BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES**

NAME: Andrey Grigoriev

eRA COMMONS USER NAME (credential, e.g., agency login): AGRIGORIEV

POSITION TITLE: Professor; Dept of Biology and Center for Computational and Integrative Biology, Rutgers University

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Engineering Physics Institute, Moscow, USSR	M.Sc.	1986	Physics
VNIIGenetika, Moscow, USSR	Ph.D.	1995	Molecular Biology
Max-Planck Institute for Molecular Genetics Berlin, Germany	Postdoctoral training	1995-98	Computational Biology, Genomics

A. Personal Statement

I have the training, expertise, leadership and motivation required to successfully execute my responsibilities in the proposed research project. I have worked in the field of computational biology since 1987. I bring to this project my skills in algorithm development, including machine learning (from genome analysis to computer game development) research, experience with large-scale genome studies, sequence analysis and connecting genomics, transcriptomics, proteomics and cancer, and most recently, small RNA, as well as significant current graduate teaching/advising effort and extensive project management experience. Both in academic and in industrial settings, I led a large number of successful projects, collaborated with other researchers in a dozen of countries, and produced several peer-reviewed publications from each project (or comparable industrial deliverables since publishing activity is often restricted in the industry). After 12 years in biotech industry, I am very familiar with multiple aspects of target and drug development in pharma. Based on this extensive previous experience, I am aware of the importance of frequent communication among project members and of having a realistic research plan, timeline, and budget. Upon transition from the industry to academia in 2009-2010, I launched a new PhD program "Computational and Integrative Biology" (CIB) at Rutgers-Camden Center for Computational and Integrative Biology (CCIB) as its first Director (in 2010). By now >60 MS/PhD students have been enrolled into the program. Five of these students have graduated from my lab with PhD and continue to work in bioinformatics research (in Columbia U, UPenn, Children's Hospital of Philadelphia, Purdue, USDA). Members of my current laboratory range from undergraduate to postdoc level, with a number of MS and PhD graduate students. At Rutgers I have received grants from the NIH, NSF (including research, equipment and commercialization grants) and other sponsors and served as a panelist for the NSF and the NIH.

B. Positions and Honors

2010-present Professor, Biology Department, Rutgers University-Camden, USA
2008-2009 Principal Consultant, SystemsBio Consulting, Munich, Germany
2001-2008 Scientific Director, Bioinformatics, GPC Biotech, Martinsried, Germany
1998-2001 Group leader, Bioinformatics, GPC Biotech, Martinsried, Germany
1995-1998 Scientist, Max-Planck-Inst. for Molecular Genetics, Berlin, Germany
1992-1995 Higher Executive Officer, Imperial Cancer Research Fund, London, UK
1986-1992 Scientist, VNIIGenetika, Moscow, USSR

Awards/Honors: Winner of 4 gold, 2 silver and 2 bronze medals for strategy game algorithms at the Computer Olympiads, London, UK and Maastricht, Netherlands (1989-1991)
2018 Audience award for the best presentation at the NSF iCorp grant meeting
2018-2019 Arts and Sciences Annual Faculty Fellowship (on small RNA), Rutgers-Camden
2020 Rutgers Open and Affordable Textbooks Award
2022 Chancellor's Award for Research Excellence
2022 Keynote talk at Oxford Global conference "NextGen Omics US", Boston, USA

Editorial Engagements: Guest Editor of Special Issue "Functional Roles of Short Fragments of RNA" of International Journal of Molecular Sciences (ongoing).

C. Contribution to Science

My work in computational biology over 35 years is summarized in 5 sections with the most relevant publications that I co-authored in each section. These sections are organized chronologically so the most recent work (and most relevant for the small RNA proposal) is at the end. This is less than half of my papers. A larger, but still incomplete **List of Published Work in MyBibliography** (not all of my papers/books are in PubMed, including those on various computational aspects, such as artificial intelligence in gaming): http://www.ncbi.nlm.nih.gov/sites/myncbi/andrey.grigoriev.1/bibliography/47382148/public/?sort=date&direction=ascending

1. **Genome mapping**. As a graduate student, I contributed to several early high-profile genome mapping projects (a-c), developing and utilizing the first open source publicly available software package for physical mapping that I co-authored with R. Mott (d,e) in the group of Hans Lehrach.

a. Hoheisel, J. D., Maier, E., Mott, R. F., McCarthy, L., Grigoriev, A. V., Schalkwyk, L. C., Nizetic, D., Francis, F. and Lehrach, H. (1993) High resolution cosmid and P1 maps spanning the 14Mb genome of the fission yeast Schizosaccharomyces pombe. **Cell** 73, 109-120.
b. Maier, E., Hoheisel, J. D., McCarthy, L., Mott, R. F., Grigoriev, A. V., Monaco, A. P., Larin, Z. and Lehrach, H. (1992) Complete coverage of the Schizosaccharomyces pombe genome in yeast artificial chromosomes. **Nature Genetics** 1, 273-277.
c. Crollius, H.R., Ross, M.T., Grigoriev, A., Knights, C.J., Holloway, E., Misfud, J., Li, K., Playford, M., Gregory, S.G., Humphray, S.J., Coffey, A., J., See, C.G., Marsh, S., Vatcheva, R., Kumlien, J., Labella, T., Lam, V., Rak, K.H., Todd, K., Mott, R., Graeser, D., Rappold, G., Zehetner, G., Poustka, A., Bentley, D.R., Monaco, A.P., Lehrach, H. (1996) An integrated yac map of the human X chromosome. **Genome Research** 6 (10), 943-955.
d. Mott, R. F., Grigoriev, A. V., Maier, E., Hoheisel, J. D., and Lehrach, H. (1993) Algorithms and software tools for ordering clone libraries: appplication to the mapping of the genome of Schizosaccharomyces pombe, **Nucleic Acids Research** 21, 1965-1974.

2. **Nucleotide composition analyses**. During my postdoc, I devised novel methods of analysis of nucleotide composition that links it with genome features (a, b). This has been utilized by others in sequencing 1000s of novel unstudied genomes of prokaryotes (including many pathogens) and in studies of mitochondrial mutation and is currently a standard tool for finding replication origins in newly sequenced genomes, often part of bioinformatics curricula in multiple programs and textbooks (c). Many predicted origins (including those in organisms for which no information on replication models was available, such as archaea) have been confirmed by experimental validation in other labs. I further extended this approach to the evolutionary analysis of viral genomes and bacteriophages, to showing how the same principles apply across single-stranded RNA genomes and how mutational patterns can vary across the same genome, as is the case with the coronavirus, and that could be used to prioritize potential targets of antiviral drugs (d). We also have expanded our compositional analysis to metagenomics studies (e).

a. Grigoriev, A. (1998) Genome Arithmetic. **Science**, 281, 1923.
b. Grigoriev, A. (1998) Analyzing genomes with cumulative skew diagrams. **Nucleic Acids Res**., 26, 2286-2290.

c. Grigoriev, A. How do replication and transcription change genomes. (2011) In Pevzner, P., Shamir, R., (Eds.) **Bioinformatics for Biologists**. Cambridge University Press. A textbook project supported by HHMI

d. Grigoriev, A. (2004) Mutational patterns correlate with genome organization in SARS and other coronaviruses. **Trends in Genetics**, 20, 131-135.

3. **Work in industry, protein domains, expression and interaction**. During my work in the industry, I led several proprietary unpublished projects such as the Cancer Genes Database and analysis of splicing vs. protein function in cancer. Although work in the industry limits publications, part of this research was published as a study on correlation of borders of exons and protein domains across the animal kingdom (a-c). These results have provided perhaps the most compelling multi-genome evidence for the exon shuffling theory. I also published the first study connecting genome-wide patterns of gene expression with protein-protein interactions (d) and the very first attempt to elucidate the complexity of the network of interacting proteins using the data from the very noisy datasets (e). These approaches have become a standard benchmark for most of the new large-scale interaction screens in multiple organisms.

a. Liu, M. and Grigoriev, A. (2004) Protein domains correlate strongly with exons in multiple eukaryotic genomes - evidence of exon shuffling? **Trends in Genetics**, 20, 399-403.

b. Liu, M., Walch, H., Wu, S., and Grigoriev, A. (2005) Significant expansion of exon-bordering protein domains during animal proteome evolution. **Nucleic Acids Res**. 33, 95-105.

c. Liu, M., Wu, S., Walch, H., and Grigoriev, A. (2005) Exon-domain correlation and its corollaries. **Bioinformatics** 21, 3213-3216.

d. Grigoriev, A. (2001) A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. **Nucleic Acids Res**., 29, 3513-3519.

4. **Genome variants**. In the genome variation field, we began with developing a tool finding copy-number variants (a), and with the NSF (ABI Innovation grant) and NIH (R15 grant) funding developed GROM, a variant caller that is much faster than standard approaches (b); a patent application was filed and an i-CORP grant for potential commercialization was awarded by the NSF. We also have used GROM to compare multiple genomes of woolly mammoths and elephants and used this analysis to train machine learning models for variant finding in very noisy ancient DNA datasets (c). GROM performed well in the DREAM Challenge competitions (d) and has been utilized for finding structural variants in rice (e) and other organisms.

a. Smith SD, Kawash JK, Grigoriev A. (2015) GROM-RD: resolving genomic biases to improve read depth detection of copy number variants. **PeerJ**, 3:e836

b. Smith, S., Kawash, J., Grigoriev, A. (2017) Lightning-fast genome variant detection with GROM. **GigaScience** 6(10), 1-7.

c. Kawash, J., Smith, S., Karaiskos, S., Grigoriev, A. (2018) ARIADNA: Machine Learning Method for Ancient DNA Variant Discovery. **DNA Research** 25(6):619-627.

d. Lee, AY, et al. (2018) Combining accurate tumor genome simulation with crowdsourcing to benchmark somatic structural variant detection. **Genome Biology** 19:188. (co-authored as ICGC-TCGA DREAM Somatic Mutation Calling Challenge Participant)

5. **Regulatory RNA**. This sub-section is probably the most relevant to the current proposal. My most recent interests and contributions include the analyses of small regulatory RNAs. My lab was the first to identify a number of novel phenomena and to provide computational framework for small RNA analysis. In collaboration with Nancy Bonini of UPenn, we discovered novel patterns of microRNA dynamics with age (a). Further computational analyses revealed potential involvement of tRNA and rRNA fragments in similar regulatory pathways (b). We also utilized published datasets and were the first to identify interaction domains and targets of tRFs and rRFs (c,d,f) and these results will be used in the current work. Finally, an overview of my theoretical work on the origins of RNA interference from tRNA fragments is in (e).

a. Abe, M., Naqvi, A., Hendriks, G. J., et al. (2014). Impact of age-associated increase in 2'-O-methylation of miRNAs on aging and neurodegeneration in Drosophila. **Genes & Development**, 28, 44-57. (co-corresponding author)

b. Karaiskos, S., Naqvi, A., Swanson, K., Grigoriev, A. (2015) Age-driven modulation of tRNA-derived fragments in Drosophila and their potential targets. **Biology Direct**, 10 (1), 51.

c. Guan, L., Karaiskos, S., and Grigoriev, A. (2020) Inferring targeting modes of Argonaute-loaded tRNA fragments. **RNA Biology** 17 (8), 1070–1080.
d. Guan, L., Grigoriev, A. (2021) Computational meta-analysis of ribosomal RNA fragments: potential targets and interaction mechanisms. **Nucleic Acids Res.** 49(7), 4085–4103.

D. Research Support

Current Research Support

Rutgers Grigoriev (PI) 1/1/2023-7/31/2024
Using small RNA and machine learning to study complex RNA regulation in planarians.
Role: Principal Investigator

Goal: Establish AI framework for analyzing small RNA regulation in planarian regeneration.

DGE-2152059 Brannigan (co-PI) 4/1/2022-3/31/2027
Codes For Life - Artificial Intelligence and Sustainable Software for Biomolecular Interactions.
Role: co-PI

Goal: Development and implementation of bold, new, potentially transformative models for STEM graduate education.

MCB-2027611 Grigoriev (PI) 5/1/2020-4/30/2023
NSF
Factors Contributing to Sequence Conservation in the SARS-CoV-2 Genome

Goal: to study patterns of nucleotide substitution and conservation in the context of RNA structure, interactions, potential coronavirus therapeutics

1R15CA220059 Grigoriev (PI) 06/01/2018 - 05/31/2023
NIH
Variant identification for cancer genomics
Role: Principal Investigator

Goal: to develop algorithms for fast identification of all types of variants in cancer genomes