ARTICLE

Unsupervised discovery of solid-state lithium ion conductors

Ying Zhang[1], Xingfeng He [2], Zhiqian Chen [3], Qiang Bai[2], Adelaide M. Nolan [2], Charles A. Roberts [1], Debasish Banerjee[1], Tomoya Matsunaga[1], Yifei Mo [2,4]* & Chen Ling[1]*

Although machine learning has gained great interest in the discovery of functional materials, the advancement of reliable models is impeded by the scarcity of available materials property data. Here we propose and demonstrate a distinctive approach for materials discovery using unsupervised learning, which does not require labeled data and thus alleviates the data scarcity challenge. Using solid-state Li-ion conductors as a model problem, unsupervised materials discovery utilizes a limited quantity of conductivity data to prioritize a candidate list from a wide range of Li-containing materials for further accurate screening. Our unsupervised learning scheme discovers 16 new fast Li-conductors with conductivities of 10^{-4}–10^{-1} S cm^{-1} predicted in ab initio molecular dynamics simulations. These compounds have structures and chemistries distinct to known systems, demonstrating the capability of unsupervised learning for discovering materials over a wide materials space with limited property data.

[1] Toyota Research Institute of North America, Ann Arbor, MI 48105, USA. [2] Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742, USA. [3] Department of Computer Science, Virginia Tech, 7054 Haycock Road, Falls Church, VA 2043, USA. [4] Maryland Energy Innovation Institute, University of Maryland, College Park, MD 20742, USA. *email: yfmo@umd.edu; chen.ling@toyota.com

The fast conduction of lithium (Li) ions in a solid is a phenomenon of significant scientific interest and technological importance. The room-temperature Li-ion conductivities (σ_{RT}) in poorly conductive and fast conducting materials can differ by more than twenty orders of magnitude[1,2]. The high σ_{RT} in electrode and electrolyte materials are essential for high power/rate performance of batteries. In particular, replacing the flammable liquid electrolyte used in commercial Li-ion batteries with a fast Li-conducting solid electrolyte, to produce an all-solid-state battery, provides improved safety, excellent stability, and long cycling life[1,2]. Although there are several thousands of known lithium-containing compounds, fast Li$^+$ conduction with σ_{RT} close to 10^{-3}–10^{-2} S cm^{-1}, comparable to the level in liquid electrolytes, is a rare property held by only a few solid-state Li-ion conductors (SSLCs), such as lithium thiophosphates (e.g., $Li_7P_3S_{11}$[3] and $Li_{10}GeP_2S_{12}$[4], LGPS), garnet (e.g., $Li_7La_3Zr_2O_{12}$[5], LLZO), NASICON (e.g., $Li_{1.3}Al_{0.3}Ti_{1.7}(PO_4)_3$[6], LATP), perovskite (e.g., $Li_{0.5}La_{0.5}TiO_3$[7], LLTO), Li_3N[8], and argyrodite[9] (e.g., Li_6PS_5Cl) (Fig. 1a). Since these known SSLCs do not meet all desired attributes required for the commercialization of all-solid-state batteries[10], there is significant interest in discovering new SSLC materials with high σ_{RT}. The challenges in predicting new SSLCs are largely a result of the diverse chemistries and structures of SSLCs, and current computational predictions and laboratory syntheses are often performed on a limited number of candidates[1,2]. SSLCs have compositions ranging from oxides and sulfides to nitrides and mixed halides, and a diverse set of crystalline structures including perovskite, argyrodite, garnet, and NASICON, and newly discovered structures, such as LGPS and $Li_7P_3S_{11}$. Over the past few years, first-principles computation has played an important role in the successful prediction of a number of novel SSLCs[11–15]. Recent studies have determined a number of key physical factors required for fast Li-ion diffusion, such as anion lattice packing[13], lattice dynamics[16,17], frustration of the mobile-ion sublattice[18,19], and concerted ion migration[14]. So far, transforming a theory into a predictive model to explore a vast composition-structure space of many materials remains a significant challenge.

Machine learning (ML) has emerged as a technique for materials discovery thanks to its capability of recognizing complex patterns in data[20–28] by representing materials with critical descriptors such as the combination of chemistry, composition, and crystal structure that yields desired materials properties. While significant research progress has been achieved by improving the materials descriptors over the years[29–37], the applications of ML for materials discovery is in general plagued by two significant challenges. First, a ML model requires training on a sufficient amount of data to capture the correlation between a desired physical property and the features of materials. Unfortunately, only a few select materials exhibit the property of interest, as is often the case in materials discovery. In many cases, even the data for materials with poor properties is scarce, due to lack of interest in performing and reporting these measurements. For example, most solids with poor ionic conductivity do not have conductivity data. The second challenge is that the parameterization of a ML model is highly susceptible to variances and errors in property data[38]. In the case of SSLCs, the conductivity obtained through experimental measurements can vary by a few orders of magnitude due to factors including synthesis method, sample preparation, and measurement technique[39]. For SSLCs, it is challenging to train a ML model of Li$^+$ conductivity from only a few compounds with known values of σ_{RT} with large variances and to make reliable predictions for thousands of compounds. This scarcity of high-quality property data greatly limits the applicability of supervised ML models to capture and predict complex

structure-property relationships over a broader space of materials beyond known examples.

Unlike supervised learning models, which require well-labeled training data, unsupervised learning can be readily applied to vast datasets regardless of whether any properties or labels exist. As a technique to draw inferences from features of data without explicitly labeled properties, unsupervised learning has been applied in materials science for feature extraction, pattern recognition, clustering, and phase mapping[40–44]. However, the application of unsupervised learning to directly discover new materials with enhanced properties has rarely been explored[27]. As shown in this study, unsupervised learning, through training on a broad range of materials, can draw boundaries between good and poor examples, identifying candidates similar to good examples, which are then further verified by more accurate first-principles calculations. This new approach using unsupervised learning for materials discovery has multiple advantages. Switching the target of ML from predicting the property (e.g., σ_{RT}) in supervised learning to grouping materials in unsupervised learning alleviates the issues of poor data quality and accuracy. Rather than predicting the targeted materials property accurately for each candidate, the output from unsupervised learning is a significantly narrowed list of materials candidates for subsequent exploration by more accurate first-principles calculations, thus significantly reducing the cost for an expensive high-throughput first-principles screening by utilizing a limited quantity of low-quality data. In addition, unsupervised learning uses unlabeled data and readily expands the applicability of the ML model to the entire materials space.

In this study, we propose an unsupervised learning scheme for guiding materials discovery, and demonstrate it for materials discovery of SSLCs. We apply unsupervised learning to screen all known Li-containing compounds from the Inorganic Crystal Structure Database. Our trained unsupervised learning models cluster Li-containing compounds into groups of SSLCs with high conductivity and other groups of materials with poor ionic conduction. Using ab initio molecular dynamics (AIMD) simulation to quantify σ_{RT} for predicted compounds[45], 16 new candidates having σ_{RT} exceeding 10^{-4} S cm^{-1} are identified, and three of them have σ_{RT} exceeding 10^{-2} S cm^{-1}, on par with known SSLCs with highest σ_{RT}. As proposed and demonstrated, our new approach of ML-guided materials discovery circumvents the data scarcity challenges, identifies new materials using a small number of known examples, and provides unique insight on structure-property relations.

Results

Scheme of the unsupervised discovery of SSLCs. We illustrate our scheme of the unsupervised discovery of SSLC materials in Fig. 1. In order to train the unsupervised model, a quantitative representation of the complex materials structure (Fig. 1a) is required as input. Instead of using a combination of hand-picked features, we used digital diffraction patterns of the crystal structure. Specifically, a representation for each crystal structure was built based on Bragg's law to map the three-dimensional periodic crystal lattice into a set of X-ray diffraction intensities at a fixed set of 2θ values (Method and Fig. 1b)[35,46,47]. Here, we only considered the anion lattice of the crystal structure, relying on the knowledge that anion configuration and Li$^+$-anion interactions significantly affect Li sites, diffusion channels, and the energy landscape of Li migration[1,13,15]. The anionic lattice was set to S anion and was scaled to the same atomistic volume, so that the representation was invariant to lattice parameter or the chemical constituent (Method). The resulting representation, called modified X-ray diffraction (mXRD), is unambiguously defined for
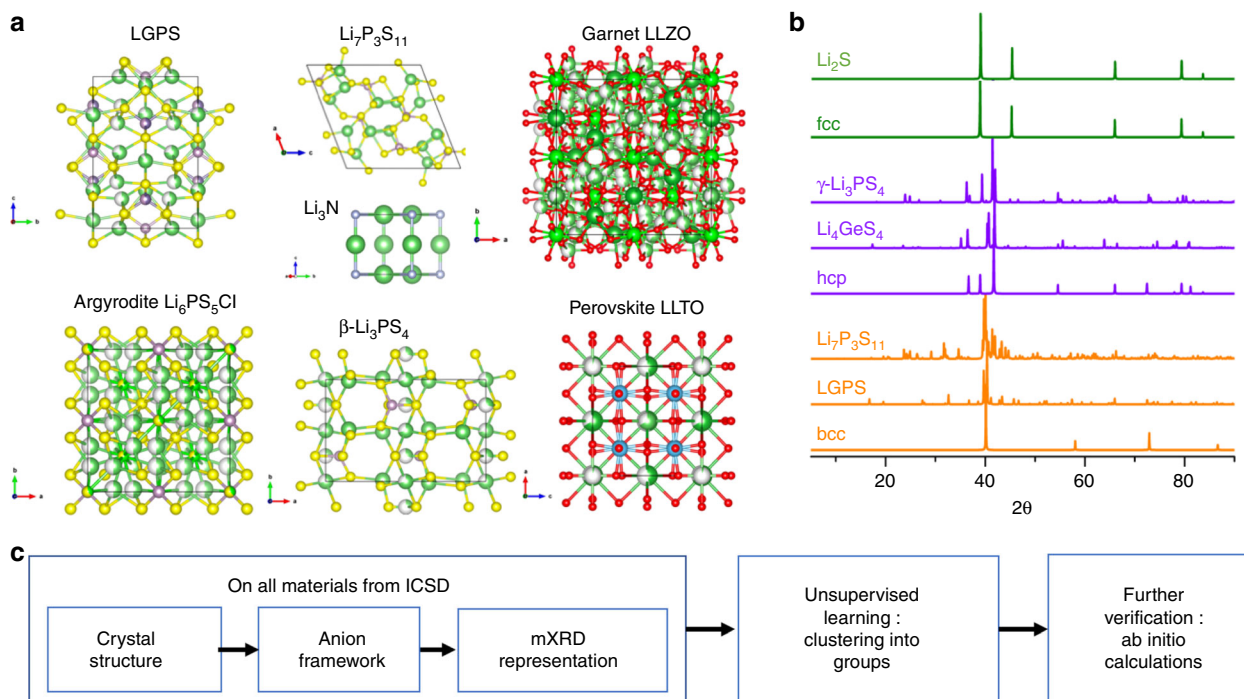
Fig. 1 Schematics of the unsupervised discovery of solid-state Li-ion conductors. **a** Crystal structures of known SSLCs, showing a large diversity of structure and chemistry. **b** mXRD patterns of selected materials in comparison to those of ideal fcc (face centered cubic), hcp (hexagonal close packed), bcc (body centered cubic) lattices. **c** Workflow of an unsupervised learning guided discovery of SSLCs

every anion lattice (Fig. 1b), fully capturing the anionic crystal structure information. Here, we performed our unsupervised discovery on 2986 compounds that contain lithium but not transition metals. Since some compounds have the same structure, one representative structure was used. A dataset of 528 representative anionic structures and their mXRDs were performed for the unsupervised learning (Method).

Unsupervised clustering of Li-containing compounds. We performed clustering, a common unsupervised learning technique, to group materials with similar mXRD representations. We first generated a model (named C_1) based on the agglomerative hierarchical clustering method to train a bottom-up grouping of the mXRD dataset (Method and Fig. 2a). The grouping showed a good quality of clustering as the mXRDs shared similar characteristics within the same groups (Fig. 2d) and different groups were well differentiated (Supplementary Note 1, Supplementary Fig. 1 and 2). More importantly, a visible clustering of SSLC materials is found using this model (Fig. 2b). Most known SSLCs with σ_{RT} close to 10^{-3}–10^{-2} S cm^{-1}, despite being structurally distinctive, were clustered into two groups in the center of the dendrogram out of a total seven groups, including LGPS, Li$_7$P$_3$S$_{11}$, LLZO, and Li$_3$N, in group VI, and argyrodite, β-Li$_3$PS$_4$, LLTO in group V. LATP, as an exception in group VII, lay close to the boundaries of group V and VII and its mXRD pattern still exhibited some similarity with group V. In addition, statistical analysis of σ_{RT} within the group quantitatively confirmed the correlation on σ_{RT} (Supplementary Note 1, Supplementary Figs 3 and 4). The violin plot of σ_{RT} of group V and VI showed significantly higher σ_{RT} (Fig. 2c), and the majority of compounds outside of group V and VI had σ_{RT} significantly below 10^{-4} S cm^{-1}. A statistically significant difference of σ_{RT} of the two groups V and VI versus the rest groups was proved by the t-test (Supplementary Note 2).

As confirmed by quantitative correlation between the groups and σ_{RT}, our unsupervised learning model captured the physical dependence of fast solid-state Li$^+$-diffusion on anion lattice. To critically assess the robustness of clustering in capturing the observed physical correlation, we performed three different clustering techniques. In addition to the aforementioned model, we trained a second model (named C_2) to create a top-down grouping by recursively applying divisive spectral clustering (Method). These two models were purely based on the mXRD dataset of anion lattices without seeing any labeled σ_{RT} data. In our third grouping, the model (named C_3) used the limited available σ_{RT} information to optimize the clustering of known SSLC examples (Supplementary Note 4). Despite the differences in the clustering methodologies of three models, the observed aggregation of fast-conducting examples was mostly consistent. Known SSLCs largely overlapped among the groups generated by these three models (Supplementary Notes 3–5, Supplementary Table 1 and Supplementary Figs 5–9). In particular, LGPS, Li$_7$P$_3$S$_{11}$, LLZO, and Li$_3$N were always clustered to the same group by all three models. Our results from three distinct models confirmed the reliability of clustering fast-conducting versus poor-conducting materials based on unsupervised learning using mXRD representations of the structures.

Physical insights from unsupervised learning. The clustering of SSLCs by mXRD provides new insight into the understanding of crystal structures exhibiting fast-ion conduction. While Li-ion diffusion in solids has been shown to correlate with various parameters, such as lattice volume[1], anion chemistry[48], bond ionicity[25,48], phonon mode[16,17], and Li coordination number[25], no single unified theory explains the similarity among highly distinctive crystal structures of all SSLCs. Our unsupervised clustering quantitatively confirmed the similarity among the mXRD patterns of anion lattice of SSLCs. The mXRD encodes the
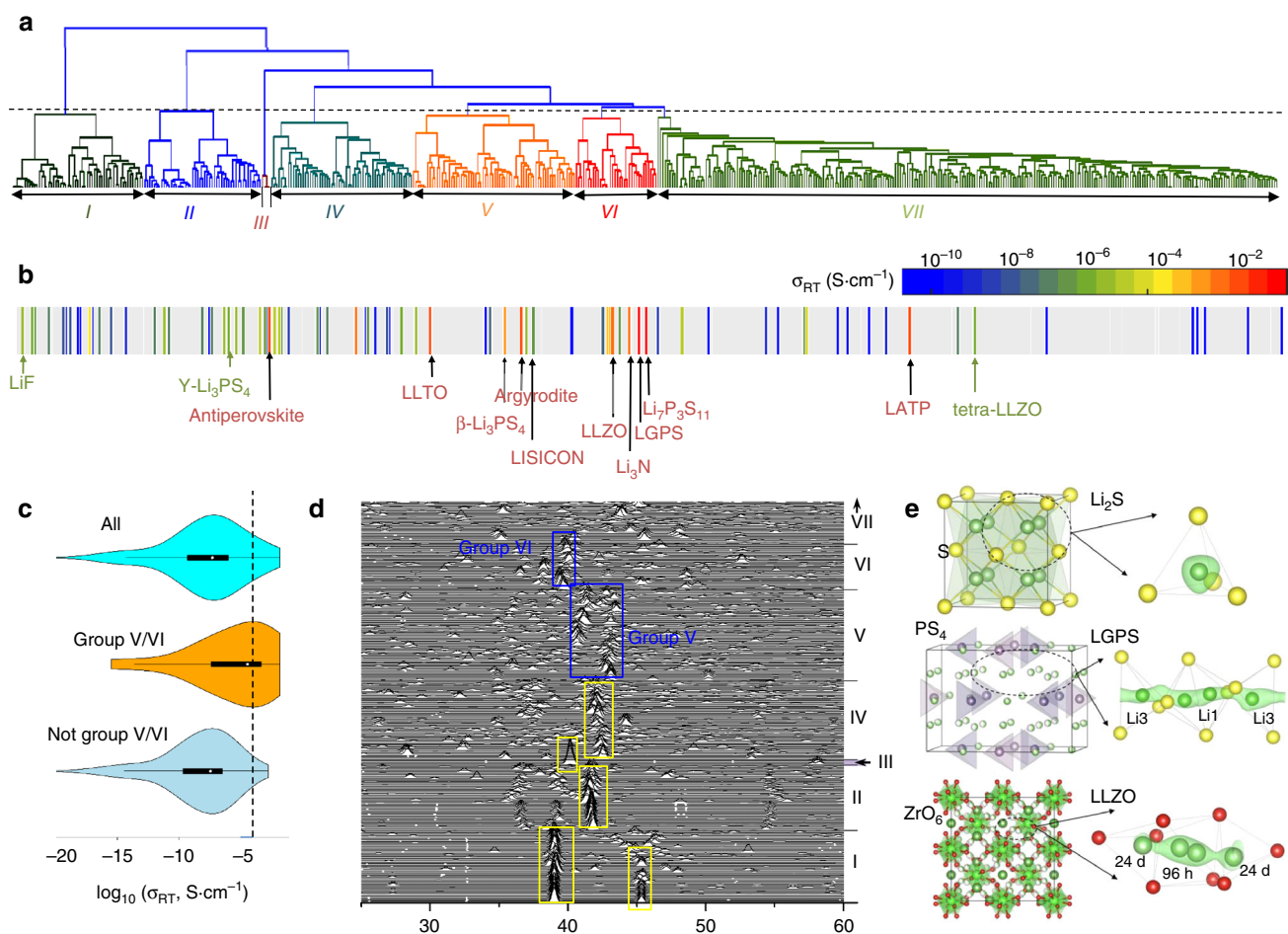
Fig. 2 Unsupervised clustering of all Li-containing compounds. **a** Bottom-up tree diagram (dendrogram) generated using the agglomerative hierarchical clustering method. The dashed line shows the position where all compounds are partitioned into seven groups, marked as I–VII from left to right and distinguished by different colors. **b** Mapping the dendrogram to the conductivity reveals the grouping of known solid-state Li-ion conductors in group V and VI. The color bar shows the scale of σ_{RT}. The gray color indicates the conductivity has not been measured for the corresponding compound. **c** Violin plots of σ_{RT} data grouped in the grouping. The outer shells of the violins bound all data, narrow horizontal lines bound 95% of the data, thick horizontal lines bound 50% of the data, and white dots represent medians. The dashed line shows the position of $\sigma_{RT} = 10^{-4}$ S cm^{-1}. **d** mXRD of all materials in group I–VI and a part of group VII. The colored boxes mark the positions of main characteristic peaks for each group. **e** Crystal structures (left) and (right) Li sites (green sphere) determined by local anion (yellow/red sphere) configuration, corresponding to isosurfaces (green) of Li probability density from AIMD simulations. Li$_2$S (top) with highly symmetric anion lattice and ordered Li sublattice versus LGPS (middle) and LLZO (bottom) SSLCs with distorted anion lattices and disordered Li sublattices

symmetry and ordering of the anionic lattice and showed strong correlation with ionic conductivity (Fig. 2). Given the information of lattice volume and anion chemistry critical for ion diffusion were removed from the mXRD descriptor, the resulted clustering of Li-conducting phases suggests that the long-range periodicity of the anion lattice as encoded in mXRD plays a fundamental role in Li-ion diffusion. By analyzing the structural origin of the clustered groups, (Supplementary Note 6), we found the materials in Group I, II, and III correspond to highly symmetrical fcc (face centered cubic), hcp (hexagonal close packed), and bcc anion lattices, respectively. For these anion lattices, Li ions are symmetrically confined in highly symmetric tetrahedral or octahedral sites of anions (as an example, Fig. 2e for Li$_2$S), and migrate among these well-defined sites[13]. Groups IV, V, and VI show a moderate level of variance, which can be understood as mild distortion of the anion lattices. The distortion of anion lattices disturbs Li$^+$ bonding environments and causes Li$^+$ to deviate from highly symmetric locations to geometrically frustrated configurations. For example, in LGPS and LLZO, the distorted

anion polyhedra generate multiple positions to host Li ions, observed as the spread Li-ion probability density observed in AIMD simulations (Fig. 2e), which were represented as partially occupied Li sites (e.g., Li1 and 96 h sites in LGPS and LLZO, respectively) from diffraction experiments[4,5]. Having multiple positions for Li$^+$ to occupy may lead to a degeneracy of Li sublattice energy and an entropically-enabled disordered-Li sublattice migrating among metastable configurations[18,19]. Therefore, as observed in their mXRD representations, the SSLCs clustered in group V and VI exhibit the characteristics of moderately distorted anion lattices, which is closely related to disordered Li sublattice for fast Li-ion conduction. The materials in Group VII, as reflected by the high standard deviation of mXRD peaks, correspond to the least symmetric and highly disordered anion lattices (Supplementary Figs 10–12). The highly disordered anion lattices in these materials may locally trap Li ions and impede Li-ion percolation across the crystal structure (Supplementary Fig. 13), resulting in the low conductivities observed for compounds in this group.

SSLC confirmed by AIMD simulations. Given the successful clustering of known SSLC materials by unsupervised learning models, the other structures clustered into the same groups are expected to exhibit fast Li-ion conduction. To further assess the conductivity of these compounds discovered from the unsupervised grouping, we conducted AIMD simulations, which have been demonstrated as a highly accurate and predictive computation approach for calculating Li ion conductivity[11,14,15,45]. From the screening of initial 2989 compounds from the ICSD, we narrowed the evaluation of the ion-conduction property down to 82 unique compounds, which were from the intersection of these fast-conducting groups in the aforementioned three models. Thus, our unsupervised learning scheme successfully reduced a high throughput screening of thousands of compounds to a focused exploration of <100 candidates with much higher success probability. Among these, we rediscovered LiZnPS$_4$, which was previously discovered by the bcc-anion-packing rule and was confirmed with an experimental σ_{RT} of 5.7×10^{-4} S cm^{-1}[11–13]. According to AIMD simulations (Fig. 3), 16 more candidates are predicted to have σ_{RT} higher than 10^{-4} S cm^{-1}. In particular, three new materials systems, Li$_8$N$_2$Se, Li$_6$KBiO$_6$ and Li$_5$P$_2$N$_5$, have σ_{RT} exceeding 10^{-2} S cm^{-1}, a conductivity higher than that of the best known SSLCs. A list of these materials and the calculated Li$^+$ conduction properties are summarized in Supplementary Tables 3–4 and Supplementary Fig. 14. Figure 3 plots the predicted σ_{RT} and activation energy of newly discovered SSLCs (filled symbols), in comparison with σ_{RT} reported in the past few decades (open symbols, Supplementary Table 2). The newly discovered SSLCs are in the upper left corner of Fig. 3, which corresponds to high σ_{RT} of $>10^{-5}$ S cm^{-1} and low E_a of 0.17–0.45 eV. More importantly, these SSLCs comprise new structures, chemistries, and compositions significantly different from known SSLCs, demonstrating the capability of our crystal-structure-based unsupervised learning model to discover materials beyond existing chemistries.

Discussion

A fraction of compounds uncovered by our grouping did not show fast Li-ion diffusion in AIMD simulations (Supplementary Tables 5–7). Among these compounds, a majority exhibit too small of a percolation radius for Li-ion migration, a blocking of diffusion network by other cations, or a poorly connected diffusion network. The inclusion of these compounds was attributed to the fact that our unsupervised models were trained solely on the anionic geometry without considering factors such as the effects of other cations. In addition, some compounds with low ionic conductivity may be further optimized via doping or tuning Li concentration. Future extension of our scheme should attempt to include features in addition to the anion lattice for more accurate prediction.

For these Li-ion conductors to be utilized as solid electrolytes for solid-state Li-ion batteries, other materials properties, such as electrochemical window, interface compatibility, and mechanical properties[1,2,10,15,16], are also required. We employed the first-principles computation techniques established in the previous studies[10,15] to evaluate the thermodynamic intrinsic electrochemical window of these newly identified ion conductors (Supplementary Fig. 15). Consistent with the general trend identified in the previous studies[15], most of the materials have limited electrochemical windows. Many identified nitrides are stable with Li metal in agreement with the previous computation study[49], while other compounds are not stable against Li metal or at low potential due to the reduction of cations. The identified fluorides have a
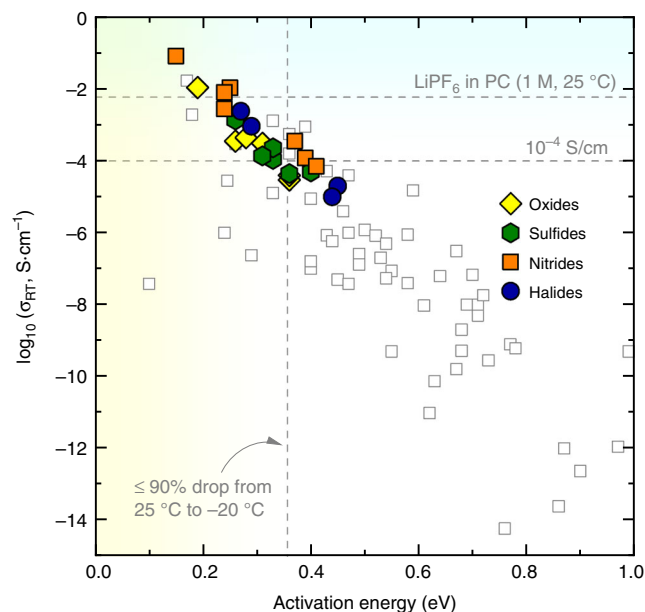

Fig. 3 Ion conducting properties of newly predicted versus known solid-state Li-ion conductors. The open symbol shows the experimental conductivity reported in the literature (Supplementary Table 2 and references therein). The horizonal dashed lines show the room temperature conductivity of 10^{-4} and 0.006 S cm^{-1}. The latter value is the conductivity of 1 M LiPF$_6$ in propylene carbonate (PC) solution. The vertical dashed line shows the activation energy at 0.356 eV, which corresponds to one order of magnitude change of conductivity when the temperature drops from 25 to −20 °C

very high oxidation limit of >6 V, which may be ideal for stable protection of high-voltage cathodes. The oxides have decent electrochemical windows but most have relatively low ionic conductivities of ~10^{-4} S cm^{-1} (except for Li$_6$KBiO$_6$). The identified sulfides have narrow windows but these two sulfides may have significantly better air/moisture stability than currently used thio-phosphates. In summary, while our discovery does not identify an ionic conductor that outcompetes current solid electrolytes, the potential choices of fast ion conductors with improvements in certain aspects (such as stability against Li metal, high voltage, or air) are predicted from the computation discovery. The properties and applicabilities of these materials in solid-state batteries may require further computational or experimental studies and optimizations.

In summary, the unsupervised learning models succeeded in distinguishing fast Li-conducting and poor Li-conducting materials, leading to the prediction of sixteen new compounds as solid-state Li-ion conductors with room-temperature conductivities higher than 10^{-4} S cm^{-1} with a few new compounds exceeding 10^{-2} S cm^{-1}. These newly discovered candidates have highly different structures and chemical compositions from current known fast Li-ion conductors, demonstrating the effectiveness of our unsupervised learning approach for discovering new materials over a wide materials space. This novel unsupervised learning approach also reveals the unique structure–property relationship between anion lattice and Li$^+$ conduction over a large materials space. Whereas the supervised learning has been widely adopted in the majority of machine-learning studies for materials, our unsupervised learning scheme, which narrows a high-throughput screening to a focused prioritized list by utilizing a limited amount of low-quality data,

presents a different approach of using ML for materials discovery, and is generally applicable for other physical properties.

Methods

Data preprocessing. The raw data of crystalline structures were exported from the Inorganic Crystalline Structure Database (ICSD) in the format of *cif* files[50]. The range of analysis in the current study includes all compounds containing Li but not transition metal species, except Sc, Y, La, Ti, and Zr. The exclusion of transition metal species is based on the consideration that compounds containing transition metal ions are usually redox active and hence may not be suitable for application as solid-state electrolytes. These filters yielded a total of 2986 ICSD entries (ver. November 2016). The representative structures for each entry was identified either as the "chemical_name_structure_type" flag in the *cif* files or as the chemical formula if this flag was not set explicitly. The entries that were structurally similar in the hierarchical clustering were further filtered to remove duplicates in the training set. The final training set included 528 unique representative structures for the unsupervised learning analysis.

Representation. The anionic sublattice of Li-containing compounds is uniquely represented in the X-ray diffraction pattern based on Bragg's law. For the diffraction from (hkl) plane, the angle is determined by

$$2 \sin(\theta) = \lambda \cdot d_{hkl}^{-1} \qquad (1)$$

where the interplane distance d_{hkl} is a function of the size and shape of the unit cell

$$d_{hkl}^2 = h^2 a^{*2} + k^2 b^{*2} + l^2 c^{*2} + 2hka^*b^* \cos(\gamma) + 2hla^*c^* \cos(\beta) + 2klb^*c^* \cos(\alpha) \qquad (2)$$

The intensity is determined by the amplitude of light scattered from the lattice plane

$$F_{hkl} = \sum_{j=1}^{m} N_j f_j \exp[2\pi i(hx_i + ky_i + lz_i)] \qquad (3)$$

where the sum runs over all atoms of the unit cell on (hkl) plane, N_j is the fraction of every equivalent position that is occupied by atom j at coordinates (x_j, y_j, z_j). The scattering factor f_j is a product describing the interaction of the X-ray with the electrons around an atom. Using Eqs. 1, 2, and 3, the X-ray diffraction of a periodic lattice is determined by the size and shape of unit cell, as well as the position and identity of atoms on a given plane. The following procedure was employed to obtain the XRD representation of the geometry of anion sublattice of the crystalline structure. First, we removed all cations from the crystalline structure, keeping only the anionic sublattice in the unit cell. Second, we substituted the remaining anions for a unitary species (e.g., S^{2-}), removing the influence of the scattering factor f on the diffraction intensity. Third, the unit cell was isotropically expanded or compressed to a pre-determined volume per anion of 40 Å^3, removing the effect of unit cell size on the position of the diffraction peaks. After these initial steps, the X-ray diffraction of modified lattice encodes only information for the geometry and topology of anion sublattice. The calculation of diffraction pattern is then performed at a fixed set of 2θ values from 0 to 89.98° at a step size of 0.1° using the *pymatgen* package[51], generating a 900-dimensional vector for each diffraction pattern[51]. We confirmed the results of hierarchical clustering was consistent when the step size was increased to 0.02° (Supplementary Note 7). A Gaussian smearing was then performed to normalize the integrated intensity of diffraction to a unitary value.

Unsupervised learning. We used the *Dendrogram* function from the *SciPy* package to perform agglomerative hierarchical clustering (AHC)[52]. In AHC, each sample starts in its own cluster, and the clusters merge progressively according to the similarity metric as one moves up the hierarchy. The output from AHC is a bottom-up hierarchical tree diagram (dendrogram). The Euclidean distance (L2) between two diffraction profiles was used as the similarity metric and Ward linkage was used to measure the cluster dissimilarity[53]. The same clustering results were also obtained using the *hclust* package in *R*.

In addition to the hierarchical clustering, we used the *kernlab* package in *R* to perform spectral clustering. Spectral clustering uses the eigenvalues of the similarity matrix of the data to divide the samples in to K groups, where K is a manually selected integer[54]. To create hierarchical grouping results, we recursively applied the bisectional divide ($K = 2$) on the larger portion from the previous grouping, and obtained a divisive top-down hierarchical diagram after the clustering.

First-principles calculation. All Density Functional Theory (DFT) calculations were performed using the Vienna Ab initio Simulation package (VASP) within the projector augmented-wave approach and Perdew–Burke–Ernzerhof (PBE) generalized-gradient approximation (GGA) functionals[55–57]. The parameters in static DFT calculations were consistent with the *Materials Project*[58]. Ab initio molecular dynamics (AIMD) simulations were performed in supercell models using non-spin-polarized DFT calculations with a Γ-centered k-point. The time step was set to 2 fs. The initial structures were statically relaxed and were set to an initial temperature of 100 K. The structures were then heated to targeted

temperatures at a constant rate by velocity scaling during 2 ps. During the estimation of Li ion diffusion, NVT ensemble using Nosé–Hoover thermostat was adopted. The total time of AIMD simulations were in the range of 100 ps to 1000 ps until the diffusivity was converged. The ionic diffusivity and conductivity were calculated following established method in previous study[45].

References

1. Bachman, J. C. et al. Inorganic solid-state electrolytes for lithium batteries: mechanisms and properties governing ion conduction. *Chem. Rev.* **116**, 140–162 (2016).
2. Zhang, Z. et al. New horizons for inorganic solid state ion conductors. *Energy Environ. Sci.* **11**, 1945–1976 (2018).
3. Mizuno, F., Hayashi, A., Tadanaga, K. & Tatsumisago, M. New, highly ion-conductive crystals precipitated from Li$_2$S–P$_2$S$_5$ glasses. *Adv. Mater.* **17**, 918–921 (2005).
4. Kamaya, N. et al. A lithium superionic conductor. *Nat. Mater.* **10**, 682–686 (2011).
5. Kumazaki, S. et al. High lithium ion conductive Li$_7$La$_3$Zr$_2$O$_{12}$ by inclusion of both Al and Si. *Electrochem. Commun.* **13**, 509512 (2011).
6. Aono, H., Sugimoto, E., Sadaoka, Y., Imanaka, N. & Adachi, G.-Y. Ionic conductivity and sinterability of lithium titanium phosphate system. *Solid State Ion.* **40/41**, 38–42 (1990).
7. Ibarra, J. et al. Influence of composition on the structure and conductivity of the fast ionic conductors La$_{2/3−x}$Li$_{3x}$TiO$_3$ (0.03≤x≤0.167). *Solid State Ion.* **134**, 219–228 (2000).
8. Li, W. et al. Li$^+$ ion conductivity and diffusion mechanism in α-Li$_3$N and β-Li$_3$N. *Energy Environ. Sci.* **3**, 1524–1530 (2010).
9. Boulineau, S., Courty, M., Tarascon, J.-M. & Viallet, V. Mechanochemical synthesis of Li-argyrodite Li6PS5X (X = Cl, Br, I) as sulfur-based solid electrolytes for all solid state batteries application. *Solid State Ion.* **221**, 1–5 (2012).
10. Zhu, Y., He, X. & Mo, Y. Origin of outstanding stability in the lithium solid electrolyte materials: insights from thermodynamic analyses based on first-principles calculations. *ACS Appl. Mater. Interfaces* **7**, 23685–23693 (2015).
11. Richards, W. D., Wang, Y., Miara, L. J., Kim, J. C. & Ceder, G. Design of Li$_{1+2x}$Zn$_{1-x}$PS$_4$, a new lithium ion conductor. *Energy Environ. Sci.* **9**, 3272–3278 (2016).
12. Suzuki, N. et al. Synthesis and electrochemical properties of I4̄-Type Li$_{1+2x}$Zn$_{1-x}$PS$_4$ solid electrolyte. *Chem. Mater.* **30**, 2236–2244 (2018).
13. Wang, Y. et al. Design principles for solid-state lithium superionic conductors. *Nat. Mater.* **14**, 1026–1031 (2015).
14. He, X., Zhu, Y. & Mo, Y. Origin of fast ion diffusion in super-ionic conductors. *Nat. Commun.* **8**, 15893 (2017).
15. Nolan, A. M., Zhu, Y., He, X., Bai, Q. & Mo, Y. Computation-accelerated design of materials and interfaces for all-solid-state lithium-ion batteries. *Joule* **2**, 2016–2046 (2018).
16. Muy, S. et al. Tuning mobility and stability of lithium ion conductors based on lattice dynamics. *Energy Environ. Sci.* **11**, 850–859 (2018).
17. Muy, S. et al. Lithium conductivity and Meyer–Neldel rule in Li$_3$PO$_4$–Li$_3$VO$_4$–Li$_4$GeO$_4$ lithium superionic conductors. *Chem. Mater.* **30**, 5573–5582 (2018).
18. Kweon, K. E. et al. Structural, chemical, and dynamical frustration: origins of superionic conductivity in closo-borate solid electrolytes. *Chem. Mater.* **29**, 9142–9153 (2017).
19. Kozinsky, B. et al. Effects of sublattice symmetry and frustration on ionic transport in garnet solid electrolytes. *Phys. Rev. Lett.* **116**, 055901 (2016).
20. Ren, F. et al. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Sci. Adv.* **4**, eaaq1566 (2018).

21. Tran, K. & Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nat. Catal.* **1**, 696–703 (2018).

22. Lu, S. et al. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* **9**, 3405 (2018).

23. Yuan, R. et al. Accelerated discovery of large electrostrains in $BaTiO_3$-based piezoelectrics using active learning. *Adv. Mater.* **30**, 1702884 (2018).

24. Xue, D. et al. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **7**, 11241 (2016).

25. Sendek, A. D. et al. Holistic computational structure screening of more than 12 000 candidates for solid lithium-ion conductor materials. *Energy Environ. Sci.* **10**, 306–320 (2017).

26. Zhang, Y. & Ling, C. A strategy to apply machine learning to small datasets in materials science. *NPJ Comput. Mater.* **4**, 25 (2018).

27. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).

28. Meredig, B. et al. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014).

29. Behler, J. & Parrinello, M. Generalized neural network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).

30. Rupp, M., Tkachenko, A., Müller, K.-R., Lilienfeld, V. & Anatole, O. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).

31. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **96**, 019902 (2012).

32. Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017).

33. Ward, L. et al. Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B* **96**, 024104 (2017).

34. Isayev, O. et al. Universal fragment descriptors for predicting electronic properties of inorganic crystals. *Nat. Commun.* **8**, 15679 (2017).

35. Ziletti, A., Kumar, D., Scheffler, M. & Ghiringhelli, L. M. Insightful classification of crystal structures using deep learning. *Nat. Commun.* **9**, 2775 (2018).

36. Zhu, L. et al. A fingerprint based metric for measuring similarities of crystalline structures. *J. Chem. Phys.* **144**, 034203 (2016).

37. Winter, R., Montanari, F., Noé, F. & Clevert, D. A. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem. Sci.* **10**, 1692–1701 (2019).

38. Park, J., Howe, J. D. & Sholl, D. S. How reproducible are isotherm measurements in metal–organic frameworks? *Chem. Mater.* **29**, 10487–10495 (2017).

39. Thangadurai, V., Narayanan, S. & Pinzaru, D. Garnet-type solid-state fast Li ion conductors for Li batteries: critical review. *Chem. Soc. Rev.* **43**, 4714–4727 (2014).

40. Balachandran, P. V., Theiler, J., Rondinelli, J. M. & Lookman, T. Materials prediction via classification learning. *Sci. Rep.* **5**, 13285 (2015).

41. Isayev, O. et al. Materials cartography: representing and mining materials space using structural and electronic fingerprints. *Chem. Mater.* **27**, 735–743 (2015).

42. Zhou, Q. et al. Learning atoms for materials discovery. *PNAS* **115**, E6411–E6417 (2018).

43. Long, C. J. et al. Rapid structural mapping of ternary metallic alloy systems using the combinatorial approach and cluster analysis. *Rev. Sci. Instrum.* **78**, 072217 (2007).

44. Kireeva, N. et al. Generative topographic mapping (GTM): universal tool for data visualization, structure-activity modeling and dataset comparison. *Mol. Inform.* **31**, 301–312 (2012).

45. He, X., Zhu, Y., Epstein, A. & Mo, Y. Statistical variances of diffusional properties from ab initio molecular dynamics simulations. *NPJ Comput. Mater.* **4**, 18 (2018).

46. Iwasaki, Y., Kusne, A. G. & Takeuchi, I. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. *NPJ Comput. Mater.* **3**, 4 (2017).

47. Park, W. B. et al. Classification of crystal structure using a convolutional neural network. *IUCrJ* **4**, 486–494 (2017).

48. Kraft, M. A. et al. Influence of lattice polarizability on the ionic conductivity in the lithium superionic argyrodites Li_6PS_5X (X = Cl, Br, I). *J. Am. Chem. Soc.* **139**, 10909–10918 (2017).

49. Zhu, Y., He, X. & Mo, Y. Strategies based on nitride materials chemistry to stabilize Li metal anode. *Adv. Sci.* **4**, 1600517 (2017).

50. Hellenbrandt, M. The inorganic crystal structure database (ICSD)—present and future. *Crsytallogr. Rev.* **10**, 17–22 (2014).

51. Ong, S. P. et al. Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).

52. Jones, E., Oliphant, T. & Peterson, P. SciPy: Open source scientific tools for Python. http://www.scipy.org/ (2001).

53. Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classfic.* **31**, 274–295 (2014).

54. Luxburg, U. V. A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416 (2007).

55. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).

56. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).

57. Kresse, G. & Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

58. Jain, A. et al. The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).

Acknowledgements

Author contributions

C.L. and Y.M. supervised the project. C.L. conceived the idea of unsupervised learning for materials discovery. Y.Z., Z.C. and C.L. performed the unsupervised learning and discussed the results with D.B. and T.M. Y.M. designed the computation workflow of AIMD simulation and X.H., Q.B. and C.L. performed the AIMD simulation. C.L. and Y.M. analyzed the results and wrote the manuscript with the help from A.N. and C.R. All authors participated in discussing the results and finalizing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information