

6 Zusammenfassung und Perspektive

6.1 Zusammenfassung

Die Array-Technologie hat bereits in weiten Gebieten der Biologie und Medizin Einzug gehalten. Sie ist ein wesentlicher und unverzichtbarer Bestandteil der heutigen Forschung in der Biotechnologie. Eine Vielzahl an anwendungsbezogenen Variationen dieser effizienten Hochdurchsatz-Technologie wurde in den letzten Jahren entwickelt und zum Einsatz gebracht.

In dieser Arbeit wurde zur Analyse zweier Leukämie-Zell-Linien, Jurkat und NKL, die bewährte Methode des Oligonucleotid Fingerprintings genutzt, einer Methode des Sequenzierens durch Hybridisieren. Anhand dieser Methode sollten die in jeder Zell-Linie exprimierten Gene identifiziert sowie ein differentielles Display der Genexpression beider Zell-Linien erstellt werden. Weiterhin sollte das ONF genutzt werden, die Redundanz der exprimierten Sequenzen zu verringern und somit die Bibliotheken zu normalisieren. Es sollte eine Vorauswahl getroffen werden, um Unigene-Sets für weiterführende Experimente zu erstellen und neue Kandidatengene zu finden.

Im Rahmen dieser Arbeit wurden zwei cDNA-Bibliotheken aus etablierten Zell-Linien leukämischer Patienten erstellt. Dabei handelt es sich zum einen um eine akute T-Zell-Leukämie (T-ALL) und zum anderen um eine akute NK-Zell-Leukämie (NK-LGL). Die Bibliothek der T-Zell-Linie Jurkat enthält etwa 23.100 Klone, die Bibliothek der NK-Zell-Linie NKL etwa 34.000 Klone. Die Analyse beider leukämischer Zell-Linien umfaßte folgende Abschnitte:

Die gerichtete Klonierung der cDNA-Fragmente in einen Vektor, der ein Derivat des Expressionsvektors pQE30 darstellt, ermöglicht in weiterführenden Untersuchungen die Expression His₆-markierter (His-tag) Proteine beider Bibliotheken.

Die inserierten Fragmente beider Bibliotheken wurden mittels PCR amplifiziert und unter Verwendung eines Roboters (*spotting*) als hoch dichte Arrays auf Nylonmembranen transferiert (*high density filter*).

Die Oligonucleotid Fingerprinting-Hybridisierungen beider Bibliotheken wurden mit 225 radioaktiv markierten Oligonucleotiden durchgeführt. Aufgrund der sich aus den Hybridisierungen ergebenden Fingerprints und des paarweisen Vergleichs dieser Fingerprints miteinander konnten die Klone der einzelnen Bibliotheken in Gruppen gleicher bzw. ähnlicher Sequenzen, sogenannter Cluster, die für ein und dasselbe Gen codieren, geordnet werden (*clustering*).

Basierend auf der bekannten Sequenz der hybridisierten Oligonucleotide ergibt sich ein experimenteller Fingerprint aller Klone der Bibliothek. Aus allen Klonen, die ein Cluster bilden, wurde jeweils ein Consensus-Fingerprint des entsprechenden Clusters ermittelt. Diese Consensus-Fingerprints wurden mit theoretischen Fingerprints verglichen, die anhand von *in silico*-Hybridisierungen der Oligos von Genen aus den Datenbanken erstellt wurden, wodurch eine Identifizierung der meisten Cluster möglich wurde. So können in der Datenbank bekannte Gene schnell identifiziert und solche Cluster, die keinen signifikanten Treffer aufweisen, als eventuelle neue Kandidatengene ermittelt werden.

Anschließend an das Clustering der einzelnen Bibliotheken wurde das Co-Clustering beider Bibliotheken durchgeführt, so daß ein Vergleich der Genexpression-Pattern beider Zell-Linien möglich wurde. Damit konnte ein differentiell Display erstellt werden, das Gemeinsamkeiten und Unterschiede hinsichtlich der Expression verschiedenster Gene widerspiegelt. So konnten insgesamt 24.057 Co-Cluster ermittelt werden. Davon sind 1.409 Co-Cluster solche mit einer Größe von mindestens 5 Mitgliedern, 2.814 mit einer Anzahl von weniger als 5 Klonen. 19.832 Klone bildeten Cluster mit nur einem Mitglied, verbleiben also als Unikate (Singletons).

Die Anzahl der Mitglieder eines Clusters, die Clustergröße, spiegelt das relative Expressionsniveau des entsprechenden Clusters und somit des vom Cluster repräsentierten Gens in der Zelle wider. So kann man von der Anzahl der Mitglieder in einem Cluster direkt auf die relative Häufigkeit schließen, mit der das Gen in der Zelle vorkommt. So gibt es nur wenige Cluster, die Gene mit hoher Expression repräsentieren und eine etwas größere Anzahl an Clustern mit mittlerer Expression. Die meisten Cluster stehen für Gene mit einer geringeren, fein regulierten Expression.

Von den 1.409 Co-Clustern mit mehr als 5 Mitgliedern werden 538 Co-Cluster in beiden Bibliotheken signifikant differentiell repräsentiert. Von diesen 538 differentiell exprimierten Co-Clustern kommen 324 NKL-spezifisch und 214 Jurkat-spezifisch vor. 135 Co-Cluster mit mehr als 5 Mitgliedern konnten bisher nicht identifiziert werden und stellen somit eventuell neue Kandidatengene für weitere Studien dar. Von diesen 135 unbekanntem Co-Clustern werden 43 differentiell exprimiert. 19 Cluster davon kommen differentiell in der Zell-Linie Jurkat zur Expression, 24 in der NKL-Linie.

Zur Validierung der Co-Clustering-Ergebnisse wurden Northern-Hybridisierungen durchgeführt. Im Allgemeinen konnten die Erkenntnisse hinsichtlich der Tendenz der differentiellen Expression bestätigt werden, darunter beispielsweise auch die Gene für CD3E ϵ oder NKG2-D, die für ihre ausschließliche Expression in T- bzw. NK-Zellen in der Literatur bekannt sind. Die genauen Verhältnisse der Expressionshäufigkeit aus den Clustering-Analysen jedoch ließen sich nicht in jedem Fall einwandfrei nachweisen, was insbesondere auf die kleineren Cluster

zutrifft, Cluster mit weniger als 20 enthaltenen Klonen. Das bedeutet, daß das Oligonucleotid Fingerprinting besonders effizient im Fall hoch exprimierter Gene ist. Die Ergebnisse aus den Clustering-Analysen bedürfen also zur exakten Bestimmung der Expressionshöhe der Validierung durch solche direkten Methoden wie der Northern-Hybridisierungen.

Anhand der Clustering-Ergebnisse wurde aus den Consensus-Klonen eines jeden Clusters der beiden Bibliotheken Jurkat und NKL jeweils ein nicht redundantes Set an Sequenzen (cDNA-Klonen) erstellt (*rearranging*). Das Jurkat- bzw. NKL-Unigene-Set besteht aus 10.506 bzw. 13.409 Klonen. Die ursprünglichen Bibliotheken konnten auf diese Weise 2.25-fach (Jurkat) bzw. 2.7-fach (NKL) normalisiert werden.

Das Jurkat- sowie das NKL-Unigene-Set stellen bedeutende Tools zweier Leukämie-Zell-Linien, die Zellen des Immunsystems sind, für weitere, speziellere Untersuchungen im Bereich der Immunologie bereit. Komplex-Hybridisierungen dieser Unigene-Sets ermöglichen es, die Genexpression verschiedener Zellpopulationen oder Gewebe direkt miteinander zu vergleichen und können zum Screening von Leukämiepatienten in Studien zur molekularen Charakterisierung der Medikamentenresistenz genutzt werden und um Targets für neue Medikamente zu finden.

6.2 Perspektiven

Die Erkenntnisse und Ergebnisse dieser Arbeit bieten eine Reihe an weiteren aussichtsreichen Anwendungsmöglichkeiten.

Im Fall der Natural Killer-Zellen wurden bisher nur wenige Zell-Linien etabliert. Von diesen Zellen ist bisher weniger bekannt als vergleichsweise von T- oder B-Lymphocyten. NK-Zell-spezifische Gene können in einem Unigene-Set angeordnet und charakterisiert werden. Die Untersuchungen dieser Arbeit sowie die sich daran anschließenden Experimente können dazu beitragen, die Funktion normaler und entarteter NK-Zellen aufzuklären.

Neben den bekannten Genen konnten auch 135 namentlich noch nicht in der Datenbank vertretene Sequenzen (neue Gene) gefunden werden. Diese und auch solche Klone, die keinen signifikanten Treffer in der Datenbank aufweisen, können mittels eines Rearranging-Roboters in Microtiterplatten neu angeordnet und in weiteren Untersuchungen, wie der Sequenzierung, näher charakterisiert werden. Das ermöglicht das Identifizieren neuer Gene in Bezug auf den Pathomechanismus der Leukämieerkrankungen und anderer Erkrankungen des Immunsystems. So können in naher Zukunft neue Gene immunologischer Relevanz gefunden werden.

Da die Klonierung der Fragmente in einen Expressionsvektor erfolgte, ist es möglich, in einigen Klonen eine Proteinexpression der klonierten Fragmente zu induzieren. Etwa 20% der Fragmente sollten im richtigen Leserahmen (*open reading frame*, ORF) kloniert vorliegen. Im falschem ORF klonierte Fragmente produzieren keine Proteine oder nur unvollständige, da es zum frühzeitigen Translationsabbruch aufgrund häufiger auftretender Stopcodons kommen kann. Beide Bibliotheken können so auch genutzt werden, die Proteinexpression in Leukämiezellen zu untersuchen oder aber auch, um Antikörper zu generieren, beispielsweise für die Krebs-Vaccination.

Das T-Zell-spezifische Unigene-Set stellt ein wertvolles Tool für weitere Analysen leukämischer Zellen zur Verfügung. So kann das Unigene-Set Anwendung finden in einer Reihe verschiedenster Experimente, wie dem Screening neuer Antikörper und Medikamente oder auch der Komplex-Hybridisierung.

Da beide Zell-Linien ursprünglich von Leukämiepatienten isoliert wurden, enthalten die Bibliotheken auch leukämiespezifisch exprimierte Gene, die in Zellen eines gesunden Gewebes nicht aktiviert oder aber inhibiert sind. Komplex-Hybridisierungen mit RNA normaler Lymphocyten einerseits und Lymphocyten von Patienten andererseits können Aufschluß über eine eventuelle Beteiligung verschiedener Gene an der Carcinogenese, insbesondere der Leukämogenese geben.

Mit Hilfe von Komplex-Hybridisierungen des Unigene-Sets mit Lymphocyten-RNA verschiedener Patientengruppen können detaillierte genetische Profile erstellt werden, mit denen die Krankheiten genauer diagnostiziert und klassifiziert werden können. Diese genetischen Profile geben auch die Möglichkeit zu neuen Definitionen neuer pathologischer Subklassen, die mit den bisherigen klassischen klinischen Methoden nicht erkennbar sind. Weiterhin wird es möglich sein, neue Marker für die Empfänglichkeit bestimmter Individuen und Organismen bestimmten immunologischen Krankheiten gegenüber zu entdecken.

Hybridisierungen von Lymphocyten-RNA von Leukämiepatienten vor und nach einer spezifischen Therapie können dazu beitragen, den Therapieerfolg der Behandlung zu überwachen und zu überprüfen. Ebenso können mit solchen Untersuchungen minimal residuale Zellen in behandelten Patienten nachgewiesen werden. Weitere Anwendungen liegen in der molekularen Charakterisierung der Grundlagen der Medikamentenresistenz oder aber auch in der Identifizierung von Targets für Medikamente. Damit können der Therapieerfolg und die Überlebenschancen der Patienten erheblich verbessert werden.

6.3 Summary

The array technology is an already well accepted tool in the field of biology and medicine. In many cases it became an essential component of a modern molecular biology. A vast number of applications for microarray technology has been developed in recent years thus creating a separate field of science.

In this work we took advantage of oligonucleotide fingerprinting technique to analyze two leukemic cell lines, Jurkat and NKL. The aim of this study was to identify gene expression pattern in each of these cell lines as well as to create a differential expression pattern for both cell lines. The oligonucleotide fingerprinting was also used to create an unigene set, a collection of non-redundant cDNA clones for further downstream applications, e.g. complex hybridization studies. Furthermore this approach allowed us to identify new candidate genes present in both libraries.

The intention behind this work was to characterize the expression profile of two cell types involved in the leukemic process. Moreover it has been aimed to create a powerful research tool, the unigene set, for functional genomics studies of the leukemic process. We subjected RNA from two leukemic cell lines to oligonucleotide fingerprinting procedure. The analysis comprised following steps:

Directed cloning of the cDNA fragments in an expression vector pQE30 allowing for expression of His₆-tagged proteins.

The cloned fragments of both libraries were amplified by PCR and spotted onto Nylon membranes using a spotting robot.

Oligonucleotide fingerprinting hybridizations were performed using 225 radioactively labeled oligoprobes. According to the pattern of the hybridizations and the pairwise comparison of the individual libraries the clones were clustered into groups -so called clusters- representing collection of identical or similar sequences coding for the same gene.

On the base of the known sequence of the oligoprobes an experimental fingerprint for the analyzed clones was created. Within each cluster a consensus fingerprint was calculated. The consensus fingerprints were compared with the theoretical fingerprints created by *in silico* hybridizations of the oligoprobes with sequences available in public database. By that approach it was possible to annotate most of the genes to particular clusters. Those experimental fingerprints, which have not had any match in the database, have been considered as potential new candidate genes.

Following the clustering of the individual libraries a co-clustering procedure of both libraries was performed in order to create differential expression pattern of both cell lines. In total 24.057 co-clusters were obtained. 1.409 co-clusters contained more than 5 clones. 2.814 co-clusters consisted of less than 5 members. 19.832 clones had unique fingerprints thus were classified as singletons (clusters with a single member).

The number of clones from both libraries within each particular cluster (cluster size) provides information about their relative abundance thus reflecting the expression level of the particular gene. Usually only a few clusters contain high numbers of clones indicating high expression level of the respective genes. Most of the genes are expressed at medium or low levels. Hence the number of clusters increases with decreasing size of a cluster.

In a present study, 538 co-clusters out of 1.409 co-clusters with more than 5 members were found to be significantly differently represented between these two libraries. 324 co-clusters were more frequently present in NKL, whereas 214 co-clusters were found to be differently expressed in Jurkat cells. 135 co-clusters revealed no match in database and therefore they have been considered as potentially new candidate genes. A number of 43 co-clusters of these 135 potentially new genes are expressed differentially in both libraries, 19 are expressed differentially in Jurkat and 24 in NKL.

Co-clustering results have been confirmed in Northern blot analyses in selected cases. As examples serve CD3E ϵ gene and NKG2-D, which have been confirmed to be specifically expressed in T and NK cells, respectively. This remained in agreement with the data described in literature.

In the case of smaller clusters (below 20 members) we observed that their cell-type-specificity could not be confirmed in all cases by Northern blot analysis. This suggests that the co-clustering procedure is most efficient in the case of highly expressed genes.

On the base of clustering results non-redundant unigene set was created for each analyzed library. The unigene set represents a collection of non-redundant cDNA clones, which were selected from each cluster on the base of the closest similarity to the consensus fingerprint of that cluster. The Jurkat and NKL unigene sets consist of 10.506 and 13.409 clones, respectively. Thus the initial libraries could be normalized 2.25-fold in case of Jurkat and 2.7-fold in case of NKL cells.

Jurkat- and NKL-specific unigene sets comprise an unique research tool in a field of immunology. The next potential step in this investigation would be an usage of the two unigene

sets for screening of leukemia patients toward drug targeting and molecular characterization of drug resistance.

6.4 Perspectives

The results presented in this work suggest a number of possible applications.

In case of Natural Killer cells there are only a very few cell lines established. Therefore, the information concerning NK-specific gene expression obtained in this study can very much contribute to further research on function of normal and malign NK cells. NK cell specific cDNA clones selected throughout this project can be rearranged in an unigene set thus creating a valuable tool for expression studies.

We found 135 genes, which so far have not been described. Further characterization of those genes could contribute to elucidation of pathomechanism of leukemic process.

Since cDNA fragments are cloned in an expression vector it is possible to induce protein expression in some of the cloned fragments. About 20% of the fragments should be cloned in the correct reading frame. Fragments cloned in incorrect frame will not produce any protein or incomplete proteins due to premature termination of translation due to increased frequency of stop codons. Both libraries can be used to investigate protein expression in leukemic cells as well as to produce antibodies against cellular antigens, e.g. for anti-cancer vaccines.

The T cell specific Jurkat unigene set provides a valuable tool for further investigation and analysis of leukemic cells. It may be utilized in a number of various experiments concerning the screening of new antibodies and drugs as well as in complex hybridizations.

Both cell lines were originally isolated from leukemic patients. Therefore genes specifically expressed in leukemic diseases are included, which will not be activated or are inhibited in healthy individuals. Complex hybridizations with RNA of normal lymphocytes on one hand and lymphocytes isolated from patients on the other hand can provide information about genes involved in carcinogenesis, particularly in leukemogenesis.

By means of complex hybridization of the unigene set with RNA isolates from patients, detailed genetic profiles of the immune competent cells can be drawn up. With the help of these genetic profiles, the disease can be precisely diagnosed and classified. This in turn allows to define new pathological subclasses, which were so far not detectable with the previous classical clinical

methods. Furthermore, it will be possible to discover new molecular markers of susceptibility of particular individuals to immunologically related diseases.

Hybridizations with RNA derived from lymphocytes of leukemic patients prior and after a specific therapy may contribute to verify or monitor an applied treatment. Detection of cancer-specific genes could also be helpful for detection of residual leukemic cells in treated patients. One can also envisage molecular characterization of drug resistance but also identification of new drug targets for leukemia treatment. Thus the efficacy of the therapy and survival rate of patients can be improved.