miRNA expression landscapes in stem cells, tissues and cancer

Volkan Cakir[1*], Henry Wirth[1,2*], Lydia Hopp[1,2] and Hans Binder[1,2#]

[1] Interdisciplinary Centre for Bioinformatics, University of Leipzig, Germany
[2] Leipzig Research Center for Civilization Diseases; Universität Leipzig

* contributed equally
# corresponding author
E-mail: binder@izbi.uni-leipzig.de

Key words: gene expression analysis, self organizing maps, miRNA/mRNA coexpression, gene set enrichment analysis

Abstract
miRNAs play critical roles in the regulation of gene expression with two major functions: degrading mRNA in a sequence specific manner or repressing its translation. Publicly available data sets on miRNA and mRNA expression in embryonal and induced stem cells, human tissues and solid tumors are analyzed in this case study using self organizing maps (SOM) to characterize miRNA expression landscapes in the context of cell fate commitment, tissue specific differentiation and its dysfunction in cancer. The SOM portraits of the individual samples clearly reveal groups of miRNA specifically overexpressed without the need of additional pairwise comparisons between the different systems. Sets of miRNA differentially over and underexpressed in different systems are provided. The individual portraying of the expression landscapes enables a very intuitive, image-based perception which clearly promotes the discovery of qualitative relationships between the systems studied. We see perspectives for broad applications of this method in standard analysis to many kinds of high throughput data of single miRNA and especially combined miRNA/mRNA data sets.

1. Introduction

miRNAs play critical roles in the regulation of gene expression. Just as the set of transcription factors (TFs) in a given cell type constitutes a 'code' that specifies cellular differentiation via mRNA activity, so 'miRNA codes' are likely to have conceptually similar roles in the regulation of gene activity (1). Both TFs and miRNA are trans-acting factors that exert their activity through composite cis-regulatory elements. The resulting miRNA/mRNA coexpression might come true either in concert, in anti-concert or also in a more complex fashion. Here, the two major functions of miRNAs are degrading the mRNA or repressing its translation. The situation is usually complex because a particular miRNA will have multiple mRNA targets and because multiple miRNAs could target the same mRNA.

The most widely studied mechanism of regulation involves binding of a miRNA to the target mRNA. miRNAs regulate their targets by triggering mRNA degradation or translational repression. As a result, translation of the target mRNA is inhibited and the mRNA may be destabilized. The inhibitory effects of miRNAs have been linked to diverse cellular processes including malignant proliferation, apoptosis, development, differentiation, and metabolic processes. The negative relationship between miRNAs and their targets suggests that the regulatory effect of a miRNA could be determined from the expression levels of its targets. Note also that miRNA do not need a perfect alignment with their targets to act. In consequence one miRNA may regulate several mRNAs or one mRNA may be regulated by several miRNA. This way miRNAs potentially regulate approximately 30% of all genes encoding human proteins and appear to interfer in a wide range of cell functions, such as cell generation, differentiation, and proliferation.

The opposite effect, namely direct correlations between miRNA and mRNA expression can be caused by the fact that intronic miRNAs are usually coordinately expressed with their host gene mRNA, implying that they derive from a common transcript, and that analysis of host gene expression can be used to probe the spatial and temporal localization of intronic miRNA (2). In addition, also proximal pairs of miRNA are often coexpressed. It was found that an abrupt transition in the correlation between pairs of expressed miRNA occurs at a distance of 50 kb, implying that miRNA separated by less than 50 kb typically derive from a common transcript.

Exact understanding of how miRNA regulate gene expression is vital to the field of miRNA research. Systematic analysis of miRNA expression landscapes and also of miRNA/mRNA coexpression patterns thus constitute a basal objective in miRNA research which complements miRNA gene and target discovery. In this contribution we analyzed miRNA expression using self-organizing maps, a machine learning clustering technique based on neural network. The method was described in detail in the accompanying chapter (3). In this chapter we apply the SOM portraying method to discover combined miRNA and mRNA expression landscapes in the context of cell fate decisions and stemness, fully differentiated human tissues and diseased cancer samples of different origin in order to illustrate the performance of the method in form of an extended case study. To our best knowledge this method was applied here for the first time to analyze miRNA expression data.

2. Analyzing miRNA/mRNA coexpression – a short overview

In general, miRNA/mRNA coexpression can be studied in two different ways: Firstly, correlation between the expression values of miRNA and mRNA species are directly analyzed. For example, in such studies on human brain biopsies the authors report that the distribution of correlation coefficients for all possible mRNA-miRNA pairs exceeds a random distribution at their tails at high positive and negative values of the correlation coefficient (4). Part of the negative correlations selected tends to predict targets and positive correlations tend to predict physically proximate pairs as expected (see (3) and references cited therein). In contrast, other studies report that miRNA activity shows very weak correlation with mRNA expression which indicates more complex regulation mechanisms between miRNAs and their target genes (5).

The second type of coexpression studies pursues a more indirect approach based on databases which collect miRNA-mRNA target relationships (see overview in (6)). These data are obtained either via in-silico target prediction of miRNA binding motifs (see data bases: PITA (7), PICTAR (8) and TargetScan (9)), via meta analyses of experimentally validated miRNA target genes (TarBase (10)), miRecords (11) and miR2Disease (11)) or via text mining of biomedical abstracts (miRSel (12)). The

collected data constitute sets of mRNA target species for individual miRNA (or families of miRNA) which are subsequently used in gene expression enrichment analyses. This approach searches for significant expression changes of the target sets compared with appropriate random sets (13). Enriched sets suggest that their expression is potentially regulated by the associated miRNA.
The computational methods predict about 40 mRNA targets per miRNA on the average while only 2 – 8 targets are associated with each miRNA in the experimentally validated and text mining data bases. The latter approaches usually consider less miRNA in total (from 93 in miRecords to 176 in miR2Disease) than the former ones (from 163 in PICTAR to 640 in PITA). In silico miRNA target prediction is usually not very accurate with fairly high false positive and/or false negative rates.

3. Methods and data

3.1. Expression analysis using self organizing maps

SOM is a neural network algorithm widely used to categorize large, high-dimensional data sets (14). In bioinformatics, SOM has been successfully applied to gene expression analysis (15) enabling characterization of genome wide expression landscapes in a sample specific way (16-17). Our implementation of the method, called SOM-cartography or SOM-portraying, transforms large and heterogeneous sets of expression data into an atlas of sample-specific portraits which can be directly compared in terms of similarities and dissimilarities (see (3) for a detailed description of the method). This global view on the behavior of defined modules of correlated and differentially expressed genes is more intuitive than ranked lists of hundreds or thousands of individual genes usually obtained in standard expression analysis. Particularly, SOM analysis is featured by several important benefits: (i) it provides an individual visual identity for each sample; (ii) it reduces the dimension of the original data; (iii) it preserves the information richness of the molecular portraits allowing the detailed, multivariate explorative comparisons between samples, and (iv) its output can be treated as a new, complex object for next level analysis in terms of visual recognition. Here we will apply the method to different data sets to portray the miRNA expression landscapes, to characterize the similarities between the different samples studied and to extract lists of miRNA relevant in the context of stem cells, differentiated tissues and cancer. miRNA data are complemented with associated mRNA expression landscapes.

3.2. Data sets

We analyzed following three data sets:
The *WILSON-data* set refers to expression data of 697 miRNA from embryonic stem cells (ESC), fibroblasts and derived induced pluripotent stem cells (IPS) obtained in a microarray study (18).
The *LIANG-tissue* set contains expression values of 175 miRNA measured in 24 human tissues by standard TaqMan qPCR assay (19). Expression values are given in units of the threshold cycle (CT) defined as the fractional cycle number at which the fluorescence exceeds a fixed threshold. Four human miRNAs (miR-30e, miR-92, miR-92N, and miR-423) that were least variable among the tissues in this study were used to normalize the miRNA expression. SOM training was performed using log CT values after quantile normalization. The obtained miRNA expression landscapes in human tissues were compared with mRNA expression data analyzed previously (17).
The *LU-cancer* set contains miRNA and mRNA measurements from the same samples of seven healthy and tumor tissues (colon, kidney, bladder, prostate, uterus, lung, breast) (20). miRNA were measured using a bead-based profiling method: Oligonucleotide-capture probes complementary to miRNAs of interest were coupled to carboxylated 5-micron polystyrene beads impregnated with variable mixtures of two fluorescent dyes for ‘barcoding’, each representing a single miRNA. The abundance of 217 miRNAs was measured after hybridization, amplification by PCR and staining. mRNA expression was determined using microarrays.

4. Case studies

4.1. Stem cells

miRNAs are important regulators for embryonic stem cell (ESC) self-renewal, pluripotency and differentiation (21). The SOM portraits of human fibroblasts, pluripotent stem cells induced from them (IPS) and ESC show a simple and consistent spot pattern (Figure 4.1a and e): Overexpressed genes in fibroblasts and underexpressed in both types of stem cells (ESC and IPS) aggregate into one spot A. Genes with antagonistic activity cluster into spot B showing high expression in ESC and IPS and low expression in somatic cells. IPSs reveal another moderately overexpressed spot C in addition to this 'stemness' spot which differentiates between both types of stem cells. It becomes underexpressed in ESC and fibroblasts.

The sample similarity tree (Figure 4.1b) reveals close similarity but not identity between the miRNA expression landscapes of ESC and IPS which, in turn, clearly differs from that of the somatic fibroblasts. The cell-type specific spots A – C contain about one to two dozen miRNAs per spot. More than 400 miRNAs remain not regulated and cluster within the invariant central area of the map (see population map, Figure 4.1c).

Analysis of the spots shows that miRNA of the mir-302 and -17 families are upregulated in ESC and IPS (spot B) and miRNAs of the let-7 family are upregulated in fibroblasts (spot A) in agreement with the results of ref. (18) (see Table 4.1). It has been argued that increased reprogramming in response to let-7 inhibition is mediated by let-7 target genes, such as c-Myc and Lin28 (21). Lin28 is also repressed by miR-125, which is abundantly expressed in differentiated cells and changes in concert with let-7 in our data set. Presumably inhibiting the activity of both miR-125 and let-7 miRNA may result in additional beneficial effects during reprogramming, due to robust activation of Lin28 expression. Another miRNA from spot A, mir-145, induces ESC differentiation by inhibiting the expression of key pluripotency/reprogramming factors, such as Sox2, Oct4, Klf4, and c-Myc. Other miRNAs from spot A (mir-24, mir-23 and mir-21) either inhibit cell proliferation by targeting important cell cycle regulators, such as c-Myc, and E2F2 (mir-24) or suppress TGF-β/Activin signaling. miRNAs from the mir-302 family are found in spot B showing antagonistic activation compared with spot A. It has been demonstrated that Sox2 and Oct4 bind the miR-302 promoter and are essential for expression of miR-302 in human ESC.

Hence, our SOM analysis identifies signature groups of miRNA in ESC, IPS and differentiated fibroblasts. The SOM portraits clearly assign the expression landscapes to one of the tree cell types studied. A similar pattern was observed for mRNA expression (22).

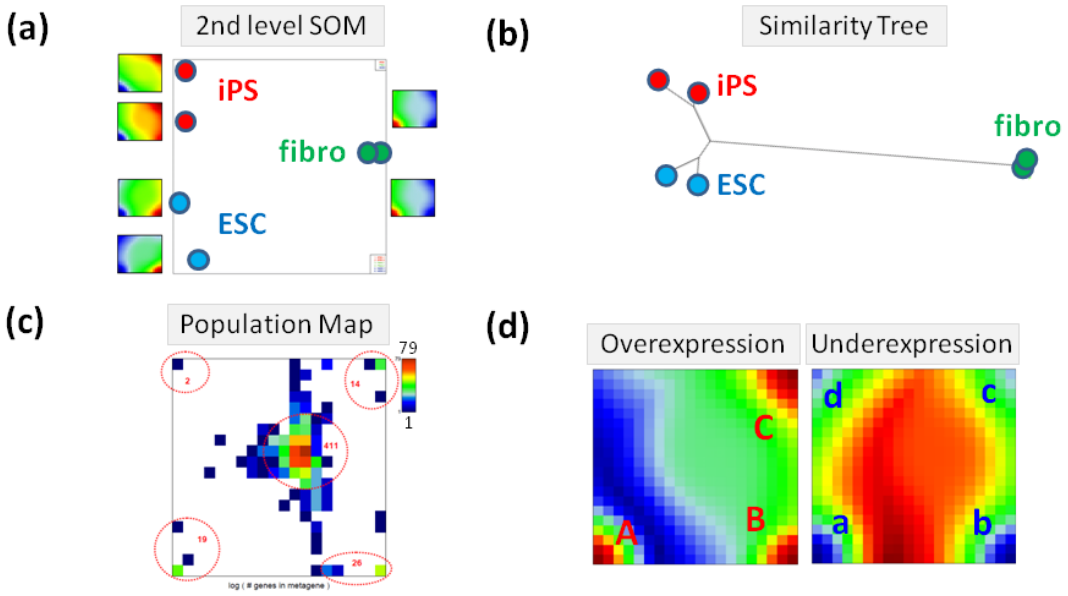

Figure 4.1: miRNA expression portraying of stem cells (WILSON-data set): 2nd level SOM arrange the 1st level SOM portraits (see the small images) of stem cells (ESC and IPS) and of differentiated cells (fibroblasts) according to their mutual similarities (panel a). The two mosaics per cell system refer to biological replicates (18). The similarity tree (panel b) expresses the Euclidean distances between the different samples in-scale. The population map (panel c) illustrates the number of miRNA per tile in the 1st level SOM mosaic. The over- and underexpression summary maps (panel d) assign the spots detected.

Table 4.1: miRNA regulated in stem cells

spot[a]	UP[b]	DN[b]	miRNA within the spot cluster[c]
A	fibro	ESC, iPS	let-7d; -let-7f; -let-7e; -let-7a; -145; -100; -let-7i; -29a; -let-7c; -199a-3p; -125b; -143; -222; -23b; -23a; -24; -34c-3p; -21; -26a
B	ESC, (iPS)	Fibro	mir-106a; -302a*; -17; -302a; -302d; -302b; -638; -93; -106b; -20a; -302c*; -19b; -25; -663; -19a; -103; -130a; -107; -20b; -16; -182; -183
C	iPS	fibro, (ESC)	mir-149*; -18b; -92b; -92a; -30c; -15b; -18a; -151-5p; -30b; -923; -302c; -148a; -363; -200c; -361-5p; -335; -454; -1
d	fibro	ESC	mir-27a; -27b; -125a-5p
	Invariant		mir-768-5p; -15b*; -150*; -219-2-3p; -517c; -504; -612; -629*; -30a*; -652; -519d; -342-5p; -330-3p; -624*; -193b*; -296-3p; -93*; -187*; -196a

[a] see Figure 4.1e for spot assignment

[b] systems showing up- or downregulation of expression compared with the mean expression of each miRNA

[c] mirRNA are sorted with decreasing concordance

4.2. Human tissues

The expression of 345 human miRNA was studied in a spectrum of 40 normal human tissues that included specimens derived from brain, muscle, circulatory, respiratory, lymphoid, gastrointestinal, urinary, reproductive, and endocrine systems. Their SOM portraits in Figure 4.2 reveal a much more diverse spot texture compared with the simple pattern observed in stem and somatic cells discussed in the previous subsection. The different tissues are grouped into eleven tissue categories in analogy with the classification of human tissues in a previous study on mRNA expression (17). Portraits of the same category mostly look very similar, hence reflecting similar miRNA expression landscapes. For example, tissues derived from different parts of heart (atrium versus ventricle) resemble that of skeletal muscle. Portraits of tissues from the gastrointestinal system ('digestion': stomach, small intestine, and colon), immune system (spleen and lymph node), female sexual organs (ovary, uterus, and cervix), and respiratory/epithelial tissues (lung and trachea) show mostly consistent spot patterns in each category.

Similarity-tree analysis identifies five larger clusters of tissues containing muscle, gastrointestinal, epithelial, brain and sexual organs which group along different branches of the tree (see Figure 4.3, the 'plusses' indicate that these clusters usually include also tissues of other categories such as adipose tissue which are found in the' muscle^{+}' branch of the tree). The similarities between the samples can be attributed to characteristic groups of spots collected in the respective summary map (see Figure 4.4,a). Spot analysis characterizes the miRNA expression landscape more in detail. The spot expression heatmap (Figure 4.4b) shows the mean expression of each of the spots in all tissues: For example, high expression levels of spots B and C are observed in colon samples and high expression of spots A, E and F in muscle samples. Over- and underexpression spot abundance analysis provides further details (Figure 4.4c): E.g., spot B collects miRNAs which are highly expressed in colon and immune system tissues whereas spot C is populated more with miRNAs which are highly expressed in many tissues, however with slight preference for colon and epithelium. In turn, miRNAs from both spots B and C are underexpressed in adipose, muscle and nervous tissues.

One also sees that, e.g., miRNAs highly expressed in adipose tissue accumulates within spot A only, whereas endocrine tissues show a broad distribution of highly expressed miRNAs over different spots. The 'adipose' spot A is also observed in a series of other tissues such as pericardium (muscle), bladder (epithelium), ovary, uterus and cervix (all sexual organs) reflecting accumulation of fat in different regions of the body. Contrarily, miRNA in spot A are strongly down regulated in fat-poor tissues such as blood, brain, thymus and, to a less degree, kidney.

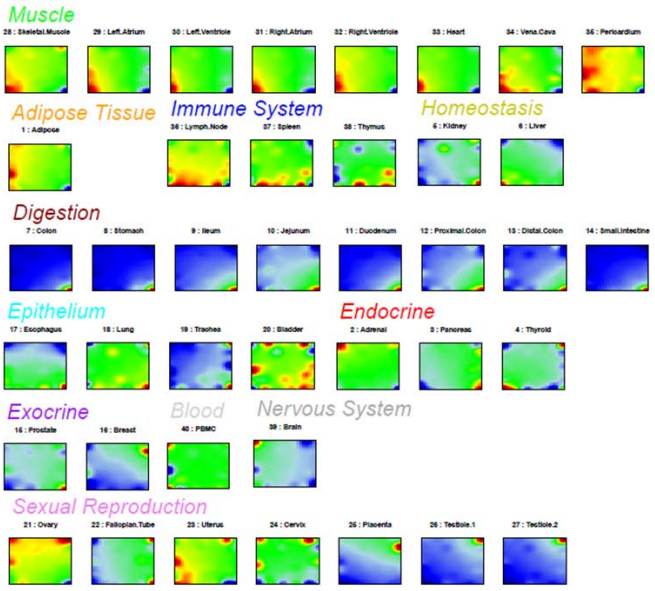

Figure 4.2: miRNA expression portraits of human tissues (LIANG-data set). Tissues are grouped into 11 categories.

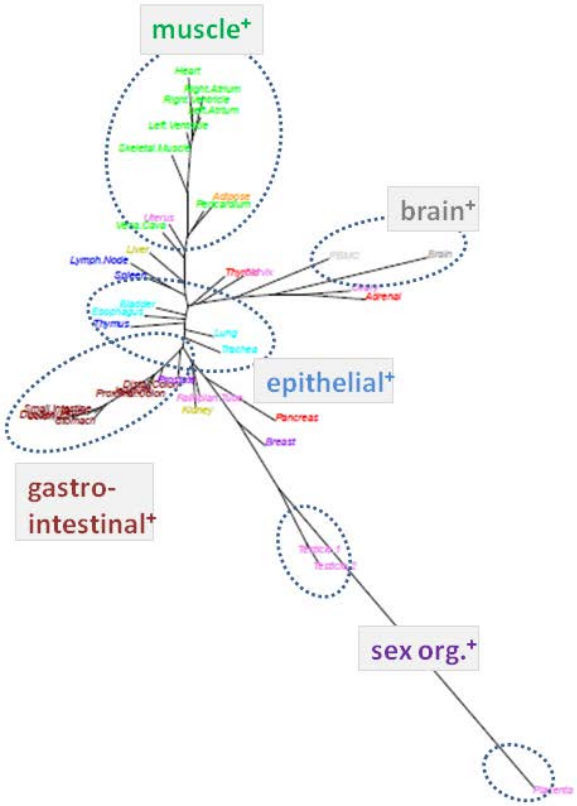

Figure 4.3: miRNA expression in human tissues (LIANG-data set). The neighbor joining tree illustrates similarities between the SOM expression portraits. One finds essentially five clusters which accumulate along different branches as indicated by the dotted ellipses.

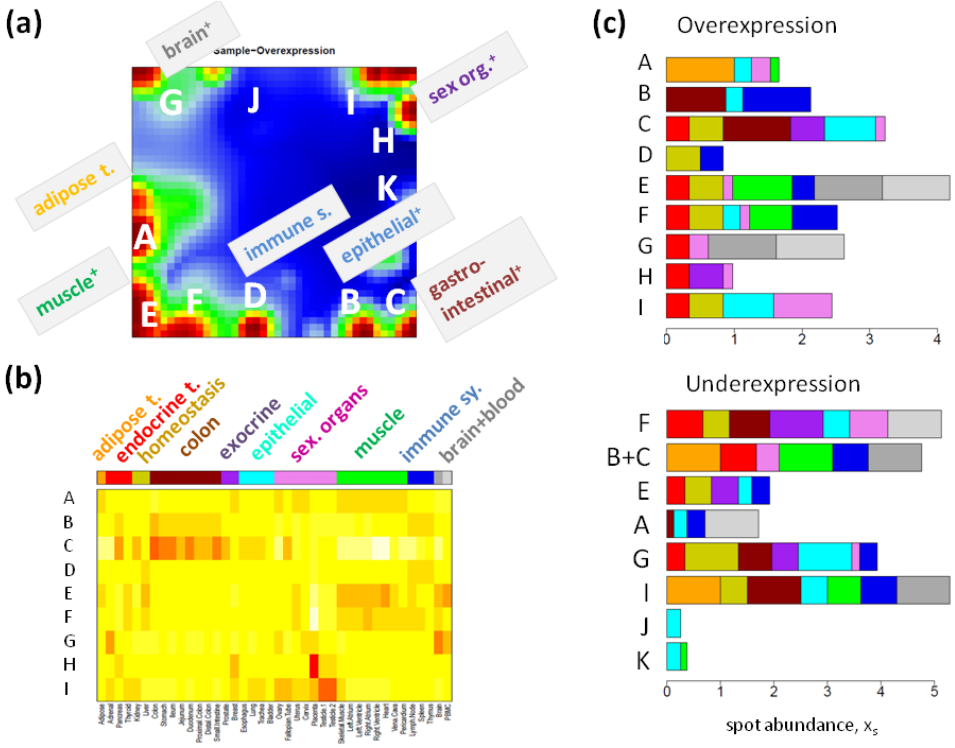

Figure 4.4: Spot characteristics of miRNA expression in human tissues (LIANG-data set): The overexpression spot summary map shows that each cluster is characterized by specific spot patterns (panel a). The spot expression heatmap (panel b) shows the mean expression of the spots detected in all tissues (red-to-white indicates high-to-low). Over- and underexpression spot abundance in ten tissue categories (panel c, the colors of the bars are assigned in the legend on top of the heatmap). For example, spot A is especially overexpressed in adipose tissue (orange bar) which, in turn, is characterized by underexpression of spots B, C and I.

The population and variance maps in Figure 4.5 reveal that the expression of 34 miRNAs remains almost invariant in the tissues studied. Most of the miRNA (54 species) are found in spots I and H upregulated in sexual organs, endocrine and epithelial tissues. The variance of miRNA expression is maximal in placenta due to extraordinarily high expression of genes from spot H (Figure 4.5c). The miRNA species found in each of the spots are listed in Table 4.2.

According to the particular spot abundance one can identify miRNAs expressed in specific tissues with minimal or no expression in other tissues such as miR-9/219 in brain (spot G) whereas miR-124a/124b (spot E) are also strongly expressed in muscle. miRNAs of the let-7 family are also found in this partly ubiquitous spot which also contains miR-1 and miR-133a/b showing high expression in different parts of the heart and skeletal muscle as well as in vena cava, and thyroid. These are hollow organs composed of smooth muscle-containing wall, such as the gastrointestinal system, suggesting that expression of miR-1 and miR-133a/b might mark some features shared by different muscle types.

Note also that selected miRNA of the mir-302 family highly expressed in ESC are among the invariant species in the tissue series. Our spot analysis thus provides an opportunity to extract tissue-specific expression patterns of groups of miRNA which can be further analyzed with respect to the genomic location of their coding sequences and to their involvement into common pathway activities.

Figure 4.6 shows the clustering pattern of mRNA expression in different tissues as seen by our previous SOM analysis (17). On one hand, part of the tissue categories such as muscle or nervous tissues form also separate branches due to category-specific mRNA expression. On the other hand, similarities between tissues and also tissue categories are different compared with the miRNA expression patterns (compare with Figure 4.3). For example, miRNA expression of brain shares partly similarity with miRNA expression in muscle (spot E) whereas mRNA expression in nervous tissue is highly specific and different from mRNA expression in muscle. Note however that the tissue data sets used for mRNA and miRNA analysis are partly different with respect to the tissues included which may distort direct comparison. Figure 4.7 matches the tissues used in both data sets. One sees, for example, that nervous tissues are highly underrepresented in the LIANG-miRNA data set. On the other hand, other tissue categories such as muscle, digestion and immune systems tissues are well represented in both data sets.

Figure 4.7 also compares the tissue-specifics of the miRNA and mRNA expression landscapes in terms of their metagene entropies. It reveals interesting differences: miRNA expression is more specific for epithelial and immune systems tissues (low entropy) than the respective mRNA expression patterns. For endocrine and reproductive organs this relation reverses.

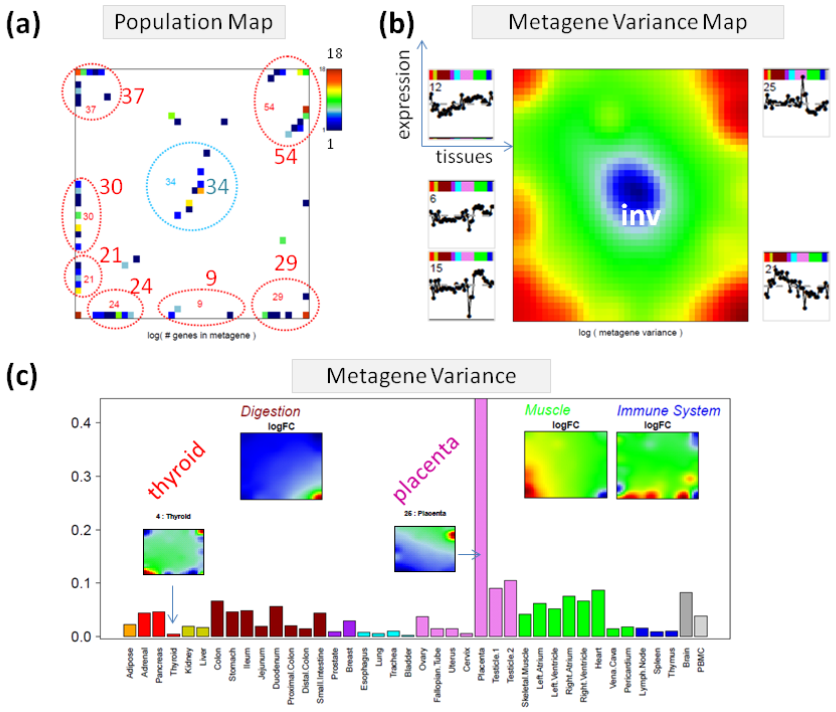

Figure 4.5: Population and variance map of expression portraits of human tissues (panel a and b, respectively) and their tissue-specific metagene variance (panel c). miRNAs with highly variant expression profiles accumulate in the spots located along the border of the map (the number of miRNAs per spot are given in the population map). 34 miRNAs with almost invariant profiles form the central blue spot in the variance map. Selected spot profiles are shown nearby the respective spots. The metagene variance of placenta is maximal due to one strong overexpression spot H (see Figure 4.4 and Figure 4.5).

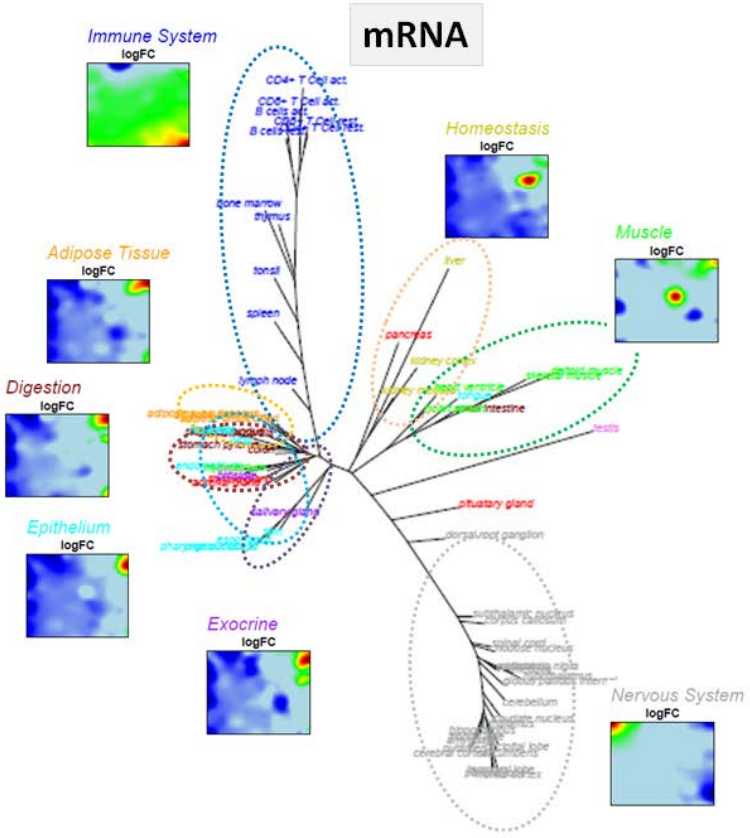

Figure 4.6: mRNA expression in human tissues. The neighbor joining tree illustrates similarities between the SOM metagene expression of 67 different tissues grouped into 8 categories. They preferentially accumulate along specific branches as indicated by the ellipses. The respective mean portraits are shown in the figure (for the full gallery of all single tissue portraits see ref. (17)). Compare with the miRNA-expression tree of human tissues in Figure 4.4.

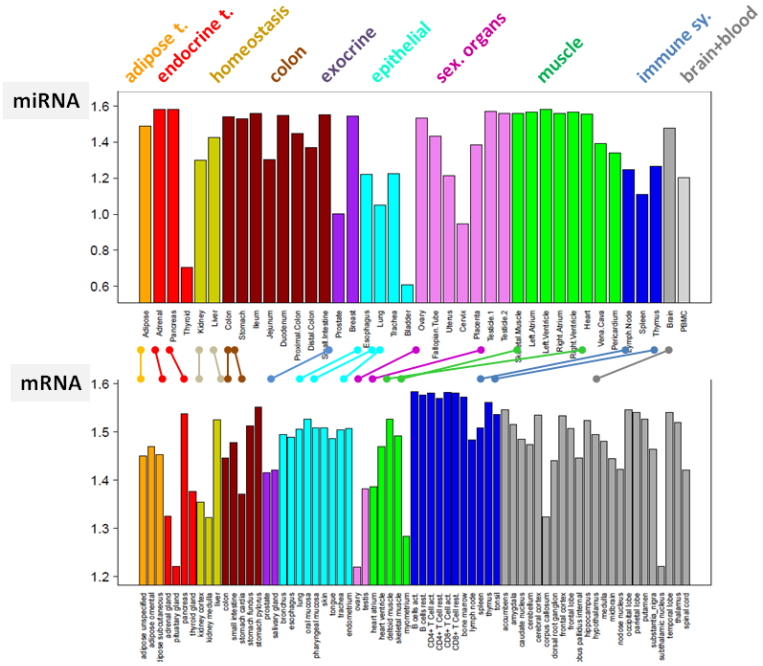

Figure 4.7: Metagene entropies of the miRNA (top) and mRNA (bottom) expression portraits of human tissues. Identical tissues in both parts of the figure are connected by the 'dumbbell'-lines.

Table 4.2: miRNA regulated in human tissues

spot[a]	group of tissues[b]	UP[c]	DN[c]	miRNA within the spot[d]
A	muscle$^+$, adipose	adipose, cervix	blood, thymus, placenta	mir-224; -452*; -452; -193b; -335; -365; -362; -199b; -27aN; -425; let-7a_control; -23a; -126*; -214; -27a; -126; -152; -374; -27b
B	epithelial$^+$, gastro-intestinal$^+$	immune system, digestion	trachea, thyroid	mir-142-3p; -338; -363; -146; -146b; -20b; -301
C	gastro-intestinal$^+$	epithel, digestion, kidney, prostate	muscle, adipose, nervous system	mir-215; -375; -192.1; -194; -192; -141.1; -141N; -141; -200a; -200c; -200b; -200cN; -200bN; -31.1; -31; -429
D	epithelial$^+$, immune s.	homeostasis, immune system	thyroid, trachea, prostate, cervix	mir-92N; -17-5p; -183; -92; -18; -15b; -106a; -15a; -25; -106b
E	muscle$^+$	muscle, homeostasis; blood, brain	Immune system, sexual repr., exocrine	mir-122a; -124b; -124a; cel-124; -30e-3p; -133a; -302d; -302b; -1; -302a; -422a; -128a; -133b; -197; -491; -378; -328; -181b.1; -138; -128b; -340; -190; -423; -490; -129; -107; -422b; -425.1; -22; -296, cel-2, cel-lin-4, -159a; -104; -108; -105; -136
F	muscle$^+$	muscle, immune system, homeostasis	blood, digestion, sexual repr., exocrine, kidney	mir-188.1; -95; -150; -155; -181c; -511; -139.1; -505; -342; -101; -193; -345; let-7i; let-7e; let-7iN; -17-3p; -324-3p; let-7f

G	brain+	blood, brain	thyroid, homeostasis, digestion, breast	mir-9; -9*; -137; -323; -433; -219; -204; -153; -485-3p; -432; -127; -382; -370; -383; -149; -134; -379; -132; -299; -125b
H	sex. organs+	Immune system, placenta, breast	thymus, trachea, prostate, bladder	mir-517c; -516-5p; -518c; -519d; -512-3p; -520h; -518e; -525*; -520g; -517b; -517a; -519c; -518b; -518a; -518f; -515-3p; -515-5p; -525; -520a*; -512-5p; -519e*; -520d; -372
I	sex. organs+	sexual repr., endothelial	adipose, digestion,liver, spleen, nervous tissues	mir-514; -508; -206; -509; -449; -506; -34c; -202*; -202; -34cN; -34bN; -34b.1; -510; -196a; -513; -196b; -424; -135a; -432*; -10b; -199a; -34aN; -199-s
J		bladder	digestion, trachea	mir-451
K		bladder	trachea, thyroid	mir-29a.1
inv				mir-448; -453; -488; -492; -496; -498; -504; -518f*; -519a; -520d*; -410; -412; -524*; -526c; -527; -384; -371; -377; -517*; -326; -329; -33; -337; -368; -373*; -376b; -380-3p; -380-5p; -450; -503; -526b*; -325; -325N; -381; -302c*; -96; -519b; -302b*; -220; -299-3p; -302a*; -208

[a] see Figure 4.3 for spot assignment

[b] tissue categories

[c] systems showing up- or downregulation of expression compared with the mean expression of each miRNA

[d] mirRNA are sorted with decreasing Pearssons correlation between their expression profile and that of the respective metagene

4.3. Cancer

Cancer is a disease of gene function in most respects caused by genetic and epigenetic alterations affecting many molecular pathways involving both canonical protein-coding ‘mRNA’-genes as well as noncoding ‘miRNA’-genes. Aberrant miRNA expression signatures can serve as a hallmark of cancer where miRNA-genes can function as oncogenes and tumor supressors. Global up- (23) as well as downregulation (20) of miRNA activity in cancer compared with normal tissues has been reported previously. Thus, an evaluation of changes in miRNA expression could provide insight into mechanisms of cancer genesis and progression.

In ref. (3) we identified groups of miRNAs differently expressed in healthy and tumor tissues. Table 4.3 lists the miRNA taken from the spot-clusters extracted from the miRNA-SOM images. Many of them are previously reported as differentially expressed in different cancers (compare with Table 1 in (24), Table 1 in (25) and also refs. (26-27)). Upregulated miRNA in cancer act as OncoMiRs whereas miRNA acting as tumor supressors are often downregulated. For example, miRNA of the let-7, mir-126 and mir-130 families are known tumor supressors. They accumulate in spot A which is downregulated in colon, bladder, lung and breast cancer. Spot F collects mir-34: miR-34a inhibits the expression of multiple oncogenes (e.g., c-Met, Notch-1/Notch-2 and CDK6) by binding to their 3’-UTR and suppressing, e.g., tumor growth in human gliomas. Many miRNA found in the T_UP (tumor up) spot H are known OncoMiRs (e.g., mir-181, -200, -146). Other OncoMiRs such as mir-10 and -196 are specifically upregulated in the tumor-specific spot C. Another upregulated miRNA in spot H, mir-296, is related to angionesis commonly activated in many solid tumors (25).

The analogous analysis provides groups of mRNAs together with their functional context using gene set enrichment analysis. Table 4.4 lists the top-enriched gene sets taken from the gene ontology category ‘biological process’ in the overexpression spots of the mRNA SOM portraits together with the top-10 concordant genes in each spot.

In the next step we combined both data sets in terms of spot-spot correlation analysis as described in the methodical section of ref. (3). The pairwise correlation heatmap in Figure 4.8 visualizes positive and negative correlations in red and blue color, respectively. Each negative correlation might refer to downregulation of miRNA and upregulation of mRNA in cancer or vice versa. The blue tiles thus

include miRNA acting as tumor suppressors (indicated by minuses in Figure 4.8) and as oncoMiRs (plusses) as well. The miRNA regulated in the respective spots of the SOM portraits and the top-enriched gene set in the respective mRNA spots are given in Figure 4.8. For example, tumor suppressor miRNA of the let-7, mir-126 and mir-130 families are clearly identified together with upregulated mRNA showing enrichment of processes such as 'translation' and 'receptor activity'. OncoMiRs such as mir-10, -196 and -128 are related to mRNA downregulation in the context of 'translation' and 'proteolysis'.

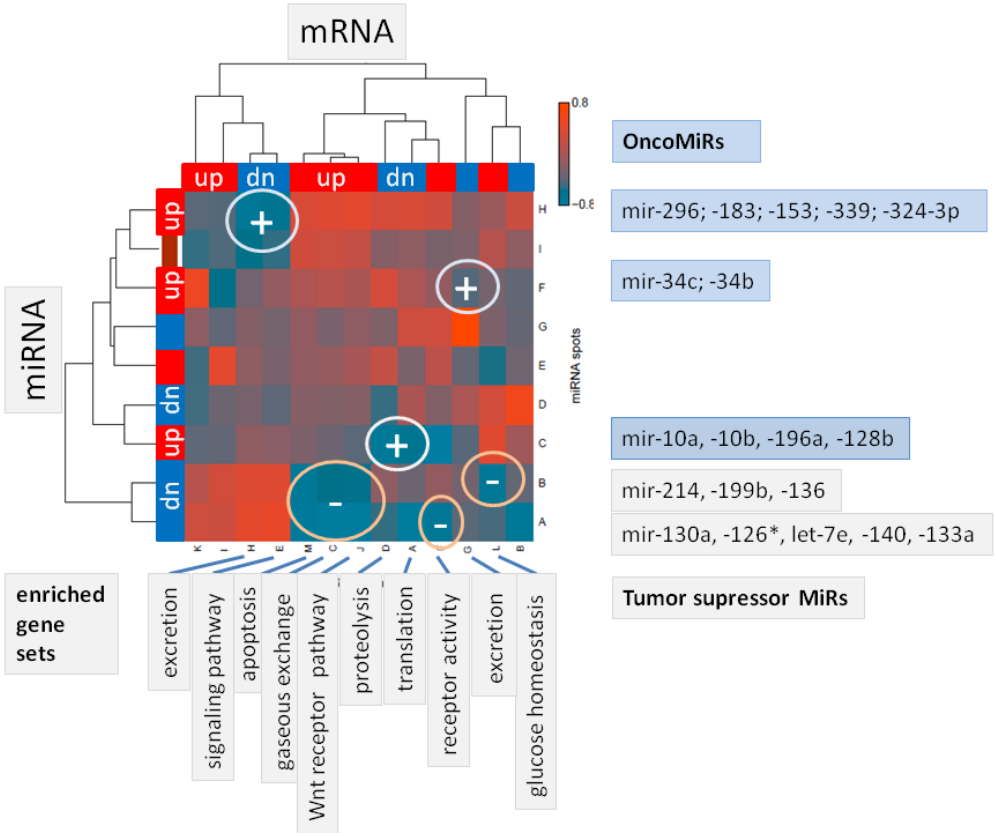

Figure 4.8: Pairwise miRNA/mRNA-spot covariance map of the LU-data set. The spots are taken from the respective spot summary maps of miRNA and mRNA overexpression shown in (3). The map color codes the covariance of all pairwise combinations of spots using their mean metagene expression profiles (blue to red refers to -0.8 <cov <+0.8). The spot letters are assigned in Figure 3.2. Their preferential up- or downregulation in cancer is indicated by the bars at the left and top borders of the map by red or blue, respectively. Minuses and plusses indicate potential tumor suppressors (miRNA_DN and mRNA_UP) and oncoMiRs (miRNA_UP and mRNA_DN), respectively. The respective top-ranked miRNAs and enriched gene sets are taken from Table 4.3 and Table 4.4, respectively.

As second option of joint miRNA and mRNA analysis we trained the SOM based on the combined covariance-features (cov-analysis, see (3) for details). It quantifies the degree of concerted miRNA and mRNA expression in each sample. Each spot in the obtained portraits thus refers to miRNA/mRNA metagene pairs which, in turn, are each associated with lists of single miRNA and mRNA. Table 4.5 provides the particular miRNAs collected in each of the 'antiexpression' spots identified in the combined cov-portraits shown in ref. (3). The mRNA-species collected in each of the spots are characterized using gene sets enrichment analysis using gene sets of the gene ontology category 'biological process'.

Note that spots of the cov-map refer to combined miRNA- and mRNA-metagenes, each representing lists of single miRNA and mRNA genes. One and the same metagene related list of miRNAs (and of mRNAs) can be found in different spots: For example, mir-1, -133a and -99a appear together in antiexpression spots c, f, g, i and l (as indicated by {...} in Table 4.5) however in combination with different mRNA metagenes and thus with different functional themes such as 'glycogen biosynthesis'

(spot c), 'protein glycosylation via asparagine' (f and g), 'tissue development' (i) and 'fibrinolysis' (f). On the other hand, similar lists of mRNA giving rise to enrichment of genes of the set 'protein glycosylation via asparagine' appear in combination with different miRNA-lists such as {mir-196a; -128b; -10b; -10a; -190} (spot e), {mir-1, -133a and -99a } (f and g) and {-100; -125b; -195} (g). Finally, the cov-spots of antiexpression are predominantly observed either in cancer (e.g., spots c and i) or in healthy (f and g) tissues. Hence, the combined SOM provides very detailed and diverse view on interrelations between miRNA and mRNA expressions. Its resolution clearly exceeds that of the spot-spot correlation analysis discussed above.

Table 4.3: miRNA regulated in normal and tumor samples (LU-data set).

spot[a]	[b]	UP[c]	DN[c]	miRNA within the spot[d]
H	T_UP	T_bladder, _uterus; _lung; _breast; _kidney; (_colon)	all N_tissues	mir-296; -183; -153; -339; -324-3p; -181c; -182; -200b; -208; -146a; -200a; -141; -210; -328
E	T_UP	T_bladder, _prostate, _lung	N_colon, _kidney	mir-205; -144; -323
F	T_UP	T_uterus; N_lung		mir-34c; -34b
C	T_UP	T_kidney		mir-10a, -10b, -196a, -128b
A	T_DN	N_uterus, N_lung	T_colon, _bladder, _lung, _breast	mir-130a, -126*, let-7e, -140, -133a, let-7d, -99a, -1, -126, -189
B	T_DN	N_prostate, N_breast	T_kidney,	mir-214, -199b, -136
D	T_DN	N_colon, N_kidney, T_colon	T_bladder, T_prostate, T_uterus, T_lung, T_breast	mir-194, -215, -192
G	T_DN	N_bladder		mir-216, -217

[a] spot-letters are assigned in the miRNA summary maps given in ref. (3)
[b] up-or downregulated mainly in tumor (T) and normal (N) tissue. T_DN spots are usually also N_UP spots and vice versa.
[c] particular samples showing this spot with up- (UP) or down- (DN) regulated genes
[d] miRNAs are ordered with decreasing significance according to the concordance t-score

Table 4.4: mRNA regulated in cancer (LU-data set)

spot[a]	[b]	UP[c]	DN[c]	enriched gene sets in the spot[d]	top-10 concordant genes in the spot [e]
F	T_UP	T_kidney, T_bladder, T_uterus, T_lung, T_breast	T_prostate, N_prostate	receptor activity (-6); keratinization (-5); cation transport (-4);	CDH15, LTK, SYN1, CCL22, DDX11, ARFRP1, HLA-DOA, NHLH1, PSMB6, PTPRN
L	T_UP	T_bladder	T_colon	glucose homeostasis (-4); defense response to Gram−negative bacterium (-4)	KIF14, RP4-669P10.16, RBL1, CDC7, PSG7, ADD2, HIST1H4E, SAG, CLCN5, CD44
J	T_UP	T_lung		Wnt receptor signaling pathway (-5); negative regulation of cyclin−dependent protein kinase activity (-4); humoral immune response (-4)	ZBTB33, FMO3, HBQ1, MMP10, PIM2, SERPINC1, HTR2C, POU2AF1, CCR6, KIAA0368
M	T_UP	T_uterus		apoptosis (-4)	ATIC, WFDC2, AIMP2, MFSD10, IRAK1, LTBR, NPIPL3, TUBG1, FLAD1, IER3
C	T_DN	T_lung, N_lung	T_colon, T_prostate, N_prostate	respiratory gaseous exchange (-8); response to hyperoxia (-7); complement activation, classical pathway (-5)	MNDA; SFTPA1; PRR4; SFTPC; SFTPA2; SFTPB; NRGN; IL1RL1; SCGB1A1; ITGB2

a1	T_DN	T_colon	T_lung, N_colon	translation (-10); viral transcription (-6); mRNA metabolic process (-5)	NFKB1, RPS16, ATP6V1G2-DDX39B, RP11-40H20.2, EEF1A1P11, SRSF3, HMGN2P17, SERINC3, PTGES3, ACTN1
a2	T_DN	N_uterus, N_lung N_breast	T_colon, T_kidney, T_bladder, T_uterus, T_breast	platelet degranulation (-9); muscle contraction (-6); complement activation (-6)	SPARCL1, CNN3, PURA, UQCRB, NFE2L2, LEPROT, SAFB, TJP1, FNTA, DSTN, ATP2A2
D	T_DN			proteolysis (-11); digestion (-5); lipid catabolic process (-4)	PNLIPRP2, CTRB2, REG3A, PRSS1, CEL, CELA2A, CPA1, PNLIP, FGL1, AMYP1
E	T_DN	N_kidney, N_bladder	T_prostate	muscle cell homeostasis (-4); intracellular receptor mediated signaling pathway (-4)	MCAM; SPEG; MFAP5; CAMK2G; TLE2; MCAM; PDLIM7; MATN2; MYOC; LEFTY2
B	T_DN	T_prostate, T_breast	T_colon, T_kidney, T_breast	pituitary gland development (-5); nucleosome assembly (-5); cholesterol biosynthetic process (-4)	CD38; ALDH1A3; SC5DL; IDI1; ACPP; CCK; PPP3CA; KLK2:MSMB; TGFB3
G	T_DN	T_bladder	T_colon, T_prostate	Excretion (-7); oxidation−reduction process (-6); metabolic process (-6)	CLCNKB; CYP4A11;RIPK1; SLC1A6; MST1P9; SERPINF2; PLG CYFIP2; HRG; FGB
H	T_DN	T_lung	T_prostate	cellular response to organic cyclic compound (-5); excretion (-4); midgut development (-4)	STIL; TTLL12; DRD3; TOMM34; CKMT1B; EEA1; GUCA2B; MEP1A; HNF1A; CCBL1

[a] spot-letters are assigned in the mRNA summary maps given in (3)
[b] up-or downregulated mainly in tumor (T). T_DN spots are usually N_UP spots and vice versa.
[c] particular samples showing this spot with up- (UP) or down- (DN) regulated genes
[d] top enriched gene sets of the Gene Ontology category 'Biological Process'. The number in the brackets is the logged p-value according to Fishers exact test, e.g. -6 means $p \sim 10^{-6}$.
[e] top genes in the list of most concordant mRNAs according to the concordance t-score

Table 4.5: Combined miRNA/mRNA expression in tumor and normal tissues (LU-data set)

spot[a]	N_DN[b]	T_DN[b]	miRNA in the spot[c]	enriched gene sets in the spot [d]
a	N_bladder		{mir-196a; -128b; -30a-3p; -30b; -30c; -10b; -10a; -190}; -204; -135a; -135b; -124a; -187	cellular protein metabolic process; viral transcription; translational termination; aerobic respiration; translation
b		T_prostate	{mir-199b; -136; -214; -199a; -199a*}; -17-3p	cholesterol metabolic process; hemidesmosome assembly
c		T_lung	{mir-1; -133a; -99a}; {-199b; -136; -214; -199a; -199a*}; -338; -100; -125b; -195; -130a; -189; {let-7d; let-7e; -126*; -140; -126; -101}; -215; -194; -192; -106b; -142-3p; -142-5p	neuromuscular synaptic transmission; positive regulation of glycogen biosynthetic process; DNA damage response, signal transduction by p53 class mediator resulting in induction of apoptosis; ion transport; synaptic transmission, cholinergic; transferrin transport; smoothened signaling pathway; ATP hydrolysis coupled proton transport; transport; proton transport
d	N_breast		{mir-205; -323; -144; -10a; -215; -194; -192}	hemidesmosome assembly; RNA splicing; protein N-linked glycosylation via asparagine
e	N_prostate		{mir-196a; -128b; -10b; -10a; -190}	protein N-linked glycosylation via asparagine
f	N_lung		{mir-1; -133a; -99a}; {-100; -125b; -195; -130a; -189}	protein N-linked glycosylation via asparagine
g	N_uterus		{mir-1; -133a; -99a}; -205;{ -100; -125b; -195}; -323; -144; -200a; -200b; -141	protein N-linked glycosylation via asparagines; cellular protein metabolic process; Leydig cell differentiation
h	N_lung	T_breast	{mir-205; -323; -144; -215; -194; -192}; -200a; -200b; -141	aerobic respiration; response to reactive oxygen species

i		T_uterus	{mir-1; -133a; -99a}; {-100; -125b; -195: -130a; -189}; -197; {let-7d; let-7e; -126*; -140; -126; -101}; -182; -200a; -200b; -141	tissue development; epidermis development; rRNA processing; RNA metabolic process
j	N_colon, N_uterus	T_colon,	mir-199a; -199a*; -17-3p; -335; -181b; -181c	methylation; single fertilization
k		T_bladder, T_breast	mir-7; -106b; -142-3p; -142-5p	hemidesmosome assembly; cholesterol transport; cholesterol metabolic process; cell junction assembly
l	N_prostate , N_breast		{mir-1; -133a; -99a};{ -100; -125b; -195; -130a; -189}; -197; -335; {let-7d; let-7e; -126*; -140; -126; -101}	fibrinolysis; complement activation; collagen fibril organization; complement activation, classical pathway

[a] 'underexpression' spots of the combined map shown in ref. (3)

[b] systems with downregulated spots

[c] miRNA extracted from the miRNA/mRNA metafeature-combinations taken from the respective spot. The brackets collect miRNAs referring to one miRNA metagene (see text).

[d] enriched gene sets taken of the gene ontology category 'biological process' in the list of mRNA extracted from the miRNA/mRNA metafeature-combinations taken from the respective spot

5. Summary and Conclusions

Case studies illustrating the potency of the portraying approach using Self Organizing Maps were presented. Sample portraits of embryonic stem cells, induced pluripotent stem cells and differentiated fibroblasts clearly reveals groups of miRNA specifically overexpressed without the need of additional pairwise comparisons between the different systems. More than one dozen miRNA are found to be differentially expressed between ESC and IPS reflecting the inequivalence of both kinds of stem cells with respect to miRNA activity.

The much more heterogeneous series of human tissues splits roughly into five groups (brain, muscle, epithelial, gastrointestinal, sexual organs) according to their miRNA expression landscapes. They are characterized by about one dozen distinguishable expression modules. This diversity of miRNA expression is comparable with the diversity of mRNA expression patterns in human tissues despite the much less number of miRNA species available. Mixed samples with miRNA signatures of different tissues can be clearly identified. For example pericardium combines signatures of muscle and adipose tissues. Interestingly, immune systems tissues are clearly separated from other tissue types according to their mRNA expression signature in contrast to the miRNA expression patterns.

The joint analysis of miRNA and mRNA in healthy and tumor samples allows identifying potential OncoMirs and Tumor supressors Mirs, their potential mRNA targets and functional annotations according to positive and negative correlations between both entities and enrichment of sets of genes of known function.

In summary, these analyses demonstrate that the individual portraying of the expression landscape of each sample is highly sophisticated because it provides unmistakable fingerprints of the underlying expression phenotypes. It enables a very intuitive, image-based perception which clearly promotes the discovery of qualitative relationships between the samples in the absence of existing hypotheses. We see perspectives for broad applications of this method in standard analysis of single miRNA and especially combined miRNA/mRNA data sets.

Acknowledgements:
This publication is supported by LIFE Center for Civilization Diseases, Universität Leipzig. LIFE is funded by the European ERDF fund and by the Free State of Saxony.

6. References

1. Hobert, O. (2004), Common logic of transcription factor and microRNA action, *Trends in Biochemical Sciences* **29,** 462-68.

2. Baskerville, S., and Bartel, D. P. (2005), Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes, *RNA* **11,** 241-47.
3. Wirth, H., Cakir, V., Hopp, L., and Binder, H. (2013), Analysis of miRNA expression using machine learning, *Methods of Molecular Biology* **accompanying chapter in this volume**.
4. Liu, T., et al. (2007), Detection of a MicroRNA Signal in an <italic>In Vivo</italic> Expression Set of mRNAs, *PLoS ONE* **2,** e804.
5. Liang, Z., Zhou, H., Zheng, H., and Wu, J. (2011), Expression levels of microRNAs are not associated with their regulatory activities, *Biology Direct* **6,** 43.
6. Alexiou, P., Maragkakis, M., and Hatzigeorgiou, A. G. (2010), Online resources for microRNA analysis, *Journal of Nucleic Acids Investigation* **2,** e4.
7. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007), The role of site accessibility in microRNA target recognition, *Nat Genet* **39,** 1278-84.
8. Krek, A., et al. (2005), Combinatorial microRNA target predictions, *Nat Genet* **37,** 495-500.
9. Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P. (2009), Most mammalian mRNAs are conserved targets of microRNAs, *Genome Research* **19,** 92-105.
10. Papadopoulos, G. L., Reczko, M., Simossis, V. A., Sethupathy, P., and Hatzigeorgiou, A. G. (2009), The database of experimentally supported targets: a functional update of TarBase, *Nucleic Acids Research* **37,** D155-D58.
11. Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009), miRecords: an integrated resource for microRNA–target interactions, *Nucleic Acids Research* **37,** D105-D10.
12. Naeem, H., Kuffner, R., Csaba, G., and Zimmer, R. (2010), miRSel: Automated extraction of associations between microRNAs and genes from the biomedical literature, *BMC Bioinformatics* **11,** 135.
13. Cheng, C., and Li, L. M. (2008), Inferring MicroRNA Activities by Combining Gene Expression with MicroRNA Target Prediction, *PLoS ONE* **3,** e1989.
14. Kohonen, T. (1982), Self-organized formation of topologically correct feature maps, *Biological Cybernetics* **43,** 59-69.
15. Törönen, P., Kolehmainen, M., Wong, G., and Castrén, E. (1999), Analysis of gene expression data using self-organizing maps, *FEBS Letters* **451,** 142-46.
16. Eichler, G. S., Huang, S., and Ingber, D. E. (2003), Gene Expression Dynamics Inspector (GEDI): for integrative analysis of expression profiles, *Bioinformatics* **19,** 2321-22.
17. Wirth, H., Loeffler, M., von Bergen, M., and Binder, H. (2011), Expression cartography of human tissues using self organizing maps, *BMC Bioinformatics* **12,** 306.
18. Wilson, K. D., Venkatasubrahmanyam, S., Jia, F., Sun, N., Butte, A. J., and Wu, J. C. (2009), MicroRNA Profiling of Human-Induced Pluripotent Stem Cells, *Stem Cells and Development* **18,** 749-57.
19. Liang, Y., Ridzon, D., Wong, L., and Chen, C. (2007), Characterization of microRNA expression profiles in normal human tissues, *BMC Genomics* **8,** 166.
20. Lu, J., et al. (2005), MicroRNA expression profiles classify human cancers, *Nature* **435,** 834-38.
21. Mallanna, S. K., and Rizzino, A. (2010), Emerging roles of microRNAs in the control of embryonic stem cells and the generation of induced pluripotent stem cells, *Developmental Biology* **344,** 16-25.
22. Binder, H., Fasold, M., Hopp, L., Cakir, V., von Bergen, M., and Wirth, H. (2011), Genomic and molecular phenotypic portraits – exploring the 'OMes' with individual resolution, *HIBIT2011 proceedings*.
23. Israel, A., Sharan, R., Ruppin, E., and Galun, E. (2009), Increased MicroRNA Activity in Human Cancers, *PLOS one* **4,** e6045.
24. Visone, R., and Croce, C. M. (2009), MiRNAs and Cancer, *The American Journal of Pathology* **174,** 1131-38.
25. Hamano, R., Ishii, H., Miyata, H., Doki, Y., and Mori, M. (2010), Role of microRNAs in solid tumors, *Journal of Nucleic Acids Investigation* **2,** e2.
26. Osada, H., and Takahashi, T. (2007), MicroRNAs in biological processes and carcinogenesis, *Carcinogenesis* **28,** 2-12.
27. Raia, R., and Calin, G. A. (2010), Non-coding RNAs and cancer:m icroRNAs and beyond, *Journal of Nucleic Acids Investigation* **2,** e5.

28. Chi, S. W., Zang, J. B., Mele, A., and Darnell, R. B. (2009), Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps, *Nature* **460,** 479-86.