mapMECFS Data Formats

Questions: mapmecfs@rti.org

Cytokine Data File Format

Column 1: Cytokine name. Column name must be 'Molecule'. These must be UNIQUE and cannot be blank ('').

2nd column and onwards are the participant IDs. These must be UNIQUE and match the phenotype file.

Molecule	MMC000001	MMC000002	MMC000003	MMC000004	MMC000005	MMC000006
sCD40L	323.7634	358.5415	310.0712	287.3893	310.4637	272.7896
EGF	49.41162	52.74591	53.72772	52.69184	55.8378	50.60854
FGF2	59.65343	66.97424	56.47696	52.93076	64.11999	50.43223
FLT3LG	4.916391	5.421188	5.302512	5.090445	5.06062	5.160972
CX3CL1	59.60498	47.14341	59.57045	51.01518	47.53699	49.6409
CSF3	34.46174	36.51572	35.93429	34.82774	35.61065	37.4167
CSF2	7.263878	6.987126	9.665282	8.22497	6.403682	8.167706
GRO	619.442	633.6399	631.4036	586.3567	602.4217	607.1942

Data starts on line 2

Each column contains measurements from 1 sample. Missing values are allowed ('NA' or leave empty).

Note: The cytokine data file is formatted as a tab-separated file

Gene Expression at the Gene Level
Data File Format

Column 1: Gene Name. Column name must be 'Molecule'. Other gene identifiers are accepted, although only gene symbol can identify synonyms of the identifier. This is REQUIRED. These must be UNIQUE and cannot be blank ('').

2nd column and onwards are the participant IDs. These must be UNIQUE and match the phenotype file.

Molecule	GSM364557	GSM364558	GSM364559	GSM364560	GSM364561	GSM364562	GSM364563
RFC2	6.0963772	6.085353311	6.136004761	6.116115706	6.189968524	6.262979864	6.116189938
HSPA6	7.347443749	7.217139972	7.681249781	7.385276973	7.116387337	7.140547201	7.448098019
PAX8	8.612818125	8.678530368	8.579275409	8.562063105	8.556573863	8.665905532	8.737066487
GUCA1A	4.047615838	3.970998339	4.016488499	3.836390674	3.996242126	3.969002165	4.007806877
NA	9.427775107	9.85612436	9.889555061	9.465435829	9.377903462	9.398643413	9.336165021
THRA	6.189940593	6.230631973	6.234009818	6.102569085	6.17315022	6.314285593	6.325875474
PTPN21	4.522562583	4.641474268	4.788068106	4.449960825	4.453870578	4.567019976	4.716097445
CCL5	10.12644618	10.66701745	10.03099084	10.32407976	10.32033785	9.56852563	10.04493409
CYP2E1	4.216807308	4.18582031	4.128110484	4.145729571	4.183883642	4.205720899	4.187756977
EPHB3	6.525101838	6.529757372	6.525074897	6.457801129	6.578782525	6.417345647	6.621183006
ESRRA	7.692959277	7.850628002	8.05249897	7.767810134	7.81920264	7.791135584	7.805728671
CYP2A6	6.447467842	6.584006236	6.628121155	6.495184149	6.35174817	6.454191542	6.54368911
GAS6	8.950749758	8.947117698	9.329527196	8.903042624	8.886195262	8.876640138	9.025508086

Data starts on line 2

Each column contains measurements from 1 sample. Missing values are allowed ('NA' or leave empty).

Note: The gene expression data file is formatted as a tab-separated file.

Metabolomics Data File Format

Column 1: InChiKey for the metabolite. Column name must be 'InChiKey'. These must be UNIQUE.

Column 2: Metabolite name. Column name must be 'Molecule'. These must be UNIQUE.

Column 3: CHEBI ID for the metabolite. Column name must be 'database_identifier'. These must be UNIQUE

4th column and onwards are the participant IDs. These must be UNIQUE and match the phenotype file.

InChiKey	Molecule	database_identifier	CFS11serum	CFS12serum	CFS13serum
CVSVTCORWBXHQV-UHFFFAOYSA-N	Creatine	CHEBI:16919	8.8	6.2	12.3
FDGQSTZJBFJUBT-UHFFFAOYSA-N	Hypoxanthine	CHEBI:17368	2.2	3.3	2.5
BDAGIHXWWSANSR-UHFFFAOYSA-M	Formate	CHEBI:30751	5.8	6.3	3.2
DDRJAANPRJIHGJ-UHFFFAOYSA-N	Creatinine	CHEBI:16737	18.1	16.2	13
QTBSBXVTEAMEQO-UHFFFAOYSA-M	Acetate	CHEBI:15366	6.3	12.1	4.1
KWIUHFFTVRNATP-UHFFFAOYSA-N	Betaine	CHEBI:17750	10.9	16.2	30.4
AGPKZVBTJJNPAG-WHFBIAKZSA-N	L-Isoleucine	CHEBI:17191	32.8	20.4	15.7
COLNVLDHVKWLRT-QMMMGPOBSA-N	L-Phenylalanine	CHEBI:17295	19	16.4	13.3
QIVBCDIJIAJPQS-VIFPVBQESA-N	L-Tryptophan	CHEBI:16828	14.1	13.8	11.5
OUYCCCASQSFEME-QMMMGPOBS…	L-Tyrosine	CHEBI:17895	22.6	17	12.9
CKLJMWTZIZZHCS-REOHCLBHSA-N	L-Aspartate	CHEBI:17053	16.6	15.9	14.3

Data starts on line 2

Each column contains measurements from 1 sample. Missing values are allowed ('NA' or leave empty).

Note: The metabolomics data file is formatted as a tab-separated file.

Methylation Data File Format

Column 1: CpG ID. Column name must be 'Molecule'. This is REQUIRED. These must be UNIQUE

Further columns are the participant IDs. These must be UNIQUE and correspond to match the phenotype file.

Molecule	MMC000001	MMC000002	MMC000003	MMC000004
cg27409831	0	0	0	0
cg04127068	0.14	0.0952380952380952	0.0418410041841004	0
cg01801315	1	1	1	1
cg00996420	0.136363636363636	0.0645161290322581	0.025	0.166666666666667
cg10509260	0.125	0.444444444444444	0.0213523131672598	0.25
cg22027930	0.5	0.6	0.823529411764706	0.841772151898734
cg26224280	0.647058823529412	0.584033613445378	0.4375	0.566666666666667
cg10186410	0.861111111111111	0.39344262295082	0.527777777777778	0.4
cg14313350	0.6	0.321428571428571	0.1	0.5
cg20274783	0.794871794871795	0.54356846473029	0.358490566037736	0.94392523364486
cg17705549	0.0576923076923077	0	0	0
cg15168383	0.805555555555556	0.8	0.821138211382114	0.7

Data starts on line 2

Each column contains measurements from 1 sample. Missing values are allowed ('NA' or leave empty).

Note: The methylation data file is formatted as a tab-separated file.

miRNA Data File Format

Column 1: miRBase ID for the miRNA. Column name must be 'Molecule'. This is REQUIRED. These must be UNIQUE and cannot be blank ('')

2nd column and onwards headings are the participant IDs. These must be UNIQUE and match the phenotype file.

Molecule	MMC000002	MMC000003	MMC000004	MMC000005	MMC000006
hsa-let-7a-3p	1.935975	1.235798	3.32162	2.390687	3.555187
hsa-let-7a-5p	16112.81177	16109.30814	18.62037	16716.45319	3355.52361
hsa-let-7b-3p	NA	NA	NA	NA	NA
hsa-let-7b-5p	3208.833	1835.4028	129.2342	7650.9067	3144.8213
hsa-let-7c	2558.24485	483.09237	2166.13565	2438.23284	1004.3033
hsa-let-7d-3p	2.599237	0.7952846	1.4535083	4.08737	3.8224884
hsa-let-7d-5p	193.165056	2376.767968	1462.675032	2471.328113	2390.837382
hsa-let-7e-3p	NA	NA	NA	NA	NA
hsa-let-7e-5p	9.8327	164.67912	416.1686	238.87236	420.96276
hsa-let-7f-1-3p	NA	NA	NA	NA	NA
hsa-let-7f-2-3p	NA	NA	NA	NA	NA
hsa-let-7f-5p	1192.97879	6041.61075	4085.86993	3104.54578	1151.99194

Data starts on line 2

Each column contains measurements from 1 sample. Missing values are allowed ('NA' or leave empty).

Note: The miRNA data file is formatted as a tab-separated file.

Phenotype File Format

Column 1: ID for each participant. Column header must be 'ParticipantID'. Column is REQUIRED. These must be UNIQUE and match the data file.

Column 2: Phenotype of interest. Column header must be 'Phenotype'. Column is REQUIRED.

Column 3: The source or tissue the sample was extracted from. Column header must be 'Sample_Source'. Column is required. Missing is allowed ('NA' or empty string).

Columns are optional. You can substitute with any other information.

ParticipantID	Phenotype	Sample_Source	characteristics_ch1.0.twin pair	characteristics_ch1.1.sex
GSM402241	unaffected	PBLs	228340	female
GSM402242	CFS	PBLs	228340	female
GSM402243	unaffected	PBLs	220263	female
GSM402244	CFS	PBLs	220263	female
GSM402245	unaffected	PBLs	235495	female
GSM402246	ICF	PBLs	235495	female
GSM402247	unaffected	PBLs	227565	male
GSM402248	ICF	PBLs	227565	male
GSM402249	CFS	PBLs	232496	female
GSM402250	unaffected	PBLs	232496	female
GSM402251	CFS	PBLs	230813	female

Data starts on line 2

Each column contains data for 1 participant. Missing values are allowed ('NA' or leave empty).

Note: The phenotype file is formatted as a tab-separated file.