31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Developing a genomic-based point-of-care diagnostic system for rheumatoid arthritis and multiple sclerosis

Fanis G. Kalatzis, Nikolaos Giannakeas, Themis P. Exarchos, *Member, IEEE*, Leandro Lorenzelli, Andrea Adami, Massimiliano Decarli, Sara Lupoli, Fabio Macciardi, Sofia Markoula, Ioannis Georgiou and Dimitrios I. Fotiadis, *Senior Member, IEEE*

Abstract— **In this paper the methodology of designing a genomic-based point-of-care diagnostic system composed of a microfluidic Lab-On-Chip, algorithms for microarray image information extraction and knowledge modeling of clinico-genomic patient data is presented. The data are processed by genome wide association studies for two complex diseases: rheumatoid arthritis and multiple sclerosis. Respecting current technological limitations of autonomous molecular-based Lab-On-Chip systems the approach proposed in this work aims to enhance the diagnostic accuracy of the miniaturized LOC system. By providing a decision support system based on the data mining technologies, a robust portable integrated point-of-care diagnostic assay will be implemented. Initially, the gene discovery process is described followed by the detection of the most informative SNPs associated with the diseases. The clinical data and the selected associated SNPs are modeled using data mining techniques to allow the knowledge modeling framework to provide the diagnosis for new patients performing the point-of-care examination. The microfluidic LOC device supplies the diagnostic component of the platform with a set of SNPs associated with the diseases and the ruled-based decision support system combines this genomic information with the clinical data of the patient to outcome the final diagnostic result.**

I. INTRODUCTION

THE evolution of point-of-care diagnostics during the 21st century is going to be motivated by the advancements of information and communication technologies, microfluidics, microelectronics and genome wide association studies. Being composed of miniaturized devices, the point-of-care diagnostics of the next generation will focus on the early prognosis and diagnosis of many complex diseases. Portable devices capable to monitor the blood glucose concentration are known since 1970s [1]. In the 90s in-vitro point-of-care (POC) devices were born with main characteristic the usage at the bedside of a patient [2]. During the last years POC molecular diagnostic systems are implemented, targeting areas of genetic testing, pharmacogenomics and infectious diseases [3]. Nowadays, there are challenges which should be faced towards the development of molecular-based POC platforms although the need of such miniaturized Lab-On-Chip (LOC) systems is considered the most suitable and appropriate technology for portable POC systems [4]. According to a recent study [5] the micro-meters technology of the LOC systems followed by the advancement of nano-technology and microelectronics aims to bridge the gaps for the future diagnostic POC systems.

Modern point-of-care diagnostic systems will be supported by artificial intelligent algorithms in order to provide more accurate and efficient diagnostic results at the bedside of the patient. These portable systems are going to make exhaustive usage of "smart" algorithms in order to enhance the point-of-care diagnosis through the automated application of decision support processes; such combined diagnostic assays will assist the primary care community to administer the patient symptoms at any point of need. Decision support process is a crucial component of POC systems since they can assist novice practitioners and nurses in their diagnostic and clinical judgment [6]. Although there are limitations for the development of autonomous genomic-based POC systems, the information and communication technological achievements of the last years can efficiently provide reliable diagnostic results at the POC; it is required the integration of current LOC technologies with decision support systems operated by handheld or portable devices.

In this paper the methodology for the implementation of a genomic-based LOC system aiming at the early prognosis and diagnosis of the Rheumatoid Arthritis (RA) and Multiple Sclerosis (MS) diseases at the POC is presented. In parallel with the microfludic LOC sensors the integrated system is

Manuscript received April 7, 2009. This work is funded by the European Commission (project POCEMON: Point-Of-Care MONitoring and Diagnostics for Autoimmune Diseases FP7-ICT-2007-216088).

F. G. Kalatzis, N. Giannakeas, T. P. Exarchos and D. I. Fotiadis* are with the Unit of Medical Technology and Intelligent Information Systems, Dept. of Material Science & Engineering, University of Ioannina, GR-45110, Ioannina, Greece (e-mail: tkalatz@cc.uoi.gr, me01310@cc.uoi.gr, exarchos@cc.uoi.gr, *Corresponding author; phone: 0030-26510-98803; fax: 0030-26510-97092; e-mail: fotiadis@cs.uoi.gr).

L. Lorenzelli, A. Adami and M. Decarli are with the Fondazione Bruno Kessler, Materials and Microsystems Area - BioMEMS Research Unit, Via Sommarive 18, 38050 Trento, Italy (email: lorenzel@fbk.eu, andadami@fbk.eu, mdecarli@fbk.eu).

S. Lupoli is with the University of Milan c/o Filarete Foundation, Viale Ortles, 22/4 - 20139 Milan Italy (e-mail: lupoli.sara@hsr.it, fabio.macciardi@unimi.it).

F. Macciardi is with the Dept of Science & Biomedical Technologies - School of Medicine, University of Milan via F.lli Cervi, 93 - 20090 Segrate (Milano) - Italy (e-mail: fabio.macciardi@unimi.it).

S. Markoula is with the Medical School, University of Ioannina, GR-45110, Ioannina, Greece (e-mail: smarkoula@grads.uoi.gr).

I. Georgiou is with the Genetics Unit, Dept of Obstetrics and Gynaecology, Medical School, GR-45110, Ioannina, Greece. (e-mail: igeorgiou@cc.uoi.gr).

also composed of genomic information extraction, disease-gene association, genotypic and phenotypic knowledge modeling which are the main aspects discussed in this work.

II. METHODOLOGY OF SYSTEM DESIGN

The key components of the methodology, which are presented in the forthcoming sections, are: (a) the gene discovery and association, (b) the algorithms for information extraction, (c) the modeling of the knowledge and (d) the fabrication and realization of the microfludic LOC device to analyze the genomic material at the POC. The hardware and software platform which will combine the aforementioned components will provide the portable point-of-care RA and MS diagnostic assay.

A. Gene Discovery and Gene Association

During the discovery phase, a Whole Genome Association Study on cases (patients with the disease) and controls (patients without the diseases) to identify HLA (Human Leukocyte Antigens) system and other potentially relevant susceptibility genes, related to the RA and MS diseases is performed. The discovery phase of RA applied to a homogeneous North-European population (with 800 cases and a similar number of controls). The discovery followed by a confirmatory phase, where the "best" SNPs (Single Nucleotide Polymorphisms) are evaluated in a second independent and much larger sample (2000 cases affected by RA or MS respectively and a similar number of controls), from three separate cohorts representing Northern, Central and Southern European Populations.

The discovery genotyping phase for RA, using Illumina HumanCNV-370 (Illumina, San Diego, USA), is concluded [6]. Plink [7] is used for Quality Control (QC) of genotyping data, single marker association analysis and correction using permutations. Although the analysis is ongoing, the first results exhibit a strong positive association with the HLA region of chromosome 6. Initial findings pointed also to several other susceptibility genes across the genome and potentially involved in the susceptibility to the disorder other than confirming the role of HLA. Along with the genetic material (i.e. DNA), clinical data are collected from cases allowing the development of the knowledge modeling component of the POC system. The "best" SNPs that will result from the discovery and confirmatory phases will be used as genomic diagnostic markers in the miniaturized LOC device, thus representing an essential component of the diagnostic system.

B. Algorithms for Information Extraction

The data produced by the gene discovery phase are analyzed in order to provide information which can be used to relate the genomic characteristics (SNPs) of the disease with the phenotypic data of a patient, performing the point-of-care diagnostic test. More specifically, the input of this part of the system is the microarray images generated by the scanning of the Illumina Beadchips [8]. The process produces numerical data as measurements of the hybridization of the SNPs. An automated software component has been developed to automatically perform the following microarray image analysis procedures: Spot Addressing/Gridding, Segmentation and intensity extraction.

Spot addressing and gridding is a procedure where each individual microarray spot of the image must be isolated and our interest is concentrated on an automated procedure which is applied to all the produced images. The main characteristic of the generated image is the hexagonal grid [9] where the spots of the image are located. Our approach consists of four automated steps of block finding, hybridized spot detection, non-hybridized spot detection and gridding.

The block finding algorithm is based on a holistic approach where the elements of each row and column of the image are summed resulting in the projections of the image in the vertical and the horizontal direction. These projections of the image are processed in order to detect the points of the image that separate the blocks. For this reason, a median filtering is employed and then the gradient of the image is extracted. The split points maximize the differences of the gradient. As a result, the image is split into a number of sub-images which contain each separated block. Each block of

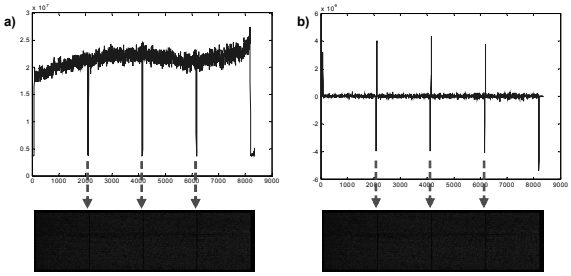

Fig. 1. a) The vertical projection of the image, b) the gradient of the vertical image projection.

the image is processed separately. Figure 1 shows the vertical projection and its gradient of an Illumina microarray image which contains four blocks.

The hybridized spot finding algorithm detects all the objects of each block. This procedure converts the image to a binary one using the Otsu method [10] and then the centre of the mass is estimated for each 8-connected object in the block.

To find the positions of the "empty" spots (i.e. the non-hybridized spots) an new algorithm is used. The algorithm of Growing Concentric Hexagon (GCH) start from an already detected spot of the image and a hexagon is grown around it, estimating the position of each empty spot on its contour. The steps of the algorithm are:

1. Select randomly a spot of the image
2. Grow the hexagonal region around this spot until a neighbour spot is found
3. Compute the mean distance between the detected neighbouring spots and the central spot

4. Generate a hexagon with a radius equal to the above distance. Six spots are allocated on the angles of the hexagon, and a number of spots on the edges according to the below formula:

$$\#of\ spots = 6 \times (index\_of\_hex - 1) + 6, \quad (1)$$

where index_of_hex defines the level of the hexagon around the centre spot.

5. Examine the generated spots of the image. If there is already a hybridized spot in the 3x3 neighbourhood of pixels, the generated spot is eliminated and the real hybridized spot remains.

Once the position of all the spots of the image is estimated, a Voronoi diagram [11] is employed. As a result, each Voronoi cell contains only one spot.

To segment each Voronoi cell, the K-means algorithm [12] is applied for each pixel. Thus, the signal and the background pixel in the area are separated. A set of features is used to feed the Clustering algorithm [13]. The background-corrected values for each spot of the image are computed for both the green and the red channel and the ratio of the hybridization for the two samples is calculated.

C. Modeling of the Knowledge

A key component of the point-of-care diagnostic system concerns the modeling of the knowledge coming from clinico-genomic patient data. These data were collected during the genome wide association studies from RA and MS cases and controls. The purpose of this step is two-fold: a) to extract new and potentially useful knowledge and b) provide prediction methods for the diagnosis as well as the detection of susceptibility for MS and RA diseases. In order to extract new and potentially useful knowledge, cluster analysis [14] and association rule mining techniques [15] are applied to the collected datasets. In order to create predictive models for the diagnosis and the detection of susceptibility of the RA & MS diseases, classification techniques [16] are applied.

The architecture for the modelling of knowledge of the system is shown in Fig 2. Starting from collected and analyzed clinico-genomic data, an important part of the architecture concerns the informative feature selection process. The selection of the most informative clinical features and SNPs is based on mathematic measures like the correlation of the features with the target outcome as well as the consistency of the informative set of features with the target outcome. Feature selection is used: (a) before clustering in order to keep only the important features for developing useful profiles, (b) before association rule mining in order to restrict the rules only to those that have as antecedent of consequent the informative features and (c) before the classification, in order to build more accurate and less complicated prediction models for MS and RA.

Cluster analysis is an unsupervised procedure, meaning that the outcome (e.g. MS or RA, or normal) is not used as input in the clustering procedure. During the clustering procedure, records of patients which are close in terms of a Euclidian distance are grouped together into the same cluster. Algorithms for this task include the k-means algorithm, the EM algorithm and the fuzzy c-means algorithm. The identified profiles are stored in the knowledge repository. Association rule mining aims at discovering hidden knowledge within the data in the form of rules. The evaluation of rules is performed using the support confidence framework and only rules with high support and confidence are considered by the experts for storing them in the knowledge repository. For the extraction of rules, the well known a-priori and FP-growth algorithms are employed [17].

Classification is one of the most important functionalities of the knowledge modeling framework, since the development of classification models, will constitute the prediction models for MS and RA. These models will be finally used to provide predictions for new patients performing an examination at the point-of-care. Extensive testing is performed in order to identify algorithms with high prediction accuracy. Our preliminary studies show that a combination of algorithms (Artificial Neural Networks, Decision Trees and Support Vector Machines) that produces highly accurate results accompanied with interpretation of the decisions is the best for the needs of the diagnostic point-of-care platform.

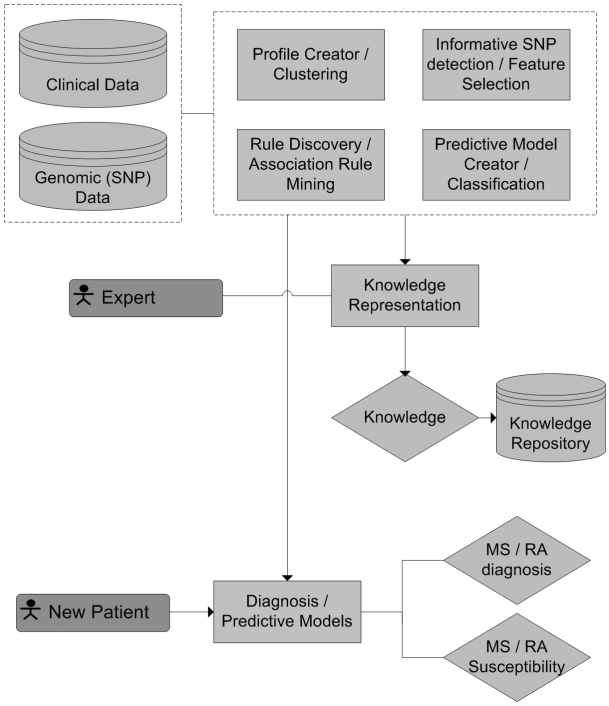

Fig. 2. The architecture of knowledge modeling component.

D. Fabrication of the microfluidic LOC device

In the literature, several works have been devoted to the realisation of integrated LOC devices for DNA analysis. A recent review on the topic [18] shows the state of the art with

respect to commercial or semi-commercial systems for point of care diagnosis.

The first part, namely the microfluidic module, includes the microfluidic channels for sample injection, reaction chamber for PCR amplification and the detector chamber, as well as the integrated microheaters and thermometers for the control of temperature in different parts of the chip. The LOC reader is required to feed the microfluidic module with sample, reagents, power supply and control signals for the management of the fluidic components and microheaters. The second part of the LOC chip will be the detector module, based on an array of piezoresistive microcantilevers [19,20]. The detector arrays are implemented using technology based on Silicon-On-Insulator (SOI) wafer in order to provide the required low thickness (340nm, single crystal Silicon). Readout is based on implanted piezoresistors. The micro cantilever is functionalized using DNA probes relevant for the detection of MS and RA. The number of cantilevers is chosen on the basis of the technological limitation related to the realization of the devices, and on the basis of the number of SNPs to be detected, according to the results of probe selection from the genome wide association studies. It is likely that a replication of each probe on multiple cantilevers is required to increase the statistical significance of results, with a factor depending on experimental results. Different steps of realization and testing are expected. In the first step, in order to test the detector functionality, detector arrays with a small number of elements (3 x 3 arrays) is tested, while in the second step devices with up to sixty sensors are used to investigate the possibility to realize devices with higher density. The LOC reader provides the bias of sensors and the readout of the analog signals (differential potential on Wheatstone bridges) coming from the sensors.

III. Conclusions

It is expected that the methodology presented in this work allows efficient genomic point-of-care diagnostics to be applied at the bedside of the patient for the early prognosis and diagnosis of the RA and MS complex disorders. Early results by combining clinical and genomic data through artificial intelligent algorithms showed that higher predictive accuracy can be achieved. Considering that the limitations of autonomous molecular-based LOC diagnostic devices - due to the relative small number of SNPs that they can host - can be overcome in the near future, the application of "smart" algorithms and data mining processes will always suggest better prognostic and diagnostic added-value at the point-of-care.

References

[1] M. J. Stansell and S. J. Stansell, " Portable self-powered system for the rapid measurement of blood electrolytes during aeromedical evacuation," Aerosp. Med., vol. 8, pp. 879-884, Aug. 1970.

[2] K. A. Erickson and P/ Wilding, "Evaluation of a novel point-of-care system, the i-STAT portable clinical analyzer," Clinical Chemistry, vol. 39, n o. 2, pp. 283-,287, Sep. 1992.

[3] C. A. Holland and F. L. Kiechle, "Point-of-care molecular diagnostic systems — past, present and future," Current Opinion in Microbiology, vol. 8, pp 504–509, Aug. 2005.

[4] M. Biehl and T. Velten, "Gaps and Challenges of Point-of-Care Technology," IEEE Sensors Journal, vol. 8, no. 5, pp. 593-600, May 2008.

[5] G. M. Whitesides, "The origins and the future of microfluidics," Nature, vol. 442, no. 7101, pp. 368–373, Jul. 2006.

[6] J. F. Steemers, K. L. Gunderson, "Illumina Inc.," Pharmacogenomics, vol. 6 no. 7, pp. 777-782, 2005.

[7] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses," Am J Hum Genet.,vol. 81, no. 3, pp. 559-575, Sep. 2007.

[8] R. Shen, J.B. Fan, D. Campbell, W. Chang, J. Chen, et al., "High-throughput SNP genotyping on universal bead arrays", Mutation Research, vol. 573, pp. 70-82, 2005.

[9] V.L. Galinsky, "Automatic registration of microarray images. I. Hexagonal grid," Bioinformatics, vol. 19, pp. 1832-1836, 2003.

[10] N. Otsu, "A Threshold Selection Method for Gray-Levels Histograms", IEEE Transactions on Pattern Analysis Machine Intelligent, vol. 13, pp. 583-598, 1979.

[11] F. Aurenhammer, "Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure," in ACM Comp. Surveys, vol. 23, pp. 345-405, 1991.

[12] N. Giannakeas, P.S Karvelis and D.I. Fotiadis, "A Classification-Based Segmentation of cDNA Microarray Images using Support Vector Machines," 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 875-878, 2008.

[13] N. Giannakeas, and D. I. Fotiadis, "An Automated Method for Gridding and Segmentation Of cDNA Microarray Images," Computerized Medical Imaging and Graphics, Vol. 33 (1), pp. 40-49, 2009.

[14] Y. Goletsis, T. P. Exarchos, N. Giannakeas, D. I. Fotiadis, "Intelligent patient profiling for diagnosis, staging and treatment selection in colon cancer", BIBE 2008, 2008.

[15] G. Potamias, L. Koumakis, A. Kanterakis, V. Moustakis, D. Kafetzopoulos, M. Tsiknakis, Artificial Intelligence and Innovations 2007: from Theory to Applications. Boston: Springer, pp. 137-145, 2007.

[16] J. Pittman, E. Huang, J. Nevins and M. West, "Prediction tree models in clinico-genomics," Bulletin of the International Statistical Institute, vol. 54, pp. 76-79, 2003.

[17] P.-N. Tan , M. Steinbach, V. Kumar, Introduction to Data Mining (First Edition). Boston: Addison-Wesley Longman Publishing Co., 2005.

[18] M. G. Dobson, P. Galvin, D. E. Barton, "Emerging technologies for point-of-care genetic testing," Expert Rev. Mol. Diagn., vol. 7, no.4, pp. 359-370, 2007.

[19] R. Mukhopadhyay, M. Lorentzen, J. Kjems, F. Besenbacher, "Nanomechanical Sensing of DNA Sequences Using Piezoresistive Cantilevers," Langmuir, vol. 21, pp. 8400-8408, 2005.

[20] S.M. Yang, C. Chang. T.I. Yin, P.L. Kuo, "DNA Hybridization measurements by self-sensing piezoresistive microcantilevers in CMOS biosensors," Sensors and Actuators B, vol. 130, pp. 674-681, 2008.