Confluence is a research project funded by the National Cancer Institute (NCI) Intramural Research Program at the Division of Cancer Epidemiology and Genetics (DCEG), to build a large research resource for the scientific community by 2021, to study the genetic architecture of breast cancer.

Breast cancer affects over 1.7 million women and globally causes more than 500,000 deaths [1]. Genome-wide association studies (GWAS) are making important contributions to our understanding of the etiology of this disease [2-7], to our ability to identify individuals at different levels of risk in the population for stratified prevention [8-12], and are providing clues on the genetic underpinnings of disease prognosis, response to treatment and toxicities [13-17]. Although current GWAS have identified over 180 common susceptibility loci, heritability analyses indicate that breast cancer is a highly polygenic disease with thousands of common genetic variants of small effects, and that increasing sample sizes will generate new discoveries. In addition, our knowledge primarily derives from studies of women of European or East Asian ancestry, and the most common breast cancer subtypes, primarily estrogen receptor positive breast cancers. Thus, additional research in understudied populations and of uncommon subtypes, such as aggressive triple negative tumors, is particularly needed.

Confluence aims to address this research gap by building a large research resource of over 300,000 cases and 300,000 controls of different races/ethnicities—doubling current sample sizes to study the genetic architecture of breast cancer. This will be accomplished by the confluence of existing and new genome-wide genotyping data to be generated through this project.

Specific aims

1) To discover susceptibility loci and advance knowledge of the etiology of breast cancer overall and by subtypes
2) To develop polygenic risk scores and integrate them with known risk factors for personalized risk assessment for breast cancer overall and by subtypes
3) To discover loci for breast cancer prognosis, long-term survival, response to treatment, and second breast cancer

This resource will also allow us to address a broad range of scientific questions in breast cancer genetics and will serve as the basis for further studies that will require the collection of additional data.

Eligibility

To be eligible to participate, studies with cases of *in situ* or invasive breast cancer (females or males) must have:

- Existing genome-wide genotyping data, or germline DNA available for new genotyping, or blood/buccal samples for germline DNA isolation and genotyping.
- Core phenotype data (age, sex, ethnicity, family history, and tumor ER status and grade). Note: complete core data is not required, if it has not been collected by the study.
- Ethics approval and consent for genetic studies and data sharing.

Studies can have a wide range of study designs, including case-control studies, prospective cohorts, clinical case series, clinical trials, or special cohorts such as retrospective cohorts of *BRCA1/2* mutation carriers, or carriers of mutations in other established breast cancer susceptibility genes (e.g. *ATM*, *CHEK2*, *PALB2*).

Study participation

For studies participating in existing breast cancer GWAS consortia (e.g. Breast Cancer Association Consortium (BCAC), Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA), the African-Ancestry Breast Cancer Genetic Study (AABCGS), Latin America Genomics Breast Cancer Consortium (LAGENO-BC), and Male Breast Cancer GWAS), they can participate in Confluence through their existing consortium. Studies not already in a consortia can participate by joining an existing consortium, forming a new group/consortia or through a direct collaboration with DCEG, NCI.

INTERESTED IN PARTICIPATING? Complete Study Inventory at https://dceg.cancer.gov/Confluence
HAVE QUESTIONS? Email us at ConfluenceProject@mail.nih.gov

Genotyping

Studies can contribute existing genome-wide genotype data from already scanned samples or new samples to be scanned. Contribution of existing genotype data can be done through the consortia after approval from individual studies. For studies that require new genotyping, Confluence will cover the costs of sample shipment and materials (plates/tubes), DNA extractions (if needed), DNA quantitation/QC, return of left-over DNA (if requested), genotyping (considering Illumina Infinium Global Screening Array (GSA) or Multi-Ethnic Genotyping (MEGA) Array) and return of genotyping files to contributing studies. However, Confluence will not be able to cover the costs for sample retrieval, preparation, and aliquoting by individual studies.

Governance

The organizational structure is designed to ensure close involvement of participating studies and consortia in the governance, oversight and operations of Confluence and will include the following:

1. *Scientific Steering Committee (SSC)* co-chaired by the DCEG Deputy Director (Dr. Montserrat Garcia-Closas) and the BCAC lead (Prof. Doug Easton) and includes representatives of all participating consortia, and other large contributing studies or groups of studies (see page 3). The SSC mission is to bring together representatives of different collaborative groups, provide scientific expertise, contribute to the development of the research plan and provide oversight of the research resource for use by the wider scientific community. The SSC will report to the director of DCEG (Dr. Stephen Chanock), the funding source for Confluence.
2. *External Advisory Group* will be formed by international experts in GWAS and advocates to provide logistical and scientific advice to Confluence.

DCEG will be responsible for the overall coordination of Confluence, including management, integration and analyses by participating groups and consortia. However, each consortium will be responsible for the management and governance of data from their member studies, according to their rules and regulations.

Scientific Review

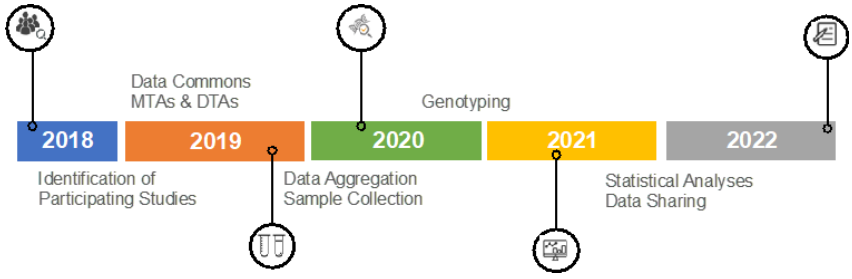
The Confluence project Research Protocol has been reviewed by the NCI intramural review process that is used by projects funded by the NCI Intramural Research Program. The *NCI Board of Scientific Counselors for Clinical Sciences and Epidemiology* will provide external review of the progress and scientific output of Confluence.

Data Sharing Plan

A. Controlled data access through the Confluence Data Platform by eligible researchers: Access to individual-level data for specific analyses will be possible through the Confluence Data Platform that will be securely hosted in a Cloud environment. This platform will be designed to facilitate data intake, access, governance, visualization and analyses of data following FAIR principles [18], and will be compatible with individual study IRB and consortia policies. This approach will greatly facilitate collaborative analyses across multiple groups in a shared analytical space. Data ownership will stay with the individual studies. For studies contributing data to Confluence through a consortium, the custodian of the data will be the consortium data coordinating center and data access will be governed by the consortium Data Access Coordinating Committee (DACC). A Memorandum of Understanding (MoUs) between the institution providing the data (i.e. individual studies) and the data coordinating center will establish the terms under which data and samples will be shared, and will enable the data coordinating center to provide access of the data to eligible researchers for specific projects, on behalf of the institution providing the data under the following process. The primary approach for data analyses will be to move the code to the data in the Cloud (no downloads).

B. Public data access through data archive (e.g. dbGAP, EGA): In accordance to the NIH Genomic Data Sharing policy, individual–level genotyping data generated using funds from Confluence must be submitted for public access to an NCI-approved Data Archive such as the NIH database of Genotypes and Phenotypes (dbGaP), or the European Genome-phenome Archive (EGA), along with associated core phenotype data.

Approximate Timeline



Confluence Project
Scientific Steering Committee

Member		Affiliation	Representing/coordinating
Montserrat	Garcia-Closas	Division of Cancer Epidemiology and Genetics, USA	Co-Chair (DCEG)
Doug	Easton	Cambridge University, England	Co-Chair (BCAC)
Jonas	Almeida	Division of Cancer Epidemiology and Genetics, USA	Data Science
Antonis	Antoniou	Cambridge University, England	CIMBA/Statistical Genetics
Jenny	Chang-Claude	German Cancer Research Center DKFZ, Heidelberg, Germany	BCAC risk factor working group
Nilanjan	Chatterjee	Johns Hopkins University, USA	Statistical genetics
Georgia	Chenevix-Trench	QIMR-Berghofer, Australia	CIMBA
Fergus	Couch	Mayo Clinic, USA	ENIGMA
Laura	Fejerman	University of California in San Francisco (UCSF), USA	LAGENO-BC
Judy	Garber	Harvard University, USA	Clinical Trials
Liz	Gillanders	Division of Cancer Control and Prevention Sciences, NCI, USA	DCCPS/NCI extramural
Chris	Haiman	University of South California, USA	AABCGS
Pete	Kraft	Harvard University, USA	Prospective cohorts/ Statistical genetics
Roger	Milne	University of Melbourne, Australia	BCAC DACC Chair
Nick	Orr	Queen's University Belfast, North Ireland	Male breast cancer studies
Julie	Palmer	Boston University, USA	AABCGS
Paul	Pharoah	Cambridge University, England	BCAC pathology/survival Working Group
Marjanka	Schmidt	Netherlands Cancer Institute, The Netherlands	BCAC pathology/survival Working Group
Jacques	Simard	University of Laval, Canada	PERSPECTIVE I&I
Wei	Zhang	Vanderbilt University, USA	ABCC and AABCGS

References

1. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2015. **136**(5): p. E359-86.
2. Michailidou, K., et al., *Association analysis identifies 65 new breast cancer risk loci.* Nature, 2017. **551**(7678): p. 92-94.
3. Milne, R.L., et al., *Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer.* Nat Genet, 2017. **49**(12): p. 1767-1778.
4. Cai, Q., et al., *Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1.* Nat Genet, 2014. **46**(8): p. 886-90.
5. Fejerman, L., et al., *Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25.* Nat Commun, 2014. **5**: p. 5260.
6. Feng, Y., et al., *Characterizing Genetic Susceptibility to Breast Cancer in Women of African Ancestry.* Cancer Epidemiol Biomarkers Prev, 2017. **26**(7): p. 1016-1026.
7. Orr, N., et al., *Genome-wide association study identifies a common variant in RAD51B associated with male breast cancer risk.* Nat Genet, 2012. **44**(11): p. 1182-4.
8. Garcia-Closas, M., N.B. Gunsoy, and N. Chatterjee, *Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer.* J Natl Cancer Inst, 2014. **106**(11).
9. Kuchenbaecker, K.B., et al., *Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers.* J Natl Cancer Inst, 2017. **109**(7).
10. Maas, P., et al., *Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States.* JAMA Oncol, 2016. **2**(10): p. 1295-1302.
11. Mavaddat, N., et al., *Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes.* Am J Hum Genet, 2018.
12. Pashayan, N., et al., *Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model.* JAMA Oncol, 2018. **4**(11): p. 1504-1510.
13. Garcia, M.E., et al., *Abstract 2271: Pathway analysis suggests biological processes driven by germline genetic associations with breast cancer prognosis.* Cancer Research, 2018. **78**(13 Supplement): p. 2271-2271.
14. Gavin, P.G., et al., *Association of Polymorphisms in FCGR2A and FCGR3A With Degree of Trastuzumab Benefit in the Adjuvant Treatment of ERBB2/HER2-Positive Breast Cancer: Analysis of the NSABP B-31 Trial.* JAMA Oncol, 2017. **3**(3): p. 335-341.
15. Grossberg, A.J., et al., *Association of Transforming Growth Factor beta Polymorphism C-509T With Radiation-Induced Fibrosis Among Patients With Early-Stage Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial.* JAMA Oncol, 2018. **4**(12): p. 1751-1757.
16. Guo, Q., et al., *Identification of novel genetic markers of breast cancer survival.* J Natl Cancer Inst, 2015. **107**(5).
17. Sucheston-Campbell, L.E., et al., *Genome-wide meta-analyses identifies novel taxane-induced peripheral neuropathy-associated loci.* Pharmacogenet Genomics, 2018. **28**(2): p. 49-55.
18. Wilkinson, M.D., et al., *The FAIR Guiding Principles for scientific data management and stewardship.* Scientific Data, 2016. **3**: p. 160018.