The Company of **Biologists**

RESEARCH ARTICLE

Lung necrosis and neutrophils reflect common pathways of susceptibility to *Mycobacterium tuberculosis* in genetically diverse, immune-competent mice

Muhammad K. K. Niazi[1], Nimit Dhulekar[2], Diane Schmidt[3], Samuel Major[3], Rachel Cooper[3], Claudia Abeijon[3], Daniel M. Gatti[4], Igor Kramnik[5], Bulent Yener[2], Metin Gurcan[1] and Gillian Beamer[3,*]

ABSTRACT

Pulmonary tuberculosis (TB) is caused by *Mycobacterium tuberculosis* in susceptible humans. Here, we infected Diversity Outbred (DO) mice with ~100 bacilli by aerosol to model responses in a highly heterogeneous population. Following infection, 'supersusceptible', 'susceptible' and 'resistant' phenotypes emerged. TB disease (reduced survival, weight loss, high bacterial load) correlated strongly with neutrophils, neutrophil chemokines, tumor necrosis factor (TNF) and cell death. By contrast, immune cytokines were weak correlates of disease. We next applied statistical and machine learning approaches to our dataset of cytokines and chemokines from lungs and blood. Six molecules from the lung: TNF, CXCL1, CXCL2, CXCL5, interferon-γ (IFN-γ), interleukin 12 (IL-12); and two molecules from blood – IL-2 and TNF – were identified as being important by applying both statistical and machine learning methods. Using molecular features to generate tree classifiers, CXCL1, CXCL2 and CXCL5 distinguished four classes (supersusceptible, susceptible, resistant and non-infected) from each other with approximately 77% accuracy using completely independent experimental data. By contrast, models based on other molecules were less accurate. Low to no IFN-γ, IL-12, IL-2 and IL-10 successfully discriminated non-infected mice from infected mice but failed to discriminate disease status amongst supersusceptible, susceptible and resistant *M.-tuberculosis*-infected DO mice. Additional analyses identified CXCL1 as a promising peripheral biomarker of disease and of CXCL1 production in the lungs. From these results, we conclude that: (1) DO mice respond variably to *M. tuberculosis* infection and will be useful to identify pathways involving necrosis and neutrophils; (2) data from DO mice is suited for machine learning methods to build, validate and test models with independent data based solely on molecular biomarkers; (3) low levels of immunological cytokines best indicate a lack of exposure to *M. tuberculosis* but cannot distinguish infection from disease.

KEY WORDS: Tuberculosis, *Mycobacterium tuberculosis*, Neutrophils, Necrosis, CXCL1, CXCL2, CXCL5, Machine learning, Diversity outbred, DO

[1]Department of Biomedical Informatics, The Ohio State University, Columbus, 43210 OH, USA. [2]Department of Computer Science and Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, 12810 NY, USA. [3]Department of Infectious Disease and Global Health, Cummings School of Veterinary Medicine, Tufts University, Grafton, 01536 MA, USA. [4]The Jackson Laboratory, Bar Harbor, 04662 ME, USA. [5]Department of Medicine, Boston University School of Medicine, Boston, 02215 MA, USA.

*Author for correspondence (gillian.beamer@tufts.edu)

INTRODUCTION

Over the past two decades, inbred laboratory mice have been used to identify important host responses to *Mycobacterium tuberculosis*. However, studies using inbred mice might not generalize to all mice or to all humans. Benefits of using a diverse experimental population are the reduction of erroneous interpretations due to strain-specific effects and the detection of new phenotypes that are not present in standard inbred laboratory strains (Yang et al., 2007, 2011). Therefore, we have started using Diversity Outbred (DO) mice as another tool to understand susceptibility to *M. tuberculosis*. The genetic diversity of the DO population is on par with that of the human population, and the DO genomes are suited for high-resolution genetic mapping. For these reasons, DO mice have been used to identify novel genetic associations with a variety of traits (Harrison et al., 2009; Churchill et al., 2012; Svenson et al., 2012; Ferris et al., 2013; Logan et al., 2013; Recla et al., 2014), but they have not yet been used extensively as a model of *M. tuberculosis* infection.

Heterogeneity in the DO population is due to eight distinctive founder strains that were chosen for maximal genetic diversity. Five founder strains are inbred laboratory strains and three are wild-derived strains (Churchill et al., 2012; Svenson et al., 2012). Only three of the eight founder strains (C57BL/6, 129 and A/J) have been used in *M. tuberculosis* research, primarily used to identify requirements for TH1-mediated immunological resistance and to provide some insight into susceptibility (Medina and North, 1998; Jagannath et al., 2000; Chackerian and Behar, 2003; Beamer and Turner, 2005). *M. tuberculosis* infection of the other five founder strains (NOD/LtJ, NZO/HlLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ) has not yet been reported. To our knowledge, DO mice have been used in only one *M. tuberculosis* infection study (Gopal et al., 2013) and one aging study (Harrison et al., 2014). Thus, additional work is needed. Here, we infected DO mice with *M. tuberculosis* by aerosol and evaluated morbidity, bacterial burden and a set of immunological and inflammatory molecules in lungs, antigen-stimulated blood cultures and plasma. Multiple methods (simple and complex statistical analyses, and machine learning) were used to identify and rank important molecular features. The features were then used to generate tree classification models to discriminate disease status using only molecular biomarkers. Finally, tree classification models were tested using data from a completely independent experiment.

Decades of research show a complex host response to *M. tuberculosis*, involving many cell types, molecules and signaling pathways. Resistance to mycobacterial infection in humans, and to *M. tuberculosis* in mice, requires TH1 polarized cell-mediated immunity, including the cytokines IFN-γ, IL-12, TNF

Disease Models & Mechanisms

TRANSLATIONAL IMPACT

Clinical issue
Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis*, which typically affects the lungs. Most infections remain latent, and only about 10% of infected individuals develop active TB, which, if left untreated, can be fatal. Animal models are a useful research tool to understand mechanisms that determine susceptibility to *M. tuberculosis*. This work models TB in a genetically heterogeneous experimental mouse population, called Diversity Outbred mice, which might better model the human population than standard inbred strains of mice that are commonly used in biomedical research.

Results
The results show that some immune responses and disease features in Diversity Outbred mice reflect disease features of human pulmonary TB – including lung and granuloma necrosis, and the presence of neutrophils. Furthermore, the authors developed and tested models that could discriminate disease status with 77% accuracy based on the detection of three neutrophil chemokines: CXCL1, CXCL2 and CXCL5. One of these chemokines, CXCL1, has the most potential to be a peripheral plasma biomarker of lung disease and of CXCL1 lung levels in individuals with TB. Additionally, as in humans, pro-inflammatory, cell-mediated, TH1 antigen-specific immune responses do not reflect TB disease state very well, but low immunological responses are very useful to identify non-exposed individuals.

Implications and future directions
This work is expected to influence the field, by showing that Diversity Outbred mice can model pathologic changes observed in human pulmonary TB, and therefore can be useful to identify and understand mechanisms of necrosis and neutrophil involvement in human TB. Future work using these mice will be exceptionally valuable when larger data sets are generated both for genetic studies, aimed at identifying novel genes and loci associated with increased susceptibility, and for longitudinal biomarker studies, in particular those focused on markers and models that can predict outcome before disease ensues.

and IL-2. Additionally, tuberculosis (TB) disease might be accelerated by the immune suppressive cytokine IL-10. When absent or blocked, these cytokines alter the survival of mice, indicating that their presence is important for outcome (Gong et al., 1996; Flynn and Chan, 2001; Turner et al., 2002; Beamer et al., 2008a,b; Cooper, 2009; Higgins et al., 2009; Zhang et al., 2011). Similarly, humans with TH1 deficiencies are at increased risk of mycobacterial disease (Casanova and Abel, 2002; Redford et al., 2011; Liang et al., 2014), and there is evidence that IL-10 is associated with pulmonary TB in humans as well (Gong et al., 1996; Zhang et al., 2011).

A perplexing problem is that pulmonary TB, the contagious form of TB, most commonly occurs in immune-competent individuals who remain immunologically responsive to mycobacterial antigens and do not have TH1 defects (Hunter et al., 2006; Modlin and Bloom, 2013; Andersen and Woodworth, 2014; Nunes-Alves et al., 2014; Vilaplana and Cardona, 2014). Thus, although TH1 immunity is a requirement for resistance, its presence does not prevent disease in susceptible individuals. For many years, lung and granuloma necrosis, and neutrophils have been recognized as negative indicators of pulmonary TB in humans (Barnes et al., 1988; Berry et al., 2010; Leong et al., 2011). In fact, many studies implicate neutrophils or neutrophil-like cells, and chemokines that recruit neutrophils, as causal mediators of lung damage in mice (Keller et al., 2006; Lyadova et al., 2010; Nandi and Behar, 2011; Yang et al., 2012; Gopal

et al., 2013; Major et al., 2013; Dorhoi et al., 2014; Nouailles et al., 2014). Therefore, we assessed necrosis, neutrophils and the neutrophil chemokines CXCL1, CXCL2 and CXCL5 to determine whether these markers of TB disease show phenotypic variation in *M.-tuberculosis*-infected DO mice.

Following aerosol infection, DO mice indeed varied in survival, morbidity, bacterial burden, granuloma composition, immunological cytokines and neutrophil chemokines. Nearly half of the DO mouse population rapidly developed morbidity associated with necrosis of lung tissue and granulomas, with neutrophil influx. Strong and significant disease correlates were lung neutrophil chemokines (CXCL1, CXCL2, CXCL5), TNF and dead cells. By contrast, IL-12, IL-2, IFN-γ and IL-10 were weak, or were not, disease correlates. In the periphery (plasma), only CXCL1 was a strong and statistically significant correlate of disease and of lung CXCL1, whereas CXCL5 was not, and neither were any other cytokines.

Statistical analyses, data mining and machine learning approaches identified important molecular features that were then used to classify status of *M.-tuberculosis*-infected mice in the absence of disease indicators (survival, weight loss, bacterial burden). Lung TNF, CXCL1, CXCL2, CXCL5, IFN-γ, IL-12 and blood IL-2 and TNF were consistently identified as important. The most accurate models relied on CXCL1, CXCL2 and CXCL5 to classify 'supersusceptible', 'susceptible', 'resistant' and 'non-infected' mice. TNF was an important feature, but models that included TNF were less accurate because they confused resistant mice with supersusceptible mice. Immune cytokines were highly successful in discriminating non-infected from infected mice, but could not distinguish disease status.

In summary, of the molecules that we measured, neutrophil chemokines CXCL1, CXCL2 and CXCL5 best distinguished TB disease in highly susceptible DO mice, whereas low and/or a lack of immune cytokines identified non-infected individuals. Furthermore, the classification model performed well using completely independent experimental data, resulting in 77% accuracy. Similar to humans, necrosis and neutrophils are TB disease features, and CXCL1 is a good peripheral biomarker in DO mice. (Gopal et al., 2013). Overall, these results indicate DO mice are another useful murine model of pulmonary TB, in particular to understand the mechanisms that contribute to neutrophil recruitment and necrosis during early *M. tuberculosis* infection.

RESULTS

DO mice respond variably to atomized *M. tuberculosis*
All mice gained weight for at least two weeks following infection with ~100 *M. tuberculosis* bacilli (supplementary material Fig. S1). Afterwards, nearly half of the DO mice developed morbidity requiring euthanasia before 35 days, resulting in significantly reduced survival compared to the founder C57BL/6J strain (Fig. 1A). Reduced survival strongly reflected the daily rate of weight loss (not shown), as observed in other heterogeneous mice (Harrison et al., 2014). As expected, the percentage of the peak body weight at euthanasia and lung *M. tuberculosis* burden had a strong, significant inverse correlation (Fig. 1B) such that high bacterial burden associated with weight loss. Interestingly, these effects were independent of absolute body weight before infection or at peak body weight (not shown). Thus, body weight alone does not appear to protect against TB.

It is common in human TB studies to stratify individuals by using clinical symptoms or disease severity, and we applied this approach to DO mice. Three susceptibility classes were detected
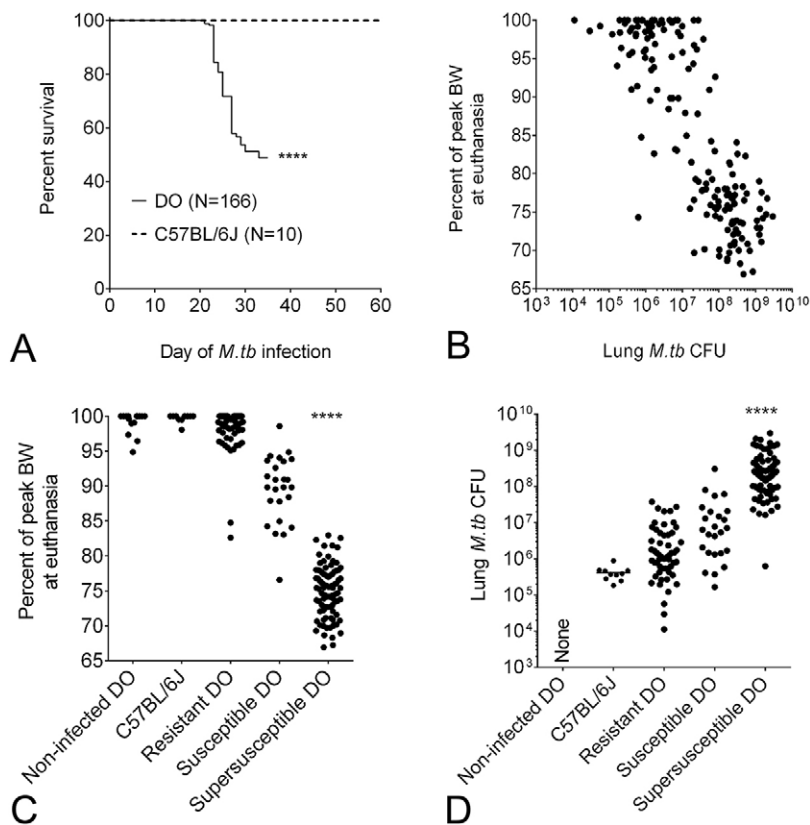
Fig. 1. *M. tuberculosis* infection and TB disease indicators in mice. Female 8-week-old non-sibling DO mice (*N*=166) and C57BL/6J mice (*N*=10) were infected with ~100 *M. tuberculosis* (*M.tb*) bacilli by aerosol. Survival was assessed by euthanasia owing to morbidity or by euthanasia at day 35 of infection, whichever came first. Survival of DO mice compared with that of the parental C57BL/6J strain was analyzed by using a Log-rank test, ********P*<0.001 (A). The percentage of peak body weight (BW) and lung *M. tuberculosis* CFU were strong, significant negative correlations in DO mice (Spearman r −0.79, *P*<0.001, 95% CI −0.84 to −0.72) (B). The percentage of peak BW (C) and lung *M. tuberculosis* burden (D) from non-infected, resistant, susceptible and supersusceptible (described in the Materials and Methods) DO mice, and C57BL/6J mice were analyzed by using ANOVA with Tukey's post-test, ********P*<0.0001. Data are combined from two independent experimental infections: *N*=97 DO mice in the first experiment, *N*=69 DO mice and *N*=10 C57BL/6J mice in the second experiment.

during infection – supersusceptible, susceptible and resistant. Fig. 1C,D show that supersusceptible mice lost the most body weight and had the highest lung bacterial burdens; susceptible mice were intermediate for both; and resistant mice exhibited the lowest values for both. Fig. 2 shows common lung lesions for each susceptibility class. The lungs of supersusceptible DO mice typically contained large lesions, occupying approximately 75% of lung tissue, which were effaced by coalescing foci with central coagulation necrosis of lung alveolar septae and intra-alveolar macrophages [resembling caseous necrosis (Leong et al., 2011; Marzo et al., 2014) and necrotizing tuberculous pneumonia (Hunter, 2011)], often with a peripheral rim of thrombosed septal capillaries and intra-alveolar neutrophilic debris. By contrast, the lungs of resistant mice typically contained small granulomas, occupying approximately 25-50% of lung tissue, comprising macrophages and abundant perivascular lymphocytes, but little necrosis or neutrophils, resembling the *M.-tuberculosis*-resistant C57BL/6 founder strain (Vesosky et al., 2010; Niazi et al., 2014). Susceptible DO mice showed intermediate patterns with small regions of coagulation and/or caseous necrosis that were surrounded by mixed inflammatory cells (macrophages, lymphocytes, neutrophils), but lacked thrombosis.

These findings link the susceptibility of DO mice to patterns (necrosis, neutrophils) that are similar to those observed in humans with active TB and in other susceptible inbred mice (Barnes et al., 1988; Eruslanov et al., 2005; Berry et al., 2010; Eum et al., 2010; Lyadova et al., 2010; Hunter, 2011; Leong et al., 2011; Harper et al., 2012; Gopal et al., 2013; Major et al., 2013; Hunter et al., 2014). Taken together, these results indicate that lung necrosis and neutrophil influx are common pathways that are likely to reflect multiple different mechanisms or interconnected pathways.

Lung and peripheral correlates of TB disease in *M.-tuberculosis*-infected DO mice

Fig. 3 depicts the cytokine and chemokine data for each class of DO mice (supersusceptible, susceptible, resistant and non-infected) and for the C57BL/6J founder strain for comparison. Using data from infected DO mice, the following chemokines and cytokines were tested for correlations with disease indicators (percentage of peak body weight at euthanasia, bacterial burden, survival): lung, blood and plasma. Lastly, lung and blood and plasma cytokines and chemokines were correlated with each other. Lung cytokine and chemokine correlations with disease indicators are shown in Table 1 [percentage of peak body weight at euthanasia and *M. tuberculosis* colony-forming units (CFU)]. Lung cytokine and chemokine correlations with survival were nearly identical to those for the other indicators (not shown). Statistically significant, very strong or strong disease correlates in DO mouse lungs were dead cells, CXCL1, CXCL2, CXCL5 and TNF, which supports previous studies from inbred mice and humans (Bekker et al., 2000; Gopal et al., 2013; Nouailles et al., 2014). Lung cytokines (IFN-γ, IL-2, IL-12, IL-10) were weak correlates of disease with variable statistical significance.

Table 1 also shows correlations of blood and plasma cytokines and chemokines with disease indicators. Plasma CXCL1 was a moderately strong, statistically significant correlate. By contrast, plasma CXCL5 was not a correlate, probably because some resistant mice had high levels (Fig. 3). CXCL2 was not detectable in plasma. Blood cytokines and cytokine ratios (IFN-γ:TNF, IFN-γ:IL-10 and IL-2:IFN-γ) were weak or very weak correlates of disease, although many had highly statistically significant relationships. Importantly, all these results in DO mice mimic studies in humans with TB, such that CXCL1 appears to be a good peripheral biomarker of lung disease (Gopal et al., 2013), and cytokines and the ratios are weak but
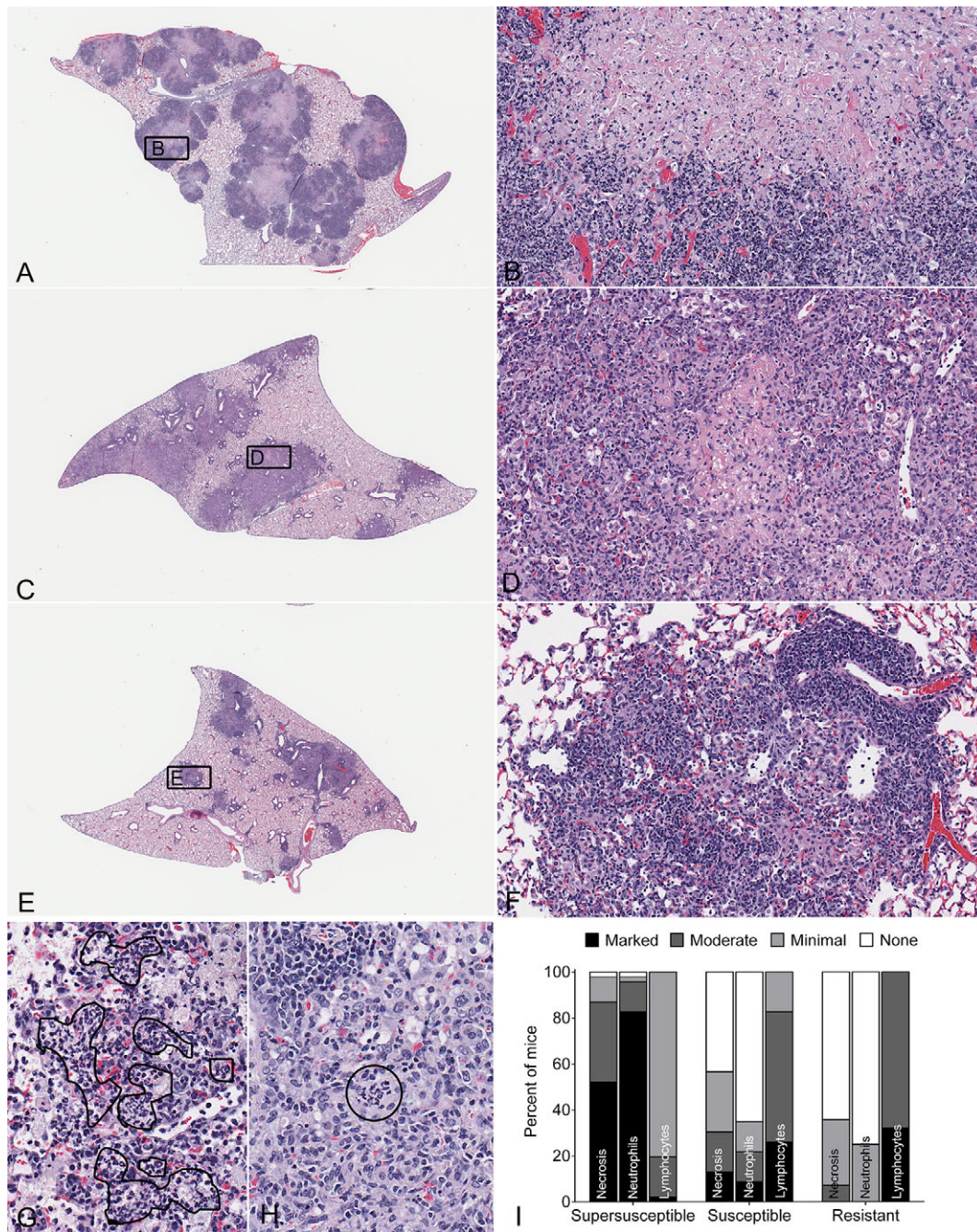
Fig. 2. Lung lesions in *M.-tuberculosis*-infected DO mice. Female 8-week-old non-sibling DO mice (*N*=97) were infected with ∼100 *M. tuberculosis* (*M.tb*) bacilli by using an aerosol. Representative hematoxylin and eosin (H&E)-stained lung sections are shown. Supersusceptible (A) with inset showing substantial lung tissue, macrophage and neutrophil necrosis with capillary thrombosis magnified (B); susceptible (C) with inset showing a small region of necrosis surrounded by inflammatory cells that lacks thrombosis (D); resistant (E) with inset showing a small, non-necrotic lesion with an abundance of perivascular lymphocytes (F). Black lines surround neutrophils in alveolar spaces (G) and within macrophage-rich granulomas (H). Magnifications are 10× (A,C,E); 200× (B,D,F); and 400× (G,H). Two lung lobes from each mouse were scored for relative severity of each lesion type by a board-certified veterinary pathologist (G.B.) without knowledge of the groups, and the data was compiled (I).

significant correlates (Jamil et al., 2007; Millington et al., 2007; Sahiratmadja et al., 2007; Suter-Riniker et al., 2011; Hur et al., 2013).

Correlations of lung cytokines and chemokines with those of blood and plasma were also examined. From the molecules that we measured, only plasma CXCL1 correlated with lung CXCL1 levels (supplementary material Fig. S2A) and, as expected from the data presented in Fig. 3, supersusceptible mice had significantly higher levels of CXCL1 in lung and plasma than susceptible, resistant or non-infected DO mice

(*P*<0.001, ANOVA with Tukey's post-test). All other cytokines and chemokines, and cytokine ratios had weak or no correlations with the levels in the lungs or blood and plasma (not shown), including CXCL5 (supplementary material Fig. S2B). The paucity of correlations was surprising, so we further probed the relationships of lung cytokine and chemokines with those in the blood and plasma by performing multidimensional analyses [2D-correlation coefficients (Noda, 1993)] and quadratic discriminant analysis (Lachenbruch, 1975; Scholkopft and
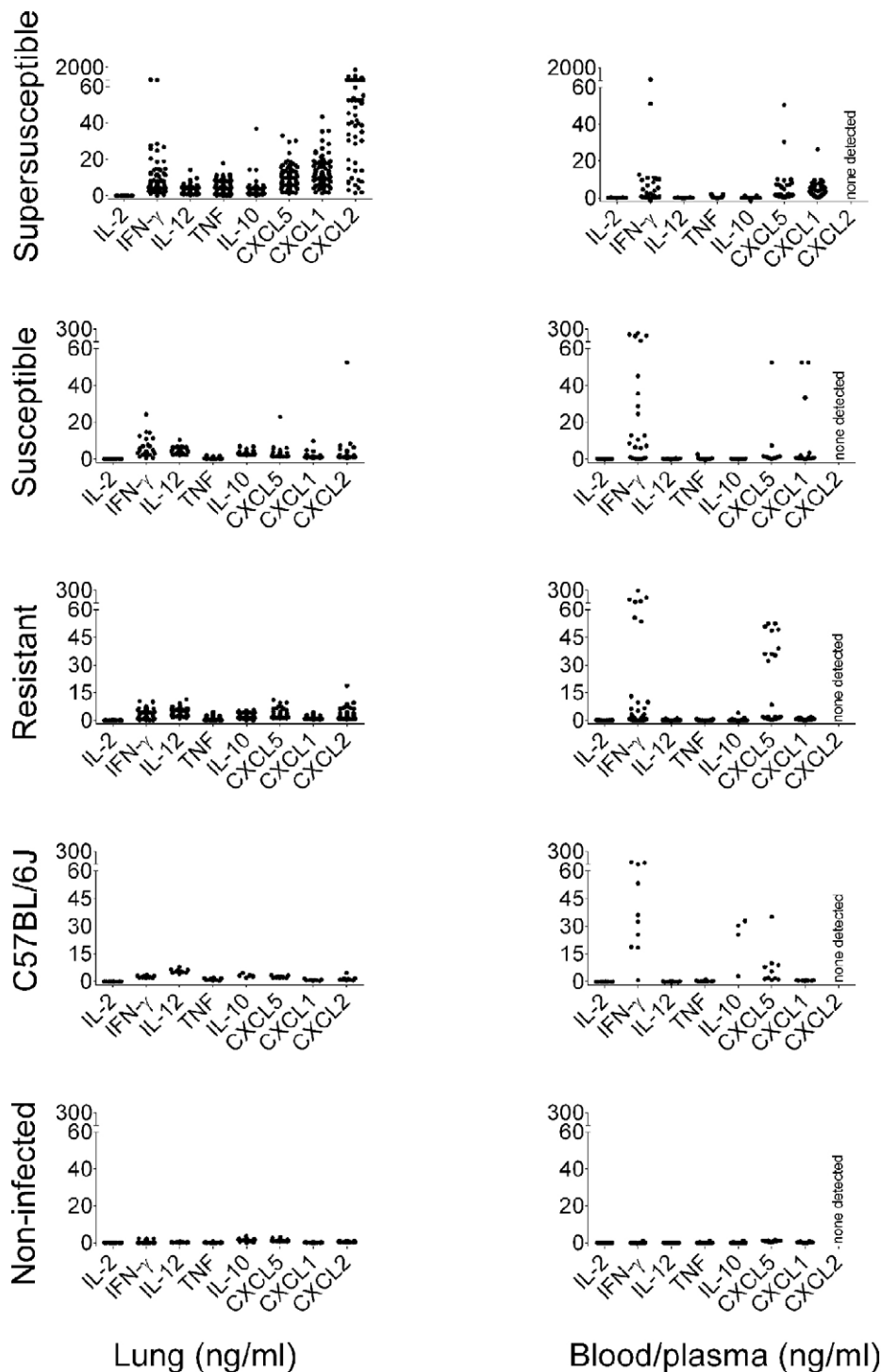
Fig. 3. Molecular profiles of lung, and of blood and plasma in *M.-tuberculosis*-infected mice. Female 8-week-old non-sibling DO mice (*N*=166) and C57BL/6J (*N*=10) mice were infected with ~100 *M. tuberculosis* (*M.tb*) bacilli by aerosol. Cytokines and chemokines were quantified in homogenized lung. Blood cytokines (IL-2, IFN-γ, IL-12, TNF, IL-10) were quantified after stimulation with antigen. Neutrophil chemokines (CXCL1, CXCL5) were quantified in plasma, but CXCL2 was not detectable. From right to left, molecular features are grouped as follows: T-cell cytokines (IL-2, IFN-γ), macrophage cytokines (IL-12, TNF, IL-10) and neutrophil chemokines (CXCL1, CXCL5, CXCL2). Each dot represents the average of duplicate or triplicate samples from one mouse. The y-axes are defined at the bottom of the figure. Data are combined from two independent experiments.

Mullert, 1999). These analyses also failed to detect significant relationships between lung and blood and plasma profiles, indicating that combined lung and peripheral responses do not mirror each other.

To summarize, the results demonstrate that lung neutrophil chemokines and necrosis are strong and statistically significant TB disease correlates in *M.-tuberculosis*-infected DO mice but that immune cytokines in the lung are not. Of the molecules that we measured, plasma CXCL1 appears to be the best biomarker for neutrophil-associated lung damage, which has been suggested previously in mice and in humans (Gopal et al., 2013).

Neutrophil chemokines can classify the disease status of *M.-tuberculosis*-infected mice

Statistical analyses and machine learning methods consistently identified similar cytokines and chemokines as being important from the 15 molecular features that we measured. ANOVA followed by Tukey's multiple comparison (MCT) tests identified cytokines and chemokines that could separate pairs of disease classes by using statistical significance (Table 2). These same cytokines and chemokines are inferred from direct observations (Fig. 3) and correlations (Table 1) such that supersusceptible DO mice had high lung TNF, CXCL1, CXCL2

Table 1. TB disease correlates in DO mice

	Correlation (r)	Strength of r	95% CI	Sample size	*P*-value of r
Correlations of lung features and percent of peak body weight at euthanasia					
CXCL1	−0.75	Strong	−0.81 to −0.66	146	<0.0001
CXCL2	−0.74	Strong	−0.81 to −0.66	146	<0.0001
CXCL5	−0.56	Strong	−0.82 to −0.50	146	<0.0001
TNF	−0.62	Strong	−0.71 to −0.51	166	<0.0001
Dead cells	−0.60	Strong	−0.71 to −0.43	80	<0.0001
IFN-γ	−0.30	Weak	−0.44 to −0.15	163	<0.0001
IL-2	−0.24	Weak	−0.42 to −0.04	97	0.0183
IL-12	0.36	Weak	0.21 to 0.48	161	0.02
IL-10	0.04	Very weak	−0.12 to 0.20	150	0.61
IFN-γ:TNF	0.41	Moderate	0.27 to 0.53	161	<0.0001
IFN-γ:IL-10	−0.26	Weak	−0.41 to −0.09	144	0.0016
IL-2:IFN-γ	0.15	Very weak	−0.006 to 0.30	94	0.90
Correlations of lung features and *M. tuberculosis* lung CFU					
CXCL1	0.83	Very strong	0.77 to 0.87	146	<0.0001
CXCL2	0.81	Very strong	0.74 to 0.86	146	<0.0001
CXCL5	0.62	Strong	0.50 to 0.71	146	<0.0001
TNF	0.66	Strong	0.56 to 0.74	166	<0.0001
Dead cells	0.59	Moderate	0.41 to 0.70	80	<0.0001
IFN-γ	0.32	Moderate	0.17 to 0.45	163	<0.0001
IL-2	0.19	Very weak	−0.17 to 0.38	97	0.06
IL-12	−0.30	Weak	−0.44 to −0.15	161	<0.0001
IL-10	0.009	Very weak	−0.15 to 0.17	150	0.91
IFN-γ:TNF	−0.43	Moderate	−0.55 to −0.29	163	<0.0001
IFN-γ:IL-10	0.25	Weak	0.08 to 0.40	144	0.0031
IL-2:IFN-γ	−0.06	Very weak	−0.21 to 0.10	163	0.45
Correlations of plasma and blood features, and percentage of peak body weight at euthanasia					
CXCL1	−0.50	Moderate	−0.61 to −0.037	159	<0.0001
CXCL5	−0.15	Very weak	−0.01 to 0.29	159	0.065
TNF	0.31	Weak	0.15 to 0.45	160	<0.0001
IFN-γ	0.25	Weak	0.09 to 0.30	161	0.0014
IL-2	0.30	Weak	0.15 to 0.44	161	<0.0001
IL-12	0.15	Very weak	−0.008 to 0.30	160	0.0557
IL-10	−0.07	Very weak	−0.07 to 0.23	161	0.34
IFN-γ:TNF	0.21	Weak	0.05 to 0.38	137	0.0106
IFN-γ:IL-10	0.24	Weak	0.06 to 0.40	126	0.0075
IL-2:IFN-γ	0.32	Weak	0.16 to 0.46	150	<0.0001
Correlations of plasma and blood features, and *M. tuberculosis* lung CFU					
CXCL1	0.53	Moderate	0.40 to 0.63	159	<0.0001
CXCL5	−0.10	Very weak	−0.25 to 0.06	159	0.23
TNF	−0.27	Weak	−0.41 to −0.12	160	0.0005
IFN-γ	−0.24	Weak	−0.38 to −0.08	161	0.0023
IL-2	−0.22	Weak	−0.36 to −0.06	161	0.0060
IL-12	−0.07	Very weak	−0.23 to 0.086	160	0.35
IL-10	−0.09	Very weak	−0.24 to 0.07	161	0.27
IFN-γ:TNF	−0.18	Very weak	−0.35 to −0.01	137	0.0313
IFN-γ:IL-10	−0.29	Weak	−0.45 to −0.12	126	0.0009
IL-2:IFN-γ	−0.25	Weak	−0.40 to −0.09	150	0.0022

Female 8-week-old non-sibling DO mice (*N*=166) were infected with ∼100 *M. tuberculosis* bacilli by using an aerosol. Cytokines and chemokines in lungs were directly quantified in homogenized tissue. Blood cytokines (IL-2, IFN-γ, IL-12, TNF, IL-10) were quantified after antigen stimulation of heparinized whole blood. Neutrophil chemokines (CXCL5, CXCL1) were directly quantified in plasma. Spearman correlation coefficients (r) were calculated and interpreted as very weak (0-0.19), weak (0.20-0.39), moderate (0.40-0.59), strong (0.60-0.79) or very strong (0.80-1.0). The 95% confidence interval (95% CI), sample size and *P*-values are reported.

and CXCL5 levels, and a high plasma CXCL1 concentration. Non-infected mice had low to no IFN-γ, IL-12, IL-2 or IL-10. However, ANOVA followed by MCT failed to identify molecules that could distinguish resistant and susceptible mice from each other. A limitation of ANOVA is that some statistical results do not make biological sense. For example, non-infected DO mice could not be identified by using ANOVA and MCT because the confidence interval for *M. tuberculosis* CFUs in

resistant mice extended below zero. Thus, although ANOVA followed by MCT corroborated previous results, statistical analyses alone cannot discriminate the four classes of DO mice (supersusceptible, susceptible, resistant and non-infected) because there is no embedded classification method. This led us to use, for the first time in *in vivo* experimental *M. tuberculosis* research, machine learning methods to build testable classification models.

Table 2. Ability of single features to distinguish pairs of disease classes by using ANOVA followed by MCT

Feature	Supersusceptible (SS)	Susceptible (S)	Resistant (R)	Non-infected (N)
% of peak BW	All	All	SS, S	SS, S
M. tuberculosis CFU	All	SS	SS	SS
Lung CXCL1	All	SS	SS	SS
Lung CXCL2	All	SS	SS	SS
Lung CXCL5	R, N	–	SS	SS
Lung TNF	All	SS	SS	SS
Lung IFN-γ	N	N	N	All
Lung IL-12	N	N	N	All
Lung IL-10	N	N	N	All
Lung IL-2	Insufficient data from test experiment			
Plasma CXCL1	–	–	–	–
Plasma CXCL5	–	–	–	–
Blood TNF	All	SS	SS	SS
Blood IFN-γ	–	–	–	–
Blood IL-12	S	SS	–	–
Blood IL-10	–	–	–	–
Blood IL-2	–	N	–	S

The rows correspond to features, whereas the columns represent the classes. Cells within each row show the features that separate pairs of classes with statistical confidence determined using ANOVA followed by Tukey's post-test ($P<0.05$). For example, the feature body weight for the supersusceptible (SS) class is significantly different from all other classes, as is body weight for the susceptible (S) class. Body weight for resistant and non-infected classes are also significantly different from SS and S, but not from each other. In other words, body weight alone cannot discriminate R from N. –, the markers could not separate any classes from each other due to overlapping confidence intervals.

By using machine learning methods, we extracted the same molecular features that were strong disease correlates or had some statistical ability to distinguish classes: lung CXCL1, CXCL2, CXCL5, TNF, IFN-γ, IL-12; and two blood features – IL-2 and TNF. Although our data also contained three disease indicators (survival, percentage of peak body weight at euthanasia and *M. tuberculosis* CFU), these were purposefully excluded because the goal was to identify molecules capable of classifying disease status and discrimination of a non-infected state. For building and testing classification models, only data from individual mice with complete data – i.e. all 15 molecular features – were used. Thus, the training and model validation steps used $N=70$ DO mice from the first experiment that had been infected with 127 ± 68 *M. tuberculosis* Erdman bacilli and $N=8$ non-infected DO mice. Multiple machine learning approaches were used, listed in the Materials and Methods. Of these, classification trees were pursued because they produced the most accurate models. Validation for the model used the standard leave-one-out cross-validation for training data (Fig. 4B).

However, performance rates in training cannot be generalized. A more stringent approach to evaluate performance (accuracy in classifying disease status of DO and C57BL/6J mice) is to test the model using completely independent molecular data. Therefore, we tested the models using data from a completely independent experiment, again using only mice for which all 15 molecular parameters were available. This included $N=55$ DO mice and $N=5$ C57BL/6J mice from the second experiment that had been infected with 97 ± 61 bacilli and $N=3$ non-infected DO mice.

When tested using independent data, classification trees using the strongest disease correlates [CXCL1, CXCL2, CXCL5 (Fig. 4) and TNF, CXCL1, CXCL2, CXCL5 (not shown)] gave the best results for classifying the disease status of DO mice – 77% and 58%, respectively. Of note, all C57BL/6J mice were correctly classified as resistant by all models. This reflects the widely recognized phenotype of C57BL/6 mice as compared with many other inbred strains of mice (reviewed in Beamer and Turner, 2005). Trees with TNF were problematic because they confused resistant with
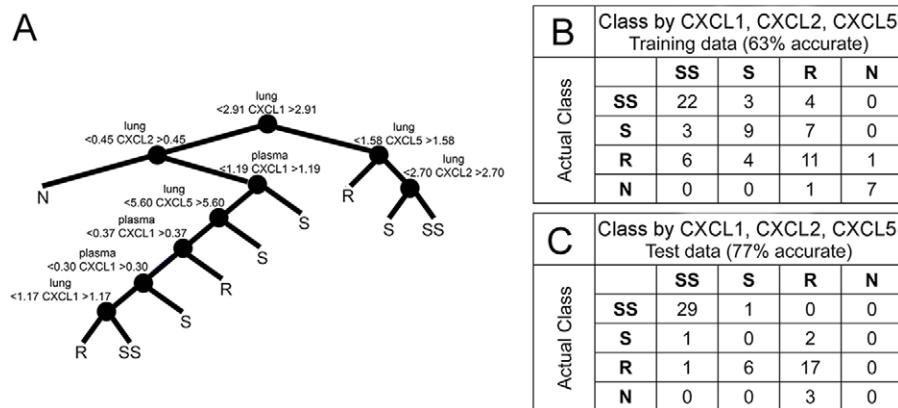

Fig. 4. Classification tree based on neutrophil chemokines CXCL1, CXCL2, CXCL5. Female 8-week-old non-sibling DO mice and C57BL/6J mice were infected with ∼100 *M. tuberculosis* bacilli by aerosol. From the first experiment, complete data for all 15 molecular parameters used for model generation and training were obtained from 70 *M.-tuberculosis*-infected DO mice and eight non-infected DO mice. The best-performing classification tree used only lung and plasma CXCL1, CXCL2, CXCL5 (A), which was then validated by the leave-one-out method (B). After validation, performance of the classification tree was stringently tested using data from the second independent experiment, which included 52 *M.-tuberculosis*-infected DO mice, three non-infected DO mice and five *M.-tuberculosis*-infected C57BL/6J mice (C), with complete data from all 15 molecular parameters.

Table 3. Models based on neutrophil CXC-chemokines better classify disease status of DO mice than models based on immune cytokines

Source	Features	Accuracy (%)	Comments
Lung, plasma	CXCL1, CXCL2, CXCL5	63; 77	This is Fig. 4. Best at discriminating SS and N. Sometimes confuses S and R.
Lung, blood and plasma	CXCL1, CXCL2, CXCL5, TNF	71; 58	Good at discriminating SS. Confuses R with SS.
Lung	CXCL1, CXCL2, CXCL5, TNF	55; 53	Confuses SS, S and R. Good at discriminating N.
Lung, blood and plasma	CXCL1, CXCL2, CXCL5, TNF, IFN-γ, IL-12, IL-10	61; 55	Good at discriminating SS and N. Confuses S and R.
Lung	CXCL1, CXCL2, CXCL5, TNF, IFN-γ, IL-12, IL-10	50; 58	Confuses SS, S and R. Good at discriminating N.
Blood and plasma	CXCL1, CXCL5, TNF, IFN-γ, IL-12, IL-10, IL-2	57; 60	Good at discriminating SS. Confuses S, R and N.
Lung, blood and plasma	CXCL1, CXCL2, CXCL5, TNF, IFN-γ, IL-12, IL-10, IL-2	61; 55	Good at discriminating SS and N. Confuses S and R.
Lung, blood	IFN-γ, IL-12, IL-10, IL-2	38; 47	Confuses SS, S and R. Good at discriminating N.
Lung	IFN-γ, IL-12, IL-10	53; 45	Confuses SS, S and R. Good at discriminating N.

Training data were acquired from the first experiment of *M.-tuberculosis*-infected (*N*=70) and non-infected (*N*=8) DO mice for which all 15 molecular features had been quantified. Trees were validated by the leave-one-out method and then tested using independent data from 52 *M.-tuberculosis*-infected DO mice, three non-infected DO mice and five C57BL/6J mice for which all 15 molecular features had been quantified. Results are summarized from classification tree models and confusion matrices that included various combinations of neutrophil chemokines (CXCL1, CXCL2, CXCL5), TNF and immune molecules (IFN-γ, IL-12, IL-10, IL-2). Accuracy is reported as overall accuracy for the discrimination of four disease classes (supersusceptible, susceptible, resistant and non-infected mice) in training and testing, separated by a semi-colon respectively. Comments summarize the strengths and weakness of the models generated.

supersusceptible DO mice, which could reflect the different roles of TNF as a resistance requirement and disease correlate in inbred mice and humans (Law et al., 1996; Bekker et al., 1998, 2000; Juffermans et al., 1998; Ray et al., 2009; Fallahi-Sichani et al., 2010), and in DO mice (Table 1).

All combinations of important molecules were used to generate and test additional classification trees but performance was not improved, ranging between 38 and 61% (Table 3). Models with accuracy of >50% showed TNF to be a dominant feature. Models organized by anatomic site (i.e. lung alone, lung plus blood and plasma, blood and plasma alone) or by function (i.e. immune cytokines IL-2, IL-12, IFN-γ, IL-10) with or without TNF, and with or without CXCL1, CXCL2 and CXCL5 did not improve accuracy above that achieved using CXCL1, CXCL2 and CXCL5 (Fig. 4). Models using cytokine ratios also failed to discriminate supersusceptible, susceptible, resistant and non-infected DO mice (not shown).

To summarize, from the 15 molecule parameters that we measured, the disease status of DO mice and C57BL/6 mice was best classified using three neutrophil chemokines CXCL1, CXCL2 and CXCL5. This classification tree model was particularly good at discriminating supersusceptible DO mice from all other categories, confirming inferences from inspecting raw data (Fig. 3) and statistical analyses (correlations and ANOVA-MCT); this model also has the benefit of being performance-tested with independent data. However, the model was not perfect. It misclassified some resistant and susceptible DO mice, and misclassified non-infected DO mice as resistant in testing. We next investigated whether the format of the data influenced the selection of important molecular features and assessed how DO mice were misclassified.

Some DO mice that were misclassified on the basis of neutrophil chemokines have unique phenotypes

To determine whether the magnitude of cytokine and chemokine responses influenced feature selection, data were normalized. This had no effect, and the same molecules were identified as being important (not shown).

The numbers of misclassifications were too small to apply data mining in order to extract information, so all results were manually inspected to identify recording errors (none were found) and to

seek biologically relevant explanations. Because of the method by which classification tree models are generated, it is not possible to examine 'misclassifications' in the training data. Therefore, only misclassifications in the test data were evaluated. Thirteen of the 55 DO test mice were misclassified by using CXCL1, CXCL2, CXCL5 (Fig. 4). Three were non-infected mice that were misclassified as resistant. This error was due to higher lung CXCL2 levels in the test (average 0.8 ng/ml) than the training (average 0.4 ng/ml) data. Thus, these non-infected mice were erroneously classified as resistant because of inter-experiment variability. Additional training data from non-infected mice would reduce this type of misclassification.

Seven resistant DO mice were misclassified. Of these seven, two were misclassified because of high CXCL1 and CXCL5, possibly owing to macrophage production without substantial neutrophil recruitment, which was corroborated by using microscopy analyses. The remaining five resistant mice were misclassified after the first node, despite having typical lung lesions. Three susceptible mice were misclassified after the first node, and these mice had a variety of lung lesions. Only one supersusceptible mouse was misclassified. This mouse was clearly different based on the molecule profile (lung CXCL1 less than 2.91 ng/ml) and microscopy analyses. Instead of neutrophil infiltration and necrosis, morbidity reflected massive lung infiltration by non-necrotic macrophages.

Overall, the examination of misclassified mice was helpful. Misclassified non-infected mice reflected inter-experiment variability, which could be improved with additional training data. Two misclassified resistant mice better fit the susceptible class, showing that the model might help to re-classify ambiguous cases. It was unclear how susceptible mice were misclassified because the numbers were too small to draw meaningful conclusions. The one misclassified supersusceptible mouse had a unique phenotype – marked infiltration by non-necrotic macrophages without neutrophils, which suggests another, but less common, response to *M. tuberculosis* that leads to rapid morbidity.

DISCUSSION

We investigated survival, morbidity, bacterial load, immune and inflammatory molecules in *M.-tuberculosis*-infected and non-infected DO mice, which is a genetically heterogeneous

experimental population. It is another useful mouse model for TB research because the genetic diversity in the DO population mimics the diversity in the human population. The most relevant findings from the studies here are that many of the DO responses to *M. tuberculosis* infection model human responses, including lung and granuloma necrosis and neutrophil influx; production of inflammatory molecules; production of immunological cytokine responses, indicative of antigen-specific TH1 immunity; and production of immune-suppressive cytokines. Furthermore, like humans, neutrophil inflammation and necrosis are strong features of TB disease, whereas immune cytokines are weak or very weak correlates of disease.

We included IFN-γ, IL-12, TNF, IL-2, IL-10 in these DO studies because these cytokines significantly alter survival when absent or blocked in inbred mice and because they have been extensively investigated in humans with TB. Like humans, *M.-tuberculosis*-infected DO mice produce immune cytokines, but these molecules are poor disease correlates and could not be used to discriminate disease status. CXCL1, CXCL2 and CXCL5 were included because CXC-chemokines might contribute to detrimental inflammation and neutrophil influx, or might be biomarkers of lung disease (Gong et al., 1996; Flynn and Chan, 2001; Turner et al., 2002; Beamer et al., 2008a,b; Cooper, 2009; Higgins et al., 2009; Zhang et al., 2011; Gopal et al., 2013; Nouailles et al., 2014). In particular, plasma CXCL1 has been suggested as a good peripheral biomarker of lung disease (Gopal et al., 2013). Likewise, in our studies we observed that lung CXC chemokines strongly correlate with disease and that CXCL1 appears to be the best peripheral biomarker of disease in a genetically diverse population. CXCL5 was not a good peripheral biomarker, and CXCL2 was below the limit of detection in plasma and could not be assessed as a peripheral biomarker.

To further utilize the DO data, we used supervised machine learning methods to generate and test tree classification models with the intent of identifying molecules that could determine TB disease status and identify the non-infected state. The best model successfully discriminated four classes (supersusceptible, susceptible, resistant and non-infected) from each other with 77% accuracy using only CXCL1, CXCL2 and CXCL5. Performance was excellent in distinguishing supersusceptible mice from all other classes, but performed less well in distinguishing susceptible, resistant and non-infected mice. Non-infected mice were misclassified owing to inter-experimental variability, which can be corrected with additional data. The reasons for misclassification of susceptible and resistant mice were not clear, indicating that additional molecular features are needed to improve discrimination.

We also analyzed immune cytokines that are important for resistance (TNF, IFN-γ, IL-12, IL-2) and one cytokine that contributes to susceptibility (IL-10) (Gong et al., 1996; Flynn and Chan, 2001; Turner et al., 2002; Beamer et al., 2008a,b; Cooper, 2009; Higgins et al., 2009; Zhang et al., 2011). All were poor correlates of disease with the exception of TNF, which is likely to reflect its dual role as a marker of resistance and as a disease correlate. All models using immune cytokines from any anatomic location (lung and periphery) failed to classify the disease status of *M.-tuberculosis*-infected DO mice but were very good at discriminating non-infected mice. These findings are important because they reflect meta-analyses from human data and the recent literature reviews by experts, which conclude that a lack of TH1 responses identifies a lack of exposure to *M. tuberculosis*, whereas the presence of TH1 immunity can neither rule in or out TB disease in humans (Modlin and Bloom, 2013; Andersen and Woodworth, 2014; Nunes-Alves et al., 2014; Rangaka et al., 2012). Thus, the DO

population could be a good experimental model to help identify better markers of immunological resistance to *M. tuberculosis*.

Performance of the tree classification models could have been improved by computational approaches – such as collapsing the data into two classes (supersusceptible and "other"), replicating data to yield groups of equivalent numbers and combining data into a single training set. However, we did not pursue these approaches because they do not reflect experimental or biological reality. The concept of generating and testing models with independent data might be relevant to human studies where subjects are assigned to clinical categories by TB disease severity (i.e. diagnoses of latent infection, active pulmonary TB, extra pulmonary TB etc.). The responses of each diagnostic category are then compared with each other and analyzed by using statistical methods. However, as we observed here, statistical analyses might not produce testable or accurate models if multiple classes are present and there is abundant or complex data. Therefore, we suggest that machine learning methods are another useful tool to discriminate TB disease status in humans.

DO mice have been used previously in only two studies (Gopal et al., 2013; Harrison et al., 2014), but the breadth of DO responses to *M. tuberculosis* was not the focus. Here, the survival of supersusceptible DO mice mimics that of immune-deficient C57BL/6 mice that cannot produce TH1 responses (reviewed in Flynn and Chan, 2001; Cooper, 2009 and references therein). However, neither DO mice nor the eight founder strains have these immune defects. Furthermore, we demonstrated that DO mice are fully capable of producing TH1 immune cytokines locally and systemically in response to *M. tuberculosis*. Thus, susceptibility in the DO population is not due to inadequate immunity but instead reflects TB disease pathways involving necrosis and neutrophils, and possibly endothelial damage.

The remarkable susceptibility of some DO mice also indicates that genetic material from founder strains enhances susceptibility to *M. tuberculosis*. Based on on-going studies, genetic material associated with increased susceptibility comes from four founder strains – NOD/LtJ, NZO/HlLtJ, PWK/PhJ and WSB/EiJ (Christopher Sassetti, personal communication). However, further research is needed to identify novel susceptibility loci, and the DO population is ideally suited because the genomes are constructed for high-resolution genetic mapping (Zhang et al., 2012; Logan et al., 2013). DO susceptibility should not involve the *Ipr1* polymorphism in C3HeB/FeJ mice (Pan et al., 2005; Tosh et al., 2006; Pichugin et al., 2009) as the polymorphism is absent from DO founders (Mouse Phylogeny Viewer and the Sanger Mouse Genome Database accessed on October 6th 2014 by D.G. in Dr Gary Churchill's laboratory, The Jackson Laboratory, Bar Harbor, ME). Therefore, future genetic studies using DO mice should identify additional genes and loci that contribute to susceptibility.

An important finding from these studies is that neutrophils and neutrophil chemokines are correlates of TB disease in *M.-tuberculosis*-infected DO mice. These same patterns reflect pathologic features of human pulmonary TB. Interest in neutrophil responses and consequences during *M. tuberculosis* infection in human and mouse studies has been increasing slowly over the past few decades and is now accelerating (Barnes et al., 1988; Keller et al., 2006; Eum et al., 2010; Gopal et al., 2013; Major et al., 2013). The majority of studies implicate neutrophils or neutrophil-like cells as causal mediators of lung damage (Keller et al., 2006; Lyadova et al., 2010; Nandi and Behar, 2011; Yang et al., 2012; Gopal et al., 2013; Major et al., 2013; Dorhoi et al., 2014; Nouailles et al., 2014), but the underlying molecular mechanisms and pathways are not yet

well defined. The fundamental biological properties of neutrophils as short-lived, innate immune cells that secrete proteases, oxidants, and inflammatory cytokines and chemokines (Elkington et al., 2011; Palanisamy et al., 2011; Amulic et al., 2012; Wilkinson et al., 2012; Gopal et al., 2013; Dorhoi et al., 2014; Nouailles et al., 2014) support this concept that neutrophils are destructive to lung tissue. However, experimental evidence in mouse *M. tuberculosis* studies is not that straightforward. It appears that neutrophils play a variety of roles that differ depending on the host, the dose of *M. tuberculosis* and the duration of *M. tuberculosis* infection. For example, depletion of neutrophils during early infection improves survival of *M.-tuberculosis*-susceptible DBA/2 mice but only with high-dose infection, and the same depletion protocol has no effect on resistant C57BL/6 mice (Keller et al., 2006). This supports the hypothesis that neutrophils are detrimental under certain conditions. By contrast, our unpublished observations indicate that neutrophil depletion during chronic infection accelerates TB disease in CBA/J mice, suggesting that neutrophils are protective under other conditions. Given that some DO mice recruit neutrophils to the lungs and have high levels of CXCL1, CXCL2 and CXCL5, DO mice provide another useful model to understand the signals that attract neutrophils and to define the context(s) in which neutrophils are detrimental or protective. This will be important for a more complete understanding of TB pathogenesis across a genetically diverse population, and also to assess outcomes of interventions aimed at reducing neutrophil-mediated damage.

A limitation of our study could have been the use of light microscopy to identify immune and inflammatory responses in lung sections. This approach maintains the tissue and granuloma architecture, providing a wealth of visual information but at the expense of rigorous identification and quantification of specific immune cells (e.g. T cells, B cells, macrophages, dendritic cells, and all of their subsets etc.) that participate in host responses to *M. tuberculosis*. We did not attempt identification of immature neutrophils (also known as 'bands', 'band cells' or 'band neutrophils') because the morphology might overlap with immature myeloid-derived suppressor cells that have been recently described in lethal murine TB (Tsiganov et al., 2014).

Given the remarkable susceptibility of some DO mice to *M. tuberculosis*, future studies with a lower infectious dose will be important to slow lung damage and to improve detection of individual granulomas. This approach successfully identified multiple susceptibility loci and one gene *Ipr1* (mouse) that alters macrophage necrosis and increases susceptibility, and helped to identify SP110 (the human homolog to mouse *Ipr1*) polymorphisms in TB-infected individuals (Kramnik et al., 1998, 2000; Pan et al., 2005; Tosh et al., 2006; Yan et al., 2006; Kramnik, 2008; Pichugin et al., 2009; Sissons et al., 2009). However, the SP110 polymorphisms do not explain all TB cases (Thye et al., 2006; Png et al., 2012) and additional studies are needed.

A mouse model might not recapitulate all characteristics of human pulmonary TB. However, mice are important animals to investigate experimental questions that cannot be addressed in humans. Here, DO mice are advantageous because the population provides immunologically intact, genetically diverse individuals, and the phenotypic disease spectrum overlaps with humans. Using machine learning to build models that can classify disease status is new and useful, particularly where statistical analyses are not successful. In this context, machine learning will be especially useful to move beyond classification of TB disease status (which we describe here for the first time using only molecular correlates of disease) to studies that are capable of predicting TB disease outcome

before it happens. That is a powerful approach and is achievable by sampling individual *M.-tuberculosis*-infected DO mice over time using peripheral host and mycobacterial biomarkers, followed by data mining, machine learning and independent testing.

MATERIALS AND METHODS
Ethics statement
The Cummings School of Veterinary Medicine (CSVM) at Tufts University has a letter of Assurance on file with the Office of Laboratory Animal Welfare. CSVM follows the Public Health Service Policy on Humane Care and Use of Laboratory Animals; the Guide for the Care and Use of Laboratory Animals; and the U.S. Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research and Training. CSVM is also accredited by AAALACi, complies with the USDA Animal Welfare Regulations and is a USDA APHIS registered Research Facility (14-R-0065). Investigators follow guidelines pertinent for the species. Here, experimental procedures were approved by Tufts University IACUC protocols (G2012-53; G2012-151; G2015-33). Biosafety Level 3 (BSL3) work was approved by Tufts University Institutional Biosafety Committee registration (GRIA04; GRIA10).

Infection with *Mycobacterium tuberculosis* and quantification of lung bacilli
Female J:DO (009376) and C57BL/6J (000664) mice (The Jackson Laboratory, Bar Harbor, ME, USA) were maintained under BSL3 conditions with sterile food, bedding and water in the New England Regional Biosafety Laboratory (North Grafton, MA, USA). At 8 weeks of age, mice were infected by aerosol exposure using a CH Technologies, Inc (Westwood, New Jersey, USA) machine. Exposure delivered 127±68 *M. tuberculosis* Erdman bacilli to the lungs (*N*=97 DO mice) that resulted in complete data from 70 individual DO mice to train machine learning models. A separate experiment delivered 97±61 bacilli to the lungs (*N*=69 DO and *N*=10 C57BL/6J mice) that resulted in complete data from *N*=52 DO mice and *N*=5 C57BL/6J mice to test machine learning models with the independent molecular data. Lung bacillary burden was determined following homogenization in sterile PBS using gMACS M-tubes (Miltenyi Biotech, Cambridge, MA, USA), plating serial dilutions on OADC-supplemented 7H11 agar and counting CFUs after 3 weeks at 37°C.

Identification of disease classes
Mice were euthanized when morbidity met IACUC-approved early removal criteria, or at day 35 of infection, whichever came first. Morbidity criteria included: weakness, respiratory difficulty, ruffled fur and a body condition score of less than 2 (Ullman-Cullere and Foltz, 1999). Body weight was monitored throughout, but weight loss alone was not sufficient for early euthanasia nor was body weight used as a threshold. Mice that met early removal criteria before day 35 were classified as supersusceptible. Susceptible mice had no signs of morbidity before day 35, but retrospective analysis identified some weight loss. Resistant mice survived 35 days of *M. tuberculosis* infection without any signs of morbidity, and retrospective analysis identified stable weight or weigh gain. Age- and sex-matched non-infected control mice, housed identically, were euthanized on day 35. Supersusceptible and resistant classes were fundamentally based on survival, whereas susceptible mice were those that survived to day 35 without external evidence of morbidity but had lost weight upon retrospective analysis. These classes and non-infected DO mice (*N*=8 for model training and *N*=3 for model testing) were used as the ground truth for establishing and training algorithms.

Cytokine measurements
ELISAs for TNF, IFN-γ, IL-12, IL-2 and IL-10 were performed using antibody pairs and standards or OptEIA kits (BD Biosciences, San Jose, CA, USA) on serially diluted homogenized lung or from antigen-stimulated whole blood as described, with the exception that blood was diluted 1:5 (Beamer et al., 2008a,b). Samples at or below the IFN-γ level of detection were repeated using the eBioscience Ready-Set-Go ELISPOT kit (San Diego, CA, USA) (Beamer et al., 2011). CXCL1, CXCL2 and CXCL5 were

quantified using R&D Systems ELISA kits (Minneapolis, MN, USA) on diluted homogenized lung and plasma.

Enumeration of dead cells
Dead cells were enumerated following collagenase and DNase digestion of the lungs, and trypan blue staining on single cell suspensions (Beamer et al., 2012).

Light microscopy
Lung lobes were inflated and fixed with 10% neutral buffered formalin, processed, embedded in paraffin, cut at 5 μm and stained with hematoxylin and eosin at the CSVM Histology Laboratory. Two serial sections, 100 μm apart, were examined by a board-certified veterinary pathologist (G.B.) without knowledge of the groups with regards to the relative extent of necrosis, neutrophils, macrophages and lymphocyte influx, and estimates of the proportion of effaced lung tissue.

Statistical analyses
GraphPad Prism 6.0 was used for correlations. Data had non-normal distributions, so Spearman correlation coefficients were calculated and identified as very weak (0-0.19), weak (0.20-0.39), moderate (0.40-0.59), strong (0.60-0.79), or very strong (0.80-1.0) and considered statistically significant if $P<0.05$. MATLAB® was used for analysis of variance (ANOVA) followed by Tukey's MCT to identify which features had significantly different means between any two classes, defined as $P<0.05$.

Machine learning
Machine learning can help identify complex relationships by exploring data in multidimensional space (Tibshirani, 1996; Kornaropoulos et al., 2014). We applied standard machine learning methods for model training and testing, using mice with all parameters measured (N=78 mice for training and N=60 mice for testing). Five supervised [classification tree, relief attribute evaluation, consistency subset, spectral feature selection and a custom principal components algorithm (Kira and Rendell, 1992; Liu and Setiono, 1996; Coppersmith et al., 1999; Rokach and Maimon, 2005; Zheng and Liu, 2007)]; and two unsupervised [brute force and standard principal component analysis (Boutsidis et al., 2008)] methods were used. Of all methods, classification trees and relief attribute evaluation were the most informative because these methods ranked molecular features by importance. Classification trees were pursued for validation and testing with independent cytokine and chemokine molecular data because these trees provided the most accurate models.

Author contributions
I.K., B.Y., M.G. and G.B. conceived experiments, analyzed data, interpreted results, wrote the manuscript and prepared figures; M.K.K.N. performed machine learning, statistical analyses and wrote the paper; N.D. performed machine learning; R.C.

performed data analysis; G.B., D.S., S.M. and C.A. performed technical assays. D.G. performed bioinformatics analyses. All authors had final approval of the manuscript.

Supplementary material
Supplementary material available online at
http://dmm.biologists.org/lookup/suppl/doi:10.1242/dmm.020867/-/DC1

References
Amulic, B., Cazalet, C., Hayes, G. L., Metzler, K. D. and Zychlinsky, A. (2012). Neutrophil function: from mechanisms to disease. *Annu. Rev. Immunol.* **30**, 459-489.

Andersen, P. and Woodworth, J. S. (2014). Tuberculosis vaccines–rethinking the current paradigm. *Trends Immunol.* **35**, 387-395.

Barnes, P. F., Leedom, J. M., Chan, L. S., Wong, S. F., Shah, J., Vachon, L. A., Overturf, G. D. and Modlin, R. L. (1988). Predictors of short-term prognosis in patients with pulmonary tuberculosis. *J. Infect. Dis.* **158**, 366-371.

Beamer, G. L. and Turner, J. (2005). Murine models of susceptibility to tuberculosis. *Arch. Immunol. Ther. Exp.* **53**, 469-483.

Beamer, G. L., Flaherty, D. K., Assogba, B. D., Stromberg, P., Gonzalez-Juarrero, M., de Waal Malefyt, R., Vesosky, B. and Turner, J. (2008a). Interleukin-10 promotes *Mycobacterium tuberculosis* disease progression in CBA/J mice. *J. Immunol.* **181**, 5545-5550.

Beamer, G. L., Flaherty, D. K., Vesosky, B. and Turner, J. (2008b). Peripheral blood gamma interferon release assays predict lung responses and *Mycobacterium tuberculosis* disease outcome in mice. *Clin. Vaccine Immunol.* **15**, 474-483.

Beamer, G. L., Cyktor, J., Carruthers, B. and Turner, J. (2011). H-2 alleles contribute to antigen 85-specific interferon-gamma responses during *Mycobacterium tuberculosis* infection. *Cell. Immunol.* **271**, 53-61.

Beamer, G. L., Cyktor, J., Flaherty, D. K., Stromberg, P. C., Carruthers, B. and Turner, J. (2012). CBA/J mice generate protective immunity to soluble Ag85 but fail to respond efficiently to Ag85 during natural *Mycobacterium tuberculosis* infection. *Eur. J. Immunol.* **42**, 870-879.

Bekker, L.-G., Maartens, G., Steyn, L. and Kaplan, G. (1998). Selective increase in plasma tumor necrosis factor-alpha and concomitant clinical deterioration after initiating therapy in patients with severe tuberculosis. *J. Infect. Dis.* **178**, 580-584.

Bekker, L.-G., Moreira, A. L., Bergtold, A., Freeman, S., Ryffel, B. and Kaplan, G. (2000). Immunopathologic effects of tumor necrosis factor alpha in murine mycobacterial infection are dose dependent. *Infect. Immun.* **68**, 6954-6961.

Berry, M. P. R., Graham, C. M., McNab, F. W., Xu, Z., Bloch, S. A. A., Oni, T., Wilkinson, K. A., Banchereau, R., Skinner, J., Wilkinson, R. J. et al. (2010). An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973-977.

Boutsidis, C., Mahoney, M. W. and Drineas, P. (2008). Unsupervised feature selection for principal components analysis. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Casanova, J.-L. and Abel, L. (2002). Genetic dissection of immunity to mycobacteria: the human model. *Annu. Rev. Immunol.* **20**, 581-620.

Chackerian, A. A. and Behar, S. M. (2003). Susceptibility to Mycobacterium tuberculosis: lessons from inbred strains of mice. *Tuberculosis* **83**, 279-285.

Churchill, G. A., Gatti, D. M., Munger, S. C. and Svenson, K. L. (2012). The Diversity Outbred mouse population. *Mamm. Genome* **23**, 713-718.

Cooper, A. M. (2009). Cell-mediated immune responses in tuberculosis. *Annu. Rev. Immunol.* **27**, 393-422.

Coppersmith, D., Hong, S. J. and Hosking, J. R. M. (1999). Partitioning nominal attributes in decision trees. *Data Min. Knowl. Discov.* **3**, 197-217.

Dorhoi, A., Yeremeev, V., Nouailles, G., Weiner, J., III, Jörg, S., Heinemann, E., Oberbeck-Müller, D., Knaul, J. K., Vogelzang, A., Reece, S. T. et al. (2014). Type I IFN signaling triggers immunopathology in tuberculosis-susceptible mice by modulating lung phagocyte dynamics. *Eur. J. Immunol.* **44**, 2380-2393.

Elkington, P. T., Ugarte-Gil, C. A. and Friedland, J. S. (2011). Matrix metalloproteinases in tuberculosis. *Eur. Respir. J.* **38**, 456-464.

Eruslanov, E. B., Lyadova, I. V., Kondratieva, T. K., Majorov, K. B., Scheglov, I. V., Orlova, M. O. and Apt, A. S. (2005). Neutrophil responses to *Mycobacterium tuberculosis* infection in genetically susceptible and resistant mice. *Infect. Immun.* **73**, 1744-1753.

Eum, S.-Y., Kong, J.-H., Hong, M.-S., Lee, Y.-J., Kim, J.-H., Hwang, S.-H., Cho, S.-N., Via, L. E. and Barry, C. E.III. (2010). Neutrophils are the predominant

infected phagocytic cells in the airways of patients with active pulmonary TB. *Chest* **137**, 122-128.

Fallahi-Sichani, M., Schaller, M. A., Kirschner, D. E., Kunkel, S. L. and Linderman, J. J. (2010). Identification of key processes that control tumor necrosis factor availability in a tuberculosis granuloma. *PLoS Comput. Biol.* **6**, e1000778.

Ferris, M. T., Aylor, D. L., Bottomly, D., Whitmore, A. C., Aicher, L. D., Bell, T. A., Bradel-Tretheway, B., Bryan, J. T., Buus, R. J., Gralinski, L. E. et al. (2013). Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. *PLoS Pathog.* **9**, e1003196.

Flynn, J. L. and Chan, J. (2001). Immunology of tuberculosis. *Annu. Rev. Immunol.* **19**, 93-129.

Gong, J. H., Zhang, M., Modlin, R. L., Linsley, P. S., Iyer, D., Lin, Y. and Barnes, P. F. (1996). Interleukin-10 downregulates Mycobacterium tuberculosis-induced Th1 responses and CTLA-4 expression. *Infect. Immun.* **64**, 913-918.

Gopal, R., Monin, L., Torres, D., Slight, S., Mehra, S., McKenna, K. C., Fallert Junecko, B. A., Reinhart, T. A., Kolls, J., Báez-Saldaña, R. et al. (2013). S100A8/A9 proteins mediate neutrophilic inflammation and lung pathology during tuberculosis. *Am. J. Respir. Crit. Care Med.* **188**, 1137-1146.

Harper, J., Skerry, C., Davis, S. L., Tasneen, R., Weir, M., Kramnik, I., Bishai, W. R., Pomper, M. G., Nuermberger, E. L. and Jain, S. K. (2012). Mouse model of necrotic tuberculosis granulomas develops hypoxic lesions. *J. Infect. Dis.* **205**, 595-602.

Harrison, D. E., Strong, R., Sharp, Z. D., Nelson, J. F., Astle, C. M., Flurkey, K., Nadon, N. L., Wilkinson, J. E., Frenkel, K., Carter, C. S. et al. (2009). Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392-395.

Harrison, D. E., Astle, C. M., Niazi, M. K. K., Major, S. and Beamer, G. L. (2014). Genetically diverse mice are novel and valuable models of age-associated susceptibility to Mycobacterium tuberculosis. *Immun. Ageing* **11**, 24.

Higgins, D. M., Sanchez-Campillo, J., Rosas-Taraco, A. G., Lee, E. J., Orme, I. M. and Gonzalez-Juarrero, M. (2009). Lack of IL-10 alters inflammatory and immune responses during pulmonary Mycobacterium tuberculosis infection. *Tuberculosis (Edinb.)* **89**, 149-157.

Hunter, R. L. (2011). Pathology of post primary tuberculosis of the lung: an illustrated critical review. *Tuberculosis (Edinb.)* **91**, 497-509.

Hunter, R. L., Olsen, M. R., Jagannath, C. and Actor, J. K. (2006). Multiple roles of cord factor in the pathogenesis of primary, secondary, and cavitary tuberculosis, including a revised description of the pathology of secondary disease. *Ann. Clin. Lab. Sci.* **36**, 371-386.

Hunter, R. L., Actor, J. K., Hwang, S. A., Karev, V. and Jagannath, C. (2014). Pathogenesis of post primary tuberculosis: immunity and hypersensitivity in the development of cavities. *Ann. Clin. Lab. Sci.* **44**, 365-387.

Hur, Y.-G., Gorak-Stolinska, P., Ben-Smith, A., Lalor, M. K., Chaguluka, S., Dacombe, R., Doherty, T. M., Ottenhoff, T. H., Dockrell, H. M. and Crampin, A. C. (2013). Combination of cytokine responses indicative of latent TB and active TB in Malawian adults. *PLoS ONE* **8**, e79742.

Jagannath, C., Hoffmann, H., Sepulveda, E., Actor, J. K., Wetsel, R. A. and Hunter, R. L. (2000). Hypersusceptibility of A/J mice to tuberculosis is in part due to a deficiency of the fifth complement component (C5). *Scand. J. Immunol.* **52**, 369-379.

Jamil, B., Shahid, F., Hasan, Z., Nasir, N., Razzaki, T., Dawood, G. and Hussain, R. (2007). Interferon gamma/IL10 ratio defines the disease severity in pulmonary and extra pulmonary tuberculosis. *Tuberculosis (Edinb.)* **87**, 279-287.

Juffermans, N. P., Verbon, A., van Deventer, S. J. H., van Deutekom, H., Speelman, P. and van der Poll, T. (1998). Tumor necrosis factor and interleukin-1 inhibitors as markers of disease activity of tuberculosis. *Am. J. Respir. Crit. Care Med.* **157**, 1328-1331.

Keller, C., Hoffmann, R., Lang, R., Brandau, S., Hermann, C. and Ehlers, S. (2006). Genetically determined susceptibility to tuberculosis in mice causally involves accelerated and enhanced recruitment of granulocytes. *Infect. Immun.* **74**, 4295-4309.

Kira, K. and Rendell, L. (1992). The feature selection problem: traditional methods and a new algorithm. Proceedings of the National Conference on Artificial Intelligence John Wiley & Sons Ltd.

Kornaropoulos, E. N., Niazi, M. K. K., Lozanski, G. and Gurcan, M. N. (2014). Histopathological image analysis for centroblasts classification through dimensionality reduction approaches. *Cytometry A* **85**, 242-255.

Kramnik, I. (2008). Genetic dissection of host resistance to Mycobacterium tuberculosis: the sst1 locus and the Ipr1 gene. *Curr. Top. Microbiol. Immunol.* **321**, 123-148.

Kramnik, I., Demant, P. and Bloom, B. B. (1998). Susceptibility to tuberculosis as a complex genetic trait: analysis using recombinant congenic strains of mice. *Novartis Found. Symp.* **217**, 120-131; discussion 132-127.

Kramnik, I., Dietrich, W. F., Demant, P. and Bloom, B. R. (2000). Genetic control of resistance to experimental infection with virulent Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. USA* **97**, 8560-8565.

Lachenbruch, P. A. (1975). *Discriminant Analysis*. New York: Wiley Online Library.

Law, K., Weiden, M., Harkin, T., Tchou-Wong, K., Chi, C. and Rom, W. N. (1996). Increased release of interleukin-1 beta, interleukin-6, and tumor necrosis factor-

alpha by bronchoalveolar cells lavaged from involved sites in pulmonary tuberculosis. *Am. J. Respir. Crit. Care Med.* **153**, 799-804.

Leong, F. J. W.-M., Eum, S., Via, L. E. and Barry, C. E.III. (2011). Pathology of tuberculosis in the human lung. In *A Color Atlas of Comparative Pathology of Pulmonary Tuberculosis* (ed. F. J. Leong, V. Dartois and T. Dick), pp. 53-81. New York: CRC Press.

Liang, B., Guo, Y., Li, Y. and Kong, H. (2014). Association between IL-10 gene polymorphisms and susceptibility of tuberculosis: evidence based on a meta-analysis. *PLoS ONE* **9**, e88448.

Liu, H. and Setiono, R. (1996). A probabilistic approach to feature selection-a filter solution. Proceedings of International Conference on Machine Learning. **96**, 319-327.

Logan, R. W., Robledo, R. F., Recla, J. M., Philip, V. M., Bubier, J. A., Jay, J. J., Harwood, C., Wilcox, T., Gatti, D. M., Bult, C. J. et al. (2013). High-precision genetic mapping of behavioral traits in the diversity outbred mouse population. *Genes Brain Behav.* **12**, 424-437.

Lyadova, I. V., Tsiganov, E. N., Kapina, M. A., Shepelkova, G. S., Sosunov, V. V., Radaeva, T. V., Majorov, K. B., Shmitova, N. S., van den Ham, H.-J., Ganusov, V. V. et al. (2010). In mice, tuberculosis progression is associated with intensive inflammatory response and the accumulation of Gr-1^{dim} cells in the lungs. *PLoS ONE* **5**, e10469.

Major, S., Turner, J. and Beamer, G. (2013). Tuberculosis in CBA/J mice. *Vet. Pathol.* **50**, 1016-1021.

Marzo, E., Vilaplana, C., Tapia, G., Diaz, J., Garcia, V. and Cardona, P.-J. (2014). Damaging role of neutrophilic infiltration in a mouse model of progressive tuberculosis. *Tuberculosis* **94**, 55-64.

Medina, E. and North, R. J. (1998). Resistance ranking of some common inbred mouse strains to *Mycobacterium tuberculosis* and relationship to major histocompatibility complex haplotype and Nramp1 genotype. *Immunology* **93**, 270-274.

Millington, K. A., Innes, J. A., Hackforth, S., Hinks, T. S. C., Deeks, J. J., Dosanjh, D. P. S., Guyot-Revol, V., Gunatheesan, R., Klenerman, P. and Lalvani, A. (2007). Dynamic relationship between IFN-gamma and IL-2 profile of Mycobacterium tuberculosis-specific T cells and antigen load. *J. Immunol.* **178**, 5217-5226.

Modlin, R. L. and Bloom, B. R. (2013). TB or not TB: that is no longer the question. *Sci. Transl. Med.* **5**, 213-216.

Nandi, B. and Behar, S. M. (2011). Regulation of neutrophils by interferon-γ limits lung inflammation during tuberculosis infection. *J. Exp. Med.* **208**, 2251-2262.

Niazi, M. K. K., Beamer, G. and Gurcan, M. N. (2014). Detecting and characterizing cellular responses to *Mycobacterium tuberculosis* from histology slides. *Cytometry A* **85**, 151-161.

Noda, I. (1993). Generalized two-dimensional correlation method applicable to infrared, Raman, and other types of spectroscopy. *Appl. Spectroscopy* **47**, 1329-1336.

Nouailles, G., Dorhoi, A., Koch, M., Zerrahn, J., Weiner, J., III, Faé, K. C., Arrey, F., Kuhlmann, S., Bandermann, S., Loewe, D. et al. (2014). CXCL5-secreting pulmonary epithelial cells drive destructive neutrophilic inflammation in tuberculosis. *J. Clin. Invest.* **124**, 1268-1282.

Nunes-Alves, C., Booty, M. G., Carpenter, S. M., Jayaraman, P., Rothchild, A. C. and Behar, S. M. (2014). In search of a new paradigm for protective immunity to TB. *Nat. Rev. Microbiol.* **12**, 289-299.

Palanisamy, G. S., Kirk, N. M., Ackart, D. F., Shanley, C. A., Orme, I. M. and Basaraba, R. J. (2011). Evidence for oxidative stress and defective antioxidant response in guinea pigs with tuberculosis. *PLoS ONE* **6**, e26254.

Pan, H., Yan, B.-S., Rojas, M., Shebzukhov, Y. V., Zhou, H., Kobzik, L., Higgins, D. E., Daly, M. J., Bloom, B. R. and Kramnik, I. (2005). Ipr1 gene mediates innate immunity to tuberculosis. *Nature* **434**, 767-772.

Pichugin, A. V., Yan, B.-S., Sloutsky, A., Kobzik, L. and Kramnik, I. (2009). Dominant role of the *sst1* locus in pathogenesis of necrotizing lung granulomas during chronic tuberculosis infection and reactivation in genetically resistant hosts. *Am. J. Pathol.* **174**, 2190-2201.

Png, E., Alisjahbana, B., Sahiratmadja, E., Marzuki, S., Nelwan, R., Adnan, I., van de Vosse, E., Hibberd, M., van Crevel, R., Ottenhoff, T. H. M. et al. (2012). Polymorphisms in SP110 are not associated with pulmonary tuberculosis in Indonesians. *Infect. Genet. Evol.* **12**, 1319-1323.

Rangaka, M. X., Wilkinson, K. A., Glynn, J. R., Ling, D., Menzies, D., Mwansa-Kambafwile, J., Fielding, K., Wilkinson, R. J. and Pai, M. (2012). Predictive value of interferon-gamma release assays for incident active tuberculosis: a systematic review and meta-analysis. *Lancet Infect. Dis.* **12**, 45-55.

Ray, J. C. J., Flynn, J. L. and Kirschner, D. E. (2009). Synergy between individual TNF-dependent functions determines granuloma performance for controlling Mycobacterium tuberculosis infection. *J. Immunol.* **182**, 3706-3717.

Recla, J. M., Robledo, R. F., Gatti, D. M., Bult, C. J., Churchill, G. A. and Chesler, E. J. (2014). Precise genetic mapping and integrative bioinformatics in Diversity Outbred mice reveals Hydin as a novel pain gene. *Mamm. Genome* **25**, 211-222.

Redford, P. S., Murray, P. J. and O'Garra, A. (2011). The role of IL-10 in immune regulation during M. tuberculosis infection. *Mucosal. Immunol.* **4**, 261-270.

Rokach, L. and Maimon, O. (2005). Top-down induction of decision trees classifiers-a survey. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* **35**, 476-487.

Sahiratmadja, E., Alisjahbana, B., de Boer, T., Adnan, I., Maya, A., Danusantoso, H., Nelwan, R. H. H., Marzuki, S., van der Meer, J. W., van Crevel, R. et al. (2007). Dynamic changes in pro- and anti-inflammatory cytokine profiles and gamma interferon receptor signaling integrity correlate with tuberculosis disease activity and response to curative treatment. *Infect. Immun.* **75**, 820-829.

Scholkopft, B. and Mullert, K.-R. (1999). Fisher discriminant analysis with kernels. *Neural Netw. Signal Process. IX.* **1**, 41-48.

Sissons, J., Yan, B.-S., Pichugin, A. V., Kirby, A., Daly, M. J. and Kramnik, I. (2009). Multigenic control of tuberculosis resistance: analysis of a QTL on mouse chromosome 7 and its synergism with sst1. *Genes Immun.* **10**, 37-46.

Suter-Riniker, F., Berger, A., Mayor, D., Bittel, P., Iseli, P. and Bodmer, T. (2011). Clinical significance of interleukin-2/gamma interferon ratios in Mycobacterium tuberculosis-specific T-cell signatures. *Clin. Vaccine Immunol.* **18**, 1395-1396.

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

Thye, T., Browne, E. N., Chinbuah, M. A., Gyapong, J., Osei, I., Owusu-Dabo, E., Niemann, S., Rusch-Gerdes, S., Horstmann, R. D. and Meyer, C. G. (2006). No associations of human pulmonary tuberculosis with Sp110 variants. *J. Med. Genet.* **43**, e32.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B*, **58**, 267-288.

Tosh, K., Campbell, S. J., Fielding, K., Sillah, J., Bah, B., Gustafson, P., Manneh, K., Lisse, I., Sirugo, G., Bennett, S. et al. (2006). Variants in the SP110 gene are associated with genetic susceptibility to tuberculosis in West Africa. *Proc. Natl. Acad. Sci. USA* **103**, 10364-10368.

Tsiganov, E. N., Verbina, E. M., Radaeva, T. V., Sosunov, V. V., Kosmiadi, G. A., Nikitina, I. Y. and Lyadova, I. V. (2014). Gr-1dimCD11b+ immature myeloid-derived suppressor cells but not neutrophils are markers of lethal tuberculosis infection in mice. *J. Immunol.* **192**, 4718-4727.

Turner, J., Gonzalez-Juarrero, M., Ellis, D. L., Basaraba, R. J., Kipnis, A., Orme, I. M. and Cooper, A. M. (2002). In vivo IL-10 production reactivates chronic pulmonary tuberculosis in C57BL/6 mice. *J. Immunol.* **169**, 6343-6351.

Ullman-Cullere, M. H. and Foltz, C. J. (1999). Body condition scoring: a rapid and accurate method for assessing health status in mice. *Lab. Anim. Sci.* **49**, 319-323.

Vesosky, B., Rottinghaus, E. K., Stromberg, P., Turner, J. and Beamer, G. (2010). CCL5 participates in early protection against *Mycobacterium tuberculosis*. *J. Leukoc. Biol.* **87**, 1153-1165.

Vilaplana, C. and Cardona, P.-J. (2014). The lack of a big picture in tuberculosis: the clinical point of view, the problems of experimental modeling and immunomodulation. The factors we should consider when designing novel treatment strategies. *Front. Microbiol.* **5**, 55.

Wilkinson, T. S., Morris, A. C., Kefala, K., O'Kane, C. M., Moore, N. R., McAuley, D. F., Dhaliwal, K., Walsh, T. S., Haslett, C., Sallenave, J.-M. et al. (2012). Ventilator-associated pneumonia is characterized by excessive release of neutrophil proteases in the lung. *Chest* **142**, 1425-1432.

Yan, B.-S., Kirby, A., Shebzukhov, Y. V., Daly, M. J. and Kramnik, I. (2006). Genetic architecture of tuberculosis resistance in a mouse model of infection. *Genes Immun.* **7**, 201-210.

Yang, H., Bell, T. A., Churchill, G. A. and Pardo-Manuel de Villena, F. (2007). On the subspecific origin of the laboratory mouse. *Nat. Genet.* **39**, 1100-1107.

Yang, H., Wang, J. R., Didion, J. P., Buus, R. J., Bell, T. A., Welsh, C. E., Bonhomme, F., Yu, A. H.-T., Nachman, M. W., Pialek, J. et al. (2011). Subspecific origin and haplotype diversity in the laboratory mouse. *Nat. Genet.* **43**, 648-655.

Yang, C.-T., Cambier, C. J., Davis, J. M., Hall, C. J., Crosier, P. S. and Ramakrishnan, L. (2012). Neutrophils exert protection in the early tuberculous granuloma by oxidative killing of mycobacteria phagocytosed from infected macrophages. *Cell Host Microbe* **12**, 301-312.

Zhang, J., Chen, Y., Nie, X.-B., Wu, W.-H., Zhang, H., Zhang, M., He, X.-M. and Lu, J.-X. (2011). Interleukin-10 polymorphisms and tuberculosis susceptibility: a meta-analysis [Review article]. *Int. J. Tuberc. Lung Dis.* **15**, 594-601.

Zhang, W., Korstanje, R., Thaisz, J., Staedtler, F., Harttman, N., Xu, L., Feng, M., Yanas, L., Yang, H., Valdar, W. et al. (2012). Genome-wide association mapping of quantitative traits in outbred mice. *G3* **2**, 167-174.

Zheng, Z. and Liu, H. C. (2007). Spectral feature selection for supervised and unsupervised learning. ICML Proceedings of 24th International Conference on Machine Learning, Corvalis, Oregon, USA. 1151-1157.