SARA BALLOUZ, PH. D.

Cold Spring Harbor Laboratory
Stanley Institute for Cognitive Genomics
1 Bungtown Road
Cold Spring Harbor, NY 11724, USA

sballouz@cshl.edu
Tel: +1(516) 422 4128

AREAS OF INTEREST

Disease genetics, complex diseases, common and rare variation, rare disorders, cognitive disorders, developmental disorders, cardiovascular disorders, X-linked disorders, X-inactivation, X-skew, transcriptomics and functional genomics, microarray, RNA-sequencing, single-cell RNA-sequencing, model organisms, robustness, method development, differential expression, co-expression, differential co-expression.

RESEARCH AIMS AND GOALS

My central scientific interest has been to understand the genetic architecture of disease. With data from the genome, transcriptome, epigenome and proteome increasing exponentially, robust tools and practices need to be established to analyse this deluge, in particular if to be applied to personalized medicine. My research goal is to develop methods to analyse transcriptomic data and integrate it with other "omics" data to discover what is common and distinct among diseases and disorders, hence allowing an understanding of the basic biology and eventual translation into focused treatments. I believe the use of meta-analysis, data mining and machine learning tools, with a strong focus on quality control, reproducibility and statistical robustness is the best way to leverage the current expanding and untapped data to inform the n=1.

QUALIFICATIONS

Institution	Degree	Year	Field of study
University Of New South Wales	BSc/ BEng (Honours, First Class)	10/2008	Genetics/ Bioinformatics
Victor Chang Cardiac Research Institute and the University Of New South Wales	PhD	01/2013	Bioinformatics

CURRENT APPOINTMENT

2013 – present Postdoctoral Fellow, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
Advisor: Jesse Gillis

RELATED PROFESSIONAL EXPERIENCE

2018 **Postdoc R bootcamp instructor,**
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

2018 **Undergraduate Research Program R bootcamp instructor,**
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

2014-2017 **Postdoc Liaison Committee Member**,
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

2012 **Guest Lecturer**, Computing for Engineers, School of Computer Science,
University of New South Wales, Sydney, NSW, Australia

2009 – 2012 **Demonstrator**, Computing for Engineers, School of Computer Science,
University of New South Wales, Sydney, NSW, Australia

2009 **Demonstrator**, Human Computer Interaction, School of Computer Science and Engineering,
University of New South Wales, Sydney, NSW, Australia

2008 **Honours Student**, School of Biological Sciences,
University of New South Wales, Sydney, NSW, Australia

2006 – 2008 **Student Research Assistant**,
Victor Chang Cardiac Research Institute, Sydney, NSW, Australia

2005 **Mentor,** UNSW Primary Girls Science and Technology Workshops,
University of New South Wales, Sydney, NSW, Australia

CURRICULUM VITAE

SARA BALLOUZ, PH. D.

PUBLICATIONS SUMMARY

H-index	10	Source: Google Scholar
Citations	312	IF: Impact factor at year of publication

PUBLICATIONS

Original research articles	**IF**	**Citations**
1. **Ballouz, S.**, Dobin, A., Gingeras, T.R., Gillis, J. (2018). *The fractured landscape of RNA-seq alignment: The default in our STARs*. **Nucleic Acids Research.** gky325	10.162	2
2. Crow, M., Paul, A., **Ballouz, S.**, Huang, Z.J., Gillis, J. (2018). *Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor*. **Nature Communications.** 9 (1), 884.	12.124	16
3. **Ballouz, S.**, and Gillis, J. (2017). *Strength of functional signature correlates with effect size in autism*. **Genome Medicine**. 9, 64	7.071	2
4. **Ballouz, S.**, Weber, M., Pavlidis, P., Gillis, J. (2017). *EGAD: ultra-fast functional analysis of gene networks*. **Bioinformatics,** 33, 612-614.	7.307	10
5. **Ballouz, S.**, Pavlidis, P., Gillis, J. (2017). *Using predictive specificity to determine when gene set analysis is biologically meaningful*. **Nucleic Acids Research** 45, e20.	10.162	7
6. O'Meara, M.J.*, **Ballouz, S.***, Shoichet, B.K., Gillis, J. (2016). *Ligand Similarity Complements Sequence, Physical Interaction, and Co-Expression for Gene Function Prediction*. **PloS One** 11, e0160098.	2.806	8
7. Crow, M., Paul, A., **Ballouz, S.**, Huang, Z.J., Gillis, J. (2016). *Exploiting single-cell expression to characterize co-expression replicability*. **Genome Biology** 17, 101.	11.908	18
8. **Ballouz, S.**, Gillis J. (2016). *AuPairWise: A Method to Estimate RNA-Seq Replicability through Co-expression.* **PLoS Computational Biology.** 12(4): e1004868.	4.542	4
9. Verleyen, W., **Ballouz, S.**, Gillis, J. (2016). *Positive and negative forms of replicability in gene network analysis.* **Bioinformatics.** 32 (7): 1065-1073.	7.307	8
10. **Ballouz, S.**, Gillis, J. (2015). *Guidance for RNA-seq co-expression network construction and analysis: safety in numbers*. **Bioinformatics.** 31 (13): 2123-2130.	5.766	73
11. Verleyen, W., **Ballouz, S.**, Gillis, J. (2015). *Measuring the wisdom of the crowds in network-based gene function inference*. **Bioinformatics.** 31(5):745-52.	5.766	12
12. Grover, M.P., **Ballouz, S.**, Mohanasundaram, K.A., George, R.A., Goscinski, A., Crowley, T.M., Sherman, C.D., Wouters, M.A. (2015). *Novel therapeutics for coronary artery disease from genome-wide association study data.* **BMC Medical Genomics**. 8(Suppl 2):S1.	2.726	8
13. Gillis, J., **Ballouz, S.**, Pavlidis, P. (2014). *Bias tradeoffs in the creation and analysis of protein-protein interaction networks*. **Journal of Proteomics**. 100:44-54.	3.888	30
14. Grover, M.P., **Ballouz, S.**, Mohanasundaram, K.A., George, RA, Sherman, CD, Crowley, T.M. and Wouters, M.A. (2014). *Identification of novel therapeutics for complex diseases from genome-wide association data.* **BMC Medical Genomics**, 7(1):1.	2.873	24
15. **Ballouz, S.**, Liu, J.Y., Oti, M., Gaeta, B., Fatkin D., Bahlo M., Wouters, M.A. (2014). *Candidate disease gene prediction using Gentrepid: application to a genome-wide association study on coronary artery disease*. **Molecular Genetics & Genomic Medicine.** 2:44-57.	2.728	9
16. **Ballouz, S.**, Liu, J.Y., George, R.A., Bains, N., Liu, A., Oti, M., Gaeta, B., Fatkin, D., Wouters, M.A. (2013). *Gentrepid V2. 0: a web server for candidate disease gene prediction*. **BMC Bioinformatics.** 14:1-9.	2.672	6
17. **Ballouz, S.**, Liu, J.Y., Oti, M., Gaeta, B., Fatkin, D., Bahlo, M., Wouters, M.A. (2011). *Analysis of genome-wide association study data using the protein knowledge base.* **BMC Genetics.** 12(1):98.	2.475	11
18. **Ballouz, S.**, Francis, A.R., Lan, R., Tanaka, M.M. (2010). *Conditions for the evolution of gene clusters in bacterial genomes*. **PLoS Computational Biology**. 6(2):e1000672.	5.515	29

19. Teber, E., Liu, J.Y., **Ballouz, S.**, Fatkin, D., Wouters, M.A. (2009). *Comparison of automated candidate gene prediction systems using genes implicated in type 2 diabetes by genome-wide association studies.* **BMC Bioinformatics.** 10 (Suppl 1):S69. 3.428 18

Book chapters

20. Oti, M., **Ballouz, S.**, Wouters, M.A. (2011). *Web tools for the prioritization of candidate disease genes.* **Methods in Molecular Biology** (Clifton, NJ). 760:189. 21

In review

21. **Ballouz, S.**, Doerfel, M., Crain, J., Crow, M., Faivre, L., Keegan, C.E., Kitsiou-Tzeli, S., Tzetis, M., Lyon, G.J., Gillis, J. (2017*). Not by systems alone: identifying functional outliers in rare disease pedigrees*. bioRxiv, 128439

SCIENTIFIC PRESENTATIONS

Ballouz, S., Doerfel, M., Crain, J., Crow, M., Lyon, G.J., Gillis, J. (2018). *Not by systems alone: Identifying functional outliers in rare disease pedigrees*
Poster presentation at **ASHG 2018**, San Diego, CA, USA.

Ballouz, S., Gillis, J. (2018). *The fractured landscape of RNA-seq alignment: the default in our STARs.*
Poster presentation at **ISMB 2018**, Chicago, IL, USA.

Ballouz, S., Gillis, J. (2018). *Conservation of transcriptional variation across human, mouse and … armadillo?!*
Oral presentation at **Lorne Genome 2018**, Lorne, VIC, Australia.
Poster presentation at **Systems Biology: Global Regulation of Gene Expression 2018,** CSHL, NY, USA.

Ballouz, S., Doerfel, M., Crain, J., Crow, M., Lyon, G.J., Gillis, J. (2017). *Going rogue: outlier gene expression drives rare disease in the TAF1 syndrome cohort*.
Oral presentation at **Systems Biology: Global Regulation of Gene Expression 2017**, CSHL, NY, USA.
Poster presentation at **Society for Neuroscience 2017**, Washington, DC, USA.

Ballouz, S., Gillis, J. (2016). *Assessment of functional convergence across study designs in autism*.
Poster presentation at **The Biology of Genomes 2016**, CSHL, NY, USA.

Ballouz, S., Gillis, J. (2015). *Distinguishing biological from technological signals in the functional interpretation of neuropsychiatric disease genes.*
Poster presentation at **World Congress of Psychiatric Genetics 2015**, Toronto, Canada.

Ballouz, S., Gillis, J. *AuPairWise: biologically focused RNA-seq quality control using co-expression.*
Poster presentation at **Genome Informatics 2015**, CSHL, NY, USA.
Poster presentation at S**ystems Biology: Global Regulation of Gene Expression 2016**, CSHL, NY, USA.
Poster presentation at **ISMB 2016**, Ontario, FL, USA.

Ballouz, S., Gillis, J. (2014). *Guidance for RNA-seq co-expression network construction and analysis*
Poster presentation at **ISMB 2014**, Boston, MA, USA.
Poster presentation at **Biological Data Science 2014**, CSHL, NY, USA.

Ballouz, S., Gillis, J. (2014). *Inferring function using non-coding RNA co-expression networks*.
Poster presentation at **Systems Biology: Global Regulation of Gene Expression 2014,** CSHL, NY, USA.

Ballouz, S., Gillis, J. (2013). *RNA-seq co-expression network meta-analysis*.
Poster presentation at **Genome Informatics 2013**, CSHL, NY, USA.

Ballouz, S., Liu, J.Y., Oti, M., Gaeta, B., Fatkin, D., Bahlo, M., Wouters, M.A. *A holistic approach to studying complex diseases using Gentrepid.*
Poster presentation at **Human Genome Meeting (HGM) 2012**. Sydney, Australia.
Poster presentation at **12th International Federation of Human Genetics Congress 2011**. Combined International Congress of Human Genetics (ICHG) and The American Society of Human Genetics (ASHG) meeting. Montreal, Canada.

Ballouz, S., Liu, J.Y., Oti, M., Gaeta, B., Fatkin, D., Bahlo, M., Wouters, M.A. (2012). *Candidate disease gene prediction for complex diseases using Gentrepid.*

Poster presentation at **33rd Lorne Genome Conference 2012.** Lorne, Australia.

Ballouz, S., Liu, J.Y., George, R., Bains, N.P., Liu, A., Oti, M., Gaeta, B., Fatkin, D., Wouters, M.A. (2011). *Gentrepid: a web server for candidate disease gene prediction.*

Oral presentation at **Asia Pacific Bioinformatics Network's 10th InCoB - 1st ISCB Asia Joint Conference 2011 (InCoB/ISCB-Asia).** Kuala Lampur, Malaysia.

Ballouz, S., Liu, J.Y., Oti M, Gaeta, B., Sparrow, D., Dunwoodie, S., Wouters, M.A. (2011). *Mouse Gentrepid, a webserver tool for candidate gene prediction in mice and humans.*

Oral presentation at **1st Mouse Genetics conference.** Combined International Mammalian Genome Society (IMGS), Complex Trait Community (CTC) and the Genetics Society of America (GSA) meeting. Washington DC, USA.

Ballouz, S., Liu, J.Y., Oti M, Gaeta, B., Fatkin, D., Bahlo, M., Wouters, M.A. (2011). *Employment of the protein knowledge base for data analysis in genome-wide association studies.*

Oral presentation at **8th GeneMappers 2011.** Hobart, Australia.

Poster presentation at **32nd Lorne Genome Conference 2011.** Lorne, Australia.

Ballouz, S., Liu, J.Y., Oti, M., Gaeta, B., Fatkin, D., Bahlo, M., Wouters, M.A. *Application of the Gentrepid candidate gene prediction system to Genome-Wide Association Studies.*

Poster presentation at **BioInfoSummer 2010.** Melbourne, Australia.

Oral presentation at **Genetics Society of AustralAsia (GSA) 2010.** Canberra, Australia.

Poster presentation at **19th St Vincent's & Mater Health Sydney Research Symposium 2009.** Sydney, Australia.

Teber, E.T., Liu, J.Y., **Ballouz, S.,** Fatkin, D., Wouters, M.A. (2009). *Comparison of automated candidate gene prediction systems using genes implicated in type 2 diabetes by genome-wide association studies.*

Oral presentation at **Genetics Society of AustralAsia (GSA) 2009.** Brisbane, Australia

INVITED TALKS

2018	Solving for X: gene co-expression networks for X-linked disease analysis	Pasteur Institute, Paris, France
2018	Rare disease and identical quadruplets as models for transcriptional identity	Monash - Bioinformatics Seminar Series, Melbourne, VIC, Australia
2018	Rare disease and identical quadruplets as models for transcriptional identity	Victor Chang Cardiac Research Institute - Special Seminar, Sydney, NSW, Australia

GRANTS AND AWARDS

2018-2019	AWS Cloud Credits for Research grant
2014	ISMB Travel Award
2009-2011	Australian Postgraduate Award, UNSW
2009-2011	Supplementary Engineering Award, UNSW

PEER REVIEW EXPERIENCE

Journals Bioinformatics, Gene, Nature Communications, eLife, Journal of Theoretical Biology, BMC Systems Biology, BioData Mining, Journal of Molecular Neuroscience (4/year)

STUDENT SUPERVISION

Mentoring of 2 rotation students and 2 undergraduate research students (2014-2018)

CURRENT PROFESSIONAL MEMBERSHIPS

2017	Society for Neuroscience
2011-	American Society of Human Genetics
2011-	International Society for Computational Biology