Supplementary Information

Table S1. Evaluation of the CSGSA CCC discrimination model (OPLSDA)

		CSGSA (CCC discrimination)		Sum	Sens. or Spec.	
		Positive	Negative			
(**a**) Cut off = 0, test set classification						
CCC	Advanced	47	17	64	73.4%	53.9%
	Early	50	66	116	43.1%	
EM		3	69	72	95.8%	
Sum		100	152	252		
(**b**) Cut off = 0, corrected based on prevalence						
CCC	Advanced	22	8	30	73.4%	52.2%
	Early	30	40	70	43.1%	
EM		413	9488	9900	95.8%	
Sum		465	9535	10000		
PPV or NPV		11.2%	99.5%			

CSGSA: comprehensive serum glycopeptide spectra analysis; CCC: clear cell carcinoma; PPV: positive predictive value; NPV: negative predictive value; EM: endometrioma

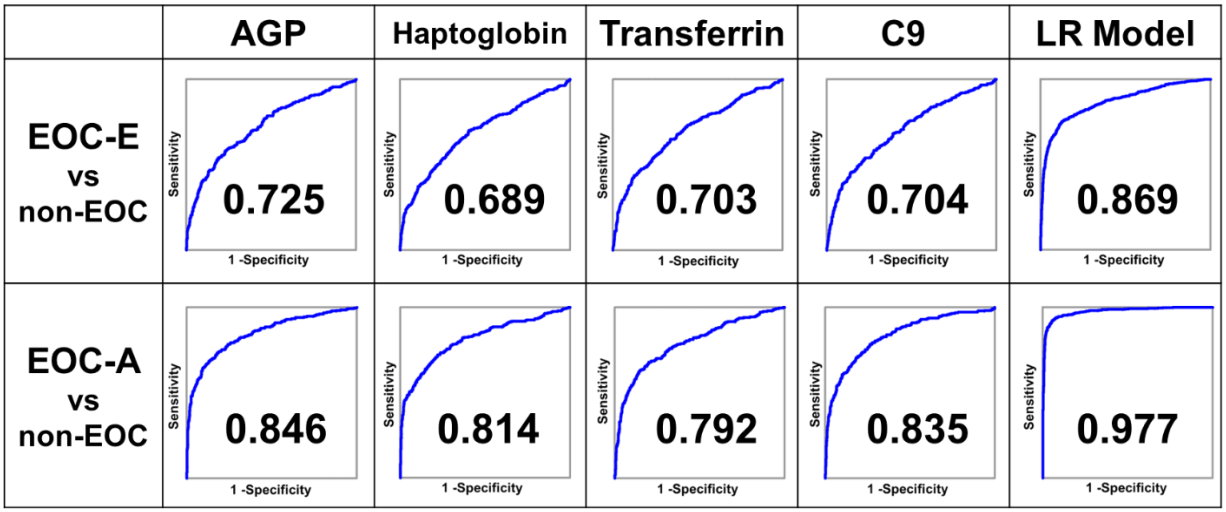

Figure S1: ROC Curves and AUC Values for Key Glycopeptides and Combination Model
This figure displays the ROC curves and AUC values for four key glycopeptides: alpha-1-acid glycoprotein (AGP) with a fully sialylated tetraantennary glycan with fucose at asparagine position 103, haptoglobin with a fully sialylated triantennary glycan with three fucoses at asparagine position 241, transferrin with a biantennary glycan with sialic acid at asparagine position 432, and complement C9 with a fully sialylated biantennary glycan at asparagine position 415. Additionally, results from a logistic regression-based combination model incorporating CA125, HE4, and these four glycopeptides are also depicted (LR Model). The model is applied to different stages of EOC: EOC-E (early stage) and EOC-A (advanced stage).

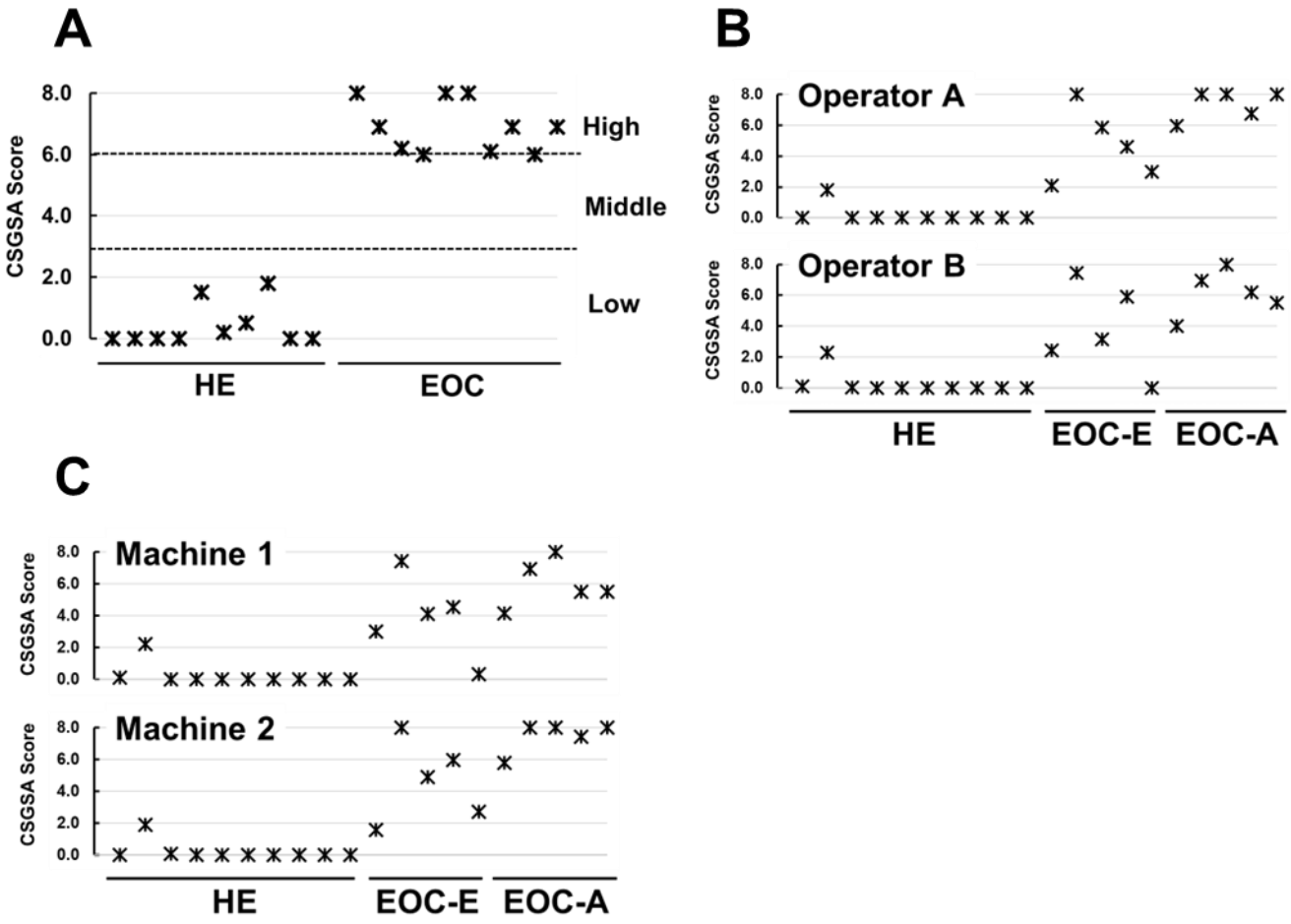

Figure S2. Reproducibility, inter-operator error, and inter-machine error
Reproducibility, inter-machine error, and inter-operator error were assessed through glycopeptide expressions, CSGSA scores, and OPLSDA scattering.
(**A**) Reproducibility was assessed by analyzing HE and EOC samples, with the process repeated 10 times for each sample.
(**B**) Inter-operator error was assessed by two operators using 10 HE samples and 10 EOC samples (including 5 EOC-E and 5 EOC-A).
(**C**) Inter-machine error was assessed by comparing the analysis of 10 HE samples and 10 EOC samples using two different mass spectrometers.

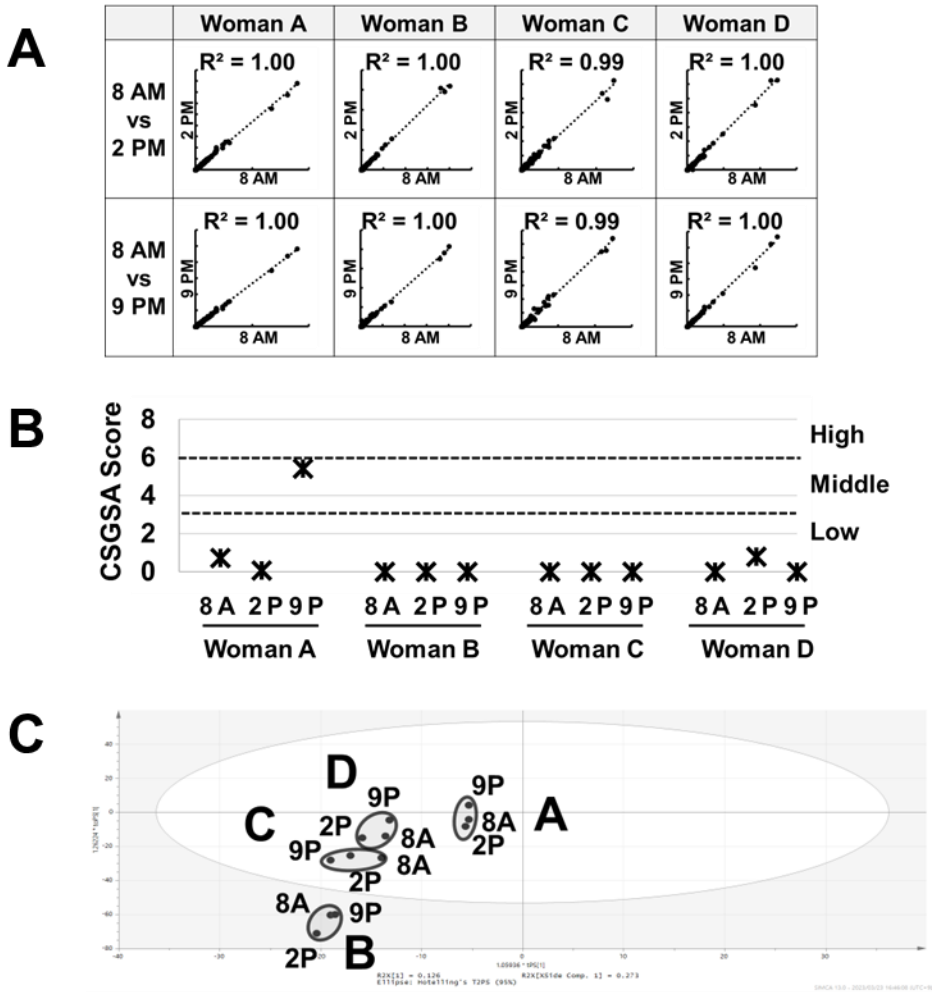

Figure S3. Effects of circadian rhythm and diet

Sera were collected from four healthy female volunteers at 8 AM, 2 PM, and 9 PM, and changes in glycopeptide expressions, CSGSA scores, and OPLSDA spots were assessed.

(**A**) Scatter plots depicting glycopeptide expressions between the different conditions.

(**B**) CSGSA scores obtained from the CNN model.

(**C**) OPLSDA score plots.

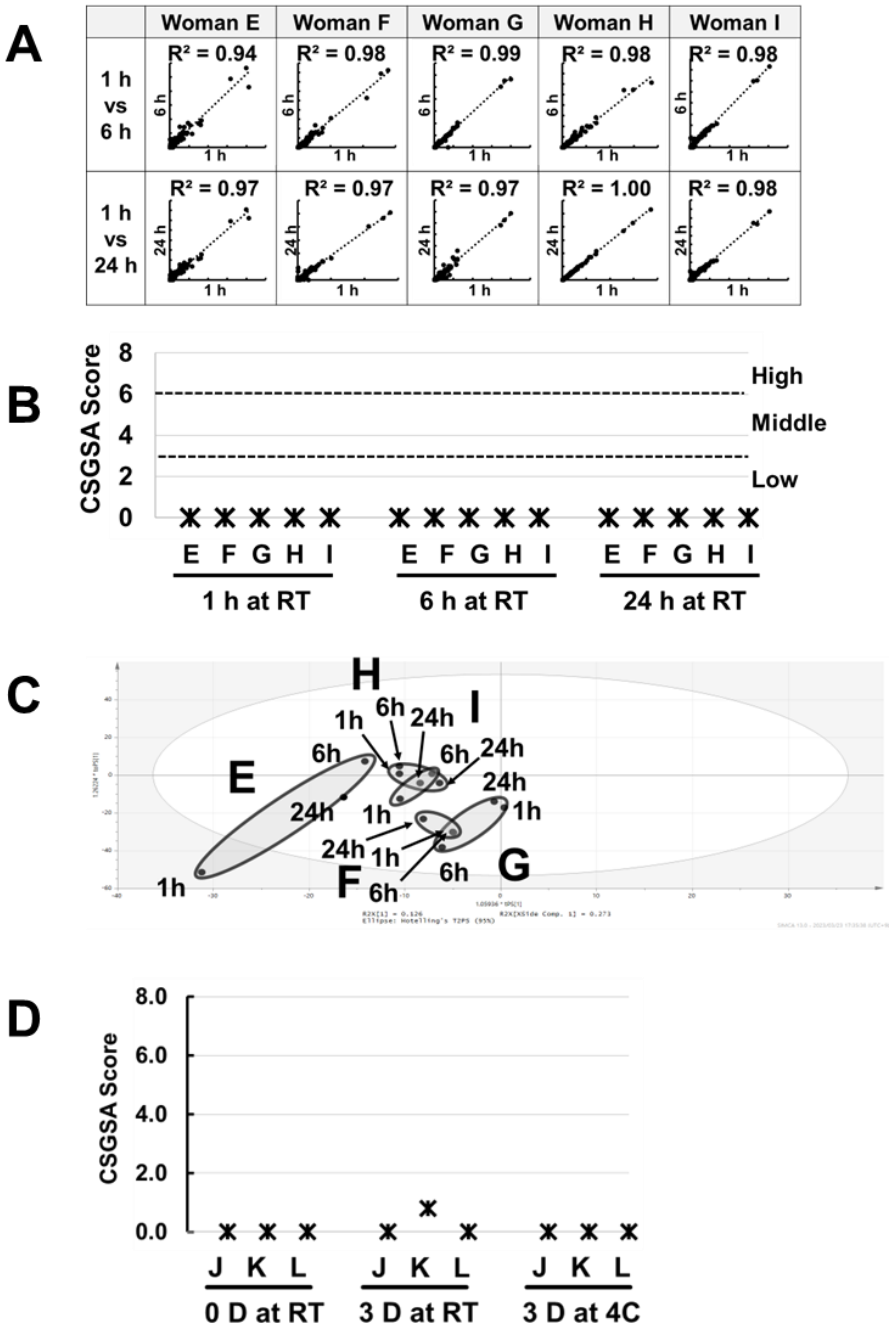

Figure S4. Short-term stability in blood

Blood samples were collected from five healthy woman volunteers and left at room temperature (22 to 25 ℃), for 1, 6, and 24 h in a tube.

(**A**) Scatter plots depicting glycopeptide expressions between 1 h and 6 or 24 h.

(**B**) CSGSA scores obtained from the CNN model.

(**C**) OPLSDA score plots.

(**D**) Blood samples collected from three healthy woman volunteers were left at room temperature for 3 days, and CSGSA scores obtained from the CNN model were assessed.

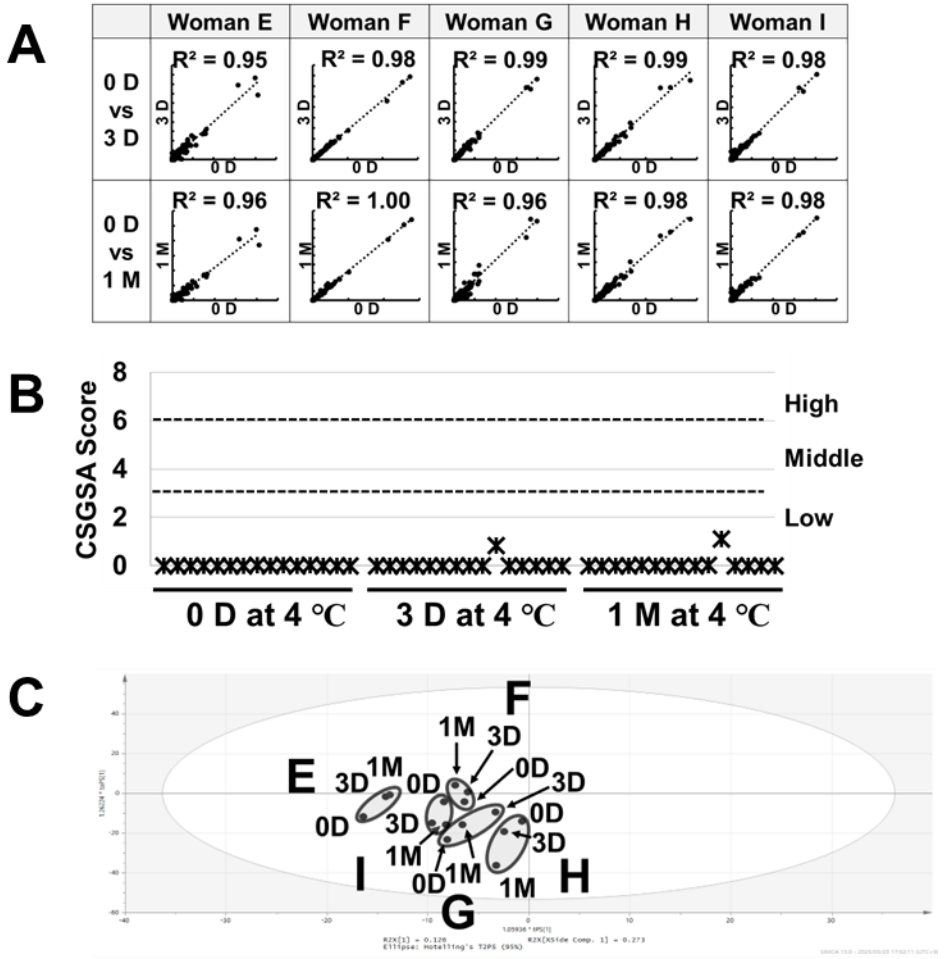

Figure S5. Long-term storage stability in serum at 4 °C

Sera were collected from five healthy woman volunteers and kept at 4 °C for 3 days and 1 month.

(**A**) Scatter plots depicting glycopeptide expressions between 0 day and 3 days or 1 month.

(**B**) CSGSA scores obtained from the CNN model.

(**C**) OPLSDA score plots.

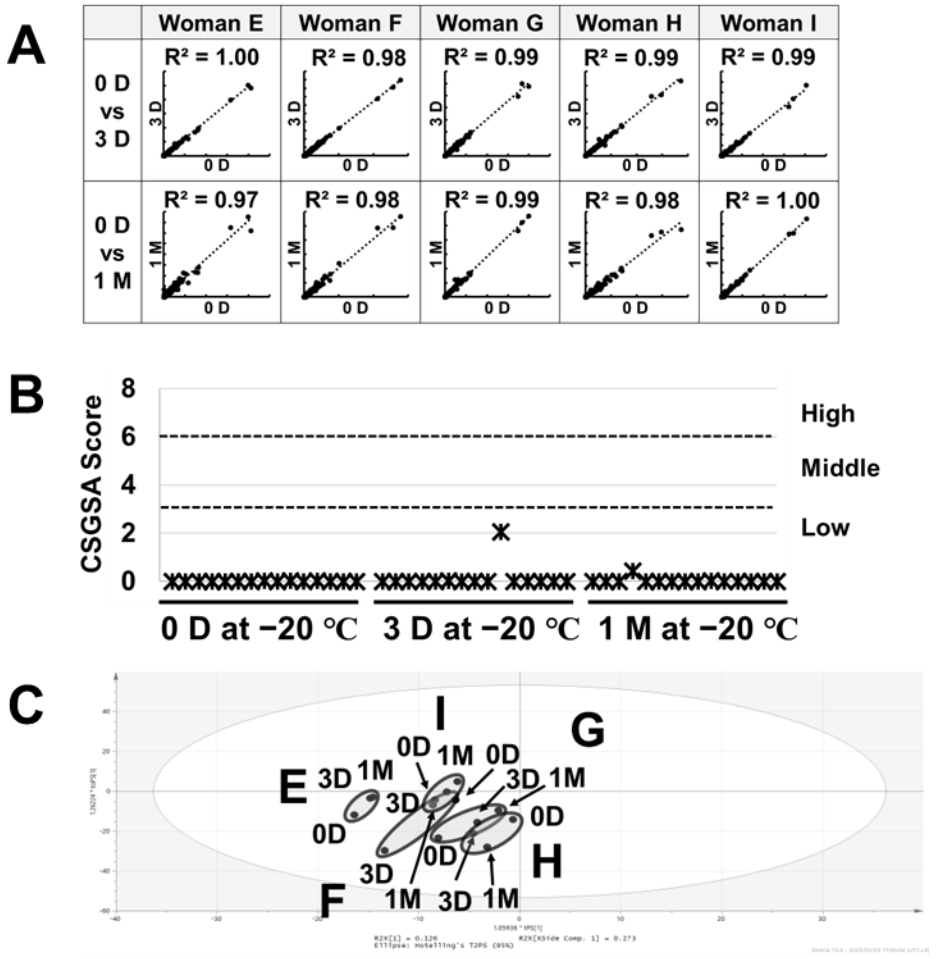

Figure S6. Long-term storage stability in serum at -20 °C

Sera were collected from five healthy woman volunteers and kept at -20 °C for 3 days and 1 month.

(**A**) Scatter plots depicting glycopeptide expressions between 0 day and 3 days or 1 month.

(**B**) CSGSA scores obtained from the CNN model.

(**C**) OPLSDA score plots.

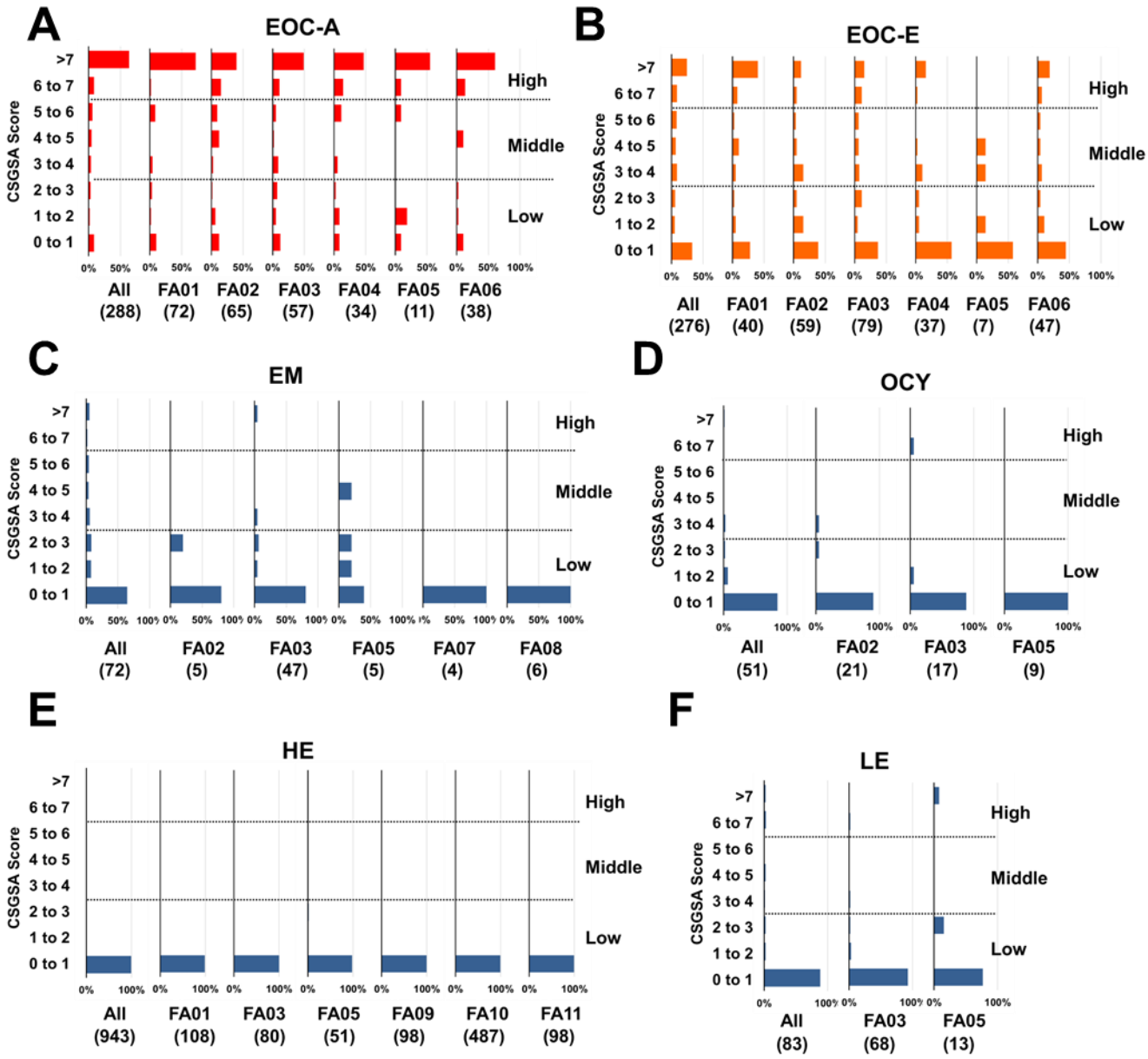

Figure S7. Inter-facility differences

Histograms of CSGSA scores were generated for each facility and disease category, including
(**A**) advanced-stage epithelial ovarian cancer (EOC-A),
(**B**) early-stage epithelial ovarian cancer (EOC-E),
(**C**) endometrioma (EM),
(**D**) ovarian cyst (OCY),
(**E**) healthy women (HE), and
(**F**) uterine myoma (LE).

Cut-offs were set at “3” and “6”, classifying patients into high, middle, and low-risk groups. The FA + number is an identifier of the facilities, and the number of patients is indicated in parentheses.

Estimation of Sample Size:

1. We assumed an alpha error of 0.05 and a beta error of 0.2, corresponding to a detection power of 80%.
2. The Z-value for the alpha error was set at 1.96, and for the beta error at 0.842.
3. We predicted that the average expression levels of glycopeptides would differ by no more than half the standard deviation between the control and case groups.
4. The rational minimum number was calculated using the following formula:

   Rational Minimum Number
   $= 2 \times (\text{Z value of alpha error} + \text{Z value of beta error})^2 \times (\text{Standard deviation} / \text{Predicted difference between controls and cases})^2$
   $= 2 \times (1.96 + 0.842)^2 \times 2^2 = 63.$