Supplementary Methods

Sample Collection (NHS/NHSII)

In 1989-1990, 32,826 NHS participants aged 43-69y contributed blood samples, as previously described[1]. Briefly, each woman arranged to have her blood drawn and shipped it overnight to our laboratory where we processed it and archived aliquots of white blood cell, red blood cell, and plasma in liquid nitrogen freezers (≤-130°C). In 2000-2002, 18,473 of these women aged 53-80y donated a second blood sample using a similar protocol. In the NHSII, 29,611 women aged 32-54y donated blood and urine samples in 1996-1999. Of these, 18,521 women who had not used oral contraceptives, been pregnant or breastfed in the previous six months provided samples timed within the menstrual cycle, targeting the early follicular (days 3 to 5 of the cycle) and mid-luteal (7 to 9 days prior to expected start of next cycle) phases. The remaining women donated a single untimed sample. Follicular plasma was separated and frozen by the participants and returned with the luteal sample; luteal and untimed samples were collected and shipped following the same protocol as NHS. Follow-up in the blood subcohorts is high (NHS 97% in 2010; NHSII 96% in 2011).

Case and Control Selection (NHS/NHSII)

Incident cases of breast cancer were identified after blood collection among women who had no reported cancer (other than non-melanoma skin). Cases were diagnosed between 2000 and 2010 (NHS) or 1999-2011 (NHSII). Overall, 940 cases were included in the NHS, 592 of whom had donated blood samples in both the first (1989-1990) and second (2000-2002) blood collections. In the NHSII (1996-1999),

1,057 cases were included. Breast cancer cases were reported (invasive cases: NHS=748; NHSII=780) and confirmed by medical record reviews (NHS=914; NHSII=1015) or verbally by the nurse (NHS=26; NHSII=42). Given the high confirmation rate by medical record for breast cancer in this cohort (99%), all cases are included in this analysis.

In NHS, one control was matched per case by the following factors (at both collections for subjects with 2 samples): age (+/- 2y), menopausal status and postmenopausal hormone therapy (HT) use at blood collection and diagnosis (premenopausal, postmenopausal and not taking HT, postmenopausal and taking HT, and unknown), and month (+/- 1mo), time of day (+/- 2h), and fasting status at blood collection (<8 h after a meal or unknown; >8h). In NHSII, one control was matched per case on the same factors as NHS, with the addition of race/ethnicity (African American, Asian, Hispanic, Caucasian, other) and luteal day (+/- 1d; timed samples only).

Covariate Information (NHS/NHSII)

Data on breast cancer risk factors, including anthropometric measures, reproductive history, and lifestyle factors, were collected from questionnaires administered biennially and at the time of blood collections. Case characteristics, including invasive vs. *in situ*, histologic grade, estrogen and progesterone receptor (ER, PR), and human epidermal growth factor receptor 2 (HER2) status, were extracted from pathology reports. As previously described[2], immunohistochemical results for ER, PR, and HER2, read manually by a study pathologist, were included for cases with available tumor tissue included in TMAs.

The Women's Health Study (WHS)

The WHS is an ongoing longitudinal cohort of 39,876 female US health professionals aged ≥45 years without history of prior cancer (except non-melanoma skin cancer) or cardiovascular disease at enrollment. The WHS began as a randomized placebo-controlled trial of low-dose aspirin, β-carotene, and vitamin E for the primary prevention of cardiovascular disease and cancer and ran from 1993 to 2004 (NCT00000479),[3,4] with no overall effect of the randomized interventions on cancer incidence.[5-7] Participants reported several characteristics related to health, lifestyle, reproductive history, medical diagnoses, family history of cancer, current medication use, and other factors via questionnaires on an annual basis. Usual diet was assessed using a 131-item semi-quantitative food frequency questionnaire (FFQ),[8,9] from which we derived the Alternative Health Eating Index 2010 (aHEI-2010) dietary quality score for each individual, as previously described.[10] Physical activity was captured on the baseline questionnaire as the average time per week spent engaged in recreational activity domains (e.g., walking, running, bicycling), flights of stairs climbed daily, and usual walking pace.[11]

Medical record review was completed for 95% of self-reported cancer cases and confirmation among those with available records was 82%.[5] In the present analysis we include confirmed invasive breast cancer cases through December 31, 2018. Written informed consent was obtained from all participants and the study protocol was approved by the Institutional Review Board of the Brigham and Women's Hospital (Boston, MA) and all research was performed in accordance with relevant guidelines and regulations.

Associations among WHS women were evaluated restricting to fasting participants, by menopausal status at blood collection and among postmenopausal women by time between blood collection and diagnosis.

BCAAs Measurement

In the NHS/NHSII, BCAAs were assayed through a metabolomic profiling platform at the Broad Institute using a liquid chromatography tandem mass spectrometry (LC-MS) method designed to measure polar metabolites such as amino acids, amino acids derivatives, dipeptides, and other cationic metabolites[12-14]. Pooled plasma reference samples were included every 20 samples and results were standardized using the ratio of the value of the sample to the value of the nearest pooled reference multiplied by the median of all reference values for the metabolite. Samples from the two cohorts were run together, with matched case-control pairs (as sets) distributed randomly within the batch, and the order of the case and controls within each pair randomly assigned. Therefore, the case and its control were always directly adjacent to each other in the analytic run, thereby limiting variability in platform performance across matched case-control pairs. Hydrophilic interaction liquid chromatography (HILIC) analyses of water-soluble metabolites in the positive ionization mode were conducted using an LC-MS system comprised of a Shimadzu Nexera X2 U-HPLC (Shimadzu Corp.; Marlborough, MA) coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific; Waltham, MA). Metabolites were extracted from plasma (10 μL) using 90 μL of acetonitrile/methanol/formic acid (74.9:24.9:0.2 v/v/v) containing stable isotope-labeled internal standards (valine-d8, Sigma-Aldrich; St. Louis, MO; and phenylalanine-d8, Cambridge Isotope Laboratories; Andover, MA). The samples were

centrifuged (10 min, 9,000 x g, 4°C), and the supernatants were injected directly onto a 150 x 2 mm, 3 μm Atlantis HILIC column (Waters; Milford, MA). The column was eluted isocratically at a flow rate of 250 μL/min with 5% mobile phase A (10 mM ammonium formate and 0.1% formic acid in water) for 0.5 minute followed by a linear gradient to 40% mobile phase B (acetonitrile with 0.1% formic acid) over 10 minutes. MS analyses were carried out using electrospray ionization in the positive ion mode using full scan analysis over 70-800 m/z at 70,000 resolution and 3 Hz data acquisition rate. Other MS settings were: sheath gas 40, sweep gas 2, spray voltage 3.5 kV, capillary temperature 350°C, S-lens RF 40, heater temperature 300°C, microscans 1, automatic gain control target 1e6, and maximum ion time 250 ms. Metabolite identities were confirmed using authentic reference standards or reference samples. NHS samples were run together, followed by NHSII samples. The coefficients of variation (CV) among blinded quality control samples (N=638) were <15% (isoleucine=7.3-11.4%, leucine=7.0-12.1%, valine=6.2-9.8%). BCAAs were not affected by delayed processing of blood samples and showed good within person stability over 1-2 years (intra-class correlation≥0.55)[15].

In the WHS, baseline blood samples from a subset of 28,345 women were shipped on ice via overnight courier to the central laboratory where they were processed and stored at -170 degrees C in vapor liquid nitrogen. Aliquots of the EDTA plasma samples were shipped on dry ice blinded to outcome status to LipoScience, Inc, now LabCorp® (Raleigh, NC). Isoleucine, leucine, and valine were measured by proton nuclear magnetic resonance (^{1}H NMR) spectroscopy using a 400MHz NMR platform, as described for the NMR LipoProfile® IV test.[16,17] The derived BCAA signal amplitudes

were converted to μmol/L, with intra- and inter-assay coefficients of variation of isoleucine 5.9-6.1%, leucine 4.5-4.9%, and valine 1.5-2.1%.

Gene Expression

Archived formalin-fixed paraffin-embedded breast cancer tissue blocks were collected and assembled into tissue microarrays (TMAs)[18]. RNA was extracted from multiple cores of 1 or 1.5 mm taken from formalin fixed paraffin-embedded (FFPE) tumor (n=1-3 cores) and normal-adjacent (n=3-5 cores) tissues using the Qiagen AllPrep RNA isolation kit[19]. Normal-adjacent tissues were obtained >1 cm away from the edge of the tumor. A detailed protocol has been published previously[20-22] (microarray accession number: GSE115577). In brief, we profiled transcriptome-wide gene expression using Affymetrix Glue Grant Human Transcriptome Array 3.0 (hGlue 3.0) and Human Transcriptome Array 2.0 (HTA 2.0) microarray chips (Affymetrix, Santa Clara, CA, USA). We used robust multi-array average to perform normalization (RMA; Affymetrix Power Tools (ATP)), log-2 transformed the data, and conducted sample quality control with Affymetrix Power Tools probeset summarization based metrics[20,21]. A total of 882 tumor tissues from invasive breast cancer cases passed quality control. Among these, 120 NHSII (68% premenopausal at blood collection) samples and 101 NHS samples (all postmenopausal) had gene expression and BCAA levels measured. For genes that were mapped by multiple probes, we selected the most variable probe to represent the gene. Our current analyses included 17,791 (70%) genes that were profiled in both platforms. Technical variabilities by batch were controlled using *ComBat*, an empirical Bayes method used to control for known batch effects[23]. Genes with low expression (<25th percentile) were removed from the analyses.

Estradiol and C-peptide

Estradiol assays were performed at Quest Diagnostics (San Juan Capistrano, CA) as previously described[24] by radioimmunoassay following extraction and celite column chromatography (NHSII, luteal measures N=278 cases, 280 controls; follicular measures N=268 cases, 264 controls) or at the Mayo Clinic (Rochester, MN) by LC-MS/MS as previously described[25] (NHS distant measures, N=145 cases, 143 controls; NHS proximate measures , N=119 cases, 115 controls). C-peptide was assessed using ELISA (Diagnostic Systems Laboratory, Webster, TX) in the laboratory of Dr. Michael Pollak (McGill University, Montreal, Quebec, Canada) (NHSII, N=290 cases, 289 controls; NHS distant collection measures N=202 cases, 205 controls; NHS proximate collection measures N=124 cases, 120 controls).

Gene Set Testing Procedure

Using a competitive gene set testing procedure, Correlation Adjusted Mean Rank (CAMERA), on the gene expression data, we explored functional enrichment of biological pathways associated with BCAA[26]. Additional details can be found in the supplememntary materials. We chose the 50 “hallmark” gene sets from the Molecular Signature Database (MSigDB)[27]. We controlled for age and year of diagnosis, BMI at blood draw, alcohol consumption, menopausal status, menopausal hormone use, ER status, race, as well as matching factors. We chose an inter-gene correlation of 0.01.

References

1. Tworoger, S.S., Sluss, P. & Hankinson, S.E. Association between plasma prolactin concentrations and risk of breast cancer among predominately premenopausal women. *Cancer research* **66**, 2476-2482 (2006).
2. Sisti, J.S. (2015).

3. Buring, J.E., Hennekens, C.H. & Group, f.t.W.s.H.S.R. The Women's Health Study: Summary of the study design. *J Myocardial Ischemia* **4**, 27-29 (1992).
4. Ridker, P.M.*, et al.* A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *The New England journal of medicine* **352**, 1293-1304 (2005).
5. Lee, I.M., Cook, N.R., Manson, J.E., Buring, J.E. & Hennekens, C.H. Beta-carotene supplementation and incidence of cancer and cardiovascular disease: the Women's Health Study. *Journal of the National Cancer Institute* **91**, 2102-2106 (1999).
6. Cook, N.R.*, et al.* Low-dose aspirin in the primary prevention of cancer: the Women's Health Study: a randomized controlled trial. *Jama* **294**, 47-55 (2005).
7. Lee, I.M.*, et al.* Vitamin E in the primary prevention of cardiovascular disease and cancer: the Women's Health Study: a randomized controlled trial. *Jama* **294**, 56-65 (2005).
8. Willett, W.C.*, et al.* Reproducibility and validity of a semiquantitative food frequency questionnaire. *American journal of epidemiology* **122**, 51-65 (1985).
9. Freedman, L.S.*, et al.* Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American journal of epidemiology* **180**, 172-188 (2014).
10. Chiuve, S.E.*, et al.* Alternative dietary indices both strongly predict risk of chronic disease. *The Journal of nutrition* **142**, 1009-1018 (2012).
11. Mora, S., Lee, I.M., Buring, J.E. & Ridker, P.M. Association of physical activity and body mass index with novel and traditional cardiovascular biomarkers in women. *Jama* **295**, 1412-1419 (2006).
12. Mascanfroni, I.D.*, et al.* Metabolic control of type 1 regulatory T cell differentiation by AHR and HIF1-α. *Nature medicine* **21**, 638 (2015).
13. O'sullivan, J.F.*, et al.* Dimethylguanidino valeric acid is a marker of liver fat and predicts diabetes. *The Journal of clinical investigation* **127**, 4394-4402 (2017).
14. Paynter, N.P.*, et al.* Metabolic predictors of incident coronary heart disease in women. *Circulation* **137**, 841-853 (2018).
15. Townsend, M.K.*, et al.* Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clinical chemistry* **59**, 1657-1667 (2013).
16. Wolak-Dinsmore, J.*, et al.* A novel NMR-based assay to measure circulating concentrations of branched-chain amino acids: Elevation in subjects with type 2 diabetes mellitus and association with carotid intima media thickness. *Clinical biochemistry* (2018).
17. Jeyarajah, E.J., Cromwell, W.C. & Otvos, J.D. Lipoprotein particle analysis by nuclear magnetic resonance spectroscopy. *Clinics in laboratory medicine* **26**, 847-870 (2006).
18. Tamimi, R.M.*, et al.* Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. *Breast Cancer Research* **10**, R67 (2008).
19. Tamimi, R.M.*, et al.* Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. *Breast Cancer Res* **10**, R67 (2008).
20. Wang, J.*, et al.* Alcohol consumption and breast tumor gene expression. *Breast Cancer Res* **19**, 108 (2017).
21. Heng, Y.J.*, et al.* Molecular mechanisms linking high body mass index to breast cancer etiology in post-menopausal breast tumor and tumor-adjacent tissues. *Breast Cancer Res Treat* (2018).
22. Kensler, K.H.*, et al.* PAM50 Molecular Intrinsic Subtypes in the Nurses' Health Study Cohorts. *Cancer Epidemiol Biomarkers Prev* (2018).
23. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).

24. Eliassen, A.H.*, et al.* Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. *Journal of the National Cancer Institute* **98**, 1406-1415 (2006).
25. Zhang, X., Tworoger, S.S., Eliassen, A.H. & Hankinson, S.E. Postmenopausal plasma sex hormone levels and breast cancer risk over 20 years of follow-up. *Breast cancer research and treatment* **137**, 883-892 (2013).
26. Wu, D. & Smyth, G.K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res* **40**, e133 (2012).
27. Liberzon, A.*, et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).

Supplementary Table 1: Multivariate ORs (95% CIs) of breast cancer according to quartile of plasma branched chain amino acids by BMI at blood collection in the NHSII and NHS by blood collection

		Premenopausal[1] women in NHSII					
		Q1	Q2	Q3	Q4	p-trend	p-interaction
Fasting samples		BMI<25: 406 cases & 373 controls; BMI ≥25: 309 cases & 342 controls					
Isoleucine	BMI <25	1.00 (ref)	1.07 (0.74-1.55)	0.74 (0.49-1.11)	0.96 (0.59-1.55)	0.45	0.33
	BMI ≥25	1.00 (ref)	0.77 (0.46-1.28)	0.70 (0.42-1.16)	0.61 (0.37-1.01)	0.05	
Leucine	BMI <25	1.00 (ref)	1.08 (0.75-1.56)	1.01 (0.68-1.49)	0.90 (0.55-1.46)	0.77	0.11
	BMI ≥25	1.00 (ref)	0.56 (0.33-0.95)	0.77 (0.47-1.25)	0.49 (0.29-0.80)	0.02	
Valine	BMI <25	1.00 (ref)	1.08 (0.76-1.56)	0.98 (0.65-1.49)	0.98 (0.60-1.60)	0.95	0.08
	BMI ≥25	1.00 (ref)	0.41 (0.23-0.71)	0.48 (0.28-0.80)	0.44 (0.26-0.74)	0.03	
Total BCAAs	BMI <25	1.00 (ref)	1.18 (0.82-1.69)	0.88 (0.59-1.32)	1.08 (0.67-1.76)	0.99	0.13
	BMI ≥25	1.00 (ref)	0.64 (0.38-1.08)	0.67 (0.40-1.10)	0.54 (0.32-0.90)	0.03	

		Postmenopausal[2] women in NHS with distant sample collection					
		Q1	Q2	Q3	Q4	p-trend	p-interaction
Fasting samples		BMI <25: 312 cases & 377 controls; BMI ≥25: 311 cases & 246 controls					
Isoleucine	BMI <25	1.00 (ref)	0.91 (0.60-1.37)	0.91 (0.58-1.42)	1.18 (0.74-1.88)	0.57	0.38
	BMI ≥25	1.00 (ref)	0.61 (0.33-1.11)	0.72 (0.41-1.25)	0.73 (0.42-1.27)	0.54	
Leucine	BMI <25	1.00 (ref)	0.98 (0.64-1.48)	0.83 (0.53-1.30)	1.31 (0.82-2.10)	0.49	0.42
	BMI ≥25	1.00 (ref)	0.72 (0.40-1.30)	0.78 (0.43-1.39)	0.82 (0.48-1.42)	0.72	
Valine	BMI <25	1.00 (ref)	0.84 (0.56-1.27)	0.82 (0.53-1.27)	1.45 (0.89-2.35)	0.30	0.28
	BMI ≥25	1.00 (ref)	1.15 (0.62-2.11)	0.80 (0.44-1.47)	0.82 (0.46-1.45)	0.25	
Total BCAAs	BMI <25	1.00 (ref)	0.94 (0.63-1.42)	0.88 (0.56-1.39)	1.28 (0.80-2.06)	0.45	0.34
	BMI ≥25	1.00 (ref)	0.91 (0.49-1.68)	0.83 (0.46-1.47)	0.84 (0.48-1.47)	0.54	

		Postmenopausal[2] women in NHS with proximate sample collection					
		Q1	Q2	Q3	Q4	p-trend	p-interaction
Fasting samples		BMI <25: 200 cases & 237 controls; BMI ≥25: 313 cases & 276 controls					
Isoleucine	BMI <25	1.00 (ref)	1.80 (1.07-3.03)	1.79 (1.01-3.17)	2.26 (1.18-4.31)	0.01	0.62
	BMI ≥25	1.00 (ref)	1.07 (0.59-1.94)	1.24 (0.70-2.19)	1.38 (0.78-2.45)	0.20	
Leucine	BMI <25	1.00 (ref)	0.94 (0.56-1.56)	1.69 (0.96-2.96)	1.46 (0.76-2.79)	0.09	0.45
	BMI ≥25	1.00 (ref)	1.52 (0.86-2.71)	1.36 (0.80-2.33)	1.33 (0.78-2.28)	0.55	
Valine	BMI <25	1.00 (ref)	1.52 (0.91-2.53)	1.13 (0.63-2.01)	2.19 (1.15-4.17)	0.05	0.58
	BMI ≥25	1.00 (ref)	1.24 (0.70-2.22)	1.07 (0.61-1.88)	1.48 (0.86-2.55)	0.17	
Total BCAAs	BMI <25	1.00 (ref)	1.08 (0.65-1.81)	0.97 (0.55-1.71)	2.07 (1.09-3.93)	0.09	0.48
	BMI ≥25	1.00 (ref)	1.42 (0.79-2.55)	1.55 (0.88-2.74)	1.50 (0.86-2.62)	0.21	

[1]predominantly premenopausal at blood collection(see Table 1 and Figure 1 for details)

[2]predominantly postmenopausal women at blood collection and diagnosis (see Table1 and Figure 1 for details)

Supplementary Table 2: Multivariate ORs (95% CIs) of breast cancer per 1SD increase in plasma branched chain amino acids by BMI at blood collection in the NHSII and NHS by blood collection

Premenopausal[1] women in NHSII			
Fasting samples		ER+: 395 cases; ER-: 87 cases	p-het
Isoleucine	ER+	0.86 (0.71-1.04)	0.60
	ER-	0.82 (0.58-1.15)	
Leucine	ER+	0.83 (0.68-1.01)	0.44
	ER-	0.95 (0.67-1.35)	
Valine	ER+	0.85 (0.71-1.02)	0.91
	ER-	0.84 (0.60-1.17)	
Total BCAAs	ER+	0.86 (0.72-1.04)	0.60
	ER-	0.89 (0.63-1.24)	

Postmenopausal[2] women in NHS with distant sample collection			
Fasting samples		ER+: 381 cases; ER-: 58 cases	p-het
Isoleucine	ER+	1.04 (0.87-1.23)	0.54
	ER-	1.02 (0.64-1.61)	
Leucine	ER+	1.06 (0.89-1.25)	0.13
	ER-	1.05 (0.68-1.64)	
Valine	ER+	0.92 (0.78-1.08)	0.98
	ER-	1.14 (0.76-1.72)	
Total BCAAs	ER+	1.00 (0.84-1.19)	0.42
	ER-	1.05 (0.68-1.62)	

Postmenopausal[2] women in NHS with proximate sample collection			
Fasting samples		ER+: 323 cases; ER-: 54 cases	p-het
Isoleucine	ER+	**1.26 (1.02-1.57)**	0.95
	ER-	1.10 (0.73-1.68)	
Leucine	ER+	1.15 (0.94-1.42)	0.67
	ER-	1.16 (0.78-1.76)	
Valine	ER+	1.22 (0.99-1.50)	0.81
	ER-	1.24 (0.83-1.86)	
Total BCAAs	ER+	1.19 (0.96-1.47)	0.93
	ER-	1.14 (0.75-1.73)	

[1]predominantly premenopausal at blood collection(see Table 1 and Figure 1 for details)

[2]predominantly postmenopausal women at blood collection and diagnosis (see Table1 and Figure 1 for details)

Supplementary Table 3. Gene expression pathway analysis.

NHS							
		BCAAs					
		Isoleucine		Leucine		Valine	
Pathway	N of genes	Direction	FDR	Direction	FDR	Direction	FDR
Hallmark estrogen response early	185	Up	5.93E-09	Up	5.93E-09	Up	4.58E-09
Hallmark estrogen response late	177	Up	1.84E-08	Up	9.21E-08	Up	2.33E-08

NHSII							
		BCAAs					
		Isoleucine		Leucine		Valine	
Pathway	N of genes	Direction	FDR	Direction	FDR	Direction	FDR
HALLMARK_DNA_REPAIR	137	Up	4.50E-03	Up	3.89E-03	Up	8.58E-04
HALLMARK_E2F_TARGETS	154	Up	1.81E-06	Up	7.12E-07	Up	1.17E-06
HALLMARK_G2M_CHECKPOINT	150	Up	1.81E-06	Up	4.52E-06	Up	6.39E-06
HALLMARK_INTERFERON_ALPHA_RESPONSE	76	Up	5.97E-21	Up	9.30E-22	Up	5.94E-22
HALLMARK_INTERFERON_GAMMA_RESPONSE	159	Up	1.35E-17	Up	9.93E-19	Up	2.05E-17
HALLMARK_MITOTIC_SPINDLE	172	Up	7.44E-05	Up	2.71E-04	Up	1.02E-04
HALLMARK_MTORC1_SIGNALING	174	Up	2.31E-03	Up	1.30E-02	Up	4.72E-02
HALLMARK_MYC_TARGETS_V1	193	Up	7.31E-06	Up	3.08E-05	Up	6.22E-04
HALLMARK_PI3K_AKT_MTOR_SIGNALING	92	Up	1.87E-03	Up	2.70E-03	Up	4.94E-03