Supplementary information

Applying modern –Omic technologies to the NCLs

Rachel A Kline[3,4], Thomas M Wishart [3,4], Kevin Mills[1,2], Wendy E Heywood[1,2*&]

[3]Roslin Institute and Royal (Dick) Veterinary studies, University of Edinburgh, Edinburgh, UK

[4] The Euan MacDonald Centre for Motor Neuron Disease Research, University of Edinburgh, Edinburgh, UK

*All authors contributed equally

[&] Corresponding Author

Methods for NCL common pathway analysis

For this analysis, we selected proteomic datasets over other –omics methods due to their relative breadth and availability (outlined in Table 1) compared to transcriptomic or metabolomic studies. Nevertheless, an expanded approach bridging the conserved dysregulation we report at the protein level to other elements of NCL cell biology would be ideal, and should be considered in light of the relative novelty of the application of –omics technologies to these disorders.

Raw data files were downloaded from their respective sources (Table 1) and converted to fold-change values. Relative expression ratios between disease model and wildtype as represented by fold-change compared to wildtype were extracted for further analysis.

In order to further filter our list, we then applied a 20% (1.2-fold) cut-off threshold in order to determine which proteins were changed by a magnitude which we have determined previously to harbour the greatest biological relevance to disease in vivo ([1] [2]. Proteins exhibiting subthreshold changes in one or more dataset, or those exhibiting ambiguous changes between studies (ie. increased in disease compared to wildtype in one study, and decreased in disease compared to wildtype in another study) were eliminated from further analysis.

Next, it was of interest to see how comprehensively we could detect conserved changes at the single-protein level between independent studies. For this reason, unique identifications were sorted by whether they had been successfully identified, changed >20% with the same directionality, and whether they had been detected in 1, 2, or 3 or more unique studies. A sub-list of proteins exhibiting a conserved directional change in expression in disease versus wildtype status, by >20%, in 3 or more unique studies was isolated. A list of top 30 proteins increased across 3 or more studies (Table S1) or decreased across 3 or more conditions (Table S2) are available in supplementary materials. Protein changes are ranked by magnitude of change as calculated by the mean of relative expression ratio across the >3 studies in which they were detected.

At this point, it was clear that conserved changes were detectable at the individual protein level between different proteomic studies of NCL models, regardless of tissue type or model characteristics. It was therefore of interest to determine whether these conserved changes could then be contextualised into higher-order networks via a pathway-analysis based approach.

For the following pathway analysis, we sought to align our input sources as much as possible in order to reduce the possibility of falsely identifying “alterations” that may be attributed to differential expression between tissues or timepoints, or indeed simple differences in molecular anatomy between regions of the same tissue [3, 4]. For this reason, the following pathway analyses are derived from comparisons between datasets of whole (brain) tissue of presymptomatic mouse models of CLN1, CLN2 and CLN3 disease, with the exception of the PPT1 interactome characterised by Scifo et al., 2015[5], which lacks associated expression profile values and therefore does not impact the following reported directionality-based scores.

IPA: Canonical Pathways: For details of canonical pathway analysis, please refer to Figure 1A&1B as well as an example of previous application by Llavero et al., 2017.

IPA: Upstream Regulators: For details of upstream regulatory analysis, please refer to Figure 1C and Table 2 as well as an example of previous application by Llavero et al., 2017.

Supplementary Table 1.

Top 30 proteins increased in expression by >20% in disease model compared to wildtype control across three or more unique datasets derived from existing studies of *in vivo* CLN1, CLN2 and CLN3 disease models. All values are represented as fold change in disease model with respect to wildtype (control) protein expression. For filtering and ranking criteria, refer to in-text methods.

Gene	Gene Name	Sleat CSF_CLN1 Sleat et al., 2017	Sleat CSF_CLN2 Sleat et al., 2017	Sleat CSF_CLN3 Sleat et al., 2017	Sleat Brain_CLN1 Sleat et al., 2017	Sleat Brain_CLN2 Sleat et al., 2017	Sleat Brain_CLN3 Sleat et al., 2017	Segal-Salto Brain_CLN1 Palmitolyated Segal-Salto et al., 2017	Llavero Brain_CLN3 Llavero et al., 2017	MEAN	Number of values
ALDH1A2	aldehyde dehydrogenase 1 family member A2	5.58	3.78	4.8						4.72	3
CALB1	calbindin 1		6.55	4.09		1.3			1.22	3.98	4
ACOT1	acyl-CoA thioesterase 1		5.64	2.68	1.46					3.26	3
PSAP	prosaposin				5.23	1.23	1.93		1.4	2.8	4
S100A11	S100 calcium binding protein A11	3.65		2.56	1.9					2.7	3
PEA15	proliferation and apoptosis adaptor protein 15		3.97			1.25			1.13	2.61	3
F13A1	coagulation factor XIII A chain	3.47	2.22	2.81	1.86					2.59	4
PGAM2	phosphoglycerate mutase 2		4.51		2.79	1.33	1.54			2.54	3
FLNC	filamin C			3.42	2.26	1.62				2.43	3
LCP1	lymphocyte cytosolic protein 1	3.29	2.28	1.94	2.21					2.43	4
ALDH1A1	aldehyde dehydrogenase 1 family member A1		3.9	1.72	1.44					2.35	3
CD44	CD44 molecule (Indian blood group)				3.72	1.7	1.57			2.33	3
PNP	purine nucleoside phosphorylase	2.24	3.1	2.19	1.61					2.29	4
DDAH2	dimethylarginine dimethylaminohydrolase 2	2.17	2.89	1.58	1.59					2.06	4
TNC	tenascin C				2.12	2.2	1.77			2.03	3
DPP7	dipeptidyl peptidase 7		2.06		2.09		1.72			1.96	3
S100A6	S100 calcium binding protein A6				3.23	1.25	1.37			1.95	3
ACYP2	acylphosphatase 2	1.78	2.52		1.46					1.92	3
MAPK1	mitogen-activated protein kinase 1		2.42		1.23				1.67	1.83	3
FHL1	four and a half LIM domains 1		2.49	1.53	1.39					1.8	3
VCL	vinculin	1.8		2.06	1.53					1.8	3
CMBL	carboxymethylenebutenolidase homolog	1.78	2.11	1.48						1.79	3
SELENBP1	selenium binding protein 1	1.91	2.41	1.56	1.84		1.21			1.79	5
PGLS	6-phosphogluconolactonase	1.78	2.07		1.47					1.77	3
HSPA1A/H	heat shock protein family A (Hsp70) member 1A	1.75	2.09		1.44					1.76	3
HEXB	hexosaminidase subunit beta		2.12		1.34		1.67			1.71	3
SH3BGRL3	SH3 domain binding glutamate rich protein like	2.18			1.45		1.27			1.63	3
HSPB1	heat shock protein family B (small) member 1				2.05	1.34	1.28			1.56	3
GALK1	galactokinase 1				1.57	1.41	1.26			1.41	3
FLNB	filamin B	1.41		1.33	1.46					1.4	3
PHGDH	phosphoglycerate dehydrogenase				1.54	1.26	1.29			1.36	3
WDR1	WD repeat domain 1				1.32	1.21		15.8	1.38	1.27	4

Supplementary Table 2.

Top 30 proteins decreased in expression by >20% in disease model compared to wildtype control across three or more unique datasets derived from existing studies of *in vivo* CLN1, CLN2 and CLN3 disease models. All values are represented as fold change in disease model with respect to wildtype (control) protein expression. For filtering and ranking criteria, refer to in-text methods.

Gene	Gene Name	Sleat CSF_CLN1	Sleat CSF_CLN2	Sleat CSF_CLN3	Sleat Brain_CLN1	Sleat Brain_CLN2	Sleat Brain_CLN3	Segal-Salto Brain_CLN1 Palmitolyated	Llavero Brain_CLN3	MEAN	Number of values
		Sleat et al., 2017	Sleat et al., 2017	Sleat et al., 2017	Sleat et al., 2017	Sleat et al., 2017	Sleat et al., 2017	Segal-Salto et al., 2017	Llavero et al., 2017		
SYT2	synaptotagmin 2				-4.35	-5.56			-1.3	-3.74	3
NECAB1	N-terminal EF-hand calcium binding protein 1				-3.57	-2.17				-2.87	2
HAPLN2	hyaluronan and proteoglycan link protein 2	-3.33		-3.13	-1.92	-1.82			-2.34	-2.51	5
HSP90AB1	heat shock protein 90 alpha family class B member 1				-1.28			-4.88	-1.23	-2.46	3
CLTC	clathrin heavy chain				-1.59	-1.32	-1.21	-5.52		-2.41	4
ALDH3A1	aldehyde dehydrogenase 3 family member A1	-2.78	-2.08	-2.08						-2.31	3
GSK3B	glycogen synthase kinase 3 beta				-1.54	-1.28		-4.11		-2.31	3
APOL2	apolipoprotein L2				-2.04	-2.56	-2.22			-2.28	3
KLK6	kallikrein related peptidase 6	-2.04	-2.22	-2.56						-2.28	3
SNCA	synuclein alpha				-1.72	-1.64	-1.39		-4.13	-2.22	4
PRRT2	proline rich transmembrane protein 2				-2.78	-1.92	-1.92			-2.21	3
GAD2	glutamate decarboxylase 2				-2.7	-2.7			-1.22	-2.21	3
SYNPO	synaptopodin				-2.33	-2.33	-1.75			-2.14	3
STX1B	syntaxin 1B				-3.13	-2.17	-2		-1.23	-2.13	4
NDUFB8	NADH:ubiquinone oxidoreductase subunit B8				-3.33	-1.7			-1.29	-2.11	3
SYNGR1	synaptogyrin 1				-3.03	-1.7	-1.52			-2.08	3
FABP4	fatty acid binding protein 4	-2.33	-2.17	-1.96		-1.72				-2.05	4
CDIP1	cell death inducing p53 target 1				-2.22	-2.13	-1.79			-2.05	3
ACTR1B	ARP1 actin related protein 1 homolog B				-1.39	-1.3	-1.25	-4.19		-2.03	4
CLSTN1	calsyntenin 1		-2.33	-2.27	-1.47					-2.02	3
ACAT1	acetyl-CoA acetyltransferase 1				-1.21	-1.27		-4.19	-1.21	-1.97	4
APP	amyloid beta precursor protein		-2.08	-1.96	-1.59					-1.88	3
CABP1	calcium binding protein 1				-2.17	-1.67	-1.7			-1.85	3
PRDX5	peroxiredoxin 5	-2.94				-1.33			-1.22	-1.83	3
ABLIM2	actin binding LIM protein family member 2				-1.92	-2	-1.56			-1.83	3
RAP2A	RAP2A, member of RAS oncogene family				-2.17	-1.72	-1.59			-1.83	3
ELAVL2	ELAV like RNA binding protein 2				-2.08	-1.82	-1.52			-1.81	3
CNTNAP2	contactin associated protein like 2			-1.85	-2.33	-1.59	-1.45			-1.8	4
HPRT1	hypoxanthine phosphoribosyltransferase 1				-2.04	-1.89	-1.41			-1.78	3

Supplementary Table 3.

Top 5 predicted upstream causal "master regulators" of proteomic changes across CLN1-3 disease states derived from independent analyses of brain in CLN1 (Sleat et al, Tikka et al, Segal-Salto et al.), CLN2 (Sleat et al), and CLN3 (Llavero Hurtado et al., Sleat et al) in vivo models as well as the neuronal PPT1 interactome characterised by Scifo et al. Predicted activation z-score is calculated by weighing the predicted expression change of target molecules as defined by Ingenuity Knowledge Database against the actual expression change of target molecules reported in input dataset(s). Values in columns represent the activation z-scores of specific upstream regulator calculated to each input proteomic dataset. An activation z-score >2 or <-2 is considered statistically significant. P-value of overlap is derived from a Fisher's exact test are derived from a Fisher's Exact Test calculating overlap between molecules in each respective input dataset and number of molecules comprising a canonical pathway defined by Ingenuity Systems Database (in this case, known downstream interactors of RICTOR). As the Scifo et al. dataset represents an interactome characterization of CLN1/PPT1, no expression profile is available and therefore no activation z-score calculation is possible.

Upstream Regulators	Tikka et al Presymptomatic Thalamus_CLN1	Tikka et al Endstage Thalamus_CLN1	Sleat Brain_CLN1	Sleat Brain_CLN2	Sleat Brain_CLN3	Segal-Salto Brain_CLN1 Palmitolyated	Scifo et al_ CLN1 Interactome	Llavero Brain_CLN3
	Tikka et al., 2016	Tikka et al., 2016	Sleat et al., 2017	Sleat et al., 2017	Sleat et al., 2017	Segal-Salto et al., 2017	Scifo et al., 2015	Llavero et al., 2017
RICTOR	-2	N/A	6.582	7.057	6.505	N/A	N/A	4.259
mono-(2-ethylhexyl)phthalate	N/A	N/A	-3.738	-5.411	-4.366	N/A	N/A	-3.643
PPARGC1A	N/A	N/A	-4.682	-6.329	-4.27	-1.635	N/A	-2.579
INSR	N/A	N/A	-4.633	-5.391	-4.586	N/A	N/A	-4.034
ST1926	N/A	N/A	3.893	5.048	4.025	0.378	N/A	1.514
5-fluorouracil	N/A	N/A	4.457	4.718	3.819	0.493	N/A	1.919
KDM5A	N/A	N/A	5.602	5.835	5.209	N/A	N/A	2.058
STK11	N/A	N/A	2.594	5.485	4	N/A	N/A	2.168
BDNF	N/A	N/A	-5.583	-2.647	-3.356	-0.632	N/A	1.973
beta-estradiol	N/A	-0.618	-2.255	-3.818	-3.54	-2.315	N/A	-0.862

[1] T.M. Wishart, T.M. Rooney, D.J. Lamont, A.K. Wright, A.J. Morton, M. Jackson, M.R. Freeman, T.H. Gillingwater, Combining comparative proteomics and molecular genetics uncovers regulators of synaptic and axonal stability and degeneration in vivo, PLoS genetics, 8 (2012) e1002936.

[2] M. Llavero Hurtado, H.R. Fuller, A.M.S. Wong, S.L. Eaton, T.H. Gillingwater, G. Pennetta, J.D. Cooper, T.M. Wishart, Proteomic mapping of differentially vulnerable pre-synaptic populations identifies regulators of neuronal stability in vivo, Sci Rep, 7 (2017) 12412.
[3] S.L. Eaton, S.L. Roche, M. Llavero Hurtado, K.J. Oldknow, C. Farquharson, T.H. Gillingwater, T.M. Wishart, Total protein analysis as a reliable loading control for quantitative fluorescent Western blotting, PloS one, 8 (2013) e72457.
[4] E.J.N. Groen, E. Perenthaler, N.L. Courtney, C.Y. Jordan, H.K. Shorrock, D. van der Hoorn, Y.T. Huang, L.M. Murray, G. Viero, T.H. Gillingwater, Temporal and tissue-specific variability of SMN protein levels in mouse models of spinal muscular atrophy, Human molecular genetics, 27 (2018) 2851-2862.
[5] E. Scifo, A. Szwajda, J. Debski, K. Uusi-Rauva, T. Kesti, M. Dadlez, A.C. Gingras, J. Tyynela, M.H. Baumann, A. Jalanko, M. Lalowski, Drafting the CLN3 protein interactome in SH-SY5Y human neuroblastoma cells: a label-free quantitative proteomics approach, Journal of proteome research, 12 (2013) 2101-2115.