PLOS | ONE

RESEARCH ARTICLE

A Highly Efficient Gene Expression Programming (GEP) Model for Auxiliary Diagnosis of Small Cell Lung Cancer

Zhuang Yu[1☯‡*], Haijiao Lu[1☯‡], Hongzong Si[2], Shihai Liu[3], Xianchao Li[4], Caihong Gao[1], Lianhua Cui[5], Chuan Li[6], Xue Yang[1], Xiaojun Yao[7]

1 The Affiliated Hospital of Qingdao University, Department of Oncology, Qingdao, Shandong, P.R. China, **2** Institute for Computational Science and Engineering, Laboratory of New Fibrous Materials and Modern Textile, the Growing Base for State Key Laboratory, Department of Pharmacy, Qingdao University, Qingdao, Shandong, P.R. China, **3** The Affiliated Hospital of Qingdao University, The Central Laboratory, Qingdao, Shandong, P.R. China, **4** Department of Pharmacy, Qingdao University, Qingdao, Shandong, P.R. China, **5** Department of Public Health, Qingdao University Medical College, Qingdao, Shandong, P.R. China, **6** The Affiliated Hospital of Qingdao University, Department of Thoracic Surgery, Qingdao, Shandong, P.R. China, **7** Department of Chemistry, Lanzhou University, Lanzhou, Gansu, P.R. China

☯ These authors contributed equally to this work.
‡ These authors are co-first authors on this work.
* yuzhuang2002@163.com

Citation: Yu Z, Lu H, Si H, Liu S, Li X, Gao C, et al. (2015) A Highly Efficient Gene Expression Programming (GEP) Model for Auxiliary Diagnosis of Small Cell Lung Cancer. PLoS ONE 10(5): e0125517. doi:10.1371/journal.pone.0125517

Academic Editor: Lanjing Zhang, University Medical Center of Princeton/Rutgers Robert Wood Johnson Medical School, UNITED STATES

Received: May 20, 2014

Accepted: March 24, 2015

Published: May 21, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Jieping Wu foundation: 320.6750.13210 and Jieping Wu foundation: 320.6753.1219. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Background

Lung cancer is an important and common cancer that constitutes a major public health problem, but early detection of small cell lung cancer can significantly improve the survival rate of cancer patients. A number of serum biomarkers have been used in the diagnosis of lung cancers; however, they exhibit low sensitivity and specificity.

Methods

We used biochemical methods to measure blood levels of lactate dehydrogenase (LDH), C-reactive protein (CRP), Na^+, Cl^-, carcino-embryonic antigen (CEA), and neuron specific enolase (NSE) in 145 small cell lung cancer (SCLC) patients and 155 non-small cell lung cancer and 155 normal controls. A gene expression programming (GEP) model and Receiver Operating Characteristic (ROC) curves incorporating these biomarkers was developed for the auxiliary diagnosis of SCLC.

Results

After appropriate modification of the parameters, the GEP model was initially set up based on a training set of 115 SCLC patients and 125 normal controls for GEP model generation. Then the GEP was applied to the remaining 60 subjects (the test set) for model validation. GEP successfully discriminated 281 out of 300 cases, showing a correct classification rate for lung cancer patients of 93.75% (225/240) and 93.33% (56/60) for the training and test sets, respectively. Another GEP model incorporating four biomarkers, including CEA, NSE,

LDH, and CRP, exhibited slightly lower detection sensitivity than the GEP model, including six biomarkers. We repeat the models on artificial neural network (ANN), and our results showed that the accuracy of GEP models were higher than that in ANN. GEP model incorporating six serum biomarkers performed by NSCLC patients and normal controls showed low accuracy than SCLC patients and was enough to prove that the GEP model is suitable for the SCLC patients.

Conclusion

We have developed a GEP model with high sensitivity and specificity for the auxiliary diagnosis of SCLC. This GEP model has the potential for the wide use for detection of SCLC in less developed regions.

Introduction

Lung cancer is a major cause of cancer death worldwide, representing about 12.7% (1.6 million cases) of all new cancer cases each year and 18.2% (1.4 million deaths) of all cancer deaths[1]. It has a poor prognosis, with a 15% 5-year survival rate, and more than 75% of patients are diagnosed at late stages of the disease[2,3]. Small cell lung cancer (SCLC) is one of the major types of lung cancer, with the highest degree of malignancy. Current therapy methods, such as chemotherapy, radiotherapy, and surgery are very limited for the treatment of late stage SCLC. Although tremendous effort and progress have been made in the treatment of lung cancer, recent advances in early detection have led to small improvements in prognosis[4]. Therefore, an effective screening method for the early diagnosis of SCLC is critically important for increasing clinical diagnosis effectiveness and outcome of this disease.

Many different techniques have been used in the detection of lung cancers, including Chest Radiograph (x-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Sputum Cytology, and bronchoscopy[5]. In recent years, whole-body positron-emission tomography (PET) has emerged to simplify and improve the evaluation of patients with this type of tumor[6]. However, these techniques are invasive, expensive, and/or time-consuming. For example, bronchoscopy can cause damage to the bronchus and lung. In addition, these detection methods are not sufficiently sensitive and specific enough in most cases[7,8] and misdiagnosis of indolent tumors, due to the low specificity of these methods, may lead to unnecessary surgical treatments[9,10]. In order to avoid overtreatment of the disease, non-invasive blood tests have been widely used in clinical settings for screening of SCLC. Biomarkers are molecules in blood, other body fluids, or tissues that can be used to evaluate the normal and abnormal conditions of human beings. Biomarkers can complement or replace radiological examinations for the screening of cancers or routine clinical visits[11,12]. In lung cancer, biomarker evaluations have been conducted in serum, tissue, and sputum [12]. Several serum biomarkers, including the carcinoembryonic antigen (CEA), the cytokeratin 19 fragment (CYFRA 21–1), the tissue polypeptide antigen (TPA), the squamous cell carcinoma antigen (SCC), the cancer antigen 125 (CA-125), the cancer antigen 153 (CA-153), the pro-gastrin-releasing peptide (ProGRP), the cancer antigen 199 (CA-199), tumor-associated glycoprotein 72–3 (TAG-72.3) and neuron-specific-enolase (NSE), have shown usefulness for diagnosis of lung cancers[13][14][15]. Nevertheless, each of them has failed to demonstrate the requisite sensitivity and specificity as a diagnostic tool to warrant clinical development[8]. The combination of a number of biomarkers may improve the diagnostic efficiency of cancers[16]. However, the combined use of

tumor biomarkers is not widely used, especially in small hospitals and in less developed countries, because of the high cost of equipment and reagents. In this study, we have found the combination of economical efficiency and correlative serum such as LDH, CRP, Na^+, Cl^-, which can be obtained by common biochemical detection method and don't need exorbitant agentias or facilities. In a rural and impoverished area, using the approach, a fundamental serum test could warn people who are at higher risk of suffering from cancer and to do an indepth health examination such as CT, PET-CT and so on.

Therefore, new technology is urgently needed to find the association information between a large set of biomarkers and for the early detection of lung cancer. In recent years, with the development of science and technology, computer-aided design has become an auxiliary tool for the diagnosis of human cancers. Nowadays, machine learning methods, such as artificial neural networks (ANNs), decision trees, the naive bayesian (NB) algorithm, and support vector machines (SVM) have been utilized in the diagnosis and prognosis prediction of cancers[17]. For instance, ANNs of different EGFR microdeletion mutations have been used to improve the diagnosis efficiency of non-small cell lung cancer (NSCLC)[18]. The ANN model combined with six tumor biomarkers, including CEA, gastrin, NSE, sialic acid (SA), Cu/Zn, and Ca, was used to successfully differentiate lung cancer from benign lung disease, a normal control, and gastrointestinal cancers[19]. A previous study has shown that NB techniques are useful for diagnosis and to generate treatment recommendations and predict the 1-year-survival rate in lung cancer patients[20]. The combination of protein characteristics and attribute weighting models with a support vector machine (SVM) was used to discriminate SCLC and NSCLC[21]. These methods have led to the development of classifiers that are capable to discriminate between cancer and non-cancer samples. The ANNs, SVMs and NBs have been widely used for classification problems[17][20][22]. The ANNs have the ability to fulfill the statistical that contain linear, logistic and nonlinear regression, but it is hard for ANNs to understand the structure of algorithm ， due to that ANNs are a "black-box" technology and hence, they can hardly discover how to operate the classification. Otherwise, generous attributes cause overfitting easily [17]. Contrast to ANNs, in SVM the overfitting hardly occur, but the training is slow when inputing large number of data. The NB is very easy to discern but like ANN excessive attributes can misinform the classification[17][23]. Recently, a novel evolutionary algorithm called Gene Expression Programming (GEP) which is an automatic programming approach first introduced by Ferreira[24] was studied for auxiliary diagnosis of cancers. GEP has the advantages of flexibility and the power to explore the entire search space, which comes from the separation of genotype and phenotype and has the visualization data model. It is easy to implement and point out why GEP can not work via parameter adjustment [24][25][26]. One particular study has manifested the superior value of GEP in predicting the adverse events of radical hysterectomy in cervical cancer patients with an accuracy of 71.96% [27]. In our fundamental research, classification of lung tumors was made based on biomarkers (measured in 120 NSCLC and 60 SCLC patients) by setting up optimal biomarker joint models with GEP algorithm [28]. However, there is little relevant data regarding GEP applied to lung cancer so far.

In this study, we developed a prediction model using the GEP method to improve the diagnostic efficacy of SCLC. A number of biomarkers have previously been demonstrated to be useful for lung cancer diagnosis. Our GEP model suggested a novel multi-analysis of serum biomarkers for the early detection of SCLC.

Materials and Methods

Patients and controls

In total 430 cases, including 145 SCLC patients, 130 non-small cell lung cancer (NSCLC) patients and 155 non-cancer controls, were enrolled from the Affiliated Hospital of Qingdao University between July of 2006 and May of 2013. The diagnosis of 145 SCLC patients was based on biopsy and histopathology, and they were proven to be untreated primary lung cancers (Fig 1), the 130 NSCLC patients were diagnosed with primary tumor in stage I, II before surgery. Histological diagnosis of primary lung cancer was established according to the revised classification of lung tumors by the World Health Organization and the International Association for Lung Cancer Study[29].

The SCLC group included 94 male and 51 female patients, aged between 33 and 78 years old. The control group was composed of 155 non-cancer cases, which underwent examinations proving their health (86 males and 69 females). The NSCLC patients (69 males and 61 females) were included in the negative control to show the difference from SCLC, we selected 130 cases from 155 non-cancer cases as the healthy control. Research approval was obtained from the corresponding ethics committee and written informed consent was obtained from all participants. Samples and health information were labeled using unique identifiers to protect subject confidentiality (Tables 1 and 2).

Selection of six serum biomarkers

We selected six biomarkers that are closely related to lung cancers, especially to SCLC, and that have been widely used in the screening of SCLC. The indexes we chose have been incorporated

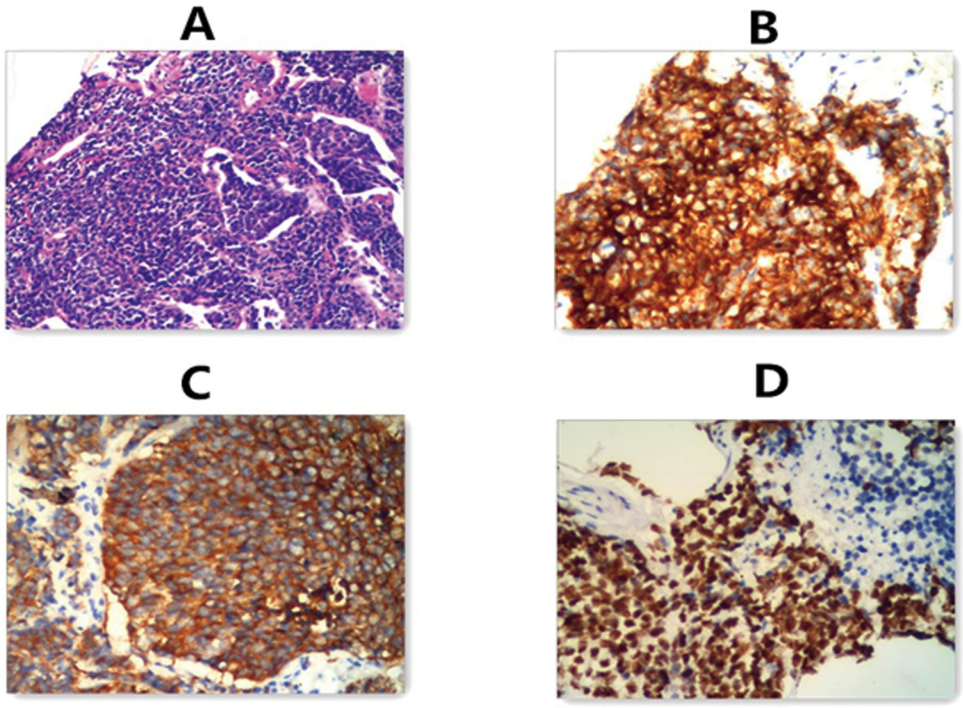

Fig 1. Histopathologic test of SCLC patients. A. hematoxylin-eosin staining of biopsy specimen slice. B. CD56(+) findings in immunohistochemical method. C. Syn (+) findings in immunohistochemical method. D. TTF-1(+) findings in immunohistochemical method

doi:10.1371/journal.pone.0125517.g001

Table 1. Demographic and clinical profiles of SCLC patients and controls included in this study ($\bar{x} \pm s$).

Demographic profile**	Controls (n = 155)	SCLC (n = 145)	*p-value**
Age (years)	56.23±8.72	57.92±9.46	*0.270*
Range (age)	29–81	33–78	-
Sex (F/M)	69/86	51/94	*0.099*
SCLC	-	145	-
Stage (L/E)	-	74/71	-
Smoking	86/69	92/53	*0.161*

*Statistics were conducted using the independent-Samples T Test and chi-square test.

**F = female and M = for male. L = limited stage and E = extensive stage.

doi:10.1371/journal.pone.0125517.t001

into the GEP model. Based on previous clinical examination, the serum levels of LDH and CRP in SCLC patients are significantly higher than in healthy controls, but the serum level of sodium and chloride are significantly lower than that in normal controls. The serum level of LDH, which is commonly elevated in neoplastic disorders, has been suggested as a powerful tumor marker for many years. Therefore, these markers have significant meaning in SCLC. For example, lung cancer patients, especially SCLC patients, the Syndrome of Inappropriate Anti Diuretic Hormone secretion (SIADH) is considered to be the leading cause for hyponatraemia and hypochloraemia and can be induced by comorbidity such as lung cancer. Also, the major osmotic active substances that in the extracellular fluid principal contain serum sodium and its accompanying anions chloride[30][31]. There are also numerous reports on the association between chronic inflammation and cancer[32]. CRP is a nonspecific acute-phase inflammatory response serum marker produced by hepatocytes under the regulation of interleukin (IL)-6 [33]. CEA and NSE are the most common biomarkers used in lung cancer screening in hospital[34][35].

Measurements of serum biomarkers

Blood (10 ml) was collected in serum separator tubes, processed immediately, and separated by centrifugation at 3,000 rpm at room temperature for 10 minutes. The separated serum was then aliquoted and stored at -80°C for the measurement of the six biomarkers mentioned above. CEA and NSE were determined by electro-chemiluminescence immunoassay (ECLIA), using the Roche E601 chemical luminescence immunity analyzer with the auxiliary reagent kit (Dongying J&M Chemical Co., Ltd., China). LDH, CRP, Na^+, and Cl^- were measured by

Table 2. Demographic and clinical profiles of NSCLC patients and controls included in this study ($\bar{x} \pm s$).

Demographic profile	Controls (n = 130)	NSCLC (n = 130)	*p-value**
Age (years)	56.24±8.94	57.75±10.69	*0.17*
Range (age)	29–81	21–80	-
Sex			
Male	64	69	*0.385*
Female	66	61	
Stage (I,II)	-	130	-
Smoking	64/66	72/58	*0.987*

*Statistics were conducted using the independent-Samples T Test and chi-square test.

doi:10.1371/journal.pone.0125517.t002

polyacrylamide gel electrophoresis (PAGE), immunoturbidimetry (ITM), and ion selective electrode methods, respectively, using the Hitachi 7600–020 automatic biochemical analyzer (Beijing Leadman Biochemical Technology Company, Beijing, China). Results were presented as mean values of duplicates after the subtraction of background values. The normal critical values of LDH (99–245 u/l), CRP (0-3mg/l), Na^+ (136–146 mmol/l, Cl^- (96-108mmol/l), CEA (0–3.4 ng/ml), and NSE (0-17ng/ml) were used as standards.

Gene expression programming (GEP) models

GEP is an evolutionary algorithm introduced by Ferreira in 2001[25]. It can emulate biological evolution based on computer programming. With the assumption of being, in some way, a natural development of genetic programming (GP) preserves few properties of genetic algorithms (GA)[36][37]. The GEP algorithm inherits the advantages of GA and GP, but overcomes their disadvantages. In contrast to GP, the chromosomes in GEP are not represented as trees, but as linear strings of fixed length, with features taken from GA. GEP adopts a simple linear fixed-length manner to describe individuals; it is therefore easy to use a nonlinear tree structure to solve complicated nonlinear problems, thus achieving the purpose of using simple coding to solve complex problems[38]. GEP uses characteristic linear chromosomes, which are composed of the genes structurally organized in the head and the tail. Head may contain functional elements like {Q, +, −, ×, /} or terminal elements like, "Q" is the statistical function of square root. The size of the tail (t) is computed as t = h (n-1) + 1, where n is the maximum number of parameters required in the function set[39]. When the representation of each gene is given, the genotype is established. It is then converted to the phenotype expression tree (ET). The chromosomes function is used as a genome and is modified by means of mutation, transposition, root transposition, gene transposition, gene recombination, and one- and two-point recombination. The flowchart of a gene expression algorithm (GEA) is shown in Fig 2. [24].

The algorithm begins with the random creation of the chromosomes in the initial population. Then the chromosomes are expressed and the fitness of each individual is evaluated. According to fitness, reproduction with modification is made, the individuals are then selected and the results lead to new traits. Additionally, the individuals of this new generation are subjected to the same developmental process: expression of the genomes, confrontation of the selection environment, and reproduction with modification. It is repeated for a certain number of generations until a satisfying solution has been found. It is important that the individuals are selected and copied into the next generation according to the fitness by roulette wheel sampling with elitism. This guarantees the survival and cloning of the best individual to the next generation. Each GEP gene contains a list of symbols with a fixed length that can be any element from a function set [36]:

$$\{+; -; *; /; \leq; \geq; >; =; <; sqr; sqrt; \exp; \ln; \cos; \sin; \tan\} \quad (1)$$

The optimum fitness is:

$$fitness(i) = \frac{TP + TN}{TP + FN + TN + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

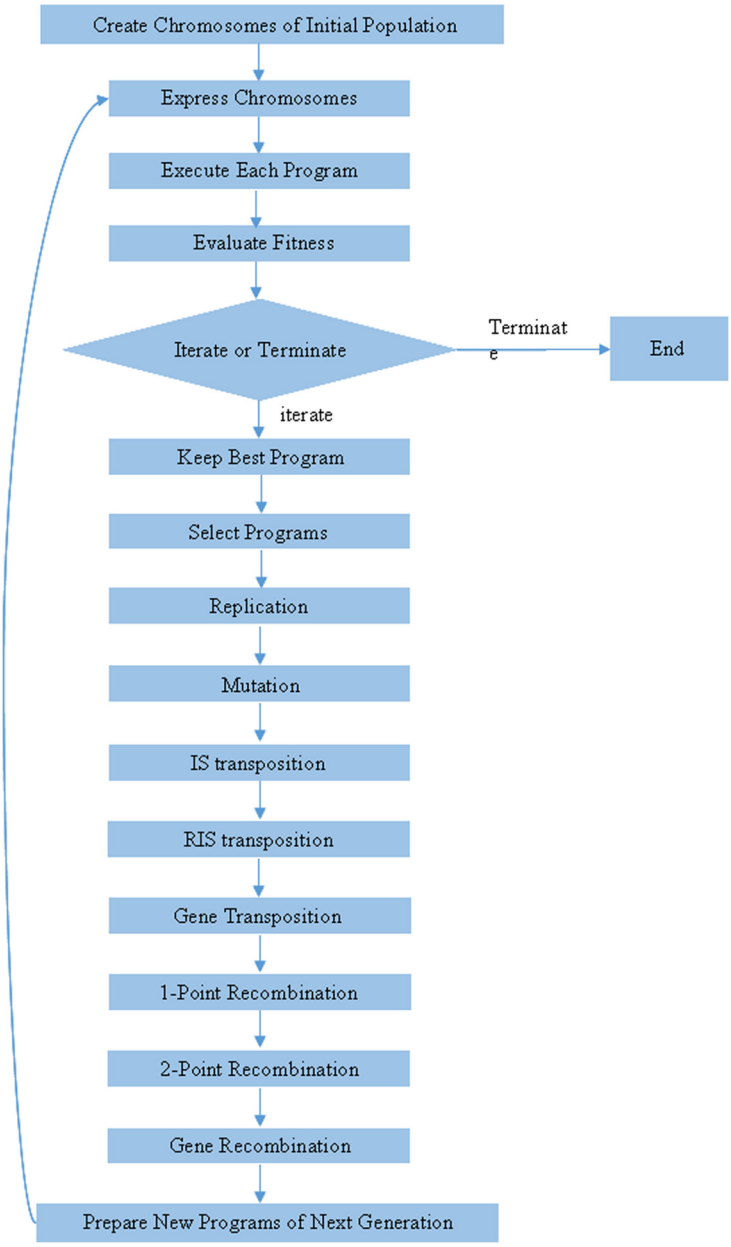

Fig 2. The flowchart of the GEP modeling in this study.

doi:10.1371/journal.pone.0125517.g002

TP, TN, FP, FN are the number of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs), respectively.

The theory of ANN models

Artificial Neural Networks (ANNs) that has the ability of classification is a mathematical model, which original designed to imitate human neural system. Multiple neurons interconnect to each other and arranged in to a wiring layer. ANNs use complicated layers (called hidden layers) to deal input and output, the input where each neuron represents an independent variable. ANNs contain a series of different architectures including Multilayer Perceptron

(MLP) and Radial Basis Function (RBF) [17][39]. MLP employs the back-propagation learning algorithm and a non-linear function to transmit the sum. RBF network activates neuron in hidden layer through radial basis function which has two parameters: the center location of the function and its bias. In RBF network, the hidden layer accepts input data via an unsupervised form[40].

Statistical analyses

Statistical analyses were performed using SPSS 16.0. Differences between groups were calculated by means of a nonparametric Wilcoxon test (Mann–Whitney U test), independent-Samples T Test and chi-square test. *P values* < 0.05 were considered to be statistically significant.

Detection capability comparison

The Receiver Operating Characteristic (ROC) curves were used to describe sensitivity of biomarkers, alone and combined, which were graphed by "R programming project 2.15–1". Using ANNs to compare the detection capability, we can ascertain the optimal algorithm.

Ethics statement

Research approval was obtained from the Ethics Committee of Qingdao University Medical College and written informed consent was obtained from all participants. The study were followed by the STARD (Standards for Reporting of Diagnostic Accuracy) checklist to improve the accuracy and completeness of reporting of studies of diagnostic accuracy[41].

Results

Demographic and clinical profiles, as well as the serum levels of six biomarkers of SCLC patients and normal controls

The clinical characteristics of SCLC patients and normal controls were summarized in Table 1, the NSCLC patients and controls were in Table 2. No significant differences of age and smoking history were observed between these two groups. To establish a novel multiple-analysis of serum biomarkers for efficient screening of SCLC, a set of six biomarkers were selected and their serum concentrations were determined by 145 lung cancer patients and 155 control subjects (S1 Dataset). SCLC patients exhibited significantly higher concentrations of serum LDH, CRP, CEA, and NSE than normal controls ($p < 0.001$), whereas the concentrations of Na^+ and Cl^- were significantly lower than in normal controls ($p < 0.001$) (Table 3). There are significant differences in the concentrations of LDH, Na, Cl and NSE between SCLC and NSCLC means

Table 3. Serum levels of six biomarkers in SCLC patients and control subjects.

biomarker	Controls (n = 155)		SCLC (n = 145)		Z-value	*P-value**
	Median	Range	Median	Range		
LDH(u/l)	146	55–397	180	3–801	-6.506	*<0.0001*
CRP(mg/l)	1.36	0.04–18.2	6.18	0.04–117.96	-8.57	*<0.0001*
Na^+ (mmol/l)	142.47	127–146.83	140	101.4–146.1	-6.614	*<0.0001*
Cl^- (mmol/l)	105	98–111	102	78–137.8	-7.328	*<0.0001*
CEA(ng/ml)	2.07	0.2–14.66	4.29	0.08–181	-7.421	*<0.0001*
NSE(ng/ml)	12.44	6.76–38.19	24.27	1.07–370	-5.081	*<0.0001*

* Statistics were conducted using the non-parametric Wilcoxon test (Mann–Whitney U test).

doi:10.1371/journal.pone.0125517.t003

Table 4. Serum levels of six biomarkers in SCLC patients and NSCLC patients.

biomarker	NSCLC(n = 130)		SCLC (n = 145)		Z-value	*P-value**
	Median	Range	Median	Range		
LDH(u/l)	159.98	10–540	180	3–801	-5.043	*<0.0001*
CRP(mg/l)	19.55	0–145	6.18	0.04–117.96	-0.515	*0.607*
Na^+ (mmol/l)	141.57	134–146	140	101.4–146.1	-4.777	*<0.0001*
Cl^- (mmol/l)	102.60	1–110	102	78–137.8	-4.351	*<0.0001*
CEA(ng/ml)	52.66	0–781	4.29	0.08–181	-2.857	*0.010*
NSE(ng/ml)	13.34	1–40	24.27	1.07–370	-4.728	*<0.0001*

* Statistics were conducted using the non-parametric Wilcoxon test (Mann–Whitney U test).

doi:10.1371/journal.pone.0125517.t004

that these biomarkers are particularly suitable for SCLC (Table 4). The correlation analysis depended on Spearman rank correlation analysis was to exclude potential confounders, the correlation coefficient which is close to "1" means repetitive in the GEP models, the six biomarkers perform their mission well and have significant role respectively. (Table 5).

ROC curves analyses to represent sensitivity/specificity of each biomarker and their combinations

The ROC curves to discover the sensitivity/specificity in each biomarker were determined by comparison with the area under the curve, we found the result in serum sodium and serum chloride were lower than any other biomarkers (Fig 3), then build models dividing two groups to confirm whether Na^+ and Cl^- are meaningful in the detection of lung cancer patients and controls. Model 1 has united all the six biomarkers and model 2 has conjoined four biomarkers that remove serum sodium and serum chloride. The striking difference of the performance in model 1 and model 2 was graphed in Fig 4, the model 1 with 6 biomarkers in the ROC curve has a significant advantage (Fig 4).

GEP modeling

GEP model 1 incorporating six serum biomarkers. A software known as "Automatic Problem Solver 3.0" was used to run the algorithm. The GEP modeling randomly selected four of five partitions as a training set (240 subjects) for model generation, including 115 SCLC patients and 125 normal controls. Next, the GEP parameters were modified to test the remaining 60 subjects for model validation. The concentration of six biomarkers was input into the GEP model to calculate its detection sensitivity and specificity for the discrimination of SCLC and

Table 5. The correlation analysis of the biomarkers were depended on Spearman rank correlation analysis (r = correlation coefficient, P value *of 0 <0.0001*).

	LDH	CRP	Na	Cl	CEA	NSE
LDH	1	0.302	0.006	0.049	0.289	0.295
CRP		1	0.161	0.199	0.093	0.063
Na			1	0.705	0.025	0.054
Cl				1	0.038	0
CEA					1	0.109
NSE						1

doi:10.1371/journal.pone.0125517.t005

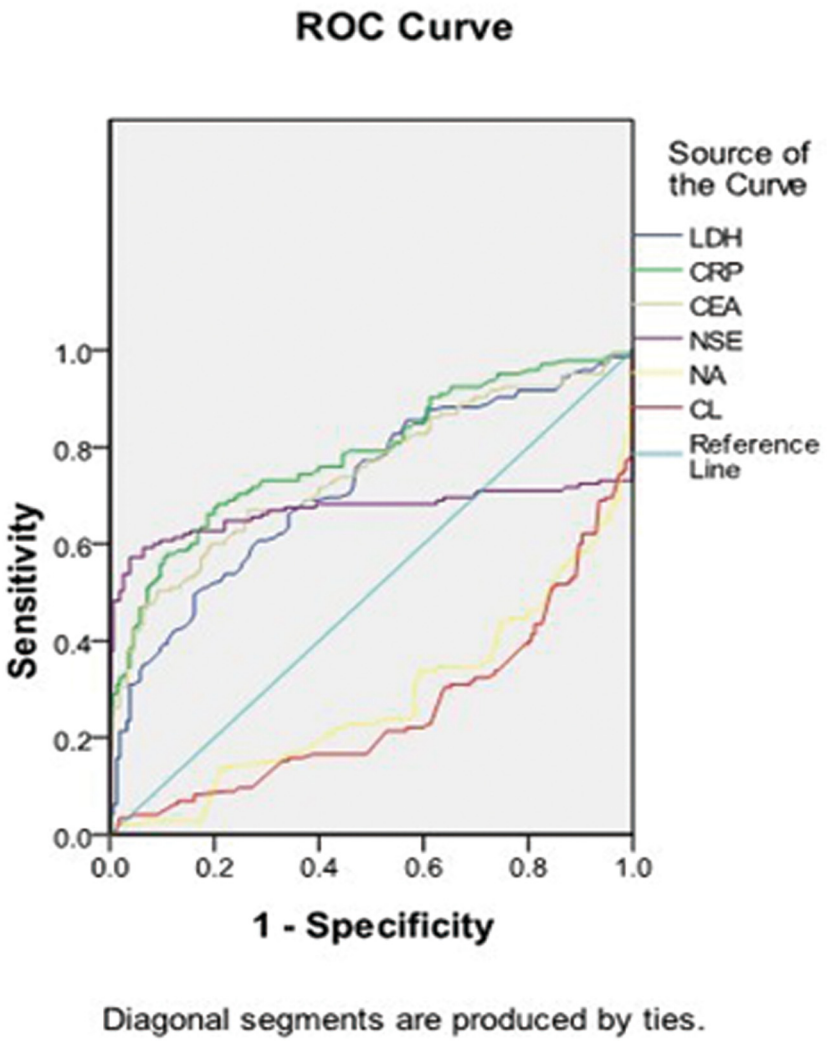

Fig 3. ROC curves analyses to represent sensitivity/specificity of each biomarker, and the Area Under the Curve represents: LDH = 0.717, CRP = 0.786, CEA = 0.748, NSE = 0.670, Na^+ = 0.279, Cl^- = 0.255.

doi:10.1371/journal.pone.0125517.g003

normal controls. GEP model 1 used all six biomarkers as inputs and the algorithm was:

$$y = \frac{x_0 - x_4}{x_5 \times \sqrt{\log_{10} x_0}} + \frac{e^{apsLogi(x_3)} + x_4}{apsLogi(x_5)} + x_2 \times apsLogi(apsLogi(\frac{1}{apsLogix_0} + x_4)) + e^{apsLogi(\log_{10} x_3)} + x_1 - x_2 + x_5$$

If the calculated value of "y" equal to or greater than the rounding threshold, then the record is classified as "1", "0" otherwise. The variables x_0, x_1, x_2, x_3, x_4, and x_5 represented the biomarkers LDH, CRP, Na^+, Cl^-, CEA, and NSE, respectively.

Patients suffered from lung cancer were marked as class "1", while the healthy subjects were marked as class "0". The serum concentrations of LDH, CRP, Na^+, Cl^-, CEA, and NSE were used as inputs in model 1. The general experiment setup was summarized in Table 6. This model successfully discriminated 281 out of 300 subjects, which represented a determination coefficient of 93.75% (225/240) and 93.33% (56/60) for training and test sets, respectively (S1 Dataset).

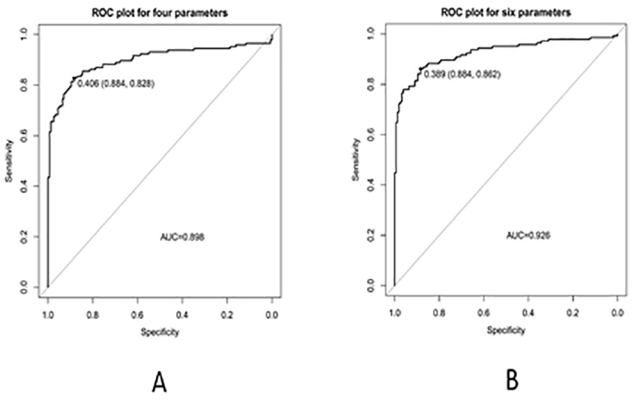

Fig 4. Comparison of the performance (sensitivity) from combined biomarkers, A is trained with six biomarkers and B is trained with four biomarkers. The sensitivity trained by six biomarkers combination performed better than four biomarkers.

doi:10.1371/journal.pone.0125517.g004

GEP model 2 including four biomarkers. While the performance of model 1 with 6 biomarkers was good, we wanted to ascertain whether the numbers of biomarkers could be decreased to only four, which could significantly reduce the cost and time for SCLC screening. In model 2, we only chose the markers that were widely used in the detection of lung cancer, including LDH, CRP, CEA, and NSE, with the same function set described above.

The algorithm of GEP model 2 was:

$$y = \sqrt{e^{x_3} \times x_3 \times x_2} + \sqrt{apsLogi(x_0)} + x_0 \times x_3 \times (x_1 + x_2 - 1) - \log_{10} x_3$$
$$+x_0^2 \times (\log_{10} x_0 + x_2 - x_3 + e^{\frac{x_1}{x_3}})$$

$$apsLogi(x) = \frac{1}{1 + e^{-x}}$$

If the calculated value of "y" equal to or greater than the rounding threshold, then the record

Table 6. SCLC detection rate of GEP model 1 and model 2.

	Model 1		Model 2	
	Training set n = 240	Test set n = 60	Training set n = 240	Test set n = 60
Accuracy	93.75%	93.33%	93.75%	91.67%
Sensitivity	92.17%	93.33%	89.57%	86.67%
Specificity	95.20%	93.33%	97.60%	96.67%
Error	6.25%	6.67%	6.25%	8.33%
CC	0.87	0.87	0.88	0.84
MSE	0.06	0.07	0.06	0.08
RAE	0.13	0.12	0.13	0.17
MAE	0.06	0.06	0.06	0.08
RSE	0.25	0.27	0.25	0.33

CC = Correlation Coefficient; MSE = Mean Squared Error; RAE = Root Mean Squared Error; MAE = Mean Absolute Error; RSE = Relative Squared Error.

doi:10.1371/journal.pone.0125517.t006

is classified as "1", "0" otherwise. In this model, variables x_0, x_1, x_2, and x_3 were biomarkers LDH, CRP, CEA, and NSE, respectively.

The accuracy of GEP model 2 was 91.66% and the sensitivity was 86.67% in the test set, which was lower than that in model 1 (Table 7). All trainings were made in triplicate to assure that the best architecture was chosen. We have made other combinations to make sure the model 1 is the optimized biomarker panel that acquired the supreme predicted value.

Development of model by Artificial Neural Networks

In order to compare the classification power between GEP and ANN, IBM SPSS Statistics 18.0 was applied to build ANNs (MLP and RBF models) prediction models. The model1 and model2 were as same to GEP. SCLC patients and controls (0 or 1) were input as a dependent variable as GEP models. Using model 1, MLP indicated accuracy of 85.4%, 80.0% and in RBF acquired an accuracy of 80.0%, 78.3% for training and test phase, respectively. In addition, in model 2 the correct classification rate for MLP represented the identification of 83.3% and 83.3% and for RBF was for 84.2%, 83.3% among training and testing stages, respectively. The software have been ran three times and covariant was different arrange to select the best (Table 8) (Fig 5).

Compared to ANNs, the GEP algorithm proves the supreme predictive rate which has significant strengths. The ROC curve and GEP model showed that the model 1 is the adequate combination to distinguish lung cancer patients from high-risk people.

Table 7. Parameter settings for the GEP algorithm.

Parameter	Settings
General	
Chromosomes	100
Genes	5
Head size	8
Gene size	17
Linking function	Addition
Function set	+ - * / Exp Sqrt Log Logi Inv
Complexity increase	
Generations without change	200
Number of tries	3
Max. complexity	5
Genetic operators	
Mutation rate	0.044
Inversion rate	0.1
IS transposition rate	0.1
RIS transposition rate	0.1
One-point recombination rate	0.3
Two-point recombination rate	0.3
Gene recombination rate	0.1
Gene transposition rate	0.1
Numerical constants	
Constants per gene	10
Data type	Floating-point
Lower bound	-10
Upper bound	10

doi:10.1371/journal.pone.0125517.t007

Table 8. The detection capability of ANN models in SCLC patients and normal controls.

	Model 1		Model 2	
	Training set	Test set	Training set	Test set
	n = 240	n = 60	n = 240	n = 60
Accuracy(MLP)	85.4%	80%	83.3%	83.3%
Accuracy(RBF)	80.0%	78.3%	84.2%	83.3%

doi:10.1371/journal.pone.0125517.t008

GEP model 1 incorporating six serum biomarkers performed by limited stage and extensive stage. The optimal GEP model 1 was used to make a comparison between early and late SCLC (74 limited stage and 71 extensive stage). We selected 74 cases from the 155 non-cancer cases as the healthy control. Firstly, in order to explore the early SCLC, as the above method GEP model randomly selected four of five partitions as a training set (118 subjects) for model generation, including 59 early SCLC patients and 59 normal controls. Remaining 30 cases (15 early SCLC and 15 normal controls) were for model validation. It can be observed that the early SCLC acquired the accuracy of 92.37% (109/118) and 90% (27/30) for training and test set, respectively. Secondly, for late SCLC, 116 subjects (57 late SCLC and 59 normal controls) for model generation and 29 cases for model validation, it represented the accuracy of 96.52% (112/116), 91.30% (27/29) for training and test set, respectively. The results showed that the accuracy of late SCLC in GEP model 1 was performed better than early SCLC and total 145 SCLC, but the early SCLC accuracy was close to the result of 145 SCLC, it was still a good performance (S3 Dataset) (S4 Dataset).

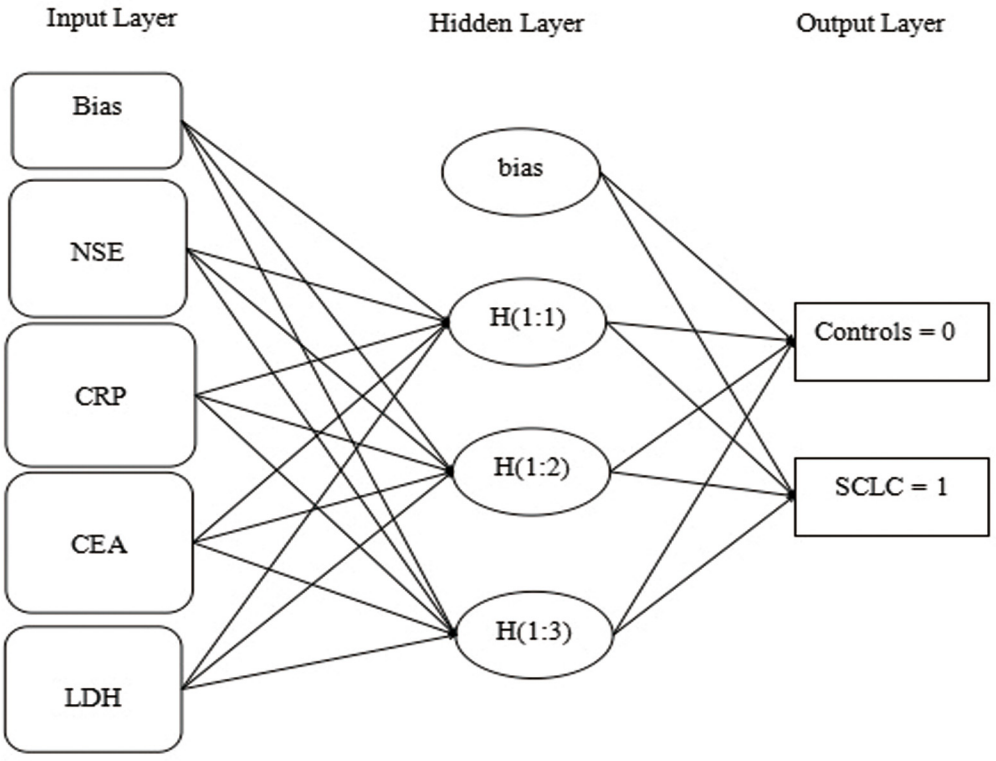

Fig 5. The structure of the ANNs implemented.

doi:10.1371/journal.pone.0125517.g005

Table 9. The detection capability of GEP model 1 with six biomarkers in SCLC and NSCLC patients.

	SCLC		NSCLC	
	Training set	Test set	Training set	Test set
	n = 240	n = 60	n = 208	n = 52
Accuracy	93.75%	93.33%	87.50%	86.53%
Sensitivity	92.17%	93.33%	81.73%	84.62%
Specificity	95.20%	93.33%	93.26%	88.46%

doi:10.1371/journal.pone.0125517.t009

GEP model 1 performed by NSCLC patients and normal controls. To confirm the GEP model 1 test, NSCLC patients have been included in the negative control with healthy subjects. As the above method, GEP randomly selected 208 subjects (104 NSCLC patients and 104 normal controls) for model generation, 52 subjects (26 NSCLC patients and 26 normal controls) for model validation respectively. It indicated that the accuracy of 87.5% (182/208), 86.5% (45/52) for training and test set, respectively. Meanwhile, the results were significantly worse than SCLC patients and were enough to prove that the GEP model is suitable for the SCLC patients (Table 9) (S2 Dataset).

Discussion

SCLC accounts for approximately 13–18% of all lung cancers, with diverse incidences in different countries[42]. Without treatment, it has the most aggressive clinical course of all lung cancer types, with survival from 2 to 4 months[43]. Diagnosis of SCLC at its early stage is challenging, because it is usually asymptomatic until advanced stages, which causes poor prognosis[44]. This emphasizes the significance of a reliable early-stage diagnosis method to prolong lives[45].

Various methods have been used for the detection of SCLC, such as thoracic radiography, sputum cytology, and CT. The efficacy of these tools has been evaluated in clinical trials and it turns out that thoracic radiography and sputum cytology have low sensitivity for early-stage detection of SCLC [46,47]. Although CT imaging has emerged as an effective technique for the diagnosis of many human diseases, the most prominent limitation of CT imaging for the detection of lung cancers is the high rate of mistaken benign pulmonary nodules as lung cancers [48,49]. In addition, CT imaging examination is still costly for most people in developing countries and medical insurance agencies would not approve the use of CT scans as a surveillance strategy for lung cancers.

Biological markers can be easily detected in biological fluids using minimally invasive procedures, which can significantly enhance the detection rate of a number of human cancers. Several tumor markers, such as α-fetoprotein (AFP), prostate specific antigen (PSA), and cancer antigen125 (CA125), have been proven to be highly sensitive and effective for the screening of liver, prostate, and ovarian cancers [50]. Each biomarker has low diagnostic because of limited sensitivity and specificity which is partially owing to the heterogeneous of the disease [15,51]. Many tumor markers are not used alone for routine tumor screening because of low detection rates and unacceptable false-positive diagnoses [52]. In this study, some conventional and economical markers such as LDH, CRP, Na^+, Cl^- and other two tumor biomarkers(CEA, NSE) were selected based on previous studies to establish the GEP model for the detection of SCLC. These biomarkers can be easily tested, even in developing regions, using two kits. For example, LDH and CRP, two important inflammation markers, are routinely tested in most hospitals in China, let alone electrolyte solution Na^+, Cl^-.

A previous study conducted by Flores, *et al.*[44,53,15] included 63 lung cancer patients, 87 non-cancer controls. The ANN model was trained with a set of biomarkers (Cyfra 21.1, CEA, CA125 and CRP) and achieved a correct classification rate of 88.9%, 93.3% and 90% in training, validation and testing phases, respectively. Feng, *et al.*[19] reached a prediction rate of 87.3% for the detection of lung cancers in a test phase using an ANN model with the above six biomarkers and 19 additional parameters, such as risk factors, symptoms, smoking, chemical exposure, kitchen environment, etc. Another study reached 90% specificity for the detection of lung cancer in the training set, based on a three-biomarker panel comprised of macrophage migration inhibitory factor (MIF), prolactin (PRL), and thrombospondin (THSP)[12]. According to the characteristic of "black-box" in ANN, we did not know how an ANN learns to perform its classification, merely giving a final results cause we fail to discern why it did not work[17]. Nevertheless, the GEP perform well even if there is large sophisticated data and offer a visual formula model. In our study, using the ROC curve to detect each sensitivity/specificity, we perceived that the area under the curve of Na^+ and Cl^- is lower than others and the six biomarkers emerged the best. Then in GEP model 1, incorporating six biomarkers, successfully distinguished 281 of 300 tested samples with an accuracy of 93.75% (225/240) and 93.33% (56/60) for the training and test sets, respectively. Model 2, including four biomarkers, had slightly lower accuracies of 93.75% and 91.67% for the training and test sets, respectively. To confirm the excellent result in GEP, we repeated the models on ANN, Our results exhibited that the accuracy of GEP models performed higher than that in ANNs. The six biomarkers combined in MLP indicated a standout result that seem as to in GEP. Therefore, when compare the detection capability of ROC curve, GEP and ANN, GEP was proved to be the best algorithm which depend on model 1, otherwise, GEP model 2 with four biomarkers may be more suitable for screening SCLC in regions with extremely low incomes.

To confirm the GEP model 1 test, NSCLC patients have been included in the negative control with healthy subjects**,** the results were significantly worse than SCLC patients, also, there are significant differences in the concentrations of LDH, Na, Cl and NSE between SCLC and NSCLC. The results were enough to prove that the GEP model is suitable for the SCLC patients. Furthermore, The optimal GEP model 1 was used to make a comparison between early and late SCLC patients, the early SCLC model 1 acquired the accuracy of 92.37% (109/118) and 90% (27/30) for training and test set. It was close to total data accuracy. Meanwhile, the late SCLC model 1 performed well than early stage which represented the accuracy of 96.52% (112/116), 91.30% (27/29) for training and test set, respectively. The particular poor physical condition of the advanced patients may explain it. Generally, the accuracy of early stage was still good and GEP model 1 can be the optimal test for SCLC early detection.

In clinical examination, the serum CRP and LDH levels in SCLC patients are significantly higher than healthy people, but the serum sodium is much lower. The clinical significance of serum level of LDH has been proven to be a strong and independent predictive factor of median survival, both in limited and extensive disease stages of SCLC[54]. In addition, the correlation between inflammation and cancer risk has been reported in many studies. For example, tumor growth causes inflammation in tumor tissues, which can be regarded as an indicator of immune response to tumor antigens. In addition, cancer cells can increase the production of inflammatory cytokines, causing increased CRP levels in cancer patients[55,56]. CRP is a non-specific acute-phase inflammatory response serum marker produced by hepatocytes and regulated by interleukin IL-6. The association between CRP and lung cancer has been widely investigated[57]. CEA is also an independent prognostic indicator associated with reduced survival in SCLC[34]. NSE has been regarded as the most sensitive tumor marker for SCLC at the time of diagnosis[35]. The close association of these biomarkers with SCLC is an important factor leading to the outstanding performance of our GEP models.

It has been reported that the syndrome of inappropriate antidiuretic hormone (SIADH) is the leading cause of hyponatremia and hypochloremia in hospitalized patients. Malignancy pulmonary diseases, particularly SCLC, usually lead to SIADH[31]. Therefore, clinical characteristics of hyponatremia and hypochloremia promote the search for underlying lung cancers by testing serum levels of biomarkers[58,59]. The other reasons for hyponatremia are the persistent natriuresis and inappropriately low aldosterone levels caused by increased levels of atrial natriuretic peptides (ANPs)[60], as well as involvement of the adrenal gland or the brain through metastases[61]. Thus, model l, which included Na^+ and Cl^-, as well as LDH, CRP, CEA, and NSE, had a slightly better performance for the detection of SCLC than model 2. GEP model 2 did not affect the detection accuracy by much, perhaps because paraneoplastic syndrome caused by SCLC is not common in clinical settings. However, GEP model 2 might be improved by adding more samples, new subjects, or keeping the normalizing criteria to train the model. In addition, clinical information, such as other biomarkers, nodules, and hemoptysis, etcetera, could also be included to improve the GEP performance.

In summary, we developed an effective GEP model incorporating six biomarkers to screen SCLC patients. This model and measurements of six biomarkers are convenient, economical, and can be widely used in less developed area. However, this model should be further tested and improved with more SCLC patients in different hospitals and regions. With the emergence of new predictive tumor markers, we are no longer going to select several determinate tumor markers joint detection, but going to obtain larger sample size, higher amount of information and larger scale on gene and protein levels. Moreover, due to the intricate parameter selection, the parameters in GEP algorithm may be optimized in the later research. Also, those serology tests couldn't replace lung biopsy to confirm a diagnosis, rather they serve as a good screening tool to auxiliary diagnosis. In addition, the economic cost of this GEP model needs to be comprehensively evaluated before it is widely applied in the clinical screening for SCLC.

Supporting Information

S1 Dataset. The six biomarkers data which were used in GEP models between SCLC patients and controls.
(XLS)

S2 Dataset. The six biomarkers data which were used in GEP models between NSCLC patients and SCLC patients.
(XLSX)

S3 Dataset. The six biomarkers serum concentrations of limited stage SCLC that used in GEP models.
(XLSX)

S4 Dataset. The six biomarkers serum concentrations of extensive stage SCLC that used in GEP models.
(XLSX)

Author Contributions

Conceived and designed the experiments: ZY HJL HZS SHL. Performed the experiments: HJL CL XJY. Analyzed the data: HJL HZS XCL. Contributed reagents/materials/analysis tools: HZS XCL. Wrote the paper: HJL SHL XY. Participate the submission: CHG HJL LHC.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, et al. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127: 2893–2917. doi: 10.1002/ijc.25516 PMID: 21351269
2. Patz EF Jr, Campa MJ, Gottlin EB, Kusmartseva I, Guan XR, et al. (2007) Panel of serum biomarkers for the diagnosis of lung cancer. J Clin Oncol 25: 5578–5583. PMID: 18065730
3. Ghosal R, Kloer P, Lewis KE (2009) A review of novel biological tools used in screening for the early detection of lung cancer. Postgrad Med J 85: 358–363. doi: 10.1136/pgmj.2008.076307 PMID: 19581246
4. National Lung Screening Trial Research T, Aberle DR, Adams AM, Berg CD, Black WC, et al. (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365: 395–409. doi: 10.1056/NEJMoa1102873 PMID: 21714641
5. Grondin SC, Liptay MJ (2002) Current concepts in the staging of non-small cell lung cancer. Surg Oncol 11: 181–190. PMID: 12450554
6. Pieterman RM, van Putten JW, Meuzelaar JJ, Mooyaart EL, Vaalburg W, et al. (2000) Preoperative staging of non-small-cell lung cancer with positron-emission tomography. N Engl J Med 343: 254–261. PMID: 10911007
7. Sone S, Li F, Yang ZG, Honda T, Maruyama Y, et al. (2001) Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. Br J Cancer 84: 25–32. PMID: 11139308
8. Bajtarevic A, Ager C, Pienz M, Klieber M, Schwarz K, et al. (2009) Noninvasive detection of lung cancer by analysis of exhaled breath. BMC Cancer 9: 348. doi: 10.1186/1471-2407-9-348 PMID: 19788722
9. Reich JM (2008) A critical appraisal of overdiagnosis: estimates of its magnitude and implications for lung cancer screening. Thorax 63: 377–383. doi: 10.1136/thx.2007.079673 PMID: 18364449
10. Marshall E (2008) Medicine. A bruising battle over lung scans. Science 320: 600–603. doi: 10.1126/science.320.5876.600 PMID: 18451275
11. Nolen BM, Langmead CJ, Choi S, Lomakin A, Marrangoni A, et al. (2011) Serum biomarker profiles as diagnostic tools in lung cancer. Cancer Biomark 10: 3–12. doi: 10.3233/CBM-2012-0229 PMID: 22297547
12. Cho WC (2007) Potentially useful biomarkers for the diagnosis, treatment and prognosis of lung cancer. Biomed Pharmacother 61: 515–519. PMID: 17913444
13. Molina R, Auge JM, Escudero JM, Marrades R, Vinolas N, et al. (2008) Mucins CA 125, CA 19.9, CA 15.3 and TAG-72.3 as tumor markers in patients with lung cancer: comparison with CYFRA 21–1, CEA, SCC and NSE. Tumour Biol 29: 371–380. doi: 10.1159/000181180 PMID: 19060513
14. Flores-Fernández JM, Herrera-López EJ, Sánchez-Llamas F, Rojas-Calvillo A, Cabrera-Galeana PA, et al. (2012) Development of an optimized multi-biomarker panel for the detection of lung cancer based on principal component analysis and artificial neural network modeling. Expert Systems with Applications 39: 10851–10856.
15. Fernández A, Martínez A, Gaspar M, Filella X, Molina R, et al. (2007) Marcadores tumorales serológicos. Química Clínica 26: 77–85.
16. Chu XY, Hou XB, Song WA, Xue ZQ, Wang B, et al. (2011) Diagnostic values of SCC, CEA, Cyfra21-1 and NSE for lung cancer in patients with suspicious pulmonary masses: a single center analysis. Cancer Biol Ther 11: 995–1000. PMID: 21483235
17. Cruz JA, Wishart DS (2006) Applications of machine learning in cancer prediction and prognosis. Cancer Inform 2: 59–77. PMID: 19458758
18. Adetiba E, Ibikunle FA (2011) Ensembling of EGFR Mutations' based Artificial Neural Networks for Improved Diagnosis of Non-Small Cell Lung Cancer. International Journal of Computer Applications 20: 39–47.
19. Feng F, Wu Y, Wu Y, Nie G, Ni R (2012) The effect of artificial neural network model combined with six tumor markers in auxiliary diagnosis of lung cancer. Journal of medical systems 36: 2973–2980. doi: 10.1007/s10916-011-9775-1 PMID: 21882004
20. Sesen MB, Kadir T, Alcantara R-B, Fox J, Brady M (2012) Survival prediction and treatment recommendation with Bayesian techniques in lung cancer. American Medical Informatics Association. pp. 838.
21. Hosseinzadeh F, KayvanJoo AH, Ebrahimi M, Goliaei B (2013) Prediction of lung tumor types based on protein attributes by machine learning algorithms. SpringerPlus 2: 238. doi: 10.1186/2193-1801-2-238 PMID: 23888262
22. Hirales Casillas CE, Flores Fernández JM, Camberos EP, Herrera López EJ, Pacheco GL, et al. (2014) Current status of circulating protein biomarkers to aid the early detection of lung cancer. Future Oncology 10: 1501–1513. doi: 10.2217/fon.14.21 PMID: 25052758

23. Cheng J, Greiner R (1999) Comparing Bayesian network classifiers. Morgan Kaufmann Publishers Inc. pp. 101–108.
24. Zhou C, Xiao W, Tirpak TM, Nelson PC (2003) Evolving accurate and compact classification rules with gene expression programming. Evolutionary Computation, IEEE Transactions on 7: 519–531.
25. Ferreira C, Gepsoft U (2008) What is Gene Expression Programming.
26. Ferreira C (2002) Gene expression programming in problem solving. Soft Computing and Industry: Springer. pp. 635–653.
27. Kusy M, Obrzut B, Kluska J (2013) Application of gene expression programming and neural networks to predict adverse events of radical hysterectomy in cervical cancer patients. Medical & biological engineering & computing 51: 1357–1365.
28. Lu S-HL (2014) Prediction of lung cancer based on serum biomarkers by gene expression programming methods. Asian Pacific journal of cancer prevention: APJCP 15: 9367. PMID: 25422226
29. Müller-Hermelink HK, Engel P, Kuo T (2004) Pathology & Genetics, Tumours of the Lung, Pleura, Thymus and Heart. Lyon, France: IARC Press.
30. Vantyghem M-C, Balavoine A-S, Wémeau J-L, Douillard C (2011) Hyponatremia and antidiuresis syndrome. Elsevier. pp. 500–512.
31. Bucher C, Tapernoux D, Diethelm M, Büscher C, Noser A, et al. (2014) Influence of weather conditions, drugs and comorbidities on serum Na and Cl in 13000 hospital admissions: Evidence for a subpopulation susceptible for SIADH. Clinical biochemistry 47: 618–624. doi: 10.1016/j.clinbiochem.2013.12.021 PMID: 24389078
32. Balkwill F, Mantovani A (2001) Inflammation and cancer: back to Virchow? The lancet 357: 539–545. PMID: 11229684
33. Heikkilä K, Ebrahim S, Lawlor DA (2007) A systematic review of the association between circulating concentrations of C reactive protein and cancer. Journal of epidemiology and community health 61: 824–833. PMID: 17699539
34. Yang X, Wang D, Yang Z, Qing Y, Zhang Z, et al. (2011) CEA is an independent prognostic indicator that is associated with reduced survival and liver metastases in SCLC. Cell biochemistry and biophysics 59: 113–119. doi: 10.1007/s12013-010-9121-0 PMID: 20945115
35. Erbaycu AE, Gunduz A, Batum O, Ucar ZZ, Tuksavul F, et al. (2010) Pre-treatment and treatment-induced neuron-specific enolase in patients with small-cell lung cancer: an open prospective study. Archivos de Bronconeumología ((English Edition)) 46: 364–369.
36. Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection: MIT press.
37. Salomon R (1996) Re-evaluating genetic algorithm performance under coordinate rotation of benchmark functions. A survey of some theoretical and practical aspects of genetic algorithms. BioSystems 39: 263–278. PMID: 8894127
38. Si H, Gao H, Yao X, Hu Z (2011) Study the chromatographic hydrophobicity index based on gene expression programming. Journal of Computational Science & Engineering 1: 22–31. doi: 10.1186/1753-6561-8-S6-S4 PMID: 25374613
39. Han XR, Li XC, Si HZ, Ge CZ, Gao H, et al. (2014) QSAR Study of the Anti-Cancer Activity of 38 Compounds in Different Cancer Cell Lines Based on Gene Expression Programming. Trans Tech Publ. pp. 1291–1294.
40. Kayaer K, Yıldırım T (2003) Medical diagnosis on Pima Indian diabetes using general regression neural networks. pp. 181–184.
41. Bossuyt PM, Reitsma JB, E Bruns D, Gatsonis CA, Glasziou PP, et al. (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Clinical chemistry and laboratory medicine 41: 68–73. PMID: 12636052
42. Wahbah M, Boroumand N, Castro C, El-Zeky F, Eltorky M (2007) Changing trends in the distribution of the histologic types of lung cancer: a review of 4,439 cases. Annals of diagnostic pathology 11: 89–96. PMID: 17349566
43. Lekic M, Kovac V, Triller N, Knez L, Sadikov A, et al. (2012) Outcome of small cell lung cancer (SCLC) patients with brain metastases in a routine clinical setting. Radiology and oncology 46: 54–59. doi: 10.2478/v10019-012-0007-1 PMID: 22933980
44. Ghosal R, Kloer P, Lewis K (2009) A review of novel biological tools used in screening for the early detection of lung cancer. Postgraduate medical journal 85: 358–363. doi: 10.1136/pgmj.2008.076307 PMID: 19581246

45. Sun T, Zhang R, Wang J, Li X, Guo X (2013) Computer-aided diagnosis for early-stage lung cancer based on longitudinal and balanced data. PloS one 8: e63559. doi: 10.1371/journal.pone.0063559 PMID: 23691066

46. Chanin TD, Merrick DT, Franklin WA, Hirsch FR (2004) Recent developments in biomarkers for the early detection of lung cancer: perspectives based on publications 2003 to present. Current opinion in pulmonary medicine 10: 242–247. PMID: 15220746

47. Hirsch FR, Mulshine JL (1998) Prevention and early detection of lung cancer-clinical aspects. Clinical and Biological Basis of Lung Cancer Prevention: Springer. pp. 1–14.

48. Welch HG, Woloshin S, Schwartz LM, Gordis L, Gøtzsche PC, et al. (2007) Overstating the evidence for lung cancer screening: the International Early Lung Cancer Action Program (I-ELCAP) study. Archives of internal medicine 167: 2289–2295. PMID: 18039986

49. Wilson DO, Weissfeld JL, Fuhrman CR, Fisher SN, Balogh P, et al. (2008) The Pittsburgh Lung Screening Study (PLuSS) outcomes within 3 years of a first computed tomography scan. American journal of respiratory and critical care medicine 178: 956–961. doi: 10.1164/rccm.200802-336OC PMID: 18635890

50. Sun Z, Fu X, Zhang L, Yang X, Liu F, et al. (2004) A protein chip system for parallel analysis of multi-tumor markers and its application in cancer detection. Anticancer research 24: 1159–1166. PMID: 15154641

51. Türeci Ö, Mack U, Luxemburger U, Heinen H, Krummenauer F, et al. (2006) Humoral immune responses of lung cancer patients against tumor antigen NY-ESO-1. Cancer letters 236: 64–71. PMID: 15992994

52. Schneider J (2006) Tumor markers in detection of lung cancer. Advances in clinical chemistry 42: 1–41. PMID: 17131623

53. Wu Y, Wu Y, Wang J, Yan Z, Qu L, et al. (2011) An optimal tumor marker group-coupled artificial neural network for diagnosis of lung cancer. Expert Systems with Applications 38: 11329–11334.

54. Hermes A, Gatzemeier U, Waschki B, Reck M (2010) Lactate dehydrogenase as prognostic factor in limited and extensive disease stage small cell lung cancer–a retrospective single institution analysis. Respiratory medicine 104: 1937–1942. doi: 10.1016/j.rmed.2010.07.013 PMID: 20719490

55. Chaturvedi AK, Caporaso NE, Katki HA, Wong H- L, Chatterjee N, et al. (2010) C-reactive protein and risk of lung cancer. Journal of Clinical Oncology 28: 2719–2726. doi: 10.1200/JCO.2009.27.0454 PMID: 20421535

56. Lee S, Choe J-W, Kim H-K, Sung J (2011) High-sensitivity C-reactive protein and cancer. Journal of Epidemiology 21: 161–168. PMID: 21368452

57. Pine SR, Mechanic LE, Enewold L, Chaturvedi AK, Katki HA, et al. (2011) Increased levels of circulating interleukin 6, interleukin 8, C-reactive protein, and risk of lung cancer. Journal of the National Cancer Institute 103: 1112–1122. doi: 10.1093/jnci/djr216 PMID: 21685357

58. Hannon M, Thompson CJ (2010) The syndrome of inappropriate antidiuretic hormone: prevalence, causes and consequences. European Journal of Endocrinology 162: S5–S12. doi: 10.1530/EJE-09-1063 PMID: 20164214

59. Seute T, Leffers P, ten Velde GP, Twijnstra A (2004) Neurologic disorders in 432 consecutive patients with small cell lung carcinoma. Cancer 100: 801–806. PMID: 14770437

60. Chute JP, Taylor E, Williams J, Kaye F, Venzon D, et al. (2006) A metabolic study of patients with lung cancer and hyponatremia of malignancy. Clinical cancer research 12: 888–896. PMID: 16467103

61. Yokosuka K, Kawashima T, Okada N, Wakabayashi T, Kawashima S, et al. (2008) Impaired consciousness caused by a metastatic adrenal tumor of pulmonary adenocarcinoma. Internal Medicine 47: 109–112. PMID: 18195500