BIOGRAPHICAL SKETCH

NAME: Ding, Li

eRA COMMONS USER NAME (agency login): DINGLI

POSITION TITLE: Associate Professor

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Fudan University, Shanghai	BS	07/1991	Biology
University of Utah School of Medicine, Salt Lake City, UT	PHD	08/1998	Biochemistry
Stanford University, Palo Alto, CA	Postdoctoral Fellow	05/2000	Biochemistry

A. Personal Statement

The long term goal of my research is to combine my strengths in biology and bioinformatics to answer fundamental questions in biological science and human diseases. My current research focuses on understanding the genetic basis of human diseases, through innovative computational and experimental approaches. My lab has developed important bioinformatics tools, including VarScan, BreakDancer, TIGRA-SV, Pindel-C, MSIsensor, HotSpot3D, and MuSiC, which are widely used for human genetics and cancer proteogenomic studies. We have aggressively pursued technological innovation and have been actively developing enabling tools in proteomics using methodologies that increasingly combine the power of genomics and proteomics to discover novel disease markers. I have been at the forefront of human genetics and cancer genomics research and with a long record of significant discovery: conducted and led the very first large-scale lung cancer genomics study (2006-2008), pioneered drug-resistant subclone/mutation discovery during tumor metastasis/relapse (2009-2012), used the Pan-cancer approach to gain statistical power for rapid, massive cancer gene discovery (starting from 2012), developed integrative approaches to study combined effects of germline and somatic mutations simultaneously (Starting from 2012), and utilized global proteomic and phosphoproteomic data for druggable protein discovery (Starting from 2012). Innovative use of blood sample sequencing data uncovered pre-existing leukemic mutations in important leukemia-associated genes in individuals without hematological diseases (starting from 2013). In summary, I have developed an outstanding program to advance biology and treatment, as demonstrated by a track record of innovative approaches and ground-breaking discoveries. Relevant publications selected from over 120 total publications are listed below.

1. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MD, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, **Ding L.** Mutational landscape and significance across 12 major cancer types. **Nature.** 2013 Oct 17;502(7471):333-9. PMCID: PMC3927368.
2. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, McMichael JF, Schmidt HK, Yellapantula V, Miller CA, Ozenberger BA, Welch JS, Link DC, Walter MJ, Mardis ER, Dipersio JF, Chen F, Wilson RK, Ley TJ, **Ding L.** Age-related mutations associated with clonal hematopoietic expansion and malignancies. **Nat Med.** 2014 Dec;20(12):1472-8. PMCID: PMC4313872.
3. Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MD, Huang KL, Wyczalkowski MA, Jayasinghe R, Banerjee T, Ning J, Tripathi P, Zhang Q, Niu B, Ye K, Schmidt HK, Fulton RS, McMichael JF, Batra P, Kandoth C, Bharadwaj M, Koboldt DC, Miller CA, Kanchi KL, Eldred JM, Larson DE, Welch JS, You M, Ozenberger BA, Govindan R, Walter MJ, Ellis MJ, Mardis ER, Graubert TA, DiPersio JF, Ley TJ, Wilson RK, Goodfellow PJ, Raphael BJ, Chen F, Johnson KJ, Parvin JD, **Ding L**. Patterns and functional implications of rare germline variants across 12 cancer types. **Nat Commun.** 2015 Dec 22;6:10086 PMCID: PMC4703835
4. Niu B, Scott AD, Sengupta S, Bailey MH, Batra P, Ning J, Wyczalkowski MA, Liang WW, Zhang Q, McLellan MD, Sun SQ, Tripathi P, Lou C, Ye K, Mashl RJ, Wallis J, Wendl MC, Chen F, **Ding L.** Protein-structure-guided discovery of functional mutations across 19 cancer types. **Nat Genet**. 2016 Jun. PMID: 27294619

B. Positions and Honors

Positions and Employment

2005 - 2012	Group leader, Medical Genomics Group, The Genome Institute, Washington University in St. Louis, MO
2008 - present	Assistant Director, McDonnell Genome Institute, Washington University in St. Louis, MO
2012 - 2015	Assistant Professor, Department of Medicine, Department of Genetics, St. Louis, MO
2015 - present	Associate Professor, Department of Medicine, Washington University School of Medicine, St. Louis, MO

Other Experience and Professional Memberships

Member, American Association for the Advancement of Science (AAAS)
Member, American Association for Cancer Research (AACR)
Member, American Society of Human Genetics (ASHG)
Research Member, Siteman Cancer Center
Reviewer, Cancer Research UK
Review Panel, UK-ICGC program (prostate and oesophagus projects)
Source Evaluation Group, NCI Cancer Genomics Data Commons
Review Panel/Reviewer, NIH Cancer Genetic Study Section
Review Panel/Reviewer, NIH microbiome sciences and the associated informatics SEP
Co-chair, ICGC Pan Cancer Mutation Calling Group
Co-chair, TCGA Pan Cancer Atlas Oncogenic Process Group
Co-chair, TCGA Sarcoma Analysis Working Group
Steering Committee Member, Clinical Proteomic Tumor Analysis Consortium (CPTAC)
Member, CPTAC Data Analysis Working Group
Co-chair, Pan Cancer Atlas Germline Group
Co-chair, Pan Cancer Atlas Driver Group
Steering Committee Member, NCI Genomic Data Commons (GDC)

Honors

2008	Tomorrow's PI, Genome Technology
2010	Chair for Functional and Cancer Genomics Session Session, The Biology of Genomes
2011	Chair for Population and Personal Genomics Session, Genome Informatics
2013	The Hottest Scientific Researchers of 2012 , Thomson Reuters
2014	Chair for Genomic Alterations of Tumors Session, American Society of Human Genetics
2014	The World’s Most Influential Scientific Minds 2014 , Thomson Reuters
2015	Chair for TCGA Fourth Annual Scientific Symposium, NIH

C. Contribution to Science

1. **Develop and Refine Computational Tools for Genome Research and Clinical Applications.** We have developed an enterprise system for analyzing genomics and proteomics data, with particular emphasis on detection and interpretation. It handles tracking, support, etc. and includes powerful detection and interpretation tools like VarScan, BreakDancer, SomaticSniper, HotSpot3D, Pindel-C, MSIsensor, MuSiC and others, which are widely used for both individual projects and large-scale collaborations, e.g. TCGA, CPTAC, and ICGC. We have further developed these tools under the banner of the “Turnkey Variant Analysis Project” (TVAP) (http://tvap.genome.wustl.edu/). All TVAP programs are publically available through GitHub or SourceForge. We recently summarized the broad landscape of such tools in several reviews.
   a. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, **Ding L.** VarScan: variant detection in massively parallel sequencing of individual and pooled samples. **Bioinformatics.** 2009 Sep 1;25(17):2283-5. PMCID: PMC2734323.

b. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, Koboldt DC, Mooney TB, Callaway MB, Dooling D, Mardis ER, Wilson RK, **Ding L.** MuSiC: identifying mutational significance in cancer genomes. **Genome Res**. 2012 Aug;22(8):1589-98. PMCID: PMC3409272.

c. **Ding L,** Wendl MC, McMichael JF, Raphael BJ. Expanding the computational toolbox for mining cancer genomes. **Nat Rev Genet**. 2014 Aug;15(8):556-70. PMCID: PMC4168012.

d. Ye K, Wang J, Jayasinghe R, Lameijer EW, McMichael JF, Ning J, McLellan MD, Xie M, Cao S, Yellapantula V, Huang KL, Scott A, Foltz S, Niu B, Johnson KJ, Moed M, Slagboom PE, Chen F, Wendl MC, **Ding L**. Systematic discovery of complex insertions and deletions in human cancers. **Nat Med** 2015 Dec 14; PMID: 26657142

2. **Employ the Latest Genomics/Proteomics Technologies to Understand DNA/RNA/protein Interactions and Conduct Protein Biomarker Discovery.** I have been working with Drs. David Fenyö and Ben Raphael's labs to develop computational approaches for integrating genomics and proteomics data for identifying altered networks and pathways in ovarian, colorectal, and breast cancer cohorts. There is of great potential for learning the significant, cancer-specific molecular alterations that have biological and clinical implications. In addition, we have been investigating phosphoproteomics data to evaluate phosphorylation status at the locus and gene levels in tumor samples. One ongoing effort is to discern activated pathways in human tumors using proteomics data and discover druggable targets using protein/phosphoprotein outlier analysis and test treatment responses using patient-derived xenograft models (manuscript under review).

   a. Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, Davies SR, Wang S, Wang P, Kinsinger CR, Rivers RC, Rodriguez H, Townsend RR, Ellis MJ, Carr SA, Tabb DL, Coffey RJ, Slebos RJ, Liebler DC. Proteogenomic characterization of human colon and rectal cancer. **Nature.** 2014 Sep 18;513(7518):382-7. PMCID: PMC4249766. (**Ding L** is a CPTAC consortium member)

   b. Leiserson MD, Vandin F, Wu HT, Dobson JR, Eldridge JV, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, Lawrence MS, Gonzalez-Perez A, Tamborero D, Cheng Y, Ryslik GA, Lopez-Bigas N, Getz G, **Ding L**, Raphael BJ. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. **Nat Genet**. 2015 Feb;47(2):106-14. PMCID:PMC4444046

   c. Ruggles KV, Tang Z, Wang X, Grover H, Askenazi M, Teubl J, Cao S, McLellan MD, Clauser KR, Tabb DL, Mertins P, Slebos R, Erdmann-Gilmore P, Li S, Gunawardena HP, Xie L, Liu T, Zhou JY, Sun S, Hoadley KA, Perou CM, Chen X, Davies SR, Maher CA, Kinsinger CR, Rodland KD, Zhang H, Zhang Z, **Ding L**, Townsend RR, Rodriguez H, Chan D, Smith RD, Liebler DC, Carr SA, Payne S, Ellis MJ, Fenyo D. An analysis of the sensitivity of proteogenomic mapping of somatic mutations and novel splicing events in cancer. **Mol Cell Proteomics** 2015 Dec 2; PMCID:PMC4813688

   d. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F, Kawaler E, Mundt F, Krug K, Tu Z, Lei JT, Gatza ML, Wilkerson M, Perou CM, Yellapantula V, Huang KL, Lin C, McLellan MD, Yan P, Davies SR, Townsend RR, Skates SJ, Wang J, Zhang B, Kinsinger CR, Mesri M, Rodriguez H, **Ding L**, Paulovich AG, Fenyo D, Ellis MJ, Carr SA & the NCI CPTAC. Proteogenomics connects somatic mutations to signaling in breast cancer. **Nature 2016 June;** 534,55–62

3. **Reveal Molecular Processes and Components Underpinning Tumor Initiation, Clonal Evolution, and Metastasis/Relapse.** 1) Somatic variants associated with cancers initiation and progression: At the launch of the Tumor Sequencing Project (TSP), circa 2005, most institutes were using ABI instruments to investigate a few candidate genes in a handful of cancer samples. TSP pushed the envelope by systematically characterizing 188 lung adenocarcinoma genomes. With colleagues from 19 different institutes, I led the sequencing and analysis of 623 selected candidate genes between 2006 and 2008. This is considered a pilot for TCGA. 2) Relapse and metastasis: Relapse and metastasis are signatures of malignancy and are the most common causes of cancer-related death, yet the genetic changes underlying these phenomena are poorly understood. I formed clinical collaborations for some of their earliest genomic studies, including the first observations of subclonal mutations in heterogeneous breast tumors and demonstration of clonal evolution and selection of resistant subclones by chemotherapy treatments in AML. 3) Driver genes and mutations contributing to individual and multiple cancer types: My lab developed a suite of tools collectively called MuSiC for cancer driver mutation/gene discovery and for revealing clinical

associations using large-scale cancer data sets. Utilizing MuSiC, we discovered 127 significantly mutated cancer genes in over 3,000 tumors from 12 major cancer types. Due to its novelty and broad implications, this study was widely covered in both scientific literature (Skipper, Nature Review Genetics, 2013, Ashworth and Hudson, Nature, 2013) and mainstream media (Wall Street Journal, Oct. 16th, 2013; Bloomberg, Oct. 16th, 2013; The Economist, Jan. 14th, 2014). It is representative of my commitment to collaborate with investigators world-wide to answer critical questions in biomedical research.

a. **Ding L,** Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. **Nature.** 2008 Oct 23;455(7216):1069-75. PMCID: PMC2694412.

b. **Ding L**, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. **Nature.** 2010 Apr 15;464(7291):999-1005. PMCID: PMC2872544.

c. **Ding L**, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, Shen D, Harris CC, Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, McGrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, DiPersio JF. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. **Nature.** 2012 Jan 11;481(7382):506-10. PMCID: PMC3267864.

d. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MD, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, **Ding L.** Mutational landscape and significance across 12 major cancer types. **Nature.** 2013 Oct 17;502(7471):333-9. PMCID: PMC3927368.

4. **Establish Interactions between Germline and Somatic genomes to Reveal Joint Contributions to Cancer Predisposition, Initiation, and Progression.** Major advancements have been made in cataloging somatic variations in cancer genomes, but companion analysis for germline changes remains challenging. Recently, we analyzed germline and tumor sequence data from 4,034 samples representing 12 cancer types. Our study, for the first time, revealed a large number of rare germline mutations enriched in the tumors across all 12 cancer types. Further, we found that thirteen genes had a significantly elevated burden of mutations across all 12 cancer types, including BRCA1, BRCA2, ATM, BRIP1, and PALB2 and an additional 21 genes had suggestive evidence of an increased burden, comprising 8.3% of total cancer cases studied. We also investigated pre-existing mutations in hematopoietic stem cells to understand their relevance to cancer mutations and development. We further demonstrated that age-related hematopoietic clonal mosaicism. The implication is elevated incidence of hematological malignancy from the expansion of such mutant clones could occur as life expectancy increases.

a. Kanchi KL, Johnson KJ, Lu C, McLellan MD, Leiserson MD, Wendl MC, Zhang Q, Koboldt DC, Xie M, Kandoth C, McMichael JF, Wyczalkowski MA, Larson DE, Schmidt HK, Miller CA, Fulton RS, Spellman PT, Mardis ER, Druley TE, Graubert TA, Goodfellow PJ, Raphael BJ, Wilson RK, **Ding L.** Integrated analysis of germline and somatic variants in ovarian cancer. **Nat Commun**. 2014;5:3156. PMCID: PMC4025965.

b. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, McMichael JF, Schmidt HK, Yellapantula V, Miller CA, Ozenberger BA, Welch JS, Link DC, Walter MJ, Mardis ER, Dipersio JF, Chen F, Wilson RK, Ley TJ, **Ding L**. Age-related mutations associated with clonal hematopoietic expansion and malignancies. **Nat Med.** 2014 Dec;20(12):1472-8. PMCID: PMC4313872.

c. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, Hedges D, Ma X, Zhou X, Yergeau DA, Wilkinson MR, Vadodaria B, Chen X, McGee RB, Hines-Dowell S, Nuccio R, Quinn E, Shurtleff SA, Rusch M, Patel A, Becksfort JB, Wang S, Weaver MS, **Ding L**, Mardis ER, Wilson RK, Gajjar A, Ellison DW, Pappo AS, Pui CH, Nichols KE, Downing JR. Germline mutations in predisposition genes in pediatric cancer. **N Engl J Med** 2015 Dec 10;373(24):2336-2346 PMC Journal – In Process

d. Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MD, Huang KL, Wyczalkowski MA, Jayasinghe R, Banerjee T, Ning J, Tripathi P, Zhang Q, Niu B, Ye K, Schmidt HK, Fulton RS, McMichael JF, Batra P, Kandoth C, Bharadwaj M, Koboldt DC, Miller CA, Kanchi KL, Eldred JM, Larson DE, Welch JS, You M, Ozenberger BA, Govindan R, Walter MJ, Ellis MJ, Mardis ER, Graubert TA, DiPersio JF, Ley TJ, Wilson RK, Goodfellow PJ, Raphael BJ, Chen F, Johnson KJ, Parvin JD, **Ding L**. Patterns and

functional implications of rare germline variants across 12 cancer types. **Nat Commun** 2015 Dec 22;6:10086 PMC Journal – In Process PMCID: PMC4703835

D. Research Support

Ongoing Research Support

U01 HG006517-04, NHGRI - Ding, Li 02/01/2012-12/31/2016
A Turnkey System for High-throughput Variant Discovery and Interpretation (NCE)
The goal of this project is to make the analysis tools and next-generation pipelines currently in place in large genome centers available to the wider community, both individually and as part of a complete informatics solution.
Role: PI

R01 CA178383-01A1, NCI - Ding, Li 09/08/2014-08/31/2017
Virus Discovery and Characterization In Large-Scale Cancer Sequencing Data
We propose to develop a set of computational methods and analysis strategies for systematic discovery of integrated and episomal, DNA and RNA oncoviruses. We will perform simultaneous analysis of viruses and somatic/germline alterations in the host genome.
Role: PI

R01 CA180006-03, NCI - Ding, Li 02/012013-01/31/2017
Cancer Susceptibility Variant Discovery In High Throughput Sequencing Data
We will develop a computational pipeline for the identification and interpretation of germline alterations in cancer including single nucleotide variants, insertions and deletions (indels), copy number variations, and structural variants. This pipeline will be initially used to systematically analyze whole genome, exome, and RNA-sequencing data from over 5,000 cancer cases already generated by several major efforts and individual research groups and additional cases that will be made publicly available in the next several years.
Role: PI

5U24CA16003502, NIH-NCI (Chen/Ellis/Giddings/Townsend) 4/01/2011-07/31/2016
Cancer Proteomic Center at Washington University and University of North Carolina
The goal of this project is to assist the CPTAC in discovering new biomarkers, verifying their clinical applicability, and ultimately, helping translate selected biomarkers into clinical practice to reduce mortality from cancer.
Role: Co-Investigator

1U41HG007497, NIH - Lee 09/20/2013-08/31/2016
An Integrative Analysis of Structural Variation for the 1000 Genomes Project
We propose to pool expertise from various research groups to provide an integrative analysis of SVs by combining rigorous computational algorithmic development with extensive experimental validation.
Role: Co-investigator

1R01CA172652 , NIH - Chen, K 04/01/2013-03/31/2017
Delineating Heterogeneous Structural Complexity in Cancer Genomes
To fully harness the power of NGS and to facilitate advances toward personalized medicine, we propose to develop a set of novel computational tools for detecting structural variants in heterogeneous cancer genomes.
Role: Co-Investigator

2P01CA101937, NIH-NCI -Ley 09/19/2003-03/31/2018
Genomics of Acute Myelogenous Leukemia
The primary goal in this project is to utilize high throughput genomics technologies to define the commonly mutated target genes in AML that are relevant to clinical outcome.
Role: Co-Investigator