Supplementary Data

MetaboHunter Testing

Different NMR spectra, combinations of spectra as well as spectra of a chemical mixture were used for testing different methods provided as part of MetaboHuneter. Validation of results was performed in all cases by analyzing results of searches using complete, unbinned spectra as well as externally determined peak lists (peaks were determined using MNova software). The Human Metabolomics Database (HMDB: Wishart et al., 2009) contains different information about many metabolites including their NMR spectra and these spectra were used for the majority of tests presented here. In the examples shown below both metabolites with large and the ones with small number of spectral peaks are included. Also, we have included examples of metabolites with some non-unique peaks. Results of the search for spectra of several individual metabolites are shown in Table 1. For the analysis of MetaboHunter's ability to assign metabolites from the spectra of mixtures we have combined spectra for 13 metabolites. Assuming that there is no chemical interaction between molecules, the NMR spectrum of a mixture is a direct sum of spectra from the components. For this test we have selected 13 water soluble, *i.e.* hydrophilic metabolites that were consistently observable in standard 1D ^{1}H NMR of cell culture metabolomics (Duarte et al., 2009; Griffin et al., 2002; Yang et al., 2007; Gotschalk et al., 2008; Tiziani et al., 2009). The results of assignments for this spectrum are presented in Table 2. Finally, MetaboHunters's assignment power was investigated against the NMR measurements of a mixture of 5 metabolites performed in house on a 200MHz NMR instrument. The peak database used by MetaboHunter includes measurements from 500MHz and 600MHz instruments. Thus validation with spectra from 270MHz instrument shows the validity of methods provided under MetaboHunter platform for assignment of data coming from different NMR instruments. The results of this analysis are shown in Table 3.

Table 1. Results of assignment of spectra obtained for selected metabolites from HMDB. The assignment was performed with MetaboHunter and HMDB. Search was performed starting from the complete spectra as well as the peak lists. In the examples we have selected metabolites with highly involved spectra and with overlapping peaks with other molecules in the database. Presented are the search scores, number of peaks found relative to the number of peaks in the database (in brackets) and finally the rank of the metabolite in the complete list of found metabolites (bold). Results from searches with MetaboHunter (MH1: Highest number of matched peaks method; MH3: Greedy selection of metabolites method) and HMDB are shown.

Input spectra - source	**Metabolite name**	**Spectral Search**		**Peaks search**		**HMDB**
		MH1	**MH3**	**MH1**	**MH3**	
HMDB00895	Acetylylcholine	67(4/5) **#4**	67(4/5) **#3**	67(4/5) **#1**	67(4/5) **#1**	5/5 **#3**
HMDB00429	17-b-Estradiol	97(86/88) **#1**	97(86/88) **#1**	NF	NF	129/78 **#1**
HMDB0067	Cholesterol	99(92/92) **#2***	99(92/92) **#1**	96(89/92) **#1**	96(89/92) **#1**	112/78 #6
HMDB00122	D-Glucose	97(38/38) **#1**	97(38/38) **#1**	95(37/38) **#1**	95(37/38) **#1**	(54/54) #6

* In this search Campesterol (plant equivalent of cholesterol) was found as #1 with 109/109 peaks. Doing search just for mammalian metabolites would lead to cholesterol being #1 find.

Table 2. Results of a metabolite search for a spectrum obtained as a combination of spectra for 13 metabolites obtained from HMDB. Spectra of individual metabolites were corrected to equal concentration (115mM). Results from searches with MetaboHunter (MH1: Highest number of matched peaks method; MH3: Greedy selection of metabolites method) and HMDB are outlined.

HMDB ID	Metabolite	Spectral search		Peaks search		HMDB
		MH1	MH3	MH1	MH3	
HMDB00097	Choline	88(7/7) **#8**	88(7/7) **#3**	88(7/7) **#12**	88(7/7) **#4**	NF
HMDB00125	Glutathione	76(22/28) **#105**	NF	76(22/28) **#118**	NF	NF
HMDB01310	L-Alanine	71(5/6) **#201**	NF	86(6/6) **#21**	86(6/6) **#5**	6/6 **#37**
HMDB03339	L-Glutamic Acid	82(18/21) **#38**	NF	91(20/21) **#4**	NF	NF
HMDB00641	L-Glutamine	86(18/20) **#13**	86(18/20) **#4**	90(19/20) **#5**	NF	NF
HMDB00687	L-Leucine	88(15/16) **#7**	88(15/16) **#2**	88(15/16) **#9**	88(15/16) **#3**	NF
HMDB00883	L-Valine	85(11/12) **#18**	85(11/12) **#7**	84(11/12) **#27**	NF	NF
HMDB00168	L-Asparagine	85(11/12) **#19**	NF	85(11/12) **#28**	NF	NF
HMDB00172	L-Isoleucine	83(25/29) **#23**	NF	83(25/29) **#29**	NF	NF
HMDB001311	L-Lactic Acid	67(2/2) #308	33(1/2) **#46**	92(2/2) **#20**	NF	NF
HMDB00162	L-Proline	88(30/33) **#6**	NF	91(31/33) **#3**	NF	NF
HMDB01259	Succinic Acid	NF	NF	50(1/1) **#571**	50(1/1) **#51**	[*1]51/51 **#2**
HMDB00251	Taurine	86(6/6) #15	86(6/6) **#5**	86(6/6) #23	86(6/6) #7	[*2]6/6 #**35**

*Similar metabolite match found: 1. Arginosuccinic acid; 2. Hypotaurine

Table 3. Experimental mixtures of 5 metabolites measured with a 200MHz spectrometer. The list of metabolites and their concentrations as well as the results from searches using

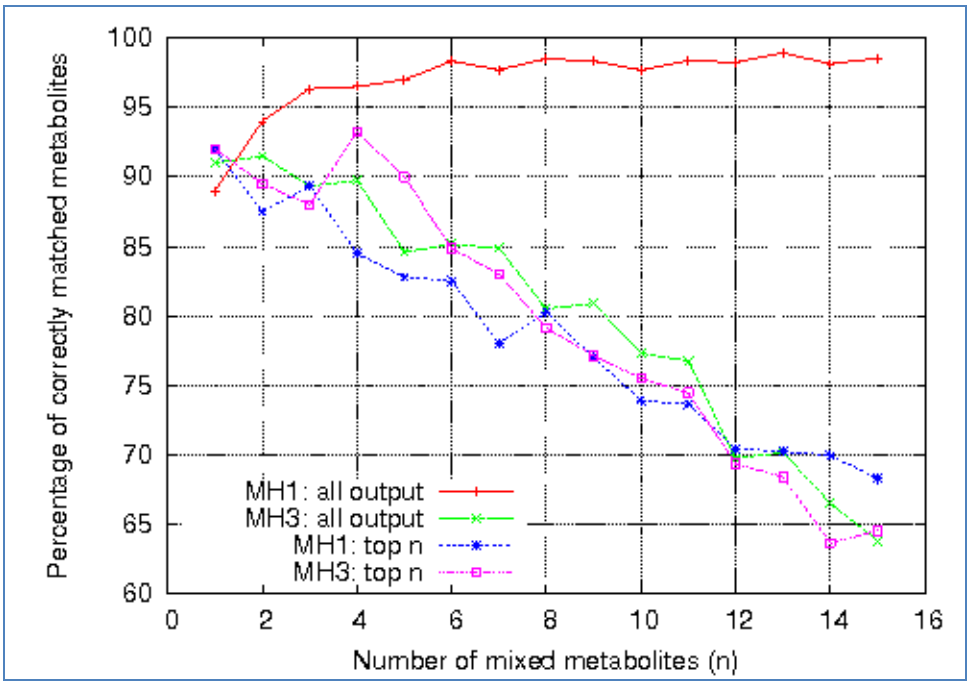
MetaboHunter (MH1: Highest number of matched peaks method; MH3: Greedy selection of metabolites method) and HMDB are provided.

Metabolite	**mM**	**Spectral Search**		**Peak Search**		**HMDB**
		MH1	**MH3**	**MH1**	**MH3**	
Creatine	**4.04179**	* 100(4/3) **#1**	* 100(4/3) **#1**	* 100(3/3) **#1**	* 100(4/3) **#1**	2/2 **#3**
Glucose	**5.2397**	95(37/38) **#4**	NF	95(10/38) **#17**	NF	39/54 **#18**
Citrate	**1.06704**	80(4/4) **#55**	80(4/4) **#3**	20(1/4) **#50**	NF	3/4 **#16**
Phosphocholine	**0.20411**	75(3/3) **#68**	NF	NF	NF	NF
Acetylcholine	**0.10789**	67(4/5) **#108**	NF	17(1/5) **#76**	NF	5/5 **#10**

*Phosphocreatine was assigned as the first match. C**H**$_3$ and C**H**$_2$ peaks observable in Phosphocreatine are identical as the corresponding peaks in Creatine

A test of the search engines employed in MetaboHunter for the analysis of NMR peak information was performed also by automatic search of mixtures of peak lists for n=1:15 metabolites. The accuracy of this search, *i.e*. its ability to return metabolites included in the mixtures, is automatically determined from investigating whether the metabolites are included: a) in the complete list and b) in the top n metabolites listed (Figure 1). When the complete lists of proposed assignments are analyzed over 97% of all metabolites in the mixture are found regardless of the number of molecules in the mixture. If only the top scored assignments are included the accuracy reduces with the increase of number of molecules in the mixture. In this case over 90% of the time the top metabolite corresponds to the peak list for n=1 and the accuracy goes down to ~70% for analysis of n=15 case.

Figure 1. Results of metabolite searches for mixtures of HMDB peaks corresponding to *n* metabolites, where n=1:15. The curves labeled “MHX: all output“ represent average percentages of correctly matched metabolites over 100 runs, when all matches regardless of their position are considered, whereas the curves labeled “MHX: top *n*” represent average percentages of correctly matched metabolites over 100 runs, when only top *n* matches were selected.



MetaboHunter provides an excellent tool for spectral assignment from 1D NMR data with several different methods provided. In the outlined examples MetaboHunter search methods found metabolites included in the spectra with good accuracy. In many cases metabolites in question were obtained with the top, highest score, in the list, however, in other cases they were still listed but with lower scores. Due to the large peak overlaps between different metabolites this can be expected and can only be dealt with either by doing multidimensional NMR or by utilizing a spectrum modeling from a more accurate and more focused metabolite database.

References:

Duarte,I.F. et al. (2009) Analytical approaches toward successful human cell metabolome studies by NMR spectroscopy. *Anal Chem* **81**, 5023-5032.

Gottschalk,M. et al. (2008) Metabolomic studies of human lung carcinoma cell lines using in vitro 1H NMR of whole cells and cellular extracts. *NMR Biomed* **21**, 809-819.

Griffin,J.L. et al. (2002) Spectral profiles of cultured neuronal and glial cells derived from HRMAS 1H NMR spectroscopy. *NMR Biomed* **15**, 375-384.

Tiziani,S. et al. (2009) Metabolic profiling of drug response in acute myeloid leukemia cell lines. *Plos One* **4**, e4251.

Wishart,D.S. et al. (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37***(Database issue), D603-610.*

Yang,C. et al. (2007) Comparative metabolomics of breast cancer. *Pacif Symp Biocomp* **12**, 181-192.