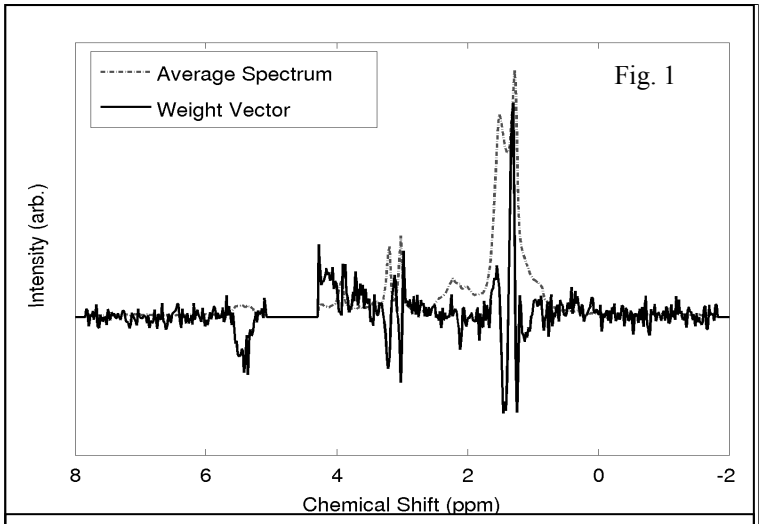
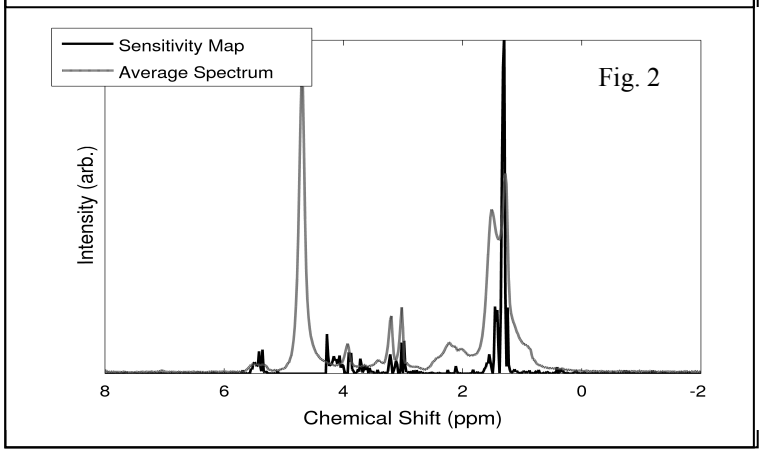
Muscle Spectroscopy shows IMCL, Creatine and Choline are Biomarkers for Adolescent Type 1 Diabetes

T. W. Ash[1], G. B. Williams[1], F. Regan[2], S. G. Harding[1], T. Saukkonen[2], D. B. Dunger[2], T. A. Carpenter[1], and A. Sleigh[1]
[1]Wolfson Brain Imaging Centre, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom, [2]Department of Paediatrics, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom

Introduction:
We have previously shown, using magnetic resonance spectroscopy (MRS), that intramyocellular lipid (IMCL) is increased in adolescent type 1 diabetics (1). Here we reanalyze the data using statistical classifiers on the entire spectrum, to be able to a) automatically classify a spectrum as being from a diabetic or control subject and b) to highlight the most discrimant parts of the spectrum for making such classifications.

Methods:
Scan Acquisition:
10 Type 1 diabetics (mean age 18.0±3.8 years) and 9 controls (mean age 16.9±0.9 years) were scanned on a Bruker Medspec S300 3.0T scanner. The voxel size was 1.5x1.5x1.5cm in the right soleus muscle (avoiding fasciae places, visible fat). 64 averages were acquired with echo time (TE) 35ms and repetition time (TR) 5s.
Data Pre-Processing:
Spectra were phased using jMRUI (2) and the real component was saved for analysis. Spectra were normalized to have the same creatine peak area and chemical shift value (an internal reference), the water peak was removed and upper and lower bounds on spectra were applied. This left 459 remaining chemical shift bins (width 0.02ppm) which were all input as independent variables to the classification algorithm.
Support Vector Machines:
We used support vector machines to classify our spectra. Linear, Polynomial and Gaussian kernels were tested, as were values of regularization parameter 'C' from 1 to 10000. Cross-validation of performance was conducted via 'leave-one-out tests', where training was performed upon 18 of the spectra, and testing on the 'left-out' subject's spectrum. For the case of the linear kernel the 'weight vector', $\mathbf{w} = \Sigma\alpha_i x_i$, and a sensitivity map (3) was recorded. The training and classification was performed using the SVM_light software package (4).





Figs.1, 2 Average weight vector (above, dark grey) and sensitivity map (below, dark grey) over all leave one out tests compared to average spectrum (light grey).

Results and discussion:
For the linear, polynomial ($2<p<5$) and Gaussian ($0.5<\sigma<10$) kernels, a 95% classification accuracy in leave one out tests was demonstrated, showing that a large amount of discriminatory information is present in the spectra of both groups. By comparison with a binomial distribution, we can reject the null hypothesis that these are chance decisions with $p=4x10^{-5}$. Splitting the groups even with prior knowledge of the whole dataset solely on the basis of IMCL CH_2 group peak area achieves a separation of 84% at best.
The averaged weight vector (fig. 1) and sensitivity map (fig. 2) demonstrate regions of the spectra used in the classification. As diabetics were classified as '1' and controls as '-1', positive values in the weight vector correspond to regions of the spectrum that were higher in diabetic spectra than in controls. For the sensitivity map, values simply indicate how sensitive the model was to changes in intensity at each chemical shift, with no sign information. The large positive signals located at 1.3ppm are consistent with that of IMCL, and support our previous study. The peaks near there with slightly higher chemical shift could be misleading, as EMCL peak area and position is very dependent on exact voxel positioning and physical orientation of muscle relative to the scanner. The weight vector surrounding the creatine and choline peaks (chemical shifts 3.02 and 3.20 ppm) was unexpected. As the spectra were normalized relative to creatine area, the negative peaks with positive side lobes seem to indicate a broadening of these peaks in diabetic subjects. This was then corroborated by an unpaired t-test on calculated line widths of these peaks, showing both choline and creatine to be widened significantly in our diabetic group.
This multivariate technique is capable of more sensitive discrimination between groups than is possible with conventional single parameter spectrum metrics. However, as a multivariate technique, we cannot assign individual confidences to regions of the spectrum using solely this technique, and in order to gain such information we have to use univariate measures in conjunction.

Conclusions:
Using support vector machines, we created a model with cross-validation accuracy of 95% in detecting the difference between muscle spectra of Type 1 diabetics and age and BMI matched controls. This multivariate technique has highlighted regions of the spectrum that vary between groups that would not normally have been tested for differences in traditional analyses, such as the creatine and choline peaks, as well as reaffirming the place of IMCL as a biomarker for the disease.

Acknowledgements: We thank the MRC for funding this research.

References:
1) Sleigh A et al. *Proc. 23rd ESMRMB Annual Scientific Meeting* 2006; 318. **2) Vapnik VN,** *Springer*, 1982.
3) Kjems U et. al. *Neuroimage 200;, 772-786.* **4) Joachims T** *in Advances in Kernel Methods: Support Vector Learning, MIT Press 1999.*