LETTER

Reply to Yablonka et al.: Parity in data processing is essential in correlation analyses

Our recent study (1) has compelled us to reach a conclusion that differs from the widely held idea that chemical features of odors can be systematically mapped to spatial locations in the olfactory bulb (2–4). In an alternative analysis using our data, Yablonka et al. (5) claim to observe substantial correlations between the physicochemical descriptors and odor-evoked responses, which challenge our conclusion. This discrepancy, as we show below, results from imparity in data processing and data selection.

First, our analyses were based on the equitable treatment of all parameters without biasing toward a particular method. We processed two sets of variables consistently to avoid potential skewing of data. In our study, we computed cosine distances (similar to Pearson) for both odor-evoked responses and chemical structures. The choice was made to minimize variability among experiments. Importantly, for high-dimensional data, Euclidean distance does not distinguish between differences along a single axis from those across many dimensions, whereas cosine and Pearson distances do. Regardless, using Euclidean distance in fact produces a weaker correlation than cosine distance (Fig. 1*A*). Yoblonka et al. (5) take a different approach. They use Pearson distance for odor response, but Euclidean distance for structural descriptors. The observed correlation can be explained by the imparity in distance metrics.

Second, contrary to what Yablonka et al. (5) suggest, we have performed descriptor normalization. When normalization was performed among the odors used in the experiments, it had little impact on correlation. In contrast, Yablonka et al. (5) perform normalization among 1,370 odors. This practice would be valid if responses to the 1,370 odors were sampled and the data were transformed similarly. However, our dataset was only from the sampling of 67 odors. We show here that normalizing against odors outside the experimental set severely skews the distribution and, importantly, diminishes the distance values among odors (Fig. 1 *B*–*E*). The redistribution leads to higher correlation values (Fig. 1 *F* and *G*). Thus, the imparity in data normalization and comparison have led to unintended skewing of distance measurements and exaggerated correlation values.

Third, our study suggested that chemotopy breaks down when tested against large odor sets. Consistent with this notion, Yablonka et al. (5) show that eliminating odors leads to increases in correlation. This observation, however, hardly suggests a high correlation between odor descriptor and response. In contrast, it suggests that the eliminated odors do not conform to the rules derived from other odors. In our analysis, we also eliminated stimuli with little response (empty arrays). Because we used responses across different concentrations to construct the response vector for an odor, fewer odors were eliminated. Constructing larger vectors also provides a critical test for the predictive power of odor descriptors. If structure predicts response, strong correlations should emerge. We did not observe strong correlations.

Finally, most of the correlation values from the Yablonka et al. (5) analysis are relatively small (<0.3). Although they are significant statistically, their power in predicting odor responses is not. Therefore, although a subset of chemically similar odors can activate a common set of glomeruli, the rule cannot be generalized.

***Limei Ma*[a,1], *Qiang Qiu*[a], *Stephen Gradwohl*[a], *Aaron Scott*[a], *Elden Q. Yu*[a], *Richard Alexander*[a], *Winfried Wiegraebe*[a], *and C. Ron Yu*[a,b,1]**

[a]*Stowers Institute for Medical Research, Kansas City, MO 64110; and* [b]*Department of Anatomy and Cell Biology, University of Kansas Medical Center, Kansas City, KS 66160*

1. Ma L, et al. (2012) Distributed representation of chemical features and tunotopic organization of glomeruli in the mouse olfactory bulb. *Proc Natl Acad Sci USA* 109: 5481–5486.
2. Mori K, Sakano H (2011) How is the olfactory map formed and interpreted in the mammalian brain? *Annu Rev Neurosci* 34:467–499.
3. Johnson BA, Leon M (2007) Chemotopic odorant coding in a mammalian olfactory system. *J Comp Neurol* 503:1–34.
4. Haddad R, et al. (2008) A metric for odorant comparison. *Nat Methods* 5:425–429.
5. Yablonka A, Sobel N, Haddad R (2012) Odorant similarity in the mouse olfactory bulb. *Proc Natl Acad Sci USA* 109:E2916–E2917.
6. Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD Odor coding by a Mammalian receptor repertoire, 10.1126/scisignal.2000016.

Author contributions: L.M. and C.R.Y. designed research; L.M. and Q.Q. performed research; L.M., S.G., A.S., and E.Q.Y. contributed new reagents/analytic tools; L.M., Q.Q., E.Q.Y., R.A., W.W., and C.R.Y. analyzed data; and L.M. and C.R.Y. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: cry@stowers.org or lim@stowers.org.

www.pnas.org/cgi/doi/10.1073/pnas.1212266109

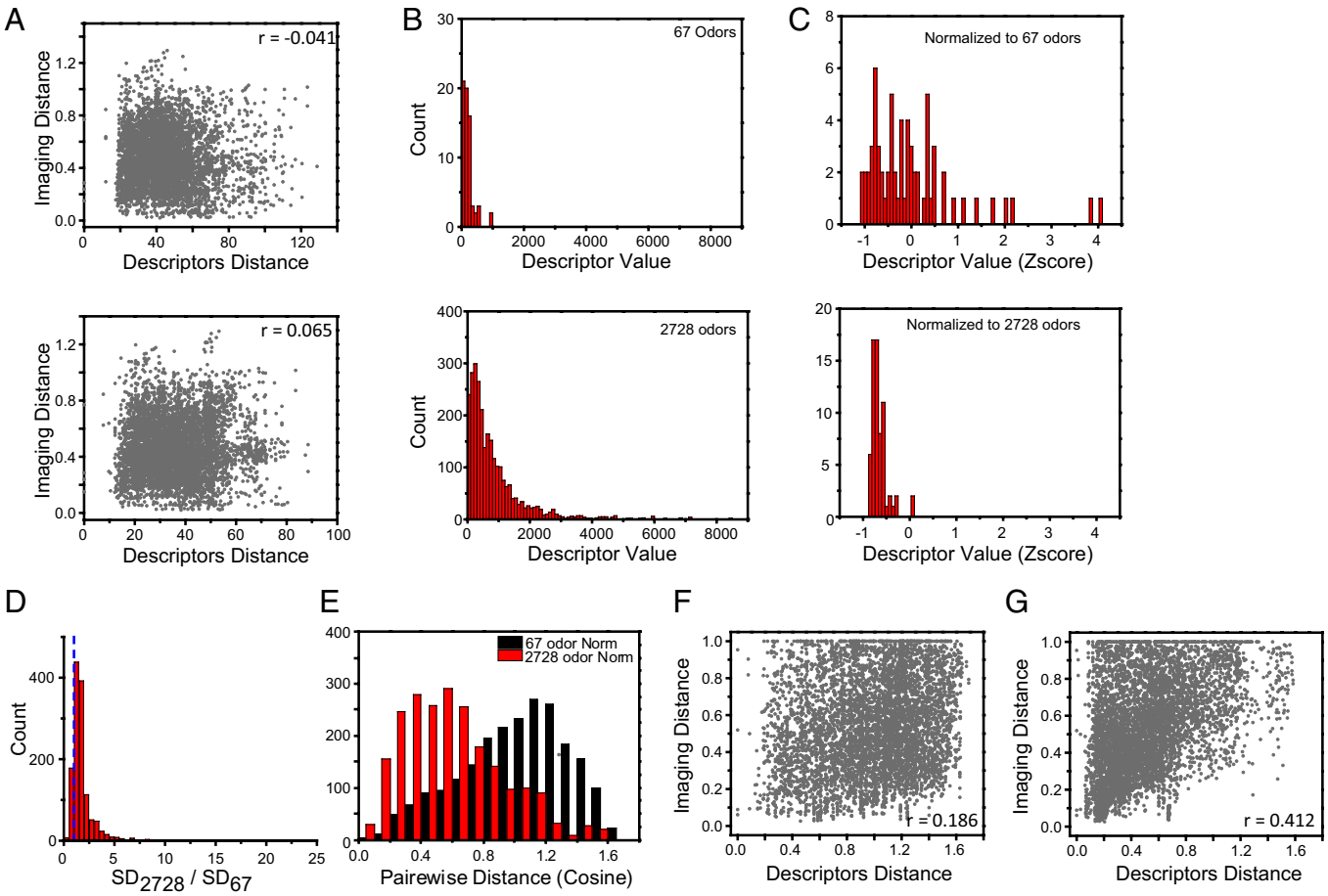

Fig. 1. (*A*) Correlation analysis of data using Euclidean distance for both sets of variables. *Upper*: Descriptor distance normalized within the 67 odors used in the experiments. *Lower*: Descriptors normalized to a set of 2,728 odors (6). (*B*) An example of the analysis performed on descriptors. Descriptor 78 is shown here. The histogram of the 67 odors follows a very different distribution (*Upper*) from the one obtained from the 2,728 odors (*Lower*). (*C*) After normalization with the mean and standard deviation (SD) within the 67 odors (*Upper*) or 2,728 odors (*Lower*), the 67 odors have two different distributions on descriptor 78. Descriptor 78 becomes less useful for differentiating odors after being normalized to the full odor set, but it still contributed in the odor distance. It made all odors similar (*Lower*). (*D*) Distribution histogram for ratios of SD for all descriptors. The SD for each descriptor is calculated using 67 odors or 2,728 odors. Vertical line indicates ratio of 1. (*E*) Odor distance distribution for the 67 odors after normalization to themselves (black) or to the full set of odors (red). The latter shifts the distance distribution leftward. (*F* and *G*) Correlation between glomeruli response and odor structure when odor distance is normalized among the experimental set (*F*) or to the large set (*G*). Cosine distances are used in both axes.