HUMBOLDT-UNIVERSITÄT ZU BERLIN

Faculty of Mathematics and Natural Sciences I

Department of Biology

Diploma Thesis

Optimizing cDNA Phage Display for High Throughput Biomarker Discovery

Sunniva Förster

born July 11th, 1984 in Ulm

Conducted at the Max Planck Institute for Molecular Genetics,
Department of Vertebrate Genomics,
Prof. Dr. Hans Lehrach in the Laboratory of Dr. Zoltán Konthur

Berlin, December 2011

Berlin December 2011

Contents

Abbreviations

AA	Amino acid
Ab	Antibody
ABTS	2,2'-azino-bis (3-ethylbenzothiazoline-6-sulphonic acid)
AD	Alzheimer's Disease
Amp	Ampicillin
AP	Alkaline Phosphatase
APP	Amyloid Precursor Protein
ATP	Adenosine triphosphate
AU	Arbitrary Units
bla	β-Lactamase gene
Blastn	Nucleotide Blast
blastp	Protein Blast
bp	Base pair
BSA	Bovine serum albumin
cDNA	copy DNA
CDS	Coding Sequence
cfu	colony forming units
Cm	Chloramphenicole
CSF	Cerebrospinal fluid
Ct	cross threshold
DDT	Dithiothreitol
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside-triphosphate
E.coli	Escherichia coli
EBV	Epstein Barr virus
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
EST	Expressed Sequence Tag
GAPDH	Glycerinaldehyde-3-phosphate dehydrogenase
Glu	Glucose
H	Healthy
h	Hour
HGNC	HUGO Gene Nomenclature Committee
HRP	Horseradish Peroxidase
IL-2	Interleucin-2
IPTG	Isopropyl-beta-D-thiogalactopyranoside
Kann	Kanamycin
kb	Kilo base pair
kDA	Kilo Dalton

L	litre
lac	Lactose
ln	natural logarithm
log	common logarithm
M	Molar
M	Marker
mA	Milli ampere
min	Minute
mL	Milli litre
mM	Milli molar
MM	Master Mix
mRNA	Messenger RNA
MS	Multiple sclerosis
NCBI	National Center for Biotechnology Information
No.	Number
ODx	Optical density (absorbance) in x nm
ORF	Open reading frame
ori	Origin of replication
PAA	Polyacrylamide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline + Tween20
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PI	Protease inhibitor
PPI	Protein Protein Interaction
qPCR	Quantitative PCR
RefSeq	Reference sequence
RNA	Ribonucleic acid
rpm	Rounds per minute

1 Abstracts

1.1 Zusammenfassung in deutscher Sprache

Für viele Krankheiten besteht ein enormer Bedarf für eine frühzeitige und sensitive Diagnostik. Basierend auf Phage Display wurde in dieser Diplomarbeit ein Hochdurchsatzverfahren für die Identifikation von serologischen Biomarkern erfolgreich etabliert. Durch die Kombination eines verbesserten Phagemid Expressionsystems mit Robotertechnologie und 454 Sequenzierung wurde Phage Display als Methode für die Biomarkentwicklung optimiert. Morbus Alzheimer und Multiple Sklerose sind schwerwiegende Erkrankungen des zentralen Nervensystems, die in dieser Studie Modellcharakter haben. Antikörper aus den Blutseren von jeweils fünf Multiple Sklerose Patienten, fünf Alzheimer Patienten und fünf gesunden Probanden dienten dem Auffinden von Peptidantigenen die sich potentiell als Biomarker eignen. Außerdem wurden in diesem Zusammenhang verschiedene *E.coli*-Stämme und Immunglobulinklassen systematisch evaluiert. Der Stamm ER2738 erwies sich hierbei als der geeignetste für die Anwendung in Phage Display. Bei den Immunglobulinen zeigte sich, dass IgA Antikörper ebenso gut funktionieren wie IgG Antikörper und komplementäre Ergebnisse liefern. Es empfiehlt sich daher beide parallel zu verwenden. Um alle Kombinationen der verwendeten Stämme, Seren und Antikörperklassen vergleichen zu können wurden 270 Proben analysiert. Basierend auf jeweils einem Patientenserum für Alzheimer, Multiple Sklerose sowie dem Serum eines gesunden Probanden wurden 18 Proben davon 454-sequenziert und detailliert ausgewertet. Obwohl die Mehrzahl der gefundenen Peptide synthetisch ist, wurden ausreichend humane Proteine gefunden, die einen deutlichen Krankheitsbezug zeigen. Durch die gute Vernetzung dieser Proteine mit weiteren krankheitsrelevanten Proteinen in einem Netzwerk konnten auch neue Anhaltspunkte für die Erforschung der biologischen Krankheitsmechanismen aufgezeigt werden. Die vielversprechenden Ergebnisse dieser Diplomarbeit zeigen, dass sich die optimierte Phage Display Methode hervorragend für die Entwicklung neuer Biomarker eignet. Sie empfiehlt sich daher und aufgrund ihrer hervorragenden Skalierbarkeit zukünftig für großangelegte Studien.

1.2 Abstract

For many diseases, there is an enormous need for more timely and sensitive means of diagnosing. In this study, we successfully established a high through-put method for identifying serological biomarkers based on phage display. By combining an optimized phagemid expression system with robotics

and 454 sequencing, phage display was optimized for biomarker development. Alzheimers disease and Multiple Sclerosis are severe illnesses of the central nervous system and are used as model diseases in this study. Antibodies from blood sera of five Alzheimers disease patients, five Multiple Sclerosis patient and five healthy controls served to find peptide antigens that might be suitable biomarker candidates. In this context, several different *E.coli* strains and different classes of immunoglobulins were systematically evaluated. The *E.coli* strain ER2738 turned out to be the strain most suited for phage display. The analysis of different classes of immunoglobulins showed that IgA immunoglobulins work as well as IgGs while producing complementary results. This means that it is advisable to use both in this kind of study. In total, 270 samples were prepared, to allow comparing all combinations of strains, sera and classes of immunoglobulins. Based on one serum from each disease condition, 18 of these were deep sequenced and analysed in detail. Although most of the peptides founds are synthetic, the peptides resembling human proteins were enough to demonstrate a clear disease relevance. Furthermore, a protein-protein interaction network focused on the hits of this study as well as other disease-relevant proteins reveals new leads for researching the biological mechanisms of the diseases. In conclusion, the the optimized phage display method established in this study is ideally suited for the development of novel biomarkers. This and the fact that it scales excellently make it lend itself for future large-scale studies.

2 Introduction

2.1 Development of serological biomarkers to improve diagnosis for Multiple sclerosis and Alzheimer's disease

Alzheimer's Disease and Multiple sclerosis are disorders of the central nervous system with different cause and pathogenesis. Alzheimer's disease is the prime example for a neurodegenerative disorder and involves the accumulation of misfolded proteins in the brain while multiple sclerosis is primarily an immune-mediated demyelinating disease. At the same time, both of them share some aspects of neurodegeneration and inflammation on the molecular level (reviewed in Lassmann (2011)). Hallmarks in Alzheimer's disease are the accumulation of amyloid plaques and neurofibrillary tangles, accompanied by the formation of brains lesions. However these conditions are not unique and occur also in other neurodegenerative disease like Parkinson and Huntington's disease (reviewed in Muchowski (2002)). Neurodegenerative aspects involving axon pathology play a role in multiple sclerosis as well (reviewed in Trapp & Nave (2008)). A good biomarker should be able to discriminate between Alzheimer's and multiple sclerosis, therefore these diseases are ideal models in a biomarker study.

The number of people affected by Alzheimer's disease is predicted to quadruple until 2050 making it more and more burdening for health care systems (World Alzheimer's report, 2011). Alzheimer's patients suffer from slow decline in memory, thinking and reasoning skills, finally leading to complete dementia. Late onset Alzheimer's disease usually affects patients above 65 years of age (Koch *et al.* (2007)). Early onset Alzheimer's disease begins before 65 and is a comparatively rare disease of clear genetic etiology (Rademakers *et al.* (2003)). A definitive diagnosis of Alzheimer's can only be made post mortem because it requires cutting apart the brain. Diagnosis in living patients relies mainly on clinical criteria that evaluate symptoms, imaging procedures and differential diagnosis to exclude other causes of dementia (reviewed in Lopez *et al.* (2011)). Although there is no cure yet, reliable diagnosis in an early stage of the disease is desirable for patients. It is also likely that diagnosis in a pre-symptomatic state will open new treatment perspectives (Ray *et al.* (2007)). Reversing neurological damage is very difficult, if not impossible therefore prophylactic intervention and attempts to slow progression seem most promising (reviewed in Thal (2006); Gandy (2011)).

The development of precise and cost efficient biomarkers is an intensive field

of research. The most promising biomarker candidates for neurodegenerative diseases have been identified in the cerebrospinal fluid (CSF). The only test validated to support clinical diagnosis in Alzheimer's disease is based on the detection of elevated levels of tau, phosphorylated tau and the 42 amino acid form of ß-Amyloid in CSF (reviewed in Blennow & Hampel (2003)).

Definite diagnosis of multiple sclerosis is complex, time consuming and expensive as well. Onset in multiple sclerosis is usually between 20 and 40 years (Qiu *et al.* (2009)). Multiple sclerosis occurs in relapsing episodes and begins with vague symptoms like double vision, facial palsy or mild cognitive impairment. The disease progression continues with relapses and is unpredictable, sometimes leading to complete paralysis in severe cases. The symptoms can be ascribed to the destruction of myelin caused by T-cell attacks and interrelated chronic inflammation (reviewed in Goverman (2011)). Treatments with immunomodulators and corticosteroids to slow down disease progression are most efficient when administered in an early state of disease. A paradigm in the diagnosis of MS is detection of disease dissemination in space and time and the exclusion of other diseases with similar neurological manifestations (Charil *et al.* (2006)). The heterogeneous disease progression and unpredictable incidence of relapses makes it difficult to find unifying disease criteria. Diagnostic criteria are therefore constantly modified and substantial improvements in MRT-imaging techniques have contributed to increasing reliability of MS diagnosis (Trojano (2011)). Testing the CSF is often made to support multiple sclerosis suspicion and oligoclonal IgG bands can be detected in approximately 75% of the patients. However, biomarkers able to define heterogeneity of multiple sclerosis and classify patients into subgroups would help to render more precise diagnoses (Bielekova & Martin (2004)).

A leaky blood brain barrier has been related to multiple sclerosis and Alzheimer's disease, therefore disease specific proteins are likely to be present not only in the CSF but also in blood (reviewed in Veerhuis *et al.* (2011)). Most patients will agree to provide blood samples and contrary to lumbar puncture a blood sample can be obtained non-invasively (Takeda *et al.* (2010)). This is of advantage especially for elderly patients who might suffer from Alzheimer's disease. In addition, a simpler diagnosis may also result in an early diagnosis, because CSF is typically obtained only after observing symptoms. Phage display is a very promising method for the discovery of serological biomarkers and is therefore used in this work in an attempt to improve diagnosis of AD and MS.

2.2 Phage Display is a method for discovering novel biomarkers

In this work the human antibody repertoire is used as a pool for the identification of peptide antigens that are suitable as serological biomarkers. Sera from multiple sclerosis and Alzheimer's patients were screened for disease specific antibodies with a phage display-based approach in an attempt to isolate seroreactive peptides. Autoantibodies from patient serum have proven to be useful in the development of diagnostic biomarkers for autoimmune diseases like systemic lupus erythematosus (Migliorini *et al.* (2005)). They also have been used to define and classify autoimmune diseases and serve as markers for autoimmune disease activity and progression (reviewed in Leslie *et al.* (2001)). Although Alzheimer's is not an autoimmune disorder, screening the blood of patients at risk of Alzheimer's might yield antibodies to proteins that are different between affected and unaffected people. Multiple studies provide evidence that autoantibodies can be found in the blood of Alzheimer's patients (Nagele *et al.* (2011); Colasanti *et al.* (2010)). Although the underlying mechanisms are not yet fully understood, a leaky blood brain barrier is thought to be involved (Ryu & McLarnon (2009)).

In this work a cDNA library derived from human brain mRNA is used as a pool of antigens to be screened. Brain tissue has a broad expression spectrum and in addition to ubiquitously expressed proteins, the tissue contains brain-specific proteins important for the investigation of neurodegenerative diseases (Ramsköld *et al.* (2009)). The expressed antigens from the brain cDNA library are fused to bacteriophage capsid proteins in phage display. To find disease specific interaction partners from patient blood that are suitable as biomarkers, an affinity selection technique called biopanning is used. The pool of expressed cDNA interacts with target antibodies from patient blood serum. Best binders are then isolated by undergoing several rounds of biopanning. In this process repetitive washing and binding to the target gradually increases affinity. To reveal the identity of enriched binders the fact that phenotype and genotype are coupled in bacteriophages is exploited. This allows easy propagation of enriched binders as well as the sequencing of their genes.

Phage display with cDNA libraries has been criticized for containing many short and unnatural peptides suggesting instead the use of ORF enriched libraries (Caberoy *et al.* (2010)). However, it is reasonable to assume that antibodies are formed against healthy condition proteins but biomolecules chemically modified in a disease specific way. Fully synthetic phage display libraries are one way to implement this finding. A recent study revealed two candidate biomarkers with high sensitivity and specificity for predicting Alzheimer's disease based on the use of a fully synthetic peptoid library (Reddy

et al. (2011)). The size of cDNA libraries can theoretically approximate the human genome. Short peptides and out of frame constructs together with fully expressed natural proteins combine the advantage of having high diversities and the chance to include disease specifically folded epitopes.

Synthetic antigens often mimic binding properties of native antigens although they have a different primary structure (reviewed in Papini (2009)). Commonly the affinity of these mimotopes to their target exceeds that of native antibodies. Prominent examples are the anti-citrullinated peptide antibodies used as biomarkers for rheumatoid arthritis (Anzilotti *et al.* (2006)). Synthetic cyclic peptides rendering the citrulline moiety more available to serum antibodies were used to develop the widely used anti-CCP tests for diagnosis of rheumatoid arthritis (reviewed in Taylor *et al.* (2011)).

Taken together, the combination of natural, semi-natural and artificial epitopes makes phage display of cDNA fragments a very promising approach for finding novel candidate biomarkers. Phage display of peptide libraries has been already used for biomarker discovery in cancer research. Autoantibody signatures that combine a couple of reactive short peptide sequences revealed biomarker candidates with high specificities and sensitivities (Wu *et al.* (2010)).

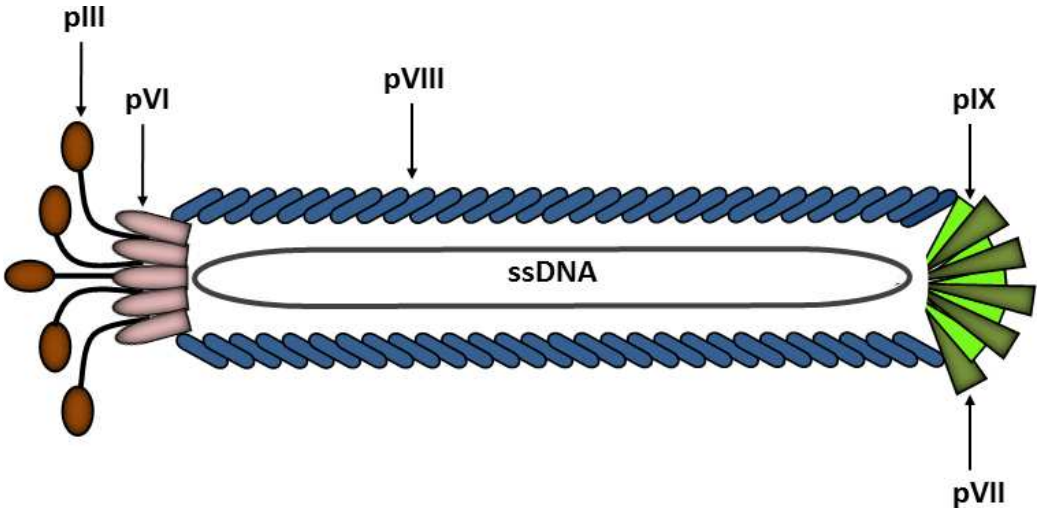

Figure 1: Schematic representation of bacteriophage M13.
The rod shaped bacteriophage is approximately 5 nm in diameter and 1 µm long. The structural proteins pIII, pVI, pVIII, pVII and pIX make up the phage coat. The major coat protein pVIII is abundant on the phage surface; 2700-3000 copies cover the entire length of the phage DNA. Five copies of pVII and pIX are required for packaging of phages. The five copies of pIII and its accessory protein pVI are needed for host infection. (Image adapted from Georgieva & Konthur (2011)).

The by far most widespread bacteriophage used for phage display is M13. This filamentous phage infects *E.coli* cells and exploits the host cell metabolism for progeny. M13 is a non-lytic phage that allows the host to survive albeit with a slower growth rate. This is a big advantage compared to the use of lytic phages because they can be easily obtained and purified from the supernatant of the growth medium with a simple PEG precipitation step. The lytic lifestyle of other phages like T7, T4 or lambda that are also used in phage display would lead to the contamination of the growth medium with proteases and other undesirable components that are difficult to remove and likely to interfere with downstream applications (Danner & Belasco (2001)). With minimal constraint on DNA insert size and high viral titers (10^{11}-10^{13} particles per ml) M13 is ideally suited for phage display. However the use of *E.coli* as host strain renders posttranslational modifications to proteins impossible. This disadvantage is outweighed by the easy handling of *E.coli* laboratory strains. For infection with bacteriophages the F-pilus needs to be presented. Bacteriophages attach to this hair-like appendage that retracts upon attachment. The phage capsid depolymerizes into the *E.coli* membrane and as a direct consequence the single stranded phage DNA is released into the cytoplasm where the minus strand is synthesized and the plasmid is replicated by means of the rolling-circle mechanism. Upon transcription and translation of the phage genes the phage proteins self-assemble and protrude from the host cell.

Many successful phage display systems are based on the use of phagemids. These vectors have a plasmid origin of replication (ORI), a filamentous phage ORI, a phage packaging signal and an antibiotic resistance as selection marker. They serve as cloning vectors that are replicated in high copy numbers in *E.coli* cells. They typically have an antibiotic resistance as selection marker and contain the desired insert under the control of a regulated promoter. The use of phagemids has great advantages because they can be easily transformed and propagated in *E.coli* cells while yielding high amounts of DNA (Breitling *et al.* (1991); Barbas *et al.* (1991))

To induce plasmid translation of phage proteins and assembly of phage particles a super-infection with helper phage is needed. A widely used helper phage is M13K07. This phage is a M13 derivative that encodes for a complete set of phage proteins but has a mutation in the origin of replication (ORI). The infection of a phagemid bearing *E.coli* cell with helper phage provides the necessary proteins to induce phage assembly. Both, helper phage DNA and the phagemid are packaged into virions. The mutated ORI of M13K07 leads to a far less efficient packaging of its DNA so that most phage particles bear the insert containing phagemids.

To select expressed inserts for a specific target biopanning is used in

Phage Display. In this work biopanning is used to enrich phages presenting antigens that interact specifically with patient antibodies. The basic principle of biopanning is to wash away unspecific binders and to keep binders with high affinity to the target. Phages that bind to the target are then propagated in *E.coli* cells. In subsequent rounds, the washing stringency is increased. This procedure, termed panning round, can be repeated any number of times. In every additional panning round the diversity of the phage pool decreases while the specificity of binders increases (reviewed in Konthur & Crameri (2003)). The number of panning rounds depends on the experimental setup and the research interest.

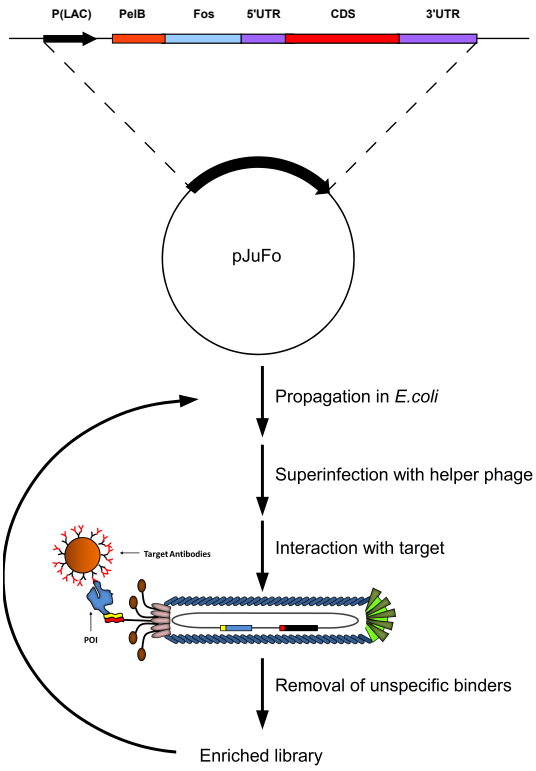

Figure 2: Schematic overview on essential biopanning steps.
The pJuFo cDNA library is transformed into *E.coli* cells. Upon superinfection with helper phage the bacteriophages presenting the recombinant proteins of interest (POI) on their surface are assembled. As target antibodies from patient sera, coupled to magnetic beads, are used. After binding to the target the less specific binders are washed away. This panning round leads to a library enriched for binders and can be repeated to achieve increasing specificity of binders with target affinity.

In this work patient antibodies were used to select disease relevant peptides

and proteins. The antibodies were immobilized on magnetic beads and served as solid phase. In comparison to other solid phases commonly used in phage display biopanning, magnetic beads have the highest binding capacity and are ideally suited for automated phage display. Magnetic pins compatible with 96-well plates are used to transfer the beads to the buffers and solutions required in the biopanning protocol. The magnet continually moves the beads through the solution and thereby improves binding kinetics. The simultaneous handling of ninety-six samples and the use of a robot increases the reproducibility and the throughput capacity of the experiment.

2.3 Phage Display of cDNA fragments with the pJuFo phagemid system

George P. Smith first showed a displayed fusion protein on filamentous bacteriophages to be functional in 1985 (Smith (1985)). In a proof-of-principle experiment, the endonuclease EcoRI was displayed on filamentous phage f1. *E.coli* cells were transformed with recombinant DNA and began releasing phage particles presenting EcoRI on their surface. It was important to use a gene with a well-known sequence because any stop codon would interrupt gIII translation and render phage particles non-infective. Smith suggested using his method to isolate desired clones from a library of random inserts by using an antibody as bait. Phage display gained popularity rapidly, but instead of isolating the desired phage-displayed antigens with an antibody the method was used instead for engineering antibodies (McCafferty *et al.* (1990); Breitling *et al.* (1991)). The immunoglobulin variable domains were displayed on M13 bacteriophages. Antibody fragments that bind to a desired antigen were selected. These were the first recombinant antibodies that were not derived from immunization of laboratory animals. A great advantage of this technique is that antibodies can be produced for therapeutic use, because immune reaction in patients is lower than with animal antibodies. However, displaying antigens, as suggested by Smith in 1985, for example from cDNA libraries, proved difficult. Frequent stop codons in the 3' region of cDNA render a fusion to the amino terminus of pIII unusable. An elegant way of making the pIII-fusion approach available for screening cDNA libraries was the invention of the pJuFo system (Crameri & Suter (1993)). In this system the protein of interest is not fused to pIII directly but is fused to a leucin zipper-forming gene. The gIII gene is fused to the corresponding other half of the leucin zipper. The two proteins whose leucin zipper-forming regions were used for the pJuFo system were the proto-oncogenes Jun and Fos. The Jun leucin zipper domain is fused to the gIII gene and Fos is fused to the

N-terminus of the cDNA fragment. Both constructs are expressed separately and bind to each other during phage assembly in the periplasm. Although the pJuFo phagemid system makes the fusion of unknown proteins from a cDNA library to pIII feasible in principal, it is not guaranteed that the correct, protein coding, strand is expressed for each cDNA.

Prior to this work a cDNA library made with Gateway technology was shuttled from pDONR222 into three different pJuFo phagemid vectors. These vectors termed pJuFo A, pJuFo B and pJuFo C differ in one nucleotide in front of the insert cloning site. The pJuFo B vector expresses the start of the insert in frame with the Fos gene and pJuFo A and pJuFo C have one nucleotide less or more, respectively. Shuttling and transforming a number of inserts at least an order of magnitude higher than the complexity of the entry library guarantees that most inserts are present in all three pJuFo vectors. This means that all three frames are expressed as Fos fusion proteins. This novel approach addresses the problem of out-of-frame cDNA inserts and potentially increases the number of expressed peptides by a factor three.

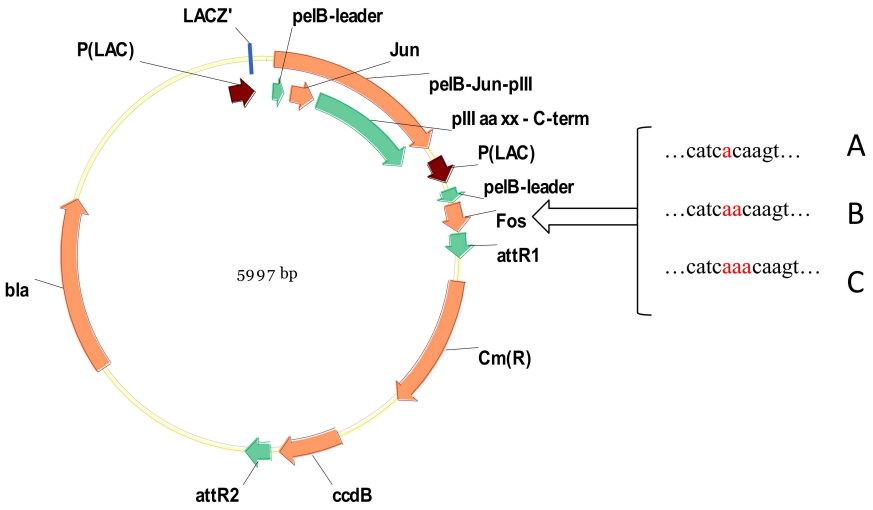

Figure 3: The three frame variants of the pJuFo phagemid.
The three variants, A, B and C of pJuFo differ in the additional nucleotide at the junction to Fos, they are marked in red. Expression of the Fos-fusion protein is driven by a Lac promoter. The Jun-pIII fusion and Fos are under the control of a pelB-leader sequence for periplasmatic targeting. As selection marker a Ampicillin resistance cassette is used (bla).

Which proteins are expressed and displayed on the phage surface depends on the cDNA inserts. The cDNA library used in this work was constructed with the Clone Miner I Kit (Invitrogen). The reverse transcriptase used for

cDNA synthesis creates DNA fragments that typically do not correspond to complete mRNAs. Frequent causes of short fragments are partially degraded template mRNA, short mRNAs or problems inherent in the PCR reaction. PolyA priming starts at the 3'UTR of the gene therefore all cDNA fragments will contain 3'UTR, only a fraction CDS and even less constructs will contain the whole gene including 5'UTR. The cDNA inserts were cloned into the pJuFo vectors directionally.

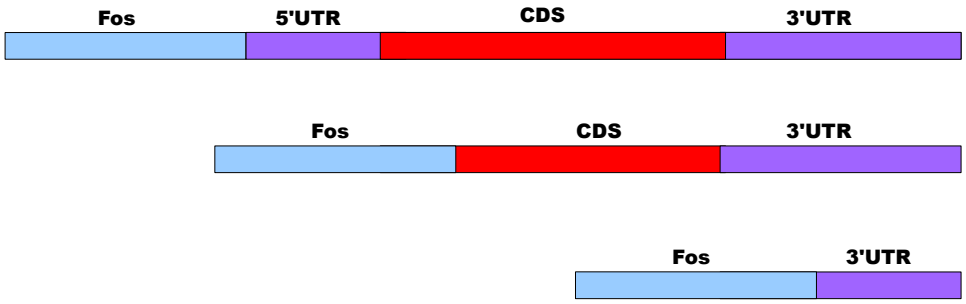

Figure 4: The Fos fusion with cDNA can result in the expression of natural and synthetic peptides.
Hypothetical peptides expressed and presented on the bacteriophages are drawn. Fos is painted in blue, the untranslated regions (UTR) in purple and the coding sequence (CDS) in red. In the upper panel a schematic representation of the Fos fusion with a cDNA encoding for a full length mRNA is shown. Expression of this construct leads to naturally occurring human proteins. Abortion products during cDNA synthesis can also lead to the cloning of fragments with partially present CDS (middle panel), expressed as peptide fragments. When 3'UTR (lower panel) is expressed, the resulting peptides are completely artificial.

The pJuFo phage display system is the most widespread system for Phage Display of cDNA libraries and was successfully used for a multitude of applications (reviewed in Georgieva & Konthur (2011)). It has been used to screen for allergy-causing proteins from many common allergens (reviewed in

Rhyner *et al.* (2004)). It has also been used to find autoantigens in autoimmune disorders (Kemp *et al.* (2002)) and cancer (Fosså *et al.* (2004)). The functionality of enzymes displayed with the pJuFo system was demonstrated in 2002 by (Brunet *et al.* (2002)). They showed a displayed polymerase to be catalytically active, hence functional. Weichel et al. expressed PhoA, an alkaline phosphatase from *E. coli*, and checked the catalytic rate. They could show that the catalytic activity is indistinguishable from free PhoA (Weichel *et al.* (2008)).

2.4 Sequencing of enriched phage display libraries on a Roche FLX-454 Genome sequencer system

To analyze enriched binders from phage display the Roche FLX-454 sequencing platform was used. The 454-technology is a pyrosequencing method based on the principle of sequencing by synthesis. Single stranded template DNA is immobilized and amplified on streptavidin beads and placed on a chip. A DNA polymerase is used to complete the single stranded fragments. The four nucleotides Adenin, Cytosin, Guanin and Thymin are supplied sequentially. During incorporation of a complementary dNTP, pyrophosphate (PPi) is released and is quantitatively turned into ATP by ATP sulfurylase. ATP is used as a substrate for the conversion of luciferin to oxyluciferin, which releases a chemiluminescent signal. Detection of the light signals on each chip position allows calling up to 400 million bases per sequencing run. Compared to other Second Generation Sequencing methods, for example Illumina sequencing, 454 is of advantage in this work. The readlength is much higher which is useful for the determination of the reading frame and allows to discriminate between the three pJuFo variants used in this work. Because the three vectors differ only in one nucleotide only reads spanning this position and the insert can be used.

Compared to Sanger sequencing this approach has several advantages. Second generation sequencing techniques have a much lower cost per base sequenced, allowing sequence coverage that would otherwise be prohibitively expensive. To circumvent the restrictions in terms of minimum number of reads per sequencing run a barcoding system can be used to distinguish reads originating from different samples. This is far less laborious than isolating plasmid from hundreds of single clones like those needed for Sanger sequencing. A rate limiting step in Phage Display experiments is the analysis of enriched binders (Dias-Neto *et al.* (2009)). The capacity to analyse samples from high throughput phage display experiments thereby increases to a point where monitoring the enrichment process in great detail becomes possible. For highly

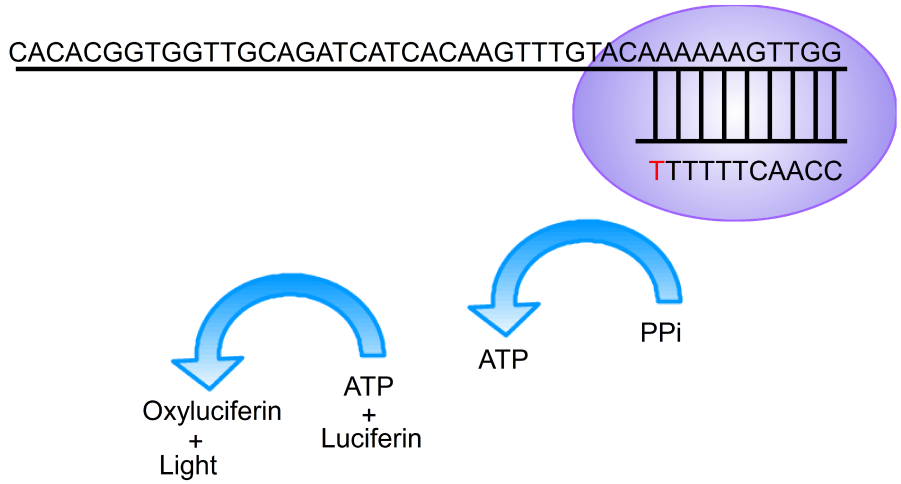

Figure 5: Schematic representation of the sequencing process used in 454 sequencing.

The method is based on the principle of sequencing-by-synthesis. The template DNA strand (top) is bound by a sequence-specific primer oligonucleotide that gets elongated gradually. Upon incorporation of a dNTP into the nascent reverse DNA strand, pyrophosphate is released. Thus pyrophosphate is turned into ATP by ATP sulfurylase. The ATP in turn is utilized by firefly luciferase to activate Luciferin, which releases a chemiluminescent pulse. Since at each point in time, there is only one kind of dNTP supplied, the DNA sequence can be calculated by the sequence and intensity of the light pulses.

enriched libraries obtained after several panning rounds, deep sequencing allows the identification of low abundant clones that would otherwise go unnoticed. Thus a more complete picture of the binders is obtained, revealing more potential biomarkers.

2.5 Objectives

Alzheimer's disease and multiple sclerosis are both severe disorders that are diagnosed with unsatisfactory preciseness to date. Alzheimer's disease can only be confidently confirmed post mortem and also for multiple sclerosis satisfactory biomarkers do not exist. There is an urgent need for biomarkers that are not only able to render definite diagnosis, possibly in a preclinical state but are also capable of monitoring disease progression and the impact of medical treatment. Extensive research for identifying biomarkers did not result in marketable products that contribute substantially to improved diagnostic criteria for these diseases. Therefore it is important to develop high-throughput methods for the systematic discovery of novel biomarker

candidates. In this study the focus of research was to optimize phage display as a method for the identification of peptides specifically reacting with antibodies from blood serum of diseased patients. The novel approach in this work, combining cDNA Phage display with robotic technology, Next Generation Sequencing and an improved phagemid system makes Phage Display available for large scale biomarker studies.

3 Materials and Methods

3.1 Materials

3.1.1 E. coli strains

Strain name	Genotype	Source
XL-1 Blue	F- mcrAΔ (mrr-hsd RMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 endA1 recA1 deoR Δ(ara, leu)7697 araD139 galU galK nupG rpsL λ-	Invitrogen, Carlsbad, USA
TG1	F- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(rB-, mB-) ara14 galK2 lacY1 proA2 rpsL20(Smr) xyl5 Δleu mtl1	Invitrogen, Carlsbad, USA
ER2738	supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5(rK– mK–) [F' traD36 proAB lacIqZΔM15]	Stratagene, La Jolla, USA

3.1.2 Bacteriophages

Bacteriophage	Source
Helperphage M13K07	New England Biolabs (NEB), Ipswich, UK
Hyperphage M13K07ΔpIII	Progen Biotechnik, Heidelberg, GER

3.1.3 Plasmids

Plasmid	Creator/Provider
pJuFo A	AG Konthur
pJuFo B	AG Konthur
pJuFo C	AG Konthur
pUC18	Invitrogen, Carlsbad, USA

3.1.4 Primers

Primers were provided by Eurofins MWG Operon, Ebersberg and BioTeZ, Berlin.

Primer name	Sequence, orientation 5'-3'	Annealing temperature (°C)
FOS_seq	GCTCTGCGGTGGTTTGACCG	66
M13 UP (-21)	CGACGTTGTAAAACGACGACGGCCAGT	76
XbaI_sfiA_rev	TGATAGGCCGAGGCGGCCCTCGATTGC GGCCGCTTAACTC	65
Fos_sfiA_Fw	ACTCTGGCCATTACGGCCACGGTGGTT GCAGATCATC	60
q_Tub_fw	GCTGTACCGTGGAGATGTG	60
q_Tub_rev	CCACAGTGGGAGGCTGGTAG	59
q_GAPDH_fw	ACCAGGTGGTCTCCTCTGAC	58
q_GAPDH_rev	TGCTGTAGCCAAATTCGTTG	59
q_pJuFo_fw	AGTCGTGTCTTACCGGGTTG	60
q_pJuFo_rev	GGCGCTTTCTCATAGCTCAC	61

3.1.5 Antibodies

Antibody	Species of Origin	Dilution	Provider
HRP/Anti-M13	Monoclonal Mouse	1:5000	GE Healthcare
Anti-pIII	Monoclonal Mouse	1:1000	MoBiTech
Anti-human IgG	Polyclonal Rabbit	2:3	Thermo Scientific
Anti-human IgA	Polyclonal Goat	2:3	Thermo Scientific
Anti-mouse IgG	Polyclonal Goat	2:3	Thermo Scientific
HRP/Anti-rabbit IgG	Monoclonal donkey	1:3000	Amersham
HRP/Anti-mouse IgG	Polyclonal goat	1:500	Millipore
Anti-Ubiquitin	Monoclonal Mouse	10:1	Calbiochem
Anti-GAPDH	Monoclonal Mouse	10:1	Research Diagnostics Inc.
Anti-Tubulin	Monoclonal Mouse	10:1	Biocarta

3.1.6 Alzheimer's Patient Sera

Lot#	age	gender	weight [kg]	height [cm]	date of diagnosis	sort of test	date of test	date MRT	date CT	medication	genetic predisposition for neurogenerative disease	date of donation	material
191491	72	m	85	180	26.1.2006	CERAD	26.1.2006		3.1.2006	Aricept 10 mg	no	10.9.2009	serum
191492	79	f	65	165	8.1.2008	CERAD	30.5.08+ 21.7.09	16.7.2008	15.5.2008	Aricept 10 mg	no	10.9.2009	serum
191493	76	m	64	168	22.1.2008	CERAD	22.1.2008	16.3.2007	3.8.2007	Aricept 10 mg	no	10.9.2009	serum
191494	80	m	80	176	19.5.2009	CERAD	22.4.2009		8.5.2009	Aricept 10 mg	no	10.9.2009	serum
191724	72	m	80	165	1.11.2003	MMST, clocktest, DEMTEC	8.10.2009	28.1.2004		Aricept 10 mg	no	8.10.2009	serum

3.1.7 Multiple Sclerosis Patient Sera

Lot#	gender	age	weight	height [m]	date of diagnosis	subtype	medication	MRT/result	lumbar puncture/IEF	last exerbation	evoked potentials
170171	f	38	50	1.63	1998	RR-MS	Copaxone	positive	positive	1998	
170172	f	39	67	1.64	1992	RR-MS	Copaxone	positive	positive	2002	pathologic
170175	f	47	85	1.7	1986	RR-MS	Betaferon	positive		2007	
170176	f	44	90	1.6	2000	RR-MS	Copaxone	positive	repeated Liquor negative	2005	
170178	f	42			2005	RR-MS	Rebif 44	positive	positive	2007	normal

3.1.8 Healthy Proband Sera

Lot#	age	gender	sort of test	date of test	date of donation	material
151002	75	m	CERAD	17.10.05	17.10.05	serum
151006	77	f	CERAD	17.10.05	17.10.05	serum
151010	75	f	CERAD	17.10.05	17.10.05	serum
151014	76	f	CERAD	18.10.05	18.10.05	serum
151015	73	f	CERAD	18.10.05	18.10.05	serum

3.1.9 Kit-systems

Kit-system	Application	Provider
QIAprep Spin Miniprep Kit (250)	Plasmid Purification out of bacteria culture	QIAGEN, Hilden, GER
QIAquick PCR Purification Kit (250)	DNA Purification after PCR	QIAGEN, Hilden, GER
PCR-Reinigungs-Kit	DNA purification after PCR	Seqlab, Göttingen, GER
Gene MATRIX PCR/DNA Clean-UP Purification Kit	DNA Purification after emulsion PCR	Roboklon, Berlin, GER
CN/DAB	Western Blot detection system	Pierce, USA
HSII Phusion Polymerase PCR Kit	PCR	Finnzymes Reagents, FIN
SYBR Green Master Mix (5 ml)	qPCR	Applied Biosystems

3.1.10 culture media

Liquid 2xYT-Broth (per 1 l)
16 g Bacto Tryptone
10 g Bacto Yeast Extract
5 g NaCl
pH 7.2 with NaOH
add ddH_2O up to 1 l
autoclave

Liquid TB-Broth (per 1 l)
12 g Bacto Tryptone
24 g Bacto Yeast Extract
4 ml Glycerol
add ddH_2O up to 900 ml
autoclave
adjust to 900 ml with a filter sterilized solution of 0.17 M KH_2PO_4 and 0.72 M K_2HPO_4

Liquid SOB (per 1 l)
20 g Bacto Tryptone
5 g Bacto Yeast Extract
0.5 g NaCl
add ddH_2O up to 1 l
autoclave
10 ml filter-sterilized 1 M $MgCl_2$
10 ml filter-sterilized 1 M $MgSO_4$

Liquid SOC (per 100 ml)
1 ml filter-sterilized 2 M glucose
Add SOB medium (autoclaved) up to 100 ml

Phage Top Agar (per 1 l)
10 g bacto agar
5 g NaCl
10 g Bacto Tryptone
autoclave

3.1.11 Culture media additives

Antbiotics	Provider
Ampicillin	Carl Roth, Karlsruhe, GER
stock solution: 50 mg/ml in 50 % EtOH	
working concentration: 100 µg/ml	
Kanamycin	Roche, Basel, Switzerland
Stock solution: 30 mg/ml	
Working concentration: 60 µg/ml	
Glucose	
stock solution: 40 %	
working concentration: 2 %	Alfa Aesar, Ward Hill, USA
10x HMFM (4x solution 1 : 1x solution 2)	Solution 1:
	3.65 mM $MgSO_4$ x $7H_2O$
	36 mM $(NH_4)_2SO_4$
	5 mM Tri-Natriumcitrat x $2H_2O$
	44 % Glycerin
	Solution 2:
	132 mM KH_2PO_4
	270 mM KH_2PO_4

3.1.12 Buffers and solutions

Unless otherwise noted, all used chemicals have been bought at the companies Sigma-Aldrich, Merck, Roche and Roth.

Buffer and solution	Components and/or concentrations
PBS (pH 7.4)	0.0027 M KCl
	0.137 M NaCl
	0.01 M phosphate buffer
PBS-T	PBS + 0.2 % Tween-20
PTM	PBS-T + 3 % non-fat dry milk powder
ABTS substrate solution (pH 4.3)	10 mg
	2.2'-azino-bis(3-ethylbenzthiazoline-6-sulphonic acid)
	10 mL 50 mM Trisodiumcitrate
	10 mL 50 mM Citric acid
	10 µl 30% hydrogen peroxide
Agarose	5-10 g/L (in 1x TAE or 0.5x TBE)
TAE (10x; pH 8.18-8.29)	48.4 g/l Tris-base
	10.9 g/l Glacial Acetic Acid
	2.92 g/l EDTA
TBE (5x; pH 8.13-8.23)	54.0 g/l Tris-base
	27.5 g/l Boric acid
	2.92 g/l EDTA
10x HMFM (4x solution 1 : 1x solution 2)	Solution 1:
(Hogness-Modified-Freezing-Medium)	3.65 mM $MgSO_4$ x $7H_2O$
	36 mM $(NH_4)_2SO_4$
	5 mM Tri-Natriumcitrat x $2H_2O$
	44 % Glycerin
	Solution 2:
	132 mM KH_2PO_4
	270 mM KH_2PO_4

3.1.13 Chemicals

Chemicals	Concentration
Glycerol	stock solution: 87 %
	working concentration: 30 %
Ethanol	stock solution: 100 %
	working concentration: 100 % or 70 %
Ethidiumbromide	working concentration: 0.00005 % in TAE or TBE
NH4OAc	stock solution: 7.5 M
Dry ice	
Nitrogen	
Isobutanol	
ABIL WE 09	working concentration: 7 %
Mineral oil	working concentration: 20 %
Tegosoft DEC	working concentration: 73 %
Polyethylene glycol 6000	working concentration: 20 %
NaCl	working concentration: 2.5 M
Urea	working concentration: 8 M
HCl	working concentration: 0.1 M (adjusted with glycin)
TRISbase	working concentration: 2 M

3.1.14 Other compounds and solutions

Product	Provider
6x Loading dye (R6011)	Fermentas, St. Leon-Rot
GeneRuler 100bp	Fermentas, St. Leon-Rot
Gene Ruler 1kb	Fermentas, St. Leon-Rot
Precision Plus Protein Ladder	Biorad
MyOne Dynabeads	Invitrogen, Oslo Norway
dNTP	NEB, Ipswich, UK
BSA	NEB, Ipswich, UK
Glyco Blue	Applied Biosystems, Carlsbad, USA
Roche Complete	Roche, Diagnostics, USA

3.1.15 Laboratory equipment

Product	Provider
Centrifuge 5424	Eppendorf
Centrifuge 5810 R	Eppendorf
Rotilabo®-centrifuge with butterfly-rotor	Roth
Peltier-Kühlbrutschrank IPP500	Memmert
Thermo Shaker PST-60HL-4	Biosan
Thermo Shaker PST-60HL Plus	Biosan
4430 Incubator Shaker	Innova
Thermomixer comfort, 1,5 ml and 2 ml	Eppendorf
Rocky Typ RT-1S	CTF Labortechnik
NANOpure DIamond Reinstwassersystem	Werner
Bi-Destilliergerät	
Destamat Bi 18 E	Heraeus
Autoklav Typ 23	Melag
Autoclave FNR 4336E	Tecnomara
Gel Doc 2000 Gel Documentation System	BioRad
UVL-21 Black - Ray Lamp	Ultraviolet Products
IKA MS1 Minishaker Vortexer	IKA
Self-Contained Ice Flaker 120Kg - AF20	Scotsman
PTC-200 Thermo Cycler	MJ Research
BioPhotometer	Eppendorf
Microflow Laminar Flow Workstation	Kendro Laboratory Products
MicroPulserTM Electroporator	BioRad
AV812 Adventurer Pro Precision Balance	Scale Solutions
Magnetic Stirrer Hot Plate Stuart SM3	Stuart Scientific
MR 3001 K	Heidolph
pH-/Redox-/Temperatur – Messgerät	Greisinger electronic
GMH 3510	
Electrophoresis Power Supply: EPS 200,	Pharmacia Biotech
EPS 301	
Agarose gel electrophoresis chamber	MPI for Molecular Genetics
(small, large)	
SUB Universal water bath 14	Grant
Spectramax 200	Molecular Devices
Micromat	AEG
Privileg 8020	Privileg
Pipetboy acu	Integra Biosciences
96 Pin Replicator Microplate Replicator	Boekel

3.1.16 Laboratory Consumables

Material	Provider
Reaction tubes for microlitre systems (0.5 ml; 1.5 ml; 2.0 ml)	Sarstedt
PCR 8er-SoftStripes 0.2 ml farblos	Biozyme Scientific
PLASTIBRAND® pipette Tipps (0.1-2 µl, 2-200 µl, 1000 µl)	Sigma-Aldrich
Euroset (1.25 ml)	Eppendorf
SafeSeal-Tips, Premium Line (0.1-2 µl, 2-200 µl, 1000 µl)	Biozyme Scientific
disposable pipettes, sterile (5 ml, 10 ml, 25 ml)	Corning
Gene Pulser/Micropulser Electroporation Cuvettes (1mm)	BioRad
Corning® PP Centrifuge Tubes (15 ml, 50 ml)	
Corning	
Centrifuge tubes, PP, conical bottom, natural (50 ml, 15 ml),	Greiner Bio One
Culture tube, PP, round bottom (5 ml, 14 ml)	Greiner Bio One
Parafilm® M	Brand
Petri dishes	Greiner Bio One
Cryo.s™ Vial (2 ml)	Greiner Bio One
Rotilabo® Einmal-Küvetten, PS Halbmikro, 1.5 ml	Roth
UVette®, Individually Packaged Single, 220-1600 mm	Eppendorf
Lazy-L Spreader™	Sigma-Aldrich
Inoculating loops	Nunc
Einmal-Skalpelle Cutfix®	B Braun
Erlenmeyer Flask, 1 l, 0.2 µm vent cap	Corning
96 well conical Bottom, PP, natural, 0.5 ml	Nunc
96-well PVC (flexible) flat-bottom plate	Becton Dickinson
NuPAGE 4-12% Bis-Tris Gel	Invitrogen
Nunc-Immuno™ Platten	Nunc, Thermo Scientific
Nitrile Examination Gloves	Blossom

3.1.17 Software

Software	Application	Provider
Primer3Plus	Primer Design	Open source
Perl	454-sequence processing	CPAN.org
R	ELISA data processing	CRAN.org
SDS 2.4	qPCR Sequence Detection	Applied Biosystems
Photoshop CS4	Image processing	Adobe
Microsoft Office	Text editing	Microsoft
Adobe Acrobat Professional 8	PDF generation	Adobe
SoftMax Pro 4.8	Absorption measurement	MindVision Software
Bioperl	454-sequence translation	Free Software Foundation
Apcluster	Affinity propagation clustering	CRAN.org
Venn package	Venn diagram	CRAN.org
MSOffice	Excel and Access	Microsoft Office
Lyx	Text formatting	Open source
BLAST	Alignment of 454-reads with a RefSeq database	NCBI
Textpad 5.4.2	Text Editor for programming	Helios Software solutions
Cytoscape 2.8.1	Protein-Protein network construction	Open source
BindIt 3.1	Kingfisher remote control	Thermo Scientific
MIRA3	Contig assembly	Bastien Chevreux

3.2 Methods

3.2.1 Analytic PCR Protocol

PCR was used to amplify DNA with specific Primer pairs. Template DNA was placed in a PCR reaction tube first and a master mix was prepared as indicated in Table1. All additives were kept on ice and the polymerase was added last.

The tubes were mixed and spun down for 5 s before being put in a thermal cycler. Hot start and heated lid were used to reduce non-specific amplification. The PCR program was run as shown in Table 1.

Table 1: Standard Master Mix for PCR reactions

PCR additive	Volume/20µl	Final concentration
5x HF buffer	4 µl	1x
dNTPs(2.5 mM)	1.6 µl	200 µM
Phusion HS II (2 U/µl)	0.2 µl	0.02 U/µl
DMSO	0.2 µl	1%
Forward primer (10µM)	2 µl	0.5 µM
Reverse primer (10 µM)	2 µl	0.5 µM
bidest. water	10 µl	

Table 2: Standard PCR Cycler Program

Reaction step	Time	Temperature
1. Denaturing	45 s	98 °C
2. Denaturing	10 s	98 °C
3. Annealing	20 s	Tm-3°C
4. Elongation	30 s/kb	72 °C
Repeat step 2-4 for 30 cycles		
6. Final Elongation	10 min	72 °C
7. Final hold	∞	4°C

3.2.2 Screening library insert size with colony PCR

Colony PCR is a method to screen bacterial colonies without isolating plasmid DNA. To check the insert size of the pJuFo cDNA library a PCR on single clones was performed. 10 µl water were pipetted into a PCR reaction tube. The clones were placed directly in the water. Afterwards the colonies were heated to 98°C for 5 min before the Master Mix was added.

3.2.3 Annealing Temperature Optimization with Gradient PCR

Gradient PCR is used to find the optimal annealing temperature for a primer set. As the annealing temperature is critical for the success of the emulsion PCR prior to library sequencing, several primer pairs were tested in a gradient PCR. The gradient was set from 58°C-68°C. PCR program and master mix were set equal to the standard PCR protocol (compare Table 3.2.1).

3.2.4 Emulsion PCR

For unbiased amplification from library plasmid, an emulsion PCR approach was used. The hydrophilic PCR master mix and the hydrophobic oil mix

were prepared separately on ice (Table 3). To establish a stable water-in-oil emulsion 100 µl master mix and 600 µl oil mix were vortexed for 5 minutes in the cold room at maximum speed. The emulsion was split into twelve 50 µl PCR reaction tubes that were put immediately in the PCR cycler. The same PCR program was uses as for standard PCR (Table 2).

Table 3: Master Mix for emulsion PCR reactions

PCR additive	Volume per 700 µl	Final concentration
PCR master mix (100 µl)		
5xHF Buffer	20 µl	1x
dNTPs (2.5 mM)	8 µl	200 µM
BSA (10 mg/ml)	2.5 µl	0.25 µg/µl
Phusion HS II (2 U/µl)	0.5 µl	0.02 U/µl
Forward primer (10 mM)	1 µl	0.5 µM
Reverse primer (10 mM)	1µl	0.5 µM
Oil mix (600 µl)		
ABIL WE09	42 µl	7 %
Mineral oil	120 µl	20 %
Tegesoft DEC	438 µl	73 %

3.2.5 Emulsion PCR product clean-up

For purification of emulsion PCR products the components of a Gene MATRIX PCR/DNA Clean-Up Purification Kit (Roboklon) were used with an adjusted protocol. The spin-columns were activated with 40 µl Buffer DX. To break the emulsion, 600 µl PCR were mixed with 1 ml Isobutanol. The solution was vortexed until it turned transparent. 400 µl of Buffer orange-DX was added and mixed vigorously. The sample was centrifuged at 12000 rpm for five minutes. This step separated the reaction mixture into an upper, hydrophobic phase and a lower orange colored hydrophilic phase. The upper phase was carefully pipetted off and the water phase was transferred to an activated spin-column. Afterwards the column was centrifuged at 12000 rpm for one minute. The flow-through was discarded, and the column was washed twice. For the first washing 500 µl wash buffer DX1 was added and centrifuged at 12000 rpm for one minute. The column was washed another time with 650 µl wash buffer DX2. To remove traces of wash buffer the centrifugation was repeated once again. The DNA was then eluted with 900 µl elution-DX buffer and stored at -20°C.

3.2.6 Agarose gel electrophoresis

Agarose gel electrophoresis was routinely used for checking the size of DNA fragments. The agarose was dissolved in 0.5x TBE buffer in a microwave. A final concentration of 1.2% agarose and 0.5 µg/ml ethidium bromide was used. The solution was poured into a tray and used when solidified. The samples were mixed with 6x loading buffer prior to being loaded in the pockets of the Gel. As electrode buffer 0.5x TBE was used. For electrophoretic separation 8 V/cm were applied.

3.2.7 Plasmid Preparation

For plasmid isolation a QIA prep Spin Miniprep Kit was used. Bacterial cultures were incubated over-night at 37°C, 180 rpm. The cells were harvested by centrifugation for 30 min at 2000 rpm. The supernatant was discarded and the remaining pellet resuspended in 250 µl buffer P1. Next 250 µl lysis buffer P2 was added and mixed by inverting the tube five times. After three minutes incubation time the lysis reaction was stopped with 350 µl buffer N3. The sample was centrifuged for 30 min at 12000 rpm. The supernatant was transferred to a spin-column and centrifuged for 1 min. The flow-through was discarded and the column was washed with 750 µl buffer PE. The column was centrifuged for one minute at 12000 rpm, the flow through discarded and centrifuged once again to remove traces of wash buffer. For elution 50 µl buffer EB was pipetted on the resin of the column and incubated for 10 min. The column was centrifuged once again at 12000 rpm for one minute. When library complexity was an issue 20 µl of library cells were used to inoculate 6 ml adequate medium.

3.2.8 Preparation of electrocompetent *E.coli* cells

For library transformation purposes *E.coli* cells with high transformation efficiency were prepared. The strains were taken from glycerol stocks and streaked out on selective media. Single colonies were picked into 50 ml salt free TB medium on the following day. The culture was then incubated over-night at 37°C, 180 rpm. The next day 200 ml TB medium containing appropriate antibiotics were inoculated with 100 µl of the overnight culture and grown to an OD 600 of ~0.5. Afterwards the culture was split into four Falcon tubes and chilled on ice for 30 min. The cells were then harvested by centrifugation for 15 min at 2°C and 4000 rpm. From this point on the experiment was done in the cold lab with great precaution that the culture temperature did not exceed 4°C by continually handling it on ice. In the next step the medium supernatant was discarded and the pellet was resuspended in 20 ml ice-cold

water. The cells were repeatedly centrifuged at 2°C and 4000 rpm for ten minutes. The pellet was resuspended in 10 ml Millipore water followed by another centrifugation step. Subsequently the Volume was reduced to 5 ml and the cells were centrifuged another time. Finally the cell pellet was taken up into 3 ml 10% Glycerol precooled to 4°C. The competent cells were split into 50 µl aliquots. The cells were used directly for library transformation or were frozen for long term storage. The aliquots were put at -20°C for one hour and transferred to -80°C afterwards.

3.7 Electrotransformation of the cDNA library

To transform electrocompetent *E.coli* cells with pJuFo cDNA library plasmid a Gene Pulser (Biorad) was used. The plasmid DNA was dialyzed against 2% Glycerol for one hour prior to electroporation. An aliquot of 50 µl electrocompetent *E.coli* cells was thawed on ice and mixed with 1 µg plasmid DNA. The cells were incubated with the plasmid for one minute on ice and then transferred to a precooled cuvette (0.1 cm). The cuvette was placed in the chamber slide and pulsed with 1.8 kV for 5 ms. The cuvette was removed and mixed with 1 ml of SOC medium (37°C) instantaneously. Then the transformants were grown for one hour at 37°C, 300rpm. To increase colony number five electroporations were made for eachof the three strains (TG1, XL-1 and ER2738) and were spread on ten 2YT/Amp/Glu bioassay dishes (Nunc). As a positive control, cells were transformed with pUC18. The cells were incubated overnight at 37°C. The next day a confluently grown cell layer was collected with a lazy spreader and transferred to 500 ml 2YT/Amp/Glu liquid medium. The suspension was grown for 60 min at 37 °C, 300 rpm. For long term storage glycerol stocks were prepared. The cells were centrifuged at 2000 rpm for 30 minutes and resuspended in 2YT containing 30% glycerol. They were stored in 500 µl aliquots at -80°C.

3.2.9 Propagation of helper phage M13K07

The helper phage M13K07 is a M13 derivative and provides components for phage assembly and replication. It can be propagated in any F'Episome positive *E.coli* strain. For selection purposes it carries a kanamycin resistance. To produce large amounts of M13K07 phage particles *E.coli* TG1 cells were used as host strain. The cells were grown to mid log phase (OD 600, 0.5-0.7) prior to infection. Helperphage from a stock solution with a titer of $1x10^{11}/ml$ was used to prepare a 1 in 10 dilution series with PBS. From each of the dilutions 100 µl were combined with 100 µl of *E.coli* cell suspension and 4 ml Top Agar. The mixture was poured on 2YT agar plates and grown over

night at 37°C. The next day single plaques were selected and grown in 4 ml 2YT/Kan until turbid. Subsequently 400 ml liquid culture was inoculated with 1ml of phage infected *E.coli* culture. The culture was grown at 30°C, 600 rpm for at least 16 h. To isolate the helper phage from the culture supernatant a PEG precipitation was made. The phage containing *E.coli* culture was centrifuged for 30 min at 4000 rpm and 40 ml of the supernatant were mixed with 10 ml PEG/NaCl respectively. The phages were allowed to precipitate for at least one hour on ice and were then centrifuged for another hour at 4°C, 4000 rpm until a white phage pellet formed. Each pellet was resuspended in 1 ml PBS and centrifuged at 12000 rpm to separe from remaining *E.coli* cells the supernatant was transferred to a new 1.5 ml tube. The procedure was repeated until no *E.coli* pellet was visible anymore. To get rid of any remaining *E.coli* cells in the samples they were pasteurized at 70°C for 60 s. The helper phage was stored at 4°C.

3.2.10 Production of recombinant phages

For production the initial phage library used in selection round 1, phagemid-bearing libraries Hyperphage was used. Hyperphage has a pIII deletion, thus no wild type pIII is displayed, increasing the amount of displayed protein significantly (Rondot *et al.* (2001)). To induce the assembly of phages fused to recombinant protein the pJuFo cDNA library glycerol stocks from TG1, ER2738 and XL-1 blue strains were used. The glycerol stock was thawed and added to 400ml 2YT/Amp/Glu. The cells were grown to an OD of 0.5 when $1x10^{12}$ Hyperphage particles were added. The culture was incubated for 30 min at 37°C to allow the infection to take hold. The culture was then shaken for one hour at 200 rpm. For induction of the Lac promoter the medium was changed. The cells were pelleted by centrifugation and resuspended in 600 ml 2YT/Amp/Kan without glucose. The liquid culture was incubated for 16 h at 30°C, 200 rpm. The following PEG precipitation of the phages was performed equal to the helperphage M13K07 preparation. To avoid degradation of recombinant proteins the samples were handled on ice. The phage pellet was resuspended in 950 µl PBS, 40 µl protease inhibitor (Roche complete) and 10 µl BSA. The recombinant phages were used for the Phage Display experiment within 48 h after production. For the preparation of phages in the consecutive selection rounds, M13K07 was used as a helper phage.

3.2.11 Western Blot analysis of recombinant phages

The expression of recombinant library proteins was confirmed on western blot. For the SDS-PAGE a NuPAGE® Novex 4-12% Bis-Tris Gel (1.0 mm, 12 well) was used. Samples were mixed with non-reducing Lämmli buffer and heated to 95°C for five minutes before being loaded on the gel. Electrophoretic separation was performed at 200 V for 20 min in MES buffer. The Gel was then assembled to a wet blot sandwich to transfer the protein to a nitrocellulose membrane. The gel was placed on the PVDF membrane and placed between two buffer wetted whatman papers. This sandwich was put in the blotting chamber. The chamber was filled with TRIS/Glycin buffer and a current of 400 mA was applied for one hour. The membrane was blocked for one hour with blocking buffer. As primary antibody mouse anti-pIII was used diluted 1:1000 in blocking buffer and incubated for one hour. The membrane was washed three times with PBS-T. HRP-coupled anti-mouse IgG was used as secondary antibody, diluted 1:5000 in blocking buffer and left on the blot for one hour. The membrane was then washed for another three times and developed with CN/DAB substrate.

3.2.12 Determining the Titer of M13 Bacteriophages

To measure the infectivity of a phage preparation a titration assay was established. A tenfold serial dilution was made up to 10^{-12}. 100 µl of the dilution was mixed with 100 µl *E.coli* TG1 cells (OD 0.5). The assay was incubated for 30 min at room temperature to allow the infection to take place. From every well, a droplet of 10 µl was spotted on a 2YT/Kan/Glu agar plate and on 2YT/Amp/Glu. Negative controls were made with PBS, uninfected culture and phage solution. The infected *E.coli* cells were incubated over night at 37°C. For analysis of the assay colony forming units (cfu) were counted and extrapolated:

$$Phagetiter[\tfrac{cfu}{ml}] = number\ of\ colonies \times Factor\ of\ dilution\ series \times Volume\ dilution\ factor$$

The phagetiter calculated on Amp/Glu plates represents the amount of phagemid-bearing particles, while the colonies on Kan/Glu represent the phage particles bearing the helper phage genome

3.2.13 Coupling Antibodies to Tosylactivated Dynabeads

The magnetic beads used for biopanning (compare section 3.2.14) in phage display were coupled to a primary antibody prior to use. Therefore the beads (100 mg/ml) were resuspended thoroughly and 250 µl of the suspension was

transferred to a 1.5 ml tube. The tube was placed on a magnetic stand and the clear supernatant was discarded. The beads were washed five times in 500 µl coating buffer to remove any traces of inhibiting preservative. The beads where then mixed with 416 µl antibody solution corresponding to 1 mg antibody and 208 µl $(NH_4)_2SO_4$ (1 M). The beads were incubated rotating slowly at 37°C for 24 h. Subsequently the supernatant was removed, replaced with 625 µl blocking buffer and placed back on the rotator for another 12 h. After incubation the beads were washed three times with washing buffer and resuspended with 50 µl to a final concentration of 50 µg/µl. Not more than one day before their use as bait in the, panning process the beads were loaded with serum antibodies. For each selection round 50 µg beads are needed so a 5 µg/µl stock solution was prepared. 500 µg of the primary antibody loaded bead solution was placed on a magnet stand, the clear supernatant was pipetted off and 80 µl PBS with 20 µl serum was added. The bead suspension was then incubated for one hour, slowly rotating at RT. The beads were then washed three times in 500 µl PBS and used for biopanning within 24 hours.

3.2.14 Biopanning

The panning step in phage display is carried out to find specific binders for a target. The enriched phages from a panning round are used for the subsequent panning round so that a gradual enrichment can be achieved. The process was conducted in a semi automated manner on a Kingfisher 96 magnetic particle processor (Figure 6) .

Figure 6: King Fisher Flex Magnetic Particle Processor (Thermo Scientific)

A total of four panning rounds was performed for each *E.coli* strain (XL-1, ER2738, TG1). The strains were kept temporally and spatially separate to avoid any cross-contamination. For the first panning round Hyperphage, presenting library proteins on its surface was used. In subsequent panning round the enriched libraries were produced with M13K07 as helper phage.

To reduce unspecific binding the phages were pre-absorbed prior to their use: The phage solution was mixed at the rate of 1:1 with 2x PTM and 10 µl uncoated beads were added. They were then incubated for one hour on an overhead shaker. The robot was loaded with NuncU96 plates. The plates were loaded as indicated in Figure 8. First, a "Phage Plate" was prepared. The wells of the first six columns were filled with 200 µl of pre-absorbed phage library, diluted in 1x PTM. The "Bead Plate" contained magnetic beads coupled with patient IgG or patient IgA antibodies from disease specific sera. The "Wash Plate" was loaded in the same manner with 200 µl PBS-T per well. For each consecutive panning round one additional "Wash Plate" was prepared. The "*E.coli* Plate" contained culture at an OD of 0.5. The culture had to be ready for loading at a given time so the beads could be transferred to the cells for infection.

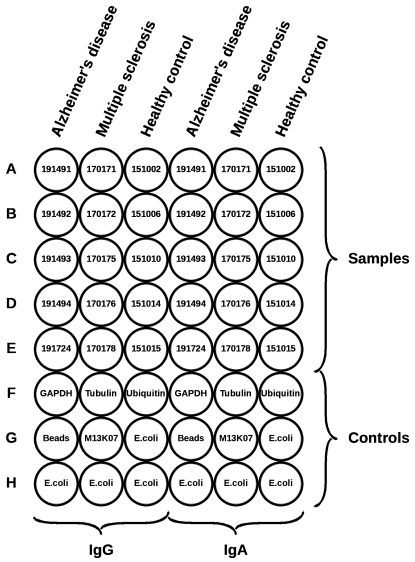

Figure 7: Plate loading scheme for Panning.
The first three columns were loaded with human IgG coupled beads from patient sera. The sera of patients with the same disease were loaded in the same column. Serum numbers are indicated in the circles. Columns 3-6 were loaded with beads coupled to IgA from patient sera. The bottom three rows contain controls. To monitor the enrichment process, mouse antibodies against the housekeeping proteins GAPDH, Tubulin and Ubiquitin were used. As negative controls non-serum loaded Beads only, M13K07 together with non-serum loaded beads and *E.coli* cells only were used.

The "*E.coli* Plate" was taken out of the robot after infection. In every well 20 µl of 10x Amp/Glu was pipetted and the plate was shaken at 37°C for 2.5 hours. Afterwards, the culture was diluted with 200 µl of 2YT/Amp/Glu

Table 4: Setup of the Kingfisher Panning Program Plate

Plate name	Working Step	Head position	Time (min)
"Bead Plate"	Blocking magnetic beads with PTM	Mix	60
"Phage Plate"	Interaction of beads with phages	Mix	60
"Wash Plate 1"	Wash 1 of magnetic beads in PBST	Mix	10
"Wash Plate 2"	Wash 2 of magnetic beads in PBST	Mix	10
"Wash Plate 3"	Wash 3 of magnetic beads in PBST	Mix	10
"Wash Plate 4"	Wash 4 of magnetic beads in PBST	Mix	10
"Release Plate"	Waiting position for magnetic beads until E. coli culture plate is ready for infection	Up	10
"*E.coli* Plate"	Infection of culture with phages	Up	60

per well leading to a total Volume of 400 µl. 200 µl of the culture were transferred to a new NuncU96 plate. Only one of these two plates was infected with helper phage M13K07 in the next step, the other one was kept as a backup plate. Every well was infected with $1x10^{10}$ helper phages followed by an incubation time of half an hour at 37°C. To allow the expression of lac repressed proteins the medium had to be changed afterwards. Therefore the culture was centrifuged for 10 min at 2000 rpm, the supernatant was pipetted off carefully and was replaced by 250 µl 2YT/Amp/Kan. The expression plate was incubated overnight at 30°C, 2000 rpm while the backup plate was shaken at 37°C, 2000 rpm.

3.2.15 Polyclonal phage ELISA

To visualize the interactions of enriched phages with serum antibodies an ELISA experiment was set up. The newly isolated phages from the previous panning round were directly used for ELISA to avoid proteolytic degradation. To ensure comparability between phages from different panning rounds the ELISA was semi-automated using a Kingfisher robot. Identical to the Panning procedure, NuncU96 micro well plates were filled with 200 µl each well. The "Bead Plate" was prepared as indicated in Figure 8. Furthermore, a "Phage Plate" and two "Wash Plates" were prepared. An "Antibody Plate" was filled with 200 µl anti-M13-HRP, diluted 1:5000 in PTM. A Nunc MaxiSorp ELISA plate was blocked for one hour with 200 µl 1x PTM per well and washed three times with PBS prior to its usage. The plates were placed in the Kingfisher instrument as indicated in 5. After the robot cycle was finished the plate was placed on a magnet and the supernatant was removed. The beads were then covered with 200 µl ABTS substrate, prepared according the manufacturer's

instructions. The plate was developed for 20 min at 25°C before the green supernatant was manually transferred to an ELISA Maxisorp plate by pipette for measuring. The plate placed in an Elisa reader (Applied Biosystems) at a wavelength of 405 nm.

Table 5: Setup of the Kingfisher robot program for ELISA

Plate name	Working Step	Head	Time (min)
"Bead Plate"	Blocking magnetic beads with PTM	Mix	60
"Phage Plate"	Interaction of beads with phages	Mix	60
"Wash Plate 1"	Wash 1 of magnetic beads in PBST	Mix	10
"Wash Plate 2"	Wash 2 of magnetic beads in PBST	Mix	10
"Antibody Plate"	Incubation of bead binding Phages with M13 HRP	Mix	60

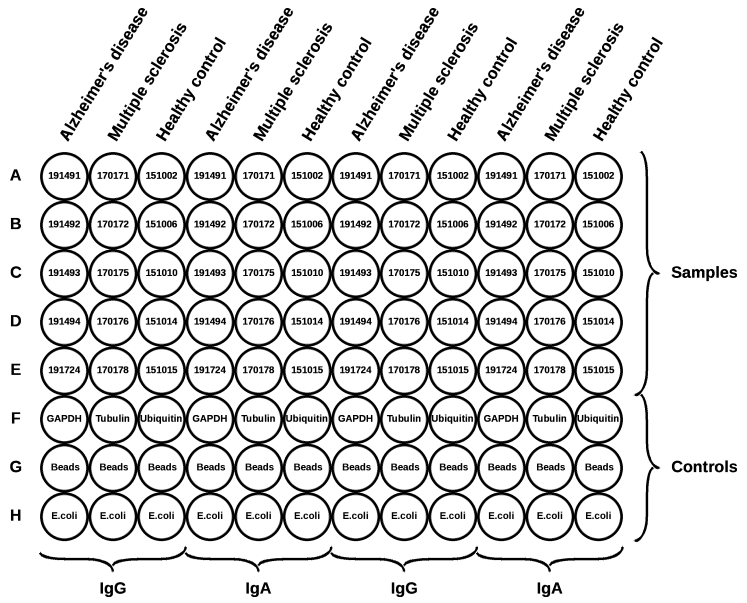

Figure 8: Plate loading scheme for ELISA.
The first six rows of ELISA plate were loaded analogous to the Panning plate. The second half of the plate was loaded with beads without serum antibodies interacting with phages for background measuring.

3.2.16 Quantitative real time PCR for quantifying enrichment of housekeeping genes from Phage display

To have a positive control for the specific enrichment of genes in Biopanning, antibodies against the housekeeping proteins GAPDH, Ubiquitin and Tubulin

were used to selectively enrich these genes from the library. These genes are highly abundant in a complex cDNA library and an increasing concentration from panning round to panning round is expected when they are panned against antibodies that are designed to recognize them. To measure the enrichment quantitatively qPCR was used. SYBR green was used as a reporter system. SYBR green is a dye that becomes fluorescent when it binds to double stranded DNA. This feature can be used to record the amplification of template during a PCR reaction. It is assumed that in the exponential phase of PCR the amount of template is doubled in every cycle. If the template concentration was high in a sample the exponential phase is reached quicker. A standard curve based method was used for quantification with an Applied Biosystems qPCR instrument. For making the standard curve a plasmid sample confirmed to contain the full ORF of the housekeeping gene, with a known concentration was used as a reference. A dilution series was prepared starting with 1.000.000 template molecules and diluting them in 1:4 steps to as little as 4 plasmid molecules. In the same manner a dilution series was prepared for the pJuFo plasmid. One primer pair specifically amplifying a hundred base pair fragment of the gene and another primer pair amplifying a hundred base pair on the pJuFo plasmid backbone were designed. Plasmid from library samples was amplified with both primer pairs. The concentration of housekeeping genes was not known and to make sure that the measured values are in the range of the standard curves different template concentrations were tested. To avoid pipetting errors doublets of every sample were pipetted. On the basis of the standard curves the amount of housekeeping genes in the samples can be calculated. The enrichment factors were obtained by dividing the measured amount of gene with the value for the total amount of plasmid.

3.8 Affinity Propagation Clustering of ELISA data

Affinity propagation clustering was used for finding patterns in the ELISA data. The algorithm builds clusters around exemplars. An exemplar represents the prime example for a member of a cluster. Exemplars are found by randomly selecting a subset of samples as prospective exemplars and then refining the set in an iterative manner. As input the algorithm needs measures of similarity between samples (Frey & Dueck (2007)). Samples were unambiguously defined by strain, patient serum and antibody class. Four data points were measured for each sample to show the enrichment after the panning rounds. The quotient of the absorption values between two consecutive rounds was calculated and logarithmized. This factor represents the relative enrichment between two rounds, so for each sample three factors were calculated. These

values were entered in a three dimensional coordinate system with every dimension corresponding to one factor. Similarity was defined as negative Euclidian distance between two points. These similarities were used as input data for the AP cluster implementation in R (Bodenhofer *et al.* (2011)). Background data were not considered in clustering.

3.2.17 Data Processing from 454 Sequencing

The sequences were received in SFF file format from the Roche 454 platform. They were converted to FASTA format with sff_extract 0.2.8 . The read length of the raw data were shown in a histogram. The reads were then sorted according to their vector sequence. Reads that could not be assigned to pJuFoA, pJuFoB or pJuFoC were disregarded for further analyses. The sequences were then trimmed by removing the vector sequence from the insert sequence. The trimmed sequences where then subjected to a BLAST against human RefSeq RNA from August 2011. The BLAST hits for each read were considered as matches after applying the following heuristic criteria: Hits that scored less than 50 were excluded. Hits with a score lower than 100 were included only if more than half of the read matched. For the other hits the identity was taken as a decision criterion. Hits with identities smaller than the identity from the hit with the maximum score were not incorporated. Reads were assembled to contigs using MIRA3 (Chevreux *et al.* (2004)). To find specifically enriched genes the RefSeq's for each library were counted. A frame analysis was made on the basis of the CDS information of the RefSeq's. The CDS of the sequenced gene had to be in-frame with the Fos sequence of the vector backbone to be considered as in-frame. To characterize the initial library in terms of cDNA quality all reads were drawn in reference to the average matched mRNA.

3.2.18 Protein-Protein Network construction

For analysis of the phage display hits on protein level, the contig sequences were translated into all three forward frames using the BioPerl Translate function and searched for the peptide motif GCRSS, which is known to be expressed in frame with the Fos gene fragment so they are are displayed on the phage surface. Peptides that did not contain the motif within the first 90 amino acid residues were discarded as out-of-frame or otherwise faulty. The remaining sequences were prepared for protein BLAST by removing the vector sequence as well as any sequence following the first stop codon. All genes matching in the BLAST with a score of at least 50 and an identity of at least 80% were considered hits and were used to query the ConsensuspathDB

protein interaction database, release 20, for protein-protein interactions. The hits and their first-degree interactors were used to query the database again, this time for any interaction within the set. (This step is necessary to find all interactions between first-degree neighbours.) The resulting protein-protein interaction network was visualized using Cytoscape (Cline *et al.* (2007)). All first-degree neighbors were manually curated for disease relevance. Nodes with similar interaction patterns and without obvious disease relevance were merged into meta-nodes for greater clarity.

4 Results

The aim of this study was the development of a novel method suited for the systematic high-throughput discovery of biomarkers. In a proof-of-principle phage display screen the method was used to find peptide antigens that specifically react with patient antibodies. Phage display is ideally suited as method for screening such biomarkers, because phages can be easily produced in high numbers and there is little constraint in terms of insert complexity. The employment of a semi-automated robot device and Next Generation Sequencing technology make Phage Display available for screening large numbers of patients. Part of the optimization of the method was the evaluation of different bacteriophage host strains. Multiple sclerosis and Alzheimer's were used as model diseases in this study. Biomarkers are urgently needed to diagnose and monitor progression of these disorders. Good biomarkers should not only be able to discriminate between healthy and diseased conditions but also between Alzheimer's and multiple sclerosis. As peptide pool for screening patient antibodies a cDNA library was used in this study, because, compared to ORF-enriched libraries or synthetic libraries, a cDNA library combines the advantages of both of them. While short synthetic peptides can be ideally suited as biomarkers with high sensitivity and specificity, peptides derived from the CDS of cDNA inserts can also shed light on disease mechanisms.

In this study a human brain cDNA library was used as peptide target for the interaction with patient antibodies. As expression vector for this library the pJuFo phagemid was used in three versions each expressing one frame as a Fos-fusion protein. The library was available in each of the three vectors increasing the chance of each cDNA insert to be expressed in the right reading frame. Three different *E.coli* strains were used as bacteriophage hosts for phage display. The three versions of the pJuFo phagemid were combined. Phagemid bearing *E.coli* strains had to be infected with a helper phage to induce the assembly of bacteriophages that present a foreign library protein on their surface. In this study Hyperphage was used as helper phage in the first selection round because it increases the proportion of protein displayed on the phage surface. In all experiments the phage library samples expressed in each of the three *E.coli* strains were kept completely separate to allow the identification of the *E.coli* strain that is best suited for biomarker screening.

In an effort to characterize the cDNA library used in this study the phagemids were isolated from each of the strains and inserts amplified in an Emulsion-PCR. Sequencing on a FLX-454 Genome Sequencer platform served the purpose of determining library complexity and insert quality. To isolate phage clones from the library that express potential biomarkers, antibodies from patient sera were applied to specifically interact with the phage library.

In an attempt to answer the question which host strain is the best in a more defined experiment, antibodies to the housekeeping proteins GAPDH, Tubulin and Ubiquitin were used for interaction with the phage library, in addition to the patient antibodies. To monitor the enrichment process, the amount of enriched housekeeping genes was measured using quantitative PCR after each panning round. To validate the results, single clones were Sanger sequenced. The pool of binders enriched with patient serum was sequenced on the FLX-454 platform and evaluated with bioinformatic methods. A subset of disease specific inserts that were expressed in frame and not present in healthy controls was analyzed in silico. Therefore a protein-protein interaction network was buildt to show their biological relevance. This approach does not only identify the most promising biomarker candidates but also opens new leads for investigating disease etiology.

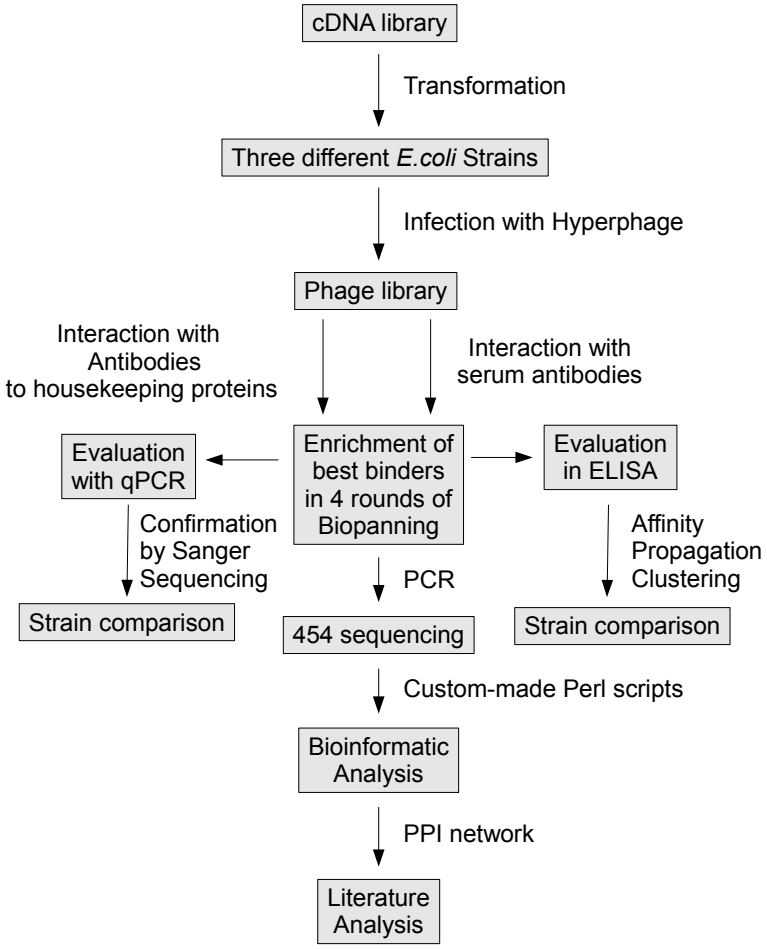

Figure 9: Classification of experimental procedures in the phage display based biomarker screen.

Depicted is a schematic diagram of the experimental workflow and classification of the results. The cDNA library was transformed into three different *E.coli* strains. Expressed peptides were presented on the phage surface. Binders to patient antibodies and housekeeping gene antibodies were enriched in 4 rounds of biopanning. For host strain comparison housekeeping gene concentration was measured with qPCR and enrichment of patient antibodies was monitored in ELISA. For analysis of potential biomarkers the enriched binders of panning round four were sequenced on a FLX-454 sequencing platform and evaluated with bioinformatics tools. Highly promising candidates were used for the construction of a disease specific protein-protein network.

4.1 Characterization of a human brain cDNA library

A human brain cDNA library previously established in the lab was used for phage display. The library had not been characterized or been used in other experiments before. The library had been constructed with a CloneMiner I cDNA library construction kit (Invitrogen) that uses Gateway technology for cloning. Initial biological material was commercially purchased mRNA, derived from a 54 year old donor who died from liver cancer. No other diseases are known for this patient. For library construction the cDNA was shuttled via BP reaction into entry vector pDONR222 and $1.27 * 10^5$ primary clones were generated. These inserts were shuttled into pJuFo A, B and C vectors, respectively and pooled to create a single library. A number of $9.6 * 10^6$ clones after the LR reaction suggests that, statistically, each insert was cloned into all three vectors. The three pJuFo vectors differ in one additional nucleotide between Fos and insert and thus potentially express cDNA inserts in each forward reading frame. Three different *E.coli* strains, XL-1 blue, ER2738 and TG1, were transformed with the library. These strains bear the F-Episome and are commonly used in phage display. To sustain the complexity of the library it was of major importance to optimize transformation efficiency. The critical step for achieving high clone numbers is to use electrocompetent cells with extremely high transformation efficiency. Hence *E.coli* cells with an efficiency of more than 10^9 transformants per µg plasmid (pUC18) were prepared. Deviant from standard protocols the cells were used for transformation directly, without deep-freezing them to avoid loss of efficiency (compare Seidman *et al.* (2001)). Dialysis of library plasmid, to free it from traces of disturbing salts, further improved electroporation efficiency. Clones obtained from ten transformation reactions were pooled to maximize clone numbers. Transformation resulted in $9.6 * 10^9$ transformants for TG1, $5 * 10^9$ for XL-1 blue and $2 * 10^{14}$ for ER2738. In summary, the brain cDNA library contained $1.27 * 10^5$ clones, this equates to the maximum of possible inserts. The number of clones after shuttling into pJuFo was about an order of magnitude higher than that after library preparation. The number of clones after transforming the phage display host strains is orders of magnitude greater still. This suggests that losses of complexity are minimal. To estimate the average insert size a colony PCR was made before sequencing the library.

Colony PCR shows that the library contains different inserts and 48 out of 90 clones result in a product. Sizes are in the range between 250 bp and 1200 bp. This is surprisingly low because the size fragmentation during library construction should exclude inserts below 500 bp and average size envisaged with the used kit is 1500 bp. Results from colony PCR are qualitative

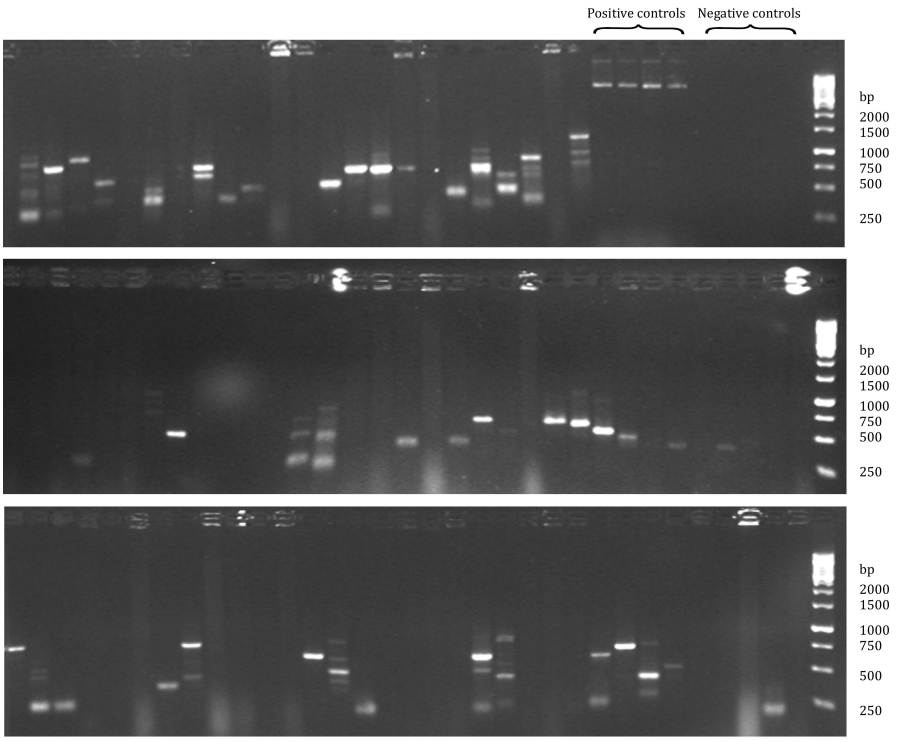

Figure 10: The cDNA library contains inserts at different sizes.
Colony PCR on single clones picked after library retransformation was made with primers on vector backbone. Overall 48 clones with an insert size higher than 250 bp can be observed. Most inserts give a band in the range between 250 bp and 750 bp. The largest observed insert size is 1200 bp. The last eight lanes on the right side of the upper panel show negative and positive controls.

and do not statistically represent library size distribution. Therefore the cDNA library inserts had to be sequenced for more detailed information. To get a high sequencing coverage 454 Next Generation Sequencing was used for that purpose. Library phagemids isolated from the three *E.coli* strains were sequenced separately. The cDNA inserts had to be amplified and a sfi I cleavage site was added prior to sequencing as an adaptor for ligation to sequencing beads. PCR on such complex samples can result in a bias because unpredictable reasons can lead to higher amplification of some inserts compared with others. To circumvent this problem emulsion PCR was used (Figure 11). In this method the PCR Master Mix is emulsified in mineral oil. The resulting water-in-oil emulsion leads to a compartmentalization of PCR reactions in emerging micelles. This physical separation into distinct compartments leads to a more controlled reaction, similar to the use of a

single test tube for every single insert (Schütze *et al.* (2011)).

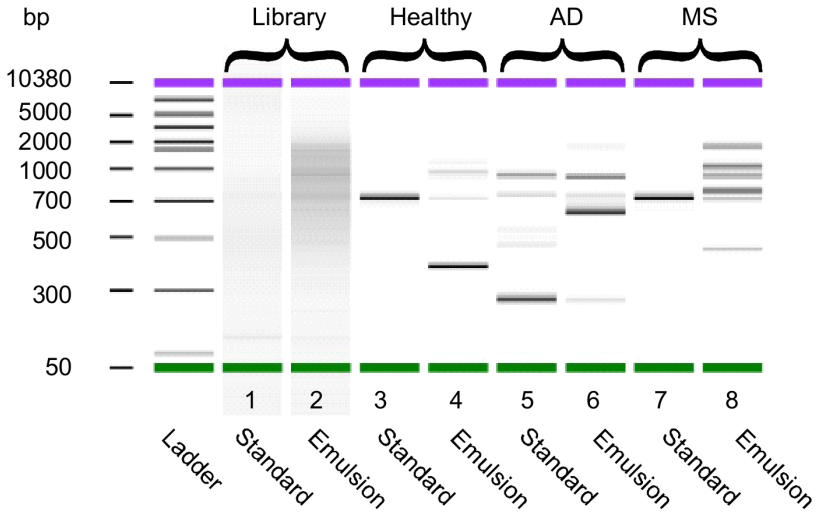

Figure 11: Emulsion PCR preserves library complexity.
All Samples were analyzed on a Bioanalyzer 7500 DNA chip prior to sequencing. Results for ER278 derived libraries that were enriched with IgG as target in panning are shown. The ladder covers the range from 50 to 10380 bp. Lane 1 shows the initial library sample prepared with standard PCR and Lane 2 shows the same library prepared under standard analytical PCR conditions. In Lane 3 and 4 the same can be seen for the Healthy sample. Lane 4 and 5 show the comparison for an Alzheimer's sample and in Lane 7 and 8 a multiple sclerosis sample can be seen.

Results from the bioanalyzer chip for the initial library indicate that fragments in the whole ladder range can be detected as smeared bands. The comparison between emulsion PCR and standard PCR demonstrates that more fragments were amplified with emulsion PCR. For example the MS sample shows six bands in emulsion PCR instead of only one in standard PCR. It was therefore decided to amplify also the other libraries with emulsion PCR for sequencing. After these preparation steps the libraries were sequenced with a 454 Genome Sequencer FLX System. Each library was tagged with a barcode so they could be sequenced in one run and remain still distinguishable. Sequencing resulted in 7191 reads for the library derived from ER2738, 3155 reads for the TG1 library and 4045 reads were obtained for XL-1 blue. The reads were assembled to contigs. Building contigs was made with the MIRA 3 EST Sequence Assembler for 454 Data. Similar sequences likely to belong to the same gene were merged during assembly.

The assembly of contigs changes the size distribution (compare Figure 12) and reduces redundant sequences. Overlapping short reads are merged

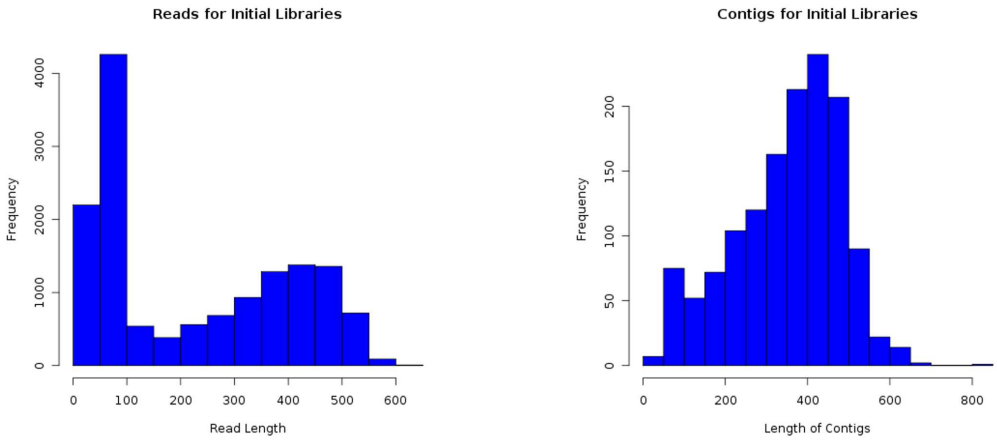

Figure 12: Building Contigs shifts length Distribution towards longer sequences.
Overlapping sequences were assembled. Size distribution is shifted towards longer reads because short reads are merged into longer sequences. The peak below hundred base pairs in the raw sequences, resulting mainly from vector backbone sequence, is lower in the contigs.

to longer sequences. This approach has advantages in further sequencing analysis. Sequencing errors can be compensated to some extent, so contigs are more reliable and cannot only be mapped confidently on DNA and Protein level but can be also analyzed in terms of reading frame. Read length does not necessarily reflect insert size because average sequencing length for 454 sequencing is 400-500 bp with the used system. This means that longer inserts will always produce reads and contigs shorter than their actual length. The polyA tail can be found in 3097 of the total 14391 sequences (22%). Due to the directional sequencing approach the start of the sequence is always the 5'-end of the insert. Since cDNA synthesis starts at 3'end of the mRNA this information is sufficient to characterize the insert very well and the beginning of 5'UTR or CDS can always be mapped, if present, while some 3'UTR sequence information might be lost.

Reads and Contigs were mapped separately to human RefSeq RNA using nucleotide blast. This algorithm is not widely used for analyzing 454 data because of its low speed and high computational demands. Nevertheless it is perfectly suited for the small dataset analyzed in this work because of its preciseness. Reads that have an identity higher than 90% and a score higher than 100 were arbitrarily chosen as cutoff value for reads mappable to RefSeq RNA. The score indicates how well the sequences match and the identity is the percentage of identical bases. To get more information about the quality

of the human brain cDNA library the relative position of the match on the respective mRNA was analyzed (compare Figure 13).

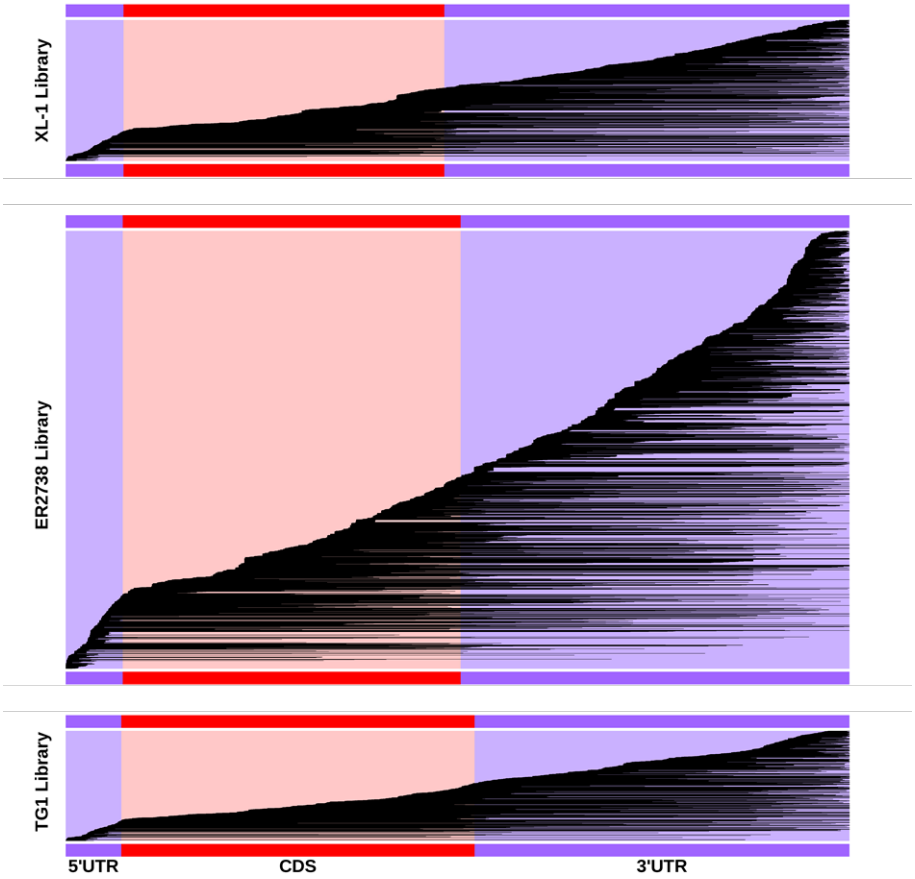

Figure 13: Positions of Mapped Reads on their respective mRNA.
The average length of all mapped mRNAs was calculated. Untranslated regions have a purple background and the CDS region is depicted in red. The upper panel shows reads from 454 sequencing of the brain cDNA library transformed into XL-1. The central panel shows reads from the ER2738 derived library and the lower panel shows sequences of the library in TG1.

Table 6: Results from Mapping Reads and Contigs to RefSeq RNA

	Number of	
	Reads	Contigs
Total	14391 (100%)	1408 (100%)
Mappable to RefSeq RNA[a]	7148 (50%)	988 (70%)
Mappable to protein coding RNA[b]	4664 (65%)	936 (95%)
3'UTR only	2515 (54%)	478 (51%)
CDS available total	1879 (40%)	395 (42%)
CDS in frame	561 (30%)	115 (29%)
CDS without 5'UTR	395 (70%)	77 (67%)

[a]Reads with an identity >90 and score>100 are considered mappable
[b]Subgroups are set apart with text indentation, percentages are calculated for subgroups

Mapping reads resulted in 7148 BLAST Hits and 988 Hits for contigs. The percentage of hits that match to Refseq RNA is higher for contigs (Table6). This result is a consequence of reduced low quality reads in contigs. Reads match to protein coding RNA for 65% of the mappable RefSeq RNAs and contigs match for 95%. The other hits match to RefSeq RNA with no CDS information available. The percentage of inserts that contain CDS is particularly interesting. When only reads mappable to protein coding RNA are considered 40% of them and 42% of the contigs have CDS information. Only CDS expressed in frame with the Fos gene is expected to be translated. This is the case for 30% of the reads and 29% of the contigs that have CDS information. Inserts that include CDS and 5'UTR can potentially bear stop codons. Statistically there are 64 codons and three of them are stop codons, thus the chance is 5%. A relatively high amount of sequences, 54% of the reads and 51% of the contigs match to 3'UTR only.

In summary comparison between reads and contigs shows similar results. From analyzing the relative position of the reads on their relative mRNA it can be learned that more than half of the inserts match only to 3'UTR. When these inserts are translated, synthetic peptides are displayed on the phages. The amount of inserts that have CDS in the correct frame is low. However it is likely that only a fraction inserts were sequenced. The next issue addressed was the complexity of the library. For this analysis different genes found in each of the three strains were counted and the overlap between them was used to estimate complexity (adapted from Rosner, 2006).

Figure 14: Venn diagram showing unique genes for the library in each strain and the overlap.
Overall 1389 different genes could be matched to RefSeq genes. The red circle shows the number of genes from the XL-blue sample. The green circle represents genes from the TG1 sample and the yellow circle contains the number of genes for the ER2738 sample. 79 genes are found in all three strains. 163 genes are specific for XL-1 blue, 94 genes can be found only in TG1 and 424 genes are unique for ER2738. The overlap between XL-1 and ER2738 is 251. The overlap between ER2738 and TG1 is 220 and between TG1 and XL-1 it is 185.

In all three samples taken together 1389 different genes were found. The genes common to all three samples are shown in Figure 14. This number is composed of 520 genes from XL-1 blue, 420 genes from TG1 and 816 genes from ER2738. XL-1 blue and TG1 share 185 common genes, ER2738 and TG1 share 220 and ER2738 and XL-1 share 251. The number of genes that can be found in all three strains is 79. It is not known how complete the coverage from sequencing was, but it is reasonable to assume that the library used in this study is larger in terms of different mRNAs. To estimate the actual complexity, the contigs were mapped to human RefSeq RNA identifiers. Assuming an equal probability of finding each cDNA represented in the library, the expected overlap between two strains can be used to extrapolate original sample size. The overlap is equal to the product of the size of the two samples divided by the number of possible genes. For example, the number of different genes found for XL1-blue is 520 (red circle in Figure 14), the number of different genes found for ER2738 is 816. The number of genes

found in both samples is 251 (first column in Table 7). Consequentially the complexity can be estimated as (520×816)/251=1691. The same can be done for each combination of two samples, resulting in somewhat lower estimates of 1558 ER2738 and TG1 or 1181 for XL-1 blue and TG1, respectively. The overlap of any two samples can also be regarded as a single sample. It must then be multiplied by the third sample and divided by the three-way overlap (right half of Table 7). Doing this analysis results in similar estimates for complexity. Complexity is 1334, 1373 or 1911 different genes after using this calculation. The fact that estimates differ, depending on which samples are considered, and that estimates are higher for larger sample sizes suggests that the assumption of finding all cDNAs in equal proportions is violated. This is consistent with the type of library used, as cDNA is produced from a pool of mRNAs of different abundance and there was no equilibration, either inherent in the method of library preparation (as with ORF libraries) or extraneous like in cDNA equilibration methods (like in Bonaldo *et al.* (1996)).

Table 7: Estimation of library complexity on the basis genes overlapping in sequenced samples

	XL-1 vs.	ER2738 vs.	TG1 vs.	XL-1&ER2738 vs.	ER2738&TG1 vs.	TG1&XL-1 vs.
Size of	ER2738	TG1	XL-1	TG1	XL-1	ER2738
Sample 1	493	816	420	251	220	185
Sample 2	816	420	493	420	493	816
Overlap	251	220	185	79	79	79
Estimated Complexity	1603	1558	1119	1334	1373	1911

Library complexity on gene level does not necessarily reflect complexity on protein level. The question which library peptides were expressed on the bacteriophages is crucial. This is the pool of molecules that is available for interaction with patient sera and is essential for the outcome of the phage display experiment. For analyzing the libraries on protein level only contigs were used because sequencing errors can lead to unreliable frame information. The sequences were translated into all three possible frames and trimmed after the first stop codon. Analysis was restricted to Fos-specific peptide motif GCRSS containing sequences and all other sequences were discarded. It should be emphasized that this is not necessarily the frame for the expression of the corresponding RefSeq but the frame phages use to express the Fos-fusion peptide. Further the size distribution of library peptides was analyzed (Figure 15).

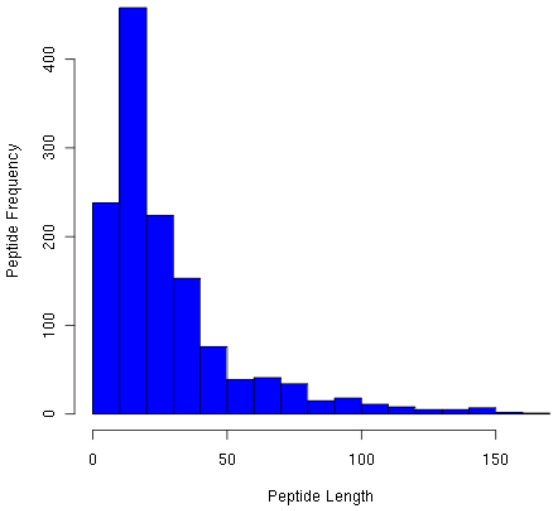

Figure 15: The distribution of peptide size shows that most peptides are smaller than 50 amino acids.
Overall 1331 peptides with more than 5 amino acids are expressed. Average peptide length is 29 amino acids. 1144 peptides are smaller than 50 amino acids, 146 are below hundred and 42 proteins are more than hundred amino acids long.

Peptide size distribution of expressed proteins shows that the vast majority of peptides are below 50 amino acids (Figure 15). Size distribution of translated contigs does not provide information about library complexity on the protein level. For that reason peptides were mapped against the human Swiss-Prot database with BLASTp. The threshold for the identity was set to 30. Peptides with an identity higher than 95% are very likely to be human proteins. Overall 125 peptides that fulfill these criteria were identified. Nevertheless hits with smaller identities are interesting but much more difficult to analyze. They can be mimotopes that are potentially recognized from patient antibodies with high specificity. A high percentage of cDNA is expressed as synthetic peptides in the studied library. These inserts contain 3'UTR, are out of frame or have a stop codon that prevents CDS translation. To estimate complexity at protein level is challenging because many assumptions have to be made. Neither the distribution of gene frequencies on mRNA level is known nor any bias occurring during library construction and sequencing procedure. The estimated complexity of 1483 different genes on average (Table 7) is likely to be three times as high on protein level because peptides are translated in all three frames. Furthermore

it can be argued that even if two cDNA fragments match the same RefSeq at the nucleotide level, the fact that they are different in size, can result in the folding of different epitopes. A library that combines a large amount of synthetic peptides can be of advantage for a biomarker screen. Fully synthetic small peptides are for example highly successful in the diagnosis of rheumatoid arthritis (reviewed in van Venrooij *et al.* (2011)).

4.2 Phage library expression in different *E.coli* strains

Phage display is a method for selecting binders to a chosen target. Bacteriophages are exploited to present recombinant library proteins on their surface. In this work the pJuFo phagemid system was used for that purpose. Phage display of cDNA libraries is not compatible with most standard phage display systems because naturally occurring translational stop codons in the 3'-regions would prevent pIII translation. Moreover, the insert has to be in frame with pIII and the leader sequence. To circumvent these problems, the cDNA inserts are expressed in all three reading frames and displayed as indirect phage coat protein fusions. Therefore the cDNA fragments are fused to the 3'end of the of a leucin zipper domain while the minor coat protein pIII is expressed separately as fusion protein with the Jun domain of the Jun gene (Crameri & Suter (1993)). These molecules show a very high affinity towards each other and hetero-dimerize in the periplasm. This physical linkage between pIII and the recombinant peptide is additionally strengthened by two disulfide bridges. In this work the *E.coli* strains XL-1 blue, TG1 and ER2738 were used as host strains for phage expression and the cDNA library was transformed in each of the three strains. The pJuFo phagemid carries a pelB leader sequence for periplasmic targeting, two Lac repressors, a kanamycin resistance as well as the Jun fusion with gIII and the Fos-cDNA fusion. Upon super-infection with a helper phage the assembly of recombinant phages can take place. The helper phage encodes the full complement of capsid-encoding genes that are missing from the phagemid. This M13 derivative is defective in its origin of replication and therefore its genome is packaged less efficiently than that one of the recombinant phages. To test the display of recombinant proteins on M13 bacteriophages a western blot was made. The samples had to be prepared under non-reducing conditions for this approach to protect the disulfide bridges that stabilize the connection between pIII and the recombinant Fos fusion peptide. The peptides were separated and detected with an antibody directed against pIII. Protein bands resulting from fusion peptides were expected at larger sizes than wild type pIII.

Analysis of recombinant bacteriophages on western blot (Figure 16)

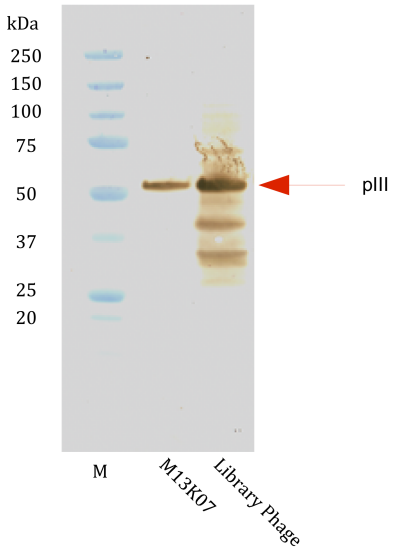

Figure 16: Western blot analysis shows recombinant pIII fusion proteins.
The western blot (4-12% gradient gel) shows recombinant peptides that were expressed in ER2738 as host strain and displayed as fusion with pIII on bacteriophage M13. As helper phage M13K07 was used. In the left lane the marker can be seen (M). Next to the marker M13K07 was loaded as control ($6.5 * 10^9$ cfu) for library phage ($5 * 10^{10} cfu$). Minor coat protein pIII was detected at 53 kD. Recombinant peptides were detected in the range between 100 kDa and 50 kDa. Bands at sizes below wild type pIII are detected between 50 kDa and 25 kDa.

demonstrates that additional bands could be detected compared with helper phage M13K07. Peptides fused to pIII can be detected at sizes as large as 100 kDa. The fact that proteins can be detected below wild type pIII can be probably attributed to degradation products. The wild type pIII band is visible not only in M13K07 helper phage but also in recombinant M13. This is an inherent consequence of the used helper phage M13K07. Competition of wild type pIII and recombinant pIII incorporation during phage assembly leads to a mixture of both variants on the phage coat. As little as 1-10 % of the phages in a library present the recombinant peptide fused to one of the five pIII proteins (Chasteen *et al.* (2006)). The titration of the phages resulted in $1.3 * 10^{12}$ cfu on Amp/2YT/Glu plates and $4.2 * 10^8$ cfu on Kan/2YT/Glu. This ratio is expected for the use of helper phage. However using Hyperphage could be of advantage in this work. This M13 derivative has a deletion in gIII. When Hyperphage is used for superinfection of phagemid bearing *E.coli* cells no wild type pIII can be packaged. This results in a higher amount of recombinant protein per phage particle in comparison with M13K07 .

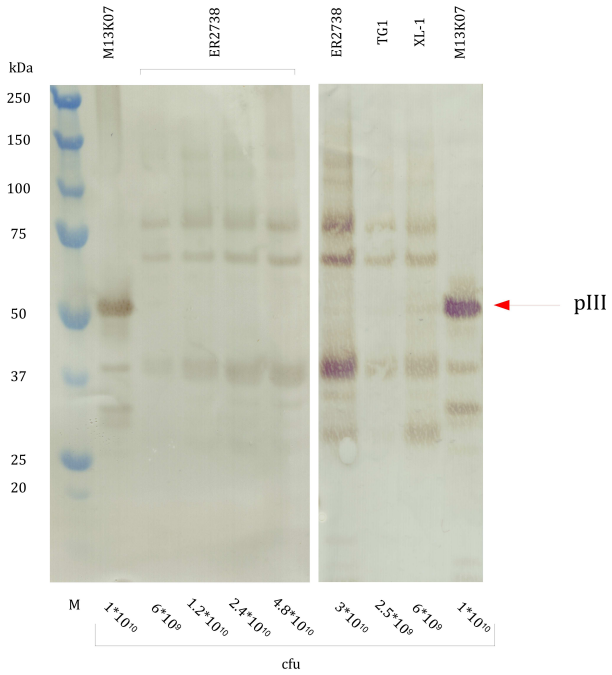

Figure 17: Hyperphage presents fusion proteins on all his five pIII-coat proteins.

The western blot (4-12% gradient gel) shows recombinant peptides from phage libraries that were produced in different host strains. Hyperphage was used as helper phage. On the left panel a dilution series of ER2738 derived phages were loaded. The right panel shows the peptides of phage libraries produced the three different *E.coli* host strains ER2738, TG1 and XL-1 blue. The minor coat protein pIII is detected at 53 kDa, in the first lane of the left panel and the last lane of the right panel.

Results from the western blot analysis of library phages produced with Hyperphage demonstrate that there is no wild type pIII band in these samples (compare Figure 17). Two additional bands, one at 40 kDa one at approximately 30 kDa can be seen in the M13K07 sample, these are probably degradation products of pIII. The patterns for the library produced from ER2738, TG1 and XL-1 host cells are similar. Peptides higher than wild type pIII were detected as a smear with distinct bands at 65 kDa and 80 kDa. Hyperphage was used only for the first panning round in the phage display experiment because the infectivity of recombinant Hyperphage is lower than the one of M13K07 (Rondot *et al.* (2001)). In the first panning round the probability that a low abundant patient antibody finds its antigen is low. This can be improved by using Hyperphage because all five pIII peptides are occupied with a recombinant peptide. Multivalent display of

antigens amplifies specific enrichment through avidity effects. However in biopanning selection round 2-4 helper phage M13K07 was used to prevent a loss of diversity due decreased phage infectivity.

4.3 Coupling of patient antibodies to magnetic beads

The patient antibodies used as targets for the phage libraries were coupled to magnetic beads as solid phase in biopanning. The beads were 1 µm in size and surface activated with a Tosyl-group. Serum IgA and serum IgG from blood serum were used separately as targets during biopanning. Therefore one fraction of beads was coupled to goat antibodies directed against human IgA and another fraction was coupled to rabbit anti-human IgG. The amino groups of these primary antibodies react with the Tosyl-group on the bead surface. The resulting covalent bond is relatively stable and assures that antibodies are not lost during extensive washing procedures in biopanning. A third fraction of primary antibodies was coupled to anti-mouse antibodies. GAPDH, Ubiquitin and Tubulin antibodies of mouse origin were used to monitor the enrichment process during biopanning. These housekeeping genes should be ubiquitously present in the cDNA library. Therefore it should be possible to enrich them with their corresponding antibodies. The success of the coupling reaction to the beads was tested qualitatively on a western blot. Beads were loaded with serum antibodies and washed thoroughly. Upon boiling with SDS-buffer (including, DTT and ß-Mercaptoethanol) antibodies were reductively cleaved from the beads. The antibodies were separated on a SDS-PAGE and detected with anti-human IgG and anti-mouse IgG.

Results from western Blot analysis demonstrate that the coupling of the primary as well as the binding of serum antibodies to them was successful (Figure 18). Although same quantities of antibodies were used, human IgG results in a stronger band than human IgA. This is likely a result of different secondary antibodies used for detection. In summary it was confirmed that beads were coupled to the target antibodies and phages presented recombinant peptides. This made it possible to let them interact and gradually select for the best binders in the biopanning procedure.

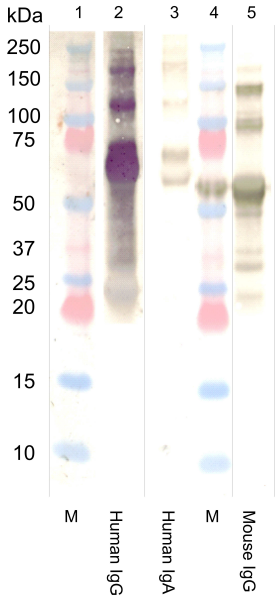

Figure 18: Immunoblot demonstrates successful coupling of antibodies to magnetic beads.

The Figure shows a western Blot (4-12% PAA gradient gel electrophoresis) of human serum antibodies cleaved from magnetic beads. Grey lines were used to indicate where the western blot was cut for the use of different detection antibodies. Lane 1 and 4 show molecular markers. In lane 2 IgG antibodies from human serum can be seen and in lane 3 IgA antibodies. In lane 5 mouse IgA was detected. The heavy chain is visible between 50 and 75 kDa for human IgG and human IgA. It is slightly smaller for mouse IgG. The light chain is detected at approximately 20 kDa and not detectable for human IgA.

4.4 Performance of enriched binders from phage display in ELISA

To monitor gradual increase of binding affinity over the course of several rounds of phage display, enriched binders from phage display were analyzed in an ELISA experiment. The enriched phage library was bound to serum antibodies and immobilized on magnetic beads. Washing thoroughly guaranteed that only binders to serum antibodies were retained. To determine the background binding affinity, the enriched phages were bound to beads that were coupled only to the primary anti-human antibodies and not loaded with serum antibodies. A secondary antibody coupled to HRP and directed against M13 bacteriophages was used to detect all binders that remain after washing. Addition of ABTS and hydrogen peroxide to ELISA samples initiated a colorimetric reaction. HRP was quantitatively converted to a green reaction

product that was measured at 405 nm. The binding capacities of enriched peptides to serum antibodies were measured after each panning round. Enrichment profiles that emerged after four phage display panning rounds for each sample could then be compared. Samples are defined by host strain, antibody class and patient serum number unequivocally. Increasing absorption values in subsequent panning rounds indicate an enrichment of binders. Evaluation of the enrichment profiles revealed that different enrichment patterns can be observed (compare Appendix **??**). To test the hypothesis that enrichment profiles are specific for the different *E.coli* host strains the data were clustered. The manual analysis of data is always influenced by a variety of cognitive biases like observer-expectancy effects or the clustering illusion effect. Humans tend to see significant clusters in small samples from random distributions and the variability in samples is generally underestimated. With the help of a computing algorithm these effects can be minimized. Affinity propagation clustering is perfectly suited for the unbiased identification of clusters in a dataset. This method has been successfully proven to solve problems like identifying clusters of faces, semantic entities or segments of DNA that reflect expression properties of genes.

The most representative member of a cluster is called exemplar in Affinity Propagation Clustering. In the beginning all samples are considered to be potential examplars and by iteratively exchanging real-valued messages between the samples the clusters emerge gradually. As input data the algorithm needs to be supplied with similarity values. For clustering ELISA data the similarity values were calculated based on the relative enrichment between consecutive panning rounds. To avoid the problem that very high values are given too much weight the enrichment factors were logarithmized. The three enrichment factors were used to calculate a characteristic similarity value. This value is best described as point in a three-dimensional coordinate system with the enrichment factors as axis (compare Figure 20).

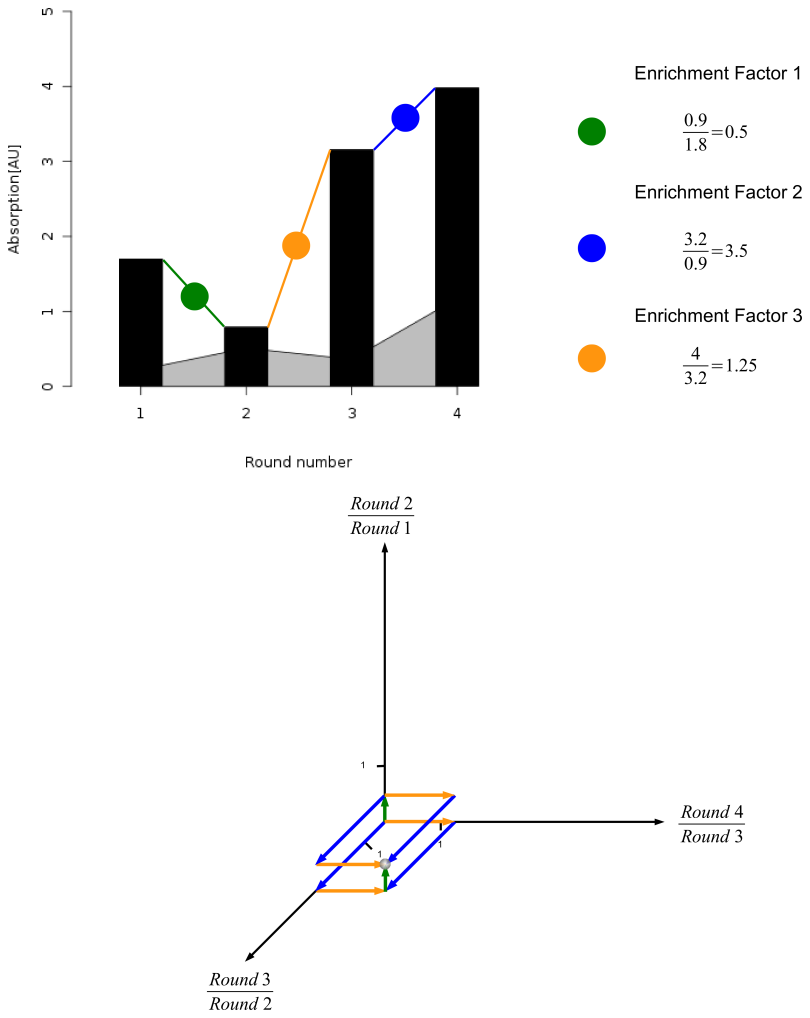

Figure 19: Main principle of calculating similarity values for affinity propagation clustering.
The upper panel shows the measured ELISA values for an exemplary phage display sample to illustrate how enrichment factors are calculated. Enrichment factors are depicted in green, orange and blue circles and are calculated as ration between two panning rounds. These values can then be plotted in a three dimensional coordinate system (lower panel) to obtain a characteristic similarity value that serves as input data for the affinity propagation cluster algorithm.

By calculating the negative Euclidian distances between the enrichment values for all three pairs of consecutive rounds (Figure 19), similarity values were obtained for each two-way combination of samples. Samples with similar enrichment patterns were clustered around an exemplar.

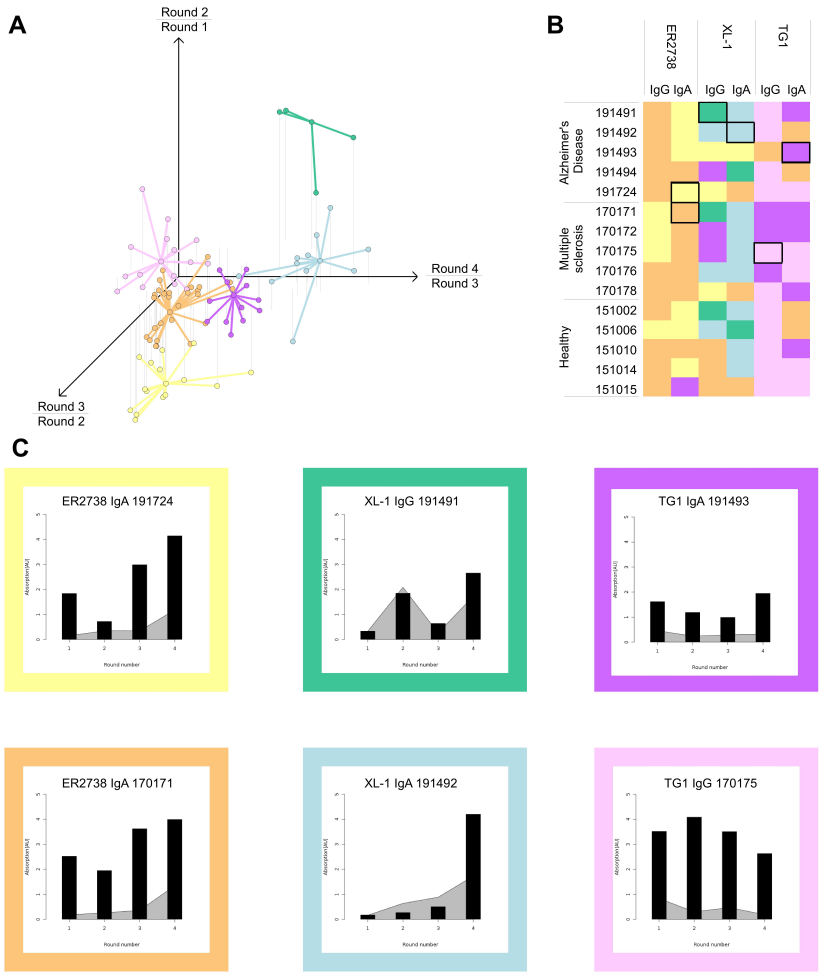

Figure 20: *E.coli* strains show characteristic enrichment patterns in ELISA.

Elisa profiles are classified by enrichment factors using Affinity Propagation Clustering. Negative Euclidian distances of log-transformed enrichment factors (A) were used as similarities. Colors represent clusters. (B) Cluster distribution among samples. Exemplars are indicated by black boxes. ELISA profiles for the exemplars are shown in (C). Black bars are the relative absorbance over four panning rounds, grey areas indicate background binding activity in control experiments without patient serum.

The Hypothesis that phage display with different host strains results in characteristic enrichment patterns, is supported by the affinity propagation clustering analysis (Figure 20). Samples from the same cluster belong mainly to one strain while antibody class or diseases do not seem to play a role (Figure 20 B). The exemplars (Figure 20 C) show characteristic enrichment patterns. Typical for ER2738 is a relatively high absorption value in the

first panning round that decreases in the second panning round and increases again in round 3 and 4 (see Figure 20 C, yellow and orange boxes). In XL-1 blue absorption values stay low till the fourth round when binders increase erratic (Figure 20 C, turquoise box). Another cluster for XL-1 shows constant low enrichment factors for all samples, although values rise in round 2 and 4. High background in these samples indicates that this is caused from unspecific binders (Figure 20 C, blue box). The TG1 characteristic patterns are either relatively low values slightly increasing in round 4 (Figure 20 C, purple box) or high values slightly decreasing in round 4 (Figure 20 C, pink box).

In summary results from ELISA show that binders could be enriched in most of the samples. More specific properties of the binders can be revealed by sequencing them. Different enrichment patterns depending on which host strain was used, confirm the hypothesis that the choice of host strain influences phage display experiments. One and the same library transformed in three different host strains results in different enrichment patterns after panning.

4.5 Enrichment of housekeeping genes

In an effort of trying to learn which of the three E. coli strains employed in this study is best suited for phage display, at least in combination with the system used here, four rounds of biopanning with anti-Tubulin antibodies have been analyzed quantitatively by qPCR. It can be expected that in every subsequent panning round the fraction of Tubulin in the sample increases. Quantitative real time PCR is a precise and sensitive method to determine the concentration of Tubulin in each sample. It is of interest if enrichment of Tubulin differs in the three phage display host strains XL-1 blue, ER2738, TG1 used in this experiment. To determine the concentration of Tubulin in the samples standard curves of known concentrations were used to calculate the amount of Tubulin and vector backbone in each sample (Figure 21). One primer pair was designed to bind on the pJuFo backbone and was used to measure the overall plasmid concentration in a sample. Another primer pair was designed to amplify the Tubulin isoform A1A, which was confirmed to be present in the initial cDNA library by sequencing. Primers were designed to amplify a fragment in the 3'end region of the Tubulin CDS to make sure that even relatively short Tubulin inserts would be measured. A standard curve was made on the basis of a dilution series for a Tubulin ORF clone. Another standard curve was made for the pJuFo backbone.

These samples served as references with known concentrations. Derived standard curves were used to convert measured fluorescent differences into concentrations. To calculate the relative amount of plasmid with Tubulin

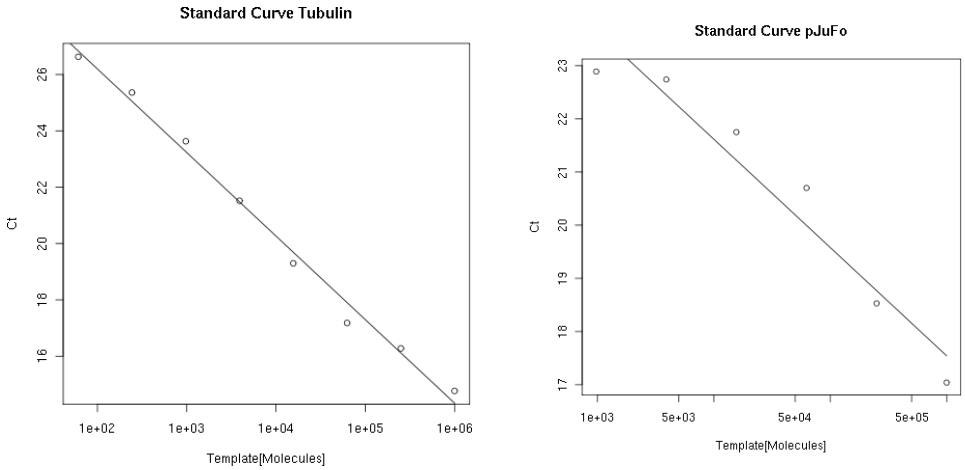

Figure 21: Ct values behave log-linear over a broad range of concentrations.

Standard curve for tubulin qPCR are shown. Ct values for known amounts of tubulin (left) and pJuFo backbone (right) DNA are plotted against the respective amounts. The values are best fitted by a logarithmic curve. The formulas and coefficients of determination are shown in the upper right corners.

as inserts from overall plasmids in a sample the quotient of their respective Tubulin and vector backbone amounts was calculated.

Enrichment factors were defined as quotients of each samples, the relative amount of Tubulin and the relative amount of Tubulin in the respective non-enriched library sample. Figure 22 shows that there is no enrichment of tubulin in any of the samples. In fact, most samples show no tubulin-specific PCR at all. This indicates that the antibody either did not bind the tubulin fragment contained in the non-enriched samples, because they are too short for example, or it did bind more strongly to some other peptide contained in the library, perhaps another tubulin isoform that the PCR primers do not bind to.

Results from qPCR show that there is no enrichment of tubulin at all. Enrichment factors are below 10% of the initial library concentration. This drop of Tubulin concentration is contrary to the expectations. In a successful enrichment the enrichment factors should exceed the initial library concetration in all four rounds. Plausible concentrations were measured for standard curves and initial libraries. This result indicates that no enrichment of Tubulin took place during panning. All three housekeeping genes used in phage display, GAPDH, Ubiquitin and Tubulin were sequenced to see if quantification with qPCR is indicated for one of these genes.

Results from sequencing demonstrate that none of the enriched genes is

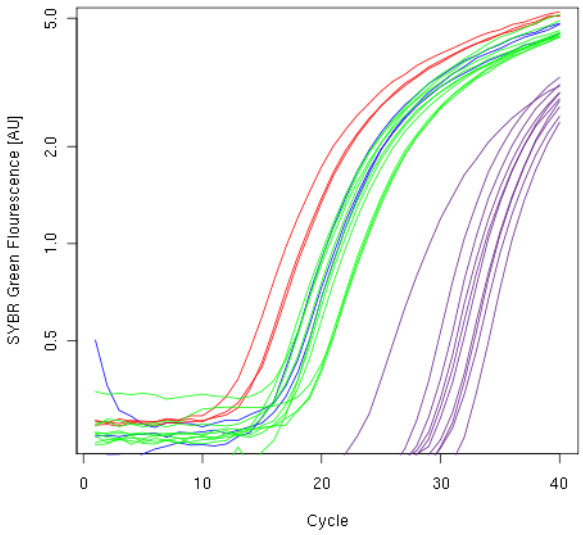

Figure 22: Tubulin is present in the brain cDNA library but not in enriched libraries.

Amounts of tubulin-specific PCR product as measured by SYBR Green fluorescence is plotted against 40 cycles of qPCR. Red curves are tubulin-specific products of brain cDNA library samples used as input for the phage display experiments, blue curves are pJuFo backbone-specific primers. Purple and green curves show the same for panning rounds 1-4, respectively.

the intended housekeeping gene. Five clones per strain and housekeeping gene is by no means statistically significant but makes a substantial enrichment seem implausible. It is striking however that enriched genes differ when panned with the different host strains. For example panning with antibodies against GAPDH resulted in the enrichment of Adenylate Kinase 2 (AK2) from using ER2738 as host strain, ATP Synthase 6 (ATP6) from XL-1 and BEX2 from TG1 (compare Table 8).

It is not known against which epitope the antibodies used in this experiment are directed. It is well possible that an epitope in the 5'region of the CDS is recognized. This would render enrichment impossible. An interesting question is if the enriched binders are mimotopes for their respective housekeeping proteins. No evident similarity could be found by comparing these proteins. However, the identification of mimotopes is not trivial and would require knowledge of the antibody epitopes used in panning.

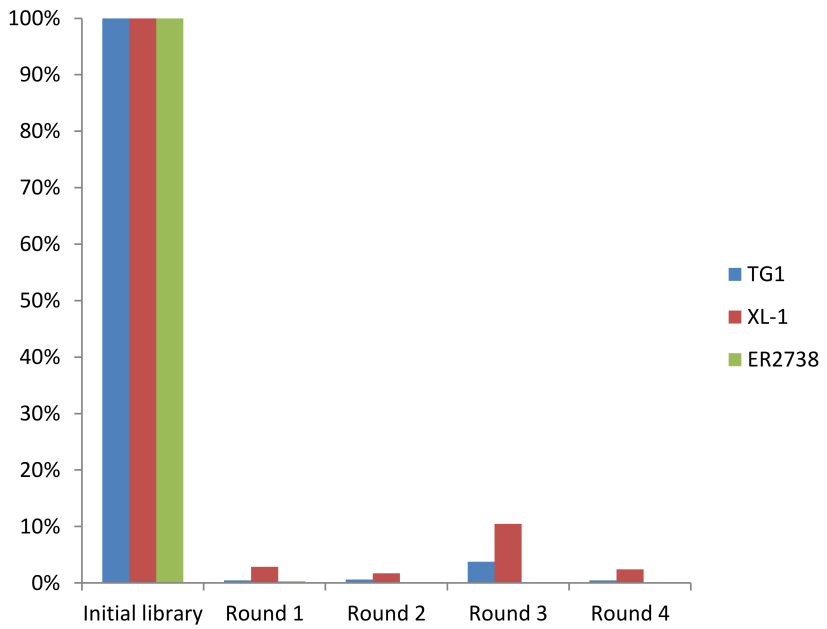

Figure 23: Enrichment factors for Tubulin show a decrease in gene concentration.

Tubulin concentrations in in initial libraries were set to 100%. Blue bars are enrichment factors for TG1, red bars for XL-1 blue and green bars ER2738 phage display host strains.

Table 8: Sequences of three housekeeping genes (HGNC Symbols) Tubulin, GAPDH, ER2738 from phage display panning in host strains TG1, XL-1 and ER2738.

GAPDH ER2738	Tubulin ER2738	Ubiquitin ER2738	GAPDH XL-1	Tubulin XL-1	Ubiquitin XL-1	GAPDH TG1	Tubulin TG1	Ubiquitin TG1
AK2	DPP7	ZHX1	ATP6	Mt DNA	TAGLN	RPL15	-	BEX2
AK2	DPP7	ZHX1	ATP6	ZFIVE20	TOPORS	RPL15	-	BEX2
AK2	DPP7	ZHX1	ATP6	ZFIVE20	TAGLN	RPL15	-	BEX2
AK2	DPP7	ZHX1	ATP6	Mt DNA	TAGLN	RPL15	CUL4B	BEX2
AK2	DPP7	ZHX1	ATP6	SRPR	TAGLN	RPL15	RAB3A	BEX2

4.6 454 Sequencing and Characterization of Enriched Binders

Analysis of phage display hits on nucleotide level

A subset of enriched libraries from selection round four was chosen to be sequenced and analyzed in detail. Serum 191492 from a patient with Alzheimer's disease, serum 170171 from a patient with multiple sclerosis and serum 151010 from a healthy donor were of special interest because they are under investigation in our research group. These sera were used in a similar phage display experiment based on the use of full-ORF libraries and were furthermore tested in screenings on protein macroarrays (Georgieva, personal communication). In this study the IgA and IgG immunoglobulin fractions were used as targets to isolate peptide antigens with Phage Display. As host strains for phage expression the *E.coli* strains XL-1, TG1 and ER2738 were used. To identify the optimal host strain and to see if both antibody classes work well in phage display they were studied in all combinations. Thus six samples per patient and a total of 18 samples sequenced on a 454-FLX sequencing platform (Roche). Another 208 samples based on four patients with MS, four Alzheimer's patients and four healthy controls were prepared in this work as well and are ready for sequencing. A total of 117.223 reads were obtained all samples taken together. The reads were assembled to contigs (see Section 4.1). In Table 9 numbers of reads and contigs for each sample are summarized.

Table 9: Reads from enriched binder samples were assembled into contigs.
The number of reads and the number of contigs is shown for each sample.

		Alzheimer's Serum 191492		Multiple sclerosis Serum 170171		Healthy Serum 151010	
Host strain	Class	Reads	Contigs	Reads	Contigs	Reads	Contigs
TG1	IgG	757	16	1075	20	827	115
	IgA	8926	66	7034	54	7350	40
XL-1 blue	IgG	8023	72	951	18	8926	11
	IgA	10822	83	7654	115	11000	98
ER2738	IgG	1880	36	2418	25	14851	23
	IgA	7388	108	8289	88	9052	475

The comparison of reads and contigs in different samples shows that the number of reads obtained from sequencing ranges from 757 to 14851. These fluctuations are not intended but are inherent in the second generation sequencing approach because number of sequences derived from one run

cannot always be precisely predicted. The number of contigs that could be assembled does not seem to correlate with the number of reads. After removing vector sequences and polyA tails, reads and contigs were mapped separately to a human RefSeq RNA database using BLASTn.

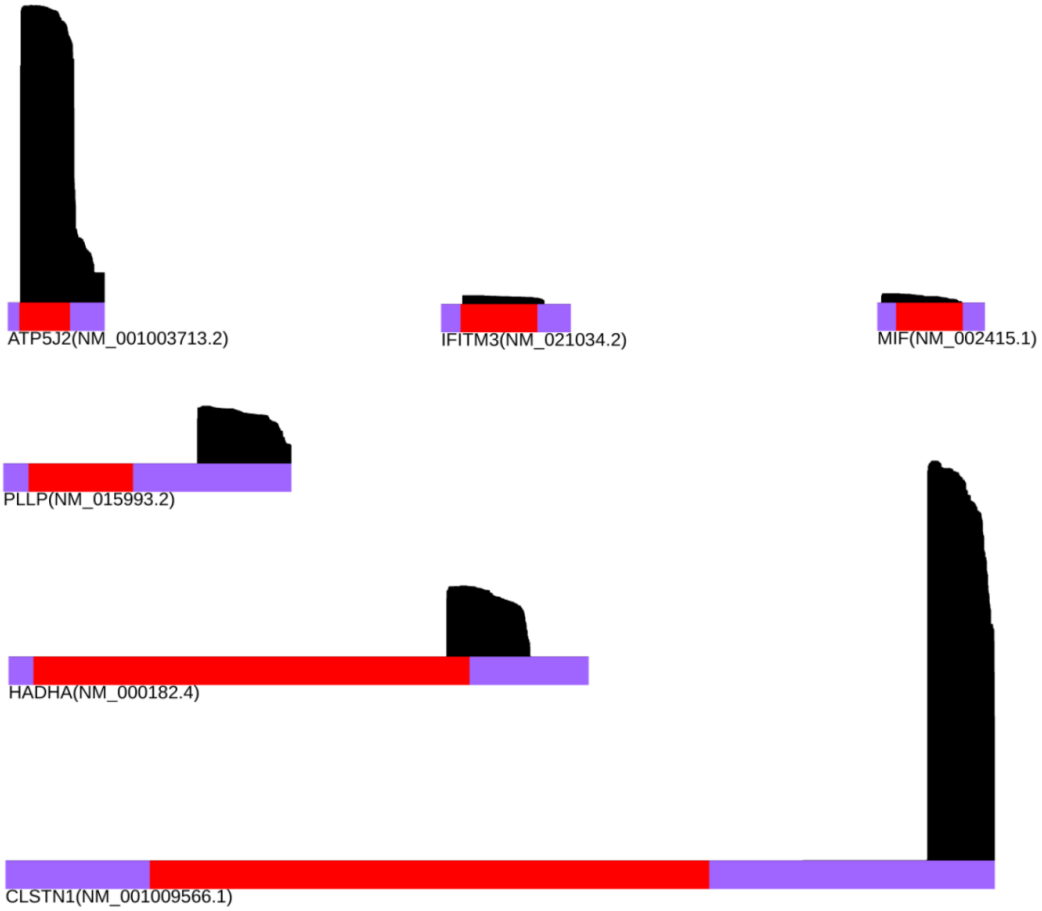

Figure 24: The mapping of enriched sequences to human RefSeq shows that in most of the cases only one construct per gene was sequenced. Exemplary hits were selected to show the match on their respective mRNA. The ENTREZ Symbols and RefSeq identifiers are written below the hits. Untranslated regions are depicted in purple and coding sequence in red. Each 454-sequencing read is represented by a one pixel line. Black regions thus represent the sequencing coverage and relative position on mRNA.

The mapped reads that matched to human RefSeq RNA were classified into two categories, hits that only match to the 3'UTR region and hits that match to the CDS region. Reads that match at least partially to the CDS of mappable human mRNA were potentially disease relevant peptides. Therefore

the literature was analyzed for these hits. Hits that match only to 3'UTR were considered to be synthetic peptides. Overall 105 binders from sequenced phage display samples could be mapped to RefSeq RNA. Binders that occur in more than one sample were counted seperately for this purpose. Of these hits 46 were derived from samples with IgG as target and 59 from IgA. Interestingly these hits are complementary. Of the 82 different genes that were found in total, 38 and 53 were found in IgG and IgA samples exclusively, whereas only 9 genes genes were found in both. In most other biomarker screens the IgA subclass is not included. However results of this study indicate that IgA immunoglobulins should not be disregarded in future experiments.

For multiple sclerosis 12 hits were found to match CDS and 22 hits matched only to 3'UTR (compare Appendix 6 and Table 10). The ATPase synthase f subunit ATP5J2, two Humanin isoforms (MTNR2L2, MTNR2L2) and Transgelin (TAGLN) were also found in Healthy samples and are therefore not suitable as biomarker candidates. The Humanin isoform MTRNL8 is abundant in the initial library before enrichment and can be found in Alzheimer's samples as well. Therefore Humanin is probably an unspecific off-target binder. ATP5J2 is the subunit f of the mitochondrial ATP-synthase complex and plays a key role in maintaining cellular energy levels. An increased energy demand as response to demyelinating events and impaired nerve impulse conduction affects neuron integrity and is likely to contribute to MS pathology (reviewed in Campell and Mahad 2010). Even though overall mitochondrial content is increased in demyelinated axons, ATP5J along with other electron transport chain gene transcripts are decreased in MS motor cortex (Dutta *et al.* (2006); Zambonin *et al.* (2011)).

Another mitochondrial protein found is an iron sulfur subunit of respiratory complex II encoded by the SDHB gene. An accumulation of iron in the central nervous system is often observed in multiple sclerosis patients. The underlying mechanism is unknown. It can be speculated that iron containing catalytic subunits, for example SDHB, might be involved in this process (reviewed in Williams *et al.* (2011)). Decreased activity of the 20S proteasome and a possible role of the 26S proteasome in multiple sclerosis were recently reported (Zheng & Bizzozero (2011)). PSMC6, the ATPase base subunit of the 26S Proteasome was highly enriched in MS samples.

Infection with Epstein-Barr virus (EBV) as a strong risk factor for the development of multiple sclerosis is a well-established hypothesis. The exact mechanism is not well understood but antibodies against EBV-antigens can be found in the blood of multiple sclerosis patients (reviewed in Ascherio & Munger (2010)). RABAC1 (also known as PRA1) directly interacts with the Epstein-Barr virus protein BHRF1. Strikingly, BHRF1 is a structural and functional homologue of anti-apoptotic BCL-2 (Li *et al.* (2001)).

The subunit DCTN1 has been involved in axonal transport defects in a model of multiple sclerosis, suggesting an indirect role in disease progression (Kreutzer *et al.* (2011)).

NF-κB signaling plays a key role in the pathology of autoimmune diseases and is a potential therapeutic target in multiple sclerosis (Yan & Greer (2008)). The LTSZ gene identified in this study is negatively regulated by transcription factor NF-κB in glioma cells (Cho *et al.* (2008)). It also regulates the nuclear transport of ß-catenin induced by NF-κB (Thyssen *et al.* (2006)). Ribosomal protein 7 is a known auto-antigen in patients suffering from autoimmune diseases (Mikecz et al., 1995; Donauer et al. 1999). Disease-specific epitope-recognition patterns of anti L7 autoantibodies could be identified in patients with systemic lupus erythematous, systemic sclerosis and rheumatoid arthritis (Neu *et al.* (1997)). The usefulness of RPL-7 in diagnosing multiple sclerosis remains to be elucidated.

No involvement in multiple sclerosis-relevant mechanisms was found for SPATS2 and RPS26. However, this does not disqualify them as biomarker candidates and further experiments are necessary to investigate seroreactivity of all enriched peptides.

Table 10: Multiple sclerosis hits. Reads were mapped to human RefSeq RNA using BLAST.
For each gene, the number of reads in the enriched samples (#Round 4) and the non-enriched libraries (#Library) is shown, as well as whether the gene has been found in the healthy controls. The antibody class (AB class) used in biopanning is also indicated. A short description of the proteins functional relevance has been given with a reference in the form of a PubmedID.

Entrez Symbol	#Library	#Round4	Healthy	AB class	Short description	PubmedID
ATP5J2	2	3760		IgA	ATP synthase subunit f, mitochondrial enzyme	12471886
DCTN1	0	378		IgG	Involved in many neurodegenerative diseases	21420428
LZTS2	0	3		IgA	Regulates the Nuclear Export of ß-Catenin	17000760
MTRNR2L2	30	161	X	IgA+IgG	Isoform of neuroprotective Humanin	19477263
MTRNR2L1	9	20	X	IgA+IgG	Isoform of neuroprotective Humanin	19477263
MTRNR2L8	65	159		IgA+IgG	Isoform of neuroprotective Humanin	19477263
PSMC6	0	733		IgA	ATPase base subunit of the 26S Proteasome	9473509
RABAC1	0	2		IgG	Interaction with anti-apoptotic EBV protein	11373297
RPL21	5	31		IgA	L21E family of ribosomal proteins	8332502
RPL7	24	41		IgA	Autoantigen in autoimmune diseases	7856252
RPS26	2	11		IgA	S26E family of ribosomal proteins	10938081
SDHB	0	62		IgA	Iron-sulfur subunit of respiratory complex II	21446022
SPATS2	0	2		IgA	Role in pancreatic beta-cell function	20119533
TAGLN	2	3570	X	IgA+IgG	Actin binding protein, cancer	18378184

Table 11: Alzheimers disease hits. Reads were mapped to human RefSeqRNA using BLAST.
For each gene, the number of reads in the enriched samples (#Round 4) and the non-enriched libraries (#Library) is shown, as well as whether the gene has been found in the healthy controls. The antibody class (AB class) used in biopanning is also indicated. A short description of the proteins functional relevance has been given with a reference in the form of a PubmedID.

Entrez Symbol	#Library	#Round4	Healthy	AB class	Short description	Pubmed ID
DDIT3		6		IgG	Endoplasmic reticulum stress response	14685163
FAM195B		264		IgA	Unknown function	
GNAS	2	16		IgG	Stimulatory G-protein subunit, imprinting	19407493
HADHA		409		IgA	Long-chain fatty acids metabolism	10816122
HBB	2	13		IgA	ß-chain of hemoglobin	
IFITM3	5	52		IgA	Interferon-inducible transmembrane protein	21166591
LOC 100134397		38		IgG	Unknown function	
LOC 100291917		72		IgG	Hypothetical protein LOC100291917	
LOC 642131		38		IgG	Hypothetical protein LOC 642131	
LYRM4		479		IgA	Essential component of iron-sulfur cluster biogenesis	12947415
MIF		70		IgA	Increased levels in CSF of AD patients	20200619
MTRNR2L1	9	654	X	IgG+IgA	Isoform of neuroprotective Humanin	19477263
MTRNR2L2	30	338	X	IgG+IgA	Isoform of neuroprotective Humanin	19477263
MTRNR2L8	65	158		IgG+IgA	Isoform of neuroprotective Humanin	19477263
MVP		92		IgG	Signal transduction and immune defense	18759128
PRSS2	4	2		IgG	Trypsin 2, associated with pancreatitis	21410075
RABAC1		10		IgG	Lipid transport and cell migration	20592422
TIAF1		83		IgA	TGF-beta-1-induced antiapoptotic factor	21368882
TMSB10	4	53	X	IgA	Regulates actin cytoskeleton organization	16249193
TUBB2A		849		IgA	Influences microtubule polymerization and Tau	16006664

In Alzheimer's disease samples, enriched after four rounds of biopanning, 14 hits matched to CDS and 17 only to 3'UTR (see Table 11 and Appendix 6). Among the enriched binders containing CDS one antigen with known direct role in Alzheimer's disease pathology was identified. TGF-ß1 induced antiapoptotic factor (TIAF1) produces aggregates that lead to caspase-dependent apoptosis. Lee et al. could show that polymerized TIAF1 directly interacts with amyloid fibrils. It also leads to the dephosphorylation of amyloid precursor protein Lee *et al.* (2010). This event is known to induce the cascade that leads to the formation of Amyloid beta and disease-characteristic plaques.

Another interesting biomarker candidate is the macrophage inhibitory factor (MIF). This immune modulator was confirmed to be associated with beta amyloid plaques by immunolabeling of glial cells from APP mice and human neuronal cell lines (Bacher *et al.* (2010)). Highly relevant for this study is the fact that significantly higher levels of MIF could be identified in the CSF of Alzheimer's patients compared to age-matched controls (Craig-Schapiro *et al.* (2011); Popp *et al.* (2009)).

Beta tubulin (TUBBA2) is abundant in brain tissue and contributes to the formation of hyper-phosphorylated tau in neurofibrillary tangles (Puig *et al.* (2005)). Monomers of the alpha and beta chain of tubulin are cross-linked by microtubule-associated tau protein and directly interact with amyloid precursor protein as well (Chau *et al.* (1998); Islam & Levy (1997)). Hyperphosphorylation of microtubule-associated tau and the accumulation of ß-amyloid isoform 42 are well-documented phenomena in the development of Alzheimer's pathology (reviewed O'Brien & Wong (2011)).

A role of Beta globin (HBB) in Alzheimer's disease has not been shown yet. However, the hemoglobin alpha chain could be identified as binding to ß-Amyloid together with MIF, ß-Tubulin, GAPDH, serum albumin, and glutamin synthetase in a complex (Oyama *et al.* (2000)).

DDIT3 (also known as CHOP/GADD153) is a pro-apoptotic transcription factor involved in endoplasmic reticulum stress and can be associated with Alzheimer's disease development (reviewed in Oyadomari & Mori (2004)). Mutations in Presenilin 1 increase DDIT3 protein translation and decrease levels of antiapoptotic Bcl-2 (Milhavet *et al.* (2002)). Mutations in presenilin are the most common cause of familial early onset Alzheimer's (reviewed in Brouwers *et al.* (2008)). A recent study demonstrates a reverse of Alzheimer pathology hallmarks in an animal model after silencing DDIT3 (Prasanthi *et al.* (2011)).

RABAC1 was found in MS samples as well but not in healthy samples, this might point to more general role in neurodegenerative processes but no direct link to Alzheimer's could be found. Major vault protein (MVP) has been associated with drug resistance and plays a role in nucleo-cytoplasmic transport (Yu *et al.* (2002)). It was shown to interact with PTEN, a dual protein tyrosin and lipid phosphatase that is discussed to play a role in Alzheimer's disease in several studies (Sonoda *et al.* (2010); Rickle *et al.* (2006); Kwak *et al.* (2010)).

Table 12: Healthy Controls. Reads were mapped to human RefSeqRNA using BLAST.
For each gene, the number of reads in the enriched samples (#Round 4) and the non-enriched libraries (#Library) is shown. The antibody class (AB class) used for biopanning is also indicated. A short description of the proteins functional relevance has been given with a reference in the form of a PubmedID.

Entrez Symbol	#Library	#Round4	AB class	Short description	Pubmed ID
C10orf116	28	49	IgA	Unknown Function	
CPE	2	102	IgG	Exopeptidase, activate neuropeptides	11373297
DYNLRB1	2	44	IgA	Part of the Dynein Motor complex	19935668
GAPDH	8	43	IgA	Glycolysis	20727968
GRM3	2	3083	IgA	Glutamate Neurotransmission, Schizophrenia	18541626
METRN	0	9	IgG	Neurotrophic factor	19259827
MTRNR2L1	9	81	IgA+IgG	Isoform of neuroprotective Humanin	19477263
MTRNR2L2	30	3	IgA+IgG	Isoform of neuroprotective Humanin	19477263
PPP1R14A	0	6	IgG	Signal Transduction, Cancer	19846560
TAGLN	2	1160	IgA+IgG	Actin binding, Cancer	18378184
TMSB10	4	3	IgA	Regulates actin cytoskeleton organization	16249193

In Healthy samples, enriched after four rounds of biopanning, 11 hits matched to CDS and 10 only to 3'UTR (see Table 12 and Appendix 6). MTRNR2L1, MTRNR2L2, TAGLN and TMSB10 were found in disease samples as well. Unspecific binding, for example to the surface of 96-well polystyrene plates is not unusual in phage display and disqualifies these hits as biomarkers. The other hits were analyzed for their function in Alzheimer's disease and multiple sclerosis. Direct disease relevance for Alzheimer's could be only found for GAPDH (reviewed in Butterfield *et al.* (2010)). Furthermore DYNLRB1 is a subunit of the Dynein motor complex, a protein involved in various neurodegenerative diseases, included Alzheimer's disease (reviewed in Eschbach & Dupuis (2011)). PPP1R14A is a subunit of protein phosphatase 1 and plays an important role in signal transduction pathways. CPE and GRM3 play a role in neurological disorders but not specifically in MS or AD. Literature analysis of hits from Healthy samples includes few hits that can be associated specifically with AD and none with MS.

An additional interesting result from analysing enriched library samples is that IgA and IgG reveal different hits. Humanin isoforms (MTRNR) and Transgelin (TAGLN) were found in samples based on IgG and IgA, further supporting that these are unspecific binders.

Taken together, results from mapping sequences from enriched library samples show several interesting hits that could be associated with well-established disease mechanisms. This does not qualify them necessarily as the best biomarkers but indicates that the chosen phage display-based approach was successful in enriching disease-relevant binders. The largest enrichments were achieved for synthetic proteins. Notably, many of the enriched genes were not sequenced in the initial cDNA library. This underlines the possibility of enriching even low abundant proteins from the phage pool and indicates that the initial library was much bigger than the sequenced fraction.

4.6.1 Frame Analysis of Phage Display hits

Phage display hits were translated and in frame peptides subjected to a protein BLAST. Hits that match to a SwissProt database query with an identity of more than 80% and a score higher than 50 were considered to be expressed peptides.

Table 13: Overview of phage display hits matching SwissProt proteins. Inserts from the enriched samples have been translated to obtain the sequence of displayed peptides and mapped against the human SwissProt database entries using BLAST. For each UniprotID the number of reads in the Alzheimers disease (AD), multiple sclerosis (MS) and healthy control (H) samples is shown, as is the number of amino acid residues of the longest contig (Peptide Length). The strain is given as indicated as 1 for ER2738, 2 for TG1 and 3 for XL-1.

UniprotKD ID	ENTREZ Symbol	Strain	AZ	MS	H	Peptide Length
ATPK_HUMAN	ATP5J2	1		1672		125
CBPE_HUMAN	CPE	1			102	26
DCTN1_HUMAN	DCTN1	1		378		126
DHSB_HUMAN	SDHB	2		62		129
DLRB1_HUMAN	DLRB1	1			44	79
G3P_HUMAN	GAPDH	1			43	36
ECHA_HUMAN	HDHA	1	390			38
GRM3_HUMAN	GRM3	1			2173	72
HMN8_HUMAN	HMN8	3	7			30
HUNIN_HUMAN	HUNIN	1,3		26	47	16
IFM3_HUMAN	IFITM3	1	52			129
LYRM4_HUMAN	LYRM4	1	477			91
LZTS2_HUMAN	LZTS2	1		3		44
METRN_HUMAN	METRN	1			9	147
MIF_HUMAN	MIF	1	68			137
MVP_HUMAN	MVP	2	92			160
NU1M_HUMAN	MT-ND1	1	2196			45
PPP14A_HUMAN	PPP14A	1			12	136
PRS10_HUMAN	PRS10	2		676		53
GRM3_HUMAN	GRM3	1			1293	102
RL7_HUMAN	RPL7	1		41		113
RS26_HUMAN	RS26	1		11		29
SPAS2_HUMAN	SPAS2	3		2		69
TAF5_HUMAN	TAF5	2		3		14
TAGL_HUMAN	TAGLN	3		3386	1166	41
TBB2B_HUMAN	TUBB2B	1	101			14
TIAF1_HUMAN	TIAF1	3	249			41
TRY2_HUMAN	PRSS2	1	2			47
TYB10_HUMAN	TMSB10	1	53			64

Results from analysis of enriched phage display binders on protein level (Table 13) reveal fewer hits than found in nucleotide BLAST. This is expected because only hits that contain CDS are likely to be identified on protein level.

It is conceivable that 3'UTR containing mRNAs are translated into a familiar human protein and expressed on bacteriophages although this scenario is less likely. For example the synthetic peptide PLLP matches to TAF5 and LOC100131754 to NU1M. Furthermore some of the CDS containing sequences were too short to be found or were out of frame. Sequences with 5'UTR in front of the CDS potentially had a stop codon that hampers expression.

A total of 25 proteins could be identified from protein BLAST. The size ranges from 14 to 160 amino acids; the average length is 73 amino acids. In the initial library (see Figure 15, Chapter 4.1) only three percent of the proteins were larger than 100 amino acids, in the enriched libraries 32 percent fulfill this criterion. In the initial library 42% of all mappable hits contain CDS. Hits from enriched MS and AD samples taken together have 40% of the sequences with CDS. This result shows that it was possible to enrich relatively long fragments compared to the length distribution of initial library sequences. The same fraction of sequences as in initial libraries contains CDS. On protein level 18 out of 26 MS and AD hits can be found, thus are expressed in frame. This corresponds to 69% of all hits with CDS. Compared to the expected value of about 33% (29% in the initial libraries), this result indicates that in-frame sequences are substantially enriched.

Peptides confirmed to be expressed in the right reading frame were mainly derived from using ER2738 as host strain and (Table 14 and Table 13). ER2738 revealed the highest number of binders that could be associated with a biological function. TG1 a satisfying host strain as well and revealed five peptides. The lowest number, only two peptides, were found in samples with XL-1 as host strain.

Table 14: Number of Phage display hits used in network analysis

	ER2738	TG1	XL-1
MS	5	4	1
AZ	8	1	1
Total	13	5	2

4.6.2 Network analysis of Phage Display Hits

A large part of the enriched binders contains inserts expressing meaningful human proteins. To explore the hypothesis that these proteins might have functional relevance for the respective diseases, the interaction partners for these proteins were downloaded from the ConsensusPathDB database (Kamburov *et al.* (2009)). The information was supplemented with findings from literature and analyzed in terms of relevance for the respective disease. It

is assumed that proteins that share a functional connection should be closely connected to each other and to other disease-relevant genes (Chuang *et al.* (2007)). Central nodes in this network can be explored as novel disease-relevant candidates. It can be hypothesized that expressed proteins that could be mapped to human proteins play a biological role in disease. Very often disease relevant proteins play a central role in protein networks by interconnecting many proteins with known pathological function. Even when disease relevance is not obvious for a single protein, the embedment in a network can provide the missing links. As basis for network construction protein-protein interactions from the Consensus Path Database (CPDB) were used. This database combines the data from currently 24 public resources. Only direct interactors of phage display hits that were disease-relevant or connected at least two phage display hits were used as nodes for the network. Disease relevance was decided upon literature search from available publications.

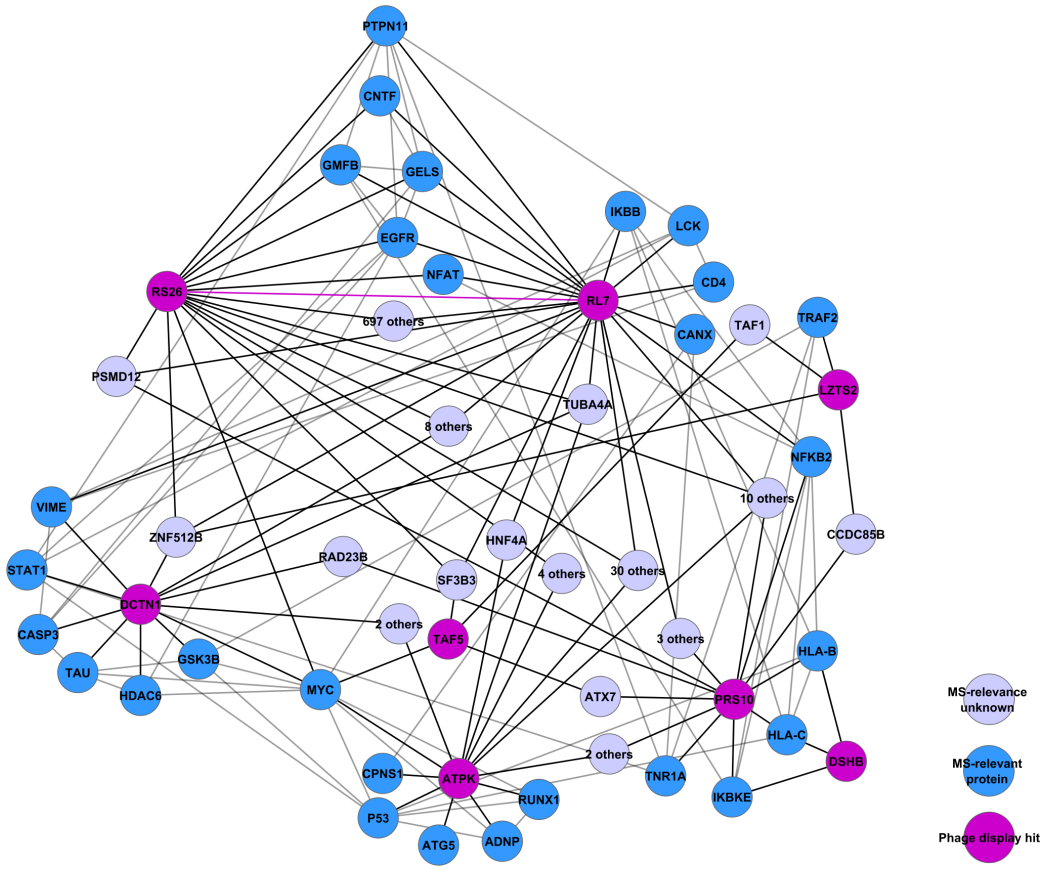

Figure 25: Multiple Sclerosis phage display hits are highly connected and closely related to known disease-relevant genes.
Protein-protein interaction network showing the phage display hits in the context of their interaction partners. Proteins are represented by nodes, links represent interactions. Purple nodes are phage display hits, dark blue nodes are proteins relevant for multiple sclerosis, and other proteins are light blue. For clarity, proteins without a known relevance in multiple sclerosis have been subsumed and links have been colored according to the number of incident phage display hits.

The protein interaction network for the enriched binders from multiple sclerosis samples (Figure 25) does indeed show some clustering of disease-relevant proteins around the enriched binders. The etiology of MS is likely to result from a combination of environmental, genetic and infectious contributors, although the underlying mechanism are not fully understood. As symptoms are very heterogeneous in patients, more than one pathway might be involved. A large body of experimental evidence is found for a T-cell mediated attack against autologous myelin in the CNS. This autoimmune response is thought to be mediated by CD4+ T-cells. CD4 is a trans-membrane protein

presented on T-cells in complex with the T cell receptor (TCR) and coreceptor CD3 and initiates a signalling cascade. CD4 is an interaction partner of RL7, one of the hits in this study.

The network nodes LCK, IKBB, NFKB2, IKBKE, PTPN11, RUNX1, MYC and NFAT, are involved in the TCR signalling pathway. NFAT forms a complex with Runx1, Ap-1 and NF-kappaB and drives expression of interleukin-2 (Hermann-Kleiter & Baier (2010)). Interleukin-2 causes T-cells to become auto-aggressive and therefore is highly relevant for MS (Gallo *et al.* (1992); Zandian *et al.* (2011)).

Cilliary neurotrophic factor (CNTF) interacts with RL7 and RS26 and is increased in MS cortical neurons where it acts as survival factor during inflammatory attacks (Dutta *et al.* (2007)). RL7 additionally interacts with tumor necrosis factor receptor TNR1A. This protein is encoded by TNFRSF1A gene, a susceptibility locus for multiple sclerosis (reviewed in Kümpfel & Hohlfeld (2009)). TNR1A acts as receptor for TNF-α and several network nodes are involved in TNF signaling. Interaction with TRADD and TRAF2 for example starts a signaling cascade that leads to NF-kappaB activation. Binding of FADD instead of TRADD on the other hand activates apoptosis.

Caspase-3 plays a central role in executing apoptosis and caspase inhibitors are discussed as potential drug targets for MS (Ahmed *et al.* (2002); Cid *et al.* (2003)). The expression of proapoptotic transcription factor p53 was found to be upregulated in oligodendrocytes of active MS lesions (Wosik *et al.* (2003)). Statin was found to interact with DCTN1 and plays a well characterized role in CD4+ T-cell differentiation and interferon signaling (reviewed in O'Shea *et al.* (2011)). More specifically STAT1 is upregulated in macrophages of MS patients and abnormal phosphorylation is thought to affect expression of interferon mediated genes (Feng *et al.* (2002); Christophi *et al.* (2009)). A very interesting protein in this context is the ATPK interactor ATG-5. This protein is required for autophagy but upon calpain cleavage turns into a proapoptotic signal (Yousefi *et al.* (2006)). ATG-5 expression is specifically increased in T-cells of MS patients and calpain has been implicated in myelin degradation as well (Shields *et al.* (1999); Alirezaei *et al.* (2009)). Recent evidence suggests that calpain is a promising drug target in MS as its inhibition might slow down T-cell proliferation (Smith *et al.* (2011)).

The proteins GSK3B, Tau, Vime, ADNP and CANX are all implicated in MS but play a role in other neurological diseases as well. GSK3 for example is important for T-cell differentiation and plays an important role in regulation of the immune response (Reviewed in Beurel (2011)). Tau is most prominent in Alzheimer's disease but abnormally phosphorylated tau plays a role in MS as well (Anderson *et al.* (2010)). Genetic risk factors in MS are mainly attributed to the MHC locus. The human leukocyte antigen

(HLA) genes located on the major histocompatibility complex have been consistently reported to predispose for MS susceptibility. HLA-C and HLA-B both interact with the phage display hits PRS10 and DSHB. HLA genes are of major importance for MS (reviewed in Svejgaard (2008)).

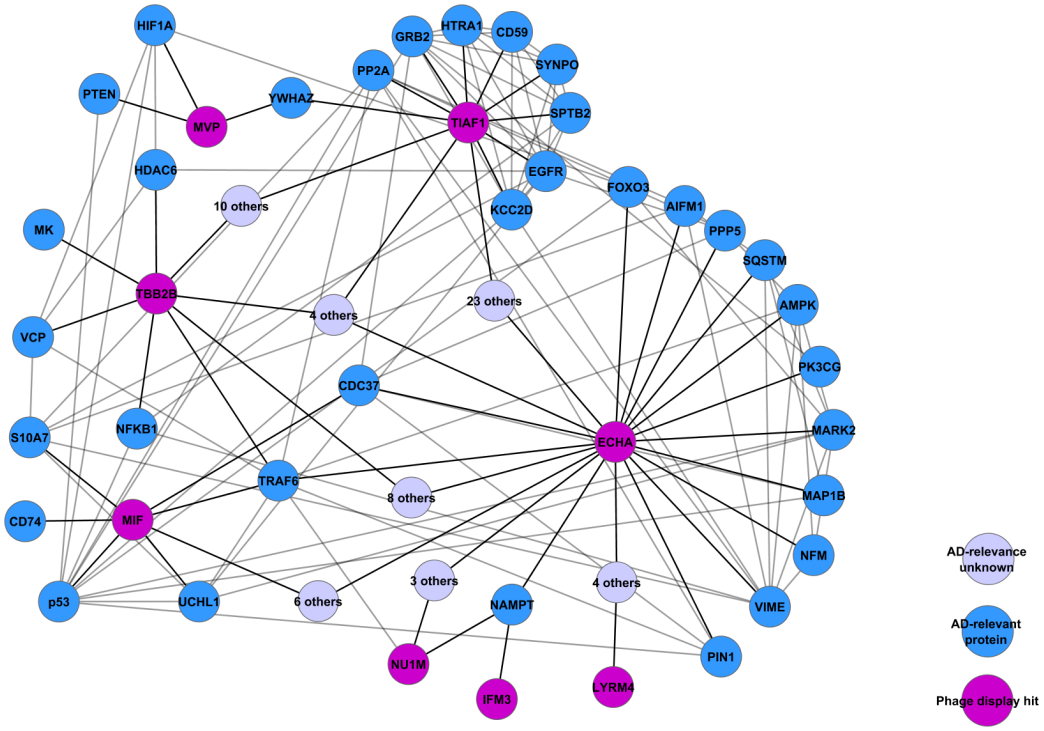

Figure 26: Alzheimer's disease phage display hits are highly connected and closely related to known disease-relevant genes.
Protein-protein interaction network showing the phage display hits in the context of their interaction partners. Proteins are represented by nodes, links represent interactions. Purple nodes are phage display hits, dark blue nodes are proteins relevant for Alzheimer's disease, and other proteins are light blue. For clarity, proteins without a known relevance in Alzheimer's disease have been subsumed and links have been colored according to the number of incident phage display hits.

The Network of Alzheimer's phage display hits shows that TIAF1, TUBB2B and ECHA are very well connected to each other and interact with numerous disease relevant proteins. For example, a highly relevant and well characterized protein in Alzheimer's disease is Pin1. The peptidyl-prolyl cis/trans isomerase Pin1 has been associated with impaired tau processing and amyloid precursor protein regulation (Pastorino *et al.* (2006); Balastik *et al.* (2007)). This finding is supported by Pin1 knockout mice that show a phenotype similar to Alzheimer's disease (Liou *et al.* (2003)).

The phage display hits LYRM4, IFM3 and NU1M on the other hand are not connected with nodes that have been implicated in disease pathology so far. NU1M also known as complex 1 of the mitochondrial respiratory chain and IFM3 both interacts with NAMPT/visfatin. A connection to AD by inducing oxidative stress is conceivable (Adams (2008)). Neurodegeneration and cognitive decline in AD is a result of massive neuronal death. This process is thought to be a result of ß-Amyloid induced neurotoxicity, inflammatory processes, complement activation and oxidative stress (reviewed in Donev *et al.* (2009)). For example complement regulatory protein CD59 was found to interact with the phage display hit TIAF1 and has been implicated in AD (Yang *et al.* (2000)).

The proteins EGFR, GRB2, FOXO3, HIF1A, PTEN, AMPK and PI3K play a role in mTor/Akt signaling, a pathway strongly associated with onset of AD pathology (Pratt *et al.* (2011); Pei & Hugon (2008); Bhaskar *et al.* (2009)). Interestingly GRB2 also interacts with amyloid precursor protein and Presenilin 1 directly. (Nizzari *et al.* (2007); Raychaudhuri & Mukhopadhyay (2010)). AMPK is thought to play an important role in AD signaling and has been shown to repress tau phosphorylation and ß-amyloid production (reviewed in Salminen *et al.* (2011)). HIF1A and FOXO3 have been associated with Alzheimer because they play a role in response to oxygen stress (Manolopoulos *et al.* (2010); Niatsetskaya *et al.* (2010)).

Many AD relevant signalling events converge in apoptosis. In this respect it is not surprising that proapototic p53 can be found in the AD network (reviewed in Bamberger & Landreth (2002)). Caspase independent neuronal death in Alzheimer's disease might be mediated by apoptosis-inducing factor 1 (AIFM1). This mitochondrial oxido-reductase co-localizes with neurofibrillary tangles in the hippocampus of post-mortem AD brains (Yu *et al.* (2010)). Decreased expression levels of AIFM1 were measured in the cortex of AD brains (Reix *et al.* (2007)).

The two phage display hits ECHA and TIAF1 are extremely well connected through 23 interactors. Noticeably these network nodes have the highest number of AD relevant interactors. TIAF1 interacts with SYNPO, HtrA, CD59 and SPTB2. Interestingly TIAF1 is also connected to MVP through the 14-3-3 protein YWHAZ. Several studies confirm a role of this protein as tau protein effector (Santpere *et al.* (2007); Umahara *et al.* (2004); Mateo *et al.* (2008)).

The postsynaptic protein synaptopodin (Synpo) was found to be downregulated in cortices of Alzheimer's brains and might contribute to a loss of synaptic function and neurodegeneration in AD (Reddy *et al.* (2005)). SPTB2, Vimentin and HtrA are proteins that are thought to play a role in the ß-Amyloid pathway (Grau *et al.* (2005); Levin *et al.* (2009); Sihag & Cataldo

(1996)).

ECHA interacts with a considerable number of proteins that have been associated with tauopathies. MARK2 phosphorylates tau, as does KCC2D (CAMKII), and MIF interactor PP2A (Segu *et al.* (2008); Wang *et al.* (2007)). PP5 dephosphorylates tau and PP5 levels were found to be reduced by 20% in AD neocortex (Liu *et al.* (2005)). Tau itself was not found to be a direct interaction partner of phage display hits in the network but experimental evidence suggests that MAP1B, a member of the same protein family as Tau might have a similar role in tauopathies (reviewed in Riederer (2007)). It has been also implicated in ß-Amyloid(42) induced apoptosis (Fifre *et al.* (2006)). In a similar way tau neuronal filament medium peptide (NFM) is hyperphosphorylated in AD brains (Rudrabhatla *et al.* (2010)).

The phage display hit MIF interacts with S10A7, UCHL1, PP2A, p53 and CD74. It is connected to TBB2B and ECHA by TRAF6 as well as to CD37. The Hsp90 kinase co-chaperone CD37 has been recently implicated in regulating tau stability (Jinwal *et al.* (2011)). S10A7 has been suggested as biomarker for Alzheimer's by Qin *et al.* (2009) because it is specifically increased in the CSF and brains of AD patients. TRAF6 is an important signaling molecule that transmits signals from the TNF and Toll like receptors. Neurotrophin (p75) belongs to the TNF superfamily and subsequent to ligand binding recruits TRAF6. In a current model of Alzheimer's disease this leads to poly-ubiquitination of TRAF6 and subsequent binding of Presenilin-1 leads to the cleavage of p75 (Powell *et al.* (2009)). Presenilin and p75 are both well characterized key players in Alzheimer's disease (Kovacs *et al.* (1996); Wang *et al.* (2011)). In a mouse model of AD injection of ubiquitin C-terminal hydrolase L1 (Uch-L1) restored synaptic function (Gong *et al.* (2006)). CD74 has a receptor binding site for MIF and is found at increased levels in microglia, neurofibrillary tangles and amyloid-ß plaques of AD patients (Bryan *et al.* (2008)). This finding contributes to increasing evidence that an inflammatory component is involved in AD pathology (Parachikova *et al.* (2007)).

Considering the immunoregulatory function of macrophage inhibitory factor MIF it is not surprising that its interactors seem to be involved in inflammatory aspects of AD. HDAC6, the valosin containing protein (VCP) and Midkine (MK) interact with TBB2B. MK is a heparin binding cytokine enriched in the serum of Alzheimer's patients (Salama *et al.* (2005)). It strongly binds to ß-amyloid and is highly abundant in senile plaques of Alzheimer's disease patients (Salama *et al.* (2005); Muramatsu *et al.* (2011); Yasuhara *et al.* (1993)). Further binders to ß-amyloid in the network are TIAF1, MIF, TBB2B and TRY2. TRY2 also known as PRSS2 has not been previously associated with AD. The AAA ATPase VCP has been implicated in Alzheimer relevant processes like the regulation of the ubiquitin proteasome

system (Weihl *et al.* (2009); Halawani *et al.* (2010)). Recent evidence also points to a role in tau phosphorylation at Ser262/356 (Dolan *et al.* (2011)). Inhibitors of the histone deacetylase HDAC6 are currently under discussion as potential AD drugs (reviewed in Kazantsev & Thompson (2008)).

In conclusion, both the Alzheimers and the MS hit-centered protein interaction networks, show a strong connection of the phage display hits to each other and to other disease-relevant genes. This suggests that the phage display biomarker screen produced highly useful results that might contain clues for new research angles as well as novel biomarker candidates.

5 Discussion

The aim of this work was to establish a method for biomarker discovery by phage display. Bacteriophages displaying peptides based on a cDNA library were prepared in three different *E.coli* strains and panned against IgG and IgA fractions from 15 different human sera, five of which come from Alzheimer's disease patients, five from multiple sclerosis patients and five from healthy controls. Enrichment of binders was monitored by ELISA and analyzed for strain-specific enrichment patterns. To validate the enrichment and demonstrate the potential of the enriched binders as biomarker candidates, 18 selected samples were sequenced on a 454 second generation sequencing platform.

5.1 Characterization of the cDNA library

Starting with a cDNA library prepared as Gateway entry vectors (Invitrogen) with a number of $1.27 * 10^5$ clones, inserts were previously subcloned in our group into the pJuFo ABC system at a much higher number of $9.6 * 10^6$ clones. The three *E.coli* strains assayed in this work, XL-1 blue, ER2738 and TG1, were transformed with the pJuFo library in numbers sufficient to preserve the libraries complexity. To assess the complexity in terms of different genes, samples from the three strains were sequenced and compared. For this comparison, the samples were assumed to be independent random draws from a common pool and that the probability of finding each insert is equal. These assumptions lead to an estimate of less than 2000 different genes. The real number might be higher than that, because the assumptions made oversimplify the problem and tend to underestimate the overlap. Some genes might be represented in high copies so rare inserts are less likely to be sequenced. In complexity estimations from EST sequencing experiments it is often assumed that mRNA abundance follows a Poisson distribution (Wang *et al.* (2005)). However, the situation might be more complex than that and many different factors can potentially have a biasing impact on the sequencing results. Differences in mRNA abundance levels, cDNA synthesis, adapter ligation to enriched library inserts and the PCR during 454 sequencing can potentially lead to the preference of some inserts over others. Having all these potential sources of bias in mind a simple method for complexity estimation was chosen to get a rough idea of the library composition and diversity used in phage display later on. This means that any violations of the underlying assumptions make the overlap larger than expected, which in turn means that the number of genes is an underestimate. Also, since the number of constructs is probably considerably higher than the number of genes and each insert can

be expressed from three different vectors, pJuFo A, pJuFo B and pJuFo C, in different frames, the number of expressed peptides is likely to be at least an order of magnitude higher than this. Mapping of the sequences obtained from the library samples suggests that at least half of the inserts express 3'UTR only. This, and the fact that even the CDS-containing inserts are expressed in all three frames, means that the displayed peptides are mostly artificial (at least 85%). Many successful phage display experiments use completely synthetic libraries (reviewed in Castel *et al.* (2011)). At the same time, the cDNA approach covers natural peptides, which allows drawing conclusions about the usefulness of the phage display experiment based on how well the enriched binders reflect the respective diseases.

5.2 Evaluation of the phage display host strains ER2738, TG1, XL-1 blue

To monitor enrichment of binders during biopanning and find out which strain is suited best for that purpose an ELISA was made for each round. The ELISA binding profiles show clear differences between the three *E.coli* strains. To obtain a classification that is as free from any subjective factors as possible, the binding differences between the four panning rounds for each sample were used to cluster the samples. The enrichment profiles in the ELISA show clear strain-specific patterns. One would expect to see an increasing ELISA signal in every subsequent panning round because of cumulating binding specificity. This was the case in most of the XL-1 derived phages. The signal is below the background level for the first three panning rounds and does not show a strong signal until the fourth round. The most common profile, seen most often in ER2738, shows a decrease in signal between the first two panning rounds. This is likely to be a result of using Hyperphage as helper phage in the first panning round and M13K07 in subsequent rounds. Hyperphage in contrast to M13K07 presents the recombinant fusion protein on each of its pIII coat proteins and does not have wild type pIII. This results in more antibodies binding to the target and thus a stronger ELISA signal in the first round. In round three and four the signals in these profiles increase again. In the TG1-specific profile signals increase only slightly or even decrease compared to round one. In all samples there was a signal above the background level supporting the concept that it was possible to enrich specific binders. The fact that ELISA signals give strain specific patterns from round one to four in the three different strains substantiates the hypothesis that the choice of host strain influences in which panning round enrichment occurs. Another question in this context is whether the binders that are enriched from each of

the sera are the same in all three host strains. When panning a library with such a complex compound like patient serum many factors influence which binders are enriched. Therefore an antibody directed against the housekeeping genes GAPDH, Ubiquitin or Tubulin was used instead of patient serum. In a similar approach (Walter *et al.* (2001)) specific enrichment of GAPDH and Ubiquitin was confirmed in ELISA but not in PCR or Sanger sequencing. Quantification of GAPDH enrichment with qPCR showed that there was some GAPDH in the initial library but that it was lost during phage display. Sanger sequencing confirmed that indeed neither the enrichment of GAPDH nor that one of Ubiquitin and Tubulin was successful. The best explanation is that the antibodies used in this experiment were directed against an epitope in the N-terminal region of the protein. Since cDNA synthesis starts at the 3'-end of the mRNA, primers specific for a region close to the 3'end of the coding sequence were chosen, so they would be able to detect any inserts able to express at least parts of the GAPDH protein. This would explain why GAPDH could be detected in the initial library using qPCR, but was not bound by the antibodies. The longest GAPDH fragment found in the initial library contains less than one quarter of the coding sequence, agreeing well with this line of reasoning. It is also possible that enriched clones are mimotopes that have a structural similarity to GAPDH that is not trivial to recognize. Antibodies usually recognize epitopes of 2-8 amino acids length (Moreau *et al.* (2006). As the epitope from the antibodies used in this experiment is not known it is extremely difficult to predict whether the enriched binders contain this epitope. Enrichment of mimotopes also called phagotopes from random peptide libraries is a frequent event and these molecules were shown to specifically bind to the antigen (Davies *et al.* (1999); Königs *et al.* (2000)). Phage display of a peptide phage library with a well characterized monoclonal antibody revealed 18 phagotopes after three rounds of biopanning (Cook *et al.* (1998)). In another study phagotopes were isolated that could be useful for improved differential diagnosis of the autoimmune disease diabetes 1 (Devendra *et al.* (2004); Ola *et al.* (2006)). To asses which host strain was most useful in this study results from 454 sequencing were used for answering this question. Experiments to test the performance of phage display peptides as biomarkers were not made therefore the analysis had to be restricted to natural peptides that have been described in literature. The highest amount of meaningful proteins that were analyzed in a protein-protein network was derived from ER2738 as host strain. Compared with five peptides from TG1 samples and two peptides from XL-1 samples the relatively high number of 13 peptides was isolated from ER2738. This can have diverse causes that can have their origin in every single step of the experiment. For example highest clone numbers were achieved after transforming the library in this strain and

also estimated complexity was highest in this strain. Furthermore enrichment profiles from ELISA were most consistent in this strain. This indicates that it would be most promising to use this strain for larger biomarker screens with more patients.

5.3 Sequence analysis of enriched libraries

Phage displayed proteins that are naturally expressed human proteins allows to investigate if they are involved in disease relevant molecular mechanisms. The enrichment of binders with a broad literature basis that implicates them in disease or demonstrates their presence in patient serum allows drawing conclusions about how well the phage display experiment worked. Since the binders that correspond to human proteins are highly disease-relevant, it is probable that the synthetic peptides are suitable biomarkers as well. The phenomenon that a humeral immune response is elicited against such cryptic epitopes is well described (reviewed in Chatterjee & Pal (2009)) . In a recent publication antibodies against cryptic epitopes were shown to stain breast cancer tissue selectively leading the authors to speculate that somatic mutations causing expression of frame-shifted transcript might be common in late-onset diseases that are not caused by infectious agents (Fischer *et al.* (2010)).

The CDS containing phage display hits revealed in this study were evaluated for molecular mechanisms involved in disease progression not regarding frame and length of CDS. Some of the hits were highly enriched and up to 4000 reads were sequenced while others were represented by only two reads. The number of reads for a hit is not necessarily correlated with its usefulness as a biomarker and depends on other factors like abundance in the initial library or propagation properties of the phage clone. This demonstrates the advantage of using second generation sequencing in this experiment. Inserts of low abundance would have been overlooked easily using Sanger sequencing. Most hits appear to be disease specific and only 4 out of 65 hits can be found in healthy samples as well. These hits should be excluded as biomarker candidates because incidence in healthy controls makes it very likely that biomarker specificity in a study with more patients would be low. In an analysis of phage display hits on protein level, 69% of CDS containing hits could be recovered. Sequences were translated and mapped to the SwissProt protein database using BLAST. It was hypothesized that synthetic peptides might match human protein sequence but not human mRNA sequence, because of the redundancy in protein translation. In other words, a piece of untranslated region or an mRNA in the wrong frame might, by chance, resemble a natural peptide when expressed. This was the case

for two phage display hits: LOC100131754 and PLLP. In the case of highly enriched 3'UTR only containing Plasmolipin (PLLP) translation on protein level demonstrates that the expressed peptide is highly similar to TAF5, which is of disease relevance. The peptide expressed from LOC100131754 mRNA is similar to NU1M. These are the only cases of protein hits that are not present in the nucleotide-based data set. It is likely that other CDS containing hits were expressed correctly, even if they contain a premature stop codon or a frameshift. Unusual translational events like transcriptional recoding, ribosome hopping or stop codon read-through are well documented events in bacteriophages (reviewed in Baranov *et al.* (2006)). In M13 cDNA phage display the percentage of frameshifted clones can be as high as 90% after three rounds of panning, indicating that non ORF clones might even have an advantage over ORF expressing clones (Cárcamo *et al.* (1998)). Studies with reporter genes cloned after the frameshifted sequence demonstrate that they can indeed be expressed correctly in more than one reading frame (Govarts *et al.* (2010); Goldman *et al.* (2000)). Furthermore sequencing errors are frequent in 454 sequencing and up to 20% of the sequences may contain erroneous insertions (Prabakaran *et al.* (2011)). PCR steps in sample preparation procedures prior to sequencing are additional error sources. In rare cases additional nucleotides are incorporated or mispriming takes place. It is therefore well possible that frameshifted clones like HBB, RL21 or FAM195B are nevertheless displayed on phage surfaces. Therefore, their relevance for disease should not be disregarded nor should they be excluded as potential biomarkers. For AD samples a direct disease relevant link could be established for the hits TIAF1, DDIT3/CHOP/GADD153 and MIF based on several publications (Oyadomari & Mori (2004); Popp *et al.* (2009); Bacher *et al.* (2010); Lee *et al.* (2010)). Without sequencing more patient samples the possibility that these links were found coincidently cannot be ruled out because unfortunately no comprehensive list of proteins involved in AD exists. A larger number of hits would allow applying statistical methods like enrichment analysis but would also require establishing a list of definite AD proteins. The fact that a relatively small number of hits was found in the samples can be most likely attributed to the fact that only binders of panning round four were sequenced. Analysis of samples from the other panning rounds would not only monitor the enrichment process better but would also provide a higher number of candidates to evaluate. Multiple sclerosis is thought to be an autoimmune disease with neurodegenerative aspects. The majority of AD publications involves ß-amyloid or tau pathology directly or indirectly. For MS no such overwhelming dominance of a protein is known but molecules involved in T-cell response against heterologous myelin are often studied in this context. The MS phage display hits ATP5J2, DCTN1 are most likely

to play a disease relevant role. RPL7 is a known auto-antigen for other autoimmune diseases. In comparison to AD phage display hits literature on MS hits is sparse. The hits LTSZ, RPS26, RABAC1 and SDHB could be linked to MS indirectly. To confirm that hits play a disease relevant role a protein-protein network was established for both AD and MS.

5.4 Network analysis of phage expressed proteins

Results of the protein network analysis of AD and MS phage display hits support the hypothesis that, in addition to being biomarker candidates, the hits represent biologically meaningful results. The phage display hits are well connected among and many of the proteins that connect phage display hits are clearly implicated in disease pathology. Nodes that connect the different phage display hits were only described when a disease relevant link was found. Although these interactors could be of future research interest the purpose of network analysis in this study was only to confirm that phage display hits interact with proteins that were already described to be of importance for MS or AD. For example phage display hit ECHA has not been associated with AD before, but network analysis shows that this protein interacts with the highly relevant AD protein PIN1 as well as a set of other known disease contributors. To examine the role of this protein that has been described as enzyme of fatty acid metabolism regarding AD relevance seems to be highly promising. Other hits like TIAF1, TUBB2B and MIF have been implicated in AD directly and the network analysis further supports this role. Surprisingly TRY2 has not yet been described as AD relevant in literature although it interacts with APP. A role of the hits IFM3, LYRM4, TYB10 and NU1M remains elusive also in the network analysis. An interesting ECHA interactor in the AD network is SQSTM1 not only because it has been implicated in AD in various studies. At the same time this protein was a highly enriched phage display hit that contains 3'UTR only. It can be speculated that under pathological circumstances the stop codon at the interjection between CDS and 3'UTR is read through so a longer protein is translated. Such a mechanism occurs for example in the rare genetic disease epidermolysis bullosa simplex (Gu *et al.* (2003)). The MS network shows interactions of phage display hits with many nodes of immunological importance. RL7 is a known antigen in different autoimmune diseases but the underlying molecular mechanism has not been described to date. Interaction of this protein with the immune signaling molecules CD4, LCK and IKBB suggests a possible role in T-cell response. RS26 is well connected to RL7 and also not described as being disease relevant so far. This could be an interesting hit as well. PRS10 was found to interact with T-Cell receptor TNRA1, which is one of the best

described MS proteins and a genetic risk factor for MS. DSHB is a good example for a protein that wouldn't have been recognized as potentially disease relevant without the network analysis but interactions with HLA-A, HLA-C, IKBKE provide a strong link to multiple sclerosis. DCTN1, LZTS2, ATPK and TAF5 were found to interact with a series of signaling molecules that can be involved in immune system pathways. The network analysis done in this study could open new leads for AD and MS research and suggests experimentally examining proteins interacting with many phage display hits in respect to disease mechanisms.

6 Outlook and Conclusion

A novel method of screening for biomarker candidates has been developed and shown to be extremely useful in a proof of principle screen. In the context of this screen, three different *E.coli* strains were tested and evaluated in terms of usefulness for phage display. Three different plasmids, expressing the three forward reading frames of the inserts, were successfully used to multiply the number of expressed peptides. Binders were enriched in four subsequent rounds and the enrichment was monitored by ELISA. The enrichment profiles were used to cluster the samples revealing strain-specific enrichment profiles. The best and least specific profiles were observed mostly in ER2738 samples. Some XL-1 blue samples showed a more delayed enrichment profile. Other profile types, characterized by a loss of binding strength between subsequent rounds, were observed mostly in TG1 samples. To control the enrichment in greater detail and obtain sequences of biomarker candidates, samples from the fourth panning round were sequenced on a 454 sequencing platform. The sequencing results reflect the proteins involved in the respective disease quite well. This makes the phage display results appear very reliable. The enrichment of naturally occurring peptides is most pronounced in ER2738 and almost absent for TG1. This finding, in combination with the ELISA results, makes ER2738 seem more suited for phage display experiments than XL-1 and TG1. The next step in validating the method and the resulting biomarker candidates is the sequencing of samples from other patients. 216 additional samples from twelve additional patients or healthy controls have been prepared in the course of this study. Inserts overlapping between different patients suffering from the same disease are the best biomarker candidates and should be validated in ELISA experiments with many more patient sera. Ideally, a single peptide would be seroreactive in all samples of the respective disease but in no other case. In reality, the best biomarkers are often combinations of several different peptides. The proteins that interact with many of the

phage display hits should also be analyzed for their biological relevance in the context of the disease. In summary, a new method has been developed and applied to successfully screen for biomarker candidates that are now ready for further experimental exploration.

References

Adams J James D (2008) Alzheimer's disease, ceramide, visfatin and NAD. *CNS & Neurological Disorders Drug Targets* **7**: 492–498. PMID: 19128206

Ahmed Z, Doward AI, Pryce G, Taylor DL, Pocock JM, Leonard JP, Baker D, Cuzner ML (2002) A role for caspase-1 and -3 in the pathology of experimental allergic encephalomyelitis : inflammation versus degeneration. *The American Journal of Pathology* **161**: 1577–1586. PMID: 12414506

Alirezaei M, Fox HS, Flynn CT, Moore CS, Hebb ALO, Frausto RF, Bhan V, Kiosses WB, Whitton JL, Robertson GS, Crocker SJ (2009) Elevated ATG5 expression in autoimmune demyelination and multiple sclerosis. *Autophagy* **5**: 152–158. PMID: 19066443 PMCID: 2779564

Anderson JM, Patani R, Reynolds R, Nicholas R, Compston A, Spillantini MG, Chandran S (2010) Abnormal tau phosphorylation in primary progressive multiple sclerosis. *Acta Neuropathologica* **119**: 591–600. PMID: 20306268

Anzilotti C, Merlini G, Pratesi F, Tommasi C, Chimenti D, Migliorini P (2006) Antibodies to viral citrullinated peptide in rheumatoid arthritis. *The Journal of Rheumatology* **33**: 647–651. PMID: 16511941

Ascherio A, Munger KL (2010) Epstein-barr virus infection and multiple sclerosis: a review. *Journal of Neuroimmune Pharmacology: The Official Journal of the Society on NeuroImmune Pharmacology* **5**: 271–277. PMID: 20369303

Assinder SJ, Stanton JL, Prasad PD (2009) Transgelin: an actin-binding protein and tumour suppressor. *The International Journal of Biochemistry & Cell Biology* **41**: 482–486. PMID: 18378184

Bacher M, Deuster O, Aljabari B, Egensperger R, Neff F, Jessen F, Popp J, Noelker C, Reese JP, Al-Abed Y, Dodel R (2010) The

role of macrophage migration inhibitory factor in Alzheimer's disease. *Molecular Medicine (Cambridge, Mass)* **16**: 116–121. PMID: 20200619

Balastik M, Lim J, Pastorino L, Lu KP (2007) Pin1 in Alzheimer's disease: multiple substrates, one regulatory mechanism? *Biochimica et biophysica acta* **1772**: 422–429. PMID: 17317113 PMCID: 1868500

Bamberger ME, Landreth GE (2002) Inflammation, Apoptosis, and Alzheimer's Disease. *The Neuroscientist* **8**: 276 –283

Baranov PV, Fayet O, Hendrix RW, Atkins JF (2006) Recoding in bacteriophages and bacterial IS elements. *Trends in Genetics* **22**: 174–181

Barbas r C F, Kang AS, Lerner RA, Benkovic SJ (1991) Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 7978–7982. PMID: 1896445

Berger W, Steiner E, Grusch M, Elbling L, Micksche M (2009) Vaults and the major vault protein: novel roles in signal pathway regulation and immunity. *Cellular and Molecular Life Sciences: CMLS* **66**: 43–61. PMID: 18759128

Beurel E (2011) Regulation by glycogen synthase kinase-3 of inflammation and T cells in CNS diseases. *Frontiers in Molecular Neuroscience* **4**: 18. PMID: 21941466

Bhaskar K, Miller M, Chludzinski A, Herrup K, Zagorski M, Lamb BT (2009) The PI3K-Akt-mTOR pathway regulates Abeta oligomer induced neuronal cell cycle events. *Molecular Neurodegeneration* **4**: 14. PMID: 19291319

Bhat KS, Morrison SG (1993) Primary structure of human ribosomal protein S21. *Nucleic Acids Research* **21**: 2939. PMID: 8332502

Bielekova B, Martin R (2004) Development of biomarkers in multiple sclerosis. *Brain: A Journal of Neurology* **127**: 1463–1478. PMID: 15180926

Blennow K, Hampel H (2003) CSF markers for incipient Alzheimer's disease. *Lancet Neurology* **2**: 605–613. PMID: 14505582

Bodenhofer U, Kothmeier A, Hochreiter S (2011) APCluster: an R package for affinity propagation clustering. *Bioinformatics (Oxford, England)* **27**: 2463–2464. PMID: 21737437

Bodzioch M, Lapicka-Bodzioch K, Zapala B, Kamysz W, Kiec-Wilk B, Dembinska-Kiec A (2009) Evidence for potential functionality of nuclearly-encoded humanin isoforms. *Genomics* **94**: 247–256

Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Research* **6**: 791–806. PMID: 8889548

Breitling F, Dübel S, Seehaus T, Klewinghaus I, Little M (1991) A surface expression vector for antibody screening. *Gene* **104**: 147–153. PMID: 1916287

Brouwers N, Sleegers K, Van Broeckhoven C (2008) Molecular genetics of Alzheimer's disease: an update. *Annals of Medicine* **40**: 562–583. PMID: 18608129

Brunet E, Chauvin C, Choumet V, Jestin J (2002) A novel strategy for the functional cloning of enzymes using filamentous phage display: the case of nucleotidyl transferases. *Nucleic Acids Research* **30**: e40

Bryan KJ, Zhu X, Harris PL, Perry G, Castellani RJ, Smith MA, Casadesus G (2008) Expression of CD74 is increased in neurofibrillary tangles in Alzheimer's disease. *Molecular Neurodegeneration* **3**: 13–13. PMID: 18786268 PMCID: 2565661

Butterfield DA, Hardas SS, Lange MLB (2010) Oxidatively modified glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and Alzheimer's disease: many pathways to neurodegeneration. *Journal of Alzheimer's Disease: JAD* **20**: 369–393. PMID: 20164570

Caberoy NB, Zhou Y, Jiang X, Alvarado G, Li W (2010) Efficient identification of tubby-binding proteins by an improved system

of T7 phage display. *Journal of Molecular Recognition: JMR* **23**: 74–83. PMID: 19718693

Campbell GR, Ziabreva I, Reeve AK, Krishnan KJ, Reynolds R, Howell O, Lassmann H, Turnbull DM, Mahad DJ (2011) Mitochondrial DNA deletions and neurodegeneration in multiple sclerosis. *Annals of Neurology* **69**: 481–492. PMID: 21446022

Cárcamo J, Ravera MW, Brissette R, Dedova O, Beasley JR, Alam-Moghé A, Wan C, Blume A, Mandecki W (1998) Unexpected frameshifts from gene to expressed protein in a phage-displayed peptide library. *Proceedings of the National Academy of Sciences* **95**: 11146 –11151

Castel G, Chtéoui M, Heyd B, Tordo N (2011) Phage display of combinatorial peptide libraries: application to antiviral research. *Molecules Basel Switzerland* **16**: 3499–3518

Charil A, Yousry TA, Rovaris M, Barkhof F, De Stefano N, Fazekas F, Miller DH, Montalban X, Simon JH, Polman C, Filippi M (2006) MRI and the diagnosis of multiple sclerosis: expanding the concept of "no better explanation". *Lancet Neurology* **5**: 841–852. PMID: 16987731

Chasteen L, Ayriss J, Pavlik P, Bradbury ARM (2006) Eliminating helper phage from phage display. *Nucleic Acids Research* **34**: e145. PMID: 17088290 PMCID: 1693883

Chatterjee S, Pal JK (2009) Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biology of the Cell / Under the Auspices of the European Cell Biology Organization* **101**: 251–262. PMID: 19275763

Chau MF, Radeke MJ, de Inés C, Barasoain I, Kohlstaedt LA, Feinstein SC (1998) The microtubule-associated protein tau cross-links to two distinct sites on each alpha and beta tubulin monomer via separate domains. *Biochemistry* **37**: 17692–17703. PMID: 9922135

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* **14**: 1147–1159. PMID: 15140833

Cho HH, Song JS, Yu JM, Yu SS, Choi SJ, Kim DH, Jung JS (2008) Differential effect of NF-kappaB activity on beta-catenin/Tcf pathway in various cancer cells. *FEBS Letters* **582**: 616–622. PMID: 18242184

Christophi GP, Panos M, Hudson CA, Christophi RL, Gruber RC, Mersich AT, Blystone SD, Jubelt B, Massa PT (2009) Macrophages of multiple sclerosis patients display deficient SHP-1 expression and enhanced inflammatory phenotype. *Lab Invest* **89**: 742–759

Chuang H, Lee E, Liu Y, Lee D, Ideker T (2007) Network-based classification of breast cancer metastasis. *Molecular Systems Biology* **3**: 140. PMID: 17940530

Cid C, Alvarez-Cermeño JC, Regidor I, Plaza J, Salinas M, Alcázar A (2003) Caspase inhibitors protect against neuronal apoptosis induced by cerebrospinal fluid from multiple sclerosis patients. *Journal of Neuroimmunology* **136**: 119–124. PMID: 12620650

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols* **2**: 2366–2382

Colasanti T, Barbati C, Rosano G, Malorni W, Ortona E (2010) Autoantibodies in patients with Alzheimer's disease: pathogenetic role and potential use as biomarkers of disease progression. *Autoimmunity Reviews* **9**: 807–811. PMID: 20656067

Cook AD, Davies JM, Myers MA, Mackay IR, Rowley MJ (1998) Mimotopes identified by phage display for the monoclonal antibody CII-C1 to type II collagen. *Journal of Autoimmunity* **11**: 205–211. PMID: 9693968

Craig-Schapiro R, Kuhn M, Xiong C, Pickering EH, Liu J, Misko TP, Perrin RJ, Bales KR, Soares H, Fagan AM, Holtzman DM (2011) Multiplexed immunoassay panel identifies novel CSF biomarkers for Alzheimer's disease diagnosis and prognosis. *PloS One* **6**: e18850. PMID: 21526197

Crameri R, Suter M (1993) Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. *Gene* **137**: 69–75. PMID: 8282202

Danner S, Belasco JG (2001) T7 phage display: A novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proceedings of the National Academy of Sciences* **98**: 12954 –12959

Davies JM, Rowley MJ, MacKay IR (1999) Phagotopes derived by antibody screening of phage-displayed random peptide libraries vary in immunoreactivity: studies using an exemplary monoclonal antibody, CII-C1, to type II collagen. *Immunology and Cell Biology* **77**: 483–490. PMID: 10571668

Devendra D, Galloway TS, Horton SJ, Wilkin TJ (2004) Exploring the idiotypes of insulin antibodies as markers for remission in Type 1 diabetes. *Diabetic Medicine: A Journal of the British Diabetic Association* **21**: 1316–1324. PMID: 15569135

Dias-Neto E, Nunes DN, Giordano RJ, Sun J, Botz GH, Yang K, Setubal JaC, Pasqualini R, Arap W (2009) Next-Generation Phage Display: Integrating and Comparing Available Molecular Tools to Enable Cost-Effective High-Throughput Analysis. *PLoS ONE* **4**: e8338

Dolan PJ, Jin YN, Hwang W, Johnson GVW (2011) Decreases in valosin-containing protein result in increased levels of tau phosphorylated at Ser(262/356). *FEBS Letters* **585**: 3424–3429. PMID: 21983102

Donev R, Kolev M, Millet B, Thome J (2009) Neuronal death in Alzheimer's disease and therapeutic opportunities. *Journal of Cellular and Molecular Medicine* **13**: 4329–4348

Dutta R, McDonough J, Chang A, Swamy L, Siu A, Kidd GJ, Rudick R, Mirnics K, Trapp BD (2007) Activation of the ciliary neurotrophic factor (CNTF) signalling pathway in cortical neurons of multiple sclerosis patients. *Brain* **130**: 2566 –2576

Dutta R, McDonough J, Yin X, Peterson J, Chang A, Torres T, Gudz T, Macklin WB, Lewis DA, Fox RJ, Rudick R, Mirnics

K, Trapp BD (2006) Mitochondrial dysfunction as a cause of axonal degeneration in multiple sclerosis patients. *Annals of Neurology* **59**: 478–489. PMID: 16392116

Eaton S, Bursby T, Middleton B, Pourfarzam M, Mills K, Johnson AW, Bartlett K (2000) The mitochondrial trifunctional protein: centre of a beta-oxidation metabolon? *Biochemical Society Transactions* **28**: 177–182. PMID: 10816122

Eschbach J, Dupuis L (2011) Cytoplasmic dynein in neurodegeneration. *Pharmacology & Therapeutics* **130**: 348–363. PMID: 21420428

Eto M (2009) Regulation of cellular protein phosphatase-1 (PP1) by phosphorylation of the CPI-17 family, C-kinase-activated PP1 inhibitors. *The Journal of Biological Chemistry* **284**: 35273–35277. PMID: 19846560

Feng X, Petraglia AL, Chen M, Byskosh PV, Boos MD, Reder AT (2002) Low expression of interferon-stimulated genes in active multiple sclerosis is linked to subnormal phosphorylation of STAT1. *Journal of Neuroimmunology* **129**: 205–215. PMID: 12161037

Fifre A, Sponne I, Koziel V, Kriem B, Yen Potin FT, Bihain BE, Olivier J, Oster T, Pillot T (2006) Microtubule-associated protein MAP1A, MAP1B, and MAP2 proteolysis during soluble amyloid beta-peptide-induced neuronal apoptosis. Synergistic involvement of calpain and caspase-3. *The Journal of Biological Chemistry* **281**: 229–240. PMID: 16234245

Fischer E, Kobold S, Kleber S, Kubuschok B, Braziulis E, Knuth A, Renner C, Wadle A (2010) Cryptic epitopes induce high-titer humoral immune response in patients with cancer. *Journal of Immunology (Baltimore, Md: 1950)* **185**: 3095–3102. PMID: 20660712

Fosså A, Alsoe L, Crameri R, Funderud S, Gaudernack G, Smeland EB (2004) Serological cloning of cancer/testis antigens expressed in prostate cancer using cDNA phage surface display. *Cancer Immunology, Immunotherapy: CII* **53**: 431–438. PMID: 14747957

Frey BJ, Dueck D (2007) Clustering by passing messages between data points. *Science (New York, NY)* **315**: 972–976. PMID: 17218491

Gallo P, Pagni S, Piccinno MG, Giometto B, Argentiero V, Chiusole M, Bozza F, Tavolato B (1992) On the role of interleukin-2 (IL-2) in multiple sclerosis (MS). IL-2-mediated endothelial cell activation. *Italian Journal of Neurological Sciences* **13**: 65–68. PMID: 1345742

Gandy S (2011) Perspective: Prevention is better than cure. *Nature* **475**: S15

Georgieva Y, Konthur Z (2011) Design and screening of M13 phage display cDNA libraries. *Molecules (Basel, Switzerland)* **16**: 1667–1681. PMID: 21330956

Gerber J, Mühlenhoff U, Lill R (2003) An interaction between frataxin and Isu1/Nfs1 that is crucial for Fe/S cluster synthesis on Isu1. *EMBO Reports* **4**: 906–911. PMID: 12947415

Goldman E, Korus M, Mandecki W (2000) Efficiencies of translation in three reading frames of unusual non-ORF sequences isolated from phage display. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* **14**: 603–611. PMID: 10698976

Gong B, Cao Z, Zheng P, Vitolo OV, Liu S, Staniszewski A, Moolman D, Zhang H, Shelanski M, Arancio O (2006) Ubiquitin Hydrolase Uch-L1 Rescues beta-Amyloid-Induced Decreases in Synaptic Function and Contextual Memory. *Cell* **126**: 775–788

Govarts C, Somers K, Stinissen P, Somers V (2010) Frameshifting in the p6 cDNA phage display system. *Molecules (Basel, Switzerland)* **15**: 9380–9390. PMID: 21173723

Goverman JM (2011) Immune tolerance in multiple sclerosis. *Immunological Reviews* **241**: 228–240. PMID: 21488900

Grau S, Baldi A, Bussani R, Tian X, Stefanescu R, Przybylski M, Richards P, Jones SA, Shridhar V, Clausen T, Ehrmann M (2005) Implications of the serine protease HtrA1 in amyloid precursor protein processing. *Proceedings of the National Academy of Sciences* **102**: 6021 –6026

Gu L, Kim S, Ichiki Y, Park J, Nagai M, Kitajima Y (2003) A usual frameshift and delayed termination codon mutation in keratin 5 causes a novel type of epidermolysis bullosa simplex with migratory circinate erythema. *The Journal of Investigative Dermatology* **121**: 482–485. PMID: 12925204

Halawani D, Tessier S, Anzellotti D, Bennett DA, Latterich M, LeBlanc AC (2010) Identification of Caspase-6-Mediated Processing of the Valosin Containing Protein (p97) in Alzheimer's Disease: A Novel Link to Dysfunction in Ubiquitin Proteasome System-Mediated Protein Degradation. *The Journal of Neuroscience* **30**: 6132 –6142

Harrison PJ, Lyon L, Sartorius LJ, Burnet PWJ, Lane TA (2008) The group II metabotropic glutamate receptor 3 (mGluR3, mGlu3, GRM3): expression, function and involvement in schizophrenia. *Journal of Psychopharmacology (Oxford, England)* **22**: 308–322. PMID: 18541626

Hermann-Kleiter N, Baier G (2010) NFAT pulls the strings during CD4+ T helper cell effector functions. *Blood* **115**: 2989 –2997

Huang Q, Qian Z, Li H (2010) A comparative study of the urinary trypsinogen-2, trypsinogen activation peptide, and the computed tomography severity index as early predictors of the severity of acute pancreatitis. *Hepato-Gastroenterology* **57**: 1295–1299. PMID: 21410075

Islam K, Levy E (1997) Carboxyl-terminal fragments of beta-amyloid precursor protein bind to microtubules and the associated protein tau. *The American Journal of Pathology* **151**: 265–271. PMID: 9212751

Jinwal UK, Trotter JH, Abisambra JF, Koren J, Lawson LY, Vestal GD, O'Leary JC, Johnson AG, Jin Y, Jones JR, Li Q, Weeber EJ, Dickey CA (2011) The Hsp90 Kinase Co-chaperone Cdc37 Regulates Tau Stability and Phosphorylation Dynamics. *Journal of Biological Chemistry* **286**: 16976 –16983

Ka S, Rifai A, Chen J, Cheng C, Shui H, Lee H, Lin Y, Hsu L, Chen A (2006) Glomerular crescent-related biomarkers in a murine model of chronic graft versus host disease. *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis*

and Transplant Association - European Renal Association **21**: 288–298. PMID: 16249193

Kamburov A, Wierling C, Lehrach H, Herwig R (2009) ConsensusPathDB–a database for integrating human functional interaction networks. *Nucleic Acids Research* **37**: D623–D628

Kardon JR, Vale RD (2009) Regulators of the cytoplasmic dynein motor. *Nature Reviews Molecular Cell Biology* **10**: 854–865. PMID: 19935668

Kazantsev AG, Thompson LM (2008) Therapeutic application of histone deacetylase inhibitors for central nervous system disorders. *Nature Reviews Drug Discovery* **7**: 854–868

Kelsey G (2009) Epigenetics and imprinted genes: insights from the imprinted Gnas locus. *Hormone Research* **71 Suppl 2**: 22–29. PMID: 19407493

Kemp EH, Herd LM, Waterman EA, Wilson AG, Weetman AP, Watson PP (2002) Immunoscreening of phage-displayed cDNA-encoded polypeptides identifies B cell targets in autoimmune disease. *Biochemical and Biophysical Research Communications* **298**: 169–177. PMID: 12379236

Kinosita J K, Yasuda R, Noji H (2000) F1-ATPase: a highly efficient rotary ATP machine. *Essays in Biochemistry* **35**: 3–18. PMID: 12471886

Koc EC, Burkhart W, Blackburn K, Moseley A, Koc H, Spremulli LL (2000) A proteomics approach to the identification of mammalian mitochondrial small subunit ribosomal proteins. *The Journal of Biological Chemistry* **275**: 32585–32591. PMID: 10938081

Koch M, Mostert J, Heersema D, De Keyser J (2007) Progression in multiple sclerosis: Further evidence of an age dependent process. *Journal of the Neurological Sciences* **255**: 35–41

Königs C, Rowley MJ, Thompson P, Myers MA, Scealy M, Davies JM, Wu L, Dietrich U, Mackay CR, Mackay IR (2000) Monoclonal antibody screening of a phage-displayed random peptide library reveals mimotopes of chemokine receptor CCR5: implications for the tertiary structure of the receptor and for

an N-terminal binding site for HIV-1 gp120. *European Journal of Immunology* **30**: 1162–1171. PMID: 10760806

Konthur Z, Crameri R (2003) High-throughput applications of phage display in proteomic analyses. *TARGETS* **2**: 261–270

Kovacs DM, Fausett HJ, Page KJ, Kim T, Moir RD, Merriam DE, Hollister RD, Hallmark OG, Mancini R, Felsenstein KM, Hyman BT, Tanzi RE, Wasco W (1996) Alzheimer-associated presenilins 1 and 2 : Neuronal expression in brain and localization to intracellular membranes in mammalian cells. *Nat Med* **2**: 224–229

Kreutzer M, Seehusen F, Kreutzer R, Pringproa K, Kummerfeld M, Claus P, Deschl U, Kalkul A, Beineke A, Baumgärtner W, Ulrich R (2011) Axonopathy Is Associated with Complex Axonal Transport Defects in a Model of Multiple Sclerosis. *Brain Pathology (Zurich, Switzerland)* PMID: 21988534

Kümpfel T, Hohlfeld R (2009) Multiple sclerosis. TNFRSF1A, TRAPS and multiple sclerosis. *Nature Reviews Neurology* **5**: 528–529. PMID: 19794511

Kwak Y, Ma T, Diao S, Zhang X, Chen Y, Hsu J, Lipton SA, Masliah E, Xu H, Liao F (2010) NO signaling and S-nitrosylation regulate PTEN inhibition in neurodegeneration. *Molecular Neurodegeneration* **5**: 49. PMID: 21067594

Lassmann H (2011) Mechanisms of neurodegeneration shared between multiple sclerosis and Alzheimer's disease. *Journal of Neural Transmission (Vienna, Austria: 1996)* **118**: 747–752. PMID: 21373761

Lee M, Lin S, Chang J, Schultz L, Heath J, Hsu L, Kuo Y, Hong Q, Chiang M, Gong C, Sze C, Chang N (2010) TGF-beta induces TIAF1 self-aggregation via type II receptor-independent signaling that leads to generation of amyloid beta plaques in Alzheimer's disease. *Cell Death & Disease* **1**: e110. PMID: 21368882

Leslie D, Lipsky P, Notkins AL (2001) Autoantibodies as predictors of disease. *Journal of Clinical Investigation* **108**: 1417–1422

Levin EC, Acharya NK, Sedeyn JC, Venkataraman V, D'Andrea MR, Wang H, Nagele RG (2009) Neuronal expression of vimentin in the Alzheimer's disease brain may be part of a generalized dendritic damage-response mechanism. *Brain Research* **1298**: 194–207. PMID: 19728994

Li LY, Shih HM, Liu MY, Chen JY (2001) The cellular protein PRA1 modulates the anti-apoptotic activity of Epstein-Barr virus BHRF1, a homologue of Bcl-2, through direct interaction. *The Journal of Biological Chemistry* **276**: 27354–27362. PMID: 11373297

Liou Y, Sun A, Ryo A, Zhou XZ, Yu Z, Huang H, Uchida T, Bronson R, Bing G, Li X, Hunter T, Lu KP (2003) Role of the prolyl isomerase Pin1 in protecting against age-dependent neurodegeneration. *Nature* **424**: 556–561. PMID: 12891359

Liu F, Iqbal K, Grundke-Iqbal I, Rossie S, Gong C (2005) Dephosphorylation of Tau by Protein Phosphatase 5. *Journal of Biological Chemistry* **280**: 1790 –1796

Liu H, Wu C, Kao H, Huang Y, Liang Y, Chen C, Yu J, Chang Y (2011) Proteome-wide dysregulation by PRA1 depletion delineates a role of PRA1 in lipid transport and cell migration. *Molecular & Cellular Proteomics: MCP* **10**: M900641MCP200. PMID: 20592422

Lopez OL, McDade E, Riverol M, Becker JT (2011) Evolution of the diagnostic criteria for degenerative and cognitive disorders. *Current Opinion in Neurology* **24**: 532–541. PMID: 22071334

Manolopoulos KN, Klotz L, Korsten P, Bornstein SR, Barthel A (2010) Linking Alzheimer's disease to insulin resistance: the FoxO response to oxidative stress. *Molecular Psychiatry* **15**: 1046–1052. PMID: 20966918

Maran C, Tassone E, Masola V, Onisto M (2009) The Story of SPATA2 (Spermatogenesis-Associated Protein 2): From Sertoli Cells to Pancreatic Beta-Cells. *Current Genomics* **10**: 361–363. PMID: 20119533

Mateo I, Sánchez-Juan P, Rodríguez-Rodríguez E, Infante J, Fernández-Viadero C, Peña N, Berciano J, Combarros O (2008)

14-3-3 zeta and tau genes interactively decrease Alzheimer's disease risk. *Dementia and Geriatric Cognitive Disorders* **25**: 317–320

McCafferty J, Griffiths AD, Winter G, Chiswell DJ (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**: 552–554. PMID: 2247164

Migliorini P, Baldini C, Rocchi V, Bombardieri S (2005) Anti-Sm and anti-RNP antibodies. *Autoimmunity* **38**: 47–54. PMID: 15804705

Milhavet O, Martindale JL, Camandola S, Chan SL, Gary DS, Cheng A, Holbrook NJ, Mattson MP (2002) Involvement of Gadd153 in the pathogenic action of presenilin-1 mutations. *Journal of Neurochemistry* **83**: 673–681. PMID: 12390529

Moreau V, Granier C, Villard S, Laune D, Molina F (2006) Discontinuous epitope prediction based on mimotope analysis. *Bioinformatics* **22**: 1088 –1095

Muchowski PJ (2002) Protein misfolding, amyloid formation, and neurodegeneration: a critical role for molecular chaperones? *Neuron* **35**: 9–12. PMID: 12123602

Muramatsu H, Yokoi K, Chen L, Ichihara-Tanaka K, Kimura T, Muramatsu T (2011) Midkine as a factor to counteract the deposition of amyloid beta-peptide plaques: in vitro analysis and examination in knockout mice. *International Archives of Medicine* **4**: 1. PMID: 21223602

Nagele E, Han M, Demarshall C, Belinka B, Nagele R (2011) Diagnosis of Alzheimer's disease based on disease-specific autoantibody profiles in human sera. *PloS One* **6**: e23112. PMID: 21826230

Neu E, Hemmerich PH, Peter HH, Krawinkel U, von Mikecz AH (1997) Characteristic epitope recognition pattern of autoantibodies against eukaryotic ribosomal protein L7 in systemic autoimmune diseases. *Arthritis and Rheumatism* **40**: 661–671. PMID: 9125248

Niatsetskaya Z, Basso M, Speer RE, McConoughey SJ, Coppola G, Ma TC, Ratan RR (2010) HIF prolyl hydroxylase

inhibitors prevent neuronal death induced by mitochondrial toxins: therapeutic implications for Huntington's disease and Alzheimer's disease. *Antioxidants & Redox Signaling* **12**: 435–443. PMID: 19659431

Nizzari M, Venezia V, Repetto E, Caorsi V, Magrassi R, Gagliani MC, Carlo P, Florio T, Schettini G, Tacchetti C, Russo T, Diaspro A, Russo C (2007) Amyloid precursor protein and Presenilin1 interact with the adaptor GRB2 and modulate ERK 1,2 signaling. *The Journal of Biological Chemistry* **282**: 13833–13844. PMID: 17314098

O'Brien RJ, Wong PC (2011) Amyloid precursor protein processing and Alzheimer's disease. *Annual Review of Neuroscience* **34**: 185–204. PMID: 21456963

Ola TO, Biro PA, Hawa MI, Ludvigsson J, Locatelli M, Puglisi MA, Bottazzo GF, Fierabracci A (2006) Importin beta: a novel autoantigen in human autoimmunity identified by screening random peptide libraries on phage. *Journal of Autoimmunity* **26**: 197–207. PMID: 16549322

O'Shea JJ, Lahesmaa R, Vahedi G, Laurence A, Kanno Y (2011) Genomic views of STAT function in CD4+ T helper cell differentiation. *Nature Reviews Immunology* **11**: 239–250. PMID: 21436836

Oyadomari S, Mori M (2004) Roles of CHOP/GADD153 in endoplasmic reticulum stress. *Cell Death and Differentiation* **11**: 381–389. PMID: 14685163

Oyama R, Yamamoto H, Titani K (2000) Glutamine synthetase, hemoglobin alpha-chain, and macrophage migration inhibitory factor binding to amyloid beta-protein: their identification in rat brain by a novel affinity chromatography and in Alzheimers disease brain by immunoprecipitation. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1479**: 91–102

Papini AM (2009) The use of post-translationally modified peptides for detection of biomarkers of immune-mediated diseases. *Journal of Peptide Science* **15**: 621–628

Parachikova A, Agadjanyan MG, Cribbs DH, Blurton-Jones M, Perreau V, Rogers J, Beach TG, Cotman CW (2007) Inflammatory changes parallel the early stages of Alzheimer disease. *Neurobiology of Aging* **28**: 1821–1833. PMID: 17052803

Pastorino L, Sun A, Lu P, Zhou XZ, Balastik M, Finn G, Wulf G, Lim J, Li S, Li X, Xia W, Nicholson LK, Lu KP (2006) The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-[beta] production. *Nature* **440**: 528–534

Pei J, Hugon J (2008) mTOR-dependent signalling in Alzheimer's disease. *Journal of Cellular and Molecular Medicine* **12**: 2525–2532. PMID: 19210753

Popp J, Bacher M, Kölsch H, Noelker C, Deuster O, Dodel R, Jessen F (2009) Macrophage migration inhibitory factor in mild cognitive impairment and Alzheimer's disease. *Journal of Psychiatric Research* **43**: 749–753. PMID: 19038405

Powell JC, Twomey C, Jain R, McCarthy JV (2009) Association between Presenilin-1 and TRAF6 modulates regulated intramembrane proteolysis of the p75NTR neurotrophin receptor. *Journal of Neurochemistry* **108**: 216–230. PMID: 19012753

Prabakaran P, Streaker E, Chen W, Dimitrov DS (2011) 454 antibody sequencing - error characterization and correction. *BMC Research Notes* **4**: 404. PMID: 21992227

Prasanthi JRP, Larson T, Schommer J, Ghribi O (2011) Silencing GADD153/CHOP Gene Expression Protects against Alzheimer's Disease-Like Pathology Induced by 27-Hydroxycholesterol in Rabbit Hippocampus. *PloS One* **6**: e26420. PMID: 22046282

Pratt KG, Zimmerman EC, Cook DG, Sullivan JM (2011) Presenilin 1 regulates homeostatic synaptic scaling through Akt signaling. *Nat Neurosci* **14**: 1112–1114

Puig B, Ferrer I, Ludueña RF, Avila J (2005) BetaII-tubulin and phospho-tau aggregates in Alzheimer's disease and Pick's disease. *Journal of Alzheimer's Disease: JAD* **7**: 213–220; discussion 255–262. PMID: 16006664

Qin W, Ho L, Wang J, Peskind E, Pasinetti GM (2009) S100A7, a Novel Alzheimer's Disease Biomarker with Non-Amyloidogenic alpha-Secretase Activity Acts via Selective Promotion of ADAM-10. *PLoS ONE* **4**: e4183

Qiu C, Kivipelto M, von Strauss E (2009) Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. *Dialogues in Clinical Neuroscience* **11**: 111–128. PMID: 19585947

Rademakers R, Cruts M, Van Broeckhoven C (2003) Genetics of early-onset Alzheimer dementia. *TheScientificWorldJournal* **3**: 497–519. PMID: 12847300

Ramsköld D, Wang ET, Burge CB, Sandberg R (2009) An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Computational Biology* **5**: e1000598. PMID: 20011106

Ray S, Britschgi M, Herbert C, Takeda-Uchimura Y, Boxer A, Blennow K, Friedman LF, Galasko DR, Jutel M, Karydas A, Kaye JA, Leszek J, Miller BL, Minthon L, Quinn JF, Rabinovici GD, Robinson WH, Sabbagh MN, So YT, Sparks DL, *et al.* (2007) Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nature Medicine* **13**: 1359–1362. PMID: 17934472

Raychaudhuri M, Mukhopadhyay D (2010) Grb2-mediated alteration in the trafficking of AbetaPP: insights from Grb2-AICD interaction. *Journal of Alzheimer's Disease: JAD* **20**: 275–292. PMID: 20164575

Reddy MM, Wilson R, Wilson J, Connell S, Gocke A, Hynan L, German D, Kodadek T (2011) Identification of Candidate IgG Biomarkers for Alzheimer's Disease via Combinatorial Library Screening. *Cell* **144**: 132–142

Reddy PH, Mani G, Park BS, Jacques J, Murdoch G, Whetsell J William, Kaye J, Manczak M (2005) Differential loss of synaptic proteins in Alzheimer's disease: implications for synaptic dysfunction. *Journal of Alzheimer's Disease: JAD* **7**: 103–117; discussion 173–180. PMID: 15851848

Reix S, Mechawar N, Susin SA, Quirion R, Krantic S (2007) Expression of cortical and hippocampal apoptosis-inducing factor (AIF) in aging and Alzheimer's disease. *Neurobiology of Aging* **28**: 351–356. PMID: 16504343

Rhyner C, Weichel M, Flueckiger S, Hemmann S, Kleber-Janke T, Crameri R (2004) Cloning allergens via phage display. *Methods* **32**: 212–218

Rickle A, Bogdanovic N, Volkmann I, Zhou X, Pei J, Winblad B, Cowburn RF (2006) PTEN levels in Alzheimer's disease medial temporal cortex. *Neurochemistry International* **48**: 114–123. PMID: 16239049

Riederer BM (2007) Microtubule-associated protein 1B, a growth-associated and phosphorylated scaffold protein. *Brain Research Bulletin* **71**: 541–558. PMID: 17292797

Rondot S, Koch J, Breitling F, Dübel S (2001) A helper phage to improve single-chain antibody presentation in phage display. *Nature Biotechnology* **19**: 75–78. PMID: 11135557

Rudrabhatla P, Grant P, Jaffe H, Strong MJ, Pant HC (2010) Quantitative phosphoproteomic analysis of neuronal intermediate filament proteins (NF-M/H) in Alzheimer's disease by iTRAQ. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* **24**: 4396–4407. PMID: 20624930

Ryu JK, McLarnon JG (2009) A leaky blood-brain barrier, fibrinogen infiltration and microglial reactivity in inflamed Alzheimer's disease brain. *Journal of Cellular and Molecular Medicine* **13**: 2911–2925. PMID: 18657226

Salama RHM, Muramatsu H, Shimizu E, Hashimoto K, Ohgake S, Watanabe H, Komatsu N, Okamura N, Koike K, Shinoda N, Okada Si, Iyo M, Muramatsu T (2005) Increased midkine levels in sera from patients with Alzheimer's disease. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* **29**: 611–616. PMID: 15866365

Salminen A, Kaarniranta K, Haapasalo A, Soininen H, Hiltunen M (2011) AMP-activated protein kinase: a potential player in

Alzheimer's disease. *Journal of Neurochemistry* **118**: 460–474. PMID: 21623793

Santpere G, Puig B, Ferrer I (2007) Oxidative damage of 14-3-3 zeta and gamma isoforms in Alzheimer's disease and cerebral amyloid angiopathy. *Neuroscience* **146**: 1640–1651. PMID: 17445990

Schütze T, Rubelt F, Repkow J, Greiner N, Erdmann VA, Lehrach H, Konthur Z, Glökler J (2011) A streamlined protocol for emulsion polymerase chain reaction and subsequent purification. *Analytical Biochemistry* **410**: 155–157. PMID: 21111698

Segu L, Pascaud A, Costet P, Darmon M, Buhot M (2008) Impairment of spatial learning and memory in ELKL Motif Kinase1 (EMK1/MARK2) knockout mice. *Neurobiology of Aging* **29**: 231–240. PMID: 17196307

Seidman CE, Struhl K, Sheen J, Jessen T (2001) Introduction of plasmid DNA into cells. *Current Protocols in Molecular Biology / Edited by Frederick M Ausubel [et Al]* **Chapter 1**: Unit1.8. PMID: 18265047

Shields DC, Schaecher KE, Saido TC, Banik NL (1999) A putative mechanism of demyelination in multiple sclerosis by a proteolytic enzyme, calpain. *Proceedings of the National Academy of Sciences* **96**: 11486 –11491

Siegrist F, Ebeling M, Certa U (2011) The small interferon-induced transmembrane genes and proteins. *Journal of Interferon & Cytokine Research: The Official Journal of the International Society for Interferon and Cytokine Research* **31**: 183–197. PMID: 21166591

Sihag RK, Cataldo AM (1996) Brain beta-spectrin is a component of senile plaques in Alzheimer's disease. *Brain Research* **743**: 249–257. PMID: 9017252

Smith AW, Doonan BP, Tyor WR, Abou-Fayssal N, Haque A, Banik NL (2011) Regulation of Th1/Th17 cytokines and IDO gene expression by inhibition of calpain in PBMCs from MS patients. *Journal of Neuroimmunology* **232**: 179–185

Smith G (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**: 1315 –1317

Sonoda Y, Mukai H, Matsuo K, Takahashi M, Ono Y, Maeda K, Akiyama H, Kawamata T (2010) Accumulation of tumor-suppressor PTEN in Alzheimer neurofibrillary tangles. *Neuroscience Letters* **471**: 20–24. PMID: 20056128

Svejgaard A (2008) The immunogenetics of multiple sclerosis. *Immunogenetics* **60**: 275–286. PMID: 18461312

Takeda S, Sato N, Rakugi H, Morishita R (2010) Plasma beta-amyloid as potential biomarker of Alzheimer disease: possibility of diagnostic tool for Alzheimer disease. *Molecular bioSystems* **6**: 1760–1766. PMID: 20567751

Tanahashi N, Suzuki M, Fujiwara T, Takahashi E, Shimbara N, Chung CH, Tanaka K (1998) Chromosomal localization and immunological analysis of a family of human 26S proteasomal ATPases. *Biochemical and Biophysical Research Communications* **243**: 229–232. PMID: 9473509

Taylor P, Gartemann J, Hsieh J, Creeden J (2011) A Systematic Review of Serum Biomarkers Anti-Cyclic Citrullinated Peptide and Rheumatoid Factor as Tests for Rheumatoid Arthritis. *Autoimmune Diseases* **2011**: 1–18

Thal LJ (2006) Prevention of Alzheimer disease. *Alzheimer Disease and Associated Disorders* **20**: S97–99. PMID: 16917204

Thyssen G, Li T, Lehmann L, Zhuo M, Sharma M, Sun Z (2006) LZTS2 is a novel beta-catenin-interacting protein and regulates the nuclear export of beta-catenin. *Molecular and Cellular Biology* **26**: 8857–8867. PMID: 17000760

Trapp BD, Nave K (2008) Multiple Sclerosis: An Immune or Neurodegenerative Disorder? *Annual Review of Neuroscience* **31**: 247–269

Tristan C, Shahani N, Sedlak TW, Sawa A (2011) The diverse functions of GAPDH: views from different subcellular compartments. *Cellular Signalling* **23**: 317–323. PMID: 20727968

Trojano M (2011) Multiple sclerosis: The evolving diagnostic criteria for multiple sclerosis. *Nat Rev Neurol* **7**: 251–252

Umahara T, Uchihara T, Tsuchiya K, Nakamura A, Iwamoto T, Ikeda K, Takasaki M (2004) 14-3-3 proteins and zeta isoform containing neurofibrillary tangles in patients with Alzheimer's disease. *Acta Neuropathologica* **108**: 279–286

van Venrooij WJ, van Beers JJBC, Pruijn GJM (2011) Anti-CCP antibodies: the past, the present and the future. *Nature Reviews Rheumatology* **7**: 391–398. PMID: 21647203

Veerhuis R, Nielsen HM, Tenner AJ (2011) Complement in the brain. *Molecular Immunology* **48**: 1592–1603. PMID: 21546088

von Mikecz A, Hemmerich P, Neu E, Peter HH, Krawinkel U (1994) [Ribosomal protein L7 as an autoantigen in patients with systemic lupus erythematosus, mixed collagen disease and rheumatoid arthritis]. *Zeitschrift Für Ärztliche Fortbildung* **88**: 501–503. PMID: 7856252

Walter G, Konthur Z, Lehrach H (2001) High-throughput screening of surface displayed gene products. *Combinatorial Chemistry & High Throughput Screening* **4**: 193–205. PMID: 11281835

Wang J, Grundke-Iqbal I, Iqbal K (2007) Kinases and phosphatases and tau sites involved in Alzheimer neurofibrillary degeneration. *The European Journal of Neuroscience* **25**: 59–68. PMID: 17241267

Wang JZ, Lindsay BG, Cui L, Wall PK, Marion J, Zhang J, dePamphilis CW (2005) Gene capture prediction and overlap estimation in EST sequencing from one or multiple libraries. *BMC Bioinformatics* **6**: 300. PMID: 16351717

Wang Y, Wang X, Lu J, Li Q, Gao C, Liu X, Sun Y, Yang M, Lim Y, Evin G, Zhong J, Masters C, Zhou X (2011) p75NTR Regulates Abeta Deposition by Increasing Abeta Production But Inhibiting Abeta Aggregation with Its Extracellular Domain. *The Journal of Neuroscience* **31**: 2292 –2304

Weichel M, Jaussi R, Rhyner C, Crameri R (2008) Display of E. coli Alkaline Phosphatase pIII or pVIII Fusions on Phagemid

Surfaces Reveals Monovalent Decoration with Active Molecules. *The open biochemistry journal* **2**: 38–43

Weihl CC, Pestronk A, Kimonis VE (2009) Valosin-containing protein disease: inclusion body myopathy with Paget's disease of the bone and fronto-temporal dementia. *Neuromuscular disorders : NMD* **19**: 308–315. PMID: 19380227 PMCID: 2859037

Williams R, Buchheit CL, Berman NEJ, Levine SM (2011) Pathogenic implications of iron accumulation in multiple sclerosis. *Journal of Neurochemistry* PMID: 22004421

Wosik K, Antel J, Kuhlmann T, Brück W, Massie B, Nalbantoglu J (2003) Oligodendrocyte injury in multiple sclerosis: a role for p53. *Journal of Neurochemistry* **85**: 635–644. PMID: 12694389

Wu L, Chang W, Zhao J, Yu Y, Tan X, Su T, Zhao L, Huang S, Liu S, Cao G (2010) Development of autoantibody signatures as novel diagnostic biomarkers of non-small cell lung cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* **16**: 3760–3768. PMID: 20501620

Yan J, Greer JM (2008) NF-kappa B, a potential therapeutic target for the treatment of multiple sclerosis. *CNS & Neurological Disorders Drug Targets* **7**: 536–557. PMID: 19128210

Yang LB, Li R, Meri S, Rogers J, Shen Y (2000) Deficiency of complement defense protein CD59 may contribute to neurodegeneration in Alzheimer's disease. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **20**: 7505–7509. PMID: 11027207

Yasuhara O, Muramatsu H, Kim SU, Muramatsu T, Maruta H, McGeer PL (1993) Midkine, a novel neurotrophic factor, is present in senile plaques of Alzheimer disease. *Biochemical and Biophysical Research Communications* **192**: 246–251. PMID: 8476427

Yousefi S, Perozzo R, Schmid I, Ziemiecki A, Schaffner T, Scapozza L, Brunner T, Simon H (2006) Calpain-mediated cleavage of Atg5 switches autophagy to apoptosis. *Nat Cell Biol* **8**: 1124–1132

Yu W, Mechawar N, Krantic S, Quirion R (2010) Evidence for the involvement of apoptosis-inducing factor-mediated caspase-independent neuronal death in Alzheimer disease. *The American Journal of Pathology* **176**: 2209–2218. PMID: 20228227

Yu Z, Fotouhi-Ardakani N, Wu L, Maoui M, Wang S, Banville D, Shen S (2002) PTEN Associates with the Vault Particles in HeLa Cells. *Journal of Biological Chemistry* **277**: 40247 –40252

Zambonin JL, Zhao C, Ohno N, Campbell GR, Engeham S, Ziabreva I, Schwarz N, Lee SE, Frischer JM, Turnbull DM, Trapp BD, Lassmann H, Franklin RJM, Mahad DJ (2011) Increased mitochondrial content in remyelinated axons: implications for multiple sclerosis. *Brain: A Journal of Neurology* **134**: 1901–1913. PMID: 21705418

Zandian M, Mott KR, Allen SJ, Chen S, Arditi M, Ghiasi H (2011) IL-2 suppression of IL-12p70 by a recombinant HSV-1 expressing IL-2 induces T-cell auto-reactivity and CNS demyelination. *PloS One* **6**: e16820. PMID: 21364747

Zheng J, Bizzozero OA (2011) Decreased activity of the 20S proteasome in the brain white matter and gray matter of patients with multiple sclerosis. *Journal of Neurochemistry* **117**: 143–153. PMID: 21235577

Appendix

Graphs from polyclonal Phage ELISA

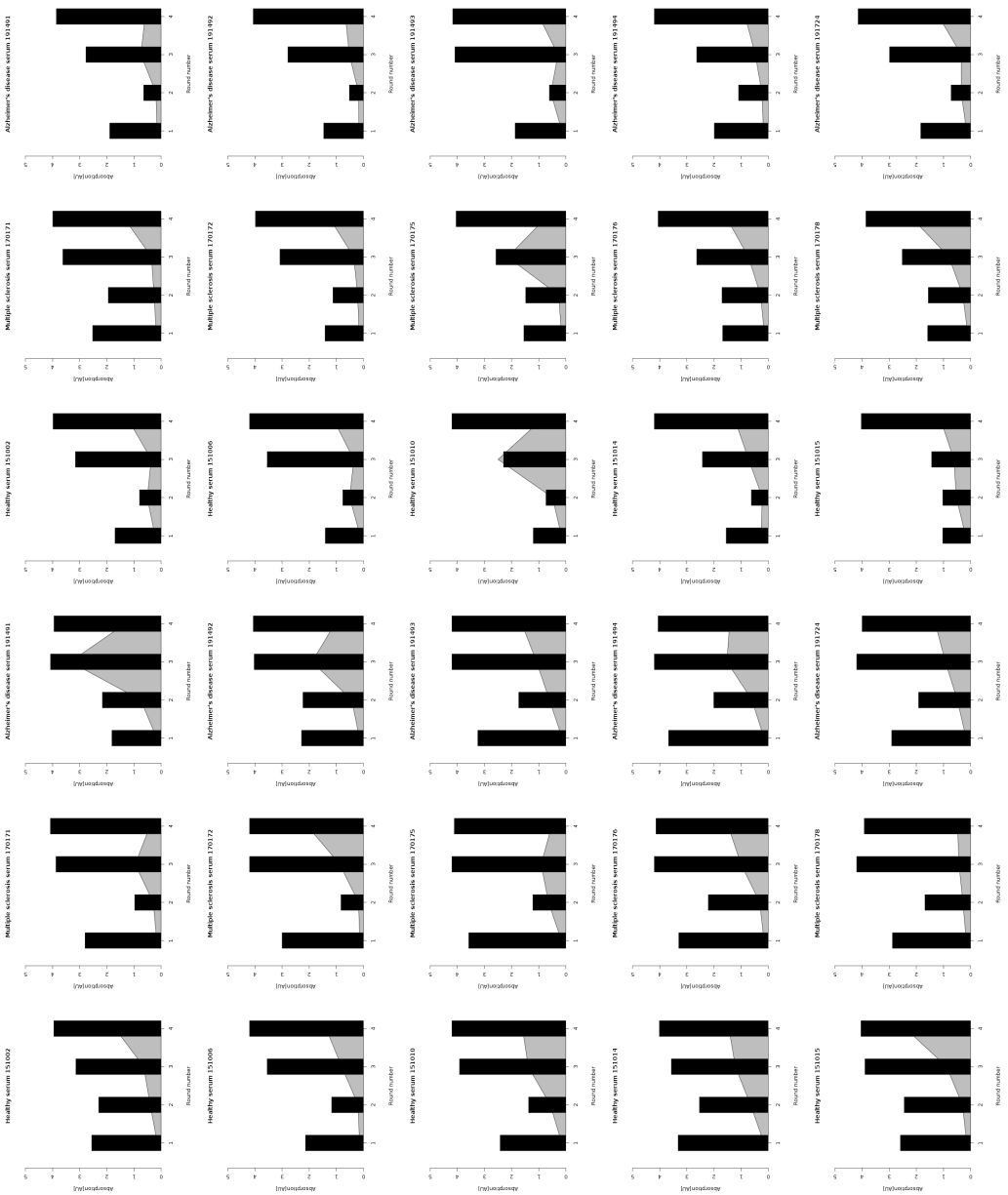

Figure 27: Enrichment profiles for ER2738

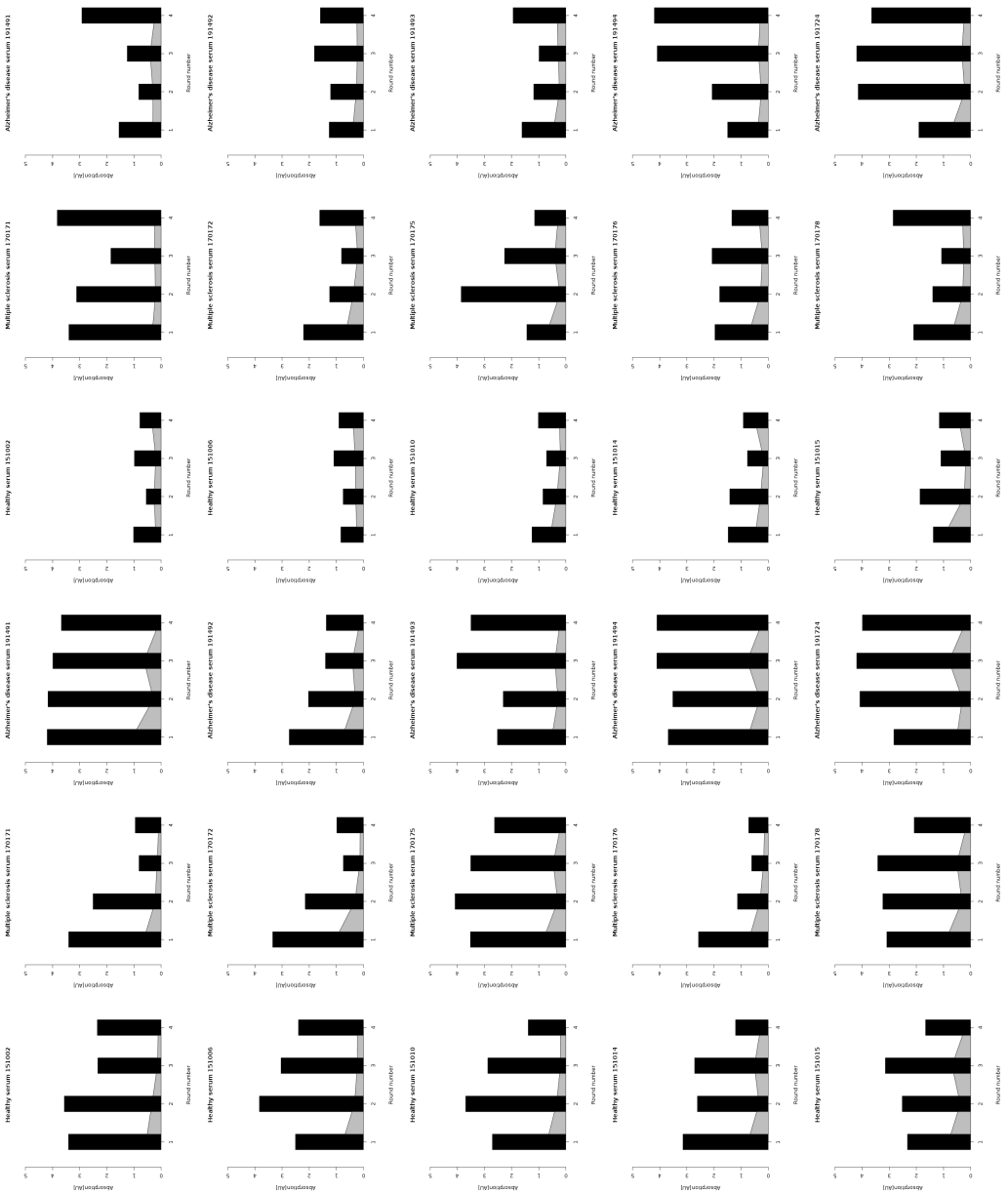

Figure 28: Enrichment profiles TG1

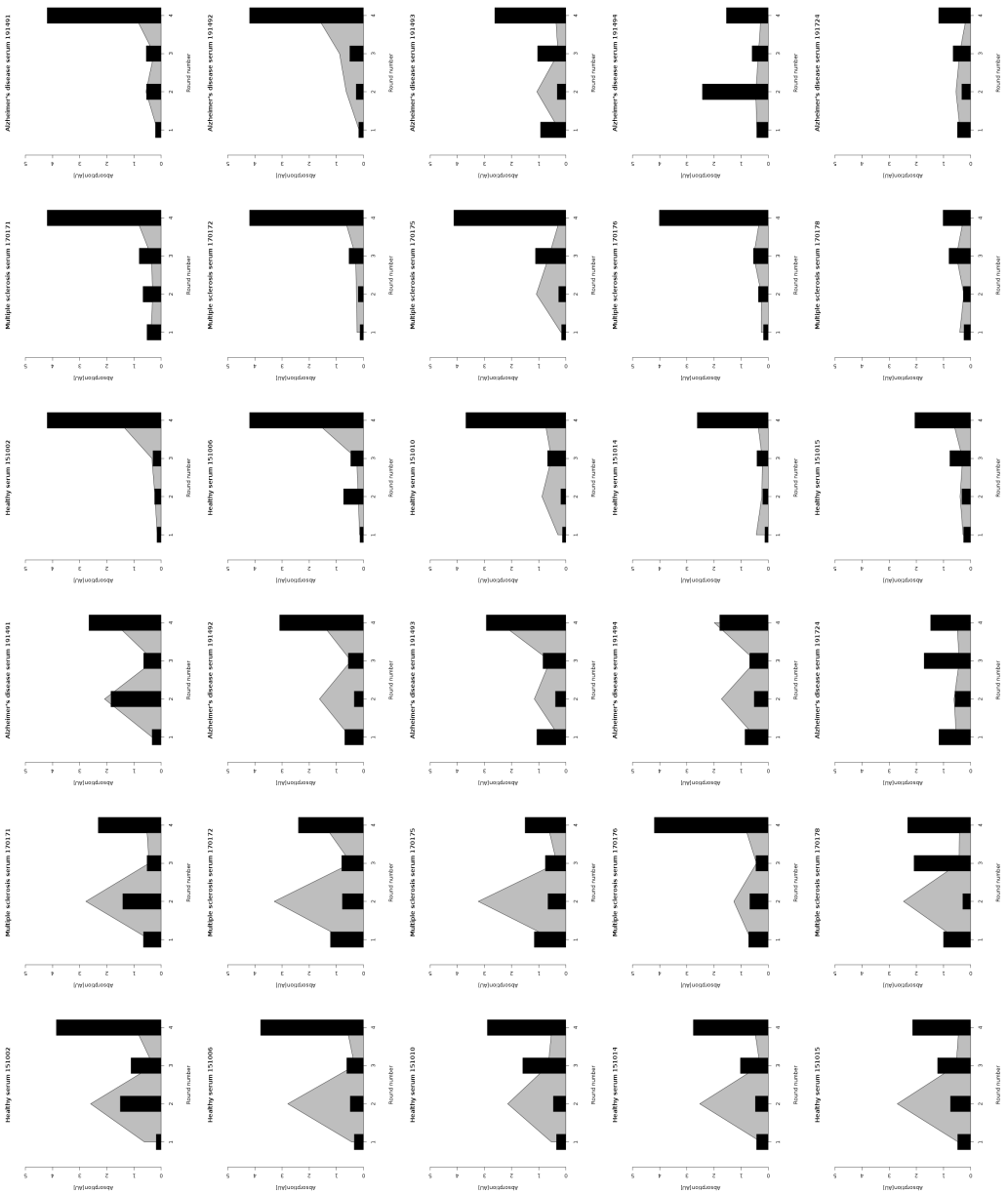

Figure 29: Enrichment profiles for XL-1

List of enriched genes from selection round four

ENTREZ Symbol	Disease	#initial library	#enriched	CDS
RPS26	MS	2	11	X
MTRNR2L1	MS	9	20	X
LZTS2	MS		3	X
PSMC6	MS		733	X
RABAC1	MS		2	X
RPL21	MS	5	31	X
MTRNR2L8	MS	65	159	X
RPL7	MS	24	41	X
MTRNR2L2	MS	30	161	X
SDHB	MS		62	X
SPATS2	MS		2	X
ATP5J2	MS	2	3760	X
TAGLN	MS	2	3570	X
DCTN1	MS		378	X
PKM2	MS		1309	
LOC100506583	MS	2	198	
RPL21P28	MS		6	
LOC100507680	MS		11745	
LOC100507713	MS		435	
PLLP	MS	14	946	
LOC100505479	MS	72	31	
PICALM	MS	3	8	
PEA15	MS	3	67	
ANKRD54	MS		14	
AIF1L	MS	3	2184	
ZNFX1	MS		6	
PPP1CB	MS	5	4	
INPP5K	MS		30	
RN28S1	MS		2	
CALM2	MS	2	358	
TMSB4X	MS	3	33	
CLSTN1	MS		4162	
CNP	MS		9	
KIAA1045	MS		24	
EDEM1	MS		900	
WDR47	MS		6	

GNL1	MS		202	
DYNLRB1	H	2	44	X
MTRNR2L2	H	30	3	X
GAPDH	H	8	43	X
GRM3	H	2	3083	X
CPE	H	2	102	X
C10orf116	H	28	49	X
METRN	H		9	X
MTRNR2L1	H	9	81	X
TAGLN	H	2	1160	X
TMSB10	H	4	3	X
PPP1R14A	H		6	X
LOC100505479	H	72	11	
GABRA4	H		66	
NAALAD2	H		15007	
ZBTB38	H	3	21	
SREK1	H		16010	
MTRNR2L10	H	37	34	
NTRK2	H		59	
BRP44L	H		4	
STMN4	H	2	45	
MTRNR2L3	H	37	34	
MVP	AZ		92	X
IFITM3	AZ	5	52	X
HBB	AZ	2	13	X
TIAF1	AZ		83	X
DDIT3	AZ		6	X
PRSS2	AZ	4	2	X
GNAS	AZ	2	16	X
FAM195B	AZ		264	X
HADHA	AZ		409	X
LOC100134397	AZ		38	X
LOC100291917	AZ		72	X
TMSB10	AZ	4	53	X
LOC642131	AZ		38	X
LYRM4	AZ		479	X
TUBB2A	AZ		849	X
MIF	AZ		70	X
MTRNR2L1	AZ	9	654	X
MTRNR2L2	AZ	30	338	X
MTRNR2L8	AZ	65	158	X
RABAC1	AZ		10	X
CYCS	AZ		4	
RTN3	AZ		36	
TXNIP	AZ	2	100	
ZDHHC17	AZ		5	

EDEM1	AZ		408
SRP19	AZ		8
SQSTM1	AZ	9	3987
SPTBN1	AZ		1996
JAKMIP3	AZ		2596
LOC100131754	AZ	20	2196
QSOX2	AZ		426
PPP4R1	AZ		6
MAP7	AZ		20
PPME1	AZ		383
PKM2	AZ		28
NPTXR	AZ		11
MYO18A	AZ		166
SOD1	AZ	9	41

List of Figures

List of Figures

List of Tables

List of Tables

Acknowledgements

First of all I want to thank Prof. Dr. Wolfgang Lockau for accepting the supervision of this work.

I want to thank Dr. Zoltán Konthur for giving me the oppurtunity to work on an exciting topic in his research group and for his valuable advice. I also want to thank all group members for their help in daily lab routine, especially my tutor Yuliya Georgieva.

My special thanks goes to Friederike Braig and Arndt Grossmann for their support and being good friends in the lab during all complications along the way.

Last but not least I want to thank Don Finkle for helpful tips with the english language.

Eigenständigkeitserklärung

Hiermit versichere ich, Sunniva Förster, dass ich die vorliegende Diplomarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Berlin, den 8.12.2011