**BD4BC Summary**

Neither Electronic Health Records (EHR) nor genomics were practical, scalable or affordable technologies a decade ago. Currently, both are undergoing massive expansion, a change brought on by the falling price of technology and by a growing interest in their application to practical problems encountered by clinicians, researchers, industry and government officials. For example, through these technologies, hundreds to thousands of variations in genes, proteins or other molecules can be identified within an individual patient or among a group of patients. If these genomic data can be linked to public and private databases, researchers could sort out the impact of these variations and whether the variants are known to be associated with disease processes.

It is reasonable to predict that both the EHR and genomics technology will be ubiquitous a decade from now - routinely integrated with each other for purposes of research, clinical care, and patient engagement. Furthermore, the "Internet of Things" – devices connected to one another through the internet (e.g. smartphones, tablets, fitness trackers, apps) – has exploded with roughly 3.5 devices connected to the internet per person in 2015 (World population = 7.2 billion; connected devices = 25 billion). These devices are producing and collecting vast amounts of data that could combine with EHR and –omics data to inform the healthcare industry and improve both biomedical research and patient care. However, integration of data has not yet occurred, and until it does, the promise of these technologies will not be fulfilled.

Komen envisions a world in which healthcare is a seamless web of information: patients are informed about their data and are empowered to share it and participate in their health care, data systems are linked and easily accessible, genomics (and other –omics) are universally available and user-friendly, and EHR are connected to other sources of data and provide evidence-based support for clinical decision-making. In this world, many, if not all, would participate in clinical research and the research enterprise would be able to mine these data to address critical questions. Most importantly, fewer people will die from breast cancer and the quality of life will be higher for those living with the disease.

In order to move toward the improvement of breast cancer research and clinical care through data integration and utilization, Komen convened BD4BC: Big Data for Breast Cancer with generous support from the Robertson Foundation. The meeting was held on October 8-9, 2015 at Rockefeller University in New York, NY.

A total of 99 individuals attended BD4BC including leaders in the fields of clinical genetics and genomics, bioinformatics, computational biology, health IT, bioethics, healthcare delivery, learning health systems, cancer epidemiology, breast oncology and other breast cancer providers and breast cancer patient advocates (see attached Participant List). By bringing

leaders in these fields together in a common setting, Komen hoped to increase interactions across these fields and generate new collaborations and partnerships that would accelerate progress.

The meeting was structured to encourage a dialogue around the use of big data for breast cancer research and clinical care that would generate ideas and help Komen identify a way forward. A series of lectures helped to frame the conversation by presenting the following topics[1]:

- *Big Data: Better Treatment - The work of the Early Breast Cancer Trialists' Collaborative Group* presented by Professor Sir Rory Collins from the University of Oxford
- *Innovation in Cancer Therapy* presented by Dr. Henry Friedman from Duke University
- *Big Data for Breast Cancer: A Patient/Advocate Perspective* presented by Dr. Jane Perlmutter (patient advocate)
- *Leveraging Electronic Health Records and Big Data to Create a Data-Fluent Culture for Cancer Medicine* presented by Dr. Mia Levy from the Vanderbilt-Ingram Cancer Center at Vanderbilt University
- *Privacy, Informed Consent, Data Access and Transparent Analysis: PIC_DATA and the Challenges Ahead for Data-Sharing and Breast Cancer Research* presented by Dr. Bob Cook-Deegan from Arizona State University and the Duke Global Health Institute
- *Reflections on Precision Medicine* presented by Dr. Charles Sawyers from Memorial Sloan-Kettering Cancer Center
- *Quantitative Analysis of Oncologic Images* presented by Sir Mike Brady from the University of Oxford

On the second day, participants broke up into self-selected working groups. A total of five working group topics were explored across two sessions. One working group topic was explored in both sessions because the planning committee felt it was important enough and would be of sufficient interest to warrant repeating. Each working group was moderated by a designated leader in that area. Moderators were selected from among the BD4BC Planning Committee members and Komen's Scientific Advisors (Scientific Advisory Board members and Komen Scholars). The working group topics and moderators were:

- *From Bits to Biology: Accelerating Breast Cancer Research via Big Data*
  Moderator: Dr. Gordon Mills (Komen Scholar)
- *Risky Business: Informatics and Risk Modeling for Breast Cancer*
  Moderator: Dr. Elad Gil (BD4BC Planning Committee Member)

[1] All lectures were video recorded and will be available through the BD4BC website and/or Komen's YouTube channel.

- *APIs, Platforms, and Open Data: the Value of Sharing*
  Moderator: Dr. Joe Gray (Komen Scholar)
- *One Size Does Not Fit All: Data Mining for Better Breast Cancer Outcomes*
  Moderator: Dr. Karen Gelmon (Komen Scientific Advisory Board Member)
- *EHRs and Analytics: Improving the Quality and Consistency of Patient Care* (repeated in both working group sessions)
  Moderator: Dr. Mia Levy (BD4BC Planning Committee Member)

Each working group was tasked with the goal of addressing the following questions and coming away with next steps as well as ideas for demonstration projects, when feasible. The questions to be explored included:

- What data are currently available to address the topic?
- What data are needed to address the topic?
- What tools are currently available to address the topic?
- What tools are needed to address the topic?
- Why hasn't this been done yet?
- What are the next steps?

Following the working group sessions, the moderators met with Komen's Chief Scientific Advisors, Dr. George Sledge and Dr. Eric Winer, to report their findings and identify common themes as well as next steps. Each of the working group moderators reported a summary of their group's discussions to all participants (summary slides attached).

Dr. Eric Winer closed the meeting by summarizing the discussion and common themes:

- Big data was defined as the integration of EHR, administrative databases, large data repositories, and genomics and other –omics data.
  - While several of these data sources may be large datasets, it is the integration of data across these sources that make "big data".
- If integrated, big data could be used to generate hypotheses for research, but could also provide definitive answers for health outcomes, quality assessment, practice patterns and complex biologic questions.
- The integration of these data is appealing because:
  - Not all questions can currently be answered in clinical trials because of cost and feasibility.
  - Breast cancer is not a single disease and many subsets may be quite rare making traditional clinical research quite challenging.
  - Big data provide a potentially inexpensive and readily available alternative to clinical trials and clinical studies, if the data are captured and available.

- However, great care needs to be taken because there is huge potential for bias due to issues related to data quality, overlap and mining techniques. (i.e. garbage in – garbage out)
- Big data is still limited in its ability to answer outcome questions. It is not an entity unto itself or a panacea; it is a tool that we need to determine how to use.
- Patients will need to be engaged on multiple levels in data collection. However, many will need to be educated about value of big data.
    - There was great interest in patient-reported outcomes across all groups
- There is a great need for access to these data sources and the integration of data across sources. However, a grassroots movement is needed since there is currently little incentive to share or make data accessible.

Participants felt that Komen could lead in a number of key areas including:

- Assessing patient attitudes about privacy and informed consent
- Influencing the development of future policies to make data accessible and enable integration
- Promoting the collection of patient-reported outcomes data
- Convening scientists and those who "hold the keys" to big data (e.g. payers, EHR providers, patients, policy makers)
- Educating and engaging patients to participate in research
- Issuing RFAs for projects that demonstrate the feasibility of merging and mining data from different sources

Next steps:

- Komen's contractor, Eric Rosenthal, will draft a white paper based on the findings and recommendations from the meeting for submission to *Clinical Cancer Research* (Draft due to Komen on October 31, 2015).
    - If *Clinical Cancer Research* declines publishing the white paper, a submission will be made to the *Journal of Oncology Practice*.
- Recommendations from the meeting will be presented to Komen's Scientific Advisory Board for consideration while planning for the next program cycle (i.e. RFAs and other scientific programs)
- Komen leadership will review recommendations from the meeting and determine strategic priorities.
- As needed, fundraising campaigns will be developed to support priority areas.

BD4BC Summary Attachments:

1) Planning Committee Members
2) Final Participant List
3) Final Agenda
4) Working Group Summary Slides

**BD4BC Planning Committee**

| Name | Org |
|---|---|
| Sir John Bell | Oxford University |
| Elad Gil, PhD | Color Genomics |
| Todd R. Golub, MD | Dana-Farber Cancer Institute |
| Cheryl Jernigan | Patient Advocate (Komen SAB) |
| Mia Levy, MD, PhD | Vanderbilt-Ingram Cancer Center |
| George Sledge, Jr. , MD | Stanford University (Komen CSA) |
| Eric Winer, MD | Dana-Farber Cancer Institute (Komen CSA) |
| Nancy Brinker (*Ex officio*) | Founder, Susan G. Komen |
| Judy Salerno (*Ex officio*) | President & CEO, Susan G. Komen |
| Ruth Brenner (*Ex officio*) | Robertson Foundation |

# BD4BC Participant List

| Name | Institution |
|---|---|
| Abernethy, Amy | Flatiron Health |
| Alderson, Lisa | Invitae |
| Aran, Dvir | University of California, San Francisco |
| Atreya, Ravi | Vanderbilt University |
| Badve, Sunil | Indiana University |
| Barbouche, Michael | Forward Health Group |
| Bator, Julia | Robertson Foundation |
| Beck, Andrew | Harvard University |
| Bhathal, Hirdey | ZibdyHealth |
| Boone, Chris | Health Data Consortium |
| Brady, Mike | University of Oxford (speaker) |
| Branch, Vernal | Patient Advocate |
| Brenner, Ruth | Robertson Foundation |
| Brinker, Eric | Susan G. Komen |
| Brinker, Nancy | Founder, Susan G. Komen |
| Buist, Diana | Group Health Research Institute |
| Carey, Lisa | University of North Carolina |
| Ceille, Kaye | Komen BOD, ZipCar |
| Chandran, Uma | University of Pittsburgh |
| Collins, Rory | University of Oxford (speaker) |
| Collyar, Deborah | Patient Advocate |
| Cook-Deegan, Robert | Arizona State University; Duke University (Speaker) |
| Cox Turner, Stephanie | Komen Staff |
| Crocker, Erica | Komen Staff |
| Culloo, Beth | Komen Staff |
| Custard, Linda | Komen BOD |
| Dai, Hongyue | M2Gen |
| Darnell, Robert | New York Genome Center |
| Deus, Helena | Foundation Medicine |
| Fitzpatrick, Kevin | CancerLinQ, LLC |
| Friedman, Henry | Duke University (Speaker) |
| Gelmon, Karen | BC Breast Cancer Agency; Komen SAB (working group leader) |
| Geoghegan, Cindy | Patients and Partners, LLC, Project Data Sphere, LLC |
| Gil, Elad | Color Genomics (Planning Committee; working group leader) |
| Gilbert, Alan | GE |
| Goodner, Jared | Akido Labs |
| Gordon, Hugh | Akido Labs |
| Gray, Joe | Oregon Health and Sciences University (working group leader) |
| Griffith, Obi | Washington University |
| Habel, Laurel | Kaiser Permanente |

# BD4BC Participant List

| | |
|---|---|
| Haibe-Kains, Benjamin | University of Toronto |
| Hassett, Michael | Dana-Farber Cancer Institute |
| Hatzis, Christos | Yale University |
| Hirsch, Jonathan | Syapse |
| Hoadley, Katherine | University of North Carolina |
| Hood, John | Robertson Foundation |
| Hoshino, Ayuko | Memorial Sloan-Kettering |
| Hubbard, Tom | NEHI |
| Hudis, Clifford | Memorial Sloan-Kettering |
| Jernigan, Cheryl | Patient Advocate; Komen SAB (planning committee) |
| Kean, Marcia | Feinstein Kean Healthcare (FKH) |
| Kovacic, Stefanie | Komen Staff |
| Kurian, Allison | Stanford University |
| Lee, Adrian | University of Pittsburgh Medical Center |
| Lehar, Joseph | Google Life Sciences |
| Levy, Mia | Vanderbilt-Ingram Cancer Center (speaker; working group leader; planning committee) |
| Li, Rongling | NIH eMERGE Network |
| London, Jack | Thomas Jefferson University |
| Lupien, Mathieu | Princess Margaret Cancer Center |
| MacConaill, Laura | Dana-Farber Cancer Institute |
| Mann, Joshua | Share for Cures |
| Margolin, Adam | Oregon Health and Sciences University |
| media | Alice Park, Time Magazine |
| media | Robert Langreth, Bloomberg News |
| Meyer, Anne-Marie | UNC Lineberger Comprehensive Cancer Center |
| Miller, Robert | American Society for Clinical Oncology (ASCO) |
| Mills, Gordon | UT MD Anderson Cancer Center (working group leader) |
| Nicholas, Joanne | Komen Contractor, PR |
| Norton, Larry | BCRF; Memorial Sloan-Kettering Cancer Center |
| O'Donnell, Tim | Icahn School of Medicine at Mount Sinai |
| Patterson, Suzanne | Komen Staff |
| Perlmutter, Jane | Patient Advocate (speaker) |
| Pugh, Trevor | Princess Margaret Cancer Center |
| Rader, Andrea | Komen Staff |
| Ramirez, Amelie | UT Health Science Center San Antonio; SAB |
| Reeves, Gillian | University of Oxford |
| Reffey, Stephanie | Komen Staff |
| Robertson, Alex | Robertson Foundation |
| Robertson, Julian | Founder, Robertson Foundation |
| Robson, Mark | Memorial Sloan-Kettering |
| Rosenthal, Eric | Komen Contractor, Journalist |

# BD4BC Participant List

| | |
|---|---|
| Rubin, Daniel | Stanford University |
| Salerno, Judy | Komen President & CEO |
| Sander, Chris | Memorial Sloan-Kettering |
| Sawyers, Charles | Memorial Sloan-Kettering (speaker) |
| Seib, Aaron | NATE - The National Association for Trusted Exchange |
| Shak, Steven | Genomic Health |
| Sledge, George | Stanford University; Komen CSA |
| Sotiriou, Christos | Jules Bordet Institute |
| Spears, Patty | Patient Advocate; Komen AIS |
| Stern, Dr. & Mrs. | Robertson Foundation |
| Stovall, Carrie | Komen Staff |
| Tantawi, Linda | CEO, Komen NYC |
| Tessier-Lavigne, Marc | President, Rockefeller Univ |
| Tirado-Ramos, Alfredo | UT Health Science Center San Antonio |
| Troester, Melissa | University of North Carolina |
| Winer, Eric | Dana-Farber Cancer Institute; Komen CSA |
| Wolodzko, Victoria | Komen staff |
| Zirkle, Maryan | Patient-Centered Outcomes Research Institute (PCORI) |

# BD4BC

## BIG DATA FOR BREAST CANCER

**October 8-9, 2015**

susan g. komen®

#BD4BC

Susan G. Komen®, with support from the

ROBERTSON FOUNDATION

is convening content experts from all relevant fields to discuss the challenges and opportunities regarding the integration of Electronic Health Records (EHR) with genomics, proteomics and other "-omics" to improve breast cancer research and treatment.

# BD4BC has Gone Mobile!

Download the official mobile app of **#BD4BC** and maximize your meeting experience:

- Search **BD4BC** in Apple, Android, Blackberry and Windows app stores!
- View/add agenda sessions to personal calendar
- Receive meeting alerts and updates in real time
- View attendee directory
- Digital infobooth
- Venue map

## The Impact of a Promise

Susan G. Komen is the world's largest breast cancer organization, funding more breast cancer research than any other nonprofit outside the U.S. government, while providing real-time help to those facing the disease. Since its founding in 1982, Komen has funded more than $889 million in research and provided $1.95 billion in funding to screening, education, treatment and psychosocial support programs serving millions of people in more than 30 countries worldwide. Komen was founded by Nancy G. Brinker, who promised her sister, Susan G. Komen, that she would end the disease that claimed Suzy's life.

## General Meeting Information

### Location

Rockefeller University, 1230 York Avenue, New York, NY 10065

### Conference Video Policy

General sessions will be recorded and may be posted on **komen.org/bd4bc** following the meeting. If you do not wish to be on video or photographed please stop by the registration desk to make arrangements.

### No Smoking

In keeping with Susan G. Komen's policy and our vision of a world without breast cancer, the BD4BC meeting is strictly a non-smoking event. Thank you.

## Shuttle Information

Shuttles will be provided between the Bentley Hotel and Rockefeller University throughout the meeting. Please reference the schedule below.

### Thursday, October 8

12:00 p.m. - 2:00 p.m. - shuttles will make continuous loops between the Bentley Hotel and Rockefeller University.

Shuttles will also be available immediately following the conclusion of dinner for transport to the Bentley Hotel.

### Friday, October 9

6:45 a.m. - 8:00 a.m. - shuttles will make continuous loops between the Bentley Hotel and Rockefeller University.

Shuttles will also be available immediately following the conclusion of the meeting to Newark, JFK and LaGuardia airports and Penn Station. Shuttles will be outside the front entrance to Welch Hall.

# BD4BC

## AGENDA | THURSDAY, OCTOBER 8

12:30 p.m. - 1:30 p.m.
**Lunch**
▶ **Abby Aldrich Rockefeller Dining Room**

1:30 p.m. - 2:00 p.m.
**Transition to Carson Auditorium**

2:00 p.m. - 2:15 p.m.
**Welcome Session**
▶ **Carson Auditorium**
Nancy G. Brinker
Founder, Susan G. Komen
Marc Tessier-Lavigne, Ph.D
President, Rockefeller University

2:15 p.m. - 2:45 p.m.
**Overview**
▶ **Carson Auditorium**
George W. Sledge, Jr., M.D.
Stanford University School of Medicine

2:45 p.m. - 5:30 p.m.
**Framing the Conversation**
▶ **Carson Auditorium**

**Big Data: Better Treatment**
Professor Sir Rory Collins
University of Oxford

**Innovation in Cancer Therapy**
Henry S. Friedman, M.D.
Duke University

**Big Data for Breast Cancer: A Patient/Advocate Perspective**
Jane Perlmutter, Ph.D.
Patient Advocate

**Leveraging Electronic Health Records and Big Data to Create a Data-Fluent Culture for Cancer Medicine**
Mia Levy, M.D., Ph.D.
Vanderbilt-Ingram Cancer Center

**Privacy, Informed Consent, Data Access and Transparent Analysis: PIC_DATA and the Challenges Ahead for Data Sharing and Breast Cancer Research**
Robert Cook-Deegan, M.D.
Duke Global Health Institute

5:30 p.m. - 6:00 p.m.
**Continuing the Conversation: Key Questions**
▶ **Carson Auditorium**

6:00 p.m.
**Reception and Dinner**
▶ **Welch Hall**

**Introduction**
Judy Salerno, M.D., M.S.
President and CEO, Susan G. Komen®

**Reflections on Precision Medicine**
Charles L. Sawyers, M.D.
Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center

# BD4BC

## AGENDA | FRIDAY, OCTOBER 9

**7:00 a.m. - 8:00 a.m.** **Breakfast Buffet**
- Welch Hall – Great Hall

**8:00 a.m. - 8:30 a.m.** **General Session**
- Welch Hall – Great Hall

**8:30 a.m. – 10:00 a.m.** **Working Group Session #1**

**From Bits to Biology:**
**Accelerating Breast Cancer Research via Big Data**
- Welch Hall – Great Hall

Moderator: Gordon Mills, M.D., Ph.D.

**Risky Business:**
**Informatics and Risk Modeling for Breast Cancer**
- Adler Room

Moderator: Elad Gil, Ph.D.

**EHRs and Analytics:**
**Improving the Quality and Consistency of Patient Care**
- Audubon Room

Moderator: Mia Levy, M.D., Ph.D.

**10:00 p.m. – 10:30 p.m.** **Break**

**10:30 p.m. – 12:00 p.m.** **Working Group Session #2**

**APIs, Platforms, and Open Data: the Value of Sharing**
- Adler Room

Moderator: Joe Gray, Ph.D.

**One Size Does Not Fit All:**
**Data Mining for Better Breast Cancer Outcomes**
- Welch Hall – Great Hall

Moderator: Karen Gelmon, M.D.

**EHRs and Analytics:**
**Improving the Quality and Consistency of Patient Care**
- Audubon Room

Moderator: Mia Levy, M.D., Ph.D.

**12:00 p.m. – 12:30 p.m.** **Lunch**
- Atrium, Research Building

**2:00 p.m. - 3:30 p.m.** **Moving Forward:**
**Breakout Group Reports and Final Recommendations**
- Carson Auditorium

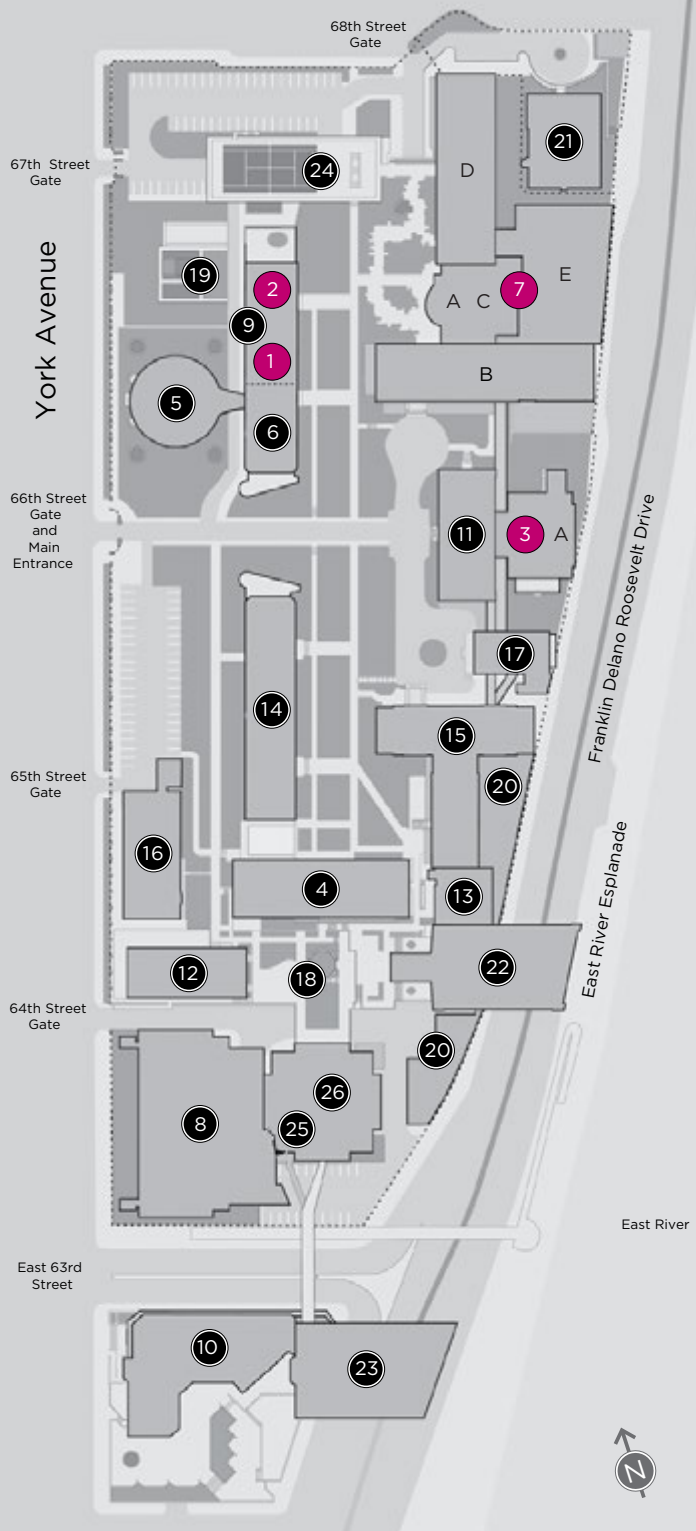
Moderator: Eric P. Winer, M.D.

**3:30 p.m. - 4:00 p.m.** **Closing Remarks**
- Carson Auditorium

Eric P. Winer, M.D.
Dana-Farber Cancer Institute

# The Rockefeller University Campus Map

1. **Abby Aldrich Rockefeller Hall**
2. **Abby Dining Room**
3. **Bass Center**
   - **A. Welch Hall (Markus Library)**
4. Bronk Laboratory
5. Caspary Auditorium
6. Caspary Hall
7. **Collaborative Research Center**
   - **A. Carson Family Auditorium**
   - B. Flexner Hall
   - C. Greenberg Building
   - D. Smith Hall
   - E. Smith Hall Annex
8. Comparative Bioscience Center
9. Faculty and Students Club
10. Faculty House
11. Founder's Hall
12. Fricke Hall
13. Gasser Hall
14. Graduate Students Residence
15. Hospital (Heilbrunn Outpatient Research Center)
16. IT Pavilion
17. Nurses Residence
18. Peggy Rockefeller Plaza
19. Philosophers Garden
20. Power House
21. Presidents House
22. Rockefeller Research Building
23. Scholars Residence
24. Tennis Court
25. Weiss Café
26. Weiss Research Building



## Susan G. Komen® Scientific Advisory Board

### Chief Scientific Advisors

**George Sledge, Jr., M.D.**
Stanford University School of Medicine
Stanford, CA

**Eric Winer, M.D.**
Dana-Farber Cancer Institute
Boston, MA

### Scientific Advisory Board Members

**Carlos Arteaga, M.D.**
Vanderbilt-Ingram Cancer Center
Nashville, TN

**Myles Brown, M.D.**
Dana-Farber Cancer Institute
Boston, MA

**Powel Brown, M.D., Ph.D.**
The University of Texas MD Anderson Cancer Center
Houston, TX

**Karen Gelmon, M.D.**
British Columbia Cancer Agency
Vancouver, Canada

**Cheryl Jernigan, CPA, F.A.C.H.E.**
Advocate in Science
Komen Greater Kansas City
Kansas City, MO

**Amelie Ramirez, Dr.PH**
The University of Texas Health Science Center at San Antonio
San Antonio, TX

## Susan G. Komen Board of Directors

Jane Abraham

Nancy G. Brinker
Founder

Kaye Ceille

Linda Custard
Chairman

Alan Feld

Janet Franitz
Affiliate Board Representative

Dan Glennon

Melissa Maxfield

Dr. Olufunmilayo (Funmi) Olopade

Connie O'Neill

John D. Raffaelli

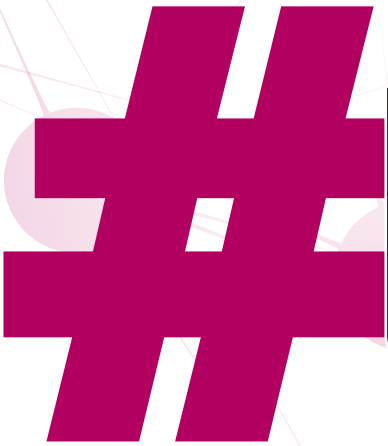
Meghan Shannon
Affiliate Board Representative

Linda A. Wilkins

# BD4BC

BIG DATA FOR BREAST CANCER

@SusanGKomen



**Follow Along on Social Using #BD4BC**

We encourage you to engage with Komen and other attendees by joining the BD4BC social conversation, using **#BD4BC**. Share inspiring mission moments, interesting and educational facts, and kudos to those making a difference.

susan g. komen®

komen.org

# From Bits to Biology: Accelerating Breast Cancer Research via Big Data

Moderator: Gordon Mills

# From Bits to Biology

## Accelerating breast cancer research via Big Data

**Define Big Data**

- **Big Data is not an entity or panacea, it is a tool that we need to determine how to use**
- **Unsupervised big data: collect all data on all patients, model systems and bench data**
  - **Emergent properties**

- **Patient driven/oriented question driven demonstration study**
  - **To demonstrate utility of the process and importantly learn the limitations and challenges required to do an "all of all" study in an effective manner**
  - **Link question to type, quantity and quality of data**

# From Bits to Biology

## Accelerating breast cancer research via Big Data

- Patient driven/oriented demonstration study
  - Tractable question
    - Near term impact
  - Build on current resources
    - Data already available
    - Resources such as registries “SEER”
    - Randomized trial linked to multi scale longitudinal data
  - High quality outcomes data critical
    - Develop and implement approaches to obtain longitudinal treatment and outcomes data
    - Compare reliability
      - Clinical trials
      - EHR
      - Patient entered  support evaluated and improve Aps

# From Bits to Biology

## Accelerating breast cancer research via Big Data

- Patient driven/oriented demonstration study
  - Link question to be asked to the types, quantity and quality of data that can be obtained and aggregated
  - Patient, model and bench data
  - Types of research questions that could benefit from a big data approach today
    - Patient driven question
    - Germline influence on outcomes
    - Rare diseases
    - Rare events Unusual responder/resistor

# From Bits to Biology

## Accelerating breast cancer research via Big Data

### Data analysis and utility

- Support approaches to aggregated, visualize,
  patient, model and bench data
- Support or engage software development and implementation
  Google, Apple, patient oriented apps

- Scalability, sustainability, incentivization, culture, interoperability

- How will we pay for it

# Risky Business: Informatics and Risk Modeling

Moderator:  Elad Gil

# Outcome of session

Provide big data centric model of breast cancer risk (start simple, with existing data sets):

- Clinician output:
  - Who is truly low risk?
  - Who is truly at risk:
    - 5 year, lifetime
    - By subtype
- Patient output:
  - Tool that allows individual to understand their risk
  - What it means
  - What they can do about it

Non-goal: risk models related to  treatment/care

# Data to be used

Use retrospective data only

- Less expensive
- Data may already exist
- OK if data does not perfectly overlap for all data sets as long as some overlaps (big data)

Data types:

- Data includes radiological imaging, outcomes
- Additional data: personal & family history, breast density, genomics etc.

# Big Data Approach Has Some Differences From Traditional Risk Modeling

Data sets that can be used

Data mining

# Example Potential Data Sources

- DMIST, TMIST
- American College Radiology
- Kaiser
- Breast Cancer Surveillance/Screening Consortium
- WHI
- Geissinger
- NHS: UK, NL
- BCFR
- KONFAB
- Large health systems with historical imaging data + outcomes

# Next steps

- **Look at existing data sets and define key information that each data set has**
- **Define protocol for future**
  - Ensure e.g. T-MIST and other collect info that could feed into model
- **RFP from funding sources:**
  - Apply to data consortium to contribute data and participate in study
- **Long term funding:**
  - Create large, normalized, high quality data set definitively

# APIs, Platforms and Open Data: The Value of Sharing

Moderator:  Joe Gray

# APIs, platforms and open data – the value of sharing

- What do we want to share?  Study design to support causal inference.
- Linking to model systems – biological co-trials to support inference development
- Data standards – Supporting community efforts to define
- Platforms – central vs. federated.  Local vs. central control.
- Data security – who decides?
- Compute speed – answers in time to support clinical decision making
- Transmission speed – could we move everything if we wanted to.
- Community learning (training and test sets)
- Total cost of operation.  Keeping at the state of the art

# Data platforms, APIs, open data

- Komen advocate for reduced data security restrictions (Common rule, HIPPA mitigation)
- Komen organize a sustainable business model to stimulate platform developers
- Komen issue RFA for a demonstration patient centric clinical trial
  - Goals
    - Test hypothesis that patient entered data is high quality and sustainable over the long term
    - Make cohort available to platform developers
  - Specific question – For example, 100 pts with HER2+ disease treated with Herceptin
    - Outcome markers, difference in sites of metastasis based on ER status
  - Organize data (EHR, outcome, Omics, Image)
  - Collect patient outcome and Dx info using diverse input platforms including disruptive technologies
  - Return information to the patients as an incentive to enter more data

# One Size Does not Fit All
# Data Mining for Better Breast Cancer Outcomes

Moderator: Karen Gelmon

# Working Group Structure

- Define Topic
- What data are currently available?
- What data are needed?
- What tools are currently available?
- What tools are needed?
- Why hasn’t this been done yet?
- What are the next steps?

# Patients

- Engaged in the process
- Privacy issues / de-indentified
- Patient related outcomes
- Engaging them through other activities – Komen races
- Long term follow up
- Access electronically
- Keep communicating, Keep participating
- Concerns expressed about bias – are the persons who contribute representative?
- Concerns about quality of data – need to use tools that can enhance data

# Databases

- Existing – clinical trials, tumor registries, SEER, others
- Linking of data to genomics, imaging, patient related outcomes
- Evolving data – need to reconfigure
- Tells us little about individual / Utility
- Not the correct people looking at the data
- Persons able to tell us which questions can be answered
- Bringing in different people/ smarter people/ different disciplines/ folks not like us

# Deliverables

- Outcome – persons living longer and BETTER
- Komen as the facilitator of linking databases to get at outcomes that cannot be assessed in other ways
- Komen as advocate around the privacy issues and linking issues that are hampering research
- Opportunistic with research
- Links to productivity/ health outcomes/ health economics/
- Participate with one's data to SAVE LIVES

# Working Group Session: EHR and Analytics

Moderator: Mia Levy

Presented by Alfredo Tirado-Ramos

# main discussion points

- Komen strengths
- Komen challenges
- Use cases
- Opportunities

# main discussion points

- **Komen strengths**
  - Activate patients
  - Educate patients/advocates with accurate messages
  - Grants for research
  - Policy change advocacy

# main discussion points

- **Komen challenges (i)**
  - Data normalization and modeling
  - Access to EMRs
  - Providers willing to share with hesitation
  - Messy data
  - Many platforms
  - Data hording
  - It is unknown what data is important to patients
  - Sustainability

# main discussion points

- **Komen challenges (ii)**
  - Should Komen push for the creation of new types of retrospective/prospective data collection tools?
  - Do providers believe the data analytics results?
  - Need large sample sizes with diversity
  - Need full deep datasets from EMRs

# main discussion points

- **Use cases (i)**
  - Quality metric extraction & validation
    - What is important to patients?
  - Patient centered vs patient driven
    - Evaluate rare disease market place solutions
    - Tools and data are lacking
    - Data elements need to be defined
  - Educate “free the data”, “give me my data”, “allow me to donate”

# main discussion points

- **Use cases (ii)**
  - Short term: institution donated data
    - varies in practice
    - quality as defined by other groups
    - what is EHR and what data is in it?
    - how do I get across to my data?
    - what are patients willing to do?
  - Long term: patient donated data
    - what is important outcomes for patients
    - demonstrate value of big data via demo projects

# main discussion points

- **Opportunities**
  - Facilitate patient access to records
  - Manage the hype
  - Facilitate data integration across relevant sources
  - Data informing personalized screening strategies and risk reduction

# Summary

Eric Winer
Komen Chief Scientific Advisor
Dana-Farber Cancer Institute

# What Is BIG DATA?

- Electronic health records
- Administrative data bases
- Other large data repositories
- Genomics and other -omics
- Ideally, both clinical data and -omic data
- Integrate as much as possible

# How Can We Use Big Data?

- Hypothesis generation
- Definitive findings

- Outcomes
- Quality assessment
- Practice patterns
- Biologic questions

# Why Is BIG DATA Appealing?

- Not all questions can be answered in clinical trials because of cost and feasibility
- Breast cancer is not a single disease and many subsets may be quite rare
- Potentially inexpensive and readily available alternative to trials if the data are captured and available
- BUT….great care needs to be taken because of huge potential for bias
- Above all else do no harm = GARBAGE IN, GARBAGE OUT

# Common Themes

- BIG DATA is still limited in its ability to answer outcome questions…a tool, not a panacea
- Many believe we need to very cautious
- We will need to engage patients on multiple levels in data collection and many may need to be educated about value of BIG DATA

# Common Themes (cont.)

- In the short term, role for clinical studies which encourage patients to enroll on line and are not institutionally based
- Great interest in patient-reported outcomes
- We need to work hard to link clinical data, omics, and patient-reported outcomes

# What Can Komen Do?

- Assess patient attitudes about privacy and informed consent
- Influence the development of future policies
- Promote collection of patient-reported outcomes data
- Bring scientists and those who “hold the keys” to big data to come together
- Engage patients to participate in research
- Issue RFA

# Thanks

- Planning Committee
- Rockefeller University
- Nancy Brinker and Julian Robertson
- Komen staff
- Stephanie Reffey (a workhorse and saint)

- All of you