Age-specific discrimination of blood plasma samples of healthy and ovarian cancer prone mice using laser-induced breakdown spectroscopy

Noureddine Melikechi[1,*], Yuri Markushin[1], Denise C. Connolly[2], and Jeremie Lasue[3], Ebo Ewusi-Annan[1] and Sokratis Makrogiannis[1]

[1]*Optical Science Center for Applied Research, Delaware State University 1200 N. Dupont Highway Dover, DE 19901, USA*
[2]*Fox Chase Cancer Center, Medical Science Division, 333 Cottman Avenue Philadelphia, PA 19111-2497, USA*
[3]*Université de Toulouse, UPS-OMP, IRAP, 9 Av. Colonel Roche, BP 44346, F-31028 Toulouse cedex 4, France*

[*]*nmelikechi@desu.edu*

Abstract: Epithelial ovarian cancer (EOC) mortality rates are strongly correlated with the stage at which it is diagnosed. Detection of EOC prior to its dissemination from the site of origin is known to significantly improve the patient outcome. However, there are currently no effective methods for early detection of the most common and lethal subtype of EOC. We sought to determine whether laser-induced breakdown spectroscopy (LIBS) and classification techniques such as linear discriminant analysis (LDA) and random forest (RF) could classify and differentiate blood plasma specimens from transgenic mice with ovarian carcinoma and wild type control mice. Herein we report results using this approach to distinguish blood plasma samples obtained from serially bled (at 8, 12, and 16 weeks) tumor-bearing TgMISIIR-TAg transgenic and wild type cancer-free littermate control mice. We have calculated the age-specific accuracy of classification using 18,000 laser-induced breakdown spectra of the blood plasma samples from tumor-bearing mice and wild type controls. **When the analysis is performed in the spectral range 250 nm to 680 nm using LDA, these are 76.7 ($\pm$2.6)%, 71.2 ($\pm$1.3)%, and 73.1 ($\pm$1.4)%, for the 8, 12 and 16 weeks. When the RF classifier is used, we obtain values of 78.5 ($\pm$2.3)%, 76.9 ($\pm$2.1)% and 75.4 ($\pm$2.0)% in the spectral range of 250 nm to 680 nm, and 81.0 ($\pm$1.8)%, 80.4 ($\pm$2.1)% and 79.6 ($\pm$3.5)% in 220 nm to 850 nm.** In addition, we report, the positive and negative predictive values of the classification of the two classes of blood plasma samples. The approach used in this study is rapid, requires only 5 μL of blood plasma, and is based on the use of unsupervised and widely accepted multivariate analysis algorithms. These findings suggest that LIBS and multivariate analysis may be a novel approach for detecting EOC.

Abbreviations: LIBS, Laser-Induced Breakdown Spectroscopy; EOC, Epithelial Ovarian Cancer; Linear Discrimination Analysis, Random Forest

Key Words: Laser-Induced Breakdown Spectroscopy, Epithelial Ovarian Cancer, Linear Discrimination Analysis, Random Forest, TgMISIIR-TAg DR-26 mouse model

I. Introduction

Despite intensive research efforts, cancer remains a lethal disease that affects millions of people around the world. In 2014, it accounted for the deaths of nearly 600,000 Americans, or 1,600 people per day [1, 2]. Sadly, despite many efforts and progress in cancer diagnosis and treatment, the human cost of this disease remains very high. A key challenge is to be able to detect and diagnose in its early stages of development. Early detection provides medical professionals much better possibilities for interventions to prevent or halt further progression, which would lead to a reduction in mortality and morbidity. Thus, for a large number of cancers, it is essential to develop and implement tests that provide signs of the disease prior to its spread from the site of origin. In addition, such tests would be even more useful if they are minimally invasive and can be made widely available.

Despite rich and diverse research efforts, early detection and diagnosis of epithelial ovarian cancer (EOC) remains a significant challenge [3, 4]. In the United States, approximately 22,000 women are diagnosed with EOC each year. This disease is the leading cause of death among gynecological cancers with ~14,000 deaths annually. Worldwide, it is estimated that there are 220,000 new cases of EOC and 140,000 deaths caused by the disease annually. EOC typically develops in the absence of specific symptoms; therefore, the majority (>70%) of cases are discovered when the disease is already at an advanced stage [5, 6]. The 5-year survival rate for patients diagnosed with advanced (stage III and IV) cancers is 30-40%. However, if EOC is diagnosed at an early stage, this rate ranges from 60% to 90%, depending on the degree of tumor differentiation [7, 8]. These statistics demonstrate that successful early detection and diagnosis of EOC would be particularly useful for improving prognosis and decreasing the mortality rate.

One of the important and promising strategies for earlier cancer diagnosis includes identification and detection of cancer-associated biomarkers through non- or minimally-invasive approaches. Such biomarkers would be ideally produced either by the tumor itself or by the surrounding tissues and can be found in a variety of biological fluids. Also important is the ability to yield specific and sensitive signatures accurate enough for the early detection of the disease. Changes in biomarkers present in biological fluids are particularly attractive because they can reveal the state of a cell. Cancer Antigen 125 (CA-125) is a protein widely used as a biomarker for detection of recurrent EOC [9, 10]. Recently, researchers reported the simultaneous detection of multiple candidate biomarkers for the early detection of EOC instead of a single one leads to increased specificity and sensitivity for cancer detection [11, 12] and the use of multicolor quantum dots and metal nanoparticles for tagging biomarkers for diagnostics of cancer [13, 14]. Recently, we reported on a technique, Tag-laser induced breakdown spectroscopy (Tag-LIBS), which relies on tagging CA-125 with metal microparticles in blood serum prior to performing LIBS measurements [15-17]. We have shown that this approach provides an estimated near single molecule per particle efficiency of CA-125 in human blood plasma and can, in principle, be generalized for the sensitive and simultaneous detection of multiple biomarkers in biomedical fluids and/or tissues.

Another approach that has attracted much attention is the use of classification techniques on spectroscopic data of biological samples acquired using methods such as surface-enhanced laser desorption and ionization (SELDI) [10], matrix-assisted laser desorption/ionization (MALDI) [18] and at the genomic level, DNA chip technology [19, 20].

LIBS has been used by a number of researchers to classify various samples [see for example, 22]. A number of review papers have been devoted to the theoretical, experimental and applications of the technique [23-26]. In this paper, we report on the first use of laser–induced breakdown spectroscopy (LIBS) for the blind classification of blood samples extracted from wild type and transgenic mice with ovarian tumors. The blood samples used for this study were collected from transgenic TgMISIIR-TAg mice in which transgene positive females develop bilateral epithelial ovarian cancer with metastatic spread to peritoneal organs and the formation of ascites [27]. The blood samples collected from wild type (transgene negative) female littermates were used as controls. Unlike Tag-LIBS, this approach does not require a specific sample preparation prior to performing LIBS measurements and classification.

II. Transgenic mice, sample preparation and experimental set-up

The C57BL/6 TgMISIIR-TAg transgenic mice used for this study have been described [8, 20, 28, 29]. Female C57BL/6 TgMISIIR-TAg transgenic mice develop bilateral ovarian tumors with 100% penetrance [28, 29]. Tumors develop with variable latency and female TgMISIIR-TAg transgenic mice survive an average of 152 days [28, 30]. The size of ovaries remains relatively normal (<10 mm^3) for the first 70–85 days of life, and undergoes a substantial increase in growth between 100 and 150 days of age, reaching a volume of about 50 mm^3 in a median time of 110 days. Although ovaries appear ostensibly normal in structure and size in young mice, postmortem examination of the ovaries shows the presence of small numbers of TAg+ tumor cells as early as 28 days of age [28, 29]. Although the size of the ovaries remains normal for the first 2-3 months, the early presence of tumor cells in one-month old mice suggests that potential cancer signatures may exist in the blood samples of young transgenic mice. For this reason, we collected blood from mice at 8, 12 and 16 weeks of age.

All of the blood specimens used in this study were collected from mice bred at Fox Chase Cancer Center under an Institutional Animal Care and Use Committee (IACUC) approved protocol. Male C57BL/6 TgMISIIR-TAg transgenic mice were bred with wild type C57BL/6 female mice to produce female C57BL/6 TgMISIIR-TAg transgenic and wild type littermate control mice. Prior to blood collection, mice were anesthetized with 2-3% isoflurane in O$_2$. Blood was successively collected from each mouse at 8, 12 and 16-weeks of age. Blood was collected from the retro-orbital sinus (three draws/mouse, alternating eyes) using heparinized Natelson blood collecting tubes and following collection, the blood collection tubes were centrifuged to compact cells. The tubes were cut just above the cellular component and the plasma is transferred to sterile micro-centrifuge tubes. Plasma specimens were stored at -80°C until all of the specimens were collected at which time they were shipped on dry ice for subsequent LIBS measurement and analysis.

A total of 56 mice (n=28 each, transgenic and wild type mice) were bred and bled at 8, 12 and 16 weeks of age yielding a total of 168 blood samples. Six specimens contained insufficient material, leaving 162 specimens for analysis in this study: 28 transgenic "cancer" and 26 wild type "non-cancer" blood plasma samples (Table 1). The blood plasma samples were divided into 3 groups according to age of mice at moment of blood collection: groups 1, 2 and 3 consisted of blood samples from 8, 12 and 16 week-old mice respectively. With the blood plasma samples grouped, we further created 6 sets of samples for each age group for analysis in the LIBS

chamber. Each set consisted of a total of 9 age-specific samples, 4 or 5 of which were selected from blood samples of wild-type mice and 5 or 4 (for a total of 9) from transgenic mice and 1 empty filter, i.e. with no blood deposit on it.

Prior to performing LIBS measurements, we prepared the surface on which the blood plasma samples were to be deposited by cutting the top part (about 21 mm) of the polyvinylidene fluoride (PVDF) filter (EMD-Millipore), we used a pipette (Fisherbrand Finnpipette II Adjustable-Volume Pipetters, Single Channel, 0.5-10 µL) to extract 5 µL from each blood plasma sample and deposited it on the top surface of the PVDF prepared filter. **These filters were selected because their porous structure and hydrophilicity allow for a more uniform spatial distribution of the mice blood plasma over a large surface area than other substrates we considered (e.g. glass).** Third, the filters were dried using a 40-Watts Tungsten lamp (UL certified portable lamp, model 1400, DA) for 5 minutes. This partially dries the liquid content of the blood plasma and increases the concentration of dissolved compounds of blood, minimizes the blood splashing caused by propagation of acoustic shock wave following plasma formation. To minimize contamination from the environment, these steps were performed in an isolated air-filtered preparation room. The sample preparation sample is illustrated in Figure 1.

The LIBS spectra of the various blood plasma samples were obtained using the previously described experimental set-up [31]. Essentially, it consists of a CPA-Series Ti-Sapphire 150-*fs* laser (Clark-MXR, Model 2210) operating at 775 nm with pulse energy about 1.54 mJ per pulse focused on the sample surface to generate a laser-induced micro plasma by using a fused silica Bi-Convex lens of 50 mm focal length. The focused laser spot diameter at the sample's surface was about 100 µm. A motorized stage with x-y translation was used to automatically translate the sample to a new location following each laser ablation. To reduce the effect from the surrounding atmosphere, measurements were performed in a chamber kept at a constant pressure of helium of 764 Torr. **Helium gas was selected over argon and air, because we find that breakdown in helium leads to more uniform plasma, compared to breakdown in air or argon.** The loading capacity for our set-up is 10 samples, which for this study consisted of 9 samples of blood plasma from mice of the same age group and of 1 empty filter. Optical emission from the plasma was collected by a fiber collimation lens at 45$^{\text{o}}$ with respect to the laser beam and focused onto a 50 μm core-diameter optical fiber. The other end of the fiber was coupled into an Echelle spectrograph, (Andor Technology, ME 5000). Dispersed light from the spectrograph was recorded with a thermoelectrically cooled iStar Intensified Charge Coupled Device (ICCD) camera (Andor Technology, DH734-18F O3) designed to operate in the wavelength region ranging from 190 to 975 nm in a single acquisition, with a spectral resolution power of 4000 corresponding to 4 pixels FWHM which corresponds to a spectral resolution of 0.013 to 0.056 nm respectively. *However, as the detection system was calibrated within the spectral range from 220 nm to 850 nm, only this range was considered in the analysis.* Emission from the plasma was collected 50 ns after the laser pulse with an integration time of 700 µs by an onboard digital delay generator of the spectrograph.

To ensure genotype-specific *blind* classification of the samples of each set, we randomized its 9 samples prior to performing a LIBS measurement. In other words, as the spectra were collected the information whether the blood plasma sample originated from a transgenic or a wild type mouse was not used. For this study, each LIBS spectrum was collected from a single laser pulse ablation of a single spot on the surface of the filter. Therefore, for each sample, we collected a

series of 100 consecutive LIBS spectra obtained from 100 non-overlapping spots of the filters with blood plasma samples and from the control samples. Thus, we analyzed a total of 180 (60 specimens per age group) including 18 empty filters as controls. A total of 18,000 LIBS emission spectra were collected for this study. These measurements took approximately 30 minutes per set of 10 samples for a total of 21 hours (for all 6 sets of the 3 age groups).

III. Processing of LIBS data

To reduce the impact of the presence of the filters on the analysis of the data, we collected 100 LIBS spectra from each empty filter used for a total of 6 per age group. These spectra were averaged and subtracted, data point-by-data point, from each of the remaining blood plus filter spectra of the set. This step provides a way to reduce the effect of the background signal related to the chemical composition of the surface of the filters used. **To improve precision and reproducibility and reduce statistical errors of the LIBS measurements per sample, total integrated spectra that did not fall within one standard deviation** about the mean integrated spectra for the 100 shots **were rejected. Each of the remaining spectra was normalized by its total integrated area.** In addition, we have taken great care to monitor and control the experimental parameters such as laser energy, spatial characteristics of the laser, and sample preparation process. In Figure 2, we show as an example a LIBS spectrum of blood plasma samples from 16 weeks old transgenic mice (Fig. 2 a) and a LIBS spectrum of empty filter (Fig. 2 b) (Spectral range 220- 850 nm).

Classification of the spectra from blood plasma from the two classes of mice was performed using Linear Discriminant Analysis (LDA) [21]. LDA is a supervised learning approach that assumes normal class-conditional distribution models to find the separating hyperplanes between two or more classes. Features are projected to linear vector subspaces and subsequently classified. To determine the class of an unknown sample, score values for the various classes are computed using the score functions and data features and the sample is assigned using maximum likelihood decision rules. In this work, wavelength dependent intensities of the spectra were not used directly as features. Instead, features were extracted from the raw spectra using the statistical dependency (SD) between features and associated class labels using a quantized feature space [32]. This was done to mitigate the contribution of non-discriminatory data points, reduce the dimensionality of the original dataset and avoid over fitting. We also tried the RELIEFF algorithm for feature selection but we adopted the SD approach because of the similarity in results obtained from the two methods [33]. All calculations were done using MATLAB [33]. We have performed our analysis over the spectral range 250 nm to 680 nm and in the range of 250 nm to 278 nm. The former wavelength range was limited to 680 nm because of the limiting computer power available. The latter range was selected because we noticed that the LIBS spectrum of an empty filter is weak compared to those of the blood plasma samples dominated by Fe emission lines (Figure 3). Although for this study, the identification of the emission lines is not necessary, we have performed it for few emission lines using the NIST table and the one previously developed by our research group [34].

Furthermore, we developed random forest-based classifiers to separate transgenic from wild type mice. RF is an ensemble learning method for classification. It constructs a set of decision trees $\{h_k(x; \Theta_k), k = 1, \ldots, N\}$ -where x is the input vector and N is the number of

trees- from subsets of the training set determined by random vectors Θ_k. RF also uses random feature selection for node splitting of each tree. These techniques decide the class after applying voting to the predicted classes by the individual trees for classification, or by calculating the mean prediction for regression. They address the overfitting tendency of the decision trees and have shown robustness with respect to noise [35]. In [35] it is shown that the generalization error asymptotically converges to an upper bound as the number of individual tree classifiers increases.

In this experiment we performed feature selection by first computing the information gain with respect to each class for each feature. Information gain is defined as

$$InfoGain(class, feature) = H(class) - H(class|feature) \quad (1)$$

where H is the entropy of each class given by $H(class) = -p(class) \cdot log p(class)$. The gain in information content is equal to the reduction of entropy between the two classes when we use a specific feature. After computing the information gain for each feature we rank all $InfoGain$ values and we select the top performing features by applying a threshold to these values.

IV. Results and analysis

To estimate the reproducibility of the spectra of the blood samples, we report, as an example, the shot-to-shot ratios of the intensities of the emission lines of the Na D lines (588.96 nm and 589.58 nm) for 100 single shots and for a set of averaged 100 spectra. These are measured to be about 10% and 2% respectively. For the empty filters, the shot-to-shot reproducibility measured by taking the ratios of the intensities of two emission lines of F, detected only in the empty filters at 623.96 nm and 634.88 nm, is found to be about 2% while the averaged over 100 shots better than 1%. These differences may be due to fluctuations related partly to the relative packing density of the samples [36].

The main goal of this work is to determine if LIBS spectra can be used to automatically identify the one of the two classes, transgenic (cancerous) or wild type (non-cancerous) mice, each blood sample originates from and quantify the accuracy of the result. Our approach does not require information on the origin (transgenic or wild type) of specific samples for the LIBS measurements or for the LDA or random forest analysis. To quantify our results, we define *classification accuracy* as the percent ratio of the number of correctly classified samples to the total number of analyzed blood samples within each age group, and *positive* and *negative* predictive value (PPV and NPV) as the percent ratios of the number of samples correctly classified as originating from transgenic mice and those originating from wild type mice respectively.

IV. 1. LDA Results and Analysis

Figures 4 and 5 shows the LDA scores for the LIBS spectra of the samples for 8- (Figs. 4a and 5a), 12 – (Figs. 4b and 5b) and 16 – (Figs. 4c and 5c) weeks in the 250 -278 nm and 250 nm – 680 nm spectral ranges respectively. Labeling of the data points was achieved subsequent to the generation of the LDA plots and was based on information on the origin of the samples provided

by Connolly's laboratory. Following the labeling step, the position of a decision line that separates the samples originating from the transgenic and the wild type mice **was determined automatically using the Fisher Discriminant approach which places the boundary in the direction that maximizes the separation between the groups and minimizes the variability within each group [37]. The solid black** lines shown represent decision boundaries separating the 2 classes of samples. **For each plot, LDA-1 is the score for the wild type and LDA-2 is the score for the transgenic classes.**

We further evaluated the classification performance measures of classification accuracy, positive predictive value (PPV), and negative predictive value (NPV) using leave-one-out cross-validation (CV) [38, 39] based on the LIBS data analyzed by LDA. The values obtained of the classification accuracy for the transgenic and the wild type mice samples are shown in Figure 6. The values obtained of the classification accuracy for the transgenic and the wild type mice samples are shown in Figure 6. This data includes the classification accuracy for the spectral ranges 250-680 nm (Figure 6a) and 220-850nm (Figure 6b) using leave-one-out cross-validation. **When the analysis is performed in the spectral range 250 nm to 680 nm, mean classification accuracy is 76.7 (±2.6)%, 71.2 (±1.3)%, %, and 73.1 (±1.4)%, for the 8, 12 and 16 weeks old mice respectively using LDA and become 76.4 (±2.9)%, 69.4 (±2.6)%, %, and 68.3 (±2.8)% for 220 to 850 nm. Our LDA analysis over the full range of data (up to 850 nm) did not improve the classification accuracy. We also measured the classifier performance in the range of 250-278 nm and obtained values of 62.9 (±1.2)%, 60.9 (±1.6)%, and 61.8 (±1.7)% respectively.** This spectral range was considered because, as shown in Figure 3, the LIBS spectra of the empty filters are relatively weak compared to those of the blood plasma samples.

IV. 2. Random Forest Results and Analysis

We evaluated the performance of random forest classifier after dimensionality reduction using the information gain measure. We report results from leave-one-out-cross-validation in Figure 7 by using the spectral range of 250-680nm (Figure 7a) and 220-850nm (Figure 7b) respectively. In 250-680nm this technique produces classification accuracy of 78.5 (±2.3)%, 76.9 (±2.1)% and 75.4 (±2.0)% for 8, 12 and 16 weeks old mice respectively. The corresponding performance rates for the range of 220-850nm are 81.0 (±1.8)%, 80.4 (±2.1)% and 79.6 (±3.5). Finally, the classification accuracy in 250-278nm was 65.6 (±1.6)%, 64.8 (±1.8)% and 66.2 (±1.8)% for 8, 12 and 16 weeks old mice. In Figure 7 we observe that the PPV and NPV values follow similar trends with the classification accuracy. The improved performance of random forest classification may be because random forest can successfully approximate nonlinear decision functions in contrast to the LDA that searches for linear decision hyperplanes in multi-dimensional vector spaces.

IV. 3. LDA and RF classification of the larger group formed by the six sets

Furthermore, for each age group we have attempted to separate the cancerous from the non-cancerous populations using the larger datasets formed by all six sets. For the spectral range from 250 to 680 nm, with LDA using leave-one-out cross-validation, the classification accuracy is found to be 67 %, 63 %, and 62 % for the 8, 12, and 16 weeks old respectively. For the same

spectral range, using RF, these values become 56.5 %, 67.8 %, and 70.9%. This shows that when considering all six sets of the same age group, using RF the classification accuracy improves slightly with age; that is specimens of the oldest age group (16 weeks) is more accurately classified than specimens from the younger age groups. This observation supports the idea that plasma specimens will exhibit greater difference from control specimens as tumor burden increases in the animals. However, when all six sets are considered as a single one, the classification accuracy for each age group decreases. We suspect that this decrease to be caused by the fact that after collecting LIBS spectra from one blood plasma set, the LIBS chamber had to be re-opened to load the new batch of blood plasma samples.

III. Conclusions and Perspectives

The use of multivariate analysis to classify LIBS spectra of various samples is known to be a powerful and useful strategy. Using a transgenic mouse model that predictably develops EOC, we have extended this approach to automatically classify blood plasma samples from wild type and transgenic mice **using random forest analysis** and found that using the spectral data around in the range of 220 to 680 nm, the accuracies obtained for all three age groups (8, 12 and 16 weeks old mice) are **76.7 (±2.6)%, 71.2 (±1.3)% and 73.1 (±1.4%)**. **These increase to 81.0 (±1.8)%, 80.4 (±2.1)% and 79.6 (±3.5)% using RF in 220 nm to 850 nm.** Additional work remains to be done to identify the specific atomic and/or ionic lines that contribute to the separation. Nevertheless, the strategy used appears to show considerable promise.

The classification accuracies reported in this work may potentially be improved by using multivariate classification tools other than LDA and RF. However, we suspect that it is critical that more blood samples be interrogated by LIBS at the same time (larger sets) irrespective of age of the mice. We are attempting to enhance the separation of the cancerous and non-cancerous populations using an improved experimental chamber and a variety of multivariate techniques for dimensionality reduction and classification. Such techniques include PLS-DA, to neural networks, Support Vector Machines or other ensemble classification techniques (see for example: [40]), with training set cross-validation and random test set model validation. An additional interesting perspective is to proceed to physical interpretation of the selected set of wavelengths that yields the best classification results.

Recent studies have shown that epithelial ovarian cancer development in mice is similar to that in humans [27, 41]. Due to the lack of availability of matched blood specimens from patients with early and late stage disease, genetically engineered mouse models of epithelial ovarian cancer are invaluable tools to evaluate novel methodologies for early disease collection. Moreover, such models are routinely used to collect useful information about factors involved in malignant transformation, invasion and metastasis, and the response to therapy [28, 42-44]. Like humans, mice with ovarian tumors exhibit few or no symptoms until the disease progresses to advanced stage [41]. Therefore, the work and conclusions obtained in this work can be very useful and have the potential to lead to development of analogous methods applicable to early stage cancer diagnostic in humans.

References

[1] American Cancer Society (2014). Cancer facts & figures 2014. American Cancer Society, Atlanta.

[2] R.L. Siegel, K.D. Miller, A. Jemal. Cancer statistics, 2015. CA Cancer J Clin. (2015), 65, 5-29.

[3] S. Brandner, J. Müller-Nordhorn, W. Stritter, C. Fotopoulou, J. Sehouli, C. Holmberg, *Soc. Sci. Med.* (2014) *119*, 123–130.

[4] K. E. Brain, S. Smits, A. E. Simon, L. J. Forbes, C. Roberts, et al. Ovarian cancer symptom awareness and anticipated delayed presentation in a population sample. *BMC Cancer* (2014) *14*, 171.

[5] M.L. Hensley, M. Castiel, M.E. Robson. Screening for ovarian cancer: what we know, what we need to know. Oncology (Huntingt). 2000, 14, 1601–1616.

[6] C.H. Holschneider, J.S. Berek. Ovarian cancer: epidemiology, biology, and prognostic factors. Semin Surg Oncol. (2000) 19, 3–10.

[7] V. McGuire, A.S. Whittemore, R. Norris, I. Oakley-Girvan. Survival in epithelial ovarian cancer patients with prior breast cancer. American Journal of Epidemiology, (2000) 152, 528-532.

[8] D.C. Connolly, R. Bao, A.Y. Nikitin, K.C. Stephens, T.W. Poole, X. Hua, S.S. Harris, B.C. Vanderhyden, T.C. Hamilton. Female mice chimeric for expression of the simian virus 40 TAg under control of the MISIIR promoter develop epithelial ovarian cancer. Cancer Res (2003) 63, 1389-1397.

[9] L. A. Beer, H. Wang, Hsin-Yao Tang, Z. Cao, T. Chang-Wong, J L. Tanyi, R. Zhang, Q. Liu, and D. W. Speicher; Identification of Multiple Novel Protein Biomarkers Shed by Human Serous Ovarian Tumors into the Blood of Immunocompromised Mice and Verified in Patient Sera, PLoS One. 2013; 8(3): e60129. doi: 10.1371/journal.pone.0060129. Epub 2013 Mar 27. PMID: 23544127

[10] W. Zhu, X. Wang, Y. Ma, M. Rao, J. Glimm, J.S. Kovach. Detection of cancer specific markers amid massive mass spectral data. PNAS, (2003), 100, 14666–14671.

[11] G. Mor, I. Visintin, Y. Lai, H. Zhao, P. Schwartzet et al. Serum protein markers for early detection of ovarian cancer. PNAS, (2005) 102, 7677–7682.

[12] K. Rubenstein. Multiplex Assays: Evolving Technologies, Applications and Future Directions. In: Insight Pharma Reports, 07.13.2010. Available fromhttp://www.insightpharmareports.com/uploadedFiles/Reports/Reports/Multiplex_Assays/Sample_Pages.pdf

[13] A.M. Smith, S. Dave, S. Nie, L. True, X. Gao. Multicolor quantum dots for molecular diagnostics of cancer. Expert Review of Molecular Diagnostics, (2006), 6, 231-244.

[14] R.G. Freeman, P.A. Raju, S.M. Norton, I.D Walton, P.C. Smith, et al Use of Nanobarcodes Particles in Bioassays. In: Methods in Molecular Biology, S.J. Rosenthal, D.W. Wright (Eds.), (2005) 73-84, Humana Press Inc, Totowa

[15] Y. Markushin, P. Sivakumar, D. Connolly, N. Melikechi. Tag-femtosecond laser-induced breakdown spectroscopy for the sensitive detection of cancer antigen 125 in blood plasma. Anal Bioanal Chem., (2014) 407, 1849-1855.

[16] S. C. Jantzi, V. Motto-Ros, F. Trichard, Y. Markushin, N. Melikechi, A. De Giacomo. Sample treatment and preparation for laser-induced breakdown spectroscopy. Spectrochimica Acta Part B, (2016) 115, 52–63.

[17] Y. Markushin and N. Melikechi. Sensitive Detection of Epithelial Ovarian Cancer Biomarkers Using Tag-Laser Induced Breakdown Spectroscopy, in: Ovarian Cancer - Basic Science Perspective, Edited by S. A. Farghaly, InTech, (2012) Chapter 9, 153 – 170.

[18] I-C. Lu, C. Lee, Y.-T. Lee, and C.-K. Ni. Ionization Mechanism of Matrix-Assisted Laser Desorption/Ionization. Annual Review of Analytical Chemistry. (2015) 8, 21 -39

[19] M. J. Heller. DNA Microarray Technology: Devices, Systems, and Applications. Annual Review of Biomedical Engineering. (2002) 4, 129-153. DOI: 10.1146/annurev.bioeng.4.020702.153438

[20] D.C. Connolly, H.H. Hensley. Xenograft and Transgenic Mouse Models of Epithelial Ovarian Cancer and Non Invasive Imaging Modalities to Monitor Ovarian Tumor Growth In situ -Applications in Evaluating Novel Therapeutic Agents. Curr Protoc Pharmacol (2009) 45, 14.12.1-14.12.26.

[21] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York (2009), 106p.

[22] V. K. Unnikrishnan, K. S. Choudhari, S. D. Kulkarni, R. Nayak, V. B. Kartha, C. Santhosh. Analytical predictive capabilities of Laser Induced Breakdown Spectroscopy (LIBS) with Principal Component Analysis (PCA) for plastic classification. RSC Adv., (2013) 3, 25872-25880.

[23] M Capitelli, A. Casavola, G. Colonna, A. De Giacomo, Laser-induced plasma expansion: theoretical and experimental aspects, Spectrochimica Acta Part B (2004) 59, 271-289.

[24] R. Noll, Laser-Induced Breakdown Spectroscopy, Springer-Verlag, Heidelberg (2012), 544 p.

[25] G. Cristoforetti, E. Tognoni, L.A. Gizzi. Thermodynamic equilibrium states in laser-induced plasmas: From the general case to laser-induced breakdown spectroscopy plasmas, Spectrochimica Acta B (2013) 90, 1-22.

[26] L. Radziemski, D. Cremers. A brief history of laser-induced breakdown spectroscopy: From the concept of atoms to LIBS 2012, Spectrochimica Acta B (2013) 87, 3-10.

[27] D.C. Connolly, R. Bao, A.Y. Nikitin, K.C. Stephens, T.W. Poole, et al. Female Mice Chimeric for Expression of the Simian Virus 40 Tag under Control of the MISIIR Promoter Develop Epithelial Ovarian Cancer. Cancer Research (2003) 63, 1389-1397.

[28] H. Hensley, B.A. Quinn, R.L. Wolf, S.L. Litwin, S. Mabuchi, S.J. Williams, C. Williams, T.C. Hamilton, D.C. Connolly. Magnetic Resonance Imaging for Detection and Determination of Tumor Volume in a Genetically Engineered Mouse Model of Ovarian Cancer. Cancer Biol Ther (2007) 6, 1717-1725.

[29] B.A. Quinn, F. Xiao, L. Bickel, L. Martin, X. Hua, A. Klein-Szanto, D.C. Connolly. Development of a syngeneic mouse model of epithelial ovarian cancer. J Ovarian Res (2010) 3:24, 1-17.

[30] H. Hensley, B.A. Quinn, R.L. Wolf, S.L. Litwin, S. Mabuchi, et al. Magnetic resonance imaging for detection and determination of tumor volume in a genetically engineered mouse model of ovarian cancer. Cancer Biol Ther. (2007) 6, 1717-1725.

[31] P. Sivakumar, L. Taleh, Y. Markushin, N. Melikechi, J. Lasue. An experimental observation of the different behavior of ionic and neutral lines of iron as a function of number density in a binary carbon–iron mixture. Spectrochimica Acta Part B Atomic Spectrosc. (2013) 82, 76–82.

[32] J. Pohjalainen, O. Rasanen & S. Kadioglu: Feature Selection Methods and Their Combinations in High-Dimensional Classification of Speaker Likability, Intelligibility and Personality Traits, Computer Speech & Language (2013) 29, 1-21.

[33] MATLAB version 9.0.0. Natick, Massachusetts: The MathWorks Inc., 2016.

[34] S. Rock, A. Marcano, Y. Markushin, C. Sabanayagam, and N. Melikechi. "Elemental Analysis of Laser Induced Breakdown Spectroscopy Aided by an Empirical Spectral Database", Applied Optics (2008) 47, G99-G104.

[35] L. Breiman (2001), "Random Forests", *Machine Learning*, vol. 45, no. 1, 2001, pp. 5-32.

[36] P. Sivakumar, L. Taleh, Y. Markushin, N. Melikechi. Packing density effects on the fluctuations of the emission lines in laser-induced breakdown spectroscopy. Spectrochimica Acta Part B Atomic Spectrosc. (2014) 92, 84–89.

[37] C. Bishop, Pattern Recognition and Machine Learning, Springer, New York (2009), 186p.

[38] P Eusebi, Diagnostic Accuracy Measures, Cerebrovasc Dis (2013) 36, 267–272.

[39] S. Banoo, D. Bell, P. Bossuyt, A. Herring, D. Mabey, et al. Evaluation of diagnostic tests for infectious diseases: general principles. Nature Reviews Microbiology (2010) S16-S28.

[40] A.J. Izenman. Modern multivariate statistical techniques. New York: Springer, 2008.

[41] B. A. Quinn, F. Xiao, L. Bickel, L. Martin, X. Hua, A. Klein-Szanto, D. C. Connolly. Development of a syngeneic mouse model of epithelial ovarian cancer. J Ovarian Res. (2010) 3, 24. doi: 10.1186/1757-2215-3-24

[42] G. Gritsina, F. Xiao, S.W. O'Brien, R. Gabbasov, M.A.Maglaty, R.H. Xu, R.J. Thapa, Y. Zhou, E. Nicolas, S. Litwin, S. Balachandran, L.J. Sigal, D. Huszar, D.C. Connolly. Targeted Blockade of JAK/STAT3 Signaling Inhibits Ovarian Carcinoma Growth. Mol Cancer Ther (2015) 14, 1035-1047.

[43] H.H. Hensley, N.A Roder, S.W. O'Brien, L.E. Bickel, F. Xiao, S. Litwin, D.C. Connolly. Combined in vivo molecular and anatomic imaging for detection of ovarian carcinoma-associated protease activity and integrin expression in mice. Neoplasia (2012) 14, 451-462.

[44] H. Liu, F. Xiao, I.G. Serebriiskii, S.W. O'Brien, M.A. Maglaty, I. Astsaturov, S. Litwin, L.P. Martin, D.A. Proia, E.A. Golemis, D.C. Connolly. Network analysis identifies an HSP90-central hub susceptible in ovarian cancer. Clin Cancer Res (2013) 19, 5053-5067.

Table 1. List of MISIIRTAg DR26 blood samples.

	ID	Tag	*4 weeks*	*8 weeks*	*12 weeks*
1	K2821	-	N/A	N/A	401
2	K2869	-	N/A	N/A	402
3	K2981	+	203	303	403
4	K2982	-	204	304	404
5	K2987	+	205	305	405
6	K2988	-	206	306	406
7	K2989	+	207	307	407
8	K2995	+	208	308	408
9	K2996	-	209	309	409
10	K3002	+	210	310	410
11	K3008	-	211	311	411
12	K3010	+	212	312	412
13	K3011	+	213	313	413
14	K3016	-	214	314	414
15	K3066	+	215	315	415
16	K3067	-	216	316	416
17	K3068	-	217	317	417
18	K3069	-	218	318	418
19	K3252	-	219	319	419
20	K3253	-	220	320	420
21	K3308	-	221	321	421
22	K3309	+	222	322	422
23	K3310	-	223	323	423
24	K3314	-	224	324	424
25	K3415	-	225	325	425
26	K3416	+	226	326	426
27	K3417	-	227	327	427
28	K3418	+	228	328	428
29	K3419	-	229	329	429

30	K3421	+	230	330	430
31	K3422	+	231	331	431
32	K3423	+	232	332	432
33	K3424	-	233	333	433
34	K3463	+	234	334	434
35	K3464	+	235	335	435
36	K3466	-	236	336	436
37	K3471	+	237	337	437
38	K3472	+	238	338	438
39	K3473	+	239	339	439
40	K3474	-	240	340	440
41	K3475	-	241	341	441
42	K3476	-	242	342	442
43	K3477	-	243	343	443
44	K3478	+	244	344	444
45	K3479	-	245	345	445
46	K3549	-	246	346	446
47	K3550	+	247	347	447
48	K3551	-	248	348	448
49	K3552	+	249	349	449
50	K3553	+	250	350	450
51	K3554	+	251	351	451
52	K3555	+	252	352	452
53	K3556	+	253	353	453
54	K3557	-	254	354	454
55	K3558	+	255	355	N/A
56	K3653	+	256	356	N/A

Figure 1

Blood sample preparation

a *b* *c*
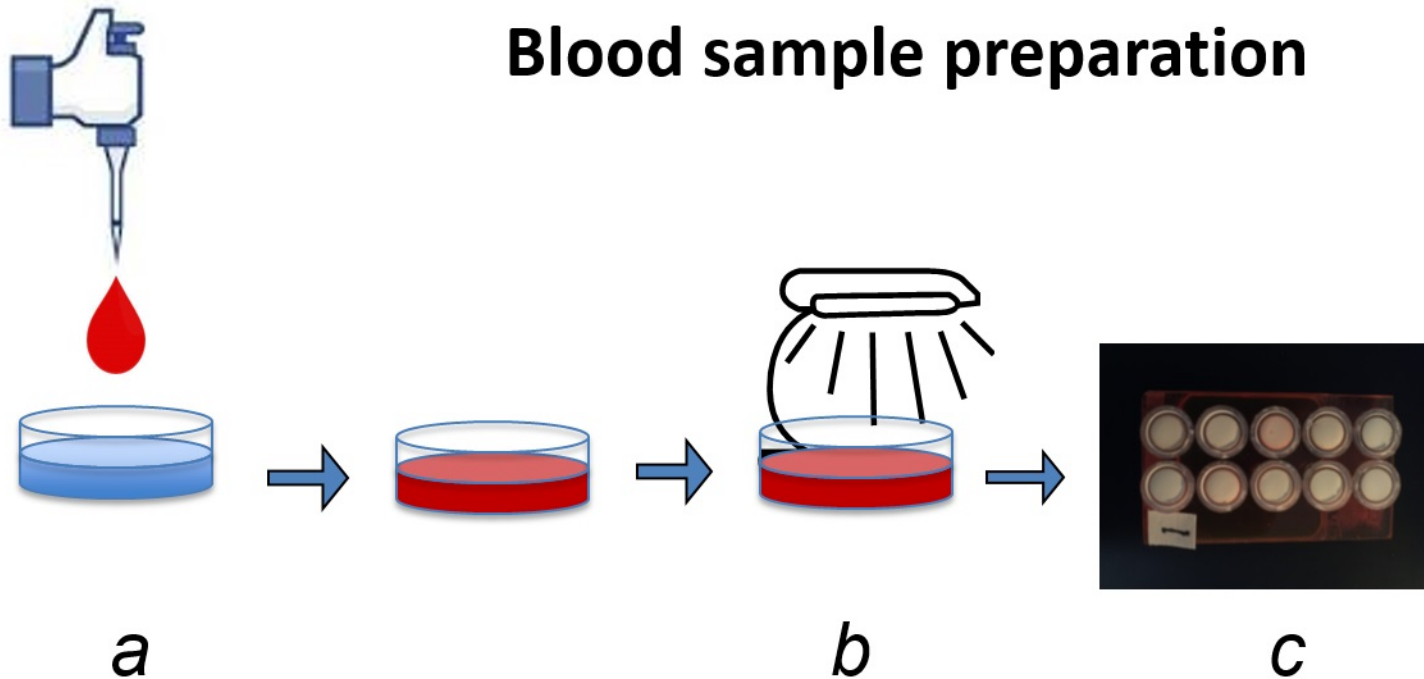
Figure 2

Figure 3

Figure 4a

Figure 4b
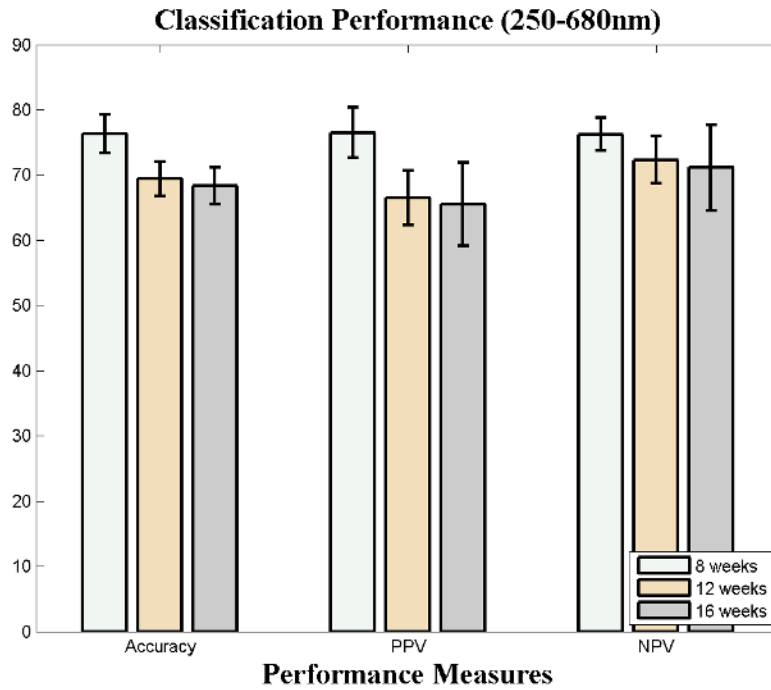
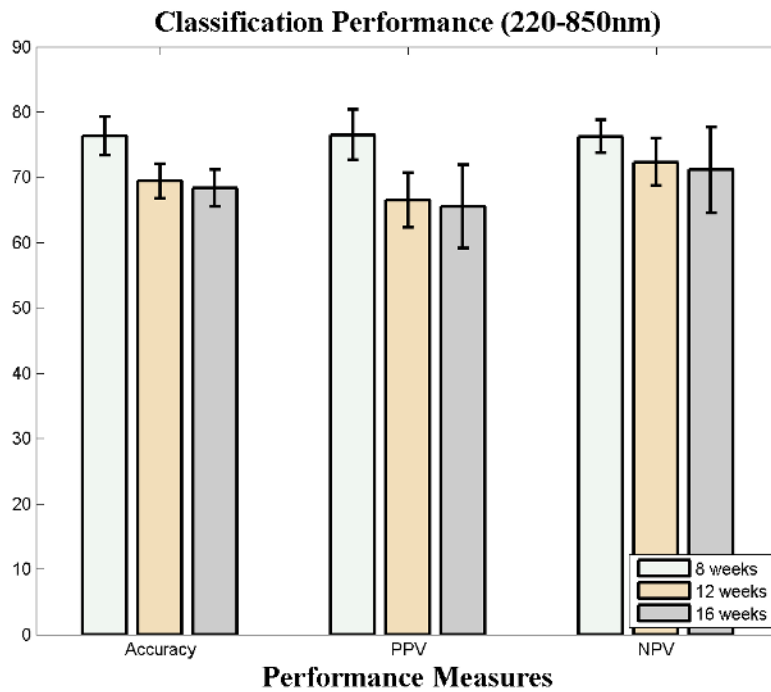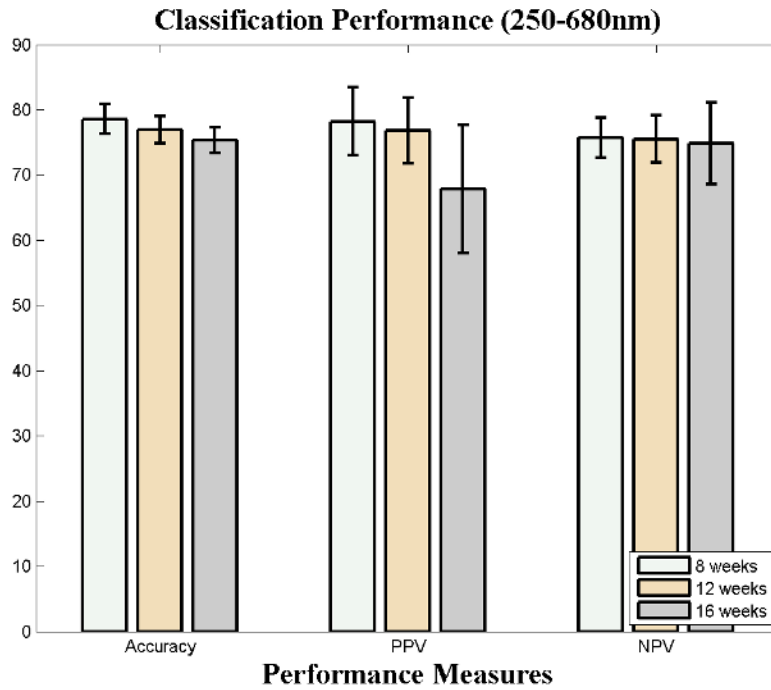Figure 4c
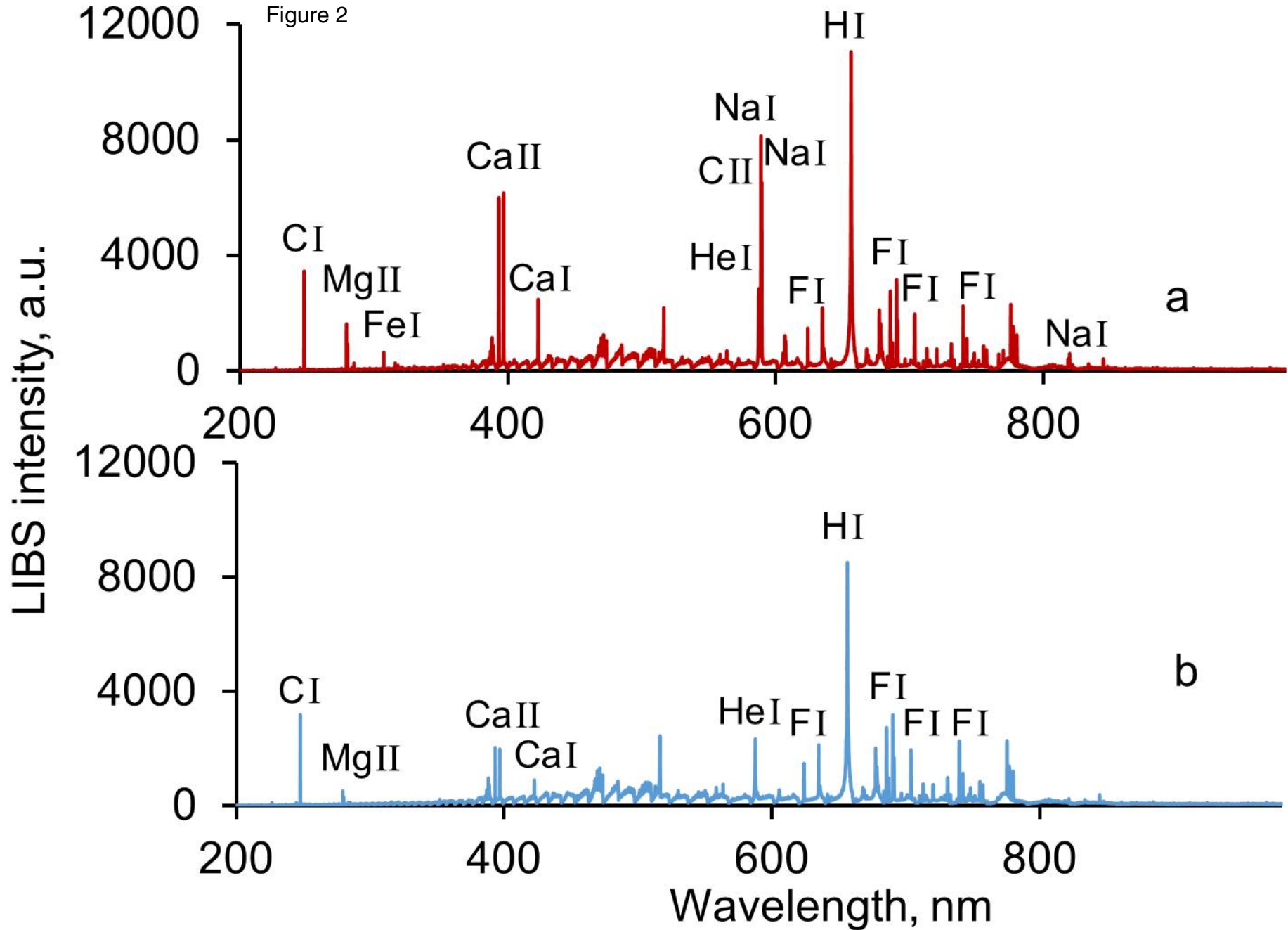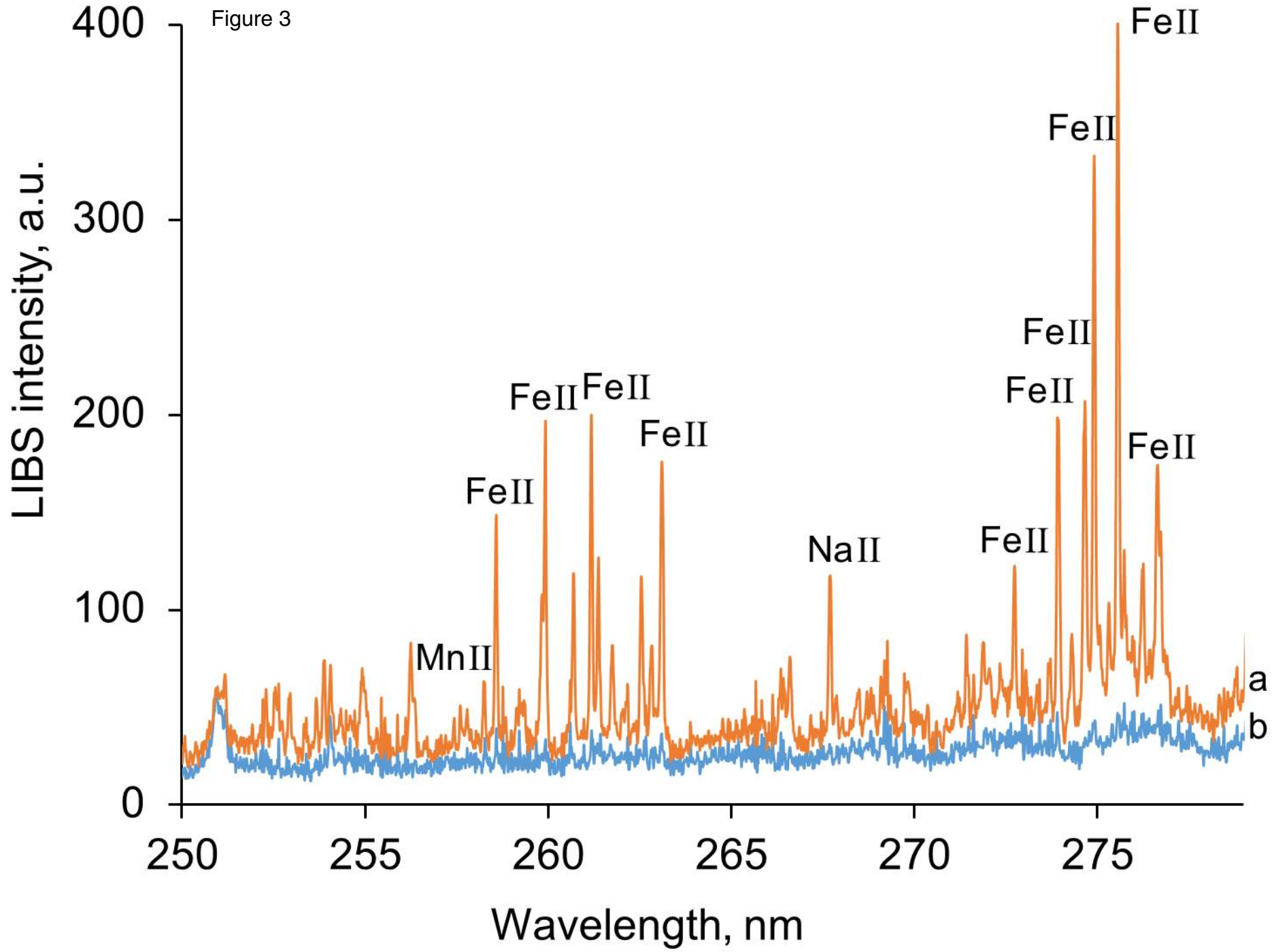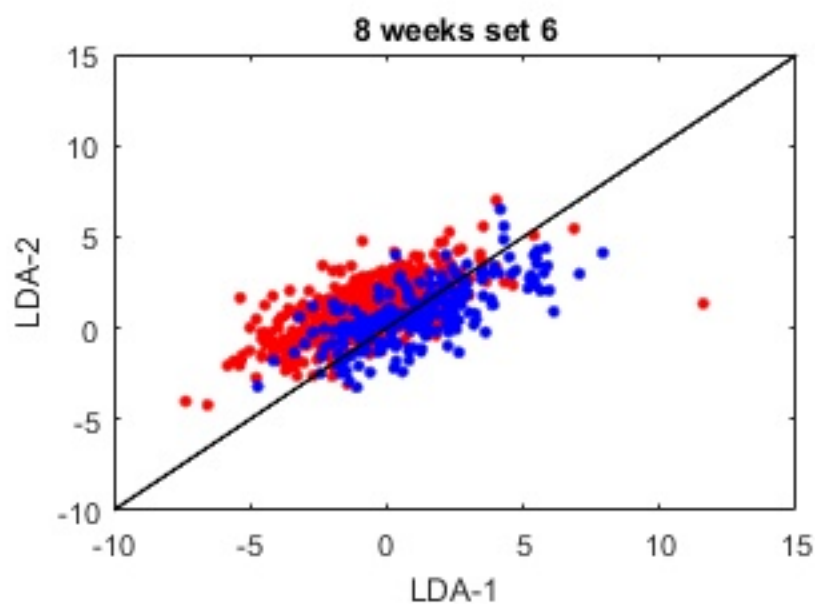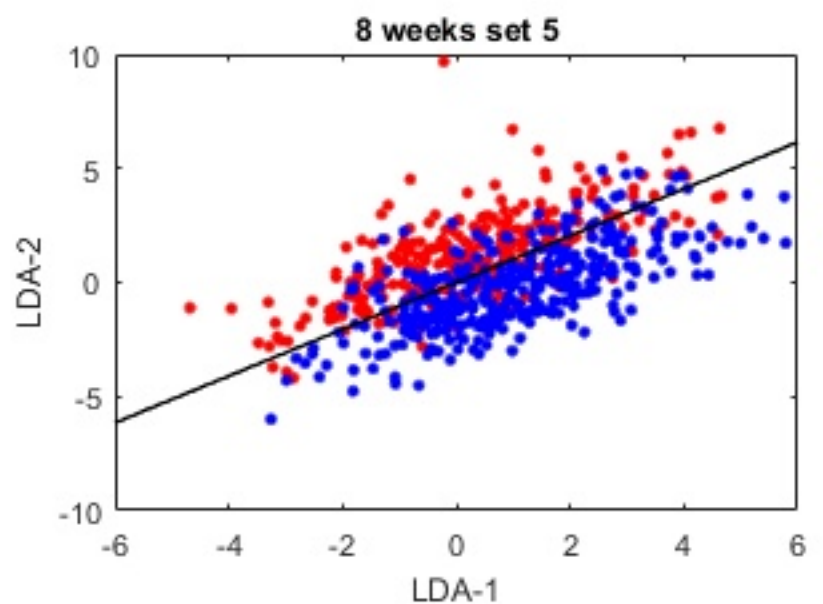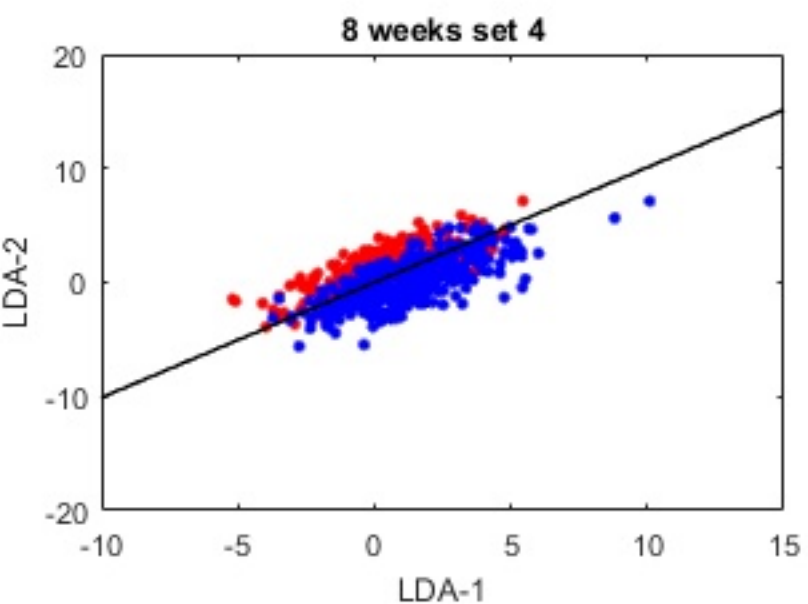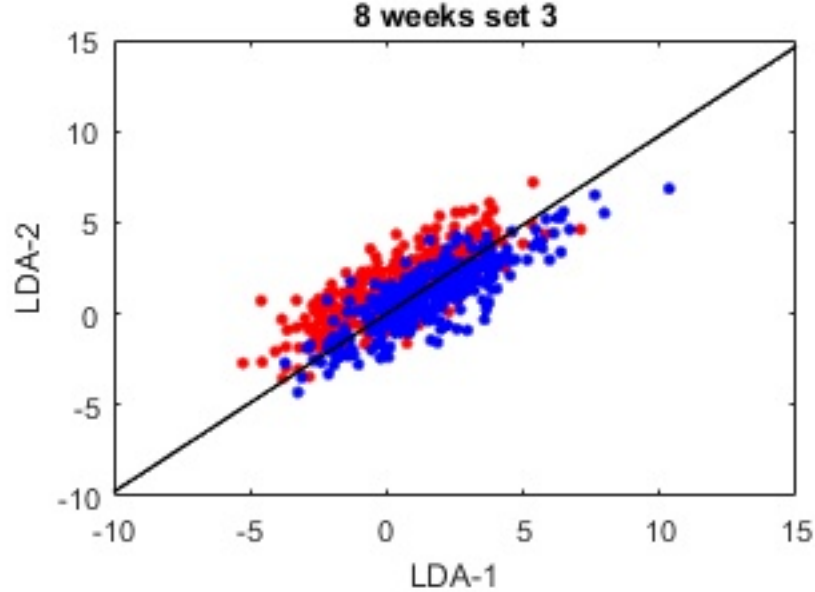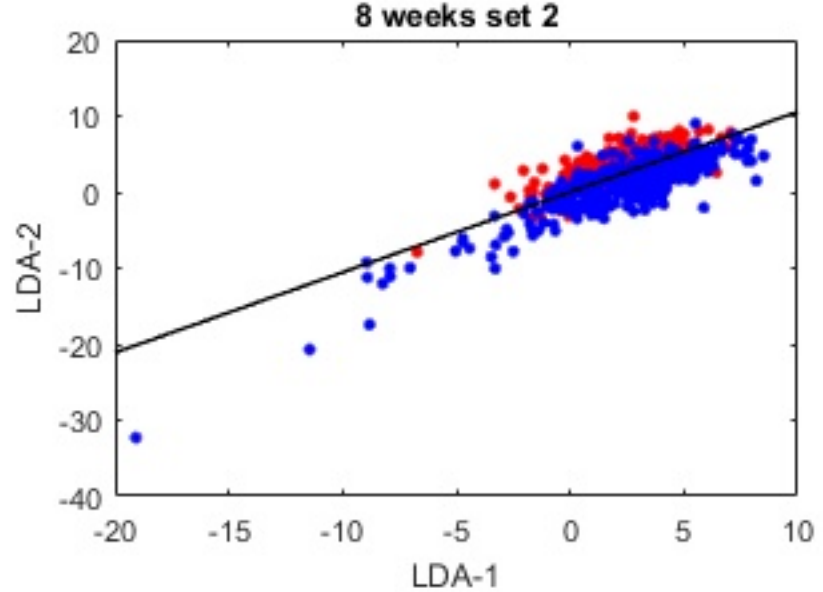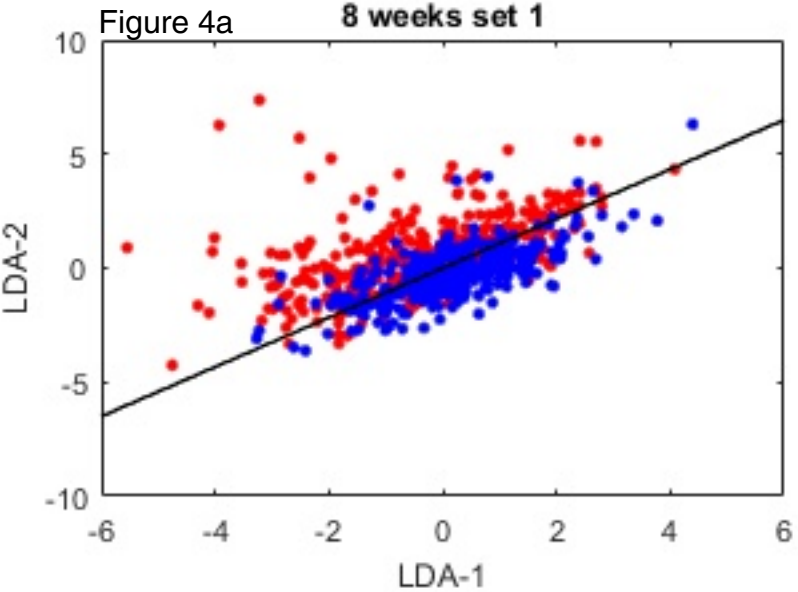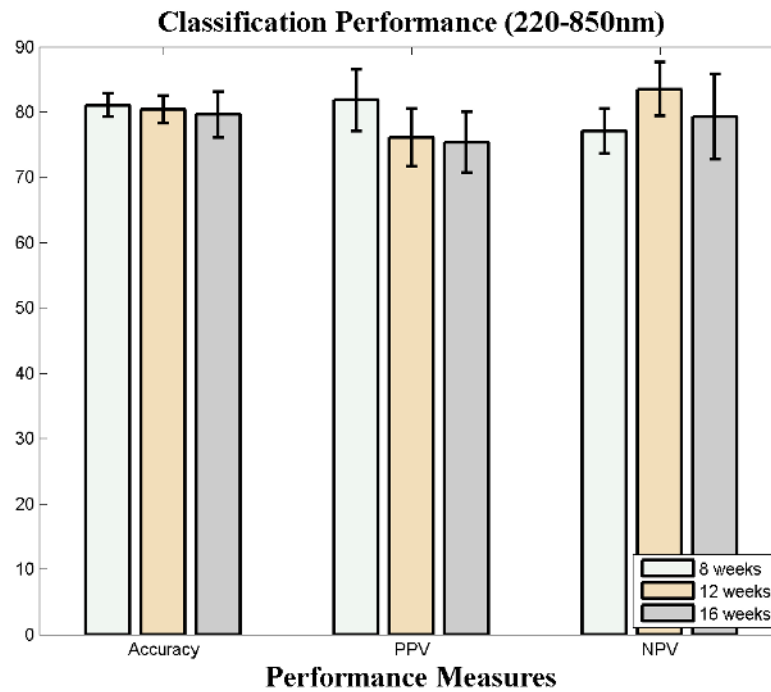
Figure 5a
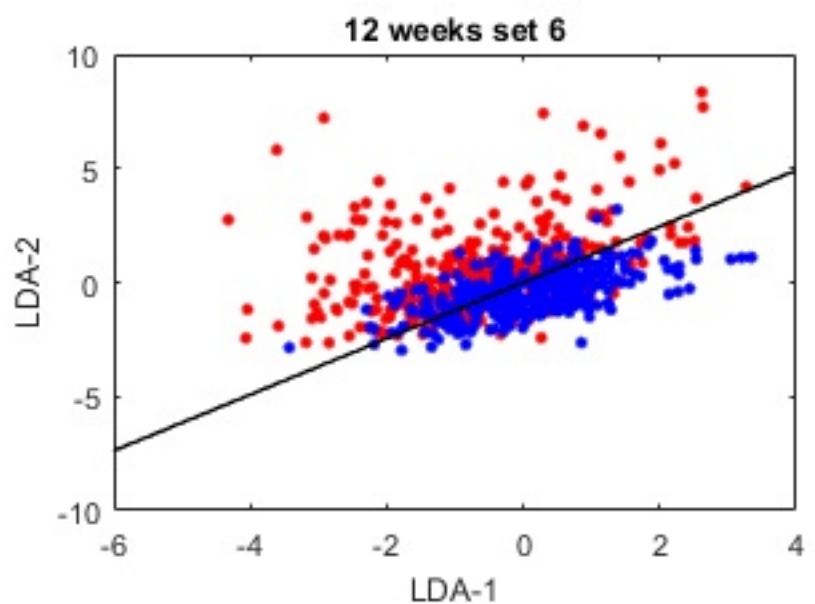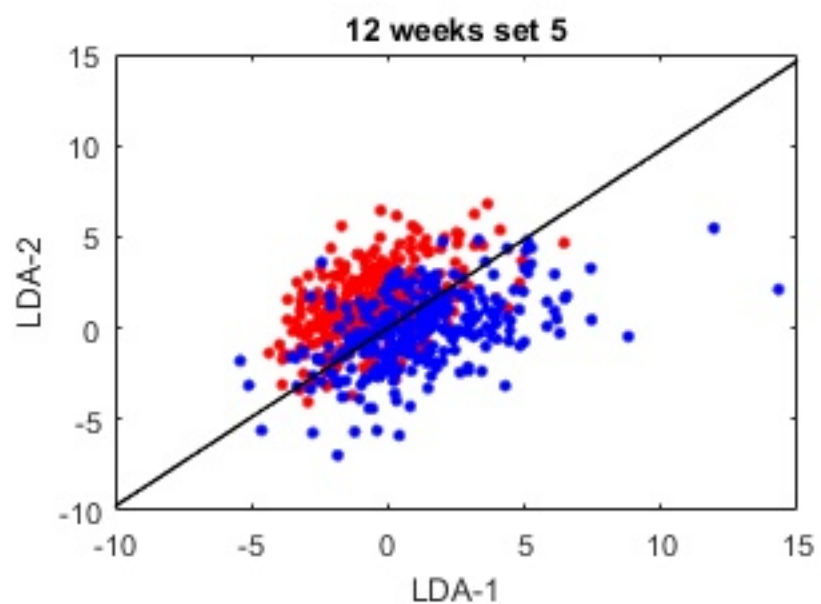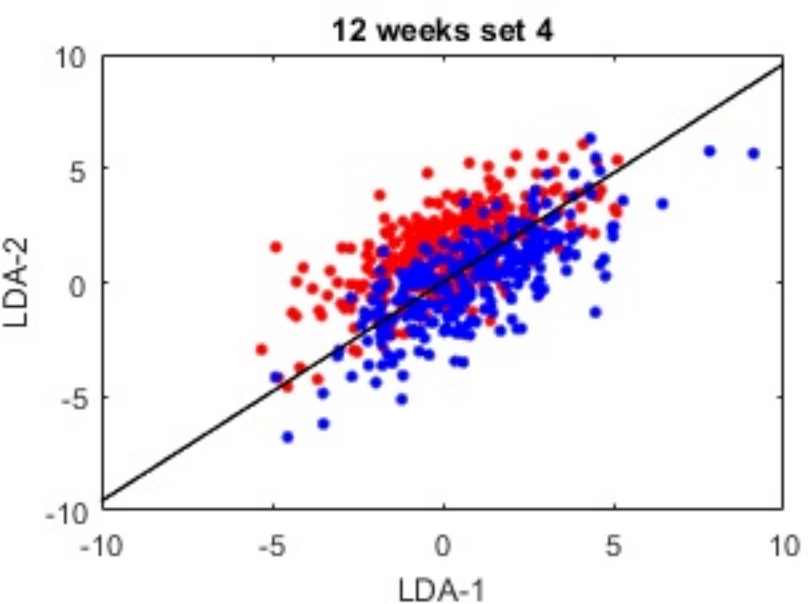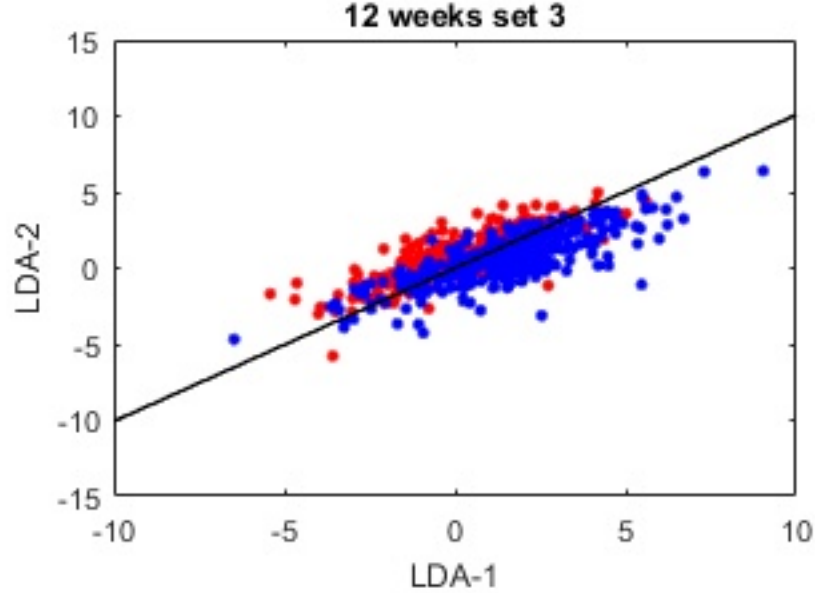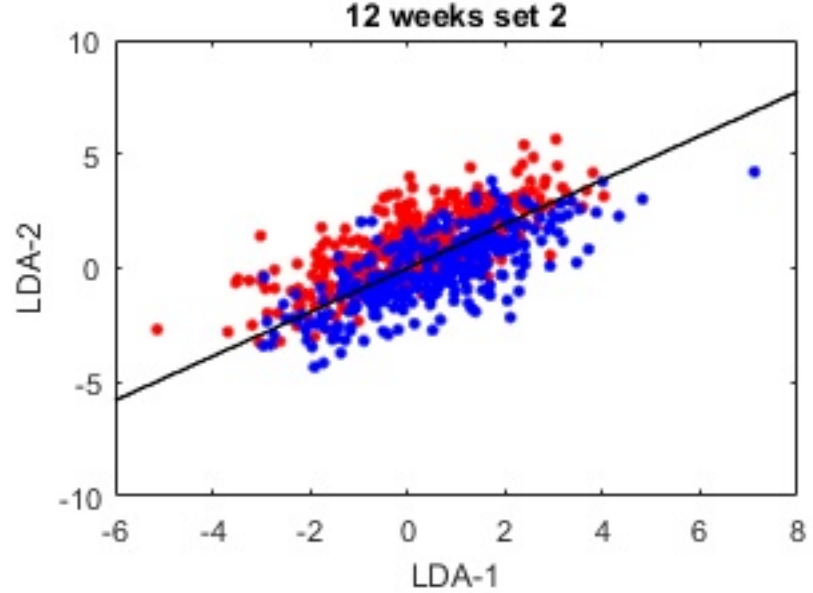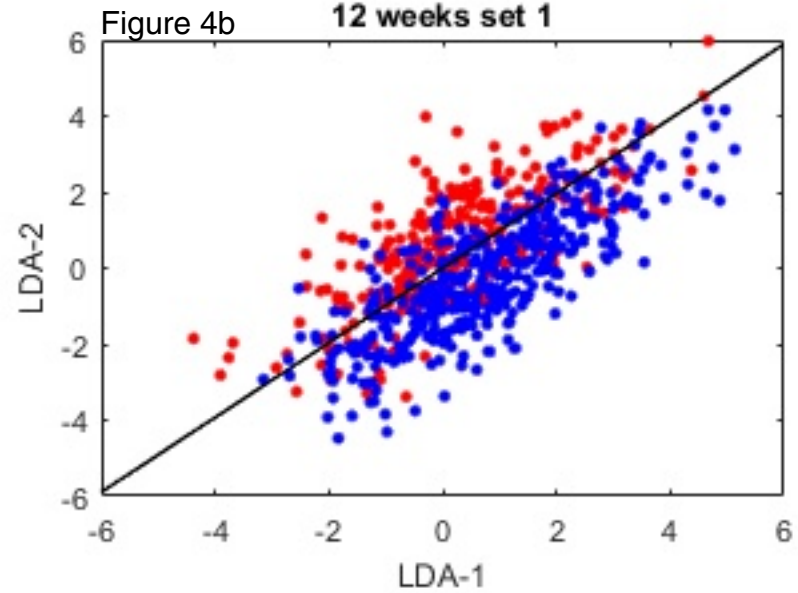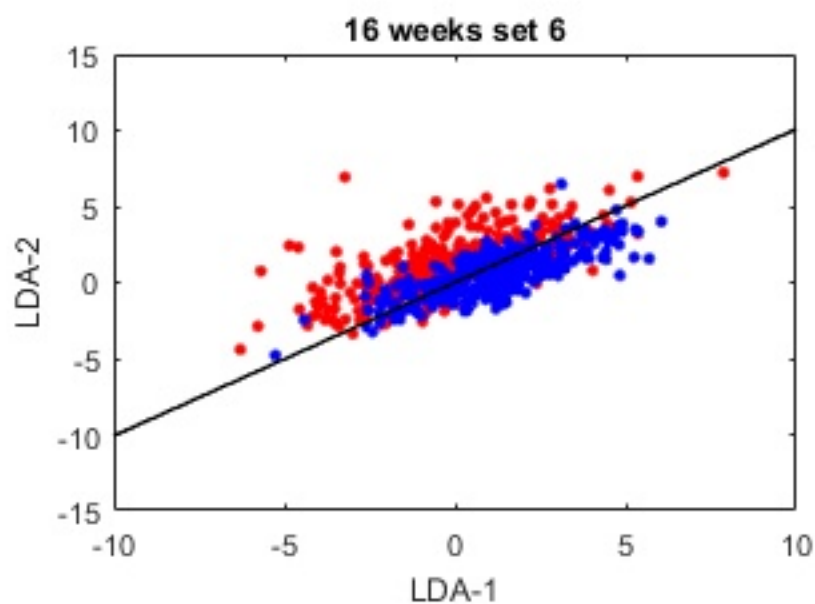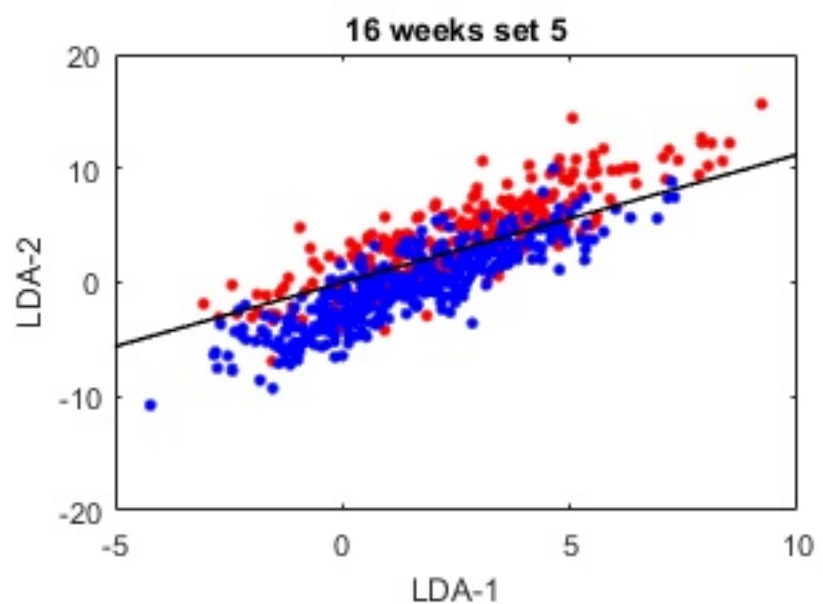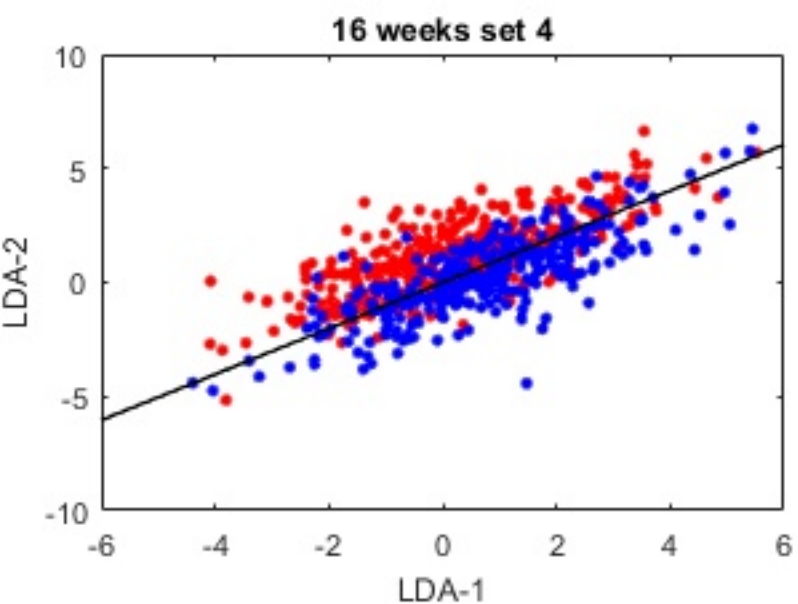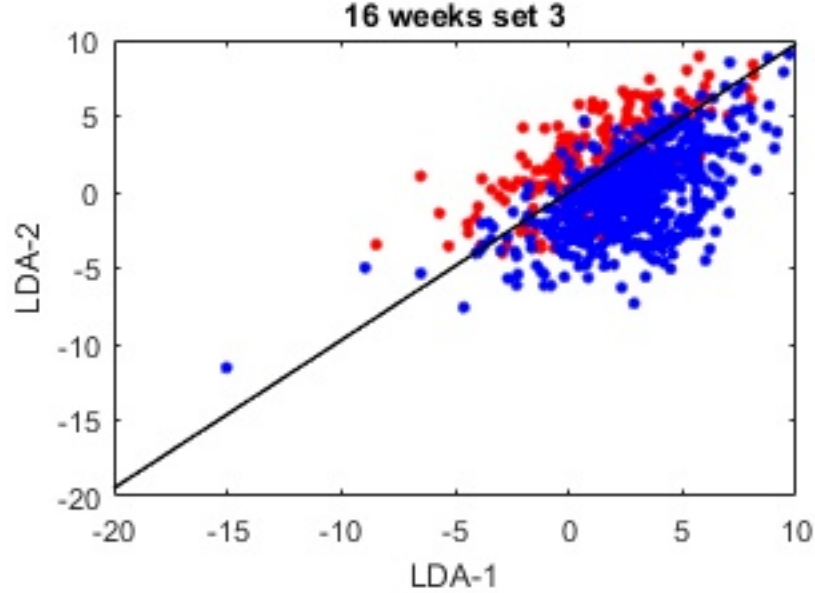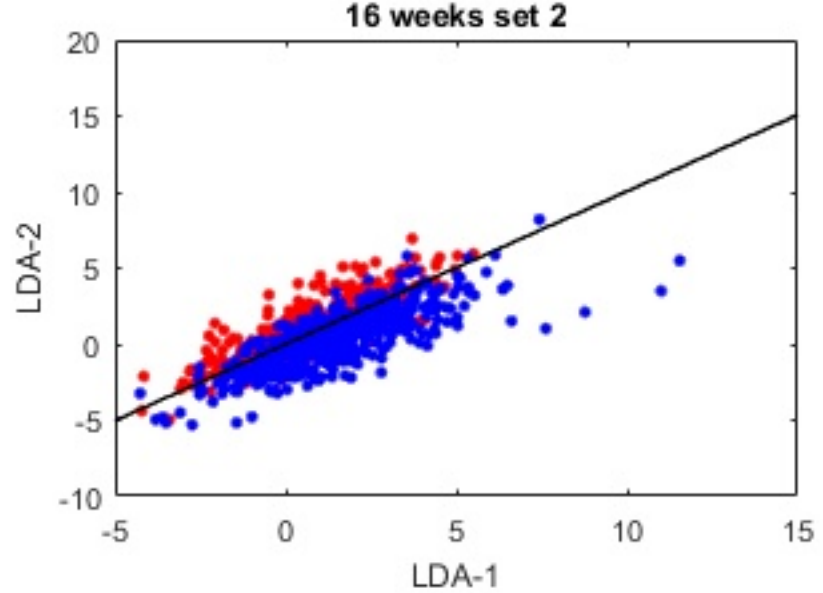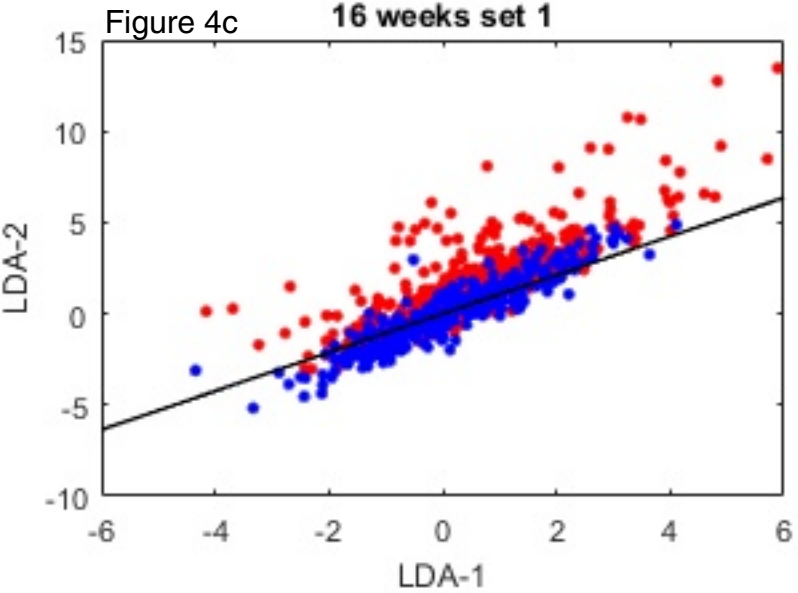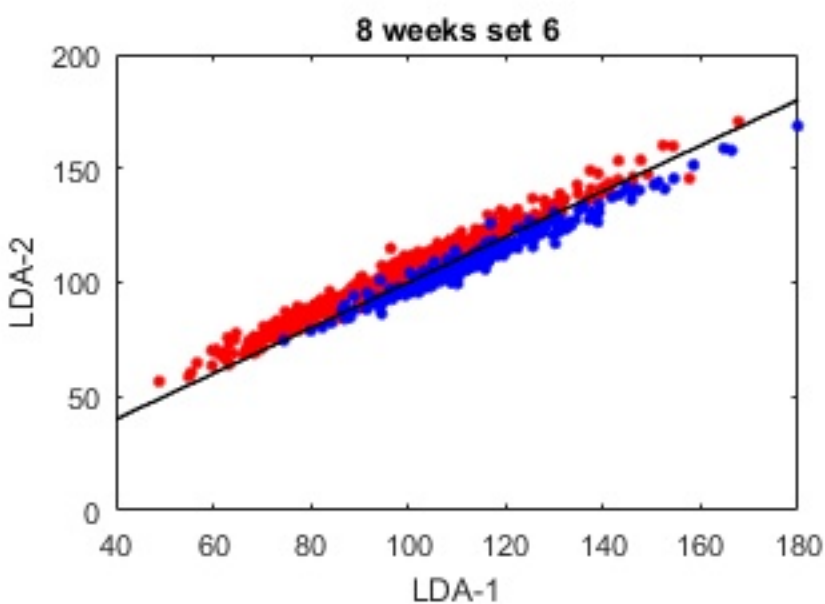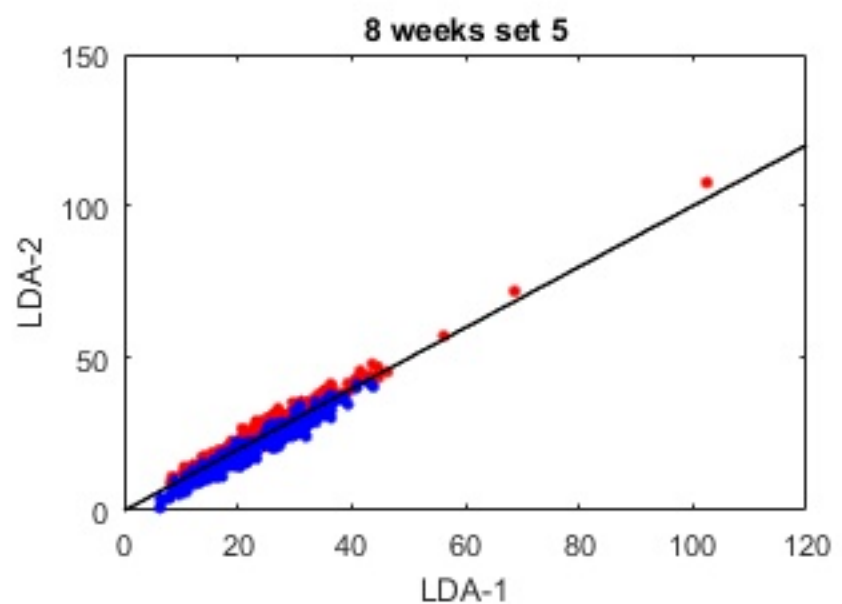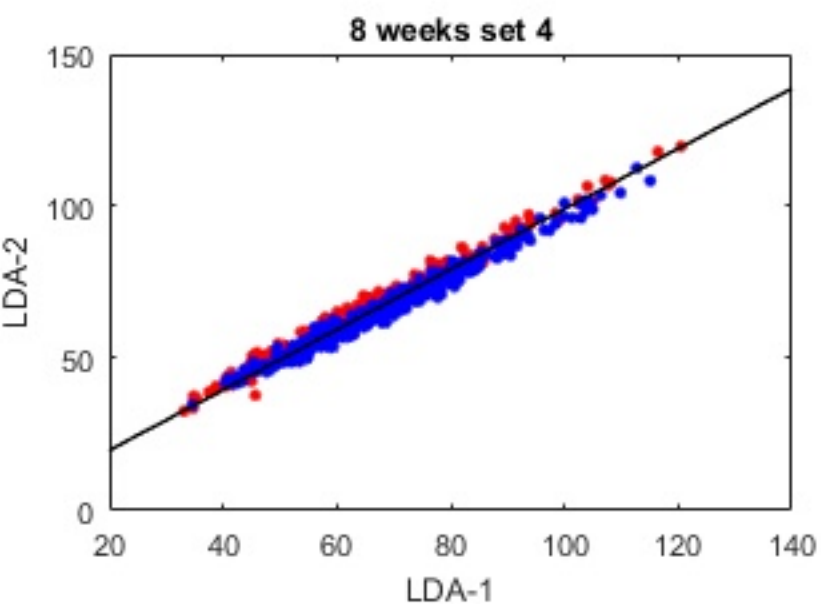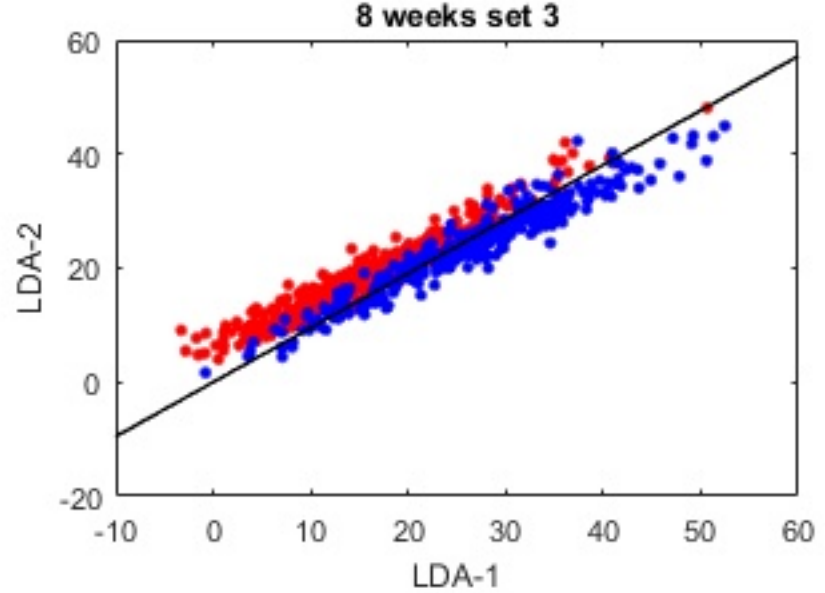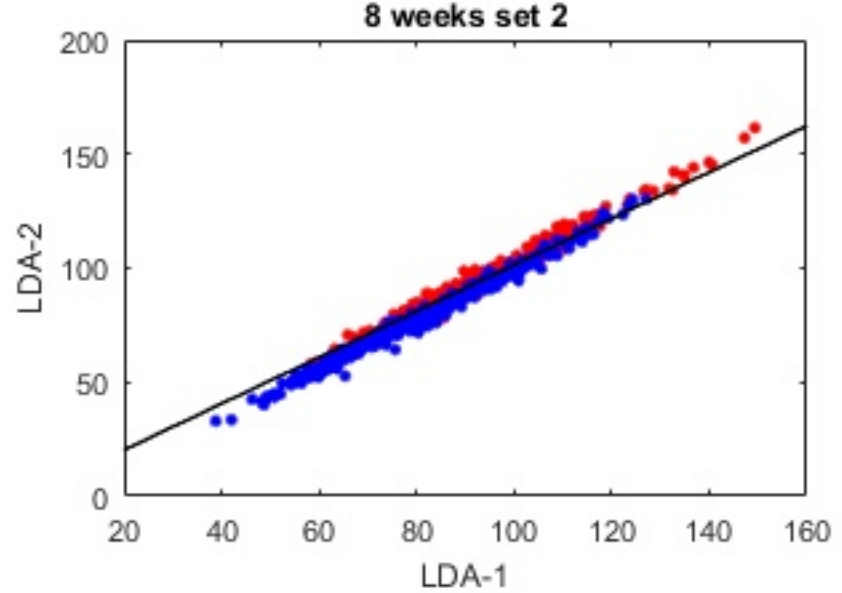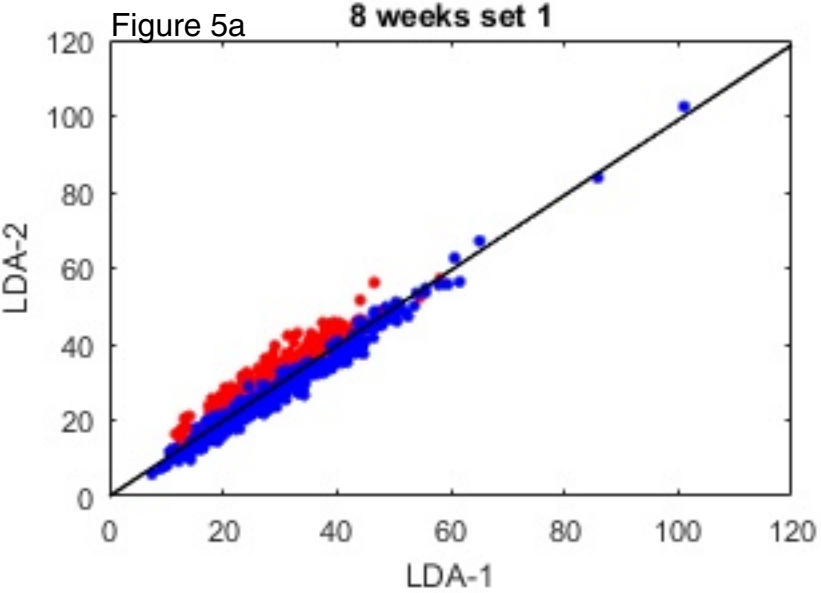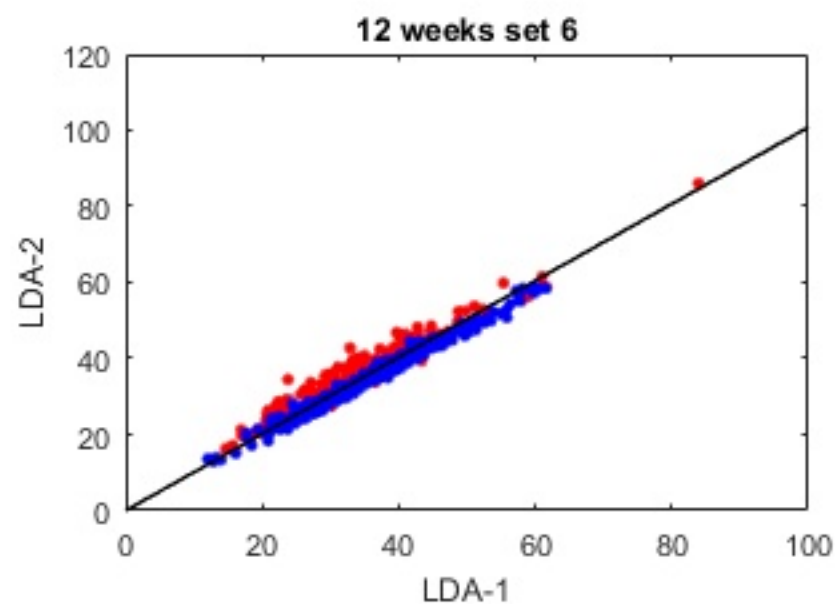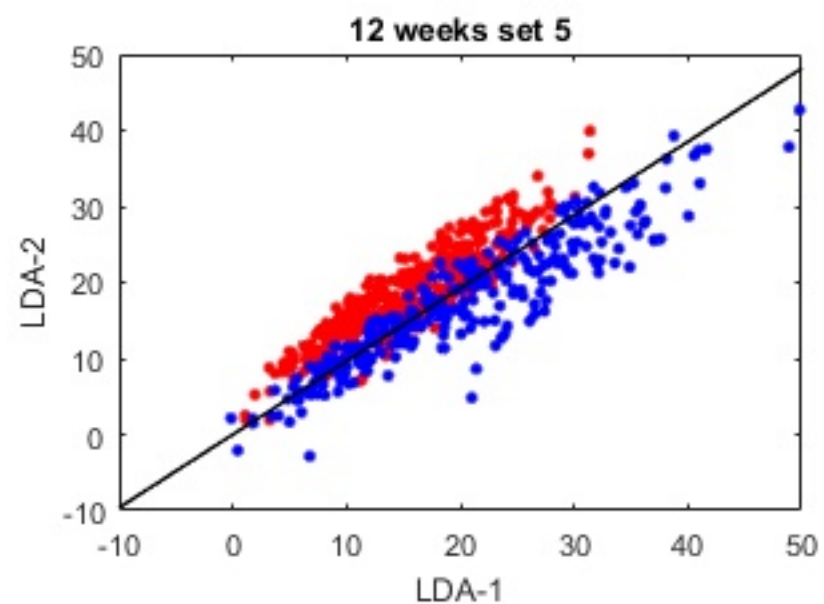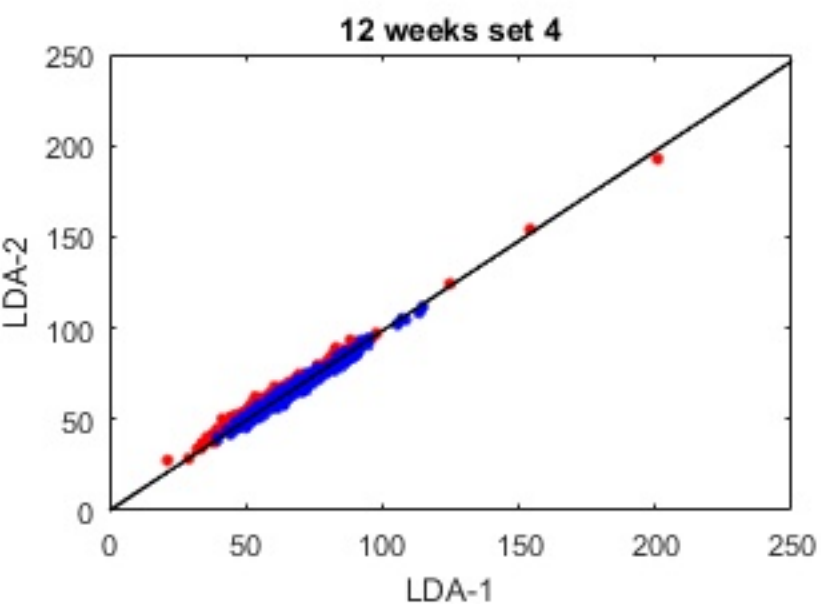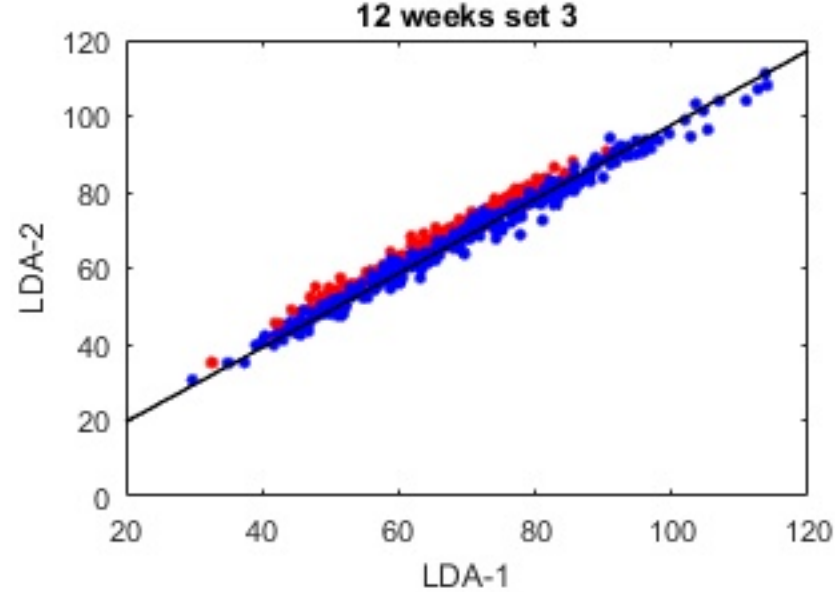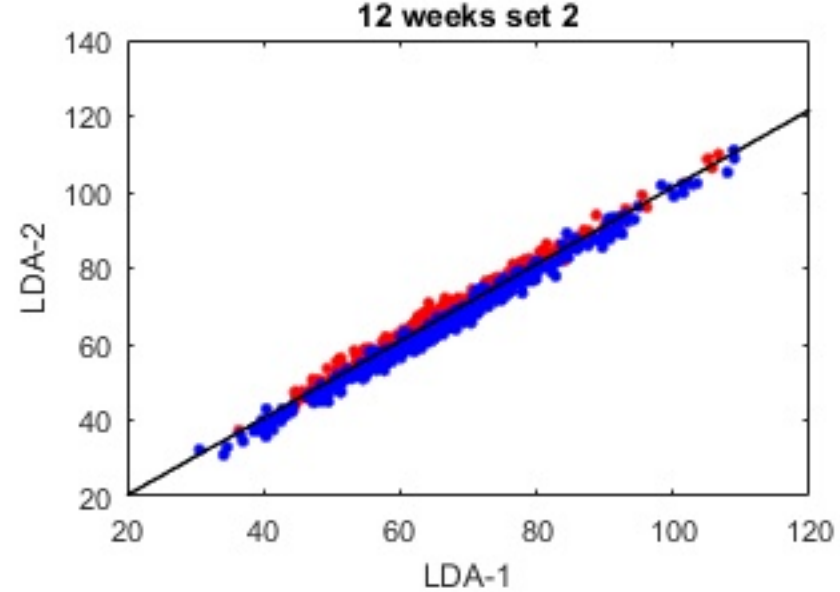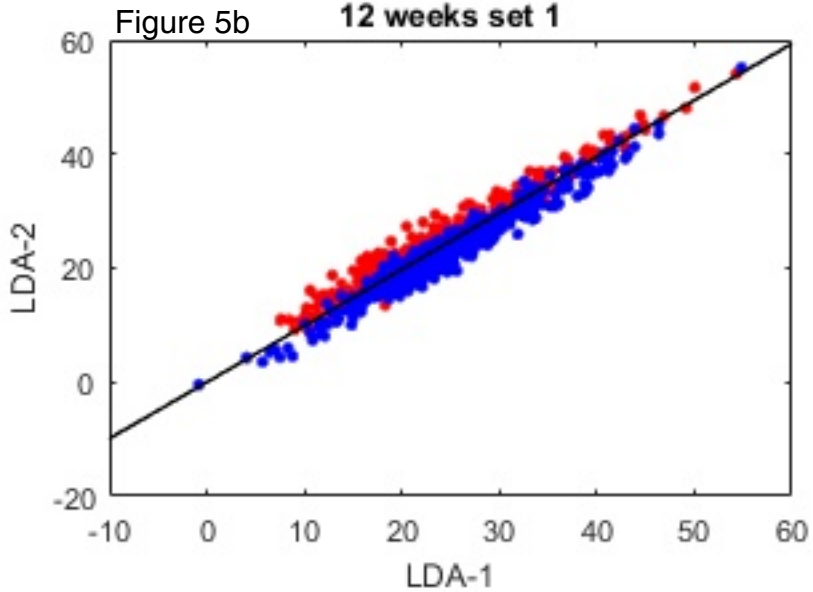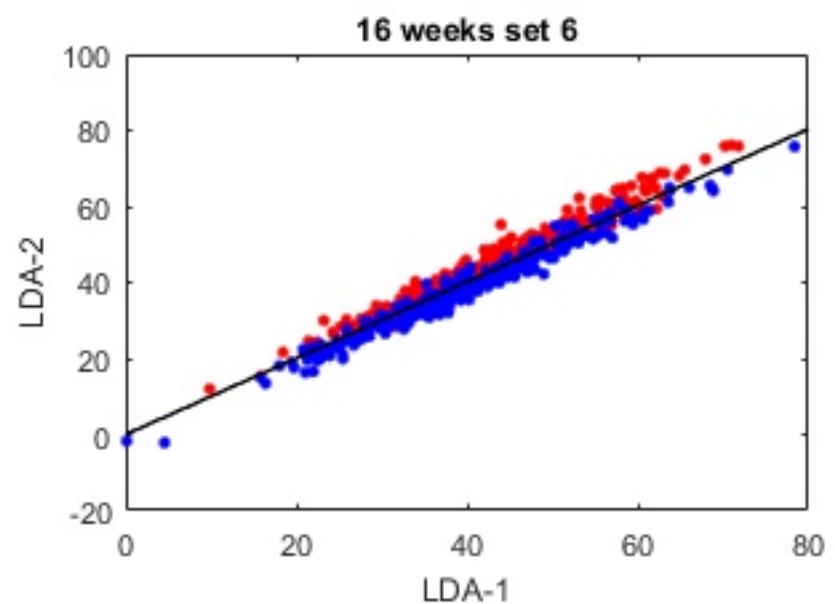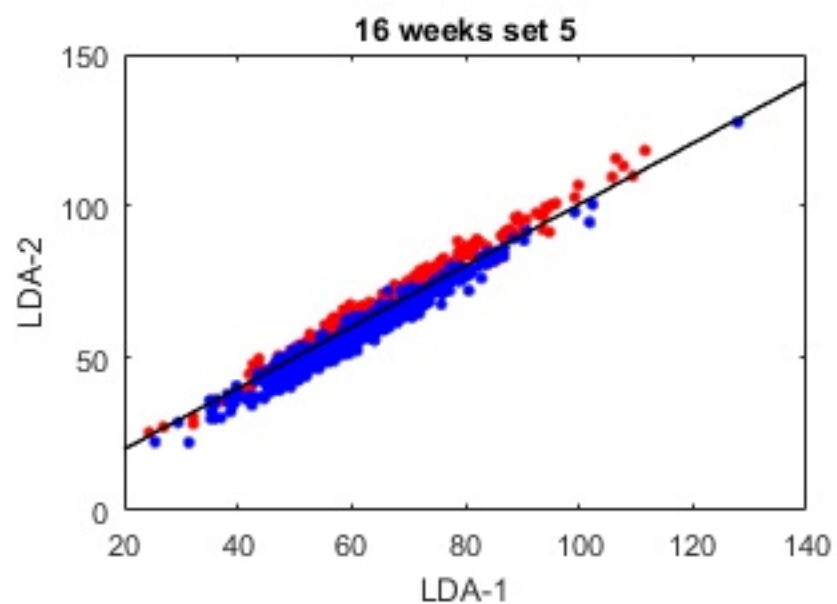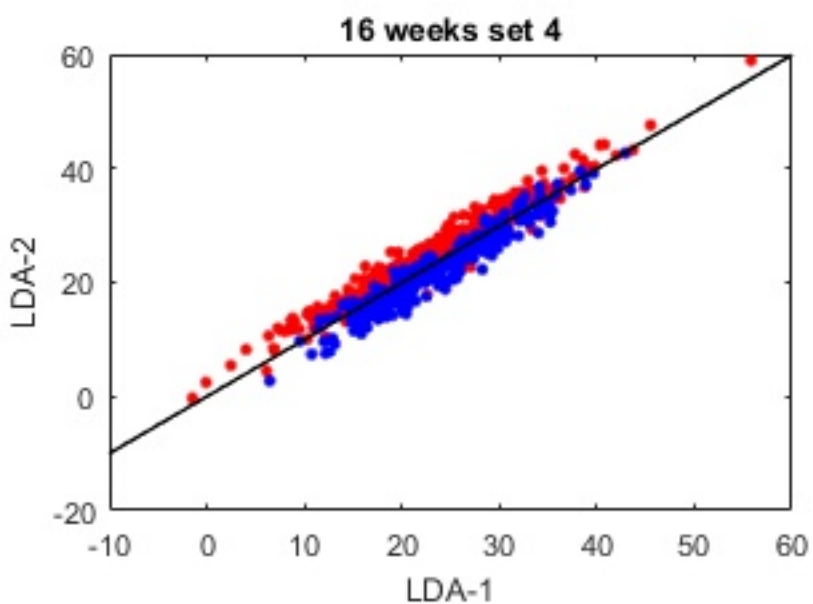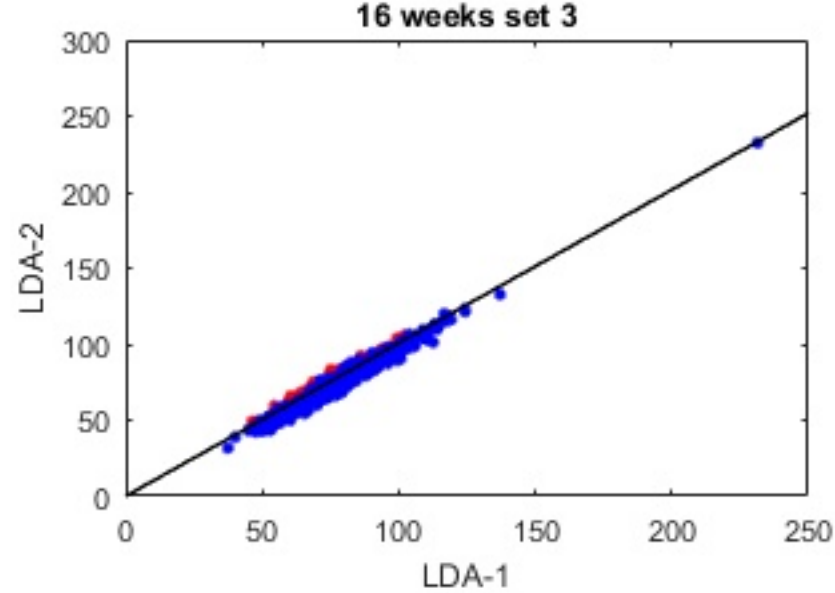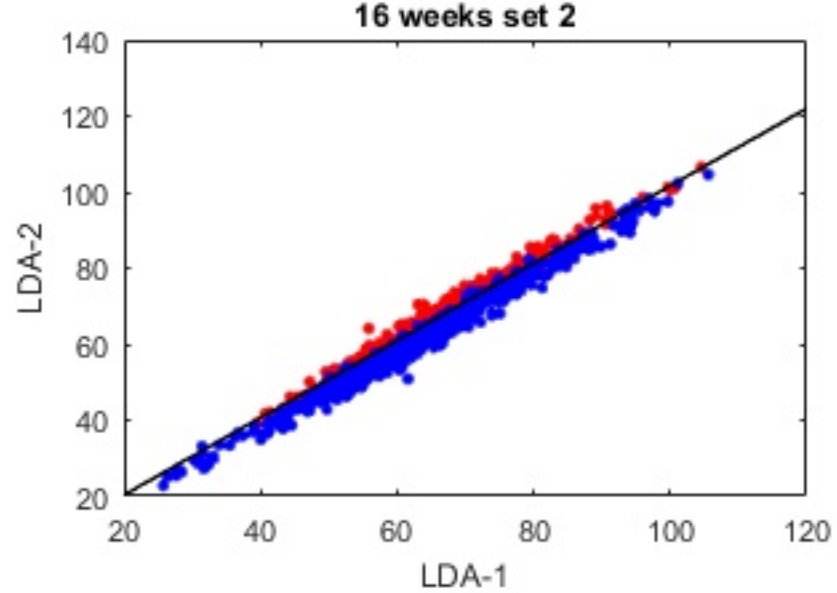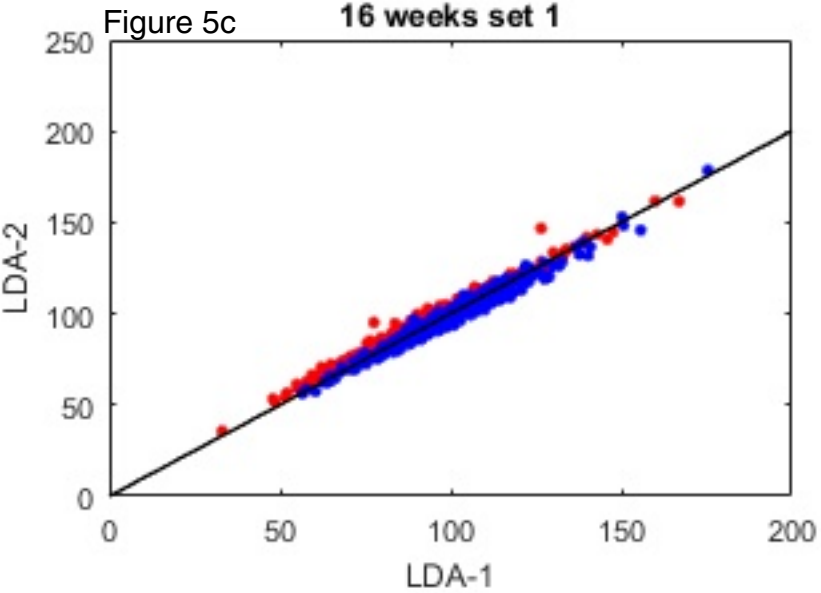
Figure 5b

Figure 5c

(a)

(b)

Figure 6a-b. Classification accuracy, positive predictive value (PPV), and negative predictive value (NPV) produced by the LDA technique using leave-one-out cross-validation for mice ages of 8 weeks, 12 weeks, and 16 weeks in wavelength ranges of (a) 250-680nm and (b) 220-850nm.

(a)

(b)

Figure 7a-b. Classification accuracy, positive predictive value (PPV), and negative predictive value (NPV) produced by the RF technique using leave-one-out cross-validation for mice ages of 8 weeks, 12 weeks, and 16 weeks in wavelength ranges of (a) 250-680nm and (b) 220-850nm.