

Technological Workshop. Omics Data Integration

Integrative Omics Data Analysis, an Overview

Alex Sánchez

March 29, 2017

Genetics, Microbiology & Statistics Department
University of Barcelona
Statistics and Bioinformatics Unit (UEB)
Vall Hebron Institut de Recerca



Table of Contents

1 Introduction and Context

- From Biology to Omics
- Why should we integrate data?

2 Methods and Tools for IODA

- Multivariate Statistical Methods
- Principal Components Analysis
- Predictive Models for Integrative Analysis
- Network Methods for Integrative Analysis
- Using Biological Knowledge for Integrative Analysis
- Applying Integrative GSEA

3 Examples and case studies

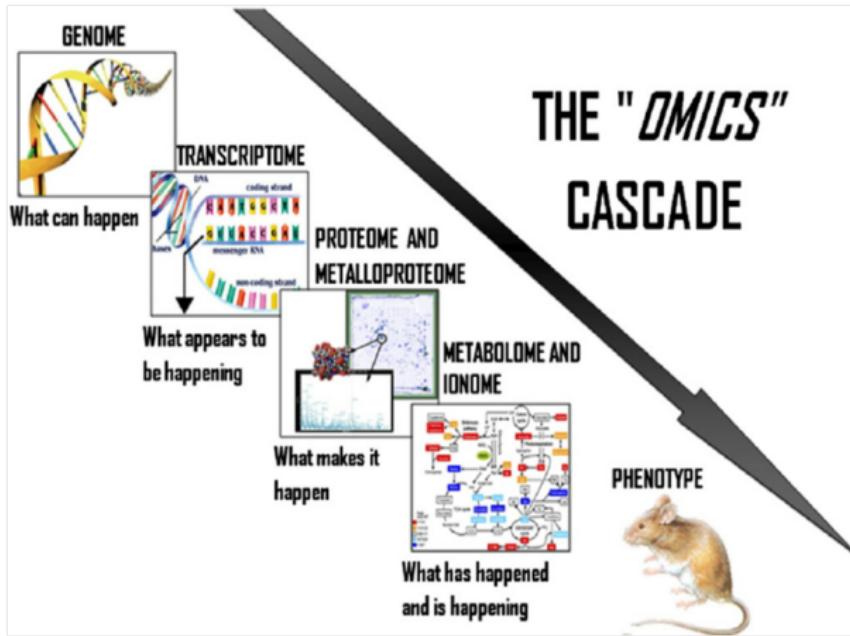
4 Summary and Take-home

What is “omics” ?

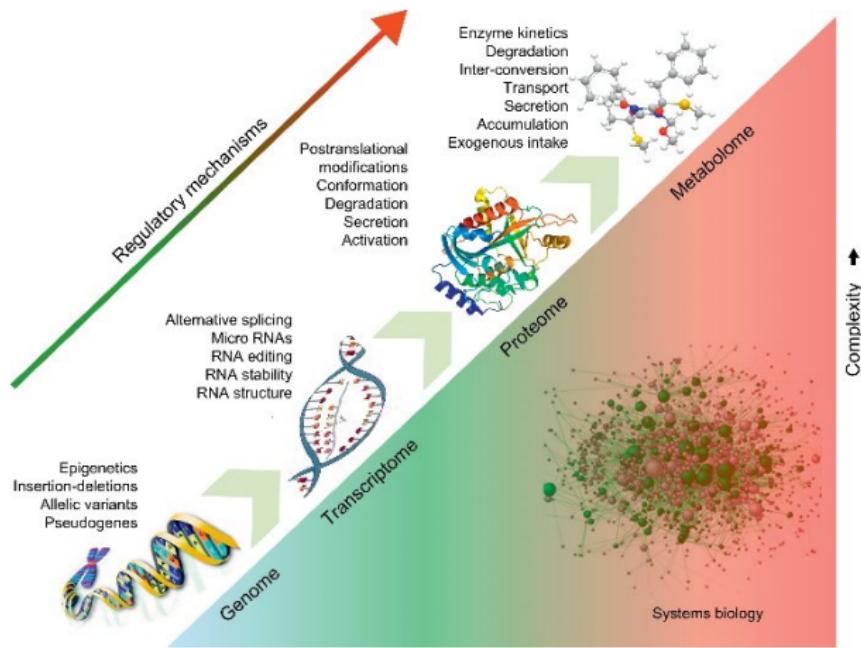
- In biological context , the suffix “omics” is used to refer to the study of large sets of biological molecules (Smith et al., 2005)
- The study of different components participating and/or regulating complex biological processes, triggered the development of several fields that, together, are described with the term OMICS.



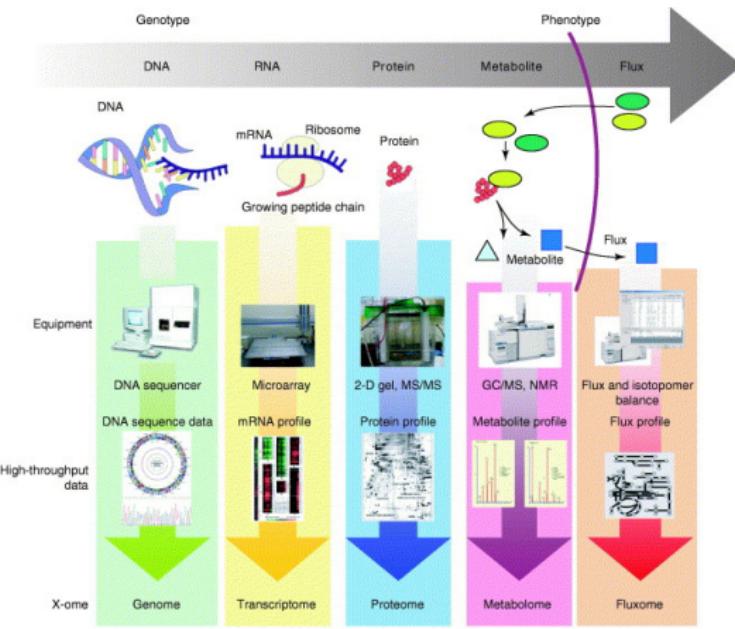
The Omics Cascade: “Omics”



The Omics Cascade: “Omes” +RegulOME

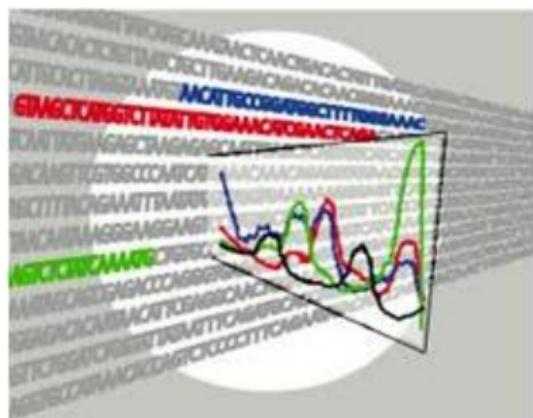


We study “Omics” with “Omics” technologies



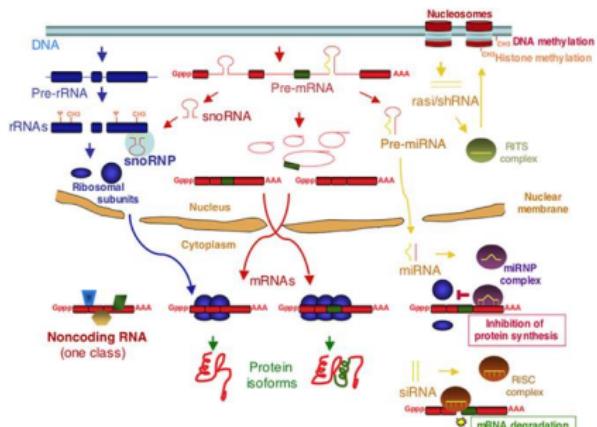
Genomics

Genomics is a discipline in genetics that applies recombinant DNA, DNA sequencing methods, and bioinformatics to sequence, assemble, and analyze the function and structure of genomes (the complete set of DNA within a single cell of an organism)



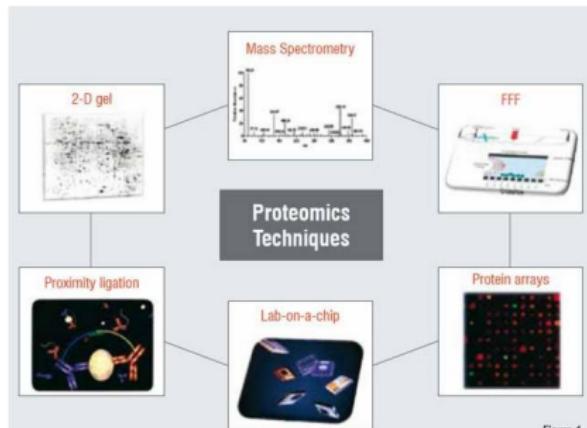
Transcriptomics

- The transcriptome is the set of all RNA molecules, in one or a population of cells.
- Transcriptomics, examines expression levels of mRNAs in a given cell population, often using high-throughput techniques: microarrays or NGS.



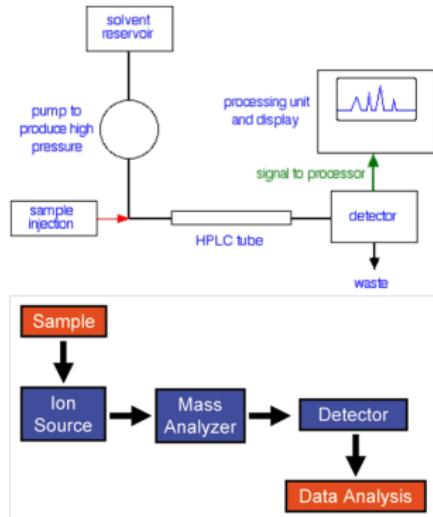
Proteomics

- The large-scale study of proteins (the proteome), particularly their structure and function.
- Relies on a wide spectra of techniques
 - 2D gel based
 - Mass Spectrometry (MS)
 - Seldi-TOF (MS)
 - Protein Arrays
 - ...



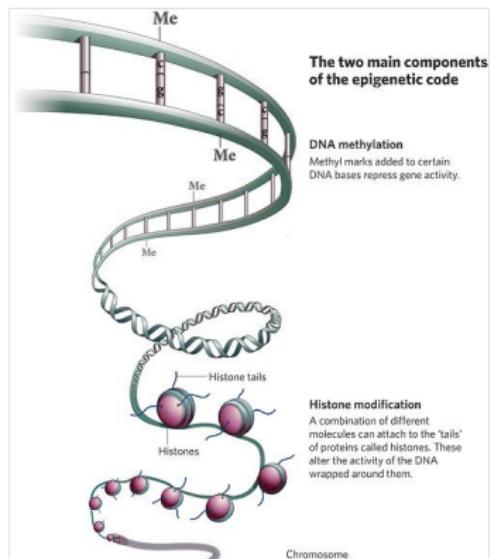
Metabolomics

- Comprehensive and simultaneous systematic determination of
 - metabolite levels in the metabolome and
 - their changes over time as a consequence of stimuli.
- Relies on
 - Separation techniques: GC, CE, HPLC, UPLC
 - Detection techniques: NMR, MS

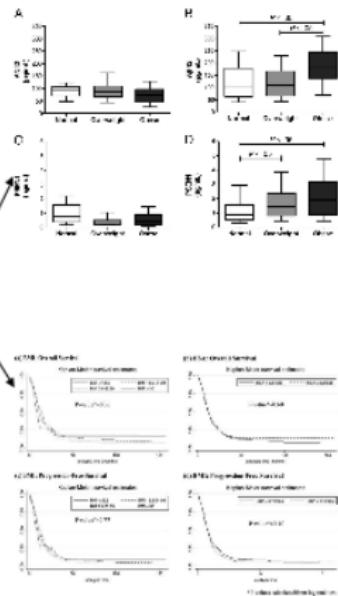
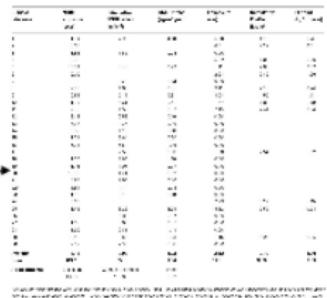
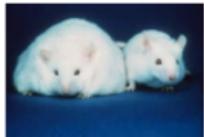


Epigenetics and Epigenomics

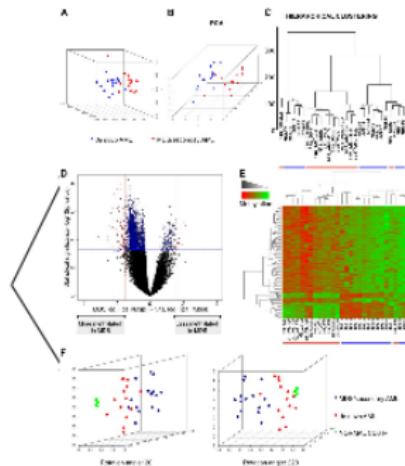
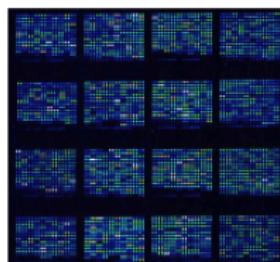
- *Epigenetics* is the study of changes in the phenotype or gene expression caused by other mechanisms than changes in the underlying DNA sequence.
 - DNA methylation
 - Histone modifications
- Epigenetics refers to the study of single genes or sets of genes. Epigenomics refers to global analyses of epigenetic changes across



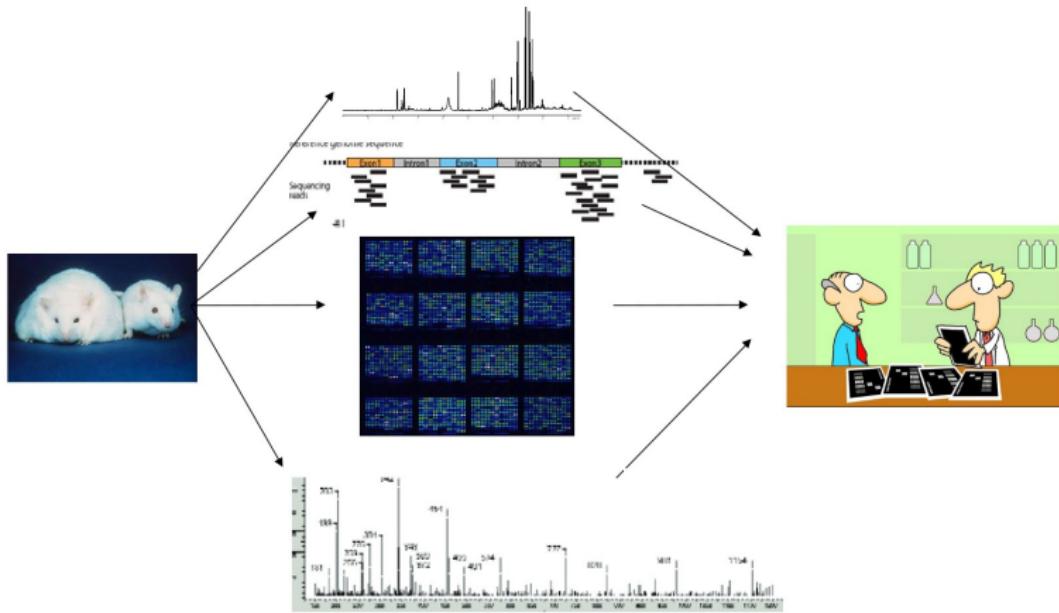
How we made studies in 1996



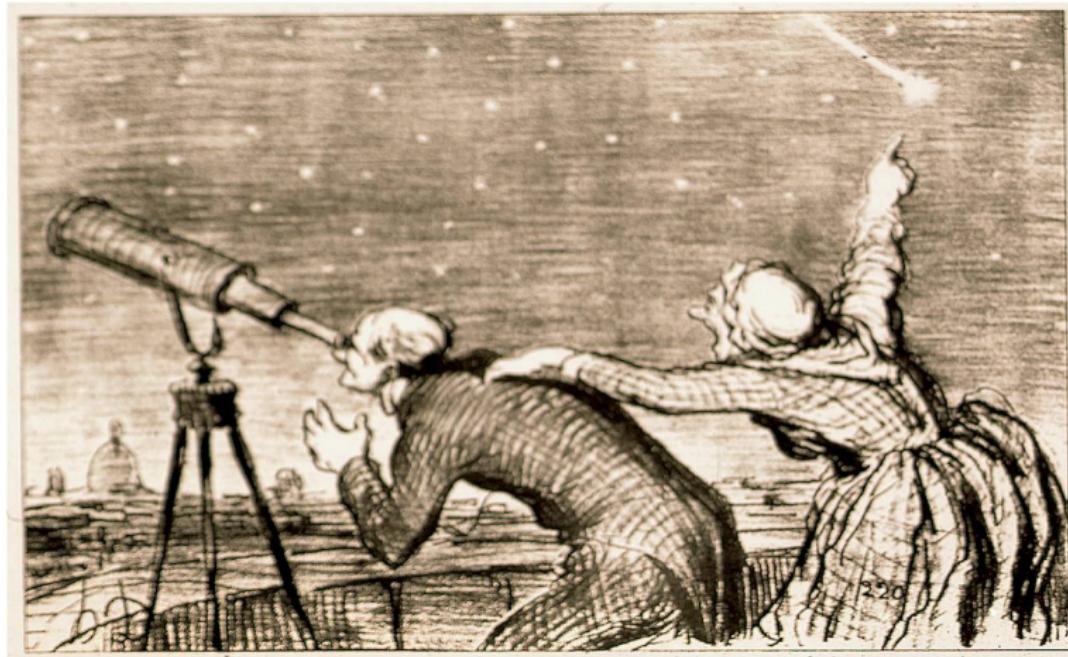
How we made studies in 2006



How we would like to make studies in 2016



Focussing only on one platform risks missing an obvious signal

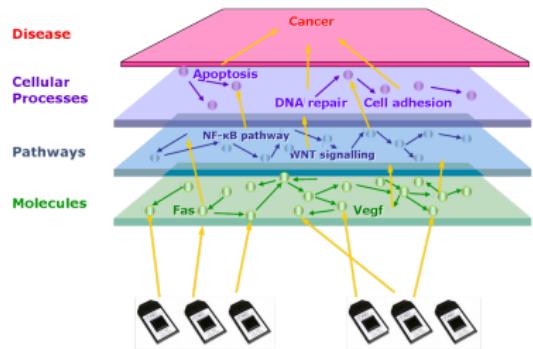
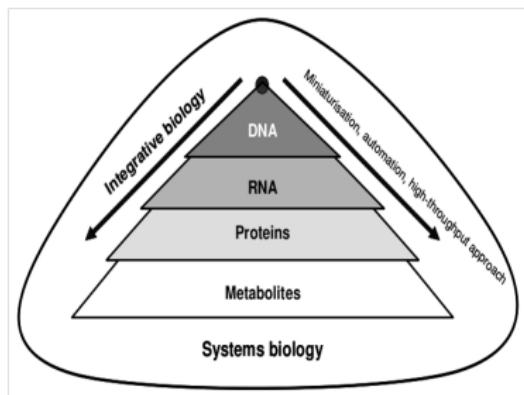


Integrative Omics Data Analysis

- The idea that efficient integration of data from different OMICS can greatly facilitate the discovery of true causes and states of disease is rapidly pervading the biomedical community (Joyce and Palsson, 2006).
- The aims of integrative analysis is the deciphering of complex biological relationships empowered by the combined use of distinct pieces of information that represent a, probably partial, view of the different levels at which these processes happen.

From componentwise to global approaches

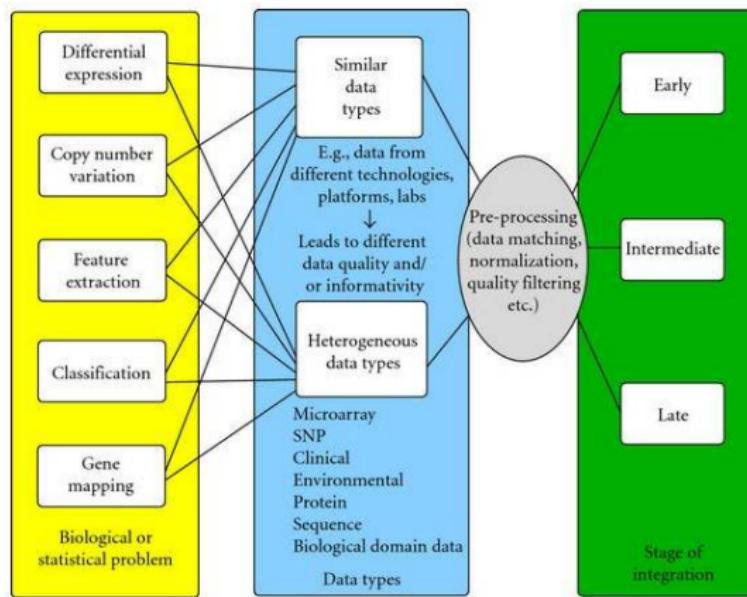
- We expect (Gomez-Cabrero et al., 2014) that the integrated collection and analysis of diverse types of data,
- jointly modeled and analyzed in a systems biology approach,
- can shed light on the global functioning of biological systems.



So what is Data Integration?

- “*Data integration*” may mean different things (Hamid et al., 2009).
 - Computational combination of data
 - Combination of studies performed independently
 - Simultaneous analysis of multiple variables on multiple datasets.
 - Not to mention any possible approach for homogeneously querying heterogeneous data sources
- **Integrative analysis** may be preferable

There are many types of integrative analysis



There are many methods ...

BMC Bioinformatics



Research article

Open Access

A structured overview of simultaneous component based data integration

Katrijn Van Deun¹, Age K Smilde², Mariët J van der Werf³, Henk AL Kiers⁴
and Iven Van Mechelen¹

Vol. 26 ISMB 20
doi:10.1093/bio

Multivariate multi-way analysis of multi-source data

Illiia Huopaniemi^{1,*}, Tommi Suvitala¹, Janne Nikkilä^{1,2}, Matej Oresic³ and
Samuel Kaski^{1,*}

Advances in Integrative Causal Analysis

Presenter: Ioannis Tsamardinos, Ph.D., Associate Professor and
Sofia Triantafillou, Ph.D. candidate.

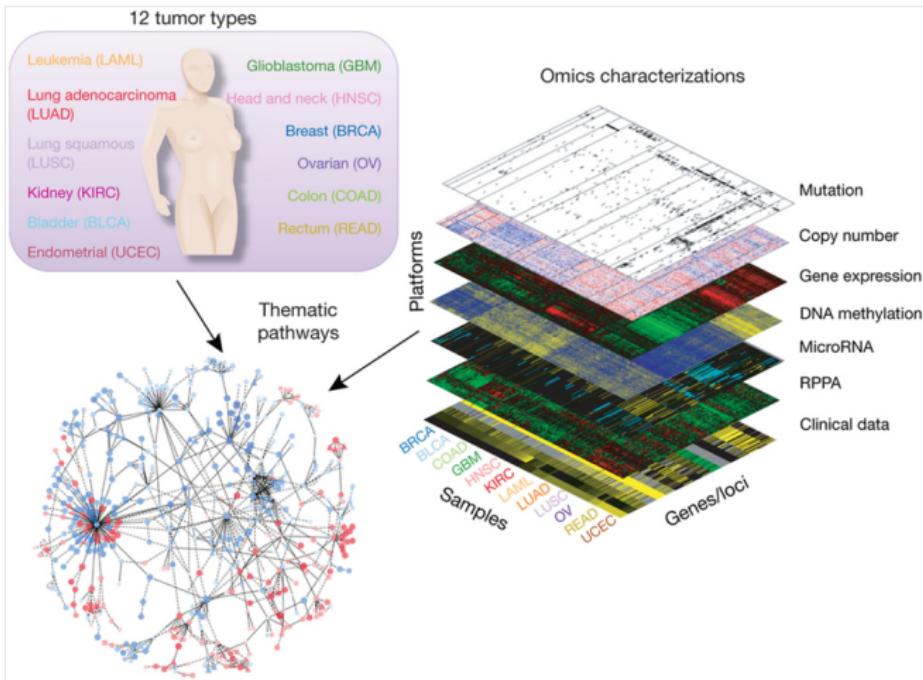
Low level data fusion methods searching for common
and distinctive biological information in hetero omics
data-sets

Frans M. van der Kloet¹, Patricia Sebastián-León², Ana Conesa²,
Age K. Smilde¹ and Johan A. Westerhuis¹

There are many tools...



There are many data sources...



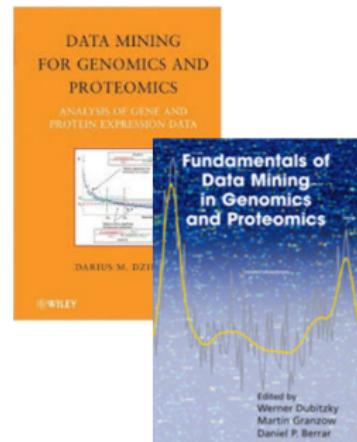
So what?

- We (try to) provide a transversal walk on methods
 - Dimension Reduction Approaches.
 - Sparse regression and classification.
 - Network methods.
 - Adding Biological knowledge
- and tools
 - MixOmics, OmicADE, PaintOmics, giTools,
- and provide some examples of applications and use cases
 - miRNA-mRNA.
 - mRNA-methylation.
 - proteomic-transcriptomic.
 - clinical and omics-integration

Methods and Tools for IODA

Multivariate statistics in genomics

- Multivariate methods have pervaded the field of genomics since its very beginning
 - The (in)famous clustering (HC, heatmaps)
 - Matrix factorizations / Dimension reduction (PCA, SVD, CoA)
 - Discriminant Analyses (LDA – > DLDA, ...)



To cite but a few.

Best approach for omics data analysis?

- Classical Statistics
 - Multiple regression
 - Discriminant analysis
 - ANOVA
- Data tables are long and lean



- Assumptions
 - Independent variables
 - More observations than variables
 - Multivariate normality
 - Interested in one dependent
 - Few missings
- DO NOT hold for many omics data

The nature of omics data

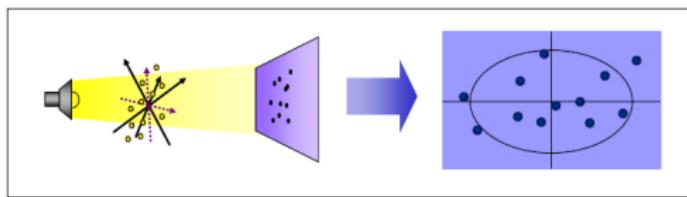
- Omics data are diverse
 - They measure distinct characteristics
 - GC/MS spectrum, Expression, Concentration
- Although they have aspects in common
 - Most of them are high throughput
 - Many variables (K) measured simultaneously
 - Relatively expensive, ethical limitations, regulations
 - Few samples (N) analyzed



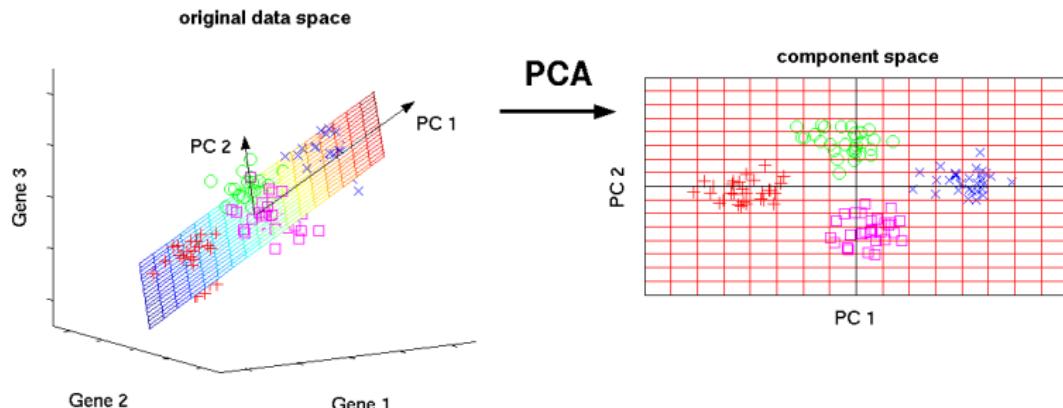
$$K \gg N$$

A Better Way

- Multivariate analysis by projection (dimension-reduction, matrix decomposition) methods (Meng et al., 2016):
 - Looks at ALL the variables together
 - Avoids loss of information
 - Finds underlying trends = “latent variables”
 - More stable models



Principal Components Analysis



Representation in reduced dimension (2 first PCs) improves visualization

Examples

The screenshot shows a detailed view of a research article from the journal *Science*. The article is titled "The Transcriptional Program of Sporulation in Budding Yeast" by S. Chu, J. DeRisi, M. Eisen, J. Mulholland, D. Botstein, P. O. Brown, and I. Herskowitz. The page includes navigation links like "Article Views", "Abstract", "Full Text", and "Full Text (PDF)". It also displays the journal's masthead, search bar, and sidebar information.

Science The World's Leading Journal of Original Scientific Research, Global News, and Commentary.

AAAS.ORG | FEEDBACK | HELP | LIBRARIANS | All Science Journals | Enter Search Term | CRAI: UNIVERSITAT DE BARCELONA | ALERTS | ACCESS RIGHTS

AAAS NEWS SCIENCE JOURNALS CAREERS MULTIMEDIA COLLECTIONS

Science Home Current Issue Previous Issues Science Express Science Products My Science About the Journal

Home > Science Magazine > 23 October 1998 > Chu et al., 282 (5389): 699-705

Article Views Abstract Full Text Full Text (PDF) A correction has been published

RESEARCH ARTICLE

Science 23 October 1998;
Vol. 282 no. 5389 pp. 699-705
DOI: 10.1126/science.282.5389.699

< Prev | Table of Contents | Next >

The Transcriptional Program of Sporulation in Budding Yeast

S. Chu,¹ J. DeRisi,¹ M. Eisen,¹ J. Mulholland,¹ D. Botstein,¹ P. O. Brown,¹ I. Herskowitz²

Pac Symp Biocomput. Author manuscript; available in PMC 2009 Apr 17.
Published in final edited form as:
Pac Symp Biocomput. 2000 : 455–466.

PMCID: PMC2669932
NIHMSID: NIHMS97353

PRINCIPAL COMPONENTS ANALYSIS TO SUMMARIZE MICROARRAY EXPERIMENTS: APPLICATION TO SPORULATION TIME SERIES

Soumya Raychaudhuri,^{*} Joshua M. Stuart,^{*} and Russ B. Altman,^Ψ

Author information ► Copyright and License information ►

Author Manuscript

Examples



Journal of Proteomics

Volume 75, Issue 13, 16 July 2012, Pages 3938–3951



Batch effects correction improves the sensitivity of significance tests in spectral counting-based comparative discovery proteomics

Josep Gregori^{a,b}, Laura Villarreal^a, Olga Méndez^a, Alex Sánchez^{b,c}, José Baselga^a, Josep Villanueva^a.

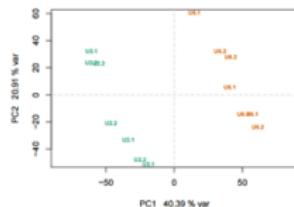


+ Show more

R Packages alpha Home Blog All packages

Search R packages

Search



Exploratory Data Analysis of LC-MS/MS data by spectral counts

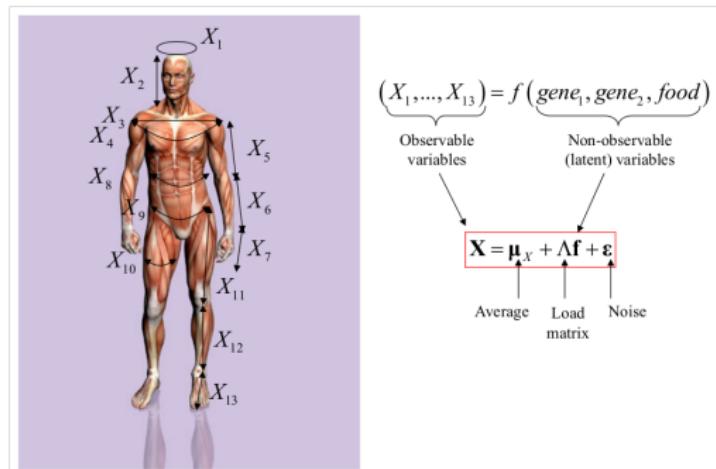
Exploratory data analysis to assess the quality of a set of LC-MS/MS experiments, and visualize de influence of the involved factors.

Author	Josep Gregori, Alex Sanchez, and Josep Villanueva
Date of publication	None
Maintainer	Josep Gregori <josep.gregori@gmail.com>
License	GPL-2
Version	1.2.0



Factor Analysis

- Related to Principal Components Analysis
- It explains observed variables as linear combinations of a set of unobservable or *latent factors* in the data
- Used in practice to model/estimate a few factors (< # variables).



Examples

Cancer Cell

Volume 17, Issue 1, 19 January 2010, Pages 98–110



Article

Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*

Roel G.W. Verhaak^{1, 2, 17}, Katherine A. Hoadley^{3, 4, 17}, Elizabeth Purdom⁷, Victoria Wang⁸, Yuan Qi^{4, 5}, Matthew D. Wilkerson^{4, 5}, C. Ryan Miller^{4, 6}, Li Ding⁹, Todd Golub^{5, 10}, Jill P. Mesirov¹, Gabriele Alexe¹, Michael Lawrence^{1, 2}, Michael O'Kelly^{1, 2}, Pablo Tamayo¹, Barbara A. Weir^{1, 2}, Stacey Gabriel¹, Wendy Winckler^{1, 2}, Supriya Gupta¹, Lakshmi Jakkula¹¹, Heidi S. Feller¹¹, J. Graeme Hodgson¹², C. David James¹², Jann N. Sarkaria¹³, Cameron Brennan¹⁴, Ari Kahn¹⁵, Paul T. Spellman¹¹, Richard K. Wilson⁹, Terence P. Speed^{7, 16}, Joe W. Gray¹¹,



OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

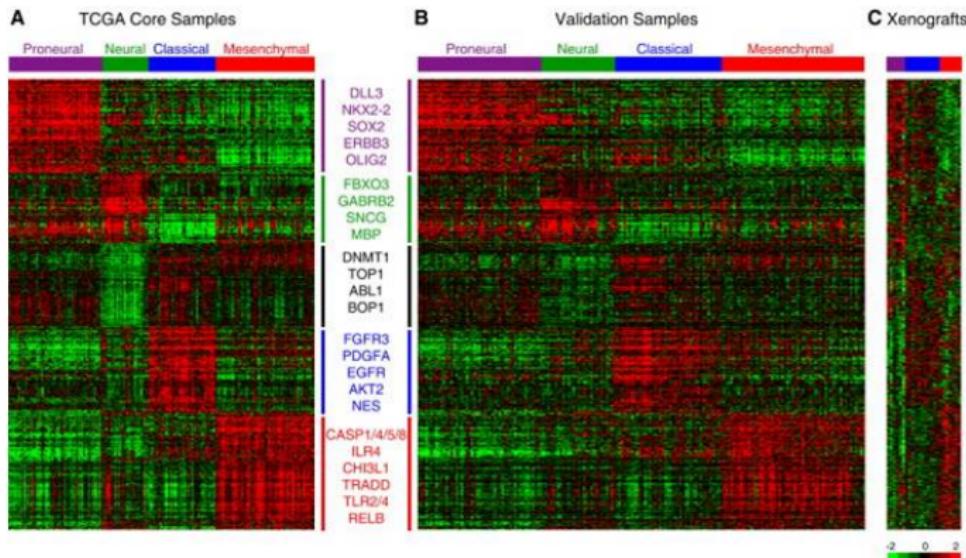
Unifying Gene Expression Measures from Multiple Platforms Using Factor Analysis

Xin Victoria Wang , Roel G. W. Verhaak, Elizabeth Purdom, Paul T. Spellman, Terence P. Speed

Published: March 11, 2011 • DOI: 10.1371/journal.pone.0017691

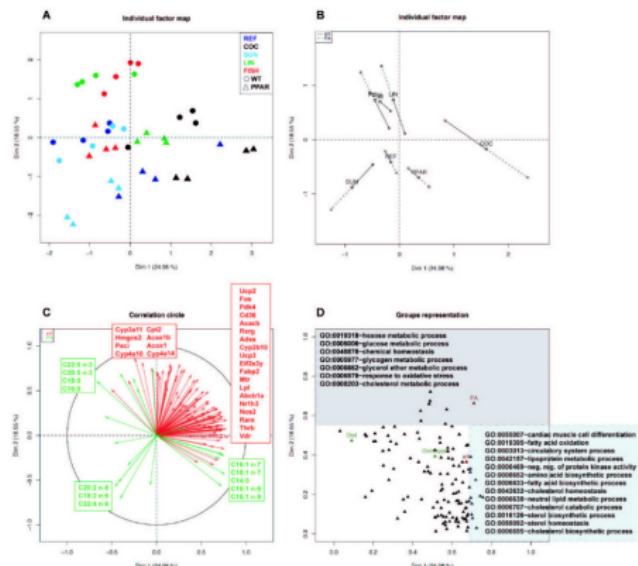
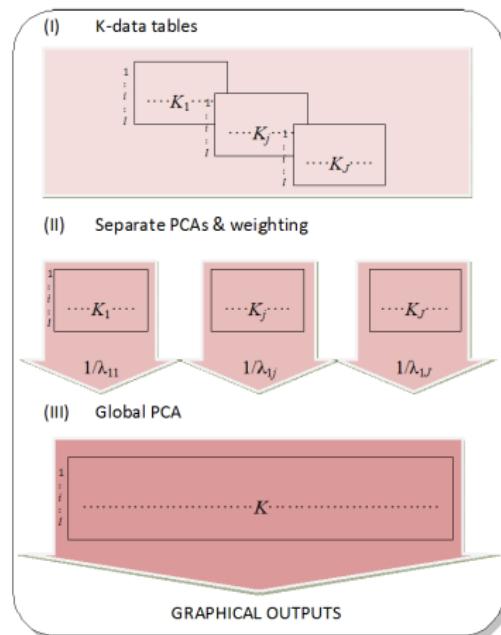


Gene Expression Data Identify 4 Gene Expression Subtypes



Verhaak et al. (2010)

Multiple Factor Analysis



MFA Applications



American Diabetes Association. diabetes

Home | Current Issue | Archive | Contact Us | Subscribe | Help | Alerts | ☰

Institution: SECCIO DEBIOLOGIA

CD14 Modulates Inflammation-Driven Insulin Resistance

José Manuel Fernández-Real^{1,2,3}, Sofía Pérez del Pulgar^{1,2,3},
Elodie Luche^{4,5}, José María Moreno-Navarrete¹, Aurelie Waget^{4,5},
Matteo Serino^{4,5}, Eleonora Sorianello³, Alex Sánchez-Plá^{6,7},

Integrative Analysis of the Relationship Between Insulin Resistance and Gut Microbiota



Bioinformatics in Personalized Medicine

10th Spanish Symposium, JSI 2010, Torremolinos, Spain, October 27-29, 2010. Revised Selected Papers
Series: »Lecture Notes in Computer Science, Vol. 6620
Subseries: »Lecture Notes in Bioinformatics

Multivariate Methods for the Integration and Visualization of Omics Data

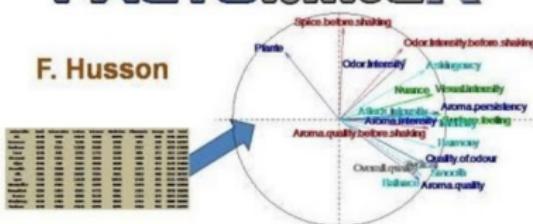
Alex Sánchez^{1, **}, José Fernández-Real², Esteban Vegas¹, Francesc Carmona¹, Jacques Amar², Remy Burcelin³, Matteo Serino³, Francisco Tinahones⁴ M. Carmen Ruiz de Villa¹, Antonio Minarro¹ and Ferran Reverter¹

FactoMineR: A tool for Multiple Factor Analysis

Multiple Factor Analysis with

FACTOMINER

F. Husson



<http://factominer.free.fr/>

FUN AgroCampus Ouest Exploratory Multivariate Data Analysis

Home • All courses • Exploratory Multivariate Data Analysis

Exploratory Multivariate Data Analysis

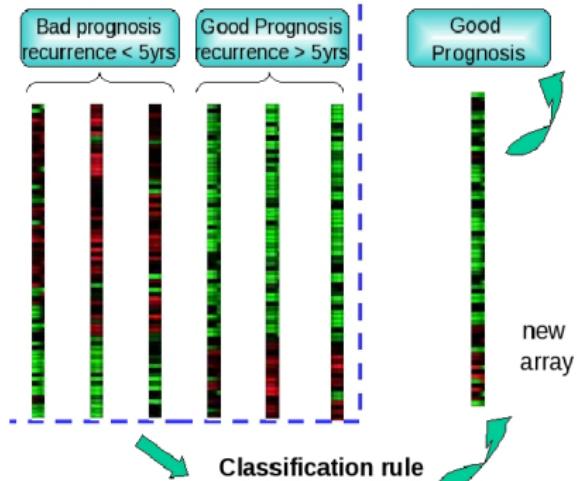
Thematics Mathématiques et statistiques sciences pour l'ingénieur

A screenshot of the Exploratory Multivariate Data Analysis course page. It shows a 3D biplot with a play button, a 2D biplot, and a data matrix. The course navigation bar includes Home, All courses, Exploratory Multivariate Data Analysis, Thematics, Mathématiques et statistiques, sciences pour l'ingénieur, and a search bar.

Predictive Models for Integrative Analysis

Predictive Models to Select Biomarkers

- Goal: build classifiers or predictors to:
 - Classify tumors. Assign new sample to one group.
 - Predict future state such as relapse, metastasis, etc.
- Problem: Thousands of variables. Difficult to select only a few
- Approach: *Penalized methods.*



Penalized Methods

- Many classical methods for building classifiers do not work well with thousands of variables (genes, proteins, etc).
- A few do well (PLS, SVM)
- Penalized methods use a trick to simplify variable selection.
 - LASSO, Elastic Net.
 - sparse-PLS
 - sPLS-LDA, O2PLS

Method	Penalty
LASSO	$\sum_{j=1}^p \beta_j < t$
Adaptive LASSO	$\sum_{j=1}^p \left(\beta_j / \hat{\beta}_j \right) < t$
Elastic net	$\sum_{j=1}^p \beta_j < t_1$ and $\sum_{j=1}^p \beta_j^2 < t_2$

Example of Penalized Methods in a Bladder Cancer Study



RESEARCH ARTICLE

Integration Analysis of Three *Omics* Data Using Penalized Regression Methods: An Application to Bladder Cancer

Silvia Pineda^{1,2}, Francisco X. Real^{3,4}, Manolis Kogevinas⁵, Alfredo Carrato⁶, Stephen J. Chanock⁷, Núria Malats^{1*}, Kristel Van Steen^{2,8*}

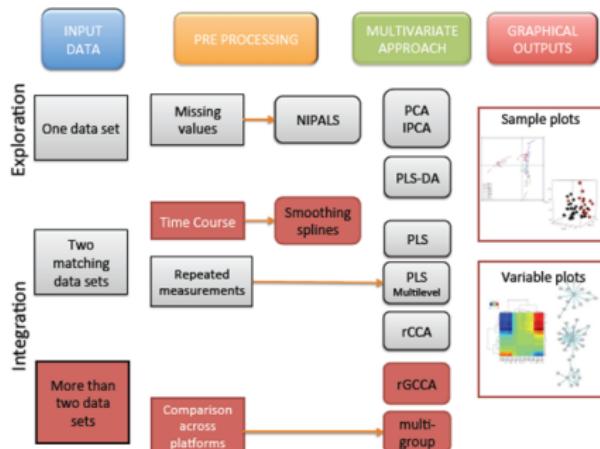
1 Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain, **2** Systems and Modelling Unit-BIO3, Montefiore Institute, Liège, Belgium, **3** Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain, **4** Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain, **5** Centre for Research in Environmental Epidemiology (CREAL) and Parc de Salut Mar, Barcelona, Spain, **6** Servicio de Oncología, Hospital Universitario Ramón y Cajal, Madrid, and Servicio de Oncología, Hospital Universitario de Elche, Alicante, Spain, **7** Division of Cancer Epidemiology and Genetics, National Cancer Institute, Department of Health and Human Services, Bethesda, Maryland, United States of America, **8** Systems Biology and Chemical Biology, GIGA-R, Liège, Belgium



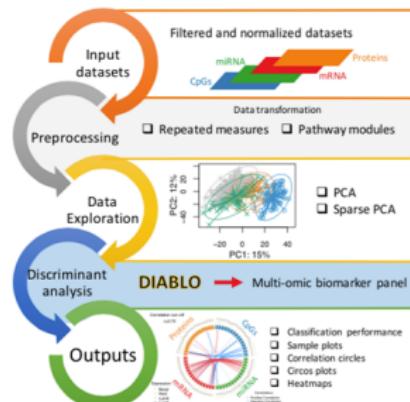
CrossMark
click for updates

mixOmics: Integrative Omics Analysis with sparse methods

Which method for which task ...



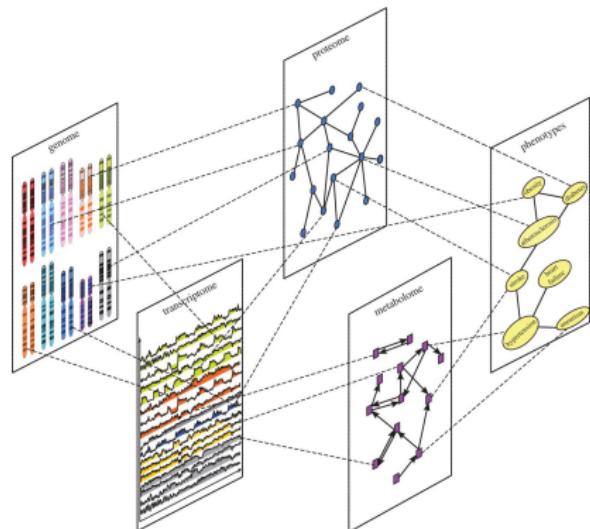
Analysis pipeline



www.mixomics.org

Network Based Integrative Analysis

- Combining networks originated in different omics/datasets seems a natural approach.
- It allows for combination of different data sources, and a more intuitive interpretability,
- At the same time it keeps the original features of study.
- Warning: *For a given network one must always understand its nodes and edges!*



Gligorijevi and Prulj (2015)

Network submodule discovery

- Information about *canonical pathways*, *protein-protein*, *protein-DNA*, *predicted miRNA-gene interactions* is collected from databases.
- Reconstruct molecular networks.
- Map statistics (e.g. z-scores) of omics data on the NW.
- Apply graph based algorithms to identifying relevant sub-networks (Ideker et al., 2002; Dittrich et al., 2008) .
- Optimal solution available via integer linear programming [69].
- Allows combination of gene expression with other data modalities (e.g. clinical information).

Example of Integrative Network Module discovery



[Bioinformatics](#). 2008 Jul 1; 24(13): i223–i231.

PMCID: PMC2718639

doi: [10.1093/bioinformatics/btn161](https://doi.org/10.1093/bioinformatics/btn161)

Identifying functional modules in protein–protein interaction networks: an integrated exact approach

Marcus T. Dittrich,^{1,2,*†} Gunnar W. Klau,^{3,4,*†} Andreas Rosenwald,⁵ Thomas Dandekar,¹ and Tobias Müller^{1,*}

[Author information ▶](#) [Copyright and License information ▶](#)

Using Biological Knowledge in Integrative Analysis

Omics Analysis using Biological Knowledge

- Distinct multivariate methods provide different approaches to classification and visualization of omics data,
- Sparse methods useful if regularizations are required
- But even *in these cases* results are difficult to interpret biologically.
- We have adopted another approach: rely on *biological knowledge*
 - To analyze the data
 - To enhance the analysis

How is biological knowledge represented



Databases



Ontologies



Gene Sets (Signatures)

Gene Set Enrichment Analysis

- Gene Set Enrichment Analysis is a successful approach for Pathway Analysis.
- Imagine we are interested in a given Gene Set, say *genes related with tumorigenesis in a certain type of breast cancer*.
- This particular gene set had only one out of the many genes in it called differentially-expressed.
- However, when we look at the distribution of t-statistics, or the distribution of the effect size, we note that there's a slight shift to the center of the distribution.
- GSEA provides methods and tools to compute summary statistics that can be used to test effects such as this.

Integrative Gene Set Enrichment Analysis

- *Stage I integration*
 - First integrate (joint data matrices)
 - Compute gene-to-phenotype association scores using all available data types (say, using logistic regression or other linear model)
- *Meta-Analytic approach or Stage II integration*
 - Start performing GSA on each data set separately
 - Use, say, Wilcoxon p-values and take their geometric average, or take the smallest one across all data types (some consensus measurement).
Alternative use some meta-analysis method for combining p-values.

Poisson et al. BMC Bioinformatics 2011, **12**:459
<http://www.biomedcentral.com/1471-2105/12/459>



RESEARCH ARTICLE

Open Access

Integrative set enrichment testing for multiple omics platforms

Laila M Poisson¹, Jeremy M Taylor² and Debashis Ghosh^{3*}

Examples and Case Studies

Integrative Analysis of Expression and Methylation I

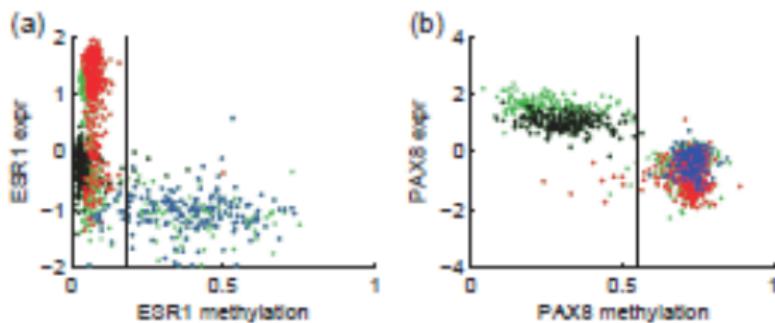
- Methylation of CpG dinucleotides in the promoter of genes involved in the oncogenic process has been shown to be a key process contributing to tumor initiation and/or progression.
- Essentially (and especially in cancer) methylation acts by inhibiting gene expression that is, *the more methylated is a gene the more repressed is its expression*
- A challenging problem is *finding genes that are regulated by methylation*.

Integrative Analysis of Expression and Methylation II

Integrative analysis of methylation and gene expression data in TCGA (Liu and Qiu, 2012).

- Considering the relation between methylation and expression in cancer (the higher methylation the lower the expression...)
- leads to expecting that scatterplots depicting the relation between methylation and expression show a negative correlation.
- This is so and indeed genes known to be regulated by methylation used to show an L-shape pattern in these plots.
- This work implements an algorithm based on Conditional Mutual Information to select genes with L-shape pattern as a surrogate for being regulated by methylation.

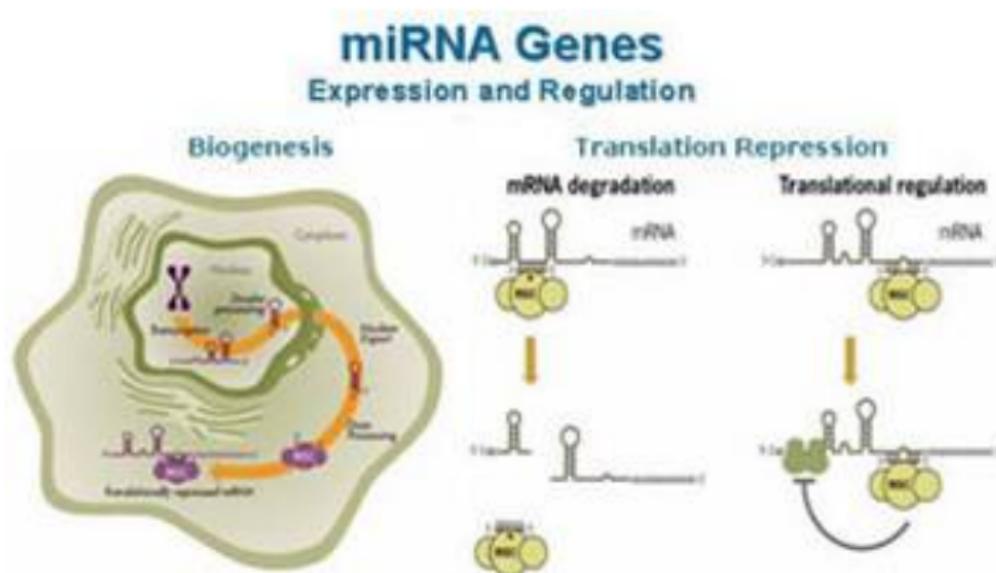
Integrative Analysis of Expression and Methylation III



Integrative Analysis of mRNA and miRNA I

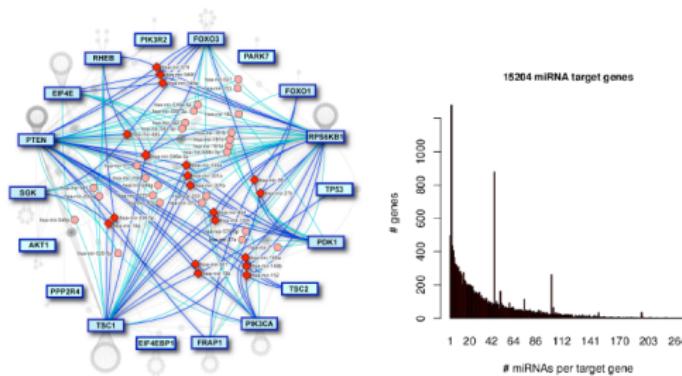
- ① miRNAs are small RNA molecules (22nt) that are key regulators of the expression of certain genes
- ② These genes are called their mRNA targets.
- ③ Only one-of-several transcription-regulating mechanisms. Others mechanisms: Alternative splicing, polyadenylation.
- ④ Although there are exceptions the general statement is that miRNAs act to repress the expression of their targets.

Integrative Analysis of mRNA and miRNA II



Integrative Analysis of mRNA and miRNA III

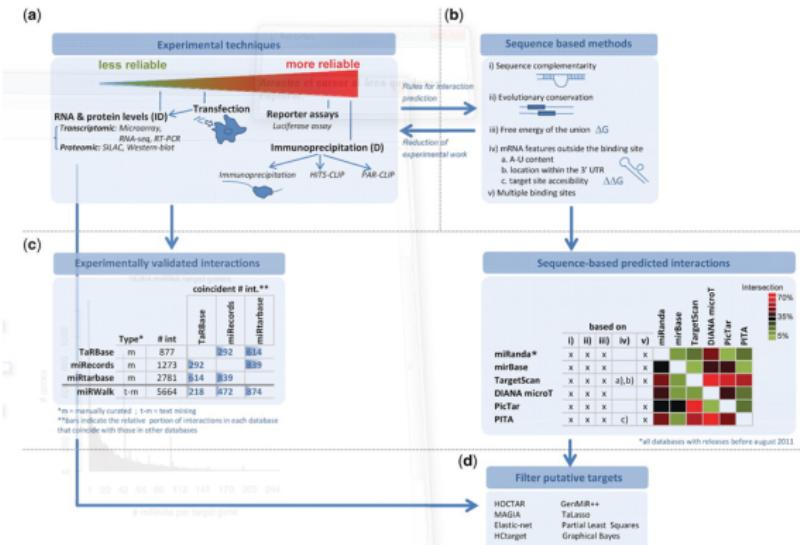
Every mRNA may target many genes,
Any gene may be the target of many miRNAs



Methods and tools for mRNA and miRNA integration I

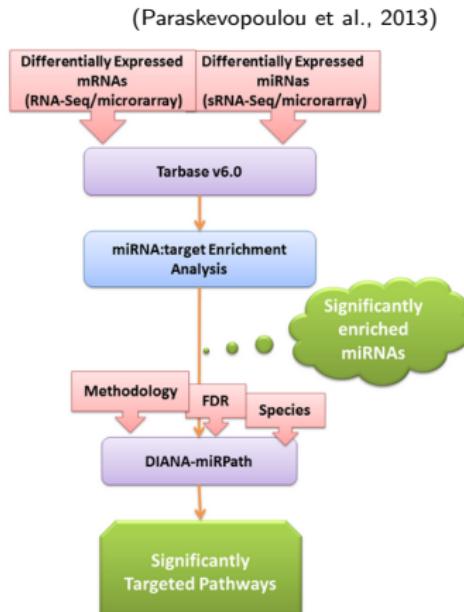
Methods and tools for mRNA and miRNA integration II

(Muniategui et al., 2013)

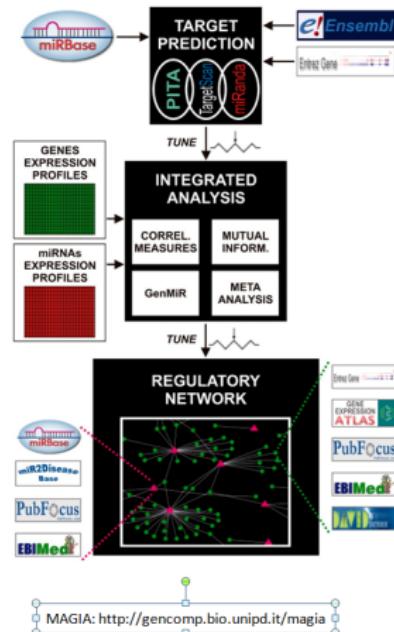


Muniategui A et al. Brief Bioinform (2012): Combining experimental and computational tools for deciphering miRNA functions and targets.

Methods and tools for mRNA and miRNA integration III



(Sales et al., 2010)



Integrating clinical and Omics data I

Journal List > Brief Bioinform > PMC4364065



Brief Bioinform. 2015 Mar; 16(2): 280–290.
Published online 2014 Mar 7. doi: 10.1093/bib/btu008

PMCID: PMC4364065

Translational research platforms integrating clinical and omics data: a review of publicly available solutions

Vincent Canuel^{1,*}, Bastien Rance^{2,3*}, Paul Avillach⁴, Patrice Degoulet⁵, and Anita Burgun⁶

[Author information](#) ► [Article notes](#) ► [Copyright and License information](#) ►

This article has been cited by other articles in PMC.

Abstract

Go to:

The rise of personalized medicine and the availability of high-throughput molecular analyses in the context of clinical care have increased the need for adequate tools for translational researchers to manage and explore these data. We reviewed the biomedical literature for translational platforms allowing the management and exploration of clinical and omics data, and identified seven platforms: BRISK, caTRIP, cBio Cancer Portal, G-DOC, iCOD, iDASH and transSMART. We analyzed these platforms along seven major axes. (1) The community axis regrouped information regarding initiators and funders of the project, as well as availability status and references. (2) We regrouped under the information content axis the nature of the clinical and omics data handled by each system. (3) The privacy management environment axis encompassed functionalities allowing control over data privacy. (4) In the analysis support axis, we detailed the analytical and statistical tools provided by the platforms. We also explored (5) interoperability support and (6) system requirements. The final axis (7) platform support listed the availability of documentation and installation procedures. A large heterogeneity was observed in regard to the capability to manage phenotype information

Brief Bioinform

Translational research platforms (Canuel et al., 2015).



Integrating clinical and Omics data II

The screenshot shows the cBioPortal homepage. At the top, there's a navigation bar with links for HOME, DATA SETS, WEB API, MATLAB, TUTORIALS, FAQ, NEWS, TOOLS, ABOUT, and VISUALIZE YOUR DATA. Below the navigation is a message about visualization, analysis, and download of large-scale cancer genomics data sets. It also includes publication guidelines and citation information. The main area features a search interface for 'Select Cancer Study' with a dropdown menu showing categories like All, Adrenocortical Carcinoma, and Cholangiocarcinoma. To the right, there's a 'What's New' section with a link to sign up for email alerts, a 'Data Sets' section showing 148 studies with a colorful heatmap, and an 'Example Queries' section with examples like RAS/RAF alterations in colorectal cancer and BRCA1 and BRCA2 mutations in ovarian cancer.

Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal (Gao et al., 2013).

Summary and Take-home

- There is no universal “IODA” method (No power-ring, yet)
- All data are not equally informative
 - Gene expression and what else
- (Type of) Objective must be kept in mind
 - Biological question should be first
 - Explanatory / Predictive
- Promising approaches are those that
 - Allow inclusion of biological information
 - Provide hints for interpretability
 - Implementations are available
- There are more mathematical/statistical tools than end-user bioinformatical solutions: Opportunity for developers

Acknowledgements



Thanks For your Attention



REFERENCES I

- Canuel, V., Rance, B., Avillach, P., Degoulet, P., and Burgun, A. (2015). Translational research platforms integrating clinical and omics data: a review of publicly available solutions. *Briefings in Bioinformatics*, 16(2):280–290.
- Cao, K.-A. L., Martin, P. G., Robert-Grani, C., and Besse, P. (2009). Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics*, 10(1):34. 00098.
- Dittrich, M. T., Klau, G. W., Rosenwald, A., Dandekar, T., and Mller, T. (2008). Identifying functional modules in proteinprotein interaction networks: an integrated exact approach. *Bioinformatics*, 24(13):i223–i231.
- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., and Schultz, N. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science signaling*, 6(269):pl1.
- Gastinel, L. N. (2012). Principal Component Analysis in the Era of Omics Data.

REFERENCES II

- Gligorijevi, V. and Prulj, N. (2015). Methods for biological data integration: perspectives and challenges. *Journal of the Royal Society, Interface*, 12(112).
- Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., Ballestar, E., Bongcam-Rudloff, E., Conesa, A., and Tegnér, J. (2014). Data integration in the era of omics: current and future challenges. *BMC Systems Biology*, 8(Suppl 2):i1.
- Hamid, J. S., Hu, P., Roslin, N. M., Ling, V., Greenwood, C. M. T., and Beyene, J. (2009). Data Integration in Genetics and Genomics: Methods and Challenges. *Human Genomics and Proteomics*, 1(1). 00024.
- Ideker, T., Ozier, O., Schwikowski, B., and Siegel, A. F. (2002). Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics (Oxford, England)*, 18 Suppl 1:S233–240.
- Joyce, A. R. and Palsson, B. . (2006). The model organism as a system: integrating 'omics' data sets. *Nature Reviews. Molecular Cell Biology*, 7(3):198–210.

REFERENCES III

- Liu, Y. and Qiu, P. (2012). Integrative analysis of methylation and gene expression data in TCGA. In *Proceedings 2012 IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS)*, pages 1–4.
- Meng, C., Zeleznik, O. A., Thallinger, G. G., Kuster, B., Gholami, A. M., and Culhane, A. C. (2016). Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in Bioinformatics*, 17(4):628–641.
- Muniategui, A., Pey, J., Planes, F. J., and Rubio, A. (2013). Joint analysis of miRNA and mRNA expression data. *Briefings in Bioinformatics*, 14(3):263–278.
- Paraskevopoulou, M. D., Georgakilas, G., Kostoulas, N., Vlachos, I. S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T., and Hatzigeorgiou, A. G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Research*, 41(Web Server issue):W169–173.

REFERENCES IV

- Sales, G., Coppe, A., Bisognin, A., Biasiolo, M., Bortoluzzi, S., and Romualdi, C. (2010). MAGIA, a web-based tool for miRNA and Genes Integrated Analysis. *Nucleic Acids Research*, 38(Web Server issue):W352–W359.
- Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., Miller, C. R., Ding, L., Golub, T., Mesirov, J. P., Alexe, G., Lawrence, M., O'Kelly, M., Tamayo, P., Weir, B. A., Gabriel, S., Winckler, W., Gupta, S., Jakkula, L., Feiler, H. S., Hodgson, J. G., James, C. D., Sarkaria, J. N., Brennan, C., Kahn, A., Spellman, P. T., Wilson, R. K., Speed, T. P., Gray, J. W., Meyerson, M., Getz, G., Perou, C. M., Hayes, D. N., and Cancer Genome Atlas Research Network (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, 17(1):98–110.