Supplementary Information

Within-person ovarian hormone variation and structural connectome topology: a three-visit diffusion MRI study

Zhijing Chen[1], Yiling Qiu[1], Guiyin Chen[2], Qingwen Chen[3,*], Yan Tan[1,*]

[1] Department of Gynaecology and Obstetrics, Guangzhou Red Cross Hospital (Guangzhou Red Cross Hospital of Jinan University), No. 396 Tongfu Middle Road, Haizhu District, Guangzhou 510220, China

[2] Department of Emergency Medicine, The Fifth Teaching Hospital of Xinjiang Medical University, No. 118, Henan West Road, Xincheng District, Urumqi City, Xinjiang Uygur Autonomous Region, 83001, China

[3] Department of Neurosurgery, Liwan Central Hospital of Guangzhou, No. 35 Liwan Road, Liwan District, Guangzhou, Guangdong 510170, China.

*** Co-corresponding authors:**

Qingwen Chen. Email: chen_qingwen0518@163.com

Yan Tan. Email: fish01water@163.com

Other author email addresses: Zhijing Chen, 305194792@qq.com; Yiling Qiu, Qiuyl10@126.com; Guiyin Chen, 253816094@qq.com.

Funding: This study was supported by Science and Technology Program of Guangzhou (2025A03J3375).

Supplementary Methods

These sections describe the study design, endocrine measurements, imaging, network construction and statistical analyses in detail. Supplementary Figures S1–S5 and Supplementary Tables S1–S8 present quality-control, sensitivity and exploratory results. Supplementary references use S-prefixed numbers independent of the main bibliography. Supplementary Data 1 contains the session-level observations and numerical results underlying the figures and tables.

Study design, source cohort and analysis hierarchy

We conducted a secondary longitudinal analysis of the Hormone Health Study brain anatomy and structure release. The data follow the Brain Imaging Data Structure standard [S1] and are available as OpenNeuro ds005360, version 1.0.1, under CC0 [S2]. This release contains 22 of the 30 source-study participants who completed three MRI sessions and consented to public sharing. We used all available participants and sessions, without selecting a subgroup according to endocrine concentrations, graph values or statistical results. The observational unit was a session nested within a participant. The participant, rather than the individual scan, was the independent sampling unit for cluster resampling.

Source-study eligibility, assessed by self-report, required ages of 18–30 years, nulliparity and relatively regular menstrual cycles lasting 21–40 days [S2,S3]. Participants had not used hormonal or implant contraception during the preceding three months and had never undergone hormonal therapy. They reported no plans to start contraception or become pregnant during the following year. Exclusion criteria covered MRI contraindications and clinically significant reproductive, cardiovascular, neurological or psychiatric conditions. The source criteria allowed mild-to-moderate mood or attention-deficit disorders, so eligibility should not be interpreted as excluding every psychiatric diagnosis. Sex was recorded as female in the public data. These characteristics describe recruitment for the source study, not additional eligibility rules introduced in this reanalysis.

Each participant contributed one visit during menses, one near estimated ovulation and one during the mid-luteal phase. Scheduling used self-reported menses, at-home ovulation testing and an estimated midpoint between ovulation and the next predicted menses [S3]. We retained the source phase labels without retrospectively assigning phases from hormone concentrations. Acquisition visit order did not necessarily match the biological order of these phases, so the model included visit order separately. The source protocol received University of California, Santa Barbara Human Subjects Committee approval, protocol 40-22-0249. All participants provided written informed consent for the source study [S3], and the public subset comprises participants who consented to open sharing [S2]. The present analysis used only the publicly released, deidentified data under CC0 and involved no new participant contact, biological sampling or collection of directly identifying information. The secondary use of these lawfully obtained public data followed the public-data provisions of Article 32(1) of China's Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects (2023).

The primary question, two exposures and two outcomes were specified in the analysis plan before the corresponding hormone inference. Fixed-density graphs were defined during hormone-blind graph processing. Other extensions were specified after the primary results but before inspection of their own association results and remained subordinate to the primary four-test family. The term “prespecified” refers to this analysis sequence and does not imply public preregistration. Additional estimator checks were performed after the primary analysis; the primary model, cohort and outcomes were not changed in response to the findings.

Input completeness, image integrity and BIDS conformity were checked before processing.

Endocrine sampling and assays

Blood samples were obtained before MRI at each of the three visits, as described in the public dataset documentation [S2]. After 30 minutes of clotting at room temperature, samples were centrifuged at 2,100 rpm for 10 minutes. Serum aliquots were stored at −80 °C and processed by the Endocrine Technologies Core at the Oregon National Primate Research Center. E2 and P4 were measured using ultra-high-performance liquid chromatography with heated electrospray ionization and tandem triple-quadrupole mass spectrometry. The instrument was a Shimadzu Nexera-LCMS-8060, and the reported dynamic range for both steroids was 0.002–20 ng/mL.

Lower quantification limits were 0.002 ng/mL for E2 and 0.010 ng/mL for P4. Reported intra-assay variation was 2.1% and 12.3%, respectively. LH and FSH were measured using a Roche cobas e411 automated clinical immunoassay platform, with an assay range of 0.1–200 mIU/mL. We used the concentrations supplied in the release rather than re-estimating assay values. E2 was expressed in pg/mL, P4 in ng/mL, and both gonadotropins in mIU/mL in the analysis table. These units were retained when applying $\log(1 + x)$, whose numerical transformation depends on the units of x.

MRI acquisition and diffusion-input selection

The source acquisition used a Siemens 3 T Prisma scanner with a 64-channel head/neck coil [S2]. T1-weighted MPRAGE images had 0.9-mm isotropic voxels, repetition time 2,500 ms, echo time 2.22 ms and inversion time 851 ms. The field of view was 241 mm and flip angle 7°. Diffusion imaging comprised four spherical and four linear b-tensor q-space trajectory imaging runs. Audited linear inputs contained one b0 volume and five shells at b = 100, 500, 1,000, 1,500 and 2,000 s/mm². The corresponding shell counts were 6, 6, 10, 16 and 30 diffusion-weighted volumes, giving 69 volumes per session. Diffusion repetition time was 6,308 ms, echo time 80 ms, field of view 230 mm and flip angle 90°. Nominal voxel dimensions were 2 × 2 × 4 mm, with parallel-imaging factor 2.

Conventional fibre-direction estimation used only the four linear b-tensor runs and their acquired b0 volume. Spherical-encoding runs were not combined with these inputs because their geometry and field-of-view centres differed. Gradient tables, image dimensions, b values, volume counts and phase-encoding metadata were audited before processing. All selected runs used j− phase encoding and a total readout

time of 0.0395495 s. Reverse-phase-encoding images and field maps were absent from the released acquisition. This absence determined the susceptibility-correction strategy. It is treated as a limitation, not as evidence that distortion was negligible.

Longitudinal anatomy and diffusion preprocessing

Processing began with a three-participant, nine-session pilot, followed by the full 66-session cohort after the pilot quality gates. Thermal noise was reduced using Marchenko–Pastur principal-component analysis [S4]. Gibbs-ringing correction was not automatically applied to the partial-Fourier diffusion acquisition. Synb0-DisCo generated a T1-informed, undistorted b0-like target because an acquired reverse-phase-encoding reference was unavailable [S5]. The synthetic target and acquired b0 were used for TOPUP susceptibility estimation. FSL eddy subsequently corrected motion and eddy-current effects while incorporating the estimated susceptibility field [S6]. The final analysis used a linear eddy field model; the pilot had used a quadratic field model.

The eddy settings included a linear field model, initialization seed 1234, b-range tolerance 10 and replacement of outlier slices. Shelled-data handling, contrast-to-noise maps and residual outputs were enabled. Rotated gradient directions were retained for downstream modelling. N4 bias-field correction was applied after eddy [S7]. Software versions were FSL 6.0.7.22, Synb0-DISCO 3.1 and ANTs 2.6.2.dev1-gff61a3d. The correction settings and gradient rotation are reported because they form part of the measurement pipeline.

T1-weighted images were processed with FreeSurfer 7.4.1. Each session first underwent cross-sectional reconstruction, after which the three visits informed an unbiased within-participant anatomical template [S8]. Session-specific longitudinal reconstructions were initialized from that template. Cortical parcellation used the Desikan–Killiany atlas [S9]. Corrected diffusion images were rigidly aligned to the corresponding longitudinal anatomy using boundary-based registration. The inverse transform placed anatomical constraints and parcellation labels into diffusion space. Nearest-neighbour interpolation preserved integer node identities, while tissue images used the interpolation specified in the anatomical processing workflow.

Required cortical surfaces, subcortical segmentations and quality-control mosaics were checked for all sessions. Node indices and abbreviations were fixed before graph computation.

Fibre modelling, tractography and edge weights

White-matter, grey-matter and cerebrospinal-fluid response functions were estimated consistently for each session using the Dhollander procedure. The same cohort-average responses were then supplied to every session's reconstruction. Multi-shell multi-tissue constrained spherical deconvolution estimated tissue-specific fibre-orientation distributions [S10]. Multi-tissue log-domain intensity normalization was performed in MRtrix3 3.0.4 [S11]. Five-tissue-type anatomy and a grey–white-matter-interface seed

image were derived from the longitudinal FreeSurfer reconstruction. Tissue alignment, response stability and fibre-orientation quality were checked before final tractography.

Anatomically constrained iFOD2 probabilistic tractography generated 10 million accepted streamlines per session [S12]. Tracking used grey–white-matter-interface seeding, backtracking and cropping at that interface. The random seed was 20260730, with single-thread tracking to avoid scheduling-dependent stochastic differences. Retained MRtrix defaults included a 45° maximum angle and an ACT fibre-orientation cutoff of 0.05. Step size was half the isotropic-equivalent voxel size; minimum and maximum lengths were two and 100 times that reference, respectively. On the diffusion grid, these corresponded to approximately 1.267 mm, 5.069 mm and 253.455 mm. The same settings applied across sessions rather than being tuned to a participant's apparent connectome quality.

SIFT2 assigned a continuous weight to each streamline by relating the reconstructed tractogram to fibre-orientation integrals [S13]. The procedure used anatomical constraints and retained its default regularization; grey-matter scaling for single-tissue reconstruction was not enabled. Endpoints were assigned to anatomical nodes using an explicit 4-mm radial search. Separate matrices recorded streamline counts, summed SIFT2 weights and mean streamline lengths. The primary graph analyses used SIFT2-weighted matrices from the final processing of all 66 sessions.

Node definition, normalization and primary graph metrics

The MRtrix FreeSurfer conversion table defined 84 cortical, subcortical and cerebellar nodes. This Desikan84 definition combined 68 cortical parcels with 16 bilateral subcortical/cerebellar regions in a fixed order. Every matrix was required to be finite, non-negative, symmetric and zero-diagonal. An undirected edge existed wherever its SIFT2 weight was positive. For each session, we divided every weight by the sum of the unique upper-triangular weights. Total undirected weight therefore equalled one, while binary edge presence was unchanged. The unthresholded normalized graph was the primary representation.

Weighted global efficiency followed the inverse-shortest-path definition [S14]. Positive strength w was converted to graph length 1/w, and all-pairs Dijkstra distances were computed. Global efficiency was the mean inverse distance over ordered pairs of distinct nodes; an unreachable pair contributed zero. The implementation explicitly used edge weights rather than the unweighted NetworkX global-efficiency function. Graph length in this calculation is an inverse-weight cost, not the physical length of a reconstructed streamline. Raw-matrix efficiency was retained as a supporting measure, but it was not a second primary outcome.

Weighted modularity Q was estimated using Louvain optimization with resolution $\gamma = 1$ [S15]. For each primary graph, optimization was repeated 100 times with deterministic seeds 1234–1333. We retained the highest-Q partition, with the earliest seed resolving an exact tie. This was a selected partition, not a consensus partition assembled across runs. The Q distribution, community-count range and agreement

with the selected partition were retained as stability diagnostics. Neither resolution nor partition selection was changed according to a hormone association.

Raw SIFT2 total weight, binary density and mean tract length were retained as quality-control or sensitivity variables.

Alternative graph definitions and quality control

Fixed-density graphs retained the strongest edges at target densities of 10%, 15%, 20% and 25%. Stable descending sorting and node-pair order resolved weight ties, and thresholding preceded total-weight renormalization. These alternative graphs used 25 Louvain repetitions, seeds 1234–1258, rather than the 100 repetitions used for primary graphs. Disconnected thresholded graphs were permitted, with unreachable pairs contributing zero to efficiency. Connectivity and isolated-node counts were reported alongside estimates so that threshold-related failures would remain visible.

A separate length-aware group-consistency mask was motivated by work on spatially informed consensus thresholding [S16,S17]. For each of the 3,486 possible edges, consistency was the fraction of sessions with positive weight. Representative length was the mean streamline length across sessions where that edge was present. Candidate edges were sorted by representative length and divided into ten deterministic, approximately equal-count bins. Within each bin, the retained count was the rounded sum of edge consistencies. The most consistent edges were selected, with node-pair order resolving ties. This procedure used reconstructed streamline length, not Euclidean distance between parcel centroids.

The common binary mask was multiplied into each raw SIFT2 matrix before total-weight normalization. Efficiency and modularity were recomputed, using 100 Louvain repetitions with seeds 1234–1333 for the masked graphs. The mask approximated cohort-average edge counts across length strata and was built without hormone or phase information. A common mask does not guarantee identical observed session density because a retained edge can be absent in an individual session. The node definition also remained fixed rather than selecting an atlas after observing results [S18].

Quality control covered input integrity, anatomical surfaces, diffusion motion and outliers, registration, fibre orientations, tractography coverage and final matrix structure. Graph distributions and within-participant trajectories were inspected before merging hormone information for inference. A disconnected primary graph, isolated node, non-finite metric or incorrect matrix size was a hard failure. Distributional and longitudinal-range flags were review targets rather than automatic exclusions. Earlier cautions were carried forward into sensitivity analyses even when downstream images were visually acceptable. Technical acceptance therefore meant usable outputs with documented cautions; it did not mean that every imaging artifact had been excluded.

Every session was screened for image completeness, finite intensities, gradient consistency, eddy outputs, relative and absolute motion, outlier slices, T1–b0 alignment, brain-mask coverage and left–right orientation. Full-resolution visual review found no gross orientation error, brain truncation or failed

registration. Motion was characterized by mean and maximum relative displacement and the percentage of outlier slices. Relative motion was retained as a primary covariate rather than an exclusion rule.

Primary longitudinal model and multiplicity

E2 and P4 were analysed separately after log(1 + x) transformation. For each hormone, we subtracted the participant's three-session mean transformed concentration to form the within-person deviation. Deviations were divided by the pooled sample SD across all 66 centred observations. Participant-average transformed concentrations were standardized using the 22 unique participant means. This separation distinguishes within-person associations from between-person differences [S19]. Each graph outcome and relative-motion measure was standardized across 66 sessions using sample SD. Visit order was coded 1–3 and centred at 2.

For each hormone–outcome pair, the Gaussian mixed model included within-person hormone, participant-average hormone, centred visit order and relative motion as fixed effects. A participant random intercept represented repeated-measures dependence. The primary coefficient was the standardized within-person hormone association. We used restricted maximum likelihood with L-BFGS optimization, followed by deterministic Powell and conjugate-gradient fallbacks if needed. Wald z tests, standard errors and 95% confidence intervals were retained, together with raw-outcome changes per within-person hormone SD. The four two-sided E2/P4-by-efficiency/modularity tests formed one Benjamini–Hochberg false-discovery-rate family, with $q < 0.05$ as the decision threshold [S20].

The fixed-effect formula was z(outcome) ~ hormone_within_z + hormone_between_z + visit_order_centered + relative_motion_z, with a separate participant random intercept.

Resampling, diagnostics and participant influence

Participant-cluster bootstrap resampled the 22 participants with replacement 2,000 times for each primary association. Each selected participant contributed all three visits, and repeated selections received separate bootstrap identifiers. We reported percentile 95% intervals and the proportion of coefficients sharing the primary sign. These intervals were not adjusted for multiplicity. For the supplementary permutation check, each participant's three within-person exposure values were independently shuffled 5,000 times. Outcomes, between-person exposure, visit order and motion remained fixed. The empirical two-sided p value was the add-one count of absolute permuted coefficients at least as large as observed, divided by 5,001. Four-test BH adjustment was applied separately to these empirical p values.

Diagnostics included convergence, random-intercept and residual variances, intraclass correlation, design condition number and variance-inflation factors. Residual checks included Shapiro–Wilk, Jarque–Bera, Breusch–Pagan, residual–fitted correlation and standardized residual extremes. Diagnostic p values were flags for interpretation, not automatic instructions to change the primary model. Influence was evaluated by omitting each participant in turn. Targeted analyses separately omitted sub-06/ses-03 and participants sub-06, sub-13, sub-27, sub-28 or sub-32. sub-06/ses-03 was red-flagged in the pilot for displacement

spikes above 4 mm with the initial quadratic eddy field model (maximum 1.57 mm with the final linear field model) and was assigned a predeclared exclusion sensitivity analysis. The participant targets carried quality-control flags for motion, graph distribution or longitudinal range. Bootstrap, permutation and omission results supplemented the primary estimate; none could substitute for its declared four-test correction.

Secondary and exploratory analyses

The four primary hormone–outcome combinations were repeated across four densities, producing a separate 16-test BH family. Length-aware graphs supplied another four-test family. LH and FSH separately replaced the steroid exposure in the primary specification, giving four secondary tests. Phase models used menses as the reference, with participant random intercepts and adjustment for visit order and motion. Estimated ovulation versus menses, mid-luteal versus estimated ovulation, and mid-luteal versus menses were evaluated for each outcome. These six contrasts formed one family. Density, mask, gonadotropin and phase analyses used 2,000 cluster bootstraps and 5,000 within-participant exposure or phase-label permutations per estimate.

Four secondary global metrics were computed from unthresholded graphs: weighted clustering, transitivity, characteristic path length and strength assortativity. Clustering used mean Onnela node clustering after maximum-weight normalization. Transitivity used weighted triangle intensity relative to ordered connected triples, with the same maximum-weight normalization. Path length averaged finite Dijkstra distances after mapping normalized strength to inverse-weight length. Assortativity was the edge-weighted correlation of endpoint strengths, representing each undirected edge in both directions. Their E2/P4 models formed an eight-test family. Log-transformed raw total SIFT2 weight and observed density formed a separate four-test control family. These families used 1,000 cluster bootstraps and 2,000 within-participant permutations per estimate.

Exploratory node models evaluated normalized strength, weighted node efficiency and weighted participation relative to each session's selected primary Louvain partition. Each of 84 nodes was tested separately for E2 and P4, producing 504 models. BH correction was applied within each 84-test hormone–metric family. The network-based statistic (NBS) addressed connected sets of suprathreshold edges [S21]. Edge outcomes were log(1 + raw SIFT2 weight), standardized across sessions. Mass-univariate models included participant fixed effects, within-person hormone, visit order and motion. Participant effects absorbed time-invariant terms, so a separate between-person hormone coefficient was unnecessary.

NBS formed positive and negative connected components separately at $|t| \geq 3.1$, with extent defined as the number of edges. Five thousand within-participant hormone permutations supplied a joint null distribution. Each draw applied the same within-participant session permutation to E2 and P4, preserving their joint ordering. Each permutation retained the maximum component extent across both hormones and both signs. Component p values used add-one exceedance counts, thereby assessing family-wise error

across the joint search. Uncorrected edge p values were descriptive, and component findings could not establish significance for every constituent edge. The threshold was not searched retrospectively to favour a component.

Observational exposure shuffling depends on within-participant exchangeability relative to fixed covariates and is not a randomized-exposure experiment.

Post-primary checks, unassessed sensitivities and reproducibility

Two additional estimator families were retained after the primary results: random-intercept-plus-hormone-slope mixed models and participant-fixed-effect models with participant-clustered standard errors. Each family contained four tests and received separate BH correction. Random-effect covariance eigenvalues and optimizer warnings were retained alongside convergence indicators. These checks were post-primary, and no Bayesian or regularized random-slope analysis had been specified before primary inference. Sensitivity analyses using a Schaefer-100 parcellation or eddy correction without the synthetic susceptibility reference could not be completed for the full cohort. Although acquisition times were recorded in the public dataset, no acquisition-time-adjusted model was fitted. Atlas dependence, alternative susceptibility correction and time-of-day confounding therefore remain incompletely assessed.

Graph and statistical analyses used Python 3.10.12, NumPy 2.2.6, pandas 2.3.3 and SciPy 1.15.3. Additional versions were NetworkX 3.4.2, matplotlib 3.10.9, statsmodels 0.14.6 and Patsy 1.0.1. Statistical resampling used seed 20260817, distinct from the tractography seed. The processing and analysis scripts are available from the corresponding authors upon reasonable request. Because graph measures depend on reconstruction and pipeline choices [S22], all declared families and cautions were retained in reporting.

Container execution used Apptainer 1.5.3. Statistical resampling used deterministic substreams of the stated base seed.

Supplementary Results

All 66 primary graphs were connected without isolated nodes (Supplementary Figure S1). At 10% fixed density, 13 session graphs were disconnected and contained isolated nodes (Supplementary Figure S4). All graphs were connected at 15%, 20% and 25%. The length-aware procedure considered 3,190 edges present at least once, forming ten bins of 319 candidate edges. Its final mask retained 2,035 of 3,486 possible edges, with mask density 0.583764. All masked graphs were connected without isolated nodes, but a common mask did not imply identical observed session density. The 20 alternative-graph model estimates are reported in Supplementary Table S3.

All 66 sessions met graph-retention criteria: 38 were classified PASS and 28 PASS WITH CAUTION, with no hard failure or exclusion. Hormone distributions showed periovulatory E2/LH elevation and mid-luteal P4 elevation (main Table 1). Primary-model ICCs were 0.749–0.847: most conditional outcome

variance lay between participants. These values are model variance shares, not directly measured test–retest reliability.

All four primary models converged (Supplementary Figure S3; Supplementary Table S2). Maximum predictor VIF was 1.053 and condition numbers were 1.37–1.38. Breusch–Pagan tests did not indicate heteroscedasticity ($p = 0.145$–0.961). Shapiro–Wilk p values were 0.045–0.231, and Jarque–Bera tests flagged three models at $p < 0.05$; maximum absolute standardized residuals were 3.42–3.62. No residual-based exclusion was introduced. The 2,000 bootstrap and 5,000 permutation fits were valid for every primary model (Supplementary Figure S2).

The E2–efficiency coefficient remained negative across all 22 participant omissions (Supplementary Table S1). Omitting sub-27 attenuated it most ($\beta = -0.101$) but did not reverse its sign. All 924 extended leave-one-participant-out fits and 252 targeted sensitivity fits were valid. No prespecified sensitivity converted a corrected null family into a discovery.

None of the LH/FSH, phase-contrast, secondary-topology or total-weight/density analyses yielded a corrected finding. The four LH/FSH estimates and six phase contrasts are reported in Supplementary Table S4; the eight secondary-topology and four control estimates are reported in Supplementary Table S5. These results are also shown in Supplementary Figure S5.

Node-level testing produced 504 converged models and no $q < 0.05$ in any of the six hormone–metric families (Supplementary Table S6). The minimum q was 0.077632 in the E2–efficiency family. NBS produced 12 candidate components (Supplementary Table S7); minimum p_FWER was 0.215557, and no component survived joint family-wise correction.

Post-primary participant-fixed-effect E2–efficiency was $\beta = -0.1337$ (95% CI −0.2529 to −0.0145), with nominal $p = 0.0280$ and separate-family $q = 0.1118$. The random-slope estimate was $\beta = -0.1004$ (95% CI −0.2542 to 0.0533; $q = 0.4008$). Its random-effect covariance was singular and the Hessian was not positive definite. Neither estimator family yielded a corrected discovery. These checks cannot replace the primary random-intercept analysis or establish equivalence. All eight post-primary estimates and model cautions are presented in Supplementary Table S8.

Supplementary Table S1. Participant influence and targeted exclusion analyses

Association	Omission β range	Same sign as primary	Omissions with p ≥ 0.05	Targeted-exclusion β range
E2 / efficiency	−0.159 to −0.101	22/22	9/22	−0.145 to −0.101
P4 / efficiency	−0.131 to −0.082	22/22	21/22	−0.131 to −0.082
E2 / modularity Q	−0.007 to 0.081	21/22	22/22	0.039 to 0.081
P4 / modularity Q	0.041 to 0.100	22/22	21/22	0.056 to 0.100

Each participant-omission analysis retained 21 participants and 63 sessions. The sign column counts estimates with the same direction as the full-cohort primary estimate; p values are nominal and are not a criterion for corrected discovery. Six targeted exclusions were evaluated per model: session 03 of sub-06, and participants sub-06, sub-13, sub-27, sub-28 or sub-32 individually. The session exclusion retained 22 participants and 65 sessions; each participant exclusion retained 21 participants and 63 sessions. All 88 omission fits and 24 targeted fits converged. β denotes the standardized within-person coefficient.

Supplementary Table S2. Primary mixed-model diagnostics

Association	ICC	Shapiro–Wilk p	Jarque–Bera p	Breusch–Pagan p	Maximum |standardized residual|
E2 / efficiency	0.749	0.0927	0.0198	0.6302	3.62
P4 / efficiency	0.751	0.0822	0.0543	0.1446	3.54
E2 / modularity Q	0.847	0.2315	0.0426	0.9606	3.42
P4 / modularity Q	0.840	0.0450	0.0023	0.8435	3.57

All four models converged without optimizer warnings (22 participants; 66 sessions). Design condition numbers ranged from 1.37 to 1.38; the maximum predictor variance-inflation factor was 1.053. ICC is the random-intercept variance divided by the sum of random-intercept and residual variance; it is not a test–retest reliability estimate. Diagnostic p values are unadjusted screening measures. Residual diagnostics did not trigger observation removal or replacement of the primary model. See Supplementary Figure S3.

Supplementary Table S3. Hormone associations across alternative graph definitions

Graph	Outcome	Hormone	β [95% CI]	p	BH q
10% density	Efficiency	E2	−0.128 [−0.254 to −0.002]	0.0468	0.187
10% density	Efficiency	P4	−0.111 [−0.237 to 0.016]	0.0869	0.213
10% density	Modularity Q	E2	0.055 [−0.030 to 0.139]	0.2050	0.273
10% density	Modularity Q	P4	0.033 [−0.051 to 0.118]	0.4403	0.470
15% density	Efficiency	E2	−0.129 [−0.253 to −0.006]	0.0401	0.187
15% density	Efficiency	P4	−0.103 [−0.228 to 0.022]	0.1053	0.213
15% density	Modularity Q	E2	0.039 [−0.076 to 0.154]	0.5065	0.506
15% density	Modularity Q	P4	0.085 [−0.027 to 0.197]	0.1352	0.240
20% density	Efficiency	E2	−0.130 [−0.255 to −0.005]	0.0411	0.187
20% density	Efficiency	P4	−0.104 [−0.231 to 0.022]	0.1051	0.213
20% density	Modularity Q	E2	0.047 [−0.063 to 0.158]	0.3976	0.454
20% density	Modularity Q	P4	0.074 [−0.034 to 0.181]	0.1810	0.263
25% density	Efficiency	E2	−0.131 [−0.257 to −0.006]	0.0404	0.187
25% density	Efficiency	P4	−0.104 [−0.231 to 0.022]	0.1063	0.213
25% density	Modularity Q	E2	0.051 [−0.056 to 0.158]	0.3484	0.429
25% density	Modularity Q	P4	0.073 [−0.032 to 0.178]	0.1724	0.263
Length-aware	Efficiency	E2	−0.132 [−0.258 to −0.006]	0.0394	0.157
Length-aware	Efficiency	P4	−0.105 [−0.232 to 0.022]	0.1047	0.209
Length-aware	Modularity Q	E2	0.041 [−0.063 to 0.145]	0.4417	0.442
Length-aware	Modularity Q	P4	0.063 [−0.039 to 0.164]	0.2288	0.305

All models used 22 participants and 66 sessions, participant random intercepts, and adjustment for participant-average transformed hormone, visit order and relative motion. Fixed-density and length-aware estimates received separate BH corrections across 16 and four tests, respectively. Confidence intervals are two-sided Wald intervals. At 10% density, 13 graphs were disconnected; no graphs were excluded, and unreachable pairs contributed zero to efficiency. All graphs at the other densities and all length-aware graphs were connected. The length-aware mask retained 2,035 of 3,486 possible edges. E2, estradiol; P4, progesterone; BH, Benjamini–Hochberg. See Figure 3 and Supplementary Figure S4.

Supplementary Table S4. Secondary hormone associations and menstrual-phase contrasts

Family	Exposure / contrast	Outcome	β [95% CI]	p	BH q
LH/FSH	LH	Efficiency	−0.068 [−0.202 to 0.066]	0.3210	0.919
LH/FSH	FSH	Efficiency	−0.019 [−0.155 to 0.116]	0.7815	0.919
LH/FSH	LH	Modularity Q	−0.014 [−0.122 to 0.093]	0.7963	0.919
LH/FSH	FSH	Modularity Q	−0.006 [−0.113 to 0.102]	0.9194	0.919
Phase	O − M	Efficiency	−0.255 [−0.578 to 0.067]	0.1209	0.363
Phase	L − O	Efficiency	−0.002 [−0.339 to 0.336]	0.9924	0.992
Phase	L − M	Efficiency	−0.257 [−0.567 to 0.053]	0.1038	0.363
Phase	O − M	Modularity Q	0.047 [−0.217 to 0.311]	0.7297	0.876
Phase	L − O	Modularity Q	0.066 [−0.210 to 0.342]	0.6395	0.876
Phase	L − M	Modularity Q	0.112 [−0.140 to 0.365]	0.3832	0.766

All models used 22 participants and 66 sessions, participant random intercepts, and adjustment for visit order and relative motion. Continuous-hormone models additionally included participant-average transformed hormone. Continuous-hormone β values are standardized within-person coefficients; phase β values are differences in standardized outcome units. Four LH/FSH tests and six phase contrasts received separate BH corrections. Confidence intervals are two-sided Wald intervals. LH, luteinizing hormone; FSH, follicle-stimulating hormone; M, menses; O, estimated ovulation; L, mid-luteal. See Supplementary Figure S5.

Supplementary Table S5. Additional global topology and control outcomes

Family	Outcome	Hormone	β [95% CI]	p	BH q
Topology	Clustering	E2	−0.075 [−0.194 to 0.044]	0.2167	0.340
Topology	Clustering	P4	−0.050 [−0.169 to 0.069]	0.4142	0.414
Topology	Transitivity	E2	−0.051 [−0.158 to 0.057]	0.3550	0.406
Topology	Transitivity	P4	−0.061 [−0.167 to 0.044]	0.2551	0.340
Topology	Path length	E2	0.121 [−0.037 to 0.279]	0.1323	0.340
Topology	Path length	P4	0.130 [−0.026 to 0.286]	0.1031	0.340
Topology	Assortativity	E2	0.075 [−0.034 to 0.183]	0.1779	0.340
Topology	Assortativity	P4	0.073 [−0.035 to 0.181]	0.1840	0.340
Control	Total weight	E2	0.037 [−0.039 to 0.113]	0.3401	0.477
Control	Total weight	P4	−0.035 [−0.111 to 0.040]	0.3578	0.477
Control	Density	E2	0.089 [−0.042 to 0.221]	0.1841	0.477
Control	Density	P4	0.030 [−0.103 to 0.164]	0.6589	0.659

All models used 22 participants and 66 sessions, participant random intercepts, and adjustment for participant-average transformed hormone, visit order and relative motion. β is the standardized within-person hormone coefficient. Total SIFT2 weight was transformed as log(1 + weight) before outcome standardization. The eight topology tests and four total-weight/density control tests received separate BH corrections. Confidence intervals are two-sided Wald intervals. Metric definitions are given in the Supplementary Methods. See Supplementary Figure S5.

Supplementary Table S6. Exploratory nodal analyses by hormone–metric family

Hormone	Nodal metric	Tests	Converged	β range	Minimum BH q	Nodes with q < 0.05
E2	Efficiency	84	84	−0.231 to 0.132	0.0776	0
E2	Participation	84	84	−0.231 to 0.149	0.4377	0
E2	Strength	84	84	−0.227 to 0.134	0.1795	0
P4	Efficiency	84	84	−0.167 to 0.094	0.3146	0
P4	Participation	84	84	−0.266 to 0.180	0.1430	0
P4	Strength	84	84	−0.116 to 0.101	0.6808	0

Each of the six families included all 84 nodes, yielding 504 models in the same 22 participants and 66 sessions. Models used participant random intercepts and adjusted for participant-average transformed hormone, visit order and relative motion. BH correction was applied separately within each 84-test family. Ranges summarize all standardized within-person coefficients; minimum q values describe the family and do not establish a significant node. No node survived correction. See Figure 4a,b.

Supplementary Table S7. All candidate network-based statistic components

Hormone	Direction	Component	Edges	Nodes	Joint FWER p
E2	Positive	1	5	6	0.2156
E2	Positive	2	3	4	0.5495
E2	Negative	3	1	2	1.0000
E2	Negative	4	1	2	1.0000
E2	Negative	5	1	2	1.0000
E2	Negative	6	1	2	1.0000
E2	Positive	7	1	2	1.0000
E2	Positive	8	1	2	1.0000
E2	Positive	9	1	2	1.0000
P4	Negative	1	2	3	0.8812
P4	Negative	2	1	2	1.0000
P4	Positive	3	1	2	1.0000

Components were formed separately for each hormone and sign at $|t| \geq 3.1$; component numbering is within each hormone. Extent is the number of edges, and node counts can overlap across components. The 5,000 within-participant permutations generated one joint maximum-extent distribution across both hormones and both signs. Corrected p values used add-one exceedance counts. No component met $p < 0.05$ after this joint family-wise error-rate (FWER) correction. Component membership does not confer significance on an individual edge. See Figure 4c,d.

Supplementary Table S8. Post-primary estimator sensitivity analyses

Estimator	Association	β [95% CI]	p	BH q
Participant fixed effects	E2 / efficiency	−0.134 [−0.253 to −0.014]	0.0280	0.112
Participant fixed effects	P4 / efficiency	−0.105 [−0.258 to 0.048]	0.1775	0.355
Participant fixed effects	E2 / modularity Q	0.047 [−0.102 to 0.196]	0.5374	0.537
Participant fixed effects	P4 / modularity Q	0.065 [−0.060 to 0.189]	0.3089	0.412
Random intercept + slope	E2 / efficiency*	−0.100 [−0.254 to 0.053]	0.2004	0.401
Random intercept + slope	P4 / efficiency	−0.110 [−0.253 to 0.033]	0.1328	0.401
Random intercept + slope	E2 / modularity Q†	0.016 [−0.095 to 0.127]	0.7780	0.778
Random intercept + slope	P4 / modularity Q	0.058 [−0.060 to 0.177]	0.3329	0.444

These analyses followed the primary results and did not replace the primary random-intercept models. Each estimator family contained four tests with separate BH correction. All eight models used the same 22 participants and 66 sessions and adjusted for visit order and relative motion. Random-slope models also included participant-average transformed hormone; participant fixed effects absorbed time-invariant between-person terms, with participant-clustered standard errors. Confidence intervals are two-sided Wald intervals. All fits met their numerical convergence criterion. * E2–efficiency random-slope covariance was singular and the Hessian was not positive definite. † E2–modularity random-slope covariance was singular and the estimate was flagged as lying on the parameter-space boundary. These cautions limit interpretation despite numerical convergence.

Supplementary References

S1. Gorgolewski, K. J. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044 (2016). https://doi.org/10.1038/sdata.2016.44.

S2. Rizor, E. J. et al. Hormone Health Study (HHS): Brain Anatomy and Structure (OpenNeuro ds005360, version 1.0.1). *OpenNeuro* https://doi.org/10.18112/openneuro.ds005360.v1.0.1 (2024).

S3. Rizor, E. J. et al. Menstrual cycle-driven hormone concentrations co-fluctuate with white and gray matter architecture changes across the whole brain. *Hum. Brain Mapp.* **45**, e26785 (2024). https://doi.org/10.1002/hbm.26785.

S4. Veraart, J. et al. Denoising of diffusion MRI using random matrix theory. *NeuroImage* **142**, 394-406 (2016). https://doi.org/10.1016/j.neuroimage.2016.08.016.

S5. Schilling, K. G. et al. Synthesized b0 for diffusion distortion correction (Synb0-DisCo). *Magn. Reson. Imaging* **64**, 62-70 (2019). https://doi.org/10.1016/j.mri.2019.05.008.

S6. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage* **125**, 1063-1078 (2016). https://doi.org/10.1016/j.neuroimage.2015.10.019.

S7. Tustison, N. J. et al. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310-1320 (2010). https://doi.org/10.1109/TMI.2010.2046908.

S8. Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* **61**, 1402-1418 (2012). https://doi.org/10.1016/j.neuroimage.2012.02.084.

S9. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968-980 (2006). https://doi.org/10.1016/j.neuroimage.2006.01.021.

S10. Jeurissen, B., Tournier, J. D., Dhollander, T., Connelly, A. & Sijbers, J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage* **103**, 411-426 (2014). https://doi.org/10.1016/j.neuroimage.2014.07.061.

S11. Tournier, J. D. et al. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* **202**, 116137 (2019). https://doi.org/10.1016/j.neuroimage.2019.116137.

S12. Smith, R. E., Tournier, J. D., Calamante, F. & Connelly, A. Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *NeuroImage* **62**, 1924-1938 (2012). https://doi.org/10.1016/j.neuroimage.2012.06.005.

S13. Smith, R. E., Tournier, J. D., Calamante, F. & Connelly, A. SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *NeuroImage* **119**, 338-351 (2015). https://doi.org/10.1016/j.neuroimage.2015.06.092.

S14. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* **52**, 1059-1069 (2010). https://doi.org/10.1016/j.neuroimage.2009.10.003.

S15. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008). https://doi.org/10.1088/1742-5468/2008/10/P10008.

S16. Betzel, R. F., Griffa, A., Hagmann, P. & Mišić, B. Distance-dependent consensus thresholds for generating group-representative structural brain networks. *Netw. Neurosci.* **3**, 475-496 (2019). https://doi.org/10.1162/netn_a_00075.

S17. Roberts, J. A., Perry, A., Roberts, G., Mitchell, P. B. & Breakspear, M. Consistency-based thresholding of the human connectome. *NeuroImage* **145**, 118-129 (2017). https://doi.org/10.1016/j.neuroimage.2016.09.053.

S18. Zalesky, A. et al. Whole-brain anatomical networks: does the choice of nodes matter? *NeuroImage* **50**, 970-983 (2010). https://doi.org/10.1016/j.neuroimage.2009.12.027.

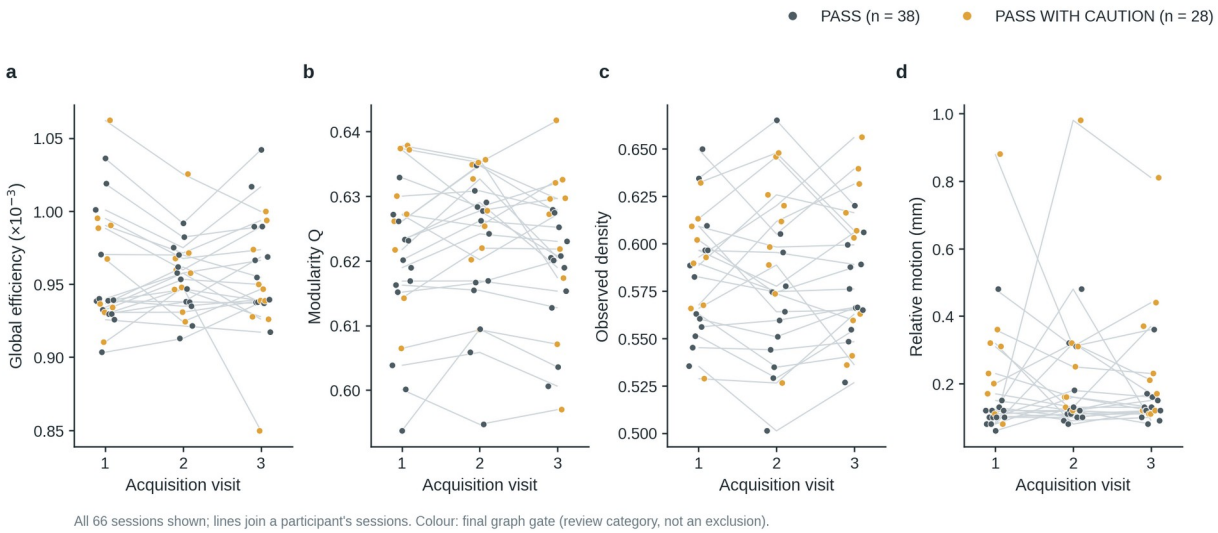
S19. Curran, P. J. & Bauer, D. J. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu. Rev. Psychol.* **62**, 583-619 (2011). https://doi.org/10.1146/annurev.psych.093008.100356.

S20. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289-300 (1995). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

S21. Zalesky, A., Fornito, A. & Bullmore, E. T. Network-based statistic: identifying differences in brain networks. *NeuroImage* **53**, 1197-1207 (2010). https://doi.org/10.1016/j.neuroimage.2010.06.041.

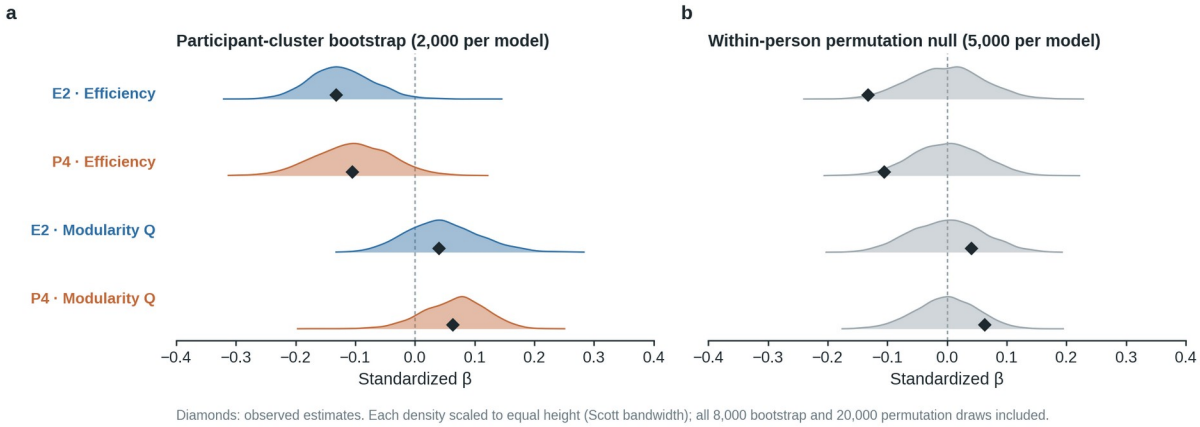
S22. Roine, T. et al. Reproducibility and intercorrelation of graph theoretical measures in structural brain connectivity networks. *Med. Image Anal.* **52**, 56-67 (2019). https://doi.org/10.1016/j.media.2018.10.009.

Supplementary Figure S1. Retained graph and motion measurements



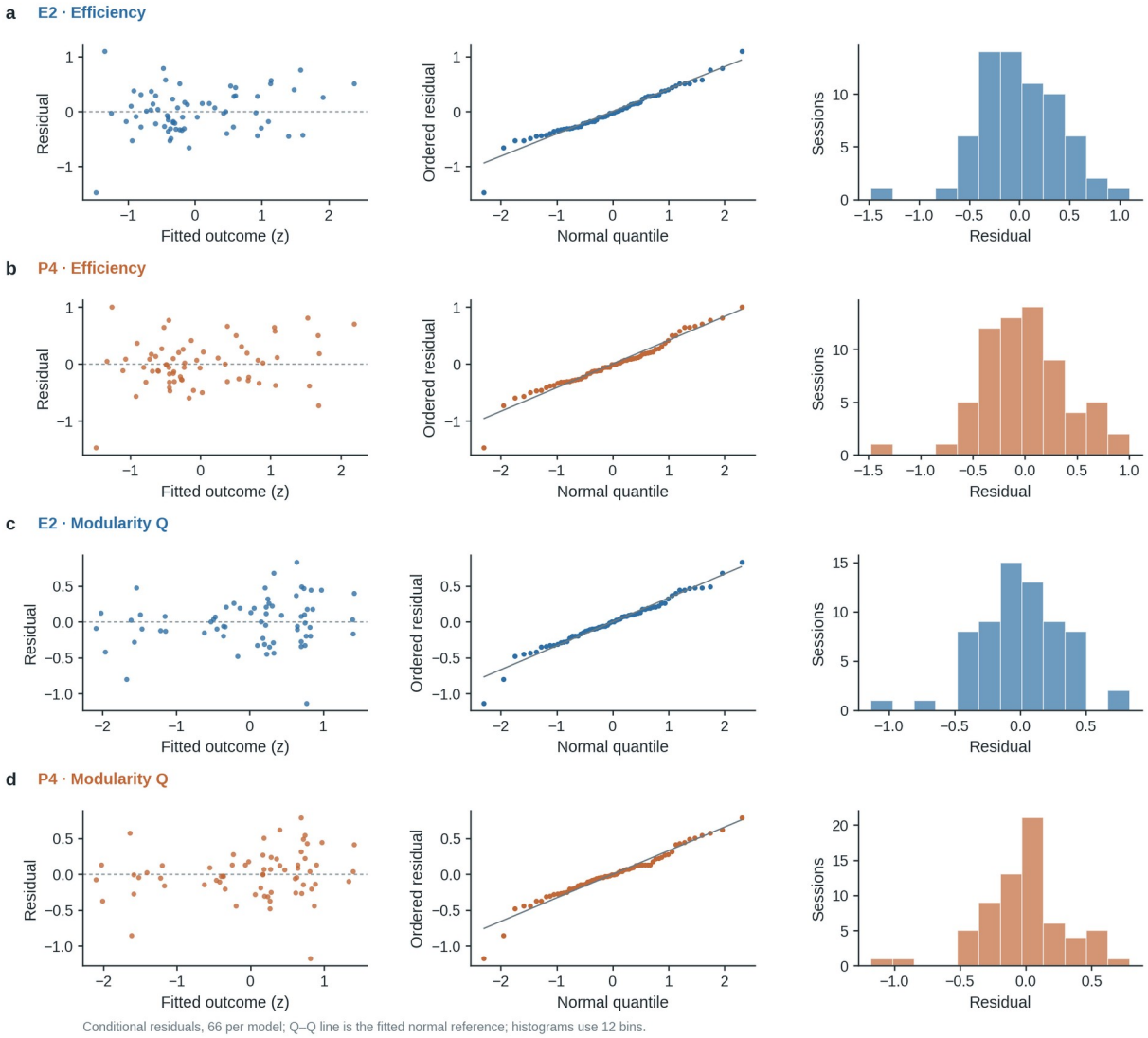
Supplementary Figure S1 | Retained graph and motion measurements. Acquisition-order trajectories for normalized efficiency (**a**), modularity Q (**b**), observed density (**c**) and relative motion (**d**). Every point is one of 66 sessions; lines join the three sessions from a participant. Colour reflects the final graph gate: PASS (grey), 38 sessions; PASS WITH CAUTION (amber), 28. PASS WITH CAUTION is a review category; it is neither an exclusion nor a complete list of earlier imaging flags. All primary graphs are connected with no isolated nodes. Efficiency tick values use units of 10^{-3}. No inferential comparison is performed in this descriptive figure.

Supplementary Figure S2. Bootstrap and permutation distributions



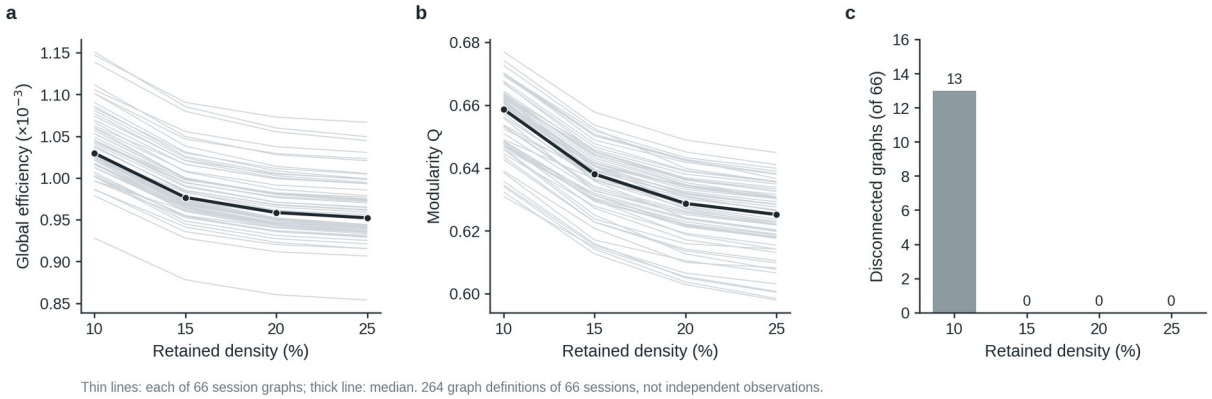
Supplementary Figure S2 | Bootstrap and permutation distributions. a, Participant-cluster bootstrap distributions, 2,000 coefficients per association (coloured by hormone). **b**, Within-person exposure-permutation null distributions, 5,000 coefficients per association (grey). The distributions include all 8,000 bootstrap and 20,000 permutation coefficients. Display uses Gaussian kernel densities with Scott bandwidth; each density is separately normalized to equal maximum height, so heights do not compare uncertainty or probability across rows. Diamonds show observed primary coefficients. Bootstrap intervals are unadjusted; permutation p values receive a separate four-test BH correction. Neither primary family yields $q < 0.05$. Exposure shuffling assumes the within-participant exchangeability described in the Methods.

Supplementary Figure S3. Primary-model residual diagnostics



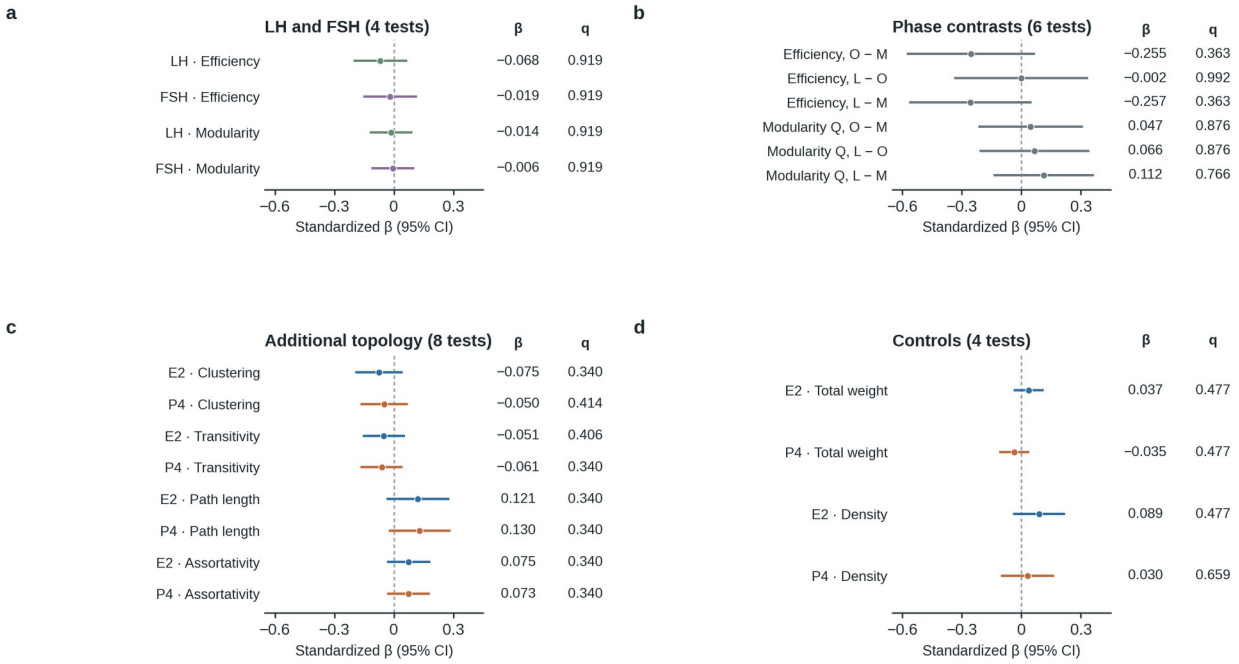
Supplementary Figure S3 | Primary-model residual diagnostics. Rows **a–d** correspond to E2–efficiency, P4–efficiency, E2–modularity and P4–modularity. Columns show conditional residuals against fitted standardized outcomes, normal quantile–quantile plots, and residual histograms using 12 bins. Each model contributes 66 conditional residuals. The Q–Q line is the fitted normal reference, not an additional model. Tail cautions did not trigger post-result exclusions. Numerical diagnostic summaries are given in Supplementary Table S2.

Supplementary Figure S4. Fixed-density graph quality control



Supplementary Figure S4 | Fixed-density graph quality control. a,b, Efficiency and modularity across 10%, 15%, 20% and 25% density. Thin lines show each of the 66 session graphs; the thick line connects session medians at each density. Efficiency tick values use units of 10^{-3}. **c**, Number of disconnected session graphs at each density: 13 at 10%, zero at the other thresholds. No graph was omitted. These 264 graph measurements represent repeated definitions of 66 sessions, not 264 independent observations. Corresponding model estimates are given in Supplementary Table S3.

Supplementary Figure S5. Secondary hormone, phase, topology and control results



Separate BH family per panel; none yields q < 0.05. M, menses; O, estimated ovulation; L, mid-luteal. 22 participants, 66 sessions.

Supplementary Figure S5 | Secondary hormone, phase, topology and control results. a, LH/FSH models, four tests. **b**, Phase contrasts, six tests; M, menses; O, estimated ovulation; L, mid-luteal. **c**, Additional topology models, eight tests. **d**, Raw-total-weight and density controls, four tests. Points are standardized estimates and horizontal lines are two-sided Wald 95% confidence intervals; columns give estimates and BH q values. Colour identifies the exposure: E2 blue, P4 brown, LH green, FSH purple; phase contrasts are grey. All models contain participant random intercepts and adjust for visit order and motion; continuous-hormone models additionally include participant-average hormone. Each panel has its own declared BH family. None yields q < 0.05. All use the same 22 participants and 66 sessions. Numerical estimates are given in Supplementary Tables S4 and S5.