

Brian Pratt <bspratt@proteinms.net>

Re: Regarding Lipid Quan using Skyline

Brian Pratt <bspratt@proteinms.net>
To: Seul Kee Byeon <seulkee@jhmi.edu>
Cc: "brendanx@proteinms.net" <brendanx@proteinms.net>

Wed, Jul 25, 2018 at 12:01 PM

We're on the right track. So there is a directory or file somewhere that's called SGA_SERUM, and contains a bunch of search result file? If so, can you zip up the entire directory and send it to me? That zipped-up directory might be the starting point for our automation. With a commonly used tool like LipidSearch we'd want to put some effort into removing manual steps from there to Skyline.

Thanks

Brian

On Tue, Jul 24, 2018 at 3:02 PM, Seul Kee Byeon <seulkee@jhmi.edu> wrote:

Hi Brian

I apologize for such late reply. I don't know if you are aware but our lab is moving in about 2 weeks so I've been analyzing and organizing the data of all projects I've been working on so we can have one last meeting with each collaborator before we go. So I always got carried away with other stuff and never managed to finish this e-mail to you.

To answer your questions,

1) So each of these zip files represents the output of a single LipidSearch run - were they originally in some kind of directory hierarchy? I'm trying to imagine what exactly we would point our reader at (a directory with subdirectories, or a list of directories, etc?)

-> Yes, each zip file contains output of a single LipidSearch run from different files. I selected all files and carried out the LipidSearch identification under the same condition, with a single click for each mode. Control and SGA were all analyzed at the same time but I just placed the result files in the different folder. That's why there are separate zip files for control and patients. But I carried out the identification with a single LipidSearch run for EACH mode.

PC in negative I did separately because identification of PC in positive ion mode using HCD is soooooo poor. In order to determine their structure, I had to analyze them in negative ion and I did separately from other classes for my convenience.

Then I placed the result files under separate folders/directory for my convenience as shown below.

I placed them under "SGA serum" folder and created subfolders for different groups and different ion modes.

The screenshot shows the Skyline software interface. On the left is a file tree with folders such as 'AD_CSF', 'Barth syndrome', and 'SGA serum'. The 'SGA serum' folder is circled in blue and labeled with a '2'. On the right is a data table with columns: Name, RawData, Type, and ExpType. The table contains 21 rows of data, all of which are 'Product' type and 'LC' ExpType. The 'Name' column contains various lipid identifiers, including 'Barth SM LSM 30' and 'Barth POS general'. The 'RawData' column lists the corresponding raw data files.

	Name	RawData	Type	ExpType
1	Barth SM LSM 30	MF_02_CE30.raw	Product	LC
2	Barth SM LSM 30	LF_02_CE30.raw	Product	LC
3	Barth SM LSM 30	E1_02_CE30.raw	Product	LC
4	Barth SM LSM 30	CD_02_CE30.raw	Product	LC
5	Barth SM LSM 30	104_02_CE30.raw	Product	LC
6	Barth SM LSM 30	018_02_CE30.raw	Product	LC
7	Barth SM LSM 30	007_02_CE30.raw	Product	LC
8	Barth POS general	MF_03.raw	Product	LC
9	Barth POS general	MF_02_CE30.raw	Product	LC
10	Barth POS general	MF_01.raw	Product	LC
11	Barth POS general	LF_03.raw	Product	LC
12	Barth POS general	LF_02_CE30.raw	Product	LC
13	Barth POS general	LF_01.raw	Product	LC
14	Barth POS general	E1_03.raw	Product	LC
15	Barth POS general	E1_02_CE30.raw	Product	LC
16	Barth POS general	E1_01.raw	Product	LC
17	Barth POS general	CD_03.raw	Product	LC
18	Barth POS general	CD_02_CE30.raw	Product	LC
19	Barth POS general	CD_01.raw	Product	LC
20	Barth POS general	104_03.raw	Product	LC
21	Barth POS general	104_02_CE30.raw	Product	LC

2) What steps exactly did you take when producing that first spreadsheet you sent? That's the kind of thing we'd want to automate, but bringing along a bit more information from the input files.

-> So what I did was

- 1) I combined the identified lipids from ALL result files (while deleting the duplicated ones) to come up with a comprehensive list of identified lipids
- 2) found the calculated m/z values (used "VLOOKUP" function in excel)
- 3) found the Top RT time (used "VLOOKUP" function in excel)
- 4) filtered out lipids under the following conditions, regardless of their grades. Everyone has their own criteria for filtering out and this is just mine.

A. lipids with odd number of carbons in fatty acyl chains (fatty acyl chains with odd number of carbons are not synthesized in human body. They're reported to come from foods people take. Since I mostly study

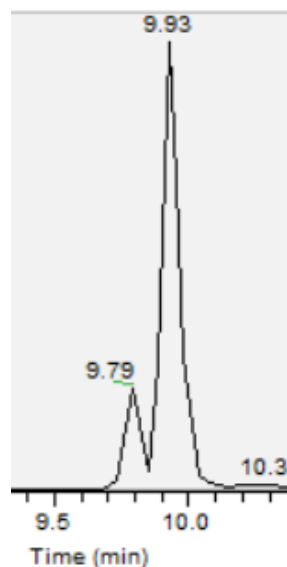
biomarkers, I automatically filter these out. The ones with 17:0 or 17:1 fatty acyl chains in that spreadsheet are internal standards that I specifically added for normalization. That's why there are in that file even though they are odd numbered chains)

B. Number of carbons in each fatty acyl chains larger than 28 (28:0, 32:1 etc.) or smaller than 12.

C. unusual combination of fatty acyl chains such as 18:5. High degree of unsaturation bonds such as 4,5,6 are mostly found in carbon length of 20~22.

* But I do manual examination and include unusual chain such as 24:5 if the ions in MS/MS spectra are all correct including relative intensities.

So one of the things was Top RT was that there were some species such as lyso-PC where LipidSearch kept giving me inconsistent RT across files. Many lysolipids (especially PC) elute in two peaks because they're regioisomers and LipidSearch sometimes give the RT based on the 1st peak on some files while it gives the RT based on the 2nd peak on some files. So I had to manually make sure that correct RT was selected.



Let me know if I misunderstood the intention of your questions.

Thanks.

- Seul Kee

From: Brian Pratt <bspratt@proteinms.net>

Sent: Friday, July 20, 2018 12:29 PM

To: Seul Kee Byeon

Cc: brendanx@proteinms.net

Subject: Re: Regarding Lipid Quan using Skyline

Hi Seul Kee,

That looks promising. Our usual goal for working with search results is to require no user manipulation of the files - it's easier for the user, and easier for us since we don't have to worry about anything other than standard file formats.

So each of these zip files represents the output of a single LipidSearch run - were they originally in some kind of directory hierarchy? I'm trying to imagine what exactly we would point our reader at (a directory with subdirectories, or a list of directories, etc?).

What steps exactly did you take when producing that first spreadsheet you sent? That's the kind of thing we'd want to automate, but bringing along a bit more information from the input files.

Thanks

Brian Pratt

On Thu, Jul 19, 2018 at 3:45 PM, Seul Kee Byeon <seulkee@jhmi.edu> wrote:

Hi Brian

I used LipidSearch from Thermo for identification and the attached files are the result files.

Since LipidSearch doesn't provide combined result from multiple files, I have to run the identification individually and combine the results myself.

So the one I sent you earlier (m/z and retention times table) are the combined files all samples with false positives filtered out.

Control zip files contain identifications made from files named 02, 05, 06, 07 and 08 while SGA zip files contain those from 01, 03, 04, 09, and 10.

As fragments for identification of fatty acyl chains for PC are not observed well under HCD fragmentation, I had to analyze the structures in negative ion mode even though the peak intensities are lower than

positive ion mode. So you can pretty much ignore results for PC made in positive ion mode since most of their grades are C.

You can consider the "grade" as the identification confidence.

A is when both fatty acyl chains and a head group of lipid are identified

B when one fatty acyl chain and head group identified

C when only head group is identified and so on.

I usually select those with grades A or B.

Let me know if you have any questions.

Thank you.

- Seul Kee

From: Brian Pratt <bspratt@proteinms.net>

Sent: Thursday, July 19, 2018 3:11 PM

To: Santosh Renuse

Cc: brendanx@proteinms.net; Seul Kee Byeon

Subject: Re: Regarding Lipid Quan using Skyline

Hi Seul Kee,

In terms of search results we're also interested in things like identification confidence, details on file+scans in which IDs were made, and so forth.

For reference, here's some information of the things we collect for spectral libraries built from search results:

<https://skyline.ms/wiki/home/software/BiblioSpec/page.view?name=BiblioSpec%20input%20and%20output%20file%20formats>

Best Regards,

Brian Pratt

On Thu, Jul 19, 2018 at 12:02 PM, Santosh Renuse <santosh@jhmi.edu> wrote:

Thanks Brian and Brendan for your prompt response.

Seul Kee, can you please give more information to Brian regarding LipidSearch results.

Thanks,

Santosh

From: Brian Pratt <bspratt@proteinms.net>
Sent: Thursday, July 19, 2018 2:58 PM
To: Santosh Renuse <santosh@jhmi.edu>
Subject: Re: Regarding Lipid Quan using Skyline

Hi Santosh,

Thanks for these files (still downloading!).

Those search result files are pretty bare - is it possible to get more information out of LipidSearch, such as molecular formula, or InChI etc? We certainly can work with mz alone, but with a chemical formula we can look at isotope distributions etc.

Thanks,

Brian Pratt

On Thu, Jul 19, 2018 at 11:02 AM, Brendan MacLean <brendanx@proteinms.net> wrote:

Thanks, Santosh. I have added Brian Pratt our small molecule lead to this thread. He has been working with others on lipid quantification. I am sure he will be interested in seeing to output of your "discovery" pipeline to better integrate it into our targeted validation workflows.

On Thu, Jul 19, 2018 at 10:42 AM, Santosh Renuse <santosh@jhmi.edu> wrote:

Hi Brendan,

Hope you are doing great !

As discussed with you regarding using Skyline for Lipid Quantitation, I have uploaded result files (from LipidSearch software) and raw files on following ftp location:

Sample_Sample#_Replicate (Con_1_01 etc.). You could use SHORT list file for lipid quan for testing.

<ftp.ibioinformatics.org>

username – public

password – iob4ever

Please let us know if you need any more information.

Sincerely,

Santosh

--

Santosh Renuse, Ph.D.

McKusick-Nathans Institute of Genetic Medicine

Johns Hopkins University School of Medicine

560 Miller Research Building

[733 N. Broadway](#)

[Baltimore, MD 21205](#)

Phone: 410-502-6665

Cell: 443-301-4129