

## Parallel Patient Karyotype Information System using Multi-threads

Chantana CHANTRAPORNCHAI<sup>1,\*</sup>, Sathapat NAVAPANITCH<sup>1</sup>, Chidchanok CHOKSUCHAT<sup>2</sup>

<sup>1</sup> Department of Computer Engineering, Faculty of Engineering, Kasetsart University, 10900 Bangkok, Thailand

<sup>2</sup> Department of Computing, Faculty of Science, Silpakorn University, 73000 Nakorn Pathom, Thailand

E-mail(\*): fengcnc@ku.ac.th

\* Author to whom correspondence should be addressed; Tel.: +662-797-0999; Fax: +662-579-2775

Received: 15 July 2015 / Accepted: 30 August 2015 / Published online: 5 September 2015

### Abstract

Human cytogenetic data are the typical laboratory results from hospitals. Karyogram is used to show the chromosome characteristics. The characteristics are written as karyotype strings. For a particular patient, there may be many records of karyotype strings due to several visits. These data for many patients are increasingly large and must be stored properly for further investigation and analysis. This research introduces the information system for the hospital for keeping the karyotypes of patients and applies the parallel method for searching required karyotypes, extracting related patient information. Particularly, we exploit the technology of Node.js with multithreads while splitting queries to search in parallel. The search method is integrated to the cytogenetic information system which is aimed to use for studying karyotypes of leukemia patients.

**Keywords:** Cytogenetic; Karyotypes; Information system; Node.js; Multi-threading

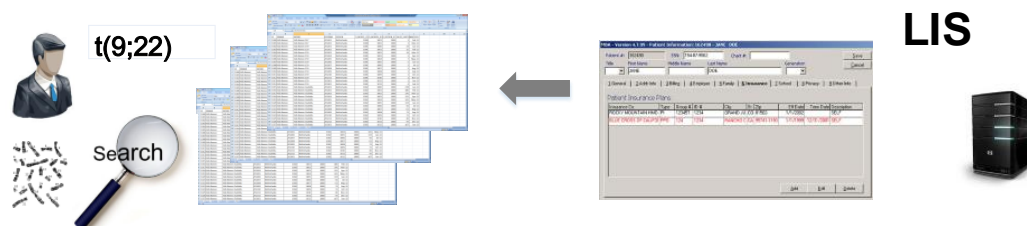
### Introduction

For a typical hospital in Thailand, patient data are a very large. They are stored based on the commercial software such as LIS ([http://www.clinfowiki.org/wiki/index.php/Laboratory\\_Information\\_Systems\\_%28LIS%29](http://www.clinfowiki.org/wiki/index.php/Laboratory_Information_Systems_%28LIS%29)). The representation of data is very proprietary since the software is commercial. For a patient visit, the patients' record has to be stored to the specific laboratories for certain test. The test results are submitted back to the hospital information system.

In one of the cytogenetic laboratory, the system such as Ikaros by Meta System Cooperation (<http://www.metasystems-international.com/ikaros>) is used to analyze the chromosome images. The diagnosis is given by the cytogenetic experts. Every year, the patients' laboratory test result files are exported from the hospital LIS to the experts' laboratory to keep the previous diagnosis records for future analysis.

It is necessary for the cytogenetic experts to keep these data since they have to use them to compare with new cases and analyze the possibility of aberrations. The analysis results need to be kept for follow-up and for tracking the states of patients. These data are accumulated and increasing every year. Also, the format of the data exported from the LIS keeps changing since the new version of LIS is consistently updated. The cytogenetic experts need to browse through each

Excel file to find the similar cases and the diagnosis records in Figure 1.



**Figure 1.** Process of finding typical karyotypes

In this work, a karyotype information system for patients with leukemia is developed. The goal is to gather existing information, store in the proper database, helping the cytogenetic experts to query by types of aberration, by specific string as well as to provide the ability to maintain the aberration types and patient records. Accumulating the data will give benefits to classification, predictions and future analysis.

#### *Previous Work*

Many works in medical informatics utilize database in the organization [1]. For example, Smith et al. developed a prototype of intranet electronic health records in Royal Sussex County hospital, U.K. Similarly, Joubert et al. provided the integration of medical information based on UMLS [2]. The system integrates information from PubMed, Medline, Intranet hospital and Faculty of Medicine of Nice (France). Haskew et al. applied the cloud-based Electronic Medical Record (EMR) in Western Kenya. They focused on the impact on the maternity care [3]. Alexandra et al. presented the web applications for disability to do the online survey tools [4].

Existing work related to cytogenetic data was around. Most of them are focusing on the standard database of aberration where the experts use to search for a kind of aberration such as Lucia which is a commercial website of Laboratory imaging in Czech Republic (<http://www.lucia.cz/en/front-page>). They presented the data about leukemia and presented the software to analyze karyotype features and can be searched according to the registered aberration.

Another example is an atlas of cytogenetic corpus, at <http://atlasgeneticsoncology.org/Anomalies/Anomliste.html>. This presents the standard aberration into 8 kinds. For a newer one, it will give links to the related publication that discovered it. Chromosomal Rearrangements In Diseases (dbCRID) (<http://dbcrd.biolead.org/>) is a general disease chromosome database. The data is accumulated from 1,172 sites.

Cydas is software for analyzing the karyotype strings (<http://www.cydas.org/>) [5]. The software provides web services for querying karyotyping using a regular expression and can draw ideograms. (<http://www.cydas.org/Docs/ISCNAnalyser/Analysis.html>) ISCN 1995 was used for the representation. It has a feature of data mining looking for frequent patterns and it is also linked to NCBI (<http://www.ncbi.nlm.nih.gov/>) or Ensemble (<http://www.ensembl.org/index.html>) for meaning for the abbreviation using CGI. The software was suitable for the study of cytogenetic. However, it was obsolete with the old development platform and old web service protocol due to the discontinuation of funding. Some part of the code is not downloadable such as about the web service. Thus, it is hard to modify from this software implementation.

Schatz et al. mentioned many interesting researches in bioinformatics clouds such as CloudBLAST ([http://ammatsun.acis.ufl.edu/amwiki/index.php/CloudBLAST\\_Project](http://ammatsun.acis.ufl.edu/amwiki/index.php/CloudBLAST_Project)) which uses a cluster of VMs with MPI and MapReduce to perform BLAST operations [6]. CloudBioLinux [7] (<http://gcid.jcvi.org/projects/gsc/CLOUD/docs/CHPC-Morning-Session-1.pdf>) presented a pre-configure VM with bioinformatics software packages. Dai et al. presented a survey in clouds for Bioinformatics, considering each layer of clouds and the example applications in each layer [8]. Emeakaroha et al. demonstrated the scientific workflow management on clouds and the application in bioinformatics [9]. Wagener et al. used Extensible Messaging and Presence Protocol (XMPP), for

asynchronous cloud service finding and invocation [10]. They tested on Bioclipse and Taverna as an example of chem-informatics and bioinformatics services. All the above works focused on genome sequencing [11], genome alignment, genome assembly, ([http://www.genomenetwork.org/resources/whats\\_a\\_genome/Chp2\\_1.shtml](http://www.genomenetwork.org/resources/whats_a_genome/Chp2_1.shtml)) [12] DNA, RNA, BLAST, etc. Most of these works are tested on Amazon EC2, and private clouds.

**Table 1.** Previous works classified in three categories: general information systems, cytogenetic search, and cloud platforms

System type	Characteristics
Database	Intranet electronic health records [1] Medical information integration [2] Cloud-based Electronic Medical Record (EMR) [3] Web applications for disability to do the online survey tools [4]
Cytogenetic data	Lucia software for leukemia chromosome analysis ( <a href="http://www.lucia.cz/en/front-page">http://www.lucia.cz/en/front-page</a> ) Atlas of cytogenetic corpus ( <a href="http://atlasgeneticsoncology.org/Anomalies/Anomliste.html">http://atlasgeneticsoncology.org/Anomalies/Anomliste.html</a> ) Chromosomal Rearrangements In Diseases (dbCRID) ( <a href="http://dbcrd.bioclead.org/">http://dbcrd.bioclead.org/</a> ) Cydas: karyotype string analysis ( <a href="http://www.cydas.org/">http://www.cydas.org/</a> ) [5]
Cloud-based platform	CloudBLAST: MPI and MapReduce for BLAST operations ( <a href="http://ammatsun.acis.ufl.edu/amwiki/index.php/CloudBLAST_Project">http://ammatsun.acis.ufl.edu/amwiki/index.php/CloudBLAST_Project</a> ) [6] CloudBioLinux, VM for bioinformatics software packages ( <a href="http://gcid.jcvi.org/projects/gsc/CLOUD/docs/CHPC-Morning-Session-1.pdf">http://gcid.jcvi.org/projects/gsc/CLOUD/docs/CHPC-Morning-Session-1.pdf</a> ) Scientific workflow management for bioinformatics [9] Extensible Messaging and Presence Protocol for chem-informatics and bioinformatics services [10]

In this work, the concepts of both karyotypes and database are utilized. The goal is to develop the prototype for searching cytogenetic data based on patients' karyotyping on case study on the leukemia patients. Unlike previous work which focuses on developing information systems about general medical records, or building the general karyotype database, our target system is the patient karyotype information. The system were built to serve in practices and to help accumulate aberration records for knowledge acquisition. The multi-thread search is employed to serve the growing number of data records. The project is being continued towards the semantic search of karyotypes with cloud-based services. NCBI or Ensemble web services will be integrated to give general information about specific karyotype.

## Material and Method

### Requirement Analysis

The case study hospital had a lot of laboratory departments and there is a central patient database maintained by outsourcing. The laboratory's results are included in LIS which is linked to the patient's record. When a patient needs the laboratory test, the physician issues the test request form. The laboratory expert performs the test and key in the diagnosis results.

For a cytogenetic laboratory, the cytogenetic experts conduct the laboratory test. The chromosomes are visualized and rearranged by the software so that they are in the format that is easy to recognize. Then, the experts observe the abnormal chromosomes. Three experts' opinions are needed to fully agree on the type of aberration. Then the karyotype strings and observation are

recorded along with the patient record. In this task, a lot of literatures may be reviewed since the experts need to consults cytogenetic encyclopedias or journal reports about the kind of aberrations. Importantly, they need to look up through the previous records about the similar kinds of aberration. Typically, the patient records are exported from the LIS system given to the department once a year in an Excel file. The experts need to look through these files in many aspects. For example,

1. Is there any prior aberration similar to this one?
2. Is this the first patient laboratory result or it is the result after some consecutive treatments?
3. Is this aberration in a group of common leukemia karyotypes?
4. Are there any relatives to the patients that have the similar symptoms?

If the previous laboratory results are kept in a database, it will be helpful to the experts to search the specific or pattern karyotypes of previous records, the patient names and relatives as well as the follow up records etc. Using the only string search in Excel is not sufficient due to many reasons. 1) Since data issued are from many years, it is very tedious to search through each file for given pattern. The file format may change every year; then, merging these files are difficult to do manually. 2) Also, the expert may define a type aberration, for example, the common aberration for Chronic Myelogenous Leukemia (CML) which is found to contain a reciprocal translocation between one chromosome 9 and one chromosome 22. This translocation is denoted designated  $t(9;22)$  (Shaffer et.al. 2013) in the karyotype ISCN representation [13]. For example,  $t(9;22)$  may be included in the string  $t(9;22)(q34;q11.2)$  which is the case for CML as well. 3) The physician may request the records of all previous CML patients. 4) On the other hands, if there are new commonly found aberrations leading for a particular symptom, the doctor may get a new implication from time to time. 5) The patient treatment results can be viewed by the changing of karyotypes between the given periods where the treatment is applied.

### Application Design

To serve the main purpose, first the database is designed to store the exported patient data. The database contains the patient investigation records in Figure 2. Also, the diagnosis data keeps the common karyotype strings as they relate to the doctor's comments. The imported file is in the CSV format given by MS Excel. The first line in the file contains the keywords used to each field in Table *patient\_record* for importing. These are fields such as KARYOTYPE, DIAG, COMMENT, SID, AGE, YEAR, SEX, DR, TESTNAME, etc. The import menu is in Figure 3.

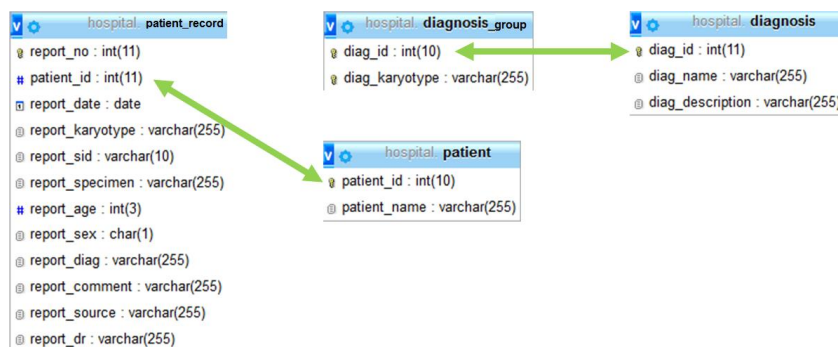


Figure 2. Database storing patients' records

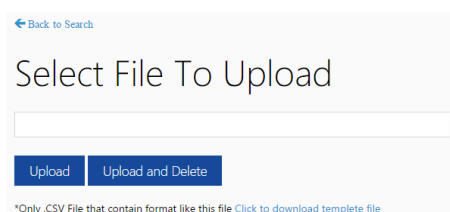
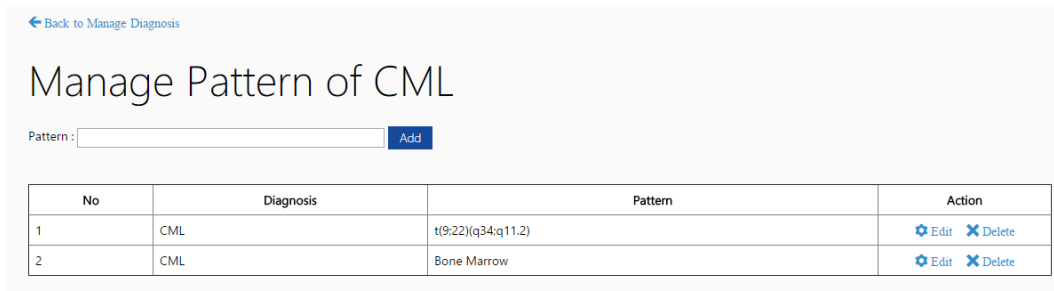


Figure 3. Import patients' data

The user interface presents the case where the user can select to find multiple groups of diseases by checkbox or finding the karyotype using keywords in (1). Also, selecting CML checkbox, referred to the karyotype string containing t(9;22)(q34,q11.2). Acute Nonlymphocytic Leukemia (AML, ANLL), Acute Lymphocytic Leukemia (ALL) etc. can be identified by several strings such as in Table 2. The other kinds of diagnosis grouping can be added (in Figure 4) later such as Acute Promyelocytic Leukemia, Myelodysplastic Syndrome, Myelodysplastic Diseases, Myeloproliferative Neoplasms, or in the genetic disorder, prenatal kind: Global Delayed Development(GDD), Autism Spectrum Disorder (ASD), Premature Ovarian Failure (POF) etc.

**Table 2.** Example aberration diagnosis

AML,ANLL	ALL
t(1;3)(p36.1;q21)	t(1;11)(p32;q23)
t(3;21)(q26.2,q22)	t(1;14)(p32;q11.2)
t(6;9)(p23;q34)	t(1;14)(p33;q11.2)
t(7;11)(p15;p15)	t(1;14)(p34;q11.2)
t(8;21)(q22;q22)	t(1;19)(q23;p13.3)
t(9;11)(p21;q23)	t(4;11)(q21;p15)
t(9;11)(p22;q23)	t(4;11)(q21;q23)
t(10;11)(p11-15;q13-23)	...
t(10;17)(q23;q21)	
t(11;17)(q23;q21)	
...	



**Figure 4.** New grouping of aberration

Each group of diagnosis is kept along with the strings in the database. For instance, CML is kept in the filed `diag_name` (Table diagnosis in Figure 2), and its strings are kept in `diag_karyotypes` (Table diagnosis\_group). If there are many karyotypes, they are concatenated with the separators (e.g. t(1;3)(p36.1;q21)# t(3;21)(q26.2,q22)#...). When querying by groups, the `diag_karyotypes` are split into subqueries to be search independently (as described in the next section).

Figure 5 presents the user interface where the user can select to search by types such as CML, AML etc. It will search the whole karyotype strings which are defined by Figure 4.

*Parallel Search Query*

Node.js® (<https://nodejs.org/>) is an event-driven, non-blocking I/O platform suitable for data-oriented applications and real-time applications. It is based on JavaScript runtime and can be applied to many device’s platforms. It is used to create the web servers and supports many networking protocols and security functions. The important characteristic is the non-blocking feature which allows the parallel execution and call-back approaches for asynchronous communications. Thus, it is possible to create concurrent server applications with the event-driven programming model.

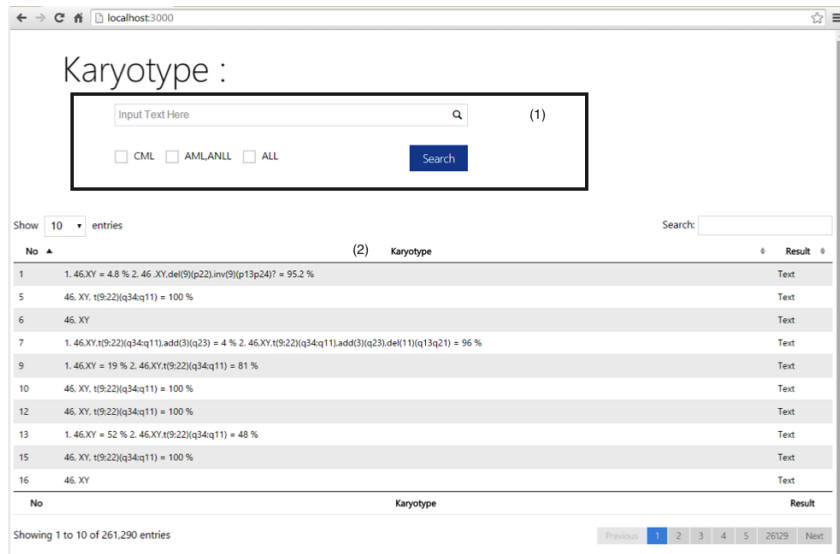


Figure 5. Main user interface

In Figure 6, the client makes a HTTP request to Node.js where Node.js can support multithreading. Node.js extracts rows from database, formats the results in JSON and sends back to the client.

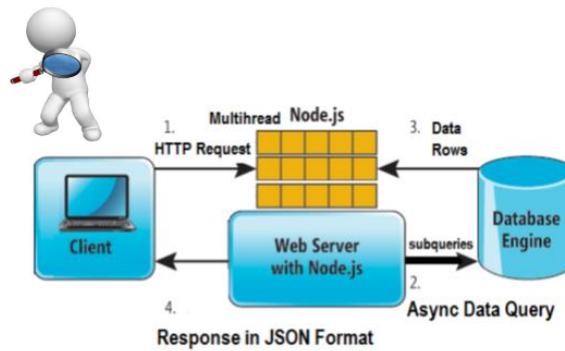


Figure 6. Workflow of query using Node.js

Multithreading is a necessary concept nowadays in modern computers especially running on a multi-core processor. This multicore processor has two or more actual processing elements which can execute program instructions in parallel. Multiple instructions can be executed at the same time with multithreading programs. The program spawns threads which are executed in a real hardware core. With the current version of Node.js (<https://nodejs.org/api/cluster.html>), it is possible to utilize the cores by launching a cluster of node processes.

In the implementation, many threads of Node.js are running on the multicore using Node.js cluster module or jxcore Library (<http://jxcore.com/home/>). The threads are created to service the user query. The user query may contain the many subqueries such as

```
SELECT      karyotype , report_no
FROM        patient_record
WHERE       karyotype LIKE '% {Pattern[1]} %' or
           karyotype LIKE '% {Pattern[2]} %' or
           karyotype LIKE '% {Pattern[3]} %'
```

which can be split into 3 subqueries for each pattern.

The threads are spawned for each subquery and the results are gathered back as in Figure 7. The results are gathered back and displayed. They are merged to eliminate the redundant records.

Pagination is used for viewing to get only limited number of records, saving load times while using Ajax plugin to load subsequent results.

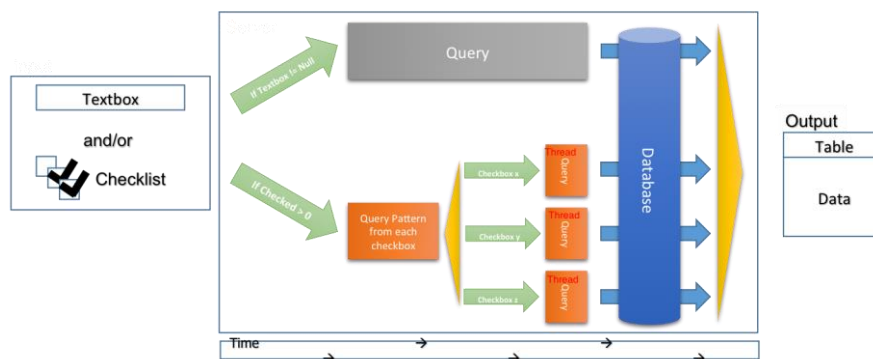


Figure 7. Overall Process

## Results and Discussion

The sample cytogenetic data from one of the hospital in Thailand is collected. The data is collected annually and is aggregated thousands of records where each record contains about 3,000 characters. The database to store the data is designed using MySQL. The raw data in Excel is imported and cleaned. First, the performance of multithreading query of Node.js is tested; then the applications user feedback and the revised version are presented.

### Performance

The test machine is Intel i5 4460 @3.20GHz 4 Core with 8 GB RAM, and HDD 1TB 7200 RPM SATA3. The execution time of queries for this data set is measured. It has 530,188 records in total. The sample query yields the found results of 59,160 records. Figure 8(a) shows the execution time of the query. The X-axis is the number of queries. The baseline is the execution time of when the queries are executed in sequence. The multi-query is the part where the connection pool with Node.js and asynchronous parallel modules are used. For the baseline, when the number of queries is double, the execution time is double and when the number of queries is triple, the execution time is triple. When the number of queries increases from 1 to 3, the execution is not increased. The execution time increases only by 70% when the number of queries increases from 3 to 6. For the other cases, the number of queries is double, the execute time is also double. Thus, the best case of these tests are about running 3 – 6 queries, giving us the better saving in execution time, not triple or double execution time for the 4-core test hardware. The overall speedup from the baseline is about 3.5 in Figure 8(b). This shows the advantage of using Node.js on the multicore machine.

### Integrations from User Comments

The application is installed at the cytogenetic laboratory in a given hospital. The users are the experts in such a laboratory. The user comments are as follows.

1. The cytogenetic experts like to view the patients' records based on specific fields beside karyotypes such as by age, sex, patient ID, patients' last name, etc.
2. It is important to view the record of the patients especially in the first visit only in some case.
3. Sometimes, the experts may want to look up the specific karyotype rather than the karyotype that contains the specific string.
4. The records found need to be exported into an Excel file.
5. For the patients' data, there are three kinds of results, based on the test name e.g, Genetic disorder analysis (LY), Leukemia Chromosome analysis (BM), Prenatal chromosome analysis (AF). It is sometimes required to look up the data based on the test name as well.

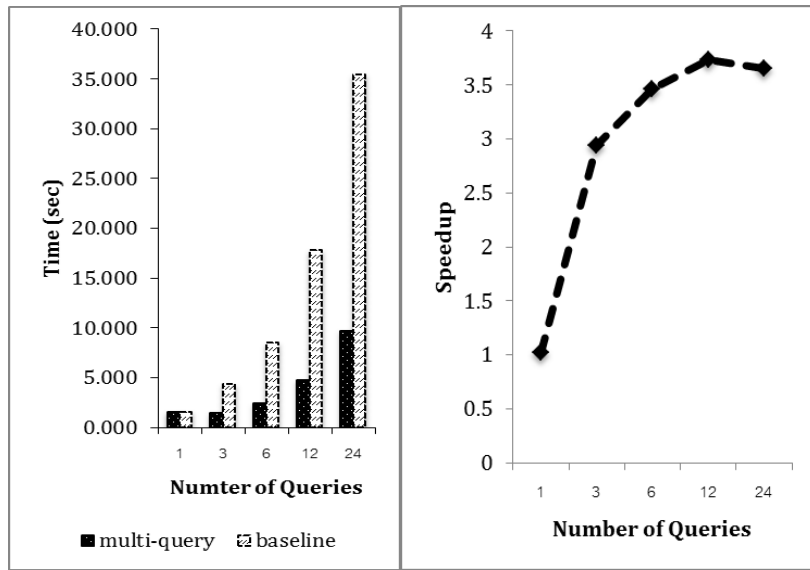


Figure 8. Timing results (a) Comparison to baseline (b) Speedup

This additional features help to search based on patients’ details such as searching by last name to find the relatives or family member previously visited having the similar symptoms. Also, searching by the earliest date makes it possible to find the first record of the patients’ visit.

The revised user interface is shown in Figure 9. At (1), the experts can select to match an exact string or match only substring of karyotypes. The patient records may be searched by dates (2), by gender (3) or by age (4). It is possible to search only specific test with specific type at (5). Figure 10 is the displayed results which can be sorted by columns. The results can be exported to the CSV file.

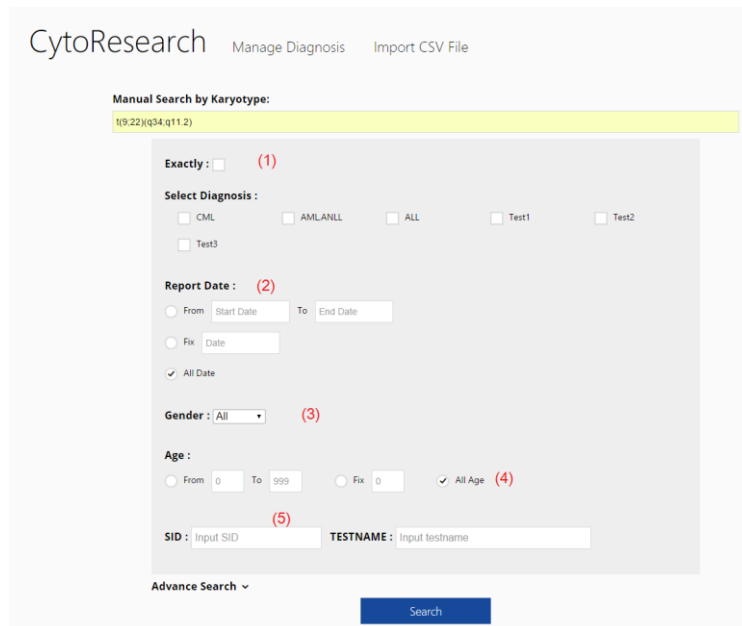


Figure 9. The revised user interface: (1) to match an exact string or match only substring of karyotypes; (2) date of the report; (3) gender; (4) age



CytoResearch Manage Diagnosis Import CSV File

Manual Search by Karyotype:

Advance Search

Display 20 records Filter:

DATE	HN	NAME	AGE	GEN	KARYOTYPE REPORT
01-06-2006	3914873	fpupP	34	Female	Staining : GTG BPHS Approx. 400 level_x000D_Analyzed : 46.XY,t(9;22)(q34;q11.2) [18] / 46.XX,t(9;22)(q34;q11.2) 46.XX,t(9;22)(q34;q11.2) = 10 cells_x000D_Interpretation : 46.XX,t(9;22)(q34;q11.2) / 46.XX,t(9;22)(q34;q11.2)
25-07-2006	0	No Patient Name	38	Male	Staining : GTG BPHS Approx. 400 level_x000D_Analyzed : 46.XY,t(9;22)(q34;q11.2) [13] / 47.idem,+der(22)t(9;22) cells / 47.idem,+der(22)t(9;22) = 8 cells_x000D_Interpretation : 46.XY,t(9;22)(q34;q11.2) / 47
21-09-2006	4024667	sBCzf	22	Male	Staining : GTG BPHS Approx. 300-400 level_x000D_Analyzed : 46.XY,t(9;22)(q34;q11.2) [14] ( 14 metaphases on 9 cell_x000D_Interpretation : 46.XY,t(9;22)(q34;q11.2) = 100 %
06-10-2006	4030157	tMfjf	44	Male	Staining : GTG BPHS Approx. 300 level_x000D_Analyzed : 46.XY,t(9;22)(q34;q11.2) [2] / 47.idem,+der(22)t(9;22) Counted : 46.XY,t(9;22)(q34;q11.2) = 1 cell / 47.idem,+der(22)t(9;22) = 3 cells / 46.XY = 1cell_x00
12-12-2006	3971727	UgInf	38	Female	Staining : GTG BPHS Approx. 400 level_x000D_Analyzed : 46.XX,t(9;22)(q34;q11.2) [20]_x000D_Counted : 46.XX Interpretation : 46.XX,t(9;22)(q34;q11.2) = 100 %
21-02-2007	4062105	OuhBT	49	Male	Staining : GTG BPHS Approx. 400 - 550 level_x000D_Analyzed : 46.XY,t(9;22)(q34;q11.2) [20]_x000D_Counted : Interpretation : 46.XY,t(9;22)(q34;q11.2) = 100 %
13-03-2007	0	No Patient Name	45	Female	Staining : GTG BPHS Approx. 400 level_x000D_Analyzed : 46.XX,t(9;22)(q34;q11.2) [20]_x000D_Counted : 46.XX : 46.XX,t(9;22)(q34;q11.2) = 100 %
22-06-2007	4062105	OuhBT	49	Male	Staining : GTG BPHS Approx. 400 level_x000D_Analyzed : 46.XY,t(9;22)(q34;q11.2) [20]_x000D_Counted : 46.XY Interpretation : 46.XY,t(9;22)(q34;q11.2) = 100 %
14-09-2007	4030157	tMfjf	0	Male	Staining : GTG BPHS Approx. 300-400 level_x000D_Analyzed : 45.XY,t(11;11)(q25;q21).add(6)(q25).del(7)(p15).t(16;-17,+mar4_x000D_45.XY,del(4)(q21q21).del(7)(p15).t(7;12)(p15;p13).t(9;22)(q34;q11.2).add(9)_x000D_(p

Showing 1 to 20 of 500 records

Figure 10. The displayed results

## Concluding Remarks

This work is aimed at helping cytogenetic experts to search for karyotype patient information. Current karyotype diagnosis records are scattered and the formats are different/varied. Collecting them properly makes it easy to maintain while gives convenience to the experts to find previous aberration records, and further analyze new cases.

Cytogenetic patient information system has been developed to serve the work in the cytogenetic laboratory. The patient diagnosis data are exported from the LIS and collected in a database. The database is designed so that the exported data can be mapped directly to its fields. The users can perform the search on specific karyotypes or groups of karyotypes by aberration. The types of aberration can be added and updated since new abnormal karyotype may occur every day. The physician can query to follow up records of patient's karyotypes to see improvement of the treatment. The search engine supports multithreading to serve the case where data is large. Current implementation records data to serve the leukemia karyotype diagnosis. The system is currently serviced at cytogenetic laboratory after several modifications. Related patient information can be retrieved by report dates, sex, names and others. The linked to the literature of each aberration based on NCBI is the future work to aid the experts finding new aberration information. The extension to derive the semantic search of aberration using ontology is also developed.

## List of abbreviations

LIS	Laboratory Information System
Ikaros	Commercial software to arrange chromosome in ideograms ( <a href="http://www.metasystems-international.com/ikaros">http://www.metasystems-international.com/ikaros</a> )
EMR	Electronic Medical Record
CGI	Common Gateway Interface
XMPP	Extensible Messaging and Presence Protocol
ANLL	Acute nonlymphocytic leukemia
ALL	Acute lymphocytic leukemia

CML	Chronic myeloid leukemia
AML	Acute Myeloid Leukemia
ALL	Acute lymphoblastic leukemia
HTTP	Hypertext Transfer Protocol

### Conflict of Interest

The authors declare that they have no conflict of interest.

### Acknowledgements

This work is supported in part by Kasetsart University Research and Development Institute. We are also thankful to Assoc. Prof. Dr. Budsaba Rerkamnuaychoke, Human Genetic Laboratory, Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, for the test data and user experience testing in the laboratory.

### References

1. Smith, RA, Arvanitis TN, English M, Vincent R. An intranet database for pacemaker patients. *Int J Med Inform* 1997;47:79-82.
2. Joubert M, Aymard S, Fieschi D, Volot F, Staccini P, Robert JJ, et al. ARIANE: Integration of Information Databases within a Hospital Intranet. *Int J Med Inform* 1998;49:297-309.
3. Haskew J, Røa G, Saitoc K, Turner KK, Odhiambo G, Wamaee A, et al. Implementation of a Cloud-Based Electronic Medical Record for Maternal and Child Health in Rural Kenya. *Int J Med Inform* 2015;84:349-354.
4. Cojan-Carlea NA, Cojocaru MC, Popa F, Berteanu M. Developing an Online Web Application for Disability Evaluation on a Medical, Social and Computerized Level. *Appl Med Inform* 2015;36(1):45-50.
5. Hiller B, Bradtke J, Balz H, Rieder H. CyDAS: a Cytogenetic Data Analysis System. *Bioinformatics* 2005;21(7):1282-1283.
6. Schatz MC, Langmead B, Salzberg SL. Cloud Computing and the DNA Data Race. *Nature Biotechnology* 2010;28:691-693.
7. Krampis K, Booth T, Chapman B, Tiwari B, Bica M, Field D, et al. Cloud BioLinux: Pre-configured and On-Demand Bioinformatics Computing for the Genomics community. *BMC Bioinformatics* 2012;13:42. doi:10.1186/1471-2105-13-42.
8. Dai L, Gao X, Guo Y, Xiao J, Zhang Z. Bioinformatics Clouds for Big Data Manipulation. 2012 Nov 28. *Biol Direct* 2012;7:43. doi:10.1186/1745-6150-7-43.
9. Emeakaroha VC, Maurer M, Stern P, Labaj PP, Brandic I, Kreil DP. Managing and Optimizing Bioinformatics Workflows for Data Analysis in Clouds. *J Grid Computing* 2013;11(3):407-428.
10. Wagener J, Spjuth O, Willighagen EL, Wikberg JE. XMPP for Cloud Computing in Bioinformatics Supporting Discovery and Invocation of Asynchronous Web Services. *BMC Bioinformatics* 2009;10:279. doi: 10.1186/1471-2105-10-279.
11. Genome sequencing [Internet]. Genome, News Network, 2004; [updated 2004; cited 2015 July]. Available from: [http://www.genomenewsnetwork.org/resources/whats\\_a\\_genome/Chp2\\_1.shtml](http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp2_1.shtml).
12. Krampis N. Cloud Computing Solutions for Genomics Across Geographic, Institutional and Economic Barriers. [Internet] Available from: <http://gcid.jcvi.org/projects/gsc/CLOUD/docs/CHPC-Morning-Session-1.pdf>. [cited 2015 July]
13. Shaffer LG, McGowan-Jordan, J, Schmid M. eds. *ISCN 2013: An International System for Human Cytogenetic Nomenclature*. Karger Medical and Scientific; 2013.