12-1-2019

Incorporating In-Source Fragments Improves Metabolite Identification Accuracy in Untargeted LCMS and LCMS/MS Datasets.

Brian C Searle

Jacob C Lippincott

Phillip M Seitzer

each community, were utilized to determine the additional bias introduced by amplification and/or sequencing reagents.

Significant differences were noted in alpha and beta diversity of samples extracted using each protocol, particularly as compared to the expected values, consistent with previous findings. Additionally, there were unexpected genera found within these communities indicating contamination uniquely introduced by each protocol. Trends were noted among different kits, in either over- or under-representation of specific genera of communities. Similar trends were observed in the results of sequencing controls, indicating the contamination and/or bias introduced after extraction. Overall, these findings illustrate the importance of utilizing controls at multiple stages of the process, to better assess the impact of the findings of any human samples.

63

Implementation of Multiplex Staining, Imaging and Analysis as a Standardized Service for Researchers

Shanmugapriya Selvaraj[1], Valeria Mezzano[1], Branka Brukner Dabovic[1], Cynthia Loomis[1]

[1]NYU Langone Medical Center

Several modalities of multiplex immunofluorescence histology currently available require significant time and resources to implement. Many research laboratories develop questions benefiting from multiplex staining and analysis but do not have the human resources and/or equipment to perform the assay. Furthermore, pilot studies use similar metrics to evaluate multiplex histology data making them ideal for a core laboratory setup. The objective of this study was to establish a semi-automated workflow for multiplex immunofluorescence staining and initial quantification of cell populations in whole slide microscopy scans. The requirements for the workflow included: A. minimal transfer of decision making from the researcher to core personnel (Semi-Automation), B. modifiable in terms of antigen targets and tissue types with minimal disruption to the process and C. reproducible across samples submitted at different time periods (eg patient samples). Methods: Leica Bond RX auto stainer was used for staining. Akoya Biosciences Phenoptics platform was used for multiplex staining, imaging was performed with the Vectra3 multispectral imaging system followed by InForm quantification. Further analysis was conducted with the statistical analysis package R. Cross platform imaging was tested with a Zeiss 880 confocal microscope. Results: Five investigators are using the workflow for their projects that range from clinical pilot studies to basic science with animal models. Our workflow for one slide involves 24 hours for staining, one hour for scanning setup, 10 minutes for whole slide scanning, up to 2 hours for automated sampling, and initial analysis setup 2 hours. Conclusions: the key steps for establishing a robust workflow include antibody optimization using a combination of chromogenic (for determination of specificity/sensitivity) and fluorescent testing in multiplex panels (to determine appropriate dilution according to corresponding fluorophore). Explicit statement from the researchers regarding what to quantify (populations of interest, areas of interest) optimizes analysis output.

64

Incorporating In-Source Fragments Improves Metabolite Identification Accuracy in Untargeted LCMS and LCMS/MS Datasets

Brian C. Searle[1], Jacob C. Lippincott[2], Phillip M. Seitzer[2]

[1]Institute for Systems Biology, [2]Proteome Software

In untargeted metabolomics experiments library search engines detect metabolites using several features, including precursor mass, isotopic distribution, retention time, and MS2 fragmentation. Matching acquired MS2 to library spectra is vital as numerous compounds share molecular formulas, resulting in identical precursor measurements and similar retention times. However, many metabolomics experiments are still collected using LC-MS only, and even in LC-MS/MS experiments many precursors lack MS2 spectra due to the stochastic nature of data dependent acquisition. We observe that when metabolites ionize they can produce unanticipated MS1 features resulting from neutral losses, in-source fragmentation, multimerization, and adducts. Here we present a new approach to leverage these measurements to identify metabolites when MS2 spectra are of low quality or not available. We processing datasets of 75 known standards mixed with whole yeast lysates to strip them of their MS2 scans to produce a gold-standard MS1-only data set of a complex metabolome with known targets. For each dataset we determined the proportion unambiguous annotations (where the correct annotation had a higher score than other potential annotations) and unmistakable annotations (where the correct annotation was the only valid annotation detected). We found that incorporating in-source fragments improved these metrics for both MS1-only (increasing from 60% to 73% unambiguous and 40% to 65% unmistakable matches) and MS2 datasets (from 79% to 84% unambiguous and 41% to 60% unmistakable). Unexpectedly, in these data we observed that the MS2 spectra were less useful than in-source fragment data for improving identification accuracy. We believe this is largely because the low-resolution iontrap MS2 spectra collected in this experiment show significant noise, which diminishes spectral match scores and allows other candidates to outscore the correct identifications. We suspect that noise is less likely to affect MS1 peak groups because they are generated from data aggregated across multiple high-resolution MS1 scans.

65

Infectious Disease Metagenomics: Error Mitigation and Best Practices for the Clinical Routine Use of Metagenomic Sequencing

Nur A. Hasan[1], Huai Li[2], Brian Fanelli[2], Arne Materna[2], Manoj Dadlani[2], Rita R. Colwell[3]

[1]CosmosID, University of Maryland, [2]CosmosID, [3]University of Maryland, Johns Hopkins Bloomberg School of Public Health, CosmosID

Shotgun metagenomic sequencing is increasingly adopted by the biomedical community for clinical infection