**Point-to-point list**

We have carefully read all reviewer comments and addressed each single point appropriately. We are convinced that the reviewer comments substantially improved the quality of the manuscript and the application. However, we still would like to keep the focus of the manuscript on the joint visualization of cross-platform microarray data. The InCroMAP application itself provides many more features, but discussing all of them would change the focus and scope of the manuscript.

The changes in the manuscript are highlighted in red. Besides them, we have changed the FAQ and the documentation of InCroMAP (e.g., by adding a section about the import of data from GEO, using GEO2R). We further added a screencast (short video tutorial) on the homepage that explains how to load data and perform a first analysis with InCroMAP. Changes on the application are not directly visible. Therefore, we created a small list that summarizes the major changes:

* Edges are now running above nodes (can be changed in the preferences).
* The fill color of DNA methylation bars was changed to grey, in order to make edges running above them visible.
* All tabs are movable (drag’n’drop) and closeable.
* It is possible to import enrichment results from other applications, such as GSEA or DAVID.
* InCroMAP can now read probeset identifiers from different microarray platforms (supported manufacturers: Affymetrix, Agilent, Illumina).
* Enhanced automatic detection of input file content (simplified the import of experimental data).

In the following, we would like to discuss all reviewer comments in detail. Therefore, we colored the reviewer comments in red and our answers/actions in black below.

Reviewer: 1   
Comments to the Author   
General comments:   
A given gene's mRNA can be up- or down-regulated under a plethora of conditions, and sample characteristics in experiment designs will reflect that variability. What is not clear to me from the paper how for gene X associated with pathway Y the authors decide that the expression data for gene X from a given experiment is relevant to pathway Y. Is it simply through the X-Y association? Or is it that if pathway Y is activated in tumor condition, experiments are chosen that compare normal to tumor? A brief description of the method used to achieve the above would help.

The available expression data and number of conditions therein is determined by the input file of the user. Available pathways are defined in the KEGG database. The association between a pathway and genes from the input datasets is made by the application, based on the intersection of genes contained in the pathway and the dataset.

Therefore, the user chooses the pathway of interest out of a list of available pathways together with one observation (i.e., experimental condition) from the input dataset. The pathway is then visualized and modified according to the selected input dataset and observation.

To help the user picking a pathway from the list, the application provides a pathway enrichment feature on any input data type. This will rank the list of all available pathways, based on expression changes observed for a particular experimental condition.

|  |  |
| --- | --- |
| Untitled-2.png |  |
| Figure 1: Example of association between pathway and input dataset. Left picture shows a datasets with two observations (called ‘Ctnnb1’ and ‘Ras’). The user can choose one of those observations and a pathway to visualize. The right picture shows a pathway enrichment that ranks all available pathways, based on their importance for the selected observation (here: ‘Ctnnb1’). | |

We have added a new sentence to the beginning of the methods section to explain how relevant pathways are identified.

The term 'differential expression' is also mentioned in the paper - what are we comparing to what in the context of the pathway? Normal versus disease, activated versus passive pathway, treatment versus no treatment?

The term ‘differential expression’ is indeed ambiguous and just states that some gene has a different expression value under any two conditions. Depending on the biological context of the analyzed dataset this can be, for example, ‘normal vs. disease’ or ‘tumor vs. non-tumor’ or ‘any treatment vs. any control’.

In the end, it is not important for the visualization technique if the input data was a comparison of different tumors or treatments. It works with a comparison of any two expression values, independent if the sample was a tumor sample, disease or whatever.

Figure 1 (left part) shows an input dataset we used for generating the example figures. Here, we had two observations called ‘Ctnnb1’ and ‘Ras’, which were calculated from a comparison between tumor and healthy tissue. This is important for the interpretation of the figures by the user, but for the visualization procedure itself, only the value is important.

In section 2.3 we have 'Each of these boxes is then labelled according to the respective protein form and colored based on the underlying expression data, as described previously for mRNA datasets'. Given that   
- proteomics experimental data is mostly qualitative (signal = protein present; no signal = protein may or may not be there), and   
- different protein modifications are mapped to the same gene   
how is quantiative expression data for different protein forms obtained?

Quantitative expression data for different protein modifications can be obtained by reverse-phase protein arrays, which utilize specific antibodies for protein detection. The relevant references are cited in the same paragraph (Pirnia et al., 2009; Yates et al., 2009). For general quantitative mass spectrometry, several other approaches exist (see, e.g., “Ong, S.-E. & Mann, M. *Mass spectrometry-based proteomics turns quantitative.* Nat Chem Biol 1, 252-262 (2005)”).

However, also binary MS data can be visualized. Imagine a table with each row corresponding to a protein and one column with a ‘1’ if a protein is expressed and ‘0’ if not. As described in the publication, the result would be a small box below each gene, corresponding to the protein. This box would be white for all zeros (no signal) and red for all ones (signal).

Reviewer: 2   
Comments to the Author   
The methods and tools described in the paper « Pathway-based visualization of cross-platform microarray datasets » are dedicated to the analysis and visualization of different kind of expression/functional data at the pathway level. For that, 4 types of data (mRNA and miRNA expression data, protein expression data and methylation data) can be mapped on Kegg pathways, and all types of data can be visualized simultaneously. The java tool first allows the user to identify Kegg pathways (and also GO annotations and MSig DB annotations) associated with the dataset of interest through an enrichment analysis. In a second step, expression and methylation data can be mapped on Kegg pathways, and finally, these different data can be integrated together.   
The paper is clear and well-written, and the tool is functioning correctly. Such integration of functional data is missing in the field.

However, a step forward in the development of the tool would improve a lot its significance for biologists and bioinformaticians. The main pitfalls are indeed that providing the tabular files for the 4 types of data in the correct formats is not easy. This could be overcome by 1) A better integration with existing tools dedicated to the analysis of expression data and 2) An extension of the gene identifiers recognized.

We completely agree with both points. The drawback is that there is no standard file format for processed data, especially not across multiple array platforms. However, it is very common to produce a single table (e.g., in excel) containing the fold-changes for each gene and experimental condition. This said, the common denominator of most stored processed array data are tabular files. However, the provided metadata columns (e.g., identifiers) depend on the employed platform and software.

Therefore, we have put a huge effort in a flexible and easy import of tabular data files. For example, the column separator char (mostly a tab or comma) or if the file contains headers is detected automatically. Even some of the content is inferred automatically: columns that contain identifiers with unique regular expressions are automatically assigned (for example, Ensembl or RefSeq identifiers).

Currently, InCroMAP supports the most popular database identifiers for genes: EntrezGene, Ensembl, NCBI Reference Sequence (RefSeq), KEGG Genes and Gene Symbols.

Nevertheless, we completely agree that the import of files should be as easy as possible and InCroMAP should support many identifier databases. Thus, to ease the import of non-annotated microarray data, we further added probe identifiers from Affymetrix, Agilent and Illumina in the latest release. Furthermore, we added a section to the FAQ on the homepage and in the documentation that describes how to import data by using the GEO2R tool directly from GEO. In addition to this, our group is currently working on a data preprocessing webservice for microarray data from diverse platforms. This pipeline will support probe identifiers and will have a special interface for InCroMAP.

From the statistics point of view, an improvement of the methods to 1) Perform GSEA like analysis to identify pathway associated with expression data instead of a simple enrichment test, and 2) The possibility to adjust the background while doing the enrichment, would also be valuable. These improvements would be of big interest to the community as the tool could provide a free public alternative to Ingenuity. But it has to be really user-friendly and connected to common tools for that.

We agree that it would be desirable to perform GSEA enrichments directly in InCroMAP and offer options to change various parameters, like the background of an overrepresentation test. But the implementation of these features would go beyond the scope of this publication. The central focus lies on the integration of data from different platforms and their visualization. The enrichment feature of InCroMAP should provide a starting point of important pathways to visualize.

However, we offer users the possibility to use other tools (for instance, the official GSEA tool or DAVID) for gene set enrichments and process the results in InCroMAP. Therefore, we have added a novel import function that is able to read the enrichment results from GSEA, DAVID and some other applications.

Comments on paper/method Title.   
As the tool also provide an enrichment analysis, it can be mentioned in the title. The tool also permit to add proteomics protein expression by MS, which is not microarray based. Hence, the title may be not really adapted.

We decided to focus the publication on the joint visualization of data from multiple platforms in the context of higher-order pathways. Enrichment analysis is only briefly mentioned in the manuscript; therefore we think that it should not be in the title.

You are correct that the manuscript also applies for correctly formatted non-microarray datasets. Especially on the protein level, expression is commonly measured by other experimental techniques (such as mass spectrometry). However, mRNA expression, microRNA expression and DNA methylation changes are typically detected using microarrays. Further, we developed and tested the application based on microarray datasets. Therefore we decided to name the publication “[…] visualization of cross-platform microarray datasets”, even though the described methodology could be applied to datasets from other high-throughput techniques.

Introduction.   
In the same way, the first part of the introduction focuses on microarray platform developments. Focusing first on the shift from gene/protein centered large-scale analyses to pathway-based analyses, as it's developed in the second part of the introduction would be more clear. For instance, it could be said that most mRNA expression analyses were previously done by first extracting up and down regulated genes and then computing enrichments on these genes, whereas nowadays analyses mostly follow GSEA likes methods based on differentially expressed pathways or molecular signatures.

In the revised manuscript, we now mention that in current microarray studies differential expression is not only inspected on the level of individual genes, but also in the context of pathways and molecular signatures. We also describe the basic concepts of enrichment analysis methods used for the functional annotation of differentially expressed genes. Thank you for proposing these changes.

The introduction should define more precisely what the authors mean by “visualization” of large-scale datasets. Expression changed on pathway and methylation changes on genomes are 2 visualizations, but with different goals. Here the visualization seem to be more a pathway analysis and visualization.

In the revised manuscript we more clearly state that both region- and network-based visualization techniques may provide two complementary means to gain new biological insights. Furthermore, we more clearly characterize the visualization approach proposed in the manuscript.  
  
Methods.  
The step of defining correct tabular files for the tool is too complicated. Indeed, it would be of great importance to be able to use other kind of identifiers than Ensembl/Entrez, especially while loading microarray datasets (e.g affy probes...). As the aim of the tool is not to do the workflows analyses from raw microarray dataset, one solution could be an better integration with existing tools. For instance, searching for datasets in GEO <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20437>   
it's possible to use the GEO2R tool for the analysis. The results could be exported directly in InCroMAP. This is currently not possible because of different identifiers. Other microarray databases probably also provide these analysis tools. R/Bioconductor packages permit to do this changes.

Microarray datasets are often provided as excel spreadsheets with different identifiers and columns with expression data. For example, gene symbols are very often contained in these files. InCroMAP can perfectly read and import these excel spreadsheets, if the user selects CSV or “Text (tab separated)” from excels save-as dialog. This makes it very easy to prepare tabular files for the tool.

We agree with you that the support of platform-specific identifiers would further ease the import of microarray data. Therefore, we now added a support for Affymetrix, Agilent and Illumina probe identifiers in the latest release.

Thank you for mentioning GEO2R. We were not yet aware of this tool, which can be perfectly used in conjunction with InCroMAP. The results of GEO2R are tabular files (always with gene symbols) exactly as InCroMAP requires them. We have added a section to the InCroMAP FAQ (on the homepage and in the documentation) that explains how to import data from GEO into InCroMAP, using GEO2R.

2.1 Pathway visualization   
Is it possible to change the transparency of the background Kegg picture? 70% is a bit too much and it's hard to see the information.

Yes it is. Please open the Preferences dialog by selecting “Edit” and “Preferences” from the menu bar of InCroMAP. From the preferences dialog, please select the “Graph options” tab and there is an option called “Brighten KEGG background image” that accepts any percentage between 0 and 100.

2.2 Visualization of messenger RNA   
Is it possible for the user to chose between the different way of summarizing info for a node corresponding to multiple genes?

Yes it is. The first time mRNA data is visualized, a dialog with various visualization options is displayed. For convenience, the decisions are stored and used as default in future sessions.

However, the preferences dialog that can be opened with “Edit” and “Preferences” provides an option called “gene center signals by” that allows to change the way of summarizing information for a node to either “Mean”, “Median”, “Maximum, “Minimum”, “Maximum distance to zero” (i.e., maximum absolute value). Furthermore, by deselecting the “Remember gene center decision” checkbox, it is possible to force the application to ask each time before visualization. More information can be found in the InCroMAP user’s guide.

Here I don't understand why the authors chose to summarize information for multiple genes on one node, whereas for protein expression, the details are provided below. I could be nice to be able to see the details for each gene associated to the node, each of which being colored according to its expression (in the same way as for protein expression visualization), while clicking on the node for instance.

Summarizing multiple values is always problematic, because it leads to a loss of information. On the other hand, showing more information leads to confusing graphs and reduces the clarity.

Therefore, we decided to keep some information and display more details in the tooltip of a node (simply mouse-over a node and it will appear). Further, we created a detail panel that appears whenever a node is clicked. Parts of it are shown in Figure 3 in the manuscript. This panel presents all information about the selected node, the corresponding genes and their associated expression values, as well as some plots of them (e.g., DNA methylation regions).

We still decided to include the protein modification data in separate boxes, because these are measurements for different protein phospho-forms. The expression of an unmodified and a phosphorylated protein should not be summarized. In contrast to summarized genes, they clearly differ in their biological function.

2.3 Visualization of methylation   
Overall, the different way to represent the different expression/methylation data are well chosen. A way to clarify the methylation data (the rectangle are not often clear when hypo/hyper methylations are small) could be to put 2 triangles linked by their peak, one for hypomethylation, one for hyper.

Excellent idea, this would provide information about the highest and the lowest peak at the same time. Unfortunately, our graph-drawing library is limited and realizing this kind of visualization is not supported. However, we really like the idea and will try to include it in a future release of InCroMAP.

Figure 2. The big picture is hard to understand. It would have been nice to have an example case, for instance showing overexpression associated with hypermethylation, and discussed with hypothesis derived from the InCroMAP analysis.

We provided this figure to give readers an impression how a whole pathway with expression data from four different platforms looks like. In our opinion, it is important to provide at least one exemplary result of the methodology for cross-platform data visualization, described in the manuscript.

An example for hypermethylation, leading to a suppression of mRNA is provided in Figure 3 (high DNA methylation in peak in the promoter and down-regulation of the corresponding mRNA).

We think that discussing further results from this dataset in more detail goes beyond the scope of this methodological paper. We are not intending to publish the dataset (consisting of mRNA, DNA methylation, microRNA and protein microarrays for many samples) or results from it.

Discussion can be shorten a bit.  
The enrichment method used in the tool is not described in the paper. I assumed it's a classical Fisher test. However, the setting of the background against which the dataset of interest is compared is of fundamental importance in these analyses. The current version of the tool doesn't permit to change the background (full genome ?). As adding the ability to compute the enrichment analysis on different background, or asking the user to provide a background gene/protein list can be very complicated, one idea would be for the tool to be able to load enrichment results of Kegg pathways from the DAVID database   
<http://david.abcc.ncifcrf.gov/>   
that is dedicated to enrichment analyses and provide many different backgrounds, for instance for Affy microarrays.   
Furthermore, fisher test enrichment analyses that imply to select a threshold in expression microarrays, and consider only genes above/under this threshold, are less used nowadays. People are using more and more GSEA-like statistics. Maybe the tool could also be able to load GSEA result files.

The paper focuses on how to jointly visualize data from mRNA, microRNA, DNA methylation and protein arrays. The tool itself provides many more features, but they should not be part of this manuscript.

However, InCroMAP currently performs enrichments, using a hypergeometric test. Since we calculate the p-values for defined gene groups (i.e., pathways) in the KEGG database, the background is the total number of genes in the KEGG PATHWAY database.

We like your idea of providing import functionalities for DAVID or GSEA. To this end, we added a new menu item in `File´ -> `Import enrichment result´ that allows users to use any external application for the enrichment if they need more features than InCroMAP provides.

Comments on Java tool   
  
General comments   
Overall, the tool is functioning correctly, no problem to launch it neither to load datasets. However, it could be improved to be more user-friendly. It's hard to understand where to click to do a simple analysis without ready the documentary file. And users that just want to have a quick look at their data to know if they continue with the tool may not spend time on documentation.

We have added several ways to import datasets in InCroMAP. Drag’n´drop is an option, selecting `File´ and `open´ or using the toolbar are already three different starting points. Then, most properties of the file format and content are detected automatically, which really eases the import of tabular files, in contrast to other tools where some fixed file format must be obeyed.

Furthermore, the toolbar on top of the tabs is changing dynamically, depending on the currently selected tab to always give users hints what type of analysis they could do next. Additionally, all dialogs and other boxes are automatically provided with reasonable default selections.

Concluding, we spent a huge amount of time in the user-friendliness of the tool, but we are always happy for further suggestions. It is our aim to improve the usability and make the tool as easy and flexible as possible.

Therefore, we have now added a screencast on the homepage, which quickly tells users how to prepare and open files with InCroMAP and run a first analysis on these files.

In this context, example files available directly on the tool would be of great help, to see the formats to upload and how the tool is working and displaying results in a few minutes.

There are example files for each of the four input formats at the bottom of the InCroMAP download page.

Further, the documentation has a “How to get started” section that explains in detail all required file formats and how to get first results from InCroMAP. There are further examples for single and integrated analysis in the documentation.

Additionally, we now created a screencast (i.e., video tutorial) that explains shortly how to open datasets and perform a first analysis with InCroMAP.

It could be nice to be able to search in the pathway list.

It is possible to simply start typing and the list will jump to the typed pathway (e.g., try typing `MAPK` and it jumps to the MAP Kinase pathway).

The list of genes/proteins used for the enrichment could be of interest for the users.

Genes that are causing pathways to show up in the result of an enrichment are mentioned in the last column of the enrichment table.

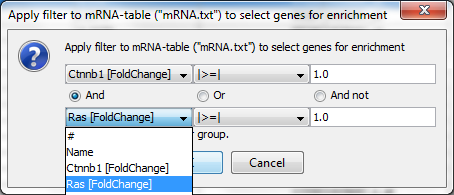
Genes from a dataset that are above a certain threshold (i.e., fold change cutoff) can easily be identified by sorting the dataset according to the fold change (by clicking on the header).

It can be nice also to be able to combine more than 2 observation for the enrichment.

This functionality is also provided.

We invented a procedure called `integrated enrichment´ that essentially performs an enrichment analysis, based on a cross-platform dataset of mRNA, miRNA, DNA methylation and protein data. Therefore, it combines the observations from those datasets. More information is provided, e.g., in the documentation.

If one wants to combine two observations from the same dataset, the InCroMAP filter dialog can be configured as in the following example:



This will perform an enrichment of all genes that have an absolute fold change in “Ctnnb1” of >= 1.0 and an absolute fold change in “Ras” of >= 1.0 (“Ctnnb1” and “Ras” are different observations in a single dataset).

(-1) is not clear, explicitly say that the gene is not found.

Do you mean the `-1´ that is shown in the table if no gene id could be found when reading a file? We changed this now to `Not found´.

Would need a way to close tabs, to try another enrichment for instance. Difficult to navigate among the many results.

By selecting “File => Close” from the menu bar, the current tab is closed. We further added a right mouse click menu to the tabbed pane. Simply click the current tab heading with the right mouse and select `close current tab´.

We further put some effort to extend the default Java tabbed pane capabilities to make the tabs movable. You can drag and drop all tabs to change their order.

I did not manage to use the function “color pathway according to enrichment”, and did not understand what it is supposed to represent (all the pathway is enriched in the dataset, no individual nodes?)

This feature is made mainly for pathways that contain many nodes, representing other pathways. For example the `Metabolic pathways´ map, which is some kind of overview pathway. Visualizing this pathway and having performed an enrichment in InCroMAP allows you to use this feature and change the color of the referenced pathways, according to their significance in the enrichment.

An example is given in the following pictures:

|  |  |
| --- | --- |
|  |  |

On the left side, the unmodified pathway is shown. On the right, some pathway-reference nodes have changed their color according to their p-value in an enrichment. A more saturated color represents a lower p-value.

However, please note that the manuscript focuses on how to integratively visualize cross-platform datasets in pathways. InCroMAP contains many more features, but describing all of them is beyond the scope of the manuscript.

Reviewer: 3   
Comments to the Author   
This manuscript describes a pathway visualization method that can   
combine KEGG pathways with experimental data including gene   
expression, protein modifications, DNA methylation, and miRNA   
expression and regulation. These are all important regulatory   
signals. A tool that enables visual assessment of all of this data in   
a pathway context would be a valuable contribution to the field. The   
system described appears to summarize regulatory signals effectively   
so that the final visualization is informative but not too complex.   
With that said, some parts of the manuscript are unclear, and some   
portions of the system's functionality merit concern.   
  
Major concerns:   
  
1. It appears that edges are drawn from the center of each pathway   
node, and when the node is expanded with protein modification states,   
these interactions can become badly obscured. For example, Figures   
1a-c show an interaction between Ntrk and Grb2, but in Figure 1d-f   
this interaction is obscured by Egfr's protein modification data, and   
would be easu to miss. Even Figure 1a shows an interaction between   
Rasgrf and Kras that is obscured by Rasgrp. Since Rasgrf and Rasgrp   
both have the same type of interaction with Kras, the two arrows could   
be mistaken for one. It would be better if the edges were rendered on   
top of nodes (perhaps with a fainter line type) than obscured by them.   
The authors might have decided against this for fear of obscuring the   
node names, but for accurate pathway representation, it is arguably   
more important to show the topology accurately than to show the names   
clearly.

Many thanks for this comment. We completely agree that in some cases, the interactions are getting obscured by edges, running below nodes.

To resolve this issue, edges are now drawn on top of nodes, by default. We further added an option to the preferences that allows users to change this behavior. Additionally, we have added an option that enables an edge-routing algorithm which should avoid some edge-node collisions.

The figures in the manuscript have been changed to reflect those changes.

2. Figure 1e shows that interaction coordinates that end at the left   
side of pathway nodes are obscured when DNA methylation boxes are   
added to the left side of the nodes. This results in losing   
information on the type of interaction. When methylation data is   
shown, the edges should end at the left side of the methylation box,   
not underneath it.

We completely agree. Unfortunately, our graph drawing library is limited and does not support changing the end point of edges. However, we agree that this issue must be resolved. Therefore, we have changed the fill color of the methylation boxes to grey and put the edges on top of nodes. These changes resolve this issue and make the edge endings visible.

The renewed figures in the manuscript now also reflect those changes.

Moderate concerns:   
  
1. In Figure 1, we see that the system adjusted the graphical coordinates of Egfr to include the protein modification data, and there may be other cases where the system modifies graphical coordinates. However, there is no discussion of the limitations of this functionaltiy. For example, is there a limit to the number of protein modification states that can be included for a given node, or in a given section of the pathway?

The layout is an important topic when drawing pathways. We have had many discussions about how to best layout graphs. In the end, no layout is as good as a hand-drawn layout. And because all KEGG pathways are layouted manually, we decided to keep this layout. This also enables us to draw the original KEGG picture in the background and overlay it with the interactive pathway nodes.

If novel nodes are added to the graph (e.g., microRNAs), we use an organic layout algorithm to place the novel nodes in the existing graph.

Our representation of DNA methylation and protein modification data may lead in some cases to an overlap of this information with other nodes. If desired, overlaps can be avoided by manual adjustment of the node positions. However, we agree that an ideal solution keeps the original layout and tries to perform automatically a minimum amount of modifications to avoid overlaps between all nodes and annotations.

Therefore, we have added a new option to the latest release of InCroMAP that is called `Relayout nodes after size change´. This will automatically adjust the layout of nodes with protein or DNA methylation data. However, the default is set to false because we recommend users to fix the layout by hand (the graph is interactive) in order to obtain optimal results.

Our protein modification dataset had a maximum of 4 to 5 modifications per protein. The application itself puts no limitations on the number of modifications to visualize. However, the user can always manually change the graph or activate the automatic re-layout option to avoid collisions with other nodes.

2. Figure 2 shows groups of nodes, but this functionality is not discussed in the text.

The pathways are translations from KGML formatted xml files from the KEGG database. This KGML format includes group nodes, which are either complexes (mostly protein complexes) or gene/protein families.

We agree that the manuscript should mention all node types and added some more explanations in section 2.1.

3. When DNA methylation data is shown, and when one pathway node is represented by multiple genes, how is a node methylation level determined from the various gene methylation levels? This is not discussed.

By default, each gene’s methylation level is determined by the probe with absolute maximum fold change and the gene with the maximum differential methylation observed among the respective genes determines the level of the node. In other words, by default, the highest peak is displayed.

We think that this is a good default, because the intention is to give researchers a hint if methylation changes are present at all. The application will give many details (including region plots for all genes) if the corresponding node is selected.

We have changed section 2.4 of the manuscript to now explain this matter.  
  
Minor concerns:   
  
1. In the first paragraph of Section 2.1, the text reads: "Some basic errors are corrected automatically and appropriate shapes, colors and labels are inferred". More detail is needed.

KEGG has some peculiarities. The organism-specific pathway maps are derived from reference pathways and this leads sometimes to nodes, representing genes that are not present in the current organism. These nodes are removed. Further, the name that is given in the XML file from KEGG is mostly not the best choice.

However, these things are quite technical and they are discussed in the referenced KEGGtranslator publication. We removed the mentioned sentence and replaced it by an explanation of the different node shapes. This will provide more information to readers than going into the technical peculiarities of the KEGG database. We still include the KEGGtranslator citation for more information on the details.

2. The second paragraph in Figure 2.2 describes ways in which expression data from multiple genes can be combined for coloring one pathway node, in cases where one node on the KEGG pathway represents multiple genes in the genome. The text describes multiple options, and while it suggests that they are options within the method, it does not state so clearly.

We have added a new sentence that now states clearly that these are options within the application.  
  
3. In Section 2.2, the sentence beginning "For this purpose, we propose to map the negative logarithm of the p-values to a color gradient" is unclear. If this functionality is available, then "propose" is the wrong verb. If it's not available, it should not be mentioned.

It is available and we have changed `propose´ to `decided´.  
  
4. Protein modification data is shown for some but not all genes. What determines this, the contents of the protein dataset?

Yes. Protein modifications were measured with reverse-phase protein arrays. Since each protein modification requires a separate specific antibody, the corresponding datasets are mostly limited to 100-200 proteins.

The relevant references for reverse-phase protein arrays are cited in the manuscript (Pirnia et al., 2009; Yates et al., 2009).  
  
5. Page 4 Lines 14-38 describe methods to derive gene methylation signals from individual methylation probes, but it is not clear what functionality is actually available in this system.

We have changed the section to more clearly reflect what functionality is presently available in InCroMAP.