

Xianjun Dong, Ph.D.**Address:** 65 Landsdowne Street, Rm 301A, Cambridge, MA 02139 USA**Phone:** 617-768-8691**Fax:** 617-768-8520**Email:** xdong@rics.bwh.harvard.edu**Website:** <http://www.sterding.com>**PART I: GENERAL INFORMATION**

Education

1998 - 2002 B.Sc., Biomedical Engineering, Southeast University, China
 2002 - 2005 M.Sc., Biomedical Engineering, Southeast University, China
 2005 - 2010 Ph.D., Bioinformatics, University of Bergen, Norway
 Advisor: Drs. Boris Lenhard and Rein Aasland

Postdoctoral Training

09/10-09/13 Postdoctoral Associate, Program of Bioinformatics and Integrative Biology, University of
 Massachusetts Medical School (Dr. Zhiping Weng)

Appointments at Hospitals/Affiliated Institutions

11/13-present Instructor, Department of Neurology, Harvard Medical School
 09/13-present Assistant Scientist in Neurology, Brigham and Women's Hospital (Dr. Clemens Scherzer)

Other Professional Positions

2005 - 2005 Manager Trainee / Process Engineer, P&G

Major Administrative Leadership Positions

2013 - present Bioinformatics team lead in Neurogenomics Laboratory, Harvard Medical School
 2007 - 2007 Organizer of the "EnsEMBL in a Nutshell" Bioinformatics Workshop, University of Bergen
 2005 - 2007 President, Chinese Students and Scholars Association (CSSA) of Bergen, Norway

Professional Societies

2006 - present International Society for Computational Biology (ISCB), Member
 2013 - present American Society for Human Genetics (ASHG), Member
 2015 The 10th Annual DNA Day Essay Contest, Judge

Honors and Prizes

2000 2nd prize of the National Undergraduate Mathematical Contest in Modeling
 2000 *Liu Yonglin* Fellowship Award, Southeast University, China
 2000 Distinguished Student with 1st grade Scholarship, Southeast University, China
 2001 1st prize of 2001 "Sony Cup" National Undergraduate Electronic Design Contest, China
 (winning by designing a robot-controlled electronic car)
 2001 **Meritorious** winner (top 10%) in the international Mathematical Contest in Modeling
 (MCM) (winning by modeling the growth control of Zebra mussels in the Great Lakes)
 2001 Distinguished Student with 1st grade Scholarship, Southeast University, China
 2002 Distinguished Graduate Student Award, Southeast University, China
 2009 Full Scholarship, "Chromatin Domains and Insulators" Workshop, Spain
 2009 Travel Fellowship, MCB Research School, University of Bergen, Norway
 2010 Chinese Government Award for Outstanding Self-financed Students Abroad, Ministry of
 Education, China (1 of 500 annual awardees in the world wide)
 2015 Reviewers' Choice – top 10% best posters in the American Society of Human Genetics
 (ASHG) meeting, US

Service to Professional Publications**- Invited reviewer -** (Total: **26** times for **15** journals)

2011 - present *PLoS Computational Biology* (1 paper)
 2011 - present *Genome Research* (1 paper)
 2012 - present *Nature Biotechnology* (1 paper)
 2013 - present *PLoS ONE* (2 paper)
 2013 - present *Genome Biology* (2 paper)
 2014 - present *Bioinformatics* (5 paper)
 2014 - present *PLoS Genetics* (1 paper)
 2014 - present *Evolutionary Bioinformatics* (2 paper)
 2014 - present *International Journal of Biological Sciences* (3 paper)
 2014 - present *Journal of Biomedical Informatics* (2 paper)
 2014 - present *Movement Disorders* (1 paper)
 2014 - present *The Pharmacogenomics Journal* (1 paper)
 2015 - present *Current Bioinformatics* (2 paper)
 2015 - present *Neurogenetics* (1 paper)
 2015 - present *Journal of Bioinformatics and Computational Biologys* (1 paper)

Funding Information**- Past -**

2008 - 2012 Title: EDAC: ENCODE Data Analysis Center
Grant number: U01 HG004695-04
Goal: The goal of this project is to perform integrative analysis on ChIP-Seq, FAIRE-Seq and RNA-Seq data as a part of the ENCODE data analysis center.
Role: Bioinformatics engineer

- Current -

2011 - 2016 Title: Epigenetic Markers in Huntington's Disease Brain
Grant number: R01 NS073947
Goal: The goal of this project is to map, genome-wide HD disease-related changes in histone methylation markings in prefrontal neurons.
Role: Bioinformatics analyst

2013 - 2016 Title: A Next Generation of Biomarkers for Incipient Huntington's Disease
Grant number: U01 NS082080
Goal: The goal of this project is to establish a national RNA biobank resource and develop digital biofluid markers useful for tracking disease progression.
Role: Co-investigator

- Current unfunded -

2015-2016 Title: Discovery of a new microcosm of circular RNAs in neurons of human brain
Grant: BWH Institute for the Neurosciences – Neuroscience Seed Grants
Goal: The goal of this project is to define circular RNAs and validate its function in human dopamine neurons.
Role: PI

PART II: REPORT OF TEACHING, TRAINING, AND PRESENTATIONS

Teaching of Students in Courses

2007 EnsEMBL in a Nutshell, Bergen Center of Computational Science, University of Bergen, Norway
 Role: Lecturer
 National-wide bioinformatics students / researchers, 8-hr sessions per day for 5 days

Laboratory and Other Research Supervisory and Training Responsibilities

2013 - present Bioinformatics training and supervision of the bioinformatics staff, students, engineers
 Daily mentorship for 3 years

Formally Supervised Trainees / Students

2009 - 2012 Yogita Sharma (graduate student in University of Bergen, Norway)
 Mentor, co-supervisor for her research project & thesis, and co-author of one article in *BMC Bioinformatics*

2013 - 2014 Carmen Portenkirchner (MD candidate in Paracelsus Medical University, Austria)
 Thesis co-advisor, research advisor

2013 - 2014 Dr. Shuilin Jin (visiting research scholar from Harbin Institute of Technology, China)
 Co-supervisor, Bioinformatics trainer

2014 - 2015 Alyssa Ehrlich (1st year graduate student in Harvard University)
 Research advisor for her proposal and her internship in the laboratory

2014 - 2015 Ruoting Wang (undergraduate in University of Massachusetts)
 Supervisor, research advisor for his summer internship in the neurogenomics laboratory

2015 - 2015 Will Nemirovsky (junior high school student in BB&N)
 Supervisor, research advisor for his summer bioinformatics training in Harvard

2015 - 2016 Dr. Yunfei Bai (visiting scholar from Southeast University, China)
 Co-supervisor, Bioinformatics trainer

2015 - 2017 Tao Wang (PhD Joint Training Program student, Harbin Institute of Technology, China)
 Supervisor, research advisor for his two-year bioinformatics training in Harvard

2016 - 2016 Ellen DeGennaro (graduate student in Harvard-MIT HST PhD program)
 Supervisor, research advisor for her proposal and her internship in the laboratory

Invited Presentations**- Local -**

2011 Correlate histone modifications and gene expression / BIB seminar (UMass Medical School)

2012 Redefining the piRNA-producing loci of the mouse testis as genes / Invited Bioinformatics talk (UMass Medical School)

2014 From ENCODE to PD Brain Map: Bioinformatics in Big Data Research / The Center for Neurologic Diseases' ADPD seminar series (Harvard Institute of Medicine, BWH)

2015 The BRAINCODE consortium: Decoding neuronal genome function in human brain / The Center for Neurologic Diseases' ADPD seminar series (Harvard Institute of Medicine, BWH)

- Regional -

2007 Remnants of lost bystander genes under non-coding selection pressure in human-zebrafish synteny blocks (BREW 2007, Norway)

2012 Modeling gene expression with chromatin features (ENCODE conference, MIT)

2013 Studying gene regulation using comparative genomics and epigenomics (DFCI / HMS)

2013 Studying gene regulation using comparative genomics and epigenomics (BWH / HMS)

- National -

2011 Modeling gene expression with chromatin features / Invited talk (ENCODE conference, Stanford University)

2013 Long-range gene regulation in vertebrates genome (Berkeley Lawrence Laboratory)

- International -

- 2010 Translog, a web browser for studying the expression divergence of homologous genes (The 8th Asia Pacific Bioinformatics Conference, India)
- 2010 Web resources to study the expression divergence of homologous genes (Huazhong Agricultural University, China)
- 2010 Evolutionary mechanisms of developmental long-range gene regulation in vertebrate genomes (GIS, Singapore)
- 2013 Studying gene regulation using comparative genomics and epigenomics (Southeast University / Huazhong Agricultural University / Tongji University, China)
- 2015 NGS and its application in translational medicine (Tongji Hospital, China)

PART III: REPORT OF ACTIVITIES AND INNOVATIONS

Technological and Other Scientific Innovations

- Synorth A website designed for genome research, allowing to explore the evolution of synteny and long-range regulatory interactions between vertebrate genomes
URL: <http://synorth.genereg.net>
Published as first author on *Genome Biology* (Impact factor: 10.30), 13 citations
- Translog A website designed to study the expression divergence of homologous genes
URL: <http://translog.genereg.net>
Published as first author on *BMC Bioinformatics* (Impact factor: 3.02)
- Factorbook A Wiki-based database for transcription factor-binding data generated by the ENCODE consortium
URL: <http://factorbook.org>
Published as co-author on *Nucleic Acid Research* (Impact factor: 8.28), 4 citations

Educational Activities to the Community

- 2012 Innovation Month in the Worcester Public Schools: How to extract DNA from an onion
Claremont Academy, Worcester
<http://www.umassmed.edu/news/2012/education/gsbs-students-take-science-on-road.aspx>
- 2008 – present Founder of bioinformatics blog, “One Tip Per Day” (<http://onetipperday.blogspot.com/>), which has 200 visit per day on average and more than 520,000 visits in total

Media Coverage

- ENCODE consortium, Dong et al. *Nature*, 2012
CNN: <http://www.cnn.com/2012/09/05/health/encode-human-genome/>
BBC: <http://www.bbc.com/news/health-19202141>
Scientific American: <http://www.scientificamerican.com/article/junk-dna-encode/>
Evolution News: http://www.evolutionnews.org/2012/09/junk_no_more_en_1064001.html
New York Times: <http://www.nytimes.com/2012/09/06/science/far-from-junk-dna-dark-matter-proves-crucial-to-health.html>
LA Times: <http://articles.latimes.com/2012/sep/05/science/la-sci-dna-encode-20120906>
The Guardian: <http://www.theguardian.com/science/2012/sep/05/genes-genome-junk-dna-encode>
Washington Post: http://www.washingtonpost.com/national/health-science/junk-dna-concept-debunked-by-new-analysis-of-human-genome/2012/09/05/cf296720-f772-11e1-8398-0327ab83ab91_story.html
- Dong et al. *Genome Biology*, 2012
Genomeweb: <http://www.genomeweb.com/blog/week-genome-biology-59>
Biomed Central: <http://blogs.biomedcentral.com/bmcblog/2012/09/07/human-genomics-comes-of-age-encode-open-access-and-biomed-central/>
Umassmed News: <http://www.umassmed.edu/news/news-archives/2012/09/umms-faculty-integral-to-consortium-decoding-human-genome/>

PART IV: PUBLICATIONS

Original Articles:

1. Mungpakdee S, Seo HC, Angotzi AR, **Dong X**, Akalin A, et al. Differential evolution of the 13 Atlantic salmon Hox clusters. *Molecular biology and evolution*. 2008; 25(7):1333-43. PMID: 18424774
2. **Dong X**, Fredman D, Lenhard B. Synorth: exploring the evolution of synteny and long-range regulatory interactions in vertebrate genomes. *Genome biology*. 2009; 10(8):R86. PMID: 19698106, PMCID: PMC2745767
3. Akalin A, Fredman D, Arner E, **Dong X**, Bryne JC, et al. Transcriptional features of genomic regulatory blocks. *Genome biology*. 2009; 10(4):R38. PMID: 19374772, PMCID: PMC2688929
4. **Dong X**, Akalin A, Sharma Y, Lenhard B. Translog, a web browser for studying the expression divergence of homologous genes. *BMC bioinformatics*. 2010; 11 Suppl 1:S59. PMID: 20122234, PMCID: PMC3009532
5. **Dong X**, Navratilova P, Fredman D, Drivenes Ø, Becker TS, et al. Exonic remnants of whole-genome duplication reveal cis-regulatory function of coding exons. *Nucleic acids research*. 2010; 38(4):1071-85. PMID: 19969543, PMCID: PMC2831330
6. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS biology*. 2011; 9(4):e1001046. PMID: 21526222, PMCID: PMC3079585
7. Yildirim O, Li R, Hung JH, Chen PB, **Dong X**, et al. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*. 2011; 147(7):1498-510. NIHMSID: NIHMS343144 PMID: 22196727, PMCID: PMC3252821
8. **Dong X**, Greven MC, Kundaje A, Djebali S, Brown JB, et al. Modeling gene expression using chromatin features in various cellular contexts. *Genome biology*. 2012; 13(9):R53. PMID: 22950368, PMCID: PMC3491397
9. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome research*. 2012; 22(9):1798-812. PMID: 22955990, PMCID: PMC3431495
10. Cheng C, Alexander R, Min R, Leng J, Yip KY, et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome research*. 2012; 22(9):1658-67. PMID: 22955978, PMCID: PMC3431483
11. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489(7414):57-74. NIHMSID: NIHMS381381, PMID: 22955616, PMCID: PMC3439153
12. Wang J, Zhuang J, Iyer S, Lin XY, Greven MC, et al. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic acids research*. 2013; 41(Database issue):D171-6. PMID: 23203885, PMCID: PMC3531197
13. Li XZ, Roy CK, **Dong X**, Bolcun-Filas E, Wang J, et al. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Molecular cell*. 2013; 50(1):67-81. NIHMSID: HHMIMS469552 PMID: 23523368, PMCID: PMC3671569
14. Hoss AG, Kartha VK, **Dong X**, Latourelle JC, Dumitriu A, et al. MicroRNAs located in the Hox gene clusters are implicated in huntington's disease pathogenesis. *PLoS genetics*. 2014; 10(2):e1004188. PMID: 24586208, PMCID: PMC3937267
15. Haberle V, Li N, Hadzhiev Y, Plessy C, Previti C, et al. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature*. 2014; 507(7492):381-5. PMID: 24531765
16. **Dong X**, Tsuji J, Labadorf A, Roussos P, Chen JF, et al. The Role of H3K4me3 in Transcriptional Regulation Is Altered in Huntington's Disease. *PloS one*. 2015; 10(12):e0144398. PMID: 26636336, PMCID: PMC4670094
17. Bai G, Cheung I, Shulha HP, Coelho JE, Li P, et al. Epigenetic dysregulation of hairy and enhancer of split 4 (HES4) is associated with striatal degeneration in postmortem Huntington brains. *Human molecular genetics*. 2015; 24(5):1441-56. PMID: 25480889, PMCID: PMC4321450
18. Locascio JJ, Eberly S, Liao Z, Liu G, Hoising AN, et al. Association between α -synuclein blood transcripts and early, neuroimaging-supported Parkinson's disease. *Brain*. 2015; 138(Pt 9):2659-71. PMID: 26220939, PMCID: PMC4643625

Reviews:

1. Fredman D, **Dong X**, Lenhard B. Making enhancers from spare parts of the genome. *Genome biology*. 2011; 12(12):138. PMID: 22206586, PMCID: PMC3334608
2. **Dong X**, Weng Z. The correlation between histone modifications and gene expression. *Epigenomics*. 2013; 5(2):113-6. NIHMSID: NIHMS598474, PMID: 23566087, PMCID: PMC4230708

Public Science:

1. **Dong X**. The brainstorm. *IEEE Potential*. 2009 Vol. 28, No.1.

Thesis:

1. **Dong X**. Evolutionary mechanisms of developmental long-range gene regulation in vertebrate genomes. 2010.

Abstracts, Poster Presentations Presented at Professional Meetings:

1. **Dong X**, Xie J, Lu Z. PatternExplorer2.0: a tool to explore related genes with random pattern sequence in genome. *The International Forum on Biochip Technologies (IFBT2004)*, Beijing, China, 2004.
2. **Dong X**, Zhang L, Lu Z. Why SARS virus only infects the lung of human? A model to analyze the association of virus attachment proteins and cell receptors. *The 2nd International Forum on Post-Genome Technologies (2'IFPT)*, Nanjing, 2004.
3. Xie J, Sun X, Lu Z, Xue W, **Dong X**, Lu Z. gWord: A Tool for Genome-Wide Word Search and Count. *Proceeding of Bioinformatics and Biomedical Engineering*, 2007. ICBBE 2007.
4. **Dong X**. Remnants of lost bystander genes under non-coding selection pressure in human-zebrafish synteny blocks. *Bioinformatics Research and Education Workshops (BREW)*, Bergen, Norway, 2007. **(Oral)**
5. **Dong X**, Fredman D, Engström P, Lenhard B. Reconstructing the Evolutionary History of Genomic Regulatory Blocks in Vertebrates. *The 15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB)*, Austria, 2007.
6. **Dong X**, Akalin A, Fredman D, Lenhard B. Distribution of highly conserved non-coding elements in *Danio rerio*. *The 5th Asia Pacific Bioinformatics Conference (APBC)*, Hong Kong, 2007.
7. **Dong X**, Sheng Y, Gupta P, Lenhard B. Computational prediction of ncRNAs as target genes in genomic regulatory blocks. *The 6th Asia Pacific Bioinformatics Conference (APBC)*, Japan, 2008.
8. **Dong X**, Fredman D, Lenhard B. GRB Evolution Browser: a web resource for exploring evolutionary behaviours of vertebrate genomic regulatory blocks. *The 4th EMBO Conference: From Functional Genomics to Systems Biology*, Germany, 2008.
9. **Dong X**, Akalin A, Fredman D, Lenhard B. Evolutionary dynamics of genomic regulatory blocks: HCNEs and HARs, the content and the boundaries. *The 7th Asia Pacific Bioinformatics Conference (APBC)*, China, 2009.
10. **Dong X**. Translog, a web browser for studying the expression divergence of homologous genes. *The 8th Asia Pacific Bioinformatics Conference (APBC)*, Bangalore, India, 2010. **(Oral)**
11. **Dong X**. Predicting gene expression using histone modifications. ENCODE Conference, Stanford, 2011.
12. **Dong X**. Correlating histone modifications and gene expression. *Systems Biology: Global Regulation of Gene Expression*, CSHL, 2012.
13. **Dong X**. Modeling gene expression with chromatin features. ENCODE Conference, MIT, 2012. **(Oral)**
14. Liao Z, Zheng B, **Dong X**, Vanderburg CR, Locascio JJ, Pochet N, Regev A, Portenkirchner C, Altmann P, Jin S, Beach TG, Adler CH and Scherzer CR. The PD Brain Map: Mapping the Transcriptional Architecture of Dopamine Neurons in Human Brain. *The American Society of Human Genetics (ASHG)*, Boston, 2013.
15. **Dong X**, Liao Z, Gritsch D, Guennewig B, Hadzhiev Y, Bai Y, Liu G, Zheng B, Blauwendraat C, Hu Q, Adler CH, Hedreen JC, Frosch MP, Faull R, Nelson PT, Locascio J, Vanderburg C, Rizzu P, Cooper AA, Heutink P, Beach TG, Mueller F, Mattick JS, Scherzer CR. BRAINCODE: How does the Human Genome Function in Specific Brain Neurons? *The American Society of Human Genetics (ASHG)*, Baltimore, 2015.

PART V: NARRATIVE REPORT

My goal is to understand how the human genome functions in brain of healthy and Parkinson by developing and applying computational methods to integrate and interpret -omics data. I feel extremely lucky to be a young scientist in this booming era of biomedical research –both for brain research and for computational biology. The rapid development of next-generation sequencing (NGS) techniques has dramatically decreased the time and cost of sequencing a personal genome; it took more than a decade and over \$3 billion ten years ago and just in one week and under \$3000 today. A tremendous volume of data is constantly being generated, covering various aspects of biomedicine, such as genomics, transcriptomics, epigenomics, proteomics, and metabolomics. This deluge of data is challenging both in the way we analyze it and the way we interpret it. This becomes particularly challenging and exciting when it comes to human brain and brain diseases. Human brain is our most complex organ, conducting various advanced functions with more than 86 billion neurons and trillions of connections. How is a single copy of genome wired into neurons with different functions? What happened to the Parkinson's brain, in DNA, RNA, and protein level? And how? Can we do something to predict and/or prevent it? Answering these questions becomes possible until nowadays when sequencing and computing techniques meet at their best-ever time. This is exactly why I plan to tackle in my research as a computational biologist specialized in human brain research.

In my current appointment, I spent most (80%) of my time in academic research in the Neurogenomics Laboratory; for most of that time I have leaded at least two projects. The remainder of my time is divided between administrative duties as the bioinformatics team lead (5%) and training the fellows as an Instructor (15%).

My current research focuses mainly on the neurodegenerative diseases, including Parkinson's disease (PD) and Huntington's disease (HD), to delineate their genetic architecture using various NGS techniques and integrative bioinformatics methods. With Dr. Clemens Scherzer at BWH, I have started two projects -- one is to create a national RNA biobank resource and develop digital biofluid markers useful for tracking HD progression (funded by NIH); the other is to provide a high-definition view of transcription activity of brain neurons via ultra-deep sequencing of the transcriptome of healthy controls and PD patients. We are wrapping up a manuscript for the latter project.

My publication record reflects my growing national and international standing in the field. I have 21 publications in top peer-reviewed journals including *Nature*, *Cell*, *Molecular Cell*, *Genome Research*, *Genome Biology*, *Nucleic Acids Research*, with more than 5,400 citations so far. Briefly, I've discovered the first set of enhancers (which are usually found in non protein-coding regions) in protein-coding exons (Dong et al. *Nucleic Acids Research* 2009). I've also served as an active member to the ENCODE Project (ENCyclopedia Of DNA Elements), where my co-author work ("An integrated encyclopedia of DNA elements in the human genome", *Nature*, 2012; 757 citations in 2013 alone) was recognized as "Top 10 Breakthroughs of the Year 2012" in *Science*. I've developed and made publicly available three web-based tools useful for studying gene regulation, Synorth (Dong et al. *Genome Biology* 2009), Translog (Dong et al. *BMC Bioinformatics* 2010), and Factorbook.org (Wang et al. *Genome Research* 2012). I've also designed sophisticated statistical models to quantify the association between gene expression and chromatin features in various cellular contexts, by integrating various genomic, epigenomics, and transcriptomics data generated from NGS techniques such as ChIP-Seq, DNase-Seq, RNA-Seq, CAGE etc. (Dong et al. *Genome Biology* 2012; Dong et al. *Epigenomics* 2013).

I love teaching and communicating with people. I have served as mentor and co-advisor for junior PhD students since my last year of PhD study. I also now co-supervise the bioinformatics visiting scientists/students in Dr. Scherzer's lab, and have trained/supervised eight (8) students since I joined in Dr. Scherzer's lab in 2013. Moreover, I like to share knowledge to the community; the bioinformatics blog I founded (<http://onetipperday.sterding.com>) has received ~200 visits per day and >520,000 visits since its opening in 2008. In 2007, I organized a bioinformatics workshop in Norway with over 50 national participants. It was the introduction of the Ensembl (<http://www.ensembl.org>) outreach to Norway and was considered extremely useful by local researchers. I am also enthusiastic in supporting educational activities in my local community, for example, by teaching hands-on science experiments in local elementary schools.

In summary, with a demonstrated record of accomplished and productive research in the area of genomics, I believe that my expertise and experience in bioinformatics have prepared me to the new role as an Assistant Professor in neurogenomics.