Integrated analysis and visualization of heterogeneous cross-platform microarray datasets

Clemens Wrzodek[1,*], Elif Unterberger[2], Johannes Eichner[1] Michael Schwarz[2], Andreas Zell[1]

1 Center for Bioinformatics Tuebingen (ZBIT), University of Tuebingen, Sand 1, 72076 Tübingen, Germany

2 Department of Toxicology, Institute of Experimental and Clinical Pharmacology and Toxicology, University of Tuebingen, Wilhelmstrasse 56, 72074 Tübingen, Germany

∗ **E-mail: clemens.wrzodek@uni-tuebingen.de**

Abstract

In the last decade, measuring transcriptome data has become the most popular method for genome-wide high-throughput analysis of diverse biological samples. But nowadays, genome-wide evaluation of regulatory effects, e.g., from DNA methylations or microRNAs has become more and more important to explain the observed transcriptomic changes. Furthermore, high-throughput technologies are available that facilitate the generation of omics data from multiple different layers of gene regulation. These datasets are usually analyzed with respect to the underlying platform. Cross-platform effects, such as a hypermethylated DNA region that leads to suppression of a microRNA, which in turn increases the amount of a targeted mRNA are hard to discover. Applications and methods that allow for an integrated analysis of data from multiple heterogeneous microarray platforms are very rare.

We here propose multiple methods and visualizations for a joint analysis of biological samples, whose expression has been measured with multiple different platforms. All described analysis techniques are implemented in the InCroMAP application, available from `http://www.cogsys.cs.uni-tuebingen.de/software/InCroMAP/`. The usefulness of all methods is demonstrated on a cross-platform dataset that comprises mRNA, microRNA, DNA methylation and protein modification data from *Ctnnb1*- and *Ha-ras*-mutated tumors. **TODO: Wäre super wenn wir hier noch eine tolle (abstract-style) Erklärung hätten, warum grade die Analyse der Daten von diesen Tumoren so interessant ist. Dafür habe ich noch 60 bis 100 Wörter platz gelassen.** We show that the integrated analysis of all four data types provides much deeper biological insights into the mechanisms of those tumor samples then independent analyses of all platforms.

Author Summary

Note: Dies ist eine PLoS comp. biol. spezifische, subjektive Kurzfassung die für jeden Wissenschaftler verständlich sein sollte. PLoS sagt: "Please keep the Author Summary between 150 and 200 words. Use first person." Derzeit sind es ca. 220 Wörter...

A frequent method to discover the impact of a certain stimulus on a cell or organism is comparing treated samples to untreated control samples. We employed this technique to discover the differences between healthy and tumor samples from male mice with mutations in the *Ctnnb1* or *Ha-ras* gene. Multiple heterogenous platforms have been employed to gain a broad insight into the biochemical changes in the tumor tissue across multiple layers. With these platforms, changes in messenger RNA, microRNA and protein modification expression can be detected as well as changes in DNA methylation.

Many tools and methods are available for an analysis of the individual platforms. However, methods for cross-platform analysis or visualizations are very rare. Especially, no techniques are available that focus on integrated analysis of DNA methylation, mRNA, microRNA and protein modification data. Therefore, we developed multiple strategies that ease cross-platform analysis of data from the same set of biological samples. These strategies can also be employed for a subset of the mentioned platforms

and include tabular integrations, an extension of gene-set enrichment analysis to multiple platforms and pathway-based visualization techniques.

All of them are implemented into an easy-to-use application that bridges the gap from single dataset analysis to an integrated analysis of cross-platform data. The developed methods, included in the application, have been used to analyze the mentioned tumor datasets and some results are highlighted in the context of this publication.

Introduction

TODO: CLEMENS: Vor Abgabe suchen und vereinheitlichen von

- **gene and genome nomenclature**
- ***p*-value**
- **fold-change**
- **Ctnnb1**
- **Ha-ras**
- **up-regulated and down-regulated (Bindestrich vs. kein Bindestrich)**
- **microRNA // miRNA; (phospho-) protein // protein modification**
- **BE vs AE (analyses vs. analyzes)**
- **irgendwo erklren dass unsere fold-changes immer log2 based sind!**
- **Anfuehrungsstriche // doppelte vs. einfache**
- **im Material u Methoden teil kurzer abschnitt ber fold-changes log2 und thresholds die verwendet wurden. Weiterhin, welche miRNA db's.**

In the past, the microarray technology has primarily been used to detect changes in genome-wide gene expression of various biological systems. Usually, a set of biological samples is tested under various experimental and corresponding control conditions. This allows to calculate a relative change in gene expression between treatments and controls, but also between different treatments. Thus, for example, comparisons between differently mutated tumors or animal groups treated with different types of tumor promoters are possible [1]. It has previously been shown that in mice, treatment with the anticonvulsant phenobarbital following a single ip injection of the mutagenic initiator N-diethylnitrosamine (DEN) predominantly leads to liver tumors harboring activating point mutations in the gene encoding the transcription factor β-Catenin, *Ctnnb1* [2, 3]. In contrast, tumors isolated from animals treated with DEN only rather carry *Ha-ras*- or *B-raf*- mutations, both of which lead to constant activation of the MAP-kinase signaling pathway [2]. In the study described by Stahl *et al.*, the microarray approach was used to explore the effect of phenobarbital-treatment on gene expression in mouse liver tumors as compared to non-tumor control tissue [4].

Today, the application of high-throughput techniques evolved from gene expression to a broad range of other genomic, proteomic and epigenomic features. MicroRNAs (miRNAs) are short non-coding RNAs that regulate gene-expression by potentially binding to complementary mRNA targets [5]. MiRNA expression can be measured very similarly to mRNA expression, by defining specific probesets for those RNAs. There are several microarray platforms available today to detect the expression of miRNAs [6]. When studying the proteome in context of signaling networks, it is very important to know the phosphorylation state of proteins. In many cases, only phosphorylated proteins become active and propagate

a signal, e.g., in a pathway. The expression of basic and phosphorylated proteins can be measured with reverse-phase protein arrays [7]. Epigenetic alterations are often detected by measuring DNA methylation of cytosines. It is known that DNA methylation is the only covalent modification of DNA and hypermethylation of cytosines in gene promoters often leads to gene silencing of corresponding genes [8, 9]. The microarray technology has been extended to approaches that tile the promoter regions of defined genes to detect the degree of methylation in those regions [10].

Using all mentioned microarray technologies (mRNA, miRNA, protein modification, and DNA methylation) on the same set of biological samples leads to high-dimensional, heterogeneous cross-platform datasets. A visualization of the genomic and gene-regulatory context for all described data types is provided in Figure 1. For most of those individual platforms, many microarray analysis methods are available that go beyond the usual calculation and comparison of fold changes or other statistical measures. But some interesting biological findings are not covered by single platform analysis. The biological system of a cell includes all levels of regulation and omics data expression and therefore, many conclusions can only be drawn when studying multiple genomic layers in parallel.

We here present methods that ease the integrated study of of multiple high-throughput datasets, in which several microarray platforms have been used to investigate various genomic features in the same set of biological samples. The usefulness of those methods is demonstrated by application to a cross-platform dataset of mRNA, miRNA, protein modification, and DNA methylation data, obtained from *Ha-ras*- or *Ctnnb1*-mutated liver tumors in *mus musculus*. These algorithms are going beyond single gene or protein expression analysis (i.e., complementary to locus-specific approaches) and thus into an area of computational biology with a very sparse research density. Interactions detected in heterogeneous datasets have the power to, e.g., reveal novel and very promising biomarker candidates.

Results and Discussion

Gene-based integration of heterogeneous datasets

All proposed methods are validated on a dataset of 19 mouse samples: 10 tumors (3 *Ha-ras*-mutated and not treated with phenobarbital (PB), and 7 *Ctnnb1*-mutated after PB-treatment) and 9 corresponding control tissue samples from the same animals (3 untreated, 6 PB-treated). Messenger RNA, microRNA, DNA methylation and protein modification expression datasets have been measured for all of those samples. We propose to integrate all datasets on gene-level. This is obvious for messenger RNA data, because mapping all probes on genes is a common task. For protein datasets, this is also straightforward, except that different protein modifications should not be mixed. Mapping DNA methylation data on genes requires to introduce a window around the transcription start site (TSS) of each gene. The Nimblegen chip design provides probes for approximately $8,000$ bps downstream and $3,000$ bps upstream of each TSS. But including all probes within this window would lead to a huge amount of peaks for one gene, which is difficult to interpret. It has been shown that usually, the proximal promoter region has a stronger effect on a gene's regulation than more distant regions [11, 12]. Furthermore, many CpG islands are close to the TSS or even reach into the 5' region of a gene [13]. And DNA methylation of CpG islands in promoter regions is crucial for the regulation of gene expression [14]. Putting all together, we decided to assign all probes to a gene that are within $-2,000$ bps and $+500$ bps of the TSS.

MicroRNAs infer into gene-regulation by binding to complementary mRNA targets [15]. Mapping of miRNA data on genes can be realized by taking the locus of miRNA transcripts or by mapping miRNAs on their corresponding mRNA targets. Both approaches are eligible and have been confirmed by numerous publications [16, 17]. However, the locus approach is difficult for integrated data analysis, because genes from neighboring loci are not necessarily co-regulated. And integration of different datasets is performed rather on a functional relationship than on a region-based relationship. Therefore, mapping from miRNA to genes is performed by mapping all probes from an miRNA to all genes, whose mRNAs are targets of

the miRNA. For this purpose, numerous public databases are available that contain information about miRNAs and corresponding mRNA targets [17]. We worked with a union of three miRNA target databases that only contain experimentally validated tarkets: miRecords v3, miRTarBase v2.4 and TarBase v5.0c [18–20].

Methods for integrated data analysis

Having all platforms integrated, we propose numerous methods for integrated representation, visualization and analysis of all platforms.

Data pairing

A first approach that is already suitable for any two datasets is 'data pairing'. Data pairing allows viewing two heterogeneous datasets at once. In the first step, each value from dataset one is mapped on the matching value in dataset two. Then, both values are displayed next to each other in a tabular view. This procedure allows a first glance on integrated data that comes from two platforms. It is especially useful for showing microRNA and mRNA target interactions: The miRNA is shown, together with p-value and fold change on the left-hand side, whereas the targeted mRNA is shown, together with p-value and fold change on the right-hand side. In the middle between both datasets, the relation is displayed. This includes the source of this target mapping (miRNA target database or prediction algorithm), the confidence of this interaction and the inferred miRNA effect on the mRNA. The InCroMAP application includes all required algorithms to pair any two datasets - including an automated target annotation for miRNA data, all described gene-centering preprocessing steps and the actual creation and rendering of the paired table. An example can be seen in Figure 2.

Paired analysis of mRNA and miRNA expression data in *Ctnnb1*-mutated tumors revealed miR-495 to be the second most strongly up-regulated miRNA in this tumor type, whereas its expression remains unchanged in *Ha-ras*-mutated tumors. Annotation of experimentally validated targets showed that miR-495 regulates the gene expression of the *Onecut*-transcription factor HNF-6 [21], which itself is down-regulated (see Figure 2). HNF-6 is known to stimulate the expression of glucose-6-phosphatase (*G6pc*) by binding at its promoter region [22]. The proposed mode of action of miRNA-mediated expression regulation is that an miRNA binds to a complementary mRNA sequence, thus either leading to its degradation or inhibiting its translation at the ribosome. The assumption that miR-495 inhibits the translation of HNF-6 and thus the expression of *G6pc* is a possible explanation for the lack of *G6pc* on protein level in mouse liver tumors [23]. The fact that miR-495 is only up-regulated in *Ctnnb1*- but not in *Ha-ras*-mutated tumors whereas both are deficient in G6PC protein indicates that different regulatory mechanisms apply to the two tumor types.

Tabular integration of multiple heterogeneous datasets

Integration of multiple datasets into a tabular view is similar to data pairing, except that more than two datasets are aligned and displayed next to each other. One important objective of multiple data integration is showing the relevant information and hiding irrelevant information, as a table might get very big if all information, contained in a paired data table is shown. Thus, one row is created for each gene and one column for each platform. A single summary value for each gene and platform is calculated and placed in the table. The calculation of the summary value largely depends on the platform and user preferences. There are arguments for taking the probe or peak with maximum differential expression but still, many researchers prefer taking the mean, e.g., for mRNA datasets. The InCroMAP application, in which this integrated analysis method is implemented, allows for selecting the preferred method to calculate the summary value for each platform.

To provide researchers more information than the single summary value of each platform and gene, a special tabular representation is used that allows to collapse and expand rows in a table. Each gene-node can be expanded to see all platforms, providing data for this gene. Every platforms itself can be further expanded to show more detailed information. This includes, all probes and their expression values for mRNA data or the corresponding protein and protein modification expression for protein datasets. Expanded DNA methylation nodes show all peaks in the promoter region of this gene or the single probes upon further expansion. Gene-based expansion of miRNA datasets will show all miRNAs that are targeting this gene, together with probe-level expression values if the miRNAs are further expanded. Figure 3 illustrates an example of this interactive tabular view.

The tabular integration of all platforms reflecting *Ctnnb1*-mutated tumors confirms a strong hyper-methylation in the promoter region of *Egfr*, resulting in an mRNA decrease (see Figures 3 and 4) [24]. Down-regulation of *Egfr* is usually a cause for cancer [25]. Interestingly, neither the basic protein nor any phosphorylated variant of EGFR shows strong expression changes.

Integrated gene set enrichment

A common data analysis method, especially for mRNA data, is the gene set enrichment analysis. This analysis is performed by creating a gene list (typically by selecting the genes, showing huge differential expression in treatment versus control) and searching for significant enrichments therein. Therefore, predefined gene sets are required, which can be pathways, gene ontology terms, genes that are regulated by certain transcription factors, or anything else. As a result, researchers can, for example, figure out which pathways are targeted by a certain treatment or compare targeted pathways of genotoxic and non-genotoxic compounds.

We propose a cross-platform extension of this procedure for multiple, integrated heterogeneous datasets. The method is depicted in Figure 5. As usually, platform-specific low-level processing and gene-centering procedures are performed for each platform. Each gene is then being assigned a p-value and fold change. Based on these values, a mixed gene pool of significant differentially expressed genes is created by picking and merging all genes, associated to probes on any platform that show a huge deviation between treatment and control. This includes genes with strong methylation changes in their promoters, target of differentially expressed microRNAs, protein modifications with strong differential expression, and genes that show strong changes in their mRNA level. On this mixed gene pool, an enrichment analysis is now performed by comparing it to predefined gene sets and using an hypergeometric distribution to calculate a p-value for each gene set. A more detailed description of this step can be found in numerous other publications [26, 27]. As a result, terms (e.g., pathways) are returned that are enriched across all platforms and not only effects of, for example, mRNA or miRNA data analysis. This provides a much broader insight into pathway changes in the organism, upon treatment with a specific compound, than enrichment analysis of a single platform.

Figure 6 shows a comparison between mRNA-only and an integrated gene set enrichment, using mRNA, miRNA, DNA methylation and protein modification data from *Ctnnb1*-mutated tumors. It is clearly visible that the integrated enrichment shows significant changes in many cancer related pathways, whereas the mRNA enrichment shows most significant changes in metabolic pathway maps. Having differentially regulated cancer-related pathways as a results in tumor tissue shows the comprehensiveness of an integrated enrichment, in contrast to platform-specific results, showing in detail which pathways are differentially regulated in this particular platform.

Pathway-based visualization of integrated datasets

Still today, only few visualization techniques are available for integrated datasets. A wide-spread data visualization technique is, creating region-based tracks in a specific file format (common formats are BED or WIG) and using the UCSC genome browser [28]. This technique is suitable to get an insight

into specific genome regions of interest, but it is not specialized on integrated data analysis and fails to give overall impressions. Especially if interesting genes lay on different chromosomes, it is not possible to visualize them together and if researchers do not already have genes of interest, the method fails completely to give a starting point in the visualized data.

Following up the common pathway enrichment analysis, integrated visualization of differentially regulated pathways and measured microarray data gives an overview of how compounds influence certain signaling networks. By changing node colors, shapes, adding new nodes to graphs or adding new labels, much information can be visualized directly in a pathway. Combinations of those possibilities allow visualizing mRNA fold change, miRNA fold change, protein modification expression data and DNA methylation information directly in a pathway.

To get a comprehensive overview of the regulation of metabolic processes and affected pathways in a microarray dataset, we propose to modify KEGG's 'Metabolic Pathways' map. This pathway visualizes compounds, enzymes and secondary pathways, which are involved in important metabolic processes in an organism (see Figure 7). To visualize specific microarray data and differentially regulated metabolic processes in this pathway, the color of various pathway elements is changed. All down-regulated enzymes (depicted as edges between compounds) are colored blue and all up-regulated enzymes are colored red. Stronger differential expression is indicated by more saturated colors. With this procedure, the expression of all contained enzymes is visualized, e.g., based on an mRNA dataset. Further, the metabolic overview pathway contains multiple rectangular nodes which are references to secondary metabolic pathways. The color of those referenced pathways can be changed to reflect the p-value of this pathway in an enrichment analysis. In other words, the color of referenced pathways can be changed to a more saturated color if the pathway is significantly differentially regulated in a microarray dataset and to a brighter color if it is less significantly differentially regulated. The resulting picture is an overview which metabolic processes and enzymes are up- or down-regulated in any input microarray dataset. See Figure 7 for an example.

The described method can further be extended to visualize data from multiple platforms in any particular pathway. For most color-based visualizations, we define blue to indicate down-regulation and red for up-regulation. More saturated colors indicate stronger differential expressions and white is used to visualize no differential expression. Grey is used to show that no data is available from the input dataset. Messenger RNA expression is typically available for the majority of nodes in a pathway. Hence, the background color of every node in a pathway is changed to reflect the mRNA expression change. Protein expression is visualized by adding a small colored box below each node. If multiple measurements are available for differentially modified proteins (e.g., acetylated or phosphorylated isoforms), a separate box is added for each modification. Each box is labeled according to the proteins modification.

Integrating microRNA data into the pathway visualization is more difficult, as pathways usually consist of protein coding genes and compounds. MicroRNAs are not included in pathways and thus, a connection from each miRNA to the nodes in a pathway must be established. As already described, miRNAs are integrated with other platforms by querying miRNA target databases and put the miRNAs in relation to their corresponding targets. The same approach is used for the pathway-based visualization. Each miRNA that has a target within the current pathway is added as small rectangle, which is colored according to miRNA expression. The connection to the target is depicted by a line from the miRNA to the node, corresponding to the target mRNA.

DNA methylation data can be interpreted as a trajectory in a defined window for each gene. This information must be summarized, in order to crete a brief overview for each gene. Since most researchers want to know if a gene is rather hyper- or hypomethylated, we recommend to inspect and visualize the DNA methylation peaks. Visualizing the mean or median is not very informative because small local peaks can already have a strong influence on gene expression. Hence, to get a single summary value for each gene, any peak detection algorithm can be applied to a DNA methylation dataset (see, e.g., the user's guide of Nimblegen's SignalMap software [29]) or the peak can be approximated by taking the probe with maximum differential expression on a normalized and smoothed DNA methylation dataset.

The summary value for DNA methylation data is visualized as a black bar in a rectangular shaped box with a predefined width on the left side of each pathway node. This black bar is drawn from the middle of the box to the left to indicate hypomethylation and from the middle to the right side to indicate hypermethylation. The total size of the bar is proportional to the summary value, i.e., the maximum DNA methylation peak. The aim of this visualization is giving a first hint if a gene promoter is differentially methylated. In the InCroMAP application, the gene can be selected to get an additional, more detailed plot of the actual DNA methylation trajectory in the corresponding promoter region (see Figure 4).

The global overview visualization of metabolic pathways in *Ha-ras*- and *Ctnnb1*-mutated tumors on mRNA level provided a first insight into the profound metabolic changes taking place in the tumors. The visualization, depicted in Figure 7, shows characteristic perturbations in the metabolism of *Ha-ras*-mutated as opposed to *Ctnnb1*-mutated tumors. Major transcriptional changes take place in key pathways of energy metabolism, such as glycolysis and gluconeogenesis, the citric acid cycle or the urea cycle. In general, many key enzymes of gluconeogenesis (PCK1, G6PC, etc.) are down-regulated whereas the expression of glucokinase (*Gck*), which catalyzes the first step of glycolysis, is up-regulated in *Ctnnb1* mutated tumors. Figure 9 shows mRNA fold changes and maximum DNA methylation peaks within parts of the glycolysis and gluconeogenesis pathway for both tumors. Further, there are some interesting examples in which promoter methylation correlates well with mRNA expression, for example the phosphoglucomutase (*Pgm*), *G6Pc* or glucose phosphate isomerase 1 (*Gpi1*). Also, some key enzymes of the citric acid cycle, such as the isocitrate dehydrogenase 3 alpha (IDH3a) or the citrate synthase (CS), are up-regulated in *Ctnnb1*-mutated tumors and down-regulated in *Ha-ras*-mutated tumors. This might indicate that the *Ctnnb1*-mutated tumor type uses glucose as fuel rather than synthesizing it *de novo*. Furthermore, the key enzymes of the urea cycle as well as several enzymes involved in amino acid catabolism are characteristically down-regulated in *Ctnnb1*-mutated tumors which is consistent with previous findings [30].

All of the above described visualization techniques allow for a joint visualization of mRNA, miRNA, DNA methylation and protein modification data (see Figure 8 for a detailed example of the described visualizations). The WNT signaling pathway constitutes a further interesting example for a joint visualization of the mentioned platforms (see Figure 10). It shows that gene and protein expression of the pathway targets and transcription factors *Myc*, *Jun*, *Fosl1* (fos-like antigen 1) and *Ccnd1* (Cyclin D1) are up-regulated. Interestingly, various β-Catenin inhibitory genes show an increased expression in the tumors with activating *Ctnnb1*-mutations which again indicates activation of negative regulatory mechanisms. These *Wnt*-inhibitors include *Gsk3b* and *Axin2*, which together form the β-Catenin degradation complex, *Nkd2* (naked cuticle homologue 2) and *Wnt* inhibitory factor 1 (*Wifi1*).

Materials and Methods

Tumor material

The samples used for the analyses presented here were taken from a previous experiment by Marx-Stoelting *et al.* [31]. In brief, male C3H mice received a single dose of DEN and were subsequently kept on a diet containing PB and a PB-free control diet, respectively. *Ctnnb1*-mutated tumors were isolated from the phenobarbital (PB)-treated mice whereas *Ha-ras*-mutated tumors were isolated from the animals which received a PB-free diet. Normal tissue samples were taken from the same livers.

Microarray Platforms

We quantified genome-wide mRNA transcript expression using the Affymetrix MG430 2.0 platform. Micro RNA expression was profiled using the Agilent G4472A platform. DNA methylation states of gene promoters were detected using Nimblegen HD2.1 Deluxe Promoter tiling arrays (MM9). These arrays

cover sequence regions spanning from -8kbps to +3kbps relative to the transcription start site (TSS) of a certain gene. Each sequence region is tiled with hundreds of probes that are about 50bps in size. For the quantitative analysis of protein expression and modification we employed Zeptosens ZeptoMARK reverse-phase protein arrays. Owing to specific antibodies these arrays facilitate the distinction between different modifications of a protein (e.g., phosphorylated and unphosphorylated protein forms), as well as the accurate and reproducible quantification of proteins in multiple samples.

Low-level data processing

As the microarray platforms employed for this study differ in their design and application, the applied preprocessing steps have to be adapted to the individual characteristics of each platform. Typically, these preprocessing steps involve the quality control, normalization, annotation, and summarization (i.e., mapping from probes to genes) of the data. While these steps mostly require platform-specific methods, the statistical analysis was to a large extent standardized for all platforms. In the following, we describe how the above-mentioned preprocessing steps were conducted for each platform.

Messenger RNA

The Affymetrix mRNA expression data (CEL files) containing the raw probe intensities were normalized using the Robust Multichip Average (RMA) method and the quality of the experiments was assessed using diverse plots and statistics implemented in the package arrayQualityMetrics for R/Bioconductor [32]. On the basis of extensive quality controls, we concluded that all arrays had sufficient quality.

A moderated t-statistic was chosen to detect differentially expressed genes (implemention from limma package for R/Bioconductor) [33]. In order to correct for testing multiple genes and to ensure a false discovery rate less than 0.05, the Benjamini-Hochberg method was applied. Additionally, fold change cutoffs of 2 and 0.5 were used to select upregulated and downregulated genes, respectively.

Micro RNA

Background correction and summarization of the probe signal levels corresponding to a individual miRNAs were performed using the Agilent Feature Extraction (AFE) image analysis algorithm implemented in the AgiMicroRna package for R/Bioconductor [34]. Probe sets which were in all samples flagged as not expressed by the AFE software, were excluded from further analysis. Afterwards, the mean fold changes were computed for each replicate group and p-values were computed using a moderated t-test with FDR correction as previously described for mRNA expression data.

In order to investigate miRNA-mRNA interactions, each miRNA represented on the microarray was linked to its mRNA targets based on experimentally confirmed interaction data extracted from public databases (miRecords v3 [18], miRTarBase v2.4 [19], TarBase v5.0c [20]). Interactions were merged in a non-redundant manner by removing duplicates and mapping all gene identifiers to common NCBI Gene IDs.

DNA methylation

In order to correct for intensity biases between the two channels, namely input DNA, and methylated DNA (Me-DNA) enriched by immunoprecipitation (IP), we used locally weighted scatterplot smoothing (LOESS) by polynomial regression. Subsequently, we used quantile normalization to correct for effects caused by experimental variation. For the normalization within arrays (i.e., LOESS) and the normalization between arrays (i.e., quantile normalization), we employed the limma package for R Bioconductor [33]. To alleviate probe-specific effects, caused by differing probe affinities, the normalized probe signal levels were smoothed, by computing a weighted mean across the intensities of neighboring probes (implemented in MEDME package for R/Bioconductor) [35]. We determined MeDIP enrichment

levels, by computing the log-ratios from the red channel (enriched Me-DNA) to the green channel (input DNA) probe intensities. As these MeDIP enrichment levels were shown to be non-linearly related with absolute DNA methylation levels, we used the MEDME (Model for Experimental Data with MeDIP Enrichment) approach proposed by Pelizzola *et al.* This procedure involves fitting a sigmoidal model on a fully methylation dataset needed for calibration [35]. Based on this model absolute methylation levels (AML) were inferred from the computed log-ratios for each probe. Then, we computed mean fold-changes for each sample group (i.e., tumor type) and determined p-values using the same statistical method as for mRNA expression data.

In order to reveal interactions between differential DNA methylation and gene expression, we summarized the measurements of probes which interogate DNA methylation in the same proximal promoter. Proximal promoters were defined as regions spanning from -2kbps to 0.5kbps relative to the TSS. As changes in DNA methylation are typically focused to small regulatory regions, we used the highest peak as a summary value for the DNA methylation status of a promoter.

Protein modification data

Each analyzed sample was represented in four different protein concentrations in duplicated spots on Zeptosens ZeptoMARK arrays. Using the Zeptosens ZeptoVIEWPro software, the spot intensities were determined from the microarray images. Subsequently, linear extrapolation was used to extend the background corrected mean intensities of the duplicated spots to the highest of the four concentrations and the mean fluorescence intensity (MFI) was computed. To compensate for the effect of non-specific binding caused by the secondary antibody, we computed blank-corrected signals by substracting the MFI determined from the corresponding blank spots, which were treated solely with the secondary antibody. The blank-corrected MFI was then normalized by multiplication with protein stain factors, which were determined from a protein stain array measuring the relative amount of protein for each spot. The blank-corrected, normalized spot intensities were exported as a spreadsheet file.

All analytes, i.e., proteins or protein modifications quantified in the liver samples, were annotated with UniProt IDs to facilitate the interpretation of the data in the context of signaling pathways. Next, the data was log-transformed and centered around the median intensity values observed for control samples to ease interpretability and to obtain a symmetrical scale of protein and protein modification levels. Measurements for which the background noise (i.e., signal of secondary antibody) was higher than the combined foreground and background signal (i.e., signal of primary and secondary antibody), were treated as missing values (¡ 1%) and imputed using the k-Nearest-Neighbor (kNN) algorithm. For the detection of differential protein expression, we employed the same procedure as previously described for mRNA data.

Implementation and availability

All described methods are implemented in an easy-to-use tool with a graphical user interface (GUI), called InCroMAP. This Java^{TM} application supports importing processed mRNA, microRNA, DNA methylation and protein modification datasets. Interactive methods for single dataset analysis are provided as well as all described cross-platform analysis and visualization methods. All provided methods can be customized with many options that control, e.g., how expression values from multiple probes for one gene should be summarized or the minimum fold change value that is required for the most saturated color in a pathway-based visualization. The application uses and includes KEGGtranslator to visualize pathways from the KEGG database [36, 37] and is freely available under the LGPL version 3 license from `http://www.cogsys.cs.uni-tuebingen.de/software/InCroMAP/`.

Acknowledgments

We gratefully acknowledge contributions from Andreas Dräger and Finja Büchel, as well as the whole MARCAR consortium.

References

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, et al. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531–537.

2. Aydinlik H, Nguyen TD, Moennikes O, Buchmann A, Schwarz M (2001) Selective pressure during tumor promotion by phenobarbital leads to clonal outgrowth of beta-catenin-mutated mouse liver tumors. Oncogene 20: 7812–7816.

3. Rignall B, Braeuning A, Buchmann A, Schwarz M (2011) Tumor formation in liver of conditional β-catenin-deficient mice exposed to a diethylnitrosamine/phenobarbital tumor promotion regimen. Carcinogenesis 32: 52–57.

4. Stahl S, Ittrich C, Marx-Stoelting P, Khle C, Ott T, et al. (2005) Effect of the tumor promoter phenobarbital on the pattern of global gene expression in liver of connexin32-wild-type and connexin32-deficient mice. Int J Cancer 115: 861–869.

5. He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet 5: 522–531.

6. Hoheisel JD (2006) Microarray technology: beyond transcript profiling and genotype analysis. Nat Rev Genet 7: 200–210.

7. Pirnia F, Pawlak M, Thallinger GG, Gierke B, Templin MF, et al. (2009) Novel functional profiling approach combining reverse phase protein microarrays and human 3-D ex vivo tissue cultures: expression of apoptosis-related proteins in human colon cancer. Proteomics 9: 3535–3548.

8. Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell 128: 683–692.

9. Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. Cell 128: 669–681.

10. Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, et al. (2006) Microarray-based DNA methylation profiling: technology and applications. Nucleic Acids Res 34: 528–542.

11. Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM (2006) Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Res 16: 1–10.

12. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, et al. (2005) A high-resolution map of active promoters in the human genome. Nature 436: 876–880.

13. Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196: 261–282.

14. Razin A, Cedar H (1991) DNA methylation and gene expression. Microbiol Rev 55: 451–458.

15. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281–297.

16. Luk JM, Burchard J, Zhang C, Liu AM, Wong KF, et al. (2011) DLK1-DIO3 genomic imprinted microRNA cluster at 14q32.2 defines a stemlike subtype of hepatocellular carcinoma associated with poor survival. J Biol Chem 286: 30706–30713.

17. Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG (2009) Lost in translation: an assessment and perspective for computational microRNA target identification. Bioinformatics 25: 3049–3055.

18. Xiao F, Zuo Z, Cai G, Kang S, Gao X, et al. (2009) miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res 37: D105-10.

19. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, et al. (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res 39: D163-9.

20. Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG (2009) The database of experimentally supported targets: a functional update of TarBase. Nucleic Acids Res 37: D155-8.

21. Simion A, Laudadio I, Prévot PP, Raynaud P, Lemaigre FP, et al. (2010) MiR-495 and miR-218 regulate the expression of the Onecut transcription factors HNF-6 and OC-2. Biochem Biophys Res Commun 391: 293–298.

22. Beaudry JB, Pierreux CE, Hayhurst GP, Plumb-Rudewiez N, Weiss MC, et al. (2006) Threshold levels of hepatocyte nuclear factor 6 (HNF-6) acting in synergy with HNF-4 and PGC-1alpha are required for time-specific gene expression during liver development. Mol Cell Biol 26: 6037–6046.

23. Weber G, Cantero A (1955) Glucose-6-phosphatase activity in normal, pre-cancerous, and neoplastic tissues. Cancer Res 15: 105–108.

24. Montero AJ, Díaz-Montero CM, Mao L, Youssef EM, Estecio M, et al. (2006) Epigenetic inactivation of EGFR by CpG island hypermethylation in cancer. Cancer Biol Ther 5: 1494–1501.

25. Zhang H, Berezov A, Wang Q, Zhang G, Drebin J, et al. (2007) ErbB receptors: from oncogenes to targeted cancer therapies. J Clin Invest 117: 2051–2058.

26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545–15550.

27. Backes C, Keller A, Kuentzer J, Kneissl B, Comtesse N, et al. (2007) GeneTrail–advanced gene set enrichment analysis. Nucleic Acids Res 35: W186–W192.

28. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome Res 12: 996–1006.

29. NimbleGen Systems Inc (2006) SignalMap User's Guide. Version 1.9. Available from NimbleGen at www.nimblegen.com/products/lit/signalmap1.9usersguide.pdf. Accessed 2012 Mar 22.

30. Stahl S, Ittrich C, Marx-Stoelting P, Khle C, Altug-Teber O, et al. (2005) Genotype-phenotype relationships in hepatocellular tumors from mice and man. Hepatology 42: 353–361.

31. Marx-Stoelting P, Mahr J, Knorpp T, Schreiber S, Templin MF, et al. (2008) Tumor promotion in liver of mice with a conditional Cx26 knockout. Toxicol Sci 103: 260–267.

32. Kauffmann A, Gentleman R, Huber W (2009) arrayqualitymetrics–a bioconductor package for quality assessment of microarray data. Bioinformatics 25: 415–416.

33. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article3.

34. Lpez-Romero P (2011) Pre-processing and differential expression analysis of agilent microrna arrays using the agimicrorna bioconductor library. BMC Genomics 12: 64.

35. Pelizzola M, Koga Y, Urban AE, Krauthammer M, Weissman S, et al. (2008) Medme: an experimental and analytical methodology for the estimation of dna methylation levels based on microarray derived medip-enrichment. Genome Res 18: 1652–1659.

36. Wrzodek C, Dräger A, Zell A (2011) KEGGtranslator: visualizing and converting the KEGG PATHWAY database to various formats. Bioinformatics 27: 2314–2315.

37. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, et al. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34: D354–D357.

Figure Legends

Tables

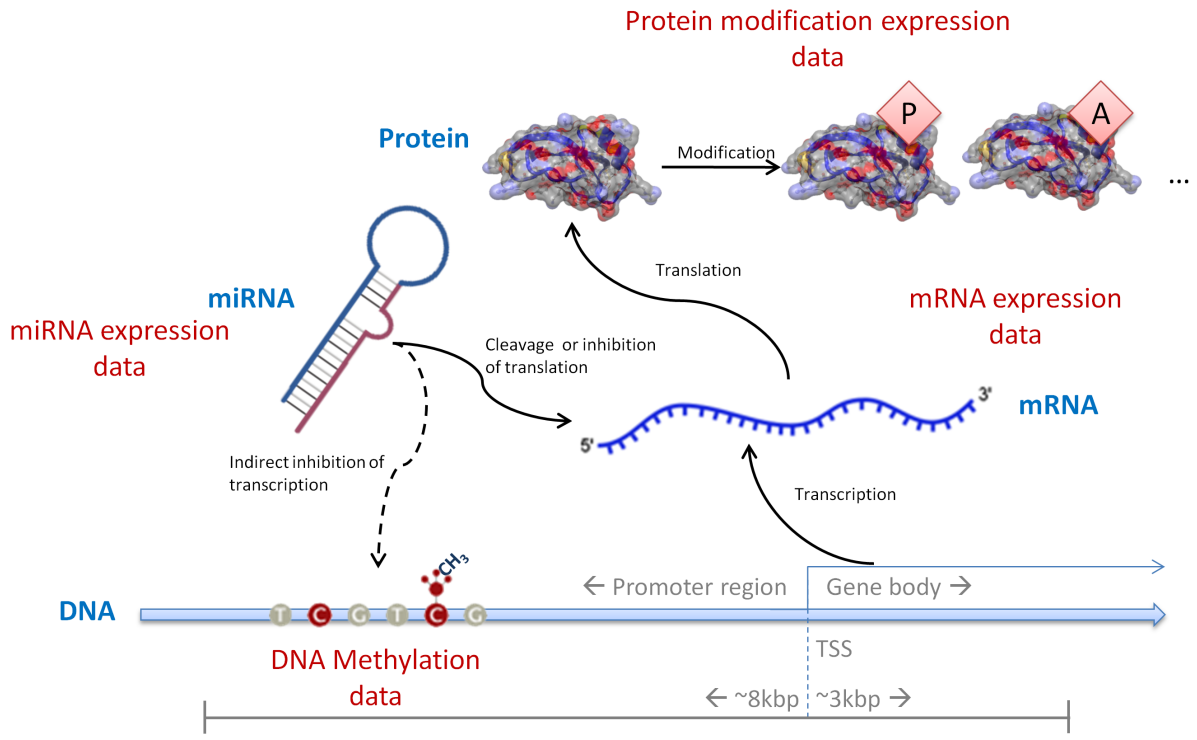
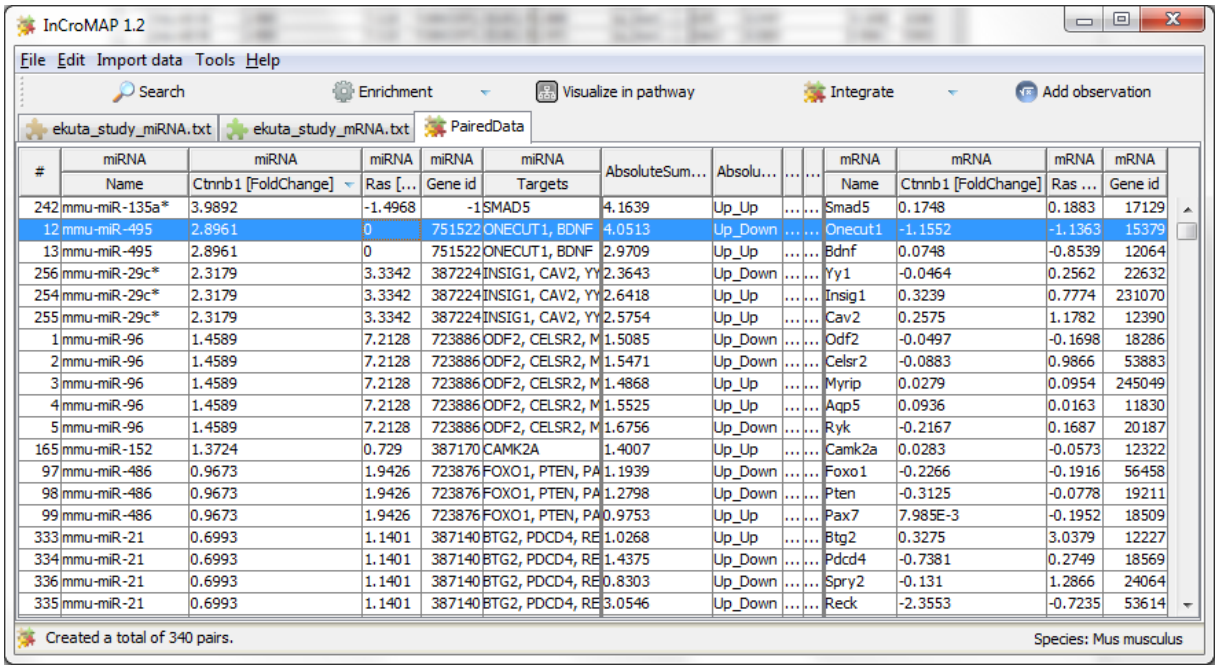


Figure 1. Visualization of different genomic and regulatory layers for which we propose cross-platform analysis and visualization methods. At the bottom of the figure, a DNA sequence is depicted for which methylation data is available. This DNA is transcribed to an mRNA. The transcription might be regulated by methylated regions on the gene promoter. Furthermore, miRNAs might inhibit the translation from mRNA to a protein. Both, mRNA and miRNA expression can be measured with microarrays. In the end, translated proteins might get modified, e.g., by phosphorylation or acetylation. The expression of basic proteins and specific modifications can be determined, e.g., by using reverse-phase protein arrays with several antibodies.



#	miRNA Name	miRNA Ctnnb1 [FoldChange]	miRNA Ras [...	miRNA Gene id	miRNA Targets	AbsoluteSum...	Absolu...	...	...	mRNA Name	mRNA Ctnnb1 [FoldChange]	mRNA Ras ...	mRNA Gene id
242	mmu-miR-135a*	3.9892	-1.4968	-1	SMAD5	4.1639	Up_Up	...	...	Smad5	0.1748	0.1883	17129
12	mmu-miR-495	2.8961	0	751522	ONECUT1, BDNF	4.0513	Up_Down	...	...	Onecut1	-1.1552	-1.1363	15379
13	mmu-miR-495	2.8961	0	751522	ONECUT1, BDNF	2.9709	Up_Up	...	...	Bdnf	0.0748	-0.8539	12064
256	mmu-miR-29c*	2.3179	3.3342	387224	INSIG1, CAV2, YY	2.3643	Up_Down	...	...	Yy1	-0.0464	0.2562	22632
254	mmu-miR-29c*	2.3179	3.3342	387224	INSIG1, CAV2, YY	2.6418	Up_Up	...	...	Insig1	0.3239	0.7774	231070
255	mmu-miR-29c*	2.3179	3.3342	387224	INSIG1, CAV2, YY	2.5754	Up_Up	...	...	Cav2	0.2575	1.1782	12390
1	mmu-miR-96	1.4589	7.2128	723886	ODF2, CELSR2, M	1.5085	Up_Down	...	...	Odf2	-0.0497	-0.1698	18286
2	mmu-miR-96	1.4589	7.2128	723886	ODF2, CELSR2, M	1.5471	Up_Down	...	...	Celsr2	-0.0883	0.9866	53883
3	mmu-miR-96	1.4589	7.2128	723886	ODF2, CELSR2, M	1.4868	Up_Up	...	...	Myrip	0.0279	0.0954	245049
4	mmu-miR-96	1.4589	7.2128	723886	ODF2, CELSR2, M	1.5525	Up_Up	...	...	Aqp5	0.0936	0.0163	11830
5	mmu-miR-96	1.4589	7.2128	723886	ODF2, CELSR2, M	1.6756	Up_Down	...	...	Ryk	-0.2167	0.1687	20187
165	mmu-miR-152	1.3724	0.729	387170	CAMK2A	1.4007	Up_Up	...	...	Camk2a	0.0283	-0.0573	12322
97	mmu-miR-486	0.9673	1.9426	723876	FOXO1, PTEN, PA	1.1939	Up_Down	...	...	Foxo1	-0.2266	-0.1916	56458
98	mmu-miR-486	0.9673	1.9426	723876	FOXO1, PTEN, PA	1.2798	Up_Down	...	...	Pten	-0.3125	-0.0778	19211
99	mmu-miR-486	0.9673	1.9426	723876	FOXO1, PTEN, PA	0.9753	Up_Up	...	...	Pax7	7.985E-3	-0.1952	18509
333	mmu-miR-21	0.6993	1.1401	387140	BTG2, PDCD4, RE	1.0268	Up_Up	...	...	Btg2	0.3275	3.0379	12227
334	mmu-miR-21	0.6993	1.1401	387140	BTG2, PDCD4, RE	1.4375	Up_Down	...	...	Pdcd4	-0.7381	0.2749	18569
336	mmu-miR-21	0.6993	1.1401	387140	BTG2, PDCD4, RE	0.8303	Up_Down	...	...	Spry2	-0.131	1.2866	24064
335	mmu-miR-21	0.6993	1.1401	387140	BTG2, PDCD4, RE	3.0546	Up_Down	...	...	Reck	-2.3553	-0.7235	53614

Figure 2. Pairing of miRNA (left-hand side) and mRNA (right-hand side) datasets reveals miR-495 as a potential regulator of the HNF-6 coding gene *Onecut1* in *Ctnnb1*-mutated tumors. Data pairing can be used to create a tabular representation of any two datasets. In this example, the effect of differentially expressed microRNAs on corresponding mRNA targets is visualized. The left-hand part of the table shows the miRNA data, sorted descending by $\log_2$ fold change expression in *Ctnnb1*-mutated tumors. The right-hand side of the table shows the targeted mRNAs, together with their fold change. Between both parts, some convenient information is shown, i.e., the absolute sum of miRNA and target mRNA fold changes in *Ctnnb1*-mutated tumors and a textual representation of miRNA and corresponding mRNA expression. This table shows that miR-495 is up-regulated and the corresponding mRNA target, *Onecut1*, is down-regulated.

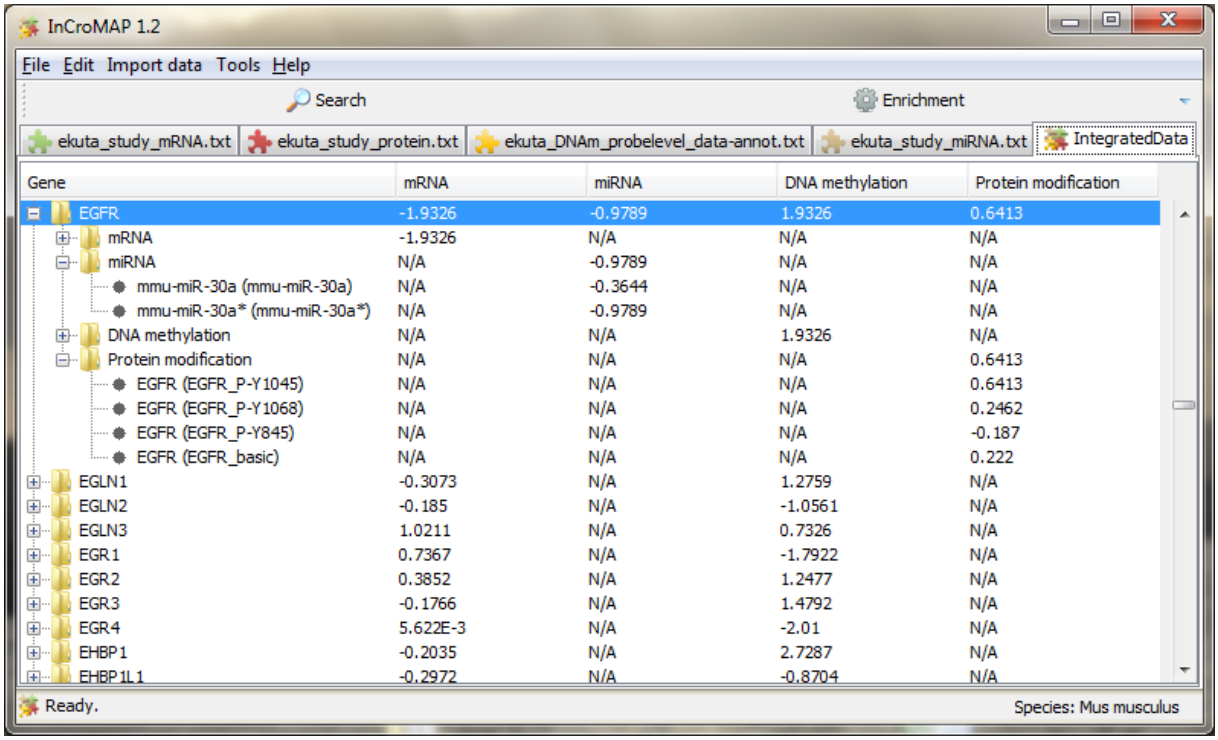


Figure 3. Multiple integration of four different platforms from *Ctnnb1* mutated tumors shows an mRNA decrease of *Egfr* as a potential effect of DNA methylation increase in the promoter region. This tabular visualization integrates data from four different platforms in an expandable, gene-based manner. On the first layer, each row corresponds to one gene and each column to one platform. A summary value is displayed for each gene and platform. If expanded, the second layer shows different groups for mRNA, miRNA, DNA methylation and protein modification data. If these are further expanded, the single probes, miRNAs targeting this mRNA or protein modifications are shown, together with the corresponding expression fold change. 'N/A' indicates either 'not applicable', e.g., the protein expression value for an mRNA probe, or 'data not available'. This tabular cross-platform integration of $\log_2$ fold-changes shows that the promoter region of *Egfr* has a maximum DNA methylation peak of 1.93 and a minimum mRNA expression of -1.93. The mRNA is the target of two microRNAs and next to the basic protein, the expression of three different phosphoforms has been measured. A subsequent analysis of DNA methylation in the promoter region of *Egfr* is depicted in Figure 4.

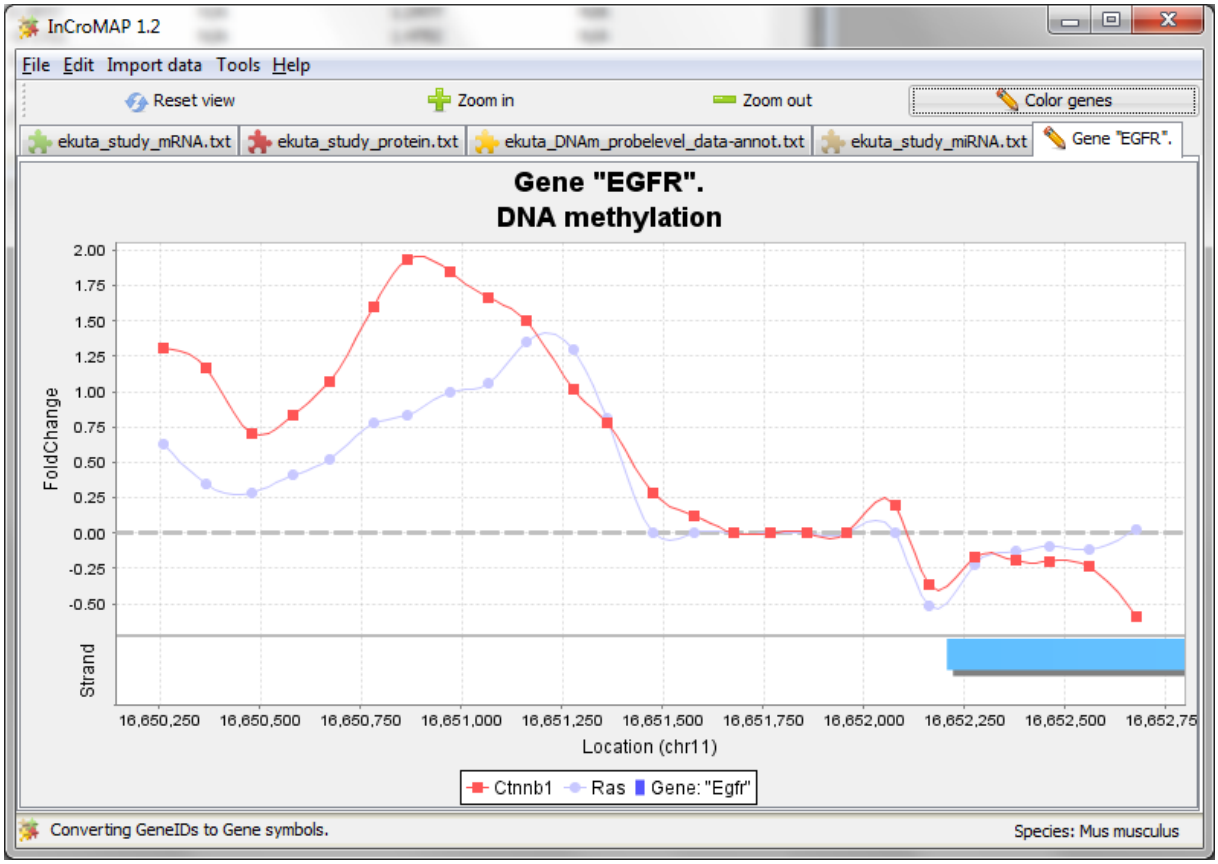


Figure 4. DNA methylation in the promoter region of *Egfr* in *Ctnnb1*-mutated tumors. This picture depicts the DNA methylation and mRNA expression changes for a region surrounding the transcription start site of *Egfr* by $-2,000$ and $+500$ bps. For DNA methylation, the fold changes of all probes in this region are visualized and connected by a curve. The gene body of *Egfr* is depicted on the forward strand with a big box on the lower part of the picture. The blue-color indicates the mRNA expression of *Egfr* in *Ctnnb1*-mutated tumors (blue means down-regulation, red would indicate an up-regulation and color saturation denotes the intensity of the differential expression). In this example, the mRNA has a maximum differential $\log_2$ fold change of -1.93.

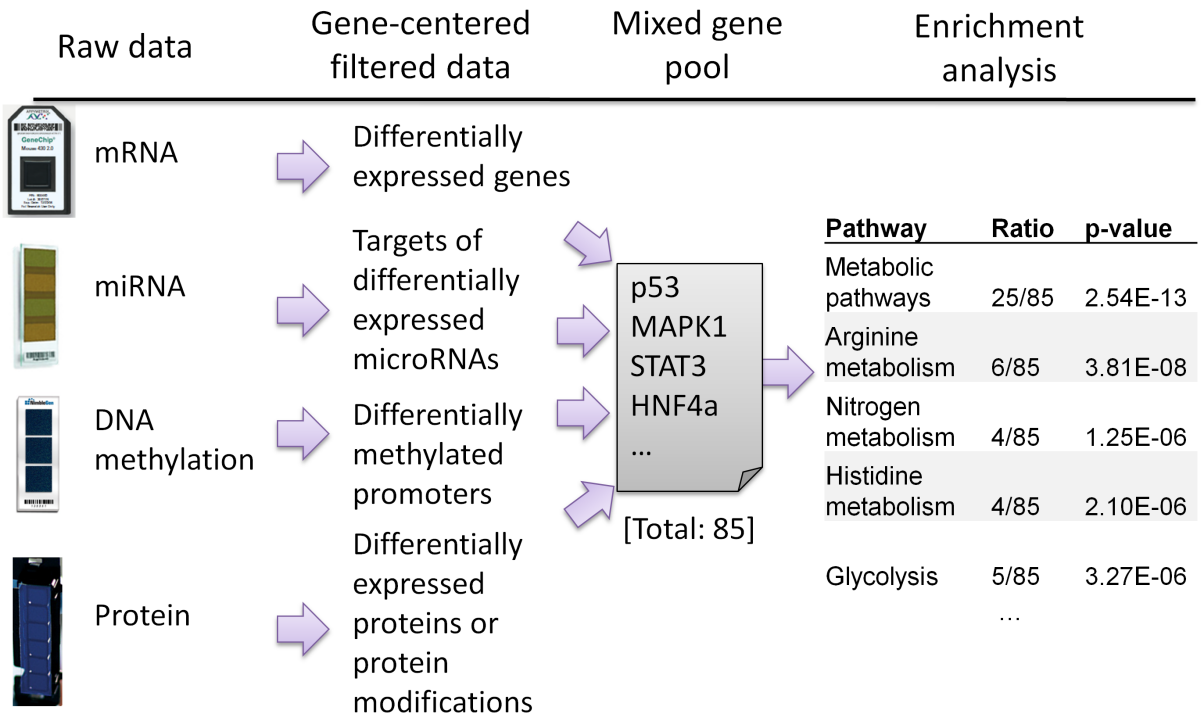


Figure 5. Proposed procedure for a cross-platform gene-set enrichment analysis. This flowchart depicts the major steps, involved in the creation of an integrated enrichment. The first two steps involve the evaluation of different datasets and extracting genes with strong platform-based changes as opposed to their corresponding control. In the third step, a mixed gene pool is created by joining all those gene lists (in this example, the result is a list of 85 genes). Finally, an enrichment is performed by comparing this list of genes to predefined gene sets and calculating a p-value, e.g., by using a hypergeometric distribution (see Subramanian *et al.* for more information [26]).

#	ID	Name	List ratio	BG ratio	P-value	Q-value
1	path:mmu05215	Prostate cancer	31/3209	90/18080	6.543E-5	0.0145
2	path:mmu05200	Pathways in cancer	80/3209	326/18080	4.21E-4	0.0467
3	path:mmu04350	TGF-beta signaling pathway	27/3209	85/18080	6.877E-4	0.0509
4	path:mmu00910	Nitrogen metabolism	11/3209	23/18080	7.084E-4	0.0393
5	path:mmu05216	Thyroid cancer	13/3209	30/18080	7.428E-4	0.033
6	path:mmu05221	Acute myeloid leukemia	20/3209	57/18080	8.312E-4	0.0308
7	path:mmu04110	Cell cycle	36/3209	128/18080	1.179E-3	0.0374
8	path:mmu05213	Endometrial cancer	18/3209	52/18080	1.678E-3	0.0466
9	path:mmu05210	Colorectal cancer	21/3209	65/18080	1.888E-3	0.0466
10	path:mmu04540	Gap junction	26/3209	88/18080	2.37E-3	0.0526
11	path:mmu00250	Alanine, aspartate and glutamate metabolism	13/3209	34/18080	2.64E-3	0.0533
12	path:mmu05218	Melanoma	22/3209	72/18080	3.064E-3	0.0567
13	path:mmu05222	Small cell lung cancer	25/3209	87/18080	3.957E-3	0.0676
14	path:mmu05214	Glioma	20/3209	66/18080	4.838E-3	0.0767
15	path:mmu04976	Bile secretion	21/3209	71/18080	5.289E-3	0.0783
16	path:mmu04510	Focal adhesion	48/3209	200/18080	5.459E-3	0.0757
17	path:mmu05219	Bladder cancer	14/3209	43/18080	8.303E-3	0.1084
18	path:mmu00561	Glycerolipid metabolism	16/3209	52/18080	8.801E-3	0.1085

Integrated enrichment
(mRNA, miRNA, DNAm, protein)

#	ID	Name	List ratio	BG ratio	P-value	Q-value
1	path:mmu00910	Nitrogen metabolism	10/937	23/18080	7.702E-8	1.502E-5
2	path:mmu04976	Bile secretion	15/937	71/18080	2.275E-6	2.218E-4
3	path:mmu01100	Metabolic pathways	95/937	1203/18080	9.317E-6	6.056E-4
4	path:mmu00330	Arginine and proline metabolism	12/937	56/18080	1.936E-5	9.437E-4
5	path:mmu04110	Cell cycle	19/937	128/18080	2.318E-5	9.041E-4
6	path:mmu00260	Glycine, serine and threonine metabolism	9/937	34/18080	3.647E-5	1.185E-3
7	path:mmu00250	Alanine, aspartate and glutamate metabolism	9/937	34/18080	3.647E-5	1.016E-3
8	path:mmu03320	PPAR signaling pathway	13/937	82/18080	2.127E-4	5.185E-3
9	path:mmu00480	Glutathione metabolism	10/937	54/18080	3.14E-4	6.804E-3
10	path:mmu00340	Histidine metabolism	7/937	28/18080	3.835E-4	7.479E-3
11	path:mmu05146	Amoebiasis	15/937	116/18080	6.413E-4	0.0114
12	path:mmu04350	TGF-beta signaling pathway	12/937	85/18080	9.885E-4	0.0161
13	path:mmu00561	Glycerolipid metabolism	9/937	52/18080	9.887E-4	0.0148
14	path:mmu04512	ECM-receptor interaction	12/937	86/18080	1.09E-3	0.0152
15	path:mmu04540	Gap junction	12/937	88/18080	1.317E-3	0.0171
16	path:mmu04964	Proximal tubule bicarbonate reclamation	5/937	20/18080	2.593E-3	0.0316
17	path:mmu05200	Pathways in cancer	28/937	326/18080	2.905E-3	0.0333
18	path:mmu05216	Thyroid cancer	6/937	30/18080	3.182E-3	0.0345

mRNA enrichment

Figure 6. Comparison of an integrated and mRNA enrichment, based on data from *Ctnnb1*-mutated tumors. This comparison shows the difference between a KEGG pathway enrichment that is based only on mRNA data (right-hand side) and the result of an integrated enrichment (left-hand side). The integrated enrichment includes genes with differentially expressed mRNA, targets of differentially expressed microRNAs, differently expressed protein or protein modifications and genes with strong DNA methylation changes in their promoters. It is clearly visible that the integrated enrichment provides a much broader insight into the pathway-changes in a tumor than using only a single dataset: Many cancer-related pathways are significantly impaired in the integrated enrichment, whereas the mRNA enrichment lists mainly metabolic pathways, which does only reflect the transcriptional changes.

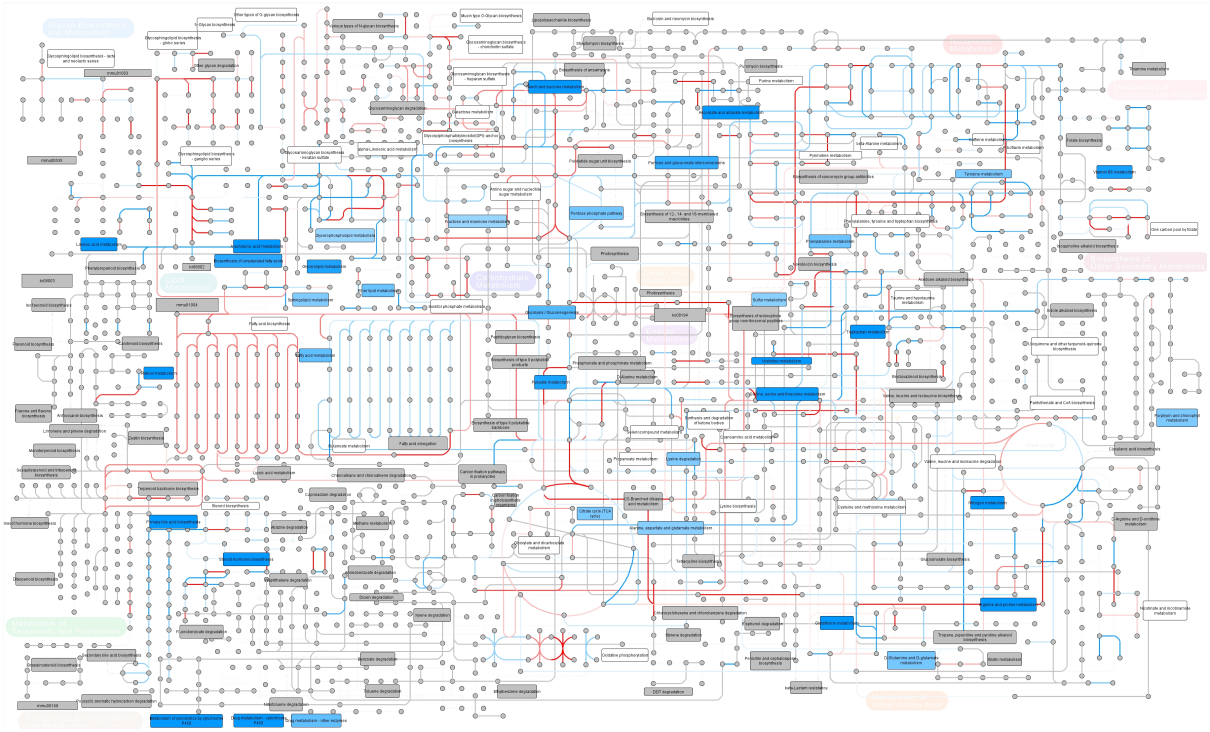


Figure 7. Overview of transcriptional metabolic changes in *Ha-ras*- and *Ctnnb1*-mutated tumors. This visualization unifies pathway information, gene-set enrichment results and mRNA expression data to a single metabolic overview picture. It is based on the mRNA differences between *Ha-ras*- and *Ctnnb1*-mutated tumors. Compounds are depicted by grey circles and edges between them represent different enzymes. These edges are colored red, if the corresponding mRNA is higher expressed in *Ha-ras*-mutated tumors and blue, if the corresponding mRNA is higher expressed in *Ctnnb1*-mutated tumors. Grey indicates no differential expression. The bigger rectangles are references to other KEGG pathways and have been colored by significance (p-value) as a result of a pathway enrichment on the fold change differences between *Ha-ras*- and *Ctnnb1*-mutated tumors: A more saturated blue color indicates a lower p-value, white indicates p-values > 0.05 and grey is used if no differential expression between both tumor types has been observed in the respective pathway. You can find a full-sized version of this picture in the supplement.

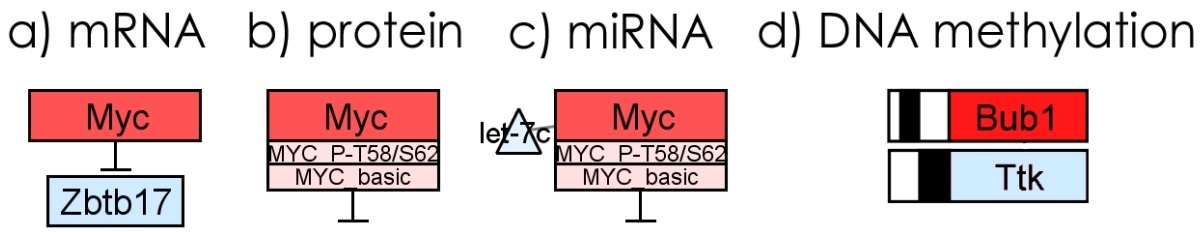


Figure 8. Pathway-based visualization of different platforms. This figure shows how data from different platforms can be jointly visualized in a pathway. A) Messenger RNA expression changes are visualized as node color. Red is used to indicate up-regulation, blue is used for down-regulation and white indicates no differential expression (i.e., fold-change is zero). More saturated colors indicate stronger differential expression. B) Protein modification data is visualized by adding small boxes below the node. In this example, one box is used for the basic protein and one for a phosphorylated isoform. 'MYC P-T58/S62' indicates an antibody that detects changes specifically for MYC proteins that are phosphorylated on Threonine 58 and Serine 62. The color scheme is the same as for the mRNA. C) MicroRNAs are added to pathways as small triangles that are connected to their potential mRNA targets. MicroRNA expression data is then visualized by changing to color of the microRNA as described above. D) DNA methylation data is summarized by taking the maximum differential peak in the corresponding gene promoter. This peak is visualized by a black bar that stretches from the middle to the left to indicate hypomethylation and from the middle to the right to indicate hypermethylation. In the depicted examples, *Bub1* shows a hypomethylation ($\log_2$ fold change of approximately 1) and *Ttk* shows a strong hypermethylation ($\log_2$ fold change of ≥ 1.5). Figure 10 shows an example in which all four visualization techniques are used to create a joint pathway-based visualization of data from heterogenous platforms.

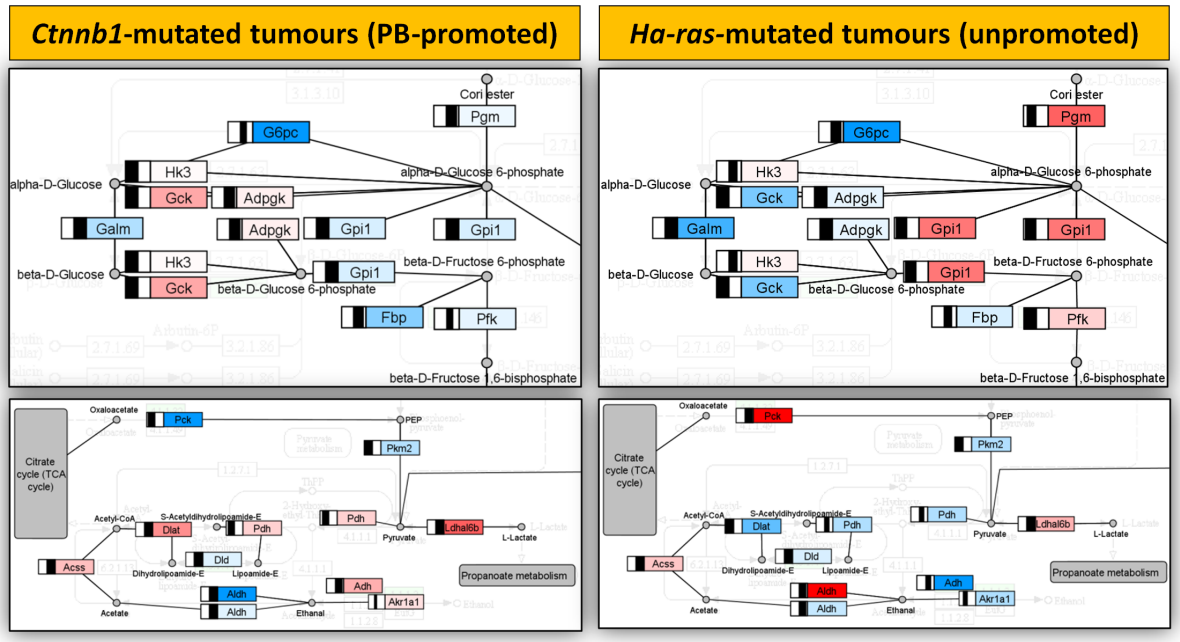


Figure 9. Integrated view of transcriptional and epigenetic changes in parts of the Glycolysis/ Gluconeogenesis pathway during profiling of *Ha-Ras* and *Ctnnb1* mutated tumors. The pictures on the left show mRNA and DNA methylation changes in the *Ctnnb1*-mutated tumors, which is mediated by phenobarbital (PB). For comparison, the same parts of the pathway are visualized on the right-hand side with data from *Ha-ras*-mutated tumors, which were not treated with PB. Briefly, nodes are colored according to mRNA expression changes (red means up-regulation and blue means down-regulation) and the black bar left of each node indicates DNA methylation changes (please see Figure 8 for a more detailed legend).

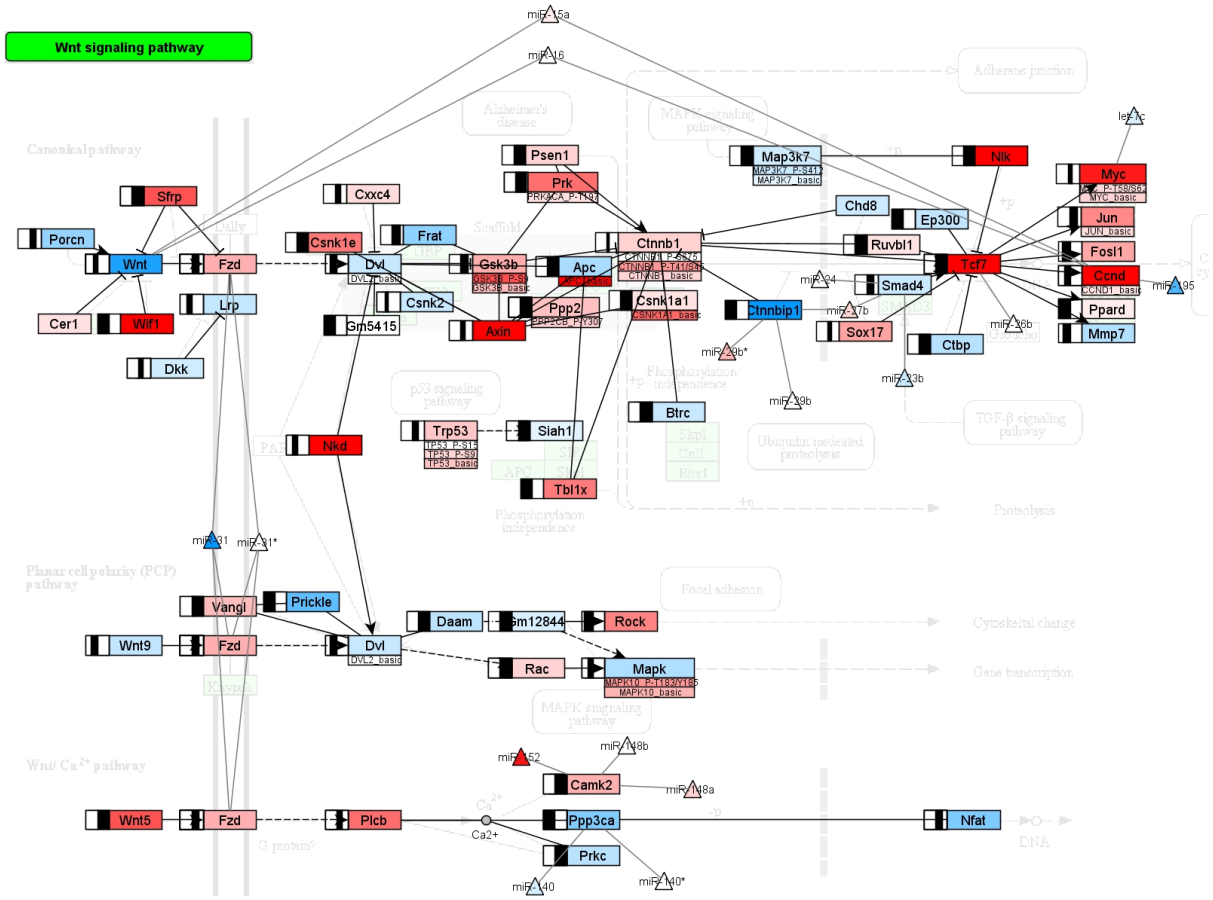


Figure 10. Integrated visualization of DNA methylation, protein modification, microRNA and messenger RNA changes in the WNT signaling pathway of *Ctnnb1* mutated tumors. Characteristic perturbations in Ctnnb1 mutated tumors across multiple layers of gene-regulation are depicted in this picture. In general, red means up-regulation and blue means down-regulation. More saturated colors indicate stronger differential regulation. Messenger RNA expression is shown directly as node color, microRNAs are connected to their mRNA targets and colored according to microRNA expression and protein modification data is shown in small separate boxes below the actual nodes. DNA methylation is indicated with a black bar in a black surrounded box, ranging from the middle to the left to indicate hypomethylation and to the right to indicate hypermethylation.