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: Allen, Mary A

eRA COMMONS USER NAME (credential, e.g., agency login): maallen3

POSITION TITLE: Research Associate Professor

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
Spring Arbor University	B.A.	05/00	Biochemistry
University of Wisconsin-Madison	M.S.	04/06	Cellular and Molecular Biology
University of Colorado-Boulder	Ph.D.	11/2010	Molecular, Cellular and Developmental Biology
University of Colorado - Denver	Post-Doc	2011-2014	Computational Biology
University of Colorado-Boulder	Post-Doc	2014-2016	Down Syndrome

A. Personal Statement

I am keenly interested in transcriptional dysregulation in Down syndrome. In particular, I am focused on how the extra copy of chromosome 21 alters transcription factor activity. My recent research efforts in this regard are focused on (a) RUNX1 as a chromosome 21 encoded transcription factor that is critical to hematopoiesis; (b) how Down syndrome cells respond differently to environmental perturbation, such as heat shock; and (c) the impact of RCAN1 on mitochondrial phenotypes observed in Down syndrome. I specialize in combining computational biology and molecular biology to understand the conditions that characterize Down syndrome.

Because my current research direction involves performing nascent transcription analysis on cells from individuals with Down Syndrome. I'm also interested in understanding how non-coding enhancer RNAs (eRNAs) affect transcription factor function. Since eRNAs are short-lived, they are best detected via nascent transcription analysis. Hence, I have worked extensively with the PRO-seq and GRO-seq protocols for nascent transcripts. But nascent protocols are complex. Therefore, I have created and taught a workshop on GRO-seq, training more than 12 other labs on the GRO-seq procedure.

Furthermore, I have been intimately involved in developing novel algorithms to analyze nascent transcription. Beyond my experience with GRO-seq, I have worked with RNA for most of my molecular career. I have used most basic RNA techniques, including RNA-seq, nuclear run-on, primer extension, RACE, qRT-PCR, in vitro transcription, northern blotting, and RNA Immunoprecipitation. I am willing to share information and help any individuals needing assistance in areas of my expertise. I spent my early career firmly studying transcription and regulators of transcription.

Twenty years ago, when I started to learn data analysis, I had little opportunity for personalized training. Therefore, I am a proponent of sharing my expertise with others. I believe those of us who have had to climb up the mountain of data analysis/bioinformatics the hard way should let down ropes to allow for faster and greater access to data. For this reason, I designed and participated in teaching a workshop on Next generation sequencing analysis at the University of Colorado Boulder. This workshop is now an annual free workshop open to all. Moreover, we have now received funding to allow individuals from all over the world investigating Down syndrome to come and learn computation skills.

Beyond my technical skills, I am devoted to improving the lives of individuals with trisomy 21 through biomedical and computational research. To that end, and with invaluable support from the Anna and John J. Sie Foundation and the BioFrontiers Institute, I was able to establish and direct the Crnic Institute Boulder Branch (CBB), at CU Boulder. The CBB is affiliated with the Linda Crnic Institute for Down syndrome research, at CU Anschutz, and aims to integrate and support the efforts from 20+ laboratories at CU Boulder studying a wide spectrum of pathophysiological alterations associated with Down syndrome, through a multidisciplinary and collaborative view.

Finally, I have extensive training in data integrity and quality data collection and analysis. I am the Faculty Director of the Responsible Conduct of Research (RCR) Education Program. Quality data collection is often an overlooked aspect of big data. Understanding metrics of what makes a data set a good one is the key to big data training. Training to keep a clean data set is crucial to efficient data analysis.

PI	MY ROLE	AWARDED TOTAL	TITLE	AGENCY	AWARD START DATE	AWARD END DATE
ALLEN, MARY ANN	PI	$1,890,079 $191,486 Supplement	Function of RUNX1 in diverse Down syndrome tissues	National Heart, Lung, and Blood Institute/NIH/DHHS	2/1/21	3/31/25
DONALDSON, ZOE REBECCA	Co-I	$2,602,973	The Neuromolecular Basis of Adaptation to Bond Loss	National Institute of Mental Health/NIH/DHHS	2/7/22	11/30/26
ALLEN, MARY ANN	PI	$58,328	Data Science for Diverse Scholars in Down Syndrome Research (DS3)	Children's Hospital of Philadelphia, supplement to NIH U01	5/2/22	4/30/23
ALLEN, MARY ANN	PI	$65,329	Data Science for Diverse Scholars in Down Syndrome Research (DS3)	Children's Hospital of Philadelphia, supplement to NIH U01	5/1/23	4/30/24
ALLEN, MARY ANN	PI	$418,813	Data Science for Diverse Scholars in Down Syndrome Research (DS3)	NIH R25 (subaward of University of Colorado Denver)	6/1/24	5/31/29

Lapses in publications: I did not publish as a Wisconsin graduate student (2002-2006). In 2005, my fellow graduate students and I discovered my principal investigator was guilty of falsification of data. My doctoral committee, therefore, recommended I not publish any of my work from that laboratory. The lapse in publication was unfortunate, but this experience is the primary reason for my passion in

responsible conduct of research. I also did not publish in both 2012 and 2015, because I was on maternity leave. Finally, I would also like to note that I am geographically constrained because I am a single mother, which clarifies why I have stayed in Colorado for both a Post-doc and Faculty position.

B. Positions, Scientific Appointments, and Honors

2023-present Assistant Director of the Crnic Boulder Branch
2023-present Research Associate Professor
2022-present Instructor and Organizer, Data Science in Down syndrome for Diversity Scholars (DS3)
2017-2023 Research Assistant Professor
2017-present Faculty Director of the Responsible Conduct of Research (RCR) Education Program
2017-present Instructor Responsible Conduct of Research (GRAD 5000), CU- Boulder
2016-2017 Research Associate, CU- Boulder
2016-2018 Lecturer, Bionomics and Genomics (MCDB 4520/5520), CU- Boulder
2016 Organizer and Instructor, GRO-seq workshop, CU- Boulder
2016, 2011 Lecturer, Bioinformatics I (CBPS 7711), CU- Boulder
2014-present Organizer and Instructor, Short Read sequencing course, CU- Boulder
2000-2002 Research Technician, University of Michigan- Ann Arbor
1997-1999 Teaching Assistant, Spring Arbor University
1999 Physics Instructor, The DaVinci Institute

Other Experience and Professional Memberships
2021-present Trisomy 21 Research Society
2013 ORI at 20: Reassessing Research Integrity-A Leadership Conference (invited speaker)
2007-2008 Faculty Representative for Graduate Students, MCDB CU- Boulder
2007-2008 Chair of the Graduate Student Symposium Committee, MCDB CU- Boulder
2006 Participated in an ethics panel titled: "Truth and consequences: What happened when a group of University of Wisconsin students learned their advisor faked data." University of Illinois at Urbana-Champaign

Honors
2014-2016 *Sie* Postdoctoral Fellowship
2011-2014 Computational Bioscience Program Postdoctoral Fellowship

C. Contributions to Science

Complete List of Published Work in MyBibliography:
http://www.ncbi.nlm.nih.gov/sites/myncbi/1pCsbbdigcs/bibliography/40637057/public/?sort=date&direction=ascending

1. **Contributed to better understanding transcription termination and RNA processing.** In *C. elegans*, there is a process known as trans-splicing that creates new 5' ends and caps on most RNAs. *C. elegans* organizes its genes into operons, transcripts in tandem that are cleaved apart by 3' end cleavage and trans-splicing. In several pieces of my work, I explored which genes control trans-splicing and transcription termination. I computationally demonstrated that there are sequences between the genes in operons that are statically over-represented, and a collaborator proved they function in operon processing. I also found several genes that, when removed by RNAi, reduce transcription termination.
    a. M Melnick, P Gonzales, J Cabral, **MA Allen**, RD Dowell, CD Link. Heat shock in C. elegans induces downstream of gene transcription and accumulation of double-stranded RNA. PLoS ONE 2019. 14 (4), e0206715.
    b. **Allen MA**, Hillier L, Waterston RB and Blumenthal T. A global analysis of C. elegans trans-splicing. Genome Research 2011. 21(2): 255-264. PMC3032929
    c. Cui M*, **Allen MA***, Larsen A, Macmorris M, Han M, and Blumenthal T. Genes involved in pre-mRNA 3'-end formation and transcription termination revealed by a lin-15 operon Muv suppressor screen. Proceedings of the National Academy of Sciences of the United States of America, 2008. 105(43): 16665-16670. (*Contributed equally to this work.) PMC2571909

d. Lasda E.L., **Allen MA** and Blumenthal T. Polycistronic pre-mRNA processing in vitro: snRNP and pre-mRNA role reversal in trans-splicing. Genes & Development 2010. 24(15): 1645–1658. PMC2912562

2. **Determined functions of transcription factors in eRNA production.** p53 is arguably the most well studied transcription factor and is mutated in half of all cancers. Many labs have tried to determine the direct targets of p53, however use of steady state assays (such as RNA-seq and microarray) result in mixtures of primary and secondary effects. Therefore, to determine the primary targets of p53 I used GRO-seq to directly measure nascent transcription. I found many new p53 targets and showed that p53 is a transcriptional activator. I also demonstrated a major transcriptional target of p53 are p53-responsive eRNAs. I completed both the molecular and computational analysis of GRO-seq. In related work, I used GRO-seq to analyze the transcription of HIF1A target genes. I extended a similar approach to assist Sasse in the understanding of glucocorticoid receptor and Cardiello to look at heat shock.

   a. Cardiello, J. F., Westfall, J., Dowell, R. D., & **Allen, M. A.** (2023). Characterizing Primary transcriptional responses to short term heat shock in paired fraternal lymphoblastoid lines with and without Down syndrome. *bioRxiv*, 2023-01.
   b. S. K. Sasse, M. Gruca, **M A. Allen**, V Kadiyala, T Song, F Gally, A Gupta, M. A. Pufall, R D. Dowell, and Anthony N. Gerber. Nascent transcript analysis of glucocorticoid crosstalk with TNF defines primary and cooperative inflammatory repression. Genome Res. 2019; 29: 1753-1765. PMC6836729
   c. J.G. Azofeifa, **MA Allen**, J.R. Hendrix, T Read, J.D. Rubin, R.D. Dowell. Enhancer RNA profiling predicts transcription factor activity. Genome research. 2018. PMID: 29449408
   d. **M. A. Allen**, Z. Andrysik, V.L. Dengler, H.S. Mellert, A. Guarnieri, J.A. Freeman, K.D. Sullivan, M.D. Galbraith, X. Luo, W.L. Kraus, R.D. Dowell, and J.M. Espinosa (2014) Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. eLife 2014;3:e02200. PMC4033189

3. **Created computational algorithms to improve analysis of nascent transcription and ATAC-seq.** Most GRO-seq analysis to date has been done on annotated genes. However, much of the unannotated genome is transcribed. For example, enhancer RNAs, eRNAs are non-coding transcripts, which may influence transcription of other genes but are not annotated. Because nascent transcription analysis is so new, tools for analysis of non-coding regions did not exist. I have assisted in the development of an algorithm that determines from data alone all transcribed regions of the genome. Using this algorithm, I re-analyzed my earlier work and found hundreds of unannotated regions that are activated by p53. This tool is now available for download for other labs to use. In a recently submitted paper we used a mathematical model to analyze hundreds of publicly available nascent transcription datasets. This work uncovered a tight relationship between eRNA activity and TF binding function. These tools will allow the nascent transcription field to analyze these data in a much more powerful way. Additionally, we have extended some of these tools to analyze ATAC-seq.

   a. S. Hunter, R.F. Sigauke, J.T. Stanley, **M.A. Allen**, R.D. Dowell. Protocol variations in run-on transcription dataset preparation produce detectable signatures in sequencing libraries. BMC genomics 23 (1), 1-18.
   b. I Tripodi, **M Allen**, R Dowell. Detecting differential transcription factor activity from ATAC-Seq data. 2018: Molecules 23 (5), 1136. PMC6099720
   c. JG Azofeifa, **MA Allen**, JR Hendrix, T Read, JD Rubin, RD Dowell. Enhancer RNA profiling predicts transcription factor activity. Genome research. 2018. PMID: 29449408
   d. ME Lladser, JG Azofeifa, **MA Allen**, RD Dowell. RNA Pol II transcription model and interpretation of GRO-seq data. J Math Biol. 2016 May 3. PubMed PMID: 27142882.

4. **Directed education in Responsible Conduct or Research.** Many years ago, when I was in my fourth year of graduate school, my fellow graduate students discovered that our thesis advisor had engaged in misconduct by falsifying and fabricating data in two grant applications. We informed the university, and my advisor resigned. This event was a turning point in my life. Since then, I have dedicated myself to educating others about responsible conduct of research. I have written several papers on this topic. I have also presented both locally and nationally to scientific boards and graduate students. Currently, I am the University of Colorado- Boulder *Responsible Conduct of Research* coordinator. I believe we

must first prevent misconduct by educating the new generations of researchers. Unfortunately, some misconduct will always occur, and in those circumstances, we must respond to protect those affected by the misconduct and help them to progress beyond the event. In so doing, we get the greatest value out of scientific research.

a. **Allen MA**, Train students to navigate ethical swamps. Nature 2019. 568 (7751): 145. PMID:30971841
b. **Allen MA**, Dowell RD. Retrospective reflections of a whistleblower- Opinions on misconduct responses. Accountability in Research 2013. 20(5-6): 339-348. PMID:24028481

5. **Added to field's knowledge of function of the Hedgehog pathway and basal cell carcinoma in skin.** The sonic hedgehog pathway is necessary for skin development. I explored several genes in this pathway and their effect on basal cell carcinoma and sebaceous gland development.

a. Grachtchouk M, Pero J, Yang S, Ermilov A, Michael LE,Wang A, Wilbert D, Patel R, Ferris J, Diener J, **Allen M**, Lim S, Syu L, Verhaegen M, and Dlugosz AA. Basal cell carcinomas in mice arise from hair follicle stem cells and multiple epithelial progenitor populations. Journal of Clinical Investigation 2011. 121: 1768–1781. PMC3083781
b. **Allen M**, Grachtchouk M, Sheng H, Grachtchouk V, Wang A, Wei L, Liu J, Ramirez A, Metzger D, Chambon P, Jorcano J, and Dlugosz AA. Hedgehog signaling regulates sebaceous gland development. The American Journal of Pathology 2003. 163(6): 2173–2178. PMC1892397