanjilur Rahman, Mehedy Hasan Abir, Mohd. Faijanur Rob Siddiquee, Hossain Ahmed, Nova Rahman, Firzan Nainu, Elly Wahyudin, Saikat Mitra, Kuldeep Dhama, Mahmoud M. Habiballah, Shafiul Haque, Ariful Islam, and Mohammad Mahmudul Hassan. Multidrug resistance in cancer: Understanding molecular mechanisms, immunoprevention and therapeutic approaches. *Frontiers in Oncology*, 12, 6 2022.

[46] Foram U. Vaidya, Abu Sufiyan Chhipa, Vinita Mishra, Vishal Kumar Gupta, Shiv Govind Rawat, Ajay Kumar, and Chandramani Pathak. Molecular and cellular paradigms of multidrug resistance in cancer. *Cancer Reports*, 5, 12 2022.

[47] Heinz Hammerlindl and Helmut Schaider. Tumor cell-intrinsic phenotypic plasticity facilitates adaptive cellular reprogramming driving acquired drug resistance. *Journal of Cell Communication and Signaling*, 12:133–141, 3 2018.

[48] Theodor H Lippert, Hans-Jörg Ruoff, and Manfred Volm. Intrinsic and acquired drug resistance in malignant tumors. the main reason for therapeutic failure. *Arzneimittel-Forschung*, 58:261–4, 2008.

[49] Jing Jin, Xu Wu, Jianhua Yin, Mingxing Li, Jing Shen, Jing Li, Yueshui Zhao, Qijie Zhao, Jingbo Wu, Qinglian Wen, Chi Hin Cho, Tao Yi, Zhangang Xiao, and Liping Qu. Identification of genetic mutations in cancer: Challenge and opportunity in the new era of targeted therapy. *Frontiers in Oncology*, 9, 4 2019.

[50] Xinming Li, Mingdong Li, Meiying Huang, Qianyi Lin, Qiuping Fang, Jianjiang Liu, Xiaohui Chen, Lin Liu, Xuliang Zhan, Huisi Shan, Deshuai Lu, Qinlan Li, Zesong Li, and Xiao Zhu. The multi-molecular mechanisms of tumor-targeted drug resistance in precision medicine. *Biomedicine Pharmacotherapy*, 150:113064, 6 2022.

[51] Xuan Wang, Haiyun Zhang, and Xiaozhuo Chen. Drug resistance and combating drug resistance in cancer. *Cancer drug resistance (Alhambra, Calif.)*, 2:141–160, 2019.

[52] Yuan Li, Zhenning Wang, Jaffer A. Ajani, and Shumei Song. Drug resistance and cancer stem cells. *Cell Communication and Signaling*, 19:19, 12 2021.

[53] Imoh S. Okon and Ming-Hui Zou. Mitochondrial ros and cancer drug resistance: Implications for therapy. *Pharmacological Research*, 100:170–174, 10 2015.

[54] Bruno Perillo, Marzia Di Donato, Antonio Pezone, Erika Di Zazzo, Pia Giovannelli, Giovanni Galasso, Gabriella Castoria, and Antimo Migliaccio. Ros in cancer therapy: the bright side of the moon. *Experimental Molecular Medicine*, 52:192–203, 2 2020.

[55] Zi-Ning Lei, Qin Tian, Qiu-Xu Teng, John N. D. Wurpel, Leli Zeng, Yihang Pan, and Zhe-Sheng Chen. Understanding and targeting resistance mechanisms in cancer. *MedComm*, 4, 6 2023.

[56] Viktoria Manstein, Chul Yang, Diane Richter, Natalia Delis, Vida Vafaizadeh, and Bernd Groner. Resistance of cancer cells to targeted therapies through the activation of compensating signaling loops. *Current Signal Transduction Therapy*, 8:193–202, 4 2014.

[57] Steven A Rosenzweig. Acquired resistance to drugs targeting receptor tyrosine kinases. *Biochemical pharmacology*, 83:1041–8, 4 2012.

[58] K G Paulson, V Voillet, M S McAfee, D S Hunter, F D Wagener, M Perdicchio, W J Valente, S J Koelle, C D Church, N Vandeven, H Thomas, A G Colunga, J G Iyer, C Yee, R Kulikauskas, D M Koelle, R H Pierce, J H Bielas, P D Greenberg, S Bhatia, R Gottardo, P Nghiem, and A G Chapuis. Acquired cancer resistance to combination immunotherapy from transcriptional loss of class i hla. *Nature communications*, 9:3868, 9 2018.

[59] Yufang Shi, Liming Du, Liangyu Lin, and Ying Wang. Tumour-associated mesenchymal stem/stromal cells: emerging therapeutic targets. *Nature reviews. Drug discovery*, 16:35–52, 1 2017.

[60] Elizabeth S Nakasone, Hanne A Askautrud, Tim Kees, Jae-Hyun Park, Vicki Plaks, Andrew J Ewald, Miriam Fein, Morten G Rasch, Ying-Xim Tan, Jing Qiu, Juwon Park, Pranay Sinha, Mina J Bissell, Eirik Frengen, Zena Werb, and Mikala Egeblad. Imaging tumor-stroma interactions during chemotherapy reveals contributions of the microenvironment to resistance. *Cancer cell*, 21:488–503, 4 2012.

[61] Arturo Álvarez Arenas, Ana Podolski-Renic, Juan Belmonte-Beitia, Milica Pesic, and Gabriel F. Calvo. Interplay of darwinian selection, lamarckian induction and microvesicle transfer on drug resistance in cancer. *Scientific Reports*, 9:9332, 6 2019.

[62] Ravi Salgia and Prakash Kulkarni. The genetic/non-genetic duality of drug 'resistance' in cancer. *Trends in Cancer*, 4:110–118, 2 2018.

[63] Matthew H Bailey, Collin Tokheim, Eduard Porta-Pardo, Sohini Sengupta, Denis Bertrand, Amila Weerasinghe, Antonio Colaprico, Michael C Wendl, Jaegil Kim, Brendan Reardon, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*, 173(2):371–385, 2018.

[64] Zhe Dai, Xu yu Gu, Shou yan Xiang, Dan dan Gong, Chang feng Man, and Yu Fan. Research and application of single-cell sequencing in tumor heterogeneity and drug resistance of circulating tumor cells. *Biomarker Research*, 8:60, 12 2020.

[65] Jurica Levatić, Marina Salvadores, Francisco Fuster-Tormo, and Fran Supek. Mutational signatures are markers of drug sensitivity of cancer cells. *Nature Communications*, 13:2926, 5 2022.

[66] Paul Hasty and Cristina Montagna. Chromosomal rearrangements in cancer. *Molecular Cellular Oncology*, 1:e29904, 1 2014.

[67] Devon A. Lukow and Jason M. Sheltzer. Chromosomal instability and aneuploidy as causes of cancer drug resistance. *Trends in Cancer*, 8:43–53, 1 2022.

[68] Nicholas A. Willis, Emilie Rass, and Ralph Scully. Deciphering the code of the cancer genome: Mechanisms of chromosome rearrangement. *Trends in Cancer*, 1:217–230, 12 2015.

[69] Svetlana N Aleksakhina, Aniruddh Kashyap, and Evgeny N Imyanitov. Mechanisms of acquired tumor drug resistance. *Biochimica et biophysica acta. Reviews on cancer*, 1872:188310, 12 2019.

[70] Zuan-Fu Lim and Patrick C. Ma. Emerging insights of tumor heterogeneity and drug resistance mechanisms in lung cancer targeted therapy. *Journal of Hematology Oncology*, 12:134, 12 2019.

[71] Charles Darwin. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. *London: John Murray (1st ed.)*, 1859.

[72] P C Nowell. The clonal evolution of tumor cell populations. *Science (New York, N.Y.)*, 194:23–8, 10 1976.

[73] LUKÁŠ LACINA, MATÚŠ ČOMA, BARBORA DVOŘÁNKOVÁ, ONDŘEJ KODET, NIKOLA MELEGOVÁ, PETER GÁL, and KAREL SMETANA. Evolution of cancer progression in the context of darwinism. *Anticancer Research*, 39:1–16, 1 2019.

[74] Roberto Vendramin, Kevin Litchfield, and Charles Swanton. Cancer evolution: Darwin and beyond. *The EMBO Journal*, 40, 9 2021.

[75] Zheng Hu, Zan Li, Zhicheng Ma, and Christina Curtis. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nature genetics*, 52:701–708, 7 2020.

[76] Josh N. Vo, Yi-Mi Wu, Jeanmarie Mishler, Sarah Hall, Rahul Mannan, Lisha Wang, Yu Ning, Jin Zhou, Alexander C. Hopkins, James C. Estill, Wallace K. B. Chan, Jennifer Yesil, Xuhong Cao, Arvind Rao, Alexander Tsodikov, Moshe Talpaz, Craig E. Cole, Jing C. Ye, Sikander Ailawadhi, Jesus G. Berdeja, Craig C. Hofmeister, Sundar Jagannath, Andrzej Jakubowiak, Amrita Krishnan, Shaji Kumar, Moshe Yair Levy, Sagar Lonial, Gregory J. Orloff, David Siegel, Suzanne Trudel, Saad Z. Usmani, Ravi Vij, Jeffrey L. Wolf, Jeffrey A. Zonder, P. Leif Bergsagel, Daniel Auclair, Hearn Jay Cho, Dan R. Robinson, and Arul M. Chinnaiyan. The genetic heterogeneity and drug resistance mechanisms of relapsed refractory multiple myeloma. *Nature Communications*, 13:3750, 6 2022.

[77] Yucheng Dong, Qi He, Xinyu Chen, Fan Yang, Li He, and Yongchang Zheng. Extrachromosomal dna (ecdna) in cancer: mechanisms, functions, and clinical implications. *Frontiers in Oncology*, 13, 6 2023.

[78] Lauren T. Pecorino, Roel G.W. Verhaak, Anton Henssen, and Paul S. Mischel. Extrachromosomal dna (ecdna): an origin of tumor heterogeneity, genomic remodeling, and drug resistance. *Biochemical Society Transactions*, 50:1911–1920, 12 2022.

[79] S. Bins, W.S. van de Geer, J. van Riet, N. Steeghs, H.M.W. Verheul, C. van Herpen, P.O. Witteveen, V.C.G. Tjan-Heijnen, M.P. Lolkema, E.E. Voest, S. Sleijfer, E. Cuppen, H.J.G. van de Werken, and R.H.J. Mathijssen. Comprehensive pan-cancer analysis of somatic mutations in drug transporters to reveal acquired and intrinsic drug resistance in 3149 metastatic cancer patients. *Annals of Oncology*, 30:v28–v29, 10 2019.

[80] Xiao Zhu, Shi Li, Bairui Xu, and Hui Luo. Cancer evolution: A means by which tumors evade treatment. *Biomedicine Pharmacotherapy*, 133:111016, 1 2021.

[81] Robert A. Beckman and Lawrence A. Loeb. Rare mutations in cancer drug resistance and implications for therapy. *Clinical Pharmacology Therapeutics*, 108:437–439, 9 2020.

[82] Ran Friedman. Drug resistance in cancer: molecular evolution and compensatory proliferation. *Oncotarget*, 7:11746–11755, 3 2016.

[83] Rebecca A. Burrell and Charles Swanton. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Molecular Oncology*, 8:1095–1111, 9 2014.

[84] Jennifer Tsoi, Lidia Robert, Kim Paraiso, Carlos Galvan, Katherine M. Sheu, Johnson Lay, Deborah J.L. Wong, Mohammad Atefi, Roksana Shirazi, Xiaoyan Wang, Daniel Braas, Catherine S. Grasso, Nicolaos Palaskas,

Antoni Ribas, and Thomas G. Graeber. Multi-stage differentiation defines melanoma subtypes with differential vulnerability to drug-induced iron-dependent oxidative stress. *Cancer Cell*, 33:890–904.e5, 5 2018.

[85] Yapeng Su, Wei Wei, Lidia Robert, Min Xue, Jennifer Tsoi, Angel Garcia-Diaz, Blanca Homet Moreno, Jungwoo Kim, Rachel H Ng, Jihoon W Lee, Richard C Koya, Begonya Comin-Anduix, Thomas G Graeber, Antoni Ribas, and James R Heath. Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 114:13679–13684, 12 2017.

[86] Charles C. Bell and Omer Gilan. Principles and mechanisms of non-genetic resistance in cancer. *British Journal of Cancer*, 122:465–472, 2 2020.

[87] Masashi Mikubo, Yoshiaki Inoue, Geoffrey Liu, and Ming-Sound Tsao. Mechanism of drug tolerant persister cancer cells: The landscape and clinical implication for therapy. *Journal of Thoracic Oncology*, 16:1798–1809, 11 2021.

[88] Michael Ramirez, Satwik Rajaram, Robert J. Steininger, Daria Osipchuk, Maike A. Roth, Leanna S. Morinishi, Louise Evans, Weiyue Ji, Chien-Hsiang Hsu, Kevin Thurley, Shuguang Wei, Anwu Zhou, Prasad R. Koduru, Bruce A. Posner, Lani F. Wu, and Steven J. Altschuler. Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nature Communications*, 7:10690, 2 2016.

[89] Florian Rambow, Aljosja Rogiers, Oskar Marin-Bejar, Sara Aibar, Julia Femel, Michael Dewaele, Panagiotis Karras, Daniel Brown, Young Hwan Chang, Maria Debiec-Rychter, Carmen Adriaens, Enrico Radaelli, Pascal Wolter, Oliver Bechter, Reinhard Dummer, Mitchell Levesque, Adriano Piris, Dennie T Frederick, Genevieve Boland, Keith T Flaherty, Joost van den Oord, Thierry Voet, Stein Aerts, Amanda W Lund, and Jean-Christophe Marine. Toward minimal residual disease-directed therapy in melanoma. *Cell*, 174:843–855.e19, 8 2018.

[90] Paola Falletta, Luis Sanchez-Del-Campo, Jagat Chauhan, Maike Effern, Amy Kenyon, Christopher J Kershaw, Robert Siddaway, Richard Lisle, Rasmus Freter, Matthew J Daniels, Xin Lu, Thomas Tüting, Mark Middleton, Francesca M Buffa, Anne E Willis, Graham Pavitt, Ze'ev A Ronai, Tatjana Sauka-Spengler, Michael Hölzel, and Colin R Goding. Translation reprogramming is an evolutionarily conserved driver of phenotypic plasticity and therapeutic resistance in melanoma. *Genes development*, 31:18–33, 1 2017.

[91] Nikolaos A Dallas, Ling Xia, Fan Fan, Michael J Gray, Puja Gaur, George van Buren, Shaija Samuel, Michael P Kim, Sherry J Lim, and Lee M Ellis. Chemoresistant colorectal cancer cells, the cancer stem cell phenotype, and increased sensitivity to insulin-like growth factor-i receptor inhibition. *Cancer research*, 69:1951–7, 3 2009.

[92] Eugene Tulchinsky, Oleg Demidov, Marina Kriajevska, Nickolai A Barlev, and Evgeny Imyanitov. Emt: A mechanism for escape from egfr-targeted therapy in lung cancer. *Biochimica et biophysica acta. Reviews on cancer*, 1871:29–39, 1 2019.

[93] Rebecca H Chisholm, Tommaso Lorenzi, Alexander Lorz, Annette K Larsen, Luís Neves de Almeida, Alexandre Escargueil, and Jean Clairambault. Emergence of drug tolerance in cancer cell populations: an evolutionary outcome of selection, nongenetic instability, and stress-induced adaptation. *Cancer research*, 75:930–9, 3 2015.

[94] Jennifer Landsberg, Judith Kohlmeyer, Marcel Renn, Tobias Bald, Meri Rogava, Mira Cron, Martina Fatho, Volker Lennerz, Thomas Wölfel, Michael Hölzel, and Thomas Tüting. Melanomas resist t-cell therapy through inflammation-induced reversible dedifferentiation. *Nature*, 490:412–6, 10 2012.

[95] Arnav Mehta, Yeon Joo Kim, Lidia Robert, Jennifer Tsoi, Begoña Comin-Anduix, Beata Berent-Maoz, Alistair J Cochran, James S Economou, Paul C Tumeh, Cristina Puig-Saus, and Antoni Ribas. Immunotherapy resistance by inflammation-induced dedifferentiation. *Cancer discovery*, 8:935–943, 8 2018.

[96] Richard W Burkhardt. Lamarck, evolution, and the inheritance of acquired characters. *Genetics*, 194:793–805, 8 2013.

[97] Swagata Adhikari, Apoorva Bhattacharya, Santanu Adhikary, Vipin Singh, Shrikanth S. Gadad, Siddhartha Roy, and Chandrima Das. The paradigm of drug resistance in cancer: an epigenetic perspective. *Bioscience Reports*, 42, 4 2022.

[98] Takuma Hayashi and Ikuo Konishi. Correlation of anti-tumour drug resistance with epigenetic regulation. *British Journal of Cancer*, 124:681–682, 2 2021.

[99] Nan Wang, Ting Ma, and Bin Yu. Targeting epigenetic regulators to overcome drug resistance in cancers. *Signal Transduction and Targeted Therapy*, 8:69, 2 2023.

[100] https://www.who.int.

[101] Freddie Bray and Isabelle Soerjomataram. *The Changing Global Burden of Cancer: Transitions in Human Development and Implications for Cancer Prevention and Control*. 2015.

[102] Shiquan Sun, Jiaqiang Zhu, Ying Ma, and Xiang Zhou. Accuracy, robustness and scalability of dimensionality reduction methods for single-cell rna-seq analysis. *Genome Biology*, 20:269, 12 2019.

[103] Lawrence N. Shulman, Walter Willett, Amy Sievers, and Felicia M. Knaul. Breast cancer in developing countries: Opportunities for improved survival. *Journal of Oncology*, 2010:1–6, 2010.

[104] Rebecca L Siegel, Kimberly D Miller, Hannah E Fuchs, and Ahmedin Jemal. Cancer statistics, 2022. *CA: a cancer journal for clinicians*, 72:7–33, 1 2022.

[105] Walburga Yvonne Joko-Fru, Elima Jedy-Agba, Anne Korir, Olufemi Ogunbiyi, Charles P Dzamalala, Eric Chokunonga, Henry Wabinga, Shyam Manraj, Anne Finesse, Nontuthuzelo Somdyala, Biying Liu, Paul McGale, Ahmedin Jemal, Freddie Bray, and Donald Maxwell Parkin. The evolving epidemic of breast cancer in sub-saharan africa: Results from the african cancer registry network. *International journal of cancer*, 147:2131–2141, 10 2020.

[106] Yeonju Kim, Keun-Young Yoo, and Marc T Goodman. Differences in incidence, mortality and survival of breast cancer by regions and countries in asia and contributing factors. *Asian Pacific journal of cancer prevention : APJCP*, 16:2857–70, 2015.

[107] Nuket Eliyatkin, Evrim Yalcin, Baha Zengel, Safiye Aktaş, and Enver Vardar. Molecular classification of breast carcinoma: From traditional, old-fashioned way to a new age, and a new way. *Journal of Breast Health*, 11:59–66, 4 2015.

[108] Seho Park, Ja Seung Koo, Min Suk Kim, Hyung Seok Park, Jun Sang Lee, Jong Seok Lee, Seung Il Kim, and Byeong-Woo Park. Characteristics and outcomes according to molecular subtypes of breast cancer as classified by a panel of four biomarkers using immunohistochemistry. *Breast (Edinburgh, Scotland)*, 21:50–7, 2 2012.

[109] Fatma Khinaifis Al-thoubaity. Molecular classification of breast cancer: A retrospective cohort study. *Annals of Medicine and Surgery*, 49:44–48, 1 2020.

[110] Karen S Johnson, Emily F Conant, and Mary Scott Soo. Molecular subtypes of breast cancer: A review for breast radiologists. *Journal of Breast Imaging*, 3:12–24, 1 2021.

[111] Shani Paluch-Shimon, Nathan I Cherny, Elisabeth G E de Vries, Urania Dafni, Martine J Piccart, Nicola Jane Latino, and Fatima Cardoso. Application of the esmo-magnitude of clinical benefit scale (v.1.1) to the field of early breast cancer therapies. *ESMO open*, 5:e000743, 9 2020.

[112] Guan-Qun Zhou, Jia-Wei Lv, Ling long Tang, Yan-Ping Mao, Rui Guo, Jun Ma, and Ying Sun. Evaluation of the national comprehensive cancer network and european society for medical oncology nasopharyngeal carcinoma surveillance guidelines. *Frontiers in Oncology*, 10, 2 2020.

[113] Michaela J Higgins and Vered Stearns. Understanding resistance to tamoxifen in hormone receptor-positive breast cancer. *Clinical chemistry*, 55:1453–5, 8 2009.

[114] Oğuz Lafcı, Pınar Celepli, Pelin Seher Öztekin, and Pınar Nercis Koşar. Dce-mri radiomics analysis in differentiating luminal a and luminal b breast cancer molecular subtypes. *Academic radiology*, 30:22–29, 1 2023.

[115] Jennifer M. Matro, Tianyu Li, Massimo Cristofanilli, Melissa E. Hughes, Rebecca A. Ottesen, Jane C. Weeks, and Yu-Ning Wong. Inflammatory breast cancer management in the national comprehensive cancer network: The disease, recurrence pattern, and outcome. *Clinical Breast Cancer*, 15:1–7, 2 2015.

[116] Zorka Inic, Milan Zegarac, Momcilo Inic, Ivan Markovic, Zoran Kozomara, Igor Djurisic, Ivana Inic, Gordana Pupic, and Snezana Jancic. Difference between luminal a and luminal b subtypes according to ki-67, tumor size, and progesterone receptor negativity providing prognostic information. *Clinical Medicine Insights: Oncology*, 8:CMO.S18006, 1 2014.

[117] Ayad Ahmad Mohammed. The clinical behavior of different molecular subtypes of breast cancer. *Cancer Treatment and Research Communications*, 29:100469, 2021.

[118] Giovanni Ciriello, Rileen Sinha, Katherine A. Hoadley, Anders S. Jacobsen, Boris Reva, Charles M. Perou, Chris Sander, and Nikolaus Schultz. The molecular diversity of luminal a breast tumors. *Breast Cancer Research and Treatment*, 141:409–420, 10 2013.

[119] Stéphanie Cornen, Arnaud Guille, José Adélaïde, Lynda Addou-Klouche, Pascal Finetti, Marie-Rose Saade, Marwa Manai, Nadine Carbuccia, Ismahane Bekhouche, Anne Letessier, Stéphane Raynaud, Emmanuelle Charafe-Jauffret, Jocelyne Jacquemier, Salvatore Spicuglia, Hugues de The, Patrice Viens, François Bertucci, Daniel Birnbaum, and Max Chaffanet. Candidate luminal b breast cancer genes identified by genome, gene expression and dna methylation profiling. *PLoS ONE*, 9:e81843, 1 2014.

[120] Uma Krishnamurti and Jan F Silverman. Her2 in breast cancer: a review and update. *Advances in anatomic pathology*, 21:100–7, 3 2014.

[121] Maria Cristina Figueroa-Magalhães, Danijela Jelovac, Roisin Connolly, and Antonio C Wolff. Treatment of her2-positive breast cancer. *Breast (Edinburgh, Scotland)*, 23:128–136, 4 2014.

[122] Nermine H. Zakaria, Doaa Hashad, Marwa H. Saied, Neamat Hegazy, Alyaa Elkayal, and Eman Tayae. Genetic mutations in her2-positive breast cancer: possible association with response to trastuzumab therapy. *Human Genomics*, 17:43, 5 2023.

[123] Libo Yang, Feng Ye, Longlong Bao, Xiaoyan Zhou, Zhe Wang, Peizhen Hu, Nengtai Ouyang, Xiaojuan Li, Yi Shi, Gang Chen, Peiyi Xia, Meiying Chui, Wencai Li, Ying Jia, Yueping Liu, Junjun Liu, Junyi Ye, Zhe Zhang, and Hong Bu. Somatic alterations of tp53, erbb2, pik3ca and ccnd1 are associated with chemosensitivity for breast cancers. *Cancer science*, 110:1389–1400, 4 2019.

[124] Eric Hahnen, Jan Hauke, Christoph Engel, Guido Neidhardt, Kerstin Rhiem, and Rita K. Schmutzler. Germline mutations in triple-negative breast cancer. *Breast Care*, 12:15–19, 2017.

[125] Suvimol Niyomnaitham, Napa Parinyanitikul, Ekkapong Roothumnong, Worapoj Jinda, Norasate Samarnthai, Taywin Atikankul, Bhoom Suktitipat, Wanna Thongnoppakhun, Chanin Limwongse, and Manop Pithukpakorn. Tumor mutational profile of triple negative breast cancer patients in thailand revealed distinctive genetic alteration in chromatin remodeling gene. *PeerJ*, 7:e6501, 2 2019.

[126] Sandra Diaz Casas, Eder Lancheros García, Andrés Sanchéz Campo, Ricardo Sanchez Pedraza, Vivian Roman Vasquez, Sara D Mendoza, Javier Angel Aristizabal, Carlos Lehmann Mosquera, Carlos Duarte Torres, and Juan C Vergel. Clinical behavior of triple negative breast cancer in a cohort of latin american women. *Cureus*, 6 2019.

[127] Paola Zagami and Lisa Anne Carey. Triple negative breast cancer: Pitfalls and progress. *npj Breast Cancer*, 8:95, 8 2022.

[128] Katrina R Bauer, Monica Brown, Rosemary D Cress, Carol A Parise, and Vincent Caggiano. Descriptive analysis of estrogen receptor (er)-negative, progesterone receptor (pr)-negative, and her2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the california cancer registry. *Cancer*, 109:1721–8, 5 2007.

[129] Lisa A. Carey, Charles M. Perou, Chad A. Livasy, Lynn G. Dressler, David Cowan, Kathleen Conway, Gamze Karaca, Melissa A. Troester, Chiu Kit Tse, Sharon Edmiston, Sandra L. Deming, Joseph Geradts, Maggie C. U. Cheang, Torsten O. Nielsen, Patricia G. Moorman, H. Shelton Earp, and Robert C. Millikan. Race, breast cancer subtypes, and survival in the carolina breast cancer study. *JAMA*, 295:2492, 6 2006.

[130] Brian D Lehmann, Joshua A Bauer, Xi Chen, Melinda E Sanders, A Bapsi Chakravarthy, Yu Shyr, and Jennifer A Pietenpol. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of clinical investigation*, 121:2750–67, 7 2011.

[131] Antonio Marra, Dario Trapani, Giulia Viale, Carmen Criscitiello, and Giuseppe Curigliano. Practical classification of triple-negative breast cancer: intratumoral heterogeneity, mechanisms of drug resistance, and novel therapies. *npj Breast Cancer*, 6:54, 10 2020.

[132] Liang Li, Shuchen Liu, Liyuan Liu, Zhongbing Ma, Man Feng, Chunmiao Ye, Wenzhong Zhou, Yongjiu Wang, Lu Liu, Fei Wang, Lixiang Yu, Fei Zhou, Yujuan Xiang, Shuya Huang, Qinye Fu, Qiang Zhang, Dezong Gao, and Zhigang Yu. Impact of phosphorylated insulin-like growth factor-1 receptor on the outcome of breast cancer patients and the prognostic value of its alteration during neoadjuvant chemotherapy. *Experimental and Therapeutic Medicine*, 8 2018.

[133] Elizabeth Sakach, Ruth O'Regan, Jane Meisel, and Xiaoxian Li. Molecular classification of triple negative breast cancer and the emergence of targeted therapies. *Clinical Breast Cancer*, 21:509–520, 12 2021.

[134] Francesc Font-Clos, Stefano Zapperi, and Caterina A. M. La Porta. Classification of triple negative breast cancer by epithelial mesenchymal transition and the tumor immune microenvironment. *Scientific Reports*, 12:9651, 6 2022.

[135] Clinton Yam, Sendurai A. Mani, and Stacy L. Moulder. Targeting the molecular subtypes of triple negative breast cancer: Understanding the diversity to progress the field. *The Oncologist*, 22:1086–1093, 9 2017.

[136] G. Gambardella, G. Viscido, B. Tumaini, A. Isacchi, R. Bosotti, and D. di Bernardo. A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. *Nature Communications*, 13:1714, 3 2022.

[137] Shaked Slovin, Annamaria Carissimo, Francesco Panariello, Antonio Grimaldi, Valentina Bouché, Gennaro Gambardella, and Davide Cacchiarelli. Single-cell rna sequencing analysis: A step-by-step overview. *Methods in molecular biology (Clifton, N.J.)*, 2284:343–365, 2021.

[138] Julio Aguila Benitez, Shangli Cheng, and Qiaolin Deng. Revealing allele-specific gene expression by single-cell transcriptomics. *The International Journal of Biochemistry Cell Biology*, 90:155–160, 9 2017.

[139] Xiliang Wang, Yao He, Qiming Zhang, Xianwen Ren, and Zemin Zhang. Direct comparative analyses of 10x genomics chromium and smart-seq2. *Genomics, Proteomics Bioinformatics*, 19:253–266, 4 2021.

[140] Ashraful Haque, Jessica Engel, Sarah A. Teichmann, and Tapio Lönnberg. A practical guide to single-cell rna-sequencing for biomedical research and clinical applications. *Genome Medicine*, 9:75, 12 2017.

[141] Xiaoqing Yu, Farnoosh Abbas-Aghababazadeh, Y. Ann Chen, and Brooke L. Fridley. *Statistical and Bioinformatics Analysis of Data from Bulk and Single-Cell RNA Sequencing Experiments*, pages 143–175. 2021.

[142] Simone Picelli, Omid R Faridani, Åsa K Björklund, Gösta Winberg, Sven Sagasser, and Rickard Sandberg. Full-length rna-seq from single cells using smart-seq2. *Nature Protocols*, 9:171–181, 1 2014.

[143] Daniel M. DeLaughter. The use of the fluidigm c1 for rna expression analyses of single cells. *Current Protocols in Molecular Biology*, 122, 4 2018.

[144] https://www.10xgenomics.com.

[145] Allon M Klein, Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A Weitz, and Marc W Kirschner. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161:1187–1201, 5 2015.

[146] https://www.illumina.com.

[147] Sergio Mora-Castilla, Cuong To, Soheila Vaezeslami, Robert Morey, Srimeenakshi Srinivasan, Jennifer N. Chousal, Heidi Cook-Andersen, Joby Jenkins, and Louise C. Laurent. Miniaturization technologies for efficient single-cell library preparation for next-generation sequencing. *SLAS Technology*, 21:557–567, 8 2016.

[148] Simone Picelli. Single-cell rna-sequencing: The future of genome biology is now. *RNA Biology*, 14:637–650, 5 2017.

[149] Barton E. Slatko, Andrew F. Gardner, and Frederick M. Ausubel. Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology*, 122, 4 2018.

[150] Alexander Dobin, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gingeras. Star: ultrafast universal rna-seq aligner. *Bioinformatics (Oxford, England)*, 29:15–21, 1 2013.

[151] Matthew D Young and Sam Behjati. Soupx removes ambient rna contamination from droplet-based single-cell rna sequencing data. *GigaScience*, 9, 12 2020.

[152] Christopher S. McGinnis, Lyndsay M. Murrow, and Zev J. Gartner. Doubletfinder: Doublet detection in single-cell rna sequencing data using artificial nearest neighbors. *Cell Systems*, 8:329–337.e4, 4 2019.

[153] Greg Finak, Andrew McDavid, Masanao Yajima, Jingyuan Deng, Vivian Gersuk, Alex K. Shalek, Chloe K. Slichter, Hannah W. Miller, M. Juliana McElrath, Martin Prlic, Peter S. Linsley, and Raphael Gottardo. Mast: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell rna sequencing data. *Genome Biology*, 16:278, 12 2015.

[154] Aaron T.L. Lun, Davis J. McCarthy, and John C. Marioni. A step-by-step workflow for low-level analysis of single-cell rna-seq data with bioconductor. *F1000Research*, 5:2122, 10 2016.

[155] Laleh Haghverdi, Aaron T L Lun, Michael D Morgan, and John C Marioni. Batch effects in single-cell rna-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, 36:421–427, 5 2018.

[156] Maximilian Sprang, Miguel A. Andrade-Navarro, and Jean-Fred Fontaine. Batch effect detection and correction in rna-seq data using machine-learning-based automated assessment of quality. *BMC Bioinformatics*, 23:279, 7 2022.

[157] Yuqing Zhang, Giovanni Parmigiani, and W Evan Johnson. Combat-seq: batch effect adjustment for rna-seq count data. *NAR Genomics and Bioinformatics*, 2, 9 2020.

[158] Ilya Korsunsky, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po ru Loh, and Soumya Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with harmony. *Nature Methods*, 16:1289–1296, 12 2019.

[159] Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck, Shiwei Zheng, Andrew Butler, Maddie J. Lee, Aaron J. Wilk, Charlotte Darby, Michael Zager, Paul Hoffman, Marlon Stoeckius, Efthymia Papalexi, Eleni P. Mimitou, Jaison Jain, Avi Srivastava, Tim Stuart, Lamar M. Fleming, Bertrand Yeung, Angela J. Rogers, Juliana M. McElrath, Catherine A. Blish, Raphael Gottardo, Peter Smibert, and Rahul Satija. Integrated analysis of multimodal single-cell data. *Cell*, 184:3573–3587.e29, 6 2021.

[160] Joshua D. Welch, Velina Kozareva, Ashley Ferreira, Charles Vanderburg, Carly Martin, and Evan Z. Macosko. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell*, 177:1873–1887.e17, 6 2019.

[161] Mohammad Lotfollahi, F. Alexander Wolf, and Fabian J. Theis. scgen predicts single-cell perturbation responses. *Nature Methods*, 16:715–721, 8 2019.

[162] Malte D. Luecken, M. Büttner, K. Chaichoompu, A. Danese, M. Interlandi, M. F. Mueller, D. C. Strobl, L. Zappia, M. Dugas, M. Colomé-Tatché, and Fabian J. Theis. Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods*, 19:41–50, 1 2022.

[163] Hoa Thi Nhu Tran, Kok Siong Ang, Marion Chevrier, Xiaomeng Zhang, Nicole Yee Shin Lee, Michelle Goh, and Jinmiao Chen. A benchmark of batch-effect correction methods for single-cell rna sequencing data. *Genome Biology*, 21:12, 12 2020.

[164] Andrew Butler, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, 36:411–420, 5 2018.

[165] Rahul Satija, Jeffrey A Farrell, David Gennert, Alexander F Schier, and Aviv Regev. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, 33:495–502, 5 2015.

[166] Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *Cell*, 177:1888–1902.e21, 6 2019.

[167] Saket Choudhary and Rahul Satija. Comparison and evaluation of statistical error models for scrna-seq. *Genome Biology*, 23:27, 1 2022.

[168] Christoph Hafemeister and Rahul Satija. Normalization and variance stabilization of single-cell rna-seq data using regularized negative binomial regression. *Genome Biology*, 20:296, 12 2019.

[169] Lukas Heumos, Anna C. Schaar, Christopher Lance, Anastasia Litinetskaya, Felix Drost, Luke Zappia, Malte D. Lücken, Daniel C. Strobl, Juan Henao, Fabiola Curion, Hananeh Aliee, Meshal Ansari, Pau Badia i Mompel, Maren Büttner, Emma Dann, Daniel Dimitrov, Leander Dony, Amit Frishberg, Dongze He, Soroor Hediyeh-zadeh, Leon Hetzel, Ignacio L. Ibarra, Matthew G. Jones, Mohammad Lotfollahi, Laura D. Martens, Christian L. Müller, Mor Nitzan, Johannes Ostner, Giovanni Palla, Rob Patro, Zoe Piran, Ciro Ramírez-Suástegui, Julio Saez-Rodriguez, Hirak Sarkar, Benjamin Schubert, Lisa Sikkema, Avi Srivastava, Jovan Tanevski, Isaac Virshup, Philipp Weiler, Herbert B. Schiller, and Fabian J. Theis. Best practices for single-cell analysis across modalities. *Nature Reviews Genetics*, 24:550–572, 8 2023.

[170] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 9 1933.

[171] Malte D Luecken and Fabian J Theis. Current best practices in single-cell rna-seq analysis: a tutorial. *Molecular Systems Biology*, 15, 6 2019.

[172] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2:559–572, 11 1901.

[173] D D Lee and H S Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–91, 10 1999.

[174] G Hinton L Van der Maaten. Visualizing data using t-sne. *Journal of machine learning research*, 2008.

[175] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. 2 2018.

[176] Emma Pierson and Christopher Yau. Zifa: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome Biology*, 16:241, 12 2015.

[177] Ruizhi Xiang, Wencan Wang, Lei Yang, Shiyuan Wang, Chaohan Xu, and Xiaowen Chen. A comparison for dimensionality reduction methods of single-cell rna-seq data. *Frontiers in Genetics*, 12, 3 2021.

[178] Wei Dong, Charikar Moses, and Kai Li. Efficient k-nearest neighbor graph construction for generic similarity measures. pages 577–586. ACM, 3 2011.

[179] Daniel N. Baker, Nathan Dyjack, Vladimir Braverman, Stephanie C. Hicks, and Ben Langmead. Fast and memory-efficient scrna-seq k -means clustering with various distances. pages 1–8. ACM, 8 2021.

[180] Soumita Seth, Saurav Mallik, Tapas Bhadra, and Zhongming Zhao. Dimensionality reduction and louvain agglomerative hierarchical clustering for cluster-specified frequent biomarker discovery in single-cell sequencing data. *Frontiers in Genetics*, 13, 2 2022.

[181] Lijia Yu, Yue Cao, Jean Y. H. Yang, and Pengyi Yang. Benchmarking clustering algorithms on estimating the number of cell types from single-cell rna-sequencing data. *Genome Biology*, 23:49, 2 2022.

[182] Zhijin Wu and Hao Wu. Accounting for cell type hierarchy in evaluating single cell rna-seq clustering. *Genome Biology*, 21:123, 12 2020.

[183] Shixiong Zhang, Xiangtao Li, Jiecong Lin, Qiuzhen Lin, and Ka-Chun Wong. Review of single-cell rna-seq data clustering for cell-type identification and characterization. *RNA*, 29:517–530, 5 2023.

[184] Vladimir Yu Kiselev, Kristina Kirschner, Michael T Schaub, Tallulah Andrews, Andrew Yiu, Tamir Chandra, Kedar N Natarajan, Wolf Reik, Mauricio Barahona, Anthony R Green, and Martin Hemberg. Sc3: consensus clustering of single-cell rna-seq data. *Nature Methods*, 14:483–486, 5 2017.

[185] Bo Wang, Daniele Ramazzotti, Luca De Sano, Junjie Zhu, Emma Pierson, and Serafim Batzoglou. Simlr: A tool for large-scale genomic analyses by multi-kernel learning. *Proteomics*, 18, 1 2018.

[186] Dominic Grün, Anna Lyubimova, Lennart Kester, Kay Wiebrands, Onur Basak, Nobuo Sasaki, Hans Clevers, and Alexander van Oudenaarden. Single-cell messenger rna sequencing reveals rare intestinal cell types. *Nature*, 525:251–5, 9 2015.

[187] Peijie Lin, Michael Troup, and Joshua W. K. Ho. Cidr: Ultrafast and accurate clustering through imputation for single-cell rna-seq data. *Genome Biology*, 18:59, 12 2017.

[188] Amit Zeisel, Ana B Muñoz-Manchado, Simone Codeluppi, Peter Lönnerberg, Gioele La Manno, Anna Juréus, Sueli Marques, Hermany Munguba, Liqun He, Christer Betsholtz, Charlotte Rolny, Gonçalo Castelo-Branco, Jens Hjerling-Leffler, and Sten Linnarsson. Brain structure. cell types in the mouse cortex and hippocampus revealed by single-cell rna-seq. *Science (New York, N.Y.)*, 347:1138–42, 3 2015.

[189] Justina žurauskienė and Christopher Yau. pcareduce: hierarchical clustering of single cell transcriptional profiles. *BMC Bioinformatics*, 17:140, 12 2016.

[190] Lihong Peng, Xiongfei Tian, Geng Tian, Junlin Xu, Xin Huang, Yanbin Weng, Jialiang Yang, and Liqian Zhou. Single-cell rna-seq clustering: datasets, models, and algorithms. *RNA Biology*, 17:765–783, 6 2020.

[191] Chunxiang Wang, Xin Gao, and Juntao Liu. Impact of data preprocessing on cell-type clustering based on single-cell rna-seq data. *BMC Bioinformatics*, 21:440, 12 2020.

[192] Ren Qi, Jin Wu, Fei Guo, Lei Xu, and Quan Zou. A spectral clustering with self-weighted multiple kernel learning method for single-cell rna-seq data. *Briefings in bioinformatics*, 22, 7 2021.

[193] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:P10008, 10 2008.

[194] Nicolas Dugué and Anthony Perez. Direction matters in complex networks: A theoretical and applied study for greedy modularity optimization. *Physica A: Statistical Mechanics and its Applications*, 603:127798, 10 2022.

[195] Ruiqing Zheng, Min Li, Zhenlan Liang, Fang-Xiang Wu, Yi Pan, and Jianxin Wang. Sinnlrr: a robust subspace clustering method for cell type detection by non-negative and low-rank representation. *Bioinformatics*, 35:3642–3650, 10 2019.

[196] F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome Biology*, 19:15, 12 2018.

[197] V. A. Traag, L. Waltman, and N. J. van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific Reports*, 9:5233, 3 2019.

[198] Matthew Chung, Vincent M. Bruno, David A. Rasko, Christina A. Cuomo, José F. Muñoz, Jonathan Livny, Amol C. Shetty, Anup Mahurkar, and Julie C. Dunning Hotopp. Best practices on the differential expression analysis of multi-species rna-seq. *Genome Biology*, 22:121, 12 2021.

[199] Charlotte Soneson and Mauro Delorenzi. A comparison of methods for differential expression analysis of rna-seq data. *BMC Bioinformatics*, 14:91, 12 2013.

[200] Luca Alessandrì, Maddalena Arigoni, and Raffaele Calogero. Differential expression analysis in single-cell transcriptomics. *Methods in molecular biology (Clifton, N.J.)*, 1979:425–432, 2019.

[201] Hai C. T. Nguyen, Bukyung Baik, Sora Yoon, Taesung Park, and Dougu Nam. Benchmarking integration of single-cell differential expression. *Nature Communications*, 14:1570, 3 2023.

[202] Jordan W. Squair, Matthieu Gautier, Claudia Kathe, Mark A. Anderson, Nicholas D. James, Thomas H. Hutson, Rémi Hudelle, Taha Qaiser, Kaya J. E. Matson, Quentin Barraud, Ariel J. Levine, Gioele La Manno, Michael A. Skinnider, and Grégoire Courtine. Confronting false discoveries in single-cell differential expression. *Nature Communications*, 12:5692, 9 2021.

[203] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15:550, 12 2014.

[204] Yunshun Chen, Aaron T L Lun, and Gordon K Smyth. From reads to genes to pathways: differential expression analysis of rna-seq experiments using rsubread and the edger quasi-likelihood pipeline. *F1000Research*, 5:1438, 2016.

[205] Matthew E Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43:e47, 4 2015.

[206] Vamsi K Mootha, Cecilia M Lindgren, Karl-Fredrik Eriksson, Aravind Subramanian, Smita Sihag, Joseph Lehar, Pere Puigserver, Emma Carlsson, Martin Ridderstråle, Esa Laurila, Nicholas Houstis, Mark J Daly, Nick Patterson, Jill P Mesirov, Todd R Golub, Pablo Tamayo, Bruce Spiegelman, Eric S Lander, Joel N Hirschhorn, David Altshuler, and Leif C Groop. Pgc-1-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, 34:267–273, 7 2003.

[207] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102:15545–15550, 10 2005.

[208] Sonja Hänzelmann, Robert Castelo, and Justin Guinney. Gsva: gene set variation analysis for microarray and rna-seq data. *BMC Bioinformatics*, 14:7, 12 2013.

[209] Karsten Krug, Philipp Mertins, Bin Zhang, Peter Hornbeck, Rajesh Raju, Rushdy Ahmad, Matthew Szucs, Filip Mundt, Dominique Forestier, Judit Jane-Valbuena, Hasmik Keshishian, Michael A. Gillette, Pablo Tamayo, Jill P. Mesirov, Jacob D. Jaffe, StevenA. Carr, and D.R. Mani. A curated resource for phosphosite-specific signature analysis. *Molecular Cellular Proteomics*, 18:576–593, 3 2019.

[210] Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts, Joost van den Oord, Zeynep Kalender Atak, Jasper Wouters, and Stein Aerts. Scenic: single-cell regulatory network inference and clustering. *Nature Methods*, 14:1083–1086, 11 2017.

[211] Melania Franchini, Simona Pellecchia, Gaetano Viscido, and Gennaro Gambardella. Single-cell gene set enrichment analysis and transfer learning for functional annotation of scrna-seq data. *NAR Genomics and Bioinformatics*, 5, 1 2023.

[212] Gennaro Gambardella and Diego Di Bernardo. A tool for visualization and analysis of single-cell rna-seq data based on text mining. *Frontiers in genetics*, 10:734, 2019.

[213] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. *Mining of Massive Datasets*. Cambridge University Press, 11 2014.

[214] S. E. Robertson and K. Sparck Jones. Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27:129–146, 5 1976.

[215] Wouter Saelens, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. A comparison of single-cell trajectory inference methods. *Nature biotechnology*, 37:547–554, 5 2019.

[216] Wendy Xueyi Wang and Julie L. Lefebvre. Morphological pseudotime ordering and fate mapping reveal diversification of cerebellar inhibitory interneurons. *Nature Communications*, 13:3433, 6 2022.

[217] Chun Zeng, Francesca Mulas, Yinghui Sui, Tiffany Guan, Nathanael Miller, Yuliang Tan, Fenfen Liu, Wen Jin, Andrea C. Carrano, Mark O. Huising, Orian S. Shirihai, Gene W. Yeo, and Maike Sander. Pseudotemporal ordering of single cells reveals metabolic control of postnatal cell proliferation. *Cell Metabolism*, 25:1160–1175.e11, 5 2017.

[218] Sean C. Bendall, Kara L. Davis, El ad David Amir, Michelle D. Tadmor, Erin F. Simonds, Tiffany J. Chen, Daniel K. Shenfeld, Garry P. Nolan, and Dana Pe'er. Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. *Cell*, 157:714–725, 4 2014.

[219] Luca Albergante, Evgeny M. Mirkes, Huidong Chen, Alexis Martin, Louis Faure, Emmanuel Barillot, Luca Pinello, Alexander N. Gorban, and Andrei Zinovyev. Robust and scalable learning of complex dataset topologies via elpigraph. 4 2018.

[220] Yusen Ye, Lin Gao, and Shihua Zhang. Circular trajectory reconstruction uncovers cell-cycle progression and regulatory dynamics from single-cell hi-c maps. *Advanced Science*, 6, 12 2019.

[221] Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang, Stefan Mundlos, Lena Christiansen, Frank J. Steemers, Cole Trapnell, and Jay Shendure. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566:496–502, 2 2019.

[222] Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J Lennon, Kenneth J Livak, Tarjei S Mikkelsen, and John L Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, 32:381–386, 4 2014.

[223] Zhicheng Ji and Hongkai Ji. Tscan: Pseudo-time reconstruction and evaluation in single-cell rna-seq analysis. *Nucleic Acids Research*, 44:e117–e117, 7 2016.

[224] Kelly Street, Davide Risso, Russell B. Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19:477, 12 2018.

[225] F. Alexander Wolf, Fiona K. Hamey, Mireya Plass, Jordi Solana, Joakim S. Dahlin, Berthold Göttgens, Nikolaus Rajewsky, Lukas Simon, and Fabian J. Theis. Paga: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, 20:59, 12 2019.

[226] Louise Deconinck, Robrecht Cannoodt, Wouter Saelens, Bart Deplancke, and Yvan Saeys. Recent advances in trajectory inference from single-cell omics data. *Current Opinion in Systems Biology*, 27:100344, 9 2021.

[227] Kai Kretzschmar and Fiona M. Watt. Lineage tracing. *Cell*, 148:33–45, 1 2012.

[228] Szu-Hsien Sam Wu, Ji-Hyun Lee, and Bon-Kyoung Koo. Lineage tracing: Computational reconstruction goes beyond the limit of imaging. *Molecules and cells*, 42:104–112, 2 2019.

[229] G S Stent. Developmental cell lineage. *The International journal of developmental biology*, 42:237–41, 1998.

[230] C D Stern and S E Fraser. Tracing the lineage of tracing cell lineages. *Nature cell biology*, 3:E216–8, 9 2001.

[231] Edwin G. Conklin. Mosaic development in ascidian eggs. *Journal of Experimental Zoology*, 2:145–223, 5 1905.

[232] CHARLES OTIS WHITMAN. Memoirs: The embryology of clepsine. *Journal of Cell Science*, s2-18:215–315, 7 1878.

[233] J E Sulston, E Schierenberg, J G White, and J N Thomson. The embryonic cell lineage of the nematode caenorhabditis elegans. *Developmental biology*, 100:64–119, 11 1983.

[234] Peter A Lawrence and Michael Levine. Mosaic and regulative development: two faces of one coin. *Current biology : CB*, 16:R236–9, 4 2006.

[235] C B Kimmel, R M Warga, and T F Schilling. Origin and organization of the zebrafish fate map. *Development (Cambridge, England)*, 108:581–94, 4 1990.

[236] Sarah-Ann Aelvoet, Jesus Pascual-Brazo, Sarah Libbrecht, Veerle Reumers, Rik Gijsbers, Chris Van den Haute, and Veerle Baekelandt. Long-term fate mapping using conditional lentiviral vectors reveals a continuous contribution of radial glia-like cells to adult hippocampal neurogenesis in mice. *PLOS ONE*, 10:e0143772, 11 2015.

[237] Liyun Liu, Hongrui Zhu, Yahui Liao, Wei Wu, Lei Liu, Li Liu, Ying Wu, Fan Sun, and Hou-Wen Lin. Inhibition of wnt/-catenin pathway reverses multi-drug resistance and emt in oct4+/nanog+ nsclc cells. *Biomedicine pharmacotherapy = Biomedecine pharmacotherapie*, 127:110225, 7 2020.

[238] D L Turner and C L Cepko. A common progenitor for neurons and glia persists in rat retina late in development. *Nature*, 328:131–6.

[239] Nick Barker, Johan H van Es, Jeroen Kuipers, Pekka Kujala, Maaike van den Born, Miranda Cozijnsen, Andrea Haegebarth, Jeroen Korving, Harry Begthel, Peter J Peters, and Hans Clevers. Identification of stem cells in small intestine and colon by marker gene lgr5. *Nature*, 449:1003–7, 10 2007.

[240] M. P. Alcolea and P. H. Jones. Lineage analysis of epidermal stem cells. *Cold Spring Harbor Perspectives in Medicine*, 4:a015206–a015206, 1 2014.

[241] Hui Zong, J. Sebastian Espinosa, Helen Hong Su, Mandar D. Muzumdar, and Liqun Luo. Mosaic analysis with double markers in mice. *Cell*, 121:479–492, 5 2005.

[242] Tamily A Weissman and Y Albert Pan. Brainbow: New resources and emerging biological applications for multicolor genetic labeling and analysis. *Genetics*, 199:293–306, 2 2015.

[243] Johannes Hirrlinger, Anja Scheller, Petra G Hirrlinger, Beate Kellert, Wannan Tang, Michael C Wehr, Sandra Goebbels, Andreas Reichenbach, Rolf Sprengel, Moritz J Rossner, and Frank Kirchhoff. Split-cre complementation indicates coincident activity of different genes in vivo. *PloS one*, 4:e4286, 2009.

[244] Petter Portin. Studies on gynandromorphs induced with the claret-nondisjunctional mutation of drosophila melanogaster. an approach to the timing of chromosome loss in cleavage mitoses. *Heredity*, 41:193–203, 10 1978.

[245] Shannon L Griswold and Peter Y Lwigale. Analysis of neural crest migration and differentiation by cross-species transplantation. *Journal of visualized experiments : JoVE*, 2 2012.

[246] Weike Pei, Thorsten B Feyerabend, Jens Rössler, Xi Wang, Daniel Postrach, Katrin Busch, Immanuel Rode, Kay Klapproth, Nikolaus Dietlein, Claudia Quedenau, Wei Chen, Sascha Sauer, Stephan Wolf, Thomas Höfer, and Hans-Reimer Rodewald. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature*, 548:456–460, 8 2017.

[247] Yuval Rinkevich, Paul Lindau, Hiroo Ueno, Michael T Longaker, and Irving L Weissman. Germ-layer and lineage-restricted stem/progenitors regenerate the mouse digit tip. *Nature*, 476:409–13, 8 2011.

[248] Hugo J Snippert, Laurens G van der Flier, Toshiro Sato, Johan H van Es, Maaike van den Born, Carla Kroon-Veenboer, Nick Barker, Allon M Klein, Jacco van Rheenen, Benjamin D Simons, and Hans Clevers. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing lgr5 stem cells. *Cell*, 143:134–44, 10 2010.

[249] C Walsh and C L Cepko. Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. *Science (New York, N.Y.)*, 255:434–40, 1 1992.

[250] Calum Gabbutt, Nicholas A Wright, Ann-Marie Baker, Darryl Shibata, and Trevor A Graham. Lineage tracing in human tissues. *The Journal of Pathology*, 257:501–512, 7 2022.

[251] Mackenzie Bloom, Camille Malouf, Alejo Rodriguez-Fraticelli, Adam C. Wilkinson, Vijay G. Sankaran, and Ana Cvejic. Exploiting somatic mutations to decipher human blood production: a natural lineage-tracing strategy. *Experimental Hematology*, 121:2–5, 5 2023.

[252] Michael Spencer Chapman, Anna Maria Ranzoni, Brynelle Myers, Nicholas Williams, Tim H H Coorens, Emily Mitchell, Timothy Butler, Kevin J

Dawson, Yvette Hooks, Luiza Moore, Jyoti Nangalia, Philip S Robinson, Kenichi Yoshida, Elizabeth Hook, Peter J Campbell, and Ana Cvejic. Lineage tracing of human development through somatic mutations. *Nature*, 595:85–90, 7 2021.

[253] Nicole A Suarez, Angela Macia, and Alysson R Muotri. Line-1 retrotransposons in healthy and diseased human brain. *Developmental neurobiology*, 78:434–455, 5 2018.

[254] Lennart Kester and Alexander van Oudenaarden. Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell*, 23:166–179, 8 2018.

[255] Leif S Ludwig, Caleb A Lareau, Jacob C Ulirsch, Elena Christian, Christoph Muus, Lauren H Li, Karin Pelka, Will Ge, Yaara Oren, Alison Brack, Travis Law, Christopher Rodman, Jonathan H Chen, Genevieve M Boland, Nir Hacohen, Orit Rozenblatt-Rosen, Martin J Aryee, Jason D Buenrostro, Aviv Regev, and Vijay G Sankaran. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell*, 176:1325–1339.e22, 3 2019.

[256] Bastiaan Spanjaard, Bo Hu, Nina Mitic, Pedro Olivares-Chauvet, Sharan Janjuha, Nikolay Ninov, and Jan Philipp Junker. Simultaneous lineage tracing and cell-type identification using crispr–cas9-induced genetic scars. *Nature Biotechnology*, 36:469–473, 5 2018.

[257] Hamim Zafar, Chieh Lin, and Ziv Bar-Joseph. Single-cell lineage tracing by integrating crispr-cas9 mutations with transcriptomic data. *Nature Communications*, 11:3055, 6 2020.

[258] Fuyi Chen, Joel Rosiene, Alicia Che, Albert Becker, and Joseph LoTurco. Tracking and transforming neocortical progenitors by crispr/cas9 gene targeting and piggybac transposase lineage labeling. *Development*, 1 2015.

[259] Maximilian Amberger and Zoltán Ivics. Latest advances for the sleeping beauty transposon system: 23 years of insomnia but prettier than ever. *BioEssays*, 42, 11 2020.

[260] Wenjun Kong, Brent A. Biddy, Kenji Kamimoto, Junedh M. Amrute, Emily G. Butka, and Samantha A. Morris. Celltagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution. *Nature Protocols*, 15:750–772, 3 2020.

[261] Aaron McKenna, Gregory M Findlay, James A Gagnon, Marshall S Horwitz, Alexander F Schier, and Jay Shendure. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science (New York, N.Y.)*, 353:aaf7907, 7 2016.

[262] Yang Yang, Shuo Li, Yujiao Wang, Yi Zhao, and Qiu Li. Protein tyrosine kinase inhibitor resistance in malignant tumors: molecular mechanisms and future perspective. *Signal transduction and targeted therapy*, 7:329, 9 2022.

[263] Anna Alemany, Maria Florescu, Chloé S Baron, Josi Peterson-Maduro, and Alexander van Oudenaarden. Whole-organism clone tracing using single-cell sequencing. *Nature*, 556:108–112, 4 2018.

[264] Gillian M Keating. Afatinib: A review in advanced non-small cell lung cancer. *Targeted oncology*, 11:825–835, 12 2016.

[265] Po-Han Lin, Ling-Ming Tseng, Yi-Hsuan Lee, Shou-Tung Chen, Dah-Cherng Yeh, Ming-Shen Dai, Liang-Chih Liu, Ming-Yang Wang, Chiao Lo, Stanley Chang, Kien Thiam Tan, Shu-Jen Chen, Sung-Hsin Kuo, and Chiun-Sheng Huang. Neoadjuvant afatinib with paclitaxel for triple-negative breast cancer and the molecular characteristics in responders and non-responders. *Journal of the Formosan Medical Association*, 121:2538–2547, 12 2022.

[266] Alexandra Canonici, Alacoque L. Browne, Mohamed F. K. Ibrahim, Kevin P. Fanning, Sandra Roche, Neil T. Conlon, Fiona O'Neill, Justine Meiller, Mattia Cremona, Clare Morgan, Bryan T. Hennessy, Alex J. Eustace, Flavio Solca, Norma O'Donovan, and John Crown. Combined targeting egfr and src as a potential novel therapeutic approach for the treatment of triple negative breast cancer. *Therapeutic Advances in Medical Oncology*, 12:175883591989754, 1 2020.

[267] Tia H. Turner, Mohammad A. Alzubi, and J. Chuck Harrell. Identification of synergistic drug combinations using breast cancer patient-derived xenografts. *Scientific Reports*, 10:1493, 1 2020.

[268] https://www.cellecta.com.

[269] Suntae Kim, Ji Hye Im, Wan Kyu Kim, Young Jae Choi, Ji-Yoon Lee, Sang Kyum Kim, Sun Jo Kim, Sung Won Kwon, and Keon Wook Kang. Enhanced sensitivity of nonsmall cell lung cancer with acquired resistance to epidermal growth factor receptor-tyrosine kinase inhibitors to phenformin: The roles of a metabolic shift to oxidative phosphorylation and redox balance. *Oxidative medicine and cellular longevity*, 2021:5428364, 2021.

[270] María Merino Salvador, Marta Gómez de Cedrón, Juan Moreno Rubio, Sandra Falagán Martínez, Ruth Sánchez Martínez, Enrique Casado, Ana Ramírez de Molina, and María Sereno. Lipid metabolism and lung cancer. *Critical reviews in oncology/hematology*, 112:31–40, 4 2017.

[271] V C Russo, W J Azar, S W Yau, M A Sabin, and G A Werther. Igfbp-2: The dark horse in metabolism and cancer. *Cytokine growth factor reviews*, 26:329–46, 6 2015.

[272] Lai-Feng Wei, Xue-Fen Weng, Xu-Chun Huang, Yu-Hui Peng, Hai-Peng Guo, and Yi-Wei Xu. Igfbp2 in cancer: Pathological role and clinical significance (review). *Oncology reports*, 45:427–438, 2 2021.

[273] Yufang Ma, Nan Tang, Reid C Thompson, Bret C Mobley, Steven W Clark, Jann N Sarkaria, and Jialiang Wang. Insr/igf1r pathway mediates resistance to egfr inhibitors in glioblastoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 22:1767–76, 4 2016.

[274] C Maris, N D'Haene, A-L Trépant, M Le Mercier, S Sauvage, J Allard, S Rorive, P Demetter, C Decaestecker, and I Salmon. Igf-ir: a new prognostic biomarker for human glioblastoma. *British journal of cancer*, 113:729–37, 9 2015.

[275] Jingsheng Yuan, Zhijie Yin, Kaixiong Tao, Guobing Wang, and Jinbo Gao. Function of insulin-like growth factor 1 receptor in cancer resistance to chemotherapy (review). *Oncology Letters*, 10 2017.

[276] Zhennan Tian, Guodong Yao, Hongtao Song, Yang Zhou, and Jingshu Geng. Igf2r expression is associated with the chemotherapy response and prognosis of patients with advanced nsclc. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*, 34:1578–88, 2014.

[277] Yongik Lee, Yian Wang, Michael James, Joseph H Jeong, and Ming You. Inhibition of igf1r signaling abrogates resistance to afatinib (bibw2992) in egfr t790m mutant lung cancer cells. *Molecular carcinogenesis*, 55:991–1001, 5 2016.

[278] Kunyu Shi, Guan Wang, Junping Pei, Jifa Zhang, Jiaxing Wang, Liang Ouyang, Yuxi Wang, and Weimin Li. Emerging strategies to overcome resistance to third-generation egfr inhibitors. *Journal of hematology oncology*, 15:94, 7 2022.

[279] Jingxian Zhang, Jia Jia, Feng Zhu, Xiaohua Ma, Bucong Han, Xiaona Wei, Chunyan Tan, Yuyang Jiang, and Yuzong Chen. Analysis of bypass signaling in egfr pathway and profiling of bypass genes for predicting response to anticancer egfr tyrosine kinase inhibitors. *Molecular bioSystems*, 8:2645–56, 10 2012.

[280] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. 6 2019.

[281] Zhirong Wu, Yuanjun Xiong, Stella Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance-level discrimination. 5 2018.

[282] Tete Xiao, Xiaolong Wang, Alexei A. Efros, and Trevor Darrell. What should not be contrastive in contrastive learning. 8 2020.

[283] Applications of support vector machine (svm) learning in cancer genomics. *Cancer Genomics Proteomics*, 15, 1 2018.

[284] Ashanira Mat Deris, Azlan Mohd Zain, and Roselina Sallehuddin. Overview of support vector machine in modeling machining performances. *Procedia Engineering*, 24:308–312, 2011.

[285] Armin Shmilovici. *Support Vector Machines*, pages 257–276. Springer-Verlag.

[286] Will Macnair, Revant Gupta, and Manfred Claassen. psupertime: supervised pseudotime analysis for time-series single-cell rna-seq data. *Bioinformatics (Oxford, England)*, 38:i290–i298, 6 2022.

[287] Agata Michna, Herbert Braselmann, Martin Selmansberger, Anne Dietz, Julia Hess, Maria Gomolka, Sabine Hornhardt, Nils Blüthgen, Horst Zitzelsberger, and Kristian Unger. Natural cubic spline regression modeling followed by dynamic network reconstruction for the identification of radiation-sensitivity gene association networks from time-course transcriptome data. *PloS one*, 11:e0160791, 2016.

[288] Yuping Gong, Muralikrishna Duvvuri, Michael B Duncan, Jian Liu, and Jeffrey P Krise. Niemann-pick c1 protein facilitates the efflux of the anticancer drug daunorubicin from cells according to a novel vesicle-mediated pathway. *The Journal of pharmacology and experimental therapeutics*, 316:242–7, 1 2006.

[289] Line Groth-Pedersen, Marie Stampe Ostenfeld, Maria Høyer-Hansen, Jesper Nylandsted, and Marja Jäättelä. Vincristine induces dramatic lysosomal changes and sensitizes cancer cells to lysosome-destabilizing siramesine. *Cancer research*, 67:2217–25, 3 2007.

[290] R.Brooks Robey, Judith Weisz, Nancy B. Kuemmerle, Anna C. Salzberg, Arthur Berg, Dustin G. Brown, Laura Kubik, Roberta Palorini, Fahd Al-Mulla, Rabeah Al-Temaimi, Anna Maria Colacci, Chiara Mondello, Jayadev Raju, Jordan Woodrick, A.Ivana Scovassi, Neetu Singh, Monica Vaccari, Rabindra Roy, Stefano Forte, Lorenzo Memeo, Hosni K.

Salem, Amedeo Amedei, Roslida A. Hamid, Graeme P. Williams, Leroy Lowe, Joel Meyer, Francis L. Martin, William H. Bisson, Ferdinando Chiaradonna, and Elizabeth P. Ryan. Metabolic reprogramming and dysregulated metabolism: cause, consequence and/or enabler of environmental carcinogenesis? *Carcinogenesis*, 36:S203–S231, 6 2015.

[291] Zachary E Stine, Zachary T Schug, Joseph M Salvino, and Chi V Dang. Targeting cancer metabolism in the era of precision oncology. *Nature reviews Drug discovery*, 21(2):141–162, 2022.

[292] Chuncao Xu, Lei Zhang, Daifei Wang, Shiqin Jiang, Di Cao, Zhongxiang Zhao, Min Huang, and Jing Jin. Lipidomics reveals that sustained srebp-1-dependent lipogenesis is a key mediator of gefitinib-acquired resistance in egfr-mutant lung cancer. *Cell Death Discovery*, 7:353, 11 2021.

[293] Rimsha Munir, Jan Lisec, Johannes V. Swinnen, and Nousheen Zaidi. Lipid metabolism in cancer cells under metabolic stress. *British Journal of Cancer*, 120:1090–1098, 6 2019.

[294] Evelien Rysman, Koen Brusselmans, Katryn Scheys, Leen Timmermans, Rita Derua, Sebastian Munck, Paul P Van Veldhoven, David Waltregny, Veerle W Daniëls, Jelle Machiels, Frank Vanderhoydonc, Karine Smans, Etienne Waelkens, Guido Verhoeven, and Johannes V Swinnen. De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer research*, 70:8117–26, 10 2010.

[295] Carlo Cenciarelli, Hany E S Marei, Manuela Zonfrillo, Pasquale Pierimarchi, Emanuela Paldino, Patrizia Casalbore, Armando Felsani, Angelo Luigi Vescovi, Giulio Maira, and Annunziato Mangiola. Pdgf receptor alpha inhibition induces apoptosis in glioblastoma cancer stem cells refractory to anti-notch and anti-egfr treatment. *Molecular cancer*, 13:247, 11 2014.

[296] Koichi Azuma, Akihiko Kawahara, Kahori Sonoda, Kazutaka Nakashima, Kousuke Tashiro, Kosuke Watari, Hiroto Izumi, Masayoshi Kage, Michihiko Kuwano, Mayumi Ono, and Tomoaki Hoshino. Fgfr1 activation is an escape mechanism in human lung cancer cells resistant to afatinib, a pan-egfr family kinase inhibitor. *Oncotarget*, 5:5908–19, 8 2014.

[297] Hideki Terai, Kenzo Soejima, Hiroyuki Yasuda, Sohei Nakayama, Junko Hamamoto, Daisuke Arai, Kota Ishioka, Keiko Ohgino, Shinnosuke Ikemura, Takashi Sato, Satoshi Yoda, Ryosuke Satomi, Katsuhiko Naoki, and Tomoko Betsuyaku. Activation of the fgf2-fgfr1 autocrine pathway: a novel mechanism of acquired resistance to gefitinib in nsclc. *Molecular cancer research : MCR*, 11:759–67, 7 2013.

[298] Victor Cy Mak, Xinran Li, Ling Rao, Yuan Zhou, Sai-Wah Tsao, and Lydia Wt Cheung. p85 alters response to egfr inhibitor in ovarian cancer through p38 mapk-mediated regulation of dna repair. *Neoplasia (New York, N.Y.)*, 23:718–730, 7 2021.

[299] Alessandro Leonetti, Sugandhi Sharma, Roberta Minari, Paola Perego, Elisa Giovannetti, and Marcello Tiseo. Resistance mechanisms to osimertinib in egfr-mutated non-small cell lung cancer. *British journal of cancer*, 121:725–737, 10 2019.

[300] Lihua Huang and Liwu Fu. Mechanisms of resistance to egfr tyrosine kinase inhibitors. *Acta Pharmaceutica Sinica B*, 5:390–401, 9 2015.

[301] Robert Roskoski. Targeting erk1/2 protein-serine/threonine kinases in human cancers. *Pharmacological research*, 142:151–168, 4 2019.

[302] Guangzhi Ma, Yunfu Deng, Luxi Qian, Karin A Vallega, Guojing Zhang, Xingming Deng, Taofeek K Owonikoko, Suresh S Ramalingam, Douglas D Fang, Yifan Zhai, and Shi-Yong Sun. Overcoming acquired resistance to third-generation egfr inhibitors by targeting activation of intrinsic apoptotic pathway through mcl-1 inhibition, bax activation, or both. *Oncogene*, 41:1691–1700, 3 2022.

[303] Chien-Hui Weng, Li-Yu Chen, Yu-Chin Lin, Jin-Yuan Shih, Yun-Chieh Lin, Ruo-Yu Tseng, An-Chieh Chiu, Yu-Hsuan Yeh, Chi Liu, Yi-Ting Lin, Jim-Min Fang, and Ching-Chow Chen. Epithelial-mesenchymal transition (emt) beyond egfr mutations per se is a common mechanism for acquired resistance to egfr tki. *Oncogene*, 38:455–468, 1 2019.

[304] Hirokazu Taniguchi, Tadaaki Yamada, Rong Wang, Keiko Tanimura, Yuta Adachi, Akihiro Nishiyama, Azusa Tanimoto, Shinji Takeuchi, Luiz H Araujo, Mariana Boroni, Akihiro Yoshimura, Shinsuke Shiotsu, Isao Matsumoto, Satoshi Watanabe, Toshiaki Kikuchi, Satoru Miura, Hiroshi Tanaka, Takeshi Kitazaki, Hiroyuki Yamaguchi, Hiroshi Mukae, Junji Uchino, Hisanori Uehara, Koichi Takayama, and Seiji Yano. Axl confers intrinsic resistance to osimertinib and advances the emergence of tolerant cells. *Nature communications*, 10:259, 1 2019.

[305] Donghwa Kim, Duc-Hiep Bach, Yan-Hua Fan, Thi-Thu-Trang Luu, Ji-Young Hong, Hyen Joo Park, and Sang Kook Lee. Axl degradation in combination with egfr-tki can delay and overcome acquired resistance in human non-small cell lung cancer cells. *Cell death disease*, 10:361, 5 2019.

[306] Sander Bekeschus. Acquired cancer tyrosine kinase inhibitor resistance: Ros as critical determinants. *Signal transduction and targeted therapy*, 6:437, 12 2021.

[307] Giovanni Y Di Veroli, Chiara Fornari, Dennis Wang, Severine Mollard, Jo L Bramhall, Frances M Richards, and Duncan I Jodrell. Combenefit: an interactive platform for the analysis and visualization of drug combinations. *Bioinformatics*, 32(18):2866–2868, 2016.

[308] https://portals.broadinstitute.org/gpp/public.

[309] https://github.com/gambalab/gficf.

[310] Vitalii Kleshchevnikov, Artem Shmatko, Emma Dann, Alexander Aivazidis, Hamish W King, Tong Li, Rasa Elmentaite, Artem Lomakin, Veronika Kedlian, Adam Gayoso, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature biotechnology*, 40(5):661–671, 2022.

[311] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[312] https://github.com/gambalab/scastral.

[313] https://github.com/gambalab/dropseq.

[314] Alfredo A Kalaitzis and Neil D Lawrence. A simple approach to ranking differentially expressed gene expression time courses through gaussian process regression. *BMC Bioinformatics*, 12:180, 12 2011.

[315] Rashidi Dzul Keflee, Kok Hoong Leong, Satoshi Ogawa, Jerome Bignon, Mun Chiang Chan, and Kin Weng Kong. Overview of the multifaceted resistances toward egfr-tkis and new chemotherapeutic strategies in non-small cell lung cancer. *Biochemical Pharmacology*, page 115262, 2022.

Author's publications

1. Franchini, M., Pellecchia, S., Viscido, G., Gambardella, G. (2023). Single-cell gene set enrichment analysis and transfer learning for functional annotation of scRNA-seq data. NAR Genomics and Bioinformatics, 5(1), lqad024.

2. Pellecchia, S., Viscido, G., Franchini, M. et al. Predicting drug response from single-cell expression profiles of tumours. BMC Med 21, 476 (2023). https://doi.org/10.1186/s12916-023-03182-1

3. Pellecchia, S., Franchini, M., Viscido, G., Arnese, R., Gambardella, G. (2023). Single cell lineage tracing reveals subclonal dynamics of anti-EGFR therapy resistance in triple negative breast cancer. bioRxiv, 2023-04.