# An Interactive Macrophage Signal Transduction Map Facilitates Comparative Analyses of High-Throughput Data [Extended Methods section]

Pia Wentker*, Martin Eberhardt*, Florian S. Dreyer*, Wilhelm Bertrams[†], Martina Cantone*, Kathrin Griss[†‡], Bernd Schmeck[†§], and Julio Vera*[2]

* Laboratory of Systems Tumor Immunology, Department of Dermatology, Friedrich-Alexander University Erlangen-Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany

[†] Institute for Lung Research / iLung, German Centre for Lung Research, Universities of Giessen and Marburg Lung Centre, Philipps-University Marburg, Marburg, Germany

[‡] Department of Internal Medicine/Infectious Diseases and Pulmonary Medicine, Charité University Medicine Berlin, Berlin, Germany

[§] Department of Medicine, Pulmonary and Critical Care Medicine, University Medical Centre Giessen and Marburg, Philipps-University Marburg, Marburg, Germany

**This document contains an extended methods section for the above article published in *The Journal of Immunology* in 2017 with DOI [tba]. Sections that are missing in or substantially diverge from the corresponding ones in the article are highlighted in yellow.**

## Extended Material and Methods

### *Reconstruction, verification and annotation of the $M\Phi$ signal transduction map*

For reconstruction of the map, the NCBI MEDLINE archive was queried for "macrophage", "signaling" and "infection". The obtained literature was browsed for pathways and proteins described to be important for macrophage activation and the subsequent immune response (e.g. pathogen recognition and cytokine production). This information was then graphically reconstructed using CellDesigner [v4.3 and v4.4; (1)].

Furthermore, previously published models for macrophage signaling pathways were searched in the databases BioModels [r27; (2)] and Panther Pathways [r9.0; (3)]. The following models were fully or partially re-used: BioModels MODEL1203220000, MODEL2463683119; Panther P00048, P00052, P00054, P00035, P00036, P00046, P05918. In addition, the following pathways were downloaded in May 2014 from www.macrophages.com and partially integrated: "Interleukin-4 -10 & -13 Pathways", "Macrophage Activation Extended", "NF-kappa-B Signaling", "Non-TLR Pathogen Detection", "p53 Signaling Pathway", "Toll Like Receptor & Kinase Pathways", "VEGF and TGFB Pathways". The map furthermore contains excerpts of maps previously published by Oda *et al.* (4), Oda and Kitano (5), and by Raza *et al.* (6, 7).

Next, all reactions in the curated map were verified by searching experimental evidence for their relevance in mouse or human cells. The acquired publications were used to annotate the reactions in the map using CellDesigner's MIRIAM [minimum information requested in the annotation of models, (8)] support. In addition, genes were annotated with Ensembl IDs [r75 *et sqq.*; (9, 10)], miRNAs with miRBase IDs [r20 *et sqq.*; (11)], proteins with UniProt IDs [r2014_7 *et sqq.*; (12)], complexes with BioGrid IDs [r3.2.110 *et sqq.*; (13)], and simple molecules and ions with ChEBI IDs [*r118 et sqq.*; (14)]. If

information could be neither annotated nor graphically presented, text notes were added. Due to the fact that most of the experimental evidence in immunology is obtained in, with, or starts from murine cells and models, most molecular annotations refer to the species mouse.

***Quality control of the curated map***

To validate the map's quality, a random sample of 25 reactions with modifiers (~3 % of all reactions) was selected from the final version. Two postdoctoral researchers each with 5 y of experience in molecular immunology independently graded the correspondence between each depicted reaction and its referenced literature, paying attention to errors and unsupported details. The degree of correspondence was classified into one of six groups and summary statistics were calculated.

***Expansion of the curated map with selected database knowledge***

The signaling map was transformed into a Graph Modelling language [GML; (15)] representation by an algorithm implemented in the programming language Python [v2.7; (16)] (available from https://github.com/marteber/miRNexpander). This automated process included the following conceptual steps: (i) molecule representatives (i.e. genes, mRNAs, proteins including activation states and post-translational modifications, complexes, simple molecules including pathogen-derived molecules, ions, and miRNAs) were extracted from the original map file, identifiers were assigned based on their molecular identities, and identical molecules were united in a single entity; (ii) reactions were extracted from the map file, and were categorized into stimulatory, neutral, and inhibitory; (iii) molecules were connected according to their participation in reactions, in

which modulators (enzymes, inhibitors etc.) were connected to the respective reaction products; (iv) MIRIAM annotations were extracted from the map file and inserted into the network; (v) molecules and reactions that would affect later analyses by elongating pathway lengths were removed (i.e. genes, mRNAs, simple molecules, ions, phenotypes, degradation, transcription, translation, and transport), and removed pathway parts were replaced by shortened abstract reactions; (vi) monomeric complex subunits that were not yet included in the map were supplemented in the network by introduction of an unbound molecule interacting with the respective complex.

For expansion of the network, initially missing annotation for molecules, particularly in the case of complex subunits, was automatically retrieved from the databases NCBI Gene and UniProt, manually curated and finally introduced computationally into the network. Furthermore, the databases miRTarBase [r4.5; (17)] plus miRecords [r4; (18)], TRANSFAC [r2015.1; (19)] and RegPhos [r2.0; (20)] were queried automatically for miRNA:mRNA, transcription factor (TF):gene, and kinase:phosphoprotein interactions, respectively. For selecting database knowledge that is related to molecules in the network, the retrieved database entries were filtered. First, Uniprot IDs were extracted from the map and translated into gene symbols and mouse genome informatics (MGI) accessions. Second, a list of genes associated with relevant gene ontology (GO) biological process terms was assembled by merging (i) most specific GO terms associated with MGI accessions of genes or proteins in the CellDesigner map (algorithmic assembly), and (ii) GO terms extracted from QuickGO (21) with key words such as "infection" or "macrophage" (manual assembly). For the algorithmic assembly, GO terms were fetched out of the Jax Lab Bioinformatics database (22), and the list of GO terms was shortened by removing all terms that had lower-level neighbors in the GO

hierarchy (23); see the accompanying PDF for a list of the assembled GO terms. Third, the merged set of GO terms was used to query the Jax Lab database to reverse-associate them with gene symbols which were employed for filtering database entries. After filtering, the miRNA:mRNA and kinase:phosphoprotein interactions were manually curated by evaluating the referenced publications, whereas this step was skipped for the filtered TF:gene interactions due to their large number. Finally, the correct and relevant database entries were introduced into the network together with initially absent corresponding molecules, and the generated network was written to a GML file.

The converted and expanded network (GML file) was imported into and analyzed with Cytoscape [v3.2.1; (24)]. The molecules and reactions from the CellDesigner map were converted into nodes and edges in the Cytoscape network, respectively. For the graphical representation of the expanded network based on cellular compartments, the Cytoscape plug-in Cerebral [v1.2; (25)] was used.

***Integration of expression data***

The databases Gene Expression Omnibus (26) and ArrayExpress (27) were queried for expression data sets that meet the following criteria: (i) cell type: human MΦs, monocytes, or PBMCs; (ii) context: bacterial lung infection; (iii) time point: early, non-chronic phase of infection; (iv) data: publicly available; processing and analysis technically feasible. The data sets GSE61535 (human monocyte-derived MΦs infected with *L. pneumophila* strain AA100 for 1 h at MOI = 20:1 and incubated for another 8 h) and E-MEXP-3805 (human monocyte-derived MΦs infected with *M. tuberculosis* strain H37Rv for 2 h at MOI 1:1, then incubated with fresh medium for another 18 h) were the

best matches. Furthermore, we used an expression data set (GSE77506) of human monocyte-derived MΦs infected for 16 h with *S. pneumoniae* strain D39 (MOI = 1:10). Only human data sets were used for the case study to improve its clinical relevance. In addition, six data sets related to LPS stimulation under different experimental conditions (GSE4712, GSE8621, GSE46903, GSE50542, GSE28880, and E-MEXP-3469) as well as one related to reactive oxygen species (ROS) challenge (GSE15457) were added to the web platform.

All data sets were pre-processed and analyzed with the statistics tool R (28) with the package limma (29). For all unprocessed sets, data were background-corrected, quantile-normalized and log2-transformed before feature selection with control of the false discovery rate (30) was performed. All resulting expression-related values were assigned to the molecules in the network based on the latter's gene symbol annotation.

***Identification of perturbed subnetworks and integration of drugs***

For the identification of perturbed subnetworks, the Cytoscape v3.2 plug-in jActiveModules [v3.1, (31)] was employed. jActiveModules combines a statistical scoring system with a search algorithm based on simulated annealing and was applied to each of the 44 data sets separately. The number of modules to detect was set to 3, the overlap threshold (i.e. maximal fraction of overlap between detected network regions) was set to 0.1, the search depth (i.e. distance of module growth steps starting from differentially regulated factors) to 1, and the maximal depth to 1 (i.e. maximal number of iterations). The three obtained modules were merged into one subnetwork that is representative for the corresponding data set.

For the case study on Lpn, Mtb, and Spn, we relaxed the above parameter settings to allow for some factor overlap between the scenarios. To achieve that, the module count per scenario was increased to 10, the overlap threshold to 0.5, and the maximal depth to 3. The 10 modules were then combined, yielding one subnetwork per infection scenario. From these three, the shared factors and reactions were extracted, and the resulting network was termed regulatory core. The Venn diagram of the three subnetworks was produced in R with the package VennDiagram (32).

The Cytoscape plug-in CyTargetLinker [v3.0.1; (33)] was used to attach drugs to their target proteins based on information listed in the database DrugBank 4.2 (34).

***Development and testing of the web platform***

The web platform was built around the Cytoscape.js graph theory and network rendering engine (35). It takes advantage of modern CSS- and JavaScript-based Web technologies and integrates libraries of the open-source community. Please see the Web page at https://vcells.net/macrophage for details.

Testing was performed with modern versions of the browsers available to us, i.e. Firefox, Chrome, Safari, and Edge on their eligible hosts (MS Windows, MacOS and Unix). Note that Internet Explorer is not supported even in its latest release.

***Addition of GO term and Reactome associations to the Web platform***

On the Web platform, GO terms (36) and Reactome pathways (37) can be selected to highlight genes associated to them. The terms and pathways that are selectable were chosen because they are associated to three or more genes in the respective map.

**Literature**


1. Kitano, H., A. Funahashi, Y. Matsuoka, and K. Oda. 2005. Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* 23: 961–966.
2. Le Novère, N., B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, and M. Hucka. 2006. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* 34: D689-691.
3. Mi, H., B. Lazareva-Ulitsky, R. Loo, A. Kejariwal, J. Vandergriff, S. Rabkin, N. Guo, A. Muruganujan, O. Doremieux, M. J. Campbell, H. Kitano, and P. D. Thomas. 2005. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* 33: D284-288.
4. Oda, K., T. Kimura, Y. Matsuoka, M. Muramatsu, and H. Kitano. 2004. Molecular interaction map of Macrophage. *AfCS Res. Rep. Online* 2: 12.
5. Oda, K., and H. Kitano. 2006. A comprehensive map of the toll-like receptor signaling network. *Mol. Syst. Biol.* 2.
6. Raza, S., K. A. Robertson, P. A. Lacaze, D. Page, A. J. Enright, P. Ghazal, and T. C. Freeman. 2008. A logic-based diagram of signalling pathways central to macrophage activation. *BMC Syst. Biol.* 2: 36.
7. Raza, S., N. McDerment, P. A. Lacaze, K. Robertson, S. Watterson, Y. Chen, M. Chisholm, G. Eleftheriadis, S. Monk, M. O'Sullivan, A. Turnbull, D. Roy, A. Theocharidis, P. Ghazal, and T. C. Freeman. 2010. Construction of a large scale integrated map of macrophage pathogen recognition and effector systems. *BMC Syst. Biol.* 4: 63.
8. Le Novère, N., A. Finney, M. Hucka, U. S. Bhalla, F. Campagne, J. Collado-Vides, E. J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J. L. Snoep, H. D. Spence, and B. L. Wanner. 2005. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* 23: 1509–1515.
9. Flicek, P., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. Hunt, N. Johnson, T. Juettemann, A. K. Kähäri, S. Keenan, E. Kulesha, F. J. Martin, T. Maurel, W. M. McLaren, D. N. Murphy, R. Nag, B. Overduin, M. Pignatelli, B. Pritchard, E. Pritchard, H. S. Riat, M. Ruffier, D. Sheppard, K. Taylor, A. Thormann, S. J. Trevanion, A. Vullo, S. P. Wilder, M. Wilson, A. Zadissa, B. L. Aken, E. Birney, F. Cunningham, J. Harrow, J. Herrero, T. J. P. Hubbard, R. Kinsella, M. Muffato, A. Parker, G. Spudich, A. Yates, D. R. Zerbino, and S. M. J. Searle. 2014. Ensembl 2014. *Nucleic Acids Res.* 42: D749-755.
10. Cunningham, F., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, A. K. Kähäri, S. Keenan, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, B. Overduin, A. Parker, M. Patricio, E. Perry, M. Pignatelli, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, B. L. Aken, E. Birney, J. Harrow, R. Kinsella, M. Muffato, M. Ruffier, S. M. J. Searle, G. Spudich, S. J. Trevanion, A. Yates, D. R. Zerbino, and P. Flicek. 2015. Ensembl 2015. *Nucleic Acids Res.* 43: D662-669.
11. Griffiths-Jones, S., R. J. Grocock, S. van Dongen, A. Bateman, and A. J. Enright. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34: D140-144.

12. Apweiler, R., A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L.-S. L. Yeh. 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 32: D115-119.
13. Stark, C., B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34: D535-539.
14. Degtyarenko, K., P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcántara, M. Darsow, M. Guedj, and M. Ashburner. 2008. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.* 36: D344-350.
15. University Passau. Graph Modelling Language. http://www.fim.uni-passau.de/en/fim/faculty/chairs/theoretische-informatik/projects.html. .
16. Sanner, M. F. 1999. Python: a programming language for software integration and development. *J. Mol. Graph. Model.* 17: 57–61.
17. Hsu, S.-D., F.-M. Lin, W.-Y. Wu, C. Liang, W.-C. Huang, W.-L. Chan, W.-T. Tsai, G.-Z. Chen, C.-J. Lee, C.-M. Chiu, C.-H. Chien, M.-C. Wu, C.-Y. Huang, A.-P. Tsou, and H.-D. Huang. 2011. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 39: D163-169.
18. Xiao, F., Z. Zuo, G. Cai, S. Kang, X. Gao, and T. Li. 2009. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.* 37: D105-110.
19. Matys, V., O. V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A. E. Kel, and E. Wingender. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* 34: D108-110.
20. Huang, K.-Y., H.-Y. Wu, Y.-J. Chen, C.-T. Lu, M.-G. Su, Y.-C. Hsieh, C.-M. Tsai, K.-I. Lin, H.-D. Huang, T.-Y. Lee, and Y.-J. Chen. 2014. RegPhos 2.0: an updated resource to explore protein kinase-substrate phosphorylation networks in mammals. *Database J. Biol. Databases Curation* 2014: bau034.
21. Binns, D., E. Dimmer, R. Huntley, D. Barrell, C. O'Donovan, and R. Apweiler. 2009. QuickGO: a web-based tool for Gene Ontology searching. *Bioinforma. Oxf. Engl.* 25: 3045–3046.
22. Jax Lab Bioinformatics. 2014. Resource for GO term-gene associations. Retrieved from *ftp://ftp.informatics.jax.org/pub/reports/gene_association.mgi* on October 31, 2014.
23. Gene Ontology Consortium. 2014. GO term database (lite version). Retrieved from *http://archive.geneontology.org/latest-lite* on August 16, 2014.
24. Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498–2504.
25. Barsky, A., J. L. Gardy, R. E. W. Hancock, and T. Munzner. 2007. Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinforma. Oxf. Engl.* 23: 1040–1042.
26. Edgar, R., M. Domrachev, and A. E. Lash. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30: 207–210.
27. Kolesnikov, N., E. Hastings, M. Keays, O. Melnichuk, Y. A. Tang, E. Williams, M. Dylag, N. Kurbatova, M. Brandizi, T. Burdett, K. Megy, E. Pilicheva, G. Rustici, A.

Tikhonov, H. Parkinson, R. Petryszak, U. Sarkans, and A. Brazma. 2015. ArrayExpress update--simplifying data submissions. *Nucleic Acids Res.* 43: D1113-1116.
28. R Core Team. 2015. *R: A language and environment for statistical computing.*,. R Foundation for Statistical Computing, Vienna, Austria.
29. Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43: e47–e47.
30. Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 57: 289–300.
31. Ideker, T., O. Ozier, B. Schwikowski, and A. F. Siegel. 2002. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinforma. Oxf. Engl.* 18 Suppl 1: S233-240.
32. Chen, Hanbo. 2016. *VennDiagram: Generate High-Resolution Venn and Euler Plots*,.
33. Kutmon, M., T. Kelder, P. Mandaviya, C. T. A. Evelo, and S. L. Coort. 2013. CyTargetLinker: a cytoscape app to integrate regulatory interactions in network analysis. *PloS One* 8: e82160.
34. Law, V., C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z. T. Dame, B. Han, Y. Zhou, and D. S. Wishart. 2014. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42: D1091-1097.
35. Franz, M., C. T. Lopes, G. Huck, Y. Dong, O. Sumer, and G. D. Bader. 2015. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* btv557.
36. Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25: 25–29.
37. Croft, D., A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M. R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, and P. D'Eustachio. 2014. The Reactome pathway knowledgebase. *Nucleic Acids Res.* 42: D472–D477.