

Article

Annotating Cancer Variants and Anti-Cancer Therapeutics in Reactome

Marija Milacic ¹, Robin Haw ^{1,*}, Karen Rothfels ¹, Guanming Wu ¹, David Croft ², Henning Hermjakob ², Peter D'Eustachio ³ and Lincoln Stein ¹

¹ Informatics and Bio-computing Platform, Ontario Institute for Cancer Research, Toronto, ON, M5G0A3, Canada; E-Mails: MORlic-Milacic@oicr.on.ca (M.M.); Karen.Rothfels@oicr.on.ca (K.R.); guanmingwu@gmail.com (G.W.); lincoln.stein@gmail.com (L.S.)

² European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK; E-Mails: croft@ebi.ac.uk (D.C.); hhe@ebi.ac.uk (H.H.)

³ Department of Biochemistry, NYU School of Medicine, New York, NY 10016, USA; E-Mail: Peter.D'Eustachio@nyumc.org

* Author to whom correspondence should be addressed; E-Mail: robin.haw@oicr.on.ca; Tel.: +1-647-260-7985; Fax: +1-416-977-1118.

Received: 8 October 2012; in revised form: 31 October 2012 / Accepted: 2 November 2012 /

Published: 8 November 2012

Abstract: Reactome describes biological pathways as chemical reactions that closely mirror the actual physical interactions that occur in the cell. Recent extensions of our data model accommodate the annotation of cancer and other disease processes. First, we have extended our class of protein modifications to accommodate annotation of changes in amino acid sequence and the formation of fusion proteins to describe the proteins involved in disease processes. Second, we have added a disease attribute to reaction, pathway, and physical entity classes that uses disease ontology terms. To support the graphical representation of “cancer” pathways, we have adapted our Pathway Browser to display disease variants and events in a way that allows comparison with the wild type pathway, and shows connections between perturbations in cancer and other biological pathways. The curation of pathways associated with cancer, coupled with our efforts to create other disease-specific pathways, will interoperate with our existing pathway and network analysis tools. Using the Epidermal Growth Factor Receptor (EGFR) signaling pathway as an example, we show how Reactome annotates and presents the altered biological behavior of EGFR variants due to their altered kinase and ligand-binding properties, and the mode of action and specificity of anti-cancer therapeutics.

Keywords: pathway database; pathway visualization; network visualization; cancer annotation; EGFR signaling

1. Introduction

The development of a malignantly transformed cell from a normal cell is a complex multi-step process that remains incompletely understood [1,2]. “Bottom-up” studies of relevant processes such as control of cell division, cell migration, tissue remodeling, and cell death have allowed the identification and characterization of many individual genes whose malfunction due to mutation or misregulation is associated with malignant transformation [3–6]. More recently, the development of high-throughput studies that exploit the availability of whole-genome sequencing has enabled “top-down” studies to systematically catalogue somatically mutated genes and altered patterns of gene expression in individual tumors [7,8]. These studies have confirmed the importance of genes identified as key players in the “bottom-up” studies, but have also suggested roles for additional genes and gene combinations not previously associated with processes relevant to malignancy.

Pathway databases have been effectively used to annotate our “bottom-up” understanding of molecular details of processes relevant to cell growth, differentiation, migration, and death. Here, we describe one such database, Reactome, focusing on extensions to this basic annotation strategy to allow the capture of details of disease processes, and on the development of data analysis tools to support the annotation and interpretation of gene sets identified in top-down studies.

Reactome is an open-source, open access, curated and peer-reviewed biological knowledgebase of human reactions, pathways and processes that serves as a platform for pathway visualization and analysis [9–12]. Reactome provides information about proteins and small molecules and how they participate in pathways to coordinate cellular events. The Reactome database employs a reductionist data model, which represents biology as reactions that convert input physical entities into output physical entities. The Reactome definition of a “reaction” is broad, including binding, dissociation, translocation and degradation, in addition to biochemical transformations of proteins and small molecules. Reactions are linked in causal chains to form pathways which in turn are grouped to represent larger biological processes like intermediary metabolism, innate immunity, solute transport, GPCR signal transduction, and apoptosis [13,14].

Reactome curators, in collaboration with outside expert researchers, annotate new pathways. The molecular details of every reaction are traceable to experimental evidence in the primary literature. If an event has not been directly studied in human systems, the appropriate non-human reaction is annotated and the homologous human one is inferred from it. Every pathway module is peer-reviewed by an additional expert. New and revised modules are publicly released to the Reactome website every quarter. Pathways, reactions, protein and small molecule entities are cross referenced with accession numbers and identifiers to a number of well-established databases, including NCBI Gene [15], Ensembl [16] and UniProt databases [17], UCSC Genome Browser [18], and ChEBI [19]. Physical entities and events are further linked to “Molecular Function”, “Biological Process” and “Cellular Component” ontology terms found in Gene Ontology (GO) [20].

Currently, the pathways in Reactome cover about 25% of the gene products encoded in the human genome, and contain the normal versions of many pathways that can be abnormally activated in cancer, such as “Signaling by EGFR” [21], “Signaling by FGFR” [22], “Signaling by NOTCH” [23], “PIP3 Activates AKT Signaling” [24], “RAF/MAP Kinase Cascade” [25]. We have also annotated a number of pathways that can be inactivated in cancer, such as pathways involving TP53: “Apoptosis” [26] and “Cell Cycle Checkpoints” [27], as well as pathways involving the RB1 protein family: “Mitotic G1-G1/S phases” [28].

Here, we use the epidermal growth factor receptor (EGFR), fibroblast growth factor receptor (FGFR) and PI3K/AKT signaling pathways to illustrate Reactome annotation of cancer pathways. EGFR and FGFR are transmembrane receptor tyrosine kinases. EGFR is activated by several growth factors, including the epidermal growth factor (EGF) [29]. FGFR family members (FGFR1, FGFR2, FGFR3 and FGFR4) are activated by 18 of 22 existing human fibroblast growth factors (FGFs), with each FGFR showing different affinity for individual FGFs [30]. Growth factor binding induces a conformational change that enables dimerization and trans-autophosphorylation on C-tail tyrosine residues of EGFR [31] and FGFRs [32–34]. Phosphorylated tyrosines in the C-tails of EGFR and FGFR serve as docking sites for downstream effectors that, upon binding to phosphorylated receptors, activate intracellular signaling cascades that regulate cellular proliferation, differentiation, and survival [30,35,36]. One of the intracellular signaling cascades downstream of EGFR and FGFRs is PI3K/AKT signaling [37,38]. PI3K class IA enzymes are heterodimers composed of a regulatory subunit p85 (encoded by PIK3R1, PIK3R2 or PIK3R3) and a catalytic subunit p110 (encoded by PIK3CA, PIK3CB or PIK3CD) [39]. The catalytic p110 subunit of PI3K becomes activated when inhibitory contacts with the p85 regulatory subunit are relieved by p85 binding to phosphorylated adaptor proteins recruited to activated EGFR or FGFRs [40,41]. Active PI3K class I enzymes phosphorylate PIP2 (phosphatidylinositol-4,5-bisphosphate), converting it into PIP3 (phosphatidylinositol-3,4,5-trisphosphate), a reaction negatively regulated by PTEN phosphatase [42]. PIP3 serves as a second messenger that activates AKT (AKT1, AKT2 or AKT3) [43]. AKT family members are cytosolic and nuclear serine/threonine protein kinases involved in phosphorylation-mediated regulation of numerous proteins involved in cell survival and growth [39].

EGFR, FGFRs, PIK3CA, PIK3R1 and AKT1 are proto-oncogenes, frequently activated in cancer through gain-of-function mutations and/or overexpression. PTEN is an established tumor suppressor gene, with a frequent loss of function in cancer [44]. Gain-of-function mutations in EGFR [45,46] and FGFRs [47–51] usually act by conferring ligand-independent activation or by increasing tyrosine kinase catalytic activity. Mutations in PIK3R1 or PIK3CA abolish inhibitory interactions between the regulatory and catalytic subunit of PI3K [52–56], resulting in PI3K activity in the absence of growth factor stimulation. AKT1 gain of function mutations increase AKT1 affinity for PIP2, allowing AKT1 activation in the absence of PI3K activity and PIP3 generation [57]. PTEN loss-of-function mutations usually affect the phosphatase domain, impairing PTEN catalytic activity and removal of PIP3 [58].

Small molecule therapeutics and recombinant antibodies are being developed as potential treatments for cancers driven by increased activity of EGFR, FGFR and/or PI3K/AKT. Gefitinib and erlotinib, small tyrosine kinase inhibitors, are approved for the clinical treatment of cancers harboring specific EGFR mutations. A recombinant antibody, cetuximab, is approved for the clinical treatment of

cancers that overexpress wild-type EGFR [59]. Small molecules that inhibit the catalytic activity of FGFRs [60], PI3K and AKT [61] are currently undergoing clinical trials or are in pre-clinical development.

We have extended the Reactome data model and enhanced the web tools to permit the annotation and visualization of the altered biological behavior of protein variants. These enhancements can be applied to any molecular abnormality due to germline or somatic mutation, as well as to abnormalities due to expression of foreign proteins encoded by genomes of infectious agents like viruses or intracellular parasites.

2. Results and Discussion

2.1. Annotation of Cancer-Perturbed Pathways

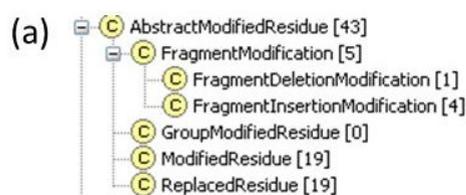
Pathways that stimulate cell growth, cell division and survival, and maintenance of undifferentiated state are activated in cancer through gain-of-function mutations in participating proto-oncogenes and/or their overexpression. On the other hand, pathways that negatively regulate cell division, growth and survival, or promote cellular differentiation are inactivated through loss-of-function mutations in tumor suppressor genes and/or their downregulation. To capture these two groups of cancer effectors, we have added new classes of data to the Reactome database.

2.1.1. Extension of Protein Modifications to Accommodate Annotation of Changes in Amino Acid Sequence

The protein modification class in the Reactome data model was constructed to support annotation of covalent co- and post-translational modifications of proteins such as the phosphorylation of serine residues. To allow for annotation of mutant proteins, two new subclasses of modifications were introduced: Replaced Residue and Fragment Modification (Figure 1a). The Replaced Residue class is used to annotate amino acid substitutions in a protein sequence. A Replaced Residue instance associates a specific coordinate of a protein sequence with two PSI-MOD ontology [62] attributes: the first identifies the amino acid found at that position in the normal protein and the second attribute identifies the amino acid that replaces it in the mutant protein. For example, the most frequently found mutation in EGFR is the substitution of a leucine residue at position 858 with an arginine residue in the catalytic domain of EGFR. This mutation disrupts autoinhibitory interactions, facilitating adoption of an active conformation [63]. The Reactome record for EGFR L858R (Figure 1b) indicates this amino acid substitution.

The FragmentModification subclass includes two subclasses, FragmentInsertionModification and FragmentDeletionModification. FragmentInsertionModification is used to annotate insertions of amino acid residues in a protein sequence. FragmentDeletionModification is used to annotate removal of amino acid residues. PIK3R1 Y463_L466del is a variant of the PI3K regulatory subunit p85alpha found in endometrial cancer (Figure 1c). This PIK3R1 mutant lacks four amino acid residues in the inter-SRC homology 2 (iSH2) domain. PIK3R1 is able to bind the catalytic subunit of PI3K, PIK3CA (p110alpha), but does not inhibit it, resulting in the constitutive activity of PI3K, in the absence of growth factors [64]. The deletion coordinates are indicated in the Reactome record for PIK3R1 Y463_L466del mutant.

Figure 1. Annotation of cancer mutations. Reactome records are not displayed in their entirety due to space limitations. **(a)** Subclasses of protein modifications contained in Reactome class Abstract Modified Residue. Currently, Reactome website displays all subclasses of protein modifications in the single field “Post-translational modification(s)”. Future changes to the website will allow chemical modifications to be distinguished from effects of mutations. **(b)** Reactome record for EGFR L858R caused by a missense mutation that replaces leucine residue at position 858 with arginine. **(c)** Reactome record for PIK3R1 Y463_L466del caused by an in-frame intragenic deletion in PIK3R1 that removes amino acid residues from position 463 to position 466, as captured by the Fragment Deletion Modification instance. **(d)** Reactome record for BCR-FGFR1 fusion protein. Truncation of the wild-type BCR protein sequence is shown by altered end coordinate. FGFR1 fragment fused to BCR is annotated as an insertion using FragmentInsertionModification class.



(b)

Name	EGFR L858R mutant Epidermal Growth Factor Receptor L858 EGFR Leu858Arg mutant
Stable identifier	REACT_116811.1
Links to corresponding entries in other databases	COSMIC:12979 COSMIC:6224 UniProt:P00533
CLASS	EntityWithAccessionedSequence
Other identifiers related to this sequence	EGFR_HUMAN, O00688, O00732, P062
Reference entity	UniProt:P00533 EGFR
Coordinates in the reference sequence	25 .. 1210
Post-translational modification(s)	L-leucine 858 replaced with L-arginine

(c)

Name	PIK3R1 Y463_L466del mutant PIK3R1 Tyr463_Leu466del mutant
Stable identifier	REACT_148365.1
Links to corresponding entries in other databases	COSMIC:87228 UniProt:P27986
CLASS	EntityWithAccessionedSequence
Other identifiers related to this sequence	P85A_HUMAN, D3DWA0, Q15747, Q4
Reference entity	UniProt:P27986 PIK3R1
Coordinates in the reference sequence	1 .. 724
Post-translational modification(s)	Deletion of residues 463 to 466

(d)

Name	BCR-FGFR1 fusion mutant BCR(1-584)-FGFR1(419-822) fusion t(8:22) BCR-FGFR1 fusion
Stable identifier	REACT_125113.1
Links to corresponding entries in other databases	UniProt:P11274
CLASS	EntityWithAccessionedSequence
Other identifiers related to this sequence	BCR_HUMAN, P78501, Q12842, Q4LE80, Q6NZ13
Reference entity	UniProt:P11274 BCR
Coordinates in the reference sequence	1 .. 584
Post-translational modification(s)	Insertion of residues 429 to 822 at 585 from UniProt:P11362 FGFR1

The FragmentModification class can also be used to annotate fusion proteins. For example, the translocation t(8;22)(p11;q11) in chronic myeloid leukemia produces a BCR-FGFR1 fusion that consists of the first four exons of BCR and exons 9–18 of FGFR1 [65]. The BCR-FGFR1 fusion protein is annotated as an Entity with Accessioned Sequence (Figure 1d) that consists of a truncated BCR protein, starting at position 1 and ending at position 584 of the reference UniProt sequence P11274 (human BCR). Then, a FragmentInsertionModification instance defines insertion of amino acids 429–822 of the UniProt reference sequence P11362 (human FGFR1) at position 585 of BCR (Figure 1d).

On the Reactome website, selecting a physical entity or an event node by clicking on a pathway diagram brings up a record for that particular instance in the details pane, which appears by clicking on the yellow triangle at the bottom of the Pathway Browser page. Selecting EGFRvIII in the diagram (Figure 2a), brings up Reactome information on this mutant protein, as well as interactive cross references that direct users to other Reactome website pages or other databases of interest (Figure 2b). Each cancer-related disease variant record cross-references available records in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (Table 1) [66]. The EGFRvIII record displayed on Reactome website links to COSMIC record 21351, which provides information on nucleotide sequence changes and tumor samples in which this mutation was reported.

2.1.2. Associating Disease Attributes with Physical Entities and Events

All physical entities related to disease variants, such as proteins, sets of proteins, and protein complexes are tagged with disease attributes (Table 1), using a term from the Disease Ontology (DO) [67]. This DO record provides, when possible, a link to the synonymous disease record in the National Cancer Institute Thesaurus (NCIt) [68]. The disease attribute of the physical entity is assigned to all reactions and pathways in which it participates.

Besides providing information on disease involvement of specific proteins and directing users to more detailed disease descriptions, a disease attribute annotation enables users to search Reactome database for proteins and events associated with a specific disease. For example, in Figure 2b, a DO instance “adult glioblastoma multiforme” is associated with EGFRvIII. Clicking on the “adult glioblastoma multiforme” link displayed on Reactome website (Figure 2b) provides a DO identifier for this disease instance (3075) and also lists all other proteins in Reactome database whose mutant forms are associated with adult glioblastoma multiforme (Figure 2c). Thus, Reactome provides cancer researchers with a quick access to cancer type-specific disease variants and information on the mechanism of action for each variant annotated.

2.1.3. Mode of Action and Specificity of Anti-Cancer Therapeutics

The Reactome data model allows for annotation of small molecules and antibodies used as anti-cancer therapeutics, as well as the annotation of their specific mode of action. We have annotated nine small tyrosine kinase inhibitors (TKIs) used to inhibit EGFR kinase activity in cancer [59,69], as well as the recombinant antibody cetuximab [70] (Figure 3a). In addition, we annotated five benzaquinoid ansamycins that inhibit the HSP90 chaperone protein that stabilizes EGFR mutant proteins [71], twelve anti-FGFR TKIs [60], one anti-FGFR recombinant antibody [72],

ten small molecules that inhibit the catalytic subunit of PI3K [61], and three small molecules that inhibit AKT [61] (Table 2).

Figure 2. Disease information presented interactively on the Reactome website. (a) Selecting an entity in the pathway diagram, EGFRvIII mutant in this case and opening the Reactome details pane by clicking on the yellow triangle at the bottom of the Pathway Browser page brings up a record for the selected instance. A pathway hierarchy displayed on the left hand side shows how the selected instance is related to the rest of the pathway content. (b) Reactome record, displayed in the details pane, provides information on the selected entity, including cross-references to other databases such as COSMIC, UniProt, GO. (c) Cross-reference to Disease Ontology: clicking on a disease attribute, such as “adult glioblastoma multiforme”, provides a Disease Ontology (DO) identifier for this disease instance (3075) and lists all proteins in Reactome database associated with adult glioblastoma multiforme.

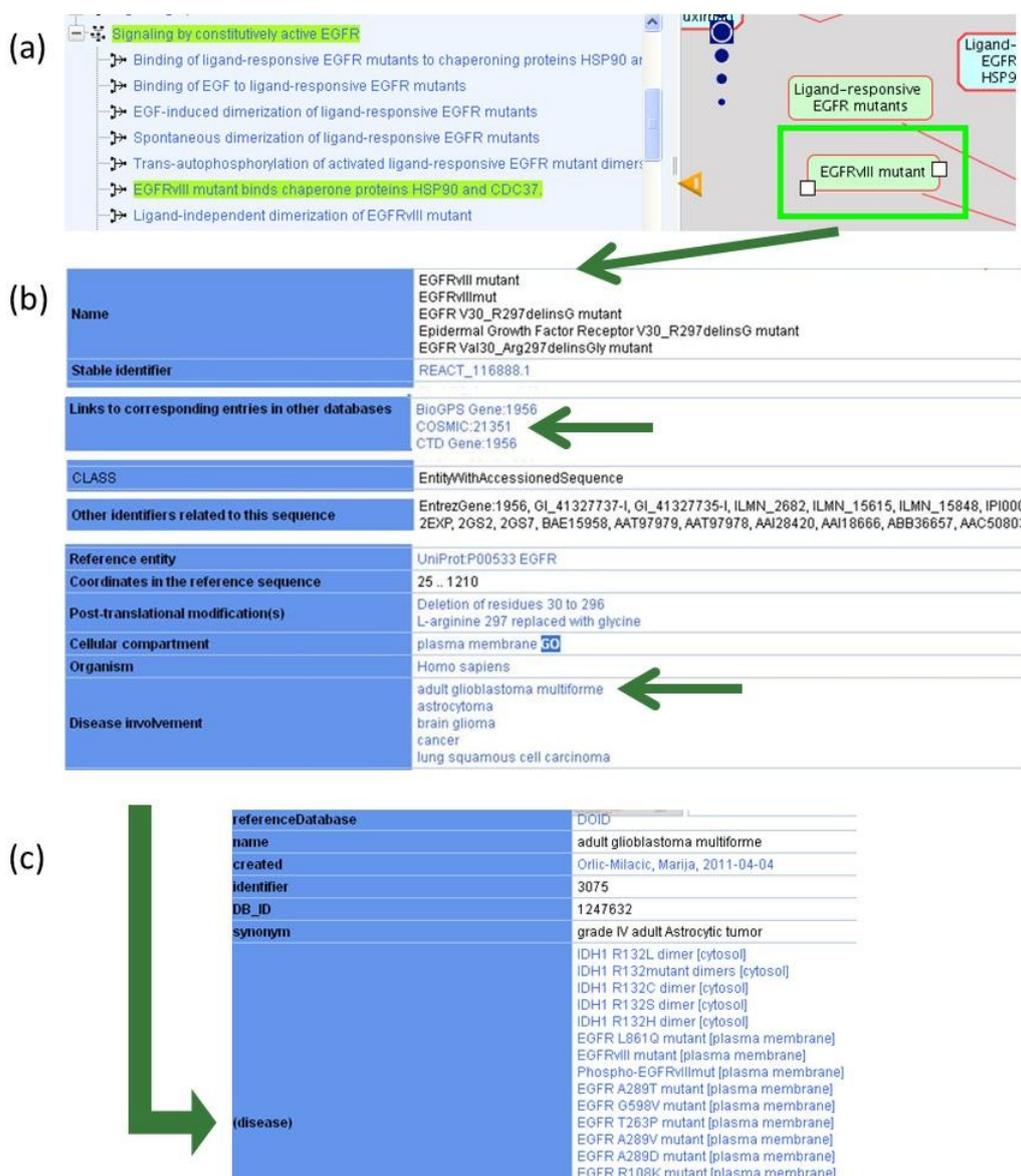
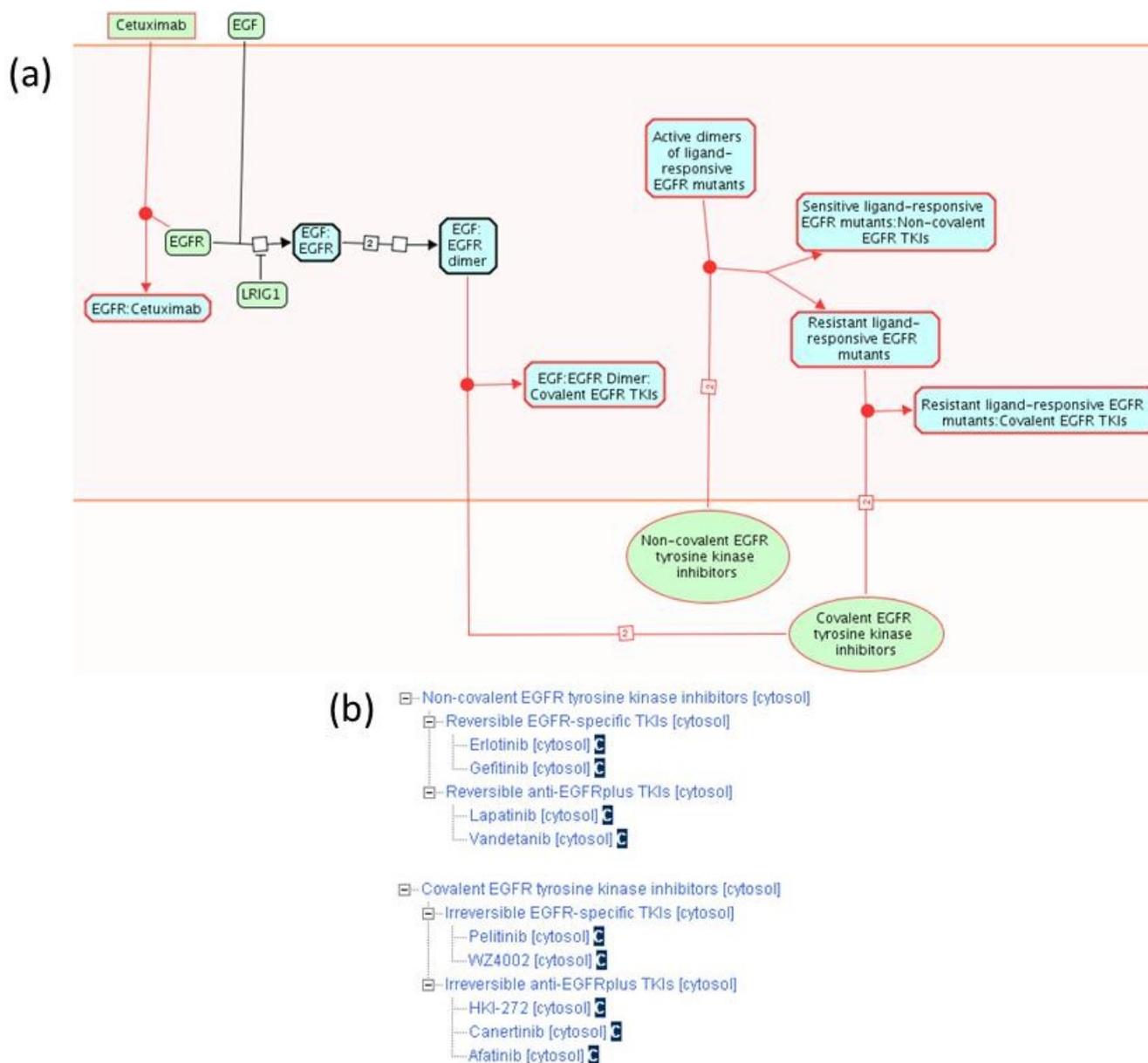


Figure 3. Mode of action and specificity of anti-EGFR cancer therapeutics. (a) Anti-EGFR therapeutics differ in their specificity for EGFR cancer variants, as well as in their mode of action (non-covalent vs. covalent binding). (b) Classification of EGFR-binding small tyrosine kinase inhibitors (TKIs) according to spectrum and reversibility of their binding.



For each anti-EGFR TKI, we specify whether it associates with the EGFR catalytic domain through formation of a covalent (irreversible) bond or through a non-covalent interaction (reversible). We also specify whether a TKI is EGFR-specific or whether it can inhibit other receptor tyrosine kinases besides EGFR (EGFRplus). Each small molecule instance we annotate is associated with the Chemical Entities of Biological Interest (ChEBI) database identifier [19]. On the Reactome website, a link to a corresponding ChEBI record is displayed after the name of each small molecule. Clicking on the ChEBI link associated with gefitinib (Figure 3b) directs the user to the gefitinib information in ChEBI, displaying its molecular structure and additional information not directly captured by Reactome.

EGFR cancer mutants in Reactome are classified into sets based on their sensitivity to various TKIs (Figure 3a). Ligand responsive EGFR mutants sensitive to non-covalent TKIs can be inhibited by low concentrations of non-covalent (reversible) TKIs that do not significantly affect the function of wild-type EGFR and therefore produce minimal side effects. Ligand responsive EGFR mutants resistant to non-covalent TKIs can be inhibited by covalent (irreversible) TKIs. As can be seen from the diagram (Figure 3a), concentrations of irreversible TKIs that inhibit EGFR mutants also inhibit the function of the wild-type protein, causing more severe side effects, as described in event summations. Cetuximab is used for treatment of cancers that overexpress wild-type EGFR protein, usually due to amplification of the EGFR locus [59,70].

Table 1. Cancer-related disease variants published by Reactome to date. A total of ~150 cancer mutants have been published since the start of the project in December 2010.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
EGFR A289D mutant	21685	Missense	Glioblastoma	Signaling by EGFR in Cancer
EGFR A289T mutant	21686	Missense	Glioblastoma, oligodendroglioma	Signaling by EGFR in Cancer
EGFR A289V mutant	21687	Missense	Glioblastoma	Signaling by EGFR in Cancer
EGFR D770_N771insNPG mutant		Insertion	Lung cancer	Signaling by EGFR in Cancer
EGFR D770_N771insNPH mutant	48920	Insertion	Lung cancer	Signaling by EGFR in Cancer
EGFR E746_A750del mutant	6223, 129800, 6225	Deletion	Breast, head and neck, kidney, lung, ovarian, salivary gland and thyroid cancer	Signaling by EGFR in Cancer
EGFR E746_A750del; T790M mutant		Deletion; Missense	Lung cancer	Signaling by EGFR in Cancer
EGFR E746_S752delinsV mutant	18492, 18426, 12384, 85797	Deletion	Lung cancer	Signaling by EGFR in Cancer
EGFR E746_T751delinsA mutant	20845, 12678, 13549	Deletion	Head and neck, lung cancer	Signaling by EGFR in Cancer
EGFR G598V mutant	34167, 21690	Missense	Glioblastoma	Signaling by EGFR in Cancer
EGFR G719A mutant	6239, 13448	Missense	Lung cancer	Signaling by EGFR in Cancer
EGFR G719C mutant	6253, 20881	Missense	Lung cancer	Signaling by EGFR in Cancer
EGFR G719S mutant	6252, 13983	Missense	Colorectal, lung cancer	Signaling by EGFR in Cancer
EGFR L747_A750delinsP mutant	13562, 12382, 12422	Deletion	Head and neck, lung cancer	Signaling by EGFR in Cancer

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
EGFR L747_P753delinsS mutant	13564, 12370	Deletion	Head and neck, lung, prostate cancer	Signaling by EGFR in Cancer
EGFR L747_S752del mutant	13984, 6255	Deletion	Lung cancer	Signaling by EGFR in Cancer
EGFR L747_T751del mutant	24432, 12369, 6254, 23571	Deletion	Lung cancer	Signaling by EGFR in Cancer
EGFR L747_T751delinsP mutant	24573, 12383, 22944	Deletion	Lung cancer	Signaling by EGFR in Cancer
EGFR L858R mutant	6224, 12979	Missense	Breast, lung, ovarian, stomach, thymus and thyroid cancer, mesothelioma	Signaling by EGFR in Cancer
EGFR L858R;T790M mutant		Missense; Missense	Lung cancer	Signaling by EGFR in Cancer
EGFR L861Q mutant	6213, 13173	Missense	Lung cancer, glioblastoma	Signaling by EGFR in Cancer
EGFR M766_A767insASV mutant		Insertion	Lung cancer	Signaling by EGFR in Cancer
EGFR R108K mutant	21683, 34166	Missense	Glioblastoma	Signaling by EGFR in Cancer
EGFR T263P mutant	21684	Missense	Glioblastoma	Signaling by EGFR in Cancer
EGFR V738_K739insKIPVAI mutant		Insertion	Lung cancer	Signaling by EGFR in Cancer
EGFRvIII mutant	21351	Deletion	Lung cancer, glioblastoma	Signaling by EGFR in Cancer
BCR-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
CNTRL-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
CPSF6-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
CUX1-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
FGFR1 K656E mutant	35673	Missense	Glioblastoma	Signaling by FGFR in Disease
FGFR1 N546K mutant	19176	Missense	Glioblastoma, hypochondroplasia	Signaling by FGFR in Disease
FGFR1 P252S mutant		Missense	Melanoma	Signaling by FGFR in Disease
FGFR1 P252T mutant	12834	Missense	Lung cancer	Signaling by FGFR in Disease

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
FGFR1 R576W mutant	19177	Missense	Glioblastoma	Signaling by FGFR in Disease
FGFR1OP-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
FGFR1OP2-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
FGFR1c P252R mutant		Missense	Pfeiffer syndrome	Signaling by FGFR in Disease
FGFR2 K660E mutant	36909	Missense	Endometrial cancer	Signaling by FGFR in Disease
FGFR2 K660M mutant	49175	Missense	Cervical cancer	Signaling by FGFR in Disease
FGFR2 K660N mutant	49173	Missense	Endometrial cancer, Crouzon syndrome	Signaling by FGFR in Disease
FGFR2 L764fs*4STOP mutant		Frameshift	Endometrial cancer	Signaling by FGFR in Disease
FGFR2 N549H mutant		Missense	Crouzon syndrome	Signaling by FGFR in Disease
FGFR2 N549K mutant	36912, 36902	Missense	Endometrial cancer	Signaling by FGFR in Disease
FGFR2 S267P mutant		Missense	Stomach cancer, Crouzon syndrome	Signaling by FGFR in Disease
FGFR2 W290C mutant	41286	Missense	Lung cancer, Pfeiffer syndrome	Signaling by FGFR in Disease
FGFR2b P253R mutant	49170	Missense	Endometrial cancer, acrocephalosyndactylia	Signaling by FGFR in Disease
FGFR2b S252W mutant	36903, 41289	Missense	Endometrial, ovarian cancer, acrocephalosyndactylia, craniosynostosis	Signaling by FGFR in Disease
FGFR2b S373C mutant	36905	Missense	Endometrial cancer	Signaling by FGFR in Disease
FGFR2b Y376C mutant	36904, 41290	Missense	Endometrial, ovarian cancer	Signaling by FGFR in Disease
FGFR2c A314D mutant	49171	Missense	Endometrial cancer	Signaling by FGFR in Disease
FGFR2c A314S mutant		Missense	Bone development disease	Signaling by FGFR in Disease
FGFR2c A315S mutant		Missense	Syndactyly	Signaling by FGFR in Disease
FGFR2c A315T mutant	30777	Missense	Endometrial cancer	Signaling by FGFR in Disease

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
FGFR2c P253R mutant	49170	Missense	Endometrial cancer, acrocephalosyndactylia	Signaling by FGFR in Disease
FGFR2c S252W mutant	41289, 36903	Missense	Endometrial, ovarian cancer, acrocephalosyndactylia, craniosynostosis	Signaling by FGFR in Disease
FGFR2c S372C mutant		Missense	Beare-Stevenson cutis gyrata syndrome	Signaling by FGFR in Disease
FGFR2c W290G mutant		Missense	Crouzon syndrome, Pfeiffer syndrome	Signaling by FGFR in Disease
FGFR2c Y375C mutant		Missense	Beare-Stevenson cutis gyrata syndrome	Signaling by FGFR in Disease
FGFR3 795fs*139STOP mutant		Frameshift	Multiple myeloma, thanatophoric dysplasia	Signaling by FGFR in Disease
FGFR3 A391E mutant	721	Missense	Bladder cancer, Crouzon syndrome	Signaling by FGFR in Disease
FGFR3 G370C mutant	716, 35897	Missense	Bladder cancer, thanatophoric dysplasia	Signaling by FGFR in Disease
FGFR3 G380R mutant	24842, 24812	Missense	Bladder cancer, multiple myeloma, achondroplasia	Signaling by FGFR in Disease
FGFR3 G382D mutant	727	Missense	Multiple myeloma	Signaling by FGFR in Disease
FGFR3 K650E mutant	719, 35899	Missense	Bladder, testicular cancer, multiple myeloma, thanatophoric dysplasia	Signaling by FGFR in Disease
FGFR3 K650M mutant	720, 85791	Missense	Bladder, testicular cancer, multiple myeloma, thanatophoric dysplasia	Signaling by FGFR in Disease
FGFR3 K650N mutant		Missense	Bladder, testicular cancer, hypochondroplasia	Signaling by FGFR in Disease
FGFR3 K650Q mutant	726	Missense	Bladder cancer, hypochondroplasia	Signaling by FGFR in Disease
FGFR3 K650T mutant	731	Missense	Bladder, testicular cancer, hypochondroplasia	Signaling by FGFR in Disease
FGFR3 R248C mutant	714, 35896	Missense	Bladder cancer, multiple myeloma, thanatophoric dysplasia	Signaling by FGFR in Disease
FGFR3 S371C mutant	17461, 35898	Missense	Bladder cancer, thanatophoric dysplasia	Signaling by FGFR in Disease
FGFR3 Y373C mutant	718, 29428	Missense	Bladder cancer, multiple myeloma, thanatophoric dysplasia	Signaling by FGFR in Disease

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
FGFR3b G697C mutant	24802	Missense	Head and neck cancer	Signaling by FGFR in Disease
FGFR3b S249C mutant	715, 29427	Missense	Bladder, cervical, head and neck, prostate cancer, thanatophoric dysplasia	Signaling by FGFR in Disease
FGFR3c P250R mutant		Missense	Acrocephalosyndactylia, craniosynostosis	Signaling by FGFR in Disease
FGFR4 N535D mutant		Missense	Rhabdomyosarcoma	Signaling by FGFR in Disease
FGFR4 N535K mutant		Missense	Rhabdomyosarcoma	Signaling by FGFR in Disease
FGFR4 V550E mutant		Missense	Rhabdomyosarcoma	Signaling by FGFR in Disease
FGFR4 V550L mutant		Missense	Rhabdomyosarcoma	Signaling by FGFR in Disease
FGFR4 Y367C mutant		Missense	Breast cancer	Signaling by FGFR in Disease
LRRFIP1-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
MYO18A-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
TRIM24-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
ZMYM2-FGFR1 fusion mutant		Translocation	Myeloid leukemia	Signaling by FGFR in Disease
IDH1 R132C mutant	28747, 41294	Missense	Glioblastoma	The citric acid (TCA) cycle and respiratory electron transport
IDH1 R132H mutant	28746, 41291	Missense	Glioblastoma	The citric acid (TCA) cycle and respiratory electron transport
IDH1 R132L mutant	28750	Missense	Glioblastoma	The citric acid (TCA) cycle and respiratory electron transport
IDH1 R132S mutant	28748	Missense	Glioblastoma	The citric acid (TCA) cycle and respiratory electron transport

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
PIK3CA E542K mutant	760, 29329	Missense	Bladder, breast, cervical, colorectal, endometrial, esophageal, gallbladder, head and neck, kidney, liver, lung, ovarian, penis, pharynx, pituitary, skin sweat gland, stomach, thyroid cancer, glioblastoma, lymphocytic leukemia	PI3K/AKT Signaling in Cancer
PIK3CA E542Q mutant	17442	Missense	Breast, colorectal, endometrial, lung cancer	PI3K/AKT Signaling in Cancer
PIK3CA E542V mutant	762	Missense	Breast, colorectal, endometrial, ovarian cancer	PI3K/AKT Signaling in Cancer
PIK3CA E545A mutant	12458	Missense	Breast, colorectal, endometrial, esophageal, lung, ovarian, prostate, thyroid cancer, glioblastoma, hepatoblastoma, synovial sarcoma	PI3K/AKT Signaling in Cancer
PIK3CA E545G mutant	764	Missense	Bladder, breast, colorectal, endometrial, head and neck, larynx, pituitary, stomach cancer, myeloid leukemia, non-Hodgkin lymphoma, retinoblastoma	PI3K/AKT Signaling in Cancer
PIK3CA E545K mutant	763, 29328	Missense	Bladder, breast, cervical, colorectal, endometrial, esophageal, gallbladder, head and neck, kidney, lung, ovarian, pancreatic, penis, pharynx, skin, stomach, sweat gland, thyroid cancer, melanoma, glioblastoma, medulloblastoma, myeloma, pituitary adenoma	PI3K/AKT Signaling in Cancer

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
PIK3CA E545Q mutant	27133	Missense	Bladder, breast, colorectal, esophageal, head and neck, ovarian, thyroid cancer	PI3K/AKT Signaling in Cancer
PIK3CA E545V mutant	144201	Missense	Ovarian cancer	PI3K/AKT Signaling in Cancer
PIK3CA H1047L mutant	776, 30744	Missense	Bladder, breast, colorectal, endometrial, head and neck, liver, lung, ovarian, pharynx, thyroid cancer, glioblastoma, non-Hodgkin lymphoma, pituitary adenoma	PI3K/AKT Signaling in Cancer
PIK3CA H1047R mutant	775, 29325	Missense	Bladder, breast, cervical, colorectal, endometrial, gallbladder, head and neck, liver, lung, ovarian, pancreatic, pharynx, prostate, stomach, thyroid cancer, glioblastoma, medulloblastoma, melanoma, meningioma, neuroectodermal tumor, non-Hodgkin lymphoma, pituitary adenoma	PI3K/AKT Signaling in Cancer
PIK3CA H1047Y mutant	774, 29326	Missense	Breast, colorectal, endometrial, lung, ovarian cancer, glioblastoma	PI3K/AKT Signaling in Cancer
PIK3CA M1043I mutant	773, 29313, 94984	Missense	Bladder, breast, cervical, colorectal, endometrial, lung, ovarian, thyroid cancer, glioblastoma	PI3K/AKT Signaling in Cancer
PIK3CA M1043T mutant	12463	Missense	Ovarian, stomach cancer, glioblastoma	PI3K/AKT Signaling in Cancer
PIK3CA M1043V mutant	12591, 30743	Missense	Breast, colorectal, endometrial, head and neck, lung, ovarian, pharynx cancer, glioblastoma	PI3K/AKT Signaling in Cancer

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
PIK3CA Q546E mutant	6147	Missense	Breast, cervical, colorectal, endometrial cancer	PI3K/AKT Signaling in Cancer
PIK3CA Q546H mutant	24712, 30740	Missense	Cervical, colorectal, endometrial cancer	PI3K/AKT Signaling in Cancer
PIK3CA Q546K mutant	766, 30738	Missense	Breast, colorectal, endometrial, lung, ovarian, stomach cancer, lymphocytic leukemia	PI3K/AKT Signaling in Cancer
PIK3CA Q546L mutant	25041, 85754	Missense	Breast, colorectal, gallbladder, head and neck cancer	PI3K/AKT Signaling in Cancer
PIK3CA Q546P mutant	767	Missense	Breast, colorectal, endometrial, ovarian cancer, glioma	PI3K/AKT Signaling in Cancer
PIK3CA Q546R mutant	12459, 30739	Missense	Breast, colorectal, endometrial, prostate, stomach cancer	PI3K/AKT Signaling in Cancer
PIK3CA R38C mutant	744	Missense	Colorectal cancer	PI3K/AKT Signaling in Cancer
PIK3CA R38G mutant	40945	Missense	Glioblastoma	PI3K/AKT Signaling in Cancer
PIK3CA R38H mutant	745, 49022	Missense	Breast, colorectal, endometrial cancer	PI3K/AKT Signaling in Cancer
PIK3CA R38S mutant	87310	Missense	Stomach cancer	PI3K/AKT Signaling in Cancer
PIK3R1 D560H mutant	125378	Missense	Pharynx cancer	PI3K/AKT Signaling in Cancer
PIK3R1 D560Y mutant	335765	Missense	Glioblastoma	PI3K/AKT Signaling in Cancer
PIK3R1 G376R mutant	35827, 132923	Missense	Endometrial cancer, glioblastoma	PI3K/AKT Signaling in Cancer
PIK3R1 H450_E451del mutant	39296	Deletion	Endometrial cancer, glioblastoma	PI3K/AKT Signaling in Cancer

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
PIK3R1 K459del mutant	87216	Deletion	Endometrial cancer	PI3K/AKT Signaling in Cancer
PIK3R1 N564D mutant	42912	Missense	Colorectal, endometrial cancer, glioblastoma	PI3K/AKT Signaling in Cancer
PIK3R1 N564K mutant	35808	Missense	Glioblastoma	PI3K/AKT Signaling in Cancer
PIK3R1 R574_T576del mutant	87219	Deletion	Endometrial cancer	PI3K/AKT Signaling in Cancer
PIK3R1 R574I mutant	85927	Missense	Colorectal cancer	PI3K/AKT Signaling in Cancer
PIK3R1 R574T mutant	87544	Missense	Bladder, breast cancer	PI3K/AKT Signaling in Cancer
PIK3R1 Y463_L466del mutant	87228	Deletion	Endometrial cancer	PI3K/AKT Signaling in Cancer
AKT1 E17K mutant	33765, 34142	Missense	Breast, colorectal, ovarian cancer	PI3K/AKT Signaling in Cancer
PTEN R130G mutant	5219	Missense	Endometrial, lung, ovarian cancer, glioblastoma	PI3K/AKT Signaling in Cancer
PTEN R130Q mutant	5033	Missense	Breast, colorectal, endometrial, ovarian, thyroid cancer, glioma, histiocytoma	PI3K/AKT Signaling in Cancer
PTEN R130L mutant	5216	Missense	Breast, endometrial cancer, Cowden syndrome	PI3K/AKT Signaling in Cancer
PTEN C124S mutant	5224, 5271	Missense	Endometrial, thyroid cancer, glioblastoma	PI3K/AKT Signaling in Cancer
PTEN C124R mutant		Missense	Thyroid adenoma, Cowden syndrome	PI3K/AKT Signaling in Cancer
PTEN R173H mutant	5039	Missense	Endometrial, ovarian cancer, glioma	PI3K/AKT Signaling in Cancer

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
PTEN R173C mutant	5089, 24682	Missense	Endometrial cancer, glioblastoma, lymphocytic leukemia, melanoma	PI3K/AKT Signaling in Cancer
PTEN R173P mutant	12735	Missense	Testicular cancer	PI3K/AKT Signaling in Cancer
PTEN S170N mutant	5045	Missense	Endometrial cancer, glioblastoma	PI3K/AKT Signaling in Cancer
PTEN S170R mutant		Missense	Bannayan-Riley-Ruvalcaba syndrome	PI3K/AKT Signaling in Cancer
PTEN H123Y mutant	5078	Missense	Endometrial cancer	PI3K/AKT Signaling in Cancer
PTEN G129E mutant	28917	Missense	Endometrial cancer	PI3K/AKT Signaling in Cancer
PTEN G129R mutant	5092	Missense	Thyroid cancer, glioblastoma	PI3K/AKT Signaling in Cancer
PTEN H93Y mutant	5043	Missense	Endometrial cancer, glioma, medulloblastoma	PI3K/AKT Signaling in Cancer
PTEN H93A mutant		Missense	Cancer	PI3K/AKT Signaling in Cancer
PTEN H93R mutant	5060	Missense	Glioblastoma, autism spectrum disorders	PI3K/AKT Signaling in Cancer
PTEN H93D mutant	5283	Missense	Endometrial cancer	PI3K/AKT Signaling in Cancer
PTEN H93Q mutant	5186	Missense	Glioblastoma	PI3K/AKT Signaling in Cancer
PTEN R130P mutant	5277	Missense	Breast, endometrial, glioblastoma	PI3K/AKT Signaling in Cancer
PTEN C124F mutant	13578	Missense	Lung cancer	PI3K/AKT Signaling in Cancer

Table 1. Cont.

Disease Variant	COSMIC Identifier(s)	Mutation Type	Disease	Reactome Pathway Name
PTEN C124Y mutant	5140	Missense	Stomach cancer	PI3K/AKT Signaling in Cancer
PTEN S170I mutant	5218	Missense	Glioblastoma	PI3K/AKT Signaling in Cancer
PTEN S170G mutant	5063	Missense	Glioblastoma	PI3K/AKT Signaling in Cancer
PTEN G129V mutant	5276	Missense	Endometrial cancer	PI3K/AKT Signaling in Cancer
PTEN R130* mutant	21342, 5152	Nonsense	Cervical, colorectal, endometrial, lung, ovarian, prostate, thyroid cancer, glioblastoma, medulloblastoma, leiomyosarcoma	PI3K/AKT Signaling in Cancer
PTEN R233* mutant	5154, 21343	Nonsense	Cervical, colorectal, endometrial, lung, ovarian cancer, glioblastoma, histiocytoma, lymphocytic leukemia,	PI3K/AKT Signaling in Cancer
PTEN R335* mutant	5775, 5151	Nonsense	Head and neck, stomach cancer, glioblastoma, melanoma, Burkitt lymphoma, lymphocytic leukemia	PI3K/AKT Signaling in Cancer

Table 2. Anti-cancer therapeutics published by Reactome to date. A total of 39 small molecule inhibitors and 2 recombinant antibodies have been published since the start of the project in December 2010.

Anti-Cancer Therapeutic	Reference Molecule Identifier	Specificity	Reactome Pathway Name
17-AAG	ChEBI ^a :64153	HSP90	Signaling by EGFR in Cancer
17-DMAG	ChEBI:65324	HSP90	Signaling by EGFR in Cancer
Afatinib	ChEBI:61390	EGFR, ERBB2	Signaling by EGFR in Cancer
Canertinib	ChEBI:61399	Pan-ERBB	Signaling by EGFR in Cancer
Cetuximab	Recombinant antibody	EGFR	Signaling by EGFR in Cancer
Erlotinib	ChEBI:114785	EGFR	Signaling by EGFR in Cancer
Gefitinib	ChEBI:49668	EGFR	Signaling by EGFR in Cancer
Geldanamycin	ChEBI:5292	HSP90	Signaling by EGFR in Cancer
HKI-272	ChEBI:61390	EGFR, ERBB2	Signaling by EGFR in Cancer

Table 2. Cont.

Anti-Cancer Therapeutic	Reference Molecule Identifier	Specificity	Reactome Pathway Name
Herbimycin A	ChEBI:5674	HSP90	Signaling by EGFR in Cancer
Lapatinib	ChEBI:49603	EGFR, ERBB2	Signaling by EGFR in Cancer
Pelitinib	ChEBI:38927	EGFR	Signaling by EGFR in Cancer
Vandetanib	ChEBI:49960	EGFR, VEGFR	Signaling by EGFR in Cancer
WZ4002	ChEBI:61400	EGFR	Signaling by EGFR in Cancer
IPI-504	Pending	HSP90	Signaling by EGFR in Cancer
AZ 2171	ChEBI:556867	FGFR, PDGFR, VEGFR, KIT	Signaling by FGFR in Disease
Brivanib	ChEBI:443041	FGFR, VEGFR	Signaling by FGFR in Disease
Brivanib alaninate	ChEBI:270995	FGFR, VEGFR	Signaling by FGFR in Disease
Dovitinib	ChEBI:594834	FGFR, FLT3, VEGFR, PDGFR, KIT, CSFR	Signaling by FGFR in Disease
E3810	Pending	FGFR, VEGFR	Signaling by FGFR in Disease
E7080	ChEBI:816009	FGFR VEGFR, PDGFR	Signaling by FGFR in Disease
Masitinib	ChEBI:63450	FGFR3, PDGFR, KIT	Signaling by FGFR in Disease
GP369	Recombinant antibody	FGFR2b	Signaling by FGFR in Disease
Midostaurin	ChEBI:63452	FGFR, FLT3, PDGFR, VEGFR, KIT, PKCA	Signaling by FGFR in Disease
PD173074	ChEBI:63448	Pan-FGFR	Signaling by FGFR in Disease
AZD4547	ChEBI:63453	Pan-FGFR	Signaling by FGFR in Disease
BGJ398	ChEBI:63451	Pan-FGFR	Signaling by FGFR in Disease
SU5402	ChEBI:63449	FGFR, VEGFR	Signaling by FGFR in Disease
GSK1059615	Pending	Pan-PI3K	PI3K/AKT Signaling in Cancer
BEZ235	Pending	PI3K Class I, mTOR	PI3K/AKT Signaling in Cancer
BGT226	Pending	PI3K Class I, mTOR	PI3K/AKT Signaling in Cancer
BKM120	Pending	PI3K Class I	PI3K/AKT Signaling in Cancer

Table 2. Cont.

Anti-Cancer Therapeutic	Reference Molecule Identifier	Specificity	Reactome Pathway Name
XL765	Pending	PI3K Class I, mTOR	PI3K/AKT Signaling in Cancer
XL147	Pending	PI3K Class I	PI3K/AKT Signaling in Cancer
GDC0941	ChEBI:65326	PI3K Class I	PI3K/AKT Signaling in Cancer
PX-866	ChEBI:65345	PIK3CA, PIK3CD, PIK3CG	PI3K/AKT Signaling in Cancer
LY294002	ChEBI:65329	Pan-PI3K	PI3K/AKT Signaling in Cancer
wortmannin	ChEBI:52289	Pan-PI3K	PI3K/AKT Signaling in Cancer
Perifosine	ChEBI:428891	AKT	PI3K/AKT Signaling in Cancer
MK2206	ChEBI:716367	AKT	PI3K/AKT Signaling in Cancer
Triciribine	ChEBI:65310	AKT	PI3K/AKT Signaling in Cancer

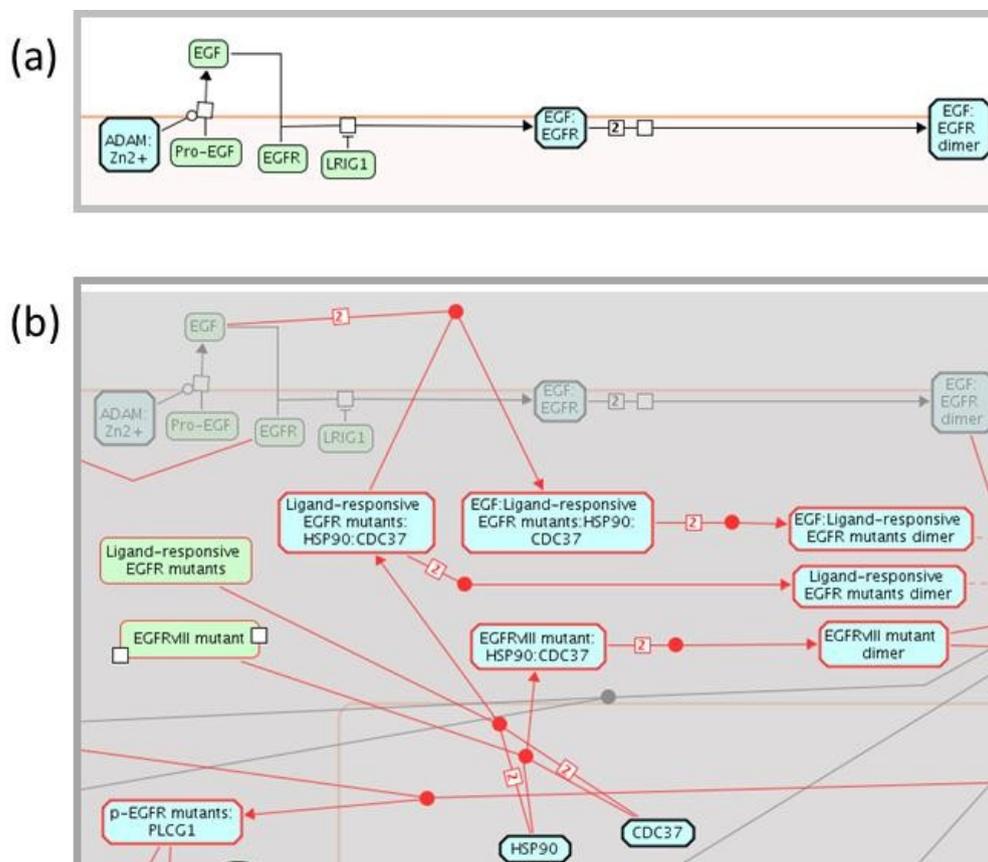
2.2. Other Disease Pathways in Reactome

In addition to cancer, Reactome also collects and provides information on communicable diseases. Currently featured infection-related Reactome pathways are “HIV Infection”, “Influenza Infection”, “Botulinum Neurotoxicity”, and “Latent Infection with Mycobacterium tuberculosis”. The pathway “Signaling by FGFR in Disease” contains, besides information on FGFR in cancer, the information on FGFR mutations and their functional implication in various developmental disorders, such as Pfeiffer syndrome and Crouzon syndrome. Reactome has recently published “Abnormal Metabolism in Phenylketonuria” and “Mucopolysaccharidoses” pathways, thereby introducing metabolic genetic diseases.

2.3. Enhancing the Reactome Pathway Browser for Display of Disease Variants

The Reactome Pathway Browser, based upon the Systems Biology Graphical Notation (SBGN) [73], permits the navigation and analysis of Reactome data, in a similar manner to Google Maps. SBGN is a standard graphical representation of biological pathway and network models. The Pathway Browser was adapted to enable display of disease variants and disease-related events involving proteins. A pathway diagram is shared between a wild-type pathway, for example “Signaling by EGFR”, and the corresponding disease pathway, “Signaling by EGFR in Cancer”. A disease attribute, attached to events involving cancer, instructs the browser to hide disease events when a user selects a wild-type pathway view (Figure 4a). When a user selects a disease pathway view, disease events appear in the diagram while all normal events are shaded gray. All disease events and physical entities with disease tags are outlined in red for easier visualization (Figure 4b).

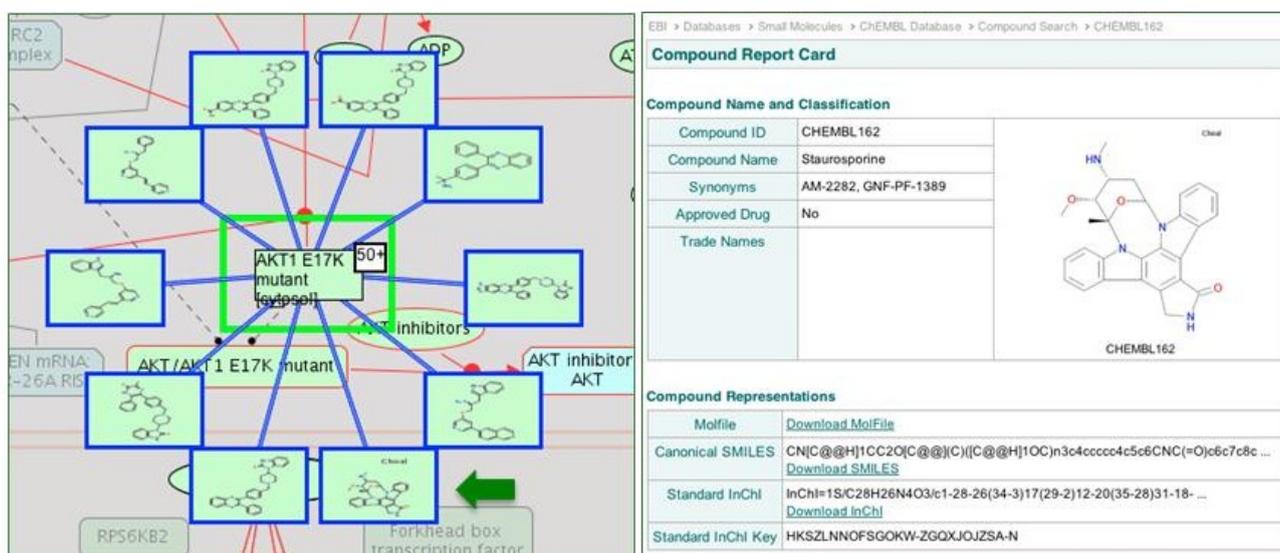
Figure 4. Display of wild-type and disease pathway diagrams. **(a)** A cancer disease attribute, assigned to events involved in cancer, instructs the browser to hide disease events when a user selects to view a wild-type pathway. **(b)** When a user selects to view a disease pathway, disease events appear in the wild type diagram, while all normal events are shaded. All disease events and physical entities with disease tags are outlined in red for easier visualization.



Physical entity and reaction nodes within the pathway diagrams are interactive. Clicking on either feature displays specific information and additional links out to external databases in the “Details” Panel, which opens by clicking on the yellow triangle at the bottom of the Pathway Browser page (Figure 2). Context sensitive menus, accessible through the right click on a selected entity, provide additional information about the physical entity in the pathway: a catalogue of other pathways in Reactome in which the selected entity participates; a list of the entities that contribute to the macromolecular complex; a catalogue of interactors of the selected entity; and the option to export a list of interactors of the selected entity. The latter two features of the context sensitive menu increase protein coverage and associated variant annotations. The Molecular Interaction Overlay (MI Overlay), accessible through “Analyze, Annotate & Upload” button of the Pathway Browser, displays proteins interacting with the manually annotated protein components of a Reactome pathway. This network overlay tool employs PSICQUIC (Proteomics Standard Initiative Common QUery InterfaCe) to apply an interactive display of interaction data from an external database such as IntAct [74] into Reactome pathway diagrams. Other sources of interaction data include protein-protein and protein-drug/small molecule interactions; a user-supplied list can also be displayed. By displaying interaction data from

ChEMBL, a database of bioactive drug-like molecules (Figure 5) [75], the MI Overlay feature provides an opportunity to identify protein variant-drug interactions, identify novel cancer targets or off-target effects, or pharmaceuticals that can moderate perturbed reactions or pathways experimentally.

Figure 5. AKT1 E17K mutant-small molecule interactions. When ChEMBL is selected as the interaction database, the MI Overlay displays small molecules from ChEMBL as interactors of AKT1 E17K variant protein of the PI3K/AKT Signaling in Cancer pathway. The nodes of the mini network are interactive; clicking the node to the left of the green arrow will link out to the Staurosporine protein kinase inhibitor record at ChEMBL.



2.4. Reactome Cancer-Perturbed Pathways Support Pathway Visualization and Analysis

The Pathway Browser provides an intuitive and interactive pathway visualization system, promoting a variety of web-based data analyses of user-supplied experimental data. The Pathway Analysis tool provides two alternate functions to analyze lists of genes. First, in the identifier (ID) mapping mode, a user-supplied set of gene or protein identifiers can be mapped to Reactome events. Second, in the overrepresentation analysis mode, users can determine which pathways are statistically overrepresented in a gene/protein list. The Expression Analysis tool will aid with the biological interpretation of large-scale cancer genome sequencing, genomics and proteomics experiments. For example, this tool allows users to visualize expression data (or any other numeric value, e.g., differential expression) superimposed on the Reactome pathway diagram. Reactome applies an orthology-based computational algorithm to curated human data to infer pathways in 22 diverse model organisms. The Species Comparison tool allows users to visually compare and contrast human pathways with these predicted model organism pathways. As additional cancer-perturbed pathways are added to Reactome, this method of “inferred” curation will provide a platform from which to study molecular disease mechanisms across the evolutionary spectrum. Reactome data is available for downloading and manipulation by third party visualization and analysis tools, including Cytoscape, Vanted and CellDesigner [76–78].

3. Experimental Section

Using the previously curated human EGFR pathway, which included a number of annotations for EGFR and downstream signaling by SHC1, GRB2, PLCG1 and CBL, as a template from which to extend the EGFR pathway, we imported this dataset into the Reactome Curator Tool [11]. Briefly, the curator tool provides Reactome curators with all the necessary tools to access the Reactome database and annotate data in agreement with the Reactome data model. Curators identified research articles and reviews in PubMed that were relevant to the annotation of the cancer-perturbed EGFR, FGFR and PI3K/AKT pathways. Once publications had been reviewed, a list of cancer-related proteins, small molecules and macromolecular complexes was prepared. Additional queries were performed in UniProt and ChEBI to identify the reference entity proteins and small molecules, respectively that would be used to construct the reactions of the cancer-perturbed EGFR pathway. Additional attributes of a reaction were captured. For example, details of the input and output entity(s), the catalytic or regulatory protein(s), the cellular location(s) of the reactants, a textual summation describing the reaction and the supporting literature reference(s). The Disease Ontology terms that match literature references and COSMIC records for annotated cancer variants were assigned as disease attributes to physical entities and events involving these mutant proteins. Oncogenic overexpression of proteins as a consequence of gene amplification is usually not explicitly shown in pathway diagrams, but is captured in text summations that accompany cancer pathways.

4. Conclusions

Reactome is a highly reliable, curated database of biological pathways. Through our website, we provide access to pathway and network data analysis tools for visualizing pathway data and interpreting experimental data sets. All Reactome data and software is openly available with no licensing required.

In view of the potential applicability of pathway and network analyses to identify and characterize novel cancer targets, Reactome has integrated and expanded the pathway gene product-function annotation and pathway curation to promote comprehensive and effective characterization of cancer targets, their related relationships and pathways. Our curation efforts thus far have focused on the EGFR pathway (including the EGFR, ERBB2, ERBB3, ERBB4 receptors), FGFR and PI3K-AKT signaling and their downstream effector genes. Reactome curators will enhance our curation of other cancer-perturbed pathways, such as apoptosis, cell cycle checkpoints, and other signaling pathways, including BMP, PDGF, NOTCH, VEGF, WNT, Rho-GTPase, and TGF-beta. Furthermore, as the Ontario Institute for Cancer Research and its partners in the International Cancer Genome Consortium (ICGC) [79,80] sequence various tumor genomes, new cancer-related candidate pathways will be identified and curated into Reactome. Existing Reactome pathways are updated on a regular basis, and additional cancer variants and anti-cancer drugs implicated in EGFR, FGFR and PI3K/AKT pathways will be included as information on their function becomes available.

Reactome is not the only pathway database to curate pathway data relevant to cancer and disease. Cancer-perturbed signaling pathways can be found in KEGG, Panther, MetaCyc, and NCI-PID [81–84]. The Reactome data model, however, provides a more detailed framework for the curation of the

knowledge relevant to cancer-related pathways, a visualization environment to display pathway data, and a suite of analysis tools for the interpretation of experimental cancer data sets.

A number of other bioinformatics databases such as Mouse Genome Informatics (MGI) [85] and Comparative Toxicogenomics Database (CTD) [86] have established disease curation pipelines, employing OMIM. OMIM is a detail-orientated database of disease annotation, widely used by the clinical community, but it lacks the structure and features of an ontology that would otherwise make it a perfect data source to systematically reference disease. Curation of human disease requires an establishment of a widely accessible and structured vocabulary (or ontology) that consists of knowledge that is familiar to Reactome's end user, flexible to future Reactome annotation updates, and open to semantic reasoning. One such ontology is the Disease Ontology. Reactome will continue to work with the research community to support the development and continuous improvement of human disease ontologies and will link out to the relevant cancer and disease-related databases, to advance our own annotation consistency. In future versions of Reactome, we may also cross-reference NCIt [68] directly for cancer-related physical entities and events. The Disease Ontology does provide NCIt identifiers when possible, but disease terms captured by the Disease Ontology and NCIt do not completely overlap. Cross-referencing different ontologies will make our disease annotations more comprehensive and stable. Since some amount of overlap exists between disease terms in any disease ontology, the overlap is reflected in our current annotation of disease attributes. This is not ideal and we are developing guidelines to standardize the use of disease terms in Reactome. As far as anti-cancer therapeutics are concerned, we do not capture their approval for clinical use other than in text summations, as this is outside the scope of Reactome project. However, cross-referencing a drug database, such as PharmaGKB [87] would provide Reactome users with easy access to clinically relevant drug information, and is currently under our consideration.

We are working on further improvements to the Reactome pathway browser to produce more compact images and to be able to share one diagram between the wild-type pathway and several disease pathways with different etiologies. Furthermore, we are making additions to the Molecular Interaction Overlay to promote visual linkages between pathway entities and disease annotations, such as OMIM. Network-based methods have been used extensively in genomic and proteomic studies to analyze challenging and complex datasets. Reactome provides the Functional Interaction (FI) network plug-in for Cytoscape, which can identify network patterns related to diseases, including cancer [88]. Future expansion of the FI network with interactions based upon Reactome cancer-related pathways should significantly improve coverage, enhance the functionality of the analysis, and enrich the functional annotations supported by the FI network plug-in. Reactome will continue to develop novel and useful technologies for the querying, visualization and analysis of experimental datasets, in the context of not only normal but also disease pathways.

Acknowledgments

The data, data model, and data analysis tools described in this article are the product of the collaborative work of curators and software developers at the Ontario Institute for Cancer Research (Lincoln Stein, Michael Caudy, Marc Gillespie, Robin Haw, Marija Milacic, Bruce May, Karen Rothfels, Heeyeon Song, Joel Weiser, Guanming Wu), the European Bioinformatics Institute (Henning

Hermjakob, David Croft, Antonio Fabregat-Mundo, Phani Garapati, Bijay Jassal, Steven Jupe) and the NYU School of Medicine (Peter D'Eustachio, Lisa Matthews, Veronica Shamovsky). We are grateful to the many scientists who collaborated with us as authors and reviewers to build the content on the knowledgebase, and to our colleagues at GO, ChEBI, and UniProt. We are especially grateful to Lisa Matthews and Marc Gillespie for helpful discussions of disease annotation strategies and to two anonymous reviewers for their comments on an earlier version of this manuscript. This work was supported by grants from the National Human Genome Research Institute at the National Institutes of Health [grant number U41 HG003751], the European Union FP7 project “LipidomicNet” [grant agreement number 202272, and Ontario Research (GL2) Fund. Funding for this open access charge: National Human Genome Research Institute at the National Institutes of Health [grant number U41 HG003751].

References

1. Vogelstein, B.; Fearon, E.R.; Hamilton, S.R.; Kern, S.E.; Preisinger, A.C.; Leppert, M.; Nakamura, Y.; White, R.; Smits, A.M.; *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **1988**, *319*, 525–532.
2. Shah, S.P.; Roth, A.; Goya, R.; Oloumi, A.; Ha, G.; Zhao, Y.; Turashvili, G.; Ding, J.; Tse, K.; Haffari, G.; *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **2012**, *486*, 395–399.
3. Hahn, S.A.; Schutte, M.; Hoque, A.T.; Moskaluk, C.A.; da Costa, L.T.; Rozenblum, E.; Weinstein, C.L.; Fischer, A.; Yeo, C.J.; Hruban, R.H.; *et al.* DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **1996**, *271*, 350–353.
4. Hockenbery, D.; Nunez, G.; Millman, C.; Schreiber, R.D.; Korsmeyer, S.J. Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death. *Nature* **1990**, *348*, 334–336.
5. Lew, D.J.; Dulic, V.; Reed, S.I. Isolation of three novel human cyclins by rescue of G1 cyclin (Cln) function in yeast. *Cell* **1991**, *66*, 1197–1206.
6. Motokura, T.; Bloom, T.; Kim, H.G.; Juppner, H.; Ruderman, J.V.; Kronenberg, H.M.; Arnold, A. A novel cyclin encoded by a bcl1-linked candidate oncogene. *Nature* **1991**, *350*, 512–515.
7. Govindan, R.; Ding, L.; Griffith, M.; Subramanian, J.; Dees, N.D.; Kanchi, K.L.; Maher, C.A.; Fulton, R.; Fulton, L.; Wallis, J.; *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **2012**, *150*, 1121–1134.
8. Greenman, C.; Stephens, P.; Smith, R.; Dalgliesh, G.L.; Hunter, C.; Bignell, G.; Davies, H.; Teague, J.; Butler, A.; Stevens, C.; *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **2007**, *446*, 153–158.
9. Croft, D.; O’Kelly, G.; Wu, G.; Haw, R.; Gillespie, M.; Matthews, L.; Caudy, M.; Garapati, P.; Gopinath, G.; Jassal, B.; *et al.* Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res.* **2011**, *39*, D691–D697.
10. Matthews, L.; Gopinath, G.; Gillespie, M.; Caudy, M.; Croft, D.; de Bono, B.; Garapati, P.; Hemish, J.; Hermjakob, H.; Jassal, B.; *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **2009**, *37*, D619–D622.

11. Vastrik, I.; D'Eustachio, P.; Schmidt, E.; Gopinath, G.; Croft, D.; de Bono, B.; Gillespie, M.; Jassal, B.; Lewis, S.; Matthews, L.; *et al.* Reactome: A knowledge base of biologic pathways and processes. *Genome Biol.* **2007**, *8*, R39.
12. Joshi-Tope, G.; Gillespie, M.; Vastrik, I.; D'Eustachio, P.; Schmidt, E.; de Bono, B.; Jassal, B.; Gopinath, G.R.; Wu, G.R.; Matthews, L.; *et al.* Reactome: A knowledgebase of biological pathways. *Nucleic Acids Res.* **2005**, *33*, D428–D432.
13. Gillespie, M.; Shamovsky, V.; D'Eustachio, P. Human and chicken TLR pathways: Manual curation and computer-based orthology analysis. *Mamm. Genome* **2010**, *22*, 130–138.
14. Jassal, B.; Jupe, S.; Caudy, M.; Birney, E.; Stein, L.; Hermjakob, H.; D'Eustachio, P. The systematic annotation of the three main GPCR families in Reactome. *Database (Oxford)* **2010**, *2010*, baq018.
15. Maglott, D.; Ostell, J.; Pruitt, K.D.; Tatusova, T. Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Res.* **2011**, *39*, D52–D57.
16. Flicek, P.; Aken, B.L.; Ballester, B.; Beal, K.; Bragin, E.; Brent, S.; Chen, Y.; Clapham, P.; Coates, G.; Fairley, S.; *et al.* Ensembl's 10th year. *Nucleic Acids Res.* **2010**, *38*, D557–D562.
17. Consortium, T.U. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2012**, *40*, D71–D75.
18. Fujita, P.A.; Rhead, B.; Zweig, A.S.; Hinrichs, A.S.; Karolchik, D.; Cline, M.S.; Goldman, M.; Barber, G.P.; Clawson, H.; Coelho, A.; *et al.* The UCSC Genome Browser database: Update 2011. *Nucleic Acids Res.* **2011**, *39*, D876–D882.
19. De Matos, P.; Alcantara, R.; Dekker, A.; Ennis, M.; Hastings, J.; Haug, K.; Spiteri, I.; Turner, S.; Steinbeck, C. Chemical Entities of Biological Interest: An update. *Nucleic Acids Res.* **2010**, *38*, D249–D254.
20. Consortium, G.O. The Gene Ontology in 2010: Extensions and refinements. *Nucleic Acids Res.* **2010**, *38*, D331–D335.
21. Reactome-Signaling by EGFR. Available online: http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_9417/ (accessed on 20 September 2012).
22. Reactome-Signaling by FGFR. Available online: http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_9470/ (accessed on 20 September 2012).
23. Reactome-Signaling by NOTCH. Available online: http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_299/ (accessed on 20 September 2012).
24. Reactome-PIP3 Activates AKT Signaling. Available online: http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_75829/ (accessed on 20 September 2012).
25. Reactome-RAF/MAP Kinase Cascade. Available online: http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_634/ (accessed on 20 September 2012).
26. Reactome-Apoptosis. Available online: http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_578/ (accessed on 20 September 2012).
27. Reactome-Cell Cycle Checkpoints. Available online: http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_1538/ (accessed on 20 September 2012).
28. Reactome-Mitotic G1-G1/S phases. Available online: http://www.reactome.org/cgi-bin/eventbrowser_st_id?ST_ID=REACT_21267/ (accessed on 20 September 2012).

29. Sherrill, J.M.; Kyte, J. Activation of epidermal growth factor receptor by epidermal growth factor. *Biochemistry* **1996**, *35*, 5705–5718.
30. Eswarakumar, V.P.; Lax, I.; Schlessinger, J. Cellular signaling by fibroblast growth factor receptors. *Cytokine Growth Factor Rev.* **2005**, *16*, 139–149.
31. Ferguson, K.M. Structure-based view of epidermal growth factor receptor regulation. *Annu. Rev. Biophys.* **2008**, *37*, 353–373.
32. Furdui, C.M.; Lew, E.D.; Schlessinger, J.; Anderson, K.S. Autophosphorylation of FGFR1 kinase is mediated by a sequential and precisely ordered reaction. *Mol. Cell* **2006**, *21*, 711–717.
33. Hart, K.C.; Robertson, S.C.; Donoghue, D.J. Identification of tyrosine residues in constitutively activated fibroblast growth factor receptor 3 involved in mitogenesis, Stat activation, and phosphatidylinositol 3-kinase activation. *Mol. Biol. Cell* **2001**, *12*, 931–942.
34. Mohammadi, M.; Dikic, I.; Sorokin, A.; Burgess, W.H.; Jaye, M.; Schlessinger, J. Identification of six novel autophosphorylation sites on fibroblast growth factor receptor 1 and elucidation of their importance in receptor activation and signal transduction. *Mol. Cell. Biol.* **1996**, *16*, 977–989.
35. Avraham, R.; Yarden, Y. Feedback regulation of EGFR signalling: Decision making by early and delayed loops. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 104–117.
36. Schlessinger, J. Common and distinct elements in cellular signaling via EGF and FGF receptors. *Science* **2004**, *306*, 1506–1507.
37. Ong, S.H.; Hadari, Y.R.; Gotoh, N.; Guy, G.R.; Schlessinger, J.; Lax, I. Stimulation of phosphatidylinositol 3-kinase by fibroblast growth factor receptors is mediated by coordinated recruitment of multiple docking proteins. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 6074–6079.
38. Rodrigues, G.A.; Falasca, M.; Zhang, Z.; Ong, S.H.; Schlessinger, J. A novel positive feedback loop mediated by the docking protein Gab1 and phosphatidylinositol 3-kinase in epidermal growth factor receptor signaling. *Mol. Cell. Biol.* **2000**, *20*, 1448–1459.
39. Manning, B.D.; Cantley, L.C. AKT/PKB signaling: Navigating downstream. *Cell* **2007**, *129*, 1261–1274.
40. Burke, J.E.; Vadas, O.; Berndt, A.; Finegan, T.; Perisic, O.; Williams, R.L. Dynamics of the phosphoinositide 3-kinase p110delta interaction with p85alpha and membranes reveals aspects of regulation distinct from p110alpha. *Structure* **2011**, *19*, 1127–1137.
41. Mandelker, D.; Gabelli, S.B.; Schmidt-Kittler, O.; Zhu, J.; Cheong, I.; Huang, C.H.; Kinzler, K.W.; Vogelstein, B.; Amzel, L.M. A frequent kinase domain mutation that changes the interaction between PI3Kalpha and the membrane. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 16996–17001.
42. Maehama, T.; Dixon, J.E. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.* **1998**, *273*, 13375–13378.
43. Scheid, M.P.; Marignani, P.A.; Woodgett, J.R. Multiple phosphoinositide 3-kinase-dependent steps in activation of protein kinase B. *Mol. Cell. Biol.* **2002**, *22*, 6247–6260.
44. Hollander, M.C.; Blumenthal, G.M.; Dennis, P.A. PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nat. Rev. Cancer* **2011**, *11*, 289–301.
45. Greulich, H.; Chen, T.H.; Feng, W.; Janne, P.A.; Alvarez, J.V.; Zappaterra, M.; Bulmer, S.E.; Frank, D.A.; Hahn, W.C.; Sellers, W.R.; *et al.* Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS Med.* **2005**, *2*, e313.

46. Fernandes, H.; Cohen, S.; Bishayee, S. Glycosylation-induced conformational modification positively regulates receptor-receptor association: A study with an aberrant epidermal growth factor receptor (EGFRvIII/DeltaEGFR) expressed in cancer cells. *J. Biol. Chem.* **2001**, *276*, 5375–5383.
47. Wesche, J.; Haglund, K.; Haugsten, E.M. Fibroblast growth factors and their receptors in cancer. *Biochem. J.* **2011**, *437*, 199–213.
48. Weiss, J.; Sos, M.L.; Seidel, D.; Peifer, M.; Zander, T.; Heuckmann, J.M.; Ullrich, R.T.; Menon, R.; Maier, S.; Soltermann, A.; *et al.* Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci. Transl. Med.* **2010**, *2*, 62ra93.
49. Turner, N.; Grose, R. Fibroblast growth factor signalling: From development to cancer. *Nat. Rev. Cancer* **2010**, *10*, 116–129.
50. Cappellen, D.; de Oliveira, C.; Ricol, D.; de Medina, S.; Bourdin, J.; Sastre-Garau, X.; Chopin, D.; Thiery, J.P.; Radvanyi, F. Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. *Nat. Genet.* **1999**, *23*, 18–20.
51. Neilson, K.M.; Friesel, R. Ligand-independent activation of fibroblast growth factor receptors by point mutations in the extracellular, transmembrane, and kinase domains. *J. Biol. Chem.* **1996**, *271*, 25049–25057.
52. Sun, M.; Hillmann, P.; Hofmann, B.T.; Hart, J.R.; Vogt, P.K. Cancer-derived mutations in the regulatory subunit p85alpha of phosphoinositide 3-kinase function through the catalytic subunit p110alpha. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15547–15552.
53. Jaiswal, B.S.; Janakiraman, V.; Kljavin, N.M.; Chaudhuri, S.; Stern, H.M.; Wang, W.; Kan, Z.; Dbouk, H.A.; Peters, B.A.; Waring, P.; *et al.* Somatic mutations in p85alpha promote tumorigenesis through class IA PI3K activation. *Cancer Cell* **2009**, *16*, 463–474.
54. Huang, C.H.; Mandelker, D.; Schmidt-Kittler, O.; Samuels, Y.; Velculescu, V.E.; Kinzler, K.W.; Vogelstein, B.; Gabelli, S.B.; Amzel, L.M. The structure of a human p110alpha/p85alpha complex elucidates the effects of oncogenic PI3Kalpha mutations. *Science* **2007**, *318*, 1744–1748.
55. Miled, N.; Yan, Y.; Hon, W.C.; Perisic, O.; Zvelebil, M.; Inbar, Y.; Schneidman-Duhovny, D.; Wolfson, H.J.; Backer, J.M.; Williams, R.L. Mechanism of two classes of cancer mutations in the phosphoinositide 3-kinase catalytic subunit. *Science* **2007**, *317*, 239–242.
56. Zhao, J.J.; Liu, Z.; Wang, L.; Shin, E.; Loda, M.F.; Roberts, T.M. The oncogenic properties of mutant p110alpha and p110beta phosphatidylinositol 3-kinases in human mammary epithelial cells. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18443–18448.
57. Carpten, J.D.; Faber, A.L.; Horn, C.; Donoho, G.P.; Briggs, S.L.; Robbins, C.M.; Hostetter, G.; Boguslawski, S.; Moses, T.Y.; Savage, S.; *et al.* A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **2007**, *448*, 439–444.
58. Han, S.Y.; Kato, H.; Kato, S.; Suzuki, T.; Shibata, H.; Ishii, S.; Shiiba, K.; Matsuno, S.; Kanamaru, R.; Ishioka, C. Functional evaluation of PTEN missense mutations using *in vitro* phosphoinositide phosphatase assay. *Cancer Res.* **2000**, *60*, 3147–3151.
59. Pao, W.; Chmielecki, J. Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. *Nat. Rev. Cancer* **2010**, *10*, 760–774.

60. Greulich, H.; Pollock, P.M. Targeting mutant fibroblast growth factor receptors in cancer. *Trends Mol. Med.* **2011**, *17*, 283–292.
61. Liu, P.; Cheng, H.; Roberts, T.M.; Zhao, J.J. Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat. Rev. Drug Discov.* **2009**, *8*, 627–644.
62. Montecchi-Palazzi, L.; Beavis, R.; Binz, P.A.; Chalkley, R.J.; Cottrell, J.; Creasy, D.; Shofstahl, J.; Seymour, S.L.; Garavelli, J.S. The PSI-MOD community standard for representation of protein modification data. *Nat. Biotechnol.* **2008**, *26*, 864–866.
63. Yun, C.H.; Boggon, T.J.; Li, Y.; Woo, M.S.; Greulich, H.; Meyerson, M.; Eck, M.J. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **2007**, *11*, 217–227.
64. Urick, M.E.; Rudd, M.L.; Godwin, A.K.; Sgroi, D.; Merino, M.; Bell, D.W. PIK3R1 (p85alpha) is somatically mutated at high frequency in primary endometrial cancer. *Cancer Res.* **2011**, *71*, 4061–4067.
65. Demiroglu, A.; Steer, E.J.; Heath, C.; Taylor, K.; Bentley, M.; Allen, S.L.; Koduru, P.; Brody, J.P.; Hawson, G.; Rodwell, R.; *et al.* The t(8;22) in chronic myeloid leukemia fuses BCR to FGFR1: Transforming activity and specific inhibition of FGFR1 fusion proteins. *Blood* **2001**, *98*, 3778–3783.
66. Forbes, S.A.; Bindal, N.; Bamford, S.; Cole, C.; Kok, C.Y.; Beare, D.; Jia, M.; Shepherd, R.; Leung, K.; Menzies, A.; *et al.* COSMIC: Mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **2011**, *39*, D945–D950.
67. Schriml, L.M.; Arze, C.; Nadendla, S.; Chang, Y.W.; Mazaitis, M.; Felix, V.; Feng, G.; Kibbe, W.A. Disease Ontology: A backbone for disease semantic integration. *Nucleic Acids Res.* **2012**, *40*, D940–D946.
68. De Coronado, S.; Wright, L.W.; Fragoso, G.; Haber, M.W.; Hahn-Dantona, E.A.; Hartel, F.W.; Quan, S.L.; Safran, T.; Thomas, N.; Whiteman, L. The NCI Thesaurus quality assurance life cycle. *J. Biomed. Inform.* **2009**, *42*, 530–539.
69. Lee, J.C.; Vivanco, I.; Beroukhim, R.; Huang, J.H.; Feng, W.L.; DeBiasi, R.M.; Yoshimoto, K.; King, J.C.; Nghiemphu, P.; Yuza, Y.; *et al.* Epidermal growth factor receptor activation in glioblastoma through novel missense mutations in the extracellular domain. *PLoS Med.* **2006**, *3*, e485.
70. Li, S.; Schmitz, K.R.; Jeffrey, P.D.; Wiltzius, J.J.; Kussie, P.; Ferguson, K.M. Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* **2005**, *7*, 301–311.
71. Stebbins, C.E.; Russo, A.A.; Schneider, C.; Rosen, N.; Hartl, F.U.; Pavletich, N.P. Crystal structure of an Hsp90-geldanamycin complex: Targeting of a protein chaperone by an antitumor agent. *Cell* **1997**, *89*, 239–250.
72. Bai, A.; Meetze, K.; Vo, N.Y.; Kollipara, S.; Mazsa, E.K.; Winston, W.M.; Weiler, S.; Poling, L.L.; Chen, T.; Ismail, N.S.; *et al.* GP369, an FGFR2-IIIb-specific antibody, exhibits potent antitumor activity against human cancers driven by activated FGFR2 signaling. *Cancer Res.* **2010**, *70*, 7630–7639.
73. Le Novere, N.; Hucka, M.; Mi, H.; Moodie, S.; Schreiber, F.; Sorokin, A.; Demir, E.; Wegner, K.; Aladjem, M.I.; Wimalaratne, S.M.; *et al.* The Systems Biology Graphical Notation. *Nat. Biotechnol.* **2009**, *27*, 735–741.

74. Aranda, B.; Achuthan, P.; Alam-Faruque, Y.; Armean, I.; Bridge, A.; Derow, C.; Feuermann, M.; Ghanbarian, A.T.; Kerrien, S.; Khadake, J.; Kerssemakers, J.; *et al.* The IntAct molecular interaction database in 2010. *Nucleic Acids Res.* **2010**, *38*, D525–D531.
75. Overington, J. ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Warr. *J. Comput. Aided Mol. Des.* **2009**, *23*, 195–198.
76. Funahashi, A.; Tanimura, N.; Morohashi, M.; Kitano, H. CellDesigner: A process diagram editor for gene-regulatory and biochemical networks. *BioSilico* **2003**, *1*, 159–162.
77. Killcoyne, S.; Carter, G.W.; Smith, J.; Boyle, J. Cytoscape: A community-based framework for network modeling. *Methods Mol. Biol.* **2009**, *563*, 219–239.
78. Junker, B.H.; Klukas, C.; Schreiber, F. VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* **2006**, *7*, 109.
79. Hudson, T.J.; Anderson, W.; Artez, A.; Barker, A.D.; Bell, C.; Bernabe, R.R.; Bhan, M.K.; Calvo, F.; Eerola, I.; Gerhard, D.S.; *et al.* International network of cancer genome projects. *Nature* **2010**, *464*, 993–998.
80. International Cancer Genome Consortium. Available online: <http://www.icgc.org> (accessed on 20 September 2012).
81. Caspi, R.; Altman, T.; Dale, J.M.; Dreher, K.; Fulcher, C.A.; Gilham, F.; Kaipa, P.; Karthikeyan, A.S.; Kothari, A.; Krummenacker, M.; *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **2010**, *38*, D473–D479.
82. Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **2010**, *38*, D355–D360.
83. Mi, H.; Dong, Q.; Muruganujan, A.; Gaudet, P.; Lewis, S.; Thomas, P.D. PANTHER version 7: Improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* **2010**, *38*, D204–D210.
84. Schaefer, C.F.; Anthony, K.; Krupa, S.; Buchoff, J.; Day, M.; Hannay, T.; Buetow, K.H. PID: The Pathway Interaction Database. *Nucleic Acids Res.* **2009**, *37*, D674–D679.
85. Bello, S.M.; Richardson, J.E.; Davis, A.P.; Wieggers, T.C.; Mattingly, C.J.; Dolan, M.E.; Smith, C.L.; Blake, J.A.; Eppig, J.T. Disease model curation improvements at Mouse Genome Informatics. *Database (Oxford)* **2012**, *2012*, bar063.
86. Davis, A.P.; King, B.L.; Mockus, S.; Murphy, C.G.; Saraceni-Richards, C.; Rosenstein, M.; Wieggers, T.; Mattingly, C.J. The Comparative Toxicogenomics Database: Update 2011. *Nucleic Acids Res.* **2011**, *39*, D1067–D1072.
87. Whirl-Carrillo, M.; McDonagh, E.M.; Hebert, J.M.; Gong, L.; Sangkuhl, K.; Thorn, C.F.; Altman, R.B.; Klein, T.E. Pharmacogenomics knowledge for personalized medicine. *Clin. Pharmacol. Ther.* **2012**, *92*, 414–417.

88. Wu, G.; Feng, X.; Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* **2010**, *11*, R53.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).