

Gene Expression Chromosomal Correlations in Tumors of Mesodermal Origin: The Case of Rhabdomyosarcoma and Acute Lymphoblastic Leukemia

Viktorina Papadimitriou¹, George I. Lambrou^{*}

^{1,2}Graduate Program “Metabolic Bones Diseases”, National and Kapodistrian University of Athens, Medical School, Mikras Asias 75, 11527, Goudi, Athens, Greece

²First Department of Pediatrics, Choremeio Research Laboratory, National and Kapodistrian University of Athens, Thivon & Levadeias, 11527, Goudi, Athens, Greece

Abstract-since the advent of high throughput methodologies, like microarrays, the load of genomic data has increased geometrically and along with that the need for computational methods which will interpret these data. In the present work we have studied the common gene expression patterns between two tumor cell types of mesodermal origin. In particular, we have attempted to find causal relations between gene expression levels with respect to chromosomal location. We have found that several genes manifested significant relations, using regression analysis and as such they could pose interesting targets for further investigations. This type of analysis can lead to the understanding of the common mechanisms that transform physiological cells to malignant, as well as it reveals a new holistic way to understand the dynamics of tumor onset as well as the mechanistic of oncogenic drivers. Such approaches could prove to be useful in the prediction of genomic targets that could be further studied in order to unravel the mechanics of tumor ontogenesis.

Keywords-Acute Lymphoblastic Leukemia, Chromosomal correlations, Mesoderm, Microarrays, Rhabdomyosarcoma.

I. INTRODUCTION

Acute lymphoblastic leukemia (ALL) and rhabdomyosarcoma (RMS) are two type of tumors, which originate from the embryonic mesoderm. ALL is the most frequent malignancy, which appears during childhood. Acute leukemia originates from the undifferentiated lymphoblast, which does not develop into the mature lymphoid cell, giving rise to a tumor. RMS is a rare cancer in childhood. This represents 5-8% of all tumors in childhood, but the sarcomas of head and neck are 12% of all neoplasias in childhood. RMS originates from myoblast or cells that will form the skeletal muscle. These are different from the smooth cells. RMS can be created at any part of the body, which has skeletal muscle but frequently appears in the head and neck. The embryonal form of RMS is most common at birth, which consists of spindle cells and botryoid form with better prognosis, but the alveolar form mainly appears in childhood and adolescence. ALL and RMS consist of cells that are undifferentiated, immortal and with the potential to divide infinitely. Myoblasts originate from the dorsal (paraxial) mesoderm, while blood cells derive from the lateral mesoderm that gives rise to the splachnic mesoderm and this to the hemangioblastic tissue. ALL and many alveolar RMS in childhood present chromosomal translocations, like the PAX3-FKHR, that is an indicator of poor prognosis and associated with metastasis [1].

During embryogenesis, blood cells originate from two sites. From the ventral mesoderm near the yolk sac, which gives rise to the intra-embryonic hematopoietic precursors, but the hematopoietic cells that last throughout the entire life time of an organism are derived from the mesodermal area surrounding the aorta. This differentiation is regulated with a network of various genes, which leads to two similar cell types with others functions and roles in the body. There are several factors that affect gene regulation, which means that aberrations in the regulatory network would lead to tumor cells.

Further on, it has been previously reported that correlation in gene expression and in particular, chromosomal correlation implies common gene regulation [1, 2]. Yet, it is also known that correlation does not imply causality. In that sense, it is of great importance if we would be able to infer gene regulatory mechanisms from chromosomal expression levels.

The present study concerns the extension of a previous work, where gene expression analysis was extensively investigated in two cell lines. The T-cell acute lymphoblastic leukemia CCRF-CEM cell line and the rhabdomyosarcoma TE-671 cell line. We used these cell lines to investigate common patterns of cellular function between the two systems. These two cell types would normally had differentiated into cells of blood and muscle cells but in an unknown stage this normal differentiation stopped and began an informal, uncontrolled proliferation which created the malignancy. The intervening stages and led in tumor genesis are unknown.

Thus, in the present work we attempted to investigate common gene regulation based on chromosomal correlations. Yet, we have tried to move a step forward and find probable etiological relations in gene expression patterns. The present work focuses towards approaches that utilize biological along with mathematical tools, in order to unravel mechanisms in gene expression patterns. Further on, it moves towards a direction that attempts to interpret gene regulation not from the differential point of view but from common gene expression patterns.

II. MATERIALS AND METHODS

2.1 Cell Cultures

The CCRF-CEM (ALL) and the TE-671 (RMS) cell lines were used as the model, both obtained from the European Collection of Cell Cultures (ECACC). The CCRF-CEM cell line, a CD4⁺ [3] and CD34⁺ presenting cell line [4], was initially obtained from the peripheral blood of a 2 year old Caucasian female. It was diagnosed as lymphosarcoma which progressed later on to acute leukemia [5]. The child was undergone irradiation therapy and chemotherapy prior to obtaining the cell line. Although remission was achieved at various stages, the disease progressed rapidly [5]. The cell line was observed to have undergone minor changes after long-term culture, except for the presence of dense granules in the nucleoli [6]. Finally, the CCRF-CEM cell line has been reported to manifest autocrine catalase activity which participates to its mechanisms of growth and progression [7]. The TE-671 cell line was initially reported to have been obtained from a cerebellar medulloblastoma, before irradiation therapy, of a six-year old Caucasian female [8] and characterized later on [9]. However, it is today known that this cell line is parental if not identical to the RD [10] rhabdomyosarcoma cell line. However, several reports still refer to this cell line as medulloblastoma [11, 12]. The experimental process of cell treatment has been described previously [1].

2.2 Experimental Procedures

Experimental procedures have been described previously in detail by our group [1]. In brief, factors estimated, included cell proliferation, cell cycle distribution and microarray experimentation. For the assay of mRNA levels two sets of microarray chips were used: cDNA microarray chips (4.8k genes) obtained from TAKARA (IntelliGene™ II Human CHIP 1) [13] and microarray chips (9.6k genes) from the Institut fuer Molekularbiologie und Tumorforschung, Microarray Core Facility of the Philipps-Universitaet, Marburg Germany (IMT9.6k). Microarray experimentation has been previously described in our previous work [1]. The microarray data have been submitted to the GEO Database under the Accession Number GSE34522.

2.3 Microarray Data Analysis

Microarray data pre-processing analysis was performed with ImaGene®v.6.0 Software (BioDiscovery Inc, CA) and ARMADA software (National Hellenic Research Foundation, Athens Greece) [14]. Data were collected from exported text file and data pre-processing was performed using the Microsoft Excel® environment. Background correction has been performed by using the robust loess-based background correction (rLsBC) approach, as proposed by *Sifakis et al.* (2011) [15]. The background corrected signal intensities were further normalized in order to mitigate the effect of extraneous, non-biological variation in the measured gene expression levels. Genes were tested for their significance in differential expression using a z-test. Genes were considered to be significantly differentially expressed if they obtained a *p*-value <0.05. The False Discovery Rate (FDR) was calculated as previously described [16-18]. There was a FDR of 1% for *p*<0.05 for the IntelliGene microarray chip, and a FDR of 9% for *p*<0.01 for the IMT 9.6k microarray chip. Calculating the FDR for the combination of both platforms gives a FDR of 6% for *p*<0.01.

2.4 Chromosomal Regressions and Data Analysis

Chromosome mapping was performed with Genesis 1.7.2 (Technische Universitaet-Graz, Austria) using Pearson's correlation, Spearman's rank order correlation [2, 19, 20] and WebGestalt web-tool (Vanderbilt University, The Netherlands, <http://bioinfo.vanderbilt.edu/gotm/>) [21]. 2D Chromosomal regressions have been performed among the similarly expressed genes between RMS (TE-671) and ALL (CCRF-CEM). 3D chromosomal regressions were performed between similarly expressed genes of RMS and ALL as well as we added the third dimension, which was the \log_2 transformed ratio of the respective genes. Regressions and data analysis were performed with the Matlab® computational environment (The Mathworks Inc. Natick, MA). For linear 2D regressions the binomial equation was used of the form $f(x)=y=ax^2+bx+c$, while for the 3D regressions the equation of the form $f(x,y)=z=ay+bx+c$.

III. RESULTS

Regression analysis was performed in order to find probable causal relations in gene expression with respect to cell type and chromosomal distribution. In particular, we have used genes that manifested similar expression levels, meaning that they were those that were not differentially expressed. We have performed two different kinds of regression analysis. One included a 2D regression of the form $y=ax^2+bx+c$ and the second of the form $z=ay+bx+c$. 2D regressions manifested significant relations between the two cell types and in particular for chromosome 1 with an $R^2=0.82$ (Fig. 1A), chromosome 3 with an $R^2=0.82$ (Fig. 1B), chromosome 8 with an $R^2=0.89$ (Fig. 1C), chromosome 13 with an $R^2=0.71$ (Fig. 1D), chromosome 14 with an $R^2=0.91$ (Fig. 1E), chromosome 15 with an $R^2=0.69$ (Fig. 1F), chromosome 16 with an $R^2=0.89$ (Fig. 1G), chromosome 17 with an $R^2=0.81$ (Fig. 1H), chromosome 21 with an $R^2=0.88$ (Fig. 1I), chromosome 22 with an $R^2=0.97$ (Fig. 1J) and chromosome X with an $R^2=0.71$ (Fig. 1K).

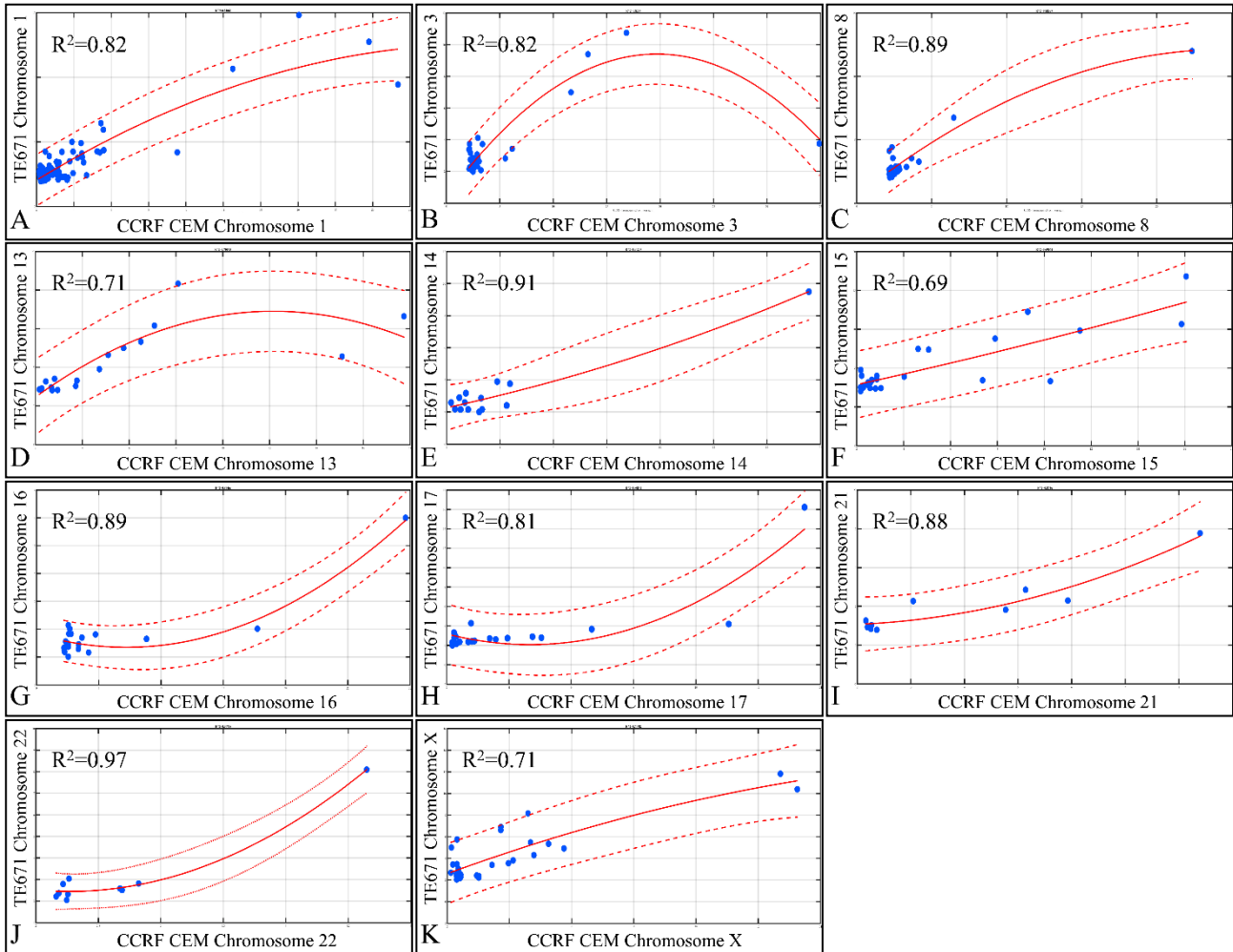


FIGURE 1. REGRESSIONS OF GENE EXPRESSION DATA BETWEEN CCRF-CEM AND TE-671 CELLS, WITH RESPECT TO CHROMOSOMAL EXPRESSION. SIGNIFICANT RELATIONS WERE FOUND FOR CHROMOSOMES 1 (A), 3 (B), 8 (C), 13 (D), 14 (E), 15 (F), 16 (G), 17 (H), 21 (I), 22 (J) and X (K).

Although 2D regressions manifested significant relations with regard to gene expression, still the revealed functions did not have the complete set of the definition of a function. In other words, for every $x \in D(f)$, there are two y , which can be described as $\exists x_1, x_2 \in D(f): x_1 \neq x_2 \Rightarrow f(x_1) = f(x_2)$, or $y_1 = y_2$, where $D(f)$ is the definition domain of f , and x, y are the expression levels of each gene. The question that came up from this observation, was that the manifested correlations could provide more information if another dimension could be added to the calculations. The solution to this question came from the observation that the \log_2 transformed ratio of TE-671 gene expression levels over the CCRF-CEM gene expression levels, could be added as the third dimension. Thus, we have formulated a new relation $z=ay+bx+c$ or $f(x,y)=ay+bx+c$, where z can

be formulated as $z = \log_2 \left(\frac{x}{y} \right)$. Thus, the aforementioned equations become $\log_2 \left(\frac{x}{y} \right) = ay + bx + c$. The results of the 3D regression are presented in **Fig. 2**.

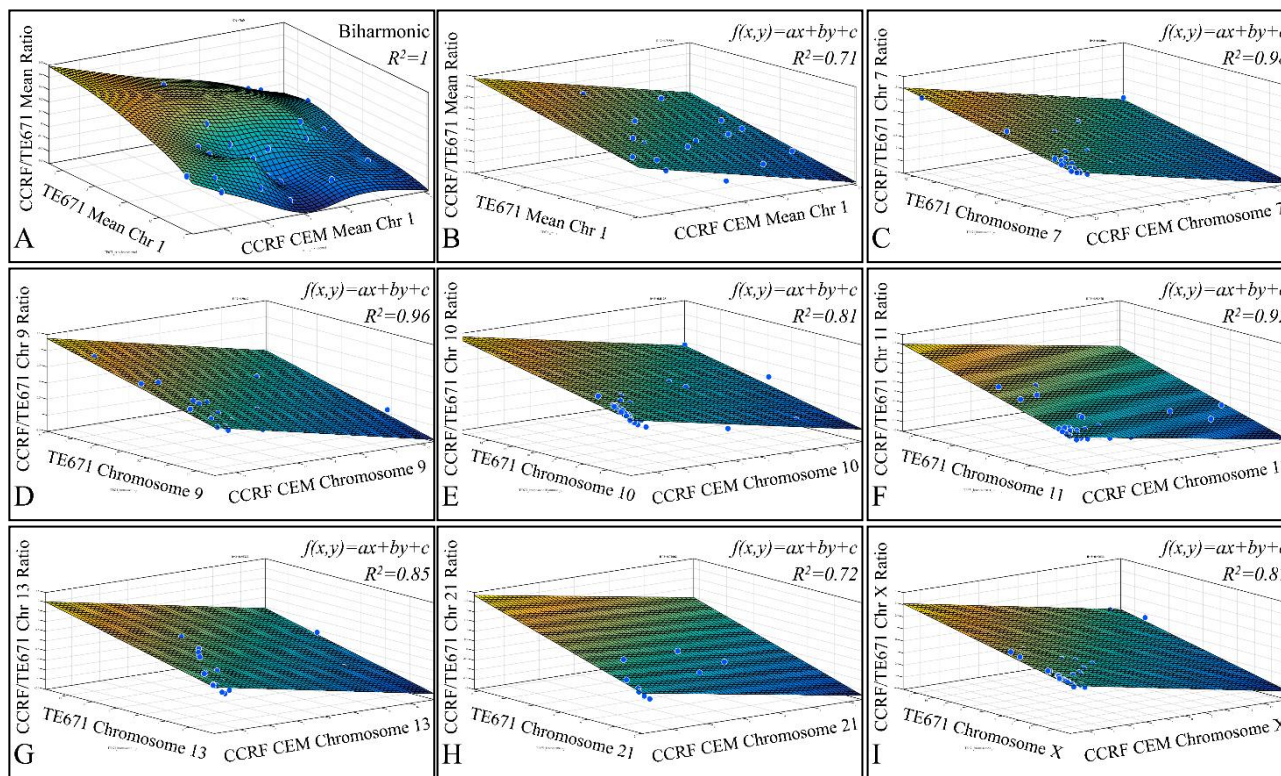


FIGURE 2. 3D REGRESSIONS OF GENE EXPRESSION DATA BETWEEN CCRF-CEM AND TE-671 CELLS AND THEIR RESPECTIVE LOG2 TRANSFORMED RATIO WITH RESPECT TO CHROMOSOMAL EXPRESSION. SIGNIFICANT RELATIONS WERE FOUND FOR CHROMOSOMES 1 (A) WITH BIHARMONIC FUNCTION, 1 (B) WITH A LINEAR FUNCTION, 7 (C), 9 (D), 10 (E), 11 (F), 13 (G), 21 (H) and X (I).

In particular, 3D regression analysis manifested significant relations between the two cell types and their ratio, and in particular for chromosome 1 with an $R^2=1$ (**Fig. 2A**) and with the use of the biharmonic function, chromosome 1 with an $R^2=0.71$ (**Fig. 2B**) with the use of the aforementioned linear function, chromosome 7 with an $R^2=0.98$ (**Fig. 2C**), chromosome 9 with an $R^2=0.96$ (**Fig. 2D**), chromosome 10 with an $R^2=0.81$ (**Fig. 2E**), chromosome 11 with an $R^2=0.92$ (**Fig. 2F**), chromosome 13 with an $R^2=0.85$ (**Fig. 2G**), chromosome 21 with an $R^2=0.72$ (**Fig. 2H**) and chromosome X with an $R^2=0.87$ (**Fig. 2I**).

It appeared that relations in three dimensions manifested better results for the specified chromosomes as compared to two dimensional regressions. This observation, shows that probably a third dimension is necessary in order to comprehend gene regulatory mechanisms.

IV. DISCUSSION

In the present study we used computational and mathematical approaches in order to answer the question whether, a causal relation between gene expression levels of two cell types, could be found based on their chromosomal distribution. We have found that such significant relations could be found both using 2D as well as 3D regression methods. Especially, 2D regressions manifested very good relations among cell types with respect to chromosomes, yet without manifesting a clear causal relation. On the other hand, 3D regressions manifested better results suggesting that causal multi-dimensional relations probably exist. At the same time, the fact that certain genes and chromosomes manifest significant correlations between two different cell types, yet of similar origin, implies that they could probably consist common regulatory mechanisms of tumor ontogenesis. In our previous work we have reported that ANXA4 (Chromosome 2), NP25 (Chromosome 3), VEGFC (Chromosome 4), PDLIM7 and THBS4 (Chromosome 5), CUL7 (Chromosome 6), CD36 (Chromosome 7), BNIP3L

(Chromosome 8), CDC2 and IL2RA (Chromosome 10), MAB21L1 and FOXO1 (Chromosome 13), TCL1A (Chromosome 14), FEM1B (Chromosome 15), CFDP1 and MMP2 (Chromosome 16), RTTN (Chromosome 18), PDCD5 (Chromosome 19), , NNAT and ZNF313 (Chromosome 20), MBNL3, PLAC1, RPS6KA3 and CD40LG (Chromosome X) were genes found to participate in cell differentiation and embryonal processes[1]. These genes were found to manifest significant correlations both in the 2D as well as in 3D regressions. This finding consist of an important observation since it is probable that these genes could possess a common machinery, either through their gene regulation or aberrations, that drives tumor ontogenesis for the two cell types under investigation.

V. CONCLUSION

The present approach attempted to find common regulatory mechanisms in gene expression patterns of two cell types. In particular, we have attempted to discover possible causal relations in gene expression patterns. We have found evidence that such correlations could exist and it is probable that certain genes could be of great significance in tumor ontogenesis. Such approaches could prove to be useful in the prediction of genomic targets that could be further studied in order to unravel the mechanics of tumor ontogenesis.

ACKNOWLEDGEMENTS

VP: This work was funded in part by the National and Kapodistrian University of Athens, Medical School. **GIL:** This work was funded in part by the National and Kapodistrian University of Athens, Medical School. **Authors' Contributions:** **VP:** performed computational experiments, drafted the manuscript, proof-read the manuscript. **GIL:** conceived the idea, performed computational experiments, wrote Matlab code, performed data analysis and drafted the manuscript, gave final permission for submission.

COMPETING INTEREST, DISCLOSURES AND CONFLICT OF INTEREST

The authors have nothing to disclose and no conflict of interest

REFERENCES

- [1] G. I. Lambrou, A. Zaravinos, M. Adamaki, D. A. Spandidos, F. Tzortzatou-Stathopoulou, and S. Vlachopoulos, "Pathway simulations in common oncogenic drivers of leukemic and rhabdomyosarcoma cells: a systems biology approach," *Int J Oncol*, vol. 40, pp. 1365-90, May 2012.
- [2] B. A. Cohen, R. D. Mitra, J. D. Hughes, and G. M. Church, "A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression," *Nat Genet*, vol. 26, pp. 183-6, Oct 2000.
- [3] L. Miranda, J. Wolf, S. Pichuantes, R. Duke, and A. Franzusoff, "Isolation of the human PC6 gene encoding the putative host protease for HIV-1 gp160 processing in CD4+ T lymphocytes," *Proc Natl Acad Sci U S A*, vol. 93, pp. 7695-700, Jul 23 1996.
- [4] C. Naujokat, O. Sezer, H. Zinke, A. Leclere, S. Hauptmann, and K. Possinger, "Proteasome inhibitors induced caspase-dependent apoptosis and accumulation of p21WAF1/Cip1 in human immature leukemic cells," *Eur J Haematol*, vol. 65, pp. 221-36, Oct 2000.
- [5] G. E. Foley, H. Lazarus, S. Farber, B. G. Uzman, B. A. Boone, and R. E. McCarthy, "Continuous Culture of Human Lymphoblasts from Peripheral Blood of a Child with Acute Leukemia," *Cancer*, vol. 18, pp. 522-9, Apr 1965.
- [6] B. G. Uzman, G. E. Foley, S. Farber, and H. Lazarus, "Morphologic variations in human leukemic lymphoblasts (CCRF-CEM cells) after long-term culture and exposure to chemotherapeutic agents. A study with the electron microscope," *Cancer*, vol. 19, pp. 1725-42, Nov 1966.
- [7] P. A. Sandstrom and T. M. Buttke, "Autocrine production of extracellular catalase prevents apoptosis of the human CEM T-cell line in serum-free medium," *Proc Natl Acad Sci U S A*, vol. 90, pp. 4708-12, May 15 1993.
- [8] R. M. McAllister, H. Isaacs, R. Rongey, M. Peer, W. Au, S. W. Soukup, *et al.*, "Establishment of a human medulloblastoma cell line," *Int J Cancer*, vol. 20, pp. 206-12, Aug 15 1977.
- [9] H. S. Friedman, S. H. Bigner, R. D. McComb, S. C. Schold, Jr., J. F. Pasternak, D. R. Groothuis, *et al.*, "A model for human medulloblastoma. Growth, morphology, and chromosomal analysis in vitro and in athymic mice," *J Neuropathol Exp Neurol*, vol. 42, pp. 485-503, Sep 1983.
- [10] R. M. McAllister, J. Melnyk, J. Z. Finkelstein, E. C. Adams, Jr., and M. B. Gardner, "Cultivation in vitro of cells derived from a human rhabdomyosarcoma," *Cancer*, vol. 24, pp. 520-6, Sep 1969.
- [11] C. M. Yeung, B. S. An, C. K. Cheng, B. K. Chow, and P. C. Leung, "Expression and transcriptional regulation of the GnRH receptor gene in human neuronal cells," *Mol Hum Reprod*, vol. 11, pp. 837-42, Nov 2005.
- [12] E. S. Chu, T. K. Wong, and C. M. Yow, "Photodynamic effect in medulloblastoma: downregulation of matrix metalloproteinases and human telomerase reverse transcriptase expressions," *Photochem Photobiol Sci*, vol. 7, pp. 76-83, Jan 2008.
- [13] H. W. Chung, S. W. Park, J. B. Chung, J. K. Kang, J. W. Kim, H. S. Kim, *et al.*, "Differences in genetic expression profiles between young-age and old-age gastric adenocarcinoma using cDNA microarray for endocrine disruptor study," *Oncol Rep*, vol. 12, pp. 33-9, Jul 2004.
- [14] A. Chatziioannou, P. Moulos, and F. N. Kolisis, "Gene ARMADA: an integrated multi-analysis platform for microarray data implemented in MATLAB," *BMC Bioinformatics*, vol. 10, p. 354, 2009.

-
- [15] E. G. Sifakis, A. Prentza, D. Koutsouris, and A. A. Chatzioannou, "Evaluating the effect of various background correction methods regarding noise reduction, in two-channel microarray data," *Comput Biol Med*, Nov 8.
- [16] Y. Klipper-Aurbach, M. Wasserman, N. Braunsiegel-Weintrob, D. Borstein, S. Peleg, S. Assa, *et al.*, "Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus," *Med Hypotheses*, vol. 45, pp. 486-90, Nov 1995.
- [17] J. D. Storey and R. Tibshirani, "Statistical significance for genomewide studies," *Proc Natl Acad Sci U S A*, vol. 100, pp. 9440-5, Aug 5 2003.
- [18] J. D. Storey and R. Tibshirani, "Statistical methods for identifying differentially expressed genes in DNA microarrays," *Methods Mol Biol*, vol. 224, pp. 149-57, 2003.
- [19] A. Sturn, J. Quackenbush, and Z. Trajanoski, "Genesis: cluster analysis of microarray data," *Bioinformatics*, vol. 18, pp. 207-8, Jan 2002.
- [20] F. Reyat, N. Stransky, I. Bernard-Pierrot, A. Vincent-Salomon, Y. de Rycke, P. Elvin, *et al.*, "Visualizing chromosomes as transcriptome correlation maps: evidence of chromosomal domains containing co-expressed genes--a study of 130 invasive ductal breast carcinomas," *Cancer Res*, vol. 65, pp. 1376-83, Feb 15 2005.
- [21] B. Zhang, D. Schmoyer, S. Kirov, and J. Snoddy, "GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies," *BMC Bioinformatics*, vol. 5, p. 16, Feb 18 2004.