The vast amount of available experimental data generating annotated motifs (Milo et al., 2002), global network structure (Albert and Barabasi, 2002), network clustering (Holme et al., 2005) and network centralities (Wuchty and Stadler, 2003). Particularly, centralities are node parameters that can identify nodes having a relevant position in the overall network architecture. Cytoscape (Cline et al., 2007; Shannon et al., 2003) is an excellent visualization and analysis tool with the analysis features greatly enhanced by plug-ins. Plug-ins such as NetworkAnalyzer (Assenov et al., 2008) computes some node centralities but does not allow direct integration with experimental data. Applications such as VisANT (Hu et al., 2005) and Centibin (Junker et al., 2006) calculate centralities, although they either calculate fewer centralities or are not suitable to integration with experimental data (see Supplementary Table S1 for a comparative evaluation). CentiScaPe is the only Cytoscape plug-in that computes several centralities at once. In CentiScaPe, computed centralities can be easily correlated between each other or with biological parameters derived from the experiments in order to identify the most significant nodes according to both topological and biological properties. Functional to this capability is the scatter plot by value options, which allows easy correlating node centrality values to experimental data defined by the user.

2 SYSTEM OVERVIEW

CentiScaPe computes several network centralities only for undirected networks. Computed parameters are: Average Distance, Diameter, Degree, Stress, Betweenness, Radiality, Closeness, Centroid Value and Eccentricity. Plug-in help and online files are provided with definition, description, biological significance and computational complexity for each centrality (Supplementary Tables S2 and S3, Centralities Tutorial). Min, max and mean values are given for each computed centrality. Multiple networks analysis is also supported. Centrality values appear in the Cytoscape attributes panel. Centralities can be deactivated. The user can select the more/equal option for some centralities, the less/equal option for others and can join them with AND-OR operators. This feature can immediately answer to questions as: ‘Which are the nodes having high Betweenness and Stress but low Eccentricity?’ Notably, the threshold can also be modified by hand to gain in resolution. Once the nodes have been

© The Author(s) 2009. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5/uk) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
selected according to their node-specific values, the corresponding subgraph can be extracted and displayed using normal Cytoscape core features. Two kind of graphical outputs are supported: plot by centrality and plot by node, both allowing analysis that are not possible with other centralities tools. The user can correlate centralities between them or with experimental data, such as, for example, gene expression level or protein phosphorylation level (plot by centralities), and can analyze all centralities values node by node (plot by node) (Fig. 1). Graphics can be saved to a jpeg file.

The plot by centrality visualization is an easy and convenient way to discriminate nodes and/or group of nodes that are most relevant according to a combination of two selected parameters. It shows correlation between centralities and/or other quantitative node attributes, such as experimental data from genomic and/or proteomic analysis. The result of the plot by centrality option is a chart where each individual node, represented by a geometrical shape, is mapped to a Cartesian axis. In the horizontal and vertical axis, the values of the selected attributes are reported. Most of the relevant nodes are easily identified in the top-right quadrant of the chart. Figure 1 (Supplementary Fig. S1) shows a plot of centroid values over intensity of protein tyrosine phosphorylation in Tyrosine. Proteins with high centroid value and high level of phosphorylation are easily identified in the top-right quadrant of the graph. Pointing the mouse over the geometrical shapes in the plot shows the corresponding node ID and attribute values (see Section 3 and Supplementary Materials).

3 A REAL WORLD EXAMPLE: CENTRALITIES IN THE HUMAN KINO-PHOSPHATOME

We tested CenSiCaPe on the human kino-phosphatome. A global human protein interactome dataset, including 11 120 nodes and 84 776 unique undirected interactions (IDs = HGNC), was compiled from public databases (HPRD, BIND, DIP, IntAct, MINT, BioGRID; Supplementary file GLOBAL-HGNC.stf). A subset of this network was extracted consisting of only known interactions between human protein kinases and phosphatases. The resulting sub-network, a kino-phosphatome network, consisted of 549 nodes and 3844 unique interactions (Supplementary Table S4 and Kino-Phosphatome.stf), with 406 kinases and 143 phosphatases. The kino-phosphatome network did not contain isolated nodes. We used CenSiCaPe to calculate centrality parameters. A first general overview of the global topological properties of the kino-phosphatome network comes from the min, max and average values of all computed centralities along with the diameter and the average distance of the network (Supplementary Table S5). For instance, an average degree =13.5 with an average distance of 3 may suggest a highly connected network, in which proteins are strongly functionally interconnected.

Computation of network centralities allowed a first ranking of human kinases and phosphatases according to their central role in the network (Supplementary Table S6, reporting node-by-node values of different centralities). To facilitate the identification of nodes with the highest scores we applied the ‘plot by centrality’ feature of CenSiCaPe. Plotting degree over degree (Supplementary Fig. S3) shows that the distribution is not uniform, with the majority of nodes having a similar low degree and very few having very high degree. This is consistent with the known scale-free architecture of biological networks (Jeong et al., 2000). The scale-free topology of the kino-phosphatome network was also confirmed with Network Analyzer (Assenov et al., 2008). A total of 186 nodes (164 kinases and 22 phosphatases) displayed a degree over the average. The top-10 degree values (64–102) were all kinases, with MAPK1 showing the highest degree (102). Notably, MAPK1 displayed the highest score for all computed centralities (Fig. 1), suggesting its central regulatory role in the kino-phosphatome. In contrast, PTPN1 had the highest degree, 46, between all phosphatases (top 31 among all nodes) and had a rather high score also for other centralities (Supplementary Fig. S4). Thus, degree analysis suggests that MAPK1 and PTPN1 are the most central kinase and phosphatase, respectively.
A total of 242 nodes (206 kinases and 36 phosphatases) displayed that tyrosine phosphorylation induced in PMNs by chemoattractants in tyrosine showing a higher centrality values. This could suggest topological position in the network with proteins phosphorylated in threonine (Supplementary Fig. S8) or have been analyzed. The plots allow immediately evidencing that tyrosine residues induced by fMLP triggering in human PMNs phosphorylation induced in PMNs by fMLP. Plot of centroid computed centrality and of experimental data regarding protein values were plotted over values of protein phosphorylation. Here, PhosphoTyr.NA, PMN-PhosphoThr.NA). Experimental data were available in Supplementary Material: PMN-PhosphoSer.NA, PMN-protein phosphorylation was evaluated by using the Kinexus 1 min with the classical chemoattractant fMLP (100 nM) and (Supplementary file Phosphorylation-Experiment for description). To further support this suggestion we analyzed also the centroid. Plotting centroid over centroid provided a linear distribution and as for the degree, also here the distribution was not uniform (Supplementary Figs S5 and S3). Average centroid was -393. A total of 242 nodes (206 kinases and 36 phosphatases) displayed a centroid over the average. The top-10 centroid values (-79 to -18) were all kinases, with MAPK1 showing the highest centroid value (18). PTPPN1 had the highest centroid value, -154, between all phosphatases (top 22 among all nodes). Thus, as for the degree, also the centroid value analysis suggests a scale-free distribution, with MAPK1 and PTPPN1 being the most central kinase and phosphatase, respectively. This conclusion was also easily evidenced by plotting the degree over the centroid (Supplementary Fig. S6). From the analysis, a non-linear distribution of nodes is evident, with few dispersed nodes occupying the top-right quadrant of the plot (i.e. high degree and high centroid): these nodes can potentially represent particularly important regulatory kinases and phosphatases.

This kind of analysis can be iterated by evaluating all other centralities. To extract the most relevant nodes, we used CenTiSCaPe to select all nodes having all centralities values over the average. Upon filtering, we obtained a kino-phosphatome sub-network consisting of 97 nodes (82 kinases and 15 phosphatases) and 962 interactions (Supplementary Fig. S7, Table S7 and K-F sub-network.sif). This sub-network possibly represents a group of highly interacting kinases and phosphatases displaying a critical role in the regulation of protein phosphorylation in human cells. Further analysis with CenTiSCaPe or other analysis tools, such as MCODE (Bader and Hogue, 2003) or network analyzer, performing a Gene Ontology categorization of nodes in large complex networks derived from experimental data.

4 CONCLUSIONS
CenTiSCaPe is a versatile and user-friendly bioinformatic tool to integrate centrality-based network analysis with experimental data. CenTiSCaPe is completely integrated into Cytoscape and the possibility of treating centralities as normal attributes permits to enrich the analysis with the Cytoscape core features and with other Cytoscape plug-ins. The analysis obtained with the Boolean-based result panel, the ‘plot by node’ and the ‘plot by centrality’ options give meaningful results not accessible to other tools and allow easy further, more focused, experimental verifications. Combination of CenTiSCaPe with other bioinformatics tools may help to analyze high-throughput genomic and/or proteomic experimental data and may facilitate the decision process.

REFERENCES

To further support this suggestion we analyzed also the centroid. Plotting centroid over centroid provided a linear distribution and as for the degree, also here the distribution was not uniform (Supplementary Figs S5 and S3). Average centroid was -393. A total of 242 nodes (206 kinases and 36 phosphatases) displayed a centroid over the average. The top-10 centroid values (-79 to -18) were all kinases, with MAPK1 showing the highest centroid value (18). PTPPN1 had the highest centroid value, -154, between all phosphatases (top 22 among all nodes). Thus, as for the degree, also the centroid value analysis suggests a scale-free distribution, with MAPK1 and PTPPN1 being the most central kinase and phosphatase, respectively. This conclusion was also easily evidenced by plotting the degree over the centroid (Supplementary Fig. S6). From the analysis, a non-linear distribution of nodes is evident, with few dispersed nodes occupying the top-right quadrant of the plot (i.e. high degree and high centroid): these nodes can potentially represent particularly important regulatory kinases and phosphatases.

This kind of analysis can be iterated by evaluating all other centralities. To extract the most relevant nodes, we used CenTiSCaPe to select all nodes having all centralities values over the average. Upon filtering, we obtained a kino-phosphatome sub-network consisting of 97 nodes (82 kinases and 15 phosphatases) and 962 interactions (Supplementary Fig. S7, Table S7 and K-F sub-network.sif). This sub-network possibly represents a group of highly interacting kinases and phosphatases displaying a critical role in the regulation of protein phosphorylation in human cells. Further analysis with CenTiSCaPe or other analysis tools, such as MCODE (Bader and Hogue, 2003) or network analyzer, performing a Gene Ontology categorization of nodes in large complex networks derived from experimental data.

4 CONCLUSIONS
CenTiSCaPe is a versatile and user-friendly bioinformatic tool to integrate centrality-based network analysis with experimental data. CenTiSCaPe is completely integrated into Cytoscape and the possibility of treating centralities as normal attributes permits to enrich the analysis with the Cytoscape core features and with other Cytoscape plug-ins. The analysis obtained with the Boolean-based result panel, the ‘plot by node’ and the ‘plot by centrality’ options give meaningful results not accessible to other tools and allow easy further, more focused, experimental verifications. Combination of CenTiSCaPe with other bioinformatics tools may help to analyze high-throughput genomic and/or proteomic experimental data and may facilitate the decision process.

REFERENCES