nature portfolio

Corresponding author(s): David W. Cescon (dave.cescon@uhn.ca)
Mitchell J. Elliott (mitchell.elliott@uhn.ca)

Last updated by author(s): Dec 22, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The following methods were performed for variant analysis of the whole-exome sequencing: Deduplicated and filtered bam files were obtained from NeoGenomics. Tumor whole-exome sequencing analysis for identification of somatic mutations was performed independent of the RaDaR® ctDNA workflow (described below). Base quality score recalibration (BQSR) was performed using GATK (v3.8) with the BaseRecalibrator and PrintReads tools.31,32 The recalibrated BAM files were sorted and indexed using Samtools (v1.3.1).33 Variant calling was performed with Mutect2 with the tumor-only option in GATK (v4.1.8.1). Variants were called against the human reference genome (Homo sapiens assembly 38) with specified capture regions (NeoGenomics panel) and using a germline resource and a panel of normals for filtering. Orientation bias filtering and contamination estimation were applied during variant calling. The resulting variant call format (VCF) files were further filtered using GATK's FilterMutectCalls tool. The filtered VCF files were annotated using VEP (v98) and converted to mutation annotation format (MAF) using vcf2maf (v1.6.17).38,39 Variants were filtered using in-house scripts with the following criteria: (1) depth >100, (2) variant allele frequency (VAF) >5% to filter sequencing errors (3) VAF not in 40-60% or >90% to filter germline mutations (4) filter for 'Missense_Mutation', 'Nonsense_Mutation', 'Frame_Shift_Ins', 'Frame_Shift_Del', 'Splice_Site', 'In_Frame_Ins', 'In_Frame_Del' (5) allele frequency <1% in the GNOMAD v.3.0 database.40 For investigation of top variants, filtering was performed for known driver genes listed in the COSMIC Cancer Gene Census (v100).34 Oncoplots were generated with maftools (v2.18.0)35 and RColorBrewer (v1.1-3) in R (v4.3.2).
Data analysis	Data analysis for ctDNA testing was provided by NeoGenomics through the standard RaDaR assay workflow.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The whole exome data generated in this study have been deposited in the European Genome-phenome Archive (EGA) under controlled access to ensure that data use is not for profit or commercial purposes. Access can be obtained by submitting a data access request via the EGA portal [https://ega-archive.org/access/request-data/how-to-request-data/] under the dataset ID: EGAD50000001133. Datasets can be accessed directly via: https://ega-archive.org/datasets/EGAD50000001133 and the study can be found at: https://ega-archive.org/studies/EGAS50000000771. The code for selecting variants to include in the RaDaR assay panels are not available due to data privacy laws and copyright restriction. The summary of the variants selected in personalized RaDaR assays and the results of each timepoint can be found in the Supplementary Tables. Given the single institution and geographically restricted nature of the cohort, in concordance with the REB approved clinical data sharing guidelines at the host institution, additional de-identified participant data are available for academic purposes on request from the corresponding authors, Dr. David W. Cescon (dave.cescon@uhn.ca) or Dr. Mitchell J. Elliott (mitchell.elliott@uhn.ca). Requests will be reviewed within 60 days by the host institution and any additional information needed will be discussed with the requesters.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Data has been reported by sex assigned at birth. Breast cancer is a disease that predominantly effects individuals with 46XX karyotype (assigned female at birth) but one male participant was included in the study. Data on self-identified gender was not collected for this study.
Reporting on race, ethnicity, or other socially relevant groupings	Data on race, ethnicity, or other socially relevant groupings was not collected for this study.
Population characteristics	Covariates are summarized in Table 1.
Recruitment	Patients diagnosed with EBC of any receptor subtype receiving standard of care NAT at the Princess Margaret Cancer Centre were enrolled prospectively from October 2016 onwards for serial blood collection and banking.
Ethics oversight	This study was approved by the University Health Network Research Ethics Board (REB#: 17-5962 and #23-5446) and all participants provided informed consent prior to study enrollment and participation.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The final sample size was based on the total number of evaluable patients at the data cutoff. The final population was n=119.
Data exclusions	142 participants enrolled from October 2016 to January 2021 were identified for analysis. Eleven participants were excluded due to a lack of sufficient baseline tumor tissue for genomic DNA extraction and WES (assay requirement). Assays were successfully generated in 93% (122/131) of participants. Three participants, for which panels were designed, were identified to have developed metastatic disease early in their course of treatment (prior to surgery) and were excluded from the primary analysis. 119 participants had evaluable data and were included for analysis in the primary cohort.
Replication	These data are consistent with the findings of other independent studies in this area. The nature of this study does not permit replication.
Randomization	No randomization was pursued in this descriptive study.
Blinding	No blinding was pursued in this descriptive study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT03702309
Study protocol	The full Liquid Biopsy Evaluation and Repository Development at Princess Margaret (LIBERATE) cohort (NCT03702309) study protocol can be requested from the corresponding author.
Data collection	Data was collected through retrospective manual chart review of those participating in the study. Enrollment began in August 2017 with the last participant included in this analysis enrolled in January 2021.
Outcomes	Predefined analyses included assessment of baseline sensitivity, evaluation of ctDNA detection through each participant's clinical timeline, and exploratory assessment of prognostic association with ctDNA detection and clinical outcomes in the entire cohort and in those with HER2-positive and HER2-negative disease. The recurrence-free interval (RFI) was defined as the time from pathologic diagnosis of invasive breast cancer to localized or metastatic recurrence of invasive breast malignancy or the development of a new primary invasive breast cancer. Participants who did not have a recurrence by the last follow up were censored at their last follow up date. All ctDNA testing (RaDaR assay, NeoGenomics) was blinded to all clinical data and performed via the locked assay workflow.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A