GeCIP Detailed Research Plan Form

Background

The Genomics England Clinical Interpretation Partnership (GeCIP) brings together researchers, clinicians and trainees from both academia and the NHS to analyse, refine and make new discoveries from the data from the 100,000 Genomes Project.

The aims of the partnerships are:

1. To optimise:

- clinical data and sample collection
- clinical reporting
- data validation and interpretation.

2. To improve understanding of the implications of genomic findings and improve the accuracy and reliability of information fed back to patients. To add to knowledge of the genetic basis of disease.

3. To provide a sustainable thriving training environment.

The initial wave of GeCIP domains was announced in June 2015 following a first round of applications with expressions of interest in January 2015. In April 2016 we invited the inaugurated Cancer GeCIP domains to develop research plans for 'Gear 2' working closely with Genomics England. Within the Cancer Main Programme, the 'Gear 2' phase of the project refers to recruitment of specific cohorts of patients, inclusion of biopsy tissue (diagnostic/recurrence) and ctDNA in selected cohorts and the initiation of clinical trials in early stage (adjuvant/consolidation) setting. These will be used to ensure that the plans are complimentary and add real value across the GeCIP portfolio and address the aims and objectives of the 100,000 Genomes Project. They will be shared with the MRC, Wellcome Trust, NIHR and Cancer Research UK as existing members of the GeCIP Board to give advance warning and manage funding requests to maximise the funds available to each domain. However, formal applications will then be required to be submitted to individual funders. They will allow Genomics England to plan shared core analyses and the required research and computing infrastructure to support the proposed research. They will also form the basis of assessment by the Project's Access Review Committee, to permit access to data.

Domain leads are asked to complete all relevant sections of the GeCIP Detailed Research Plan Form, ensuring that you provide names of domain members involved in each aspect so we or funders can see who to approach if there are specific questions or feedback and that you provide details if your plan relies on a third party or commercial entity. You may also attach additional supporting documents as relevant (optional).

Additional members can apply to join the GeCIP domain by completing the form on our website found here: http://www.genomicsengland.co.uk/join-a-gecip-domain/.

Genomics England Clinical Interpretation Partnership (GeCIP) Detailed Research Plan Form

Application Summary	
GeCIP domain name	**Breast cancer**
Project title *(max 150 characters)*	**Breast cancer research in the 100,000 Genomes Project**

Objectives. *Set out the key objectives of your research. (max 200 words)*

1. Clinical aspects and translation
2. Training
3. Germline genetics and genetic epidemiology
4. Tumour somatic genetics discovery
5. Rapid translation of genetic variants
6. Prediction of prognosis in early breast cancer

Lay summary. *Information from this summary may be displayed on a public facing website. Provide a brief lay summary of your planned research. (max 200 words)*

Breast cancer is a cancer that develops from breast tissue, usually manifesting itself as a lump in the breast, a change in breast shape or skin texture. Cancer most commonly develops in the cell that line the milk ducts (ductal carcinomas) and lobules (lobular carcinomas) that supply the ducts with milk. Survival rates vary by tumour type, extent of the disease, and the patient's age, but tend to be high, with around 80-90% of patients being alive five years after their diagnosis. The breast cancer GeCIP domain hopes to use the 100,000 Genomes Project to better understand how spelling mistakes in the DNA (variants) of participants before the tumour formed might have contributed to their breast cancer forming, or might alter their risk if their lifestyle or other factors would traditionally put them at a higher risk. The group will also investigate how variants that are present in the tumour only influence how the tumour develops and how it reacts to treatment. The group hopes to identify patients where alternative treatments can be repurposed and made more effective given the genetics of themselves or their tumour.

Expected start date	**Q2 2017**
Expected end date	**Q2 2020**

Lead Applicant(s)	
Name	Nicholas Turner
Post	Academic Consultant Medical Oncologist
Department	Breakthrough Breast Cancer Research Centre
Institution	Royal Marsden Hospital and Institute of Cancer Research
Current commercial links	
Deputy Lead Applicant(s)	
Name	Clare Turnbull
Post	Clinical Lead for Cancer Data, Genomics England Senior Researcher in Cancer Genomics, Institute of Cancer Research Reader in Genomic Medicine, Queen Mary University of London Honorary Consultant in Clinical Cancer Genetics, Guys and St Thomas' NHS Trust
Department	
Institution	As above
Current commercial links	

GeCIP domain - Expression of interest

Full proposal

Title	Breast cancer

Research plans. *Give details of the analyses and experimental approaches, study designs and techniques that will be used and timelines for your analysis. Describe the major challenges of the research and the steps required to mitigate these.*

1) Clinical aspects and translation

The Breast GeCIP Domain recognises the overriding importance of ensuring the 100,000 Genomes Project transforms the NHS, and offers the highest quality of data interpretation. We have a number of areas of clinical priority:

a) Clinical reporting

We will generate subdivisions of the GeCIP that will advise on the data to be fed back to individual clinicians and patients, in collaboration with the interpretation, validation and feedback domain. This will be an interactive process with the aim to enhance reporting as the programme progresses.

For germline genetics we will focus on:

- Analysis of germline data for mutations in genes currently included in the (looked for) secondary findings gene list that are associated with predisposition to breast cancer (BRCA1 and BRCA2), as well as other genes in this list (MLH1, MSH2, MSH6, PMS2, EPCAM, APC, MUTYH, VHL, MEN1, RET, RB1)
- In addition, and in collaboration with the inherited cancers domain, we shall develop analysis of germline data for mutations in additional genes implicated in and currently clinically tested for in breast cancer predisposition (TP53, CDH1, PTEN, STK11, PALB2).

For somatic genetics we will focus on:

Development of a database of actionable, and validated genes/mutations that will be released to individual clinicians, with a secondary list of recurrent genetic events of uncertain clinical significance that could be returned to clinicians on specific request.
Clare Turnbull will participate in the interpretation, validation and feedback domain.

b) Clinical data collection

The GeCIP will work to ensure that all relevant clinicopathological data is collected on patients, to enable all future potential research applications, and to facilitate future data interpretation for routine clinical use. We will aim to work with existing electronic health records research infrastructure, such as the Oxford and Farr Institutes, to build a large data set to facilitate all analysis proposed.

For patients with metastatic breast cancer, knowledge of sites of metastasis will facilitate research into genetic events that are associated with site specific metastasis, in particular CNS involvement, members of the GeCIP are establishing a national CNS biobank (PRIMROSE), that may lead to risk stratification with a view to targeted screening as well as the identification of possible therapeutic target. Where feasible we will also collect information on outcome and response to therapy in the metastatic setting, for linkage with genomic data.

c) Selection of tumours for sequencing –

The breast members of the GMCs represented in the GeCIP Domain strongly endorse generating workable pathways for routine collection of additional core biopsies at diagnosis or at surgery under generic consent, working with diagnostic pathology and radiology so as not to disrupt the

diagnostic process. This will enhance the collection of material both to facilitate WGS and support additional research, including the taking of fresh frozen biopsies and potential direct DNA extraction from fresh tissue.

The focus of recruitment should be on sequencing patients presenting with primary breast cancers, or those presenting with de novo metastatic disease at diagnosis. To facilitate subsequent outcome research we would propose to cap the number of patients with small (<2cm) node negative cancers. We propose to sequence multiple lesions in multifocal breast cancer. Patients presenting with relapsed disease, will be a separate focus of study, and these will be sequenced when the archival primary can be obtained along with a recurrent disease sample, including a defined number of patients with multiple recurrent disease samples available through the treatment pathway.

d) Therapeutic trials

We will aim to work with industrial partners to generate clinical trials platforms in early breast cancer using WGS data generated from the 100,000 project. The Breast CSG is strongly represented on the GeCIP Domain to lead on the development of therapeutic trials. In early stage triple negative breast cancer it has been agreed that GEL will collaborate with a breast GeCIP clinical trial, cTRAK - A randomised trial utilising **c**tDNA mutation **t**racking to detect minimal **r**esidual disease **a**nd trigger intervention in patients with moderate and high ris**k** early stage triple negative breast cancer (TNBC). This collaboration will utilise GEL sequencing of TNBC, when previously conducted at GMCs, to identify mutations to track in ctDNA. In metastatic breast cancer, members of the GeCIP Domain are leading a proposed multi-arm platform trial for metastatic breast cancer (plasmaMATCH). We will also explore potential interactions with diagnostic companies, for example PAM50 for breast cancer subtyping, to build richness to the breast cancer dataset.

e) Standardised approved diagnostic approaches –

Members of the GeCIP (Gonzalez, Jones) have extensive experience in developing standardized diagnostic approaches to confirm WGS findings, to ensure potential clinically significant findings translate to patient benefit.

2) Training

This is a highly important part of the GeCIP. Dr Ellen Copson will be the training director. She chairs the medical oncology SAC onco-genetics training working party. We will develop training for trainees based on resources from the 100K genomes dataset particularly aimed at identifying the oncogenetic training needs. The training director will form a point of contact for interested trainees, to facilitate introduction to the domain. Both through existing and future funding we will ensure that there is widespread availability of PhD, Clinical Fellows and Post Docs to work on the 100K genome project.

3) Germline genetics and genetic epidemiology

Research on germline genetic variation will focus on the following areas

- Discovery of novel coding and non-coding genomic variants associated with increased risk of breast cancer (case-control analysis against remainder of 100KGP cohort)
- Functional characterisation of novel germline variants (both coding and non-coding), to develop evidence regarding likely pathogenicity.
- Analysis of germline data for mutational profile of DNA repair genes conferring intermediate penetrance breast cancer risk BRIP1, ATM, CHEK2, RAD50, RAD51D (not currently clinically tested for).
- Study the association between germline changes in breast cancer susceptibility genes (BCSG)

and somatic events that occur in breast tumours, and tumour characteristics. This may reinforce the candidacy of a putative BCSG and secondly, may indicate previously-unconsidered therapeutic avenues.

- Cancer risk characterisation for variants shown to be associated with breast cancer risk: kin-cohort/segregation analysis studies and prospective follow-up studies using linkage with clinical and other flagging data. Where possible, we will investigate whether risks are modified by other common genetic variants associated with breast cancer risk. The aim will be to ultimately use assembly of registers and data to fully elucidate the mechanism of carcinogenesis in mutation carriers, the risks and identify potential therapeutic targets for primary cancer prevention in mutation carriers.
- Integration of data/information on risks, pathogenicity and frequencies into comprehensive breast cancer risk prediction algorithms that can be used to predict future breast cancer risks.

We will collect a rich risk factor genetic epidemiology data set to investigate for association between genetic variants and body mass index, exogenous hormone exposure, socioeconomic status, breast density and outcome.

GeCIP Domain members (Garcia-Closas and Pharoah) lead a Horizon 2020 funded grant that will perform targeted sequencing on 10,000 women with breast cancer, allowing for robust cross-data set collaboration and validation. In addition GeCIP Domain members (Copson) have recruited a cohort of young women with breast cancer in POSH (Prospective study of Outcomes in Sporadic versus Hereditary breast cancer in young breast cancer patients) providing a further source for validation.

4) Tumour somatic genetics discovery

We will explore the following areas of research into tumour somatic genetics

- Discovery of novel coding and non-coding somatic variants in breast cancer, with prioritisation of genetic variants for functional analysis.
- Investigate associations between standard clincopathological parameters and somatic genetic variants, and association with special histological subtypes (e.g. lobular, metaplastic) and with inflammatory breast cancer.
- Refine mutational signatures of breast cancer, define associations including the clinicopathological and treatment response, and associations with germline and somatic genetic variants.
- Use novel image based analysis tools to measure breast cancer microenvironmental heterogeneity, and the association with somatic genetic variation. For example, we will investigate for somatic and germline variants that associate with lymphocytic infiltration.
- Use the whole genome sequencing data combined with high depth targeted resequencing to describe the extent of clonal heterogeneity in individual cancers, including in a subset resequencing of multiple core biopsies to describe spatial heterogeneity.

Validation cohorts are available for lobular (300 cases of pure LCIS no invasion, 1000 cases of LCIS with ILC collected through the GLACIER study Roylance and Sawyer), special histological types of breast cancer (Reis-Filho and Natrajan) and chemotherapy response in metastatic BRCA1/2 and Triple Negative Breast Cancer in the TNT Trial (Tutt and Bliss).

Multi-omic analysis

Through collection of frozen, or RNAlater, biopsies we will be collecting samples for multi-omic analysis of epigenetic changes, RNAseq for gene-expression and microRNA, and proteomics.

For example we will define how exposure driven epigenetic variation in the genome is associated

with somatic genetic variation (Flanagan), and through additional research investigate for mutations that may change transcription factor specific DNA motif in epigenetically defined regulatory regions, with the potential to discover novel regulatory mutations. We will explore how recurrent mutations in SF3B1 (a spliceosomal component complex protein) in breast cancer are associated with alternative splicing of a conserved set of genes that could be therapeutically targeted.

5) Rapid translation of genetic variants

A key aim of the GeCIP will be to translate genetic discovery into functional studies in vitro and in vivo, to confirm functional significance and identify variants with potential therapeutic implications. We will identify the molecular biologists in the UK with the most appropriate expertise to follow-up particular genetic variants, for example function of non-coding variants (Fletcher) and synthetic lethal interactions (Lord). The functional significance of recurrent genomic alterations on breast cancer progression will be investigated in the MCF10 progression series and findings validated using vivo orthotopic and PDX models.

We propose to use data from the 100K Genomes Project to uncover novel therapeutic approaches from analysis of gene interactions, for example patterns of mutually exclusive and co-occurring mutations in breast cancer, using computational tools such as MEMO3 and DAISY4. Such interactions can identify functional overlap and potential synthetic lethal interactions, potentially allowing for novel therapeutic approaches to be devised.

6) Prediction of prognosis in early breast cancer

Over 95% of women that present with early breast cancer are diagnosed without evidence of macroscopic metastases. Prognostication is a major importance in breast cancer, and the 100K genomes project presents a major opportunity to address the impact of genetic variation on outcome. In addition, for patients receiving neoadjuvant therapy, the project presents an opportunity to examine factors that predict for pathological complete response (pCR) to chemotherapy, a strong surrogate for long term outcome. In due course we will examine the following associations with outcome

- Somatic and germline genetic variation, mutations signatures and disease free survival.
- Impact of clonal heterogeneity and disease free survival.
- Prediction of pCR to chemotherapy and the genetics of extreme response (pCR to hormone therapy)
- Prediction of radiation resistance and local recurrence.
- Influence of germline genetic variation on sites of metastasis (Bone, Lung, Liver, CNS).
- ctDNA in predicting risk of recurrence. We have demonstrated proof of principle that in patients who have completed treatment for early breast cancer, persistent detection of ctDNA predicts a very high risk of early relapse (Turner ASCO 2014). The WGS of the 100K presents a major opportunity to extend this work, identifying multiple genetic variants in particular for structural variants, that can be tracked in plasma. We will collect serial plasma samples in selected GMCs to develop robust ctDNA criteria to predict recurrence.

Pharmacogenomics of treatment toxicity

The GeCIP Domain considers the importance of 100K data set in examining germline variants that predict of toxicity to treatment, in particular chemotherapy. We will be keen to work with GE to establish robust data collection practices for acute (such as neutropenia, decline in cardiac function, peripheral neuropathy) and later (cardiac, leukaemia) toxicities. Acute and late toxicity extreme phenotypes will be relatively few in number, and GeCIP Domain members have well-annotated cohorts that will be essential for validation (Abraham/Caldas). We will subsequently aim to identify germline epigenetic variation that may predict for toxicity. GeCIP Domain member

(Pirmohamed) is deputy director of the MRC centre for drug safety science and has led a number of international studies in pharmacogenomics.

7) Evolution of genetics of metastatic breast cancer

We propose that patients presenting with metastatic breast cancer or locally recurrent breast cancer with a biopsy of recurrent tumour may supply paired recurrence and archival primary tissue for sequencing. In a small subset of patients with metastatic breast cancer, specific GMCs will collect serial sampling through the metastatic course to study clonal evolution, with parallel plasma sample collection for ctDNA analysis. This research will enable in-depth inference of the evolutionary history of breast cancer (through so-called molecular archaeology approaches) on any cases with multiple samples.

To facilitate molecular archeology approaches we will investigate in additional research combining bulk tumour sequencing with single-cell genomics (and transcriptomics) to dissect tumour evolution and subclonal architecture in detail (Van Loo, Voet).

Funding for research

The academic members of the GeCIP Domain have extensive existing funding, and we will be looking to apply for structured substantial funds from GeCIP funders (MRC, Wellcome, NIHR, CRUK), or other funders, to facilitate the research ideas presented in the outline application.

Engagement of patient representation groups

The GeCIP Domain values patient public involvement, and we have approached the Independent Cancer Patients Voice to join the GeCIP Domain to provide input into both clinical and research aspects of the proposal.

Collaborations including with other GeCIPs. *Outline your major planned academic, healthcare, patient and industrial collaborations. This should include collaborations and data sharing with other GeCIPs.*

GMCs members on GeCIP

East of England NHS (Abraham, Baird, Caldas, Earl, Tischkowitz)
Oxford NHS (Harris, Schuh)
West Midlands NHS (Francis, Rea)
South West Peninsula NHS (Ellard and Ferguson)
South London NHS (Sawyer and Simpson) Wessex NHS (Copson)
North West Coast NHS (Palmieri and Pirmohamed)
Imperial College Health Partners NHS (Coombes, Krell, Ring, Stebbing, Turner)
University College London Partners NHS (Roylance, Schmid)

Breast Cancer Clinical Studies Group

Mark Beresford, Royal United Hospitals Bath NHS
Judith Bliss, Institute of Cancer Research
David Cameron, University of Edinburgh
Ellen Copson, University Hospital Southampton
Helena Earl, University of Cambridge
Adele Francis, University Hospital Birmingham
Stuart McIntosh, Queens University Belfast
Iain MacPherson, Glasgow University
Carlo Palmieri, Liverpool University Daniel Rea, University of Birmingham
Alistair Ring, Royal Marsden Hospital
Peter Schmid, Queen Mary University of London
Abeer Shaaban, University of Birmingham

Academic members
Jean Abraham, University of Cambridge
Antonis Antoniou, University of Cambridge
Richard Baird, University of Cambridge
Fedor Berdichevski, University of Birmingham
Carlos Caldas, CRUK Cambridge Institute
David Cameron, University of Edinburgh
Charles Coombes, Imperial College London
Douglas Ferguson, Royal Devon and Exeter NHS
James Flanagan, Imperial College London
Olivia Fletcher, Institute of Cancer Research
Monste Garcia-Closas, Institute Cancer Research
Adrian Harris, University of Oxford
Jonathan Krell, Imperial College London
Chris Lord, Institute of Cancer Research
Luca Magnani, Imperial College London
Serena Nik-Zainal, Wellcome Sanger Institute
Kai Ren Ong, Birmingham Women's Hospital
Nicholas Orr, Institute of Cancer Research
Paul Pharoah, University of Cambridge
Munir Pirmohamed, University of Liverpool
Rebecca Roylance, Queen Mary University
Elinor Sawyer, King's College London
Michael Simpson, King's College London
Justin Stebbing, Imperial College London
Marc Tischkowitz, University of Cambridge
Andrew Tutt, King's College London & ICR
Thierry Voet, Wellcome Trust Sanger Institute

Bioinformaticians and statistics
Jean Baptiste Cazier, University of Birmingham
Maggie Cheang, Institute of Cancer Research
Anita Grigoriadis, King's College London
Lucy Kilburn, Institute of Cancer Research
Peter Van Loo, Cancer Research UK London Research Institute

Molecular pathology
David Gonzales de Castro, Institute of Cancer Research
Louise J Jones, Queen Mary University
Rachael Natrajan, Institute of Cancer Research
Anna Schuh, University of Oxford

International collaborators
William Foulkes, McGill University Montreal Canada
Church Perou, University of North Carolina USA
Jorge Reis-Filho, Memorial Sloan Kettering Cancer Centre New York USA

The Breast GeCIP Domain will create an inclusive environment that will allow collaborative research to thrive. The GeCIP Domain has representation from all Genomic Medicine Centres, and the breast NCRI Clinical Studies Group, to ensure that all GMCs can contribute to the clinical and

research opportunities in the GeCIP, and benefit from the substantial contribution the GMCs will make to the overall project.

The GeCIP Domain brings together clinicians, clinical researchers, geneticists, molecular biologists, bioinformaticians, statisticians and patient representatives to ensure the appropriate skills are in place to implement all aspects of research with the 4000 tumour/normal pairs. We have international collaborators with appropriate links to international initiatives such as TCGA, and members of the GeCIP Domain have well annotated breast cancer collections to ensure findings identified in the GeCIP can rapidly be validated in independent external data-sets.

Gear 2 substudy proposals

Gear 2 Eligibility Criteria

Pre- and post-operative cfDNA collection - Triple-negative (or ER and HER2-negative if PR not available) breast cancer undergoing primary surgery or ER or HER2 positive BC with >=4 axillary lymph nodes positive on imaging.
Multiregion (multiple samples/cores taken ex-vivo from a single resected tumour mass) - Up to 4 regions of surgically resected breast cancers, from primary and/or from axillary lymph nodes.
Multitumour (disparate sites of tumour deposition collected synchronously) - Metastatic disease - multiple sites of metastasis up to 4 sites including patients with prior treatment.
Longitudinal - Paired samples pre-and post- administration of neoadjuvant chemotherapy or endocrine therapy; Progression to metastatic breast cancer following primary breast cancer.

BR01: Genetics of tumours post-neoadjuvant therapy.

The factors resulting in resistance to primary hormone and chemotherapy remain remarkably obscure despite decades of research. Paired WGS of treatment naïve and residual disease after standard neo-adjuvant therapy presents a substantial opportunity to investigate whether individual genetic events or mutational signatures are enriched in residual disease as a mechanisms of resistance.

BR02: Sequencing of metastatic breast cancer with retrospective primary sequencing.

Although a large number of primary breast cancers have been subject to WGS, few metastatic breast cancers have been sequenced. Patients who have primary breast cancer sequenced in GEL are eligible to have the metastasis sequenced at a later date. Yet the time from primary diagnosis to recurrence is frequently many years in breast cancer.
To increase the number of primary-metastasis pairs sequenced, we would like to allow patients who present with metastatic disease be eligible for GEL, if a fresh frozen or similar sample from the original primary is available. Many GMCs have tissue banks that store fresh frozen material from patients presenting with primary breast cancer.

Detailed research plan – Cancer Main Programme Gear 2 Substudies

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**BR01: Genetics of tumours post-neoadjuvant therapy.**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Sequencing of patients who have undergone a full therapeutic course of standard neoadjuvant therapy, endocrine or chemotherapy, for primary breast cancer. Residual disease in the breast suitable for sampling at surgery as determined by routine pre- operative imaging.
Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	1
# cores per tumour (if multi-region biopsying proposed)	1
Follow-up samples following first ascertainment	1 (post therapy)
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To inform the genetic evolution of primary breast cancer under neoadjuvant therapy
Scientific case and insights that will be gained from this cohort (more details, as indicated)	The factors resulting in resistance to primary hormone and chemotherapy remain remarkably obscure despite decades of research. Paired WGS of treatment naïve and residual disease after standard neo-adjuvant therapy presents a substantial opportunity to investigate whether individual genetic events or mutational signatures are enriched in residual disease as a mechanisms of resistance.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Full proposal (total max 1500 words per Gear 2 Substudy)	
Title *(max 150 characters)*	**BR02: Sequencing of metastatic breast cancer with retrospective primary sequencing.**
Cohort details and scientific case	
Cohort eligibility definition (disease type, sub-type, presentation, stage, treatment, clinical characteristics, epidemiological characteristics)	Presentation with metastatic breast cancer having had prior treatment for primary breast cancer. Archival fresh frozen, or similar, primary breast cancer stored in tissue bank.

Samples per patient at first ascertainment (primary tumour, LNs, metastatic sites) *It is assumed that in addition there will be one germline sample per patient.*	2
# cores per tumour (if multi-region biopsying proposed)	1
Follow-up samples following first ascertainment	None
Purpose of analysis WGS and clinical data from this cohort of patients (brief)	To increase the number of primary-metastasis pairs available to inform the evolution of somatic genetic events between primary and recurrent breast cancer.
Scientific case and insights that will be gained from this cohort (more details, as indicated)	Although a large number of primary breast cancers have been subject to WGS, few metastatic breast cancers have been sequenced. Patients who have primary breast cancer sequenced in GEL are eligible to have the metastasis sequenced at a later date. Yet the time from primary diagnosis to recurrence is frequently many years in breast cancer. To increase the number of primary-metastasis pairs sequenced, we would like to allow patients who present with metastatic disease be eligible for GEL, if a fresh frozen or similar sample from the original primary is available. Many GMCs have tissue banks that store fresh frozen material from patients presenting with primary breast cancer.
Alignment to clinical trials	
Is it proposed that this sub-study is aligned to an existing clinical trial/sample collection study? Study details (incl Phase, commercial partner, geographic recruitment, remuneration). Is co-recruitment to this trial/study optional/mandatory for recruitment to this cohort eligibility?	No
Is this sub-study a new therapeutic trial?	No

Data and informatics requirements
All GeCIP Domains have contributed to the construction of the data model for their tumour type and have contributed to ongoing efforts within Genomics England to develop the clinical and research informatics infrastructure.

Data access and security	
GeCIP domain name	Breast cancer
Project title *(max 150 characters)*	Gear 2 Substudies

Applicable Acceptable Uses. *Tick all those relevant to the request and ensure that the justification for selecting each acceptable use is supported above.*

☐ ***X*** *Clinical care*

☐ ***X*** *Clinical trials feasibility*

☐***X*** *Deeper phenotyping*

☐***X*** *Education and training of health and public health professionals*

☐***X*** *Hypothesis driven research and development in health and social care - observational*

☐***X*** *Hypothesis driven research and development in health and social care - interventional*

☐***X*** *Interpretation and validation of the Genomics England Knowledge Base*

☐***X*** *Non hypothesis driven R&D - health*

☐***X*** *Non hypothesis driven R&D - non health*

☐***X*** *Other health use - clinical audit*

☐***X*** *Public health purposes*

☐ *Tool evaluation and improvement*

Information Governance

☐***X*** The lead for each domain will be responsible for validating and assuring the identity of the researchers. The lead may be required to support assurance and audit activities by Genomics England.

Any research requiring access to the embassy will be required to complete IG Training and read and sign a declaration form. Access will only be granted once these requirements have been met.