

Development of the Informatics Infrastructure for the MCC

Hai Hu, PhD
Senior Director of Biomedical Informatics
Deputy Chief Scientific Officer
Windber Research Institute

Big Science, Team Science

55. Fedele, C. G. et al. Inositol polyphosphate 4-phosphatase II regulates PI3K/Akt signaling and is lost in human basal-like breast cancers. *Proc. Natl. Acad. Sci. USA* **107**, 22231–22236 (2010).

56. Gewzner, C. et al. Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling. *Cancer Cell* **16**, 115–125 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Sheth and S. Lucas for administrative coordination of TCGA activities, and C. Gunter for critical reading of the manuscript. This work was supported by the following grants from the USA National Institutes of Health: U24CA143883, U24CA143888, U24CA143890, U24CA143799, U24CA143835, U24CA143845, U24CA143882, U24CA143867, U24CA143866, U24CA143848, U24CA144025, U54HG003799, P50CA116201 and P50CA082223. Additional support was provided by the Susan G. Komen for the Cure, of the USA Department of Defense through the Henry M. Jackson Foundation for the Advancement of Military Medicine, and the Breast Cancer Research Foundation. The views expressed in this paper are those of the authors and do not reflect the official policy of the Department of Defense, or US Government.

Author Contributions TCGA research network contributed collectively to this study. Biopsies were provided by tissue source sites and processed by a Biopsies Core Resource. Data generation and analyses were performed by genome sequencing centres, cancer genome characterization centres, and genome data analysis centres. RPPA analysis was performed at the MD Anderson Cancer Center in association with the genome data analysis centre. All data were released through the Data Coordinating Center. Project activities were coordinated by NCI and NHGRI project teams. We also acknowledge the following TCGA investigators of the Breast Analysis Working Group who contributed substantially to the analysis and writing of this manuscript: Project leaders, C.M.P., M.J.E., manuscript coordinator, C.M.P., K.A.H.; data coordinator, K.A.H.; analysis coordinator, C.M.P., K.A.H.; DNA sequence analysis, D.C.K., L.D.; mRNA microarray analysis, K.A.H., C.F.; mRNA sequence analysis, A.G.R., A.C.; DNA methylation analysis, S. Malik, S. Mahurkar, P.W.L.; copy number analysis, A.D.C., M.M.; protein analysis, T.L.K., R.G.W.W., G.B.M.; pathway/integrated analysis, C.J.C., C.Y., J.M.S., C.G.S., G.C., S.R., L.S.; biopsies core resource, T.L., J.B., J.M.G.; pathology and clinical expertise, T.A.K., H.H., R.J.M., J.N.L., T.S., F.W.

Author Information All of the primary sequence files are deposited in CGHub (<https://cghub.ucsf.edu/>); all other data including mutation annotation files are deposited at the Data Coordinating Center (<http://cancergenome.nih.gov/>). Sample lists, data matrices and supporting data can be found at http://tcga-data.nci.nih.gov/docs/publications/bca_2012/. The data can be explored via the ISB Reguome Explorer (<http://explorer.cancergenome.gov/>) and the cBio Cancer Genomics Portal (<http://www.cbioportal.org/>). Data descriptions can be found at <https://www.ncbi.nlm.nih.gov/txdisplay/TCGA/TCGA+data+Primer+and+Supplementary+Methods+Distributions+and+permissions+information+is+available+at+www.nature.com/reprints>. This paper is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence, and the online version of this paper is freely available to all readers. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M.P. (cpereou@med.unc.edu).

The Cancer Genome Atlas Network

Genome sequencing centres: Washington University in St. Louis: Daniel C. Koboldt¹, Robert S. Fulton¹, Michael D. McLellan¹, Heather Schmidt¹, Joelle Kalicki-Verzell¹, Joshua F. McMichael¹, Louchda L. Fulton¹, David J. Dooling¹, Li Ding^{1,2}, Elaine R. Mardis^{1,2,3}, Richard K. Wilson^{1,2,3}

Genome characterization centres: BC Cancer Agency Adria Allari⁴, Miruna Balasandram⁴, Yaron S. N. Butterfield⁴, Rebecca Carlsen⁴, Candace Carter⁴, Andy Chu⁴, Eric Chuah⁴, Hye-Jung E. Chun⁴, Robin J. N. Coope⁴, Noreen Dhalla⁴, Ranabir Guin⁴, Carrie Hirst⁴, Martin Hirst⁴, Robert A. Holt⁴, Darlene Lee⁴, Haiyan L. Li⁴, Michael Mayo⁴, Richard A. Moore⁴, Andrew J. Mungall⁴, Erin Plesance⁴, A. Gordon Robertson⁴, Jacqueline E. Scheel⁴, Arash Shafiqi⁴, Payal Singhania⁴, Jared R. Slododan⁴, Dominik Stoil⁴, Angela Tan⁴, Nina Theissen⁴, Richard J. Vanhoose⁴, Natasja Wye⁴, Thomas Zeng⁴, Yongjun Zhao⁴, Inanc Biro⁴, Steven J. J. Jones⁴, Marco A. Marra⁴, **Broad Institute** Andrew D. Cherniack⁵, Gordon Sakseena⁵, Robert C. Onofrio⁵, Nam H. Pho⁵, Scott L. Carter⁵, Steven E. Schumacher⁵, Barbara Tabak⁵, Bryan Hernandez⁵, Jeff Getz⁵, Huy Nguyen⁵, Andrew Crenshaw⁵, Kristin Ardlie⁵, Ramona Beroukhi⁵, Wendy Winkler⁵, Gad Getz⁵, Stacey B. Gabriel⁵, Michael Meyerson^{5,6,7}, **Brigham & Women's Hospital & Harvard Medical School** Lynda Chin^{8,9,10}, Peter J. Park¹, Raju Kucherlapati¹¹, **University of North Carolina, Chapel Hill** Katherine A. Hoadley^{11,12}, J. Todd Amundson^{11,12}, Cheng Fan¹¹, Fridi J. Turman¹¹, Yan Shi¹¹, Ling Li¹¹, Michael D. Topal^{11,12}, Kiang He Li¹¹, Hann-Hsiang Chao^{11,12}, Alex Pratt^{11,12}, Grace Q. Silva^{11,12}, Michael Adami¹¹, Iglecia^{11,12}, Wei Zhao^{11,12}, Jerry Usary¹¹, Jonathan S. Berg^{11,12}, Michael Adams¹¹, Jessica Booker¹¹, Junyun Wu¹¹, Anisha Gulabani¹¹, Tom Bodenheimer¹¹, Alan P. Hoyle¹¹, Janae V. Simons¹¹, Matthew G. Soloway¹¹, Lisle E. Mose¹¹, Stuart R. Jefferys¹¹, Saisamant Balu¹¹, Joel S. Nelson¹¹, Heidi Hayes¹¹, Charles M. Perou^{11,12}, **University of Southern California/Johns Hopkins** Simeen Malik¹³, Swapna Mahurkar¹³, Hui Shen¹³, Dennis T. Weisenberger¹³, Timothy Triche Jr¹³, Phillip H. Law¹³, Muz S. Bootwalla¹³, Dennis T. Maginotte¹³, Benjamin P. Berman¹³, David J. Van Den Berg¹³, Stephen B. Bayliff¹³, Peter W. Laird¹³

Genome data analysis: Baylor College of Medicine Chad J. Creighton^{22,23}, Lawrence A. Donohoue^{22,23,24,25}, **Broad Institute** Gad Getz²⁶, Michael Noble²⁶, Doug Voet²⁶, Gordon Sakseena²⁶, Nils Gehlenborg^{26,27}, Daniel DiCara²⁶, Junhua Zhang²⁶, Hailie Zhang²⁶, Chang-Jun Wu²⁶, Spring Tingchun Liu²⁶, Michael S. Lawrence²⁶, Lihua Zou²⁶, Andrey Sveshenkov²⁶,裴林 Sun²⁶, Peter Stojanov²⁶, Rui Jing²⁶, Jueok Cho²⁶, Raktim Sinha²⁶, Richard W. Park²⁶, Marc-Danie Nazaire²⁶, Jim Robinson²⁶, Helga Thorvaldsdottir²⁶, Jill Mesirov²⁶, Peter J. Park^{26,29,30}, Lynda Chin^{26,27}

Institute for Systems Biology Sheila Reynolds³¹, Richard B. Krebsberg³¹, Brady Bernard³¹, Ryan Brastler³¹, Timo Erkkila³¹, Joke Lee³¹, Yevgeniy Thorosov³¹, Wei Zhang³¹, Ilya Shmulevich³¹, **Memorial Sloan-Kettering Cancer Center** Giovanni Ciriello³¹, Nils Weinhold³⁴, Nikolaus Schultz³⁴, Jianjiong Gao³¹, Ethan Cerami³⁴, Brian R. Olsen³⁴, Anders Jacobsen³⁴, Pileen Sinha³⁴, B. Arman Aksoy³⁴, Yevgeniy Antipin³⁴, Boris Repp³⁴, Ronglai Sheng³⁴, Barry S. Taylor³⁴, Marc Ladanyi³⁴, Chris Sander³⁴, **Oregon Health & Science University** Pavana Anant³⁵, Paul T. Spellman³⁵, **The University of Texas MD Anderson Cancer Center** Yiling Lu^{36,37}, Wenbin Lu³⁶, Roel R. G. Verhaak³⁶, Gordon B. Mills^{36,39}, Rehan Akbani³⁶, Nianxiang Zhang³⁶, Bradley M. Broom³⁶, Todd D. Cassantini³⁶, Chris Wakefield³⁶, Anna K. Urhun³⁶, Keith Baggerly³⁶, Kevin Coombes³⁶, John N. Weinstein³⁶, **University of California, Santa Cruz/Buck Institute** David Haussler^{41,42}, Christopher C. Benz⁴¹, Joshua M. Stuart⁴¹, Stephen C. Benz⁴¹, Jingchun Zhu⁴¹, Christopher C. Szeto⁴¹, Gary K. Scott⁴¹, Christina Yai⁴¹, Evan O. Pauli⁴¹, Daniel Carlin⁴¹, Christopher Wong⁴¹, Artem Sokolov⁴¹, Janita Thuber⁴¹, Sean Mooney⁴¹, Sam Ng⁴¹, Theodore C. Golstarian⁴¹, Kyle Elliott⁴¹, Mia Griffith⁴¹, Christopher Wilks⁴¹, Singer Ma⁴¹, Brian Craft⁴¹, MCI Chunhua Yan⁴¹, Ying-Hui⁴¹, Daoud Meerzaman⁴⁴

Biopsies core resource: Nationwide Children's Hospital/Biopsies Core Resource Julie M. Gaster-Foster⁴⁵, Jay Bowen⁴⁵, Nilsa C. Ramirez⁴⁵, Aaron D. Black⁴⁵, Robert E. Pyatt⁴⁵, Peter White⁴⁵, Erik J. Zmuda⁴⁵, Jessica M. Tala⁴⁵, Maria Lichtner⁴⁵, Robert Brooker⁴⁵, Myra M. Georger⁴⁵, Paul T. Spellman⁴⁵, Holly A. Harper⁴⁵, Kristen M. LeRae⁴⁵, Lisa J. Wise⁴⁵, Teresa R. Tabler⁴⁵, Cynthia Beck⁴⁵, Thomas Barr⁴⁵, Melissa Hart-Kothari⁴⁷

Tissue source sites: ABS-IUPUI Katie Tarvin⁴⁸, Charles Saller⁴⁹, George Sandusky⁵⁰, Mitchell Chertkov⁵¹, Christina Mary Y. Isard⁵¹, Hengren Tang⁵¹, Roberto Tardano⁵¹, Christine Czerwinski⁵¹, Nicholas Petrelli⁵¹, **Corning** Oleg Dzhanshyan⁵², Mikhail Abramov⁵³, Olga Voronina⁵³, Olga Potapova⁵³, **Duke University Medical Center** Jeffrey R. Marks⁵³, **The Greater Poland Cancer Center** Wiktorina M. Suchorska⁵³, Dawid Murawa⁵³, Witold Kycler⁵³, Matthew Thies⁵³, Konstanty Korski⁵³, Anadiusz Spychala⁵³, Pawel Murawa⁵³, Jacek J. Brzezinski⁵³, Hanna Parz⁵³, Radoslaw Labiak⁵³, Marek Teresiak⁵³, Honorata Tatka⁵³, Ewa Leporowska⁵³, Marja Bogusz-Czerniewicz^{53,57}, Julian Malicki^{53,57}, Andrzej Mackiewicz^{53,57}, Maciej Wzorniewicz^{53,57}, **LSIBio** Jian Van Lee⁵⁸, Bernard Kohler⁵⁸, Nguyen Viet Tien⁵⁸, Richard Thorpe⁵⁸, Nguyen Van Bang⁵⁸, Howard Susman⁵⁸, Bui Duc Phu⁵⁸, Richard Hapak⁵⁸, Nguyen Phi Hung⁵⁸, Tran Viet The Phung⁵⁸, Huynh Thi Thanh⁵⁸, Khuram Zaki Khan⁵⁸, **Modesto Patangano**, **International Genomics Consortium** Robert Penny⁵⁹, David Malley⁵⁹, Eric Curley⁵⁹, Candace Shelton⁵⁹, Peggy Yeh⁵⁹, **Mayo Clinic** James N. Ingle⁶⁰, Fergus J. Couch⁶⁰, Wilma L. Linglie⁶⁰, **MSKCC** Tara A. King⁶⁰, **MD Anderson Cancer Center** Ana Maria Gonzalez-Angulo⁶⁰, Gordon B. Mills⁶⁰, Mary D. Dyer⁶⁰, Shuying Liu⁶⁰, Xiaolong Meng⁶⁰, **Modesto Patangano**, **University of California San Francisco** Frederic Waldman^{61,72}, Hubert Stoppier⁶¹, **University of North Carolina** W. Kimryn Rathmell⁶¹, Leigh Thorne^{61,74}, Mei Huang^{61,74}, Lou Boice^{61,74}, Ashley Hill⁶¹, **Roswell Park Cancer Institute** Carl Morrison⁶¹, Carmelo Gaudio⁶¹, Wiam Bishara⁶¹, **University of Miami** Kelly Daily⁶¹, Sophia G. Egert⁶¹, Mark D. Pogorzelski⁶¹, Carmen Gomez-Fernandez⁶¹, **University of Pittsburgh** Raju Ravi⁶¹, Rohit Bhargava⁶¹, Adam Brufsky⁶¹, **Walter Reed National Military Medical Center** Craig D. Shriver⁶¹, Jeffrey A. Hooke⁶¹, Jamie Leigh Campbell⁶¹, Richard J. Murali⁶¹, Hai Hu⁶¹, Stella Somari⁶¹, Caroline Larson⁶¹, Brenda Deyarmin⁶¹, Leonid Kvecher⁶¹, Albert J. Kovatich⁶¹

Disease working group: Matthew J. Ellis^{82,83}, Tari A. King⁶⁹, Hai Hu⁶⁹, Fergus J. Couch⁶⁸, Richard J. Murali⁶⁹, Thomas Strickie⁸⁴, Kevin White⁶⁹, Olufunmilayo Olopade⁶⁹, James N. Ingle⁶⁸, Chunqing Luo⁶⁹, Yaqin Chen⁶⁹, Jeffrey R. Marks⁶⁹, Frederic Waldman⁶⁹, Maciej Wzorniewicz⁶⁹, Ron Bose^{82,83}, Li-Wei Chang⁶⁹, Andrew H. Beck¹⁰, Ana Maria Gonzalez-Angulo⁸⁰

Data coordination centre: Todd Pihl⁶², Mark Jensen⁶⁷, Robert Steir⁶⁷, Ari Kahn⁶⁷, Anna Chu⁶⁷, Prachi Kothiyal⁶⁷, Zhining Wang⁶⁷, Eric Snyder⁶⁷, Joan Pontius⁶⁷, Brenda Ayala⁶⁷, Mark Backus⁶⁷, Jessica Walton⁶⁷, Julien Baboussi⁶⁷, Dominique Berton⁶⁷, Michael Nicholas⁶⁷, Deepak Srivastava⁶⁷, Rohini Ramani⁶⁷, Stanley Garshick⁶⁷, Peter Kigonyi⁶⁷, Shelley Allosky⁶⁷, Rashmi Sanbhanwal⁶⁷, Sean Barletta⁶⁷, David Pot

Project team: **National Cancer Institute** Margi Sheth⁶⁸, John A. Demchok⁶⁸, Kenna R. Mills Shaw⁶⁸, Liming Yang⁶⁸, Greg Eley⁶⁸, Martin L. Ferguson⁶⁸, Roy W. Tamuzue⁶⁸, Jianhan Zhang⁶⁸, Laura A. L. Dillon⁶⁸, Kenneth Buehler⁶⁸, Peter Fielding⁶⁸, **National Human Genome Research Institute** Bradley A. Ozenberger⁶¹, Mark S. Guyer⁶¹, Heidi J. Sofia⁶¹, Jacqueline D. Paich⁶¹

¹The Genome Institute, Washington University, St. Louis, Missouri 63108, USA. ²Department of Genetics, Washington University, St. Louis, Missouri 63110, USA. ³Stem Cell Center, Washington University, St. Louis, Missouri 63110, USA. ⁴Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z, Canada. ⁵The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁶Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁷Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁸Departments of Cancer Biology and Medical Oncology, and the Center for

RESEARCH ARTICLE

Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁹Department of Medical Oncology and the Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ¹⁰Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹¹Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ¹²The Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹³Department of Genetics, Harvard Medical School and Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ¹⁴Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁵Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁶Fehelmann School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁷Institute for Pharmacogenetics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁸Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹⁹Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ²⁰USC Epigenome Center, University of Southern California, Los Angeles, California 90033, USA. ²¹Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, Maryland 21231, USA. ²²Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA. ²³Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. ²⁴Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. ²⁵Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas 77030, USA. ²⁶The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA. ²⁷Institute for Applied Cancer Science, Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. ²⁸Department of Genomic Medicine, University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. ²⁹Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ³⁰Informatics Program, Children's Hospital, Boston, Massachusetts 02115, USA. ³¹Institute for Systems Biology, Seattle, Washington 98109, USA. ³²Tampere University of Technology, Tampere, Finland. ³³Cancer Genomics Core Laboratory, MD Anderson Cancer Center, Houston, Texas 77030, USA. ³⁴Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³⁵Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³⁶Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ³⁷Oregon Health and Science University, 3181 Southwest Sam Jackson Park Road, Portland, Oregon 97239, USA. ³⁸Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ³⁹Center for Molecular Markers, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁴⁰Department of Biostatistics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁴¹Department of Biomolecular Engineering and Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. ⁴²Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA.

⁴³Buck Institute for Research on Aging, Novato, California 94945, USA. ⁴⁴Center for Bioinformatics and Information Technology, National Cancer Institute, Rockville, Maryland 20852, USA. ⁴⁵The Ohio State University College of Medicine, Department of Pathology, Columbus, Ohio 43205, USA. ⁴⁶The Ohio State University College of Medicine, Department of Pediatrics, Columbus, Ohio 43205, USA. ⁴⁷The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA. ⁴⁸ABS Inc., Indianapolis, Indiana 46204, USA. ⁴⁹ABS Inc., Wilmington, Delaware 19801, USA. ⁵⁰Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. ⁵¹Heleen F. Graham Cancer Center, Christiansa Care, Newark, Delaware 19715, USA. ⁵²Moscow City Clinical Oncology Dispensary 1 and the Central IHC Laboratory of the Moscow Health Department, Moscow 105005, Russia. ⁵³Russian Cancer Research Center, Moscow 115478, Russia. ⁵⁴Careline Inc., South San Francisco, California 94080, USA. ⁵⁵Department of Surgery, Duke University Medical Center, Durham, North Carolina 27710, USA. ⁵⁶The Greater Poland Cancer Centre, Poznan 61-866, Poland. ⁵⁷Poznan University of Medical Sciences, Poznan 61-701, Poland. ⁵⁸LSIBio, LLC, Chestertown, Maryland 21620, USA. ⁵⁹Ministry of Health, Hanoi, Vietnam. ⁶⁰LSIBio, LLC, Karachi, Pakistan. ⁶¹Hue Central Hospital, Hue City, Vietnam. ⁶²Stanford University Medical Center, Stanford, California 94305, USA. ⁶³Center for Minority Health Research, University of Texas, MD Anderson Cancer Center, Houston, Texas 07703, USA. ⁶⁴National Cancer Institute, Hanoi, Vietnam. ⁶⁵Ho Chi Minh City Cancer Center, Vietnam. ⁶⁶Can Tho Cancer Center, Can Tho, Vietnam. ⁶⁷International Genomics Consortium, Phoenix, Arizona 85004, USA. ⁶⁸Mayo Clinic, Rochester, Minnesota 55905, USA. ⁶⁹Department of Surgery, Breast Service, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁷⁰Department of Breast Medical Oncology, The University of Texas, MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁷¹University of California at San Francisco, San Francisco, California 94143, USA. ⁷²Cancer Diagnostics, Nichols Institute, Quest Diagnostics, San Juan Capistrano, California 92675, USA. ⁷³Heleen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California 94115, USA. ⁷⁴UNIC Tissue Procurement Facility, Department of Pathology, UNC Lineberger Cancer Center, Chapel Hill, North Carolina 27599, USA. ⁷⁵Department of Pathology, Roswell Park Cancer Institute, Buffalo, New York 14263, USA. ⁷⁶Department of Pathology, University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Miami, Florida 33136, USA. ⁷⁷University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. ⁷⁸Magge-Womens Hospital of University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15213, USA. ⁷⁹Walter Reed National Military Medical Center, Bethesda, Maryland 20899-5600, USA. ⁸⁰Windber Research Institute, Windber, Pennsylvania 15693, USA. ⁸¹MDR Global, LLC, Windber, Pennsylvania 15693, USA. ⁸²Breast Cancer Program, Washington University, St. Louis, Missouri 63110, USA. ⁸³Department of Internal Medicine, Division of Oncology, Washington University, St. Louis, Missouri 63110, USA. ⁸⁴Institute for Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. ⁸⁵Center for Clinical Cancer Genetics, The University of Chicago, Chicago, Illinois 60637, USA. ⁸⁶Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, USA. ⁸⁷SRSA International, 4300 Far Lakes Court, Fairfax, Virginia 22033, USA. ⁸⁸The Cancer Genome Atlas Program Office, Center for Cancer Genomics, National Cancer Institute, Bethesda, Maryland 20852, USA. ⁸⁹TCGA Consultant, Scotts Creek, LLC, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

WRI as part of the Data Analysis Working Group: H.H., R.J.M.

WRNMMC/WRI as a Tissue Source Site: Craig Dr. Shriver, Jeffrey A. Hooke, Jamie Leigh Campbell, Richard J. Murali, Hai Hu, Stella Somari, Caroline Larson, Brenda Deyarmin, Leonid Kvecher, Albert J. Kovatich

WRI as part of the Disease Working Group: Hai Hu, Richard J. Murali, Chunqing Luo, Yaqin Chen

Big Data—Data Silos or Data Center



MCC Clinical Services and Research Programs

Clinical Departments and Services

- Dermatologic Oncology/*Mohs* Surgery
- Hematology Oncology
- Gynecologic Oncology
- Orthopaedic Oncology
- Pediatric Oncology
- Radiation Oncology
- Surgical Oncology

Disease Specialties

- Breast Cancer
- Gynecologic Cancers
- Pediatric Cancers
- Prostate Cancer

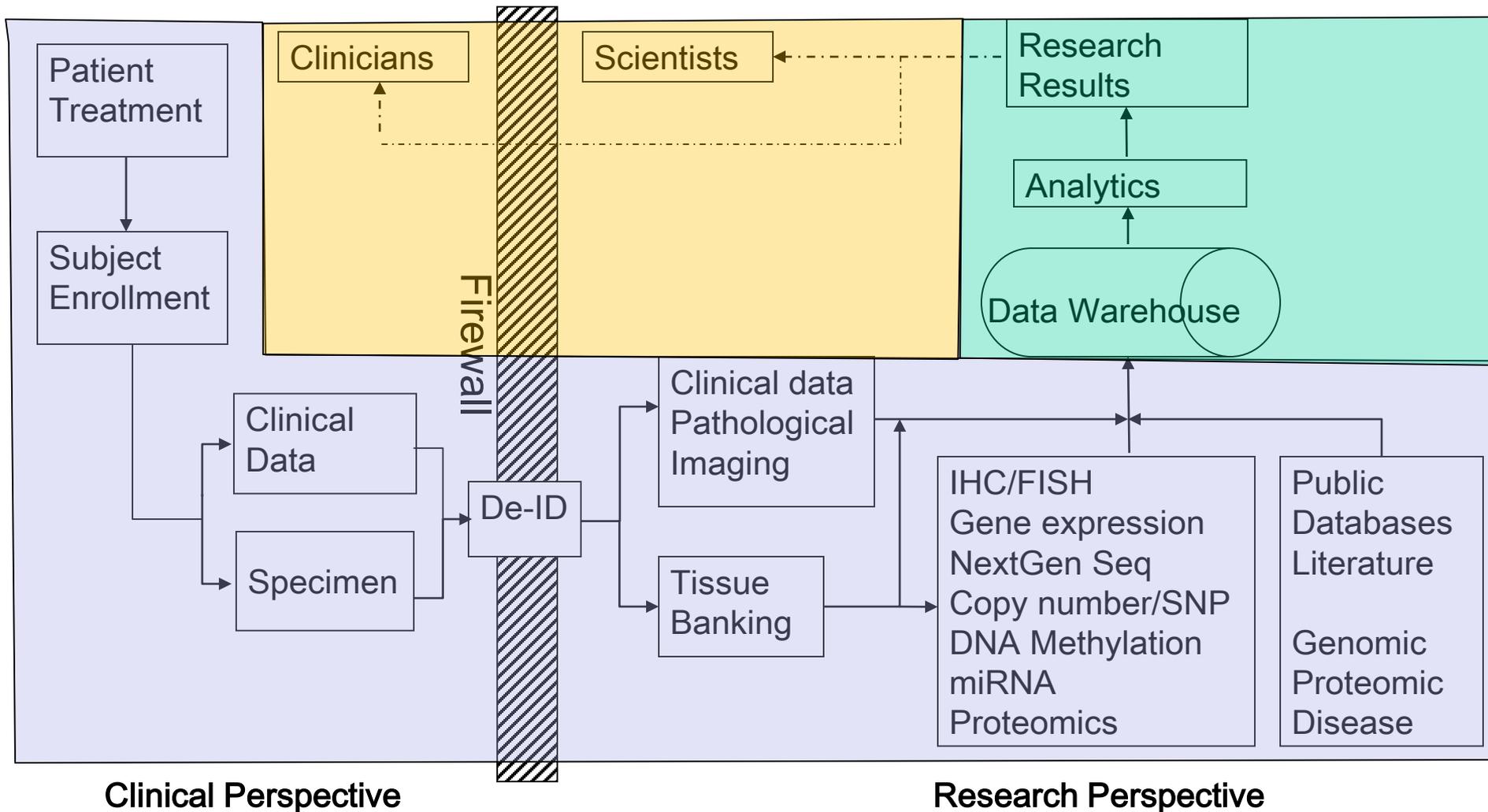
Patient and Family Support Services

- Support and Counseling Services

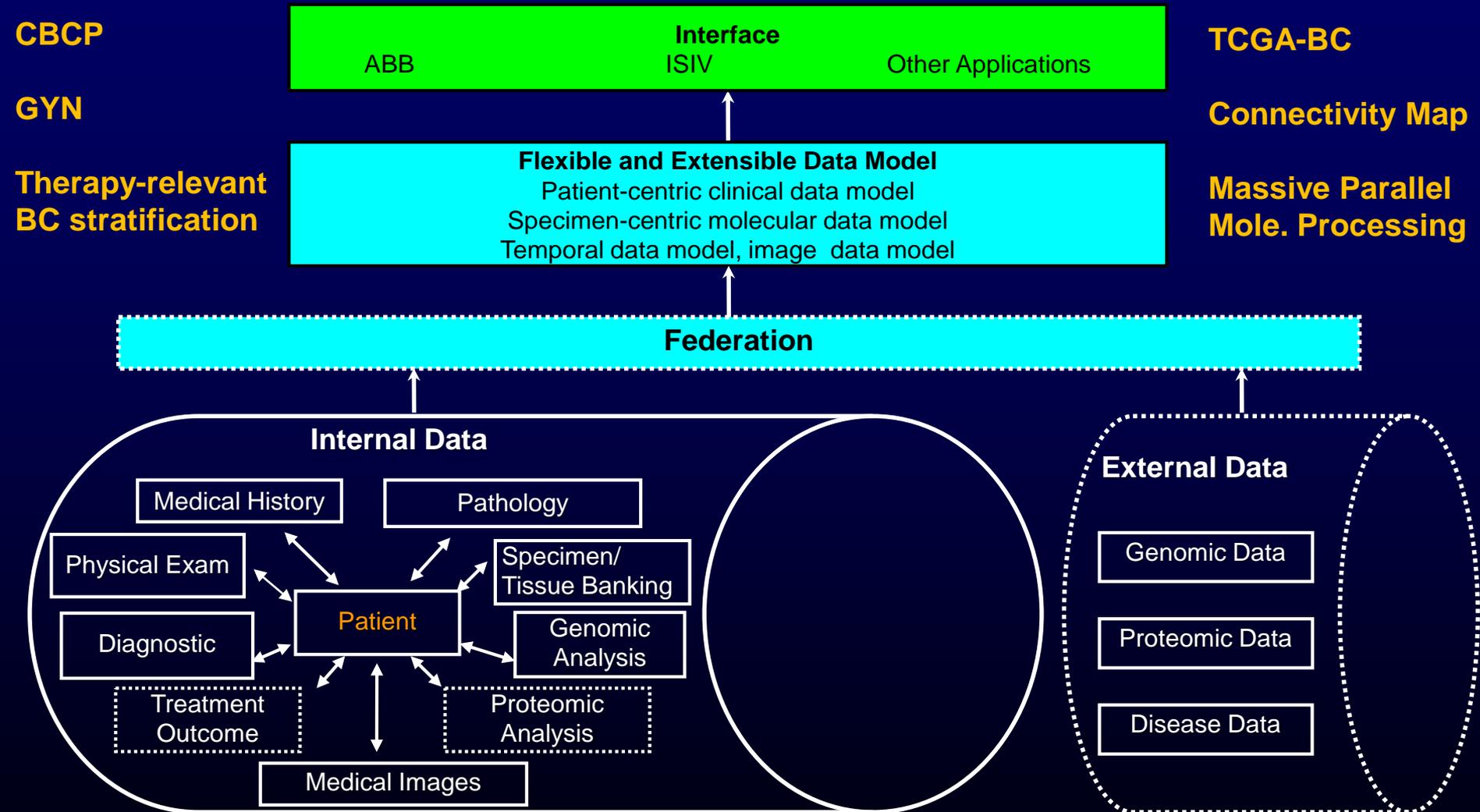
Translational Research Programs and Clinical Trials:

- Women's Malignancies Program (CBCP, GYN, and more)
- Military Population Sciences and Epidemiology Program
- Urologic Malignancies Program (CPDR, and more)
- Inflammation, Infection, Immunity, and Stroma Program (new)

Informatics Infrastructure



Data Warehouse for Translational Research (DW4TR)



Phase I. Vision and Conceptual Design

Vision Definition

Meeting with MCC Executives

Conceptual Design

Meetings with leaders of functional units

Major inputs, processes, and outputs

Interactions between functional units

Existing systems—strengths and weaknesses

Expected infrastructure

Learn from successful NCI CCCs

Big Picture

Phase II. Detailed System Design/Use Case Development

Meeting with technical staff of each functional unit

Clinicopathologic data

Questionnaires and data forms

- A general form for common data elements
- Followed by disease-specific ones

AHLTA/EMR

Biospecimen handling and banking

Molecular study platforms

Clinical Trials

Data tracking

Different levels of details for different types of data

Data Warehousing

What to integrate and how

How data will be used

Scale of data to be managed

System hosting and management

Locally hosted at MCC, or at USUHS

Remotely hosted, offsite

CLOUD

Military firewall

Details
Details
Details

Phase III. Implementation Methods and Execution

Evaluation of

MCC existing systems, extensibilities, and team capabilities

Open source systems

Commercial systems

Timeline and the cost

System implementation model

Build—internal and external

Purchase—external

Combination

Internal development:

- Use of existing infrastructure and open source solutions
- Need to build/contract a development team

External efforts

- Will develop Request For Proposals covering requirements/use cases, and criteria for evaluation
- Invite for on-site presentation and demonstration

Implementation in stages

Develop a thin application layer to reflect the conceptual design of the whole system, enable expansion

Focus on one cancer and go deep

Extend to other cancer studies one by one

Implementation

Summary

Background: Big science
Big data

Experiences: System design
Data tracking
Data warehousing

Development: Big picture
Details
Implementation

Project Team

WRI Biomedical Informatics

Hai Hu, Deputy CSO and Sr. Director

Leonid Kvecher, Data Manager

Ed Elston, IT Manager

Yuanbin (Kevin) Ru, Bioinformatics Scientist

Informatics Management Consultants, LLC

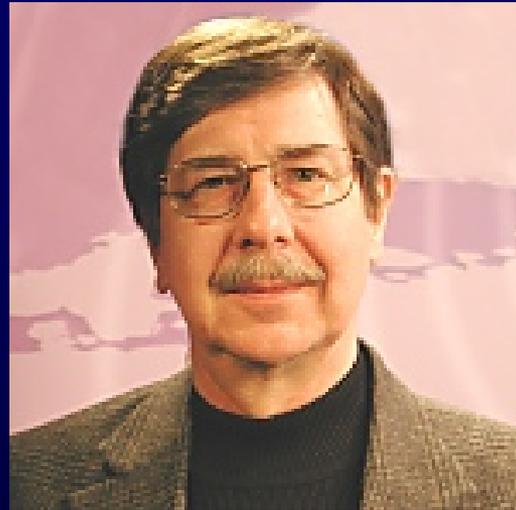
Marc Nodell, President, and the team

MCC

COL Craig D. Shriver, Director

Al J. Kovatich, Jeff A. Hooke

and the team



**Richard J. Mural, WRI Chief Scientific Officer
Died of cancer on May 29, 2014**

Thank you all!

h.hu@wriwindber.org