Getting personal: advancing personalized oncology through computational analysis of membrane proteins

Gorostiola González, M.

Chapter 3

Computational characterization of membrane proteins as anticancer targets: Current challenges and opportunities

Marina Gorostiola González, Pepijn Rakers, Willem Jespers, Adriaan P. IJzerman, Laura H. Heitman, Gerard J.P. van Westen
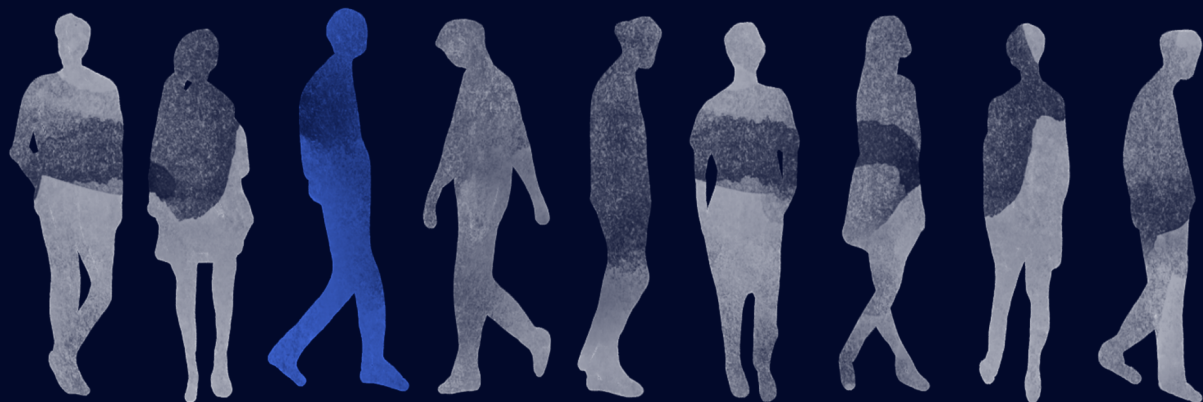
Cancer remains a leading cause of mortality worldwide and calls for novel therapeutic targets. Membrane proteins are key players in various cancer types but present unique challenges compared to soluble proteins. The advent of computational drug discovery tools offers a promising approach to address these challenges, allowing for the prioritization of "wet lab" experiments. In this review, we explore the applications of computational approaches in membrane protein oncological characterization, particularly focusing on three prominent membrane protein families: receptor tyrosine kinases (RTKs), G protein-coupled receptors (GPCRs), and solute carrier proteins (SLCs). We chose these families due to their varying levels of understanding and research data availability, which leads to distinct challenges and opportunities for computational analysis. We discuss the utilization of multi-omics data, machine learning, and structure-based methods to investigate aberrant protein functionalities associated with cancer progression within each family. Moreover, we highlight the importance of considering the broader cellular context and, in particular cross-talk between proteins. Despite existing challenges, computational tools hold promise in dissecting membrane protein dysregulation in cancer. With advancing computational capabilities and data resources, these tools are poised to play a pivotal role in identifying and prioritizing membrane proteins as personalized anticancer targets.

3

Introduction

Cancer remains one of the main causes of death worldwide, being responsible for nearly 10 million deaths each year[1]. There is therefore continuous need for novel biomarkers and disease-modifying targets. These biomarkers can be leveraged for diagnosis and progression tracking, while the identified targets can be the focus of effective and safe drug development efforts. The etiology of this multifaceted disease often involves aberrant functionality in specific proteins, resulting in increased cellular proliferation and a decrease in standard checkpoints[2]. Notably, membrane proteins have emerged as central players in the development of the most prevalent cancer types[3–5]. Unfortunately, their study presents additional challenges compared to their soluble counterparts, as has been extensively reviewed[5–7]. On the bright side, the rise of computational methodologies applied to drug discovery in the past decades has provided researchers with a new set of tools to study these protein classes[4,8]. These computational pipelines allow scientists to accelerate and streamline the identification of challenging targets and novel hits by prioritizing experiments and reducing the "wet" experimental burden[9,10].

Computational methods have applications at multiple levels of the oncological drug discovery pipeline, as highlighted in **Chapter 2**[11]. Machine learning (ML) and other statistical models can be used to analyze a wealth of multi-omics data and pinpoint the driver mutations in proteins that may be responsible for the onset of a tumor[12]. These approaches can also be used in the in silico characterization of the effect of point mutations on protein stability, function, and pharmacology[13]. Cancer-related mutations can also be analyzed structurally (as done in **Chapter 6**[14]), and ML models such as AlphaFold make it possible even when structural data is not available[15]. Furthermore, pharmacophore or quantitative structure-activity relationship (QSAR) models can be used to find chemical structures that either inhibit or activate selectively the target of interest[16,17]. Once a satisfactory selection of candidate molecules is found, structure-based approaches such as molecular docking or molecular dynamics (MD) simulations can help further refine the favorable protein-drug, or in the case of biological drugs, protein-protein interactions[18]. Such detailed knowledge, which many times can only be obtained with computational approaches, is key to enabling personalized oncological treatments[19].

Computational drug discovery pipelines can ease some experimental challenges in membrane protein research, but they face their own issues, mainly stemming from limited data availability due to experimental difficulties[5,8,10]. Different membrane protein families have varying degrees of experimental investigation, particularly in the context of cancer, which together with differences in structural and functional characteristics leads to family-specific challenges. This review will primarily focus on three membrane protein families: receptor tyrosine kinases (RTKs), G protein-coupled receptors (GPCRs), and solute carrier proteins (SLCs), which exhibit differing levels of general understanding and their connection to cancer[20–22]. RTKs have been extensively studied, particularly in the context of cancer, with a wealth of research available[23]. GPCRs, on the other hand, have received significant attention in drug discovery, but their connection to cancer has only recently become the subject of investigation[20]. In contrast, SLCs have been relatively ly understudied in general[24]. These trends explain the amount of literature linking each

of these protein families with cancer, computational drug discovery, and both (**Figure 3.1**) despite GPCRs and SLCs being the two largest families of membrane proteins.
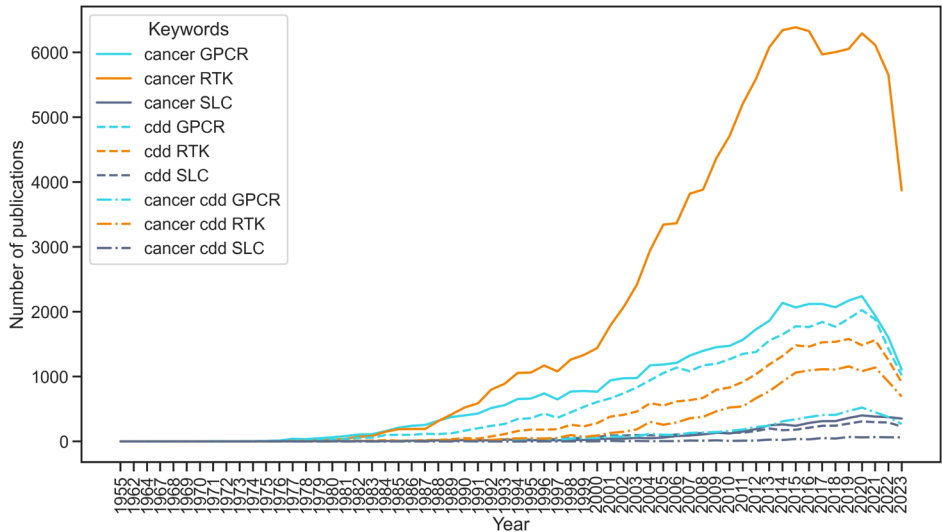

Figure 3.1. Number of publications in PubMed linking three membrane protein families – Receptor tyrosine kinases (RTK), G protein-coupled receptors (GPCR), and Solute carriers (SLC) – to cancer, computational drug discovery (CDD), and the combination of both. The search was computed using the Python package paperscraper[25]. The list of keywords for each term was: cancer – "cancer"; CDD – "computational", "computational drug discovery", "artificial intelligence", "deep learning", "machine learning", "expert systems", "QSAR", "PCM", "molecular dynamics", "docking", "molecular modeling", "FEP"; RTK – "RTK", "receptor tyrosine kinase"; GPCR – "GPCR", "G protein-coupled receptor"; SLC – "SLC", "solute carrier". Data was retrieved in November 2023, therefore the number of publications related to years 2020-2023 shows a drop corresponding to publication embargoes and delayed publication dates.

In the upcoming sections, we first expand on the key experimental and computational challenges particular to the study of membrane proteins. Then, we outline the primary structural and functional characteristics of RTKs, GPCRs, and SLCs, focusing particularly on alterations that are associated with the progression of cancer. Subsequently, we explore the use of multi-omics, ML, and structure-based methods to investigate these anomalies for each protein family, highlighting their inherent challenges. Finally, we place these membrane protein families within the broader landscape of cellular biology, focusing in particular on inter-family crosstalk. To conclude, we delineate potential avenues to further improve the computational characterization of membrane proteins as anticancer targets.

Key experimental and computational challenges in the study of membrane proteins

Membrane proteins are embedded into the cell membrane yielding an extracellular part, a transmembrane (TM) part, and an intracellular part (**Figure 3.2**). It is known that the

3

interactions between the lipids in the cell membrane and the membrane protein are crucial for the protein to acquire the right structure and to function properly[26]. This means that removing the membrane proteins from the membrane lipids, necessary for many experimental studies such as protein structural determination, brings additional challenges compared to cytosolic proteins[27,28]. To alleviate a part of this problem, researchers have come up with intricate membrane mimetics that try to imitate the natural environment of the membrane protein instead of using detergents, such as nanodiscs, lipic cubic phase, and styrene malic acid lipid particles[28]. Another problem is the often low expression level of membrane proteins compared to cytosolic proteins. As such, prokaryotes such as Escherichia coli may be used to overexpress a certain membrane protein, but this often leads to aggregation, lack of proper posttranslational modifications, and misfolding of the membrane protein in the process[27].
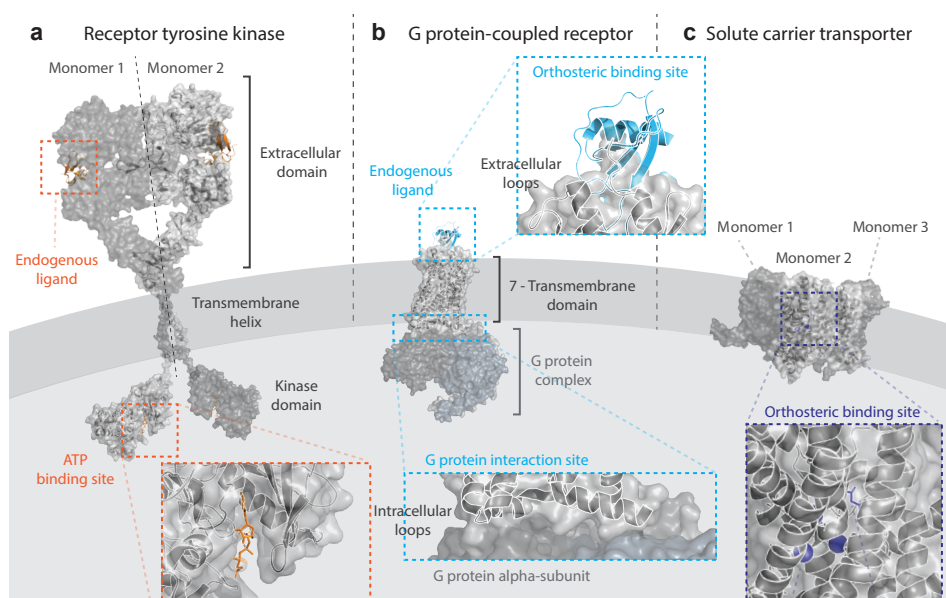

Figure 3.2. Structural models of three membrane protein family members and their main interacting partners. **a)** Receptor tyrosine kinases, represented by a dimer model of the epidermal growth factor – EGFR. The model was constructed using the following determined structures from the RCSB Protein Data Bank (PDB): 3NJP for the dimeric extracellular domain in complex with endogenous ligand EFG; 2M20 for the dimeric transmembrane helix; and 2GS6 for the monomeric kinase domain in complex with ATP. **b)** G protein-coupled receptors, represented by a model of the chemokine receptor CCR2 bound to its endogenous ligand CCL2 and the Gi protein complex. The PDB code used was 7XA3. **c)** Solute carrier transporters, represented by a trimeric model of the glutamate transporter *SLC1A3*/EAAT1 in complex with its endogenous substrate L-Aspartate and in coordination with three sodium ions needed for transport. The PDB code was 7AWM. The models were built using Pymol[29].

Crystallization itself is more difficult for membrane proteins, particularly due to protein instability outside of the membrane. This means that additional measures are needed, such as the addition of stabilizing molecules, or the construction of protein orthologues, which in turn reduces the throughput of structural characterization[28]. While X-ray crystallization has been the classical method to determine the structure of membrane

proteins, cryo-EM has been on the rise, allowing near-native structural determination with resolutions that have vastly increased in the past years[30]. The aforementioned challenges do not apply equally to the protein classes that are reviewed here, as demonstrated by their structural coverage. At the time of writing this review, over 2,000 structures were available for 41 out of 58 human RTKs, although in most cases they only represent the soluble kinase domain[31]. Moreover, 187 out of the 826 GPCRs in the human genome have been structurally determined with a total of 1,160 receptor structures in different conformations[32]. However, due to their dynamic nature, under 5% of human SLC protein structures are available – although the number is increasing rapidly thanks to cryo-EM structures[33]. Computational tools, and in particular AlphaFold, are increasingly expanding the availability of good quality predicted protein structures, although this task is less accurate for protein families with fewer structures available for training, which is the case for membrane proteins[34].

Besides finding the right structure of membrane proteins, it requires additional computing power to be able to simulate the behavior of membrane proteins in their natural environment, with detailed simulations requiring large computing clusters to be achieved[35]. The supplemental amount of interactions that come with the oligomerization of membrane proteins adds to this computing power restraint[36]. Choices have to be made about how the membrane is represented and set up, which lipids are used, and how many atoms are used in the simulation, to name a few issues. Trade-offs have to be made between the amount of detail in the simulation and the time scale at which the simulation takes place[35]. The ever-increasing computing power and efficiency of simulation algorithms help in overcoming these obstacles and, in the last decades, the number of applications has been on the rise, as we review in the following sections.

Receptor tyrosine kinases

RTKs are characterized by a single TM helix, with an extracellular region that recognizes a ligand and an intracellular tyrosine kinase domain (**Figure 3.2a**)[22]. This global structure of RTKs is highly evolutionary conserved, both within the human genome and between different species[37]. The main activation of RTKs is through dimerization or oligomerization after binding a ligand. Often these ligands themselves have a dimeric nature to assist in the activation process[37]. After dimerization, the RTKs are able to auto-phosphorylate each other's kinase domains. The phosphorylated kinase domain then allows proteins that contain an SH2 domain to bind. The intracellular kinase domain can phosphorylate downstream kinases, which can induce a range of different cellular effects such as cellular differentiation, growth, and proliferation[22]. Aberrant activation of some of these pathways plays a key role in multiple cancers. One example is the Ras/MAPK pathway, which contains the extracellular-signal-regulated ERK5 and p38 kinases downstream. ERK5 is known to play a role in tumor invasion, while p38 is able to regulate the activity of the transcription factor p53, a central protein that is found to be mutated in over half of all tumors[38,39]. Another RTK signaling pathway that often plays a role in cancer is the PI3K/Akt/mTOR pathway. PI3K is phosphorylated by RTKs, after which PI3K phosphorylates Akt. When Akt is activated, it is able to

3

phosphorylate transcription factors such as mTOR which can increase cell survival and cellular proliferation[38].

The RTK family is the most strongly associated with cancer of the three membrane protein families that are reviewed here. Famous examples of aberrant RTK signaling in cancer are epidermal growth factor receptors HER2 in breast cancer[40] and EGFR in multiple cancers[41,42]. For a more detailed overview of the role of RTKs in cancer, the review by Du et al. is recommended[22]. Many inhibitors of RTKs are currently used in clinics to treat several oncological diseases. As of January 1st, 2024, 43 small molecule RTK inhibitors have been approved for anticancer indications, besides 37 inhibitors targeting other kinase families[43]. Apart from small-molecule drugs, biologics are also used in clinics to combat aberrant RTK signaling. A famous example is the HER2 inhibitor trastuzumab, which is effective in the treatment of HER2+ breast cancer[44].

The main alterations in RTK structure and function leading to cancer development result in increased kinase activity (**Figure 3.3a**). Alterations in the kinase domain can increase the stability of the RTK dimer, leading to constitutive activation independent of the ligand[22]. This can be the result of point mutations that are then considered to be drivers of oncogenicity. Indeed in the case of EGFR, 90% of the mutations in lung cancer are found in the genetic regions that contain the kinase domain[45]. Several mutant driver prediction tools, which rely on different ML tools, are available to forecast the pathogenic effect of these mutations, although their level of agreement is limited. Interestingly, it is higher in RTKs compared to other kinase families, possibly due to the wealth of available training data[46]. Structurally, the effect of mutations in the kinase domain is easier to study, since the intracellular kinase domain of RTKs can be determined experimentally as a soluble protein. For example, structure-based approaches have been able to shed light on the mechanisms of constitutive activation triggered by the D816V mutation in the kinase domain of c-Kit[47]. However, recent studies have highlighted the importance of mutations in the extracellular and TM domains of RTKs[48,49]. Despite the different characteristics of the oncogenic mutations across domains, driver prediction ML models have been able to predict with equal success the oncogenicity of mutations in the extracellular and kinase domains[50]. Moreover, structural studies have allowed us to understand that these mutations in non-kinase domains trigger constitutive ligand-free activation via covalent extracellular or TM dimerization[50,51].

Constitutive activation can also be the effect of chromosomal rearrangements leading to fusion proteins, which are very common in RTKs. Targeting these fusion proteins in anticancer therapies is very promising because they are not present in healthy cells. Fusion genes can be detected from sequencing data, although the characterization of their oncogenic and druggability potential, and therefore their clinical relevance, is not trivial[52]. Computational analyses of genomic, transcriptomics, and drug sensitivity data can be used to prioritize oncogenic[53] and actionable RTK gene fusions[54]. From validated genetic fusions, ML models can also be constructed to further improve detection, which at the moment still has a problem of high false positive rates. For example, the method developed by Hafstað et al. showed that including an ML-based filter on top of RNA-seq-based fusion detecting algorithms improved the true positive detection

rate, although domain-specific information needed to be provided to the model[55]. Deep learning models have also been developed to predict oncogenicity starting from the fusion protein sequence without providing any oncogenic domain-specific features[56].
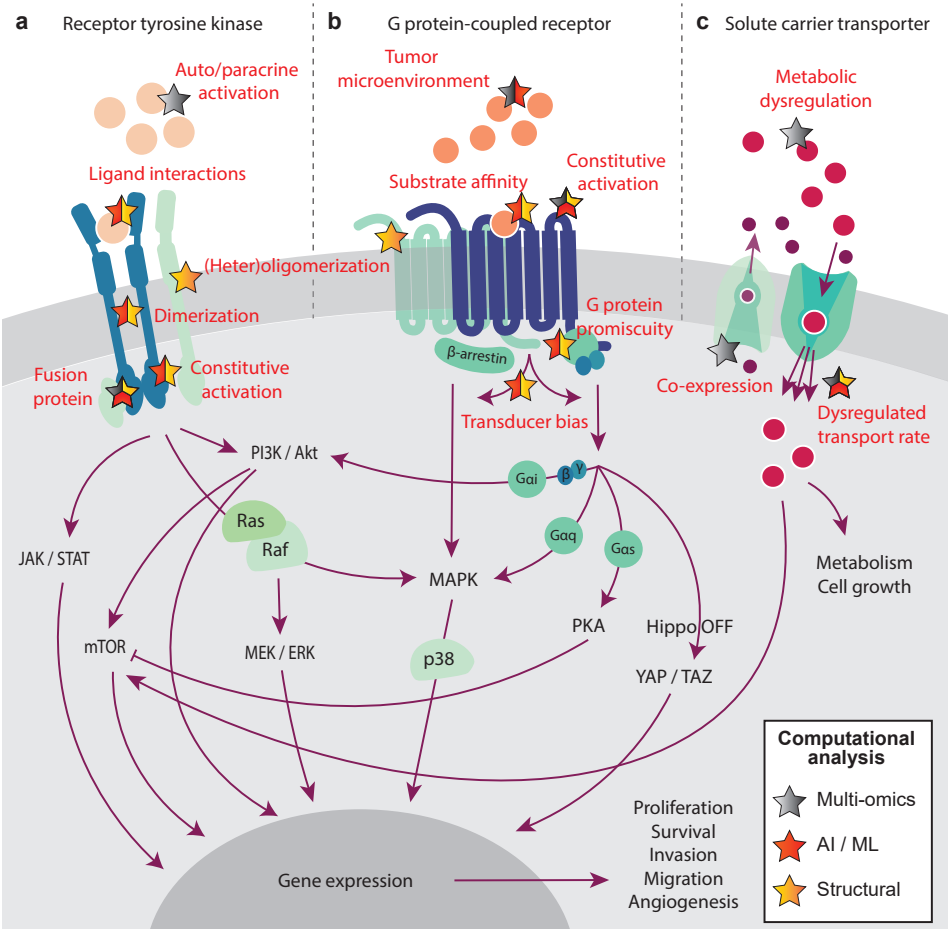

Figure 3.3. Functional and structural alterations in membrane proteins leading to cancer progression that can be characterized by one or several computational methods, including multi-omics analyses (grey star), artificial intelligence or machine learning algorithms (AI/ML, red star), and structure-based approaches (yellow star). In particular, three membrane protein families are explored: **a)** Receptor tyrosine kinases - RTK, activated by endogenous ligands represented by light orange spheres **b)** G protein-coupled receptors – GPCR activated by endogenous ligands represented by orange spheres, and **c)** Solute carriers - SLC that transport substrates represented by pink and purple spheres.

Given the historical focus on kinase domains, the study of the effect of cancer alterations in substrate affinity on the extracellular domain of the RTK is not very extended. In fact, there are very few structures of full-length RTKs containing the TM and extracellular domains[57]. However, these can be very useful in designing monoclonal antibodies targeting the ligand-binding region[58]. Structural characterization has also enabled

3

the determination of the mechanism behind the inhibitory synergy of Pertuzumab and Trastuzumab, showing with cryo-EM structures that they do induce cooperative binding[59]. ML models based on structural signatures have been leveraged to improve the design of mAbs[60,61]. The structural analysis of the kinase domain, on the other hand, has been very useful to identify and prioritize small molecules targeting this domain, and to explain the reasons for resistance[62]. The wealth of experimental data for kinase inhibitors has made it possible to use ML models to screen not only potency for wild-type[63] and mutant RTKs[64] but also clinical responses associated with gene expression signatures[65]. Beyond small molecule screening, ML models have also been employed to generate de novo RTK inhibitors by combining 2D and 3D features of known kinase inhibitors[66]. While most drugs are initially designed to bind to a specific target, some drugs bind to multiple kinases, which can be predicted through poly-pharmacology ML models[67]. The combination of ML and structure-based methods was also leveraged to identify potent small molecules that block the dimerization of the kinase domain[18]. What is more, basing these models on structural features has enabled the prediction of drug response toward specific cancer-related RTK mutants[68].

Finally, pathogenic mutations in RTKs can induce aberrant dimerization or oligomerization that leads to increased signaling. These alterations can be studied through structural analyses. For example, MD simulations helped identify ephrin type-A receptor EphA4 melanoma mutation L920F in the C-terminus as the destabilizing factor leading to receptor trimerization instead of dimerization[69]. Similarly, MD simulations showed that oncogenic mutation V536E in the TM domain of platelet-derived growth factor receptor PDGFRA is responsible for stabilizing a tetrameric conformation responsible for constitutive activation. Further dimerization alterations leading to cancer, such as heterodimerizations of EGFR with other RTKs, have been studied with MD, for which only the kinase domain is needed[70]. Beyond receptor-specific abnormalities, increased RTK signaling in cancer can also be triggered by autocrine and paracrine activation. This is the result of ligand overexpression, which in turn can be studied by multi-omics computational approaches[71].

G protein-coupled receptors

Ever since the initial characterization of the rhodopsin structure by Schertler et al.[72], a GPCR structure is recognized by its seven TM α-helices, collectively the 7TM domain (**Figure 3.2b**). Additionally, they include an N-terminus, three extracellular loops (ECL), three intracellular loops (ICL), and a C-terminus[73]. Of the three ECLs, ECL2 is usually the longest loop and the most structurally diverse between different GPCRs. An exception is the highly conserved disulfide bond between ECL2 and the TM3 α-helix[74]. Whereas ECL2 is of paramount importance to change the conformation of the GPCR after ligand binding, contacts with the ligand binding to the orthosteric pocket usually happen with ECL1 or ECL3[74]. The orthosteric binding pocket is in the extracellular side of the 7TM domain and usually comprises TM3, TM5, TM6, and TM7[75]. Additionally, many allosteric pockets have been described for GPCRs[76]. Upon activation of the GPCR by a ligand, the structure suffers a conformational rearrangement that primarily involves

TM5 and TM6[75,77]. This conformational change often induces the heterotrimeric G protein that is bound to the ICL2 and ICL3 of the GPCR to exchange its bound GDP for GTP, activating the G protein in the process[75]. Alterations in the GPCR structure due to mutations can lead to e.g., constitutive activation of the receptor, where it remains in an active state in the absence of the (endogenous) agonist[74].

GPCRs are the most commonly targeted proteins by drugs, with estimations showing that 35% of all developed drugs have a GPCR as their target[78]. Multiple GPCRs have been extensively associated with cancer, for example, the thyrotropin receptor in thyroid adenomas[79], estrogen receptors GPER1 and GPR30 in breast cancer[80,81], and gonadotropin-releasing hormone receptor GnRH in prostate cancer. In fact, hormonal therapy targeting GnRH is used in the clinic to combat prostate cancer[82]. The smoothened receptor SMO, which is part of the Hedgehog pathway is also currently targeted by small molecule antagonists in the treatment of basal cell carcinomas[83]. Furthermore, there is a range of GPCR antagonists that have been or are currently tested in (pre)clinical trials such as Ki16198 LPA receptor inhibitor for pancreatic cancer[84] and astrasentan endothelin receptor inhibitor in prostate cancer[85]. The astrasentan trial however, like multiple other GPCR antagonists that were developed, had to be ended due to unwanted side-effects occurring in patients[85]. Many clinical candidates target GPCRs involved in the tumor microenvironment (TME), such as chemokine receptor CXCR4, which are promising targets in immunotherapy[20,83]. For an extensive overview of the current state of GPCR targeting drugs in oncology, the reviews by Arang and Gutkind and Usman *et al.* are recommended[83,86].

Computational analyses have proven relevant in identifying the role of GPCRs in the TME and their potential role in immunotherapy. For example, the computational analysis of multi-omics data helped pinpoint chemokine receptor axes relevant to particular cancer types and, more importantly, the epigenetic mechanisms responsible for their overexpression (**Figure 3.3b**)[87]. Beyond cancer types, GPCR expression signatures extracted with ML models have also been shown to allow head and neck cancer patient stratification into subtypes leading to differential sensitivity to immunotherapy[88]. A similar approach enabled the classification of melanoma patients based on survival and response to immunotherapy based on combined GPCR-TME multi-omics data[89]. These applications have a high potential to define GPCRs as immune biomarkers to help in cancer treatment and patient stratification.

On the tumor side, computational tools can help study the constitutive activation of GPCRs, which may lead to the onset of cancer by inducing downstream cellular pathways[79,90]. An example is the frizzled receptors, which indirectly activate the Wnt pathway, a pathway that is strongly linked to the progression of cancer[91,92]. Genomic analyses have been able to identify oncogenic mutational drivers among GPCR genes, particularly SMO[93]. However, most GPCR mutants do not share the characteristics of classical drivers, such as a high prevalence. With a much lower mutation prevalence than in RTKs, identifying GPCR drivers needed the integration of multi-omics data[94], or the characterization of multi-gene oncodrivers[95], for which computational tools have been crucial. GPCRs of interest in cancer have also been pinpointed based solely on dysregulated

3

expression in different cancer types, for which there does not seem to be a pan-cancer common profile[96]. Given the challenges of predicting the oncogenic status of GPCR mutations, many authors have opted to study the structural impact of cancer-related mutations on the receptor's stability and activation mechanism. This can be useful to pinpoint novel potential biomarkers, such as the olfactory receptor OR2T7 destabilizing mutation D125V in glioblastoma[97], or to gain further insights into the activating and binding mechanisms of established anticancer targets, such as CXCR4[98] or SMO[99], to improve the development of targeted therapies.

The combination of structure-based and ML tools has also made it possible to get an insight into the different mechanisms behind GPCR involvement in cancer. The principal pathways leading to aberrant GPCR signaling in cancer concern G protein promiscuity and biased signaling[85]. Canonically, every GPCR preferentially activates one of the main four subtypes of G protein α subunits – $G_{\alpha i}$, $G_{\alpha q}$, $G_{\alpha s}$, and $G_{\alpha 12/13}$. However, some G proteins are more important than others in the development of cancer, which explains why certain cancer-related GPCR mutations lead to G protein promiscuity[85,86]. ML models that predict the probability of different GPCR variants binding to the different G_α protein subunits have been developed, where the GPCR embeddings were generated from the receptor's sequence[100]. This method also allowed the assessment of bias towards different signaling partners beyond G proteins, by considering also β-arrestins as interacting partners, which could have very important implications in the development of biased ligands that trigger preferentially one signaling pathway. However, structurally there does not seem to be a clear conformation basis for transducer biases[101], which would introduce an important risk of side effects to these therapies.

Similarly to RTKs, the formation of heterodimers has been shown to trigger aberrant GPCR signaling in cancer[102]. The structural analysis of the homo- and heterodimerization patterns and stability therefore introduces novel avenues for treatment. However, in the case of GPCRs, the lipidic environment seems to be extremely determinant in the formation of GPCR oligomers[103], which introduces additional experimental and computational constraints in the mechanistic analyses[104]. Surpassing the technical challenges, however, can help gain insights into cancer-related mutants leading to distinct di/oligo-merization patterns that in turn result in biased signaling[105].

Despite the challenges to computationally assess GPCR oncogenic mechanisms due to the limited availability of training data, the wealth of data collected in non-oncological GPCR drug discovery campaigns is a very good starting point for the discovery of anticancer therapies targeting GPCRs. There are several examples in this area, such as structure-based virtual screenings of novel small molecules targeting free fatty acid receptor FFAR4 for colorectal cancer[106], or adhesion receptor ADGRF5 for breast cancer[107,108]; or ligand-based screenings of small molecules targeting oxoeicosanoid receptor OXER1 that signal specifically through $G_{\alpha i}$ and/or $G_{\beta\gamma}$ for prostate cancer[109]. Beyond providing a wealth of data for novel hit identification, approved GPCR therapies can considered to be repurposed for oncological applications. To this end, the analysis of omics data can assist in identifying GPCRs with approved drugs that play an important role in cancer survival, such as dopamine receptor 2 in osteosarcoma[110]. Structure-based

and ML applications are also common in drug repurposing and can equally be leveraged for cancer applications. However, in the case of GPCRs, there are many risks of off-target effects and unintended implications in the cancer phenotype[111], thus omics-aware approaches are preferred.

Solute carriers

In contrast to the GPCR and RTK families, which both have recognizable basic structures, the SLCs family consists of very diverse proteins. A structure that many SLCs do have in common consists of 10 to 14 TM α-helices (**Figure 3.2c**)[112]. However, due to the difficulties of obtaining SLC protein structures, they are mostly classified on the basis of their known sequence. SLCs are normally considered of the same family when they have an overlap in sequence of at least 20%[112]. ML approaches have been relevant in classifying SLCs into families[113]. While SLC families are structurally diverse, they remain highly evolutionary conserved within Bilaterian species, with glucose transporters being conserved within all eukaryotes[114]. SLCs do not only have a high sequential and structural variety, but their transport mechanisms are also very diverse both conformationally and dynamically, which poses a big strain for structure-based methods[24].

SLCs transport differing molecules through the cell membrane, such as ions, lipids, and carbohydrates[115]. As Warburg *et al.* noticed back in 1927, the metabolism in tumor cells differs from that in non-tumor cells[116]. This change in metabolism is achieved through, among other things, changes in the expression of SLCs in the cell[117]. A well-known change in SLC expression in cancer is the upregulation of the glucose transporters to meet the increased demand for glucose in tumor cells[118]. However, no drugs are currently targeting SLCs in an oncological setting. Liu *et al.* performed a preclinical study in which a glucose transporter GLUT1 inhibitor was able to inhibit cancer cell growth, but this compound was not pursued any further[119]. The monocarboxylate transporter MCT1 inhibitor AZD3965 was tested in a phase 1 clinical trial on patients suffering from advanced stages of cancer, but this drug did not enter phase 2 clinical trials[120].

Multi-omics analyses of SLCs in cancer have been crucial in detecting aberrant mechanisms leading to dysregulated transport rates (**Figure 3.3c**). Most commonly, abnormal expression of SLCs beyond glucose transporters has been associated with increased transport of metabolites and building blocks necessary for cancer development, which has been used to identify SLCs as prognostic biomarkers in pan-cancer[121,122] and cancer-specific studies[123–125]. In this context, ML has also helped further discriminate between the genes with the biggest effect within the cancer signatures[126]. What is more, these transcriptomic analyses have been able to identify SLC co-expression patterns that effectively influence cancer development and that can be used as more precise biomarkers than unique SLC signatures[127,128]. Moreover, additional omics data types can help further classify tumor subtypes and make sense of the mechanisms leading to cancer development, for example by linking expression profiles to genomic[128,129] or metabolomic data[130,131]. The latter provides an additional advantage since metabolic dysregulation is a good candidate for faster biomarker detection in liquid biopsies[132].

3

ML and structure-based approaches have also been able to elucidate the role of point mutations triggering changes in transport function, although not frequently in cancer. Polymorphisms in many SLC families are related to several non-oncological diseases, such as cystinuria or ataxia, as well as drug sensitivity, and have mostly been studied in this context[133]. Several analyses have demonstrated the effect of point mutations in SLC structural and functional changes, as well as the potential risk posed by rare uncharacterized mutations[133–136]. These methods can be further explored in the context of cancer-related mutations, which have also been shown to affect transport efficiency and conformational dynamics in **Chapter 6**[14]. Structural changes can further be exploited to design and virtually screen SLC targeting compounds, as demonstrated for organic anion transporting polypeptide – OATPs[137]. ML-based virtual screening is also possible, but a relative lack of bioactivity data for SLCs is still a big drawback compared to other protein families more broadly explored, such as RTKs and GPCRs[137,138].

Crosstalk between membrane proteins

The fact that proteins do not exist in isolation is one of the most difficult aspects to tackle in cancer research. In turn, crosstalk between proteins can lead to compensation mechanisms, synergistic effects, and therapy resistance. Protein interplay has been extensively characterized for different members of the same family, for example triggering synergy by co-expression in SLCs[128], or the activation of compensatory networks by RTKs as a mechanism of drug resistance[139]. However, the crosstalk can also happen with members from other membrane protein families (**Figure 3.4**), which in turn opens opportunities for novel therapeutic avenues that can also be explored computationally.

Crosstalk between GPCRs and other membrane proteins can lead to oncogenic events. For example, the insulin receptor is known to interact with multiple GPCRs to commence the mTOR pathway[140]. Multiple transactivations between GPCRs and EGFR that induce oncogenic pathways are described, such as GPR30 and EGFR to activate the oncogenic MAPK and PI3K/Akt pathways or the protease-activated receptor 1 and EGFR in breast cancer[141,142]. Moreover, activated RTKs in cancer have been shown to activate GPCR signaling pathways via direct interaction with G proteins[143]. Computationally, structure-based approaches can be used to gain insights into the mechanisms leading to aberrant signaling[143]. Where RTKs and GPCRs are often concerned with activating downstream proteins to exert an effect, the prime task of SLCs is to transport molecules through the cell membrane, meaning SLCs themselves cannot activate oncogenic pathways. There are however interactions between RTKs and GPCRs with SLCs that aid tumor cells. For example, EGFR is known to be able to stabilize the Glucose transporter SGLT1 in tumor cells to increase cell survival[144].

Further crosstalk with SLCs is characterized by shared substrates. This is the case for many GPCR ligands, whereby the expression of SLCs serves as a regulatory mechanism for GPCR ligand availability[145]. An example is monocarboxylate transporter MCT1, which is able to efflux succinate from the cell. This succinate is then able to bind the succinate receptor SUCNR1 (a GPCR), inducing a proinflammatory response[145]. ML models can identify novel and approved small molecules with shared GPCR and SLC

targets[146] that can be exploited for drug repurposing or poly-pharmacology approaches in cancer when linked to multi-omics cancer analyses.
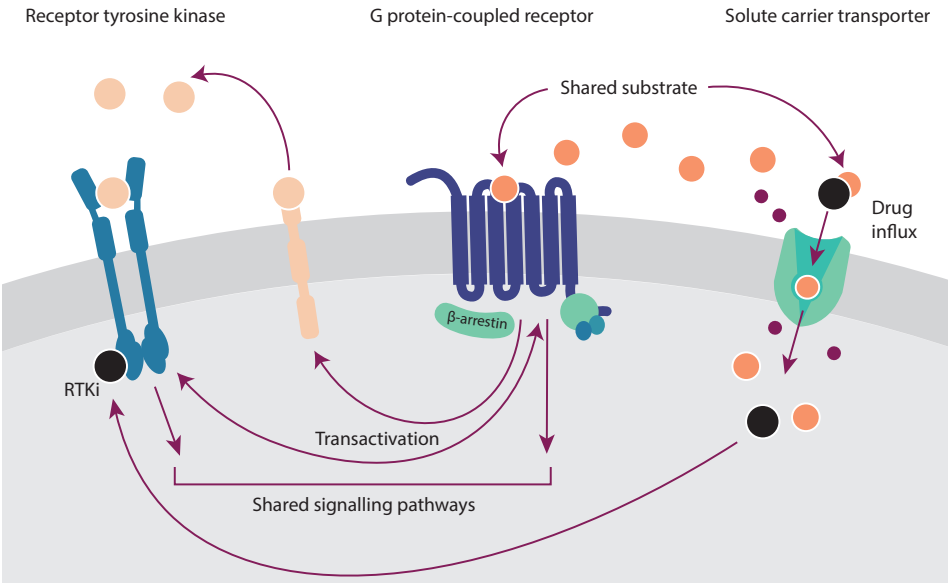

Figure 3.4. Crosstalk between three membrane protein families – Receptor tyrosine kinases (RTK), G protein-coupled receptors (GPCR), and Solute carriers (SLC) – in cancer. The expression of RTK endogenous ligands, represented by light orange spheres, can be induced by GPCRs. GPCR ligands, represented by orange spheres, can be also the substrates of SLCs. SLCs also transport RTK inhibitors (RTKi) into the cell.

Of particular relevance in cancer treatment is the transport of many anticancer drugs, including RTK inhibitors such as sunitinib, via SLCs. Thus, alterations in SLCs are a prominent cause of therapy resistance, which can be explored via multi-omics and drug sensitivity analyses[147,148]. Of note, alterations in SLCs - together with other genes - can be responsible not only for resistance to targeted therapies but also for first-line chemotherapy[149]. As a result of these alterations in membrane transporters, not only drug sensitivity is affected, but also prognosis[150], which can help stratify populations for treatment selection.

Conclusions

Membrane proteins are very promising anticancer targets, but their study is hindered by experimental challenges. Computational tools used in drug discovery pipelines can help overcome some of these challenges, although they are not free of their own obstacles. In particular, the main bottleneck is the lack of experimental data to train ML algorithms or to apply and validate structure-based approaches on. Not all membrane protein families, however, suffer equally from these issues. Historically relevant families such as RTKs have a vast wealth of experimental cancer data and many approved anticancer small molecules, which provide an excellent starting point for ML applications.

3

Moreover, the kinase domain can be experimentally determined and simulated as a soluble kinase, decreasing the threshold for structure-based approaches. This, however, means that the TM and extracellular domains of RTKs are rather unexplored computationally, even though targeting these domains could be key to avoiding off-target effects. Pharmacologically relevant families underexplored in cancer research, such as GPCRs, lack cancer-related data but they compensate for that in non-oncological data. In fact, many computational approaches have been used to study and bring GPCR-targeting molecules to the market. These tools and knowledge can be easily repurposed for oncological applications, although their relevance for this particular applicability domain should be backed up by multi-omics analyses. Finally, in membrane protein families where the lack of experimental data is very prominent, such as SLCs, family-wide tools should be explored that leverage data from other membrane protein families. These can facilitate the prediction of the effect of mutations in TM domains in particular[151,152], or assess the relevance of soluble counterparts of membrane proteins in experimental and computational approaches[153]. Regardless of the wealth of data available for each membrane protein family, they can all benefit from additional computational approaches that consider a holistic view of the tumor and its environment. Some examples include the prediction of the effect of mutations in gene expression[154], or the occurrence of mutant signatures as latent drivers[155], which could be further explored to prioritize personalized cancer therapies[156]. Moreover, the extrapolation of methods beyond their conventional use cases, for example, the application of ML algorithms to analyze structural complexes, can help circumvent some of the classical bottlenecks and assist in the design of novel therapies[157]. In conclusion, computational tools can help analyze the relevance and mechanisms behind membrane protein dysregulation in cancer and will be crucial tools for prioritizing anticancer targets and improved therapies with increasing amounts of data and computational power.

References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **71**, 209–249 (2021).

2. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).

3. Kampen, K. R. Membrane Proteins: The Key Players of a Cancer Cell. *J Membrane Biol* **242**, 69–74 (2011).

4. Lin, C.-Y. *et al.* Membrane protein-regulated networks across human cancers. *Nat Commun* **10**, 3131 (2019).

5. de Jong, E. & Kocer, A. Current Methods for Identifying Plasma Membrane Proteins as Cancer Biomarkers. *Membranes* **13**, 409 (2023).

6. Sojo, V., Dessimoz, C., Pomiankowski, A. & Lane, N. Membrane Proteins Are Dramatically Less Conserved than Water-Soluble Proteins across the Tree of Life. *Molecular Biology and Evolution* **33**, 2874–2884 (2016).

7. Hedin, L. E., Illergård, K. & Elofsson, A. An Introduction to Membrane Proteins. *J. Proteome Res.* **10**, 3324–3331 (2011).

8. Sowlati-Hashjin, S., Gandhi, A. & Garton, M. Dawn of a New Era for Membrane Protein Design. *BioDesign Research* **2022**, 9791435 (2022).

9. Rahman, M. M. *et al.* Emerging Promise of Computational Techniques in Anti-Cancer Research: At a Glance. *Bioengineering* **9**, 335 (2022).

10. Sadybekov, A. V. & Katritch, V. Computational approaches streamlining drug discovery. *Nature* **616**, 673–685 (2023).

11. Gorostiola González, M., Janssen, A. P. A., IJzerman, A. P., Heitman, L. H. & van Westen, G. J. P. Oncological drug discovery: AI meets structure-based computational research. *Drug Discovery Today* **27**, 1661–1670 (2022).

12. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371–385 (2018).

13. Rodrigues, C. H. *et al.* DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Research* **46**, W350–W355 (2018).

14. Gorostiola González, M. *et al.* Molecular insights into disease-associated glutamate transporter (EAAT1 / SLC1A3) variants using in silico and in vitro approaches. *Frontiers in Molecular Biosciences* **10**, 3389 (2023).

15. Zheng, L. *et al.* MoDAFold: a strategy for predicting the structure of missense mutant protein based on AlphaFold2 and molecular dynamics. *Briefings in Bioinformatics* **25**, bbae006 (2024).

16. Burggraaff, L. *et al.* Successive statistical and structure-based modeling to identify chemically novel kinase inhibitors. *Journal of Chemical Information and Modeling* **60**, 4283–4295 (2020).

17. Weng, C.-W. *et al.* Pharmacophore-based virtual screening for the identification of the novel Src inhibitor SJG-136 against lung cancer cell growth and motility. *Am J Cancer Res* **10**, 1668–1690 (2020).

18. Mohanan, A., Melge, A. R. & Mohan, C. G. Predicting the Molecular Mechanism of EGFR Domain II Dimer Binding Interface by Machine Learning to Identify Potent Small Molecule Inhibitor for Treatment of Cancer. *Journal of Pharmaceutical Sciences* **110**, 727–737 (2020).

19. Sakellaropoulos, T. *et al.* A Deep Learning Framework for Predicting Response to Therapy in Cancer. *Cell Reports* **29**, 3367-3373.e4 (2019).

20. Wu, V. *et al.* Illuminating the Onco-GPCRome: Novel G protein-coupled receptor-driven oncocrine networks and targets for cancer immunotherapy. *Journal of Biological Chemistry* **294**, 11062–11086 (2019).

21. Lavoro, A. *et al.* In silico analysis of the solute carrier (SLC) family in cancer indicates a link among DNA methylation, metabolic adaptation, drug response, and immune reactivity. *Front. Pharmacol.* **14**, 1191262 (2023).

22. Du, Z. & Lovly, C. M. Mechanisms of receptor tyrosine kinase activation in cancer. *Molecular Cancer* **17**, 58 (2018).

23. Saraon, P. *et al.* Receptor tyrosine kinases and cancer: oncogenic mechanisms and therapeutic approaches. *Oncogene* **40**, 4079–4093 (2021).

24. Wang, W., Gallo, L., Jadhav, A., Hawkins, R. & Parker, C. G. The Druggability of Solute Carriers. *Journal of Medicinal Chemistry* **63**, 3834–3867 (2020).

25. Born, J. & Manica, M. Trends in Deep Learning for Property-driven Drug Design. *Curr Med Chem* **28**, 7862–7886 (2021).

26. Lee, A. G. How lipids affect the activities of integral membrane proteins. *Biochim Biophys Acta* **1666**, 62–87 (2004).

27. Kermani, A. A. A guide to membrane protein X-ray crystallography. *FEBS J* **288**, 5788–5804 (2021).

28. Errasti-Murugarren, E., Bartoccioni, P. & Palacín, M. Membrane Protein Stabilization Strategies for Structural and Functional Studies. *Membranes (Basel)* **11**, 155 (2021).

29. The PyMOL Molecular Graphics System, Version 2.5.2 Schrödinger, LLC.

30. Piper, S. J., Johnson, R. M., Wootten, D. & Sexton, P. M. Membranes under the Magnetic Lens: A Dive into the Diverse World of Membrane Protein

3

Structures Using Cryo-EM. *Chem. Rev.* **122**, 13989–14017 (2022).

31. Kooistra, A. J. *et al.* KLIFS: a structural kinase-ligand interaction database. *Nucleic Acids Res* **44**, 365–371 (2015).

32. Pándy-Szekeres, G. *et al.* GPCRdb in 2018: Adding GPCR structure models and ligands. *Nucleic Acids Research* **46**, D440–D446 (2018).

33. Schlessinger, A., Zatorski, N., Hutchinson, K. & Colas, C. Targeting SLC transporters: small molecules as modulators and therapeutic opportunities. *Trends in Biochemical Sciences* **48**, 801–814 (2023).

34. Jambrich, M. A., Tusnady, G. E. & Dobson, L. How AlphaFold2 shaped the structural coverage of the human transmembrane proteome. *Sci Rep* **13**, 20283 (2023).

35. Goossens, K. & De Winter, H. Molecular Dynamics Simulations of Membrane Proteins: An Overview. *J. Chem. Inf. Model.* **58**, 2193–2202 (2018).

36. Škerle, J. *et al.* Membrane Protein Dimerization in Cell-Derived Lipid Membranes Measured by FRET with MC Simulations. *Biophys J* **118**, 1861–1875 (2020).

37. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor-tyrosine kinases. *Cell* **141**, 1117–1134 (2010).

38. Regad, T. Targeting RTK signaling pathways in cancer. *Cancers* **7**, 1758–1784 (2015).

39. Hoyos, D., Greenbaum, B. & Levine, A. J. The genotypes and phenotypes of missense mutations in the proline domain of the p53 protein. *Cell Death Differ* **29**, 938–945 (2022).

40. Browne, B. C., O'Brien, N., Duffy, M. J., Crown, J. & O'Donovan, N. HER-2 signaling and inhibition in breast cancer. *Curr Cancer Drug Targets* **9**, 419–438 (2009).

41. Wang, Z. *et al.* Mechanistic insights into the activation of oncogenic forms of EGF receptor. *Nat Struct Mol Biol* **18**, 1388–1393 (2011).

42. Lopez-Gines, C. *et al.* New pattern of EGFR amplification in glioblastoma and the relationship of gene copy number with gene expression profile. *Mod Pathol* **23**, 856–865 (2010).

43. Roskoski, R. Properties of FDA-approved small molecule protein kinase inhibitors: A 2024 update. *Pharmacol Res* **200**, 107059 (2024).

44. Kreutzfeldt, J., Rozeboom, B., Dey, N. & De, P. The trastuzumab era: current and upcoming targeted HER2+ breast cancer therapies. *Am J Cancer Res* **10**, 1045–1067 (2020).

45. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nature Reviews Cancer* **7**, 169–181 (2007).

46. Akula, S. *et al.* Large-scale pathogenicity prediction analysis of cancer-associated kinase mutations reveals variability in sensitivity and specificity of computational methods. *Cancer Medicine* **12**, 17468–17474 (2023).

47. Raghav, P. K., Singh, A. K. & Gangenahalli, G. A change in structural integrity of c-Kit mutant D816V causes constitutive signaling. *Mutat Res* **808**, 28–38 (2018).

48. Nair, S. *et al.* Novel EGFR ectodomain mutations associated with ligand-independent activation and cetuximab resistance in head and neck cancer. *PLoS One* **15**, e0229077 (2020).

49. Cleary, J. M. *et al.* FGFR2 Extracellular Domain In-Frame Deletions Are Therapeutically Targetable Genomic Alterations That Function as Oncogenic Drivers in Cholangiocarcinoma. *Cancer Discov* **11**, 2488–2505 (2021).

50. Ishiyama, N. *et al.* Computational and Functional Analyses of HER2 Mutations Reveal Allosteric Activation Mechanisms and Altered Pharmacologic Effects. *Cancer Research* **83**, 1531–1542 (2023).

51. Wagner, A. *et al.* Identification of Activating Mutations in the Transmembrane and Extracellular Domains of EGFR. *Biochemistry* **61**, 2049–2062 (2022).

52. Latysheva, N. S. & Babu, M. M. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Research* **44**, 4487–4503 (2016).

53. Hernandez, A. *et al.* In silico validation of RNA-Seq results can identify gene fusions with oncogenic potential in glioblastoma. *Sci Rep* **12**, 14439 (2022).

54. Li, J. *et al.* A functional genomic approach to actionable gene fusions for precision oncology. *Science Advances* **8**, eabm2382 (2022).

55. Hafstað, V. *et al.* Improved detection of clinically relevant fusion transcripts in cancer by machine learning classification. *BMC Genomics* **24**, 783 (2023).

56. Lovino, M., Urgese, G., Macii, E., Di Cataldo, S. & Ficarra, E. A Deep Learning Approach to the Screening of Oncogenic Gene Fusions in Humans. *Int J Mol Sci* **20**, 1645 (2019).

57. Diwanji, D., Thaker, T. & Jura, N. More than the Sum of the Parts: Towards Full-Length Receptor Tyrosine Kinase Structures. *IUBMB Life* **71**, 706–720 (2019).

58. Cruz, V. L. *et al.* Binding Affinity of Trastuzumab and Pertuzumab Monoclonal Antibodies to Extracellular HER2 Domain. *International Journal of Molecular Sciences* **24**, 12031 (2023).

59. Hao, Y., Yu, X., Bai, Y., McBride, H. J. & Huang, X. Cryo-EM Structure of HER2-trastuzumab-pertuzumab complex. *PLOS ONE* **14**, e0216095 (2019).

60. Shanehsazzadeh, A. *et al.* Unlocking de novo

antibody design with generative artificial intelligence. Preprint at *BioRxiv* https://doi.org/10.1101/2023.01.08.523187 (2023).

61. Balakrishnan, N., Baskar, G., Balaji, S., Kullappan, M. & Krishna Mohan, S. Machine learning modeling to identify affinity improved biobetter anticancer drug trastuzumab and the insight of molecular recognition of trastuzumab towards its antigen HER2. *Journal of Biomolecular Structure and Dynamics* **40**, 11638–11652 (2022).

62. Majumdar, S., Di Palma, F., Spyrakis, F., Decherchi, S. & Cavalli, A. Molecular Dynamics and Machine Learning Give Insights on the Flexibility–Activity Relationships in Tyrosine Kinome. *J. Chem. Inf. Model.* **63**, 4814–4826 (2023).

63. Reid, T.-E., Fortunak, J. M., Wutoh, A. & Wang, X. S. Cheminfomatic-based Drug Discovery of Human Tyrosine Kinase Inhibitors. *Curr Top Med Chem* **16**, 1452–1462 (2016).

64. Weng, C.-W. *et al.* Hybrid Pharmacophore- and Structure-Based Virtual Screening Pipeline to Identify Novel EGFR Inhibitors That Suppress Non-Small Cell Lung Cancer Cell Growth. *Int J Mol Sci* **23**, 3487 (2022).

65. Ferrato, M. H. *et al.* Machine learning classifier approaches for predicting response to RTK-type-III inhibitors demonstrate high accuracy using transcriptomic signatures and ex vivo data. *Bioinform Adv* **3**, vbad034 (2023).

66. Zhavoronkov, A. *et al.* Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol* **37**, 1038–1040 (2019).

67. Ma, X. H. *et al.* Virtual screening of selective multitarget kinase inhibitors by combinatorial support vector machines. *Mol Pharm* **7**, 1545–1560 (2010).

68. Robichaux, J. P. *et al.* Structure-based classification predicts drug response in EGFR-mutant NSCLC. *Nature* **597**, 732–737 (2021).

69. Light, T. P. *et al.* A cancer mutation promotes EphA4 oligomerization and signaling by altering the conformation of the SAM domain. *Journal of Biological Chemistry* **297**, 100876 (2021).

70. Zhu, M., Wang, D. D. & Yan, H. Genotype-determined EGFR-RTK heterodimerization and its effects on drug resistance in lung Cancer treatment revealed by molecular dynamics simulations. *BMC Mol Cell Biol* **22**, 34 (2021).

71. Lam, I., Pickering, C. M. & Mac Gabhann, F. Context-Dependent Regulation of Receptor Tyrosine Kinases: Insights from Systems Biology Approaches. *Wiley Interdiscip Rev Syst Biol Med* **11**, e1437 (2019).

72. Schertler, G. F. X., Villa, C. & Henderson, R. Projection structure of rhodopsin. *Nature* **362**, 770–772 (1993).

73. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).

74. Wheatley, M. *et al.* Lifting the lid on GPCRs: the role of extracellular loops. *Br J Pharmacol* **165**, 1688–1703 (2012).

75. Weis, W. I. & Kobilka, B. K. The Molecular Basis of G Protein–Coupled Receptor Activation. *Annual Review of Biochemistry* **87**, 897–919 (2018).

76. Hedderich, J. B. *et al.* The pocketome of G-protein-coupled receptors reveals previously untargeted allosteric sites. *Nat Commun* **13**, 2567 (2022).

77. Schwartz, T. W., Frimurer, T. M., Holst, B., Rosenkilde, M. M. & Elling, C. E. Molecular Mechanism of 7tm Receptor Activation—a Global Toggle Switch Model. *Annual Review of Pharmacology and Toxicology* **46**, 481–519 (2006).

78. Sriram, K. & Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol Pharmacol* **93**, 251–258 (2018).

79. Parma, J. *et al.* Somatic mutations in the thyrotropin receptor gene cause hyperfunctioning thyroid adenomas. *Nature* **365**, 649–651 (1993).

80. Yang, H., Wang, C., Liao, H. & Wang, Q. Activation of GPER by E2 promotes proliferation, invasion and migration of breast cancer cells by regulating the miR-124/CD151 pathway. *Oncology Letters* **21**, 432 (2021).

81. Steiman, J., Peralta, E. A., Louis, S. & Kamel, O. Biology of the estrogen receptor, GPR30, in triple negative breast cancer. *The American Journal of Surgery* **206**, 698–703 (2013).

82. Cook, T. & Sheridan, W. P. Development of GnRH antagonists for prostate cancer: new approaches to treatment. *Oncologist* **5**, 162–168 (2000).

83. Usman, S., Khawer, M., Rafique, S., Naz, Z. & Saleem, K. The current status of anti-GPCR drugs against different cancers. *J Pharm Anal* **10**, 517–521 (2020).

84. Komachi, M. *et al.* Orally active lysophosphatidic acid receptor antagonist attenuates pancreatic cancer invasion and metastasis in vivo. *Cancer Sci* **103**, 1099–1104 (2012).

85. Chaudhary, P. K. & Kim, S. An Insight into GPCR and G-Proteins as Cancer Drivers. *Cells* **10**, 3288 (2021).

86. Arang, N. & Gutkind, J. S. G Protein-Coupled receptors and heterotrimeric G proteins as cancer drivers. *FEBS Lett* **594**, 4201–4232 (2020).

87. Qualliotine, J. R. *et al.* A Network Landscape of HPVOPC Reveals Methylation Alterations as Significant Drivers of Gene Expression via an Immune-Mediated GPCR Signal. *Cancers (Basel)* **15**, 4379 (2023).

88. Huang, C., Zhu, F., Zhang, H., Wang, N. & Huang, Q. Identification of S1PR4 as an immune modulator for favorable prognosis in HNSCC

3

through machine learning. *iScience* **26**, 107693 (2023).

89. Shen, K. *et al.* Prediction of survival and immunotherapy response by the combined classifier of G protein-coupled receptors and tumor microenvironment in melanoma. *Eur J Med Res* **28**, 352 (2023).

90. Vecchio, E. A. *et al.* Ligand-Independent Adenosine A2B Receptor Constitutive Activity as a Promoter of Prostate Cancer Cell Proliferation. *J Pharmacol Exp Ther* **357**, 36–44 (2016).

91. Lappano, R. & Maggiolini, M. GPCRs and cancer. *Acta Pharmacologica Sinica* **33**, 351–362 (2012).

92. Clevers, H. Wnt/beta-catenin signaling in development and disease. *Cell* **127**, 469–480 (2006).

93. Sherman, M. A. *et al.* Genome-wide mapping of somatic mutation rates uncovers drivers of cancer. *Nat Biotechnol* **40**, 1634–1643 (2022).

94. Li, S. *et al.* Multi-omics integration analysis of GPCRs in pan-cancer to uncover inter-omics relationships and potential driver genes. *Comput Biol Med* **161**, 106988 (2023).

95. Raimondi, F. *et al.* Rare, functional, somatic variants in gene families linked to cancer genes: GPCR signaling as a paradigm. *Oncogene* **38**, 6491–6506 (2019).

96. Suteau, V. *et al.* Identification of Dysregulated Expression of G Protein Coupled Receptors in Endocrine Tumors by Bioinformatics Analysis: Potential Drug Targets? *Cells* **11**, 703 (2022).

97. Sharp, A. K. *et al.* Biophysical insights into OR2T7: Investigation of a potential prognostic marker for glioblastoma. *Biophysical Journal* **121**, 3706–3718 (2022).

98. Rebolledo-Bustillo, M. *et al.* Structural Basis of the Binding Mode of the Antineoplastic Compound Motixafortide (BL-8040) in the CXCR4 Chemokine Receptor. *Int J Mol Sci* **24**, 4393 (2023).

99. Kumari, S., Mitra, A. & Bulusu, G. Structural dynamics of Smoothened (SMO) in the ciliary membrane and its interaction with membrane lipids. *Biochim Biophys Acta Biomembr* **1864**, 183946 (2022).

100. Matic, M. *et al.* PRECOGx: exploring GPCR signaling mechanisms with deep protein representations. *Nucleic Acids Research* **50**, W598–W610 (2022).

101. Seyedabadi, M., Gharghabi, M., Gurevich, E. V. & Gurevich, V. V. Structural basis of GPCR coupling to distinct signal transducers: implications for biased signaling. *Trends Biochem Sci* **47**, 570–581 (2022).

102. Coke, C. J. *et al.* Simultaneous Activation of Induced Heterodimerization between CXCR4 Chemokine Receptor and Cannabinoid Receptor 2 (CB2) Reveals a Mechanism for Regulation of Tumor Progression. *J Biol Chem* **291**, 9991–10005

(2016).

103. Gahbauer, S., Pluhackova, K. & Böckmann, R. A. Closely related, yet unique: Distinct homo- and heterodimerization patterns of G protein coupled chemokine receptors and their fine-tuning by cholesterol. *PLOS Computational Biology* **14**, e1006062 (2018).

104. Di Marino, D., Conflitti, P., Motta, S. & Limongelli, V. Structural basis of dimerization of chemokine receptors CCR5 and CXCR4. *Nat Commun* **14**, 6439 (2023).

105. Paradis, J. S. *et al.* Computationally designed GPCR quaternary structures bias signaling pathway activation. *Nat Commun* **13**, 6826 (2022).

106. Pal, A., Curtin, J. F. & Kinsella, G. K. In silico and in vitro screening for potential anticancer candidates targeting GPR120. *Bioorg Med Chem Lett* **31**, 127672 (2021).

107. Muthiah, I., Rajendran, K. & Dhanaraj, P. In silico molecular docking and physicochemical property studies on effective phytochemicals targeting GPR116 for breast cancer treatment. *Mol Cell Biochem* **476**, 883–896 (2021).

108. Muthiah, I., Rajendran, K., Dhanaraj, P. & Vallinayagam, S. In silico structure prediction, molecular docking and dynamic simulation studies on G Protein-Coupled Receptor 116: a novel insight into breast cancer therapy. *Journal of Biomolecular Structure and Dynamics* **39**, 4807–4815 (2021).

109. Panagiotopoulos, A. A., Konstantinou, E., Pirintsos, S. A., Castanas, E. & Kampa, M. Mining the ZINC database of natural products for specific, testosterone-like, OXER1 antagonists. *Steroids* **199**, 109309 (2023).

110. Tan, M. *et al.* Prediction and Identification of GPCRs Targeting for Drug Repurposing in Osteosarcoma. *Frontiers in Oncology* **12**, 828849 (2022).

111. Cornwell, A. C. & Feigin, M. E. Unintended Effects of GPCR-Targeted Drugs on the Cancer Phenotype. *Trends in Pharmacological Sciences* **41**, 1006–1022 (2020).

112. Schlessinger, A. *et al.* Comparison of human solute carriers. *Protein Sci* **19**, 412–428 (2010).

113. Meixner, E. *et al.* A substrate-based ontology for human solute carriers. *Mol Syst Biol* **16**, e9652 (2020).

114. Höglund, P. J., Nordström, K. J. V., Schiöth, H. B. & Fredriksson, R. The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilaterian species. *Mol Biol Evol* **28**, 1531–1541 (2011).

115. Colas, C., Ung, P. M. U. & Schlessinger, A. SLC transporters: Structure, function, and drug discovery. *MedChemComm* **7**, 1069–1081 (2016).

116. Warburg, O., Wind, F. & Negelein, E. The metabolism of tumors in the body. *J Gen Physiol* **8**, 519–530 (1927).

117. El-Gebali, S., Bentz, S., Hediger, M. A. & Anderle, P. Solute carriers (SLCs) in cancer. *Molecular Aspects of Medicine* **34**, 719–734 (2013).

118. Scafoglio, C. *et al.* Functional expression of sodium-glucose transporters in cancer. *Proc Natl Acad Sci U S A* **112**, E4111-4119 (2015).

119. Liu, Y. *et al.* A small-molecule inhibitor of glucose transporter 1 downregulates glycolysis, induces cell-cycle arrest, and inhibits cancer cell growth in vitro and in vivo. *Mol Cancer Ther* **11**, 1672–1682 (2012).

120. Lin, L., Yee, S. W., Kim, R. B. & Giacomini, K. M. SLC transporters as therapeutic targets: Emerging opportunities. *Nature Reviews Drug Discovery* **14**, 543–560 (2015).

121. Xie, M. *et al.* Systematic pan-cancer analysis identifies SLC35C1 as an immunological and prognostic biomarker. *Sci Rep* **13**, 5331 (2023).

122. Li, J. *et al.* A pan-cancer analysis revealed the role of the SLC16 family in cancer. *Channels (Austin)* **15**, 528–540 (2021).

123. Zhu, J. *et al.* Identification of a Six-Gene SLC Family Signature With Prognostic Value in Patients With Lung Adenocarcinoma. *Frontiers in Cell and Developmental Biology* **9**, 803198 (2021).

124. Zhao, X., Jin, L., Liu, Y., Liu, Z. & Liu, Q. Bioinformatic analysis of the role of solute carrier-glutamine transporters in breast cancer. *Annals of Translational Medicine* **10**, 777–777 (2022).

125. Sun, T., Bi, F., Liu, Z. & Yang, Q. SLC7A2 serves as a potential biomarker and therapeutic target for ovarian cancer. *Aging* **12**, 13281–13296 (2020).

126. Samaržija, I., Trošelj, K. G. & Konjevoda, P. Prognostic Significance of Amino Acid Metabolism-Related Genes in Prostate Cancer Retrieved by Machine Learning. *Cancers* **15**, 1309 (2023).

127. Zhang, Y. *et al.* Co-expression pattern of SLC transporter genes associated with the immune landscape and clinical outcomes in gastric cancer. *Journal of Cellular and Molecular Medicine* **27**, 4181–4194 (2023).

128. Zhou, R. *et al.* Integrative analysis of co-expression pattern of solute carrier transporters reveals molecular subtypes associated with tumor microenvironment hallmarks and clinical outcomes in colon cancer. *Heliyon* **10**, e22775 (2024).

129. Zhang, P. *et al.* SLC31A1 Identifying a Novel Biomarker with Potential Prognostic and Immunotherapeutic Potential in Pan-Cancer. *Biomedicines* **11**, 2884 (2023).

130. Danzi, F. *et al.* To metabolomics and beyond: a technological portfolio to investigate cancer metabolism. *Sig Transduct Target Ther* **8**, 137 (2023).

131. Poplawski, P. *et al.* Coordinated reprogramming of renal cancer transcriptome, metabolome and secretome associates with immune tumor infiltration. *Cancer Cell International* **23**, 2 (2023).

132. Wang, W. *et al.* Cancer metabolites: promising biomarkers for cancer liquid biopsy. *Biomarker Research* **11**, 66 (2023).

133. Schaller, L. & Lauschke, V. M. The genetic landscape of the human solute carrier (SLC) transporter superfamily. *Human Genetics* **138**, 1359–1377 (2019).

134. Koleske, M. L. *et al.* Functional genomics of OCTN2 variants informs protein-specific variant effect predictor for Carnitine Transporter Deficiency. *Proceedings of the National Academy of Sciences* **119**, e2210247119 (2022).

135. Pasquadibisceglie, A., Quadrotta, V. & Polticelli, F. In Silico Analysis of the Structural Dynamics and Substrate Recognition Determinants of the Human Mitochondrial Carnitine/Acylcarnitine SLC25A20 Transporter. *International Journal of Molecular Sciences* **24**, 3946 (2023).

136. Wu, Q. *et al.* Ataxia-linked SLC1A3 mutations alter EAAT1 chloride channel activity and glial regulation of CNS function. *Journal of Clinical Investigation* **132**, e154891 (2022).

137. Tuerkova, A. *et al.* Identifying Novel Inhibitors for Hepatic Organic Anion Transporting Polypeptides by Machine Learning-Based Virtual Screening. *J. Chem. Inf. Model.* **62**, 6323–6335 (2022).

138. Burggraaff, L. *et al.* Identification of novel small molecule inhibitors for solute carrier SGLT1 using proteochemometric modeling. *Journal of Cheminformatics* **11**, 15 (2019).

139. Xu, A. M. & Huang, P. H. Receptor tyrosine kinase coactivation networks in cancer. *Cancer Res* **70**, 3857–3860 (2010).

140. Kisfalvi, K., Rey, O., Young, S. H., Sinnett-Smith, J. & Rozengurt, E. Insulin potentiates Ca2+ signaling and phosphatidylinositol 4,5-bisphosphate hydrolysis induced by Gq protein-coupled receptor agonists through an mTOR-dependent pathway. *Endocrinology* **148**, 3246–3257 (2007).

141. Prossnitz, E. R. & Maggiolini, M. Mechanisms of estrogen signaling and gene expression via GPR30. *Mol Cell Endocrinol* **308**, 32–38 (2009).

142. Arora, P., Cuevas, B. D., Russo, A., Johnson, G. L. & Trejo, J. Persistent transactivation of EGFR and ErbB2/HER2 by protease-activated receptor-1 promotes breast carcinoma cell invasion. *Oncogene* **27**, 4434–4445 (2008).

143. Kalogriopoulos, N. A. *et al.* Receptor tyrosine kinases activate heterotrimeric G proteins via phosphorylation within the interdomain cleft of Gαi. *Proceedings of the National Academy of Sciences* **117**, 28763–28774 (2020).

144. Weihua, Z. *et al.* Survival of cancer cells is

3

maintained by EGFR independent of its kinase activity. *Cancer Cell* **13**, 385–393 (2008).

145. Sijben, H. J., Superti-Furga, G., IJzerman, A. P. & Heitman, L. H. Targeting solute carriers to modulate receptor-ligand interactions. *Trends Pharmacol Sci* **43**, 358–361 (2022).

146. Oh, J., Ceong, H.-T., Na, D. & Park, C. A machine learning model for classifying G-protein-coupled receptors as agonists or antagonists. *BMC Bioinformatics* **23**, 346 (2022).

147. van de Geer, W. S. *et al.* Identifying somatic changes in drug transporters using whole genome and transcriptome sequencing data of advanced tumors. *Biomedicine & Pharmacotherapy* **159**, 114210 (2023).

148. Qu, Y.-Y., Guo, R.-Y., Luo, M.-L. & Zhou, Q. Pan-Cancer Analysis of the Solute Carrier Family 39 Genes in Relation to Oncogenic, Immune Infiltrating, and Therapeutic Targets. *Frontiers in Genetics* **12**, 757582 (2021).

149. Buttarelli, M. *et al.* Identification of a novel gene signature predicting response to first-line chemotherapy in BRCA wild-type high-grade serous ovarian cancer patients. *Journal of Experimental & Clinical Cancer Research* **41**, 50 (2022).

150. Alves, R. *et al.* Genetic Variants of ABC and SLC Transporter Genes and Chronic Myeloid Leukaemia: Impact on Susceptibility and Prognosis. *International Journal of Molecular Sciences* **23**, 9815 (2022).

151. Pires, D. E. V., Rodrigues, C. H. M. & Ascher, D. B. mCSM-membrane: predicting the effects of mutations on transmembrane proteins. *Nucleic Acids Research* **48**, W147–W153 (2020).

152. Ge, F. *et al.* MutTMPredictor: Robust and accurate cascade XGBoost classifier for prediction of mutations in transmembrane proteins. *Comput Struct Biotechnol J* **19**, 6400–6416 (2021).

153. Ma, L. *et al.* CrMP-Sol database: classification, bioinformatic analyses and comparison of cancer-related membrane proteins and their water-soluble variant designs. *BMC Bioinformatics* **24**, 360 (2023).

154. Avsec, Ž. *et al.* Effective gene expression prediction from sequence by integrating long-range interactions. *Nature Methods* **18**, 1196–1203 (2021).

155. Yavuz, B. R., Tsai, C.-J., Nussinov, R. & Tuncbag, N. Pan-cancer clinical impact of latent drivers from double mutations. *Commun Biol* **6**, 202 (2023).

156. Mateo, L. *et al.* Personalized cancer therapy prioritization based on driver alteration co-occurrence patterns. *Genome Medicine* **12**, 78 (2020).

157. Qi, X., Zhao, Y., Qi, Z., Hou, S. & Chen, J. Machine Learning Empowering Drug Discovery: Applications, Opportunities and Challenges. *Molecules* **29**, 903 (2024).