**Background**

The spectroscopy data was acquired at CAMH using a 3 Tesla GE Discovery MR750 scanner. Data was collected for 2 regions of interest (ROI) for each patient: The Dorsolateral prefrontal cortex (DLPFC), and the subgenual anterior cingulate cortex (sgACC). The repetition time (TR) and the echo time (TE) were set to 2000ms and 35ms respectively for both ROIs, with a total scan time being 5 minutes and 4 seconds. For the DLPFC, the voxel size was X(15mm), Y (30mm), Z(30mm) with a total volume of 13.5cm³ and X(20), Y(30), Z(15) with a volume of 9.0cm³ for the sgACC. To correctly orient the voxels to align with the skull (DLPFC) and the corpus callosum (sgACC) they were both obliqued multiple times. The sample size varied slightly depending on quality of data available. For the DLPFC, there were n=19 patients and n=18 controls. For the sgACC, n=22 patients and n=20 controls.

For CAMH, the data was already run through LCModel. The spectroscopy files that came off the GE scanner had the file extension ‘.7’. Two of these files were generated per subject, one for each ROI. Metabolic concentrations were analyzed with LCModel. Due to the limitations of a 3 Tesla scanner, only 6 metabolites were analyzed, these being:

1. Glutamate (GLU)
2. Glutamate + Glutamine (GLX)
3. Inositol (INS)
4. Creatine + Phosphocreatine (Cr + PCr)
5. Glycerophosphocholine + Phosphocholine (GPC + PCh)
6. N -Acetylaspartate + N- Acetylaspartylglutamate (NAA + NAAG)

The specific parameters used for LCModel are as follows: number of averages(NEX) = 8.0, TE = 35, and TR = 2000. LCModel produced a table containing a list of metabolites with their respective absolute concentrations (usually unknown units), % standard deviations (Cramer-Rao lower bounds) expressed in percent of the estimated concentration, and concentration ratios relative to Creatine and Phosphocreatine(Cr + PCr). These ratios are not used in my analysis but are a general standard since it was previously assumed that the concentration of Cr + PCr is generally similar within individuals.

**Technical Details**

Preprocessing was completed using an in-house semi-automated pipeline written in BASH and MATLAB R2015b. The pipeline depends on spm8, Gannet2.0, LCModel, and print\_raw\_headers (a proprietary program distributed by GE).

The first step for the user is to run Master1.sh through terminal, which requires the path of the input data folder, a name of the desired output folder, and the path to the executable print\_raw\_headers software mentioned above. This script first extracts data from different sources (folders and/or zip files) and organizes them into a structured format. It runs on the assumption that scans from different years will have different naming conventions and organization styles. Next, print\_raw\_headers creates MRS header files in plain text format from the raw spectroscopy data. These headers are created with the extension ‘\*.7.hdr’ (short for .7.header, where the ‘.7’ file, also called a pfile, is the file extension of raw GE spectroscopy data). Header files contain important information which is extracted when needed by other parts of this pipeline. Examples of important information include: name of ROI, prescribed series number (this is the name of a folder which contains the dicom acting as a reference point for placing the voxel on the brain), and finally image prescription number (shows which dicom from the series folder is the one used for voxel placement if multiple dicoms are present in the series folder).

This Master1.sh script internally connects to a MATLAB script which runs GannetLoad and GannetMask\_GE in that order (both included in the gannet2.0 package). GannetLoad requires the path of the ‘.7’ file for each ROI. GannetMask\_GE requires paths to the pfile, dcmdir1\*, dcmdir2, (both are created by the script in previous steps) and outputs of GannetLoad. The outputs of this script include 2x ROI mask nifti files (showing each ROI’s voxel in 2D space), 2x top-side-back view jpgs highlighting where the voxel is placed on the brain (one for sgacc and one for dlpfc), and 1x nifti T1 weighted image.

Next, the user must perform a qualitative QC by viewing each jpg image created by the previous step and ensuring that the orientation of the voxel is correct**.** If the voxel is not oriented as shown in the ‘Protocol.pdf’ document, the user must go into the file GannetMask\_GE and change the ratio of the voxel size. Refer to the document ‘QC1\_instructions’ for details on how to correctly do this. After changing the numbers run the MATLAB script again and repeat until the voxel orientation is displayed in the images as it is shown in the protocol document.

The second half of this analysis pipeline is executed by running the second bash script Master2.sh. This script calculates the contribution of CSF to each MRS acquisition voxel, and removes its effect. The measured metabolite concentrations are valid if the MRS acquisition voxel resides entirely in white matter, as the CSF contains no metabolites. However, the MRS voxels used to target the DLPFC and sgACC typically contain some CSF, and this amount will vary between participants. Therefore, it is crucial to calculate the ratio of white matter to csf in the voxel, and divide the metabolite concentration by this value:

.

Csf fraction is calculated using FSL’s Brain Extraction Tool (BET). The details of how this works are shown in the document ‘Steps4FSL\_2017-05-12’ written by Dr. Sofia Chavez. Finally, this bash script links to a MATLAB script (csv\_creation.m) which first performs CSF correction on each ROI per subject using the formula shown above, along with the table text file from LCModel which contains the metabolite concentrations, and the csf fraction found using BET. It discards the irrelevant information in the LCModel table files (such as headers and lipid concentrations), and then appends subject id, ROI, uncorrected csf concentration, csf fraction, and corrected csf concentration into a compact csv file per ROI (2 per subject). At the end of this process each subject will have 2 notable folders: one for each ROI. They each contain the table from LCModel, as well as the newly created csv.

The last step is another quality control. Each folder will have a plot **(figure 1).** From this plot note that the black baseline at the bottom should not have reoccurring dips and the red line should match the black plot to a relatively high degree of accuracy. Dips in the baseline or a plot mismatch are indications that the data isn’t accurate. It can indicate artifacts in the data, incomplete water suppression, or inaccuracies in the simulated spectra model. For the quantitative checks, under MISC OUTPUT section, the signal to noise ratio (SNR) should be greater than 10, and the full-width half-maximum (FWHM) should be less than 0.1 ppm. The FWHM is a rough estimate of the linewidth in the spectrum. Refer to the document ‘Interpreting MRS Results’ by Dr.Sofia Chavez for more information.

\*To learn more about dcmdir1 and dcmdir2, read the document ‘StepsForVoxelROI…pdf’

**Figure 1**

