Combined tissue and fluid proteomics with Tandem Mass Tags to identify circulating biomarkers of microglial cell activation in CSF from patients with Alzheimer's disease.

Claire L Russell, Amanda Heslegrave[†], Vikram Mitra, Henrik Zetterberg[†◊], Jennifer M Pocock[††], Malcolm A Ward, Ian Pike.*

Proteome Sciences plc. Institute of Psychiatry, Kings College London, South Wing Laboratories, De Crespigny Park, London, SE5 8AF

SHORT TITLE – Combined tissue and fluid proteomics with Tandem Mass Tags

KEYWORDS – biomarkers, tandem mass tags, microglia, CSF

ABSTRACT

RATIONALE - Ideal biomarkers are present in readily accessible samples such as plasma or CSF and are directly derived from the diseased tissue. As such they are likely to be of relatively low abundance as they are diluted into a large fluid volume. Traditional unbiased proteomic approaches for biomarker discovery have struggled to detect low abundance markers due to the high dynamic range of plasma and CSF proteins, the predominance of plasma albumin and other hyper-abundant proteins and the use of data dependent acquisition mass spectrometry. The use of data independent acquisition partially address sensitivity but are limited to analysis of single samples and require prior production of spectral libraries from an appropriate sample.

METHODS - To overcome these limitations and improve biomarker discovery in peripheral fluids, we have developed TMTcalibrator; a novel mass spectrometry workflow that allows the use of isobarically labeled diseased tissue digests in parallel with an appropriate set of labeled body fluids to increase the chance of identifying low abundance, tissue-derived biomarkers. Here the use of isobaric Tandem Mass Tags® allows suitable tissue or cell line samples to be labelled and mixed into equivalently labelled samples of peripheral biofluids at a range of concentrations that provide a multi-point calibration curve to improve statistical analysis, and whose total amount is sufficient to ensure the vast majority of data dependent MS/MS acquisitions are made on tissue or cell-derived peptides.

RESULTS - For each peptide that is also present in the chosen body fluid, even when at a low abundance, the fragmentation of the unique TMT® tag allows the relative abundance to be determined relative to the calibration curve created by the reference tissue calibrator. A further key aspect of this method is the use of Synchronous Precursor Selection (SPS) a form of MS3 analysis that eliminates quantitative interference from co-isolated ion species that is only available on the Orbitrap Fusion Tribrid. The use of SPS allows direct analysis of non-depleted, unfractionated CSF samples with the complete profiling of 6 individual patient samples requiring only 15 hours of mass spectrometer time, equivalent to 1.5 h per sample.

CONCLUSIONS - Here we apply the TMTcalibrator approach to identify markers of microglia activation in the cerebrospinal fluid of Alzheimer's disease (AD) patients.

INTRODUCTION

Biomarkers are indicators of pathological processes and biological events, with the levels of these markers changing as a result of disease and also in response to pharmacological intervention. As the brain is in direct contact with the cerebrospinal fluid (CSF), and pathological changes in the brain often result in altered biochemical composition of the CSF, this makes it an ideal source for biomarkers of neurological disorders. Two key features of an ideal biomarker are high specificity for disease versus non-disease and high sensitivity to distinguish between disease types. In Alzheimer's disease (AD), where much of the early CSF biomarker research was focused, three "core" CSF biomarkers (amyloid β_{1-42}, total tau and phosphorylated tau) are routinely used to diagnose AD. All three of these CSF biomarkers demonstrate high levels of sensitivity (falling within the 80-90% criteria specified by the National Institute of Neurological and Disorders and Stroke and the Alzheimer Disease and Related Disorders Work Group) but struggle to differentiate AD from other forms of dementia and neurological disorders. For example, CSF amyloid β_{1-42} levels are decreased in AD but are also reportedly lower in Lewy body dementia (LBD), fronto-temporal dementia (FTD), vascular dementia (VaD), amyotrophic lateral sclerosis (ALS) and Creutzfeldt-Jakob disease (CJD)[1]. Similarly, total tau levels are raised in AD but are also found elevated following stroke, traumatic brain injury, FTD, VaD and CJD[1]. Additional biomarkers that reflect the pathological process of AD are required to improve AD diagnosis, and to detect changes earlier in the disease process. In this study, we have analysed peptides present in the CSF from AD and non-AD patients that are common to activated microglia cells, with the aim of identifying potential biomarker candidates that are specific to the early changes in the AD brain.

Microglia are a type of glial cell, and the resident macrophage-like cells of the central nervous system (CNS)[2]. Microglial cells perform a variety of different functions within the CNS related to the immune response and the maintenance of cellular homeostasis. Much interest has focused on the role of microglia in neurodegenerative diseases (particularly AD) in recent times due to an increasing number of genetic and biomarker discovery programs and growing evidence to suggest that inflammation has a role in disease pathogenesis[3]. An early

event in neurodegenerative pathology involves activation of microglial cells[4]. Microglia sense the environment around them and can be activated by various factors. This activation can take different forms, and the so called M1 activation state causes the release of pro-inflammatory cytokines such as TNF-α, IL-1β and reactive oxygen/nitrogen species ROS/NOS[5]. This may ultimately cause neuronal damage. The M2 activation state is thought of as protective and anti-inflammatory by blocking the release of pro-inflammatory cytokines, ingesting debris, promoting tissue repair and releasing neurotropic factors[5]. We hypothesize that the activated microglia will secrete distinct proteins into the brain and subsequently into CSF during the early stages of AD pathology. The challenge is how to find these molecules in a biofluid with a high protein dynamic range, and hyper-abundant proteins such as albumin. To answer this question, we have applied a novel Tandem Mass Tag® (TMT®) mass spectrometry (MS) approach; TMTcalibrator, to look for markers of microglia activation within the CSF of AD patients.

BV2 cells are a microglial cell line derived from raf/myc-immortalised murine neonatal microglia[6] and have been used extensively to study microglial activation. Studies have shown their utility as a surrogate for primary microglial cultures which are costly in terms of animals used and time consuming to prepare[7,8]. Interferon-gamma (IFNγ) and lipopolysaccharide (LPS) treatment were used to activate the BV2 cell line[9,10,11], followed by cell lysis and tryptic digestion, and labeling of peptides with four TMT10plex tags to form a 4-point calibrator. This calibrator was mixed into six TMT® 10plex labelled CSF samples; 3 from AD patients and 3 from non-AD controls. By virtue of reporter ions produced specifically by each TMT reagent upon fragmentation, we can determine which activated microglial proteins are also found in CSF and which, if any, are differentially expressed in AD compared with non-AD individuals with impaired cognitive function.

METHODS

REAGENTS

All reagents for the cell culture and digestion of the samples were purchased from Sigma Aldrich (Dorset, UK) unless stated. Tandem Mass Tags® (Thermo Scientific), Acetonitrile (Fisher Scientific, Loughborough, UK) Trypsin (Roche Diagnostics, West Sussex, UK).

PROCEDURES

Samples –BV2 cells were seeded into 6 well plates at 80,000 cells per well and maintained at 37°C at 5% CO_2. Cells were cultured for 24 hours in Dulbecco's modified eagle medium (Gibco®, Life Technologies) with 10% fetal bovine serum (Gibco®) supplemented with 2mM L-glutamine, 100U/mL penicillin and 100mg/mL streptomycin, gentamycin and mycoplasma removal agent (AbD Serotec). To activate the cells, 2µg/ml LPS and 10ng/ml IFNγ (R&D Systems) treatment was applied for 24 hours prior to harvesting. Cells were lysed in 8M urea, 75mM NaCl, 50mM Tris (pH 8.2) before sonication (20 x 1sec 20% amplitude) on ice and

centrifuged at 12,500g for 10mins at 4°C. The protein concentration of the cell lysate was estimated by Bradford assay. Samples were aliquoted to prevent freeze-thawing and stored at -80°C.

In-Solution Tryptic Digest and TMT labelling – Following solubilzation and denaturation in 100mM TEAB buffer and 0.1% SDS, the CSF samples and the BV2 cell-line (calibrator) were reduced with 1mM TCEP at 55 °C for 60 min and alkylated with 7.5mM iodoacetamide at room temperature for 60 min. Trypsin was used for digestion at an approximate 1:25 weight ratio of trypsin-to-total protein and incubated at 37 °C overnight (~18 h). Following digestion, the peptides were labeled with one of the TMT 10plex reagents at 15mM and incubated for 60 min at room temperature. To quench the TMT® reaction 0.25% hydroxylamine was added to each sample and incubated for 15 min. The samples were then combined as detailed in the text and incubated for a further 15 min. The sample was desalted using RP18 columns and excess reagents and SDS were removed by SCX purification. Peptides were eluted in 75% ACN + 400mM NH_4Oac and dried to completion. For MS analysis the sample was resuspended in 2% ACN + 0.1% formic acid.

LC-MS/MS - Quantitative analysis was performed using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) in the positive mode with EASYnLC1000 system and a 50cm EASY-Spray column (Thermo Scientific). The column temperature was maintained at 40°C, and the peptides were separated at a flow rate of 200 nL/min. Peptides were eluted from the column over a 300 min gradient, from 3% to 30% solvent B (acetonitrile with 0.1% formic acid) in 280 min, followed by an increase to 80% solvent B in 10mins, which was held for a further 10mins. Solvent A was water with 0.1% formic acid. Wash and blank LC–MS/MS runs preceded the analysis.

A Top10 CID-HCD MS3 SPS method with a 2 second cycle time was utilized and the parameters were as follows: MS; Spray voltage - 2000V; ion transfer tube temperature - 275 °C; detector – Orbitrap; scan range (m/z) - 400-1400; resolution – 120000; AGC target - $5x10^5$. MS/MS (CID); Isolation mode – quadropole; collision energy – 35%; detector – Ion trap; AGC target - $4x10^{3.}$ MS3 (HCD); detector – Orbitrap; collision energy – 55%; scan range – 100-1000 m/z; resolution power – 30000; AGC target - $6x10^4$.

Bioinformatics - LC-MS/MS data was initially processed within Proteome Discoverer 1.4 and an Excel spreadsheet was created by exporting directly from the software. A selection of automated scripts, written in R, were subsequently used to produce the graphical outputs and final table of results.

Cerebrospinal fluid samples – CSF samples were from patients who sought medical advice because of cognitive impairment. Patients were designated as normal or AD according to CSF biomarker levels using cutoffs that are 90% sensitive and specific for AD[12]: total tau (T-tau) >350 ng/L, phospho-tau (P-tau) >80 ng/L and Aβ42 <530 ng/L. None of the biochemically normal subjects fulfilled these criteria. CSF T-tau, P-tau and Aβ42 levels were determined using INNOTEST enzyme-linked immunosorbent assays (Fujirebio, Ghent, Belgium) by board-certified laboratory technicians according to protocols approved by the

Swedish Board for Accreditation and Conformity Assessment (SWEDAC). The study was approved by the regional ethics committee at the University of Gothenburg.

RESULTS AND DISCUSSION

In order to enhance the MS data acquisition from the cell line calibrator, TMTcalibrator has been designed to ensure the calibrator proteins, in this case the BV2 cell line proteins, dominate the overall protein content of the TMT 10plex sample. As we are limited by the amount of protein that can be injected onto the EASY-Spray column a balance between dominant calibrator and analytical samples must be achieved. We designed the 4-point TMTcalibrator to be spiked into the total combined CSF samples at a ratio of 1:4:6:10. These are arbitrary values allowing us to determine a suitable protein load from the six CSF samples for analysis. The six individual CSF samples were combined at a 2:2:2:2:2:2 ratio, falling within the 4-point calibrator range. As the maximum load on column is 2µg, the true protein load per channel can be determined by dividing the maximum 2µg by the total number of arbitrary values (x) needed, which in this case is 33 (calibrator channels plus CSF channels). If $33x$ equals 2µg, $1x$ is therefore 0.06µg and this can then be used to determine the total amount of protein from each of the 10 channels contributing to the TMTcalibrator analytical sample. Just 0.12µg of each CSF sample is required ($2x$), along with a 4-point calibrator comprised of 0.06µg ($1x$), 0.24µg ($4x$), 0.36µg ($6x$) and 0.6µg ($10x$). This results in a total of 0.72µg of protein from all the CSF samples and 1.26µg of protein from the BV2 cell line calibrator – the calibrator proteins are therefore 1.75 fold more prevalent than the total CSF protein load.

The TMTcalibrator sample was analysed in triplicate on the Oribtrap Fusion Tribrid using an MS3 SPS method over a 300 minute gradient. The raw data files from all three MS injections were searched using SequestHT and Mascot in Proteome Discoverer 1.4, against the Uniprot human protein database. Results for the triplicate injections were merged resulting in one large data set for analysis. After applying a 5% false discovery rate (FDR) to filter the data in Proteome Discoverer, the number of peptide and protein IDs from calibrator-derived only, CSF sample derived only, and peptides common to both the microglia cell line calibrator and CSF samples, was determined. The design of the TMTcalibrator experiment ensured that the MS acquisition was driven by the more prevalent microglia cell line peptides, and this was evident in the peptide IDs (Figure 1). No peptides specific to CSF samples were detected, confirming the premise that the TMTcalibrator approach circumvents the issues surrounding highly dominant CSF proteins. Instead, the peptides identified were either specific to the microglia cell line (2,317), or common to CSF and microglial (11,150). These shared peptides were the focus for further analysis to isolate any markers of microglial activation present in the CSF. All peptides with signal for all 10 TMT® reporter ions were taken forward for processing through a set of in-house bioinformatics scripts developed in R.

The reporter ion signal intensities from the six CSF channels were normalized by sum-scale normalization, a mathematical approach to remove experimental bias. The process involves summing the intensity values for all analytes measured in a given sample and then calculating the median value across all the summed values. The median value is divided by each summed

value to create a correction factor which is multiplied to the original intensity values to give the normalized sum scaled measurement[13,14]. Following normalization the variability across the CSF dataset was investigated using Principal Component Analysis (PCA). Group separation of AD and control CSF was seen at the peptide and protein level, forming the first principal component, and accounting for 44.6% of the total variation seen in the dataset. The second principal component explained 19.6% of the variability in the dataset, and this corresponded to the biological variation within the groups (Figure 2).

To further filter the data, we normalized the cell-line calibrator data to one of the TMTcalibrator channels (TMT129). This was carried out for the four channels dedicated to calibrator, which allowed us to filter for an expected TMTcalibrator signal signature or pattern. Following normalization, the ratio of the signal intensity in channels 129, 130e, 130 and 131 (calibrator channels) was expected to be 1:4:6:10. To measure a correspondence between expected calibrator signal intensities to observed calibrator intensities, an R^2 value across all four channels was calculated for every single peptide. PSM level intensities for the four TMTcalibrator channels show marginal deviation from expected linear calibrator signal intensity ratio of 1:4:6:10 (Figure 3). 39.3% of the peptides had a calibrator R^2 of ≥ 0.95 and were taken on for further data analysis.

The normalized dataset, filtered for the presence of an expected linear calibrator signal intensity was further filtered based on the significance following a 2-sample t-test across the two CSF groups for every single peptide and protein. Peptides with a p value ≤ 0.05 were considered significant. The log2 ratio of the average AD CSF signal intensities (channels 128e, 128 and 129e) to control CSF signal intensities (channels 126, 127e and 127) was also normalized to the 129 TMTcalibrator channel. This provided a means of identifying which of the peptide sequences common to the activated microglia cell line and CSF samples were differentially regulated between control and AD CSF. We identified 84 unique peptide sequences from 77 proteins that were significantly up-regulated at least 60% (log2 ratio $\leq$ -0.7) in AD CSF compared to controls, and 34 unique peptides from 26 proteins that were significantly down-regulated by at least 60% (log2 ratio $\geq$ 0.7).

When the list of regulated peptides in the CSF of AD patients is analysed, the benefit of TMTcalibrator over the traditional proteomics approach can be seen. As stated, the maximum protein load on the 75µm diameter analytical LC column per injection is 2ug. In the TMTcalibrator design the 2µg is split unequally between the 4-point calibrator and the six CSF samples (1:4:6:10 ratio for the calibrator, 2:2:2 for AD CSF and 2:2:2 for non-AD CSF). Without the addition of the calibrator, the six CSF samples could be labelled with TMT® in a multiplex experiment, with 0.33µg combined from each sample per inject. This would not overcome the problems associated with high abundant CSF proteins, and these would dominate the MS acquisition, preventing lower abundant markers from being detected. To highlight the different results seen after inclusion of the cell line calibrator, the same six CSF samples used in the TMTcalibrator study were equally combined and a total load of 2µg was analysed in triplicate using the same method. Following the same bioinformatics processing, the results were filtered and analyzed to identify significantly regulated peptides in the AD

CSF compared to control CSF samples. 73 proteins were significantly regulated by at least 60% (up or down regulation) but only 16 of these overlapped with significantly regulated peptides in the TMTcalibrator+ study (Figure 4). The majority of the proteins identified in the CSF samples are known to be secreted proteins, annotated as such under their Uniprot accession entries (such as plasminogen, mimecan and hemopexin). Only 5 proteins are found which are localized to the nucleus, whereas with the addition of the calibrator, 29 of the proteins from the significantly regulated list are known to reside within the nucleus. A large proportion of the significantly regulated proteins found without calibrator are proteins that are commonly observed in CSF studies including serum albumin, complement proteins (C3, C5 and C7), fibrinogen gamma chain and Ig gamma chain[11]. These proteins are likely representative of disease related changes of low specificity rather than the cellular changes specific to AD. They are some of the most abundant proteins found in the CSF and are reported in the majority of CSF proteomic analysis studies[15]. In comparison, the majority of peptides derived when the BV2 cell line is used to drive the experiment are cellular and may represent a cellular response specific to disease. These include peptides from dynactin (Figure 5a), cofilin (Figure 5b), alcohol dehydrogenase, filamin-A, myosin proteins and Ras-related proteins. This supports the use of TMT calibrator workflow, and demonstrates that the inclusion of an appropriate calibrator, in this case the BV2 cell line, allows for a more targeted and thorough analysis of biofluids from case and control subjects.

Another alternative to the multiplexing would be to analyze the samples individually which would allow 2µg of each individual CSF sample to be injected however, due to the lack of a reliable calibration curve such an approach would not allow for reliable quantitation.

The successful application of TMTcalibrator in this study has provided a novel list of potential biomarker candidates likely originating from activated microglia and circulating in CSF. Whilst these markers of microglial activation need further validation (experiments are ongoing at the time of submission), this study has demonstrated the benefits of the new TMTcalibrator workflow and the results suggest this is a suitable and efficient method of detecting low abundant peptides within biological fluids. The use of TMTcalibrator in further biomarker discovery studies should be considered to overcome some of the limitations commonly associated with more conventional approaches.

#