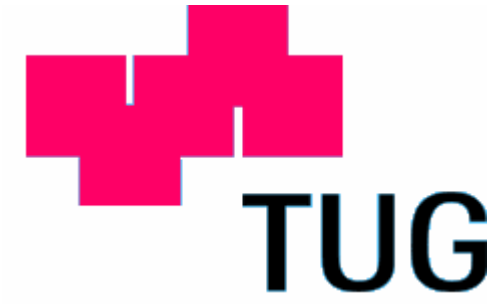
Thomas Burkard

Gene expression analyses of 3T3-L_1 cell lines during differentiation

Master thesis



Institute of Biomedical Engineering
University of Technology, Graz

Supervisor: Univ.-Prof. Dipl.-Ing. Dr. techn. Zlatko Trajanoski
Head of Institute: Univ.-Prof. Dipl.-Ing. Dr. techn. Gert Pfurtscheller

Graz, May 2003

FOR MY FIANCÉE BEATRIX BUDER

ABSTRACT

In this study, the differentiation from fibroblast like cells to mature adipocytes was investigated using genome-scale cDNA-microarray. Three independent differentiation experiments were examined at eight time points. Confluent 3T3-L1 cells were induced to differentiate with insulin, isobutylmetyhlxanthine and dexamethasone. Dye swapping was used to hybridize 51 adipocyte specific chips with 27648 expressed sequence tags (ESTs). The data was normalized and analyzed with clustering algorithms. Additional relevant information about the arrayed ESTs was retrieved from public databases and the complementarity between the regulatory microRNAs and the mRNAs was investigated.

It was shown that 800 genes have at least 2-fold over- or under-expression in more than three time points. The clusters of co-expressed genes show many expected transcriptional profiles of fat metabolism associated enzymes, transcription factors and signal transduction factors. It was proposed that regulatory mechanisms between the transforming growth factor beta3 (TGF beta3) and decorin as well as between the myc proto-oncogene protein (c-myc) and the CCAAT/Enhancer binding protein alpha (C/EBP alpha) play a role in adipogenesis.

Keywords: adipogenesis, transcriptional profiling, adipose specific cDNA-microarray, 3T3-L1 cell line

KURZFASSUNG

In dieser Studie wurde die Differenzierung von Fibroblasten-ähnlichen Zellen zu reifen Fettzellen mit der cDNA-Microarray Technik untersucht. Drei unabhängige Differenzierungsexperimente wurden an acht Zeitpunkten untersucht. Hierzu wurden konfluente 3T3-L1 Zellen mit Insulin, Isobutylmethylxanthin und Dexamethason induziert. Mit Hilfe der „Dye-Swapping"-Technik wurden 51 fettzellenspezifische Chips, die 27648 „Expressed Sequence Tags" (ESTs) enthielten, hybridisiert. Die Daten wurden normalisiert und mit Clustering-Algorithmen analysiert. Weiters wurden zu den ESTs relevante Informationen aus öffentlichen Datenbanken abgerufen und wurde die Komplementarität zwischen microRNAs and mRNAs untersuchen.

Es wurde gezeigt, dass 800 Gene in mehr als drei Zeitpunkten mindestens 2-fach über- oder unterexprimiert wurden. Die Cluster von coexprimierten Genen zeigten viele erwartete Transkriptionsprofile von Fettmetabolismus assoziierten Enzymen, Transkriptionsfaktoren und Signaltransductionsfaktoren. Es wurde weiters vorgeschlagen, dass Regulierungsmechanismen zwischen dem Transforming Growth Factor beta3 (TGF beta3) und Decorin sowie zwischen dem Myc proto-oncogene Protein (c-myc) und dem CCAAT/Enhancer Binding Protein alpha (C/EBP alpha) in der Adipogenese eine Rolle spielen.

Schlüsselwörter: Adipogenese, Transkriptionsprofil, Fettzellen-spezifische cDNA-Microarray, 3T3-L1 Zelllinie

ACKNOWLEDGMENT

I would like to thank my supervisor Professor Zlatko Trajanoski for his guidance of this work. He made it possible to make this study in the interdisciplinary field.

Special thanks are due to my tutor Hubert Hackl, for his guidance, open ears and patience. It was a pleasure to work together with him and learn from the intensive discussions. Many thanks also to my colleagues of the bioinformatics group at the Institute of Biomedical Engineering in Graz. They had always time for my various queries and guided me through my cell culture beginnings.

I am especially grateful to my family and my beloved companion Beatrix, who have supported me during my whole work.

CONTENTS

1 Introduction 1
1.1 Obesity 1
1.2 Adipose tissue, 3T3-L_1 cells and Adipogenesis 3
1.3 Transcriptional Profiling with cDNA-Microarray 8
2 Objectives 11
3 Methods 12
3.1 Cell Culture 12
3.1.1 Chemical formation 12
3.1.2 Inoculation 14
3.1.3 Maintenance 14
3.1.4 Freezing 15
3.1.5 Differentiation Experiment 15
3.2 RNA Isolation 17
3.3 RNA quality control 18
3.3.1 Photometer 18
3.3.2 BioAnalyzer 18
3.4 Microarray assay 19
3.4.1 Aminoallyl Labeling and coupling aa-cDNA to Cy dye Esters 20
3.4.2 Prehybridization and Hybridization 21
3.4.3 Washing the slide 22
3.5 Computational Analysis 22
3.5.1 Scanning 22
3.5.2 Image Analysis and data filtering 23
3.5.3 Normalization and data retrieval 24
3.5.4 Microarray clustering - Genesis 25
4 Perl Progams 26
5 Results 31
5.1 Cell culture 31
5.2 RNA concentration and integrity 32
5.3 Microarrays 35
5.4 Clustering 38

5.5 Data retrieval 41
6 Discussion 42
Bibliography 46
Appendix A – Experimental Flowchart 50
Appendix B – Microarray assay 51

ABBREVIATIONS

18S	18 Svedberg ribosomal RNA
28S	28 Svedberg ribosomal RNA
3T3-L_1	preadipocytes (fibroblast-like mouse cell line)
A/C/G/T/U	Adenosine/Cytidine/Guanosine/Thymidine/Uridine
BAT	Brown Adipose Tissue
bHLH	basic-Helix-Loop-Helix
BMI	Body Mass Index
C/EBP	CCAAT/Enhancer Binding Proteins
cDNA	complementary DNA
c-myc	Myc proto-oncogene protein
CO_2	carbon dioxide
ddH_2O	MilliQ water
DMEM	Dulbecco's Modified Eagle Medium
DMEM+	Culture medium
DMSO	DiMethylSulfOxid
DNA	DesoxyriboNucleic Acid
ECM	ExtraCellular Matrix
EDTA	EthyleneDiamineTetraAcetic acid
EST	Expressed Sequence Tag
FGF-10	Fibroblast Growth Factor 10
IBMX	IsoButylMethylXanthine
miRNA	micro RNA
mRNA	messenger RNA
N_2	Nitrogen
NaN	Not a Number
PBS	Phophate Buffered Saline
PCR	Polymerase Chain Reaction
PMT	PhotoMultiplier Tube
PPAR	Peroxisome Proliferator-Activated Receptor
RNA	RiboNucleic Acid
RXR	Retinoic X Receptor

SREBP-1c	Sterol Regulatory Element-Binding Protein 1c
TGFß	Transforming Growth Factor beta
TIFF	Tagged Image File Format
TIGR	The Institute for Genome Research
UCP-1	UnCoupling Protein 1
WAT	White Adipose Tissue
WHO	World Health Organization

1 INTRODUCTION

1.1 OBESITY

Obesity is defined medically as a state of increased body weight, more specifically adipose tissue, of sufficient magnitude to produce adverse health consequences. Body weight and composition, and the storage of energy as triglyceride in adipose tissue, are determined by the interaction between genetic, metabolic, environmental, and psychosocial factors. These influences ultimately act by changing the energy balance equation, that is, the long-term balance between energy intake and expenditure [1]. This energy balance equation is the thermodynamic background, which says that the whole energy which goes into a system (individual) must leave it again or will get stored in high energy molecules.

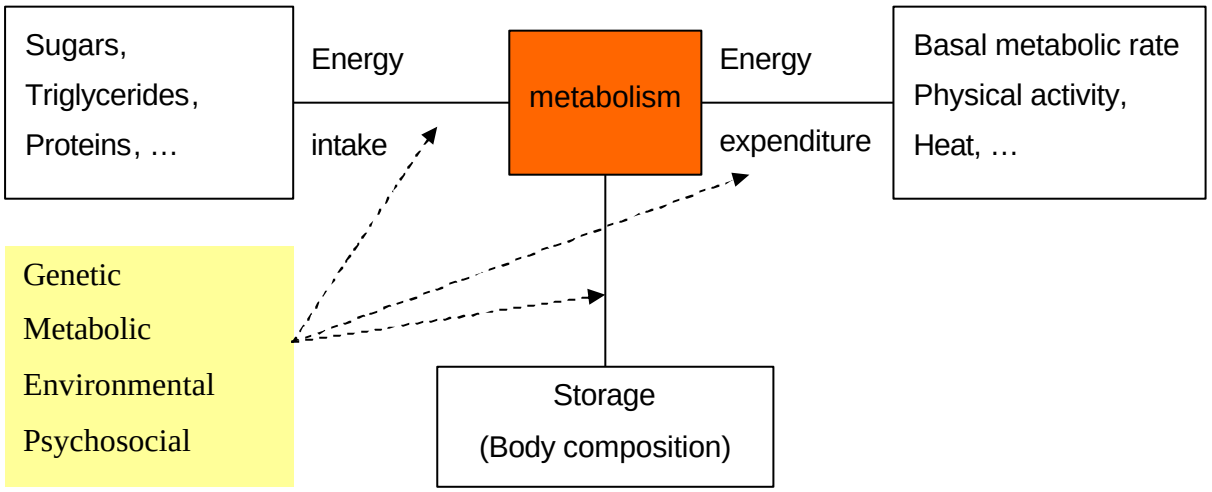


Figure 1 Energy Balance with storage of extra energy. Genetic, metabolic, environmental, and psychosocial factors influence the energy balance which leads to a storage of fat in adipose tissue.

Any disorder in the energy balance can lead to overweight or obesity. To diagnose obesity the National Institute of Health issues a clinical guideline. The body mass index (BMI) serves to estimate the degree of overweight:

$$BMI = \frac{bodymass(kg)}{[height(m)]^2}$$

For adult persons the body mass index should be between 18.5 kg/m^2 and 24.9 kg/m^2. The experts proposed for overweight individuals a BMI between 25.0 kg/m^2 and 29.9 kg/m^2. We speak of obesity, if the BMI is above 30 kg/m^2. The BMI overestimates the adipose portion for very muscular persons whereas it underestimates the fat tissue for elderly persons, who have lost much muscle mass. For children and adolescents between 2 and 20 years a gender- and age-specific BMI-table exists [2].

The World Health Organization (WHO) reported 2002 the most serious world health risks, which are defined as "a probability of an adverse outcome or a factor that raises this probability" but don't include infectious diseases. The ten leading risk factors globally, in terms of the burden of disease they cause, are: underweight; unsafe sex; high blood pressure; tobacco consumption; alcohol consumption; unsafe water, sanitation and hygiene; iron deficiency; indoor smoke from solid fuels; high cholesterol and obesity. In developed countries and low-mortality developing countries, overweight is the fifth most serious risk factor. A very interesting view is the comparison between undernutrition and overnutrition burdens. On the one hand there are 170 million children in poor countries who are underweight – and over three million of them die each year as a result – on the other hand there are more than one billion adults worldwide who are overweight and at least 300 million who are clinically obese. Among these, about half a million people die from obesity-related diseases every year in the USA and Canada (220000 deaths) and in 20 countries of the Western Europe (320000 deaths) [3]. In the USA there are 55% of adults overweight and nearly a quarter are obese [4].

Table 1 Associated diseases with obesity (adapted from [7])

Obesity is associated with increased risk of:	
• Premature death • Type 2 diabetes • Heart disease • Stroke • Hypertension • Gallbladder disease • Osteoarthritis • Sleep apnea • Asthma • Breathing problems • Cancer (endometrial, colon, kidney, gallbladder and breast cancer)	• High blood cholesterol • Complication of pregnancy • Menstrual irregularities • Hirsutism (presence of excess body and facial hair) • Stress incontinence (urine leakage caused by weak pelvic-floor muscles) • Increased surgical risk • Psychological disorders such as depression • Psychological difficulties due to social stigmatization

The impacts of obesity are of social and economic relevance. The costs for health care are exploding. In the United States of America the total economic cost of obesity in 2000 was estimated with $117 billion. This sum can be divided in $61 billion direct health care costs and $56 billion indirect costs [5]. The direct health care costs refer to preventive, diagnostic and treatment services. The indirect costs are results of wage lost by people unable to work because of morbidity [6].

Overweight and obesity is associated with higher risks of morbidity and premature death. There are more than 30 medical conditions associated with obesity [8]. Some of these conditions are listed in Table 1. The body mass index is a direct measure for the risk to gain an associated disease. This correlation is shown in Figure 2.

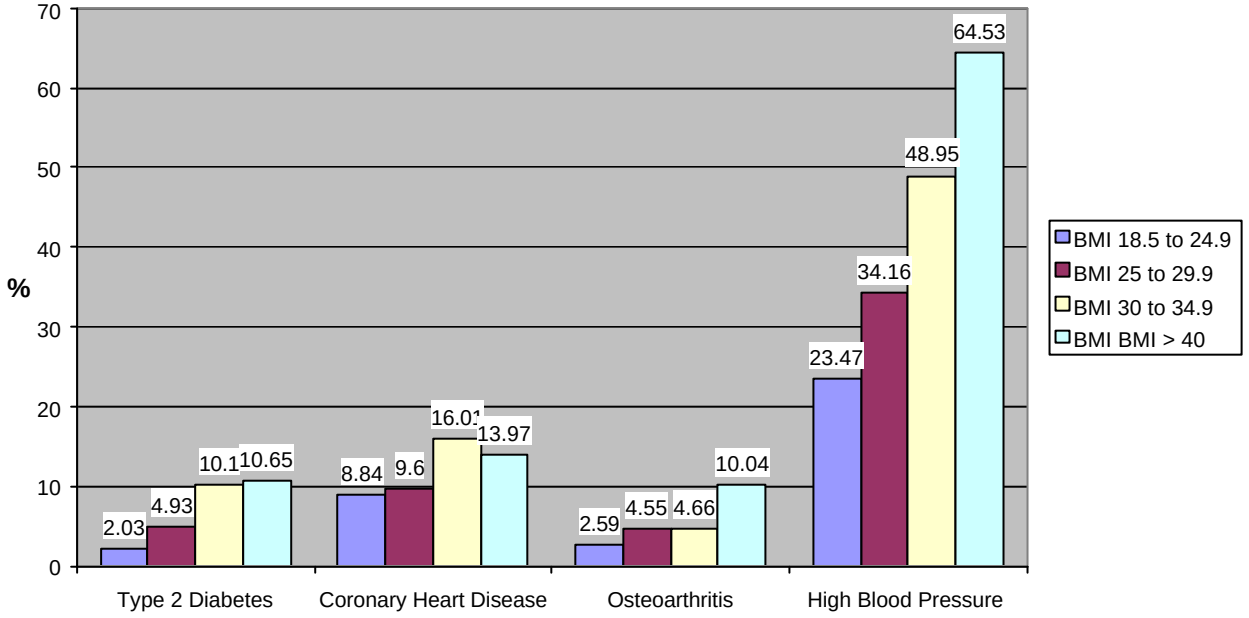


Figure 2 The body mass index can be correlated with the amount of morbid men (data from [8])

1.2 ADIPOSE TISSUE, 3T3-L_1 CELLS AND ADIPOGENESIS

In all mammals white and brown adipocytes are organized in a multi-depot organ called the adipose organ [9]. The mouse adipose organ is mainly composed of anterior and posterior subcutaneous depots which are located at the roots of the limbs and is composed of several visceral depots which are located in tight connection with mediastinic and abdominal organs. Histologically it is composed of white adipose tissue and brown adipose tissue (Figure 3) which are joined together

without any anatomical boundary between them. The relative amounts of the tissues depend on strain, age, sex, environmental and nutritional conditions [10].

The main function of white adipose tissue is the storage and mobilization of extra energy. Sex-differences in white adipose tissue location are interesting and probably have a functional significance during the evolution. Women have about 50% more adipose tissue than men. The hip-tight depot supplies the fetus during the latter part of gestation and also the new born child with energy. Therefore this typical female localization of fat tissue has also a specific female function especially during times of varying food supply. Men accumulate more of their depot fat in visceral regions. The fat is drained to the vena porta and can easily mobilize. Prehistoric man, surviving through hunting and quick escape, had great need for easily mobilized energy that quickly reached the liver and entered the circulation as lipoproteins. In our days, fat depots of a normal person contain energy for one month of total starvation, which was maybe in earlier days an evolutionary and genetic selection advantage [11].

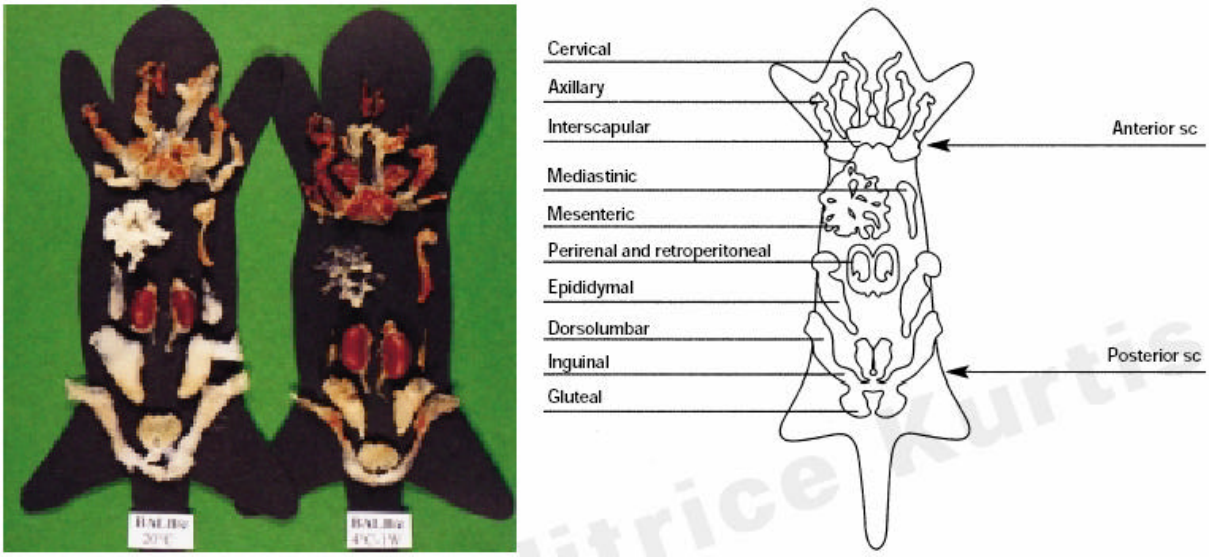


Figure 3 Adipose organ of mouse at 20°C and -4°C and the description of its anatomic locations. WAT (white) and BAT (brown) (adapted from [10])

The brown adipose tissue serves primarily to dissipate energy instead of storing it [12]. Through uncoupling the oxidative phosphorylation by the uncoupling protein-1 (UCP-1) the proton gradient across the inner mitochondrial membrane dissipates [13]. So heat is generated through fat burning. Due to this fact brown adipose tissue plays probably a role in protection against obesity [14].

Since the discovery of leptin in the mid-1990s, attention has been focused on the role of proteins secreted by adipose tissue. Research in the past decade has demonstrated that many peptide and non-peptide substances are secreted from adipose tissue. Most of these factors act in an autocrine/paracrine manner to regulate adipocytes metabolism. After secreting and releasing into the bloodstream, they act as endocrine signals at multiple distant sites to regulate energy homeostasis [15], [16]. The big group of secreted substances can be classified as metabolic modulators, vascoactive factors, growth factors, complement system, binding proteins, hormones, cytokines and many others. The importance of adipose tissue in vertebrate and mammalian evolution is now apparent, and its importance in mammalian physiology is beginning to be appreciated [17].

A well characterized model of adipocyte differentiation is the cell line 3T3-L_1. This fibroblast-like cell line is derived from non-adipose tissue and can convert efficiently into lipid-accumulating cells in culture [18], [19]. The 3T3 (Swiss albino) is a cell line derived from whole mouse embryos. L_1 is a continuous substrain of 3T3 (Swiss albino) developed through clonal isolation [20] . These cell lines do not express any of the structural and functional markers of adipocytes in the growing phase, but show instead properties consistent with their non-transformed fibroblast nature. This means for instance that they synthesize collagen of type I and III [21] and possess a well-organized fibrillar network of fibronectin and well-developed microfilaments. The differentiation process occurs spontaneously after growth arrest in the confluent culture. The cells develop the enzyme activities of fatty acid synthesis, triglyceride synthesis and the utilization of exogenous lipid. The phenotype changes to rounded cell shape characteristic of adipocytes with multilocular lipid droplets and a central positioned cell nucleus.

Chronic treatment with hyperphysiological concentrations of insulin accelerates the rapidity and efficiency of the adipose conversion in 3T3-L_1 cell lines [22], [23]. Two mechanisms are suggested for the need of this high concentration: the involvement of cell surface receptors for type I insulin-like growth factor have lower affinity for insulin or a high degradation rate of insulin in the culture medium [24]. Newly synthesized fatty acids are the major part of accumulated triglycerides [22]. Therefore the accumulation of lipids is inhibited if the medium contains no biotin, even in the presence of exogenous lipids and elevated cellular LPL activity [25]. The phosphodiesterase inhibitor isobutylmetyhlxanthin stimulates the conversion to

adipocytes, which suggests a role for cyclic nucleotides in the control of differentiation [26]. The glucocorticoid dexamethasone also induces the adipocyte differentiation efficiently.

The adipose differentiation itself is a very complex process. The acquisition of the phenotype is characterized by chronological changes in the expression profile. The process can be divided in phases of growth arrest, post-confluent mitoses, clonal expansion and terminal differentiation. It is also reflected by the expression of early, intermediate and late mRNA/protein markers and lipid accumulation [27]. The regulation and the changes occur primarily at the transcriptional level, although posttranscriptional regulation is known for some genes [28], [29]. An overview of the complex processes is given in Figure 4.

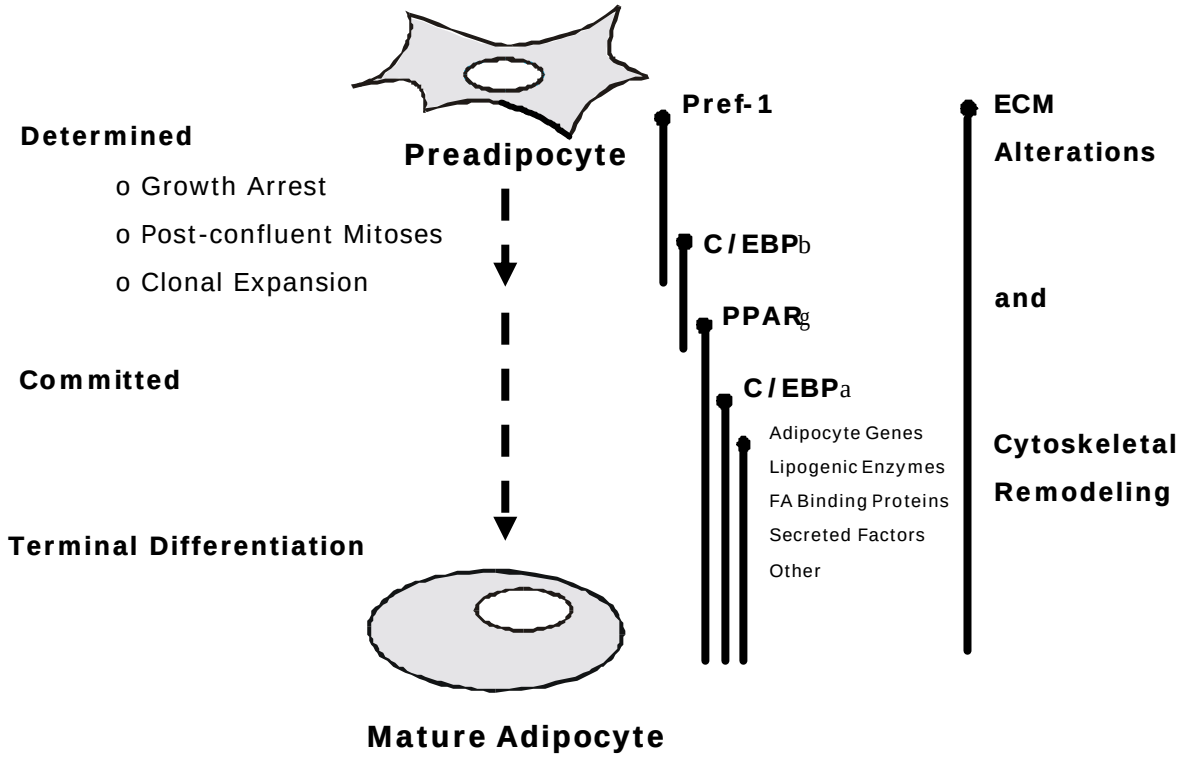


Figure 4 Chronological changes of the biological processes and transcriptional cascade during adipocyte differentiation (adapted from [27])

To initiate adipose differentiation growth arrest is required not cell confluence or cell-cell contact per se [30], [31]. The transcription factors CCAAT/enhancer binding protein a (C/EBP-a) and peroxisome proliferators-activated receptor-? are playing an important role in later phases of adipogenesis. At low concentration these two transcription factors also seem to be involved cooperatively in growth arrest [32],

[33]. Before late mRNA markers appear the cells must receive an appropriate combination of mitogenic and adipogenic signals. These lead to at least one round of DNA replication and cell doubling [30]. Transient changes in retinoblastoma proteins (pRB, p107, p130) modulate the E2F/DP complex to inactivate growth-promoting transcriptional activities, which is a main difference between clonal expansion and preconfluent cell growth [27].

A summary of the transcriptional pathways is given in Figure 5. The most important transcription factor classes are Cys4 zincfinger, leucine zippers and bHLH. Peroxisome proliferators-activated receptors (PPARs) belong to type II nuclear hormone receptor family and form heterodimers with RXR [34], [35]. The PPARs are activated by a variety of dissimilar compounds [27]. The heterodimer PPAR-RXR regulates transcription by binding a response element consisting of a direct repeat of the nuclear receptor hexameric DNA recognition motif (PuGGTCA) spaced by one nucleotide [36]. CCAAT/enhancer binding proteins have a basic transcriptional activation domain and an adjoining leucine zipper motif, which provides the ability for homo- and heterodimerization [27]. Fine-tuning of the control of adipocyte gene expression involves specific patterns in dimerization between the isoforms [37]. bHLH are transcription factors with a basic-helix-loop-helix domain. One member is the SREBP-1c/ADD1 which is expressed abundantly in adipose tissue [38] and binds DNA at an E-box motif (CANNTG) as well as a SRE-1 motif [39].

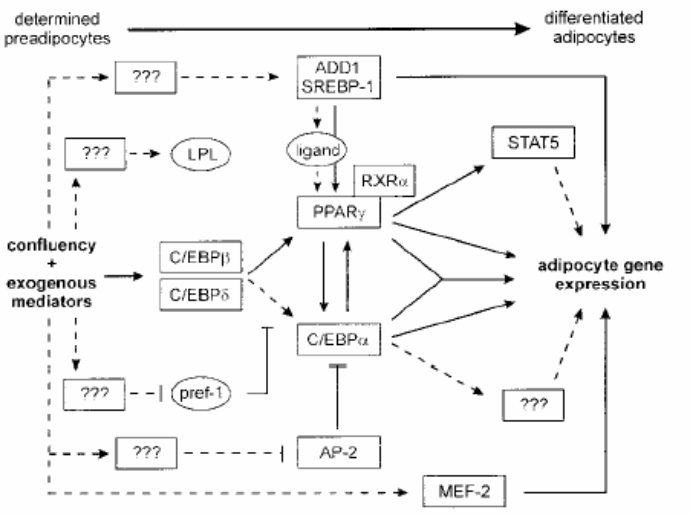


Figure 5 Model illustrating linear and parallel transcriptional pathways of adipogensis. Solid lines: direct or indirect transcriptional events; broken lines: interactions that are less well understood; Squares boxes: transcription factors; question marks: unknown factors; (adapted from [40])

1.3 TRANSCRIPTIONAL PROFILING WITH CDNA-MICROARRAY

In the era after the genome sequencing it becomes important to understand the context of the AGTC-sequence and to characterize the gene products and regulating networks on the level of transcription and posttranscription. A relative robust, easy and fast technique to get a snapshot of the transcriptome, or at least a great part of the transcriptome, at a specific condition and a specific time point is the production of microarrays. The most used systems today are complementary DNA (cDNA) and oligonucleotide microarrays. To analyze the transcriptome of cells or tissues the messenger RNA or total RNA is extracted. The RNA is reversely transcribed to cDNA, which is labeled with a fluorescence dye. The labeled probe is mixed with a different labeled reference and together the mixture is co-hybridized to the arrayed genes on the slide. With a laser scanner the relative fluorescence intensities of the sample and reference are measured. In Figure 6 an overview of the main steps in cDNA-microarray is given.

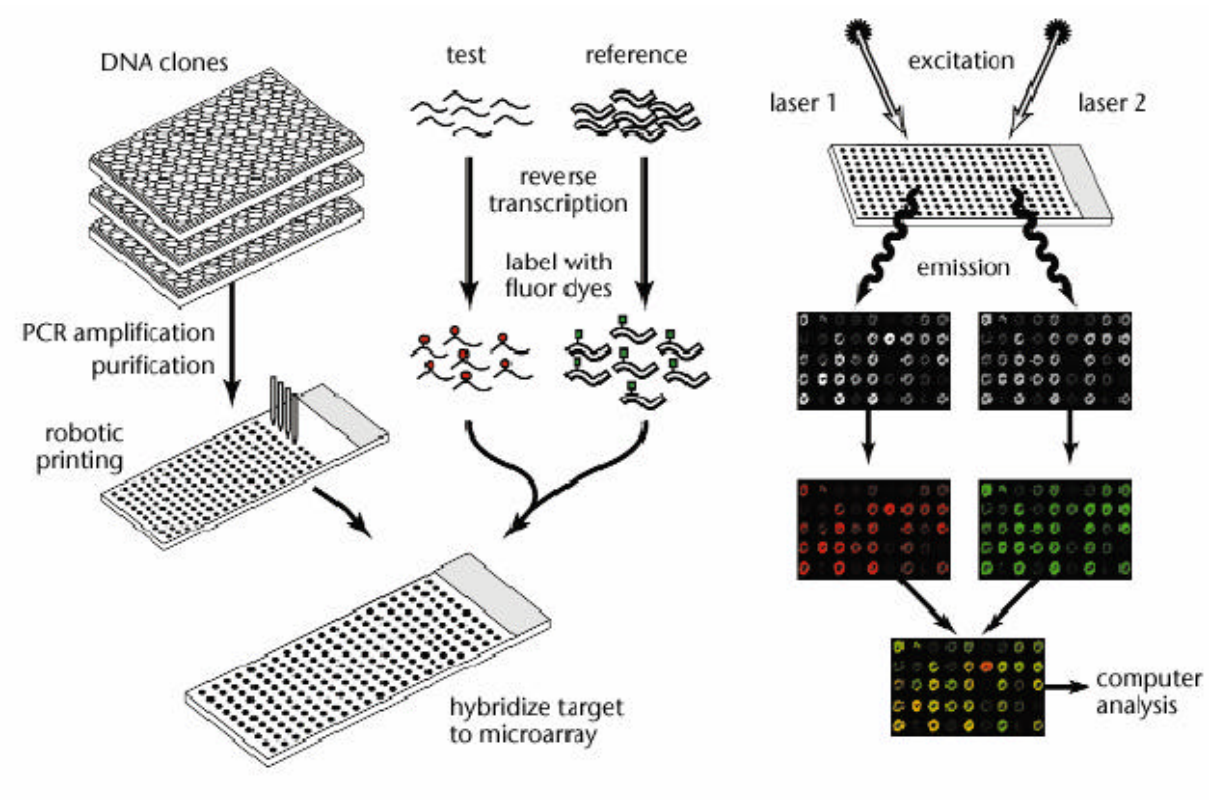


Figure 6 cDNA microarray procedure (adapted from [41])

The arrayed materials in the cDNA-microarrays are products of a polymerase chain reaction (PCR) generated from cDNA libraries and clone collections. A robot spots the probe at a defined location on specially coated glass slides with high throughput. Spots are typically 100-300µm in size and are spaced by the same distance apart. On the surface of a conventional microscope slide more than 30000 cDNAs can be fitted. Spotted arrays are more flexible in the choice of arrayed elements than oligonucleotid chips. An additional advantage is the possibility to spot unidentified and unsequenced clones form libraries, which can be useful for gene discovery [42].

Oligonucleotide arrays can be synthesized *in situ* or spotted. Therefore short 20-25mers are arrayed. The advantages are the possibility of detection of closely related genes or splice variants. The less specific hybridization and reduced sensitivity can be circumvented through longer oligonucleotides (50-100mers) [43]. However, the oligo-slides are very expensive.

Richter et al [44] compared different fluorescent tag DNA labeling methods. The aminoallyl labeling method was classified as combining excellent accuracy with very good sensitivity and medium work intensity. Thereby aminoallyl modified nucleotides are incorporated during the reverse transcription. In a following step the cDNA is coupled to Cy3 or Cy5 fluorescent dyes. For very small amounts of input material T7 RNA polymerase amplification showed to be the best method.

Due to fluorescence bias and other interfering effects, the microarray data must be normalized. The method of choice is the technique of dye swapping. Thereby two identical microarrays are hybridized. Only one difference occurs: One slide is hybridized with a Cy3 labeled reference and a Cy5 labeled sample and the other vice versa. If the arrays are treated equally, the intensity ratio for each spot is corrected by the geometric mean [56]. Beside this technique which doubles the laboratory work and costs, there are many other normalization methods, which have a solid statistical background.

To interpret the huge amount of data in microarray experiments computational clustering methods are needed. These algorithms identify pattern by looking either for similarities or for differences. The methods can be divided in unsupervised and supervised approaches. Supervised approaches need additional biological information about specific genes that are functionally related to guide the clustering algorithm [45], [46].

Two methods are used in this master thesis and therefore will be examined here a little bit closer. Hierarchical clustering is an iterative method and has the advantage that it is simple and the result can be easily visualized. At beginning a pairwise distance matrix is calculated for all of the genes. In average-linkage hierarchical clustering the average distance is calculated between each point in a cluster and all other points in another cluster. The two closest clusters are subsequently merged. The whole process repeats by recalcutating the distances between the new cluster and all other clusters [45], [46].

K-mean clustering usually uses the Euclidean properties of the vector space. After choosing the number of clusters (*K*) the vector space is divided into *K* parts. After calculating the center points of each subspace the partitions are adjusted by assigning the vectors to the closest centre. This is repeated iteratively until either the partitioning stabilizes or the given number of iterations is exceeded [45], [46].

In most microarray studies different groups of patients or a time series of a biological process are observed. Clustering the patients can give information over the mRNA profile of a disease and can show differences between ill and healthy individuals. The analysis of a biological process shows the changes of the transcriptome in a time interval. From this information transcriptional networks can be built. One very important assumption for this analysis is that co-regulated genes are also co-expressed and therefore they have the same profile over time.

2 OBJECTIVES

The long term objective is to delineate transcriptional networks of adipogenesis for a better understanding of the causes of obesity. For a large scale gene expression study the powerful technology of cDNA microarray will be used to get a quick insight of a very interesting part of the genome wide transcriptome under specific conditions. Therefore a fat-specific microarray is available, which contains many adipose specific ESTs. Analyzing a couple of time points in differentiation from fibroblast like 3T3-L_1 cells to mature adipocytes will lead to an expression profile of the biological processes. The assumption that co-expressed genes are also co-regulated or at least share some regulatory mechanism is obvious. Therefore intensive computational analyzes will identify clusters of genes which are co-expressed and maybe co-regulated. In the vast amount of resulting data it should be possible to identify good known genes of adipogenesis as well as new genes involved in the differentiation. These new genes could be new targets for further investigations in antagonizing obesity. Also the mechanism of peroxisome proliferators-activated receptors should be explored intensively as well as downstream regulated genes from PPARs.

Specific goals of this master thesis are listed below:

- Designing a large scale microarray experiment
- Performing the differentiation experiment of the 3T3-L_1 cell line to mature adipocytes with high throughput
- RNA isolation and hybridization experiment
- Data analysis

3 METHODS

3.1 CELL CULTURE

One cryo-tube with 3T3-L_1 cells (The Global Bioresource Center™, ATCC# CL-173) in the 3rd passage (-196°C) was obtained from the Institute for Molecular Biology, Biochemistry and Microbiology, Graz (IMBM). For practice with cell cultures and testing of the new laboratory, the cells were brought to the 4th passage and frozen on fluid N_2 over several days. Out of the 4th passage a stock of frozen cells have been built till the 11th passage. For differentiation the 6th, 7th and 8th passage were used.
For all these cell culture experiments the cell number was not determined, because the state of confluence was an obvious characteristic and all actions happened nearly before or at confluence.

3.1.1 CHEMICAL FORMATION

All chemicals were mixed under laminar flow.

Culture Medium (DMEM+): The substances of Table 2 were mixed together. This medium could be stored up to 2 weeks at 4°C. Especially L-Glutamine was instable and broke down to ammonia. Serum was added for additional macromolecular growth-factors and the addition of antibiotics substituted the immune system [47].

Table 2: Culture Medium (DMEM+) [48]

Substance ..... Substances needed for medium
c ..... Concentration of substance in final medium
V_S ..... Volume of substance to add

Substances	c	V_S
Dulbecco's modified Eagle Medium with 0.11G/L NA Pyr with Pyridoxine (Gibco™ Invitrogen Corporation, #41966-029)	-	500 mL
Foetal Bovine Serum heat inactivated (Gibco™ Invitrogen Corporation, # 10500064)	10%	56.8 mL
Penicillium-Streptomycin 10000 units/mL (Gibco™ Invitrogn Corporation, # 15140122)	1%	5.68 mL
L-Glutamine 200mM (100x) (Gibco™ Invitrogen Corporation, # 25030024)	1x	5.68 mL
		Σ 568.16 mL

1x Trypsin: 10x Trypsin-EDTA (Gibco™ Invitrogen Corporation, # 15400054) was diluted with 1x Phosphate Buffered Saline, pH 7.2 (Gibco™ Invitrogen Corporation, # 20012019).

10mM 3-Isobutyl-1-Methylxanthine (IBMX) stock: To 21.65 mg IBMX (Sigma®, # I7018) 2mL ddH_2O and 100 µL 1N sodium hydroxide solution (Roth, # K021.1) was added. It was mixed well until it became a clear liquid. Additional 7.6 mL ddH_2O were added. The whole solution was sterile filtrated through a 22µm filter and partitioned into aliquots. The solution was stored at -20°C.

50µM Dexamethasone stock: 1.88 mg Cyclodextrin encapsulated Dexamethasone water soluble (Sigma®, # D2915, 69mg Dexamethasone/g) were solved in 6.6 mL ddH_2O. The whole solution was sterile filtrated through a 22µm filter, partitioned into aliquots and stored at -20°C.

1mg/mL Insulin stock: 3.34 mg Insulin from Bovine Pancreas (Sigma®, # I6634) were solved in 0.0125N Hydrochloric acid (pH2.0), sterile filtrated, partitioned and stored at -20°C.

1.7mM DL-Pantothenic acid stock: 4.31 mg DL-Pantothenic Acid Hemicalcium Salt (Sigma®, # P6292) were solved in 10.75 mL ddH_2O, sterile filtrated, partitioned and stored at -20°C.

3.3mM D-Biotin stock: 8.38mg D-Biotin (Sigma®, # B4639) were solved entirely in 1mL N,N-Dimethylformamide (Sigma®, #68-12-2) and filled up with 9.5mL Dulbecco's modified Eagle Medium. After sterile filtrating it was partitioned and stored at -20°C.

Induction medium: The substances of Table 3 were mixed together and the solution was sterile filtrated with a Stericup™ – Vacuum driven disposable filtration system (Millipore Corporation).

Table 3: Induction medium [50]

Substance ..... Substances and concentration in final induction medium
Stock solution ..... Concentration of the stock solution (Chapter 3.1.1)
Volume to add ..... Volume to add for 270mL induction medium

Substance	Stock solution	Volume to add
Culture Medium	-	270 mL
Panthotenacid (0.17µM)	1.7 mM	27 µL
Biotin (3.3µM)	3.3 mM	270 µL
Insulin (1µg/mL)	1 mg/mL	270 µL
IBMX (100µM)	10 mM	2700 µL
Dexamethasone (0,25µM)	50 µM	1350 µL

3.1.2 INOCULATION

Ten 75cm^2 tissue culture flasks (BD Falcon®, # 351005) were prepared with 12mL culture medium (37°C). Than one cryo-tube was taken out of the fluid N_2 and very quickly defrosted in the water bath. Before the freezing medium was totally defrosted, the cells were quickly resuspended. An aliquot (1:10) of the suspension was added to each prepared culture flask. It was crucial to dilute the cells quickly with culture medium because the high concentration of dimethylsulfoxid (DMSO) in the freezing medium was toxic for the cells.

3.1.3 MAINTENANCE

The cells were incubated at 37°C in an atmosphere of 100% humidity and 5% CO_2. The culture medium was exchanged every Monday, Wednesday and Friday. The medium contained enough nutrients as long as the indicator was not changing to yellow. On the same days a part of the culture flasks were observed under the

microscope for infections and density of the cells. It was important, that the cells got not confluent, because otherwise the cells were induced for differentiation. Therefore the cells were frozen or passaged shortly before they were confluent (approximately after 5 days). To passage the 80-90% confluent cells, one cell culture flask was split to ten new culture flasks.

3.1.4 FREEZING

To create a stock of frozen cells, the 80-90% confluent cells of one culture flask were frozen in 1 mL freezing medium, which consisted of 90% culture medium and 10% DMSO (Aldrich®, # 47,230-1). The cells were rinsed with 1x Phosphate Buffered Saline (pH 7.2, PBS) and detached with 1x Trypsin. The reaction of Trypsin was stopped with culture medium and the cell suspension was centrifuged with 1500rpm for 3 minutes. The pellet was resuspended with chilled freezing medium and transferred to a NUNC cryo-tube (NUNC™ , # 375353). To freeze the cells slowly they were frozen in a Cryo 1C Freezing Container (NALGENE™, # 5100-0001) at -80°C. This box was filled with Isopropanol and guarantied the freezing rate of 1°C/1minute. After one night the tube was stored on fluid N_2 at -196°C.

3.1.5 DIFFERENTIATION EXPERIMENT

In the following section the differentiation procedure from 3T3-L_1 cells to adipocytes is shortly described. One cryo-tube with 3T3-L_1 cells in the 4th passage was taken out of the fluid N_2 and the cells were inoculated (see Chapter 3.1.2). After 4 days maintaining (see Chapter 3.1.3), the cells of one culture flask were split to ten new ones and maintained for additional 4 days. The 80-90% confluent cells of six cell culture flasks were split to sixty-six 100mm Tissue Culture Dishes (IWAKI brand, #3020-100). It was important that the Petri dishes were treated for tissue cultures, otherwise the cells easily detached from the dish surface, as happened in one pilot survey. The cells could be reattached by supplementing the culture medium with 1% Bovine serum albumin (Sigma®, # A9418). Because the detachment could have an adverse effect on the transcriptome, the reattached cells were only used for verifying, if the cell line differentiates to adipocytes.

The 3T3-L_1 cells (6^{th} passage) of the 66 culture dishes were maintained for four days. To achieve the reference RNA, the subconfluent cells (80-90%) of 36 Petri dishes were washed twice with Phosphate buffered Saline and were lysed with 2mL TRIzol® Reagent (Invitrogen™ life technologies, # 15596-026) per dish. TRIzol® Reagent is based on phenol and guanidine isothiocyanate to ensure the integrity of the RNA and to inactivate RNases. The cell homogenate was stored at -70°C.

3 days later the cells of the remaining 30 dishes became confluent. To guarantee the growth arrest of all cells through contact inhibition, they were maintained for additional 2 days. For the time point 0 hours the cells of three dishes were lysed with 2mL TRIzol® Reagent per dish and stored at -70°C. The remaining cells were induced with 10mL induction medium (see Chapter 3.1.1). The induced cells were incubated at 37°C in an atmosphere of 100% humidity and 5% CO_2. At the time points 6h, 12h, 24h, 48h and 72h after induction the differentiating cells of three tissue culture dishes were dissolved in 2mL TRIzol® Reagent per dish and stored at -70°C. Because of the worse homogenization of fat rich cells in organic reagents, 3mL TRIzol® Reagent was used at the time points 7d and 14d after induction.

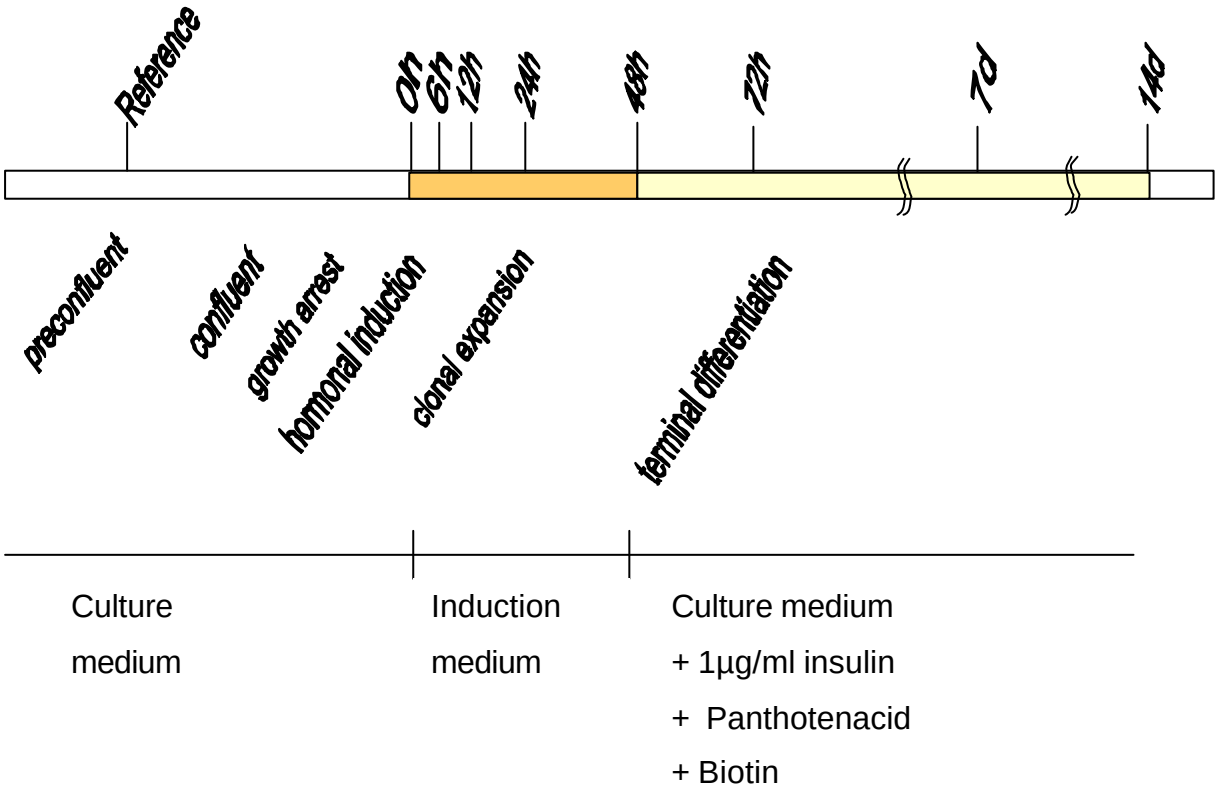


Figure 7 Time schema of differentiation experiment

Two days past the induction and every following second day the medium was exchanged with culture medium supplemented with 0.17µM Panthotenacid, 3.3µM Biotin and 1µg/mL Insulin. After induction the cells began to differentiate to round adipocytes which detached from the dish surface very easily, therefore the cells had to be handled with additional care.

To evidence the differentiation, the fat droplets in the cells were stained with Oil Red O (ICN Biomeducal Inc., # 1320-06-5) and the cytoplasm was counter stained with Light Green SF Yellowish (ICN Biomeducal Inc., # 5141-20-8) [49] at some time points. Microscope images were taken with the Olympus CKX41 microscope (Olympus) equipped with the digital camera Olympus CAMEDIA C-4040Zoom (Olympus).

Parallel to the differentiation experiment subconfluent cells were maintained. Out of these subconfluent cells the second and third differentiation experiments were carried out at the 7^{th} and 8^{th} passage. The splitting to 66 Petri dishes of the second experiment was accomplished when the reference RNA of the first run was taken. The same time difference was between the second and third experiment. The differentiation procedure was the same as in the first experiment.

3.2 RNA ISOLATION

To separate the RNA from the rest of the cell components, RNA isolation had been carried out. After defrosting and homogenizing with the homogenizer RW16 basic (IKA Labortechnik), the extraction of total RNA was accomplished as in the TRIzol® Reagent manual [51] described. The main steps will be described briefly below.

Around 36mL of reference or 6 time points (each ~6mL) were processed at once. The limiting factor was the capacity of the centrifuge for twenty-four 2mL vials. To each vial with around 1.7mL homogenate, 350µL Chloroform was added. The vials were vigorously shaken for 15 seconds, followed by a 10 minutes rest. Subsequently the separation was accelerated by centrifuging with 16000g for 15 minutes at 18°C. The inter phase and the organic phase contained normally most of DNA, proteins, polysaccharides and lipids. The RNA remained exclusively in the upper aqueous phase which was transferred cautiously to a new vial. For further purification the RNA was precipitated with 1mL 2-Propanol, 99.5% HPLC grade (Aldrich, Cat#

27,049-0). The precipitate was washed with 70% ethanol (MERCK, Cat# K21482643307) and a second time with absolute ethanol. The RNA was air-dryed for 30 minutes and subsequently solved in DEPC treated H_20 (Research Genetics®, # 800-533-4363).

3.3 RNA QUALITY CONTROL

3.3.1 PHOTOMETER

The measurement of the total RNA concentration and a first check of the quality and purity were performed on a spectrophotometer. 3µL of each probe were diluted with 297µL ddH_2O (1:100) and the absorption was immediately measured on the UV/VIS Spectrophotometer DU® 520 (Beckman) with the program Warburg-Christian. The concentration of the samples was adjusted with DEPC H_2O to approximately 1.5-2.0µg/µL. To verify the new concentration, a 1:100 dilution of the resulting solution was measured twice on the photometer. The adjusted total RNA solution (1.5-2.0µg/µL) was stored at -70°C.
Also the ratio of absorption at 260nm to the absorption at 280nm was obtained for a first estimation of quality and purity.

3.3.2 BIOANALYZER

For an accurate quality and purity control of the RNA, it was analyzed with the BioAnalyzer 2100 (Agilent Technologies, Paris, France). RNAs, DNAs, proteins and cells can be analyzed with different BioAnalyzer chips. These chips contain micro-channels filled with gel in which the molecules are separated by electrophoreses. For total RNA a concentration of the samples between 25ng/µL and 500ng/µL was needed.
First the electrodes were washes with 350µL RNase AWAY® (Molecular BioProducts, Cat# 7003) followed by washing with 350µL DEPC water for 10 seconds. A RNA 6000 Nano Chip (Agilent, # 5065-4476) was filled with a gel-dye mix on the Chip Priming Station. Into the ladder and sample wells 5µL of RNA 6000 Nano Marker was

pipetted. The chip was loaded with the RNA 6000 ladder and 12 RNA probes. After vortexing for 1 minute the run was started immediately.

3.4 MICROARRAY ASSAY

For hybridization 51 adipose specific chips were used. These chips were designed and spotted in cooperation with the Institute of Biomedical Engineering (Graz, Austria) and The Institute for Genomic Research (TIGR, Rockville, USA). These chips were 27648 elements mouse cDNA microarrays which contained four different categories of arrayed ESTs:

- 627 Adipose specific elements: These genes are important for adipocyte biology and the fat metabolism (Transcription factors, co-activators and –repressors...). Expressed sequence tags (ESTs) and genes shown differentially expressed in 3T3-L1 preadipocytes and by leptin in ob/ob mice.
- 15k NIA mouse clone set: Genes expressed in early embryonic development stages (National Institute of Aging).
- 11k BMAP clones: Subtracted cDNA libraries from 10 brain regions of adult mouse brain, spinal cord, and retina (mouse strain c57BL6J). (Brain Molecular Anatomy Project NIH & University of Iowa)
- Control spots: Genomic DNA, *Arabidopsis thaliana* genes, DMSO, Cot1 DNA, Salmon Sperm DNA

For reasons of normalization and minimization of errors, each probe was labeled twice by the dye swap technique. Therefore each probe was labeled once with Cy3 and once with Cy5 and hybridized with the corresponding labeled reference (Figure 8). This has led to 48 hybridized microarrays.
Additional the three references were hybridized among each other, to verify the consistency of their expression profiles.
In the following chapters the procedures of labeling and hybridization will shortly be discussed. It was accomplished as in the TIGR Standard Operating Procedure [53], [54] described with a few little changes. A detailed protocol and a closer description of the used chemicals can be found in Appendix B.

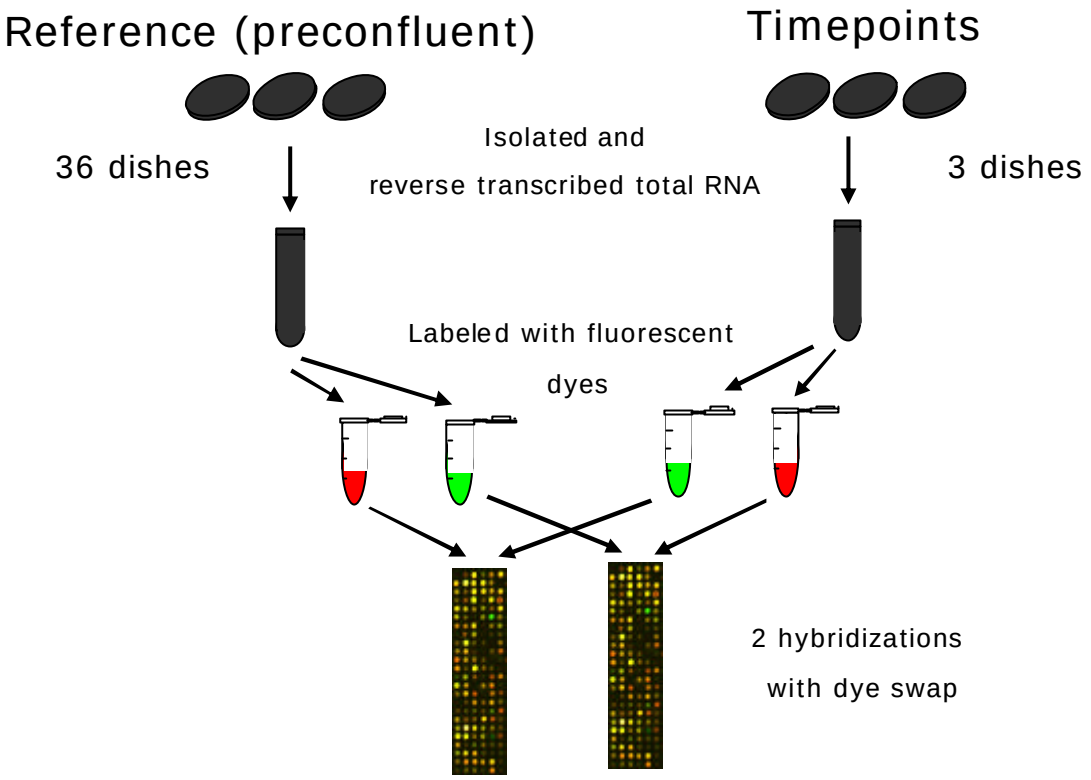


Figure 8 Microarray hybridizations with the dye swapping technique

3.4.1 AMINOALLYL LABELING AND COUPLING AA-CDNA TO CY DYE ESTERS

There were two main steps in this procedure. First aminoallyl-dUTP was incorporated into the reverse transcribed cDNA, second the incorporated aminoallyl-dUTP was coupled to a specific Cy dye ester (Figure 9).

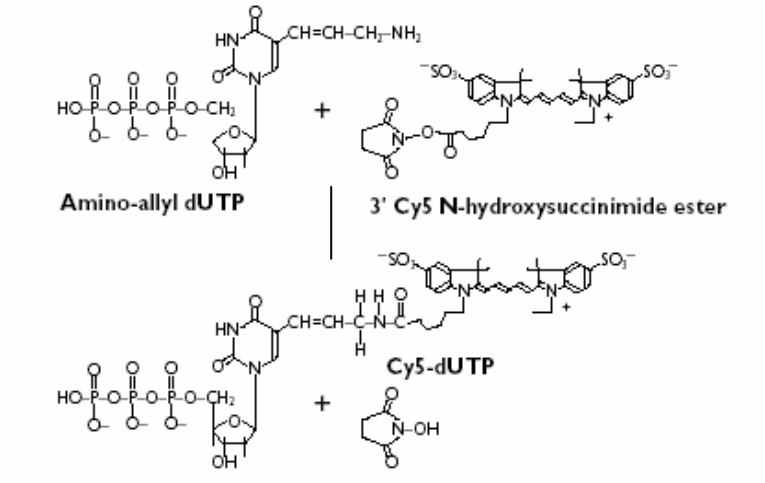


Figure 9 Coupling reaction of the NHS-Cy5 ester to aminoallyl-dUTP (adapted from [55])

24µg total RNA was used for the reverse transcription. After one night the cDNA was completely synthesized. Due to the fact that the remaining RNA could interfere the hybridization, it was hydrolyzed with 1M NaOH. The aminoallyl-labeled cDNA was cleaned up with a QIAquick PCR Purification Kit (QIAGEN, # 28104). Instead of the supplied washing and elution buffer, which contained amines and therefore would interfere with the coupling reaction, phosphate buffers were used for washing (5mM KPO_4, pH 8.0, 80% EtOH) and eluting (4mM KPO_4, pH 8.5). The purified cDNA was dried on the speed Vac and solved in 0.1M sodium carbonate buffer (pH 9.1) for coupling with the appropriate NHS-ester-Cy-dye. After the dye-coupling in the dark, the fluorescence labeled cDNA was an additional time purified with the QIAquick PCR Purification Kit to remove the uncoupled dye and dried on the speed vac. This time all QIAgen supplied buffers were used.

3.4.2 PREHYBRIDIZATION AND HYBRIDIZATION

To prevent unspecific binding of the labeled cDNA on the chip, the microarray was prehybridized with preheated prehybridization buffer (42°C), which contained Bovine serum albumin. For further use the slides were washed five times in MilliQ water and once with isopropanol. After the last washing step they were dried immediately by centrifuging at 1500rcf for 2 minutes. The quality of the washed slides was checked by scanning a preview on the GenePix 4000B (Axon Instruments) scanner. If the quality was not satisfying, the washing procedure was repeated.

For hybridization the Cy3 probe was resuspended together with the Cy5 reference for one slide and for the dye swapped slide the Cy5 probe was resuspended together with the Cy3 reference. To block non-specific hybridization mouse COT1-DNA and Poly(A)-DNA was added to the suspension. To denature the cDNA, the mixture was held at 95°C for 3 minutes and snap-cooled on ice for additional 30 second. The hybridization was performed in hybridization chambers, which were filled with 20µL water against desiccation. The labeled cDNA suspension was put on the slide and spread under a microscope glass coverslip. The hybridization was performed at 42°C for 18-20 hours in the dark.

3.4.3 WASHING THE SLIDE

The very important washing step removed not and unspecific bound cDNA but it had not to solve the hybridized cDNA. Therefore it was crucial to follow strictly the steps of the protocol in the appendix, especially in the very stringent solutions. After the washing the microarray was immediately dried by centrifuging and as soon as possible scanned.

3.5 COMPUTATIONAL ANALYSIS

3.5.1 SCANNING

Scanning of the cDNA-Chips was performed on the GenePix 4000B (Axon Instruments) scanner with the GenePix Pro 4.1 (Axon Instruments) software. All 51 slides were scanned twice with different Photomultiplier gains:

- High Photomultiplier gain: The gain was adjusted to the number of saturated spots. There should be no more than approximately 7 saturated spots (white) in each 24x24 block. With this setting most spots in a wide range of fluorescence intensities could be detected.
- Low Photomultiplier gain: Some strong regulated genes were extremely over-expressed in the reference or the probe. This resulted in saturated spots. To detect these very interesting spots, the slide was scanned with a low photomultiplier gain, so that nearly no spot was saturated all over the slide. The outcome of this was that spots with low intensity were no more detectable but the very strong overexpressed genes became detectable.

The settings of the channels 635nm (corresponds to Cy5) and 532nm (corresponds to Cy3) were chosen, so that the overall impression of the microarray image was neither red nor green but yellow. For normalization it was crucial to scan the dye swapped slides with the same parameters.
Further hardware settings had been: Power 100% for both channels, Pixel size 10µm, Lines to average: 1, Focus position 0µm. The results were saved in three

TIFF-image (tag image file format), which contained the images of each channel and a combination of both channels.

3.5.2 IMAGE ANALYSIS AND DATA FILTERING

To extract the intensity of each spot and color and to cross-link the resulting intensity ratios with the clone descriptions, image analysis had to be accomplished. The microarray images were analyzed with the GenePix Pro 4.1 (Axon Instruments) software. A GAL-file, which defined the coordinates of the genes on the slide and contained the gene annotations, was loaded and the grid was laid over the image. By analyzing the program calculated many different values for each spot (e.g. Local background of both wavelength around the spot, Mean of intensities over the spot, Median of intensities, …), which were saved in a TAB-delimited table of a GPR-files.
For a reliable microarray dataset filters were run over the dataset to mark maybe good, neutral and bad spots.
Maybe good spots were marked with the value 100. These were spots which were difficult to handle in further data processing because the intensity was in one channel not distinguishable from the background but in the other channel the intensity was very high. These ratios could be very valuable, because of the strong regulation due to the fact of the big difference between the two channels. By the following criteria these spots were marked:

```
([% > B532+1SD] <= 55 And [F635 Mean - B635] > 1000) Or
([% > B635+1SD] <= 55 And [F532 Mean - B532] > 1000)
```

The first expression `[% > B532+1SD] <= 55` checked if 55% of the pixels within the spot had a lower fluorescence intensity in the 532nm channel than the background plus once the standard deviation. These were spots with very low intensities. The second expression `[F635 Mean - B635] > 1000` checked if the intensity in the other channel subtracted by the local background gave a clear signal. In the second line only the wavelength for the criteria were switched.

With the following criteria bad spots were marked with the value -100 in the GPR-file:

```
[Flags]          = [Bad]        Or
[Flags]          = [Absent]     Or
[Flags]          = [Not Found]  Or
[F635 % Sat.]    > 10           Or
[F532 % Sat.]    > 10           Or
[Sum of Medians] < 1000         Or
[Sum of Means]   < 1000         Or
([% > B635+1SD] < 55 And [F532 Mean - B532] <= 1000) Or
([% > B532+1SD] < 55 And [F635 Mean - B635] <= 1000) Or
(([F532 Mean]-[F532 Median])/[F532 Mean])> 0.2      Or
(([F635 Mean]-[F532 Median])/[F532 Mean])> 0.2
```

The first three lines examined, if the analyzing tool had already defined the spots as bad, absent or not found. The next two lines checked, if more than 10% of the pixels in one channel were saturated. In case of saturation they were marked as bad. The following four lines marked spots with low intensities in both channels as bad. The last two lines checked if the spot was homogeneous. If the mean and the median differed too much, the spots were no more homogenous, in the majority of cases.
All defined values for comparison were empirical values.
All remaining spots were not marked. The ratios of means of these neutral spots were used for further calculation.

3.5.3 NORMALIZATION AND DATA RETRIEVAL

In the microarray experiments very huge sets of data were produced. 2654208 ratios were measured in the three time series with eight time points each, dye swapped hybridization and scanning at low and high photomultiplier gain. For a better handling of this big set of data the programming language Perl 5.8.0 (ActiveState) was chosen. A description of all programmed and used Perl programs can be found in chapter 4.
To normalize the data the program Vcalc_1.pl was run with all three time series. For the calculation of the standard deviation and estimator factor the ratios had to be in the interval of 0.25 to 4. Subsequently the programs Average_1.pl and Relevant_genes.pl were run. The parameters for Relevant_genes.pl were chosen

that the expression profile had to have at least four time points with a absolute value of the log_2-ratio of at least one.

To retrieve relevant information for the arrayed ESTs, the program Retrieve_unigene_7.pl was run with all EST accession numbers of the slide. Therefore the EST accession numbers were divided in six batches of 5000 numbers each and the data for all six batches were retrieved parallel to reduce the time. 110 mouse miRNAs and 230 miRNAs from other species were retrieved from the genbank and from the paper [58] and checked for complementary with the retrieved mRNA sequences.

3.5.4 MICROARRAY CLUSTERING - GENESIS

To identify co-expressed genes, clustering analysis was performed with the software Genesis 1.2.0 [57]. Two different methods were applied. First Hierarchical Clustering with average linkage was accomplished for the genes. A big advantage of this method is the possibility to visualize the data by a hierarchical tree indicating the relationship between the genes. With the hierarchical tree co-expressed clusters of up-regulated genes were marked and set public. This allows to identify the previous clustered genes in clusters of other methods. Second a k-mean algorithm was carried out. The number of expected clusters (k-value) was specified with 10. The resulting clusters were further investigated, if the clustered genes of the hierarchical clustering method also cluster together with the k-mean algorithm.

4 PERL PROGAMS

The following programs were written in the programming language Perl 5.8.0 (ActiveState) to normalization the huge dataset and to retrieve information from public databases. Perl is a retronym for “Practical Extraction and Report Language” and is optimized for problems which are about 90% working with text and about 10% everything else [52].

Vcalc_1.pl:

This program normalizes the data and prints the $\log_2$-ratios of one time series out in a gene vs. timepoint table, which is TAB-delimited. Therefore the program needs a parameter file, which defines the locations of the GPR-files, the output TXT-file and the total number of scanned spots. Normalization is performed by dye swapping, which has many advantages against the global normalization. It normalizes the data, spot by spot and tries to eliminate the dye bias fom Cy3 and Cy5. The data is normalized by the geometric mean [56] between the ratios of two dye swapped slides:

$$\log_2\left(\frac{\overline{P_i}}{\overline{R_i}}\right)=\log_2\left(\sqrt{\frac{P_{Gi}}{R_{Ri}}\cdot\frac{P_{Ri}}{R_{Gi}}}\right)=\log_2\left(\sqrt{r_i\cdot s_i}\right)\qquad with\qquad r_i=\frac{P_{Gi}}{R_{Ri}}\ \ and\ \ s_i=\frac{P_{Ri}}{R_{Gi}}$$

Equation 1 $\log_2$ of the normalized intensity ratios

P_i ..... Intensity of the i-th arrayed probe
R_i ..... Intensity of the i-th arrayed reference
G ..... Green labeled
R ..... Red labeled
r_i ..... i-th ratio of the first dye swapped slide
s_i ..... i-th ratio of the second dye swapped

Beside the elimination of the dye bias, it averages although all other errors in the spot, which is an advantage of the replication. The $\log_2$ is calculated for a better visualization because if the $\log_2$ changes by one, the intensity ratio doubles.

For validating of outliers, where normalization should not be accomplished, the mean and standard deviation of the distribution of the ratios is calculated:

$$\left| \frac{r_i}{s_i} - c \right| \geq 2s \Rightarrow outlier$$

with $$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{r_i}{s_i} - c \right)^2$$

and $$c \approx \frac{1}{N} \sum_{i=1}^{N} \frac{r_i}{s_i}$$

Equation 2 Validation of outliers

N ..... Number of genes
s ..... Standard deviation
c ..... Estimated factor

The calculation of the estimated factor c and the standard deviation can cause numerical problems. Therefore the intensity ratios should be in a limited range for the calculation of these two variables. This range can also be specified in the parameter file.

Outliers, data of maybe good marked genes, data of bad marked genes and data where one value is missing in one slide (no calculation possible) are marked in the resulting table with 'NaN'. If it is not possible to calculate the geometric mean out of the slide scanned with high photomultiplier gain, the program tries to calculate it from the associated microarray which was scanned with a low photomultiplier gain.

Therefore it calculates for the slides with low photomultiplier gain an own standard deviation and estimated factor because the distribution of the ratios is mostly different from that with low photomultiplier gain. For a better visualization of this effect the distribution of the $\log_2$-ratios is shown, which is only a transformation of the data (Figure 10).

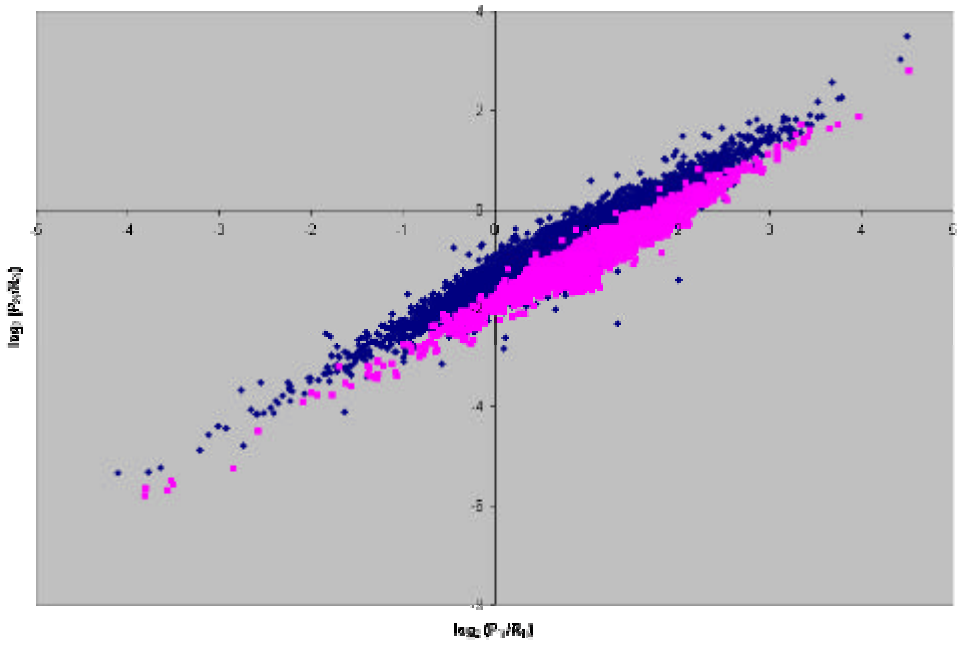


Figure 10 Distribution of the $\log_2$-ratios between the technical replicates of the dye swapped slides (first experiment, timepoint 0h). Here it is visualized that the distribution of slides scanned with high photomultiplier gain (blue) is not the same as with low photomultiplier gain (magenta).

Additional to the result file a LOG-file with relevant parameters, standard deviations, estimated factors and numbers of valid spots for each time point is saved.

Vcalc_1_good_marked.pl:

This program does the same as Vcalc_1.pl but classifies the spots which are marked with 'Maybe good' as valid ratios for calculation.

Average_1.pl:

This program calculates the arithmetic mean over the time series for each gene at a specific time point. The output TXT-file is formatted for the Genesis clustering software. Therefore it needs a parameter file with the locations of the result files from Vcalc_1.pl, the name of the output file and the total number of spots on one slide. It also creates a LOG-file with all important parameters and the numbers of valid ratios.

Relevant_gene.pl

This script produces a new set of data, which includes strong regulated genes. Therefore the user must define in how many time points the absolute value of the log_2-ratio must be higher than a specified value.

Retrieve_unigene_7.pl

The microarray GAL-file defines the position of each expressed sequence tag (EST) on the slide. As additional information the genbank accession numbers (gb) of the ESTs and the TIGR Gene indices annotations are saved. This is not enough information for an intensive research of gene networks and other microarray analyses. So it is very important to get as much information to each arrayed gene as possible. Therefore this Perl program retrieves important information and numbers for an easier accession of public databases.

The first step is to access an unigene-entry through the genbank accession number (gb) of the expressed sequence tags (ESTs). This is important because the information of the mRNA and of the geneproduct is needed for further analyses and not the information of the EST. The unigene cluster builds the link between the EST and the mRNA, because it assigns the ESTs through clustering to mRNA entries. At the Unigene webpage it is possible to retrieve the UniGene accession number as well as the number for the LocusLink page. The LocusLink contains the location of the gene in the genome, gene ontology terms and many other interesting information. The next task is to parse the accession number and the Geninfo identifier (gi) of the associated mRNA and the Protein entries. If a mRNA with a RefSeq number exits, the genbank datasheet for the RefSeq entry is accessed. If there is no RefSeq entry, the first mRNA entry in the unigene cluster leads to the genbank data. Furthermore the script checks if for the mRNA entries a RIKEN full-length clone exists. If so the clone ID is saved. Furthermore it retrieves the annotation of the mRNA, the number of basepairs, the position of the coding sequence and the mRNA sequence itself. It also checks if the sequence has a full complementary region to the regulatory elements of miRNAs. miRNAs are ~20bp long sequences, which can bind the mRNA and inhibits the translation or increases the degradation of the mRNA. For this task a set of miRNAs have to be pre-specified.

average_no_replicates.pl:

With this program it is possible to average genes within one set of data, if genes have the same TIGR annotation and if they are in the same unigene cluster. This process has the advantage that the clusters in the clustering analysis get not weighted by multiple entries of the same gene. Furthermore it reduces the number of missing values and many values get averaged again over spots with different ESTs. The result is that the hybridization of RNA to the arrayed ESTs is no longer depending on the homology to only one EST. The disadvantage is that the annotation and the belonging to unigene cluster depend on the reliability of the data in databases, which is often questionable.

Control_multiple_unigene_for_dataset:

This program generates a new dataset, which contains only genes with multiple equal UniGene annotations within the dataset. If this program is run after the script average_no_replicates.pl, it makes it possible to investigate the expression profile of genes with the same unigene annotation but different TIGR annotations. If the expression profile is equal, it can be assumed that the arrayed ESTs belong to the same gene.

5 RESULTS

5.1 CELL CULTURE

The maintenance of the 3T3-L_1 cell culture was performed bacterial infections or detachment of the cells. The cells grew fast, especially after adaptation. Therefore the cells had to be out of the cryo-preservation for at least two passages before differentiation was initiated. The used differentiation cocktail led to rapid differentiation of most preadipocytes. Microscope images of one differentiation are shown in Figure 11 as an representative example.

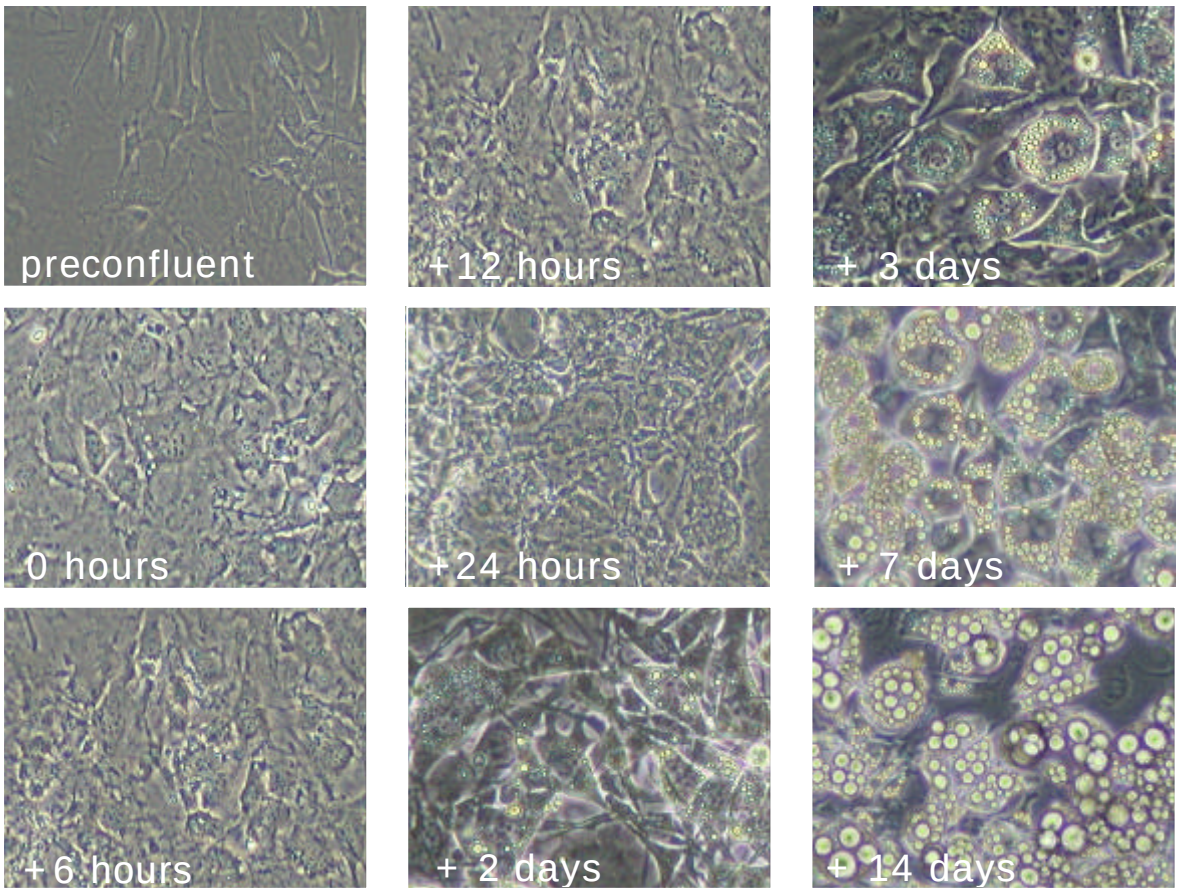


Figure 11 Microscope images of 3T3-L_1 cells during the differentiation to mature adipocytes representing all three differentiation experiments.

The preconfluent cells were fibroblast like shaped and the bottom of the culture dish was visible, which is an indication, that there was no contact-inhibition. Contact

inhibition was ensured by maintaining the confluent cells for additional two days. At induction (0 hours) the cells were fibroblast-like shaped and confluent. Changes in cell morphology were observed after 24 hours. The first visible fat inclusions were identified two days after induction. Fat droplets were verified by oil red O staining. At the same time the morphology was an intermediate between the fibroblast-like preadipocytes and the round adipocytes. These differentiating cells accumulated in small islands all over the culture dish, which is an indication for cell to cell. After 7 and 14 days the mature adipocytes were spread all over the culture dishes, with many huge fat inclusions, visualized by oil red O staining (Figure 12).

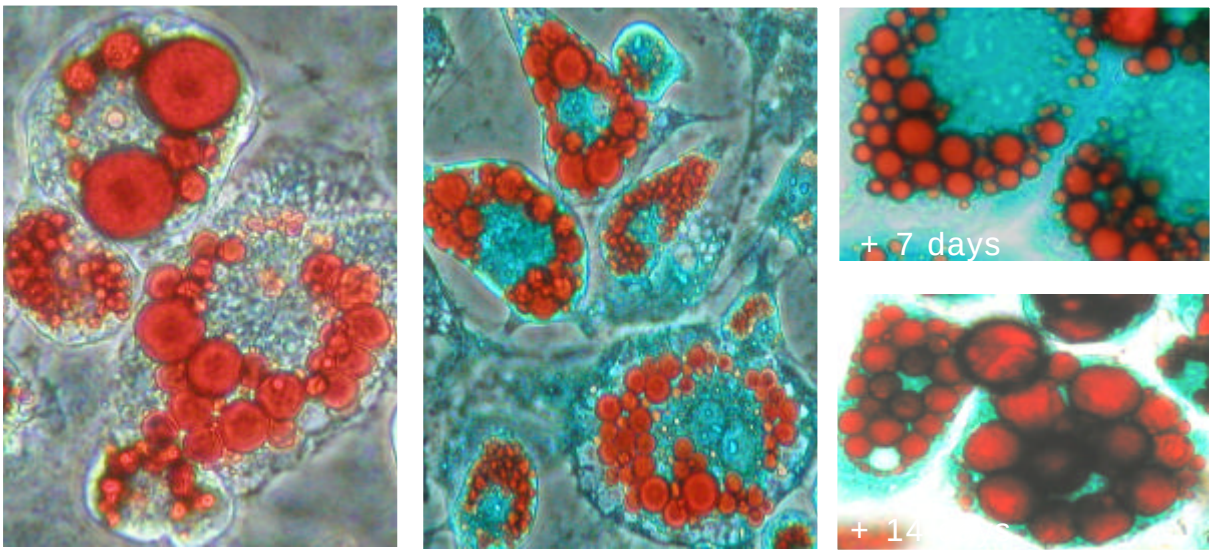


Figure 12 Oil red O stained fat droplets in mature adipocytes (7 and 14 days after induction). The cytoplasm was counterstained with Light Green SF Yellowish. This images represent the whole differentiation experiments.

5.2 RNA CONCENTRATION AND INTEGRITY

The concentration of the extracted RNA was measured with the spectrophotometer. The absorption was measured twice. The results of one time series are summarized in Table 4. Similar results were obtained from the other two experiments. The concentration was always in the range of 1.4µg/µL to 2.2µg/µL. Samples of the first time series were too much diluted. Therefore they were evaporated on the speed vac and diluted a second time. The A_{260}/A_{280}-ratio was around 1.5 (For pure RNA the ratio shall be between 1.6 and 2.0) (Table 4). Maybe the reason lies in the

considerable influence of the pH. Since the water is not buffered the pH can vary greatly [59]. For example lower pH results in lower ratios and reduced sensitivity to protein contamination [60]. For accurate ratios the probes must be measured in buffered solutions. Nevertheless the extinction coefficient for RNA is based on dilution in water. Therefore the probe was measured in water and the RNA quality and purity was qualified with the BioAnalyzer.

Table 4 Measurement of the time series V3 on the spectrophotometer representing all 27 RNA samples

Probe-ID..... Identification of Probe
V_{final}..... Final volume after dilution in µL
A_{260}/A_{280} ..... Ratio of the absorption at 260nm and 280nm
c_i..... Measured RNA concentration in µg/µL
c_m ..... Averaged RNA concentration in µg/µL

Probe-ID	V_{final}	A_{260}/A_{280}	C_1	c_2	c_m
V3 time 0h diluted	*66*	*1.580*	*1.73*	*1.85*	*1.79*
V3 time 6h diluted	*73*	*1.592*	*1.58*	*1.66*	*1.62*
V3 time 12h diluted	*75*	*1.706*	*2.10*	*2.08*	*2.09*
V3 Ref. Pool diluted	*434*	*1.621*	*1.97*	*2.01*	*1.99*
V3 time 24h diluted	*104*	*1.554*	*1.68*	*1.65*	*1.66*
V3 time 48h diluted	*132*	*1.512*	*1.76*	*1.71*	*1.73*
V3 time 72h diluted	*122*	*1.531*	*1.75*	*1.76*	*1.75*
V3 time 7d diluted	*106*	*1.598*	*1.98*	*2.04*	*2.01*
V3 time 14d diluted	*89*	*1.587*	*2.05*	*2.05*	*2.05*

Analysis with the BioAnalyzer 2100 from Agilent Technologies showed high quality and integrity of all RNA samples. For good quality and integrity the ratio of 28S ribosomal RNA to 18S ribosomal RNA shall be approximately two. This criterion is validated in all samples. In digested samples this ratio would shift to lower values, due to the higher amount of small fragments. A representative data is visualized in Figure 13 and Figure 14. Beside the quality of the RNA the concentration was estimated. As in any electropherogram the RNA concentration could not very accurate determined with this method. Especially with the small amounts of RNA used for the BioAnalyzer run the error is high. Therefore the values vary from the accurate values obtained by the spectrophotometer.

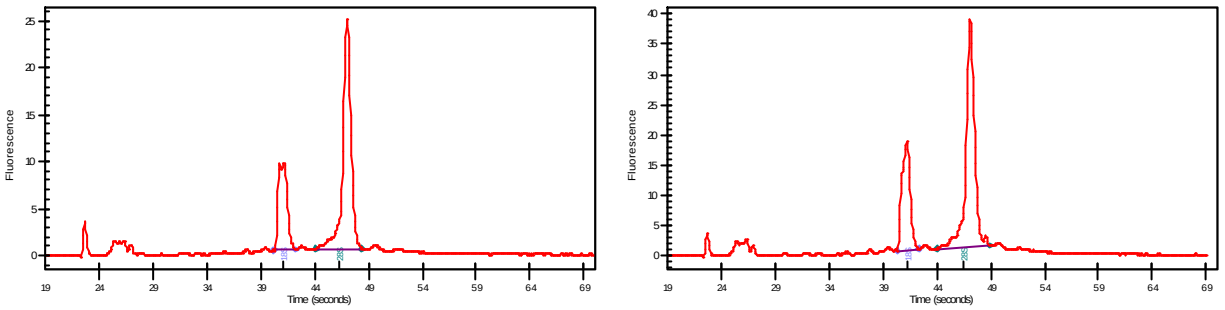


Figure 13 Quality and integrity control of the extracted RNA with the BioAnalyzer 2100 from Agilent Technologies which represents all 27 RNA samples. The graph (Time (seconds) vs. Fluorescence) shows from left to right following peaks: loading dye, 5S, 18S and 28S ribosomal RNA.

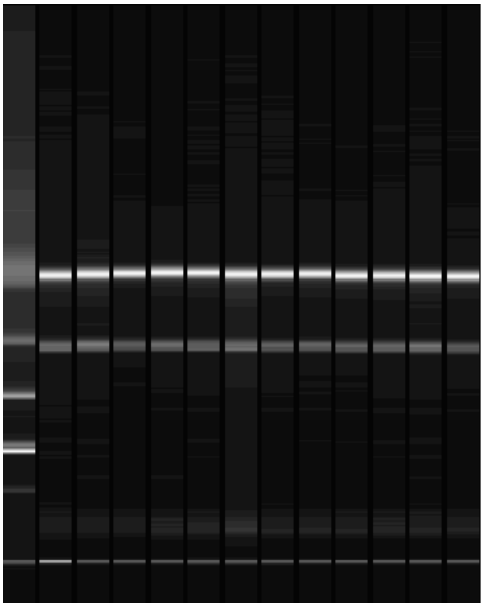


Figure 14 Artificial electropherogram of extracted RNA samples drawn by the BioAnalyzer 2100 software. The first lane corresponds to a defined RNA ladder. The remaining lanes show the samples of well 1 to 12 (see **Table 5**) This data represents all 27 RNA samples.

Table 5 BioAnalyzer data of extracted RNA corresponding to the electropherogram (Figure 14)

Probe ID..... Identification of sample
Well ..... Well on the BioAnalyzer chip (Figure 14)
18S ..... Area of 18S ribosomal RNA in percent of the total area
28S ..... Area of 28S ribosomal RNA in percent of the total area
28S/18S ..... Ratio of 28S ribosomal RNA to 18S ribosomal RNA
c(RNA)..... Concentration of total RNA in µg/µL

Probe ID	Well	18S	28S	28S/18S	c(RNA)
V2 time 7d diluted	1	21.48	43.20	2.01	2.26
V2 time 14d diluted	2	21.82	40.69	1.86	3.88
V3 Ref. Pool diluted	3	22.03	47.30	2.15	2.24
V3 time 0h diluted	4	21.26	42.83	2.01	3.21
V3 time 6h diluted	5	21.53	43.36	2.01	1.73
V3 time 12h diluted	6	14.01	28.25	2.02	2.43
V3 time 24h diluted	7	21.90	44.74	2.04	2.32
V3 time 48h diluted	8	21.14	42.23	2.00	2.55
V3 time 72h diluted	9	20.95	42.20	2.01	2.92
V3 time 7d diluted	10	20.94	42.03	2.01	2.70
V3 time 14d diluted	11	21.26	42.34	1.99	2.04
V3 Ref. Pool diluted	12	20.42	44.44	2.18	1.97

5.3 MICROARRAYS

The first quality control of the microarray was viewing the scanned image. The scanned slides had resulted in images with bright spots and low background (Figure 15a). The samples were labeled well. Only one sample of the last time series hadn't labeled at the time point 14 days. Therefore the associated two slides haven't been used for further analyses.

Few slides had in the upper right block problems with the arrayed spots (Figure 15b). This couldn't be an artifact of wrong handling during hybridization. Maybe this resulted from empty wells of expressed sequence tags during the spotting process. Few slides showed another adverse effect. Thereby columns of high intensity spots alternated with columns of low intensity spots (Figure 15). Due to the systematic pattern it was also a problem of spotting. Most likely the surface of the slide was uneven.

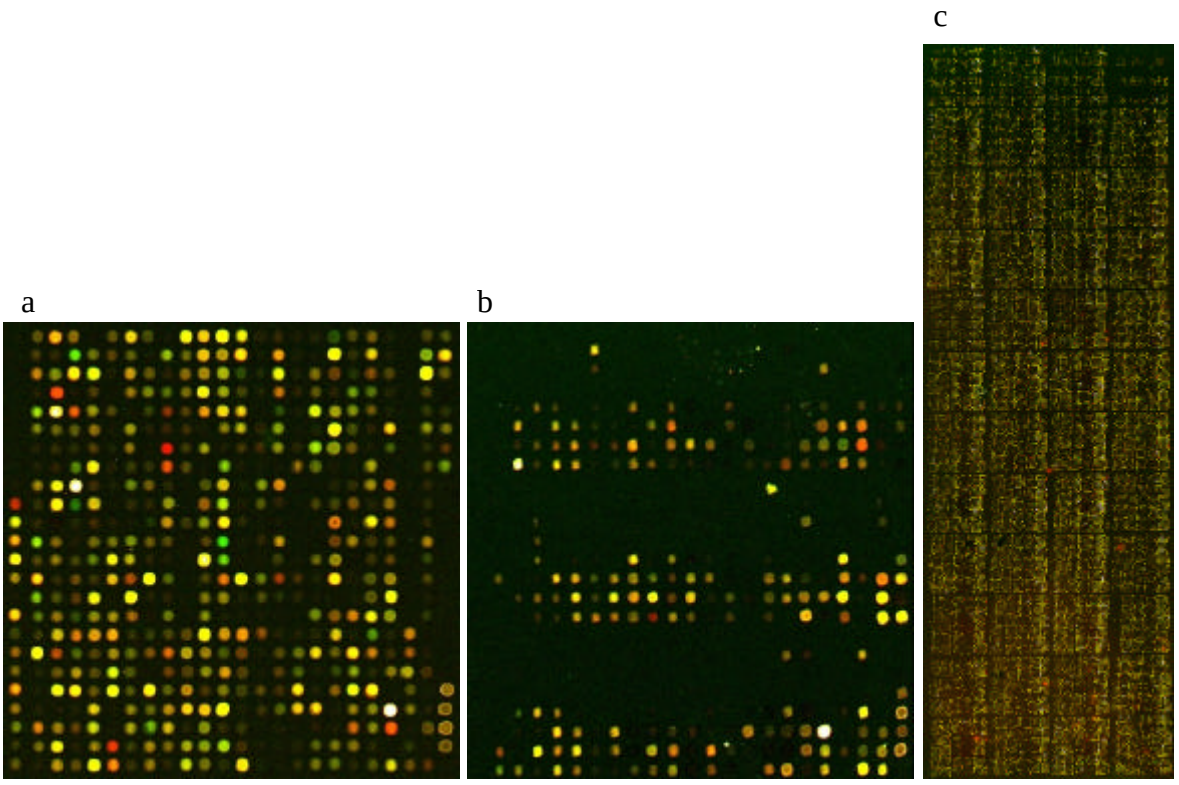


Figure 15 Quality of microarray images. (a) Most of the 51 hybridized slides are of good quality with bright spots and low background as shown in this block. (b) Few slides have spotting problems in the upper right corner. (c) Few slides have spotting problems, where columns of high intensity alternate with columns of low intensity.

MA-plots and histograms (data not shown) were used to control the microarray quality. All slides showed good quality.

The consistency of the references (preconfluent stage of cells before differentiation) was checked by hybridizing the reference RNAs of the independent experiments among each other. After scanning, the slides were normalized by a global normalization method. Figure 16 shows only minor deviations of the reference intensities. Due to this fact the references are very similar.

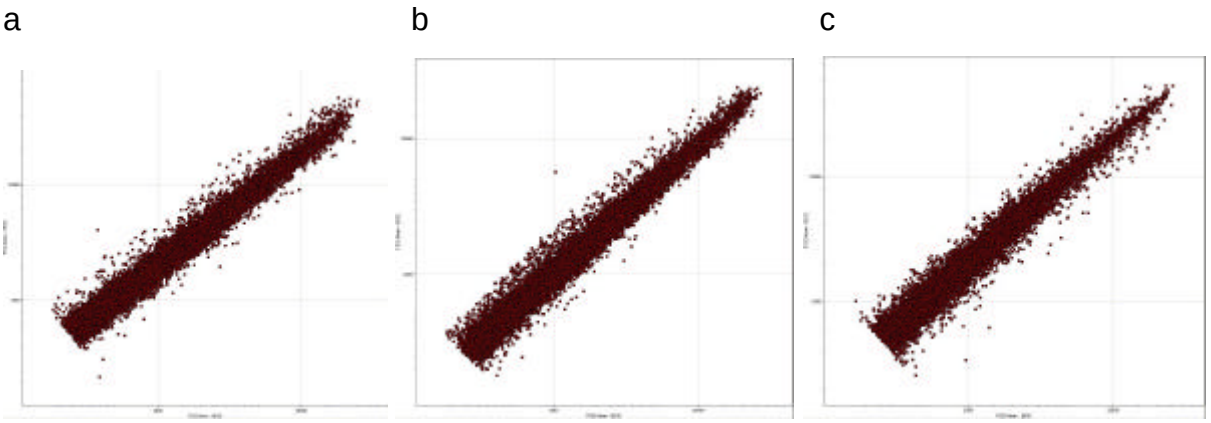


Figure 16 Control of reference consistency. The log-intensities of the normalized references of the three time series are plotted. (a) Reference 2 vs. Reference 3, (b) Reference 3 vs. Reference 4, (c) Reference 4 vs. Reference 2.

After calculating the geometric mean of the dye swapped slides and the arithmetic mean of the three time series the resulting dataset consists of 221184 data points, which is assembled by:

- 80049 data points with no ratio
- 31972 data points resulting from one time series
- 32467 data points averaged by two time series
- 76696 data points averaged by all three time series

Out of this dataset a reliable subset was created for relevant genes. Genes with a value in all time points and in at least four time points strongly up or down regulated compared to the reference (at least double or half of reference) were used for further investigations. The resulting subset consists of 800 genes.

5.4 CLUSTERING

To identify co-expressed genes computational clustering methods were performed. The assumption that co-expressed genes are also co-regulated or at least share some regulatory mechanism is obvious. The subset of 800 differential expressed genes was clustered with Genesis [57]. First Hierarchical clustering with average-linkage was performed. The relation between the genes was measured and visualized in a dendogram (Figure 17a). The over-expressed genes could be grouped into six different clusters. All 800 genes were also clustered with the K-mean method (K=10). This method forms five clusters of mainly up-regulated genes and five clusters of mainly down-regulated genes (Figure 17). Two clusters of K-mean were observed closer (Figure 18).

a b

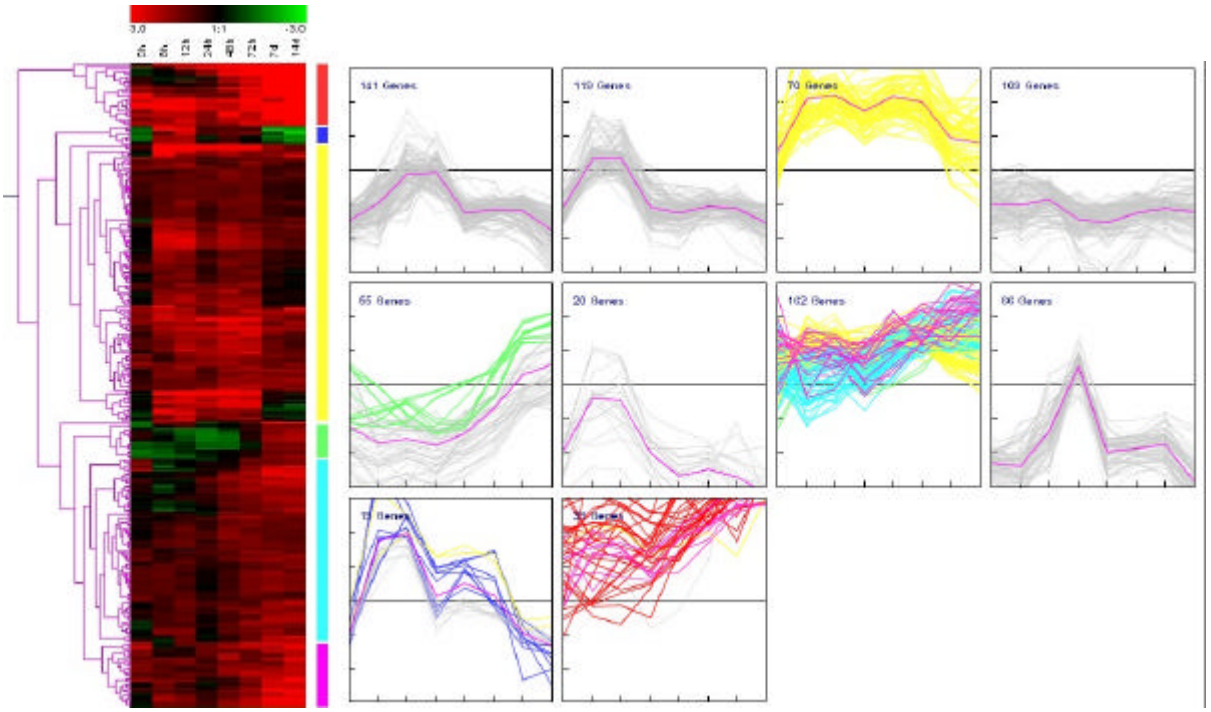
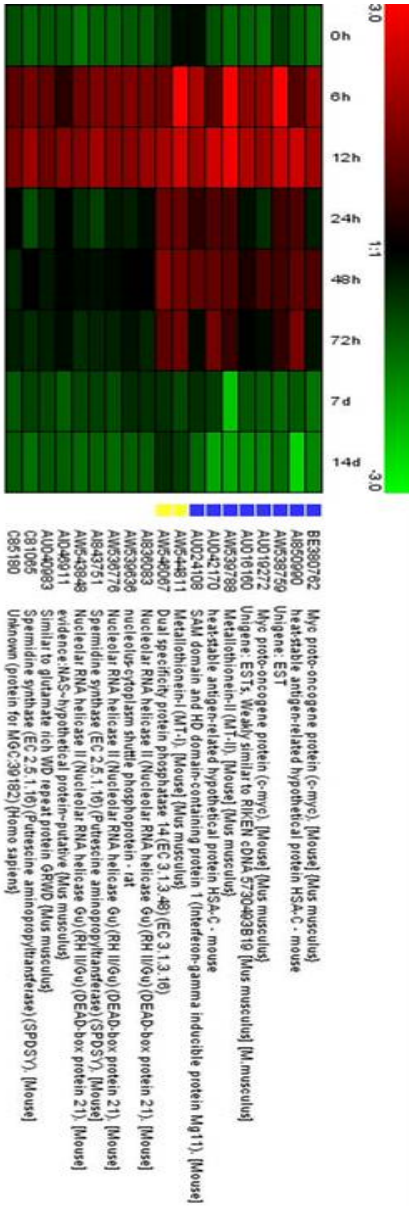


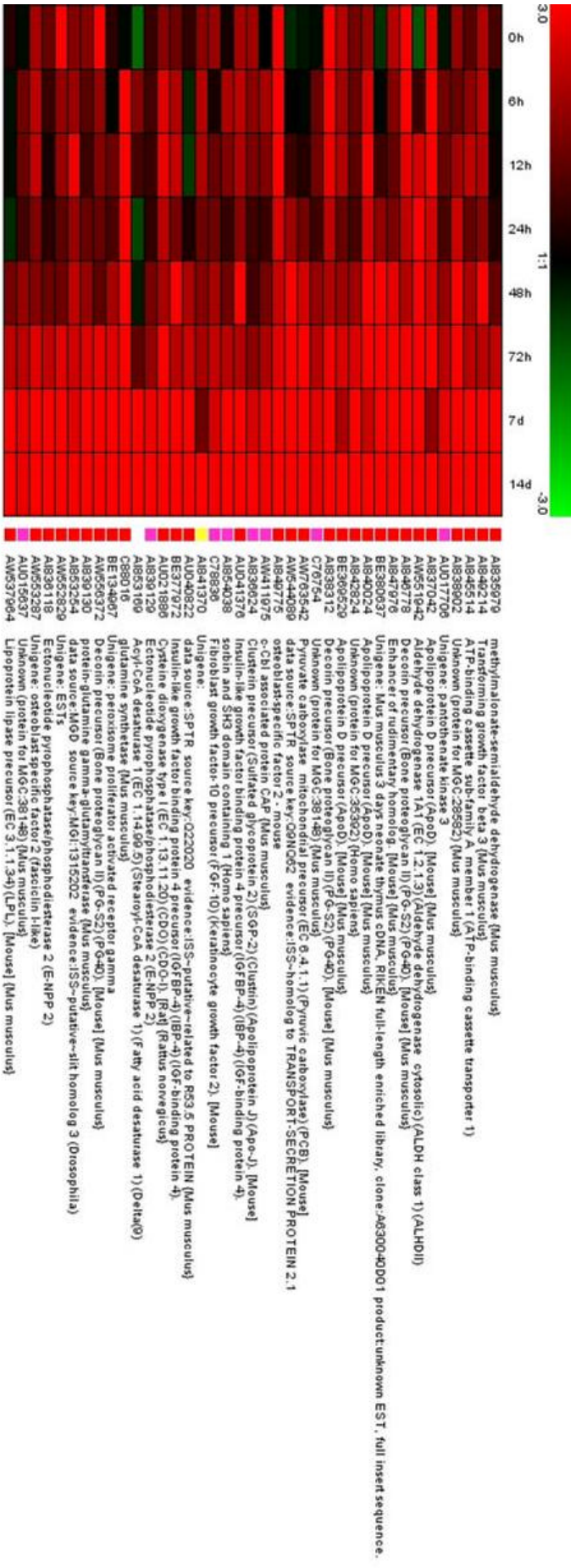
Figure 17 (a) Dendogram of hierarchical clustering and (b) graphs of K-mean clustering (K=10) of a subset of 800 differential expressed genes during adipogenesis. In (a) only mainly upregulated genes during adipogenese are shown which can be grouped into six different clusters. The genes of the colored clusters also cluster together with the K-mean clustering method. The gray lines in (b) are genes which are mainly down regulated in hierarchical clusters

Figure 18 The 9th (a) and 10th (b) K-mean cluster of strong differential expressed genes during adipogenesis (Figure 17). The TIGR-annotation of the ESTs is shown. If this annotation was not available, the annotation from the unigene cluster was obtained, which are shown by a leading Unigene-label. The color beside the gene annotation is a label of the belonging hierarchical cluster which is shown in Figure 17.

a



b



With the the Gene Ontology annotations and literature research, the proteins of the 10th K-mean cluster (Figure 18b) can be grouped by functions:

Table 6 Profil expectations and functional annotations of the genes in the 10th K-mean cluster (Figure 18b)

Expected up-regulation during adipogenesis	
Transcriptional factors	PPAR gamma
Triglycerol/lipid metabolism	Stearoyl-Coenzyme A desaturase 1 Pyruvate carboxylase Lipoprotein lipase (metabolism/catabolism/transport) Panthothenat kinase 3 (CoA biosynthesis)
Transport	Apolipoprotein A ATP-binding cassette transporter 1 (cholesterol)
Regulation of cell growth	Insulin-like growth factor binding protein 4
Nitrogen metabolism	Glutamine synthase
Regulation of blood pressure	Angiotensinogen
Signal transduction	Fibroblast growth factor 10 (FGF-10)
Unexpected up-regulation during adipogenesis	
Growth factor	Transforming growth factor beat 3 (TGFß3)
Cell adhesion	Osteoblast-specific factor 2 (fasciclin I-like)
Up-regulation during adipogenesis	
Metabolism	Aldehyde dehydrogenase 1A1 Methylmalonate-semialdehyde dehydrogenase
Proteolysis and Peptidolysis	Protein-glutamine gamma- glutamyltransferase
Extracellular matrix	Decorin
Nucleotide metabolism	Ectonucleotide pyrophosphatase/phosphodiesteras 2
L-cysteine catabolism to taurine	Cysteine dioxygenase type I
Unknown function	enhancer of rudimentary homology, C-Cbl associated protein CAP, sorbin and SH3 domain containing 1, all other EST entries in this cluster

Only three very interesting genes of known functions belong to the 9th K-mean cluster (Figure 18a): Myc proto-oncogene protein (c-myc), Metallothionein-I (MT-I) and Nucleolar RNA helicase II (RH II/Gu, DEAD-box protein 21).

5.5 DATA RETRIEVAL

Relevant data was retrieved from databases in the world wide web for all 27648 arrayed expressed sequence tags on the microarray.

There could be allocated:

- 24700 UniGene entries
- 17890 LocusLink entries
- 19865 entries with at least one information about the messenger RNA
- 16220 interval positions for the coding sequence of the gene product
- 16370 entries of the protein identification number
- no total complementarity to known micro RNAs

If the 27648 arrayed ESTs are searched for ESTs with multiple unigene entries it shows that there are:

- 16024 unique unigene entries
- 13927 arrayed ESTs have at least a second same entry
- 5243 unique entries in the 13927 multiple entries

6 DISCUSSION

During this master thesis the first large scale cDNA-microarray experiment with an adipose specific chip was performed. Due to the fact that adipogenesis is mainly regulated on the transcriptional level, cDNA-microarray assay is the method of choice to investigate the transcriptional networks and changes during the biological process. The 3T3-L_1 cell line, which is a very well characterized *in vitro* model system for adipocyte differentiation, was maintained to the 6^{th}, 7^{th} and 8^{th} passage and the biological process of differentiation to adipocytes, induced by insulin, isobutylmethylxanthin and dexamethasone, was investigated over a time interval of fourteen days. Three independent time series of different passages were examined. The morphological changes from the fibroblast-like 3T3-L_1 cells to round adipocytes were documented with microscope images (Figure 11). The proof of the fat accumulation in adipocytes could be done through oil-red O staining of the fat droplets and counterstaining of the cytoplasm with Light Green SF Yellowish (Figure 12). Most cells differentiated very well which was verified by the large amount of accumulated triglycerides. The cells differentiated beginning from many small islands in the cell culture dish, which grew in size. This appearance is an indication for autocrine/paracrine signaling pathways during adipogenesis. Another side effect of this mechanism is that not all cells of one cell culture dish are in the same phase of differentiation at the same time. This results into a bell-shaped distribution curve for the stages of differentiation. Certainly there is a maximum for one stage at one time but for further investigations the cross-talk of other stages should be considered.

To analyze the changes in the mRNA profile during adipocyte differentiation total RNA was extracted from preconfluent cultures and at eight time points in three different passages. Even though the 260nm/280nm absorption ratio on the spectrophotometer was mostly too low (Table 4), the integrity and quality showed to be very well on the BioAnalyzer (Table 5). The low ratio may be the result of the dilution with water instead of buffer. Accurate RNA concentrations are measured in water, because the extinction coefficient was calculated for this condition.

The cDNA-microarray assay was performed with the dye swapping technique. This technique doubles the laboratory work and the costs but reduces the technical variability of the method. 51 cDNA-microarrays were hybridized with good quality and

bright signals (Figure 15). The references between the three time series have been consistent (Figure 16). The final dataset consisted of 800 genes which were strongly differentially regulated during adipocyte differentiation and have no missing values in the time interval.

For first data analysis these genes are especially interesting, because of the high up and down regulation in at least four time points (Figure 17). Certainly they can also contain stress factors and cell culture specific factors but the main amount is of adipogenesis nature. For comprehensive research of the transcriptional network transient up and down regulated genes must be investigated too because some factors will be only relevant for a short time during adipocyte differentiation as C/EBP delta which is only significantly up regulated at six hours after induction.

Nevertheless this subset of strong regulated genes gives important information about adipogenesis. A closer look at the strongly up regulated genes in the 10th K-mean cluster (see Figure 17a, Table 6) shows expected and unexpected effects of adipogeneses. The early adipogenesis marker Lipoprotein Lipase (LPL) follows as well as the late central triglycerol metabolism enzyme Stearoyl-Coenzyme A desaturase 1 and other triglyceride metabolism associated enzymes an expected transcription profile during adipogenesis. Also transport proteins for fatty acids and cholesterol are as expected in this mainly upregulated cluster.

One very important transcription factor of adipocyte differentiation is PPAR gamma. The expected increase of concentration during adipogenesis could be verified. The assumption that at least a few of the genes in the 10th K-mean cluster are regulated by PPAR gamma is obvious. Another important information lies in the overexpression of fibroblast growth factor 10 (FGF-10) because without the signal transduction of FGF-10 the expression of C/EBP beta and the subsequent differentiation of the cells is inhibited [61].

The overexpression of transforming growth factor beta 3 (TGF-beta3) during adipogenesis at all time points was not expected. Although expression of TGF-beta3 was found in mature adipose tissue [62], TGF-beta is a potent inhibitor of adipogenesis [27]. Beside TGF-beta3 Decorin is overexpressed with the same time-expression profile. It is proposed that decorin is a bidentate ligand attached to two parallel neighboring collagen molecules in the fibril, helping to stabilize fibrils and orient fibrillogenesis [63]. The protein core can also bind TGF-beta [64]. Recent research showed that interaction between decorin and TGF-beta plays an important

role in the myogenesis [64] and that decorin reverses the repressive effect of autocrine-produced TGF-beta on mouse macrophage activation [66]. All this observations result in the assumption that the interaction between TGF-beta and decorin plays also an important role in adipogenesis.

The 9th K-mean cluster (Figure 18a) consists of genes which are strongly downregulated at the induction (0h) and in the late phase or mature adipocytes (7h and 14h) but are strongly upregulated in the early states of adipogenesis (6h, 12h). This group comprises three very interesting genes: Myc proto-oncogene protein (c-myc), Metallothionein-I (MT-I) and Nucleolar RNA helicase II (RH II/Gu, DEAD-box protein 21). C-myc can play a role in cell growth and maintenance, regulation of transcription and regulation of cell cycle. Expression of c-myc prohibits the normal induction of C/EBP alpha and prevents adipogenesis [67]. C/EBP alpha plays a role in early growth arrest and in terminal differentiation. This represents also the profile of c-myc, because at low concentration of c-myc (beginning and end) the transcription factor C/EBP alpha can be induced and expressed properly.

Metallothionein is a metal-binding and stress-response protein which may have an antioxidant role. This protein is a known secretory protein of adipocytes [68]. The high overexpression during the early stages of adipogenesis indicates that the induction to the adipocyte differentiation is correlated with stress for the cells. At the same time Nucleolar RNA helicase II is strongly overexpressed that unwinds double-stranded RNA and introduces secondary structure to a single-stranded RNA. It is suggested that the helicase plays a role in ribosomal RNA production [69] which is needed for the translational system to synthesize proteins. If the protein synthesis can be directly correlated with the transcription of the Nucleolar RNA helicase II, a suggestion of the protein synthesizing activity at the eight time points can be achieved from the transcriptional profile.

A further subject of this thesis was the retrieval of important information for the arrayed expressed sequence tags. This should ease the further investigation of the huge amount of the microarray data because the ESTs give no information of the function of the belonging genes and proteins. The most data was retrieved from the NCBI databases. The information should always handled with care, because the information belongs to mainly unsupervised data which means that everybody can upload wrong data. This became especially obvious as the mRNA sequences were investigated.

The mRNA was searched for total complementary to the regulatory microRNA elements. Total complementary from mRNAs to microRNAs couldn't be observed as predicted by Tuschel et al [58].

The affiliation from the arrayed expressed sequence tags to their messenger RNA and therefore also to the protein and their function is often maintained with extensive computational methods. Sometimes it happens that an EST is clustered to a false mRNA and therefore everything else is false too. It should also be kept at the back of one's mind that many databases are unsupervised and that there are many wrong sequences and annotations. Another aspect is that the ESTs are only short sequences which can sometimes build wrong hybridizations with the mRNA. Due to these facts microarray data always must be confirmed by other methods such as Real-Time-PCR which is not part of this thesis but which will be studied in the near future.

The transcriptome and the proteome of cells sometimes correlate not very well. Certainly for a complete description of adipocyte differentiation it is important to investigate also the posttranscriptional influences such as post-tranlational modification of proteins. Nevertheless adipogenesis is proposed to be regulated mainly on the transcriptional level ([28], [29]), which makes it a good candidate for microarray research.

In summary, in this thesis the transcriptional level of adipogenesis was analyzed for the first time with large-scale focused microarray.

BIBLIOGRAPHY

[1] Spiegelman BM, Flier JS: Obesity and the regulation of energy balance, *Cell* 23; 104(4): 531-43 (2001)

[2] National Institute of Health (NIH): Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults--The Evidence Report, National Institutes of Health, *Obes Res* 6: 51-209 (1998)

[3] World Health report 2002: Reducing risks, promoting healthy life, Geneva, World Health Organization (2002)

[4] Chopra M, Galbraith S, Darnton-Hill I: A global response to a global problem: the epidemic of overnutrition, *Bulletin of the World Health Organization* 80: 952-958 (2002)

[5] U.S. Department of Health and Human Services: The Surgeon General's call to action to prevent and decrease overweight and obesity, [Rockville, MD]: U.S. Department of Health and Human Services, Public Health Service, Office of the Surgeon General (2001)

[6] Wolf AM, Colditz GA: Current estimates of the economic cost of obesity in the United States, *Obes Res* 6(2): 97-106 (1998)

[7] National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK): Statistics related to obesity and overweight, Online, **www.niddk.nih.gov/health/nutrit/pubs/statobes.htm** (2001)

[8] American Obesity Association, Online, **www.obesity.org** (2002)

[9] Cinti S: The adipose organ, *Kurtis Ed.*, Milano (1999)

[10] Cinti S: Adipocyte differentiation and transdifferentiation: Plasticity of the adipose organ, *J Endocrinol. Invest.*, 25: 823-835 (2002)

[11] Björntorp P: Adipose Tissue, *eJIFCC*, vol 12, no 3, Online, **www.ifcc.org/ejifcc/vol12no3/adipose** (2000)

[12] Lowell BB, Flier JS: Brown adipose tissue, ß3-adrenergic receptors, and obesity, *Annu. Rev. Med*. 48: 307-316 (1997)

[13] Garlid KD, Jaburek M, Jezek P: The mechanism of proton transport mediated by mitochondrial uncoupling proteins, *FEBS Lett.* 428: 10-14 (1998)

[14] Rosen ED, Walkey CJ, Puigserver P, Spiegelman BM: Transcriptional regulation of adipogenesis, *Genes & Development* 14: 1293-1307 (2000)

[15] Kim S, Moustaid-Moussa N: Secretory, endocrine and autocrine/paracrine function of the adipocyte, *J. Nutr.* 130: 3110S-3115 (2000)

[16] Yang Y, Song H, Li R, Zhou L, Zhu Z, Hu R, Han Z, Chen J: The gene expression profiling of human visceral adipose tissue and its secretory function, *Biochem. Biophys. Res. Comm.* 300: 839-846 (2003)

[17] Vernon RG, Houseknecht KL: Adipose tissue: Beyond an Energy Reserve, Ruminant Physiology: Digestion, Metabolism, Growth and Reproduction, *CAB International* 171-186 (2000)

[18] Green H, Kehinde O: Sublines of mouse 3T3 cells that accumulate lipid, *Cell* 11: 169-172 (1974)

[19] Green H, Kehinde O: Spontaneous heritable changes leading to increased adipose conversion in 3T3 cells, *Cell* 7: 105-113 (1976)

[20] Green H, Meuth M: An established pre-adipose cell line and its differentiation in cluture, *Cell* 3: 127-133 (1974)

[21] Goldberg B: Collagen synthesis as a marker for cell type in mouse 3T3 line, *Cell* 11: 169-172 (1977)

[22] Wise LS, Green H: Studies of lipoprotein lipase during the adipose conversion of 3T3 cells, *Cell* 13: 233-242 (1978)

[23] Hirugan A, Sato M Mitsui H: Establishment of a clonal cell line that differentiates into adipose cells in vitro, *In Vitro* 16: 685-693 (1980)

[24] Rubin CS, Hirsch A, Fung C, Rosen OM: Development of hormone receptors and hormonal responsiveness in vitro, *J. biol. Chem*., 253: 7570-7578 (1978)

[25] Kuri-Harcuch W, Green H: Adipose conversion of 3T3 cells depends on a serum factor, *Proc. Natl. Acad. Sci. USA* 75: 6107-6109 (1978)

[26] Björntorp P, Karlsson M, Gustafsson L, Smith U, Sjoström L, Cigolini M, Storck G, Pettersson P: Quantitation of different cells in the epididymal fat pad of the rat, *J. Lipid Res*., 20: 97-106 (1979)

[27] Gregoire FM, Smas CM, Sook Sul H: Understanding Adipose Differentiation, *Physiological Reviews* 78;3: 783-809 (1998)

[28] Moustaid N, Sul HS: Regulation of expression of the fatty acid synthase gene in 3T3-L_1 cells by differentiation and triiodothyronine, *J. Biol. Chem.* 266: 18550-1854 (1991)

[29] Wilkinson WO, Min HY, Claffey KP, Satterberg BL, Spiegelman BM: Control of the adipsin gene in adipocyte differentiation. Identification of distinct nuclear factors binding to single- and double-stranded DNA, *J. Biol. Chem*. 265: 477-482 (1990)

[30] Pairault J, Green H: A study of the adipose conversion of suspended 3T3 cells by using glycerophosphate dehydrogenase as differentiation marker*, Proc. Natl*. *Acad. Sci. USA* 76: 5128-5142 (1979)

[31] Vanderstraeten-Gregoire F : La Différentiation In Vitro des Precurserus des Cellules Adipeuses de Rat (PhD Theses), Louvain: Université Catholique de Louvain (1989)

[32] TimcHenko NA, Wilde M, Nakanishi M, Smith JR, Darlington GJ : CCAAT/enhancer-binding protein alpha (C/EBP alpha) inhibits cell proliferation through p21 (WAF-1/CIP-1/SID-1) protein, *Genes Dev*. 10: 804-815 (1996)

[33] Altiok S, Xi SM, Spiegelman BM: PPARgamma induces cell cycle withdrawal: inhibition of E2F/DP DNA-binding activity via down-regulation of PP2A, *Genes Dev*. 11: 1987-1998 (1997)

[34] Juge-Aubry CE, Gorla-Bajszczak A, Pernin A, Lemberger T, Wahli W, Burger AG, Meier CA: Peroxisome proliferator-activated receptor mediates cross talk with thyroid hormone receptor by competition for retinoid X receptor. Possible role of a leucine zipper-like heptad repeat, *J. Biol. Chem*. 270: 18117-18122 (1995)

[35] Schoonjans K, Staels B, Auwerx J: The peroxisome proliferator activated receptors (PPARs) and their effects on lipid metabolism and adipocyte differentiation. *Biochem. Biophys. Acta*. 1302: 93-109 (1996)

[36] Tontonoz P, Hu E, Devine J, Beale EG, Spiegelman BM: Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor. *Cell* 79: 1147-1156 (1994)

[37] Cao Z, Umek RM, McKnight SL: Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. *Genes Dev.* 5: 1538-1552 (1991)

[38] Kim JB, Spiegelman BM: ADD1/SREBP1 promotes adipocyte differentiation and gene expression linked to fatty acid metabolism. *Genes Dev*. 10: 1096-1107 (1996)

[39] Kim JB, Spotts GD, Halvorsen YD, Shih HM, Ellenberger T, Towle HC, Spiegelman BM: Dual DNA binding specifity of ADD1/SREBP1 controlled by a single amino acid in the basic helix-loop-helix domain. *Mol. Cell. Biol*. 15: 2582-1588 (1995)

[40] Morrison RF, Farmer SR: Insights into the Transcriptional Control of Adipocyte Differentiation, *J. Cell. Biochem. Suppl*. 32/33: 59-67 (1999)

[41] Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM: Expression profiling using cDNA microarray*, Nature Genetic Suppl*. 21: 10-14 (1999)

[42] Schulze A, Downward J: Navigating gene expression using microarrays – a technology review, *Nature Cell Biology* 3: E190-E195 (2001)

[43] Kane MD et al.: Assessment of the sensitivity and specifity of oligonucleotide (50mers) microarrays, *Nucleic Acids Res*. 28: 4552-4557 (2000)

[44] Richter A, Schwager C, Hentze S, Ansorge W, Hentze MW, Muckenthaler: Comparison of fluorescent tag DNA labeling methods used for expression analysis by DNA microarrays, *Biotechniques* 33/3: 620-630 (2002)

[45] Brazma A, Vilo J: Gene expression data analysis, *FEBS Letters* 480: 17-24 (2000)

[46] Quackenbush J: Computational analysis of microarray data, *Nature Review Genetics* 2: 418-427 (2001)

[47] Lindl T: *Zell- und Gewebekulturen, 4. Auflage* Heidelberg-Berlin, Spektrum Akademischer Verlag (2000)

[48] European Collection of Cell Cultures, Online, **www.ecacc.org.uk**, ECACC No. 86052701

[49] Prokesch A: Large Scale Gene Expression Profiling during Adipogenesis of Murine Bone Marrow Stromal Cells, *Master thesis*, Graz (2002)

[50] LeLay S, Lefrere I, Trautwein C, Dugail I, Krief S: Insulin and sterol regulatory element binging protein-1c regulation of gene expression in 3T3-L_1 Adipocytes: Identification of C/EBPß as a SREBP-1c target, *J Biol Chem* M203913200 (2002)

[51] GibcoBRL life technologies™: TRIzol® Reagent manual, Online, **http://www.lifetech.com/content/sfs/manuals/15596026.pdf**

[52] Schwartu R, Phoenix T: *Learning Perl, 3rd Edition*, Bejing-Cambridge-Farnham-Köln-Paris-Sebastopol-Taipei-Tokyo, O'Reilly (2001)

[53] Hasseman J: Aminoallyl Labeling of RNA for Microarrays, Standard Operating Procedure, *The Institue for Genomic Research, Rockville, USA* (2002)

[54] Hasseman J: Microarray Labeled Probe Hybridization, Standard Operating Procedure, *The Institue for Genomic Research, Rockville, USA* (2002)

[55] Ambion: Amino Allyl MessageAmp^{TM} a RNA Kit, Online, **www.ambion.com** (2002)

[56] Quackenbush J: Microarray data normalization and transformation, *Nature Genetic Supplements* 32 (2002)

[57] Sturn A et al., Bioinformatics (2002)

[58] Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T: New microRNAs from mouse and human, *RNA* 9: 175-179 (2003)

[59] Qiagen: QiagenNews, Online, *Issue 5,* **www.qiagen.com**, (1997)

[60] Wilfinger W, Mackey M, Chomczynski P: Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity, *BioTechniques* 22: 474 (1997)

[61] Sakaue H, Konishi M, Ogawa W, Asaki T, Mori T, Yamasaki M, Takata M, Ueno H, Kato S, Kasuga M, Itoh N: Requirement of fibroblast growth factor 10 in development of white adipose tissue, *Genes and Development* 16, 8: 908-912 (2002)

[62] Miller DA, Lee A, Matsui Y, Chen EY, Moses HL, Derynck R: Complementary DNA cloning of the murine transforming growth factor-beta 3 (TGF beta 3) precursor and the comparative expression of TGF beta 3 and TGF beta 1 messenger RNA in murine embryos and adult tissues, *Mol Endocrinol* 3;12: 1926-34 (1989)

[63] Scott JE: Proteodermatan and proteokeratan sulfate (decorin, lumican/fibromodulin) proteins are horseshoe shaped. Implications for their interactions with collagen, *Biochemistry* 9;35(27): 8795-9 (1996)

[64] Hildebrand A, Romaris M, Rasmussen LM, Heinegard D, Twardzik DR, Border WA, Ruoslahti E: Interaction of the small interstitial proteoglycans biglycan, decorin and fibromodulin with transforming growth factor beta, *Biochem J* 1;302(Pt 2): 527-34 (1994)

[65] Riquelme C, Larrain J, Schonherr E, Henriquez JP, Kresse H, Brandan E: Antisense inhibition of decorin expression in myoblasts decreases cell responsiveness to transforming growth factor beta and accelerates skeletal muscle differentiation, *J Biol Chem* 2;276(5): 3589-96 (2001)

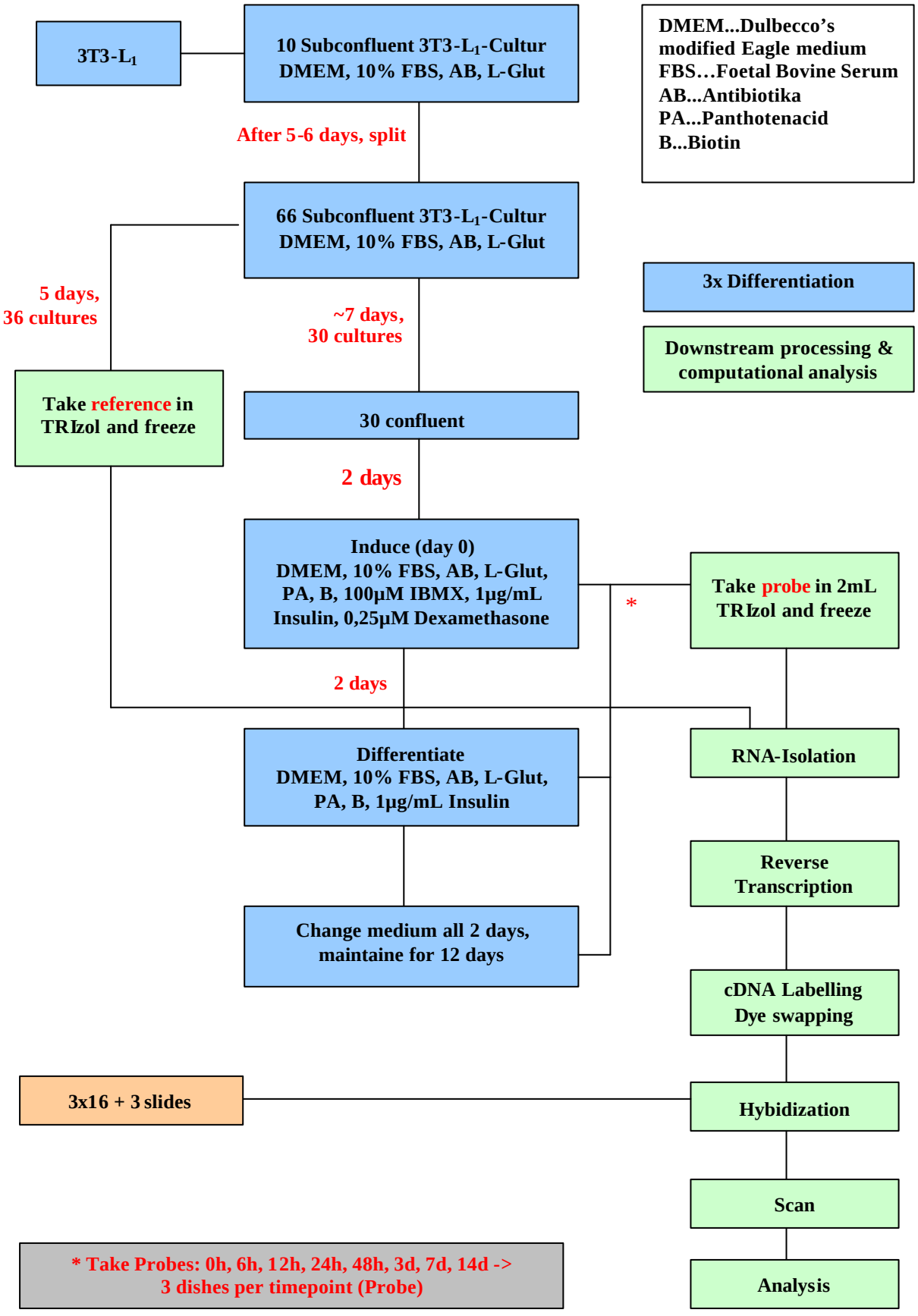
[66] Comalada M, Cardo M, Xaus J, Valledor AF, Lloberas J, Ventura F, Celada A: Decorin reverses the repressive effect of autocrine-produced tgf-Beta on mouse macrophage activation, *J Immunol* 1;170(9): 4450-6 (2003)

[67] Freytag SO, Geddes TJ: Reciprocal regulation of adipogenesis by Myc and C/EBP alpha, *Science* 17;256(5055): 379-82 (1992)

[68] Trayhurn P, Beattie JH: Physiological role of adipose tissue: white adipose tissue as an endocrine and secretory organ, *Proc Nutr Soc* 60(3): 329-39 (2001)

[69] Valdez BC, Perlaky L, Henning D: Expression, cellular localization, and enzymatic activities of RNA helicase II/Gu(beta), *Exp Cell Res* 10;276(2): 249-63 (2002)

APPENDIX A – EXPERIMENTAL FLOWCHART



APPENDIX B – MICROARRAY ASSAY

Aminoallyl Labeling

- To 24µg total RNA (in DEPC water) add 2µL Random Hexamer primers (3mg/mL) and bring to the final volume up to 18.5µL with DEPC water
- Mix well and incubate at 70°C for 10 minutes
- Snap-freeze in ice for 1 minute, centrifuge briefly at 15,700 rcf and continue at room temperature
- Add:

5X First Strand Buffer	6 µL
0.1M DTT	3 µL
50X aminoallyl-dNTP mix	0.6 µL
SuperScript II RT (200U/µL)	2 µL

- Mix and incubate at 42°C overnight
- Centrifuge briefly at 15,700 rcf
- To hydrolyze RNA, add:

1M NaOH	10 µL
0.5M EDTA	10µL

- Mix and incubate at 65°C for 15 minutes
- Centrifuge briefly at 15,700 rcf
- Add 10µL of 1M HCl to neutralize pH.

Qiagen Cleanup

1. Mix cDNA reaction with 300µL (5X reaction volume) buffer PB (Qiagen supplied) and transfer to QIAquick column

2. Place the column in a 2 mL collection tube (Qiagen supplied) and centrifuge at ~15,700rcf for 1 minute. Empty collection tube.
3. To wash add 750µL phosphate wash buffer to the column and centrifuge at 15,700rcf for 1 minute
4. Empty the collection tube and repeat the wash and centrifugation step (3.)
5. Empty the collection tube and centrifuge an additional 1 minute at maximum speed.
6. Transfer the column to a new 1.5 mL tube and carefully add 30µL phosphate elution buffer to the centre of the column membrane
7. Incubate for 1 minute at room temperature
8. Elute by centrifugation at ~15,700 rcf for 1 minute
9. Elute a second time into the same tube by repeating the step 6-8
10. Transfer the liquid to a 0.5mL tube
11. Dry sample in a speed vac (Run: 45 minutes; Heat: 30 minutes, 42°C)

Coupling aa-CDNA to Cy Dye Ester

1. Resuspend aminoallyl-labeled cDNA in 4.5µL of 0.1M sodium carbonate buffer, pH 9.1
2. Add 4.5µL of the appropriate NHS-ester Cy dye
3. Incubate the reaction for 1 hour in the dark at room temperature

Reaction Purification II: Removal of uncoupled dye

1. To the reaction add 35µL 100mM NaOAc pH 5.2
2. Add 250µL (5x reaction volume) Buffer PB (Qiagen supplied)
3. Place a QIAquick spin column in a 2mL collection tube (Qiagen supplied), apply the sample to the column, and centrifuge at ~15,700 rcf for 1 minute. Empty collection tube.
4. To wash, add 750µL Buffer PE (Qiagen supplied) to the column and centrifuge at ~15,700 rcf for 1 minute
5. Empty collection tube and centrifuge column for an additional 1 minute at maximum speed

6. Place column in a clean 1.5mL microfuge tube and carefully add 30µL Buffer EB (Qiagen supplied) to the center of the column membrane
7. Incubate for 1 minute at room temperature
8. Elute by centrifugation ~15,700 rcf for 1 minute
9. Elute a second time into the same tube by repeating steps 6-8
10. Transfer the liquid to a autoclaved 0.5mL tube
11. Dry sample in a speed vac (Run: 45 minutes; Heat: 30 minutes, 42°C)

Prehybridization

1. Preheat prehybridization buffer ~30 minutes before use
2. Place the printed slides which will be used for the hybridization in a Coplin jar containing preheated prehybridization buffer and incubate at 42°C for 45 minutes
3. Fill five Coplin jars with MilliQ water and another with isopropanol
4. With forceps carefully grasp slide by the labeled end and vertically dip slide into the first Coplin jar (water) ten times. Repeat with the next four coplin jars filled with water
5. Finally dip slide ten times into the Coplin jar filled with isopropanol
6. To dry centrifuge slide immediately with 1500 rcf, 2 minutes
7. For quality check put slides into scanner and scan a preview

Hybridization

1. Preheat 1X hybridization buffer (50% formamide, 5X SSC and 0.1% SDS) at 42°C
2. Resuspend labeled probe in 24µL of 1X hybridization buffer
3. Transfer the liquid to the appropriate reference and resuspend it
4. To block non-specific hybridization add:

COT1-DNA (20µg/µL) .....1 µL
Poly(A)-DNA (20µg/µL).....1 µL

5. To denature, heat the probe mixture at 95°C for 3 minutes and snap cool on ice for 30 seconds
6. Centrifuge the probe mixture at maximum angular velocity for 1 minute. Keep at room temperature and use immediately
7. Put 10µL ddH_2O to the wells in the hybridization chamber
8. Place a prehybridized microarray slide into the hybridization chamber
9. Pipette the labeled probe mixture to the slide surface near one end of the array pint area keeping bubbles to a minimum
10. Take a microscope glass coverslip, dust it with compressed air
11. Holding the coverslip over the array print area, lower the end nearest the pool of cDNA probe until solution wicks to the surface of the coverslip. Drop the other end onto a pipette tip
12. Gradually lower the end of the coverslip by pulling the pipette tip out
13. Cover and seal the chamber
14. Wrap the chamber in foil and incubate in a 42°C water bath for 16-20 hours (we did 20 hours). Newer flip the chamber.

Washing the slide

1. Preheat Washing buffer #1 at 42°C
2. Remove foil and unseal hybridization chamber. Remove the slide from the chamber, taking care not to disturb the coverslip
3. To remove coverslip submerge slide in a dish containing Washing buffer #1 and incubate 2 minutes at 42°C.
4. Pull the slide slowly out, so the coverslip remains in the liquid
5. Place the slide in a staining dish containing wash buffer #1 (42°C preheated) and agitate for 2 minutes
6. Wash the slide in a staining dish with wash buffer #2 by agitating for 4 minutes by room temperature
7. Wash the slide in a staining dish with wash buffer #3 by agitating for 4 minutes by room temperature and repeat this step by agitating for 2.5 minutes.
8. Dip five times in two Coplin jars filled with water
9. To dry centrifuge slide immediately with 1500 rcf, 2 minutes

10. Place slide in a light tight slide box until they can be scanned, preferably as soon as possible

Chemical for Microarray assay

Reagent		Amount for one hybridization		Amount for 50 hybridizations	
50X aminoallyl-dNTP mix	50X	1.2	µl	60	µl
dATP (100mM)	25 mM	0.3	µl	15	µl
dCTP (100mM)	25 mM	0.3	µl	15	µl
dGTP (100mM)	25 mM	0.3	µl	15	µl
dTTP (100mM)	15 mM	0.18	µl	9	µl
aa-dUTP (100mM)	10 mM	0.12	µl	6	µl
Phosphate wash buffer		3	ml	150	ml
Ethanol 95%		2.53	ml	126.5	ml
K2HPO4 1M		14.25	µl	712.5	µl
KH2PO4 1M		0.75	µl	37.5	µl
Phosphate elution buffer	4mM	120	µl	6000	µl
K2HPO4 1M		0.45	µl	22.5	µl
KH2PO4 1M		0.024	µl	1.2	µl
Na2CO3 pH=9,0	0,1 M	9	µl	450	µl
Na2CO3		97	µg	4850	µg
HCl 12N		0.1	µl	5	µl
Prehyb buffer	1X	60	ml	3000	µl
BSA	1%	0.6	g	30	g
20X SSC	5X	15	ml	750	ml
SDS 10%	0.10%	0.6	ml	30	ml
1X Hyb buffer	1X	24	µl	1200	µl
20X SSC	5X	6	µl	300	µl
SDS 10%	0.10%	0.24	µl	12	µl
Formamide (100 %)	50%	12	µl	600	µl
Cot1-DNA	20 µg/µl	1	µl	50	µl
Cot1-DNA		20	µg	1000	µg
poly(A)-DNA	20 µg/µl	1	µl	50	µl
poly(A)-DNA		20	µg	1000	µg
Wash buffer #1		310	ml	15500	ml
20X SSC	1X	15.5	ml	775	ml
SDS 10%	0.20%	6.2	ml	310	ml
Wash buffer #2		250	ml	12500	ml
20X SSC	0,1 X	1.25	ml	62.5	ml
SDS 10%	0.20%	5	ml	250	ml
Wash buffer #3		500	ml	25000	ml
20X SSC	0,1 X	1.25	ml	62.5	ml

Product	Company	Catalog number
Random Hexamer Primer (3 mg/ml)	Invitrogen	48190-011
DTT*	Invitrogen	18064-071
5X First Strand Buffer*	Invitrogen	18064-071
SuperScript II RT	Invitrogen	18064-071
NaOH (1M)	Roth (Lactan)	K021.1
EDTA	Sigma	E-7889
HCl (1M)	Roth (Lactan)	K025.1
QIAquick PCR Purification Kit**	QIAGEN****	28104
Cy3 mono reactive dye	Amersham Biosciences	PA23001
Cy5 mono reactive dye	Amersham Biosciences	PA25001
DMSO	Sigma	472301
Cot1-DNA	Invitrogen	18440016
poly(A)-DNA 18 mer	HVD Life Sciences (MWG)	1210-100100
Isopropanol	Sigma	D8418
dATP (100mM)	Invitrogen	10216018
dCTP (100mM)	Invitrogen	10217016
dGTP (100mM)	Invitrogen	10218014
dTTP (100mM)	Invitrogen	10219012
aa-dUTP (100mM)	Sigma	A0410
Ethanol 95%	VWR	1.08543.0250
K2HPO4 1M	Sigma	P-8709
KH2PO4 1M	Sigma	P-8584
Na2CO3	Merck	1063920500
HCl 12N	Merck***	1003171000
BSA	Sigma	A-9418
20X SSC	Sigma	S-6639
SDS 10%	Invitrogen	15553-035
SDS/Natriumlaurylsulfat	Roth (Lactan)	4360.2
Formamide (100 %)	Sigma	F-9037
DEPC Treated H2O	Invitrogen	750024

* Available together with SuperScript II RT
** incl. columns, wash and elutionbuffer
*** used solutions, now available by VWR
**** distributed by VWR