Supplementary Note 1: Algorithm for handling the 3D point cloud

We developed a simple algorithm to calculate the normal directional vector $\mathbf{n}$ and the patch area A of the demolding surface in the 3D point cloud data set that considers spatially closest 4 points. Example plot of the generated values are shown in Supplementary Figure 1.

–Algorithm for $\mathbf{n}$ and A from the 3D point cloud:

1. For each point $i \in 1, ..., n$, search the closest 4 points and sort them according to the right-hand rule about the center point. Index of the 4 points are given as $j = 1, 2, 3, 4$.

2. Construct 4 triangles, each of which consists of 3 points, the center point i, j, and $(j+1)$. Calculate the unit normal vector of each triangle as

$$\mathbf{n}_{j,(j+1)} = (\mathbf{v}_{i,j} \times \mathbf{v}_{i,(j+1)})/|\mathbf{v}_{i,j} \times \mathbf{v}_{i,(j+1)}| \tag{1}$$

   where $\mathbf{v}_{i,j}$ is a vector from point i to point j, $(j = 1, 2, 3, 4)$. Note that when $j = 4$, $(j+1)$ is assumed 1.

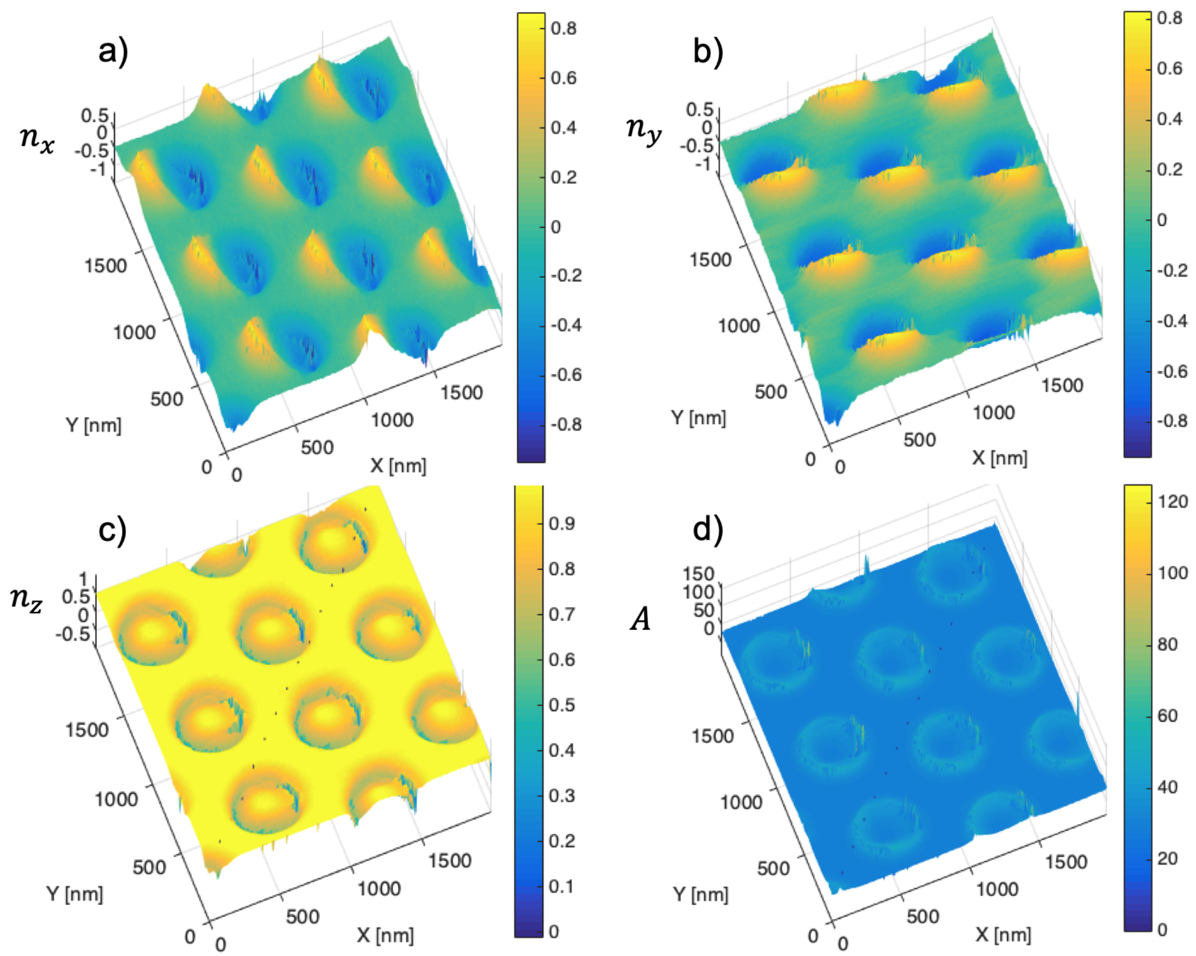
3. Calculate the averaged normal vector $\mathbf{n}^{(i)}$ and the surface area $A_{(i)}$ associated with the point i (as illustrated in Fig. 1b of main article)

$$\mathbf{n}^{(i)} = \frac{1}{4}\sum_{j=1}^{4} \mathbf{n}_{j,(j+1)}, \tag{2}$$

$$A_{(i)} = \frac{1}{8}\sum_{j=1}^{4} |\mathbf{v}_{i,j} \times \mathbf{v}_{i,(j+1)}| \tag{3}$$

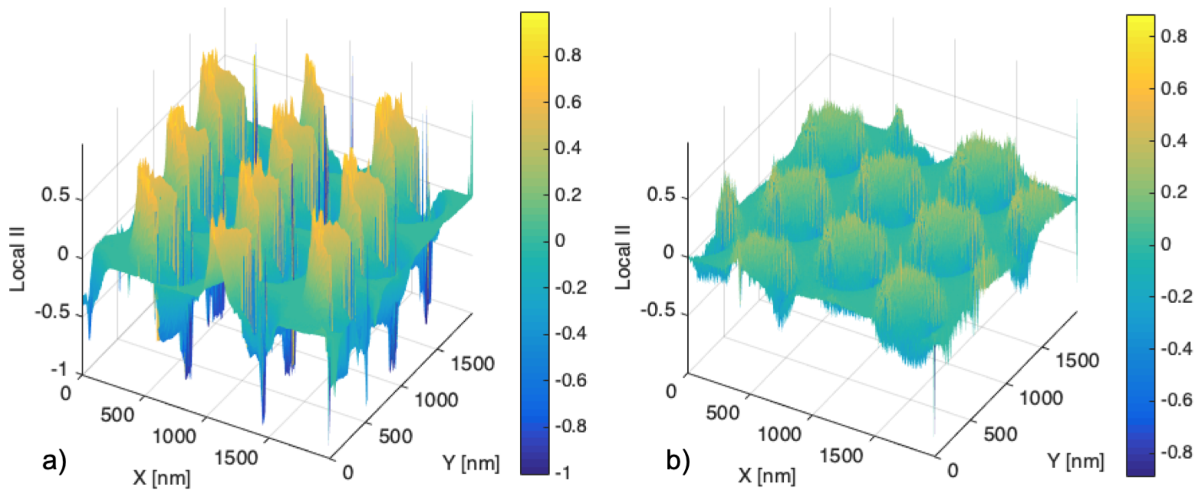
Supplementary Note 2: Convolved information index with various information ranges

As explained in the main article, the role of the convolved information index ($\overline{II}^{(l)}$) is to offer a quantified scalar that represents physical information within the radius of information range



Supplementary Figure 1. Example plots of basic information. **a-c** The calculated unit normal vector $\mathbf{n} = (n_x, n_y, n_z)$; **d** patch area A [nm^2] around a point. Raw data of Fig. 2a, d in (*1*) are used.

($L^{(l)}$) of the 3D point cloud. Also, the spatial integration of the ”local” information index is necessary to capture the interactions of physical actions over nearby spaces. For instance, friction is not a point-point action, but a surface-surface action in reality. The local index has peaks and lacks smoothness over the space as Supplementary Figure 2 shows an example of local index without spatial integration. Since we used the Gaussian averaging process, if the radius is small, we see many peaked values due to greater importance on the close points’ physical quantities. On the contrary, if the radius is large, we can embrace a broader range’s information such as

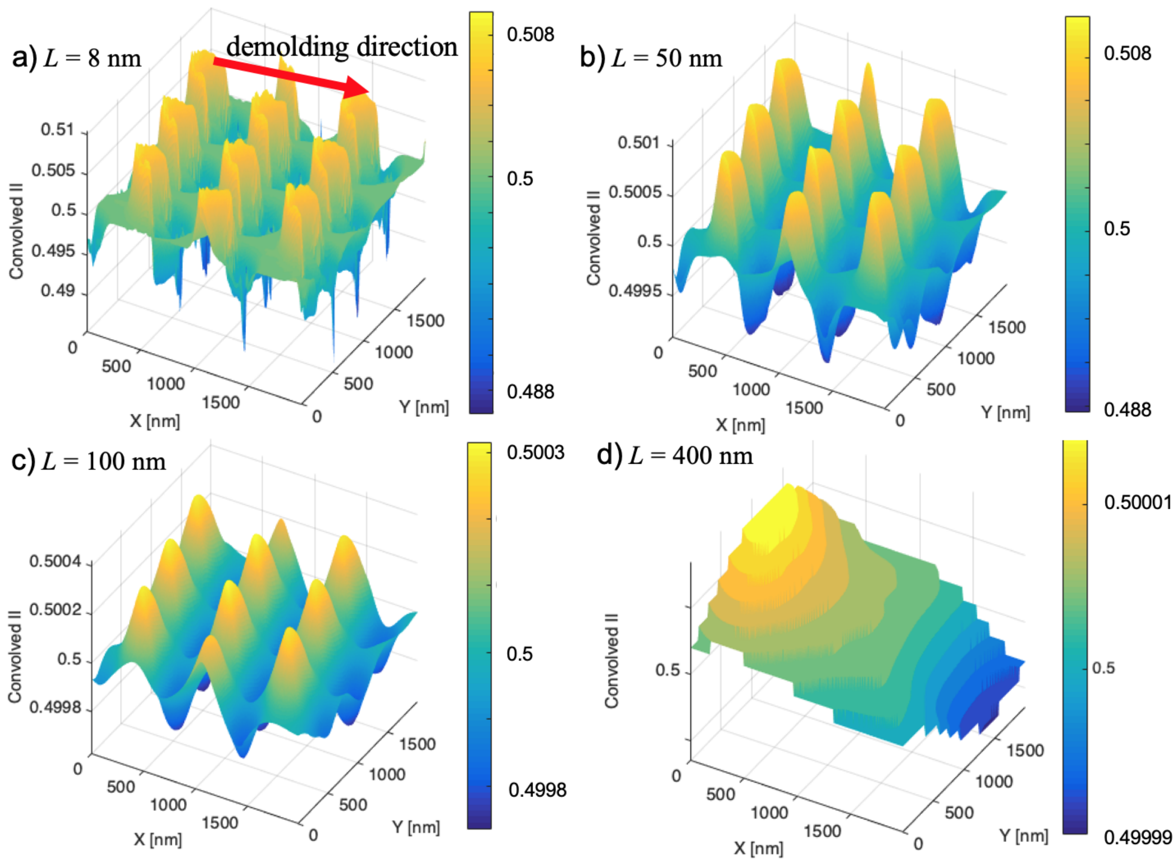


Supplementary Figure 2. Example plots of the local information index (II_{local}). **a** II_{local} calculated with raw data from the experiment (Fig. 2a, d of (*1*)) of rightward demolding parallel to the X, and **b** from the experiment of diagonal demolding (Fig. 2b, e of (*1*)).

the global demolding direction. Supplementary Figure 3 compares several cases of ($\overline{II}^{(l)}$) with varying radii. This case corresponds to the horizontal (parallel to the X axis) demolding case in the main article. To some extent, $\overline{II}^{(l)}$ provides new generic coordinates for machine learning (ML) to train and explore. Also, physical principles and scientists' experiences can be infused into the ML by introducing new forms of information indices. As long as it can quantify physical knowledge through a distinguishable value, any form of information index can be suitable in the proposed framework. Then, ML's learning and searching power will take care of the complexity and high dimensionality of the search spaces of the introduced information indices. Thus, this framework suggests a synergistic bridge from ML methods to mechanics, physics, and broader science.

Supplementary Note 3: Identified best-so-far link functions

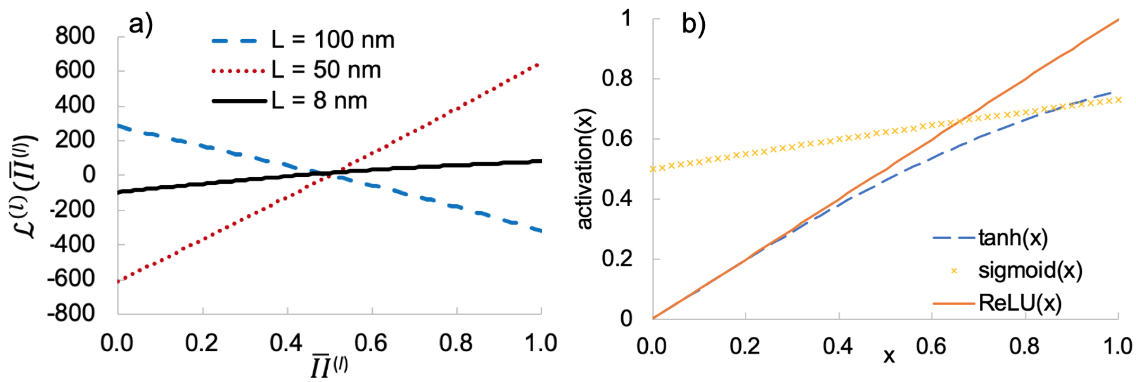
Supplementary Figure 4 compares the identified link functions (LFs) and the popular activation functions in ML methods. Both plots are shown in the meaningful ranges of $\overline{II}^{(l)} \in [0, 1]$ for



Supplementary Figure 3. Example plots of convolved information index $\overline{II}^{(l)}$ with varying $L^{(l)}$. **a** With $L^{(l)} = 8$ nm, there are many peaks and localities; **b-c** The peaks are gradually smoothed out with a large range $L^{(l)} \geq 8$ nm. With radii of 50 nm and 100 nm, the smoothing effect is becoming strong; **d** With a too large radius 400 nm, the information index is over-smoothed. Raw data of Fig. 2a, d in (*1*) are used.

facile comparison. It should be noted that the shapes of typical activation functions (Supplementary Figure 4b) are pre-defined and fixed throughout learning, but the LFs of this study are constantly evolving. Although our particular feasibility tests gave simple shapes (Supplementary Figure 4a) to LFs, it may take highly nonlinear shapes by virtue of the adopted cubic spline basis.

The decomposition of LF into cubic spline basis helps elucidate dominant relation between II and target physics. For instance, the LF of $L^{(2)} = 50$ nm is enlarged in Supplementary Figure 5 along with its detailed, decomposed basis while those of $L^{(3)} = 100$ nm is presented in Supplementary Figure 6. Comparing these decomposed basis appears to help identify linear or nonlinear dominance of each convolved information index and their contributions to the target response.



Supplementary Figure 4. Comparison of the identified link functions (LFs) and typical activation functions. **a** Identified best-so-far LFs, $\mathcal{L}^{(l)}(\overline{II}^{(l)})$ with $L^{(1)} = 8$ nm, $L^{(2)} = 50$ nm, and $L^{(3)} = 100$ nm; **b** For comparison, example plots of typical activation functions of hyperbolic tangent (tanh), sigmoid (or logistic), and rectified linear unit (ReLU) are shown.

Supplementary Note 4: Statistical explanation for feasibility of the proposed framework

We assume the observed electric potential data set $\Delta\mathbf{V} = \{\Delta V_{(i)}\}_{i=1}^{n}$ to be composed of n independent and identically distributed samples. At current time (t) in the sequence of a Bayesian update, consider the global log-likelihood (denoted as L_n) that consists of log-likelihoods of

Supplementary Table 1. Identified internal parameters of the best-so-far link functions

l	$L^{(l)}$	$\mathbf{a}^{(l)} = \{a_1, a_2, a_3, a_4, a_5\}^{(l)}$: cubic splines	$\mathbf{x}^{*(l)} = \{x_1^*, x_2^*, x_3^*\}^{(l)}$: knots
1	8	$\{-82.3529, 176.471, -2647.06, -2105.06, -1870.59\}$	$\{0.020915, 0.335948, 0.79085\}$
2	50	$\{-623.529, 1258.82, -1000, -2552.94, -2717.65\}$	$\{0.183007, 0.419608, 0.746405\}$
3	100	$\{294.118, -600, -2176.47, -435.294, 2952.94\}$	$\{0.103268, 0.562092, 0.734641\}$

individual data points

$$L_n(\theta^{(t)}; \Delta V_{(1)}, ..., \Delta V_{(n)}) = \sum_{i=1}^{n} \ln[P_\theta(\Delta V_{(i)})]. \tag{4}$$

Referring to Eq. (12) in the main article, the individual material point's contribution to the raw cost, denoted as $\mathcal{E}_{(i)}$, under the assumption of multiplicative combination of n_l LFs is given by

$$\mathcal{E}_{(i)} = [\Delta V_{(i)} - \Delta V_{(i)}(\theta^{(l)})]^2 = \left[\Delta V_{(i)} - \prod_{l=1}^{n_l} \mathcal{L}^{(l)}(\overline{II}^{(l)}_{(i)};\ \theta^{(l)})\right]^2 \tag{5}$$

In the adopted fitness proportionate probability (FPP) in Eq. (16) of main article, the probability of next parent selection is simply proportional to the inverse error, and thus the log-likelihood and the individual error terms has a clear relation as

$$\ln[P_\theta(\Delta V_{(i)})] \propto \ln(\mathcal{E}^{-1}_{(i)}) = -2\ln\left|\Delta V_{(i)} - \prod_{l=1}^{n_l} \mathcal{L}^{(l)}(\overline{II}^{(l)}_{(i)};\ \theta^{(l)})\right| \tag{6}$$

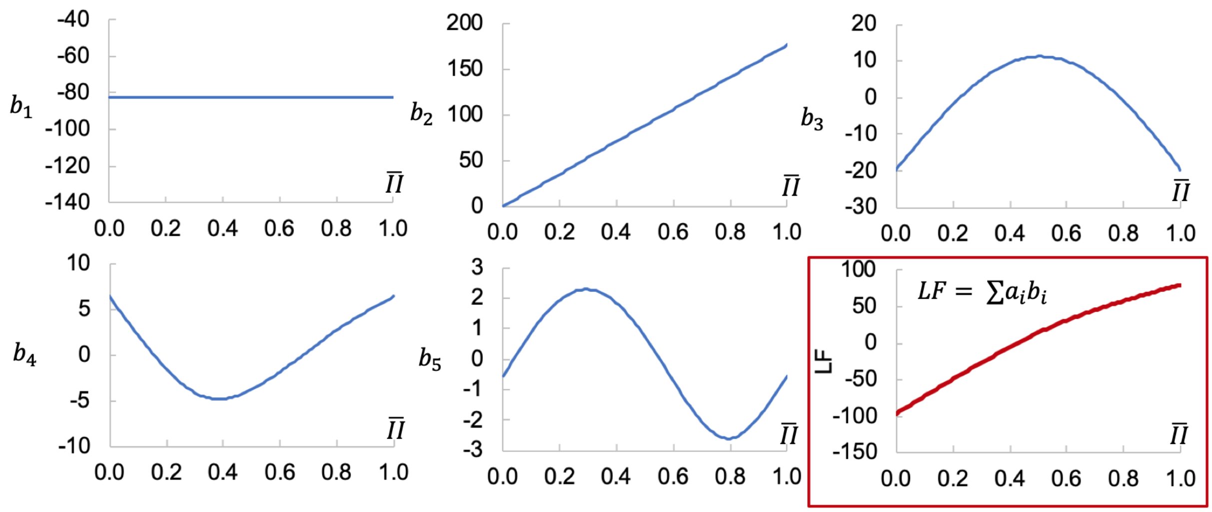
$$= -2\ln\left|\Delta V_{(i)}\right|\left|1 - \frac{\prod_{l=1}^{n_l} \mathcal{L}^{(l)}(\overline{II}^{(l)}_{(i)}; \theta^{(l)})}{\Delta V_{(i)}}\right| \tag{7}$$

$$= -2\ln|\Delta V_{(i)}| - 2\ln\left|1 - \frac{\prod_{l=1}^{n_l} \mathcal{L}^{(l)}(\overline{II}^{(l)}_{(i)}; \theta^{(l)})}{\Delta V_{(i)}}\right| \tag{8}$$

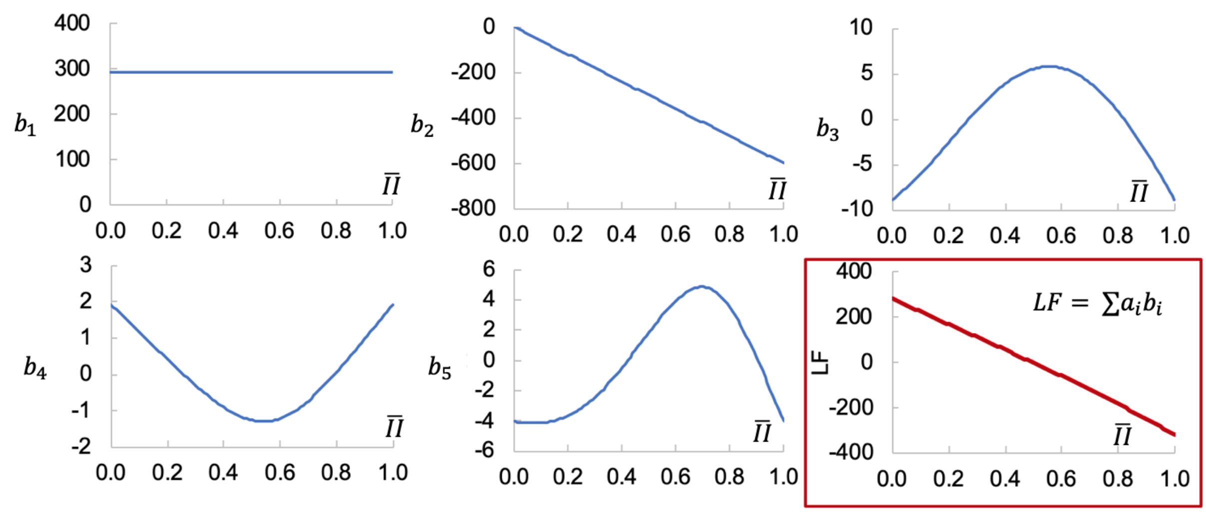
Neglecting the observational noise effects, the first term of Eq. (8) is assumed to be constant. The adopted evolutionary algorithm continues to improve the link functions such that individuals, candidates for $\theta^{(l)}$, with large errors (i.e. $\left|\prod_{l=1}^{n_l} \mathcal{L}^{(l)}(\overline{II}^{(l)}_{(i)}; \theta^{(l)})/\Delta V_{(i)}\right| > 1.0$) will gradually disappear via the FPP-based spawning, natural selection, and inheritance. Therefore, this evolution eventually promotes $\left|\prod_{l=1}^{n_l} \mathcal{L}^{(l)}(\overline{II}^{(l)}_{(i)}; \theta^{(l)})/\Delta V_{(i)}\right| \rightarrow 1.0$, which will lead the second term of Eq. (8) to 0. Therefore, the proposed framework may translate into a maximization of the log-likelihood of the entire data points:

$$L_n(\theta^{*(t)}; \Delta \mathbf{V}) = \sup_{\theta \in \boldsymbol{\Theta}} \sum_{i=1}^{n} \ln[P_\theta(\Delta V_{(i)})] \tag{9}$$

where $\theta^{*(t)}$ means the best-so-far parameters of hidden rules up to the current time (t) which will evolve at $(t+1)$ via a Bayesian update; $\boldsymbol{\Theta}$ is a parameter space.



Supplementary Figure 5. Decomposition of the cubic spline basis. The best-so-far link function (LF) after Bayesian learning with raw data sets of Fig. 2a, d and Fig. 2b, e in (*1*). The right-bottom panel shows the weighted summation of 5 bases b_1 through b_5 of LF ($L = 8$ nm) with a_i's in Supplementary Table 1. A slightly nonlinear, positively linear relation is identified by this LF.



Supplementary Figure 6. Decomposition of the cubic spline basis. The best-so-far link function (LF) after Bayesian learning with raw data sets of Fig. 2a, d and Fig. 2b, e in (*1*). The right-bottom panel shows the weighted summation of 5 bases b_1 through b_5 of LF ($L = 100$ nm) with a_i's in Supplementary Table 1. A negatively linear relation (i.e. b_2) appears dominant in this LF.

Supplementary References

1. Li, Q., Peer, A., Cho, I., Biswas, R. & Kim, J. Replica molding-based nanopatterning of tribocharge on elastomer with application to electrohydrodynamic nanolithography. *Nat. Commun.* **9** (2018).