Darya **Vanichkina**

PhD, Genomics and Bioinformatics · Specialist, Molecular Biology and Biochemistry

(+61) 420 - 889 - 939 | ✉ d.vanichkina@gmail.com | 🏠 daryavanichkina.com | 🐙 dvanic | 💼 daryavanichkina | 🐦 @dvanichkina | 📞 darya_van

Summary

Data scientist with a biology background and experience in big data, machine learning, and statistics. Passionate about explaining and teaching quantitative skills, programming and reproducible research methodologies to other researchers and non-technical audiences.

Currently a genomics and bioinformatics postdoctoral research officer at the Centenary Institute of Cancer Medicine and Cell Biology investigating RNA regulation in cancer and normal biology.

Interested in working on new analytics challenges in genomics, bioinformatics, medical & health science, as well as those outside of the life sciences domain.

Work Experience

Centenary Institute of Cancer Medicine and Cell Biology
Sydney, Australia

Postdoctoral research officer (supervisor: Professor John Rasko)   *Oct. 2015 - PRESENT*

- Investigating publicly available dataset of Glioblastoma Multiforme patient samples to assess whether intron retention events could be used to predict cancer grade or treatment outcomes.
- Carrying out tool and methodology development to use long read sequencing for intron retention detection.
- Implemented and deployed Juicer and Juicebox HiC analytics pipeline on Amazon AWS resources to analyse chromatin conformation data to assess whether mutations leading to changes in levels of a specific protein cause changes in chromatin interactions.
- Established RNA sequencing analysis pipeline using Artemis HPC. Was awarded and manage Artemis Grand Challenge allocation. Train students and other staff to take advantage of available resources.

Institute of Molecular Bioscience, University of Queensland
Brisbane, Australia

PhD student (supervisors: Dr Ryan Taft, Prof John Mattick, Dr Cas Simons, A/Prof Ernst Wolvetang)   *Dec. 2010 - Mar. 2016*

- Characterised the mouse and human trascriptome in response to neuronal depolarisation using microarrays, identifying *Gomafu* as a ncRNA with a critical role in normal activity altered in schizophrenia.
- Characterised the transcriptome of stem cells derived from Ataxia Telangiectasia patients and cells differentiated into cerebellar-like neurons to reveal recapitulation of aspects of the disease state in only the differentiated cells.
- Developed a novel method of identifying the targets of miRNA, a regulatory small RNA class, which demonstrated a ~10x improvement over previously existing similar approaches.

Karolinska Institutet
Stockholm, Sweden

Visiting researcher (supervisor: A/Prof Goncalo Castelo-Branco)   *Sep. 2014 - Dec. 2014*

- Awarded a competitive Boehrinher Ingelheim Travel Grant to carry out a collaborative project at the Department of Medical Biochemistry and Biophysics.
- Compared the mouse oligodendrocyte precursor cell transcriptome from the brain and spinal cord at two developmental stages to assess whether there were differences in cell differentiation capabilities.
- Reanalysing publically available microarray data demonstrated that a previous study using a cuprizone OPC activation model was erroneous, as a result of normalisation challenges, in postulating that adult OPCs were similar to oligodendrocytes until activation, whereby they became similar to neonatal OPC, and instead adult oligodendrocytes were being specifically affected by drug treatment.

Laboratory of Structure and Function of Human Genes, Shemyakin-Ovchinnikov Institute
Moscow, Russia

Senior (Masters) Thesis (supervisor: Dr Tatyana Azhikina)   *Sep. 2009 - Sep. 2010*

- Using RT-qPCR demonstrated that expression of human PIWIL homologues is lower in various types of germ cell tumours than in normal tissue taken from the same patient, contradicting studies describing the overexpression of these proteins specifically in most seminomas as compared to NSGCT.
- Conducted the first study of PIWIL expression in human embryonic material, demonstrating that these genes are transcribed at specific stages during lung and esophageal development, at much higher levels than characteristic of the respective adult tissues.

Department of Molecular Biology, Lomonosov Moscow State University
Moscow, Russia

Junior (Undergraduate) Thesis (supervisor: Dr Tatyana Azhikina)   *Apr. 2008 - Aug. 2009*

- Characterised a novel transcript, cell type specific transcript from *chr19* using Step-out RACE and bioinformatic analyses, showing a correlation between chromatin conformation, DNA methylation and transcription.

Department of Molecular Biology, Lomonosov Moscow State University *Moscow, Russia*

Research Assistant (supervisor: Dr Tatyana Kalebina) *Sep. 2004 - Dec. 2007*

- Using electron, fluorescent and confocal microscopy techniques visualized DNA transfer through the cell wall in pre- apoptotic yeast cells under the influence of amiodarone.
- Assisted senior faculty in conducting lectures and seminars, helped prepare grant applications and reports, and was responsible for reagent and equipment purchasing, book keeping, and inventory management.

Skills

General

- **Designing, running & coordinating** independent and collaborative research projects.
- **Programming languages**: R (including the tidyverse, data.table, openxlsx, RColorBrewer, shiny, knitr, e1071, randomForest, caret, glmnet, ROCR & other libraries), python (web scraping, regex, NumPy, SciPy, pandas, Jupyter notebooks, networkX, scikit-learn, sqlite3), SQL, MongoDB, version control using git, UNIX shell scripting, HTML, CSS & LaTeX.
- **HPC**: Developing analytics pipelines on university research clusters (PBS-Pro based: Artemis, Karolinska's UPPMAX, UQ's Barrine, NECTAR etc) and Amazon AWS (EC2, S3, Glacier).
- **Complex data integration**: Acquisition, analysis & integration of large datasets, including 250+ cancer patient sample DNA, RNA, epigenetic information and metadata from TCGA.
- **Data visualisation**: ggplot2, lattice & base graphics in R; matplotlib & seaborn in python.
- **Data mining**: Twitter, PubMed, SRA, DDBJ, ENA.
- **Machine Learning & Statistics**: classic descriptive and inferential statistics, supervised & unsupervised learning approaches.
- **Research data management**: Assessing research data management requirements & preparing RDMPs. Submission of data to GEO.
- **Communication**: Engaging with researchers without backgrounds in data science to design experiments that allow questions of interest to be answered in a statistically valid manner. Participating in Hacky hours and ResBaz events.
- **Writing**: Preparing manuscripts collaboratively for publication in peer-reviewed journals & contributing to grant applications, including developing budgets, ethics and compliance checks; preparing risk assessments.
- **Other**: Jekyll & github pages for lab and project landing pages; Adobe Photoshop, Illustrator & InDesign for image and poster creation, editing & publishing; electronic lab notebook use and maintenance.

Biology-specific

- **RNA sequencing data analysis**: differential gene & isoform expression analysis, alternative splicing, novel transcript annotation, repeat expression, RNA editing investigation, small RNA expression.
- **Microarray data analysis**: differential gene expression and genotyping.
- **Functional analysis**: single enrichment, gene-set based & ontology analysis.
- **Biostatistics**: NGS and microarray normalisation, regression, multiple testing correction.
- **Integrative genomics analysis**: variant calling, chromatin conformation (HiC) & ChIP-Seq analysis.
- **Systems biology and biological networks analysis:** Cytoscape, Metacore GeneGO & Ingenuity Pathway Analysis.
- **Data visualisation**: UCSC, Ensembl and ECR genome browsers.
- **Bioconductor (R) packages**: limma, DESeq2, edgeR, topGO, GenomicRanges, rtracklayer, rsamtools, biomaRt, GenomicFeatures, affy & others.
- **NGS analysis software**: FastQC, samtools, bowtie & bowtie2, STAR, bwa, BLAST,BLAT, IGV, Novoalign, TopHat2, Cufflinks, MATS, MISO, Monocle, eXpress, RSEM, salmon, kallisto, sailfish, HTSeq-count, featureCounts, GATK, Picard, Trinity, Iso-Seq, Quiver.
- **GUI tools**: Generating Galaxy workflows and using CLC Genomics for small analytics tasks.
- **Molecular biology techniques**: nucleic acid and protein purification, RT end point & quantitative PCR, RACE, Western & Northern blotting, NGS library preparation, cell culture & molecular cloning, *in situ* hybridization, immunofluorescence & immunohistochemistry.
- **Microscopy**: brightfield, confocal and electron - sample preparation, imaging and image analysis.
- **Ecology**: analysis of population and distribution data for small mammals and plants, to assess species distribution, ecosystem utilisation and status.

Education

University of Queensland *St. Lucia, Australia*

PhD Genomics and Bioinformatics *Dec. 2010 - Mar. 2016*

- Thesis: *Marvellous complexity: investigating the transcriptional landscape of the mammalian nervous system using RNA sequencing*.
  Supervised by Prof John Mattick, Dr Guy Barry, Dr Ryan Taft, A/Prof Ernst Wolvetang and Dr Cas Simons.

Lomonosov Moscow State University *Moscow, Russia*

Specialist degree in Biochemistry with a major in Molecular Biology *Sep. 2005 - Jun. 2010*

- Degree conferred with Highest Honours.

Teaching, community outreach and other roles

Journal of Open Research Software   *London, UK*

ASSOCIATE EDITOR   *Feb. 2016 - PRESENT*

- Identify and work with submission reviewers and assess suitability of manuscripts & licences submitted to JORS across life sciences and chemistry .

Software and Data Carpentry   *Australia*

INSTRUCTOR & LESSON MAINTAINER   *Jun. 2015 - PRESENT*

- Serve as instructor or lead instructor for R, python, OpenRefine, shell, SQL, & git course materials, teaching reproducible research and automation techniques to scientists across all disciplines and career stages in Australia.
- Maintain Data Carpentry cloud computing for genomics lesson, coordinating the contributions of the community.

School of Chemistry and Molecular Bioscience, University of Queensland   *St Lucia, Australia*

TUTOR, LEAD TUTOR & COURSE DEVELOPER   *Feb. 2011 - Jul. 2015*

- Delivered wet and dry lab seminars and practicals for undergraduate and masters-level bioinformatics, genetics, microbiology, cell and molecular biology courses (MICR3003, MICR3004, BIOL2200, BIOL2202, BIOL3004).
- Supervised other tutors for Cell & Molecular biology (BIOL2200), Genomics and Bioinformatics (BIOL3004) and Microbiology (MICR3004) courses.
- Developed flipped learning materials (YouTube videos) for Genomics and Bioinformatics (BIOL3004) course, including modules on CLC Genomics use.

Institute for Molecular Bioscience, University of Queensland   *St Lucia, Australia*

IMB SCIENCE AMBASSADOR   *Aug. 2011 - Jul. 2015*

- Represented the IMB to visitors, including school and community groups, visiting academic, industry and government dignitaries.
- Was involved in organisation and coordination of institute public outreach events.

Australian Academy of Technology and Engineering   *Queensland, Australia*

ATSE WONDER OF SCIENCE YOUNG SCIENCE AMBASSADOR   *2014*

- Visited primary schools in rural Queensland to encourage students to pursue study in STEM.
- Delivered class lessons for these students in line with the curriculum to engage them with the content taught.

Students of the Institute of Molecular Bioscience Student Association (SIMBA)   *Queensland, Australia*

PRESIDENT   *Aug. 2011 - Aug. 2012*

- Led a team of 6 postgraduate students involved in organising education, social and interdepartmental events at the IMB.
- Organised 1-2 events per month, including catering to over 150 staff and students for fundraisers and trivia nights.
- Served as a liaison between undergraduate, masters and PhD students and the Institute's executive on matters of WH&S, candidature issues and other concerns.

Club of Young Biologists, Moscow Zoo & Club of Young Biologists, Zvenigorod Biological Station, Lomonosov Moscow State University   *Moscow, Russia*

MENTOR & ADVISOR   *Sep. 2004 - Dec. 2010*

- Supervised middle and high school students in carrying out ecological, zoological and botanical research at the Zvenigorod biological station, Briansky Les & Central Forest National Parks.
- Developed and delivered an introductory cell and molecular biology course for club students without a background in chemistry.
- Served as examiner in cell biology for club olympiads.
- Mentored students in the clubs to pursue higher education and STEM careers.

Honours & Awards

2014	**Graduate Teaching Assistant (GTA)**, University of Queensland	*St Lucia, QLD*
2014	**Boehringer Ingelheim Travel Grant**, Boehringer Ingelheim Fond	*Germany*
2014	**Student Poster Prize**, 35th Lorne Genome Conference	*Lorne, Victoria*
2013	**1st place, 3 minute thesis competition**, Institute for Molecular Bioscience, University of Queensland	*St Lucia, QLD*
2011	**ANZ Trustees Scholarship for Medical Research**, ANZ Bank	*St Lucia, QLD*
2009	**R.B. Khesin Award for Outstanding Junior Thesis**, Department of Molecular Biology, Lomonosov Moscow State University	*Moscow, Russia*

Publications

- Vanichkina DP, Schmitz U, Wong JJ, Rasko JEJ (2018) **Challenges in defining the role of intron retention in normal biology and disease.** *Seminars in Cell and Developmental Biology* 75:40-49
- Vanichkina DP, Choudhary A, Ender C, Crawford J, Baillie G, Calcino A, Ru K, Taft RJ (2017) **Identification of miR-29b targets using 3-cyanovinylcarbazole containing mimics.** *RNA* doi: 10.1261/rna.064923.117
- Vanichkina DP, Marques S, van Bruggen D, Floriddia E, Munguba H, Vareimo L, Giacomello S, Falcao AM, Meijer M, Samudyata, Codeluppi S, Bjorklund A, Linnarsson S, Hjerling-Leffler J, Taft RJ, Castelo-Branco G (2017) **Single-cell transcriptomic profiling of progenitors of the oligo-dendrocyte lineage reveals transcriptional convergence during development.** *bioarXiv* doi: 10.1101/186445
- Nayler SP, Powell JE,3, Vanichkina DP, Korn O, Wells CA, Kanjhan R, Sun J, Taft RJ,, Lavin MF, Wolvetang EJ. (2017) **Human iPSC-Derived Cere-bellar Neurons from a Patient with Ataxia-Telangiectasia Reveal Disrupted Gene Regulatory Networks** *Frontiers in Cellular Neuroscience* 11:327
- Kalebina TS, Sokolov SS, Selyakh IO, Vanichkina DP, Severin FF (2015) **Amiodarone induces cell wall channel formation in yeast Hansenula polymorpha** *Springerplus* 4:453
- Barry G, Briggs JA, Vanichkina DP, Poth EM, Beveridge NJ, Ratnu VS, Nayler SP, Nones K, Hu J, Bredy TW, Nakagawa S, Rigo F, Taft RJ, Cairns MJ, Blackshaw S, Wolvetang EJ, Mattick JS (2014). **The long non-coding RNA *Gomafu* is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing.** *Molecular psychiatry* 19, pp. 486 - 494
- Lybaek H, de Bruijn D, Engelsman-van Dijk den AH, Vanichkina D, Nepal C, Brendehaug A, and Houge G (2013) **RevSex duplication-induced and sex-related differences in the SOX9 regulatory region chromatin landscape in human fibroblasts.** *Epigenetics* 9(**3**) pp. 416-427

Conference Presentations

2018	**Training the trainer: Software Carpentry Instructor Training,** ANDS-NECTAR-RDS Workshop: Powering up your 2018 (data skills) training	*Sydney, NSW*
2016	**Big Data Biology for Pythonistas: Getting in on the Genomics Revolution,** PyConAu	*Melbourne, Victoria*
2016	**Temporal and developmental dynamics of the oligodendrocyte precursor transcriptome,** Sydney Bioinformatics Research Symposium	*Sydney, NSW*
2016	**Temporal and developmental dynamics of the oligodendrocyte precursor transcriptome,** 37th Lorne Genome Conference	*Lorne, Victoria*
2014	**Ataxia-telangiectasia iPS-derived neuronal transcriptome,** 35th Lorne Genome Conference, *Poster prize winner*	*Lorne, Victoria*
2012	**Activity-dependent transcriptional dynamics in mouse cortical and human iPS - derived neurons,** 33rd Lorne Genome Conference	*Lorne, Victoria*
2010	**A Novel Non-coding Human Transcript from Locus ATP4A-HAUS5 (Chr19),** 14th International Pushchino School-Conference for Young Scientists "Biology- Science of the XXI century"	*Puschino, Russia*
2010	**A Novel Non-coding Human Transcript from Locus ATP4A-HAUS5 (Chr19),** XVII International student, postgraduate and young scientist conference "Lomonosov"	*Moscow, Russia*
2010	**Characterization of a Novel Transcript from Human Locus ATP4A-KIAA0841 (Chr19),** XXII Winter School for Young Scientists"Promising Areas of Physico-chemical Biology and Biotechnology"	*Moscow, Russia*