The presence and configurations of defects are primary components determining materials functionality. Their population and distribution are often nonergodic and dependent on synthesis history, and therefore rarely amenable to direct theoretical prediction. Here, dynamic electron beam–induced transformations in Si deposited on a graphene monolayer are used to create libraries of possible Si and carbon vacancy defects. Deep learning networks are developed for automated image analysis and recognition of the defects, creating a library of (meta) stable defect configurations. Density functional theory is used to estimate atomically resolved scanning tunneling microscopy (STM) signatures of the classified defects from the created library, allowing identification of several defect types across imaging platforms. This approach allows automatic creation of defect libraries in solids, exploring the metastable configurations always present in real materials, and correlative studies with other atomically resolved techniques, providing comprehensive insight into defect functionalities.

INTRODUCTION

Materials form the basis of modern civilization, ranging from structural materials in buildings and machines to semiconductors underpinning information technologies to functional materials in batteries, fuel cells, sensors, and other applications. For many materials classes, functionality is ultimately determined by the presence and configurations of defects, ranging from simple substitutional defects in solid solutions to complex extended defects involved in crystallographic shear phases, dislocations, etc.

In the case of two-dimensional (2D) materials, a growing number of examples illustrate the possibility of electron beam–induced introduction of defects (1–6), some of which offer the possibility of atomically precise positioning within the lattice (7–9). These capabilities may open pathways for the atomically precise fabrication of physically interesting and technologically relevant structures such as molecular magnets (10), quantum dots (11–14), spin diodes (15, 16), and ballistic transport field-effect transistors (17–19). To achieve the level of control necessary to fabricate these structures, a deeper understanding of beam-induced modifications and a broad grasp of what defect configurations are experimentally possible must be prioritized.

Clearly, the use of high-throughput computations to create libraries of materials is not new, as exemplified by Materials Genome Initiative (20). In this, large databases of first-principles calculations were created and subsequently mined for materials with interesting properties or for establishing structure-property relationships. However, while powerful for simple materials, the number of possible structures for extended defects makes direct computational search infeasible. An alternative approach for materials design is based on artificial intelligence–assisted experimental workflows, where thermodynamic databases and density functional theory (DFT) calculations are integrated to achieve property predictions and establish synthesis-property relationships. The corresponding databases include Materials Data Facility (21), Citrination (22), Dark Reactions (23), and Materials Innovation Network (24). However, to date, these efforts use macroscopic experimental descriptors and mesoscopic structural descriptors only. Very recently, Rajan et al. (25) created a theoretical library of nanosized pore defects in 2D materials and explored their energetics.

Given advances in imaging, it is now possible to image many systems with atomic resolution via electron or scanning probe microscopy (26–30). The key to understanding and using these data is to extract the different configurations of the atomic species, as their distributions (fluctuations) encode important aspects that govern the system’s response to thermodynamic perturbations. However, this requires locating and classifying the defects themselves in an automated manner because of the numbers required for proper statistics. The task is even more important given the need for complementary data on the same defect configurations. For instance, imaging of atomic structure of defects is relatively straightforward in scanning transmission electron microscopy (STEM). However, the finer details of electronic structure at or around the defect necessitate scanning tunneling microscopy (STM) measurements. Because of the differences in imaging mechanisms and limitations of the experimental platforms, correlative atomically resolved STEM-STM studies are extremely complex. Furthermore, the contrast in STM and STEM data is completely different, so the complementary nature of both techniques cannot always be readily used. STEM measures lattice structure (i.e., the ideal image is the atomic nuclei, and the intensity in the image is roughly proportional to the square of the atomic number), while STM measures electronic properties [i.e., the ideal image is the (local) integrated density of states, and the image intensity reflects spatial distribution of electronic states at the Fermi level]. What is needed, is a combined experimental-theoretical approach to extract defects, classify them, and then perform first-principles calculations to explore their electronic structure. The latter can immediately be used to simulate the expected STM images and provides direct insight into the effects of the defect on the material’s properties.

Here, we demonstrate an experimental approach for creating defect libraries from STEM data and determine defect statistics in the Si-graphene system via advanced image recognition tools. We further use these as an input for the DFT calculations to establish
RESULTS AND DISCUSSION
To create libraries of defects in graphene, we used active manipulation via a 100-kV electron beam to disperse silicon (Si) impurities over the graphene lattice (7). This active sputtering approach allowed us to create multiple metastable configurations of Si-C complexes in graphene. We then acquired images using a 60-kV electron beam, which minimizes damage to the graphene lattice and allows a closer inspection of the formed defect structures. Figure 1A shows an example of high-angle annular dark field (HAADF) STEM image of a graphene sample with amorphous Si-C regions, which serve as the source of Si dopant atoms. Individual Si impurities embedded in the graphene lattice are seen.

To analyze the created defect structures and to generate a library of defects, we have developed a deep learning–based analytical workflow that searches for and identifies various defect configurations in the available “raw” experimental data and categorizes them on the basis of the number of Si atoms in each configuration. As a first step, we used two deep fully convolutional neural networks (see Materials and Methods for details). The goal of the first neural network, a “defect sniffer,” is to identify anomalies and irregularities (that is, defects) in the periodic structure of a crystalline material (in this case, graphene). The second neural network, an “atom finder,” identifies the elemental species and positions of atoms in a localized region (typically 2 nm by 2 nm) around the detected defect. This two-stage deep learning approach resembles the logic of a human operator who first identifies points of interest (irregularities and “anomalies”) in a larger-scale experimental image and subsequently examines specific features (e.g., details of atomic structure surrounding the detected irregularity) in more detail. Thus, it can also potentially be applied in the future for a realization of an intelligent (“self-driving”) microscope. Notably, while this two-stage approach can be, in principle, replaced by just a single deep learning model, the classical generalization versus accuracy trade-off will likely increase the error of finding positions of atoms and atomic defects. We therefore believe that the current approach (with two specialized models rather than one very general model) is an optimal solution for this type of problem.

The defect sniffer neural network was trained using several weakly labeled experimental images [that is, a structure is either a defect (vacancy or impurity) or not, without classifying defects into specific subcategories at this stage], using the fact that each defect is associated with violation of ideal periodicity of the lattice. The details of the procedure for constructing a training set for this type of problem can be found elsewhere (31). It was trained to find point defects only (vacancies and Si dopants), while avoiding areas with extended defects (such as larger holes and amorphous Si-C regions). Density-based spatial clustering analysis (32) was applied to the output of a neural network to group multiple detected point defects located close to each other (maximum radius \( \varepsilon \) set to 0.5 nm). The regions of interest (in terms of the presence of Si point defects) identified by this approach for data from Fig. 1A are shown in Fig. 1B.

The atom finder neural network was trained using simulated STEM images. Recently, several studies (33–35) have demonstrated that deep neural networks trained using simulated data can successfully generalize to real experimental atom-resolved and molecule-resolved STM and STEM data. The theoretical data were produced by either using a MultiSlice algorithm (36) for STEM image simulations or representing atoms as Gaussian blobs. Both approaches for model training gave similar results in terms of defect finding and identification for the current 2D system. Each theoretical image was augmented to account for model uncertainties such as local distortions (±15% displacements in atomic positions) and instrumental factors such as global image or scan distortion and different levels of noise.

The output of the second neural network was mapped onto the graph structure representing the classical chemical bonding picture (35). Specifically, we first identified the centers of the mass for each “blob” (atom) in the pixel-wise maps from the output(s) of a neural network and then used the information about the type and position of each atom to construct the graph nodes. The nodes were then connected by the graph edges that represent bonds. To construct the bonds, we used chemistry-based hard constraints on the maximum possible coordination number of each type of atom, as well as maximum or minimum allowed length of atomic bond between corresponding pairs of atoms. We note that this analysis assumes that a sample has a quasi-2D configuration. If there is no information (e.g., from image “tags”) available on the scan size, then it is possible switch to a calibration based on the peak in the distribution of nearest-neighbor distances, which are predominantly C—C bonds.

Fig. 1. STEM data: Original and processed by neural networks. (A) STEM overview image. Brighter areas typically correspond to amorphous Si-C regions, from which Si atoms can be dispersed into clean graphene lattice, leading to various Si defect structures (in terms of number of Si atoms and their bonding to each other and to nearby lattice atoms). (B) Results of applying a defect sniffer (deep neural network followed by density-based spatial clustering) to data in (A). White markers show regions of interest as identified by a defect sniffer. Notice that it ignores amorphous Si-C regions and returns information on the location of point Si defects, only (C) 2 nm by 2 nm area cropped around one of the identified defects (white box in (B)). This cropped image serves as an input to atom finder model. (D) The output of an atom finder network for the image in (C), where red blobs correspond to C lattice atoms and a green blob is associated with Si impurity.
Here, it was crucial to identify accurately the environment of each Si impurity, which included the information on the lengths and angles of bonds with the neighboring carbon atoms in the host graphene lattice. To the best of our knowledge, the current methods for atom finding based on 2D Gaussian fitting in atomically resolved data do not allow a reliable extraction of atomic positions for images where atoms do not appear as local maxima (such as in Fig. 1C). As a result, there is no ground truth for the experimental data against which our deep learning–based atom-finding approach can be tested. Hence, we used images of graphene lattice simulated with the MultiSlice algorithm that were not part of the model training or testing and therefore can serve as a validation dataset. The number of atoms in the validation images was comparable to that in the experimental images used in the current study (in the second, “atom finding,” stage). Each image was corrupted by unique combination of blurring and noise (no two images were the same). The average (per each image) deviation of the predicted positions from the “ground truth” positions of atoms for different combinations of noise and blurring is shown in Fig. 2A. We generally found that the trained model can determine robustly atomic positions until the level of noise becomes so high that the lattice contours become barely distinguishable to the human eye (this corresponds to regions with $0 > 130$ and $\lambda' < 110$ in Fig. 2A). In Fig. 2B to D), we show an example of the model output for data where the image resolution, the number of atoms in an image, and the level of noise are comparable to those in current experiments. The deviation of atomic center positions from “true” positions for these data is mostly below 10 pm, which is within the instrumental error for detection of atomic position in single-image STEM data with low signal-to-noise ratio (37, 38). One of the reasons why a neural network shows such a good performance when tasked with finding atoms is because it is “aware” of the system (lattice) under study (i.e., trained specifically to look for atoms in this type of lattice, ideal and distorted). We believe that because getting the atomic positions accurately is crucial for this particular type of pixel-wise classification problem, the approach for testing a network’s accuracy described above is preferable to the existing ones from the domain of classical machine or deep learning.

The total number of analyzed images was approximately 500, with each image containing either single or multiple Si impurities. The image size varied between 2 and 16 nm, and the image resolution was in the range between 256 pixels by 256 pixels and 2048 pixels by 2048 pixels. Notably, many of the same types of Si defects can exist without 100-kV e-beam “sputtering” but with a much smaller density, which is not sufficient for creating a library of defects. The representative examples for each category were then selected with the assistance of a human expert. Structures that could not be confirmed independently (even on a simple qualitative level) by a human expert eye, either because of very high levels of noise in the image or because of ambiguity associated with

![Fig. 2. Evaluation of accuracy in finding the atomic positions via a deep learning model for different levels of noise.](image)

(A) The deviation of the predicted positions (average value for all atoms in the image) from the ground truth positions of atoms for different combinations of noise ($\lambda'$) and blurring ($\sigma'$). The $\lambda'$ and $\sigma'$ are the scaled parameters of Poisson and Gaussian distributions, respectively. See the Supplementary Materials for the details and code of how the noisy data were created and analyzed. (B) Simulated image of graphene lattice associated with the red cross in (A). (C) Same image overlaid with atomic coordinates. Different colors of circular markers show different degrees of atomic positions displacement from ground truth coordinates. (D) The histogram showing atomic displacements for every atomic position displayed in (C). Most of the deviations are below 10 pm, which is within instrumental uncertainty for atomic position extraction.
the absence of information in the $z$ direction for certain structures, were not considered or counted. Last, the input (structural) data for the first principles–based studies of the detected defect were obtained by cropping a square “cell” with ~30 atoms around the defect of interest (for the chosen categories) and embedding them into a larger standard graphene supercell.

Figures 3 and 4 illustrate a constructed library of Si-containing defects in graphene. We define a Si defect as a Si atom and its first coordination “sphere” or Si atoms connected either directly or through a single C atom to each other and their first coordination sphere. Using this definition for a defect, we began categorizing Si defects on the basis of the number of Si atoms that each defect contained. The defects containing a single Si are shown in Fig. 3, and the defects containing two and more Si impurities are shown in Fig. 4. Our approach also allowed a further split of these categories into subcategories based on coordination number, variation in bond lengths in a single defect, proximity to other defects [e.g., multi-atom lattice vacancies located within a specific radius (one lattice constant, two lattice constants, etc.) from the defect center of mass]. Notably, because our analysis was performed after the STEM experiments were completed, it does

**Fig. 3. Constructing libraries of defects from STEM data on graphene with Si impurities via deep learning–based analysis of raw experimental data.** (A and B) Defects containing a single three-fold coordinated Si atom (count, 465) (A) and a single four-fold coordinated Si atom (count, 16) (B). (C and D) Histograms of Si—C bond lengths (C) and Si—C bond angles (D) for three-fold Si defect. (E and F) Examples of distorted threefold Si defect located next to a multivacancy (E) and at topological defect (F). The number of distorted (SD $s = ( \Sigma dx^2/n)^{1/2}$ in C—Si—C bond angles above 15) and undistorted three-fold Si defects was 231 and 234, respectively.
not account for the possibility that certain defects appear more frequently than others simply because of human operator (confirmation) bias and specific interest in certain structures. Hence, the frequency of different defect occurrences in the analyzed images may not necessarily reflect the relative frequency of their occurrence in the physical system. However, automatization of experimental acquisition will preclude these problems in the future.

The first category is a defect that has only one Si atom (Fig. 3). It contains two subcategories, based on number of bond connections to nearby lattice atoms, corresponding to classical three and four-coordinated Si atoms (Fig. 3, A and B, respectively) (39). Our model was able to identify 465 three-fold coordinated Si atoms and 16 four-fold coordinated Si atoms. This approach also enables an automated calculation of the bond lengths and bond angles in each detected defect. The histograms of bond lengths and angles for all detected three-fold Si defects are shown in Fig. 3 (C and D). We next defined a distorted three-fold defect as a defect where an SD $s = (\sum dx^2/n)^{1/2}$ of three C—Si—C bond angles is larger than 15. This allowed the identification of additional subcategories of three-fold defect corresponding to Si impurities located adjacent to hole or multivacancies (Fig. 3E) or to a topological defect (Fig. 3F), which are characterized by a substantially stronger distortion of C—Si—C bond angles (note that both missing atoms in honeycomb lattice and topological reconstructions with an enlarged size of carbon rings are seen as vacancies in our analysis).

We proceed to the discussion of defects containing more than a single Si atom. Figure 4 shows the identified structures of defect with two and more Si atoms. This includes multiple different dimer structures, in which each Si atom in the dimer has three-fold or four-fold coordination (Fig. 4, A to D and F) and a two-Si defect where Si atoms are connected to each other via a C atom (Fig. 4E). In the case with a single atom, the two-Si defects located next to the hole or multivacancy are characterized by a stronger distortion in C—Si—C bonds compared to when they are observed on a clean lattice (see Fig. 4, A and D). A third category includes Si defects with three or more Si atoms (Fig. 4, G to I). This category contains trimer (Fig. 4G) and tetramer (Fig. 4H) arrangements of Si impurities and more complex Si-C clusters containing five and more Si that are usually observed at the edges of graphene vacancies and holes.

![Fig. 4. Atomic defects containing two and more Si atoms.](image-url)
(A to F) Defects with two Si atoms. Si dimer structures with each Si having two C in its first coordination sphere (count, 6) (A and D); one Si having three C atoms and another one having two C atoms (count, 4) (B); each Si having two C atoms plus one "shared" C atom (count, 3) (C); each Si connected to three C atoms (count, 5) (F); nondimer structure, where two Si atoms are connected via C atom (count, 3) (E). (G to I) Defects with three and more Si atoms: three Si (count, 50) (G), four Si (count, 6) (H), and five or more Si (count, 56) (I).
This comprises the library of observed Si atomic defects in monolayer graphene. We made the images labeled according to the classification scheme described above, available through the Citrination platform (40). One immediate application of this library for material exploration is to compare the experimentally observed structures to structures calculated with DFT at different levels of strain using the averaged experimental coordinates as an initial input for geometric relaxation. By calculating and plotting the strain components, together with translation and rotation transformations, needed for transforming the experimental structures into an ideal theoretical structure (see fig. S5) or vice versa, one can estimate a “window of stability” for a particular defect type.

To demonstrate a further applicability of this library and additional insight into the electronic structure of defects in graphene with Si, we perform DFT-based calculations of electronic structure and experimental atomic-scale STM measurements on the same sample. The STM images obtained at low-bias voltages reflect spatial distribution of electronic states around the Fermi level and can be used to probe local perturbation in a material’s electronic structure caused by defects. An example of an experimental STM image obtained over a relatively large field of view on the sample from STEM measurements is shown in Fig. 5. The STM data typically showed the presence of atomically resolved regions with a lateral size of about 3 to 6 nm separated by regions for which no atomic resolution could be achieved. We assigned the latter to amorphous C-Si regions, which were commonly seen in the STEM experiment on this and other samples (7). An example of a point defect causing perturbation in the electronic structure of a nearby graphene region is shown in the magnified image in Fig. 5B.

Notably, two important limitations are expected in the interpretation of STM images from a STEM sample. First, because the sample was transferred through ambient atmosphere, contamination of the sample surface with oxygen functional groups is possible. In addition, in the STM setup used for this study, it was not possible explore the exact same region that was examined in the STEM experiment.

Despite these limitations, we were able to capture a number of atomic-scale STM images with characteristic point defects and compare them to the structures derived from the analysis of the STEM data. Note that a proper interpretation of STM data requires comparison with theoretical calculations of local density of states at the Fermi level for the structures of interest. For this purpose, we performed DFT calculations for the first two categories of defects. We limited ourselves only to single and dimer Si defects occurring on a clean graphene surface. Specifically, we first performed full 3D relaxation of the 2D projection of atomic coordinates found in the analysis of the STEM data. Then, the relaxed coordinates were used to calculate the electronic structure for each type of defect. The STM data for point defects were then compared with STM images calculated from DFT data within Tersoff-Hamman approach (41, 42).

We compared the experimental STM data with DFT calculations of integrated local density of states associated with two bands above and below the Fermi level (if these bands existed within an energy range of −0.5 to +0.5 eV). We note that of the structures analyzed, only the three-fold Si impurity showed an out-of-the-plane displacement in the final relaxed geometry, whereas the rest of the structures remained flat. Comparison of point-like defects obtained in the STM at low-bias voltage (0.1 to 0.3 V) to simulations of three- and fourfold Si defects (Fig. 6, A and B) suggests that we were able to capture the former in the STEM experiment. We next compared the dimer-like structures observed in the STM experiment (Fig. 6, C and D) to the DFT simulations of dimer structures observed in Fig. 4. We found that the closest match is the Si2C6 (Fig. 4F) structure that is characterized by two well-defined bright spots associated with two Si impurities in the relevant energy range. Either the rest of the tested dimer structures did not produce a well-defined structure with two bright spots or the distance between the spots was smaller than that in the case of STM experimental data (see the Supplementary Materials). Last, in fig. S6, we show two additional structures that are possible candidate matches with those displayed in Fig. 4 (G to I), containing three and more Si atoms. The interpretation of the exact nature of these complex defects will be the subject of future work.

CONCLUSIONS AND OUTLOOK
To summarize, we combined active electron beam manipulation of impurities in graphene and deep machine learning to construct libraries of Si defect structures. By performing STM experiments on the same sample, supplemented by first-principles calculations, we were able to link together experimental results that characterize a material’s structure (STEM) and functionalities (STM). Our work shows a pathway toward the creation of comprehensive libraries of atomic defects and their functionalities based on experimental observations from multiple atomically resolved probes generating synergistic information and performing correlative analysis of structure-property relationships on the level of single atomic defects.

We propose that an important part of both theoretically and experimentally driven materials prediction will be libraries of atomic defects. For the theoretical effort, these libraries can significantly confine the region of the chemical space to explore by focusing the effort on the experimentally observed defect classes, as opposed to all those theoretically possible. For experiment, the defect populations and correlations can be used as direct input to machine learning schemes. Last, this information can shed light on defect equilibria, solid-state reaction pathways, and other fundamental parameters of materials.

MATERIALS AND METHODS
Sample preparation
Graphene was grown via atmospheric pressure chemical vapor deposition on a Cu foil. (43) The Cu foil was then spin-coated with
poly(methyl-methacrylate) (PMMA) as a mechanical stabilizer. The Cu foil was dissolved in a bath of ammonium persulfate–
deionized (DI) water (0.05 g/ml). The graphene/PMMA film
was transferred to hydrogen chloride diluted in DI water bath
to remove the ammonium persulfate, followed by a DI water
rinse. The graphene/PMMA film was placed on a transmission electron
microscopy grid and annealed on a hot plate at 150°C for ~20 min
before the grid to the graphene. The PMMA was subsequently
dissolved away in an acetone bath, followed by an isopropyl alcohol
rinse. Last, the grid was annealed in an Ar–O2 [450 standard
cubic centimeters per minute (SCCM)/45 SCCM] environment at
500°C for 1.5 hours to prevent e-beam–induced hydrocarbon
deposition in the STEM (44, 45).

STEM experiment
STEM imaging was performed at room temperature using a Nion
UltraSTEM U100 microscope operated at 60 kV. The images were
acquired in HAADF imaging mode and were introduced to the deep
learning network without any postprocessing.

STM experiment
The STM experiments were performed at room temperature using an
Omicron VT-STM microscope equipped with a Nanonis controller in
an operating pressure of 2 × 10−10 Torr. The STM images were ob-
tained with mechanically cut Pt/Ir alloy tips.

DFT calculations
The DFT calculations were carried out using the VASP (Vienna Ab
Initio Simulation Package; 5.4.1) (46, 47) using the projector-augmented
wave (PAW) method (48, 49). The electron-ion interactions were
-described using standard PAW potentials, with valence electron config-
urations. A kinetic energy cutoff on the plane waves was set to 400 eV,
and the “accurate” precision setting was adopted. A Γ-centered k-point
mesh of 9 by 9 by 1 was used for the Brillouin zone integrations. The
convergence criteria for the electronic self-consistent loop were set
to 10−5 eV.

Data analysis
The deep neural networks were implemented using the Keras deep
learning library (https://keras.io/). The neural networks had an
encoder-decoder structure [SegNet-like architecture (50)] that allowed
a pixel-level classification of input images. For the atom finder network,
the encoder part consisted of six convolutional layers, all activated by a
rectified linear unit function. The convolutional filters (kernels) in all
the layers were of the size 3 by 3 and stride 1. The first convolutional
layer had 64 filters, and second and third layers had 128 filters each,
and the fourth to sixth layers had 256 filters each. The max pooling
layers were placed after the first, third and sixth layers. The decoder part
contained the same blocks of convolutional layers in reversed order
with a bilinear interpolated upsampling instead of the max pooling
units. The Adam optimizer (51) was used with categorical cross-
entropy as the loss function. For the “defect sniffer” network, the
encoder part consisted of three convolutional layers with 64, 128,
and 256 filters of the size 3 by 3 and stride 1, activated by a rectified
linear unit function. The max pooling layers were placed in between
the layers. The decoder part contained the same blocks of convolu-
tional layers in reversed order with the nearest-neighbor upsampling
between them. The focal loss (52) was used for improving identifica-
tion of the defect structures. To optimize this loss, the Adam optimizer
was used. The accuracy scores on a test set for atom finder and defect
sniffer were ≈96% and ≈99%, respectively. The graph structures were
constructed using NetworkX library (https://networkx.github.io/).
The latest version of a workflow for creating and exploring the defect libraries
is available at https://github.com/pcrscopy/AICrystallographer/tree/master/LibraryNet.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/
content/full/5/9/eaaw8989/DC1
Fig. S1. DFT-simulated distribution of electronic charge density for Si1C4 defect.
Fig. S2. DFT-simulated distribution of electronic charge density for Si2C4 defect.
Fig. S3. DFT-simulated distribution of electronic charge density for Si2C5 (type 2) defect.
Fig. S4. DFT-simulated distribution of electronic charge density for Si1C3 defect.
Fig. S5. Strain analysis for SiC3 defect.
Fig. S6. More complex defect structures observed in the STM experiment on STEM sample of
graphene.
Jupyter notebook (atom-finding accuracy versus noise)

REFERENCES AND NOTES
1. N. Jiang, E. Zarkadoula, P. Narang, A. Maksov, I. Kravchenko, A. Borisievich, S. Jesse,
S. V. Kalinin, Atom-by-atom fabrication by electron beam via induced phase
transformations. MRS Bull. 42, 653–659 (2017).
2. X. Zhao, J. Kotakoski, J. C. Meyer, E. Sutter, P. Sutter, A. V. Krisheninnikov, U. Kaiser,
W. Zhou, Engineering and modifying two-dimensional materials by electron beams.
MRS Bull. 42, 667–676 (2017).

Ziatdinov et al., Sci. Adv. 2019;5:eaaw8989 27 September 2019
19. B. Obradovic, R. Kotlyar, F. Heinz, P. Matagne, T. Rakshit, M. D. Giles, M. A. Stettler, M. L. Green, C. L. Choi, J. R. Hattrick-Simpers, A. M. Joshi, I. Takeuchi, S. C. Barron, E. Campo, G. Liang, N. Neophytou, D. E. Nikonov, M. S. Lundstrom, Performance projections for large Spatial Database with Noise (A4AIP, 1996), vol. 96, pp. 226–231.

20. M. Ziatdinov, A. Maksov, S. V. Kalinin, Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2. npj Comput. Mater. 5, 12 (2019).

21. J. Madsen, P. Liu, J. Kling, J. B. Wagner, H. T. Hansen, A. S. Admasu, J. Kim, S.-W. Cheong, A. Bhattacharya, D. G. Schlom, T. M. McQueen, R. Hovden, L. F. Kourkoutis. Image registration of low signal-to-noise cryo-STEM data. Ultramicroscopy 191, 56–65 (2018).

22. M. Ziatdinov, D. Yvick, A. Kalinin, S. V. Kalinin, Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2. npj Comput. Mater. 5, 12 (2019).

23. J. Barthel, Dr. Probe: A software for high-resolution STEM image simulation. Ultramicroscopy 193, 1 (2018).

24. B. Obradovic, B. B. Bagari, C. B. Clement, E. Waite, B. H. Goodrow, D. J. Baek, J. P. Shekellton, C. Pasco, H. Nair, N. Schreiber, J. Hoffman, A. S. Admasu, J. Kim, S.-W. Cheong, A. Bhattacharya, D. G. Schlom, T. M. McQueen, R. Hovden, L. F. Kourkoutis. Image registration of low signal-to-noise cryo-STEM data. Ultramicroscopy 191, 56–65 (2018).

25. A. B. Yankovich, B. Berkels, W. Dahmen, P. Binev, S. I. Sanchez, S. A. Bradley, A. Li, I. Sizaffurska, P. M. Vayle, Picometre-precision analysis of scanning transmission electron microscopy images of platinum nanocatalysts. Nat. Commun. 5, 4155 (2014).

26. Q. M. Ramasse, C. R. Seabourne, D. M. Kepaptsooglou, R. Zan, A. J. Norquist, Machine-learning-assisted materials discovery using a citration platform to examine data import, storage, and access. Adv. Theory Simul. 1, 1800037 (2018).

27. M. Ziatdinov, D. Yvick, A. Kalinin, S. V. Kalinin, Deep learning of atomically resolved scanning transmission electron microscopy images: Chemical identification and tracking local transformations. ACS Nano 11, 12742–12752 (2017).

28. J. Barthel, Dr. Probe: A software for high-resolution STEM image simulation. Ultramicroscopy 193, 1 (2018).

29. B. H. Savitzky, A. B. Bagari, C. B. Clement, E. Waite, B. H. Goodrow, D. J. Baek, J. P. Shekellton, C. Pasco, H. Nair, N. Schreiber, J. Hoffman, A. S. Admasu, J. Kim, S.-W. Cheong, A. Bhattacharya, D. G. Schlom, T. M. McQueen, R. Hovden, L. F. Kourkoutis. Image registration of low signal-to-noise cryo-STEM data. Ultramicroscopy 191, 56–65 (2018).

30. A. B. Yankovich, B. Berkels, W. Dahmen, P. Binev, S. I. Sanchez, S. A. Bradley, A. Li, I. Sizaffurska, P. M. Vayle, Picometre-precision analysis of scanning transmission electron microscopy images of platinum nanocatalysts. Nat. Commun. 5, 4155 (2014).

31. Q. M. Ramasse, C. R. Seabourne, D. M. Kepaptsooglou, R. Zan, A. J. Norquist, Machine-learning-assisted materials discovery using a citration platform to examine data import, storage, and access. Adv. Theory Simul. 1, 1800037 (2018).

32. M. Ziatdinov, D. Yvick, A. Kalinin, S. V. Kalinin, Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2. npj Comput. Mater. 5, 12 (2019).

33. J. Madsen, P. Liu, J. Kling, J. B. Wagner, H. T. Hansen, A. S. Admasu, J. Kim, S.-W. Cheong, A. Bhattacharya, D. G. Schlom, T. M. McQueen, R. Hovden, L. F. Kourkoutis. Image registration of low signal-to-noise cryo-STEM data. Ultramicroscopy 191, 56–65 (2018).

34. A. B. Yankovich, B. Berkels, W. Dahmen, P. Binev, S. I. Sanchez, S. A. Bradley, A. Li, I. Sizaffurska, P. M. Vayle, Picometre-precision analysis of scanning transmission electron microscopy images of platinum nanocatalysts. Nat. Commun. 5, 4155 (2014).

35. Q. M. Ramasse, C. R. Seabourne, D. M. Kepaptsooglou, R. Zan, A. J. Norquist, Machine-learning-assisted materials discovery using a citration platform to examine data import, storage, and access. Adv. Theory Simul. 1, 1800037 (2018).

36. A. B. Yankovich, B. Berkels, W. Dahmen, P. Binev, S. I. Sanchez, S. A. Bradley, A. Li, I. Sizaffurska, P. M. Vayle, Picometre-precision analysis of scanning transmission electron microscopy images of platinum nanocatalysts. Nat. Commun. 5, 4155 (2014).
Author contributions: S.V.K. and M.Z. conceived the project. O.D. performed the STEM measurements. M.Z. performed the STM measurements and the data analysis. B.G.S. performed the DFT calculations. R.K.V. and S.J. aided in the interpretation of results. X.L. aided in data analysis. S.V.K., M.Z., and R.K.V. wrote the paper with inputs from all the coauthors.

Competing interests: All authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Citation: M. Ziatdinov, O. Dyck, X. Li, B. G. Sumpter, S. Jesse, R. K. Vasudevan, S. V. Kalinin, Building and exploring libraries of atomic defects in graphene: Scanning transmission electron and scanning tunneling microscopy study. Sci. Adv. 5, eaaw8989 (2019).