scientific reports

Check for updates

OPEN

Identification of core carcinogenic elements based on the age-standardized mortality rate of lung cancer in Xuanwei Formation coal in China

Zailin Chen[1,2,3]✉, Xianfeng Cheng[1,2], Xingyu Wang[3], Shijun Ni[3], Qiulian Yu[1,2] & Junchun Hu[4]

In this study, the core carcinogenic elements in Xuanwei Formation coal were identified. Thirty-one samples were collected based on the age-standardized mortality rate (ASMR) of lung cancer; Si, V, Cr, Co, Ni, As, Mo, Cd, Sb, Pb, and rare earth elements and yttrium (REYs) were analyzed and compared; multivariate statistical analyses (CA, PCA, and FDA) were performed; and comprehensive identification was carried out by combining multivariate statistical analyses with toxicology and mineralogy. The final results indicated that (1) the high-concentration Si, Ni, V, Cr, Co, and Cd in coal may have some potential carcinogenic risk. (2) The concentrations of Cr, Ni, As, Mo, Cd, and Pb meet the zoning characteristics of the ASMR, while the Si concentration is not completely consistent. (3) The REY distribution pattern in Longtan Formation coal is lower than that in Xuanwei Formation coal, indicating that the materials of these elements in coal are different. (5) The heatmap divides the sampling sites into two clusters and subtypes in accordance with carcinogenic zoning based on the ASMR. (6) PC1, PC2, and PC3 explain 62.629% of the total variance, identifying Co, Ni, As, Cd, Mo, Cr, and V. (7) Fisher discriminant analysis identifies Ni, Si, Cd, As, and Co based on the discriminant function. (8) Comprehensive identification reveals that Ni is the primary carcinogenic element, followed by Co, Cd, and Si in combination with toxicology. (9) The paragenesis of Si (nanoquartz), Ni, Co, and Cd is an interesting finding. In other words, carcinogenic elements Ni, Co, Cd, and Si and their paragenetic properties should receive more attention.

Lung cancer is one of the most common malignant tumors; it originates in the lung bronchial mucosa or glands and severely endangers human health and life[1–3]. Many countries have reported that the incidence rates and mortality from lung cancer have increased significantly in recent years, with higher incidences in males than females and higher incidences in urban areas than in rural areas. The age-standardized mortality rate (ASMR) values of lung cancer in 2018 were $27.1/10^5$ for men and $11.2/10^5$ for women, with an average of $19.2/10^5$[4]. The ASMRs of lung cancer in China were $27.91/10^5$, $40.32/10^5$ for men, $16.08/10^5$ for women, $30.33/10^5$ in cities, and $26.66/10^5$ in rural areas from 2004 to 2018[5].

Northeast Yunnan (Xuanwei-Fuyuan) in China is an area that has high morbidity and mortality related to lung cancer[6], and this area shows distinctive characteristics: (1) it is a typical rural area; (2) female incidence rates and mortality are high; (3) the pathological features are characterized by high proportions of lung adenocarcinoma[7–9] and squamous cell carcinoma[10–13]; (4) the rates are four to eight times (4–8)[14] the national average, and the incidence rate remains high[6,8]; and (5) areas with high morbidity from lung cancer are highly consistent with the development and application range of Xuanwei Formation coal[15–17].

The burning of bituminous coal in Northeast Yunnan has been associated with the region's high reported incidence of lung cancer[18–20]. However, the specific cause is still a mystery[21–27]. Currently, the lung cancer incidence has shown no substantial relationship with smoking[28] in this region. The polycyclic aromatic hydrocarbons (PAHs) present in coal[19–22,29] may be a cause, but the putative dose–response curves cannot fully explain the high

[1]Engineering Center of Yunnan Education Department for Health Geological Survey and Evaluation, Kunming 652501, China. [2]Yunnan Land and Resources Vocational College, Kunming 652501, China. [3]College of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China. [4]Coal Geology Prospecting Institute of Yunnan Province, Kunming 650218, China. ✉email: 327634884@qq.com

morbidity from lung cancer[30–32]; in addition, poor geographical reproducibility is a major problem[33]. Nanoquartz particles[33,34] may also be a cause, but some studies have shown that silica is not the main cause[19,24,35]. Therefore, these controversial nanoquartz particles are worth discussing again. In addition, the ASMR of lung cancer in Xuanwei-Fuyuan has geographical differences. It is urgent to explore the geographical differences of carcinogenic elements in coal, which is also the innovation of this paper.

Furthermore, according to current research results, high levels of potentially toxic elements exist in Northeast Yunnan in some coal mines[36–39] and in the environment[17,27,40–44]; specifically, mainly Mn, Ti, Ni, V, Cr, Co, Cu, Sr, Zn, As, Mo, Cd, Pb, Cs, and Sb are present (Ni, V, Cr, Co, As, Cd, Mo, Pb, and Sb are carcinogenic[15,45–48], but due to spatial complexity, it is not clear which elements play decisive roles[49]). Rare earth elements and yttrium (REYs) (La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Y, Ho, Er, Tm, Yb, Lu, and Y) are a group of elements that have similar geochemical properties[50,51], including La, Ce, Pr, Nd, Sm, and Eu (light REYs (LREYs)) and Gd, Tb, Dy, Y, Ho, Er, Tm, Yb, and Lu (heavy REYs (HREYs))[52,53], which are often used to identify rock characteristics and trace chemical processes[54]. To better assess carcinogenic elements, the behavior of REYs in coal must be understood[55].

Multivariate statistical analyses[56,57], including correlation analysis (CA)[58], principal component analysis (PCA)[59,60], and Fisher discriminant analysis (FDA)[61–63], have been widely used as tools to identify sources and determine the main influential factors from compositional data. For example, Jin et al.[56] applied PCA and CA to identify potential sources in soil and dust at children's playgrounds in Beijing; Ranjbar et al.[60] utilized CA to establish the relationship between variables and PCA to reduce the dataset to several determining factors; He et al.[62] used PCA and FDA to model, assess, and classify ecological and environmental quality and the impacts of coal mining; Bi et al.[61] utilized the FDA model for mine water inrush sources. Hence, these methods can be used to reveal core carcinogenic elements in combination with toxicology.

In this study, the concentration and mineralogical characteristics of Si, Ni, V, Cr, Co, As, Mo, Cd, Pb, Sb, and REYs in coal were obtained to achieve the following objectives: (1) to understand the general characteristics of carcinogenic elements; (2) to understand the geographical differences in carcinogenic elements and REYs; (3) to identify the core carcinogenic elements; and (4) to fully explore carcinogenic information.

Methodology

Study area

Xuanwei city is located in the Wumeng Mountains in northeastern Yunnan Province, China. The geographical coordinates are 103° 35′ 30″ to 104° 40′ 50″ N and 25° 53′ 30″ to 26° 44′ 50″ E, with a total area of 6069.88 km^2 and a population of 1.53×10^6 (at the end of 2015). The production activities are mainly agricultural. Xuanwei city is one of the main coal production bases in Yunnan Province and has many small coal mines. The reported and predicted coal resources are 3.85×10^9 tons, and the raw coal output is 2.7×10^6 tons/year[16].

Fuyuan County is located in the Wumeng Mountains in northeastern Yunnan Province. The geographical coordinates are 25° to 25°58' N and 103°58' to 104°49' E, with a total area of 3251 km^2 and a population of 0.83×10^6 (at the end of 2017). The economy is dominated by agricultural production (the agricultural population accounts for 92.93%), the industrial foundation is weak, and coal resources are abundant. The geological reserves of coal are 14.102×10^9 tons, the reported reserves are 6.457×10^9 tons, and the reported reserves of anthracite are 3.88×10^9 tons. The raw coal output is more than 5×10^6 tons/year[16].

Recent research shows that higher morbidity and mortality related to lung cancer exist in the entire coal-producing (burning) area of northeast Yunnan (Xuanwei-Fuyuan), and the problem remains serious and complex[6,41,64,65]. Hence, the geographical distribution of the ASMRs of lung cancer (Fig. 1) was plotted based on the latest data[6,8,66]. In this study, villages and towns were divided into five categories (normal, low, medium, high, and ultrahigh); among them, normal areas had no development or use of coal mines. The ASMRs of lung cancer in most villages and towns (Table S13) (II, III, and IV) were more than two times higher than that in China ($27.91/10^5$)[5], and those in some areas were more than four to eight times the national rate, with distinct geographical differences. Furthermore, the ASMR of lung cancer in zone I was low and could be used as excellent data for comparison. Hence, zones I, II, III, and IV were the focus of this paper.

Sampling and analysis

Sample collection

Thirty-one (31) coal samples were collected based on the ASMR of lung cancer (Table S14) to explore the carcinogenic information carried in the coal samples[15]; information for other samples was compiled from the literature (some samples were from discontinued coal mines) (Tables S1, S2). Among them, the Xiaohebian and Bailongshan coals belong to the Longtan Formation, which is used for comparison. The coals were collected in sample bags, transported to the laboratory, ground, passed through a 200-mesh sieve, and prepared for use.

Analysis and quality control

Major elements. SiO_2 was measured by an APL ADVANTXP + X instrument (X-ray fluorescence spectroscopy, 200 mesh)[15,52], and then SiO_2 (%) was changed to Si (mg/kg). The steps were as follows: (1) the organic matter was removed and then analyzed (1.0000 g of samples were thrown into a platinum crucible (5% Au + 95% Pt), whereupon the sample was placed in a muffle furnace (650 °C with the temperature raised for 1 h)); (2) the glass sample (dry sample) was made with a mixed flux (dehydrating agents NH_4NO_3 and LiBr and melting agent $Li_2B_4O_7 + LiBO_2 + LiF$); (3) finally, the glass sample was measured by X-ray fluorescence (XRF). The test accuracy (TA < 1%) and the detection limit (DL < 0.09%) met the requirements. The quality control experiment was considered satisfactory (RSD < 5%).

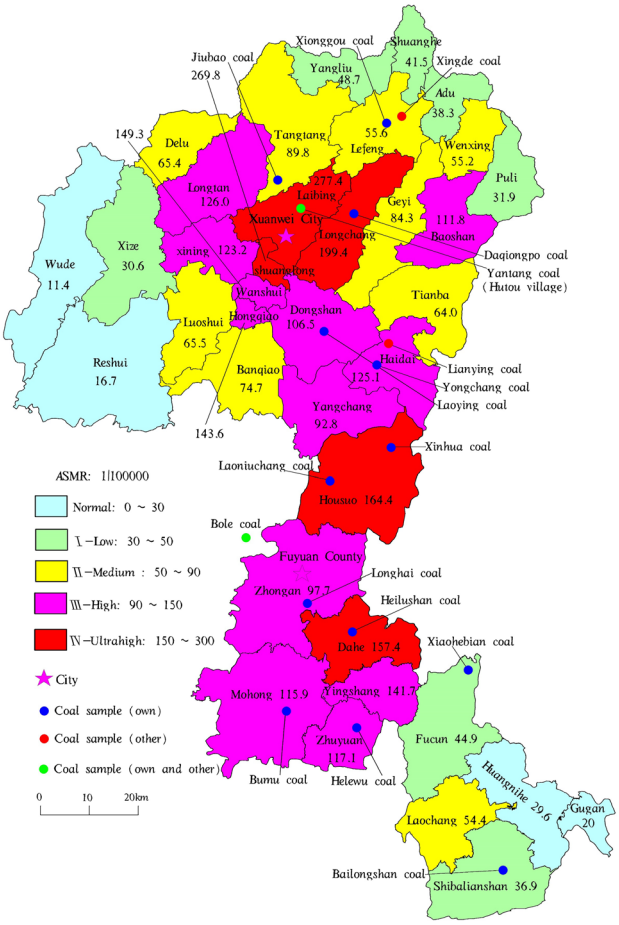

Figure 1. Geographical distribution of lung cancer ASMRs and coal sampling[6,8,66].

Trace elements. Trace elements were digested by a high-temperature closed digestion method[67], and the concentrations were determined by inductively coupled plasma–mass spectrometry (ICP—MS). The digestion steps, analytical steps, and quality control were based on our previous study[15,52]. The steps were as follows: (1) 0.10000 g of samples and 2 drops of ultrapure water were placed into a numbered 25 ml polytetrafluoroethylene liner (PTFEL) ((including 1 blank and 3 standard samples (GSS-1, GSS-3, and GSS-4)); (2) 2 ml HNO_3 and 1 ml HF were added into the above PTFEL and placed into the digestion tank (DK); (3) the DK was tightened and placed in the oven (100 °C for 1 h, and 180 °C for 29 h); (4) the PTFEL was taken out from DK and then placed on an electric heating plate (130 °C, steamed dry); (5) the residual HF was removed (1 ml HNO_3 was added and then steamed at 130 °C twice); (6) a standard sample (1 ml 1000 ng/ml Rh internal standard) was added to continue digestion (1 ml HNO_3 and 5 ml ultrapure water were added to the PTFEL, and the DK was tightened and placed in the oven (130 °C for 1 h, and 140 °C for 4 h)); (7) dilution and filtration were performed (1 ml HNO_3 was added, the volume was fixed to 10 ml with ultrapure water, and the sample was filtered with a 0.45 μM filtration membrane); (8) dilution was performed and determined by ICP—MS (1 ml of solution was pipetted into a centrifuge tube and diluted to 10 ml with ultrapure water). The quality control met the requirements (RSD < 5%).

Optical microscopy. The thin and polished sections of coal were processed to observe the occurrence state of silicon by a polarizing microscope[37,68] (Leica DM2700P and Jiangnan XPL-2) with reflected light and transmitted light[15], and the magnifications were adjusted as needed (50–800 times).

Scanning electron microscopy (SEM). The coal samples were processed and analyzed by a Hitachi S4800 (field emission scanning electron microscope)[52], and the accuracy met the analytical and testing requirements (component analysis < 0.01%). First, the test bench and ceramic scissors (a degreasing cotton ball dipped in 95% ethanol) were wiped, and the sample was placed on a test bench containing conductive tape. Second, Au was sprayed on the sample to improve its conductivity. Afterward, the test bench was transferred to the scanning electron microscope compartment, and the microscopic morphology of the particles was tested.

Electron probe microanalysis (EMPA). The quantitative analysis of in situ elements (Si, Ni, Co, and Cd) was completed by using an electron probe microanalyzer (JXA-8230 of JEOL)[69]. The analysis was completed using a

JEOL JXA-8230 instrument, the voltage and current were 15 kV and 50 nA, the peak analysis time of Si, Ni, Co, and Cd was 30 ms, and the background analysis time was 30 ms.

Identification of core carcinogenic elements

Correlation analysis (CA)

CA is used to comprehend the degree of resemblance and evaluate the relationships between carcinogenic elements and sources[70]. Furthermore, the latest research proposes heatmaps of correlation coefficients to exclude the dependency on variable variability[71,72]. In this study, the core carcinogenic elements were identified by analyzing the correlation between carcinogenic elements and REYs.

Principal component analysis (PCA)

Principal component analysis (PCA) aims to transform a set of potentially relevant variables into a set of linearly uncorrelated new variables through orthogonal transformation, which can retain the original information within the expressed information. Principal component analysis data processing plays a role in effectively eliminating correlations among high-dimensional data[73], reducing the data dimensions, and simplifying the data structure.

Fisher discriminant analysis (FDA)

The Fisher discriminant method (the description is in the supplementary data S1) was proposed in 1936[74], and it has no specific requirements for the overall distribution. In addition, it is a linear discriminant method[75,76] that can discriminate among a small number of samples. It projects high-dimensional data points[77] into low-dimensional space (one-dimensional straight line)[61] so that the data points can become denser, and this can overcome the "curse of dimensionality" caused by high dimensionality. The principle of projection is to separate the population[78] as much as possible, determine the discriminant analysis function according to the principle of maximum distance[79,80] between classes and minimum distance within classes, and then classify and distinguish the new samples.

Statistical analysis

The range of the elements, standard deviation (SD), median, mean, skewness, kurtosis, coefficient of variation (CV) and $\log_{10}(x+1)$ (Si: $\log_{10}(x/1000+1)$) functions were calculated via Microsoft Excel 2019[52]. As a note, log-transformation of each element is sufficient to put the data into normal distribution[60,72,81] and thus meets the requirements of data processing. Statistical tests, including CA, PCA, and FDA, were performed using SPSS 25. The heatmap was implemented using R version 4.2.0.

Ethical approval

All authors have read, understood, and complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Results

Elemental concentrations

Carcinogenic element concentrations

The concentrations of the investigated elements are listed in Tables S3 and S4, and the descriptive statistics for each in the Xuanwei Formation coal are shown in Table 1. The numerical value of each carcinogenic element exhibits a wide range. The mean concentrations of carcinogenic elements are 108,035 ± 40,748 mg/kg for Si, 101.45 ± 68.021 mg/kg for V, 35.81 ± 21.54 mg/kg for Cr, 23.29 ± 7.01 mg/kg for Co, 31.48 ± 11.70 mg/kg for Ni, 3.31 ± 4.86 mg/kg for As, 2.06 ± 1.42 mg/kg for Mo, 0.86 ± 0.75 mg/kg for Cd, 0.62 ± 0.63 mg/kg for Sb, and 15.13 ± 7.39 mg/kg for Pb.

The concentrations of Si, V, Cr, Ni, Co, As, Mo, Cd, Pb, and Sb are approximately 2.73, 2.89, 2.33, 3.29, 2.30, 0.87, 0.67, 3.44, 0.74, and 1.00 times the corresponding coal concentrations in China[37], and the concentrations of Si, Ni, Cr, Co, Cd, Sb, and Pb are higher than those in Longtan Formation coal (I-(low)), indicating that Si,

	Si	V	Cr	Co	Ni	As	Mo	Cd	Sb	Pb
Max	195,067	328.33	94.74	46.60	69.75	30.00	6.67	2.90	3.39	44.79
Min	40,413	15.12	9.00	8.45	13.78	0.34	0.27	0.05	0.06	2.99
Median	114,380	85.14	32.42	24.00	29.55	2.51	1.78	0.62	0.46	14.42
Mean	108,035	101.45	35.81	23.29	31.48	3.31	2.06	0.86	0.62	15.13
SD	40,748	68.02	21.54	7.01	11.70	4.86	1.42	0.75	0.63	7.39
Kurtosis	− 0.40	2.34	0.94	2.62	3.14	26.70	2.66	1.29	9.68	6.71
CV (%)	0.38	67.05	60.14	30.10	37.18	146.96	69.15	87.33	101.99	48.85
Skewness	0.09	1.37	1.09	0.62	1.51	4.89	1.53	1.44	2.64	1.85
Chinese coal	39,527	35.10	15.40	7.08	13.70	3.79	3.08	0.25	0.84	15.10
I-(low)	56,560	110.12	26.41	14.79	15.36	4.30	2.51	0.31	0.53	10.30

Table 1. Statistical results of the carcinogenic element concentrations of Xuanwei Formation coal (mg/kg). *Max* maximum; mean, *SD* standard deviation, *CV (%)* coefficient of variance.

Cr, Co, Ni, and Cd may have some potential carcinogenic risk. Therefore, Si, Cr, Co, Ni, and Cd are given more attention in the following discussion.

Carcinogenic element comparison

According to a comparison of zones, the concentrations of Cr, Ni, As, Mo, Cd, and Pb in the coal of zone IV are higher than those in zones II and III (Table S15), indicating that the carcinogenic elements in coal are different.

The SiO_2 content of the Xuanwei Formation coal is more than twice the average value of Chinese coal and more than twice that of the adjacent Longtan Formation coal. It is mainly closely related to quartz, which may lead to a high content of quartz particles in the local environment and indoor air, increasing the risk of local residents being exposed to quartz particles and causing pneumoconiosis. However, it is not completely consistent with the zoning characteristics of the ASMR, suggesting that there may be other collaborating factors. The concentrations of Cr, Ni, and Cd not only exceed those of Chinese coal and Longtan Formation coal but also meet the zoning characteristics of the ASMR. These may be the root cause of the difference in the ASMR.

REY comparison

The REY geochemical distribution model can directly reflect the differences in coal. In this study, the REY distribution curve was plotted based on the upper continental crust (UCC)[82,83]. The REY distribution patterns in zones I, II, III, and IV are characterized by LREY enrichment, weak negative Ce anomalies, and weak negative Y anomalies[63] (Fig. S1). The REY distribution pattern in zone I (Longtan Formation coal) is lower than those in zones II, III, and IV (Xuanwei Formation coal), and there are also significant differences within the Xuanwei Formation coal, indicating that the materials of these elements in coal are different. Previous studies have revealed that the source distance and weathering intensity of the Emeishan basalt are factors[15,84–86] that constrain the differences in the concentrations of REYs and carcinogenic elements. La, Ce, Pr, Nd, Eu, Gd, Tb, Dy, and REYs in the coal of zone IV are significantly lower than those in zones II and III, indicating that these elements (Table S11) can become environmental geochemical indicators for studying coal toxicity.

The geochemical parameters of REYs can reflect their degree of enrichment and material sources (Table S12). The degree of HREY fractionation ($(Gd/Yb)_N$) is highly consistent with the ASMR zoning of lung cancer (the smaller the fractionation degree is, the higher the ASMR) based on the UCC. Moreover, the smaller the LREY/HREY and REY are, the higher the ASMR, indicating that these parameters can also be used as important indicators for predicting the carcinogenic elements present in Xuanwei Formation coal.

Core carcinogenic element identification in the Xuanwei Formation coal

Core carcinogenic element identification (CA)

Correlation analysis (including cluster analysis) was performed to determine the relationship between 25 elements (Si, Ni, V, Cr, Co, As, Mo, Cd, Sb, Pb, and REYs) and sampling sites (Fig. 1, Tables S18, S19, S20, S21) through heatmapping (Fig. 2). The right vertical dendrogram presents the clustering of the sampling sites (ASMRs of lung cancer zones). The horizontal dendrogram symbolizes the clustering of carcinogenic elements and REYs according to their similarities (the clustering basis for rows and columns that was chosen was "manhattan", and the clustering method that was chosen was "mcquitty").

The vertical tree graph on the right side of the correlation heatmap displays the clustering of sampling points, while the horizontal tree graph displays the clustering of elements. Overall, the carcinogenic elements in the study area can be divided into two clusters. Cluster A22 consists of Cd, Sb, As, and Mo, with a negative correlation with a high ASMR of lung cancer; among them, Cd and Sb (A2211) is a subtype indicating their similar geochemical behavior, probably with sulfur (S) compounds, while another subtype, As and Mo (A222), represents hydrothermal influence[87]. Cluster A1 consists of Si, V, Pb, Cr, Ni, and Co, with a positive correlation with a high ASMR of lung cancer; subtype A11 consists of Si, V, and Ce, indicating that the influence of sea water on coal seams during coalification due to Ce anomalies is controlled by the seawater content[88–90]; and subtype A12 consists of Pb, Cr, La, Nd, Ni, Co, and Y, indicating that carcinogenic elements inside come from the weathering of the Emeishan basalt[91–93] during coalification. This phenomenon also exists objectively in other coal areas worldwide[94–99]. The heatmap divides sampling sites into two clusters and several subtypes in accordance with carcinogenic zoning based on the ASMR, indicating the effectiveness of ASMR partitioning in this paper. In addition, they are elements with significant concentration centers in Xuanwei and Fuyuan[100].

Hence, Si, V, Cr, Co, and Ni are identified; Pb has a weak correlation; and abnormalities in Y and Ce[68,101] can be important indicators for predicting carcinogenic elements in coal.

Core carcinogenic element identification (PCA)

To verify the above conclusions, a PCA of 10 carcinogenic elements in coal was performed. All factors were obtained with eigenvalues > 1[60] and then rotated using the varimax method in SPSS 25 (Kaiser normalization), and finally, the rotation converged in 5 iterations[102]. The PCA results revealed five factors (82.633% of the total variance (TV)), and PC5 was used as a reference only for its eigenvalues < 1. The first three principal components (PCs) explain 27.444% (PC1), 18.664% (PC2), and 16.521% (PC3) of the total variance in carcinogenic element concentrations.

PCA reduced the dataset to[60,103] major factors to explore the source of carcinogenic elements detected in coal. Moreover, the varimax rotation method[104], Kaiser–Meyer–Olkin (KMO)[105], and Bartlett's sphericity test[106] were used. In addition, the dataset was standardized and transformed using the $\log_{10}(x+1)$-scale before PCA. The principal components (PC1, PC2, and PC3), loadings of variables (LV), eigenvalues (EV), and their respective variances (RV) are displayed in Table S16. In our study, three PCs were extracted, accounting for 62.629% of the total variance, revealing the main carcinogenic elements in Xuwanwei Formation coal. The factor loadings

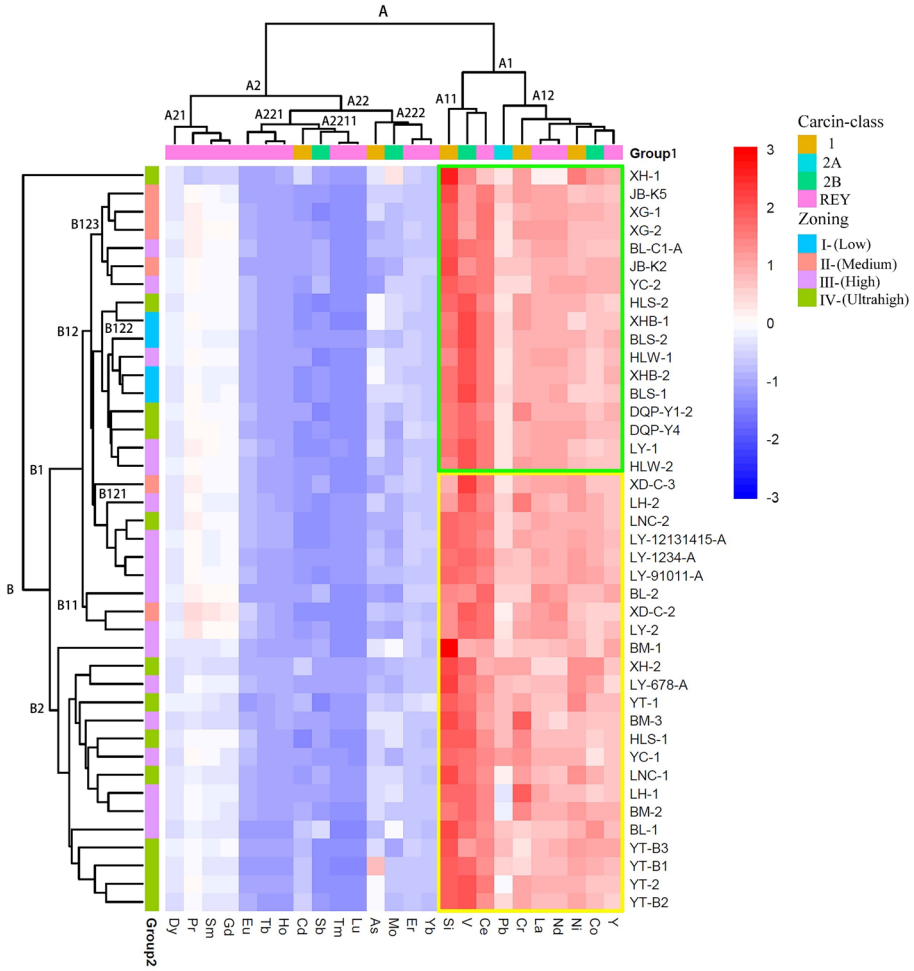

Figure 2. Heatmap of the correlation between sampling sites and studied elements.

were divided into "strong", "medium", and "weak" in terms of the absolute loading values of > 0.75, 0.75–0.50, and 0.50–0.30, respectively.

PC1 consists of Ni and Co (Table S16), explains 27.444% of the TV and 2.744% of the EV, and has strong positive loadings of Co (0.893) and Ni (0.840). Considering their high concentrations and siderophile features, Ni and Co probably originated from the weathering of the Emeishan basalt[91–93,107] during coalification, which is highly consistent with the CA results.

PC2 consists of As, Cd, and Mo (Table S16) and explains 18.664% of the TV and 1.866% of the EV. These carcinogenic elements are probably present due to the contribution of hydrothermal activity during coalification[36,97,108].

PC3 consists of Cr and V (Table S16), explains 16.521% of the TV and 1.652% of the EV, and has strong positive loadings of Cr (0.862) and V (0.847). These carcinogenic elements are siderophile elements and originated from the weathering of the Emeishan basalt and the influence of sea water during coalification, which is highly consistent with CA.

Among them, PC1, PC2, and PC3 explain 62.629% of the total variance. Hence, Co, Ni, As, Cd, Mo, Cr, and V are identified.

Core carcinogenic element identification (FDA)

The Fisher discriminant function was calculated for four groups of samples (I, II, III, and IV). Table S5 shows that the significance probability of Co, Ni, As, Cd, and Si is less than 0.05 (rejecting the original hypothesis), indicating that the Co, Ni, As, Cd, and Si included in the discriminant function play a role in determining the correct classification. Therefore, Co, Ni, As, Cd, and Si were selected for Fisher discriminant analysis.

The significance test results of the discriminant function show that (Table S6) the Wilks' lambda value of the function from 1 to 3 is 0.320, the chi-squared value is 40.460, the degree of freedom (Df) is 15, and the significance probability is 0.000. The discriminant function has reference significance.

Table S7 shows that the value of the box's M is 114.604, which meets the calculation requirements (> 0.05). Consequently, all kinds of covariance matrices were considered equal and met the requirements for the test results (Yang et al. 2017). At the same time, the significance probability[109] of the F test is less than 0.05, indicating that the error probability of the discriminant function is small.

The variance percentage can be used as an interpretation of the discriminant equation. The variance percentage of discriminant function 1 is 83.3% (Table S8), so this function could discriminate most samples. The structure matrix represents the intragroup correlations between the discriminant variable and the standardized canonical discriminant function (Table S9). According to the absolute size of the intrafunction correlation, it consists of Ni (0.722), Si (0.515), Cd (0.416), As (0.107), and Co (0.471). Based on the absolute size of the intrafunction correlation, Ni shows the largest correlation, followed by Co. The function and combination characteristics serve as important bases for identifying the carcinogenic elements Ni, Si, Cd, As, and Co.

According to the coefficients and constant terms of discriminant function 1, function 2, and function 3 (Table S10), two groups of functions and comprehensive results were obtained (Fig. 3), which clearly distinguishes each group:

$$\text{Function 1:}\quad Y1 = -0.821\text{Co} + 6.705\text{Ni} - 0.774\text{As} + 2.443\text{Cd} + 3.596\text{Si} - 15.938 \tag{1}$$

$$\text{Function 2}:\quad Y2 = 0.396\text{Co} - 0.974\text{Ni} + 4.350\text{As} - 0.132\text{Cd} - 0.420\text{Si} - 0.62 \tag{2}$$

$$\text{Function 3:}\ Y3 = 9.851\text{Co} - 4.862\text{Ni} + 1.103\text{As} - 3.756\text{Cd} + 0.404\text{Si} - 6.726 \tag{3}$$

The comprehensive result map of the Fisher discriminant function of carcinogenic elements established in this study can be used to easily and quickly distinguish the carcinogenic characteristics of the coal in the study area and provide a quantitative method for deepening the understanding of the environmental geochemical characteristics of coal. However, a single mathematical statistical discrimination procedure cannot be used as a sufficient condition for establishing coal carcinogenic characteristics, and other indicators need to be used to confirm the findings.

Discussion

The CA, PCA, and FDA had high discrimination accuracy. However, if the toxicological characteristics of carcinogenic elements and the mineralogical characteristics of silicon are not considered, the effect of some carcinogenic elements can be exaggerated, and the results of multivariate statistical analyses are distorted. Therefore, toxicology is also the focus of this paper.

Toxicological characteristics of carcinogenic elements

Si, Ni, V, Cr, Co, As, Mo, Cd, Pb, and Sb were filtered out based on the findings of substantial previous scientific studies[17,27,33,41–44,110]. Nevertheless, the correlations between these carcinogenic elements and lung cancer as well as their toxicological characteristics[45,47,48] need to be considered (Table S17).

Some important information can be obtained from Table S17: (1) crystalline silica, Cr^{6+}, Ni, As, and Cd are classified as class I carcinogens; and (2) the toxicological characteristics of crystalline silica, Co, Ni, Cr^{6+}, and Cd are connected with lung cancer (Table S17). However, the concentration of Cr^{6+} in local coal[111] and the environment[102] seems not to be the main cause. Hence, Si, Co, Ni, As, and Cd should be given more attention.

Mineralogical characteristics of silicon

Amorphous silica is commonly present in nature and has little or no chronic adverse pulmonary effects[112–116], such as in sedimentary rocks[117,118], hot spring systems[119,120], and soil[121]. Crystalline silica is a class I carcinogen[112].

The occurrence state of silicon in the coal of the Xuanwei Formation is mainly authigenic quartz[36,37] (Fig. 4), except for terrigenous clastic quartz and pyroclastic quartz. In addition, respirable silica refers to silica particles

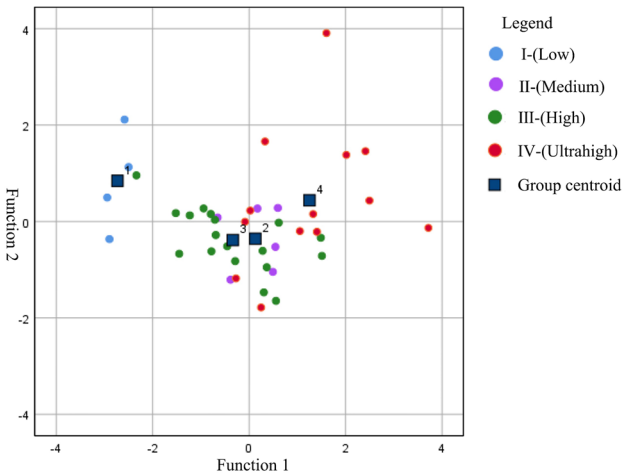

Figure 3. Comprehensive results of the discriminant function.

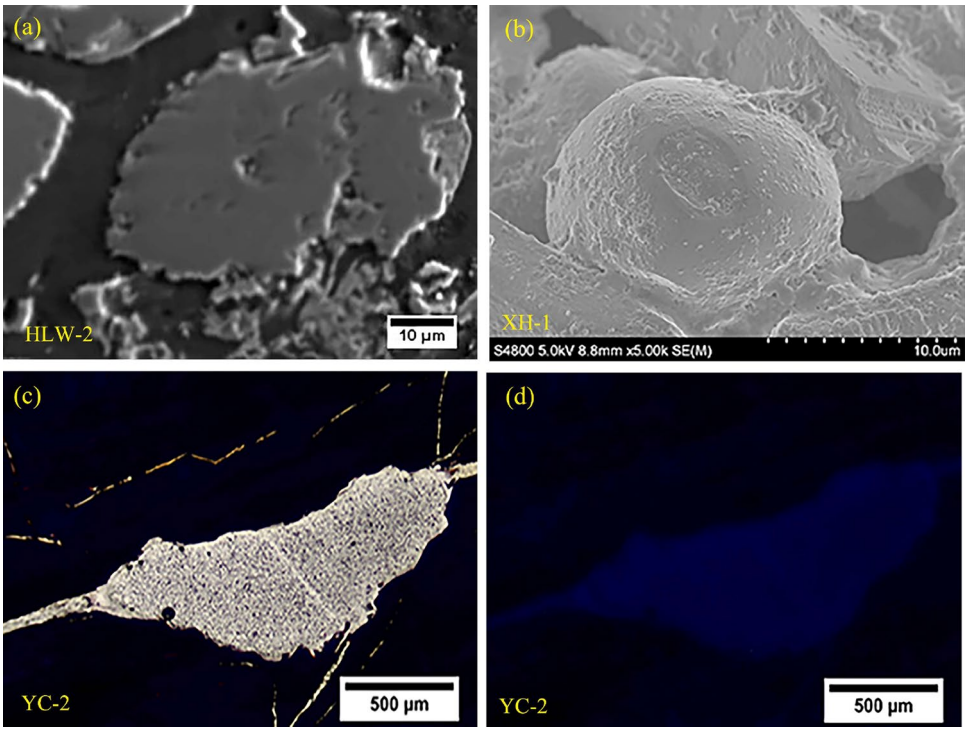

Figure 4. Occurrence state of authigenic silica. (**a**) Backscattered electron (BSE) image; (**b**) electron micrographs from silica nanoparticles; (**c**) plane-polarized light, amorphous silica; (**d**) cross-polarized light, complete extinction of amorphous silica.

less than 10 μm in diameter[122], widely existing in the Xuanwei Formation coal (Fig. 4b). Therefore, we should pay attention to the crystalline quartz in coal and the conditions under which it is converted into respirable quartz[122].

Comprehensive identification

We reduced Si, Ni, V, Cr, Co, As, Mo, Cd, Pb, and Sb to Si, Co, Ni, As, and Cd according to the toxicological characteristics. Furthermore, the concentration, CA, PCA, and FDA methods identified the carcinogenic elements. However, the core carcinogenic elements were still unclear. Therefore, it was necessary to combine concentration, CA, PCA, and FDA with toxicology and mineralogy to achieve comprehensive identification because each method has its limitations.

Table 2 shows that Ni was the primary core carcinogenic element, followed by Co, Cd, and Si, which was consistent with the conclusions of our previous research[102]. Hence, accurate and effective identification of core carcinogenic elements (Ni, Co, Cd, and Si) can support the development and use of local coal in the future.

Interestingly, Si (nanoquartz), Ni, Co, and Cd are highly paragenetic (Fig. 5), consistent with the EDX spectrum of silica particles in air[24], which seems to enhance the carcinogenic activity of local coal. More studies are needed to better comprehend the role of carcinogenic elements and crystalline quartz paragenesis in carcinogenesis in this area.

Discussion on the causes of cancer risk

Quartz is a class I carcinogen, and its carcinogenic mechanism is that it leads to pulmonary fibrosis (pneumoconiosis, silicosis)[123–125] after inhalation into the lungs and induces cancer. Pneumoconiosis related to coal miners is common worldwide[6,126], and the Xuanwei Fuyuan area is no exception[109,127]. However, the incidence of lung cancer in Xuanwei and Fuyuan is not only among coal miners but also among those who use local coal

Method	Si	V	Cr	Co	Ni	As	Mo	Cd	Sb	Pb
Concentration	√		√	√	√			√		
IARC classification	√			√	√			√		
CA	√	√	√	√	√					
PCA		√	√	√	√	√	√	√		
FDA	√			√	√	√		√		

Table 2. Comprehensive identification table.

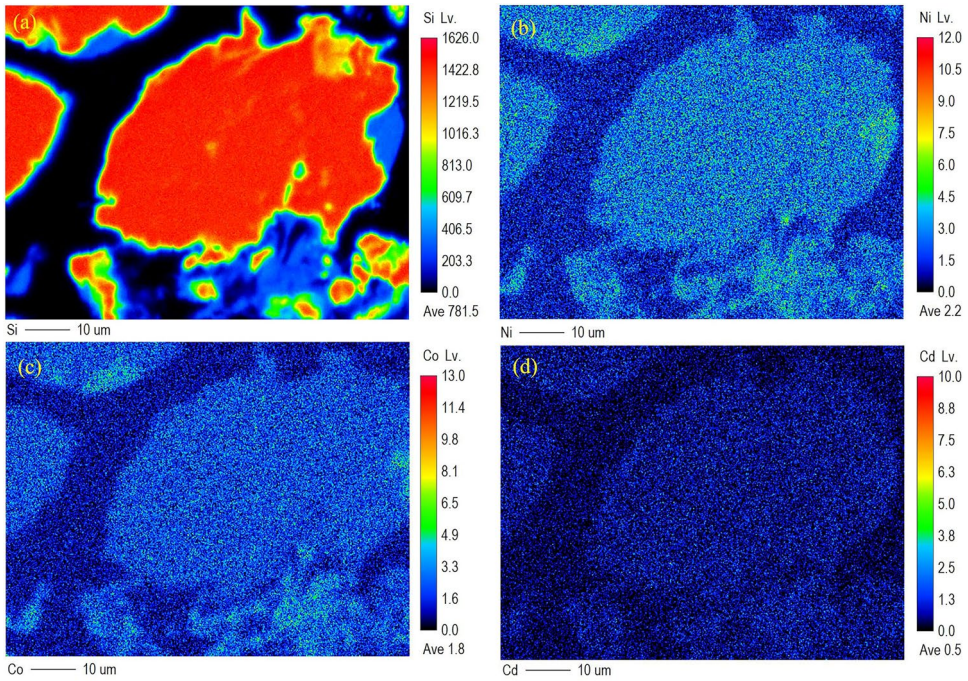

Figure 5. X-ray maps for the distribution of Si, Ni, Co, and Cd in coal (HLW-2) (Lv represents the signal strength level).

for heating or cooking, indicating a synergistic effect of factors other than quartz. Li[128] believes that residents in the Xuanwei Fuyuan area have been exposed to quartz particles and that the inflammatory response is the key factor leading to lung injury[88]. Quartz particles easily penetrate the bronchial epithelial cell membrane and enter the cell but cannot enter the nucleus. At present, most scientists believe that local nanoquartz particles are synergistic carcinogens and synergists[19,24,35].

Exposure to nickel in the environment is related to human lung cancer and nasal cancer[104,129], and research has shown that its triggering of DNA damage (cell cycle imbalance) is an important carcinogenic mechanism. In vitro and in vivo experiments show that nickel can produce reactive oxygen species by binding DNA, mediating DNA damage and inhibiting DNA repair[130,131]. Coal mines that have high Ni contents include the Candiota and Colchester low-grade coal mine in the Shenbei lignite in China and Brazil, respectively, and the Kosovo lignite mine in Serbia[37,96,98,99,132]. Notably, the lung cancer incidence in these areas is also worrisome[94,95,108]. The correlation between nickel and lung cancer is evident worldwide[133–135]. The carcinogenic mechanism of Ni in coal remains to be explored. Morbidity from lung cancer has been linked with exposure to high contents of nickel compounds[136,137], including sulfidic, oxidic, water-soluble, and insoluble metallic nickel[138–141]. Among them, water-soluble nickel has a greater oral absorption and is the most important risk factor. Furthermore, studies have shown that a dose-related association[139] of cumulative exposure to water-soluble nickel compounds could lead to lung cancer[139,142]. At present, a high nickel content and high nickel water solubility have been found in the PM_{10} of Hutou village (Fig. 1) from coal development and use, which probably supports this view. Epithelial-mesenchymal transitional lung injury is the mechanism of nickel-induced lung diseases[137,143]. Coincidentally, the pathological feature of lung cancer in Fuyuan and Xuanwei is the damage of epithelial cells, with a high proportion of squamous cell carcinoma[144] and adenocarcinoma[10,12]. Nickel appears to be the most dangerous carcinogen in the local area.

Cobalt exposure in the environment is a recognized cause of human interstitial lung disease (which can later develop into diffuse pulmonary fibrosis)[145], generally occurring in hard metal and bonded diamond tool industries. There have also been reports of interstitial lung disease related to coal mines, but further work is needed to determine whether it is related to cobalt in coal[146,147].

Cadmium is a known human lung carcinogen, and its main carcinogenic mechanism is damage to lung epithelial cells[148]. In vitro studies have revealed possible toxicokinetic pathways, such as increased oxidative stress, changes in transcription factor activity and inhibition of DNA repair[149]. Particles with aerodynamic diameters less than 10 μm can be used as carriers of cadmium, which affects lung health. The increasing trend of lung cancer mortality related to Cd in coal has also been reported worldwide[150]. The pathological characteristics of lung cancer in Xuanwei Fuyuan are a high proportion of adenocarcinoma and squamous cell carcinoma[13]. From a pathological point of view, the lung cancer is due to the abnormal proliferation of adenoid epithelium and squamous epithelium, which may be related to Cd in atmospheric particles.

Therefore, this paper infers that the synergetic carcinogenic mechanism of nanoquartz particles (Si), Ni, Co and Cd in Xuanwei Formation coal is that "nanoquartz particles and Co damage lung tissue cells (inflammatory

reaction), and Ni and Cd damage lung nuclei (mediate DNA damage and inhibit DNA repair)". This conclusion still needs further verification through medical experiments in the future.

Conclusions

The current research results provide information on the characteristics of core carcinogenic elements in coal from the Xuanwei Formation. The results demonstrated the following:

(1) The concentrations of Si, Ni, V, Cr, Co, and Cd were higher than those in Chinese coal and Longtan Formation coal; (2) the heatmap of correlation identified Si, V, Cr, Co, and Ni; PCA identified Co, Ni, As, Cd, Mo, Cr, and V; FDA identified Ni, Si, Cd, As, and Co; (3) comprehensive identification revealed that Ni was the primary carcinogenic element, followed by Co, Cd, and Si in combination with toxicology; and (4) the paragenesis of Si (nanoquartz), Ni, Co, and Cd in coal may increase the possibility of carcinogenesis.

Data availability

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Received: 22 July 2023; Accepted: 14 December 2023
Published online: 02 January 2024

References

1. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**(21), 2129–2139 (2004).
2. Rudin, C. M., Brambilla, E., Faivre-Finn, C. & Sage, J. Small-cell lung cancer. *Nat. Rev Dis. Primers.* **7**(1), 1–20 (2021).
3. Sharma, R. Mapping of global, regional and national incidence, mortality and mortality-to incidence ratio of lung cancer in 2020 and 2050. *Int. J. Clin. Oncol.* **27**, 665–675 (2022).
4. Bray, F. *et al.* Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **1**, 1 (2018).
5. Liu, D. H., Jiang, D. M., Zhou, X. M., *et al.* Comparison of lung cancer mortality between rural and urban areas in the mainland of China from 2004 to 2018. *Shanghai J. Prev. Med.* **33**(10), 893–898 (2021) (**In Chinese**).
6. Cha, S., Chen, Y., & Shao, Y. Yunnan death cause monitoring data set (2018) (Yunan university press, 2021).
7. Devarakonda, S., Morgensztern, D. & Govindan, R. Genomic alterations in lung adenocarcinoma. *Lancet Oncol.* **16**(7), e342–e351 (2015).
8. Li, J. *et al.* Geostatistical analysis of village-level lung cancer mortality from 2010 to 2019 in Fuyuan County Yunnan Province. *China Cancer* **30**(10), 8 (2021).
9. Xu, J. Y. *et al.* Integrative proteomic characterization of human lung adenocarcinoma. *Cell* **182**(1), 245-261.e17 (2020).
10. Chen, Y. *et al.* Distinct epithelial growth factor receptor mutation profile in non-small-cell lung cancer patients from the Xuanwei area of China. *Mol. Clin Oncol.* **4**(5), 749–755 (2016).
11. Wang, N. J. *et al.* Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **108**(43), 17761–17766 (2011).
12. Wu, H. *et al.* Gene expression profiling of lung adenocarcinoma in Xuanwei China. *Eur. J. Cancer Prev.* **25**(6), 508–517 (2016).
13. Zhang, H. *et al.* Genomic evidence of lung carcinogenesis associated with coal smoke in Xuanwei area, China. *Natl. Sci. Rev.* **8**(12), 152 (2021).
14. Chen, G. *et al.* The mortality patterns of lung cancer between 1990 and 2013 in Xuanwei China. *Lung Cancer.* **90**(2), 155–160 (2015).
15. Chen, Z., Shi, Z., Ni, S. & Hu, J. Environmental geochemical characteristics of Xuanwei Formation coal and their controlling geological factors, with comments on the relationship with lung cancer incidence and distribution. *Geochem. Explor. Environ. Anal.* **23**(2), 1–15 (2023).
16. Luo, J., Yuan, X., Lin, Y. C., *et al. Accumulation law and resource potential of coal and coalbed methane in Yunnan Province* (Geological Publishing House, 2015) (**in Chinese**).
17. Shao, L., Zhang, M., Feng, X., Jiao, J., Wang, W., & Zhou, Y. Study on toxicology of particulate matter from coal combustion and geological origin of lung cancer epidemic in Xuanwei area. *Coal Geol. Explor.* 1–14 (2023) (**In Chinese**).
18. Barone-Adesi, F. *et al.* Risk of lung cancer associated with domestic use of coal in Xuanwei, China: Retrospective cohort study. *BMJ* **345**, e5414–e5414 (2012).
19. Vermeulen, R. *et al.* Constituents of household air pollution and risk of lung cancer among never-smoking women in Xuanwei and Fuyuan China. *Environ. Health Perspect.* **127**(9), 097001 (2019).
20. Xiao, Y., Shao, Y., Yu, X. & Zhou, G. The epidemic status and risk factors of lung cancer in Xuanwei City, Yunnan Province. *China. Front. Med.* **6**(4), 388–394 (2012).
21. Downward, G. S. *et al.* Heterogeneity in coal composition and implications for lung cancer risk in Xuanwei and Fuyuan countis China. *Environ. Int.* **68**, 94–104 (2014).
22. Downward, G. S. *et al.* Polycyclic aromatic hydrocarbon exposure in household air pollution from solid fuel combustion among the female population of Xuanwei and Fuyuan Counties China. *Environ. Sci. Technol.* **48**(24), 14632–14641 (2014).
23. Downward, G. S. *et al.* Outdoor, indoor, and personal black carbon exposure from cookstoves burning solid fuels. *Indoor Air* **26**(5), 784–795 (2015).
24. Downward, G. S. *et al.* Quartz in ash, and air in a high lung cancer incidence area in China. *Environ. Pollut.* **221**, 318–325 (2017).
25. Kim, C. *et al.* Smoky coal, tobacco smoking, and lung cancer risk in Xuanwei China. *Lung Cancer.* **84**(1), 31–35 (2014).
26. Yang, M. *et al.* Association of cancer screening and residing in a coal-polluted East Asian region with overall survival of lung cancer patients: A retrospective cohort study. *Sci. Rep.* **10**(1), 1 (2020).
27. Zhang, M. *et al.* Hemolysis of PM_{10} on RBCs in vitro: An indoor air study in a coal-burning lung cancer epidemic area. *Geosci. Front.* **13**(1), 101176 (2021).
28. Mumford, J. L. *et al.* Lung cancer and indoor air pollution in Xuan Wei China. *Science.* **235**, 217–220 (1987).
29. Chuang, J., Cao, S., Xian, Y., Harris, D. & Mumford, J. Chemical characterization of indoor air of homes from communes in Xuan Wei, China, with high lung cancer mortality rate. *Atmos. Environ. Part A. General Topics* **26**(12), 2193–2201 (1992).
30. Chapman, R. S. *et al.* The epidemiology of lung cancer in Xuan Wei, China: Current progress, issues, and research strategies. *Arch. Environ. Health.* **43**(2), 180–185 (1988).
31. Lan, Q., Chapman, R. S., Schreinemachers, D. M., Tian, L. & He, X. Household stove improvement and risk of lung cancer in Xuanwei. *China. J. Natl. Cancer I.* **94**(11), 826–835 (2002).

32. Large, D. J. *et al.* Silica-volatile interaction and the geological cause of the Xuan Wei Lung cancer epidemic. *Environ. Sci. Technol.* **43**(23), 9016–9021 (2009).
33. Tian, L. Coal Combustion Emissions and Lung Cancer in Xuan Wei, China. Ph.D. thesis, University of California, Berkeley (2005).
34. Fruijtier-Pölloth, C. The toxicological mode of action and the safety of synthetic amorphous silica—A nanostructured material. *Toxicology* **294**(2–3), 61–79 (2012).
35. He, J., Cai, Y., Lv, J. & Zhang, L. Primary investigation of quartz as a possible carcinogen in coals of Xuanwei and Fuyuan, high lung cancer incidence area in China. *Environ. Earth Sci.* **67**(6), 16791684 (2012).
36. Dai, S. *et al.* Mineralogical and compositional characteristics of Late Permian coals from an area of high lung cancer rate in Xuan Wei, Yunnan, China: Occurrence and origin of quartz and chamosite. *Int. J. Coal. Geol.* **76**(4), 318–327 (2008).
37. Dai, S. *et al.* Origin of minerals and elements in the Late Permian coals, tonsteins, and host rocks of the Xinde Mine, Xuanwei, eastern Yunnan China. *Int. J. Coal. Geol.* **121**, 53–78 (2014).
38. Li, X. Mineral matter Characteristic and sources of volcanic ash in the Late Permian coal-bearing strata from Xuanwei, eastern Yunnan. China University of mining and Technology (Beijing) (2015) (**In Chinese**).
39. Zheng, X. Mineral Matter in Lopingian Coals from Eastern Yunnan Province and Its Response to the Regional Geological Evolution. China University of mining and Technology (Beijing) (2018) (**In Chinese**).
40. Fan, J. S. Physicochemical characteristics of indoor PM_{10} and $PM_{2.5}$ in Xuanwei lung cancer area. China University of mining and Technology (Beijing) (2013) (**In Chinese**).
41. Feng, X. *et al.* Particle-induced oxidative damage by indoor size-segregated particulate matter from coal-burning homes in the Xuanwei lung cancer epidemic area, Yunnan Province China. *Chemosphere* **256**, 127058 (2020).
42. Lü, S. *et al.* Single particle aerosol mass spectrometry of coal combustion particles associated with high lung cancer rates in Xuanwei and Fuyuan China. *Chemosphere.* **186**, 278–286 (2017).
43. Shao, L. *et al.* Particle-induced oxidative damage of indoor PM_{10} from coal burning homes in the lung cancer area of Xuan Wei China. *Atmos. Environ.* **77**, 959–967 (2013).
44. Zhou, L. Study on the Physicochemistry and Toxicology of Indoor PM_{10} in the High Incidence Areas of Xuanwei Lung Cancer. China University of mining and Technology (Beijing) (2010) (**in Chinese**).
45. Ghaffari, H. R. *et al.* The concentration of potentially hazardous elements (PHEs) in drinking water and non-carcinogenic risk assessment: A case study in Bandar Abbas Iran. *Environ. Res.* **201**, 111567 (2021).
46. IARC (International Agency for Research on Cancer website). Agents classified by the IARC Monographs, Volumes 1–124 (2019).
47. Nordberg, G., Fowler, B., & Nordberg, M. *Handbook on the toxicology of metals* (Academic Press, 2015).
48. Saleh, H. N. *et al.* Carcinogenic and non-carcinogenic risk assessment of heavy metals in groundwater wells in Neyshabur Plain Iran. *Biol. Trace Elem. Res.* **190**, 251–261 (2019).
49. Papadimitriou, F. *Spatial complexity. Theory, mathematical methods and applications* (Springer, 2020).
50. Cheisson, T. & Schelter, E. J. Rare earth elements: Mendeleev's bane, modern marvels. *Science* **363**(6426), 489–493 (2019).
51. Laveuf, C. & Cornu, S. A review on the potentiality of Rare Earth Elements to trace pedogenetic processes. *Geoderma* **154**(1–2), 1–12 (2009).
52. Chen, Z., Shi, Z., Ni, S. & Cheng, L. Characteristics of soil pollution and element migration associated with the use of coal in Hutou Village, Yunnan Province China. *Ecol. Indic.* **139**, 108976 (2022).
53. Wang, Z., Shu, J., Wang, Z., Qin, X. & Wang, S. Geochemical behavior and fractionation characteristics of rare earth elements (REEs) in riverine water profiles and sentinel Clam (Corbicula fluminea) across watershed scales: Insights for REEs monitoring. *Sci. Total Environ.* **803**, 150090 (2022).
54. Haley, B. A., Klinkhammer, G. P. & McManus, J. Rare earth elements in pore waters of marine sediments. *Geochim. Cosmochim. Acta.* **68**(6), 1265–1279 (2004).
55. Desrosiers, M. *et al.* Toxicokinetics in rats and modeling to support the interpretation of biomonitoring data for rare-earth elements. *Environ. Int.* **155**, 106685 (2021).
56. Jin, Y. *et al.* Assessment of sources of heavy metals in soil and dust at children's playgrounds in Beijing using GIS and multivariate statistical analysis. *Environ. Int.* **124**, 320–328 (2019).
57. Kükrer, S. & Mutlu, E. Assessment of surface water quality using water quality index and multivariate statistical analyses in Saraydüzü Dam Lake Turkey. *Environ. Monit. Assess.* **191**(2), 1 (2019).
58. Kynčlová, P., Hron, K. & Filzmoser, P. Correlation between compositional parts based on symmetric balances. *Math. Geosci.* **49**, 777–796 (2017).
59. Islam, M. S., Hossain, M. B., Matin, A. & Islam Sarker, M. S. Assessment of heavy metal pollution, distribution and source apportionment in the sediment from Feni River estuary Bangladesh. *Chemosphere* **202**, 25–32 (2018).
60. Ranjbar, J. A., Riyahi, B. A., Shadmehri, T. A. & Jadot, C. Spatial distribution, ecological and health risk assessment of heavy metals in marine surface sediments and coastal seawaters of fringing coral reefs of the Persian Gulf Iran. *Chemosphere* **185**, 1090–1111 (2017).
61. Bi, Y. *et al.* Discriminant analysis of mine water inrush sources with multi-aquifer based on multivariate statistical analysis. *Environ. Earth. Sci.* **80**(4), 1 (2021).
62. He, H., Tian, C., Jin, G. & Han, K. Principal component analysis and Fisher discriminant analysis of environmental and ecological quality, and the impacts of coal mining in an environmentally sensitive area. *Environ. Monit. Assess.* **192**(4), 1 (2020).
63. Li, J. *et al.* Enrichment of Nb-Ta-Zr-W-Li in the Late carboniferous coals from the Weibei coalfield, Shaanxi North China. *Energies* **13**(18), 4818 (2020).
64. Liu, L., Liu, X., Ma, X., Ning, B. & Wan, X. Analysis of the associations of indoor air pollution and tobacco use with morbidity of lung cancer in Xuanwei China. *Sci. Total. Environ.* **135**, 232 (2020).
65. Wong, J. Y. Y. *et al.* Lung cancer risk by geologic coal deposits: A case-control study of female never-smokers from Xuanwei and Fuyuan China. *Int. J. Cancer.* **144**(12), 2918–2927 (2018).
66. Liu, X. *et al.* Epidemiological Features of lung cancer mortality between 1990 and 2016 in Xuanwei City Yuannan Province. *ACTA Acad. Med. Sin.* **41**(3), 1 (2019).
67. Zhang, Y. *et al.* Cd isotope fractionation during simulated and natural weathering. *Environ. Pollut.* **216**, 9–17 (2016).
68. Dai, S., Graham, I. T. & Ward, C. R. A review of anomalous rare earth elements and yttrium in coal. *Int. J. Coal Geol.* **159**, 82–95 (2016).
69. Rinaldi, R. & Llovet, X. Electron probe microanalysis: A review of the past, present, and future. *Microsc. Microanal.* **21**(05), 1053–1069 (2015).
70. Tokalioglu, S. Determination of trace elements in commonly consumed medicinal herbs by ICP-MS and multivariate analysis. *Food. Chem.* **134**(4), 2504–2508 (2012).
71. Lučić, M., Mikac, N., Bačić, N. & Vdović, N. Appraisal of geochemical composition and hydrodynamic sorting of the river suspended material: Application of time-integrated suspended sediment sampler in a medium-sized river (the Sava River catchment). *J. Hydrol.* **597**, 125768 (2021).
72. Reimann, C., Filzmoser, P., Hron, K., Kynčlová, P. & Garrett, R. G. A new method for correlation analysis of compositional (environmental) data—a worked example. *Sci. Total Environ.* **607–608**, 965–971 (2017).

73. Zhu, Z. B. & Song, Z. H. A novel fault diagnosis system using pattern classification on kernel FDA subspace. *Expert Syst. Appl.* **38**(6), 6895–6905 (2011).
74. Fisher, R. A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188 (1936).
75. Shin, H. An extension of Fisher's discriminant analysis for stochastic processes. *J. Multivar. Anal.* **99**(6), 1191–1216 (2008).
76. Volpi, M., Petropoulos, G. P. & Kanevski, M. Flooding extent cartography with Landsat TM imagery and regularized kernel Fisher's discriminant analysis. *Comput. Geosci.* **57**, 24–31 (2013).
77. Clarke, R. *et al.* The properties of high-dimensional data spaces: Implications for exploring gene and protein expression data. *Nat. Rev. Cancer* **8**(1), 37–49 (2008).
78. Huang, P. H. & Wang, X. Y. Piper-PCA-Fisher recognition model of water inrush source: A case study of the Jiaozuo mining area. *Geofuids* **1**, 1–10 (2018).
79. Elias de Almeida, V., Douglas de Sousa Fernandes, D., Henrique Gonçalves Dias Diniz, P., de Araújo Gomes, A., Véras, G., Kawakami Harrop Galvão, R., & Cesar Ugulino Araujo, M. Scores selection via Fisher's discriminant power in PCA-LDA to improve the classification of food data. *Food Chem.* **363**, 130296 (2021).
80. Li, B., Wu, Q. & Liu, Z. Identification of mine water inrush source based on PCA-FDA: Xiandewang coal mine case. *Geofluids* **2020**, 1–8 (2020).
81. Zuo, R., Wang, J., Xiong, Y. & Wang, Z. The processing methods of geochemical exploration data: Past, present, and future. *Appl. Geochem.* **132**, 105072 (2021).
82. Condie, K. C. Chemical composition and evolution of the upper continental crust: Contrasting results from surface samples and shales. *Chem. Geol.* **104**(1–4), 1–37 (1993).
83. Kamber, B. S., Greig, A. & Collerson, K. D. A new estimate for the composition of weathered young upper continental crust from alluvial sediments, Queensland Australia. *Geochim. Cosmochim. Acta.* **69**(4), 1041–1058 (2005).
84. Xiao, L. *et al.* Distinct mantle sources of low-Ti and high-Ti basalts from the western Emeishan large igneous province, SW China: Implications for plume-lithosphere interaction. *Earth. Planet. Sci. Lett.* **228**(3–4), 525–546 (2004).
85. Xu, Y., Chung, S.-L., Jahn, B. & Wu, G. Petrologic and geochemical constraints on the petrogenesis of Permian-Triassic Emeishan flood basalts in southwestern China. *Lithos* **58**(3–4), 145–168 (2001).
86. Xu, Y. G., He, B., Chung, S. L., Menzies, M. A. & Frey, F. A. Geologic, geochemical, and geophysical consequences of plume involvement in the Emeishan flood-basalt province. *Geology* **32**(10), 917 (2004).
87. Breuer, C. & Pichler, T. Arsenic in marine hydrothermal fluids. *Chem. Geol.* **348**, 2–14 (2013).
88. Pattan, J. N., Pearce, N. J. G. & Mislankar, P. G. Constraints in using Cerium-anomaly of bulk sediments as an indicator of paleo bottom water redox environment: A case study from the Central Indian Ocean Basin. *Chem. Geol.* **221**(3–4), 260–278 (2005).
89. Tostevin, R. *et al.* Effective use of cerium anomalies as a redox proxy in carbonate-dominated marine settings. *Chem. Geol.* **438**, 146–162 (2016).
90. Wilde, P., Quinby-Hunt, M. S. & Erdtmann, B. D. The whole-rock cerium anomaly: A potential indicator of eustatic sea-level changes in shales of the anoxic facies. *Sediment. Geol.* **101**(1–2), 43–53 (1996).
91. Ali, J. R., Fitton, J. G. & Herzberg, C. Emeishan large igneous province (SW China) and the mantle-plume up-doming hypothesis. *J. Geol. Soc.* **167**(5), 953–959 (2010).
92. Bai, Z. J., Zhong, H., Naldrett, A. J., Zhu, W. G. & Xu, G. W. Whole-Rock and Mineral Composition Constraints on the Genesis of the Giant Hongge Fe-Ti-V Oxide Deposit in the Emeishan Large Igneous Province Southwest China. *Econ. Geol.* **107**(3), 507–524 (2012).
93. Zhou, M. F. *et al.* Two stages of immiscible liquid separation in the formation of Panzhihua-type Fe-Ti-V oxide deposits. *SW China. Geosci. Front.* **4**(5), 481–502 (2013).
94. Hallal, A. & Gotlieb, S. Time trends in cancer mortality in Rio Grande do Sul, Brazil, 1979–1995. *Rev. bras. Epidemiol.* **4**(3), 168–177 (2001).
95. Ilic, M. & Ilic, I. Cancer mortality in Serbia, 1991–2015: An age-period-cohort and joinpoint regression analysis. *Cancer. Commun.* **38**, 10 (2018).
96. Kittner, N., Fadadu, R. P., Buckley, H. L., Schwarzman, M. R. & Kammen, D. M. Trace metal content of coal exacerbates air-pollution-related health risks: The case of lignite coal in Kosovo. *Environ. Sci. Technol.* **52**(4), 2359–2367 (2018).
97. Ren, D. Y., Zhao, F. H., Dai, S. F., Zhang, J. Y., & Luo, K. L. Trace element geochemistry of coal. science press, beijing (2006) (**In Chinese**).
98. Ruppert, L. *et al.* Origin and significance of high nickel and chromium concentrations in Pliocene lignite of the Kosovo Basin Serbia. *Int. J. Coal. Geol.* **29**(4), 235–258 (1996).
99. Solari, J. A., Fiedler, H. & Schneider, C. L. Modelling of the distribution of trace elements in coal. *Fuel* **68**(4), 536–539 (1989).
100. Xie, X. J., Cheng, Z. Z., & Zhang, L. S. Seventy-six Element Geochemical Atlas of Southwest China. Geological Publishing House (2008) (**In Chinese**).
101. Bau, M. *et al.* Discriminating between different genetic types of marine ferro-manganese crusts and nodules basedon rare earth elements and yttrium. *Chem. Geol.* **381**, 1–9 (2014).
102. Chen, Z. *et al.* Assessment of toxic elements in road dust from Hutou Village, China: Implications for the highest incidence of lung cancer. *Environ. Sci. Pollut. Res.* **28**, 1850–1865 (2021).
103. Wang, J., Liu, G., Liu, H. & Lam, P. K. S. Multivariate statistical evaluation of dissolved trace elements and a water quality assessment in the middle reaches of Huaihe River, Anhui China. *Sci. Total Environ.* **583**, 421–431 (2017).
104. Guo, H. *et al.* Nickel carcinogenesis mechanism: Cell cycle dysregulation. *Environ. Sci. Pollut. Res.* **28**, 4893–4901 (2021).
105. Goni, M. D. *et al.* Development and validation of knowledge, attitude and practice questionnaire for prevention of respiratory tract infections among Malaysian Hajj pilgrims. *BMC Public Health* **20**, 1–10 (2020).
106. Cüce, H., Kalipci, E., Ustaoğlu, F., Dereli, M. A. & Türkmen, A. Integrated spatial distribution and multivariate statistical analysis for assessment of ecotoxicological and health risks of sediment metal contamination, Ömerli Dam (Istanbul, Turkey). *Water Air Soil Pollut.* **233**(6), 199 (2022).
107. Dai, S. *et al.* Metalliferous coal deposits in East Asia (Primorye of Russia and South China): A review of geodynamic controls and styles of mineralization. *Gondwana. Res.* **29**(1), 60–82 (2016).
108. Dai, S. *et al.* Geochemistry of trace elements in Chinese coals: A review of abundances, genetic types, impacts on human health, and industrial utilization. *Int. J. Coal. Geol.* **94**, 3–21 (2012).
109. Hassainia, F., Petit, D. & Montplaisir, J. Significance probability mapping: The final touch int-statistic mapping. *Brain. Topogr.* **7**(1), 3–8 (1994).
110. Tian, L. *et al.* Nanoquartz in Late Permian C1 coal and the high incidence of female lung cancer in the Pearl River Origin area: A retrospective cohort study. *BMC Public Health* **8**, 398 (2008).
111. Chen, Y. Study on the Regularity of Leaching, Releasing and Transfer of Carcinogen Chromium of Coal and Its Products in Rich Coal of Yunnan Province. Kunming University of Science & Technology (Kunming) (2008) (**In Chinese**).
112. Croissant, J. G., Butler, K. S., Zink, J. I. & Brinker, C. J. Synthetic amorphous silica nanoparticles: Toxicity, biomedical and environmental implications. *Nat. Rev Mater.* **5**(12), 886–909 (2020).
113. Hoy, R. F. & Chambers, D. C. Silica-related diseases in the modern world. *Allergy* **75**(11), 2805–2817 (2020).
114. Hempt, C. *et al.* The impact of synthetic amorphous silica (E 551) on differentiated Caco-2 cells, a model for the human intestinal epithelium. *Toxicol. in Vitro.* **67**, 104903 (2020).

115. Leemann, A., Shi, Z. & Lindgård, J. Characterization of amorphous and crystalline ASR products formed in concrete aggregates. *Cem. Concr. Res.* **137**, 106190 (2020).
116. Zheng, Y. *et al.* Disordered hyperuniformity in two-dimensional amorphous silica. *Sci. Adv.* **6**(16), 826 (2020).
117. Liesegang, M. & Milke, R. Australian sedimentary opal-A and its associated minerals: Implications for natural silica sphere formation. *Am. Miner.* **99**(7), 1488–1499 (2014).
118. Yanchilina, A. G., Yam, R., Kolodny, Y. & Shemesh, A. From diatom opal-A $\delta^{18}O$ to chert $\delta^{18}O$ in deep sea sediments. *Geochim. Cosmochim. Acta* **268**, 368–382 (2020).
119. Jones, B. & Renaut, R. W. Microstructural changes accompanying the opal-A to opal-CT transition: New evidence from the siliceous sinters of Geysir, Haukadalur Iceland. *Sedimentology* **54**(4), 921–948 (2007).
120. Rodgers, K. A. *et al.* Silica phases in sinters and residues from geothermal fields of New Zealand. *Earth Sci. Rev.* **66**(1–2), 1–61 (2004).
121. Clarke, J. The occurrence and significance of biogenic opal in the regolith. *Earth Sci. Rev.* **60**(3–4), 175–194 (2003).
122. Yang, X. *et al.* Phase transformation of silica particles in coal and biomass combustion processes. *Environ. Pollut.* **292**, 118312 (2022).
123. Castranova, V. & Vallyathan, V. Silicosis and coal workers' pneumoconiosis. *Environ. Health Perspect.* **108**(suppl 4), 675–684 (2000).
124. Leung, C. C., Yu, I. T. S. & Chen, W. Silicosis. *The Lancet* **379**(9830), 2008–2018 (2012).
125. Pollard, K. M. Silica, silicosis, and autoimmunity. *Front. Immunol.* **7**, 1 (2016).
126. Hall, N. B., Blackley, D. J., Halldin, C. N. & Laney, A. S. Current review of pneumoconiosis among US coal miners. *Curr. Env. Hlth. Rep.* **6**, 137–147 (2019).
127. Gu, H. K. *et al.* Analysis on epidemiological characteristics of pneumoconiosis in Qujing City from 2006–2014. *Occup. Health.* **32**, 7 (2016).
128. Li, G. J. Relationship between high incidence of lung cancer among non-smoking women and naturally occurring silica in bituminous coal and the possible mechanism of carcinogenesis in Xuanwei, China (Kunming Medical University, Kunming, 2013) (**In Chinese**).
129. Bolek, E. C., Erden, A., Kulekci, C., Kalyoncu, U. & Karadag, O. Rare occupational cause of nasal septum perforation: Nickel exposure. *Int. J Occup. Med. Environ.* **30**(6), 963–967 (2017).
130. Latvala, S., Vare, D., Karlsson, H. L. & Elihn, K. In vitro genotoxicity of airborne Ni-NP in air-liquid interface. *J. Appl. Toxicol.* **37**(12), 1420–1427 (2017).
131. Scanlon, S. E., Scanlon, C. D., Hegan, D. C., Sulkowski, P. L. & Glazer, P. M. Nickel induces transcriptional down-regulation of DNA repair pathways in tumorigenic and non-tumorigenic lung cells. *Carcinogenesis* **38**(6), 627–637 (2017).
132. Ren, D., Xu, D. & Zhao, F. A preliminary study on the enrichment mechanism and occurrence of hazardous trace elements in the Tertiary lignite from the Shenbei coalfield China. *Int. J. Coal. Geol.* **57**(3–4), 187–196 (2004).
133. Lightfoot, N., Berriault, C. & Semenciw, R. Mortality and cancer incidence in a nickel cohort. *Occup. Med.* **60**(3), 211 (2010).
134. Qu, H. M. *et al.* Trend analysis of cancer mortality in the Jinchang Cohort, China, 2001–2010. *Biomed. Environ. Sci.* **28**(5), 6 (2015).
135. St-Jean, A. *et al.* Nickel and associated metals in New Caledonia: Exposure levels and their determinants. *Environ. Int.* **118**, 106–115 (2018).
136. Grimsrud, T. K., Berge, S. R., Haldorsen, T. & Andersen, A. Exposure to different forms of nickel and risk of lung cancer. *Am. J Epidemiol.* **156**(12), 1123–1132 (2002).
137. Lee, H. W., Jose, C. C. & Cuddapah, S. Epithelial-mesenchymal transition: Insights into nickel-induced lung diseases. *Semin. Cancer Biol.* **76**, 99–109 (2021).
138. Christiani, D. C. Ambient air pollution and lung cancer: Nature and nurture. *Am. J. Resp. Crit. Care.* **204**(7), 752–753 (2021).
139. Grimsrud, T. K. Exposure to different forms of nickel and risk of lung cancer. *Am. J Epidemiol.* **156**(12), 1123–1132 (2002).
140. Prueitt, R. L., Li, W., Chang, Y. C., Boffetta, P. & Goodman, J. E. Systematic review of the potential respiratory carcinogenicity of metallic nickel in humans. *Crit. Rev. Toxicol.* **50**(7), 605–639 (2020).
141. You, D. J., Lee, H. Y., Taylor-Just, A. J., Linder, K. E. & Bonner, J. C. Sex differences in the acute and subchronic lung inflammatory responses of mice to nickel nanoparticles. *Nanotoxicology* **14**(8), 1058–1081 (2020).
142. Oller, A. R. Respiratory carcinogenicity assessment of soluble nickel compounds. *Environ. Health Perspect.* **110**(suppl 5), 841–844 (2002).
143. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* **119**(6), 1420–1428 (2009).
144. Li, R. *et al.* The characteristics of lung cancer in Xuanwei County: A review of differentially expressed genes and noncoding RNAs on cell proliferation and migration. *Biomed. Pharmacother.* **119**, 109312 (2019).
145. Adams, T. N., Butt, Y. M., Batra, K. & Glazer, C. S. Cobalt related interstitial lung disease. *Resp. Med.* **129**, 91–97 (2017).
146. Beer, C. *et al.* A systematic review of occupational exposure to coal dust and the risk of interstitial lung diseases. *Eur. Clin. Respir. J.* **4**(1), 1264711 (2017).
147. Demedts, M. *et al.* Interstitial lung diseases: An epidemiological overview. *Eur. Respir. J.* **18**(32), 2s–16s (2001).
148. García-Esquinas, E. *et al.* Cadmium exposure and cancer mortality in a prospective cohort: The strong heart study. *Environ. Health Perspect.* **122**(4), 363–370 (2014).
149. Person, R. J. *et al.* Chronic cadmium exposure in vitro induces cancer cell characteristics in human lung cells. *Toxicol. Appl. Pharmacol.* **273**(2), 281–288 (2013).
150. Hendryx, M., O'Donnell, K. & Horn, K. Lung cancer mortality is elevated in coal-mining areas of Appalachia. *Lung Cancer* **62**(1), 1–7 (2008).

Acknowledgements

The authors would like to thank the China University of Mining and Technology-Beijing, Kunming Medical University, Utrecht University, etc., for its fruitful work in the research area, which has brought us inspiration. The authors are indebted to the reviewers for the very constructive comments that helped to improve the manuscript.

Author contributions

The study conception was led by Z.C. and S.N.; data collection and data analysis were performed by Z.C., X.W., Q.Y., and J.H.; the first draft of the manuscript was written by Z.C. and X.C. X.C. provided technical support for this study. All the authors have read and approved the final manuscript.

Funding

This study was supported by the Engineering Center of Yunnan Education Department for Health Geological Survey and Evaluation (9135009009), the Sichuan Natural Science Foundation Project (22NSFSC0191 and 22NSFSC3990), the Everest Scientific Research Program of the Chengdu University of Technology

(80000-2022ZF11419), and the Opening Fund of the Provincial Key Lab of Applied Nuclear Techniques in Geosciences (gnzds201905 and gnzds202201).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-49975-5.

Correspondence and requests for materials should be addressed to Z.C.