

ARTICLE

<https://doi.org/10.1038/s41467-019-10533-1>

OPEN

Materials informatics for the screening of multi-principal elements and high-entropy alloys

J.M. Rickman^{1,2}, H.M. Chan², M.P. Harmer², J.A. Smeltzer², C.J. Marvel², A. Roy³ & G. Balasubramanian³

The field of multi-principal element or (single-phase) high-entropy (HE) alloys has recently seen exponential growth as these systems represent a paradigm shift in alloy development, in some cases exhibiting unexpected structures and superior mechanical properties. However, the identification of promising HE alloys presents a daunting challenge given the associated vastness of the chemistry/composition space. We describe here a supervised learning strategy for the efficient screening of HE alloys that combines two complementary tools, namely: (1) a multiple regression analysis and its generalization, a canonical-correlation analysis (CCA) and (2) a genetic algorithm (GA) with a CCA-inspired fitness function. These tools permit the identification of promising multi-principal element alloys. We implement this procedure using a database for which mechanical property information exists and highlight new alloys having high hardnesses. Our methodology is validated by comparing predicted hardnesses with alloys fabricated by arc-melting, identifying alloys having very high measured hardnesses.

¹ Department of Physics, Lehigh University, Bethlehem, PA 18015, USA. ² Department of Materials Science and Engineering, Lehigh University, Bethlehem, PA 18015, USA. ³ Department of Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, PA 18015, USA. Correspondence and requests for materials should be addressed to J.M.R. (email: jmr6@lehigh.edu)

The study of high-entropy (HE) (or multi-principal element) alloys, typically comprising five or more elements, is a relatively new area of materials research that has attracted intense interest in recent years^{1–6} as, in many cases, these systems possess unexpected and superior mechanical properties relative to those of conventional alloys^{7–10} as well as enhanced oxidation resistance and magnetic properties^{11,12}. In general, these alloys exhibit crystal structures and phases that are believed to be entropically stabilized due to the large number of elements present. In this regard, the early work by Yeh and co-workers^{6,13} revealed that a veritable cocktail of metallic elements resulted in a far greater degree of solid–solid solubility than is achievable in mixtures comprising fewer components. Despite substantial progress in this area, however, there remain significant challenges in understanding the origin of these superior properties from a fundamental point of view¹⁴. For example, the role of complex strengthening mechanisms, including solid–solution strengthening, in determining measured alloy hardness is still a subject of vigorous debate.

It is therefore of considerable interest to identify element combinations and associated compositions leading to high-strength, high-hardness alloys. However, given a large palette of possible elements, the number of potential HE alloys that may be fabricated is exceedingly large. Unfortunately, only a relatively small subset of these alloys is expected to have desirable properties. Thus, to obviate fruitless experimentation in the search for this desirable subset, it is imperative to identify, a priori, candidate systems that are likely to have a high degree of solid–solid solubility and enhanced thermomechanical properties.

In recent years, several groups have devised computational strategies to identify and characterize HE alloys^{15–17}. For example, Troparevsky et al.¹⁷ used “high-throughput” density-functional theory calculations of formation energies to predict which combination of elements is likely to form an HE alloy. In addition, Senkov et al.¹⁸ employed a calculated phase diagram (CALPHAD)-based combinatorial approach to screen computationally a large number of candidate metal alloys for those forming only solid–solution phases. Beyond these studies, other workers have formulated predictive metrics, such as average melting temperature and average valence electron concentration, to characterize relatively simple correlations between the metrics and the propensity to form HE alloys. Most recently, Sarker et al. proposed using a new descriptor, the entropy-forming ability, to predict HE alloys having high hardnesses and validated its use by combining first-principles and experimental synthesis¹⁹. While quite useful, these methods have not really explored the complex inter-relationships among many metrics and the thermomechanical properties of HE alloys nor the use of this information to accelerate alloy design.

With this in mind, we describe here the use of data analytics to accelerate the discovery of new, useful multi-component alloys by determining maximal correlations among metrics and alloy properties in experimental databases and the exploitation of this information to search for and identify promising HE alloy candidates. In particular, by combining a multiple regression analysis or its generalization, a canonical correlation analysis (CCA), with a genetic algorithm (GA) optimization strategy, we will explore systematically the HE alloy chemistry/composition space to highlight those alloys that have beneficial mechanical properties (e.g., high hardness). A CCA is a very general technique for quantifying relationships between two sets of variables, most parametric tests of significance being essentially special cases of CCA²⁰, and it has been used recently to: explore, for example, correlations between ceramic powder chemistry and the resulting microstructure of a dense, sintered ceramic^{21,22}. In this context, a GA is employed to construct many virtual

candidate alloys that evolve from one generation to the next by processes that mimic reproduction and mutation, and in which survival to the next generation is dependent upon a measure of fitness. The fitness measure governing alloy selection will be determined from the aforementioned multiple regression/CCA results. This computational approach allows one to screen hypothetical alloys relatively quickly, thereby accelerating the identification and design of new alloy systems. We implement this procedure using a database comprising 82 HE alloys for which reliable mechanical property information is available to highlight new alloys having potentially high hardnesses.

Results

Metrics. We began by compiling a list of $M = 82$ experimentally fabricated HE alloys for which there are measured (Vicker’s) hardnesses^{8,10,13,23–35}. For each system, we also used (or computed) the values of alloy metrics that have been employed in previous studies to predict the potential for high solid solubility. In particular, for each element i having an atomic radius of r_i , melting temperature, $(T_m)_i$, Young’s modulus, E_i , and valence electron concentration, VEC_i , we considered the following quantities calculated in terms of the molar compositions, c_i , for each of the N constituent elements:

$$\text{Radius asymmetry, } \delta = \sqrt{\sum_{i=1}^N c_i \left(1 - \frac{r_i}{\bar{r}}\right)^2}$$

$$\text{Enthalpy of mixing, } \Delta H_{\text{mix}} = 4 \sum_{i=1, j \neq i}^N (\Delta H_{\text{mix}})_{ij} c_i c_j$$

$$\text{Ideal entropy of mixing, } \Delta S_{\text{mix}} = -R \sum_{i=1}^N c_i \ln c_i$$

$$\text{Mean melting temperature, } T_m = \sum_{i=1}^N c_i (T_m)_i$$

$$\text{Entropy/enthalpy ratio, } \Omega = \frac{T_m \Delta S_{\text{mix}}}{|\Delta H_{\text{mix}}|}$$

$$\text{Young’s modulus asymmetry, } \varepsilon = \sqrt{\sum_{i=1}^N c_i \left(1 - \frac{E_i}{\bar{E}}\right)^2}$$

$$\text{Valence electron concentration, } VEC = \sum_{i=1}^N c_i VEC_i$$

In these expressions, the bar denotes the composition-weighted mean, $(\Delta H_{\text{mix}})_{ij}$ is the regular solution enthalpy of mixing associated with elements i and j calculated from Miedema’s model³⁶, and R is the gas constant.

Parallel coordinate plot and correlation analysis. It is of interest to identify significant relationships between the alloy metrics (i.e., predictor variables) and the corresponding thermomechanical properties (i.e., outcome variables) of the alloy. For this purpose, four such outcomes were examined, namely: (1) the presence of body-centered cubic (bcc) solid solution(s) only, (2) the presence of face-centered cubic (fcc) solid solution(s) only, and (3) the presence of intermetallic (IM) phase(s), and (4) the Vicker’s hardness (H). To establish a basis for comparison, it is useful to relate both predictor and outcome variables to those of a reference alloy, conveniently chosen here to be CoCrFeNiCu, and to denote the ratio of a given variable to that of a reference variable with a prime. This choice was made given that the reference alloy is relatively well studied. Figure 1 summarizes these normalized metrics and outcomes for the 82 HE alloys in the form of a parallel coordinate plot. This type of plot replaces the conventional d -dimensional orthogonal Cartesian axes by a set of d parallel axes, with each d -dimensional point represented by a polyline in parallel coordinates, and is very useful for displaying high-dimensional data^{37–39} and identifying some global trends. As is evident from the figure, there does not appear to be any strong correlation between phase behavior and hardness, whereas, for a subset of alloys, there is some correlation between valence electron concentration and hardness.

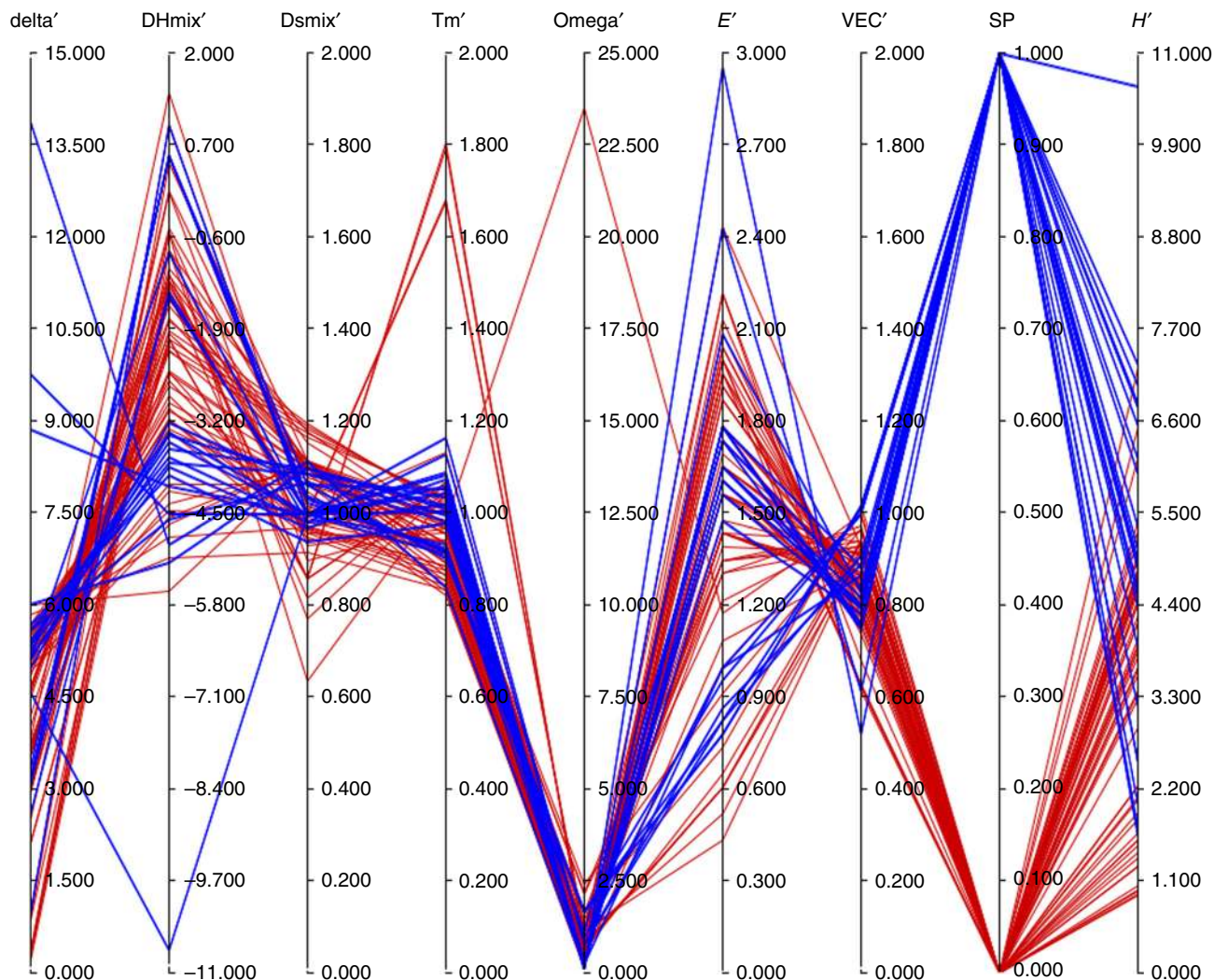


Fig. 1 Parallel coordinate plot. A parallel coordinate plot displaying both normalized metrics and property data for a set of high-entropy (HE) alloys. The phase behavior of the alloys is summarized in the variable single phase (SP) (unnormalized), with $SP = 1$ indicating alloys comprising a single, solid-solution (either fcc or bcc) and $SP = 0$ otherwise. Solid-solution alloys are shown in blue, with the remaining alloys having intermediate phases, etc., shown in red

To be more quantitative, we employ a multiple regression analysis to assess the relative importance of various materials characteristics whose complex interplay dictates properties. Our analysis begins with a vector of predictor variables, x , and outcome variables, y , taken as

$$\begin{aligned} x &= \{\delta', \Delta H'_{\text{mix}}, \Delta S'_{\text{mix}}, T'_m, \Omega', \varepsilon', \text{VEC}'\}, \\ y &= \{H'\}, \end{aligned} \quad (1)$$

where y consists of only H' for a multiple regression. (A more general CCA analysis that is useful when there are additional outcome variables in y is described below in the “Discussion” section). It should be noted that correlations, some of which are significant, exist among the predictor variables as can be seen by inspection of a color map of the correlation matrix (Fig. 2) given below.

From this analysis, the correlation coefficient was found to be 0.79 ± 0.07 with $p \approx 0$. An examination of the resulting canonical weights (i.e., the α_i summarized in Table 1) reveals that three of the predictor variables, namely $\Delta S'_{\text{mix}}$, $\Delta H'_{\text{mix}}$, and VEC' , are especially relevant in determining the hardness.

More specifically, from Table 1 it can be seen that $\Delta S'_{\text{mix}}$ are positively correlated with the H , suggesting that compositional disorder enhances the hardness. By contrast, $\Delta H'_{\text{mix}}$ and VEC' are negatively correlated with the hardness, suggesting in the first case that phase separation is associated with lower hardness values (HV). In the latter case, the negative correlation of VEC' with hardness is consistent with recent findings that a low VEC' is associated with an improvement in strength via the promotion of a bcc phase⁴⁰.

Finally, one may ask whether the three principal metrics, $\Delta S'_{\text{mix}}$, $\Delta H'_{\text{mix}}$, and VEC' , embody most of the physics that determines the hardness. If so, one would expect that an analysis based only on this reduced set of variables would satisfactorily predict the hardness. To assess the adequacy of this (null) hypothesis, we conducted a correlation analysis of this reduced model and computed the resulting F -statistic that compares the sum-of-squares errors for the reduced and full models⁴¹. The resulting p -value of 0.81 suggests that the reduced model is indeed adequate. It should be emphasized, though, that further microstructural (and possibly computational) analysis is needed to confirm these statistical insights.

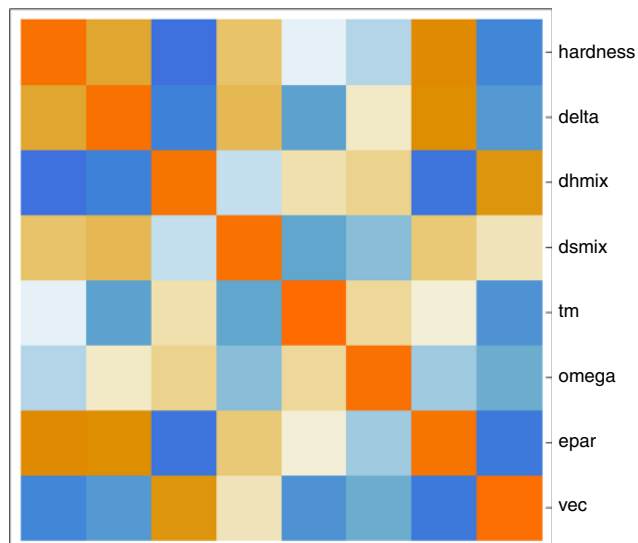


Fig. 2 Correlation matrix map. A color map of the correlation matrix, with both alloy metrics and the hardness labeled. The color scale ranges from red to blue, with dark red indicating a correlation of +1 and dark blue a correlation of -1. No correlation (0) corresponds to white

Table 1 The standardized canonical weights found in the CCA with hardness as the sole output variable

δ'	$\Delta H'_{\text{mix}}$	$\Delta S'_{\text{mix}}$	T'_m	Ω'	ϵ'	VEC'
-0.047	-0.553	0.285	-0.095	-0.108	0.057	-0.476

CCA canonical correlation analysis, VEC valence electron concentration

Figure 3 shows the values for the correlated variates, $V^{(1)}$ and $W^{(1)} = (H' - \langle H' \rangle) / \sigma$, for each of the M alloys and, in addition, a regression line that highlights the relationship between the variates. (The prime here denotes a value normalized by the reference alloy, the angle brackets denote an average over a dataset comprising the $M = 82$ HE alloys, and σ denotes the standard deviation in H' over the dataset). Despite the adequacy of the reduced model, for greater accuracy the multiple regression analysis for the full model will be used below to identify promising alloys.

Alloy identification using a GA. Having obtained canonical variates from the above analysis, the variate $V^{(1)}$ was used as a fitness function in a GA to find candidate alloys having high hardness. While it may be possible to narrow the search space to some degree by analytical optimization, the non-linear dependence of some of the alloy metrics on composition suggested a numerical solution using all of the metrics. More specifically, to construct candidate, 5-element alloys, a 16-element palette and 16 molar compositions per element were employed to represent $\binom{16}{5} (16)^5$ independent alloys. The alloys in this palette are Co, Cr, Fe, Ni, Al, Cu, Mn, Ti, Mo, Nb, Ta, V, W, Zr, Zn and Sn.

The calculation began with $N_c = 500$ randomly selected chromosomes, each a 40-bit string encoding the chemistry and composition of an alloy having distinct elements, with the i th chromosome having a fitness f_i . Successive generations were produced with a series of evolutionary processes, including fitness selection, recombination, and mutation. In particular, from each

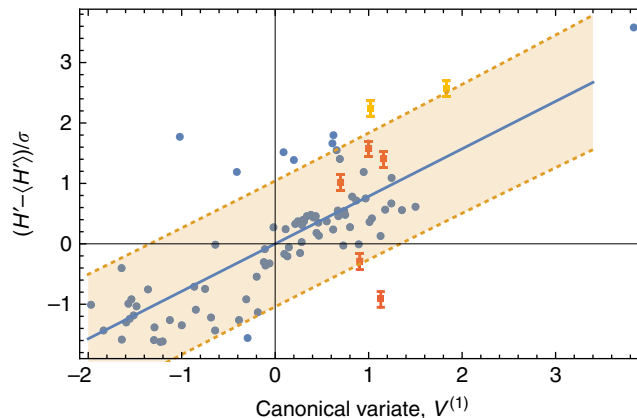


Fig. 3 Correlation analysis. The canonical variate, $V^{(1)}$, versus the canonical variate, $W^{(1)} = (H' - \langle H' \rangle) / \sigma$, for each alloy (blue circle) and the corresponding regression line (solid blue). The Vicker's hardness is H and σ denotes the standard deviation in normalized hardness. The dotted lines delimit a shaded 90% (single observation) prediction band. Also shown are the measured (normalized) micro-hardnesses for 7 fabricated alloys (5 orange squares, with the 2 alloys having hardnesses in excess of 1000 hardness values (HV) in yellow.) The error bars were estimated from approximately 20 separate hardness measurements per sample. The standard error of the estimate, as obtained using the data for the synthesized alloys, is approximately 259.2 HV

generation, N_c chromosomes were selected with probability $p_i = f_i / \sum_j f_j$ for the i th chromosome. (Thus a given chromosome may be, and often is, chosen more than once.) These selected chromosomes were then subjected to a recombination process in which pairs are chosen with probability $p_r = 0.05$, and then, upon randomly selecting bits, chromosomal strands are exchanged between a given pair. Finally, a mutation process was modeled in which randomly selected bits, chosen with probability $p_m = 0.001$, were changed from 0 to 1, or vice versa. If any of these processes resulted in a chromosome with one or more repeated elements, a new chromosome comprising distinct elements was generated. It should be noted that some tuning of the parameters p_r and p_m was required to create an algorithm in which successive generations have markedly improved fitness distributions. In practice, one observes the evolution of the fitness function over successive generations and then adjusts the parameters so that there is substantial improvement over many generations. Multiple trials consisting of at least 1000 generations were used to identify superior alloys.

The utility of these may also depend, at least in part, on their ductility. Given the inherent tradeoff between hardness and ductility in most systems, it is useful to also screen for ductility in this procedure. This may be accomplished via the implementation of a multi-objective (Pareto) optimization GA in which a composite objective function is formulated from a weighted sum of two objective functions, namely, one for the hardness (as obtained from the canonical variate calculation) and one for the ductility⁴². As a proxy for ductility measurements, it is useful to employ the empirical Pugh criterion⁴³ that correlates ductility with the ratio $P = \mu/B$, where in this context μ is the composition-weighted mean shear modulus and B is the corresponding bulk modulus. Ductile behavior is then associated with $P < 0.5$. In this work, we have chosen to simply evaluate P for promising candidates, as summarized below.

Figure 4 displays the cumulative distribution of fitnesses both at the beginning of the GA and after 1000 generations. As is evident from the figure, the population of alloys better matches

the hardness criterion after many iterations. From these simulations, we identified those alloys that are associated with the largest fitness values as candidates for high-hardness alloys.

Comparison with synthesized alloys. To assess the predictive capability of our methodology, we fabricated and hardness tested seven candidate alloys that were identified as promising from the aforementioned GA. Given that the GA produces an extremely fit population of high-hardness alloys that represent at optimality an extreme extrapolation from the $M = 82$ dataset, we conservatively select from this population a subset of alloys whose predicted hardnesses are somewhat beyond the range of the existing dataset (i.e., having fitness values that are about 10–15% greater than the maximum found with the multiple regression). Our view is that, used in this way, the GA systematically creates a large pool of potential candidates having relatively high hardnesses for a range of fitness values that represents a conservative extrapolation of the data. Parameter tuning in the GA is then performed to create this large pool. This extrapolation can be tested step by step, and if desired, the new alloys discovered in this way may be then added to the existing data to bootstrap the extrapolation. The candidate HE alloys were synthesized by arc-melting compressed pellets of elemental, high-purity powders (Sigma-Aldrich, purity $\geq 99.9\%$). (Powdered metals were used to limit the occurrence of macro-segregation of elements and to improve homogeneity. Also, arc-melted pellets were remelted four times to ensure homogeneity.) The resulting as-cast, solid pieces were then subjected to micro-hardness testing with a 100 g load to determine their respective Vicker's HV, the values taken to be the mean of 20 independent tests per piece.

The resulting scaled hardnesses are also displayed in Fig. 3 as a function of the canonical variate, $V^{(1)}$. It can be seen from the figure that about 5 out of 7 of the hardnesses are within the pictured 90% prediction interval. The highest measured hardnesses were found for $\text{Co}_{33}\text{W}_{07}\text{Al}_{33}\text{Nb}_{24}\text{Cr}_{03}$ (1084 ± 37 HV) and $\text{Ti}_{18}\text{Ni}_{24}\text{Ta}_{12}\text{Cr}_{22}\text{Co}_{24}$ (1011 ± 20 HV) that have corresponding Pugh ductility ratios of $P = 0.38$ and $P = 0.47$, respectively (For comparison, the Vicker's HV for alloys such as Ti-5Al, Ti-5V, and 75Au-7Cu-13Ag-5Co (wt. %) are 532.3, 429.3 and 128, respectively^{44,45}). Indeed, for each of the 7 candidate alloys identified here, $P < 0.5$. (However, for the hardest alloy a close inspection of the hardness indentations did reveal some degree of brittleness as cracking occurred near indentation corners in addition to observable cooling cracks.) The HV for the 7 synthesized alloys are summarized in the table below.

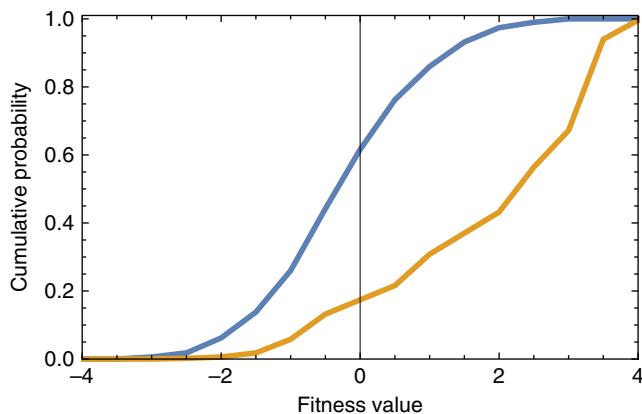


Fig. 4 Genetic algorithm (GA) fitness distribution. The cumulative distribution of fitnesses both at the beginning (shown in blue) of the GA and after 1000 generations (shown in orange). Clearly, the alloy distribution better matches the hardness criterion after many iterations

In summary, given the number of candidate alloys considered here and the inherent scatter in the original data, this analysis provides a convenient means to predict alloys with enhanced hardness with 90% confidence.

Microstructural and microchemical characterization. To understand better the observed plastic response of the synthesized alloys, the microstructure of the samples with compositions $\text{Co}_{33}\text{W}_{07}\text{Al}_{33}\text{Nb}_{24}\text{Cr}_{03}$ and $\text{Ti}_{39}\text{W}_{04}\text{Nb}_{31}\text{Ta}_{04}\text{Co}_{22}$, the alloys having the highest and lowest HV (see Table 2), respectively, were examined using scanning electron microscopy (SEM). Figures 5 and 6 depict the microstructure of $\text{Co}_{33}\text{W}_{07}\text{Al}_{33}\text{Nb}_{24}\text{Cr}_{03}$ using secondary electron intensity, as well as the associated compositional maps of the constituent elements obtained by X-ray energy-dispersive spectroscopy (EDS). It can be seen that the microstructure is dendritic (which is consistent with a cast alloy), with a relatively small volume fraction (13 vol. %) of inter-dendritic phase. The EDS data show that the dendritic phase is enriched in Nb and W. Given that there is no obvious partitioning of the Nb and W, it is assumed that they form a solid solution. This behavior would not be unexpected given that both Nb and W are bcc and exhibit complete solid solubility⁴⁶. The SEM also reveals that the inter-dendritic phase exhibits a two-phase, eutectic-like morphology. Of the two phases, one is Nb rich, whereas the other contains higher proportions of Al, Co, and Cr, with little Nb. Unlike for the case of the dendrite body, it does not appear that there is strong preferential association of the W with the Nb-rich component.

Table 2 The Vicker's hardness values (in HV) for the synthesized alloys

Alloy	Hardness (HV)	Hardness (HV) predicted
$\text{Co}_{33}\text{W}_{07}\text{Al}_{33}\text{Nb}_{24}\text{Cr}_{03}$	1084 ± 37	825
$\text{Ti}_{18}\text{Ni}_{24}\text{Ta}_{12}\text{Cr}_{22}\text{Co}_{24}$	1011 ± 20	677
$\text{Co}_6\text{W}_9\text{Al}_{36}\text{Mo}_{38}\text{Ni}_{11}$	725 ± 47	618
$\text{Ni}_{47}\text{Co}_{02}\text{Ta}_{12}\text{Ti}_9\text{Nb}_{30}$	815 ± 43	702
$\text{Ti}_{44}\text{Ni}_{02}\text{Nb}_{21}\text{Cr}_{21}\text{Co}_{12}$	422 ± 13	656
$\text{Ti}_{32}\text{Nb}_9\text{Ta}_{01}\text{Cr}_{19}\text{Co}_{39}$	856 ± 29	673
$\text{Ti}_{39}\text{W}_{04}\text{Nb}_{31}\text{Ta}_{04}\text{Co}_{22}$	277 ± 12	697

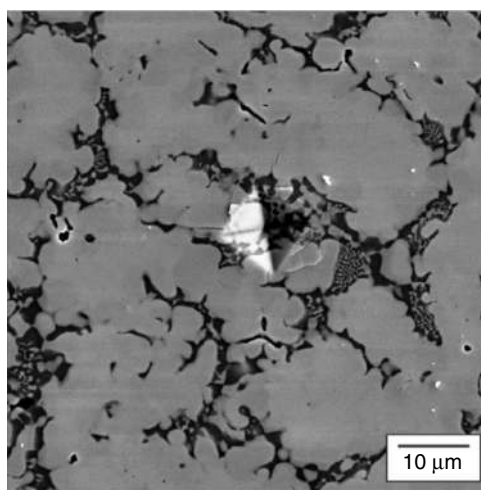


Fig. 5 Micrograph of a hard alloy. A scanning electron microscopic micrograph in a region near a microhardness indent in the $\text{Co}_{33}\text{W}_{07}\text{Al}_{33}\text{Nb}_{24}\text{Cr}_{03}$ sample that was synthesized by arc-melting. The figure evinces a dendritic microstructure with eutectic-like inter-dendritic regions

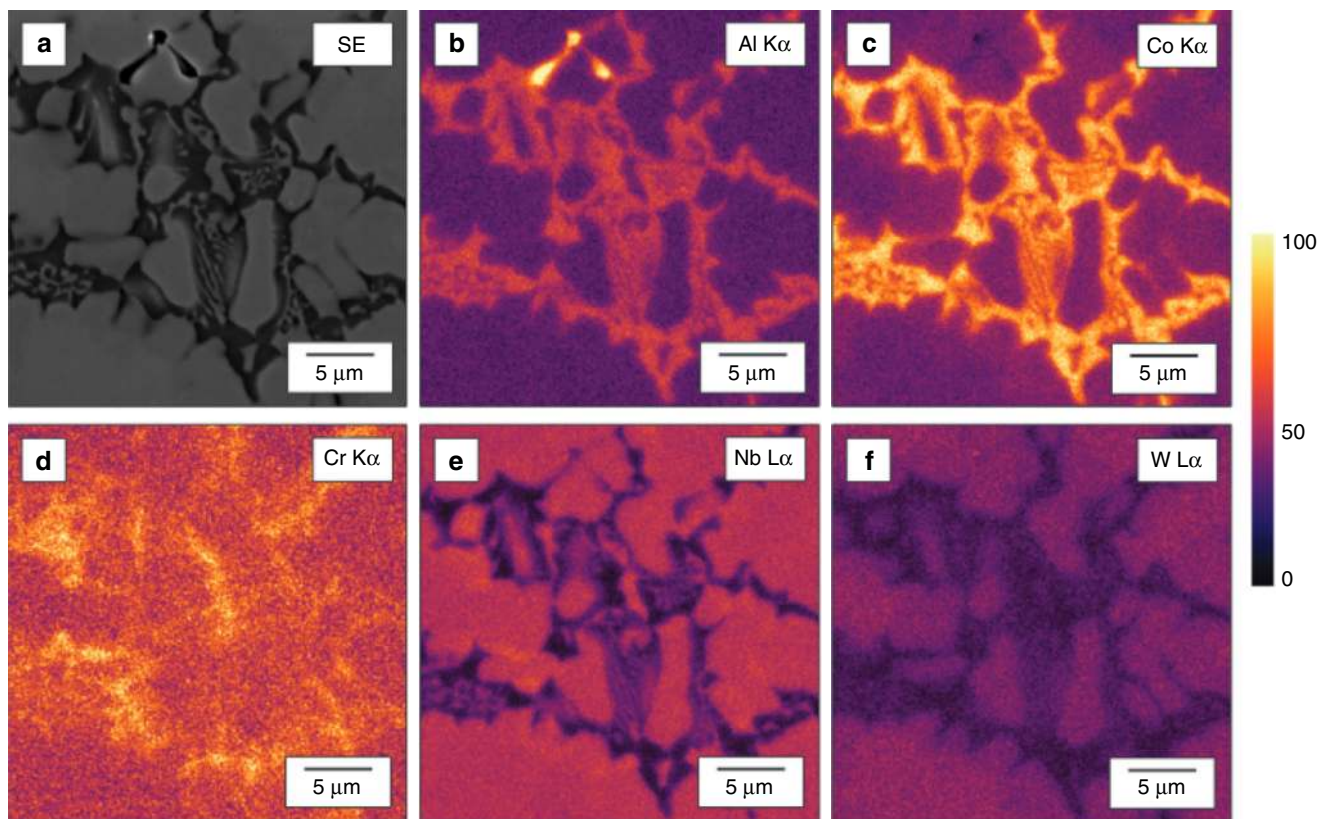


Fig. 6 Energy-dispersive spectroscopic (EDS) map of a hard alloy. An X-ray EDS map of the as-cast microstructure of $\text{Co}_{33} \text{W}_7 \text{Al}_{33} \text{Nb}_{24} \text{Cr}_{03}$. **a** Secondary electron image and X-ray intensity maps produced by integrating **b** Al $\text{K}\alpha$, **c** Co $\text{K}\alpha$, **d** Cr $\text{K}\alpha$, **e** Nb $\text{L}\alpha$, and **f** W $\text{L}\alpha$ X-ray signals. The color scale represents the relative intensity based on the number of counts per pixel. The map suggests that the dendritic region of the microstructure is rich in Nb and W while the eutectic inter-dendritic region comprises two phases, one rich in Nb and the other primarily composed of Al, Co, and Cr

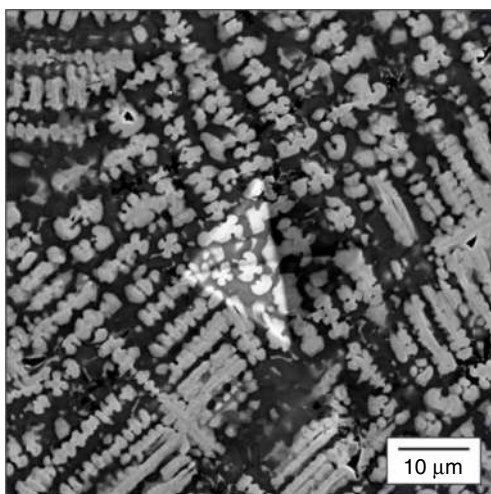


Fig. 7 Micrograph of a soft alloy. A scanning electron microscopic micrograph in a region near a microhardness indent in $\text{Ti}_{39} \text{W}_4 \text{Nb}_{31} \text{Ta}_{04} \text{Co}_{22}$ in a sample that was synthesized by arc-melting. The figure shows a dendritic microstructure similar to that of traditionally cast alloys

The corresponding images for the softest alloy are shown in Figs 7 and 8. This microstructure of $\text{Ti}_{39} \text{W}_4 \text{Nb}_{31} \text{Ta}_{04} \text{Co}_{22}$ is also dendritic, but the volume fraction of inter-dendritic material is much higher here (53 vol. %). In this alloy, the dendrites are enriched in Nb, Ta, and W, whereas the inter-dendritic regions contain a greater proportion of Co and Ti. In the SEM image

(Fig. 7), the slight differences in contrast at the edges of the dendrites suggest the possibility of coring. This observation is confirmed by the compositional maps where it can be seen that, for a given dendrite structure, the spatial extent of Nb and Ta enrichment is the greatest while W is confined to the dendrite inner core (see, in particular, Fig. 8g). Thus one infers that, although all three elements are present in the dendritic regions, the spatial distribution is inhomogeneous.

The marked difference in hardness between the two alloys is very interesting, particularly given that the compositional make-ups of the dendritic phases are not that dissimilar. In both cases, there is a significant component of Nb and W. In the case of the softest alloy, Ta is also present. One possible explanation is that the high hardness of $\text{Co}_{33} \text{W}_7 \text{Al}_{33} \text{Nb}_{24} \text{Cr}_{03}$ is attributable to the significant volume fraction of the Nb–W phase, which is solid solution strengthened. One could suggest that the same degree of strengthening is not present in $\text{Ti}_{39} \text{W}_4 \text{Nb}_{31} \text{Ta}_{04} \text{Co}_{22}$ either due to the non-uniform distribution of the elements or the additional presence of Ta. Of course, the potential contribution of alloying elements present at relatively low concentrations cannot be excluded.

Finally, we have also assessed the ductility of each synthesized alloys by optical microscopic inspection of the (20) microhardness indentations per alloy for signs of cracks at the corners of the indentations. It was found that 2 of the 7 systems showed moderate to considerable ductility (i.e., $\gtrsim 50\%$ of the indentations without cracks), including the relatively hard alloys $\text{Ti}_{32} \text{Nb}_9 \text{Ta}_{01} \text{Cr}_{19} \text{Co}_{39}$ and $\text{Co}_6 \text{W}_9 \text{Al}_{36} \text{Mo}_{38} \text{Ni}_{11}$, while the hard alloy $\text{Ti}_{18} \text{Ni}_{24} \text{Ta}_{12} \text{Cr}_{22} \text{Co}_{24}$ showed limited ductility (i.e., $\approx 15\text{--}20\%$ of the indentations without cracks). In addition, the relatively low hardness alloys, $\text{Ti}_{39} \text{W}_4 \text{Nb}_{31} \text{Ta}_{04} \text{Co}_{22}$ and $\text{Ti}_{44} \text{Ni}_{02} \text{Nb}_{21}$

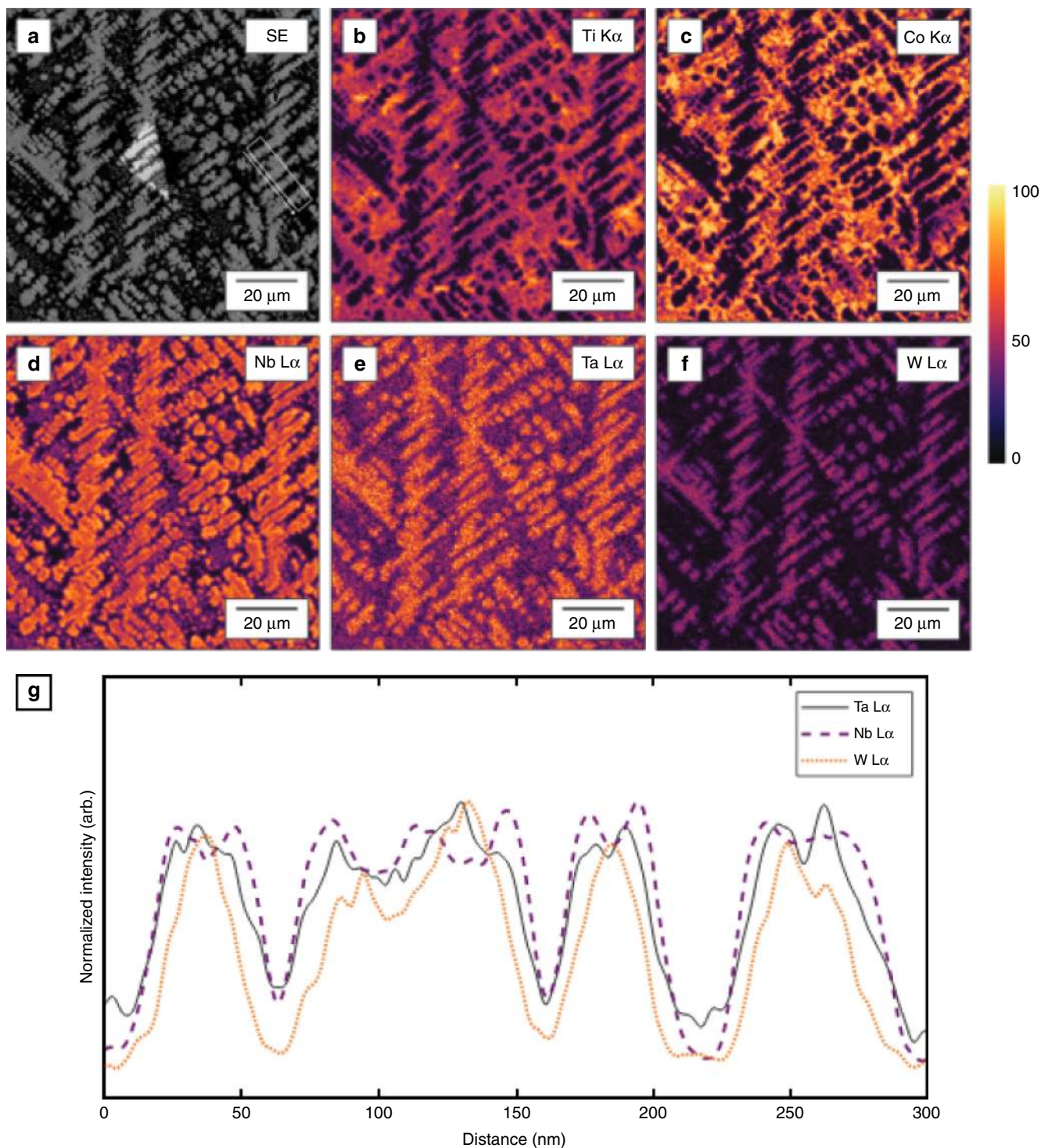


Fig. 8 Energy-dispersive spectroscopic (EDS) map of a soft alloy. An EDS map of the as-cast microstructure of $\text{Ti}_{39}\text{W}_{04}\text{Nb}_{31}\text{Ta}_{04}\text{Co}_{22}$ and evidence of coring. **a** Secondary electron image and X-ray intensity maps produced by integrating **b** Ti K α , **c** Co K α , **d** Nb L α , **e** Ta L α , and **f** W L α X-ray signals. The color scale represents the relative intensity based on the number of counts per pixel. The dendrites were found to be rich in Nb, Ta, and W, while the interdendritic phase is composed primarily of Co and Ti. Evidence of dendrite coring can be seen in **g**, a line profile showing the normalized intensity of W L α , Ta L α , and Nb L α X-ray signals taken from the box in **a** with an arrow indicating the direction of the line profile from left to right

$\text{Cr}_{21}\text{Co}_{12}$, were found to be quite ductile (i.e., $\approx 95\%$ of the indentations without cracks), which is generally consistent with observations that Group IV element additions, such as Ti, are often associated with increased ductility⁴⁷. The remaining alloys were found to be brittle. Given the empirical nature of the Pugh criterion, it is perhaps unsurprising that it is a rough predictor of ductility in

this context. Nevertheless, with this methodology we have identified several alloys that are both hard and relatively ductile.

Discussion

We present here a data analytics approach to screen efficiently multi-principal component alloys using a multiple regression

analysis and a GA. Our methodology permits the exploration of the multi-dimensional chemistry/composition space that is associated with these alloys and thereby the identification of alloys having beneficial mechanical properties. In particular, two alloys were discovered with extremely high HV in excess of 1000 HV, and these predictions were validated by micro-hardness testing of alloys that were fabricated via arc-melting. The most promising alloys were also screened to ensure that they possessed reasonable ductility.

Given the analysis described above, it is intriguing to consider the role of minor elements in determining alloy hardness, especially for $\text{Co}_{33}\text{W}_{07}\text{Al}_{33}\text{Nb}_{24}\text{Cr}_{03}$. Our methodology permits a detailed investigation of the impact of small composition changes in a reduced parameter space containing, for example, these minor elements. We therefore examined the behavior of the fitness function obtained from the regression for small changes in either Cr alone or W alone. It was found that the maximum hardness occurred for a Cr composition close to 0.07 and for a W composition close to 0.04. Thus, for experimental guidance, it is possible to fine tune the predictions made here by considering small compositional perturbations.

The methodology summarized here can, of course, be employed to examine the influence of the predictor set on hardness and phase behavior, as summarized in Eq. (2) below. This analysis is useful if one wishes to obtain hard alloys that are also likely to be a single solid solution. To perform this analysis, one employs a CCA with an extended set of output variables, \vec{y} , (see below) for the calculations of canonical weights and variates. Moreover, the CCA and GA methodologies are quite general and can therefore be applied to properties other than hardness. As noted in the “Introduction” section, a CCA and its relative, a Monte Carlo CCA, have been used in a very different context to establish correlations between ceramic powder chemistry and the resulting microstructure of a dense, sintered ceramic. A similar CCA analysis has also been used to relate the microstructure to the optoelectronic properties of thin-film solar cells²². Thus, in this context, the CCA can also be generalized to identify alloys having other useful thermomechanical and kinetic properties (e.g., yield strength, electrical conductivity, corrosion resistance) or, as described above, to Pareto-optimize multiple properties simultaneously using a multi-objective GA. Moreover, this approach can also be employed to investigate the impact of processing on property measurements. For example, recent work suggests that a nominally single-phase, HE alloy, namely, $\text{Mo}_{25}\text{Nb}_{25}\text{Ta}_{25}\text{W}_{25}$, synthesized by mechanical alloying comprised, in fact, multiple phases. Its associated complex microstructure, along with any subsequent annealing of the system, may dictate, at least in part, its measured mechanical properties (Smeltzer, J. A. unpublished work (2018)). The investigation of such processing/property correlations is the subject of ongoing work.

Finally, to illustrate the use of the CCA to examine multiple output variables, we consider the following set of predictor (x) and output (y) variables, respectively, for the HE alloy example.

$$\begin{aligned} x &= \{\delta', \Delta H'_{\text{mix}}, \Delta S'_{\text{mix}}, T'_m, \Omega', \epsilon', \text{VEC}'\}, \\ y &= \{\text{fcc}, \text{bcc}, \text{IM}, H'\}, \end{aligned} \quad (2)$$

where fcc, bcc, and IM denote the presence of a fcc solid solution, a bcc solid solution, or intermediate phase(s), respectively. One then finds linear combinations (known as canonical variates) $V = \sum_i \alpha_i x_i$ and $W = \sum_i \beta_i y_i$, where the coefficients α_i and β_i are called canonical weights, such that these combinations are maximally correlated. For this purpose, one constructs the associated correlation (or covariance) matrix, Σ , and defines the

operators σ_1 and σ_2 from the blocks of Σ as^{48,49}

$$\begin{aligned} \sigma_1 &= \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}, \\ \sigma_2 &= \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}. \end{aligned} \quad (3)$$

The use of the correlation matrix here implies that one is examining relationships among standardized variables⁴⁸. The square roots of the eigenvalues of σ_1 and σ_2 are the canonical correlations. Moreover, the eigenvectors corresponding to the maximum eigenvalue are the desired canonical variates that maximize the correlation. We note that the CCA methodology has been extended to identify non-linear variable combinations that are highly correlated^{22,50,51}.

We performed a CCA using the HE alloy dataset described above. From this analysis, one finds three significant pairs of canonical variates having correlation coefficients 0.82, 0.63 and 0.53, respectively, with associated errors of approximately ± 0.06 . The significance of these results is determined by employing a hypothesis test in which one tests the null hypothesis that a given pair of variates is uncorrelated via the construction of an appropriate statistic, such as Wilks lambda⁵². The small calculated p -values found here are each < 0.001 , indicating that a correlation between variates exists. From this investigation, one can identify cases in which high hardness is associated with solid solution behavior and other cases in which high hardness is associated with the presence of IM phases. An examination of the variates indicates that the first variate (i.e., the one having the largest correlation coefficient) is most associated with the hardness and accounts for nearly 50% of the variability. The remaining variates primarily reveal associations between the alloy metrics and the phase behavior variables with only a weak dependence on hardness. Thus this general CCA links the alloy metrics with both properties and phase information and is especially useful to workers who distinguish between HE alloys and multi-principal element alloys based on whether the system is a solid solution.

Methods

Data analytics. We employ here a multiple regression analysis and its generalization, a supervised learning strategy known as CCA, and then use the output from this analysis to construct a fitness function as an input to a GA. The aim of the CCA is to identify those linear combinations of predictor variables that are maximally correlated with the outcome variables. This is accomplished by constructing from products of blocks of the correlation (or covariance) matrix an operator whose eigenvalues (λ_i , where i runs from 1 to the number of variates pairs, m) capture the degree of correlation between combinations of predictor and outcome variables and whose eigenvectors determined the relative weights of the variables^{20,22,48}. The correlation coefficient associated with a given variate i is $\sqrt{\lambda_i}$ and the associated variability is $\lambda_i / \sum_{j=1}^{mv} \lambda_j$. In this context, a GA is a fitness function optimizer that begins with a random distribution of chromosomes (i.e., alloys) represented as fixed-length bit strings and, via evolutionary processes, produces successive generations that are better trial solutions to the problem (i.e., having good mechanical properties), as determined from the fitness function^{53–55}. Some tuning is required with respect to the frequency of these evolutionary processes, such as recombination and mutation, to achieve a well-functioning algorithm.

Microstructural and microchemical characterization. Microstructural characterization was conducted using SEM. Secondary electron imaging and X-ray EDS were performed on a FEI Scios Dual Beam FIB/SEM operated at an accelerating voltage of 20 kV and beam current of 13 nA. Secondary electron images were collected using an Everhart-Thornley detector. X-ray EDS maps were collected using an EDAX Octane Elite detector and exported via the EDAX Team software.

Hardness testing. Vicker's indentation hardness testing was conducted on each as-cast alloy using a LEICO LM 248AT model hardness tester. Each indent was performed on a polished surface, and a 100 g load was applied for 10 s. At least 20 indents were conducted for each alloy using a 2×10 array, and the error was calculated to reflect 95% certainty using the Student's t distribution.

Data availability

The authors will make available, upon request, the data used in the applications described in this work. It is understood that the data provided will not be for commercial use.

Code availability

The authors will make available, upon request, the code used in the applications described in this work. It is understood that the code will not be for commercial use.

Received: 12 November 2018 Accepted: 13 May 2019

Published online: 13 June 2019

References

- Tsai, M. H. & Yeh, J. W. Effect of stacking fault energy on the creep behavior of Ni-base superalloy. *Mater. Des.* **64**, 316–323 (2014).
- Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mat.* **122**, 448–511 (2017).
- Pickering, E. J. & Jones, N. G. High-entropy alloys: a critical assessment of their founding principles and future prospects. *Int. Mater. Rev.* **61**, 183–202 (2016).
- Gao, M. C., Yeh, J.-W., Liaw, P. K. & Zhang, Y. *High-Entropy Alloys: Fundamentals and Applications* (Springer, New York, NY, 2016).
- Murty, B. S., Yeh, J. W. & Ranganathan, S. *High-Entropy Alloys* (Butterworth-Heinemann, Oxford, 2014).
- Yeh, J. W., Chen, Y. L., Lin, S. J. & Chen, S. K. High-entropy alloys - a new era of exploitation. *Mater. Sci. Forum* **560**, 1–9 (2007).
- Hsu, C. Y. et al. On the superior hot hardness and softening resistance of AlCoCr_xFeMo_{0.5}Ni high-entropy alloys. *Mater. Sci. Eng. A* **528**, 3581–3588 (2011).
- Senkov, O. N., Wilks, G. B., Scott, J. M. & Miracle, D. B. Mechanical properties of Nb₂₅Mo₂₅Ta₂₅W₂₅ and V₂₀Nb₂₀Mo₂₀Ta₂₀W₂₀ refractory high entropy alloys. *Intermetallics* **19**, 698–706 (2011).
- Schuh, B. et al. Mechanical properties, microstructure and thermal stability of a nanocrystalline CoCrFeMnNi high-entropy alloy after severe plastic deformation. *Acta Mater.* **96**, 258–268 (2015).
- Tian, F. et al. Empirical design of single phase high-entropy alloys with high hardness. *Intermetallics* **58**, 1–6 (2015).
- Kozak, R., Sologubenko, A. & Steurer, W. Single-phase high-entropy alloys - an overview. *Z. Krist.* **230**, 55–68 (2015).
- Ma, S. G. & Zhang, Y. Effect of Nb addition on the microstructure and properties of AlCoCrFeNi high-entropy alloy. *Mater. Sci. Eng. A* **532**, 480–486 (2012).
- Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
- Zhang, Y. et al. Microstructures and properties of high-entropy alloys. *Prog. Mater. Sci.* **61**, 1–93 (2014).
- Singh, P. et al. Design of high-strength refractory complex solid-solution alloys. *npj Comput. Mater.* **4**, 16 (2018).
- Sharma, A. et al. Cuckoo searching property specific optimal compositions of multicomponent alloys by molecular simulations. *Scr. Mater.* **130**, 292–296 (2017).
- Troparevsky, M. C., Morris, J. R., Kent, P. R. C., Lupini, A. R. & Stocks, G. M. Criteria for predicting the formation of single-phase high-entropy alloys. *Phys. Rev. X* **5**, 011041 (2015).
- Senkov, O. N., Miller, J. D., Miracle, D. B. & Woodward, C. Accelerated exploration of multi-principal element alloys with solid solution phases. *Nat. Commun.* **6**, 6529 (2015).
- Sarker, P. et al. High-entropy high-hardness metal carbides discovered by entropy descriptors. *Nat. Commun.* **9**, 4980 (2018).
- Knapp, T. R. Canonical correlation analysis: a general parametric significance-testing system. *Psychol. Bull.* **85**, 410–416 (1978).
- Lawrence, A., Rickman, J. M., Harmer, M. P. & Rollett, A. D. Parsing abnormal grain growth. *Acta Mater.* **103**, 681–687 (2016).
- Rickman, J. M., Wang, Y., Rollett, A. D., Harmer, M. P. & Compson, C. Data analytics using canonical correlation analysis and Monte Carlo simulation. *npj Comput. Mater.* **3**, 26 (2017).
- Yang, X. & Zhang, Y. Prediction of high-entropy stabilized solid-solution in multi-component alloys. *Mater. Chem. Phys.* **132**, 233–238 (2012).
- Yeh, J.-W. et al. Formation of simple crystal structures in Cu-Co-Ni-Cr-Al-Fe-Ti-V alloys with multiprincipal metallic elements. *Metall. Mater. Trans. A* **35A**, 2533–2536 (2004).
- Cantor, B., Chang, I., Knight, P. & Vincent, A. Microstructural development in equiatomic multicomponent alloys. *Mat. Sci. Eng. A* **357–377**, 213–218 (2004).
- Tung, C.-C. et al. On the elemental effect of AlCoCrCuFeNi high-entropy alloy system. *Mater. Lett.* **61**, 1–5 (2007).
- Senkov, O. N., Wilks, G. B., Miracle, D. B., Chuang, C. P. & Liaw, P. K. Refractory high-entropy alloys. *Intermetallics* **18**, 1758–1765 (2010).
- Chen, M.-R. et al. Effect of vanadium addition on the microstructure, hardness and wear resistance of Al_{0.5}CoCrCuFeNi high-entropy alloy. *Metall. Mater. Trans. A* **37**, 1363–1369 (2006).
- Chou, H.-P., Chang, Y.-S., Chen, S.-K. & Yeh, J.-W. Microstructure, thermophysical and electrical properties of Al_xCoCrFeNi (0 ≤ x ≤ 2) high-entropy alloys. *Mat. Sci. Eng. B* **163**, 184–189 (2009).
- Senkov, O. N., Wilks, G. B., Scott, J. M. & Miracle, D. B. Mechanical properties of Nb₂₅Mo₂₅Ta₂₅W₂₅ and V₂₀Nb₂₀Mo₂₀Ta₂₀W₂₀ refractory high entropy alloys. *Intermetallics* **19**, 698–706 (2011).
- Guo, S., Ng, C. & Liu, C.-T. Anomalous solidification microstructures in Co-free Al_xCrCuFeNi₂ high-entropy alloys. *J. Alloy Compd.* **557**, 77–81 (2013).
- Kao, Y.-F., Chen, T.-J., Chen, S.-K. & Yeh, J.-W. Microstructure and mechanical property of as-cast, homogenized, and deformed Al_xCoCrFeNi (0 ≤ x ≤ 2) high-entropy alloys. *J. Alloy Compd.* **448**, 57–64 (2009).
- Li, C., Li, J. C., Zhao, M. & Jiang, Q. Effect of alloying elements on microstructure and properties of multiprincipal elements high-entropy alloys. *J. Alloy Compd.* **475**, 752–757 (2009).
- Senkov, O. N., Senkova, S. V., Woodward, C. & Miracle, D. B. Low-density, refractory multi-principal element alloys of the Cr-Nb-Ti-V-Zr system: microstructure and phase analysis. *Acta Mater.* **61**, 1545–1557 (2013).
- Zhou, Y., Jin, X., Du, X. Y., Zhang, L. & Li, B. S. Comparison of the structure and properties of equiatomic and non-equiatomic multicomponent alloys. *Mater. Sci. Technol.* **34**, 988–991 (2018).
- Takeuchi, A. & Inoue, A. Classification of bulk metallic glasses by atomic size difference, heat of mixing, period of constituent elements and its application to characterization of the main alloying element. *Mater. Trans. JIM* **46**, 2817–2829 (2005).
- Inselberg, A. The plane with parallel coordinates. *Vis. Comput.* **1**, 6991 (1985).
- Inselberg, A. *Parallel Coordinates: VISUAL Multidimensional Geometry and its Applications* (Springer, Berlin, 2009).
- Rickman, J. M. Data analytics and parallel-coordinate materials property charts. *npj Comp. Mater.* **3**, 206 (2017).
- Chen, R. Composition design of high entropy alloys using the valence electron concentration to balance strength and ductility. *Acta Mater.* **144**, 129–137 (2018).
- Mooney, D. & Swift, R. *A Course in Mathematical Modeling* (Mathematical Association of America, Washington, DC, 1999).
- Konak, A., Coit, D. W. & Smith, A. E. Multi-objective optimization using genetic algorithms: a tutorial. *Reliab. Eng. Syst. Safe* **91**, 992–1007 (2006).
- Pugh, S. F. XCII. Relations between the elastic moduli and the plastic properties of polycrystalline pure metals. *Lond. Edinb. Dubl. Philos. Mag.* **45**, 823–843 (1954).
- Lim, H.-S. et al. Evaluation of surface mechanical properties and grindability of binary Ti alloys containing 5 wt % Al, Cr, Sn, and V. *Metals* **7**, 487 (2017).
- Suss, R. et al. 18 Carat yellow gold alloys with increased hardness. *Gold Bull.* **37**, 3–4 (2004).
- Li, Y., Li, C., Du, Z. & Guo, C. Thermodynamics optimization of the Ni-Si-W ternary system. *CALPHAD* **43**, 112–123 (2013).
- Senkov, O. N., Miracle, D. B., Chaput, K. J. & Couzinie, J. P. Development and exploration of refractory high entropy alloys - review. *J. Mater. Res.* **33**, 3092–3128 (2018).
- Jobson, J. D. *Applied Multivariate Data Analysis*, Vol. II (Springer-Verlag, New York, 1992).
- Gittins, R. *Canonical Analysis: A Review with Applications in Ecology* (Springer-Verlag, New York, 1985).
- Akaho, S. A kernel method for canonical correlation analysis. In International Meeting of Psychometric Society (Osaka, 2001).
- Balakrishnan, S., Puniyani, K. & Lafferty, J. Sparse additive functional and kernel CCA. In *Proc. 29th International Conference on Machine Learning* (eds. Langford, J. and Pineau, J.) pp. 911–918 (Omnipress, Edinburgh, Scotland, 2012).
- Thorndike, R. M. *Correlational Procedures for Research* (Gardner Press, Springer, New York, NY, 1978).
- Whitley, D. A genetic algorithm tutorial. *Stat. Comput.* **4**, 65–85 (Springer, New York, NY, 1994).
- Michalewicz, Z. *Genetic Algorithms+Data Structures=Evolution Programs*, 3rd edn (Springer, New York, 1996).
- Sivanandam, S. N. & Deepa, S. N. *Introduction to Genetic Algorithms* (Springer-Verlag, Berlin, 2008).

Acknowledgements

The authors acknowledge support from the Office of Naval Research under Grant Numbers N00014-18-1-2181 and N00014-18-1-2484. They also acknowledge support from the Nano/Human Interface initiative at Lehigh University.

Author contributions

J.M.R. is the primary author (and guarantor) of this work. He developed the formalism and did most of the analysis. H.M.C., M.P.H. and G.B. contributed substantially to the discussion of the correlation of hardness with alloy metrics. J.S., C.J.M. and A.R. performed the mechanical testing and the microstructural analysis and also contributed to interpretation of the results.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-10533-1>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019