MUGA Example Data

Daniel M. Gatti

11 October 2013

1 Introduction

The data in this package contains phenotype and genotype data from Diversity Outbred (DO) mice. The Mouse Universal Genotyping Array (MUGA) that was developed by the University of North Carolina at Chapel Hill [1]. The array contains 7,864 markers and was developed to genotype Collaborative Cross and Diversity Outbred mice. It may also be used to genotype other multi-founder mouse crosses.

The data in this package is from Svenson et.al, Genetics, 2012 [2]. Briefly, 150 mice (75 F and 75 M) were placed on either a chow or high fat diet at wean. They were phenotyped at early and late time points and sacrificed by 30 weeks of age. Tail tips were taken, DNA was extracted and the mice were genotypes on the MUGA.

2 Data Description

There are 10 data files available in this package. Any of them may be accessed using data(<name>).

```
> library(MUGAExampleData)
```

FinalReport1 Raw genotype file for the samples in this study.

FinalReport2 Raw genotype file for the samples in this study.

Samples1 Sample IDs for the samples in FinalReport1.

Samples2 Sample IDs for the samples in FinalReport2.

call.rate.batch The allele call rates and batch IDs for each sample.

x The X allele intensities extracted from FinalReport1 and FinalReport2.

y The Y allele intensities extracted from FinalReport1 and FinalReport2.

geno The allele calls extracted from FinalReport1 and FinalReport2.

model.probs The DO founder haplotype probabilities for each sample at each marker.

pheno The phenotype data for this study.

2.1 FinalReport1 and 2

These are raw text files containing the MUGA genotyping output as it is received from GeneSeek. There are nine rows of header information. Each line is tab delimited and contains the following columns.

SNP Name MUGA SNP ID.

Sample ID Sample name. Will match the sample name in Samples1 or 2.

Allele1 - Forward Allele call for probe 1.

Allele2 - Forward Allele call for probe 2.

X Normalized X allele intensity

Y Normalized Y allele intensity

GC Score GC Score

Theta X and Y intensites converted to θ coordinate.

X Raw Untransformed X allele intensity.

Y Raw Untransformed Y allele intensity.

R X and Y intensites converted to ρ coordinate.

2.2 Samples1 and 2

These are raw text files containing the sample names from the MUGA genotyping from GeneSeek. Each line is tab delimited and contains the following columns.

Index Sample index

Name Sample Name

ID Sample ID (may be the same as the name).

Gender Sample sex.

Plate Plate ID on which sample was run.

Well Well in which sample was run.

Group Sample group.

Parent1 Parent1.

Parent2 Parent2.

Replicate Replicate ID

SentrixPosition Sample position code.

2.3 call.rate.batch

Data.frame containing allele call rate and batch information for each sample.

sample Sample ID.

call.rate The proportion of successful allele calls.

batch A batch identifier to distinguish batch 1 and 2.

2.4 x and y

These are numeric matrices that contain the X and Y allele intensity data extracted from the FinalReport1 and FinalReport2 files. The dimensions are 141 samples by 7,854 markers. Although there were 150 samples in the study, only 141 were genotyped for technical reasons.

The rows are names by sample ID and the columns are named by the SNP ID.

2.5 geno

This is a character matrix that contains the allele calls extracted from the FinalReport1 and FinalReport2 files. The dimensions are 141 samples by 7,854 markers. Each cell contains either 'A', 'C', 'G',' H', 'T' or '-'. Although there were 150 samples in the study, only 141 were genotyped for technical reasons.

The rows are names by sample ID and the columns are named by the SNP ID.

2.6 model.probs

This is a numeric three dimensional array containing the founder haplotype contributions for each sample at each marker. The dimensions are 141 samples by 8 founders by 7,854 markers. Cell (i, j, k) contains the proportion of j at locus k for sample i. The founders are labeled A through H and are explained below.

A	A/J
B	C57BL/6J
C	129S1/SvImJ
D	NOD/ShiLtJ
E	NZO/H1LtJ
F	CAST/EiJ
G	PWK/PhJ
H	WSB/EiJ

```
> data(model.probs)
> model.probs[1,,1:5]

  UNC010515443 UNC010001943 UNC010515539 UNC010515556 UNC010002207
A 9.598007e-51 3.220165e-16 1.380275e-17 1.458903e-17 1.640058e-16
B 8.411587e-09 8.419524e-46 1.083905e-19 8.248867e-39 4.575537e-17
C 1.877777e-08 9.333647e-09 5.353990e-12 1.109631e-22 4.553043e-17
D 5.000000e-01 5.000000e-01 5.000000e-01 5.000000e-01 5.000000e-01
E 5.000000e-01 5.000000e-01 5.000000e-01 5.000000e-01 5.000000e-01
F 7.889177e-40 7.543572e-22 1.074099e-17 1.087953e-17 9.531830e-17
G 4.537672e-39 6.875965e-21 1.950897e-16 1.948297e-16 2.054865e-16
H 1.054247e-08 3.493217e-12 1.563549e-15 9.469679e-17 1.257905e-17
```

2.7 pheno

This is the phenotype data for the samples in this study. There are 149 samples in rows and 142 columns. The first six columns contain sample information and the remaining columns contain phenotype measurements.

```
> data(pheno)
> pheno[1:5,1:6]
```

```
    Sample Sex  Gen Diet Coat.Color WBC1
F01    F01   F G4L2   hf     agouti 4.95
F02    F02   F G4L2   hf      black 5.59
F03    F03   F G4L2   hf      white 6.54
F04    F04   F G4L2   hf     agouti 6.02
F05    F05   F G4L2   hf     agouti 5.64
```

Column Name	Description	Timepoint (weeks)
Sample	Sample ID	NA
Sex	Sample Sex	NA
Gen	DO outbreeding generation and litter	NA
Diet	Diet, either chow or hf	NA
Coat.Color	Coat color coded as agouti, black or white	NA
WBC1	White Blood Cell counts (1000 cells / μl)	10
RBC1	Red Blood Cell counts (1000 cells / μl)	10
mHGB1	Measured Hemoglobin	10
HCT1	Hematocrit	10
MCV1	Mean Corpuscular Volume	10
MCH1	Mean Corpuscular Hemoglobin	10
MCHC1	Mean Corpuscular Hemoglobin Concentration	10
CHCM1	Corpuscular Hemoglobin Concentration Mean	10
RDW1	Red blood cell distribution width	10
HDW1	Hemoglobin distribution width	10
PLT1	Platelet counts	10
MPV1	Mean platelet volume	10
perc.NEUT1	Percent neutrophils	10
perc.LYM1	Percent lymphocutes	10
perc.MONO1	Percent monocytes	10
perc.EOS1	Percetn Eosinophils	10
Retic1	Reticulocyte counts	10
cHGB1	Calulcated hemoglobin	10
ct.NEUT1	Neutrophil counts	10
ct.LYM1	Lymphocyte counts	10
ct.MONO1	Monocyte counts	10
ct.EOS1	Eosinophil counts	10
WBC2	White Blood Cell counts (1000 cells / μl)	22
RBC2	Red Blood Cell counts (1000 cells / μl)	22
mHGB2	Measured Hemoglobin	22
HCT2	Hematocrit	22
MCV2	Mean Corpuscular Volume	22
MCH2	Mean Corpuscular Hemoglobin	22
MCHC2	Mean Corpuscular Hemoglobin Concentration	22
CHCM2	Corpuscular Hemoglobin Concentration Mean	22
RDW2	Red blood cell distribution width	22
HDW2	Hemoglobin distribution width	22
PLT2	Platelet counts	22
MPV2	Mean platelet volume	22
perc.NEUT2	Percent neutrophils	22
perc.LYM2	Percent lymphocutes	22
perc.MONO2	Percent monocytes	22
perc.EOS2	Percetn Eosinophils	22
Retic2	Reticulocyte counts	22
cHGB2	Calulcated hemoglobin	22
ct.NEUT2	Neutrophil counts	22
ct.LYM2	Lymphocyte counts	22
ct.MONO2	Monocyte counts	22
ct.EOS2	Eosinophil counts	22
HR	Heart rate (beats/min)	13
HRV	Heart rate variability	13
PQ	P to Q wave time	13
PR	P to R wave time	13
QRS	Q, R S wave time	13
QTC	Q to T wave time, corrected	13
RR	RR wave	13
ST	S to T wave time	13

Column Name	Description	Timepoint (weeks)
QTc.dispersion	Q to T, corrected dispersion	13
pNN50...6ms.	Mean number of time that teh NN signal exceeds 6 ms	13
rMSSD	root mean squared standard deviation	13
CHOL1	Total serum cholesterol	8
TG1	Serum triglycerides	8
HDLD1	Serum high density lipoprotein	8
NEFA1	Serum non-esterified fatty acids	8
Lipase1	Serum lipase	8
Glucose1	Serum glucose	8
Phosphorus1	Serum phosphorus	8
Calcium1	Serum calcium	8
GLDH1	Serum glutamate dehydrogenase	8
BUN1	Blood urea nitrogen	8
FRUC1	Serum fructose	8
CHOL2	Total serum cholesterol	19
TG2	Serum triglycerides	19
HDLD2	Serum high density lipoprotein	19
NEFA2	Serum non-esterified fatty acids	19
Lipase2	Serum lipase	19
Glucose2	Serum glucose	19
Phosphorus2	Serum phosphorus	19
Calcium2	Serum calcium	19
GLDH2	Serum glutamate dehydrogenase	19
BUN2	Blood urea nitrogen	19
FRUC2	Serum fructose	19
non.fast.Phosphorous	Non-fasted serum phosphorus	
non.fast.Calcium	Non-fasted serum calcium	
non.fast.ALB2	Non-fasted serum albumin	
non.fast.CREX	Non-fasted serum creatinine	
Subject.Length1	Length (cm)	12
Weight1	Weight (g)	12
BMD1	Bone Mineral Density	12
BMC1	Bone Minearal Content	12
B.Area1	Bone Area	12
T.Area1	Total Area	12
X..Fat1	Percent fat	12
TTM1	Total tissue mass (g)	12
LTM1	Lean tissue mass (g)	12
Subject.Length2	Length (cm)	21
Weight2	Weight (g)	21
BMD2	Bone Mineral Density	21
BMC2	Bone Minearal Content	21
B.Area2	Bone Area	21
T.Area2	Total Area	21
X..Fat2	Percent fat	21
TTM2	Total tissue mass (g)	21
LTM2	Lean tissue mass (g)	21
urine.microalbumin1	Urine microalbumin	
urine.Glucose1	Urine glucose	
urine.creatinine1	Urine creatinine	
urine.microalbumin2	Urine microalbumin	
urine.Glucose2	Urine glucose	
urine.creatinine2	Urine creatinine	

Column Name	Description	Timepoint (weeks)
heart.wt	Heart weight (g)	24 - 30
spleen.wt	Spleen weight (g)	24 - 30
kidney.wt.L	Left kidney weight (g)	24 - 30
kidney.wt.R	Right kidney weight (g)	24 - 30
BW.3	Body weight (g)	3
BW.4	Body weight (g)	4
BW.5	Body weight (g)	5
BW.6	Body weight (g)	6
BW.7	Body weight (g)	7
BW.8	Body weight (g)	8
BW.9	Body weight (g)	9
BW.10	Body weight (g)	10
BW.11	Body weight (g)	11
BW.12	Body weight (g)	12
BW.13	Body weight (g)	13
BW.14	Body weight (g)	14
BW.15	Body weight (g)	15
BW.16	Body weight (g)	16
BW.17	Body weight (g)	17
BW.18	Body weight (g)	18
BW.19	Body weight (g)	19
BW.20	Body weight (g)	20
BW.21	Body weight (g)	21
BW.22	Body weight (g)	22
BW.23	Body weight (g)	23
BW.24	Body weight (g)	24
BW.25	Body weight (g)	25
BW.26	Body weight (g)	26
BW.27	Body weight (g)	27
BW.28	Body weight (g)	28
BW.29	Body weight (g)	29
BW.30	Body weight (g)	30
INSULIN	Body weight (g)	17
LEPTIN	Body weight (g)	17

References

[1] F. A. Iraqi, M. Mahajne, Y. Salaymah, H. Sandovski, H. Tayem, K. Vered, L. Balmer, M. Hall, G. Manship, G. Morahan, K. Pettit, J. Scholten, K. Tweedie, A. Wallace, L. Weerasekera, J. Cleak, C. Durrant, L. Goodstadt, R. Mott, B. Yalcin, C. Hill, D. L. Aylor, R. S. Baric, T. A. Bell, K. M. Bendt, J. Brennan, J. D. Brooks, R. J. Buus, J. J. Crowley, J. D. Calaway, M. E. Calaway, A. Cholka, D. B. Darr, J. P. Didion, A. Dorman, E. T. Everett, M. T. Ferris, W. F. Mathes, C. P. Fu, T. J. Gooch, S. G. Goodson, L. E. Gralinski, S. D. Hansen, M. T. Heise, J. Hoel, K. Hua, M. C. Kapita, S. Lee, A. B. Lenarcic, E. Y. Liu, H. Liu, L. McMillan, T. R. Magnuson, K. F. Manly, D. R. Miller, D. A. O'Brien, F. Odet, I. K. Pakatci, W. Pan, F. P. de Villena, C. M. Perou, D. Pomp, C. R. Quackenbush, N. N. Robinson, N. E. Sharpless, G. D. Shaw, J. S. Spence, P. F. Sullivan, W. Sun, L. M. Tarantino, W. Valdar, J. Wang, W. Wang, C. E. Welsh, A. Whitmore, T. Wiltshire, F. A. Wright, Y. Xie, Z. Yun, V. Zhabotynsky, Z. Zhang, F. Zou, C. Powell, J. Steigerwalt, D. W. Threadgill, E. J. Chesler, and

A. Gary. The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics*, 190(2):389–401, Feb 2012.

[2] K. L. Svenson, D. M. Gatti, W. Valdar, C. E. Welsh, R. Cheng, E. J. Chesler, A. A. Palmer, L. McMillan, and G. A. Churchill. High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics*, 190(2):437–447, Feb 2012.