Interaction Networks and perturbed cellular Functions in Schizophrenia

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

Konrad Klockmeier

from Karl-Marx-Stadt

2019

1. Gutachter: Prof. Dr. Peter R. Hiesinger

2. Gutachter: Prof. Dr. Erich E. Wanker

Date of PhD defense: 15th November 2019

Table of content

1. Abstract

Schizophrenia (SCZ) is a devastating psychiatric disease with a worldwide prevalence of approximately 1%. It is therefore one of the leading causes of public health burden. Twin studies suggest that the impact of heritable factors causing SCZ is very high. Thus, in the recent past huge efforts were made to identify the genetic factors responsible for the disease. Many genes and genomic variations have already been associated with schizophrenia, but the interplay of these genes, as well as the precise mechanism of how they are involved in the development of schizophrenia is still not fully understood.

To get a deeper insight into the role of SCZ associated genes I used protein-protein interaction analyses combined with bioinformatical methods. My goal was to answer three mayor questions:

The first question was, if schizophrenia associated proteins form clusters within protein-protein interaction networks and how these clusters are involved in functional processes. For that reason, a propagation-based algorithm was invented that identified five clusters with high potential for SCZ relevance. The two highest scoring clusters represented known synaptic complexes and were validated with LuTHy assays.

The second question was, if there is a potential SCZ relevance of a set of 39 protein coding candidate genes of a small exome sequencing study and if their importance could be prioritized. Therefore, a protein-protein interaction network was created, using the HIPPIE database, including all medium high confident interactions of these genes. In a next step the density of SCZ associated proteins within the created network were compared to all HIPPIE proteins, not already included in the created network and their connectivity to SCZ related proteins. Chi-squared tests revealed indeed a significant enrichment of schizophrenia associated proteins within the created candidate protein-protein interaction network. In order to rank candidate genes, the browser based ToppNet tool was used.

The third question should shed light on the functional role of ZNF804A. This protein had repeatedly been associated with schizophrenia before, but its functional role remained unclear. By following the hypothesis "guilt by association", a proteome

scaled Y2H screen was preformed and 18 new ZNF804A interacting proteins had been been identified with functional enrichment for RNA binding, the circadian clock and inflammation pathways. By using DULIP and LuTHy assays, 67% of identified ZNF804A interactions were validated. The functional implications of ZNF804A with the most promising interaction partner STAT2 were further analyzed. STAT2 is a key protein of the intracellular interferon response and ZNF804A was identified to co-translocate with STAT2 into the nucleus upon interferon induction. Overexpression, as well as CRISPR/Cas9 induced knock down of ZNF804A indicated a potential modulating role of ZNF804A in STAT2 mediated interferon response.

The results of my work help to better understand the role of SCZ related genes and their interplay. Additionally, my studies demonstrate that protein-protein interaction analyses are able to gather information on different levels and are a key tool set to reveal the molecular implications of genes associated with schizophrenia.

2. Zusammenfassung

Schizophrenie ist eine dramatische, psychische Erkrankung, die weltweit schätzungsweise 1% der Bevölkerung betrifft. Es ist damit eine der gravierendsten Belastungen für die öffentliche Gesundheit. Zwillingsstudien implizieren einen hohen Einfluss von genetischen Komponenten für die Manifestation von Schizophrenie. Dies ist der Grund, warum große Anstrengungen in den letzten Jahren unternommen wurden, um diese genetischen Faktoren zu identifizieren. Viele Gene und genetische Variationen wurden bereits mit Schizophrenie assoziiert, jedoch ist deren Zusammenwirken und der Mechanismus, durch den sie zur Entwicklung von Schizophrenie führen, noch immer nicht vollständig verstanden.

Um ein tiefergehendes Verständnis für die Rolle von Schizophrenie assoziierten Genen zu erlangen, habe ich sowohl Protein-Protein Interaktionsanalysen als auch bioinformatische Analysen verwendet. Mein Ziel war die Beantwortung von drei Hauptfragen.

Die erste Frage war, ob Schizophrenie assoziierte Proteine Cluster in einem Protein-Protein Interaktionsnetzwerk bilden und wie diese Cluster involviert sind in funktionelle Prozesse, welche für Schizophrenie relevant sein könnten. Aus diesem Grund wurde ein auf Ausbreitung basierter Algorithmus entwickelt, durch welchen fünf Cluster identifiziert wurden, mit wahrscheinlich hoher Relevanz für Schizophrenie. Die zwei Cluster mit den höchsten Assoziationswerten waren bekannte synaptische Komplexe und wurden mit LuTHy Experimenten validiert.

Die zweite Frage war, ob ein Set von 39 Schizophrenie-Kandidaten Genen eine potenzielle Relevanz für Schizophrenie hat und diese nach ihrer Relevanz zu priorisieren. Zu diesem Zweck wurde erneut ein Protein-Protein Interaktionsnetzwerk auf der Grundlage der HIPPIE Datenbank erstellt, mit allen medium- bis hoch validen Interaktionen der Kandidaten Gene. Die Dichte an Schizophrenie assoziierten Proteinen in dem erstellten Netzwerk wurde dann verglichen mit der Dichte an Schizophrenie assoziierten Proteinen in der restlichen HIPPIE Datenbanknetzwerk. Mittels Chi-quadrat Test wurde eine signifikante Anreicherung von Schizophrenie assoziierten Proteinen im kreierten Kandidaten

Protein-Protein Interaktionsnetzwerk ermittelt. Eine Priorisierung der Protein codierenden Gene wurde mittels des Onlinetools ToppNet erreicht.

Die dritte Frage sollte die funktionelle Rolle von ZNF804A weiter aufklären. ZNF804A ist ein Protein, welches wiederholt mit Schizophrenie assoziiert wurde, dessen funktionelle Rolle jedoch noch relativ unbekannt ist. Der Hypothese „Schuldig durch Assoziation" folgend, wurde eine proteomweite Hefe-Zwei-Hybrid Analyse durchgeführt und 18 neue ZNF804A interagierende Proteine identifiziert. Diese wiesen eine funktionelle Anreicherung für RNS Bindung, circadianen Rhythmus und Infektions-Signalwege auf. Mit Hilfe von DULIP und LuTHy Experimenten konnten 67% der identifizierten ZNF804A Interaktionen validiert werden. Der funktionelle Einfluss von ZNF804A auf den vielversprechendsten Interaktionspartner STAT2 wurde tiefergehend analysiert. STAT2 ist ein wichtiger Mediator der intrazellulären Antwort auf Interferon. ZNF804A zeigte nach Induktion mit Interferon die Eigenschaft zusammen mit STAT2 in den Zellkern zu migrieren. Überproduktion sowie die CRISPR/Cas9 induzierte Eliminierung von ZNF804A zeigten eine regulierende Rolle von ZNF804A auf die STAT2 vermittelte Antwort der Zelle auf Interferon.

Die Resultate meiner Arbeit helfen, die Rolle von Schizophrenie assoziierte Gene und ihr Zusammenspiel besser zu verstehen. Zusätzlich demonstriert sie, dass Protein-Protein Interaktionsanalysen ein bedeutsames Werkzeug sind, um auf verschiedenen Ebenen Informationen über die molekularen Implikationen von Schizophrenie assoziierten Genen zu erlangen.

3. Introduction

3.1 Schizophrenia

Scientific researchers identified hallmarks of functionality and new therapy strategies for many fields of molecular medicine in the last couple of decades. One field still in the beginnings of being the focus of researchers are neuropsychiatric diseases like Schizophrenia (SCZ), Bipolar Disorder (BD) and Major Depression (MD). One of the most devastating among these is schizophrenia. SCZ is not a discrete illness but rather a complex neuropsychiatric disorder with multiple interacting genetic and environmental causes and a wide variety of symptoms and disease progressions.[1–4]

The estimated prevalence of SCZ is approximately 1% worldwide and is therefore one of the leading causes of public health burden.[5,6] The economic burden of SCZ in the United States alone was estimated $156 billion in 2013.[7]

This estimation included health care costs of $38 billion, direct non-health care costs of $9 billion, and indirect costs of $117 billion.[7] Additionally, to the mental health burden for the patients, the life expectancy of subjects with schizophrenia is about 20 years shorter compared to the general population, which means that schizophrenia causes more life time loss than most cancers and other physical illnesses.[8,1] The onset of schizophrenia is typically in adolescence or early adulthood.[1]

3.2 Symptoms of SCZ

Schizophrenia is a diverse disorder and there is not a single symptom that is essential for the disease, but the term schizophrenia rather describes a syndrome that is characterized by delusions, as well as negative symptoms (described below).[2,9]

Individuals with SCZ can develop a variety of manifestations and symptoms but in general it is a chronic psychotic disorder that disrupts the thoughts and flattens the affect of patients.[2] The symptoms are classified as positive (perceptions healthy people don't have), negative (abilities/perceptions patients lack) and cognitive.[1,4,10]

3.2.1 Positive symptoms

Positive symptoms can be described as "psychotic behaviors not seen in healthy people" such as delusions, hallucinations, lack of insight about their mental state and varying degrees of abnormal motor behavior up to catatonia.[11]

3.2.2 Negative symptoms

The most common negative symptoms are reduced or diminished emotional expression (emotional blunting/flattened affect) and decreased goal driven behavior or drive (avolition).[4] Other possible negative symptoms are a general lack of unpromoted content in speech (alogia) and the reduced or inability to experience pleasure (anhedonia).[4]

3.2.3 Cognitive symptoms

Cognitive symptoms are the newest category of symptom classification and are non-specific. Cognitive symptoms include disorganized speech, thought, and/or attention, often impairing the ability of the patient to communicate.[9]

Psychiatrists often categories SCZ patients, regarding the presence and dominance of different symptoms in four major subgroups of SCZ (Table 1).[10]

Table 1: Categories of SCZ and their symptoms

TYPE	PROMINENT SYMPTOMS
PARANOID	• Prominent delusions or hallucinations
HEBEPHRENIC	• Constantly flattened or inadequate affect • Lack of goal directed behavior • Prominent thought disorder
CATATONIC	• Long lasting catatonic behaviour
SIMPLE	• Considerable loss of drive • Progressive sevirety of negative symptoms • Dominant decline in social and academic/employment performance

Schizophrenia is considered a complex disorder with multiple factors contributing to the pathogenesis. Among these are environmental factors such as apoxia during birth or birth complications in general, pre-natal infections, for example with toxoplasmosis or drug abuse.[12,13] Additionally, a high heritability of estimated 60-80% has been consistently shown by family and twin studies.[14] This points to a major role for inherited genetic variants in the etiology of schizophrenia. The connection between how the heritable factors of SCZ lead to the development of the disease is still not clear and numerous hypothesis are debated to be causative.

3.2.4 Hypothesis of SCZ

Several neurotransmitter systems and functional networks within the brain have been found to be affected in patients with schizophrenia, but the question is still not completely resolved if those are causative for the disease or if they are simply consequences of disease progression or of treatment. Three major neurotransmitter based hypothesis have arose in the years, namely the dopamine-, glutamatergic- and the GABAergic hypothesis.

3.2.5 Dopamine hypothesis

The most widely accepted neurotransmitter-based hypothesis of SCZ is the dopamine hypothesis.[15,16] In this concept, symptoms of SCZ may result from excess dopaminergic neurotransmission, mainly in the strial and mesolimbic regions of the brain.[17] This excess is believed to lead to positive symptoms and a deficit of dopamine transmission in prefrontal brain regions, which may result in the appearance of negative symptoms.[18] The main dopamine pathway plays a major role in reward-motivation behavior.[19] Other pathways in which dopamine is involved are motor control, as well as the release of different hormones. The dopamine hypothesis is also favored by the SCZ research community because all common antipsychotic medications available are at least partial agonist of the D2 receptor, which is the main site of dopaminergic action.[20]

3.2.6 Glutamatergic hypothesis

There are reports as far back as 1949 of patients with schizophrenia being treated with glutamic acid.[21] The glutamatergic hypothesis postulates, that the dopaminergic dysfunction in SCZ is secondary to an underlying glutamatergic dysfunction. Following this hypothesis, a hypofunction of glutamate signaling in the

cortical-striatal projections leads to an opening effect in the thalamo-cortical loop resulting in a sensory overload which induces psychotic symptoms, including the dopaminergic dysregulation, suggesting that dopamine levels might play a major role in SCZ.[18,22] Glutamate receptors are grouped into ionotropic ligand-gated in-channels and metabotrophic g-protein-coupled receptors, which are subdivided into the alpha-amino-3-hydroxy-5-methyl-4- isoxazole-propionic acid (AMPA) kaniate and N-methyl d-aspartate (NMDA)-receptors.[23] Especially the latter are hypothesized to play a key role in the development of SCZ. This Hypothesis arose from the observation that non-competitive NMDA receptor antagonists led to psychological effects, which closely resemble some positive, as well as negative symptoms that occur in schizophrenia. Those drugs include phencyclidine (PCP), dizocilpine (MK-801) and ketamine which are mainly known as addictive substances.[24–26] In general, the excitatory neurotransmission in the brain is primarily glutamatergic, with estimation of glutamatergic neurons utilizing between 60-80% percent of total brain metabolic activity.[27]

3.2.7 Synaptic pruning hypothesis

The Synaptic pruning hypothesis arises from the fact that SCZ is typically beginning in adolescence or early adulthood. At this time a process called synaptic pruning, a process that shapes the neuronal network to a fully functional state by reducing the synapse density in the central nervous system, leads to the ability of abstract thinking.[28] The reduction of synapse density is caused by a controlled synaptic elimination. An increasing number of researchers speculated that SCZ may be a consequence of intensively increased elimination process.[29–31] In consequence of the reduced synaptic interconnectivity, brain functions would be severally distracted which might lead to the development of SCZ.

3.2.8 Immune hypothesis

Multiple lines of evidence indicate that severe infections and immune disorders represent an additional risk factor for the development of SCZ.[32,33] The hypotheses that prenatal infections are increasing the risk of developing SCZ for the offspring is caused by the observation that the risk of developing SCZ is increased for persons born during influenza epidemics.[34] However, not only influenza seems to increase the risk for the development for SCZ. Several studies showed an

increased risk for SCZ in the offspring of patients suffering from respiratory, genital and reproductive tract infections as well as several viral induced disorders and especially for *Toxoplasmosis gondii*.[35–40] Epidemiologicalists associated SCZ with influenza after analyzing population data. Additionally, signs of inflammation are commonly found in brains of SCZ patients. Animal models have shown that stimulations of the maternal immune system during pregnancy by viral or bacterial agents lead to SCZ–like symptoms in the resulting offspring. Furthermore, a large-scale epidemiological psychiatric study showed that severe infections and autoimmune disorders increase the risk of SCZ and schizophrenia spectrum disorders for the patient itself.[41–44]

Additionally, degradation products of inflammatory substances have been described in brain, as well as cerebral spine fluid of about 50% of SCZ patients.[45,46] As a consequence, antibiotics and anti-inflammatory agents have been tested to treat schizophrenia. However, the success was limited

3.3 Identifying the genetics of SCZ

Most of the hypotheses regarding SCZ are based on clinical and psychological observations, but the molecular and genetic mechanisms, leading to or increasing the risk to develop SCZ are still not clearly identified. The heritable factors, leading to SCZ are estimated to be very high, according to twin studies, additionally emphasizing the importance to unravel the genetics of SCZ.[14]

Therefore, several large scaled genomic studies have been carried out in order to identify SCZ candidate genes in mainly three major study types. The most effort up to day was spend in genome wide association studies (GWAS).[11,47] Particularly, given that the GWAS approach provides opportunities to study common genetic variations across the entire genome without an *a priori* hypotheses, many novel but seemingly less relevant risk variants with functions beyond the known disease biology have been discovered.[19] In the largest study till now, the PGC (Psychiatric Genomic Consortium) analyzed over 37,000 schizophrenic patients and compared them to over 110,000 control individuals. They were able to identify 108 genome wide significant loci within the SCZ patient cohort. Unfortunately, the

effect size of each of the identified SNIPs (Single-Nucleotide Polymorphism) is low, as well as the penetrance of each single SNIP.[3]

Of the 108 loci, 40% include single protein coding gene regions and a total of 75% include protein coding genes. For an additional 8% genomic regions were within 20kb of a gene. The highest scoring associated loci was an extended MHC region (Major Histocompatibility Complex), which is coding for antigens, presented on the surface of cells, important for the acquired immune system to recognize foreign molecules. Other prominent genes within the range of identified SNIPS included DRD2 (Dopamine Receptor D2), the target of all effective antipsychotic drugs. Genes involved in glutamatergic neurotransmission and synaptic transmission like GRM3 (Glutamate Metabotropic Receptor 3), GRIN2A (Glutamate Ionotropic Receptor NMDA Type Subunit 2A), SRR (Serine Racemase) and GRIA1 (Glutamate Ionotropic Receptor AMPA Type Subunit 1), as well as genes encoding for voltage-gated calcium channel subunits.

Genetic variations with a higher effect size and penetrance are copy number variation (CNV), identified in CNV studies. In total 8 CNVs were successfully associated with SCZ.[11,48] The drawback of this kind of studies is, that most CNVs span over multiple genes and it is not easy to identify which of those genes are most important or if at all relevant for the disease. Only one CNV is described for SCZ that affects a single gene, i.e. NRXN1 (Neurexin 1).[49] The CNV with the highest penetrance of is 22q11.2 del a partial deletion of the chromosome 22.[3,48] This specific CNV is also described as 22q11.2 Deletion Syndrome and besides SCZ, individuals carrying this genetic variations have a risk to develop a psychotic disorder in general of 41%.[50] The last study type of studies used to identify SCZ candidate genes is the whole genome sequencing, a rapidly growing field in the world-wide society of genomic research. The two largest studies cover the two sides of SCZ exome sequencing.[51,52] Former and colleagues focused their efforts on affected families and sequenced affected patients and their ancestors to identify de novo mutations.[51] Purcell and colleagues on the other hand compared huge cohorts of patients with control individuals to identify enrichments in missense mutations in their exome sequencing work.[52] Even though the sample sizes of those studies is impressive, no significant SCZ point mutation was identified up to day via exome sequencing. Together, these types of studies identified hundreds of

genes potentially associated SCZ with rather low individual impact for the disease. Although this was a huge step towards understanding SCZ, those genes did not directly indicate one or more common mechanisms dysregulated in SCZ. In order to discover new drugs and improve the treatment of SCZ, identifying those mechanisms is of the most importance.

A disease is rarely the consequence of a single gene abnormality, which is especially true for SCZ. Complex diseases and disorders are rather the reflection of the interplay of multiple molecular processes.[53] This includes all physiological interactions in the cell including protein-proteins, regulatory protein-DNA and metabolic interactions.[53] Specifically, disease associated proteins tend to interact with proteins, associated with the same disease or cluster in the same local neighborhood in an PPI (Protein-Protein Interaction) network, as well as in networks of other relationship propensities.[53] This led to the hypothesis that disease proteins form so called disease modules in a given networks.

3.4 Protein-protein interactions (PPI) – Proteins don't act alone

One powerful approach for inferring information about protein function is to identify protein-protein interactions, as proteins rarely act alone in the cell. Rather they connect spatiotemporally with other proteins to perform their specific tasks.[54] Consequently, identifying the protein interaction partners of a protein provides crucial insides into its function in health and disease.

There are many ways and methods to study PPIs available and still more are developed. In general, two main methods are suitable to identify PPIs in a systematical and proteome wide matter up to day, the yeast-two hybrid (Y2H) method and mass-spectrometry (MS).[55] Besides those two approaches there are different methods available to study PPIs. In the following, methods are summarized used in this thesis.

3.4.1 Yeast two-hybrid (Y2H)

The development of the Y2H in 1989 by Stanley Fields and Ok-Kyu Song enabled scientist the first time to analyze binary interactions in such an efficient way, that soon afterwards the first proteome scaled approaches to map binary interaction

were performed.[56–58] The general principle of Y2H is based on the reconstitution of a transcription factor which was split in two halves and fused to two proteins of interest. Fields and Song invented this method by splitting the transcription factor GAL4, a yeast native protein, in its N-terminal DNA binding domain (DBD) and the C-terminal transcriptional activation domain (TAD).[56] Each domain is separately fused to proteins of a potential binary PPI pair of interest. The DBD domain binds to the upstream activation sequence (UAS) but is unable to trigger transcriptional activation. Only after the two proteins of interest interact in the yeast nucleus, the TAD domain is in close proximity to the DBD domain, resulting in the transcription of reporter genes. Reporter genes activated by the restored transcription factor are proteins which allow easy detection of positive interactions such as enzymes in an essential amino acid pathway as auxotrophic markers or for the ß-galactosidase (ß-GAL) enzyme, allowing to quantify ß-Gal induced blue color change in growing yeast colonies. As auxotrophic markers, HIS3 or URA3 are used and enable to select for colony growth on selective media. But a variety of different variants of this principle were established since the discovery and are now available and in use all over the world.[59,60]

Although Y2H is still the most efficient method to map binary interactions, nowadays the Y2H is widely believed to have a relatively high rate to detect false positive PPIs. This might be attributed to the lack of appropriate correction and usage of controls in the past, while benchmarking approaches showed that different Y2H variants exhibit false positive rates between 1-4.5%, which is comparable to other PPI methods.[61–63] Nevertheless, a clear downside of Y2H is the arbitrary environment of the nucleus of yeast cells in which the PPIs must take place. Therefore, for example the interaction between proteins of which one or both are membrane bound is nearly impossible.

3.4.2 Luminescence-based co-immunoprecipitation assays

Luminescence-based co-immunoprecipitation assays uses the principle of co-immunoprecipitation (CoIP). In classical CoIP experiments positive interactions and CoIP efficiencies are determined via western-blotting. This represents a time-consuming effort which is limited by the scale in which PPIs can be analyzed, as well by its quantifiably. The LUMIER (LUminescence-based Mammalian

IntEractome Mapping) assay was developed to overcome these limitations.[64] In LUMIER assays, one protein of interest (prey) is fused to the *Renilla* luciferase (RL) and are co-produced in mamallian cells with FLAG-tagged proteins (bait). After lysis, the FLAG tag of the bait protein is used for immuno precipitation (IP) and the RL tagged prey protein is co-immunoprecipitated if both proteins interact. The final read out is the RL enzymatic activity, providing a semi-quantitative read out. One of the most recent variations of the Lumier assay is DULIP (DUal Luminescence-based co-Immuno Precipitation).[65] DULIP uses a set of two luciferases, the prey protein is fused to the fire fly luciferase and the bait protein is fused to a PA-*Renilla* luciferase. Both constructs are co-produced in mammalian cells and the luminescence of both luciferases can be measured separately in lysates after co-IP. This two-luciferase strategy enables to quantify the pray, as well as the bait fusion protein of interest. With the usage of a tandem construct of both luciferases for normalization purposes, as well as the intense usage of controls, DULIP enables to rank PPIs and estimate binding affinities.[65]

3.4.3 Bioluminescence Resonance Energy Transfer (BRET)

BRET is a physical phenomenon similar to fluorescence resonance energy transfer (FRET). In both cases energy is transferred from either an exited donor fluorophore for FRET or an enzymatically active donor luciferase in case of BRET to another molecule, the acceptor, by long range dipole-dipole interactions in a non-radioactive way Lakowicz2006. Both, BRET and FRET can only occur when donor and acceptor are in very close proximity (<10nm) which makes them good principles to use for PPI detection.

Both, FRET and BRET assays give semi-quantitative values, not only dependent on interaction strength, but also mostly dependent on the distance of donor to acceptor. The use of saturation experiments, where the quantity of one interaction partner is kept constant while the production of the other interaction partner is stepwise increased, allow the calculation of relative binding affinities in cells.

A very advanced BRET based method, recently established is LuTHy (bioLUminescence-based Two-HYbrid technologies).[66] LuTHy uses the very bright NanoLuc luciferase as a donor-tag together with mCitrine as a fluorescence acceptor-tag, providing reliable and sensitive PPI detection at very low detection

levels. The major novelty of LuTHy is the combination of the classical in cell BRET readout for detecting PPIs with a subsequent second readout, similar to DULIP, using the principle of co-immunoprecipitation.[66]

3.5 Guilt by association – PPI networks infer information on SCZ associated proteins

The identification of PPIs for given disease related proteins are the basis to identify mentioned disease modules. Examples for the identification of disease modules or simple PPI network analysis are plenty and were also done for SCZ. For example, Former *et al.* mapped their identified synaptic SCZ associated proteins from the biggest de novo study up to day via the SynSyNet database (http://bioinformatics.charite.de/synsys/) to create a mainly postsynaptic network which displays the localization and interplay between the identified proteins.[51] Ganapathiraju *et al.* developed a computational model to discover PPIs, which they state to be highly accurate.[67] They used their model to create a SCZ specific PPI network on the basis of the GWAS results of the PGC.[11,67] Many of the PGC identified genes lack a high coverage of their PPIs, therefor the approach by Ganapathiraju described 504 new PPIs for the PGC SCZ protein dataset. For example, seventeen proteins without described PPIs were identified to have a total of 54 predicted PPIs. Schwarz et al used PPI network analysis to analyze common function of as well as PGC dataset based SCZ related genes and found it to be significantly enriched for synaptic function.[67] In a similar approach, Luo *et al.* analyzed SCZ associated proteins of different sources using PPI network analysis. They discovered that top SCZ susceptibility genes form a highly interconnected PPI network with an enrichment for nucleosome assembly genes.[68]

3.6 PPI cluster generation – bioinformatical approaches to find densely connected groups of proteins

Since the beginning of high throughput PPI identification in the late 1990s and the early 2000s a variety of different computer based algorithms have been developed to identify clusters of connected proteins within a PPI network.[69] Clusters can be

described in context of PPI networks as groups of proteins that share a larger number of interactions. The most commonly used can be categorized into mainly three major groups. 1) Local neighborhood Density search (LD); 2) Cost-based Local search (CL) and 3) Flow Simulation (FS).[69]

3.6.1 Local neighborhood density search

Methods of this sub category are optimized to find dense subgraphs (networks) within the PPI network to maximize the density of each found subnetwork.

Molecular Complex Detection (MCODE)[70]

As one of the first computational methods Bader and Hogue developed MCODE for complex detection from PPI networks. MCODE works in principle in two stages, vertex (for PPI network proteins) weighting and complex prediction. In the first step, MCODE measures the density of interactions within the neighborhood of each vertex (protein), which is also called its "cliquishness". A clique is a complex of for example proteins in which each protein is connected to all the other proteins within the   same   complex. A half click would be a complex in which all proteins would be connected to at least half of all the other proteins within this complex. This MCODE strategy amplifies the weighting of densely connected regions within the network using the *k-core* concept. A *k*-core is the most densely connected subgraph of a network. In the second step the highest scoring vertex is used as a seed and ads additional vertexes with a given percentage of the initial seed weight until no more vertexes surpasses these criteria and no vertex can be added. Then the algorithm uses the next highest weighted vertex as seed and repeats this process. It is to note that each vertex added to a complex is excluded from being added to another one, therefore the result are non-overlapping complexes.

CFinder[71]

The principle of CFinder is the idea that clusters can be interpreted as a conglomerate of smaller fully connected subgraphs (clicks) that overlap in some nodes (proteins). CFinder extracts all clicks within a network and merges neighboring clicks. The parameter used is k for the initial clique size (k-clique). The

higher the k-clique, the higher the stringency because clicks are estimated to be adjacent if they share at least k-1 nodes.

3.6.2 Cost based local search

Approaches based in this category divide the network in subgraphs within a network by a cost function that directs the search towards the best result

RNSC[72]

This approach computes all possible clusters and ends up with those that minimize a cost function that reflects the number of inter-cluster and intra-cluster edges. The network is initially randomly participated in multiple subnetworks. In the next step the algorithm moves nodes from one subnetwork to another until a cost function is optimized. All clusters not surpassing a minimum cluster size or minimum density are excluded. Finally, a p-value is calculated using GO functional annotation for each cluster, measuring the functional homogeneity of a cluster. All clusters with p-values above the set threshold are also discarded.

3.6.3 Flow simulation

Methods based on flow simulation use mainly random walk algorithms to mimic the spread of information between proteins in a network.

Macarov Clustering (MCL)[73]

The MLC algorithm performs a high number of random walks (called flow) of a defined step size from each node in the network. Densely connected regions within the network are more often passed by random walks and therefore have a higher flow.
Additionally, two operators called expansion and inflation control the random walks. Where the expansion models the spread of the flow via step-size and inflation making the flow in dense regions thicker and in sparse regions thinner.

PRINCE[74]

PRINCE start with a set of known proteins, called priors, in addition to a given network. Those priors can be disease associated proteins or proteins of a defined functionality. Like the MLC algorithm, PRINCE starts with a random walk algorithm but instead of using random starting points within the network, the random walks

start at the priors within the network. With this method each protein is given a score (propagation score) which represent the closeness and connectivity of each protein of the network to priors. In a next step clusters are generated via greedy search with a subsequent elimination of nodes if the elimination increases the overall scoring of the cluster while still maintaining the connectivity of the complex. The output are not clusters solely based on network properties but are instead disease associated clusters, if the priors represent known disease proteins, making this approach a powerful tool to identify protein clusters with a distinct association. Over time the lab of Roded Sharan further improved their clustering strategy and for example used random prior sets to define propagation score cut offs for each prior size and used propagation of random sets to assign p-values to each protein of the network, enabling to only include significantly higher propagation scoring proteins for a given prior set for further clustering.

NETBAG+ [75]

Gilman *et al.* invented a very interesting approach on cluster generation. Instead of a usual PPI network, Gilman used a network of proteins where the weighted edges represented the likelihood that genes share a genetic phenotype. Like PRINCE, NETBAG+ was based on a set of input genes. The cluster algorithm is based on a greedy search, starting at the input gene to find high scoring clusters and a significance of each cluster was reached by comparing to data collected with random input genes. Small generated clusters with five or less genes were removed.

3.7 Ranking and prioritizing newly identified SCZ related genes

The process of identification of SCZ relevant genes is still not far from booing complete and is an ongoing process.[3,11,51,52] Especially the field of exome sequencing is growing and increased sample sizes is needed to identify SCZ associated mutations that show genome wide significance.[3,76] For such studies it is very difficult to distinguish between disease relevant and irrelevant mutations.[77] Filtering those findings via predicting the disruptiveness of found mutations, as well as the potential phenotypic impact is a good method to reduce the amount of

potential SCZ mutations. Additionally, to these filtering strategies, the ranking of identified candidate genes is of high interest.[77]

Again, PPI network analysis are a great tool to identify candidate proteins that are highly connected to previously described SCZ genes. For this purpose easy-to-use free access online tools already exist for example ToppGene.[78]

Besides identifying the interplay of SCZ candidate genes and the implication in potential functional processes disturbed in SCZ via bioinformatic strategies, some of the identified SCZ candidate genes are still poorly studied and their functional implication in health and disease is unclear.[3,11]

As described above PPI analysis are a great tool to infer information about a target protein of unknown function. Since the research field of SCZ is in the process to identify numerous genes with implications in the process of developing SCZ, there is also the need to identify the mode of action this identified proteins influence SCZ. One prominent example is the first gene, reaching genome wide significands in GWAS, the zinc-finger protein 804A (ZNF804A).[11,47]

3.8 ZNF804A – A gene with clear association with SCZ and unclear functional role

The first genetic variant that achieved genome-wide significands in schizophrenia GWAS studies was the SNIP rs1344706 in the second intron of zinc-finger protein 804A ZNF804A.[47] The genome wide significance was further enhanced when both schizophrenia and bipolar disorder were considered as the associated phenotypes.[47] Rs1344706 was later repeatedly associated in further GWAS studies.[11,79–81]

ZNF804A is a protein coding gene on chromosome 2 encoding 4 exons and 1209 amino acids and is ubiquitously expressed, including the human brain.[82] Besides its name giving zinc-finger domain, no functional domain is described for ZNF804A. Since the initial identification of rs1344706 as a risk SNP for SCZ, additional intronic SNPs within ZNF804A were identified: rs7597593 and rs11693094.[11,79,83]

After its genome wide association, ZNF804A and especially rs1344706 was the focus of many studies, trying to unravel the potential impact for SCZ. Hill *et al.* identified analyzed the allelic expression of *ZNF804A* of rs1344706 carriers vs controls in fetal brains and identified a significant increased *ZNF804A* allelic expression in second trimester fetal brains.[84] Tao *et al.* conducted a patient vs control study using next generation sequencing. Interestingly, they did not find significant expression changes for rs1344706 variant carriers regarding the full length ZNF804A transcript in fetal brain.[85] Rather they identified a truncated version of ZNF804A, missing the first two exons, coding for 135 amino acids (*ZNF804A^E3E4*) which was abundantly expressed in the brain and was able to show that the rs1344706 risk variant lead to reduced expression of this newly identified variant *ZNF804A^E3E4* in fetal brain.[85] They further revealed that SCZ patients showed a significant lower *ZNF804A^E3E4* expression compared to healthy controls.[85]

Knockdown of *ZNF804A* in human neural progenitor cells (hNPCs) and in developing neurons derived from pluripotent stem cells indicated altered expression of genes involved in cytokine signaling, adhesion, synapse formation and neurite outgrowth.[86,87]

Dean *et al.* was the first to analyze the subcellular localization of ZNF804A in diverse human and rat neuronal models and were able to show co-localization of endogenous ZNF804A with the synaptic marker protein PSD-95 (postsynaptic density protein 95) in somatodendritic compartments.[88]

Clearly, ZNF804A has a distinct effect in the human brain with implications in developmental processes, whereby the mechanism and the distinct role of ZNF804A is unclear.

The research of ZNF804A and rs1344706 in the past years wasn't only restricted to the molecular implications. Additionally, many studies focused their interest on the effect on the brain itself, using brain imaging techniques. The rs1344706 variant was shown to reduce the connectivity in the dorsolateral prefrontal cortex (DLPFC) between hemispheres during working memory task as well as in resting state and induced cognitive state triggered by emotional recognition.[89–91]

Increased connectivity between the right DLPFC and the left hippocampal formation (HF) was exclusively observed for working memory tasks.[91]

Taken together, these data provided further evidence that *ZNF804A* modulates cortical network connectivity during executive cognition, further encouraged by studies analyzing theory of mind task which showed significant changes for the dorsomedial prefrontal cortex and the temporoparietal cortex.[92,93] Theory of mind tasks measure the ability to infer the mental state as part of the social cognition.

3.9 STAT2 – A new and very interesting identified ZNF804A **interactor**

The focus of the present thesis study regarding ZNF804A was to elucidate its cellular functionality via the identification of its PPIs. A protein library based yeast two-hybrid screens to create a PPI network for ZNF804A was performed and identified 18 new interaction partners of ZNF804A. PPIs were intensively validated using DULIP and LuTHy assays, and identified STAT2 as the most promising hit. The functional implication for ZNF804A that results of its interaction with STAT2 was the further focus of this thesis.

Signal Transducer and Activator of Transcription 2 (STAT2) is the most structurally and functionally divergent member of the STAT family.[94] The STAT2 gene contains 24 exons and the encoded protein has a molecular weight of 113 kDa.[94] The domain structure consists of seven different domains, present in all members of the STAT family (summarized in Figure 1).[94]
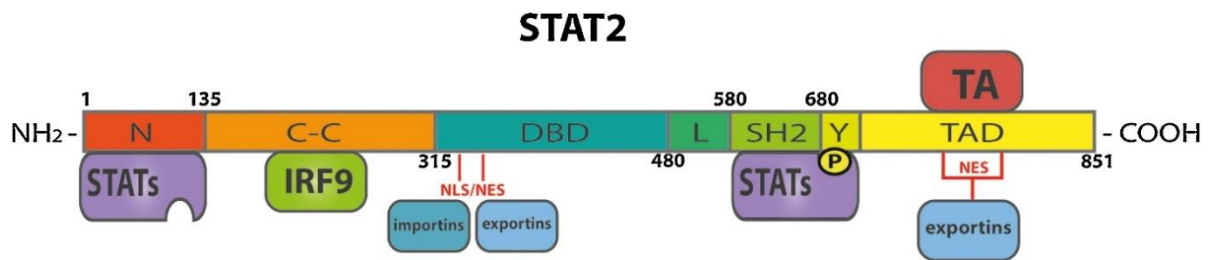
Figure 1: **Domain-structure of STAT2.** *STAT2 is composed of 7 domains: N-terminal domain (N), Coiled-coil domain (C-C), DNA-binding domain (DBD), linker domain (L), Src Homology 2 domain (SH2), tyrosine phosphorylation site (Y) and the transcriptional activation domain (TAD).The N domain, as well as the SH2 domain serve as interaction survaces for the mutual interaction with other STAT (STATs) proteins. The C-C domain is the binding side for IRF9. The c-terminal TAD domain is a platform for transcriptional activation (TA). A constantly active nuclear export signal (NES) is located within the TAD domain. Within the DBD domain an additional NES, as well as a nuclear localization signal (NLS) is localized which depends on the dimerization of STAT2.Figure adapted from Blaszczyk et al..94*

Functionally, STAT2 is a key mediator of mainly type I (IFNβ and IFNα), but also III interferons (IFNs).[94] Especially type I interferons can be produced by nearly every cell in the body, but the major producers are plasmacytoid dendritic cells, which are able to produce up to thousand times more INF-I then other cell types.[95] Interferons are a subset of cytokines and regulate cell proliferation, apoptosis and inflammation but also take part in the first line of defense against viral infections. In the main STAT2 associated pathway, IFNβ and IFNα bind to a heterodimeric transmembrane receptor composed of Interferon-alpha/beta receptor alpha and beta chain (IFNAR1 and IFNAR2) subunits.[94] This initiates a signaling cascade through the Janus kinase-signal transducer and activator of transcription (JAK-STAT) pathway, a canonical cascade used by many cytokines and growth factors to transduce their signals to the nucleus and activate target genes.[94] After receptor activation by INF STATs are recruited to the receptors and successively phosphorylated by JAK1 or TYK2 mediating STAT homodimers or heterodimers that move into the nucleus and activate transcription[94,96]. In the canonical pathway of IFN-I-mediated signaling, phosphorylation of STAT1 on Tyr701 and STAT2 on Tyr690 leads to heterodimerization and interaction with IRF9.[94] Only STAT2 is able to bind IRF9, making it unique among the STAT proteins and crucial as well.[97] The STAT1/STAT2 complex together with IRF9, also known as

IFN-I-stimulated gene factor 3 (ISGF3), then translocate into the nucleus and binds the IFN-I-stimulated response element (ISRE) leading to the activation of transcription of over 300 interferon stimulated genes.[98] Apart from the ISGF3 complex, STAT2-containing heterodimers of STAT2/STAT3 and STAT2/STAT6 were described.[99,100]

Recent studies showed that STAT2 is also capable of forming homodimers when phosphorylated in response to IFN-I.[101] These STAT2 homodimers were shown to interact with IRF9 and form an ISGF3-like complex STAT2/IRF9 that activates transcription of ISRE-containing genes in response to IFNα.[101,102]

The potential impact of ZNF804A on STAT2 function was analyzed and it was found that not only ZNF804A and STAT2 together translocate into the nucleus upon interferon2α treatment but also a ZNF804A dependent reduction in STAT2 mediated target gene expression using an ISRE reporter assay, as well as qPCR.

This finding suggests, that ZNF804A may play a regulatory role in STAT2 mediated interferon response and a reduced ZNF804A expression, as observed in rs1344706 carriers, might lead to a miss-regulated immune defense of neuronal cells upon infection.

4. Results

Schizophrenia is not a disease caused by a distinct set of mutated genes, it is rather a multigenetic burden. In the recent past, great efforts were made to identify genes with potential impact on SCZ, leading to the identification of hundreds of genes.

Although the search is still not finished, another important question came up during the identification process: How do those genes interact, and which processes are the most affected and disturbed during SCZ development? PPI analysis is a power full tool to study the interaction of potential disease relevant genes and to identify clusters of proteins, which are densely connected and enriched for disease associated proteins. Those clusters could be indicative for disturbed processes or in other ways causative for the development of SCZ.

4.1 Creating a bioinformatical tool to predict SCZ relevant clusters

The goal of the first part of this thesis was to establish a computer-based method to identify protein clusters potentially relevant for SCZ. Therefore, in cooperation with Franziska Degenhardt and Markus Nöthen, a list of 283 SCZ associated proteins were created, containing (1) protein coding genes from SCZ-associated CNV loci,[48] (2) protein coding genes with multiple reports for mutations[51,52] and (3) genes with SNPs from GWAS.[11] Because most of the GWAS identified SNPs are not intragenomically, the first genes to genomically up and down stream of SNPs were used (Table 18).

In order to identify highly connected SCZ related clusters, a highly connected PPI network had to be constructed. Multiple data bases curating PPIs are accessible for scientist today.[103] The HIPPIE (Human Integrated Protein-Protein Interaction rEference) database is a constantly updated and growing PPI curating resources which incorporated some of the most prominent available PPI databases.[104–107]

However, the biggest advantage of the HIPPIE database represents the so called HIPPI confidence score, scoring every curated PPI a value based on the amount of reliable evidence supporting them.[105,107] The HIPPIE database include 14.077 proteins, connected via 266.092 PPIs of at least medium high confidence

(suggested HIPPIE score of 0.63). In order to create a SCZ specific network, a second PPI network was created by reducing the interactions of the first network to first neighbor PPIs of the 286 SCZ associated proteins. The resulting SCZ specific PPI network (containing 260 of 286 proteins) still contained 2673 proteins connected via 3532 (Figure 2 A).

This result clearly indicated that without bioinformatical methods, no clear implication can be made with such big data sets. Therefore, a random walk-based propagation algorithm to identify densely connected and SCZ related PPI clusters was established (in cooperation with Arnon Mazza and Roded Sharan). The algorithm assigns every protein in a given PPI network a propagation score, based on their closeness and connectivity to priors used (Figure 2 B). In this case the list of 286 SCZ associated proteins was used as priors and as a PPI network the HIPPIE database.
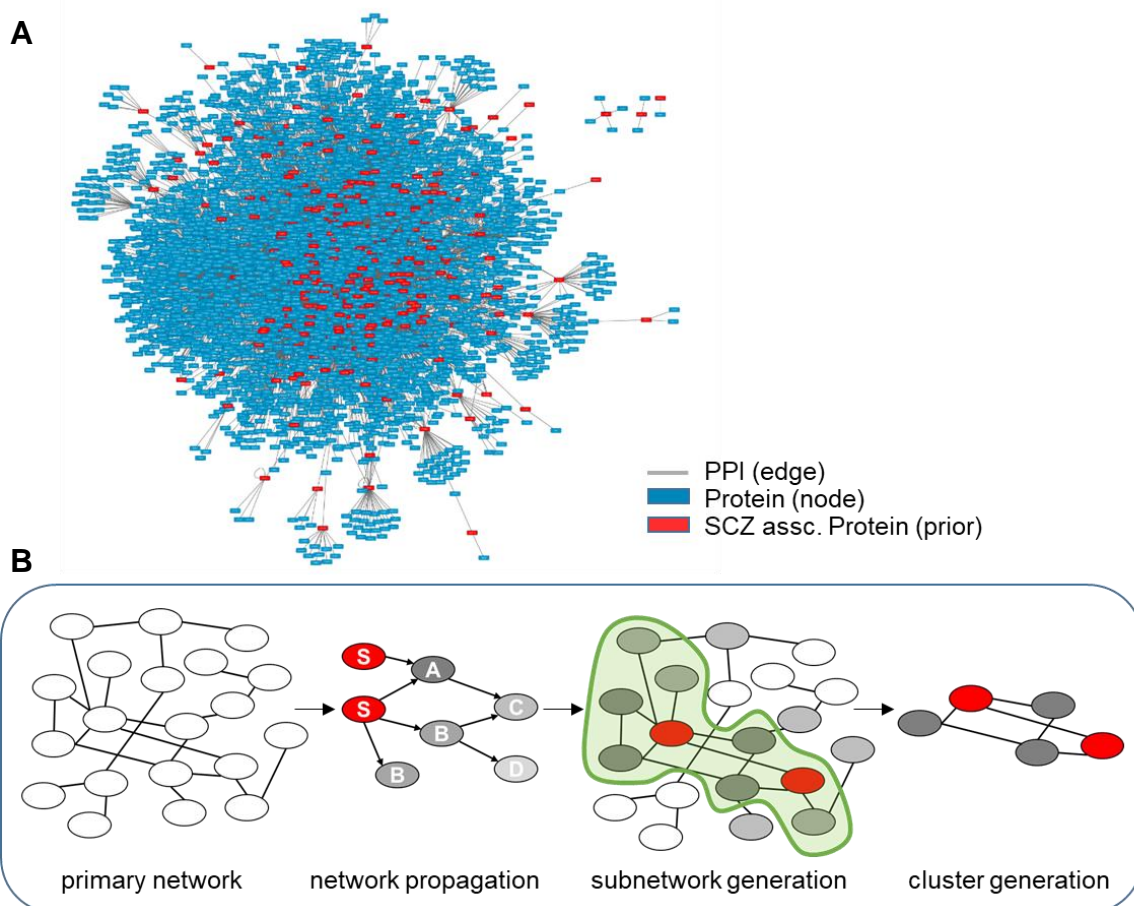

*Figure 2: **Propagation based identification of SCZ PPI cluster from large scale PPI network.** (a) Medium high confident (HIPPIE confident score of at least 0.63) HIPPIE PPI network containing 2673 protein nodes and 3532 PPI edges. Grey*

lines represent PPIs. Blue rectangles represent proteins. (b) HIPPI based PPI network of all first neighbor interactions of SCZ associated proteins (table x) with at least a HIPPIE confident score of 0.6). Red rectangles represent proteins, previously associated with SCZ. (c) Schematic flow chart of algorithm based SCZ cluster generation. Ellipses represent protein nodes and black lines PPIs. Red ellipses represent SCZ associated priors. Shades of grey, as well as letters from A to D indicate propagation scores (propagation score: A>B>C>D). Green outlining indicating propagation score based PPIs used for subnetwork creation.

As depicted in the flow chart (Figure 2 B), the algorithm follows three steps: 1) network propagation of a given network (HIPPIE confidence 0.63) using a set of priors (283 SCZ related proteins), 2) subnetwork generation using propagation scores, 3) cluster generation using propagation scores, protein connectivity and additionally expression values in the prefrontal cortex for each protein. To generate a SCZ based subnetwork, based on the propagation scores (and therefore in dependence of the connectivity to SCZ genes) a cut of those propagation scores had to be made to consider a protein for the inclusion in the subnetwork. Two problems arose from making this decision. The first problem is, that propagation scores highly depend on the prior sizes used. The higher the number of priors used for the propagation of a network of the same size, the higher the average propagation score of each protein of the network. The other inherent problem of using simply cut offs are, that some proteins have generally a high number of PPIs. Those proteins generally end up with high propagation scores, even though they would not be necessarily enriched for interactions with priors. To circumvent these factors, propagation of a 1000 random sets of the same prior size of 286 were run to determine the random propagation score for each protein in the given network. Comparing propagation scores when using the 283 SCZ associated proteins as priors to those scores of random prior propagations allowed the selection of proteins with a significant SCZ prior propagation score (p<0.05). Out of those, a disease subnetwork was formed (Figure 2 B). In order to search for highly connected protein clusters within the subnetwork with high propagation scores, in a third step, a cluster algorithm had to be set up.

For the cluster search three parameters were considered: 1) the overall propagation score, 2) the connectivity within the clusters and 3) the expression in the brain. In cooperation with Arnon Mazza the cluster algorithm was set to search for clusters of proteins which were at least a half click (a click is a cluster where

every protein is connected to all other proteins of the cluster) and had a maximal sum of propagation score. For the third parameter the Alan Human Brain Atlas dataset of post mortem brain gene expression was used because it represents one of the biggest publicly available human gene expression data sets.[82] The gene expression levels in the prefrontal cortex were used for the cluster generation, because it has been found to be the brain area with clear alterations in SCZ.[108] Arnon Mazza implemented the expression value as an edge value, given protein PPI pairs, where one protein is not expressed in the brain a lowered score, and pairs, where both proteins are not expressed, an even lower score.

4.1.1 Bioinformatical identification of five highly connected SCZ associated clusters

As a result, five high scoring clusters were identified (Figure 3Figure 3: **Identified SCZ related PPI clusters.** Rectangle represent proteins, previously associated with SCZ. Circles represent proteins without previous SCZ association. Grey lines represent PPIs.). The highest scoring cluster consist of post synaptic proteins, including CRIPT (Cysteine-rich PDZ-binding protein), DLG1 (Discs large homolog 1), DLG2 (Discs large homolog 2), DLG4 (Discs large homolog 4), DLGAP1 (Disks large-associated protein 1), GRIN2A (Glutamate receptor subunit epsilon-1) and SEMA4C (Semaphorin-4C). Three proteins of this cluster were previously identified to be candidate genes for SCZ: DLG1 and DLG2, as well as GRIN2A.[11,51,52] DLG proteins are defined by containing at least one PDZ domain which enables them to homo and heteromultimerise and together with other PDZ containing proteins as CRIPT to form the scaffold backbone of the post synaptic density.[109,110] Similarly, DLGAP1, a protein of the MaGuk (Membrane Associated Guanylate Kinase) family, acts as a postsynaptic density scaffold protein, linking glutamate receptors together.[111] GRIN2A and SEMA4C encode for postsynaptic receptor subunits. The expression product of GRIN2A acts as a subunit of the glutamate-gated NMDA receptors. Large PSDs are especially characteristic of type 1 glutamatergic excitatory synapses and represent a core structure of neuronal connectivity.[110]
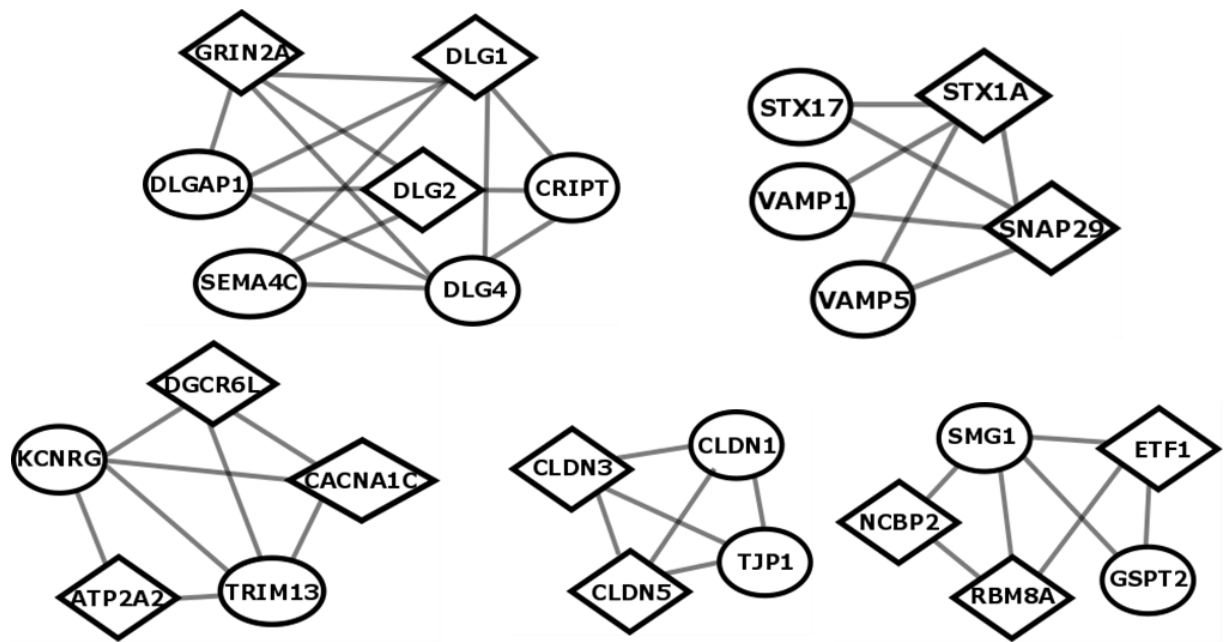
*Figure 3: **Identified SCZ related PPI clusters.** Rectangle represent proteins, previously associated with SCZ. Circles represent proteins without previous SCZ association. Grey lines represent PPIs.*

The second cluster contained STX1A (Sytaxin-1A), STX17 (Sytaxin-17), SNAP29 (Synaptosomal-associated protein 29), VAMP1 (Vesicle-associated membrane protein 1) and VAMP5 (Vesicle-associated membrane protein 1). All of these proteins are a part of the SNARE complex (SNARE proteins) in the presynapse and represent a very well described protein machinery.[112,113] SNARE proteins were originally identified for their ability to regulate vesicle release at mature synapses.[114] In fact, the spontaneous and calcium guided interaction between members of the SNARE family allows for the release of neurotransmitters at the synaptic cleft, which leads to efficient synaptic transmission.[115] Both, STX1 as well as SNAP29 are associated with SCZ via their inclusion in CNVs.[48]

The third cluster represents potassium channel associated proteins (CACNAC1C (Calcium channel, voltage-dependent, L type, alpha 1C subunit), KCNRG (Potassium channel regulator)), as well as the Ca^{2+}-ATPase ATP2A2 (ATPase Sarcoplasmic/Endoplasmic Reticulum Ca2+ Transporting 2) and DGCR6L (DiGeorge Syndrome Critical Region 6-Like Protein). The most interesting protein is CACNA1C as a subunit of L-type calcium channels (LTCCs), which are critical mediators of experience-dependent plasticity in the brain.[116] CACNA1C was associated not only with SCZ through GWAS, but also with bipolar disorder, autism

spectrum disorder, major depression disorder and attention hyperactivity disorder.[117–120]

Besides CACNA1C, ATP2A2 and DGCR6L are both SCZ associated genes, associated by being genes within SCZ related CNVs.[48,121]

Cluster 4 is composed of only tight junction proteins consisting of CLDN1, -3 and -5 (Claudin 1, -3 and -5) as well as TJP1 (Tight Junction Protein 1). Both CLDN3 and CLDN5 are associated with SCZ via inclusion in CNVs.[48]

Cluster 5 represents a more diverse group of proteins. The via CNVs with SCZ associated RBM8A (RNA Binding Motif Protein 8A) as well as NCBP2 (Nuclear Cap Binding Protein Subunit 2) are both involved in splicing and SMG1 (Nonsense Mediated MRNA Decay Associated PI3K Related Kinase) in nonsense mediated mRNA decay.[48] GSTP2 (G1 To S Phase Transition 2) and ETF1 (Eukaryotic Translation Termination Factor 1) on the other hand represent translational elongation and release factors respectively.

4.1.2 Validating identified SCZ associated cluster PPIs using LuTHy

The further goal was to analyze the identified clusters, which were suspected to be disease modules. The first step was to not only validate cluster PPIs, but also analyzing all possible PPIs within the cluster. The second step was to then analyze the impact of mutations described for SCZ and their impact on the cluster connectivity. This approach might indicate disease relevant loss of functions or other processes with implications for drug development.

For the time and resource consuming PPI validation, the two highest scoring clusters were chosen as to be the most promising.

To analyze cluster PPIs in the most comprehensive way, the newly developed dual readout BRET based LuTHy method was used.[66] In LuTHy assays NanoLuc luciferase tagged proteins are co-produced with PA-mCitrine tagged proteins. In intact HEK293 cells, after addition of NanoLuc substrate, activated luciferases of prey proteins are able to transfer their energy via bioluminescence resonance energy transfer radiant free to mCitrine tags of interacting bait proteins. For BRET to occur, mCitrine and NanoLuc tags of prey and bait proteins need to be in very close proximity (under 10 nm).[122] Due to the ~80 times brighter NanoLuc luciferase,

compared to other luciferases, LuTHy requires only very low expression of NanoLuc tagged prey proteins.[123] Bait PA taq enable the assay to perform a second, co IP based readout.[66] All cluster encoding genes were cloned N-terminally into both, the LuTHy vectors (pcmyc-NL-GW, pPA-mCit-GW). 50 ng of NanoLuc fusion constructs were co transfected together with 150 ng of PA-mCitrine fusion constructs using PEI transfection agency in HEK293 cells in 96 well format. Each used prey and bait construct were additionally tested against bait (xx) and prey (xx) control constructs respectively to exclude unspecific binding. In order to normalize measured values cross plates and between screens on every plate a NanoLuc-mCitrin fusion construct was transfected. All screens were performed in three biological replicates with three technical replicates and all PPIs were tested in both direction (every protein as bait and as prey) (Figure 6). As cut offs, to determine PPIs as positive, even stricter scores for intracellular BRET and IP based measurements of 0.04 cNIR and 0.02 cLuC were set respectively then published (Figure 6 B-E).[66]
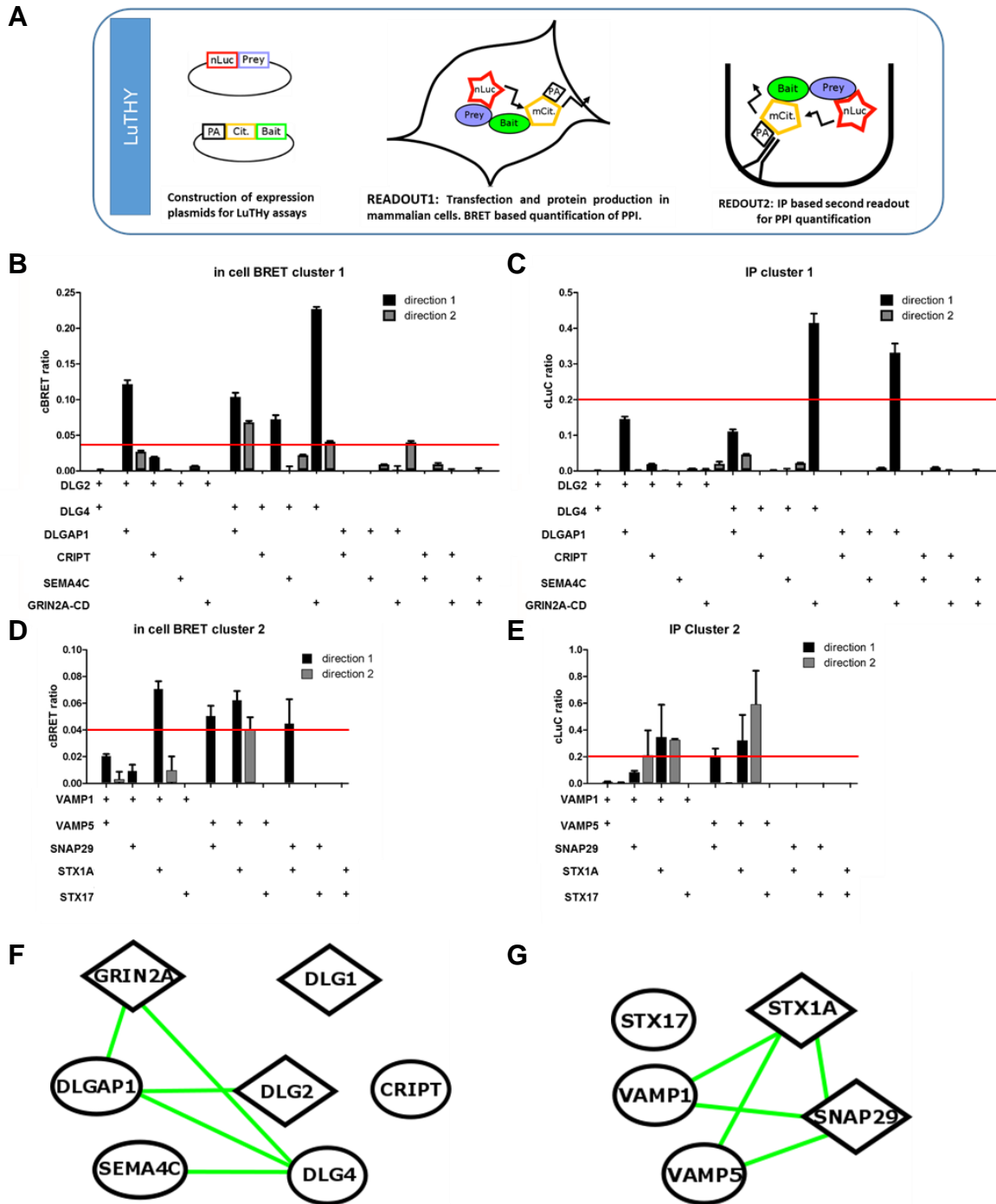
*Figure 4: **SCZ related cluster validation using LuTHy.** (a) Flow chart representation of the LuTHy assay. nLuc: nanoLuciferase, PA: podoplanin antibody tag, Cit: mCitrine. Arrows represent either transferred energy or measurable emission of light. (b-c) First and second readout of cluster 1 validation using LuTHy displayed as cBRET and cLuC respectively. (d-e) First and second readout of cluster 2 validation using LuTHy displayed as cBRET and cLuC respectively. Direction 1 represent bait vs prey pairing (for example for the first interaction, cluster 1 DLG2 as bait and DLG4 as prey), where direction 2 represent prey vs bait pairing (for example for the first interaction, cluster 1 DLG2 as prey and DLG4 as bait). All experiments were done in triplicates with three biological replicates. Error bars represent standard deviation. Red lines indicate cut of scores for considering PPIs as positive. (f-g) Depiction of validated cluster 1 and 2 PPIs respectably. Rectangle represent proteins, previously associated with SCZ. Circles represent proteins without previous SCZ association. Green lines represent validated PPIs.*

4.1.2.1 Cluster 1 validation

First of all, expression valuations of all cluster 1 proteins, cloned in both LuTHy vectors, revealed a lack of expression for proteins encoded by GRIN2A, as well as by DLG1 (data not shown). For GRIN2A therefore a construct was used for LuTHy screening that encoded only the cytosolic C-terminal cytosolic domain (GRIN2A-CD: amino acids 838-1464), because nearly all PPI methods tend to have problems to recover PPIs with transmembrane proteins. Unfortunately, DLG1 had to be excluded from further analysis. The intracellular LuTHy readout was able to validate 38% of previously described cluster PPIs (Figure 4 B). If for this consideration DLG1 would have been excluded, the validation rate would even have had increased to 50%. The highest cNIR values were measured for the interaction of DLG4 with the cytosolic domain of GRIN2A. CRIPT did not show any interactions. The IP based readout additionally validated the interaction between GRIN2A-CD and DLG4, as well as between GRIn2A-CD and DLGAP1, but did not increase the coverage of overall validated PPIs (Figure 4 C). Taken together, a total of 38% of previously described PPIs of the first cluster were successfully validated, unfortunately without any detected interactions for the not expressed DLG1, as well as for CRIPT. No previously unknown PPIs were identified.

4.1.2.2 Cluster 2 validation

For cluster two, all proteins were well expressed. In cell BRET revealed interactions between VAMP1-STX1, VAMP5-SNAP29, as well VAMP5-STX1A and SNAP29-STX1 without any PPIs with extraordinary cNIR values. A total of 57% of previously described PPIs were detectable in cell using LuTHy (Figure 4 D). The second IP based readout additionally validated the interactions of VAMP1-STX1 and VAMP5-STX1. It also validated the interaction of VAMP5-SNAP29, increasing the overall validation rate of previously described PPIs to 71% (Figure 4 E). This shows the importance of using different PPI methods to get the largest possible validation coverage. All not covered previously described PPIs are those of STX17 of which none were able to be recovered by any LuTHy readout.

4.1.2.3 Attempt to analyze the impact of SCZ associated mutations on cluster connectivity

Due to the lack of DLG1 expression, for the analysis of mutation effects on cluster PPIs, only GRIN2A had described SCZ associated coding mutations with

accessible location at that time.[51,52] Therefore, using primer directed mutation, SCZ mutation was introduced into the cytosolic construct of GRIN2A and tested again in LuTHy assays without the detection of significant changes in PPI strength for any interaction (data not shown).

Although no SCZ mutation driven changes in cluster protein connectivity were identified, the strategy might be valuable for future approaches, after more and more SCZ related coding mutations are discovered.

In summary, five SCZ specific clusters were identified that can be considered disease modules using PPI network propagation. The two highest scoring clusters had a clear synaptic association as depicted (Figure 6) and PPIs were successfully validated using LuTHy assays. A first attempt to evaluate the impact of described coding SCZ mutations was made for GRIN2A without detectable impact on cluster connectivity.
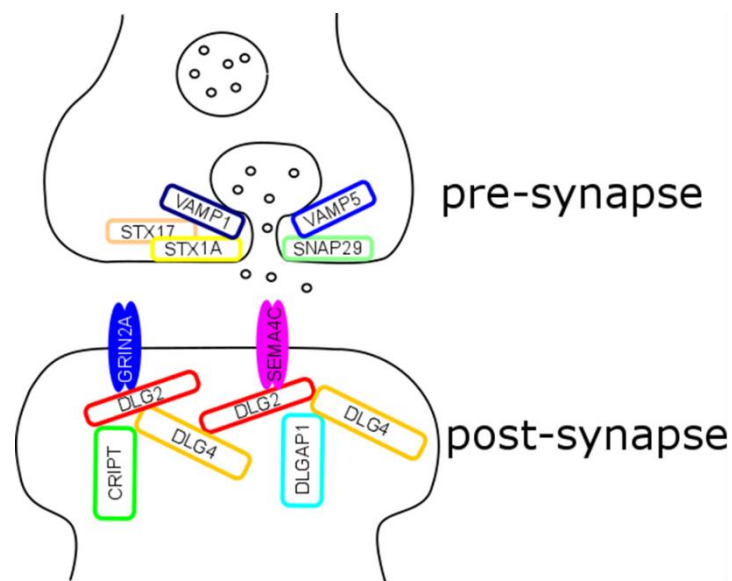

Figure 5: **SCZ cluster protein location in the synapse.** *Schematic depiction of all proteins of identified cluster1 and 2. Cluster 1 is located in the post synapse as mainly parts of the post synaptic density. Cluster 2 proteins are part of the SNARE complex, mediating vesicle fusion. Depiction only visualizes broad localization and not necessary interactions.*

4.2 Network based ranking and prioritization of SCZ susceptibility genes of a focused exome sequencing study

Although a high number of SCZ associated genes were already identified in the recent past, the identification process of relevant SCZ genes is still an ongoing process, especially in the field of exome sequencing. The challenge of exome sequencing studies is that they usually recover a lot of gene alterations besides those primarily impacting SCZ. Therefore, these studies need to perform an intense filtering to increase the chance for the recovery of relevant SCZ genes. But even then, the ranking of those identified genes is still unclear.

As a cooperation partner for Franziska Degenhardt, Anna Koller and Markus Nöthen we were engaged in an exome sequencing project that examined three psychiatric diseases affected families. In total, 20 individuals (n = 9 patients with SCZ; n = 3 individuals with major depression; and n = 8 unaffected relatives) underwent an extensive diagnostic interview conducted by an experienced research psychiatrist and donated DNA samples for sequencing (Figure 7).

After intensive filtering and the additional inclusion of identified CNV in the affected family members, a total of 39 protein coding genes were identified as being SCZ candidate genes (methods: Table 19). The two questions that were tried to be answer, using bioinformatics PPI methods were: A) Are those genes enriched for interacting with previously reported SCZ candidate genes, as an indicator for disease relevance? And B) Which of those genes are the most promising SCZ candidates?
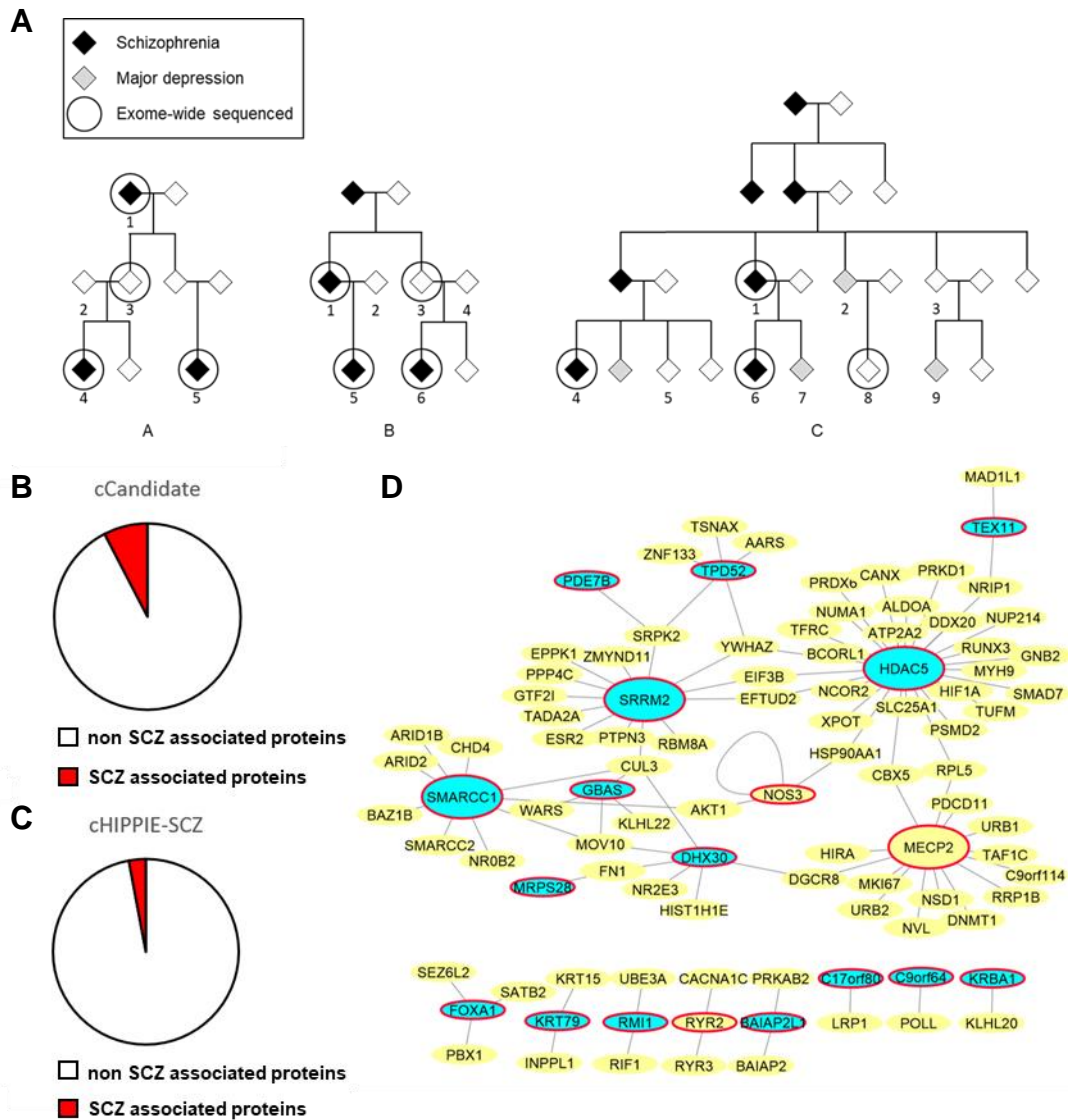
Figure 7: **Family three based SCZ candidate protein ranking.** (a) family tree of three severely SCZ affected families from Würzburg (WB-set). Besides diagnosed SCZ (black rectangle), also major depression dignoses are indicated (grey rectangle). Indicated family members (circled) were exome wide sequenced. (b-c) Pie chart depiction of the proportion of previously with SCZ associated proteins in cCandidate and cHIPPIE-SCZ respectively. (d) Cytoscape PPI network of identified proteins with previously with SCZ associated proteins. Red bordering: WB-set proteins, yellow: previous SCZ annotation, turquoise: No previous SCZ association. Enlarged circles indicate highly connected WB-set proteins.

Previous investigations indicate that proteins involved in the same disease have an increased tendency to interact with each other, suggesting that the 39 SCZ candidate proteins identified by Koller *et al.* might preferentially associate with previously reported SCZ candidate proteins in protein-protein interaction (PPI) networks.[53] To address this question, a focused PPI network for the 39 SCZ-associated candidate protein coding genes was generated, using again medium

high confidence interactions from the HIPPIE interaction database with a HIPPIE score ≥ 0.63. This network was termed cCandidate. It connects 1.183 proteins via 1.416 interactions, including 38 of the initial 39 proteins. The protein L3MBTL1 was not included in cCandiate, because it was not present in the HIPPIE database.

4.2.1 PPI network analysis reveals significant connectivity of candidate proteins to known SCZ related proteins

Next, together with Koller *et al.* a compiled another comprehensive list of 1,284 previously reported SCZ-associated genes using the available literature information. It contained all genes: 1) located within SCZ-associated CNV loci;[48,124] 2) with a reported *de novo* mutation in a patient with SCZ;[51,125–134] and 3) in which a genome-wide significant single nucleotide polymorphism was reported in the PGC SCZ GWAS.[11] It is to note, that this list was not filtered for protein coding and non-coding genes and was compiled two years after the list used in the first part of this thesis. Finally, the abundance of the previously reported SCZ-associated proteins was assessed in the cCandidate interaction data set and in an independent medium high-confidence PPI data set (termed cHIPPIE_without_cCandidate: 9.283 proteins connected via 65.533 interactions), which represents all PPIs of proteins included in HIPPIE crossing the confidence score of 0.63 and were not included in cCandidate.

Strikingly, in comparison to the cHIPPIE_without_cCandidate network, in the cCandidate network a significantly enrichment of reported SCZ candidate proteins (chi squared test; p < 0.0001) was detected (depicted as pie diagrams Figure 8), indicating that the newly identified proteins preferentially interact with previously reported SCZ-associated proteins.

4.2.2 Bioinformatical ranking of SCZ candidate proteins

In order to investigate the second question, about which of those genes are the most promising SCZ candidates, a subnetwork of cCandidate was created, only displaying the interactions between candidate proteins and the previously reported SCZ-associated proteins. To exclusively display the links between these proteins an additional focused PPI network was generated (fig). This network connects 99 human proteins via 99 interactions and links 16 newly identified proteins to 83 previously reported SCZ-associated proteins (Figure 7 and Supplementary Table 10). Three proteins, MECP2 (methyl CpG binding protein), RYR2 (Ryanodine receptor 2) and NOS3 (Nicotin oxide synthase 3) were previously reported as SCZ relevant but additionally identified by Koller *et al.*.

The proteins HDAC5 (Histone deacetylase 5), MECP2, SRRM2 (Serin/argenin repetitive matrix protein 2) and SMARCC1 (SWI/SNF Related,Matrix Associated, Actin Dependent Regulator Of Chromatin Subfamily C Member 1) displayed as hubs (proteins with a high number of interactions) in this focused network and therefore might play an important role in pathogenesis. Based on their central position in the network these proteins might be very interesting for further analysis. Interestingly, HDAC5, MECP2 and SMARCC1 are involved in chromatin remodeling, which might be an interesting process to focus on in a follow up study to evaluate the impact for those affected patients and SCZ in general.[135–137]

Finally, an independent network-based approach to prioritize the SCZ-associated proteins identified by Koller *et al.* was performed. The browser based ToppNet tool were used which allows to train a PPI network of 17.064 interactions with a training set of proteins, in this case the list of previously reported SCZ genes. In a second step, a test set was run over the trained PPI network, using the implemented k-step Makarov algorithm with a step size of 6.[138] All overlapping proteins between the 38 identified candidate genes of this study with previously identified SCZ genes were excluded from the trainings set however.

The results of ToppNet revealed very high scores for the proteins HDAC5, MECP2 and SMARCC1 (Table 2).

Taken together, the by Koller *et al.* identified candidate genes were validated to be enriched for SCZ interactions. Additionally, candidate genes were prioritized and HDAC5, MECP2 and SMARCC1 identified as the most promising candidates for further SCZ studies.

Table 2: Prioritization of SCZ candidate genes using ToppNet

RANK	ID	PROTEIN	SCORE
1	10014	HDAC5	0.463
2	23524	SRRM2	0.303
3	4204	MECP2	0.226
4	6599	SMARCC1	0.186
5	3169	FOXA1	0.129
6	22907	DHX30	0.100
7	6262	RYR2	0.076
8	56159	TEX11	0.070
9	4846	NOS3	0.068
10	7163	TPD52	0.067
11	26013	L3MBTL1	0.067
12	28957	MRPS28	0.055
13	55971	BAIAP2L1	0.048
14	84267	C9orf64	0.039
15	57544	TXNDC16	0.033
16	84626	KRBA1	0.033
17	338785	KRT79	0.032
18	55028	C17orf80	0.029
19	79088	ZNF426	0.027
20	80010	RMI1	0.024
21	26575	RGS17	0.021
22	80312	TET1	0.016
23	64412	GZF1	0.013
24	56965	PARP6	0.010
25	23483	TGDS	0.010
26	27115	PDE7B	0.004
27	285973	ATG9B	0.004
28	91355	LRP5L	0.003
29	339761	CYP27C1	0.002
30	139135	PASD1	0.002
31	85416	ZIC5	0.001

4.3 ZNF804A – PPI network exploration discovers potential functional implications for SCZ

The identification and prioritisation of SCZ candidate genes is not the final step in the task of understanding the melecular mechanisms of SCZ. In order to unreavel the cause of SCZ, it is also of high priority to identify the role of each SCZ associated gene in health and disease. The first single nucleotide polymorphism that showed genome wide significanz association for SCZ, but also bipolar disorder was rs1344706.[47] This specific SNIP is located within the second intron of the *ZNF804A* gene. In silico modeling has revealed a C2H2 containing zinc-finger domain, encoded within the second exone of ZNF804A (Figure 9 A).[139] Zinc-finger domains (ZNF) are discribed as DNA binding and regulating RNA transcription. To further characterise the secondary structure of ZNF804A browser based tools (NLS Mapper, NetNLS) were used to identify NLS and NES within the ZFN804A protein and identified two NLS sequences and one NES (Figure 9 A). Although destinct affects on brain function and structure have been discribed for rs1344706, the function of ZNF804A is relativly unknown.

4.3.1 Y2H identifies 18 new and unique PPI partners

The hypothesis of guild by association for proteins, which means that interactors of proteins involved in a specific process very likely also contribute to the same process, making the approach to identify PPIs for proteins of intrest very appealing. One central project goal was to follow this hypothesis and identify ZNF804A PPI partners and gather therefore further insights into ZNF804A functionallity. In order to achive this goal, ZNF804A had to be tested against (in the best case) the whole proteome. Automated and robot based Y2H is one of the few PPI methods, able to achive such a goal in a timly and costly manner. A first Y2H screen of ZNF804A against a library of over 17.000 prey proteins was previously performed in the Wanker lab by co-workers and was able to identify 84 PPIs for ZNF804A. The initial screen used besides the full lentgh ZNF804A protein also two fragments of ZNF804A, ZNF804A-1-609 and ZNF804A-601-1209 (Figure 9 B). Since the first innitial screen several control steps of the Y2H methology have been improved significantly. For example, to eliminate all false positive PPIs I established a routine of testing potential interaction partners against a non-interactive protein: mCherry, as well as a PCR, followed by a sequenzing protocoll to ensure the identity of all

positive tested proteins. Therefore, a secondary screen to validate all protein-interaction partners of ZNF804A, tested positive in the primary screen was performed. For determining interactions as positive stringent cut offs were used, as well as all newly established improvments to the Y2H methodelogy. Additionally, to increase the mapping of potential interaction domains of ZNF804A, the number of ZNF804A fragments used in the secondary screen were increased to a total number of seven (Figure 9 B).
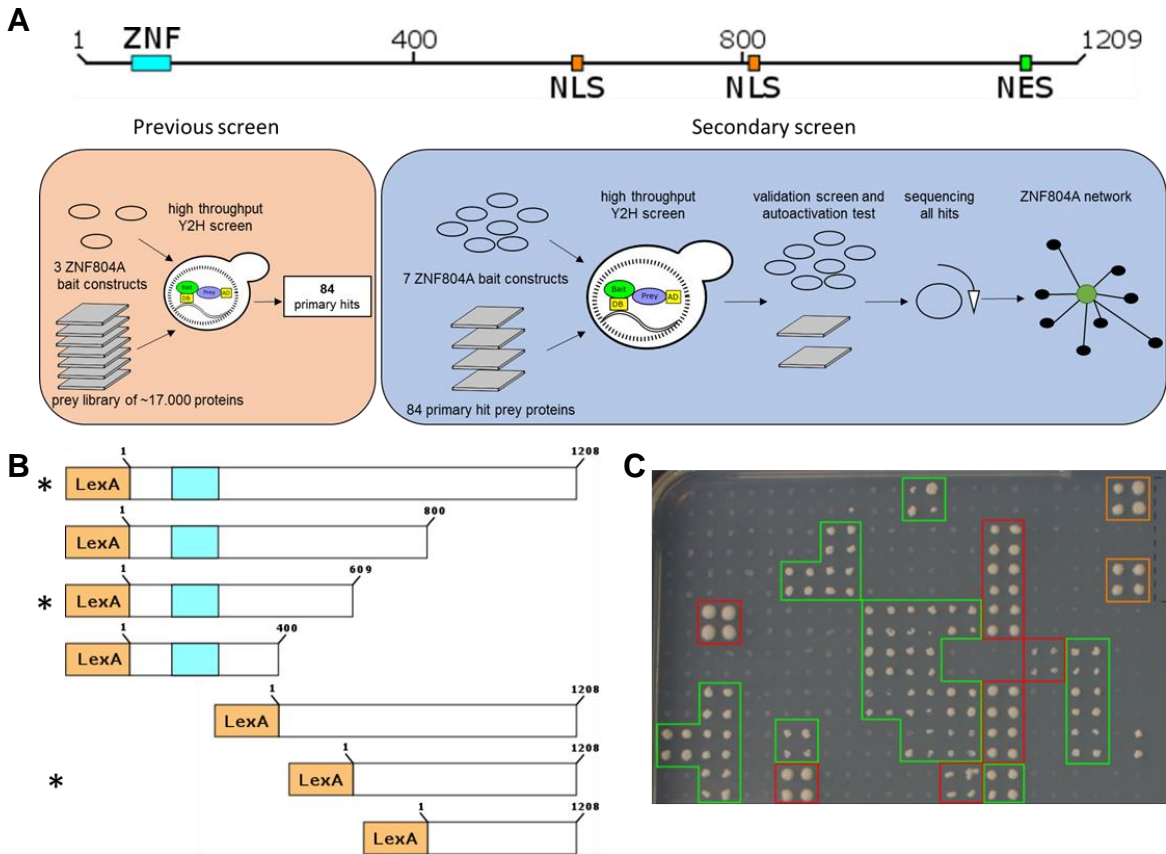
Figure 9: **Y2H screening of ZNF804A protein-protein interactors.** *(a) Schematic domain structure of ZNF804A. Blue rectangle: Zinc finger domain (ZNF), orange rectangles: Nuclear localization signals (NLS), green rectangle: Nuclear export signal (NES). (b) Flow diagram of the previous and secondary ZNF804A Y2H screen. (c) Depiction of ZNF804A bait constructs used in Y2H screens. Orange rectangles: LexA tags (LexA), blue rectangles: Zinc finger domains. Asterisks indicate ZNF804A bait constructs used in primary screen. (d) Representative image of one biological replicate of one bait (ZNF804A_1-400) in technical quadruplicates against the set of 84 primary identified prey proteins. Orange squares: Positive controls, green squares: Positive tested interactions, Red squares: false positive tested interactions.*

All ZNF804A fragments were cloned into N-terminal bait pBTM116-D9 vector and chemically transfected into MATa yeasts. All initialy identified pray proteins were isolated from the library that was used for the initial screen and used as preys for the validation screen (prey proteins in pACT4-DM vector and MATα). For interaction mating, bait and prey yeast were co-spotted (mated) on YPD agar plates and grown for two days. Grown mated colonies were transferred into liquid YPD media and spotted on two different selective media: SD2 and SD4. SD2 requires mated yeast to form colonies and therefore, is used to validate successful matings. SD4 on the other hand, allows colony growth only if auxotrophic markers are expressed that are under the control of GAL4 and therefore only if the LexA/Gal4 transcription factor is restored due to bait and prey protein interact. For that reason, SD4 colony growth are the readout for positive protein-protein interaction.

The use of the robot based Y2H screening method enabled to mate all MATa yeast strains producing the seven different ZNF804A fragments systematically four times on YPD agar plates against all 84 yeast strains expressing human prey proteins identified in the previous screen in an efficient and rapid manner (Figure 9 A). For the mating process, MATa strains containing pBTM116-D9 encoding bait proteins are one to one mixed with MATα strains containing pACT4-DM encoding prey proteins in a one bait pairing with one prey manner on YPD agar and were incubated for three days at 37 °C.

Every mating was replicated four times and mated colonies were re-plated on selective media (SD4). Interacting bait and prey proteins were finally identified through growth assays on SD4 (Figure 9 C). To ensure that all bait-prey pairings were successfully mated, they were additionally plated on SD2 media, which is selective for successful mating. Only pairs tested positive for at least eleven out of sixteen times (4 matings x 4 replications) were considered as positive interactions. Due to the stringent cut off, a total of 34 potential interactions were identified at this step (Figure 9 A). To exclude false positive interactions, all potential interactors were tested for auto-activation and were mated against MATα producing only GAL4, as well as MATα producing GAL4-mCherry. In a similar fashion, all ZNF804A constructs were tested against MAT producing only LexA, as well as MAT producing LexA-mCherry. In this process 12 prey proteins were tested as

auto active and therefore false positive. None of the bait constructs were tested as auto active. Finally, to confirm prey identity, positive colonies were picked and sequenced with prey vector specific primers. For all seven constructs together, a total of 18 unique and previously unknown ZNF804A protein interaction partner proteins were identified in Y2H mating approach and sequencing identity confirmed (Figure 10 and Table 3:). All n-terminal constructs, harboring the zinc Finger domain (ZNF804A-1-400, ZNF804A-1-609, ZNF804A-1-800) interacted with only one protein: STAT2. The smallest c-terminal construct of ZNF804A (ZNF804A-800-1209) showed the most interactions with a total number of 17 (Figure 10 and Table 3:), indicating that the last 400 AS of ZNF804A harbor a PPI domain. The slightly larger c-terminal construct ZNF804A 601-1209 shared 13 of the 17 interaction partners with ZNF804A-800-1209, only missing RBM4 (RNA Binding Motif Protein 4), KRTAP8-1 (Keratin Associated Protein 8-1), STAM2 (Signal Transducing Adaptor Molecule 2) and TSC1 (TSC Complex Subunit 1) (Figure 10 and Table 3:). The largest c-terminal construct ZNF804A-400-1209 only recovered two interactions of the 17 of the smallest fragment ZNF804A-800-1209; FAM46A (Family With Sequence Similarity 46 Member A) and PTS (6-Pyruvoyltetrahydropterin Synthase) (Figure 10 and Table 3:). The full length construct of ZNF804A was unable to recover any PPIs (Figure 10).

Taken together, 18 reliable and previously unknown ZNF804A interaction partners were identified using automated Y2H technology.

Table 3: Identified ZNF804A protein-protein interactors

ZNF804A interactors	full name	ZNF804A fragment
DAZAP2	DAZ Associated Protein 2	ZNF804A-800-1209
		ZNF804A-601-1209
EYA3	EYA Transcriptional Coactivator And Phosphatase 3	ZNF804A-800-1209
FAM46A	Family With Sequence Similarity 46 Member A	ZNF804A-800-1209
		ZNF804A-601-1209
		ZNF804A-400-1209
KRTAP19-1	Keratin Associated Protein 19-1	ZNF804A-800-1209
		ZNF804A-601-1209

KRTAP19-7	Keratin Associated Protein 19-7	ZNF804A-800-1209
KRTAP8-1	Keratin Associated Protein 8-1	ZNF804A-800-1209
		ZNF804A-601-1209
LASP1	LIM And SH3 Protein 1	ZNF804A-800-1209
		ZNF804A-601-1209
LITAF	Lipopolysaccharide Induced TNF Factor	ZNF804A-800-1209
		ZNF804A-601-1209
PTS	6-Pyruvoyltetrahydropterin Synthase	ZNF804A-800-1209
		ZNF804A-601-1209
		ZNF804A-400-1209
QKI	Quaking	ZNF804A-800-1209
		ZNF804A-601-1209
RBM4	RNA Binding Motif Protein 4	ZNF804A-800-1209
		ZNF804A-601-1209
RBM4B	RNA Binding Motif Protein 4B	ZNF804A-800-1209
		ZNF804A-601-1209
RBM46	RNA Binding Motif Protein 46	ZNF804A-800-1209
		ZNF804A-601-1209
SEMA4G	Semaphorin 4G	ZNF804A-800-1209
		ZNF804A-601-1209
STAM2	Signal Transducing Adaptor Molecule 2	ZNF804A-800-1209
STAT2	Signal Transducer And Activator Of Transcription 2	ZNF804A-1-800
		ZNF804A-1-609
		ZNF804A-1-400
TSC1	TSC Complex Subunit 1	ZNF804A-800-1209
UBE2I	Ubiquitin Conjugating Enzyme E2 I	ZNF804A-800-1209
		ZNF804A-601-1209
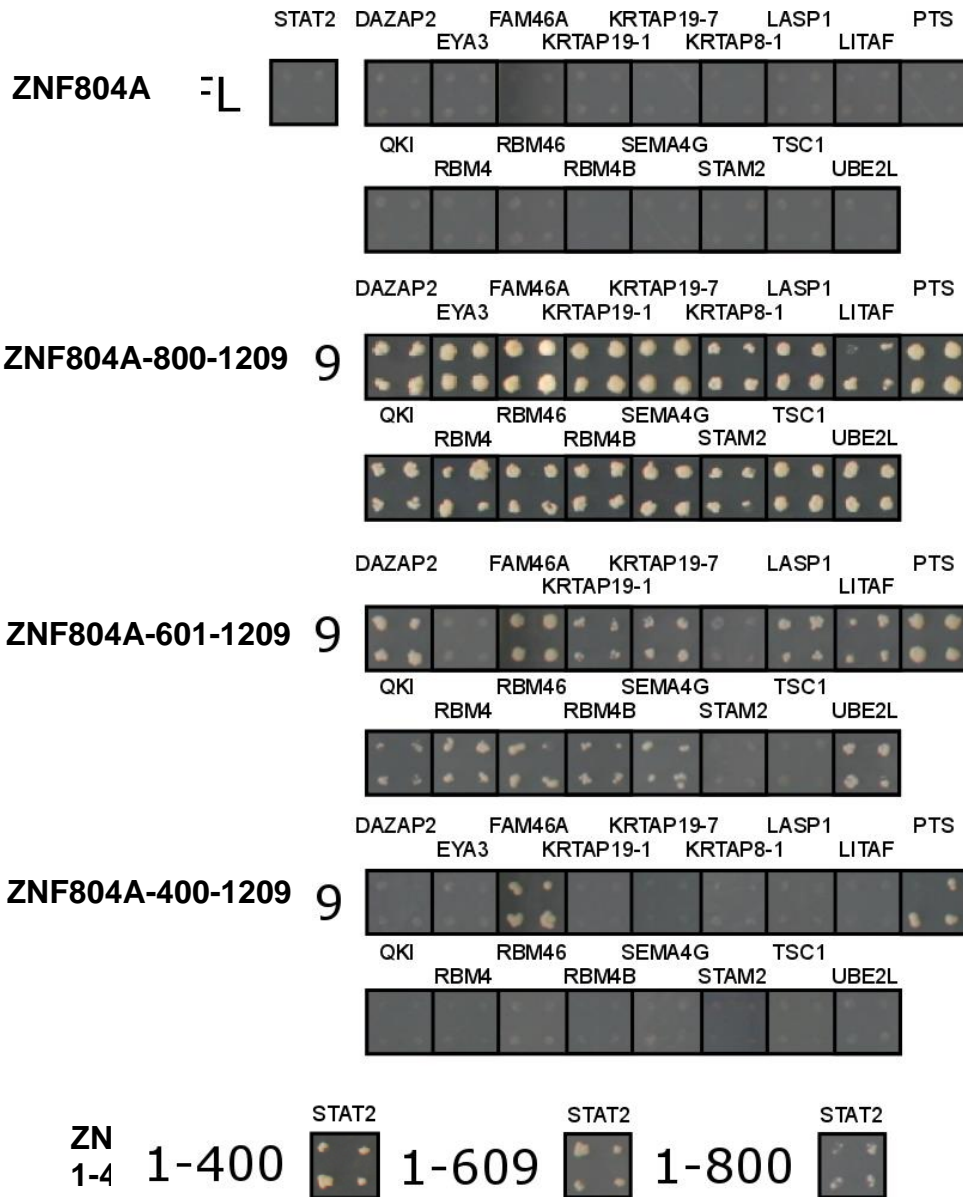
Figure 10: **Y2H colony growth assay identifies 18 previously unknown ZNF804A interactors.** Depicted are representative quadruplicates colony growth measurement of all identified ZNF804A interacting proteins with all fragments of ZNF804A used on SD4.

4.3.2 ZNF804A interactors are associated with RNA binding, the circadian clock and inflammation pathways

ZNF804A is one of the most consistent genes associated with SCZ. Therefore, the first logical step for the evaluation of the 18 new interactors of ZNF804A was to identify if one or more of them might be associated with SCZ in the dataset of the biggest SCZ genetic study up to date: The PGC2 GWAS dataset. Although, none of the interactors (except for ZNF804A itself) was genome wide significantly associated, it was still possible that at least one or two only did not surpass the restrictive genome wide correction. Thus, in cooperation with Franziska Degenhardt and Anna Koller a set-based test of all 18 identified interactors over the whole PGC2 GWAS data set was performed, only correcting for the set of 18 interactors instead of the whole genome (Table 4).

The only ZNF804A interactor, showing significant enriched SNIPs accumulation in the PGC2 patient data set, compared to controls was QKI (Quaking) (Table 4). QKI is a well-known and described SCZ candidate gene, first described as candidate gene for SCZ by Aberg *et al.*.[140] QKI regulated mRNA splicing, export of RNA from the nucleus, protein translation, as well as mRNA stability. It is expressed in various isoforms and regulates several cellular processes as embryogenesis, blood vessel development, glial cell fate determination, apoptosis, as well as protein translation.[141] For SCZ specifically, it is suggested, that induced myelin and oligodendrocyte dysfunction may be the role of QKI isoforms.[140]

Table 4: Set based test of ZNF804A interactors based on PGC2 dataset.

	NSNPS	PVALUE	PVALUE CORRECTED
DAZAP2	11	0,15284715	2,7512487
EYA3	223	0,42057942	7,57042956
FAM46A	13	0,37062937	6,67132866
KRTAP19-1	-	-	-
KRTAP19-7	-	-	-
KRTAP8-1	2	0,53946054	9,71028972
LASP1	164	0,07992008	1,43856144
LITAF	177	0,42557443	7,66033974
PTS	11	0,44555445	8,0199801
QKI	326	0,000368	0,006624
RBM4	24	0,00375996	0,06767928
RBM46	87	0,80919081	14,5654346
RBM4B	16	0,0189981	0,3419658
SEMA4G	23	0,08991009	1,61838162
STAM2	147	0,01079892	0,19438056
STAT2	25	0,06893107	1,24075926
TSC1	107	0,74725275	13,4505495
UBE2I	109	0,02959704	0,53274672

To further evaluate the general association with SCZ of identified ZNF804A interactors, the Schizophrenia Gene Resource 2 (SZGR 2.0) Database was used. SZGR 2.0 contains schizophrenia associated genes from multiple datasets for common variants, *de novo* mutations, copy number variants, differentially expressed genes and differentially methylated genes.[142] Over 60% (DAZAP2 (DAZ Associated Protein 2), EYA3 (EYA transcriptional coactivator and phosphatase 3), KRTAP19-7 (Keratin Associated Protein 19-7), LASP1 (LIM And SH3 Protein 1), PTS, QKI, RBM4, RBM4B (RNA Binding Motif Protein 4 B), SEMA4G (Semaphorin 4G), STAT2 and STAM2) of identified ZNF804A PPI partners were identified to be significantly associated with SCZ. Although it is to mention, that the SZGR 2.0 database sets a very low threshold for SCZ association and therefore associates

a lot more genes with SCZ than other data sources. Most of these SCZ genes were associated with SCZ through at least one SNIP in the PGC2 dataset reaching nominal significance (Figure 11 and Table 5).

Nevertheless, over 60% of the 18 completely unknown protein-protein interactions for ZNF804A were classified to be SCZ associated, underlining the impact of ZNF804A itself for SCZ.

Table 5: SCZ associated PPI partners of ZNF804A (based on SCRG 2.0)

GENE ID	SYMBOL	OFFICIAL FULL NAME	SCZ ASSOCIATION
5936	RBM4	RNA binding motif protein 4	Common variant
83759	RBM4B	RNA binding motif protein 4B	Common variant
9802	DAZAP2	DAZ associated protein 2	Common variant Differently methylated
2140	EYA3	EYA transcriptional coactivator and phosphatase 3	Common variant Differently methylated
3927	LASP1	LIM and SH3 protein 1	Common variant Literature co-occurance Literature co-occurance
5805	PTS	6-pyruvoyltetrahydropterin synthase	GO_Annotation_neuronal Meta analysis Literature co-occurance
9444	QKI	Quacen	Common variant Differently methylated Literature co-occurance
337974	KRTAP19-7	keratin associated protein 19-7	Common variant
10254	STAM2	signal transducing adaptor molecule 2	Common variant Meta analysis Differently methylated
57715	SEMA4G	semaphorin 4G	Common variant GO_Annotation_neuronal
6773	STAT2	signal transducer and activator of transcription 2	Common variant

To get a general inside in the functional involvement of the newly identified ZNF804A interactors, a GO analysis was performed using the browser based DAVID tool. The two most prominent biological processes implicated were RNA splicing and mRNA processing, mainly due to the amount of RBMs (RNA Binding

Motive proteins) within the ZNF804A interactors (Table 6). More interestingly, the third biological process, identified by GO to be enriched, was entrainment of circadian clock by photoperiod associated (Table 6). Circadian rhythm disturbance is a common feature in psychiatric disorders, especially in SCZ with today undissolved underlying cause.[143] Other significant enriched biological processes were myelination as well as positive and negative regulation of translation (Table 6). Myelination is discussed to be a potential key disrupted process in the developing brain, leading to a high risk of developing SCZ.[144] The GO category of molecular function didn't reveal new insides, but validated the RNA and nucleotide binding properties of identified ZNF804A interactors (Table 7).

Table 6: GO biological processes analysis of identified ZNF804A interactors

BIOLOGICAL PROCESS	GENES	PVALUE
RNA SPLICING	3	6,0E-3
MRNA PROCESSING	3	7,0E-3
ENTRAINMENT OF CIRCADIAN CLOCK BY PHOTOPERIOD	2	1,4E-2
MYELINATION	2	3,2E-2
REGULATION OF TRANSLATION	2	3,7E-2
NEGATIVE REGULATION OF TRANSLATION	2	4,1E-2

Table 7: GO molecular function analyzes of identified ZNF804A interactors

MOLECULAR FUNCTION	GENES	PVALUE
INTERMEDIATE FILAMENT	3	4,3E-3
CYTOPLASM	11	5,0E-3
NUCLEOPLASM	6	8,4E-2

To strengthen the initial functional annotation, intensive literature research was performed and revealed three major functional groups within the ZNF804A interactors.

The biggest prevalent group, similar to the GO annotation, were RNA binding proteins, represented by DAZAP2, FAM46A, QKI, RBM4, RBM4B and RBM46, suggesting a regulatory role of ZNF804A for target gene expression, as also reported previously for ZNF804A.[87,145] QKI, as described above, is clearly one of the most relevant proteins of the identified ZNF804A interactors, simply because of its broad functional implications and clear association with SCZ. DAZAP2 functions as a TCF-4 (a well-known SCZ associated gene) interacting partner.[146]

The knockdown of DAZAP2 was shown to reduce the activity of Wnt-signaling, as well as altering the expression of Wnt-signaling target genes.[147] The Wnt signaling pathway is an evolutionarily conserved pathway that regulates various aspects of for example cell fate determination, cell migration, cell polarity and neural patterning, mainly described during embryonic development.[148] FAM46A belongs to the FAM46 family of poly A polymerases and is not very well described.[149]

As said before, RBM proteins carry an RNA binding domain and are therefore described as RNA binding proteins. RBM4 is mostly described as a spicing factor, interestingly involved in neuronal development and neurite outgrowth.[150,151]

RBM4B, and RBM46 are not very well explored proteins. RBM4B is suggested to play a role in the developing central nervous system and RBM46 is described as mainly important for trophectoderm differentiation.[152,153]

The second large functional group of proteins, also validating findings of the GO annotation, were described to play a role in the circadian rhythm or clock, including DAZAP2, EYA3, RBM4, RBM4B, STAT2 and TSC1 (Tuberous sclerosis 1). EYA3 is involved in pathways of DNA double-strand break repair and response and is a target for cancer research.[154] TSC1 is an mTOR suppressor and therefore interesting in cancer research, but was also associated with epilepsy, mental retardation and autism spectrum disorders.[155–157]

48

The third very interesting functional group includes genes which are associated with inflammatory pathways, represented by DAZAP2, LITAF, STAT2 and TSC1. Besides DAZAP2 involvement in Wnt signaling reductions, it is also described to be inhibitory for example for IL-25 and IL-17 signaling.[158,159]

Studies in mice showed, that LITAF deficits lead to a delayed response in proinflamatory cytokines and a prolonged persistence of anti-inflammatory cytokines after exposure to lipopolysaccharide.[160,161]

STAT2 is a transcription factor and together with STAT1 critical for the signal transduction pathway of type I interferons as described in the introduction. Briefly, after INF induction, it is described to translocate from the cytoplasm into the nucleus to bind to the ISRE promoter, inducing ISRE controlled target gene expression and a so called anti-viral-state of the cell.[94]

In its role as mTOR suppressor, TSC1 serves as a key regulator of inflammatory responses after bacterial stimulation in monocytes, macrophages and primary dendritic cells. Considering the inflammation hypothesis for SCZ, this group of ZNF804A interactors was of special interest.[162]

GO analysis highly depend on the quantity of input genes and the 18 identified ZNF804A interaction partners as described above therefore had a reduced informative value. To overcome this limitation, a HIPPIE network which included all direct interactors (medium high HIPPIE confident score of ≥0.63) of ZNF804A and its 18 identified PPI partners was created. The resulting network included 950 proteins and 1090 PPIs (Figure 11 A).
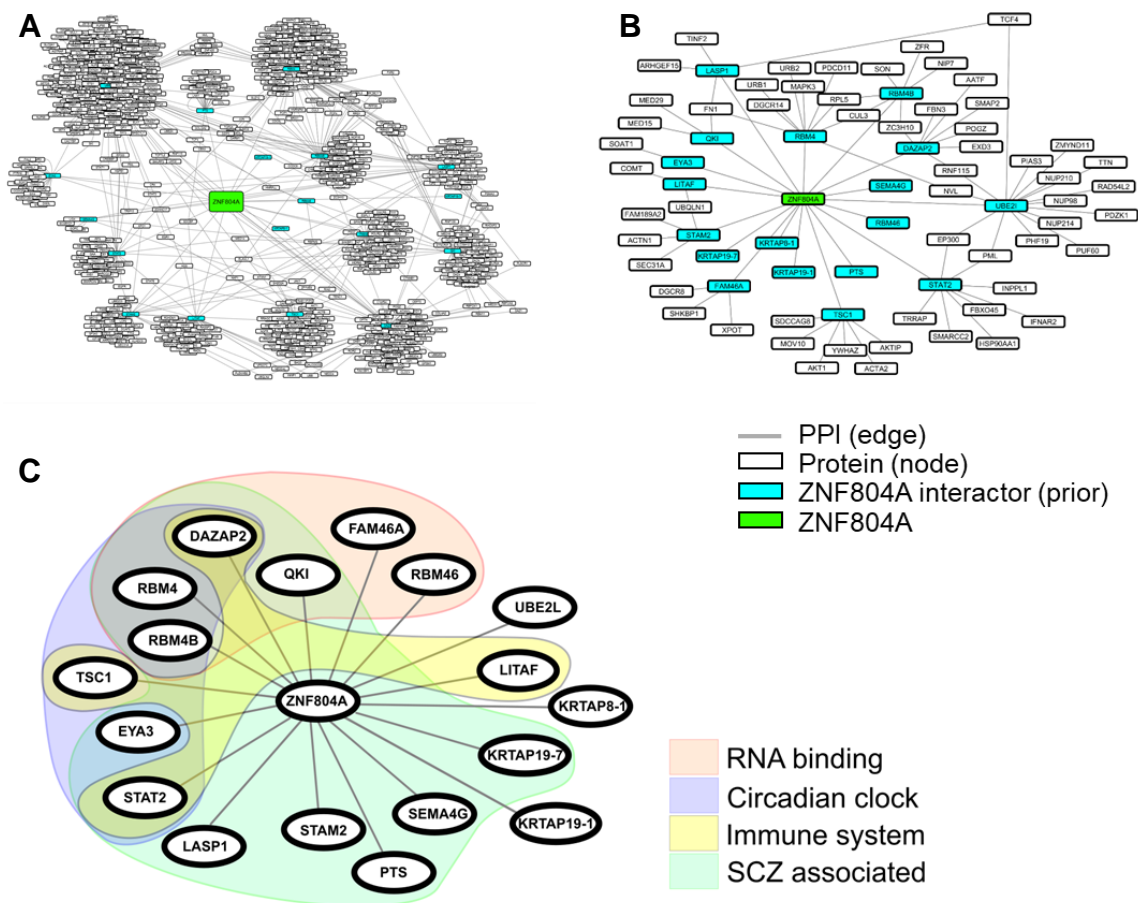
*Figure 11: **ZNF804A interactors interact with SCZ associated proteins.** (a) HIPPIE based network of all medium high confidence of ≥0.63 of all Y2H identified ZNF804A interactors. The network consists of 950 proteins and 1090 interactions. (b) Reduced HIPPIE based network of medium high confidence to SCZ related proteins to Y2H identified ZNF804A protein interactors. The network consists of 76 proteins and 86 interactions. Green square: ZNF804A, blue squares: Identified ZNF804A interactors, white squares: Proteins of the HIPPIE database.*

To condense the network to be SCZ specific, the network was reduced again to only contain proteins associated with SCZ, besides the initial ZNF804A interactors. As a source of known SCZ interactors, the list of SCZ related proteins created for the cluster algorithm was used. The resulting SCZ specific network consisted of 76 proteins and 86 PPIs (Figure 11 B). GO analysis was run for the SCZ specific ZNF804A HIPPIE based network similar to before (Table 8). The Go annotations did not rank any biological processes involved in circadian clock highly, but heavily represented viral processes such as viral process, viral transcription and intracellular transport of virus. This led to the assumption, that the cellular immune system, more specifically the viral response, might be a promising biological process for further analysis on the functionality of ZNF804A.

50

Table 8: GO biological processes analysis of SCZ specific network based on identified ZNF804A interactors

BIOLOGICAL PROCESSS	NUMBER	P-VALUE
REGULATION OF CELLULAR RESPONSE TO HEAT	7	5,20E-07
GENE SILENCING BY RNA	7	5,20E-06
TRNA EXPORT FROM NUCLEUS	5	7,70E-06
VIRAL PROCESS	9	2,40E-05
PROTEIN SUMOYLATION	6	1,00E-04
REGULATION OF GLUCOSE TRANSPORT	4	3,00E-04
MITOTIC NUCLEAR ENVELOPE DISASSEMBLY	4	7,20E-04
VIRAL TRANSCRIPTION	5	1,00E-03
INTRACELLULAR TRANSPORT OF VIRUS	4	1,10E-03
REGULATION OF SIGNAL TRANSDUCTION BY P53 CLASS MEDIATOR	5	1,50E-03
PLATELET FORMATION	3	2,60E-03
PROTEIN EXPORT FROM NUCLEUS	3	6,30E-03
MRNA EXPORT FROM NUCLEUS	4	7,50E-03

To summarize, the identified ZNF804A interactors play important roles in multiple SCZ relevant processes with the cellular immune system as one of the most interesting for analysis of ZNF804A implications.

Up to this point, all ZNF804A analyses were based on the results of the Y2H screen. As every PPI method, the Y2H has its limitations. As described earlier, stringent cut off rates for positive PPI consideration were used and corrections for false positives were made. Still the best way to validate true PPI interactions is by using different PPI methods. Therefore, to strengthen and validate the Y2H based ZNF804A interaction data, two independent and quantitative PPI detection methods in mammalian cells were used: DULIP and LuTHy.

4.3.3 DULIP: Immunoprecipitation based assay validates of 56% of ZNF804A PPIs

The DULIP assay is based on the co-immunoprecipitation of firefly tagged prey proteins by PA-*Renilla* tagged bait proteins (Figure 12 A). After successful IP, using the bait PA-tag, the bait firefly, as well as the prey *Renilla* luminescence can be measured separately. IP based PPI methods have the tendency to recover mainly high affinity interactions and are dependent on high target protein production. For that reason, the ZNF804A fragments ZNF804A-1-400 and ZNF804A-800-1209, which together were able to recover all ZNF804A PPIs in Y2H, were chosen for the DULIP validation. Both ZNF804A fragments, as well as all identified ZNF804A interactors were systematically cloned into n-terminal DULIP vectors (Figure 12 A). pPA-RL-ZNF804A and pFL-V5-ZNF804A-800-1209 were co-transfected with all of their interactors identified in Y2H (STAT2 for fragment ZNF804A-1-400 and 17 proteins for ZNF804A-800-1209 fragment) fused to either PA-RL (run1) or FL-V5 (run2) respectively in HEK293 cells in 96 well format. All transfections were performed in triplicates and in 3 biological triplicates. After 48 h of co-expression, cells were lysed and lysates transferred into antibody coated 384 well plates and washed surely. The PA-tag of expressed bait proteins, bound to the heavy chain of coated antibodies, reflected by *Renilla* signal. Co-immunoprecipitated prey proteins were quantified by firefly luminescence. To take expression levels into consideration, lysates were additionally measured separately in an uncoated 384 well plate. Interactions were excluded if at least one target protein lacked considerable expression. To exclude background binding of either target protein, all used constructs were tested against pPA-RL-mCherry and pFL-V5-mCherry as negative controls. Through quantification of luciferase activities of precipitated interacting proteins and background correction, normalized interaction ratios (cNIRs) were calculated for all tested interactions (Figure 12 B).

These ratios are an indication of interaction strength, allowing the distinction of potentially high from low affinity interactions.[65] A stringent cNIR cut off of four to define DULIP-positive PPIs was used. Under these conditions ten out of 18 (56%) tested Y2H interactions were validated in DULIP assays (Figure 12 B and C). High cNIR values indicate high interaction strength. This was particular represented by

the interaction between ZNF804A-800-1209 with RBM46 as well as for the N-terminally ZNF804A fragment ZNF804A-1-400 with STAT2. Y2H validation through other PPI methods has been published to recover between 10-34% of the interactions, using reference sets.[65,66,163] Therefore, DULIP validation rate of 56% indicates that the identified ZNF804A PPI network is of high confidence. Furthermore, the by DULIP suggested high interaction strength of ZNF804A to STAT2, the only N-terminal interactor of ZNF804A, strengthened the interest to analyze a potential involvement of ZNF804A in the STAT2 mediated immune system pathway.
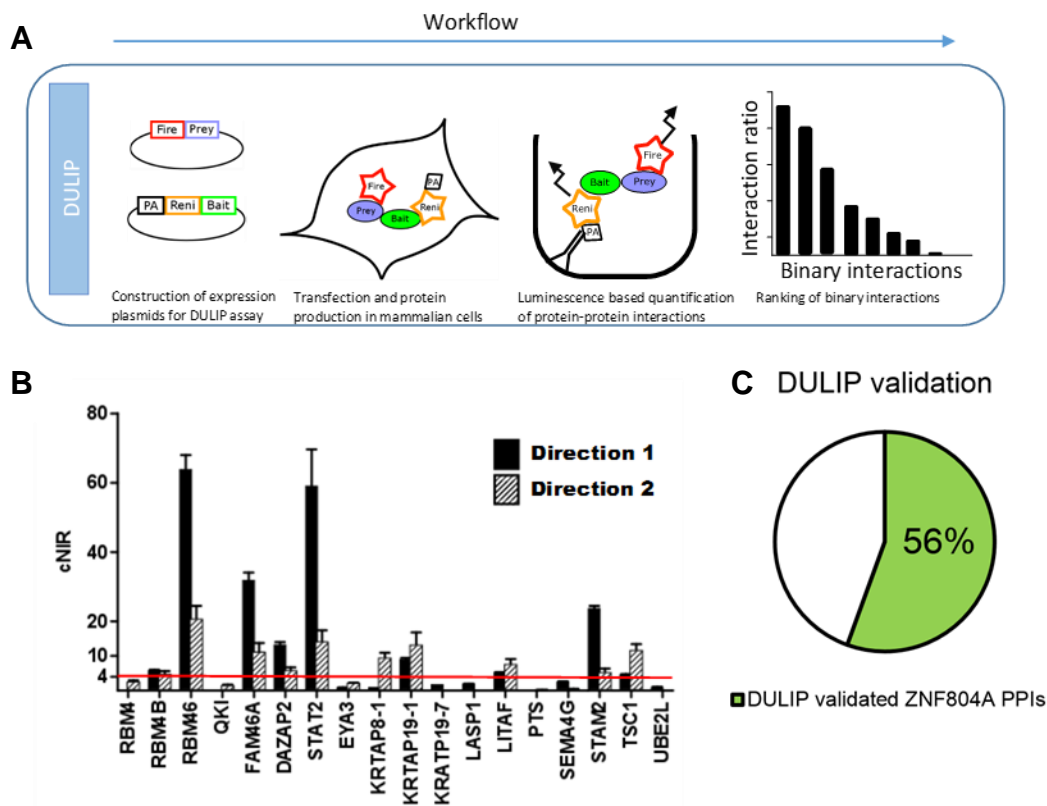
*Figure 12: **DULIP validation of identified ZNF804A interactors.** (a) Flow chart of DULIP validation. (b) Luminescence-based interaction ratios from the DULIP validations of ZNF804A interactons. Values are displayed as a bar diagram (means ± SEM of three biological replicates). PPIs surpassing the cNIR threshold (red line) are considered positive. ZNF804A was tested as bait (direction 1) as well as prey (direction 2). (c) Pie diagram summarizing the overall achieved validation rate of ZNF804A.*

4.3.4 LUTHI: Verifying ZNF804A interactions using ZNF804A full length proteins in intact cells

Up to this point no PPIs were identified nor validated with the full length ZNF804A protein. To overcome this limitation, in cell LuTHy assays were used to validate interactions with the full length protein.[66] Co-IP based readout was not utilized for validation purposes.

ZNF804A-NanoLuc fusion proteins were screened against all 18 with Y2H identified PPI partners fused to PA-mCitrine in triplicates and in four independent repetitions. Similar to DULIP, background correction was achieved by also screening every pcmyc-NL- and pPA-mCit construct against NanoLuc and PA-mCitrine only vector controls respectively. With this, normalized BRET efficiency (cBRET) were determined for all tested interactions (Figure 13 A).[66] A restrictive threshold of 0.2 was used to determine interactions as positive.[66] Additionally, all possible interactions were excluded if at least one interaction partner was not expressed as published.[66]

Eight out of 18 interactions, identified in Y2H for ZNF804A, surpassed the set cBRET threshold and were successfully validated with LuTHy (Figure 13 B). This validation rate of 44% also exceeds those published and, because ZNF804A full length constructs were used, strengthens the relevance of identified PPIs for endogenously expressed ZNF804A. STAT2, one of the two most prominent DULIP validated interactors of ZNF804A, was also verified with LuTHy, further supporting the hypothesis that ZNF804A might play a role in the INF/STAT2 pathway (Figure 13 A). RBM46 was not validated using LuTHy.
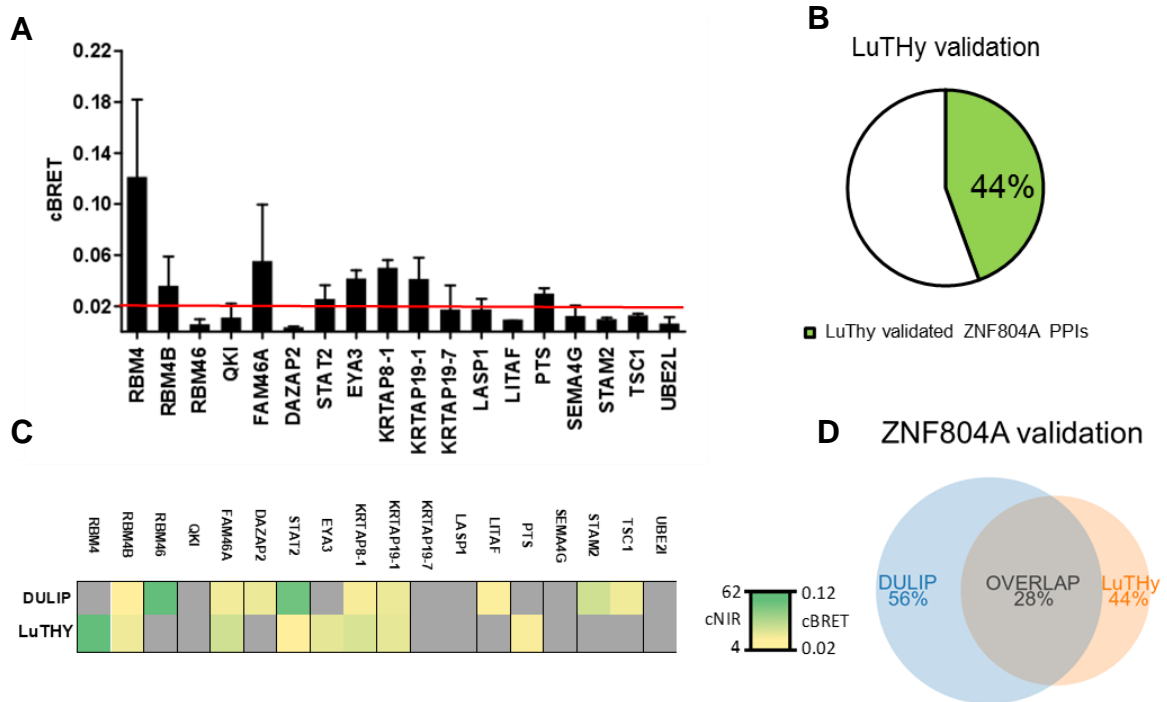
*Figure 13: **LuTHy validation of identified ZNF804A interactors.** (a) cBRET ratios of ZNF804A interactons analysed with LuTHy. Values are displayed as a bar diagram (means ± SEM of three biological replicates). PPIs surpassing the cBRET threshold (red line) are considered positive. (b) Pie diagram summarizing the achieved validation rate of ZNF804A using LuTHY. (c) Summary of DULIP and LuTHy validated PPIs in table format. Coloring indicates interaction strength, specified by cNIR and cBRET values (yellow to green). Grey coloring indicates non validated interactions.(d) Venn-diagram depicting the achived validation rates using DULIP and LuTHY as well as the overlap of both.*

4.3.5 Dual method based validation approach validates 67% of identified ZNF804A PPIS

Taken together, twelve out of 18 Y2H identified ZNF804A interactions were successfully validated with either DULIP and/or LuTHy assays, representing a validation rate of 67% (Figure 13 C and D). A total of five interactions (28%) were validated with DULIP as well as with LuTHY (STAT2, RBM4B, KRTAP8-1, KRTAP19-1, FAM46A), depicted in Venn diagram (Figure 13 D). Every PPI method has its limitation and it is unlikely to recover 100% of interactions with two different methods. One example is the potential loss of low affinity interactors, due to the IP

55

based character of DULIP after lysis of cells and wash steps after IP. On the other hand, for the occurrence of BRET in LuTHy assays the orientation of the tags and the total distance is of importance. Y2H is very sensitive but has difficulties to recover interactions with membrane bound or very large proteins. Therefore, multi-assay approaches strengthen the overall validity of identified PPIs.

Taken together, the high validation rate of 67% led to the assumption that the identified ZNF804A indeed were relevant for ZNF804A biological function.

4.3.6 STAT2 – A promising ZNF804A interactor

To elucidate the potential functional implications of ZNF804A, I decided to focus on one specific interaction and examined the potential role of ZNF804A in the pathway, implicated with this interactor. I focused on an interaction, that has been identified with all used methods as being most reliable. Additionally, the implicated pathway had to be associated with SCZ and had to be feasible for me to analyze. From the five interactions, tested to be positive by all three used PPI methods (RBM4B, FAM46A, KRTAP8-1, KRTAP19-1 and STAT2) the interaction between ZNF804A and STAT2 stood out. It has been the only PPI taking place within the n-terminal region of ZNF804A and it had one of the highest cNIR scores in DULIP assays. Most importantly, STAT2 as a key mediator of the cellular INF triggered immune answer, was by far the most promising target for further analyses, especially taking the immune system hypothesis of SCZ into consideration. More specifically, STAT2 is a transcription factor, critical to signal transduction pathways of type I interferons. STAT2 resides primarily in the cytoplasm. After INFα/β binding to IFN receptors (IFN-R) homodimerise, recruit janus kinases proteins, which lead to STAT2 tyrosine phosphorilation. Phosphorilated STAT2 homo or hetero dimerises mainly with STAT1. Dimerised STAT2/STAT2 or STAT2/STAT1 complexes bind IRF9 and form the ISGF3 (Interferon-stimulated gene factor 3) complex. The formed complex translocates into the nucleus where it initiates the expression of ISRE promotor controlled viral response genes. A potential modulating function of ZNF804A in this pathway would have highly important implications for its role in SCZ.

4.3.6.1 Identification of ZNF804A binding sides to STAT2

First, to specify the binding sides of the ZNF804A n-terminus to STAT2, a peptide array, composed of peptides derived from the first 400 amino acids of ZNF804A were designed. Since the domain architecture of ZNF804A is mainly unknown, the 400 amino acids were divided into 15 residues long peptides with overlap of five residues for the array design (table in the method section: Table 17). INVATIS Celluspot arrays were used with 96 peptide spots in duplicates on one array. As a control, alongside the ZNF804A peptides, the with STAT2 high affinity interacting domain (as 205-390) of the interferon-regulatory factor associated (IRF9) were spotted in a similar fashion. IRF9 belongs to the IRF family, which plays important roles in immune responses to viral infections, cytokine signaling, cell growth regulation and hematopoietic development.[164] After interferon dependent STAT2 dimerization, IRF9 binds to STAT2/STAT2 or STAT2/STAT1 complexes and co-migrates into the nucleus to activate ISRE reporter dependent transcription response. The principle of the peptide array is the utilization of the binding affinity of isolated protein of interest to the spotted peptides. The higher the binding affinity to a specific spot, the more protein will be bound and stick after washing steps and can finally be detected via immuno detection.

Before peptide array assays were performed it had to be clarified that the STAT2 antibody to be used was specific for the detection of STAT2 and not cross detecting other proteins of the STAT family. Therefore, all seven proteins of the STAT family (STAT1, STAT2, STAT3, STAT4, STAT5a, STAT5b and STAT6) were cloned in N-terminal GFP-tag vectors and transfected into HEK293 cells and analyzed in western blot using STAT2 antibody (STAT2(C-20): sc-476, SantaCruz). For all transfected cell samples, as well as for untransfected controls, endogenous STAT2 protein was detected in comparable levels at the expected size of 113 kDa (Figure 14 A). Only in GFP-STAT2 transfected cell samples an additional band was detected at expected size of STAT2 fused to GFP of about 130 kDa. Therefore, the STAT2 antibody was verified to specifically detect only STAT2 protein, either untagged or tagged to GFP, but did not detect any other GFP-tagged STAT family proteins (Figure 14 A).

In a next step STAT2 protein had to be isolated. Thus, STAT2, as well as mCherry as a control, was cloned into N-terminal HIS tag encoding qQLink gateway vectors. Both constructs were transformed into E. coli BL21 (DE3) Rosetta expression clones, protein were produced in culture flasks and afterwards isolated. To load similar amounts of proteins of HIS-STAT2 and His-mCherry control, concentrations were calculated by dilution series loaded on SDS-gels and visualized by western blot (Figure 14 B) as well as with NanoDrop measurement.
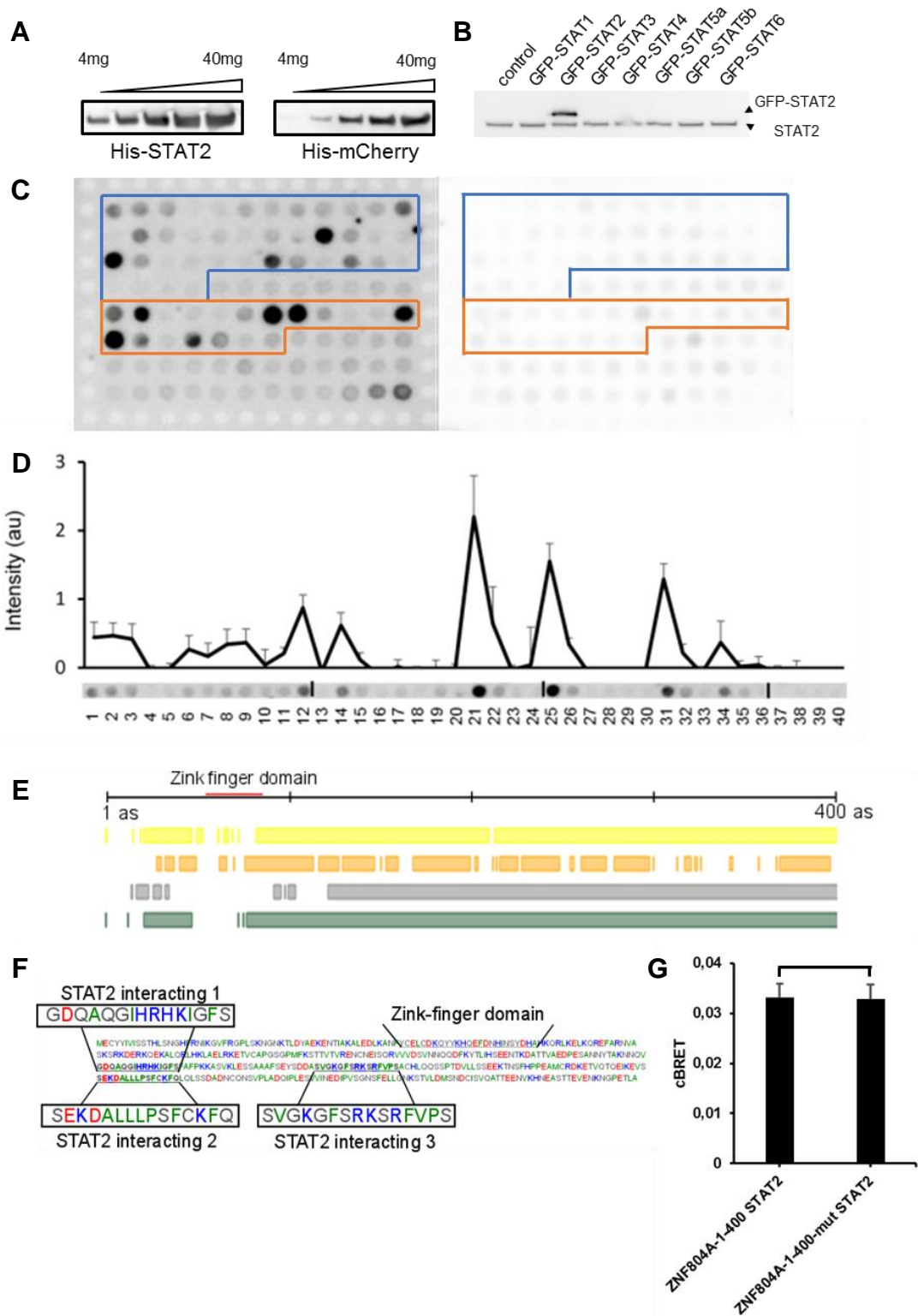
Figure 14: ZNF804A interacts with STAT2 like a intrinsically disordered protein. (a) WB validation of His tagged STAT2 and mCherry proteins, isolated from bacterial culture. For detection purposes STAT2 antibody (STAT2(C-20): sc-476, SantaCruz) and mCherry antibody (living color mCherry, Clontech) were used. (b) WB validation of the specificity of the STAT2 antibody to its target STAT2. Lysates of with GFP fused STAT family proteins were analyzed and detected with STAT2

antibody. In all samples the endogenous STAT2 levels were detected, indicated by arrowhead labeled with STAT2. Only transfected GFP-STAT2 HEK293 cell lysates showed an additional band (indicated by arrowhead labeled GFP-STAT2). (c) Peptide array of the first 400 AS of ZNF804A and incubated with His-STAT2 (left) or His-mCherry (right). For detection, STAT2 antibody was used. (d) The corresponding quantification of spot intensities of three independent replicates (intensities background corrected ± STD). (e) Evaluation of natively structured regions using PredictProtein.org analysis tools. Colored bars indicate predicted natively unstructured regions using the algorithms: Yellow: Schlessinger, Yachdav et al.[165], Orange: Schlessinger, Punta et al.[166], Grey: Schlessinger, Lui et al.[167], Green: Schlessinger, Punta 2 et al..[168] (f) Display of GenScript peptide property calculator tool analysis of the first 400 amino acids of ZNF804A. Indicated are the zink finger domain, as well as the three strongly STAT2 binding regions. Colors are indicating polarity of as; blue: Basic residues (R, K and H), green: Hydrophobic uncharged residues (F, I, L, M, V, W, A and P), red: Acidic residues (D and E), black: Other residues (G, S, T, C, N, Q and P). (g) Comparison of the interaction strength between STAT2 and the n-terminal region of ZNF804A with (ZNF804A-1-400) or without the ZNF domain (ZNF804A-1-400-mut) using LuTHy assays (means ± SEM of three biological replicates).

For the actual peptide array, blocked arrays were incubated with 20µM HIS-STAT2, HIS-mCherry or without any protein over night and developed with STAT2 specific antibody (STAT2(C-20): sc-476, SantaCruz) (Figure 14 C) .

Each black spot on the developed array represents a ZNF804A (Figure 14 C (left)) or IRF9 (Figure 14 C (right)) derived peptide that was detected by STAT2 specific antibodies. Arrays incubated with HIS-mChery control protein showed no specific detection pattern. Therefore, STAT2 antibody binding to unspecific proteins or directly to spotted array peptides was excluded as a likely possibility (Figure 14 A (right)). In contrast, incubation with HIS-STAT2 protein led to a clear and reproducible detection pattern for spotted ZNF804A and IRF9 peptides. For IRF9, STAT2 detection showed clear and strong binding motives for STAT2 with residues spanning over two to four peptide spots (20-40 as) mainly at amino acid 205-225, 265-285 and 325-340 (Figure 14 C). This was the expected result for a domain dependent protein interaction, because domain-based interactions span normally over several amino acids to form an interaction surface. The detection pattern for the ZNF804A peptides was different. Instead of detecting groups of spots, indicating interaction motives, three strong and distant small binding areas were detected (15-20 as) (Figure 14 D). This pattern was most likely emerged due to an intrinsically disordered character of the interaction surface of ZNF804A to STAT2. The interaction with a intrinsically disordered protein mainly

depends on ionic amino acid bonds.[169,170] As covered above, the n-terminal region of ZNF804A is primarily disordered, besides its zink finger domain. All the detected high binding spots of STAT2 lay outside of this zink finger domain. To confirm that binding regions of ZNF804A to STAT2 indeed are following properties of natively unstructured proteins, the GenScript peptide property calculator tool was used (Figure 14 E). The two most strongly interacting regions of ZNF804A are highly ionic with an isoelectric point of 9.69 and 12.53 respectively, further underlining that indeed STAT2 and ZNF804A are most likely interacting via a intrinsically disordered interaction interface (Figure 14 F). In order to verify the intrinsically disordered nature of ZNF804A the PredictProtein.org database was used. All three implemented algorithms determined the n-terminal domain of ZNF804A as natively unstructured, except for the zinc finger domain (Figure 14 E).

To confirm that the zink finger domain of ZNF804A does not contribute to the PPI between ZNF804A and STAT2, a BRET assay was performed with PA-mCitrine-STAT2 and NanoLuc-ZNF804A-1-400, as well as with a NanoLuc-ZNF804A construct lacking the zink finger domain. The cBRET ratio and therefore the interaction strength between ZNF804A fragments and STAT2 were not significantly changed by the domain deletion (Figure 14 G).

In summary, the interaction between ZNF804A and STAT2 was identified to be independent from the ZNF domain and follow the binding pattern of PPIs of an interaction with an intrinsically disordered protein. STAT2 seems to bind distinctly to three small and polar amino acid sequences within the n-terminus of ZNF804A.

4.3.6.2 ZNF804A and STAT2 co-migrate into the nucleus after INF treatment and have the tendency to form perinuclear speckles

After specifying the binding patterns of ZNF804A and STAT2, the next step was to analyze the interaction between ZNF804A and STAT2 in a cellular context. Thus, both proteins had to be stained via antibodies, followed by confocal microscopy analysis. To determine a feasible cell line, human HAP, SH-SY5Y, SH-EP, HEK293 cell lines, a sample of primary neurons as well as rat PC12 and two samples of rat hippocampal cells were analyzed in western blot using ZNF804A and STAT2 specific antibodies (Figure 15A-B).

61
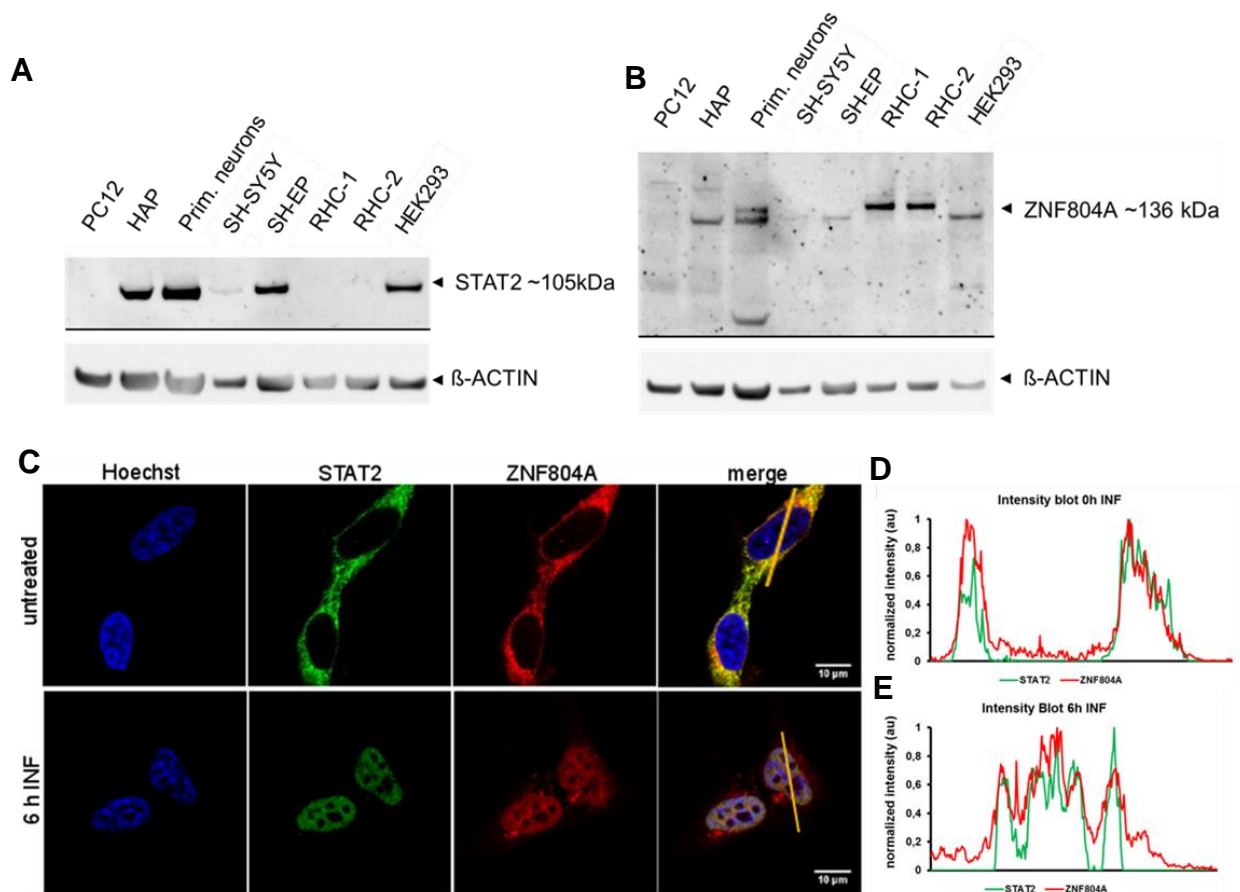
*Figure 15: **ZNF804A co-translocate with STAT2 into the nucl**eus. (a-b) Western blot evaluation of endogenous ZNF804A (a) and STAT2 (b) expression in different cell lines respectively. ßActin served as a loading control. Arrowheads indicate detected proteins and sizes. (c) Immuno flouorescence confocal microscopy pictures of SH-EP cells treated with or without INF2α for 6 h. Blue: Hoechst 33342 staining of the nuclei (Ex/Em 353/483), red: ZNF804A antibody staining, detected with Alexa FLuor® 568 secondary antibody (Ex/Em 573/603), green: STAT2 antibody staining and detected with Alexa FLuor® 488 secondary antibody (Ex/Em 495/519), yellow lines indicate areas of 2D intensity blot analysis (d-e). (d-e) 2D intensity blots of STAT2 and ZNF804A on indicated areas to the left. Intensities are normalized to maximum intensities.*

The STAT2 antibodies were able to detect clear signals in all samples in the expected size of about 105 kDa, except for cells with rat origin (PC12 and RHC) and only a faint signal for SH-SY5Y cells (Figure 15 A).

For ZNF804A, the sizes and number of detected protein bands differed between cell samples, but in all cases at least a faint signal was detected (Figure 15) . This indicates that ZNF80A might exist in different splice variants. The lack of STAT2 detectability in cells of rat origin excluded those for further consideration. Because

of a big cell size, as well as the neuronal origin and general good transfection rates, SH-EP cells were chosen for all following in cell microscopy based analysis.

In order to detect endogenous STAT2 as well as ZNF804A with immuno fluorescence, proper fluorescence tagged secondary antibodies had to be used. For STAT2 and ZNF804A Alexa FLuor® 568 secondary antibody (Ex/Em 573/603) anti-rabbit and Alexa FLuor® 488 secondary antibody (Ex/Em 495/519) anti-mice secondary antibodies were used respectively. The co-staining of STAT2 as well as ZNF804A in fixed SH-EP cells revealed clear co-localization mainly in the cytosol (Figure 15 C). It is known, that STAT2 is retained mostly in the cytosol until interferon signaling triggers its translocation into the nucleus. Due to its zinc finger domain and implications in RNA binding and splicing, the expectations for ZNF804A was a primarily nuclear, rather than a cytosolic localization. In a next step cells were treated with 100 U/ml interferon 2α (INF2α) for 6 h to trigger expected STAT2 translocation into the nucleus. As expected, STAT2 translocated into the nucleus after 6 h of INF2α treatment (Figure 15 C). Strikingly, ZNF804A co-translocated into the nucleus and displayed a high co-localization with STAT2 (Figure 15 D). This finding not only validates the interaction of ZNF804A with STAT2 but also gives a potent indication that ZNF804A has a functional impact on STAT2 mediated INF response.

To better understand the interplay of STAT2 and ZNF804A, genes were cloned into n-terminal GFP and N-terminal mCherry vectors respectively, co-transfected and overproduced in SH-EP cells for 48 h.

Confocal microscopy of co-transfected GFP-STAT2 and mCherry-ZNF804A SHEP cells revealed that co over production of both tagged proteins lead to accumulation of GFP-STAT2 and mCherry-ZNF804A into perinuclear speckles (Figure 16). Formed speckles displayed no visual localization in the nucleus and showed a high degree of co-localization of mCherry-ZNF804A and GFP-STAT2 signal (Figure 16 A-B). Neither mCherry-ZNF804A, nor GFP-STAT over production alone displayed such perinuclear accumulation, pointing towards that perinuclear aggregates are the effect of STAT2 and ZNF804A interaction (Figure 16 A). INF2α treatment did not lead to translocation into the nucleus (data not shown).

To substantiate this finding, filter retardation experiments with lysate of HEK293 cells transfected with GFP-STAT2 and different concentrations of mCherry-ZNF804A or mCherry as a control were performed. Filter retardation assays only retain large protein accumulations, insoluble to SDS and resistant to boiling. After developing the filter retardation membranes with STAT2 antibody, concentration dependent increase of mCherry-ZNF804A led to a clear increase of STAT2 detectable signal (Figure 16 C-D). In contrast, mCherry did not show a concentration dependent effect on accumulation of STAT2 on filter membranes, indicating that speckles formation is indeed specific for co-transfection of STAT2 and ZNF804A (Figure 16 C-D).

Taken together, ZNF804A co-localizes with STAT2 and co-translocates into the nucleus after INFα treatment. Overexpression revealed the tendency of STAT2 and ZNF804A to co-aggregate into perinuclear speckles, which is specific and concentration dependent for ZNF804A and STAT2. Although this might be an arbitrary effect, it is another clear sign for the interaction strength between ZNF804A and STAT2.
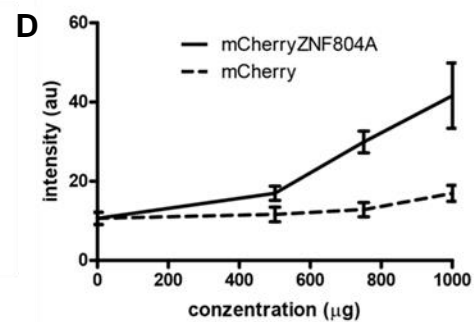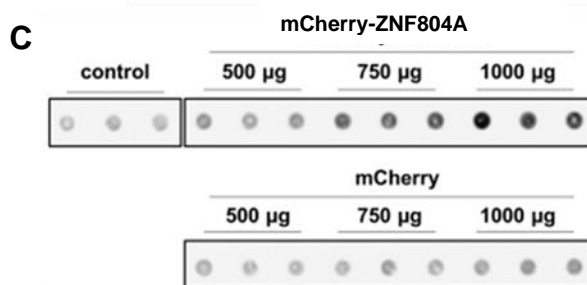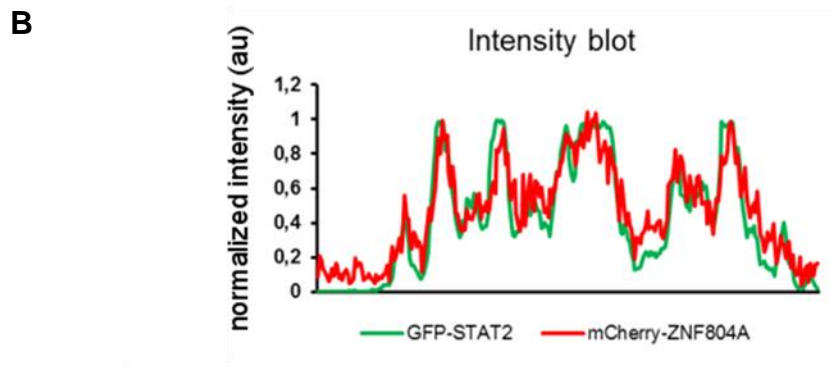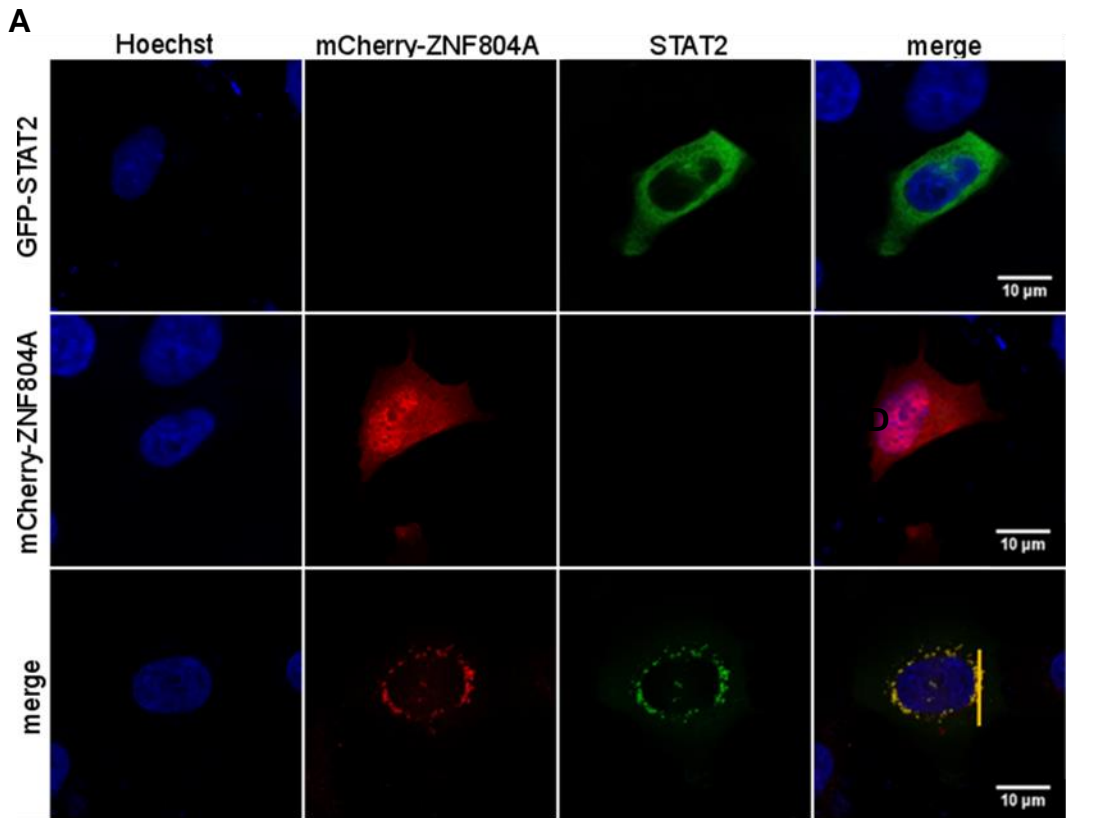
Figure 16: **Co-transfected GFP-STAT2 and mCherryZN6F804A form perinuclear specales**. (a) Confocal microscopy pictures of SH-EP cells transfecter with either or both GFP-STAT2 and mCherryZNF804A. Blue: Hoechs33342 staining of the nuclei (Ex/Em 353/483), red: mCherryZNF804A flouorescence signal (Ex/Em 353/483), green: GFP-STAT2 flouorescence signal (Ex/Em =580/610), yellow line indicate the area of 2D intensity blot analysis in b. (b) 2D intensity blot of GFP-STAT2 and mCherryZNF804A signal on indicated line above. Intensities are normalized to maximal intensities. (c) Filter retardatipon blots

of 1000 ug GFT-STAT2 transfected SHEP cells co-transfected with either mCherry-ZNF804A, mCherry or with empty vector (control). STAT2 antibody used for detection. (d) Graphical evaluation of GFP-STAT2 aggregate formation in dependency of mCherryZNF804A or mCherry concentration.

4.3.6.3 ZNF804A modulates STAT2 mediated INF transcriptional response

After the identification of ZNF804A co-translocation into the nucleus with STAT2 due to INF2α treatment, the further focus was to evaluate the potential impact of ZNF804A on STAT2 function. As mentioned, STAT2 is a key regulator of interferon mediated ISRE promoter target gene expression. Therefore, an ISRE reporter assay was established (QIAGEN) to evaluate potential effects of ZNF804A on STAT2 functionality. The ISRE reporter assay utilized two luciferases, firefly luciferase and *Renilla* luciferase. The firefly luciferase was transcriptionally under the control of the ISRE promoter as the main STAT2 promoter target. The *Renilla* luciferase was under the constitutively active CNV promotor control. These two readout strategies allowed the evaluation of changes in ISRE activation via firefly signal, while normalizing to the *Renilla* signal and therefore, correct for different transfection efficiencies and cell densities (Figure 17 A).

The ISRE assay also allowed the correction for other possible factors arbitrarily increasing the firefly signal. To do so, all conditions had to be additionally tested with the constrictively active *Renilla* vector together with a firefly vector without any promoter sequence (Figure 17 A). The transfection of both luciferases under the control of the constitutive CMV promoter served as an additional positive transfection control. To determine the optimal INFα concentration, HEK293 cells were co-transfected with ISRE reporter constructs first. The response to the 24 h treatment was measured with increasing amounts of INFα. The EC 65 was determined to be 50 U/ml INFα (Figure 17 B). This concentration was used for all following assays that measured potential changes of ISRE reporter response to INF2α treatment (Figure 17 C-E).
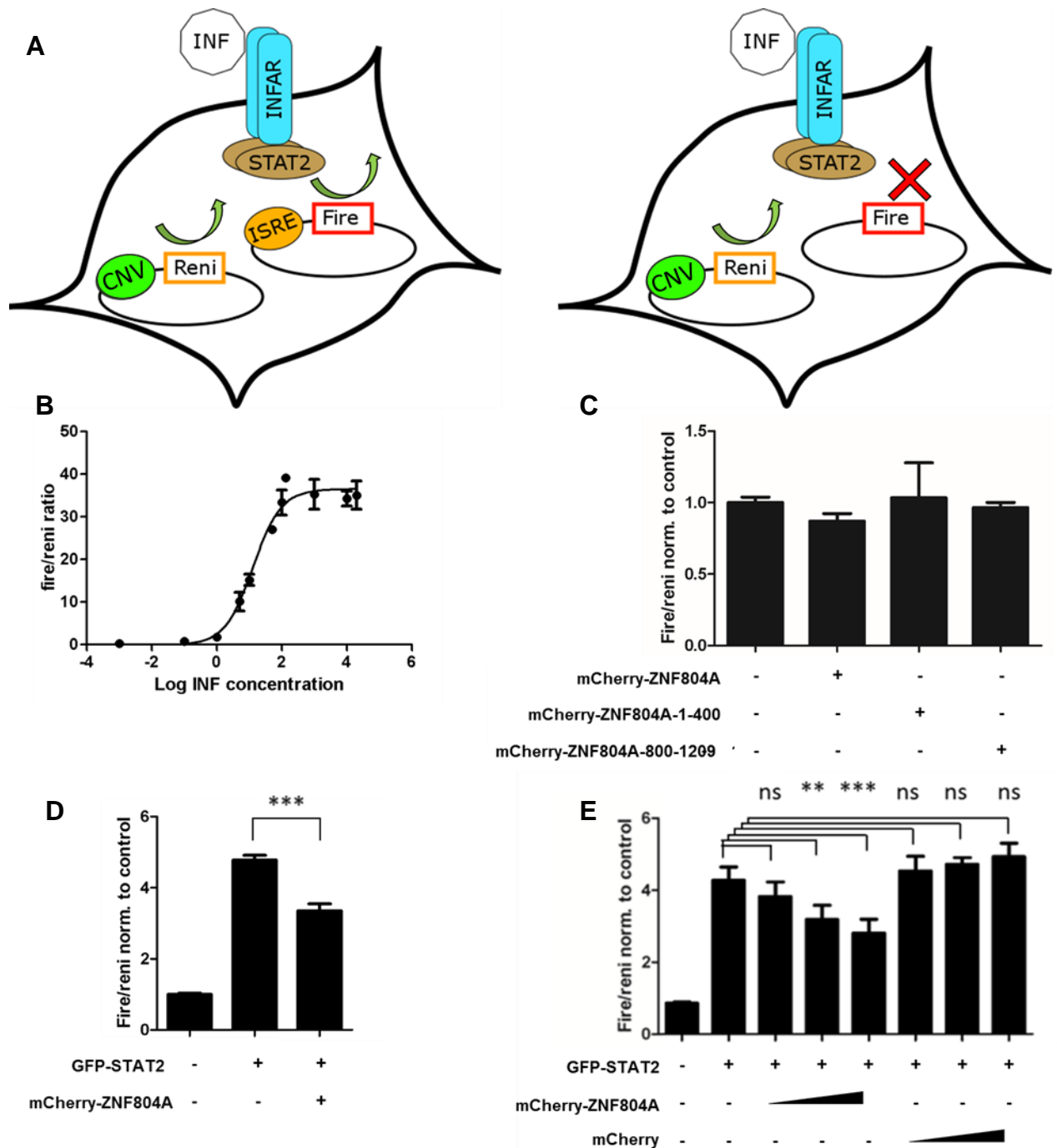
Figure 17: **Evaluation of the effect of ZNF804A on ISRE response**. (a) Schematic depiction of ISRE assay. Left: Depiction of co-transfected cell with both luciferase constructs, Renilla luciferase (Reni) under the consticutive CNV promoter (CNV) and firefly luciferase (Fire) under the ISRE promoter control. Right: Control co-transfection with Renilla luciferase (Reni) under the constitutive CNV promoter (CNV) and the firefly luciferase (Fire) without any promoter. Green arrows indicate expression. Red mark indicate no expression. (b) Determination of the EC65 for HEK293 cell response after 24 h of INF induction, measured by ISRE assay. (c) Measurement of ISRE reporter response to 24h INF treatment and transfection of diffrent mCherry-ZNF804A constructs (100 ng). (d) Measurement of ISRE reporter response to 24h INF treatment and co-transfection of GFP-STAT2 (100 ng) together with mCherry-ZNF804A (100 ng). (e) Measurement of ISRE reporter response to 24h INF treatment and co-transfection of GFP-STAT2 (100 ng) together with either mCherry-ZNF804A (50 ng, 75 ng and 100 ng) or mCherry (50 ng, 75 ng and 100 ng).

To evaluate the impact of ZNF804A on ISRE reporter response, not only the full length construct of ZNF804A was overproduced in HEK293, but also the STAT2 interacting n-terminal construct ZNF804A-1-400 and the non-interacting-terminal constructs ZNF804A-800-1209. To confirm expression of ZNF804A variants microscopically ZNF804A constructs were used as n-terminal mCherry fusions. All three fusion constructs, as well as mCherry alone as control were co-transfected (100 ng) with IRSE reporter constructs for 24 h in HEK293 cells. Afterwards they were treated with INF2α for additionally 24 h. After lysis both luciferases signals were measured separately. No significant changes were detectable after 24h of INF2α treatment for neither ZNF804A constructs nor mCherry (Figure 17 C). The lack of a significant effect of ZNF804A on IRSE reporter response could have been due to other members of the STAT family compensating potential effects on endogenous STAT2. To circumvent this possibility, in a second set of assays, mCherry-ZNF804A was co-transfected together with GFP-STAT2, as well as the reporter constructs. The overproduction of GFP-STAT2 in co-transfected reporter cells led to a fourfold increase of ISRE response (Figure 17 D). The additional co-transfection of mCherry-ZNF804A led to a significant reduction of GFP-STAT2 triggered ISRE response (p≤ 0.001) (Figure 17 C). Co-transfection of mCherry on the other hand did not change GFP-STAT2 mediated ISRE response (Figure 17 C).

The effect of mCherry-ZNF804A on GFP-STAT2 mediated ISRE response also showed a clear concentration dependency (Figure 17 D). Co-transfection of GFP-STAT2 together with mCherry did not change ISRE reporter response (Figure 17 D). Taking together, ISRE reporter analyses revealed a clear negative regulatory effect of ZNF804A overproduction on STAT2 overproduction induced ISRE response.

To confirm this potential role of ZNF804A, four genes (*IFIT1* (Interferon Induced Protein With Tetratricopeptide Repeats 1)*, ISG15* (ISG15 Ubiquitin-Like Modifier)*, MX1* (MX Dynamin Like GTPase 1) and *RASD2* (RASD Family Member 2)) were selected for qPCR validation. All four were previously reported to be regulated by STAT2, even in the absence of STAT1.[171] For qPCR validations primers were chosen from the Hravard PrimerBank (methods: Table 27: PrimerTable 27) while GBDH served as a control. All four of these genes are specific antiviral response proteins, heavily upregulated by Interferon. The direct mechanisms by which these genes respond to viral infections often vary with the species of the virus (*RSAD2, MX1*) and/or is not fully understood (*ISG15*). The main function of antiviral genes is to inhibit viral protein translation (*IFIT1, MX1*), as well as
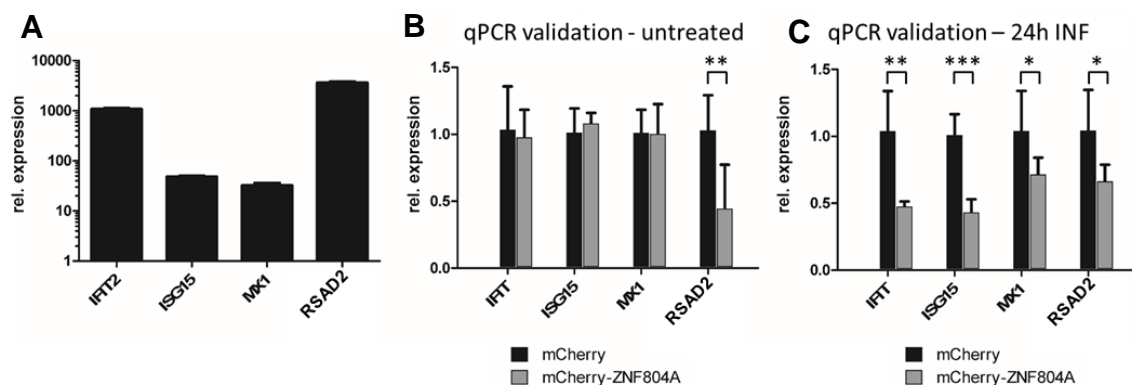
*Figure 18: **qPCR quantification of ZNF804A impacting STAT2 target gene expression.** (a) Relative increase of STAT2 target gene expression after 24 h of INF2α treatment. Bar diagram displayed with logarithmic scaled y ax and as means of triplicates of 3 biological replicates (error bars = SEM). (b-c) qPCR evaluation of STAT2 target gene expression of mCherryZNF804A transfected HEK293 cells normalized to mCherry transfected cells without (b) or after 24 h INF2alpha treatment (c). All bars are displays of triplicates of 3 biological replicates (error bars = SEM). (p-values: *≤0.05, **≤0.01, ***≤0.001)*

primary viral transcription (*MX1*).[172–174]

First of all, the induction by INF2α for every target gene selected was determined. Therefore, HEK293 cells were treated with or without INFα (50 U/ml) for 24 hours. Afterwards the total mRNA was isolated and quantified for selected target genes in treated and untreated cells. For all four selected target genes a dramatic

increase in expression were measured after INF2α treatment.

The expression of MX1 was increased 32-fold, while RASD2 was expressed 3.600 fold higher. (Figure 18 A). To evaluate the effect of ZNF804A on STAT2 target expression, again mCherry-ZNF804A and as a control mCherry transfected HEK293 cells were used. Selected STAT2 target gene expression was analyzed using qPCR.

The only significant (p=0.0068) reductive effect was measured for RSAD2 with a decrease below 50% for mCherry-ZNF804A transfected cells compared to mCherry transfection (Figure 18 B). In a second set of experiments the same parameters were used but cells were treated with INFα (50 U/ml) 24 h before lysis and qPCR. In contrast to the first qPCR evaluation, INFα treatment resulted in a significant reduction of all STAT2 target genes (*IFIT1* to 47%, p=0.001; ISG15 to 43%, p<0.0001; MX1 to 29%, p=0.034; RSAD2 to 71%, p=0.018) for mCherry-ZNF804A transfected cells, compared to mCherry transfection (Figure 18 C). Unlike the ISRE assays, in qPCRs no additional GFP-STAT2 was transfected. This reduces the possibility of ZNF804A effects on STAT2 functionality being only an artifact due to overproduction of proteins in analyzed cells.

4.3.6.3 CRISPR/Cas9: Knock down of ZNF804A confirms modulating effect on STAT2 pathway

To further eliminate the risk that measured effects with ISRE and qPCR for ZNF804A were arbitrary effects of overproducing tagged proteins, a HEK293 ZNF804A knock out cell line was established using the CRISPR/Cas9 technology. Therefore, a guide RNA (gRNA) that targeted the second exon of ZNF804A was designed. The guide was cloned into the pSPcas9(BB)-2A-Puro plasmid that also encodes the CRISPR and Cas9 protein, as well as a pyromicine resistance. HEK293 cells were transfected with the designed pSPcas9 ZNF804A gRNA vector. After two days of growth they were selected for pyromycin resistance for additional three days. To ensure that only transfected cells survive the selection period, untransfected cells were always maintained as additional samples alongside. Three days of selection were sufficient to ensure no untransfected cells survived. Surviving cells were seeded into 96 well format with 0.5 cells per well. Growing single cell colonies were later isolated and cultured as separate cell lines.

70
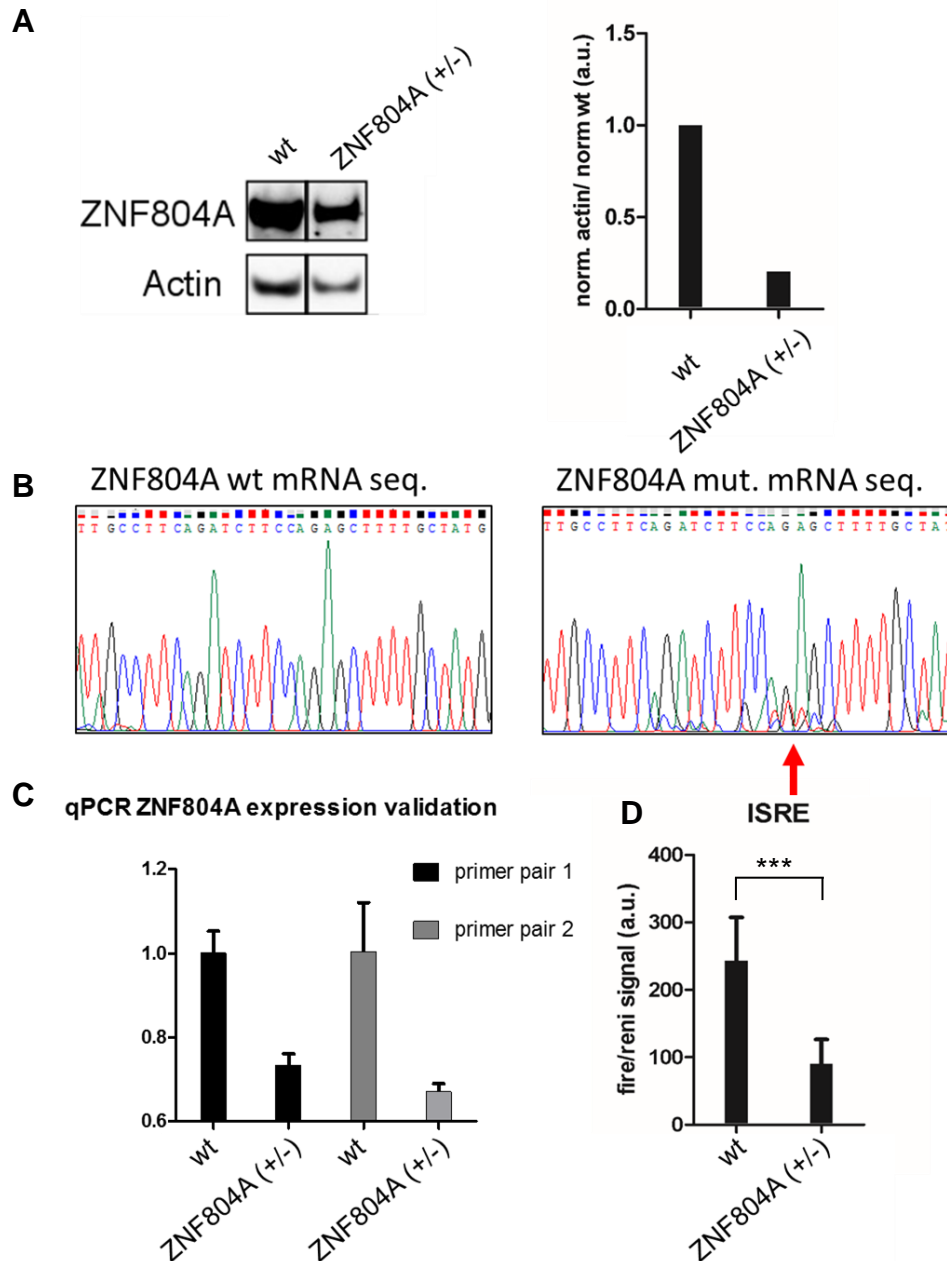
*Figure 19: **CRISPR/Cas9 knock down of ZNF804A.** (a) Western blot validation of endogenous ZNF804A production levels. Left lane wild type HEK293 cell lysate, right lane CRISPR/Cas9 treated HEK293 cells for ZNF804A knock out. Top lane ZNF804A specific antibody (SynapticSystems) was used to quantify ZNF804A production. Actin antibody was used as a loading control (down lane). Tot he left: Evaluation of relative ZNF804A production of western blot in (a) as bar diagram. (c) Chromatogram of CRISR/Cas9 ZNF804A gRNA target sequins for wild type HEK293 (left) and CRISPR/Cas9 treated HEK293 cell line. (d) qPCR evaluation of ZNF804A expression displayed as bar diagram using two pairs of qPCR primers Bars represent means of triplicates of 3 biological replicates (error bars = SEM). (e) ISRE reporter assay evaluation of 24 h INF2alpha treated wild type and CRISPR/Cas9 treated HEK293 cells. Bars represent normalized luminescence ratios of triplicates of 3 biological replicates (error bars = SEM). (***= p≤ 0.001)*

All cell lines were analyzed for reduced ZNF804A signal via western blot analysis using ZNF804A antibody and sequenced. For one cell line a reduced detectability of ZNF804A, compared to wild type cells was detectable (Figure 18 A). Normalization to the actin loading control indicated a reduction of nearly 80%. The fact that ZNF804A signal was still detectable, indicate that the cell line was heterozygote for ZNF804A knock out (-/+). The corresponding sequencing showed an unsolvable sequence in the area of the estimated Cas9 cleavage side, which is also a clear indication for overlapping sequence signals and therefore, for a heterogeneous cell line.

To further confirm this result, qPCRs were performed to evaluate the ZNF804A expression for the newly established ZNF804A (+/-) cell line. For this purpose, two primer pairs from the Harvard PrimerBank with target sequences within the second exon of ZNF804A were used. GABDH served as a control. Again, ZNF804A (+/-) cell line showed strong decrease in ZNF804A expression for both primer pairs additionally validating heterozygous knock out of ZNF804A in ZNF804A (+/-) cell line (Figure 18 B).

In summary, a heterozygous knock out cell line for ZNF804A (ZNF804A (+/-)) were successfully established and validated via western blot and sequencing.

To evaluate the impact of heterozygote knock out of ZNF804A on STAT2 mediated response, again ISRE assays were performed. Heterozygote ZNF804A (+/-) cell line as well as wt HEK293 cells were transfected with ISRE reporter constructs, according to manufacturer recommendations. The ZNF804A (+/-) cell line showed a significant decrease in ISRE response (p> 0.001) (Figure 18 C).

Taken together, functional analyses indicated a clear effect of ZNF804A on STAT2 mediated INFα response.

5. Discussion

SCZ is caused by multiple genes as well as environmental factors. Numerous genes were associated with SCZ in mainly the past decade through GWAS, CNV studies, exome sequencing approaches and also expression and methylation studies to a lessor extend.

Because none of the associated genomic variants show a high prevalence in patient cohorts, nor is their effect size high enough to pin point single genes as major drivers of the disease.[3] This results in an inability to immediately convert the identified genetic inside into plausible angles for the development of new and effective SCZ therapy strategies.

To circumvent this factor, scientists started to search for possibilities to relate patient genotype to its phenotype. Because of the low abundance of single genomic variations and a clear association of symptoms to genomic background is very difficult to achieve. Additionally, genes are often associated across psychiatric disorder boundaries.[175]

The concept to still achieve a correlation is called endophenotype. Endophenotypes are characteristics or behaviors that are intermediate between genotype and phenotype.[176,177] Those characteristics can be tics or also specific changes in tasks during brain imagine. In undiagnosed relatives, carrying the genomic variants, often those characteristics are also identifiable, underlining the closer relation between endophenotypes and genotypes, rather than genotypes and disease.[178,179]

Several endophenotypes were already identified. Base of those endophenotypes are for example psychophysiological inhibitory responses to internal and external stimuli, which are considered to be dependent on genetic rather than environmental factors.[180–182]

An inherent constriction for the identification of endophenotypes is that the effect size of genetic influence on endophenotypes is quite small compared to the effect size of symptoms of the disorder itself, making endophenotypes difficult to identify.[183]

A possible way to increase the resolution for the identification of endophenotypes might be to group potential with SCZ associated genes to functional modules. One way to achieve that goal was taken by the Consortium on the Genetics of Schizophrenia (COGS). The COGS analyzed endophenotypes in 300 affected families, as well as in 2471 individuals. Within their findings, they were able to group their identified genes into a network of 42 genes which describe 12 related endophanotypes.[184]

The COGS benefitted from the large sample size of analyzed patients, relatives and controls. Another strategy to identify potential groups of genes influencing potential endophanotypes is presented in this thesis and doesn't require large patient and control cohorts: The identification of disease modules via PPI network analysis. Identifying disease modules might increase the a priori expectation of SCZ related genes, relevant for a specific endophenotype. Additionally, previously not with SCZ associated proteins are included in those modules, which might lead to a better functional understanding of SCZ associated genes, as well as potential new targets for SCZ association.

In the present study, five SCZ related potential disease modules were identified with promising implications for SCZ. The two highest scoring clusters represent well known synaptic protein complexes. In the context of endophenotypes, analyzing those cluster proteins combined for endophenotypic implications would be a promising perspective. The potential insides that can be gathered with identified modules would further increase if SCZ related mutations and their influence on cluster connectivity would be analyzed. The most comprehensive example for an approach that used side directed mutation for PPI analysis was performed by Sahni *et al.*. They analyzed over a thousand described coding mutations for mendelian disorders and their connectivity within a PPI network. They were able to conclude that about two thirds of coding disease mutations indeed change the PPI binding pattern of the corresponding protein.[185] In this thesis a first effort in this direction as made. Although no significant changes were identified, the potential to identify changes increases when more and more mutations are identified. Overall, identifying potential disease modules might be key to functionally group SCZ related genes with a potential for the development of new and specific drug therapies for SCZ.

As mentioned, the amount of with SCZ associated genes is still rising and huge efforts are undertaken to further identify and validate SCZ associated genes. One fast growing field is exome sequencing. An inherent difficulty of exome sequencing studies is that an overwhelming amount of variations between each individual, even when patients and their relatives were analyzed. Degenhardt *et al.* analyzed 3 multiple with SCZ affected family trees and used intense filtering regarding the likelihood of each variation to be impactful for SCZ. Filtering narrowed down the number of candidate genes to 38 with the highest potential to be of high impact for SCZ. As an additional approach to further priorities identified candidate genes and provide a ranking, PPI network based approaches presented in this study. The strategy of using previously with SCZ associated genes proved to successfully confirming a significant enrichment in connectivity of the identified candidate genes to SCZ associated proteins. The connectivity of each candidate gene to SCZ related genes additionally served as a prioritization parameter for each candidate genes.

Overall, using PPI network analysis to prioritize identified candidate genes for genetic studies of SCZ are a promising approach.

ZNF804A is one of the most reliably with SCZ associated genes, mainly via largescale GWAS.[11,47] Despite its clear association with SCZ, the cellular function of ZNF804A is relatively unknown. Following the principe of guilt by association, the approach undertaken in this study to shed light on the cellular role of ZNF804A was to identify the direct PPI partners of ZNF804A in an unbiased proteome scaled Y2H screen. 18 previously unknown ZNF804A protein protein interactors were identified and identity verified via sequencing. The validation rate of Y2H is published to range around 10-34%.[48,65,66] The validation approach of this study used two independent mamallian cell-based validation methods and was able to recover 67% of Y2H identified interactors. The high validation rate underlining the validity of identified ZNF804A PPIs as well as indicate a correlation between strict cut of rates and the usage of extensive controls for Y2H with a high confidence for identified PPIs.

A very similar approach was undertaken by Zhou *et al.*.[186] They performed an unbiased Y2H screen using the full length ZNF804A and identified 22 new ZNF804A interacting proteins. [186] None of the identified ZNF804A interacting proteins overlapped with identified ZNF804A presented in this thesis. This phenomenon is common within Y2H screen approaches. The first two published human proteome scaled Y2H networks showed very little overlap.[57,58] This doesn't necessarily mean that one or both of this studies had a high amount of false positive identified PPIs. For example, used prey libraries differ between groups and even shared proteins between libraries can differ in expression levels, splice variants or can even include only protein fragments. Additionally, differences in methodology like cut of scores for considering PPIs as positive may differ between groups or settings. All those factors can lead to differences in the outcomes of PPI evaluations. Interestingly, ZNF804A PPI partners identified by Zhou *et al.* were significant enriched in being involved in protein translational processes where as in this thesis identified PPI partners most significantly were involved in RNA splicing. Also, both sets of identified ZNF804A partners did not showed higher amounts of intra-set PPIs, rather then inter-set PPIs (using the HIPPIE data base), making it more likely that they are indeed two different sets of ZNF804A interactors (Figure 20).
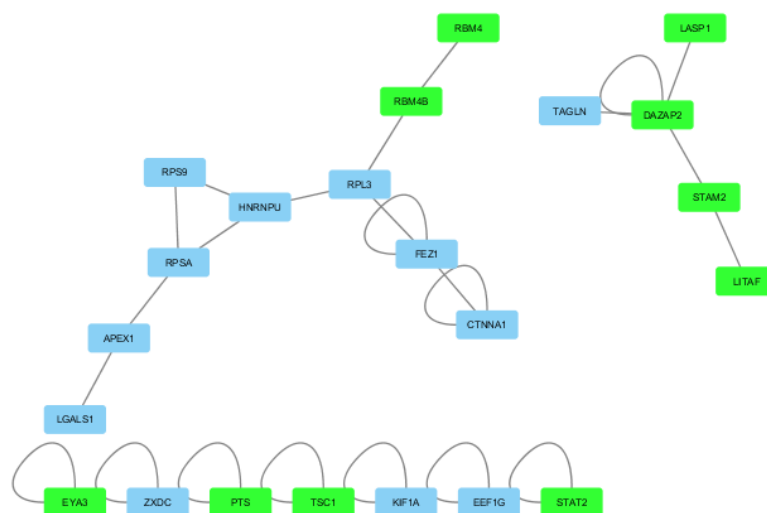
Figure 20: **Depiction of interactions between proteins identified to interact with ZNF804A of this thesis, as well as published by Zhou et al.**.[186] *The HIPPIE data based was used as a PPI resource and all interactions are of medium high confidence (HIPPIE score ≥ 0.63). Green: ZNF804A interactors identified in this thesis. Blue: ZNF804A interactors identified by Zhou et al..*

A possible explanation for this, as well as the differences in identified ZNF804A PPI partners between these two approaches might be the usage of different constructs. Zhou *et al.* identified their ZNF804A interaction partners using the full length ZNF804A protein, whereas in this thesis presented the Y2H results were based on smaller ZNF804A fragments. Although, in validation steps using LuTHy assays, identified ZNF804A PPIs were partially validated with full length ZNF804A constructs, the initial identification process based on ZNF804A fragments might shifted the exploration process of Y2H assays towards another subset of ZNF804A interactors.

This interpretation is strengthened by the fact, that more PPIs were identified with smaller c-terminal construct of ZBF804A, compared to larger c-terminal constructs. This may be the consequence of steric reasons, accessibility and conformation. Therefore, Zhou *et al.* might have missed c-terminal interactors of ZNF804A, where as in this thesis presented Y2H screening might have missed interactions that depend on the full ZNF804A protein. In functional analysis, Zhou *et al.* identified a primary localization of ZNF804A within the cytoplasm. Within their PPI analyze they, based on their PPI data and the cytosolic localization, consequently concluded that ZNF804A may play a crucial role in protein translation.[186] As in this thesis described, ZNF804A was also found to be primarily cytosolic localized under untreated conditions using immuno fluorescence microscopy. After INF treatment, ZNF804A showed a clear translocation into the nucleus. This implies, that ZNF804A might play two functional roles. The first one, as Zhou *et al.* identified, as a protein translation controlling protein. The second cellular function would be a transcriptionally controlling role, potentially in dependency on INF in conjunction with STAT2. These two potential roles of ZNF804A, identified via Y2H based PPI analysis might be two future angels to analyze the role of ZNF804A within the development of SCZ and consequently for the development of new SCZ drugs.

The focused analysis in this thesis of the binding of STAT2 to ZNF804A revealed that this interaction relies on ionic amino acids on the n-terminal surface of ZNF804A, indicating an interaction with an intrinsically disordered protein. This indication was supported by computational prediction of the domain structure of the n-terminus of ZNF804A. In the context of INF2α induced ZNF804A and STAT2 co-translocation into the nucleus, the interaction might be an induced interaction

and ZNF804A might mainly interact with dimerized STAT2 complexes. This possibility, as well how ZNF804A might interact with formed ISGF3 complexes is a very interesting investigation to further understand the modalities of this specific interaction and the precise role of ZNF804A in the STAT2 pathway.

It is naturally difficult to analyze the impact of an PPI functionally in a cellular context. The established strategies are mainly to over express or knock down one or both proteins. The inherent problems with both strategies are, that the over expression of a protein, as well as the knock down has a wide variety of effects, that might overshadow the impact of the specific PPI of interest.

IN the context of the results of this thesis, over expression, as well as knock down experiments of ZNF804A displayed a reduction in STAT2 mediated INFα response in ISRE assays. Expected would have been that both assay strategies resulted in opposite influences on STAT2 mediated INFα response. This is a prime example of the difficulties to evaluate the impact of a specific PPI by analyzing knock down and overexpression strategies. The context of the confocal microscopy experiments for overexpression of fluorescence tagged ZNF804A and STAT2 compared to immune fluorescence detection of endogenous proteins gave a plausible explanation for the unexpected results. Co-transfection of fluorescence tagged ZNF804A was observed to form perinuclear speckles together with tagged STAT2, whereas transfected fluorescence tagged single proteins didn't show those tendencies. Therefore, the most plausible explanation for the effect of ZNF804A on STAT2 mediated INFα response was that overexpression of tagged ZNF804A co-accumulates with STAT2 and prevents its proper translocation into the nucleus. Without the nuclear localization, STAT2 seemed not to be able to properly perform its role in the INF response pathway. The occurrence of those speckles seemed to be tied to co-overexpression of tagged constructs and therefore to most likely not represent a biological relevant effect. The knock down of ZNF804A on the other hand indicated that the function of STAT2 as an INF induced transcription factor is significantly diminished.

To more precisely confirm that the PPI between ZNF804A and STAT2 is important for the INFα induced STAT2 activity, in a follow up study, side directed mutation of the ZNF804A interaction sides to STAT2 using CRISPR/CAS9 would be very

inside full. The results of the performed peptide array gave the first guideline which amino acids are the most importance for the interaction with STAT2. Also, the indication that the polarity might be of high importance.

The identified role of ZNF804A in the cellular immune system might have significant impact for the development of SCZ. As described in the introduction, prenatal infections are significantly associated with the chance of developing SCZ.[32,33,37,39–41,187,188] Also, the most significant hit of the largest SCZ GWAS study(lit) up to day was localized in an MHC region (Major Histocompatibility Complex), a region coding for antigens, presented on the surface of cells, important for the acquired immune system to recognize foreign molecules. In general the immune hypothesis of SCZ is rising in interest for researchers.[32] Therefore, the potential role of ZNF804A in the cellular immune system might be of special interest.

6. Methods

6.1 Protein-protein interaction assays

6.1.1 Y2H assay

The Y2H interaction mating assays were performed as previously published.[57] Briefly, bait constructs were transformed into yeast strains L40ccua (MATa) and for prey constructs a library of ~17.000 pre-transformed yeast L40ccα (MATα) strains were used. For interaction mating, on L-HAUT q-trays grown MATα colonies were spotting robot based (Kbiosystems) mixed into 100 µl cultures of MATa yeast strains in 96-well microtiter plates. The yeast mixtures were afterwards spotted onto YPD agar plates using a spotting robot. After 48 h of mating at 30°C, yeast colonies were picked up via spotting robot and transferred into 100 ul selective liquid medium (SD2-Leu-Trp) in 96-well microtiter plates.

Finally, for determine positive PPIs, mated yeasts were spotted onto SD4 (-Leu-Trp-Ura-His), as well as SDII (-Leu-Trp) selective agar plates. After incubation for 6 days at 30°C, agar plates were imaged, and yeast colony growth assessed by visual inspection. Media are listed below (Table 9).

Table 9: Y2H media

YPD liquid medium	Bacto peptone Bacto yeast-extract Glucose	20 g/l 10 g/l 20 g/l
SD liquid medium (amino acids were added according to requirement of yeast dependency and read out)	Yeast Nitrogen Base Glucose Adenine 100x Histidine 100x Leucine 100x Tryptophan 100x Uracil 100x	6.7 g/l 20 g/l 10 ml 10 ml 10 ml 10 ml 10 ml

6.1.2 DULIP assay

HEK293 cells were reversely transfected in 96 well microtiter plates at a density of 3.5×10^4 cells per well. 48 h after transfection cells were lysed in 100 µl HEPES lysis buffer (Table 10) for 30 minutes at 4°C. The amount of produced PA-RL- and FL-tagged fusion proteins was monitored by measuring the respective luciferase activities in crude cell lysates in 384-well microtiter plates. For that, to 10 µl of cell lysate, 20 µl PBS and 10 µl of Dual-Glo® luciferase reagent (Promega) were added. After 10 minutes the firefly activity was measured using an Infinite® M1000 (Tecan) plate reader. In order to stop the firefly luciferase activity and measure the *Renilla* luciferase activity , 10 µl of the Dual-Glo® Stop & Glow® reagent (Promega) were added, incubated for 15 minutes and the activity was measured. Additionally, 50 µl of the cell lysate were incubated for 3 hours at 4°C in IgG pre-coated 384-well microtiter plates. Plates were pre-coated with sheep gamma globulin (Dianoca), blocked with 1% BSA in carbonate buffer (70 mM $NaHCO_3$, 30 mM Na_2CO_3, pH 9.6) for 1 h and finally incubated with rabbit anti-sheep IgGs (Dianova) overnight. After cell lysate incubation in antibody coated plates, all wells were washed three times with HEPES lysis buffer and 30 µl of PBS were added to each well.

For luminescence measurements, Dual-Glo® kit was used following manufacturer instructions.

Measurement of firefly and *Renilla* luminescence activity was performed using an Infinite® M1000 (Tecan) plate reader. DULIP data analysis were described elsewhere.[65]

6.1.3 LuTHy assay

HEK293 cells were reversely transfected in white 96 well microtiter plates at a density of 3.5×10^4 cells per well. For low expressing ZNF804A donor constructs, DNA ratios of donor and acceptor were used at a 1:3 ratio, with 50 ng donor and 150 ng of acceptor. In other cases, a ratio of 1:10 was chosen, with 10 ng donor and 100 ng of acceptor. 48 h after transfection, 96 well plates were inserted into a Tecan Infinite® M1000Pro microtiter plate reader and mCitrine fluorescence was measured in intact cells at Ex/Em: 500 nm/530 nm. Afterwards, 5 µM coelenterazine-h (NanoLight, 301) was added and cells incubated for 10 min. Again, using a Tecan Infinite® M1000Pro microtiter plate reader and the NanoLuc

emission was measured with the BLUE1 filter (370-480 nm), the emission of mCitrine due to BRET using the GREEN1 filter (520-570 nm), as well as the total short-WL and long-WL luminescence without using a filter. The chosen integration time was 100-1000 ms.

After luminescence measurements in intact cells, the cells were lysed in 80 µl HEPES-phospo lysis buffer (*Table* 8) for 30 minutes at 4°C. Production of fusion proteins was monitored by measuring the fluorescence activity and luciferase activity in cell lysates. To 10 µl of the cell lysate, coelenterazine-h was added to a final concentration of 5 µM and the luminescence activity measured as before in a microplate reader. In parallel, 15 µl of the cell lysates were incubated for 3 hours at 4°C in IgG pre-coated small-volume 384-well microtiter plates. Plates were coated with sheep gamma globulin in carbonate buffer (70 mM NaHCO3, 30 mM Na2CO3, pH 9.6) for 3 h at room-temperature, blocked with 1% BSA in carbonate buffer before they were incubated with rabbit anti-sheep IgGs in carbonate buffer overnight at 4°C. Cell-lysates were incubated for 3 h at 4°C, after which, all wells were washed three times with HEPES-phospho lysis buffer and the mCitrine fluorescence activity measured. Afterwards, 15 µl of PBS containing 5 µM coelenterazine-h were added to each well and the NanoLuc luminescence activity measured as described above. Data analysis were described elsewhere.[66]

Table 10: HEPES lysis buffers

Ingredients	Concentration
HEPES	50 mM
NaCl	150 mM
Glycerol	10%
NP-40	1%
Deoxycholate	0.5%
NaF	20 mM
MgCl₂	1.5 mM
EDTA	1 mM
DTT (not included for WB)	1 mM
Benzonase (Roche)	1 U
PMSF (not included for DULIP)	1 mM
Glycerol-2-phosphate (not included for DULIP)	25 mM
Sodium orthovanadate (not included for DULIP)	1 mM

6.2 Cell based assays

6.2.1 Cell culture and transfection

The human embryonic kidney cell line 293 (HEK293), as well as SHEP cells were grown in high-glucose (4.5 g/L) supplemented with 10% heat inactivated fetal bovine serum at 37°C and 5% CO_2. Cells were subcultured every 2-4 days. For FRET and DULIP experiments, the HEK cells were transfected with linear polyethyleneimine (25 kDa, Polysciences) using the reverse transfection method according to the manufacturer's instructions. For BRIP, ISRE, western blot and qPCR assay transfections, HEK cells forward transfected using Lipofectamin™ (Thermo Fisher) following manufacturer instructions. For microscopic imaging, SHEP cells were transfected with GeneJet according to manufacturer's instructions on coated coverslips. Transfections were performed with a total DNA amount of 200 ng per well in a 96-well plate, 600 ng per well in 24 well-plates or 2 µg per well in 6 well. Plasmids with high expression value were sometimes transfected in lower concentrations. In that case pcDNA3.1(+) was used as carrier DNA to fill the total DNA amount to comparable levels. For luminescence measurements cells were examined 48 h or 72 h after transfection.

6.2.2 Confocal microscopy

For confocal microscopy, SH-EP cells were cultured in 24-well cell culture plates ($7x10^4$ cells/well) on fibronectin coated coverslips.

Imaging of transfected proteins

After 24 h, cells were transfected with 300 ng DNA using GenJet as manufacture described. SH-EP cells were fixed with 2 % PFA for 15 mins at RT. In order to stain nuclei, PFA included a final concentration of 1:5000 Hoechst 33342 (Sigma), which stains DNA and is visualizable at a wavelength of Ex/Em 353/483 nm. Fixed cover slips were transferred top down to conventional microscope slides using Darko fluorescence mounting medium (Dako) before image acquisition with a TSC SP8 (Leika) confocal microscope. GFP and mCherry fluorescent images were acquired at excitation wavelengths of 488 nm and 380 nm, respectively.

Imaging of endogenous proteins

Untransfected SH-EP cells were fixed with 2 % PFA for 15 mins at RT. In order to stain nuclei, PFA included a final concentration of 1:5000 Hoechst 33342 (Sigma). Cells were washed two times with PBS-T and permeabilized for 10 minutes with 0.1% Triton, 1% BSA solution. Afterwards samples were blocked for 30 minutes (1% BSA -PBS-T). Primary antibodies were applied in 1% BSA-PBS-T solution for 1 h. After three washing steps with PBS-T, secondary antibodies were applied in 1% BSA-PBS-T for 45 minutes. After three additional washing steps with PBS-T, samples were mounted using Dako fluorescence mounting medium (Dako).

For confocal microscopy TSC SP8 (Leika) confocal microscope was used. Co-localization analysis was performed in all cases using ImageJ software.

6.2.3 ISRE assay

For ISRE assays, cells were reverse transfected in 96 well plates, using PEI, described in section cell transfection, with either 100 ng of reporter constructs or 100ng of reporter controls (both luciferases auto active) together with 200ng of target constructs or empty vector or 100 ng of each target construct in case of two target gene transfections. On every plate also positive control with only positive control vectors were transfected. After 48 h cells were treated with INF2α if mentioned and after 72 h after transfection, cells were lysed using 20 µl HEPES lysis buffer (Table 10) for 30 minutes at 4°C. In order to measure firefly signals, lysates were incubated with 100 µl of Dual-Glo® luciferase reagent (Promega) were added. After 10 minutes the firefly activity was measured using an Infinite® M1000 (Tecan) plate reader. In order to stop the firefly luciferase activity and measure the *Renilla* luciferase activity, 100 µl of the Dual-Glo® Stop & Glow® reagent (Promega) were added, incubated for 15 minutes and the activity was measured. To access relative firefly expression, ratios of firefly to *Renilla* were calculated for every well. Negative control conditions were analyzed to exclude general effects on expression values due to arbitrary expression. Finally, means of technical replicates were calculated with standard deviations and normalized to an according control condition.

6.2.4 Yeast transformation (96 well format)

Cell pellets of 30 ml yeast cultures in YPD media (OD_{600}=0.6) were resuspended in 10 ml 1xTE buffer (pH=8). After additional centrifugation, 1.1 ml LiAC solution 1 was added (Table 11) to the pellet. To 500 ng vector DNA and 2 µl carrier DNA, 12 µl of yeast in LiAC solution 1 were added, as well as 50 µl LiAC solution 2 (Table 11). After 30 minutes of incubation at 30 °C, 8 µl DMSO were added and heat shocked for 7 minutes at 42 °C.

Table 11: Y2H transformation

LiAC solution 1	LiAC solution 2
1 ml LiAC (1M)	1.5 ml LiAC (1M)
0.5 ml 10xTE buffer	1.5 ml 10xTE buffer
5 ml sorbitol	10 ml PEG (60%)
3.5 ml H_2O	2 ml H_2O

6.3 Molecular Methods

6.3.1 Plasmid preparation

Plasmid DNA was isolated according to the manufacturer instructions (QIAprep Spin miniprep, Quiagen). Briefly, the bacterial samples were centrifuged for 7 minutes at 4°C and 4000 rpm. The supernatant was discarded, and the pelleted bacteria were resuspended in 250 µl Buffer P1 and transferred to a new microcentrifuge tube. 250 µl Buffer P2 was added to lyse the cells. After inverting the tubes 5 times 350 µl Buffer N3 was added and the tubes were inverted again thoroughly. After the suspension became cloudy, the samples were centrifuged for 10 minutes at 13000 rpm in a standard table-top microcentrifuge. The supernatant was then applied to the QIAprep spin columns and again centrifuged for 1 minute at 13000 rpm. For washing the QIAprep spin column 750 µl PE buffer were added and removed by centrifugation.

Finally, the DNA was eluted by adding 40 µl dH20 (incubating for one minute) and centrifuged into a standard 1,5 ml microcentrifuge tube.

6.3.2 Coating of coverslips

Acid washed coverslips were incubated with 10µg/ml fibronectin and (if un-transfected cells were spotted on coverslips) 10µg/ml polylysine for 4 h at 37°C. Afterwards, coverslips were washed twice (once with water and once with PBS) and cells were able to be directly seeded.

6.3.3 Filter assays

Dot blot assays (DB) is used for immunodetection of proteins without electrophoretic size separation. Samples are directly spotted on a nitrocellulose membrane with a pore size of 0.1 µm with a 96-well vacuum apparatus. Dot blot assays are native, therefore samples are prepared in PBS. To assemble the vacuum apparatus, wartman papers and the membrane were equilibrated in PBS, the membrane placed on top of the Whatman papers and both together in between the apparatus. After washing the membrane with 100 µl PBS, protein samples (25 ug, diluted in PBS) were applied and finally, the membrane was washed twice with 100 µl PBS. Membrane blocking and immunological protein detection was done as described for western blots.

To only retain insoluble protein aggregates on non-binding membranes, filter retardation assays (FRA) were performed. FRA uses, unlike dot blot assays where the total protein amount is spotted on a membrane, filter retardation 0.2 µm cellulose acetate membrane to enable size-dependent retardation. The assay was performed similar to dot blot assays, however, the membrane was equilibrated in 0.1 % SDS, samples were prepared in equal volumes of denaturing buffer (4 % SDS and 100 mM DTT) and all washing steps were carried out with 100 µl 0.1 % SDS. Additionally, samples were boiled for 5 minutes at 95°C prior to application. Membrane blocking and immunological protein detection was done as described for western blots.

Dot intensities of DB and FRA membraned were quantified using AIDA image analyzer software.

6.3.4 Polymerase Chain Reaction

Phusion (Thermo) polymerase chain reaction (PCR) was used to amplify various defined double stranded DNA fragments from template DNA using pairs of short oligonucleotide primers, flanking the sequence of interest. Target specific primers were designed and ordered from BIOTEZ. For cloning or detection purposes it was sometimes necessary to implement restriction sites or small nucleotide sequences, coding for small peptide tags to the 5´and 3´ends of the potential PCR fragments, respectively. Therefore, overhang primers were designed and used. PCR ingredients and PCR–program is shown in table (Table 12 and Table 13).

Table 12: PCR ingredients

Ingredients	Concentration
DNA-Template	100 ng
dNTP´s	200 µM
Primer forward	0.5 µM
Primer reverse	0.5 µM
Phusion polymerase	0.5 U
5x GC-Buffer	1x
ddH$_2$O	To 50µl

Table 13: PCR-program

	Temperature	Time
Denaturation	98 °C	1 min
Denaturation	98 °C	10 sec
Annealing	58-70 °C	20 sec
Elongation	72 °C	35 sec
Elongation-final	72 °C	4 min
Store	4 °C	forever

‹---x34---›

6.3.5 Gateway™ cloning reactions

Gateway™ LR cloning reactions were performed to transfer inserts from entry clones into expression vectors via site-specific recombination. Each reaction was composed of 1 µl destination vector (75 ng/µl), 1 µl entry vector (25 ng/µl), 0.5 µl Gateway™ LR Clonase II Enzyme mix (Invitrogen) and 2.5 µl TE buffer (pH 8.0). To allow site specific recombination, the reaction was incubated at 25°C for 2 h.

6.3.6 Sequencing

cDNA sequencing of cloned constructs and PCR fragments was performed by Source BioScience, Stratec and LGC-Genomics.

6.3.7 Gel electrophoresis of DNA fragments

For separation and size determination of DNA fragments, samples were loaded on a gel with 0.8 % (w/v) agarose in TBE buffer supplemented with 0.5 µg/ml ethidium bromide together with a DNA ladder that has known band sizes and concentrations.

Electrophoresis was performed at constant current (220 mA) in TBE buffer. Gels were analyzed using a UV lamp and size was estimated by comparison with the standard.

6.3.8 Determination of DNA concentration

To determine the concentration of nucleic acids, the absorbance of isolated DNA and RNA was measured at 260 and 280 nm with a NanoDrop spectrophotometer. The measurement at A260 reads the concentration of DNA/RNA. A value of 1.0 is equal to ~50 µg/ml of double stranded DNA. The A260/A280 ratio determines the purity of DNA/RNA with an optimal ratio of A260/280 of 1.8-2.1.

6.3.9 Expression and purification of HIS fusion protein

BL21 DE3 Rosetta2 BL21-RP *E. coli* strain were transformed with His-tagged STAT2 or mCherry and was inoculated overnight in 50 ml LB-amp medium at 37°C shaking. With each overnight culture, two l LB-amp were inoculated 1:100 on the next day. Bacteria culture were grown again at 37°C shaking until an optical density of 0.6 (OD_{600} nm) and the protein expression were induced at 18°C with 1mM IPTG. After three hours, cells were centrifuged down (4000 rpm, 20 minutes), the

supernatant discarded and the resulting pellet shock frozen at -80°C. Frozen pellets from 500 ml culture were suspended in 10 ml His lysis buffer and incubated for 30 minutes at 4 °C. Lysates were sonicated on ice eight times for 10 sec with intermediate breaks for one minute between sonication steps. Afterwards, 1% Triton was added, and lysates were incubation for five minutes ice, cell debris was centrifuged down (15000 rpm, 30 min, 4°C) and pellets discarded. Supernatant was incubated for 2 h at 4 °C with Ni-nutrotriacetate (Ni-NTA) agarose matrix (200 ul/ 50 ml initial culture). After two washing steps with His wash buffer, His elution buffer was added and again incubated at 4 °C for 30 minutes (100 ul washing buffer/200 ul Ni-NTA slurry). After centrifugation, the solution was stored at -80 °C. Buffers are listed below (Table 14).

Table 14: His tagged protein isolation buffers

His lysis buffer	Tris-HCl pH 8.0	100 mM
	NaCl	200 mM
	Imidazole-HCl pH 8.0	25 mM
	Complete Protease Inhibitor (fresh)	tablet / 50ml dilution
	Benzonase (fresh)	10000^{-1} v/v
His wash buffer	Tris-HCl pH 8.0	100 mM
	NaCl	200 mM
	Imidazole-HCl pH 8.0	50 mM
His elution buffer	Tris-HCl pH 8.0	100 mM
	NaCl	200 mM
	Imidazole-HCl pH 8.0	500 mM

6.3.10 Measurement of protein concentrations

Protein concentration was determined using the BCA method, which is based on a color change in proportion to protein concentration. It is to mention that the color change heavily depends on the concentration of trypthophan. For the analysis of protein lysates 10 µl sample (1:10 dilution) was incubated with 22 µl BCA solution (Pierce) at 37°C for 30 minutes in 384-well plate. Afterwards, protein concentration was determined by measuring the absorption at 562 nm spectrophotometrically using a Tecan plate reader. As reference a standard curve of defined BSA concentrations in the appropriate buffer was used on every plate.

6.3.11 Denaturizing SDS-PAGE

Polyacrylamide gel electrophoresis (PAGE) is used for the separation of macromolecules according to their electrophoretic mobility, in this case proteins. The mobility depends mainly on the molecular weight, but also conformation and electric charge of analyzed molecules. The use of the detergent SDS leads to denaturation of secondary structures of proteins and covers the protein inherent charge. The addition of DTT causes breakage of disulfide bonds, resulting together with SDS, in full denaturation of proteins and leads to a separation, depending almost exclusively regarding their molecular weight in SDS-PAGE.

Protein samples were diluted to 25 ug total amount and prepared with 1x SDS-PAGE LDS Sample Buffer, as well as with 50 mM DTT detergent and boiled at 95°C for 5 min. Afterwards, samples were loaded on Invitrogen SDS-PAGE Bis-Tris 4-12% gels, together with 10 µl SeeBlue® Plus2 (Thermo-Fisher) as a size marker. Electrophoresis was performed at 185 V for 35 minutes with 1x MES SDS running Buffer.

6.3.12 Western Blot

Protein samples, separated by SDS-PAGE were transferred onto a nitrocellulose membrane using wet blotting system from BioRad® with 1x Transfer buffer at 100 V for 120 min. The membrane was blocked in 3 % milk PBS-T for 1 h at room temperature followed by incubation with primary

antibody in 3 % milk PBS-T overnight at 4°C. Membranes were washed 3x with PBS-T and incubated with secondary antibodies coupled to horse- radish-

peroxidase, diluted 1:2000 for 1 h at room temperature. Specificity of secondary antibodies depended on species in which primary antibodies were raised and were chosen accordingly. Membranes were again washed, 2x with PBS-T and 1x with PBS for each 5 min. Antibody staining was visualized by chemiluminescence of secondary antibody HRP signal after addition of WesternBright Quantum with an image reader (LAS 3000, Fujifilm).

6.3.13 pSpCas9 golden gate assemply

Oligos, encoding for sgRNAS were diluted to 100 µM final concentration within ligation reaction (1 µl 10x T4 ligation buffer, 1 µl T4 polynucleotide kinase, 10 µl total)

Ligation reaction were incubated at 37 °C for 30, followed by incubation at 95 °C for 5 minutes. Afterwards, reactions were ramped down to 25 °C at 5 °C steps each minute. 1:200 diluted sgRNA oligos were set up for pSpCas9 (100 ng) vector insertion as described in the table below (Table 15). Reactions were incubated for 6 cycles of first step 37 °C for 5 minutes, followed by 21 °C fo 5 minutes.

Table 15: pSpCas9 ligation

Ingredients	Amount
pSpCas9 (100ng/µl)	1 µl
Diluted oligos	2 µl
Tango buffer, 10x	2 µl
DTT, 10 mM	1 µl
ATP, 10 mM	1 µl
FastDigest BdsI	1 µl
ddH$_2$O	11.5

In the last step, ligation reactions were cleaned of any residual linearized DNA by digestion with PlasmidSafe as listed in the table below (Table 16). Reactions were incubated at 37 °C for 30 minutes, followed by 79 °C for 30 minutes.

Table 16: PlasmidSafe exonuclease digestion

Ingredients	Amount
Ligation reaction	11 µl
PlasmidSafe buffer, 10x	1,5 µl
ATP, 10 mM	1,5 µl
PlasmidSafe exonuclease	1 µl

6.3.14 CRISPR/Cas9

HEK293 cells were maintained according to the manufacturer's recommendations. Cells are cultured in D10 medium supplemented with 10% (vol/vol) FBS at 37 °C and 5% CO_2. For transfection, dissociated HEK293 cells were plates onto 24-well plates in D10 medium without antibiotics 24 h before transfection. Cells were seeded at a density of 1.3×10^5 cells per well in a total volume of 500 µl. Cells were transfected with Lipofectamine 2000 according to the manufacturers' instructions using 500 ng of sgRNA containing CRISPR/Cas9 Plasmids (pSPcas9(BB)-2A-Puro). After 24h of incubation, the cell culture medium was supplemented with 500 µl of warm D10 medium together with 2 µg ml^{-1} puromycine. After 72 h puromycine selection, cells were seeded into 96 well plates with a density of 0.5 cells/well for single cell derived cell lines. Successful single cell line creation was checked via fluorescence microscope during incubation in 96 well plates and sequencing after cell lines were established.

6.3.15 RNA isolation

RNA isolation was performed using the RNeasy® Mini kit (Qiagen) following the manufacturer instructions. Cells were lysed by adding 350 µl Buffer RTL (10 µl ß-mercaptorthanol was added to 1 ml Buffer RTL). One volume of 70% ethanol was added and mixed. 700 µl of sample were transferred to RNeasy Mini kit columns and centrifuged (8000 x g, 15 seconds). After DNase 1 digestion (adding 10 µl DNase 1 and 70 µl Buffer RD for 15 minutes), columns were washed 3 times (1x 700 µl Buffer RW1 and 2x 500 µl Buffer RPE). RNA was eluted with RNase-free water (40 µl).

6.3.16 cDNA synthesis

The cDNA synthesis was performed according to manufacturer instructions using the High Capacity RNA-to-cDNA Kit from Applied Biosystems as followed. For the reactions in a total volume of 10 µl, 5 µl 2x reaction buffer and 0.5 µl 20x enzyme mix, as well as 1 µg of isolated RNA were mixed and filled up with nuclease free water. For the cDNA synthesis reaction, a thermocycler was used and heated to 37°C for 60 minutes. The reaction was stopped by heating to 95°C for 5 minutes.

6.3.17 qPCR

qPCR was performed in ABI Prism 384-well clear optical reaction plates (4309849 Thermo Fischer) with a total reaction volume of 12 ul (primers 0.2 µM each, cDNA 200 µg and 6 µl SYBR Green (Qiagen). GABDH served as a control on every plate. Sealed plates were measured using Applied Biosystems ViiA7.

6.3.18 Peptide array

Ordered peptide arrays (Intavis) were blocked with 3% milk PBS-T for 1 h at room temperature. Blocked peptide arrays were incubated with either 20 µl HIS-STAT2, HIS-mCherry or no additional protein, each in 3% milk PBS-T for 2h over night at 4°C. Afterwards, arrays were washed 3x with PBS-T for 10 min. For detection purposes, arrays were incubated with either STAT2 (1:1000) or mCherry antibody (1:500) in 3% milk PBS-T for 2 h at room temperature. After washing 3x with PBS-T for 10 min, 3% milk with HRP fused secondary antibodies (1:5000)(Table 22), were incubated on arrays for 1 h. After final washing with 2x PBS-T and 1x PBS, WesternBright Quantum was added and HRP tags were detected using an image reader (LAS 3000 Fujifilm).

Table 17: Summary of peptides spotted on peptide array

protein of origin	pos. in protein	spot number	oligo sequence
ZNF804A	1-15	1	M-E-C-Y-Y-I-V-I-S-S-T-H-L-S-N
ZNF804A	10-25	2	T-H-L-S-N-G-H-F-R-N-I-K-G-V-F
ZNF804A	20-35	3	I-K-G-V-F-R-G-P-L-S-K-N-G-N-K
ZNF804A	30-45	4	K-N-G-N-K-T-L-D-Y-A-E-K-E-N-T
ZNF804A	40-55	5	E-K-E-N-T-I-A-K-A-L-E-D-L-K-A
ZNF804A	50-65	6	E-D-L-K-A-N-F-Y-C-E-L-C-D-K-Q
ZNF804A	60-75	7	L-C-D-K-Q-Y-Y-K-H-Q-E-F-D-N-H
ZNF804A	70-85	8	E-F-D-N-H-I-N-S-Y-D-H-A-H-K-Q
ZNF804A	80-95	9	H-A-H-K-Q-R-L-K-E-L-K-Q-R-E-F
ZNF804A	90-105	10	K-Q-R-E-F-A-R-N-V-A-S-K-S-R-K
ZNF804A	100-115	11	S-K-S-R-K-D-E-R-K-Q-E-K-A-L-Q
ZNF804A	110-125	12	E-K-A-L-Q-R-L-H-K-L-A-E-L-R-K
ZNF804A	120-135	13	A-E-L-R-K-E-T-V-C-A-P-G-S-G-P
ZNF804A	130-145	14	P-G-S-G-P-M-F-K-S-T-T-V-T-V-R
ZNF804A	140-155	15	T-V-T-V-R-E-N-C-N-E-I-S-Q-R-V
ZNF804A	150-165	16	I-S-Q-R-V-V—V-D-S-V-N-N-Q-Q-D
ZNF804A	160-175	17	N-N-Q-Q-D-F-K-Y-T-L-I-H-S-E-E
ZNF804A	170-185	18	I-H-S-E-E-N-T-K-D-A-T-T-V-A-E
ZNF804A	180-195	19	T-T-V-A-E-D-P-E-S-A—N-N-Y-T-A
ZNF804A	190-205	20	N-N-Y-T-A-K-N-N-Q-V-G-D-Q-A—Q
ZNF804A	200-215	21	G-D-Q-A—Q-G-I-H-R-H-K-I-G-F-S
ZNF804A	210-225	22	K-I-G-F-S-F-A-F-P-K-K-A-S-V-K
ZNF804A	220-235	23	K-A-S-V-K-L-E-S-S-A-A-A-F-S-E
ZNF804A	230-245	24	A-A-F-S-E-Y-S-D-D-A-S-V—G-K-G
ZNF804A	240-255	25	S-V—G-K-G-F-S-R-K-S-R-F-V-P—S
ZNF804A	250-265	26	R-F-V-P—S-A-C-H-L-Q-Q-S-S-P—T
ZNF804A	260-275	27	Q-S-S-P—T-D-V-L—L-S-S-E-E-K-T
ZNF804A	270-285	28	S-E-E-K-T-N-S-F-H-P-P-E-A-M-C
ZNF804A	280-295	29	P-E-A-M-C-R-D-K-E-T-V-Q-T-Q-E
ZNF804A	290-305	30	V-Q-T-Q-E-I-K-E-V-S-S-E-K-D-A
ZNF804A	300-315	31	S-E-K-D-A-L-L-L-P-S-F-C-K-F-Q
ZNF804A	310-325	32	F-C-K-F-Q-L-Q-L—S-S-D-A-D-N-C
ZNF804A	320-335	33	D-A-D-N-C-Q-N-S-V-P-L-A-D-Q-I
ZNF804A	330-345	34	L-A-D-Q-I-P-L-E-S-V-V-I-N-E-D
ZNF804A	340-355	35	V-I-N-E-D-I-P-V-S-G-N-S-F-E-L
ZNF804A	350-365	36	N-S-F-E-L-L-G-N-K-S-T-V-L-D-M
ZNF804A	360-375	37	T-V-L-D-M-S-N-D-C-I-S-V-Q-A-T
ZNF804A	370-385	38	S-V-Q-A-T-T-E-E-N-V-K-H-N-E-A
ZNF804A	380-395	39	K-H-N-E-A-S-T-T-E-V-E-N-K-N-G
ZNF804A	390-400	40	E-N-K-N-G-P-E-T-L-A
IRF9	205-220	49	R-S-L-E-F-L-L-P-P-E-P-D-Y-S-L

IRF9	215-2300	50	P-D-Y-S-L-L-L-T-F-I-Y-N-G-R-V
IRF9	225-240	51	Y-N-G-R-V-V-G-E-A-Q-V-Q-S-L-D
IRF9	235-250	52	V-Q-S-L-D-C-R-L-V-A-E-P-S-G-S
IRF9	245-260	53	E-P-S-G-S-E-S-S-M-E-Q-V-L-F-P
IRF9	255-270	54	Q-V-L-F-P-K-P-G-P-L-E-P-T-Q-R
IRF9	265-280	55	E-P-T-Q-R-L-L-S-Q-L-E-R-G-I-L
IRF9	275-290	56	E-R-G-I-L-V-A-S-N-P-R-G-L-F-V
IRF9	285-300	57	R-G-L-F-V-Q-R-L-C-P-I-P-I-S-W
IRF9	295-310	58	I-P-I-S-W-N-A-P-Q-A-P-P-G-P-G
IRF9	305-320	59	P-P-G-P-G-P-H-L-L-P-S-N-E-C-V
IRF9	315-330	60	S-N-E-C-V-E-L-F-R-T-A-Y-F-C-R
IRF9	325-340	61	A-Y-F-C-R-D-L-V-R-Y-F-Q-G-L-G
IRF9	335-350	62	F-Q-G-L-G-P-P-P-K-F-Q-V-T-L-N
IRF9	345-360	63	Q-V-T-L-N-F-W-E-E-S-H-G-S-S-H
IRF9	355-370	64	H-G-S-S-H-T-P-Q-N-L-I-T-V-K-M
IRF9	365-380	65	I-T-V-K-M-E-Q-A-F-A-R-Y-L-L-E
IRF9	375-390	66	R-Y-L-L-E-Q-T-P-E-Q-Q-A-A-I-L
IRF9	385-393	67	Q-A-A-I-L-S-L-V

aa (amino acids)

6.4 Bioinformatic analysis

6.4.1 Construction of SCZ-relevant network for cluster analysis

In order to identify complexes that are related to SCZ, a focused network had to be created, using the HIPPIE database (http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/) as a basis, that includes PPI network regions surrounding previously described SCZ proteins (Table 18). To find such regions, network propagation algorithm[74] were applied, that starts at the known SCZ proteins as priors and ranks all other network proteins by computing their propagation scores. Discrete formula described elswere.[189] Before selecting the most promising proteins from the list of ranked proteins by their score, the propagation algorithm was run over 1000 random sets of the same size as the SCZ list. Each protein prior score was ranked with respect to its scores on the random data (excluding the random instances where that gene was selected for the prior). This provided a P-value for every protein and allowed the selection of only significant ones (a threshold of 0.01 was used).[189] The detection process for protein complexes was described elsewhere.[189]

Table 18: SCZ associated and protein coding genes for cluster identification

AATF	DCN	ITIH4	MYH11	SATB2	VPS45
ABCC1	DDX52	KCNQ5	MYO19	SCARF2	VRK2
ABCC6	DFNA5	KCNV1	NAB2	SCG5	WBSCR22
ABHD11	DGCR14	KCTD13	NBEA	SDCCAG8	WBSCR27
ACACA	DGCR2	KDM3B	NCBP2	SEC22B	WDR53
ACP6	DGCR6	KDM4A	NDE1	SENP5	YPEL3
AIFM3	DGCR6L	KIAA0430	NEGR1	SEPT5	ZDHHC19
AKT3	DGCR8	KIAA1244	NFATC3	SERPIND1	ZDHHC8
ALDOA	DGKI	KIF18A	NIPAL3	SEZ6L2	ZNF74
ANKRD35	DGKZ	KIF22	NLGN4X	SF3B1	ZNF804A
ARL3	DLG1	KLF13	NMUR2	SHMT2	ZNHIT3
ARVCF	DLG2	KLHL22	NOTCH2NL	SLC25A1	UBE3A
ASCL1	DNAJC30	LAMA4	NRGN	SLC38A7	UBXN7
ATP2A2	DOC2A	LAT2	NRXN1	SLC7A4	UFD1L
BAIAP2	DPP4	LHX1	NT5C2	SNAP29	VIPR2
BANK1	DPYD	LIMK1	NXPH4	SNAP91	VPS37D
BAZ1B	DRD2	LPHN2	OCA2	SND1	VPS45
BCL11B	DUSP14	LZTR1	OTUD7B	SNRPN	VRK2
BCL7B	EFNB1	MAD1L1	P2RX6	SNURF	WBSCR22
BCL9	EIF4H	MANEA	PAK2	SPN	WBSCR27
BDH1	ELN	MAPK3	PAK6	SRPK2	WDR53
C11orf87	EPHX2	MAZ	PDE4DIP	SRR	YPEL3
C16orf45	ERCC4	MED15	PDZK1	STAG1	ZDHHC19
C22orf39	ETF1	MLXIPL	PEX10	STX1A	ZDHHC8
C2orf47	FBXO45	MMP16	PEX11B	SULF2	ZNF74
C3orf43	FKBP6	MPHOSPH9	PHC2	SYNGAP1	ZNF804A
C3orf49	FMN1	MRM1	PHF7	SYNRG	ZNHIT3
CA8	FMO5	MRPL40	PI4KA	SZT2	UBE3A
CACNA1C	FURIN	MTMR10	PIAS3	TADA2A	UBXN7
CACNA1S	FUT9	ITGA10	PIGX	TAF13	UFD1L
CCDC68	FXR1	ITIH3	PIK3C2B	TANC1	VIPR2
CCDC85A	FZD9	ITIH4	PJA1	TAOK2	VPS37D
CD14	GABRA5	KCNQ5	PODXL	TBC1D5	VPS45
CD160	GABRB3	KCNV1	POLL	TBX6	VRK2
CDC45	GALNT10	KCTD13	POLR3C	TCF20	WBSCR22
CDIPT	GATAD2A	KDM3B	PPP1R16B	TCF4	WBSCR27
CENPM	GIGYF2	KDM4A	PPP4C	TCTEX1D2	WDR53
CEP19	GJA5	KIAA0430	PRKAB2	TEP1	YPEL3
CHD1L	GJA8	KIAA1244	PRKD1	TFRC	ZDHHC19
CHRNA3	GNB1L	KIF18A	PRRG2	THAP7	ZDHHC8
CHRNA5	GP1BB	KIF22	PRRT2	THOC7	ZNF74
CHRNA7	GPM6A	KLF13	PSPC1	TLE3	ZNF804A
CLCN3	GPR89B	KLHL22	PTGIS	TM4SF19	ZNHIT3
CLDN3	GREM1	LAMA4	PTK2B	TMEM191A	UBE3A
CLDN4	GRIA1	LAT2	PTPRF	TMPRSS5	UBXN7
CLDN5	GRIN2A	LHX1	QPRT	TMTC1	UFD1L
CLIP2	GRM3	LIMK1	RANBP1	TOX	VIPR2
CLTCL1	GTF2I	LPHN2	RBM8A	TRIM50	VPS37D
CLU	GTF2IRD1	LZTR1	RERE	TRIM8	VPS45
CNKSR2	HCN1	MAD1L1	RFC2	TRMT2A	VRK2
CNOT1	HERC2	MANEA	RGS6	TRPM1	WBSCR22
CNTN4	HIC2	MAPK3	RIMBP3	TSNARE1	WBSCR27
COMT	HIRA	MAZ	RIMBP3C	TSSK2	WDR53
COQ10B	HIRIP3	MED15	RIMS1	TXNIP	YPEL3
CORO1A	HNF1B	MLXIPL	RIPK2	TXNRD2	ZDHHC19
CREB3L1	HSPA8	MMP16	RNF115	UACA	ZDHHC8
CRKL	HUWE1	MPHOSPH9	RNF168	UBE3A	ZNF74
CSMD1	IMMP2L	MRM1	RORA	UBXN7	ZNF804A
CUL3	INO80E	MRPL40	RPS6KA3	UFD1L	ZNHIT3
CYFIP1	ITGA10	MTMR10	RTN4R	VIPR2	
CYP26B1	ITIH3	MVP	RYR3	VPS37D	

6.4.2 Construction of SCZ-relevant network for candidate gene prioritization

The interaction networks cCandidate and cHIPPIE_without_cCandidate were generated using high confidence protein-protein interactions (PPIs) with a HIPPIE score of ≥ 0.63 from the HIPPIE database (http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/). The focused network cCandidate connected 38 of the 39 (Table 19) identified candidate proteins to partner proteins in the HIPPIE database. L3MBTL1 was not included in cCandidate, as this protein was not present in the HIPPIE database.

Table 19: List of candidate genes

Gene symbol	Protein name
GBAS	Glucosylceramidase Beta
PASD1	PAS Domain Containing Repressor 1
ADAMTS19	ADAM Metallopeptidase With Thrombospondin Type 1 Motif 19
TEX11	Testis Expressed 11
FRG2C	FSHD Region Gene 2 Family Member C
RYR2	Ryanodine Receptor 2
TMEM74B	Transmembrane Protein 74B
MECP2	Methyl-CpG Binding Protein 2
ANKLE1	Ankyrin Repeat And LEM Domain Containing 1
ZIC5	Zic Family Member 5
OR2B11	Olfactory Receptor Family 2 Subfamily B Member 11
TXNDC16	Thioredoxin Domain Containing 16
SRRM2	Serine/Arginine Repetitive Matrix 2
RMI1	RecQ Mediated Genome Instability 1
CYP27C1	Cytochrome P450 Family 27 Subfamily C Member 1
C9orf64	Chromosome 9 Open Reading Frame 64
ZNF426	Zinc Finger Protein 426
LRP5L	LDL Receptor Related Protein 5 Like
TGDS	TDP-Glucose 4,6-Dehydratase
KIAA0226	RUN And Cysteine Rich Domain Containing Beclin 1 Interacting Protein
HDAC5	Histone Deacetylase 5
ATG9B	Autophagy Related 9B
L3MBTL1	L3MBTL1, Histone Methyl-Lysine Binding Protein
FAM185A	Family With Sequence Similarity 185 Member A
BAIAP2L1	BAI1 Associated Protein 2 Like 1
TET1	Tet Methylcytosine Dioxygenase 1
PDE7B	Phosphodiesterase 7B

GZF1	GDNF Inducible Zinc Finger Protein 1
C17orf80	Chromosome 17 Open Reading Frame 80
KRT79	Keratin 79
RGS17	Regulator Of G Protein Signaling 17
NOS3	Nitric Oxide Synthase 3
PARP6	Poly(ADP-Ribose) Polymerase Family Member 6
FOXA1	Forkhead Box A1
KRBA1	KRAB-A Domain Containing 1
SMARCC1	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin Subfamily C Member 1
DHX30	DExH-Box Helicase 30
MRPS28	Mitochondrial Ribosomal Protein S28
TPD52	Tumor Protein D52

6.4.3 ToppNet analysis

To prioritize identified SCZ candidate proteins, browser based ToppNet online tool was used (https://toppgene.cchmc.org/network.jsp).ToppNet uses a PPI data base (17 064 interactions). In order to prioritize target proteins, the PPI network had to be trained by using the list of SCZ related genes (training set), described in the results section (4.2.2 Bioinformatical ranking of SCZ candidate proteins). As a test set, the list of 38 candidate proteins were used. All proteins encoded by previously reported SCZ-associated genes that were also identified in the present study were removed from the training set. For the prioritization, k-Step Markov algorithm with a step size of 6 was used.

6.4.4 Chi square test and calculations of significance

For all calculations for significance, as well as chi-squared test, GraphPad Prism software was used.

6.4.5 Network visualization

For the visualization of PPI networks, Cytoscape 3.1 software was used.

6.4.6 Other browser-based methods

GO analysis were performed using the browser based DAVID tool (https://david.ncifcrf.gov/).

GenScript peptide property calculator was used for analyzing n-terminal ZNF804A amino acid properties.

Nuclear localization and nuclear export sequences were identified using NLS mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) and NetNES (http://www.cbs.dtu.dk/services/NetNES/) browser tools.

PredictProtein was used to determine the structure of the n-terminus of ZNF804A (https://predictprotein.org/).

Gene expression data was taken from the post mortal data set of the Allen Human Brain Atlas (http://portal.brain-map.org/).

For the analysis of SCZ association, the SCGR 2.0 data base was use (https://bioinfo.uth.edu/SZGR/). summarizes curated SCZ associations for identified ZNF804A interactors (Table 20).

Table 20: SCZ associations of ZNF804A interactors

Gene ID	Symbol	SCZ association
5936	RBM4	Common variant
83759	RBM4B	Common variant
9802	DAZAP2	Common variant
		Differently methylated
2140	EYA3	Common variant
		Differently methylated
3927	LASP1	Common variant
		Literature co-occurance
		Literature co-occurance
		GO_Annotation_neuronal
5805	PTS	Meta analysis
		Literature co-occurance
		Common variant
9444	QKI	Differently methylated
		Literature co-occurance
337974	KRTAP19-7	Common variant
		Common variant
10254	STAM2	Meta analysis
		Differently methylated
57715	SEMA4G	Common variant
		GO_Annotation_neuronal
6773	STAT2	Common variant

7. Consumables and instruments

Chemical	Distributor
Adenosine 5′-triphosphate sodium salt	Sigma-Aldrich
Agarose	Invitrogen
Ammonium persulfate (APS)	BioRad
Ampicillin-sodium salt	Sigma-Aldrich
Bacto peptone	Becton Dickinson
Bacto tryptone	Becton Dickinson
Bacto yeast extract	Becton Dickinson
BenchMark™ Protein Ladder	Invitrogen
Chloramphenicol	Sigma-Aldrich
Coelenterazine-h	NanoLight
Complete™ protease inhibitor cocktail	Roche
Diatomaceous Earth, Celite® Analytical	Sigma-Aldrich
Dimethylsulfoxide (DMSO)	Sigma-Aldrich
Dithiothreitol (DTT)	Serva
DNA ladder, Ready-Load™ 1 Kb Plus	Invitrogen
Ethidium bromide solution	Sigma-Aldrich
Ethylenediamine tetraacetic acid (EDTA)	Merck Eurolab GmbH
Fibronectin	Sigma
Gamma Globulin, Sheep	Dianova
Glycerol	Merck
IgG	Jackson immunoresearch
Imidazole	Fluka
Interferone2α	Sigma-Alberichc
Isopropyl-beta-D-thiogalactopyranoside	Fermentas
Kanamycin A monosulfate	Sigma-Aldrich
MG-132	Calbiochem
N,N,N',N'-tetramethylethylenediamine	Serva
NP-40 (IGEPAL CA 630)	Sigma-Aldrich
Polyethylenimine (PEI), linear 25kDa	Polysciences
Polyoxyethylensorbitan-Monolaureat	Sigma-Aldrich
Protein Standard Unstained, NativeMark™	Invitrogen
Protein ladder pre-stained, PageRulerTM	Fermentas
p-t-Octylphenyl-polyoxyethylen (Triton X-	Sigma-Aldrich
Sample buffer LDS, 4x NuPAGE®	Invitrogen
Spectinomycin	Sigma-Aldrich
Tetracycline	Sigma-Aldrich
TrypanBlue solution (0.4 %)	Sigma-Aldrich

All other chemicals, salts, buffer components, solvents, acids and bases were purchased from Carl Roth GmbH.

Table 21: Enzymes

Enzyme	Distributor
Benzonase purity grade II	Merck
Bdsl	Thermo
LR Clonase® II enzyme mix	Invitrogen
Polymerase, (Phusion) Hot Start	Thermo
Plasmid-Safe DNase	Biozym
Restriction enzymes	New England Biolabs
T4 DNA Ligase	Fermentas
T7 ligase	New England Biolabs
DNase 1	Qiagen

Table 22: Antibodies

Antibody	Species	Dilution	Information	Distributor
anti-ß-Actin	mice	1:20.000	clone AC-15	Sigma
anti-GFP	mice	1:20.000	ab290	Abcam
anti-mCherry	mice	1:20.000	living color	Clonetech
anti-STAT2	rabbit	1:20.000	(c20) sc-476	Santa Cruz
anti-ZNF804A	mice	1:1000		SynSys
anti-mouse IgG	horse	1:5000	HRP-linked	Cell Signalling
anti-rabbit IgG	goat	1:5000	HRP-linked	Cell Signalling
Alexa Fluor® 488 IgG	donkey	1:500	anti-mouse (Ex495/Em519)	Invitrogen
Alexa Fluor® 568 IgG	goat	1:500	anti-rabbit (Ex578/Em603)	Invitrogen

Table 23: Kits

Kit	Distributor
Dual Glo Luciferase Assay System	Promega
High Capacity RNA-to-cDNA	Applied Biosystems
ISRE Reporter Assay Kit	Qiagen
Luciferase Assay System, Dual-Glo®	Promega
Miniprep Kit, QIAprep Spin	Qiagen
Miniprep Kit, RNeasy®	Qiagen
NuPAGE Novex® Bis-Tris gel-system	Invirogen
WesternBright Quantum	Advansta

Table 24: Instruments

Device	Specificity	Distributor
Blotting device	Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell	BioRad
Imiger for blots	LAS-3000 Photo Imager	Fujifilm
Centrifuges	5810R	Eppendorf
	Avanti J-26 XP	Beckman Coulter
	Biofuge Pico	Heraeus
	Sorvall® Evolution™ RC	Thermo Scientific
Electrophoresis chambers	DNA chamber	BioRad BioRad
	PAGE XCell SureLock™ Mini	Invitrogen
	PAGE Criterion Cell	BioRad BioRad
Gel documentation systems	Gene Genius UV Imager (DNA)	Syngene
	Transilluminator	UVP
	Las-4000 (SDS-PAGE)	Fuji
Incubators	37°C incubator	Memmert
	HT culture shaker	INFORCE
	Innova 4430 Incubator Shaker	New Brunswick Scientific
	Cell culture incubator BBD6220	Heraeus
Microscopes	CK30	Olympus

	IX70	Olympus
	Fluoview 1000	Olympus
	TSC SP8 confocal laser scanning	Leica
Spectrophotometer	BioPhotometer 6131	Eppendorf
	NanoDROP 8000	Thermo Scientific
	Ultrospec 3000	Pharmacia Biotech
Plate readers	Infinite M200 Microplate Reader	TECAN
	Infinite M1000 Microplate Reader	TECAN
Power supplies	EPS 301	Amersham
	Power Pac 300 and 1000	BioRad
Scales	Genius ME2359	Sartorius
	MC1	Sartorius
	PCB 6000-1	Kern & Sohn GmbH
Thermal block	Thermomixer 533 comfort and 5436	Eppendorf
Thermocyclers	PTC200	MJ Research
	C1000™	BioRad
	ViiA7	Applied Biosystms
Vortex Mixer	Vortex-Genie 2®	Scientific Industries
Y2H robots		Kbiosystems

Table 25: Bacterial strains, yeast strains and human cell lines

Name	**Species**
BL21 (DE3) Rosetta (Novagen)	*E. coli*
Mach1™ T1® (Invitrogen)	*E. coli*
L40ccua	MATa, yeast
L40ccα	MATα, yeast
HEK 293	human kidney cell line
SH-EP	human neuroblastoma cell line

Table 26: Expression vectors

Name	Use	Source
pcmyc-NL-GW	LuTHy vector with n-terminal NanoLuc fusion protein. Gateway compatible.	AG Wanker
pPA-mCit-GW	LuTHy vector with n-terminal PA-mCitrine fusion protein. Gateway compatible.	AG Wanker
pEGFP-C1-vector	Mammalian expression vector with n-terminal GFP-tag	Clontech
pBTM116-D9	Gateway compatible plasmid for expressing bait proteins for yeast two hybrid analysis.	AG Wanker
pACT4-DM	Gateway compatible plasmid for expressing prey proteins for yeast two hybrid analysis.	AG Wanker
pPA-RL-GW	DULIP vector with n-terminal PA-*Renilla* luciferase fusion protein. Gateway compatible.	AG Wanker
pFL-V5-GW	DULIP vector with n-terminal firefly luciferase fusion protein. Gateway compatible.	AG Wanker
mCherry-C1-vector	Mammalian expression vector with n-terminal mCherry-tag	Clontech
pSPcas9(BB)-2A-Puro	Vector for CRISPR/Cas9 knock out experiments.	Addgene
pQLinkN	Gateway compatible plasmid for His-tag protein expression.	Addgene

Table 27: Primer

Name	Sequence 5´ to 3´
GRIN2A-mutation forward	p-CTG TCG ACA AAC CTA GGG AGC TA
GRIN2A-mutation reverse	p-TGC TAT CGT AGG AAT GCT GAC G
gRNA-ZNF804A forward	CAC CGA AAT ACC ATA GCA AAA GCT C
gRNA-ZNF804A reverse	AAA CGA GCT TTT GCT ATG GTA TTT C
ZNF804A-qPCR-1 forward	GGA CAC TTT CGC AAC ATC AAG G
ZNF804A-qPCR-1 reverse	CTT CCA GAG CTT TTG CTA TGG TA

ZNF804A-qPCR-2 forward	AGT GGC CCC ATG TTC AAA TCA
ZNF804A-qPCR-2 reverse	CCA CAA CAA CTC GTT GGG AAA T
Ifit1 forward	TTG ATG ACG ATG AAA TGC CTG A
Ifit1 reverse	CAG GTC ACC AGA CTC CTC AC
Isg15 forward	CGC AGA TCA CCC AGA AGA TCG
Isg15 reverse	TTC GTC GCA TTT GTC CAC CA
Mx1 forward	GTT TCC GAA GTG GAC ATC GCA
Mx1 reverse	CTG CAC AGG TTG TTC TCA GC
Rsad2 forward	GTT TCC GAA GTG GAC ATC GCA
Rsad2 reverse	CTG CAC AGG TTG TTC TCA GC
ZNF804A gRNA oligo forward	CAC CGA AAT ACC ATA GCA AAG CTC
ZNF804A gRNA oligo reverse	AAA CGA GCT TTG CT ATG GTA TTT C
Primer-prey-check forward	GAT GAA GAT ACC CCA CCA AAC C
Primer-prey-check reverse	GTG CAC GAT GCA CAG TTG AAG

8. References

1. van Os, J. & Kapur, S. Schizophrenia. *Lancet (London, England)* **374,** 635–645 (2009).

2. Schultz, S. H., North, S. W. & Shields, C. G. Schizophrenia: a review. *Am. Fam. Physician* **75,** 1821–1829 (2007).

3. Kim, Y., Zerwas, S., Trace, S. E. & Sullivan, P. F. Schizophrenia Genetics: Where Next? *Schizophr. Bull.* **37,** 456–463 (2011).

4. Patel, K. R., Cherian, J., Gohil, K. & Atkinson, D. Schizophrenia: overview and treatment options. *P T* **39,** 638–645 (2014).

5. Jablensky, A. The 100-year epidemiology of schizophrenia. *Schizophr. Res.* **28,** 111–125 (1997).

6. Jin, H. & Mosweu, I. The Societal Cost of Schizophrenia: A Systematic Review. *Pharmacoeconomics* **35,** 25–42 (2017).

7. Cloutier, M. *et al.* The Economic Burden of Schizophrenia in the United States in 2013. *J. Clin. Psychiatry* **77,** 764–771 (2016).

8. Laursen, T. M., Nordentoft, M. & Mortensen, P. B. Excess early mortality in schizophrenia. *Annu. Rev. Clin. Psychol.* **10,** 425–448 (2014).

9. Lehman, A. F. *et al.* Practice guideline for the treatment of patients with schizophrenia, second edition. *Am. J. Psychiatry* **161,** 1–56 (2004).

10. Picchioni, M. M. & Murray, R. M. Schizophrenia. *BMJ* **335,** 91–95 (2007).

11. Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).

12. Nimgaonkar, V. L., Prasad, K. M., Chowdari, K. V, Severance, E. G. & Yolken, R. H. The complement system: a gateway to gene-environment interactions in schizophrenia pathogenesis. *Mol. Psychiatry* **22,** 1554–1561 (2017).

13. van Os, J., Kenis, G. & Rutten, B. P. F. The environment and schizophrenia. *Nature* **468,** 203–212 (2010).

14. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60,** 1187–1192 (2003).

15. Meltzer, H. Y. & Stahl, S. M. The dopamine hypothesis of schizophrenia: a review. *Schizophr. Bull.* **2,** 19–76 (1976).

16. Davis, K. L., Kahn, R. S., Ko, G. & Davidson, M. Dopamine in schizophrenia: a review and reconceptualization. *Am. J. Psychiatry* **148,** 1474–1486 (1991).

17. Brisch, R. *et al.* The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: old fashioned, but still in vogue. *Front. psychiatry* **5,** 47 (2014).

18. Stone, J. M., Morrison, P. D. & Pilowsky, L. S. Review: Glutamate and dopamine dysregulation in schizophrenia — a synthesis and selective review. *J. Psychopharmacol.* **21,** 440–452 (2007).

19. Bressan, R. A. & Crippa, J. A. The role of dopamine in reward and pleasure

behaviour – review of data from preclinical research. *Acta Psychiatr. Scand.* **111,** 14–21 (2005).

20. Miyamoto, S., Duncan, G. E., Marx, C. E. & Lieberman, J. A. Treatments for schizophrenia: a critical review of pharmacology and mechanisms of action of antipsychotic drugs. *Mol. Psychiatry* **10,** 79 (2004).

21. KITZINGER, H. & ARNOLD, D. G. A preliminary study of the effects of glutamic acid on catatonic schizophrenics. *Rorschach Res. Exch. J. Proj. Tech.* **13,** 210–218 (1949).

22. Howes, O., McCutcheon, R. & Stone, J. Glutamate and dopamine in schizophrenia: an update for the 21st century. *J. Psychopharmacol.* **29,** 97–115 (2015).

23. Scheefhals, N. & MacGillavry, H. D. Functional organization of postsynaptic glutamate receptors. *Mol. Cell. Neurosci.* **91,** 82–94 (2018).

24. Krystal, J. H. *et al.* Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. *Arch. Gen. Psychiatry* **51,** 199–214 (1994).

25. Morgan, C. J. A. & Curran, H. V. Acute and chronic effects of ketamine upon human memory: a review. *Psychopharmacology (Berl).* **188,** 408–424 (2006).

26. Javitt, D. C. Glutamate and schizophrenia: phencyclidine, N-methyl-D-aspartate receptors, and dopamine-glutamate interactions. *Int. Rev. Neurobiol.* **78,** 69–108 (2007).

27. Rothman, D. L., Behar, K. L., Hyder, F. & Shulman, R. G. In vivo NMR studies of the glutamate neurotransmitter flux and neuroenergetics: implications for brain function. *Annu. Rev. Physiol.* **65,** 401–427 (2003).

28. Woo, T.-U. W. Neurobiology of schizophrenia onset. *Curr. Top. Behav. Neurosci.* **16,** 267–295 (2014).

29. Feinberg, I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? *J. Psychiatr. Res.* **17,** 319–334

30. Keshavan, M. S., Anderson, S. & Pettegrew, J. W. Is schizophrenia due to excessive synaptic pruning in the prefrontal cortex? The Feinberg hypothesis revisited. *J. Psychiatr. Res.* **28,** 239–265 (1994).

31. McGlashan, T. H. & Hoffman, R. E. Schizophrenia as a disorder of developmentally reduced synaptic connectivity. *Arch. Gen. Psychiatry* **57,** 637–648 (2000).

32. Muller, N., Weidinger, E., Leitner, B. & Schwarz, M. J. The role of inflammation in schizophrenia. *Front. Neurosci.* **9,** 372 (2015).

33. Watkins, C. C. & Andrews, S. R. Clinical studies of neuroinflammatory mechanisms in schizophrenia. *Schizophr. Res.* **176,** 14–22 (2016).

34. Brown, A. S. & Derkits, E. J. Prenatal infection and schizophrenia: a review of epidemiologic and translational studies. *Am. J. Psychiatry* **167,** 261–280 (2010).

35. Babulas, V., Factor-Litvak, P., Goetz, R., Schaefer, C. A. & Brown, A. S. Prenatal exposure to maternal genital and reproductive infections and adult schizophrenia. *Am. J. Psychiatry* **163,** 927–929 (2006).

36. Sorensen, H. J., Mortensen, E. L., Reinisch, J. M. & Mednick, S. A. Association

between prenatal exposure to bacterial infection and risk of schizophrenia. *Schizophr. Bull.* **35,** 631–637 (2009).

37. Pearce, B. D. Schizophrenia and viral infection during neurodevelopment: a focus on mechanisms. *Mol. Psychiatry* **6,** 634–646 (2001).

38. Brown, A. S. *et al.* Serologic evidence of prenatal influenza in the etiology of schizophrenia. *Arch. Gen. Psychiatry* **61,** 774–780 (2004).

39. Buka, S. L., Cannon, T. D., Torrey, E. F. & Yolken, R. H. Maternal exposure to herpes simplex virus and risk of psychosis among adult offspring. *Biol. Psychiatry* **63,** 809–815 (2008).

40. Brown, A. S. *et al.* Maternal exposure to toxoplasmosis and risk of schizophrenia in adult offspring. *Am. J. Psychiatry* **162,** 767–773 (2005).

41. Meyer, U. & Feldon, J. Prenatal exposure to infection: a primary mechanism for abnormal dopaminergic development in schizophrenia. *Psychopharmacology (Berl).* **206,** 587–602 (2009).

42. Meyer, U., Schwarz, M. J. & Muller, N. Inflammatory processes in schizophrenia: a promising neuroimmunological target for the treatment of negative/cognitive symptoms and beyond. *Pharmacol. Ther.* **132,** 96–110 (2011).

43. Benros, M. E. *et al.* Autoimmune diseases and severe infections as risk factors for schizophrenia: a 30-year population-based register study. *Am. J. Psychiatry* **168,** 1303–1310 (2011).

44. Benros, M. E., Mortensen, P. B. & Eaton, W. W. Autoimmune diseases and infections as risk factors for schizophrenia. *Ann. N. Y. Acad. Sci.* **1262,** 56–66 (2012).

45. Korschenhausen, D. A., Hampel, H. J., Ackenheil, M., Penning, R. & Muller, N. Fibrin degradation products in post mortem brain tissue of schizophrenics: a possible marker for underlying inflammatory processes. *Schizophr. Res.* **19,** 103–109 (1996).

46. Wildenauer, D. B. *et al.* Analysis of cerebrospinal fluid from patients with psychiatric and neurological disorders by two-dimensional electrophoresis: identification of disease-associated polypeptides as fibrin fragments. *Electrophoresis* **12,** 487–492 (1991).

47. O'Donovan, M. C. *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.* **40,** 1053–1055 (2008).

48. Rees, E. *et al.* Analysis of copy number variations at 15 schizophrenia-associated loci. *Br. J. Psychiatry* **204,** 108–114 (2014).

49. Kirov, G. *et al.* Neurexin 1 (NRXN1) deletions in schizophrenia. *Schizophr. Bull.* **35,** 851–854 (2009).

50. Schneider, M. *et al.* Psychiatric disorders from childhood to adulthood in 22q11.2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. *Am. J. Psychiatry* **171,** 627–639 (2014).

51. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506,** 179–184 (2014).

52. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506,** 185–190 (2014).

53. Menche, J. *et al.* Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347,** 1257601 (2015).

54. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92,** 291–294 (1998).

55. Causier, B. Studying the interactome with the yeast two-hybrid system and mass spectrometry. *Mass Spectrom. Rev.* **23,** 350–367 (2004).

56. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340,** 245–246 (1989).

57. Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122,** 957–968 (2005).

58. Rual, J.-F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437,** 1173–1178 (2005).

59. Buntru, A., Trepte, P., Klockmeier, K., Schnoegl, S. & Wanker, E. E. Current Approaches Toward Quantitative Mapping of the Interactome. *Front. Genet.* **7,** 74 (2016).

60. Hamdi, A. & Colas, P. Yeast two-hybrid methods and their applications in drug discovery. *Trends Pharmacol. Sci.* **33,** 109–118 (2012).

61. Braun, P. *et al.* An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6,** 91–97 (2009).

62. Chen, Y.-C., Rajagopala, S. V., Stellberger, T. & Uetz, P. Exhaustive benchmarking of the yeast two-hybrid system. *Nature methods* **7,** 667–8; author reply 668 (2010).

63. Huang, H. & Bader, J. S. Precision and recall estimates for two-hybrid screens. *Bioinformatics* **25,** 372–378 (2009).

64. Barrios-Rodiles, M. *et al.* High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* **307,** 1621–1625 (2005).

65. Trepte, P. *et al.* DULIP: A Dual Luminescence-Based Co-Immunoprecipitation Assay for Interactome Mapping in Mammalian Cells. *J. Mol. Biol.* **427,** 3375–3388 (2015).

66. Trepte, P. *et al.* LuTHy: a double-readout bioluminescence-based two-hybrid technology for quantitative mapping of protein-protein interactions in mammalian cells. *Mol. Syst. Biol.* **14,** e8071 (2018).

67. Ganapathiraju, M. K. *et al.* Schizophrenia interactome with 504 novel protein-protein interactions. *NPJ Schizophr.* **2,** 16012 (2016).

68. Luo, X. *et al.* Protein-Protein Interaction and Pathway Analyses of Top Schizophrenia Genes Reveal Schizophrenia Susceptibility Genes Converge on Common Molecular Networks and Enrichment of Nucleosome (Chromatin) Assembly Genes in Schizophrenia Susceptibility Loci. *Schizophr. Bull.* **40,** 39–49 (2013).

69. Pizzuti, C. & Rombo, S. E. Algorithms and tools for protein-protein interaction networks clustering, with a special focus on population-based stochastic methods. *Bioinformatics* **30,** 1343–1352 (2014).

70. Bader, G. D. & Hogue, C. W. V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4,** 2 (2003).

71. Niklas, N. *et al.* cFinder: definition and quantification of multiple haplotypes in a mixed sample. *BMC Res. Notes* **8,** 422 (2015).

72. King, A. D., Przulj, N. & Jurisica, I. Protein complex prediction with RNSC. *Methods Mol. Biol.* **804,** 297–312 (2012).

73. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30,** 1575–1584 (2002).

74. Vanunu, O., Magger, O., Ruppin, E., Shlomi, T. & Sharan, R. Associating genes and protein complexes with disease via network propagation. *PLoS Comput. Biol.* **6,** e1000641 (2010).

75. Gilman, S. R. *et al.* Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat. Neurosci.* **15,** 1723–1728 (2012).

76. Need, A. C. *et al.* Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am. J. Hum. Genet.* **91,** 303–312 (2012).

77. Isakov, O., Perrone, M. & Shomron, N. Exome sequencing analysis: a guide to disease variant detection. *Methods Mol. Biol.* **1038,** 137–158 (2013).

78. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37,** W305-11 (2009).

79. Riley, B. *et al.* Replication of association between schizophrenia and ZNF804A in the Irish Case-Control Study of Schizophrenia sample. *Mol. Psychiatry* **15,** 29–37 (2010).

80. Steinberg, S. *et al.* Expanding the range of ZNF804A variants conferring risk of psychosis. *Mol. Psychiatry* **16,** 59–66 (2011).

81. Williams, H. J. *et al.* Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol. Psychiatry* **16,** 429–441 (2011).

82. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489,** 391–399 (2012).

83. Zhang, F. *et al.* Evidence of sex-modulated association of ZNF804A with schizophrenia. *Biol. Psychiatry* **69,** 914–917 (2011).

84. Hill, M. J. & Bray, N. J. Evidence that schizophrenia risk variation in the ZNF804A gene exerts its effects during fetal brain development. *Am. J. Psychiatry* **169,** 1301–1308 (2012).

85. Tao, R. *et al.* Expression of ZNF804A in human brain and alterations in schizophrenia, bipolar disorder, and major depressive disorder: a novel transcript fetally regulated by the psychosis risk variant rs1344706. *JAMA psychiatry* **71,** 1112–1120 (2014).

86. Hill, M. J., Jeffries, A. R., Dobson, R. J. B., Price, J. & Bray, N. J. Knockdown of the psychosis susceptibility gene ZNF804A alters expression of genes involved in cell adhesion. *Hum. Mol. Genet.* **21,** 1018–1024 (2012).

87. Chen, J. *et al.* ZNF804A Transcriptional Networks in Differentiating Neurons Derived from Induced Pluripotent Stem Cells of Human Origin. *PLoS One* **10,** e0124597 (2015).

88. Deans, P. J. M. *et al.* Psychosis Risk Candidate ZNF804A Localizes to Synapses and Regulates Neurite Formation and Dendritic Spine Structure. *Biol. Psychiatry* **82,** 49–61 (2017).

89. Rasetti, R. *et al.* Altered cortical network dynamics: a potential intermediate phenotype for schizophrenia and association with ZNF804A. *Arch. Gen. Psychiatry* **68,** 1207–1217 (2011).

90. Esslinger, C. *et al.* Neural mechanisms of a genome-wide supported psychosis variant. *Science* **324,** 605 (2009).

91. Esslinger, C. *et al.* Cognitive state and connectivity effects of the genome-wide significant psychosis variant in ZNF804A. *Neuroimage* **54,** 2514–2523 (2011).

92. Walter, H. *et al.* Effects of a genome-wide supported psychosis risk variant on neural activation during a theory-of-mind task. *Mol. Psychiatry* **16,** 462–470 (2011).

93. Mohnke, S. *et al.* Further evidence for the impact of a genome-wide-supported psychosis risk variant in ZNF804A on the Theory of Mind Network. *Neuropsychopharmacology* **39,** 1196–1205 (2014).

94. Blaszczyk, K. *et al.* The unique role of STAT2 in constitutive and IFN-induced transcription and antiviral responses. *Cytokine Growth Factor Rev.* **29,** 71–81 (2016).

95. Liu, Y.-J. IPC: Professional Type 1 Interferon-Producing Cells and Plasmacytoid Dendritic Cell Precursors. *Annu. Rev. Immunol.* **23,** 275–306 (2005).

96. Darnell, J. E. J., Kerr, I. M. & Stark, G. R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* **264,** 1415–1421 (1994).

97. Haque, S. J. & Williams, B. R. Identification and characterization of an interferon (IFN)-stimulated response element-IFN-stimulated gene factor 3-independent signaling pathway for IFN-alpha. *J. Biol. Chem.* **269,** 19523–19529 (1994).

98. Martinez-Moczygemba, M., Gutch, M. J., French, D. L. & Reich, N. C. Distinct STAT structure promotes interaction of STAT2 with the p48 subunit of the interferon-alpha-stimulated transcription factor ISGF3. *J. Biol. Chem.* **272,** 20070–20076 (1997).

99. Ghislain, J. J., Wong, T., Nguyen, M. & Fish, E. N. The interferon-inducible Stat2:Stat1 heterodimer preferentially binds in vitro to a consensus element found in the promoters of a subset of interferon-stimulated genes. *J. Interferon Cytokine Res.* **21,** 379–388 (2001).

100. Gupta, S., Jiang, M. & Pernis, A. B. IFN-alpha activates Stat6 and leads to the formation of Stat2:Stat6 complexes in B cells. *J. Immunol.* **163,** 3834–3841 (1999).

101. Bluyssen, H. A. & Levy, D. E. Stat2 is a transcriptional activator that requires sequence-specific contacts provided by stat1 and p48 for stable interaction with DNA. *J. Biol. Chem.* **272,** 4600–4605 (1997).

102. Kraus, T. A., Lau, J. F., Parisien, J.-P. & Horvath, C. M. A hybrid IRF9-STAT2 protein recapitulates interferon-stimulated gene expression and antiviral response. *J. Biol. Chem.* **278,** 13033–13038 (2003).

103. Zahiri, J. *et al.* Protein complex prediction: A survey. *Genomics* (2019).

doi:10.1016/j.ygeno.2019.01.011

104. Schaefer, M. H. *et al.* Adding Protein Context to the Human Protein-Protein Interaction Network to Reveal Meaningful Interactions. *PLOS Comput. Biol.* **9,** 1–11 (2013).

105. Alanis-Lobato, G., Andrade-Navarro, M. A. & Schaefer, M. H. HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.* **45,** D408–D414 (2016).

106. Suratanee, A. *et al.* Characterizing Protein Interactions Employing a Genome-Wide siRNA Cellular Phenotyping Screen. *PLOS Comput. Biol.* **10,** 1–13 (2014).

107. Schaefer, M. H. *et al.* HIPPIE: Integrating Protein Interaction Networks with Experiment Based Quality Scores. *PLoS One* **7,** 1–8 (2012).

108. Dienel, S. J. & Lewis, D. A. Alterations in cortical interneurons and cognitive function in schizophrenia. *Neurobiol. Dis.* (2018). doi:10.1016/j.nbd.2018.06.020

109. Nagano, T., Jourdi, H. & Nawa, H. Emerging roles of Dlg-like PDZ proteins in the organization of the NMDA-type glutamatergic synapse. *J. Biochem.* **124,** 869–875 (1998).

110. Boeckers, T. M. The postsynaptic density. *Cell Tissue Res.* **326,** 409–422 (2006).

111. Rasmussen, A. H., Rasmussen, H. B. & Silahtaroglu, A. The DLGAP family: neuronal expression, function and role in brain disorders. *Mol. Brain* **10,** 43 (2017).

112. Karmakar, S., Sharma, L. G., Roy, A., Patel, A. & Pandey, L. M. Neuronal SNARE complex: A protein folding system with intricate protein-protein interactions, and its common neuropathological hallmark, SNAP25. *Neurochem. Int.* **122,** 196–207 (2019).

113. Sudhof, T. C. & Rothman, J. E. Membrane fusion: grappling with SNARE and SM proteins. *Science* **323,** 474–477 (2009).

114. Chen, Y. A. & Scheller, R. H. SNARE-mediated membrane fusion. *Nat. Rev. Mol. Cell Biol.* **2,** 98–106 (2001).

115. Sudhof, T. C. Neurotransmitter release: the last millisecond in the life of a synaptic vesicle. *Neuron* **80,** 675–690 (2013).

116. Kabir, Z. D., Martinez-Rivera, A. & Rajadhyaksha, A. M. From Gene to Behavior: L-Type Calcium Channel Mechanisms Underlying Neuropsychiatric Symptoms. *Neurotherapeutics* **14,** 588–613 (2017).

117. Ament, S. A. *et al.* Rare variants in neuronal excitability genes influence risk for bipolar disorder. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 3576–3581 (2015).

118. Nyegaard, M. *et al.* CACNA1C (rs1006737) is associated with schizophrenia. *Molecular psychiatry* **15,** 119–121 (2010).

119. Rao, S. *et al.* Common variants in CACNA1C and MDD susceptibility: A comprehensive meta-analysis. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **171,** 896–903 (2016).

120. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet (London, England)* **381,** 1371–1379 (2013).

121. Zhuo, C., Hou, W., Lin, C., Hu, L. & Li, J. Potential Value of Genomic Copy

Number Variations in Schizophrenia. *Front. Mol. Neurosci.* **10,** 204 (2017).

122. Förster, T. Zwischenmolekulare energiewanderung und fluoreszenz. *Ann. Phys.* **437,** 55–75 (1948).

123. Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7,** 1848–1857 (2012).

124. Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* **49,** 27–35 (2017).

125. Smedemark-Margulies, N. *et al.* A novel de novo mutation in ATP1A3 and childhood-onset schizophrenia. *Cold Spring Harb. Mol. case Stud.* **2,** a001008 (2016).

126. Gulsuner, S. *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154,** 518–529 (2013).

127. Takata, A. *et al.* Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. *Neuron* **82,** 773–780 (2014).

128. Wang, Q. *et al.* Increased co-expression of genes harboring the damaging de novo mutations in Chinese schizophrenic patients during prenatal development. *Sci. Rep.* **5,** 18209 (2015).

129. Ambalavanan, A. *et al.* De novo variants in sporadic cases of childhood onset schizophrenia. *Eur. J. Hum. Genet.* **24,** 944–948 (2016).

130. McCarthy, S. E. *et al.* De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol. Psychiatry* **19,** 652–658 (2014).

131. Xu, B. *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* **44,** 1365–1369 (2012).

132. Xu, B. *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat. Genet.* **43,** 864–868 (2011).

133. Girard, S. L. *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat. Genet.* **43,** 860–863 (2011).

134. Guipponi, M. *et al.* Exome sequencing in 53 sporadic cases of schizophrenia identifies 18 putative candidate genes. *PLoS One* **9,** e112745 (2014).

135. Shahbazian, M. D. & Grunstein, M. Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.* **76,** 75–100 (2007).

136. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320,** 1224–1229 (2008).

137. Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* **8,** 14648 (2017).

138. Chen, J., Aronow, B. J. & Jegga, A. G. Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics* **10,** 73 (2009).

139. Hess, J. L. & Glatt, S. J. How might ZNF804A variants influence risk for schizophrenia and bipolar disorder? A literature review, synthesis, and bioinformatic analysis. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **165B,** 28–40 (2014).

140. Aberg, K. *et al.* Human QKI, a new candidate gene for schizophrenia involved in

myelination. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **141B,** 84–90 (2006).

141. Artzt, K. & Wu, J. I. STAR trek: An introduction to STAR family proteins and review of quaking (QKI). *Adv. Exp. Med. Biol.* **693,** 1–24 (2010).

142. Jia, P., Han, G., Zhao, J., Lu, P. & Zhao, Z. SZGR 2.0: a one-stop shop of schizophrenia candidate genes. *Nucleic Acids Res.* **45,** D915–D924 (2017).

143. Karatsoreos, I. N. Links between Circadian Rhythms and Psychiatric Disease. *Front. Behav. Neurosci.* **8,** 162 (2014).

144. Whitford, T. J., Ford, J. M., Mathalon, D. H., Kubicki, M. & Shenton, M. E. Schizophrenia, myelination, and delayed corollary discharges: a hypothesis. *Schizophr. Bull.* **38,** 486–494 (2012).

145. Girgenti, M. J., LoTurco, J. J. & Maher, B. J. ZNF804a regulates expression of the schizophrenia-associated genes PRSS16, COMT, PDE4B, and DRD2. *PLoS One* **7,** e32404 (2012).

146. Luo, S.-Q. *et al.* The effects of promoter methylation on downregulation of DAZAP2 in multiple myeloma cell lines. *PLoS One* **7,** e40475 (2012).

147. Lukas, J. *et al.* Dazap2 modulates transcription driven by the Wnt effector TCF-4. *Nucleic Acids Res.* **37,** 3007–3020 (2009).

148. Komiya, Y. & Habas, R. Wnt signal transduction pathways. *Organogenesis* **4,** 68–75 (2008).

149. Kuchta, K. *et al.* FAM46 proteins are novel eukaryotic non-canonical poly(A) polymerases. *Nucleic Acids Res.* **44,** 3534–3548 (2016).

150. Su, C.-H., Hung, K.-Y., Hung, S.-C. & Tarn, W.-Y. RBM4 Regulates Neuronal Differentiation of Mesenchymal Stem Cells by Modulating Alternative Splicing of Pyruvate Kinase M. *Mol. Cell. Biol.* **37,** (2017).

151. Tarn, W.-Y. *et al.* RBM4 promotes neuronal differentiation and neurite outgrowth by modulating Numb isoform expression. *Mol. Biol. Cell* **27,** 1676–1683 (2016).

152. Lin, J.-C. *et al.* RBM4 promotes pancreas cell differentiation and insulin expression. *Mol. Cell. Biol.* **33,** 319–327 (2013).

153. Wang, C., Chen, Y., Deng, H., Gao, S. & Li, L. Rbm46 regulates trophectoderm differentiation by stabilizing Cdx2 mRNA in early mouse embryos. *Stem Cells Dev.* **24,** 904–915 (2015).

154. Tadjuidje, E. & Hegde, R. S. The Eyes Absent proteins in development and disease. *Cell. Mol. Life Sci.* **70,** 1897–1913 (2013).

155. Wesseling, H., Elgersma, Y. & Bahn, S. A brain proteomic investigation of rapamycin effects in the Tsc1(+/-) mouse model. *Mol. Autism* **8,** 41 (2017).

156. Cho, C. H. Frontier of epilepsy research - mTfOR signaling pathway. *Exp. Mol. Med.* **43,** 231–274 (2011).

157. van Eeghen, A. M. *et al.* Understanding relationships between autism, intelligence, and epilepsy: a cross-disorder approach. *Dev. Med. Child Neurol.* **55,** 146–153 (2013).

158. Monin, L. & Gaffen, S. L. Interleukin 17 Family Cytokines: Signaling Mechanisms, Biological Activities, and Therapeutic Implications. *Cold Spring Harb. Perspect. Biol.* **10,** (2018).

159. Zepp, J. A. *et al.* TRAF4-SMURF2-mediated DAZAP2 degradation is critical for IL-25 signaling and allergic airway inflammation. *J. Immunol.* **194,** 2826–2837 (2015).

160. Merrill, J. C., You, J., Constable, C., Leeman, S. E. & Amar, S. Whole-body deletion of LPS-induced TNF-alpha factor (LITAF) markedly improves experimental endotoxic shock and inflammatory arthritis. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 21247–21252 (2011).

161. Tang, X., Metzger, D., Leeman, S. & Amar, S. LPS-induced TNF-alpha factor (LITAF)-deficient mice express reduced LPS-induced cytokine: Evidence for LITAF-dependent LPS signaling pathways. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 13777–13782 (2006).

162. Zhang, B., Zou, J., Rensing, N. R., Yang, M. & Wong, M. Inflammatory mechanisms contribute to the neurological manifestations of tuberous sclerosis complex. *Neurobiol. Dis.* **80,** 70–79 (2015).

163. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159,** 1212–1226 (2014).

164. Honda, K. & Taniguchi, T. Toll-like receptor signaling and IRF transcription factors. *IUBMB Life* **58,** 290–295 (2006).

165. Schlessinger, A., Yachdav, G. & Rost, B. PROFbval: predict flexible and rigid residues in proteins. *Bioinformatics* **22,** 891–893 (2006).

166. Schlessinger, A., Punta, M. & Rost, B. Natively unstructured regions in proteins identified from contact predictions. *Bioinformatics* **23,** 2376–2384 (2007).

167. Schlessinger, A., Liu, J. & Rost, B. Natively unstructured loops differ from other loops. *PLoS Comput. Biol.* **3,** e140 (2007).

168. Schlessinger, A., Punta, M., Yachdav, G., Kajan, L. & Rost, B. Improved disorder prediction by combination of orthogonal approaches. *PLoS One* **4,** e4433 (2009).

169. Mészáros, B., Tompa, P., Simon, I. & Dosztányi, Z. Molecular Principles of the Interactions of Disordered Proteins. *J. Mol. Biol.* **372,** 549–561 (2007).

170. Dogan, J., Gianni, S. & Jemth, P. The binding mechanisms of intrinsically disordered proteins. *Phys. Chem. Chem. Phys.* **16,** 6323–6331 (2014).

171. Sarkis, P. T. N., Ying, S., Xu, R. & Yu, X.-F. STAT1-independent cell type-specific regulation of antiviral APOBEC3G by IFN-alpha. *J. Immunol.* **177,** 4530–4540 (2006).

172. Meyer, T. & Horisberger, M. A. Combined action of mouse alpha and beta interferons in influenza virus-infected macrophages carrying the resistance gene Mx. *J. Virol.* **49,** 709–716 (1984).

173. Krug, R. M., Shaw, M., Broni, B., Shapiro, G. & Haller, O. Inhibition of influenza viral mRNA synthesis in cells expressing the interferon-induced Mx gene product. *J. Virol.* **56,** 201–206 (1985).

174. Fensterl, V. & Sen, G. C. Interferon-induced Ifit proteins: their role in viral pathogenesis. *J. Virol.* **89,** 2462–2468 (2015).

175. Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* **13,** 537–551 (2012).

176. Gottesman, I. I. Schizophrenia genesis: The origins of madness. *Schizophrenia genesis: The origins of madness.* xiii, 296-xiii, 296 (1991).

177. Gottesman, I. I. & Gould, T. D. The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions. *Am. J. Psychiatry* **160,** 636–645 (2003).

178. Liu, M. *et al.* Psychophysiological endophenotypes to characterize mechanisms of known schizophrenia genetic loci. *Psychol. Med.* **47,** 1116–1125 (2017).

179. McGue, M. The end of behavioral genetics? 2008. *Behav. Genet.* **40,** 284–296 (2010).

180. Earls, H. A., Curran, T. & Mittal, V. A Meta-analytic Review of Auditory Event-Related Potential Components as Endophenotypes for Schizophrenia: Perspectives From First-Degree Relatives. *Schizophr. Bull.* **42,** 1504–1516 (2016).

181. Swerdlow, N. R., Gur, R. E. & Braff, D. L. Consortium on the Genetics of Schizophrenia (COGS) assessment of endophenotypes for schizophrenia: an introduction to this Special Issue of Schizophrenia Research. *Schizophr. Res.* **163,** 9–16 (2015).

182. Cadenhead, K. S., Light, G. A., Geyer, M. A., McDowell, J. E. & Braff, D. L. Neurobiological measures of schizotypal personality disorder: defining an inhibitory endophenotype? *Am. J. Psychiatry* **159,** 869–871 (2002).

183. DiLalla, L. F., McCrary, M. & Diaz, E. A review of endophenotypes in schizophrenia and autism: The next phase for understanding genetic etiologies. *Am. J. Med. Genet. C. Semin. Med. Genet.* **175,** 354–361 (2017).

184. Braff, D. L. The importance of endophenotypes in schizophrenia research. *Schizophrenia research* **163,** 1–8 (2015).

185. Sahni, N. *et al.* Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161,** 647–660 (2015).

186. Zhou, Y. *et al.* Interactome analysis reveals ZNF804A, a schizophrenia risk gene, as a novel component of protein translational machinery critical for embryonic neurodevelopment. *Mol. Psychiatry* **23,** 952–962 (2018).

187. Benros, M. E. *et al.* A nationwide study on the risk of autoimmune diseases in individuals with a personal or a family history of schizophrenia and related psychosis. *Am. J. Psychiatry* **171,** 218–226 (2014).

188. Oommen, K. J., Johnson, P. C. & Ray, C. G. Herpes simplex type 2 virus encephalitis presenting as psychosis. *Am. J. Med.* **73,** 445–448 (1982).

189. Mazza, A., Klockmeier, K., Wanker, E. & Sharan, R. An integer programming framework for inferring disease complexes from network data. *Bioinformatics* **32,** 3855 (2016).

9. Publications

1. Stroedicke, M. *et al.* Systematic interaction network filtering identifies CRMP1 as a novel suppressor of huntingtin misfolding and neurotoxicity. *Genome Res.* **25,** 701–713 (2015).

2. Wagner, A. S. *et al.* Self-assembly of Mutant Huntingtin Exon-1 Fragments into Large Complex Fibrillar Structures Involves Nucleated Branching. *J. Mol. Biol.* **430,** 1725–1744 (2018).

3. Ast, A. *et al.* mHTT Seeding Activity: A Marker of Disease Progression and Neurotoxicity in Models of Huntington's Disease. *Mol. Cell* **71,** 675–688.e6 (2018).

4. Buntru, A., Trepte, P., Klockmeier, K., Schnoegl, S. & Wanker, E. E. Current Approaches Toward Quantitative Mapping of the Interactome. *Front. Genet.* **7,** 74 (2016).

5. Mazza, A., Klockmeier, K., Wanker, E. & Sharan, R. An integer programming framework for inferring disease complexes from network data. *Bioinformatics* **32,** 3855 (2016).

6. Trepte, P. *et al.* LuTHy: a double-readout bioluminescence-based two-hybrid technology for quantitative mapping of protein-protein interactions in mammalian cells. *Mol. Syst. Biol.* **14,** e8071 (2018).

7. Trepte, P. et al. DULIP: A Dual Luminescence-Based Co-Immunoprecipitation Assay for Interactome Mapping in Mammalian Cells. J. Mol. Biol. 427, 3375–3388 (2015)

Selbständigkeitserklärung

Name: Klockmeier	
Vorname: Konrad	
Geb.am:	
Matr.Nr.:	

Ich erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe.

Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch bei keiner anderen Universität als Prüfungsleistung eingereicht.

Datum:_____   Unterschrift:_____