TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

NOVEL TOOLS FOR IDENTIFICATION OF ONCOGENIC DRIVER MUTATIONS

Deepankar Chakroborty

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1679 | MEDICA – ODONTOLOGICA | TURKU 2022

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

NOVEL TOOLS FOR IDENTIFICATION OF ONCOGENIC DRIVER MUTATIONS

Deepankar Chakroborty

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1679 | MEDICA – ODONTOLOGICA | TURKU 2022

University of Turku

Faculty of Medicine
Institute of Biomedicine
Medical Biochemistry and Genetics
Turku Doctoral Programme of Molecular Medicine (TuDMM)

Supervised by

Professor Klaus Elenius, MD, PhD
Institute of Biomedicine and
Medicity Research Laboratories
University of Turku
Turku Bioscience
University of Turku and Åbo Akademi
Turku, Finland

Professor Laura Elo, PhD
Institute of Biomedicine
University of Turku
Turku Bioscience
University of Turku and Åbo Akademi
Turku, Finland

Reviewed by

Professor Juha Kere MD, PhD
Department of Biosciences and Nutrition
Karolinska Institutet
Stockholm, Sweden

Professor George Steven Bova, MD
Faculty of Medicine and Health Technology
Tampere University
Tampere, Finland

Opponent

Professor Rene Bernards, PhD
Department of Molecular Carcinogenesis
Netherlands Cancer Institute
Amsterdam, Netherlands

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

Cover Image: Deepankar Chakroborty

ISBN 978-951-29-9097-9 (PRINT)
ISBN 978-951-29-9098-6 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama Oy, Turku, Finland 2022

"যদি তোর ডাক শুনে কেউ না আসে তবে একলা চলো রে।"
(Jôdi Tor Daak Shune Keu Naa Ashe Tôbe Ekla Chôlo Re)

If they answer not to thy call, walk alone.

- Rabindranath Tagore, 1905

UNIVERSITY OF TURKU
Faculty of Medicine
Institute of Biomedicine
Medical Biochemistry and Genetics
DEEPANKAR CHAKROBORTY: Novel Tools for Identification of Oncogenic Driver Mutations
Doctoral Dissertation, 175 pp.
Turku Doctoral Programme of Molecular Medicine
November 2022

ABSTRACT

Genetic alterations contribute to the development and pathogenesis of several human cancers. These mutations accumulate in a cancer tissue over the course of time due to the instability of the cancer genome. Large-scale sequencing efforts have enabled identification of an abundance of these somatic mutations, and the amount of data is constantly increasing due to the improved accessibility of next-generation sequencing technologies. From this multitude of cancer-associated somatic mutations, a large majority are predicted to be inconsequential "passenger" mutations, (i.e., mutations which do not confer a selective growth advantage to the cancer cells); and only a handful have been validated as "driver" mutations (i.e., mutations playing a critical role in the development or maintenance of cancer). These driver mutations also function as predictive markers for survival, therapeutic efficacy, and often make the cancer cells susceptible to therapeutic intervention.

Identification of driver mutations is an integral part of biomarker discovery in cancer research, and my thesis aimed to address this by developing a screening platform and a database. The in vitro Screen for Activating Mutations (iSCREAM) is a high-throughput screening workflow which was established with Epidermal Growth Factor Receptor (EGFR) as a model. The screen was validated by detection of known activating mutations like EGFR L858R. A previously known EGFR variant of unknown significance (VUS), EGFR A702V, was discovered in the screen and was functionally characterized to be an activating mutation. The iSCREAM screening methodology was further used to systematically study ERBB4, another gene in the EGFR family of receptor tyrosine kinases. We detected ERBB4 VUS R687K, and E715K in the screen and identify them as activating mutations. The ERBB4 mutations were characterized for their effect on ERBB4 phosphorylation, their sensitivity to various tyrosine kinase inhibitors, and their tumorigenicity was evaluated with *in vivo* allografts.

The Database Of Recurrent Mutations (DORM), was prepared by analyzing a public registry of somatic mutations and preparing a catalog of the mutations identified from genome-wide studies to recapitulate the "real-world" frequency of all the recurrent ($n > 1$) somatic mutations. DORM allows limiting the scope of search to 38 tissue types and supports advanced queries using regular expressions. The easy-to-use database and its backend are written to be very responsive and fast in comparison to contemporary public cancer databases.

Taken together, the findings and resources presented in this thesis establish grounds for further studies with other tyrosine kinases and potentially enable diversification into new niches.

KEYWORDS: EGFR, ERBB4, high-throughput screen, functional genomics, activating mutations, cancer, database, recurrent mutations

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Bioläketieteen laitos
Lääketieteellinen biokemia ja genetiikka
DEEPANKAR CHAKROBORTY: Uusia työkaluja syöpää aiheuttavien mutaatioiden tunnistamiseksi
Väitöskirja, 175 s.
Molekyylilääketieteen tohtoriohjelma
Marraskuu 2022

TIIVISTELMÄ

Geneettiset muutokset vaikuttavat useiden ihmisen syöpien syntyyn ja kehittymiseen. Syöpäkudokseen geenimutaatioita kertyy yhä enemmän ajan kuluessa syövän genomisen instabiliteetin vuoksi. Laajamittaisten sekvensointihankkeiden avulla on pystytty tunnistamaan paljon erilaisia somaattisia eli hankinnallisia mutaatioita ja sekvensointitulosten määrä kasvaa jatkuvasti uuden sukupolven sekvensointi-tekniikoiden (engl. *next generation sequencing*, NGS) paremman saatavuuden ansiosta. Näistä lukuisista syöpään liittyvistä somaattisista mutaatioista suurin osa on potilaan ennusteen kannalta merkityksettömiä "matkustajamutaatioita" (engl. passenger mutation) eli mutaatioita, jotka eivät anna valikoivaa kasvuetua syöpäsoluille. Vain muutamia somaattisia mutaatioita on validoitu "ajajamutaatioiksi" (engl. driver mutation) eli mutaatioiksi, joilla on kriittinen rooli syövän kehittymisessä tai ylläpitämisessä. Nämä ajajamutaatiot toimivat usein eloonjäämisen sekä hoidon tehon ennusteellisina markkereina ja usein myös herkistävät syöpäsoluja hoidoille.

Ajajamutaatioiden tunnistaminen on olennainen osa syövän biomarkkereiden tutkimusta. Väitöskirjatyöni tavoitteena oli kehittää ajajamutaatioiden seulonta-alusta ja tietokanta. Aktivoivien mutaatioiden *in vitro* -seulonta (engl. *in vitro Screen for Activating Mutations,* iSCREAM) on tehoseulontamenetelmä, jonka kehittämistyössä käytettiin mallina epidermaalista kasvutekijäreseptoria (EGFR) koodaavaa geeniä. iSCREAM-seulonnalla tunnistettiin jo tunnettuja aktivoivia EGFR-mutaatioita, kuten L858R, mikä validoi menetelmän toimivuuden. Seulontamenetelmällä tunnistettiin ja karakterisoitiin myös uusi EGFR-geenin aktivoiva mutaatio, A702V, jonka oletettu toimintamekanismi selvitettiin. iSCREAM-seulontamenetelmää hyödynnettiin tässä työssä myös EGFR-reseptorityrosiinikinaasiperheen toisen geenin, ERBB4-geenin, systemaattiseen tutkimiseen, jonka avulla löydettiin uusina aktivoivina mutaatioina ERBB4 R687K ja E715K. Näiden ERBB4-mutaatioiden vaikutusta ERBB4:n fosforylaatioon ja lääkeherkkyyteen erilaisille tyrosiinikinaasiestäjille karakterisoitiin, ja niiden tuumorigeenisyys validoitiin *in vivo* -allografteissa.

Toistuvien mutaatioiden tietokanta (engl. *Database Of Recurrent Mutations*, DORM) luotiin analysoimalla somaattisten mutaatioiden julkista rekisteriä ja laatimalla luettelo genominlaajuisissa tutkimuksissa tunnistetuista mutaatioista, jotta kaikkien toistuvien (n > 1) somaattisten mutaatioiden "todellinen" esiintymistiheys voitaisiin laskea. DORM mahdollistaa haun rajoittamisen 38:aan kudostyyppiin ja tukee edistyneempiä kyselyjä säännöllisten lausekkeiden (engl. *regular expression*) avulla. Helppokäyttöinen tietokanta ja sen taustajärjestelmä kehitettiin hyvin reagoivaksi ja nopeaksi nykyisiin julkisiin syöpätietokantoihin verrattuna.

Tässä työssä esitetyt havainnot ja resurssit luovat yhdessä perustan jatkotutkimuksille muilla tyrosiinikinaaseilla ja ovat mahdollisesti laajennettavissa muillekin tutkimusalueille.

AVAINSANAT: EGFR, ERBB4, tehoseulonta, funktionaalinen genomiikka, aktivoivat mutaatiot, syöpä, tietokanta, toistuvat mutaatiot

Abbreviations

AACR-GENIE	American Association for Cancer Research for the project Genomics, Evidence, Neoplasia, Information, Exchange
AKT	AKT serine/threonine kinase 1
AREG	amphiregulin
ATP	adenosine triphosphate
BCR-ABL	breakpoint cluster region - c-abl oncogene
BRAF	b-raf proto-oncogene, serine/threonine kinase
BRCA1	BRCA1 DNA repair associated (also known as Breast Cancer 1)
BRCA2	BRCA2 DNA repair associated (also known as Breast Cancer 2)
BTC	betacellulin
COSMIC	Catalogue Of Somatic Mutations In Cancer
CRC	colorectal cancer
CYT	cytoplasmic
DAG	diacylglycerol
DMS	deep mutational scanning
DNA	deoxyribonucleic acid
DORM	Database Of Recurrent Mutations
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EGFRvIII	EGFR variant III
EMA	European Medicines Agency
EMPIRIC	Exceedingly Methodical and Parallel Investigation of Randomized Individual Codons
ENU	N-ethyl-N-nitrosourea
EPGN	epigen (epithelial mitogen)
EREG	epiregulin
ERK	mitogen-activated protein kinase 1
FCS	fetal calf serum
FDA	United States Food and Drug Administration
GRB2	growth factor receptor-bound protein 2
HB-EGF	heparin-binding EGF-like growth factor
HER	human EGF receptor

HNSCC	head and neck squamous cell carcinoma
ICD	intracellular domain
ICGC	International Cancer Genome Consortium
IL-3	interleukin-3
iSCREAM	*in vitro* screen for activating mutations
JAK	Janus kinases
JM	juxtamembrane
MANO	Mixed All Nominated mutants in One assay
MAPK	mitogen activated protein kinase
MAVE	Multiplexed Assays of Variant Effects
MEK	mitogen activated protein kinase kinase
MITE	mutagenesis by integrated tiles
MSK-IMPACT	Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets
NGS	next-generation sequencing
NRG	neuregulin (heregulin)
NSCLC	non-small cell lung cancer
ORF	open reading frame
PALS	Programmed Allelic Series
PARP	poly(ADP-ribose) polymerase
PCAWG	Pan-Cancer Analysis of Whole Genomes
PCR	polymerase chain reaction
PDK1	3-phosphoinositide dependent kinase 1
PI3K	phosphatidyl inositol 3-kinase
PIP2	phosphatidylinositol-4,5-phosphate
PIP3	phosphatidylinositol-3,4,5-phosphate
PKC	protein kinase C
PLC-ɣ	phospholipase-C gamma
POPCODE	Precision Oligo-Pool Based Code Alteration
PROVEAN	Protein Variation Effect Analyzer
PTB	phosphotyrosine-binding
RIP	regulated intramembrane proteolysis
RNA	ribonucleic acid
RTK	receptor tyrosine kinase
SH2	src homology 2
SHC	SHC adaptor protein 1
SIFT	Sorting Intolerant From Tolerant
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SOS	son of sevenless
STAT	signal transducers and activators oF transcription

SV	structural variant
TACE	tumor necrosis factor-α-converting enzyme
TCGA	The Cancer Genome Atlas
TGF-α	transforming growth factor alpha
TKI	tyrosine kinase inhibitor
VUS	variants of unknown significance
YAP1	Yes1 associated protein

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Chakroborty, D., Kurppa, K. J., Paatero, I., Ojala, V. K., Koivu, M., Tamirat, M. Z., Koivunen, J. P., Jänne, P. A., Johnson, M. S., Elo, L. L., & Elenius, K. An unbiased in vitro screen for activating epidermal growth factor receptor mutations. *Journal of Biological Chemistry,* 2019*;* 294(24): 9377–9389.

II Chakroborty, D.*, Ojala, V. K.*, Knittle, A. M., Drexler, J., Tamirat, M. Z., Ruzicka, R., Bosch, K., Woertl, J., Schmittner, S., Elo, L. L., Johnson, M. S., Kurppa, K. J., Solca, F., & Elenius, K. An unbiased functional genetics screen identifies rare activating ERBB4 mutations. *Cancer Research Communications*, 2022; 2(1): 10–27.

III Chakroborty, D., Paatero I., Kurppa, K.J., Elenius, K. Database of recurrent mutations (DORM), a webtool to browse recurrent somatic mutations in human cancers. *biorxiv*, 2022; (2022.11.21.517363).

* These authors contributed equally to article II.

Table of Contents

1 Introduction

Alterations in the genome are the fundamental cause of cancer. Advances in genomics gave us tools to peer into cancer genomes and identify key alterations that drive the initiation, maintenance, metastasis, and relapse of cancer. However, this process of characterizing cancer genomes led to the accumulation of numerous variants in the databases that have never been functionally characterized. These variants are referred as Variants of Unknown Significance (VUS). There are primarily two reasons for the apparent "neglect": practical infeasibility of analyzing large number of variants, and lack of representation in the global population (i.e., low frequency of prevalence in the cancer samples analyzed with DNA sequencing). Together, they lead to several scientists preferring simple economics over pure science, as the upfront costs of characterizing a rare VUS outweighs the potential benefits it can have to the very small number of patients who have the variant. This thesis focuses on these problems of VUS and presents two tools that help in characterizing VUS of a gene/protein of interest, and in fairly representing the true share of VUS in the domain of cancer-associated somatic mutations.

The thesis presents the development of a high-throughput screen for gain-of-function mutations in oncogenic kinases. The workflow was used to study EGFR and ERBB4 which are members of the ERBB family of receptor tyrosine kinases. The four ERBB receptors are transmembrane cell-surface proteins that receive extracellular stimulus (by one of eleven ligands) and activate an assortment of well-orchestrated intracellular signaling pathways in a context-dependent manner. These proteins are an integral part of mammalian growth and development; however, genetic alterations have been reported to cause aberrant ERBB signaling in several human cancers. These findings led to the development of inhibitors of ERBB signaling and their successful use in the clinic. The work in this thesis identified several activating mutations in the ERBB receptors (*EGFR* and *ERBB4*) which were previously reported to be VUS, thereby, improving the understanding about ERBB biology.

This thesis also presents a database that was developed to exhibit the true population (i.e., cancer samples that have been analyzed with DNA sequencing across the globe and are aggregated by the COSMIC cancer registry (cancer.sanger.ac.uk)) frequencies of thousands of cancer-associated somatic mutations. The contemporary databases are resource intensive and have a significant delay in processing a user's

query (article III, Figure 1) and have a strong bias (section 6.4, Table 6) towards a selected set of genes and mutations (targeted-panel sequencing). The database developed in this thesis circumvents these issues and is presented as a demonstrably fast and easy-to-use webtool (article III, Figure 1).

2 Review of the Literature

2.1 Genomic alterations are fundamental drivers in human cancers

At the fundamental level, propagation of any terrestrial life relies on replication of its genetic material by a nucleic acid polymerase. In case of mammalian cells, it is carried out by a DNA polymerase that although has a very high fidelity, also has a non-zero error rate (Loeb and Monnat, 2008). These natural errors introduce random changes into the genome of a cell during cell cycle (in the S-phase) which are interchangeably called DNA alterations or mutations. This imperfect, but, completely natural process drives molecular evolution, and over geological timescales may cause speciation (Wilson, 1985). Somatic evolution of cancer shares the basic principles with natural evolution, but unlike natural evolution, the cancer evolution happens within a short human life span. In humans, the DNA polymerase, with an error-rate of ≤ 1 incorrect base-incorporations per 150,000 nucleotides, coupled with an exceptional DNA repair machinery creates less than one mutation per cell division in non-malignant somatic cells (Araten et al., 2005; Cagan et al., 2022; Korona et al., 2011; Roberts and Kunkel, 1988). The accuracy is phenomenal, however when we factor in the billions of cells dividing every day in the human body, the prospects of genomic integrity look bleak. Fortunately, a majority of these accumulated alterations are inconsequential passengers, however, a remaining few are drivers (i.e., DNA changes that play a critical role in the development and/or maintenance of cancer) (Martincorena and Campbell, 2015; Vogelstein et al., 2013). Pan-cancer analysis of whole genomes (PCAWG) revealed a tumor to contain an average of 4-5 driver events (Campbell et al., 2020; Martincorena et al., 2017). Origins of cancer at a cellular level are monoclonal in nature, however, over time, it evolves polyclonally (Fearon et al., 1987; Greaves and Maley, 2012; Linder and Gartler, 1965). The cellular clones, within a cancer tissue, harboring various driver alterations undergo rounds of competitive selection to improve their survival fitness by genetic and non-genetic mechanisms in a process called clonal expansion (Burrell and Swanton, 2014; Gerstung et al., 2020; Greaves and Maley, 2012; McFarland et al., 2014; Podlaha et al., 2012; Posada, 2015).

2.1.1 Effect of DNA alterations on cellular biochemistry

As mentioned earlier, most of the acquired genetic alterations or mutations over the course of time in a cancer tissue are predicted to be passengers (Vogelstein et al., 2013), and this is (in part), due to the genomic context where a particular mutation gets introduced. Proto-oncogenes are genes whose products are involved in normal cell growth, proliferation, and differentiation. However, acquired activating mutations can lead to loss of homeostatic regulation, which renders them oncogenic (Slamon, 1987). For instance, *BRAF* is a gene encoding the B-Raf proto-oncogene (BRAF), a serine/threonine kinase, which regulates cell proliferation via the mitogen activated protein kinase (MAPK) pathway. An acquired DNA change in codon 600 in exon 15 of BRAF (c.1799T>A), creates a GTG > GAG change, which creates a single amino acid substitution of valine 600 to glutamate (Val600Glu or V600E). BRAF V600E is the most frequently observed genetic alteration in patients with metastatic melanoma (Greaves et al., 2013). In the clinic, patients harboring this somatic alteration benefit from BRAF-inhibitors like vemurafenib (Chapman et al., 2011; Sosman et al., 2012; Yang et al., 2010).

On the other hand, tumor-suppressor genes are often involved in regulation of cell cycle and DNA repair. Acquired mutations often inactivate or destroy their physiological function, thereby, compromising the integrity of the cellular genome (Weinberg, 1991). For instance, *BRCA1* and *BRCA2* are genes encoding BRCA1 DNA Repair Associated (BRCA1) and BRCA2 DNA Repair Associated (BRCA2) proteins that are critical in restoring the original DNA sequence at double-strand DNA breaks by homologous recombination repair. Germline and somatic mutations in *BRCA1/2* have been reported in patients with breast and ovarian cancer (Berchuck et al., 1998; Ford et al., 1995; Zhang et al., 2011). The deleterious BRCA gene mutations disrupt their biological activity rendering the cells harboring them more susceptible to acquiring mutations (Kinzler and Vogelstein, 1997; Venkitaraman, 2002; Welcsh et al., 2000). Pharmacological inhibitors of Poly (ADP-Ribose) Polymerase 1 and 2, called PARP inhibitors, pioneered the development of drugs based on the strategy of Synthetic Lethality (Bryant et al., 2005; Farmer et al., 2005). The premise of Synthetic Lethality is, that a defect in either of the two genes (and/or their gene products) alone has little to no effect on cell survival, however, defects in both of the genes together is lethal (Bridges, 1922; Kaelin, 2005; Lucchesi, 1968). In tumors harboring BRCA1/2 alterations, PARP inhibitors stall the DNA replication fork, triggering a DNA damage response. Homologous recombination repair (which involves BRCA1/2 proteins) is the optimal choice for repairing such lesions. However, these cells deficient in BRCA1/2 activity resort to DNA repair mechanisms (e.g., non-homologous end joining) which generate large-scale genomic rearrangements that could be unfavorable for cell survival (Farmer et al., 2005; Lord and Ashworth, 2017).

In addition to directly affecting the biological function of the cancer-associated genes, the acquisition of mutations in the regulatory elements of the tumor-suppressors and oncogenes, can alter the expression of their products thereby impairing the regulation of cellular signaling (Elliott and Larsson, 2021; Khurana et al., 2016; Weinhold et al., 2014).

2.2 Genetic alterations in human cancers

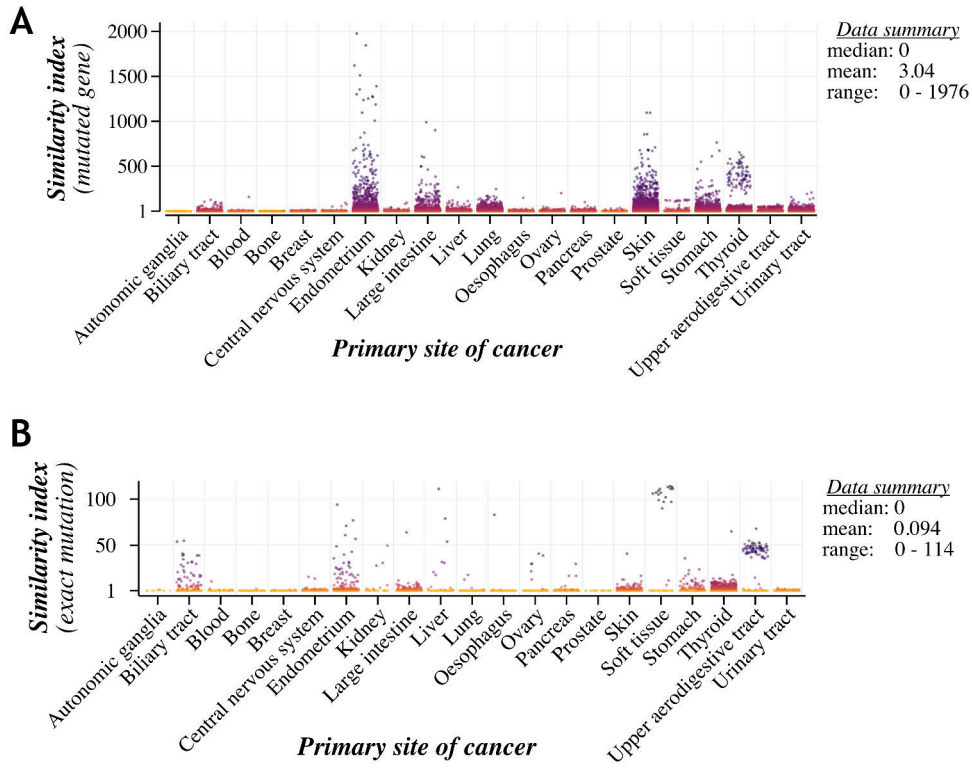


Figure 1. Heterogeneity across primary tumors analyzed from genome-scale DNA sequencing data. The y-axis shows similarity index between two cancer samples chosen randomly from the pool of samples of the same primary site (x-axis). Similarity index is defined as number of shared A) coding alterations in the same gene or B) exact same coding alteration. The individual points show data for 10,000 random pairs of samples from the same primary site (for cancers with more than 200 samples). Only non-zero values have been plotted. (Source of data: COSMIC release v96, cancer.sanger.ac.uk)

Cytogenetic studies revealed the heterogenous nature of human cancers at chromosomal resolution (Pathak, 1990; Wolman, 1986). Further characterization of cancer tissues with genomics revealed that the collection of mutations accumulated by a cancer tissue over the course of time serves like a unique fingerprint, and two

tumors even from the same cancer type might only share a few mutations (**Figure 1**) (Gerlinger et al., 2012; Mcgranahan and Swanton, 2017; Meacham and Morrison, 2013). The heterogeneity has been recognized for a long time (Huxley, 1958; Nervi et al., 1982; Roberts and Tattersall, 1990) however, in depth characterizations and mapping the clonal lineages have happened in the past decade (Jamal-Hanjani et al., 2017; Mitchell et al., 2018; TRACERx Consortium et al., 2020). Furthermore, varying degrees of heterogeneity has been observed between the primary and metastasized tumor samples (Fidler, 1978; Nguyen et al., 2022; Priestley et al., 2019), sometimes, even from the same individual (Naxerova et al., 2017). This is primarily because the metastatic cell population(s) originate from the primary tumor, but as they are in a perpetual state of molecular evolution, with time, they accumulate more changes in their DNA (Turajlic and Swanton, 2016). Tumor evolution is a complex process and is under active deconvolution, evidence exists of cases, where the overall genetic diversity is higher in the primary tumor in comparison to tumors harvested from metastases (Gundem et al., 2015; Woodcock et al., 2020).

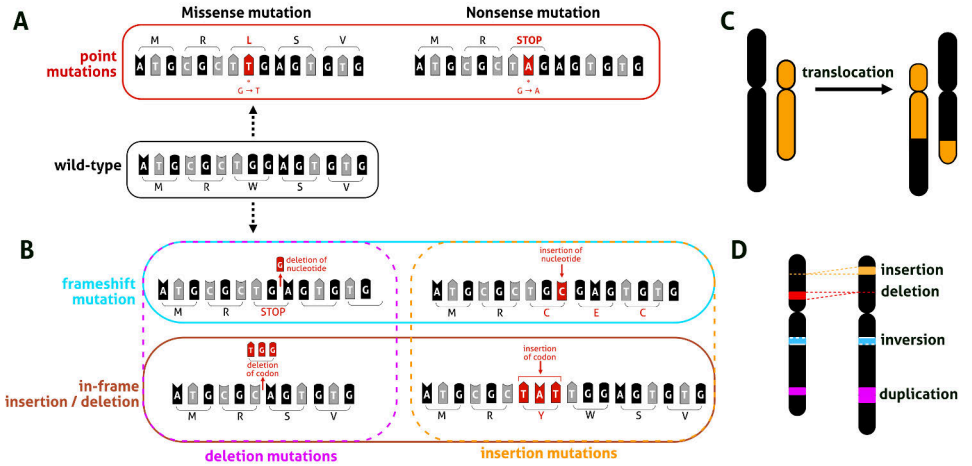


Figure 2. Major types of possible alterations in a DNA and their consequence on the protein sequence. A) Point mutations creating missense or nonsense mutations (in addition to synonymous changes). B) Insertions and deletions can span up to several codons resulting in an altered reading frame. C) Translocation events swap large chunks of genetic information between chromosomes. D) Loci of a chromosome getting altered by insertion, deletion, inversion, and duplication events. (Source: original diagram, panel C-D adapted from (Balachandran and Beck, 2020))

DNA replication and repair are complicated processes involving a large number of proteins, but in principle the mutations can be categorized into three main categories based on the size of the DNA fragment involved (Yates et al., 2017):

1) single nucleotide variants (SNVs): These are created when a nucleotide at a position is changed to one of the other three nucleotides (**Figure 2**). If introduced in a coding region, depending on the new codon, it can change the amino acid sequence in the translated protein. e.g., BRAF V600E in melanoma (Yang et al., 2010), and EGFR L858R in lung adenocarcinoma (Lynch et al., 2004; Paez et al., 2004; Pao et al., 2004).

2) insertions or deletions (indels): These are incorporations or removal of nucleotides, which are typically small (1-5 bp) and less frequently involve larger fragments of DNA (100 bp - several kilobases) (**Figure 2**). If introduced within a coding sequence, these alterations can cause anything from introducing/removing a couple of amino acids up to removing some exons which can affect the translated protein. e.g., EGFR exon 19 deletions in lung adenocarcinoma (Lynch et al., 2004; Paez et al., 2004; Pao et al., 2004).

3) structural variants (SVs): These are changes in the orientation, location, or number of copies of larger chunks of genomic DNA (**Figure 2**). These typically span several kilobases and include gene deletions, duplications as well as larger chromosomal rearrangements like translocations and inversions, and marker chromosomes. e.g., EGFR amplification in glioblastoma multiforme (Libermann et al., 1985), HER2 amplifications in breast cancer (Zeillinger et al., 1989), ALK-fusions in lung cancer (Soda et al., 2007), and the Philadelphia chromosome translocation [t(9;22)(q34;q11)] creating the BCR-ABL fusion in chronic myelogenous leukemia (CML) (Nowell and Hungerford, 1960; Tough et al., 1961).

4) changes in cellular ploidy: These are aberrant changes in structure and number of chromosomes in daughter cells after undergoing a cell division with improper segregation of chromosomes. Aneuploidy results in changes in individual chromosomes e.g., trisomy 12 in chronic lymphocytic leukemia (Losada et al., 1991) and trisomy 8 in acute myeloid leukemia (Philip, 1975; Virtaneva et al., 2001). Polyploidy refers to changes in entire sets of chromosomes. These phenomenon give rise to somatic mosaicism in cancer (Atkin and Ross, 1960; Pellman, 2007; Storchova and Pellman, 2004).

2.3 Identification of driver mutations

As mentioned earlier, driver mutations are the pathogenic alterations, which, in the case of oncogenes, can prove to be potential vulnerabilities of the cancer tissues (Garraway and Lander, 2013; Vogelstein et al., 2013). They have the potential to unravel opportunities for developing targeted inhibitors (Chapman et al., 2011; Hong et al., 2020; Miller et al., 2012) as well as uncovering interesting mechanisms of signal transduction (Brewer et al., 2013; Duensing et al., 2004). Moreover, the detection of driver mutations may help in selection of treatment modalities, improve the clinical actionability and accuracy of prognostic predictions (Lievre et al., 2006; Wood et al.,

2016; Zehir et al., 2017). These reasons are compelling enough to set up a search for driver mutations among the abundance of the variants of unknown significance.

2.3.1 Characterizing cancer genomes with next generation sequencing

The advent of next-generation sequencing (NGS) technologies and their increasing affordability and accessibility enabled analysis of the genetic alterations across large cohorts of cancer genomes (cancer genomes in the magnitude of 10^5 have been characterized (source: gdc.cancer.gov, dcc.icgc.org, database.hartwigmedicalfoundation.nl, and jcga-scc.jp) however annually there are new diagnoses and cancer-associated deaths in the magnitude of 10^6 (Siegel et al., 2022)). The large-scale sequencing efforts have helped identify large number of cancer-associated mutations and helped establish precision medicine (Hyman et al., 2017; Jameson and Longo, 2015). Among the large number of reported mutations in a specific protein, a handful are observed significantly more than other mutations, these are called hotspot mutations (Chang et al., 2016). However, in addition to these hotspots, the next-generation sequencing data also contains millions of mutated nucleotides reported from thousands of cancer genomes, which are classified as infrequent non-hotspot mutations (Chang et al., 2016). Functional classification of a majority of these mutations is incomplete without relevant clinical data and lacks pre-clinical validation, thereby, creating an abundance of Variants of Unknown Significance (VUS) (Dienstmann et al., 2015; "The future of cancer genomics," 2015).

2.3.2 Computational prediction of the consequence of mutations

With the accumulation of data over time, computational methods have been developed to predict the consequences of mutations based on sequence alignment, evolutionary conservation of amino acid residues and protein's structural information. Methods such as Sorting Intolerant From Tolerant (SIFT), PolyPhen-2, and Protein Variation Effect Analyzer (PROVEAN) have been around and in development for a little over two decades (Adzhubei et al., 2010; Choi et al., 2012; Ng and Henikoff, 2001) with consistent improvements by training the models with acquired information from the clinic and functional data. However, for predicting effects of VUS and novel mutations, these tools still lack the sensitivity and specificity needed for clinical decision making (Ernst et al., 2018). More recently, the 'protein folding problem' (Dill et al., 2008) was addressed with AlphaFold, a computational deep learning algorithm that predicts the three-dimensional structure of a polypeptide (16 - 2700 amino acids long) solely based on the sequence. For modelling a polypeptide, the algorithm

incorporates homologous structures (when available), but can also predict structures with atomic accuracy in cases where no homologous structural data is available (Jumper et al., 2021). Further developments could enable accurate modelling of the consequences of somatic mutations on a protein's structure.

2.4 Functional genetics screens

A fundamental issue raised during the characterization of tumor samples by sequencing, was the determination of the VUS. Scientists, having realized the problem, resorted to conducting functional analyses to assign significance to individual VUS (Farrugia et al., 2008). The idea, of performing a functional analysis with variants of a protein generated by systematic mutagenesis dates back several decades. An early example being the alanine-scanning mutagenesis of the human growth hormone to identify residues that modulate its binding to the human growth hormone receptor (Cunningham and Wells, 1989). Systematic mutagenesis required the authors to pre-select which mutations are worth studying. Saturation mutagenesis, on the other hand, is based on the principle of creating a library consisting of all the possible mutations of particular regions in a protein (**Figure 3**) (Derbyshire et al., 1986; Myers et al., 1985).

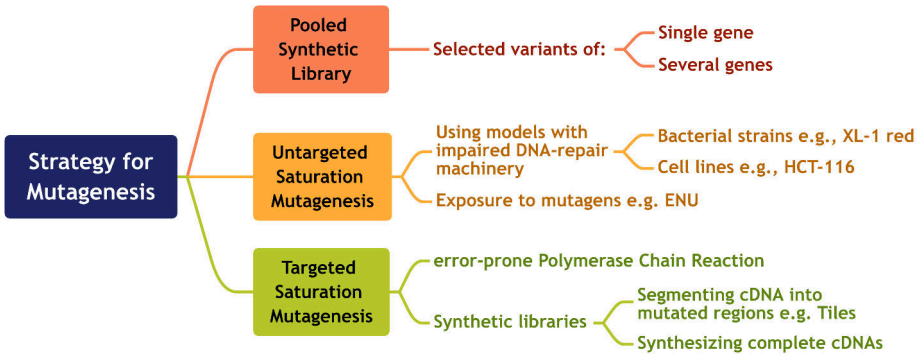


Figure 3. Various approaches for generating libraries of mutations to perform a high-throughput characterization of the variants using a functional genetics screen. (Source: original diagram)

2.4.1 Pooled synthetic libraries

The wealth of mutational data generated by the next generation sequencing has made it feasible to identify candidate mutations and perform functional screens using synthetic pooled libraries to characterize variants (Berger et al., 2015; Ikegami et al., 2020; Kohsaka et al., 2017; Nagano et al., 2018). Here, the expression library consists

of cDNA constructs tagged with unique barcodes. These barcodes enhance the reliability of detection for the individual mutations in the assays because the recognition of a particular mutation is based on several (random) nucleotides of the barcode. By contrast, calling point mutations from the individual cDNAs would be based on mismatch at a single nucleotide. There have been efforts that aimed at characterization of large number of selected mutations, such as the namesake "mixed-all-nominated-mutants-in-one" assay, or the MANO-method (Kohsaka et al., 2017) developed at the laboratory of Dr. Hiroyuki Mano. The screening protocol was used to characterize 101 EGFR mutations (Kohsaka et al., 2017) and 55 ERBB2 mutations (Nagano et al., 2018) which were transduced individually into cell models (Ba/F3 and NIH-3T3) and pooled together to create a "cell library". This library of cells was cultured in a competitive setting to evaluate the transforming potential, and the drug sensitivity of the mutations included in the screen. Similarly, by colleagues at the laboratory of Dr. Jesse Boehm, a pooled screen was carried out using 194 somatic mutations identified in primary lung adenocarcinoma. Including the controls, 352 barcoded open reading frames (ORFs) were generated that were pooled into batches of 70-80 ORFs, introduced into their cell model (SALE-Y cells, which are human small-airway lung epithelial cells immortalized by an activating variant of Yes1 associated protein (YAP1)) injected subcutaneously into immunocompromised mice. The tumors were studied for their genetic composition to unravel the changes in the mutational composition of the pooled ORFs (Berger et al., 2015).

2.4.2 Untargeted saturation mutagenesis

One of the methods of constructing a saturation mutagenesis library is using mutator strains of *Escherichia coli* like XL1-Red (Greener and Callahan, 1994), which are deficient in three of the primary DNA repair pathways in bacteria (a mutated *mutS* impairs error-prone mismatch repair (Radman et al., 1980), mutated *mutD* abrogates 3'-5' exonuclease activity of DNA polymerase III (Scheuermann et al., 1983), and mutated *mutT* incapacitates the bacteria to hydrolyze 8-oxo-dGTP (Cox, 1976)). Propagating plasmids encoding proteins in these cells lead to misincorporations of bases during DNA replication, which over time creates a mutation library. However, there is limited control over mutation frequency, and the entire plasmid is subjected to mutagenesis which can create unwanted biases during the functional screen. Another relatively simple approach for saturation uses exposure of the cells expressing the desired protein to the DNA-alkylating agent *N*-ethyl-*N*-nitrosourea (ENU) which introduces random point mutations (Bradeen et al., 2006). Due to the simplicity of the approach the method is used to identify secondary on-target mutations that confer resistance to targeted therapeutics (Bradeen et al., 2006; Ercan et al., 2015; Kobayashi et al., 2017; Kosaka et al., 2017; Tiedt et al., 2011). A similar approach involves using cells with mismatch repair deficiency e.g., the HCT-116

colon cancer cells, as a background for functional screens (Girdler et al., 2008; Wacker et al., 2012). Both these methods share an undesirable drawback of not being strategies of targeted mutagenesis, i.e., in their process both of the methods indiscriminately mutate the target cells' genomes thereby, compromising its integrity (Glaab and Tindall, 1997; Russell et al., 1979).

2.4.3 Targeted saturation mutagenesis

In order to maintain the integrity of the host-cell's genome and to ensure that only the DNA sequence for the target protein is mutated several methods have been developed for targeted-saturation mutagenesis. The simplest method for generating a mutation library is error-prone Polymerase Chain Reaction (PCR) amplification of the target cDNA (Cadwell and Joyce, 1994), which is an inexpensive and rapid way to generate a mutation library. However, as various thermostable DNA polymerases have inherent biases favoring certain types of DNA base substitutions, using cocktails of polymerases like Mutazyme II which is a commercial blend of a novel mutant of *Taq* DNA polymerase and Mutazyme I (an exonuclease-deficient Pfu DNA polymerase) which produces a less biased mutational spectrum (Cline and Hogrefe, 2000). Another recent development has been the generation of a saturation mutation library by using a mutated HIV Reverse Transcriptase containing the Met230Ile point mutation which, in comparison to the wild type enzyme, has a reduced replication fidelity and an improved overall mutational bias (Yenerall et al., 2021).

However, as the mutations incorporated during the cycles of an error-prone PCR or the viral replication are non-deterministic and random in nature, alternative approaches were developed recently to better control the composition of the mutation library. Methods like Deep Mutational Scanning (DMS) (Fowler et al., 2010; Fowler and Fields, 2014) and Exceedingly Methodical and Parallel Investigation of Randomized Individual Codons (EMPIRIC) (Hietpas et al., 2012) were pioneers, which were soon followed by Mutagenesis by Integrated tiles (MITE) (Melnikov et al., 2014), inverse PCR (Jain and Varadarajan, 2014), Programmed Allelic Series (PALS) (Kitzman et al., 2015) and Precision Oligo-Pool based Code alteration (POPCode) (Weile et al., 2017). These approaches differ in their biochemical implementation and coverage strategy, but, for the incorporating mutations in the library, all of them rely on using synthesized oligonucleotides of varying lengths, scale of synthesis which influence the cost and complexity of designing the library. More recently, advancements in gene synthesis have helped commercial production of synthetic mutation libraries (Plesa et al., 2018).

In the recent years, several contemporary approaches have been developed which leverage libraries built using targeted saturation mutagenesis to conduct functional characterization of the variants. These methods are referred by the umbrella term Multiplexed assays of variant effects (MAVE) (Starita et al., 2017). In practice, a

library of mutated variants of a particular gene are synthesized, cloned into an expression vector, and then introduced into the target cell model(s) where the encoded protein carries out a function that has a phenotype (i.e., mutations have variation in growth rate, drug response, fluorescence, biochemistry, etc.) which can be used to select desired populations. This selection enriches the cells with certain variants and may deplete other variants in the cell population. The cells are harvested, and the mutation library is extracted from the initial and final populations to determine the variant allele frequency of each mutation in the library using high-depth next-generation sequencing (NGS). The fold changes in the frequency of each mutation serves as an indicator of its function (Weile and Roth, 2018).

2.4.4 Ba/F3 cells as a model for kinase activity

The Ba/F3 cells are a murine Interleukin-3 (IL-3) dependent pro-B cells and have been widely used to the study the biochemistry of kinases and kinase inhibitors. The cells were derived in 1984 from an experiment aimed at isolating IL-3 dependent precursors of B-cells from murine bone marrow. The derived cells were devoid of T-cell specific markers and expressed B-cell lineage specific antigens, while lacking the expression of IgM or kappa chains on the cell surface (Palacios et al., 1984). The cells were characterized to be completely dependent on IL-3 for mediating growth and proliferation (exogenous IL-3, or conditioned media of WEHI-3 cells were equally good) and died when cultured in the absence of IL-3 (Palacios et al., 1984). The origin of the Ba/F3 cells, however, is shrouded in a little mystery, as a discrepancy surfaced in 2014 regarding their identity. A group performed single nucleotide polymorphism (SNP) profiling of commonly used mouse strains and cell lines and discovered the Ba/F3 cells to resemble the C3H mouse strain instead of the originally reported BALB/c (Didion et al., 2014). These findings were recapitulated by the RIKEN institute, which has been distributing the cell line since 1992, with a simple sequence length polymorphism (SSLP) analysis (Nakamura, Yukio, 2020).

However, despite this controversy, their utility as an essential biochemical model is widely-accepted due to their interesting properties that facilitate rapid high-throughput quantitative experiments. Ba/F3 cells divide rapidly and grow in suspension, in an inexpensive media formulation (RPMI-1640 supplemented with serum, glutamine, antibiotics and conditioned media of WEHI-3 cells is sufficient) (Palacios et al., 1984). They readily take foreign DNA by means of viral transductions or a simple electroporation, and the variation in their cell proliferation serves an immediate qualitative and quantitative readout. Furthermore, their acute dependence on IL-3 was demonstrated, in 1988, to be surmountable by expression of BCR-ABL tyrosine kinase (Daley and Baltimore, 1988). In addition to, characterizing the fusion product of *BCR-ABL* (generated by the Philadelphia chromosomal translocation) as an oncogenic alteration, this experiment also laid the foundation for using the Ba/F3

transformation assays in studying kinases (Daley and Baltimore, 1988; Warmuth et al., 2007). Ba/F3 cells have been extensively used in studying activating mutations in tyrosine kinases and analyzing their sensitivity and resistance to various tyrosine kinase inhibitors as their dependence on IL-3 can be compensated by activation kinases (Adam et al., 2006; Bradeen et al., 2006; Ercan et al., 2015; Jiang et al., 2005; Shimamura et al., 2006; Thress et al., 2015).

2.5 The Epidermal Growth Factor Receptor family

Kinases are cellular enzymes (Enzyme Commission group 2.7) that catalyze the transfer of a phosphate moiety from adenosine triphosphate (ATP) to their respective substrate which may be either proteins, lipids, carbohydrates, or cellular metabolites. Out of more than 518 known protein kinases in the human genome, 90 are protein tyrosine kinases (EC 2.7.10) and selectively catalyze the phosphorylation of tyrosine (Tyr, Y) residues on their substates (Manning et al., 2002; Robinson et al., 2000). From these, 55 are transmembrane receptors, referred as receptor tyrosine kinases (RTKs) (EC 2.7.10.1), and are sub-categorized into 19 families (Lemmon and Schlessinger, 2010; Robinson et al., 2000; Wheeler and Yarden, 2015).

2.5.1 ERBB receptors

The ERBB or Epidermal Growth Factor (EGF) Receptor family of RTKs consists of the eponymous receptor EGFR (Ullrich et al., 1984; Ushiro and Cohen, 1980), also referred as *ERBB1* due to sequence homology to the retroviral oncogene *v-erb-B* (Downward et al., 1984), and its three homologs *ERBB2* (Coussens et al., 1985), *ERBB3* (Kraus et al., 1989), and *ERBB4* (Plowman et al., 1993a). Additionally, these are also referred in the literature as Human EGF Receptors (HER1, HER2, HER3 and HER4). The ERBB receptors are transmembrane glycoproteins of approximately 180 kDa and consist of an extracellular segment with a ligand binding domain, a transmembrane alpha helix, and an intracellular segment comprising of a bilobular kinase domain which is followed by a C-terminal tail. These four parts, together, facilitate the signal transduction with the transmembrane domain serving as a membrane anchor for the receptors, the extracellular domain initiates the signal transduction by binding an appropriate ligand and inducing receptor dimerization. This brings the individual kinase domains in proximity, which starts the series of phosphorylation events (subject to availability of ATP) on key tyrosine residues on both molecules in the active dimer (Lemmon et al., 2014; Roskoski, 2014). In addition to the dimerization of the extra-cellular domain, the activation of the ERBB-kinase requires asymmetric dimerization of the kinase lobes (details in section 2.5.3). The ERBB family members share the signaling modality mentioned above with two noteworthy exceptions; a) the extracellular segment of ERBB2 is devoid of a ligand-

binding domain and is constitutively in a conformation favoring receptor dimerization (Cho et al., 2003; Garrett et al., 2003), and b) the intrinsic kinase activity of the ERBB3 intracellular segment is heavily impaired due to the lack of key catalytic residues (Guy et al., 1994; Shi et al., 2010).

2.5.2 Ligands of ERBB receptors

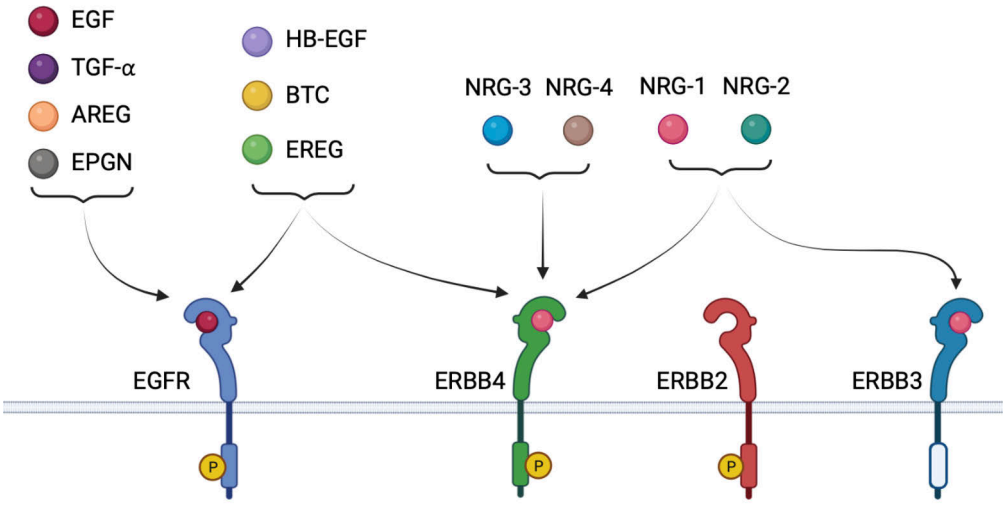


Figure 4. The ERBB family of proteins are transmembrane cell-surface receptor tyrosine kinases. Four ERBBs, and their 11 ligands are depicted and arranged by their specificity for various ERBB receptors. The black arrows indicate the binding of the set of ligands binds to the indicated ERBB receptor. ERBB2 does not have a ligand and is in a confirmation favorable for dimerization with another ERBB receptor. Grey box indicates the impaired kinase domain of ERBB3. (Created in Biorender, adapted from (Lemmon et al., 2014))

The four individual ERBB receptors are activated by the 11 EGF-like ligands (Figure 4) which are synthesized as transmembrane precursors and are released as soluble ligands by undergoing proteolytic cleavage. Four of these ligands exclusively bind to EGFR, namely EGF, amphiregulin (AREG), epigen (EPGN) and the transforming growth factor alpha (TGFA) (Cohen, 1962; Riese et al., 1996b; Strachan et al., 2001). Betacellulin (BTC), epiregulin (EREG), and heparin binding EGF-like growth factor (HBEGF) can bind to both EGFR and ERBB4 (Elenius et al., 1997b; Riese et al., 1996a, 1998). Among the four neuregulins (NRG1, NRG2, NRG3, NRG4), the NRG1 and NRG2 ligands can bind to both ERBB3 and ERBB4 (Carraway III et al., 1997; Carraway et al., 1994; Chang et al., 1997; Plowman et al., 1993b), while the NRG3 and NRG4 ligands exclusively bind to ERBB4 (Harari et al., 1999; Zhang et al., 1997). To date, no EGF-like ligands have been reported to bind ERBB2. This was clarified by crystal structures showing that ERBB2 cannot bind ligands because it is in a pre-

activated confirmation which primes it for dimerization with other ERBBs (Cho et al., 2003; Garrett et al., 2003).

Despite a certain degree of apparent functional redundancy in the biological responses induced by the ligands of ERBB receptors, they are known to induce varying degrees of cell signaling responses of by the virtue of differences in their receptor-bound confirmation, the affinity and duration of their binding to the receptor, availability in various tissues, and their ability to induce various homo- and hetero-dimeric configurations of the ERBB receptors (Freed et al., 2017; Macdonald-Obermann and Pike, 2014; Singh et al., 2016; Sweeney et al., 2000; K. J. Wilson et al., 2012).

2.5.3 Activation of ERBB receptors

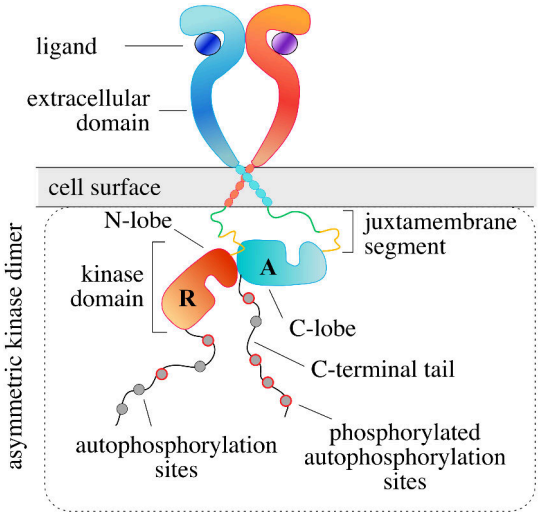


Figure 5. A schematic diagram showing activation of ERBB receptors. Two ligand-bound extracellular domains of ERBB receptors dimerize and lead to the asymmetric dimerization of the intracellular kinase domains. The N-lobe of the receiver kinase (R) and the C-lobe of the activator kinase (A) interactions stabilize the receiver kinase which then transactivates the activator subunit. Subsequently, several tyrosine residues on both the receiver and activator unit are phosphorylated which enables recruitment of molecules facilitating downstream signaling. (Source: original diagram based on Chakroborty et al., 2022; Kurppa et al., 2016.)

Under physiological conditions ERBB receptor activation is mostly dependent on availability of ligands in extracellular space, because, with the exception of ERBB2, all ERBB proteins exist in a closed autoinhibitory confirmation concealing the dimerization arm (located in subdomain II) (Cho and Leahy, 2002; Garrett et al., 2002). Ligand-binding to subdomains I and III induces major confirmational changes resulting in a stabilized active form of the receptor. These changes uncover the

dimerization arm and prime the receptor for interacting with another ERBB receptor in the active confirmation (Garrett et al., 2002; Liu et al., 2012; Ogiso et al., 2002). The dimerization of two extracellular domains leads to activation of the intracellular kinase domains by the formation of a head-to-tail asymmetric dimer (Figure 5) (Endres et al., 2013; Qiu et al., 2008; Zhang et al., 2006). This phenomenon is in contrast to the mechanism of activation of most protein kinases, where the phosphorylation of the activation loop is known to enhance kinase activity (Cargnello and Roux, 2011; Russo et al., 1996; Taylor and Kornev, 2011). ERBB receptors do not require the phosphorylation of activation loop and are activated by the allosteric dimerization of two ERBB kinase units (Jura et al., 2011). Here, one serves as an allosteric "activator kinase", the N-terminal lobe of which comes in contact with the C-terminal lobe of the other kinase unit (referred as "receiver kinase") (Figure 5). These interactions are essential for stabilizing the active confirmation of the receiver kinase, which phosphorylates the activator kinase in *trans* and leads to a cascade of *trans*- phosphorylation events on the C-terminal tails of the receptors. This allosteric mechanism of activation does not require kinase activity, and while underscoring the diversity of heterodimeric configurations of ERBB proteins, it also explains how ERBB3 (deficient in intrinsic kinase activity) forms functional heterodimers exclusively with other ERBBs (Aertgeerts et al., 2011; Endres et al., 2013; Jura et al., 2011, 2009a, 2009b; Qiu et al., 2008; Zhang et al., 2006).

2.5.4 Signal transduction pathways

The active ERBB RTK has phosphorylated intracellular domain and C-terminus tail and these phosphorylated tyrosine residues create docking sites for various cell signaling molecules containing the Src Homology 2 (SH2) or phosphotyrosine-binding (PTB) domains. This allows the ERBB receptors to mediate signaling initiated by an extracellular stimulus (ligand binding) towards the cell's nucleus. The recruited proteins are either phosphorylated by ERBB kinase or by their own intrinsic kinase activity which leads to the activation of specific signal transduction pathway (Lemmon and Schlessinger, 2010). The C-terminal tails of the four ERBB family members are the least homologous region among the four RTKs and harbor unique sets of motifs that enable recruitment of various cell signaling and adaptor molecules, thereby, inducing a diverse cellular response (Schulze et al., 2005). Moreover, the ability of the ERBB family of proteins to form homo- and heterodimers further diversifies the qualitative and quantitative nature of the signaling that is mediated by these proteins (Olayioye et al., 2000, 1998; Riese et al., 1995). The majority of canonical ERBB-mediated signal transduction occurs via these pathways, namely mitogen activated protein kinase (MAPK) pathway, the phosphatidyl inositol 3-kinase (PI3K) pathway, the phospholipase-C gamma (PLC-ɣ) pathway and the signal

transducers and activators of transcription (STAT) pathway (Olayioye et al., 1999) (Figure 6).

For activating the MAPK signaling, either of the adaptor proteins growth factor receptor-bound protein 2 (GRB2) or Src Homology 2 Domain-Containing (SHC), can recruit son of sevenless (SOS) to the cell membrane. SOS activates small GTPases of the Ras-family which in turn activate the chain of phosphorylation events down the chain of serine-threonine kinase starting at Raf (MAPK kinase kinase), then MEK (MAPK kinase), and terminating at extracellular signal-regulated kinases (ERK or MAPK). ERK then translocates to the nucleus and activates several downstream transcription factors (McKay and Morrison, 2007; Schulze et al., 2005; Yarden and Sliwkowski, 2001).

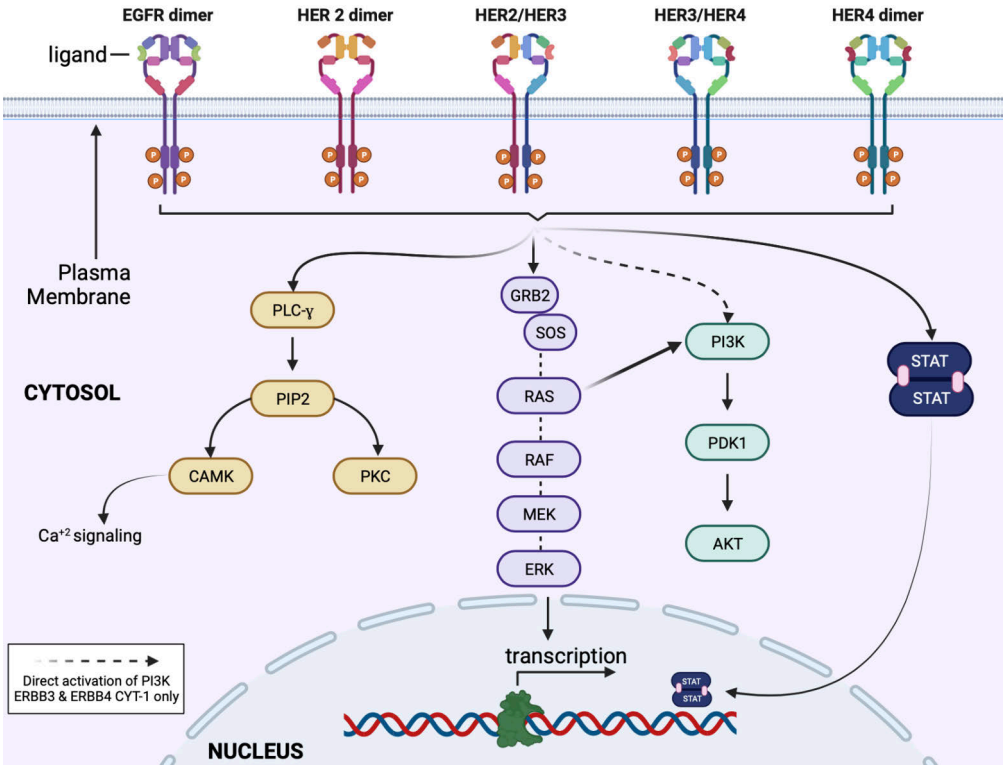


Figure 6. Signaling pathways activated by ERBB receptors. Dashed line shows direct recruitment of PI3K by the YXXM motif of ERBB3 and ERBB4. (Created in Biorender, adapted from (Yarden and Sliwkowski, 2001)).

The ERBB3 and ERBB4 receptors also harbor YxxM motif (where x can be any amino acid) in their C-terminus tail, and on tyrosine phosphorylation it can recruit the p85 regulatory subunit of PI3K via its SH2 domain and invoke the activation of PI3K-protein kinase B (AKT) pathway. The PI3K complex phosphorylates

phosphatidylinositol-4,5-phosphate (PIP2) to phosphatidylinositol-3,4,5-phosphate (PIP3) which recruits 3-phosphoinositide dependent kinase 1 (PDK1) and AKT to the plasma membrane. PDK1 activates AKT by phosphorylation which subsequently phosphorylates its downstream targets regulating cell survival, apoptosis, and cell migration (Fruman and Rommel, 2014; Schulze et al., 2005; Sepp-Lorenzino et al., 1996; Soltoff et al., 1994). In addition to direct recruitment, ERBB receptors can also invoke the PI3K-AKT pathway by direct activation of the p110 subunit via a GTP-bound RAS (Rodriguez-Viciana et al., 1994).

The phosphotyrosine residues on ERBBs can be recognized by phospholipase C gamma (Margolis et al., 1989; Peles et al., 1991; Vecchi et al., 1996) which catalyzes breakdown of PIP2 into two second messengers, diacylglycerol (DAG) and inositol triphosphate (IP3). DAG activates Protein Kinase C (PKC) whereas, IP3 stimulates the release of calcium (Ca^{2+}) from smooth endoplasmic reticulum. The PLC-ɣ pathway regulates cell migration, proliferation and survival (Owusu Obeng et al., 2020; Yang et al., 2013).

The ERBB receptors can activate STAT transcription factors by activating Janus Kinases (JAK), which then phosphorylate the STATs proteins enabling them to dimerize via their SH2 domains. Among the ERBB family, EGFR activates STAT1, STAT3, STAT5a and STAT5b, while ERBB4 activates STAT5a and STAT5b (Olayioye et al., 1999; Schulze et al., 2005). Phosphorylated dimeric STATs translocates to the cell nucleus activates transcription of genes mediating cell proliferation, apoptosis and differentiation (Quesnelle et al., 2007; Rawlings et al., 2004).

2.5.5 Alternative splicing of ERBB4

A peculiar feature about *ERBB4* is that under normal physiological conditions, it is the only ERBB gene to be processed by alternative splicing producing four distinct isoforms. Two isoforms are created by inclusion of either exon 16 (23 aa) or 15 (13 aa), creating Juxtamembrame-a (JM-a) or JM-b isoforms (Elenius et al., 1997a). The differences in the protein sequences enable only the JM-a isoform to undergo regulated intramembrane proteolysis (RIP), allowing the ectodomain to be shed and the release of a soluble Intracellular domain (ICD). The process of RIP involves two sequential proteolytic cleavage events performed by tumor necrosis factor-α-converting enzyme (TACE) followed by ɣ-secretase complex (Elenius et al., 1997a; Rio et al., 2000). The cleaved and solubilized ICD can translocate to the nucleus, owing to the presence of a nuclear localization sequence, and has been shown to regulate transcription (Komuro et al., 2003; Ni et al., 2001; Paatero et al., 2012; Sardi et al., 2006; Sundvall et al., 2010; Williams et al., 2004).

The two cytoplasmic (CYT) isoforms are created by either inclusion (CYT-1) or exclusion (CYT-2) of exon 26 (16 aa) which encodes the CYT-1 specific region (Elenius et al., 1999). The CYT-1 amino acid sequence contains a YXXM motif,

which constitutes a direct binding site for PI3K (Elenius et al., 1999), and includes an additional PPxY motif (where x can be any amino acid), serving as a binding site for WW domain-containing ubiquitin ligases (Sundvall et al., 2008b). Thus, through the combination of two JM- and two Cyt- sequences, these four ERBB4 isoforms are created, namely, ERBB4 JM-a CYT-1, JM-b CYT-1, JM-a CYT-2, or JM-b CYT-2 (Elenius et al., 1999, 1997a; Junttila et al., 2000).

These four isoforms also exhibit differences in their expression and function in various tissues. The JM-a isoforms are abundant in epithelial tissues like kidney, salivary gland and testis, while skeletal muscle and heart tissue predominantly expresses the JM-b isoforms (Junttila et al., 2005; Veikkolainen et al., 2011). In the brain tissue, the two JM- isoforms are expressed in different brain regions (Elenius et al., 1997a; Junttila et al., 2005; Veikkolainen et al., 2011). The CYT isoforms on the other hand do not exhibit preferential expression between tissues, and have been reported to be co-expressed, albeit at different levels (Junttila et al., 2005; Veikkolainen et al., 2011).

2.6 Aberrant ERBB signaling in human cancers

The initial reports of nucleotide sequence homology between ERBBs and oncogenes i.e., *EGFR* and avian *v-erbB* (Downward et al., 1984), and *ERBB2* and murine *neu* (Schechter et al., 1984) suggested an involvement in tumorigenesis. Over the years, scores of investigations have discovered and characterized the involvement of the ERBB RTKs and their ligands in the development and maintenance of human cancers of epithelial origin. Aberrant ERBB signaling manifests by means of a multitude of mechanisms, most common being overactivation of the pathway due to receptor and/or ligand overexpression, activating mutations, and gene amplification events. In addition to being driving alterations, these aberrant changes in ERBB signaling network serve as therapeutic vulnerabilities and predictive biomarkers in several cancer types (Hynes, 2016; Kiavue et al., 2020; Lucas et al., 2022; Segers et al., 2020; Uribe et al., 2021; Yarden and Pines, 2012).

2.6.1 EGFR in cancer

A well-studied case of aberrant *EGFR* signaling is of the non-small cell lung cancer (NSCLC) patients (Figure 7) harboring activating mutations in the intra-cellular domain of the EGFR RTK. The most-well studied of these are the exon19 deletions (Δex19) and the EGFR Leu858Arg (L858R) missense substitution. These mutations were initially discovered to be predictive of clinical response to EGFR-targeted therapy in NSCLC and were later shown to constitutively activate the EGFR tyrosine kinase (Lynch et al., 2004; Paez et al., 2004; Pao et al., 2004). Over time, several other EGFR mutations in the exon 18 – exon 21 have been proven to be driving alterations

in NSCLC (Beau-Faller et al., 2014; Kobayashi et al., 2015; Kobayashi and Mitsudomi, 2016). These mutations often produce the same qualitative phenotype of activating MAPK and PI3K-AKT pathways, but are also, predictably, different in many ways, including their sensitivity to EGFR-targeted Tyrosine Kinase Inhibitors (TKIs) (Kobayashi and Mitsudomi, 2016). Several TKIs have been approved by the United States Food and Drug Administration (FDA) and European Medicines Agency (EMA) for treating patients with NSCLC harboring the predictive mutations (see the details in section 2.6.4). In gliomas, EGFR amplifications are common (Figure 7), and are often accompanied by chromosomal arrangements that result in generation of in-frame deletion variants of EGFR, EGFRvIII (lacks exons 2 – 7) being the most frequently observed (An et al., 2018; Libermann et al., 1985). EGFRvIII, which lacks the key residues (aa 6 – 273), is a driving alteration in over 30% of gliomas. This region is essential for ligand binding and causes EGFRvIII to adopt a constitutively active confirmation (An et al., 2018). EGFR is also a therapeutic target in metastatic colorectal cancer (CRC) and head-and-neck squamous cell carcinoma (HNSCC) where anti-EGFR therapeutic monoclonal antibodies have shown clinical efficacy (Bonner et al., 2006; Cunningham et al., 2004).

2.6.2 ERBB2 and ERBB3 in cancer

Amplification of *ERBB2* leading to its overexpression was initially reported in a subset of breast carcinomas (Figure 7). These findings lead to the development of anti-ERBB2 monoclonal antibody (mAb) trastuzumab. Since then, these alterations have also been detected in gastric, colorectal, esophageal, endometrial and some ovarian cancer samples. Furthermore, mutations in the ERBB2 kinase have also been identified in a small number of NSCLC patients as well as lung squamous cell carcinoma (Figure 7). ERBB2-targeting TKIs and mAbs are approved for clinical use in breast cancer patients with ERBB2 amplification.

ERBB3 has recently been under the spotlight for its role in cancer and is known to function as a favored co-receptor to oncogenic ERBB2 (Kiavue et al., 2020; Sithanandam and Anderson, 2008). It has been demonstrated in several models that ERBB3 facilitates neuregulin mediated resistance (Hegde et al., 2013; Miyake et al., 2020; T. R. Wilson et al., 2012; Kimio Yonesaka et al., 2015; K Yonesaka et al., 2015). ERBB3 over expression and mutations have also been reported in several types of cancers (Figure 7) (Mota et al., 2017), but, despite preclinical data showing their oncogenic properties, attempts of inhibiting ERBB3 mutations with neratinib, a pan-ERBB tyrosine kinase inhibitor, was unsuccessful in patients with ERBB3 mutations (Hyman et al., 2018). The rationale for using a pan-ERBB inhibitor against the ERBB3 pseudokinase, was inhibiting the other ERBB receptors that heterodimerize with an active ERBB3. However, unaffected by these findings, ERBB3 remains a target for investigation with active clinical and pre-clinical development (Kiavue et

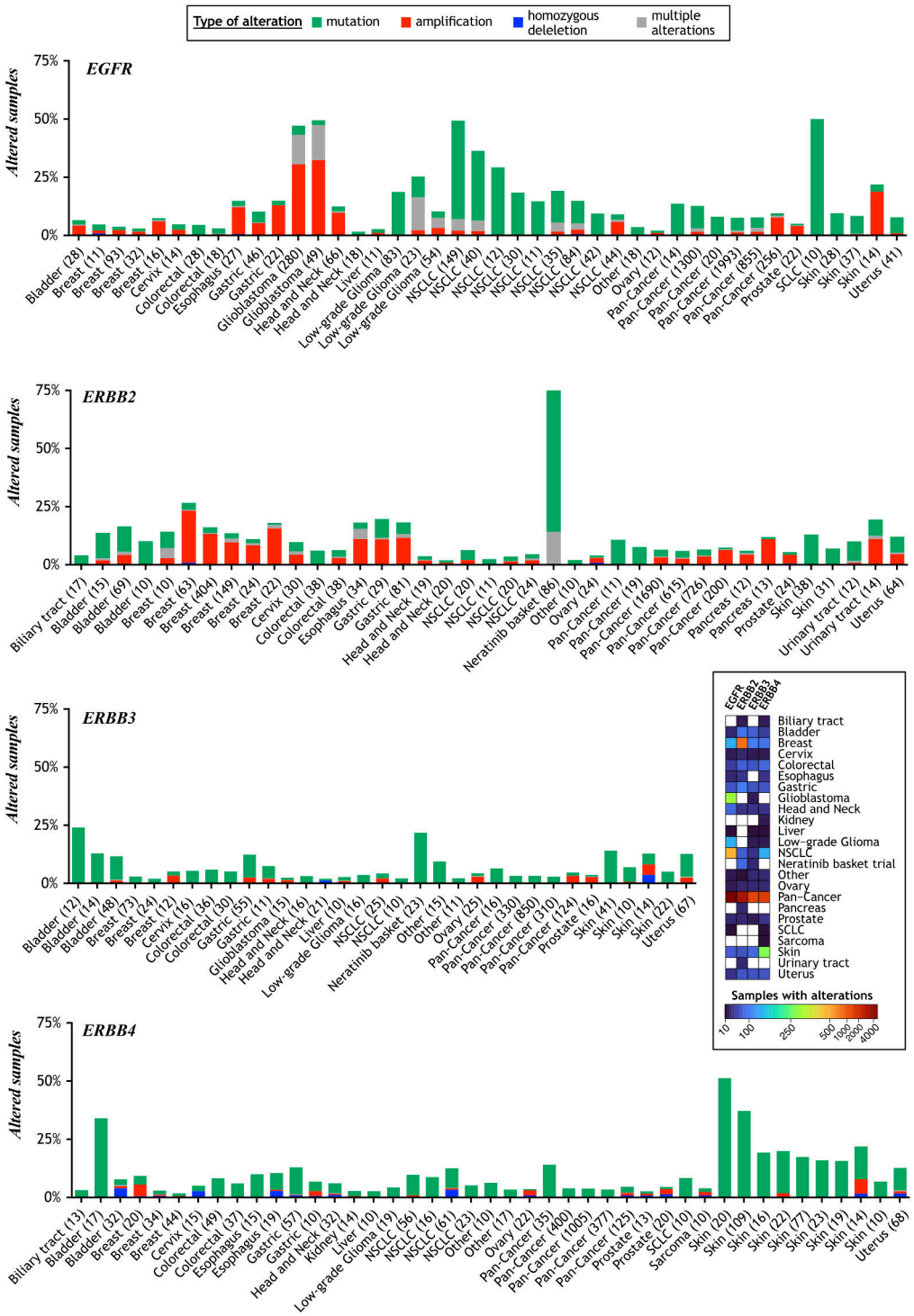
al., 2020) with anti-ERBB3 (α-ERBB3) patritumab being granted a breakthrough drug designation in December, 2021 (fda.gov).

2.6.3 ERBB4 in cancer

The role of ERBB4 in cancer is ambiguous with reports demonstrating its role, in different cellular contexts and experimental settings, as a typical oncogene (Kurppa et al., 2016; Nakamura et al., 2016; Prickett et al., 2009; Tvorogov et al., 2009), or akin to a tumor suppressor (Mill et al., 2011; Penington et al., 2002; Vidal et al., 2007; Williams et al., 2003). ERBB4 is expressed in several types of human cancers (Hollmén and Elenius, 2010), and alterations in ERBB4 are common in melanoma and bladder cancer (Figure 7). ERBB4 expression associates with favorable prognosis in well differentiated, estrogen- and progesterone receptor-positive and ERBB2 negative breast cancer (Bacus et al., 1996; Kew et al., 2000; Knowlden et al., 1998; Sundvall et al., 2008a). On the other hand, activation of ERBB4 in glioblastoma patients, and increased expression of ERBB4 in triple negative breast cancer patients associates with poor prognosis (Donoghue et al., 2018; Kim et al., 2016). ERBB4 expression has also been shown to be elevated in some colorectal cancer (CRC) tumors, and overexpression of ERBB4 was shown to enhance growth of CRC xenografts (Williams et al., 2015). ERBB4 has also been linked to chemotherapeutic resistance in NSCLC, sarcomas and in ovarian cancer (Merimsky et al., 2002, 2001; Saglam et al., 2017). Inhibition of ERBB4 signaling, by blocking its ligands, has been shown to improve response to chemotherapy in *in vitro* models of NSCLC (Hegde et al., 2013). Furthermore, ERBB4 has been shown to mediate acquired resistance to ERBB2 inhibition in breast cancer cells and in *MMTV-Neu* transgenic mice (Canfield et al., 2015).

There are a limited number of studies that take the ERBB4 splice variants (discussed in 2.5.5) into consideration. For instance, in medulloblastoma and serous ovarian cancer, the JM-a CYT-1 isoform was associated with unfavorable survival (Ferretti et al., 2006; Paatero et al., 2013). While the JM-a CYT-2 isoform was the dominant form expressed in glioblastoma patients but did not have prognostic significance (Donoghue et al., 2018). Taken together, it is evident that the function of ERBB4 in cancer is highly nuanced and context-sensitive (Segers et al., 2020).

► **Figure 7.** Prevalence of cancer-associated somatic alterations among the ERBB family of receptors (four panels), in various cancer datasets. The y-axis indicates the percentage of samples with alterations; the x-axis lists the primary histology of the dataset and the total number of altered cases within parentheses. Samples with amplification and deletion events are shown in red and blue respectively. Samples with small mutations are shown in green, and samples with more than one type of alteration are shown in grey. NSCLC = non-small cell lung cancer. Heatmap (inset) shows total number of samples with alterations in indicated ERBBs across the indicated primary sites. White color indicates absence of reported genetic alterations. (Source: original analysis and figure; data from cBioportal / 03.05.2022).


Type of alteration
mutation
amplification
homozygous deleletion
multiple alterations
EGFR
ERBB2
ERBB3
ERBB4
Altered samples
75%
50%
25%
0%
Biliary tract
Bladder
Breast
Cervix
Colorectal
Esophagus
Gastric
Glioblastoma
Head and Neck
Kidney
Liver
Low-grade Glioma
NSCLC
Neratinib basket trial
Other
Ovary
Pan-Cancer
Pancreas
Prostate
SCLC
Sarcoma
Skin
Urinary tract
Uterus
Samples with alterations
1.0
100
250
500
1000
2000
4000

2.6.4 ERBB inhibitors

Imatinib was the first ever small-molecule tyrosine kinase inhibitor approved by the FDA for use in patients with Philadelphia chromosome-positive leukemia (Kantarjian et al., 2002; Sawyers, 2002). It was soon followed by the approval of the first EGFR-targeted TKI, gefitinib in 2003 and then erlotinib in 2004 (fda.gov) (Figure 8). Gefitinib was withdrawn from the market in 2005, following a of lack of statistically significant benefit in survival compared to best supportive care (Thatcher et al., 2005). Around the same time, on-target sensitizing mutations were identified in EGFR, which, in addition to being oncogenic drivers, also conferred sensitivity to these TKIs (Lynch et al., 2004; Paez et al., 2004). Subsequent phase 3 trials with these inhibitors employing patient selection based on EGFR mutation status, demonstrated significantly longer progression-free survival in patients harboring activating EGFR mutations in exon 21 (L858R) or exon 19 deletions (Mok et al., 2009; Rosell et al., 2012). The clinical indications for the two inhibitors were eventually revised to incorporate this information. Lapatinib, a dual EGFR and ERBB2 inhibitor, was approved for combination therapy in patients with advanced breast cancer (Geyer et al., 2006). These molecules were reversible inhibitors competing for the ATP-binding pocket in the ERBB catalytic site.

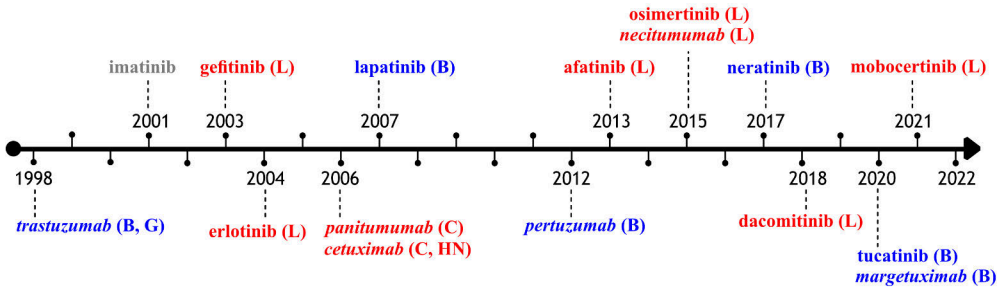


Figure 8. Timeline of FDA approval for indicated ERBB tyrosine kinase inhibitors approved for EGFR (red) and ERBB2 (blue). The monoclonal antibodies are italicized. Pan-ERBB TKIs are shown in bold. The clinically relevant cancer type is indicated in parentheses where B = breast cancer, C = colorectal cancer, G = gastric cancer, HN= head and neck cancer and L = non-small cell lung cancer. Imatinib (grey) is shown to anchor the timeline in the context of tyrosine kinase inhibitors in clinical oncology. (Source: original diagram, data from fda.gov).

The early reports of acquired resistance to the 1st gen inhibitors, which happened primarily by the acquisition and emergence of EGFR T790M gatekeeper-mutation, (Kobayashi et al., 2005), promoted the development of several covalent (irreversible) ATP-competitive inhibitors like the FDA-approved afatinib, dacomitinib, and neratinib. These 2nd generation TKIs covalently attach to the EGFR Cys 797 (and homologous residues in other ERBBs) within the catalytic site of the

kinase domain (Li et al., 2008; Rabindran et al., 2004; Solca et al., 2012). These inhibitors, although designed for EGFR, exhibit a potent pan-ERBB inhibition (Davis et al., 2011) and are thus referred as pan-ERBB inhibitors. However, despite pre-clinical evidence, these inhibitors were ineffective in the clinic for treating patients developing resistance due to EGFR T790M, largely due to the compounds possessing a high affinity for wild-type EGFR (Solca et al., 2012) and the requirement of higher concentrations of the TKIs to effectively block EGFR T790M (Kim et al., 2012). These findings lead to the development of the third generation of EGFR TKIs, like the mutant-selective osimertinib which is effective in blocking EGFR L858R/T790M and exon19del/T790M while sparing the wild-type EGFR (Cross et al., 2014). However, as covalent binding to the EGFR Cys 797 residue is a key part of its efficacy, resistance eventually emerges on therapy progression with the acquisition of mutations in this particular residue, abrogating the TKI activity (Thress et al., 2015).

Decades ago, during the infancy of targeted oncology, efforts were being made to block ERBB signaling using engineered antibodies which were good at blocking ligand binding (for EGFR), preventing dimerization, as well as in triggering an immune response (Hudziak et al., 1989; Saleh et al., 1999, p. 225). Trastuzumab (α-ERBB2) (Cobleigh et al., 1999; Slamon et al., 2001) received FDA-approval in 1998 and was followed by EGFR-targeting monoclonal antibodies by the FDA in 2006, namely, panitumumab (anti-EGFR or α-EGFR) (Van Cutsem et al., 2007), cetuximab (α-EGFR) (Bonner et al., 2006; Cunningham et al., 2004). These mAbs were followed by the α-ERBB2 pertuzumab and margetuximab, and the α-EGFR necitumumab being approved for clinical use over the years (Figure 8).

In the recent years, the focus has been to block mutations against which the FDA-approved drugs are ineffective, such as the insertions in exon 20 of EGFR and ERBB2. Mobocertinib, was granted the FDA approval in 2021 for NSCLC harboring EGFR exon 20 insertions, and currently poziotinib is under consideration by the FDA after promising phase II results in NSCLC patients harboring ERBB2 exon 20 insertions (Elamin et al., 2022; Le et al., 2022). Another frontier has been exploring the efficacy of the ERBB TKIs in genomically-selected “basket” trials, where the treatment is guided by genomic biomarkers in cancers of different histology (Hyman et al., 2018; Lopez-Chavez et al., 2015). Single-agent EGFR and ERBB2 blockade has been effective in the clinic, however, addressing the ambivalence in the validity of ERBB3 and ERBB4 as clinical drug targets warrants more pre-clinical characterization and subsequent exploratory clinical studies. The ERBB TKIs are an active domain of research and development in oncology with new molecules and several combination therapies undergoing clinical investigation (clinicaltrials.gov).

2.7 Databases of cancer-associated somatic mutations

Large-scale projects aiming at comprehensive genomic characterization of cancer tissues e.g., The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium (ICGC), and the targeted sequencing efforts, such as the cohorts analyzed by the MSK-IMPACT protocol (Cheng et al., 2015) at the Memorial Sloan Kettering Cancer Center generate large amounts of data. The art of efficient dispersal of information relies on the sequential process of data aggregation, integration, and presentation. The goal is to facilitate data exploration and analysis by the end-user, which in the case of data generated by cancer genomics, are scientists and clinicians who often lack an expertise in scripting. To resolve this, several efforts have been made with the most prominent and internationally acclaimed ones being the cBioPortal for Cancer Genomics (https://cbioportal.org) (Cerami et al., 2012; Gao et al., 2013), the Catalogue Of Somatic Mutations In Cancer (COSMIC) database (Forbes et al., 2017; Tate et al., 2019), the ICGC data portal (Zhang et al., 2011) and the data portal from the American Association for Cancer Research for the project Genomics, Evidence, Neoplasia, Information, Exchange creating the AACR-GENIE data portal (The AACR Project GENIE Consortium, 2017). These data portals provide rich access to the relevant information and often provide a set of tools for performing some data analysis (Cerami et al., 2012; Gao et al., 2013).

3 Aims of the Study

Comprehensive genetic characterization of cancer tissues was enabled with the advent of next-generation sequencing technologies. This has opened up a gateway for peering into the cancer genomes, and in the process, improved the understanding about the molecular players and the driver events involved in the life cycle of a cancer. In addition to generating a small list of driver mutations, the large-scale sequencing efforts generated a long "tail" consisting of "variants of unknown significance" (VUS), which are cancer-associated somatic mutations with uncharacterized functional significance.

These VUS have largely been overlooked amid the gold rush of finding hotspot activating mutations and producing pharmaceuticals for blocking them. However, studies revealed the occurrence of several mutations, that even though are infrequent in the population, are in fact, activating mutations. The problem of identification of driver mutations is further complicated by the fact that most of the cancer-associated somatic mutations are predicted to be non-functional "passengers". Together, these facts motivated the development of a screening workflow that is based on creating an environment that allows an unbiased competition among all the coding variants of a transgene.

On the other hand, the large volumes of data have created a different problem which is analogous to finding a needle in the haystack. Contemporary databases enabling access to the abundance of cancer genomics data suffer from being slow, consume significant bandwidth, and require multiple steps to get to key pieces of information. Discussions in molecular tumor boards, for instance, do not benefit from comprehensive data about one-off cases; they require well-summarized information to steer the decision-making process. Furthermore, inclusion of substantial proportions of data from targeted sequencing panels skews the apparent distribution of mutation frequencies away from reality. Therefore, there was a need for a database that would evade this inherent selection bias of targeted screens, and to enable fast and easy access to information about realistic mutation frequencies in human cancers.

The specific aims of this thesis were as follows:

1. To establish an unbiased high-throughput functional genetics screen for activating mutations.

2. To validate the workflow using a library of Epidermal Growth Factor Receptor (EGFR) mutations.
3. To utilize the screening workflow and study libraries of ERBB4 mutations, a receptor tyrosine kinase in the EGFR family.
4. To develop and deploy a fast and easy-to-use database that lists recurrent cancer-associated somatic mutations.

4 Materials and Methods

4.1 Expression Plasmids

The cDNA pieces encoding wild-type human EGFR, ERBB4 JM-a CYT-1, and ERBB4 JM-a CYT-2 were cloned individually into pDONR221 (Tol2Kit (Kwan et al., 2007)) with the Gateway recombination cloning strategy by using the BP clonase II mix (Invitrogen, Cat #11789-020) based on manufacturer's instructions. These plasmids were used as respective templates; *EGFR WT* (Greulich et al., 2005) (Addgene plasmid #11011; http://n2t.net/addgene:11011; RRID:Addgene_11011), *pcDNA3.1neo(-)-ERBB4JM-aCYT-1* (Maatta et al., 2006) and *pcDNA3.1neo(-)-ERBB4JM-aCYT-2* (Maatta et al., 2006). The *attB*-flanked PCR-amplicons required for the BP recombination reaction, were generated by a PCR with Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Cat #F-534L) and using primers described in (Table 1) and following the manufacturer's protocol. The PCR master mix was supplemented with 1.5% DMSO, and the oligos (1 µM final concentration) were annealed for 20 s at 72°C for *EGFR,* and 68°C for *ERBB4*.

Table 1. List of primers used for cloning wild-type inserts from individual plasmids into the pDONR221 using Gateway BP cloning.

Name	Primer sequence	Used in
attB1-EGFR	ggggacaagtttgtacaaaaaagcaggcttcaccatgcgaccctccgggacgg	I
attB2-EGFR	ggggaccactttgtacaagaaagctgggtttcatgctccaataaattcactgctttgtg	I
attB1-ERBB4	ggggacaagtttgtacaaaaaagcaggcttcaccatgcgaccggctacaggact	II
attB2-ERBB4	ggggaccactttgtacaagaaagctgggttttacaccacagtattccggtgtc	II

These Gateway cloning reactions created *pDONR221-EGFR*, *pDONR221-ERBB4JM-aCYT-1,* and *pDONR221-ERBB4JM-aCYT-2* and they were used to create individual retroviral mammalian expression plasmids with a Gateway LR recombination reaction with LR clonase II mix (Invitrogen, Cat #11791-020) and *pBABEpuro-gateway* (Greulich et al., 2012) (Addgene plasmid #51070; http://n2t.net/addgene:51070; RRID:Addgene_51070). Additionally, *pDONR221-eGFP* (Yang et al., 2011) (Addgene plasmid #25899; http://n2t.net/addgene:25899; RRID:Addgene_25899), was used to create *pBABEpuro-gateway-eGFP*, to use a control for retroviral transduction with a LR reaction.

The expression plasmids encoding the indicated *EGFR* and *ERBB4* mutations were generated by using site-directed whole-plasmid mutagenesis using the *pDONR221* constructs described above. The mutant cDNAs were subsequently cloned to *pBABEpuro-gateway* to create retroviral mammalian expression plasmids. This two-step site-directed mutagenesis protocol was designed to avoid the unintended consequences of an inadvertent mutation getting incorporated in the backbone of our large (> 8 kb) expression plasmids. Table 2 below lists the respective primers, templates and expression plasmids used to create the indicated EGFR and ERBB4 mutations.

Table 2. Plasmids used in this study. The source plasmid *pDONR221* was used as a template for whole-plasmid site-directed mutagenesis and the destination plasmid was the mammalian expression plasmid where the mutant cDNA was moved with a Gateway LR reaction. The *pcDNA3.1* and *pMSCV-PGK-Puro-IRES-GFP* and plasmid derivatives were used directly for transfection.

Isoform	Mutation	FW Primer	REV Primer	Source	Used in	Reference (source plasmid)
Canonical	EGFR N604I	aacatcaccctg gtctgg	ttctcccatgact cctgc	pDONR221-EGFR	I	(Chakroborty et al., 2019)
	EGFR A702V	caagttctcttga ggatctt	gttgggagcttct ccact	pDONR221-EGFR	I	
	EGFR T790M	atcatgcagctc atgccc	gagttgcacggt ggaggtg	pDONR221-EGFR	I	
	EGFR C797S	ggcagcctcctg gactatg	gaagggcatga gctgcgt	pDONR221-EGFR	I	
	EGFR L858R	gggcgggcca aactgct	aaaatctgtgat cttgacatgctgc	pDONR221-EGFR	I	
	EGFR D956V	gtctatcatccag cacttgacc	gtctatcatccag cacttgacc	pDONR221-EGFR	I	
	EGFR P1170H	aaccatgactac cagcaggac	gtccaggctaat ttggtggc	pDONR221-EGFR	I	
JM-a CYT-2	ERBB4 Y52C	caagtgctatga aaactgtgaggt	cgcaaggctcg gtactg	pDONR221-ERBB4JM-aCYT-2	II	(Chakroborty et al., 2022)
	ERBB4 R124K	tacaaaaaaga tggaaactt	gtttaaaaatatt gccaag	pDONR221-ERBB4JM-aCYT-2	II	
	ERBB4 R687K	cttgaaaagatt cttggaaac	gctcttttctttttga tgc	pDONR221-ERBB4JM-aCYT-2	II	
	ERBB4 E715K	aaaaaaactga gctgaagag	caaaatacgaa gttgagct	pDONR221-ERBB4JM-aCYT-2	II	
	ERBB4 G741R	ctgaaagagaa actgtgaag	gtacccaaatac ctttataaac	pDONR221-ERBB4JM-aCYT-2	II	
	ERBB4 G802D	catgactgcctgt tggag	gggcataagttg agtaaccag	pDONR221-ERBB4JM-aCYT-2	II	
	ERBB4 M993I	gtataaagcttc ccagtcc	gatcatcaccct gaataactag	pDONR221-ERBB4JM-aCYT-2	II	
	ERBB4 V1172F	ctttttttctcgga gaaaaa	ggttctcctccac tggat	pDONR221-ERBB4JM-aCYT-2	II	
	ERBB4 G1217R	cttgcgaaaag ctgagtacc	gtgttggcaaag gtgttg	pDONR221-ERBB4JM-aCYT-2	II	
	ERBB4 K1218N	ggaaacgctga gtacctgaag	caaggtgttggc aaaggtg	pDONR221-ERBB4JM-aCYT-2	II	

Isoform	Mutation	FW Primer	REV Primer	Source	Used in	Reference (source plasmid)
JM-a CYT-1	ERBB4 R124K	Commercially ordered from Genscript (genscript.com) to be generated and inserted into *pMSCV-PGK-Puro-IRES-GFP* backbone.			II	(Zuber et al., 2011)
	ERBB4 R687K				II	
	ERBB4 E715K				II	
	ERBB4 G741R				II	
	ERBB4 K935I				II	
JM-a CYT-2	ERBB4 E715K				II	
	ERBB4 E715K	aaaaaaactga gctgaagag	caaaatacgaa gttgagct	pcDNA3.1ERBB4JM-aCYT-2-V954R-HA	II	(Kurppa et al., 2016)
Canonical	EGFR K721R	gctatcaggga attaagag	gacgggaatttt aactttc	pcDNA3.1EGFR-HA	II	(Merilahti et al., 2017)
	ERBB2 K753M	atcatggtgttga ggga	ggccactggaa ttttca	pcDNA3.1ERBB2-HA	II	

4.2 Generating expression libraries for random mutants of *EGFR* and *ERBB4*

The expression library for human *EGFR*, *ERBB4 JM-a CYT-1*, and *ERBB4 JM-a CYT-2* mutants was generated by an error-prone PCR (ep-PCR) by using Genemorph II (Agilent Technologies, Cat #200550), which is a cocktail of two DNA-polymerases with an intrinsic error rate higher than that of DNA polymerases used in conventional PCRs e.g., *Taq* and *Pfu* DNA-polymerase. The randomly mutated cDNA fragments for *EGFR* or the two isoforms of *ERBB4* were created by using the primer pair 5′-ttgatgcctggcagttcccta-3′ (binds 78 bp upstream of *attL1* in the plasmid), and 5′-atcttgtgcaatgtaacatcagagatt-3′ (binds 80 bp downstream of *attL2* in the plasmid), and either *pDONR221-EGFR*, *pDONR221-ERBB4JM-aCYT-1,* or *pDONR221-ERBB4JM-aCYT-2* as the template (plasmid containing 4 µg of the cDNA insert) and 10 cycles of ep-PCR.

The PCR amplicons (EGFR: 4,043 bp, ERBB4 JM-a CYT1: 4,337 bp, and ERBB4 JM-a CYT-2: 4,289 bp) were run on 1% agarose gel (at 70 V), the corresponding band was excised using a scalpel blade in a UV transilluminator. The DNA was purified with Nucleospin Gel and PCR Clean-up kit (Macherey Nagel, Cat #740609.250) and quantified with NanoDrop (Thermo Fisher, Cat #ND-ONE-W). The purified amplicons were cloned directly into *pBABEpuro-gateway* with a LR reaction (as described above) to create expression libraries comprising of random EGFR and ERBB4 mutants. The mutant fragments were not cloned into an "entry" clone (e.g., a *pENTR* or *pDONR* derivative) to prevent losses in the representation of the mutant ORFs incurred during the sub-cloning process. The product of LR reaction was transformed into chemically competent *ccdb*-sensitive *E. coli* after inactivation of the clonase with Proteinase K (as indicated by the manufacturer).

4.3 Cell culture and generation of stable lines

Phoenix-Ampho cells (Swift et al., 2001) (modified HEK-293 cells for producing amphotropic viruses, a gift from Garry Nolan), Ba/F3 cells (murine lymphoid cells, a gift from David M. Weinstock) and BEAS-2B cells (human bronchial epithelial cells, ATCC CRL-9609) were cultured in RPMI-1640 (Lonza, Cat #12-167 or Gibco Cat #21875-091), supplemented with 10% Fetal Calf Serum (FCS) (Biowest, Cat #S1810), 1-2 mM L-glutamine (Lonza), and 50 U/mL penicillin-streptomycin (Lonza). The growth media of Ba/F3 cells was supplemented with conditioned media of the WEHI-3B cells (source of Interleukin-3 (IL-3), final concentration 5%), unless indicated otherwise. A549 (human lung adenocarcinoma cells harboring KRAS G12S mutation), NCI-H661 (human lung adenocarcinoma cells), NIH-3T3 cells (mouse fibroblasts), Platinum-E cells (Cell Biolabs, Cat #RV-101), and COS-7 cells (Green monkey kidney cells) were cultured in Dulbecco's modified Eagle's medium (Lonza) supplemented with 10% FCS, 50 U/mL penicillin-streptomycin, 2 mM L-glutamine.

To generate Ba/F3 expressing EGFR and ERBB4 variants and NIH-3T3 cells expressing ERBB4 variants, the Phoenix-Ampho packaging cells (Swift et al., 2001) were transfected with retroviral plasmids (*pBABEpuro-gateway*) encoding the ERBB variants or enhanced Green Fluorescent Protein (eGFP) using FuGENE 6 transfection reagent (Promega, Cat #E2692) based on manufacturer's instruction. The Phoenix cells produced viruses in RPMI media, and the supernatant was harvested 24 and 48 h after transfection. This supernatant contained retroviral particles and was incubated on 5 x 10^5 Ba/F3 or NIH-3T3 cells for 6 h for 2 consecutive days. On both days, the media (RPMI-1640 for Ba/F3 and DMEM for NIH-3T3) was diluted 1:2 to reduce viral load, after 6 h of incubation with the viruses. Cell pools with stable expression of the inserts were selected with Puromycin (Gibco, Cat #A11138-03) (final concentration for Ba/F3: 2 µg/mL, and NIH-3T3: 6 µg/mL) and were further maintained in culture in media formulations containing Puromycin (final concentration for Ba/F3: 1 µg/mL, and NIH-3T3: 3 µg/mL).

Table 3. Cell lines used in this study.

Cell Line	Type	Species	Growth media	Used in
A549	Lung adenocarcinoma	Human	DMEM	I
Ba/F3	Lymphoid (pro-B cells)	Mouse	RPMI-1640 ± 5% WEHI-3B CM	I, II
BEAS-2B	Bronchial epithelium	Human	RPMI	II
COS-7	Kidney fibroblast-like	African green monkey	DMEM	II
NCI-H661	Lung adenocarcinoma	Human	DMEM	I
NIH-3T3	Fibroblast	Mouse	DMEM	II
Phoenix-Ampho	Embryonic kidney	Human	RPMI-1640	I, II
Platinum-E	Embryonic kidney	Human	DMEM	II

Before moving the cells out of the Biosafety Level 2 cell culture facility, a portion of the cells were lysed and screened to confirm that they are devoid of retroviral ribonucleic acid (RNA). Briefly, the RNA was isolated using TRIsure (Bioline, Cat #BIO-38033) by following manufacturer's protocol, up to 1 µg of RNA was reverse transcribed to cDNA using the SensiFAST cDNA Synthesis Kit (Bioline, Cat #BIO-65054) and following manufacturer's instructions. The cDNA was used as a template in PCR reactions with primers (listed below) specific to the Group-specific antigen (*gag*), Polymerase (*pol)* retroviral genes, and murine *Actin* (as control). In isolated cases when viral cDNA was detected in a cell line, the cells were cultured for 1-3 passages in the BSL-2 facility, tested (using the methodology described above), and the transduced cells were moved out of the BSL-2 facility when they were devoid of retroviral RNA.

- ACTIN_FW 5'-atc tgg cac cac acc ttc tac aat-3'
- ACTIN_REV 5'-ccg tca ccg gag tcc atc a-3'
- GAG_FW 5'-cgc cta cgt ggg aga cgg ga-3'
- GAG_REV 5'-ccg cgt ttt gga gac ccg ct-3'
- POL_FW 5'-tat atg ggg cac ccc cgc cc-3'
- POL_REV 5'-gga ccc aca ctg tgt cgc cg-3'

Platinum-E packaging cells were transfected with retroviral pMSCV-PGK-Puro-IRES-GFP vectors encoding ERBB4 variants or an empty vector using the X-tremeGENE 9 transfection reagent (Roche, Cat #XTG9-RO). The retroviral supernatants were harvested 48 h after transfection and incubated for 72 h on BEAS-2B cells. 1 µg/mL puromycin was used to select the transduced BEAS-2B cells and maintain the cells with a stable expression.

4.4 Conducting the *in vitro* screen for activating mutations

Batches of Phoenix-Ampho cells (Swift et al., 2001) were transfected with *EGFR*, *ERBB4 JM-a CYT-1*, *ERBB4 JM-a CYT-2* random mutation library, their wild-type counterparts and the vector control (*pBABEpuro-gateway eGFP*) to produce retroviruses as described above. The viral supernatant was incubated on Ba/F3 cells and stable cell populations were selected with Puromycin as described above.

The stable cell populations were cultured in media devoid of IL-3 for two weeks, in case of the EGFR iSCREAM study (i.e. article I). For carrying out the ERBB4 iSCREAM, the stable cell populations were cultured in IL-3-depleted media that was supplemented with 10 ng/mL Neuregulin-1β (NRG1). The cells were passaged pre-confluence to maintain their growth in log-phase. The genomic DNA was harvested from the surviving cells and 100 ng of it was used to extract the cDNA inserts with a PCR (30 cycles) using Phusion (Thermo Scientific, Cat #F-534L) and Velocity (Bioline, Cat #BIO-21099) high-fidelity DNA polymerases, and the primer pair 5'-

ggg gac aag ttt gta caa aaa agc agg ctt cac cat gcg acc ctc cgg gac gg-3' and 5'-ggg gac cac ttt gta caa gaa agc tgg gtt tca tgc tcc aat aaa ttc act gct ttg tg-3' for samples from the EGFR iSCREAM; and for samples of the ERBB4 iSCREAM, the primer pair: 5'-gaa cct cct ctt tcg acc cc-3' and 5'-aag agt tct tgc agc tcg gt-3'. At the same time, 5 ng of original respective plasmid cDNA libraries for *EGFR* and *ERBB4* were used to extract the library for sequencing. Amplicons were purified (Nucleospin Gel and PCR Clean-up kit, Macherey Nagel, Cat #740609.250), and sequencing libraries were prepared with Nextera XT DNA sample Preparation Kit (Illumina, Cat #FC-131-1024), and subsequently sequenced on Illumina MiSeq instrument with 150 bp paired-end sequencing chemistry.

4.5 Functional and analytical assays

4.5.1 Cell lysis and Western blotting

Western blotting was performed to determine the expression and phosphorylation of proteins in the cell lysates. To prepare samples for analysis with Western blotting, the Ba/F3 cells that grow in suspension culture were harvested by centrifugation for 5 min at 500 x g, and washed with PBS, while the adherent cells were washed with PBS. The cells were treated with lysis buffer (10 mM Tris- HCl, pH 7.4, 1% Triton X-100, 150 mM NaCl, 1 mM EDTA, 10 mM NaF), supplemented with protease and phosphatase inhibitors (2 mM phenylmethylsulfonyl fluoride, 10 mg/mL aprotinin, 10 mg/mL leupeptin, 10 mM $Na_4P_2O_7$ and 1 mM Na_3VO_4) or with Pierce Protease Inhibitor Mini Tablets (Thermo Fisher Scientific). Protein was recovered from the lysates by centrifugation at 16000 x g for 15 min and the protein concentration in the supernatant was measured using the Bradford protein assay (Bio-Rad). The proteins (in the supernatant) were denatured by incubation at 95°C for 5 min in Laemmli loading buffer. Samples were resolved by SDS-PAGE and transferred to nitrocellulose membranes. Membranes were blocked with whole milk, and subsequently incubated with primary antibodies (as indicated in original publications). The blots were imaged using the IRDye-conjugated (corresponding to 800 nm or 680 nm) secondary antibodies and the near-IR fluorescence on the Odyssey CLx imaging system (LI-COR). Alternatively, some blots were imaged using chemiluminescence (WesternBright ECL HRP, Advansta) by incubation with HRP-conjugated secondary antibody on the ImageQuant LAS-4000 imaging system (Fuji-film).

Table 4. Primary antibodies used in this study. WB = Western blotting, FC = Flow Cytometry. CST = Cell Signaling Technologies.

Antigen	Cat# / Clone	Vendor	Type	Application	Used in
Actin	A5441	Sigma	Mouse monoclonal	WB	I, II
EGFR	4267	CST	Rabbit monoclonal	WB, FC	I, II
ERBB2	MA5-14057	Thermo	Mouse polyclonal	WB	II
ERBB3	4754	CST	Rabbit monoclonal	WB	I, II
ERBB4	E200	Abcam	Rabbit monoclonal	WB	II
GFP	ab183734	Abcam	Rabbit monoclonal	WB	II
MET	8198	CST	Rabbit monoclonal	WB	I
phospho-ERBB4	4757	CST	Rabbit monoclonal	WB	II
Phospo-EGFR	2220	CST	Rabbit polyclonal	WB	I

4.5.2 Cell viability assay

Ba/F3 cells transduced with indicated transgenes (listed in original publications) were seeded at a density of 1 x 10^5 cells/mL in the presence or absence of 5% conditioned medium from WEHI-3B cells as a source of IL-3. In case of ERBB4, the cells were analyzed in a growth media formulation containing 10 ng/mL NRG1. Cell viability was assessed using the MTT assay (Promega) where a tetrazolium salt is metabolized into a formazan product by the mitochondrial dehydrogenases. After solubilization of the formazan product the absorbance can be measured at 570 nm to assess cell viability.

4.5.3 Dose-response analysis using tyrosine kinase inhibitors

Ba/F3 cells (20,000 cells / well) expressing EGFR or ERBB4 variants were seeded in 96-well plates. Ba/F3 cells expressing ERBB4 variants were also cultured in presence of 10 ng/mL NRG1 (indicated in relevant text and figure legends in the original publication). After incubation with a series of concentrations for the indicated inhibitors (as in the original publication) for 48 to 72 h, the cell viability was analyzed with MTT assay (described in section 4.5.2). Empty vector control cells were cultured in presence of 5% WEHI-3B conditioned medium.

Table 5. Inhibitors and growth factors used in this study. SCBT = Santa Cruz Biotechnology. TKI = Tyrosine kinase inhibitor. mAb = monoclonal antibody. rh = recombinant human.

Reagent	Type	Vendor	Used in
afatinib	pan-ERBB TKI	SCBT	I, II
Erbitux	cetuximab (EGFR mAb)	Merck	I
dacomitinib	pan-ERBB TKI	Selleck Chemicals	II
erlotinib	EGFR TKI	SCBT	I, II
ibrutinib	BTK TKI	Selleck Chemicals	II
lapatinib	EGFR, ERBB2-TKI	SCBT	II
neratinib	pan-ERBB TKI	SCBT	II
poziotinib	pan-ERBB TKI	Selleck Chemicals	II
rhNRG1-β1	ERBB4 ligand	R&D systems	II

4.5.4 Analysis of EGFR expression using flow cytometry

Ba/F3 cells expressing EGFR variants (indicated in the original publication) were washed with azide-free PBS and stained with eBioscience Fixable Viability Dye eFluor 780 (Thermo Fisher, Cat #65-0865) to stain dead cells (which were later removed from the analysis). Next, the cells were fixed with 4% paraformaldehyde and permeabilized with ice-cold methanol and were then incubated with anti-EGFR (1:100, Cell Signaling Technologies, Cat #4267) and subsequently with Alexa Fluor 488-conjugated anti-rabbit (1:200, Thermo Fisher, Cat #A-11034). An LSR Fortessa flow cytometer was used with BD FACSDiva Software (v. 8.0.1) to capture flow cytometry data which was analyzed using FlowJo software (v. 10.5.3).

4.5.5 Assessment of tumorigenicity of *ERBB4* mutations *in vivo*

To assess tumorigenicity of indicated ERBB4 variants *in vivo*, 5 x 10^6 murine Ba/F3 cells in 100 µL PBS + 5% FCS) were injected subcutaneously into the left and right flanks of 6-8 week old female NMRI nude mice (BomTac:NMRI-Foxn1^{nu}). Tumor growth was monitored thrice weekly by bilateral caliper measurements and the tumor volume (V) was calculated as $V = length \times width^2 \times \pi/6$ and the data were plotted using Graphpad Prism9 to show mean ± standard error of mean. The animal studies were approved by Austrian authorities to be conducted at Boehringer Ingelheim in Austria in accordance with EU legislation at an animal facility which is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) International. These experiments were performed by our collaborators at Boehringer Ingelheim in Vienna, Austria.

4.5.6 Analysis of mRNA expression using real-time RT-PCR

Real-time RT-PCR was used to analyze EGFR and ERBB4 mRNA expression in Ba/F3 cells and mouse tumors respectively. RNA was extracted using TRIsure (Bioline, Cat #BIO-38033) and cDNA was synthesized with the SensiFast cDNA synthesis kit (Bioline, Cat #BIO-65054). The analysis was carried out using TaqMan Universal Master Mix II (Applied Biosystems, Cat #4440040) with the following primers and probes:

- human EGFR forward, 5'-cca cct gtg cca tcc aaa ct-3' (Pharmacia)
- human EGFR reverse, 5'-ggc gat gga cgg gat ctt-3' (Pharmacia)
- human EGFR probe, 5'-FAM-cca ggt ctt gaa ggc tgt cca acg aat-TAMRA-3' (Eurogentech)
- mouse GAPDH, Universal ProbeLibrary Mouse GAPDH Gene Assay 5046211001 (Sigma)
- human ERBB4 CYT forward 5'-caa cat ccc acc tcc cat cta tac-3' (Pharmacia)
- human ERBB4 CYT reverse 5'-aca ctc ctt gtt cag cag caa a-3' (Pharmacia)
- human ERBB4 CYT-2 probe 5'-FAM- aat tga ctc gaa tag gaa cca gtt tgt ata ccg aga t-TAMRA-3' (Eurogentec)
- mouse β-actin forward 5'-cta agg cca acc gtg aaa ag-3' (Eurofins Genomics)
- mouse β-actin reverse 5'-acc aga ggc ata cag gga ca-3' (Eurofins Genomics)

The QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific) was used for thermal cycling the reactions as follows: 2 min at 50°C and 10 min at 95°C, followed by 40 cycles of 15 s at 95°C and 60 s at 60°C. Samples were analyzed in triplicates, and the standard deviation of the C_T values was <5% of the mean. *EGFR* and *ERBB4* mRNA expression was quantified using mouse *GAPDH* mRNA expression as a reference with the $2^{-\Delta CT}$ method.

4.5.7 Three-dimensional growth assay

For performing three-dimensional (3D) growth assays, the wells in 96-well plates were coated with 5 mg/mL poly(2-hydroxyethyl methacrylate) (Sigma-Aldrich) in 96% ethanol with a volume of 50 µL per well. The BEAS-2B cells stably expressing the individual GFP-linked ERBB4 variants or GFP alone (described in section 4.3) were plated in quintuplicates on the poly-HEMA-coated 96-well plates at a density of 1,000 cells per well. The cells were plated in 2% Growth Factor Reduced Matrigel (Corning) in the presence of 2% FCS and 50 ng/mL NRG1. After culturing the cells for seven days, the 3D growth was quantified as the difference in the measured fluorescence intensity between day seven and the day of plating.

4.5.8 ERBB4 transactivation assay

pcDNA3.1 constructs encoding hemagglutinin (HA)-tagged wild-type, kinase-dead or kinase dimerization interface mutant ERBB receptors (specific mutations listed in original publication (article # II) text and figure) were used to transiently transfect the COS-7 cells with Lipofectamine2000 (Invitrogen) in 6-well plates. The cells were lysed 24 h after transfection and were analyzed for total ERBB4 expression and basal ERBB4 phosphorylation by Western blotting(described above).

4.6 Bioinformatics analysis

4.6.1 Processing and analyzing next-generation sequencing data (iSCREAM)

Reads in the FASTQ file were trimmed of adapter (ctgtctcttatacacatct) sequences and low-quality bases using trimmomatic (version 0.36) (Bolger et al., 2014) and the parameters recommended for paired-end sequencing (at http://usadellab.org/cms/?page=trimmomatic). The trimmed reads were aligned to human reference genome (hg19) using BWA-MEM (v 0.7.13-r1126 and 0.7.15-r1140) (Li, 2013). The generated Sequence Alignment Map (SAM) files were converted to Binary Alignment Map (BAM) files, sorted, and indexed on the fly using samtools (version 1.3.1) (Li et al., 2009). Variants were called using samtools the parameter --max-depth was set to 300000 to ensure all the reads aligned to the reference genome are used in calculations by samtools mpileup. As the Nextera XT libraries were not generated using a strand-specific protocol, bam-readcount was used to identify potential sequencing artifacts by calculating strand bias (ratio of counts of forward reads to reverse reads aligning at a particular locus) for each variant. Variants with strand bias < 0.1 or > 10 were filtered out (i.e., variants having number of reads in either the forward (5' – 3') or the reverse direction than the other). The amino acid substitutions (if any) generated by the nucleotide substitutions were deduced by using ANNOVAR (Wang et al., 2010).

To discern EGFR and ERBB4 mutations that were enriched after IL-3 depletion, a fold change statistic was calculated as the ratio of the variant allele frequency (VAF) of the mutation at the final and initial time point (these varied between two iSCREAM experiments, and are indicated in relevant section in "Results" and in the corresponding figure legends). The VAF of a mutation was defined as the ratio of the number of reads with a particular mutation to the total number of reads aligned at the particular locus. The "fitdistrplus" R-package (Delignette-Muller and Dutang, 2015) was used to summarize the fold change values as continuous distributions, which in the case of the *EGFR* screen conformed to a normal distribution defined by the parameters mean (μ) and standard deviation (σ) as: $\mu = 0.093 \pm 0.011$, $\sigma = 0.967 \pm$

0.008. For *ERBB4* screens, log-normal distributions were fitted with parameters mean-log = 0.041 ± 0.007 and standard deviation-log = 0.608 ± 0.005 for screen #1 and mean-log = 0.255 ± 0.008 and standard deviation-log = 0.691 ± 0.0057 for screen #2. The *P* values were calculated for all the mutations and correction for multiple testing was performed using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). The significance thresholds used for setting cutoffs are indicated in the original publications.

4.6.2 Processing COSMIC database

Data regarding mutations identified from genome-wide screens (whole genome, and whole exome sequencing) was acquired from COSMIC (https://cancer.sanger.ac.uk) (Forbes et al., 2017; Tate et al., 2019) as a GNU zip (GZIP) archive. The samples from targeted sequencing were deliberately excluded from our analysis to avoid selection bias and to facilitate direct comparison of population-wide frequency (tumors from 35,626 samples, and 38 primary sites) of mutations in various proteins. The columns in the table containing relevant information (indicated in manuscript) were selected using awk (Aho et al., 1988) code and this filtered data table was read and processed in R using the "data.table" package (Dowle and Srinivasan, 2021).

Mutation entries were removed if they were attributed to duplicate transcripts, had unknown consequence on the protein sequence, or were silent (i.e., synonymous) mutations. To retain only unique entries, mutation id was created using the sample name, the protein name and the amino acid change, and duplicates were removed leaving 4.8 million coding mutations. Multiple samples from the same individual (e.g., primary and metastasis, or cases of multi-region biopsies) are catalogued as individual samples in the database (as they have a unique sample name). From these, all the mutations with a single recorded occurrence were removed leaving recurrent (tissue agnostic population frequency >1) mutations. For each mutation, its cumulative frequency in the data set, as well as the frequency in cancers of various tissues (primary sites) was calculated and compiled in a table comprising the Database for Recurrent Mutations (DORM).

4.6.3 Webserver to deploy DORM

DORM is hosted on a virtual private server at the premises of University of Turku, Turku, Finland. Access to DORM is facilitated through an R "Shiny" (Chang et al., 2021) web app that allows browsing and querying DORM. DORM is accessible at https://eleniuslabtools.utu.fi/tools/DORM/Mutations/, and all the communication between a client browser and the server are encrypted and handled by an NGINX reverse-proxy (https://nginx.org/). The connection is encrypted using the latest Transport Layer Security (TLS) cryptographic protocol 1.3 (Rescorla, 2018) and an

industry standard 256-bit Advanced Encryption Standard (AES-256) (National Institute of Standards and Technology, 2001). The source code for deploying DORM as an R Shiny app is available at https://github.com/dchakro/DORM_Mutations repository.

4.6.4 Testing performance and benchmarking code blocks

Benchmarking code blocks is a staple for code-optimization and it was performed using the "microbenchmark" package (Mersmann, 2021) in R. Statistical testing comparing multiple groups was performed using Brown Forsythe and Welch ANOVA test and correction for multiple testing was done by controlling the false discovery rate using the two-stage step-up method of Benjamini, Krieger and Yekutieli (Benjamini et al., 2006) in Grahpad Prism 9. Statistical testing comparing two groups of observations was done using Welch's t-test in Graphpad Prism 9. The data were graphically presented using Graphpad Prism 9.

The performance of the websites hosting the databases was measured on Google Chrome (v. 97.0.4692.99) with Google Lighthouse (v. 8.5.0) (available in Chrome DevTools).

Lighthouse (https://github.com/GoogleChrome/lighthouse) is an open-source tool for automated auditing and assessing performance metrics. The JSON data in the lighthouse reports was parsed using the 'jsonlite' R package (Ooms, 2014) and tabulated in R. The data were graphically represented using Graphpad Prism 9. Statistical testing comparing multiple groups was performed either using Brown Forsythe and Welch ANOVA test or Kruskal-Wallis test. Correction for multiple testing was done by controlling the false discovery rate using the two-stage step-up method of Benjamini, Krieger and Yekutieli (Benjamini et al., 2006) in Grahpad Prism 9.

4.6.5 Curve fitting and statistical analysis

Using the cell-proliferation data, growth curves indicating mean ± standard deviation were plotted after fitting linear-quadratic models using Graphpad Prism 9 (www.graphpad.com). Alternatively, sigmoidal curves were fitted in R using the "drc" package (Ritz et al., 2015) and plotted using the "ggplot2" package (Wickham, 2009).

Using the dose-response data, calculation of IC_{50} values and fitting of the sigmoidal dose-response curves was done in R with the four-parameter logistic regression using the "drc" package (Ritz et al., 2015). Alternatively, the sigmoidal dose-response curves were fitted using asymmetric five-parameter non-linear regression in GraphPad Prism 9. The results are graphically displayed showing the mean ± standard deviation using the R package "ggplot2" (Wickham, 2009, p. 2) or Graphpad Prism 9. The Welch two-sample *t* test) was used to compare IC_{50} values.

5 Results

5.1 High coverage random mutation library

To perform unbiased screens for activating mutations, random mutation libraries were created with PCR using an error-prone DNA polymerase (Mutazyme II) and the cDNA templates for the respective receptor tyrosine kinases, namely: *EGFR*, *ERBB4 JM-a CYT-1* and *ERBB4 JM-a CYT-2*. To retain the library diversity and minimize potential losses that can be incurred during sub-cloning from an "entry" clone (gateway cloning), the ep-PCR was designed (Gruet et al., 2012) to generate amplicons that can be incorporated directly into a destination vector (*pBABEpuro-gateway*) with a single LR-gateway reaction (more details in section 4.2).

The average mutation frequency was roughly estimated by plating dilutions from the bacteria transformed with the mutation library on LB-agar plates and Sanger sequencing regions from 4 colonies each (I, Supplementary Fig S1). Out of the several libraries created to optimize the mutation frequency (to have on average 1 amino acid change per coding insert), one was chosen for each of the three templates that had an average mutation frequency (mutations per cDNA) of 2.67 for *EGFR*, 2.11 for *ERBB4 JM-a CYT-1*, and 2.59 for *ERBB4 JM-a CYT-2*.

5.2 Cellular models to detect enhanced ERBB signaling

Ba/F3 cells were engineered to discriminate between signaling mediated by wild-type ERBBs or their previously published activating mutations. EGFR L858R (Lynch et al., 2004; Paez et al., 2004) and ERBB4 K935I (Kurppa et al., 2016) were used as positive controls to establish and optimize the conditions for the screen. For EGFR, a complete IL-3 deprivation was enough to facilitate growth of activating mutations over wild-type (WT) EGFR constructs. The difference in growth (in IL-3 depleted media) was binary, i.e., no growth with EGFR WT vs sustained survival followed by exponential growth for EGFR L858R. In case of ERBB4, the IL-3 depleted growth media required supplementation with 10 ng/mL NRG1 to facilitate ERBB4-mediated growth. Here, there was a difference in rate of proliferation of wild-type and ERBB4 K935I, giving the activating mutations (in the screen) a window of opportunity to outcompete ERBB4 WT (and passenger mutations). In brief, the screen selects

activating mutations with ligand-independent activity in the EGFR model; and activating mutations with ligand-dependent activity in the ERBB4 model.

5.3 Identification of activating EGFR mutations

On culturing the Ba/F3 cells transduced with a random mutation library of EGFR in complete absence of IL-3, a majority of the cells experienced apoptosis during the first few days (as intended). However, the surviving cells underwent clonal expansion over 2 weeks and colonized the cell culture flask. These cells were harvested, and the human *EGFR* cDNA inserts were PCR-amplified from their genomic DNA. Additionally, the *EGFR* cDNA inserts were PCR-amplified from the initial random mutant library. These sets of amplicons were sequenced with ultra-high depth (> 100,000 X) next-generation sequencing (Illumina MiSeq) and the data was subsequently processed, normalized, and analyzed to identify the EGFR mutations that thrived during IL-3 depletion. The screen included 7,216 non-synonymous EGFR SNVs (85% of the theoretical maximum, 8,485) that were present in both the library and in the cells surviving IL-3 depletion. Of these, only 21 mutations were enriched in the surviving cell pool ($q < 0.0001$) (I, Fig. 3, Table 1). Among the enriched mutations, the biggest allele fraction was represented by EGFR L858R and L858M, which are mutations in the clinically relevant EGFR Leu858 residue (Lynch et al., 2004; Paez et al., 2004). Six mutations were selected to be validated by creating Ba/F3 cells with stable expression by using expression vectors encoding the mutations individually. Additionally, the Ba/F3 cells were transduced with expression vectors encoding EGFR WT and enhanced green fluorescent protein (eGFP) to use as controls. Out of these six, three mutations (EGFR A702V, T790M and L858R) were found to induce IL-3-independent growth in the Ba/F3 cells (I, Fig. 4). EGFR L858R and T790M have previously been reported to enhance EGFR activity (Lynch et al., 2004; Paez et al., 2004; Regales et al., 2007). EGFR A702V was identified as a novel activating mutation of EGFR in the juxtamembrane region.

5.4 Biochemical characterization of activating EGFR mutants

Upon culturing the Ba/F3 cells individually expressing these constructs in the absence of IL-3, three mutants (A702V, T790M and L858R) also showed enhanced EGFR tyrosine phosphorylation (compared to EGFR WT) (I, Fig. 5 A). Ba/F3 cells expressing these mutants when cultured in absence of IL-3 showed an increase in EGFR expression at a protein (Flow Cytometry) and mRNA (qPCR) level (I, Fig. 5 C-E). Cells expressing these three mutants were sensitive to EGFR monoclonal antibody cetuximab to a similar extent (I, Fig. 7 A) but had varying degrees of response to erlotinib and afatinib (I, Fig. 7 A-B). Consistent with the published

literature (Pao et al., 2005), cells expressing EGFR T790M were resistant to inhibition by erlotinib (I, Fig. 7 A-B).

5.5 Identification of activating ERBB4 mutations

Ba/F3 cells were transduced with the wild-type receptors (JM-a CYT-1 and JM-a CYT-2), random mutation libraries of ERBB4 (JM-a CYT-1 and JM-a CYT-2 isoforms), or empty vector were used to control for retroviral transduction. Neither of these cells survived a complete depletion of IL-3 from their growth media. However, when the IL-3 depleted growth media was supplemented with 10 ng/mL NRG1, the cells with the ERBB4 JM-a CYT-2 mutant library demonstrated an enhanced rate of proliferation in comparison to their wild-type counterpart (II, Supplementary Fig. S2B). The cells were passaged before reaching confluency to maintain them in the exponential growth phase. After ten days of culture in NRG1 supplemented growth media the growth rate of the cells harboring the library equaled that of the cells expressing the wild-type ERBB4 (II, Supp. Fig 2B). Therefore, cells frozen on day 8 were used as the end point, and their genomic DNA was extracted. The *ERBB4* cDNA inserts were PCR-amplified using primers targeting the retroviral backbone. Additionally, the *ERBB4* inserts were PCR-amplified from mutation library, and these sets of amplicons were sequenced with ultra-high depth (> 100,000 X) next-generation sequencing (Illumina MiSeq). The data were subsequently processed, normalized, and analyzed to identify the ERBB4 mutations with an enhanced rate of proliferation in presence of 10 ng/mL NRG1.

The screen included 7,396 mutations (91.7% of the theoretical maximum 8,065) that were detected both in the surviving cells and in the mutant library. Of these, ten mutations were enriched in the surviving cell pool ($q < 0.00001$) (II, Fig. 1, Supp. Fig. 4). To validate the findings of the screen, Ba/F3 cells were transduced with these ten mutations after they were individually cloned in retroviral expression vectors. Five (R687K, E715K, G741R, G802D, and M993I) out of ten mutations were able to promote IL-3 independent growth of the Ba/F3 cells in presence of 10 ng/mL NRG1 in comparison to the cells expressing ERBB4 wild-type ($q < 0.001$) (II, Fig. 2 A). When the screen was conducted again with a newly synthesized library, four (R124K, R687K, E715K,and G741R) out of the ten mutations were enriched in the surviving cell pool ($q < 0.00001$) (II, Fig. 2 D).

5.6 Biochemical characterization of activating ERBB4 mutants

The investigation of ERBB4 expression and its phosphorylation status in the Ba/F3 cells revealed ERBB4 R687K and E715K to have more basal phosphorylation (II, Fig. 2 B) as well as NRG-induced phosphorylation (II, and Fig. 2 C). Additionally, the

ERBB4 mutations were also transduced in NIH-3T3 cells (all ten hits from first screen), and in the BEAS-2B cells (R124K, R687K, E715K, and G741R, i.e., the four mutations identified in both repeats of the screen II, Fig. 1 and 2 D). Western analysis of the NIH-3T3 cells expressing ERBB4 R687K, E715K, G741R and G802D showed increased tyrosine phosphorylation in comparison to ERBB4 wild-type (II, Supplementary Fig. S7B). Furthermore, ERBB4 R124K, R687K, E715K were found to induce more ERBB4 tyrosine phosphorylation in the BEAS-2B cells (II, Fig. 3 B). However, other than the positive control, ERBB4 K935I, only ERBB4 R687K and E715K were able to promote 3D-growth of BEAS-2B cells in 2% Matrigel on poly-HEMA–coated plates (II, Fig. 3 A). When analyzed *in vivo* (NMRI nude mice), Ba/F3 cells expressing ERBB4 E715K formed tumors significantly faster than its wild-type counterpart as well as other tested mutations like ERBB4 R687K, and G741R (II, Fig. 3 C).

The Ba/F3 cells expressing the ERBB4 mutations were incubated with increasing concentrations of small-molecule pan-ERBB tyrosine kinase inhibitors (TKIs) (namely, afatinib, neratinib, and dacomitinib), as well as an EGFR-specific inhibitor erlotinib. All the tested mutations (ERBB4 R687K, E715K and G741R) were sensitive to the pan-ERBB TKIs with IC_{50} values in nanomolar range (II, Fig. 4).

5.7 ERBB4 E715K is constitutively active

ERBB4 E715K displayed enhanced activity in all the *in vitro* validation experiments and characterizations in three different cellular models (II, Fig. 2 A-C, 3 A-B and Supplementary Fig. S7), as well as had the shortest tumor latency in vivo (II, Fig. 30 C-D). Western blot analyses in Ba/F3, NIH-3T3 and BEAS-2B cells revealed ligand independent constitutive phosphorylation of the ERBB4 E715K molecule (II, Fig. 2 B, 3B, and Supplementary Fig. S7 B). Additionally, the Ba/F3 cells expressing ERBB4 E715K were the only cell population capable of sustaining NRG-independent growth (i.e., survival and subsequent proliferation in media devoid of IL-3 as well as NRG1) (II, Fig. 5 A).

Furthermore, these NRG-independent Ba/F3 cells expressing ERBB4 E715K had an upregulated protein expression of the ERBB4 protein (II, Fig. 5 B). As expected, these cells were consequently more sensitive to tyrosine kinase inhibition to pan-ERBB TKIs (namely, afatinib, dacomitinib, neratinib) in comparison to the NRG-dependent Ba/F3 cells expressing E715K (II, Fig. 5 C).

In addition to the dimerization of the extracellular domains, the kinase domains of ERBB family members also form an asymmetric dimer where the N-terminal lobe of the "activator kinase" comes in contact with the C-terminal lobe of the "receiver kinase" (II, Supplementary Figure S10). ERBB4 E715K was shown to be a more potent receiver kinase in homodimeric (i.e., with ERBB4) and heterodimeric

configurations (with EGFR, ERBB2, and ERBB3) (II, Fig. 5 F) (assay performed by transient over-expression of ERBB-family members in COS-7 cells).

5.8 Activating mutations situated at the ERBB kinase dimerization interface

Structural modelling (detailed methodology described in the original publications) revealed that EGFR A702V (identified from the EGFR iSCREAM), and ERBB4 E715K (identified from ERBB4 iSCREAM), are both situated in the N-terminal lobe of the receiver kinase (I, Fig. 6 and II Fig. 5 D). These mutations exist at the surface of the receiver kinase which forms interactions with the C-terminal lobe of an activator kinase that are essential for the activation of the ERBB kinase (Qiu et al., 2008; Zhang et al., 2006). EGFR A702V-I941 interaction was modelled to strengthen hydrophobic interactions with the activator kinase (I, Fig. 6). By contrast, the ERBB4 E715K-E934 interaction was predicted to strengthen the kinase dimer via ionic interactions (II, Fig. 5 E). Additional molecular dynamics simulation (detailed methodology in the original publications) also revealed overall improvements in stability of EGFR A702V and ERBB4 E715K in their respective ERBB homodimers in comparison to their wild-type counterparts (I, Supplementary Fig. S8, and II Fig. 5 E).

5.9 Database of recurrent mutations

The increasing utility and accessibility of next-generation sequencing has led to an accumulation of large amounts of data (Campbell et al., 2020). This information-rich data is presented on expansive databases which are designed to present comprehensive information, often requiring multiple operations (examples of these tasks are a) selecting data sets with non-overlapping samples, b) searching the mutations of interest manually and counting number of rows, c) using a spreadsheet software to calculate number of cases with desired alterations, etc.) from a user to obtain key pieces of information like the number of occurrences of a particular mutation in a dataset or a cancer type of interest. Therefore, the Database Of Recurrent Mutations (DORM) (III, Fig. 2), which is a database that lists recurrent mutations (tissue-agnostic population frequency > 1) was created. The DORM database (eleniuslabtools.utu.fi/tools/DORM/Mutations) has faster response times, high performance score (Google Lighthouse) and becomes responsive to user input within seconds (III, Figure 1). The feature-rich database supporting searching for coding alterations several genes, the encoded amino acid substitutions as well as advanced queries (with regular expressions) (III, Figure 4).

The raw data was obtained from the Catalogue Of Somatic Mutations In Cancer (COSMIC) public release v95, and processed to remove duplicates and identify

unique coding mutations and uniquely mutated amino acid residues (III, Supplementary Fig. S1). This information is presented using two R-shiny webtools giving the user a fast and easy-to-navigate platform to browse recurrent mutations (III, Figure 3). The results can also be filtered by a specific cancer tissue; and the search scope can be limited to a selected tissue (e.g., searching KRAS mutations in lung vs pancreas). A user can also perform advanced search queries using regular expressions (III, Figure 4 C) to combine several search terms e.g., searching for "EGFR|ERBB" lists all the mutations in the four receptor tyrosine kinases in the Epidermal Growth Factor Receptor family, and searching for "`RAS\> C\>`" lists the mutations in any residue of the RAS-family of proteins that create a Cysteine (Cys, C). In addition to the code for the web-interface that presents the database (https://github.com/dchakro/DORM_Mutations), the data processing pipeline (https://github.com/KE-group/generate_DORM) contains highly-optimized (II, Supplementary Fig. S2), and parallelized code (i.e., improved performance by utilizing multiple cores of modern CPUs) for fast and automated processing of new COSMIC data releases.

When the performance of DORM was compared (using Google Lighthouse, https://github.com/GoogleChrome/lighthouse) to contemporary cancer genomics databases such as the cBioPortal (Cerami et al., 2012; Gao et al., 2013), AACR Genie (The AACR Project GENIE Consortium, 2017), ICGC data portal (Zhang et al., 2011) and COSMIC (Forbes et al., 2017; Tate et al., 2019). These databases (namely, AACR GENIE, COSMIC, cBioPortal and ICGC data portal), were found to have high latency in processing user's queries and were resource intensive (III, Figure 1 and Supplementary Fig. S3). Furthermore, COSMIC, cBioPortal and AACR Genie contain duplicate entries (because a few samples are incorporated in multiple studies). DORM (and COSMIC) allows searching individual mutations and amino acid residues directly, but DORM is the only database that can process a search term containing both a HUGO gene symbol and a specific mutation, e.g., "KRAS G12C, EGFR". In addition to information about a specific amino acid change, DORM also aggregates information at the level of individual amino acid residues (i.e., the results for BRAF V600 include various amino acid substitutions at the specific residue such as BRAF V600E/K/M/R/G/D) and the information is available at eleniuslabtools.utu.fi/tools/DORM/Residues.

5.10 The long tail of infrequent recurrent mutations

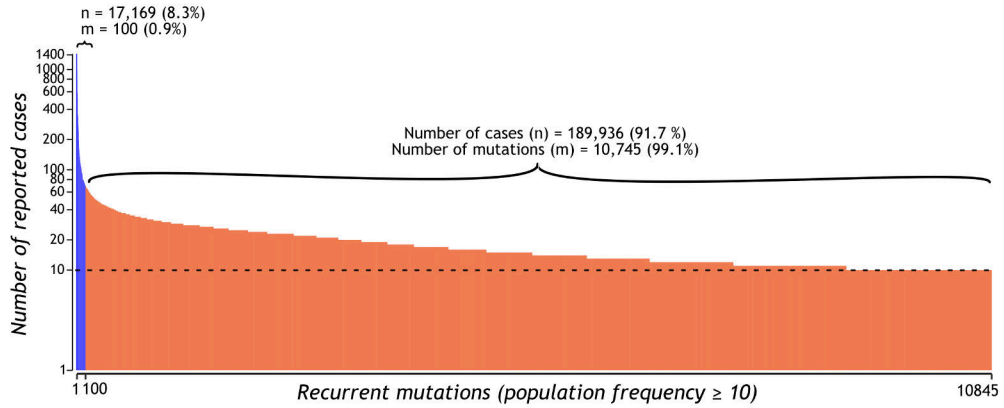


Figure 9. The distribution of recurrent mutations (population frequency ≥ 10). The top 100 recurrent mutations are highlighted in violet. The number of cases (n) are shown on y-axis (log scale) and number of mutations (m) (on x-axis) comprising the two marked zones (purple and orange, annotated in figure). The percentages are calculated from the data that is displayed in the graph and not for the entire data set (containing 4.8 million mutations). (Source: original analysis and figure. Data from COSMIC v95 available at cancer.sanger.ac.uk).

The COSMIC v95 data release was processed (details in section 4.6.2 and 5.9) and 4.82 million unique coding mutations were identified. From these, 2.93 million (61%) were observed just once in the dataset (n = 35,462 samples). Among the 1.88 million unique recurrent mutations (i.e., mutations with tissue-agnostic population frequency > 1) the ten most recurrent mutations (i.e., rank 1-10), collectively, were observed in 17% of the samples (III, Figure 2), while the next 90 mutations (rank 11-100) were observed in 18.3% of samples all together. However, in the context of sheer numbers these two groups are comprised of just 0.34% (rank 1-10) and 0.56% (rank 11-100) of the total number of recurrent mutations. The observed frequency of the recurrent mutations spans a wide range, i.e., the first mutation on the list of most recurrent mutations (BRAF V600E) is reported in 1,432 samples, the 100th (NRAS Q61L) in just 70 (III, Figure 2). However, it is astounding that there exists a "long tail" to this histogram (Figure 9), as there are in total 10,745 mutations with an observed tissue-agnostic population frequency between 10 and 70. The bulk (91.71%) of this histogram (total n = 207,105 occurrences) is present in the tail (of 10,745 mutations) which is comprised of infrequent mutations (n = 189,936 occurrences, with population frequency: mean = 19; median = 14; 3rd quartile = 22 samples). In addition to this, there is an appreciable variation in the distribution of these infrequent recurrent mutations between various cancer types, for instance, among the cancers with at least a hundred samples, the highest percentage of recurrent mutations was in thyroid

cancers (89.6%; *n* = 139,883 mutations; sample size = 989), while the lowest percentage of recurrent mutations was in adrenal cancer samples (17.5%; *n* = 5,229 mutations; sample size = 282). This trend is not a direct result of the background rate of mutation in these cancers as the samples with highest tumor mutational load (defined as the number of mutations per sample) are from cancers of the endometrium (mean (μ) = 562 mutations/sample) and skin (μ = 508) (III, Figure 4 D). The adrenal cancer samples however, have a low background mutation rate (μ = 19) (III, Figure 4 D).

6 Discussion

6.1 Strengths and future prospects of the iSCREAM methodology

The iSCREAM workflow (I, Figure 2) was established and validated using a random mutant library containing over 7,200 Epidermal Growth Factor Receptor (EGFR) cDNA variants. After characterization of selected hits from the screen, EGFR A702V emerged as a novel activating mutation. Additionally, the unbiased screens identified clinically actionable and relevant mutations like EGFR L858R and EGFR T790M (article I). ERBB4, another member of the EGFR family was studied with iSCREAM, to identify ERBB4 R687K and ERBB4 E715K as potent activating mutations from a library of almost 7,400 cDNA variants (article II). The iSCREAM methodology was successful in a high-throughput characterization of libraries of variants and in reducing the "search space" and generating candidates for validation. Demonstrating reproducibility among randomness (i.e., finding same mutations between the screens performed with two independently synthesized libraries with distinct mutational composition), four out of ten hits in the first screen (II, Figure 1) with the ERBB4 RTK were found to be enriched in a second screen (II, Figure 2D) which was performed with a newly synthesized library.

The iSCREAM methodology was successful in demonstrating that there are driver mutations that are present among the infrequent cancer-associated somatic mutations (i.e., the crowd of mutations at the base of the lollipop diagrams (in I, Figure 1 and II, Supplementary Fig. S1 B). The work described here, with *EGFR* and *ERBB4*, indicates that there are novel driver mutations still to be characterized from the large lists of variants of unknown significance. The iSCREAM methodology developed here can be used for the biochemical characterization of other kinases too. This is in part due to the robustness (incorporates a variety of mutations at a controlled rate) of our library synthesis strategy, and also because the Ba/F3 cells have already been used to study activating mutations in several other kinases such as BRAF, KIT, FLT3, FGFR2, FGFR3 BCR-ABL, NTRK1 (Bradeen et al., 2006; Byron et al., 2013; Chen et al., 2005; Fuse et al., 2017; O'Hare et al., 2009; Shi et al., 2016; Warmuth et al., 2007; Weisberg et al., 2002; Whittaker et al., 2010) in addition to mutations in the ERBB family members (Greulich et al., 2012; Jaiswal et al., 2013; Jiang et al., 2005; Wang et al., 2016).

Essentially, the iSCREAM methodology described here is a tool to emulate cancer evolution in vitro and with little modifications it can be used to extend the framework further. For instance, we know that tumor microenvironment plays a significant role in tumorigenesis and maintenance; this can be achieved by using appropriate xenograft models (like the immortalized and transformed human lung airway epithelial SALE-Y cells, that have been used in high-throughput screens (Berger et al., 2015; Vichas et al., 2021)). Targeted inhibitors are known to alter the tumor population and the mutational landscape during the emergence of resistance (Dagogo-Jack and Shaw, 2018; Uribe et al., 2021), the introduction of inhibitors during or after the first enrichment can help identify on-target mutations driving de-novo resistance. Additionally, the mutation library could be built using the sensitizing mutations as the template, in attempts to generate resistant mutations and a screen will identify on-target mutations driving acquired resistance.

At its core, the iSCREAM methodology serves as a framework to link a particular genotype to a phenotypical readout, therefore, with appropriate experimental models the iSCREAM can be adapted to characterize other molecules of interest, potentially even in other diseases and biological conditions (provided there exist models that can serve as good phenotypic indicators of cell signaling activity).

6.2 Limitations of the iSCREAM methodology

The characterization of EGFR and ERBB4 random mutant libraries with iSCREAM enabled identification of activating mutations, but also produced a several false positives. This property necessitates a diligent validation of the hits identified from the screen. It is a common practice for high-throughput screens to consist of assays aimed at validation and characterization of the findings. However, a high rate of false positive discoveries in the screen causes the necessary but undesirable utilization of the technical resources on analyzing mutations that are likely to be false positives. iSCREAM could benefit from approaches geared at reducing the number of false positive hits. For instance, performing independent experiments with distinct ERBB4 random mutant libraries, 3 out of 4 mutations (common between the two screens) were confirmed as positive hits (75%). Some limitations come from the experimental models and techniques used to establish iSCREAM, for instance, the mutation frequency of 1 amino acid change / cDNA (I, Supplementary Fig. S1), means that in addition to cDNAs with 1 amino acid change, there are several synthetic inserts with multiple cDNA/amino acid changes. Though, these can be characterized with long-read sequencing (e.g., PacBio High Fidelity Circular Consensus Sequencing), the phenomenon of multiple mutations in a single cDNA opens the interesting avenue of studying the co-occurring mutations as composite mutations (Gorelick et al., 2020; Saito et al., 2020).

Delivery of transgenes via viral vectors is an essential tool in molecular biology, however there is limited control on number of integrations per cell. The iSCREAM studies published so far use retroviral vectors based on Moloney Murine Leukemia Virus (MLV) for delivering the transgene. MLVs are known to only infect mitotic cells (Roe et al., 1993), thereby reducing the transduction efficiency compared to alternatives, but it also reduces the likelihood of more than one viral integrations per cell. But, given that bulk genomic DNA is harvested and sequenced (targeted to virus), as such there is no way to discern number of viral integrations per genome per cell. Advances in single cell sequencing could perhaps help address this concern.

iSCREAM failed to facilitate enrichment of several of the previously characterized EGFR (Kobayashi and Mitsudomi, 2016; Kohsaka et al., 2017) and ERBB4 mutations (Kurppa et al., 2016; Nakamura et al., 2016). There are three main factors that could be at play: 1) co-occurrence with an inactivating/dampening mutation 2) stochastic nature of synthesis of mutations, and introduction in the cells 3) stronger mutations outcompeting the weaker mutations. However, this is in line with the original idea of emulating somatic evolution of cancer *in vitro*, where all three of these factors influence the success of "winners" (Burrell and Swanton, 2014; Posada, 2015). Moreover, as a cancer in a patient undergoes evolution on a longitudinal time scale (Abbosh et al., 2017; Jamal-Hanjani et al., 2017), while most routine next-generation sequencing in the clinic as well as conducted during iSCREAM captures a snapshot of that process. Future iterations of iSCREAM could perhaps incorporate several selection pressures (such as introduction of targeted inhibitors) and establish a longitudinal data series.

As the mutation libraries contained only point mutations and small (1-3bp insertion-deletions it effectively limits the "search space" for driver events to coding point mutations and frameshifts. There are several other types of alterations which are excluded from the analysis in iSCREAM (and admittedly, also in DORM) such as, small insertion-deletions (e.g., EGFR exon 19 deletions, and ERBB2 exon 20 insertions are known driver events), non-coding mutations (e.g., in regulatory regions), copy number aberrations (e.g., ERBB2 amplification in gastric and breast cancer), chromosomal re-arrangements, epigenetic modifications). It is of note, that a recent characterization of cancer genomes demonstrated that over 75% of samples contained on average 2-3 coding driver alterations (point mutations) (Campbell et al., 2020), which is line with previous estimates (Martincorena et al., 2017). Therefore, coding point mutations were a reasonable starting point to build screening methodology. Lastly, even though the identified hits, i.e., EGFR A702V, and ERBB4 R687K and E715K, are found in clinical samples (cbioportal.org, cancer.sanger.ac.uk), the clinical significance of these mutants and the response to ERBB-targeted inhibitors in the clinic is indeterminate as of yet.

Another set of differences between *in vivo* and *in vitro* observations can be potentially introduced due of the cell culture conditions. For instance, the three-

dimensional growth experiments with BEAS-2B cells were conducted on Matrigel as the substrate (described in section 4.5.7), but, recent studies have documented to bridge the gap between *in vivo* and *in vitro* phenotype by using novel human leiomyoma–derived matrix (Tuomainen et al., 2019; Wahbi et al., 2020). Furthermore, all of the cell culture work in this thesis was performed in normoxic conditions, however, it is known that several human tumor microenvironments have conditions of hypoxia (Bhandari et al., 2019; Brown and Wilson, 2004) and EGFR and ERBB4 activity has been shown to be affected by oxygen concentration (Lu et al., 2018; Misra et al., 2012; Paatero et al., 2012).

The iSCREAM methodology and the conclusions are dependent on several variables and their potential confounding effects were discussed above. It is worth mentioning that characterization of genes or mutations as “driver” or “non-driver” in cancer biology is a dynamic process, and it evolves with acquisition of new insights. As, in the case of Maternal Embryonic Leucine Zipper Kinase (MELK), years of accumulated evidence positioning it as an oncogene was brought into question by further characterization and meticulous dissection of the underlying biology (McDonald and Graves, 2020; Settleman et al., 2018).

6.3 Prospects of functional genomics screens

Functional screening for characterizing variants of proteins has come a long way in this past decade. There are so many alternatives (see section 2.4) all the way from synthesis of the library to the choice of cell models for characterizing the variants of a gene. The most fascinating recent discoveries have been the developments in synthetic oligos which have eased the accessibility of completely synthetic libraries (Kitzman et al., 2015; Plesa et al., 2018). The advantage of such a library is the ability to literally "program" the mutation library to have desired characteristics. This allows granular control over every aspect of the library, such as modulating and exactly controlling the mutation frequency (mutations/cDNA), incorporation of various types of mutations (e.g., can easily add specific insertion, deletion, and frameshift mutations), and incorporate promoter polymorphisms. Another interesting approach was demonstrated by Yenerall and colleagues, where they introduced a point mutation in the HIV reverse transcriptase (M230I) which increased the enzyme's intrinsic error rate by three-fold, and reduced the enzymes mutational bias without significantly altering the enzymes processivity (Svarovskaia et al., 2003; Yenerall et al., 2021). This mutant abrogated the need for the synthesis of mutation libraries, as the viral polymerase creates mutations itself. However, there are clear drawbacks, like the inability to modulate the mutation rate set by the mutant polymerase. Additionally, there is no scope to establish "biological" repeats because every transduction, starting with the wild-type cDNA, will have a unique mutational composition.

iSCREAM and similar studies are performed by introduction of mutant transgenes (libraries) into the otherwise stable genomes of cell models. However, an area of active research has been to establish systems to perform saturation editing of genomic regions, thereby enabling characterization of mutations in endogenous loci (Findlay et al., 2018, 2014). The approach utilizes CRISPR-Cas9 with and guide RNA targeted to target certain exons, then a synthetic library consisting of mutant templates of the targeted exon is able to recruit the homology-directed repair machinery to introduce the chosen variants (present in sgRNA library). The approach was used to perform functional screen by saturation editing of 13 exons of BRCA1 that encode the RING1 and BRCT domain (Findlay et al., 2018, p. 1). A different approach uses cytosine base editors to achieve sub-saturating editing of genomic regions with a simplified approach of using a single guide to target the CRISPR-base editor to introduce transition mutations at desired endogenous loci (Gaudelli et al., 2017; Hanna et al., 2021). Considering the highly heterogenous nature of cancer, these methods open up interesting avenues for characterizing variants in the genes of interest in the appropriate cellular contexts which may help improve the translational aspect of the discoveries (Haigis et al., 2019; Schneider et al., 2017).

6.4 Selection bias skews the true prevalence of mutations

Over the years, we have grown accustomed to seeing the classical lollipop figures that show the population frequency of various genetic alterations observed in a particular gene product. However, those figures are an intentional "zoomed-in" version of the real mutational landscape of cancer. In practice, hotspot mutations, though apparently frequent, are found to be a rare occurrence in comprehensive analysis of the cancer genomes (Chang et al., 2018, 2016). The survey of cancer genomes in article III, deliberately focused on genome-wide studies to evade the selection bias that can be introduced with the incorporation of targeted sequencing panels (e.g., MSK-IMPACT, and other commercially available "Cancer Panel" kits). Although these kits offer a cost-effective and efficient strategy to identify druggable drivers and predictive mutations; from the perspective of comprehensive cancer genomics, they introduce an over-representation of the selected few. This bias in the frequency of mutations in those genes manifests in the statistics by masking the true population frequency of other mutations. For instance, EGFR L858R is the hotspot driver mutation in lung adenocarcinoma, and it ranks 84th (v = 75) when analyzing genome-wide data (III, Figure 2); but, when data from the targeted screens is included in the analysis, it ranks 6th (v = 10,631). The selection bias is clearly evident when we consider that incorporating data from targeted sequencing studies increases the number of samples by over 900%, while only adding 12% mutations on top of the mutations reported by genome-wide studies (Table 6).

Table 6. Changes introduced by the integration of data from targeted sequencing studies and the data from genome-wide studies. The table shows size of indicated entities (first column) in the genome-wide data, and in the full data (i.e., genome-wide + targeted). The last column shows the change (as %) caused by addition of data from targeted sequencing. Data source: COSMIC release v95.

Category (n)	Genome-wide	Genome-wide + targeted	Change (%)
Sample count	36,224	364,241	906 %
Mutations	4,823,109	5,399,206	12 %
Recurrent mutations (*v* >1)	1,887,757	2,369,282	26 %
Mutations (*v* =1)	2,935,352	3,029,924	3 %
Top 100 recurrent mutations	17,169	220,526	1184 %

Moreover, the bulk of this addition happens to the most recurrent mutations (rank 1-100 in DORM), where a 1184% increase can be seen in their representation when data from targeted sequencing studies is included in calculations (Table 6). To summarize, the inclusion of targeted-sequencing data to compute comparative statistics on a genome-level introduces a strong selection bias that makes the histogram of the "apparent" population frequency of alterations in the cancer-associated genes front-heavy and skews it away from the "true" population frequency of alterations observed in the human population (as depicted in III, Figure 2).

6.5 Lineage diversity and variant allele specificity in ERBB family

My analysis of 4.82 million unique coding somatic mutations identified in genome-wide studies of from cancer samples revealed that a majority (61%) of these have been observed in the population only once (i.e., with a tissue-agnostic population frequency of one) (II, Figure S4 C). There have been observations of lineage diversity and variant allele specificity among some hotspots (Chang et al., 2016; Gorelick et al., 2020), wherein a particular mutation is more common in specific cancers (than what random variation would allow). This phenomenon can also be observed among the ERBB family members of receptor tyrosine kinases to varying degrees (Figure 10).

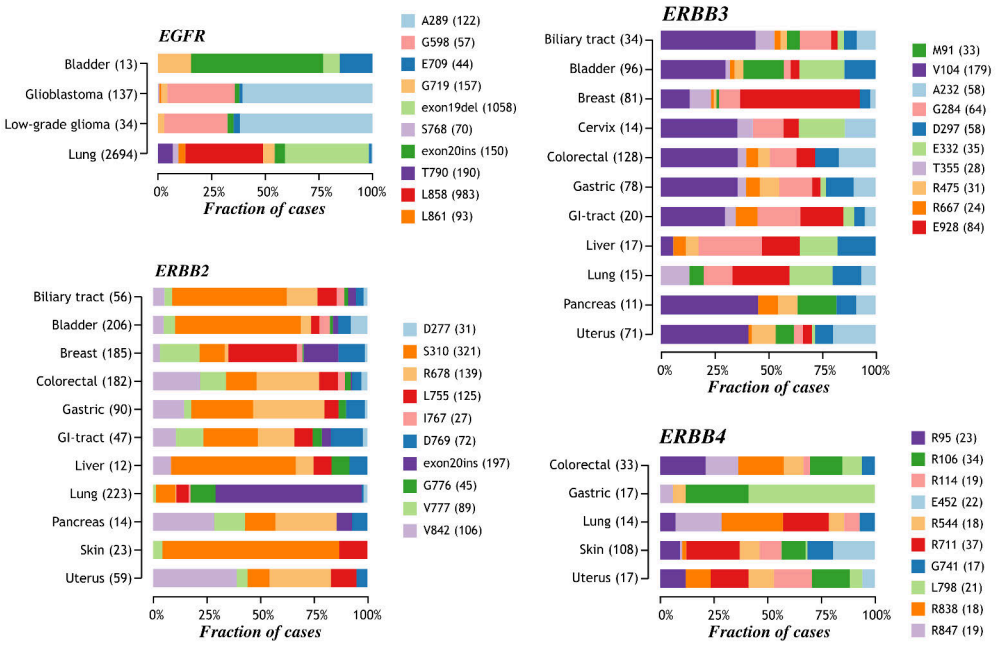


Figure 10. Fraction of cases with mutations in 10 of the most frequently observed hotspots for each member of the ERBB family of RTKs. The distribution of mutations is presented in the most commonly mutated tissue lineages (i.e., tissues with more than 10 mutations) depicting the varying degree of lineage diversity and hotspot specificity pertaining to the tissue of origin of the cancer. The number of individual mutations in each category is shown in parentheses next to the label for hotspots or tissue. (Source: Original analysis and figure. Data was sourced from cBioPortal on 11 May, 2022).

In case of EGFR, the specific association of extracellular domain alterations to glioblastoma and of the intracellular kinase domain mutations to non-small cell lung cancer is well-known (An et al., 2018; Lee et al., 2006; Paez et al., 2004; Pao et al., 2004; Vivanco et al., 2012), however the functional significance of the apparent association of other ERBB mutations to specific cancers (Figure 10) still remains to be deciphered. This effect is also observed in other oncogenes and tumor suppressors such as KRAS, PIK3CA, APC, PTEN (Chang et al., 2016; Gorelick et al., 2020), and in the case of KRAS Gly12, even specific amino acid substitutions have an appreciable lineage diversity (Chang et al., 2016).

7 Conclusions

Advances in cancer genomics have allowed identification of several driver alterations in various human cancers, however, the process has also generated a significant number of variants of unknown significance (VUS). The VUS have not been characterized, largely due to their small population frequency. This study aimed to develop tools to improve our understanding and showcase the true representation of VUS in cancer.

The following conclusions can be drawn based on the findings of this study:

1. An unbiased high-throughput screen was developed to screen functionally activating mutations in kinases. The workflow was named *in vitro* Screen for Activating Mutations (iSCREAM) and expedites screening thousands of point mutations of a gene.
2. The iSCREAM methodology was established using EGFR as a model, and the screen identified clinically relevant mutations (EGFR L858R and T790M). Moreover, a previously reported VUS (EGFR A702V), enriched in the screen, and was demonstrated to have ligand-independent activation of EGFR signaling. A702V was differentially sensitive to first and second generation of EGFR-TKIs.
3. Novel activating ERBB4 mutations were identified using iSCREAM. In functional assays, ERBB4 R687K and E715K had increased activity in three different cell models. The constitutively active ERBB4 E715K accelerated tumor growth *in vivo* (Ba/F3 allograft) and was the first reported ERBB4 variant to induce ligand-independent transformation of the Ba/F3 cells.
4. Contemporary databases of cancer-associated mutations are slow (take up to 4x-29x longer to respond to queries when compared to DORM) and are imbued with a heavy selection-bias that is due to the incorporation of data from targeted sequencing. Therefore, DORM was developed as a fast (low latency in processing queries and in becoming interactive) and robust database that presents the true population frequency of reported cancer-associated somatic mutations in protein coding genes.

5. A highly optimized, fast (processes 46.2 million variants in 30 minutes), and automated data processing pipeline was developed to compile the DORM database. The codebase will be made open source (upon publication of the manuscript III).

Together, the findings of this thesis established two tools: iSCREAM and DORM. iSCREAM is a high-throughput workflow to identify candidate activating mutations in ERBBs and other kinases. iSCREAM can be further developed to study composite mutations, on-target drug resistance, and cancer evolution, among other things. The framework for DORM can be used to analyze other databases and even extended to other diseases that can benefit from analysis of variants from genomics data.

Footnote: The term "population" in the text above refers to cancer samples that have been analyzed with DNA sequencing worldwide and are aggregated by the COSMIC cancer registry (accessible at cancer.sanger.ac.uk).

8 Acknowledgements

The work presented in this thesis was carried out at the Faculty of Medicine, Department of Medical Biochemistry and Genetics, University of Turku; Medicity Research Laboratories, University of Turku; and Turku Bioscience Centre, University of Turku and Åbo Akademi University. I am grateful to Professors Klaus Elenius, Riitta Lahesmaa, John Eriksson, and Academician Sirpa Jalkanen for providing a nurturing environment and outstanding facilities to conducting research.

I have profound gratitude for my supervisors – professors Klaus Elenius and Laura Elo; my academic journey had plenty of chances to go astray, thankfully you were awesome at "course-correction". Klaus, I owe everything I've accomplished to your unconditional support, seemingly infinite compassion (for negative data), and a consistent motivation. I thank you for the guidance, mentorship, and encouragement that helped me get through the troughs of the "*roller coaster of PhD*". Laura, thank you for integrating me into your newly formed group as a summer intern, which paved the way towards an MSc thesis project in your laboratory. Thank you for opening a door that led me so deep into the world of science and helped me realize my passion.

With a warm heart, I acknowledge the generous contributions of Drs. Ilkka Paatero & Kari Kurppa as members of my thesis committee. Thank you, for taking a genuine interest in my doctoral studies and research, and for sharing relevant practical and professional advice.

I will remain eternally grateful to Professor Juha Kere and Professor Steve Bova for diligently reviewing my thesis manuscript and for their constructive criticism, valuable feedback. I have earnest admiration and gratitude towards Professor René Bernards for accepting the invitation of serving as the respected opponent during the public defense of my PhD thesis.

I thank the Turku Doctoral Programme of Molecular Medicine (TuDMM) and its past and present director, professors Kati Elima and Noora Kotaja for accepting me in the doctoral school, and the financial support. I thank my fellow PhD students at the University of Turku, especially those in the TuDMM and DRDP programs, for the fantastic peer group and for the fun gatherings. I also thank Eeva Valve for coordinating TuDMM's operations, and Outi Irjala for being an excellent guide through the academic formalities of grad school.

The work in this thesis wouldn't have been possible without financial support and travel grants issued by the K. Albin Johanssons stiftelse, Juhani Ahon Lääketieteen Tutkimussäätiö, Finnish Cancer Organizations, Turku University Foundation, and the LIFESPAN research program. I also want to extend my sincere gratitude for contributors to the Finnish Taxes, because every year a portion of it gets utilized to educate and train thousands of students (just like me) at our universities.

I am extremely grateful to my co-authors and collaborators Drs. Kari Kurppa, Veera Ojala, Mahlet Z. Tamirat, Mark S. Johnson, Ilkka Paatero, Marika Koivu, Jasmin Drexler, Regina Ruzicka, Karin Bosch, Johanna Woertl, Susanne Schmittner, Pasi Jänne, Jussi Koivunen, and Flavio Solca for their impactful contributions to the work included in this thesis. I am deeply appreciative of the work carried out by Professor Mark Johnson and Mahlet with the structural biology of the *ERBB* mutations and Dr. Solca's team at Boehringer Ingelheim for characterization of the *ERBB4* mutations in various models including the mouse allografts.

I have heartfelt appreciation for the past and current colleagues at the Elenius laboratory who have been monumental in shaping the way I conduct science, Kaisa Aalto, Iman Farahani, Nelli Heiskanen, Juho Heliste, Maria Helkkula, Anne Jokilammi, Peppi Kirjalainen, Anna Knittle, Marika Koivu, Kari Kurppa, Johannes Merilahti, Matias Mäenpää, Elli Narvi, Veera Ojala, Fred Saarinen, Minna Santanen, Mika Savisalo, Farid Siddiqui, Zejia Song, Jori Torkkila, Maria Tuominen, and Katri Vaparanta. Furthermore, I also thank everyone associated with the Medical Bioinformatics Centre led by Professor Elo, especially those that I have spent office hours and had fascinating discussions with, Alexander Biehl, Thomas Faux, Ye Hong, Maria Jaakkola, Markku Karhunen, Sofia Khan, Riku Klén, Teemu Daniel Laajala, Mehrad Mahmmoudian, Arfa Mehmood, Esko Pakarinen, Anna Pursiheimo, Xu Qiao, Kalle Rytkönen, Fatemehsadat Seyednasrollah, Kalaimathy Singaravelu, Tomi Suomi, Veronica Suni, An Le Thi Thanh, and Ning Wang. I would also like to thank Dr. Jukka Lehtonen for helping organize high-performance computing at the Structural Bioinformatics Laboratory at Åbo Akademi University.

It fills my heart with joy on remembering the time I spent in the laboratory space shared between members of the Elenius, Heino and the Kurppa labs. I appreciate your superb company and am very thankful for help in troubleshooting. Abbi, Elina, Johanna, Juuli, Kamal, Noora, Meija, Mari, Meri, Salli, Sarang, and Sini are thanked.

In addition to the work presented in this thesis, I was fortunate to participate in a few studies carried out in close association between Professor Elo and Professor Lahesmaa. I have utmost appreciation for Riitta's dedication to high-quality science and for her way of leading fruitful collaborations. On a personal note, I have the warmest gratitude for her enthusiastic encouragement and the precious advice. I also thank Professor Timo Veromaa, for candidly sharing his experiences from the Biotech industry, and bolstering my motivation.

Conducting research work at the campus would not have been possible without the expertise and reliable technical assistance provided by Maria Tuominen, Minna Santanen, Mika Savisalo, Raili Salonen, Jukka Karhu, Ioan Iagar, and Annika Kankare. I would also like to thank Hanna Tuominen, Pia Tahvanainen, Katri Kulmala , Päivi Aalto, and Kristiina Nuutila for their secretarial assistance. Research in genomics is heavily dependent on the sequencing service providers, and I have been fortunate that we partnered with people who made my life easier. Thank you, Jukka Karhu & Dr. Juha-Pekka Pursiheimo at TCSL, Lars Paulin & Professor Petri Auvinen at BIDGEN, and Kristofer Hellberg & Sissy Strecker at GATC Genomics.

During the "*roller coaster of PhD*", I have been fortunate to be surrounded by the love and support of my Friends and Family. Two pillars of my strength, Mehrad Mahmmoudian and Vaibhav Prakash, I don't know what I have ever done to deserve your friendship and unwavering support. Mehrad, you've been there for, literally, everything significant in my life; armed with a barbecue tong and a beer when I was on cloud nine, with a drill and power tools when needed (surprisingly often), and more importantly to shoulder the heavy moments. Vaibhav, though we met in high school, from where I sit today it seems like we were kids. Thank you for the abundance of fond memories, and funny incidents. Being >5000 km away from family is tough, but Vaibhav you have stood as my proxy, you guard the stockpile of my embarrassing secrets, cover for me, give life-altering advice; you're basically a friend, confidant, guardian, and a therapist all rolled into one. Anna and Veera, thank you for bearing with me when I went on my unrestricted rants about #PhDLife, and on occasion sharing your coping mechanisms; Veera, thank you for helping me write the *tiivistelmä*; Fred, thank you for educating me about the Finnish culture and language, and for our discussions on global politics and finance; Klaus and Kari, for sharing your experiences and advice on "*How to adult?*". The *Desi* community in Turku was instrumental in keeping me from feeling homesick, especially, Ranjita (and Tuomas), Prasanna-Amruta, Nitin-Avlokita, Srikar-Sravani, Sarang-Parul, Madhukar, and Jismi. Shomya, Shailja and Gunjan, thank you for being in this "long-distance friendship" for over a decade now, and for being the reliably relatable ones.

ঠাকুমা (শ্রীমতী সাবিত্রী চক্রবর্তী) আর দিদা (শ্রীমতী ছন্দা গাঙ্গুলী), আপনাকে প্রণাম। তোমরা দুই জন আমাকে যা স্নেহ আর মমতা দিলে আমি জীবনভোর মনে রাখবো। শরীর খারাব হলে বুকে চটিয়ে, নিজের ঘুম খারাব করে কি করে ঘুম পাড়াতে আমাকে, এখন নিজের ছেলে হবার পড়ে বুঝলাম। তোমাদের কোলে মাথা রেখে জা শান্তি পাই ওটা পৃথিবীর কোনো বালিশে নেই। স্মৃতি নামের রেলগাড়িটা পেছন দিকে পালায়, দূরের থেকে শুধু ওই গাড়িটা আমাকে তোমার কাছে লোয়। I would not exist, quite literally, without my parents, Smt. Nivedita (Mumma) and Shri Nirmal (Papa). Papa, one of my earliest memories with you is you answering my annoying "why" & "how" questions about the world around us. You are the reason I found out about Science, despite not having a background in science, you were my first science teacher. Mumma, तुमसे ही सब सीखा, तुमसे ही सब पाया, तुम्हारे

चरणों में ही है सब समर्पित। हमेशा हम पे प्यार और स्नेह कि छाया रखने के लिए शुक्रिया। मेरे सोए हुए आठ (∞) और ख़राब लिखित हिंदी का बुरा ना मानने के लिए आपको नमन। Bhaiyu (Abhijeet), Kaku (Shri Purnendu), Kakima (Smt. Purnima), and Prakhar, thank you for your support, encouragement, and for being part of all of my lovely memories. सासुमाँ (श्रीमती बिंदु तिवारी) आपके निरंतर प्रोत्साहन के बिना हमारा सफल होना सम्भव नहीं था, इसके लिए हम सादेव आभारी रहेंगे।

On to the people with most substantial influence on my life, My Neil & Nikki :) Nikita, thank you for the endless love, affection, and support, before I got fascinated with Biotech, you were the one to drag me to the lectures in school. Your patience and perseverance of dealing with my regular rain-checks about weekend plans, and dedication to our family, really made it possible for me to pursue Science. Neil, you are and will always be the best thing that I have ever experienced. Coming home to your smile is what I look forward to every day. Thank you, Neil, for being my personal little *pomodoro timer*, that reminds me to take a break and have a little fun when I get glued to the keyboard.

Deepankar Chakroborty

9 References

Abbosh, C., Birkbak, N.J., Wilson, G.A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D.A., Veeriah, S., Rosenthal, R., Marafioti, T., Kirkizlar, E., Watkins, T.B.K., McGranahan, N., Ward, S., Martinson, L., Riley, J., Fraioli, F., Al Bakir, M., Grönroos, E., et al., 2017. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 545, 446–51. https://doi.org/10.1038/nature22364

Adam, M., Pogacic, V., Bendit, M., Chappuis, R., Nawijn, M.C., Duyster, J., Fox, C.J., Thompson, C.B., Cools, J., Schwaller, J., 2006. Targeting PIM Kinases Impairs Survival of Hematopoietic Cells Transformed by Kinase Inhibitor–Sensitive and Kinase Inhibitor–Resistant Forms of Fms-Like Tyrosine Kinase 3 and BCR/ABL. Cancer Res 66, 3828–35. https://doi.org/10.1158/0008-5472.CAN-05-2309

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nature methods 7, 248–49.

Aertgeerts, K., Skene, R., Yano, J., Sang, B.C., Zou, H., Snell, G., Jennings, A., Iwamoto, K., Habuka, N., Hirokawa, A., Ishikawa, T., Tanaka, T., Miki, H., Ohta, Y., Sogabe, S., 2011. Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein. Journal of Biological Chemistry 286, 18756–65. https://doi.org/10.1074/jbc.M110.206193

Aho, A.V., Kernighan, B.W., Weinberger, P.J., 1988. The AWK Programming Language, Addison-Wesley series in computer science and information processing. Addison-Wesley Publishing Company.

An, Z., Aksoy, O., Zheng, T., Fan, Q.-W., Weiss, W.A., 2018. Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. Oncogene 37, 1561–75. https://doi.org/10.1038/s41388-017-0045-7

Araten, D.J., Golde, D.W., Zhang, R.H., Thaler, H.T., Gargiulo, L., Notaro, R., Luzzatto, L., 2005. A Quantitative Measurement of the Human Somatic Mutation Rate. Cancer Research 65, 8111–17. https://doi.org/10.1158/0008-5472.CAN-04-1198

Atkin, N.B., Ross, A.J., 1960. Polyploidy in Human Tumours. Nature 187, 579–81. https://doi.org/10.1038/187579a0

Bacus, S.S., Chin, D., Yarden, Y., Zelnick, C.R., Stern, D.F., 1996. Type 1 receptor tyrosine kinases are differentially phosphorylated in mammary carcinoma and differentially associated with steroid receptors. The American journal of pathology 148, 549.

Balachandran, P., Beck, C.R., 2020. Structural variant identification and characterization. Chromosome Res 28, 31–47. https://doi.org/10.1007/s10577-019-09623-z

Beau-Faller, M., Prim, N., Ruppert, A.-M., Nanni-Metéllus, I., Lacave, R., Lacroix, L., Escande, F., Lizard, S., Pretet, J.-L., Rouquette, I., Crémoux, P. de, Solassol, J., Fraipont, F. de, Bièche, I., Cayre, A., Favre-Guillevin, E., Tomasini, P., Wislez, M., Besse, B., Legrain, M., et al., 2014. Rare EGFR exon 18 and exon 20 mutations in non-small-cell lung cancer on 10 117 patients: a multicentre observational study by the French ERMETIC-IFCT network. Annals of oncology 25, 126–31. https://doi.org/10.1093/annonc/mdt418

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57, 289–300.

Benjamini, Y., Krieger, A.M., Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93, 491–507. https://doi.org/10.1093/biomet/93.3.491

Berchuck, A., Heron, K.A., Carney, M.E., Lancaster, J.M., Fraser, E.G., Vinson, V.L., Deffenbaugh, A.M., Miron, A., Marks, J.R., Futreal, P.A., Frank, T.S., 1998. Frequency of germline and somatic BRCA1 mutations in ovarian cancer. Clinical Cancer Research 4, 2433–37.

Berger, A.H., Brooks, A.N., Wu, X., Shrestha, Y., Chouinard, C., Piccioni, F., Bagul, M., Kamburov, A., Imielinski, M., Hogstrom, L., Zhu, C., Yang, X., Pantel, S., Sakai, R., Watson, J., Kaplan, N., Campbell, J.D., Singh, S., Root, D.E., Narayan, R., et al., 2015. High-throughput Phenotyping of Lung Cancer Somatic Mutations. Cancer Cell 30, 214–28. https://doi.org/10.1016/j.ccell.2016.06.022

Bhandari, V., Hoey, C., Liu, L.Y., Lalonde, E., Ray, J., Livingstone, J., Lesurf, R., Shiah, Y.-J., Vujcic, T., Huang, X., Espiritu, S.M.G., Heisler, L.E., Yousif, F., Huang, V., Yamaguchi, T.N., Yao, C.Q., Sabelnykova, V.Y., Fraser, M., Chua, M.L.K., Kwast, T. van der, et al., 2019. Molecular landmarks of tumor hypoxia across cancer types. Nat Genet 51, 308–18. https://doi.org/10.1038/s41588-018-0318-2

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–20. https://doi.org/10.1093/bioinformatics/btu170

Bonner, J.A., Harari, P.M., Giralt, J., Azarnia, N., Shin, D.M., Cohen, R.B., Jones, C.U., Sur, R., Raben, D., Jassem, J., Ove, R., Kies, M.S., Baselga, J., Youssoufian, H., Amellal, N., Rowinsky, E.K., Ang, K.K., 2006. Radiotherapy plus Cetuximab for Squamous-Cell Carcinoma of the Head and Neck. N Engl J Med 354, 567–78. https://doi.org/10.1056/NEJMoa053422

Bradeen, H. a., Eide, C. a., O'Hare, T., Johnson, K.J., Willis, S.G., Lee, F.Y., Druker, B.J., Deininger, M.W., 2006. Comparison of imatinib mesylate, dasatinib (BMS-354825), and nilotinib (AMN107) in an N-ethyl-N-nitrosourea (ENU)-based mutagenesis screen: High efficacy of drug combinations. Blood 108, 2332–38. https://doi.org/10.1182/blood-2006-02-004580

Brewer, M.R., Yun, C.H., Lai, D., Lemmon, M.A., Eck, M.J., Pao, W., 2013. Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. Proceedings of the National Academy of Sciences of the United States of America 110, 3595–3604. https://doi.org/10.1073/pnas.1220050110

Bridges, C.B., 1922. The origin of variations in sexual and sex-limited characters. The American Naturalist 56, 51–63.

Brown, J.M., Wilson, W.R., 2004. Exploiting tumour hypoxia in cancer treatment. Nat Rev Cancer 4, 437–47. https://doi.org/10.1038/nrc1367

Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., Helleday, T., 2005. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature 434, 913–17. https://doi.org/10.1038/nature03443

Burrell, R.A., Swanton, C., 2014. The evolution of the unstable cancer genome. Current Opinion in Genetics and Development 24, 61–67. https://doi.org/10.1016/j.gde.2013.11.011

Byron, S.A., Chen, H., Wortmann, A., Loch, D., Gartside, M.G., Dehkhoda, F., Blais, S.P., Neubert, T.A., Mohammadi, M., Pollock, P.M., 2013. The N550K/H Mutations in FGFR2 Confer Differential Resistance to PD173074, Dovitinib, and Ponatinib ATP-Competitive Inhibitors. Neoplasia 15, 975-IN30. https://doi.org/10.1593/neo.121106

Cadwell, R.C., Joyce, G.F., 1994. Mutagenic PCR. Genome Research 3, S136–40.

Cagan, A., Baez-Ortega, A., Brzozowska, N., Abascal, F., Coorens, T.H.H., Sanders, M.A., Lawson, A.R.J., Harvey, L.M.R., Bhosle, S., Jones, D., Alcantara, R.E., Butler, T.M., Hooks, Y., Roberts, K., Anderson, E., Lunn, S., Flach, E., Spiro, S., Januszczak, I., Wrigglesworth, E., et al., 2022. Somatic mutation rates scale with lifespan across mammals. Nature 604, 517–24. https://doi.org/10.1038/s41586-022-04618-z

Campbell, P.J., Getz, G., Korbel, J.O., Stuart, J.M., Jennings, J.L., Stein, L.D., Perry, M.D., Nahal-Bose, H.K., Ouellette, B.F.F., Li, C.H., Rheinbay, E., Nielsen, G.P., Sgroi, D.C., Wu, C.L., Faquin, W.C., Deshpande, V., Boutros, P.C., Lazar, A.J., Hoadley, K.A., Louis, D.N., et al., 2020. Pan-cancer analysis of whole genomes. Nature 578, 82–93. https://doi.org/10.1038/s41586-020-1969-6

Canfield, K., Li, J., Wilkins, O.M., Morrison, M.M., Ung, M., Wells, W., Williams, C.R., Liby, K.T., Vullhorst, D., Buonanno, A., Hu, H., Schiff, R., Cook, R.S., Kurokawa, M., 2015. Receptor tyrosine kinase ERBB4 mediates acquired resistance to ERBB2 inhibitors in breast cancer cells. Cell Cycle 14, 648–55. https://doi.org/10.4161/15384101.2014.994966

Cargnello, M., Roux, P.P., 2011. Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases. Microbiology and Molecular Biology Reviews 75, 50–83. https://doi.org/10.1128/MMBR.00031-10

Carraway III, K.L., Weber, J.L., Unger, M.J., Ledesma, J., Yu, N., Gassmann, M., Lai, C., 1997. Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine kinases. Nature 387, 512–16. https://doi.org/10.1038/387512a0

Carraway, K.L., Sliwkowski, M.X., Akita, R., Platko, J.V., Guy, P.M., Nuijens, A., Diamonti, A.J., Vandlen, R.L., Cantley, L.C., Cerione, R.A., 1994. The erbB3 gene product is a receptor for heregulin. J Biol Chem 269, 14303–6.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N., 2012. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2, 401–4. https://doi.org/10.1158/2159-8290.CD-12-0095

Chakroborty, D., Kurppa, K.J., Paatero, I., Ojala, V.K., Koivu, M., Tamirat, M.Z., Koivunen, J.P., Jänne, P.A., Johnson, M.S., Elo, L.L., Elenius, K., 2019. An unbiased in vitro screen for activating epidermal growth factor receptor mutations. Journal of Biological Chemistry 294, 9377–89. https://doi.org/10.1074/jbc.ra118.006336

Chakroborty, D., Ojala, V.K., Knittle, A.M., Drexler, J., Tamirat, M.Z., Ruzicka, R., Bosch, K., Woertl, J., Schmittner, S., Elo, L.L., Johnson, M.S., Kurppa, K.J., Solca, F., Elenius, K., 2022. An unbiased functional genetics screen identifies rare activating ERBB4 mutations. Cancer Research Communications 2, 10–27. https://doi.org/10.1158/2767-9764.CRC-21-0021

Chang, H., Riese II, D.J., Gilbert, W., Stern, D.F., McMahan, U.J., 1997. Ligands for ErbB-family receptors encoded by a neuregulin-like gene. Nature 387, 509–12. https://doi.org/10.1038/387509a0

Chang, M.T., Asthana, S., Gao, S.P., Lee, B.H., Chapman, J.S., Kandoth, C., Gao, J., Socci, N.D., Solit, D.B., Olshen, A.B., Schultz, N., Taylor, B.S., 2016. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nature biotechnology 34, 1–11. https://doi.org/10.1038/nbt.3391

Chang, M.T., Bhattarai, T.S., Schram, A.M., Bielski, C.M., Donoghue, M.T.A., Jonsson, P., Chakravarty, D., Phillips, S., Kandoth, C., Penson, A., Gorelick, A., Shamu, T., Patel, S., Harris, C., Gao, J., Sumer, S.O., Kundra, R., Razavi, P., Li, B.T., Reales, D.N., et al., 2018. Accelerating Discovery of Functional Mutant Alleles in Cancer. Cancer Discov 8, 174–83. https://doi.org/10.1158/2159-8290.CD-17-0321

Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2021. shiny: Web Application Framework for R.

Chapman, P.B., Hauschild, A., Robert, C., Haanen, J.B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M., Hogg, D., Lorigan, P., Lebbe, C., Jouary, T., Schadendorf, D., Ribas, A., O'Day, S.J., Sosman, J. a, Kirkwood, J.M., Eggermont, A.M.M., et al., 2011. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. [WWW Document]. The New England journal of medicine. https://doi.org/10.1056/NEJMoa1103782

Chen, J., Williams, I.R., Lee, B.H., Duclos, N., Huntly, B.J.P., Donoghue, D.J., Gilliland, D.G., 2005. Constitutively activated FGFR3 mutants signal through PLCgamma-dependent and -independent pathways for hematopoietic transformation. Blood 106, 328–37. https://doi.org/10.1182/blood-2004-09-3686

Cheng, D.T., Mitchell, T.N., Zehir, A., Shah, R.H., Benayed, R., Syed, A., Chandramohan, R., Liu, Z.Y., Won, H.H., Scott, S.N., Rose Brannon, A., O'Reilly, C., Sadowska, J., Casanova, J., Yannes, A., Hechtman, J.F., Yao, J., Song, W., Ross, D.S., Oultache, A., et al., 2015. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): A hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. Journal of Molecular Diagnostics 17, 251–64. https://doi.org/10.1016/j.jmoldx.2014.12.006

Cho, H.-S., Leahy, D.J., 2002. Structure of the extracellular region of HER3 reveals an interdomain tether. Science 297, 1330–33.

Cho, H.-S., Mason, K., Ramyar, K.X., Stanley, A.M., Gabelli, S.B., Denney, D.W., Leahy, D.J., 2003. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature 421, 756–60. https://doi.org/10.1038/nature01392

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P., 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE 7, e46688. https://doi.org/10.1371/journal.pone.0046688

Cline, J., Hogrefe, H.H., 2000. Randomize gene sequences with new PCR mutagenesis kit. Strategies 13, 157–61.

Cobleigh, M.A., Vogel, C.L., Tripathy, D., Robert, N.J., Scholl, S., Fehrenbacher, L., Wolter, J.M., Paton, V., Shak, S., Lieberman, G., others, 1999. Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. Journal of clinical Oncology 17, 2639–2639.

Cohen, S., 1962. Isolation of a Mouse Submaxillary Gland Protein Accelerating Incisor Eruption and Eyelid Opening in the New-born Animal. Journal of Biological Chemistry 237, 1555–62. https://doi.org/10.1016/S0021-9258(19)83739-0

Coussens, L., Yang-Feng, T., Liao, Y., Chen, E., Gray, A., McGrath, J., Seeburg, P., Libermann, T., Schlessinger, J., Francke, U., et, al., 1985. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. Science 230, 1132–39. https://doi.org/10.1126/science.2999974

Cox, E.C., 1976. BACTERIAL MUTATOR GENES AND THE CONTROL OF SPONTANEOUS MUTATION. Annu. Rev. Genet. 10, 135–56. https://doi.org/10.1146/annurev.ge.10.120176.001031

Cross, D.A.E., Ashton, S.E., Ghiorghiu, S., Eberlein, C., Nebhan, C.A., Spitzler, P.J., Orme, J.P., Finlay, M.R.V., Ward, R.A., Mellor, M.J., Hughes, G., Rahi, A., Jacobs, V.N., Brewer, M.R., Ichihara, E., Sun, J., Jin, H., Ballard, P., Al-Kadhimi, K., Rowlinson, R., et al., 2014. AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. Cancer Discovery 4, 1046–61. https://doi.org/10.1158/2159-8290.CD-14-0337

Cunningham, B.C., Wells, J.A., 1989. High-Resolution Epitope Mapping of hGH-Receptor Interactions by Alanine-Scanning Mutagenesis. Science 244, 1081–85. https://doi.org/10.1126/science.2471267

Cunningham, D., Humblet, Y., Siena, S., Khayat, D., Bleiberg, H., Santoro, A., Bets, D., Mueser, M., Harstrick, A., Verslype, C., others, 2004. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. New England journal of medicine 351, 337–45.

Dagogo-Jack, I., Shaw, A.T., 2018. Tumour heterogeneity and resistance to cancer therapies. Nature Reviews Clinical Oncology 15, 81–94. https://doi.org/10.1038/nrclinonc.2017.166

Daley, G.Q., Baltimore, D., 1988. Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P210bcr/abl protein. Proceedings of the National Academy of Sciences of the United States of America 85, 9312–16. https://doi.org/10.1073/pnas.85.23.9312

Davis, M.I., Hunt, J.P., Herrgard, S., Ciceri, P., Wodicka, L.M., Pallares, G., Hocker, M., Treiber, D.K., Zarrinkar, P.P., 2011. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol 29, 1046–51. https://doi.org/10.1038/nbt.1990

Delignette-Muller, M.L., Dutang, C., 2015. fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software 64, 1–34.

Derbyshire, K.M., Salvo, J.J., Grindley, N.D.F., 1986. A simple and efficient procedure for saturation mutagenesis using mixed oligodeoxynucleotides. Gene 46, 145–52. https://doi.org/10.1016/0378-1119(86)90398-7

Didion, J.P., Buus, R.J., Naghashfar, Z., Threadgill, D.W., Morse, H.C., Villena, F.P.-M. de, 2014. SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy. BMC Genomics 15, 847. https://doi.org/10.1186/1471-2164-15-847 1471-2164-15-847 [pii]

Dienstmann, R., Jang, I.S., Bot, B., Friend, S., Guinney, J., 2015. Database of genomic biomarkers for cancer drugs and clinical targetability in solid tumors. Cancer Discovery 5, 118–23. https://doi.org/10.1158/2159-8290.CD-14-1118

Dill, K.A., Ozkan, S.B., Shell, M.S., Weikl, T.R., 2008. The Protein Folding Problem. Annu. Rev. Biophys. 37, 289–316. https://doi.org/10.1146/annurev.biophys.37.092707.153558

Donoghue, J., Kerr, L., Alexander, N., Greenall, S., Longano, A., Gottardo, N., Wang, R., Tabar, V., Adams, T., Mischel, P., Johns, T., 2018. Activation of ERBB4 in Glioblastoma Can Contribute to Increased Tumorigenicity and Influence Therapeutic Response. Cancers 10, 243. https://doi.org/10.3390/cancers10080243

Dowle, M., Srinivasan, A., 2021. data.table: Extension of `data.frame`.

Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J., Waterfield, M.D., 1984. Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. Nature 307, 521–27. https://doi.org/10.1038/307521a0

Duensing, A., Medeiros, F., McConarty, B., Joseph, N.E., Panigrahy, D., Singer, S., Fletcher, C.D., Demetri, G.D., Fletcher, J.A., 2004. Mechanisms of oncogenic KIT signal transduction in primary gastrointestinal stromal tumors (GISTs). Oncogene 23, 3999–4006.

Elamin, Y.Y., Robichaux, J.P., Carter, B.W., Altan, M., Gibbons, D.L., Fossella, F.V., Lam, V.K., Patel, A.B., Negrao, M.V., Le, X., Mott, F.E., Zhang, J., Feng, L., Blumenschein, G., Tsao, A.S., Heymach, J.V., 2022. Poziotinib for Patients With HER2 Exon 20 Mutant Non–Small-Cell Lung Cancer: Results From a Phase II Trial. Journal of Clinical Oncology 40, 702–9. https://doi.org/10.1200/JCO.21.01113

Elenius, K., Choi, C.J., Paul, S., Santiestevan, E., Nishi, E., Klagsbrun, M., 1999. Characterization of a naturally occurring ErbB4 isoform that does not bind or activate phosphatidyl inositol 3-kinase. Oncogene 18, 2607–15. https://doi.org/10.1038/sj.onc.1202612

Elenius, K., Corfas, G., Paul, S., Choi, C.J., Rio, C., Plowman, G.D., Klagsbrun, M., 1997a. A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific tissue distribution and differential processing in response to phorbol ester. Journal of Biological Chemistry 272, 26761–68. https://doi.org/10.1074/jbc.272.42.26761

Elenius, K., Paul, S., Allison, G., Sun, J., Klagsbrun, M., 1997b. Activation of HER4 by heparin-binding EGF-like growth factor stimulates chemotaxis but not proliferation. EMBO J 16, 1268–78. https://doi.org/10.1093/emboj/16.6.1268

Elliott, K., Larsson, E., 2021. Non-coding driver mutations in human cancer. Nature Reviews Cancer 21, 500–509.

Endres, N.F., Das, R., Smith, A.W., Arkhipov, A., Kovacs, E., Huang, Y., Pelton, J.G., Shan, Y., Shaw, D.E., Wemmer, D.E., Groves, J.T., Kuriyan, J., 2013. Conformational Coupling across the Plasma Membrane in Activation of the EGF Receptor. Cell 152, 543–56. https://doi.org/10.1016/j.cell.2012.12.032

Ercan, D., Choi, H.G., Yun, C.H., Capelletti, M., Xie, T., Eck, M.J., Gray, N.S., Jänne, P.A., 2015. EGFR mutations and resistance to irreversible pyrimidine-based EGFR inhibitors. Clinical Cancer Research 21, 3913–23. https://doi.org/10.1158/1078-0432.CCR-14-2789

Ernst, C., Hahnen, E., Engel, C., Nothnagel, M., Weber, J., Schmutzler, R.K., Hauke, J., 2018. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. BMC Med Genomics 11, 35. https://doi.org/10.1186/s12920-018-0353-y

Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., Martin, N.M.B., Jackson, S.P., Smith, G.C.M., Ashworth, A., 2005.

Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 434, 917–21. https://doi.org/10.1038/nature03445

Farrugia, D.J., Agarwal, M.K., Pankratz, V.S., Deffenbaugh, A.M., Pruss, D., Frye, C., Wadum, L., Johnson, K., Mentlick, J., Tavtigian, S.V., others, 2008. Functional assays for classification of BRCA2 variants of uncertain significance. Cancer research 68, 3523–31.

Fearon, E.R., Hamilton, S.R., Vogelstein, B., 1987. Clonal Analysis of Human Colorectal Tumors. Science 238, 193–97. https://doi.org/10.1126/science.2889267

Ferretti, E., Di Marcotullio, L., Gessi, M., Mattei, T., Greco, A., Po, A., De Smaele, E., Giangaspero, F., Riccardi, R., Di Rocco, C., Pazzaglia, S., Maroder, M., Alimandi, M., Screpanti, I., Gulino, A., 2006. Alternative splicing of the ErbB-4 cytoplasmic domain and its regulation by hedgehog signaling identify distinct medulloblastoma subsets. Oncogene 25, 7267–73. https://doi.org/10.1038/sj.onc.1209716

Fidler, I.J., 1978. Tumor heterogeneity and the biology of cancer invasion and metastasis. Cancer research 38, 2651–60.

Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C., Shendure, J., 2014. Saturation editing of genomic regions by multiplex homology-directed repair. Nature 513, 120–23. https://doi.org/10.1038/nature13695

Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., Shendure, J., 2018. Accurate classification of BRCA1 variants with saturation genome editing. Nature 562, 217–22. https://doi.org/10.1038/s41586-018-0461-z

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., Stefancsik, R., Harsha, B., Kok, C.Y., Jia, M., Jubb, H., Sondka, Z., Thompson, S., De, T., Campbell, P.J., 2017. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Research 45, D777–83. https://doi.org/10.1093/nar/gkw1121

Ford, D., Easton, D.F., Peto, J., 1995. Estimates of the gene frequency of BRCA1 and its contribution to breast and ovarian cancer incidence. American journal of human genetics 57, 1457.

Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., Fields, S., 2010. High-resolution mapping of protein sequence-function relationships. Nat Methods 7, 741–46. https://doi.org/10.1038/nmeth.1492

Fowler, D.M., Fields, S., 2014. Deep mutational scanning: a new style of protein science. Nat Methods 11, 801–7. https://doi.org/10.1038/nmeth.3027

Freed, D.M., Bessman, N.J., Kiyatkin, A., Salazar-Cavazos, E., Byrne, P.O., Moore, J.O., Valley, C.C., Ferguson, K.M., Leahy, D.J., Lidke, D.S., Lemmon, M.A., 2017. EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. Cell 171, 683-695.e18. https://doi.org/10.1016/j.cell.2017.09.017

Fruman, D.A., Rommel, C., 2014. PI3K and cancer: lessons, challenges and opportunities. Nature reviews Drug discovery 13, 140–56.

Fuse, M.J., Okada, K., Oh-hara, T., Ogura, H., Fujita, N., Katayama, R., 2017. Mechanisms of resistance to NTRK inhibitors and therapeutic strategies in NTRK1-rearranged cancers. Molecular Cancer Therapeutics 16, 2130–43. https://doi.org/10.1158/1535-7163.MCT-16-0909

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., Schultz, N., 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling 6, pl1. https://doi.org/10.1126/scisignal.2004088

Garraway, L. a, Lander, E.S., 2013. Lessons from the cancer genome. Cell 153, 17–37. https://doi.org/10.1016/j.cell.2013.03.002

Garrett, T.P., McKern, N.M., Lou, M., Elleman, T.C., Adams, T.E., Lovrecz, G.O., Zhu, H.-J., Walker, F., Frenkel, M.J., Hoyne, P.A., others, 2002. Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α. Cell 110, 763–73.

Garrett, T.P.J., McKern, N.M., Lou, M., Elleman, T.C., Adams, T.E., Lovrecz, G.O., Kofler, M., Jorissen, R.N., Nice, E.C., Burgess, A.W., Ward, C.W., 2003. The Crystal Structure of a Truncated ErbB2

Ectodomain Reveals an Active Conformation, Poised to Interact with Other ErbB Receptors. Molecular Cell 11, 495–505. https://doi.org/10.1016/S1097-2765(03)00048-0

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., Liu, D.R., 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551, 464–71.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N.Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C.R., Nohadani, M., et al., 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. New England Journal of Medicine 366, 883–92. https://doi.org/10.1056/NEJMoa1113205

Gerstung, M., Jolly, C., Leshchiner, I., Dentro, S.C., 2020. The evolutionary history of 2,658 cancers. Nature 578.

Geyer, C.E., Forster, J., Lindquist, D., Chan, S., Romieu, C.G., Pienkowski, T., Jagiello-Gruszfeld, A., Crown, J., Chan, A., Kaufman, B., Skarlos, D., Campone, M., Davidson, N., Berger, M., Oliva, C., Rubin, S.D., Stein, S., Cameron, D., 2006. Lapatinib plus Capecitabine for HER2-Positive Advanced Breast Cancer. N Engl J Med 355, 2733–43. https://doi.org/10.1056/NEJMoa064320

Girdler, F., Sessa, F., Patercoli, S., Villa, F., Musacchio, A., Taylor, S., 2008. Molecular Basis of Drug Resistance in Aurora Kinases. Chemistry & Biology 15, 552–62. https://doi.org/10.1016/j.chembiol.2008.04.013

Glaab, W.E., Tindall, K.R., 1997. Mutation rate at the hprt locus in human cancer cell lines with specific mismatch repair-gene defects. Carcinogenesis 18, 1–8.

Gorelick, A.N., Sánchez-Rivera, F.J., Cai, Y., Bielski, C.M., Biederstedt, E., Jonsson, P., Richards, A.L., Vasan, N., Penson, A.V., Friedman, N.D., Ho, Y.-J., Baslan, T., Bandlamudi, C., Scaltriti, M., Schultz, N., Lowe, S.W., Reznik, E., Taylor, B.S., 2020. Phase and context shape the function of composite oncogenic mutations. Nature 582, 100–103. https://doi.org/10.1038/s41586-020-2315-8

Greaves, M., Maley, C.C., 2012. Clonal evolution in cancer. Nature 481, 306–13. https://doi.org/10.1038/nature10762

Greaves, W.O., Verma, S., Patel, K.P., Davies, M.A., Barkoh, B.A., Galbincea, J.M., Yao, H., Lazar, A.J., Aldape, K.D., Medeiros, L.J., Luthra, R., 2013. Frequency and Spectrum of BRAF Mutations in a Retrospective, Single-Institution Study of 1112 Cases of Melanoma. The Journal of Molecular Diagnostics 15, 220–26. https://doi.org/10.1016/j.jmoldx.2012.10.002

Greener, A., Callahan, M., 1994. XL1-Red: a highly efficient random mutagenesis strain. Strategies 7, 32–34.

Greulich, H., Chen, T.H., Feng, W., Jänne, P.A., Alvarez, J.V., Zappaterra, M., Bulmer, S.E., Frank, D.A., Hahn, W.C., Sellers, W.R., Meyerson, M., 2005. Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. PLoS Medicine 2, 1167–76. https://doi.org/10.1371/journal.pmed.0020313

Greulich, H., Kaplan, B., Mertins, P., Chen, T.-H.T., Tanaka, K.E., Yun, C.-H.C., Zhang, X., Lee, S.-H., Cho, J., Ambrogio, L., Liao, R., Imielinski, M., Banerji, S., Berger, A.H., Lawrence, M.S., Zhang, J., Pho, N.H., Walker, S.R., Winckler, W., Getz, G., et al., 2012. Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of ERBB2. Proceedings of the National Academy of Sciences of the United States of America 109, 14476–81. https://doi.org/10.1073/pnas.1203201109

Gruet, A., Longhi, S., Bignon, C., 2012. One-step generation of error-prone PCR libraries using Gateway® technology. Microbial Cell Factories 11, 14. https://doi.org/10.1186/1475-2859-11-14

Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L.B., Tubio, J.M.C., Papaemmanuil, E., Brewer, D.S., Kallio, H.M.L., Högnäs, G., Annala, M., Kivinummi, K., Goody, V., Latimer, C., O'Meara, S., Dawson, K.J., Isaacs, W., Emmert-Buck, M.R., Nykter, M., Foster, C., Kote-Jarai, Z., et al., 2015. The evolutionary history of lethal metastatic prostate cancer. Nature 520, 353–57. https://doi.org/10.1038/nature14347

Guy, P.M., Platko, J.V., Cantley, L.C., Cerione, R.A., Carraway, K.L., 1994. Insect cell-expressed p180erbB3 possesses an impaired tyrosine kinase activity. Proceedings of the National Academy of Sciences of the United States of America 91, 8132–36. https://doi.org/10.1073/pnas.91.17.8132

Haigis, K.M., Cichowski, K., Elledge, S.J., 2019. Tissue-specificity in cancer: The rule, not the exception. Science 363, 1150–51. https://doi.org/10.1126/science.aaw3472

Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A., Feeley, M.N., Sanson, K.R., Baidi, Y., Koblan, L.W., Liu, D.R., Neal, J.T., Doench, J.G., 2021. Massively parallel assessment of human variants with base editor screens. Cell 184, 1064-1080.e20. https://doi.org/10.1016/j.cell.2021.01.012

Harari, D., Tzahar, E., Romano, J., Shelly, M., Pierce, J., Andrews, G., Yarden, Y., 1999. Neuregulin-4: a novel growth factor that acts through the ErbB-4 receptor tyrosine kinase. Oncogene 18, 2681–89. https://doi.org/10.1038/sj.onc.1202631

Hegde, G.V., Cruz, C.C. de la, Chiu, C., Alag, N., Schaefer, G., Crocker, L., Ross, S., Goldenberg, D., Merchant, M., Tien, J., Shao, L., Roth, L., Tsai, S.-P., Stawicki, S., Jin, Z., Wyatt, S.K., Carano, R.A.D., Zheng, Y., Sweet-Cordero, E.A., Wu, Y., et al., 2013. Blocking NRG1 and Other Ligand-Mediated Her4 Signaling Enhances the Magnitude and Duration of the Chemotherapeutic Response of Non-Small Cell Lung Cancer. Science Translational Medicine 5, 171ra18-171ra18. https://doi.org/10.1126/scitranslmed.3004438

Hietpas, R., Roscoe, B., Jiang, L., Bolon, D.N.A., 2012. Fitness analyses of all possible point mutations for regions of genes in yeast. Nat Protoc 7, 1382–96. https://doi.org/10.1038/nprot.2012.069

Hollmén, M., Elenius, K., 2010. Potential of ErbB4 antibodies for cancer therapy. Future Oncology 6, 37–53. https://doi.org/10.2217/fon.09.144

Hong, D.S., Fakih, M.G., Strickler, J.H., Desai, J., Durm, G.A., Shapiro, G.I., Falchook, G.S., Price, T.J., Sacher, A., Denlinger, C.S., Bang, Y.-J., Dy, G.K., Krauss, J.C., Kuboki, Y., Kuo, J.C., Coveler, A.L., Park, K., Kim, T.W., Barlesi, F., Munster, P.N., et al., 2020. KRAS G12C Inhibition with Sotorasib in Advanced Solid Tumors. N Engl J Med 383, 1207–17. https://doi.org/10.1056/NEJMoa1917239

Hudziak, R.M., Lewis, G.D., Winget, M., Fendly, B.M., Shepard, H.M., Ullrich, A., 1989. p185HER2 monoclonal antibody has antiproliferative effects in vitro and sensitizes human breast tumor cells to tumor necrosis factor. Molecular and cellular biology 9, 1165–72.

Huxley, J., 1958. Biological aspects of cancer. George Allen & Unwin.

Hyman, D.M., Piha-Paul, S.A., Won, H., Rodon, J., Saura, C., Shapiro, G.I., Juric, D., Quinn, D.I., Moreno, V., Doger, B., Mayer, I.A., Boni, V., Calvo, E., Loi, S., Lockhart, A.C., Erinjeri, J.P., Scaltriti, M., Ulaner, G.A., Patel, J., Tang, J., et al., 2018. HER kinase inhibition in patients with HER2- and HER3-mutant cancers. Nature. https://doi.org/10.1038/nature25475

Hyman, D.M., Taylor, B.S., Baselga, J., 2017. Implementing Genome-Driven Oncology. Cell 168, 584–99. https://doi.org/10.1016/j.cell.2016.12.015

Hynes, N.E., 2016. ErbB2: From an EGFR Relative to a Central Target for Cancer Therapy. Cancer Res 76, 3659–62. https://doi.org/10.1158/0008-5472.CAN-16-1356

Ikegami, M., Kohsaka, S., Ueno, T., Momozawa, Y., Inoue, S., Tamura, K., Shimomura, A., Hosoya, N., Kobayashi, H., Tanaka, S., Mano, H., 2020. High-throughput functional evaluation of BRCA2 variants of unknown significance. Nat Commun 11, 2573. https://doi.org/10.1038/s41467-020-16141-8

Jain, P.C., Varadarajan, R., 2014. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. Analytical Biochemistry 449, 90–98. https://doi.org/10.1016/j.ab.2013.12.002

Jaiswal, B.S., Kljavin, N.M., Stawiski, E.W., Chan, E., Parikh, C., Durinck, S., Chaudhuri, S., Pujara, K., Guillory, J., Edgar, K.A., Janakiraman, V., Scholz, R.-P., Bowman, K.K., Lorenzo, M., Li, H., Wu, J., Yuan, W., Peters, B.A., Kan, Z., Stinson, J., et al., 2013. Oncogenic ERBB3 mutations in human cancers. Cancer cell 23, 603–17. https://doi.org/10.1016/j.ccr.2013.04.012

Jamal-Hanjani, M., Wilson, G.A., McGranahan, N., Birkbak, N.J., Watkins, T.B.K., Veeriah, S., Shafi, S., Johnson, D.H., Mitter, R., Rosenthal, R., Salm, M., Horswell, S., Escudero, M., Matthews, N.,

Rowan, A., Chambers, T., Moore, D.A., Turajlic, S., Xu, H., Lee, S.-M., et al., 2017. Tracking the Evolution of Non–Small-Cell Lung Cancer. New England Journal of Medicine 376, 2109–21. https://doi.org/10.1056/NEJMoa1616288

Jameson, J.L., Longo, D.L., 2015. Precision Medicine — Personalized, Problematic, and Promising. N Engl J Med 372, 2229–34. https://doi.org/10.1056/NEJMsb1503104

Jiang, J., Greulich, H., Jänne, P.A., Sellers, W.R., Meyerson, M., Griffin, J.D., 2005. Epidermal growth factor-independent transformation of Ba/F3 cells with cancer-derived epidermal growth factor receptor mutants induces gefitinib-sensitive cell cycle progression. Cancer Research 65, 8968–74. https://doi.org/10.1158/0008-5472.CAN-05-1829

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., et al., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–89. https://doi.org/10.1038/s41586-021-03819-2

Junttila, T.T., Sundvall, M., Lundin, M., Lundin, J., Tanner, M., Härkönen, P., Joensuu, H., Isola, J., Elenius, K., 2005. Cleavable ErbB4 Isoform in Estrogen Receptor–Regulated Growth of Breast Cancer Cells. Cancer Res 65, 1384–93. https://doi.org/10.1158/0008-5472.CAN-04-3150

Junttila, T.T., Sundvall, M., Määttä, J.A., Elenius, K., 2000. ErbB4 and Its Isoforms Selective Regulation of Growth Factor Responses by Naturally Occurring Receptor Variants. Trends in Cardiovascular Medicine 10, 304–10. https://doi.org/10.1016/S1050-1738(01)00065-2

Jura, N., Endres, N.F., Engel, K., Deindl, S., Das, R., Lamers, M.H., Wemmer, D.E., Zhang, X., Kuriyan, J., 2009a. Mechanism for Activation of the EGF Receptor Catalytic Domain by the Juxtamembrane Segment. Cell 137, 1293–1307. https://doi.org/10.1016/j.cell.2009.04.025

Jura, N., Shan, Y., Cao, X., Shaw, D.E., Kuriyan, J., 2009b. Structural analysis of the catalytically inactive kinase domain of the human EGF receptor 3. Proc. Natl. Acad. Sci. U.S.A. 106, 21608–13. https://doi.org/10.1073/pnas.0912101106

Jura, N., Zhang, X., Endres, N.F., Seeliger, M.A., Schindler, T., Kuriyan, J., 2011. Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms. Molecular Cell 42, 9–22. https://doi.org/10.1016/j.molcel.2011.03.004

Kaelin, W.G., 2005. The Concept of Synthetic Lethality in the Context of Anticancer Therapy. Nat Rev Cancer 5, 689–98. https://doi.org/10.1038/nrc1691

Kantarjian, H., Sawyers, C., Hochhaus, A., Guilhot, F., Schiffer, C., Gambacorti-Passerini, C., Niederwieser, D., Resta, D., Capdeville, R., Zoellner, U., others, 2002. Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. New england journal of medicine 346, 645–52.

Kew, T.Y., Bell, J.A., Pinder, S.E., Denley, H., Srinivasan, R., Gullick, W.J., Nicholson, R.I., Blamey, R.W., Ellis, I.O., 2000. c-erbB-4 protein expression in human breast cancer. British Journal of Cancer 82, 1163–70. https://doi.org/10.1054/bjoc.1999.1057

Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M.A., Gerstein, M., 2016. Role of non-coding sequence variants in cancer. Nature Reviews Genetics 17, 93–108.

Kiavue, N., Cabel, L., Melaabi, S., Bataillon, G., Callens, C., Lerebours, F., Pierga, J.-Y., Bidard, F.-C., 2020. ERBB3 mutations in cancer: biological aspects, prevalence and therapeutics. Oncogene 39, 487–502. https://doi.org/10.1038/s41388-019-1001-5

Kim, J.-Y., Jung, H.H., Do, I.-G., Bae, S., Lee, S.K., Kim, S.W., Lee, J.E., Nam, S.J., Ahn, J.S., Park, Y.H., Im, Y.-H., 2016. Prognostic value of ERBB4 expression in patients with triple negative breast cancer. BMC Cancer 16, 138. https://doi.org/10.1186/s12885-016-2195-3

Kim, Y., Ko, J., Cui, Z., Abolhoda, A., Ahn, J.S., Ou, S.-H., Ahn, M.-J., Park, K., 2012. The EGFR T790M Mutation in Acquired Resistance to an Irreversible Second-Generation EGFR Inhibitor. Molecular Cancer Therapeutics 11, 784–91. https://doi.org/10.1158/1535-7163.MCT-11-0750

Kinzler, K.W., Vogelstein, B., 1997. Gatekeepers and caretakers. Nature 386, 761–63.

Kitzman, J.O., Starita, L.M., Lo, R.S., Fields, S., Shendure, J., 2015. Massively parallel single-amino-acid mutagenesis. Nat Methods 12, 203–6. https://doi.org/10.1038/nmeth.3223

Knowlden, J.M., Gee, J.M.W., Seery, L., Farrow, L., Gullick, W., Ellis, I., Blamey, R., Robertson, J., Nicholson, R.I., 1998. c-erbB3 and c-erbB4 expression is a feature of the endocrine responsive phenotype in clinical breast cancer. Oncogene 17, 1949–57.

Kobayashi, S., Boggon, T.J., Dayaram, T., Jänne, P.A., Kocher, O., Meyerson, M., Johnson, B.E., Eck, M.J., Tenen, D.G., Halmos, B., 2005. EGFR Mutation and Resistance of Non–Small-Cell Lung Cancer to Gefitinib. New England Journal of Medicine 352, 786–92. https://doi.org/10.1056/NEJMoa044238

Kobayashi, Y., Azuma, K., Nagai, H., Kim, Y.H., Togashi, Y., Sesumi, Y., Chiba, M., Shimoji, M., Sato, K., Tomizawa, K., Takemoto, T., Nishio, K., Mitsudomi, T., 2017. Characterization of EGFR T790M, L792F, and C797S Mutations as Mechanisms of Acquired Resistance to Afatinib in Lung Cancer. Molecular Cancer Therapeutics 16, 357–64. https://doi.org/10.1158/1535-7163.MCT-16-0407

Kobayashi, Y., Mitsudomi, T., 2016. Not all EGFR mutations in lung cancer are created equal: Perspectives for individualized treatment strategy. Cancer Sci 107, 1179–86. https://doi.org/10.1111/cas.12996

Kobayashi, Y., Togashi, Y., Yatabe, Y., Mizuuchi, H., Jangchul, P., Kondo, C., Shimoji, M., Sato, K., Suda, K., Tomizawa, K., Takemoto, T., Hida, T., Nishio, K., Mitsudomi, T., 2015. EGFR Exon 18 Mutations in Lung Cancer: Molecular Predictors of Augmented Sensitivity to Afatinib or Neratinib as Compared with First- or Third-Generation TKIs. Clinical Cancer Research 21, 5305–13. https://doi.org/10.1158/1078-0432.CCR-15-1046

Kohsaka, S., Nagano, M., Ueno, T., Suehara, Y., Hayashi, T., Shimada, N., Takahashi, K., Suzuki, K., Takamochi, K., Takahashi, F., Mano, H., 2017. A method of high-Throughput functional evaluation of EGFR gene variants of unknown significance in cancer. Science Translational Medicine 9, eaan6566. https://doi.org/10.1126/scitranslmed.aan6566

Komuro, A., Nagai, M., Navin, N.E., Sudol, M., 2003. WW Domain-containing Protein YAP Associates with ErbB-4 and Acts as a Co-transcriptional Activator for the Carboxyl-terminal Fragment of ErbB-4 That Translocates to the Nucleus *. Journal of Biological Chemistry 278, 33334–41. https://doi.org/10.1074/jbc.M305597200

Korona, D.A., LeCompte, K.G., Pursell, Z.F., 2011. The high fidelity and unique error signature of human DNA polymerase. Nucleic Acids Research 39, 1763–73. https://doi.org/10.1093/nar/gkq1034

Kosaka, T., Tanizaki, J., Paranal, R., Endoh, H., Lydon, C., Capelletti, M., Repellin, C.E., Choi, J., Ogino, A., Calles, A., Ercan, D., Redig, A., Bahcall, M., Oxnard, G.R., Eck, M., Janne, P.A., 2017. Response heterogeneity of EGFR and HER2 exon 20 insertions to covalent EGFR and HER2 inhibitors. Cancer Research 2, canres.3404.2016. https://doi.org/10.1158/0008-5472.CAN-16-3404

Kraus, M.H., Issing, W., Miki, T., Popescu, N.C., Aaronson, S.A., 1989. Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family: evidence for overexpression in a subset of human mammary tumors. Proceedings of the National Academy of Sciences of the United States of America 86, 9193–97. https://doi.org/10.1073/pnas.86.23.9193

Kurppa, K.J., Denessiouk, K., Johnson, M.S., Elenius, K., 2016. Activating ERBB4 mutations in non-small cell lung cancer. Oncogene 35, 1283–91. https://doi.org/10.1038/onc.2015.185

Kwan, K.M., Fujimoto, E., Grabher, C., Mangum, B.D., Hardy, M.E., Campbell, D.S., Parant, J.M., Yost, H.J., Kanki, J.P., Chien, C.-B., 2007. The Tol2kit: A multisite gateway-based construction kit forTol2 transposon transgenesis constructs. Developmental Dynamics 236, 3088–99. https://doi.org/10.1002/dvdy.21343

Le, X., Cornelissen, R., Garassino, M., Clarke, J.M., Tchekmedyian, N., Goldman, J.W., Leu, S.-Y., Bhat, G., Lebel, F., Heymach, J.V., Socinski, M.A., 2022. Poziotinib in Non–Small-Cell Lung Cancer Harboring HER2 Exon 20 Insertion Mutations After Prior Therapies: ZENITH20-2 Trial. Journal of Clinical Oncology 40, 710–18. https://doi.org/10.1200/JCO.21.01323

Lee, J.C., Vivanco, I., Beroukhim, R., Huang, J.H.Y., Feng, W.L., DeBiasi, R.M., Yoshimoto, K., King, J.C., Nghiemphu, P., Yuza, Y., Xu, Q., Greulich, H., Thomas, R.K., Paez, J.G., Peck, T.C., Linhart, D.J., Glatt, K.A., Getz, G., Onofrio, R., Ziaugra, L., et al., 2006. Epidermal Growth Factor Receptor

Activation in Glioblastoma through Novel Missense Mutations in the Extracellular Domain. PLoS Med 3, e485. https://doi.org/10.1371/journal.pmed.0030485

Lemmon, M. a, Schlessinger, J., 2010. Cell signaling by receptor tyrosine kinases. Cell 141, 1117–34. https://doi.org/10.1016/j.cell.2010.06.011

Lemmon, M.A., Schlessinger, J., Ferguson, K.M., 2014. The EGFR Family: Not So Prototypical Receptor Tyrosine Kinases. Cold Spring Harbor Perspectives in Biology 6, a020768–a020768. https://doi.org/10.1101/cshperspect.a020768

Li, D., Ambrogio, L., Shimamura, T., Kubo, S., Takahashi, M., Chirieac, L.R., Padera, R.F., Shapiro, G.I., Baum, A., Himmelsbach, F., Rettig, W.J., Meyerson, M., Solca, F., Greulich, H., Wong, K.-K., 2008. BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models. Oncogene 27, 4702–11. https://doi.org/10.1038/onc.2008.109

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN] 1–3. https://doi.org/10.48550/arXiv.1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–79. https://doi.org/10.1093/bioinformatics/btp352

Libermann, T.A., Nusbaum, H.R., Razon, N., Kris, R., Lax, I., Soreq, H., Whittle, N., Waterfield, M.D., Ullrich, A., Schlessinger, J., 1985. Amplification and Overexpression of the EGF Receptor Gene in Primary Human Glioblastomas. Journal of Cell Science 1985, 161–72. https://doi.org/10.1242/jcs.1985.Supplement_3.16

Lievre, A., Bachet, J.-B., Le Corre, D., Boige, V., Landi, B., Emile, J.-F., Côté, J.-F., Tomasic, G., Penna, C., Ducreux, M., others, 2006. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. Cancer research 66, 3992–95.

Linder, D., Gartler, S.M., 1965. Glucose-6-Phosphate Dehydrogenase Mosaicism: Utilization as a Cell Marker in the Study of Leiomyomas. Science 150, 67–69. https://doi.org/10.1126/science.150.3692.67

Liu, P., Cleveland, T.E., Bouyain, S., Byrne, P.O., Longo, P.A., Leahy, D.J., 2012. A single ligand is sufficient to activate EGFR dimers. Proc. Natl. Acad. Sci. U.S.A. 109, 10861–66. https://doi.org/10.1073/pnas.1201114109

Loeb, L.A., Monnat, R.J., 2008. DNA polymerases and human disease. Nat Rev Genet 9, 594–604. https://doi.org/10.1038/nrg2345

Lopez-Chavez, A., Thomas, A., Rajan, A., Raffeld, M., Morrow, B., Kelly, R., Carter, C.A., Guha, U., Killian, K., Lau, C.C., others, 2015. Molecular profiling and targeted therapy for advanced thoracic malignancies: a biomarker-derived, multiarm, multihistology phase II basket trial. Journal of clinical oncology 33, 1000.

Lord, C.J., Ashworth, A., 2017. PARP inhibitors: Synthetic lethality in the clinic. Science 355, 1152–58. https://doi.org/10.1126/science.aam7344

Losada, A.P., Wessman, M., Tiainen, M., Hopman, A., Willard, H.F., Sole, F., Caballin, M.R., Woessner, S., Knuutila, S., 1991. Trisomy 12 in chronic lymphocytic leukemia: an interphase cytogenetic study.

Lu, Y., Liu, Y., Oeck, S., Glazer, P.M., 2018. Hypoxia Promotes Resistance to EGFR Inhibition in NSCLC Cells via the Histone Demethylases, LSD1 and PLU-1Hypoxia Promotes EGFR Inhibitor Resistance. Molecular Cancer Research 16, 1458–69.

Lucas, L.M., Dwivedi, V., Senfeld, J.I., Cullum, R.L., Mill, C.P., Piazza, J.T., Bryant, I.N., Cook, L.J., Miller, S.T., Lott, J.H., Kelley, C.M., Knerr, E.L., Markham, J.A., Kaufmann, D.P., Jacobi, M.A., Shen, J., Riese, D.J., 2022. The Yin and Yang of ERBB4: Tumor Suppressor and Oncoprotein. Pharmacol Rev 74, 18–47. https://doi.org/10.1124/pharmrev.121.000381

Lucchesi, J.C., 1968. Synthetic Lethality and Semi-Lethality among Functionally Related Mutants of Drosophila Melanogaster. Genetics 59, 37–44. https://doi.org/10.1093/genetics/59.1.37

Lynch, T.J., Bell, D.W., Sordella, R., Gurubhagavatula, S., Okimoto, R.A., Brannigan, B.W., Harris, P.L., Haserlat, S.M., Supko, J.G., Haluska, F.G., Louis, D.N., Christiani, D.C., Settleman, J., Haber, D.A., 2004. Activating Mutations in the Epidermal Growth Factor Receptor Underlying

Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. The New England Journal of Medicine 350, 2129–39. https://doi.org/10.1056/NEJMoa1603827

Maatta, J.A., Sundvall, M., Junttila, T.T., Peri, L., Laine, V.J.O., Isola, J., Egeblad, M., Elenius, K., 2006. Proteolytic Cleavage and Phosphorylation of a Tumor- associated ErbB4 Isoform Promote Ligand-independent Survival and Cancer Cell Growth□D. Molecular Biology of the Cell 17, 13.

Macdonald-Obermann, J.L., Pike, L.J., 2014. Different Epidermal Growth Factor (EGF) Receptor Ligands Show Distinct Kinetics and Biased or Partial Agonism for Homodimer and Heterodimer Formation. Journal of Biological Chemistry 289, 26178–88. https://doi.org/10.1074/jbc.M114.586826

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S., 2002. The Protein Kinase Complement of the Human Genome. Science 298, 1912–34. https://doi.org/10.1126/science.1075762

Margolis, B., Rhee, S., Felder, S., Mervic, M., Lyall, R., Levitzki, A., Ullrich, A., Zilberstein, A., Schlessinger, J., 1989. EGF induces tyrosine phosphorylation of phospholipase C-II: a potential mechanism for EGF receptor signaling. Cell 57, 1101–7.

Martincorena, I., Campbell, P.J., 2015. Somatic mutation in cancer and normal cells. Science 349, 1483–89. https://doi.org/10.1126/science.aab4082

Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R., Campbell, P.J., 2017. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell 171, 1029-1041.e21. https://doi.org/10.1016/j.cell.2017.09.042

McDonald, I.M., Graves, L.M., 2020. Enigmatic MELK: The controversy surrounding its complex role in cancer. Journal of Biological Chemistry 295, 8195–8203. https://doi.org/10.1074/jbc.REV120.013433

McFarland, C.D., Mirny, L. a, Korolev, K.S., 2014. Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.1404341111

Mcgranahan, N., Swanton, C., 2017. Clonal Heterogeneity and Tumor Evolution : Past , Present , and the Future. Cell 168, 613–28. https://doi.org/10.1016/j.cell.2017.01.018

McKay, M.M., Morrison, D.K., 2007. Integrating signals from RTKs to ERK/MAPK. Oncogene 26, 3113–21. https://doi.org/10.1038/sj.onc.1210394

Meacham, C.E., Morrison, S.J., 2013. Tumour heterogeneity and cancer cell plasticity. Nature 501, 328–37. https://doi.org/10.1038/nature12624

Melnikov, A., Rogov, P., Wang, L., Gnirke, A., Mikkelsen, T.S., 2014. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. Nucleic Acids Res 42, e112–e112. https://doi.org/10.1093/nar/gku511

Merilahti, J.A.M., Ojala, V.K., Knittle, A.M., Pulliainen, A.T., Elenius, K., 2017. Genome-wide screen of gamma-secretase–mediated intramembrane cleavage of receptor tyrosine kinases. MBoC 28, 3123–31. https://doi.org/10.1091/mbc.e17-04-0261

Merimsky, O., Issakov, J., Bickels, J., Kollender, Y., Flusser, G., Soyfer, V., Schwartz, I., Inbar, M., Meller, I., 2002. ErbB-4 expression in limb soft-tissue sarcoma: correlation with the results of neoadjuvant chemotherapy. European Journal of Cancer 38, 1335–42. https://doi.org/10.1016/S0959-8049(02)00075-8

Merimsky, O., Staroselsky, A., Inbar, M., Schwartz, Y., Wigler, N., Mann, A., Marmor, S., Greif, J., 2001. Correlation between c-erbB-4 receptor expression and response to gemcitabine–cisplatin chemotherapy in non-small-cell lung cancer. Annals of Oncology 12, 1127–31. https://doi.org/10.1023/A:1011665216616

Mersmann, O., 2021. microbenchmark: Accurate Timing Functions.

Mill, C.P., Gettinger, K.L., Riese, D.J., 2011. Ligand stimulation of ErbB4 and a constitutively-active ErbB4 mutant result in different biological responses in human pancreatic tumor cell lines. Experimental Cell Research 317, 392–404. https://doi.org/10.1016/j.yexcr.2010.11.007

Miller, V.A., Hirsh, V., Cadranel, J., Chen, Y.M., Park, K., Kim, S.W., Zhou, C., Su, W.C., Wang, M., Sun, Y., Heo, D.S., Crino, L., Tan, E.H., Chao, T.Y., Shahidi, M., Cong, X.J., Lorence, R.M., Yang,

J.C.H., 2012. Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-Lung 1): A phase 2b/3 randomised trial. The Lancet Oncology 13, 528–38. https://doi.org/10.1016/S1470-2045(12)70087-6

Misra, A., Pandey, C., Sze, S.K., Thanabalu, T., 2012. Hypoxia activated EGFR signaling induces epithelial to mesenchymal transition (EMT). PloS one 7, e49766.

Mitchell, T.J., Turajlic, S., Rowan, A., Nicol, D., Farmery, J.H.R., O'Brien, T., Martincorena, I., Tarpey, P., Angelopoulos, N., Yates, L.R., Butler, A.P., Raine, K., Stewart, G.D., Challacombe, B., Fernando, A., Lopez, J.I., Hazell, S., Chandra, A., Chowdhury, S., Rudman, S., et al., 2018. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. Cell 173, 611-623.e17. https://doi.org/10.1016/j.cell.2018.02.020

Miyake, T.M., Pradeep, S., Bayraktar, E., Stur, E., Handley, K.F., Wu, S.Y., Rodriguez-Aguayo, C., Lee, J.-S., Lopez-Berestein, G., Coleman, R.L., Sood, A.K., 2020. NRG1/ERBB3 Pathway Activation Induces Acquired Resistance to XPO1 Inhibitors. Molecular Cancer Therapeutics 19, 1727–35. https://doi.org/10.1158/1535-7163.MCT-19-0977

Mok, T.S., Wu, Y., Thongprasert, S., Yang, C.-H., Chu, D.-T., Saijo, N., Sunpaweravong, P., Han, B., Margono, B., Ichinose, Y., Nishiwaki, Y., Ohe, Y., Yang, J., Chewaskulyong, B., Jiang, H., Duffield, E.L., Watkins, C.L., Armour, A.A., Fukuoka, M., 2009. Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma. New England Journal of Medicine 361, 947–57. https://doi.org/10.1056/NEJMoa0810699

Mota, J.M., Collier, K.A., Costa, R.L.B., Taxter, T., Kalyan, A., Leite, C.A., Chae, Y.K., Giles, F.J., Carneiro, B.A., 2017. A comprehensive review of heregulins, HER3, and HER4 as potential therapeutic targets in cancer. Oncotarget 8, 89284–306. https://doi.org/10.18632/oncotarget.18467

Myers, R.M., Lerman, L.S., Maniatis, T., 1985. A General Method for Saturation Mutagenesis of Cloned DNA Fragments. Science 229, 242–47. https://doi.org/10.1126/science.2990046

Nagano, M., Kohsaka, S., Ueno, T., Kojima, S., Saka, K., Iwase, H., Kawazu, M., Mano, H., 2018. High-Throughput Functional Evaluation of Variants of Unknown Significance in ERBB2. Clin Cancer Res 24, 5112–22. https://doi.org/10.1158/1078-0432.CCR-18-0991

Nakamura, Y., Togashi, Y., Nakahara, H., Tomida, S., Banno, E., Terashima, M., Hayashi, H., Velasco, M.A. de, Sakai, K., Fujita, Y., Okegawa, T., Nutahara, K., Hamada, S., Nishio, K., 2016. Afatinib against esophageal or head-and-neck cell squamous carcinoma: significance of activating oncogenic HER4 mutations in HNSCC. Molecular cancer therapeutics 15, 1988–97. https://doi.org/10.1158/1535-7163.MCT-15-0737

Nakamura, Yukio, 2020. The information about Ba/F3.

National Institute of Standards and Technology, 2001. Advanced encryption standard (AES) (No. NIST FIPS 197). National Institute of Standards and Technology, Gaithersburg, MD. https://doi.org/10.6028/NIST.FIPS.197

Naxerova, K., Reiter, J.G., Brachtel, E., Lennerz, J.K., Wetering, M. van de, Rowan, A., Cai, T., Clevers, H., Swanton, C., Nowak, M.A., Elledge, S.J., Jain, R.K., 2017. Origins of lymphatic and distant metastases in human colorectal cancer. Science 357, 55–60. https://doi.org/10.1126/science.aai8515

Nervi, C., Badaracco, G., Maisto, A., Mauro, F., Tirindelli-Danesi, D., Starace, G., 1982. Cytometric evidence of cytogenetic and proliferative heterogeneity of human solid tumors. Cytometry: The Journal of the International Society for Analytical Cytology 2, 303–8.

Ng, P.C., Henikoff, S., 2001. Predicting deleterious amino acid substitutions. Genome research 11, 863–74.

Nguyen, B., Fong, C., Luthra, A., Smith, S.A., DiNatale, R.G., Nandakumar, S., Walch, H., Chatila, W.K., Madupuri, R., Kundra, R., Bielski, C.M., Mastrogiacomo, B., Donoghue, M.T.A., Boire, A., Chandarlapaty, S., Ganesh, K., Harding, J.J., Iacobuzio-Donahue, C.A., Razavi, P., Reznik, E., et al., 2022. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. Cell 185, 563-575.e11. https://doi.org/10.1016/j.cell.2022.01.003

Ni, C.Y., Murphy, M.P., Golde, T.E., Carpenter, G., 2001. gamma -Secretase cleavage and nuclear localization of ErbB-4 receptor tyrosine kinase. Science 294, 2179–81. https://doi.org/10.1126/science.1065412

Nowell, P.C., Hungerford, D.A., 1960. A minute chromosome in human chronic granulogytic leukemia. Sceience 132, 1497.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.-H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., Yokoyama, S., 2002. Crystal Structure of the Complex of Human Epidermal Growth Factor and Receptor Extracellular Domains. Cell 110, 775–87. https://doi.org/10.1016/S0092-8674(02)00963-7

O'Hare, T., Shakespeare, W.C., Zhu, X., Eide, C.A., Rivera, V.M., Wang, F., Adrian, L.T., Zhou, T., Huang, W.S., Xu, Q., Metcalf, C.A., Tyner, J.W., Loriaux, M.M., Corbin, A.S., Wardwell, S., Ning, Y., Keats, J.A., Wang, Y., Sundaramoorthi, R., Thomas, M., et al., 2009. AP24534, a Pan-BCR-ABL Inhibitor for Chronic Myeloid Leukemia, Potently Inhibits the T315I Mutant and Overcomes Mutation-Based Resistance. Cancer Cell 16, 401–12. https://doi.org/10.1016/j.ccr.2009.09.028

Olayioye, M.A., Beuvink, I., Horsch, K., Daly, J.M., Hynes, N.E., 1999. ErbB Receptor-induced Activation of Stat Transcription Factors Is Mediated by Src Tyrosine Kinases. Journal of Biological Chemistry 274, 17209–18. https://doi.org/10.1074/jbc.274.24.17209

Olayioye, M.A., Graus-Porta, D., Beerli, R.R., Rohrer, J., Gay, B., Hynes, N.E., 1998. ErbB-1 and ErbB-2 acquire distinct signaling properties dependent upon their dimerization partner. Molecular and cellular biology 18, 5042–51.

Olayioye, M.A., Neve, R.M., Lane, H.A., Hynes, N.E., 2000. The ErbB signaling network: receptor heterodimerization in development and cancer. EMBO J 19, 3159–67. https://doi.org/10.1093/emboj/19.13.3159

Ooms, J., 2014. The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. arXiv:1403.2805 [stat.CO].

Owusu Obeng, E., Rusciano, I., Marvi, M.V., Fazio, A., Ratti, S., Follo, M.Y., Xian, J., Manzoli, L., Billi, A.M., Mongiorgi, S., Ramazzotti, G., Cocco, L., 2020. Phosphoinositide-Dependent Signaling in Cancer: A Focus on Phospholipase C Isozymes. IJMS 21, 2581. https://doi.org/10.3390/ijms21072581

Paatero, I., Jokilammi, A., Heikkinen, P.T., Iljin, K., Kallioniemi, O.-P., Jones, F.E., Jaakkola, P.M., Elenius, K., 2012. Interaction with ErbB4 Promotes Hypoxia-inducible Factor-1α Signaling. Journal of Biological Chemistry 287, 9659–71. https://doi.org/10.1074/jbc.M111.299537

Paatero, I., Lassus, H., Junttila, T.T., Kaskinen, M., Bützow, R., Elenius, K., 2013. CYT-1 isoform of ErbB4 is an independent prognostic factor in serous ovarian cancer and selectively promotes ovarian cancer cell growth in vitro. Gynecologic oncology 129, 179–87. https://doi.org/10.1016/j.ygyno.2012.12.044

Paez, J.G., Jänne, P., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M.J., Sellers, W.R., Johnson, B.E., Meyerson, M., Janne, P.A., Lee, J.C., Tracy, S., et al., 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304, 1497–1500. https://doi.org/10.1126/science.1099314

Palacios, R., Henson, G., Steinmetz, M., McKearn, J.P., 1984. Interleukin-3 supports growth of mouse pre-B-cell clones in vitro. Nature 309, 126–31.

Pao, W., Miller, V., Zakowski, M., Doherty, J., Politi, K., Sarkaria, I., Singh, B., Heelan, R., Rusch, V., Fulton, L., Mardis, E., Kupfer, D., Wilson, R., Kris, M., Varmus, H., 2004. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proceedings of the National Academy of Sciences of the United States of America 101, 13306–11. https://doi.org/10.1073/pnas.0405220101

Pao, W., Miller, V.A., Politi, K.A., Riely, G.J., Somwar, R., Zakowski, M.F., Kris, M.G., Varmus, H., 2005. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. PLoS Medicine 2, 0225–35. https://doi.org/10.1371/journal.pmed.0020073

Pathak, S., 1990. Cytogenetic abnormalities in cancer: with special emphasis on tumor heterogeneity. Cancer and Metastasis Reviews 8, 299–318.

Peles, E., Levy, R.B., Or, E., Ullrich, A., Yarden, Y., 1991. Oncogenic forms of the neu/HER2 tyrosine kinase are permanently coupled to phospholipase C gamma. The EMBO Journal 10, 2077–86. https://doi.org/10.1002/j.1460-2075.1991.tb07739.x

Pellman, D., 2007. Aneuploidy and cancer. Nature 446, 38–39. https://doi.org/10.1038/446038a

Penington, D.J., Bryant, I., Riese II, D.J., 2002. Constitutively Active ErbB4 and ErbB2 Mutants Exhibit Distinct Biological Activities. Cell Growth & Differentiation 13, 247–56.

Philip, P., 1975. Trisomy 8 in Acute Myeloid Leukaemia. Scandinavian Journal of Haematology 14, 140–47. https://doi.org/10.1111/j.1600-0609.1975.tb00307.x

Plesa, C., Sidore, A.M., Lubock, N.B., Zhang, D., Kosuri, S., 2018. Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. Science 359, 343–47. https://doi.org/10.1126/science.aao5167

Plowman, G.D., Culouscou, J.M., Whitney, G.S., Green, J.M., Carlton, G.W., Foy, L., Neubauer, M.G., Shoyab, M., 1993a. Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal growth factor receptor family. Proceedings of the National Academy of Sciences of the United States of America 90, 1746–50. https://doi.org/10.1073/pnas.90.5.1746

Plowman, G.D., Green, J.M., Culouscou, J.M., Carlton, G.W., Rothwell, V.M., Buckley, S., 1993b. Heregulin induces tyrosine phosphorylation of HER4/p180erbB4. Nature 366, 473–75. https://doi.org/10.1038/366473a0

Podlaha, O., Riester, M., De, S., Michor, F., 2012. Evolution of the cancer genome. Trends in Genetics 28, 155–63. https://doi.org/10.1016/j.tig.2012.01.003

Posada, D., 2015. Cancer Molecular Evolution. Journal of Molecular Evolution 81, 81–83. https://doi.org/10.1007/s00239-015-9695-7

Prickett, T.D., Agrawal, N.S., Wei, X., Yates, K.E., Lin, J.C., Wunderlich, J.R., Cronin, J.C., Cruz, P., Rosenberg, S.A., Samuels, Y., 2009. Analysis of the tyrosine kinome in melanoma reveals recurrent mutations in ERBB4. Nature Genetics 41, 1127–32. https://doi.org/10.1038/ng.438

Priestley, P., Baber, J., Lolkema, M.P., Steeghs, N., Bruijn, E. de, Shale, C., Duyvesteyn, K., Haidari, S., Hoeck, A. van, Onstenk, W., Roepman, P., Voda, M., Bloemendal, H.J., Tjan-Heijnen, V.C.G., Herpen, C.M.L. van, Labots, M., Witteveen, P.O., Smit, E.F., Sleijfer, S., Voest, E.E., et al., 2019. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature 575, 210–16. https://doi.org/10.1038/s41586-019-1689-y

Qiu, C., Tarrant, M.K., Choi, S.H., Sathyamurthy, A., Bose, R., Banjade, S., Pal, A., Bornmann, W.G., Lemmon, M.A., Cole, P.A., Leahy, D.J., 2008. Mechanism of Activation and Inhibition of the HER4/ErbB4 Kinase. Structure 16, 460–67. https://doi.org/10.1016/j.str.2007.12.016

Quesnelle, K.M., Boehm, A.L., Grandis, J.R., 2007. STAT-mediated EGFR signaling in cancer. Journal of Cellular Biochemistry 102, 311–19. https://doi.org/10.1002/jcb.21475

Rabindran, S.K., Discafani, C.M., Rosfjord, E.C., Baxter, M., Floyd, M.B., Golas, J., Hallett, W.A., Johnson, B.D., Nilakantan, R., Overbeek, E., Reich, M.F., Shen, R., Shi, X., Tsou, H.-R., Wang, Y.-F., Wissner, A., 2004. Antitumor Activity of HKI-272, an Orally Active, Irreversible Inhibitor of the HER-2 Tyrosine Kinase. Cancer Res 64, 3958–65. https://doi.org/10.1158/0008-5472.CAN-03-2868

Radman, M., Wagner, R., Glickman, B., Meselson, M., 1980. DNA methylation, mismatch correction and genetic stability. Progress in environmental mutagenesis 121–30.

Rawlings, J.S., Rosler, K.M., Harrison, D.A., 2004. The JAK/STAT signaling pathway. Journal of Cell Science 117, 1281–83. https://doi.org/10.1242/jcs.00963

Regales, L., Balak, M.N., Gong, Y., Politi, K., Sawai, A., Le, C., Koutcher, J.A., Solit, D.B., Rosen, N., Zakowski, M.F., Pao, W., 2007. Development of new mouse lung tumor models expressing EGFR T790M mutants associated with clinical resistance to kinase inhibitors. PLoS ONE 2, e810. https://doi.org/10.1371/journal.pone.0000810

Rescorla, E., 2018. The Transport Layer Security (TLS) Protocol Version 1.3, Request for Comments. RFC Editor. https://doi.org/10.17487/RFC8446

Riese, D., Bermingham, Y., Van Raaij, T., Buckley, S., Plowman, G.D., Stern, D.F., 1996a. Betacellulin activates the epidermal growth factor receptor and erbB-4, and induces cellular response patterns distinct from those stimulated by epidermal growth factor or neuregulin-beta. Oncogene 12, 345–53.

Riese, D., Kim, E., Elenius, K., Buckley, S., 1996b. The epidermal growth factor receptor couples transforming growth factor-α, heparin-binding epidermal growth factor-like factor, and amphiregulin to Neu, ErbB. Journal of Biological Chemistry 271, 20047–52.

Riese, D.J., Komurasaki, T., Plowman, G.D., Stern, D.F., 1998. Activation of ErbB4 by the Bifunctional Epidermal Growth Factor Family Hormone Epiregulin Is Regulated by ErbB2. Journal of Biological Chemistry 273, 11288–94. https://doi.org/10.1074/jbc.273.18.11288

Riese, D.J., Raaij, T.M. van, Plowman, G.D., Andrews, G.C., Stern, D.F., 1995. The cellular response to neuregulins is governed by complex interactions of the erbB receptor family. Molecular and cellular biology 15, 5770–76.

Rio, C., Buxbaum, J.D., Peschon, J.J., Corfas, G., 2000. Tumor necrosis factor-alpha-converting enzyme is required for cleavage of erbB4/HER4. J Biol Chem 275, 10379–87. https://doi.org/10.1074/jbc.275.14.10379

Ritz, C., Baty, F., Streibig, J.C., Gerhard, D., 2015. Dose-Response Analysis Using R. PLoS ONE 10, e0146021. https://doi.org/10.1371/journal.pone.0146021

Roberts, C.G., Tattersall, M.H., 1990. Cytogenetic study of solid ovarian tumors. Cancer genetics and cytogenetics 48, 243–53.

Roberts, J.D., Kunkel, T.A., 1988. Fidelity of a human cell DNA replication complex. Proc. Natl. Acad. Sci. U.S.A. 85, 7064–68. https://doi.org/10.1073/pnas.85.19.7064

Robinson, D.R., Wu, Y.-M., Lin, S.-F., 2000. The protein tyrosine kinase family of the human genome. Oncogene 19, 5548–57. https://doi.org/10.1038/sj.onc.1203957

Rodriguez-Viciana, P., Warne, P.H., Dhand, R., Vanhaesebroeck, B., Gout, I., Fry, M.J., Waterfield, M.D., Downward, J., 1994. Phosphatidylinositol-3-OH kinase direct target of Ras. Nature 370, 527–32. https://doi.org/10.1038/370527a0

Roe, T., Reynolds, T.C., Yu, G., Brown, P., 1993. Integration of murine leukemia virus DNA depends on mitosis. The EMBO journal 12, 2099–2108.

Rosell, R., Carcereny, E., Gervais, R., Vergnenegre, A., Massuti, B., Felip, E., Palmero, R., Garcia-Gomez, R., Pallares, C., Sanchez, J.M., Porta, R., Cobo, M., Garrido, P., Longo, F., Moran, T., Insa, A., De Marinis, F., Corre, R., Bover, I., Illiano, A., et al., 2012. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): A multicentre, open-label, randomised phase 3 trial. The Lancet Oncology 13, 239–46. https://doi.org/10.1016/S1470-2045(11)70393-X

Roskoski, R., 2014. The ErbB/HER family of protein-tyrosine kinases and cancer. Pharmacological Research 79, 34–74. https://doi.org/10.1016/j.phrs.2013.11.002

Russell, W., Kelly, E., Hunsicker, P., Bangham, J., Maddux, S., Phipps, E., 1979. Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proceedings of the National Academy of Sciences 76, 5818–19.

Russo, A.A., Jeffrey, P.D., Pavletich, N.P., 1996. Structural basis of cyclin-dependent kinase activation by phosphorylation. Nature Structural Biology 3, 696–700. https://doi.org/10.1038/nsb0896-696

Saglam, O., Xiong, Y., Marchion, D.C., Strosberg, C., Wenham, R.M., Johnson, J.J., Saeed-Vafa, D., Cubitt, C., Hakam, A., Magliocco, A.M., 2017. ERBB4 Expression in Ovarian Serous Carcinoma Resistant to Platinum-Based Therapy. Cancer Control 24, 89–95. https://doi.org/10.1177/107327481702400115

Saito, Y., Koya, J., Araki, M., Kogure, Y., Shingaki, S., Tabata, M., McClure, M.B., Yoshifuji, K., Matsumoto, S., Isaka, Y., Tanaka, H., Kanai, T., Miyano, S., Shiraishi, Y., Okuno, Y., Kataoka, K., 2020. Landscape and function of multiple mutations within individual oncogenes. Nature 582, 95–99. https://doi.org/10.1038/s41586-020-2175-2

Saleh, M.N., Raisch, K.P., Stackhouse, M.A., Grizzle, W.E., Bonner, J.A., Mayo, M.S., Kim, H.-G., Meredith, R.F., Wheeler, R.H., Buchsbaum, D.J., 1999. Combined modality therapy of A431 human

epidermoid cancer using anti-EGFr antibody C225 and radiation. Cancer biotherapy & radiopharmaceuticals 14, 451–63.
Sardi, S.P., Murtie, J., Koirala, S., Patten, B.A., Corfas, G., 2006. Presenilin-dependent ErbB4 nuclear signaling regulates the timing of astrogenesis in the developing brain. Cell 127, 185–97.
Sawyers, C.L., 2002. Imatinib induces hematologic and cytogenetic responses in patients with chronic myelogenous leukemia in myeloid blast crisis: results of a phase II study. Blood 99, 3530–39. https://doi.org/10.1182/blood.V99.10.3530
Schechter, A.L., Stern, D.F., Vaidyanathan, L., Decker, S.J., Drebin, J.A., Greene, M.I., Weinberg, R.A., 1984. The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigen. Nature 312, 513–16. https://doi.org/10.1038/312513a0
Scheuermann, R., Tam, S., Burgers, P., Lu, C., Echols, H., 1983. Identification of the epsilon-subunit of Escherichia coli DNA polymerase III holoenzyme as the dnaQ gene product: a fidelity subunit for DNA replication. Proceedings of the National Academy of Sciences 80, 7085–89.
Schneider, G., Schmidt-Supprian, M., Rad, R., Saur, D., 2017. Tissue-specific tumorigenesis: context matters. Nat Rev Cancer 17, 239–53. https://doi.org/10.1038/nrc.2017.5
Schulze, W.X., Deng, L., Mann, M., 2005. Phosphotyrosine interactome of the ErbB-receptor kinase family. Mol Syst Biol 1, 42–54. https://doi.org/10.1038/msb4100012
Segers, V.F.M., Dugaucquier, L., Feyen, E., Shakeri, H., De Keulenaer, G.W., 2020. The role of ErbB4 in cancer. Cell Oncol (Dordr) 43, 335–52. https://doi.org/10.1007/s13402-020-00499-4
Sepp-Lorenzino, L., Eberhard, I., Ma, Z., Cho, C., Serve, H., Liu, F., Rosen, N., Lupu, R., 1996. Signal transduction pathways induced by heregulin in MDA-MB-453 breast cancer cells. Oncogene 12, 1679–87.
Settleman, J., Sawyers, C.L., Hunter, T., 2018. Challenges in validating candidate therapeutic targets in cancer. eLife 7, e32402. https://doi.org/10.7554/eLife.32402
Shi, F., Telesco, S.E., Liu, Y., Radhakrishnan, R., Lemmon, M.A., 2010. ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. Proc. Natl. Acad. Sci. U.S.A. 107, 7692–97. https://doi.org/10.1073/pnas.1002753107
Shi, X., Sousa, L.P., Mandel-Bausch, E.M., Tome, F., Reshetnyak, A.V., Hadari, Y., Schlessinger, J., Lax, I., 2016. Distinct cellular properties of oncogenic KIT receptor tyrosine kinase mutants enable alternative courses of cancer cell inhibition. Proceedings of the National Academy of Sciences of the United States of America 113, E4784–93. https://doi.org/10.1073/PNAS.1610179113
Shimamura, T., Ji, H., Minami, Y., Thomas, R.K., Lowell, A.M., Shah, K., Greulich, H., Glatt, K.A., Meyerson, M., Shapiro, G.I., Wong, K.-K., 2006. Non–Small-Cell Lung Cancer and Ba/F3 Transformed Cells Harboring the ERBB2 G776insV_G/C Mutation Are Sensitive to the Dual-Specific Epidermal Growth Factor Receptor and ERBB2 Inhibitor HKI-272. Cancer Res 66, 6487–91. https://doi.org/10.1158/0008-5472.CAN-06-0971
Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A., 2022. Cancer statistics, 2022. CA A Cancer J Clinicians 72, 7–33. https://doi.org/10.3322/caac.21708
Singh, B., Carpenter, G., Coffey, R.J., 2016. EGF receptor ligands: recent advances. F1000Res 5, 2270. https://doi.org/10.12688/f1000research.9025.1
Sithanandam, G., Anderson, L.M., 2008. The ERBB3 receptor in cancer and cancer gene therapy. Cancer Gene Ther 15, 413–48. https://doi.org/10.1038/cgt.2008.15
Slamon, D.J., 1987. Proto-oncogenes and human cancers. New England Journal of Medicine.
Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., Norton, L., 2001. Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. N Engl J Med 344, 783–92. https://doi.org/10.1056/NEJM200103153441101
Soda, M., Choi, Y.L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H., Bando, M., Ohno, S., Ishikawa, Y., Aburatani, H., Niki, T., Sohara, Y., Sugiyama, Y., Mano, H., 2007. Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. Nature 448, 561–66. https://doi.org/10.1038/nature05945

Solca, F., Dahl, G., Zoephel, A., Bader, G., Sanderson, M., Klein, C., Kraemer, O., Himmelsbach, F., Haaksma, E., Adolf, G.R., 2012. Target Binding Properties and Cellular Activity of Afatinib (BIBW 2992), an Irreversible ErbB Family Blocker. Journal of Pharmacology and Experimental Therapeutics 343, 342–50. https://doi.org/10.1124/jpet.112.197756

Soltoff, S.P., Carraway, K.L., Prigent, S.A., Gullick, W.G., Cantley, L.C., 1994. ErbB3 is involved in activation of phosphatidylinositol 3-kinase by epidermal growth factor. Molecular and Cellular Biology 14, 3550–58. https://doi.org/10.1128/mcb.14.6.3550-3558.1994

Sosman, J.A., Kim, K.B., Schuchter, L., Gonzalez, R., Pavlick, A.C., Weber, J.S., McArthur, G.A., Hutson, T.E., Moschos, S.J., Flaherty, K.T., Hersey, P., Kefford, R., Lawrence, D., Puzanov, I., Lewis, K.D., Amaravadi, R.K., Chmielowski, B., Lawrence, H.J., Shyr, Y., Ye, F., et al., 2012. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. The New England journal of medicine 366, 707–14. https://doi.org/10.1056/NEJMoa1112302

Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., Fowler, D.M., 2017. Variant interpretation: functional assays to the rescue. The American Journal of Human Genetics 101, 315–25.

Storchova, Z., Pellman, D., 2004. From polyploidy to aneuploidy, genome instability and cancer. Nat Rev Mol Cell Biol 5, 45–54. https://doi.org/10.1038/nrm1276

Strachan, L., Murison, J.G., Prestidge, R.L., Sleeman, M.A., Watson, J.D., Kumble, K.D., 2001. Cloning and Biological Activity of Epigen, a Novel Member of the Epidermal Growth Factor Superfamily. Journal of Biological Chemistry 276, 18265–71. https://doi.org/10.1074/jbc.M006935200

Sundvall, M., Iljin, K., Kilpinen, S., Sara, H., Kallioniemi, O.-P., Elenius, K., 2008a. Role of ErbB4 in breast cancer. Journal of mammary gland biology and neoplasia 13, 259–68. https://doi.org/10.1007/s10911-008-9079-3

Sundvall, M., Korhonen, A., Paatero, I., Gaudio, E., Melino, G., Croce, C.M., Aqeilan, R.I., Elenius, K., 2008b. Isoform-specific monoubiquitination, endocytosis, and degradation of alternatively spliced ErbB4 isoforms. Proceedings of the National Academy of Sciences 105, 4162–67. https://doi.org/10.1073/pnas.0708333105

Sundvall, M., Veikkolainen, V., Kurppa, K., Salah, Z., Tvorogov, D., Zoelen, E.J. van, Aqeilan, R., Elenius, K., 2010. Cell death or survival promoted by alternative isoforms of ErbB4. Molecular biology of the cell 21, 4275–86. https://doi.org/10.1091/mbc.E10

Svarovskaia, E.S., Cheslock, S.R., Zhang, W.-H., Hu, W.-S., Pathak, V.K., 2003. Retroviral mutation rates and reverse transcriptase fidelity. Frontiers in Bioscience-Landmark 8, 117–34.

Sweeney, C., Lai, C., Riese, D.J., Diamonti, A.J., Cantley, L.C., Carraway, K.L., 2000. Ligand Discrimination in Signaling through an ErbB4 Receptor Homodimer. Journal of Biological Chemistry 275, 19803–7. https://doi.org/10.1074/jbc.C901015199

Swift, S., Lorens, J., Achacoso, P., Nolan, G.P., 2001. Rapid Production of Retroviruses for Efficient Gene Delivery to Mammalian Cells Using 293T Cell–Based Systems. Current Protocols in Immunology 31. https://doi.org/10.1002/0471142735.im1017cs31

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S.C., Kok, C.Y., Noble, K., Ponting, L., Ramshaw, C.C., Rye, C.E., Speedy, H.E., et al., 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Research 47, D941–47. https://doi.org/10.1093/nar/gky1015

Taylor, S.S., Kornev, A.P., 2011. Protein kinases: evolution of dynamic regulatory proteins. Trends in Biochemical Sciences 36, 65–77. https://doi.org/10.1016/j.tibs.2010.09.006

Thatcher, N., Chang, A., Parikh, P., Rodrigues Pereira, J., Ciuleanu, T., Pawel, J. von, Thongprasert, S., Tan, E.H., Pemberton, K., Archer, V., Carroll, K., 2005. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). The Lancet 366, 1527–37. https://doi.org/10.1016/S0140-6736(05)67625-8

The AACR Project GENIE Consortium, 2017. AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov 7, 818–31. https://doi.org/10.1158/2159-8290.CD-17-0151

The future of cancer genomics, 2015. . Nat Med 21, 99–99. https://doi.org/10.1038/nm.3801

Thress, K.S., Paweletz, C.P., Felip, E., Cho, B.C., Stetson, D., Dougherty, B., Lai, Z., Markovets, A., Vivancos, A., Kuang, Y., Ercan, D., Matthews, S.E., Cantarini, M., Barrett, J.C., Jänne, P.A., Oxnard, G.R., 2015. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. Nature medicine 21, 560–62. https://doi.org/10.1038/nm.3854

Tiedt, R., Degenkolbe, E., Furet, P., Appleton, B.A., Wagner, S., Schoepfer, J., Buck, E., Ruddy, D.A., Monahan, J.E., Jones, M.D., Blank, J., Haasen, D., Drueckes, P., Wartmann, M., McCarthy, C., Sellers, W.R., Hofmann, F., 2011. A Drug Resistance Screen Using a Selective MET Inhibitor Reveals a Spectrum of Mutations That Partially Overlap with Activating Mutations Found in Cancer Patients. Cancer Research 71, 5255–64. https://doi.org/10.1158/0008-5472.CAN-10-4433

Tough, IshbelM., Baikie, A.G., Harnden, D.G., King, MargaretJ., Court Brown, W.M., Buckton, KarinE., Jacobs, PatriciaA., Mcbride, J.A., 1961. CYTOGENETIC STUDIES IN CHRONIC MYELOID LEUKÆMIA AND ACUTE LEUKAÆMIA ASSOCIATED WITH MONGOLISM. The Lancet 277, 411–17. https://doi.org/10.1016/S0140-6736(61)90001-0

TRACERx Consortium, AbdulJabbar, K., Raza, S.E.A., Rosenthal, R., Jamal-Hanjani, M., Veeriah, S., Akarca, A., Lund, T., Moore, D.A., Salgado, R., Al Bakir, M., Zapata, L., Hiley, C.T., Officer, L., Sereno, M., Smith, C.R., Loi, S., Hackshaw, A., Marafioti, T., Quezada, S.A., et al., 2020. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. Nat Med 26, 1054–62. https://doi.org/10.1038/s41591-020-0900-x

Tuomainen, K., Al-Samadi, A., Potdar, S., Turunen, L., Turunen, M., Karhemo, P.-R., Bergman, P., Risteli, M., Åström, P., Tiikkaja, R., Grenman, R., Wennerberg, K., Monni, O., Salo, T., 2019. Human Tumor–Derived Matrix Improves the Predictability of Head and Neck Cancer Drug Testing. Cancers 12, 92. https://doi.org/10.3390/cancers12010092

Turajlic, S., Swanton, C., 2016. Metastasis as an evolutionary process. Science 352, 169–75. https://doi.org/10.1126/science.aaf2784

Tvorogov, D., Sundvall, M., Kurppa, K., Hollmén, M., Repo, S., Johnson, M.S., Elenius, K., 2009. Somatic mutations of ErbB4: Selective loss-of-function phenotype affecting signal transduction pathways in cancer. Journal of Biological Chemistry 284, 5582–91. https://doi.org/10.1074/jbc.M805438200

Ullrich, A., Coussens, L., Hayflick, J.S., Dull, T.J., Gray, A., Tam, A.W., Lee, J., Yarden, Y., Libermann, T.A., Schlessinger, J., Downward, J., Mayes, E.L.V., Whittle, N., Waterfield, M.D., Seeburg, P.H., 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature 309, 418–25. https://doi.org/10.1038/309418a0

Uribe, M.L., Marrocco, I., Yarden, Y., 2021. EGFR in Cancer: Signaling Mechanisms, Drugs, and Acquired Resistance. Cancers 13, 2748. https://doi.org/10.3390/cancers13112748

Ushiro, H., Cohen, S., 1980. Identification of phosphotyrosine as a product of epidermal growth factor-activated protein kinase in A-431 cell membranes. Journal of Biological Chemistry 255, 8363–65. https://doi.org/10.1016/S0021-9258(18)43497-7

Van Cutsem, E., Peeters, M., Siena, S., Humblet, Y., Hendlisz, A., Neyns, B., Canon, J.-L., Van Laethem, J.-L., Maurel, J., Richardson, G., others, 2007. 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 25, 1658–64.

Vecchi, M., Baulida, J., Carpenter, G., 1996. Selective Cleavage of the Heregulin Receptor ErbB-4 by Protein Kinase C Activation*. Journal of Biological Chemistry 271, 18989–95. https://doi.org/10.1074/jbc.271.31.18989

Veikkolainen, V., Vaparanta, K., Halkilahti, K., Iljin, K., Sundvall, M., Elenius, K., 2011. Function of ERBB4 is determined by alternative splicing. Cell Cycle 10, 2647–57. https://doi.org/10.4161/cc.10.16.17194

Venkitaraman, A.R., 2002. Cancer susceptibility and the functions of BRCA1 and BRCA2. Cell 108, 171–82.

Vichas, A., Riley, A.K., Nkinsi, N.T., Kamlapurkar, S., Parrish, P.C.R., Lo, A., Duke, F., Chen, J., Fung, I., Watson, J., Rees, M., Gabel, A.M., Thomas, J.D., Bradley, R.K., Lee, J.K., Hatch, E.M., Baine, M.K., Rekhtman, N., Ladanyi, M., Piccioni, F., et al., 2021. Integrative oncogene-dependency mapping identifies RIT1 vulnerabilities and synergies in lung cancer. Nat Commun 12, 4789. https://doi.org/10.1038/s41467-021-24841-y

Vidal, G.A., Clark, D.E., Marrero, L., Jones, F.E., 2007. A constitutively active ERBB4/HER4 allele with enhanced transcriptional coactivation and cell-killing activities. Oncogene 26, 462–66. https://doi.org/10.1038/sj.onc.1209794

Virtaneva, K., Wright, F.A., Tanner, S.M., Yuan, B., Lemon, W.J., Caligiuri, M.A., Bloomfield, C.D., De La Chapelle, A., Krahe, R., 2001. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proceedings of the National Academy of Sciences 98, 1124–29.

Vivanco, I., Robins, H.I., Rohle, D., Campos, C., Grommes, C., Nghiemphu, P.L., Kubek, S., Oldrini, B., Chheda, M.G., Yannuzzi, N., Tao, H., Zhu, S., Iwanami, A., Kuga, D., Dang, J., Pedraza, A., Brennan, C.W., Heguy, A., Liau, L.M., Lieberman, F., et al., 2012. Differential Sensitivity of Glioma- versus Lung Cancer–Specific EGFR Mutations to EGFR Kinase Inhibitors. Cancer Discovery 2, 458–71. https://doi.org/10.1158/2159-8290.CD-11-0284

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Kinzler, K.W., 2013. Cancer Genome Landscapes. Science 339, 1546–58. https://doi.org/10.1126/science.1235122

Wacker, S.A., Houghtaling, B.R., Elemento, O., Kapoor, T.M., 2012. Using transcriptome sequencing to identify mechanisms of drug action and resistance. Nat Chem Biol 8, 235–37. https://doi.org/10.1038/nchembio.779

Wahbi, W., Naakka, E., Tuomainen, K., Suleymanova, I., Arpalahti, A., Miinalainen, I., Vaananen, J., Grenman, R., Monni, O., Al-Samadi, A., others, 2020. The critical effects of matrices on cultured carcinoma cells: Human tumor-derived matrix promotes cell invasive properties. Experimental Cell Research 389, 111885.

Wang, A., Yan, X.-E., Wu, H., Wang, W., Hu, C., Chen, C., Zhao, Z., Zhao, P., Li, X., Wang, L., Wang, B., Ye, Z., Wang, J., Wang, C., Zhang, W., Gray, N.S., Weisberg, E.L., Chen, L., Liu, J., Yun, C.-H., et al., 2016. Ibrutinib targets mutant-EGFR kinase with a distinct binding conformation. Oncotarget 7, 69760. https://doi.org/10.18632/oncotarget.11951

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research 38, e164–e164. https://doi.org/10.1093/nar/gkq603

Warmuth, M., Kim, S., Gu, X., Xia, G., Adrián, F., 2007. Ba/F3 cells and their use in kinase drug discovery. Current opinion in oncology 19, 55–60. https://doi.org/10.1097/CCO.0b013e328011a25f

Weile, J., Roth, F.P., 2018. Multiplexed assays of variant effects contribute to a growing genotype–phenotype atlas. Hum Genet 137, 665–78. https://doi.org/10.1007/s00439-018-1916-x

Weile, J., Sun, S., Cote, A.G., Knapp, J., Verby, M., Mellor, J.C., Wu, Y., Pons, C., Wong, C., Lieshout, N., Yang, F., Tasan, M., Tan, G., Yang, S., Fowler, D.M., Nussbaum, R., Bloom, J.D., Vidal, M., Hill, D.E., Aloy, P., et al., 2017. A framework for exhaustively mapping functional missense variants. Mol Syst Biol 13, 957. https://doi.org/10.15252/msb.20177908

Weinberg, R.A., 1991. Tumor suppressor genes. Science 254, 1138–46.

Weinhold, N., Jacobsen, A., Schultz, N., Sander, C., Lee, W., 2014. Genome-wide analysis of noncoding regulatory mutations in cancer. Nature genetics 46, 1160–65.

Weisberg, E., Boulton, C., Kelly, L.M., Manley, P., Fabbro, D., Meyer, T., Gilliland, D.G., Griffin, J.D., 2002. Inhibition of mutant FLT3 receptors in leukemia cells by the small molecule tyrosine kinase inhibitor PKC412. Cancer Cell 1, 433–43. https://doi.org/10.1016/S1535-6108(02)00069-7

Welcsh, P.L., Owens, K.N., King, M.-C., 2000. Insights into the functions of BRCA1 and BRCA2. Trends in Genetics 16, 69–74.

Wheeler, D.L., Yarden, Y., 2015. Receptor tyrosine kinases: family and subfamilies. Springer.

Whittaker, S., Kirk, R., Hayward, R., Zambon, A., Viros, A., Cantarino, N., Affolter, A., Nourry, A., Niculescu-Duvaz, D., Springer, C., Marais, R., 2010. Gatekeeper mutations mediate resistance to BRAF-targeted therapies. Science Translational Medicine 2. https://doi.org/10.1126/scitranslmed.3000758

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Williams, C.C., Allison, J.G., Vidal, G.A., Burow, M.E., Beckman, B.S., Marrero, L., Jones, F.E., 2004. The ERBB4/HER4 receptor tyrosine kinase regulates gene expression by functioning as a STAT5A nuclear chaperone. J Cell Biol 167, 469–78. https://doi.org/10.1083/jcb.200403155

Williams, C.S., Bernard, J.K., Demory Beckler, M., Almohazey, D., Washington, M.K., Smith, J.J., Frey, M.R., 2015. ERBB4 is over-expressed in human colon cancer and enhances cellular transformation. CARCIN 36, 710–18. https://doi.org/10.1093/carcin/bgv049

Williams, E.E., Trout, L.J., Gallo, R.M., Pitfield, S.E., Bryant, I., Penington, D.J., Riese, D.J., 2003. A constitutively active ErbB4 mutant inhibits drug-resistant colony formation by the DU-145 and PC-3 human prostate tumor cell lines. Cancer Letters 192, 67–74. https://doi.org/10.1016/S0304-3835(02)00690-0

Wilson, A.C., 1985. The Molecular Basis of Evolution. Sci Am 253, 164–73. https://doi.org/10.1038/scientificamerican1085-164

Wilson, K.J., Mill, C., Lambert, S., Buchman, J., Wilson, T.R., Hernandez-Gordillo, V., Gallo, R.M., Ades, L.M.C., Settleman, J., Riese, D.J., 2012. EGFR ligands exhibit functional differences in models of paracrine and autocrine signaling. Growth Factors 30, 107–16. https://doi.org/10.3109/08977194.2011.649918

Wilson, T.R., Fridlyand, J., Yan, Y., Penuel, E., Burton, L., Chan, E., Peng, J., Lin, E., Wang, Y., Sosman, J., Ribas, A., Li, J., Moffat, J., Sutherlin, D.P., Koeppen, H., Merchant, M., Neve, R., Settleman, J., 2012. Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. Nature 487, 505–9. https://doi.org/10.1038/nature11249

Wolman, S.R., 1986. Cytogenetic heterogeneity: its role in tumor evolution. Cancer genetics and cytogenetics 19, 129–40.

Wood, K., Hensing, T., Malik, R., Salgia, R., 2016. Prognostic and predictive value in KRAS in non–small-cell lung cancer: a review. JAMA oncology 2, 805–12.

Woodcock, D.J., Riabchenko, E., Taavitsainen, S., Kankainen, M., Gundem, G., Brewer, D.S., Ellonen, P., Lepistö, M., Golubeva, Y.A., Warner, A.C., Tolonen, T., Jasu, J., Isaacs, W.B., Emmert-Buck, M.R., Nykter, M., Visakorpi, T., Bova, G.S., Wedge, D.C., 2020. Prostate cancer evolution from multilineage primary to single lineage metastases with implications for liquid biopsy. Nat Commun 11, 5070. https://doi.org/10.1038/s41467-020-18843-5

Yang, H., Higgins, B., Kolinsky, K., Packman, K., Go, Z., Iyer, R., Kolis, S., Zhao, S., Lee, R., Grippo, J.F., Schostack, K., Simcox, M.E., Heimbrook, D., Bollag, G., Su, F., 2010. RG7204 (PLX4032), a Selective BRAF V600E Inhibitor, Displays Potent Antitumor Activity in Preclinical Melanoma Models. Cancer Res 70, 5518–27. https://doi.org/10.1158/0008-5472.CAN-10-0646

Yang, Xiaoping, Boehm, J.S., Yang, Xinping, Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S.R., Alkan, O., Bhimdi, T., Green, T.M., Johannessen, C.M., Silver, S.J., Nguyen, C., Murray, R.R., Hieronymus, H., Balcha, D., Fan, C., Lin, C., Ghamsari, L., et al., 2011. A public genome-scale lentiviral expression library of human ORFs. Nat Methods 8, 659–61. https://doi.org/10.1038/nmeth.1638

Yang, Y.R., Follo, M.Y., Cocco, L., Suh, P.-G., 2013. The physiological roles of primary phospholipase C. Advances in Biological Regulation 53, 232–41. https://doi.org/10.1016/j.jbior.2013.08.003

Yarden, Y., Pines, G., 2012. The ERBB network: at last, cancer therapy meets systems biology. Nature reviews. Cancer 12, 553–63. https://doi.org/10.1038/nrc3309

Yarden, Y., Sliwkowski, M.X., 2001. Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2, 127–37. https://doi.org/10.1038/35052073

Yates, L.R., Seoane, J., Le Tourneau, C., Siu, L.L., Marais, R., Michiels, S., Soria, J.C., Campbell, P., Normanno, N., Scarpa, A., Reis-Filho, J.S., Rodon, J., Swanton, C., Andre, F., 2017. The European Society for Medical Oncology (ESMO) Precision Medicine Glossary. Annals of Oncology 30–35. https://doi.org/10.1093/annonc/mdx707

Yenerall, P., Kollipara, R.K., Avila, K., Peyton, M., Eide, C.A., Bottomly, D., McWeeney, S.K., Liu, Y., Westover, K.D., Druker, B.J., Minna, J.D., Kittler, R., 2021. Lentiviral-Driven Discovery of Cancer Drug Resistance Mutations. Cancer Research 81, 4685–95. https://doi.org/10.1158/0008-5472.CAN-21-1153

Yonesaka, K, Hirotani, K., Kawakami, H., Takeda, M., Kaneda, H., Sakai, K., Okamoto, I., Nishio, K., Jänne, P.A., Nakagawa, K., 2015. Anti-HER3 monoclonal antibody patritumab sensitizes refractory non-small cell lung cancer to the epidermal growth factor receptor inhibitor erlotinib. Oncogene 1–9. https://doi.org/10.1038/onc.2015.142

Yonesaka, Kimio, Kudo, K., Nishida, S., Takahama, T., Iwasa, T., Yoshida, T., Tanaka, K., Takeda, M., Kaneda, H., Okamoto, I., Nishio, K., Nakagawa, K., 2015. The pan-HER family tyrosine kinase inhibitor afatinib overcomes HER3 ligand heregulin-mediated resistance to EGFR inhibitors in non-small cell lung cancer. Oncotarget 6, 33602–11. https://doi.org/10.18632/oncotarget.5286

Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S.M., Hellmann, M.D., Barron, D.A., Schram, A.M., Hameed, M., Dogan, S., Ross, D.S., Hechtman, J.F., DeLair, D.F., Yao, J., Mandelker, D.L., et al., 2017. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nature Medicine 23, 703–13. https://doi.org/10.1038/nm.4333

Zeillinger, R. al, Kury, F., Czerwenka, K., Kubista, E., Sliutz, G., Knogler, W., Huber, J., Zielinski, C., Reiner, G., Jakesz, R., 1989. HER-2 amplification, steroid receptors and epidermal growth factor receptor in primary breast cancer. Oncogene 4, 109–14.

Zhang, D., Sliwkowski, M.X., Mark, M., Frantz, G., Akita, R., Sun, Y., Hillan, K., Crowley, C., Brush, J., Godowski, P.J., 1997. Neuregulin-3 (NRG3): A novel neural tissue-enriched protein that binds and activates ErbB4. Proc. Natl. Acad. Sci. U.S.A. 94, 9562–67. https://doi.org/10.1073/pnas.94.18.9562

Zhang, J., Baran, J., Cros, A., Guberman, J.M., Haider, S., Hsu, J., Liang, Y., Rivkin, E., Wang, J., Whitty, B., Wong-Erasmus, M., Yao, L., Kasprzyk, A., 2011. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. Database : the journal of biological databases and curation 2011, bar026. https://doi.org/10.1093/database/bar026

Zhang, X., Gureasko, J., Shen, K., Cole, P.A., Kuriyan, J., 2006. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. Cell 125, 1137–49. https://doi.org/10.1016/j.cell.2006.05.013

Zuber, J., Rappaport, A.R., Luo, W., Wang, E., Chen, C., Vaseva, A.V., Shi, J., Weissmueller, S., Fellman, C., Taylor, M.J., Weissenboeck, M., Graeber, T.G., Kogan, S.C., Vakoc, C.R., Lowe, S.W., 2011. An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. Genes & Development 25, 1628–40. https://doi.org/10.1101/gad.17269211

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2022

ISBN 978-951-29-9097-9 (PRINT)
ISBN 978-951-29-9098-6 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)