

Impact of BRCA mutations, age, surgical indication, and hormone status on the molecular phenotype of the human Fallopian tube

Received: 14 July 2024

Accepted: 12 March 2025

Published online: 26 March 2025

 Check for updates

Ian Beddows¹, Svetlana Djirackor¹, Dalia K. Omran², Euihye Jung², Natalie NC Shih³, Ritu Roy⁴, Aaron Hechmer⁴, Adam Olshen^{4,5}, Guillaume Adelmant^{6,7,8}, Ann Tom^{6,7,8}, Jacob Morrison¹, Marie Adams⁹, Daniel C. Rohrer¹⁰, Lauren E. Schwartz³, Celeste Leigh Pearce¹¹, Heidi Auman¹², Jarrod A. Marto^{6,7,8}, Charles W. Drescher^{13,14}✉, Ronny Drapkin^{2,15}✉ & Hui Shen¹✉

The human Fallopian tube (FT) is an important organ in the female reproductive system and has been implicated as a site of origin for pelvic serous cancers, including high-grade serous tubo-ovarian carcinoma (HGSC). We have generated comprehensive whole-genome bisulfite sequencing, RNA-seq, and proteomic data of over 100 human FTs, with detailed clinical covariate annotations. Our results challenge existing paradigms that extensive epigenetic, transcriptomic and proteomic alterations exist in the FTs from women carrying heterozygous germline *BRCA1/2* pathogenic variants. We find minimal differences between *BRCA1/2* carriers and non-carriers prior to loss of heterozygosity. Covariates such as age and surgical indication can confound *BRCA1/2*-related differences reported in the literature, mainly through their impact on cell composition. We systematically document and highlight the degree of variations across normal human FT, defining five groups capturing major cellular and molecular changes across various reproductive stages, pregnancy, and aging. We are able to associate gene, protein, and epigenetic changes with these and other clinical covariates, but not heterozygous *BRCA1/2* mutation status. This sheds new light into prevention and early detection of tumorigenesis in populations at high-risk for ovarian cancer.

The human Fallopian tube (FT), also known as the uterine tube or oviduct, plays a crucial role in female reproductive biology. It is a highly specialized organ for transporting the ovum and is the site for fertilization and early development. It is now clear that the FT is a site of origin for pelvic serous cancer, a spectrum of diseases that includes high-grade serous cancer of the ovary, FT, and peritoneum^{1–6}. A widely accepted tumorigenesis model includes stepwise progression starting

with clonal expansion of histologically normal, lowly proliferative p53-mutated, secretory epithelial cells. This is followed by progressive accumulation of additional molecular alterations, increased cellular proliferation, and cytologic atypia ultimately leading to a serous tubal intraepithelial cancer (STIC), localized almost exclusively in the fimbria of the distal FT. Invasive serous cancer arises directly from STIC cells within the FT or in the ovary/peritoneal cavity as a consequence of

A full list of affiliations appears at the end of the paper. ✉ e-mail: cdresche@fredhutch.org; rdrapkin@pennmedicine.upenn.edu; hui.shen@vai.org

exfoliation of STIC cells from the lumen of the FT^{7–11}. Phylogenetic analyses suggest the entire process spans decades with p53 signatures arising 20 or more years before serous cancer diagnosis and STIC lesions and early invasive cancer developing roughly 6.5 years and 2 years before diagnosis respectively^{9,11}.

BRCA1 and *BRCA2* mutation carriers (hereafter BRCaM) are at significantly elevated risk for serous ovarian cancer with a lifetime risk of ~15–40%, depending on the specific mutation, relative to those without pathogenic mutations (hereafter non-BRCaM). An estimated 30% of serous cancers have *BRCA1/2* functional loss^{12,13}. Risk increases with age and is reduced by breastfeeding, oral contraceptive pill use, and prior tubal ligation¹⁴. It is well established that surgical removal of FTs and ovaries substantially reduces risk and recent data suggest that removal of the FT alone without oophorectomy also reduces risk but the magnitude is uncertain^{15–17}.

The molecular mechanisms whereby germline *BRCA1/2* mutations contribute to serous cancer risk are unclear. While p53 signatures occur at roughly equal frequency in FTs from *BRCA1/2* mutation carriers and controls, incidental STIC lesions are ~40-fold more common in *BRCA1/2* mutation carriers^{18,19}. The transition from p53 signature to STIC appears to require inactivation of the functional *BRCA* allele⁹. However, some literature suggests that *BRCA* mutations contribute to molecular alterations prior to loss of heterozygosity (LOH), as non-tumorigenic, haplodeficient breast and ovarian epithelial cell lines exhibit evidence of genomic instability and an altered response to replication stress^{20–22}.

Prior molecular profiling studies of FTs from BRCaM carriers also appear to support a haploinsufficient mechanism for *BRCA*-associated serous cancer risk. Transcriptomic studies in normal FTs have identified large numbers of differentially expressed genes between BRCaM and control groups^{23–26} as well as an altered transcriptional response to post-ovulatory microenvironmental stress²⁷. Similarly, DNA methylation analysis of normal FTs from BRCaM carriers and controls posited that germline *BRCA1/2* mutations contribute to elevated ovarian cancer risk via epigenetic reprogramming at the *HOXC4* locus²⁸. Furthermore, a BRCaM-associated DNA methylation signature from normal blood was predictive of breast cancer incidence and survival²⁹, and a *BRCA1/2*-specific proteomic signature from proximal liquid biopsy was shown to be sensitive at detecting high-grade ovarian cancer³⁰. In addition, *BRCA1/2* germline mutations have been reported to be associated with accelerated aging in a host of normal tissues³¹.

While this body of literature suggests that even heterozygous *BRCA1/2* mutations are associated with molecular alterations that impact cancer risk, these studies are all limited in sample size. Most BRCaM-related studies age-match with fewer also matching on menopause status^{23,24,27,29,32}. However, most studies are limited in sample size and do not consider other confounding variables such as cell composition, hormonal status, and reason for surgery, which is inherently confounded for risk-reduction in the *BRCA1/2* populations.

Here we report results from an integrated transcriptomic, epigenomic, and proteomic analysis of histologically normal FT fimbria from *BRCA1/2* carriers and controls, testing the hypothesis that there are indeed molecular alterations in FTs of carriers that contribute to field cancerization and serous cancer risk. Our analysis focuses on the fimbrial region of the FT, where serous cancers are thought to arise. We address the limitations of prior molecular profiling studies by controlling for covariates known to influence the molecular and cellular landscape of the FT. Namely, our results demonstrate that the molecular profiles of FT fimbria from *BRCA1/2* mutation carriers and controls are not demonstrably different when cell composition and reason for surgery are accounted for. The data generated from over 100 FTs provides the research community with the first integrated, multi-omic FT data resource annotated with key epidemiologic and reproductive covariates. Our study design allows us to define cellular and molecular differences in the FT associated with non-BRCaM serous cancer risk factors including hormone status and age.

Results

Characteristics of the sample cohort

We obtained full-thickness FTs from formalin-fixed paraffin-embedded blocks collected less than 3 years prior. Only histologically normal tubes defined using standard clinical workflows and following review by a gynecologic pathologist were selected (i.e., no STIC lesions). The SEE-FIM protocol⁴ including p53 and Ki67 IHC where indicated, was followed for all *BRCA1/2*-mutated patients. Our initial focus was comparing heterozygous *BRCA1/2* pathogenic mutation carriers (BRCaM) to a control cohort without mutation (non-BRCaM). As such, the distal ends of 125 FTs were collected (70 non-BRCaM, 28 *BRCA1*, and 27 *BRCA2*; Fig. 1a; Table 1; Supplementary Data 1). The BRCaM group includes cancer-free patients with a known pathogenic mutation in *BRCA1* or *BRCA2* based on germline sequencing performed for clinical indications at a CLIA-approved laboratory (Supplementary Data 1). The non-BRCaM group includes patient with neither *BRCA1* nor *BRCA2* mutations on clinical testing ($n = 3$) and patients who did not meet clinical criteria for genetic testing ($n = 67$). Following processing and QC filtering, we had high-quality data for 105 FTs (Table 1; Fig. 1b). QC metrics did not differ by BRCA group (Supplementary Fig. 1a–e).

FTs were collected from a diverse patient population and annotated with clinical information such as age at time of salpingectomy, menopause status, and surgical indication (Table 1; Supplementary Data 1). Other clinical data available included number of pregnancies, days to last menstrual period, endometrium state from pathology, and information on contraception use. There were no significant differences in age nor menopause status between germline BRCA groups (Table 1; Fig. 1b, c). Efforts were made to recruit a diverse cohort with regard to self-reported race (Fig. 1d; Table 1). Many non-BRCaM FTs came from elective salpingectomy at the time of cesarean section; these FTs were designated as the postpartum group (Table 1).

In BRCaM patients we confirmed clinical germline variant annotations by directly observing the alternate allele or deletion in our whole-genome bisulfite sequencing data. The clinical BRCA mutations were all known pathogenic variants, being classified as frameshift, nonsense, missense, splice or large deletions (Fig. 1e, f; Supplementary Data 1). This orthogonal validation confirmed the clinical heterozygous variants, and demonstrated that there was no sample swap for the molecular profiling workflows, which is a recognized problem for large genomic studies³³. Variant calling from WGBS data did not identify any known pathogenic *BRCA1/2* mutations in the non-BRCaM FTs.

Tissue-type composition dominates the molecular profiles of the Fallopian tube

We expected cell type composition to be a main driver for variation in the molecular profiles of bulk FTs. Therefore, we first estimated stromal fraction using DNA methylation level at the *MIR200CHG* (HGNC:53161) promoter (Supplementary Fig. 2a). This promoter region (chr12:6962650–6964102) is fully methylated in mesenchymal and unmethylated in epithelial cells. Therefore, methylation fraction at this promoter is directly proportional to stroma:epithelial cell type composition within bulk samples^{34,35}. In addition, two experienced pathologists independently estimated the non-epithelial cell fraction from hematoxylin and eosin (H&E) stained sections in 13 randomly selected FTs. This further validated *MIR200CHG* methylation level as an accurate epithelial:stromal indicator ($\rho = 0.81$; $P = 7.1 \times 10^{-4}$; Supplementary Fig. 2b). RNA-expression levels for key epithelial/mesenchymal markers correlated well with the DNA methylation-based estimate (Supplementary Fig. 2c).

Based on the DNA-methylation method, stromal fraction in our samples ranged from 24% to 89% with a median of 54%. The mean epithelial fraction did not significantly differ by BRCA status, despite being slightly higher in BRCaM FTs (F test $P = 0.13$; Supplementary Fig. 2d; Table 1). The stromal fraction increases with age, indicating a

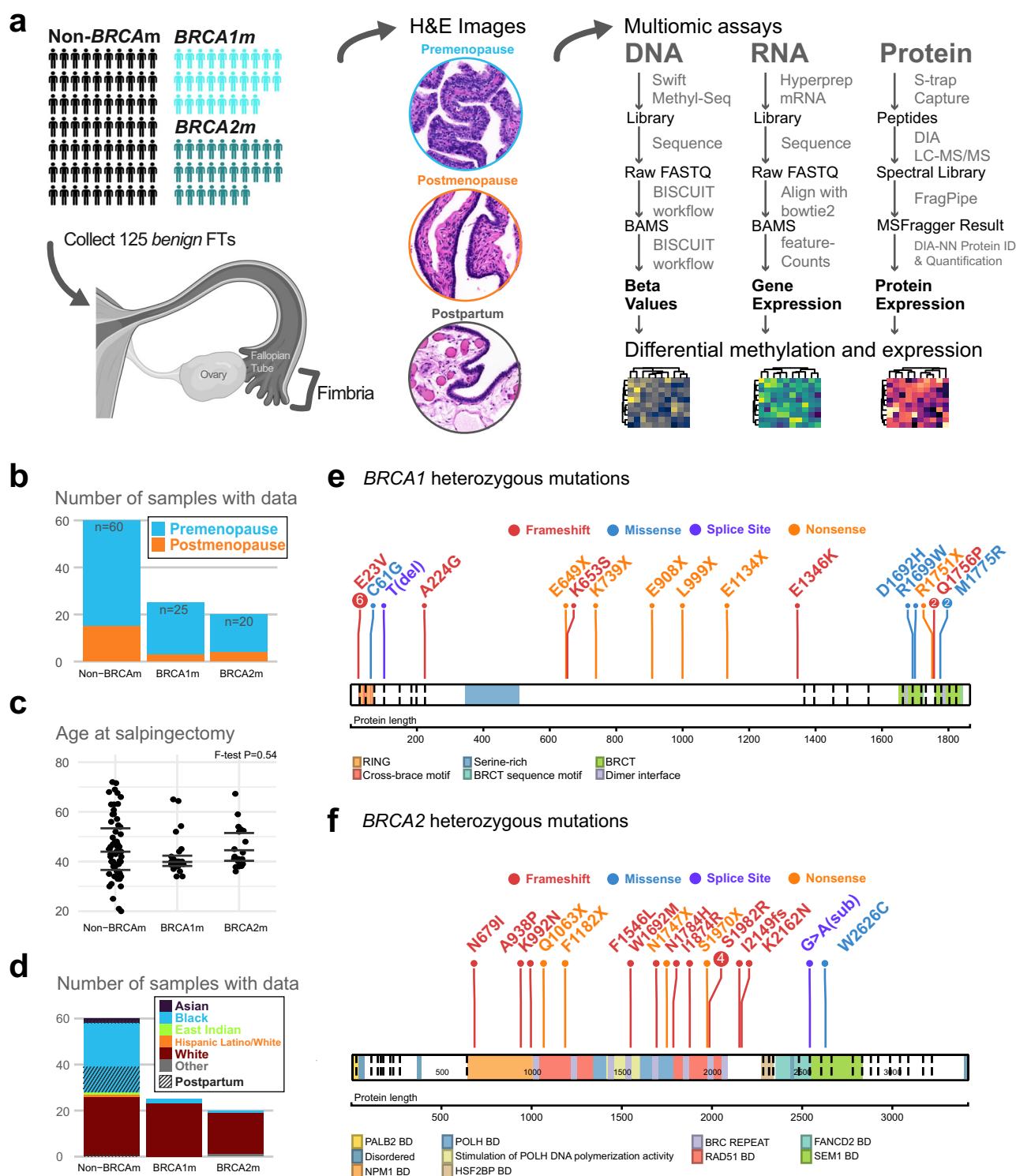


Fig. 1 | Overview of the study design and sample cohort. **a** Left: a total of 125 benign Fallopian tubes were collected: control FTs with no known pathogenic mutations (non-BRCa; $N=70$), pathogenic germline *BRCA1* mutation carriers (BRCa; $N=28$), and pathogenic germline *BRCA2* carriers ($N=27$). Middle: Some sections from fimbria were imaged following H&E staining. Right: DNA, RNA, and protein extraction from FFPE to generate whole-genome bisulfite (WGBS), mRNA, and protein quantification matrices. **b** Distribution of 105 FTs which have data for one or more of the -omics assays. Sample number is shown by *BRCA1/2* germline status colored by menopause status. From these 105 FTs, 92 have high quality data for all three assays. **c** Age distribution at the time of surgery was not different by *BRCA* status (F test P value = 0.54 from ANOVA model; *BRCA1m* $n=28$; *BRCA2m*

$n=27$; non-BRCa $n=70$). **d** Distribution of postpartum samples and self-reported race by *BRCA* status; Black FTs were overrepresented in non-BRCa and postpartum FTs were present exclusively in the non-BRCa group. **e, f** Pathogenic germline mutations in our cohort were validated in the WGBS data. Mutations in *BRCA1* included 11 frameshift, five missense, one splice site, six nonsense, and one large deletion. Mutations in *BRCA2* included 13 frameshift, one missense, one splice site, and four nonsense mutations. The specific types of mutations were not known for one *BRCA1* and one *BRCA2* patient, and one *BRCA1* sample had a large deletion spanning multiple exons; mutations for these 3 samples are not shown. Portions of this figure were created in BioRender. Beaddows, I. (2025) <https://BioRender.com/r66m378>.

Table 1 | Summary of key covariates

	Non-BRCAm	BRCAm		Sum
		BRCA1m	BRCA2m	
FTs Collected	70	28	27	125
Use DNA Methylation	59	25	19	103
Use RNAseq	53	22	19	94
Use Proteomics	55	23	18	96
Use All Assays	52	22	18	92
Cell Composition (from DNA methylation)				
Stroma Fraction ($p = 0.16$)	0.58 ± 0.16	0.48 ± 0.13	0.46 ± 0.16	
Age at Salpingectomy				
Age ($p = 0.42$)	44 ± 13	40 ± 9	42 ± 9	
Menopause Status				
Premenopause	26	19	14	59
Postmenopause	15	3	4	22
Postpartum				
Postpartum	11	0	0	11
Self-reported Race				
Asian	2	0	0	2
Black	25	2	1	28
White	25	20	16	61
Other	0	0	1	1
Reason For Salpingectomy				
Adnexal Mass	1	0	0	1
Cervical Dysplasia	2	0	0	2
Cesarean Section	11	0	0	11
Endometriosis	1	0	0	1
Gender Affirmation	2	0	0	2
Menorrhagia	3	0	0	3
Ovarian Cyst	5	0	0	5
Ovarian Cyst (Sex chord stromal tumor)	1	0	0	1
Ovarian Serous Cystadenoma	2	0	0	2
Pelvic mass	1	0	0	1
Risk Reduction	0	22	18	40
Tubal Sterilization	2	0	0	2
Benign Uterine	21	0	0	21

Table covariate counts and summary statistics are from the 92 samples with all data for RNA expression, DNA methylation, and protein expression. *P* values are from ANOVA *F* test. Values represent mean $\pm$ standard deviation.

loss of FT epithelium relative to the stromal compartment with aging ($\rho = 0.42$; $P = 0.0007$; Supplementary Fig. 2e).

Stromal fraction, age, postpartum, and BRCA status could all be independently associated with principal components (PCs) from one or more -omics assays (Supplementary Fig. 3a–l). We tested these variables together in general linear models to see what explains PC1 and 2 (Supplementary Data 2). Overall, stromal fraction (as measured from DNA) was the strongest explanatory variable for both DNA methylation PCs, as well as PC2 for both the transcriptome and the proteome (Supplementary Fig. 3a–c; Supplementary Data 2). The transcriptome and proteome PCs were most strongly associated with being postpartum or not postpartum. RNA PC1 was also associated to BRCA status, but we later show this is due to postpartum status and other surgical indications which are confounded with BRCA status. BRCA status is also associated with DNA methylation PC1, but not when adjusted for both stroma and age (Supplementary Data 2). Overall, cell

composition is the strongest explainer of variation with postpartum, age, and surgical indication also being important.

Differential methylation analysis between BRCAm and non-BRCAm groups

Overall, the measured DNA methylation patterns at heterochromatin, transcriptional start sites, and CTCF binding sites are consistent with expectations, including clear nucleosome positioning around CTCF binding sites, indicating high data quality (Fig. 2a). However, there are no differences in global DNA methylation at these sites based on BRCA status. Similarly, multiscale representation³⁶ previously used to examine DNA methylation at 10kb–10Mb scales in normal tissue samples³⁷, showed no visible changes (Fig. 2b).

To complement these large-scale analyses, we also performed differential methylated region (DMR) calling at base resolution, in search of any loci-specific differences. We tested various contrasts (Table 2) adjusting for key covariates. As postpartum status (only present in the non-BRCAm group) appeared to have a major impact on the molecular profiles (Supplementary Data 2; Supplementary Fig. 3d–f), we excluded postpartum FTs when comparing BRCA groups to avoid confounding (Table 2). In a model adjusting exclusively for cell composition, no regions had an adjusted $P < 0.05$. We further examined known loci with DNA hypermethylation in either STIC lesions or high-grade serous ovarian cancer^{38,39}. No DNA hypermethylation at these loci was observed in any BRCAm FTs (Fig. 2c).

A prior study²⁸ reported epigenomic reprogramming in BRCAm FTs driven by *HOXC4*-mediated *AID* over-expression in fimbriae. We do not observe any difference in *AID* expression in fimbriae between BRCAm and non-BRCAm FTs; *AID* expression is very low in all groups (Supplementary Fig. 4a). *HOXC4* methylation in fimbriae is not different between BRCAm and non-BRCAm, but *HOXC4* expression decreases with increasing percent stroma, suggesting epithelial expression (Supplementary Fig. 4b, c). *HOXC4* expression was not different by BRCA status when adjusting for cellular composition. Other genes previously associated with *BRCA1* variants in histologically normal FTs^{24,27} also show strong correlations to cell composition (Supplementary Fig. 4d), and do not distinguish BRCAm samples in our data (Supplementary Fig. 4e).

Differences in RNA and protein expression between BRCAm and Non-BRCAm Fallopian tubes

We tested the same models used for differential methylation to find BRCAm–non-BRCAm differences using both RNA and protein expression (Table 2). We found 159 differentially expressed genes (DEGs) and 0 differentially expressed proteins (DEPs) when adjusting for stroma fraction as well as excluding postpartum FTs (Supplementary Data 3). Of note, accounting for stroma and postpartum samples removed 95.5% of DEGs and all 90 DEPs found in the univariate BRCA model (Fig. 3a–c; Supplementary Fig. 5a; Supplementary Data 3).

Genes lower in BRCAm were overwhelmingly immune-related immunoglobulin (IG) genes (Fig. 3c, d). However, the difference appeared to be driven by a subset of non-BRCAm samples. Unlike FTs from the BRCAm cases, which exclusively came from patients electing risk-reducing surgery, many of the non-BRCAm cases presented with clinically significant pathology including uterine fibroids, endometriosis, adenomyosis, and benign ovarian masses that can be associated with a local inflammatory response, or simply increase of vasculature (Fig. 3c; Table 1). Thus, the observed difference in immune activity is more likely associated with these benign conditions in only a subset of non-BRCAm cases, rather than immune downregulation in the BRCAm cases per se (Fig. 3c). Only two non-BRCAm FTs came from non-postpartum elective sterilization, which are expected to better represent normal states. Indeed, these two FTs did not show increased immune expression, and were similar to the risk-reduction BRCAm

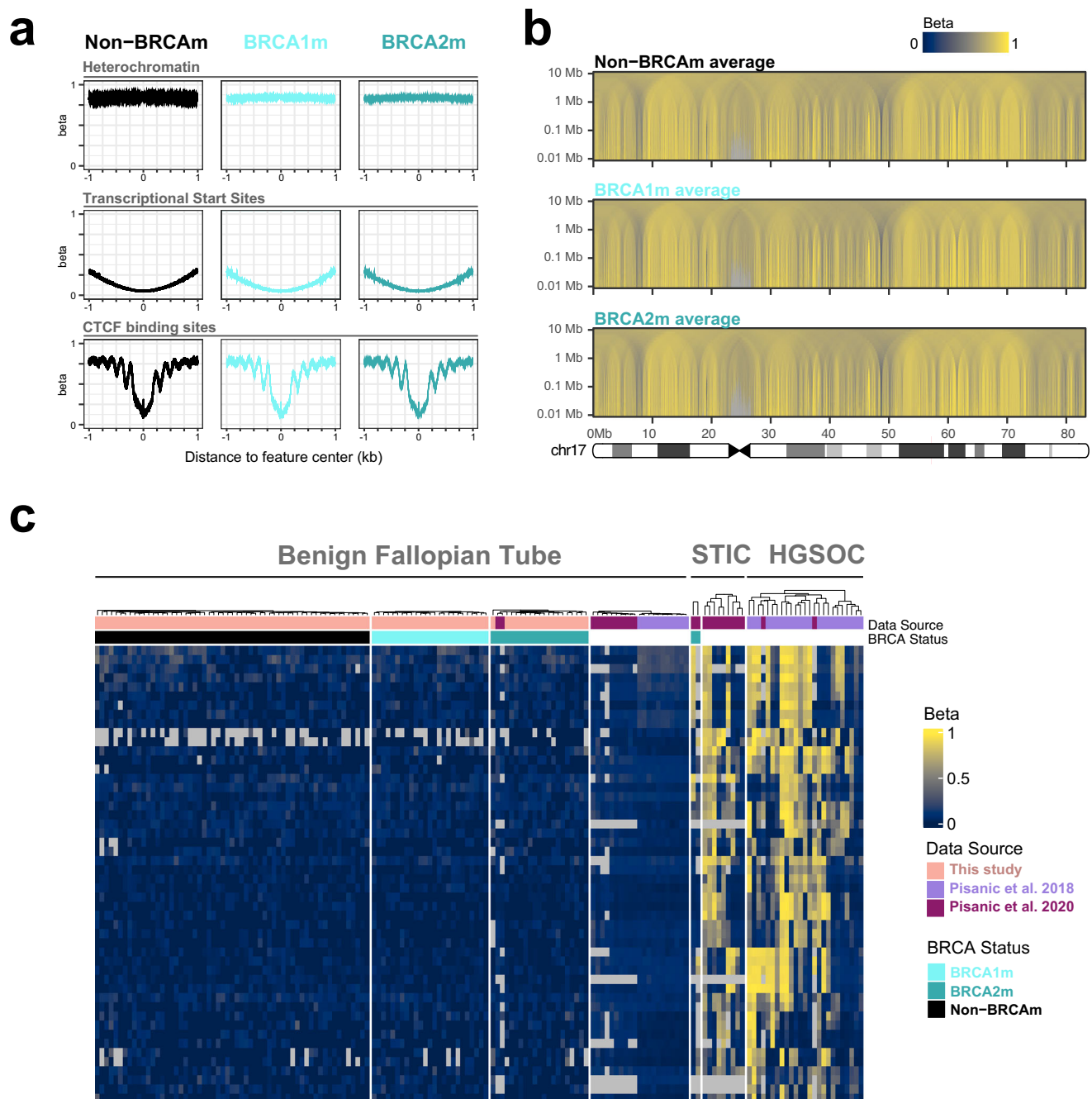


Fig. 2 | DNA methylation profiles of the cohort. a Region-centered binned averages show high DNA methylation levels for heterochromatin (top row), lack of methylation at transcriptional start sites (TSS; middle row), and nucleosome positioning around CTCF binding sites (bottom row). No difference in methylation by *BRCA* status at these features was detected. **b** Multiscale DNA methylation

averages along chromosome 17 from 10 kb to 10 Mb showing no *BRCA*-associated differences in methylation at any of these scales. **c** Heatmap showing DNA methylation levels at loci known to be hypermethylated in STIC or HGSC (rows) for samples of different origins (columns). Regardless of study origin and *BRCA* status, benign FTs do not show gain of methylation at these loci.

samples. Interestingly, many postpartum FTs also had increased immune gene expression, possibly due to increased microvasculature, consistent with increased endothelial signature in this group (Fig. 4a). A handful of DEGs upregulated in BRCAm were primarily mitochondria-encoded genes (Fig. 3e), but their fold change in BRCAm was low (fold change less than 1.5; Fig. 3c).

Protein products were only detected for 25 of the 159 DEGs, consistent with the very low expression level for many of these genes (Supplementary Fig. 5a). When these 25 proteins were examined, they reproduced the immune upregulation in a subset of non-BRCAm FTs

(Fig. 3f). Despite significant associations between gene expression and protein product (Supplementary Fig. 5b), differential gene expression patterns beyond immunoglobins were not reproduced in the protein data (Fig. 3f).

There was no difference in *BRCA1* (Fig. 3g) nor *BRCA2* (Fig. 3h) mRNA abundance between the non-BRCAm and BRCAm cases. Coverage was not high enough in the transcriptomics data to confidently quantify allelic expression. Neither *BRCA1* nor *BRCA2* were detected in the proteomics data. RNA expression of both *BRCA1* and *BRCA2* was highest in high-epithelial FTs (Supplementary Fig. 6a, b).

Table 2 | Summary of differential testing for DNA methylation, RNA, and protein expression

Contrasted Groups	Model	DMRs (Stouffer's Pval <0.05)	Differentially Expressed Genes	Differentially Expressed Proteins
BRCA DIFFERENCES				
BRCAm vs. Non-BRCAm	~BRCAm	10	3524	92
BRCAm vs. Non-BRCAm	~BRCAm + % Stroma	5	1569	0
BRCAm vs. Non-BRCAm excl. postpartum	~ BRCAm + % Stroma	0	159	0
MENOPAUSE DIFFERENCES				
Pre- vs. Post-menopause	~Menopause Status	36547	5826	966
Pre- vs. Post-menopause	~Menopause Status + % Stroma	164	696	4
Pre- vs. Post-menopause excl. postpartum	~Menopause Status + % Stroma	60	1769	0

DMR = differentially methylated regions.

Non-BRCAm related factors impact the cellular composition of primary human Fallopian tubes

Several recent studies have used single-cell technologies to obtain a detailed molecular profile of the cell types present in normal FTs^{40–43}, including extensive markers for stromal, epithelial, and immune cell types (Fig. 4a). We used the expression level of consensus marker genes (Supplementary Data 4) to define the cellular composition landscape in our bulk samples, and evaluate how they change with important covariates, especially age, menopause status, pregnancy, and menstrual phase.

As shown with DNA methylation data, the epithelial cell fraction dropped around and after menopause. Postmenopausal FTs consisted of mostly stroma and atrophic epithelium. Postpartum FTs and postmenopausal FTs were both strongly depleted for ciliated epithelial markers. Interestingly, while postpartum FTs, like the older postmenopausal FTs, were depleted of canonical epithelial markers such as *EPCAM*, they uniquely showed expression of *KRT8/18/19* and *TMEM190*, along with a subset of the secretory cell markers (e.g., *KRT7*, *PAX8*, and *MSLN*), but not others (*CSMD1* and *OVGP1*). Postpartum FTs were also strongly depleted of *ESR1* and *PGR* expression, presumably due to high circulating hormone levels (Fig. 4a). In contrast, postmenopausal FTs still expressed these hormonal receptors. Postpartum FTs also showed an increase in endothelial cell marker expression.

Based on these markers, pre-menopausal FTs clustered into two main groups: stromal-rich and epithelium-rich, with the epithelium-rich FTs further split into two groups, strongly associated the menstrual cycle state of patient-matched endometrium: proliferative/follicular, secretory/luteal, or inactive, for those patients where endometrium tissue was available (Fig. 4a). A subset of the putative luteal phase FTs that were also high in stroma grouped with the inactive samples. These samples exhibited high pericyte, muscle, and fibroblast marker expression, and also appeared to have more immune cells. Interestingly, the secretory epithelial markers themselves clustered into two groups: those expressed in the luteal phase, and those expressed in the follicular phase (Fig. 4a). An expanded secretory epithelial marker panel obtained from a high-resolution single cell method⁴⁴, which included well-known canonical markers such as *PAX8*, *MUC1* and *OVGP1*, also clustered into these two groups (Fig. 4b). This split by menstrual cycle phase in the normal FT is consistent with expression of secretory cell markers in normal endometrium⁴⁵ from both phases (Fig. 4b, left). Of these previously reported markers, secretory cells of the follicular (proliferative) phase had higher levels of *OVGP1*, *PLCB1*, *FMOD*, *PODXL*, *ANO1*, and *PKHD1L1*, while those of the luteal (secretory) phase had higher levels of *MUC1*, *C3*, and *PAX8* (Fig. 4b).

We contrasted gene expression in premenopausal luteal phase FTs vs. premenopausal follicular phase FTs and found that expression differences mirror those previously shown in cycling endometrium⁴⁶ ($\rho = 0.26$; P val < 2.2×10^{-16} ; Supplementary Fig. 7a–c; Supplementary Data 5). As seen in the endometrium, the secretory/luteal phase was associated with TGF-beta signaling, ECM-receptor interaction, as well as the Notch pathway, while the proliferative/follicular phase showed upregulation of genes involved in oxidative phosphorylation and cell cycle (Supplementary Fig. 7d).

Based on these results, we categorized the premenopause FTs into four clusters: (1) follicular, (2) luteal, (3) inactive/stroma-rich, and (4) postpartum. Postmenopause samples constitute a fifth major cluster. H&E staining confirmed the presence of the stroma-rich subset identified by RNA-seq and protein data in the inactive samples (Fig. 5a). Postpartum samples also exhibit substantially higher stromal content compared to other pre-menopause FTs (Fig. 5a). Both postpartum and inactive clusters appear to have more blood vessels compared to postmenopause samples (Fig. 5a), consistent with the molecular data (Fig. 4a).

We identified markers for the five major FT clusters independently in RNA (7495 unique genes at 5% FDR; Supplementary Fig. 8a; Supplementary Data 6) and protein (936 unique proteins at 5% FDR; Supplementary Fig. 8b; Supplementary Data 7). Many genes had $FDR < 0.05$ for >1 group, so markers were further grouped by their intersects (Supplementary Fig. 9a). Postpartum samples had the most unique markers across data types, followed by the luteal and follicular clusters (Supplementary Fig. 9a). A total of 383 unique genes demarcated the same cluster ($FDR < 0.05$) in both RNA and protein (Supplementary Fig. 9a, right). From these consistent markers, the top 10 by P value are shown for RNA (Fig. 5b) and protein (Fig. 5c). For markers that were only significant in one data modality and not the other (at 5% FDR level), similar trends in the other data type were visible (Supplementary Fig. 9b).

Overall, postpartum FTs tended to be associated with higher expression of cell adhesion molecules, including laminin genes. In addition, laminin subunits showed an interesting switch between the groups. Postpartum FTs were uniquely marked by *LAMBI*, while *LAMA4* and *LAMB2* were expressed postmenopause. The inactive cluster has high stroma, so many of its markers are shared with postmenopause (Fig. 5b); however, many proteins (e.g., *ALDH1A1* and *COL6A1*) were up only postmenopause. Premenopause follicular and luteal also shared a substantial fraction of their RNA and protein markers, although each showed their own robust unique gene expression program. Among others, follicular FTs were marked by *OVGP1* and *CPM*. PreM-L was marked by *ALDH5A1* and *SRGAP3* (Fig. 5c). Interestingly, postpartum

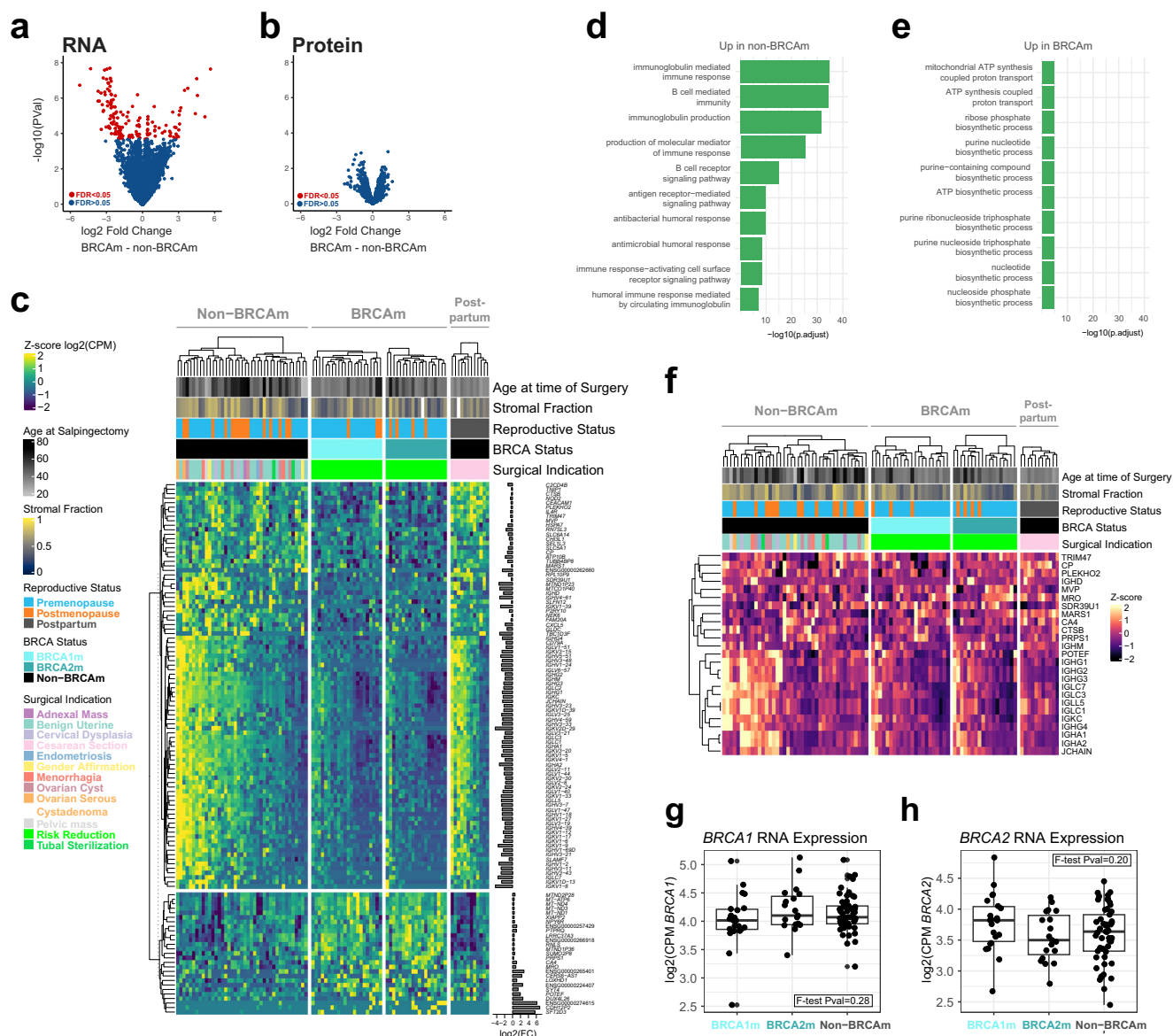


Fig. 3 | Differential gene and protein expression between non-BRCAm and BRCAm FTs. a Volcano plot for all genes between BRCAm and non-BRCAm adjusting for stromal fraction and excluding postpartum samples. X-axis shows the fold change and y-axis shows the $-\log_{10}(P\text{ value})$ for each gene from quasi-likelihood F tests. **b** As in (a) but for proteomics data with P values from moderated t tests. **c** Heatmap of 117 of 159 differentially expressed genes shown in (a) comparing non-BRCAm and BRCAm. Not shown are 42 genes with less than one counts per million (CPM) in 95% of samples. Supplementary Fig. 5a shows all 159 DEGs and their average log CPM values. Rows are grouped by up in non-BRCAm (top) or up in BRCAm (bottom); **d** Gene ontology (GO) enrichment for genes up in non-

BRCAm showing the top 10 enriched terms all of which are related to immunoglobulin mediated immunity. P values are from Fisher exact tests. **e** GO enrichment terms which are up in BRCAm shown in (c) showing relatively lower enrichment by P value for terms mostly involving core metabolism. P values from Fisher exact test. **f** Twenty-five protein products from DEGs were identified showing immune hot samples but otherwise few consistent changes. **g** Boxplot of *BRCA1* expression by *BRCA1/2* mutation status showing no difference between groups (BRCA1 $n = 22$; BRCA1 $n = 19$; non-BRCAm $n = 53$). Y-axis is log2(CPM). **h** as in (g) but for *BRCA2* expression. P values derived from F test conducted with ANOVA models.

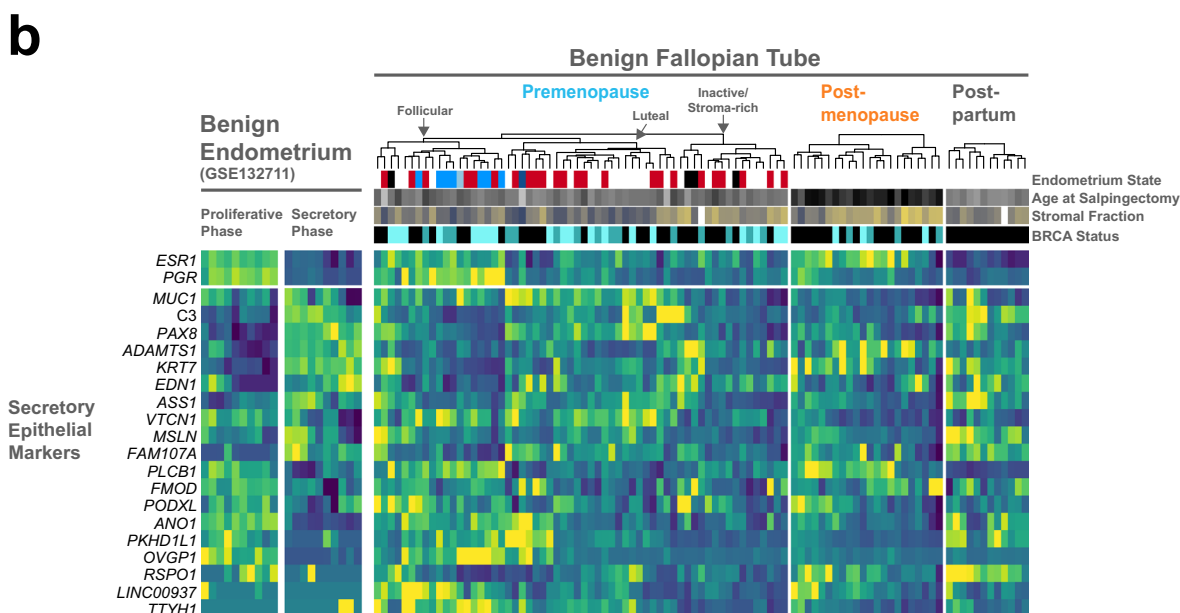
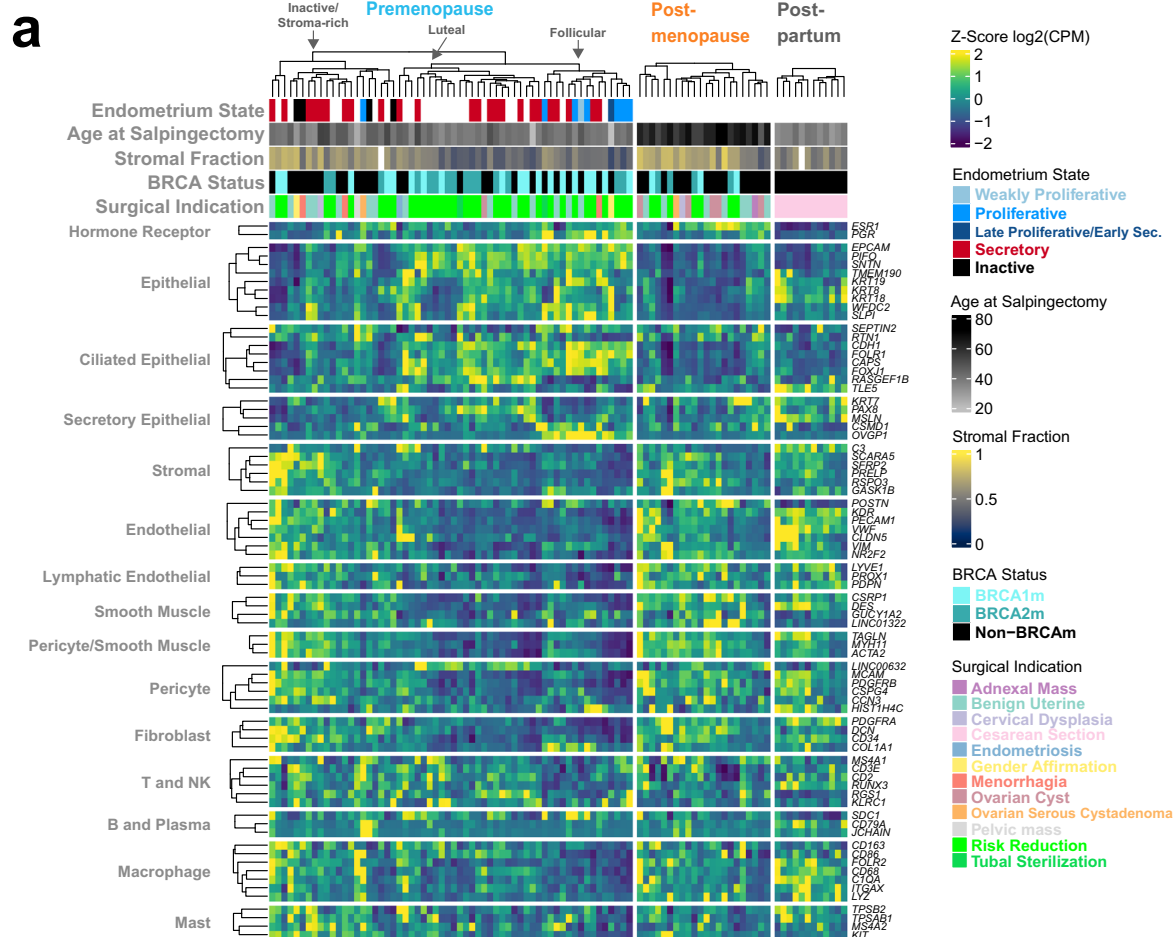
and luteal FTs also shared many markers including monoamine oxidase A (MAOA).

Aging and estrogen signaling in normal Fallopian tube

Some studies suggest an association between cellular aging and heterozygous germline *BRCA1/2* mutations in vitro³¹. As DNA methylation represents a unique and powerful tool to assess biological aging, we inferred the biological age of these samples with the Horvath Clock⁴⁷. The inferred age from our sequencing data correlated with clinical age at the time of salpingectomy ($\rho = 0.55$; $P\text{ value} = 1.2 \times 10^{-9}$; Fig. 6a). Using this inferred biological age, we did not observe any difference in acceleration/deceleration of aging between the BRCA

groups (Fig. 6b). In addition, it was previously shown that DNA methylation in a special sequence context, dubbed solo-WCGW, could be used to measure replicative history of a cell^{37,48}. Similar to the Horvath clock, the solo-WCGW levels were also consistent with a lack of difference in replicative history among the BRCAm carriers and controls (Fig. 6c).

We next investigated how aging affects estrogen signaling, as increasing *ESR1* promoter methylation has been reported to be associated with age in various tissue types^{49,50}. In the human FT, there does not seem to be observable increased *ESR1* promoter methylation with age ($\rho = 0.056$; $P\text{ value} = 0.55$; Supplementary Fig. 10a), although there is gain of methylation flanking the CpG island surrounding the



canonical TSS and A/B promoter (as annotated in Lung et al.⁵¹) in some samples independent of age (Supplementary Fig. 10b). DNA methylation level of ER binding sites does increase with age ($\rho = 0.37$; $P_{\text{val}} = 1.0 \times 10^{-4}$), but more strongly tracks tissue composition than age (Fig. 6d). When both age and stroma fraction are jointly modeled, only stroma fraction was significant ($P_{\text{val}} < 2 \times 10^{-16}$ for stroma; $P_{\text{val}} = 0.90$

for age). This suggests that ER enhancers are primarily active in the epithelium. High-epithelium samples (>70%) showed decreasing solo-WCGW methylation levels with age (Fig. 6e), while overall solo-WCGW level positively correlated with age (Supplementary Fig. 10c). This opposite trend in all samples is likely confounded by tissue composition (Fig. 6e; Supplementary Fig. 10d), as solo-WCGW

Fig. 4 | Cellular landscapes of normal human Fallopian tube samples. **a** Gene expression heatmap for known markers of various cell types present in the FT (Supplementary Data 4). Rows represent individual markers and columns represent individual FTs; markers are grouped by their cell type, and samples are clustered freely within premenopause, postmenopause or postpartum. A luteal, follicular, and inactive/stroma-rich group within the premenopause samples consistent with molecular and clinical data are indicated by arrows pointing to the root node.

b Heatmap of an expanded panel of secretory epithelial cell markers for normal FT (this study, right) and normal endometrium (GSE132711, left). From these markers and those in (a), there is a visible high stroma subgroup within the premenopausal

FTs (root node designated with an arrow) that contains most of the inactive endometrium samples. This high stroma subgroup is associated with relatively lower secretory cell marker expression than other premenopause samples. The remaining epithelium-high premenopausal FT samples showed gene expression changes similar to endometrium (i.e., *ESR1* and *PGR*), and were divided into a proliferative and a secretory cluster (root node designated with an arrow). These patterns of expression corresponded to the pathology-determined endometrium state from sample-matched endometrium. Postmenopause and postpartum samples do not show menstrual cycling but do have distinct patterns of secretory epithelial marker expression.

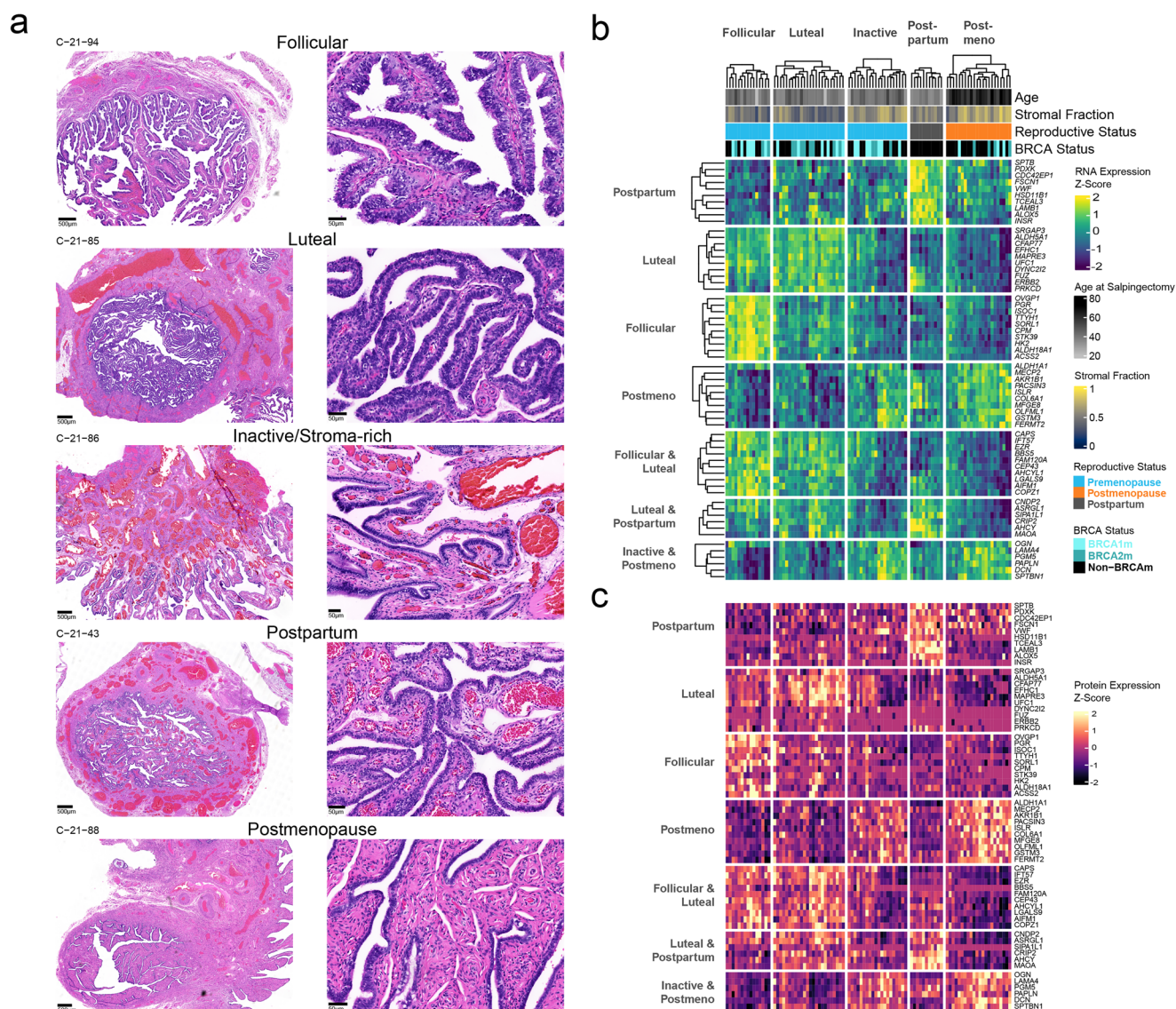


Fig. 5 | Five molecular states of normal human Fallopian tube. **a** FTs were imaged following H&E staining for all samples where available to examine cell type composition. H&E slides for five representative samples from the five expression groups identified in Fig. 4 are shown. Left panels show a cross-sectional cut of a representative distal FT (scale bar 500 μ m) and right panels are a zoom in on epithelial, stromal, and immune cells within these same slides (scale bars represent 50 μ m). **b**, **c** Marker genes for the five expression groups were identified

independently in the RNA and protein. Features with both a 1:1 match in the RNA and protein and $FDR < 0.05$ for the same marker cluster in both assays are shown for gene expression (b) and protein expression (c). These markers are grouped by the intersect of which sample group they mark (Supplementary Fig. 9a, right). Samples and features are clustered based on their RNA expression in (b), and this same order is used for the corresponding protein products (c).

methylation shows a much stronger association with stroma fraction (Supplementary Fig. 10d). These results suggest that the stromal cells retain their solo-WCGW methylation better than epithelial cells, indicating slower turnover. Indeed, these megabase blocks of loss of DNA

methylation is evident in samples with high epithelial fraction, and not in samples with high stroma fraction (Fig. 6f). Together, these results suggest that the epithelial compartment is likely where estrogen signaling and cellular turnover primarily occur in the FT.

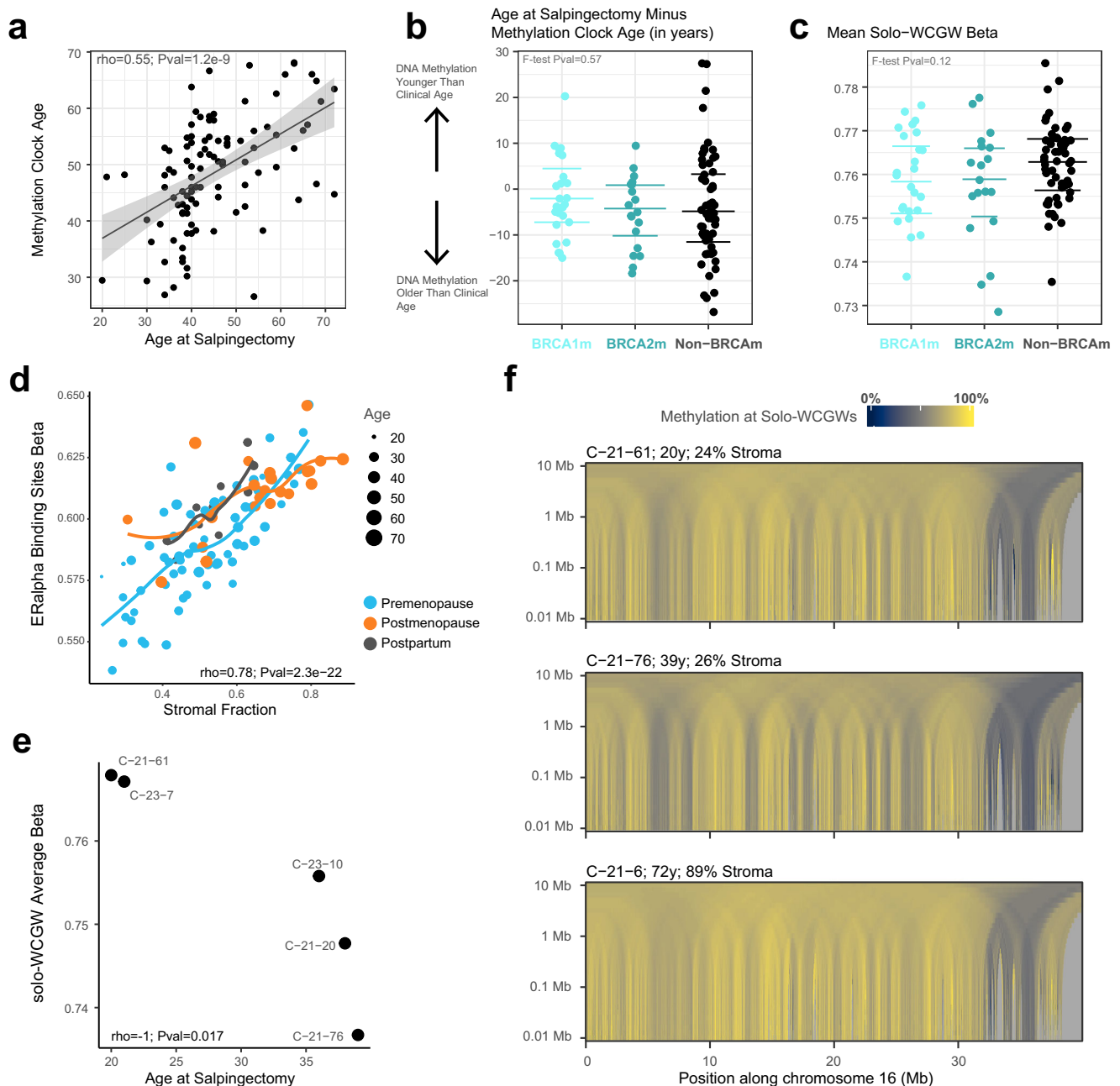


Fig. 6 | Aging in the normal Fallopian tube. **a** Clinical age at time of salpingectomy (x-axis) correlates with predicted age from the Horvath methylation clock. A linear line of best fit is shown as a black line; shading around the best fit line indicates the 95% confidence interval as done with ggplot2 function geom_smooth(method = "lm"). A Spearman rank correlation coefficient was found to be significant with $p = 1.2\text{e-}9$. **b** Differences between predicted and clinical age (as indication of accelerated or decelerated aging) are not different by *BRCA* status (F test $p = 0.57$ from ANOVA model; *BRCA1m* $n = 25$; *BRCA2m* $n = 19$; non-*BRCAM* $n = 59$). The y-axis is the difference between predicted and clinical age with horizontal lines representing the first quartile, median, and third quartile of each x-axis group. **c** DNA methylation at solo-WCGWs, which lose methylation during aging, are not different by *BRCA1/2* status (F test $p = 0.12$ from ANOVA model; *BRCA1m* $n = 25$; *BRCA2m* $n = 19$; non-*BRCAM* $n = 59$). Horizontal lines representing the first quartile,

median, and third quartile. **d** Average DNA methylation at Eralpha binding sites from JASPAR database (y-axis) positively correlates with stroma fraction (x-axis) as determined by Spearman rank correlation ($p = 2.3\text{e-}22$; $\rho = 0.78$). LOESS smoothed lines are displayed for each of the reproductive status groups with the 95% confidence interval represented by the shaded region. **e** When restricted to samples with more than 70% of epithelium, solo-WCGW methylation is inversely associated with age at time of sampling, as expected for a more homogeneous population. Spearman rank correlation coefficient ρ was -1 with $p = 0.017$. **f** Multi-scale plot showing DNA methylation profile at solo-WCGWs for the pet arm of chromosome 6 for two samples with high epithelium and different age (20 years and 39 years) and the oldest sample with high stromal content. Mega-base DNA methylation loss is associated with age in the epithelium in normal human FT, but not in the stroma compartment.

Discussion

BRCA1 and *BRCA2* are among the most widely known germline cancer susceptibility genes. With the high penetrance of *BRCA1/2* mutations for breast and ovarian cancer risk, there has been intense research

interest surrounding whether these germline mutations cause any genomic/epigenomic alterations before LOH or any visible aberration occurs; particularly in disease-relevant tissue types such as the FT for ovarian cancer, and breast tissue for breast cancer. Indeed, prior

studies claimed prevalent epigenetic, transcriptional, and other biological differences (e.g., accelerated aging) in both FTs and peripheral blood of *BRCA1/2* carriers^{23,24,26,28,29}. Our study shows that germline heterozygous *BRCA1/2* mutations per se are not associated with extensive changes in the human FT, indicating that there is likely no haploinsufficiency for *BRCA1/2* functions. LOH, or another form of secondary hit such as promoter DNA methylation for *BRCA1*, is likely required before any further molecular alterations are incurred. This is consistent with LOH being almost always observed in primary HGSC samples from *BRCA1/2* carriers^{52,53}. It is further consistent with early molecular studies showing LOH in 6/6 STIC or microinvasive cancers, but not in p53 signatures nor histologically normal FT epithelium⁵⁴.

One caveat is that we treated *BRCA1* and *BRCA2* pathogenic variant carriers as one group when comparing to control. However, *BRCA1* and *BRCA2* have distinct functions and are associated with different cancer risk⁵⁵. Analyses with *BRCA1* and *BRCA2* independently did not identify any differences to control beyond what is presented here. Another important consideration is that we are examining bulk primary samples directly from patients. One copy of *BRCA1* and *BRCA2* may be sufficient in most cells when cellular stress level is normal. However, these same cells could exhibit haploinsufficiency when challenged with genotoxic conditions, and this may explain findings of haploinsufficiency in vitro^{22,56}.

We show in this paper that cellular composition of the FT tissue has the biggest impact on readouts from bulk-omics assays. We further show that this cellular composition changes with age, menopausal status, and hormonal status (e.g., postpartum). Importantly, risk-reducing surgery following any desired reproduction is standard of care for pathogenic *BRCA1/2* carriers. The FTs from non-BRCa patients, in contrast, originate from various clinical presentations including fibroids, abnormal bleeding, or removal during c-section. These different clinical features result in a generally younger patient sample population for BRCa, as well as a lower fraction of benign conditions affecting the reproductive tract. Similar to other BRCa-related studies^{24,27}, we made an effort to age-match by purposefully collecting younger non-BRCa FTs. However, we show that even in similarly-aged premenopausal FTs, cell composition can vary widely. Therefore, the clinical covariate of age does not fully account for the effects of cell composition, hormone status, nor surgical indication.

It is also important to note the significance of surgical indications as a potential confounder in BRCa-related studies. As stated above, BRCa carriers often have their breast or FT tissues removed for prophylactic purposes, while non-BRCa FTs originate from benign conditions. We expect this confounding to be common in human sample based *BRCA1/2* studies, as it is hard to bypass. Importantly, we show that postpartum FTs have a unique signature, and that many FTs from salpingectomies for benign uterine conditions have immune activation. In this study, we have demonstrated that the remaining transcriptomic differences between BRCa and non-BRCa after adjusting for epithelium/stromal composition are likely due to slightly elevated immune reaction in the non-BRCa group. Benign uterine conditions including leiomyoma, were the most common indication for gynecologic surgery in our control group. Chronic inflammation and immune response have been associated with the pathogenesis of leiomyoma^{57,58}, and likely contribute to the inflammatory molecular signature we observed in some of our control samples. Endometriosis, another common indication for benign gynecologic surgery, is considered to be a chronic inflammatory disease, characterized by dysregulation of multiple immune pathways with macrophages, NK cells, and B cells playing major roles⁵⁹. In light of this, results that indicate immune cell differences in BRCa carriers need to be carefully interpreted. Confounding may be more pronounced with small sample sizes, such as in single cell or spatial datasets. Indeed, recent single cell studies on *BRCA1/2* mutated FTs showed changes in immune signatures^{25,60}. The similar analogy applies to BRCa-associated normal

breast studies, as normal breast tissues from non-BRCa patients is often from individuals undergoing breast reduction and will therefore often have a different cellular composition than BRCa-mutated samples predominately from prophylactic mastectomy.

This confounding can also manifest in various ways in different studies. In our study, after identifying a difference in age between the BRCa groups within our initial smaller cohort, we sought to balance the age distribution by including more young, premenopausal non-BRCa individuals. Simultaneously, we requested additional FTs from Black individuals to improve representation. However, we later recognized that postpartum surgeries were disproportionately represented as a surgical indication for salpingectomy among younger, premenopausal non-BRCa individuals; as a result, many postpartum FTs were added to our cohort at this stage. With our concurrent efforts to include younger and Black non-BRCa individuals, these postpartum FTs tend to be from Black individuals. This overlap introduced an imbalance that could lead to apparent race-associated molecular differences, which were actually confounded by surgical indications in our investigations.

Individual cells in the bulk tissue carrying germline *BRCA1/2* mutations could have LOH and associated molecular alterations, which would be undetectable in bulk assays before substantial clonal expansion. High-depth (either read depth for bulk studies, or cell number for single-cell studies) or high-fidelity methods for clonal/subclonal studies would be needed to study these early events. While these studies will not be confounded by cell composition, this study provides some guidance on other potential confounders to consider.

We showed extensive differences between follicular and luteal phase FTs in pre-menopause samples. These transcriptional variations with menstrual cycle modulation were similar to changes in the endometrium, albeit to a lower magnitude. Our study based on the correlation between cellular fraction and ER-binding enhancers shows that estrogen response is primarily limited to the epithelium in the FT, in contrast to the well-described estrogen-responsiveness of endometrial stroma⁶¹. This likely explains the magnitude differences between menstrual phases in the endometrium and FT. We show that in both the endometrium and FT, canonical secretory cell markers vary with menstrual cycle, with *OVGP1* expressed higher in the follicular phase, and *PAX8/MUC1/KRT7* in the luteal phase. Many FT and/or endometrium-based studies use these markers to drive transgene expression^{62,63}, and this phase-related difference in expression of these key genes is important to consider.

While most non-pregnant pre-menopause FT samples have more epithelium than postmenopausal and postpartum samples, we identified a subset of samples that showed low epithelium content, mostly within the luteal phase. This group included all four samples in our cohort that had an associated endometrium annotated as inactive. Inactive endometrium is often associated with the continuous use of oral contraceptive pills (OCP)⁶⁴ and hormonal intrauterine device (IUD)⁶⁵. Therefore, we hypothesize that this could be partially due to a hormonal effect. This, if true, could explain the protective effects of OCP use on ovarian cancer. We collected contraceptive use status and history and made them available as part of the clinical data (Supplementary Data 1). However, the hormonal IUD or OCP use data for these samples have not been collected in a consistent manner, making it hard to test this hypothesis. Furthermore, contraceptive use history does not explain all of these inactive cases. Gene and protein expression points to a more similar profile to postmenopause FTs, instead of the postpartum FT that are highly influenced by progesterone. The age of these patients are comparable to other premenopause patients, and are unlikely to be peri-menopausal. Another possibility is that this group may include Polycystic ovary syndrome patients⁶⁶. Understanding why this subset of epithelium-sparse FT fimbria occurs could be key to understanding fertility variations, and potential additional factors that contribute to differences in ovarian cancer risk.

Independent of the relevance for *BRCA1/2* and ovarian cancer research, this study is a comprehensive survey of FT cellular, epigenomic, transcriptomic, and proteomic landscape of histologically normal human FT, an organ with both reproductive and oncological implications. This cohort shows cellular and molecular variations associated with age, pregnancy, reproductive status, menstrual cycle phase, and surgical indication. These results are key to understanding FT biology, which in turn will inform strategies for cancer prevention and early detection.

Methods

Sample collection for high and average risk Fallopian tubes

The cases for this study were obtained with patient consent and study approval by the Institutional Review Board at the University of Pennsylvania. Tissues of the fimbriated and proximal portions of historically benign FTs were fixed in 10% neutral buffered formalin and processed on a Leica 300 ASP tissue processor to create FFPE blocks. The cases utilized were all females. There were 55 BRCA mutation carrier blocks and 70 average risk cases collected. Clinical data for these cases were obtained from electronic medical records (Supplementary Data 1).

DNA and RNA extraction and quality assessment

FFPE block tissue sections (10 μ m thick sections) were submitted to the VAI Pathology and Biorepository Core (PBC) in 4 batches over 4 years. Each batch was processed and submitted for sequencing at the Van Andel Institute Genomics Core. Each block had 8 total 10 μ m sections cut and shipped to the PBC. Four 10 μ m sections were placed in an Eppendorf tube for extraction (2 tubes per block). 2 tubes (8 total 10 μ m sections) were used for each extraction. All batches were processed using the Qiagen Allprep DNA/RNA FFPE kit following the manufacturer's protocol for recovering total RNA, including small RNAs. Deparaffinization Solution was used for the deparaffinization. Quantification and qualification of nucleic acids was performed by Qubit Fluorometer and Agilent TapeStation. Extracted RNA samples with DV200 > 30 were deemed sufficient to continue library preparation and sequencing. Extracted DNA samples with an average material size of 1000 bp or greater were deemed sufficient to continue to library preparation and sequencing.

Construction and sequencing of directional total RNA-seq libraries

Libraries were prepared by the Van Andel Genomics Core from 500 ng of total RNA using the KAPA RNA HyperPrep Kit (Kapa Biosystems, Wilmington, MA USA). Ribosomal RNA material was reduced using the QIAseq FastSelect -rRNA HMR Kit (Qiagen, Germantown, MD, USA). RNA was sheared to 300–400 bp and was converted to cDNA. cDNA fragments were ligated to IDT for Illumina TruSeq UD Indexed adapters (Illumina Inc, San Diego CA, USA). Following adapter ligation, 8 cycles of PCR were performed. Quality and quantity of the finished libraries were assessed using a combination of Agilent DNA High Sensitivity chip (Agilent Technologies, Inc.), QuantiFluor® dsDNA System (Promega Corp., Madison, WI, USA), and Kapa Illumina Library Quantification qPCR assays (Kapa Biosystems). Individually indexed libraries were pooled and 100 bp, paired-end sequencing was performed on an Illumina NovaSeq6000 sequencer to an average depth of 50 M raw paired-reads per sample. Base calling was done by Illumina RTA3 and the output of NCS was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v1.9.0.

Construction and sequencing of WGBS libraries

Libraries were prepared by the Van Andel Genomics Core from 200–250 ng of extracted DNA using the IDT xGen Methylation-Sequencing DNA Library Preparation Kit (IDT, Cat #10009824, Accel-NGS Methyl-Seq DNA Library kit (v3.0), Swift Biosciences, Cat. #30024). DNA was sheared following the manufacturer's protocol to

an average size of 250 bp, and sheared DNA was bisulfite converted using the EZ DNA Methylation-Gold kit (Zymo Research, Cat. #D5005) with an elution volume of 15 μ l. Following adapter ligation, 6 cycles of library amplification were performed. Quality and quantity of the finished library pools were assessed using a combination of Agilent DNA High Sensitivity chip (Agilent Technologies, Inc.), and QuantiFluor® dsDNA System (Promega Corp., Madison, WI, USA). 100 bp, paired-end sequencing was performed on an Illumina NovaSeq6000 sequencer using an S4, 200 bp sequencing kit (Illumina Inc., San Diego, CA, USA) with 10% PhiX included to improve base diversity. Each sample was sequenced to a minimum raw depth of 450 M reads. Base calling was done by Illumina RTA3 and the output of NCS was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v1.9.0.

FFPE sample processing for proteomic analysis

Each of the 155 samples (including multiple technical replicates for some samples from which final abundance values were averaged, see Supplementary Data 1 and 8) consisted of five 4 μ m unstained FFPE sections placed in 2 mL Eppendorf tubes. Three 1 mm zirconium beads and 125 μ l of xylenes were added to each tube. Paraffin solubilization and tissue homogenization were performed using three 30-s cycles on a Beadbeater 16 apparatus (BioSpec Products, Bartlesville, OK). 125 μ l of 2X S-Trap extraction buffer (10% sodium dodecyl sulfate [SDS], 100 mM triethyl ammonium bicarbonate [TEAB]) was added to each sample. Protein solubilization was achieved by vortexing for 15 min at room temperature. Disulfide bonds were reduced at 56 °C for 20 min with 1 mM dithiothreitol. Reduced cysteines were alkylated at room temperature for 20 min with 2.25 mM iodoacetamide. Proteins were acidified with 25 μ l of 12% phosphoric acid and precipitated with 1.6 mL S-Trap binding buffer (100 mM TEAB, 90% methanol). Protein precipitates were collected by centrifugation at 20,000 $\times g$ for 10 min at room temperature. 1.5 mL of the supernatants were discarded. The protein pellets were resuspended in the remaining 400 μ l of supernatant and transferred to the wells of a 96-well S-Trap plate (ProtiFi, Fairport, NY). Proteins captured on the S-Trap matrix were washed once with 400 μ l of 50% chloroform/50% methanol and 5 times with 400 μ l of S-Trap binding buffer using a vacuum manifold for 96-well plates. The S-Trap plates were centrifuged at 800 $\times g$ for 15 s to collect residual binding buffer to the bottom of the wells and placed on the vacuum manifold for 10 s. Captured proteins were digested overnight at 37 °C using 5 μ g of trypsin diluted in 125 μ l of S-Trap digestion buffer (50 mM TEAB). The digested peptides were recovered in 2 mL collection plates (Waters, Milford, MA) by centrifugation at 800 $\times g$ for 1 min. A second 3-h digestion was performed using 1 μ g of trypsin diluted in 100 μ l of S-Trap digestion buffer. Peptides were recovered by centrifugation at 800 $\times g$ after each addition of 80 μ l of S-Trap digestion buffer, 80 μ l of 0.2% formic acid and 80 μ l of 50% acetonitrile.

Peptide fractionation (for spectral library generation)

To create a pooled sample, 0.5% of each sample was combined in a single tube, dried to completion, and solubilized in 10 μ l of 20 mM ammonium formate pH10 containing 2% acetonitrile. Peptides were loaded onto a 7 cm $\times$ 150 μ m column packed with 5 μ m XBridge BEH130 C18 beads (Waters). Peptides were fractionated in a concatenated fashion across a 60-min gradient from 2% to 35% acetonitrile using 20 mM ammonium formate pH10 as the aqueous mobile phase and directly collected into 24 conditioned Evotips Pure (Evosep, Odense, Denmark) containing 190 μ l of 0.1% formic acid⁶⁷. Evotips were handled as recommended by the manufacturer.

LC-MS/MS analysis

Peptides were separated using a standardized 31-min gradient “Whisper” method (40SPD) with 0.1% FA and 0.1% FA/99.9% ACN as the mobile phases using an Evosep One liquid chromatography system (Evosep) connected to a 15 cm $\times$ 75 μ m column packed with 1.9 μ m

ReproSil-Pur 120 C18-AQ reverse phase beads (Dr. Maisch, Ammerbuch-Entringen, Germany) and introduced into a timsTOF fleX MALDI-2 mass spectrometer (Bruker, Billerica, MA). Spectral library generation: Fractionated peptides from the pooled sample were analyzed using a data dependent acquisition (DDA) strategy with the mass spectrometer executing 10 PASEF scans (charge 0–5, ramp time: 100 ms, target intensity: 14,500, intensity threshold: 1750, m/z range: 150–1300, mobility range: 0.6–1.6). Active exclusion was enabled with a release time of 0.4 min. Analysis of individual FFPE samples: Peptides from individual samples were analyzed using data independent acquisition (DIA) mode with the mass spectrometer executing 12 DIA-PASEF scans with 2 isolation windows per scan as optimized using the py_diaID algorithm⁶⁸ (m/z range: 300–1200, mobility range: 0.7–1.4).

Spectral library generation

The project-specific spectral library was generated from the DDA runs in the FragPipe computational environment using MSFragger v3.8⁶⁹. Search parameters included lysine formylation, variable oxidation of methionine, N-terminus methionine excision and acetylation and fixed carbamidomethylation of cysteine, and all other parameters set as default. Spectra were searched against a human UniProt database of reviewed sequences (downloaded on 8/8/2023) including common contaminants and using reversed sequences as decoys. MSBooster⁷⁰ was used to predict spectra and retention times. Peptide-spectral matches were rescored using default parameters in Percolator⁷¹ and filtered to 1% false discovery rate. Protein hits were validated using default parameters in ProteinProphet⁷². The spectral library was generated using default parameters.

Protein identification and quantitation

Protein identification and quantification was performed on each individual sample analyzed in DIA mode using DIA-NN⁷³ (v1.8.2 beta 8). The search was performed against the spectral library generated in FragPipe (see above) and included the following parameters: lysine formylation, variable oxidation of methionine, N-terminus methionine excision and fixed carbamidomethylation of cysteine, with 2 missed tryptic cleavages. Additional parameter ranges were specified: peptide length: 6–40, charge state: 2–5, precursor m/z: 300–1700, fragment ion: 200–1700. Search results were re-annotated using the UniProt database used for the MSFragger search. The parameters for the quantitation algorithm were set as follows: mass accuracy: 10 ppm, MS1 accuracy: 15 ppm with match between runs (MBR) enabled, allowed isotopologues, shared spectra exclusion. Gene-based protein inference, double-pass neural network with robust LC quantification, retention time dependent normalization and smart profiling were used as additional parameters. Quantitation values for each protein were used in analyses.

Histology

Histologic slides from the FFPE samples were imaged on the 3DHistech Panoramic MIDI II (catalog number MI2BGHM00010070001) using the Panoramic Scanner Software Version 3.0.3 at 86X magnification. SVS image files were then imported to QuPath v0.5.1 to crop and add the scale bars.

RNAseq data analysis

Expression counts from RNA-seq sequencing relied on a pipeline based on STAR v2.7.1⁷⁴ and RSEM v1.3.3⁷⁵. Specifically, prior to alignment, reads were quality filtered using fastp with default settings⁷⁶. Next, reads were aligned to an rRNA reference using bwa with the putative rRNA mapped reads removed from downstream analyses⁷⁷. Using STAR, the remaining non-rRNA reads were mapped to the GRCh38.90 reference. We utilized the expected counts for genes from RSEM.

We excluded samples with a library size of fewer than 8 million reads or less than a 70% alignment rate. We normalized counts using

the weighted trimmed mean of M values method from the Bioconductor edgeR v4.2.0 package⁷⁸. We then excluded genes with minimum count less than 100 across 94 samples with expression data. This resulted in 27,303 features for differential testing. We corrected for batch effect while preserving any *BRCA* effect in the RNA counts using the Bayes model in the ComBat_seq function of the R package sva v3.5.0^{79,80}. P values were derived from a quasi-likelihood F test following data fitting using a negative binomial generalized log-linear model as implemented with edgeR functions glmQLFit and glmQLFTest.

Marker Identification

RNA and protein markers were identified using findMarkers function from scan R package v1.32.0⁸¹. Markers were found using the log of the expression values with the following options: pval.type = “some”, direction = “up”, test.type = “t”, and min.proportion = 0.5. Upset plots were made with the complex upset v1.3.3⁸².

DNA methylation data analysis

FASTQ files were aligned using BISulfite-seq CUI Toolkit (BISCUIT) v1.2.0 wrapped within a Snakemake workflow available on github: https://github.com/huishenlab/Biscuit_Snakemake_Workflow^{83,84}. This was run using Snakemake v7.25.0+. This workflow also performs several other quality control steps. As recommended with the Accel-NGS® Methyl-Seq DNA Library Kit, we first trimmed 15 nucleotides from the end of R2 as these positions are not fully converted. Next, trimmed reads were put through our “biscuit sifter” step which aligns read pairs with BISCUIT and subsequently removes PCR duplicates before outputting a sorted, aligned, and indexed BAM file⁸⁵. BISCUIT is then used again to ultimately generate a BED file with methylation information. This BED file is read into R using the biscuiteer R package as a bsseq object. This data was then used for analyses including differential methylation, plotting heatmaps, and estimation of stromal composition.

The Snakemake workflow also outputs a region centered binned average methylation from input bed files. This takes the regions, centers them, and calculates the average of 200 bp windows in sliding steps of 5 out to 1000 bp from the region center (specified in the workflow config). This was used to show methylation around features of interest (in our case, CTCF binding sites, transcriptional start sites, and heterochromatin). The Snakemake workflow also generates a binned average methylation matrix using 10,000 bp bins. Finally, a second Snakemake workflow was used to generate the multiscale methylation plot (https://github.com/huishenlab/multiscale_methylation_plot_pipeline). This averages DNA methylation in 10 kb to 10 Mb bins and visualized using the bisplotti R package available on github (<https://github.com/huishenlab/bisplotti>).

In all BRCa aligned libraries, except two where the exact clinical mutation was not known, we had sufficient coverage in the WGBS to manually confirm the clinical mutation. We did not observe the germline mutations in the RNAseq data due to low transcript abundance. WGBS coverage at *BRCA1* averaged 18x and *BRCA2* 13x. We identified high quality heterozygous SNPs in the WGBS data that were also present in NCBI Clinical Variant Database (https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh38/) using BISCUIT; no variants of known pathogenicity were identified in the non-BRCa cohort.

Differential gene expression

A negative binomial generalized log-linear model was then fit to the filtered and batch corrected count data with edgeR v4.2.0 using the weighted trimmed mean of M values to normalize for library size and composition biases⁷⁸. Groups were compared with and without covariates as indicated in Table 2. P values were estimated using empirical Bayes quasi-likelihood F -tests and adjusted for multiple comparisons using the Benjamini-Hochberg (BH) method; adjusted P values less than 0.05 were considered significant.

Differential methylation

To identify regions of differential methylation between non-BRCAM patients and those with germline *BRCA1* or *BRCA2* mutations, we used the R package DMRcate v2.12.0⁸⁶. Our design matrix was generated using the edgeR function modelMatrixMeth⁷⁸. Subsequently, the contrast matrix, design matrix, and a bsseq object with all CpGs for all samples in any specific contrast was input into the DMRcate pre-processing function sequencing.annotate with options all.cov = TRUE and FDR = 0.05. Next, DMRcate::dmrcate was called with default options except min.cpgs=1 and a bandwidth scaling factor of two. The differentially methylated regions were then output with the extraRanges functions.

Differential protein expression

We used the R package DEP (Differential Enrichment and Analysis of Proteomics Data) v1.20.0 for differential protein expression analysis⁸⁷. Variance stabilization normalization was done with the function vsn. Two samples that had greater than 50% missing data were removed. Technical replicates were averaged. Filtered for proteins that are identified in 2 out of 3 replicates of at least one condition when testing for differential abundance. Differential testing was performed using the test_diff function from the DEP R package, which applies protein-wise linear models and empirical Bayes statistics via limma, with *p* values derived from moderated *t* tests. FDR correction was done with BH method. In total, we assayed 5957 unique proteins with 2376 fragments having values for all samples.

Statistical analyses and data visualization

The Spearman's rank correlation test was used for all reported rho and *P* values for pairwise correlations. Spearman's test was implemented with the cor.test() function from the R package stats v4.2.2. Heatmaps were generated using the R package ComplexHeatmap v2.20.2⁸⁸ with annotations as shown. Pairwise group comparisons were done using the pairwise_wilcox_test function from the rstatix v0.7.2 package. ANOVA tests were run with the rstatix v0.7.2 anova_test function. Lollipop plots of mutations in *BRCA1* and *BRCA2* were then made using ProteinPaint⁸⁹. *BRCA1* domains were derived from the ProteinPaint Web interface and *BRCA2* domains were obtained from UniProt⁹⁰. All other plots were made with ggplot2 v3.4.1+. Gene ontology term enrichment was done with clusterProfiler v4.12.0 using the enrichGO function with *p* values from Fisher exact test corrected for multiple testing using the BH method.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The RNAseq data generated in this study have been deposited in the Gene Expression Omnibus under accession code GSE274628. The whole-genome bisulfite sequencing data generated in this study have been deposited in Sequence Read Archive under accession code PRJNA1146365. Native mass spectrometry data files have been deposited in the MassIVE database under accession code MSV000094339. The protein abundance data averaged across technical replicates is available as Supplementary Data 8. Benign endometrium expression counts were from GSE132711⁴⁵ (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132711>).

References

- Perets, R. & Drapkin, R. It's totally tubula. Riding the new wave of ovarian cancer research. *Cancer Res.* **76**, 10–17 (2016).
- Bergsten, T. M., Burdette, J. E. & Dean, M. Fallopian tube initiation of high grade serous ovarian cancer and ovarian metastasis: Mechanisms and therapeutic implications. *Cancer Lett.* **476**, 152–160 (2020).
- Shih, I. M., Wang, Y. & Wang, T. L. The origin of ovarian cancer species and precancerous landscape. *Am. J. Pathol.* **191**, 26–39 (2021).
- Medeiros, F. et al. The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome. *Am. J. Surg. Pathol.* **30**, 230–236 (2006).
- Kindelberger, D. W. et al. Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: evidence for a causal relationship. *Am. J. Surg. Pathol.* **31**, 161–169 (2007).
- Callahan, M. J. et al. Primary fallopian tube malignancies in BRCA-positive women undergoing surgery for ovarian cancer risk reduction. *J. Clin. Oncol.* **25**, 3985–3990 (2007).
- McDaniel, A. S. et al. Next-generation sequencing of tubal intraepithelial carcinomas. *JAMA Oncol.* **1**, 1128–1132 (2015).
- Eckert, M. A. et al. Genomics of ovarian cancer progression reveals diverse metastatic trajectories including intraepithelial metastasis to the Fallopian Tube. *Cancer Discov.* **6**, 1342–1351 (2016).
- Labidi-Galy, S. I. et al. High grade serous ovarian carcinomas originate in the fallopian tube. *Nat. Commun.* **8**, 1093 (2017).
- Soong, T. R. et al. Evidence for lineage continuity between early serous proliferations (ESPs) in the Fallopian tube and disseminated high-grade serous carcinomas. *J. Pathol.* **246**, 344–351 (2018).
- Wu, R. C. et al. Genomic landscape and evolutionary trajectories of ovarian cancer precursor lesions. *J. Pathol.* **248**, 41–50 (2019).
- Norquist, B. M. et al. Inherited mutations in women with ovarian carcinoma. *JAMA Oncol.* **2**, 482–490 (2016).
- Bell, D. et al. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
- Friebel, T. M., Domchek, S. M. & Rebbeck, T. R. Modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: a systematic review and meta-analysis. *J. Natl. Cancer Inst.* **106**, dju091 (2014).
- Eleje, G. U. et al. Risk-reducing bilateral salpingo-oophorectomy in women with BRCA1 or BRCA2 mutations. *Cochrane Database Syst. Rev.* **8**, CD012464 (2018).
- Falconer, H., Yin, L., Grönberg, H. & Altman, D. Ovarian cancer risk after salpingectomy: a nationwide population-based study. *J. Natl. Cancer Inst.* **107**, dju410 (2015).
- Hanley, G. E. et al. Outcomes from opportunistic salpingectomy for ovarian cancer prevention. *JAMA Netw. Open* **5**, e2147343 (2022).
- Lee, Y. et al. A candidate precursor to serous carcinoma that originates in the distal fallopian tube. *J. Pathol.* **211**, 26–35 (2007).
- Samimi, G., Trabert, B., Geczik, A. M., Duggan, M. A. & Sherman, M. E. Population frequency of serous tubal intraepithelial carcinoma (STIC) in clinical practice using SEE-Fim protocol. *JNCI Cancer Spectr.* **2**, pky061 (2018).
- Iniesta, M. D., Chien, J., Wicha, M. & Merajver, S. D. One-hit effects and cancer. *Cancer Prev. Res.* **3**, 12–15 (2010).
- Konishi, H. et al. Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells. *Proc. Natl. Acad. Sci. USA* **108**, 17773–17778 (2011).
- Pathania, S. et al. BRCA1 haploinsufficiency for replication stress suppression in primary cells. *Nat. Commun.* **5**, 1–15 (2014).
- Tone, A. A. et al. Gene expression profiles of luteal phase fallopian tube epithelium from BRCA mutation carriers resemble high-grade serous carcinoma. *Clin. Cancer Res.* **14**, 4067–4078 (2008).
- Press, J. Z. et al. Identification of a preneoplastic gene expression profile in tubal epithelium of BRCA1 mutation carriers. *Neoplasia* **12**, 993–1002 (2010).
- Yu, X. et al. Dissection of transcriptome dysregulation and immune characterization in women with germline BRCA1 mutation at single-cell resolution. *BMC Med.* **20**, 283 (2022).
- Nag, J. K. et al. Involvement of Protease-Activated Receptor2 Pleckstrin Homology Binding Domain in Ovarian Cancer:

- Expression in Fallopian Tubes and Drug Design. *Biomedicines* **2024** **12**, 246 (2024).
27. George, S. H. L. et al. Identification of abrogated pathways in fallopian tube epithelium from BRCA1 mutation carriers. *J. Pathol.* **225**, 106–117 (2011).
 28. Bartlett, T. E. et al. Epigenetic reprogramming of fallopian tube fimbriae in BRCA mutation carriers defines early ovarian cancer evolution. *Nat. Commun.* **7**, 1–10 (2016).
 29. Anjum, S. et al. A BRCA1-mutation associated DNA methylation signature in blood cells predicts sporadic breast cancer incidence and survival. *Genome Med.* **6**, 47 (2014).
 30. Bahar-Shany, K. et al. Proteomic signature for detection of high-grade ovarian cancer in germline BRCA mutation carriers. *Int. J. Cancer* **152**, 781–793 (2023).
 31. Shalabi, S. F. et al. Evidence for accelerated aging in mammary epithelia of women carrying germline BRCA1 or BRCA2 mutations. *Nat. Aging* **1**, 838–849 (2021).
 32. Reed, A. D. et al. A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast. *Nat. Genet.* 1–11 <https://doi.org/10.1038/s41588-024-01688-9> (2024).
 33. Javed, N. et al. Detecting sample swaps in diverse NGS data types using linkage disequilibrium. *Nat. Commun.* **11**, 1–8 (2020).
 34. George, J. W. et al. Integrated epigenome, exome, and transcriptome analyses reveal molecular subtypes and homeotic transformation in uterine fibroids. *Cell Rep.* **29**, 4069–4085.e6 (2019).
 35. Cherniack, A. D. et al. Integrated molecular characterization of uterine carcinosarcoma. *Cancer Cell* **31**, 411–423 (2017).
 36. Knijnenburg, T. A. et al. Multiscale representation of genomic signals. *Nat. Methods* **11**, 689–694 (2014).
 37. Zhou, W. et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **1** <https://doi.org/10.1038/s41588-018-0073-4> (2018).
 38. Pisanic, T. R. et al. Methyloomic analysis of ovarian cancers identifies tumor-specific alterations readily detectable in early precursor lesions. *Clin. Cancer Res.* **24**, 6536–6547 (2018).
 39. Pisanic, T. R. et al. Methyloomic landscapes of ovarian cancer precursor lesions. *Clin. Cancer Res.* **26**, 6310–6320 (2020).
 40. Lengyel, E. et al. A molecular atlas of the human postmenopausal fallopian tube and ovary from single-cell RNA and ATAC sequencing. *Cell Rep.* **41**, 111838 (2022).
 41. Ulrich, N. D. et al. Cellular heterogeneity of human fallopian tubes in normal and hydrosalpinx disease states identified using scRNA-seq. *Dev. Cell* **57**, 914–929.e7 (2022).
 42. Weigert, M. et al. A cell atlas of the human fallopian tube throughout the menstrual cycle and menopause. *Nat. Commun.* **16**, 1–15 (2025).
 43. Dinh, H. Q. et al. Single-cell transcriptomics identifies gene expression networks driving differentiation and tumorigenesis in the human fallopian tube. *Cell Rep.* **35**, 108978 (2021).
 44. Johnson, B. K. et al. Single-cell Total RNA Miniaturized sequencing (STORM-seq) reveals differentiation trajectories of primary human fallopian tube epithelium. *bioRxiv* 2022.03.14.484332. <https://doi.org/10.1101/2022.03.14.484332> (2022).
 45. Chi, R. P. A. et al. Human endometrial transcriptome and progesterone receptor cistrome reveal important pathways and epithelial regulators. *J. Clin. Endocrinol. Metab.* **105**, e1419–e1439 (2020).
 46. Beddows, I. et al. Cell state of origin impacts development of distinct endometriosis-related ovarian carcinoma histotypes. *Cancer Res.* **184**, 26–38 (2024).
 47. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, 1–20 (2013).
 48. Endicott, J. L., Nolte, P. A., Shen, H. & Laird, P. W. Cell division drives DNA methylation loss in late-replicating domains in primary human cells. *Nat. Commun.* **13**, 1–12 (2022).
 49. Issa, J. P. J. et al. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat. Genet.* **7**, 536–540 (1994).
 50. Kwabi-Addo, B. et al. Age-related DNA methylation changes in normal human prostate tissues. *Clin. Cancer Res.* **13**, 3796–3802 (2007).
 51. Lung, D. K., Reese, R. M. & Alarid, E. T. Intrinsic and extrinsic factors governing the transcriptional regulation of ESR1. *Horm. Cancer* **11**, 129–147 (2020).
 52. Maxwell, K. N. et al. BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. *Nat. Commun.* **8**, 1–11 (2017).
 53. Shah, J. B. et al. Analysis of matched primary and recurrent BRCA1/2 mutation-associated tumors identifies recurrence-specific drivers. *Nat. Commun.* **13**, 1–19 (2022).
 54. Norquist, B. M. et al. The molecular pathogenesis of hereditary ovarian carcinoma. *Cancer* **116**, 5261–5271 (2010).
 55. Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **12**, 68–78 (2011).
 56. Galhenage, P. et al. Replication stress and defective checkpoints make fallopian tube epithelial cells putative drivers of high-grade serous ovarian cancer. *Cell Rep.* **42**, 113144 (2023).
 57. Ishikawa, H. et al. Role of inflammation and immune response in the pathogenesis of uterine fibroids: Including their negative impact on reproductive outcomes. *J. Reprod. Immunol.* **165**, 104317 (2024).
 58. Duan, Y. et al. Correlation between inflammatory marker and lipid metabolism in patients with uterine leiomyomas. *Front. Med.* **10**, 1124697 (2023).
 59. Abramiuk, M. et al. The role of the immune system in the development of endometriosis. *Cells* **11**, 2028 (2022).
 60. Brand, J. et al. Fallopian tube single cell analysis reveals myeloid cell alterations in high-grade serous ovarian cancer. *iScience* **27**, 108990 (2024).
 61. Wada-Hiraike, O. et al. Role of estrogen receptor β in uterine stroma and epithelium: Insights from estrogen receptor β -/- mice. *Proc. Natl. Acad. Sci. USA* **103**, 18350–18355 (2006).
 62. Perets, R. et al. Transformation of the Fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca; Tp53; Pten models. *Cancer Cell* **24**, 751 (2013).
 63. Wu, R. et al. Impact of oviductal versus ovarian epithelial cell of origin on ovarian endometrioid carcinoma phenotype in the mouse. *J. Pathol.* **240**, 341–351 (2016).
 64. Hee, L., Kettner, L. O. & Vejtorp, M. Continuous use of oral contraceptives: an overview of effects and side-effects. *Acta Obstet. Gynecol. Scand.* **92**, 125–136 (2013).
 65. Sheppard, B. L. Endometrial morphological changes in IUD users: a review. *Contraception* **36**, 1–10 (1987).
 66. Stener-Victorin, E. et al. Polycystic ovary syndrome. *Nat. Rev. Dis. Prim.* **10**, 1–23 (2024).
 67. Ficarro, S. B., Max Alexander, W., Tavares, I. & Marto, J. A. Open source fraction collector/MALDI spotter for proteomics. *HardwareX* **11**, e00305 (2022).
 68. Skowronek, P. et al. Rapid and in-depth coverage of the (phospho-) proteome with deep libraries and optimal window design for dia-PASEF. *Mol. Cell. Proteom.* **21**, 100279 (2022).
 69. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520 (2017).
 70. Yang, K. L. et al. MSBooster: improving peptide identification rates using deep learning-based features. *Nat. Commun.* **14**, 1–14 (2023).
 71. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4**, 923–925 (2007).

72. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).
73. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2019).
74. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
75. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma.* **12**, 1–16 (2011).
76. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
77. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. <https://arxiv.org/abs/1303.3997v2> (2013).
78. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
79. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882 (2012).
80. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform* **2**, lqaa078 (2020).
81. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).
82. krassowski/complex-upset: v1.3.5. <https://doi.org/10.5281/ZENODO.7314197>.
83. Köster, J. et al. Sustainable data analysis with Snakemake. *F1000Res* **10**, 33 (2021).
84. Zhou, W. et al. BISCUIT: an efficient, standards-compliant tool suite for simultaneous genetic and epigenetic inference in bulk and single-cell studies. *Nucleic Acids Res.* **52**, e32 (2024).
85. Morrison, J., Zhou, W., Johnson, B. K. & Shen, H. Dupsifter: a light-weight duplicate marking tool for whole genome bisulfite sequencing. *Bioinformatics* **39**, btad729 (2023).
86. Peters, T. J. et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**, 1–16 (2015).
87. Zhang, X. et al. Proteome-wide identification of ubiquitin interactions using UbiA-MS. *Nat. Protoc.* **13**, 530–550 (2018).
88. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
89. Zhou, X. et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat. Genet.* **48**, 4–6 (2016).
90. Bateman, A. et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).

Acknowledgements

This work was made possible by the generous support of the Canary Foundation which aims at improving early ovarian cancer detection (<https://www.canaryfoundation.org>), and the Gray Foundation, which aims to accelerate research, raise awareness, and improve treatment to support individuals and families who have inherited BRCA mutations and related cancers (<http://grayfoundation.org>). We thank Mei Zheng for assistance with histology. This study was also in part supported by the National Institutes of Health P50 SPORE CA228991 (R.D.), the

Department of Defense W81XWH-22-1-0852 (R.D.), the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (R.D.), the Honorable Tina Brozman Foundation for Ovarian Cancer Research (R.D. and J.A.M.), the Claneil Foundation (R.D.), the Gray Foundation (R.D. and J.A.M.), the Carl H. Goldsmith Ovarian Cancer Translational Research Fund (R.D.), the Marjorie S. Stanek and Lowell H. Dubrow Ovarian Cancer Research Center Endowed Fund (RD), the Mike and Patti Hennessy Foundation (R.D.), the Monica K. Young Foundation (R.D.), and the Bassett Center for BRCA (R.D.). We acknowledge the Van Andel Institute Genomics Core Facility (RRID:SCR_022913) and the Van Andel Institute Pathology and Biorepository (RRID:SCR_022912) for their work on this project.

Author contributions

C.W.D., R.D., H.S. conceived the study. G.A., A.T., J.A.M., D.O., E.J., N.S., L.S., C.L.P., M.A., D.R. data collection and methodology. I.B., S.D., R.R., A.H., A.O., J.M. formal analysis. H.A. project administration. I.B., H.S. C.W.D., R.D., E.J., M.A., J.A.M., A.O., G.A. original writing and methods. I.B., H.S. C.W.D., R.D. manuscript review and editing.

Competing interests

J.A.M. is a founder, equity holder, and advisor to Entact Bio, serves on the SAB of 908 Devices, and receives or has received sponsored research funding from Vertex, AstraZeneca, Taiho, Springworks, TUO Therapeutics, and Takeda. R.D. serves as an advisor to Repare Therapeutics. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-58145-2>.

Correspondence and requests for materials should be addressed to Charles W. Drescher, Ronny Drapkin or Hui Shen.

Peer review information *Nature Communications* thanks Robert Hollis, Patricia Tonin, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Department of Epigenetics, Van Andel Research Institute, Grand Rapids, MI, USA. ²Penn Ovarian Cancer Research Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ³Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁴Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA. ⁵Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA. ⁶Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. ⁷Blais Proteomics Center, Center for Emergent Drug Targets, Dana-Farber Cancer Institute, Boston, MA, USA. ⁸Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. ⁹Genomics Core, Van Andel Research Institute, Grand Rapids, MI, USA. ¹⁰Pathology and Biorepository Core, Van Andel Research Institute, Grand Rapids, MI, USA. ¹¹Department of Epidemiology, University of Michigan School of Public Health and Rogel Cancer Center, Ann Arbor, MI, USA. ¹²Canary Foundation, Palo Alto, CA, USA. ¹³Paul G. Allen Research Center, Swedish Cancer Institute, Seattle, WA, USA. ¹⁴Translational Research Program, Fred Hutchinson Cancer Center, Seattle, WA, USA. ¹⁵Basser Center for BRCA, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA. ✉e-mail: cdresche@fredhutch.org; rdrapkin@pennmedicine.upenn.edu; hui.shen@vai.org