Elucidating the genomic effects of obesity in breast cancer. A closer look at low- and middle-income countries

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe: structured template" fourni par Science Europe.

Plan Details

Plan title	Elucidating the genomic effects of obesity in breast cancer. A closer look at low- and middle-income countries
Deliverable	D1
Version	First version
Plan purpose/scope	The objective of this data management plan is to describe the source and type of the data that will be used and generated during the course of this research project in relation to the different objectives and also describe how we will manage, analyse, store and share the data with the participant centers.
Fields of science and technology (from OECD classification)	3.3 Health sciences
Language	english
Creation date	2024-05-28
Last modification date	2024-06-10
Identifier	Data Mangement plan O-BRiDGE
Associated documents (publications,reports, patents, experimental plan...), website	• IARC Data Protection Policy : IARC DP Policy 09/2021

Project Details

Project title	Elucidating the genomic effects of obesity in breast cancer. A closer look at low- and middle-income countries
Acronym	O-BRiDGE
Abstract	Breast cancer (BC) is nowadays the most common malignancy and the leading cause of cancer-related mortality among women in all continents. The number of newly diagnosed BCs is projected to grow by over 40% in 2040. A particularly large relative increase will be seen in low- and middle-income countries LMICs. This projection is solely due to the growth and aging of the population without accounting for the changes in cancer incidence due to prevalence of other associated risk factors including obesity. Levels of overweight and obesity across LMICs have been increasing rapidly, affecting particularly women in the Middle East, North of Africa, and Latin America. Obesity has been shown to be a BC risk factor for many years. However, studies about the biological and genomic impact of obesity on breast cancer are limited. There is growing evidence demonstrating that obesity drives breast cancer development through biological mechanisms associated with increased levels of DNA damage and impaired DNA damage response (DDR). The burden of obesity in LMICs urges understanding the genomic bases of obesity driven BC in these populations systematically understudied. The O-BRiDGE project aims to understand the genomic effects of obesity in DNA damage and DDR in BC by conducting integrative genomic analyses of BC patients from LMICs. Using a comprehensive genomic analysis, we propose to combine whole genome and RNA sequencing data with detailed information on demographics, anthropometric measurements, and histopathology to elucidate the biological effect of obesity in breast cancer development associated with increased DNA damage and impaired DDR in LMICs. The findings of this work will improve our understanding of the biological mechanisms of obesity as a risk factor for BC development. Our focus on women of LMICs where the burden of obesity and BC is increasing will provide additional relevant evidence for preventive interventions in these regions.
Funding	• Institut National Du Cancer : 2023-223
Start date	2024-01-13
End date	2027-05-12
Partners	• National Center of Human Genomics Research https://ror.org/004yvsb77 • Weill Cornell Medicine https://ror.org/02r109517

Research outputs :

1. O-BRiDGE Data collection (Dataset)

Contributors

Name	Affiliation	Roles
Perdomo Sandra - https://orcid.org/0000-0001-7740-8141		• Contact Person • Project Coordinator • Project coordinator

Elucidating the genomic effects of obesity in breast cancer. A closer look at low- and middle-income countries

Data description and collection or re-use of existing data

Research output description

Name	O-BRiDGE Data collection
Description	We will bring together a large and diverse collection of BC cases with different ethnic backgrounds, biological samples and complete demographic, histopathological and clinical information including different anthropometric measurements and population specific exposures. Using mutational signature and gene expression analyses, we will elucidate the DNA repair and damage mutagenic processes attributed to obesity. The findings of this work will improve our understanding of the biological mechanisms of obesity as a risk factor for breast cancer development. We will characterize the effects of obesity in DNA damage and DNA repair mechanisms that contribute to chromosome instability and increased oncogenesis. Our focus on women of LMICs where the burden of obesity and breast cancer is increasing will provide additional relevant evidence for preventive interventions in these regions. This study contains several important innovations. It will substantially contribute to our current knowledge of the biological effect of obesity in BC and provide avenues for potential prevention strategies in populations with an imminent increase in BC incidence in the next 20 years. The emerging findings from comparative studies that interrogate population diversity in tumour biology have unveiled the necessity to require diverse genetic data as a matter of rigorous scientific interrogation. No other comprehensive studies have generated new genomic (WGS and RNA-Seq) data from BC cases with diverse ancestries with well-annotated epidemiological and clinical data.
Type	Dataset
Workpackage	
Keywords	• ADN (Thésaurus INRAE) • ARN (Thésaurus INRAE) • cancer du sein (Thésaurus INRAE)
Language	english
Issued Date	
May contain personal data?	Yes
May contain sensible data?	Yes
May take ethical issues into account?	Yes

Will existing data be reused?

Justification	Centers with already recruited cases/samples (FFT, FFPE and blood) from study collections and biobanks or with paired whole genome sequencing (WGS) and gene expression data already generated and associated epidemiological and clinical data are also included in the O-BRiDGE project. Centers and studies participating under scenarios 2 and 3 must meet the following criteria to be eligible to participate: • Studies of invasive breast cancer • Available FFT, FFPE and blood samples or existing WGS and gene expression data • Core epidemiological and clinical data (as defined below) • Ethics approval and consent for genetic studies • Data sharing plan

---

How new data will be collected or produced?

Name of the method	Epidemiological and clinical data collection
Description	All information will be collected from patients that have given their specific consent to participate in the **BRiDGE** project, or their general consent to participate to parallel research studies when the consent provided is consistent with the research activities foreseen in **BRiDGE**. Ongoing recruitment or retrospective collections can contribute to **BRiDGE** if the followed protocol is consistent with these guidelines. Prospective recruitment of all cases will take place according to the following strategy: i. The cases are identified and their eligibility to the study is confirmed. ii. The cases are informed of the study and if they agree to participate, they are requested to complete the informed consent form iii. A lifestyle questionnaire is administered to the case individual by a trained interviewer (using lifestyle questionnaire and questionnaire manual). iv. Anthropometric measures are taken from the case (height, weight, waist and hip circumference) and additional clinical information is collected from the clinical notes. v. A blood sample is obtained from the case individual. When feasible, this is a fasting blood sample. vi. A tumor and non-tumor tissue sampling is performed and preserved frozen. The following core data will be required from all participating studies: (i) Demographic details (age, sex, ethnic origin, city and place of residence and educational status); (ii) History of tobacco use, including frequency and intensity; (iii) History of alcohol consumption, including frequency and intensity; (iv) Anthropometric data: height, weight and (v) Reproductive factors including parity, breastfeeding (including cumulative duration), contraceptive and menopausal hormone use, age of menarche, and age of menopause. (vi) Family history and ER status. Additional clinical information will be retrieved from clinical records.
Data Nature	Observation
Equipments, technical platforms used	• BRiDGE RedCap platform :

Documentation and data quality

What metadata and documentation (for example way of organising data) will accompagny the data?

Description	Existing demographical, clinical and lifestyle data will be reused as allowed for in consent forms of participants will be harmonised and merged with prospectively collected data. Serial number or alphanumeric code (excluding surname(s), first name(s) and registration number in the National Identification Register of Natural Persons) are used to identify subjects together with the initials of the projects the subjects were included for and a participating centre number to identify the origin of the subject. A dictionary of variables and codebook is also available.

What methods will be used to ensure their scientific quality?

Description

Information on demographic, socioeconomic, behaviours/exposures, and medical history were collected using standardised structured questionnaires. The REDCap (Research Electronic Data Capture) software is being used to accurately and securely collect questionnaire and clinical data. The data collected will be harmonised and validated and quality control checks are being conducted to correct missing or unusual data entries.

Legal and ethical requirements, codes of conduct

If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

Description

Personal data shall be processed in accordance with the UN Principles and the IARC Data Protection Policy. IARC is committed to strive for the highest level of protection as it processes personal data, and will always respect the rights and freedoms of data subjects. Any research conducted at IARC will also be subject to supervision of the IARC Ethics Committee. IARC shall ensure that the personal data is accurate, and that the processing is limited to the minimum required to achieve the tasks of the organization. IARC shall also ensure that data protection is embedded into the design of procedures and the technical setup of the organization. IARC shall also maintain a training and awareness program for IARC staff, and shall ensure that third parties acting on behalf of IARC are made aware of the data protection requirements of the organization. IARC shall act as a data controller of research data, and shall encourage its partners and collaborators to ensure full compliance with applicable data protection legislation and the highest standards of data protection. In case of the processing of personal data for research purposes, IARC shall ensure an adequate level of transparency that is proportionate to the risks to the rights and freedoms of data subjects. IARC shall also, at minimum, adhere to the principles and procedures as laid out in the IARC Data Protection Policy on the processing of personal data for research purposes.

IARC, as part of the World Health Organization, which is an intergovernmental organization and Specialized Agency of the United Nations, is not subject to any national nor EU laws or regulations, including, without limitation, national data protection laws or the GDPR. All actions, obligations and undertakings of IARC will be conducted in compliance with its Statute and international law, as well as the regulations, rules and policies adopted by its Governing Bodies, including the WHO regulatory framework, and its internal policies and procedures (all together referred to as “IARC/WHO Regulatory Framework”). As part of the IARC/WHO Regulatory Framework, IARC has implemented a data protection framework that is in line with internationally recognized standards. This includes, without being limited to, the Personal Data Protection and Privacy Principles for UN System Organizations (the “UN Principles”) adopted by the UN High-Level Committee on Management at its 36th Meeting on 11 October 2018, UN-HLCM 2018, and more specifically, the IARC Data Protection Policy.

Related references

- IARC Data Protection Policy :

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Description

All legal issues are included in the material and data transfer agreements that are signed between IARC and participant centres.

What ethical issues and codes of conduct are there, and how will they be taken into account?

Description	IARC is the leading institution of the O-BRiDGE project and has ensured that all appropriate ethical and legal documentation is in place to allow use of data and biological samples collected as part of the study. Ethical approval was obtained from all the collaborating centers involved and informed consent for data sharing and long term preservation was also collected.

Data processing and analysis

How and with what resources will the data be processed / analyzed?

Description	Data and associated metadata harmonised by O_BRiDGE will be deposited in compliance with legal and ethical requirements on public data repositories such as e.g. OpenAIRE. Any code to be deposited will utilize open access platforms such as Github. The appropriate repository has not yet been identified and will be finalized at a later time. Nevertheless, a data access committee is planned to provide access to data when data is ready to be shared with the external scientific community. Determination of conditions for access are underway and will follow similar formats as existing repositories.

Storage and backup during the research process

How will data be stored and backed up during the research?

Storage needs	Data already collected and being collected is coordinated by IARC. There is a strict data policy in place with measures and controls for: o Data access: Physical and logical access controls must be established in order to protect the Data at all times and regardless of its classification. The Data will be stored in a secure data centre on IARC premises requiring either badge or key access to its physical location. o Identity management: All Data Users requiring access to the Data must use an IARC user identity and password as the primary form of authentication. These identities are managed through a central directory maintained by ITS. o Backup to a secure on-site encrypted server: Backups of the Data will be stored in a secondary secure data centre respecting the same access controls as of the primary secure data centre.
Measures taken for data security	o Network security: physical and logical segregation of different networks and the use of modern firewalls and routers to ensure the security of the different parts of our network o Server Management: central management to ensure the latest security patches are applied, antivirus software is up-to-date and the appropriate access controls are in place in order to protect the confidentiality, integrity and availability of the Data. Only authorized ITS personnel are able to access these servers, to make modifications to their configurations or to grant access rights as appropriate. o Logging: Detailed logging of access to Data will be put in place in order to allow for a clear audit trail to be maintained of access and any modification made by each authorized Data User. o Cryptography: Data will be encrypted using the 256-bit Advanced Encryption Standard (AES-256) both at rest as well as in all backups of Data. Data will be safely stored in certified repositories for long term preservation and curation.

Data sharing and long-term preservation

How will data be shared?

Modalities of sharing	Resulting genomic data generated by CNRGH-CEA will be shared with IARC/WHO (France) and with Cornell University(United States) for further data processing and analysis. For patient associated data, only the participating centres from where Materials/Data originate will retain the link with patients' identifying information. The raw and analysed sequence data will be shared via the secure data import/export platform Globus (https://www.globus.org/data-transfer).
Reusability	

How will data be long-term preserved? Which data?

Justification	As a default, IARC shall keep data only as long as the data is needed to serve the research purposes of the institution, and as long as IARC has a legal basis to process such data. This is depending on the requirements of the research, e.g. in case that IARC follows developments in cancer over a long time span. IARC shall periodically, at least once per calendar year, review the retention of personal data for research purposes, and shall also review whether such data must be kept in a person-identifiable manner or whether data could be anonymized, further pseudonymized or aggregated.
Start date	
End date	
Final dispositions	