Electronic Supplementary Information

ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: classifying subtypes of human cancer

Georgios Theophilou[1, 2], Kássio M.G. Lima[1, 3], Pierre L. Martin-Hirsch[1, 2], Helen F. Stringfellow[2], Francis L. Martin[1*]

[1]Centre for Biophotonics, LEC, Lancaster University, Lancaster LA1 4YQ, UK; [2]Department of Obstetrics and Gynaecology, Central Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK; [3]Institute of Chemistry, Biological Chemistry and Chemometrics, Federal University of Rio Grande do Norte, Natal 59072-970, RN-Brazil

***Corresponding author:** email: f.martin@lancaster.ac.uk; phone: +44 (0)1524 510206; fax: +44 (0)1524 510217

No. of Pages = 17

No. of Tables = 5

No. of Figures = 4

Abbreviations:

- LGSC: low grade serous carcinoma
- HGSC: high grade serous carcinoma
- EC: Endometrioid carcinoma
- MC: Mucinous carcinoma
- MT: Mixed tumour
- CCC: Clear cell carcinoma
- CS: Carcinosarcoma
- DF: Discriminant function
- RMI: Risk malignancy index

Table S1. Risk Malignancy index (RMI): Women with ovarian cysts or vague abdominal symptoms undergo screening using the "Risk malignancy index". This predicts the risk of an ovarian mass being malignant and dictates further surgical or medical management.

Feature	**RMI 1**	**RMI 2**
Ultrasonic: • Bilateral lesions • Ascities • Multilocular cysts • Solid areas • Metastases	• No positive ultrasound features= 0 • 1 abnormality= 1 • 2 abnormalities= 2	• No positive ultrasound features= 0 • 1 abnormality= 1 • 2 abnormalities= 2
Premenopausal	1	1
Postmenopausal	3	4
Ca_{125}	U/ml	U/ml
RMI= Ultrasound score × Menopausal score × Ca_{125} in U/ML		

RMI	**Risk**	**Women (%)**	**Risk of cancer (%)**
<25	Low	40	<3
25-250	Moderate	30	20
>250	High	30	75

Table S2. Histopathological classification of ovarian epithelial tumours: Descriptive criteria for classification of ovarian carcinomas according to the World Health organization (WHO), 2003. Carcinosarcoma is not included in the main five categories due to its rarity.

Carcinoma subtype	Description
Serous	Composed of cells ranging in appearance from those resembling fallopian tube epithelium in well-differentiated tumours to anaplastic epithelial cells with severe nuclear atypia in poorly differentiated tumours
	Low grade / **High Grade**
	Uniform nuclei / 3-fold variability in nuclear size
	<13/10 high field powers mitotic figures / >13/10 high field powers mitotic figures
	Prominent nucleoli / Small nucleoli
	Differentiated architecture with papillary growth / Undifferentiated growth
	Numerous psammoma bodies / Few psammoma bodies
Mucinous	Resembles intestinal or endocervical epithelium
Endometrioid	Closely resembles the common variant of endometrioid carcinoma of the uterine corpus
Clear cell	Composed of glycogen-containing clear cells and hobnail cells and occasionally other histological types
Mixed surface	Composed of an admixture of two or more of the five major histological types, and the minor component(s) must comprise alone or together at least 10% of the tumour
Carcinosarcoma	Composed of both malignant epithelial and homologous (similar to Mullerian duct system) or heterologous (e.g. cartilage, bone, muscle) stromal elements

Table S3: Internal and external algorithm validation: 70% of the spectra were used to train the algorithm, 15% to test it internally and 15% to validate it externally.

	Normal	Borderline	Cancer	Total
Train	239 × 235	207 × 235	778 × 235	1224 × 235
Validation	55 × 235	45 × 235	165 × 235	265 × 235
Test	55 × 235	45 × 235	165 × 235	265 × 235

Table S4: Selected wavenumbers for SPA-LDA and GA-LDA. These wavenumbers were used to achieve classification of normal, borderline and malignant ovarian tissue.

Classification into normal, borderline and malignant ovaries	
Chemometric analysis	**Wavenumbers (cm^{-1}) selected**
SPA-LDA	900, 995, 1026,1068, 1111, 1165, 1230, 1377, 1404, 1446, 1462, 1512, 1543, 1554, 1562, 1604, 1620, 1643, 1658, 1681, 1747, 1800
GA-LDA	952, 983, 987, 1041, 1049, 1084, 1099, 1122, 1141, 1168, 1203, 1219, 1346, 1365, 1419, 1446, 1450, 1512, 1527, 1539, 1546, 1558, 1593, 1604, 1631, 1643, 1647, 1720, 759

Table S5: Selected wavenumbers for SPA-LDA and GA-LDA. These wavenumbers were used to achieve classification of ovarian carcinoma subtypes.

Ovarian carcinoma subtype classification	
Chemometric analysis	**Wavenumbers (cm^{-1}) selected**
SPA-LDA	964, 991, 1018, 1037, 1068, 1111, 1153, 1219, 1334, 1415, 1458, 1485, 1504, 1539, 1562, 1597, 1624, 1635, 1654, 1662, 1697, 1724, 1800
GA-LDA	902, 941, 964, 999, 1003, 1018, 1022, 1084, 1099, 1103, 1122, 1192, 1222, 1311, 1381, 1392, 1400, 1408, 1423, 1438, 1469, 1481, 1485, 1489, 1492, 1504, 1516, 1531, 1535, 1554, 1562, 1570, 1589, 1593, 1624, 1627, 1654, 1658, 1674, 1685, 1712, 1732, 1747, 1782

Figure S1: Optimization of Principal Component and Wavenumber selection for each of the analytical methods for paired comparison. This example uses the HGSC and LGSC classes.

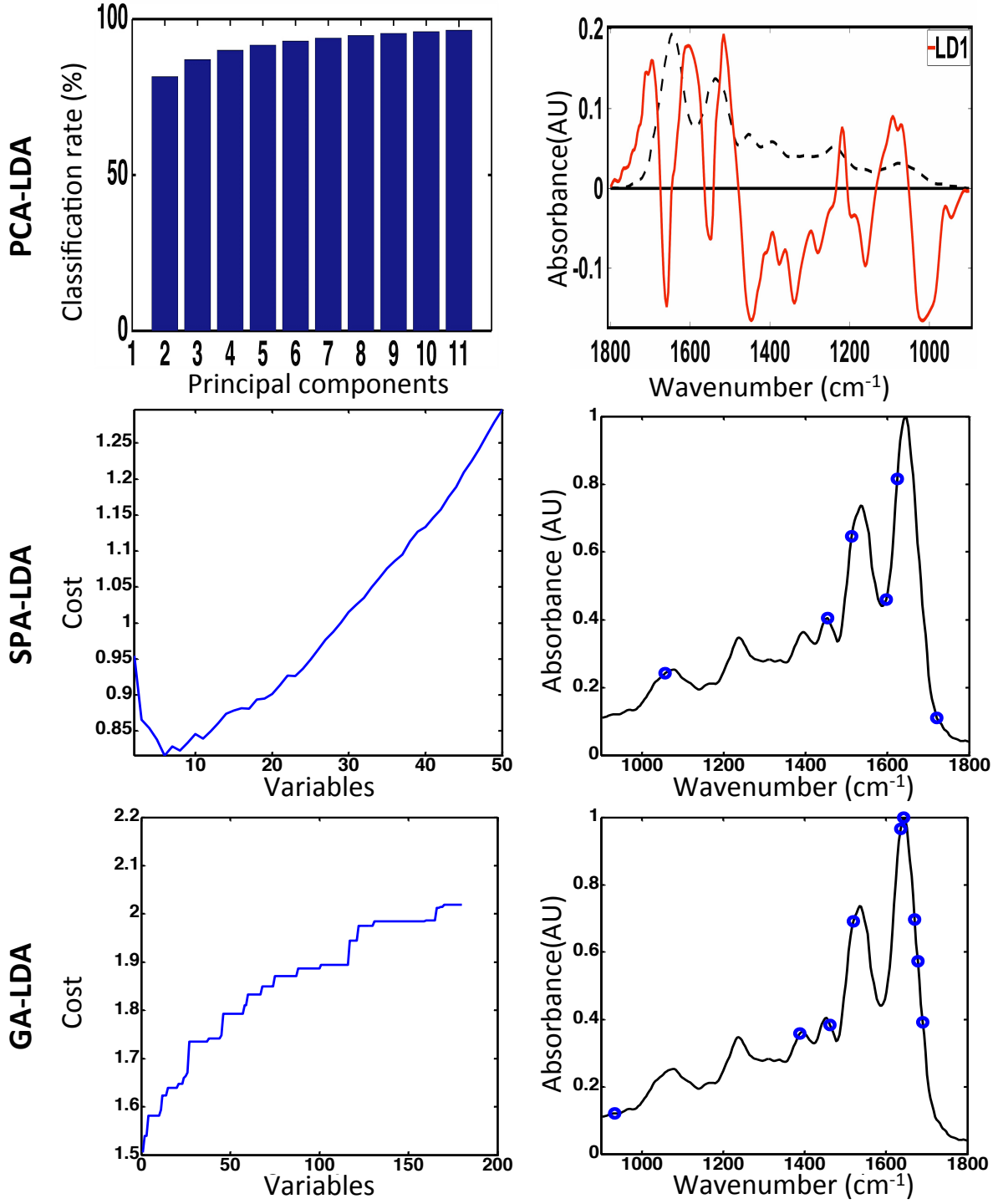

Figure S2: PCA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

Page: 1

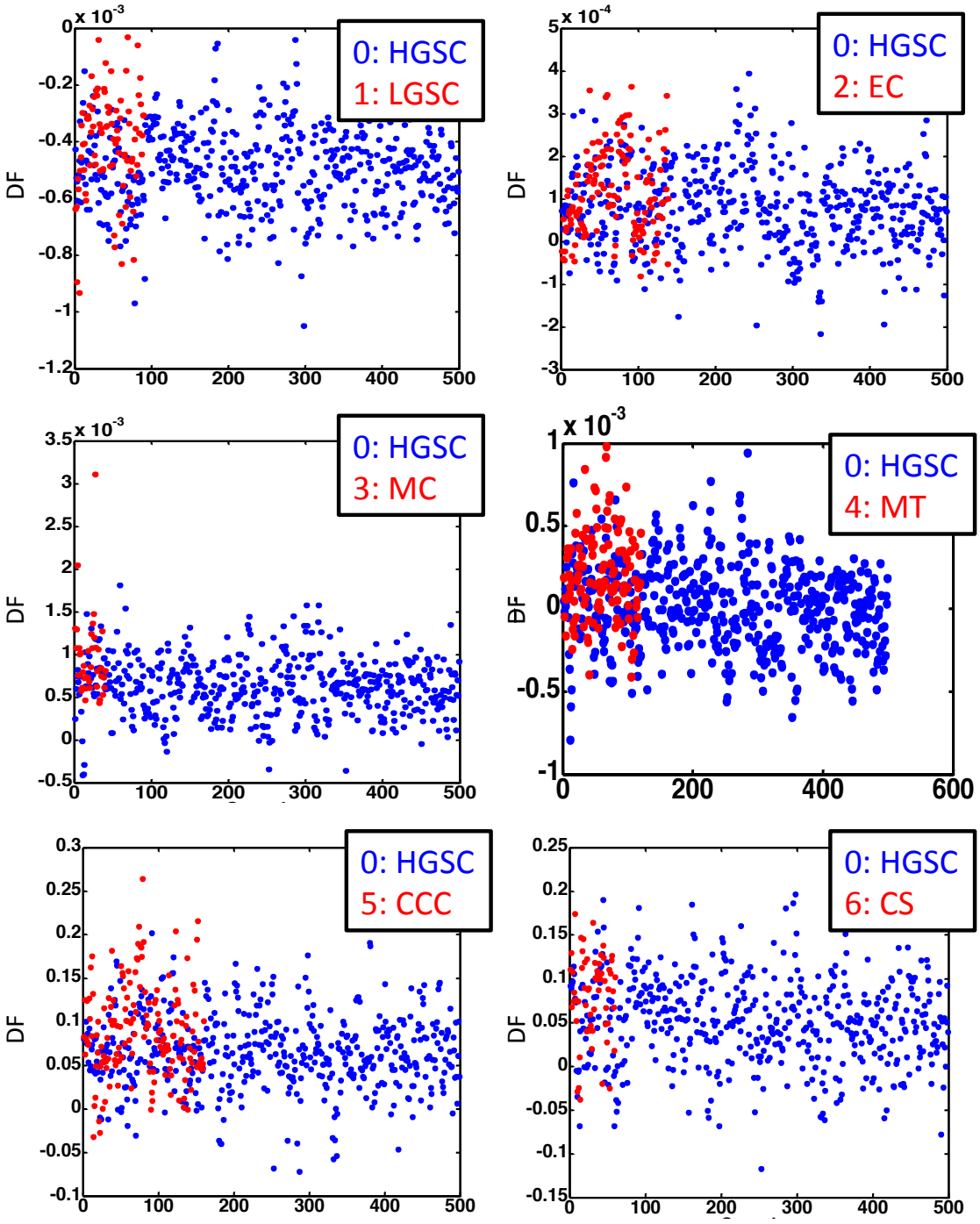

Figure S2: PCA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

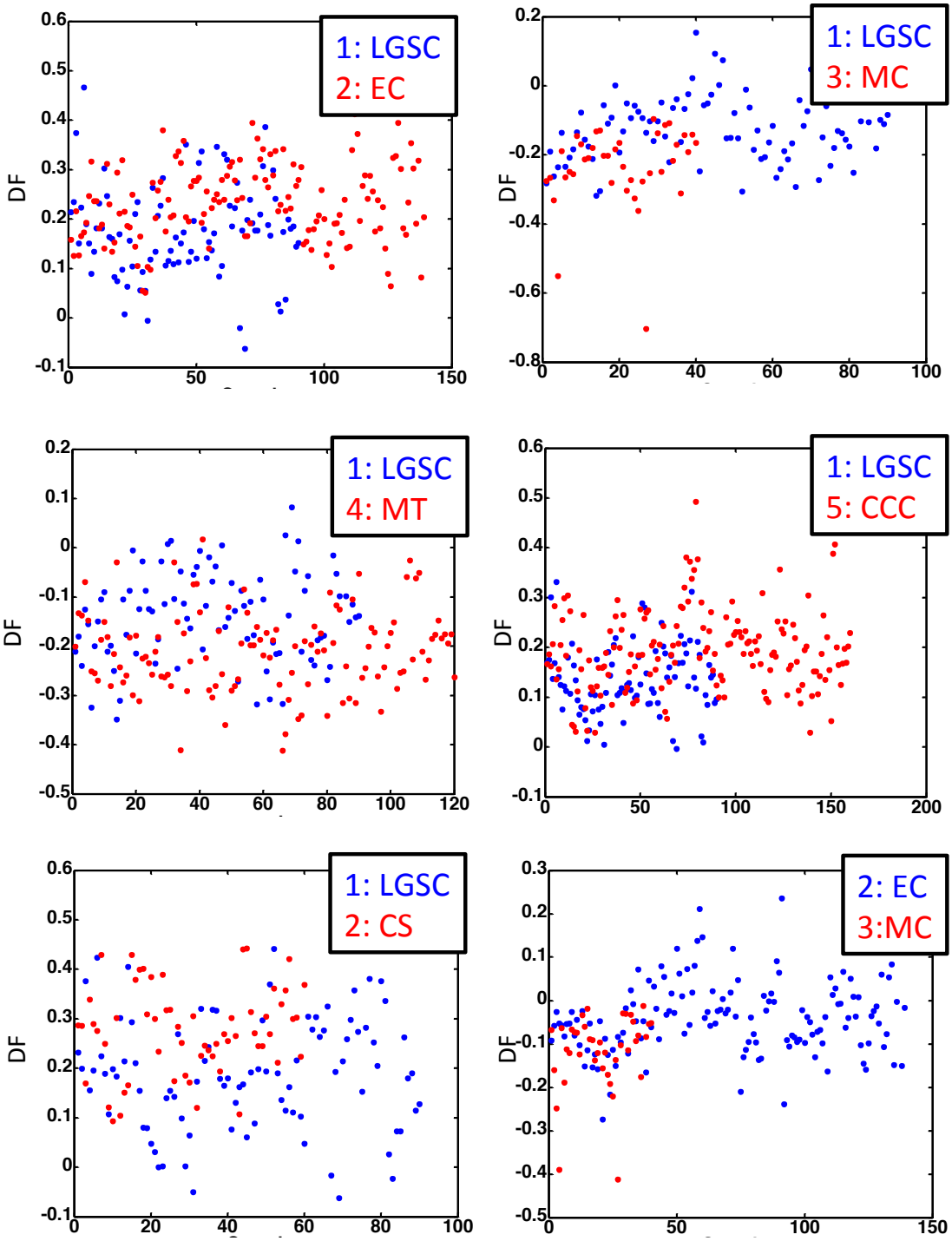

Figure S2: PCA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

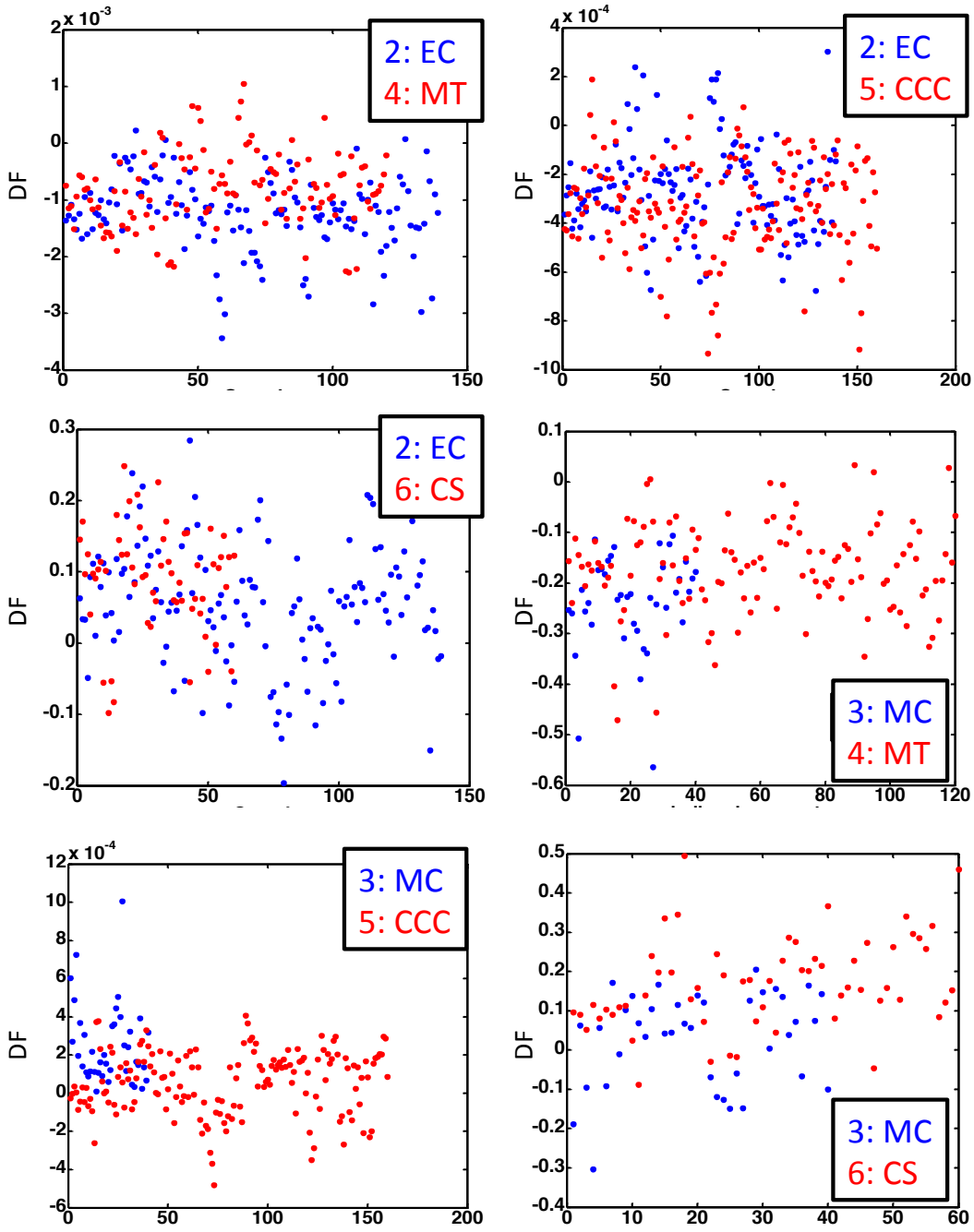

Figure S2: PCA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

Page: 4

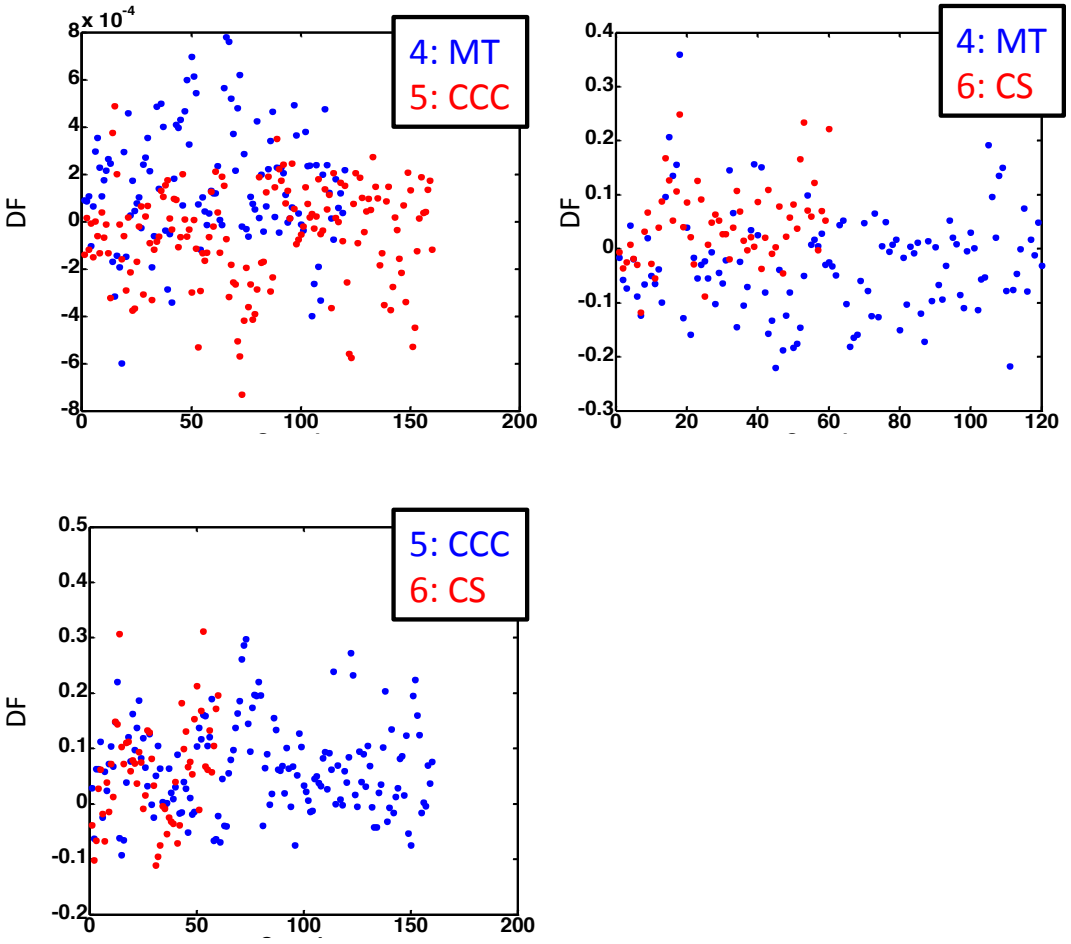

Figure S3: SPA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

Page: 1

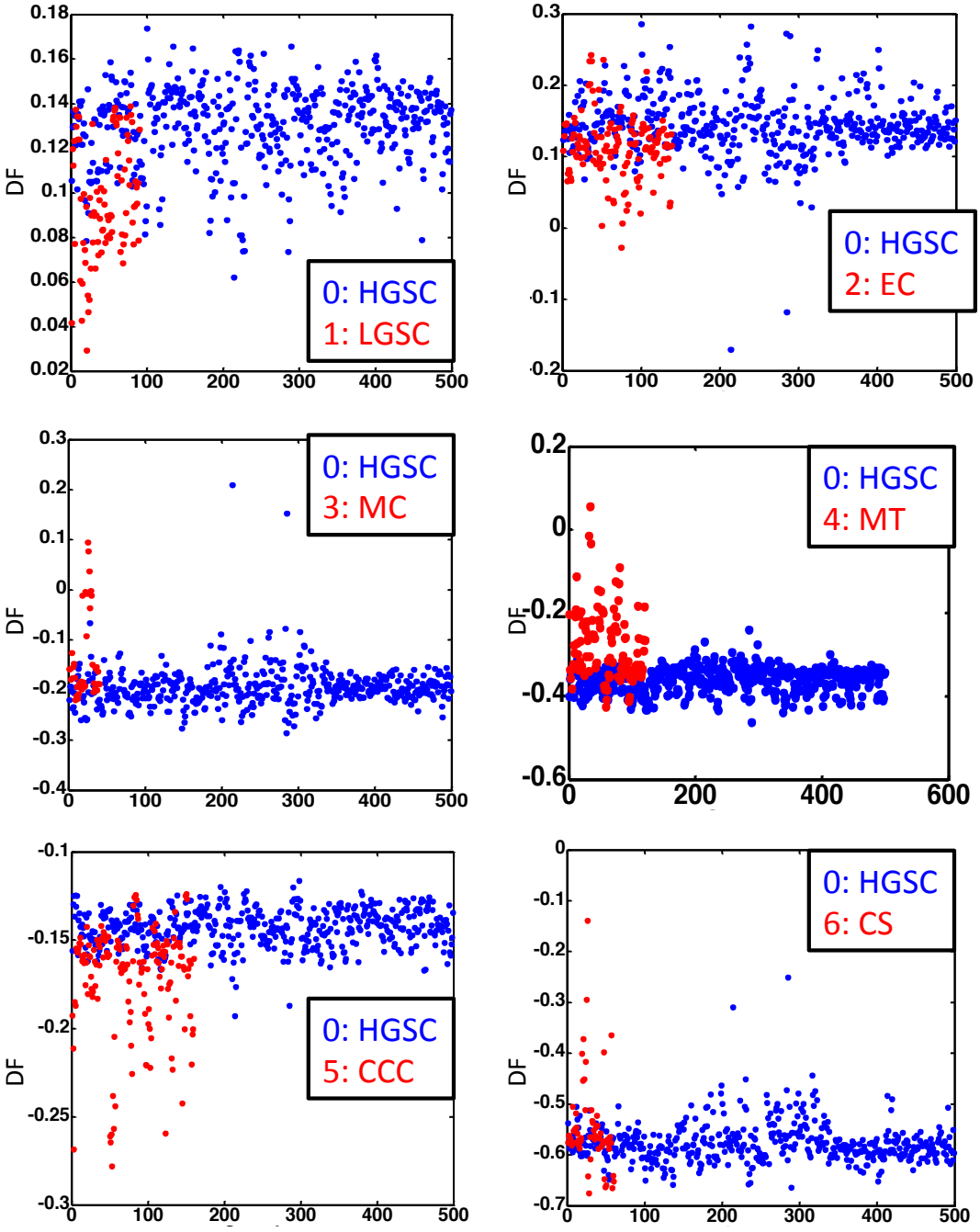

Figure S3: SPA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

Page: 2

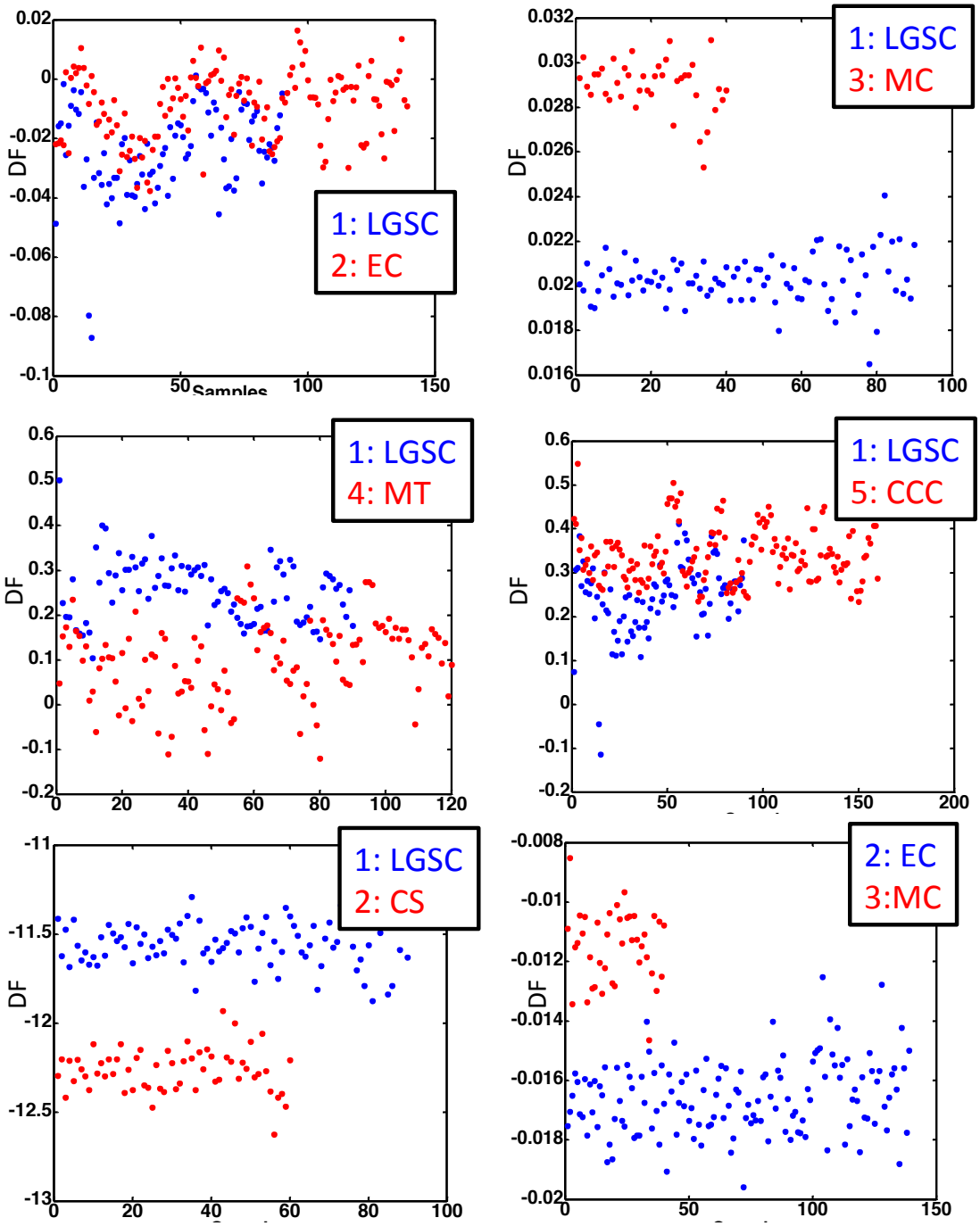

Figure S3: SPA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

Page: 3

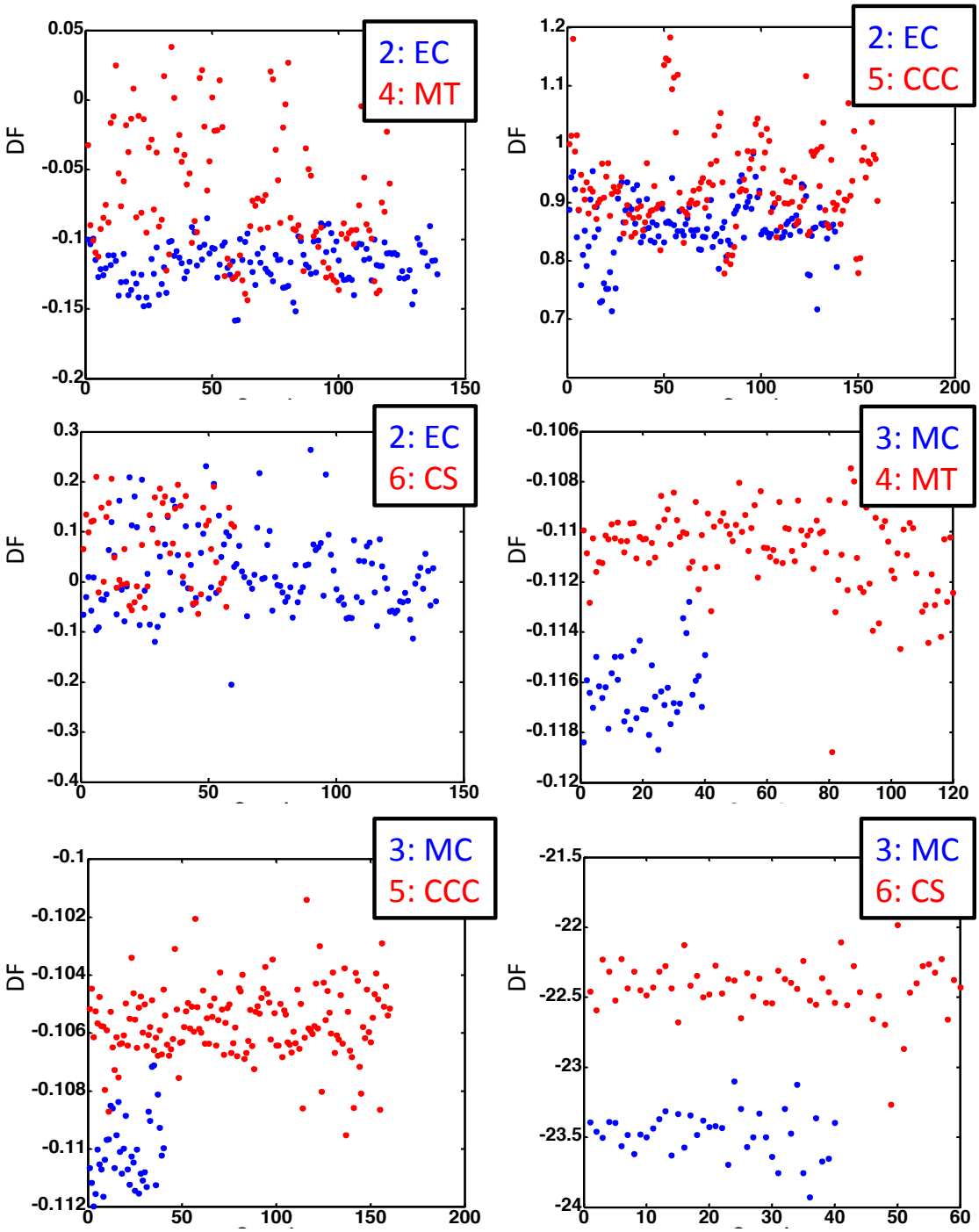

Figure S3: Dimensional **SPA-LDA** derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

Page: 4

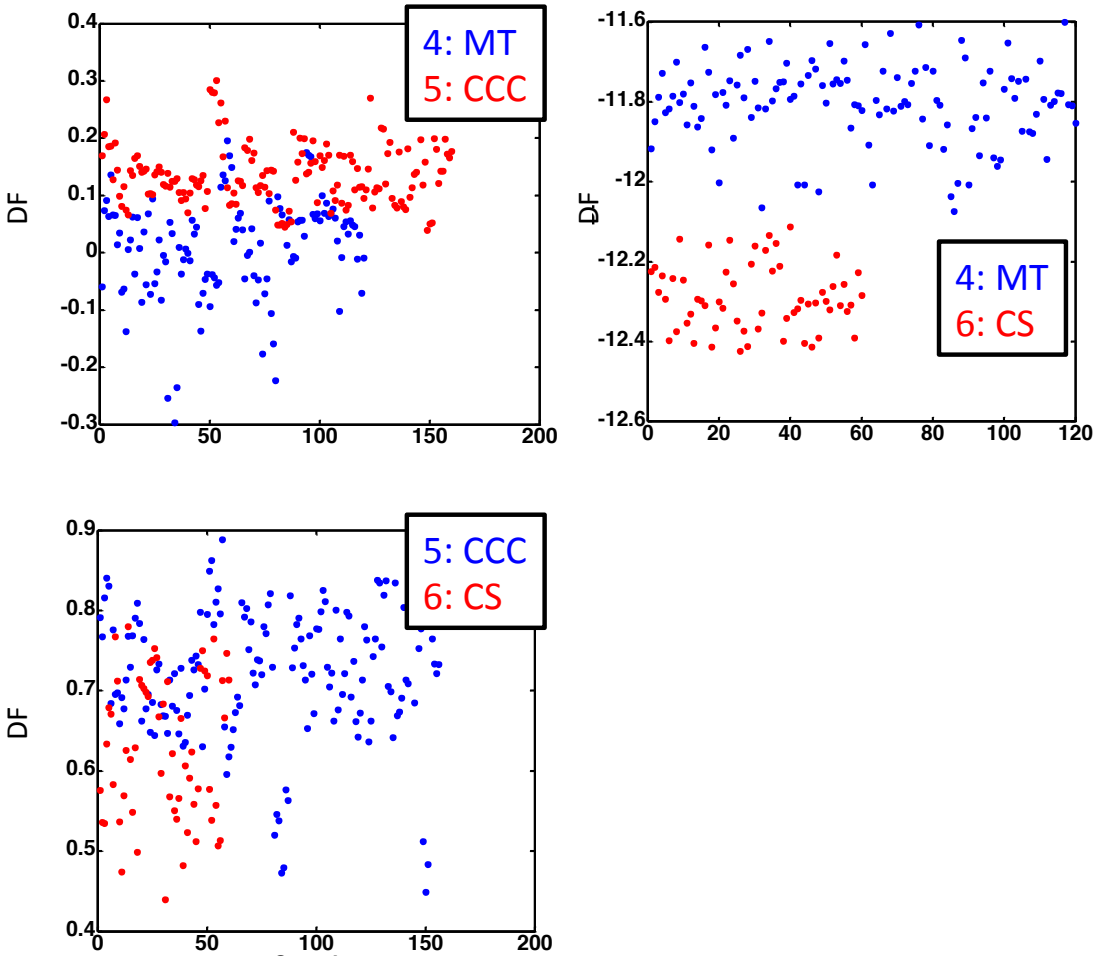

Figure S4: GA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

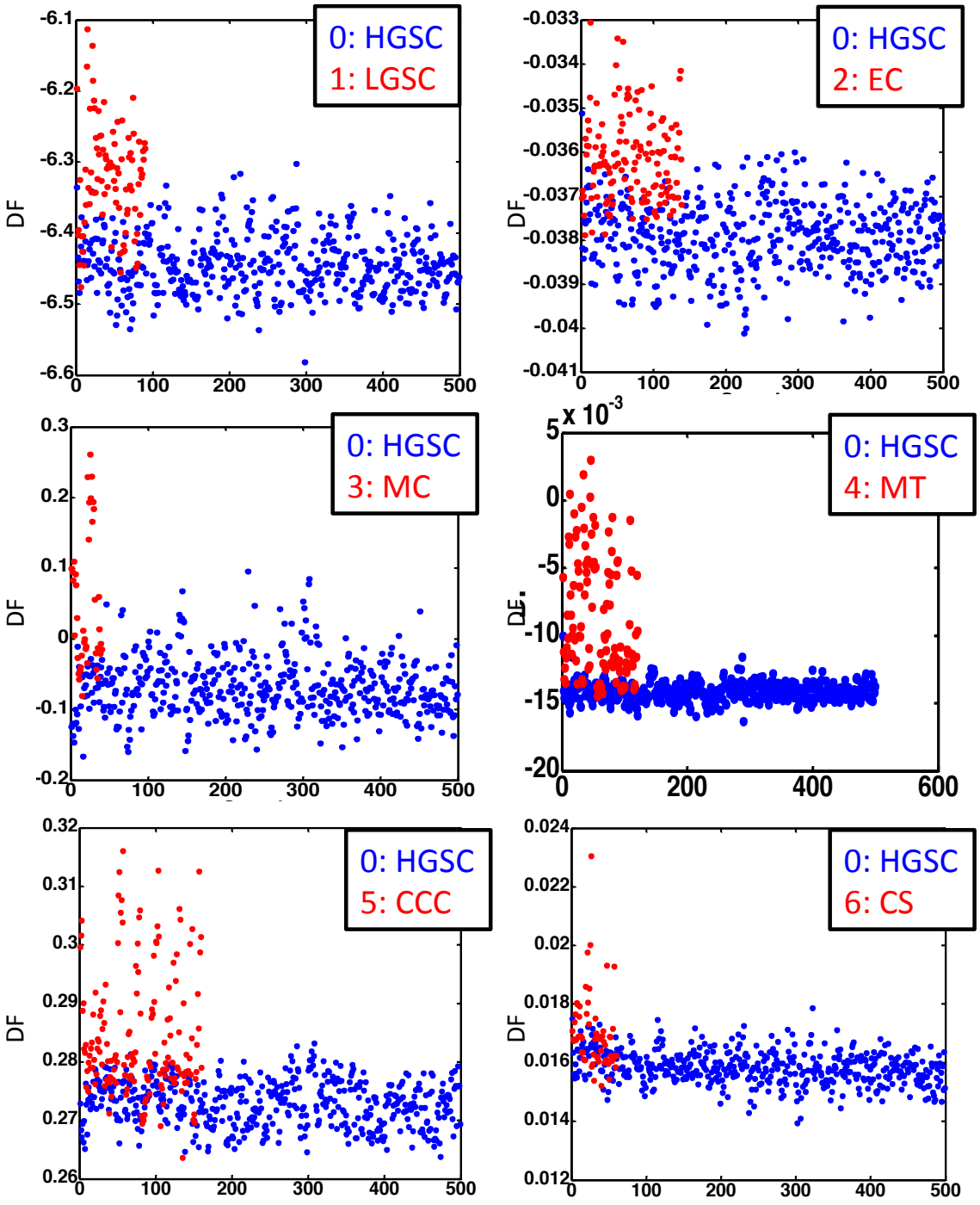

Figure S4: GA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

Page: 2

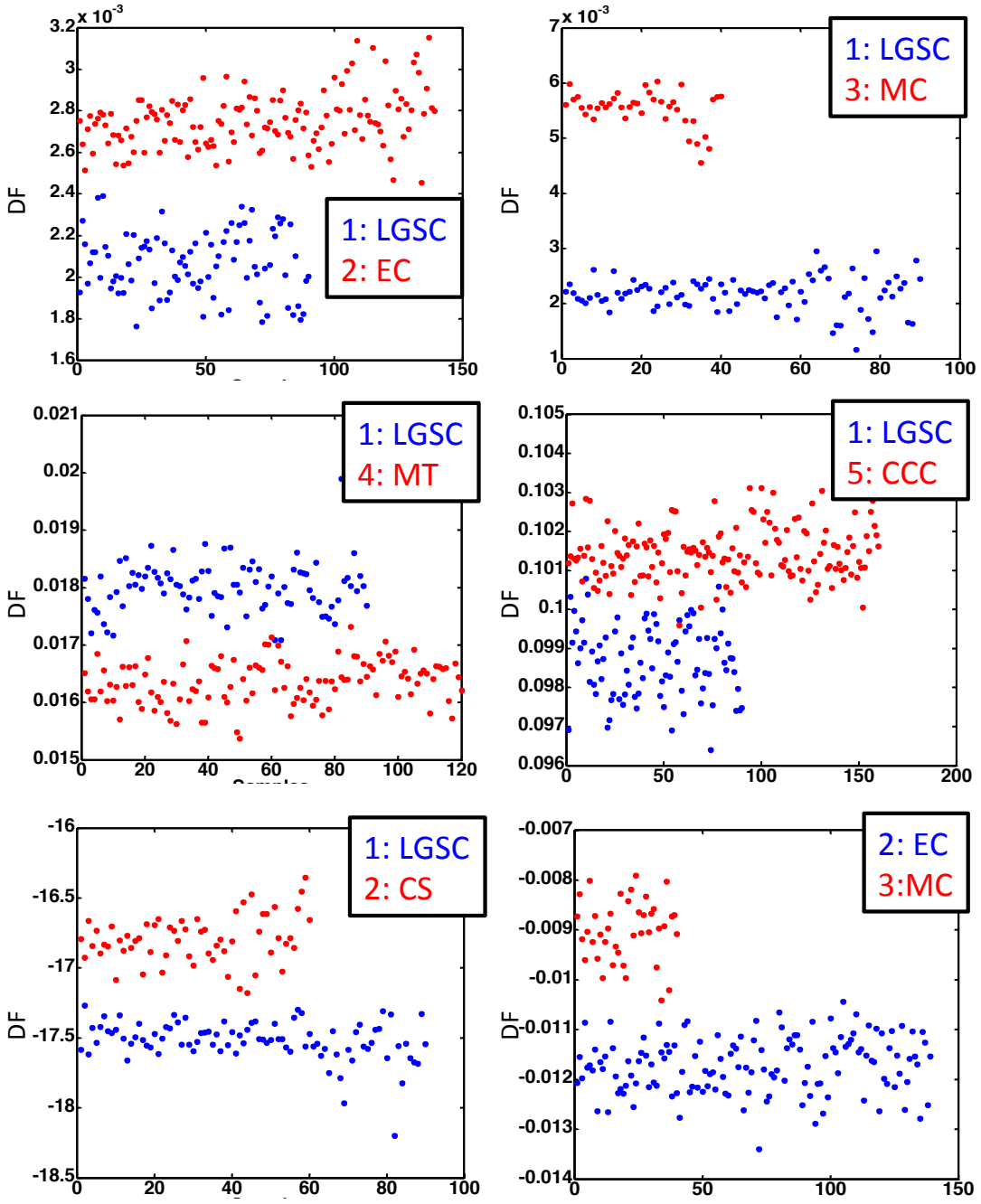

Figure S4: GA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

Page: 3

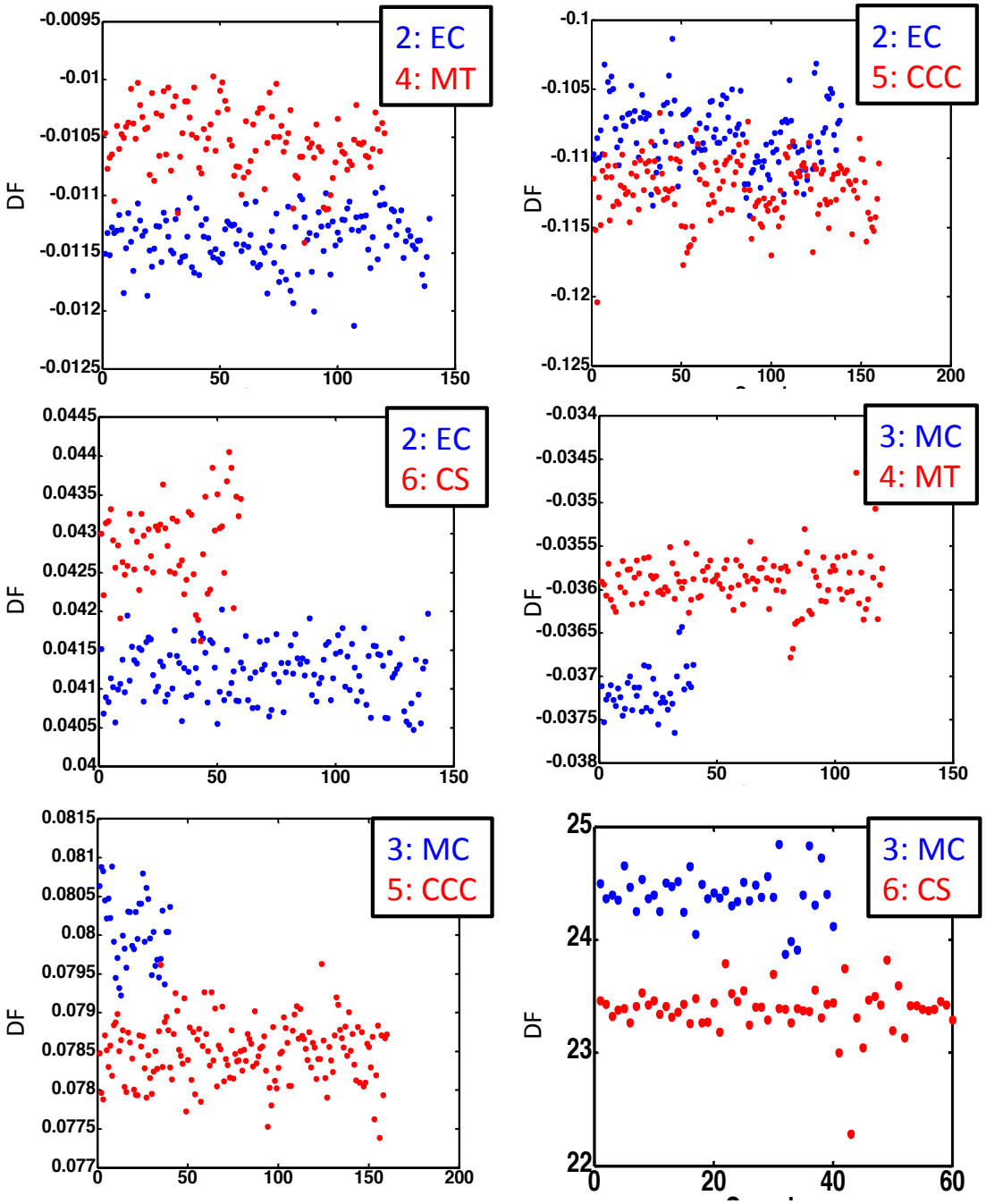

Figure S4: GA-LDA derived scores plots obtained from paired comparison of ovarian carcinoma subtypes.

Page: 4