

Genomic responses to inflammation in mouse models mimic humans: We concur, apples to oranges comparisons won't do

We read with interest the report of Takao and Miyakawa (1), who compare genomic responses to inflammatory challenges in humans and mice, reanalyzing the datasets originally reported by Seok et al. (2), but reaching a diametrically opposite conclusion. In line with its goal to establish the landscape of gene expression and regulation in the immune system, the ImmGen consortium has run focused human/mouse comparisons (3), and we also reanalyzed the datasets of Seok et al. We largely concur with the conclusions reached by Takao and Miyakawa, but wish to make a few additional points.

Takao and Miyakawa point out computational issues in the article by Seok et al., and we agree with this assessment. The Pearson correlation is a simplistic and inadequate tool for such comparisons. Tools that assess the replication rate are far better suited (e.g., Storey's π_1 commonly used for expression quantitative trait loci) (4). There was no attempt to control for batch, genetic, or origin-of-sample confounders. The assertion that diverse inflammatory stresses result in similar genomic profiles was confounded by using the same controls in all fold-change calculations (shared denominator artifact).

We did not follow Takao and Miyakawa in restricting the analysis to transcripts whose abundance changes in both species. For an unbiased perspective on relatedness, one also needs to consider those that change in only one side of the comparison. However, we still found highly significant correlations ($R = 0.38-0.45$; $P = 10^{-75}$ to 10^{-86}) between changes occurring in humans and mice. Importantly, this required correct matching of time frames, an aspect that was largely ignored in Seok et al., whose analyses mainly considered maximum fold-changes, irrespective of time. Activation of inflammatory pathways occurs in carefully orchestrated sequences, within minutes to a few hours.

For example, our analyses of the IFN response show that $\sim 90\%$ of genes induced in one species (5% false discovery rate) are also induced in the other, with very similar strength and timing of induction.

Beyond these analytical concerns, we contend that these datasets are altogether unsuitable for interspecies comparison. (i) Profiling whole blood leukocytes is inappropriate, as many of the changes in RNA abundance merely reflect varying proportions of cell types (e.g., monocytes increase markedly in blood after trauma) rather than transcriptional responses. Worse, neutrophils are far more abundant in human than mouse blood, such that human data mainly reflect responses in neutrophils and mouse data reflect responses in lymphocytes. Matching purified cells is essential for valid interspecies comparison. (ii) The time frames are badly mismatched and mostly too late (days to weeks) for the human data to yield valuable insight on therapeutic targets for acute treatment. (iii) Treatment effects in humans were ignored, but certainly confound the comparison with untreated mice. (iv) The large component of genetic variation in the human dataset was overlooked, relative to the constant background of the inbred mouse comparator.

In a sense, the validity of mouse models is a matter of "glass-half-full" perspective. However, although perhaps not their intent, the work of Seok et al. was sensationally portrayed as an empty glass in the lay press and at the science policy level. Translation of basic discoveries from animal models to human therapy faces many challenges, and it has become a mantra that treatments pioneered in mice fail in human patients. However, the issues with Seok et al. are a good illustration of the problems that often plague these attempts, with little regard to dose, time, protocol, or genetic variables. When

these are carefully matched, it is found that mouse results are predictive of many successes in immuno-inflammatory therapy (5), as illustrated by the striking success of immune checkpoint blockade in cancer therapy. More important than putting an abstract number on how much a mouse model mimics the human condition will be to accurately chart which cellular or genomic elements match in humans and animal models. This will not be achieved with loose amalgams of inflammatory diseases, cells, and conditions.

Tal Shay^{a,1}, James A. Lederer^b, and Christophe Benoist^c

^aDepartment of Life Sciences, Ben-Gurion University of the Negev, Be'er Sheva, 8410501 Israel; ^bDepartment of Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115; and ^cDivision of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115

1 Takao K, Miyakawa T (2015) Genomic responses in mouse models greatly mimic human inflammatory diseases. *Proc Natl Acad Sci USA* 112:1167–1172.

2 Seok J, et al.; Inflammation and Host Response to Injury, Large Scale Collaborative Research Program (2013) Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci USA* 110(9):3507–3512.

3 Shay T, et al.; ImmGen Consortium (2013) Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc Natl Acad Sci USA* 110(8):2946–2951.

4 Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. *Proc Natl Acad Sci USA* 100(16):9440–9445.

5 Shoda LK, et al. (2005) A comprehensive review of interventions in the NOD mouse and implications for translation. *Immunity* 23(2):115–126.

Author contributions: T.S., J.A.L., and C.B. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed. Email: talshay@bgu.ac.il.