Supporting Information

Thickness Dependence on Anodization Time

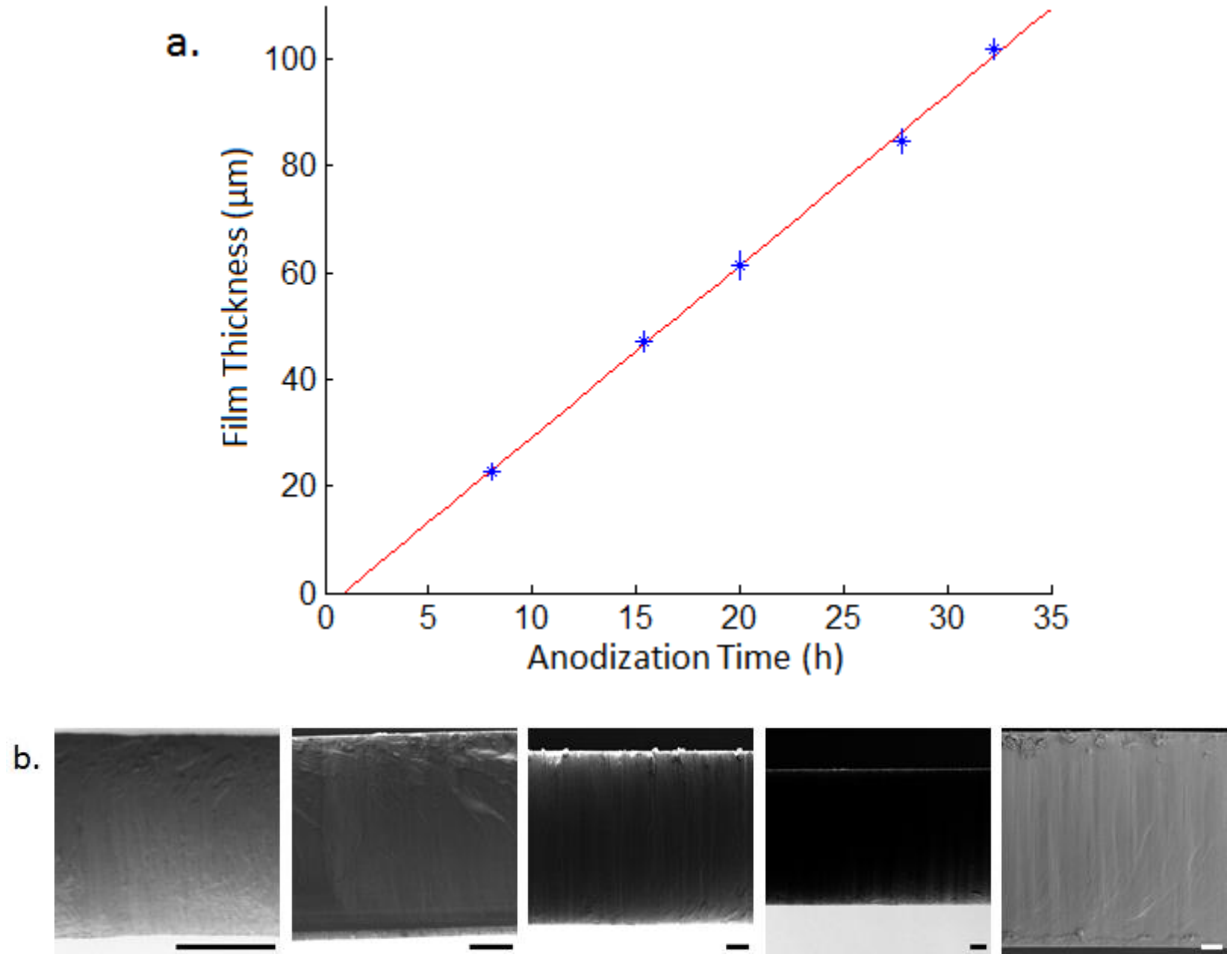


Figure S1. Plot of mean AAO film thickness versus anodization time. Uncertainties are standard deviation calculated from measurements on three samples (a). Representative cross sectional SEM images from each iteration (1st on left to 5th on right). All scale bars are 10 µm (b).

The thickness calibration plot has a slope of 3.2 µm per hour. Film thicknesses were measured by taking cross sectional SEM images (Figure S1 b).

SEM Image Analysis to Quantify Long-Range Ordering of Unit Cells

The micrographs in Figure 1 in the main text show the microscopic structure and arrangement of the unit cells over small range. In order to quantify the degree of long range ordering, several large area images of 242 μm^2 were taken of the bottom of the alumina films after each iteration. These images were acquired with just over 6 million pixels and contain around 25,000 unit cells each. A unit cell is defined by its center point and all its adjacent neighbors, which is six in most cases. However, only small regions are shown in Figures S2 to S4. As can be seen from the micrographs, there are different domains of unit cells nicely packed together with hexagonal lattice. The first step in identifying the length scale over which these domains exist is to identify and locate the center position of each unit cell. The following algorithm and the images in Figures S2 to S4 show how this was automated. The starting SEM image was smoothed with a 3 x 3 kernel to reduce the noise in the image, which helps yield a smoother binary image after thresholding according to Otsu's Method.[1] A hole-filling algorithm was applied to fill in the few small holes remaining within individual unit cells. From this binary image a new image is made with the intensity represented by the shortest distance from a unit cell pixel to a boundary pixel. This image is inverted, and all pixels with greater than three times the mean depth are treated as infinitely deep. This thresholding step prevents some unit cells with noisy outlines from being artificially broken into two separate cells, while also not artificially merging adjacent cells. Next, the watershed algorithm is applied to find the boundaries between domains and is subtracted from the previous binary image. The watershed analysis and boundary subtraction breaks apart some unit cells that were not completely isolated

from the first thresholding step going from SEM image to binary image. The highlighted areas in Figure S2, panels d and h show how these unit cells are broken apart.

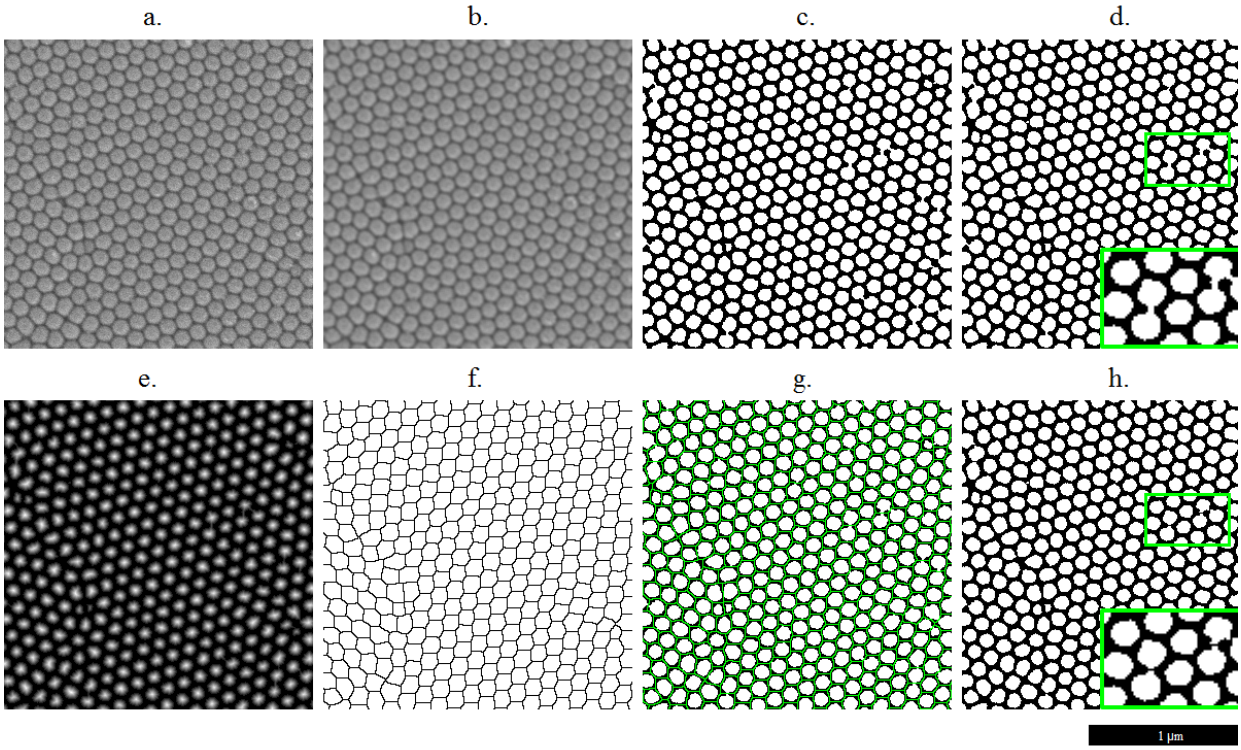


Figure S2. Images after each step of processing to go from original SEM image to finding each unit cell. Only a small subset of the image is shown due to the large size of the whole image. Original SEM image (a). Smoothed SEM image (b). Binary image after thresholding (c). Binary image after filling in any missing holes (d). Image showing distance of each pixel within a unit cell to nearest boundary (e). Binary image showing boundaries after watershed analysis (f). Image showing watershed boundaries overlaid in green on binary image from d (g). Binary image after breaking up merged unit cells by subtracting watershed boundaries from d (h). The enlarged, highlighted inlay d shows how the previously merged unit cells are broken up in h.

From the binary image shown in Figure S2, panel h, the center position and size of each unit cell is calculated. The histogram of unit cell sizes is shown in Figure S3, panel a. The histogram indicates that there is one large distribution with a tail extending all the way down to zero. The small area domains are not true unit cells and are removed by the threshold represented by the red line. This threshold was set quite low at $1/7^{th}$ of the mean unit cell size, such that the few small remaining artifacts were removed without deleting any true unit cells. This thresholding prevents these small area artifacts from inducing the analysis to label those sites as defects in future steps. Also, the selection of this threshold is not very sensitive because most of the unit cells with well below average size are part of defect clusters and will be ignored in future steps anyways.

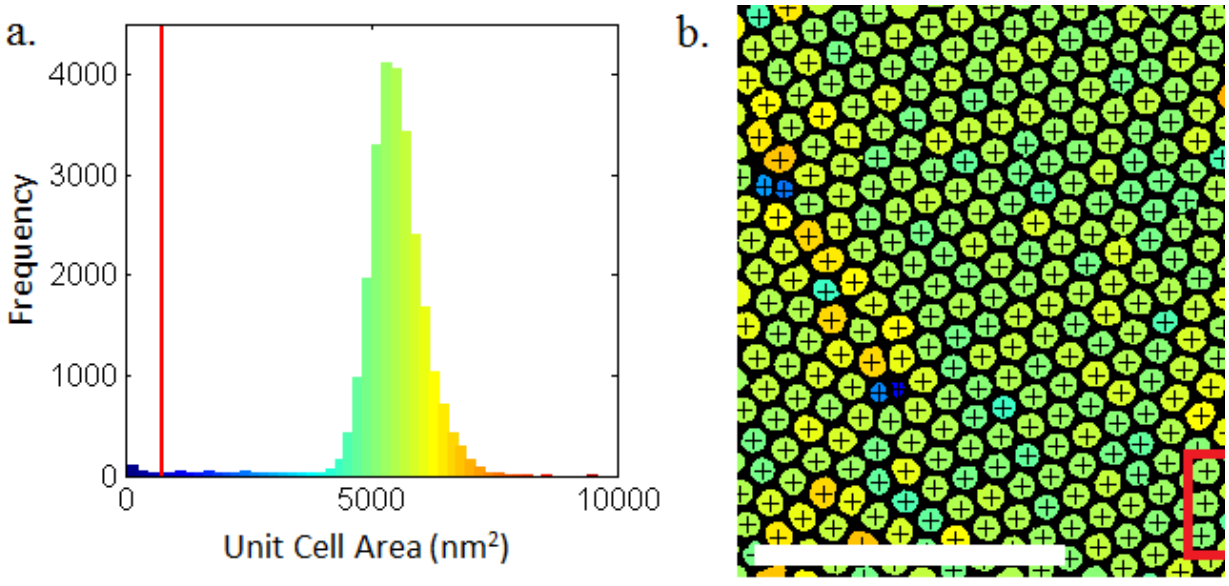


Figure S3. Histogram showing distribution of unit cell sizes (a). Image showing unit cells plotted with color corresponding to its size as shown in the histogram and with the center locations indicated by the black cross (b). The scale bar is 1 μm, and the inset to the bottom

right shows a region (red circle) that had a small, artifact unit cell that was removed with the size threshold shown on the histogram.

With all of the unit cell's center points located, Delaunay Triangulation is performed which finds the connectivity of each unit cell with its neighbors. Based on the locations and connectivity, several values can be calculated for each unit cell including: number of neighbors, closest and farthest neighbor, minimum angle and maximum angle between adjacent neighbors, standard deviation of the angles of adjacent neighbors, and mean orientation to the horizontal. The distribution of the first six of these values is shown in Figure S4, panel a. By applying a series of thresholds, many of which are redundant, the defect sites can be automatically identified. Although these thresholds are hand set, it is easy to see from the image in Figure S4, panel b, that these conditions separate the defect sites from the nicely close packed hexagonal structure elsewhere.

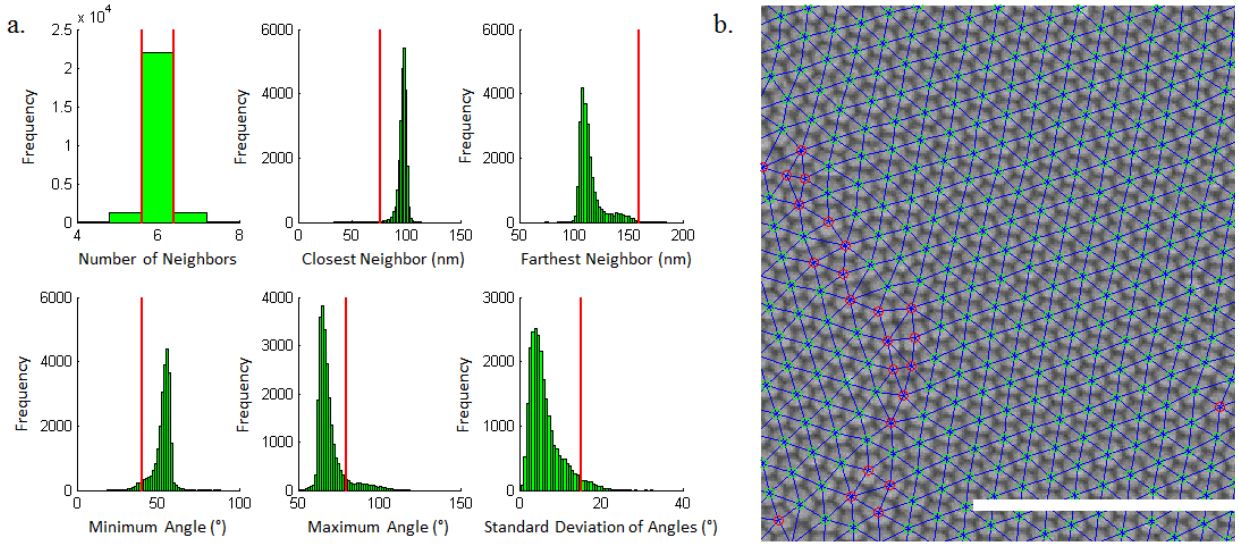
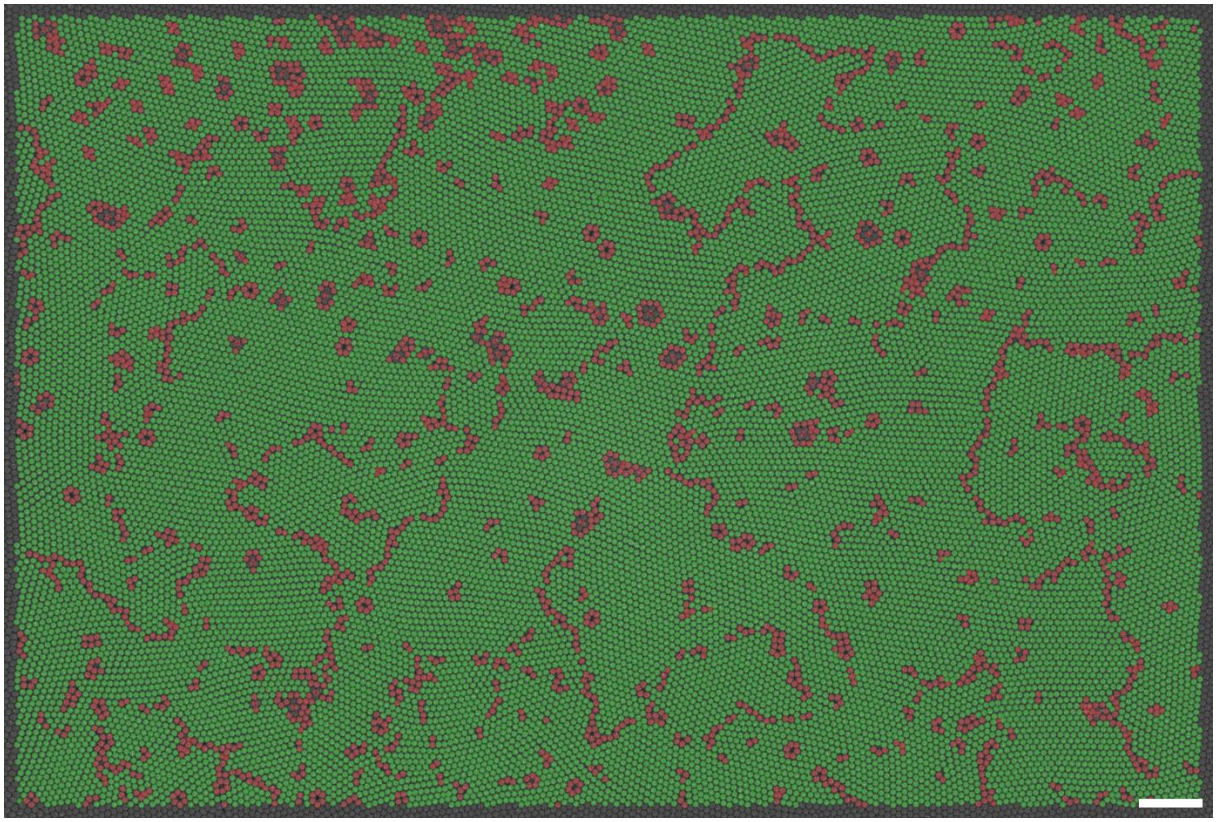


Figure S4. Histograms for the distributions of number of neighbors, closest and farthest neighbor, minimum, maximum, and standard deviation of the angles for each unit cell (a). The vertical red lines show where the threshold on each histogram is for defining a defect. The original SEM image is shown in (b) with the connectivity from the Delaunay Triangulation overlaid in blue. The cyan circles denote good, non-defect unit cells whereas the red circles mark defect unit cells from the thresholding in part a. The scale bar is 1 μm.

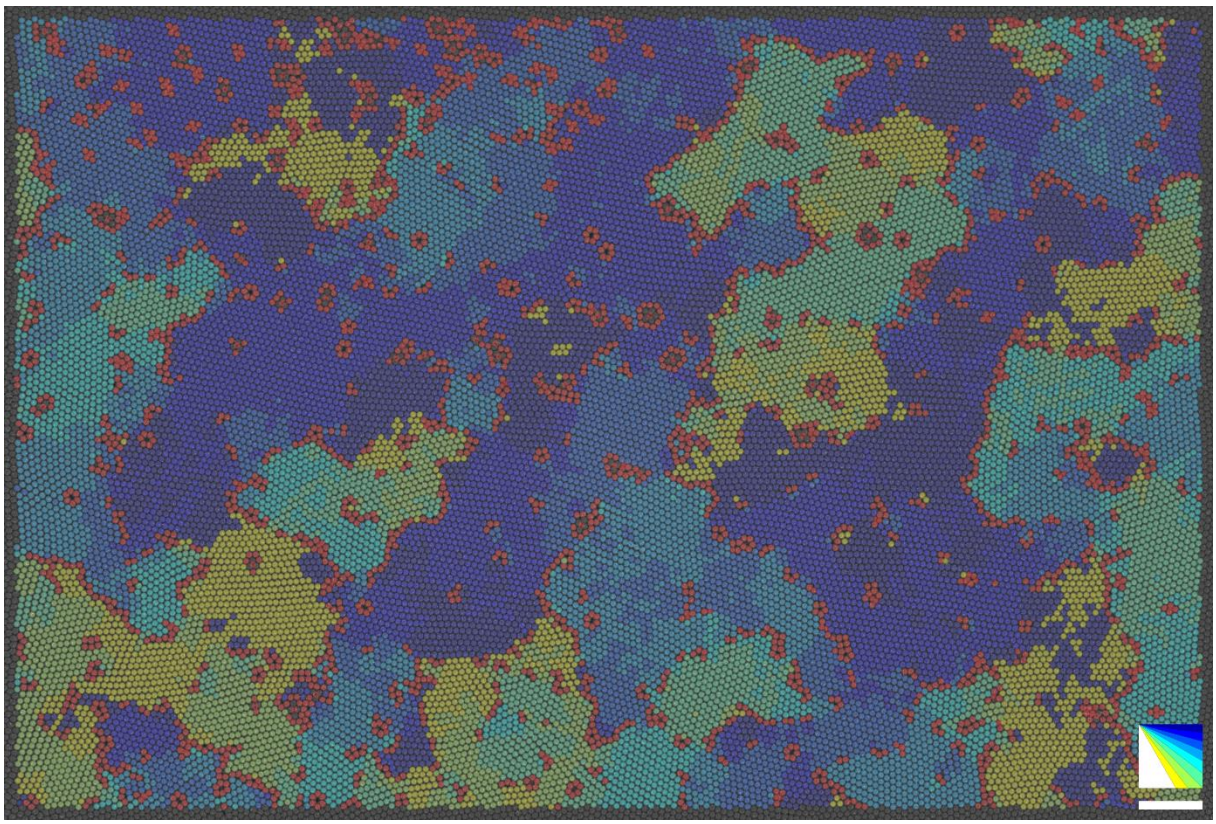
The full zoomed-out SEM image is shown in Figure S5, panel a, with the defect unit cells highlighted in red and the non-defective unit cells highlighted in green. For the next step in the analysis, the non-defective unit cells (good unit cells) are binned by their orientation with respect to the horizontal. Because the close packing is six fold symmetric, the total orientation only ranges from 0° to 60°. In Figure S5, panel b, the defect unit cells are again highlighted in red and the non-defective unit cells are highlighted according to their orientation by breaking up the 0° to 60° range into 8 bins of 7.5° each. The orientation coloring map is shown in the bottom right. The final step in the image processing analysis is to group the adjacent unit cells with the

same orientation bin into one domain, and calculate its size. This procedure is demonstrated in Figure S5, panel c, in which the defect unit cells are ignored, and the isolated domains of similar orientation are highlighted according to the orientation coloring map in the bottom right. Overlaid on this image is the size of each domain in μm^2.

a.



b.



c.

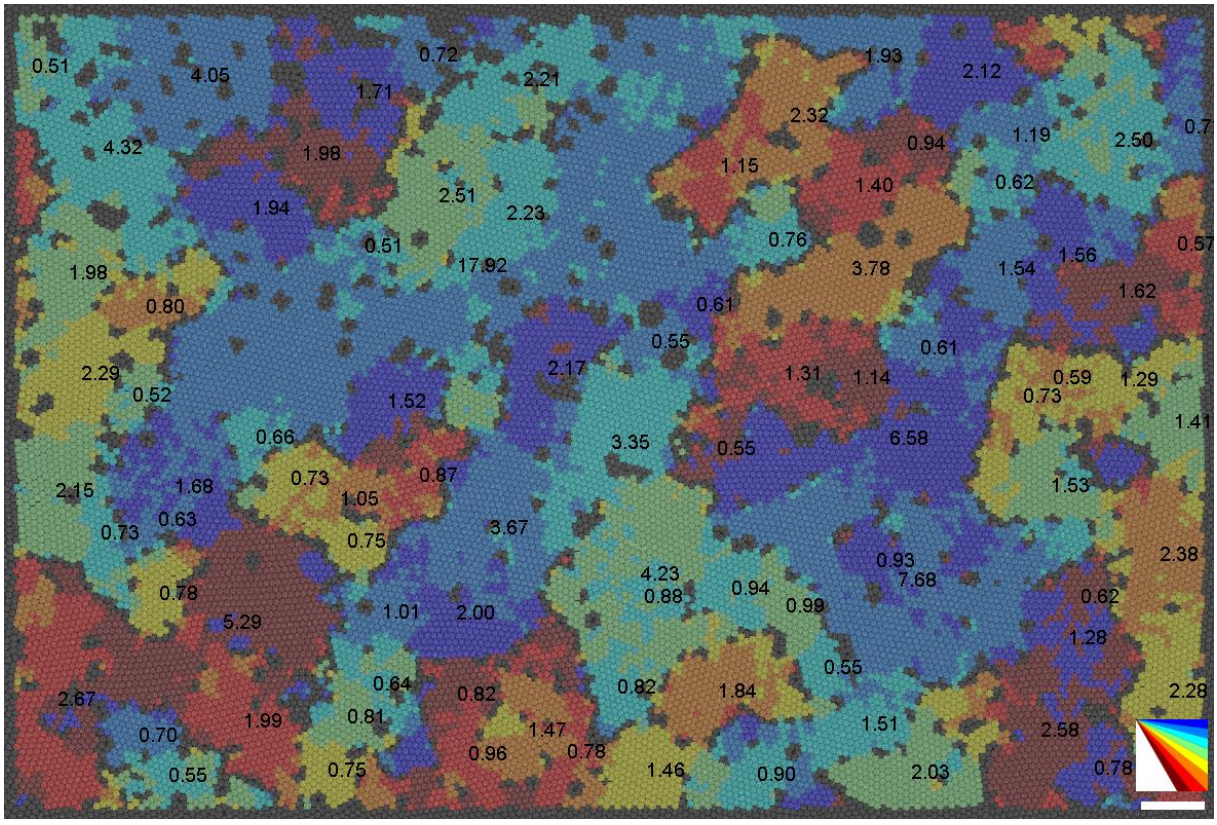
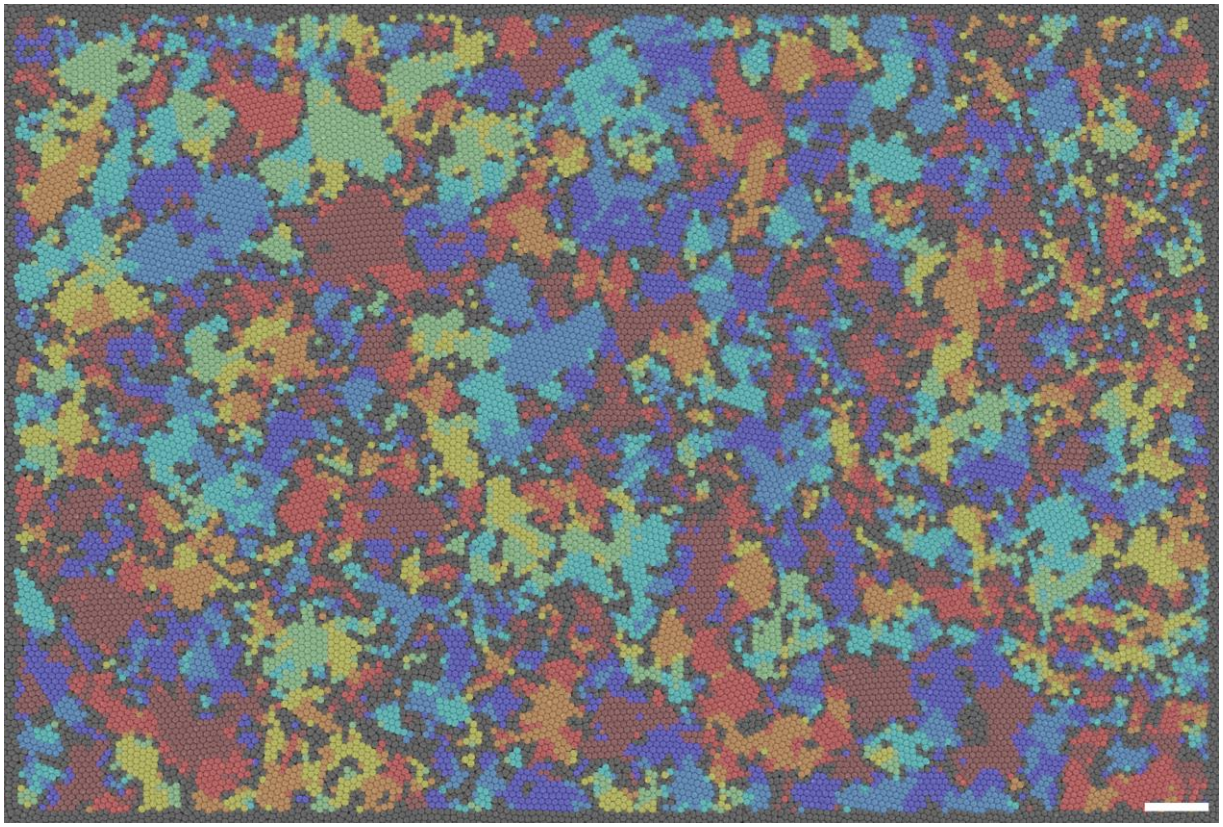


Figure S5. Full 6 megapixel SEM micrographs highlighting the selection of good, non-defect, unit cells in green and defect unit cells in red (a), highlighting the defect unit cells in red and the good unit cells according to their orientation (b), and showing the entire domain consisting of unit cells with similar orientation and the size of the grain in μm^2 (c). All scale bars are 1 μm and the binning of the 0° to 60° of possible unit cell orientations are shown above the scale bars.

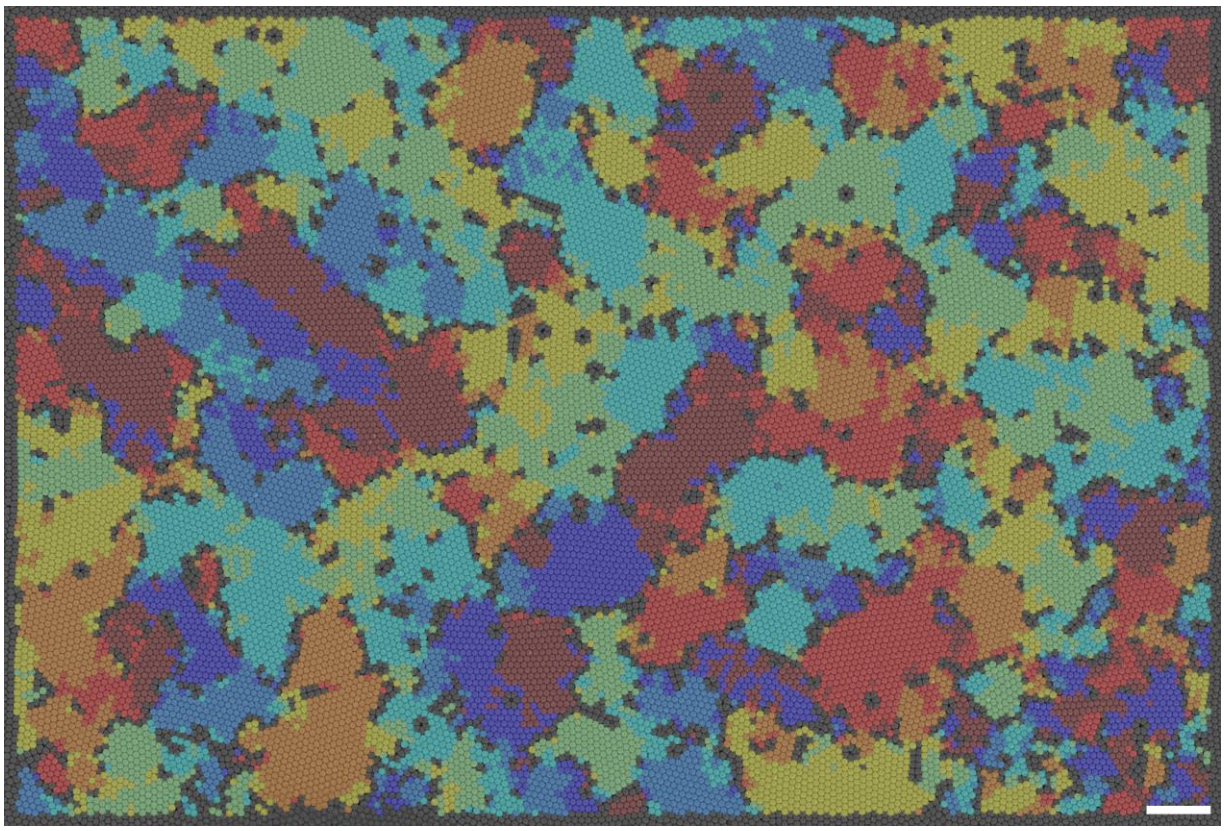
The values of the thresholds throughout the analysis changes the number of unit cells labeled as defects, which are removed from further analysis. And, the choice of angular bin size changes the size of the output domains. Neither of these adjustable parameters changes the ultimate trend

of larger domains being generated from each successive iteration. This trend is easily discernible by eye in Figure S6 which depicts an image from each iteration with the grains highlighted in color according to their orientation.

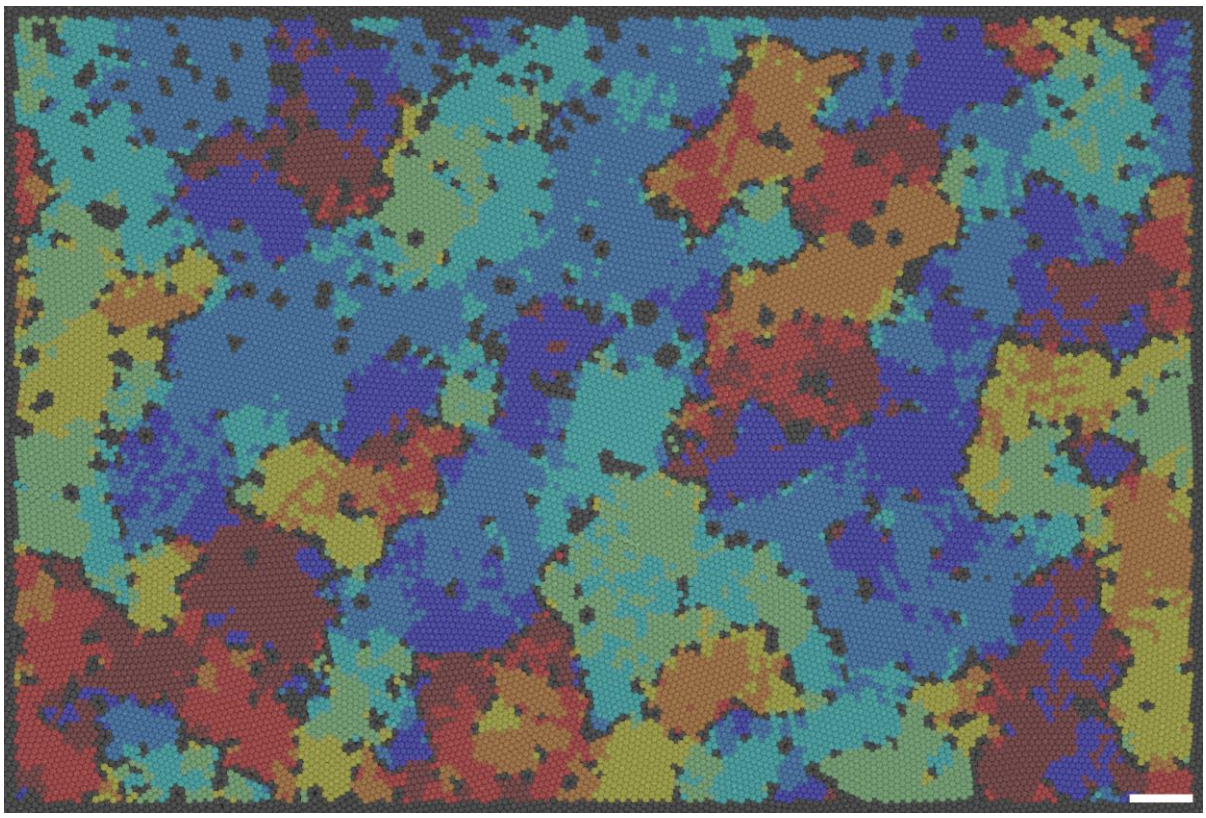
a. First Iteration



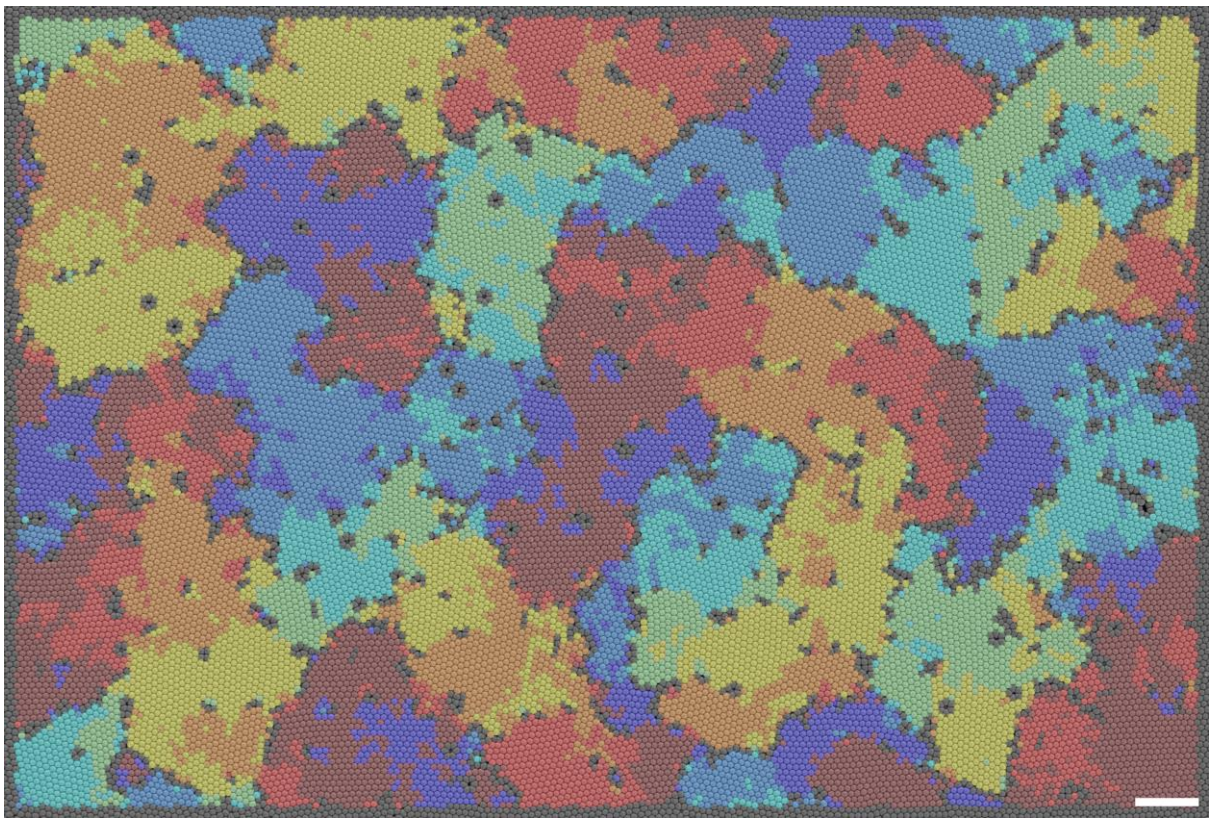
b. Second Iteration



c. Third Iteration



d. Fourth Iteration



e. Fifth Iteration

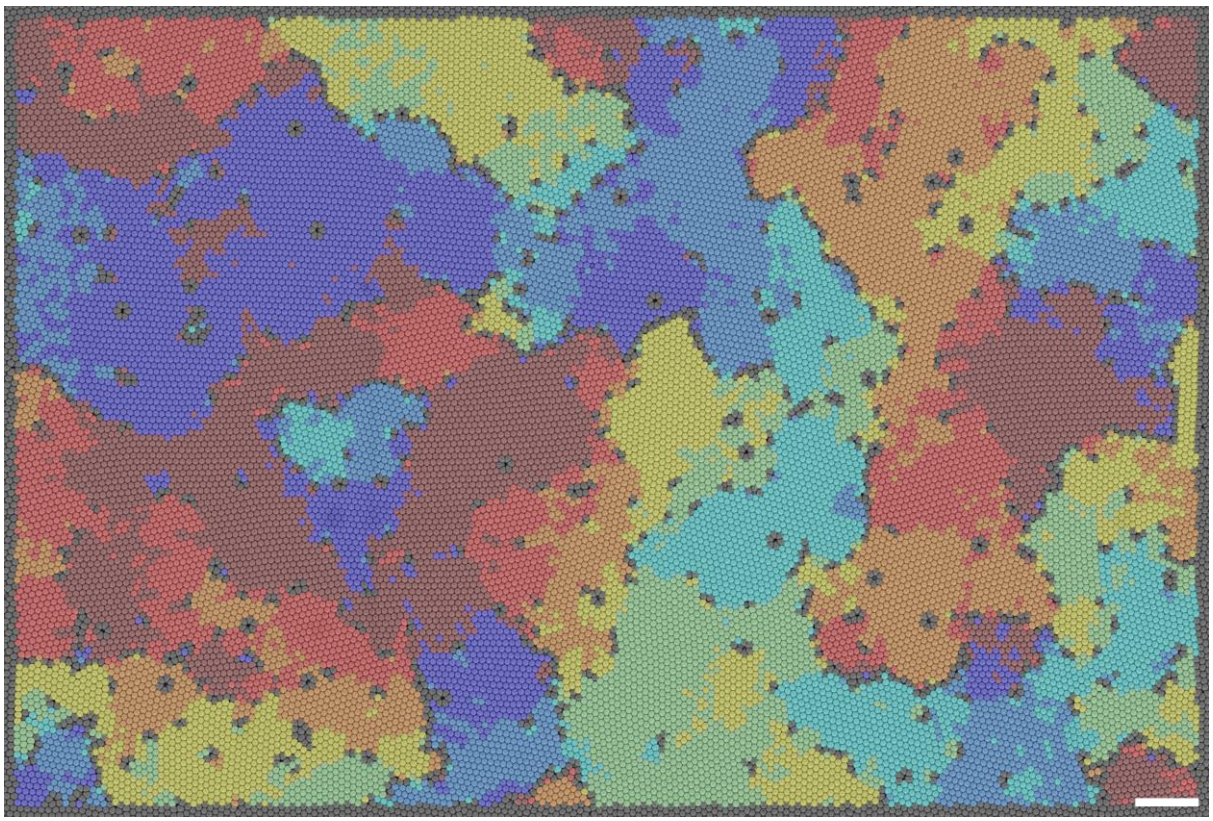


Figure S6. SEM micrographs overlaid with the grains of different orientation highlighted according to orientation after the 1st through 5th iterations in order (a-e). All scale bars are 1 μm.

Several large images were analyzed using this analysis for each iteration to build sufficient statistics to define the distribution of domain sizes. The distribution of domain sizes after each iteration is plotted as a histogram. Half square micrometer bins were used, but the first bin was always discarded because its value was easily skewed to higher values by isolated unit cells surrounded by other cells in an adjacent orientation bin. The arbitrary choice of orientations to mark the bin boundaries in the domain size analysis leads to areas of slowing varying orientation being broken up into several small domains that only consist of a few unit cells. Because of this

feature of the analysis, the first bin was discarded, which leads to a smooth distribution for the remainder of the size histogram. The histogram follows an exponential distribution which was fit to the equation $y = Ae^{-x/k}$ in order to find the characteristic domain size, k. The histogram of domain sizes after each iteration and their corresponding exponential fit is shown in Figure S7 and the plot of characteristic domain size after each iteration is shown in the main text in Figure 2.

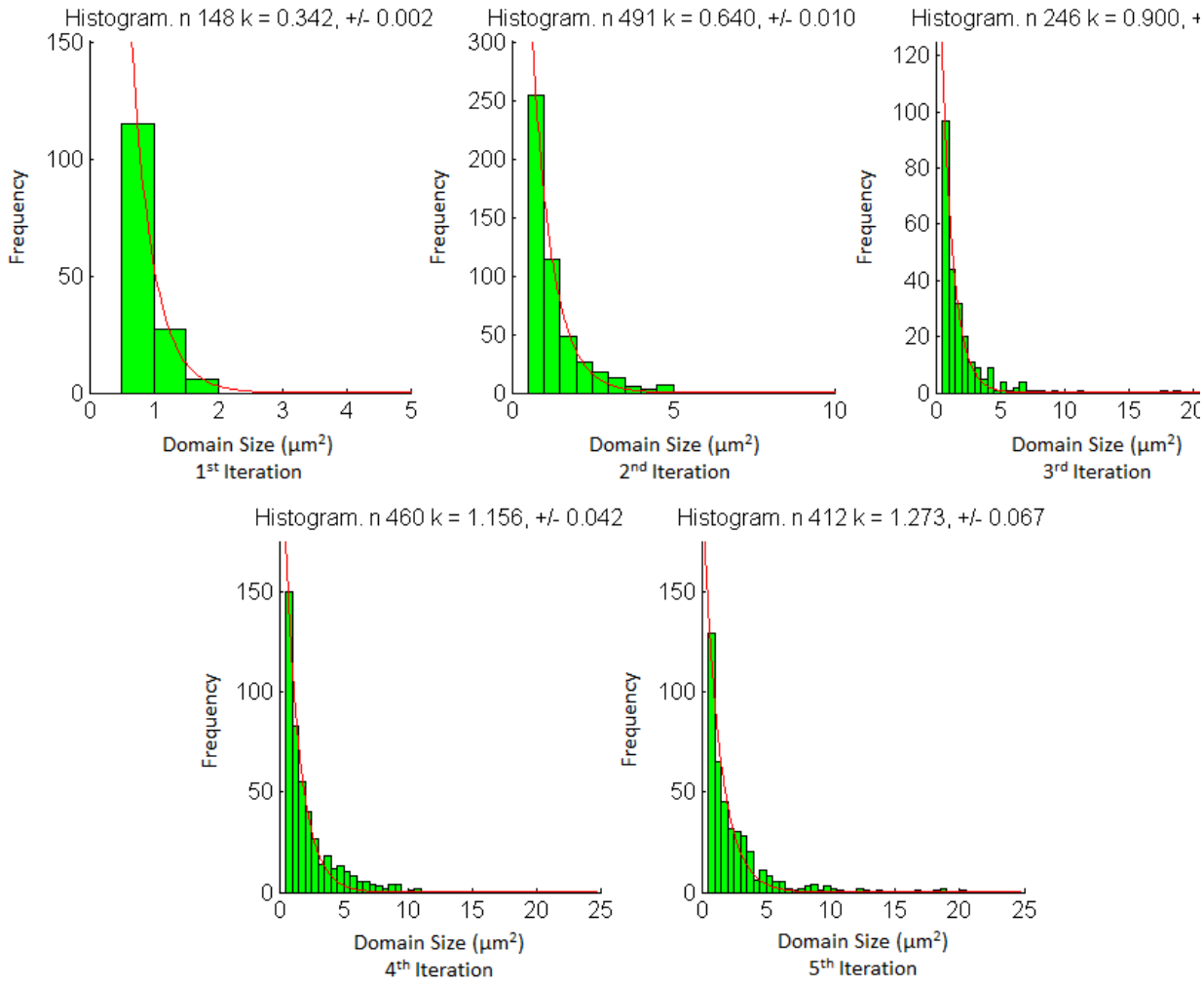


Figure S7. Histograms showing the distribution of the domain sizes after each iteration with the exponential fit shown in red.

Pore Size Analysis

A similar method that is used to find the central locations of each unit cell from the SEM images in the last section, shown in Figure S2, is used to find the distribution of pore sizes from the top side SEM images. One difference here is that higher magnification, smaller area images covering 3.5 μm^2 are used because the higher magnification allows for better definition of the pores while still being able to analyze several hundred pores. Another difference in the pore size analysis is that the output is a distribution of pore diameters calculated by assuming the area of each pore comes from a circle. The main distribution of this histogram is fit to a Gaussian to yield the mean pore size and standard deviation. The top down SEM image, the resulting binary image, and the distribution of pore sizes is shown in Figure S8. This analysis produces a larger pore diameter than taking the FWHM of the line profile across the pores but is used because it can be fully automated and uses more of the information available in the image. The resulting pore sizes after each iteration are shown in Figure S8, panel d. The first iteration has a broad distribution of small, irregular pores, but for each iteration after that, the pores are well defined and have a constant diameter, within the uncertainty.

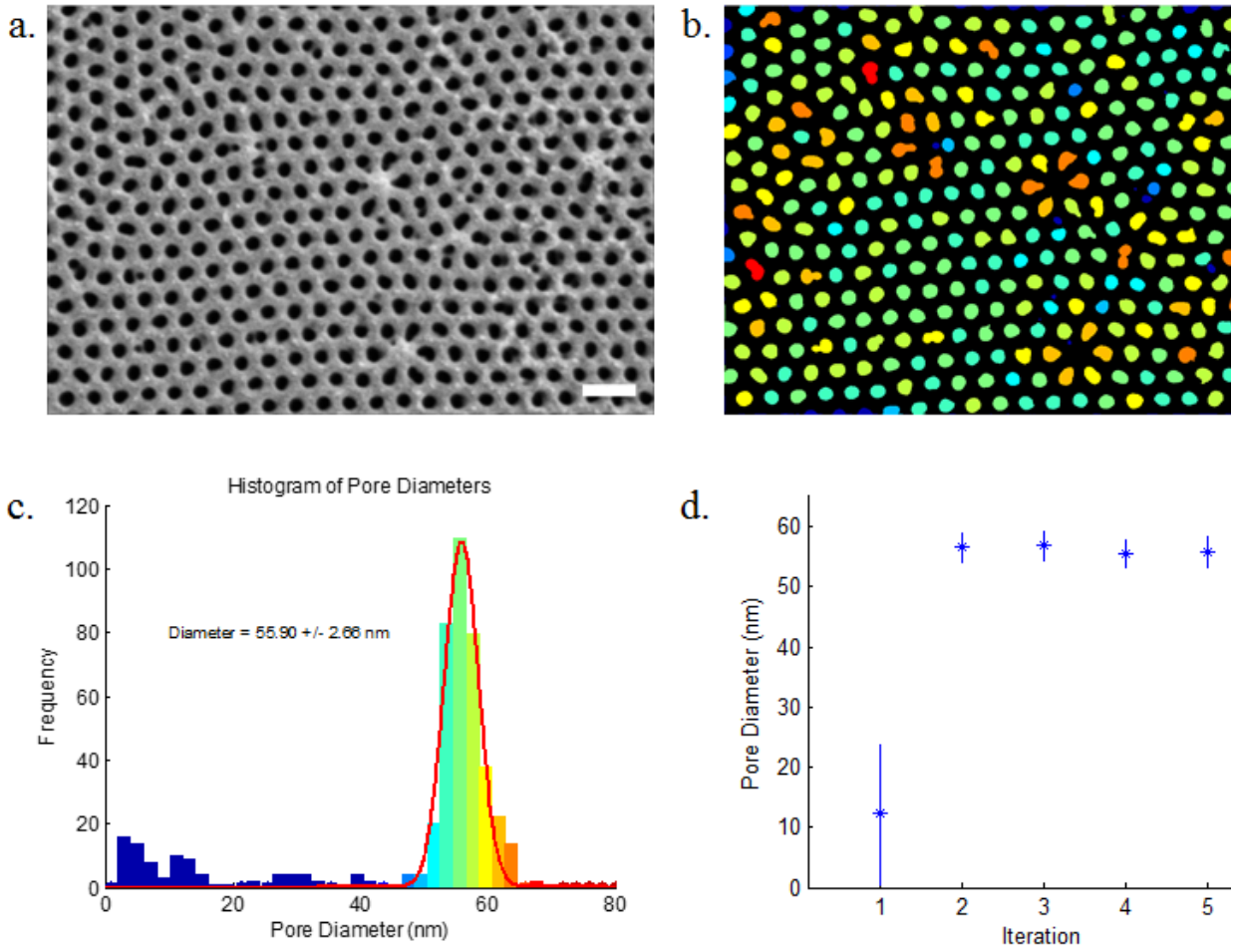


Figure S8. Top down SEM image showing the open pores of the alumina film from the 3rd iteration (a). Binary image with pores color coded according to their size (b) as shown in the histogram in (c). Plot of pore diameters after each iteration. Uncertainties are standard deviation. Scale bars are 200 nm.

References:

1. Otsu, N., A Threshold Selection Method from Gray-Level Histograms. *Systems, Man and Cybernetics, IEEE Transactions on* **1979,** *9* (1), 62-66.