DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Entourage: Visualizing Relationships between Biological Pathways using Contextual Subsets

Citation
Lex, Alexander Christian Partl, Denis Kalkofen, Marc Streit, Samuel Gratzl, Anne Mai Wassermann, Dieter Schmalstieg, and Hanspeter Pfister. 2013. Entourage: visualizing relationships between biological pathways using contextual subsets. IEEE Transactions on Visualization and Computer Graphics (InfoVis '13) 19(12): 2277–2286.

Published Version
doi:10.1109/TVCG.2013.154

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:11408595

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Entourage: Visualizing Relationships between Biological Pathways using Contextual Subsets

Alexander Lex, Christian Partl, Denis Kalkofen, Marc Streit, Samuel Gratzl,
Anne Mai Wassermann, Dieter Schmalstieg, and Hanspeter Pfister

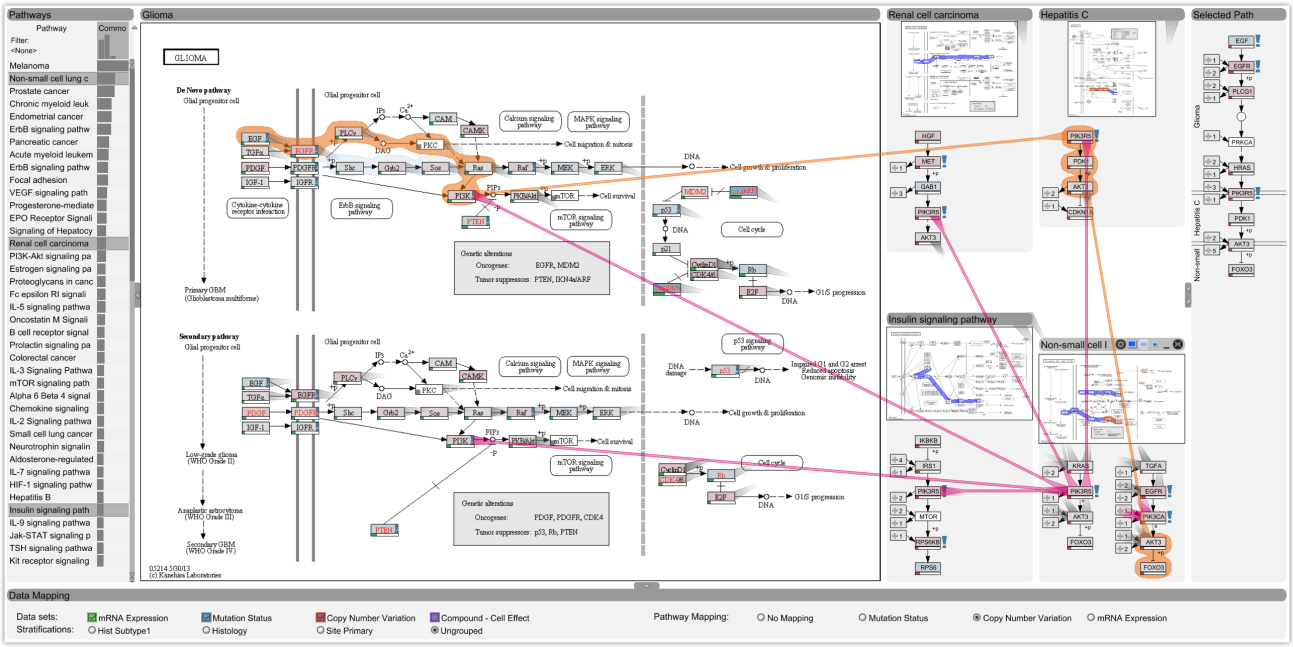


Fig. 1. Entourage showing the Glioma pathway in detail and contextual information of multiple related pathways.

Abstract—Biological pathway maps are highly relevant tools for many tasks in molecular biology. They reduce the complexity of the overall biological network by partitioning it into smaller manageable parts. While this reduction of complexity is their biggest strength, it is, at the same time, their biggest weakness. By removing what is deemed not important for the primary function of the pathway, biologists lose the ability to follow and understand cross-talks between pathways. Considering these cross-talks is, however, critical in many analysis scenarios, such as judging effects of drugs. In this paper we introduce *Entourage*, a novel visualization technique that provides contextual information lost due to the artificial partitioning of the biological network, but at the same time limits the presented information to what is relevant to the analyst's task. We use one pathway map as the focus of an analysis and allow a larger set of contextual pathways. For these context pathways we only show the *contextual subsets*, i.e., the parts of the graph that are relevant to a selection. Entourage suggests related pathways based on similarities and highlights parts of a pathway that are interesting in terms of mapped experimental data. We visualize interdependencies between pathways using stubs of visual links, which we found effective yet not obtrusive. By combining this approach with visualization of experimental data, we can provide domain experts with a highly valuable tool. We demonstrate the utility of Entourage with case studies conducted with a biochemist who researches the effects of drugs on pathways. We show that the technique is well suited to investigate interdependencies between pathways and to analyze, understand, and predict the effect that drugs have on different cell types.

Index Terms—Pathway visualization, biological networks, subsets, graphs, biomolecular data

- *Alexander Lex and Hanspeter Pfister are with Harvard University. E-mail: {alex, pfister}@seas.harvard.edu.*
- *Christian Partl, Denis Kalkofen and Dieter Schmalstieg are with Graz University of Technology. E-mail: {partl,kalkofen,schmalstieg}@tugraz.at.*
- *Marc Streit and Samuel Gratzl are with Johannes Kepler University Linz. E-mail: {marc.streit, samuel.gratzl}@jku.at.*
- *Anne Mai Wassermann is with Novartis Institutes for BioMedical Research. E-mail: anne_mai.wassermann@novartis.com.*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

1 Introduction

All living organisms can be considered as highly complex networks of biomolecules (genes, gene products, and metabolites) and biochemical reactions. It is the sum of tightly controlled and regulated interactions between these components that determines an organism's form and function. In the study of biological networks, the series of actions among biomolecules that lead to specific biological effects are commonly described as biological pathways. In other words, a pathway is a meaningful *subset* of biomolecules and reactions whose interplay fulfills a function in a cell or organism. Some pathways describe metabolic processes, e.g., the production of the amino acid *tyrosine* (*Tyrosine metabolism pathway*), whereas other pathways highlight the processes involved in a disease, such as the *Glioma* pathway, which describes the molecular mechanisms dysregulated in brain cancers.

1077-2626/13/$31.00 © 2013 IEEE Published by the IEEE Computer Society

Accepted for publication by IEEE. ©2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

This creation of subsets is, however, largely artificial, with the goal of reducing the complexity so that it can easily be comprehended by humans. While this approach in general is very successful it also has its drawbacks. By focusing only on those components that are immediately relevant to a biological process under study, other interaction partners that might only become relevant under specific circumstances, e.g., the treatment of a disease, are left out. For example, cancer is often caused by defects in multiple genes and pathways. In these cases, the identification of genes that are shared between dysregulated pathways is of high relevance because their products constitute prime targets for modulation by compounds, i.e., changing their activity using drugs. Furthermore, many drugs do not only bind to one target but exhibit activity against multiple gene products. For example, *Dovitinib (TKI258)* is a drug that targets, among others, the products of the genes *EGFR*, *FGFR1*, and *PDGFRbeta*, which are well-known proto-onco genes (normal genes that if mutated or highly expressed can potentially cause cancer). In these cases, the study of all affected genes and pathways in a common reference framework is highly desirable to better understand the drug's effect on the tumor cell. Moreover, some gene products are relevant in different cellular processes and drugs interfering with their function could potentially have multiple therapeutic indications. If a drug has already been approved because it has proven clinical safety, it is attractive to study the role of the drug's target in all possible pathways to find other disease implications and novel therapeutic uses of the drug. Lastly, it is important to consider the effects of a drug on all possible pathways to avoid undesirable side effects that result from unwanted modulations of the biological network.

In this paper we present *Entourage*, a visualization technique that allows analysts to conduct the kind of inter-pathway analysis required to answer such questions. By visualizing not only a single pathway, but including contextually relevant pathways, Entourage allows researchers to analyze the effects that modulations in one pathway might have on other interconnected processes. The primary challenge of doing so is dealing with scale. Individual pathway maps often contain several dozens, sometimes hundreds of nodes in addition to rich meta-information and are designed for full-screen, one-at-a-time viewing. While several approaches for integrating multiple of those pathway maps exists (e.g., [29, 19, 12]) we find that none are particularly successful in showing the relevant information at the right level of detail.

Our first contribution therefore is the *contextual subsets* concept that addresses this problem. By showing a small number of pathways (one or two) as *focus pathways* in detail, while showing contextual subsets of a set of related pathways, we enable analysts to see the details of the most important pathway, while keeping them informed about the interdependencies to other pathways.

Our second contribution is Entourage, which employs the contextual subsets concept, addresses practical issues in its realization, and introduces several domain specific visual encodings. In particular, we show how we determine related pathways and how we indicate potentially relevant content. Furthermore we describe our technique to visualize relationships between pathways and our approach to efficiently manage screen space. Moreover, we show how Entourage integrates advanced visualization techniques for analyzing large quantities of genomic and pharmacological data.

We demonstrate the utility of the contextual subset concept and the Entourage system in case studies on *KEGG* [13] pathways and the public *Cancer Cell Line Encyclopedia (CCLE)* [4] dataset. The CCLE dataset contains rich genetic profiling data for more than 500 cell lines (cultures of cells) in addition to pharmacological data that records responses of each of these cell lines to a set of 24 approved cancer drugs or drug candidates, which are referred to interchangeably in the following as drugs or compounds. We show that Entourage is indeed a highly valuable tool to (a) understand drug sensitivities of cell lines in light of their different genomic profiles and, consequently, distinct dysregulated pathways, and (b) explain different therapeutic indications for a single compound.

2 Domain Goals and Biological Background

While the analysis of relationships between multiple pathways is an important task in many application scenarios, our development of the Entourage technique was driven by three domain goals in drug discovery. We have been in close collaboration with an early stage drug discovery research group from a large pharmaceutical company over a period of several months, which included several meetings with larger groups of researchers and weekly meetings with one of their domain experts. In the following we describe their analysis goals and the datasets required to achieve these goals.

Understand a drug's mechanism of action and drug sensitivities of cell lines. The target of established drugs is typically known. In many cases, such drugs inhibit one or multiple gene products. However, there are fluctuations on how cell lines with distinct genomic profiles respond to the drugs [4]. Finding out why, for example, some cancer cells are killed by the drugs while others survive is one objective.

Judge side effects and safety of drugs. Although drugs are often designed to modulate only one particular biological pathway, their action on the cell and the organism as a whole must be considered in their development to better assess their safety. Being able to analyze cross-talk between pathways can help in judging the potential risks associated with a compound early on in the drug development process [32].

Identify potential for repositioning of drugs. Two alternative routes are usually taken in drug discovery: (1) developing new chemical entities and (2) finding new uses for already existing or previously failed drugs that have shown an adequate clinical safety profile [2]. The second route is usually more effective as such drugs can be approved quicker. As more and more knowledge about biological interactions and refined pathway maps become available, it is quite possible that existing drugs can be repurposed for a broader spectrum of therapeutic indications. Inter-pathway analysis can help to identify potential new therapeutic uses for approved drugs.

The data to be analyzed in these tasks can be classified into three categories: pathway data, and two forms of experimental data: genomic and pharmacologic profiling data. We have already introduced pathway data, and will now briefly explain the experimental data.

Genomic profiling data refers to datasets that measure the activity or structural variation of genes. An example of **genomic activity** is *gene expression* (or *mRNA expression*), which indicates how much of a functional gene product, such as a protein, is produced. Changes in gene activity can cause pathway dysregulation and a diseased state. One reason for a change in gene activity can be a **structural variation**. Structural variations occur on different scales. They can affect only a single base pair in the DNA or modify a whole chromosome. Two common forms of structural variation data are *copy-number variation* data, which records large scale duplications or deletions of genes, and *mutation* data, which captures smaller changes within an individual gene. Structural variation can result in changed activity or even loss of function. For example, a mutation in the gene *PTEN* results in uncontrolled signaling in a pathway promoting cell survival, which can lead to tumor growth [5]. The joint analysis of pathways and genomic experimental data makes it possible to identify such effects.

Pharmacologic profiling data essentially measures how cells react to compound exposure. A common measure is the *half maximal inhibitory concentration (IC_{50})*, which reports the concentration at which a drug achieves 50% inhibition, e.g., the drug concentration that is required to kill half of the treated tumor cells. The lower the concentration, the more effective is the drug and the more sensitive is the cell line under study to the drug treatment.

Pharmacologic and genomic profiling data are commonly jointly analyzed to, for example, identify reasons for the differential response that cell lines show to drug treatment. Integrating pathways into such an analysis can make it much more targeted, since it allows analysts to focus on the processes influenced by the compound and to explore related processes. Currently, several distinct tools are used for this analysis. Entourage is the first system to combine the different data types into one integrated interactive visualization.

3 Analysis Tasks

All of the domain goals described in the previous section have pathway analysis as pivotal parts of them. In this section we break down these domain goals into generalizable *analysis tasks*. We have elicited these tasks through interviews and feedback sessions with our collaborators. We classify the detailed tasks into two categories, the *Pathway Interconnectivity* tasks and the *Pathway-Experimental Data Linking* tasks. Note that both of those tasks are among the most critical requirements in pathway analysis [26] and are considered open problems in biological network analysis [1].

The *Pathway Interconnectivity* tasks deal with finding pathways related to each other and analyzing the relationships between pathways. The analysis tasks are:

Find related pathways. While an initial pathway is typically known for the stated goals, it is important to easily find related pathways, as cross-talk and other interdependencies are more likely between highly related pathways. We consider two pathways as related when they either share one or multiple genes (nodes), have an edge crossing from one pathway to the other, and/or if one is contained or referenced in the other.

Identify high-level relationships of pathways. When related pathways are found, it is also important to see how they are related. For example, it is interesting to see whether the same sub-process is contained in both pathways, or whether one pathway is contained or referenced within the other pathway.

Identify the role of a gene in multiple pathways. Identifying the role of a gene in other pathways is important to determine the different cellular processes that a gene is involved in, which, for example, is valuable knowledge when assessing the suitability of a gene as a drug target.

Find path intersections. As a change in gene activity, e.g., caused through a mutation or modulation by a drug, can influence the activity of subsequent genes in a path, it is important to not only look for the role of the originally altered gene but also to explore the role of genes that might be influenced by it. This can be done by exploring the relationships of the nodes downstream of the original gene, i.e., by finding pathways that intersect the path of a changed gene.

The *Pathway-Experimental Data Linking* tasks are equally important to achieve the goals stated above, as only experimental data can give insight into the effects a change, either naturally occurring or introduced by drug treatment, has on the whole cell or organism. We have in the past conducted an extensive task and requirement analysis for jointly analyzing pathways and experimental data [24]. The five elicited requirements (dealing with large scale data, dealing with heterogeneous data, resolving multi-mappings, following pathway layout constraints, and enabling topology-based and attribute-based tasks at the same time) equally hold for our domain goals. In addition, we identified the following, more specific, analysis tasks:

Identify subsets of pathways that warrant detailed investigation. As we have discussed in previous work, trying to show all experimental data on top of pathways is futile [24]. Consequently, analysts must be able to easily identify subsets of pathways (i.e., genes or paths) that warrant a closer look at the mapped experimental data.

Identify relationships between cell line responses to drug treatment and genomic data. Our collaborators would like to find out why certain cell lines react differently to the same compound treatment. These differences typically have genetic causes. Joint analysis of all the aforementioned data types can reveal the reasons for such differential behavior. This knowledge could in turn be used for targeted therapy, i.e., the identification of patients that are most likely to respond to a specific drug treatment.

4 Related Work

There are two major classes of visualizations that show interconnections between pathways. The first of these avoids sub-division by **showing the overall network as a whole**. The second **shows multiple pathways and visualizes relationships between them**.

One-Network Approaches.

As an example of the first class, *KEGG* [13] provides a high-level overview of the metabolic network (the *KEGG Atlas*) and lets analysts drill down into individual pathways. Other tools, like *iPath* [33] or *Pathway Projector* [16] use the same underlying data but improve the interaction with the atlas. While the original KEGG Atlas and iPath both use hyperlinks to replace the network overview with pathway maps on demand, Pathway Projector embeds node information directly on the all-encompassing map, thereby relying wholly on zooming and panning for navigation. Pathway Projector differs from the other two also in respect to how it represents gene nodes. The KEGG Atlas and iPath both represent genes and embedded pathways only as edges between the metabolites (the intermediate products of the metabolism), while the Pathway Projector actually shows nodes for genes, enzymes, and metabolites, which allows for direct mapping of experimental data.

All of these techniques show an incredibly large amount of data on a single screen. For example, the KEGG Atlas for *E. coli*, a comparatively simple organism, summarizes 1,365 genes, 1,813 enzymes, and 1,572 metabolites. In order to handle all this data, these techniques rely on selecting a focus, either by zooming and panning, or by changing into a different view altogether. This makes it very hard to identify interrelations to parts of the network outside of the currently visible area. All of these examples rely on a static layout, making features such as the *layout lens* [31] that pull connected nodes into focus, impossible.

Multiple Pathway Maps Approaches.

The approach of showing multiple pathways and visualizing relationships among them is taken in earlier versions of the *Caleydo* system, for example, in the *Jukebox* [29] and the *Bucket* [19] techniques. Both of them arrange multiple pathways in a 2.5D layout, while one of the pathways serves as focus. Visual links are used to connect related items on demand. As both techniques use thumbnails for context pathways, labels or even individual nodes are hard to see. Also, none of the techniques can show more than one relationship at a time.

Similar to the Caleydo techniques, Jusufi et al. extended the *Vanted* system to show multiple pathways as thumbnails [12]. They use navigation glyphs that show how individual nodes are connected to other pathways. A glyph has one petal for each possible link to other pathway groups, where the length of the petal encodes whether there is a link to a pathway group or not. These petals can be used to navigate to other pathways. The system, however, does not show any relationships on a node level.

Hybrid Approaches.

VisANT [11] takes a hybrid approach by showing a larger network as a set of meta-nodes which can be uncollapsed to reveal the underlying nodes and their cross-pathway connections. In the visualization community such meta-nodes are typically referred to as super nodes and are supported by a wide range of general purpose graph visualization frameworks such as *Tulip* [3] or *CGV* [31]. The concept of meta- or super nodes is significantly different from that of our contextual subsets as the nodes are not smart with respect to the context. A super node is either collapsed or not, whereas contextual subsets can show the elements that are contextually relevant while hiding the others.

The approach by Klukas and Schreiber [15] also employs super nodes. It uses a force-directed graph layout of abstracted pathway nodes and their relationships. Each abstract pathway node can interactively be expanded to show all the nodes of the corresponding pathway in detail. These nodes are arranged using the KEGG layout. They show all cross-pathway connections of individual nodes for multiple pathways at the same time. While this approach is reasonable for a limited set of pathways, the adding of more pathways continuously introduces clutter and reduces node size.

Rohrschneider et al. [25] use a similar approach in terms of showing multiple pathways at the same time, but use a grid-based automatic layout for the overall metabolic network. Their navigation approach is of particular interest, because they use the table-lens metaphor to switch between pathway super nodes and detailed renderings of the pathway. However, as with all super node approaches, the nodes are either expanded or collapsed but do not allow a context-based preview.

Consequently, the technique also can only provide detail and context for a very small set of pathways simultaneously.

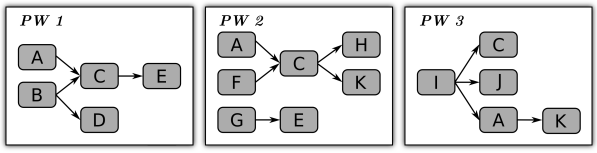
General Subset Techniques Our approach is also related to subset visualization techniques such as *VisBricks* [18] or *Portals* [23, 10]. VisBricks partition numerical datasets into subsets and show each subset with the visualization technique most suitable for the contained data and task. Portals are local regions within a visualization that show a different view on the area they cover. Olston and Woodruff employ portals to show data overlayed on maps [23], while the Hadlak et al. in-situ visualization uses portals (also) for graphs [10]. None of these techniques however, use any form of semantic context. Notice that we use the term *portal* to refer to shared nodes between pathways and that this usage is unrelated to Olston and Woodruff's term.

Visualizing Experimental Data in Pathways We use the *enRoute* technique [24] to visualize experimental data. enRoute uses path extraction and a separate linked view, which we integrated into Entourage. In contrast to Entourage, enRoute is strictly limited to a single pathway and has no notion of finding or presenting related or multiple pathways. Entourage and enRoute are complementary: the former addresses the problems of large and disjoint networks, while the latter makes it possible to visualize many node attributes.

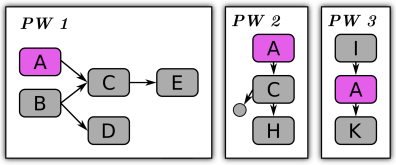
5 Visualizing Pathway Relationships

Enabling the *Pathway Interconnectivity* tasks requires a joint analysis of multiple pathways. Current techniques, however, lack the flexibility required for exploring interdependencies across pathway boundaries. The main problem one must address is scale. Current approaches either cannot show individual nodes sufficiently large or cannot show relationships between multiple pathways. We developed contextual subsets to remedy this issue.

Figure 2 illustrates the difference between a traditional multiple pathway analysis and the contextual subsets method. The traditional approach depicted in Figure 2(a) shows all nodes for all pathways. The pathways in this example share several nodes. We refer to such shared nodes as *portals* as they allow us to jump from one pathway to another. Figure 2(b) illustrates the same set of pathways using the contextual subsets technique. Instead of showing all pathways in detail, we distinguish between *focus pathways*, shown at full scale (PW 1 in Figure 2(b)), and *context pathways*, which are smaller and show only a contextually relevant subset of their graph (PW 2 and 3 in Figure 2(b)). What is contextually relevant is driven by a user-selected *focus node* (A, purple in Figure 2(b)). The context pathways only show limited subsets of their network that also contain node A. In the example shown in Figure 2(b), the most important path is shown for each occurrence of the focus node, while other branches are only indicated, as is evident in Pathway 2.



(a) Traditional Multi-Pathway Visualization



(b) Pathways as Contextual Subsets

Fig. 2. Comparison of a traditional multi-pathway approach and contextual subsets. (a) All pathways are shown at the same scale competing for display space. (b) The contextual subset technique showing one **focus pathway** (PW 1) and two **context pathways** (PW 2, PW 3). The context pathways only show paths that contain the **focus node** A.

This simple yet effective principle makes use of an observation: analysts do want to see all the details of one pathway map (their focus pathway), but do not need to see all the intricate details of other potentially involved processes (the context pathways) to judge interdependencies to their focus pathway. Entourage utilizes the observation that the focus of attention shifts serially to optimize the visible content to what is currently relevant to the analyst. The challenge we have to address is the continuous change of attention, the adaption of the analysis focus in the process of an exploration. Entourage employs a series of visual encodings and interaction techniques to make these changes as convenient and transparent as possible.

5.1 Overview

Figure 3 shows Entoruage's main components. The focus pathway takes up the majority of the space, while the context pathways are shown at the side. In this example *E2F* was selected as the focus node and the context pathways show their paths related to this node. Details on how context paths are selected are explained in Section 5.2, as are our methods to find relevant pathways.

Changes in focus are driven by user selections. However, choosing a meaningful focus is not always easy. Sometimes analysts will need to understand high-level relationships of pathways before they can set a sensible focus. Visualization is ideally suited to convey such high level relationships. Relationships between pathways are largely driven by portals as they connect two pathways. Showing portals and where they link to is therefore the most important aspect of showing high-level relationships between pathways. Figure 3 shows our approach for visualizing portals. We use a combination of stubs, which are shown for all portals at the same time, and visual links, i.e., visible edges, which are shown on request. These visual encodings efficiently convey high-level relationships between pathways and enable an analyst to set good focus points. Our visual encodings for showing relationships are explained in Section 5.3.

Finally, we need to address how to efficiently manage display space, as multiple focus and context elements compete for the limited screen real estate. We use an intelligent arrangement of pathways as well as multiple levels of detail for context pathways to optimize the display space, which are described in Section 5.4.

5.2 Determining Context Paths and Pathways

As discussed before, the contextual subsets concept is based on showing contextual information for a user-chosen focus, i.e., a focus node of a pathway. Which context information is eventually displayed depends on two factors: which paths in a pathway contain a focus node and which pathways are considered in the first place.

Determining Context Paths Context paths are selected by searching the graph for occurrences of the focus node or for immediately related nodes. Related nodes are, for example, nodes belonging to the same gene family. As it is common in nature that several distinct genes can fulfill the same role, albeit often with varying efficiency, pathway maps use both, a single label for the whole family or individual label for each of the family members. We consider these multi-mappings in our choices of relevant paths. This is the reason why occasionally differently labeled nodes are connected in Entourage.

Paths can either be unambiguous, as is the case in Pathway 3 of Figure 2(b), or contain branches, as in Pathway 2. If a path contains branches we automatically determine the branch that is likely to be most interesting by calculating the most variable branch in terms of the underlying experimental data. We do so by calculating the standard deviation across all experiments for each of the mapped datasets for every possible branch and choosing the branch that exhibits the highest deviation. As discussed in previous work [24], we preserve as much of the topology in the vicinity of paths as possible. Incoming and outgoing branches are collapsed into abstract nodes to save space, but can be extended to full-size nodes and switched-in to replace the main branch on demand. We decided against more complex attempts of linearizing larger portions of the network and including branches and cycles [22], to make the paths easy to understand for the analyst. Furthermore, we limit the length of automatically determined paths to what fits

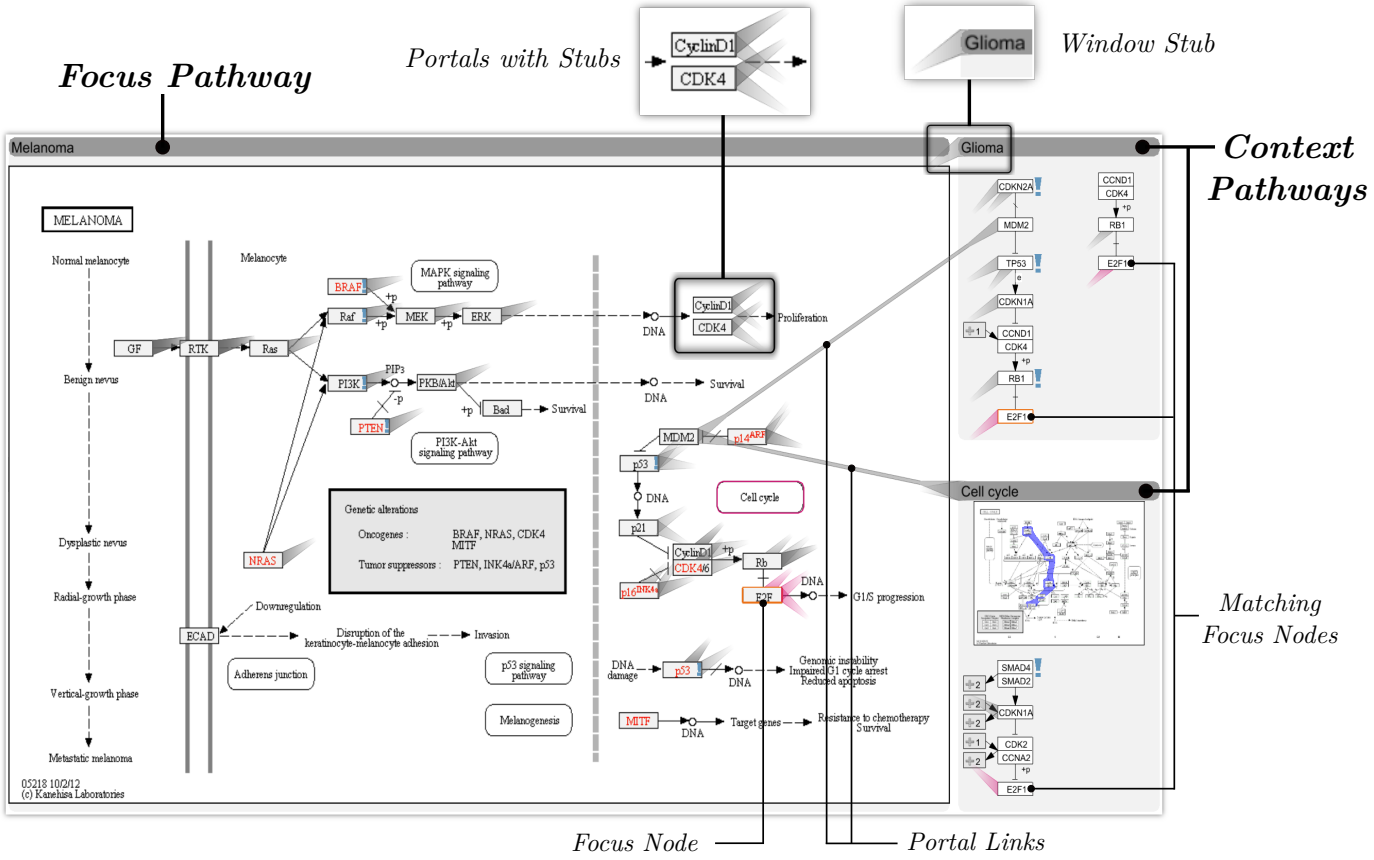


Fig. 3. The major components of Entourage. The **focus pathway** shows all details while the **context pathways** only show what is relevant in the context of the **focus node**. The insets at the top show how we indicate connections between pathways.

conveniently in the available space constraints, but give analysts the ability to extend the paths manually.

Determining Pathways Entourage shows paths only for manually selected pathways but suggests pathways that are relevant for a current focus node. Figure 1 shows a list of pathways on the left side. This list contains all pathways that contain the currently selected focus node, or a node of the same gene family. The pathways in the list are ranked by their similarity to the current focus pathway. We calculate a similarity score for each pathway by computing the number of nodes shared with the focus pathway and normalize it by its size. The score is shown as a bar next to the pathway name. To quickly determine which pathways have already been added to the workspace we mark loaded pathways using a dark gray background. In some situations analysts are interested in pathways that are generally similar to a selected pathway, without choosing a focus node. We use a similar algorithm to calculate scores of pathways in this scenario.

An alternative to the automatic, similarity based list is an alphabetic list which can be searched using keywords or regular expressions. This is especially helpful to find an entry point of an analysis. Finally, since pathway maps often embed related pathways, we enable the adding of such pathways to the workspace by clicking on embedded pathway nodes. The *Melanoma* focus pathway shown in Figure 3, for example, contains six embedded pathways indicating that these pathways play an important role in the context of the focus pathway. One of them (*Cell cycle*) is also a current context pathway, which is indicated by its purple border.

5.3 Visualizing Connections

To *find path intersections* and to enable the *identify high-level relationships* task we need to visually communicate which portal nodes connect two pathways. This requires visual encodings to (a) convey that a node (either in a focus or in a context pathway) is a portal and to (b) tell the analyst to which other portals it can be connected to.

Since related pathways often contain a substantial quantity of portal nodes, obvious approaches, such as color-coding or drawing visible edges, may easily fail. Even though objective (a) could be addressed by using a color-based highlighting of portal nodes, objective (b) would potentially require assigning many different colors to a single node. Visual links (i.e., visible edges), on the other hand, can connect a node to many others, but can result in significant clutter, given the many nodes and the dense layout of pathway maps, even if they were intelligently routed [28]. Therefore, we have chosen to primarily use stubs to encode relationships between nodes. Stubs were shown to be effective for indicating a connection without cluttering the display [7]. Figure 3 illustrates our stubs implementation. The two insets at the top show them in detail. For each pair of related portal nodes we render a pair of stubs pointing at each other. The direction of a stub thus indicates the location of its target. We attach the stubs to the side of the node closest to the target and we quickly let them fade while they are converging to a point. We also show portals only with respect to the "active" pathway, i.e., stubs only point to and from the pathway on which the mouse pointer rests. This reduces the set of portals, minimizing clutter and ambiguities while showing all relevant connections.

As context pathways only show a subset of nodes, potential portals might not be displayed. Nevertheless, we also want to communicate the presence of hidden portals. To achieve this, we show that a pathway has a relationship to the active pathway through one or multiple hidden portals by placing a stub on its window's title bar, as shown in Figure 3 (labeled *window stub*).

As is evident from the figures in this paper, stubs are excellent at indicating connections between many portal nodes without introducing a high amount of visual clutter. However, they can be ambiguous at times, especially when the angle between two stubs attached to the same node is small. To resolve potential ambiguities, we show the exact connections out of a portal node by using visual links when the mouse hovers over the portal (labeled *portal links* in Figure 3).

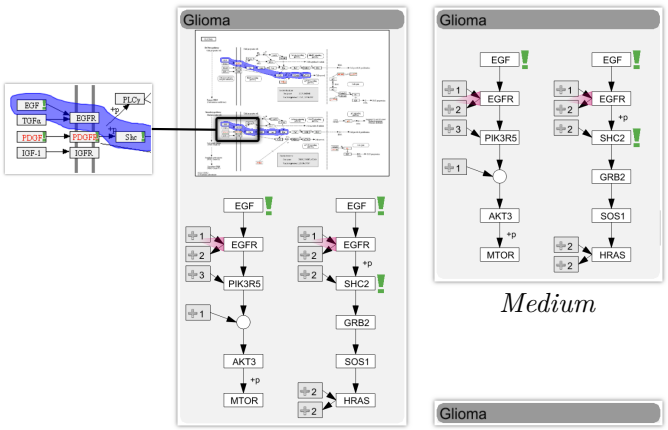


Fig. 4. The three different levels of detail of a context pathway. The highest level shows context paths plus a thumbnail of the overall pathway. Notice that the thumbnail also highlights the context paths. The medium level only shows the context paths, and the lowest level reduces the pathway to its title.

Notice that while relationships between portals are generally indicated by gray stubs, the recurring focus nodes in the different pathways are emphasized by using purple stubs or links. This combination of gray and purple stubs and on-demand visual links results in a clean visualization showing cross-connections between pathways in a minimally obtrusive way. These visual encodings also work well for comparing two focus pathways. What remains is to discuss how we can make good use of the limited screen space.

5.4 View Management

Using contextual subsets significantly reduces the number of elements that need to be displayed yet preserves the relevant context. Nevertheless it is prudent to make good use of the available screen space. Here we describe how we optimize the arrangement, size, and amount of data shown in the various pathways under analysis.

When optimizing a layout for pathway analysis, one is confronted with a range of partially conflicting goals. The first and most obvious goal is to *maximize the amount of relevant content shown*. This often conflicts with the goal to *ensure legibility of all elements*. Following the contextual subsets concept, we always use at least one focus pathway, for which we comply with the legibility goal, thus limiting the remaining space for contextual information. To deal with the varying amounts of space, we promote and demote pathways to various levels of detail and optimize the pathway layout.

Levels of Detail Our approach to efficiently layout pathways requires us to change their size. We achieve this by introducing three levels of detail for context pathways: *high*, *medium*, and *low*, which are illustrated in Figure 4. The thumbnail used in the highest level is typically large enough to convey a sense of the overall topology of the pathway. In order to aid orientation, we highlight the route of the context path(s) in the thumbnail, as shown in the inset of Figure 4. However, we consider this topological information less relevant than the actual context, which is why we omit the thumbnail if space is limited. In situations where there is not enough space to show any context paths, we resort to showing only the pathway titles. While this is not ideal, it is better than removing the pathway, since (a) it can be conveniently brought back into focus or any other level of detail and (b) it still indicates whether there is context information to be shown. Promotion or demotion of pathways between these levels of detail and the focus can be triggered manually but is also done automatically. Automatic actions can be disabled for individual pathways. This also makes a high-level comparison of two focus-pathways possible.

Layout Optimization We decided to use a rigid column-based layout to arrange pathways as opposed to a free layout, since matrix-like layouts are more space efficient when it comes to layouting rectangular shaped objects like pathway maps. Also, a column-based layout is well suited to reflect the history of the analysis process by sorting the pathways by age. Entourage can accommodate as many columns as are reasonable for a given screen resolution, but always enforces at least one context column as well as a minimum width for a column.

Our initial implementation followed the goals outlined above, always aiming to maximize the visible context information while ensuring legibility. Early feedback, however, triggered the realization that another factor is essential: *layout stability*. We observed that our collaborators were irritated by layout changes, even though they were animated. As a consequence we added the goal of *minimizing layout changes*. We also found that changing the size of a particular pathway is much less irritating than changing its position, either within or between columns. Consequently, our layout algorithm now prohibits position changes unless the focus pathway is exchanged, but permits resizing and switching between levels of detail of context pathways.

Within a context column, we maximize the vertical space between individual pathway windows. While this might not be as aesthetically pleasing as stacking them on top of each other, this strategy serves a purpose: It helps to avoid ambiguities of stubs pointing to the pathways by increasing the angle between stubs.

To fulfill our goal of maximizing the amount of relevant content displayed we promote and demote pathways intelligently. Automatic demotion of pathways is triggered when the horizontal or vertical display space is insufficient for displaying all elements at a reasonable size, while automatic promotion is triggered as space becomes available. An important decision in this regard is which pathways to demote or promote. This primarily depends on the causes of the space change. For example, if the vertical space is exceeded by the pathways in a context column, only pathways within that column have to be considered for demotion. In contrast, if there is too little horizontal space, the demotion of any pathway can potentially free up space.

To ultimately decide which of the pathways to demote or promote, we use three attributes of various priorities. The highest priority is given to pathways that contain a user-selected path (see Section 6). The second-highest priority is given to pathways that currently contain context paths. Finally, pathway "age" is considered as the lowest priority, where "young" pathways, i.e., those that were recently in focus, are given priority. We calculate a ranking of the candidate pathways based on these attributes and eventually demote the pathway with the lowest priority.

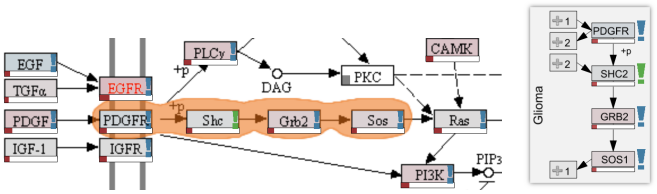
Taken together, our layout considerations guarantee a stable and predictable management of many pathways.

The techniques and encodings discussed in this section allow analysts to take a detailed look at one pathway while always keeping an eye open for cross-connections to other pathways. By showing only the information relevant to the current analysis, the important parts of the data can be shown at full scale. We thus provide an analyst with the necessary tools to address the *Pathway Interconnectivity* tasks.

6 Experimental Data Analysis

So far we have focused on how to visualize relationships between pathways considering only the pathways and the underlying network. In this section we will introduce (a) how experimental data can be leveraged to select interesting pathways and cross-connections in the first place and (b) how to visualize experimental data in the context of pathways and pharmacologic data. We will thereby address the two *Pathway-Experimental Data Linking* tasks.

The first of these tasks is to *identify subsets of pathways that warrant detailed investigation* based on experimental data. To accomplish this task, we need to provide information on which nodes are interesting in terms of the mapped experimental data. The most common approach to supplement pathways with experimental data is to color-code the nodes [20]. Other approaches include small bars, line plots, etc. For a comprehensive analysis refer to the review by Gehlenborg et al. [8]. All of these approaches aim at encoding experimental data on top of the pathways. However, such attempts are futile when dealing with large and heterogeneous datasets, but are helpful for single, homogeneous datasets [24]. Consequently, we make it possible to map individual



(a) (b)

Fig. 5. Path selection and experimental data mapping in pathways. The path highlighted in orange (a) is extracted and presented in a top-down layout (b). The node color in this example encodes the average copy number of mapped samples, while the red bars indicate the standard deviation. The exclamation marks indicate that the mapped experimental data varies considerably. The color of the exclamation marks and the standard deviation bars encodes the dataset in which the variation occurs. In this example blue marks signal variation in mutation and green marks show variation in mRNA expression data. Notice that we show cancer data in a cancer pathway, so it is not surprising that a high number of genes are mutated.

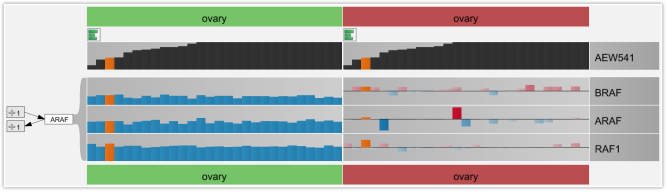


Fig. 6. Juxtaposition of pharmacological (on top) and genomic data. The pharamcological data captures the sensitivity of cell lines to drugs. The genomic data shown is mRNA expression (green, on the left) and copy number variation data (red, on the right). Orange bars are highlighted. Notice that the samples are sorted by the magnitude of their responses to the drug *AEW541*.

datasets by color coding the nodes, if desired (see Figure 5), but by default take a different approach: We point analysts to parts of a pathway that are either interesting for exploring the underlying experimental data or that are relevant to consider in other pathways. We do so by calculating the standard deviation of experimental data associated with each gene. If this deviation is higher than a threshold, we show an exclamation mark, as shown in Figure 5. The color of the exclamation mark encodes the dataset where the deviation was observed, which is also used in the data mapping view (see Figure 1). If multiple datasets show a large deviation, we encode only the largest. We chose a glyph since it nicely supplements the color-coding of nodes we use for showing average values of a single, selected dataset. This feature addresses the aforementioned task well, as typically variability in the data is of most interest.

For visualization of the actual underlying experimental data we employ the *enRoute* technique [24], which is part of the Caleydo framework. enRoute requires analysts to select a path in the network for which detailed experimental data is shown. Selected paths are highlighted using *Bubble Sets* [6]. Figure 5(a) shows a simple example of a selected path, the resulting extracted path is shown in Figure 5(b). Entourage always keeps track of the selected path in the selected path view, shown in Figure 1 on the right. By default, only the path is shown, but this view can be expanded to show enRoute, as demonstrated in Figure 7. Notice, that the enRoute view can also be shown full-screen, thereby occluding the pathways but giving more space to the experimental data analysis. Which data and which stratifications (groupings) of experiments are shown is driven by analyst choices made in the data mapping view shown in Figure 1 at the bottom.

While the original enRoute technique can only be used for paths in a single pathway, Entourage is ideally suited to select paths across pathways, as is shown in Figure 1. Notice that pathway boundaries are included in the path representation. We chose not to extend the bubble sets across pathways but instead use the visual links we also use for portals, as the connecting portal nodes are in fact the same node.

Finally, to address the *identify relationships between cell line responses to drug treatment and genomic data* task, we extended enRoute to show contextual data that is not associated with genes. Such data is shown above the gene-associated data and uses the same ordering of samples. Figure 6 shows the compound sensitivity of ovary CCLE cancer cell lines to the drug *AEW541* on top of the expression (on the left) and copy number values (on the right) associated with the *RAF* gene family (*BRAF*, *ARAF*, *RAF1*). Here, low bars indicate high sensitivity, i.e., low IC_{50} values. Notice that the samples are sorted with respect to their sensitivity to the compound, which is a simple yet effective way to search for relationships between genomic and pharmacologic data. This visual encoding can successfully address the task of associating cell line responses and genomic data. Moreover, since, for the small set of compounds the targeted processes and genes are known, it is easy to identify paths where interesting relationships between genomic and pharmacological data occur.

7 Implementation and Scalability

Entourage is part of the open source Caleydo Biomolecular Data Visualization Framework[1]. Caleydo is implemented in Java and uses JOGL [9] for rendering. Entourage will be a part of the next Caleydo release. We use an adapted version of the freely available implementation of Bubble Sets [17] for highlighting selected paths. Our Entourage implementation works with pathway maps from the KEGG [13] and WikiPathways [14] databases (see supplementary material for examples). Although we use the layouts provided by these databases in our current implementation, our technique is not limited to those and can equally be applied to automatically generated pathway layouts.

Depending on the size of the current focus pathway, Entourage can display up to ten pathways simultaneously, where one pathway is the focus pathway while the other pathways are at least in "medium" level of detail on a full HD notebook display (see supplementary material for examples). On larger, higher resolution screens, this number increases. The space for pathways can be increased by hiding currently unused support views (the data mapping, pathway list, and path views shown in Figure 1). Feedback from our collaborators indicates that this number of simultaneously explorable pathways is sufficient in all but the rarest cases and superior compared to other systems. We believe that our visual encodings are also suitable to point at interesting relationships outside an analyst's primary field of view, making it suitable for the increasingly large displays that are becoming commonplace. On conventional displays we typically limit the number of focus pathways to one in order to guarantee readability. This number, however, can be temporarily increased if detail about the structural relationship of pathways should be shown.

8 Case Studies

Entourage was developed in a user-centered design process including weekly meetings between the visualization developers and multiple domain experts. As a result of these meetings we have established the 3 domain goals and the 6 analysis tasks. We deployed various iterations of Entourage and our primary contact, a biochemist, who is also an author of this paper, used Entourage over a period of four weeks. During this period we were in constant contact with her and refined various aspects of the system. The case studies presented here report on her observations. Prior to the deployment of Entourage the team was using conventional pathway tools and had to resolve any questions concerning pathway relationships manually. For visualization of experimental data they mainly relied on tools like *TIBCO Spotfire* [30], whereas interactions between biomolecules were analyzed with network visualization tools like *Cytoscape* [27] or the KEGG web interface. Hence, Entourage allowed them to integrate two analysis steps that were previously carried out separately into one single task.

In the following we describe case studies for the domain goals concerning drugs' mechanisms of action and drug repositioning, which we found representative for demonstrating Entourage's functionality. The case studies either describe a novel observation or clearly demonstrate how a known effect can be rationalized with Entourage.

[1] http://caleydo.org

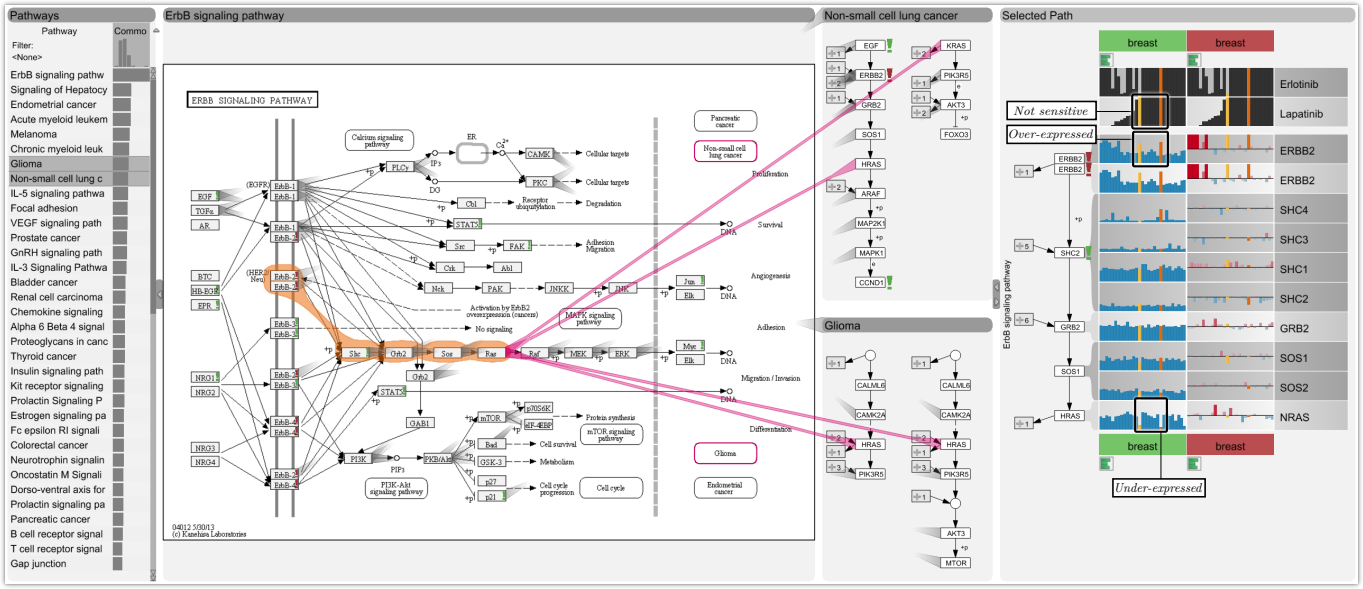


Fig. 7. The *ErbB signaling pathway* (the focus pathway) is a target of the drugs *Lapatinib* and *Erlotinib* that are used for cancer treatment. As shown in the pathway list on the left that results from a query for similar pathways, the *ErbB signaling pathway* is related to many cancer pathways. A signaling cascade from *ErbB2* to *Ras* is selected. The integrated enRoute view shows copy number and mRNA expression data for breast cancer cell lines. The sensitivity of the different cell lines to *Lapatinib* and *Erlotinib* is reported at the top. For the shown cell lines, increased copy numbers of *ErB2* (high red bars in the *ErbB2* row) result in over-expression of this gene (high blue bars). Furthermore, there is a strong relation between *ErbB2* over-expression and sensitivity to *Lapatinib* (high blue bars for gene over-expression in the *ErbB2* row coincide with low bars in the *Lapatinib* row). This means that *Lapatinib* is effective if *ErbB2* is highly expressed. There are, however, two exceptions - the highlighted cell lines (gold and orange), for which an under-expression in *Ras* downstream in the pathway is observed, likely causing *Lapatinib* to be ineffective in these cases. While this observation was made for breast cancer tissue, exploring the related context pathways by setting the focus node to *Ras* reveals that the same signaling cascade (i.e., path) is also contained in the non-small cell lung cancer pathway. Thus, it would be interesting to explore the transferability of the observed resistance pattern to this tissue type.

8.1 Relating Genomic Features to Compound Sensitivity

To explain different compound sensitivities of cell lines, our collaborator used the previously introduced *CCLE* dataset. This dataset contains data on the inhibitory effects of 24 drugs against roughly 500 cell lines from different cancer tissues and genomic data. Ideally, a drug completely inhibits the growth of these cell lines at minimal concentrations. First, she wanted to investigate factors that sensitize cell lines to the drugs *Lapatinib* and *Erlotinib* that inhibit members of the *ErbB* gene family and are used in cancer treatment. The *ErbB* family is a family of epidermal growth factor receptors that are known to play an important role in tumor growth. The drug *Lapatinib* is a dual inhibitor of *EGFR* and *ErbB2*, while *Erlotinib* is a known inhibitor of only *EGFR*, all of which belong to the aforementioned family. Due to its immediate relevance, the expert started by loading the *ErbB signaling pathway* into Entourage. By searching for related pathways she found several cancer-specific pathway maps. The pathways *Glioma* and *Non-small cell lung cancer* ranked among the top on the list (see Figure 7). She commented that this indicates that the *ErbB signaling pathway* is a key player in these diseases. For the *ErbB pathway map*, our collaboration partner was interested in the experimental data for the genes in the path that leads from *ErbB* receptors to *Myc*, a gene known to regulate cell growth. She also noticed that *ErbB2* was highlighted with a red exclamation mark indicating high variance in the copy number data. She thus selected the genes of this path for an in-depth analysis. She then looked at this path's gene expression data in the embedded enRoute view and combined it with sensitivities to *Erlotinib* and *Lapatinib*. For the analysis, cell lines were grouped by their tissue of origin (e.g., breast, ovary, liver, etc.) and sorted by sensitivity to *Lapatinib*. Her first observation, when looking at the experimental data, was that the two drugs displayed inhibitory activities across cell lines from many different tissues. The cell lines from lung, breast and three other tissues were in general most responsive. The set of cell lines that were responsive to *Erlotinib* and *Lapatinib* largely overlapped, although *Lapatinib* showed a broader spectrum of activity than *Erlotinib*. She found a strong co-occurrence between *ErbB2* mRNA over-expression and sensitivity to *Lapatinib* in lung and breast cancer cell lines, a trend that was less apparent or not observed at all for other responsive cell lines.

She then chose to focus on cell lines from breast and also investigated copy number variation for these cell lines. For most breast cancer cell lines that over-expressed *ErbB2*, high copy numbers of this gene were found, i.e., the increased expression could generally be traced back to an increased copy number. Interestingly, only two breast cancer cell lines that showed strong over-expression of *ErbB2* did not respond to *Lapatinib* treatment. The columns of these two cell lines are highlighted gold and orange in Figure 7. She then tried to find the cause for this effect and examined the expression of downstream genes in the pathway. She found that for these two cell lines, the gene *Ras* was strongly under-expressed (also shown in Figure 7).

It is straightforward to assume that this under-expression further down the path counteracts the over-activation of the pathway by increased *ErbB2* expression, explaining the resistance to *Lapatinib* treatment that reduces the effects of ErbB2 expression. Our collaborator stated that this highlights the importance of being able to analyze genomic data in a pathway context because compound sensitivities can often only be explained by the interplay of multiple genomic features.

Based on this observation she started to investigate whether other cancer-related pathways contain the same signaling cascade, i.e., path. She selected *Ras* as the focus node of her analysis, which revealed several other pathways that contain the same *ErbB* signaling cascade. Figure 7 shows an example where it is obvious, due to the many parallel stubs that the cascade is indeed identical. Our collaborator commented that it would interesting to investigate in the future whether a similar gene expression pattern in these cancer types would also entail resistance to *Lapatinib*.

8.2 Rationalizing Successful Drug Repositioning

Graft-versus-host disease (GVHD) is frequently observed after tissue or organ transplantation and is caused by immune cells that originate from the donor and were transplanted with the tissue. These immune cells perceive tissue of the recipient as foreign and attack it, thereby causing damage. The *Graft-versus-host disease* pathway identifies *TNF-alpha*, a gene involved in inflammation, as an important player

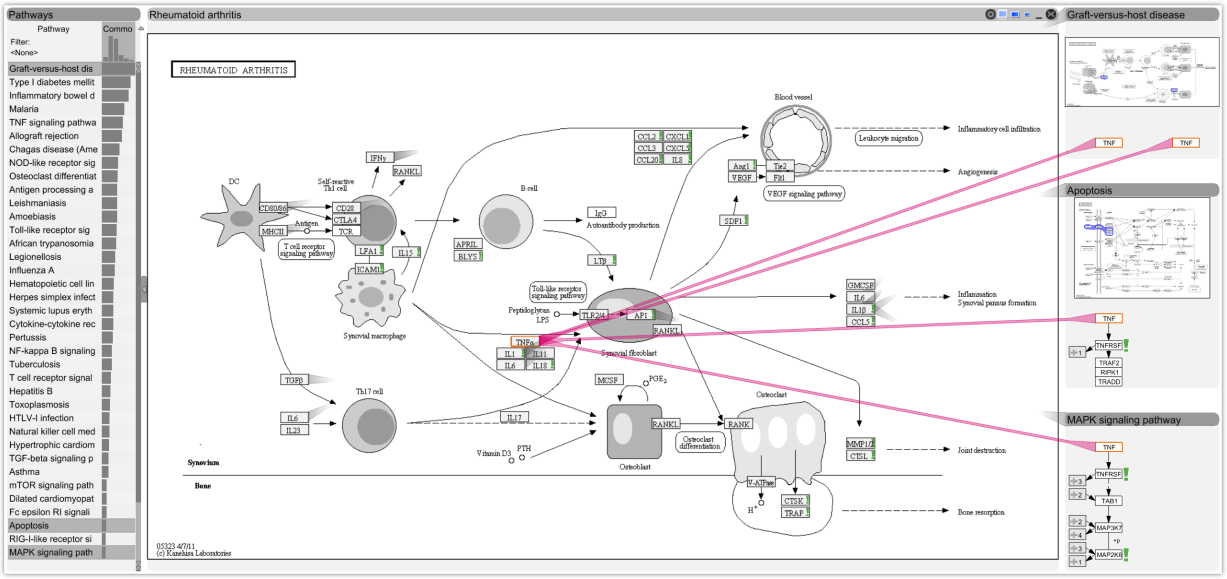


Fig. 8. *TNF-alpha* (the focus node) was originally explored as a target for the *Graft-Versus-Host Disease* (GVHD, top-right pathway). However, when tested in clinical trials, *TNF-alpha* inhibiting compounds were not effective against GVHD but could later be repositioned for the treatment of *Rheumatoid Arthritis* (focus pathway). Entourage shows Rheumatoid Athritis as closely related to the GVHD pathway (see pathway list on the left). Entourage also reveals seemingly contradictory roles of *TNF-alpha*. It is involved in cell death (Apoptosis) and also in cancer (i.e., uncontrolled cell growth) through the *MAPK signaling pathway*.

in the disease. Accordingly, molecules counter-acting (inhibiting) the effect of *TNF-alpha* have been evaluated for preventing GVHD in transplantation patients, with no success. However, the roles of *TNF-alpha* in the organism are manifold, as our collaborator was able to demonstrate using Entourage, when she chose *TNF-alpha* as her focus node, revealing all associated pathways. She found that one of the highest scoring and therefore most similar pathway to the *Graft-versus-host disease* map is the *Rheumatoid arthritis* pathway, shown as the focus pathway in Figure 8. Indeed, as clinical safety for *TNF-alpha* inhibitors had been proven in the initial trials for GVHD patients, the molecules were revisited and tested for their efficacy in patients suffering from rheumatoid arthritis. In this case, anti-*TNF alpha* therapy showed the desired clinical effect and today *TNF* inhibitors are part of the standard treatment of rheumatoid arthritis. The domain expert pointed out that Entourage also ranks the *Apoptosis*, *MAPK signaling*, and *NF-kappa B signaling* pathways, which explain the controversial role of *TNF-alpha* in cancer. While the *Apoptosis* pathway shows the process by which *TNF-alpha* leads to cell death, the other two pathways point out how *TNF-alpha* contributes to cell survival. Accordingly, the benefit of *TNF-alpha* inhibitors in anti-cancer treatment remains an open question and clinical trials are awaited to further explore the potential use of these molecules in malignancies.

9 Conclusions and Future Work

Analyzing relationships between pathways to accommodate analysis scenarios such as accounting for pathway cross-talk, repurposing drugs, and relating genomic features to drug sensitivity are challenging and to date unsolved problems. Previous approaches have aimed to show all considered pathways as a whole or even tried to represent the whole network at the same time. We have argued that doing either is not scalable and at the same time unnecessary. Our approach uses a strict focus and context approach, where the focus as well as the *relevant* context is presented at any given time. We use a combination of carefully chosen visual encodings and analytical support to help experts find the important parts of their data.

Overall, our collaboration partners were excited about the analytical capabilities of Entourage and mentioned that they perceived a significant improvement over their previous tool-chain. While, for example, the KEGG interface could be used to conduct an analysis similar to the one described in the second case study, doing so would be very tedious, as KEGG provides no support for analyzing relationships of pathways. They highly valued the ability to immediately see all relevant related processes for a pathway and being able to compare them easily and to see experimental data in the context of pathways.

We have demonstrated the utility of Entourage in two case studies highly relevant for pharmacological research. These case studies reflect current needs of pharmaceutical research, but we believe that our technique is equally applicable in domains such as systems biology or general molecular biology, as interconnections between pathways influence virtually all domains involved with biomolecular data.

Moreover, we argue that the contextual subsets approach can be applied to general graph analysis. It is conceivable that automatically created clusters of a graph can be used instead of manually partitioned pathways. We also believe that several other aspects presented in this paper can be generalized to other visualization applications. In particular our methods for visualizing relationships could be used for supplementary relationships in graphs, while our view management approach is applicable to all techniques using flexible multi-view setups.

In the future we aim to investigate how to represent compound data and its influences on the biological network. For the CCLE data, visualization of protein-compound interaction, for example, is irrelevant, since the compounds covered by this dataset are few and well understood. There are, however, similar datasets being created that contain data for hundreds or even thousands of compounds, about which there is only limited knowledge available. Early research in this area [21] highlights the important role visualization can play in this domain.

Another potential future line of inquiry is that of comparative analysis of multiple paths. Consider an example where two branches converge into a single node. Current visualization techniques are either not able to deal with the quantity of node attributes necessary to conduct a sensible analysis or fail to represent the topology efficiently, opening opportunities for interesting visualization research.

Acknowledgments

The authors thank Mark Borowsky and the anonymous reviewers. This work was supported in part by the Austrian Science Fund (J 3437-N15, P22902, P24021), the Province of Styria HTI (A3-22.M-5/2012-21 “Tumor Heterogeneity”), and the Air Force Research Laboratory and DARPA grant FA8750-12-C-0300. Anne Mai Wasserman is the recipient of a NIBR Presidential Postdoctoral Fellowship.

REFERENCES

[1] M. Albrecht, A. Kerren, K. Klein, O. Kohlbacher, P. Mutzel, W. Paul, F. Schreiber, and M. Wybrow. On open problems in biological network visualization. In D. Eppstein and E. Gansner, editors, *Graph Drawing*, volume 5849 of *Lecture Notes in Computer Science*, pages 256–267. Springer Berlin Heidelberg, 2010.

[2] T. T. Ashburn and K. B. Thor. Drug repositioning: identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, 3(8):673–683, Aug. 2004.

[3] D. Auber. Tulip: A huge graph visualization framework. In M. Jünger and P. Mutzel, editors, *Graph Drawing Software*, Mathematics and Visualization, pages 105–126. Springer Berlin Heidelberg, 2004.

[4] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, and L. A. Garraway. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, Mar. 2012.

[5] L. C. Cantley and B. G. Neel. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-Kinase/AKT pathway. *Proceedings of the National Academy of Sciences*, 96(8):4240–4245, 1999.

[6] C. Collins, G. Penn, and S. Carpendale. Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '09)*, 15(6):1009–1016, 2009.

[7] C. Collins, F. Viegas, and M. Wattenberg. Parallel tag clouds to explore and analyze faceted text corpora. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09)*, pages 91–98. IEEE Computer Society Press, 2009.

[8] N. Gehlenborg, S. I. O'Donoghue, N. S. Baliga, A. Goesmann, M. A. Hibbs, H. Kitano, O. Kohlbacher, H. Neuweger, R. Schneider, D. Tenenbaum, and A.-C. Gavin. Visualization of omics data for systems biology. *Nature Methods*, 7(3):56–68, 2010.

[9] S. Gothel, J. Gouesse, X. Rånby, R. Santina, M. Raynsford, H. Harrison, D. Ströhlein, and W. Walker. JOGL, 2013.

[10] S. Hadlak, H. J. Schulz, and H. Schumann. In situ exploration of large dynamic networks. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '11)*, 17(12):2334–2343, Dec. 2011.

[11] Z. Hu, J.-H. Hung, Y. Wang, Y.-C. Chang, C.-L. Huang, M. Huyck, and C. DeLisi. VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. *Nucleic Acids Research*, 37(Web Server issue):W115–W121, July 2009.

[12] I. Jusufi, C. Klukas, A. Kerren, and F. Schreiber. Guiding the interactive exploration of metabolic pathway interconnections. *Information Visualization*, 11(2):136–150, Apr. 2012.

[13] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40(1):D109–114, 2012.

[14] T. Kelder, M. P. v. Iersel, K. Hanspers, M. Kutmon, B. R. Conklin, C. T. Evelo, and A. R. Pico. WikiPathways: building research communities on biological pathways. *Nucleic Acids Research*, 40(D1):D1301–D1307, Nov. 2011.

[15] C. Klukas and F. Schreiber. Dynamic exploration and editing of KEGG pathway diagrams. *Bioinformatics*, 23(3):344–350, 2006.

[16] N. Kono, K. Arakawa, R. Ogawa, N. Kido, K. Oshita, K. Ikegami, S. Tamaki, and M. Tomita. Pathway projector: Web-based zoomable pathway browser using KEGG atlas and google maps API. *PLoS ONE*, 4(11):e7710, Nov. 2009.

[17] J. Krause. Bubble sets implementation, 2012.

[18] A. Lex, H.-J. Schulz, M. Streit, C. Partl, and D. Schmalstieg. VisBricks: multiform visualization of large, inhomogeneous data. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '11)*, 17(12):2291–2300, 2011.

[19] A. Lex, M. Streit, E. Kruijff, and D. Schmalstieg. Caleydo: Design and evaluation of a visual analysis framework for gene expression data in its biological context. In *Proceeding of the IEEE Symposium on Pacific Visualization (PacificVis '10)*, pages 57–64, 2010.

[20] H. Lindroos and S. G. E. Andersson. Visualizing metabolic pathways: comparative genomics and expression analysis. *Proceedings of the IEEE*, 90(11):1793–1802, 2002.

[21] E. Lounkine, P. Kutchukian, P. Petrone, J. W. Davies, and M. Glick. Chemotography for multi-target SAR analysis in the context of biological pathways. *Bioorganic & Medicinal Chemistry*, 20(18):5416–5427, Sept. 2012.

[22] M. Meyer, B. Wong, M. Styczynski, T. Munzner, and H. Pfister. Pathline: A tool for comparative functional genomics. *Computer Graphics Forum (EuroVis '10)*, 29(3):1043–1052, 2010.

[23] C. Olston and A. Woodruff. Getting portals to behave. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '00)*, pages 15–25. IEEE Computer Society Press, 2000.

[24] C. Partl, A. Lex, M. Streit, D. Kalkofen, K. Kashofer, and D. Schmalstieg. enRoute: dynamic path extraction from biological pathway maps for in-depth experimental data analysis. In *Proceedings of the IEEE Symposium on Biological Data Visualization (BioVis '12)*, pages 107–114, 2012.

[25] M. Rohrschneider, C. Heine, A. Reichenbach, A. Kerren, and G. Scheuermann. A novel grid-based visualization approach for metabolic networks with advanced Focus&Context view. In D. Eppstein and E. Gansner, editors, *Graph Drawing*, volume 5849 of *Lecture Notes in Computer Science*, pages 268–279. Springer Berlin Heidelberg, 2010.

[26] P. Saraiya, C. North, and K. Duca. Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. *Information Visualization*, 4(3):191–205, 2005.

[27] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.

[28] M. Steinberger, M. Waldner, M. Streit, A. Lex, and D. Schmalstieg. Context-preserving visual links. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '11)*, 17(12):2249–2258, Dec. 2011.

[29] M. Streit, M. Kalkusch, K. Kashofer, and D. Schmalstieg. Navigation and exploration of interconnected pathways. *Computer Graphics Forum (EuroVis '08)*, 27(3):951–958, 2008.

[30] TIBCO Software. Spotfire, 2013.

[31] C. Tominski, J. Abello, and H. Schumann. CGV—An interactive graph visualization system. *Computers & Graphics*, 33(6):660–678, Dec. 2009.

[32] J. Wang, Z.-x. Li, C.-x. Qiu, D. Wang, and Q.-h. Cui. The relationship between rational drug design and drug side effects. *Briefings in Bioinformatics*, 13(3):377–382, May 2012.

[33] T. Yamada, I. Letunic, S. Okuda, M. Kanehisa, and P. Bork. iPath2.0: interactive pathway explorer. *Nucleic Acids Research*, 39(suppl 2):W412–W415, July 2011.