Global Patterns Of Changes In The Gene Expression Associated With Genesis Of Cancer

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Ganiraju Manyam
Master of Science
IIIT-Hyderabad, 2004
Bachelor of Engineering
Bharatiar University, 2002

Director: Dr. Ancha Baranova, Associate Professor
Department of Molecular & Microbiology

Fall Semester 2009
George Mason University
Fairfax, VA
DEDICATION

To my parents
Pattabhi Ramanna and Veera Venkata Satyavathi
who introduced me to the joy of learning.

To friends, family and colleagues
who have contributed in work, thought, and support to this project.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Ancha Baranova, whose tolerance, patience, guidance and encouragement helped me throughout the study. This dissertation would not have been possible without her ever ending support. She is very sincere and generous with her knowledge, availability, compassion, wisdom and feedback.

I would also like to thank Dr. Vikas Chandhoke for funding my research generously during my doctoral study at George Mason University.

Special thanks go to Dr. Patrick Gillevet, Dr. Alessandro Giuliani, Dr. Maria Stepanova who devoted their time to provide me with their valuable contributions and guidance to formulate this project.

Thanks to the faculty of Molecular and Micro Biology (MMB) department, Dr. Jim Willett and Dr. Monique Vanhoek in embedding valuable thoughts to this dissertation by being in my dissertation committee. I would also like to thank the present and previous doctoral program directors, Dr. Daniel Cox and Dr. Geraldine Grant, for facilitating, allowing, and encouraging me to work in this project.

My heartful thanks goes to the formal and current graduate students in the MMB department Dr. Mohammed Jarrar, Dr. Michael Estep, Dr. Manpreet Randhawa, Dr. Aybike Birerdinc, Subashini Iyer, Beatrix Meltzer and Eshwar Iyer who contributed their time and effort in facilitating and providing the scientific environment for this project.

A very special thanks to my friends in galigang for their continual moral support to enhance my spirit in progressing this project.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>Abstract</td>
<td>xi</td>
</tr>
<tr>
<td>A Summary</td>
<td>xiii</td>
</tr>
<tr>
<td>1. Introduction</td>
<td></td>
</tr>
<tr>
<td>- Cancer and its genesis</td>
<td>1</td>
</tr>
<tr>
<td>- Genetic and epigenetic events causing cancer</td>
<td>2</td>
</tr>
<tr>
<td>- Diversity of human gene expression</td>
<td>8</td>
</tr>
<tr>
<td>- Bioinformatics &amp; Cancer</td>
<td>11</td>
</tr>
<tr>
<td>- The analysis of microarrays and the cancer transcriptome</td>
<td>15</td>
</tr>
<tr>
<td>- Cancer – A systems biology perspective</td>
<td>23</td>
</tr>
<tr>
<td>2. Genome wide discrimination of normal and tumor samples</td>
<td></td>
</tr>
<tr>
<td>- Rationale</td>
<td>25</td>
</tr>
<tr>
<td>- Background</td>
<td>25</td>
</tr>
<tr>
<td>- Hypothesis</td>
<td>28</td>
</tr>
<tr>
<td>- Materials and Methods</td>
<td>28</td>
</tr>
<tr>
<td>- Results and Discussion</td>
<td>33</td>
</tr>
<tr>
<td>a) Modeling strategy</td>
<td>33</td>
</tr>
<tr>
<td>b) Assessment of the global and signature-specific gene expression</td>
<td>36</td>
</tr>
<tr>
<td>c) Assessment of the global and signature-specific gene expression</td>
<td>36</td>
</tr>
<tr>
<td>d) Principal component analysis (PCA) of the distance spaces</td>
<td>41</td>
</tr>
<tr>
<td>e) Cancer – An attractor with intermediate regulatory framework</td>
<td>46</td>
</tr>
<tr>
<td>- Conclusion and Future Perspective</td>
<td>52</td>
</tr>
<tr>
<td>3. Abundance based transcriptome analysis as a tool for automated discovery of the tumor biomarkers</td>
<td>58</td>
</tr>
<tr>
<td>- Rationale</td>
<td>58</td>
</tr>
<tr>
<td>- Background</td>
<td>58</td>
</tr>
</tbody>
</table>

v
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: The attributes of two-point datasets describing paired normal and tumor tissue samples collected from the same individual.</td>
<td>29</td>
</tr>
<tr>
<td>2: Description of the two-point datasets comprised of normal and tumor samples collected from the same tissue type across a number of subjects.</td>
<td>31</td>
</tr>
<tr>
<td>3: Description of the datasets with three or more physiological groups of normal and tumor samples collected across the same subject or a number of subjects.</td>
<td>32</td>
</tr>
<tr>
<td>4: Rankings of the tumor malignancy potential by relative distance to the normal sample space of two-point paired datasets.</td>
<td>39</td>
</tr>
<tr>
<td>5: Mean, Standard Deviation and Variance calculated for the global and specific distances from individual samples to the normal sample space of paired datasets.</td>
<td>40</td>
</tr>
<tr>
<td>6: Mean, Standard Deviation and Variance calculated for the global and specific distances from individual samples to the normal sample space in population datasets.</td>
<td>42</td>
</tr>
<tr>
<td>7: Illustration of the relative importance of components and the actual loadings corresponding to the distances in specific two-point datasets.</td>
<td>48</td>
</tr>
<tr>
<td>8: PCA profiles of two-point paired datasets representing the proportion of variance observed by each component.</td>
<td>49</td>
</tr>
<tr>
<td>9: PCA profiles of two-point population datasets representing the proportion of variance observed by each component.</td>
<td>50</td>
</tr>
<tr>
<td>10: A list of broader tumor descriptors used in the EST abundance study.</td>
<td>65</td>
</tr>
<tr>
<td>11: Statistics describing the libraries, unigenes (gene clusters) and ESTs derived from normal tissues and tumors samples.</td>
<td>69</td>
</tr>
<tr>
<td>12: Tissue-specific distribution of EST clusters composed of both normal and tumor ESTs, tumor ESTs only and normal ESTs only.</td>
<td>71</td>
</tr>
</tbody>
</table>
13: Diversity, richness, evenness and variance of the diversity of ESTs in each of the normal human tissues.................................................................74

14: Diversity, richness, evenness and variance of the diversity of ESTs in each of the human tumors.................................................................75

15(a): Illustration of the tissues having increased diversity in tumors compared to normal tissue.................................................................76

15(b): Illustration of the tissues having decreased diversity in tumors compared to normal tissue.................................................................77

16: Mann-whitney test results of the M/B, M/P and P/B ratios between the telomere and centromere associated genes in all human chromosomes.................................95

17: Statistical variation in the distribution of significant p-values corresponding to genes in the telomere and centromere associated regions...............................97

18: The p-values obtained from chi-square test of homogeneity between the distribution of M/B ratios in telomere and centromere associated regions.................104

19: The p-values obtained by the chi-square test using 2x2 contingency table between the telomere and centromere associated genes........................................105
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Illustration of the distance parameters derived for the paired breast carcinoma dataset (GSE5764) using paired-plot panels</td>
<td>38</td>
</tr>
<tr>
<td>2: Distance parameters of the esophageal dataset representing normal esophagus, Barrett’s esophagus and esophagus carcinoma samples</td>
<td>43</td>
</tr>
<tr>
<td>3(a): Linear graphs depicting the relative distance of every given sample to the normal sample space in the multi-stage datasets</td>
<td>44</td>
</tr>
<tr>
<td>3(b): Illustration of linear graphs using DN metric for the multi-stage datasets GSE6764 and GSE1097</td>
<td>45</td>
</tr>
<tr>
<td>4: Three dimensional representation of principal components PC1, PC2 and PC3 in the two-point paired and population datasets</td>
<td>51</td>
</tr>
<tr>
<td>5: Attractor topology and the classical view of cancer</td>
<td>53</td>
</tr>
<tr>
<td>6: Heatmap of the EST abundance data extracted from 27 different human normal and tumor tissues of the Unigene system</td>
<td>79</td>
</tr>
<tr>
<td>7: KPP representation of the KEGG cell cycle pathway painted using genes differentially expressed in the malignant tumors</td>
<td>80</td>
</tr>
<tr>
<td>8: KPP representation of the KEGG human Calcium signaling pathway enriched in the normal tissues</td>
<td>81</td>
</tr>
<tr>
<td>9: Illustration of the proportion of genes marked as telomere associated and centromere associated in every human chromosome</td>
<td>90</td>
</tr>
<tr>
<td>10: Depiction of the percentage of genes that change their expression on each chromosome in the telomere and centromere associated regions</td>
<td>92</td>
</tr>
<tr>
<td>11: Bar graph of the proportion of GO identifiers connected with the telomere associated and centromere associated genes</td>
<td>98</td>
</tr>
</tbody>
</table>
12: M/B, M/P and P/B gene expression ratio plots over the physical length of human chromosome X

13: Plots of the quantity of upregulated and downregulated transcripts in each of the M/B ratio bins for the whole human genome

14: Distance indices for the telomeric and non-telomeric genes in the prostate carcinoma dataset
ABSTRACT

GLOBAL PATTERNS OF CHANGES IN THE GENE EXPRESSION ASSOCIATED WITH GENESIS OF CANCER

Ganiraju Manyam, PhD
George Mason University, 2009

Dissertation Director: Dr. Ancha Baranova

Cancer arises from a stepwise accumulation of genetic changes through expansion of the malignant cell clones in the population of pre-malignant cells undergoing the Darwinian selection process. In other words, cancer is an outcome of continuous and random acquisition of the changes in the genomes of individual cells. These modifications gradually and progressively change the phenotype of the normal cell making it more malignant through a loss of an overall stability of genome. To gain the comprehension of the mechanisms underlying tumor development, a number of high-throughput expression studies have been performed. The objective of the current study is to use publicly available datasets in order to analyze the most general features of the malignant cell, thus, investigating molecular phenomena common for all tumor cells, with no regard to the characteristics related to tumor’s tissue of origin. Thus, we analyzed and compared the transcript diversity patterns in tumor and normal cells, studied an expression of the genes located adjacent to the telomeres and provided an evidence for the hypothesis that tumor
state behaves as stable “attractor” state. An intermediate regulatory framework hypothesis implying a set of local ‘vantage points’ genes that control the transcription of all other genes in a semi-democratic fashion has been endorsed.
A SUMMARY

The central idea of this dissertation is to explore gene expression patterns in the cancer cells having in mind a systems biology perspective on this disease. In this study we attempted to uncover the most general features of the malignant cell, thus, investigating molecular phenomena common for all tumor cells, with no regard to the tissue-specific characteristics of individual tumors. To perform large-scale data analysis, we have developed a number of novel bioinformatics techniques and as well employed some algorithms previously developed for other purposes and published elsewhere. The results obtained in this study enhance a general understanding of cancer as an expression system comprising of the components dynamically interacting with each other. The results reported in this dissertation and novel tools for in silico analysis of the cancer cell provide novel avenues for the functional genomics and systems biology of cancer and may be of help for large-scale computational modeling of cancer.

The first chapter provides general information on cancer, the molecular events causing this pathology as well as an introduction to human gene expression and its diversity. This summarization helps to place cancer in the perspective of gene expression both for the biologists and bioinformatic researchers. A review of bioinformatics studies of cancer is aimed to introduce the readers to the in silico analysis of this disorder. An overview of
the high-throughput experimental platforms, particularly, microarrays, associated methods of the gene expression analysis, an utility of these methods for the cancer transcriptome studies is accompanied by a special emphasis on the meta-analysis of expression data. Finally, the systems biology perspective of carcinogenesis is outlined, stressing the importance of inter-disciplinary nature of these studies allowing comprehension of the malignant cell at the systems biology level.

The second chapter describes the analysis of the cancer expression system in the perspective of attractor states. First, the concept of the cell as a dynamic system and the concept of attractor states are reviewed. The modeling strategy for the distance based statistical approach is described in details. The data produced as a result of the distance analysis of the two-point and the multi-point datasets are presented in two different sections. Finally, the results of the principal component analysis (PCA) are described. These results are shedding more light on the attractor behavior of cancer.

The third chapter explores the bioinformatics methods developed to analyze transcriptional abundances. We reviewed the UniGene database system that provides information on the clusters of ESTs obtained by sequencing of cDNA libraries prepared form various human tumors and tissues. The adaptation of the Shannon’ statistics for estimation of the diversity to the gene expression systems is detailed in the Methods section. The classification of human tissues and tumor types as well as quantitative estimation of unique and common gene clusters in the normal and cancer tissues are
described in two separate sections. Diversity estimations for the various human tissues and the lists of the potential tumor biomarkers and biomarkers of normal tissue functioning (anti-cancer biomarkers) are given. Finally, we present the results of the functional analysis of the protein-coding subset of the tumor biomarkers and anti-cancer biomarkers.

The last chapter reviews the role of telomeres in cancer development and the previous studies on the telomere position effect as well as the methodology we employed to define subtelomeric and non-telomeric region of the human chromosomes. The statistical techniques allowing comparison of the expression levels for the genes located in subtelomeric region to that in other regions of the human genome are discussed and results presented. An exploratory approach allowing estimation of the optimal length of the subtelomeric regions is presented. Finally, we have used the distance analysis technique previously discussed in the second chapter to compare functional properties of the expression patterns of human genes located in these two regions.

Finally, an appendix with Supplementary information containing the data associated with the chapters two, three and four was put together. Of particular interest, this appendix contains a description of a novel publicly available tool, KEGG Pathway Painter. This tool was specifically developed to provide automated summarization of the functional information pertinent to the genes comprising “top hit” lists routinely obtained in large-scale transcriptome studies.
1. Introduction

Cancer and its genesis

Cancer can be attributed to as many as 25% of deaths in the United States, making it the second most common cause of death (ACS 2009). Cancer arises from a stepwise accumulation of genetic changes through clonal expansion events in the population of pre-malignant cells undergoing the Darwinian selection process (Weinberg 2007). In a genetics perspective, it is a micro-evolutionary phenomenon resulting in the cooperative malfunction of a number of human genes enhancing the selfish survival and metastasis of cells. Each individual tumor is an outcome of continuously acquisition of the random mutations in the genomes of individual cells, a process supported by a natural selection.

Carcinogenesis is complex process initiated in a single cell or a group of cells, progressively leading to the disruption of the normal tissue architecture. The subset of cells tolerates deleterious mutations, accelerates their proliferation and invades the surrounding tissue followed by metastasis to the distant organs. This uncontrolled growth of cellular population is often due to the damage of the molecular circuitry responsible for programmed cell death or apoptosis. The cause of carcinogenesis is unknown. There are two common theories that explain the process. The somatic mutation theory states
that DNA mutations that occur in the genes regulating cell cycle and proliferation of a single somatic cell cause its transformation into the malignant one through the disruption of the otherwise quiescence state (Varmus 2006). The tissue organization field theory suggests that proliferation is the default state of the cell and that carcinogenesis is a problem of tissue restructuring and organization (Sonnenschein and Soto 2008). To the most part, a combined understanding provided by these theories enables one to comprehend the causation for the genesis of cancer.

In a combined view, cancer is produced as the result of a Darwinian evolution of cells located in the microenvironment within the particular tissue of multicellular organism (Stratton, Campbell et al. 2009). The progression of cancer leads to an increase of the overall mutation rates due to the changes in certain genes that serve as caretakers of the genome. Cumulative load of the mutations pushes cell to proliferate faster than normal by whatever mean possible, direct or indirect (Strauss 1998). Usually, the hypermutability is observed at early stage of the tumor development, often in the premalignant cell before it successful transforms into the malignant one. Gene expression level changes observed in tumors are caused by a combination of genetic and epigenetic events (Gronbaek, Hother et al. 2007). Mutated genes tend to express ectopically, in the tissues where they are normally silenced.

**Genetic and epigenetic events causing cancer**

The genetic events causing cancerogenesis include point mutations and deletions or insertions in small as well as large DNA segments. The rearrangements of the genome
and increase/decrease of the gene/chromosome copy number are also quite common events in this process. Additionally, somatic mutations in mitochondrial genomes were also reported in many tumor types, although the precise role of these alterations in cancer progression is not well comprehended (Chatterjee, Mambo et al. 2006). Mutations can also happen due to disruption of the gene by insertion of the completely exogenous DNA, for example, some tumorigenic viruses, including HPV, EBV, HBV, HHV8 which can also possess viral oncogenes (Talbot and Crawford 2004). Of all these mutations described above, a small portion can be fixed in the cellular lineage by the process of natural selection. The mutational rates for these different kinds of genetic alterations vary. Generally, these rates tend to increase in response to the exposure to exogenous mutagens or clastogens, for example, tobacco smoke or X-ray. Mutation rates are also increased in subjects with certain inherited diseases associated with increased cancer risk, for example, xeroderma pigmentosum, Fanconi anemia or ataxia-telangiectasia (Kennedy and D'Andrea 2006; Stratton, Campbell et al. 2009). The increased rate of mutation would yield increased DNA sequence diversity, providing the selection with the raw material to choose from and support the drive for increased proliferation, invasion and cancer.

Mutations found in tumors often damage the key regulatory genes controlling cell cycle, proliferation, apoptosis, genomic stability and other biological processes or functions, which normally prevent the cancer. The key genes in this context can be subdivided into
two categories: cancer causing genes (oncogenes) and the tumor suppressor genes (often abbreviated as TSGs).

Oncogenes code for proteins that positively regulate cell proliferation and/or negatively affect apoptosis. Usually, these genes are activated by point mutation or fusion with other gene or juxtaposition to enhancer element that drive their expression to higher level (Konopka, Watanabe et al. 1985; Tsujimoto, Gorham et al. 1985). The cellular proto-oncogenes may acquire genetic mutation or increase their copy number (become amplified) and, thus, transform into oncogenes (Croce 2008). Oncogenes are identified by their tendency to increase the tumor growth rate and their ability to transform the cell morphologically. The protein products of oncogenes include transcription factors, chromatin remodelers, signal transducers, apoptosis suppressors, growth factors and their receptors.

As opposed to oncogenes, tumor suppressors are capable of the suppression of oncological transformation by negative regulation of the genes in the pathways related to oncogenesis, for example, cell cycle (Kopnin 2000) or positive regulation of apoptosis. Unlike oncogenes, tumor suppressor genes require mutations in both of their alleles in the diploid genome as they act recessively. The mutation patterns of TSGs range from single base substitutions to whole gene deletions, generally tending to abolish the functioning of the protein product of tumor suppressor gene.
In the biological knowledge bases summarizing information of the genes involved in malignization of human cells, proven tumor suppressors are relatively rare (10%) as compared to the proven oncogenes (90%) (Stratton, Campbell et al. 2009). The mutational events promoting the tumor growth may either upregulate oncogenes or downregulate the tumor suppressor genes. The overall increase in the transcription rates were noted during cancer progression which might bias in favor of the outnumbered oncogenes and push the cell through malignization threshold.

In their natural evolution, the cancer cells gradually and progressively change their phenotype toward most malignant as result of the accumulation of the mutations in the oncogenes and the tumor suppressor genes (Barrett, Oshimura et al. 1986), the rearrangements of chromosomes (Radman, Jeggo et al. 1982) and the perturbation of the gene expression levels (Nicolson 1991). Many genes and molecular pathways involved in the underlying processes are well studied (Vogelstein and Kinzler 2004), but the complex spatiotemporal patterns of interactions between the involved molecules contribute to the difficulties of the comprehension of the malignant cell system (Hornberg, Bruggeman et al. 2006). On top of the genetic modifications, epigenetic events also play a vital role in the initiation and progression of a tumor.

Epigenetic alterations comprise mitotically and meiotically heritable changes in gene expression that are not caused by changes in the primary DNA sequence. These changes are increasingly being recognized for their roles in carcinogenesis (Gronbaek, Hother et
al. 2007). Darwinian selection can act upon the phenotypic effects that are generated by epigenetic changes for cancer evolution, analogous to selection of mutations in the case of genetic alterations. Methylation is the most common epigenetic event, which often leads to gene silencing through subsequent histone deacetylation and chromatin condensation (Worm and Guldberg 2002). Hypermethylation at the promoters of tumor suppressors is a distinctive feature seen in many tumors, which can happen very early in the tumor progression. Hypermethylation is known to mediate an imbalance in many important signaling pathways (Baylin, Esteller et al. 2001; Gronbaek, Hother et al. 2007). CpG islands are the hot spots for methylation, with 50-70% of the cytosines in these sites are being methylated in human tissues (Ehrlich, Gama-Sosa et al. 1982; Esteller 2006). These cytosines often undergo spontaneous transition to thymines by deamination (Rideout, Coetzee et al. 1990). If these transitions are not corrected, they become either somatic or germline point mutations.

The change in the DNA methylation landscape of a cancer cell usually occurs in the context of other epigenetic changes. DNA methylation attracts methyl-CpG binding proteins and DNA methyltransferases. In turn, these proteins associate with histone deacetylases and histone methyltransferases, two types of enzymes playing a key role in chromatin remodeling (An 2007). Some well known oncogenes, for example, PML/RARα fusion, an archetypal chimeric oncoprotein, were shown to bring complexes of histone deacetylases (HDACs), histone methyltransferases (HMTs), and DNA methyltransferases (DNMTs) to target genes (Hormaeche and Licht 2007).
Epigenetic gene silencing has always been envisaged as a local event, silencing genes one by one. However, recent data indicate that large regions of chromosomes can be coordinately suppressed by a process termed as long range epigenetic silencing (LRES) (Frigola, Song et al. 2006; Stransky, Vallot et al. 2006; Hitchins, Lin et al. 2007). LRES can span megabases of DNA. It involves formation of a broad heterochromatin regions accompanied by hypermethylation in the contiguous clusters of CpG islands. It is not clear if LRES is initiated by one critical gene target, then spreads to cloak innocent bystanders, analogous to large chromosome deletions, or if coordinated silencing of multiple genes occur (Clark 2007).

It is important to note that unlike the genetic alterations, gene silencing by epigenetic modifications is potentially reversible. Treatment by agents that inhibit cytosine methylation and histone deacetylation can initiate chromatin decondensation, demethylation and re-establishment of gene transcription of the silenced tumor suppressor genes that, in turn, might help to restore normal phenotype. On the other hand, it is likely that application of said therapeutics will provoke further deregulation of cancer cell transcriptome, and, possibly, further malignization.

Interestingly, both genetic and epigenetic changes can be seen as either the cause for the malignization or as consequences of the tumor progression. Despite tremendous efforts to connect gene expression profiles in tumor cells to particular milestones of the tumor progression (Ellis 2003; Lee and Thorgeirsson 2004; Liu 2004; Wang 2005; Driouch,
Landemaine et al. 2007; Henrickson, Hartmann et al. 2007), the exact mechanism of the
tumorigenesis remains unclear.

**Diversity of human gene expression**

Tumors arising from different organ/tissue systems of the tumor body are often
considered as different diseases due to observed differences in their cell phenotypes and
particular prognoses. However, the difference of one type of the tumor from another type
of the tumor might be explained by the underlying differences in the tissues of their
respective origins. Each of the normal human tissues has a distinct gene expression
pattern, with an exception of the constitutively expressing housekeeping genes shared
between all the tissues (Butte, Dzau et al. 2001). Recent studies shown that even the
housekeeping genes are not necessarily expressed at the same level across all tissues;
rather, each tissue seems to have a specific expression profile of housekeeping genes.
Interestingly, housekeeping genes are less compact and are evolutionary older than
tissue-specific genes, and they evolve more slowly in terms of both coding and core
promoter sequences (Zhang and Li 2004; Zhu, He et al. 2008). Housekeeping genes
primarily use CpG-dependent core promoters, whereas the majority of tissue-specific
genes possess neither CpG-islands nor TATA-boxes in their core promoters (Zhu, He et
al. 2008).

Generally, the distinctive patterns of gene expression in various tissue types are explained
by the differences in the functions of the resulting proteins in that particular tissue.
Typically, this phenomenon is illustrated by the tissue-specific expression of mRNA encoding for secreted hormone insulin, that is present only in Langerhans’ islets of the pancreatic gland (Bliss 1982). Another important note about the expression of the human genes is that many of them are alternatively spliced. Alternative splicing leads to the expression of multiple mRNA transcripts with different sets of exons joined together. It is a prevalent phenomenon that is observed in around half of the human genes (Modrek and Lee 2002). The use of alternative promoters and the alternative exons endings (splice donor sites) may also result in alternative mRNA transcripts. There is lot of variation in the alternative transcript expression patterns across the human tissues (Landry, Mager et al. 2003; Yeo, Holste et al. 2004). The alternative splicing code that controls and coordinates the transcriptome in complex multicellular organisms remains poorly understood. It has long been argued that regulation of alternative splicing relies on combinatorial interactions between multiple proteins, and that tissue-specific splicing decisions most likely result from differences in the concentration and/or activity of these proteins (Matlin, Clark et al. 2005; Singh and Valcarcel 2005).

Among human organs, the testis shows unusually diverse gene expression pattern. In part, this peculiarity of the testicular expression signature is explained by presence of the specific gene isoforms that are transcribed as a result of the chromatin remodeling and the activation of specialized transcription complexes activated during the differentiation program of spermatogenesis (Kimmins, Kotaja et al. 2004). Along with testis, brain also shows a complex and diversified transcription. On the other extreme, tissues with a
secretory function such as pancreas, salivary gland, and stomach show more specialized
and narrow gene expression patterns (Jongeneel, Delorenzi et al. 2005; Shyamsundar,
Kim et al. 2005). The expression signatures of the splicing factor encoding genes
correlate with the degree of the expression variation seen among the human tissues
(Grosso, Gomes et al. 2008). Respectively, brain and testis, the two tissues with highest
levels of alternative splicing events, have the largest number of expressed genes encoding
for splicing factors. Additionally, SR protein kinases and small nuclear ribonucleoprotein
particle (snRNP) proteins that modulate the association of core components of the
spliceosome with the pre-mRNA were identified as most highly differentially expressed
in the particular tissues (Grosso, Gomes et al. 2008). Concerning the brain-specific
splicing factor gene expression signature, the gene list includes the brain-splicing
regulators PTB1, NOVA1, A2bp1/FOX1, and members of the CELF/BRUNOL and
ELAVL families, the non-SR splicing regulator Y-box protein 1 and the core snRNP
protein SmN. The testis-specific signature included the splicing factor 3a subunit 2
(SF3A2) and the SR protein kinases 1 and 2 (SRPK1 and SRPK2) (Grosso, Gomes et al.
2008).

Nearly all of the cancer-upregulated genes with tissue-selective expression tend to show
their selective expression in tissues, which are different from the tissue of origin of
cancer (Axelsen, Lotem et al. 2007). Interestingly, different types of cancers, including
different brain cancers arising from the same lineage, showed differences in the tissue-
selective genes they overexpressed. Cancer cells ectopically expressing such genes may
acquire phenotypic modifications that contribute to cancer cell growth and metastasis (Axelsen, Lotem et al. 2007). Of all of the genes with tissue-selective expression, those selectively expressed in testis showed the highest frequency of genes that are overexpressed in at least two types of cancer (Axelsen, Lotem et al. 2007). Respectively, in many types of human cancers the phenomenon of coordinated up-regulation of so-called cancer/testis antigens is observed; these genes are particularly expressed only in testis and in tumor tissues. The specific expression of these genes in tumors and their restriction to testis tissue, make them good candidates for the cancer vaccines (Scanlan, Simpson et al. 2004).

**Bioinformatics & Cancer**

An increase of the amount of biotechnological datasets and the demand for the maintenance and analysis of such data lead to the development of specialized scientific discipline known as bioinformatics. It can be broadly described as the application of computer technology to explore biological processes through an analysis of the large-scale and multi-dimensional data generated from numerous sources. It encompasses organized storage of the data, development of tools to analyze the data and the actual analysis of data. Bioinformatics uses both informatic and statistical tools implemented to extract and analyze information (Wu 2001). The tools utilized to analyze the data depend on the type of data and nature of biological question addressed through the use of the particular dataset. In a broader context, bioinformatics can be viewed as the management
information system (MIS) with the scope on the biotechnology and the molecular biology (Umar 2004).

Bioinformatics became an inevitable research component of the molecular biology. Due to the accelerated generation of large-scale datasets by high throughput research platforms, cancer biology also got tightly bonded to informatics. Informatics occupies a special role in the translation cancer research to aggregate and perform integrative analysis on cancer biomarkers with ultimate goal of the improvement of both prevention and therapy. Bioinformatics is being used primarily to identify cancer biomarkers, their function and molecular mechanisms underlying cancer progression. The feasibility of the targeting of the biomolecule with the drug can also be assessed by structural bioinformatics, thus, reducing the drug development timeline (Wishart 2005). The addition of the biological function and/or the associated molecular cascade of the marker gene along with the quantified transcript or protein data significantly improved the relevance of biomarkers discovered. A number of novel bioinformatics methods, for example, gene set enrichment and pathway-based analysis, already reached the level of sophistication allowing to perform the functional analysis in semi-automated manner (Subramanian, Tamayo et al. 2005; Yi, Horton et al. 2006).

So far, expression analysis spotted a number of target genes, which laid the path for the development of novel cancer diagnostics and drugs. For example, AMACR (alpha-methylacyl CoA racemase) was at first identified as prostate cancer specific protein.
Currently, it is one of the prominent biomarkers for prostate cancer, with excellent specificity and sensitivity (Jiang, Woda et al. 2004). Other studies identified specific markers for progressing of the early stage cancers that permits individualized application of the treatments to patients through precise assessment of their prognoses. An example of such biomarker is EZH2 (enhancer of zeste homolog 2), a polycomb group protein that seems to be overexpressed in metastatic prostate cancer and specifically in the progressive tumors (Varambally, Dhanasekaran et al. 2002). Using multivariate model, this protein was later identified along with E-cadherin as powerful predictor of the disease recurrence following surgery (Rhodes, Sanda et al. 2003).

Functional analysis of biomarkers identified in large-scale datasets can be performed using a wide variety of bioinformatics tools. Gene annotation is the first and foremost requirement for the initiation of functional analysis. The annotation information is provided both by various biotechnological database management organizations like NCBI (Gene), EBI (Ensembl) and by independent groups like GeneCards (Safran, Solomon et al. 2002; Curwen, Eyras et al. 2004; Maglott, Ostell et al. 2007). Pathway information systems, for example, Kyoto encyclopedia of genes and genomes (KEGG) and Biocarta provide information concerning the molecular interactions of gene products in an organized fashion (Biocarta ; Aoki-Kinoshita and Kanehisa 2007). Association of genes with their corresponding ontological terms is another way of enhancing functional analysis. The gene ontology (GO) database was developed to standardize the gene and gene product attributes across the unending list of biological databases (Ashburner, Ball
et al. 2000). Database for annotation, visualization, and integrated discovery (DAVID) is a hybrid functional analysis tool combining all these systems in a single web interface (Dennis, Sherman et al. 2003; Huang, Sherman et al. 2009). Such integrative tools enhanced the functional analysis of the genes and/or proteins while providing a holistic perspective of large-scale datasets.

Data collection and maintenance remains an important aspect of clinical research as the aggregated data improves the statistical power to ascertain the results (Mathew, Taylor et al. 2007). The cancer biomedical informatics grid (caBIG™) is an ambitious initiative to connect cancer research centers in a network that allows to share biomedical data (Kakazu, Cheung et al. 2004). This co-operative bioinformatics system brings translational cancer research into the next level, changing the whole cancer research paradigm. Its seamless architecture defines compatibility standards based on ontologies, interfaces, data elements and information models for tool development enhancing the semantic interoperability between the research groups (Cimino, Hayamizu et al. 2009).

An integrating informatics system for the annotation and exchange of array based data, caArray, was recently developed for acquisition, dissemination and aggregation of interoperable array data as a part of caBIG™. caArray enhances the translational cancer research by supporting analysis of array data by tools and services on and off the grid. The resources developed on the grid are generally applicable beyond cancer research, promoting the comprehension of other complex diseases.
The analysis of microarrays and the cancer transcriptome

From its inception in the mid-90s, microarray analysis was instrumental in the discovery of clinically relevant knowledge by associating changes in gene expression (GE) patterns with particular pathological conditions. In a typical experiment, mRNA expression profiles are generated for thousands of genes across a collection of samples that belong to either one of two classes, for example, pathological specimen vs. healthy tissue controls. The registered changes in expression of individual genes can be ordered in a ranked list. However, extraction of a biological insight from the long lists of individual genes generated in typical microarray experiment remains a significant challenge.

A common approach to the analysis of GE data involves focusing on a handful of genes at the top and bottom of the ordered list (i.e., those showing the largest difference in the expression level in up and down regulatory fashion) and attempting an interconnection of these genes into the plausible biological network or pathway. This approach, however, has major limitations that may potentially nullify the experimental results. In particular, the required correction for multiple hypotheses testing (e.g. by Bonferroni or by Benjamini-Hochberg methods) may lead to the situation when no individual gene meets the threshold for statistical significance indicating that the relevant biological differences between two sample sets are modest relative to the technical noise. This outcome is not surprising since severe over-fitting is regarded as inevitable plague of the microarray studies where the typical number of observations (dozens or, rarely, hundreds of patients’ samples) is dramatically smaller than the number of measured end-points (typically 5 to
30 thousands of genes). Alternative approaches not controlled by multiple hypothesis tests often leave scientists with extensive lists of statistically significant genes that cannot be bound together by any common biological theme. Subsequently, an interpretation of these gene lists is left to subjective opinions of the expert biologists.

One of the current standard technologies of GE analysis that partially resolves the problem of over-fitting in microarray study design is Significance Analysis of Microarrays (SAM) (Tusher, Tibshirani et al. 2001). Based on stochastic procedure of iterative cross-validation of false discovery rate, it allows eliminating a majority of spurious and potentially non-reproducible findings. This technology has already successfully replaced conventional t-tests. However, it remains a subject of the commonest drawback in all single-gene methods: statistically significant genes often lack biological meaning. Additionally, SAM, on par with other single-gene based methods of GE analysis, may miss pathway wide effects. For example, effects of the coordinated 20% increases in the expression levels of all genes that belong to the same metabolic pathway may be masked by spurious 500% decrease in an expression of a single gene with redundant function.

Identification of the pathological mechanisms underlying differentially expressed gene lists may be facilitated by pre-grouping these genes into the relevant biological pathways or networks. Thus, instead of the traditional “bottom-up” approach that relies on post factum integration of GE lists with existing literature describing the potential biological
roles of individual genes, enrichment-based analysis of gene sets may be executed. In this type of knowledge-based analysis (KBA), previously collected gene lists are used for the ranking of the functional categories or pathways and subsequent evaluation of the ranked pathways according to their gene enrichment levels. The significantly enriched pathways are further explored as primary biological themes. For this purpose, the gene sets known as signature databases were generated in accordance with the prior biological knowledge accumulated in wet-lab experiments or using computational models that identify functionally similar genes by their sequence or structural similarities.

The pioneering tool for enrichment-based analysis of microarray datasets named GSEA (Gene Set Enrichment Analysis) has been developed by a group of bioinformaticians at MIT (Subramanian, Tamayo et al. 2005). Though it’s been more than three years since the method was published, GSEA have just started to gain popularity among microarray researchers, mostly due to the time required for generation of extensive knowledge basis justifying the grouping of genes into gene sets. Today, an innovative idea of GSEA has reached its maturity. Particularly, several gene signature databases became available, including those with gene sets grouped according to their chromosomal locations, an a priori placement within certain molecular pathways, a commonality of regulatory mechanisms, and an annotation under the same gene ontology terms. These gene signature databases could be used for extracting novel and often unexpected knowledge about pathologic processes and disease mechanisms. Some of the examples of successful GSEA applications include studying differential gene expression associated with

To gain the comprehension of the underlying mechanism of tumor development, particularly the regulation of cell growth, apoptosis and genomic stability (Hanahan and Weinberg 2000), a number of high-throughput microarray initiatives have been started. Cancer cell lines have served as the primary experimental system for exploring cancer molecular biology and pharmacology. For instance, the Nation Cancer Institute developed an NCI-60 panel consisting of 60 diverse human cancer cell lines. These cell lines underwent both selected gene sequencing and gene expression profiling in order to facilitate screenings for drugs aimed at cancer therapy (Ikediobi, Davies et al. 2006; Shoemaker 2006). DNA microarrays expression analysis of these heterogeneous cell lines provided initial snapshots of the genes and related molecular pathways pertaining to the malignization in a broad sense and to the response to specific chemotherapeutic drug treatments (Efferth 2005; Shankavaram, Reinhold et al. 2007).

The contribution of microarray based studies to cancer research is not limited to the study of NCI-60 cell lines. For instance, microarray profiling is widely used for an identification of aberrant chromosomal regions and expression signatures of various primary tumors (Buness, Kuner et al. 2007; Xu, Geman et al. 2007). Numerous high-throughput experiments generated quantitative profiles of global gene expression for most of the common forms of cancer. Based on these profiles, the diagnostic signatures
were generated for specific cancer types, in many cases producing substantial insights into the tumor biology. Meta-analysis of these datasets has been attempted to extract common or generic cancer signature pattern (Xu, Geman et al. 2007). The utility of gene expression profiling for clinical settings was demonstrated by bringing the breast and lymphoma tumor signatures into the process of the clinical decision making (van de Vijver, He et al. 2002; Dave, Wright et al. 2004). Proven to be efficient means for the cancer researchers, DNA microarrays became a staple of cancer transcriptome studying labs. For example, study of Golub et al demonstrated the feasibility of cancer classification based solely on gene expression monitoring and suggested a general strategy for discovering and predicting cancer classes, independent of previous biological knowledge (Golub, Slonim et al. 1999).

The increasing number of the microarray datasets in the research of cancer and other diseases, demanded a central system to congregate this valuable profile data. The National Central for Biotechnology Information (NCBI) took an initiative to develop the database called Gene Expression Omnibus (GEO), which archives and freely distributes raw microarray files as well as other forms of large-scale data generated by the scientific community (Barrett and Edgar 2006; Barrett, Troup et al. 2007). Until recently, ArrayExpress was an analogous system to GEO developed by the European Bioinformatics Institute (EBI) (Parkinson, Kapushesky et al. 2007). Nowadays, GEO has emerged into the principal repository for microarray data storage and retrieval that covers an assortment of microarray platforms. Standard analyses of the cancer array dataset
stored in these repositories are maintained in a database known as Oncomine. This
database enables one to query for the specific up and down regulated genes across
microarray datasets according to the tumor or tissue types of interest (Rhodes, Kalyana-
Sundaram et al. 2007).

Public availability of cancer microarray datasets lead to the development of the methods
to derive common cancer signatures across datasets (Xu, Geman et al. 2007). Integrative
analyses are also performed combining sub-datasets in order to borrow information and
using statistical and clustering analysis techniques to derive relevant markers (Golub,
Slonim et al. 1999; Tusher, Tibshirani et al. 2001). The process of the combination of
expression profiles resulted from independent microarray experiments often called the
meta-analysis, analogous to meta analysis in clinical research that is directed to test single
hypothesis using material from multiple studies (Rhodes and Chinnaiyan 2005). An
example of meta-analysis of microarray datasets is the study of Wren, 1999 that included
all publicly available GEO two-channel human microarray datasets (a total of 3551
individual profiling experiments). This study was conducted to identify genes with
recurrent, reproducible patterns of co-regulation across different conditions. Patterns of
co-expression were divided into parallel (i.e. genes are up and down-regulated together)
and anti-parallel. Several ranking methods to predict a gene's function based on its top 20
co-expressed gene pairs were compared. The data matrix describing differential
expression of the human genes with unknown function was made available to the
scientific community (Wren 2009).
Pertinent to cancer, a handful of meta-analysis studies were performed. In the study of Alles et al., 2009 gene expression values from five large microarray datasets describing breast carcinoma expression patterns relative to ER status of the tumors were subjected to Gene Set Enrichment Analysis (GSEA). As expected, the expression of the direct transcriptional targets of the ER was muted in ER- tumors, but the expression of genes indirectly regulated by estrogen was enhanced. An enrichment of independent MYC- and E2F-driven transcriptional programs was also observed. A conclusion concerning increased transcriptional activity of MYC as a characteristic of basal breast cancers capable of mimicking a large part of an estrogen response in the absence of the ER was made, thus, suggesting a mechanism by which these cancers achieve estrogen-independence and providing a potential therapeutic target for this poor prognosis subgroup of breast cancer (Alles, Gardiner-Garden et al. 2009). Another study compared expression data from a diverse collection of 9 breast tumor array datasets generated on either cDNA or oligonucleotide arrays from the Oncomine database and identified genes that were universally up or down regulated with respect to ER+ versus ER- tumor status (Smith, Saetrom et al. 2008). Liang et al. analyzed miRNA target genes across Oncomine datasets profiling gene expression in the patients with lung adenocarcinoma (AD) and squamous cell carcinoma (SCC), two major histologic subtypes of lung cancer. Expression of a minimal set of 17 predicted miR-34b/34c/449 target genes was identified from a training set to classify 41 AD and 17 SCC, and correctly predicted in average 87% of 354 AD and 82% of 282 SCC specimens from total 9 independent published datasets. Expression of this signature in two published datasets of epithelial cells obtained at
bronchoscopy from cigarette smokers, if combined with cytopathology of the cells, yielded 89-90% sensitivity of lung cancer detection and 87-90% negative predictive value to non-cancer patients (Lagarde, Ver Loren van Themaat et al. 2008). Romualdi et al. used meta-analysis to define the gene expression signature of rhabdomyosarcoma, a highly malignant soft tissue sarcoma (Romualdi, De Pitta et al. 2006), and demonstrated a general downregulation of the energy production pathways, suggesting a hypoxic physiology for rhabdomyosarcoma cells.

Meta analysis of cancer microarrays was also useful in identifying the aberrant genomic loci in the tumor cells. Eight datasets including more than 1200 breast tumors were investigated to identify chromosomal regions and candidate genes possibly causal for breast cancer metastasis. By utilizing Gene Set Enrichment Analysis chromosomal regions were ranked according to their relation to metastasis. Over-representation analysis identified regions with increased expression for chromosome 1q41-42, 8q24, 12q14, 16q22, 16q24, 17q12-21.2, 17q21-23, 17q25, 20q11, and 20q13 among metastasizing tumors and reduced gene expression at 1p31-21, 8p22-21, and 14q24. Analysis of genes with extremely imbalanced expression in these regions, DIRAS3 at 1p31, PSD3, LPL, EPHX2 at 8p21-22 and FOS at 14q24 pinpointed them as candidate metastasis suppressor genes. Potential metastasis promoting genes list included RECQL4 at 8q24, PRMT7 at 16q22, GINS2 at 16q24, and AURKA at 20q13 (Thomassen, Tan et al. 2009). An association was also established between the tumor stage of the breast carcinoma and the recognized genomic regions identified by the meta analysis applied to
12 independent human breast cancer microarray studies comprising 1422 tumor samples (Buness, Kuner et al. 2007).

**Cancer – A systems biology perspective**

Understanding the biology at the system level should improve the way a medical condition is perceived and thus affects the methods to prevent, diagnose and treat it (Ahn, Tewari et al. 2006). The capability of current biotechnological methods to extract global gene expression patterns along with levels of proteins and metabolites from the same sample opened avenues for the conception of network level data. Sequencing of genomes, high-throughput data generation technologies combined with organized informatic systems further enabled the collection of comprehensive datasets providing the system-wide overview of the complex biological objects (Kitano 2002). The elucidation of the network of biological networks with their dynamic interactions forms the basic premise for systems biology in general. System-level models require quantification of the network elements as they change over time in response to various perturbations, therefore, implying the computational modeling of the dynamic interactions (Laubenbacher, Hower et al. 2009). In the context of cancer, such models would be helpful in deciphering the relevant biological network underlying particular type of malignancies.

Cancer is an outcome of the malfunctioning in the cell system with inherent genomic instability, thus it is identified as a systems biology disease. The progress in the treatment of cancer can be drastically accelerated, if the ongoing research is done in the systems biology perspective, instead of classic molecular biology based approaches (Hornberg,
Apart of intra-cellular interactions as represented by internal signal transduction in cell, inter-cellular interactions also play a vital role in the development of cancer. The horizon of system-level study of cancer should be expanded to the tissue level considering the role of inter-cellular interactions in the realistic comprehension of cancer. Challenges lie in various disciplines of sciences including mathematics, physics, chemistry and biology to develop models, generate precise data and calculate the interactions between the biological network components. Last but not the least, the contribution of informational technology is inevitable, not only for data management but also for the analyzing the simulated models of individual cancer cells and interactions within their proliferating society (cancer tissue).
2. Genome wide discrimination of normal and tumor samples

Rationale
The current study quantitatively estimates the relative importance of global and local features of gene expression regulation landscape in the process of tumor development. The work is based on the hypothesis that the cancer could be viewed as an attractor state.

Background
To date, most of the high-throughput studies of the gene expression studies are still focused on elucidation of the discriminatory gene signatures reflecting key regulatory processes participating in establishing cell phenotypes (J. Wang et al. 2003; Ben-Dor et al. 2000; Furge et al. 2004). On the other hand, a change in a cell phenotype requires coordinated interaction of a variety of genes that determine the functional identity of the cell within a population of cells (Bar-Yam et al. 2009). This notion implies an understanding that a given cell type could be represented as a dynamic system that can assume different states, thus, occupying a specific position in the multidimensional phase space spanned by the different genes (Tsuchiya, Piras et al. 2009; Tsuchiya, Selvarajoo et al. 2009; H. H. Chang et al. 2008).
The ability of gene regulatory circuits to assume multiple equilibrium states was first proposed by Max Delbruck in 1948 (cited according to S. Huang et al. 2009). In terms of dynamics, this specific position of equilibrium is called an ‘attractor’, i.e. a “stable” position characterized by a specific pattern of gene expression levels that determines the particular kind (differentiation pattern) of the cell population (S. Huang 2009). Multiple attractor states can exist. The current stable state of the cells depends on the history of the past states of cell, implicating the importance of epigenetic mechanisms in such a context. The attractor states are robust, distinct and possess self stabilizing properties. The gene expression pattern associated with a particular state could be maintained even after the original stimulus that placed the cell in the current attractor state has been removed (S. Huang 2009). Of course, the attractor state is a property of the cell population, so its location in the phase space corresponds to the average expression levels for the millions of single cells over thousands of genes. When individual gene expression levels are measured, cells could be different for each other, thus, demonstrating intra-population variance. In this sense, attractor state viewed as an analogy to the definition of the temperature in statistical mechanics that allows for evaluation of the intrinsic differences between the components of the system (Huang 2009).

Earlier studies have indicated that the differentiation destinies of the progenitor cells could be defined as high dimensional attractor states of the underlying molecular networks (H. H. Chang et al. 2008; S. Huang et al. 2007). Particularly, a study of the differentiation trajectories of blood stem cells demonstrated that specific differentiated
cell types behave as attractors (A. C. Huang et al. 2009). The same group provided some evidences of an analogous behavior of the cancer cells that are to be considered as located at the ‘periphery’ of the correspondent normal cell attractor for the same kind of tissue (S. Huang & Ingber 2006; S. Huang et al. 2009; Yuchun Guo et al. 2006). Although cancer was proposed as an attractor state of a cell as early as 1971 (S Kauffman 1971), a path to verify such a notion has been paved only recently, with an advent of the genomic technologies.

Under “attractor” paradigm, cell population is considered as a dynamic system that could be attracted to one or another “stable” state by transition that implies extensive mutual regulation of all elements of cell’s genome. This is in striking contrast with the traditional idea of a division of the mRNA transcripts into those generated by ‘housekeeping’ and ‘tissue-specific genes’, where a set of the master genes is responsible for the switch between different phenotypes. In their seminal paper Bar-Yam and colleagues describe this dichotomy. Particularly, the definitions for a ‘democratic’ (no master genes, all genes act as mutual regulators going toward a global attractor state) and an ‘autocratic’ (few master genes drive the differentiation process) regulatory landscape were introduced (Bar-Yam et al. 2009).

A possible middle ground between “democratic” and ‘autocratic” regulatory landscapes may be described as a general attractor-like behavior of the regulatory machinery with some local ‘vantage points’ representing genes most sensitive to dynamical changes of
the system. Recent study of Tsuchiya et al demonstrated biphasic nature of the cellular response to innate immune stimuli involving an acute-stochastic mode consisting of small number of sharply induced genes and a collective mode where a large number of weakly induced genes adjust their expression levels to novel “stable” state. We hypothesize that similar regulatory scenario takes place during tumor development.

**Hypothesis**

Cancer is an attractor state. Normal cell can become cancerous and progress toward malignant phenotype using an intermediate regulatory framework that combines both local and global regulatory features.

Here we propose to perform a quantitative estimation of the relative importance of global and local features of gene expression regulation landscape in the process of tumor development through an analysis of publicly available microarray data.

**Materials and Methods**

Microarray datasets were extracted from the NCBI Gene Expression Omnibus as raw data (.CEL files) by selecting the data using Oncomine browser (Barrett & Edgar 2006; Barrett et al. 2007; D. R. Rhodes et al. 2007). To exclude cross-platform variability factors, only the datasets profiled using Affymetrix oligonucleotide arrays were chosen. The chosen datasets were classified into the following three categories: 1) Two-point datasets describing paired normal and tumor tissue samples collected from the same individual (N=8); 2) Two-point datasets describing a group of normal and a group of
tumor samples collected from the same tissue type across a number of subjects (N=9); 3) Multi-point datasets describing three or more physiological groups of normal and tumor samples collected from same subject or across a number of subjects (N=7). The detailed descriptions of these datasets are given in the tables 1, 2 and 3 for each of categories, respectively.

Table 1: The table describes the attributes of two-point datasets describing paired normal and tumor tissue samples collected from the same individual.

<table>
<thead>
<tr>
<th>GEO ID</th>
<th>Sample source</th>
<th>Number of samples</th>
<th>Total Number of transcripts extracted; Total number of transcripts significant by MW test</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE5764</td>
<td>Invasive ductal (IDC) and lobular breast (ILC) carcinomas in postmenopausal patients</td>
<td>IDC (N=5) Normal ductal (N=5) ILC (N=5) Normal lobular (N=5)</td>
<td>54675; 2278 54675; 988</td>
<td>(Turashvili et al. 2007)</td>
</tr>
<tr>
<td>GSE2514</td>
<td>Pulmonary adenocarcinoma and adjacent lung tissue</td>
<td>Lung AdCa (N=20) Normal (N=19)</td>
<td>12625; 5857</td>
<td>(Stearman et al. 2005)</td>
</tr>
<tr>
<td>GSE7670</td>
<td>pulmonary adenocarcinoma and adjacent lung tissue</td>
<td>Lung AdCa (N=27) Normal (N=27)</td>
<td>22283; 8599</td>
<td>(Su et al. 2007)</td>
</tr>
<tr>
<td>GSE6344</td>
<td>Renal cell carcinoma (RCC)</td>
<td>Stage 1 tumor (N=5) Stage 2 tumor (N=5) Stage 1 normal (N=5) Stage 2 normal (N=5)</td>
<td>44760; 23701</td>
<td>(Gunz et al. 2007)</td>
</tr>
<tr>
<td>GSE781</td>
<td>Renal clear cell carcinoma (RCC)</td>
<td>Tumor (N=7) Normal (N=7)</td>
<td>44760; 11119</td>
<td>(Lenburg et al. 2003)</td>
</tr>
<tr>
<td>GSE6631</td>
<td>Head and neck squamous cell carcinoma (HNSCC)</td>
<td>Tumor (N=22) Normal (N=22)</td>
<td>12625; 2880</td>
<td>(Kuriakos e et al. 2004)</td>
</tr>
<tr>
<td>GDS1665</td>
<td>papillary thyroid carcinoma (PTC)</td>
<td>Tumor (N=9) Normal (N=9)</td>
<td>54675; 13985</td>
<td>(H. He et al. 2005)</td>
</tr>
</tbody>
</table>
Analysis was performed by R data analysis packages of Bioconductor. Affy package was used for the data processing and normalization (Reimers & Carey 2006; Gregory Alvord et al. 2007; Gautier et al. 2004). Perl scripting has been used to automate the analysis pipeline. The gene expression data were background corrected, normalized and the summarized expression values were calculated using Robust Multichip Average (RMA) method that consists of three steps: a background adjustment, quantile normalization and, finally, summarization (R. A. Irizarry et al. 2003). The expression values for individual genes in each of the cancer and normal samples were subjected to non-parametric Mann-Whitney test that extracted the transcripts with significant (P <0.05) differential expression (Mann & Whitney 1947). The global and specific expression distances (DGlobal and DSpecific) were calculated based on the whole transcripts on the chip and significantly differentially expressing transcripts as selected by Mann-Whitney test, respectively. The distance between two samples i and j corresponds to : Dij = 1 – Rij, where Rij is the Pearson correlation coefficient between the vectors correspondent to i and j samples and having as dimensions the entire set of transcripts (DGlobal) or only the gene with statistically significant expression differences (DSpecific). These distance metrics were previously adopted for the dynamical characterization of microarray data and statistical analysis of microarrays in other studies (Hayden et al. 2009; H. H. Chang et al. 2008; Tsuchiya, Selvarajoo et al. 2009).
Table 2: The table describes the two-point datasets comprised of normal and tumor samples collected from the same tissue type across a number of subjects.

<table>
<thead>
<tr>
<th>GEO ID</th>
<th>Sample source</th>
<th>Number of samples</th>
<th>Total Number of transcripts extracted; Total number of transcripts significant by MW test</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE6791</td>
<td>Gene Expression Profiles of HPV-Positive and -Negative Head/Neck Cancers</td>
<td>Normal Head/Neck (N=14) Head/Neck Cancer (N=42)</td>
<td>54675; 35778</td>
<td>(Pyeon et al. 2007)</td>
</tr>
<tr>
<td>GSE3678</td>
<td>Papillary thyroid carcinoma</td>
<td>Normal Thyroid (N=7) Papillary thyroid carcinoma (N=7)</td>
<td>54675; 5617</td>
<td>--</td>
</tr>
<tr>
<td>GSE3524</td>
<td>Oral squamous cell carcinoma (OSCC)</td>
<td>OSCC (N=16) Normal (N=4)</td>
<td>22283; 5757</td>
<td>(Toruner et al. 2004)</td>
</tr>
<tr>
<td>GSE10797</td>
<td>Transcriptomes of breast epithelium and stroma in normal reduction mammoplasty and invasive breast cancer patients.</td>
<td>Normal breast epithelium (N=5) Invasive breast cancer epithelium (N=28)</td>
<td>22277; 2491</td>
<td>(Casey et al. 2009)</td>
</tr>
<tr>
<td>GSE12345</td>
<td>Global gene expression profiling of human pleural mesotheliomas</td>
<td>Normal pleural tissue (N=8) Mesothelioma tissue (N=8)</td>
<td>54675; 5995</td>
<td>--</td>
</tr>
<tr>
<td>GSE12452</td>
<td>mRNA expression profiling of nasopharyngeal carcinoma</td>
<td>Normal nasopharyngeal tissue (N=10) nasopharyngeal carcinoma (N=31)</td>
<td>54675; 15383</td>
<td>(Dodd et al. 2006; Sengupta et al. 2006)</td>
</tr>
<tr>
<td>GSE14762</td>
<td>Renal Cell Carcinoma: Hypoxia and Endocytosis</td>
<td>Normal renal tissue (N=12) Renal carcinoma (N=10)</td>
<td>54675; 18501</td>
<td>(Y. Wang et al. 2009)</td>
</tr>
</tbody>
</table>
Table 3: The table describes the datasets with three or more physiological groups of normal and tumor samples collected across the same subject or a number of subjects

<table>
<thead>
<tr>
<th>GEO ID</th>
<th>Sample source</th>
<th>Number of samples</th>
<th>Total Number of transcripts extracted; Total number of transcripts significant by MW test</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE1420</td>
<td>Barrett's esophagus, Barrett's-associated adenocarcinomas and normal esophageal epithelium</td>
<td>Normal (N=8)</td>
<td>22283; 6552</td>
<td>(Kimchi et al. 2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Barrett's esophagus (N=8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Barrett's-associated adenocarcinoma (N = 8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE3325</td>
<td>Benign prostate, primary and metastatic prostate cancer samples</td>
<td>Benign prostate (N=6)</td>
<td>54675; 20667</td>
<td>(Sooryanarayana Varambally et al. 2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>primary prostate cancer (N=7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>metastatic prostate (N=6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Normal pancreas, chronic pancreatitis and pancreatic adenocarcinoma (microdissected)</td>
<td>Normal pancreas (N=5)</td>
<td>7129; 2289</td>
<td>(Logsdon et al. 2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chronic pancreatitis (N=5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pancreatic adenocarcinomas (n=10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE3167</td>
<td>Normal Bladder, superficial transitional cell carcinoma(sTCC), STCC with carcinoma in situ, metastatic transitional cell carcinoma, normal cystectomy and cystectomy with CIS</td>
<td>Normal Bladder (N=9)</td>
<td>22283; 13861</td>
<td>(Dyrskjøt et al. 2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>sTCC (N=15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>sTCC with CIS (N=13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>mTCC (N=13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cystectomy Normal(N=5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CIS (N=5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE6919</td>
<td>The Normal Prostate Tissue free of any pathological alteration,. Metastatic Prostate Tumor, Primary Prostate Tumor, Normal Prostate Tissue Adjacent to Tumor</td>
<td>Normal Prostate Tissue free of any pathological alteration (N=17)</td>
<td>37757; 18973</td>
<td>(Yan Ping Yu et al. 2004; Chandran et al. 2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metastatic Prostate(N=25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Primary Prostate (P=59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal Prostate Tissue Adj to Tumor (N=62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE6764</td>
<td>Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma</td>
<td>Normal liver (N=10)</td>
<td>54675; 19250</td>
<td>(Wurmbach et al. 2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dysplastic liver tissue (N=17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cirrhotic liver tissue (N=13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Very early HCC (N=8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Early HCC (N=10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Advanced HCC (N=7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Very Adv HCC (N=10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE10971</td>
<td>Gene expression data from non-malignant fallopian tube epithelium and high grade serous carcinoma.</td>
<td>Normal controls (N=12)</td>
<td>54675; 15988</td>
<td>(Tone et al. 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BRCA-1/2 mutation carriers (N=12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>High grade serous carcinoma (N=13)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Principal Component Analysis (PCA) was performed on the cancer microarray expression datasets based on the distance parameters (Roden et al. 2006). In this analysis, each sample is described by four distance based descriptors reflecting the average distance of each sample from i) cancer sample space (DC) and ii) normal sample space (DN) in both global and specific frames, therefore, producing following variables: DCglobal, DNglobal, DCspecific and DNspecific. PCA was performed using R on each of the datasets separately, in the four dimensional space represented by these parameters. The structure of correlations emerging from the analysis of the variable loadings on the extracted components allowed for a straightforward quantification of some relevant topological features of the analyzed systems.

**Results and Discussion**

**a) Modeling strategy**

The discrimination between a tumor and a normal sample can be achieved using both a summed expression change involving the entire set of mRNAs (DGlobal) and a summed expression change of the functionally important genes specifically involved in the development of the tumor state (DSpecific). In the case of the “democratic” regulatory landscape (no preferred vantage points, or particular mRNAs, specifically responding to the change of the physiological state), the discrimination would be achieved by DGlobal, while gene signature-based (DSpecific) distances should better reflect “autocratic”
landscape with a profound changes in expression of master (or signature) genes while the
great portion of mRNAs remain unaffected. In the latter case, the correlation between
genome-wide (DGlobal) and signature-based (DSpecific) distances should not be
substantial.

In case of an intermediate scenario, - a middle ground between “democratic” and
“autocratic” regulatory landscapes, - the discrimination between tumor and normal
sample calculated using DSpecific should be consistently better than the discrimination
achieved using DGlobal. However the two metrics should correlate, thus,
demonstrating both the existence of a global attractor correspondent to the cell phenotype
and reflecting the change of entire genome expression and the most influential roles for a
specific set of the tumorigenesis-related genes.

The most natural metrics for estimating the distance between expression profiles of two
biological samples is based on the Pearson correlation coefficient: the level of
concordance of any two expression vectors correspondent to two different biological
samples, x and y with n dimension (n = genes) and mean values of expressions \( \bar{x} \) and \( \bar{y} \)
corresponds to their mutual Pearson correlation, \( r(x, y) \) defined as:

\[
r(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} X_iY_i}{\sqrt{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i^2}} = \frac{X \cdot Y}{|X||Y|} = \cos \theta, \quad \text{Eq. 1}
\]
where \( X = (x_1 - \bar{x}, x_2 - \bar{x}, ..., x_n - \bar{x}) \) \( Y = (y_1 - \bar{y}, y_2 - \bar{y}, ..., y_n - \bar{y}) \) correspond to the differences from the mean expression of each gene in the X and Y sample respectively and \( \theta \) is the angle between two expression vectors. Geometrically, Eq. 1 shows the correlation coefficient can be viewed as the cosine of the angle on n-dimensional space between the two vectors of data which have been shifted by the average to have mean zero. Angle \( \theta \) is a measure of the differences between the two vectors and consequently of the difference in expression pattern of the two sample, when \( \theta = 0 \) (and consequently \( r = 1.0 \)) the two expression patterns are completely coincident, and the two vectors are parallel. In the case of \( r = 1 \) (and consequently \( \theta = 90 \) degrees) the two expression vectors are orthogonal, i.e. the expression patterns of the two samples are each other independent.

The measure \( D_{ij} = 1 - R_{ij} \) with \( R = \) Pearson correlation coefficient between i and j samples can be considered as a distance between samples. This distance could vary from 0 (\( R = 1 \)) reflecting the perfect resemblance of the two samples to 1 corresponding to maximal possible distance between two states (absence of correlation). In the case when samples are picked from two different sub-groups -- normal (N) and cancer (C) -- for each sample j analyzed two different descriptors \( D_{Cj} \) and \( D_{Nj} \) can be computed corresponding to the average distance of sample j from the spaces occupied by cancer (\( D_{Cj} \)) and normal (\( D_{Nj} \)) samples. Thus if (i) corresponds to a cancer sample \( D_{Ci} \) will be the average of all the pairwise distances of (i) vector from all the other cancer samples vectors, and consequently \( D_{Ni} \) the average of all the distances of (i) from the non-cancer samples.
When the distance is computed only over the previously extracted differentiating gene signature defined as a set of genes with expression values significantly different between Cancer and normal subgroups by Mann-Whitney test, two similarly defined but gene signature-specific distance indexes (DCspecific, DNspecific) were obtained. In entirety, four descriptors were defined for every specific sample on each dataset:

- **DCglobal**: Genome-wide distance from cancer sample space to the particular sample
- **DNglobal**: Genome-wide distance from normal sample space to the particular sample
- **DCspecific**: Signature based distance from cancer sample space to the particular sample
- **DNspecific**: Signature based distance from normal sample space to the particular sample

b) **Assessment of the global and signature-specific gene expression distances for two-point (Normal-Tumor) datasets**

In this study we used a total of 17 two-point datasets represented by normal and tumor gene expression profiles. Paired datasets (tumor and normal samples derived from the same individual) and populational datasets (tumor and normal samples were collected across a number of subjects) were considered separately. Eight paired and nine population datasets profiled using the Affymetrix platforms were chosen for the two-
point (normal-tumor) analysis (Tables 1, 2). For each dataset, the global and specific expression distances were calculated based either on the all probes present on the chip and passing the detection call (DNglobal and DCglobal) or on the genes highlighted as significantly differentially expressed according to Mann-Whitney test (DNspecific and DCspecific).

In both paired and population datasets, DC (global, specific) was greater than DN (global, specific) for most of the normal samples. The reverse was true, i.e. DC (global, specific) is less than DN (global, specific) for the tumor samples. Such a relation provides a basis an unbiased classification scheme, given a sufficiently relevant population of samples is achieved. Figure 1 depicts the four parameters as panels of paired plots for the lobular and ductal breast carcinoma dataset. The clear classification of the cancer and tumor samples using the complete chip data (global) using a simple metric like distance illustrate the differentiating power of the overall transcription. Moreover, ranking of the datasets based on global and specific distances of the tumor sample from the normal center were very similar, albeit not identical (Table 4). The conservation of global and specific distances across the datasets adds to the credibility of using this metric for diagnostic purpose.
Figure 1: The illustration depicts the distance parameters derived for the paired breast carcinoma dataset (GSE5764) using a paired-plot panels. Panel A and Panel B represent the distance parameters for the lobular and ductal carcinoma within this dataset. Each point depicts individual cancer (red) or normal (blue) sample, line reflects a linear fit for each group of samples.
Table 4: Ranknings of the tumor malignancy potential according to the relative distance to the Normal Sample Space (two-point paired datasets) 1 – lowest; 9 – highest

<table>
<thead>
<tr>
<th>DATASET</th>
<th>Mean (DGlobal) from individual tumor samples to the Normal center</th>
<th>Mean (DSpecific) from individual tumor samples to the Normal center</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE2514 (pulmonary adenocarcinoma)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GDS1665 (papillary thyroid carcinoma)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>GSE781 (RCC)</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>GSE6344 (RCC stage 2)</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>GDS2520 (HNSCC)</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>GSE6344 (RCC stage 1)</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>GSE7670 (pulmonary adenocarcinoma)</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>GSE5764 (ductal breast cancer subset)</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>GSE5764 (lobular breast cancer subset)</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

In case when distances were calculated using DNglobal, in all studied paired data, tumors were further away from the Normal Sample Space than the control samples with normal histology (Table 5). On average, for normal samples the distance to the Normal Space defined by DGloba was 0.047+/–0.045 as compared to 0.080+/– 0.034 for Tumor samples (P < 0.038) in paired datasets. Distances between individual Normal samples and the Normal Space defined by DSspecific were also significantly different from that calculated for Tumor samples (Normal: 0.044+/–0.034; Tumor: 0.138 +/- 0.063, P < 0.001). All metrics were heavily correlated to each other. This correlation indicates
strong attractor-like behavior; the discussion on this would be continued in the PCA results section. Here it is important to stress that signature-based and genome-wide approaches allow for the same level of discrimination efficiency of the data sets.

Table 5: Mean, Standard Deviation and Variance calculated for Global and Specific Distances from individual samples to the Normal Sample Space of the paired datasets

<table>
<thead>
<tr>
<th>DATASET</th>
<th>Mean +/- SD variance (DNglobal) from individual normal samples to the Normal Sample Space</th>
<th>Mean +/- SD variance (DNglobal) from individual tumor samples to the Normal Sample Space</th>
<th>Mean +/- SD variance (DNspecific) from individual normal samples to the Normal Sample Space</th>
<th>Mean +/- SD variance (DNspecific) from individual tumor samples to the Normal Sample Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE5764 (ductal breast cancer subset)</td>
<td>0.0989+/-.0111, 0.0001231</td>
<td>0.1134+/-.0196, 0.0003861</td>
<td>0.0634+/-.00595, 0.00003547</td>
<td>0.1827+/-.02951, 0.000870855</td>
</tr>
<tr>
<td>GSE5764 (lobular breast cancer subset)</td>
<td>0.1449+/-.0084, 0.0000704</td>
<td>0.1496+/-.0389, 0.00151395</td>
<td>0.1092+/-.01037, 0.00010758</td>
<td>0.2788+/-.0873, 0.00762137</td>
</tr>
<tr>
<td>GSE2514 (pulmonary adenocarcinoma)</td>
<td>0.0113+/-.0015, 0.0000023</td>
<td>0.0407+/-.0199, 0.000399112</td>
<td>0.0138+/-.00211, 0.0000044</td>
<td>0.0688+/-.03296, 0.001086227</td>
</tr>
<tr>
<td>GSE7670 (pulmonary adenocarcinoma)</td>
<td>0.0399+/-.0104, 0.000107128</td>
<td>0.0841+/-.0285, 0.000814786</td>
<td>0.0483+/-.01129, 0.000127647</td>
<td>0.1417+/-.04826, 0.002329823</td>
</tr>
<tr>
<td>GSE781 (RCC)</td>
<td>0.0187+/-.0008, 0.0000646</td>
<td>0.0624+/-.0087, 0.0000751</td>
<td>0.0234+/-.0128, 0.0001639</td>
<td>0.1247+/-.0158, 0.0002513</td>
</tr>
<tr>
<td>GDS2520 (HNSCC)</td>
<td>0.0577+/-.0015, 0.000227314</td>
<td>0.0742+/-.0141, 0.000197704</td>
<td>0.0789+/-.0186, 0.000348429</td>
<td>0.1362+/-.02979, 0.000887682</td>
</tr>
<tr>
<td>GDS1665 (PTC)</td>
<td>0.0184+/-.0002, 0.0000039</td>
<td>0.0407+/-.0133, 0.0001773</td>
<td>0.0168+/-.00210, 0.0000044</td>
<td>0.0785+/-.0276, 0.0007636</td>
</tr>
<tr>
<td>GSE6344 (RCC stage 1)</td>
<td>0.0216+/-.0019, 0.0000038</td>
<td>0.0802+/-.0058, 0.0000337</td>
<td>0.0219+/-.00208, 0.0000043</td>
<td>0.1213+/-.00702, 0.0000494</td>
</tr>
<tr>
<td>GSE6344 (RCC stage 2)</td>
<td>0.0196+/-.0022, 0.0000048</td>
<td>0.0758+/-.0096, 0.0000926</td>
<td>0.0201+/-.00265, 0.0000070</td>
<td>0.1098+/-.01424, 0.0002028</td>
</tr>
</tbody>
</table>

Similar to that in paired datasets, by DNglobal, tumors in all the population datasets were further away from the Normal Sample Space than the control samples with normal histology (Table 6). On average, for normal samples the distance to the Normal Space
defined by DGlobal was 0.0520 +/- 0.021 as compared to 0.095 +/- 0.032 for Tumor samples (P < 0.012). Distances between individual Normal samples and the Normal Space defined by DSpecific were also significantly smaller than that that calculated for Tumor samples (Normal: 0.054 +/- 0.018; Tumor: 0.154 +/- 0.029, P < 0.00078). The concordance between the populational and paired data sets allows us to exclude the hypothesis the ‘between distances’ correlation is driven by ‘individuality effects’, i.e. by the fact each single individual has a specific gene expression pattern accounting for the observed global/specific distance from tumor / distance from normal concordance.

C) Assessment of the global and signature-specific gene expression distances of Multi-stage (three or more stage) datasets

There were a total of 7 datasets describing tumor and normal samples collected from the same subject (1 dataset) or across a number of subjects (6 datasets). The development of the tumor usually involves its progression from the relatively benign to invasive and to metastatically aggressive phenotypes (Merlo et al. 2006). It is widely accepted that the gene expression signatures are able to discriminate between distinct stages of the tumor development. To explore the idea whether a summed expression change involving the entire set of mRNAs behaves similarly to the changes in signature-specific, “master” genes, we calculated DNGlobal and DNSpecific for 7 (six from NCBI GEO and one external) datasets representing normal and tumor samples that were comprised of three or more distinct physiological states of the underlying tissue.
Table 6: Mean, Standard Deviation and Variance calculated for Global and Specific distances from individual samples to the Normal Sample Space in the populational datasets

<table>
<thead>
<tr>
<th>DATASET</th>
<th>DNGlobal From individual normal samples to the Normal Sample Space (Mean +/- SD; variance)</th>
<th>DN Specific From individual tumor samples to the Normal Sample Space (Mean +/- SD; variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE6791 (cervical cancer)</td>
<td>0.05721 +/- 0.001671; 0.000279167</td>
<td>0.13059 +/- 0.02493; 0.0006216929</td>
</tr>
<tr>
<td>GSE10797 (invasive breast cancer)</td>
<td>0.08878211 +/- 0.018546943; 0.0003439891</td>
<td>0.1480193 +/- 0.04274649; 0.0018272624</td>
</tr>
<tr>
<td>GSE12345 (pleural mesothelioma)</td>
<td>0.06323201 +/- 0.01296917; 0.0001681993</td>
<td>0.0871369 +/- 0.02157966; 0.00004656815</td>
</tr>
<tr>
<td>GSE12452 (nasopharyngeal carcinoma)</td>
<td>0.05510947 +/- 0.01769130; 0.0003129822</td>
<td>0.077843 +/- 0.013091253; 0.0001713809</td>
</tr>
<tr>
<td>GSE14762 (RCC)</td>
<td>0.02229638 +/- 0.004879693; 0.000238114</td>
<td>0.09884668 +/- 0.009699626; 0.0000584726</td>
</tr>
<tr>
<td>GSE6791 (HNSCC)</td>
<td>0.05799383 +/- 0.01747614; 0.0003054155</td>
<td>0.0834743 +/- 0.01661218; 0.0002759646</td>
</tr>
<tr>
<td>GSE3678 (papillary thyroid carcinoma)</td>
<td>0.04147274 +/- 0.006705467; 0.0004496329</td>
<td>0.0560819 +/- 0.005507370; 0.000303311</td>
</tr>
<tr>
<td>GSE3524 (oral squamous cell carcinoma)</td>
<td>0.02964479 +/- 0.006389468; 0.000408253</td>
<td>0.0715533 +/- 0.01914830; 0.0003666572</td>
</tr>
</tbody>
</table>
Figure 2: Distance parameters successfully separate samples in the esophageal sample (GSE1420) dataset representing normal esophagus (blue), Barrett’s esophagus (orange) and esophagus carcinoma (red) samples.
Figure 3(a): Linear graphs depicting the relative distance of every given sample to the Normal sample space as defined by DNGlobal and DNSpecific metrics in the multi-stage datasets.
As GEO database contains only one dataset, GSE1420 (Figure 2), that is represented by paired tissue samples profiled using Affimetrix platform, we added to this study 6 datasets comprised of the samples collected across a number of individuals and profiled using the same microarray platform (Table 3). For each dataset, the global and specific expression distances were calculated as described above. In all datasets, the progression
of the disease was reflected in an increase of the distance of individual tumors from Normal Sample Space.

For each of these datasets linear graphs were generated. Each graph depicts the relative distance of every given sample to the Normal Sample Space as defined by DNGlobal and DNSpecific metrics (Figure 3). As could be seen at the Figure 1, both DNGlobal and DNSpecific place the most malignant tumors farther from the normal tissue control than the least malignant tumors or relatively benign tumors precursor states. The only case when metastatic tumors were less distant from the Normal Tissue Space than primary tumors, was the comparison of metastatic transitional cell carcinomas (TCC) of the bladder and superficial TCC with carcinoma in situ (TCC-CIS) (dataset GSE3167). This discrepancy might be explained by previous observations that the presence of concomitant CIS confers a worse prognosis in patients TCC (Shariat et al. 2007) et al., 2007). In all the cases when easy visual discrimination of the tumor and normal/benign samples could be achieved, the performances of DNGlobal and DNSpecific were comparable. These results suggest that the genome-wide metrics may help to assess the ‘degree of malignancy’ of the tumor cells.

**d) Principal component analysis (PCA) of the distance spaces**

In addition to the direct correlation between indexes, the degree of the mutual correlation between DNGlobal and DNSpecific distances could be quantified by the principal component analysis (PCA) on the four dimensional space spanned by these four indexes (DCglobal, DNGlobal, DCspecific, DNSpecific). PCA gives us an immediate quantitative
appreciation of the relative importance of the architectural modes of gene regulation. Typical results of the PCA analysis of the two-point and multipoint (one for each type) datasets are reported in Table 7. The patterns of the component loading are remarkably consistent across all the 24 (including multi-stage) datasets analyzed. The proportion of the variation observed is also similar across the datasets. The variance data for the two-point data can be observed for paired and population datasets in Tables 8 and 9, respectively.

In the four dimensional space, the PCA generated four components reflecting the variation in the data. The first component (PC1) is the largest one. In this component all the indexes enter with the same direction of correlation (loading sign). This component might reflect the presence of the attractor. The proportion of the variance it explains reflects the relative importance of attractor (cell type) driven dynamics in gene expression regulation. As all the distance indexes are positively correlated along this axis and as the distance from this attractor is equally measured by all the distance indexes adopted (DNglobal, DNspecific, DCglobal, DCspecific), this attractor corresponds to the center of distribution, and the PC1 (distance from the attractor) has the same sign as measured by any of the indexes. PC1 component explains by far major portion of information contained in the expression profiles and, given the homogeneity of signs, it reflects a topological ‘distance from a centre’ (here, a center of attractor) from which all the samples could have either lesser or higher distance independently of being cancer or normal samples.
Table 7: The table illustrates the relative importance of components and the actual loadings corresponding to the distances in the two-point datasets GDS1165 and GSE12345. The highlighted pattern of loadings is consistent across all the datasets. The results of the PCA analysis of all other datasets can be found in the Appendix.

<table>
<thead>
<tr>
<th>Two-point dataset: Papillary thyroid carcinoma dataset (GDS1665)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative importance</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
</tr>
<tr>
<td>Component Loadings:</td>
</tr>
<tr>
<td>DCGlobal</td>
</tr>
<tr>
<td>DNGlobal</td>
</tr>
<tr>
<td>DCSpecific</td>
</tr>
<tr>
<td>DNSpecific</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multi-stage dataset: Mesothelioma (GSE12345)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative importance:</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>Proportion of Variance</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
</tr>
<tr>
<td>Component Loadings:</td>
</tr>
<tr>
<td>DCGlobal</td>
</tr>
<tr>
<td>DNGlobal</td>
</tr>
<tr>
<td>DCSpecific</td>
</tr>
<tr>
<td>DNSpecific</td>
</tr>
</tbody>
</table>
The second component (PC2) puts in opposition (opposite loading signs) the distances from cancer (DC) and normal (DN) poles. The topological structure described by PC2 corresponds to the fact that normal and cancer poles do in effect occupy distinct positions in the gene expression space and thus, as for this structure, there must be a component of the distances indexes reflecting the relatively higher (lower) distance of a sample from the Normal or Tumor pole (Figure 4). The modulation driven by Tumor/Normal relative distance is definitively less important than the cell-kind attractor, as is inferred from the observation that the portion of the variance explained by PC2 is considerably lower than the portion explained by PC1. Along this component, DNspecific-DNglobal indices enter with the same loading sign, while being in opposition to the DCspecific-DCglobal pair.

Table 8: PCA profiles of two-point paired datasets representing the proportion of variance observed by each component

<table>
<thead>
<tr>
<th>Proportion of Variance / Dataset</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ductal Breast Carcinoma (GSE5764)</td>
<td>0.908</td>
<td>0.0901</td>
<td>0.00107</td>
</tr>
<tr>
<td>Lobular Breast Carcinoma (GSE5764)</td>
<td>0.882</td>
<td>0.116</td>
<td>0.00199</td>
</tr>
<tr>
<td>Pulmonary adenocarcinoma (GSE2514)</td>
<td>0.8635</td>
<td>0.1361</td>
<td>0.00022</td>
</tr>
<tr>
<td>Pulmonary adenocarcinoma (GSE7670)</td>
<td>0.917</td>
<td>0.0815</td>
<td>0.00108</td>
</tr>
<tr>
<td>Renal cell carcinoma (GSE6344)</td>
<td>0.777</td>
<td>0.2231</td>
<td>0.00023</td>
</tr>
<tr>
<td>Renal cell carcinoma (GSE781)</td>
<td>0.781</td>
<td>0.219</td>
<td>0.00055</td>
</tr>
<tr>
<td>Head and neck squamous cell carcinoma (GSE6631)</td>
<td>0.954</td>
<td>0.0436</td>
<td>0.00252</td>
</tr>
<tr>
<td>Papillary thyroid carcinoma (GSE3467)</td>
<td>0.8542</td>
<td>0.1444</td>
<td>0.00128</td>
</tr>
<tr>
<td>Esophagus Carcinoma (GSE1420)</td>
<td>0.875</td>
<td>0.124</td>
<td>0.124</td>
</tr>
</tbody>
</table>
The third component (PC3) reflects the ‘degree of autonomy’ of the signature genes from the global behavior of the cell-kind attractor. Relative strength of PC3 tells us whether signature genes possess intrinsic difference from the components of the general expression landscape or simply represent transcription units most sensitive to the common regulatory signal. Latter behavior is registered by PC2, while purely ‘democratic’ behavior of gene expression profile is registered by PC1. Intuitively, the loading pattern of PC3 component (the loadings correspond to the correlation coefficient of the original variables with the components) should have the specific (DNspecific,DCspecific) and global (DNglobal,DNspecific) indexes entering with opposite signs.

Table 9: PCA profiles of two-point population datasets representing the proportion of variance observed by each component

<table>
<thead>
<tr>
<th>Proportion of Variance / Dataset</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasive Breast (Epithelial) Carcinoma (GSE10797)</td>
<td>0.978</td>
<td>0.0192</td>
<td>0.00233</td>
</tr>
<tr>
<td>Invasive Breast (Stromal) Carcinoma (GSE10797)</td>
<td>0.986</td>
<td>0.012</td>
<td>0.0013</td>
</tr>
<tr>
<td>Cervical Carcinoma (GSE6791)</td>
<td>0.884</td>
<td>0.1153</td>
<td>0.00029</td>
</tr>
<tr>
<td>Head and Neck Carcinoma (GSE6791)</td>
<td>0.967</td>
<td>0.0319</td>
<td>0.00072</td>
</tr>
<tr>
<td>Mesothelioma (GSE12345)</td>
<td>0.892</td>
<td>0.1071</td>
<td>0.00119</td>
</tr>
<tr>
<td>Nasopharyngeal Carcinoma (GSE12452)</td>
<td>0.934</td>
<td>0.0655</td>
<td>0.00062</td>
</tr>
<tr>
<td>Oral Squamous Cell Carcinoma (GSE3524)</td>
<td>0.914</td>
<td>0.0857</td>
<td>0.00059</td>
</tr>
<tr>
<td>Renal Cell carcinoma(GSE14762)</td>
<td>0.891</td>
<td>0.1027</td>
<td>0.00669</td>
</tr>
<tr>
<td>Papillary thyroid carcinoma (GSE3678)</td>
<td>0.814</td>
<td>0.1847</td>
<td>0.00094</td>
</tr>
</tbody>
</table>
The proportion of the variation explained by fourth component (PC4) was negligible in all the cases compared to three previously discussed components. The PC4 might represent the ‘background’ noise generated by the stromovascular or other cells that may be present in the analyzed tissue samples. The PC4 would explain the smallest proportion of observed variation between sample sets. Its relatively small size reflects the strict quality controls used in the procedure of the selection of the published high-throughput datasets used in the current study.

Figure 4: Three dimensional representation of the principal components PC1, PC2 and PC3 in the two-point paired and population datasets. Normal samples are shown in blue and tumor samples are shown in red. This figure specifically highlights the classification power of PC2 (Normal/Cancer classifier) that does not require selection or validation of the minimized expression signature.
In analyzed datasets, the relative importance of cell-kind driven gene expression regulation (PC1) was ranged from 77% to 98%, while the distinction between normal and cancer poles (PC2) was ranged from 22% to 1%. The ‘degree of autonomy’ (signature genes working independently of global attractor dynamics) was represented by smallest component (PC3) being less than 1% in all datasets with an exception of esophageal dataset (GSE1420).

e) Cancer – An attractor with intermediate regulatory framework

Results of the principal component analysis could be used to discern the topological structure of cancer and cell-kind attractors. Our observations support the hypothesis of cancer being a stable attractor state in the dynamic system with intermediate regulation architecture could be described as a midpoint between “democratic” and “autocratic” regulatory landscapes. The intermediate paradigm is illustrated through an analysis of PC2 that is able to “readily sense” the difference between Normal and Cancer samples using both specific and global distance measures. Despite the fact that specific indices (gene signatures) enter as higher loadings on PC2 as compared to global distance indexes, latter indices also play a substantial role. In the case of purely ‘democratic’ architecture, PC3 would be expected to accounts for only a very small portion of variation; otherwise, at least some degree of autonomy of signature, or ‘master’, genes shall be acknowledged. Thus, after analysis of the principal components we conclude the canalization of the tumor development towards the stabilization of the cell population in the cancer attractor state follows the intermediate paradigm [not fully “democratic” or not fully “autocratic”].
It is worth noting, that the use of the distances (instead of the differences in the expression levels for individual genes) allows for an unbiased estimation of the regulatory paradigm in the living system, as each descriptive parameter of the system (global, specific, normal, tumor) is described by numerical value and evaluated as such, being not affected by the number of genes that passed some arbitrary significance threshold chosen for individual dataset. The cancer attractor model arising from the results obtained in the present study is depicted in the Figure 5.

Figure 5: Panel A describes the topology of the cell-kind and tumor attractors supported by present study. Panel B reports the classical view of cancer. The red and blue circles represent the cancer and normal attractor states as distinct poles. The rectangle represents the phase space of possible gene expression profiles, the stars are the observed samples, while the ellipse represents the general cell-kind attractor. From this model we can derive that the cells that by one or another reason leave “stable state” and depart from the normal attractor may with relatively high probability be attracted to the road toward cancer attractor without the prerequisite of getting departed from relatively strong cell-kind attractor.

As could be seen from the Figure 5, the topology of the cell-kind and tumor attractors supported by present study closely follows the Huang’s hypothesis stating that the cancer is a sub-attractor of the general cell kind attractor (S. Huang et al. 2009). The main
component defining the location of the sample in the space occupied by all samples is its distance from the general cell-kind attractor, thus the samples far removed from the normal subattractor are also distant from the cancer subattractor (PC1 component). In case of PC1, DN and DC indices are correlated and enter with the same sign into the component. The second component, PC2, discriminates if a given sample is closer to the cancer or normal sub-attractors (PC2 has opposite signs for DN and DC). Therefore, the similarity between cancer and normal samples is greater than the difference between them. In other words, prostatic cancer cell remains a prostate cell after all. Notable, this view is substantially different from the “classical” understanding of the tumorigenesis, when tumor and normal cells occupy the opposite poles of the allowed expression space (Fig.5, Panel B). If the “classical” model was correct, PC1 should have DN and DC indices entering with opposite signs reflecting negative correlation values.

A case study performed on the breast carcinoma dataset (GSE10971) may serve as a good illustration for an attractor model. The multi-stage dataset comprises luteal phase fallopian tube epithelium from BRCA1/2 mutation carriers and from normal controls as well as the samples of the high-grade adnexal serous carcinoma of the ovary. Traditional analysis of this data collected using Affymetrix microarrays highlighted specific gene signature that passed multiple test correction places. This gene signature places fallopian tube epithelium from BRCA1/2 mutation carriers close to the high-grade serous carcinoma samples (Tone et al. 2008). Our analysis of both Global and Specific distance characteristics indicated that the normal epithelial samples collected from the patients
predisposed to ovarian carcinoma have not yet embarked on the travel toward “cancer” attractor (Figure 3b). Other three-point datasets also provided clear discrimination between normal and malignant states, while providing relatively poor discrimination for the true normal and pre-malignant samples (Figure 3a). The only case when surefooted discrimination was possible at the earliest stages of the carcinogenesis was a set of samples representing the progression of the hepatocellular carcinoma (dataset GSE6764, Figure 3b). All together, our observations point that that the shift toward cancer attractor either takes place relatively late in the process of carcinogenesis or requires some time to become substantial. This observation also goes well with the hypothesis that cancer-specific changes of the expression landscape are subject to intermediate regulatory pattern, representing the middle ground between “democratic” and “autocratic” regulatory landscapes.

Conclusion and Future Perspective

Here we presented quantitatively evidence supporting the structure of the cancer attractor previously suggested by Huang and the hypothesis that cancer-specific changes of the expression landscape are subject to intermediate regulatory pattern, representing the middle ground between “democratic” and “autocratic” regulatory landscapes. The remarkable similarity of the observations made using multiple independent datasets, including these comprised of multiple types of samples demonstrates robustness of the genome-wide expression signatures as a mean to diagnose tumors. This study supports the view of the cell population as dynamic system. Moreover, the strong correlation
between the ‘distance from normal’ and ‘distance from cancer’ poles for all the analyzed samples proves existence of a cell-kind-attractor, with cancer and normal poles representing two sub-attractors.

There are a number of immediate applications of the analyses performed. First, after initial sets of normal and tumors samples for each particular cancer are analyzed to define Normal and Cancer Spaces, the classification of any new sample to be diagnosed could be achieved by calculation sample specific distance from this sample to Normal Space (DN) and Cancer Sample (DC). If DN > DC sample will be classified as cancer, If DC > DN, sample will be classified as normal. An increase in the number of the initially profiled samples will provide for better definition of the Normal and Cancer Spaces and better classification of the subsequent samples. Second, for every sample to be diagnosed, the distance from the sample to the Normal Space could be plotted linearly, and the degree of the malignancy of the given sample will be proportional to the linear distance. Importantly, relative degree of the malignancy could be assigned to the sample using whole-genome patterns of the gene expression, without the need for specific biomarkers or gene signatures. Third, the principal component analysis (PCA) on the four dimensional space spanned by four indexes (DCGlobal, DNGlobal, DCSpecific, DNSpecific) could be used for diagnostic discrimination of the samples. Each new sample to be diagnosed should be added to initial (reference) dataset of the cancer and normal tissues of the particular cell-type, PCA executed at whole dataset, then first three components (PC1, PC2, PC3) should be used for three dimensional graphing of the
results. New samples will be co-classified with the group of the samples with similar
degree of the malignancy.

Cell populations are collective dynamic systems living in a phase space where only very
specific low energy states (cell kind attractors) are compatible with survival. These
attractor states define cell differentiation. When cell departs from its cell-kind attractor,
there are only three possible scenarios. One, cell could die as a result of a profound
deregulation of its molecular networks incompatible with survival. Second, cell could be
attracted back to the normal pole of the cell-kind attractor. Third, cell could randomly fall
under the influence of the cancer pole of the cell-kind attractor, and acquire tumorigenic
properties. We are still far from the exploiting a statistical mechanics of life, but our data
suggest that, in principle, this can be done. The ‘cell kind’ barriers are energetically much
higher than the normal/cancer one, thus, offering a possibility of the ‘global reversion’ of
cancer phenotype. It might be possible to find the way to “kick” the cell out of
equilibrium, and, therefore, out of the influence of cancer pole of cell-kind attractor.
Being removed from low energy state, cell will be pushed to face three possible fates
again: death, normalization or attracting back to the cancer pole. Of course, the molecular
or other mean of the ‘global reversion’ therapy should be delivered specifically to the
cancer cells. ‘Global reversion’ therapy cannot be based on the exploitation of ‘master
key genes’, but should rely on more general means, for example, previously postulated
morphogenetic fields sharing some similarities in embryonic and cancer cells.
3. Abundance based transcriptome analysis as a tool for automated discovery of the tumor biomarkers

Rationale

The purpose of the current study is to explore the composition of the human transcriptome over a wide range of normal tissues and tumors using EST abundance analysis. We hypothesize that analysis of EST abundance might help to identify novel biomarkers of cancer initiation and progression.

Background

The enormous scale in which cancer affected mankind in the past century emphasizes an importance of both prevention and early stage detection of this devastating disease. Indeed, biomarkers project a promising future for the early stage detection of cancer (Oluwadara & Chiappelli 2009; Cazzaniga et al. 2009; A. Scott & Salgia 2008). Additionally, prognostic biomarkers provide vital information influencing therapeutic decisions (Ludwig & John N Weinstein 2005; J. J. Liang et al. 2009; Hwa et al. 2008). A number of bioinformatics and machine learning methods for the detection of biomarkers have been developed and utilized previously (Phan et al. 2009).
Unlike genomes comprised of relatively stable, species-specific DNA, tissue transcriptomes are very dynamic in nature. The functional and structural landscape of a particular cell phenotype depends heavily on the relative frequency of the transcription of individual genes. These frequencies, usually described as expression levels, are prone to change under the influence of the environmental and internal stimuli (Martínez & Reyes-Valdés 2008). Gene expression can be quantitatively measured at the transcriptional level by a number of low- to high-throughput methods.

The inventory of the human transcripts has increased dramatically in recent years, to include large number of non-coding mRNAs. Accumulating data on non-protein-coding transcripts suggest that besides the regulatory machinery driven by proteins, another yet enigmatic regulatory network of RNA molecules operates and has considerable impact on cell functions (Széll et al. 2008; Pontius et al. 2007; Waterston et al. 2002). Indeed, exonic sequences cover only 1.1% of the Human genome; the majority of not-yet-processed and spliced transcripts are represented by intronic and intergenic sequences (J. C. Venter et al. 2001). Therefore, it is not surprising that aside from mutations and polymorphisms in protein-coding genes, much of the variation between individuals, including that which may affect our predispositions to cancer, is probably due to differences in the non-coding regions of the genome (Mattick JS 2003).

In the course of an analysis of tissue-specific transcriptomes, many non-coding transcripts have been identified. Additionally, the class of the transcripts with relatively
low coding potential has been described. Later transcripts encode only short open reading frames (50–70 amino acids) and, in many cases, these ORFs lack Kozak sequences at translation start sites or are not evolutionarily conserved. Many of these non-coding transcripts RNAs with low coding potential were predicted using bioinformatic methods. For example, a study of Washietl et al, 2005 evaluated conserved genomic DNA sequences for signatures of structural conservation in base-pairing patterns with exceptional thermodynamic stability and predicted more than 30,000 structured RNA elements in the human genome (Washietl et al. 2005). Almost a 1,000 of these sequences were found to be conserved across all vertebrates (Washietl et al. 2005). Chromosome tiling experiments using DNA microarrays also demonstrated that most of the genome sequences are transcribed, and that many introns encode for the novel RNA species (Weile et al. 2007; Kapranov et al. 2005).

The role of non-coding RNAs, particularly miRNA, in the context of cancer has been recently established (Visone & Croce 2009; Conrad et al. 2006). In many cases non-coding RNA species has been shown to regulate the alternative splicing of essential proteins, a phenomenon that has great implication in inflammation, disease and cancer (Mallardo et al. 2008). Interestingly, the use of tiling microarrays revealed genome-wide hyper-transcription in mouse embryonic stem cells (ESCs), including expression of normally silent, non-coding regions. This hyper-transcription reflects the unusual "open" structure of ESC chromatin and contributes to the plasticity of the stem cells. Hyper-transcription points represent additional commonality between ESC and cancer (Efroni et
al. 2008; B. M. Turner 2008). Thus, it is logical to assume that non-coding RNAs could become valuable source for novel prognostic and diagnostic biomarkers for human malignancies.

DNA Microarrays and protein arrays, with their inherent ability to capture the diverse aspects of cancer are the most commonly used data source for the discovery of novel biomarkers. Most often, these arrays are fabricated to cover only the protein encoding genes (Ambros 2001). In a number of previous studies, meta analysis was attempted using a compendium of microarrays allowing one to mine for novel markers of cancer (Xu et al. 2007; Wren 2009). However, these high-level analyses in most part were restricted to searches for the markers performing in the context of a particular subtype of cancer. Multi-cancer analysis attempts are less common. One example of this kind of studies is analysis of expression profiles covering 60 human cancer cell lines of NCI-60 panel spanning 9 different human tissues (Shankavaram et al. 2007).

Complete transcriptome studies of cancer cells that include non-coding transcripts are warranted. One barrier to such studies is relative difficulty of obtaining expression profiles that are comprehensive enough to cover low-abundancy intra and intergenic transcripts. However, these difficulties may be solved using publicly available data describing EST sequences. Expressed Sequence Tags (ESTs) were introduced during the initiation phase of high-throughput cDNA sequencing and quickly became useful in the identification of novel genes and mapping of the genomic sequences (M. D. Adams et al.
1991). On average, these sequences are relatively short, with a length ranging from 400 to 600 bases, and are relatively inaccurate (NCBI 2002). Many genome projects utilized the EST based approach to gene mapping, resulting in an ever-increasing number of ESTs in the databases maintained by NCBI. ESTs provide a quick and inexpensive roadmap not only to identify novel genes, but also to obtain data on gene expression and gene regulation in various tissues. The highly redundant nature of ESTs makes them an excellent material for the RNA expression quantification *in silico*. The study of cancer transcriptomes using both coding and non-coding ESTs would help to comprehend the cancer cell expression patterns down to a better degree than the classic gene-coding RNA studies.

The expression patterns of both tumor tissues and their normal counterparts may be profiled by analyzing the largest publicly available expression database, the UniGene. This system is developed, maintained and updated by the National Center for Biotechnological Information at National Institute of Health (NCBI, NIH). UniGene is comprised of non-redundant set of gene-centered clusters of the expressed sequences, including mRNAs, ESTs and high throughput cDNA (HTC) sequences (NCBI 2002; D. L. Wheeler et al. 2007). To avoid false alignment within the gene clusters, the repetitive stretches of the nucleotides are masked (Schneeberger et al. 2005). High quality reads on the templates of expressed sequences with a sequence length of at least 100 base pairs are used for subsequent clustering. UniGene forms the basis for other core NCBI resources utilized in the routine searches for protein similarities and putative evolutionary
relationships and in comparisons of EST-based expression profiles of various tissues (NCBI 2002).

ESTs form the building blocks of UniGene and represent an integral part of nearly every gene cluster in the UniGene. The EST data in the UniGene database can be utilized to study the expression patterns of genes in normal tissues and tumor samples. This methodology enables a study of the abnormal expression of genes in cancer through the direct comparison of the diverse express patterns of genes within a wide range of human tissues and tumor samples and will possibly lead to the discovery of new human tumor marker candidates.

**Hypothesis**

Abundance based meta-analysis of tissue-specific transcriptomes may reveal novel diagnostic and prognostic biomarkers for human tumors. Here we attempted to profile the variation in the abundance of ESTs derived from normal tissues and tumors and available through UniGene database system and to extract candidate biomarker genes for various human malignancies.

**Materials and Methods**

To compare the patterns of the gene expression in the normal and tumor EST libraries, the UniGene build 210 of *Homo sapiens* was downloaded from the NCBI ftp portal of the Unigene database. Descriptions of the UniGene cluster data and the cDNA libraries
associated with each of these clusters were deconvoluted. This particular version of the UniGene database contained 123,687 UniGene clusters associated with 8668 cDNA libraries used to generate all contributing ESTs. Perl scripts and MySQL database management system were used to automate the data extraction and to provide for the data storage, respectively. R data analysis package with Bioconductor was utilized to perform the statistical analysis on the downloaded dataset (Gregory Alvord et al. 2007; Mark Reimers & Carey 2006).

The abundance of ESTs was used as the fundamental quantitative unit describing the expression levels similarly to the hybridization intensity of the DNA microarray. The abundance of the ESTs belonging to a particular gene cluster in a given tissue of the Unigene was defined as the ratio of a number of ESTs that belong to the particular gene cluster and are expressed in a given tissue to the total number of ESTs captured from that particular tissue. The terms “UniGene cluster”, “cluster” and “unigene” are used interchangeably in the following text. Diversity coefficients of the tissue transcriptomes as represented by abundance measures for various UniGene clusters were calculated using Shannon diversity index (Shannon 1948). This index was formally defined in a recent study addressing diversity in the context of transcriptome (Martínez & Reyes-Valdés 2008). The richness (quantity of unigenes) and the evenness of each tissue were calculated using Pielou index. A t-test was used to compare the diversities in the normal and the corresponding tumor tissues (Magurran 1988). The significance of the variation between the paired normal and cancer human tissues was tested using non-parametric
Mann-Whitney procedure (H. Mann & Whitney 1947). Metastats, an expression variation analysis tool that handles sparsely sampled features was used to compute the false discovery rate (White et al. 2009). Gene cluster related information for the significant unigenes was extracted from the data provided in the current version of the UniGene database. Functional analysis was performed using a population of genes with expressions significantly skewed toward tumors. For this analysis we used the KEGG pathway painter (KPP) specially developed for this project using the KEGG API (Kawashima et al. 2003). The detailed description of the KPP tool is given in the Appendix.

Results and Discussion

a) Development of the standard vocabulary describing human tissues

The text descriptors annotating all the tissues and cancer types that served as a source for the production of the cDNA libraries comprising UniGene were extracted. For most of these libraries, data descriptions containing information about their tissue of origin and type of the tumor were available. The complete list of tissue descriptors includes a total of 64 tissue categories.
and 51 different types of cancers originated in these tissues.

These library descriptors were compressed into broader tissue descriptors. For example, terms “hepatic cancer”, “hepatic carcinoma”, “hepatic tumor”, “liver cancer”, “liver tumor”, “HCC” etc were co-classified as “Hepatic tumors”. Re-classification of the descriptors resulted in a final vocabulary comprised of 37 different tissues and 28 types of tumors derived from these tissues (Table 10). The descriptions of cDNA libraries stored in the MySQL database were automatically updated using these broadly defined classes that were utilized in further analysis.

b) Classification of the cDNA libraries used in the study

For the purpose of this study, all cDNA libraries were classified in four different broad categories: i) Normal, ii) Cancer, iii) Non-malignant diseases (e.g. Parkinson disease, ischemia) and iv) Libraries lacking proper tissue descriptions.

For each tissue, we have calculated number of normal and cancer libraries, numbers of gene clusters (unigenes) and total number of ESTs derived from normal and cancer libraries (Table 11), as well as libraries made of the diseased tissues and libraries lacking tissue descriptions. The normal and the cancer groups represented the largest part (87%) of the expression data represented by the ESTs. There were also a substantial number of ESTs lacking proper descriptions (12.5%), while the amount of ESTs from the diseased tissues and/or benign tumors was negligible compared to the other three groups. The
diseased tissue samples were mostly restricted to the bone and muscle diseases (e.g. Duchenne dystrophy), illustrating an unevenness of efforts in characterization of transcriptome changes in non-malignant disorders (Table 12). ESTs from the non-malignant diseased tissues and tissues lacking proper descriptors were excluded from further analysis.

According to our observations, cumulative efforts aimed at the cancer transcriptomes, though, seems to be more advanced, as UniGene database contains more cancer cDNA libraries (N = 4642) than the libraries built using normal tissues (N = 2682). An overall excess of cancer cDNA libraries spans an entire dataset, with every tissue covered by twice as much of tumor-derived libraries than normal libraries. On the other hand, the total number of ESTs derived from each normal tissue, on average, is substantially higher compared to that derived from respective cancer. This paradox can be explained by the fundamental objective of the Human Genome project to catalogue all the genes in the normal human genome.

As could be seen from Table 11, the distribution of ESTs across tumor and normal tissues is far from being well-balanced. For example, categories “heart”, “vasculature”, “peritoneum”, “ear” and “pineal gland” lack tumor ESTs, while “olfactory mucosa”, “chest wall” and “penis” lack ESTs derived form normal tissues. This discrepancy was, to certain extent, normalized by taking into account per tissue abundance for each unigene.
c) Unique and Common Unigenes

After extraction the data describing ESTs and gene clusters, it became feasible to separately extract tissue-specific unigenes represented by both normal and tumor ESTs, tissue-specific unigenes built solely upon ESTs derived from tumor libraries and tissue-specific unigenes that contain only ESTs derived from normal libraries (Table 12).

For example, according to the Table 12, in the brain the number of “normal unigenes” represents the number of EST clusters where at least one EST was derived from normal brain samples; the number of “cancer unigenes” reflect the number of EST clusters, where at least one EST was derived from brain tumor samples; the number of “unique or exclusive normal unigenes” is a number of EST clusters, where all brain-specific ESTs were derived from normal brain samples; the number of “exclusive cancer unigenes” represent the number of EST clusters, where all brain-specific ESTs were derived from brain tumor samples; number of “common unigenes” is a number of EST clusters, where at least one EST was derived from normal brain samples and at least one EST was derived from brain tumor sample. Tissues with sparse representation by ESTs are removed from the table. EST clusters summarized in the Table 12 represents both the protein coding as well as non-coding genes.
Table 11: Statistics describing the libraries, unigenes (gene clusters) and ESTs derived from normal tissues and tumors samples. Normal unigenes are defined as gene clusters that contain at least one EST from the particular normal tissue; Cancer unigenes are defined as gene clusters that contain at least one EST from the cDNA library representing a particular tumor type.

As could be seen from the Table 12, there is a substantial number of unigenes that include both normal and tumor ESTs derived from the same tissue (“common unigenes”).

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Normal libraries</th>
<th>Normal unigenes</th>
<th>Normal ESTs</th>
<th>Cancer libraries</th>
<th>Cancer unigenes</th>
<th>Cancer ESTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>497</td>
<td>34604</td>
<td>787007</td>
<td>87</td>
<td>18642</td>
<td>174529</td>
</tr>
<tr>
<td>Tissues lacking descriptors</td>
<td>59</td>
<td>39259</td>
<td>352400</td>
<td>174</td>
<td>24057</td>
<td>223292</td>
</tr>
<tr>
<td>Eye</td>
<td>56</td>
<td>24976</td>
<td>160932</td>
<td>3</td>
<td>8086</td>
<td>39123</td>
</tr>
<tr>
<td>Heart</td>
<td>32</td>
<td>14619</td>
<td>717193</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Liver</td>
<td>39</td>
<td>15068</td>
<td>99207</td>
<td>54</td>
<td>13687</td>
<td>107936</td>
</tr>
<tr>
<td>Kidney</td>
<td>31</td>
<td>19817</td>
<td>107596</td>
<td>87</td>
<td>17058</td>
<td>89069</td>
</tr>
<tr>
<td>Bone</td>
<td>9</td>
<td>6157</td>
<td>16172</td>
<td>22</td>
<td>13702</td>
<td>50358</td>
</tr>
<tr>
<td>Adrenal gland</td>
<td>12</td>
<td>6029</td>
<td>19762</td>
<td>14</td>
<td>6274</td>
<td>13221</td>
</tr>
<tr>
<td>Muscle</td>
<td>19</td>
<td>15397</td>
<td>75841</td>
<td>6</td>
<td>5592</td>
<td>23898</td>
</tr>
<tr>
<td>Testis tissue</td>
<td>170</td>
<td>25782</td>
<td>147432</td>
<td>73</td>
<td>13859</td>
<td>104177</td>
</tr>
<tr>
<td>Pregnancy tissue</td>
<td>472</td>
<td>31392</td>
<td>512435</td>
<td>9</td>
<td>7091</td>
<td>30351</td>
</tr>
<tr>
<td>Pancreas</td>
<td>19</td>
<td>14908</td>
<td>105643</td>
<td>18</td>
<td>14142</td>
<td>102394</td>
</tr>
<tr>
<td>Lung tissue</td>
<td>159</td>
<td>22985</td>
<td>158569</td>
<td>630</td>
<td>23543</td>
<td>195964</td>
</tr>
<tr>
<td>Lymphatic tissue</td>
<td>93</td>
<td>23380</td>
<td>220930</td>
<td>342</td>
<td>16707</td>
<td>148729</td>
</tr>
<tr>
<td>Vasculature</td>
<td>18</td>
<td>9876</td>
<td>46530</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lower gastrointestinal tissue</td>
<td>240</td>
<td>13571</td>
<td>83030</td>
<td>1040</td>
<td>24175</td>
<td>229598</td>
</tr>
<tr>
<td>Skin</td>
<td>27</td>
<td>13031</td>
<td>74232</td>
<td>33</td>
<td>13274</td>
<td>119604</td>
</tr>
<tr>
<td>Adipose tissue</td>
<td>13</td>
<td>5468</td>
<td>12417</td>
<td>1</td>
<td>558</td>
<td>720</td>
</tr>
<tr>
<td>Mammary gland</td>
<td>330</td>
<td>11156</td>
<td>47829</td>
<td>767</td>
<td>16565</td>
<td>106111</td>
</tr>
<tr>
<td>Upper gastrointestinal tissue</td>
<td>13</td>
<td>4075</td>
<td>12363</td>
<td>262</td>
<td>12470</td>
<td>72982</td>
</tr>
<tr>
<td>Peritoneum</td>
<td>5</td>
<td>180</td>
<td>291</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Female reproductive tissue</td>
<td>27</td>
<td>12859</td>
<td>53761</td>
<td>318</td>
<td>24819</td>
<td>250839</td>
</tr>
<tr>
<td>Parathyroid</td>
<td>1</td>
<td>24</td>
<td>24</td>
<td>2</td>
<td>7045</td>
<td>20602</td>
</tr>
<tr>
<td>Prostate</td>
<td>144</td>
<td>16667</td>
<td>81107</td>
<td>165</td>
<td>14591</td>
<td>107355</td>
</tr>
<tr>
<td>Salivary gland</td>
<td>4</td>
<td>912</td>
<td>2514</td>
<td>3</td>
<td>2891</td>
<td>10355</td>
</tr>
<tr>
<td>Connective tissue</td>
<td>8</td>
<td>4770</td>
<td>16389</td>
<td>17</td>
<td>16599</td>
<td>82051</td>
</tr>
<tr>
<td>Thyroid</td>
<td>79</td>
<td>4700</td>
<td>12031</td>
<td>280</td>
<td>9706</td>
<td>33611</td>
</tr>
<tr>
<td>Pineal gland</td>
<td>4</td>
<td>3437</td>
<td>6353</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ear</td>
<td>6</td>
<td>5523</td>
<td>16378</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pituitary gland</td>
<td>7</td>
<td>5403</td>
<td>13430</td>
<td>2</td>
<td>923</td>
<td>1392</td>
</tr>
<tr>
<td>Nerve tissue</td>
<td>7</td>
<td>9582</td>
<td>25579</td>
<td>7</td>
<td>609</td>
<td>701</td>
</tr>
<tr>
<td>Bladder tissue</td>
<td>12</td>
<td>3304</td>
<td>8550</td>
<td>53</td>
<td>6611</td>
<td>18295</td>
</tr>
<tr>
<td>Olfactory mucosa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Chest wall</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Penis</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>
These common unigenes corresponds to the parts of the cellular machinery in one or another way essential for running both normal and tumor cells derived from the same tissue. Existence of a large number of such unigenes signifies the inevitable need to fulfill basic metabolic and other cellular function. However, it seems that the sizes of the gene sets supporting basic functions differ from tissue to tissue.

As inferred from the Table 12, the total number of unigenes in normal and cancer groups was classified as either exclusive or common unigenes. The portion of exclusive and common unigenes in the normal and cancer categories is illustrated as percentage of total unigenes in each category (Table 12).

The category “pregnancy-related tissues” corresponds to the largest number of exclusive normal unigenes. Our definition of “pregnancy-related tissues” includes umbilical cord, amniotic fluid and embryonic tissues. The high number of gene clusters that contain at least one EST from tissues of this category can be attributed to the fact that embryonic parts represent all tissues of the human body.

Next in the tissue list ranked by number of exclusively normal unigenes are brain and eye followed by testis. These patterns correspond to the diversity patterns previously reported for these tissues (Martínez & Reyes-Valdés 2008; Piatigorsky 1989). In case of exclusive cancer unigenes these numbers might be influenced both by an intrinsic versatility of expression and the variability among the assortment of cancer subtypes derived of the
particular tissue. Largest number of unigenes that include at least one tumor EST derived from a given were found in the categories “lower gastrointestinal tissues”, followed by “connective tissue”, “uterus’ and “ovary”.

Table 12: The table summarizes tissue-specific distribution of EST clusters composed of both normal and tumor ESTs, tumor ESTs only and normal ESTs only.

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Normal unigenes</th>
<th>Cancer unigenes</th>
<th>Exclusive Normal unigenes (ENU)</th>
<th>Exclusive Cancer unigenes (ECU)</th>
<th>Common unigenes (CU)</th>
<th>% of ENU</th>
<th>% of CNU</th>
<th>% of ECU</th>
<th>% of CCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>34604</td>
<td>18642</td>
<td>19392</td>
<td>3430</td>
<td>15212</td>
<td>56.04</td>
<td>43.96</td>
<td>18.40</td>
<td>81.60</td>
</tr>
<tr>
<td>Pregnancy – related tissues</td>
<td>31392</td>
<td>7091</td>
<td>24706</td>
<td>405</td>
<td>6686</td>
<td>78.70</td>
<td>21.30</td>
<td>5.71</td>
<td>94.29</td>
</tr>
<tr>
<td>Testis tissue</td>
<td>25782</td>
<td>13859</td>
<td>15725</td>
<td>3802</td>
<td>10057</td>
<td>60.99</td>
<td>39.01</td>
<td>27.43</td>
<td>72.57</td>
</tr>
<tr>
<td>Eye</td>
<td>24976</td>
<td>8086</td>
<td>17787</td>
<td>897</td>
<td>7189</td>
<td>71.22</td>
<td>28.78</td>
<td>11.09</td>
<td>88.91</td>
</tr>
<tr>
<td>Lymphatic tissue</td>
<td>23380</td>
<td>16707</td>
<td>10960</td>
<td>4287</td>
<td>12420</td>
<td>46.88</td>
<td>53.12</td>
<td>25.66</td>
<td>74.34</td>
</tr>
<tr>
<td>Lung tissue</td>
<td>22985</td>
<td>23543</td>
<td>8662</td>
<td>9220</td>
<td>14323</td>
<td>37.69</td>
<td>62.31</td>
<td>94.29</td>
<td>55.71</td>
</tr>
<tr>
<td>Kidney</td>
<td>19817</td>
<td>17058</td>
<td>8300</td>
<td>5541</td>
<td>11517</td>
<td>41.88</td>
<td>58.12</td>
<td>32.48</td>
<td>67.52</td>
</tr>
<tr>
<td>Prostate</td>
<td>16667</td>
<td>14591</td>
<td>7091</td>
<td>5015</td>
<td>9576</td>
<td>42.55</td>
<td>57.45</td>
<td>34.37</td>
<td>65.63</td>
</tr>
<tr>
<td>Muscle</td>
<td>15397</td>
<td>5592</td>
<td>11343</td>
<td>1538</td>
<td>4054</td>
<td>73.67</td>
<td>26.33</td>
<td>27.50</td>
<td>72.50</td>
</tr>
<tr>
<td>Liver</td>
<td>15068</td>
<td>13687</td>
<td>6263</td>
<td>4882</td>
<td>8805</td>
<td>41.56</td>
<td>58.44</td>
<td>35.67</td>
<td>64.33</td>
</tr>
<tr>
<td>Pancreas</td>
<td>14908</td>
<td>14142</td>
<td>6068</td>
<td>5302</td>
<td>8840</td>
<td>40.70</td>
<td>59.30</td>
<td>37.49</td>
<td>62.51</td>
</tr>
<tr>
<td>Lower gastrointestinal tract</td>
<td>13571</td>
<td>24175</td>
<td>2539</td>
<td>13143</td>
<td>11032</td>
<td>18.71</td>
<td>81.29</td>
<td>54.37</td>
<td>45.63</td>
</tr>
<tr>
<td>Skin</td>
<td>13031</td>
<td>13274</td>
<td>4345</td>
<td>4588</td>
<td>8686</td>
<td>33.34</td>
<td>66.66</td>
<td>34.56</td>
<td>65.44</td>
</tr>
<tr>
<td>Mammary gland</td>
<td>11156</td>
<td>16565</td>
<td>3198</td>
<td>8607</td>
<td>7958</td>
<td>28.87</td>
<td>71.33</td>
<td>51.96</td>
<td>48.04</td>
</tr>
<tr>
<td>Uterus</td>
<td>11084</td>
<td>20800</td>
<td>2097</td>
<td>11813</td>
<td>8987</td>
<td>18.92</td>
<td>81.08</td>
<td>56.79</td>
<td>43.21</td>
</tr>
<tr>
<td>Nerve tissue</td>
<td>9582</td>
<td>609</td>
<td>9298</td>
<td>325</td>
<td>284</td>
<td>97.04</td>
<td>2.96</td>
<td>53.37</td>
<td>46.63</td>
</tr>
<tr>
<td>Bone</td>
<td>6157</td>
<td>13702</td>
<td>1997</td>
<td>9542</td>
<td>4160</td>
<td>32.43</td>
<td>67.57</td>
<td>69.64</td>
<td>30.36</td>
</tr>
<tr>
<td>Adrenal gland</td>
<td>6029</td>
<td>6274</td>
<td>3287</td>
<td>3532</td>
<td>2742</td>
<td>54.52</td>
<td>45.48</td>
<td>56.30</td>
<td>43.70</td>
</tr>
<tr>
<td>Ovary</td>
<td>5503</td>
<td>14698</td>
<td>1377</td>
<td>10572</td>
<td>4126</td>
<td>25.02</td>
<td>74.98</td>
<td>71.93</td>
<td>28.07</td>
</tr>
<tr>
<td>Adipose tissue</td>
<td>5468</td>
<td>558</td>
<td>5241</td>
<td>331</td>
<td>227</td>
<td>95.85</td>
<td>4.15</td>
<td>59.32</td>
<td>40.68</td>
</tr>
<tr>
<td>Pituitary gland</td>
<td>5403</td>
<td>923</td>
<td>4905</td>
<td>425</td>
<td>498</td>
<td>90.78</td>
<td>9.22</td>
<td>46.05</td>
<td>53.95</td>
</tr>
<tr>
<td>Connective tissue</td>
<td>4770</td>
<td>16599</td>
<td>1030</td>
<td>12859</td>
<td>3740</td>
<td>21.59</td>
<td>78.41</td>
<td>77.47</td>
<td>22.53</td>
</tr>
<tr>
<td>Thyroid</td>
<td>4700</td>
<td>9706</td>
<td>1764</td>
<td>6770</td>
<td>2936</td>
<td>37.53</td>
<td>62.47</td>
<td>69.75</td>
<td>30.25</td>
</tr>
<tr>
<td>Upper gastrointestinal tract</td>
<td>4075</td>
<td>12470</td>
<td>986</td>
<td>9381</td>
<td>3089</td>
<td>24.20</td>
<td>75.80</td>
<td>75.23</td>
<td>24.77</td>
</tr>
<tr>
<td>Bladder tissue</td>
<td>3304</td>
<td>6611</td>
<td>1646</td>
<td>4953</td>
<td>1658</td>
<td>49.82</td>
<td>50.18</td>
<td>74.92</td>
<td>25.08</td>
</tr>
<tr>
<td>Salivary gland</td>
<td>912</td>
<td>2891</td>
<td>644</td>
<td>2623</td>
<td>268</td>
<td>70.61</td>
<td>29.39</td>
<td>90.73</td>
<td>9.27</td>
</tr>
<tr>
<td>Parathyroid</td>
<td>24</td>
<td>7045</td>
<td>8</td>
<td>7029</td>
<td>16</td>
<td>33.33</td>
<td>66.67</td>
<td>99.77</td>
<td>0.23</td>
</tr>
</tbody>
</table>
The percentage of unigenes contributing to the exclusive and common pools by the normal and cancerous gene clusters is also summarized in the table. In the cases of tissues with substantial number of unigenes (Eg: Brain, testis, eye, lung, kidney), the cancer unigenes tend to show a lot of commonality with the normal pool of unigenes. Interestingly, the exclusive gene clusters of the cancerous tissue were underrepresented when compared to the pool of exclusive normal unigenes. It might push the notion that the spectrum of transcriptional activity in normal tissue is higher than that of the cancer tissue. If this is the case, then the increased transcription observed in cancer might be due to amplified transcription of already active genes rather than increased transcriptional diversity.

**d) Estimation of the diversity of transcripts within normal and tumor tissues**

An introduction of the tissue-specific measures of the diversity may shed light on the versatility of the transcription of human genome in particular tissue contexts. An estimation of the diversity has been performed through quantification of the tissue labels attached to ESTs contributing to unigenes that are expressed in each tissue (or tumor) of interest.

Diversity is the composition of two fundamental components: a) Variety and b) Relative abundance. The tissue-specific “richness” coefficients (S) depict the diversity pattern in a way that reflects not only the number of unigenes *per se*, but also the expression levels (abundance) of these unigenes. Thus, the “richness” of each unigene contributes to the tissue-specific “richness”. Although as a heterogeneity measure, Shannon’s index takes
into account the evenness of the abundance of species; it is possible to calculate a separate additional measure of evenness (Magurran 1988). Evenness is the normalized Shannon and falls in the range of 0 and 1. An evenness of 1 indicates that the all unigenes contributing to tissue transcriptome are equally abundant, while an evenness of 0 indicates that there is absolutely no commonality in the abundance of unigenes. Both the normal tissue and tumors demonstrate substantial evenness of the distribution of abundances of the ESTs contributing to the gene clusters. High values of the evenness factor indicate that the process of the normalization of EST abundance successfully reduced an initial bias of the UniGene database, where tissues were unevenly covered by EST libraries.

Comparison of the diversity measures calculated for normal and cancerous tissues as represented by the abundance of contributing ESTs is presented in Tables 13 and 14. Among human tumors, the most diverse pattern of expression was observed in the category: “lower gastrointestinal cancer” comprising gastric, colonic and intestinal malignancies. The lung cancer tissue group consisting of tumors associated with lung, bronchus, pleura, trachea, larynx and pharynx displayed second most highly diverse pattern. Observed diversity patterns observed also provided interesting data concerning expression patterns in normal human tissues. According to our observations, eye tops the list of the tissues with highly diverse expression patterns, thus, confirming a conclusion inferred after the quantification of exclusively normal unigenes (Table 12). Second place in the list is occupied by testis.
Table 13: Diversity statistics describing of diversity, richness, evenness and variance of diversity reflecting the abundance of ESTs in each of the normal human tissues.

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Number of Sequences</th>
<th>Diversity H'</th>
<th>Richness S</th>
<th>Evenness E</th>
<th>Variance of H'</th>
<th>Standard Error of H'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye</td>
<td>160932</td>
<td>9.31899</td>
<td>24976</td>
<td>0.920333</td>
<td>1.18E-05</td>
<td>0.003435</td>
</tr>
<tr>
<td>Testis tissue</td>
<td>147432</td>
<td>9.250885</td>
<td>25782</td>
<td>0.91075</td>
<td>1.63E-05</td>
<td>0.004042</td>
</tr>
<tr>
<td>Pregnancy tissue</td>
<td>512435</td>
<td>9.013833</td>
<td>31392</td>
<td>0.870539</td>
<td>5.31E-06</td>
<td>0.002305</td>
</tr>
<tr>
<td>Kidney</td>
<td>107596</td>
<td>8.950741</td>
<td>19817</td>
<td>0.904636</td>
<td>2.55E-05</td>
<td>0.005045</td>
</tr>
<tr>
<td>Lung tissue</td>
<td>158569</td>
<td>8.946467</td>
<td>22985</td>
<td>0.890852</td>
<td>1.80E-05</td>
<td>0.004248</td>
</tr>
<tr>
<td>Prostate</td>
<td>81107</td>
<td>8.883301</td>
<td>16667</td>
<td>0.913808</td>
<td>2.99E-05</td>
<td>0.005467</td>
</tr>
<tr>
<td>Female reproductive tissue</td>
<td>53761</td>
<td>8.859325</td>
<td>12859</td>
<td>0.936326</td>
<td>3.19E-05</td>
<td>0.005646</td>
</tr>
<tr>
<td>Lymphatic tissue</td>
<td>220930</td>
<td>8.833965</td>
<td>23380</td>
<td>0.878159</td>
<td>1.33E-05</td>
<td>0.003644</td>
</tr>
<tr>
<td>Brain</td>
<td>787007</td>
<td>8.804831</td>
<td>34604</td>
<td>0.842429</td>
<td>4.39E-06</td>
<td>0.002095</td>
</tr>
<tr>
<td>Skin</td>
<td>74232</td>
<td>8.713065</td>
<td>13031</td>
<td>0.919576</td>
<td>2.42E-05</td>
<td>0.004923</td>
</tr>
<tr>
<td>Heart</td>
<td>77193</td>
<td>8.70495</td>
<td>14619</td>
<td>0.907764</td>
<td>3.26E-05</td>
<td>0.005707</td>
</tr>
<tr>
<td>Nerve tissue</td>
<td>25579</td>
<td>8.553971</td>
<td>9582</td>
<td>0.933061</td>
<td>7.60E-05</td>
<td>0.00872</td>
</tr>
<tr>
<td>Muscle</td>
<td>75841</td>
<td>8.494439</td>
<td>15397</td>
<td>0.876323</td>
<td>4.65E-05</td>
<td>0.006819</td>
</tr>
<tr>
<td>Mammary gland</td>
<td>47829</td>
<td>8.400288</td>
<td>11156</td>
<td>0.901344</td>
<td>6.12E-05</td>
<td>0.007825</td>
</tr>
<tr>
<td>Pancreas</td>
<td>105643</td>
<td>8.316743</td>
<td>14908</td>
<td>0.865457</td>
<td>4.31E-05</td>
<td>0.006565</td>
</tr>
<tr>
<td>LowerGastrointestinal tissue</td>
<td>83030</td>
<td>8.208109</td>
<td>13571</td>
<td>0.862587</td>
<td>4.55E-05</td>
<td>0.006743</td>
</tr>
<tr>
<td>Vascular</td>
<td>46530</td>
<td>8.08016</td>
<td>9876</td>
<td>0.879482</td>
<td>7.21E-05</td>
<td>0.008489</td>
</tr>
<tr>
<td>Adipose tissue</td>
<td>12417</td>
<td>8.06399</td>
<td>5468</td>
<td>0.936947</td>
<td>0.000141</td>
<td>0.011867</td>
</tr>
<tr>
<td>Bone</td>
<td>16172</td>
<td>7.874164</td>
<td>6157</td>
<td>0.902447</td>
<td>0.000206</td>
<td>0.014359</td>
</tr>
<tr>
<td>Ear</td>
<td>16378</td>
<td>7.80055</td>
<td>5523</td>
<td>0.905285</td>
<td>0.00017</td>
<td>0.013049</td>
</tr>
<tr>
<td>Pineal gland</td>
<td>6353</td>
<td>7.785438</td>
<td>3437</td>
<td>0.956165</td>
<td>0.000204</td>
<td>0.014297</td>
</tr>
<tr>
<td>Adrenal gland</td>
<td>19762</td>
<td>7.602619</td>
<td>6029</td>
<td>0.873429</td>
<td>0.000195</td>
<td>0.01397</td>
</tr>
<tr>
<td>Liver</td>
<td>99207</td>
<td>7.601805</td>
<td>15068</td>
<td>0.790161</td>
<td>7.87E-05</td>
<td>0.008873</td>
</tr>
<tr>
<td>Pituitary gland</td>
<td>13430</td>
<td>7.510406</td>
<td>5403</td>
<td>0.873841</td>
<td>0.000351</td>
<td>0.018742</td>
</tr>
<tr>
<td>Thyroid</td>
<td>12031</td>
<td>7.501853</td>
<td>4700</td>
<td>0.887225</td>
<td>0.000238</td>
<td>0.018382</td>
</tr>
<tr>
<td>Connective tissue</td>
<td>16389</td>
<td>7.355674</td>
<td>4770</td>
<td>0.868428</td>
<td>0.000208</td>
<td>0.014424</td>
</tr>
<tr>
<td>Bladder tissue</td>
<td>8550</td>
<td>7.272814</td>
<td>3304</td>
<td>0.897558</td>
<td>0.000324</td>
<td>0.018008</td>
</tr>
<tr>
<td>UpperGastrointestinal tissue</td>
<td>12363</td>
<td>7.166962</td>
<td>4075</td>
<td>0.862178</td>
<td>0.000325</td>
<td>0.018018</td>
</tr>
<tr>
<td>Salivary gland</td>
<td>2514</td>
<td>5.292427</td>
<td>912</td>
<td>0.776512</td>
<td>0.001936</td>
<td>0.044001</td>
</tr>
<tr>
<td>Peritoneum</td>
<td>291</td>
<td>4.893458</td>
<td>180</td>
<td>0.942326</td>
<td>0.003833</td>
<td>0.061913</td>
</tr>
</tbody>
</table>

The diversity patterns of the normal tissue and corresponding tumor tissue were compared by calculating the tissue specific T-statistics (Table 15). According to this measure, the diversity patterns in the tumor and normal lung tissue appears to be similar. With the exception on lung tissue, all other human tissue demonstrated significant changes in the diversity after malignization. The diversity pattern in the lung tumors might be overestimated due to the tumor set comprising of lung, bronchus, pleura, trachea, larynx and pharynx tissues. The lack of variation between the normal and tumor lung tissue diversity might be due to an overestimation in the diversity in the normal
tissue due to large variety of the epithelial tissue components of this organ that is comprised of lung, bronchus, pleura, trachea, larynx and pharynx.

Table 14: Diversity statistics describing of diversity, richness, evenness and variance of diversity reflecting the abundance of ESTs in each of the human tumors.

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Number of Sequences</th>
<th>Diversity H'</th>
<th>Richness S</th>
<th>Evenness E</th>
<th>Variance of H'</th>
<th>Standard Error of H'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Gastro-intestinal tissue</td>
<td>229598</td>
<td>9.028835</td>
<td>24175</td>
<td>0.894558</td>
<td>9.99E-06</td>
<td>0.003161</td>
</tr>
<tr>
<td>Lung tissue</td>
<td>195964</td>
<td>8.954499</td>
<td>23543</td>
<td>0.889527</td>
<td>1.44E-05</td>
<td>0.003793</td>
</tr>
<tr>
<td>Female reproductive</td>
<td>250839</td>
<td>8.924292</td>
<td>24819</td>
<td>0.881902</td>
<td>1.13E-05</td>
<td>0.003357</td>
</tr>
<tr>
<td>Connective tissue</td>
<td>82051</td>
<td>8.868648</td>
<td>16599</td>
<td>0.912685</td>
<td>2.56E-05</td>
<td>0.005058</td>
</tr>
<tr>
<td>Bone</td>
<td>50358</td>
<td>8.853011</td>
<td>13702</td>
<td>0.929421</td>
<td>3.74E-05</td>
<td>0.006112</td>
</tr>
<tr>
<td>Kidney</td>
<td>89069</td>
<td>8.834877</td>
<td>17058</td>
<td>0.906664</td>
<td>2.99E-05</td>
<td>0.005466</td>
</tr>
<tr>
<td>Mammary gland</td>
<td>106111</td>
<td>8.782686</td>
<td>16565</td>
<td>0.904029</td>
<td>2.19E-05</td>
<td>0.004683</td>
</tr>
<tr>
<td>Brain</td>
<td>174529</td>
<td>8.679998</td>
<td>18642</td>
<td>0.882726</td>
<td>1.71E-05</td>
<td>0.004138</td>
</tr>
<tr>
<td>Lymphatic tissue</td>
<td>148729</td>
<td>8.596777</td>
<td>16707</td>
<td>0.884116</td>
<td>1.86E-05</td>
<td>0.004308</td>
</tr>
<tr>
<td>Prostate</td>
<td>107355</td>
<td>8.504952</td>
<td>14591</td>
<td>0.887027</td>
<td>2.44E-05</td>
<td>0.004939</td>
</tr>
<tr>
<td>Testis tissue</td>
<td>104177</td>
<td>8.479037</td>
<td>13859</td>
<td>0.889096</td>
<td>2.44E-05</td>
<td>0.004942</td>
</tr>
<tr>
<td>Upper Gastro-Intestinal tissue</td>
<td>72982</td>
<td>8.407071</td>
<td>12470</td>
<td>0.891422</td>
<td>3.60E-05</td>
<td>0.006</td>
</tr>
<tr>
<td>Pancreas</td>
<td>102394</td>
<td>8.392749</td>
<td>14142</td>
<td>0.878187</td>
<td>3.09E-05</td>
<td>0.005563</td>
</tr>
<tr>
<td>Thyroid</td>
<td>33611</td>
<td>8.384825</td>
<td>9706</td>
<td>0.91333</td>
<td>7.34E-05</td>
<td>0.005969</td>
</tr>
<tr>
<td>Parathyroid</td>
<td>20602</td>
<td>8.383832</td>
<td>7045</td>
<td>0.946249</td>
<td>6.29E-05</td>
<td>0.007932</td>
</tr>
<tr>
<td>Skin</td>
<td>119604</td>
<td>8.377987</td>
<td>13274</td>
<td>0.882491</td>
<td>2.20E-05</td>
<td>0.004694</td>
</tr>
<tr>
<td>Liver</td>
<td>107936</td>
<td>8.286482</td>
<td>13687</td>
<td>0.870045</td>
<td>3.27E-05</td>
<td>0.005723</td>
</tr>
<tr>
<td>Adrenal gland</td>
<td>13221</td>
<td>8.232684</td>
<td>6274</td>
<td>0.941506</td>
<td>0.000129</td>
<td>0.011374</td>
</tr>
<tr>
<td>Bladder tissue</td>
<td>18295</td>
<td>8.165420</td>
<td>6611</td>
<td>0.928259</td>
<td>0.000104</td>
<td>0.010189</td>
</tr>
<tr>
<td>Pregnancy tissue</td>
<td>30351</td>
<td>7.966101</td>
<td>7091</td>
<td>0.898441</td>
<td>8.31E-05</td>
<td>0.009115</td>
</tr>
<tr>
<td>Eye</td>
<td>39123</td>
<td>7.833186</td>
<td>8086</td>
<td>0.870558</td>
<td>9.87E-05</td>
<td>0.009933</td>
</tr>
<tr>
<td>Muscle</td>
<td>23989</td>
<td>7.677075</td>
<td>5592</td>
<td>0.889674</td>
<td>0.00011</td>
<td>0.010511</td>
</tr>
<tr>
<td>Salivary gland</td>
<td>10355</td>
<td>7.2727</td>
<td>2891</td>
<td>0.912583</td>
<td>0.000192</td>
<td>0.013858</td>
</tr>
<tr>
<td>Nerve tissue</td>
<td>701</td>
<td>6.274774</td>
<td>609</td>
<td>0.978626</td>
<td>0.001336</td>
<td>0.036549</td>
</tr>
<tr>
<td>Pituitary gland</td>
<td>1392</td>
<td>6.217004</td>
<td>923</td>
<td>0.910566</td>
<td>0.002162</td>
<td>0.046495</td>
</tr>
<tr>
<td>Adipose tissue</td>
<td>720</td>
<td>6.033908</td>
<td>558</td>
<td>0.954074</td>
<td>0.002033</td>
<td>0.045086</td>
</tr>
</tbody>
</table>

Interestingly, “human tumor” categories were almost evenly split into those with increased expression diversity when compared to its normal tissue counterpart, and those
with decreased diversity. As could be seen from tables 15(a) and 15(b), tumors with increased diversity mostly originate from epithelial cells (carcinomas), while tumors with decreased diversity were mostly represented by the malignancies of non-epithelial tissues, e.g. brain (gliomas and astrocytomas), pituitary gland, nerve, muscle. The only exception to this list was skin. However, in our classification skin tumors included melanomas that are derived of melanocytes (non-epithelial skin components). Importantly, the changes in the diversity patterns observed in this study cannot be explained by the change in the alternative splicing, as our datapoints represent EST clusters (unigenes) rather than ESTs.

Table 15(a): The following table illustrates the tissues having increased diversity in tumors compared to normal tissue. T-statistics reflects the diversity in the EST abundance data of the Unigene system in human normal tissues and the corresponding cancerous tissues.

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Normal Number of ESTs</th>
<th>Normal Diversity</th>
<th>Normal Variance</th>
<th>Cancer Number of ESTs</th>
<th>Cancer Diversity</th>
<th>Cancer Variance</th>
<th>T-stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung tissue</td>
<td>158569</td>
<td>8.946467</td>
<td>1.80E-05</td>
<td>195964</td>
<td>8.954499</td>
<td>1.44E-05</td>
<td>-1.4110</td>
<td>0.158321</td>
</tr>
<tr>
<td>Female-reproductive</td>
<td>53761</td>
<td>8.859325</td>
<td>3.19E-05</td>
<td>250839</td>
<td>8.924292</td>
<td>1.13E-05</td>
<td>-9.8843</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Mammary gland</td>
<td>47829</td>
<td>8.400288</td>
<td>6.12E-05</td>
<td>106111</td>
<td>8.782687</td>
<td>2.19E-05</td>
<td>-10.56</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Pancreas</td>
<td>105643</td>
<td>8.831674</td>
<td>4.31E-05</td>
<td>102394</td>
<td>8.39275</td>
<td>3.09E-05</td>
<td>-8.8356</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Lower GI tissue</td>
<td>83030</td>
<td>8.208109</td>
<td>4.55E-05</td>
<td>229598</td>
<td>9.028836</td>
<td>9.99E-06</td>
<td>-110.17</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Bone</td>
<td>16172</td>
<td>7.874164</td>
<td>0.000206</td>
<td>50358</td>
<td>8.853011</td>
<td>3.74E-05</td>
<td>-62.717</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Adrenal gland</td>
<td>19762</td>
<td>7.602619</td>
<td>0.000195</td>
<td>13221</td>
<td>8.232684</td>
<td>0.000129</td>
<td>-34.975</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Liver</td>
<td>99207</td>
<td>7.801805</td>
<td>7.87E-05</td>
<td>107936</td>
<td>8.286482</td>
<td>3.27E-05</td>
<td>-64.869</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Thyroid</td>
<td>12031</td>
<td>7.501853</td>
<td>0.000338</td>
<td>33611</td>
<td>8.384826</td>
<td>7.34E-05</td>
<td>-43.537</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Connective tissue</td>
<td>16389</td>
<td>7.355674</td>
<td>0.000208</td>
<td>82051</td>
<td>8.868648</td>
<td>2.56E-05</td>
<td>-98.977</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Bladder tissue</td>
<td>8550</td>
<td>7.272814</td>
<td>0.000324</td>
<td>18295</td>
<td>8.16542</td>
<td>0.000104</td>
<td>-43.141</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Upper GI tissue</td>
<td>12363</td>
<td>7.166962</td>
<td>0.000325</td>
<td>72982</td>
<td>8.407071</td>
<td>3.60E-05</td>
<td>-65.300</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Salivary gland</td>
<td>2514</td>
<td>5.292427</td>
<td>0.001936</td>
<td>10355</td>
<td>7.2727</td>
<td>0.000192</td>
<td>-42.926</td>
<td>&lt;0.00001</td>
</tr>
</tbody>
</table>
Table 15(b): The following table illustrates the tissues having decreased diversity in tumors compared to normal tissue. T-statistics reflect the diversity in the EST abundance data of the Unigene system in human normal tissues and the corresponding cancerous tissues.

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Normal</th>
<th>Cancer</th>
<th>T-stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of ESTs</td>
<td>Diversity</td>
<td>Variance</td>
<td>Number of ESTs</td>
</tr>
<tr>
<td>Eye</td>
<td>160932</td>
<td>9.31899</td>
<td>1.18E-05</td>
<td>39123</td>
</tr>
<tr>
<td>Testis tissue</td>
<td>147432</td>
<td>9.250885</td>
<td>1.63E-05</td>
<td>104177</td>
</tr>
<tr>
<td>Pregnancy tissue</td>
<td>512435</td>
<td>9.013833</td>
<td>5.31E-06</td>
<td>30351</td>
</tr>
<tr>
<td>Kidney</td>
<td>107596</td>
<td>8.950741</td>
<td>2.55E-05</td>
<td>89069</td>
</tr>
<tr>
<td>Prostate</td>
<td>81107</td>
<td>8.883301</td>
<td>2.99E-05</td>
<td>107355</td>
</tr>
<tr>
<td>Lymphatic tissue</td>
<td>220930</td>
<td>8.833965</td>
<td>1.33E-05</td>
<td>148729</td>
</tr>
<tr>
<td>Brain</td>
<td>787007</td>
<td>8.804831</td>
<td>4.39E-06</td>
<td>174529</td>
</tr>
<tr>
<td>Skin</td>
<td>74232</td>
<td>8.713065</td>
<td>2.42E-05</td>
<td>119604</td>
</tr>
<tr>
<td>Nerve tissue</td>
<td>25579</td>
<td>8.553971</td>
<td>7.60E-05</td>
<td>701</td>
</tr>
<tr>
<td>Muscle</td>
<td>75841</td>
<td>8.449439</td>
<td>4.65E-05</td>
<td>23898</td>
</tr>
<tr>
<td>Adipose tissue</td>
<td>12417</td>
<td>8.063999</td>
<td>0.000141</td>
<td>720</td>
</tr>
<tr>
<td>Pituitary gland</td>
<td>19430</td>
<td>7.510406</td>
<td>0.000351</td>
<td>1392</td>
</tr>
</tbody>
</table>

e) An analysis of the unigenes for putative tumor biomarkers

The tissue-specific unigenes exclusively represented by cancer or normal ESTs (Table 12) represent a unique mining resource that may be used to produce biomarker candidates for a particular cancer type or for a set of related cancers. Given an enormous share occupied by these exclusive gene clusters in a wide range of human tissues, extracting the gene clusters with differences between the normal and cancer tissues, poses interesting statistical problem.

A total of 27 human tumor tissues having a corresponding normal tissue were used to identify the biomarkers. Pair-wise Mann-Whitney test was performed on the normalized EST abundance data of each Unigene gene cluster. The gene clusters which show significant variation in this pair-wise non-parametric test across the 27 tissues were filtered for further analyses. The tissue groups are marked by only their pathological
characteristic (normal or tumor) leaving their tissue identities in the filtered gene clusters. This data is scaled to illustrate EST abundance as frequency. Metastats is used on this grouped normal and tumor frequency data to extract differentially abundant features (gene clusters). The gene clusters with a significant degree of differential abundance were only considered as biomarkers.

Of the 123,687 gene clusters in the current Unigene database system, only 2863 were observed to have significant variation (P-values < 0.05) by pairwise non-parametric testing. These significant unigenes were ran through Metastats software to filter these unigenes differentiating normal and tumor samples in order to pinpoint ones associated with low false discovery rate (<0.05). This filter resulted in a final list of candidate unigenes (1751 gene clusters) comprised of 668 tumor-specific and 1083 normal-specific expressing gene clusters.

Heatmap of the EST abundance of these significant unigenes showed a clear difference between the normal and tumor groups (Figure 6). Selected unigenes were mostly represented by the protein-coding gene clusters and, to some extent, by non-coding unigenes, with 139 non-coding unigenes in tumor biomarker list and 157 unigenes in the normal (anti-cancer) list. The Genbank accession numbers for all listed biomarker candidate transcripts are presented in the Appendix.

There were 526 cancer specific and 826 normal specific protein-coding gene clusters whose expression patterns were statistically different between normal and tumor tissues.
These unigenes are potential candidate biomarkers that may help to identify various subtypes of human cancer. An entire list of these biomarkers is also given in an Appendix.

Figure 6: Heatmap of the EST abundance data from 27 different human tissues from the normal and corresponding cancer cDNA libraries of the unigene system. Blue bar at the top of heatmap identifies the normal tissues, while red bar marks their malignant counterparts.

f) Functional analysis of protein-coding unigenes identified as tumor biomarker candidates

Functional analysis of protein-coding unigenes identified as tumor biomarker candidates was performed using the molecular pathway information provided in the KEGG
knowledge base (Kanehisa et al. 2008). A novel tool, KEGG Pathway painter (KPP) was built in-house with the purpose to automatically identify the function of the candidate biomarker unigenes through high-throughput analysis of the KEGG pathway maps. The methodology underlying KPP processing and the detailed description of KPP are available in the Appendix. Briefly, the complete sets of human molecular pathways were extracted by KPP separately for the cancer-specific and normal-specific significant unigenes. The enrichment of the pathways with the genes that belong to particular functional categories was assessed using the DAVID functional analysis framework (Dennis et al. 2003; Huang et al. 2009).

Figure 7: KPP representation of the KEGG cell cycle pathway map painted using genes differentially expressed in the malignant tumors (red background) as compared to normal human tissues.
The enrichment analysis revealed that sets of the pathways highlighted by the cancer and normal biomarker candidates are substantially different. In this context, the normal specific genes could be called anti-cancer markers as they are exclusively present in the normal tissue, and are devoid in cancerous tissue. The fundamental biological processes enriched by cancer specific unigenes as represented by KPP analysis include cell cycle regulation, p53 signaling, ubiquitin mediated proteolysis and apoptosis. The enhanced proliferation activity influenced by the set of tumor biomarker candidate unigenes in the perspective of cell cycle is illustrated in Figure 7.

![CALCIUM SIGNALING PATHWAY](image)

Figure 8: Illustration of the Genes expressing more abundantly in the normal tissue (red background) compared to the cancerous tissue on the KEGG Calcium signaling pathway.
Wide set of cancer specific pathways pertaining to chronic myeloid leukemia, melanoma, glioma, pancreatic cancer, prostate cancer, colorectal cancer, bladder cancer, small cell and non-small cell lung cancer was also shown as being enriched by these cancer-specific markers. The extensive range of cancer pathways enrichment supports the cancer-specificity of the unigenes associated in the context.

The anti-cancer markers represented by the unigenes abundant in normal tissues, on the other hand, demonstrated an enrichment of the broad variety of biological pathways and networks. The cascades highlighted by this analysis include calcium signaling, MAPK signaling, insulin signaling, gonadotropin-releasing hormone (GnRH) signaling and T-cell receptor signaling. The enrichment of calcium signaling pathway by the anti-cancer markers (Specific to normal tissue) is depicted visually in figure 8. The specific emphasis identified in the context of normal tissue in comparison to cancer tissue is signaling in a broad sense. Speaking generally, our analysis redefined cancer as the loss of tissue-specific signaling, while, at the same time, cancer demonstrated gains in the proliferating system. This definition is equivalent to the classic understanding of the tumor phenomenon. The full set of the pathways enriched by the cancer and anti-cancer biomarkers could be found in the appendix.

**Conclusion**

In this study, the diversity of the human tissue specific transcriptomes was studied. After normalization, relative abundancies of the ESTs derived from normal human tissues were compared to these collected using their malignant counterparts. An analysis of the
abundance of EST in the UniGene database was proven useful as a tool for a search for candidate biomarkers for human tumors. Functional analysis of the protein-coding biomarker candidates described cancer as the loss of tissue-specific signaling with simultaneous gain in the proliferating system.
4. Effects of the tumor-specific telomere rearrangements on the adjacent gene expressions

Rationale
Telomere rearrangements may result in the disturbance of the expression levels of adjacent genes. Telomere position effect may provide a mechanism to incrementally alter phenotype of the cancer cells. Telomere-related cancer studies concentrated at the structural changes have to be complemented by studies of gene expression changes. Here we hypothesize that the telomere rearrangements commonly found in cancer cells perturb an expression of the genes adjacent to telomeres in human tumors.

Background
Telomeres are made of guanine (G) rich repetitive DNA, composed of TTAGGG motifs representing the ends of chromosomes. DNA loss incurred during the DNA replication leads to the formation of 3’ G-rich protrusions which transforms into G-quartet structure at both the chromosomal ends (Pommier, Lebeau et al. 1995). The G-rich single stranded overhangs, with the help of telomeric repeat binding factors (TRF1 and TRF2) bind to the double-stranded telomeric region, forming a lasso-like loop structure called the T(telomere)-loop (Shin, Hong et al. 2006; Verdun and Karlseder 2007). This T-loop
structure is supposed to protect the chromosome ends from recognition by recombination and double strand DNA-repair machinery.

Telomeres also undergo length alterations during tumor development. These alterations have been shown to be independent markers of cancer prognosis (Bisoffi, Heaphy et al. 2006). Normal human somatic cells stop dividing after a finite number (~50) of replication cycles, a phenomenon known as the “Hayflick limit” (Hayflick 1965). This limitation of cell cycle is associated with the reduction of the number of telomere repeats at the chromosomal ends after each cycle. The reduction of the telomeres activates p53 which prevents the damage of cells by inducing DNA damage repair and inhibiting further cell division, causing cellular senescence (Artandi and Attardi 2005). However, some cells are exempt from the replicative senescence, particularly, the immortal germ cells and some tissue stem and progenitor cells expressing special telomere restoring enzyme called telomerase (Sherr and DePinho 2000).

Telomerase is a complex ribonucleoprotein that consists of the a DNA polymerase and a RNA template components and possesses reverse transcriptase activity at the telomeres (Cech 2004). In human cells, this enzyme stabilizes telomere length by adding TTAGGG repeats onto the telomeric ends during the S phase of cell cycle (Hug and Lingner 2006). The reactivation of telomerase in the somatic cells leads to the uncontrollable cell division, which might further pave the way to cancer. In the cells with p53 mutations, an expression of telomerase lifts the “Hayflick limit” and helps to maintain genomic instability (Finkel, Serrano et al. 2007).
Telomere maintenance is essential to the proliferation of tumor cells as expression of the mutant catalytic subunit of human telomerase results in complete inhibition of telomerase activity, reducing the telomere length, which ultimately lead to the death of tumor cells (Hahn, Stewart et al. 1999). Moreover, the minimal set of molecular events required for direct tumorogenic conversion of normal human epithelial and fibroblast cells requires an abnormal expression of the telomerase catalytic subunit (hTERT) in combination with overexpression of two oncogenes, the simian virus 40 large-T oncoprotein and an oncogenic allele of H-ras (Hahn, Counter et al. 1999). Indeed, a majority of human tumor cells acquire immortality through expression of the catalytic subunit of telomerase (hTERT) (Hiyama and Hiyama 2002). Approximately 10% of human cancers do not show evidence of telomerase activity, and a subset of these maintain telomere lengths by a recombination-based mechanism termed alternative lengthening of telomeres (ALT). The ALT phenotype, relatively common in certain sarcomas and germ cell tumors, is very rare in carcinomas. This alternative mechanism of telomere maintenance does not depend on the actions of telomerase (Stewart 2005).

In addition to the telomerase that is directly responsible for addition of telomeric sequences to the ends of the chromosomes, a plethora of other proteins also play important roles in the regulation of telomere length maintenance (de Lange 2005). Simultaneous and balanced upregulation of genes encoding telomere-associated proteins in cancer cell is an unlikely event. As active telomerase alone is not sufficient for preserving normal functionality of the telomere-associated protein complex in cancer
cells, many telomerase-expressing tumors exhibit chromosomal instability triggered by dysfunctional telomeres (Gisselsson and Hoglund 2005; Calcagnile and Gisselsson 2007). Invasive tumors frequently demonstrate intra-tumoral heterogeneity of telomere lengths that include both an increase in telomere length over the normal range and telomere shortening (Meeker, Hicks et al. 2004; Hansel, Meeker et al. 2006; Maida, Kyo et al. 2006).

In eukaryotic yeast cells, particularly in *Saccharomyces cerevisiae*, genes located near telomeres undergo reversible silencing (Tham and Zakian 2002). This effect, termed as the telomere position effect (TPE), involve changes in the chromatin conformation and is dependent on both the distance from the telomere and the telomere length (Tham and Zakian 2002). The silencing effect of the genes near telomeres, and the spontaneous reactivation of these genes have been described in HeLa cells (Baur, Zou et al. 2001; Baur, Shay et al. 2004). However, the literature on possibility of TPE in mammalian cells is scarce and conclusions remain controversial. Some researchers failed to find evidence for TPE and concluded that in higher eukaryotes the gene expression is independent of telomere length (Bayne, Broccoli et al. 1994; Sprung, Sabatier et al. 1996). Others demonstrated an influence of telomeres on the expression of the adjacent transgenes within the human and mouse cell lines (Baur, Zou et al. 2001; Koering, Pollice et al. 2002; Pedram, Sprung et al. 2006). A study of the expression of endogenous genes located near telomeres in human fibroblasts revealed a discontinuous pattern of altered gene expression during senescence-associated telomere shortening (Ning, Xu et al. 2003).
Subtelomeric regions may buffer or facilitate the spreading of silencing that emanates from the telomere (Ottaviani, Gilson et al. 2008). As these regions are particularly prone to recombination in tumor cells, they may add an extra level of the complexity to the problem reviewed.

Telomere rearrangements observed in cancer cells may result in the disturbance of the expression levels of the adjacent genes, and this in genomic instability. So far, the studies of the telomeres in cancer cells were concentrated at the structural changes in the telomeres themselves and the quantification of the telomere-associated molecules, including telomerase and telomere-binding proteins TRF1, TRF2, TIN1, POT1, TPP1, Cdc13p (Hug and Lingner 2006). In our opinion, the characterization of expression levels of the genes located near telomeres in tumors and normal tissues may shed light on the effects of TPE in the physiology of cancer. Therefore, we assessed effects of the telomere position and its rearrangements in the cancer cell studying the perturbations of expression of the genes adjacent to telomeres in human tumors.

**Hypothesis**

Telomere rearrangements commonly found in cancer cells perturb an expression of the genes adjacent to telomeres in human tumors.

**Materials and Methods**

To investigate the possibility that rearrangement of telomeres in cancer cells may lead to the disturbance in the expression levels of adjacent genes, we retrieved the publicly
available data describing changes in gene expression patterns in the prostatic carcinoma, a tumor with common findings of the telomerase reactivation and telomere erosion (Vukovic, Park et al. 2003; Meeker, Hicks et al. 2004; Meeker 2006). Prostatic carcinoma dataset GDS1439 was obtained from Gene Expression Omnibus (GEO) repository (Barrett and Edgar 2006). This dataset describes raw data collected in course of the microarray profiling of primary (N=7) and metastatic (N=6) prostate carcinoma samples along with the samples of normal prostate tissue (N=6). The microarray platform employed for this study was the Affymetrix GeneChip U133 Plus 2.0 (GPL570), an oligonucleotide array covering an entire human genome with over 47,000 transcript-specific probes.

The GenBank annotated Human Genome (Build 36, Version 2) sequence was downloaded from the NCBI Genome repository (Benson, Karsch-Mizrachi et al. 2007). Contig information for each of the human chromosomes was obtained from the NCBI MapViewer (Wheeler, Barrett et al. 2007). R data analysis package with Bioconductor (Reimers and Carey 2006; Gregory Alvord, Roayaei et al. 2007) and Perl scripting language were used to perform the computational analysis.

Based on the GenBank annotation and the contig information, locations for each human gene on their respective chromosomes were calculated as absolute numeric coordinates defined as nucleotide position from the end of the available chromosome sequence to the gene. All genes were associated with their expression data based on their HUGO gene symbols. A total of 17522 out of 28479 genes annotated in the human genome at the time
of the project were in this manner assigned to the human genome map and associated with their respective expression levels. These mapped transcripts were then categorized into telomere associated and non-telomere (body of the chromosome) associated fractions based on their position. Two computational experiments were performed (see Fig 9). In experiment I, 10% of genes on either extreme of each chromosome (except acrocentric chromosomes with only one telomere), accounting to a total of 20% per chromosome (10% for acrocentric chromosome) were defined as telomere associated genes, and remaining 80% were defined as centromere associated or non-telomeric genes. In experiment II the telomere associated gene definition was narrowed down, to include only 5% on each extreme of chromosome, accounting to a total of 10% telomere associated genes per chromosome.

**Experiment I**

\[ \begin{array}{c}
\text{T} & & \text{C} & & \text{T} \\
10\% & & \text{80}\% & & 10\% \\
\end{array} \]

**Experiment II**

\[ \begin{array}{c}
\text{T} & & \text{C} & & \text{T} \\
5\% & & \text{95}\% & & 5\% \\
\end{array} \]

Figure 9: All the genes in each of the human chromosomes were subdivided into sum of their telomere associated (T) genes and rest as the centromere associated (C) genes based on the number of the genes as shown above in two experiments.
The gene expression data were background corrected, normalized and the individual gene expression values were calculated from the CEL files using Robust Multichip Average (RMA) method of background adjusted, normalized, and log transformed perfect match (PM) values (Irizarry, Bolstad et al. 2003). RMA software is a part of the R package affy that is a module of Bioconductor. It provides an interactive environment for data analysis and exploration of Affymetrix oligonucleotide array’s probe level data (Gautier, Cope et al. 2004). The mean expression value of all the samples in each sample type (B-normal, P-primary and M-metastatic) was taken as the representative expression for that gene in the particular stage of the tumor development or in normal prostatic tissue.

Results and Discussion

a) The definition of the over/underexpressed genes

The dataset GDS1439 contains gene expression profiles for three types of prostate samples: primary (N=7) and metastatic (N=6) prostatic carcinoma samples along with the samples of normal prostate tissue (N = 6). Therefore, in this study over- and underexpressed genes could be derived in three different comparisons: metastatic (M) vs. primary (P) samples, metastatic (M) vs. normal (B) samples and primary (P) vs. normal (B) samples. For the purpose of this study we termed the genes that significantly change their expression in M/B comparison. If M/B was greater than or equal to 1.4, then the gene was considered to be overexpressed; else if M/B less than or equal to 0.71 (i.e. 1/1.4), then the gene was considered to be underexpressed. A total of 528 over/under
expressed genes pertinent to the GEO dataset GDS1439 correspond to the 3% of the genes matched with GenBank annotation of the whole Human genome (Benson, Karsch-Mizrachi et al. 2007). In Experiment I, the number of over/under expressed genes present in telomere-associated regions covering 20% of coding human chromosome DNA was 15% of the total number of over/under expressed genes.

![Graph showing percentage of gene expression change](image)

Figure 10: The graph depicts the percentage of genes that change their expression on each chromosome in the telomere and centromere associated regions. On average, the percentage of the over/under expressed genes varies by a greater degree in the telomere associated region compared to the body of chromosome, when individual chromosomes are taken into account.

The percentage of centromere associated genes that change their expression was approximately the same for every chromosome, while the percentage of telomere associated genes that change their expression was chromosome specific, as demonstrated by high variation in a number of over/under expressed genes on individual chromosomes (see Fig 10). Chromosome Y was devoid of overexpressed genes in the non-telomere
associated regions. Telomeric regions of many other chromosomes were shown to lack either under or over expressed genes or both.

b) Correlation with tumor stage

The patterns of the changes of gene expression at different stages of tumor development might be different for the telomeric genes and genes in the body of chromosome. The metastatic to normal (M/B) gene expression ratio was initially introduced in order to estimate the variation of expression in telomeric as compared to centromere associated genes. We interpreted this measure as representation of expression change imbalance between telomeric and centromeric regions. Mann-Whitney test (Mann and Whitney 1947), a non-parametric testing procedure was employed for assessing the statistical significance of the difference in the gene expressions in the telomere and non-telomere associated regions in each of the chromosome.

The gene expression ratios reflecting the shift of the tissue between three studied states (M/B, M/P and P/B) were calculated as independent vectors for all genes associated with the telomeres and the body of chromosomes. The differences between the groups of corresponding ratios for telomere-associated and body of chromosome-associated genes were tested using the Mann-Whitney test under the null hypothesis that they are similar. Thus, the correlation between the location of the gene and the change in its expression was tested. The tests were performed for the whole human genome (i.e. all telomeres together vs. all bodies of chromosomes together) at once, and for chromosome separately.
The chromosomes that showed statistically significant variation in Experiment I, were again tested for variation in Experiment II. The chromosomes that revealed significant differences in both experiments were again tested for variation in the telomere and non-telomere associated regions using metastatic to primary (M/P) and primary to normal (P/B) gene expression ratios. These analyses allowed to correlate the changes in the expression for genes located in telomere and centromere associated regions with the stage of the tumor development.

In Experiment I, the variation of the M/B ratio across the telomere and centromere associated regions was significant by Mann-Whitney test (p-values < 0.005) in chromosomes 4, 5, 8, 9, 16 and X. There was a marginal significance (p-value < 0.05) in the variation for the chromosomes 2, 10 and Y as assessed by the same test. Mann-Whitney test results also showed the difference in the M/B ratio (p-values < 0.005) for chromosomes 1, 4, 5, 7, 8, 16 and X, and a marginal variation (p-value < 0.05) for chromosome 13 in Experiment II. The chromosomes that show significant variation in M/B ratio in both Experiments I and II (4, 5, 8, X, 16) were further tested for variation in the M/P ratio and P/B ratio in the telomere and non-telomere associated regions.

The expression change imbalance of the M/P ratios for the telomeric and body of chromosome associated was significant when all of these chromosomes were taken into account both in Experiment I (P < 8.53e-14) and Experiment II (P < 1.05e-11). In case of P/B ratios, expression change imbalance was not significant in almost all the cases, except for chromosomes 12, 15 and 16 using telomere definition of Experiment II (see
Table 16). Differences in P/B ratios were not significant even when all the chromosomes are compared together representing the genome.

Table 16: Telomere associated genes are more likely to change their level of expression as compared to non-telomere genes. When metastatic tumors were compared to normal prostates using (M/P) ratio, these changes were significant for 11 out of 24 chromosomes (Exp I) or 8 out of 24 chromosomes (Exp II), while in the primary tumor to normal tissue (P/B) comparisons only 3 out of 24 chromosomes demonstrated expression change imbalance. HS-Highly significant (p-value < 0.0005), S-Significant (p-value < 0.005), MS-Marginally significant (p-value < 0.05), NS-Not significant (p-value ≥ 0.05).

<table>
<thead>
<tr>
<th>Chr</th>
<th>Metastatic / Normal</th>
<th>Metastatic / Primary</th>
<th>Primary / Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experiment I</td>
<td>Experiment II</td>
<td>Experiment I</td>
</tr>
<tr>
<td>1</td>
<td>NS</td>
<td>S</td>
<td>NS</td>
</tr>
<tr>
<td>2</td>
<td>MS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>3</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>4</td>
<td>HS</td>
<td>HS</td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>HS</td>
<td>S</td>
<td>HS</td>
</tr>
<tr>
<td>6</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>7</td>
<td>NS</td>
<td>S</td>
<td>NS</td>
</tr>
<tr>
<td>8</td>
<td>HS</td>
<td>S</td>
<td>MS</td>
</tr>
<tr>
<td>9</td>
<td>S</td>
<td>NS</td>
<td>MS</td>
</tr>
<tr>
<td>10</td>
<td>MS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>11</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>12</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>13</td>
<td>NS</td>
<td>MS</td>
<td>NS</td>
</tr>
<tr>
<td>14</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>15</td>
<td>MS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>16</td>
<td>S</td>
<td>HS</td>
<td>MS</td>
</tr>
<tr>
<td>17</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>18</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>19</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>20</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>21</td>
<td>MS</td>
<td>NS</td>
<td>MS</td>
</tr>
<tr>
<td>22</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>X</td>
<td>HS</td>
<td>S</td>
<td>HS</td>
</tr>
<tr>
<td>Y</td>
<td>MS</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>
As could be seen from the Table 1, the variation in the ratios of the gene expression between the telomere and non-telomere (centromere) associated regions increases with the progression of the prostatic tumor and is the most pronounced for the metastatic tumors.

c) Variation in distribution of expression changes

To assess possible variation in distribution of individual gene expression values in each of the metastatic (M) and normal (B) samples, these values were subjected to pairwise t-test. The p-values obtained in the t-test were adjusted to control the false discovery rate (FDR) using Multtest module of R package (Pollard, Dudoit et al. Dec 2004) that utilizes Benjamini and Hochberg (BH) method (Benjamini and Hochberg 2000). The distributions of these adjusted p-values from telomere and non-telomere associated genes were also constructed using R. The overall variation in the distribution of p-values in the two regions (telomere and centromere associated) was tested using Kolmogorov-Smirnov test (Conover 1971). This variation in the p-values was also verified by t-test and rank test (Mann and Whitney 1947).

The paired t-test between the metastatic (M) and normal (B) expression data for every gene in the genome resulted in a total of 1521 significantly varying genes (p-value < 0.05) after controlling the false discovery rate. Interestingly, the distributions of these p-values in the telomere and centromere associated genes as defined in Experiment I were found to be more dense at lower cutoffs (p-value < 0.01), showing a significant variation between metastatic and normal samples. The variation between the telomere and
centromere associated genes in the context of metastatic and normal samples was assessed by testing the variation between these p-values by a number of statistical tests (Table 17). A significant variation was seen in these p-values corresponding to telomere and centromere associated genes, returning the p-values in the order of 0.001. The rank test between the first 1000 p-values of telomere and centromere associated genes show even more substantial variation described by a p-value less than 3.18796e-05.

Table 17: The p-values obtained from the t-test between the metastatic and normal expression data were subjected to various statistical tests to find the variation between the telomere and centromere associated genes. The significant p-values (< 0.05) corresponding to genes in the telomere and centromere associated regions indeed show variation in all three tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>p-values for all the genes, including not significantly changed</th>
<th>p-values &lt; 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov test</td>
<td>0.06352</td>
<td>0.003997</td>
</tr>
<tr>
<td>t-test</td>
<td>NS</td>
<td>0.001201</td>
</tr>
<tr>
<td>Rank test (Mann-Whitney test)</td>
<td>NS</td>
<td>0.000886426</td>
</tr>
</tbody>
</table>

**d) Gene Ontology analysis**

For each of the the telomere and centromere associated genes with p-values remaining significant after FDR adjustment, the gene ontology (GO) term descriptions defining its relationship to certain cellular compartment, molecular function and/or biological process were retrieved. A comparative analysis of the GO terms distributions between the telomere and non-telomere associated genes was done using the goTools package of R (Paquet and Yang 2007) and presented on Fig. 11.
Figure 11: The bar graph in each GO category represents the ratio of the number of the direct GO identifiers associated with the telomere associated (shown in black) and centromere associated (shown in grey) genes to the number of GO identifiers associated with all the genes in the Affymetrix GeneChipU133 Plus 2.0 oligonucleotide array. As could be seen from the Figure, relative abundances of the GO terms in each GO category for the telomere and centromere associated genes were similar.
A comparative analysis of the GO terms in the telomere and centromere associated genes was performed with reference to the number of probe identifiers and GO identifiers associated with all the genes in the array. GO term analysis revealed that differentially expressed telomere and non-telomere associated genes are involved in a wide variety of biological processes, carrying various molecular functions in different cellular locations. The relative abundances of the GO terms and probe identifiers in each GO category for the telomere and centromere associated genes were almost equal (see Fig 11), revealing the functional similarity of the telomere and non-telomere associated genes based on GO terms. This functional similarity points out that the variation between the telomere and non-telomere associated genes was due to the effect of their adjacent to telomeres rather than any bias in their functional importance.

e) Defining the maximal length of the subtelomeric fragment that might be influenced by telomere rearrangements in cancer cells

In the Experiments I and II genes were called “telomeric” or “centromeric” (non-telomere associated) based on the overall number of genes on each chromosome (Fig 9). However, in the perspective of TPE, a physical distance from the gene to the telomere is important. We hypothesized that the plotting the gene expression ratios obtained in M/B (metastatic to normal), M/P (metastatic to primary) and P/B (primary to normal) comparisons across the chromosome length may help to define the distances with measurable TPE.

The gene locus table of the GenBank annotated Human genome (Build 36, Version 2) sequence was downloaded from the NCBI. Exact physical coordinates for each gene have
been utilized to map M/B ratios for the genes present on Affymetrix chip across chromosomal lengths. These plots have not revealed any discernible pattern that might help to define the length of the telomeric fragment susceptible to TPE. However, the plots provided for the clear illustration of the concept that the scale of the changes in the gene expression dramatically increases with the progression of the prostatic tumors (see Fig 12 for an example of the plot).

Figure 12: This graph plots M/B, M/P and P/B gene expression ratios (y-axis) to the physical distance from the telomere (x-axis) on human chromosome X. The variation in the M/B ratio (blue) is larger than the variation in M/P ratio (purple), while the P/B ratio (yellow), is characterized by smallest variation along the length of chromosome X.
An alternative approach to find the characteristic length of the subtelomeric region susceptible to TPE is to calculate gene expression changes in a variety of different subtelomere lengths windows that were arbitrary chosen from the range of 0.2MB to 50MB. The tested arbitrary telomeric lengths were stepwise increased up to as the maximum physical length that can be assigned to telomere from the end of the chromosome to centromere of this chromosome in each of its arms. This telomere length window was defined globally for the whole genome, while chromosomes with shorter arms ended up with a smaller subtelomeric region. The genomic regions accounting for 5MB of physical length on either side of the centromere were truncated as the heterochromatic parts of the chromosomes surrounding its centromeres might also effect the gene expression levels via CPE (centromere positioning effect).

The metastatic to normal (M/B) expression ratios were used to find the variation between the telomere genes and genes in the body of chromosome for 20 gradually increasing subtelomeric windows. The analysis was performed by separately for a) all genes changing their expression, b) upregulated genes only and c) downregulated genes only. Cumulative changes in the expression levels of the genes located within given subtelomeric window and the rest of the chromosome were compared.
The distributions of the transcripts in the upregulated groups in the telomere and centromere associated regions were calculated by sorting them into bins of the range 1 - 1.1, 1.1 - 1.2, 1.2 - 1.3, 1.4 - 1.5, and > 1.5. Similar methodology was used for downregulated groups where the bins representing the ranges of 1 - 1/1.1, 1/1.1 – 1/1.2, 1/1.2 – 1/1.3, 1/1.3 – 1/1.4, 1/1.4 – 1/1.5 and < 1/1.5. The distribution plots were built for each of the subtelomere windows; Fig 13 represents the distribution of upregulated and downregulated genes for subtelomeric window of 5 Mb. Although these distributions were difficult to assess visually, statistical analysis performed on the frequency distribution help to dissect expression change imbalance between the telomere and centromere associated regions in specific subtelomeric windows. Chi-square test of
homogeneity was used to find the extent of variation between the distributions in the
telomere and centromere associated genes.

The homogeneity test was performed separately for a) all genes changing their
expression, b) upregulated genes only and c) downregulated genes only. Genes located
within given subtelomeric window and the rest of the chromosomes were compared.
Differentially expressing groups DEG1, DEG2 and DEG3 comprising of transcripts with
varying levels of changes in their expression were formed by excluding transcripts with
less pronounced expression changes. M/B ratios in the range 1/1.1 to 1.1, 1/1.2 to 1.2 and
1/1.3 to 1.3 were excluded in DEG1, DEG2 and DEG3, respectively (Table 18). Almost
all the tests seem to be significant when gene expression changes were considered wit no
regards to their directionality, except for subtelomere windows of 2, 4, 5 and 6 Mb. The
position effects for downregulated group of genes were significant when all expression
ratios were considered, but lost significance when tests were performed only for genes in
DEG1 and DEG2 groups. The upregulated group of genes demonstrated substantial
position effect in telomeres with lengths ranging from 0.5 Mb to 5Mb. The smaller DEG2
and DEG1 groups with larger differences in expression levels seem pinpoint subtelomeric
windows with length range of 2 to 4Mb, with largest effect at the subtelomeric window of
3Mb.

To study the correlation between the significance of gene expression changes and gene
location in the subtelomere region, chi-square test using a 2x2 contingency table was
performed. For each arbitrary subtelomeric window, a 2x2 contingency table consisting of
the number of significantly changing genes within the subtelomere window, significantly changing genes outside of the subtelomere window, not significantly changing genes within the subtelomere window and not significantly changing genes outside of the subtelomere window was computed for two different significance levels, DEG2 and DEG3. The chi-square tests were performed on each of these tables using a) all genes changing their expression, b) upregulated genes only and c) downregulated genes only.

Table 18: The table represents the p-values obtained from the chi-square test of homogeneity between the distribution of M/B ratios in the telomere and centromere associated regions for varying subtelomeric windows. P-values shown in grey are not significant (>=0.05). In the differentially expressing groups (DEG), the transcripts showing minimal changes of expression were excluded. In DEG1, DEG2 and DEG3 M/B ratios in the range 1/1.1 to 1.1, 1/1.2 to 1.2 and 1/1.3 to 1.3 were excluded respectively.

<table>
<thead>
<tr>
<th>Telomere Length</th>
<th>All (Up + Down regulated)</th>
<th>Downregulated</th>
<th>Upregulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 MB</td>
<td>8.47E-21</td>
<td>0.00047</td>
<td>0.549</td>
</tr>
<tr>
<td></td>
<td>7.37E-118</td>
<td>0.361</td>
<td>0.675</td>
</tr>
<tr>
<td></td>
<td>2.53E-18</td>
<td>0.284</td>
<td>0.747</td>
</tr>
<tr>
<td>40 MB</td>
<td>4.33E-20</td>
<td>3.59E-07</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>9.12E-114</td>
<td>0.017</td>
<td>0.319</td>
</tr>
<tr>
<td></td>
<td>1.42E-11</td>
<td>0.05</td>
<td>0.245</td>
</tr>
<tr>
<td>30 MB</td>
<td>3.37E-30</td>
<td>9.93E-10</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>3.53E-95</td>
<td>0.502</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>2.24E-10</td>
<td>0.999</td>
<td>0.006</td>
</tr>
<tr>
<td>25 MB</td>
<td>8.01E-33</td>
<td>2.00E-08</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>7.66E-88</td>
<td>0.099</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>1.47E-07</td>
<td>0.84</td>
<td>0.045</td>
</tr>
<tr>
<td>20 MB</td>
<td>1.35E-51</td>
<td>2.07E-10</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>1.47E-77</td>
<td>0.42</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>3.29E-08</td>
<td>0.957</td>
<td>0.144</td>
</tr>
<tr>
<td>15 MB</td>
<td>1.60E-53</td>
<td>5.85E-14</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>6.36E-67</td>
<td>0.121</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.002048</td>
<td>0.414</td>
<td>0.105</td>
</tr>
<tr>
<td>10 MB</td>
<td>2.01E-47</td>
<td>1.80E-13</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>4.33E-44</td>
<td>0.047</td>
<td>0.211</td>
</tr>
<tr>
<td></td>
<td>0.02297</td>
<td>0.158</td>
<td>0.123</td>
</tr>
<tr>
<td>9 MB</td>
<td>4.81E-51</td>
<td>1.73E-10</td>
<td>0.444</td>
</tr>
<tr>
<td></td>
<td>2.43E-42</td>
<td>0.049</td>
<td>0.346</td>
</tr>
<tr>
<td></td>
<td>0.025999</td>
<td>0.156</td>
<td>0.215</td>
</tr>
<tr>
<td>8 MB</td>
<td>2.46E-45</td>
<td>7.64E-10</td>
<td>0.748</td>
</tr>
<tr>
<td></td>
<td>2.62E-41</td>
<td>0.037</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>0.006398</td>
<td>0.072</td>
<td>0.489</td>
</tr>
<tr>
<td>7 MB</td>
<td>8.34E-44</td>
<td>9.65E-11</td>
<td>0.616</td>
</tr>
<tr>
<td></td>
<td>1.17E-33</td>
<td>0.064</td>
<td>0.472</td>
</tr>
<tr>
<td></td>
<td>0.017777</td>
<td>0.109</td>
<td>0.466</td>
</tr>
<tr>
<td>6 MB</td>
<td>1.81E-48</td>
<td>1.57E-11</td>
<td>0.534</td>
</tr>
<tr>
<td></td>
<td>4.36E-320</td>
<td>0.12</td>
<td>0.439</td>
</tr>
<tr>
<td></td>
<td>5.99E-32</td>
<td>0.283</td>
<td>0.406</td>
</tr>
<tr>
<td>5 MB</td>
<td>2.27E-54</td>
<td>1.24E-10</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>5.38E-291</td>
<td>0.289</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>2.27E-28</td>
<td>0.291</td>
<td>0.18</td>
</tr>
<tr>
<td>4 MB</td>
<td>2.03E-57</td>
<td>1.73E-10</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>1.17E-256</td>
<td>0.049</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>5.85E-28</td>
<td>0.223</td>
<td>0.013</td>
</tr>
<tr>
<td>3 MB</td>
<td>1.02E-45</td>
<td>7.64E-10</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>4.21E-213</td>
<td>0.037</td>
<td>8E-04</td>
</tr>
<tr>
<td></td>
<td>1.80E-28</td>
<td>0.387</td>
<td>4E-04</td>
</tr>
<tr>
<td>2 MB</td>
<td>9.12E-31</td>
<td>9.65E-11</td>
<td>0.616</td>
</tr>
<tr>
<td></td>
<td>7.90E-161</td>
<td>0.064</td>
<td>0.472</td>
</tr>
<tr>
<td></td>
<td>3.71E-17</td>
<td>0.109</td>
<td>0.466</td>
</tr>
<tr>
<td>1 MB</td>
<td>8.93E-16</td>
<td>1.57E-11</td>
<td>0.534</td>
</tr>
<tr>
<td></td>
<td>7.60E-87</td>
<td>0.12</td>
<td>0.439</td>
</tr>
<tr>
<td></td>
<td>1.85E-11</td>
<td>0.283</td>
<td>0.406</td>
</tr>
<tr>
<td>0.75 MB</td>
<td>2.05E-11</td>
<td>1.24E-10</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td>1.14E-73</td>
<td>0.289</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>1.77E-08</td>
<td>0.291</td>
<td>0.18</td>
</tr>
<tr>
<td>0.5 MB</td>
<td>3.39E-06</td>
<td>1.73E-10</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>4.22E-39</td>
<td>0.049</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>8.48E-05</td>
<td>0.15</td>
<td>0.02</td>
</tr>
<tr>
<td>0.25 MB</td>
<td>0.069175</td>
<td>1.9165</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>3.71E-11</td>
<td>0.14</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>0.004005</td>
<td>0.503</td>
<td>0.142</td>
</tr>
<tr>
<td>0.2 MB</td>
<td>0.012106</td>
<td>0.004893</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>2.12E-07</td>
<td>0.000139</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>0.57665</td>
<td>0.477</td>
<td>0.008</td>
</tr>
</tbody>
</table>
The combined (upregulated + downregulated) and downregulated groups seem to show good p-values (< 0.05) in both the significance levels, except for small telomere lengths 0.2 and 0.25 MB (Table 19).

Table 19: The table represents the p-values obtained by two significance levels (DEG2 and DEG3) by the chi-square test using 2x2 contingency table. The test is performed between the telomere and centromere associated genes with all the genes combined, as well as as by dividing the genes into upregulated and downregulated groups. P-values shown in grey are not significant (>=0.05).

<table>
<thead>
<tr>
<th>Telomere Length</th>
<th>DEG2 = ratios &lt; 1/1.2 and &gt; 1.2</th>
<th>DEG3 = ratios &lt; 1/1.3 and &gt; 1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Downregulated</td>
</tr>
<tr>
<td>50 MB</td>
<td>1.91E-06</td>
<td>0.002436953</td>
</tr>
<tr>
<td>40 MB</td>
<td>1.24E-08</td>
<td>3.75E-06</td>
</tr>
<tr>
<td>30 MB</td>
<td>1.96E-09</td>
<td>5.87E-08</td>
</tr>
<tr>
<td>25 MB</td>
<td>6.26E-11</td>
<td>2.23E-08</td>
</tr>
<tr>
<td>20 MB</td>
<td>8.85E-16</td>
<td>2.52E-08</td>
</tr>
<tr>
<td>15 MB</td>
<td>7.24E-14</td>
<td>3.75E-10</td>
</tr>
<tr>
<td>10 MB</td>
<td>3.83E-12</td>
<td>9.36E-10</td>
</tr>
<tr>
<td>9 MB</td>
<td>3.22E-10</td>
<td>3.13E-08</td>
</tr>
<tr>
<td>8 MB</td>
<td>3.41E-09</td>
<td>1.97E-07</td>
</tr>
<tr>
<td>7 MB</td>
<td>5.83E-10</td>
<td>9.99E-08</td>
</tr>
<tr>
<td>6 MB</td>
<td>4.43E-10</td>
<td>1.88E-08</td>
</tr>
<tr>
<td>5 MB</td>
<td>5.26E-11</td>
<td>6.51E-07</td>
</tr>
<tr>
<td>4 MB</td>
<td>1.28E-11</td>
<td>8.10E-07</td>
</tr>
<tr>
<td>3 MB</td>
<td>5.75E-09</td>
<td>2.57E-05</td>
</tr>
<tr>
<td>2 MB</td>
<td>3.86E-07</td>
<td>0.000728299</td>
</tr>
<tr>
<td>1 MB</td>
<td>0.000132</td>
<td>2.91E-05</td>
</tr>
<tr>
<td>0.75 MB</td>
<td>0.000219</td>
<td>0.000208550</td>
</tr>
<tr>
<td>0.5 MB</td>
<td>0.006767</td>
<td>0.00354004</td>
</tr>
<tr>
<td>0.25 MB</td>
<td>0.225291</td>
<td>0.01420092</td>
</tr>
<tr>
<td>0.2 MB</td>
<td>0.380347</td>
<td>0.140937114</td>
</tr>
</tbody>
</table>

The upregulated group opens up a window again in the higher differential expression level (DEG3). This window specifically corresponds to telomere length 2 to 4 MB, which seem to be significant in the chi-square test of homogeneity also. The ideal telomere length obtained (3MB) in the homogeneity test show a better p-value (0.007) than for telomere lengths 2MB (0.024) and 4MB (0.012), that are present in this window.
This test concludes that the most suitable telomere length to define subtelomeric genes for the current dataset as 3MB.

f) Functionally, the behavior of the genes located in the subtelomeric regions of human chromosomes is not different from that of the genes located in other parts of the human chromosomes

Using results of the in silico experiment described above the definition of subtelomeric region may be derived. According to our findings, TPE susceptible subtelomeric regions of human chromosomes are approximately 3Mb in length. To find out whether subtelomeric genes contribute to the cancer phenotype to larger extent than the genes located in the bodies of the chromosomes, we quantified the variation in both sets of genes by distance analysis. The distance (D) between the tissue states was calculated based on Pearson’ correlation (R), as $D = 1 - R$ (Materials and Methods, Chapter 2). This distance might be computed over the entire genome ($n = $ all the gene of chip) or over the particular set of the genes (i.e. subtelomeric genes). Fundamentally, two kinds of the global distances may be computed, one that measure the distance from the particular tissue sample to the metastatic tumor space and from the particular tissue sample to the normal tissue space, DMglobal and DBglobal, respectively.

- DMglobal: Genome-wide distance from Metastatic sample space
- DBglobal: Genome-wide distance from Normal (Benign) sample space
The patterns of the distances between prostatic tumors and normal prostate samples look similar when DB and DM are calculated using subtelomeric and non-telomeric genes as illustrated in Figure 14, panels A) and B) respectively. According to our 3-Mb definition of human telomere, the telomeric genes account for 8.4% of all human genes. The metastatic (red) are distant from the primary tumor (orange) and normal (blue) samples. The primary tumor and normal samples didn’t show much variation as compared to the metastatic tumors as the distribution of the distances between prostatic tumors and normal prostates are very similar when calculated using subtelomeric or non-telomeric gene sets, thus indicating that the changes of expression of subtelomeric genes are rather consequence than the cause of the tumor progression.
**Conclusion**

Here we demonstrate an *in silico* evidence proving telomere positioning effect (TPE) in human tumors. In studied dataset, TPE effect was found to spread over 3Mb of the subtelomeric distance. In microarray experiment profiling human prostatic tumors, a considerable increase in the number of up-regulated and down-regulated genes in telomeres compared to other regions of genome was observed. Importantly, an extent of this increase parallels progression of the prostatic tumors form normal tissue to primary tumor to metastatic carcinoma, with a steep increase at the late stage of cancer progression. The distances between prostatic tumors and normal prostates are very similar when calculated using subtelomeric or non-telomeric gene sets, thus indicating that the changes of expression of subtelomeric genes are rather consequence than the cause of the tumor progression.
Appendix A

KEGG Pathway Painter

High-throughput technologies became common tools to decipher genome-wide changes of gene expression (GE) patterns. Functional analysis of GE patterns is a daunting task as it requires often recourse to the public repositories of biological knowledge. On the other hand, in many cases researcher’s inquiry can be served by a comprehensive glimpse. The KEGG PATHWAY database is a compilation of manually verified maps of biological interactions represented by the complete set of pathways related to signal transduction and other cellular processes. Rapid mapping of the differentially expressed genes to the KEGG pathways may provide an idea about the functional relevance of the gene lists provided by the high-throughput expression experiments. Here web-based graphic tool KEGG Pathway Painter (KPP) is described. KPP paints pathways from the KEGG database using large sets of the candidate genes accompanied by “overexpressed” or “underexpressed” marks, for example, those generated by microarrays or miRNA profilings. KPP will provide fast and comprehensive visualization of the global GE changes by consolidating a list of the color-coded candidate genes into the KEGG pathways. KPP is freely available and can be accessed at http://www.cos.gmu.edu/~gmanyam/kegg/

Introduction

High-throughput technologies became common tools to decipher genome-wide changes of gene expression (GE) patterns or relative protein abundance. Typical output of these
large-scale studies is represented as a list comprised of hundreds of gene candidates with attached quantitative labels. Functional analysis of these gene lists is a daunting task as it requires often recourse to the public repositories of biological knowledge or use of expensive databases of manually curated biological annotation (Ganter and Giroux 2008). On the other hand, in many cases researcher’s inquiry can be successfully served by a comprehensive glimpse.

Functional analysis of markers identified from large-scale datasets can be performed using a wide variety of bioinformatics tools. As microarrays became a common tool to decipher global gene expression, centralized systems like Gene Expression Omnibus (GEO), ArrayExpress were developed to congregate the valuable expression profile data (Barrett and Edgar 2006; Barrett, Troup et al. 2007; Parkinson, Kapushesky et al. 2007). The combined expression profile analysis fusing various microarray datasets (termed as the meta-analysis) is useful in the process of the development of the biomarker panels for various human diseases and specifically for various types of cancers (Rhodes and Chinnaiyan 2005; McShea, Marlatt et al. 2006). Meta-analysis lead to an increase of the complexity of microarray analysis pipeline and further sophistication of subsequent functional analysis is anticipated. Recently, Gene Ontology (GO) and Pathway-based analysis became the most important entry points to the functional analysis of expression data derived from high-throughput platforms (Werner 2008).
KEGG (Kyoto Encyclopedia of Genes and Genomes) is a compendium of databases covering annotated genomes and protein interaction networks for all sequenced organisms. KEGG PATHWAY is a compilation of manually verified pathway maps displaying both the molecular interactions and the biochemical reactions (Kanehisa 2002). The recent version of this database includes a complete set of pathways related to signal transduction and other cellular processes (Kanehisa, Araki et al. 2008). The extensive collection of the pathways at KEGG can be utilized for the rapid graphical evaluation of the functional relevance of the observed changes in GE patterns. This will save the precious time of the expert biologists and bioinformatic specialists.

Pathways assembled into the KEGG database are displayed as semi-static objects that can be manipulated using tools like KGML and KEGG application programmable interface (API) (Kawashima, Katayama et al. 2003; Klukas and Schreiber 2007). KEGG API provides a routine that highlights specified genes within the particular metabolic pathway (http://www.genome.jp/kegg/tool/color_pathway.html). Gene set functional analysis tools like DAVID also returns the KEGG pathways by marking genes of interest (Dennis et al. 2003; Huang et al. 2009). The upregulated and downregulated marks can't be incorporated in such platforms. Similar task may be also executed using G-language Genome Analysis Environment (Arakawa, Kono et al. 2005). Both approaches work on the pathway by pathway basis. Another tool, Pathway express, calculates the pathway-wise impact of differentially expressed genes based on normalized fold change and depicts the pathways with differentially expressed genes (Khatri, Sellamuthu et al. 2005).
However, the fold-change approach and its associated standard t-test statistics usually produce severely over-fitted models. A number of recently developed approaches generate gene rankings dissociated from the fold change estimates (Hsiao, Worrall et al. 2004; Simon 2008). An analysis of these gene lists may benefit from the binary graphical mapping of upregulated and downregulated elements within the complete collection of pathway maps. Resulting graphical pictures may be helpful both as tool for a quick assessment of the functional relevance of a gene list and as a set of the snapshots easily convertible into the illustrative material for presentations or manuscript figures.

With this notion, here we present a web-based tool, KEGG Pathway Painter (KPP). KPP performs a batch painting of relevant pathways using the uploaded lists of up-regulated and down-regulated genes in KEGG. KPP returns a set of images that give a holistic perspective to the functional importance of the change in the GE patterns revealed by a given high-throughput experiment and facilitate the extraction of the biological insights.

**Algorithm and Implementation**

KPP accepts the up-regulated and down-regulated gene lists as two different text files containing the gene identifiers of any sequenced organism. Permitted identifiers include GenBank id, NCBI GENE id, NCBI GI accession, Unigene ID and Uniprot ID.
a) Algorithm

These gene identifiers are converted to KEGG identifiers and all the pathways associated with each of identified genes are extracted. Mapped genes are automatically consolidated within each pathway. The number of the KPP returned pathways could be filtered by either the total number of the painted genes in a given pathway or the ratio of painted genes to the total number of genes in a given pathway. The chosen pathways passing the criteria on filter are color coded according to users’ preferences. Users can browse through these high-resolution pathway images along with gene information and an archive of the painted pathways can also be saved for future reference.

b) Implementation

KPP was implemented using PERL/CGI, and KEGG API was used to communicate with kegg/pathway database. The API allows access to the resources stored in KEGG system in a interactive and user-friendly way (http://www.genome.jp/kegg/soap/). Conversion of the gene identifiers, extraction of the corresponding pathway and their painting is performed by specific API routines. The KPP processes data through direct interface to the KEGG database, and therefore, the KPP painted pathways are always up-to date with reference to KEGG knowledgebase. Genes of interest can be also highlighted with user-specified foreground and background colors for differentiating up and down regulated genes. The URL to the index of resulting output images is sent to the user by e-mail along with the job summary.
Discussion

The motivation for the development of KPP came up from the idea to build a user-friendly, platform-independent and simple tool to visualize the genes in their associated pathways. The simplicity factor of KPP was due to the acceptance of gene identifiers instead of association with a microarray platform. This isolation would enhance its utility study the quantified transcript data from RT-PCR or even to validate varies hypothesis.

Figure 1: Image of the MAPK signaling pathway painted by KPP according to the imported list of genes differentially expressed in the prostatic carcinoma as compared to normal prostate. Red and blue boxes represent up- and down-regulated genes, respectively. The genes in green background represent the specifies specific genes (Homo sapiens, in this case) in the KEGG reference pathway.
surrounding groups of genes that regulate patterns of gene expression in abnormal tissues. The utility is demonstrated using the publicly available prostate carcinoma dataset (GDS1439), from the NCBI GEO database (see Figure 1).

The major fetching point of the tool lies in its tight connection with the KEGG database, as this will allow for the pathway visualization of every sequenced organism. Although, this flexibility is at the cost of bottlenecks caused as a result of the delays during the data transfer. While the pathway painting can be performed for a given set of genes through the KEGG website (http://www.genome.jp/kegg/), the utility of KPP lies in generating the over-all glimpse up and down regulated genes of the dataset as a whole.

In summary, KPP provides fast and comprehensive visualization of the global gene expression changes by consolidating a list of the color-coded candidate genes into the KEGG pathways.
Appendix B

Enriched pathway plots generated by KPP for cancer-specific markers

The genes in light green background were human-specific.

Figure 2: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG p53-signalling pathway.
Figure 3: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG Apoptosis pathway.
Figure 4: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG pathway representing the Ubiquitin mediated proteolysis.
Figure 5: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG pathway representing the regulation of actin cytoskeleton.
Figure 6: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG colorectal cancer pathway.
Figure 7: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG Renal cell carcinoma pathway.
Figure 8: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG Pancreatic cancer pathway.
Figure 9: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG Human Glioma pathway.
Figure 10: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG Prostate cancer pathway.
Figure 11: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG Human Melanoma pathway.

Figure 12: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG Bladder cancer pathway.
Figure 13: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG chronic myeloid leukemia pathway.
Figure 14: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG small cell lung cancer pathway.

Figure 15: The figure illustrates abundant cancer-specific genes (in red background) in the KEGG Non-small cell lung cancer pathway.
Appendix C

Enriched pathway plots generated by KPP for normal-specific (anti-cancer) markers

The genes in light green background were human-specific.

Figure 16: The figure illustrates abundant normal-specific genes (in red background) in the KEGG MAPK signaling pathway.
Figure 17: The figure illustrates abundant normal-specific genes (in red background) in the KEGG ERBB signaling pathway.
Figure 18: The figure illustrates abundant normal-specific genes (in red background) in the KEGG pathway representing cytokine-cytokine reception interaction genes.
Figure 19: The figure illustrates abundant normal-specific genes (in red background) in the KEGG chemokine signaling pathway.
Figure 20: The figure illustrates abundant normal-specific genes (in red background) in the KEGG Wnt signaling pathway.
Figure 21: The figure illustrates abundant normal-specific genes (in red background) in the KEGG TGF-beta signaling pathway.
Figure 22: The figure illustrates abundant normal-specific genes (in red background) in the KEGG TOLL-like signaling pathway.
Figure 23: The figure illustrates abundant normal-specific genes (in red background) in the KEGG T-Cell receptor signaling pathway.
Figure 24: The figure illustrates abundant normal-specific genes (in red background) in the KEGG leukocyte trans-endothelial migration pathway.
Figure 25: The figure illustrates abundant normal-specific genes (in red background) in the KEGG Insulin signalling pathway.

Figure 26: The figure illustrates abundant normal-specific genes (in red background) in the KEGG GnRH signalling pathway.
Appendix D

Results of the PCA analysis of the datasets comprised of two-point paired tumor and normal samples (Table 1)

### Ductal Breast Carcinoma (GSE5764)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.261</td>
<td>0.0823</td>
<td>0.00898</td>
<td>0.00592</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.908</td>
<td>0.0901</td>
<td>0.00107</td>
<td>0.00047</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.908</td>
<td>0.9985</td>
<td>0.99953</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

**Component Loadings:**

<table>
<thead>
<tr>
<th></th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.4792611</td>
<td>0.0347449</td>
<td>-0.3525736</td>
<td>0.8029903</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.4315327</td>
<td>-0.0642264</td>
<td>-0.7019314</td>
<td>-0.5629803</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.5729407</td>
<td>0.6708277</td>
<td>0.4351537</td>
<td>-0.1799179</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.5057938</td>
<td>-0.7380095</td>
<td>0.4400289</td>
<td>-0.0767414</td>
</tr>
</tbody>
</table>

### Lobular Breast Carcinoma (GSE5764)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 “size”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.374</td>
<td>0.135</td>
<td>0.01778</td>
<td>0.00960</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.882</td>
<td>0.116</td>
<td>0.00199</td>
<td>0.00058</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.882</td>
<td>0.997</td>
<td>0.99942</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

**Component Loadings:**

<table>
<thead>
<tr>
<th></th>
<th>PC1 “size”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.4336790</td>
<td>0.0032826</td>
<td>-0.5559545</td>
<td>0.7091025</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.4196183</td>
<td>-0.0321633</td>
<td>-0.5714445</td>
<td>-0.7045120</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.5651979</td>
<td>0.7156881</td>
<td>0.4093347</td>
<td>-0.0280528</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.5624894</td>
<td>-0.6976713</td>
<td>0.4436338</td>
<td>0.0070378</td>
</tr>
</tbody>
</table>
**Pulmonary adenocarcinoma (GSE2514)**

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.0901</td>
<td>0.0358</td>
<td>0.00143</td>
<td>0.000988</td>
</tr>
<tr>
<td>Proportion of variance explained by component</td>
<td>0.8635</td>
<td>0.1361</td>
<td>0.00022</td>
<td>0.000100</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.8635</td>
<td>0.9997</td>
<td>0.99990</td>
<td>1.000000</td>
</tr>
<tr>
<td>Component Loadings:</td>
<td>PC1 “size”</td>
<td>PC2 “Normal/Cancer difference”</td>
<td>PC3 “Degree of autonomy”</td>
<td>PC4 “Noise”</td>
</tr>
<tr>
<td>DCGlobal</td>
<td>-0.4223058</td>
<td>-0.2673878</td>
<td>0.3579391</td>
<td>0.7886958</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.3479712</td>
<td>0.3425144</td>
<td>0.7661334</td>
<td>-0.4178990</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.6195546</td>
<td>-0.6111058</td>
<td>-0.2421958</td>
<td>-0.4290024</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.5627841</td>
<td>0.6616172</td>
<td>-0.4756684</td>
<td>0.1388390</td>
</tr>
</tbody>
</table>

**Pulmonary adenocarcinoma (GSE7670)**

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviation</td>
<td>0.193</td>
<td>0.0576</td>
<td>0.00663</td>
<td>0.00149</td>
</tr>
<tr>
<td>Proportion of variance explained by component</td>
<td>0.917</td>
<td>0.0815</td>
<td>0.00108</td>
<td>0.00005</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.917</td>
<td>0.9989</td>
<td>0.99995</td>
<td>1.00000</td>
</tr>
<tr>
<td>Component Loadings:</td>
<td>PC1 “size”</td>
<td>PC2 “Normal/Cancer difference”</td>
<td>PC3 “Degree of autonomy”</td>
<td>PC4 “Noise”</td>
</tr>
<tr>
<td>DCGlobal</td>
<td>-0.4241949</td>
<td>0.2328628</td>
<td>-0.4694148</td>
<td>0.7385685</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.3486490</td>
<td>-0.3041166</td>
<td>-0.6972629</td>
<td>-0.5475230</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.6355490</td>
<td>0.6064523</td>
<td>0.3269082</td>
<td>-0.3484596</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.5427522</td>
<td>-0.6967809</td>
<td>0.4319783</td>
<td>0.1825134</td>
</tr>
</tbody>
</table>
## Esophageal Cancer (GSE1420)

<table>
<thead>
<tr>
<th>Component</th>
<th>Relative importance:</th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.226</td>
<td>0.085</td>
<td>0.00990</td>
<td>0.00172</td>
<td></td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.875</td>
<td>0.124</td>
<td>0.00168</td>
<td>0.00005</td>
<td></td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.875</td>
<td>0.998</td>
<td>0.99995</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>Component Loadings:</td>
<td>PC1 “size”</td>
<td>PC2 “Normal/Cancer difference”</td>
<td>PC3 “Degree of autonomy”</td>
<td>PC4 “Noise”</td>
<td></td>
</tr>
<tr>
<td>DCGlobal</td>
<td>-0.3634003</td>
<td>0.2365433</td>
<td>0.5838462</td>
<td>0.6863753</td>
<td></td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.3856582</td>
<td>-0.2269184</td>
<td>0.6150850</td>
<td>-0.6491890</td>
<td></td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.5892720</td>
<td>0.6806972</td>
<td>-0.3663376</td>
<td>-0.2349607</td>
<td></td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.6098905</td>
<td>-0.6551386</td>
<td>-0.3828719</td>
<td>0.2285521</td>
<td></td>
</tr>
</tbody>
</table>

## Renal cell carcinoma (GSE6344)

<table>
<thead>
<tr>
<th>Component</th>
<th>Relative importance:</th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.133</td>
<td>0.0715</td>
<td>0.00230</td>
<td>0.000472</td>
<td></td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.777</td>
<td>0.2231</td>
<td>0.00023</td>
<td>0.000010</td>
<td></td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.777</td>
<td>0.998</td>
<td>0.99999</td>
<td>1.00000</td>
<td></td>
</tr>
<tr>
<td>Component Loadings:</td>
<td>PC1 “Attractor”</td>
<td>PC2 “Normal/Cancer difference”</td>
<td>PC3 “Degree of autonomy”</td>
<td>PC4 “Noise”</td>
<td></td>
</tr>
<tr>
<td>DCGlobal</td>
<td>-0.4536096</td>
<td>0.3104922</td>
<td>0.6132022</td>
<td>0.5672884</td>
<td></td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.3913748</td>
<td>-0.3780481</td>
<td>0.5147935</td>
<td>-0.6624901</td>
<td></td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.5869027</td>
<td>0.5997387</td>
<td>-0.4316633</td>
<td>-0.3309463</td>
<td></td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.5446185</td>
<td>-0.6332359</td>
<td>-0.4154967</td>
<td>0.3602296</td>
<td></td>
</tr>
</tbody>
</table>
### Renal cell carcinoma (GSE781)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.138</td>
<td>0.073</td>
<td>0.00366</td>
<td>0.000972</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.781</td>
<td>0.219</td>
<td>0.00055</td>
<td>0.000040</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.781</td>
<td>0.999</td>
<td>0.99996</td>
<td>1.000000</td>
</tr>
<tr>
<td>Component Loadings:</td>
<td>PC1 “Attractor”</td>
<td>PC2 “Normal/Cancer difference”</td>
<td>PC3 “Degree of autonomy”</td>
<td>PC4 “Noise”</td>
</tr>
<tr>
<td>DCGlobal</td>
<td>-0.4015367</td>
<td>0.1770823</td>
<td>-0.5931930</td>
<td>-0.6749313</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.3159439</td>
<td>-0.2905877</td>
<td>-0.6200694</td>
<td>0.6566979</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.6328689</td>
<td>0.6538785</td>
<td>0.3089273</td>
<td>0.2765569</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.5817428</td>
<td>-0.6757541</td>
<td>0.4101218</td>
<td>-0.1916557</td>
</tr>
</tbody>
</table>

### Head and neck squamous cell carcinoma (GDS2520)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.190</td>
<td>0.0407</td>
<td>0.00979</td>
<td>0.00126</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.954</td>
<td>0.0436</td>
<td>0.00252</td>
<td>0.00004</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.954</td>
<td>0.9974</td>
<td>0.99996</td>
<td>1.00000</td>
</tr>
<tr>
<td>Component Loadings:</td>
<td>PC1 “Attractor”</td>
<td>PC2 “Normal/Cancer difference”</td>
<td>PC3 “degree of autonomy”</td>
<td>PC4 “noise”</td>
</tr>
<tr>
<td>DCGlobal</td>
<td>-0.3839816</td>
<td>0.1435169</td>
<td>0.5972208</td>
<td>0.6894116</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.3578240</td>
<td>-0.1809713</td>
<td>0.6062387</td>
<td>-0.6867940</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.6171703</td>
<td>0.6782271</td>
<td>-0.3589338</td>
<td>-0.1739981</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.5861919</td>
<td>-0.6976105</td>
<td>-0.3833646</td>
<td>0.1508320</td>
</tr>
</tbody>
</table>
### Papillary thyroid carcinoma (GDS1665)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.0968</td>
<td>0.0398</td>
<td>0.00375</td>
<td>0.00120</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.8542</td>
<td>0.1444</td>
<td>0.00128</td>
<td>0.00013</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.8542</td>
<td>0.9986</td>
<td>0.99987</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component Loadings:</th>
<th>PC1 “Attractor”</th>
<th>PC2 “Normal/Cancer difference”</th>
<th>PC3 “Degree of autonomy”</th>
<th>PC4 “Noise”</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.4055784</td>
<td>0.1936700</td>
<td>-0.4882024</td>
<td>0.7481019</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.3365074</td>
<td>-0.2185903</td>
<td>-0.7043803</td>
<td>-0.5855164</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.6383106</td>
<td>0.6270132</td>
<td>0.3455199</td>
<td>-0.2828959</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.5610958</td>
<td>-0.7221944</td>
<td>0.3822601</td>
<td>0.1322270</td>
</tr>
</tbody>
</table>
Appendix E

Results of the PCA analysis of population datasets comprised of two-point tumor and normal samples (Table 2)

<table>
<thead>
<tr>
<th>Invasive Breast (Epithelial) Carcinoma (GSE10797)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relative importance:</strong></td>
</tr>
<tr>
<td>PC1 (Attractor)</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
</tr>
<tr>
<td><strong>Component Loadings:</strong></td>
</tr>
<tr>
<td>PC1 (Attractor)</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>DCGlobal</td>
</tr>
<tr>
<td>DNGlobal</td>
</tr>
<tr>
<td>DCSpecific</td>
</tr>
<tr>
<td>DNSpecific</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cervical Cancer (GSE6791)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relative importance:</strong></td>
</tr>
<tr>
<td>PC1 (Attractor)</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
</tr>
<tr>
<td><strong>Component Loadings:</strong></td>
</tr>
<tr>
<td>PC1 (Attractor)</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>DCGlobal</td>
</tr>
<tr>
<td>DNGlobal</td>
</tr>
<tr>
<td>DCSpecific</td>
</tr>
<tr>
<td>DNSpecific</td>
</tr>
</tbody>
</table>
### Head and Neck Cancer (GSE6791)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.169</td>
<td>0.0308</td>
<td>0.00463</td>
<td>0.0011</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.967</td>
<td>0.0319</td>
<td>0.00072</td>
<td>0.00004</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.967</td>
<td>0.9992</td>
<td>0.99996</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component Loadings:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.45602</td>
<td>-0.30736</td>
<td>0.542741</td>
<td>0.634831</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.47091</td>
<td>0.312987</td>
<td>0.525327</td>
<td>-0.63586</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.47903</td>
<td>-0.70074</td>
<td>-0.41236</td>
<td>-0.33083</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.58379</td>
<td>0.562613</td>
<td>-0.50934</td>
<td>0.288488</td>
</tr>
</tbody>
</table>

### Mesothelioma (GSE12345)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.244</td>
<td>0.0846</td>
<td>0.00892</td>
<td>0.00362</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.892</td>
<td>0.1071</td>
<td>0.00119</td>
<td>0.0002</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.892</td>
<td>0.9986</td>
<td>0.9998</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component Loadings:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.34642</td>
<td>0.145146</td>
<td>-0.63865</td>
<td>0.671609</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.34828</td>
<td>-0.08457</td>
<td>-0.59061</td>
<td>-0.72299</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.51471</td>
<td>0.780904</td>
<td>0.334231</td>
<td>-0.11643</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.70269</td>
<td>-0.60164</td>
<td>0.362763</td>
<td>0.112533</td>
</tr>
</tbody>
</table>
### Nasopharyngeal Carcinoma (GSE12452)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.184</td>
<td>0.0488</td>
<td>0.00475</td>
<td>0.00121</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.934</td>
<td>0.0655</td>
<td>0.00062</td>
<td>0.00004</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.934</td>
<td>0.9993</td>
<td>0.99996</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component Loadings:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.33784</td>
<td>0.289965</td>
<td>-0.70708</td>
<td>0.549388</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.4066</td>
<td>-0.1413</td>
<td>-0.46484</td>
<td>-0.77372</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.47991</td>
<td>0.752777</td>
<td>0.42756</td>
<td>-0.14215</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.70016</td>
<td>-0.57385</td>
<td>0.318053</td>
<td>0.281654</td>
</tr>
</tbody>
</table>

### Oral Squamous Cell Carcinoma (GSE3524)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.171</td>
<td>0.0522</td>
<td>0.00434</td>
<td>0.00237</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.914</td>
<td>0.0857</td>
<td>0.00059</td>
<td>0.00018</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.914</td>
<td>0.9992</td>
<td>0.99982</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component Loadings:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.36222</td>
<td>-0.29688</td>
<td>0.725488</td>
<td>-0.50431</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.40249</td>
<td>0.162528</td>
<td>0.419678</td>
<td>0.797156</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.43109</td>
<td>-0.78102</td>
<td>-0.42128</td>
<td>0.163367</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.72177</td>
<td>0.524838</td>
<td>-0.3465</td>
<td>-0.28901</td>
</tr>
</tbody>
</table>
### Renal Cell carcinoma (GSE14762)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.271</td>
<td>0.0922</td>
<td>0.02352</td>
<td>0.00252</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.891</td>
<td>0.1027</td>
<td>0.00669</td>
<td>0.00008</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.891</td>
<td>0.9932</td>
<td>0.99992</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component Loadings:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.48269</td>
<td>0.14762</td>
<td>-0.27471</td>
<td>-0.81838</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.33537</td>
<td>-0.40744</td>
<td>-0.75998</td>
<td>0.379411</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.64799</td>
<td>0.602155</td>
<td>0.178277</td>
<td>0.430968</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.4844</td>
<td>-0.67055</td>
<td>0.561411</td>
<td>-0.02369</td>
</tr>
</tbody>
</table>

### Papillary thyroid carcinoma (GSE3678)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.17</td>
<td>0.0807</td>
<td>0.00574</td>
<td>0.00121</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.814</td>
<td>0.1847</td>
<td>0.00094</td>
<td>0.00004</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.814</td>
<td>0.999</td>
<td>0.99996</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component Loadings:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.3045</td>
<td>0.081804</td>
<td>0.496542</td>
<td>-0.80872</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.2992</td>
<td>-0.09796</td>
<td>0.759593</td>
<td>0.569125</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.64121</td>
<td>0.702814</td>
<td>-0.27129</td>
<td>0.145953</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.63766</td>
<td>-0.69983</td>
<td>-0.32073</td>
<td>-0.02762</td>
</tr>
</tbody>
</table>
Appendix F

Results of the PCA analysis of the datasets comprised of multi-stage tumor and normal samples collected across population (Table 3)

### Ovarian (fallopian tube) carcinoma (GSE10971)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.271</td>
<td>0.0956</td>
<td>0.00778</td>
<td>0.00111</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.889</td>
<td>0.1107</td>
<td>0.00073</td>
<td>0.00001</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.889</td>
<td>0.9992</td>
<td>0.99999</td>
<td>1</td>
</tr>
</tbody>
</table>

#### Component Loadings:

<table>
<thead>
<tr>
<th></th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.4319740</td>
<td>0.1653989</td>
<td>0.5648672</td>
<td>-0.6833496</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.3371510</td>
<td>-0.2908749</td>
<td>0.6148158</td>
<td>0.6509397</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.6792346</td>
<td>0.5755423</td>
<td>-0.3745092</td>
<td>0.2591029</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.4882294</td>
<td>-0.7461810</td>
<td>-0.4033230</td>
<td>-0.2053693</td>
</tr>
</tbody>
</table>

### Bladder carcinoma (GSE3167)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.228</td>
<td>0.0694</td>
<td>0.00589</td>
<td>0.00160</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.914</td>
<td>0.0850</td>
<td>0.00061</td>
<td>0.00005</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.914</td>
<td>0.9993</td>
<td>0.99995</td>
<td>1</td>
</tr>
</tbody>
</table>

#### Component Loadings:

<table>
<thead>
<tr>
<th></th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.4219051</td>
<td>0.4471904</td>
<td>-0.6546958</td>
<td>-0.4397615</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.5576904</td>
<td>-0.3487430</td>
<td>-0.3327277</td>
<td>0.6757603</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.3752885</td>
<td>0.7052464</td>
<td>0.5163564</td>
<td>0.3084836</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.6083880</td>
<td>-0.4254722</td>
<td>0.4405017</td>
<td>-0.5047729</td>
</tr>
</tbody>
</table>
### Esophagus carcinoma (GSE1420)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.246</td>
<td>0.101</td>
<td>0.0119</td>
<td>0.00201</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.854</td>
<td>0.144</td>
<td>0.0020</td>
<td>0.00006</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.854</td>
<td>0.998</td>
<td>0.9999</td>
<td>1</td>
</tr>
</tbody>
</table>

### Component Loadings:

<table>
<thead>
<tr>
<th></th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC Global</td>
<td>-0.3604909</td>
<td>0.2376977</td>
<td>0.5853858</td>
<td>0.6861993</td>
</tr>
<tr>
<td>DN Global</td>
<td>-0.3849471</td>
<td>-0.2241357</td>
<td>0.6157891</td>
<td>-0.6499097</td>
</tr>
<tr>
<td>DC Specific</td>
<td>-0.5851186</td>
<td>0.6857592</td>
<td>-0.3637739</td>
<td>-0.2346041</td>
</tr>
<tr>
<td>DN Specific</td>
<td>-0.6160343</td>
<td>-0.6503824</td>
<td>-0.3818320</td>
<td>0.2273955</td>
</tr>
</tbody>
</table>

### Hepato Cellular carcinoma (GSE6764)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.134</td>
<td>0.0425</td>
<td>0.00439</td>
<td>0.000683</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.908</td>
<td>0.0911</td>
<td>0.00097</td>
<td>0.000020</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.908</td>
<td>0.9990</td>
<td>0.99998</td>
<td>1</td>
</tr>
</tbody>
</table>

### Component Loadings:

<table>
<thead>
<tr>
<th></th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC Global</td>
<td>-0.4871506</td>
<td>0.1924781</td>
<td>-0.5238308</td>
<td>0.6717423</td>
</tr>
<tr>
<td>DN Global</td>
<td>-0.3563782</td>
<td>-0.3680381</td>
<td>-0.5963258</td>
<td>-0.6180114</td>
</tr>
<tr>
<td>DC Specific</td>
<td>-0.6539801</td>
<td>0.5581170</td>
<td>0.3891397</td>
<td>-0.3307351</td>
</tr>
<tr>
<td>DN Specific</td>
<td>-0.4560580</td>
<td>-0.7183352</td>
<td>0.4675117</td>
<td>0.2396632</td>
</tr>
</tbody>
</table>
### Pancreas carcinoma (Logsdon et al.)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.213</td>
<td>0.106</td>
<td>0.0098</td>
<td>0.00296</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.801</td>
<td>0.198</td>
<td>0.0017</td>
<td>0.00016</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.801</td>
<td>0.998</td>
<td>0.9998</td>
<td>1</td>
</tr>
</tbody>
</table>

**Component Loadings:**

<table>
<thead>
<tr>
<th></th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.3106436</td>
<td>0.3413867</td>
<td>-0.5844092</td>
<td>-0.6673991</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.4308919</td>
<td>-0.1495312</td>
<td>-0.6043367</td>
<td>0.6532609</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.4633882</td>
<td>0.7489053</td>
<td>0.4070460</td>
<td>0.2423340</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.7092984</td>
<td>-0.5479383</td>
<td>0.3571507</td>
<td>-0.2628743</td>
</tr>
</tbody>
</table>

### Prostate carcinoma (GSE6919)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.094</td>
<td>0.0218</td>
<td>0.00202</td>
<td>0.000289</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.948</td>
<td>0.0512</td>
<td>0.00044</td>
<td>0.000010</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.948</td>
<td>0.9996</td>
<td>0.99999</td>
<td>1</td>
</tr>
</tbody>
</table>

**Component Loadings:**

<table>
<thead>
<tr>
<th></th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.4868723</td>
<td>0.2591513</td>
<td>-0.5374330</td>
<td>0.6379355</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.3937599</td>
<td>-0.3927989</td>
<td>-0.5607571</td>
<td>-0.6133626</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.6342255</td>
<td>0.5425107</td>
<td>0.4349970</td>
<td>-0.3379612</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.4535047</td>
<td>-0.6958676</td>
<td>0.4555160</td>
<td>0.3203231</td>
</tr>
</tbody>
</table>
# Prostate carcinoma (GSE3325)

<table>
<thead>
<tr>
<th>Relative importance:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.185</td>
<td>0.0962</td>
<td>0.00555</td>
<td>0.00142</td>
</tr>
<tr>
<td>Proportion of Variance explained by component</td>
<td>0.787</td>
<td>0.2127</td>
<td>0.00071</td>
<td>0.00005</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.787</td>
<td>0.9992</td>
<td>0.99995</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component Loadings:</th>
<th>PC1 (Attractor)</th>
<th>PC2 (Normal/Cancer difference)</th>
<th>PC3 (Degree of autonomy)</th>
<th>PC4 (Noise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGlobal</td>
<td>-0.4656498</td>
<td>0.1456634</td>
<td>-0.6308057</td>
<td>0.6033544</td>
</tr>
<tr>
<td>DNGlobal</td>
<td>-0.2653653</td>
<td>-0.3860657</td>
<td>-0.5525927</td>
<td>-0.6893300</td>
</tr>
<tr>
<td>DCSpecific</td>
<td>-0.7344495</td>
<td>0.4844237</td>
<td>0.3763689</td>
<td>-0.2902828</td>
</tr>
<tr>
<td>DNSpecific</td>
<td>-0.4163357</td>
<td>-0.7714072</td>
<td>0.3937917</td>
<td>0.2766288</td>
</tr>
</tbody>
</table>
Appendix G

Protein-coding genes extracted using the EST abundance analysis of the Human Unigene data as potential tumor biomarkers (Table 4).

<table>
<thead>
<tr>
<th>S.No</th>
<th>Unigene_ID</th>
<th>Gene Symbol</th>
<th>NCBI Gene_ID</th>
<th>Genomic loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hs.655285</td>
<td>ABCF1</td>
<td>23</td>
<td>6p21.33</td>
</tr>
<tr>
<td>2</td>
<td>Hs.387567</td>
<td>ACLY</td>
<td>47</td>
<td>17q12-q21</td>
</tr>
<tr>
<td>3</td>
<td>Hs.514581</td>
<td>ACTG1</td>
<td>71</td>
<td>17q25</td>
</tr>
<tr>
<td>4</td>
<td>Hs.467125</td>
<td>AP2A1</td>
<td>160</td>
<td>19q13.33</td>
</tr>
<tr>
<td>5</td>
<td>Hs.388004</td>
<td>AHCY</td>
<td>191</td>
<td>20cen-q13.1</td>
</tr>
<tr>
<td>6</td>
<td>Hs.525622</td>
<td>AKT1</td>
<td>207</td>
<td>14q32.32</td>
</tr>
<tr>
<td>7</td>
<td>Hs.531682</td>
<td>ALDH3A1</td>
<td>218</td>
<td>17p11.2</td>
</tr>
<tr>
<td>8</td>
<td>Hs.511605</td>
<td>ANXA2</td>
<td>302</td>
<td>15q21-q22</td>
</tr>
<tr>
<td>9</td>
<td>Hs.517989</td>
<td>APEH</td>
<td>327</td>
<td>3p21.31</td>
</tr>
<tr>
<td>10</td>
<td>Hs.514527</td>
<td>BIRC5</td>
<td>332</td>
<td>17q25</td>
</tr>
<tr>
<td>11</td>
<td>Hs.502659</td>
<td>RHOC</td>
<td>389</td>
<td>1p13.1</td>
</tr>
<tr>
<td>12</td>
<td>Hs.465985</td>
<td>ASNA1</td>
<td>439</td>
<td>19q13.3</td>
</tr>
<tr>
<td>13</td>
<td>Hs.707979</td>
<td>ATP1B2</td>
<td>482</td>
<td>17p13.1</td>
</tr>
<tr>
<td>14</td>
<td>Hs.406510</td>
<td>ATP5B</td>
<td>506</td>
<td>12q13.13</td>
</tr>
<tr>
<td>15</td>
<td>Hs.514870</td>
<td>ATP5F1</td>
<td>515</td>
<td>1p13.2</td>
</tr>
<tr>
<td>16</td>
<td>Hs.516966</td>
<td>BCL2L1</td>
<td>598</td>
<td>20q11.21</td>
</tr>
<tr>
<td>17</td>
<td>Hs.106880</td>
<td>BYSL</td>
<td>705</td>
<td>6p21.1</td>
</tr>
<tr>
<td>18</td>
<td>Hs.555866</td>
<td>C1QBP</td>
<td>708</td>
<td>17p13.3</td>
</tr>
<tr>
<td>19</td>
<td>Hs.377010</td>
<td>CAD</td>
<td>790</td>
<td>2p22-p21</td>
</tr>
<tr>
<td>20</td>
<td>Hs.515162</td>
<td>CALR</td>
<td>811</td>
<td>19p13.3-p13.2</td>
</tr>
<tr>
<td>21</td>
<td>Hs.516155</td>
<td>CAPG</td>
<td>822</td>
<td>2p11.2</td>
</tr>
<tr>
<td>22</td>
<td>Hs.58974</td>
<td>CCNA2</td>
<td>890</td>
<td>4q25-q31</td>
</tr>
<tr>
<td>23</td>
<td>Hs.23960</td>
<td>CCNB1</td>
<td>891</td>
<td>5q12</td>
</tr>
<tr>
<td>24</td>
<td>Hs.244723</td>
<td>CCNE1</td>
<td>898</td>
<td>19q12</td>
</tr>
<tr>
<td>25</td>
<td>Hs.82916</td>
<td>CCT6A</td>
<td>908</td>
<td>7p11.2</td>
</tr>
<tr>
<td>26</td>
<td>Hs.501497</td>
<td>CD70</td>
<td>970</td>
<td>19p13</td>
</tr>
<tr>
<td>27</td>
<td>Hs.405958</td>
<td>CDC6</td>
<td>990</td>
<td>17q21.3</td>
</tr>
<tr>
<td>28</td>
<td>Hs.524947</td>
<td>CDC20</td>
<td>991</td>
<td>1p34.1</td>
</tr>
<tr>
<td>29</td>
<td>Hs.437705</td>
<td>CDC25A</td>
<td>993</td>
<td>3p21</td>
</tr>
<tr>
<td>30</td>
<td>Hs.153752</td>
<td>CDC25B</td>
<td>994</td>
<td>20p13</td>
</tr>
<tr>
<td>31</td>
<td>Hs.95577</td>
<td>CDK4</td>
<td>1019</td>
<td>12q14</td>
</tr>
<tr>
<td>32</td>
<td>Hs.512599</td>
<td>CDKN2A</td>
<td>1029</td>
<td>9p21</td>
</tr>
<tr>
<td>33</td>
<td>Hs.1594</td>
<td>CENPA</td>
<td>1058</td>
<td>2p24-p21</td>
</tr>
<tr>
<td>34</td>
<td>Hs.516855</td>
<td>CENPB</td>
<td>1059</td>
<td>20p13</td>
</tr>
<tr>
<td>35</td>
<td>Hs.75573</td>
<td>CENPE</td>
<td>1062</td>
<td>4q24-q25</td>
</tr>
<tr>
<td>36</td>
<td>Hs.170622</td>
<td>CFL1</td>
<td>1072</td>
<td>11q13</td>
</tr>
<tr>
<td></td>
<td>Gene</td>
<td>Chromosome</td>
<td>Location</td>
<td>Notes</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>------------</td>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td>37</td>
<td>Hs.706874 CTSC</td>
<td>11q14.1-q14.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Hs.374378 CKS1B</td>
<td>1q21.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Hs.83758 CKS2</td>
<td>9q22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Hs.563509 AP1S1</td>
<td>7q22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Hs.414565 CLIC1</td>
<td>6p22.1-p21.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Hs.522114 CLTA</td>
<td>9p13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Hs.132370 CSTF2</td>
<td>Xq22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Hs.289271 CYC1</td>
<td>8q24.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Hs.706840 DCN</td>
<td>12q21.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Hs.290758 DDB1</td>
<td>11q12-q13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>Hs.484782 DFFA</td>
<td>1p36.3-p36.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>Hs.4747 DKC1</td>
<td>Xq28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>Hs.632398 DPH2</td>
<td>1p34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Hs.591664 DUSP7</td>
<td>3p21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>Hs.654393 E2F1</td>
<td>20q11.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>Hs.703174 E2F3</td>
<td>6p22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>Hs.518299 ECT2</td>
<td>3q26.1-q26.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Hs.586423 EEF1A1</td>
<td>6q14.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>Hs.520703 EEF1A1</td>
<td>6q14.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>Hs.333388 EEF1D</td>
<td>8q24.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>Hs.144835 EEF1G</td>
<td>11q12.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>Hs.129673 EIF4A1</td>
<td>17p13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>Hs.707977 EIF4A2</td>
<td>3q28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>Hs.433750 EIF4G1</td>
<td>3q27-qter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>Hs.534314 EIF5A</td>
<td>17p13-p12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>Hs.522823 EMD</td>
<td>Xq28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>Hs.517145 ENO1</td>
<td>1p36.3-p36.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>Hs.2913 EPHB3</td>
<td>3q21-qter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>Hs.299002 FBL</td>
<td>19q13.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>Hs.110849 ESRRB</td>
<td>11q13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>Hs.434059 ETV4</td>
<td>17q21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>Hs.444082 EZH2</td>
<td>7q35-q36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>Hs.302003 FANCE</td>
<td>6p22-p21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>Hs.335918 FDPS</td>
<td>1q22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>Hs.409065 FEN1</td>
<td>11q12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>Hs.524183 FKBP4</td>
<td>12p13.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>Hs.239 FOXM1</td>
<td>12p13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>Hs.524910 FTH1</td>
<td>11q13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>Hs.461047 G6PD</td>
<td>Xq28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>Hs.544577 GAPDH</td>
<td>12p13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>Hs.479728 GAPDH</td>
<td>12p13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>Hs.708288 GJA1</td>
<td>6q21-22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>Hs.487341 GNA12</td>
<td>7p22.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>Hs.185172 GNB2</td>
<td>7q21.3-q22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>Hs.523718 SFN</td>
<td>1p36.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>Hs.594634 GRINA</td>
<td>8q24.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hs.</td>
<td>Gene</td>
<td>Chromosome</td>
<td>Position</td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>83</td>
<td>466828</td>
<td>GSK3A</td>
<td>19q13.2</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>445052</td>
<td>MSH6</td>
<td>2p16</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>75782</td>
<td>GTF3C2</td>
<td>2p23.3</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>477879</td>
<td>H2AFX</td>
<td>11q23.2-q23.3</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>171280</td>
<td>HSD17B10</td>
<td>Xp11.2</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>88556</td>
<td>HDAC1</td>
<td>1p34</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>707995</td>
<td>HMGB1</td>
<td>13q12</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>181183</td>
<td>HMGN2</td>
<td>1p36.1</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>518805</td>
<td>HMGA1</td>
<td>6p21</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>703764</td>
<td>HMGA1</td>
<td>6p21</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>569017</td>
<td>SLC29A2</td>
<td>11q13</td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>436181</td>
<td>HOXB7</td>
<td>17q21.3</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>463350</td>
<td>HOXB9</td>
<td>17q21.3</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>580427</td>
<td>HPRT1</td>
<td>2p25.1</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>20521</td>
<td>PRMT1</td>
<td>19q13.3</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>707984</td>
<td>IDUA</td>
<td>4p16.3</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>654400</td>
<td>IMPDH2</td>
<td>3p21.2</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>522819</td>
<td>IRAK1</td>
<td>Xq28</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>654848</td>
<td>EIF6</td>
<td>20q12</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>2722</td>
<td>ITPKA</td>
<td>15q14-q21</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>3100</td>
<td>KARS</td>
<td>16q23-q24</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>436912</td>
<td>KIFC1</td>
<td>6p21.3</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>532793</td>
<td>KPNB1</td>
<td>17q21.32</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>533782</td>
<td>KRT8</td>
<td>12q13</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>406013</td>
<td>KRT18</td>
<td>12q13</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>449909</td>
<td>RPSA</td>
<td>3p22.2</td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>446149</td>
<td>LDHB</td>
<td>12p12.2-p12.1</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>514535</td>
<td>LGALS3BP</td>
<td>17q25</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>89497</td>
<td>LMNB1</td>
<td>5q23.3-q31.1</td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>706751</td>
<td>LTBP1</td>
<td>2p22-p21</td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>521903</td>
<td>LY6E</td>
<td>8q24.3</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>546264</td>
<td>NBR1</td>
<td>17q21.31</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>417816</td>
<td>MAGEA3</td>
<td>Xq28</td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>441113</td>
<td>MAGEA6</td>
<td>Xq28</td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>169246</td>
<td>MAGEA12</td>
<td>Xq28</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>23650</td>
<td>MAZ</td>
<td>16p11.2</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>477481</td>
<td>MCM2</td>
<td>3q21</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>179565</td>
<td>MCM3</td>
<td>6p12</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>460184</td>
<td>MCM4</td>
<td>8q11.2</td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>438720</td>
<td>MCM7</td>
<td>7q21.3-q22.1</td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>567303</td>
<td>MDM2</td>
<td>12q14.3-q15</td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>423348</td>
<td>MEN1</td>
<td>11q13</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>80976</td>
<td>MKI67</td>
<td>10q25-pter</td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>391464</td>
<td>ABCC1</td>
<td>16p13.1</td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>179718</td>
<td>MYBL2</td>
<td>20q13.1</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>524599</td>
<td>NAP1L1</td>
<td>12q21.2</td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene Symbol</td>
<td>Gene Name</td>
<td>Chromosome</td>
<td>Band</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>Hs.81469</td>
<td>NUBP1</td>
<td>4682</td>
<td>16p13.13</td>
<td></td>
</tr>
<tr>
<td>Hs.277677</td>
<td>NDUFA10</td>
<td>4705</td>
<td>2q37.3</td>
<td></td>
</tr>
<tr>
<td>Hs.148340</td>
<td>RPL10A</td>
<td>4736</td>
<td>6p21.13-p21.2</td>
<td></td>
</tr>
<tr>
<td>Hs.675285</td>
<td>NFKBIL2</td>
<td>4796</td>
<td>8q24.3</td>
<td></td>
</tr>
<tr>
<td>Hs.557550</td>
<td>NPM1</td>
<td>4869</td>
<td>5q35</td>
<td></td>
</tr>
<tr>
<td>Hs.473583</td>
<td>YBX1</td>
<td>4904</td>
<td>1p34</td>
<td></td>
</tr>
<tr>
<td>Hs.446427</td>
<td>OAZ1</td>
<td>4946</td>
<td>19p13.3</td>
<td></td>
</tr>
<tr>
<td>Hs.708130</td>
<td>OGN</td>
<td>4969</td>
<td>9q22</td>
<td></td>
</tr>
<tr>
<td>Hs.17908</td>
<td>ORC1L</td>
<td>4998</td>
<td>1p32</td>
<td></td>
</tr>
<tr>
<td>Hs.524498</td>
<td>PA2G4</td>
<td>5036</td>
<td>12q13.2</td>
<td></td>
</tr>
<tr>
<td>Hs.180909</td>
<td>PRDX1</td>
<td>5052</td>
<td>1p34.1</td>
<td></td>
</tr>
<tr>
<td>Hs.546271</td>
<td>PCB2</td>
<td>5094</td>
<td>12q13.12-q13.13</td>
<td></td>
</tr>
<tr>
<td>Hs.255093</td>
<td>PFKL</td>
<td>5111</td>
<td>21q22.3</td>
<td></td>
</tr>
<tr>
<td>Hs.494691</td>
<td>PFN1</td>
<td>5216</td>
<td>17p13.3</td>
<td></td>
</tr>
<tr>
<td>Hs.290404</td>
<td>SLC25A3</td>
<td>5250</td>
<td>12q23</td>
<td></td>
</tr>
<tr>
<td>Hs.371344</td>
<td>PIK3R2</td>
<td>5296</td>
<td>19q13.2-q13.4</td>
<td></td>
</tr>
<tr>
<td>Hs.53770</td>
<td>PKM2</td>
<td>5315</td>
<td>15q22</td>
<td></td>
</tr>
<tr>
<td>Hs.154104</td>
<td>PLAGL2</td>
<td>5326</td>
<td>20q11.21</td>
<td></td>
</tr>
<tr>
<td>Hs.591953</td>
<td>PLCB3</td>
<td>5331</td>
<td>11q13</td>
<td></td>
</tr>
<tr>
<td>Hs.592049</td>
<td>PLK1</td>
<td>5347</td>
<td>16p12.1</td>
<td></td>
</tr>
<tr>
<td>Hs.279413</td>
<td>POLD1</td>
<td>5424</td>
<td>19q13.3</td>
<td></td>
</tr>
<tr>
<td>Hs.306791</td>
<td>POLD2</td>
<td>5425</td>
<td>7p13</td>
<td></td>
</tr>
<tr>
<td>Hs.356331</td>
<td>PPIA</td>
<td>5478</td>
<td>7p13</td>
<td></td>
</tr>
<tr>
<td>Hs.183994</td>
<td>PPP1CA</td>
<td>5499</td>
<td>11q13</td>
<td></td>
</tr>
<tr>
<td>Hs.533308</td>
<td>PPP2R5D</td>
<td>5528</td>
<td>6p21.1</td>
<td></td>
</tr>
<tr>
<td>Hs.516948</td>
<td>PRCC</td>
<td>5546</td>
<td>1q21.1</td>
<td></td>
</tr>
<tr>
<td>Hs.465627</td>
<td>MAP2K2</td>
<td>5605</td>
<td>19p13.3</td>
<td></td>
</tr>
<tr>
<td>Hs.523004</td>
<td>PSAP</td>
<td>5660</td>
<td>10q21-q22</td>
<td></td>
</tr>
<tr>
<td>Hs.89545</td>
<td>PSMB4</td>
<td>5692</td>
<td>1q21</td>
<td></td>
</tr>
<tr>
<td>Hs.77060</td>
<td>PSMB6</td>
<td>5694</td>
<td>17p13</td>
<td></td>
</tr>
<tr>
<td>Hs.250758</td>
<td>PSMC3</td>
<td>5702</td>
<td>11p12-p13</td>
<td></td>
</tr>
<tr>
<td>Hs.211594</td>
<td>PSMC4</td>
<td>5704</td>
<td>19q13.11-q13.13</td>
<td></td>
</tr>
<tr>
<td>Hs.518464</td>
<td>PSMC2</td>
<td>5708</td>
<td>3q27.1</td>
<td></td>
</tr>
<tr>
<td>Hs.459927</td>
<td>PTMA</td>
<td>5757</td>
<td>2q35-q36</td>
<td></td>
</tr>
<tr>
<td>Hs.458332</td>
<td>PYCR1</td>
<td>5831</td>
<td>17q25.3</td>
<td></td>
</tr>
<tr>
<td>Hs.368157</td>
<td>PYGB</td>
<td>5834</td>
<td>20p11.2-p11.1</td>
<td></td>
</tr>
<tr>
<td>Hs.647062</td>
<td>RFC2</td>
<td>5982</td>
<td>7q11.23</td>
<td></td>
</tr>
<tr>
<td>Hs.461925</td>
<td>RPA1</td>
<td>6117</td>
<td>17p13.3</td>
<td></td>
</tr>
<tr>
<td>Hs.79411</td>
<td>RPA2</td>
<td>6118</td>
<td>1p35</td>
<td></td>
</tr>
<tr>
<td>Hs.644628</td>
<td>RPL4</td>
<td>6124</td>
<td>15q22</td>
<td></td>
</tr>
<tr>
<td>Hs.186350</td>
<td>RPL4</td>
<td>6124</td>
<td>15q22</td>
<td></td>
</tr>
<tr>
<td>Hs.571841</td>
<td>RPL7</td>
<td>6129</td>
<td>8q21.11</td>
<td></td>
</tr>
<tr>
<td>Hs.499839</td>
<td>RPL7A</td>
<td>6130</td>
<td>9q34</td>
<td></td>
</tr>
<tr>
<td>Hs.178551</td>
<td>RPL8</td>
<td>6132</td>
<td>8q24.3</td>
<td></td>
</tr>
<tr>
<td>Hs.546285</td>
<td>RPLP0</td>
<td>6175</td>
<td>12q24.2</td>
<td></td>
</tr>
<tr>
<td>Hs.109059</td>
<td>MRPL12</td>
<td>6182</td>
<td>17q25</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Description</td>
<td>Chromosome</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------------</td>
<td>-------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>Hs.370895</td>
<td>6185</td>
<td>20q12-q13.1</td>
<td></td>
</tr>
<tr>
<td>176</td>
<td>Hs.506997</td>
<td>6187</td>
<td>16p13.3</td>
<td></td>
</tr>
<tr>
<td>177</td>
<td>Hs.498569</td>
<td>6187</td>
<td>16p13.3</td>
<td></td>
</tr>
<tr>
<td>178</td>
<td>Hs.356366</td>
<td>6187</td>
<td>16p13.3</td>
<td></td>
</tr>
<tr>
<td>179</td>
<td>Hs.546286</td>
<td>6188</td>
<td>11q13.3-q13.5</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>Hs.356572</td>
<td>6189</td>
<td>4q31.2-q31.3</td>
<td></td>
</tr>
<tr>
<td>181</td>
<td>Hs.446628</td>
<td>6191</td>
<td>Xq13.1</td>
<td></td>
</tr>
<tr>
<td>182</td>
<td>Hs.378103</td>
<td>6193</td>
<td>19q13.4</td>
<td></td>
</tr>
<tr>
<td>183</td>
<td>Hs.381126</td>
<td>6208</td>
<td>5q31-q33</td>
<td></td>
</tr>
<tr>
<td>184</td>
<td>Hs.433427</td>
<td>6218</td>
<td>15q</td>
<td></td>
</tr>
<tr>
<td>185</td>
<td>Hs.226390</td>
<td>6241</td>
<td>2p25-p24</td>
<td></td>
</tr>
<tr>
<td>186</td>
<td>Hs.436687</td>
<td>6418</td>
<td>9q34</td>
<td></td>
</tr>
<tr>
<td>187</td>
<td>Hs.97616</td>
<td>6455</td>
<td>19p13.3</td>
<td></td>
</tr>
<tr>
<td>188</td>
<td>Hs.23348</td>
<td>6502</td>
<td>5p13</td>
<td></td>
</tr>
<tr>
<td>189</td>
<td>Hs.631582</td>
<td>6510</td>
<td>19q13.3</td>
<td></td>
</tr>
<tr>
<td>190</td>
<td>Hs.187946</td>
<td>6574</td>
<td>2q11-q14</td>
<td></td>
</tr>
<tr>
<td>191</td>
<td>Hs.118400</td>
<td>6624</td>
<td>7p22</td>
<td></td>
</tr>
<tr>
<td>192</td>
<td>Hs.83753</td>
<td>6628</td>
<td>20p13</td>
<td></td>
</tr>
<tr>
<td>193</td>
<td>Hs.707993</td>
<td>6662</td>
<td>17q24.3-q25.1</td>
<td></td>
</tr>
<tr>
<td>194</td>
<td>Hs.301540</td>
<td>6697</td>
<td>2p14-p12</td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>Hs.443258</td>
<td>6721</td>
<td>22q13</td>
<td></td>
</tr>
<tr>
<td>196</td>
<td>Hs.511425</td>
<td>6726</td>
<td>1q42.12</td>
<td></td>
</tr>
<tr>
<td>197</td>
<td>Hs.523680</td>
<td>6749</td>
<td>11q12</td>
<td></td>
</tr>
<tr>
<td>198</td>
<td>Hs.250822</td>
<td>6790</td>
<td>20q13.2-q13.3</td>
<td></td>
</tr>
<tr>
<td>199</td>
<td>Hs.481860</td>
<td>6897</td>
<td>5p13.2</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>Hs.519672</td>
<td>6949</td>
<td>5q32-q33.1</td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>Hs.708025</td>
<td>7009</td>
<td>12q12-q13</td>
<td></td>
</tr>
<tr>
<td>202</td>
<td>Hs.513305</td>
<td>7023</td>
<td>16p13</td>
<td></td>
</tr>
<tr>
<td>203</td>
<td>Hs.645227</td>
<td>7040</td>
<td>19q13.2</td>
<td>19q13.1</td>
</tr>
<tr>
<td>204</td>
<td>Hs.78769</td>
<td>7064</td>
<td>19q13.3</td>
<td></td>
</tr>
<tr>
<td>205</td>
<td>Hs.515122</td>
<td>7083</td>
<td>17q23.2-q25.3</td>
<td></td>
</tr>
<tr>
<td>206</td>
<td>Hs.707975</td>
<td>7132</td>
<td>12p13.2</td>
<td></td>
</tr>
<tr>
<td>207</td>
<td>Hs.524219</td>
<td>7167</td>
<td>12p13</td>
<td></td>
</tr>
<tr>
<td>208</td>
<td>Hs.654421</td>
<td>7170</td>
<td>1q21.2</td>
<td></td>
</tr>
<tr>
<td>209</td>
<td>Hs.12084</td>
<td>7284</td>
<td>16p11.2</td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>Hs.170107</td>
<td>7386</td>
<td>19q12-q13.1</td>
<td></td>
</tr>
<tr>
<td>211</td>
<td>Hs.78601</td>
<td>7389</td>
<td>1p34</td>
<td></td>
</tr>
<tr>
<td>212</td>
<td>Hs.520943</td>
<td>7458</td>
<td>7q11.23</td>
<td></td>
</tr>
<tr>
<td>213</td>
<td>Hs.707878</td>
<td>7538</td>
<td>19q13.1</td>
<td></td>
</tr>
<tr>
<td>214</td>
<td>Hs.662176</td>
<td>7643</td>
<td>19p13.1-p12</td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>Hs.234521</td>
<td>7867</td>
<td>3p21.3</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>Hs.695957</td>
<td>7913</td>
<td>6p22.3</td>
<td></td>
</tr>
<tr>
<td>217</td>
<td>Hs.283565</td>
<td>8061</td>
<td>11q13</td>
<td></td>
</tr>
<tr>
<td>218</td>
<td>Hs.631661</td>
<td>8078</td>
<td>12p13</td>
<td></td>
</tr>
<tr>
<td>219</td>
<td>Hs.524214</td>
<td>8079</td>
<td>12p13</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>Hs.513797</td>
<td>8140</td>
<td>16q24.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gene Symbol</td>
<td>Chromosome</td>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>------------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>221</td>
<td>Hs.115232</td>
<td>SF3A2</td>
<td>8175</td>
<td>19p13.3-p13.2</td>
</tr>
<tr>
<td>222</td>
<td>Hs.75238</td>
<td>CHAF1B</td>
<td>8208</td>
<td>21q22.13</td>
</tr>
<tr>
<td>223</td>
<td>Hs.401509</td>
<td>RBM10</td>
<td>8241</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>224</td>
<td>Hs.592082</td>
<td>AXIN1</td>
<td>8312</td>
<td>16p13.3</td>
</tr>
<tr>
<td>225</td>
<td>Hs.134999</td>
<td>HIST1H2AM</td>
<td>8336</td>
<td>6p22-p21.3</td>
</tr>
<tr>
<td>226</td>
<td>Hs.706783</td>
<td>RAD54L</td>
<td>8438</td>
<td>1p32</td>
</tr>
<tr>
<td>227</td>
<td>Hs.405046</td>
<td>CBX4</td>
<td>8535</td>
<td>17q25.3</td>
</tr>
<tr>
<td>228</td>
<td>Hs.708050</td>
<td>PPAP2B</td>
<td>8613</td>
<td>1pter-p22.1</td>
</tr>
<tr>
<td>229</td>
<td>Hs.371001</td>
<td>EIF3B</td>
<td>8662</td>
<td>7p22.2</td>
</tr>
<tr>
<td>230</td>
<td>Hs.492599</td>
<td>EIF3H</td>
<td>8667</td>
<td>8q24.11</td>
</tr>
<tr>
<td>231</td>
<td>Hs.86131</td>
<td>FADD</td>
<td>8772</td>
<td>11q13.3</td>
</tr>
<tr>
<td>232</td>
<td>Hs.212680</td>
<td>TNFRSF18</td>
<td>8784</td>
<td>1p36.3</td>
</tr>
<tr>
<td>233</td>
<td>Hs.412842</td>
<td>CDC123</td>
<td>8872</td>
<td>10p13</td>
</tr>
<tr>
<td>234</td>
<td>Hs.591942</td>
<td>ZNF259</td>
<td>8882</td>
<td>11q23.3</td>
</tr>
<tr>
<td>235</td>
<td>Hs.118631</td>
<td>TIMELESS</td>
<td>8914</td>
<td>12q12-q13</td>
</tr>
<tr>
<td>236</td>
<td>Hs.446522</td>
<td>RPL14</td>
<td>9045</td>
<td>3p22-p21.2</td>
</tr>
<tr>
<td>237</td>
<td>Hs.567385</td>
<td>PRC1</td>
<td>9055</td>
<td>15q26.1</td>
</tr>
<tr>
<td>238</td>
<td>Hs.54277</td>
<td>FAM50A</td>
<td>9130</td>
<td>Xq28</td>
</tr>
<tr>
<td>239</td>
<td>Hs.194698</td>
<td>CCNB2</td>
<td>9133</td>
<td>15q22.2</td>
</tr>
<tr>
<td>240</td>
<td>Hs.514590</td>
<td>HGS</td>
<td>9146</td>
<td>17q25</td>
</tr>
<tr>
<td>241</td>
<td>Hs.498248</td>
<td>EXO1</td>
<td>9156</td>
<td>1q42-q43</td>
</tr>
<tr>
<td>242</td>
<td>Hs.576875</td>
<td>DDX21</td>
<td>9188</td>
<td>10q21</td>
</tr>
<tr>
<td>243</td>
<td>Hs.442658</td>
<td>AURKB</td>
<td>9212</td>
<td>17p13.1</td>
</tr>
<tr>
<td>244</td>
<td>Hs.151787</td>
<td>EFTUD2</td>
<td>9343</td>
<td>17q21.31</td>
</tr>
<tr>
<td>245</td>
<td>Hs.350265</td>
<td>LONP1</td>
<td>9361</td>
<td>19p13.2</td>
</tr>
<tr>
<td>246</td>
<td>Hs.31442</td>
<td>RECQL4</td>
<td>9401</td>
<td>8q24.3</td>
</tr>
<tr>
<td>247</td>
<td>Hs.5258</td>
<td>MAGED1</td>
<td>9500</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>248</td>
<td>Hs.522810</td>
<td>CSAG2</td>
<td>9598</td>
<td>Xq28</td>
</tr>
<tr>
<td>249</td>
<td>Hs.118351</td>
<td>UBE3C</td>
<td>9690</td>
<td>7q36.3</td>
</tr>
<tr>
<td>250</td>
<td>Hs.81892</td>
<td>KIAA0101</td>
<td>9768</td>
<td>15q22.31</td>
</tr>
<tr>
<td>251</td>
<td>Hs.292579</td>
<td>PTDSS1</td>
<td>9791</td>
<td>8q22</td>
</tr>
<tr>
<td>252</td>
<td>Hs.5719</td>
<td>NCAPD2</td>
<td>9918</td>
<td>12p13.3</td>
</tr>
<tr>
<td>253</td>
<td>Hs.656243</td>
<td>AMMCR1</td>
<td>9949</td>
<td>Xq22.3</td>
</tr>
<tr>
<td>254</td>
<td>Hs.654972</td>
<td>RCE1</td>
<td>9986</td>
<td>11q13</td>
</tr>
<tr>
<td>255</td>
<td>Hs.497353</td>
<td>MED6</td>
<td>10001</td>
<td>14q24.2</td>
</tr>
<tr>
<td>256</td>
<td>Hs.79018</td>
<td>CHAF1A</td>
<td>10036</td>
<td>19p13.3</td>
</tr>
<tr>
<td>257</td>
<td>Hs.58992</td>
<td>SMC4</td>
<td>10051</td>
<td>3q26.1</td>
</tr>
<tr>
<td>258</td>
<td>Hs.515500</td>
<td>SAE1</td>
<td>10055</td>
<td>19q13.32</td>
</tr>
<tr>
<td>259</td>
<td>Hs.654958</td>
<td>ABCF2</td>
<td>10061</td>
<td>7q36</td>
</tr>
<tr>
<td>260</td>
<td>Hs.489284</td>
<td>ARPC1B</td>
<td>10095</td>
<td>7q22.1</td>
</tr>
<tr>
<td>261</td>
<td>Hs.73625</td>
<td>KIF20A</td>
<td>10112</td>
<td>5q31</td>
</tr>
<tr>
<td>262</td>
<td>Hs.522817</td>
<td>BCAP31</td>
<td>10173</td>
<td>Xq28</td>
</tr>
<tr>
<td>263</td>
<td>Hs.467408</td>
<td>TRIM28</td>
<td>10155</td>
<td>19q13.4</td>
</tr>
<tr>
<td>264</td>
<td>Hs.696342</td>
<td>NUTF2</td>
<td>10204</td>
<td>16q22.1</td>
</tr>
<tr>
<td>265</td>
<td>Hs.516160</td>
<td>SF3B4</td>
<td>10262</td>
<td>1q12-q21</td>
</tr>
<tr>
<td>266</td>
<td>Hs.522087</td>
<td>OPRS1</td>
<td>10280</td>
<td>9p13.3</td>
</tr>
<tr>
<td>No.</td>
<td>Accession</td>
<td>Gene</td>
<td>Chromosome</td>
<td>Start</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>267</td>
<td>Hs.20447</td>
<td>PAK4</td>
<td>10q21.33</td>
<td>10298</td>
</tr>
<tr>
<td>268</td>
<td>Hs.173162</td>
<td>COX4NB</td>
<td>10q21.33</td>
<td>10328</td>
</tr>
<tr>
<td>269</td>
<td>Hs.707307</td>
<td>PCGF3</td>
<td>10p11.2</td>
<td>10336</td>
</tr>
<tr>
<td>270</td>
<td>Hs.524390</td>
<td>TUBA1B</td>
<td>10p11.2</td>
<td>10376</td>
</tr>
<tr>
<td>271</td>
<td>Hs.705373</td>
<td>TUBA1B</td>
<td>10p11.2</td>
<td>10376</td>
</tr>
<tr>
<td>272</td>
<td>Hs.433615</td>
<td>TUBB2C</td>
<td>10p11.2</td>
<td>10383</td>
</tr>
<tr>
<td>273</td>
<td>Hs.5662</td>
<td>GNB2L1</td>
<td>10q21.33</td>
<td>10399</td>
</tr>
<tr>
<td>274</td>
<td>Hs.435850</td>
<td>LYPLA1</td>
<td>10p11.2</td>
<td>10434</td>
</tr>
<tr>
<td>275</td>
<td>Hs.655909</td>
<td>TOMM40</td>
<td>10q21.33</td>
<td>10452</td>
</tr>
<tr>
<td>276</td>
<td>Hs.104019</td>
<td>TACC3</td>
<td>10p11.2</td>
<td>10460</td>
</tr>
<tr>
<td>277</td>
<td>Hs.386390</td>
<td>TADA3L</td>
<td>10p11.2</td>
<td>10474</td>
</tr>
<tr>
<td>278</td>
<td>Hs.470804</td>
<td>UBE2E3</td>
<td>10p11.2</td>
<td>10477</td>
</tr>
<tr>
<td>279</td>
<td>Hs.371416</td>
<td>CARM1</td>
<td>10p11.2</td>
<td>10498</td>
</tr>
<tr>
<td>280</td>
<td>Hs.532851</td>
<td>RNASEH2A</td>
<td>10p11.2</td>
<td>10535</td>
</tr>
<tr>
<td>281</td>
<td>Hs.494604</td>
<td>ANP32B</td>
<td>10p11.2</td>
<td>10541</td>
</tr>
<tr>
<td>282</td>
<td>Hs.518774</td>
<td>PAICS</td>
<td>10p11.2</td>
<td>10606</td>
</tr>
<tr>
<td>283</td>
<td>Hs.514033</td>
<td>SPAG5</td>
<td>10p11.2</td>
<td>10615</td>
</tr>
<tr>
<td>284</td>
<td>Hs.144936</td>
<td>IGF2BP1</td>
<td>10p11.2</td>
<td>10642</td>
</tr>
<tr>
<td>285</td>
<td>Hs.520794</td>
<td>YKT6</td>
<td>10p11.2</td>
<td>10652</td>
</tr>
<tr>
<td>286</td>
<td>Hs.348308</td>
<td>C1orf2</td>
<td>10p11.2</td>
<td>10712</td>
</tr>
<tr>
<td>287</td>
<td>Hs.241517</td>
<td>POLQ</td>
<td>10p11.2</td>
<td>10721</td>
</tr>
<tr>
<td>288</td>
<td>Hs.263812</td>
<td>NUDC</td>
<td>10p11.2</td>
<td>10726</td>
</tr>
<tr>
<td>289</td>
<td>Hs.101937</td>
<td>SIX2</td>
<td>10p11.2</td>
<td>10736</td>
</tr>
<tr>
<td>290</td>
<td>Hs.435120</td>
<td>KIF1C</td>
<td>10p11.2</td>
<td>10749</td>
</tr>
<tr>
<td>291</td>
<td>Hs.655012</td>
<td>GIPC1</td>
<td>10p11.2</td>
<td>10755</td>
</tr>
<tr>
<td>292</td>
<td>Hs.469030</td>
<td>MTHFD2</td>
<td>10p11.2</td>
<td>10797</td>
</tr>
<tr>
<td>293</td>
<td>Hs.458598</td>
<td>UTP14A</td>
<td>10p11.2</td>
<td>10813</td>
</tr>
<tr>
<td>294</td>
<td>Hs.705916</td>
<td>CD3EAP</td>
<td>10p11.2</td>
<td>10849</td>
</tr>
<tr>
<td>295</td>
<td>Hs.310809</td>
<td>WDR3</td>
<td>10p11.2</td>
<td>10885</td>
</tr>
<tr>
<td>296</td>
<td>Hs.337295</td>
<td>STIP1</td>
<td>10p11.2</td>
<td>10963</td>
</tr>
<tr>
<td>297</td>
<td>Hs.74405</td>
<td>YWHAQ</td>
<td>10p11.2</td>
<td>10971</td>
</tr>
<tr>
<td>298</td>
<td>Hs.15591</td>
<td>COPS6</td>
<td>10p11.2</td>
<td>10980</td>
</tr>
<tr>
<td>299</td>
<td>Hs.298716</td>
<td>GCN1L1</td>
<td>10p11.2</td>
<td>10985</td>
</tr>
<tr>
<td>300</td>
<td>Hs.528834</td>
<td>NUDT21</td>
<td>10p11.2</td>
<td>11051</td>
</tr>
<tr>
<td>301</td>
<td>Hs.93002</td>
<td>UBE2C</td>
<td>10p11.2</td>
<td>11065</td>
</tr>
<tr>
<td>302</td>
<td>Hs.397638</td>
<td>WDR5</td>
<td>10p11.2</td>
<td>11091</td>
</tr>
<tr>
<td>303</td>
<td>Hs.708055</td>
<td>HNRPU1</td>
<td>10p11.2</td>
<td>11100</td>
</tr>
<tr>
<td>304</td>
<td>Hs.591363</td>
<td>ZWINT</td>
<td>10p11.2</td>
<td>11130</td>
</tr>
<tr>
<td>305</td>
<td>Hs.160958</td>
<td>CDC37</td>
<td>10p11.2</td>
<td>11140</td>
</tr>
<tr>
<td>306</td>
<td>Hs.478150</td>
<td>PDCD10</td>
<td>10p11.2</td>
<td>11235</td>
</tr>
<tr>
<td>307</td>
<td>Hs.519993</td>
<td>NRM</td>
<td>10p11.2</td>
<td>11270</td>
</tr>
<tr>
<td>308</td>
<td>Hs.567419</td>
<td>MGAT4B</td>
<td>10p11.2</td>
<td>11282</td>
</tr>
<tr>
<td>309</td>
<td>Hs.504620</td>
<td>PHB2</td>
<td>10p11.2</td>
<td>11331</td>
</tr>
<tr>
<td>310</td>
<td>Hs.632296</td>
<td>PDAP1</td>
<td>10p11.2</td>
<td>11333</td>
</tr>
<tr>
<td>311</td>
<td>Hs.531563</td>
<td>DOLK</td>
<td>10p11.2</td>
<td>22845</td>
</tr>
<tr>
<td>312</td>
<td>Hs.515610</td>
<td>SAPS1</td>
<td>10p11.2</td>
<td>22870</td>
</tr>
<tr>
<td>313</td>
<td>Hs.244580</td>
<td>TPX2</td>
<td>22974</td>
<td>20q11.2</td>
</tr>
<tr>
<td>314</td>
<td>Hs.586219</td>
<td>KIAA0194</td>
<td>22993</td>
<td>5q33.1</td>
</tr>
<tr>
<td>315</td>
<td>Hs.246112</td>
<td>ASCC3L1</td>
<td>23020</td>
<td>2q11.2</td>
</tr>
<tr>
<td>316</td>
<td>Hs.155829</td>
<td>TBC1D9B</td>
<td>23061</td>
<td>5q35.3</td>
</tr>
<tr>
<td>317</td>
<td>Hs.517670</td>
<td>TTLL12</td>
<td>23170</td>
<td>22q13.31</td>
</tr>
<tr>
<td>318</td>
<td>Hs.535901</td>
<td>BOP1</td>
<td>23246</td>
<td>8q24.3</td>
</tr>
<tr>
<td>319</td>
<td>Hs.645279</td>
<td>BOP1</td>
<td>23246</td>
<td>8q24.3</td>
</tr>
<tr>
<td>320</td>
<td>Hs.583391</td>
<td>NOMO1</td>
<td>23420</td>
<td>16p13.11</td>
</tr>
<tr>
<td>321</td>
<td>Hs.513484</td>
<td>QPRT</td>
<td>23475</td>
<td>16p11.2</td>
</tr>
<tr>
<td>322</td>
<td>Hs.436329</td>
<td>SCRIB</td>
<td>23513</td>
<td>8q24.3</td>
</tr>
<tr>
<td>323</td>
<td>Hs.32018</td>
<td>SNAPIN</td>
<td>23557</td>
<td>1q21.3</td>
</tr>
<tr>
<td>324</td>
<td>Hs.49760</td>
<td>ORC6L</td>
<td>23594</td>
<td>16q12</td>
</tr>
<tr>
<td>325</td>
<td>Hs.173464</td>
<td>FKBP8</td>
<td>23770</td>
<td>19p12</td>
</tr>
<tr>
<td>326</td>
<td>Hs.648326</td>
<td>KIF4A</td>
<td>24137</td>
<td>Xq13.1</td>
</tr>
<tr>
<td>327</td>
<td>Hs.33334</td>
<td>PRPF6</td>
<td>24148</td>
<td>20q13.33</td>
</tr>
<tr>
<td>328</td>
<td>Hs.50915</td>
<td>KLK5</td>
<td>25818</td>
<td>19q13.3-q13.4</td>
</tr>
<tr>
<td>329</td>
<td>Hs.706873</td>
<td>SIN3A</td>
<td>25942</td>
<td>15q24.2</td>
</tr>
<tr>
<td>330</td>
<td>Hs.11314</td>
<td>C20orf4</td>
<td>25980</td>
<td>20pter-q12</td>
</tr>
<tr>
<td>331</td>
<td>Hs.708014</td>
<td>SERBP1</td>
<td>26135</td>
<td>1p31</td>
</tr>
<tr>
<td>332</td>
<td>Hs.532129</td>
<td>TSPAN17</td>
<td>26262</td>
<td>5q35.3</td>
</tr>
<tr>
<td>333</td>
<td>Hs.279529</td>
<td>PRELID1</td>
<td>27166</td>
<td>5q35.3</td>
</tr>
<tr>
<td>334</td>
<td>Hs.396393</td>
<td>UBE2S</td>
<td>27338</td>
<td>1q13.43</td>
</tr>
<tr>
<td>335</td>
<td>Hs.502705</td>
<td>PRPF19</td>
<td>27339</td>
<td>11q12.2</td>
</tr>
<tr>
<td>336</td>
<td>Hs.534041</td>
<td>CTA-126B4.3</td>
<td>27341</td>
<td>22q13.2-q13.31</td>
</tr>
<tr>
<td>337</td>
<td>Hs.504249</td>
<td>DCPS</td>
<td>28960</td>
<td>11q24.2</td>
</tr>
<tr>
<td>338</td>
<td>Hs.18349</td>
<td>MRPL15</td>
<td>29088</td>
<td>8q11.2-q13</td>
</tr>
<tr>
<td>339</td>
<td>Hs.5199</td>
<td>UBE2T</td>
<td>29089</td>
<td>1q32.1</td>
</tr>
<tr>
<td>340</td>
<td>Hs.157351</td>
<td>OLA1</td>
<td>29789</td>
<td>2p31.1</td>
</tr>
<tr>
<td>341</td>
<td>Hs.279877</td>
<td>FTSJ2</td>
<td>29960</td>
<td>7p22</td>
</tr>
<tr>
<td>342</td>
<td>Hs.436341</td>
<td>DONSON</td>
<td>29980</td>
<td>1q21.2</td>
</tr>
<tr>
<td>343</td>
<td>Hs.3439</td>
<td>STOML2</td>
<td>30968</td>
<td>9p13.1</td>
</tr>
<tr>
<td>344</td>
<td>Hs.645463</td>
<td>THAP4</td>
<td>51078</td>
<td>2q37.3</td>
</tr>
<tr>
<td>345</td>
<td>Hs.463797</td>
<td>MRTO4</td>
<td>51154</td>
<td>1p36.13</td>
</tr>
<tr>
<td>346</td>
<td>Hs.381256</td>
<td>GLTP</td>
<td>51228</td>
<td>12q24.11</td>
</tr>
<tr>
<td>347</td>
<td>Hs.584908</td>
<td>MRPL37</td>
<td>51253</td>
<td>1p32.1</td>
</tr>
<tr>
<td>348</td>
<td>Hs.475387</td>
<td>CHCHD8</td>
<td>51287</td>
<td>11q13.4</td>
</tr>
<tr>
<td>349</td>
<td>Hs.558499</td>
<td>CD320</td>
<td>51293</td>
<td>19p13.3-p13.2</td>
</tr>
<tr>
<td>350</td>
<td>Hs.69499</td>
<td>TRIAP1</td>
<td>51499</td>
<td>12q24.31</td>
</tr>
<tr>
<td>351</td>
<td>Hs.386189</td>
<td>GTSE1</td>
<td>51512</td>
<td>22q13.2-q13.3</td>
</tr>
<tr>
<td>352</td>
<td>Hs.706898</td>
<td>MTP18</td>
<td>51537</td>
<td>22q</td>
</tr>
<tr>
<td>353</td>
<td>Hs.528641</td>
<td>SIRT7</td>
<td>51547</td>
<td>17q25</td>
</tr>
<tr>
<td>354</td>
<td>Hs.514216</td>
<td>SLC25A39</td>
<td>51629</td>
<td>17q12</td>
</tr>
<tr>
<td>355</td>
<td>Hs.655138</td>
<td>WBP11</td>
<td>51729</td>
<td>12p12.3</td>
</tr>
<tr>
<td>356</td>
<td>Hs.193326</td>
<td>FGFR1</td>
<td>53834</td>
<td>4p16</td>
</tr>
<tr>
<td>357</td>
<td>Hs.437060</td>
<td>CYCS</td>
<td>54205</td>
<td>7p15.2</td>
</tr>
<tr>
<td>358</td>
<td>Hs.592116</td>
<td>FAM64A</td>
<td>54478</td>
<td>17p13.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>359</td>
<td>Hs.481836</td>
<td>MTMR12</td>
<td>54545</td>
<td>5p13.3</td>
</tr>
<tr>
<td>360</td>
<td>Hs.654762</td>
<td>DDX56</td>
<td>54606</td>
<td>7p13</td>
</tr>
<tr>
<td>361</td>
<td>Hs.485449</td>
<td>GTPBP2</td>
<td>54676</td>
<td>6p21-p12</td>
</tr>
<tr>
<td>362</td>
<td>Hs.700127</td>
<td>QPCTL</td>
<td>54814</td>
<td>19q13.32</td>
</tr>
<tr>
<td>363</td>
<td>Hs.481526</td>
<td>NSUN2</td>
<td>54888</td>
<td>5p15.31</td>
</tr>
<tr>
<td>364</td>
<td>Hs.572318</td>
<td>TIPIN</td>
<td>54962</td>
<td>15q22.31</td>
</tr>
<tr>
<td>365</td>
<td>Hs.330663</td>
<td>C12orf48</td>
<td>55010</td>
<td>12q23.2</td>
</tr>
<tr>
<td>366</td>
<td>Hs.34045</td>
<td>CDCA4</td>
<td>55038</td>
<td>14q32.33</td>
</tr>
<tr>
<td>367</td>
<td>Hs.524571</td>
<td>CDCA8</td>
<td>55143</td>
<td>1p34.3</td>
</tr>
<tr>
<td>368</td>
<td>Hs.14559</td>
<td>CEP55</td>
<td>55165</td>
<td>10q23.33</td>
</tr>
<tr>
<td>369</td>
<td>Hs.655253</td>
<td>ATAD3A</td>
<td>55210</td>
<td>1p36.33</td>
</tr>
<tr>
<td>370</td>
<td>Hs.513126</td>
<td>FANCI</td>
<td>55215</td>
<td>15q26.1</td>
</tr>
<tr>
<td>371</td>
<td>Hs.267446</td>
<td>FLJ11184</td>
<td>55319</td>
<td>4q32.3</td>
</tr>
<tr>
<td>372</td>
<td>Hs.7570</td>
<td>CNO</td>
<td>55330</td>
<td>4p16.1</td>
</tr>
<tr>
<td>373</td>
<td>Hs.532968</td>
<td>HJURP</td>
<td>55355</td>
<td>2q37.1</td>
</tr>
<tr>
<td>374</td>
<td>Hs.518265</td>
<td>CDV3</td>
<td>55573</td>
<td>3q22.1</td>
</tr>
<tr>
<td>375</td>
<td>Hs.516450</td>
<td>SMPD4</td>
<td>55627</td>
<td>2q2.11</td>
</tr>
<tr>
<td>376</td>
<td>Hs.27222</td>
<td>NOLA2</td>
<td>55651</td>
<td>5q35.3</td>
</tr>
<tr>
<td>377</td>
<td>Hs.707116</td>
<td>ZNF446</td>
<td>55663</td>
<td>19q13.43</td>
</tr>
<tr>
<td>378</td>
<td>Hs.567803</td>
<td>C9orf86</td>
<td>55684</td>
<td>9q34.3</td>
</tr>
<tr>
<td>379</td>
<td>Hs.55028</td>
<td>CENPN</td>
<td>55839</td>
<td>16q23.2</td>
</tr>
<tr>
<td>380</td>
<td>Hs.656063</td>
<td>BAIAP2L1</td>
<td>55971</td>
<td>7q21.3</td>
</tr>
<tr>
<td>381</td>
<td>Hs.472667</td>
<td>CTNNB1</td>
<td>56259</td>
<td>20q11.23-q12</td>
</tr>
<tr>
<td>382</td>
<td>Hs.22678</td>
<td>C10orf2</td>
<td>56652</td>
<td>10q23.3-q24.3</td>
</tr>
<tr>
<td>383</td>
<td>Hs.283739</td>
<td>UBQLN4</td>
<td>56893</td>
<td>1q21</td>
</tr>
<tr>
<td>384</td>
<td>Hs.250456</td>
<td>DHX33</td>
<td>56919</td>
<td>17p13.2</td>
</tr>
<tr>
<td>385</td>
<td>Hs.663740</td>
<td>ARNTL2</td>
<td>56938</td>
<td>12p12.2-p11.2</td>
</tr>
<tr>
<td>386</td>
<td>Hs.283734</td>
<td>MRPL47</td>
<td>57129</td>
<td>3q26.33</td>
</tr>
<tr>
<td>387</td>
<td>Hs.268488</td>
<td>LRRC47</td>
<td>57470</td>
<td>1p36.32</td>
</tr>
<tr>
<td>388</td>
<td>Hs.158381</td>
<td>AARS2</td>
<td>57505</td>
<td>6p21.1</td>
</tr>
<tr>
<td>389</td>
<td>Hs.107382</td>
<td>DHX37</td>
<td>57647</td>
<td>12q24.31</td>
</tr>
<tr>
<td>390</td>
<td>Hs.444173</td>
<td>PHF12</td>
<td>57649</td>
<td>17q11.2</td>
</tr>
<tr>
<td>391</td>
<td>Hs.325838</td>
<td>KIAA1542</td>
<td>57661</td>
<td>11p15.5</td>
</tr>
<tr>
<td>392</td>
<td>Hs.465829</td>
<td>ZNF317</td>
<td>57693</td>
<td>NA</td>
</tr>
<tr>
<td>393</td>
<td>Hs.516826</td>
<td>TRIB3</td>
<td>57761</td>
<td>20p13-p12.2</td>
</tr>
<tr>
<td>394</td>
<td>Hs.654720</td>
<td>KIAA1967</td>
<td>57805</td>
<td>8p22</td>
</tr>
<tr>
<td>395</td>
<td>Hs.107003</td>
<td>CCNB1IP1</td>
<td>57820</td>
<td>14q11.2</td>
</tr>
<tr>
<td>396</td>
<td>Hs.708201</td>
<td>CXCL16</td>
<td>58191</td>
<td>17p13</td>
</tr>
<tr>
<td>397</td>
<td>Hs.477498</td>
<td>EEFSEC</td>
<td>60678</td>
<td>3q21.3</td>
</tr>
<tr>
<td>398</td>
<td>Hs.175613</td>
<td>CLSPN</td>
<td>63967</td>
<td>1p34.2</td>
</tr>
<tr>
<td>399</td>
<td>Hs.604789</td>
<td>MCCC2</td>
<td>64087</td>
<td>5q12-q13</td>
</tr>
<tr>
<td>400</td>
<td>Hs.706966</td>
<td>HERPUD2</td>
<td>64224</td>
<td>7p14.2</td>
</tr>
<tr>
<td>401</td>
<td>Hs.527989</td>
<td>NXN</td>
<td>64359</td>
<td>17p13.3</td>
</tr>
<tr>
<td>402</td>
<td>Hs.15825</td>
<td>NOM1</td>
<td>64434</td>
<td>7q36.3</td>
</tr>
<tr>
<td>403</td>
<td>Hs.436102</td>
<td>ISG20L1</td>
<td>64782</td>
<td>15q26.1</td>
</tr>
<tr>
<td>404</td>
<td>Hs.18946</td>
<td>MRPS26</td>
<td>64949</td>
<td>20p13</td>
</tr>
<tr>
<td></td>
<td>Gene</td>
<td>Chromosome</td>
<td>Position</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------------</td>
<td>------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>Hs.103832 UPF3B</td>
<td>Xq25-q26</td>
<td>65109</td>
<td></td>
</tr>
<tr>
<td>406</td>
<td>Hs.157160 MRPS34</td>
<td>16p13.3</td>
<td>65993</td>
<td></td>
</tr>
<tr>
<td>407</td>
<td>Hs.437059 C17orf53</td>
<td>17q21.31</td>
<td>78995</td>
<td></td>
</tr>
<tr>
<td>408</td>
<td>Hs.209979 LRFN4</td>
<td>11q13.1</td>
<td>78999</td>
<td></td>
</tr>
<tr>
<td>409</td>
<td>Hs.208912 CENPM</td>
<td>22q13.2</td>
<td>79019</td>
<td></td>
</tr>
<tr>
<td>410</td>
<td>Hs.596726 TMEM106C</td>
<td>12q13.1</td>
<td>79022</td>
<td></td>
</tr>
<tr>
<td>411</td>
<td>Hs.240170 OBFC2B</td>
<td>12q13.2</td>
<td>79035</td>
<td></td>
</tr>
<tr>
<td>412</td>
<td>Hs.632191 XTP3TPA</td>
<td>16p11.2</td>
<td>79077</td>
<td></td>
</tr>
<tr>
<td>413</td>
<td>Hs.79625 C20orf149</td>
<td>20q13.33</td>
<td>79144</td>
<td></td>
</tr>
<tr>
<td>414</td>
<td>Hs.465374 EFHD2</td>
<td>1p36.21</td>
<td>79180</td>
<td></td>
</tr>
<tr>
<td>415</td>
<td>Hs.412304 THOC6</td>
<td>16p13.3</td>
<td>79228</td>
<td></td>
</tr>
<tr>
<td>416</td>
<td>Hs.661128 KREME2</td>
<td>16p13.3</td>
<td>79412</td>
<td></td>
</tr>
<tr>
<td>417</td>
<td>Hs.521168 GCC1</td>
<td>7q32.1</td>
<td>79571</td>
<td></td>
</tr>
<tr>
<td>418</td>
<td>Hs.418233 MRPL24</td>
<td>1q21-q22</td>
<td>79590</td>
<td></td>
</tr>
<tr>
<td>419</td>
<td>Hs.371642 ADIPOR2</td>
<td>12p13.31</td>
<td>79602</td>
<td></td>
</tr>
<tr>
<td>420</td>
<td>Hs.59425 NOL9</td>
<td>1p36.31</td>
<td>79707</td>
<td></td>
</tr>
<tr>
<td>421</td>
<td>Hs.411865 IPO4</td>
<td>14q12</td>
<td>79711</td>
<td></td>
</tr>
<tr>
<td>422</td>
<td>Hs.163754 ZNF669</td>
<td>1q44</td>
<td>79862</td>
<td></td>
</tr>
<tr>
<td>423</td>
<td>Hs.496501 CXorf34</td>
<td>Xq22.1</td>
<td>79979</td>
<td></td>
</tr>
<tr>
<td>424</td>
<td>Hs.302051 MYO19</td>
<td>17q12</td>
<td>80179</td>
<td></td>
</tr>
<tr>
<td>425</td>
<td>Hs.567594 CXXC6</td>
<td>10q21</td>
<td>80312</td>
<td></td>
</tr>
<tr>
<td>426</td>
<td>Hs.440899 TTYH3</td>
<td>7p22</td>
<td>80727</td>
<td></td>
</tr>
<tr>
<td>427</td>
<td>Hs.534492 PRR7</td>
<td>5q35.3</td>
<td>80758</td>
<td></td>
</tr>
<tr>
<td>428</td>
<td>Hs.458390 INTS5</td>
<td>11q12.3</td>
<td>80789</td>
<td></td>
</tr>
<tr>
<td>429</td>
<td>Hs.373741 HM13</td>
<td>20q11.21</td>
<td>81502</td>
<td></td>
</tr>
<tr>
<td>430</td>
<td>Hs.150837 TXNDC5</td>
<td>6p24.3</td>
<td>81567</td>
<td></td>
</tr>
<tr>
<td>431</td>
<td>Hs.495229 URM1</td>
<td>9q34.11</td>
<td>81605</td>
<td></td>
</tr>
<tr>
<td>432</td>
<td>Hs.656466 ANP32E</td>
<td>1q21.2</td>
<td>81611</td>
<td></td>
</tr>
<tr>
<td>433</td>
<td>Hs.122908 CDT1</td>
<td>16q24.3</td>
<td>81620</td>
<td></td>
</tr>
<tr>
<td>434</td>
<td>Hs.444046 NETO2</td>
<td>16q11</td>
<td>81831</td>
<td></td>
</tr>
<tr>
<td>435</td>
<td>Hs.374421 CEP78</td>
<td>9q21.2</td>
<td>84131</td>
<td></td>
</tr>
<tr>
<td>436</td>
<td>Hs.643537 TRAF7</td>
<td>16p13.3</td>
<td>84231</td>
<td></td>
</tr>
<tr>
<td>437</td>
<td>Hs.19673 MAF1</td>
<td>8q24.3</td>
<td>84232</td>
<td></td>
</tr>
<tr>
<td>438</td>
<td>Hs.124015 HAGHL</td>
<td>16p13.3</td>
<td>84264</td>
<td></td>
</tr>
<tr>
<td>439</td>
<td>Hs.76662 ZDHHC16</td>
<td>10q24.1</td>
<td>84287</td>
<td></td>
</tr>
<tr>
<td>440</td>
<td>Hs.462913 MRPL45</td>
<td>17q21.2</td>
<td>84311</td>
<td></td>
</tr>
<tr>
<td>441</td>
<td>Hs.548868 THOC3</td>
<td>5q35.2</td>
<td>84321</td>
<td></td>
</tr>
<tr>
<td>442</td>
<td>Hs.631633 ELOF1</td>
<td>19p13.2</td>
<td>84337</td>
<td></td>
</tr>
<tr>
<td>443</td>
<td>Hs.631506 MCM8</td>
<td>20p12.3</td>
<td>84515</td>
<td></td>
</tr>
<tr>
<td>444</td>
<td>Hs.405925 PSRC1</td>
<td>1p13.3</td>
<td>84722</td>
<td></td>
</tr>
<tr>
<td>445</td>
<td>Hs.538286 LMNB2</td>
<td>9q34.11</td>
<td>84885</td>
<td></td>
</tr>
<tr>
<td>446</td>
<td>Hs.133122 ZDHHC12</td>
<td>12p11.1</td>
<td>84920</td>
<td></td>
</tr>
<tr>
<td>447</td>
<td>Hs.102971 ALG10</td>
<td>15q24.1</td>
<td>84948</td>
<td></td>
</tr>
<tr>
<td>448</td>
<td>Hs.71574 TIGD5</td>
<td>8q24.3</td>
<td>84993</td>
<td></td>
</tr>
<tr>
<td>449</td>
<td>Hs.334713 UBL7</td>
<td>17q25.1</td>
<td>85451</td>
<td></td>
</tr>
<tr>
<td>450</td>
<td>Hs.655493 UNK</td>
<td>17q25.1</td>
<td>85451</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>451</td>
<td>Hs.495240</td>
<td>WDR34</td>
<td>89891</td>
<td>9q34.11</td>
</tr>
<tr>
<td>452</td>
<td>Hs.19322</td>
<td>C9orf1140</td>
<td>89958</td>
<td>9q34.3</td>
</tr>
<tr>
<td>453</td>
<td>Hs.533597</td>
<td>PYGO2</td>
<td>90780</td>
<td>1q21.3</td>
</tr>
<tr>
<td>454</td>
<td>Hs.696333</td>
<td>LOC90784</td>
<td>90784</td>
<td>2p11.2</td>
</tr>
<tr>
<td>455</td>
<td>Hs.708161</td>
<td>C17orf72</td>
<td>92340</td>
<td>17q24.2</td>
</tr>
<tr>
<td>456</td>
<td>Hs.590956</td>
<td>TIMM50</td>
<td>92609</td>
<td>19q13.2</td>
</tr>
<tr>
<td>457</td>
<td>Hs.638652</td>
<td>LOC92755</td>
<td>92755</td>
<td>8p12</td>
</tr>
<tr>
<td>458</td>
<td>Hs.531614</td>
<td>BTBD14B</td>
<td>112939</td>
<td>19p13.13</td>
</tr>
<tr>
<td>459</td>
<td>Hs.101742</td>
<td>RPUSD1</td>
<td>113000</td>
<td>16p13.3</td>
</tr>
<tr>
<td>460</td>
<td>Hs.380094</td>
<td>PLCD3</td>
<td>113026</td>
<td>17q21.31</td>
</tr>
<tr>
<td>461</td>
<td>Hs.434886</td>
<td>CDA5</td>
<td>113130</td>
<td>11q12.1</td>
</tr>
<tr>
<td>462</td>
<td>Hs.591998</td>
<td>SAAL1</td>
<td>113174</td>
<td>11p15.1</td>
</tr>
<tr>
<td>463</td>
<td>Hs.534521</td>
<td>TMEM54</td>
<td>113452</td>
<td>1p35-p34</td>
</tr>
<tr>
<td>464</td>
<td>Hs.406840</td>
<td>SLC35A4</td>
<td>113829</td>
<td>5q31.3</td>
</tr>
<tr>
<td>465</td>
<td>Hs.201083</td>
<td>MAL2</td>
<td>114569</td>
<td>8q23</td>
</tr>
<tr>
<td>466</td>
<td>Hs.347524</td>
<td>C16orf75</td>
<td>116028</td>
<td>16p13.13</td>
</tr>
<tr>
<td>467</td>
<td>Hs.705716</td>
<td>TLCD1</td>
<td>116238</td>
<td>17q11.2</td>
</tr>
<tr>
<td>468</td>
<td>Hs.368934</td>
<td>C17orf45</td>
<td>125144</td>
<td>17p11.2</td>
</tr>
<tr>
<td>469</td>
<td>Hs.708197</td>
<td>PODN</td>
<td>127435</td>
<td>1p32.3</td>
</tr>
<tr>
<td>470</td>
<td>Hs.416375</td>
<td>E2F7</td>
<td>144455</td>
<td>12q21.2</td>
</tr>
<tr>
<td>471</td>
<td>Hs.135094</td>
<td>LOC146909</td>
<td>146909</td>
<td>17q21.31</td>
</tr>
<tr>
<td>472</td>
<td>Hs.632255</td>
<td>RUNDC1</td>
<td>146923</td>
<td>17q21.31</td>
</tr>
<tr>
<td>473</td>
<td>Hs.631760</td>
<td>DHRS13</td>
<td>147015</td>
<td>17q11.2</td>
</tr>
<tr>
<td>474</td>
<td>Hs.164324</td>
<td>TRIM16L</td>
<td>147166</td>
<td>17p11.2</td>
</tr>
<tr>
<td>475</td>
<td>Hs.381225</td>
<td>SPC24</td>
<td>147841</td>
<td>19p13.2</td>
</tr>
<tr>
<td>476</td>
<td>Hs.105153</td>
<td>SGOL1</td>
<td>151648</td>
<td>3p24.3</td>
</tr>
<tr>
<td>477</td>
<td>Hs.377830</td>
<td>MBOAT1</td>
<td>154141</td>
<td>6p22.3</td>
</tr>
<tr>
<td>478</td>
<td>Hs.567739</td>
<td>LOC158345</td>
<td>158345</td>
<td>9p24.1</td>
</tr>
<tr>
<td>479</td>
<td>Hs.657472</td>
<td>GPR180</td>
<td>160897</td>
<td>13q32.1</td>
</tr>
<tr>
<td>480</td>
<td>Hs.406461</td>
<td>FAM86A</td>
<td>196483</td>
<td>16p13.3</td>
</tr>
<tr>
<td>481</td>
<td>Hs.189823</td>
<td>TRIM65</td>
<td>201292</td>
<td>17q25.1</td>
</tr>
<tr>
<td>482</td>
<td>Hs.380920</td>
<td>LOC201725</td>
<td>201725</td>
<td>4q32.1</td>
</tr>
<tr>
<td>483</td>
<td>Hs.636480</td>
<td>TUBB</td>
<td>203068</td>
<td>6p21.33</td>
</tr>
<tr>
<td>484</td>
<td>Hs.706772</td>
<td>TUBB</td>
<td>203068</td>
<td>6p21.33</td>
</tr>
<tr>
<td>485</td>
<td>Hs.533655</td>
<td>TYSD1</td>
<td>219743</td>
<td>10q22.1</td>
</tr>
<tr>
<td>486</td>
<td>Hs.165607</td>
<td>C11orf82</td>
<td>220042</td>
<td>11q14.1</td>
</tr>
<tr>
<td>487</td>
<td>Hs.448226</td>
<td>RPLP0-like</td>
<td>220717</td>
<td>2p22.1</td>
</tr>
<tr>
<td>488</td>
<td>Hs.706964</td>
<td>RPLP0-like</td>
<td>220717</td>
<td>2p22.1</td>
</tr>
<tr>
<td>489</td>
<td>Hs.88523</td>
<td>C13orf3</td>
<td>221150</td>
<td>13q12.11</td>
</tr>
<tr>
<td>490</td>
<td>Hs.72363</td>
<td>C14orf80</td>
<td>283643</td>
<td>14q32.33</td>
</tr>
<tr>
<td>491</td>
<td>Hs.633835</td>
<td>LOC284072</td>
<td>284072</td>
<td>17q25.1</td>
</tr>
<tr>
<td>492</td>
<td>Hs.378885</td>
<td>PIGW</td>
<td>284098</td>
<td>17q12</td>
</tr>
<tr>
<td>493</td>
<td>Hs.381204</td>
<td>METTL2A</td>
<td>339175</td>
<td>17q23.2</td>
</tr>
<tr>
<td>494</td>
<td>Hs.103939</td>
<td>C1orf174</td>
<td>339448</td>
<td>1p36.32</td>
</tr>
<tr>
<td>495</td>
<td>Hs.459311</td>
<td>ZNF710</td>
<td>374655</td>
<td>15q26.1</td>
</tr>
<tr>
<td>496</td>
<td>Hs.512492</td>
<td>RAB15</td>
<td>376267</td>
<td>14q23.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>---</td>
<td>----------------</td>
<td>---</td>
<td>----------------</td>
</tr>
<tr>
<td>497</td>
<td>Hs.500561</td>
<td>LOC387703</td>
<td>387703</td>
<td>10q23.33</td>
</tr>
<tr>
<td>498</td>
<td>Hs.433791</td>
<td>TMEM46</td>
<td>387914</td>
<td>13q12.13</td>
</tr>
<tr>
<td>499</td>
<td>Hs.560655</td>
<td>LOC388524</td>
<td>388524</td>
<td>19p12</td>
</tr>
<tr>
<td>500</td>
<td>Hs.647251</td>
<td>LOC390183</td>
<td>390183</td>
<td>11q12.1</td>
</tr>
<tr>
<td>501</td>
<td>Hs.355210</td>
<td>LOC400019</td>
<td>400019</td>
<td>12p11.21</td>
</tr>
<tr>
<td>502</td>
<td>Hs.620821</td>
<td>VMAC</td>
<td>400673</td>
<td>19p13.3</td>
</tr>
<tr>
<td>503</td>
<td>Hs.516290</td>
<td>LOC400963</td>
<td>400963</td>
<td>2p11.2</td>
</tr>
<tr>
<td>504</td>
<td>Hs.590999</td>
<td>RPL23AP2</td>
<td>401904</td>
<td>19p13.12</td>
</tr>
<tr>
<td>505</td>
<td>Hs.73105</td>
<td>PMS2CL</td>
<td>441194</td>
<td>7p22.1</td>
</tr>
<tr>
<td>506</td>
<td>Hs.536395</td>
<td>DUXAP10</td>
<td>503639</td>
<td>14q11.2</td>
</tr>
<tr>
<td>507</td>
<td>Hs.523097</td>
<td>EIF5AL3</td>
<td>642592</td>
<td>10q22.3</td>
</tr>
<tr>
<td>508</td>
<td>Hs.645558</td>
<td>LOC642784</td>
<td>642784</td>
<td>Xq21.31</td>
</tr>
<tr>
<td>509</td>
<td>Hs.646886</td>
<td>LOC642909</td>
<td>642909</td>
<td>5q21.1</td>
</tr>
<tr>
<td>510</td>
<td>Hs.507343</td>
<td>LOC643446</td>
<td>643446</td>
<td>4p15.33</td>
</tr>
<tr>
<td>511</td>
<td>Hs.647694</td>
<td>LOC643586</td>
<td>643586</td>
<td>1p13.2</td>
</tr>
<tr>
<td>512</td>
<td>Hs.531200</td>
<td>HMGN2P6</td>
<td>643872</td>
<td>14q12</td>
</tr>
<tr>
<td>513</td>
<td>Hs.614453</td>
<td>LOC643873</td>
<td>643873</td>
<td>Xq23</td>
</tr>
<tr>
<td>514</td>
<td>Hs.647368</td>
<td>LOC644035</td>
<td>644035</td>
<td>16q22.1</td>
</tr>
<tr>
<td>515</td>
<td>Hs.632240</td>
<td>FAM83G</td>
<td>644815</td>
<td>17p11.2</td>
</tr>
<tr>
<td>516</td>
<td>Hs.632598</td>
<td>LOC645018</td>
<td>645018</td>
<td>4q31.21</td>
</tr>
<tr>
<td>517</td>
<td>Hs.632537</td>
<td>LOC645691</td>
<td>645691</td>
<td>2q33.1</td>
</tr>
<tr>
<td>518</td>
<td>Hs.647919</td>
<td>LOC646612</td>
<td>646612</td>
<td>3q22.3</td>
</tr>
<tr>
<td>519</td>
<td>Hs.571791</td>
<td>LOC647150</td>
<td>647150</td>
<td>1q31.2</td>
</tr>
<tr>
<td>520</td>
<td>Hs.289232</td>
<td>LOC648232</td>
<td>648232</td>
<td>NA</td>
</tr>
<tr>
<td>521</td>
<td>Hs.654748</td>
<td>TMEM183B</td>
<td>653659</td>
<td>3q25.1</td>
</tr>
<tr>
<td>522</td>
<td>Hs.444467</td>
<td>LOC654007</td>
<td>654007</td>
<td>7q32.3</td>
</tr>
<tr>
<td>523</td>
<td>Hs.646673</td>
<td>LOC728301</td>
<td>728301</td>
<td>5q22.3</td>
</tr>
<tr>
<td>524</td>
<td>Hs.535769</td>
<td>LOC728554</td>
<td>728554</td>
<td>5q35.3</td>
</tr>
<tr>
<td>525</td>
<td>Hs.594117</td>
<td>HMGN2P3</td>
<td>728632</td>
<td>16p12.1</td>
</tr>
<tr>
<td>526</td>
<td>Hs.596312</td>
<td>hCG_1988300</td>
<td>728638</td>
<td>8q12.3</td>
</tr>
<tr>
<td>527</td>
<td>Hs.535464</td>
<td>EIF3CL</td>
<td>728689</td>
<td>16p11.2</td>
</tr>
<tr>
<td>528</td>
<td>Hs.512314</td>
<td>LOC728891</td>
<td>728891</td>
<td>NA</td>
</tr>
<tr>
<td>529</td>
<td>Hs.652172</td>
<td>LOC729859</td>
<td>729859</td>
<td>2p11.2</td>
</tr>
</tbody>
</table>
Appendix H

Protein-coding genes extracted using the EST abundance analysis of the Human Unigene data as potential biomarkers of normal tissues (Table 5).

<table>
<thead>
<tr>
<th>S.No</th>
<th>Unigene_ID</th>
<th>Gene Symbol</th>
<th>NCBI Gene_ID</th>
<th>Genomic loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hs.506908</td>
<td>AADAC</td>
<td>13</td>
<td>3q21.3-q25.2</td>
</tr>
<tr>
<td>2</td>
<td>Hs.647097</td>
<td>ABP1</td>
<td>26</td>
<td>7q34-q36</td>
</tr>
<tr>
<td>3</td>
<td>Hs.445040</td>
<td>ACADM</td>
<td>34</td>
<td>1p31</td>
</tr>
<tr>
<td>4</td>
<td>Hs.532492</td>
<td>ACP2</td>
<td>53</td>
<td>11p11.2-p11.11</td>
</tr>
<tr>
<td>5</td>
<td>Hs.498178</td>
<td>ACTN2</td>
<td>88</td>
<td>1q42-q43</td>
</tr>
<tr>
<td>6</td>
<td>Hs.591026</td>
<td>ACVRL1</td>
<td>94</td>
<td>12q11-q14</td>
</tr>
<tr>
<td>7</td>
<td>Hs.474018</td>
<td>ADARB1</td>
<td>104</td>
<td>21q22.3</td>
</tr>
<tr>
<td>8</td>
<td>Hs.481545</td>
<td>ADCY2</td>
<td>108</td>
<td>5p15.3</td>
</tr>
<tr>
<td>9</td>
<td>Hs.593293</td>
<td>ADCY5</td>
<td>111</td>
<td>3q13.2-q21</td>
</tr>
<tr>
<td>10</td>
<td>Hs.4</td>
<td>ADH1B</td>
<td>125</td>
<td>4q21-q23</td>
</tr>
<tr>
<td>11</td>
<td>Hs.654537</td>
<td>ADH1C</td>
<td>126</td>
<td>4q21-q23</td>
</tr>
<tr>
<td>12</td>
<td>Hs.197029</td>
<td>ADORA2A</td>
<td>135</td>
<td>22q11.23</td>
</tr>
<tr>
<td>13</td>
<td>Hs.249159</td>
<td>ADRA2A</td>
<td>150</td>
<td>10q24-q26</td>
</tr>
<tr>
<td>14</td>
<td>Hs.522666</td>
<td>ALAS2</td>
<td>212</td>
<td>Xp11.21</td>
</tr>
<tr>
<td>15</td>
<td>Hs.76392</td>
<td>ALDH1A1</td>
<td>216</td>
<td>9q21.13</td>
</tr>
<tr>
<td>16</td>
<td>Hs.2533</td>
<td>ALDH9A1</td>
<td>223</td>
<td>1q23.1</td>
</tr>
<tr>
<td>17</td>
<td>Hs.111256</td>
<td>ALOX15B</td>
<td>247</td>
<td>17p13.1</td>
</tr>
<tr>
<td>18</td>
<td>Hs.102</td>
<td>AMT</td>
<td>275</td>
<td>3p21.2-p21.1</td>
</tr>
<tr>
<td>19</td>
<td>Hs.283749</td>
<td>ANG</td>
<td>283</td>
<td>14q11.1-q11.2</td>
</tr>
<tr>
<td>20</td>
<td>Hs.620557</td>
<td>ANK2</td>
<td>287</td>
<td>4q25-q27</td>
</tr>
<tr>
<td>21</td>
<td>Hs.1239</td>
<td>ANPEP</td>
<td>290</td>
<td>15q25-q26</td>
</tr>
<tr>
<td>22</td>
<td>Hs.422986</td>
<td>ANXA4</td>
<td>307</td>
<td>2p13</td>
</tr>
<tr>
<td>23</td>
<td>Hs.412117</td>
<td>ANXA6</td>
<td>309</td>
<td>5q32-q34</td>
</tr>
<tr>
<td>24</td>
<td>Hs.406238</td>
<td>AOX1</td>
<td>316</td>
<td>2q33</td>
</tr>
<tr>
<td>25</td>
<td>Hs.158932</td>
<td>APC</td>
<td>324</td>
<td>5q21-q22</td>
</tr>
<tr>
<td>26</td>
<td>Hs.244139</td>
<td>FAS</td>
<td>355</td>
<td>10q24.1</td>
</tr>
<tr>
<td>27</td>
<td>Hs.76152</td>
<td>AQP1</td>
<td>358</td>
<td>7p14</td>
</tr>
<tr>
<td>28</td>
<td>Hs.455323</td>
<td>AQP7</td>
<td>364</td>
<td>9p13</td>
</tr>
<tr>
<td>29</td>
<td>Hs.502876</td>
<td>RHOB</td>
<td>388</td>
<td>2p24</td>
</tr>
<tr>
<td>30</td>
<td>Hs.6838</td>
<td>RND3</td>
<td>390</td>
<td>2q23.3</td>
</tr>
<tr>
<td>31</td>
<td>Hs.503284</td>
<td>ARRB1</td>
<td>408</td>
<td>11q13</td>
</tr>
<tr>
<td>32</td>
<td>Hs.88251</td>
<td>ARSA</td>
<td>410</td>
<td>22q13.31-pter</td>
</tr>
<tr>
<td>33</td>
<td>Hs.24976</td>
<td>ART3</td>
<td>419</td>
<td>4p15.1-p14</td>
</tr>
<tr>
<td>34</td>
<td>Hs.460</td>
<td>ATF3</td>
<td>467</td>
<td>1q32.3</td>
</tr>
<tr>
<td>35</td>
<td>Hs.343522</td>
<td>ATP2B4</td>
<td>493</td>
<td>1q32.1</td>
</tr>
<tr>
<td></td>
<td>Symbol</td>
<td>Chromosome</td>
<td>Gene Region</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Hs.492280</td>
<td>ATP7B</td>
<td>540</td>
<td>13q14.3</td>
</tr>
<tr>
<td>37</td>
<td>Hs.33738</td>
<td>BBS2</td>
<td>583</td>
<td>16q21</td>
</tr>
<tr>
<td>38</td>
<td>Hs.410026</td>
<td>BCL2L2</td>
<td>599</td>
<td>14q11.2-q12</td>
</tr>
<tr>
<td>39</td>
<td>Hs.821</td>
<td>BGN</td>
<td>633</td>
<td>Xq28</td>
</tr>
<tr>
<td>40</td>
<td>Hs.169998</td>
<td>BST1</td>
<td>683</td>
<td>4p15</td>
</tr>
<tr>
<td>41</td>
<td>Hs.150557</td>
<td>KLF9</td>
<td>687</td>
<td>9q13</td>
</tr>
<tr>
<td>42</td>
<td>Hs.384598</td>
<td>SERPING1</td>
<td>710</td>
<td>11q12-q13.1</td>
</tr>
<tr>
<td>43</td>
<td>Hs.8986</td>
<td>C1QB</td>
<td>713</td>
<td>1p36.12</td>
</tr>
<tr>
<td>44</td>
<td>Hs.458355</td>
<td>C1S</td>
<td>716</td>
<td>12p13</td>
</tr>
<tr>
<td>45</td>
<td>Hs.591148</td>
<td>C3AR1</td>
<td>719</td>
<td>12p13.31</td>
</tr>
<tr>
<td>46</td>
<td>Hs.78065</td>
<td>C7</td>
<td>730</td>
<td>5p13</td>
</tr>
<tr>
<td>47</td>
<td>Hs.149363</td>
<td>C1orf1</td>
<td>753</td>
<td>18p11.2</td>
</tr>
<tr>
<td>48</td>
<td>Hs.155097</td>
<td>CA2</td>
<td>760</td>
<td>8q22</td>
</tr>
<tr>
<td>49</td>
<td>Hs.82129</td>
<td>CA3</td>
<td>761</td>
<td>8q13-q22</td>
</tr>
<tr>
<td>50</td>
<td>Hs.59093</td>
<td>CACNB2</td>
<td>783</td>
<td>10p12</td>
</tr>
<tr>
<td>51</td>
<td>Hs.440961</td>
<td>CAST</td>
<td>831</td>
<td>5q15</td>
</tr>
<tr>
<td>52</td>
<td>Hs.458426</td>
<td>CCK</td>
<td>885</td>
<td>3p22-p21.3</td>
</tr>
<tr>
<td>53</td>
<td>Hs.292524</td>
<td>CCNH</td>
<td>902</td>
<td>5q13.3-q14</td>
</tr>
<tr>
<td>54</td>
<td>Hs.163867</td>
<td>CD14</td>
<td>929</td>
<td>5q22-q32</td>
</tr>
<tr>
<td>55</td>
<td>Hs.374909</td>
<td>CD34</td>
<td>947</td>
<td>1q32</td>
</tr>
<tr>
<td>56</td>
<td>Hs.120949</td>
<td>CD36</td>
<td>948</td>
<td>7q11.2</td>
</tr>
<tr>
<td>57</td>
<td>Hs.633085</td>
<td>CD36</td>
<td>948</td>
<td>7q11.2</td>
</tr>
<tr>
<td>58</td>
<td>Hs.278573</td>
<td>CD59</td>
<td>966</td>
<td>11p13</td>
</tr>
<tr>
<td>59</td>
<td>Hs.191346</td>
<td>7-Sep</td>
<td>989</td>
<td>7p14.3-p14.1</td>
</tr>
<tr>
<td>60</td>
<td>Hs.690198</td>
<td>CDC42</td>
<td>998</td>
<td>1p36.1</td>
</tr>
<tr>
<td>61</td>
<td>Hs.76206</td>
<td>CDH5</td>
<td>1003</td>
<td>16q22.1</td>
</tr>
<tr>
<td>62</td>
<td>Hs.238990</td>
<td>CDKN1B</td>
<td>1027</td>
<td>12p13.1-p12</td>
</tr>
<tr>
<td>63</td>
<td>Hs.106070</td>
<td>CDKN1C</td>
<td>1028</td>
<td>11p15.5</td>
</tr>
<tr>
<td>64</td>
<td>Hs.442378</td>
<td>CDO1</td>
<td>1036</td>
<td>5q22-q23</td>
</tr>
<tr>
<td>65</td>
<td>Hs.517106</td>
<td>CEBPB</td>
<td>1051</td>
<td>20q13.1</td>
</tr>
<tr>
<td>66</td>
<td>Hs.479867</td>
<td>CENPC1</td>
<td>1060</td>
<td>4q12-q13.3</td>
</tr>
<tr>
<td>67</td>
<td>Hs.657385</td>
<td>RCNTB2</td>
<td>1102</td>
<td>13q14.3</td>
</tr>
<tr>
<td>68</td>
<td>Hs.535891</td>
<td>CHRM2</td>
<td>1129</td>
<td>7q31-q35</td>
</tr>
<tr>
<td>69</td>
<td>Hs.334347</td>
<td>CKM</td>
<td>1158</td>
<td>19q13.2-q13.3</td>
</tr>
<tr>
<td>70</td>
<td>Hs.628393</td>
<td>CLN3</td>
<td>1201</td>
<td>16p12.1</td>
</tr>
<tr>
<td>71</td>
<td>Hs.30213</td>
<td>CLN5</td>
<td>1203</td>
<td>13q21.1-q32</td>
</tr>
<tr>
<td>72</td>
<td>Hs.465929</td>
<td>CNN1</td>
<td>1264</td>
<td>19p13.2-p13.1</td>
</tr>
<tr>
<td>73</td>
<td>Hs.368921</td>
<td>COL16A1</td>
<td>1307</td>
<td>1p35-p34</td>
</tr>
<tr>
<td>74</td>
<td>Hs.4055</td>
<td>KLF6</td>
<td>1316</td>
<td>10p15</td>
</tr>
<tr>
<td>75</td>
<td>Hs.432453</td>
<td>MAP3K8</td>
<td>1326</td>
<td>10p11.23</td>
</tr>
<tr>
<td>76</td>
<td>Hs.584750</td>
<td>CREB1</td>
<td>1385</td>
<td>2q34</td>
</tr>
<tr>
<td>77</td>
<td>Hs.200250</td>
<td>CREM</td>
<td>1390</td>
<td>10p11.21</td>
</tr>
<tr>
<td>78</td>
<td>Hs.115617</td>
<td>CRHBP</td>
<td>1393</td>
<td>5q11.2-q13.3</td>
</tr>
<tr>
<td>79</td>
<td>Hs.408767</td>
<td>CRYAB</td>
<td>1410</td>
<td>11q22.3-q23.1</td>
</tr>
<tr>
<td>80</td>
<td>Hs.592192</td>
<td>CSF2RB</td>
<td>1439</td>
<td>22q13.1</td>
</tr>
<tr>
<td>81</td>
<td>Hs.108080</td>
<td>CSRP1</td>
<td>1465</td>
<td>1q32</td>
</tr>
<tr>
<td></td>
<td>Locus</td>
<td>Description</td>
<td>Build</td>
<td>Band1</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>-------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>82</td>
<td>Hs.75262</td>
<td>CTSO</td>
<td>1519</td>
<td>4q31-q32</td>
</tr>
<tr>
<td>83</td>
<td>Hs.154654</td>
<td>CYP1B1</td>
<td>1545</td>
<td>2p21</td>
</tr>
<tr>
<td>84</td>
<td>Hs.516700</td>
<td>CYP27A1</td>
<td>1593</td>
<td>2q33-qter</td>
</tr>
<tr>
<td>85</td>
<td>Hs.481980</td>
<td>DAB2</td>
<td>1601</td>
<td>5p13</td>
</tr>
<tr>
<td>86</td>
<td>Hs.279806</td>
<td>DDX5</td>
<td>1655</td>
<td>17q21</td>
</tr>
<tr>
<td>87</td>
<td>Hs.594952</td>
<td>DES</td>
<td>1674</td>
<td>2q35</td>
</tr>
<tr>
<td>88</td>
<td>Hs.155597</td>
<td>CFD</td>
<td>1675</td>
<td>19p13.3</td>
</tr>
<tr>
<td>89</td>
<td>Hs.700572</td>
<td>CYB5R3</td>
<td>1727</td>
<td>2q13.31-qter</td>
</tr>
<tr>
<td>90</td>
<td>Hs.80552</td>
<td>DPT</td>
<td>1805</td>
<td>1q12-q23</td>
</tr>
<tr>
<td>91</td>
<td>Hs.173381</td>
<td>DPYSL2</td>
<td>1808</td>
<td>8p22-p21</td>
</tr>
<tr>
<td>92</td>
<td>Hs.522074</td>
<td>TSC22D3</td>
<td>1831</td>
<td>Xq22.3</td>
</tr>
<tr>
<td>93</td>
<td>Hs.117060</td>
<td>ECM2</td>
<td>1842</td>
<td>9q22.3</td>
</tr>
<tr>
<td>94</td>
<td>Hs.784</td>
<td>EBI2</td>
<td>1880</td>
<td>13q32.3</td>
</tr>
<tr>
<td>95</td>
<td>Hs.126667</td>
<td>EDG2</td>
<td>1902</td>
<td>9q31.3</td>
</tr>
<tr>
<td>96</td>
<td>Hs.183713</td>
<td>EDNRA</td>
<td>1909</td>
<td>4q31.23</td>
</tr>
<tr>
<td>97</td>
<td>Hs.82002</td>
<td>EDNRB</td>
<td>1910</td>
<td>13q22</td>
</tr>
<tr>
<td>98</td>
<td>Hs.647061</td>
<td>ELN</td>
<td>2006</td>
<td>7q11.23</td>
</tr>
<tr>
<td>99</td>
<td>Hs.76753</td>
<td>ENG</td>
<td>2022</td>
<td>9q33-9q34.1</td>
</tr>
<tr>
<td>100</td>
<td>Hs.253903</td>
<td>STOM</td>
<td>2040</td>
<td>9q34.1</td>
</tr>
<tr>
<td>101</td>
<td>Hs.473819</td>
<td>ERG</td>
<td>2078</td>
<td>2q22.3</td>
</tr>
<tr>
<td>102</td>
<td>Hs.155729</td>
<td>ETFDH</td>
<td>2110</td>
<td>4q32-q35</td>
</tr>
<tr>
<td>103</td>
<td>Hs.361463</td>
<td>F10</td>
<td>2159</td>
<td>13q34</td>
</tr>
<tr>
<td>104</td>
<td>Hs.591133</td>
<td>FBN1</td>
<td>2200</td>
<td>15q21.1</td>
</tr>
<tr>
<td>105</td>
<td>Hs.76224</td>
<td>EFEMP1</td>
<td>2202</td>
<td>2p16</td>
</tr>
<tr>
<td>106</td>
<td>Hs.58367</td>
<td>GPC4</td>
<td>2239</td>
<td>Xq26.1</td>
</tr>
<tr>
<td>107</td>
<td>Hs.7636</td>
<td>FES</td>
<td>2242</td>
<td>15q26.1</td>
</tr>
<tr>
<td>108</td>
<td>Hs.567268</td>
<td>FGF7</td>
<td>2252</td>
<td>15q15-q21.1</td>
</tr>
<tr>
<td>109</td>
<td>Hs.435369</td>
<td>FHL1</td>
<td>2273</td>
<td>Xq26</td>
</tr>
<tr>
<td>110</td>
<td>Hs.144912</td>
<td>FMO2</td>
<td>2327</td>
<td>1q23-q25</td>
</tr>
<tr>
<td>111</td>
<td>Hs.103183</td>
<td>FMR1</td>
<td>2332</td>
<td>Xq27.3</td>
</tr>
<tr>
<td>112</td>
<td>Hs.25647</td>
<td>FOS</td>
<td>2353</td>
<td>14q24.3</td>
</tr>
<tr>
<td>113</td>
<td>Hs.370858</td>
<td>FUCA1</td>
<td>2517</td>
<td>1p34</td>
</tr>
<tr>
<td>114</td>
<td>Hs.390567</td>
<td>FYN</td>
<td>2534</td>
<td>6q21</td>
</tr>
<tr>
<td>115</td>
<td>Hs.80720</td>
<td>GAB1</td>
<td>2549</td>
<td>4q31.21</td>
</tr>
<tr>
<td>116</td>
<td>Hs.75335</td>
<td>GATM</td>
<td>2628</td>
<td>15q21.1</td>
</tr>
<tr>
<td>117</td>
<td>Hs.62661</td>
<td>GBP1</td>
<td>2633</td>
<td>1p22.2</td>
</tr>
<tr>
<td>118</td>
<td>Hs.656774</td>
<td>GBP3</td>
<td>2635</td>
<td>1p22.2</td>
</tr>
<tr>
<td>119</td>
<td>Hs.2171</td>
<td>GDF10</td>
<td>2662</td>
<td>10q11.22</td>
</tr>
<tr>
<td>120</td>
<td>Hs.97469</td>
<td>GGTA1</td>
<td>2681</td>
<td>9q33.2-q34.11</td>
</tr>
<tr>
<td>121</td>
<td>Hs.437156</td>
<td>GGTLa1</td>
<td>2687</td>
<td>22q11.23</td>
</tr>
<tr>
<td>122</td>
<td>Hs.296310</td>
<td>GJA4</td>
<td>2701</td>
<td>1p35.1</td>
</tr>
<tr>
<td>123</td>
<td>Hs.83381</td>
<td>GNG11</td>
<td>2791</td>
<td>7q21</td>
</tr>
<tr>
<td>124</td>
<td>Hs.524418</td>
<td>GPD1</td>
<td>2819</td>
<td>12q12-q13</td>
</tr>
<tr>
<td>125</td>
<td>Hs.122926</td>
<td>NR3C1</td>
<td>2908</td>
<td>5q31.3</td>
</tr>
<tr>
<td>126</td>
<td>Hs.75652</td>
<td>GSTM5</td>
<td>2949</td>
<td>1p13.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>Hs.449630</td>
<td>HBA1</td>
<td>3039</td>
<td>16p13.3</td>
</tr>
<tr>
<td>128</td>
<td>Hs.654744</td>
<td>HBA2</td>
<td>3040</td>
<td>16p13.3</td>
</tr>
<tr>
<td>129</td>
<td>Hs.705371</td>
<td>HBG1</td>
<td>3047</td>
<td>11p15.5</td>
</tr>
<tr>
<td>130</td>
<td>Hs.302145</td>
<td>HBG2</td>
<td>3048</td>
<td>11p15.5</td>
</tr>
<tr>
<td>131</td>
<td>Hs.363396</td>
<td>CFH</td>
<td>3075</td>
<td>1q32</td>
</tr>
<tr>
<td>132</td>
<td>Hs.233325</td>
<td>HFE</td>
<td>3077</td>
<td>6p21.3</td>
</tr>
<tr>
<td>133</td>
<td>Hs.118651</td>
<td>HHEX</td>
<td>3087</td>
<td>10q23.33</td>
</tr>
<tr>
<td>134</td>
<td>Hs.196952</td>
<td>HLF</td>
<td>3131</td>
<td>17q22</td>
</tr>
<tr>
<td>135</td>
<td>Hs.524430</td>
<td>NR4A1</td>
<td>3164</td>
<td>12q13</td>
</tr>
<tr>
<td>136</td>
<td>Hs.632828</td>
<td>HNRPH2</td>
<td>3188</td>
<td>Xq22</td>
</tr>
<tr>
<td>137</td>
<td>Hs.436885</td>
<td>HRC</td>
<td>3270</td>
<td>19q13.3</td>
</tr>
<tr>
<td>138</td>
<td>Hs.195040</td>
<td>HSD11B1</td>
<td>3290</td>
<td>1q32-q41</td>
</tr>
<tr>
<td>139</td>
<td>Hs.406861</td>
<td>HSD17B4</td>
<td>3295</td>
<td>5q21</td>
</tr>
<tr>
<td>140</td>
<td>Hs.520028</td>
<td>HSPA1A</td>
<td>3303</td>
<td>6p21.3</td>
</tr>
<tr>
<td>141</td>
<td>Hs.97013</td>
<td>HSPB2</td>
<td>3316</td>
<td>11q22-q23</td>
</tr>
<tr>
<td>142</td>
<td>Hs.75619</td>
<td>HYAL1</td>
<td>3373</td>
<td>3p21.3-p21.2</td>
</tr>
<tr>
<td>143</td>
<td>Hs.654563</td>
<td>ICAM3</td>
<td>3385</td>
<td>19p13.3-p13.2</td>
</tr>
<tr>
<td>144</td>
<td>Hs.312485</td>
<td>CFI</td>
<td>3426</td>
<td>4q25</td>
</tr>
<tr>
<td>145</td>
<td>Hs.47338</td>
<td>IFIT3</td>
<td>3437</td>
<td>10q24</td>
</tr>
<tr>
<td>146</td>
<td>Hs.520414</td>
<td>IFNGR1</td>
<td>3459</td>
<td>6q23.3</td>
</tr>
<tr>
<td>147</td>
<td>Hs.8867</td>
<td>CYR61</td>
<td>3491</td>
<td>1p31-p22</td>
</tr>
<tr>
<td>148</td>
<td>Hs.632790</td>
<td>IL3RA</td>
<td>3563</td>
<td>Xp22.3 or Yp11.3</td>
</tr>
<tr>
<td>149</td>
<td>Hs.194778</td>
<td>IL8RA</td>
<td>3577</td>
<td>2q35</td>
</tr>
<tr>
<td>150</td>
<td>Hs.513022</td>
<td>ISLR</td>
<td>3671</td>
<td>15q23-q24</td>
</tr>
<tr>
<td>151</td>
<td>Hs.699822</td>
<td>ISLR</td>
<td>3671</td>
<td>15q23-q24</td>
</tr>
<tr>
<td>152</td>
<td>Hs.512235</td>
<td>ITPR2</td>
<td>3709</td>
<td>12p11</td>
</tr>
<tr>
<td>153</td>
<td>Hs.121495</td>
<td>KCNE1</td>
<td>3753</td>
<td>21q22.1-q22.2</td>
</tr>
<tr>
<td>154</td>
<td>Hs.591606</td>
<td>KCNJ3</td>
<td>3760</td>
<td>2q24.1</td>
</tr>
<tr>
<td>155</td>
<td>Hs.182971</td>
<td>KPN5</td>
<td>3841</td>
<td>6q22.2</td>
</tr>
<tr>
<td>156</td>
<td>Hs.444414</td>
<td>AFF3</td>
<td>3899</td>
<td>2q11.2-q12</td>
</tr>
<tr>
<td>157</td>
<td>Hs.572535</td>
<td>LAIR1</td>
<td>3903</td>
<td>19q13.4</td>
</tr>
<tr>
<td>158</td>
<td>Hs.133421</td>
<td>LIFR</td>
<td>3977</td>
<td>5p13-p12</td>
</tr>
<tr>
<td>159</td>
<td>Hs.438236</td>
<td>ABLIM1</td>
<td>3983</td>
<td>10q25</td>
</tr>
<tr>
<td>160</td>
<td>Hs.65436</td>
<td>LOXL1</td>
<td>4016</td>
<td>15q24-25</td>
</tr>
<tr>
<td>161</td>
<td>Hs.661130</td>
<td>LOXL2</td>
<td>4017</td>
<td>8p21.3-p21.2</td>
</tr>
<tr>
<td>162</td>
<td>Hs.1116</td>
<td>LTBR</td>
<td>4055</td>
<td>12p13</td>
</tr>
<tr>
<td>163</td>
<td>Hs.406475</td>
<td>LUM</td>
<td>4060</td>
<td>12q21.3-q22</td>
</tr>
<tr>
<td>164</td>
<td>Hs.656534</td>
<td>SMAD1</td>
<td>4086</td>
<td>4q31</td>
</tr>
<tr>
<td>165</td>
<td>Hs.465087</td>
<td>SMAD7</td>
<td>4092</td>
<td>18q21.1</td>
</tr>
<tr>
<td>166</td>
<td>Hs.446125</td>
<td>MAK</td>
<td>4117</td>
<td>6p24</td>
</tr>
<tr>
<td>167</td>
<td>Hs.102788</td>
<td>MAN1A1</td>
<td>4121</td>
<td>6q22</td>
</tr>
<tr>
<td>168</td>
<td>Hs.599039</td>
<td>MCAM</td>
<td>4162</td>
<td>11q23.3</td>
</tr>
<tr>
<td>169</td>
<td>Hs.387262</td>
<td>MCF2</td>
<td>4168</td>
<td>Xq27</td>
</tr>
<tr>
<td>170</td>
<td>Hs.6790</td>
<td>DNAJB9</td>
<td>4189</td>
<td>7q31</td>
</tr>
<tr>
<td>171</td>
<td>Hs.699175</td>
<td>MEF2C</td>
<td>4208</td>
<td>5q14</td>
</tr>
<tr>
<td>172</td>
<td>Hs.170355</td>
<td>ME0X2</td>
<td>4223</td>
<td>7p22.1-p21.3</td>
</tr>
<tr>
<td></td>
<td>167 Hs.61418</td>
<td>MFAP1</td>
<td>4236</td>
<td>15q15-q21</td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>---------</td>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td>174</td>
<td>Hs.432818</td>
<td>MFAP3</td>
<td>4236</td>
<td>5q32-q33.2</td>
</tr>
<tr>
<td>175</td>
<td>Hs.3745</td>
<td>MFGE8</td>
<td>4240</td>
<td>15q25</td>
</tr>
<tr>
<td>176</td>
<td>Hs.163924</td>
<td>NR3C2</td>
<td>4306</td>
<td>4q31.1</td>
</tr>
<tr>
<td>177</td>
<td>Hs.293970</td>
<td>ALDH6A1</td>
<td>4329</td>
<td>14q24.3</td>
</tr>
<tr>
<td>178</td>
<td>Hs.357128</td>
<td>MOCS1</td>
<td>4337</td>
<td>6p21.3</td>
</tr>
<tr>
<td>179</td>
<td>Hs.79015</td>
<td>CD200</td>
<td>4345</td>
<td>3q12-q13</td>
</tr>
<tr>
<td>180</td>
<td>Hs.396566</td>
<td>MPP3</td>
<td>4356</td>
<td>17q21.31</td>
</tr>
<tr>
<td>181</td>
<td>Hs.371225</td>
<td>MSH5</td>
<td>4439</td>
<td>6p21.3</td>
</tr>
<tr>
<td>182</td>
<td>Hs.349110</td>
<td>MST1</td>
<td>4485</td>
<td>3p21</td>
</tr>
<tr>
<td>183</td>
<td>Hs.471991</td>
<td>MTF1</td>
<td>4520</td>
<td>1p33</td>
</tr>
<tr>
<td>184</td>
<td>Hs.498187</td>
<td>MTR</td>
<td>4548</td>
<td>1q43</td>
</tr>
<tr>
<td>185</td>
<td>Hs.654589</td>
<td>MYBPC1</td>
<td>4604</td>
<td>12q23.2</td>
</tr>
<tr>
<td>186</td>
<td>Hs.82116</td>
<td>MYD88</td>
<td>4615</td>
<td>3p22</td>
</tr>
<tr>
<td>187</td>
<td>Hs.440895</td>
<td>MYH3</td>
<td>4621</td>
<td>17p13.1</td>
</tr>
<tr>
<td>188</td>
<td>Hs.278432</td>
<td>MYH7</td>
<td>4625</td>
<td>14q12</td>
</tr>
<tr>
<td>189</td>
<td>Hs.517939</td>
<td>MYL3</td>
<td>4634</td>
<td>3p21.3-p21.2</td>
</tr>
<tr>
<td>190</td>
<td>Hs.463300</td>
<td>MYL4</td>
<td>4635</td>
<td>17q21-qter</td>
</tr>
<tr>
<td>191</td>
<td>Hs.556600</td>
<td>MYLK</td>
<td>4638</td>
<td>3q21</td>
</tr>
<tr>
<td>192</td>
<td>Hs.436037</td>
<td>MYOC</td>
<td>4653</td>
<td>1q23-q24</td>
</tr>
<tr>
<td>193</td>
<td>Hs.444403</td>
<td>PPP1R12B</td>
<td>4660</td>
<td>1q32.1</td>
</tr>
<tr>
<td>194</td>
<td>Hs.66180</td>
<td>NAP1L2</td>
<td>4674</td>
<td>Xq13</td>
</tr>
<tr>
<td>195</td>
<td>Hs.477693</td>
<td>NCK1</td>
<td>4690</td>
<td>3q21</td>
</tr>
<tr>
<td>196</td>
<td>Hs.522615</td>
<td>NDP</td>
<td>4693</td>
<td>Xp11.4</td>
</tr>
<tr>
<td>197</td>
<td>Hs.699288</td>
<td>NEDD9</td>
<td>4739</td>
<td>6p25-p24</td>
</tr>
<tr>
<td>198</td>
<td>Hs.191911</td>
<td>NFIA</td>
<td>4774</td>
<td>1p31.3-p31.2</td>
</tr>
<tr>
<td>199</td>
<td>Hs.77810</td>
<td>NFATC4</td>
<td>4776</td>
<td>14q11.2</td>
</tr>
<tr>
<td>200</td>
<td>Hs.656450</td>
<td>NINJ2</td>
<td>4815</td>
<td>12p13</td>
</tr>
<tr>
<td>201</td>
<td>Hs.529509</td>
<td>NKR</td>
<td>4820</td>
<td>3p23-p21</td>
</tr>
<tr>
<td>202</td>
<td>Hs.436100</td>
<td>NOTCH4</td>
<td>4855</td>
<td>6p21.3</td>
</tr>
<tr>
<td>203</td>
<td>Hs.529006</td>
<td>NPC1</td>
<td>4864</td>
<td>18q11-q12</td>
</tr>
<tr>
<td>204</td>
<td>Hs.237028</td>
<td>NPR3</td>
<td>4883</td>
<td>5p14-p13</td>
</tr>
<tr>
<td>205</td>
<td>Hs.268788</td>
<td>NRAP</td>
<td>4892</td>
<td>10q24-q26</td>
</tr>
<tr>
<td>206</td>
<td>Hs.410969</td>
<td>NTRK3</td>
<td>4916</td>
<td>15q25</td>
</tr>
<tr>
<td>207</td>
<td>Hs.563344</td>
<td>NR4A2</td>
<td>4929</td>
<td>2q22-q23</td>
</tr>
<tr>
<td>208</td>
<td>Hs.380271</td>
<td>OGG1</td>
<td>4968</td>
<td>3p26.2</td>
</tr>
<tr>
<td>209</td>
<td>Hs.31595</td>
<td>CLDN11</td>
<td>5010</td>
<td>3q26.2-q26.3</td>
</tr>
<tr>
<td>210</td>
<td>Hs.510334</td>
<td>SERPINA5</td>
<td>5104</td>
<td>14q32.1</td>
</tr>
<tr>
<td>211</td>
<td>Hs.680373</td>
<td>PDE1A</td>
<td>5136</td>
<td>2q32.1</td>
</tr>
<tr>
<td>212</td>
<td>Hs.530871</td>
<td>PDE1B</td>
<td>5153</td>
<td>12q13</td>
</tr>
<tr>
<td>213</td>
<td>Hs.256667</td>
<td>PDK2</td>
<td>5164</td>
<td>17q21.33</td>
</tr>
<tr>
<td>214</td>
<td>Hs.8364</td>
<td>PDK4</td>
<td>5166</td>
<td>7q21.3</td>
</tr>
<tr>
<td>215</td>
<td>Hs.190977</td>
<td>ENPP2</td>
<td>5168</td>
<td>8q24.1</td>
</tr>
<tr>
<td>216</td>
<td>Hs.532768</td>
<td>SERPINF1</td>
<td>5176</td>
<td>17p13.1</td>
</tr>
<tr>
<td>217</td>
<td>Hs.36473</td>
<td>PEPD</td>
<td>5184</td>
<td>19q12-q13.2</td>
</tr>
<tr>
<td>218</td>
<td>Hs.119316</td>
<td>PET112L</td>
<td>5188</td>
<td>4q27-q28</td>
</tr>
<tr>
<td>219</td>
<td>Hs.137415</td>
<td>VIT</td>
<td>5212</td>
<td>2p22-p21</td>
</tr>
<tr>
<td>220</td>
<td>Hs.307835</td>
<td>PGM5</td>
<td>5239</td>
<td>9q13</td>
</tr>
<tr>
<td>221</td>
<td>Hs.159628</td>
<td>SERPINA4</td>
<td>5267</td>
<td>14q31-q32.1</td>
</tr>
<tr>
<td>222</td>
<td>Hs.132225</td>
<td>PIK3R1</td>
<td>5295</td>
<td>5q13.1</td>
</tr>
<tr>
<td>223</td>
<td>Hs.99949</td>
<td>PIP</td>
<td>5304</td>
<td>7q34</td>
</tr>
<tr>
<td>224</td>
<td>Hs.75813</td>
<td>PKD1</td>
<td>5310</td>
<td>16p13.3</td>
</tr>
<tr>
<td>225</td>
<td>Hs.181272</td>
<td>PKD2</td>
<td>5311</td>
<td>4q21-q23</td>
</tr>
<tr>
<td>226</td>
<td>Hs.466804</td>
<td>PLA2G2A</td>
<td>5320</td>
<td>1p35</td>
</tr>
<tr>
<td>227</td>
<td>Hs.444975</td>
<td>PLAGL1</td>
<td>5325</td>
<td>6q24-q25</td>
</tr>
<tr>
<td>228</td>
<td>Hs.80776</td>
<td>PLG2D1</td>
<td>5333</td>
<td>3p22-p21.3</td>
</tr>
<tr>
<td>229</td>
<td>Hs.442498</td>
<td>FXYD1</td>
<td>5348</td>
<td>19q13.1</td>
</tr>
<tr>
<td>230</td>
<td>Hs.170839</td>
<td>PLN</td>
<td>5350</td>
<td>6q22.1</td>
</tr>
<tr>
<td>231</td>
<td>Hs.372031</td>
<td>PMP22</td>
<td>5376</td>
<td>17p12-p11.2</td>
</tr>
<tr>
<td>232</td>
<td>Hs.292996</td>
<td>PMS2L2</td>
<td>5380</td>
<td>7q11-q22</td>
</tr>
<tr>
<td>233</td>
<td>Hs.632368</td>
<td>EXOSC10</td>
<td>5394</td>
<td>1p36.22</td>
</tr>
<tr>
<td>234</td>
<td>Hs.702224</td>
<td>PRRX1</td>
<td>5396</td>
<td>1q24</td>
</tr>
<tr>
<td>235</td>
<td>Hs.287518</td>
<td>4-Sep</td>
<td>5414</td>
<td>17q22-q23</td>
</tr>
<tr>
<td>236</td>
<td>Hs.1897</td>
<td>POMC</td>
<td>5443</td>
<td>2p23.3</td>
</tr>
<tr>
<td>237</td>
<td>Hs.530077</td>
<td>PON2</td>
<td>5445</td>
<td>7q21.3</td>
</tr>
<tr>
<td>238</td>
<td>Hs.505662</td>
<td>PPP1R1A</td>
<td>5502</td>
<td>12q13.2</td>
</tr>
<tr>
<td>239</td>
<td>Hs.303090</td>
<td>PPP1R3C</td>
<td>5507</td>
<td>10q23-q24</td>
</tr>
<tr>
<td>240</td>
<td>Hs.467192</td>
<td>PPP2R1A</td>
<td>5518</td>
<td>19q13.33</td>
</tr>
<tr>
<td>241</td>
<td>Hs.280604</td>
<td>PPP3R1</td>
<td>5534</td>
<td>2p15</td>
</tr>
<tr>
<td>242</td>
<td>Hs.125503</td>
<td>MAPK10</td>
<td>5602</td>
<td>4q22.1-q23</td>
</tr>
<tr>
<td>243</td>
<td>Hs.632287</td>
<td>PRKY</td>
<td>5616</td>
<td>Yp11.2</td>
</tr>
<tr>
<td>244</td>
<td>Hs.89983</td>
<td>MASP1</td>
<td>5648</td>
<td>3q27-q28</td>
</tr>
<tr>
<td>245</td>
<td>Hs.154658</td>
<td>PSD</td>
<td>5662</td>
<td>10q24</td>
</tr>
<tr>
<td>246</td>
<td>Hs.458324</td>
<td>PTGIR</td>
<td>5739</td>
<td>19q13.3</td>
</tr>
<tr>
<td>247</td>
<td>Hs.154084</td>
<td>PYGM</td>
<td>5837</td>
<td>11q12-q13.2</td>
</tr>
<tr>
<td>248</td>
<td>Hs.377992</td>
<td>RABGGTA</td>
<td>5875</td>
<td>14q11.2</td>
</tr>
<tr>
<td>249</td>
<td>Hs.695926</td>
<td>RASA1</td>
<td>5921</td>
<td>5q13.3</td>
</tr>
<tr>
<td>250</td>
<td>Hs.591111</td>
<td>RASGRF1</td>
<td>5923</td>
<td>15q24</td>
</tr>
<tr>
<td>251</td>
<td>Hs.50223</td>
<td>RBP4</td>
<td>5950</td>
<td>10q23-q24</td>
</tr>
<tr>
<td>252</td>
<td>Hs.235069</td>
<td>RECQL</td>
<td>5965</td>
<td>12p12</td>
</tr>
<tr>
<td>253</td>
<td>Hs.78944</td>
<td>RGS2</td>
<td>5997</td>
<td>1q31</td>
</tr>
<tr>
<td>254</td>
<td>Hs.657266</td>
<td>RPL3L</td>
<td>6123</td>
<td>16p13.3</td>
</tr>
<tr>
<td>255</td>
<td>Hs.287749</td>
<td>SC5DL</td>
<td>6309</td>
<td>11q23.3</td>
</tr>
<tr>
<td>256</td>
<td>Hs.272493</td>
<td>CCL14</td>
<td>6358</td>
<td>17q11.2</td>
</tr>
<tr>
<td>257</td>
<td>Hs.57907</td>
<td>CCL21</td>
<td>6366</td>
<td>9p13</td>
</tr>
<tr>
<td>258</td>
<td>Hs.531668</td>
<td>CX3CL1</td>
<td>6376</td>
<td>16q13</td>
</tr>
<tr>
<td>259</td>
<td>Hs.598247</td>
<td>SDC2</td>
<td>6383</td>
<td>8q22-q23</td>
</tr>
<tr>
<td>260</td>
<td>Hs.522891</td>
<td>CXCL12</td>
<td>6387</td>
<td>10q11.1</td>
</tr>
<tr>
<td>261</td>
<td>Hs.275775</td>
<td>SEPP1</td>
<td>6414</td>
<td>5q31</td>
</tr>
<tr>
<td>262</td>
<td>Hs.309090</td>
<td>SFRS7</td>
<td>6432</td>
<td>2p22.1</td>
</tr>
<tr>
<td>263</td>
<td>Hs.591727</td>
<td>SGCD</td>
<td>6444</td>
<td>5q33-q34</td>
</tr>
<tr>
<td>264</td>
<td>Hs.380691</td>
<td>SLC2A4</td>
<td>6517</td>
<td>17p13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>265</td>
<td>Hs.530003</td>
<td>SLC2A5</td>
<td>6518</td>
<td>1p36.2</td>
</tr>
<tr>
<td>266</td>
<td>Hs.1964</td>
<td>SLC5A1</td>
<td>6523</td>
<td>22q13.1</td>
</tr>
<tr>
<td>267</td>
<td>Hs.149098</td>
<td>SMTN</td>
<td>6525</td>
<td>22q12.2</td>
</tr>
<tr>
<td>268</td>
<td>Hs.468274</td>
<td>SLC8A1</td>
<td>6546</td>
<td>2p23-p22</td>
</tr>
<tr>
<td>269</td>
<td>Hs.337696</td>
<td>SLC8A3</td>
<td>6547</td>
<td>14q24.1</td>
</tr>
<tr>
<td>270</td>
<td>Hs.518270</td>
<td>SLC2A1</td>
<td>6578</td>
<td>3q21</td>
</tr>
<tr>
<td>271</td>
<td>Hs.152292</td>
<td>SMARCA1</td>
<td>6594</td>
<td>Xq25</td>
</tr>
<tr>
<td>272</td>
<td>Hs.2420</td>
<td>SOD3</td>
<td>6649</td>
<td>4p15.3-p15.1</td>
</tr>
<tr>
<td>273</td>
<td>Hs.654397</td>
<td>SOS1</td>
<td>6654</td>
<td>2p22-p21</td>
</tr>
<tr>
<td>274</td>
<td>Hs.167535</td>
<td>SRP54</td>
<td>6729</td>
<td>14q13.2</td>
</tr>
<tr>
<td>275</td>
<td>Hs.117715</td>
<td>ST5</td>
<td>6764</td>
<td>11p15</td>
</tr>
<tr>
<td>276</td>
<td>Hs.80642</td>
<td>STAT4</td>
<td>6775</td>
<td>2q32.2-q32.3</td>
</tr>
<tr>
<td>277</td>
<td>Hs.437058</td>
<td>STAT5A</td>
<td>6776</td>
<td>17q11.2</td>
</tr>
<tr>
<td>278</td>
<td>Hs.25590</td>
<td>STC1</td>
<td>6781</td>
<td>8p21-p11.2</td>
</tr>
<tr>
<td>279</td>
<td>Hs.479898</td>
<td>SULT1E1</td>
<td>6783</td>
<td>4q13.1</td>
</tr>
<tr>
<td>280</td>
<td>Hs.558403</td>
<td>SUOX</td>
<td>6821</td>
<td>12q13.2</td>
</tr>
<tr>
<td>281</td>
<td>Hs.2563</td>
<td>TAC1</td>
<td>6863</td>
<td>7q21-q22</td>
</tr>
<tr>
<td>282</td>
<td>Hs.632099</td>
<td>TAGLN</td>
<td>6876</td>
<td>11q23.2</td>
</tr>
<tr>
<td>283</td>
<td>Hs.644653</td>
<td>TCF4</td>
<td>6925</td>
<td>18q21.1</td>
</tr>
<tr>
<td>284</td>
<td>Hs.124503</td>
<td>ZEB1</td>
<td>6935</td>
<td>10p11.2</td>
</tr>
<tr>
<td>285</td>
<td>Hs.446392</td>
<td>DYNLT3</td>
<td>6990</td>
<td>Xp21</td>
</tr>
<tr>
<td>286</td>
<td>Hs.89640</td>
<td>TEK</td>
<td>7010</td>
<td>9p21</td>
</tr>
<tr>
<td>287</td>
<td>Hs.592317</td>
<td>TGFBR3</td>
<td>7043</td>
<td>14q24</td>
</tr>
<tr>
<td>288</td>
<td>Hs.482390</td>
<td>TGFBR3</td>
<td>7049</td>
<td>1p33-p32</td>
</tr>
<tr>
<td>289</td>
<td>Hs.657724</td>
<td>TLR3</td>
<td>7098</td>
<td>4q35</td>
</tr>
<tr>
<td>290</td>
<td>Hs.174312</td>
<td>TLR4</td>
<td>7099</td>
<td>9q32-q33</td>
</tr>
<tr>
<td>291</td>
<td>Hs.267632</td>
<td>TMF1</td>
<td>7110</td>
<td>3p21-p12</td>
</tr>
<tr>
<td>292</td>
<td>Hs.494595</td>
<td>TMOD1</td>
<td>7111</td>
<td>9q22.3</td>
</tr>
<tr>
<td>293</td>
<td>Hs.505337</td>
<td>CLDN5</td>
<td>7122</td>
<td>22q11.21</td>
</tr>
<tr>
<td>294</td>
<td>Hs.182421</td>
<td>TNNT2</td>
<td>7125</td>
<td>20q12-q13.11</td>
</tr>
<tr>
<td>295</td>
<td>Hs.73454</td>
<td>TNNT3</td>
<td>7140</td>
<td>11p15.5</td>
</tr>
<tr>
<td>296</td>
<td>Hs.471381</td>
<td>TNS1</td>
<td>7145</td>
<td>2q35-q36</td>
</tr>
<tr>
<td>297</td>
<td>Hs.133892</td>
<td>TPM1</td>
<td>7168</td>
<td>15q22.1</td>
</tr>
<tr>
<td>298</td>
<td>Hs.108301</td>
<td>NR2C1</td>
<td>7181</td>
<td>12q22</td>
</tr>
<tr>
<td>299</td>
<td>Hs.159003</td>
<td>TRPC6</td>
<td>7225</td>
<td>11q21-q22</td>
</tr>
<tr>
<td>300</td>
<td>Hs.486292</td>
<td>TSPYL1</td>
<td>7259</td>
<td>6q22-q23</td>
</tr>
<tr>
<td>301</td>
<td>Hs.654592</td>
<td>TTN</td>
<td>7273</td>
<td>2q31</td>
</tr>
<tr>
<td>302</td>
<td>Hs.520348</td>
<td>UBC</td>
<td>7316</td>
<td>12q24.3</td>
</tr>
<tr>
<td>303</td>
<td>Hs.21899</td>
<td>SLC35A2</td>
<td>7355</td>
<td>Xp11.23-p11.22</td>
</tr>
<tr>
<td>304</td>
<td>Hs.516217</td>
<td>UGP2</td>
<td>7360</td>
<td>2p14-p13</td>
</tr>
<tr>
<td>305</td>
<td>Hs.409662</td>
<td>COL14A1</td>
<td>7373</td>
<td>8q23</td>
</tr>
<tr>
<td>306</td>
<td>Hs.133135</td>
<td>UTRN</td>
<td>7402</td>
<td>6q24</td>
</tr>
<tr>
<td>307</td>
<td>Hs.440848</td>
<td>VWF</td>
<td>7450</td>
<td>12p13.3</td>
</tr>
<tr>
<td>308</td>
<td>Hs.326420</td>
<td>WNT9B</td>
<td>7484</td>
<td>17q21</td>
</tr>
<tr>
<td>309</td>
<td>Hs.78919</td>
<td>XK</td>
<td>7504</td>
<td>Xp21.1</td>
</tr>
<tr>
<td>310</td>
<td>Hs.326801</td>
<td>ZNF711</td>
<td>7552</td>
<td>Xq21.1-q21.2</td>
</tr>
<tr>
<td>ID</td>
<td>Symbol</td>
<td>Chromosome</td>
<td>Band</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>311</td>
<td>ZNF155</td>
<td>19q13.2-q13.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>312</td>
<td>ZNF187</td>
<td>6p21.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>313</td>
<td>ZNF204</td>
<td>6p21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>314</td>
<td>ZFAND5</td>
<td>9q13-q21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>315</td>
<td>ZNF225</td>
<td>19q13.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>316</td>
<td>PRDM2</td>
<td>1p36.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>317</td>
<td>DNAL11</td>
<td>1p35.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>318</td>
<td>MFAP5</td>
<td>12p13.1-p12.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>319</td>
<td>SSPN</td>
<td>12p11.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>IFT88</td>
<td>13q12.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>321</td>
<td>HDHD1A</td>
<td>Xp22.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>322</td>
<td>JARID1D</td>
<td>Yq11-Yq11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>323</td>
<td>USP9Y</td>
<td>Yq11.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>324</td>
<td>FZD4</td>
<td>11q14.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>PIP5K1B</td>
<td>9q13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>326</td>
<td>SPARCL1</td>
<td>4q22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>327</td>
<td>CMAH</td>
<td>6p21.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>328</td>
<td>LTBP4</td>
<td>19q13.1-q13.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>329</td>
<td>RECK</td>
<td>9p13-p12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>330</td>
<td>TPST2</td>
<td>22q12.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>331</td>
<td>SORBS2</td>
<td>4q35.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>332</td>
<td>CILP</td>
<td>15q22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>333</td>
<td>ITGA10</td>
<td>1q21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>334</td>
<td>ITGA8</td>
<td>10p13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>335</td>
<td>CYP4F2</td>
<td>19pter-p13.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>336</td>
<td>CDC14B</td>
<td>9q22.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>337</td>
<td>TCAP</td>
<td>17q12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>338</td>
<td>MKNK1</td>
<td>1p33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>339</td>
<td>PLA2G4C</td>
<td>19q13.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>CDC2L5</td>
<td>7p13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>341</td>
<td>AOC3</td>
<td>17q21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>342</td>
<td>HYAL2</td>
<td>3p21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>343</td>
<td>B3GALT4</td>
<td>6p21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>344</td>
<td>CD164</td>
<td>6q21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>345</td>
<td>SNAP23</td>
<td>15q15.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>346</td>
<td>SUCLA2</td>
<td>13q12.2-q13.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>347</td>
<td>WDR22</td>
<td>14q23-q24.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>348</td>
<td>CFLAR</td>
<td>2q33-q34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>349</td>
<td>STK19</td>
<td>6p21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>350</td>
<td>PER2</td>
<td>2q37.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>351</td>
<td>STBD1</td>
<td>4q24-q25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>352</td>
<td>SELENBP1</td>
<td>1q21-q22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>353</td>
<td>CH25H</td>
<td>10q23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>354</td>
<td>DOK2</td>
<td>8p21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>PAPSS2</td>
<td>10q23-q24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>356</td>
<td>CLDN10</td>
<td>13q31-q34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>357</td>
<td>Hs.23748</td>
<td>LDB2</td>
<td>9079</td>
<td>4p16</td>
</tr>
<tr>
<td>358</td>
<td>Hs.448851</td>
<td>USP6</td>
<td>9098</td>
<td>17p13</td>
</tr>
<tr>
<td>359</td>
<td>Hs.77854</td>
<td>RGN</td>
<td>9104</td>
<td>Xp11.3</td>
</tr>
<tr>
<td>360</td>
<td>Hs.625674</td>
<td>MTMR7</td>
<td>9108</td>
<td>8p22</td>
</tr>
<tr>
<td>361</td>
<td>Hs.216226</td>
<td>SYNGR1</td>
<td>9145</td>
<td>22q13.1</td>
</tr>
<tr>
<td>362</td>
<td>Hs.443683</td>
<td>MYOM2</td>
<td>9172</td>
<td>8p23.3</td>
</tr>
<tr>
<td>363</td>
<td>Hs.66</td>
<td>IL1RL1</td>
<td>9173</td>
<td>2q12</td>
</tr>
<tr>
<td>364</td>
<td>Hs.478031</td>
<td>SLC33A1</td>
<td>9197</td>
<td>3q25.31</td>
</tr>
<tr>
<td>365</td>
<td>Hs.533986</td>
<td>ZMYM6</td>
<td>9204</td>
<td>1p34.2</td>
</tr>
<tr>
<td>366</td>
<td>Hs.612814</td>
<td>CCPG1</td>
<td>9236</td>
<td>15q21.1</td>
</tr>
<tr>
<td>367</td>
<td>Hs.654558</td>
<td>CD83</td>
<td>9308</td>
<td>6p23</td>
</tr>
<tr>
<td>368</td>
<td>Hs.647113</td>
<td>ACCN3</td>
<td>9311</td>
<td>7q35</td>
</tr>
<tr>
<td>369</td>
<td>Hs.376206</td>
<td>KLF4</td>
<td>9314</td>
<td>9q31</td>
</tr>
<tr>
<td>370</td>
<td>Hs.632339</td>
<td>TRIP11</td>
<td>9321</td>
<td>14q31-q32</td>
</tr>
<tr>
<td>371</td>
<td>Hs.95243</td>
<td>TCEAL1</td>
<td>9338</td>
<td>Xq22.1</td>
</tr>
<tr>
<td>372</td>
<td>Hs.66708</td>
<td>VAMP3</td>
<td>9341</td>
<td>1p36.23</td>
</tr>
<tr>
<td>373</td>
<td>Hs.696554</td>
<td>ITGBL1</td>
<td>9358</td>
<td>13q33</td>
</tr>
<tr>
<td>374</td>
<td>Hs.80485</td>
<td>ADIPOQ</td>
<td>9370</td>
<td>3q27</td>
</tr>
<tr>
<td>375</td>
<td>Hs.667720</td>
<td>CYP7B1</td>
<td>9420</td>
<td>8q21.3</td>
</tr>
<tr>
<td>376</td>
<td>Hs.656823</td>
<td>RASAL2</td>
<td>9462</td>
<td>1q24</td>
</tr>
<tr>
<td>377</td>
<td>Hs.180871</td>
<td>PICK1</td>
<td>9463</td>
<td>22q13.1</td>
</tr>
<tr>
<td>378</td>
<td>Hs.643357</td>
<td>ADAMTS1</td>
<td>9510</td>
<td>21q21.2</td>
</tr>
<tr>
<td>379</td>
<td>Hs.459940</td>
<td>LITAF</td>
<td>9516</td>
<td>16p13.13</td>
</tr>
<tr>
<td>380</td>
<td>Hs.437075</td>
<td>CREB5</td>
<td>9586</td>
<td>7p15.1</td>
</tr>
<tr>
<td>381</td>
<td>Hs.371240</td>
<td>AKAP12</td>
<td>9590</td>
<td>6q24-q25</td>
</tr>
<tr>
<td>382</td>
<td>Hs.594708</td>
<td>SH3PXD2A</td>
<td>9644</td>
<td>10q24.33</td>
</tr>
<tr>
<td>383</td>
<td>Hs.468426</td>
<td>SOCS5</td>
<td>9655</td>
<td>2p21</td>
</tr>
<tr>
<td>384</td>
<td>Hs.654651</td>
<td>PDE4DIP</td>
<td>9659</td>
<td>1q12</td>
</tr>
<tr>
<td>385</td>
<td>Hs.655934</td>
<td>ZNF432</td>
<td>9668</td>
<td>19q13.33</td>
</tr>
<tr>
<td>386</td>
<td>Hs.518138</td>
<td>KIAA0040</td>
<td>9674</td>
<td>1q24-q25</td>
</tr>
<tr>
<td>387</td>
<td>Hs.168762</td>
<td>ULK2</td>
<td>9706</td>
<td>17p11.2</td>
</tr>
<tr>
<td>388</td>
<td>Hs.478868</td>
<td>KIAA0226</td>
<td>9711</td>
<td>3q29</td>
</tr>
<tr>
<td>389</td>
<td>Hs.559459</td>
<td>C6orf32</td>
<td>9750</td>
<td>6p22.3-p21.32</td>
</tr>
<tr>
<td>390</td>
<td>Hs.634856</td>
<td>TOX</td>
<td>9760</td>
<td>8q12.1</td>
</tr>
<tr>
<td>391</td>
<td>Hs.482660</td>
<td>ZFYVE16</td>
<td>9765</td>
<td>5q14</td>
</tr>
<tr>
<td>392</td>
<td>Hs.79276</td>
<td>KIAA0232</td>
<td>9778</td>
<td>4p16.1</td>
</tr>
<tr>
<td>393</td>
<td>Hs.31720</td>
<td>HEPH</td>
<td>9843</td>
<td>Xq11-q12</td>
</tr>
<tr>
<td>394</td>
<td>Hs.170999</td>
<td>LBA1</td>
<td>9881</td>
<td>3p22.2</td>
</tr>
<tr>
<td>395</td>
<td>Hs.524692</td>
<td>NUAK1</td>
<td>9891</td>
<td>12q23.3</td>
</tr>
<tr>
<td>396</td>
<td>Hs.5333</td>
<td>KBTBD11</td>
<td>9920</td>
<td>8p23.3</td>
</tr>
<tr>
<td>397</td>
<td>Hs.434951</td>
<td>USP15</td>
<td>9958</td>
<td>12q14</td>
</tr>
<tr>
<td>398</td>
<td>Hs.282735</td>
<td>NR1H4</td>
<td>9971</td>
<td>12q23.1</td>
</tr>
<tr>
<td>399</td>
<td>Hs.527105</td>
<td>HNRPD1</td>
<td>9987</td>
<td>4q13-q21</td>
</tr>
<tr>
<td>400</td>
<td>Hs.13967</td>
<td>NAALADL1</td>
<td>10004</td>
<td>11q12</td>
</tr>
<tr>
<td>401</td>
<td>Hs.508148</td>
<td>ABI1</td>
<td>10006</td>
<td>10p11.2</td>
</tr>
<tr>
<td>402</td>
<td>Hs.556496</td>
<td>TANK</td>
<td>10010</td>
<td>2q24-q31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>403</td>
<td>Hs.474705</td>
<td>TOM1</td>
<td>10043</td>
<td>22q13.1</td>
</tr>
<tr>
<td>404</td>
<td>Hs.306412</td>
<td>SH2D3C</td>
<td>10044</td>
<td>9q34.11</td>
</tr>
<tr>
<td>405</td>
<td>Hs.20136</td>
<td>MAMLD1</td>
<td>10046</td>
<td>Xq28</td>
</tr>
<tr>
<td>406</td>
<td>Hs.490745</td>
<td>DNAJB6</td>
<td>10049</td>
<td>7q36.3</td>
</tr>
<tr>
<td>407</td>
<td>Hs.593923</td>
<td>DNAJB6</td>
<td>10049</td>
<td>7q36.3</td>
</tr>
<tr>
<td>408</td>
<td>Hs.498720</td>
<td>OPTN</td>
<td>10133</td>
<td>10p13</td>
</tr>
<tr>
<td>409</td>
<td>Hs.332706</td>
<td>OPTN</td>
<td>10133</td>
<td>10p13</td>
</tr>
<tr>
<td>410</td>
<td>Hs.123464</td>
<td>P2RY5</td>
<td>10161</td>
<td>13q14</td>
</tr>
<tr>
<td>411</td>
<td>Hs.655248</td>
<td>MBOAT5</td>
<td>10162</td>
<td>12p13</td>
</tr>
<tr>
<td>412</td>
<td>Hs.648603</td>
<td>LHFP</td>
<td>10186</td>
<td>13q12</td>
</tr>
<tr>
<td>413</td>
<td>Hs.470882</td>
<td>CALCRL</td>
<td>10203</td>
<td>2q32.1</td>
</tr>
<tr>
<td>414</td>
<td>Hs.559259</td>
<td>NBR2</td>
<td>10230</td>
<td>17q21</td>
</tr>
<tr>
<td>415</td>
<td>Hs.25691</td>
<td>RAMP3</td>
<td>10268</td>
<td>7p13-p12</td>
</tr>
<tr>
<td>416</td>
<td>Hs.13351</td>
<td>LANCL1</td>
<td>10314</td>
<td>2q33-q35</td>
</tr>
<tr>
<td>417</td>
<td>Hs.471619</td>
<td>NMUR1</td>
<td>10316</td>
<td>2q37.1</td>
</tr>
<tr>
<td>418</td>
<td>Hs.50282</td>
<td>RRAGB</td>
<td>10325</td>
<td>Xp11.21</td>
</tr>
<tr>
<td>419</td>
<td>Hs.504687</td>
<td>MYL9</td>
<td>10398</td>
<td>20q11.23</td>
</tr>
<tr>
<td>420</td>
<td>Hs.54403</td>
<td>CLEC10A</td>
<td>10462</td>
<td>17p13.1</td>
</tr>
<tr>
<td>421</td>
<td>Hs.584851</td>
<td>TRIM38</td>
<td>10475</td>
<td>6p21.3</td>
</tr>
<tr>
<td>422</td>
<td>Hs.523793</td>
<td>NXF1</td>
<td>10482</td>
<td>11q12-q13</td>
</tr>
<tr>
<td>423</td>
<td>Hs.332708</td>
<td>FBLN5</td>
<td>10516</td>
<td>14q32.1</td>
</tr>
<tr>
<td>424</td>
<td>Hs.448664</td>
<td>DEAF1</td>
<td>10522</td>
<td>11p15.5</td>
</tr>
<tr>
<td>425</td>
<td>Hs.696027</td>
<td>SORBS1</td>
<td>10580</td>
<td>10q23.3-q24.1</td>
</tr>
<tr>
<td>426</td>
<td>Hs.522449</td>
<td>POMT1</td>
<td>10585</td>
<td>9q34.1</td>
</tr>
<tr>
<td>427</td>
<td>Hs.369574</td>
<td>CDC42EP3</td>
<td>10602</td>
<td>2p21</td>
</tr>
<tr>
<td>428</td>
<td>Hs.480311</td>
<td>PDLIM5</td>
<td>10611</td>
<td>4q22</td>
</tr>
<tr>
<td>429</td>
<td>Hs.472227</td>
<td>POLR3F</td>
<td>10621</td>
<td>20p11.23</td>
</tr>
<tr>
<td>430</td>
<td>Hs.533977</td>
<td>TXNIP</td>
<td>10628</td>
<td>1q21.1</td>
</tr>
<tr>
<td>431</td>
<td>Hs.309288</td>
<td>CUGBP2</td>
<td>10659</td>
<td>10p13</td>
</tr>
<tr>
<td>432</td>
<td>Hs.59106</td>
<td>CGRRF1</td>
<td>10668</td>
<td>14q22.2</td>
</tr>
<tr>
<td>433</td>
<td>Hs.515048</td>
<td>AP4B1</td>
<td>10717</td>
<td>1p13.2</td>
</tr>
<tr>
<td>434</td>
<td>Hs.314246</td>
<td>ZNF271</td>
<td>10778</td>
<td>18q12</td>
</tr>
<tr>
<td>435</td>
<td>Hs.635221</td>
<td>WASF3</td>
<td>10810</td>
<td>13q12</td>
</tr>
<tr>
<td>436</td>
<td>Hs.519694</td>
<td>C5orf4</td>
<td>10826</td>
<td>5q31-q32</td>
</tr>
<tr>
<td>437</td>
<td>Hs.920</td>
<td>ARID5A</td>
<td>10865</td>
<td>2q11.2</td>
</tr>
<tr>
<td>438</td>
<td>Hs.654480</td>
<td>HCP5</td>
<td>10866</td>
<td>6p21.3</td>
</tr>
<tr>
<td>439</td>
<td>Hs.655332</td>
<td>LYVE1</td>
<td>10894</td>
<td>11p15</td>
</tr>
<tr>
<td>440</td>
<td>Hs.425144</td>
<td>MTMR11</td>
<td>10903</td>
<td>1q12-q21</td>
</tr>
<tr>
<td>441</td>
<td>Hs.486357</td>
<td>SMPDL3A</td>
<td>10924</td>
<td>6q22.31</td>
</tr>
<tr>
<td>442</td>
<td>Hs.523774</td>
<td>EHD1</td>
<td>10938</td>
<td>11q13</td>
</tr>
<tr>
<td>443</td>
<td>Hs.272168</td>
<td>SERINC3</td>
<td>10955</td>
<td>20q13.1-q13.3</td>
</tr>
<tr>
<td>444</td>
<td>Hs.75969</td>
<td>PNRC1</td>
<td>10957</td>
<td>6q15</td>
</tr>
<tr>
<td>445</td>
<td>Hs.509343</td>
<td>FERMT2</td>
<td>10979</td>
<td>14q22.2</td>
</tr>
<tr>
<td>446</td>
<td>Hs.532824</td>
<td>MAPRE2</td>
<td>10982</td>
<td>18q12.1</td>
</tr>
<tr>
<td>447</td>
<td>Hs.268551</td>
<td>RIPK3</td>
<td>11035</td>
<td>14q11.2</td>
</tr>
<tr>
<td>448</td>
<td>Hs.659219</td>
<td>C10orf10</td>
<td>11067</td>
<td>10q11.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>449</td>
<td>Hs.470646</td>
<td>RAPGEF4</td>
<td>11069</td>
<td>2q31-q32</td>
</tr>
<tr>
<td>450</td>
<td>Hs.521651</td>
<td>STMN2</td>
<td>11075</td>
<td>8q21.13</td>
</tr>
<tr>
<td>451</td>
<td>Hs.13852</td>
<td>DNAJB4</td>
<td>11080</td>
<td>1p31.1</td>
</tr>
<tr>
<td>452</td>
<td>Hs.666782</td>
<td>PRDM5</td>
<td>11107</td>
<td>4q25-q26</td>
</tr>
<tr>
<td>453</td>
<td>Hs.191510</td>
<td>BTN3A1</td>
<td>11119</td>
<td>6p22.1</td>
</tr>
<tr>
<td>454</td>
<td>Hs.159028</td>
<td>BTN2A1</td>
<td>11120</td>
<td>6p22.1</td>
</tr>
<tr>
<td>455</td>
<td>Hs.43670</td>
<td>KIF3A</td>
<td>11127</td>
<td>5q31</td>
</tr>
<tr>
<td>456</td>
<td>Hs.293411</td>
<td>AP4S1</td>
<td>11154</td>
<td>14q12</td>
</tr>
<tr>
<td>457</td>
<td>Hs.657271</td>
<td>LDB3</td>
<td>11155</td>
<td>10q22.3-q23.2</td>
</tr>
<tr>
<td>458</td>
<td>Hs.43666</td>
<td>PTP4A3</td>
<td>11156</td>
<td>8q24.3</td>
</tr>
<tr>
<td>459</td>
<td>Hs.506357</td>
<td>FAM107A</td>
<td>11170</td>
<td>3p21.1</td>
</tr>
<tr>
<td>460</td>
<td>Hs.647118</td>
<td>ABCB8</td>
<td>11194</td>
<td>7q36</td>
</tr>
<tr>
<td>461</td>
<td>Hs.105105</td>
<td>AKAP11</td>
<td>11215</td>
<td>13q14.11</td>
</tr>
<tr>
<td>462</td>
<td>Hs.373857</td>
<td>KLF12</td>
<td>11278</td>
<td>13q22</td>
</tr>
<tr>
<td>463</td>
<td>Hs.646614</td>
<td>KLF8</td>
<td>11279</td>
<td>Xp11.21</td>
</tr>
<tr>
<td>464</td>
<td>Hs.7884</td>
<td>SLCO2B1</td>
<td>11309</td>
<td>11q13</td>
</tr>
<tr>
<td>465</td>
<td>Hs.483909</td>
<td>GPR182</td>
<td>11318</td>
<td>12q13.3</td>
</tr>
<tr>
<td>466</td>
<td>Hs.8904</td>
<td>VSIG4</td>
<td>11326</td>
<td>Xq12-q13.3</td>
</tr>
<tr>
<td>467</td>
<td>Hs.439199</td>
<td>NLGN4Y</td>
<td>22829</td>
<td>Yq11.221</td>
</tr>
<tr>
<td>468</td>
<td>Hs.445030</td>
<td>RHOBTB3</td>
<td>22836</td>
<td>5q15</td>
</tr>
<tr>
<td>469</td>
<td>Hs.470457</td>
<td>COBLL1</td>
<td>22837</td>
<td>2q24.3</td>
</tr>
<tr>
<td>470</td>
<td>Hs.508010</td>
<td>FNDC3A</td>
<td>22862</td>
<td>13q14.2</td>
</tr>
<tr>
<td>471</td>
<td>Hs.188495</td>
<td>WDR37</td>
<td>22884</td>
<td>10p15.3</td>
</tr>
<tr>
<td>472</td>
<td>Hs.443109</td>
<td>ARHGEF15</td>
<td>22899</td>
<td>17p13.1</td>
</tr>
<tr>
<td>473</td>
<td>Hs.7972</td>
<td>RUFY3</td>
<td>22902</td>
<td>4q13.3</td>
</tr>
<tr>
<td>474</td>
<td>Hs.268107</td>
<td>MMRN1</td>
<td>22915</td>
<td>4q22</td>
</tr>
<tr>
<td>475</td>
<td>Hs.182982</td>
<td>GOLGA8A</td>
<td>23015</td>
<td>15q11.2</td>
</tr>
<tr>
<td>476</td>
<td>Hs.151220</td>
<td>PALLD</td>
<td>23022</td>
<td>4q32.3</td>
</tr>
<tr>
<td>477</td>
<td>Hs.98259</td>
<td>SAMD4A</td>
<td>23034</td>
<td>14q22.2</td>
</tr>
<tr>
<td>478</td>
<td>Hs.167115</td>
<td>ENDOD1</td>
<td>23052</td>
<td>11q21</td>
</tr>
<tr>
<td>479</td>
<td>Hs.155995</td>
<td>CLUAP1</td>
<td>23059</td>
<td>16p13.3</td>
</tr>
<tr>
<td>480</td>
<td>Hs.591221</td>
<td>MYCBP2</td>
<td>23077</td>
<td>13q22</td>
</tr>
<tr>
<td>481</td>
<td>Hs.584867</td>
<td>CDC2L6</td>
<td>23097</td>
<td>6q21</td>
</tr>
<tr>
<td>482</td>
<td>Hs.440414</td>
<td>SPG20</td>
<td>23111</td>
<td>13q13.3</td>
</tr>
<tr>
<td>483</td>
<td>Hs.301989</td>
<td>STAB1</td>
<td>23166</td>
<td>3p21.1</td>
</tr>
<tr>
<td>484</td>
<td>Hs.464585</td>
<td>ANKRD12</td>
<td>23253</td>
<td>18p11.22</td>
</tr>
<tr>
<td>485</td>
<td>Hs.633454</td>
<td>EXOC7</td>
<td>23265</td>
<td>17q25.1</td>
</tr>
<tr>
<td>486</td>
<td>Hs.193133</td>
<td>SASH1</td>
<td>23328</td>
<td>6q24.3</td>
</tr>
<tr>
<td>487</td>
<td>Hs.655410</td>
<td>DNAJC16</td>
<td>23341</td>
<td>1p36.1</td>
</tr>
<tr>
<td>488</td>
<td>Hs.12967</td>
<td>SYNE1</td>
<td>23345</td>
<td>6q25</td>
</tr>
<tr>
<td>489</td>
<td>Hs.343334</td>
<td>TENC1</td>
<td>23371</td>
<td>12q13.13</td>
</tr>
<tr>
<td>490</td>
<td>Hs.4014</td>
<td>ZDHHC17</td>
<td>23390</td>
<td>12q21.2</td>
</tr>
<tr>
<td>491</td>
<td>Hs.369779</td>
<td>SIRT1</td>
<td>23411</td>
<td>10q21.3</td>
</tr>
<tr>
<td>492</td>
<td>Hs.580782</td>
<td>MACF1</td>
<td>23499</td>
<td>1p32-p31</td>
</tr>
<tr>
<td>493</td>
<td>Hs.409081</td>
<td>OPN3</td>
<td>23596</td>
<td>1q43</td>
</tr>
<tr>
<td>494</td>
<td>Hs.271341</td>
<td>RABGAP1</td>
<td>23637</td>
<td>9q33.2</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Symbol</td>
<td>Chromosome</td>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>Hs.517617</td>
<td>MAFF</td>
<td>22q13.1</td>
<td>23764</td>
<td>22q13.1</td>
</tr>
<tr>
<td>Hs.533710</td>
<td>FLRT2</td>
<td>14q24-q32</td>
<td>23768</td>
<td>23768</td>
</tr>
<tr>
<td>Hs.252839</td>
<td>IFIT5</td>
<td>10q23.31</td>
<td>24138</td>
<td>24138</td>
</tr>
<tr>
<td>Hs.591976</td>
<td>PANX1</td>
<td>11q21</td>
<td>24145</td>
<td>24145</td>
</tr>
<tr>
<td>Hs.30965</td>
<td>SHC2</td>
<td>19p13.3</td>
<td>25759</td>
<td>25759</td>
</tr>
<tr>
<td>Hs.519075</td>
<td>LMOD1</td>
<td>1p32</td>
<td>25802</td>
<td>25802</td>
</tr>
<tr>
<td>Hs.653847</td>
<td>METTL7A</td>
<td>12q13.13</td>
<td>25840</td>
<td>25840</td>
</tr>
<tr>
<td>Hs.105460</td>
<td>DKFZP564O0823</td>
<td>4q13.3-q21.3</td>
<td>25849</td>
<td>25849</td>
</tr>
<tr>
<td>Hs.591288</td>
<td>C3orf17</td>
<td>3q13.2</td>
<td>25871</td>
<td>25871</td>
</tr>
<tr>
<td>Hs.655272</td>
<td>LETMD1</td>
<td>12q13.13</td>
<td>25875</td>
<td>25875</td>
</tr>
<tr>
<td>Hs.477015</td>
<td>ABI3BP</td>
<td>3q12</td>
<td>25890</td>
<td>25890</td>
</tr>
<tr>
<td>Hs.55044</td>
<td>DKFZP586H2123</td>
<td>11p13</td>
<td>25891</td>
<td>25891</td>
</tr>
<tr>
<td>Hs.12844</td>
<td>EGFL6</td>
<td>Xp22</td>
<td>25975</td>
<td>25975</td>
</tr>
<tr>
<td>Hs.466539</td>
<td>CLIP3</td>
<td>19q13.12</td>
<td>25999</td>
<td>25999</td>
</tr>
<tr>
<td>Hs.431317</td>
<td>GORASP2</td>
<td>2q31.1-q31.2</td>
<td>26003</td>
<td>26003</td>
</tr>
<tr>
<td>Hs.654657</td>
<td>SIPA1L1</td>
<td>14q24.2</td>
<td>26037</td>
<td>26037</td>
</tr>
<tr>
<td>Hs.446017</td>
<td>WSB1</td>
<td>17q11.1</td>
<td>26118</td>
<td>26118</td>
</tr>
<tr>
<td>Hs.279580</td>
<td>KIAA1279</td>
<td>10q21.3</td>
<td>26128</td>
<td>26128</td>
</tr>
<tr>
<td>Hs.655108</td>
<td>ZBTB20</td>
<td>3q13.2</td>
<td>26137</td>
<td>26137</td>
</tr>
<tr>
<td>Hs.693802</td>
<td>ZBTB20</td>
<td>3q13.2</td>
<td>26137</td>
<td>26137</td>
</tr>
<tr>
<td>Hs.494985</td>
<td>FBXW2</td>
<td>9q34</td>
<td>26190</td>
<td>26190</td>
</tr>
<tr>
<td>Hs.696160</td>
<td>PITPNC1</td>
<td>17q24.2</td>
<td>26207</td>
<td>26207</td>
</tr>
<tr>
<td>Hs.643433</td>
<td>FBXL5</td>
<td>4p15.33</td>
<td>26234</td>
<td>26234</td>
</tr>
<tr>
<td>Hs.76917</td>
<td>FBXO8</td>
<td>4q34.1</td>
<td>26269</td>
<td>26269</td>
</tr>
<tr>
<td>Hs.400095</td>
<td>HSPB8</td>
<td>12q24.23</td>
<td>26353</td>
<td>26353</td>
</tr>
<tr>
<td>Hs.78960</td>
<td>LAT52</td>
<td>13q11-q12</td>
<td>26524</td>
<td>26524</td>
</tr>
<tr>
<td>Hs.491172</td>
<td>NBEA</td>
<td>13q13</td>
<td>26960</td>
<td>26960</td>
</tr>
<tr>
<td>Hs.352656</td>
<td>GHITM</td>
<td>10q23.1</td>
<td>27069</td>
<td>27069</td>
</tr>
<tr>
<td>Hs.657015</td>
<td>SDCBP2</td>
<td>20p13</td>
<td>27111</td>
<td>27111</td>
</tr>
<tr>
<td>Hs.652367</td>
<td>PDE7B</td>
<td>6q23-q24</td>
<td>27115</td>
<td>27115</td>
</tr>
<tr>
<td>Hs.502612</td>
<td>HSPB7</td>
<td>1p36.23-p34.3</td>
<td>27129</td>
<td>27129</td>
</tr>
<tr>
<td>Hs.310893</td>
<td>CSDF2</td>
<td>22q13.2-q13.31</td>
<td>27254</td>
<td>27254</td>
</tr>
<tr>
<td>Hs.306339</td>
<td>SRPX2</td>
<td>Xq21.33-q23</td>
<td>27286</td>
<td>27286</td>
</tr>
<tr>
<td>Hs.419800</td>
<td>C10orf28</td>
<td>10q24.2</td>
<td>27291</td>
<td>27291</td>
</tr>
<tr>
<td>Hs.696468</td>
<td>RBMS3</td>
<td>3p24-p23</td>
<td>27303</td>
<td>27303</td>
</tr>
<tr>
<td>Hs.530053</td>
<td>RABGEF1</td>
<td>7q11.21</td>
<td>27342</td>
<td>27342</td>
</tr>
<tr>
<td>Hs.523230</td>
<td>POLL</td>
<td>10q23</td>
<td>27343</td>
<td>27343</td>
</tr>
<tr>
<td>Hs.696057</td>
<td>MAT2B</td>
<td>5q34-q35</td>
<td>27430</td>
<td>27430</td>
</tr>
<tr>
<td>Hs.279819</td>
<td>MAGEH1</td>
<td>Xp11.21</td>
<td>28986</td>
<td>28986</td>
</tr>
<tr>
<td>Hs.371856</td>
<td>ZC3H7A</td>
<td>16p13-p12</td>
<td>29066</td>
<td>29066</td>
</tr>
<tr>
<td>Hs.699226</td>
<td>TRA2A</td>
<td>7p15.3</td>
<td>29896</td>
<td>29896</td>
</tr>
<tr>
<td>Hs.470369</td>
<td>BAZ2B</td>
<td>2q23-q24</td>
<td>29994</td>
<td>29994</td>
</tr>
<tr>
<td>Hs.109672</td>
<td>ST6GALNAC6</td>
<td>9q34.11</td>
<td>30815</td>
<td>30815</td>
</tr>
<tr>
<td>Hs.278694</td>
<td>CD209</td>
<td>19q13.3</td>
<td>30835</td>
<td>30835</td>
</tr>
<tr>
<td>Hs.631554</td>
<td>EHD2</td>
<td>19q13.3</td>
<td>30846</td>
<td>30846</td>
</tr>
<tr>
<td>Hs.432132</td>
<td>G0S2</td>
<td>1q32.2-q41</td>
<td>50486</td>
<td>50486</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene Symbol</td>
<td>Chromosome</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>541</td>
<td>Hs.647182</td>
<td>10q26.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>542</td>
<td>Hs.44685</td>
<td>11p15.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>543</td>
<td>Hs.241545</td>
<td>10pter-q26.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>544</td>
<td>Hs.16606</td>
<td>10q24.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>Hs.579828</td>
<td>14q24.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>546</td>
<td>Hs.136309</td>
<td>1p22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>547</td>
<td>Hs.178170</td>
<td>10q22.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>548</td>
<td>Hs.631730</td>
<td>12p13.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>549</td>
<td>Hs.27018</td>
<td>15q11.2-q22.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>Hs.439474</td>
<td>5q31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>551</td>
<td>Hs.274309</td>
<td>16p11.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>552</td>
<td>Hs.428147</td>
<td>6q21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>553</td>
<td>Hs.279815</td>
<td>12q13.11-q14.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>554</td>
<td>Hs.647072</td>
<td>7q36.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>555</td>
<td>Hs.191213</td>
<td>6q25.1-q26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>556</td>
<td>Hs.163776</td>
<td>6q15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>557</td>
<td>Hs.371563</td>
<td>9q32-q34.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>558</td>
<td>Hs.705444</td>
<td>9q32-q34.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>559</td>
<td>Hs.696104</td>
<td>6p22.3-p22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>560</td>
<td>Hs.515890</td>
<td>2p23.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>561</td>
<td>Hs.534458</td>
<td>16q22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>562</td>
<td>Hs.510327</td>
<td>14q31-q32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>563</td>
<td>Hs.152913</td>
<td>4q23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>564</td>
<td>Hs.656794</td>
<td>19p13.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>565</td>
<td>Hs.125300</td>
<td>11p15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>566</td>
<td>Hs.547009</td>
<td>21q22.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>567</td>
<td>Hs.54725</td>
<td>21q22.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>568</td>
<td>Hs.702188</td>
<td>21q22.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>569</td>
<td>Hs.62880</td>
<td>22q13.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>570</td>
<td>Hs.353022</td>
<td>2p13-p15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>571</td>
<td>Hs.431081</td>
<td>4q26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>572</td>
<td>Hs.524121</td>
<td>11q24.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>573</td>
<td>Hs.567513</td>
<td>3q21.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>574</td>
<td>Hs.591901</td>
<td>9q31-q32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>575</td>
<td>Hs.558570</td>
<td>1p36.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>576</td>
<td>Hs.591145</td>
<td>12p13.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>577</td>
<td>Hs.370522</td>
<td>6p22.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>578</td>
<td>Hs.270851</td>
<td>4q31.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>579</td>
<td>Hs.269654</td>
<td>13q12.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>580</td>
<td>Hs.441975</td>
<td>17p13.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>581</td>
<td>Hs.386684</td>
<td>6q23.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>582</td>
<td>Hs.435655</td>
<td>9q22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>583</td>
<td>Hs.356216</td>
<td>1p12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>584</td>
<td>Hs.406223</td>
<td>17q12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>585</td>
<td>Hs.30141</td>
<td>15q13.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>586</td>
<td>Hs.439894</td>
<td>1p36.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>587</td>
<td>Hs.371210</td>
<td>C1orf27</td>
<td>54953</td>
<td>1q25</td>
</tr>
<tr>
<td>588</td>
<td>Hs.413123</td>
<td>C2orf42</td>
<td>54980</td>
<td>2p14</td>
</tr>
<tr>
<td>589</td>
<td>Hs.265018</td>
<td>FAM118A</td>
<td>55007</td>
<td>22q13</td>
</tr>
<tr>
<td>590</td>
<td>Hs.591900</td>
<td>STX17</td>
<td>55014</td>
<td>9q31.1</td>
</tr>
<tr>
<td>591</td>
<td>Hs.409352</td>
<td>PID1</td>
<td>55022</td>
<td>2q36.3</td>
</tr>
<tr>
<td>592</td>
<td>Hs.655317</td>
<td>C19orf60</td>
<td>55049</td>
<td>19p13.11</td>
</tr>
<tr>
<td>593</td>
<td>Hs.567523</td>
<td>DET1</td>
<td>55070</td>
<td>15q25.3</td>
</tr>
<tr>
<td>594</td>
<td>Hs.504597</td>
<td>TAPBP</td>
<td>55080</td>
<td>12p13.31</td>
</tr>
<tr>
<td>595</td>
<td>Hs.159066</td>
<td>C10orf118</td>
<td>55088</td>
<td>10q25.3</td>
</tr>
<tr>
<td>596</td>
<td>Hs.168241</td>
<td>ATG2B</td>
<td>55102</td>
<td>14q32.2</td>
</tr>
<tr>
<td>597</td>
<td>Hs.353454</td>
<td>BSDC1</td>
<td>55108</td>
<td>1p35.1</td>
</tr>
<tr>
<td>598</td>
<td>Hs.476319</td>
<td>ECHDC2</td>
<td>55268</td>
<td>1p32.3</td>
</tr>
<tr>
<td>599</td>
<td>Hs.435933</td>
<td>PHF10</td>
<td>55274</td>
<td>6q27</td>
</tr>
<tr>
<td>600</td>
<td>Hs.24545</td>
<td>ZNF444</td>
<td>55311</td>
<td>19q13.42</td>
</tr>
<tr>
<td>601</td>
<td>Hs.567532</td>
<td>NIPSNAP3B</td>
<td>55335</td>
<td>9q31.1</td>
</tr>
<tr>
<td>602</td>
<td>Hs.647079</td>
<td>GIMAP5</td>
<td>55340</td>
<td>7q36.1</td>
</tr>
<tr>
<td>603</td>
<td>Hs.525589</td>
<td>MEG3</td>
<td>55384</td>
<td>14q32</td>
</tr>
<tr>
<td>604</td>
<td>Hs.700471</td>
<td>YOD1</td>
<td>55432</td>
<td>1q32.1</td>
</tr>
<tr>
<td>605</td>
<td>Hs.654970</td>
<td>IL17RB</td>
<td>55540</td>
<td>3p21.1</td>
</tr>
<tr>
<td>606</td>
<td>Hs.525163</td>
<td>ANKRDL</td>
<td>55608</td>
<td>13q34</td>
</tr>
<tr>
<td>607</td>
<td>Hs.515169</td>
<td>TRMT1</td>
<td>55621</td>
<td>19p13.13</td>
</tr>
<tr>
<td>608</td>
<td>Hs.259605</td>
<td>PIGV</td>
<td>55650</td>
<td>1p36.11</td>
</tr>
<tr>
<td>609</td>
<td>Hs.377705</td>
<td>ZNF692</td>
<td>55657</td>
<td>1q44</td>
</tr>
<tr>
<td>610</td>
<td>Hs.561954</td>
<td>CDC37L1</td>
<td>55664</td>
<td>9p24.1</td>
</tr>
<tr>
<td>611</td>
<td>Hs.469881</td>
<td>LIMS2</td>
<td>55679</td>
<td>2q14.3</td>
</tr>
<tr>
<td>612</td>
<td>Hs.584933</td>
<td>ZNF534</td>
<td>55713</td>
<td>20q13.12</td>
</tr>
<tr>
<td>613</td>
<td>Hs.696152</td>
<td>RCOR3</td>
<td>55758</td>
<td>1q32.2-q32.3</td>
</tr>
<tr>
<td>614</td>
<td>Hs.467210</td>
<td>ZNF83</td>
<td>55769</td>
<td>19q13.3</td>
</tr>
<tr>
<td>615</td>
<td>Hs.655166</td>
<td>ChGn</td>
<td>55790</td>
<td>8p21.3</td>
</tr>
<tr>
<td>616</td>
<td>Hs.4865</td>
<td>SCN3B</td>
<td>55800</td>
<td>11q23.3</td>
</tr>
<tr>
<td>617</td>
<td>Hs.435741</td>
<td>IQWD1</td>
<td>55827</td>
<td>1q24.2</td>
</tr>
<tr>
<td>618</td>
<td>Hs.32148</td>
<td>SELS</td>
<td>55829</td>
<td>15q26.3</td>
</tr>
<tr>
<td>619</td>
<td>Hs.187635</td>
<td>C20orf19</td>
<td>55857</td>
<td>20pter-q11.23</td>
</tr>
<tr>
<td>620</td>
<td>Hs.446438</td>
<td>GPRC5C</td>
<td>55890</td>
<td>17q25</td>
</tr>
<tr>
<td>621</td>
<td>Hs.507025</td>
<td>MYNN</td>
<td>55892</td>
<td>3q26.2</td>
</tr>
<tr>
<td>622</td>
<td>Hs.699209</td>
<td>ZNF395</td>
<td>55893</td>
<td>8p21.1</td>
</tr>
<tr>
<td>623</td>
<td>Hs.193406</td>
<td>C1orf183</td>
<td>55924</td>
<td>1p13.2</td>
</tr>
<tr>
<td>624</td>
<td>Hs.529100</td>
<td>LIN37</td>
<td>55957</td>
<td>19q13.1</td>
</tr>
<tr>
<td>625</td>
<td>Hs.168799</td>
<td>METTL3</td>
<td>56339</td>
<td>14q11.1</td>
</tr>
<tr>
<td>626</td>
<td>Hs.164144</td>
<td>EIF5A2</td>
<td>56648</td>
<td>3q26.2</td>
</tr>
<tr>
<td>627</td>
<td>Hs.499960</td>
<td>SAR1A</td>
<td>56681</td>
<td>1q22.1</td>
</tr>
<tr>
<td>628</td>
<td>Hs.29106</td>
<td>DUSP22</td>
<td>56940</td>
<td>6p25.3</td>
</tr>
<tr>
<td>629</td>
<td>Hs.9315</td>
<td>OLFML3</td>
<td>56944</td>
<td>1p13.2</td>
</tr>
<tr>
<td>630</td>
<td>Hs.118241</td>
<td>CABC1</td>
<td>56997</td>
<td>1q42.13</td>
</tr>
<tr>
<td>631</td>
<td>Hs.4859</td>
<td>CCNL1</td>
<td>57018</td>
<td>3q25.32</td>
</tr>
<tr>
<td>632</td>
<td>Hs.126035</td>
<td>RPGRIP1</td>
<td>57096</td>
<td>14q11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>633</td>
<td>Hs.371794</td>
<td>ZNFX1</td>
<td>57169</td>
<td>20q13.13</td>
</tr>
<tr>
<td>634</td>
<td>Hs.645966</td>
<td>FAM91A2</td>
<td>57234</td>
<td>1q21.1</td>
</tr>
<tr>
<td>635</td>
<td>Hs.655636</td>
<td>KIAA0508</td>
<td>57244</td>
<td>1p36.32</td>
</tr>
<tr>
<td>636</td>
<td>Hs.333958</td>
<td>SLC4A10</td>
<td>57282</td>
<td>2q23-q24</td>
</tr>
<tr>
<td>637</td>
<td>Hs.656339</td>
<td>RHOJ</td>
<td>57381</td>
<td>14q23.2</td>
</tr>
<tr>
<td>638</td>
<td>Hs.525205</td>
<td>NDRG2</td>
<td>57447</td>
<td>14q11.2</td>
</tr>
<tr>
<td>639</td>
<td>Hs.21035</td>
<td>GALNTL1</td>
<td>57452</td>
<td>14q24.1</td>
</tr>
<tr>
<td>640</td>
<td>Hs.551552</td>
<td>ZNF512B</td>
<td>57473</td>
<td>20q13.33</td>
</tr>
<tr>
<td>641</td>
<td>Hs.705876</td>
<td>ZNF608</td>
<td>57507</td>
<td>5q23.2</td>
</tr>
<tr>
<td>642</td>
<td>Hs.7946</td>
<td>MTUS1</td>
<td>57509</td>
<td>8p22</td>
</tr>
<tr>
<td>643</td>
<td>Hs.657263</td>
<td>CDGAP</td>
<td>57514</td>
<td>3q13.32-q13.33</td>
</tr>
<tr>
<td>644</td>
<td>Hs.156352</td>
<td>KIAA1377</td>
<td>57562</td>
<td>11q22.1</td>
</tr>
<tr>
<td>645</td>
<td>Hs.211520</td>
<td>KIAA1432</td>
<td>57589</td>
<td>9p24.1</td>
</tr>
<tr>
<td>646</td>
<td>Hs.42586</td>
<td>GPAM</td>
<td>57678</td>
<td>10q25.2</td>
</tr>
<tr>
<td>647</td>
<td>Hs.438482</td>
<td>WDR19</td>
<td>57728</td>
<td>4p14</td>
</tr>
<tr>
<td>648</td>
<td>Hs.270869</td>
<td>ZNF410</td>
<td>57862</td>
<td>14q24.3</td>
</tr>
<tr>
<td>649</td>
<td>Hs.511251</td>
<td>SQRDL</td>
<td>58472</td>
<td>15q15</td>
</tr>
<tr>
<td>650</td>
<td>Hs.516994</td>
<td>TP53IP2</td>
<td>58476</td>
<td>20q11.22</td>
</tr>
<tr>
<td>651</td>
<td>Hs.655066</td>
<td>ZBED5</td>
<td>58486</td>
<td>11p15.3</td>
</tr>
<tr>
<td>652</td>
<td>Hs.201034</td>
<td>NTN4</td>
<td>59277</td>
<td>12q22-q23</td>
</tr>
<tr>
<td>653</td>
<td>Hs.501624</td>
<td>SIGIRR</td>
<td>59307</td>
<td>11p15.5</td>
</tr>
<tr>
<td>654</td>
<td>Hs.407694</td>
<td>ZNF350</td>
<td>59348</td>
<td>19q13.33</td>
</tr>
<tr>
<td>655</td>
<td>Hs.269764</td>
<td>BACH2</td>
<td>60468</td>
<td>6q15</td>
</tr>
<tr>
<td>656</td>
<td>Hs.42194</td>
<td>SPCS3</td>
<td>60559</td>
<td>4q34.2</td>
</tr>
<tr>
<td>657</td>
<td>Hs.187284</td>
<td>PAPPA2</td>
<td>60676</td>
<td>1q23-q25</td>
</tr>
<tr>
<td>658</td>
<td>Hs.348342</td>
<td>BRUNOL6</td>
<td>60677</td>
<td>15q24</td>
</tr>
<tr>
<td>659</td>
<td>Hs.463035</td>
<td>FKBP10</td>
<td>60681</td>
<td>17q21.2</td>
</tr>
<tr>
<td>660</td>
<td>Hs.372309</td>
<td>C10orf84</td>
<td>63877</td>
<td>10q26.11</td>
</tr>
<tr>
<td>661</td>
<td>Hs.567562</td>
<td>CIDEC</td>
<td>63924</td>
<td>3p25.3</td>
</tr>
<tr>
<td>662</td>
<td>Hs.198158</td>
<td>PBLD</td>
<td>64081</td>
<td>10pter-q25.3</td>
</tr>
<tr>
<td>663</td>
<td>Hs.487200</td>
<td>SMOC2</td>
<td>64094</td>
<td>6q27</td>
</tr>
<tr>
<td>664</td>
<td>Hs.24719</td>
<td>MOAP1</td>
<td>64112</td>
<td>14q32</td>
</tr>
<tr>
<td>665</td>
<td>Hs.591605</td>
<td>TMBIM1</td>
<td>64114</td>
<td>2p24.3-p24.1</td>
</tr>
<tr>
<td>666</td>
<td>Hs.199368</td>
<td>TINAGL1</td>
<td>64129</td>
<td>1p35.2</td>
</tr>
<tr>
<td>667</td>
<td>Hs.525597</td>
<td>DIO3OS</td>
<td>64150</td>
<td>14q32.31</td>
</tr>
<tr>
<td>668</td>
<td>Hs.319171</td>
<td>NFKBIOS</td>
<td>64332</td>
<td>3p12-q12</td>
</tr>
<tr>
<td>669</td>
<td>Hs.420830</td>
<td>HIF3A</td>
<td>64344</td>
<td>19q13.32</td>
</tr>
<tr>
<td>670</td>
<td>Hs.501289</td>
<td>IKZF5</td>
<td>64376</td>
<td>10q26</td>
</tr>
<tr>
<td>671</td>
<td>Hs.511143</td>
<td>ZFP106</td>
<td>64397</td>
<td>15q15.1</td>
</tr>
<tr>
<td>672</td>
<td>Hs.380897</td>
<td>AKTIP</td>
<td>64400</td>
<td>16q12.2</td>
</tr>
<tr>
<td>673</td>
<td>Hs.592982</td>
<td>TPSB2</td>
<td>64499</td>
<td>16p13.3</td>
</tr>
<tr>
<td>674</td>
<td>Hs.159430</td>
<td>FNDC3B</td>
<td>64778</td>
<td>3q26.31</td>
</tr>
<tr>
<td>675</td>
<td>Hs.71912</td>
<td>LMF1</td>
<td>64788</td>
<td>16p13.3</td>
</tr>
<tr>
<td>676</td>
<td>Hs.112981</td>
<td>DEPDC6</td>
<td>64798</td>
<td>8q24.12</td>
</tr>
<tr>
<td>677</td>
<td>Hs.471162</td>
<td>RAPH1</td>
<td>65059</td>
<td>2q33</td>
</tr>
<tr>
<td>678</td>
<td>Hs.235390</td>
<td>ZSCAN18</td>
<td>65982</td>
<td>19q13.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>679</td>
<td>Hs.363558</td>
<td>GRAMD3</td>
<td>65983</td>
<td>5q23.2</td>
</tr>
<tr>
<td>680</td>
<td>Hs.8035</td>
<td>RASL11B</td>
<td>65997</td>
<td>4q12</td>
</tr>
<tr>
<td>681</td>
<td>Hs.632772</td>
<td>SLC2A11</td>
<td>66035</td>
<td>22q11.2</td>
</tr>
<tr>
<td>682</td>
<td>Hs.8719</td>
<td>DUSP26</td>
<td>78986</td>
<td>8p12</td>
</tr>
<tr>
<td>683</td>
<td>Hs.241576</td>
<td>DERL1</td>
<td>79139</td>
<td>8q24.13</td>
</tr>
<tr>
<td>684</td>
<td>Hs.591453</td>
<td>MMEL1</td>
<td>79258</td>
<td>1p36</td>
</tr>
<tr>
<td>685</td>
<td>Hs.181173</td>
<td>GLB1L</td>
<td>79411</td>
<td>2q35</td>
</tr>
<tr>
<td>686</td>
<td>Hs.211511</td>
<td>TCTN1</td>
<td>79600</td>
<td>12q24.11</td>
</tr>
<tr>
<td>687</td>
<td>Hs.655660</td>
<td>RIC3</td>
<td>79608</td>
<td>11p15.4</td>
</tr>
<tr>
<td>688</td>
<td>Hs.90250</td>
<td>C4orf31</td>
<td>79625</td>
<td>4q27</td>
</tr>
<tr>
<td>689</td>
<td>Hs.459652</td>
<td>TMEM204</td>
<td>79652</td>
<td>16p13.3</td>
</tr>
<tr>
<td>690</td>
<td>Hs.655162</td>
<td>ZCCHC6</td>
<td>79670</td>
<td>9q21</td>
</tr>
<tr>
<td>691</td>
<td>Hs.458973</td>
<td>ZFHX4</td>
<td>79776</td>
<td>8q21.11</td>
</tr>
<tr>
<td>692</td>
<td>Hs.115497</td>
<td>RERGL</td>
<td>79785</td>
<td>12p12.3</td>
</tr>
<tr>
<td>693</td>
<td>Hs.524479</td>
<td>MMRN2</td>
<td>79812</td>
<td>10q23.2</td>
</tr>
<tr>
<td>694</td>
<td>Hs.665354</td>
<td>ASAM</td>
<td>79827</td>
<td>11q24.1</td>
</tr>
<tr>
<td>695</td>
<td>Hs.180402</td>
<td>ZNF671</td>
<td>79891</td>
<td>19q13.43</td>
</tr>
<tr>
<td>696</td>
<td>Hs.183390</td>
<td>ZNF613</td>
<td>79898</td>
<td>19q13.33</td>
</tr>
<tr>
<td>697</td>
<td>Hs.513296</td>
<td>FLJ14154</td>
<td>79903</td>
<td>16p13.3</td>
</tr>
<tr>
<td>698</td>
<td>Hs.694119</td>
<td>GRRP1</td>
<td>79927</td>
<td>1p36.11</td>
</tr>
<tr>
<td>699</td>
<td>Hs.189652</td>
<td>C7orf58</td>
<td>79974</td>
<td>7q31.31</td>
</tr>
<tr>
<td>700</td>
<td>Hs.522334</td>
<td>SVEP1</td>
<td>79987</td>
<td>9q32</td>
</tr>
<tr>
<td>701</td>
<td>Hs.309849</td>
<td>C14orf159</td>
<td>80017</td>
<td>14q32.12</td>
</tr>
<tr>
<td>702</td>
<td>Hs.390817</td>
<td>MYO15B</td>
<td>80022</td>
<td>17q25.1</td>
</tr>
<tr>
<td>703</td>
<td>Hs.286194</td>
<td>SLC2A6</td>
<td>80024</td>
<td>12q24.13</td>
</tr>
<tr>
<td>704</td>
<td>Hs.513343</td>
<td>ATF7IP2</td>
<td>80063</td>
<td>16p13.13</td>
</tr>
<tr>
<td>705</td>
<td>Hs.654967</td>
<td>ZNF606</td>
<td>80095</td>
<td>19q13.4</td>
</tr>
<tr>
<td>706</td>
<td>Hs.469561</td>
<td>UXS1</td>
<td>80146</td>
<td>2q12.2</td>
</tr>
<tr>
<td>707</td>
<td>Hs.288382</td>
<td>CENPT</td>
<td>80152</td>
<td>16q22.1</td>
</tr>
<tr>
<td>708</td>
<td>Hs.632527</td>
<td>SLC35F5</td>
<td>80255</td>
<td>2q14.1</td>
</tr>
<tr>
<td>709</td>
<td>Hs.167805</td>
<td>EPC1</td>
<td>80314</td>
<td>10p11</td>
</tr>
<tr>
<td>710</td>
<td>Hs.127126</td>
<td>CPEB4</td>
<td>80315</td>
<td>5q21</td>
</tr>
<tr>
<td>711</td>
<td>Hs.173716</td>
<td>ADAM33</td>
<td>80332</td>
<td>20p13</td>
</tr>
<tr>
<td>712</td>
<td>Hs.147434</td>
<td>TRAF3IP3</td>
<td>80342</td>
<td>1q32.3-q41</td>
</tr>
<tr>
<td>713</td>
<td>Hs.221597</td>
<td>SLC19A3</td>
<td>80704</td>
<td>2q37</td>
</tr>
<tr>
<td>714</td>
<td>Hs.120267</td>
<td>TSGA10</td>
<td>80705</td>
<td>2q11.2</td>
</tr>
<tr>
<td>715</td>
<td>Hs.42217</td>
<td>ITFG1</td>
<td>81533</td>
<td>16q12.1</td>
</tr>
<tr>
<td>716</td>
<td>Hs.372123</td>
<td>NDEL1</td>
<td>81565</td>
<td>17p13.1</td>
</tr>
<tr>
<td>717</td>
<td>Hs.656389</td>
<td>PLA2G12A</td>
<td>81579</td>
<td>4q25</td>
</tr>
<tr>
<td>718</td>
<td>Hs.356061</td>
<td>MAP1LC3B</td>
<td>81631</td>
<td>16q24.2</td>
</tr>
<tr>
<td>719</td>
<td>Hs.132599</td>
<td>DOCK8</td>
<td>81704</td>
<td>9p24.3</td>
</tr>
<tr>
<td>720</td>
<td>Hs.169333</td>
<td>TIGD6</td>
<td>81789</td>
<td>5q33.1</td>
</tr>
<tr>
<td>721</td>
<td>Hs.657508</td>
<td>ADAMS10</td>
<td>81794</td>
<td>19p13.3-p13.2</td>
</tr>
<tr>
<td>722</td>
<td>Hs.444229</td>
<td>ARHGAP24</td>
<td>83478</td>
<td>4q21.23-q21.3</td>
</tr>
<tr>
<td>723</td>
<td>Hs.513779</td>
<td>CRISP2L2</td>
<td>83716</td>
<td>16p24.1</td>
</tr>
<tr>
<td>724</td>
<td>Hs.696346</td>
<td>SYT15</td>
<td>83849</td>
<td>10q11.1</td>
</tr>
<tr>
<td></td>
<td>Genomic Location</td>
<td>Protein Name</td>
<td>Gene Name</td>
<td>Hs.</td>
</tr>
<tr>
<td>----</td>
<td>------------------</td>
<td>--------------</td>
<td>-----------</td>
<td>-----</td>
</tr>
<tr>
<td>725</td>
<td>13q14</td>
<td>SETDB2</td>
<td>Hs.631789</td>
<td>83852</td>
</tr>
<tr>
<td>726</td>
<td>9q31</td>
<td>FSD1L</td>
<td>Hs.136901</td>
<td>83856</td>
</tr>
<tr>
<td>727</td>
<td>22q12.1-qter</td>
<td>TBC1D10A</td>
<td>Hs.655273</td>
<td>83874</td>
</tr>
<tr>
<td>728</td>
<td>19p13</td>
<td>USHBP1</td>
<td>Hs.512773</td>
<td>83878</td>
</tr>
<tr>
<td>729</td>
<td>1p31.3</td>
<td>SGIP1</td>
<td>Hs.132121</td>
<td>84251</td>
</tr>
<tr>
<td>730</td>
<td>9q32</td>
<td>HSDL2</td>
<td>Hs.59486</td>
<td>84263</td>
</tr>
<tr>
<td>731</td>
<td>2q12.2</td>
<td>C2orf40</td>
<td>Hs.43125</td>
<td>84417</td>
</tr>
<tr>
<td>732</td>
<td>NA</td>
<td>USP38</td>
<td>Hs.480848</td>
<td>84640</td>
</tr>
<tr>
<td>733</td>
<td>11p11</td>
<td>ACCS</td>
<td>Hs.126706</td>
<td>84680</td>
</tr>
<tr>
<td>734</td>
<td>9p13.3</td>
<td>C9orf24</td>
<td>Hs.50334</td>
<td>84688</td>
</tr>
<tr>
<td>735</td>
<td>2q35</td>
<td>PLCD4</td>
<td>Hs.632528</td>
<td>84812</td>
</tr>
<tr>
<td>736</td>
<td>3p25.3</td>
<td>p25.3-p24.1</td>
<td>IL17RC</td>
<td>Hs.129959</td>
</tr>
<tr>
<td>737</td>
<td>11q23.3</td>
<td>TMEM25</td>
<td>Hs.564188</td>
<td>84866</td>
</tr>
<tr>
<td>738</td>
<td>6q22.33</td>
<td>RSPO3</td>
<td>Hs.135254</td>
<td>84870</td>
</tr>
<tr>
<td>739</td>
<td>1p34.2</td>
<td>MFSD2</td>
<td>Hs.655177</td>
<td>84879</td>
</tr>
<tr>
<td>740</td>
<td>9q34.3</td>
<td>C9orf37</td>
<td>Hs.522520</td>
<td>85026</td>
</tr>
<tr>
<td>741</td>
<td>NA</td>
<td>DIXDC1</td>
<td>Hs.655626</td>
<td>85458</td>
</tr>
<tr>
<td>742</td>
<td>11q22.3</td>
<td>ZC3H12C</td>
<td>Hs.376289</td>
<td>85463</td>
</tr>
<tr>
<td>743</td>
<td>12q13.12</td>
<td>CDDC65</td>
<td>Hs.512805</td>
<td>85478</td>
</tr>
<tr>
<td>744</td>
<td>1q25.2</td>
<td>SEC16B</td>
<td>Hs.149540</td>
<td>89866</td>
</tr>
<tr>
<td>745</td>
<td>6p21.1</td>
<td>KLC4</td>
<td>Hs.655123</td>
<td>89953</td>
</tr>
<tr>
<td>746</td>
<td>15q15.1</td>
<td>CCDC32</td>
<td>Hs.654661</td>
<td>90416</td>
</tr>
<tr>
<td>747</td>
<td>6q27</td>
<td>C6orf176</td>
<td>Hs.31917</td>
<td>90632</td>
</tr>
<tr>
<td>748</td>
<td>9p24.1</td>
<td>IL33</td>
<td>Hs.348390</td>
<td>90865</td>
</tr>
<tr>
<td>749</td>
<td>11q24.2</td>
<td>ESAM</td>
<td>Hs.173840</td>
<td>90952</td>
</tr>
<tr>
<td>750</td>
<td>1p36.11</td>
<td>UBXD5</td>
<td>Hs.145061</td>
<td>91544</td>
</tr>
<tr>
<td>751</td>
<td>19q13.41</td>
<td>MYADM</td>
<td>Hs.380906</td>
<td>91663</td>
</tr>
<tr>
<td>752</td>
<td>14q32.31</td>
<td>WDR20</td>
<td>Hs.36859</td>
<td>91833</td>
</tr>
<tr>
<td>753</td>
<td>12q24.32</td>
<td>TMEM132C</td>
<td>Hs.49599</td>
<td>92293</td>
</tr>
<tr>
<td>754</td>
<td>17q23.3</td>
<td>LYK5</td>
<td>Hs.514402</td>
<td>92335</td>
</tr>
<tr>
<td>755</td>
<td>10q25.2</td>
<td>LOC92482</td>
<td>Hs.651480</td>
<td>92482</td>
</tr>
<tr>
<td>756</td>
<td>10q11.21</td>
<td>ANUBL1</td>
<td>Hs.89029</td>
<td>93550</td>
</tr>
<tr>
<td>757</td>
<td>17p11.2</td>
<td>MYOCD</td>
<td>Hs.567641</td>
<td>93649</td>
</tr>
<tr>
<td>758</td>
<td>Xq13.1</td>
<td>ACRC</td>
<td>Hs.135167</td>
<td>93953</td>
</tr>
<tr>
<td>759</td>
<td>19q13.2</td>
<td>EGLN2</td>
<td>Hs.515417</td>
<td>112398</td>
</tr>
<tr>
<td>760</td>
<td>22q12.2</td>
<td>PIK3IP1</td>
<td>Hs.26670</td>
<td>113791</td>
</tr>
<tr>
<td>761</td>
<td>15q22.1</td>
<td>LACTB</td>
<td>Hs.410388</td>
<td>114294</td>
</tr>
<tr>
<td>762</td>
<td>NA</td>
<td>LRRC37B</td>
<td>Hs.514071</td>
<td>114659</td>
</tr>
<tr>
<td>763</td>
<td>4p16.1</td>
<td>MRFAP1L1</td>
<td>Hs.593159</td>
<td>114932</td>
</tr>
<tr>
<td>764</td>
<td>8q11.23</td>
<td>PCMTD1</td>
<td>Hs.308480</td>
<td>115294</td>
</tr>
<tr>
<td>765</td>
<td>7p22.3</td>
<td>GPR146</td>
<td>Hs.656731</td>
<td>115330</td>
</tr>
<tr>
<td>766</td>
<td>14q12</td>
<td>DHRS1</td>
<td>Hs.348350</td>
<td>115817</td>
</tr>
<tr>
<td>767</td>
<td>8q22.2</td>
<td>OSR2</td>
<td>Hs.253247</td>
<td>116039</td>
</tr>
<tr>
<td>768</td>
<td>20p13</td>
<td>HSPA12B</td>
<td>Hs.516854</td>
<td>116835</td>
</tr>
<tr>
<td>769</td>
<td>11q13.2</td>
<td>HRASLS5</td>
<td>Hs.410316</td>
<td>117245</td>
</tr>
<tr>
<td>770</td>
<td>6p21.31</td>
<td>IHPK3</td>
<td>Hs.17253</td>
<td>117283</td>
</tr>
<tr>
<td></td>
<td>Hs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>771</td>
<td>529984</td>
<td>TAGAP</td>
<td>117289</td>
<td>6q25.3</td>
</tr>
<tr>
<td>772</td>
<td>162963</td>
<td>ANTXR2</td>
<td>118429</td>
<td>4q21.21</td>
</tr>
<tr>
<td>773</td>
<td>656887</td>
<td>CPXM2</td>
<td>119587</td>
<td>10q26.13</td>
</tr>
<tr>
<td>774</td>
<td>657163</td>
<td>TARSL2</td>
<td>123283</td>
<td>15q26.3</td>
</tr>
<tr>
<td>775</td>
<td>371690</td>
<td>C18orf51</td>
<td>125704</td>
<td>18q22.3</td>
</tr>
<tr>
<td>776</td>
<td>534538</td>
<td>HSPB6</td>
<td>126393</td>
<td>19q13.12</td>
</tr>
<tr>
<td>777</td>
<td>524767</td>
<td>ZNF684</td>
<td>127396</td>
<td>1p34.2</td>
</tr>
<tr>
<td>778</td>
<td>44277</td>
<td>LRRRC39</td>
<td>127495</td>
<td>1p21.2</td>
</tr>
<tr>
<td>779</td>
<td>269546</td>
<td>FLJ40298</td>
<td>129852</td>
<td>2p16.2</td>
</tr>
<tr>
<td>780</td>
<td>233998</td>
<td>BBS5</td>
<td>129880</td>
<td>2q31.1</td>
</tr>
<tr>
<td>781</td>
<td>591615</td>
<td>RFTN2</td>
<td>130132</td>
<td>2q33.1</td>
</tr>
<tr>
<td>782</td>
<td>123933</td>
<td>OSR1</td>
<td>130497</td>
<td>2p24.1</td>
</tr>
<tr>
<td>783</td>
<td>534540</td>
<td>ZFAND2B</td>
<td>130617</td>
<td>2q35</td>
</tr>
<tr>
<td>784</td>
<td>40808</td>
<td>TMEM178</td>
<td>130733</td>
<td>2p22.1</td>
</tr>
<tr>
<td>785</td>
<td>230601</td>
<td>DNAJC19</td>
<td>131118</td>
<td>3q26.33</td>
</tr>
<tr>
<td>786</td>
<td>390823</td>
<td>IL17RE</td>
<td>132014</td>
<td>3p25.3</td>
</tr>
<tr>
<td>787</td>
<td>656937</td>
<td>CPEB2</td>
<td>132864</td>
<td>4p15.33</td>
</tr>
<tr>
<td>788</td>
<td>297814</td>
<td>ENPP6</td>
<td>133121</td>
<td>4q35.1</td>
</tr>
<tr>
<td>789</td>
<td>661876</td>
<td>LOC134466</td>
<td>134466</td>
<td>5q33.1</td>
</tr>
<tr>
<td>790</td>
<td>368203</td>
<td>DOCK11</td>
<td>139818</td>
<td>Xq24</td>
</tr>
<tr>
<td>791</td>
<td>591712</td>
<td>ASB5</td>
<td>140458</td>
<td>4q34.2</td>
</tr>
<tr>
<td>792</td>
<td>27453</td>
<td>RAB40A</td>
<td>142684</td>
<td>Xq22.1</td>
</tr>
<tr>
<td>793</td>
<td>210586</td>
<td>C13orf31</td>
<td>144811</td>
<td>13q14.11</td>
</tr>
<tr>
<td>794</td>
<td>144696</td>
<td>C14orf50</td>
<td>145376</td>
<td>14q23.3</td>
</tr>
<tr>
<td>795</td>
<td>658619</td>
<td>TMED6</td>
<td>146456</td>
<td>16q22.1</td>
</tr>
<tr>
<td>796</td>
<td>11782</td>
<td>MGC45438</td>
<td>146556</td>
<td>16p13.3</td>
</tr>
<tr>
<td>797</td>
<td>657197</td>
<td>C18orf18</td>
<td>147525</td>
<td>18p11.31</td>
</tr>
<tr>
<td>798</td>
<td>651111</td>
<td>ZNF565</td>
<td>147929</td>
<td>19q13.12</td>
</tr>
<tr>
<td>799</td>
<td>511848</td>
<td>ZNF569</td>
<td>148266</td>
<td>19q13.12</td>
</tr>
<tr>
<td>800</td>
<td>591445</td>
<td>SAMD13</td>
<td>148418</td>
<td>1p31.1</td>
</tr>
<tr>
<td>801</td>
<td>177744</td>
<td>PM20D1</td>
<td>148811</td>
<td>1q32.1</td>
</tr>
<tr>
<td>802</td>
<td>593721</td>
<td>LOC149448</td>
<td>149448</td>
<td>1q43</td>
</tr>
<tr>
<td>803</td>
<td>116254</td>
<td>C22orf15</td>
<td>150248</td>
<td>22q11.23</td>
</tr>
<tr>
<td>804</td>
<td>368312</td>
<td>FAM109B</td>
<td>150368</td>
<td>22q13.2</td>
</tr>
<tr>
<td>805</td>
<td>516176</td>
<td>SMYD1</td>
<td>150572</td>
<td>2p11.2</td>
</tr>
<tr>
<td>806</td>
<td>655700</td>
<td>ANKAR</td>
<td>150709</td>
<td>2q32.2</td>
</tr>
<tr>
<td>807</td>
<td>493819</td>
<td>C9orf19</td>
<td>152007</td>
<td>9p13-p12</td>
</tr>
<tr>
<td>808</td>
<td>484195</td>
<td>C5orf41</td>
<td>153222</td>
<td>5q35.2</td>
</tr>
<tr>
<td>809</td>
<td>289293</td>
<td>C8orf42</td>
<td>157695</td>
<td>8p23.3</td>
</tr>
<tr>
<td>810</td>
<td>190043</td>
<td>MOSPD2</td>
<td>158747</td>
<td>Xp22.2</td>
</tr>
<tr>
<td>811</td>
<td>42572</td>
<td>ALDH1L2</td>
<td>160428</td>
<td>12q23.3</td>
</tr>
<tr>
<td>812</td>
<td>13854</td>
<td>PPTC7</td>
<td>160760</td>
<td>12q24.11</td>
</tr>
<tr>
<td>813</td>
<td>525307</td>
<td>CLEC14A</td>
<td>161198</td>
<td>14q21.1</td>
</tr>
<tr>
<td>814</td>
<td>522545</td>
<td>ZNF791</td>
<td>163049</td>
<td>19p13.2-p13.13</td>
</tr>
<tr>
<td>815</td>
<td>65256</td>
<td>LGI4</td>
<td>163175</td>
<td>19q13.12-19q13.11</td>
</tr>
<tr>
<td>816</td>
<td>681239</td>
<td>CLEC4F</td>
<td>165530</td>
<td>2p13.3</td>
</tr>
<tr>
<td>No.</td>
<td>Gene ID</td>
<td>Gene Symbol</td>
<td>Chro.</td>
<td>Band</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>-------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>817</td>
<td>Hs.699317</td>
<td>PRICKLE2</td>
<td>16</td>
<td>16636</td>
</tr>
<tr>
<td>818</td>
<td>Hs.132087</td>
<td>KLHDC6</td>
<td>16</td>
<td>16638</td>
</tr>
<tr>
<td>819</td>
<td>Hs.159006</td>
<td>ZNF800</td>
<td>17</td>
<td>16885</td>
</tr>
<tr>
<td>820</td>
<td>Hs.385493</td>
<td>C10orf128</td>
<td>17</td>
<td>17037</td>
</tr>
<tr>
<td>821</td>
<td>Hs.655519</td>
<td>SYNPO2</td>
<td>17</td>
<td>17102</td>
</tr>
<tr>
<td>822</td>
<td>Hs.527874</td>
<td>CCDC131</td>
<td>17</td>
<td>19644</td>
</tr>
<tr>
<td>823</td>
<td>Hs.532469</td>
<td>PAOX</td>
<td>17</td>
<td>19674</td>
</tr>
<tr>
<td>824</td>
<td>Hs.443935</td>
<td>TTC21A</td>
<td>19</td>
<td>19922</td>
</tr>
<tr>
<td>825</td>
<td>Hs.301243</td>
<td>TIGD4</td>
<td>20</td>
<td>20179</td>
</tr>
<tr>
<td>826</td>
<td>Hs.133337</td>
<td>RWDD4A</td>
<td>20</td>
<td>20196</td>
</tr>
<tr>
<td>827</td>
<td>Hs.199777</td>
<td>RANBP3L</td>
<td>20</td>
<td>20215</td>
</tr>
<tr>
<td>828</td>
<td>Hs.482625</td>
<td>CMYA5</td>
<td>20</td>
<td>20233</td>
</tr>
<tr>
<td>829</td>
<td>Hs.585069</td>
<td>STK32A</td>
<td>20</td>
<td>20237</td>
</tr>
<tr>
<td>830</td>
<td>Hs.28780</td>
<td>ZNF449</td>
<td>20</td>
<td>20352</td>
</tr>
<tr>
<td>831</td>
<td>Hs.42400</td>
<td>USP12</td>
<td>20</td>
<td>21933</td>
</tr>
<tr>
<td>832</td>
<td>Hs.204947</td>
<td>PLAC9</td>
<td>20</td>
<td>21934</td>
</tr>
<tr>
<td>833</td>
<td>Hs.118513</td>
<td>MRGPRF</td>
<td>20</td>
<td>21992</td>
</tr>
<tr>
<td>834</td>
<td>Hs.668747</td>
<td>FLJ32682</td>
<td>20</td>
<td>22008</td>
</tr>
<tr>
<td>835</td>
<td>Hs.607594</td>
<td>FAM13C1</td>
<td>20</td>
<td>22096</td>
</tr>
<tr>
<td>836</td>
<td>Hs.147440</td>
<td>ZNF485</td>
<td>20</td>
<td>22099</td>
</tr>
<tr>
<td>837</td>
<td>Hs.427449</td>
<td>LOC221091</td>
<td>22</td>
<td>22109</td>
</tr>
<tr>
<td>838</td>
<td>Hs.412103</td>
<td>EFHA1</td>
<td>22</td>
<td>22115</td>
</tr>
<tr>
<td>839</td>
<td>Hs.25391</td>
<td>Pi16</td>
<td>22</td>
<td>22147</td>
</tr>
<tr>
<td>840</td>
<td>Hs.519904</td>
<td>RBM24</td>
<td>22</td>
<td>22166</td>
</tr>
<tr>
<td>841</td>
<td>Hs.596587</td>
<td>C7orf38</td>
<td>22</td>
<td>22178</td>
</tr>
<tr>
<td>842</td>
<td>Hs.587427</td>
<td>HOXA11S</td>
<td>22</td>
<td>22183</td>
</tr>
<tr>
<td>843</td>
<td>Hs.368944</td>
<td>JAZF1</td>
<td>22</td>
<td>22189</td>
</tr>
<tr>
<td>844</td>
<td>Hs.200100</td>
<td>C7orf41</td>
<td>22</td>
<td>22166</td>
</tr>
<tr>
<td>845</td>
<td>Hs.522863</td>
<td>C1orf15A</td>
<td>24</td>
<td>24612</td>
</tr>
<tr>
<td>846</td>
<td>Hs.407926</td>
<td>RICTOR</td>
<td>25</td>
<td>25326</td>
</tr>
<tr>
<td>847</td>
<td>Hs.339024</td>
<td>MSRB3</td>
<td>25</td>
<td>25382</td>
</tr>
<tr>
<td>848</td>
<td>Hs.693749</td>
<td>ZDHHC20</td>
<td>25</td>
<td>25383</td>
</tr>
<tr>
<td>849</td>
<td>Hs.346575</td>
<td>C19orf26</td>
<td>25</td>
<td>25505</td>
</tr>
<tr>
<td>850</td>
<td>Hs.435515</td>
<td>LOC255167</td>
<td>25</td>
<td>25516</td>
</tr>
<tr>
<td>851</td>
<td>Hs.163451</td>
<td>LOC255275</td>
<td>25</td>
<td>25527</td>
</tr>
<tr>
<td>852</td>
<td>Hs.22575</td>
<td>BCL6B</td>
<td>25</td>
<td>25587</td>
</tr>
<tr>
<td>853</td>
<td>Hs.591401</td>
<td>KANK3</td>
<td>25</td>
<td>25694</td>
</tr>
<tr>
<td>854</td>
<td>Hs.503500</td>
<td>OLFML1</td>
<td>28</td>
<td>28329</td>
</tr>
<tr>
<td>855</td>
<td>Hs.560343</td>
<td>LOC283666</td>
<td>28</td>
<td>28366</td>
</tr>
<tr>
<td>856</td>
<td>Hs.569669</td>
<td>LOC283901</td>
<td>28</td>
<td>28390</td>
</tr>
<tr>
<td>857</td>
<td>Hs.664267</td>
<td>FLJ36208</td>
<td>28</td>
<td>28394</td>
</tr>
<tr>
<td>858</td>
<td>Hs.437191</td>
<td>PTRF</td>
<td>28</td>
<td>28412</td>
</tr>
<tr>
<td>859</td>
<td>Hs.400688</td>
<td>IZUMO1</td>
<td>28</td>
<td>28435</td>
</tr>
<tr>
<td>860</td>
<td>Hs.303669</td>
<td>SLC25A42</td>
<td>28</td>
<td>28443</td>
</tr>
<tr>
<td>861</td>
<td>Hs.567816</td>
<td>FAM126B</td>
<td>28</td>
<td>28517</td>
</tr>
<tr>
<td>862</td>
<td>Hs.518059</td>
<td>C3orf64</td>
<td>28</td>
<td>28520</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>863</td>
<td>Hs.559386</td>
<td>LOC285286</td>
<td>285286</td>
<td>3p14.2</td>
</tr>
<tr>
<td>864</td>
<td>Hs.476999</td>
<td>CCDC66</td>
<td>285331</td>
<td>3p14.3</td>
</tr>
<tr>
<td>865</td>
<td>Hs.449206</td>
<td>LOC285359</td>
<td>285359</td>
<td>3q12.3</td>
</tr>
<tr>
<td>866</td>
<td>Hs.588682</td>
<td>SUMF1</td>
<td>285362</td>
<td>3p26.2</td>
</tr>
<tr>
<td>867</td>
<td>Hs.399980</td>
<td>LOC285550</td>
<td>285550</td>
<td>4p15.33-p15.32</td>
</tr>
<tr>
<td>868</td>
<td>Hs.480371</td>
<td>LOC285556</td>
<td>285556</td>
<td>4q23</td>
</tr>
<tr>
<td>869</td>
<td>Hs.403594</td>
<td>EFHA2</td>
<td>286097</td>
<td>8p22</td>
</tr>
<tr>
<td>870</td>
<td>Hs.496530</td>
<td>MGC39900</td>
<td>286527</td>
<td>Xq22.2</td>
</tr>
<tr>
<td>871</td>
<td>Hs.146059</td>
<td>TUSC5</td>
<td>286753</td>
<td>17p13.3</td>
</tr>
<tr>
<td>872</td>
<td>Hs.700799</td>
<td>LOC339290</td>
<td>339290</td>
<td>18p11.31</td>
</tr>
<tr>
<td>873</td>
<td>Hs.471067</td>
<td>LOC339483</td>
<td>339483</td>
<td>1p35.1</td>
</tr>
<tr>
<td>874</td>
<td>Hs.146730</td>
<td>KY</td>
<td>339855</td>
<td>3q22.2</td>
</tr>
<tr>
<td>875</td>
<td>Hs.379754</td>
<td>TMEM173</td>
<td>340061</td>
<td>5q31.2</td>
</tr>
<tr>
<td>876</td>
<td>Hs.444834</td>
<td>RSPO2</td>
<td>340419</td>
<td>8q23.1</td>
</tr>
<tr>
<td>877</td>
<td>Hs.21249</td>
<td>ZC3H12B</td>
<td>340554</td>
<td>Xq11.1</td>
</tr>
<tr>
<td>878</td>
<td>Hs.369380</td>
<td>MGC40069</td>
<td>348035</td>
<td>14q11.2</td>
</tr>
<tr>
<td>879</td>
<td>Hs.208673</td>
<td>NMNAT3</td>
<td>349565</td>
<td>3q23</td>
</tr>
<tr>
<td>880</td>
<td>Hs.595458</td>
<td>MAST4</td>
<td>375449</td>
<td>5q12.3</td>
</tr>
<tr>
<td>881</td>
<td>Hs.704486</td>
<td>LOC387647</td>
<td>387647</td>
<td>10p11.23</td>
</tr>
<tr>
<td>882</td>
<td>Hs.32478</td>
<td>FIBIN</td>
<td>387758</td>
<td>11p14.2</td>
</tr>
<tr>
<td>883</td>
<td>Hs.131035</td>
<td>GLTPD2</td>
<td>388323</td>
<td>17p13.2</td>
</tr>
<tr>
<td>884</td>
<td>Hs.657260</td>
<td>RP13-401N8.2</td>
<td>388358</td>
<td>20p11.21-p11.1</td>
</tr>
<tr>
<td>885</td>
<td>Hs.204449</td>
<td>ZNF470</td>
<td>388566</td>
<td>19q13.43</td>
</tr>
<tr>
<td>886</td>
<td>Hs.435013</td>
<td>VGLL3</td>
<td>389136</td>
<td>3p12.1</td>
</tr>
<tr>
<td>887</td>
<td>Hs.658041</td>
<td>MGC21881</td>
<td>389741</td>
<td>9q21.11</td>
</tr>
<tr>
<td>888</td>
<td>Hs.561708</td>
<td>FLJ45244</td>
<td>400242</td>
<td>14q32.13</td>
</tr>
<tr>
<td>889</td>
<td>Hs.153827</td>
<td>C14orf180</td>
<td>400258</td>
<td>14q32.33</td>
</tr>
<tr>
<td>890</td>
<td>Hs.187134</td>
<td>LOC400464</td>
<td>400464</td>
<td>15q26.3</td>
</tr>
<tr>
<td>891</td>
<td>Hs.641441</td>
<td>LOC400604</td>
<td>400604</td>
<td>17q21.33</td>
</tr>
<tr>
<td>892</td>
<td>Hs.61508</td>
<td>LOC400657</td>
<td>400657</td>
<td>18q22.3</td>
</tr>
<tr>
<td>893</td>
<td>Hs.668085</td>
<td>C1orf220</td>
<td>400798</td>
<td>1q25.2</td>
</tr>
<tr>
<td>894</td>
<td>Hs.173705</td>
<td>LOC401152</td>
<td>401152</td>
<td>4q26</td>
</tr>
<tr>
<td>895</td>
<td>Hs.131064</td>
<td>KLHL31</td>
<td>401265</td>
<td>6p12.1</td>
</tr>
<tr>
<td>896</td>
<td>Hs.561708</td>
<td>LOC401320</td>
<td>401320</td>
<td>7p15.1</td>
</tr>
<tr>
<td>897</td>
<td>Hs.461247</td>
<td>MRC1L1</td>
<td>414308</td>
<td>10p12.33</td>
</tr>
<tr>
<td>898</td>
<td>Hs.530380</td>
<td>FAM116B</td>
<td>414918</td>
<td>22q13.33</td>
</tr>
<tr>
<td>899</td>
<td>Hs.536319</td>
<td>IQSEC3</td>
<td>440073</td>
<td>12p13.33</td>
</tr>
<tr>
<td>900</td>
<td>Hs.512963</td>
<td>ALG11</td>
<td>440138</td>
<td>13q14.2</td>
</tr>
<tr>
<td>901</td>
<td>Hs.449880</td>
<td>LOC440434</td>
<td>440434</td>
<td>17q12</td>
</tr>
<tr>
<td>902</td>
<td>Hs.641142</td>
<td>FLJ46875</td>
<td>440918</td>
<td>2q24.1</td>
</tr>
<tr>
<td>903</td>
<td>Hs.507676</td>
<td>FLJ12993</td>
<td>441027</td>
<td>4q21.22</td>
</tr>
<tr>
<td>904</td>
<td>Hs.510098</td>
<td>C6orf217</td>
<td>441171</td>
<td>6q23.3</td>
</tr>
<tr>
<td>905</td>
<td>Hs.559067</td>
<td>ARMETL1</td>
<td>441549</td>
<td>10p13</td>
</tr>
<tr>
<td>906</td>
<td>Hs.647105</td>
<td>GIMAP6</td>
<td>474344</td>
<td>NA</td>
</tr>
<tr>
<td>907</td>
<td>Hs.596537</td>
<td>C2orf64</td>
<td>493753</td>
<td>2q11.2</td>
</tr>
<tr>
<td>908</td>
<td>Hs.661883</td>
<td>PGA4</td>
<td>643847</td>
<td>11q12.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>909</td>
<td>Hs.380698</td>
<td>WIPF3</td>
<td>644150</td>
<td>7p15.1</td>
</tr>
<tr>
<td>910</td>
<td>Hs.58690</td>
<td>LOC644192</td>
<td>644192</td>
<td>15q26.2</td>
</tr>
<tr>
<td>911</td>
<td>Hs.575741</td>
<td>LOC644431</td>
<td>644431</td>
<td>NA</td>
</tr>
<tr>
<td>912</td>
<td>Hs.683930</td>
<td>LOC644554</td>
<td>644554</td>
<td>19q13.12-q13.13</td>
</tr>
<tr>
<td>913</td>
<td>Hs.693822</td>
<td>LOC645431</td>
<td>645431</td>
<td>14q23.3</td>
</tr>
<tr>
<td>914</td>
<td>Hs.463652</td>
<td>LOC645638</td>
<td>645638</td>
<td>17q23.1</td>
</tr>
<tr>
<td>915</td>
<td>Hs.444950</td>
<td>LOC652968</td>
<td>652968</td>
<td>22q12</td>
</tr>
<tr>
<td>916</td>
<td>Hs.632434</td>
<td>LOC653513</td>
<td>653513</td>
<td>1q21.1</td>
</tr>
<tr>
<td>917</td>
<td>Hs.659982</td>
<td>LOC654841</td>
<td>654841</td>
<td>2q36-q37</td>
</tr>
<tr>
<td>918</td>
<td>Hs.655534</td>
<td>LOC728190</td>
<td>728190</td>
<td>10q23.2</td>
</tr>
<tr>
<td>919</td>
<td>Hs.535639</td>
<td>D2HGDH</td>
<td>728294</td>
<td>2q37.3</td>
</tr>
<tr>
<td>920</td>
<td>Hs.662127</td>
<td>RP11-592B15.4</td>
<td>728407</td>
<td>10q11.23</td>
</tr>
<tr>
<td>921</td>
<td>Hs.559428</td>
<td>SERF1B</td>
<td>728492</td>
<td>5q13.2</td>
</tr>
<tr>
<td>922</td>
<td>Hs.559827</td>
<td>LOC729096</td>
<td>729096</td>
<td>10q22.2</td>
</tr>
<tr>
<td>923</td>
<td>Hs.557608</td>
<td>LOC729178</td>
<td>729178</td>
<td>6q24.3</td>
</tr>
<tr>
<td>924</td>
<td>Hs.591387</td>
<td>KIAA1881</td>
<td>729359</td>
<td>19p13.3</td>
</tr>
<tr>
<td>925</td>
<td>Hs.371576</td>
<td>FAM18A</td>
<td>780776</td>
<td>16p13.13</td>
</tr>
<tr>
<td>926</td>
<td>Hs.1581</td>
<td>DDTL</td>
<td>100037417</td>
<td>22q11.23</td>
</tr>
</tbody>
</table>
REFERENCES


Bisoffi, M., C. M. Heaphy, et al. (2006). "Telomeres: prognostic markers for solid tumors." Int J Cancer a) all genes changing their expression, b) upregulated genes only and c) downregulated genes only. (10): 2255-60.


Mann, H. B. and D. R. Whitney (1947). "On a test of whether one of two random variables is stochastically larger than the other." Annals of Mathematical Statistics 18: 50-60.


200


Ganiraju Manyam was born and raised in the state of Andhra Pradesh in southern India. He graduated from Gandhi Centenary high school and then completed his higher secondary school at Pragathi Junior college in Kakinada, India. He attended Bharatiar University in Coimbatore where he received B.E degree in Computer Science & Engineering. He received his MS in Bioinformatics from International Institute of Information Technology, Hyderabad, India, in 2004. He was a Junior Research Fellow at the Centre for Cellular and Molecular Biology, India for a period of two years. He is now working on his PhD in Biosciences at Molecular and Microbiology Department, College of Science, George Mason University, USA. His research interests are in the areas of cancer genomics, data mining, machine learning and Bioinformatics.