Huntington disease patients and transgenic mice have similar pro-catabolic serum metabolite profiles

Benjamin R. Underwood,[1,5,*] David Broadhurst,[6,*] Warwick B. Dunn,[6,*] David I. Ellis,[6] Andrew W. Michell,[3] Coralie Vacher,[1] David E. Mosedale,[4] Douglas B. Kell,[6,†] Roger A. Barker,[3,†] David J. Grainger[2,4,†] and David C. Rubinsztein[1,†]

[1]Department of Medical Genetics, Cambridge Institute for Medical Research, [2]Department of Medicine, University of Cambridge, Addenbrooke's Hospital, [3]Cambridge Centre for Brain Repair, Forvie Site, Robinson Way, [4]Translational Research Unit, Papworth Hospital NHS Trust, Papworth Everard, Cambridge, [5]Suffolk Mental Health Partnership NHS Trust, Department of Psychiatry, Wedgwood House, West Suffolk Hospital, Bury St Edmunds and [6]School of Chemistry, University of Manchester, Manchester, UK

Correspondence to: Prof. David C. Rubinsztein, MB, ChB, PhD, Department of Medical Genetics, Cambridge Institute for Medical Research, Wellcome/MRC Building, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 2XY, UK
E-mail: dcr1000@hermes.cam.ac.uk
*These authors contributed equally to the study.
†The study has four equal senior authors.

There has been considerable progress recently towards developing therapeutic strategies for Huntington's disease (HD), with several compounds showing beneficial effects in transgenic mouse models. However, human trials in HD are difficult, costly and time-consuming due to the slow disease course, insidious onset and patient-to-patient variability. Identification of molecular biomarkers associated with disease progression will aid the development of effective therapies by allowing further validation of animal models and by providing hopefully more sensitive measures of disease progression. Here, we apply metabolic profiling by gas chromatography-time-of-flight-mass spectrometry to serum samples from human HD patients and a transgenic mouse model in a hypothesis-generating search for disease biomarkers. We observed clear differences in metabolic profiles between transgenic mice and wild-type littermates, with a trend for similar differences in human patients and control subjects. Thus, the metabolites responsible for distinguishing transgenic mice also comprised a metabolic signature tentatively associated with the human disease. The candidate biomarkers composing this HD-associated metabolic signature in mouse and humans are indicative of a change to a pro-catabolic phenotype in early HD preceding symptom onset, with changes in various markers of fatty acid breakdown (including glycerol and malonate) and also in certain aliphatic amino acids. Our data raise the prospect of a robust molecular definition of progression of HD prior to symptom onset, and if validated in a genuinely prospective fashion these biomarker trajectories could facilitate the development of useful therapies for this disease.

Introduction

Huntington's disease (HD) is a devastating autosomal dominant neurodegenerative condition that manifests with movement disorder, behavioural disturbance, and cognitive deterioration. Although it can present at any age, the median age of onset is 40 years, and death typically follows some 15–20 years after symptom onset. The HD mutation is a

$(CAG)_n$ trinucleotide repeat expansion at the $5'$ end of the transcript encoding huntingtin. The $(CAG)_n$ repeats are translated into a polyglutamine tract. Disease is caused by >35 CAG repeats and the age at onset correlates inversely with CAG repeat number.

In the 12 years since the Huntington's disease mutation was identified, considerable progress has been made with modelling pathogenesis in cell and animal models. Mutant huntingtin accumulates in intraneuronal aggregates (also called inclusions). Huntingtin is cleaved to form N-terminal fragments consisting of the first 100–150 residues containing the expanded polyglutamine tract and these are believed to be the toxic species found in the aggregates (DiFiglia *et al.*, 1997; Taylor *et al.*, 2002). HD pathogenesis is frequently modelled with exon 1 fragments containing expanded polyglutamine repeats, which lead to aggregate formation and toxicity in cell models and *in vivo* (Cooper *et al.*, 1998; Martindale *et al.*, 1998; Wyttenbach *et al.*, 2000). The extent to which mouse models expressing N-terminal fragments of mutant huntingtin recapitulate the human disease phenotype is unclear, although these mice have pathological and behavioural similarites to HD.

These models have been particularly powerful tools for developing therapeutic strategies and a number of compounds have emerged from studies in mice that delay disease onset and/or alleviate progression of the disease. The former strategy is appealing, since one could effectively cure many cases if one could force the age at onset of symptoms to beyond normal life expectancy. This is theoretically feasible in HD, as mutation status can be assessed presymptomatically in most cases with technical ease. However, the ability to test promising approaches based on mouse studies in patients is handicapped by a number of factors. Human trials in HD are difficult, costly and time-consuming due to the slow disease course, insidious onset and patient-to-patient variability. Any such study would require many hundreds of people to be treated for at least 3 years in order to have suitable power to detect clinically realistic outcomes.

Identification of molecular biomarkers associated with the progression of this disease would aid the development of effective therapies in at least two ways. First, they would allow validation of the animal models which have been used to select candidate therapies. If the animal models share a wide range of biomarkers with human patients, this increases the likelihood that treatments effective in the animals will also ultimately be effective in humans. Secondly, following changes in biomarkers, particularly in the presymptomatic or early symptomatic phases of the disease should simplify and shorten early clinical trials examining the efficacy of candidate therapies. Any such treatment which successfully rescued biomarkers in animal models in the early symptomatic phase and delayed disease progression could be viewed optimistically if it had a similar effect on biomarkers in early symptomatic patients. Indeed, a marker that predicted the transition to the disease state in those carrying the expanded *huntingtin* allele would be extremely

valuable as it is this group of patients that stand the most chance of benefiting from disease modifying therapies.

Unfortunately, conventional approaches to biomarker discovery have not yet yielded biomarkers that can be used for these purposes. Recently, however, it has become clear that non-hypothesis driven systems biology approaches can be used effectively in such cases (Kell, 2004; Kell and Oliver, 2004). In this experimental paradigm, the levels of many thousands of molecules (be it genes, proteins or metabolites) are measured simultaneously and pattern recognition, machine learning or statistical methods are used to select the few that are robustly linked to the development and progression of the disease under study. The study of the complete collection of metabolites in an organism is termed metabolomics (Fiehn, 2002; Goodacre *et al.*, 2004; Kell, 2004). To date, however, it has not been not possible to accurately determine the levels of all metabolites simultaneously, because of the substantial chemical and physical heterogeneity of the metabolome. As a result, subsets of metabolites are studied (a process often termed metabolic profiling). A number of analytical strategies have been successfully employed (Dunn and Ellis, 2005; Dunn *et al.*, 2005) including GC-MS, LC-MS, CE-MS, NMR and FT-IR, although each has its own limitations and advantages.

Metabolic profiling has already proved useful in a number of disease areas: markers associated with the presence and severity of coronary artery disease (Brindle *et al.*, 2002), hypertension (Brindle *et al.*, 2003), subarachnoid haemorrhage (Dunne *et al.*, 2005), pre-eclampsia (Kenny *et al.*, 2005), Type 2 diabetes (Wang *et al.*, 2005), liver cancer (Yang *et al.*, 2004), motor neuron disease (Rozen *et al.*, 2005) and ovarian cancer (Odunsi *et al.*, 2005) have all been identified by different implementations of this general approach.

The aim of the present study was to apply metabolic profiling by gas chromatography-time-of-flight-mass spectrometry (GC-TOF-MS) to serum samples from human HD patients as well as a mouse model of the disease in a non-hypothesis driven search for disease biomarkers. Since HD shares many of the above issues with other late-onset neurodegenerative conditions, progress in this field of biomarkers may provide enthusiasm for parallel approaches in Alzheimer's disease, Parkinson's disease and other chronic disorders of the CNS.

Methods
Humans
This study was performed with appropriate Local Ethical Committee approval and with informed consent from the participants. We have studied 30 HD mutation-positive (>36 uninterrupted CAG repeats) patients (10 in the presymptomatic and 20 in the early symptomatic phase) and 20 control subjects recruited from the HD clinic at Addenbrooke's Hospital. Prior to recruitment, we established the following inclusion criteria: asymptomatic gene-positive cases who were over 18 years of age, who had unified Huntington's disease rating scale (UHDRS) motor scores of ≤ 6 (Huntington Study Group, 1996; range 0–6, mean 1.9, SD 1.85)

and who were on no psychoactive medication (antidepressants, antipsychotics anticonvulsants or medication prescribed to relieve symptoms) at their last clinic visit. Symptomatic gene-positive cases had the same inclusion criteria as the asymptomatic group except that they had overt motor features of the disease and a UHDRS motor score of ≥11 (range 11–70, mean 29.9, SD 15.94). We excluded patients in the final terminal phase of their illness, patients so impaired as to be unable to give informed consent, and any patient suffering from current acute medical complaint. The control group were partners/carers of the patients (aged ≥18 years) and had no known neurological disorder and at the time of initially being screened were on no psychoactive medications.

A detailed history was taken for each patient, with particular emphasis on variables known to affect metabolite profiles. Age, ethnicity, occupation, family history (including detailed family history of HD), smoking status, alcohol consumption, exercise, sleep, tea and coffee consumption, prescription and over-the-counter medication, dietary supplementation, height, weight, body mass index, blood pressure, past medical history, menstrual status, symptom history, independence score and UHDRS motor scores were recorded. The ages of the groups were as follows: asymptomatic group = mean 47 (SD 11), range 30–66; symptomatic group = mean 54 (SD 10), range 42–77; controls = mean 52 (SD 11), range 31–76; (no significant differences between groups). The male : female ratios of the groups were 3 : 7 (asymptomatic); 9 : 11 (symptomatic) and controls 13 : 7. The body mass index (BMI) data of the different groups were as follows: asymptomatic group = range 14.7–30.2, mean 24.5, SD 4.7; symptomatic group = range 16.9–27.8, mean 23, SD 2.97; control group = range 22.3 – 34.8, mean 27.99, SD 3.49.

Some of the patients invited to take part in the study had been started on psychoactive medication subsequent to their last clinic visit. Similarly some of the subjects from the control group were on such medication. All subjects were included in the study and further statistical analyses were performed to exclude any possible effect of differences in drug use between the groups under study. There was no single medication or class of medication which caused a systematic difference between the groups. Of the asymptomatic gene-positive group seven were on no medication, one on 'other medication' and two on fluoxetine. Of the symptomatic gene-positive group 11 were medication free, 1 was on 'other medication' (i.e. not psychoactive) and 9 were on psychoactive or possible anti-HD medication (fluoxetine = 1, amitriptyline = 2, minocycline = 1, carbamazepine = 1, amisulpride = 2, ginkgo biloba = 1, co-enzyme Q10 = 1, olanzapine = 1 and venlafaxine = 1). Of the control group nine were medication free, six were on 'other medication' and four on psychoactive medication (amitriptyline = 3 and paroxetine = 1).

Venous blood samples (10 ml) were transferred to a 15 ml plain plastic tube and left at room temperature for 2–3 h to clot. The samples were then centrifuged for 5 min at 3600 *g*. Supernatant serum was removed using a Pasteur pipette and the supernatant was respun for a further 5 min at 3600 *g*. The serum was removed and aliquoted in 800 μl volumes and stored at −80°C.

Mice

HD-N171-N82Q mice (B6C3F1/J-Tg(HD82Gln)81Dbo/J, Jackson Laboratory, Bar Harbour, ME) expressing the first 171 amino acids of human huntingtin under the expression of a mouse PrP promoter were used (Schilling *et al.*, 1999). The time course of the development of symptoms in these mice has been well characterized

(Schilling *et al.*, 1999, 2001). All studies and procedures were performed under the jurisdiction of the appropriate Home Office Project and Personal Animal Licence and with local ethical committee approval. The mice were genotyped between 3 and 5 weeks of age by PCR. Non-transgenic littermates of these mice were used as the controls.

Three groups of mice were sacrificed to provide analogous groups to the human subjects (control, asymptomatic and symptomatic). The asymptomatic transgenic group were sacrificed at 8 weeks and the symptomatic were sacrificed at 15 weeks of age. A total of 9 transgenic symptomatic and 10 transgenic asymptomatic mice were sacrificed. A total of 10 non-transgenic mice were sacrificed, of which 4 were sacrificed at 8 weeks (asymptomatic littermates) and 5 were sacrificed at 15 weeks (symptomatic littermates). The symptomatic mouse group contained eight males and one female, the asymptomatic transgenic group contained seven males and three females. The control group comprised seven males and three females. The mice were anaesthetized by intraperitoneal injection of 500 μl of Pentoject (sodium pentobarbitone, Animal Ltd., York, UK). They were exsanguinated by intracardiac puncture, followed by cervical dislocation. Mouse blood was transferred into a standard 1.5 ml plastic Eppendorf tube and allowed to stand for 2 h to clot. The blood was spun at 3700 *g* for 5 min. The top 250 μl of serum was then removed and transferred to a second Eppendorf tube. The supernatant was respun for a further 5 min at 3700 *g*. The top 220 μl of the supernatant was removed and stored at −80°C.

Metabolite profiling

We selected GC-TOF-MS to construct a metabolome estimate for the serum samples in this study. GC-TOF-MS combines chromatographic separation with sensitive detection and the ability to identify metabolites through mass spectra and retention time comparisons with pure standards (although it is employed for detection of small molecular weight metabolites (<400 Da) generally only after chemical derivatization to increase volatility and thermal stability). Sample preparation for GC-TOF-MS analysis was performed as follows: 175 μl human serum or 50 μl mouse serum was spiked with 20 μl internal standard solution (1.53 mg/ml succinic d_4 acid, 2.34 mg/ml malonic d_2 acid and 1.59 mg/ml glycine d_5; Sigma-Aldrich, Gillingham, UK) and vortex-mixed for 15 s. Aliquots of 450 μl (human samples) or 150 μl (mouse samples) of acetonitrile (AR grade; Sigma-Aldrich) were added followed by vortex mixing (15 s) and centrifugation (13 385 *g*, 15 min) to deproteinize the samples. The supernatant was transferred into an Eppendorf tube and freeze dried (HETO VR MAXI vacuum centrifuge attached to a HETO CT/DW 60E cooling trap; Thermo Life Sciences, Basingstoke, UK). Two-stage sample chemical derivatization was performed on the dried sample. An aliquot of 70 μl (human samples) or 40 μl (mouse samples) of 20 mg/ml O-methylhydroxylamine solution in pyridine was added and heated at 40°C for 90 min followed by addition of 70 μl (human samples) or 40 μl (mouse samples) of 20 mg/ml MSTFA (N-acetyl-N-(trimethylsilyl)-trifluoroacetamide) and heating at 40°C for 90 min. Retention index solution of 20 μl (6 mg/ml *n*-decane, *n*-dodecane, *n*-pentadecane, *n*-nonadecane, *n*-docosane dissolved in hexane) was added and the samples were analysed using an Agilent 6890N gas chromatograph and 7683 autosampler (Agilent Technologies, Stockport, UK) coupled to a LECO Pegasus III electron impact time-of-flight mass spectrometer (LECO Corporation, St Joseph, USA). Optimized instrumental conditions for serum have been described elsewhere (O'Hagan

et al., 2005). Initial processing of raw data was undertaken using the LECO ChromaTof v2.12 software to construct a data matrix (of metabolite peak for each specific sample) using response ratios (peak area metabolite/peak area succinic-d_4 internal standard) for each metabolite peak in each sample.

This approach generates a metabolic profile composed of almost 1300 distinguishable species (uniquely defined by a combination of retention time and mass spectrum). In some cases, two or more distinguished species are derived from the same metabolite during the chemical derivatization process. As a result we estimated that some 1000 distinct metabolites contributed to the metabolic profile. Of these, we estimated that ∼60% can be tentatively assigned a molecular structure (based on database matching with s >700) with the rest unidentified or ambiguous.

Chemometric modelling

Since supervised modelling strategies (which make use of class membership data during model construction) were envisaged, a validation set of 17 samples (34% of the total) were excluded from the model building phase, and their disease and symptom status remained blind to the investigators. Initially unsupervised principal components analysis (PCA) was used to identify outliers, which identified one sample (ND130; an asymptomatic gene-positive subject) in the human dataset as a major outlier in the first three principal components and no outliers in the mouse dataset. This outlier was therefore excluded from all further analyses. Two different supervised modelling approaches were then applied. Principal components discriminant function analysis (PC-DFA) was used to build models, whose predictivity was initially optimized using 5-fold cross-validation with randomly selected internal hold-out sets (each set representative of all classes). The final model was built with all non-blind data, using the optimal latent variable structure. Projection to latent structures discriminant analysis (PLS-DA) was also used, initially optimizing the model by cross-validation based on Q^2. For both PC-DFA and PLS-DA models, the degree of overfitting was estimated by predicting the class membership of the blind validation set, testing these predictions by χ^2. Significant loadings associated with gene status or symptom status were then reported.

To avoid reporting artefacts due to differences in the gender distribution or treatment regimens of the controls versus the cases, separate PLS-DA models were constructed on the entire model set (irrespective of the gene or symptom status) independently predicting gender and drug regimen. None of these models was externally predictive of the hold-out set, nor were any of the metabolites in Tables 1 or 2 dominant loadings, confirming that our findings are associated with gene and/or symptom status, rather than with gender or drug regimen.

Results

Serum samples were prepared from 50 individuals, 30 of whom were mutation-positive for HD (>35 CAG repeats) and 20 were non-HD controls of similar age and sex distribution (see Methods). Importantly, the gene-positive subjects were all in the early stages of the disease [being either asymptomatic ($n = 10$) or early symptomatic individuals ($n = 20$)] ensuring that any metabolic differences that we identified were unlikely to be due to the major behavioural and neurological changes associated with more advanced disease.

Table 1 Metabolites associated with transgenic expression of a mini-HD gene in mice

	Metabolite	Identification
240/242	Glycerol	Product of triglyceride breakdown
280	Unidentified	?
572	Glucose	Carbohydrate metabolism
589	Monosaccharides	Carbohydrate metabolism
122	Lactate	Anaerobic metabolism
296	Urea	Nitrogen excretion
301	Unidentified	?
270/271	Malonate	?
126	Unidentified	?
222	Valine	Proteogenic amino acid
478	Pyroglutamate	Amino acid

Metabolites contributing to the metabolic signature discriminating transgenic mice expressing a 'mini-HD' gene (see Methods) in PC-DFA models are shown. Metabolite identifications are carried out by matching mass spectra to a database (see Methods), and the best match is reported. Weaker matches ($s < 600$) are indicated by a question mark. Some species (such as glycerol and malonate) may have two independent peaks in the profile resulting from the different chemical derivatization products of the sample.

Table 2 Metabolites associated with the presence of HD

	Metabolite	Identification
240	Glycerol	Product of triglyceride breakdown
555/563	Monosaccharides	Carbohydrate metabolism
136/137	Lactate ?	Anaerobic metabolism
126	*Unidentified*	?
152	Alanine	Proteogenic amino acid
228	Leucine	Proteogenic amino acids
224	*Unidentified*	?
278	2-amino-n-butyrate	Intermediate in pyrimidine metabolism
49	Ethylene glycol	Glycerol metabolite ?
1175	Alpha-hydroxybutyric acid	?
222/226	Valine	Proteogenic amino acid
364	*Unidentified*	?
562	Monosaccharide	Carbohydrate metabolism
270	Malonate	?
299	Urea	Nitrogen excretion
197	*Unidentified*	?
64	*Unidentified*	?

Metabolites contributing to the metabolic signature discriminating HD in PC-DFA models (and also in PLS-DA) are shown. Metabolite identifications are carried out by matching mass spectra to a database (see Methods), and the best match is reported. Weaker matches ($s < 600$) are indicated by a question mark. Some species (such as valine) may have two independent peaks in the profile resulting from the different chemical derivatization products of the sample. Note that the monosaccharide peaks 555/563 and 562 represent distinct molecular monosaccharide structures.

Our study design deliberately examined only a relatively small number of subjects for several reasons: most importantly, by restricting the size of the study we were able to perform a more comprehensive phenotypic and biochemical

characterization of each individual. In addition this allowed us to study a relatively homogeneous population of patients, which greatly helped us in the interpretation of the data. This reflects a trade-off between sample size and the amount of data collected on each individual for a given resource commitment. Additionally, the discriminatory power of multivariate statistical modelling increases only slowly with increasing sample size once the size of each class reaches 20–40 individuals (D.J.G., unpublished observations). Therefore a study of this size should identify most biochemical markers likely to be of clinical relevance.

We also carried out an independent analysis of the metabolic profile of a murine model of HD, in which the first 171 residues of huntingtin with 82 CAG repeats are expressed under the control of the mouse prion protein promoter. Serum from 29 mice (19 transgenic and 10 wild-type littermate controls) was prepared as for the human studies.

A metabolic profile for each individual or mouse was then constructed by GC-TOF-MS (see Methods) yielding a matrix consisting of the relative concentration of 1275 uniquely detected metabolite peaks for each of 50 subjects and 29 mice (Fig. 1). For many, but not for all, of the metabolite peaks, the chemical identity of the metabolite can be assigned by comparison of the mass spectrum with a database of more than 80 000 mass spectra (a process similar to identifying genes by sequence homology searches) or more definitively by comparison of mass spectrum and retention time with pure metabolites analysed under the same analytical conditions (Schauer *et al.*, 2005).

Multivariate pattern recognition approaches were then used to search for systematic differences in metabolite profiles between the *HD* gene-positive carriers and the controls. First, unsupervised PCA was used to survey the data for outliers. This method tests whether there is an obvious clustering of subjects by genetic or symptom status, using the entire metabolic profile. In the mouse dataset no outliers were identified, and the samples formed a single homogeneous distribution with no evidence of clustering. In contrast, in the human dataset one individual (an asymptomatic female gene-positive subject, aged 30 years) lay far away from the remaining subjects in both the first two principal components (Fig. 2A) and was therefore removed from all subsequent analyses. The remaining individuals now formed a single homogeneous distribution (Fig. 2B). However, PCA revealed no obvious clustering of the subjects by gene status or symptoms.

Supervised techniques (where the class membership of the subjects is used to build a maximally discriminating mathematical model of the data) are much more powerful at identifying class markers in large multivariate datasets (whether genetic, proteomic or metabolic in origin), providing that adequate model validation is performed to prevent overfitting (Hastie *et al.*, 2001). PC-DFA was therefore applied to build a model on approximately two-thirds of the individuals, which was used to predict the class membership of the remaining one-third (the validation set), whose status was blind to the analyst. In essence, we grouped individuals into their respective genetic and symptom categories and used the metabolite data to try to define a signature for each group. Here we used two-thirds of each group to define a signature, and the reproducibility of this signature was then tested using the remaining one-third of each group. Such a PC-DFA model of the mouse dataset is shown in Fig. 3, discriminating symptomatic (15-week-old) transgenic mice (TS) from both non-transgenic littermates (CS and CP) and pre-symptomatic transgenic mice (8-week-old animals; TP). This model correctly predicted the gene status of the blind validation set (χ^2; $P = 0.03$), and likely provides a robust description of the metabolic consequences of transgene expression and symptom progression.

In this multivariate statistical model, a metabolic signature comprising altered levels of malonate, valine, methionine, glycerol, various monosaccharides as well as an unidentified metabolite (peak no. 301) is indicative of the presence of HD-like symptoms. Consistent with these findings, these mice do develop glycosuria (however, much milder than that seen with the R6/1 HD mouse line).

A PLS-DA model was also constructed on the same dataset (not shown), which yielded similar discrimination between the symptomatic (15-week-old) transgenic animals and the littermate controls. In essence, PLS-DA and PC-DFA are different mathematical approaches used to pick signatures out of the very large metabolic dataset. Valine, glycerol and monosaccharides also possessed significant loadings in this model, confirming the likely importance of these molecules as biomarkers of HD-like symptom development in mice.

The same pattern recognition methods were applied to the human dataset. PC-DFA models tentatively discriminate HD mutation carriers (whether asymptomatic or symptomatic) from the gene-negative controls (Fig. 4A). Unlike the PC-DFA model of the mouse dataset, however, the external predictive power of this model did not reach statistical significance (χ^2; $P = 0.54$), probably reflecting the greater variation in metabolic profiles between individual humans compared with the mice, which share a common genetic background and more controlled lifestyle and diet. The metabolites which together compose the metabolic signature discriminating the gene-positive individuals from the controls are listed in Table 2. Importantly, valine, glycerol and monosaccharides are all significant loadings in both the human PC-DFA model and the mouse PC-DFA model confirming that these model loadings are unlikely to be due to overfitting. Similar loadings dominated the PLS-DA model of the same dataset (not shown).

Interestingly, the metabolites dominantly responsible for the discrimination in these statistical models of both murine and human HD are indicative of a change to a pro-catabolic phenotype early in the disease progression, with markers of increased nucleic acid breakdown (elevated 2-amino-*n*-butyrate, an intermediate in pyrimidine metabolism) and fatty acid β-oxidation (elevated glycerol and ethylene glycol). This same pattern is seen in both asymptomatic and early
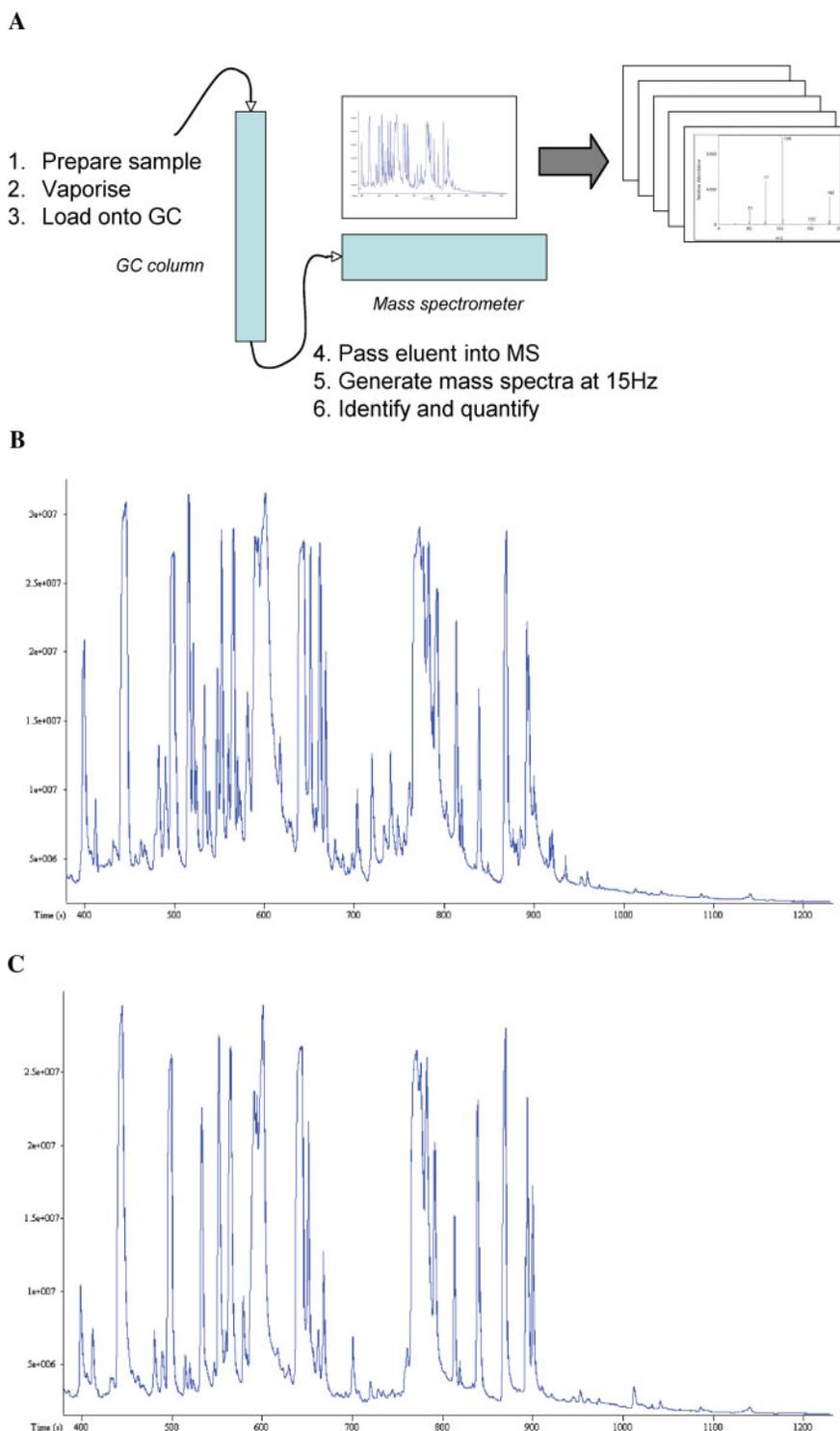
A

1. Prepare sample
2. Vaporise
3. Load onto GC

GC column

Mass spectrometer

4. Pass eluent into MS
5. Generate mass spectra at 15Hz
6. Identify and quantify

B

C

Fig. I Metabolic profiling by GC-TOF-MS. (**A**) A schematic representation of the GC-TOF-MS methodology employed in the present study to obtain a metabolome estimate. (**B**) A typical GC-TOF-MS chromatogram of a human serum sample in this study, plotting total ion current on the mass spectrometer against retention time. After this step, the data are further deconvoluted on the basis of mass to obtain a table of >1300 uniquely identified metabolites. (**C**) A typical GC-MS profile of a mouse serum sample in this study.

symptomatic patients, suggesting, therefore, that a catabolic physiology precedes any detectable clinical symptoms.

These models also provide the first indications of a metabolic signature of symptom progression. While gene-positive

subjects have lower serum valine levels than do gene-negative controls irrespective of symptom status, we see alterations in alanine, leucine/isoleucine and possibly also proline, as well as ethylene glycol and ethylamine, specifically associated
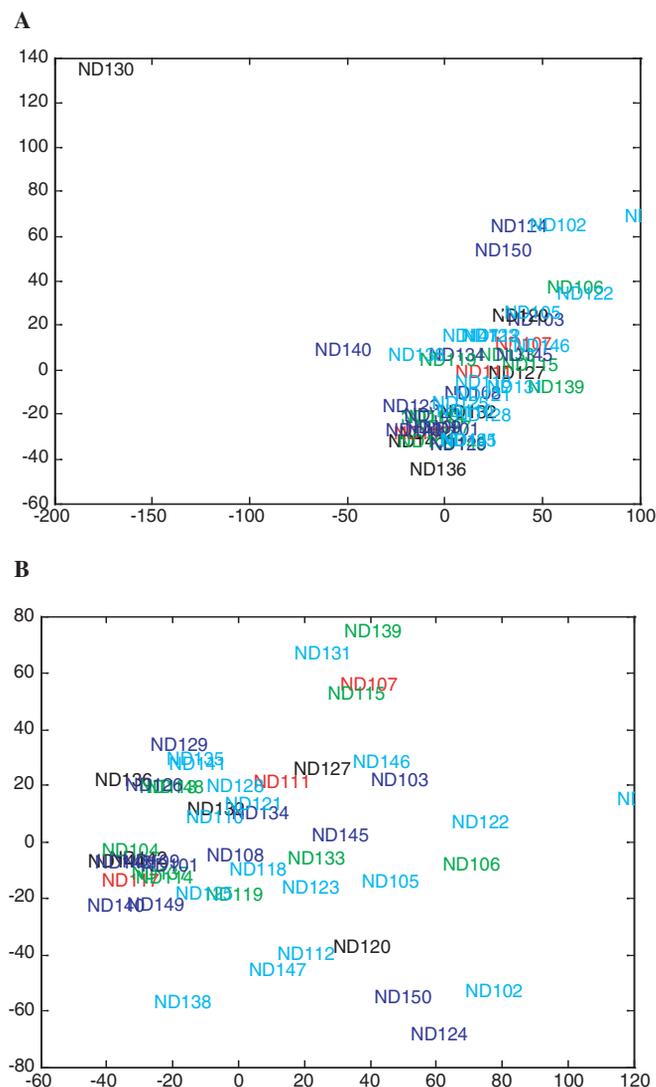
Fig. 2 Principal component analysis to identify outliers. (**A**) PCA of the human dataset ($k = 1275$, $n = 50$; $r^2x = 0.862$, $A = 3$; first two principal components shown). Observations coded according to class membership: red and cyan = controls; black = asymptomatic HD patients; blue = symptomatic HD patients; green = blind validation set. ND130 is clearly identified as a major outlier. (**B**) PCA of the human dataset with ND130 excluded. There were no outliers in the PCA plot of the mouse dataset ($k = 1275$, $n = 29$; $r^2x = 0.919$; $A = 3$; not shown)

with symptom progression. We conclude that there is likely to be a subtle but complex misregulation of amino acid metabolism in HD gene-positive individuals, which correlates with symptom status in our cross-sectional study design, suggesting a progression with time.

Discussion

Our study has revealed three interesting findings. First, we observed metabolic differences between HD patients and controls. Some of these changes (such as the complex dysregulation of amino acid metabolism and the accumulation of

ethylene glycol and its oxidation product oxalate) may be specific for the disease. Other changes may be more general markers of neurodegeneration, since they have also been identified in analogous studies of Parkinson's disease and Alzheimer's disease (D.J.G., R.A.B. *et al.*, unpublished data). For example, malonate and 2-amino-*n*-butyrate (as well as the unidentified metabolites nos 126 and 301) are also associated with Parkinson's disease. The elevation of malonate in both the HD patients and mice is potentially interesting, as intrastriatal infusion of malonate, a reversible inhibitor of the mitochondrial enzyme succinate dehydrogenase, results in striatal pathology that has similarities to that seen in HD (Beal *et al.*, 1993). Succinate dehydrogenase is a component of the tricarboxylic acid cycle and complex II of the mitochondrial electron transport chain, and decreased complex II activity has been reported in HD brains (Browne *et al.*, 1997; Gu *et al.*, 1996). However, before one can speculate that the malonate elevation may be partly contributing to pathology, further work will be required to test whether it is elevated in the CNS, to what extent it is altered and at which stages of disease. Nevertheless, the malonate data and the changes in lactate levels are consistent with the hypothesis that misregulation of energy expenditure (evidenced by altered mitochondrial electron transport activity, Gu *et al.*, 1996, and reduced GAPDH activity, Mazzola *et al.*, 2001) plays a central role in the phenotype resulting from mutant huntingtin expression.

Secondly, we see a change in biomarker profile between asymptomatic gene carriers and patients with early disease, consistent with progression. The transition of phenotypes between asymptomatic and symptomatic cases is associated with differing patterns of amino acid accumulation. Asymptomatic cases have elevated levels of alanine and leucine, while symptomatic disease is associated instead with ethylene glycol accumulation, and possibly also alpha-hydroxybutyric acid. However, all gene-positive individuals (irrespective of the symptom status) have significant reductions in valine. Thus, the complex dysregulation of amino acid metabolism changes between asymptomatic and early symptomatic cases, if validated in further studies, may be a useful independent state marker. Interestingly, as early as 1969 alterations in amino acid metabolism in HD were reported (Perry *et al.*, 1969; Philipson and Bird, 1977; Reilmann *et al.*, 1995), but these have never been established as important state markers until now. While some of these changes may reflect a pro-catabolic phenotype, it is unclear whether this accounts for the overall signature. However, it is encouraging when broad non-hypothesis driven experimental approaches yield a subset of candidate biomarkers which had previously been tentatively identified. On the basis of the current findings, we are encouraged to embark upon a series of prospective studies that will allow us to investigate longitudinal changes in different patients, and investigate the minimum time between observations that allow meaningful changes in metabolites to be observed.
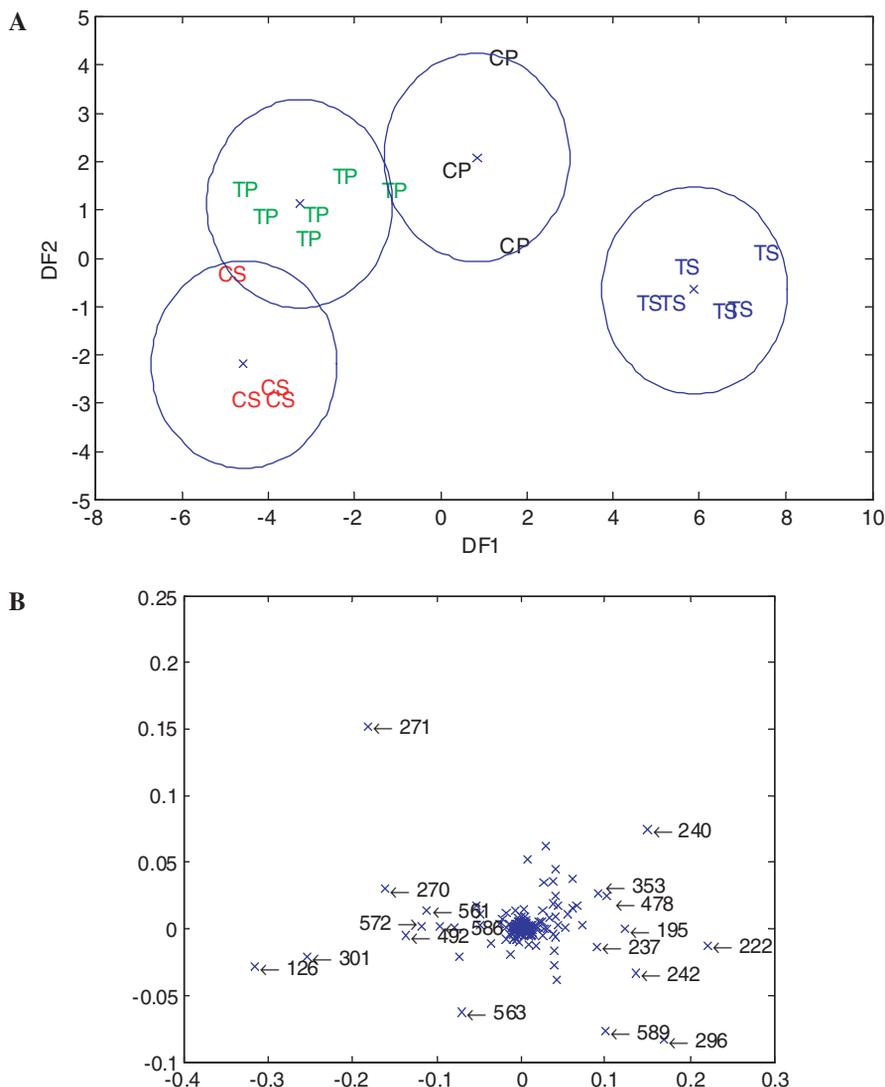
Fig. 3 PC-DFA model of the mouse serum metabolic profiles. (**A**) PC-DFA model of the mouse serum metabolic profiles, showing the discriminant functions. Each mouse is represented by a single point in model space, coded according to class membership: CP = non-transgenic littermate control at 8 weeks of age; CS = non-transgenic littermate control at 15 weeks of age; TP = transgenic mouse at 8 weeks of age; TS = transgenic mouse at 15 weeks of age when symptoms have become detectable. The circles indicate the 95% confidence regions for each class based on the unblind data. This model significantly predicts the blind validation set (χ^2; $P < 0.05$). The power of this model is illustrated, as members of the same group lie close together in the 2D plot. Each discriminant function (dimension in the plot) represents a complex linear function of the entire metabolic dataset. (**B**) Loadings of the PC-DFA model in (**A**). Loadings should be interpreted as vector quantities, with distance from the origin indicating the magnitude of their contribution to class discrimination, and the direction superimposed onto the model in (**A**) to indicate which class(es) have the greatest loading for each metabolite. Each metabolite is indicated by its arbitrary peak number. Note that the arrows indicate exactly where the point for each metabolite is in the 2D space (rather than indicating any direction). For example, metabolite 271 is higher in the TP group. To interpret the data, panels **A** and **B** should be overlaid.

Thirdly, and possibly most remarkable, we observed similar metabolite profile abnormalities in a mouse model of HD and in humans and parallel changes when comparing asymptomatic and (early) symptomatic humans and mice. It should be noted that the mouse study was analysed independently to the human data yet many of the same metabolites showed dominant changes. As far as we are aware, this is one of the first systems biology validations of a mouse model of any human disease. The convergence between the patterns seen in man and mouse also suggests the human model has some

validity, despite the lack of statistically significant external validation, and reduces the likelihood that the patterns we see are due to changes in behaviour or changes in drug treatment patterns.

HD in adults appears to be associated with a catabolic phenotype, irrespective of the symptom status. It appears that polynucleotide, fatty acid and protein metabolisms are all stimulated in the presence of a phenotype similar to that of Type I diabetes. While monosaccharide levels are raised, these are about an order of magnitude less than one would
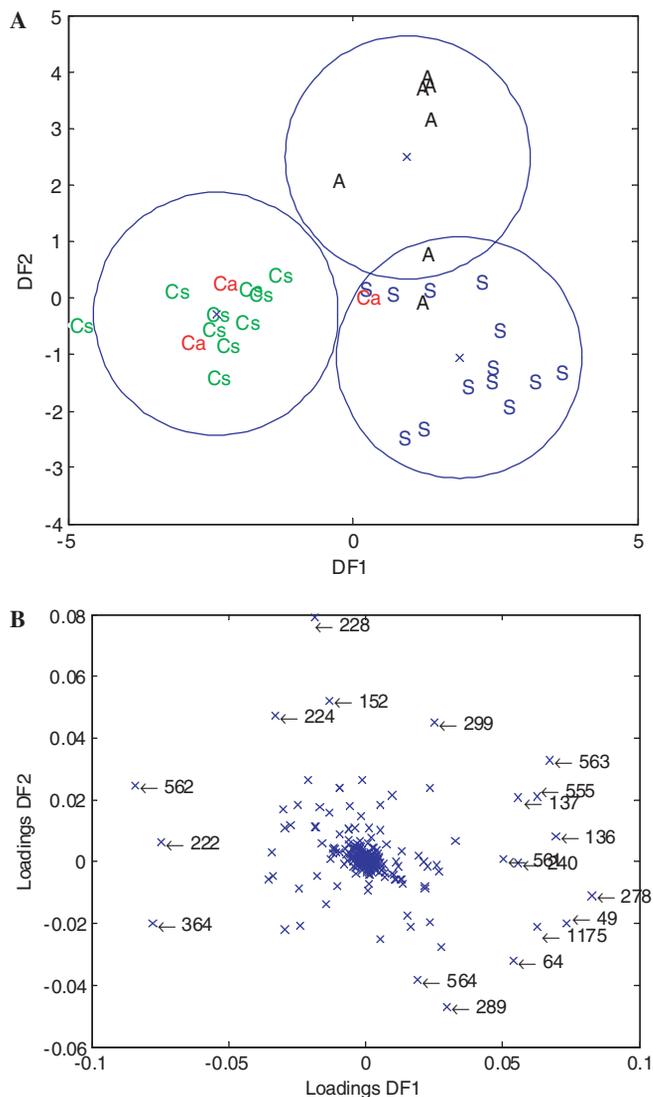
Fig. 4 PC-DFA model of the human serum metabolic profiles. (**A**) PC-DFA model of the human serum metabolic profiles, showing the first two discriminant functions. Each subject is represented by a single point in model space, colour coded according to class membership: Ca and Cs = control gene-negative subjects; A = asymptomatic gene-positive individuals; S = symptomatic gene-positive individuals. The circles indicate the 95% confidence regions for each class based on the unblind data. Prediction of the blind validation set by this model was not statistically significant (χ^2; $P > 0.05$). (**B**) Loadings of the PC-DFA model in (**A**). Loadings and arrows should be interpreted as for the model in Fig. 3.

see with diabetic patients—most cases would not be picked up as frank diabetes with conventional diagnostic criteria. Nevertheless, our data are entirely compatible with previous suggestions that HD is associated with Type I diabetes (Farrer, 1985).

The pro-catabolic state suggested by the metabolites in HD patients may account for the cachexia-like phenotypes seen late in the disease and may also account for the lower than expected weight of patients early in the disease (Djouse *et al.*,

2002) and the reported difficulties that patients have in gaining weight (Djouse *et al.*, 2002). This metabolic phenotype is also seen in the mouse model that we have studied—even though the transgene is driven by the mouse prion protein promoter, there is some pancreatic expression in this model (R. Fincham and D. Rubinsztein, unpublished data). Since the pro-catabolic phenotype is seen even in pre-symptomatic cases, it is not likely to be due to symptoms, signs or overt behavioural consequences of the disease. Our data (e.g. changes in lactate levels) therefore provide support for the hypothesis that misregulation of energy expenditure (evidenced by altered mitochondrial electron transport activity, Gu *et al.*, 1996, and reduced GAPDH activity, Mazzola *et al.*, 2001) plays a central role in the phenotype resulting from mutant huntingtin expression.

A variety of approaches can be used to sample metabolites for metabolomics studies. Gas chromatography preferentially assays volatile metabolites. Liquid chromatography will preferentially assay more compounds which are soluble in aqueous media. It generally separates compounds by their degree of hydrophobicity. While NMR is unselective by molecular structure, it is relatively insensitive. Importantly, we have coupled gas chromatography with mass spectrometry. This allows us to give a candidate molecular identity to metabolites we have identified. Furthermore, the mass spectrometry serves to deconvolute individual metabolites that may appear to almost coelute from the gas chromatography column (see Fig. 1).

In conclusion, our data raise the prospect of a robust molecular definition of progression of HD through the pre-symptomatic phase and into early symptomatic disease. This is an ideal window for therapeutic intervention. If validated in a genuinely prospective fashion in larger samples, the biomarker trajectories described here will go a long way to facilitate the development of useful therapies.

Acknowledgements

References
Beal MF, Brouillet E, Jenkins B, Henshaw R, Rosen B, Hyman BT. Age-dependent striatal excitotoxic lesions produced by the endogenous mitochondrial inhibitor malonate. J Neurochem 1993; 61: 1147–50.
Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HW, et al. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nat Med 2002; 8: 1439–44.

Brindle JT, Nicholson JK, Schofield PM, Grainger DJ, Holmes E. Application of chemometrics to 1H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. Analyst 2003; 128: 32–6.

Browne SE, Bowling AC, MacGarvey U, Baik MJ, Berger SC, Muqit MMK, et al. Oxidative damage and metabolic dysfunction in huntington's disease: selective vulnerability of the basal ganglia. Ann Neurol 1997; 41: 646–53.

Cooper JK, Schilling G, Peters MF, Herring WJ, Sharp AH, Kaminsky Z, et al. Truncated N-terminal fragments of huntingtin with expanded glutamine repeats form nuclear and cytoplasmic aggregates in cell culture. Hum Mol Genet 1998; 7: 783–90.

DiFiglia M, Sapp E, Chase KO, Davies SW, Bates GP, Vonsattel JP, et al. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science 1997; 277: 1990–3.

Djousse L, Knowlton B, Cupples LA, Marder K, Shoulson I, Myers RH. Weight loss in early stage of Huntington's disease. Neurology 2002; 59: 1325–30.

Dunn WB, Ellis DI. Metabolomics: current analytical platforms and methodologies. Trends Anal Chem 2005; 24: 285–94.

Dunn WB, Bailey NJC, Johnson HE. Measuring the metabolome: current analytical technologies. Analyst 2005; 130: 606–25.

Dunne VG, Bhattachayya S, Besser M, Rae C, Griffin JL. Metabolites from cerebrospinal fluid in aneurysmal subarachnoid haemorrhage correlate with vasospasm and clinical outcome: a pattern-recognition 1H NMR study. NMR Biomed 2005; 18: 24–33.

Farrer LA. Diabetes mellitus in Huntington disease. Clin Genet 1985; 27: 62–7.

Fiehn O. Metabolomics—the link between genotypes and phenotypes. Plant Mol Biol 2002; 48: 155–71.

Goodacre R, Vaidyanathan R, Dunn WB, Harrigan GG, Kell DG. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol 2004; 22: 245–53.

Gu M, Gash MT, Mann VM, Javoy-Agid F, Cooper JM, Schapira AH. Mitochondrial defect in Huntington's disease caudate nucleus. Ann Neurol 1996; 39: 385–9.

Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. New York: Springer-Verlag; 2001.

Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. Huntington Study Group. Mov Disord 1996; 2: 136–42.

Kell, DB. Metabolomics and systems biology: making sense of the soup. Curr Opin Microbiol 2004; 7: 296–307.

Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays 2004; 26: 99–105.

Kenny LC, Dunn WB, Ellis DI, Myers J, Baker PN, Kell DB, et al. Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning. Metabolomics 2005; 1: 227–234.

Martindale D, Hackam A, Wieczorek A, Ellerby L, Wellington C, McCutcheon K, et al. Length of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates. Nat Genet 1998; 18: 150–4.

Mazzola JL, Sirover MA. Reduction of glyceraldehyde-3-phosphate dehydrogenase activity in Alzheimer's disease and in Huntington's disease fibroblasts. J Neurochem 2001; 76: 442–9.

Odunsi K, Wollman RM, Ambrosone CB, Hutson A, McCann SE, Tammela J, et al. Detection of epithelial ovarian cancer using 1H-NMR-based metabonomics. Int J Cancer 2005; 113: 782–8.

O'Hagan S, Dunn WB, Brown M, Knowles JD, Kell DB. Closed-loop, multiobjective optimization of analytical instrumentation: gas chromatography/time-of-flight mass spectrometry of the metabolomes of human serum and of yeast fermentations. Anal Chem 2005; 77: 290–303.

Perry TL, Diamond S, Hansen S, Stedman D. Plasma-aminoacid levels in Huntington's chorea. Lancet 1969; 1: 806–8.

Phillipson OT, Bird ED. Plasma glucose, non-esterified fatty acids and amino acids in Huntington's chorea. Clin Sci Mol Med 1977; 52: 311–8.

Reilmann R, Rolf LH, Lange HW. Decreased plasma alanine and isoleucine in Huntington's disease. Acta Neurol Scand 1995; 91: 222–4.

Rozen S, Cudkowicz ME, Bogdanov M, Matson WR, Kristal BS, Beecher C, et al. Metabolomic analysis and signatures in motor neuron disease. Metabolomics 2005; 1: 101–8.

Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, et al. GC–MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett 2005; 579: 1332–7.

Schilling G, Becher MW, Sharp AH, Jinnah HA, Duan K, Kotzuk JA, et al. Intranuclear inclusions and neuritic aggregates in transgenic mice expressing a mutant N-terminal fragment of huntingtin. Hum Mol Genet 1999; 8: 397–407.

Schilling G, Jinnah HA, Gonzales V, Coonfield ML, Kim Y, Wood JD, et al. Distinct behavioral and neuropathological abnormalities in transgenic mouse models of HD and DRPLA. Neurobiol Dis 2001; 8: 405–18.

Taylor JP, Hardy J, Fischbeck KH. Toxic proteins in neurodegenerative disease. Science 2002; 296: 1991–5.

Wang, C, Kong H, Guan Y, Yang J, Gu J, Yang S, et al. Plasma phospholipid metabolic profiling and biomarkers of type 2 diabetes mellitus based on high-performance liquid chromatography/electrospray mass spectrometry and multivariate statistical analysis. Anal Chem 2005; 77: 4108–16.

Wyttenbach A, Carmichael J, Swartz J, Furlong RA, Narain Y, Rankin J, et al. Effects of heat shock, heat shock protein 40 (HDJ-2), and proteasome inhibition on protein aggregation in cellular models of Huntington's disease. Proc Natl Acad Sci USA 2000; 97: 2898–903.

Yang J, Xu G, Zheng Y, Kong H, Pang T, Lv S, et al. Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. J Chromatogr B 2004; 813: 59–65.