Program No. LB195 Phenotype and genotype comparison for C57BL/6N substrains contributing to the International Mouse Phenotyping Consortium (IMPC)

Bruce Elder[1], Eric Engelhard[2], Ann Flenniken[3], Yann Herault[4], Colin McKerlie[5], Lauryl Nutter[5], Tania Sorg-Guss[4]
[1]Charles River Laboratories, Wilmington, MA, [2]UC Davis Mouse Biology Program, Davis, CA, [3]Toronto Centre for Phenogenomics, Centre for Modeling Human Disease, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, [4]ICS-MCI, Strasbourg, France, [5]The Hospital for Sick Children at the Toronto Centre for Phenogenomics, Toronto, ON, Canada

charles river
accelerating drug development. exactly.

Experimental Biology 2013, Boston, MA
04/2013
www.criver.com

Abstract

The International Mouse Phenotyping Consortium (IMPC) is a 10-year program to produce a null mutant mouse line for every gene in the mouse genome, generate comprehensive phenotype data for each, and provide all the resources to the scientific community. The chosen mouse strain is the C57BL/6N, however there are several substrains of this mouse including C57BL/6NCrl (NCrl), C57BL/6NTac (NTac) and C57BL/6NJ (NJ). A C57BL/6N-derived embryonic stem (ES) cell line (JM8) was used to create the gene-targeted clones. Given the scope of the IMPC project, it is important to catalog phenotype and genotype similarities or differences between the C57BL/6N substrains and ES cells used by the various IMPC centers. Here we describe results from genotype analyses using the Affymetrix Mouse Diversity Genotyping array to compare single-nucleotide polymorphisms (SNP) between NCrl, NTac mice and JM8 and C2 ES cell lines. In addition, whole genome sequencing (WGS) was carried out on NCrl, NTac, NJ and JM8 genomic DNA. Phenotype analyses using the IMPC pipeline was conducted on NCrl and NTac mice at two separate locations; The Toronto Centre for Phenogenomics (TCP) and the Institut Clinique de la Souris (ICS). Substrain dfferences in certain traits were detected and the number of differences was strongly influenced by the rearing environment of the cohort, with fewer differences between NCrl and NTac cohorts derived from animals born and raised within the same institution.

Materials and Methods

Phenotype Analysis
Phenotyping was carried out at two locations; TCP and ICS.
TCP cohort was 10 males (M) and 10 females (F), each for NTac and NCrl shipped in from the vendor barriers, acclimated and tested.
ICS cohorts consisted of 4 batches.
Batches1 and 2: 20 M + 20 females from ICS internal NTac colony and the same number of NCrl shipped in from the vendor.
Batches 3 and 4: 20 M + 20 F each for NTac and NCrl animals from internal ICS colonies of each strain.
Both TCP and ICS used the IMPC phenotyping pipeline and protocols (www.mousephenotype.org)

Genotype Analysis
Affymetrix analysis: Liver genomic DNA was isolated from six NCrl and four NTac mice. In addition, ES cell genomic DNA was isolated from JM8 and C2 cell lines. All DNA samples were analyzed using standard protocols for hybridization and scanning of the Affymetrix MDG array, and SNP calls were made using Affymetrix Power Tools suite.

Whole Genome Sequencing: Data were generated from liver genomic DNA from one female mouse for NTac, NCrl and NJ samples, and one JM8.F6 cell pellet for paired-end library construction at SeqWright. 30X sequencing of each sample was carried out on the Illumina HiSeq2000 (~90Gb per sample).

Single nucleotide polymorphisms (SNPs) from each sample were identified with Illumina CASAVA v1.8.2 processing of reads mapped to the mm10 assembly and filtered for those with a coverage of >= 20 and a population percentage of >= 40%. A custom Perl script then sorted into eight sets: those unique to each of the four samples and each of the four intersecting mouse strain possibilities. Color-coded sets were then mapped to chord diagram (Figure 3) generated by Circos v0.55-2.

Insertions and deletions were identified from each sample with Illumina Genome Studio. A custom Perl script was used to generate a distance matrix by establishing the symmetrical difference of indel locations between each possible sample pair combination. This matrix was then used to create a Newick tree using the phylip (v1:3.69-1) nearest-neighbor method, and in turn, rendered using phylip drawtree (Figure 4).

For the Kernel Density Estimate plot of Chromosome 2 (Figure 5), a custom Perl script post-processed the Genome Studio indel output file into a set of chromosome coordinates for each chromosome of each sample. These files were imported into R data frames and Kernel Density Estimates (KDE) of each combined chromosome set were calculated and rendered with density and plot, respectively.

Results

Phenotype Analysis
Data generated at TCP and ICS (Cohorts 1 and 2) detected differences between NCrl and NTac mice for body weight, open field, grip strength, DEXA and hematology. For both of these studies, mice had been reared in environments different from one another; NTac and NCrl were shipped to TCP from vendor barrier rooms and ICS used NTac mice from an internal colony with imported NCrl mice from a vendor barrier.

Interestingly, ICS cohorts 3 and 4 (where the tests were conducted on animals born and raised at ICS) showed no significant differences between NTac and NCrl mice for body weight, open field and DEXA. For grip strength, TCP cohorts showed increased “four paw” strength in NCrl females relative to NTac females, with no differences between males. Figure 1 shows a significant difference in female “two paw” grip strength in ICS cohorts consistent with the “four paw” results from TCP. NCrl males showed a significant decrease in “four paw” grip strength relative to NTac males at ICS.

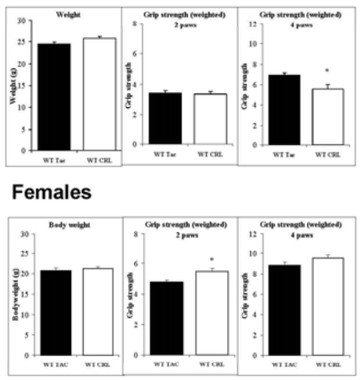

Figure 1. Grip strength in male and female C57BL/6NCrl and NTac mice at ICS. No significant differences in males were observed in TCP cohorts, but NCrl females showed significantly increased “4-paw” grip strength relative to NTac females.

Figure 2 shows significant differences in white blood cell (WBC) count between males and females from one cohort tested at ICS. In the other cohort, a similar trend was noted, although it was not statistically significant. Male NCrl animals tested at TCP also showed a significant decrease in WBC relative to NTac males however no significant differences were noted between females.

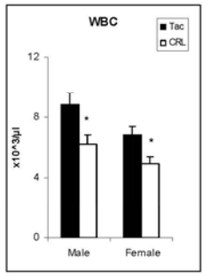

Figure 2. White blood cell count for one ICS cohort. NCrl males had significantly lower WBC counts in TCP cohorts as well, although there was no significant difference in females at TCP.

Results, cont.

A summary of results for ICS cohorts 3 and 4 is shown in Table 1. Taken together with the results from cohorts 1 and 2, along with the TCP results listed above, these data indicate that early environment has a significant impact on phenotype differences detected between NTac and NCrl mice in these studies.

Data are expressed as phenotype of CRL as compared to Taconic
increased
decreased

Phenotyping test	Male p<0,05 C3 (e&f)	Male p<0,05 C4 (i&j)	Female p<0,05 C3 (e&f)	Female p<0,05 C4 (i&j)	Male p<0,01 C3 (e&f)	Male p<0,01 C4 (i&j)	Female p<0,01 C3 (e&f)	Female p<0,01 C4 (i&j)
Energy expenditure (TSE)	Not done	No difference	Not done	No difference				
ECG	Not done	No difference	Not done	No difference				
ECHO	Not done	No difference	Not done					increase in aortic flow velocity and integral (stenotic aortic valve?)
Resp	Not done	No difference	Not done	No difference				
DEXA	No difference	No difference	No difference	No difference				
Hemato	WBC: Increased	WBC: Trend Decreased	WBC: Increased	WBC: Trend Decreased				
openfield/Distance	RAS	RAS	RAS	RAS				
openfield/Rears	RAS	RAS	RAS	RAS				
openfield/center exploration	RAS	RAS	RAS	RAS				
Openfield/Average speed	RAS	RAS	RAS	Increased				
grip/2 paws	RAS	Increased	Trend	RAS				
grip/4 paws	Decreased	RAS	Trend	RAS				
Rotarod	RAS	RAS	RAS	RAS				
Ymaze	RAS	RAS	RAS	RAS				
Startle	RAS	RAS	RAS	RAS				
PPI	RAS	RAS	RAS	RAS				
ABR Threshold	RAS/effectif faible	RAS/effectif faible	RAS/effectif faible	RAS/effectif faible				
Hpt plate	RAS	RAS	RAS	RAS				
Vision/Slit Lamp	RAS	Not performed	RAS	Not performed				
Vision/Ophtalmoscope	RAS	RAS	RAS?	RAS				
Histo	RAS	RAS	RAS?	RAS				

Table 1 Summary

Genotype Analysis:
Affymetrix analysis: There were a total of 34,647 probe sets with at least one different call among all samples. Those sets that resulted in >50% no calls within or between samples were excluded from analysis, as well as those sets where only one sample in a given set was different from all other samples. Of the remaining 1,545 probe sets, 331 correlated with sex difference. For the remaining 1,214 probe sets, 196 were identified with genotype differences among sample groups. Principle component analysis (PCA) with these 196 probe sets was used to determine relatedness among the genomic DNA samples. Not surprisingly, the C57BL/6J clustered away from the other samples while N-derived samples clustered more closely. To more closely examine the relatedness among the N-derived samples, PCA analysis was repeated using only those samples. Interestingly, results from this analysis indicated a closer clustering between NCrl and JM8 samples, along with a similar cluster between NTac and C2 genomes. However, clustering distances between all four genomes were very small (data not shown).
Whole Genome Sequencing:
Figure 3 displays the SNP distribution among the three C57BL/6N genomes as well as for the JM8.F6 cell line. A Venn diagram was generated to better illustrate overlapping and unique SNP calls for the three mouse strains. There were 12,281 SNPs shared among the strains. NJ and NTac genomes had more SNPs in common than either did with NCrl, and NCrl shared more SNPs with NTac than with NJ. Each C57BL/6N strain had approximately 5000 SNPs unique to their genome.

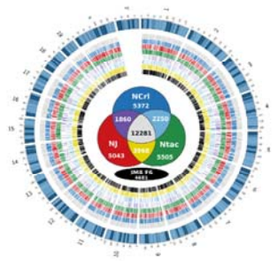

Figure 3. SNP distribution among C57BL/6NCrl, NTac, NJ mice and JM8.F6 ES cells. Numbers in Venn-diagram indicate shared SNPs (in overlaps), and unique SNPs (non-overlap) among genomes. Only unique SNPs are indicated for the JM8.F6 cell line.

Results, cont.

Analysis of indels detected in the four genomic sequences was used to further examine differences between them. Figure 4 is a non-rooted phylogenetic tree based upon the indel analysis. As with the SNPs, NTac and NJ clustered more closely to one another than they did to NCrl. The JM8.F6 cell line was more closely aligned with NCrl, a result consistent with the Affymetrix MDG array data.

Figure 4. Non-rooted indel tree showing relationship between C57BL/6NCrl, NTac, NJ mice and JM8.F6 ES cell line.

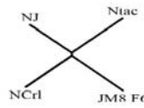

Figure 5 shows a density map of indels for chromosome 2 for NCrl, NTac and NJ mice and JM8.F6 ES cells. The JM8.F6 genome has a region with a large indel (in this case an insertion or duplication) not present in the three mouse genomes, which likely indicates a rearrangement has occurred during passage of this cell line over time.

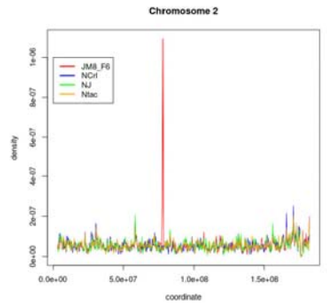

Figure 5. Density map of indels on chromosome 2 of C57BL/6NCrl, NTac, NJ mice and JM8.F6 ES cell line demonstrating a large insertion/duplication present only in the ES cell genome.

Discussion

Phenotype Analysis:
Our results demonstrate that there may be no significant differences between NTac and NCrl C57BL/6 mice for the phenotypes measured by the IMPC phenotyping pipeline, provided the early rearing environment is the same for both strains. Possible exceptions to this include grip strength and WBC, with differences being detected in most cohorts regardless of location. There are currently three more cohorts consisting of site-reared animals (one at ICS and two at TCP), being run through the entire pipeline to help determine if robust phenotypic differences are present between NTac and NCrl.

Regardless of those results, it is important to note that early rearing environment has an effect on many of the IMPC pipeline phenotypes, and this should be considered when comparing data between phenotyping centers and in experimental design for phenotyping where the strains being compared may be available from different sources.

Genotype Analysis:
Results from genome scanning using the Affymetrix MDG array along with WGS of B6N or B6N-derived genomes indicates there is a high degree of similarity among the strains and ES cell examined. While NTac and NJ tended to cluster closer to each other than to NCrl and JM8.F6, the differences are small.

These results are consistent with the phenotype data generated for this study. The large insertion/duplication detected on chromosome 2 of the JM8.F6 ES cells points to the relative instability in genomes of cell lines. It is important to note that mice generated from this resource will have been backcrossed several generations to help lessen the impact of this type of instability on IMPC phenotyping efforts, as well as for mice distributed to the scientific community.

All sequence and SNP data will be deposited in publically available databases once our analysis is complete.