Electronic Supplementary Material (ESI) for Nanoscale.
This journal is © The Royal Society of Chemistry 2023

Supplementary Information For

Explainable Machine-Learning Approach for Revealing Complex Synthesis Path–Property Relationships of Nanomaterials

Kun Jin [a, ‡], Wentao Wang [a, ‡], Guangpei Qi [a, ‡], Xiaohong Peng [b, ‡], Haonan Gao [a], Hongjiang Zhu [a], Xin He [a], Haixia Zou [a,], Lin Yang [a], Junjie Yuan [a], Liyuan Zhang [c], Hong Chen [d,*] and Xiangmeng Qu [a,*]

a Key Laboratory of Sensing Technology and Biomedical Instruments of Guangdong Province and School of Biomedical Engineering, Sun Yat-Sen University, Shenzhen 518107, China

b YueYang Central Hospital, YueYang 414000, China

c School of Petroleum Engineering, State Key Laboratory of Heavy Oil Processing China University of Petroleum (East China), Qingdao, 266580, China

d Pen-Tung Sah Institute of Micro-Nano Science and Technology, Xiamen University, Xiamen, Fujian, China 361005

Supplemental Data Items

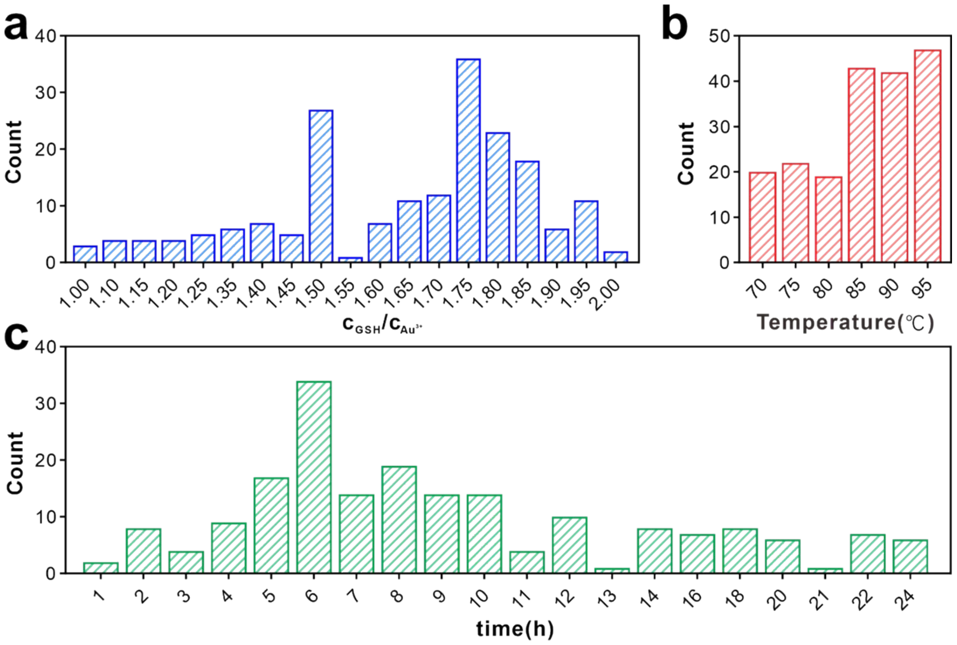

Fig S1. Histogram of the overall dataset over each feature. (a) The histogram of the temperature. (b) The histogram of the concentration of GSH. (c) The histogram of the time.

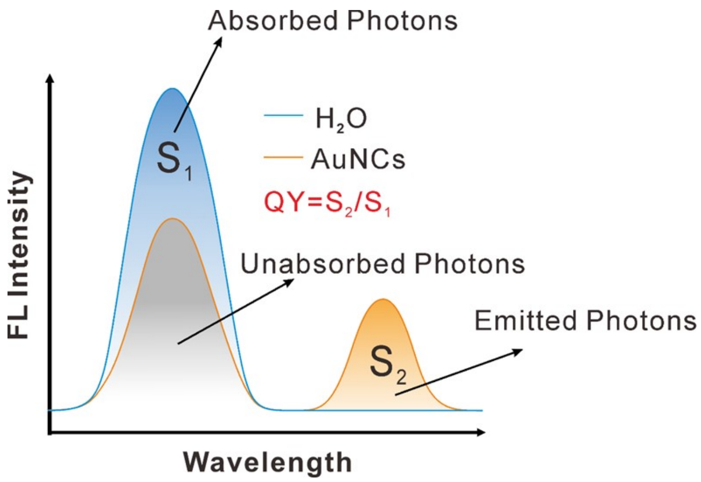

Fig S2. Schematic illustration of the calculation of absolute fluorescence QYs of AuNC.

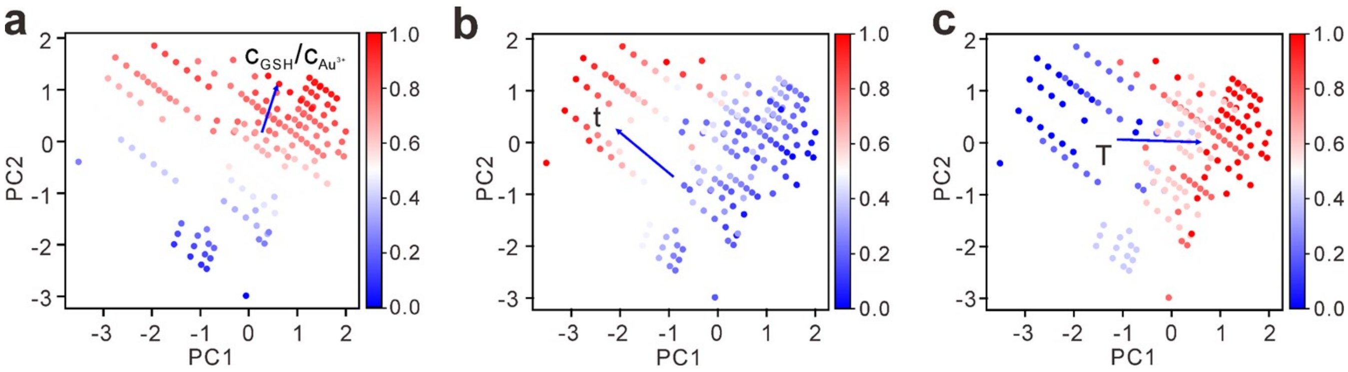

Fig S3. PCA analysis of the AuNCs data with 3 principal components. a, PCA analysis of QY data with the thiol-to-metal molar ratios ($C_{GSH}/C_{Au}{}^{3+}$) synthetic parameter. **b,** PCA analysis of QY data with the reaction time (*t*) synthetic parameter. **c,** PCA analysis of QY data with the reaction temperature (*T*) synthetic parameter. The arrows represent the distribution direction of the parameters.

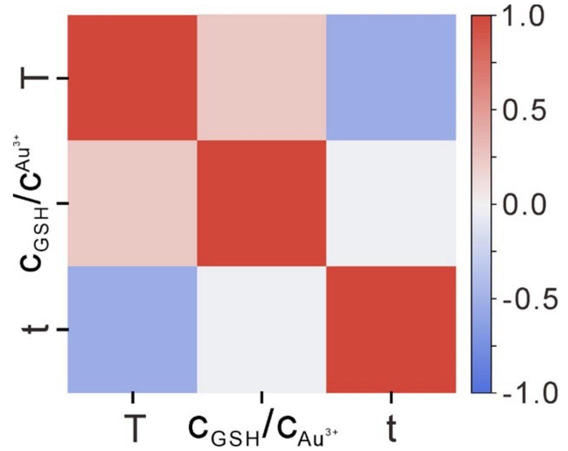

Fig S4. Pearson's correlation coefficient matrix of different synthetic parameters.

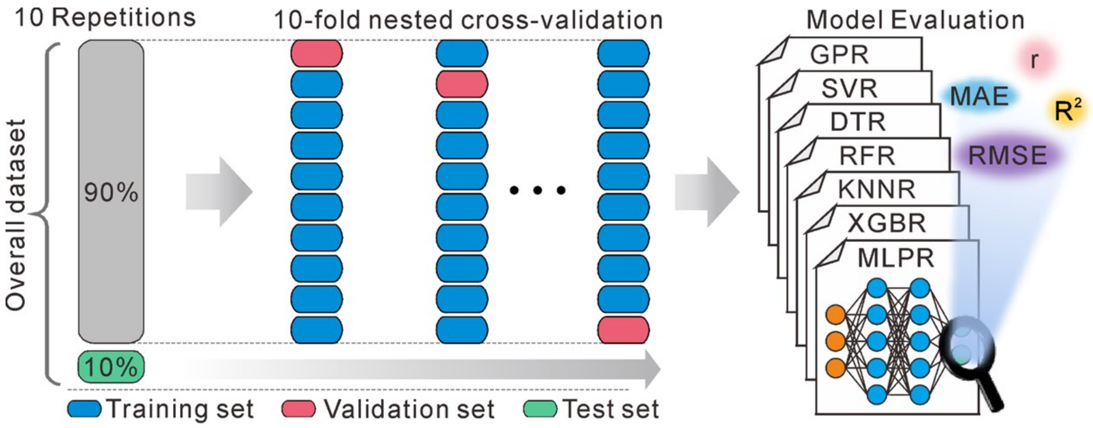

Fig S5. Scheme workflow for machine learning regression model's selection using nested cross-validation.

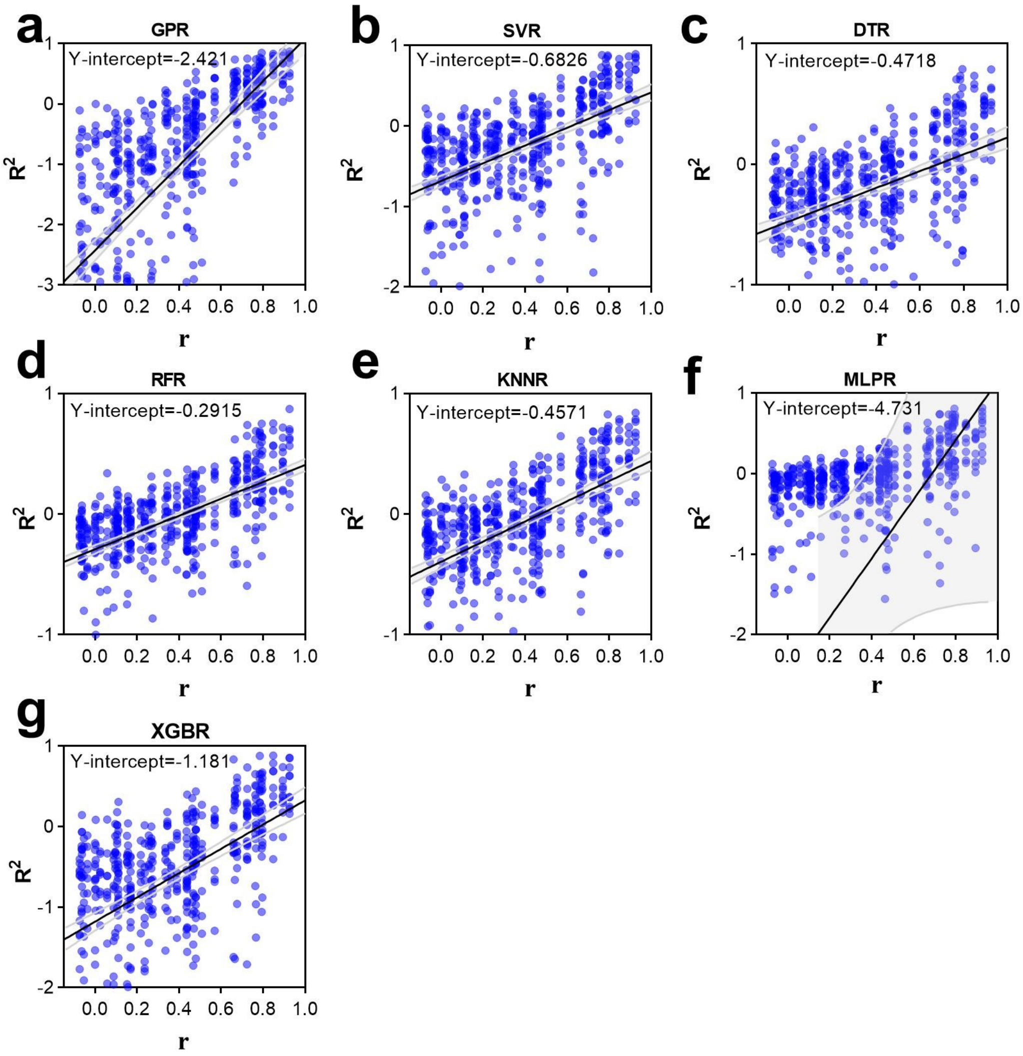

Fig S6. Overfitting test using a permutation test. The Y-axis represents the coefficient of determinations (R^2) of the models trained by the whole dataset after permutating QYs, and the X-axis represents the corresponding Pearson's correlation coefficients (r) of the raw QYs and permutated QYs. The linear regressions on the r values and the corresponding R^2 values indicate that the model is not overfitting (intercept < 0.05).

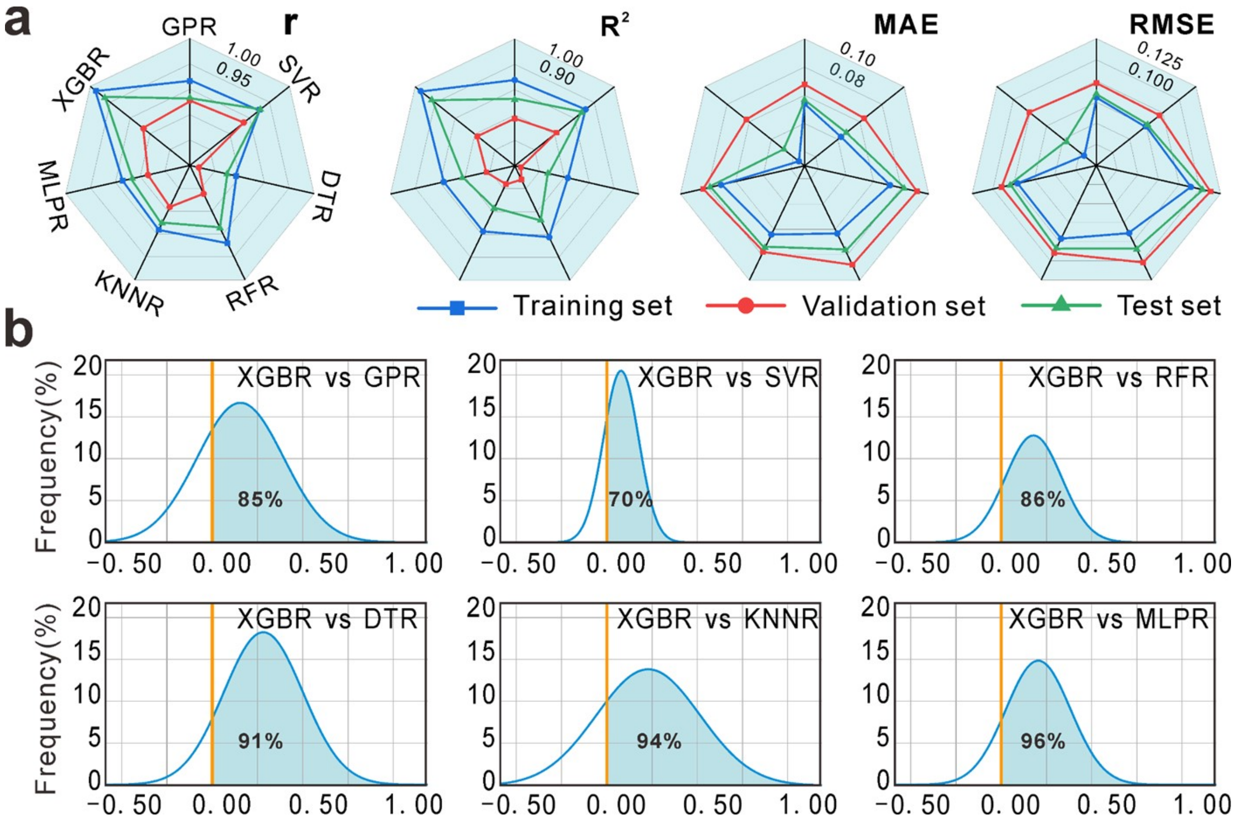

Fig S7. a, Radar-map of the mean of r, R^2, MAE, and RMSE. **b,** Bayesian correlated t-tests, revealing comparison results of 10 repetitions of cross-validation between XGBR and the other six candidate models in inferring QY. XGBR outperforms the other six candidate models significantly.

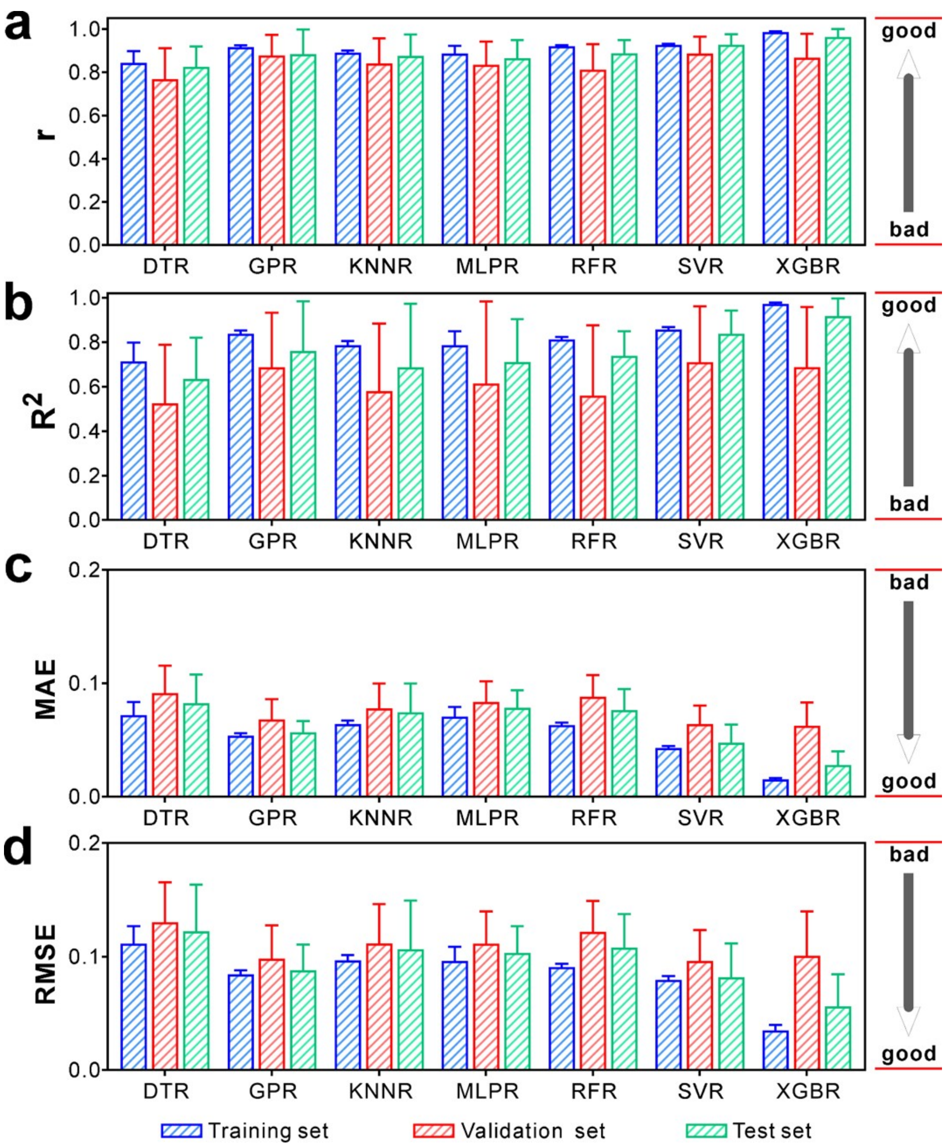

Fig S8. Model evaluation for inferring QY from the parameters more precisely. Histograms of the seven candidate models' performance indexes of the 10 repetitions × 10-fold cross-validation, including Pearson's correlation coefficient (r), coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). All values were expressed as mean ± SD (n = 10 repetitions × 10 folds = 100).

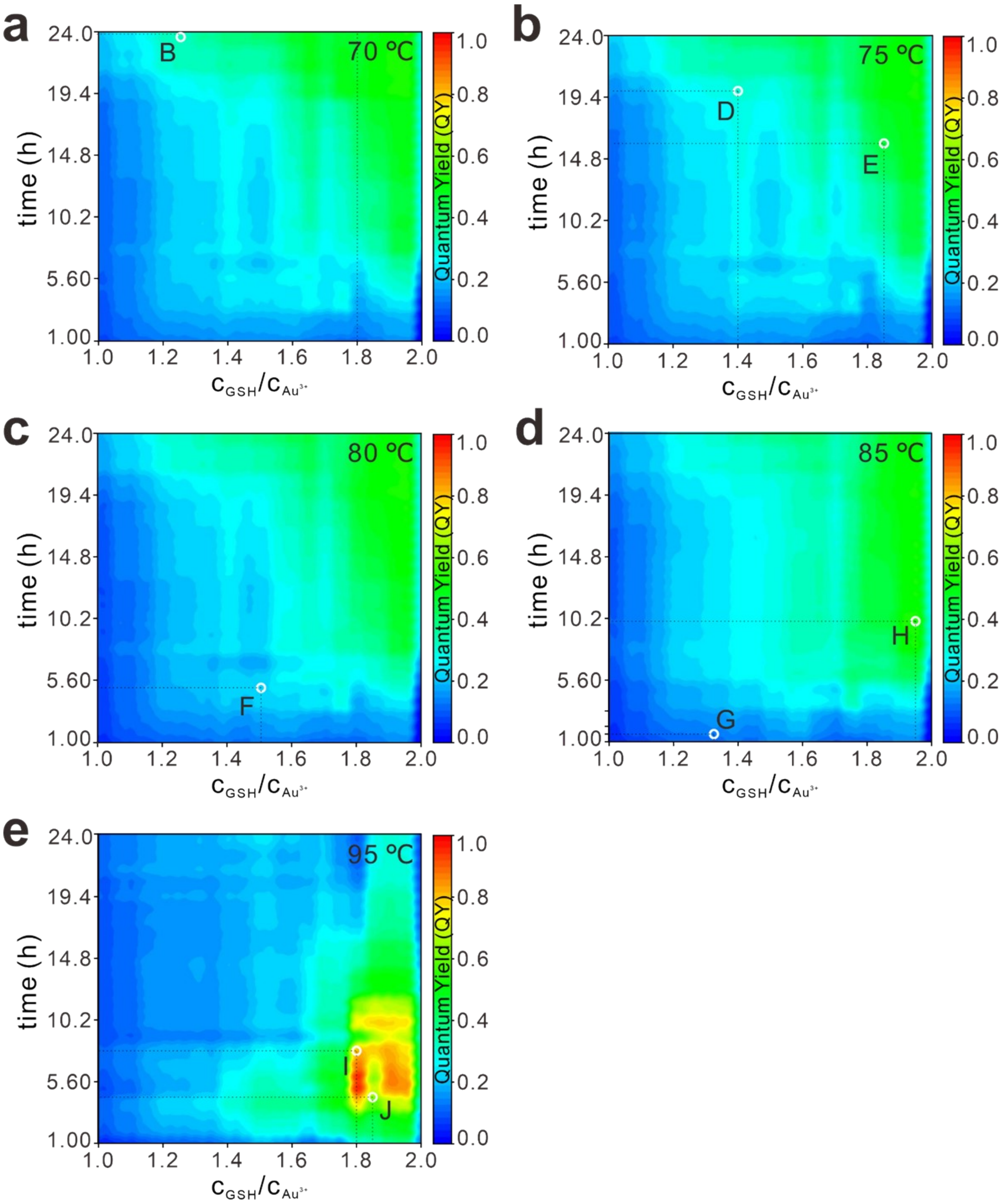

Fig S9. The synthetic phase diagram determined by the trained XGBR model at different temperatures 70℃(a), 75℃(b), 80℃(c), 85℃(d), and 95℃(e). The region with red color indicates that GSH-AuNC with higher fluorescence QY can be synthesized under this synthetic parameter. In contrast, the region with blue color indicates that GSH-AuNC with lower fluorescence QY can be synthesized under this synthetic parameter.

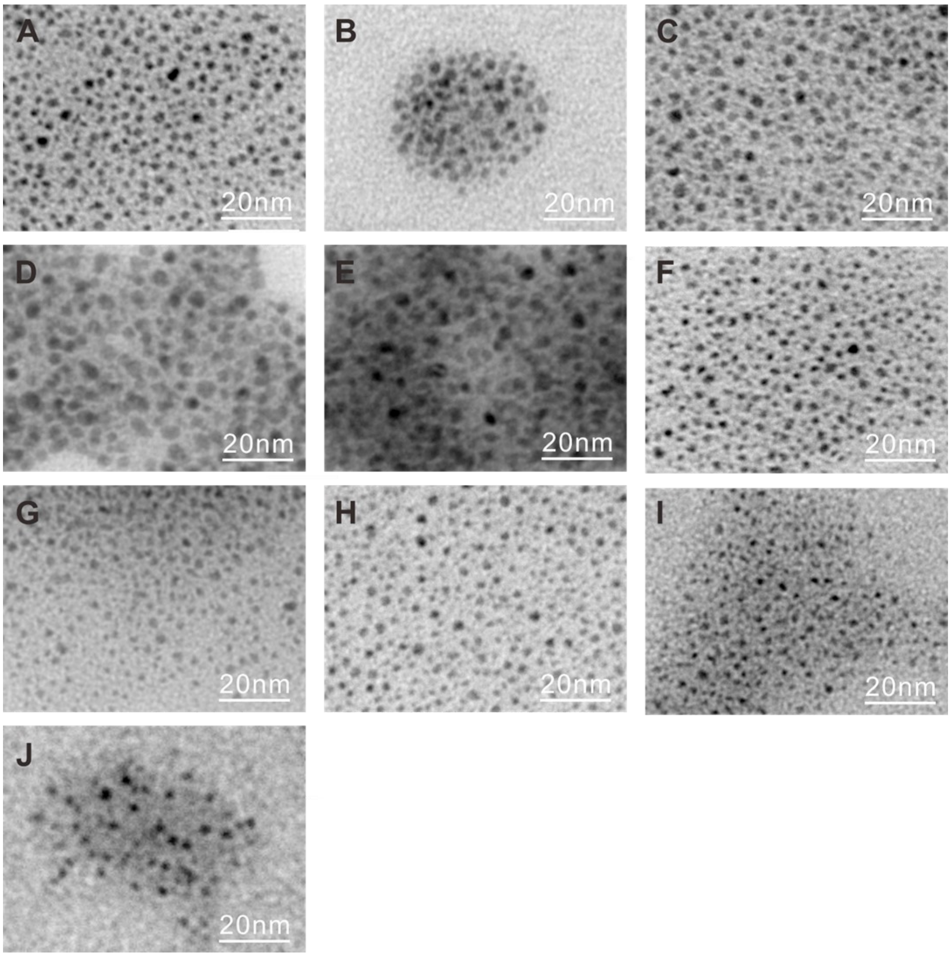

Fig S10. TEM images of reverse synthesized GSH-AuNC.

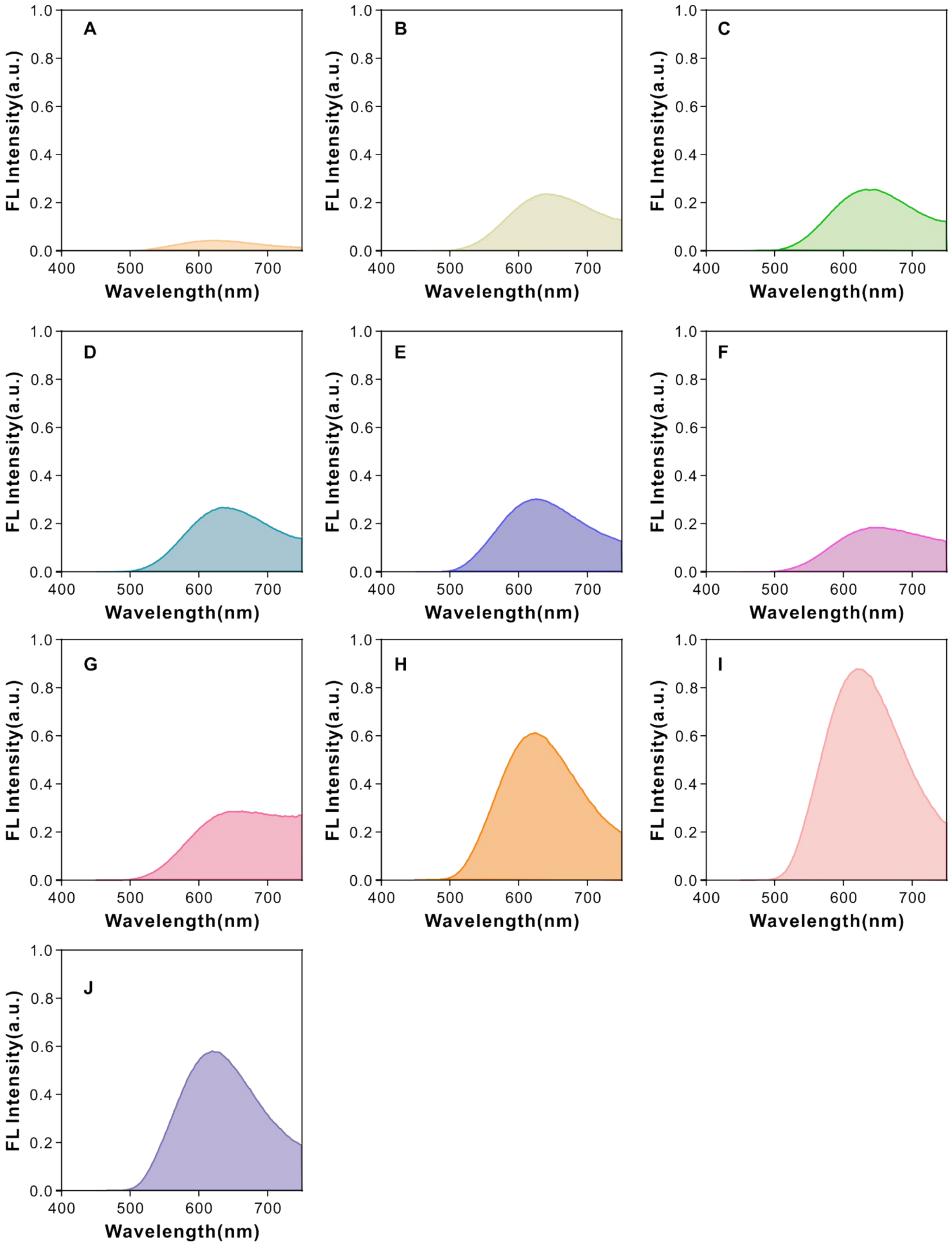

Fig S11. Fluorescence spectrum of reverse synthesized GSH-AuNC. The synthesis parameters are as follows: $C_{GSH}/C_{Au}{}^{3+}$=1.9, T=90°C, t=6h(A); $C_{GSH}/C_{Au}{}^{3+}$=1.25, T=70°C, t=24h(B); $C_{GSH}/C_{Au}{}^{3+}$=1.5, T=90°C, t=8h(C); $C_{GSH}/C_{Au}{}^{3+}$=1.4, T=75°C, t=20h(D); $C_{GSH}/C_{Au}{}^{3+}$=1.85, T=75°C, t=16h(E); $C_{GSH}/C_{Au}{}^{3+}$=1.5, T=80°C, t=5h(F); $C_{GSH}/C_{Au}{}^{3+}$=1.35, T=85°C, t=2h(G); $C_{GSH}/C_{Au}{}^{3+}$=1.95, T=85°C, t=10h(H); $C_{GSH}/C_{Au}{}^{3+}$=1.8, T=95°C, t=8h(I); $C_{GSH}/C_{Au}{}^{3+}$=1.85, T=95°C, t=4h(J).

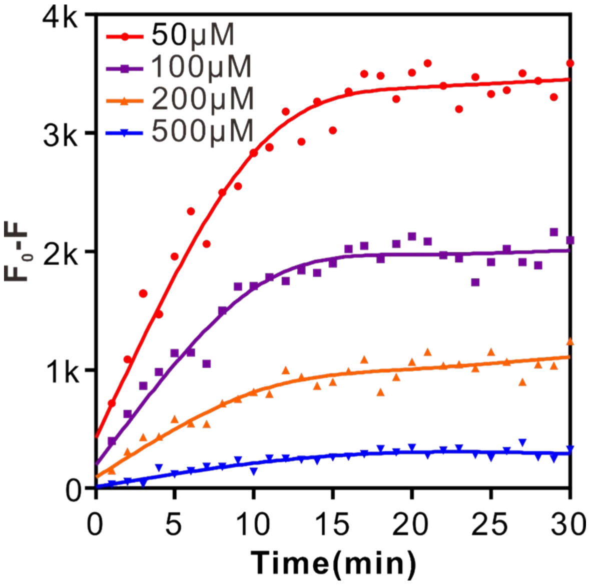

Fig S12. Kinetic process of the fluorescence intensity change (F_0-F) of AuNCs at 625 nm after adding different concentrations of Cu^{2+} (50 μM, 100 μM, 200 μM, 500 μM).

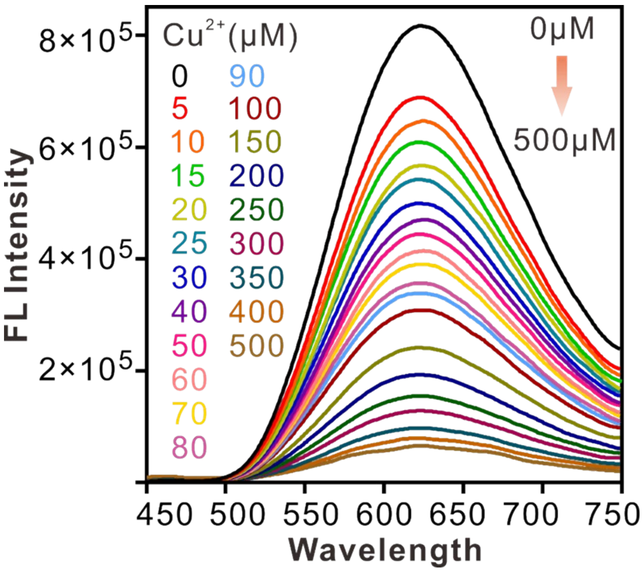

Fig S13. Fluorescence spectra of AuNCs after adding different concentrations of Cu^{2+}, respectively.

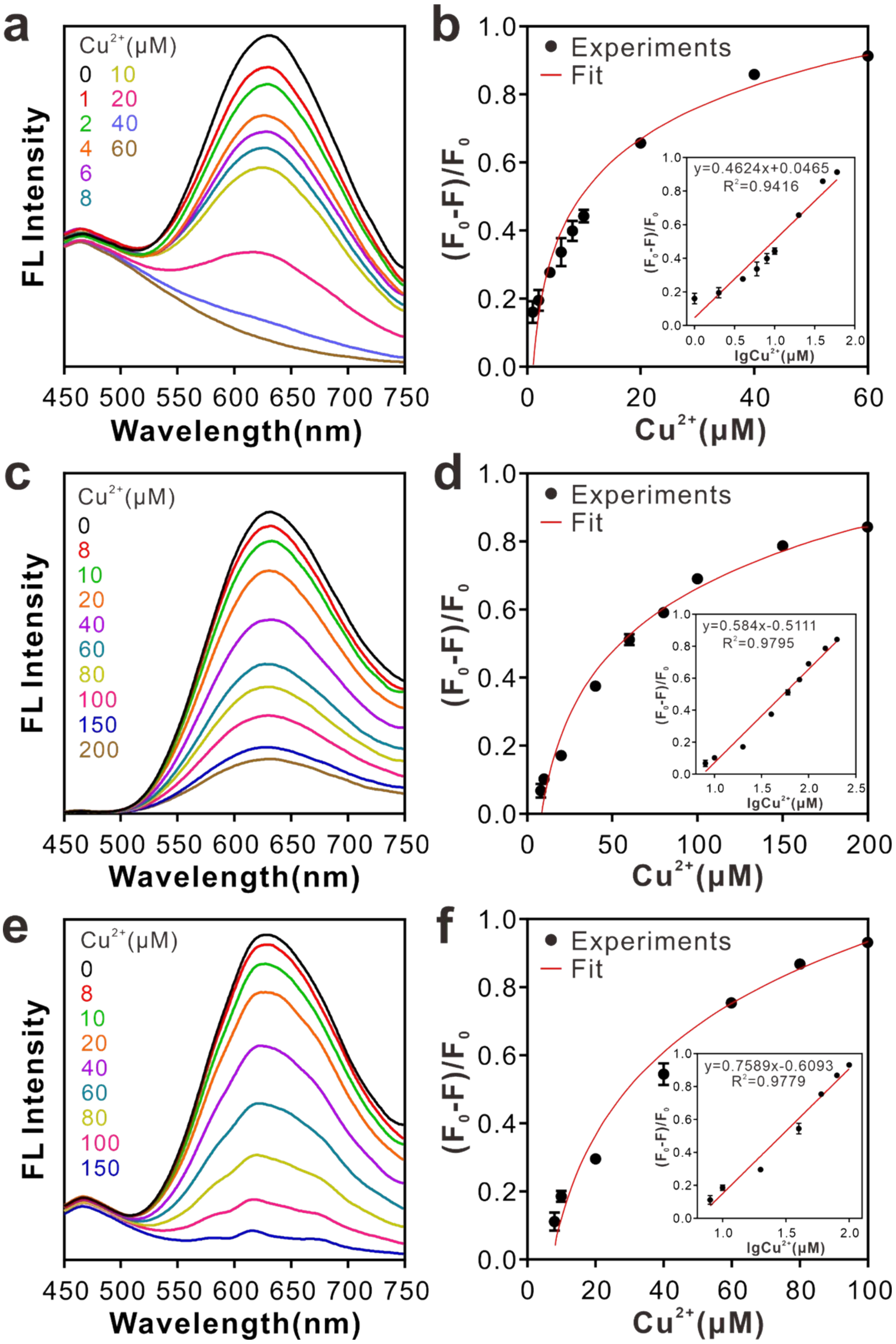

Fig S14. The detection performance of GSH-AuNCs with high fluorescence QY for Cu^{2+} in different body fluids. (a) Fluorescence spectra of AuNCs after adding AuNCs to the cerebrospinal fluid containing different concentrations of Cu^{2+}. (b) Curve fitting of the fluorescence intensity change rate ((F$_0$-F)/F$_0$) of AuNCs at 625 nm versus Cu^{2+} concentration. Inset: Linear fitting of the rate of change in fluorescence intensity ((F$_0$-F)/F$_0$). (c) Fluorescence spectra of AuNCs after adding AuNCs to saliva containing different concentrations of Cu^{2+}. (d) Curve fitting of the fluorescence intensity change rate ((F$_0$-F)/F$_0$) of AuNCs at 625 nm versus Cu^{2+} concentration. Inset: Linear fitting of the rate of change in fluorescence intensity ((F$_0$-F)/F$_0$). (e) Fluorescence spectra of AuNCs after adding AuNCs to sweat containing different concentrations of Cu^{2+}. (f) Curve fitting of the fluorescence intensity change rate ((F$_0$-F)/F$_0$) of AuNCs at 625 nm versus Cu^{2+} concentration. Inset: Linear fitting of the rate of change in fluorescence intensity ((F$_0$-F)/F$_0$).

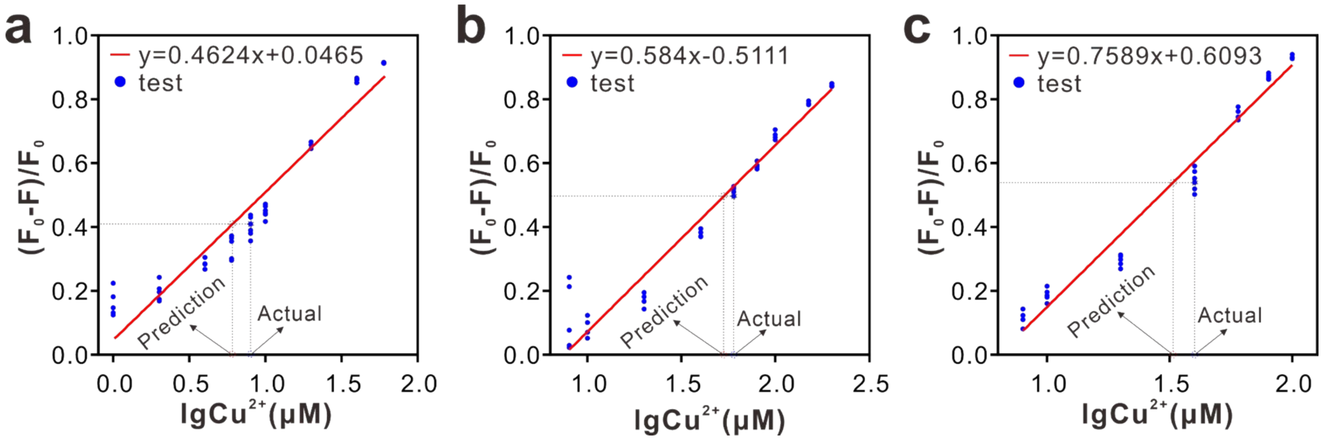

Fig S15. The predictive performance of the linear fit for bodily fluid samples containing unknown Cu^{2+} concentrations. (a) The predictive performance of the linear fit for cerebrospinal fluid samples containing unknown Cu^{2+} concentrations. (b) The predictive performance of the linear fit for saliva samples containing unknown Cu^{2+} concentrations. (c) The predictive performance of the linear fit for sweat samples containing unknown Cu^{2+} concentrations.

Table S1. Data description of the input parameters.

Parameter	Notation	Unit	Min	Max	Increment
Reaction temperature	T	°C	70	95	5
Thiol-to-metal molar ratio	$C_{GSH}/C_{Au}{}^{3+}$	-	1	2	0.1
Reaction time	t	h	1	24	1

Table S2. The input synthetic parameter variables of phase diagram.

Parameter	Notation	Unit	Min	Max	Increment
Reaction temperature	T	°C	70	95	5
Thiol-to-metal molar ratio	$C_{GSH}/C_{Au}{}^{3+}$	-	1	2	0.01
Reaction time	t	h	1	24	0.23