Genome Biology

Genome Biology

A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains

Genome Biology 2013, **14**:R82 doi:10.1186/gb-2013-14-7-r82

Michelle M Simon (m.simon@har.mrc.ac.uk)
Simon Greenaway (s.greenaway@har.mrc.ac.uk)
Jacqueline K White (jkw@sanger.ac.uk)
Helmut Fuchs (hfuchs@helmholtz-muenchen.de)
Valérie Gailus-Durner (gailus@helmholtz-muenchen.de)
Tania Sorg (tsorg@igbmc.fr)
Kim Wong (kw10@sanger.ac.uk)
Elodie Bedu (bedu@igbmc.fr)
Elizabeth J Cartwright (elizabeth.j.cartwright@manchester.ac.uk)
Romain Dacquin (romain.dacquin@etu.univ-lyon1.fr)
Sophia Djebali (sophia.djebali@inserm.fr)
Jeanne Estabel (je2@sanger.ac.uk)
Jochen Graw (graw@helmholtz-muenchen.de)
Neil J Ingham (neil.ingham@kcl.ac.uk)
Ian J Jackson (ian.jackson@igmm.ed.ac.uk)
Andreas Lengeling (andreas.lengeling@roslin.ed.ac.uk)
Silvia Mandillo (smandillo@ibc.cnr.it)
Jacqueline Marvel (jacqueline.marvel@inserm.fr)
Hamid Meziane (meziane@igbmc.fr)
Frédéric Preitner (Frederic.Preitner@unil.ch)
Oliver Puk (oliver.puk@helmholtz-muenchen.de)
Michel Roux (mjroux@igbmc.fr)
David J Adams (da1@sanger.ac.uk)
Sarah Atkins (s.atkins@har.mrc.ac.uk)
Abdel Ayadi (ayadi@igbmc.fr)
Lore Becker (lore.becker@helmholtz-muenchen.de)
Andrew Blake (a.blake@har.mrc.ac.uk)
Debra Brooker (d.brooker@har.mrc.ac.uk)
Heather Cater (h.cater@har.mrc.ac.uk)
Marie-France Champy (champy@igbmc.fr)
Roy Combe (combe@igbmc.fr)
Petr Danecek (pd3@sanger.ac.uk)
Armida di Fenza (a.difenza@har.mrc.ac.uk)
Hilary Gates (h.gates@har.mrc.ac.uk)
Anna-Karin Gerdin (akg@sanger.ac.uk)
Elisabetta Golini (egolini@ibc.cnr.it)
John M Hancock (jmhancock@gmail.com)
Wolfgang Hans (wolfgang.hans@helmholtz-muenchen.de)
Sabine M Hölter (hoelter@helmholtz-muenchen.de)

Genome Biology

Genome Biology

Tertius Hough (t.hough@har.mrc.ac.uk)
Pierre Jurdic (pjurdic@ens-lyon.fr)
Thomas M Keane (tk2@sanger.ac.uk)
Hugh Morgan (h.morgan@har.mrc.ac.uk)
Werner Müller (werner.muller@manchester.ac.uk)
Frauke Neff (frauke.neff@helmholtz-muenchen.de)
George Nicholson (george.nicholson@stats.ox.ac.uk)
Bastian Pasche (Bastian.Pasche@helmholtz-hzi.de)
Laura-Anne Roberson (lr4@sanger.ac.uk)
Jan Rozman (jan.rozman@helmholtz-muenchen.de)
Mark Sanderson (ms14@sanger.ac.uk)
Luis Santos (l.santos@har.mrc.ac.uk)
Mohammed Selloum (selloum@igbmc.fr)
Carl Shannon (cs7@sanger.ac.uk)
Anne Southwell (a.southwell@har.mrc.ac.uk)
Glauco P Tocchini-Valentini (tocchini@ibc.cnr.it)
Valerie E Vancollie (vv2@sanger.ac.uk)
Sara Wells (s.wells@har.mr.ac.uk)
Henrik Westerberg (h.westerberg@har.mrc.ac.uk)
Wolfgang Wurst (wurst@helmholtz-muenchen.de)
Min Zi (min.zi@manchester.ac.uk)
Binnaz Yalcin (Binnaz.Yalcin@well.ox.ac.uk)
Ramiro Ramirez-Solis (rrs@sanger.ac.uk)
Karen P Steel (karen.steel@kcl.ac.uk)
Ann-Marie Mallon (a.mallon@har.mrc.ac.uk)
Martin Hrabě de Angelis (hrabe@helmholtz-muenchen.de)
Yann Herault (herault@igbmc.fr)
Steve DM Brown (s.brown@har.mrc.ac.uk)

ISSN 1465-6906

Article type Research

Submission date 18 March 2013

Acceptance date 28 June 2013

Publication date 31 July 2013

Article URL http://genomebiology.com/2013/14/7/R82

Genome Biology

A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains

Michelle M Simon[1], Simon Greenaway[1], Jacqueline K White[2], Helmut Fuchs[3a], Valérie Gailus-Durner[3a], Tania Sorg[4], Kim Wong[2], Elodie Bedu[4], Elizabeth J Cartwright[5a], Romain Dacquin[6], Sophia Djebali[7], Jeanne Estabel[2], Jochen Graw[3b], Neil J Ingham[2], Ian J Jackson[8], Andreas Lengeling[9], Silvia Mandillo[10], Jacqueline Marvel[6], Hamid Meziane[4], Frédéric Preitner[11], Oliver Puk[3b], Michel Roux[4], David J Adams[2], Sarah Atkins[1], Abdel Ayadi[4], Lore Becker[3a], Andrew Blake[1], Debra Brooker[1], Heather Cater[1], Marie-France Champy[4], Roy Combe[4], Petr Danecek[2], Armida di Fenza[1], Hilary Gates, Anna-Karin Gerdin[2], Elisabetta Golini[10], John M Hancock[1], Wolfgang Hans[3a], Sabine M Hölter[3a], Tertius Hough[1], Pierre Jurdic[6], Thomas M Keane[2], Hugh Morgan[1], Werner Müller[5b], Frauke Neff[3c], George Nicholson[1], Bastian Pasche[12], Laura-Anne Roberson[2], Jan Rozman[3a], Mark Sanderson[2], Luis Santos[1], Mohammed Selloum[4], Carl Shannon[2], Anne Southwell[1],Glauco P Tocchini-Valentini[10], Valerie E Vancollie[2], Sara Wells[1], Henrik Westerberg[1], Wolfgang Wurst[3b,15,16,17], Min Zi[5a], Binnaz Yalcin[13, 14], Ramiro Ramirez-Solis[2], Karen P Steel[2], Ann-Marie Mallon[1], Martin Hrabě de Angelis[3a], Yann Herault[4] and Steve DM Brown[1]*

[1] Medical Research Council Harwell (Mammalian Genetics Unit and Mary Lyon Centre), Harwell Science Campus, OX11 0RD, UK
[2] The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, UK
[3a] Helmholtz Zentrum München, German Research Centre for Environmental Health, Institute of Experimental Genetics and German Mouse Clinic, Neuherberg, 85764, Germany
[3b] Helmholtz Zentrum München, German Research Centre for Environmental Health, Institute of Developmental Genetics, Neuherberg, 85764, Germany
[3c] Helmholtz Zentrum München, German Research Centre for Environmental Health, Institute of Pathology, Neuherberg, 85764, Germany
[4] Institut Clinique de la Souris, ICS/MCI, PHENOMIN, GIE CERBM, IGBMC, CNRS, INSERM, 1 Rue Laurent Fries, 67404 Illkirch-Graffenstaden Cedex, France
[5a] Faculty of Medical and Human Sciences, University of Manchester, Oxford Road, Manchester, MN13 9PT, UK
[5b] Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester, MN13 9PT, UK
[6] Institut de Génomique Fonctionnelle de Lyon, 32-34 Avenue Tony Garnier, 69007 Lyon, France
[7] Ecole Normale Supérieure de Lyon, 50 Avenue Tony Garnier, 69007 Lyon, France
[8] Medical Research Council Human Genetics Unit, IGMM, University of Edinburgh Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK
[9] Infection and Immunity Division, The Roslin Institute, University of Edinburgh, Easter Bush Veterinary Campus, Edinburgh, EH25 9RG, UK
[10] Consiglio Nazionale delle Ricerche- Cell Biology and Neurobiology Institute, Via E.Ramarini 32, 00015 Monterotondo Scala, Italy
[11] Department of Infection Genetics, Helmholtz Centre for Infection Research, Braunschweig, 38124, Germany
[12] Mouse Metabolic Facility of the Cardiomet Center, University Hospital, and Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland
[13] Institute of Genetics and Molecular and Cellular Biology, 67404 Illkirch, France
[14] Center for Integrative Genomics, University of Lausanne, Lausanne, CH-1015, Switzerland
[15] Chair for Developmental Genetics, Technische Universität München, Arcisstr. 21, Munich, 80333, Germany

[16] Max Planck Institute of Psychiatry, Kraepelinstrasse 2, Munich, 80804, Germany
[17] Deutsches Zentrum für Neurodegenerative Erkrankungen, Schillerstrasse 44, Munich, 80336, Germany

* Corresponding Author: Steve D.M. Brown- s.brown@har.mrc.ac.uk

Email:
Michelle M. Simon- m.simon@har.mrc.ac.uk; Simon Greenaway- s.greenaway@har.mrc.ac.uk; Jacqueline K. White- jkw@sanger.ac.uk; Helmut Fuchs- hfuchs@helmholtz-muenchen.de; Valérie Gailus-Durner- gailus@helmholtz-muenchen.de; Tania Sorg- tsorg@igbmc.fr; Kim Wong- kw10@sanger.ac.uk; Elodie Bedu- bedu@igbmc.fr; Elizabeth J. Cartwright- elizabeth.j.cartwright@manchester.ac.uk; Romain Dacquin- romain.dacquin@etu.univ-lyon1.fr; Sophia Djebali- sophia.dejbali@inserm.fr; Jeanne Estabel- je2@sanger.ac.uk; Jochen Graw- graw@helmholtz-muenchen.de; Neil J. Ingham- neil.ingham@kcl.ac.uk; Ian J. Jackson- ian.jackson@igmm.ed.ac.uk; Andreas Lengeling- andreas.lengeling@roslin.ed.ac.uk; Silvia Mandillo- smandillo@ibc.cnr.it; Jacqueline Marvel- Jacqueline.marvel@inserm.fr; Hamid Meziane- meziane@igbmc.fr; Frédéric Preitner- Frederic.Preitner@unil.ch; Oliver Puk- oliver.puk@helmholtz-muenchen.de; Michel Roux- mjroux@igbmc.fr; David J. Adams- da1@sanger.ac.uk; Sarah Atkins- s.atkins@har.mrc.ac.uk; Abdel Ayadi- ayadi@igbmc.fr; Lore Becker- lore.becker@helmholtz-muenchen.de; Andrew Blake- a.blake@har.mrc.ac.uk; Debra Brooker- d.brooker@har.mrc.ac.uk; Heather Cater- h.cater@har.mrc.ac.uk; Marie-France Champy- champy@igbmc.fr; Roy Combe- combe@igbmc.fr; Petr Danecek- pd3@sanger.ac.uk; Armida di Fenza- a.difenza@har.mrc.ac.uk; Hilary Gates- h.gates@har.mrc.ac.uk; Anna-Karin Gerdin- akg@sanger.ac.uk; Elisabetta Golini- egolini@ibc.cnr.it; John M. Hancock- jmhancock@gmail.com; Wolfgang Hans- Wolfgang.hans@helmholtz-muenchen.de; Sabine M. Hölter- hoelter@helmholtz-muenchen.de; Tertius Hough- t.hough@har.mrc.ac.uk; Pierre Jurdic- pjurdic@ens-lyon.fr; Thomas M. Keane- tk2@sanger.ac.uk; Hugh Morgan- h.morgan@har.mrc.ac.uk; Werner Müller- werner.muller@manchester.ac.uk; Frauke Neff- Frauke.neff@helmholtz-muenchen.de; George Nicholson- George.nicholson@stats.ox.ac.uk; Bastian Pasche- Bastian.Pasche@helmholtz-hzi.de; Laura-Anne Roberson- lr4@sanger.ac.uk; Jan Rozman- jan.rozman@helmholtz-muenchen.de; Mark Sanderson- ms14@sanger.ac.uk; Luis Santos- l.santos@har.mrc.ac.uk; Mohammed Selloum- Selloum@igbmc.fr; Carl Shannon- cs7@sanger.ac.uk; Glauco P. Tocchini-Valentini- tocchini@ibc.cnr.it; Valerie E. Vancollie- vv2@sanger.ac.uk; Sara Wells- s.wells@har.mrc.ac.uk; Henrik Westerberg- h.westerberg@har.mrc.ac.uk; Wolfgang Wurst- wurst@helmholtz-muenchen.de; Min Zi- min.zi@manchester.ac.uk; Binnaz Yalcin- Binnaz.Yalcin@well.ox.ac.uk; Ramiro Ramirez-Solis- rrs@sanger.ac.uk; Karen P. Steel- Karen.steel@kcl.ac.uk; Anne-Marie Mallon- a.mallon@har.mrc.ac.uk; Martin Hrabě de Angelis- Hrabe@helmholtz-muenchen.de; Yann Herault- Herault@igbmc.fr; Steve D.M. Brown- s.brown@har.mrc.ac.uk*

Abstract

Background

The mouse inbred line C57BL/6J is widely used in mouse genetics and its genome has been incorporated into many genetic reference populations. More recently large initiatives such as The International Knockout Mouse Consortium (IKMC) are using the C57BL/6N mouse strain to generate null alleles for all mouse genes. Hence both strains are now widely used in mouse genetics studies. Here we perform a comprehensive genomic and phenotypic analysis of the two strains to identify differences that may influence their underlying genetic mechanisms.

Results

We undertake genome sequence comparisons of C57BL/6J and C57BL/6N to identify SNPs, indels and structural variants, with a focus on identifying all coding variants. We annotate 34 SNPs and 2 indels that distinguish C57BL/6J and C57BL/6N coding sequences, as well as 15 structural variants that overlap a gene. In parallel we assess the comparative phenotypes of the two inbred lines utilizing the EMPReSSslim phenotyping pipeline, a broad based assessment encompassing diverse biological systems. We perform additional secondary phenotyping assessments to explore other phenotype domains and to elaborate phenotype differences identified in the primary assessment. We uncover significant phenotypic differences between the two lines, replicated across multiple centers, in a number of physiological, biochemical and behavioral systems.

Conclusions

Comparison of C57BL/6J and C57BL/6N demonstrates a range of phenotypic differences that have the potential to impact upon penetrance and expressivity of mutational effects in these strains. Moreover, the sequence variants we identify provide a set of candidate genes for the phenotypic differences observed between the two strains.

Keywords: Mouse inbred lines; sequence variation; mouse phenotyping; gene knock-out; C57BL/6

Background

The development of a comprehensive mouse embryonic stem (ES) cell mutant resource by the International Knockout Mouse Consortium, IKMC [1] is a critical step in the systematic functional annotation of the mouse genome. To date, ES cell mutant lines are available for around 15,000 mouse genes providing a very significant resource for the generation of mutant mice and their subsequent phenotypic analysis. The IKMC resource is being used by the International Mouse Phenotyping Consortium (IMPC) that plans over the next five years to generate and carry out broad based phenotyping on 5000 mouse mutant lines as the first step towards a comprehensive encyclopedia of mammalian gene function [2].

All IKMC mutant clones have been generated using a C57BL/6N ES cell line [1]. Moreover, chimaeras generated from IKMC clones as part of the IMPC programme are bred to C57BL/6N mice maintaining the mutations on an isogenic background. The use of C57BL/6N for these major functional genomics programmes brings into perspective the genetic relationship between the C57BL/6N strain and other inbred strains that have been the focus of mouse genetics research in the past. In particular, a considerable number of mouse genetic resources have been developed using the C57BL/6J strain, including a variety of reference populations such as recombinant inbred lines [3, 4], consomics [5], heterogeneous stocks [6] and the Collaborative Cross [7]. Moreover, a large number of spontaneous mutations have been identified on the C57BL/6J background. As a consequence the C57BL/6J line was the natural choice to provide the first reference sequence of the mouse genome [8, 9]. The significant usage of both the N and J substrains throughout the wider biomedical science communities underlines the need to better understand the genetic and phenotypic relationships between these two inbred strains and how they might affect our understanding of genetic mechanisms and phenotype outcomes.

The inbred C57BL/6 mouse strain was established at the Jackson Laboratory from the parental strain C57BL in 1948 at F24. In 1951, at F32, it was then passed on to the National Institutes of Health (NIH) leading to the C57BL/6N line. The C57BL/6Ntac substrain was established at F151 following the transfer of the C57BL/6N line to Taconic Farms in 1991 [10]. Thus currently C57BL/6J and C57BL/6N have been separated for around 220 generations. Early assessment of the genetic variation between C57BL/6J and C57BL/6N substrains using a panel of 1,427 SNP loci identified only 12 SNPs (0.8%) between the two strains [Medaka et al.], reflecting their close genetic relationship.

In 2011, an extensive analysis of genomic variation in 17 inbred strains catalogued an extraordinarily large number of variants, including 56.7M single nucleotide polymorphisms (SNPs), 8.8M small indels and 0.28M structural variants (SVs) across both the classical laboratory strains, as well as wild-derived lines [11]. In addition, these analyses illustrated the potential to relate sequence variation to aspects of phenotypic variation between mouse strains. Importantly, the analyses provided an insight into the molecular and genetic basis of quantitative traits that distinguish the phenotypic characteristics of inbred strains [11]. Small effect QTLs were found to be more often due to intergenic variation and are unlikely to be the result of structural variation. In contrast larger effect QTLs are usually explained by intronic variation. However, for the small proportion of QTLs of very large phenotypic effect, there is a significant enrichment of coding variation, with an increasing frequency of SVs and small indels. Although overall the proportion of SVs within the mouse genome that cause major phenotypic effects is small, it is likely that SVs that cause phenotypic change will provide significant insights into gene function [11]. This work underlines

the utility and importance of cataloguing genomic variation in the mouse and analyzing its contribution to phenotypic effects.

We focus our analysis on a detailed genomic and phenotype comparison of the C57BL/6N and C57BL/6J strains, aiming to relate the underlying genomic differences to phenotype outcomes. We expand and refine the analysis of the two inbred strain genome sequences. Importantly, using the new short-read genome sequence of the C57BL/6J generated by the Broad institute and improved analytical tools, we identify a high-quality set of variants including SNPs, small indels and SVs that distinguish the C57BL/6N and C57BL/6J strains, with a particular focus on cataloguing variation of coding sequences. Using a combination of experimental methods, we validated all coding variants and SVs generating a significantly higher quality variant dataset than generated from the 17 genomes project, with a null false positive rate. In parallel, we have undertaken a comprehensive phenotypic comparison and have examined the relationship between genome variation and phenotypic changes in these two substrains.

Results

Genome sequence comparisons of C57BL/6N and C57BL/6J mice – SNPs and small indels

We utilized the paired-end alignment of C57BL/6N to the reference genome (C57BL/6J) from the 17 Genomes project [10]. However the list of differentiating variants (SNPs, small indels and SVs) between the two genomes was newly created using novel inbuilt procedures in order to increase the likelihood of identifying accurate putative sequence changes. A key analysis step in identifying a high quality set of variants from the alignment was to utilise the newly generated short read

genome sequence of C57BL/6J generated by the Broad Institute. This enabled us to identify assembly errors in the reference sequence. In addition, we updated the variant detection method first, by using different and/or more evolved software to detect variants; second, by performing manual curation on all coding variants and finally, employing extensive validation of a large proportion of the variants (including all coding variants) to confirm the sequence predictions. These steps provided a robust dataset of high quality coding variants considerably reducing the false positive rate.

To identify SNPs and small indels differentiating C57BL/6J and C57BL/6N strains, we used the paired-end reads generated from the 17 Mouse Genomes Project [11]. We called variants using the Genome Analysis Toolkit (GATK) [12] and found 681,220 variants that distinguish C57BL/6J and C57BL/6N strains. Using the short-read genome sequence of C57BL/6J generated by the Broad Institute [13], we were able to filter out prospective sequencing errors by removing variants common to the Broad C57BL/6J sequence. This counteracts discrepancies in the reference whilst improving the false negative rate. The remaining reads were filtered with an allele ratio <0.8 (heterozygous) and covered by < 3 or >150 reads. These steps reduced significantly the list to 10,794 putative variants that was subjected to further analyses.

Using Sequenom, PyroSequencing and Sanger sequencing, we validated all coding variants and a subset of the non-coding variants, which included 762 SNPs and 169 small indels. Assays were carried out using a panel of four C57BL/6J and four C57BL/6N samples in order to confirm genotypes (see Methods). We considered a variant as validated when all four C57BL/6J and C57BL/6N samples demonstrated consistent genotypes within a substrain and variant between the substrains. During

the validation process we eliminated 363 variants due to a number of reasons, including heterozygous and inconsistent genotypes as well as PCR failures. For the remaining 568, 236 were confirmed as variant between the substrains (Additional file 1, Table S1).

Using the annotation programs NGS-SNP and Annovar [14, 15], the genomic location and other gene features were examined. The final validated sequence variants between C57BL/6J and C57BL/6N consist of 34 coding SNPs, 2 coding small indels, 146 non-coding SNPs and 54 non-coding small indels. Coding variants include 32 missense SNPs, 1 nonsense mutation, 1 splicing and 2 frameshift mutations (Table 1). We found that all variants except one (*Zp2*, chromosome 7) were private to either N or J, and not observed in any of the 16 other inbred strains recently sequenced [11](Keane et al. 2011).

Genome sequence comparisons of C57BL/6N and C57BL/6J mice – structural variants (SVs)

Again, employing the paired-end reads generated from the 17 Mouse Genomes Project [11] and a combination of four computational methods [16], we identified 551 SVs between C57BL/6J and C57BL/6N. As described elsewhere [17], we visually inspected short-read paired-end mapping at these 551 SV sites in the 17 sequenced inbred strains of mice [11] and in the Broad J sequenced genome [13]. By doing this, we were able to retain 81 of the 551 sites for further experimental analyses (470 predicted sites were false due to paired-end mapping errors). PCR and Sanger-based sequencing analyses at these 81 retained sites, allowed us to remove a further 38 sites, which were confirmed to not be polymorphic between C57BL/6J and C57BL/6N because of reference errors. At the end all 43 predicted variants were

validated as authentic SVs differentiating C57BL/6J and C57BL/6N strains (Table 2), resulting in a null false positive rate.

Fifteen of the 43 SVs overlap with a gene (Table 2), including 12 variants that lie within non-coding regions of genes, 2 variants that affect the coding region of the gene (*Vmn2r65*, Vomeronasal 2, receptor 65; *Nnt*, nicotinamide nucleotide transhydrogenase) and one that affects the entire gene *Cyp2a22* (cytochrome P450, family 2, subfamily a, polypeptide 22). Only one of the fifteen variants is known and has been already associated with a phenotype, *Nnt* [18]; the remaining 14 are novel and for several we discuss their potential biological functions below.

Using the rat as an outgroup species, we next inferred the origin of the 43 SVs between C57BL/6J and C57BL/6N, and found that 27 variants were the product of retrotransposition, 15 were non-repeat mediated SVs and 1 was a variable number tandem repeat (VNTR) (Table 2). Remarkably, almost all variants were private to either C57BL/6J or C57BL/6N (Table 2).

Comprehensive phenotypic assessment of C57BL/6N and C57BL/6J mice

In parallel to the genomic analyses, the European Mouse Disease Clinic (EUMODIC) consortium has carried out a comprehensive phenotypic comparison of the C57BL/6NTac and C57BL/6J strains. EUMODIC comprises four mouse centres [19] carrying out broad based primary phenotyping of 500 mouse mutant knockout lines generated from the European Conditional Mouse Mutagenesis (EUCOMM) and Knockout Mouse (KOMP) projects within the IKMC programme. Cohorts of mice from each mutant line enter the EMPReSSslim phenotype assessment that consists of two phenotyping pipelines together comprising 20 phenotyping platforms that are carried out from 9 to 15 weeks (Additional file 2, Figure S1). The methods for performing

each screen are detailed in Standard Operating Procedures (SOPs) that can be found in the EMPReSS database[20]. Data is acquired on 413 phenotype parameters along with 146 metadata parameters and entered into the EuroPhenome database [19]. As part of this work we have been capturing extensive control data on the baseline phenotype of C57BL/6NTac. We have taken this opportunity to also investigate the phenotype of C57BL/6J mice and to compare this to C57BL/6NTac (henceforth referred to as J and N respectively).

For each line, N and J, age-matched mice have been analysed through both EMPReSSslim (European mouse phenotyping resource of standardized screen) pipelines. Data was acquired from all four centres in the consortium for 19 of the 20 platforms from the pipeline (excluding fluorescence-activated cell sorting, FACs, analyses, see Additional file 2, Figure S1). The EMPReSSslim protocols have been rigorously standardized in the EUMODIC consortium, however there remain some differences in, for example, equipment and diet, and this is captured in the metadata sets within EuroPhenome. There will of course be other unrecognized environmental differences between centres. Collectively these may contribute to gene-environment differences and phenotype outcome, but we have not sought to systematically define these effects, instead focusing on phenotypes that are concordant between centres and are clearly robust to unrecognized environmental perturbations. Data from the N and J cohorts from each centre has been deposited in EuroPhenome and has been subjected to a statistical analysis for each centre (see Methods). It is important to note here that comparisons between N and J were performed within, not between centres. Statistical analysis of results between centres was not performed as experiments could not be completely controlled between centres due to environmental and other variables and differences in numbers of animals analysed in each centre (Additional file 3, Figure S2a-d). We thus chose to adopt an approach

that focused on strain comparisons within individual centres as opposed to generating a multi-centre statistical model that examined an overall statistical difference between the two strains. However, replication of the N and J comparison across multiple centres provides us with additional power in substantiating significant phenotypic differences between the two strains [see Methods]. In addition to the analysis of N and J through the EMPReSSslim primary phenotyping pipeline at the four centres, other partners within the EUMODIC consortium have applied a wider range of often more sophisticated phenotyping tests to gather additional information, some of which explores further and aims to substantiate the phenotypic differences revealed through EMPReSSslim.

In analyzing the data, we have focused first on identifying phenotype parameters that show a consistent and significant difference between N and J in 3 or more centres. We identified 27 phenotype parameters in this class (Figure 1a and Additional file 3, Figure S2a, e). In several cases, these differences are supported by data from secondary analysis, and we discuss these instances below. We also uncovered a second class of parameters where similar trends were seen in two centres, but no evidence of trends in the other two centres (Figure 1b and Additional file 3, Figure S2b, f). However, our statistical analysis (see Methods) indicates that for this class of parameters the overall significance of N vs J differences is low, and the trends observed should be treated with caution. In several of these cases however, the observed trends are consistent with phenotypes found in the first class. We also identified a third but small class of parameters that showed highly significant differences in 2 or more centres (Figure 2 and Additional file 3, Figure S2d, h), but surprisingly the opposite trend in one of the centres. We discuss the reasons for these anomalies, in some cases presumably arising from gene-centre interactions. The final class represents a large number of tests where we did not observe any

consistent and significant differences across the centres, concluding that these are more likely to be false positives rather than evidence for N/J differences (Additional file 3, Figure S2c, g).

Dysmorphology and Ophthalmology

We found no evidence for any major differences in morphological features between N and J, including X-ray analysis of the skeleton. However, a number of ophthalmological differences between the two strains were identified. Analysis of the general visual functions using the virtual optokinetic drum [21] revealed a reduced vision of N in comparison to J mice (0.314 cycle/degree, 95% CI 0.305-0.323 for B6N mice (n = 89), vs. 0.399 cycles/degree, 95% CI 0.394-0.404 for B6J mice (n = 128), $p < 0.001$, Student t-test). This did not reflect differences in lens opacities as a quantitative analysis using a Scheimpflug camera demonstrated transparent lenses in both strains (J: 5.0% + 0.5 % opacity, n = 10; N: 5.2 + 0.5 %, n = 10). White flecks were observed in the fundus of N mice at a high frequency, which were absent in J mice (Figure 3A). This is likely due to the presence of the *Cbr1*^{rd8} mutation in N mice as reported by Mattapallil et al., 2012 [22], though in our case the flecks were observed only in the ventral retina, with variable fleck size and affected area from one mouse to the other (Figure 3A). Further studies using topical fundus endoscopy [23] revealed that the number of main vessels was variable, ranging between 3 and 7 for veins and 3 and 8 for arteries (Figure 3B); a given mouse could have non matching numbers between the two eyes. The mean number of both veins and arteries was significantly higher ($p < 0.001$) in J compared to N mice (Figure. 3C).

Cardiovascular

Non-invasive blood pressure measurements (ESLIM_002) demonstrated that systolic arterial pressure in J mice was significantly higher than N, although the significance of the effect was found to be variable between sexes and between centres. Moreover, all centres observed that pulse rate was significantly higher in N (although heart rate under anaesthetic was found to be significantly lower in N male mice than J by a secondary partner within the consortium; reflected in a long RR and QTc interval). We also found that heart weight normalized to tibia length (ESLIM_020) was significantly lower in N mice versus J mice in 2 of the centres, and these results were independently confirmed by secondary analysis. Further studies of cardiac structure and function by echocardiography and of cardiac contractile function by haemodynamics failed to reveal any differences between N and J (data not shown).

Metabolism

For indirect calorimetry measurements (free-fed) (ESLIM_003), we observed a consistent difference between N and J for O_2 consumption, CO_2 production and heat production. J showed reduced gas exchange and lower energy expenditure (heat production or metabolic rate) compared to N, which was generally more marked in females. In secondary phenotyping with fasted indirect calorimetry, there was a trend towards lower energy expenditure in J versus N during the night period. This was possibly associated with decreased ambulatory activity in J and lower food intake in J compared to N during the night period, especially on re-feeding (data not shown). There was no consistent difference in activity in the free-fed calorimetry screen (ESLIM_003) in the two centres where activity is measured (Additional file 3, Figure S2c, g). Simplified intraperitoneal glucose tolerance tests (IPGTTs) (ESLIM_004) demonstrated impaired glucose tolerance in J vs N mice. These observations on glucose metabolism are consistent with the known deletion of the

Nnt gene specific to J mice [18] that has been shown to play a role in the regulation of the insulin response in pancreatic beta cells.

DEXA (Dual Energy X-ray Absorptiometry) body composition and bone densitometry measurements (ESLIM_005) demonstrate that N mice have increased fat mass (both absolute and normalized to weight). Furthermore, DEXA measurements indicate that J mice have increased lean mass compared to N. In two of the centres, Bone Mineral Density (BMD) measurements were higher in J male mice. However, this finding was not replicated in the third centre that undertook DEXA screens. We proceeded to undertake μCT analysis of the two strains (Figure. 4) and found that cortical thickness, cortical porosity and trabecular bone volume were unchanged between N and J. In addition, analysis of various micro-architecture parameters indicated that the overall trabecular network is similar. Finally, measurement of bone formation and resorption markers failed to reveal any differences between the two strains (Figure 4).

Neurological, behavioural and sensory

Two centres showed major and consistent differences between N and J in activity in the open field (ESLIM_007) (Figure 2) including higher activity measured by distance travelled in J mice and raised centre entries, indicative of reduced anxiety in J mice. These differences are in accord with data reported recently on a behavioural comparison of N and J [24]. Interestingly the most significant effects were confined to males in the two centres. Surprisingly, in a third centre, the reverse was seen with N mice more active than J, though these effects were seen in both males and females. A fourth centre did not detect these effects, finding no significant differences. The centres employed the EMPReSSslim SOP for the procedure, which included a requirement for similar sized arenas, but there were some operational

differences amongst the centres including use of single or multiple rooms to house the arenas; transparent or opaque sided arenas; and the absence or presence of environmental enrichment in home cages (which is known to have an effect on behavioural outcomes [25]). However, none of these variables were consistent with the differing observations between centres. We cannot rule out however influences of the gut microbiome, which might be expected to differ between centres. The gut microbiome is known to influence CNS function and behavior, mainly through the hypothalamic-pituitary-adrenal (HPA) axis [26]. We conclude that under certain conditions significant differences in open-field parameters can be revealed between N and J, but the nature of these differences is sensitive to unknown environmental conditions. It is interesting that the major contradictory finding in N vs J phenotypes was confined to a behavioural phenotyping platform. In contrast, for most other tests (aside from a few hematological and clinical chemistry parameters, see below), we did not find inconsistencies, indicating that in contrast to most phenotyping platforms, behavioural analyses can be acutely sensitive to environmental parameters.

We also carried out a light/dark transition test to compare anxiety in N and J strains (Figure. 5). We found no significant differences between N and J mice in the number of light-dark transitions or in percentage time spent in the dark compartment. However, the latency to enter the dark compartment was significantly higher in N mice. Modified SHIRPA (SmithKline Beecham, Harwell, Imperial College, Royal London Hospital, Phenotype Assessment) testing (ESLIM_008, Figure 1a) in all 4 centres indicated that male J mice showed significantly raised locomotor activity that correlates with the findings of increased distance travelled in open-field testing in some centres (see above).

We conducted a number of tests that reflect motor ability. Differences in grip-strength (ESLIM_009) were observed across all centres with J higher than N, but the parameters affected were different, with some centres reporting differences in forelimb grip-strength and some for forelimb and hindlimb grip-strength combined (Figure 1a and 1b). Rotarod testing (ESLIM_010) revealed significant differences in latency to fall across all centres, although the reduced motor ability of N was only seen for females in two of the centres. We further explored motor abilities in N and J male mice by examining motor learning performance on the rotarod over 4 days (Figure 5). While motor performance of J mice improved markedly from day 1 to day 2, the performance of N mice improved only gradually and was significantly different from day 1 measurements only from day 3 ($p<0.05$) onwards. Moreover, from day 2 to day 4 there were highly significant differences in the latency to fall between N and J. The primary testing carried out at the centres thus uncovered a potential reduced motor performance in N that was confirmed and further elaborated by more sophisticated testing of motor learning performance.

We also carried out two additional behavioural tests to further elaborate N vs J differences. Firstly, we compared the performance of N and J in the Morris Water Maze test used to assess spatial memory. N male mice show very significantly reduced performance (higher latency) compared with J (Figure 6). Secondly, we examined emotional learning or memory for an aversive event using the cue and contextual fear conditioning tests. However, here we observed no significant differences between the two strains (data not shown).

Exploration of acoustic startle and pre-pulse inhibition (ESLIM_011, Figure 1a) in the two strains revealed a variety of parameters that were significantly and consistently different across centres. Acoustic startle magnitude at 110dB as well as startle

response magnitude to pre-pulse and pulse (PP1-PP4 + pulse, see EMPReSSslim) were reduced in N mice compared with J, though this effect was not seen in females in one centre. Consistent with these observations we found that pre-pulse inhibition differs between N and J mice, with pre-pulse inhibition at PP2 and PP3 and global inhibition raised in N mice compared with J. Several other startle magnitude and pre-pulse inhibition parameters showed significant effects in one or two centres (Figure 1b and Additional file 3, Figure S2b, f) but no differences were observed in other centres. The observations on startle magnitude were not confounded by differences in hearing as we assessed auditory thresholds in J and N mice using the ABR (auditory brainstem response) test and found no differences (data not shown).

Clinical Chemistry

Extensive panels of clinical chemistry tests were performed on plasma samples collected at the end of each phenotyping pipeline. The blood sample at the end of Pipeline 1 (ESLIM_021) was collected following an overnight fast, while the sample at the end of Pipeline 2 (ESLIM_015) was a free-fed sample. Data from at least three centres agreed that urea and the electrolytes sodium, potassium and chloride are significantly higher in plasma from J mice relative to N mice (Figure 1a), though there was some clear sex-centre interactions. Data for free-fed and fasted plasma glucose levels, indicate that for each test at least two centres found plasma glucose levels to be higher in N mice compared to J mice (Figure 1b). However, blood glucose levels are known to be affected by animal handling, sample processing and the use of anaesthetics. Data presented here are from samples collected under gaseous isofluorane anaesthetic, aside from one centre where samples were collected under ketamine/xylazine injection (see Figure 1). As discussed above, due to their known impairment in insulin secretion, it seems contradictory for J mice to have lower plasma glucose levels than N mice, but the deletion in *Nnt* appears to

affect glucose clearance rates only and fasted or non-challenged J mice do not suffer from constant hyperglycaemia. Several other parameters were shown to be higher in J than in N mice in at least two centres while in each case the other centre/s reported no significant differences in the same parameters (Figure 1b); for example, free fatty acids. Two centres demonstrated that iron is significantly higher in N males and that ALP (alkaline phosphatase) is significantly higher in J males. One of these centres also found the same to be true in females (Figure 2). Data for each of these parameters from a third centre contradict these findings.

Haematology

Various haematological parameters were measured at the end of Pipeline 2 (ESLIM_016). A number of parameters showed significant changes in two centres but were not replicated in the others, including white blood counts (WBC), red blood cell counts (RBC), mean cell volume (MCV) and mean corpuscular haemoglobin (MCH) (Figure 1b). Contradictory results were obtained however for the haematocrit and mean cell haemoglobin concentration tests (Figure 2). In each case data from two centres agree while a third centre demonstrated the opposite effect. This could potentially be due to the different machine technologies employed for haematological measurements in the participating clinics as recorded in the metadata.

Immune function and allergy

We investigated a number of secondary phenotypes including host resistance to *Listeria monocytogenes* in J and N strains. Females of the J and N strains were more susceptible to *Listeria monocytogenes* infection; however, the sex difference in *Listeria* host susceptibility was less pronounced in N as compared to J. Males of the N strain showed enhanced clearance of *Listeria* on day 4 post infection as compared to

J males. This correlates with an increased pro-inflammatory response in N males on day 3 post infection as compared to J males (Figure 7).

We also tested N and J mice for DNFB (dinitrofluorobenzene)-induced contact hypersensitivity (CHS). Significant differences in the CHS response were identified between the two strains of mice with J showing an increased CHS response. Noticeably, female mice of both strains showed an increased CHS compared to male mice. Investigation of the responsiveness of NK cells to various stimuli revealed that a larger fraction of NK cells are activated by IL-12 alone or in combination with IL-2 in J mice compared to N; again this response was more significant in females (Figure 8).

Discussion

We found that there are significant phenotypic differences between N and J mice covering a number of physiological, biochemical and neurobehavioural systems. These findings have been replicated across a number of centres, indicating that the differences are robust to environmental variables and are likely to impact upon the comparative analysis of mutations in the two backgrounds in most laboratories. The phenotype differences we find between N and J will require careful consideration when comparing the effects of mutations created in the two genetic backgrounds. While the use of C57BL/6N ES cells has allowed the rapid creation of a valuable genome-wide mutation resource [1], our findings of phenotypic differences between N and J indicates that the analysis of the phenotype data that will emerge from the IKMC resource will require careful interpretation in the context of the considerable legacy of data accumulated for C57BL/6J mutations.

Phenotypic differences between N and J are likely to be accounted for at least in part by variation affecting coding sequences in the two genomes. We have catalogued and validated a total of 36 SNPs and small indels affecting coding sequences along with a total of 43 SVs between the two genomes. In total, we have identified 51 genes carrying some sequence variant or SV that might impact upon gene function. A proportion of these variants are likely to have little or no phenotypic consequence, including many missense mutations and also SVs that do not overlap coding sequence. However, given the pleiotropic nature of most genetic loci and additive and epistatic effects it seems likely that the catalogued coding variation will account for a considerable proportion of the differing phenotypes between the N and J strains.

We have proceeded to test this proposition by comparing the phenotypes arising from knockouts of the identified loci with the phenotypes observed between N and J. First, we examined the available Mammalian Phenotype (MP) ontology terms, derived by analysis of mutant (usually knockout) phenotypes, for all the loci carrying a SNP, small indel or SV between N and J. In so doing we have attempted to draw correlations between phenotypes associated with individual variant loci and the phenotype change observed between N and J. For many of the loci, as might be expected, no mutants have been characterized and phenotypic annotations (MP) are not available. For loci carrying SNPs and small indels, 14 out of the 36 loci had available MP terms. For SVs, where the SV overlaps a gene, we found MP terms for 7 out of 15 loci. In all cases, MP terms have been derived from knockout, presumably loss of function, mutations. For our analysis we have compared the loss of function phenotype found in the homozygous knockout to the phenotypic change between N and J and identified candidate loci that might underlie the observed phenotype effects. For those loci where we were able to make a comparison (see below)

information on heterozygote phenotypes was not available, and therefore the analysis is confined to homozygotes.

There are number of assumptions inherent in this analysis. First, it is not trivial to consider the direction of phenotype effect that will arise from sequence variation between N and J. While missense SNPs that are private to J or N may represent loss of function variants, this might not always be the case. SVs in the neighborhood of genes might also more often lead to loss of function effects, but other indirect effects, for example on gene regulation, might lead to gain of function. Second, knockout mutations have been generated on a variety of genetic backgrounds, often mixed, and this will confound any comparative analysis. Nevertheless, we have proceeded to compare the available phenotype terms with observed phenotype changes documented between N and J, assuming the model that private variants in N or J are likely to lead to a loss of function phenotype of the kind that would be revealed by a knockout mutation.

Of the 14 SNP and small indel variant loci with phenotypic annotations, 5 (*Crb1*, *Pdzk1*, *Pmch*, *Adcy5*, *Nlrp12*) had MP terms that overlapped with the output of the phenotype tests undertaken by EUMODIC or by secondary phenotyping (see Table 3). *Crb1* carries the *rd8* allele, a 1bp-deletion leading to a premature stop and truncated protein. CRB1 is essential for external limiting membrane integrity and photoreceptor morphogenesis in the mammalian retina [27]. The $Crb1^{rd8}$ allele leads to shortened photoreceptor inner and outer segments 2 weeks after birth and subsequent photoreceptor degeneration accompanied by regions of retinal hypopigmentation. We observed a very similar phenotype in the N mice, and in addition demonstrated significantly reduced vision in N mice compared to J using the optokinetic drum. Moreover, we found that there were significant differences in the

mean numbers of retinal veins and arteries between N and J mice. It has recently been reported that the *rd8* allele is confined to the N substrain and derived ES cells [22]. For the locus *Pdzk1*, increased circulating cholesterol levels are reported in the knockout, but we did not observe this phenotype between N and J. For three of the loci (*Pmch*, *Adcy5*, *Nlrp12*) however, we find some comparable phenotypic effects.

Pmch knockout mice display decreased circulating glucose, abnormal glucose tolerance and increased oxygen consumption. N carries a private missense variant in this gene (isoleucine to threonine). N mice display increased oxygen consumption, but higher circulating glucose levels and normal glucose tolerance compared to J. *Adcy5* knock-out mice display hypoactivity, impaired coordination, decreased vertical activity and bradykinesia. N carries a private missense variant (valine to methionine) in *Adcy5*. N mice displayed a number of behavioural changes in open field reflecting hypoactivity, including distance travelled and number of centre entries. However, these phenotype outcomes were only observed in two of the centres, with one of the centres finding opposing effects, and no changes being observed in another centre (see above). Both primary and secondary phenotyping employing the rotarod identified significantly impaired motor coordination in N mice. For both missense variants in *Pmch* and *Adcy5*, the Provean (Protein Variation Effect Analyser) predictions indicated that the changes may not have a deleterious effect on protein function [28]. NLRP12 is known to be associated with autoinflammatory disease in humans [29], and mutations in the NBS and NOD domains can cause periodic fever syndromes. *Nlrp12* knockout mice show attenuated inflammatory responses for contact hypersensitivity [30]. J mice carry a private missense variant (arginine to lysine) in *Nlrp12* that resides in a C-terminal leucine rich repeat region of the gene. However, we found that J mice show an increased response to DNFB-induced contact hypersensitivity, suggesting that the *Nlpr12* locus is not involved or, alternatively,

the missense mutation is a gain of function. Notably, most species (data not shown) carry a lysine at this position. Provean predictions indicate the mutation is not damaging to the protein.

For SVs with MP terms, 3 loci (*Chl1*, *Rptor*, *Nnt*) had MP terms that overlapped with the phenotype outputs generated in the EUMODIC pipeline (Table 3). *Chl1* knockout mice demonstrate abnormal learning and memory, including abnormal response to a novel object and spatial working memory. *Chl1* carries an intronic LINE insertion in J mice. However, N mice display impaired spatial working memory in the Morris Water Maze test in comparison to J, though it is worth noting that the poor performance of N could be explained by reduced vision that would impair ability to decipher visual reference clues. *Rptor* knockout mice demonstrate a large number of metabolic phenotypes including increased lean mass and reduced fat mass; improved glucose tolerance and decreased circulating glucose; increased oxygen consumption; and hypoactivity. *Rptor* carries an MTA insertion in J mice. J shows reduced fat mass, and increased lean mass (in two of the centres) as well as decreased circulating glucose. However, a number of phenotypes observed between N and J mice are inconsistent with a loss of function mutation at the *Rptor* locus, including poor glucose tolerance in J mice, and increased oxygen consumption and hypoactivity in N mice. Finally, J mice have been documented as carrying a large deletion at the *Nnt* locus [18] that is associated with significantly impaired glucose tolerance and this phenotype was confirmed in our N vs J comparison. It is worth noting that given the expected strong effects of the *Nnt* locus on glucose tolerance, predicted effects from mutations at other loci on glucose tolerance may be overridden, and that *Nnt* will be epistatic to other loci. So for example, as we discuss above, while *Rptor* knockout mice show improved glucose tolerance, we find that in J mice that carry an intronic MTA insertion in the *Rptor* gene there is poor glucose tolerance. This may reflect the

overriding effect of the *Nnt* deletion on glucose regulation, or alternatively that the MTA insertion at *Rptor* has no effect on gene function.

Conclusions

Functional analysis of the genetic mechanisms that underlie phenotypic traits in mouse mutants may be influenced, often profoundly, by genetic variation between individual inbred strains. For the first time, we have undertaken an analysis to detect and verify sequence variants between the two widely used mouse strains; C57BL/6N and C57BL6J. Using deep sequence data and comprehensive detection methods we have validated 51 coding variants, 34 coding SNPs, 2 indels and 15 SVs, differentiating C57BL/6N and C57BL/6J.

At the same time we have carried out a comprehensive phenotypic comparison of the two inbred strains and identified a considerable number of significant phenotype differences. While a direct analysis of the relationship between genomic variants and phenotypes are beyond this study, we have thoroughly examined the landscape of phenotypic differences between the two strains and related these where possible to the known functions of genes that are variant. The comparative examination of the phenotypic terms associated with knockout mutations and phenotype changes between N and J reveals some concordance and some discordance. These analyses are confounded by several factors including the genetic background of the knockout mice and assumptions regarding the direction of phenotype effect of variants discovered between N and J. In addition, for many SNPs and SVs there may be little or no phenotype effect. However, our findings suggest a number of variants and loci that will merit further investigation by exploring the linkage between variant segregation and phenotype in N/J intercrosses. Moreover, N/J intercrosses would

enable the identification of genetic loci underlying the many other phenotype differences between N and J, and allow us to explore the potential functional consequences of coding variation at the majority of loci for which there is as yet no functional annotation.

Materials and Methods

Sequencing and genomic analyses

Full details on the mouse strains (C57BL/6J and C57BL/6N) used for sequencing and validation are documented in Keane et al., 2011[11].

SNP and small indel identification

The paired-end alignment of C57BL/6N against the reference genome (C57BL/6J - also known as mm9/NCBIM37) [11] was used to find SNPs and small indels differentiating the two sub-strains. The raw sequence variant calls were made using the Genome Analysis Toolkit (GATK) [12] with default parameters. We adopted a filtering strategy to reduce the number of false positives and lessen the burden for validation. SNP sites that occur in the Broad J alignment, had an allele ratio <0.8 or located in a region of <3 or >150 read depth were removed from further analysis. The C57BL/6N BAM was realigned for calling small indels and the above filtering procedure adopted. Annotation of the variants was performed with Annovar [14] and/or NGS-SNP [15]. Using these annotations manual inspection of the coding variants (non-synonymous, splice donor-acceptor or frameshift sites) removed sites embedded in homopolymers and GC rich regions. The remaining coding variants and a subset of the non-coding variants were sent for Sequenom validation.

SNP and small indel validation

We designed extension and amplification primers for 762 SNPs and 169 small indels using SpectroDESIGNER. Oligonucleotides were synthesized at Metabion (Martinsried, Germany). We used the iPLEX GOLD assay of the Sequenom MassARRAY platform for genotyping these variants in eight DNA samples from three C57BL/6 substrains (replicate of four C57BL/6J and replicate of two C57BL/6NJ and

C57Bl/6NTac) and SpectroTYPER for data analysis. The resulting genotypes were then downloaded and checked for consistency in the four replicates. Inconsistent or heterozygous genotypes in either the C57BL/6J or C57BL/6N samples were discarded from further analyses.

In addition to Sequenom we used Pyrosequencing and traditional Sanger sequencing for validation. We designed primers for 22 SNPs and 10 small indels. Primers were designed with Pyrosequencing™Assay Design and oligonucleotides were synthesized at Eurofins MWG Operon. The PCR was performed using Qiagen Taq Mastermix. Samples were sent to GATC BioTech for sequencing and pyrosequencing was carried out on the PSQ 96H Pyrosequencer. In cases with insufficient DNA or poor primer design the SNPs or small indels were omitted from any further analysis.

C57BL/6N alignment files are available at the Mouse Genomes project [11]. Coding SNPs and indels are available from dbSNP [31, 32].

SV identification

Using a combination of four computational methods, as described elsewhere [16], we detected a total of 551 genome-wide structural variants (SVs) between C57BL/6J and C57BL/6N. We then visually inspected short-read sequencing data at each of these 551 unique sites using LookSeq [33], and found that 470 predicted sites were false due to paired-end mapping errors. At the remaining 81 sites, we carried out PCR and Sanger-based sequencing analyses as described below.

SV validation

Primers were designed using Primer3 and purchased from MWG (Ebersberg, Germany. Primer design strategy was dependent on the type and size of the structural variant. Three independent PCR reactions were carried out with Hotstar Taq obtained from Qiagen (Hilden, Germany). These reactions were performed as previously described [34]. LongRange PCR kit (Qiagen) was used for genomic regions >2 kbp. PCR gel images were then assessed for quality of primer design and performance of PCR reaction. PCR products were then purified in a 96-well Millipore purification plate, resuspended in 30 µl of H_2O and sequenced. All sequencing reactions were run out on an ABI3700 sequencer and assembled by using PHRED/PHRAP. PCR and Sanger-based sequencing analyses at the 81 retained sites allowed us to further remove 38 sites confirmed to not be polymorphic between C57BL/6J and C57BL/6N because of reference errors. The structural variants are available from the Database of Genomic Variants Archive (DGVa) (accession ID:estd204) [35].

Predicted effects of sequence variants

Predicted effects of amino acid substitutions on their respective proteins were performed using Provean (Protein Variation Effect Analyzer) [28]. Mammalian Phenotype (MP) ontology terms for genes containing SVs or SNPs was obtained from The Mouse Genome Database (MGD) [36].

Phenotyping

Phenotyping platforms in EMPReSSslim

All SOPs for phenotyping procedures are described within the EMPReSS database [37]. All mice were analysed through the complete phenotyping pipeline (excluding FACs and Ig analyses which were not undertaken). Subsequently subsets of mice were selected according to appropriate metadata considerations to ensure robust comparisons between strains (see *Phenotype Data Analysis* below). Unless otherwise stated C57BL/6NTac was used throughout the study. Secondary phenotyping protocols can be found at[37].

Data capture by Europhenome

Data generated from EMPReSSslim by the four centres are stored in their local Laboratory Information Management Systems (LIMS), backed by diverse database schemas running on different relational database management systems. The data is transferred to EuroPhenome in a common format Phenotype Data Markup Language (PDML), an extension of eXtensible Markup Language (XML) defined by XML schema. To assist in data export and improve standardization and data consistency we have provided a java library [38] for data export. The informaticians at the centres use this to represent the data to be exported as an object model. The library then performs the necessary validation against the European Mouse Phenotyping Resource for Standardized Screens (EMPReSS) database [37]. If this is successful the data are output to XML, compressed and placed on a file transfer protocol (FTP) site.

Each centre's FTP site is regularly checked by the EuroPhenome data capture system and any new files are uploaded. The data is again verified against the schema and EMPReSS, and further checked for consistency against existing data within EuroPhenome. The results of the upload and validation are provided to the sites in

the form of XML log files and a web interface, the EuroPhenome Tracker. If validation is successful the data is loaded into the EuroPhenome database. Data can be removed from the database by placing the files in the delete directory of the FTP site. The same process is employed to capture and validate the data prior to removal. Phenotype data may be downloaded from MRC Harwell [39].

Phenotype data analysis

Statistical analysis was carried out in a manner consistent with the Europhenome data repository [19]. In order to compare phenotype data from N and J, groups of comparable measurements from each centre were extracted from the Europhenome database. In a few instances there is more than one dataset from each centre that differs according to a critical metadata difference. Where there is more than one comparable group for a parameter, the largest one has been used. An example is for the simplified IPGTT procedure at the Helmholtz Institute where the 'type of strip' was changed from 'accu-check aviva' to 'roche'. This resulted in two sets, one of sizes (20,22) and one of (20,13), for C57BL/6N and C57BL/6J respectively. The set with the largest minimum value within the set is used: the (20,22) set. Once comparable groups of C57BL/6J and C57BL/6N have been identified, statistical tests are applied separately to male and female groups for each centre. A dataset of each group's sizes, means, standard deviations, effect sizes and the resultant p-values from the statistical test is then created for subsequent hypothesis testing during the creation of the heat maps.

Categorical data uses the Fisher's Exact Test (for 2x2 contingency tables) or a Chi-squared Test to produce a p-value. Numerical data has the Mann-Whitney U Test applied, as this is a non-parametric test suitable for all types of unimodal distributions. Two dimensional data (such as parameters where a measurement is

taken over a time course) are averaged into a single mean, which loses information but still gives an overall comparable value for that parameter.

In order to create a multi-centre heat map of statistical significances, a colour and shade is selected for each parameter/site/sex combination, with a lower p-value (therefore indicating a higher confidence in the putative difference between the strains) resulting in a darker colour. Where C57BL/6N values are higher than C57BL/6J a red colour is used, with green used for the opposite case. In categorical fields where there are no numerical values a blue colour is used to indicate a difference between the strains with no order.

In order to calculate the false positive rate when significant results are found across multiple sites, a bootstrapping re-sampling technique [40] was used to estimate the probabilities of a parameter revealing similar trends in the same direction across 3 or more centres. All heatmap squares were randomised and the number of times three or more sites had squares of the same colour was recorded. Repeated many times, and divided by the number of parameters and number of repeats, this provides a probability of that event occurring at random within the given heatmap. Comparison of this probability with the observed probability allows us to assess whether we are looking at a purely random effect, or if there is underlying structure within the data. Analysis of the class of parameters with 3 or more centres demonstrating trends in the same direction (Figure 1a) indicates that a randomized sample would show this in 0.074 of cases. This is compared to the observed rate of 0.213 indicating that this class is over-represented with respect to random. We applied the same analysis to the class of results that showed contradictory trends between centres (Figure 2). A randomized sample indicates that we would observe this pattern in 0.388 of cases, while the observed rate is 0.102 thus showing an underrepresentation and again

giving confidence in these results. The class of results demonstrating similar trends in 2 but not more centres (Figure 1b) gave similar probabilities for random (0.183) and observed (0.181).

Abbreviations

EMPReSS: European Mouse Phenotyping Resource of Standardised Screen; ES cells: Embryonic Stem cells; EUCOMM: European Conditional Mouse Mutagenesis project; FACS: Fluorescence-Activated Cell Sorting; IKMC: International Knockout Mouse Consortium; IMPC: International Mouse Phenotyping Consortium; IPGTT: Intraperitoneal Glucose Tolerance Test; KOMP: Knockout Mouse Project; MP: Mammalian Phenotype; SNP(s): Single Nucleotide Polymorphism(s); SOP: Standard Operating Procedure

Competing interests

The authors have no conflicts of interest.

Authors' contributions

MMS, SG, JKW, HF, VG-D, TS, KW, EB, EJC , SD, JE, JG, NJI, IJJ, AL, SM, JM, HMe, FP, OP, MR, PJ, BY, RR-S, KPS, A-MM, MHA, YH and SDMB contributed equally, conceived and designed the experiments and wrote the paper.
DJA, BP, JKW, HF, VG-D, TS, EB, EJC , SD, JE, JG, NJI, IJJ, AL, SM, JM, HMe, FP, OP, MR, PJ, KPS, YH, RD, SA, AA, LB, DB, HC, M-FC, RC, AF, A-KG, EG, WH, SMH, TH, WM, FN, L-AR, , JR, MSa, MSe, CS, AS, GT-V, VEV, SW, WW and MZ performed the experiments.

DJA, BP, MMS, SG, A-MM, SDMB, AB, PD, JMH, TMK, HMo, GN, LS and HW performed statistical analysis and analysed the data.

All authors read and approved the final manuscript.

Description of additional data files

Additional file 1: Table S1, SNP and small indel validation numbers

Additional file 2: Figure S1, the EMPReSSslim phenotyping pipeline.

Additional file 3: Figures S2a-h, heat maps showing phenotyping parameter differences between the phenotyping centres.

Acknowledgements

The EUMODIC project was funded by the European Commission contract numbers LSHG-CT-2006-037188 and LSHG-CT-2007-037445 and also supported by the Wellcome Trust (grant number 098051). We gratefully acknowledge the contribution of all contributors to the EUMODIC project; a list can be found at [19]. We thank Genomics-Services group at the Wellcome Trust Centre for Human Genetics, University of Oxford, for running the Sequenom assays. Binnaz Yalcin is supported by an EMBO Long-Term Fellowship and the Agence Nationale de la Recherche. The ICS was also supported by the French state aid managed by the National Agency for Research under the program of future investments (PHENOMIN, ANR-10-INBS-07).

References

1. Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A: **A conditional knockout resource for the genome-wide study of mouse gene function**. *Nature* 2011, **474**:337-342.

2. Brown SD, Moore MW: **Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium**. *Dis Model Mech* 2012, **5**:289-292.

3. Carneiro AM, Airey DC, Thompson B, Zhu CB, Lu L, Chesler EJ, Erikson KM, Blakely RD: **Functional coding variation in recombinant inbred mouse lines reveals multiple serotonin transporter-associated phenotypes**. *Proc Natl Acad Sci U S A* 2009, **106**:2047-2052.

4. Williams RW, Gu J, Qi S, Lu L: **The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis**. *Genome Biol* 2001, **2**:RESEARCH0046.

5. Gregorova S, Divina P, Storchova R, Trachtulec Z, Fotopulosova V, Svenson KL, Donahue LR, Paigen B, Forejt J: **Mouse consomic strains: exploiting genetic divergence between Mus m. musculus and Mus m. domesticus subspecies**. *Genome Res* 2008, **18**:509-515.

6. Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JN, Mott R, Flint J: **Genome-wide genetic association of complex traits in heterogeneous stock mice**. *Nat Genet* 2006, **38**:879-887.

7. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbe JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Farber CR, Forejt J, Gaile D, Garlow SJ *et al*: **The Collaborative Cross, a community resource for the genetic analysis of complex traits**. *Nat Genet* 2004, **36**:1133-1137.

8. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P *et al*: **Initial sequencing and comparative analysis of the mouse genome**. *Nature* 2002, **420**:520-562.

9. Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, She X, Bult CJ, Agarwala R, Cherry JL, DiCuccio M, Hlavina W, Kapustin Y, Meric P, Maglott D, Birtle Z, Marques AC, Graves T, Zhou S, Teague B, Potamousis K, Churas C, Place M, Herschleb J, Runnheim R, Forrest D, Amos-Landgraf J, Schwartz DC, Cheng Z,

Lindblad-Toh K, Eichler EE *et al*: **Lineage-specific biology revealed by a finished genome assembly of the mouse**. *PLoS Biol* 2009, **7**:e1000112.

10. Mekada K, Abe K, Murakami A, Nakamura S, Nakata H, Moriwaki K, Obata Y, Yoshiki A: **Genetic differences among C57BL/6 substrains**. *Exp Anim* 2009, **58**:141-149.

11. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, Furlotte NA, Eskin E, Nellaker C, Whitley H, Cleak J, Janowitz D, Hernandez-Pliego P, Edwards A, Belgard TG, Oliver PL, McIntyre RE, Bhomra A, Nicod J, Gan X, Yuan W, van der Weyden L, Steward CA, Bala S, Stalker J, Mott R *et al*: **Mouse genomic variation and its effect on phenotypes and gene regulation**. *Nature* 2011, **477**:289-294.

12. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ: **A framework for variation discovery and genotyping using next-generation DNA sequencing data**. *Nat Genet* 2011, **43**:491-498.

13. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB: **High-quality draft assemblies of mammalian genomes from massively parallel sequence data**. *Proc Natl Acad Sci U S A* 2011, **108**:1513-1518.

14. Wang K, Li M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data**. *Nucleic Acids Res* 2010, **38**:e164.

15. Grant JR, Arantes AS, Liao X, Stothard P: **In-depth annotation of SNPs arising from resequencing projects using NGS-SNP**. *Bioinformatics* 2011, **27**:2300-2301.

16. Wong K, Keane TM, Stalker J, Adams DJ: **Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly**. *Genome Biol* 2010, **11**:R128.

17. Yalcin B, Wong K, Bhomra A, Goodson M, Keane TM, Adams DJ, Flint J: **The fine-scale architecture of structural variants in 17 mouse genomes**. *Genome Biol* 2012, **13**:R18.

18. Freeman H, Shimomura K, Horner E, Cox RD, Ashcroft FM: **Nicotinamide nucleotide transhydrogenase: a key role in insulin secretion**. *Cell metabolism* 2006, **3**:35-45.

19. Morgan H, Beck T, Blake A, Gates H, Adams N, Debouzy G, Leblanc S, Lengger C, Maier H, Melvin D, Meziane H, Richardson D, Wells S, White J, Wood J, de Angelis MH, Brown SD, Hancock JM, Mallon AM: **EuroPhenome: a repository for high-throughput mouse phenotyping data**. *Nucleic Acids Res* 2010, **38**:D577-585.

20. Gates H, Mallon AM, Brown SD: **High-throughput mouse phenotyping**. *Methods* 2011, **53**:394-404.

21. Prusky GT, Alam NM, Beekman S, Douglas RM: **Rapid quantification of adult and developing mouse spatial vision using a virtual optomotor system**. *Invest Ophthalmol Vis Sci* 2004, **45**:4611-4616.

22. Mattapallil MJ, Wawrousek EF, Chan CC, Zhao H, Roychoudhury J, Ferguson TA, Caspi RR: **The Rd8 mutation of the Crb1 gene is present in vendor lines of C57BL/6N mice and embryonic stem cells, and confounds ocular induced mutant phenotypes**. *Invest Ophthalmol Vis Sci* 2012, **53**:2921-2927.

23. Paques M, Guyomard JL, Simonutti M, Roux MJ, Picaud S, Legargasson JF, Sahel JA: **Panretinal, high-resolution color photography of the mouse fundus**. *Invest Ophthalmol Vis Sci* 2007, **48**:2769-2774.

24. Matsuo N, Takao K, Nakanishi K, Yamasaki N, Tanda K, Miyakawa T: **Behavioral profiles of three C57BL/6 substrains**. *Front Behav Neurosci* 2010, **4**:29.

25. Tucci V, Lad HV, Parker A, Polley S, Brown SD, Nolan PM: **Gene-environment interactions differentially affect mouse strain behavioral parameters**. *Mamm Genome* 2006, **17**:1113-1120.

26. Grenham S, Clarke G, Cryan JF, Dinan TG: **Brain-gut-microbe communication in health and disease**. *Front Physiol* 2011, **2**:94.

27. Mehalow AK, Kameya S, Smith RS, Hawes NL, Denegre JM, Young JA, Bechtold L, Haider NB, Tepass U, Heckenlively JR, Chang B, Naggert JK, Nishina PM: **CRB1 is essential for external limiting membrane integrity and photoreceptor morphogenesis in the mammalian retina**. *Human molecular genetics* 2003, **12**:2179-2189.

28. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP: **Predicting the functional effect of amino Acid substitutions and indels**. *PLoS One* 2012, **7**:e46688.

29. Jeru I, Duquesnoy P, Fernandes-Alnemri T, Cochet E, Yu JW, Lackmy-Port-Lis M, Grimprel E, Landman-Parker J, Hentgen V, Marlin S, McElreavey K, Sarkisian T, Grateau G, Alnemri ES, Amselem S: **Mutations in NALP12 cause hereditary periodic fever syndromes**. *Proc Natl Acad Sci U S A* 2008, **105**:1614-1619.

30. Arthur JC, Lich JD, Ye Z, Allen IC, Gris D, Wilson JE, Schneider M, Roney KE, O'Connor BP, Moore CB, Morrison A, Sutterwala FS, Bertin J, Koller BH, Liu Z, Ting JP: **Cutting edge: NLRP12 controls dendritic and myeloid cell migration to affect contact hypersensitivity**. *J Immunol* 2010, **185**:4515-4519.

31. **SNP & Indel Data:** http://www.ncbi.nlm.nih.gov/projects/SNP/snp_viewTable.cgi?handle=MRCHARWELL

32. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K: **dbSNP: the NCBI database of genetic variation**. *Nucleic Acids Res* 2001, **29**:308-311.

33. Manske HM, Kwiatkowski DP: **LookSeq: a browser-based viewer for deep sequencing data**. *Genome Res* 2009, **19**:2125-2132.

34. Yalcin B, Willis-Owen SA, Fullerton J, Meesaq A, Deacon RM, Rawlins JN, Copley RR, Morris AP, Flint J, Mott R: **Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice**. *Nat Genet* 2004, **36**:1197-1202.

35. **Structural Variant Data:** http://www.ebi.ac.uk/dgva/page.php

36. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE: **The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse**. *Nucleic Acids Res* 2012, **40**:D881-886.

37. Mallon AM, Blake A, Hancock JM: **EuroPhenome and EMPReSS: online mouse phenotyping resource**. *Nucleic Acids Res* 2008, **36**:D715-718.

38. **Europhenome Java Library:** http://sourceforge.net/projects/europhenome/

39. **Phenotype Data:** http://www.har.mrc.ac.uk/nj

40. Efron B: **1977 Rietz Lecture - Bootstrap Methods - Another Look at the Jackknife**. *Ann Stat* 1979, **7**:1-26.

Figure 1. Heat maps illustrating significant differences in phenotype parameters between C57BL/6N and C57BL/6J male and female mice for each of the four centres – Helmholtz Zentrum Munchen (HMGU), Institut Clinique Souris (ICS), MRC Harwell and the Wellcome Trust Sanger Institute (WTSI). Parameter designations and parameter descriptions are from EMPReSSslim [37]. Significance levels and the direction of the effect (red and green) are defined in the key. Significant differences for categorical data are illustrated in blue. **A)** Phenotype parameters that show a significant difference between N and J in 3 or more centres. **B)** Phenotype parameters that show a significant difference between N and J in 2 centres but no evidence of trends in the other centres.

Figure 2. Heat maps illustrating significant differences in phenotype parameters between C57BL/6N and C57BL/6J male and female mice for each of the four centres – Helmholtz Zentrum Munchen (HMGU), Institut Clinique Souris (ICS), MRC Harwell and the Wellcome Trust Sanger Institute (WTSI) – and for which phenotype parameters showed significant differences in 2 or more centres, but the opposite trend in one of the centres. Parameter designations and parameter descriptions are from EMPReSSslim. Significance levels and the direction of the effect (red and green) are defined in the key.

Figure 3. Morphological and functional differences between C57BL/6N and C57BL/6J mice eyes. **(A)** While white flecks were absent from C57BL/6J fundus, they were frequently detected in the ventral retina from C57BL/6N mice, with various degree of severity, as illustrated here by three C57BL/6N fundus images. Depending on the centre, C57BL/6N mice with at least one eye affected represented 69.2% (n = 70 males + 34 females, ICS), 44.6% (n = 145 males + 158 females, GMC) or 23.0% (n = 184 males + 194 females, WTSI) of the population, while no fleck was detected in

C57BL/6J mice (ICS: 29 males; GMC: 75 males + 75 females; WTSI: 34 males + 28 females). **(B)** Both vein and artery numbers differed from mouse to mouse in both strains, usually between 3 (left) and 7 (right), with a mean around 5 (middle), as can be seen in three fundus images from C57BL/6J mice. **(C)** Quantification of veins and arteries in male C57BL/6N and C57BL/6J mice (n = 140 and n = 70 eyes, respectively). The mean number of veins per eye was 4.8 ± 0.1 for C57BL/6N (n = 122 eyes) vs. 5.3 ± 0.1 for C57BL/6J mice (n = 138 eyes). Both differences were statistically significant ($p < 0.001$, Mann-Whitney Rank Sum Test).

Figure 4. µCT analysis of distal femur revealed similar trabecular bone parameters in 14-week-old C57BL/6J vs C57BL/6N males **(A)** and females **(B)**. Cortical bone parameters from 14-week-old male midshaft femur were also unchanged between the two strains **(C)**. Measurement of serum osteocalcin and urinary deoxypyridinoline, bone formation and bone resorption markers respectively, indicate that bone turnover is identical between 14-week-old C57BL/6J and C57BL/6N **(D)**. Abbreviations: B.V./T.V., bone volume/tissue volume; Tb.N., trabecular number; Tb.Sp., Trabecular spacing; Conn-Dens., Connectivity density; S.M.I., structural model index (0 for parallel plates, 3 for cylindrical rods); D.A., degree of anisotropy; Ct.Po., cortical porosity; Ct.Th., cortical thickness; DPD, deoxypyridinoline; creat., creatinin.

Figure 5. Light dark test. Bars represent the latency to enter **(A)**, the % of time spent in the dark compartment **(B)** and the number of light/dark transitions **(C)** in C57BL/6J (n=10) and C57BL/6N (n=9) male mice, age 8-10 weeks. Data are mean ± SEM, * $p<0.05$ (t-test). **(D)** Rotarod motor learning performance over four days. Symbols and lines represent mean (± SEM) daily latencies to fall from rotating rod at acceleration from 4 to 40 r.p.m. in 300 sec of C57BL/6J (n=10) and C57BL/6N

(n=10) male mice, age 9-11 weeks. *p<0.05, **p<0.005, ***p<0.0001 J vs. N, t-test; *p<0.05, **p<0.005, ***p<0.0001 vs. Day 1, Fisher's LSD.

Figure 6. Morris Water Maze. **A)** Familiarization and Training phase learning curves. Symbols and lines represent mean (± SEM) daily latencies to reach the platform of C57BL/6J (n=10) and C57BL/6N (n=10) male mice, age 16-20 weeks. **p<0.005, ***p<0.0005 J vs. N, t-test; **p<0.005, ***p<0.0001 vs. Day 1, Fisher's LSD. **B)** Probe test. Bars represent % time spent in each quadrant on day 5 during probe test. Dotted line is set at chance level (25%). *p<0.05, **p<0.005, ***p<0.0005 vs. Correct quadrant, t-test. **C**. Representative tracks of two C57BL/6J and C57BL/6N mice paths during probe test. Dotted circle indicates former platform location.

Figure 7. Comparison of Listeria host resistance between C57BL/6J and C57BL/6N inbred strains. **A)** Kaplan-Meier survival curves of females and males of the C57BL/6J and C57BL/6N strains after i.v. infection with 2×10^4 colony-forming units (cfu) of *Listeria monocytogenes* strain EGD. **B)** Bacterial load in liver and spleen of C57BL/6J and C57BL/6N mice after i.v. infection with 2×10^4 cfu *L. monocytogenes* EGD. Organ loads were ascertained at four time points to analyze kinetics of bacterial growth. **C)** Comparison of plasma levels of interleukin 6 (IL-6), interferon inducible protein 10 (IP-10), and chemokine ligand 2 (CCL2) between the C57BL/6J and C57BL/6N mice shown in **(B)**. Concentrations of pro-inflammatory cytokines and chemokines were determined in peripheral blood samples using the Cytokine Mouse 20-Plex Panel (Invitrogen) and a LiquiChip 100 system (Qiagen). Significant differences are indicated as follows: * p<0.05, ** p<0.01 Mann-Whitney, U-test. Black bars and symbols, C57BL/6J inbred strain. White bars and symbols, C57BL/6N inbred strain.

Figure 8. A) Measurement of splenic Natural Killer (NK) cells activity of C57BL/6J (B6J) vs C57BL/6N (B6NTac) male (upper panel) and female (lower panel) mice. Splenic NK cells from C57BL/6J or C57BL/6N mice were stimulated under the indicated conditions (six mice per group). Mean (+/- SD) of IFNγ positive cells among CD3- NK1.1+ NK cells was measured by flow cytometry .

(**B**) Hapten-specific hypersensitivity.

Male or female C57BL/6J or C57BL/6N mice were sensitized by the application of 25 µl of 0.5% DNFB solution on the ventral skin. They were then challenged by the application of 5 µl of 0.15% DNFB solution on the left ear 5 days later (DNFB). Right ears were painted with vehicle (-) and used as controls. Ear thickness was measured 48h after challenge. Results representative of 3 independent experiments with six mice per group. *, $p < 0.05$; **, $p < 0.005$ (Mann-Whitney test)

Figure S1. EMPReSSslim phenotyping pipeline. The pipeline includes 20 phenotyping platforms. Data for FACs analysis of peripheral blood populations were not acquired for all centres and are not presented here.

Figure S2. A-D) Heat maps (see Figure 1 and Figure 2) displayed with numbers of C57BL/6N and C57BL/6J animals analysed for each test in each centre **E-H)** Heat maps (see Figure 1 and Figure 2) showing the effect sizes observed in each test in each centre. **A,E)** Phenotype parameters that show a significant difference between N and J in 3 or more centres. **B,F)** Phenotype parameters that show a significant difference between N and J in 2 centres but no evidence of trends in the other centres. **C,G)** Phenotype parameters for which we did not observe any significant differences across the centres. **D,H)** Phenotype parameters that showed significant differences in 2 or more centres, but the opposite trend in one of the centres.

Table 1. Coding SNPs and small indels identified between C57BL/6N and C57BL/6J

Chr	Position	B6J Base	B6N Base	Strain	Gene Name	B6J Amino Acid	B6N Amino Acid
Nonsense Polymorphism							
13	65023280	C	T	B6N	*Spata31*	Arginine	* (STOP)
Missense Polymorphisms							
1	59904011	G	A	B6N	*Bmpr2*	Arginine	Glutamine
3	95538799	T	C	B6J	*Ecm1*	Isoleucine	Valine
3	96658480	A	G	B6J	*Pdzk1*	Asparagine	Aspartic acid
4	21800831	C	G	B6J	*Sfrs18*	Arginine	Glycine
4	137777588	C	T	B6N	*Hp1bp3*	Leucine	Phenylalanine
4	140354038	A	G	B6N	*Padi3*	Leucine	Proline
4	148318468	T	C	B6J	*Casz1*	Leucine	Proline
5	90204376	C	T	B6N	*Adamts3*	Valine	Isoleucine
5	97187161	T	C	B6J	*Fras1*	Leucine	Proline
5	113191741	C	T	B6N	*Myo18b*	Arginine	Histidine
6	39350455	T	A	B6J	*Mkrn1*	Asparagine	Tyrosine
7	3222538	T	C	B6J	*Nlrp12*	Lysine	Arginine
7	63386662	G	A	B6J	*Herc2*	Glycine	Aspartic acid
7	86256240	A	C	B6J	*Acan*	Histidine	Proline
7	110121823	C	T	B6N	*Olfr577*	Valine	Isoleucine
7	127278693	G	A	B6N+Spretus	*Zp2*	Alanine	Valine
7	129311164	C	T	B6N	*Plk1*	Arginine	Tryptophan
9	24935069	C	G	B6N	*Herpud2*	Valine	Leucine
10	66700922	T	C	B6J	*Jmjd1c*	Leucine	Proline
10	78632222	A	G	B6N	*Vmn2r80*	Asparagine	Serine
10	87554578	T	C	B6N	*Pmch*	Isoleucine	Threonine
11	46036117	G	A	B6N	*Cyfip2*	Serine	Phenylalanine
11	90341985	C	T	B6N	*Stxb4*	Alanine	Threonine
13	21560172	A	G	B6J	*Nkapl*	Glycine	Arginine
13	73465884	A	G	B6J	*Ndufs6*	Valine	Alanine
13	93833534	C	G	B6J	*Cmya5*	Alanine	Proline
14	70986011	G	T	B6N	*Fam160b2*	Serine	Arginine
15	11266138	G	T	B6N	*Adamts12*	Cysteine	Phenylalanine
15	77468437	A	C	B6J	*Apol11b*	Isoleucine	Arginine
16	35291630	G	A	B6N	*Adcy5*	Valine	Methionine
17	47537359	T	C	B6J	*Guca1a*	Isoleucine	Valine
X	131227581	C	A	B6N	*Armcx4*	Alanine	Aspartic acid
Splice Site Polymorphism							
5	54280548	A	G	B6J	*Tbc1d19*	-	-
Frameshift 1bp-Deletions							
1	141133664	G	-	B6N	*Crb1*	-	-
9	65127938	G	-	B6J	*Cilp*	-	-

Table 2. Structural variants between C57BL/6N and C57BL/6J

Chr	SV start	SV stop	Ancestral event	Strain	Gene	Overlap
1	149518394	149524878	LINE Ins	B6J		
2	7325700	7330977	IAP Ins	B6J		
2	70619835	70620080	SINE Ins	B6J	*Tlk1*	intron
3	77975065	77977953	Del	B6N		
3	5049018	5055845	LINE Ins	B6J		
3	60336036	60336037	Del (large)	B6J	*Mbnl1*	intron
3	41885819	41887255	LINE Ins	B6J		
3	18484710	18484889	Del	B6N		
4	101954274	101954395	Del	B6N	*Pde4b*	intron
4	116051393	116051799	MaLR Ins	B6J	*Mast2*	intron
5	46376307	46377852	LINE Ins	B6J		
5	90356490	90356491	Del (~300 bp)	B6J		
5	146248861	146261885	Ins	B6J+others		
6	18112291	18119019	LINE Ins	B6J		
6	62964974	62972907	LINE Ins	B6J		
6	86478779	86479400	Ins	B6J		
6	103669536	103676487	LINE Ins	B6J	*Chl1*	intron
6	104207081	104214434	LINE Ins	B6J		
7	92095990	92096149	Del	B6N	*Vmn2r65*	exon
7	27636128	27748456	Ins	B6J	*Cyp2a22*	entire
7	100892501	100899058	LINE Ins	B6J		
7	139306094	139307981	MaLR Ins	B6J	*Cpxm2*	intron
8	16716381	16716382	Del (large)	B6J	*Csmd1*	intron
9	25674550	25674770	SINE Ins	B6J		
9	58544415	58546304	MaLR Ins	B6J	*2410076I21Rik*	intron
10	3039196	3039197	Del (large)	B6J		
10	29339441	29345955	LINE Ins	B6J		
10	32536420	32543464	LINE Ins	B6J	*Nkain2*	intron
10	49543303	49550645	LINE Ins	B6J		
11	104906390	104906621	Del	B6N		
11	119560391	119566827	MTA Ins	B6J	*Rptor*	intron
12	42023964	42032747	Del	B6N	*Immp2l*	intron
13	71224557	71231011	MTA Ins	B6J		
13	120164268	120164269	Del (large)	B6J	*Nnt*	exon
14	112825585	112832341	LINE Ins	B6J		
15	49554596	49554597	Ins (large)	B6N		
15	31106173	31106382	VNTR			
16	6115804	6138105	Del	B6N		
17	60286367	60286368	Ins (~2000 bp)	B6N		
18	4809271	4809272	Del (~1200 bp)	B6J		
19	12863187	12863188	Del (~1800 bp)	B6J	*Zfp91*	intron
X	15697909	15697910	Del (~400 bp)	B6J		
X	95155499	95163160	LINE Ins	B6J		

Start and stop coordinates are given for MGSCv37 of the mouse reference genome. Del, deletion; IAP, intracisternal A particle; Ins, insertion; LINE, long interspersed nuclear elements; MaLR, mammalian apparent LTR-retrotransposon; MTA, member of transcript retrotransposon; SINE, short interspersed nuclear elements; VNTR, variable number tandem repeat.

Table 3. Comparison of predicted effects of SNPs and SVs that might contribute to the phenotypic differences between C57BL/6N and C57BL/6J (see text). We have identified variant genes that show homozygote knockout phenotypes with associated MP terms that were assessed in the phenotyping pipeline, and compared these phenotypes to those observed between N and J.

Protein coding gene	C57BL/6J Amino Acid	C57BL/6N Amino Acid	SNP is Private to	PROVEAN Prediction[1]	MP Terms	B6J vs B6N[2]	B6N vs B6J[2]
Adcy5	Valine (V)	Methionine (M)	B6N	TOLERATED (-1.712)	impaired coordination_MP:0001405	NR	P
					hypoactivity_MP:0001402	NR	P
Pmch	Isoleucine (I)	Threonine (T)	B6N	TOLERATED (0.493)	decreased circulating glucose level_MP:0005560	NR	A
					abnormal glucose tolerance_MP:0005291	NR	A
					increased oxygen consumption_MP:0005289	NR	P
Pdzk1	Asparagine (N)	Aspartic Acid (D)	B6J	TOLERATED (0.95)	increased circulating cholesterol level_MP:0001556	A	NR
Nlrp12	Lysine (K)	Arginine (A)	B6J	TOLERATED (0.781)	abnormal type IV hypersensitivity reaction_MP:0002534	P	NR
Crb1	-	-	B6N	-	photosensitivity_MP:0001999	NR	P
					abnormal ocular fundus morphology_MP:0002864	NR	P
					retinal degeneration_MP:0001326	NR	P
					abnormal retina morphology_MP:0001325	NR	P
					abnormal retinal photoreceptor layer_MP:0003728	NR	P
Chl1	-	-	B6J	-	abnormal learning/ memory_ MP:0001449	A	NR
					abnormal spatial working memory_MP:0008428	A	NR
Rptor	-	-	B6J	-	increased lean body mass_MP:0003960	P	NR
					increased oxygen consumption_MP:0005289	A	NR
					hypoactivity_MP:0001402	A	NR
					decreased circulating glucose level_MP:0005560	P	NR
					improved glucose tolerance_MP:0005292	A	NR
Nnt	-	-	B6J	-	impaired glucose tolerance_MP:0005293	P	NR

[1]Threshold for intolerance is -2.3

[2]These columns indicate the direction of the phenotype effect that might be observed given the assignment of a SNP or SV as private to B6J or B6N. Only one direction will be relevant and comparable to the effects of the knockout mutation.

Key:

NR	Not relevant
P	Phenotype present
A	Phenotype ant

A

Parameter	Description	HMGU M	HMGU F	ICS M	ICS F	MRC Harwell M	MRC Harwell F	WTSI M	WTSI F
ESLIM_002_001_002	Non-Invasive blood pressure: Systolic arterial pressure								
ESLIM_002_001_003	Non-Invasive blood pressure: Pulse rate								
ESLIM_003_001_003	Calorimetry: Oxygen consumption								
ESLIM_003_001_004	Calorimetry: Carbon dioxide production								
ESLIM_003_001_006	Calorimetry: Heat production (metabolic rate)								
ESLIM_004_001_002	Simplified IPGTT: Blood glucose concentration								
ESLIM_004_001_701*	Simplified IPGTT: Glucose response AUC								
ESLIM_005_001_002	DEXA: Fat mass								
ESLIM_008_001_008	Modified SHIRPA: Locomotor activity								
ESLIM_008_001_013	Modified SHIRPA: Startle response								
ESLIM_009_001_001	Grip-strength: Forelimb grip strength measurement								
ESLIM_009_001_701*	Grip-strength: Forelimb grip strength measurement mean								
ESLIM_010_001_001	Rotarod: Latency to fall								
ESLIM_010_001_002	Rotarod: Passive rotation								
ESLIM_010_001_701*	Rotarod: Latency to fall mean								
ESLIM_011_001_006	Acoustic Startle & PPI:110dB startle magnitude								
ESLIM_011_001_007	Acoustic Startle & PPI:PP1 + pulse startle magnitude								
ESLIM_011_001_008	Acoustic Startle & PPI:PP2 + pulse startle magnitude								
ESLIM_011_001_009	Acoustic Startle & PPI:PP3 + pulse startle magnitude								
ESLIM_011_001_010*	Acoustic Startle & PPI:PP4 + pulse startle magnitude								
ESLIM_011_001_702*	Acoustic Startle & PPI:Prepulse inhibition - PP2								
ESLIM_011_001_703*	Acoustic Startle & PPI:Prepulse inhibition - PP3								
ESLIM_011_001_705*	Acoustic Startle & PPI:Global prepulse inhibition								
ESLIM_015_001_001	Clinical Chemistry: Glucose								
ESLIM_015_001_002	Clinical Chemistry: Urea								
ESLIM_015_001_004	Clinical Chemistry: Sodium								
ESLIM_015_001_005	Clinical Chemistry: Potassium								
ESLIM_015_001_006	Clinical Chemistry: Chloride								

B

Parameter	Description	HMGU M	HMGU F	ICS M	ICS F	MRC Harwell M	MRC Harwell F	WTSI M	WTSI F
ESLIM_005_001_003	DEXA: Lean mass								
ESLIM_005_001_004	DEXA: Bone Mineral Density (excluding skull)								
ESLIM_007_001_007	Open-field: Periphery resting time								
ESLIM_009_001_002	Grip-Strength: Forelimb and hindlimb grip strength measurement								
ESLIM_009_001_702*	Grip-Strength: Forelimb and hindlimb grip strength measurement mean								
ESLIM_011_001_001	Acoustic Startle & PPI:BN startle magnitude								
ESLIM_011_001_002	Acoustic Startle & PPI:PP1 startle magnitude								
ESLIM_011_001_701*	Acoustic Startle & PPI:Prepulse inhibition - PP1								
ESLIM_011_001_704*	Acoustic Startle & PPI:Prepulse inhibition - PP4								
ESLIM_012_001_002	Hot Plate: Type of response								
ESLIM_015_001_007	Clinical Chemistry: Total protein								
ESLIM_015_001_012	Clinical Chemistry: Lactate dehydrogenase								
ESLIM_015_001_015	Clinical Chemistry: Alkaline phosphatase								
ESLIM_015_001_016	Clinical Chemistry: Alpha-amylase								
ESLIM_015_001_019*	Clinical Chemistry: Free fatty acid								
ESLIM_016_001_001	Haematology: White blood cell count								
ESLIM_016_001_002	Haematology: Red blood cell count								
ESLIM_016_001_005	Haematology: Mean cell volume								
ESLIM_016_001_006	Haematology: Mean corpuscular haemoglobin								
ESLIM_020_001_002	Heart weight/tibia length: Heart weight								
ESLIM_021_001_001	Fasted Clinical Chemistry: Glucose								
ESLIM_021_001_003	Fasted Clinical Chemistry: Triglycerides								
ESLIM_021_001_004	Fasted Clinical Chemistry: Free fatty acids								

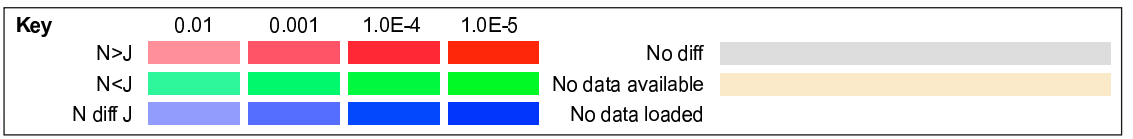


Figure 1

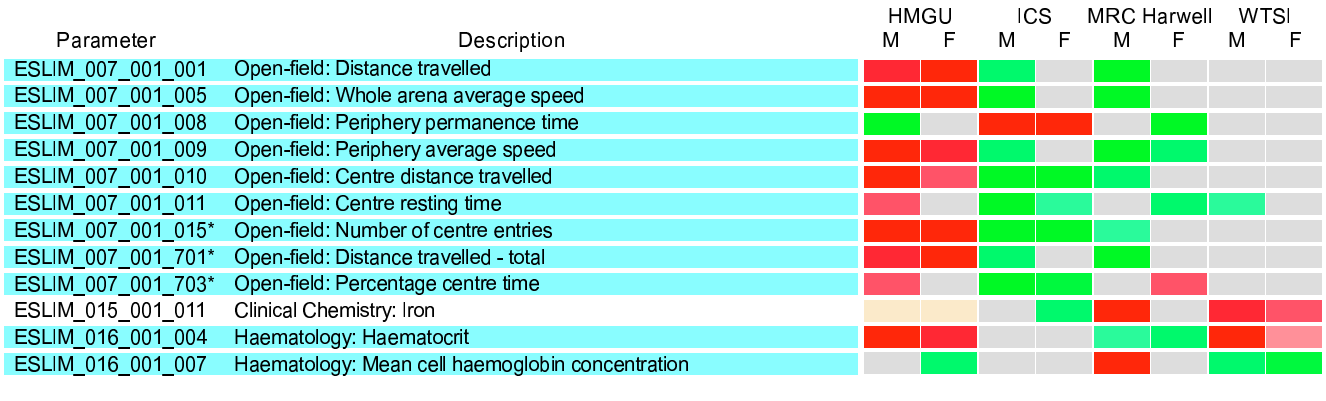
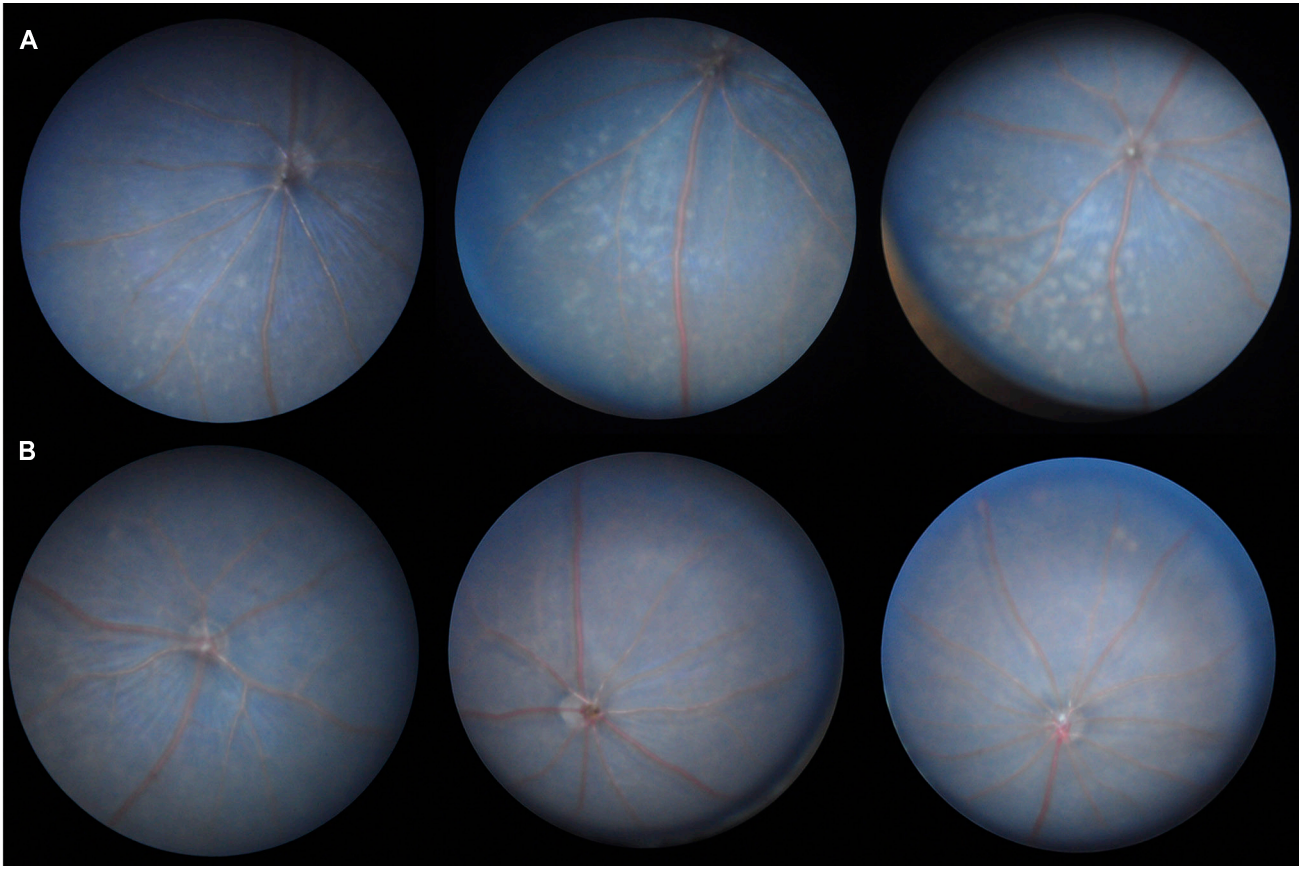

HMGU
ICS
MRC Harwell
WTSI
Parameter
Description
M F M F M F M F
ESLIM_007_001_001 Open-field: Distance travelled
ESLIM_007_001_005 Open-field: Whole arena average speed
ESLIM_007_001_008 Open-field: Periphery permanence time
ESLIM_007_001_009 Open-field: Periphery average speed
ESLIM_007_001_010 Open-field: Centre distance travelled
ESLIM_007_001_011 Open-field: Centre resting time
ESLIM_007_001_015* Open-field: Number of centre entries
ESLIM_007_001_701* Open-field: Distance travelled - total
ESLIM_007_001_703* Open-field: Percentage centre time
ESLIM_015_001_011 Clinical Chemistry: Iron
ESLIM_016_001_004 Haematology: Haematocrit
ESLIM_016_001_007 Haematology: Mean cell haemoglobin concentration
Key
0.01
0.001
1.0E-4
1.0E-5
N>J
N<J
N diff J
No diff
No data available
No data loaded

Figure 2


A
B

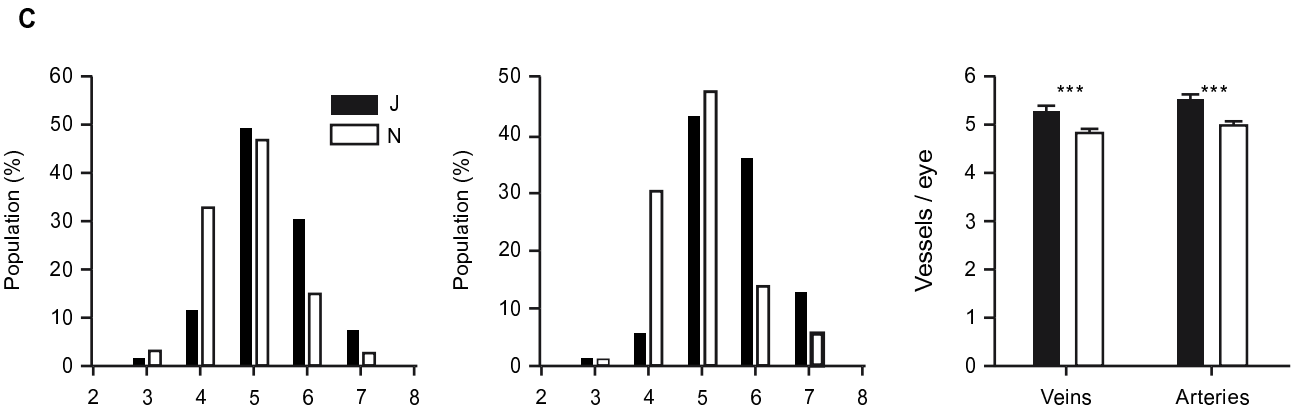
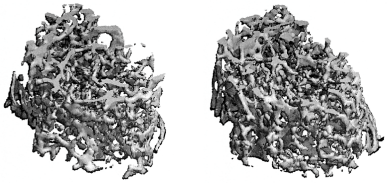

C
Population (%)
J
N
Veins / eye
Population (%)
Arteries / eye
Vessels / eye
***
***
Veins
Arteries

Figure 3

A

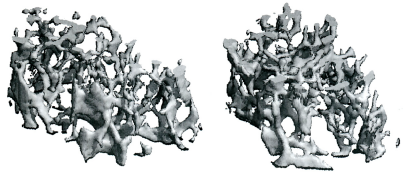
Male



	C57BL/6J	C57BL/6NTac
B.V./T.V. (%)	14,4 ± 1,4	13,2 ± 1,8
Micro-architecture parameters		
Tb.N. (1/mm)	5,25 ± 0,12	5,49 ± 0,15
Tb.Sp. (mm)	0,187 ± 0,005	0,180 ± 0,005
Tb.Th. (mm)	0,048 ± 0,001	0,040 ± 0,001
Conn-Dens. (1/mm3)	150,3 ± 16,5	155,9 ± 21,1
Structural indexes		
S.M.I. (arbitrary unit)	2,38 ± 0,14	2,67 ± 0,10
D.A. (arbitrary unit)	1,52 ± 0,04	1,51 ± 0,04

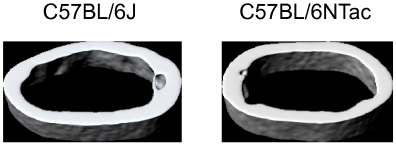
B

Female



	C57BL/6J	C57BL/6NTac
B.V./T.V. (%)	7,4 ± 0,5	7,5 ± 0,5
Micro-architecture parameters		
Tb.N. (1/mm)	3,74 ± 0,04	3,61 ± 0,14
Tb.Sp. (mm)	0,263 ± 0,003	0,277 ± 0,015
Tb.Th. (mm)	0,049 ± 0,002	0,053 ± 0,003
Conn-Dens. (1/mm3)	28,6 ± 2,7	29,2 ± 4,1
Structural indexes		
S.M.I. (arbitrary unit)	3,11 ± 0,07	3,09 ± 0,08
D.A. (arbitrary unit)	1,49 ± 0,03	1,38 ± 0,03

C



	C57BL/6J	C57BL/6NTac
Ct.Po. (%)	2,65 ± 0,25	2,57 ± 0,18
Ct.Th. (mm)	0,159 ± 0,004	0,154 ± 0,004

D

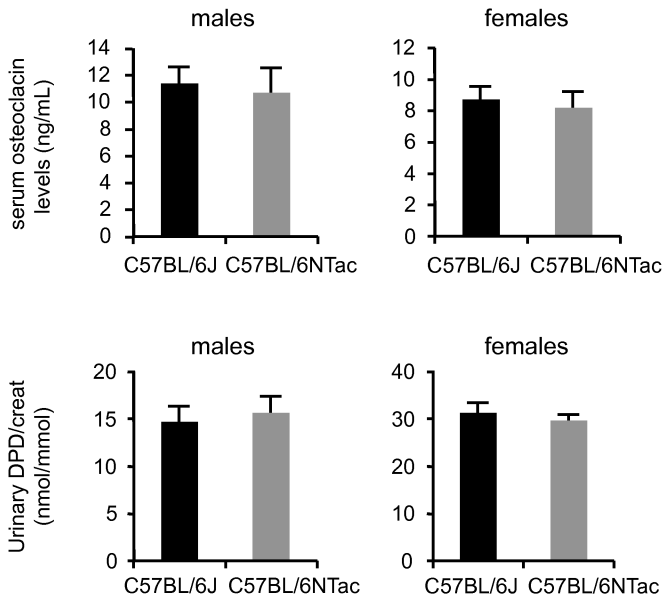


Figure 4

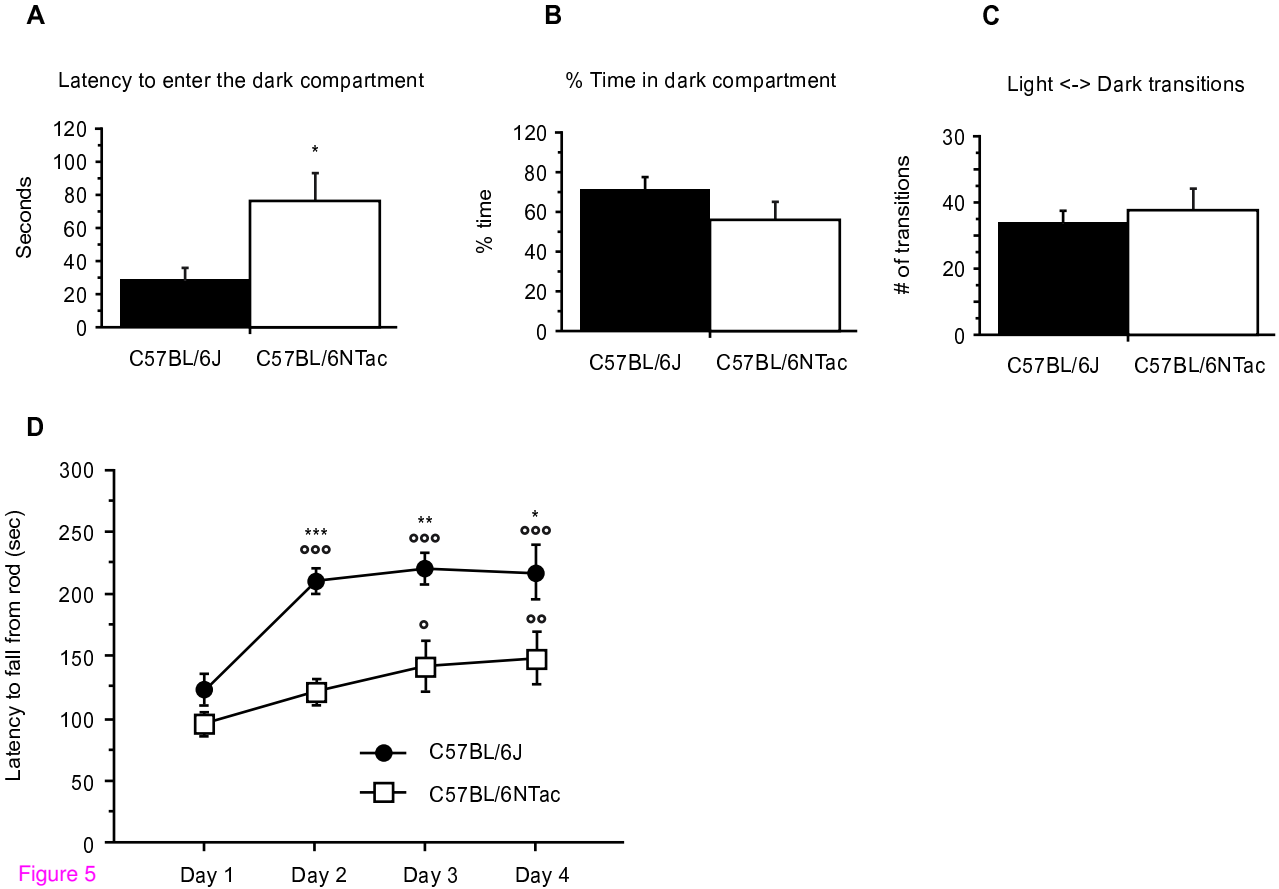
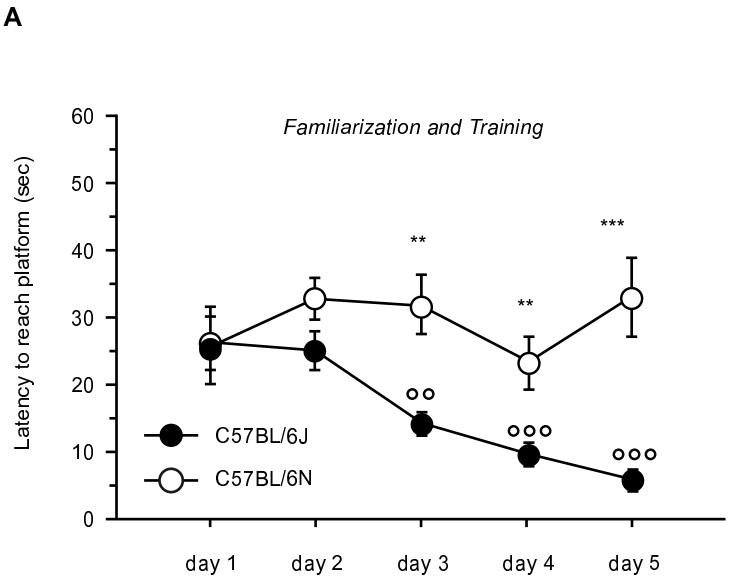
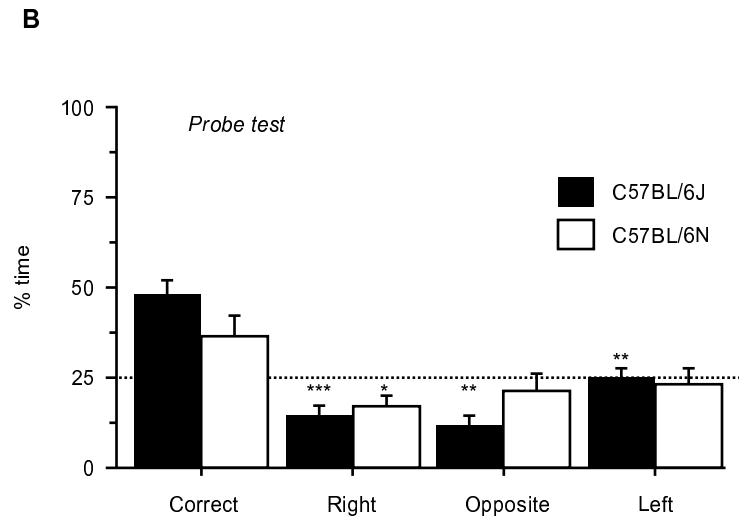

A
Latency to enter the dark compartment
Seconds
120
100
80
60
40
20
0
*
C57BL/6J
C57BL/6NTac
B
% Time in dark compartment
% time
120
80
60
40
20
0
C57BL/6J
C57BL/6NTac
C
Light <-> Dark transitions
# of transitions
30
40
20
0
C57BL/6J
C57BL/6NTac
D
Latency to fall from rod (sec)
300
250
200
150
100
50
0
***
**
*
°°°
°°°
°°°
°
°°
C57BL/6J
C57BL/6NTac
Day 1
Day 2
Day 3
Day 4

Figure 5





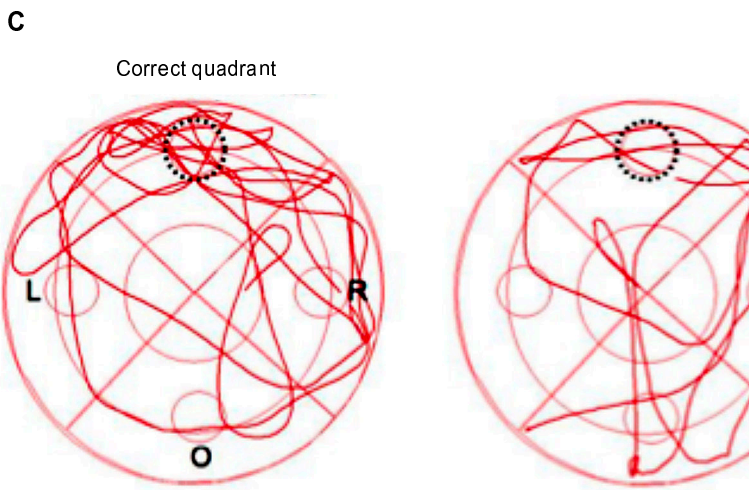


Figure 6 J4 N3

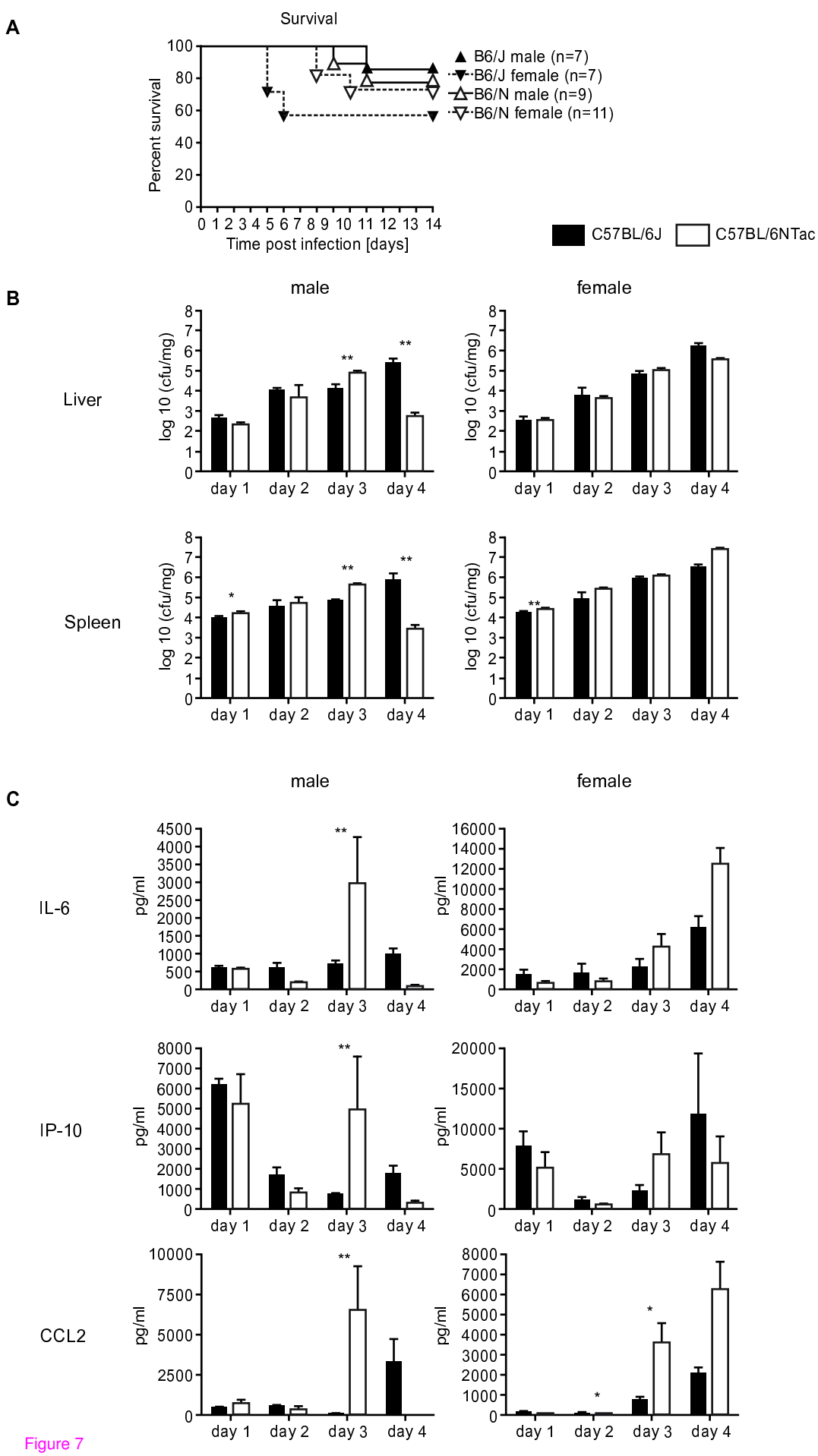
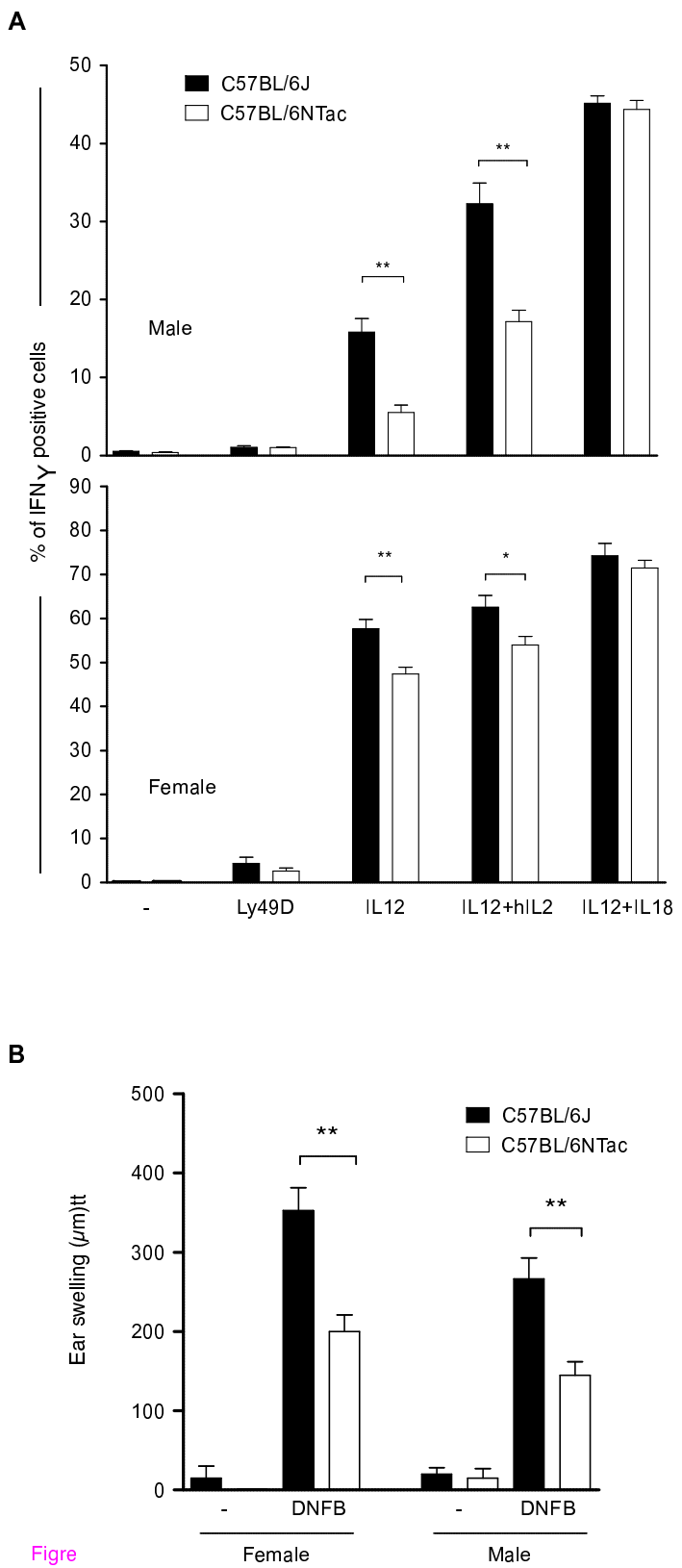

A
Survival
Percent survival
Time post infection [days]
B6/J male (n=7)
B6/J female (n=7)
B6/N male (n=9)
B6/N female (n=11)
C57BL/6J
C57BL/6NTac
B
male
female
Liver
Spleen
log 10 (cfu/mg)
day 1
day 2
day 3
day 4
C
IL-6
IP-10
CCL2
pg/ml

Figure 7


A
C57BL/6J
C57BL/6NTac
Male
Female
% of IFNγ positive cells
50
40
30
20
10
0
90
80
70
60
50
40
30
20
10
0
**
**
**
*
-
Ly49D
IL12
IL12+hIL2
IL12+IL18
B
C57BL/6J
C57BL/6NTac
Ear swelling (μm)tt
500
400
300
200
100
0
**
**
-
DNFB
-
DNFB
Female
Male

Figre

Additional files provided with this submission:

Additional file 1: TableS1.pdf, 225K
http://genomebiology.com/imedia/8046326471045638/supp1.pdf
Additional file 2: Supplementary_Figure_1.pdf, 544K
http://genomebiology.com/imedia/1086014387104073/supp2.pdf
Additional file 3: Supplementary_Figure_2.pdf, 2994K
http://genomebiology.com/imedia/1777967506104073/supp3.pdf