MEDICAL SCIENCE MONITOR

LAB/IN VITRO RESEARCH

e-ISSN 1643-3750
© Med Sci Monit, 2019; 25: 4999-5004
DOI: 10.12659/MSM.917182

Received: 2019.04.24
Accepted: 2019.06.17
Published: 2019.07.06

# A 10-Long Non-Coding RNA-Based Expression Signature as a Potential Biomarker for Prognosis of Acute Myeloid Leukemia

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

AE Ping Tang
BE Menghan Xie
B Yan Wei
C Xinsheng Xie
C Dandan Chen
AE Zhongxing Jiang

Department of Hematology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, P.R. China

**Corresponding Author:** Zhongxing Jiang, e-mail: zxjiangr@163.com
**Source of support:** This study was supported by the National Natural Science Foundation of China (No. 81400108)

**Background:** Acute myeloid leukemia (AML) is a heterogeneous form of cancer, and it is one of the dominant causes of malignancy-related mortality in patients younger than 35 years old. Therefore, the treatment must be selected based on risk stratification. However, the methods to predict the clinical outcomes of AML are insufficient. Long non-coding RNAs (lncRNAs) are unable or barely able to code for proteins and have attracted remarkable interest because of their involvement in malignancies. Previous studies have proven that some lncRNAs contribute to the development and clinical outcome of AML. Our study constructed a risk stratification system for AML that will facilitate the prediction of clinical outcomes.

**Material/Methods:** We acquired the expression profiles of lncRNAs from the TCGA database to examine their role in the clinical outcomes of AML. We designed and validated a prognostic signature-based risk score system using a sample splitting approach and Cox regression analysis to elucidate the relationship between the clinical outcomes of AML and lncRNAs.

**Results:** We selected 10 lncRNAs to predict the clinical outcome of AML and were able to successfully predict the survival of patients with AML using this 10-lncRNA expression signature.

**Conclusions:** We developed a 10-lncRNA expression signature to predict the clinical outcome of AML. This approach demonstrates remarkable prognostic and therapeutic potential for AML.

**MeSH Keywords:** Leukemia, Myeloid, Acute • Mortality • Prognosis

**Full-text PDF:** https://www.medscimonit.com/abstract/index/idArt/917182







Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

## Background

Acute myeloid leukemia (AML) is a common form of leukemia and it is a major contributor to malignancy-related mortality in people younger than 35 years [1,2]. The treatment for AML is mainly chosen according to cytogenetics, which represents the risk status [1]. Although there has been huge progress in the area of risk stratification, some patients with few risk factors ultimately encounter recurrence [3]. Hence, it is crucial to elucidate the effective markers to improve prediction of the clinical outcome of AML.

Long non-coding RNAs (lncRNAs) have poor or no capability for coding proteins and they have been recently proven to contribute to malignancies. It has been recently demonstrated that lncRNAs display various biological activities and their abnormal expression is related to the development and clinical outcome of human cancers such as AML; therefore, it can be used as a diagnostic marker [4–7].

Our research aimed to construct an lncRNA risk stratification system to facilitate the prediction of the clinical outcome of AML. We carried out AML-related RNA sequencing (RNA-seq) with the help of published data from the Cancer Genome Atlas (TCGA) projects. With the help of the sample splitting approach and Cox regression analysis, we designed and verified a prognostic 10-lncRNA signature-based risk system to elucidate the relationship between the clinical outcome of AML and lncRNAs.

## Material and Methods

### Acquisition of publicly available data from TCGA

The AML-related RNA-seq data set and specific clinical information of the follow-up patients were acquired from the TCGA database. We obtained the information of 139 patients who were additionally randomly selected into a training group (n=71, to examine the crucial lncRNAs), as well as testing group (n=68, to confirm the lncRNA signature) and the entire 139-patient cohort. We obtained 14 376 lncRNA profiles from all of the participants and they were normalized among specimens. The terminal expression of lncRNAs was determined as $\log_2$ (X+1) of the raw expression level.

### Expression levels of lncRNA in patients with AML

Since the expression of lncRNAs is comparatively repressed, it could not be not clearly profiled using lncRNA sequencing. Hence, in our study, we categorized lncRNAs as abundantly expressed if their expression levels were above zero and they occurred in over half of the total specimens.

### Examination and choice of prognostic associated lncRNA

The relationship between abundantly expressed lncRNA and overall survival of AML patients was evaluated using the training group. Our research selected target lncRNAs to find ones that are most closely associated with the clinical outcome to promote reliability and feasibility.

### Construction and validation of the risk score formula

A risk score system was constructed by enrolling every lncRNA associated with the clinical outcome in the training group. The risk score for every participant in the training group was evaluated according to the formula and, based on the result, the patients were categorized into high-risk or low-risk groups. The risk score system was additionally validated via fitting using the confirmation group and the complete group.

### Statistical analysis

We assessed the capacity of all the lncRNAs to predict event-free survival (EFS) of the patients via univariate Cox regression evaluation, which was examined between lncRNA expression displayed as $\log_2$ (X+1) and patient EFS as years. lncRNAs were regarded as significantly related to survival if $P<0.05$. The best lncRNAs were chosen by utilizing the minimal AIC criterion, which relies on robust likelihood-based survival modeling and was carried out with the help of R and Rbsurv packages [8,9]. Univariate analysis and multivariate analysis for overall survival were conducted for clinical variants and genetic mutations in all 139 cohorts. The chosen lncRNAs underwent subsequent Cox regression assessment and the risk score was built by evaluating the regression coefficients in the multivariate Cox regression and lncRNA expression. Median risk score was chosen in the training group in the form of a cut-off and, using that, the patients were categorized into a low-risk group and a high-risk group. Kaplan-Meier analysis was used to evaluate survival distinction between low-risk and high-risk patients in the training and confirmation groups. Specificity and sensitivity of lncRNA expression signature was assessed by calculating the area under the curve (AUC) of 5-year EFS. ROC, Cox regression analysis and Kaplan-Meier survival assessment was carried out using R statistical software (version 3.3.3) (Figure 1).

## Results

### Examination and selection of the lncRNAs associated with clinical outcome

We selected 7830 lncRNAs expressed in all AML patients from a pool containing 14 376 lncRNAs, as mentioned in our protocol. The remaining specimens were randomly divided into





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

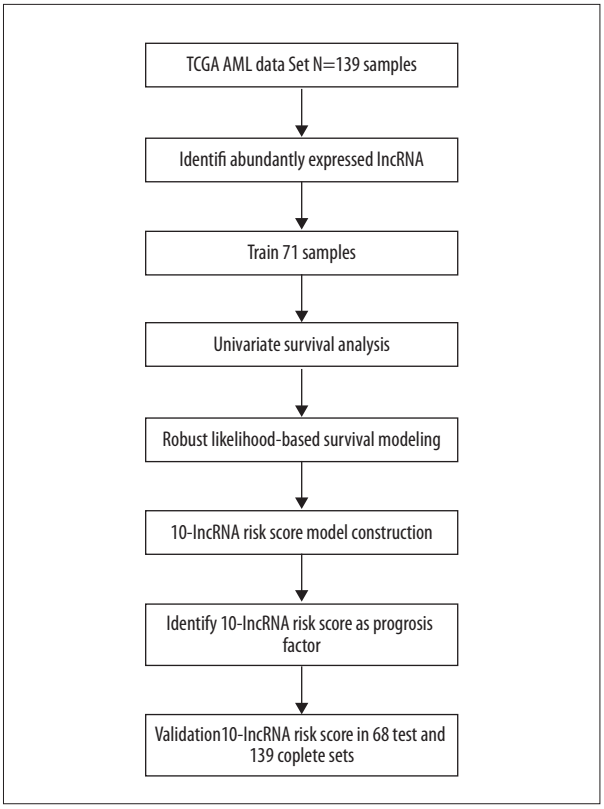


**Figure 1.** Flow chart of the protocol utilized in our study. AML – acute myeloid leukemia; TCGA – the Cancer Genome Atlas; ROC – receiver operating characteristics.

a training group and a testing group. Univariate Cox regression assessment of the training group revealed distinctly expressed lncRNA (all with $P<0.01$), which were later recognized as target lncRNAs. Finally, 10 lncRNAs associated with the clinical outcome were selected via robust likelihood-based survival analysis from the identified target lncRNAs (Table 1).

### Assessment of risk score based on the 10-lncRNAs signature and evaluation of the effective prognostic indicator for AML

We designed a 10-lncRNA signature risk system according to their Cox coefficients to better examine relationship between these 10 lncRNAs and the clinical outcome of AML:
*Risk score*=(–0.24476* expression value of AC004223.2)
+(–7e 05*expression value of AC067735.1)
+(0.36099 *expression value of DIRC3-AS1)
+(0.18098* expression value of AL355353.1)
+(–0.4052*expression value of AL645608.1)
+(0.57742* expression value of AC025430.1)
+(0.25425*expression value of AF064858.2)
+(0.31499*expression value of AL645608.5)
+(–0.20944* expression value of FP671120.3)
+(–0.3091*expression value of AC107398.3).

We subsequently examined the 10-lncRNA signature risk system of every participant in the training group and ranked them based on the risk scores. Patients were consequently categorized into high-risk and low-risk groups. AML survival was negatively related to risk scores (Figure 2A). To examine the specificity and sensitivity of the survival prediction, ROC assessment was carried out for the 10-lncRNA risk score. The majority of the cut-off points arrived at a precise classification and the AUC was 0.8765 (Figure 2B). The best cut-off point was chosen as -6.8066, which displayed the highest sensitivity and specificity. Participants were additionally categorized into high-risk (n=36) and low-risk (n=35) groups based on the best cut-off point. The Kaplan-Meier curve and log-rank test suggested a remarkable distinction in survival between risk groups (Figure 2C, $P<0.0001$).

### Survival prediction using the 10-lncRNA signature-based risk score

We verified the 10-lncRNA signature in the whole group and the testing group to ascertain our outcome of the risk system.

**Table 1.** Baseline information on the training group and testing group.

| Indicators | Training group | Testing group | P |
|---|---|---|---|
| Male (n) | 30 | 40 | 0.460 |
| Female (n) | 34 | 35 | |
| Age (years) | 31.7±8.34 | 32.3±7.95 | 0.824 |
| Height (cm) | 163.3±12.46 | 166.4±10.34 | 0.236 |
| Weight (kg) | 64.60±11.52 | 68.60±12.57 | 0.149 |
| BMI kg/m$^2$ | 24.67±4.43 | 26.17±4.72 | 0.532 |
| Diabetes (n) | 1 | 2 | 0.787 |
| Smoking (n) | 11 | 13 | 0.505 |

BMI – body mass index.





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

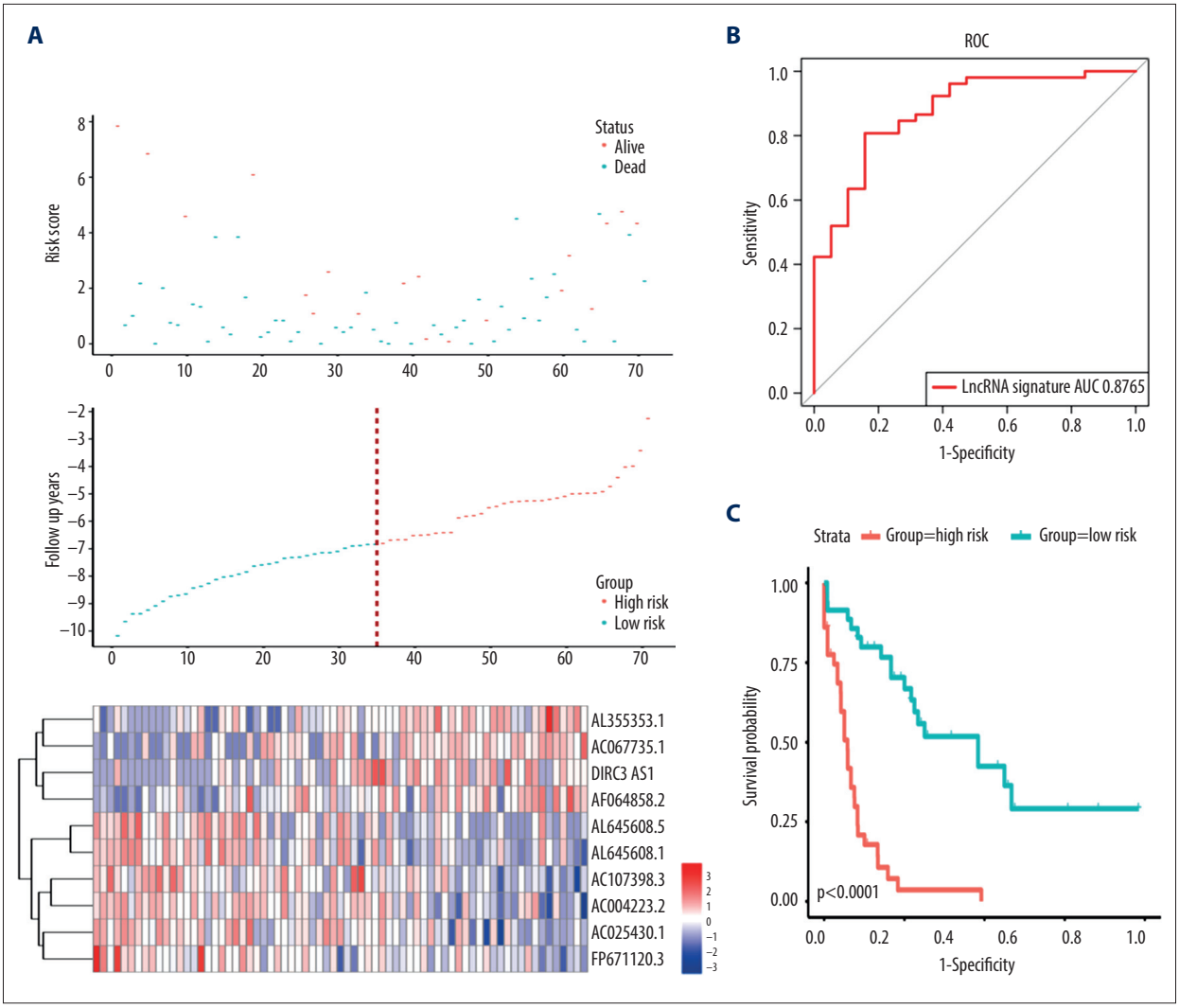


**Figure 2.** 10-LncRNA risk score assessment of the training group. (**A**) LncRNA signature risk score distribution, patients' event-free survival status, follow-up years (red point means high risk, blue point means low score) and a heat-map of the lncRNA expression profiles. Rows stand for lncRNAs, and columns stand for patients. (**B**) Receiver operating characteristic (ROC) analysis of the sensitivity and specificity of survival according to 10-lncRNA signature risk score. (**C**) Kaplan-Meier estimates of survival probability of patients from training group using the 10-lncRNA signature risk score (red line means high risk, blue line means low score).

Patients in the whole group were classified into high-risk (n=64) and low-risk (n=75) groups, with the best cut-off point chosen using identical formula. Consistent with our previous findings, Kaplan-Meier curves displayed a remarkably longer survival for the low-risk AML patients in comparison to the high-risk participants ($P$<0.0001) (Figure 3A). The separation of the testing group according to the best cut-off point resulted in a high-risk group containing 28 patients and a low-risk group consisting of 40 patients. Although the sample sizes were unequal, survival assessment displayed similar outcomes ($P$<0.05) (Figure 3B).

## Discussion

We selected 10 lncRNAs associated with AML survival – DIRC3-AS1, AC004223.2, AC067735.1, AL355353.1, AL645608.1, AC025430.1, AF064858.2, AL645608.5, FP671120.3, and AC107398.3 – from the training group. We generated a 10-lncRNA risk scoring system based on Cox coefficients. Furthermore, we revealed the best cut-off points to enable the categorization of patients as high risk and low risk using ROC assessment. In addition, not only did we establish a risk score formula, but also confirmed the best cut-off point in an internal examination group apart from the complete group. Patients suffering from AML with elevated risk score displayed





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

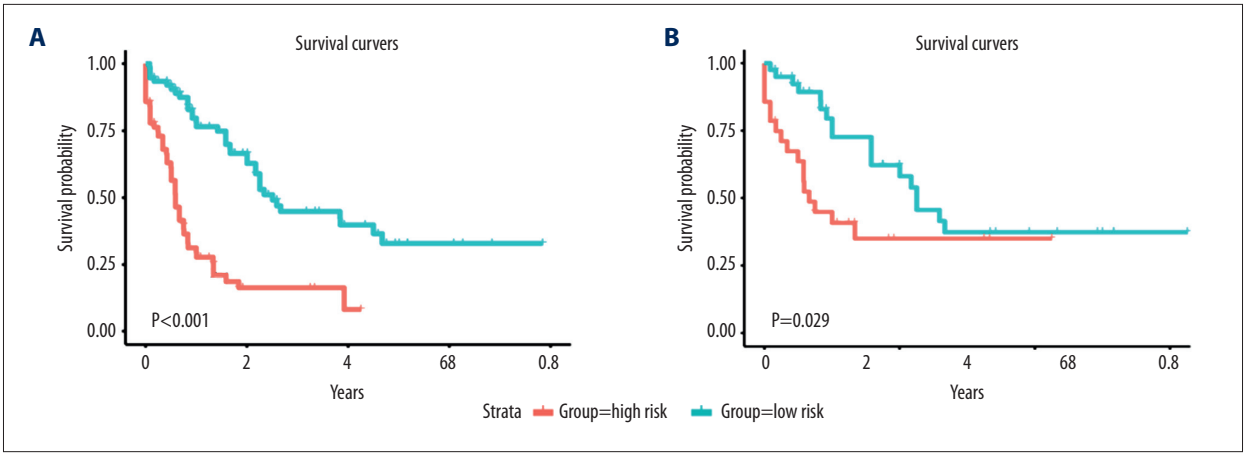


**Figure 3.** Confirmation of the 10-lncRNA signature risk score to predict survival. (**A**) Kaplan-Meier assessment of survival probability of patients from complete group using the 10-lncRNA signature risk score. (**B**) Kaplan-Meier assessment of survival probability of patients from the examination group adopting the 10-lncRNA signature risk score (red line means high risk, blue line means low score).

poorer survival and higher mortality. Our findings indicated that lncRNAs were essential contributors to etiology, development, and clinical outcome of AML.

It has been shown previously that several lncRNAs can predict the outcome of AML, including CRNDE and HOTAIRM1 [10,11]. However, these lncRNAs displayed a subtype specific expression pattern and were more appropriate for categorization of the disease state instead of the clinical outcome. In our study, we revealed 10 lncRNAs associated with clinical outcome with noticeably distinct expression in AML patients from the training group. Nevertheless, further studies are required to confirm the specific lncRNA used to predict clinical outcome. DIRC3. AS1 was the only real lncRNA out of the 10 that was not accepted widely. Our research proved its malfunction as well as its correlation with lower mortality and longer survival of patients with AML. We believe this is the first study to reveal its promising impact on malignancy outcome to our knowledge. Nevertheless, the understanding of the exact mechanism is inadequate and requires additional investigation.

The 10-lncRNA signature risk system was additionally generated relying on the lncRNAs associated with the clinical outcomes. A remarkable distinction was revealed in the survival curve between the high and the low scores, which apply to the risk score in the TCGA training group. Patients with elevated 10-lncRNA signature risk score are more likely to have shorter survival and poorer prognosis than those with low scores. Our research additionally stressed the importance of lncRNA-based risk scoring in malignancy outcome studies. Furthermore, this method circumvents the challenges in interpreting individual genes. Our research revealed the best cut-off point for various risk groups, producing remarkably distinct survival results. This cut-off value offered an innovative approach to assess patients and to predict clinical outcomes. We confirmed that the risk score system was reliably reproducible when used for predicting the survival outcomes of the complete group as well as the TCGA examination group. However, we could not assess the cause-effect link between AML outcome and the modeled risk score in this work. Consequently, further research is required to confirm the risk system using lncRNA expression signatures. As it is not enough to investigate the lncRNA signature of AML, more attention should be given to the evaluation of the 10-lncRNA signature.

## Conclusions

We examined 10 lncRNAs associated with the clinical outcomes of AML. The predictive target genes and biological activities of the lncRNAs provided additional insight into the contribution of lncRNAs in AML development. We also established a 10-lncRNA expression signature-reliant risk score system that can predict AML survival. We believe that this approach is a promising and innovative strategy for the prediction of clinical outcome and treatment of patients with AML.

### Conflict of interest

None.





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

## References:

1. Dohner H, Estey EH, Amadori S et al: Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood, 2010; 115: 453–74
2. Sill H, Olipitz W, Zebisch A et al: Therapy-related myeloid neoplasms: Pathobiology and clinical characteristics. Br J Pharmacol, 2011; 162: 792–805
3. Rollig C, Bornhauser M, Thiede C et al: Long-term prognosis of acute myeloid leukemia according to the new genetic risk classification of the European LeukemiaNet recommendations: Evaluation of the proposed reporting system. J Clin Oncol, 2011; 29: 2758–65
4. Zhu X, Tian X, Yu C et al: A long non-coding RNA signature to improve prognosis prediction of gastric cancer. Mol Cancer, 2016; 15: 60
5. Qi P, Du X: The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. Mod Pathol, 2013; 26: 155–65
6. Garzon R, Volinia S, Papaioannou D et al: Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci USA, 2014; 111: 18679–84
7. Wang P, Jin M, Sun CH et al: A three-lncRNA expression signature predicts survival in head and neck squamous cell carcinoma (HNSCC). Biosci Rep, 2018; 38(6): pii: BSR20181528
8. Renaud G, Stenzel U, Maricic T et al: deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. Bioinformatics, 2015; 31: 770–72
9. Wang JY, Tai JJ: Robust quantitative trait association tests in the parent-offspring triad design: Conditional likelihood-based approaches. Ann Hum Genet, 2009; 73: 231–44
10. Diaz-Beya M, Brunet S, Nomdedeu J et al: The lincRNA HOTAIRM1, located in the HOXA genomic region, is expressed in acute myeloid leukemia, impacts prognosis in patients in the intermediate-risk cytogenetic category, and is associated with a distinctive microRNA signature. Oncotarget, 2015; 6: 31613–27
11. Chen ZH, Wang WT, Huang W et al: The lncRNA HOTAIRM1 regulates the degradation of PML-RARA oncoprotein and myeloid cell differentiation by enhancing the autophagy pathway. Cell Death Differ, 2017; 24: 212–24





Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]