Published in final edited form as:

J Comp Neurol. 2010 October 15; 518(20): . doi:10.1002/cne.22365.

Comparison of Analytical Mathematical Approaches for Identifying Key Nuclear Magnetic Resonance Spectroscopy Biomarkers in the Diagnosis and Assessment of Clinical Change of Diseases

Jason B. Nikas^{1,2,*}, C. Dirk Keene³, and Walter C. Low^{1,4,5,*}

¹Department of Neurosurgery, Medical School, University of Minnesota, Minneapolis, Minnesota 55455

²Pharmaco-Neuro-Immunology Program, Medical School, University of Minnesota, Minneapolis, Minnesota 55455

³Department of Pathology, University of Washington, Seattle, Washington 98104

⁴Graduate Program in Neuroscience, Medical School, University of Minnesota, Minneapolis, Minnesota 55455

⁵Department of Integrative Biology and Physiology, Medical School, University of Minnesota, Minneapolis, Minnesota 55455

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is a rapidly emerging technology that can be used to assess tissue metabolic profile in the living animal. At the present time, no approach has been developed 1) to systematically identify profiles of key chemical alterations that can be used as biomarkers to diagnose diseases and to monitor disease progression; and 2) to assess mathematically the diagnostic power of potential biomarkers. To address this issue, we have evaluated mathematical approaches that employ receiver operating characteristic (ROC) curve analysis, linear discriminant analysis, and logistic regression analysis to systematically identify key biomarkers from NMR spectra that have excellent diagnostic power and can be used accurately for disease diagnosis and monitoring. To validate our mathematical approaches, we studied the striatal concentrations of 17 metabolites of 13 R6/2 transgenic mice with Huntington's disease, as well as those of 17 wild-type (WT) mice, which were obtained via in vivo proton NMR spectroscopy (9.4 Tesla). We developed diagnostic biomarker models and clinical change assessment models based on our three aforementioned mathematical approaches, and we tested all of them, first, with the 30 original mice and, then, with 31 unknown mice. Their prediction results were compared with genotyping—the gold standard. All models correctly diagnosed all of the 30 original mice (17 WT and 13 R6/2) and all of the 31 unknown mice (20 WT and 11 R6/2), with a positive likelihood ratio approximating infinity [$1/0 \rightarrow \infty$], and with a negative likelihood ratio equal to zero [$0/1 = 0$].

Indexing terms

receiver operating characteristic (ROC) curve analysis; linear discriminant analysis; logistic regression analysis; proton magnetic resonance spectroscopy; metabolomics; Huntington's disease

Nuclear magnetic resonance (NMR) spectroscopy, empowered by technological advances in the last 15 years, has brought to light a new research area called metabolomics (Viant et al., 2003; Griffin, 2003; Pears et al., 2005; Weckwerth, 2007). The axiomatic premise upon which the theory of metabolomics is predicated is the following: a healthy cell has a metabolism that is different from that of the same type of cell that has been affected by a disease or a disorder. The study of cellular metabolite concentrations, therefore, especially as it pertains to the differentiation between the normal metabolic state of cells and an abnormal state caused by a disease, is the subject matter of metabolomics. Given that organisms are cellular in nature, the potential of this new research area is boundless. In medical research, more specifically, NMR spectroscopy, which is noninvasive and can be used longitudinally, can be applied in the study of many diseases, disorders, and conditions, ranging from neurological diseases (such as Huntington's and Parkinson's) to cardiovascular diseases (such as stroke and coronary artery disease) to cancer (Graham et al., 1993; Freedman et al., 1998; Nakae et al., 2003; Jones and Waldman, 2004; Boska et al., 2005; Jenkins et al., 2005; Opstad et al., 2008).

As often happens, however, in a field with rapid technological advances, both the means to fully utilize and apply the capabilities of the new technology and the means to maximally reap the benefits resulting from it are underdeveloped at best and lacking at worst. In the case of NMR spectroscopy, a systematic mathematical methodology, whereby appropriate and soundly applied mathematical analyses are employed to elucidate all the information captured by the rapidly increasing resolution of NMR spectroscopy, is lacking.

In this study, by using the NMR spectroscopy data of the wild-type (WT) and the R6/2 transgenic mice with Huntington's disease that were presented in our previous study (Tkac et al., 2007), we employed three different analytical mathematical approaches, receiver operating characteristic (ROC) curve analysis, linear discriminant analysis (LDA), and logistic regression analysis (LRA), to accomplish the following objectives:

1. Construct mathematical diagnostic biomarker models (DBMs) from which we could develop diagnostic biomarker computer programs that could accurately diagnose R6/2 mice as a prototype for the diagnosis of diseases. Because genotyping is available for the R6/2 Huntington's disease mice, and because genotyping is the gold standard, we would use it to test the accuracy of our DBMs.
2. Develop clinical change assessment models, based on the above mathematical analyses and modeling, that could accurately measure the progress of a disease and potentially the efficacy of an applied medical treatment or intervention.

Materials and Methods

Animal methods

R6/2 Transgenic mice—All animal breeding and all animal experiments described in this study were performed in accordance with the procedures approved by the University of Minnesota Institutional Animal Care and Use Committee. Original R6/2 mice were purchased from the Jackson Laboratories (Bar Harbor, ME) and bred by crossing transgenic males and wild-type (WT) females at 5 weeks of age. Offspring were genotyped according to established procedures (Mangiarini et al., 1996; the Jackson Laboratory).

Animal preparation—Prior to the in vivo ^1H NMR (proton nuclear magnetic resonance) scanning, all animals were anesthetized and maintained thus throughout the duration of the scanning procedure. The anesthesia used was a gas mixture ($\text{O}_2:\text{N}_2\text{O} = 1:1$) containing 1.25–2.0% of isoflurane that flowed throughout the cylindrical chamber wherein the spontaneously breathing animals were placed. The temperature inside the chamber was maintained at 30°C by the circulation of warm water on the outside surface of the chamber. The duration of the ^1H NMR scanning was approximately 1 hour.

In Vivo ^1H NMR Spectroscopy—All ^1H NMR scans were conducted with a 9.4 T/31 cm magnet (MagneX Scientific, Abingdon, UK) equipped with an 11-cm gradient coil insert (300 mT/m, 500 ls) and strong custom-designed second-order shim coils (MagneX Scientific) (Tkac et al., 2004). The position of the volume of interest (VOI) was selected based on multi-slice RARE images. The VOI was centered in the left striatum at the level of the anterior commissure. The size of the VOI, which varied from 7 to $12\ \mu\text{L}$, was adjusted to fit the anatomical structure of the left striatum, as well as to exclude the lateral ventricle and thus to minimize partial volume effects (inclusion of a tissue other than the target tissue). The striatum was selected as the area of interest because it consists to a large extent of the medium spiny projection neurons, which are γ -aminobutyric acid (GABA)ergic, and which, more importantly, constitute the initial and preferential target of Huntington's disease. It is in this structure of the brain and in this neuronal population that Huntington's disease first manifests itself and its destructive force. At the end stage, following extensive neuronal cell loss in the striatum, the disease evinces itself in other brain areas, such as the cerebral cortex, globus pallidus, and—to a smaller extent—the substantia nigra, the thalamus, the cerebellum, etc. (Browne et al., 2004).

Thirty mice (17 WT and 13 R6/2) were scanned according to the aforementioned procedure. Of the 17 WT mice, 8 were 8 weeks old and 9 were 12 weeks old; of the 13 R6/2 mice, 7 were 8 weeks old and 6 were 12 weeks old. Of the 30 spectra obtained from the ^1H NMR scans, four (representative of the two groups of mice and their two age subgroups) are shown in Figure 1. Those 30 mice were used in the development of the DBMs of all three analytical mathematical approaches. In addition, 31 unknown mice (11 R6/2 and 20 WT) were also scanned according to the aforementioned procedure and were used to test all three DBMs. All of those 31 unknown mice were extraneous to the development of the three DBMs, and their status had been ascertained via genotyping.

Spectral analysis resulted in the identification and individual quantification of 15 metabolites. By combining the obtained individual absolute concentrations of creatine (Cr) and phosphocreatine (PCr), we created the Cr+PCr and PCr/Cr metabolites (variables) in order to obtain information about the total striatal creatine (free and phosphorylated), as well as about the ratio of those two metabolites. In the case of glycerophosphorylcholine (GPC) and phosphorylcholine (PC), we were not able to separate those two and obtain individual concentrations. We were able, however, to obtain the absolute concentration of the sum of GPC and PC, which represents the total striatal phosphorylated choline. All of the 15 striatal metabolites we were able to identify and quantify individually as a result of the high magnetic field spectrometer we used (9.4 Tesla), as well as the two metabolites (variables) we created, are shown in Table 1.

Because both of our animal groups (WT and R6/2) comprised two age subgroups (8- and 12-week-old mice), the time-dependent variable was collapsed, so the developed models for diagnostic biomarkers and clinical change assessment would be applicable from 8–12 weeks of age—a most important time period in the progression of the disease in R6/2 mice, as well as a significant portion of the observed lifespan of the R6/2 mice. The development of all of the subsequent mathematical analyses, therefore, was based on the data of the

aforementioned 13 R6/2 mice (seven at 8 weeks of age and six at 12 weeks of age) and 17 WT mice (eight at 8 weeks of age and nine at 12 weeks of age). For more details on animal methods, as well as on spectra obtainment and processing, please see our previous study (Tkac et al., 2007).

Statistical software and graphics

For all of our mathematical analyses, we used the statistical software of NCSS 2004 (Kaysville, UT). The graphs appearing in Figure 2 were created by using the statistical software of NCSS 2004. The graphs appearing in Figures 6-10 were created by using the graphics software Origin 8 (OriginLab, Northampton, MA).

Computer programs

Computer programs were written by using MATLAB R2007a (The MathWorks, Natick, MA).

Diagnostic biomarker models

ROC curve analysis—The receiver operating characteristic (ROC) curve analysis, which was originally developed during World War II in connection with radars and sonar equipment (Portney and Watkins, 2000), is a versatile and valuable mathematical tool, especially used in the many fields of clinical research, such as radiology, epidemiology, biomedical informatics, rehabilitation sciences, psychiatry, etc. More specifically, in the area of clinical research, ROC curve analysis can be used to assess the validity of a diagnostic instrument (Brettschneider et al., 2006; Ramankulov et al., 2007), to compare different diagnostic instruments (Rieber et al., 2006; Hong et al., 2007), and to assess the accuracy of a statistical model, such as logistic regression (Ellenberger et al., 2006; Vanderstichele et al., 2006). It can also be used to assess instruments that predict various outcomes, such as future risk, as well as to assess the ability of an outcome measure instrument to detect clinically meaningful change (Lamia et al., 2006; Schultz-Larsen et al., 2007).

ROC curve analysis employs the study of two probabilities, sensitivity and (1-specificity) in order to develop a third probability, namely, the area under the curve (AUC). Sensitivity is the probability that someone with the condition will test positive, whereas (1-specificity) is the probability that someone without the condition will test positive. Obviously, the ideal sensitivity is equal to 1, and the ideal (1-specificity) is equal to 0. The AUC, which is calculated by graphing (1-specificity) (as the independent variable) versus sensitivity (as the dependent variable), is also a probability, and it represents the discriminating power of a diagnostic instrument (Portney and Watkins, 2000). More specifically, the AUC is the probability that a subject drawn from the population with the condition will be diagnosed positive as compared with a subject drawn from the population without the condition. For example, an AUC of 0.95, would indicate that a subject with the condition has a 95% chance of being diagnosed positive, whereas a subject without the condition has a 5% chance of being misdiagnosed as positive. Instruments with an AUC = 0.75 are not clinically useful, whereas instruments with an AUC > 0.96 have excellent diagnostic (prognostic) capability (Fan et al., 2006). Because ROC curve analysis can be performed either when normality is met (by using the *binormal* method) or when normality is not met (by using the empirical method), we needed to determine first whether our independent variables (IVs), i.e., the 17 metabolite concentrations, were normally distributed with respect to our two groups.

Normality: We performed the following seven normality tests: 1) Shapiro-Wilk W; 2) Anderson-Darling; 3) Martinez-Iglewicz; 4) Kolmogorov-Smirnov; 5) D'Agostino Skewness; 6) D'Agostino Kurtosis; and 7) D'Agostino Omnibus. We required that both groups (R6/2 and WT) within each IV pass the majority of the aforementioned normality

tests (4 tests) in order for an IV to be classified as normal. We found that all of 17 IVs were normally distributed with respect to both groups.

It should be pointed out here that in ROC curve analysis (a probability theory), besides ascertaining whether normality is met, there are no other conditions. For instance, multicollinearity, although a very important condition in multivariate methods, does not apply in ROC curve analysis, in which every IV is always analyzed individually and separately from all other IVs.

We performed ROC curve analysis on all of 17 IVs for two reasons. First, we needed to ascertain which of those 17 IVs had the greatest significance in the differentiation between R6/2 and WT mice and, consequently, the greatest significance in the neuropathology of Huntington's disease. Second, we wanted to know whether we could develop diagnostic biomarkers that could accurately diagnose R6/2 mice, provided, of course, a number of the IVs had a sufficiently high AUC.

The results of the ROC analyses for all 17 metabolites are shown in Table 2. Because all of the IVs were normally distributed, we used the *binormal* method, as developed by Metz (1978) and McClish (1989) to carry out the analyses and estimate the AUC's. The graphs, as wells as the AUCs, of the top nine most significant metabolites are depicted in Figure 2.

For every metabolite, we wanted to select the best cutoff concentration value that discriminated best between the R6/2 and the WT mice. We therefore wrote a computer program that, given all the ROC curve points within an IV, would select the point at which there was the highest possible sensitivity and the lowest possible (1-specificity). For *binormal* ROC curves this point is uniquely defined as the shortest distance from the point [0,1]. It is worth mentioning that in the case of an ideal ROC curve ($AUC = 1$), the entire curve collapses to the point [0,1]. The ROC curve of Cr+PCr turned out to be such an ideal curve, and it is shown, along with those of the top nine most significant metabolites, in the results section (Fig. 2). The rank of all 17 metabolites, as determined by their AUC, is shown in Table 2 along with their cutoff concentration values and their respective sensitivity, (1-specificity), and AUC.

Because the most significant metabolite, Cr+PCr, is ideal ($AUC = 1.0$), as can be seen from Table 2, and because the next three most significant metabolites are nearly ideal (Gln, Cr, and NAA have an $AUC > 0.98$), we developed a diagnostic biomarker based on these four metabolites by using their respective cutoff concentration values. Our diagnostic biomarker computer program compares the concentration values of those four metabolites of an unknown mouse obtained from ^1H NMR spectroscopy with the respective ROC curve cutoff concentration values for these metabolites and makes a diagnosis based on whether the four metabolite concentrations are normal or not. For example, if the concentration of Cr+PCr of the unknown mouse is $> 9.344 \mu\text{mol/g}$, it is abnormal; otherwise, it is normal. If all four metabolite concentrations are normal, for a particular mouse, then the mouse is classified as normal; otherwise, it is diagnosed as R6/2.

Linear discriminant analysis—Linear discriminant analysis (LDA) is a mathematical multivariate method that, given a number of IVs, can render a differential diagnosis and make predictions about group membership (Portney and Watkins, 2000; Tabachnick and Fidell, 2007). It does so by constructing one or more predictive equations, termed discriminant functions. The number of discriminant functions is always one less than the number of the groups. A discriminant function is linear and has the following general form: $D = m_1X_1 + m_2X_2 + \dots + m_nX_n + b$. The outcome of the dependent variable is D, and $X_1, X_2, \dots, X_n$ are the independent variables. In the case of LDA, the dependent variable must be

discrete, whereas the IVs ($X_1, X_2, \dots, X_n$) must be continuous and normally distributed. LDA generates the constant b (y-axis intercept) and the weights (coefficients) $m_1, m_2, \dots, m_n$ of the discriminant function in such a way that, given a set of IVs, the discrimination between the two or more groups is maximum. LDA has been used extensively throughout all research areas (Yan and Zheng, 2008; Wang et al., 2008; Mu et al., 2008; Leng, 2008; Dai et al., 2008; Chaparro et al., 2008; Chan et al., 2008). LDA is the most stringent of all multivariate methods (Lei and Koehly, 2003), in that a number of arduous conditions must be satisfied before it can be appropriately applied. A list of those LDA conditions according to Tabachnick and Fidell (2007) is as follows:

1. All IVs must be normally distributed.
2. The dependent variable must be discrete.
3. There must be no multicollinearity between/among the IV's.
4. Homogeneity of variance: the within-group variances must be equal for every IV.
5. Homogeneity of covariance: the within-group covariances must be equal for every IV.
6. Linearity: there must exist linear relationships among all pairs of IVs within each group.
7. Sample sizes: the number of IVs should be smaller than the sample size of the smallest group.

Of these conditions, the first was addressed and met earlier in the ROC curve analysis. Our dependent variable is dichotomous, that is to say it has two groups, labeled 0 (WT) and 1 (R6/2), and it is, therefore, discrete, which satisfies the second condition.

We first performed the LDA with all of the 17 IVs in order to assess which of those were significant in the construction of the discriminant function. Looking at the variable influence report, and by using the Wilks' λ as our guide, we found that 10 of the IVs were significant in the construction of the discriminant function. In LDA, the Wilks' λ is the statistic that tells one how much of the variance of the dependent variable would not be accounted for if one were to use a particular independent variable; therefore, the smaller the Wilks' λ for a given IV, the more significant that IV is to the discriminant function. Because, therefore, we determined that 10 of the 17 IVs were the most significant to the discriminant function, we then performed the LDA by using just those 10 IVs and again examined the variable influence report to ensure that they were all significant. Those 10 metabolites are as follows in descending order of significance (ascending order of Wilks' λ): Cr+PCr, Gln, Cr, NAA, GSH, GPC+PC, PCr, mIns, PE, and Tau.

It is worth noting that these 10 metabolites are precisely the same as the top 10 metabolites determined by the ROC curve analysis by using the AUC as the criterion of importance (Table 2). This speaks to the significance of these 10 metabolites in the neuropathology of Huntington's disease in R6/2 mice given the fact that ROC curve analysis is a completely different mathematical approach from LDA.

Multicollinearity: Looking again at the variable influence report of the LDA with the 10 IVs, we noted that three IVs (Cr+PCr, Cr, and PCr) had an R^2 (Other X's) = 1.00. This is the R^2 value that one would get if the particular IV were regressed on all other remaining IVs. In essence, the R^2 (Other X's) value tells us what percentage of the particular IV's variance can be accounted for by the remaining IVs. This value should be less than 0.99; otherwise, there exists a multicollinearity issue (Hintze, 2005). To state it equivalently, a variable must contain >1% of new information—information that is not contained by the rest of the variables—in order not to cause multicollinearity issues. In our case, that the aforementioned

three IVs had an R^2 (Other X's) = 1.00 is hardly surprising because the Cr+PCr is the total creatine of the system (free + bound), and it is obviously related to both Cr and PCr. Given the redundancy that exists among those three IVs in terms of variance, we dropped PCr, the least significant (highest Wilks' λ) of the three creatine-related variables, from the LDA. We ran the LDA without the PCr, i.e., with the remaining nine IVs. We found that all R^2 (Other X's) values were less than 0.96, indicating that there was no multicollinearity among the nine remaining variables. Multicollinearity is one of the most important issues in multivariate analysis, and it can significantly impact the accuracy, and, therefore, the predictability of a model based on multivariate analysis (Concato et al., 1993; Hintze, 2005; Tabachnick and Fidell, 2007).

Sample sizes: The number of IVs used in the LDA should be smaller than the sample size of the smallest group (Tabachnick and Fidell, 2007). In our case, the size of the smaller group is 13 (13 R6/2 and 17 WT mice), and we ended up with 9 IVs. Therefore, the sample size condition is met.

Homogeneity of variance and covariance: We next tested the within-group variances in each of the nine IVs. We used the following two tests: 1) variance-ratio equal-variance test; and 2) modified-Levene equal-variance test. We required that each of the nine IVs pass both of those tests. The following IVs failed that requirement: Cr+PCr, GPC+PC, PE, and Tau. We then investigated the within-group covariances for equality in each of the nine IVs. To that end, we used the following two tests: 1) Bartlett's test; and 2) Box's M test. We required that each IV pass both of those tests. The Bartlett's test checks for equality of group covariances within an IV separately, whereas the Box's M test checks for equality of group covariances in all IVs together. The Box's M test gave us $F = 1.90$ and $F_{\text{prob}} = 0.000327$, indicating that the group covariances were significantly different when all nine IVs were examined together. The Bartlett's test revealed that four of those IVs were responsible for the inequality of the group covariances. Those four were: Cr+PCr, Gln, GPC+PC, and Tau. We subsequently transformed them so as to attain equality of group covariances. We should point out here that PE was also transformed because, as was mentioned earlier, it did not meet the homogeneity of variance condition. We then tested again all nine IVs, i.e., T(Cr+PCr), T(Gln), Cr, NAA, GSH, T(GPC+PC), mIns, T(PE), and T(Tau), using the transformed forms of the five aforementioned variables (their name is preceded by the letter "T"). The Box's M test this time gave us $F = 1.31$ and $F_{\text{prob}} = 0.080320$, indicating that there was no significant difference among the group covariances when all nine IVs were examined together. The Bartlett's test also showed that all of the group covariances were equal when each IV was examined separately. The equality of covariance condition of LDA was thus met. We then tested all of the five aforementioned transformed IVs for homogeneity of group variances. They all passed both the variance-ratio equal-variance test and the modified-Levene equal-variance test. Therefore, the homogeneity of variance condition was also met.

Linearity: We compared scatter plots of every possible combination of IV pairs, and we looked for any curvilinear or nonlinear patterns within each group. All of the scatter plots were linear.

Having thus met all the required conditions, we ran the LDA with the nine IVs (five transformed). From the predicted classification section, we noted that for one animal the value for GSH was undetermined from the spectral profile. LDA will ignore an entire row of data if one or more values are missing. In our case, that was a costly incident given that mouse #22 was missing one value out of nine. That meant that the valuable information included in those eight metabolite concentrations could not be used in the construction of the

discriminant function, and that a loss of accuracy to some extent would be incurred unless we made the effort to impute the missing value.

Missing values: We used the mean GSH value of the R6/2 mice to impute (generate) the missing value. Having thus generated the missing value, we ran the LDA, and we noted that all of the 30 rows of our data were used, and that the new discriminant function was an improvement over the previously determined function. The Wilks' λ for the discriminant function decreased from 0.047896 to 0.046941; the Eigenvalue increased from 19.8786 to 20.3032; and the canonical correlation² increased from 0.9521 to 0.9531. In LDA, canonical correlation² = 1 – Wilks' λ for the discriminant function. Therefore, the canonical correlation² tells us how much of the variance of the dependent variable is accounted for by the IVs we selected for the LDA model—in our case, 95.31%.

Normality: We retested the five transformed variables plus GSH for normality using the same tests as mentioned earlier in the ROC curve analysis. All of the six IVs under consideration passed the normality tests.

Logistic regression analysis—Logistic regression analysis (LRA) is a mathematical regression method that, given a number of independent variables, can render a differential diagnosis by calculating the probability of a group membership. More specifically, given a number of independent variables, and given a reference group (the normal group, “0”) and a target group (the group with the condition, “1”), LRA can tell us what the probability is of a given subject's belonging to the target group. LRA calculates the probability of membership in the target group via the logistic function. LRA constructs the logistic function, which is linear, from the input IVs according to the following equation:

$$Z = m_1 X_1 + m_2 X_2 + \dots + m_n X_n + b \quad (1)$$

where $X_1, X_2, \dots, X_n$ are the IVs, b is a constant (y-axis intercept), and $m_1, m_2, \dots, m_n$ are the weights (coefficients) that LRA calculates and assigns to the IVs. Unlike linear and multiple regression, which use the least squares method for calculating the best function (the function with the smallest residuals), LRA calculates the weights and constructs the best logistic function by using the method of maximum likelihood. According to that method, the best logistic function is the one that has the best odds of making accurate predictions about target group membership.

In Equation (1), Z is the outcome of the dependent variable, but, more specifically, according to the maximum likelihood method, Z is the natural logarithm of the odds of belonging to the target group (O_{TG}):

$$Z = \ln(O_{TG}) \quad (2)$$

The odds of the occurrence of an event (O) are related to the probability (P) of the occurrence of that event according to the following formula:

$$O = P / (1 - P) \quad (3)$$

From Equations (2) and (3), we can see that Z is related to the probability of belonging to the target group (P_{TG}) as follows:

$$Z = \ln[P_{TG}/(1 - P_{TG})] \quad (4)$$

From Equation (4), if we solve for P_{TG} , we get the following equation:

$$P_{TG} = 1/(1 + e^{-Z}) \quad (5)$$

Equation (5) is the one that allows us to obtain the probability of target group membership from the Z value calculated by the logistic function (Eq. 1). Once again, Z is the outcome of the dependent variable, and it is often referred to as “logit” (Portney and Watkins, 2000).

It is worth noting here that, as a probability, $P_{TG} \in [0, 1]$, that is to say it takes values from 0 to 1, and that the cutoff probability value is 0.5 (50%). If $P_{TG} > 0.5$, then the diagnosis is positive, i.e., the subject has the condition and belongs to the target group. If, on the other hand, $P_{TG} < 0.5$, then the diagnosis is negative, i.e., the subject does not belong to the target group, which means that the subject belongs to the reference group, or that the subject is normal. In the case where $P_{TG} = 0.5$, the diagnosis is incomplete, and the subject's status is undetermined; in this case, P_{TG} is equal to the chance probability. We should also point out that whereas $P_{TG} \in [0, 1]$, $Z \in (-\infty, +\infty)$, in other words, the logit's range is from $-\infty$ to $+\infty$. Next, let us consider a few examples. When $P_{TG} = 0.99$, then from Equation (4), one can obtain that $Z = 4.595$. When $P_{TG} = 0.01 \rightarrow Z = -4.595$. Finally, when $P_{TG} = 0.5 \rightarrow Z = 0$. Therefore, one can see that according to the logit (Z), the diagnostic cutoff value is 0. In other words, when $Z > 0$, the subject belongs to the target group; and when $Z < 0$, then the subject is normal. This is an observation that we will exploit in the construction of our LRA CCA (clinical change assessment) model.

LRA is one of the most widely used regression methods. The range of applications of LRA extends from natural selection and microevolution to family and other social studies and to a comparison with LDA (Morgan and Teachman, 1988; Janzen and Stern, 1998; Fan and Wang, 1999).

In LRA, the dependent variable has to be discrete. In our case, we have two groups, WT (coded “0”) and R6/2 (coded “1”), so our LRA dependent variable takes those two values, i.e., 0 and 1. Moreover, because we have only two groups, the analysis yields one logistic function. If we had more than two groups, let us say n groups, then LRA would yield $(n - 1)$ logistic functions. In that case, the analysis would be called multiple-group logistic regression (Hintze, 2005) or multiple logistic regression (Menard, 2000). The fact that the dependent variable has to be discrete is not as limiting as it may sound because a continuous variable may be converted to a discrete one (Tabachnick and Fidell, 2007). Insofar as it concerns the type of the IVs, LRA permits a significant amount of flexibility; the IVs may be continuous, discrete, dichotomous, or any combination thereof. This flexibility notwithstanding, the following LRA conditions if unmet may compromise the accuracy of LRA (Bagley et al., 2001; Tabachnick and Fidell, 2007; Hintze, 2005):

1. The dependent variable must be discrete.
2. Linearity: although the IVs are not required to have a linear relationship among themselves, those IVs that are continuous have to have a linear relationship with the logit (Z value).
3. There must be no multicollinearity between/among the IVs.

4. Sample sizes: the empirical rule here suggests that the number of cases of the target group divided by the number of the IVs chosen for the analysis is greater than 10 (Peduzzi et al., 1996; Bagley et al., 2001).

As was discussed earlier, the first condition is met, and we will now consider the remaining conditions, starting with the last condition first.

Sample sizes: Because the number of our R6/2 mice is 13, if we were to meet that empirical rule pertaining to the recommended sample sizes, we would have to use only one IV in our LRA. In other words, the only way that the ratio $13/(\# \text{ of IVs}) > 10$ is true is when $(\# \text{ of IVs}) = 1$. Selecting one IV, albeit the most significant, to carry out our LRA would not be a wise choice because according to our ROC curve analysis, there are nine IVs with an AUC > 0.8, and those are the most important ones, meaning that each one of them carries a significant amount of information. If we left one of them out of our analysis, the accuracy of our results would be reduced, but if we left eight IVs out of nine out, then the accuracy of our results would be decimated. Being fully aware that, for reasons beyond our control, we could not meet the empirical rule about sample sizes, we decided to proceed with the LRA for the following reason. A smaller than required sample size lowers the statistical power. One way to increase the statistical power is to increase, somehow, the magnitude of the observed effect. In our case, we used ^1H NMR spectroscopy with a high-intensity magnetic field of 9.4 Tesla, which gave us a high resolution, which in turn resulted in the “magnification” of the striatal metabolite differences between the WT and R6/2 mice, or, to put it differently, in the increase of the magnitude of the observed effect. Moreover, in connection with that, and in order to ensure the realization of the offset of any possible accuracy loss (due to a smaller sample size) by the magnification of the observed effect, we used in our LRA the top nine most significant IVs (with AUC > 0.8; Table 2), i.e., those IVs that had the highest discriminating power between the WT and the R6/2 mice.

Multicollinearity: As we saw in the case of LDA, there is a severe multicollinearity among the following three IVs: Cr+PCr, Cr, and PCr. That means that we cannot use all three of them together. For reasons explained in the corresponding LDA section, we left PCr out of the model.

Linearity: In order to examine the linearity of each of the remaining eight IVs with respect to the logit (Z), we plotted each of those eight IVs against the Z scores for all of our 30 mice (13 R6/2 and 17 WT). All of those eight scatter plots were linear.

Missing values: As was the case with LDA, LRA omits an entire row of data if one or more values are missing. We therefore used the same method as described above in the LDA section to impute the one missing value from our data.

The eight IVs that we used to run our LRA are, in order of significance according to Table 2: Cr+PCr, Gln, Cr, NAA, GSH, GPC+PC, mIns, and PE.

Clinical change assessment models

ROC curve analysis—Because we were interested in developing a clinical change assessment tool that would assess not only the baseline level but also the progression of the disease of a subject with Huntington's disease, we included all metabolites that were affected by the disease and accurately indicated the changes of the disease throughout its course. We chose, thus, all metabolites whose AUC > 0.85. The decision on the >85% AUC was based on considerations regarding accuracy—a predictor with an AUC > 85% is considered good to excellent. According to Table 2, the metabolites that satisfy this requirement are: Cr+PCr, Gln, Cr, NAA, GSH, GPC+PC, mIns, and PCr.

We wrote a computer program that, upon reading the ^1H NMR spectroscopy concentrations of the above eight metabolites, assesses the status of each one separately, based on their individual ROC curve cutoff concentration values, and renders a decision as to whether the given metabolite concentration is normal or not. If it is not normal, then it would proceed to calculate the percentage of its deficit based on how far its concentration value is from the ROC cutoff point. As can be seen from Figure 3, the ROC cutoff values are also the threshold points to being normal. For example, if a subject has $[\text{PCr}] < 3.792 \mu\text{mol/g}$, then its PCr level is normal, but if its $[\text{PCr}] \geq 3.792 \mu\text{mol/g}$, then it is abnormal. This ROC cutoff value of 3.792, in the example of PCr, can be used as one of the two reference points required to determine the magnitude of the % deficit of the PCr level in the case of an abnormal subject (R6/2 mouse). More specifically, it can reference a given PCr concentration to the normal value because it represents the 0% deficit point. We need, however, one more reference point in order to determine the extent of the disease progression. For this reference point we can determine the worst possible (maximum) level of PCr because that would represent the 100% deficit point. By taking the mean value of PCr for the R6/2 mice, and by adding to it 3 SD (3 standard deviations of PCr), we get a value, which according to Chebyshev's Theorem (Freund, 1984), is inclusive of 99.73% of the R6/2 population with respect to the worst (in this case, largest) PCr value. So, if an R6/2 mouse has $[\text{PCr}] = C$, then, as illustrated in Figure 3, its PCr % deficit is given by:

$$\text{Metabolite \% deficit} = [(C - R) / (C_{\max} - R)] \times 100 \quad (6)$$

As shown in Table 3, in the case of PCr, R is the ROC cutoff value for PCr ($3.792 \mu\text{mol/g}$) and $C_{\max}$ ($6.628 \mu\text{mol/g}$) is the R6/2 PCr mean + 3 SD. In the case of NAA, we need to use $C_{\min}$ ($4.642 \mu\text{mol/g}$), which is obtained by subtracting 3 SD from the mean value of NAA of the R6/2 mice, because the direction of the condition is opposite to that observed in the rest of the metabolites. (R6/2 mice have lower [NAA] values than WT ones.) Moreover, in the case of metabolites such as NAA, $R > C$ and $R > C_{\min}$. This means that both the numerator and the denominator in Equation (6) will be negative, and that the metabolite % deficit will therefore be positive. In the case of metabolites such as NAA, therefore, the above relationship still holds true provided one uses $C_{\min}$ instead of $C_{\max}$. Table 3 shows all eight metabolites used by our ROC clinical change assessment model, including the respective R6/2 R and $C_{\max}$ ($C_{\min}$ for NAA) values.

Our computer program assesses individually each of the eight metabolites based on the above method. It also calculates the overall metabolite % deficit by determining the average value of the eight individual metabolite % deficits. In order to further bolster the accuracy of the overall metabolite % deficit, the eight individual metabolite % deficits are assigned weight factors, which are the respective AUCs of the eight metabolites. So, in the case of Cr + PCr and PCr, for example, the former—owing to its AUC value, which is larger than that of PCr—has a bigger influence in the overall metabolite % deficit than the latter.

This clinical change assessment can be used to obtain a baseline level of the status of the disease, and it can also be used in the future to assess the progression of the disease.

We should point out here that this ROC CCA (clinical change assessment) model is designed to assess the individual metabolite % deficit and the overall metabolite % deficit of R6/2 mice only. Because both of those assessments are predicated on how far an R6/2 mouse is from being normal (i.e., how much larger or smaller are the concentrations of those eight metabolites of an R6/2 mouse from their respective ROC cutoff points), WT mice have, by definition, zero individual metabolite % deficit and zero overall metabolite % deficit, which makes sense physiologically.

Linear discriminant analysis—Based on the same concept as above, and using LDA, we developed an additional method to assess clinical change. As was discussed before, our LDA DBM can read the ^1H NMR spectroscopy concentrations of nine metabolites of a given mouse and return the canonical discriminant score (D) of that mouse. If $D < 0$, then the mouse is normal, and if $D > 0$, then the mouse is an R6/2 one. Just like in the case of a ROC cutoff value, the $D = 0$ here represents the threshold value to being normal. The canonical discriminant score, therefore, can also be used to assess how far an R6/2 mouse is from being normal—how far its D score is from 0. As was the case using the ROC curve analysis, here, too, we need another fixed reference point in order to measure the % deficit. By using Chebyshev's theorem again, we can calculate what the worst levels of the nine metabolites could be by adding to their mean values 3 SD (subtracting 3 SD in the cases of NAA and PE). This guarantees that we have included 99.73% of the R6/2 population in the direction of the severity of the disease. We can then use those nine metabolite concentrations as the values of a hypothetical R6/2 mouse (a mouse in the worst state of Huntington's disease) to calculate the D_{MAX} , which represents how far a canonical discriminant score can be from the normal ($D = 0$). As illustrated in Figure 4, the % deficit with respect to all nine metabolites of an R6/2 mouse with a canonical discriminant score D can thus be calculated from the following:

$$\text{Overall Metabolite \% deficit} = (D/D_{\text{MAX}}) \times 100 \quad (7)$$

This LDA clinical change assessment (LDA CCA) model relies completely on the LDA DBM, because it uses it to calculate D. In the ROC clinical change assessment model, although we took measures to minimize possible sources of error, we incorporated four more metabolites with AUCs in the 80%+ range, but that represents a potential, however small, for a lower degree of accuracy. It may very well be that the LDA model may turn out to be more accurate than the ROC model, but the LDA model can offer only a global assessment of clinical change. Whereas the ROC model can offer an individualized assessment of all the metabolites used, which may be particularly helpful and insightful in the early stages of the disease, when perhaps one or few metabolite levels may be affected, or when a therapy is involved.

As was the case with the ROC CCA model earlier, the LDA CCA model is designed to assess the overall metabolite % deficit of R6/2 mice only. Because that assessment is with respect to the cutoff point ($D = 0$), i.e., the threshold point of being normal, WT mice have, by definition, zero overall metabolite % deficit.

Logistic regression analysis—As was mentioned above, the diagnostic cutoff point according to the logit (Z value) is 0. If $Z > 0$, then the mouse belongs to the target group, i.e., it is an R6/2 mouse. If $Z < 0$, then the mouse is normal. Using the Z value, therefore, calculated from our LRA DBM, we wrote a computer program that can assess the clinical change of an animal or a patient based on our LRA model. In the same fashion we calculated D_{MAX} in the case of LDA, outlined in great detail in the immediately preceding section, we calculated the Z_{MAX} , which also represents the highest theoretically attainable logit by an R6/2 mouse in the worst state of Huntington's disease ($Z_{\text{MAX}} = 4.777$). As depicted in Figure 5, the % deficit with respect to all of the aforementioned eight metabolites of an R6/2 mouse with a logit value Z can, thus, be calculated from the following equation:

$$\text{Overall Metabolite \% deficit} = (Z/Z_{\text{MAX}}) \times 100 \quad (8)$$

Once again, both our LRA DBM and LRA CCA models use the following eight metabolites: Cr+PCr, Gln, Cr, NAA, GSH, GPC+PC, mIns, and PE.

The LRA CCA model, just like the ROC and LDA CCA models, is designed to assess the overall metabolite % deficit of R6/2 mice only. Because that assessment is with respect to the cutoff point ($Z = 0$), i.e., the threshold point of being normal, WT mice have, by definition, zero overall metabolite % deficit.

Results

Diagnostic biomarker models

ROC curve analysis—The rank of all 17 metabolites as determined by ROC curve analysis of their respective probabilities of the area under the curve (AUC) is listed in Table 2. Each metabolite's best ROC cutoff concentration is shown, along with its respective sensitivity and (1-specificity) values. According to ROC curve analysis, Cr+PCr, Gln, Cr, and NAA all have AUC values $> 98\%$, which is considered excellent in terms of discrimination and, hence, diagnostic power. Of those, Cr+PCr has an AUC = 100%, which is considered perfect. In Figure 2, one can observe that the ROC graph of Cr+PCr depicts the ideal ROC curve, i.e., one that has collapsed to the point [0,1]. The perfect ROC curve has the coordinates [0,1], that is to say that it has (1-specificity) = 0 and sensitivity = 1.

The ROC curves for the nine most important metabolites are shown in Figure 2. The thin aqua diagonal line represents the $y = x$ line, whose AUC is 50%. This line is the chance line, for every point on that line has the property of having (1-specificity) = sensitivity. This line, furthermore, represents the baseline for an AUC. A meaningful ROC curve should fall far above the chance line, that is to say that it should have an AUC $\gg 50\%$.

The seven least significant metabolites, also in accordance with their respective AUC value, are shown at the bottom in Table 2. It is worth noting that from the AUC for MM (0.50), one can quickly realize that if one were to use the macromolecules (MM) for diagnostic purposes, an R6/2 mouse would have only a 50% chance of being diagnosed as an R6/2 as compared with a WT mouse. That is the same probability as the one offered by the chance line. An equivalent way of stating this is to say that according to the MM IV, there is no discrimination between the two groups of mice (WT and R6/2). Given that in NMR spectroscopy the MM IV is used to represent mostly proteins, i.e., background tissue, which is considered to be the same for all subjects, this result provides further evidence for the accuracy of ROC curve analysis.

As was mentioned earlier, we used the top four most significant metabolites according to ROC curve analysis to develop our ROC diagnostic biomarker. We tested it with 31 mice that were extraneous to the model, that is to say that those mice were different from the 30 that we used to build the DBM. The status of those 31 mice had been previously determined by genotyping. Of those 31 mice, 20 were WT and 11 were R6/2. Our ROC DBM determined correctly the status of all of those 31 mice. Therefore, its sensitivity = 1 and its (1-specificity) = 0. The positive likelihood ratio [(+) LR] of our ROC DBM, which is given by

$$(+)\text{LR} = (\text{sensitivity}) / (1 - \text{specificity}) [43, 13], \text{ is: } (+)\text{LR} = 1/0 \rightarrow \infty$$

The negative likelihood ratio [(-) LR] of our ROC DBM, which is given by

$$(-)LR = (1 - \text{sensitivity}) / (\text{specificity}) [43, 13], \text{ is: } (-)LR = 0/1 = 0$$

All of the ROC curve analysis test results are summarized in Table 8.

Linear discriminant analysis—The variable influence section of our final LDA run is shown in Table 4. It should be noted (by the Alone λ , Alone F-value, and Alone F-prob) that all of our final nine IVs (metabolites) used in the construction of the final discriminant function are very significant. The Alone Lambda here refers to the Wilks' λ pertaining to a particular IV if that IV were to be used alone. It tells us how much of the variance of the DV (dependent variable) cannot be accounted for by a particular IV (independent variable) if that IV were to be used alone in the model. The smaller the Alone Lambda (Wilks' λ), the more significant a particular IV is to the model. The Alone F-value is the F ratio used to test the significance of the respective Alone λ , and the Alone F-prob is the probability of significance of the respective Alone F-value.

We constructed our discriminant function by using the canonical coefficients because we wanted our cutoff point between the two groups to be equal to zero. The canonical coefficients (weights) are generated in such a way so as to render the discriminant scores of the resultant discriminant function standardized, i.e., have a mean value equal to zero. Discriminant scores obtained thus are commonly referred to as canonical scores; however, in this study, we refer to them as canonical discriminant scores. The statistical results of our final discriminant function are shown in Table 5. According to the canonical correlation², which is equal to $1 - (\text{Wilks' } \lambda)$, over 95% of the variance of the canonical discriminant scores can be accounted for by the nine IVs we selected. The F-value and the probability level show the significance of the Wilks' λ , or the canonical correlation² of our discriminant function.

Our discriminant function is shown in Table 6. Those are the canonical coefficients for all of the nine IVs plus the constant (y-intercept). The accuracy of our LDA discriminant function's predictions with respect to the 30 original mice (those used to develop the function) is reported in Table 8. There were no misclassifications or erroneous predictions. The distance of separation or discrimination between the mean canonical discriminant scores of WT and R6/2 mice is 8.785 units. A visual representation of the magnitude of that discrimination between the two groups (WT = 0 and R6/2 = 1) can be seen in Figure 6, in which all 30 canonical discriminant scores are plotted in a straight line (D vs. D) in order to better visualize their separation. The mean discriminate score for WT mice was -3.81 ± 0.20 , and that for the R6/2 mice was 4.98 ± 0.33 , respectively ($P < 1 \times 10^{-6}$). As we mentioned above, the cutoff canonical discriminant score between the WT and the R6/2 mice is zero; WT mice have negative canonical discriminant scores, whereas R6/2 mice have positive canonical discriminant scores.

As we did in the case of the ROC DBM, we used the same 31 mice that were also extraneous to the LDA DBM model in order to test its accuracy. Our LDA DBM identified correctly the status of all of those 31 mice (20 WT and 11 R6/2) by reading their nine ¹H NMR spectroscopy concentration values. Therefore, for our LDA DBM, the sensitivity = 1 and the (1-specificity) = 0.

The positive likelihood ratio [(+) LR] of our LDA DBM is:

$$(+)LR = (\text{sensitivity}) / (1 - \text{specificity}) = 1/0 \rightarrow \infty$$

The negative likelihood ratio [(-) LR] of our LDA DBM is:

$$(-)LR = (1 - \text{sensitivity}) / (\text{specificity}) = 0/1 = 0$$

Because we were interested in finding out whether there was a significant difference in the canonical discriminant scores between our original seven R6/2 mice that had been scanned at 8 weeks of age and our original six R6/2 mice that had been scanned at 12 weeks of age, we performed an independent t-test (because they were two different groups), and we graphed the results, which appear in Figure 7. Those are the 13 R6/2 mice (7+6) whose data we used to develop the LDA DBM. As can be seen from Figure 7, there is a significant difference in the canonical discriminant scores between the 8-week-old R6/2 and the 12-week-old R6/2 mice (mean values of 4.18 ± 0.28 and 5.91 ± 0.36 , respectively) ($P = 0.002959$). More specifically, the canonical discriminant scores of the latter are significantly greater than those of the former. Given that a higher canonical discriminant score is indicative of a worse state with respect to the disease, this result is in agreement with the pathophysiological progression of Huntington's disease.

In other words, one would expect that the striatal metabolite profiles of the R6/2 mice would significantly change as the progression of the disease advances in the age interval from 8 to 12 weeks. By virtue of the same argument, one would expect no significant change in the striatal metabolite profiles of normal (WT) mice during the same age interval. This notion is also in agreement with our findings: there was no significant difference in the canonical discriminant scores between the 8-week-old and the 12-week-old WT mice (mean values of -4.02 ± 0.35 and -3.62 ± 0.22 , respectively; $P = 0.343401$; Fig. 7). Furthermore, as can be seen in Figure 7, all possible combinations of comparisons between the two R6/2 groups of mice and the two WT groups of mice are statistically significant ($P < 1 \times 10^{-6}$). All of the LDA test results are summarized in Table 8.

Logistic regression analysis—The logistic function constructed from our LRA is shown in Table 7. How well the constructed logistic function fits the data of the 30 mice used in the development of the model is depicted in Table 8. More specifically, Table 8 shows that our model correctly classified all of the original 30 mice.

In order to better visualize how well our two groups (R6/2 and WT) were discriminated by the logistic function of our LRA, we plotted the logit values of all 30 mice versus themselves (Z value vs. Z value). What this graph does is to plot all the Z values in the same straight line ($y = x$) so one can better visualize and compare the magnitude differences of the Z values without any spatial distractions (Fig. 8). The mean logit value for WT mice was -1.91 ± 0.10 , and that for the R6/2 mice was 1.87 ± 0.15 ($P < 1 \times 10^{-6}$). As we mentioned previously, the cutoff logit value between the WT and the R6/2 mice is zero; WT mice have negative logit values, whereas R6/2 mice have positive logit values.

We used the aforementioned logistic function to develop our LRA DBM. Just as we did with the other two DBMs, we tested its accuracy with the 31 mice that were extraneous to this DBM, as well. Our LRA DBM identified correctly the status of all of those 31 mice (20 WT and 11 R6/2) by reading their eight ^1H NMR spectroscopy concentration values. Therefore, for our LRA DBM, too, the sensitivity = 1 and the (1-specificity) = 0.

The positive likelihood ratio [(+) LR] of our LRA DBM is:

$$(+)\text{LR} = (\text{sensitivity}) / (1 - \text{specificity}) = 1/0 \rightarrow \infty$$

The negative likelihood ratio $[(-)\text{LR}]$ of our LRA DBM is:

$$(-)\text{LR} = (1 - \text{sensitivity}) / (\text{specificity}) = 0/1 = 0$$

Just as we did with the other two DBMs, we investigated for differences between our two R6/2 groups. We calculated and compared the logit values of the seven 8-week old R6/2 mice and the six 12-week-old R6/2 mice. The results are shown in Figure 9. The 12-week-old group has significantly higher logit values than the 8-week-old group (mean values of 5.91 ± 0.36 vs. 4.18 ± 0.28 , respectively) ($P = 0.03249$), and because a higher logit value signifies a worse state of the disease, this result, too, is in agreement with the pathophysiological progression of Huntington's disease. We also calculated and compared the logit values for our two WT groups (8-week-old and 12-week-old; Fig. 9). We found no significant difference in terms of their logit values (mean values of $7-1.99 \pm 0.17$ and -1.83 ± 0.11 , respectively; $P = 0.418088$), which is what one would expect, and which, furthermore, is also in agreement with the respective findings of the ROC and LDA DBMs. Furthermore, as can be seen in Figure 9, all possible combinations of comparisons between the two R6/2 groups of mice and the two WT groups of mice are statistically significant ($P < 1 \times 10^{-6}$). All of the LRA test results are summarized in Table 8.

Clinical change assessment models

ROC curve analysis—As was discussed in Materials and Methods, our ROC clinical change assessment (ROC CCA) model uses the top eight most significant metabolites, according to their AUCs derived from the ROC analysis, and calculates their individual % deficit by using Equation (6). As previously discussed in Materials and Methods, our ROC CCA model further calculates the overall metabolite % deficit based on the weighted average of the individual deficits of those eight metabolites. This may be particularly useful in a clinical setting, wherein a patient's progress, following a given therapy, can be assessed not only globally but also with respect to specific metabolites. It may be of great importance, for example, to know which metabolites have become normalized, which are progressing toward normalization, which remain unchanged, and which have become worse, as this information may be used to determine the success or failure of the future course of a specific therapy.

We next tested our ROC CCA model by using it to calculate the overall metabolite % deficit of all of our 13 R6/2 mice (7 at 8 weeks and 6 at 12 weeks), and by examining whether there was a significant difference between these two groups. As depicted in Figure 10, one can see that the overall metabolite % deficit of the 12-week-old R6/2 mice is significantly greater than that of the 8-week old-R6/2 mice. More specifically, the overall metabolite % deficit for R6/2 mice at 8 weeks of age was $19.5 \pm 1.5\%$, and was significantly more severe in 12-week-old R6/2 mice at $38.0 \pm 2.6\%$ deficit; $P = 0.000043$). This is in agreement with what was observed about those two groups by using the LDA DBM to calculate and compare their canonical discriminant scores and also by using the LRA DBM to calculate and compare their logit scores. As was mentioned earlier, WT mice have 0% overall metabolite deficit by definition because their concentration values of the eight metabolites used by the ROC CCA model are in the normal side of the respective ROC cutoff points for those eight metabolites.

Linear discriminant analysis—Our LDA clinical change assessment (LDA CCA) model uses the LDA DBM computer program to obtain the canonical discriminant score of a mouse, and if the score is positive (R6/2), then, it calculates the overall metabolite % deficit for that mouse by using Equation (7). WT mice have 0% overall metabolite deficit by definition because their canonical discriminant score is negative, i.e., in the normal range.

Here, too, we wanted to determine the overall metabolite % deficit of all 17 R6/2 mice according to our LDA CCA model, and here, too, we were interested in knowing whether there was any significant difference between the 8-week-old and the 12-week-old R6/2 mice. As shown in Figure 10, there is a significant difference between the two groups: the 12-week-old R6/2 mice have significantly higher overall metabolite % deficits as determined by their canonical discriminant scores than those of the 8-week-old R6/2 mice (mean values of $55.98 \pm 3.45\%$ deficit vs. $39.53 \pm 2.66\%$ deficit, respectively; $P = 0.002776$), which is in agreement with our findings using ROC curve analysis.

Logistic regression analysis—In a manner similar to that of the LDA CCA model, our LRA CCA model avails itself of the computer program of the LRA DBM to calculate the logit value of a mouse, and if the value is positive, which is to say if the mouse is an R6/2, then it calculates the overall metabolite % deficit for that mouse by using Equation (8). WT mice have 0% overall metabolite deficit by definition because their logit value is negative, i.e., in the normal range.

Next, we used our LRA CCA model to calculate the overall metabolite % deficit for all 17 R6/2 mice, and we compared the deficits of the 8-week-old with those of the 12-week-old mice. The results are shown in Figure 10. The 12-week-old R6/2 mice have significantly higher overall metabolite % deficits than those of the 8-week-old R6/2 mice (mean values of 46.21 ± 4.25 vs. 33.18 ± 3.34 , respectively; $P = 0.03249$). This finding is in agreement with our respective findings from both the ROC and LDA CCA models.

Discussion

The results from our study demonstrate that ROC, LDA, and LRA are robust mathematical approaches that can be used to identify biomarkers from NMR spectra for experimental Huntington's disease, and to monitor disease progression. Insofar as quantitative metabolite observations are concerned, our results are in partial agreement with Jenkins et al. (2005), who, using 4.7-Tesla NMR equipment, investigated the in vivo striatal and cortical metabolite changes in a number of Huntington's disease animal models, including R6/2. They found that in the case of 10-week-old R6/2 mice, there was a significant decrease in NAA and significant increases in Cho (choline), Tau, and Lac in comparison with the WT mice. All of the aforementioned metabolites, however, were not compared as absolute individual concentration values; they were expressed relative to the Cr concentration value. As we showed, both NAA and Cr are two of the top four most significant biomarkers in Huntington's disease—the former is significantly lower in the R6/2 mice, whereas the latter is significantly higher. Therefore, if we examine the ratio NAA/Cr, it will certainly be significantly lower in the R6/2 mice than in the WT mice. We were not able to quantify the absolute value of the free choline (Cho). Nevertheless, we were able to isolate and measure the total phosphorylated choline (GPC+PC), and, as we showed, it was significantly higher in the R6/2 mice. Regarding Lac, we did not find any significant differences between the R6/2 and the WT mice.

Interestingly enough, the aforementioned study did not observe a significant increase in Gln in the R6/2 mice; that, however, may also be attributed to comparing ratios of metabolites instead of comparing absolute individual concentrations of metabolites. As we showed, both

Gln and Cr are two of the top three most significant bio-markers in Huntington's disease in R6/2 mice, and both of them are significantly higher in the R6/2 mice as compared with the WT mice. If we therefore examined and compared the ratio Gln/Cr in both the R6/2 and WT mice, we would expect that ratio not to be significantly different between the two groups.

Behrens et al. (2002), by using microdialysis and high-performance liquid chromatography, showed that both Gln and Tau were significantly higher in 9-week-old and 12-week-old R6/2 mice than in the corresponding age groups of WT mice. They also showed that the striatal and cortical extracellular Glu was not significantly different in the two mice groups (R6/2 and WT), which is in agreement with our finding.

Other studies using NMR spectroscopy have used various mathematical approaches for the identification of biomarkers in other diseases. For instance, by using ex vivo NMR spectroscopy (600-MHz spectrometer), Odunsi et al. (2004) scanned the serum samples of patients with epithelial ovarian cancer, patients with benign ovarian disease, postmenopausal controls, and premenopausal controls. First, by employing principal component analysis (PCA) to reduce the dimensionality of their data, i.e., to identify those IVs that were significant in the construction of the first principal component; second, by using ROC curve analysis on the IVs significant to the first principal component in order to assess their discriminating power; and finally by employing logistic regression analysis with those IVs determined to be significant by the ROC curve analysis, they were able to classify correctly the serum samples of the aforementioned four groups. According to the results of our third companion study (Investigation of Principal Component Analysis in Connection with the Diagnosis of Diseases [in preparation]), in which we investigated PCA, unsupervised PCA is neither a consistent nor a reliable method for diagnostic purposes.

Furthermore, unsupervised PCA, either with the covariance matrix or with the correlation matrix, gave us a rank of significance of our IVs (as determined by the absolute value of the weights of the eigenvector) that was consistently different from all of our other mathematical approaches. We should also point out here that all of our mathematical approaches (ROC curve analysis, LDA, and LRA) were in complete agreement with respect to the rank of significance of our IVs, and that all of them classified correctly not only all of the subjects intrinsic to their development but also all of the unknown subjects (extraneous to their development), whereas unsupervised PCA failed to do so. Therefore, using unsupervised PCA to identify and select the most significant IVs may not be the optimal approach. In their study, Odunsi et al. (2004) were able to classify correctly the serum samples of the four groups they investigated, i.e., they were able to show that their analysis correctly classified the subjects that were intrinsic to the development of their analysis. That is only the first necessary step, however. The most important test in the case of a DBM is to demonstrate the capability to classify correctly unknown subjects—subjects that are extraneous to the development of the model/analysis—and that is something that the authors of that study did not show.

We should also point out here that, according to the results of our second companion study (Nikas et al., 2010), there are other data analysis/mining methods besides PCA, such as Fuzzy clustering analysis, that passed all of our three tests, thus demonstrating that they, too, may be used for the diagnosis of diseases, for which a high degree of accuracy is necessitated.

In our previous study (Tkac et al., 2007), by applying typical statistical tools, such as t-tests, we were able to make numerous discrete observations about the differences between the 8-week-old and 12-week-old R6/2 and the 8-week-old and 12-week-old WT mice in connection with the 17 metabolites detected and measured by 9.4-Tesla ^1H NMR

spectroscopy. Examining all the possible combinations of comparisons between the R6/2 and WT mice with respect to the 17 IVs and the two-group time variable (8-week-old and 12-week-old mice) is an exhaustive process. If one takes into account that in the near future, technological improvements will result in NMR spectrometers of considerably higher resolution (both in vivo and ex vivo), which would inevitably entail a considerable increase in the number of biomolecules detected and measured, then one can arrive at the conclusion that conventional statistical tools would be inadequate at best. The amount of information generated by high-resolution NMR spectroscopy (16 Tesla) would necessitate far more sophisticated analytical approaches. In that study (Tkac et al., 2007), we were able to ascertain significant differences between our two groups of mice (R6/2 and WT) by using conventional methods.

We did not have the capability, however, to 1) assess and compare the 17 IVs; 2) determine which of those were significant in the discrimination between the R6/2 and WT mice and could be used as biomarkers of Huntington's disease; 3) develop the means to diagnose unknown subjects; and 4) develop the means to monitor the progression of Huntington's disease in mice or the efficacy of an applied intervention. All of those tasks, however, are precisely what we were able to accomplish here in the present study, and that distinguishes this study from other studies of a similar nature. In fact, currently, the vast majority of studies (either in the area of neurodegenerative diseases, or in cancer diseases, or in heart or any other diseases) rely on conventional statistical tools to make a number of observations about the object of interest—but those conventional statistical tools are inherently limited. Finally, because of all of the aforementioned capabilities of our analytical mathematical approaches, and because of their universality, we can apply them not only in Huntington's disease and with in vivo NMR spectroscopy, but also in any disease using any type of technology.

Diagnostic biomarker models

We subjected our ROC, LDA, and LRA DBMs (diagnostic biomarkers) to the first two of the following three tests directly and to the third test indirectly. More specifically, because we used the aforementioned DBMs to develop CCA models, we subjected their respective CCA models to the third test.

1. All of our DBMs were tested with our original 30 mice, which were intrinsic to the development of all of them.
2. All of our DBMs were tested with the 31 unknown (test) mice, which were extraneous to their development.
3. All of our DBMs were tested with our 13 R6/2 mice indirectly, i.e., via their respective CCA models. Seven of those R6/2 mice were scanned at the age of 8 weeks and six of those R6/2 mice were scanned at the age of 12 weeks. A mathematical model, be it a DBM or a CCA, should be able to discriminate between those two groups, that is to say it should be sensitive enough to pick up the metabolomic differences between those two groups effected by the progression of the disease in the time span of 4 weeks.

With respect to the aforementioned three tests, all three DBMs presented a positive likelihood ratio approximating infinity [$1/0 \rightarrow \infty$] and a negative likelihood ratio equal to zero [$0/1 = 0$].

More specifically, the results from our study (Table 8) demonstrate that all of our diagnostic biomarker models (ROC, LDA, and LRA DBMs) were perfectly accurate after we tested them with 31 unknown mice and compared their results with genotyping. The ability to genotype these animals allowed us to validate our approach for developing our diagnostic

biomarker models and points to the potential application of ROC curve analysis, LDA, and LRA for NMR-spectroscopy-based diagnostics in the field of medicine.

Clinical change assessment models

We developed three different CCA models, all of which we found to be very accurate. The ROC CCA model can measure the overall metabolite % deficit of patients compared with the normal levels, and in addition it can measure the individual metabolite % deficit. The LDA and LRA CCA models can measure only the overall metabolite % deficit. The way that all CCA models can be used together in a complementary fashion is to use the LDA CCA and the LRA CCA models for a global monitoring of a disease or for a global monitoring of the progress of a therapy and then to use the ROC CCA model to obtain a complete profile of all individual metabolite changes, which may be used by physicians to make decisions about what treatment they should use or about whether to continue or not an ongoing treatment.

We should point out here that all three models—ROC, LDA, and LRA CCA—are fundamentally different in nature. The first one is derived from and predicated upon a theory of probabilities; the second one is the result of a theory of multivariate analysis; and the third one is the result of a theory of regression based on the concept of maximum likelihood. Moreover, the first and third models use eight metabolites; whereas the second one uses nine, of which seven are common to all. This means that each of the three models calculates the overall metabolite % deficit in accordance with its own intrinsic approach. Assessments calculated by each model are, therefore, meaningful and accurate insofar as they are used for evaluation and comparison purposes only within the domain of that same model. For example, an assessment (the overall metabolite % deficit) of a given mouse when obtained by using one model is bound to be numerically different from that obtained by using another model. The methods and the scales employed by the three models are different, but the physical quantity that is being measured is the same. Although individual assessments made by the three models are expected to be numerically different, the physiological reality is such that the three models should render identical measurements at the two limits. When, for instance, we enter the metabolite values of any normal mouse, all three models give the same answer: 0 % deficit. When we also enter the metabolite values of the mouse in the worst possible condition of Huntington's disease (R6/2 mean values $\pm$ 3 SD), all three models also give the same answer: 100 % deficit.

It stands to reason, therefore, that one should choose any of the three models and continue to use the same model throughout the course of the disease or the therapy when one is interested in monitoring the overall metabolite % deficit, that is to say when one is interested in a global assessment of the disease. As previously mentioned, however, there is a way whereby the models may be used together in a complementary fashion. Physicians, for example, may use the LDA CCA and/or the LRA CCA model to monitor globally the disease or the progress of the therapy, and, at the same time, they may use the ROC CCA model to obtain a complete profile of all individual metabolite changes in order to monitor or derive valuable information about local and specific changes, which would allow them to make decisions about the nature and future course of a treatment or intervention.

On the nature and applicability of our three mathematical analyses

Even though all three of our mathematical analyses and their respective final applications (DBMs and CCA models) were developed in connection with animal Huntington's disease utilizing ^1H NMR spectroscopy, they all transcend both Huntington's disease and NMR spectroscopy, i.e., they can all be applied in connection with any disease utilizing any type of technology.

ROC curve analysis

Pros: Of our three mathematical analyses, ROC curve analysis is the most versatile. Moreover, if normality, as it pertains to the IVs, is satisfied, then ROC curve analysis is very accurate as well. It may be used practically under any circumstances and in the investigation of any disease or disorder. Because there is no condition on the nature of the IVs, one may choose to use (for the development of DBMs and CCA models in connection with a particular disease or disorder) all continuous IVs, such as metabolite concentrations, temperature, body mass, dimensions of tumor cells, etc.; all discrete IVs, such as gender, race, metastasis of tumor, part of the body where incidence of a particular phenomenon occurs, education, occupation, etc.; or any combination thereof. As was mentioned earlier, ROC analysis is the only one of the four that may be used to compare two different DBMs, and it is also the only analysis of the three that may be used to couple two different outcome measure instruments or one outcome measure instrument and a CCA model.

Cons: The only limitation of ROC curve analysis lies in the nature of the dependent variable (DV). In this case, the DV has to be dichotomous, i.e., one that has only two outcomes, such as WT (0) or R6/2 (1) in our case. If, for example, one was interested in developing a DBM that could provide a diagnosis about the stratification of the progression of a particular disease—let us say, in which of the three stages a patient's particular cancer falls—then one may not, at least directly, use ROC analysis. One may be able to perform three different ROC analyses, one pairing stages 1 and 2, another pairing stages 2 and 3, and another pairing stages 3 and 1. That, however, could be an unrealistic option if, for example, the number of stages is five or greater.

LDA

Pros: LDA has the potential of being the most accurate of all three analyses. Moreover, it may be used not only when the DV is dichotomous but also when it is discrete in general, i.e., when two or more outcomes are possible. In the aforementioned example, if a particular type of cancer has five stages of progression, then LDA ideally lends itself to the development of a DBM that could diagnose the cancer stage of a patient. More importantly, in this case, one may use LDA to develop a CCA model that could accurately monitor the progression of the cancer from one stage to the next and/or could accurately monitor the course of an applied therapy.

Cons: LDA is undoubtedly the most difficult of the three analyses in terms of the conditions that have to be satisfied pertaining to the IVs. In some cases, it may not even be feasible to use it. Furthermore, one may not use for the development of a DBM or a CCA model any IVs that are not continuous. For example, if a particular investigation necessitates the inclusion of gender as an IV because it is a defining aspect of the disease in question, then one may not employ LDA, as gender is a dichotomous variable (M or F) and not a continuous one.

LRA

Pros: LRA is the third most versatile mathematical analysis. It imposes few conditions on the IVs, and all of them with the exception of one can relatively easily be satisfied. As in the case of LDA, one may use it when the possible outcomes are two or more, i.e., the DV can be discrete. Unlike LDA, however, LRA may be used when the IVs are all continuous, all discrete, or any combination thereof. Compared with ROC curve analysis and LDA, LRA has the following important advantage: it is the only analysis that one may use under the following circumstances. For example, let us say that there are four similar multiple system atrophy (MSA) diseases, that some of those four affect only males, that some affect only

Caucasians, and that some are genetic. Clearly, in this case, one would have to include, among others, gender, race, and family history as IVs. Because the diagnosis can have four different outcomes, ROC analysis is not very practical. Because gender, race, and family history (all discrete variables) must be included in the analysis, LDA may not be used. LRA, on the other hand, is ideally suited for these types of situations.

Cons: Albeit not imposed by theory, the empirical rule pertaining to the adequacy of the sample size is the biggest disadvantage of LRA. It is not a common occurrence for one to have more than 10 experimental observations (data points) for every IV that one would like to use in a given investigation.

Acknowledgments

We thank Ivan Tkac and Professor Janet M. Dubinsky for providing us with the spectral data of 20 unknown mice and also Professor Janet M. Dubinsky for her invaluable editorial help.

Grant sponsor: National Institutes of Health; Grant numbers: T32 DA007097; R03 NS060059.

Literature Cited

- Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol.* 2001; 54:979–985. [PubMed: 11576808]
- Behrens PF, Franz P, Woodman B, Lindenberg KS, Landwehrmeyer GB. Impaired glutamate transport and glutamate-glutamine cycling: downstream effects of the Huntington mutation. *Brain.* 2002; 125:1908–1922. [PubMed: 12135980]
- Bosca MD, Lewis TB, Destache CJ, Benner EJ, Nelson JA, Berti M, Osley RL, Endelman HE. Quantitative ^1H magnetic resonance spectroscopic imaging determines therapeutic immunization efficacy in an animal model of Parkinson's disease. *J Neurosci.* 2005; 25:1691–1700. [PubMed: 15716405]
- Brettschneider J, Petzold A, Sussmuth SD, Landwehrmeyer GB, Ludolph AC, Kassubek J, Tunami H. Neurofilament heavy-chain NfH^{SM135} in cerebrospinal fluid supports the differential diagnosis of parkinsonian syndromes. *Move Disord.* 2006; 21:2224–2227.
- Browne SE, Beal MF. The energetics of Huntington's disease. *Neurochem Res.* 2004; 29:531–546. [PubMed: 15038601]
- Chan HP, Wei J, Zhang Y, Helvie MA, Moore RH, Sahiner B, Hadjiiski L, Kopans DB. Computer-aided detection of masses in digital tomosynthesis mammography: comparison of three approaches. *Med Phys.* 2008; 35:4087–4095. [PubMed: 18841861]
- Chaparro MAE, Marinelli C, Sinito AM. Multivariate techniques as alternative statistical tools applied to magnetic proxies for pollution: a case study from Argentina and Antarctica. *Environ Geol.* 2008; 54:365–371.
- Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariate models. *Ann Intern Med.* 1993; 118:201–210. [PubMed: 8417638]
- Dai W, Theodoridis JM, Graham J, Zeller C, Huang THM, Yan P, Vass JK, Brown R, Paul J. Methylation Linear Discriminant Analysis (MLDA) for identifying differentially methylated CpG islands. *BMC Bioinformatics.* 2008; 9(337)
- Deeks JJ. Diagnostic tests 4: likelihood ratios. *BMJ.* 2004; 329:168–169. [PubMed: 15258077]
- Ellenberger C, Schweizer A, Diaper J, Kalangos A, Murith N, Katchatourian G, Panos A, Licker M. Incidence, risk factors and prognosis of changes in serum creatinine early after aortic abdominal surgery. *Intensive Care Med.* 2006; 32:1808–1816. [PubMed: 16896848]
- Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. *Can J Emerg Med.* 2006; 8:19–20.
- Fan X, Wang L. Comparing linear discriminant function with logistic regression for the two-group classification problem. *J Exp Educ.* 1999; 67:265–286.

- Freedman DS, Otvos JD, Jeyarajah EJ, Barboriak JJ, Anderson AJ, Walker JA. Relation of lipoprotein subclasses as measured by proton nuclear magnetic resonance spectroscopy to coronary artery disease. *Arterioscler Thromb Vasc Biol.* 1998; 18:1046–1053. [PubMed: 9672064]
- Freund, JE. Modern elementary statistics. 6th. Engle-wood Cliffs, NJ: Prentice Hall; 1984.
- Graham PGD, Blamire A, Rothman DL, Brass LM, Fayad PB, Petroff OA, Prichard JW. Early temporal variation of cerebral metabolites after human stroke. A proton magnetic resonance spectroscopy study. *Stroke.* 1993; 24:1891–1896. [PubMed: 8248973]
- Griffin JL. NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterization of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol.* 2003; 7:648–654. [PubMed: 14580571]
- Hintze, JL. NCSS 2004 Manual. Kaysville, UT: NCSS; 2005.
- Hong S, Ahn H, Ha SJ, Yeom HY, Seong GJ, Hong YJ. Early glaucoma detection using the Humphrey matrix perimeter, GDx VCC, stratus OCT, and retinal nerve fiber layer photography. *Ophthalmology.* 2007; 114:210–215.
- Janzen FJ, Stern HS. Logistic regression for empirical studies of multivariate selection. *Evolution.* 1998; 52:1564–1571.
- Jenkins BG, Andreassen OA, Dedeoglu A, Leavitt B, Hayden M, Borchelt D, Ross CA, Ferrante RJ, Beal MF. Effects of CAG repeat length, HTT protein length and protein context on cerebral metabolism measured using magnetic resonance spectroscopy in transgenic mouse models of Huntington's disease. *J Neurochem.* 2005; 95:553–562. [PubMed: 16135087]
- Jones RS, Waldman AD. ¹H-NMR spectroscopy evaluation of metabolism in Alzheimer's disease and vascular dementia. *Neurol Res.* 2004; 26:488–495. [PubMed: 15265265]
- Lamia B, Hellot MF, Girault C, Tamion F, Dachraoui F, Lenain P, Bonmarchand G. Changes in severity and organ failure scores as prognostic factors in onco-hematological malignancy patients admitted to the ICU. *Intensive Care Med.* 2006; 32:1560–1568. [PubMed: 16896863]
- Lei PW, Koehly LM. Linear discriminant analysis versus logistic regression: a comparison of classification errors in the two-group case. *J Exp Educ.* 2003; 72:25–49.
- Leng C. Sparse optimal scoring for multiclass cancer diagnosis and biomarker detection using microarray data. *Comput Biol Chem.* 2008; 32:417–425. [PubMed: 18722815]
- Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington Lawton M, Trotter Y, Leach H, Davies SW, Bates GP. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell.* 1996; 87:493–506. [PubMed: 8898202]
- McClish DK. Analyzing a portion of the ROC curve. *Med Decis Making.* 1989; 9:190–195. [PubMed: 2668680]
- Menard S. Coefficients of determination for multiple logistic regression analysis. *Am Stat.* 2000; 54:17–24.
- Metz CE. Basic principles of ROC analysis. *Semin Nuclear Med.* 1978; 8:283–298.
- Morgan P, Teachman JD. Logistic regression: description, examples, and comparisons. *J Marriage Fam.* 1988; 50:929–936.
- Mu T, Nandi AK, Rangayyan RM. Classification of breast masses using selected shape, edge-sharpness, and texture features with linear and kernel-based classifiers. *J Digital Imaging.* 2008; 21:153–169.
- Nakae I, Mitsunami K, Omura T, Yabe T, Tsutamoto T, Matsuo S, Takahashi M, Morikawa S, Inubushi T, Nakamura Y, Kinoshita M, Horie M. Proton magnetic resonance spectroscopy can detect creatine depletion associated with the progression of heart failure in cardiomyopathy. *J Am Coll Cardiol.* 2003; 42:1587–1593. [PubMed: 14607443]
- Nikas JB, Low WC. Clustering analyses for the diagnosis of Huntington and other diseases. Abstracts for the 17th Annual Meeting of the American Society for Neural Therapy and Repair. *Cell Transplantation.* 2010; 19:355.
- Odunsi K, Wollman RM, Ambrosone CB, Hutson A, McCann SE, Tammela J, Geisler JP, Miller G, Sellers T, Cliby W, Qian F, Keitz B, Intengan M, Lele S, Alderfer JL. Detection of epithelial ovarian cancer using ¹H-NMR-based metabonomics. *Int J Cancer.* 2005; 113:782–788. [PubMed: 15499633]

- Opstad KS, Bell BA, Griffiths JR, Howe FA. An investigation of human brain tumour lipids by high-resolution magic angle spinning ^1H NMR spectroscopy and histological analysis. *NMR Biomed.* 2008; 21:677–685. [PubMed: 18186027]
- Pears MR, Cooper JD, Mitchison HM, Mortishire-Smith RJ, Pearce DA, Griffin JL. High resolution ^1H NMR-based metabolomics indicates a neurotransmitter cycling deficit in cerebral tissue from a mouse model of Batten disease. *J Biol Chem.* 2005; 280:42508–42514. [PubMed: 16239221]
- Peduzzi P, Concato J, Kemper E, Holford TR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol.* 1996; 49:1373–1379. [PubMed: 8970487]
- Portney, LG.; Watkins, MP. *Foundations of clinical research: applications to practice.* 2nd. Upper Saddle River, NJ: Prentice Hall; 2000.
- Ramankulov A, Lein M, Kristiansen G, Loening SA, Jung K. Plasma osteopontin in comparison with bone markers as indicator of bone metastasis and survival outcome in patients with prostate cancer. *Prostate.* 2007; 67:330–e40. [PubMed: 17192877]
- Rieber J, Huber A, Erhard I, Mueller S, Schweyer M, Koenig A, Schiele TM, Theisen K, Siebert U, Schoenberg SO, Reiser M, Klauss V. Cardiac magnetic resonance perfusion imaging for the functional assessment of coronary artery disease: a comparison with coronary angiography and fractional flow reserve. *European Heart Journal.* 2006; 27:1465–71. [PubMed: 16720685]
- Schultz-Larsen K, Lomholt RK, Kreiner S. Mini-Mental Status Examination: a short form of MMSE was as accurate as the original MMSE in predicting dementia. *J Clin Epidemiol.* 2007; 60:260–267. [PubMed: 17292020]
- Tabachnick, B.; Fidell, L. *Using multivariate statistics.* 5th. Boston, MA: Pearson; 2007.
- Tkac I, Henry PG, Andersen P, Keene CD, Low WC, Gruetter R. Highly resolved in vivo ^1H NMR spectroscopy of the mouse brain at 9.4 T. *Magn Reson Med.* 2004; 52:478–484. [PubMed: 15334565]
- Tkac I, Dubinsky JM, Keene CD, Gruetter R, Low WC. Neurochemical changes in Huntington R6/2 mouse striatum detected by in vivo ^1H NMR spectroscopy. *J Neurochem.* 2007; 100:1397–1406. [PubMed: 17217418]
- Vanderstichele H, De Vreese K, Blennow K, Andreasen N, Sindic C, Ivanolu A, Hampel H, Burger K, Parnetti L, Lanari A, Padovani A, DiLuca M, Blaser M, Olsson AO, Pottel H, Hulstaert F, Vanmechelen E. Analytical performance and clinical utility of the INNOTEST PHOSPHO-TAU_(181P) assay for discrimination between Alzheimer's disease and dementia with Lewy bodies. *Clin Chem Lab Med.* 2006; 44:1472–1480. [PubMed: 17163825]
- Viant MR, Rosenblum ES, Tjeerdema RS. A powerful approach for characterizing the effects of environmental stressors on organism health. *Environ Sci Technol.* 2003; 37:4982–4989. [PubMed: 14620827]
- Wang Z, Palacios MA. Fluorescence sensor array for metal ion detection based on various coordination chemistries: general performance and potential application. *Anal Chem.* 2008; 80:7451–7459. [PubMed: 18778035]
- Weckwerth, W. *Metabolomics: methods and protocols.* Totowa, NJ: Humana Press; 2007.
- Yan X, Zheng T. Selecting informative genes for discriminant analysis using multigene expression profiles. *BMC Genomics.* 2008; 9(suppl 2):S14. [PubMed: 18831779]

Figure 1.

Representative in vivo ^1H NMR spectra measured from striata of HD transgenic R6/2 and WT mice at 8 and 12 weeks of age. **A:** Eight-week-old R6/2. **B:** Eight-week-old WT. **C:** Twelve-week-old R6/2. **D:** Twelve-week-old WT. Insets: NMR images with the selected volume of interest (VOI) centered rostrocaudally at the level of the anterior commissure. Arrows indicate changes of a specific metabolite in R6/2 mice relative to age-matched WT control. STEAM, TE = 2 ms, TR = 5 seconds, NT = 160. Processing: Gaussian multiplication ($\sigma = 0.085$ seconds), FT and zero-order phase correction. No water signal removal or baseline correction was applied. For abbreviations, see Table 1. (Adapted from Tkac et al., 2007, with permission by the Journal of Neurochemistry.)

Figure 2.

A–I: ROC curves and their respective AUC values for the top nine most significant metabolites: Cr+PCr, Gln, Cr, NAA, GSH, GPC+PC, mIns, PCr, and PE. For abbreviations, see Table 1.

$$\text{METABOLITE \% DEFICIT} = [(C-R)/(C_{\text{MAX}}-R)] * 100$$

Figure 3.

Calculation of an individual metabolite % deficit according to ROC curve analysis. C is the ^1H NMR spectroscopy concentration value of a metabolite of an R6/2 mouse. R is the ROC cutoff concentration point, i.e., the threshold concentration value that defines normal (for WT, $C < R$) and abnormal (for R6/2, $C > R$). C_{MAX} is the largest theoretical concentration value of a particular metabolite that an R6/2 mouse can have, defined as $R+3$ SD. SD is the standard deviation of all R6/2 concentration values for the particular metabolite.

Figure 4.

Calculation of the overall metabolite % deficit according to linear discriminant analysis (LDA). According to the discriminant function, the score $D = 0$ is the cutoff score, i.e., WT mice have negative scores ($D < 0$), whereas R6/2 mice have positive scores ($D > 0$). D is the canonical discriminant score of an R6/2 mouse, and D_{MAX} is the theoretically maximum R6/2 canonical discriminant score.

$$\text{OVERALL METABOLITE \% DEFICIT} = (Z/Z_{\text{MAX}}) * 100$$

Figure 5.

Calculation of the overall metabolite % deficit according to logistic regression analysis (LRA). Z is the natural logarithm of the odds of belonging to the target group (R6/2), also called logit. According to LRA, the logit $Z = 0$ is the cutoff value, so that WT mice have negative logits ($Z < 0$) and R6/2 mice have positive logits ($Z > 0$). Z is the logit of an R6/2 mouse, and Z_{MAX} is the theoretically maximum R6/2 logit value.

Figure 6.

The canonical discriminant scores of all 30 mice (17 WT and 13 R6/2) are plotted against themselves (D vs. D) in order to provide a better visualization of the separation of the two groups achieved by the LDA. All WT mice have negative canonical discriminant scores (mean score = -3.81 ± 0.20), whereas all R6/2 mice have positive canonical discriminant scores (mean score = 4.98 ± 0.33 ; $P < 1 \times 10^{-6}$).

Figure 7.

The two red bars on the left represent the mean canonical discriminant score of 7 R6/2 mice at 8 weeks and that of 6 R6/2 mice at 12 weeks, respectively. The 12-week-old R6/2 mice have significantly higher D-scores (are in a worse disease state) than the 8-week-old R6/2 mice. The two green bars on the right represent the mean canonical discriminant score of 8 WT mice at 8 weeks and that of 9 WT mice at 12 weeks, respectively. There is no statistically significant difference between the two WT groups. Furthermore, all possible combinations of comparisons between the two R6/2 groups of mice and the two WT groups of mice are statistically significant ($P < 1 \times 10^{-6}$).

Figure 8.

The logit values of 17 WT and 13 R6/2 mice are plotted against themselves (Z vs. Z) so as to provide a better visualization of the separation of the two groups achieved by LRA. All WT mice have negative logit values (mean score = -1.91 ± 0.10), whereas all R6/2 mice have positive logit values (mean score = 1.87 ± 0.15 ; $P < 1 \times 10^{-6}$).

Figure 9.

The mean logit value of 7 R6/2 mice at 8 weeks and that of 6 R6/2 mice at 12 weeks are represented respectively by the two red bars on the left. The 12-week-old R6/2 mice have significantly higher logit values (are in a worse disease state) than the 8-week-old R6/2 mice. The mean logit values of 8 WT mice at 8 weeks and that of 9 WT mice at 12 weeks are represented respectively by the two green bars on the right. There is no statistically significant difference between the two WT groups. Furthermore, all possible combinations of comparisons between the two R6/2 groups of mice and the two WT groups of mice are statistically significant ($P < 1 \times 10^{-6}$).

Figure 10.

The mean value of the overall metabolite % deficit of the 7 R6/2 mice at 8 weeks (red bars) and that of the 6 R6/2 mice at 12 weeks (blue bars) as calculated by the ROC, LDA, and LRA CCA models. The 12-week-old R6/2 mice have significantly higher % deficits than the 8-week-old R6/2 mice. For abbreviations, see Table 1.

Table 1
Names and Abbreviations of All Metabolites Detected, Measured, and Evaluated in the Study

No.	Metabolite symbol	Metabolite name
1	Cr	Creatine
2	PCr	Phosphocreatine
3	Cr+PCr	Creatine + phosphocreatine
4	PCr/Cr	Phosphocreatine / creatine
5	GABA	γ -Aminobutyric acid
6	Glc	Glucose
7	Gln	Glutamine
8	Glu	Glutamate
9	GSH	Glutathione
10	GPC+PC	Glycerophosphorylcholine + phosphorylcholine
11	Lac	Lactate
12	MM	Macromolecules
13	mIns	<i>myo</i> -Inositol
14	NAA	N-acetylaspartate
15	NAAG	N-acetylaspartylglutamate
16	PE	Phosphorylethanolamine
17	Tau	Taurine

Table 2
Rank of All Metabolites Based on Their Discriminating Power (AUC) From ROC Curve Analysis¹

ROC curve analysis over 8–12 weeks					
Metabolite	Best ROC cutoff concentration (μmol/g)	Sensitivity (Sens)	1-Specificity (1-Spec)	AUC	AUC rank
Cr+PCr	>9.344	0.99999	0.05882	1.00000	1
Gln	>4.334	0.95166	0.05882	0.98897	2
Cr	6.176	0.94199	0	0.98832	3
NAA	6.921	0.90770	0	0.98198	4
GSH	>1.313	0.86805	0.11765	0.94052	5
GPC+PC	1.847	0.83431	0.05882	0.90301	6
mIns	6.238	0.82754	0.11765	0.89978	7
PCr	3.792	0.78208	0.11765	0.87023	8
PE	<2.991	0.75247	0.23529	0.83667	9
Tau	13.251	0.64550	0.05882	0.72888	10
NAAG	1.078	0.75250	0.29412	0.69632	11
Glc	3.334	0.44391	0.23529	0.58495	12
Glu	9.020	0.60048	0.47059	0.58179	13
PCr/Cr	<0.607	0.45995	0.35294	0.53852	14
GABA	2.561	0.65132	0.52941	0.52209	15
Lac	5.858	0.37087	0.29412	0.52187	16
MM	<2.190	0.38111	0.23529	0.50067	17

¹ For abbreviations, see Table 1. Each metabolite's best receiver operating characteristic (ROC) cutoff concentration is shown, along with its respective sensitivity and (1-specificity) values.

Table 3
R6/2 Concentration Values According to the Eight Most Significant Metabolites by ROC Curve Analysis¹

ROC CCA	R6/2 ROC cutoff concentration ($\mu\text{mol/g}$) (R)	R6/2 Mean concentration ($\mu\text{mol/g}$) $\pm$ 3 SD (C_{max} or C_{min} *)
Cr+PCr	>9.344	13.088 (+3SD)
Gln	>4.334	8.941 (+3SD)
Cr	6.176	9.011 (+3SD)
NAA	6.921	4.642 (−3SD)*
GSH	> 1.313	2.183 (+3SD)
GPC+PC	1.847	4.799 (+3SD)
mIns	6.238	8.607 (+3SD)
PCr	3.792	6.628 (+3SD)

¹The top eight most significant metabolites ($\text{AUC} > 0.85$) and the respective R6/2 cutoff concentration values (R) and theoretical maximum or minimum concentration values (C_{max} or C_{min}).CCA, clinical change assessment. For other abbreviations, see Table 1.

Table 4

Linear Discriminant Analysis Report: Variable Influence¹

Variable	Removed λ	Removed F-value	Removed F-prob.	Alone λ	Alone F-value	Alone F-prob.	R-squared other X's
Cr	0.907904	2.03	0.169763	0.258382	80.37	0.000000	0.827534
T(Gln)	0.994164	0.12	0.735434	0.215629	101.85	0.000000	0.863745
NAA	0.932949	1.44	0.244574	0.295003	66.91	0.000000	0.724283
T(Cr+PCr)	0.797921	5.07	0.035821	0.071075	365.95	0.000000	0.945075
GSH	0.963671	0.75	0.395519	0.421218	38.47	0.000001	0.610893
mIns	0.994174	0.12	0.735656	0.532885	24.54	0.000031	0.579251
T(PE)	0.993737	0.13	0.726274	0.665245	14.09	0.000810	0.624892
T(Tau)	0.962931	0.77	0.390662	0.467347	31.91	0.000005	0.628281
T(GPC+PC)	0.946174	1.14	0.298832	0.382597	45.18	0.000000	0.809156

¹The final nine independent variables (IVs) used for construction of the discriminant function in linear discriminant analysis (LDA) and their significance to the function. An IV's Wilk's λ (Alone Lambda) is indirectly proportional to the significance of that IV to the discriminant function. Five of those IVs have been transformed; their names in parentheses are preceded by the letter "T". For abbreviations, see Table 1.

Table 5

Linear Discriminant Analysis Report: Canonical Variate Analysis¹

Fn	Inv(W)B Eigenvalue	Ind'l. %t	Total %	Canon. corr	Canon cort ²	F-value	Numer. DF	Denom. DF	Prob. level	Wilks' λ
1	20.303202	100.0	100.0	0.9762	0.9531	45.1	9.0	20.0	0.0000	0.046941

¹ Statistics pertaining to the discriminant function of linear discriminant analysis (LDA), Wilks' λ and the canonical correlation

² (Canon. corr.) show that >95% of the variance in the canonical discriminant scores can be accounted for by the nine IVs we selected for the construction of the discriminant function.

Table 6
Linear Discriminant Analysis (LDA) Report: Estimated LDA Function¹

Variable	Canonical coefficient
Constant	−39.0075
Cr	0.7858
T(Gln)	1.5433
NAA	−0.5980
T(Cr+PCr)	37.5859
GSH	1.1174
mIns	0.1562
T(PE)	−0.0409
T(Tau)	1.0243
T(GPC+PC)	3.8285

¹The constant (y-intercept) and the weights (coefficients) of the nine independent variables (IVs) used for the construction of the discriminant function. For abbreviations, see Table 1

Table 7
Logistic Regression Analysis (LRA) Report: Estimated LRA Function^I

Variable	Coefficient
Constant	-7.7656
Cr+PCr	1.0582
Gln	4.3641E-02
Cr	5.3906E-02
NAA	-0.4431
GSH	0.6089
GPC+PC	-0.31036
mIns	5.1964E-02
PE	-0.3581

^IThe y-intercept and the weights (coefficients) of the independent variables (IVs) used to construct the logistic function. For abbreviations, see Table 1.

Table 8
Summary of Test Results of ROC Curve Analysis, Linear Discriminant Analysis (LDA),
and Linear Regression Analysis (LRA)¹

	% Correct		
	ROC	LDA	LRA
Test 1: ID of original 30 mice			
17 WT	17/17 → 100%	17/17 → 100%	17/17 → 100%
13 R6/2	13/13 → 100%	13/13 → 100%	13/13 → 100%
Total	30/30 → 100%	30/30 → 100%	30/30 → 100%
Sensitivity	1.0	1.0	1.0
1-Specificity	0.0	0.0	0.0
(+) Likelihood ratio	1/0 → ∞	1/0 → ∞	1/0 → ∞
(-) Likelihood ratio	0/1 = 0	0/1 = 0	0/1 = 0
Test 2: ID of 31 unknown mice			
20 WT	20/20 → 100%	20/20 → 100%	20/20 → 100%
11 R6/2	11/11 → 100%	11/11 → 100%	11/11 → 100%
Total	31/31 → 100%	31/31 → 100%	31/31 → 100%
Sensitivity	1.0	1.0	1.0
1-Specificity	0.0	0.0	0.0
(+) Likelihood ratio	1/0 → ∞	1/0 → 1	1/0 → ∞
(-) Likelihood ratio	0/1 = 0	0/1 = 0	0/1 = 0
Test 3: ID of original 13 R6/2 mice			
7 R6/2 8 weeks old	7/7 → 100%	7/7 → 100%	7/7 → 100%
6 R6/2 12 weeks old	6/6 → 100%	6/6 → 100%	6/6 → 100%
Total	13/13 → 100%	13/13 → 100%	13/13 → 100%
Sensitivity	1.0	1.0	1.0
1-Specificity	0.0	0.0	0.0
(+) Likelihood ratio	1/0 → ∞	1/0 → ∞	1/0 → ∞
(-) Likelihood ratio	0/1 = 0	0/1 = 0	0/1 = 0

¹Test 1 pertains to the identification (ID) of the 30 original mice (intrinsic to all models). Test 2 pertains to the identification of the 31 unknown mice, which were extraneous to all models, and the status of which had been determined via genotyping. Test 3 pertains to the identification of the 13 original R6/2 mice with respect to their two age groups (8 and 12 weeks old).