eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

Functional analysis of ANKRD55, a multiple sclerosis risk gene with unknown function

PhD Thesis

Author: Nerea Ugidos Damboriena

Director: Koen Vandenbroeck

2018

CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

LIST OF ABBREVIATIONS

MS	Multiple sclerosis
CNS	Central nervous system
BBB	Blood–brain barrier
CSF	Cerebrospinal fluid
EAE	Experimental autoimmune encephalomyelitis
MRI	Magnetic resonance imaging
IFNβ	Interferon β
GWAS	Genome-wide association studies
CIS	Clinically isolated syndrome
RRMS	Relapsing-remitting multiple sclerosis
PPMS	Primary progressive multiple sclerosis
SPMS	Secondary progressive multiple sclerosis
NK	Natural killer
DC	Dendritic cell
HLA	Human leukocyte antigen
MHC	Major histocompatibility complex
TCR	T cell receptor
APC	Antigen-presenting cells
Th	$CD4^+$ T helper cells
TReg	FOXP3 expressing $CD4^+$ T regulatory cells
TR1	IL-10 producing $CD4^+$ T regulatory type 1 cells
RA	Rheumatoid arthritis
T1DM	Type 1 diabetes mellitus
SLE	Systemic lupus erythematosus
CD	Crohn disease
AR	Ankyrin repeat
SNP	Single-nucleotides polymorphism
OR	Odds ratio
LD	Linkage disequilibrium
eQTL	Expression quantitative trait loci
CRISPR	Clustered regularly interspaced short palindromic repeats

dCas9	Nuclease-deactivated CRISPR-associated protein
PBMC	Peripheral blood mononuclear cells
sgRNA	Single-guide RNA
VEP	Variant Effect Predictor
DSP	Dithiobis [succinimidyl propionate]
PPI	Protein–protein interactions
Y2H	Yeast two-hybrid
IP	Immunoprecipitation
nLC-MS/MS	Nano-flow liquid chromatography coupled to tandem mass spectrometry
NSAF	Normalized Spectral Abundance Factor
RP	Ribosomal proteins
SLC	Solute carrier proteins
GO	Gene Ontology
IPA	Ingenuity Pathway Analysis

PREFACE: CONTEXT AND STRUCTURE OF THE THESIS

The starting point of the work presented here was the identification of *ANKRD55* as a genetic risk factor of multiple sclerosis by the Neurogenomiks research group in 2012 and its consequent characterization of expression in the immune and nervous systems.

This thesis is organized in three parts. The first chapter includes a general introduction about multiple sclerosis, highlighting the role of the immune system and genetics, and characteristics of *ANKRD55*, a member of the ankyrin repeat domain family, associated with several autoimmune and other type of diseases. Then the aim and objectives of the thesis are specified. The second chapter contains the results and studies carried out in the context of DNA variants of *ANKRD55*. The third chapter includes the results and the analysis based on the ANKRD55 protein. More detailed introductions about functional genomics and proteomics are included in the corresponding chapters. Finally, the general conclusions derived from both chapters are presented.

Part of this work was included in the publication in the appendix. Furthermore, a manuscript detailing ANKRD55-protein based analysis performed in this thesis is currently in preparation.

CHAPTER 1: GENERAL INTRODUCTION

1. DEFINITION AND HISTORY OF MULTIPLE SCLEROSIS

Multiple sclerosis (MS) is an inflammatory demyelinating disease of the central nervous system (CNS) that results in chronic progressive disability for the majority of people with the disorder[1]. This condition has a heterogeneous presentation that can involve motor, sensory, visual and autonomic systems[2]. The variation in clinical manifestations correlates with the spatiotemporal dissemination of lesional sites of pathology within CNS[3], affecting both white matter tracts and cortical and deep gray matter[4]. These lesions are a hallmark of MS and are caused by immune cell infiltration across the blood–brain barrier (BBB) that promotes inflammation, demyelination, gliosis and neuroaxonal degeneration, leading to disruption of neuronal signaling[5].

Tracing the historical roots of MS has proven to be difficult, given the lack of knowledge of clinical-anatomic localization in neurology prior to the late 19th century. Two cases from the late 13th century described women afflicted with chronic, multifocal, and partially remitting neurologic illnesses that conceivably might have been MS[6–8]. The idea that MS is a disseminated plaque-like sclerosis was established approximately 150 years ago[9] when the features of MS were first well defined by Jean-Martin Charcot neurologist in 1868, as 'la sclérose en plaques'[10,11].

The discovery of the myelin sheath by Rudolf-Carl Virchow in 1854, the discovery of Ranvier nodes, Schwann cells and the characterization of the nerve-fiber by Louis-Antoine Ranvier, as well as the discovery of oligodendrocytes in the late 1920s, shed light on the pathophysiology of MS and provided an explanation for the wide range of clinical symptoms[12].

In 1916, James Dawson reported about MS neuropathology when he described microscopic myelin damage and blood vessel inflammation occurring in the brain thanks to the perfected new chemicals to enhance the visibility of nerve cells under the microscope by Dr. Camillo Golgi and Dr. Santiago Ramon y Cajal[13,14].

In the early 1930s, Thomas Rivers and colleagues showed that healthy monkeys injected with virus-free brain extract developed acute CNS disease with immune cell infiltration

and demyelinating lesions[15] and established what is now the most well-studied model of MS and autoimmunity, the experimental autoimmune encephalomyelitis (EAE)[12].

In 1981, magnetic resonance imaging (MRI) was first used to examine patients with MS[16]. The use of MRI in MS has since revolutionized how patients are diagnosed and further enlightened researchers on the pathophysiology of the illness[17].

In 1993, interferon β (IFNβ) was the first Food and Drug Administration (FDA) approved immunomodulatory drug shown to alter the clinical course of MS by reducing both disability and MRI lesion burden[18]. A total of 15 FDA-approved treatments for relapsing-remitting MS are now available as disease modifiers to control inflammatory lesions and clinical relapsing activity[9,19].

The completion of the human genome sequence in 2003 clearly marked the beginning of a new era for biomedical research. Moreover, it has revealed millions of genetic variants in the human genome. This has generated an unprecedented explosion of innovative analysis techniques that can take full advantage of the full sequence data and the corresponding functionality of the genome[20]. From 2005, the development of new, next-generation sequencing technologies greatly facilitated the fast sequencing of DNA. As a result, genome-wide association studies (GWAS) have revolutionized the search for genetic risk variants underlying complex diseases[20], such as MS. Currently, the greatest challenge in the 'post GWAS' era is to understand the functional consequences of these DNA variants and to accurately elucidate the biological mechanism by which these genes and variants act[21].

2. CLINICAL FEATURES AND COURSE OF MULTIPLE SCLEROSIS

There is much variability in the clinical features of MS, in the onset as well as in the progression of the disease[22]. Clinical symptoms of MS often correlate with the functional localization of impaired conduction in MS lesion[23] (Table 1). Few of the clinical features are disease-specific[2]. The most common features of MS are paralysis, sensory disturbances, lack of coordination, and visual impairment[24].

Classification of MS clinical forms is not straightforward due to the heterogeneity of the disease course and symptomatology. Therefore, these concepts are revised based on new evidence and increased understanding of MS and its pathology. The standardized definitions of the clinical courses of MS were proposed in 1996. The updated classification of MS clinical forms distinguishes three clinical patterns.

The majority of patients who develop MS begin with a single episode, termed as clinically isolated syndrome (CIS)[25], with clinical symptoms typical for a demyelinating event[26]. Not all patients have subsequent disease activity after a single CIS, but risk of another episode is increased if white matter lesions are detected by MRI[27] and so fulfilling the diagnostic criteria for *relapsing-remitting multiple sclerosis (RRMS)*[2].

Table 1. Symptoms and signs of multiple sclerosis by site

LESION SITE	SYMPTOMS	SIGNS
Cerebrum	Cognitive impairment	Deficits in attention, reasoning, and executive function (early); dementia (late)
	Hemisensory and motor	Upper motor neuron signs
	Affective (mainly depression)	
	Epilepsy (rare)	
	Focal cortical deficits (rare)	
Optic nerve	Unilateral painful loss of vision	Scotoma, reduced visual acuity, colour vision, and relative afferent pupillary defect
Cerebellum and cerebellar pathways	Tremor	Postural and action tremor, dysarthria
	Clumsiness and poor balance	Limb incoordination and gait ataxia
Brainstem	Diplopia, oscillopsia	Nystagmus, internuclear and other complex ophthalmoplegias
	Vertigo	
	Impaired swallowing	Dysarthria
	Impaired speech and emotional lability	Pseudobulbar palsy
	Paroxysmal symptoms	
Spinal cord	Weakness	Upper motor neuron signs
	Stiffness and painful spasms	Spasticity
	Bladder dysfunction	
	Erectile impotence	
	Constipation	
Other	Pain	
	Fatigue	
	Temperature sensitivity and exercise intolerance	

Source: *Compston, A. & Coles, A., 2008*[2]

RRMS is the most common form, affecting approximately 85% of patients. It is characterized by recurring bouts of relapse (discrete episode of acute worsening of a given neurological function) and remission (complete or partial recovery period) (Figure 1; black line). Relapses coincide with focal CNS inflammation and demyelination that are typically discernible as white matter lesions[28]. Eventually, improvement during each remission wanes as disability accumulates, and between 15 and 30% of patients go on to develop *secondary progressive multiple sclerosis (SPMS)* over a long-term follow up[25].

In SPMS inflammatory lesions are no longer characteristic, and progressive neurological decline is instead accompanied by CNS atrophy; that is, decreased brain volume and increased axonal loss (Figure 1; red solid and dashed lines, respectively)[28].

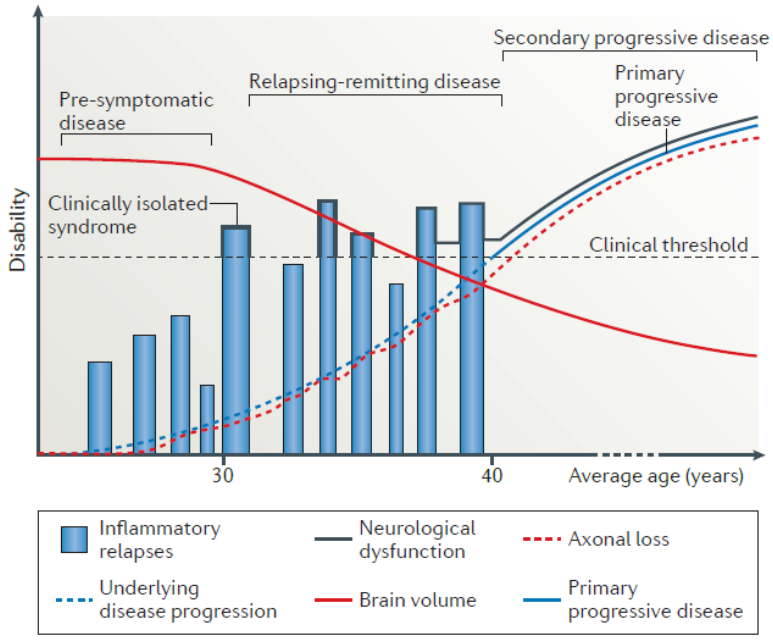


Figure 1. Evolution of multiple sclerosis. Source: Dendrou, C. A., Fugger, L. & Friese, M. A., 2015[28]

Primary progressive multiple sclerosis (PPMS) affects about 10-15% of patients with MS, which features progressive decline from the outset and an absence of relapses (Figure 1; blue line)[28].

Moreover, MRI findings that are consistent with MS have been observed in healthy people who underwent scanning for other purposes, and clinical MS develops in up to 50% of people with this so-called radiologically isolated syndrome, sometimes with a primary progressive course[29,30].

The diagnosis of MS is based on the integration of clinical, imaging, and laboratory findings[25]. Since 2000, MRI has been the key diagnostic test when patients present with a clinical syndrome that is suggestive of MS, and the most recent criteria, McDonald criteria[31].

3. EPIDEMIOLOGY

MS is one of the world's most common neurologic disorders, and in many countries it is the leading cause of nontraumatic neurologic disability in young adults[32]. It is usually diagnosed between 20 and 40 years old, with a later onset of disease for PPMS than for RRMS[33].

According to the Atlas of MS, the most extensive worldwide study of the epidemiology of MS, completed by MS International Federation (MSIF) in 2013, the estimated number of people with MS had increased from 2.1 million in 2008 to 2.3 million in 2013[32]. The disease has a heterogeneous prevalence worldwide: it is highest in North America (140/100,000 population) and Europe (108/100,000), and lowest in East Asia (2.2/100,000 population) and sub-Saharan Africa (2.1/100,000)[34]. Nevertheless, notwithstanding difficulties in surveillance, MS is almost nonexistent in black Africans[35] (Figure 2).

MS occurs more frequently in women than men (3:1 ratio)[19]. Remarkably, the prevalence of MS appears to have steadily increased over the past century, primarily in women leading to higher female to male sex ratios of MS[36,37] and in regions previously considered low incidence[38,39] . There is evidence that women generally have an earlier onset of disease, they have a slightly lower prevalence of *PPMS* and show in general less progression of disability than men[40].

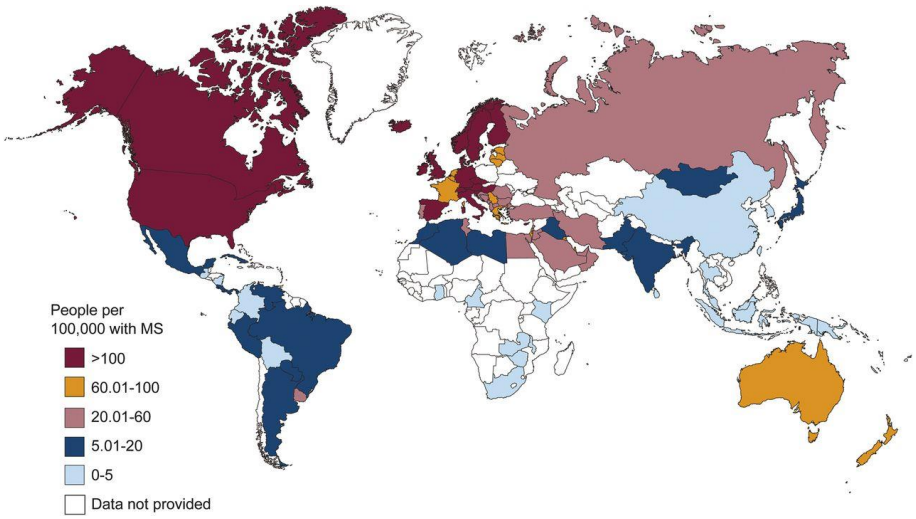


Figure 2. Global prevalence of MS. Source: Browne P *et al*, 2014[32]

4. PATHOGENESIS OF MULTIPLE SCLEROSIS

Although some suggest that MS is primarily a neurodegenerative process with secondary immune involvement[41], overwhelming evidence points to immune-mediated disease mechanisms because of the following findings: T and B lymphocytes are present in demyelinating lesions[42]; CNS antigen-specific immune responses are detected in the peripheral blood of patients with MS[43]; the disease can be suppressed by immunomodulatory therapeutics[44,45]; and genetic studies point towards the adaptive immune system[25], particularly its $CD4^+$ T cell component.

Both the innate and adaptive immune systems are known to influence the pathogenesis of MS[25]. Being the earliest defense against pathogens, the innate immune system fights against infections and protects against self or innocuous antigens[46]. Monocytes, macrophages, microglia, dendritic cells (DCs) and natural killer (NK) cells, among others, compose the innate immune system which do not have immunological memory[46]. While the innate immune system delivers an early and rapid response, the adaptive immune system develops a response that is highly specific to the encountered infectious agents and enhanced with subsequent pathogen exposures. B and T lymphocytes mediate the adaptive immune responses and provide long-term protection. T and B cells have specific antigen receptors and clonally expand after contact with antigen. B cells differentiate into plasma cells that secrete immunoglobulins[43]. The main T cell subsets are $CD4^+$ T lymphocytes, also called helper T cells that coordinate numerous immune responses, and

$CD8^+$ T lymphocytes, also known as cytotoxic T cells that play key roles in controlling intracellular pathogens but also neoplastic cells. Antigen specificity is determined by the human leukocyte antigen (HLA), which refers in particular to the major histocompatibility complex (MHC) in humans, a group of cell surface proteins that is essential for the recognition of self-cells and non-self-cells and activation of the acquired immune system to eliminate the non-self-components[47,48]. Via T cell receptor (TCR) complex, $CD4^+$ T lymphocytes recognize antigens that are presented by the MHC class II molecules, whereas $CD8^+$ T lymphocytes recognize antigens presented by MHC class I molecules[49].

Without a known predominant exogenous risk factor, it is an open question whether MS is triggered in the periphery or in the CNS. In the CNS-extrinsic model, autoreactive T cells that are activated at peripheral sites, potentially through molecular mimicry[50–52], bystander activation or the co-expression of TCRs with different specificities[53], traffic to the CNS along with activated B cells and monocytes. Once in the CNS, autoreactive T cells are locally reactivated by antigen-presenting cells (APCs) and recruit additional T cells and macrophages to establish the inflammatory lesion[25]. MS lesions contain $CD8^+$ T cells, which are mostly found at the edges of lesions, and $CD4^+$ T cells, which are found deep in the lesions[54,55]. These cells cause myelin loss, oligodendrocyte destruction, and axonal damage, leading to neurologic dysfunction. In parallel, immune-modulatory networks are triggered to limit inflammation and to initiate repair, which results in at least partial remyelination and is associated with clinical remission[56]. The extrinsic model is consistent with the method used to induce EAE in animals: emulsified CNS antigen is administered along with immune stimulants, resulting in the generation of pathogenic $CD4^+$ T helper 1 (Th1) cells and Th17 cells in the draining lymph nodes. These cells then enter the circulation and ultimately exert their effector functions within the CNS, having crossed the BBB or the blood–cerebrospinal fluid (CSF) barrier at the choroid plexus[28].

Alternatively, CNS-intrinsic events may trigger disease development, with the infiltration of autoreactive lymphocytes occurring as a secondary phenomenon. It is unclear what these CNS-intrinsic events might be, although postulated mechanisms include inflammatory responses to an as yet unknown CNS viral infection, a hypothesis based partly on the emerging appreciation of CNS immune surveillance[57,58] or to processes

leading to primary neurodegeneration, similar to those that have been implicated in Alzheimer disease or Parkinson disease[59].

4.1. THE MAJOR IMMUNE PLAYERS

4.1.1. Effector T cells in MS

The presence of T cells within CNS lesions is detectable in the early stages of MS[60], and the long-appreciated HLA associations with the disease are thought to reflect the presentation of specific CNS autoantigens to autoreactive T cells. As demyelination is a key feature of MS neuropathology, myelin protein-derived antigens have been hypothesized to be the main autoreactive targets[28]. Most autoreactive T cells are deleted by the negative selection in the thymus, resulting in central tolerance. This is further supported by regulatory mechanisms outside of primary lymphoid tissues, which are collectively known as peripheral tolerance. Breakdown of mechanisms that maintain immune tolerance to self tissues can lead to autoimmunity[61].

$CD4^+$ Th1 and Th17 cells are believed to be the pathogenic initiators of MS. In the late 1980s, $CD4^+$ T cells were first separated into the functionally distinct IFNγ producing Th1 cells, which clear extracellular pathogens, and the IL-4 producing Th2 cells, which mediate allergic responses[62]. Subsequent studies then identified $CD4^+$ Th17 cells, which play a central role in autoimmunity[63]. A pathogenic role for IFNγ in MS is supported by a trial in 1987 in which the administration of IFNγ to MS patients exacerbated their disease[64]. A role for IL-17 is supported by a double blind trial in which the administration of an anti-IL-17 compound to MS patients reduced lesion formation, as assessed by MRI[65]. Thus, both IL-17 and IFNγ appear to play a role in the human disease, which is consistent with adoptive transfer EAE experiments that have shown that both Th1 and Th17 cells can independently induce CNS autoimmunity with distinct patterns of EAE histology[56,66,67]. Consequently, skewing of T cell differentiation away from these cells subsets and towards a Th2 cell phenotype has been a prominent therapeutic concept and is considered to be a mechanism of action of the first-line, disease-modifying therapies IFNβ[68], glatiramer acetate[69] and dimethyl fumarate[28,70].

Although much research has focused on the encephalitogenic role of $CD4^+$ T cells in MS, it appears that $CD8^+$ T cells also play a significant role in human MS[56]. $CD8^+$ T cells are found in higher frequency than $CD4^+$ T cells in the white matter and in grey matter cortical demyelinating lesions, and their numbers closely correlate with axonal damage[5].

$CD8^+$ T cells recognize peptides of endogenous intracelular proteins presented in the context of MHC class I molecules and kill cells via a cell-contact-mediated process involving the activities of granzyme A and granzyme B[71]. Microglial cells have the capacity to cross present exogenous antigens, potentially leading to the high frequency of myelin-reactive $CD8^+$ T cells that has been reported in patients with MS[72]. These myelin-reactive $CD8^+$ T cells secrete IFNγ and kill cells that endogenously produced myelin. The cytotoxic function of $CD8^+$ T cells might play a central role in axonal damage, as their intracellular lytic granules are directionally positioned toward nearby axons in immunohistochemical analysis. Axonal injury has been correlated with the presence of lesional $CD8^+$ T cells in close apposition to neurons[73]. Pathogenic $CD8^+$ T cells might also contribute to pathology by secreting IFNγ and IL-17[74]. These IFNγ-, IL-17-, and granzyme B-producing $CD8^+$ T cells also potentially undergo enhanced endothelial transmigration in a BBB model with human cells and in mouse models[75,76]. Thus, $CD8^+$ T cells may not only cause oligodendrocyte death and neuronal damage, but they may also potentiate IFNγ- and IL-17-mediated pathology once inside the CNS.

4.1.2. Effector B cells in MS

Although MS is considered to be a T-cell-mediated disease, the dramatic effects produced by anti-CD20 therapy (rituximab and ocrelizumab) in MS demonstrate a central role for B cells in its pathogenesis[77]. Compared with T cells, infiltrating B cell numbers in the CNS vary more throughout disease progression. One of the classic findings in MS is the presence of oligoclonal bands in the CSF, which arise from the intrathecal synthesis of clonal IgG and are present in more than 95% of patients with MS[77].

Clonally expanded B cells are found in the brain parenchyma, meninges, and CSF of MS patients[78] and are present in the CNS at greater frequency earlier in the disease[79]. Increased B cell frequency in the CSF correlates with a rapider disease progression[80].

Numbers of antibody-secreting plasma cells are increased with age in patients with primary or secondary progressive MS[5]. Beyond their potential ability to produce autoantibodies, B cells in the CNS could act in MS by secreting chemokines/cytokines and by presenting antigen to T cells[56].

B cell depleting anti-CD20 antibodies (rituximab) have become a powerful therapy of early MS substantially reducing relapse frequency, radiologically determined lesional load and improving neurological deficiencies[81]. This transmembrane protein is expressed in different stages of B-cell differentiation, from pre-B cells to naive and memory B cells, but is absent in earlier stages (pro-B) and plasma cells[82,83]. Therefore, the role of B cells as APCs and secretors of cytokines and chemokines is very likely more important than their capacity to produce antibodies in the pathogenesis of MS.

4.1.3. Regulatory T cells in MS

Subsets of T cells that modulate immune activation and control the development of autoimmunity have been identified[84]. Two important subsets of $CD4^+$ regulatory T cells have been studied in the context of MS. Pathogenesis of the disease may be due to defects in their functions. These are (i) FOXP3 expressing $CD4^+$ regulatory T (TReg) cells[85] that contribute to their capacity to suppress T cell proliferation *in vitro* via a cell-contact-mediated mechanism[86] and (ii) IL-10 producing T regulatory type 1 (TR1) cells[87] that inhibit cell proliferation primarily via the secretion of IL-10[88].

TRegs, which make up less than 4% of circulating $CD4^+$ T cells, are referred to as 'professional' suppressor cells[89,90], selected to recognize self-antigens[91,92] and activated by self-antigens. Several studies have demonstrated that inflammation can adversely affect TReg suppression[93], as high numbers of non-suppressing TRegs have been found to co-exist with proinflammatory T cells in autoimmune target tissues at peak disease in EAE in the CNS[94].

TRegs from MS patients have a reduced suppressive capacity as measured by the co-culture of MS-derived TRegs and effector T cells[95–97]. Although increased TRegs function has been proposed as a therapy for MS, such approaches do not address whether the

TReg defect observed in MS is related to the resistance of effector T cells to TRegs suppression or to an inherent defect in the TRegs themselves. Recent studies of TRegs and effector T cells isolated from patients with a number of autoimmune diseases, including MS, have demonstrated that patient-derived effector T cells are, in fact, resistant to TReg-mediated suppression[98,99]. Thus, published studies suggest that in MS there exists both a defect in TRegs and a resistance to TRegs suppression by effector T cells[56].

Inflammatory molecules expressed in autoimmune diseases can negatively affect TRegs suppression. The pro-inflammatory cytokine IL-6, which exacerbates EAE, can alter $CD4^+$ T cell differentiation and promote the generation of Th17 cells at the expense of TRegs[100,101].

TR1 cells were first described in the mouse model of colitis and were shown to produce IL-10[102]. These TR1 cells secreted high amounts of IL-10 and killed myeloid APCs via a granzyme B-mediated mechanism. Although not as extensively studied as TRegs, TR1 cells from patients with MS have also been found to be dysfunctional. By stimulating $CD4^+$ T cells through CD46, which strongly induces IL-10, it was found that MS $CD4^+$ T cells express less IL-10 than those from healthy individuals[103]. Numerous studies have demonstrated the importance of IL-10 in ameliorating murine EAE[104]. In MS, the capacity of $CD4^+$ T cells to secrete IL-10 is associated with decreased disease activity as the expression of IL-10 and CD46 is enhanced in patients who respond to IFNβ therapy compared to cells from patients who did not respond[105].

4.1.4. Dendritic cells in MS

DCs are APCs that play an important role in promoting the activation and differentiation of T cells[106]. Human dendritic cells can be divided into myeloid and plasmacytoid cell types[107]. Dendritic cells can also be derived from monocytes under inflammatory conditions[108]. The interaction of DCs with $CD4^+$ T cells is crucial in determining T cell differentiation into either effector (Th1, Th2, Th9, and Th17) or regulatory (TRegs and TR1) T cells thus shaping the adaptive response[109–111]. DCs are also important for the activation of $CD8^+$ T cells and can induce either cytotoxic or regulatory NK cells[112]. DCs

were also shown to be the most effective APC in the CNS in mediating epitope spreading in different EAE models[113,114]. Moreover, manipulating DC function alters the T-cell repertoire, thus affecting the disease course. In MS patients, DCs are found in MS lesions[113,115] and DCs isolated from the peripheral blood of MS patients exhibit an altered phenotype with decreased or delayed expression of the activation markers in addition to their altered functionality in terms of T-cell proliferation and generation of regulatory T cells[116]. Moreover, changes in the cytokine profile and expression of costimulatory and inhibitory molecules were reported in SPMS versus RRMS[117–120].

4.1.5. Monocytes, macrophages and microglia in MS

The mononuclear phagocyte system comprises a population of cells derived from progenitor cells in the bone marrow, which differentiate to form blood circulating monocytes and then upon activation enter tissues to become resident tissue macrophages[121]. Microglial cells are considered resident macrophages of the CNS, which regulate local innate and adaptive immune responses in the CNS tissue[106].

Under basal conditions, microglia are highly dynamic and constantly patrol the CNS microenvironment[122]. Microglial activation is a hallmark of demyelinating lesions in EAE[123,124]. Microglial cells are rapidly activated in response to injury, neurodegeneration, infection, tumors and inflammation[46], which leads to an increase in their size, upregulated expression of several molecules, and initiation of effector functions, including antigen presentation, chemoattraction, and phagocytosis of debris or toxic factors[108]. Activated microglia can be injurious to the CNS, damaging oligodendrocytes and neurons by producing free radicals and proinflammatory cytokines, such as TNF[125], or provide benefits such as neuroprotection, mobilization of neural precursors, remyelination and axonal regeneration[126–129]. During inflammatory disorders, such as MS, monocytes are repeatedly recruited from the periphery, thereby reinforcing the local inflammatory reaction within the CNS. In MS, macrophages act as APCs perpetuating epitope spreading upon T cell traffic on brain through BBB[130].

4.1.6. Natural killer cells in MS

NK cells contribute to both effector and regulatory functions of innate immunity[106]. NK cells are innate immune cells involved in early host defense against infection and tumor transformation. NK cells also have regulatory properties and suppress immune responses via their capacity to kill activated, and potentially pathogenic, $CD4^+$ T cells[56]. They can also directly lyse oligodendrocytes, astrocytes, and microglIa[131]. NK cells secrete both pro-inflammatory (IFNγ, TNFα) and anti-inflammatory (IL-4, IL-10) cytokines and have been suggested to play dual roles in disease[132]. It is unclear whether the specific two types of NK cells, which are commonly defined as $CD56^{dim}CD16^{hi}$, representing 90% of NK cells in peripheral blood, highest cytotoxic activity and much lower frequency in tissue, and $CD56^{bright}$, which primarily secrete cytokines and acquire cytotoxic activity over time, predominate in tissues, play unique roles in autoimmunity[133,134].

Most lines of evidence suggest that NK cells play an immunoregulatory role in MS even though NK cells have been found in the demyelinating lesions of MS patients[135]. $CD56^{bright}$ NK cells are increased by immunomodulatory and immunosuppressant therapies, increases in NK frequency correlate with treatment response[136], reduced NK frequency has been associated with relapse[137], and *in vitro* NK functional activity increases at times of remission[138]. Daclizumab, which reduces MS relapses, is a humanized antibody against the IL-2 receptor that expands $CD56^{bright}$ NK cells[136]. A similar increase in $CD56^{bright}$ NK cell frequency has also been described in MS patients treated with IFNβ[139].

5. CAUSES OF MULTIPLE SCLEROSIS

Despite major research efforts in the past few decades, the exact cause of MS, and whether this varies from one patient to the next, still remains elusive[28]. MS is believed to have autoimmune etiology. Its pathogenesis is best explained by a multifactorial model that incorporates interactions between genetic, epigenetic, and infectious, nutritional, climatic, or other environmental influences[35]. All these factors can influence adaptive and/or innate immunity, which is thought to be the main pathway modulated by MS genetic risk variants[140].

5.1. LIFESTYLE AND ENVIRONMENTAL RISK FACTORS

It is recognized that environmental and lifestyle factors interact with genetics to influence both MS susceptibility and the course of the disease[56].

5.1.1. Latitude gradient and migration

The earliest clues to environmental influences came from studies into the effect of latitude and migration. It has long been recognized that there is a latitudinal gradient in MS[56]; a gradual decrease in incidence and prevalence of MS from north to south in the northern hemisphere, and in the opposite direction in the southern hemisphere[140].

Migration studies show that the risk of MS depends on the age at which an individual migrates: the risk of MS for those who migrate from a low-risk country to a high-risk country before adolescence is similar to that of those who are born and reside in the high-risk country[141,142].

5.1.2. Vitamin D and sunlight exposure

Vitamin D levels and exposure to ultraviolet radiation (UVR) are two environmental factors known to be negatively associated with MS susceptibility that might also help to explain the latitude effect[56].

Conversion of vitamin D to its active metabolite is dependent on URV making it very difficult to distinguish between the effects of UVR and vitamin D. UVR also protects against MS, probably both through its effects on vitamin D and through independent effects on the immune system[143].

Increased vitamin D levels, especially before the age of 20, are associated with a reduced risk of MS in later life [144], and later data on supplementation and sun exposure have supported the role of vitamin D in reducing the risk of MS[145,146]. In addition, a diet rich in fatty fish that contains vitamin D also reduces MS risk in individuals with low sunlight exposure[147]. Recently, high vitamin D levels were shown to correlate with decreased axonal damage[148] and during an IFNβ trial, higher levels of the vitamin were associated

with a lower MRI activity[149]. Vitamin D has broad effects on the immune system, including the suppression of both B cell and T cell proliferation, skewing T cells away from inflammatory responses and toward regulatory T cell responses, and promoting tolerogenic monocyte and dendritic cell phenotypes[150]. Vitamin D levels do not only influence risk for MS, but also that for other autoimmune diseases, including rheumatoid arthritis (RA), type 1 diabetes mellitus (T1DM) and systemic lupus erythematosus (SLE)[56].

5.1.3. Epstein-Barr virus infection

Many infectious agents have been proposed to have a role in MS, but one of the most interesting candidate is EBV[140]. Nearly all individuals with MS (>99%) have been found to have been infected with EBV compared with approximately 94% of age-matched controls[151]. There is an increased risk of MS in individuals who have infectious mononucleosis during adolescence and adulthood, especially so in individuals with elevated antibody titers against EBV nuclear antigen 1 (EBNA1)[56]. It is not clear how EBV infection could contribute to MS. It could have a specific effect, either through a mechanism such as molecular mimicry, or through general immune effects on B cells or other immune regulatory elements[56].

5.1.4. Smoking and use of oral tobacco

Cigarette smoking is a well-recognized environmental factor that is linked to both MS risk and disease activity[152]. The risk increases with duration and intensity, and is stronger in men than in women[25]. Besides the risk of developing MS, history of smoking is associated with a worse prognosis in MS[153], and aggravated disease course[153,154]. In addition, continued smoking is also associated with the increased risk of developing neutralizing antibodies against biologics used in treatment of MS[155,156]. Cigarette smoking increases the risk of autoimmune diseases, including RA[157] and SLE[158]. Smoking provokes lung inflammation and promotes pro-inflammatory pathways[159]. If CNS auto-antigenic cells are present in the lung, such cells might be activated to attack the CNS[160]. Smoking displays a considerable interaction with MS-associated HLA risk alleles[140]. In the Scandinavian population, having the class II HLA-DRB1*15:01 MS risk allele confers an odds ratio (OR) of ~3, and lack of HLA-A*02 confers an OR of ~1.8, resulting in a combined

OR of ~5 among non-smokers; however, in smokers, the combined OR is much higher, at ~14[161].

5.1.5. Obesity

Growing evidence strongly supports a role for obesity in the risk of MS. Obesity in early life is associated with a two-fold increase in risk in men and women[25]. For adult-onset MS the critical period regarding obesity seems to be during adolescence[162,163]. Obesity later in life is not associated with increased MS risk. The molecular pathways behind the association between adolescent obesity and MS may involve fat-related chronic inflammation. By promoting Th1 responses and decreasing the number of TReg, obesity may increase the risk of recruitment of autoreactive $CD4^+$ T cells that target CNS autoantigens[164]. Furthermore, obesity also leads to decreased bioavailability of vitamin D[165]. Also in the case of obesity, there is an interaction with MS risk HLA genes in that HLA-DRB1*15 positive and HLA-A*02 negative individuals with adolescent obesity display an OR of ~14[140,166].

5.2. THE GENETICS OF MULTIPLE SCLEROSIS

5.2.1. Evidence for a genetic component

The earliest evidence supporting a genetic influence on MS susceptibility was derived from observations of familial aggregation and differences in MS risk among ethnic groups residing in the same geographical regions[167]. People with an affected first-degree relative have a 2-4% risk of MS (as compared with approximately 0.1% risk in the general population)[9]. Equally important, siblings of an affected individual are at least seven times more likely to develop MS than the general population[168,169]. Studies that were carried out in Canada on adoptees, half-siblings and spouses seem to confirm that genetics is primarily responsible for the co-incidence of MS within families[167]. Twin studies from several populations have shown that monozygotic twins have a higher concordance rate (20–30%) compared with dizygotic twins (2–5%), providing strong support for a significant but complex genetic etiology in MS[170,171]. Taken together, the familial

recurrence and twin concordance rates for MS do not support the presence of a Mendelian trait[172], but underscore a polygenic context of genetic risk.

5.2.2. The Human Leukocyte Antigen Locus in Multiple Sclerosis

The recognition during the early 1970s of the influence of specific HLA variants within the MHC gene complex (chromosome 6p21) in MS susceptibility represented the first empirical demonstration linking disease risk with common genetic variation[35]. This 4-megabase region contains approximately 165 closely linked genes. About half of these genes have important roles in the regulation of the immune system, and include the six classical transplantation HLA genes (the class I genes HLA-A, HLA-B, and HLA-C, and the class II genes HLA-DPB1, HLA-DQB1, and HLA-DRB1)[173]. The extreme polymorphism, with more than 14,000 alleles identified to date, and extensive linkage disequilibrium (LD) characterize this gene-dense region[174]. Carriers of the HLA DRB1*15:01 allele are about three times more likely to develop MS than non-carriers[175]. The class I variants HLA-A*02 (OR~0.6) and HLA-B*44 are associated with protection from disease[1]. The presence of DRB1*15:01 and absence of HLA-A*02 has a combined OR~5. Since individuals not carrying the HLA risk alleles also can develop MS, it was early on recognized that genes outside the MHC complex may contribute partially to the genetic risk. Despite decades of candidate-gene-based efforts to find such genes, little progress was made in the identification of relevant, genuinely associated risk alleles outside the MHC before the advent of GWAS[1]. Although largely negative, these studies strongly supported the notion that MS is not caused by a small amount of mutations of large effect but is likely to be due to many small risk effects spread across the genome[176].

5.2.3. Genome-Wide Association Studies

In the early 2000s, the introduction of chip-based technologies with the capacity to genotype simultaneously hundreds of thousands of single-nucleotides polymorphism (SNP) allowed the development of a new analytical methodology known as GWAS, a hypothesis-free method in which SNP spaced across the entire genome are screened for association with a particular trait in case–control datasets composed of genetically unrelated individuals[177]. Compared to classic linkage studies that rely on extended

families, the possibility to test unrelated individuals allows collecting much larger datasets, substantially increasing the statistical power of gene-discovery studies. These studies have demonstrated that the common disease–common variant hypothesis of human diseases[178] is broadly true, where disease risk is driven by many common variants (those with minor allele frequency >5%), each of which explains a small fraction of the risk in a population[176].

Several GWAS and large-scale targeted studies followed in the ensuing 10 years showed unequivocal statistical evidence for the association of 200 autosomal susceptibility variants outside the MHC, one chromosome X variant, and 32 independent associations within the extended MHC[179] (Figure 3). This data supports a polygenic model of heritability for MS in which risk is determined by a single moderate-effect allele (OR roughly 3) and many much-smaller-effect alleles (OR <1.5)[1].

Altogether, the observed heritability explained by all associations does not amount to more than 20–30%, of which a substantial proportion can be assigned to the MHC region[35]. The remaining fraction of the risk commonly known as 'missing heritability' is likely due to either still unknown common variants characterized by much smaller effects, below the detection limits of the GWAS conducted so far, as yet undiscovered very rare variants with large ORs, or epistatic interactions between gene loci[19]. Also, likewise gene by environment interactions, cis/trans-regulators of allelic expression, population and/or disease heterogeneity, and hidden epigenetic effects may all contribute to the missing heritability[19].

Nearly all the identified associations map to noncoding regions of the genome, either in intragenic or intergenic regions. Naturally, those mapping to intragenic regions were explored first (*IL7RA, TNFSF10,* etc.) as their relationship to function is thought to relate to splicing, mRNA stability, or promoter activity of the gene in question. However, most associations lie in genomic regions distant from any known gene. In those cases, the most likely explanation is that they modify regulatory sites (open chromatin, histone modifications, enhancers, and repressors) and indirectly influence gene activity[35].

In the most recent MS GWAS, careful pathway, transcriptomic and epigenetic enrichment analyses suggest T-cell biology is a major feature of the disease, but also highlight the involvement of many other components of both adaptive and innate immunity in pathogenesis[176], a finding consistent with the notion that autoimmune mechanisms are paramount in the development of clinical MS[9].

5.2.4. Sharing with other autoimmune diseases

One salient characteristic of the MS susceptibility map is the large proportion of the associations (either the exact same variant, or within the same gene or locus) that are shared with other autoimmune diseases, such as Crohn disease (CD), T1DM, RA, and SLE, among others[35]. Interestingly, some of these shared associations confer risk for one disease but protection for another. One example is rs744166 located in an intron of the *STAT3* gene is associated with increased risk in MS[180] and decreased risk in CD[181]. Allelic heterogeneity was also reported in the *IL2RA* locus, where rs11594656 was shown to be associated with susceptibility to MS and protection against T1DM. Furthermore, rs41295061 only conferred susceptibility to T1DM but not MS[182]. A few genetic associations are shared with other neurodegenerative diseases, such as Alzheimer's and Parkinson's diseases[56].

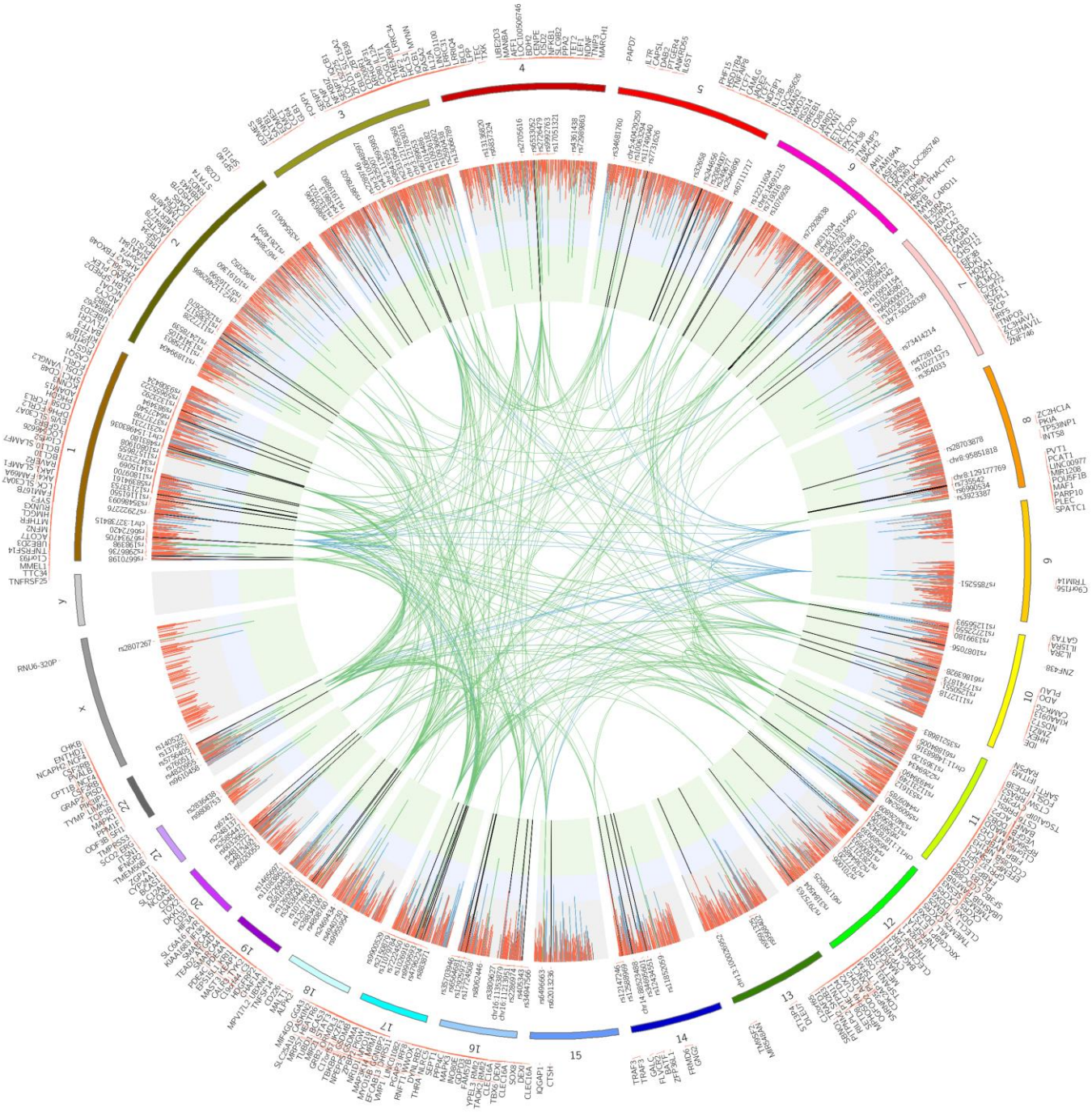


Figure 3. The genetic map of multiple sclerosis. Source: Patsopoulos, N. *et al.* 2017[179]

5.2.5. From Gene to Function

One of the most difficult steps in any GWAS is to infer the biologically relevant consequences of each statistical association[35]. First reason for this shortcoming consists in the pervasive LD along the human genome, which hinders the identification of true causative variants. LD refers to the tendency of genetic loci in physical proximity to segregate together during meiosis, leading DNA to be inherited in large blocks through generations. This peculiarity of genome architecture substantially impairs GWAS resolution since SNPs in the same LD block are inherited together as well. Thus, statistically significant GWAS risk variants are usually proxy for the real causative variants, which can be located up to several megabases away within the same LD block[19]. The second challenge is that the majority of MS risk variants appear to localise to gene regulatory regions, rather than coding sequence[183], and specifically to enhancer elements active in stimulated immune cell subsets[184]. Furthermore, a subset of associated SNPs have been shown to be expression quantitative trait loci (eQTL) in different cells of the immune system, supporting the idea that much risk is due to changes in gene regulation[176]. For example, associated variants that are eQTLs for genes such as *IFITM3, CD37,* and *CD6* in $CD4^+$ T cells strongly suggest that they have an effect in modulating adaptive immune responses. Additionally, eQTL variants in or near *CLECL1, RGS1,* and *MERTK* in $CD14^+$ monocytes suggest that the innate immune system is also involved[35]. Notably, variants in genes expressed in the CNS are less common in these studies, thus supporting a model in which genetic susceptibility to MS is mediated primarily by sustained dysregulation of immune responses over several decades before clinical symptoms appear[35].

A variety of experimental systems have been employed to study the biological functions associated with MS risk variants, ranging from patients-derived primary blood cells to animal models of disease. The first such demonstration in MS was the discovery that rs6897932, located within the exon 6, influenced the function of the gene interleukin 7 receptor alpha chain (*IL7R*)[35]. Specifically, the risk allele disrupts a splicing acceptor site and results in transcriptional skipping of exon 6 of the gene, thus altering the relative amounts of soluble and membrane-bound isoforms of the gene[185]. Recent evidence has

shown that the RNA helicase DEAD box polypeptide 39B (*DDX39B*) is also a potent activator of *IL7R* exon 6 and consequently a repressor of soluble IL7R. Moreover, the rs2523506 located in 5'UTR of DDX39B reduces translation of DDX39B mRNA and increases MS risk[186]. This example provides functional evidence that two associated loci can work together to confer susceptibility to MS[35].

A similar effect was described for the intronic rs2104286 in the gene interleukin 2 receptor alpha (IL2RA)[19]. This risk variant was also found to alter the soluble/membrane-bound ratio of IL2RA protein by driving the expression of higher levels of its soluble form[182].

Another well-characterized example is the intronic rs1800693 in the gene encoding tumor necrosis factor receptor super family 1A (*TNFRSF1A*)[35]. In this case, the risk allele promotes the skipping of exon 6 with the production of a novel soluble form of the TNF receptor which is able to inhibit TNF signaling inside the cells, mirroring somehow, the exacerbating effects of TNF-blocking drugs on MS course[187].

Regarding the MS-associated tyrosine kinase 2 (*TYK2*) rs34536443 located in the exon 21, protective allele infers decreased tyrosine kinase 2 activity, and this reduction of activity is associated with a shift in the cytokine profile favouring the secretion of Th2 cytokines. These findings suggest that the rs34536443 variant effect on MS susceptibility might be mediated by deviating T lymphocyte differentiation toward a Th2 phenotype[188].

More recently, it was reported that the exonic rs11808092 in the ecotropic viral integration site 5 (*EVI5*) gene induces changes in superficial hydrophobicity patterns of the coiled-coil domain of EVI5 protein, which, in turn, affects the EVI5 interactome[35]. In particular, this work showed that the EVI5 protein bearing the risk allele selectively interacts with sphingosine 1-phosphate lyase (SGPL1), an enzyme important for the creation of the S1P gradient, which is relevant to the adaptive immune response and the therapeutic management of MS[189].

5.2.6. ANKRD55

An intronic variant in *ANKRD55*, rs6859219, was identified as a genetic risk factor for MS[190], and subsequently this association was validated with genome-wide significance[191]. A proxy of rs6859219 (r^2 = 0.9, D'= 1), rs71624119, also was associated with MS in a comprehensive fine-mapping of autoinmune disease related genomic regions[192]. rs6859219 in *ANKRD55* emerged originally from a meta-analysis of GWAS on RA[193]. Other studies have linked *ANKRD55* to several autoimmune diseases, including RA[194,195], CD[196,197], T1DM[198], juvenile idiopathic arthritis[199], celiac disease[200] and inflammatory myopathies (polymyositis and dermatomyositis)[201,202], as well as post-traumatic stress disorder[203], Alzheimer's disease (cognitive decline)[204], and type 2 diabetes[205,206]. Recently, *ANKRD55* was associated with IgA levels[207]. Moreover, another variant near this gene was also linked with changes in N-glycosylation of IgG[208], which is altered in RA. Interestingly, defective N-glycosylation was also implicated in MS[209].

ANKRD55 locus, located on chromosome 5q11.2, codes for Ankyrin Repeat Domain Protein-55. The Ensembl Release 75 (http://www.ensembl.org)[210] of February 2014 includes six alternatively spliced transcript variants: four are protein coding (001, 002, 005 and 006), and the remaining two processed transcripts (007 and 008) without protein products (Figure 4). However, only Ensembl protein coding transcripts 001 (corresponding to the full-length form) and 005 are included in UniProt databases. Information about cells and tissues that express ANKRD55 protein, its subcellular localization, and its function is relatively scarce. Some gene expression databases, such as BioGPS (http://biogps.gnf.org)[211], GTEx (https://gtexportal.org)[212] and Gene Expression Atlas (http://www.ebi.ac.uk/gxa)[213] shows expression in appendix, testis, superior cervical ganglion, spinal cord, brain, megakaryocytes, eosinophils and $CD4^+$ T cells. The Human Protein Atlas (http://www.proteinatlas.org)[214] documents ANKRD55 protein expression in tonsil, lymph node, spleen and testis.

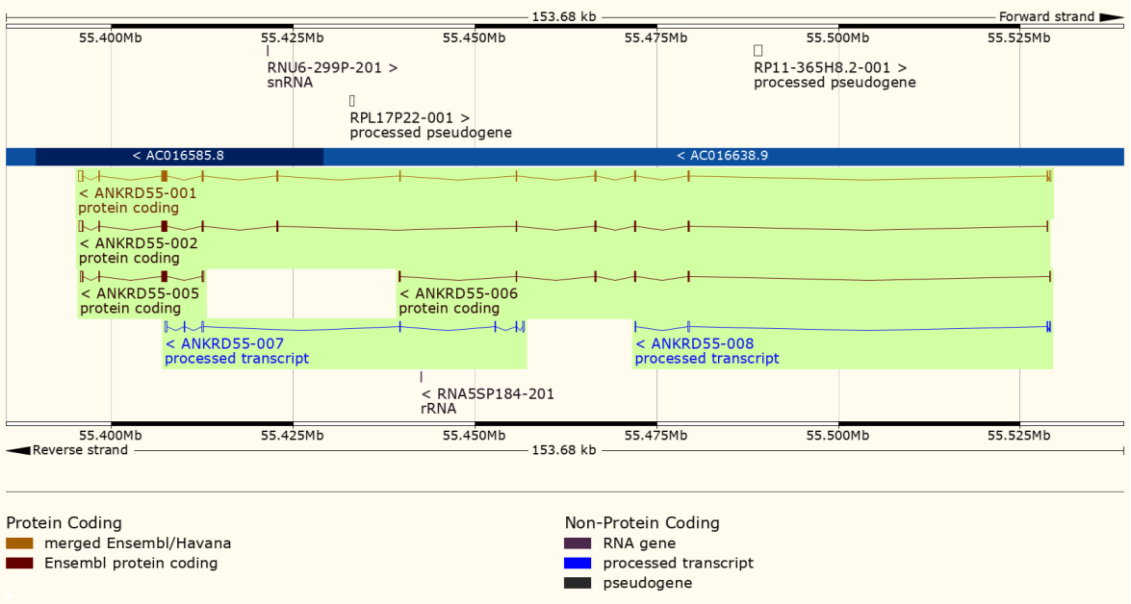


Figure 4. Representation of *ANKRD55* transcript variants. The six transcript variants described in Ensembl are shown. Gold and maroon colors depicts protein-coding transcripts, and blue color processed non protein-coding transcripts. Exons are repesented with boxes, and introns with lines. Filled boxes represent coding exons, and empty boxes UTRs. Coordinates from version GRCh37 of the human genome are shown.

The ankyrin repeat (AR), a 30-34 residue sequence motif, is one of the most common, modular, protein–protein interaction motifs in nature[215]. AR has been found in proteins with diverse functions, such as transcription, cell cycle regulation, cytoskeletal integrity, inflammatory response, development, cell–cell signaling, and various transport phenomena[216,217]. The AR is found in several biologically important proteins. The family of INK4 tumor suppressors, p15, p16, p18, and p19, as well as 53BP2, a regulator of the tumor suppressor p53, all contain ARs. The signaling protein Notch, which is involved in many cell-fate decisions during development, has seven ARs. NF-κB, a transcription factor that regulates inflammatory response is inhibited by IκB, which contains seven ARs[216].

Each AR motif exhibits a canonical helix-turn-helix conformation, in which two α-helices are arranged in an antiparallel fashion and the loop projects outward at an approximately 90° angle to facilitate the formation of hairpinlike β-sheets with neighboring loops[218–220]. Usually, a hairpinlike β-sheet structure consists of the last three residues of the preceding AR and the first four residues of the next AR. Overall, the topology of an AR motif resembles the letter L with the helices as the vertical arm and the N- and C-terminal stretches as the base[220–222]. The AR proteins that have been characterized closely resemble one another despite their different cellular functions, supporting the role of the AR as a versatile scaffold for protein–protein interactions[215]. It appears that the AR motif

is defined by its fold rather than by its function, as there is no specific sequence or structural motif that is universally recognized by AR proteins[215]. To retain this characteristic topology, some residues are well-conserved in most AR sequences while residues at other positions vary[220]. Figure 5 shows the comparison between AR consensus sequences derived from statistical analyses of AR sequences by Mosavi *et al.*[215] and the AR from ANKRD55 isoform 001 and 005.

Consensus sequence (Mosavi *et al.*, 2002)	N	G	R	T	P	L	H	L	A	A	R	N	G	H	L	E	V	V	K	L	L	L	E	A	G	A	D	V	N	A	K	D	K	
ANK 1 (26-55)	V	D	L	**T**	M	V	Y	Q	A	A	S	N	**G**	D	V	N	A	L	**T**	A	V	I	R	E	-	-	D	P	S	I	L	E	-	-
ANK 2 (60-89)	E	G	C	**T**	**P**	**L**	M	H	A	V	S	G	R	Q	A	D	T	**V**	**K**	**L**	**L**	**L**	K	M	**G**	A	N	I	N	M	-	-	-	-
ANK 3 (93-125)	Y	G	R	**T**	S	**L**	C	L	A	T	Y	L	**G**	W	L	E	G	C	V	S	**L**	**L**	R	N	**G**	A	K	H	N	I	P	D	K	-
ANK 4 (126-157)	N	G	R	L	**P**	**L**	**H**	A	A	T	A	E	P	D	M	R	L	L	**T**	**V**	**L**	**L**	Q	Q	S	N	I	S	E	I	N	H	-	-
ANK5 (161-190)	E	G	M	**T**	**P**	**L**	**H**	W	A	A	F	H	N	Q	P	Q	H	T	**Q**	M	**L**	**L**	K	K	**G**	A	D	P	T	L	-	-	-	-
ANK6 (194-223)	D	F	K	**T**	A	**L**	**H**	W	A	V	Q	S	**G**	N	R	I	L	C	**S**	I	I	**L**	S	H	H	Q	G	P	S	I	-	-	-	-
ANK7 (230-260)	S	G	K	**T**	C	V	**H**	I	A	A	A	A	**G**	F	S	D	**I**	I	**H**	E	**L**	A	R	V	-	-	-	P	E	C	N	L	Q	A
ANK8 (264-293)	D	D	R	**T**	**P**	**L**	**H**	W	A	A	A	A	**G**	K	A	E	C	**V**	**Q**	S	**L**	**L**	E	L	**G**	M	D	S	N	L	-	-	-	-
ANK9 (297-326)	N	E	S	**T**	**P**	**L**	A	Y	A	L	Y	C	**G**	H	T	A	C	**V**	**K**	**L**	**L**	S	Q	E	S	R	T	E	P	T	-	-	-	-

Figure 5. Comparison between ankyrin repeat sequences of ANKRD55 isoform 001 and 005 and consensus sequence of the ankyrin repeat proteins. ARs from isoform 001 contains nine repeats (position of each repeat defined by Uniprot is indicated in parentheses and 005 shares ANK9 repeat (9-38). The conservation degree of each aminoacid in AR is marked in different colours and is shown in consensus sequence. The structural integrity of AR motif is kept by specific residues, most of them highly conserved: 1) T-P-L-H motif (4-7 position) initiates the first α-helix, 2) V/I-V-hydrophilic-L/V-L-L motif (17-22 position) is the central piece of the second α-helix and 3) glycines (13 and 25 position) terminate and link the helices.

AIM AND OBJECTIVES

The aim of this study is to gain insight into the biological functions of ANKRD55 associated with MS and RA and increase our understanding of its role in autoimmune diseases. To address the aim of the project, we pursued the following specific objectives:

1. Identification of the main cellular source of *ANKRD55* in PBMC.
2. Determination of MS and RA-associated *ANKRD55* intronic variant as eQTL for three transcripts of *ANKRD55* in five PBMCs immune subsets.
3. Study of MS and RA-associated *ANKRD55* intronic variants, localized to a putative enhancer element, for their possible capacity to regulate the expression of *ANKRD55* transcripts and its interacting genes, *IL6ST* and *IL31RA.*
4. Analysis of endogenous and recombinant ANKRD55 subcellular localization in immune and non-immune cells.
5. Identification and validation of potential ANKRD55-interacting partners in several subcelullar compartments.

CHAPTER 2: DNA-BASED ANALYSIS

INTRODUCTION

GWAS have identified thousands of genetic variants that are associated with diseases and traits of medical importance in humans[223]. As previously mentioned, genetic variants that are associated with diseases are primarily noncoding: 'lead' GWAS SNPs are more likely to be associated with the expression levels of neighboring genes (eQTLs) than is expected by chance[224,225] and the same lead SNPs are enriched in regulatory regions marked by chromatin accessibility and modification[183,226]. Of all the genetic risk variants discovered to date, the number that impact enhancer function is estimated to far exceed the number that affect protein-coding genes or disrupt promoter function. These enhancer variants probably play an important part in common disease susceptibility by influencing transcriptional output[227]. Fine-mapping has revealed enrichment of autoimmune and inflammatory diseases associated variants in enhancer elements active in stimulated T-cell subpopulations[184]. All this suggest that the majority of disease risk is mediated by changes to gene regulation in specific cell subpopulations[228].

Gene enhancer elements are noncoding segments of DNA that play a central role in regulating transcriptional programs that control development, cell identity, and evolutionary processes[227]. The locations of enhancer elements coincide with DNase I hypersensitive regions of open chromatin flanked by nucleosomes marked with H3K4me1/2[229,230]. Enhancers can be active or repressed, and each state generally correlates with the presence of additional histone marks, such as H3K27ac and H4K16ac associated with active chromatin, or H3K27me3 and H3K9me3 associated with repressed chromatin[227,231–234] (Figure 6[235]).

eQTL enable the identification of genetic variants that influence gene expression and they can be used to quantify the impact of a possible enhancer variants within a specific cell type. These studies involve stratifying a panel of individuals based on their particular SNP genotypes and then determining whether transcript levels differ between individuals based on the specific SNP genotypes[227].

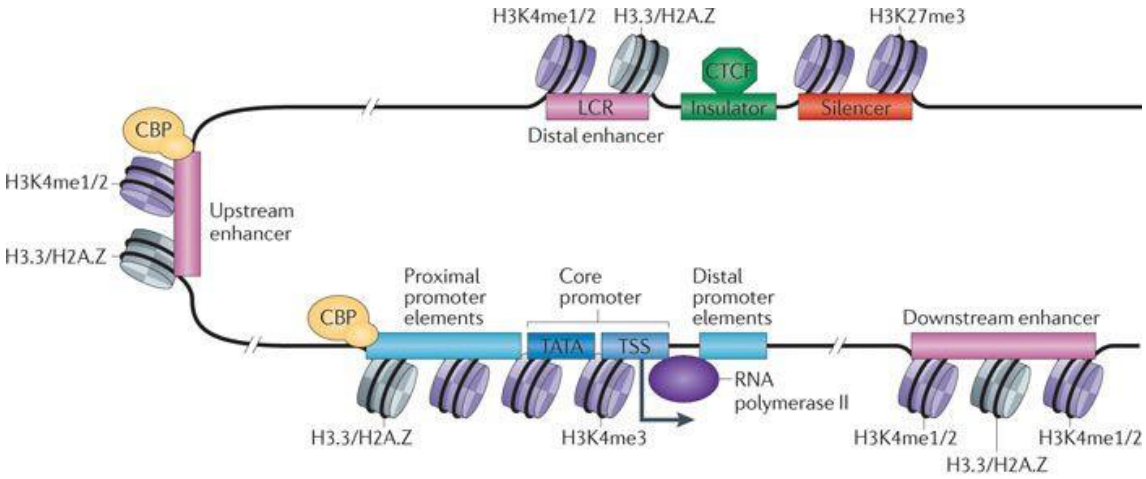


Figure 6. Transcriptional regulatory elements. The promoter is typically comprised of proximal, core and downstream elements. Transcription of a gene can be regulated by multiple enhancers that are located distantly and interspersed with silencer and insulator elements. which are bound by regulatory proteins such as CCCTC-binding factor (CTCF). Many enhancers can be defined by unique chromatin features and the binding of cyclic AMP-responsive element-binding (CREB) protein (CBP). Abbreviations: H3K4me1/2, histone H3 mono- or dimethylation at lysine 4; H3K4me3, histone H3 trimethylation at lysine 4; H3K27me3, histone H3 trimethylation at lysine 27; H3.3/H2A.Z, histone variants H3.3 and H2A.Z; LCR, locus control region; TATA, 5′-TATAAAA-3′ core DNA sequences; TSS, transcription start site. Source: Ong, C.T. & Corces, V.G., 2011[235].

CRISPR-based genome editing systems have offered a host of strategies to study non-coding genomes and evaluate the effect of DNA variants. In particular, enhancers have been a large focus of study via genome editing[236]. CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 (CRISPR-associated protein) is an RNA-mediated adaptive immune system found in bacteria and archaea, in which it protects host cells from invasion by foreign DNA elements[237]. The CRISPR/Cas9 platform requires an RNA molecule to guide the Cas9 nuclease for site specific cleavage upstream of a genomic protospacer adjacent motif (PAM) sequence[238,239]. The natural dual trans-activating CRISPR RNA (tracrRNA) – CRISPR RNA (crRNA) structure has been simplified by a synthetic single-guide RNA (sgRNA) to facilitate genome editing experiments[236,240].

In addition to using the nuclease Cas9 for editing genomic sequences, the CRISPR/Cas9 technology can be used as a sequence-specific, non-mutagenic gene regulation tool[236]. This repurposing was first demonstrated by introducing mutations into the *S. pyogenes* Cas9 in its two nuclease domains[241,242]. The resulting nuclease-deactivated Cas9 (dCas9) is unable to cleave DNA but retains the ability to specifically bind to DNA when guided by a sgRNA. dCas9 allows for direct manipulation of the transcription process without genetically altering the DNA sequence[243]. dCas9 can be fused to epigenome-modifying

protein domains to precisely modulate gene expression from gene promoters and enhancer regions[244–249]. Fusion of the KRAB (Krüppel-associated box) domain to dCas9 and subsequent targeting to a promoter or enhancer causes highly specific gene repression through the recruitment of a host of factors that deposit H3K9me3, which ultimately results in heterochromatin formation[242,248]. Conversely, fusion of the histone acetyltransferase p300 core domain to dCas9 and targeting to either promoters or enhancers induces target gene activation concomitant with the deposition of H3K27ac[249,250].

Following the identification of a functional enhancer variant, the next major challenge is to identify its target and to test the effect of the SNP on target transcript levels. Many enhancer elements are located within 100 kb of the genes that they regulate but can also be located more than a megabase away, or even on separate chromosomes. Enhancers can regulate genes or long noncoding RNAs. Most genes are regulated by more than one enhancer, and many enhancers regulate more than one target gene[227,251,252].

It is well-established that enhancers regulate gene transcription by physical interactions[253]. Higher-order three-dimensional organization of chromatin facilitates physical interactions between enhancers and their target promoters. Interactions between enhancers and their targets may occur on the same chromosome (*in cis*) or on different chromosomes (*in trans*)[227]. Chromosome conformation capture technology (Capture Hi-C) has been used for the detection of patterns of interactions between chromosomal regions[254–258]. In 2015, Capture Hi-C was used to study the chromatin interactions between associated variants for four autoimmune diseases (RA, T1DM, JIA and psoriatic arthritis) and their functional targets in human B (GM12878) and T (Jurkat) cell lines with the aim of linking disease-associated SNPs with disease-causing genes. The interactions were tested in two complementary experiments: (i) region capture targeting regions associated with disease and (ii) promoter capture obtaining promoters within 500kb of lead disease-associated SNP[254]. Both region capture and promoter capture analysis revealed that *ANKRD55* showed potential interactions with *IL6ST* and *IL31RA* (Figure 7), a possible interesting disease-associated region to follow-up[259].

As previously mentioned, *ANKRD55* has been linked to several autoimmune diseases. A non-coding variant in *ANKRD55*, rs6859219, was identified as a genetic risk factor for MS[190], which emerged originally from a meta-analysis of GWAS on RA[193]. A proxy of rs6859219, rs71624119, was also associated with MS[192] and RA[195]. Other three DNA variants (rs10213692[260], rs11377254[260] (indel) and rs7731626[194]) within *ANKRD55* were also associated with RA. In addition, preliminary chromatin immunoprecipitation (ChIP) data was suggestive of a greater enrichment for the histone mark of enhancer activity, H3K4me1, with carriage of the risk allele at both rs10065637 and rs6859219[261].

In the present work, we characterize five *ANKRD55* non-coding variants associated with several autoimmune diseases as possible enhancer variants. For this purpose, first we identified the main cellular source of *ANKRD55* expression in PBMCs and we correlated the expression of the three *ANKRD55* transcripts with rs6859219 in the five isolated immune cell subtypes. Subsequently, we analyze five *ANKRD55* predicted enhancer variants for their possible capacity to modulate the expression of *ANKRD55* transcripts and its interacting genes, *IL6ST* and *IL31RA*, using CRISPR/dCas9-based epigenome editing; this was achieved during a collaborative short-term research stay at the University of Manchester.

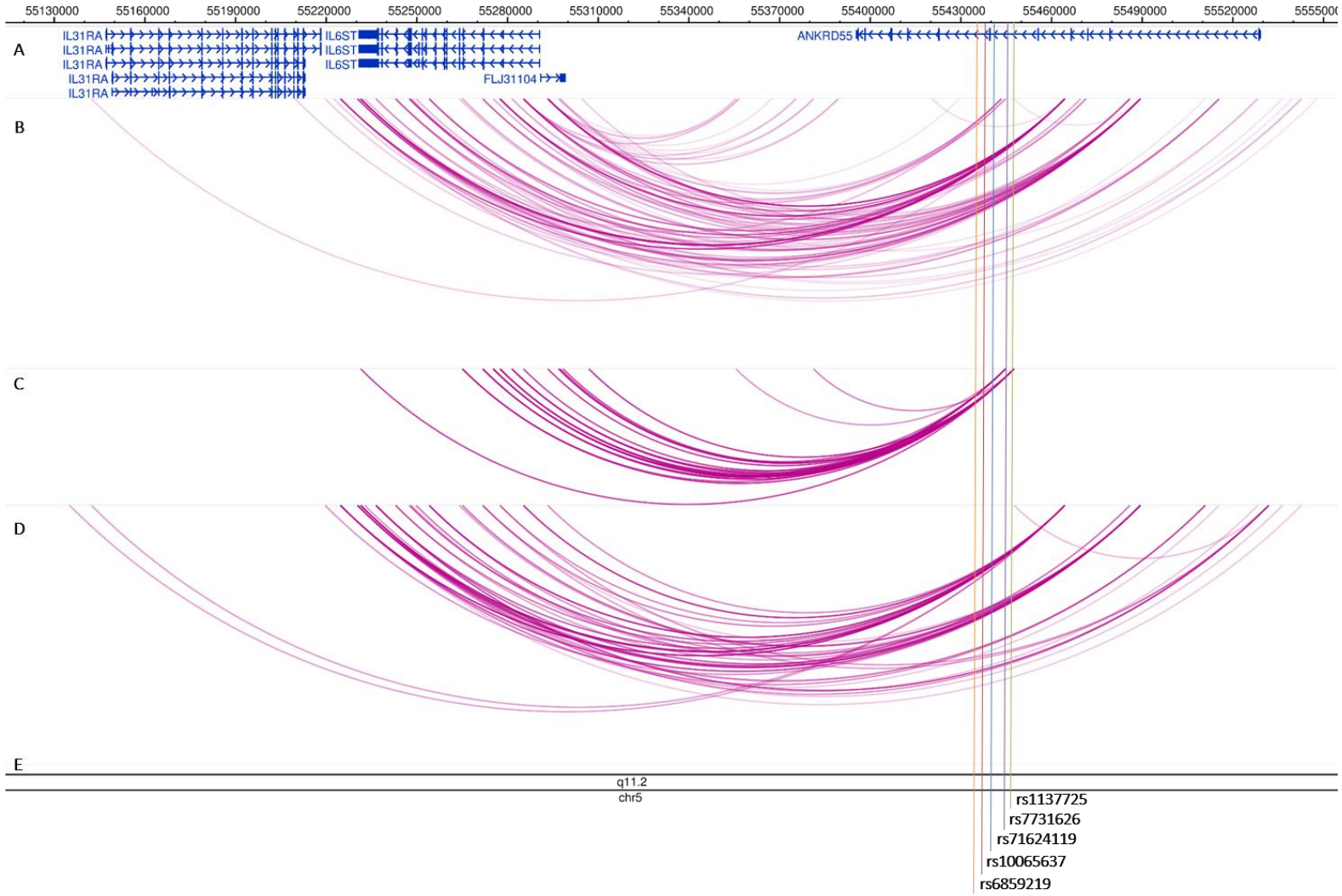


Figure 7. Long-range interactions in the 5q11 locus. (A) Refseq gene annotations. (B-E) Identified interactions in the Region and Promoter capture experiments in GM12878 (B, C) and Jurkat (D, E) cells. Genomic coordinates from Hg19 version are shown on the top the figure. *ANKRD55* variants associated with MS and/or RA positions are indicated below the figure.

MATERIALS AND METHODS

1. Primary human cell and cell lines

Human peripheral blood samples were collected after obtaining written informed consent from donors and approval by local ethics committee (Comité Ético de Investigación Clínica de Euskadi). Peripheral blood mononuclear cells (PBMC) of 23 healthy donors were provided by the Basque BioBank for Research-OEHUN (http://www.biobancovasco.org). Demographical and genotype data of controls samples are included in Table 2.

Table 2. Demographical details and ANKRD55 genotypes of healthy controls.

Individual group	Number (Female/Male)	Age (mean ± SD)	rs6859219 (AA/AC/CC)
Controls	23 (11/12)	43 ± 10,12	(6/7/10)

Human embryonic kidney 293 that expresses SV40 large T antigen (HEK293-T) and HEK293-T clone selected for high-titer lentivirus production (Lenti-X HEK293-T) (Cat. No. 632180; ClonTech) cell lines were maintained in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% inactivated fetal bovine serum (FBSi), and 1% penicillin/streptomycin (all from Sigma-Aldrich). All cultures were maintained at 37°C and 5% CO_2 in a humidified atmosphere.

2. Single guide RNAs design and cloning procedure

Five *ANKRD55* DNA variants were analyzed in 3 separated regions, as follow: region 1 (rs6859219 and rs10065637), region 2 (rs71624119) and region 3 (rs11377254 and rs7731626) (Table 3). Two sgRNAs spanning less than 250bp from the SNPs included in each region (except one sgRNA) and positive and negative controls for *ANKRD55* transcripts 001 and 005 and *IL6ST* (Table 3) were designed using Chop Chop tool (http://chopchop.cbu.uib.no/)[262] and Benchling software (https://benchling.com).

sgRNAs were ordered as pair of primers including specific restriction site from Sigma-Aldrich.

Each pair of oligos annealed and cloned directly into pLKO5.sgRNA.EFS.GFP vector (Cat. No. 57822; Addgene)[263]. Briefly, the vector was digested and dephosphorylated using FastDigest BsmBI (Thermo Fisher), verified by agarose gel electrophoresis and purified using QIAquick Gel extraction Kit (Qiagen) following manufacturer's protocol. Oligos were annealed and phosphorylated mixing in equimolar concentrations with T4 polynucleotide kinase (New England Biolabs), and ligated with purified and digested vector with T4 DNA ligase (New England Biolabs). Stbl3 competent cells were transformed with ligation reaction. 2-3 selected colonies from each construct were purified using Qiaprep spin miniprep kit (Qiagen) according to manufacturer's protocol. The insert was confirmed by Sanger sequencing.

Table 3. List of designed and cloned sgRNAs for CRISPR/dCas9 study

Region	SNP	Code	sgRNA	Sequence 5'-3'	PAM
1	rs10065637 rs6859219	2A	rs685+rs100 F	CACCGAGTATGTAGGAGGGTGCTAT	GGG
			rs685+rs100 R	AAACATAGCACCCTCCTACATACTC	
		3A	rs685 F	CACCGCCAACAGAAATCAACTGTGG	GGG
			rs685 R	AAACCCACAGTTGATTTCTGTTGGC	
2	rs71624119	4B	rs716 F1	CACCGAAAGCTAGGAGACAACACCT	TGG
			rs716 R1	AAACAGGTGTTGTCTCCTAGCTTTC	
		6B	rs716 F	CACCGAATGAAGATGCCTGATAAAG	GGG
			rs716 R	AAACCTTTATCAGGCATCTTCATTC	
3	rs7731626 rs11377254	7B	rs113 F	CACCGTTGTTCCCCGTCTTGGTTGG	CGG
			rs113 R	AAACCCAACCAAGACGGGGAACAAC	
		9A	rs113 F	CACCGCCACCACAATTTCAGTTAA	AGG
			rs113 R	AAACTTAACTGAAATTGTGGTGGC	
ANKRD55 positive control	001	B1	ANKp F	CACCGTTTCAAGGCTCGAGTGTCTG	TGG
			ANKp R	AAACCAGACACTCGAGCCTTGAAAC	
	005	C12A	ANK005p F	CACCGTAAGTGTTAAATGAGCCCGA	AGG
			ANK005p R	AAACTCGGGCTCATTTAACACTTAC	
ANKRD55 negative control		I1	ANKneg F	CACCGTTTGAGTTAGAGAGATCTAG	CGG
			ANKneg R	AAACCTAGATCTCTCTAACTCAAAC	
IL6ST positive control		6FA	IL6STp F	CACCGGGCAAAGTGGGCCTGCTAA	CGG
			IL6STp R	AAACTTAGCAGGCCCACTTTGCCC	

3. Lentivirus production

Lentivirus was produced using third-generation system with the vectors pRSV-REV, pMDLg/pRRE and pMD2.G (Cat. No: 12253, 12251 and 12259, respectively; Addgene)[264] and polyethyenimine (PEI) transfection protocol.

For this purpose, $10x10^6$ Lenti-X HEK293-T cells in DMEM/10% FBSi w/o antibiotics were plated in 15cm dish the day before transfection. Transfection day, a mixture of lentiviral transfer plasmid including specific sgRNA or dCas9^{p300} (pLKO5.sgRNA.EFS.GFP or pLV-dCas9-p300-P2A-PuroR (Cat. No. 83889; Addgene)[249], respectively) and lentiviral envelope and packaging plasmids (pRSV-REV, pMDLg/pRRE and pMD2.G) at 4:2:1:2 DNA ratio were mixed with serum-free DMEM w/o phenol red and FBSi. PEI (1µg/µl) was added to diluted DNA at 6:1 ratio of PEI (µg): total DNA (µg), vortex and incubated for 15 min at room temperature (RT). Lenti-X HEK293-T cells were transfected with the mixture. 72 h after transfection, the virus containing supernatant was harvested and concentrated using Vivaspin 20 columns (Sartorious). Transfection of empty pLKO5.sgRNA.EFS.GFP vector was carried out as negative control.

4. Stable cell line generation

First, HEK293-T cells were transduced with concentrated dCas9^{p300} lentivirus and selected with puromycin (Sigma-Aldrich) to obtain a polyclonal cell line that stably expressed the transgene. Briefly, $500x10^3$ HEK293T cells/well in DMEM/10% FBSi w/o antibiotics were plated in a six-well plate. The following day, cells were transduced with 8µl/ml polybrene (1mg/ml) and 50 µl of previously thawed dCas9^{p300} concentrated lentivirus. 72 h after transduction, the cells were collected and plated in DMEM/10% FBSi containing puromycin. The cell media was changed every 2-3 days to maintain the dose of antibiotic. Once the polyclonal cell populations grew well and expanded sufficiently, cell stocks were prepared.

Next, we transduced HEK293T-dCas9^{p300} with ten individual sgRNAs (Table 3) and negative control lentivirus as described earlier. 72 h later, GFP level of a small amount of polyclonal cell populations was measured with BD LSRFortessa X-20 (BD Biosciences) to

determine the percent transduction. Then the expanded cell populations expressing GFP were sorted using BD FACS Aria Fusion cell sorter (BD Biosciences) in flow cytometry core facility within the Faculty of Biology, Medicine and Health at the University of Manchester. HEK293T-dCas9^{p300}-B1 and C12A (*ANKRD55* transcripts positive controls) stable cell lines could not be sorted (+ GFP% 92.2 and 95.4%, respectively).

5. Isolation of immune cell subsets

PBMC samples were removed from liquid nitrogen storage, thawed, washed with phosphate-buffered saline (PBS) (Gibco) and centrifuged at 300 g for 10 min. PBMC number and viability were determined using Neubauer chamber and trypan blue staining. 20x10^6 PBMC were divided in two tubes, resuspended in PBS and centrifuged at 300 g for 10 min. Positive selection of immune cell subsets was performed by suspending 10x10^6 PBMC in 80 µl of ice-cold separation buffer (PBS, 0.5% bovine serum albumin (BSA) and 2 mM EDTA) (all from Sigma-Aldrich) and 20 µl antibody-labelled microbeads (Miltenyi Biotec) for 15 min at 4°C. PBMC were washed with 2 ml of ice-cold separation buffer and centrifugated at 300 g for 10 min at 4 °C. Magnetic separation of microbead-labelled cells was done using MS Columns (Miltenyi Biotec) following manufacturer's protocol. CD8$^+$ (T cytotoxic cells; Cat. No. 130-045-201) were purified from first tube and CD56$^+$ (natural killer (NK) cells; Cat. No. 130-050-401) were retrieved from the resulting CD8$^-$ fraction. CD19$^+$ (B cells; Cat. No. 130-050-301) were recovered from the second tube, CD14$^+$ (monocytes; Cat. No. 130-050-201) were obtained from the CD19$^-$ fraction and CD4$^+$ (helper T cells; Cat. No. 130-045-101) were subsequently obtained from the CD14$^-$ fraction.

Subset purity was determined using flow cytometric analysis using PE mouse anti-human CD4$^+$ (Cat. No. 130-098-167), PE-Vio 770 mouse anti-human CD8$^+$ (Cat. No. 130-098-060), FITC mouse anti-human CD14$^+$ (Cat. No. 130-080-701), VioBlue mouse anti-human CD19$^+$ (Cat. No. 130-098-606) and PE-Vio 770 mouse anti-human CD56$^+$ (Cat. No. 130-098-132) antibodies, as well as PE mouse IgG2a (Cat. No. 130-098-849), PE-Vio 770 mouse IgG2a (Cat. No. 130-098-564), FITC mouse IgG2a (Cat. No. 130-098-877), VioBlue mouse IgG1 (Cat. No. 130-099-756) and PE-Vio 770 mouse IgG1 (Cat. No. 130-098-563) isotype controls (all from Miltenyi Biotec).

6. Gene expression analysis

Total RNA from human immune subsets was obtained with TRI Reagent® (Sigma-Aldrich), following the manufacturer's protocol and treated with DNase I (Sigma-Aldrich) for genomic DNA removal. Total RNA extraction and DNAse I treatment from stable HEK293T-dCas9^{p300}-sgRNA expressing cell lines was done with RNeasy Micro Kit and QIAshredder (Qiagen) following the manufacturer's protocol. RNA concentration was determined using a NanoDrop spectrophotometer. RNA was reverse transcribed with High-capacity cDNA reverse transcription kit (Applied BioSystems) and qPCR was performed using Fast SYBR Green Master Mix or Power SYBR Green PCR Master Mix (Applied BioSystems) and primers for different *ANKRD55* transcripts[265], *IL6ST*[266], *IL31RA*, *GAPDH* purchased from Integrated DNA Technologies (IDT) or Sigma-Aldrich and *ACTB* (predesigned from QIAGEN or Sigma-Aldrich) (Table 4). Expression of *ANKRD55* transcripts, *IL6ST* and *IL31RA* was normalized to internal controls (*GAPDH* and *ACTB*) and analyzed using the $2^{-\Delta Ct}$ method[267]. Reactions were run in triplicate and no-template and no reverse transcription controls were included.

Table 4. List of primers used for gene expression analysis in eQTL and CRISPR/dCas9 studies

Analysis	Gene symbol - Transcript variant	Primer	Sequence 5'-3'
CRISPR/dCas9	*ANKRD55*-001	001qF	CAGCCTCAACACACACAAATGC
		001qR	TAGTTGATTATGGACGGCCCCTG
eQTL	*ANKRD55*-005	005qF	CGGGCTCATTTAACACTTACTATTTC
		005qR	CTTAGCCAGCAACAGCTCCTG
	ANKRD55-007	007qF	TCCACTATGCTCGGCTGC
		007qR	CTCGCTGATGTTCGACTGTTG
CRISPR/dCas9	*IL6ST*	IL6STF	CCGCCACATAATTTATCAGT
		IL6STR	AAGGTCTTGGACAGTGAATG
	IL31RA	IL31RAF	CTCTGCGATGTGCGGTCAA
		IL31RAR	GCCAAGTGTTTTCTCTAGGACTG
eQTL	*GAPDH*	GAPDHF	GCAACAATATCCACTTTACCAGAG
CRISPR/dCas9		GAPDHR	CACATCGCTCAGACACCAT
CRISPR/dCas9	*ACTB*	ACTBF	GGGCATGGGTCAGAAGGATT
		ACTBR	TTCTCCATGTCGTCCCAGTT

7. Statistical analysis

Analysis of the gene expression data was performed with GraphPad v.6 (GraphPad Software, San Diego, CA, USA). Paired t test for comparison of $CD4^+$ and the other immune subsets, Kruskal–Wallis test for comparison of three groups and Pearson correlation coefficient for correlation analysis between the three transcripts.

8. Bioinformatic analysis

Several MS and RA *ANKRD55* variants and their proxies were subjected to *in silico* analysis to evaluate their regulatory potential using the following bioinformatic tools: Ensembl Variant Effect Predictor (VEP)[268], RegulomeDB[269] and HaploReg[270].

HaploReg v4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php)[270] was used to search for proxy SNPs (r^2>0.8) associated with the rs6859219, rs11377254 and rs7731626 in the *ANKRD55* intronic region. HaploReg uses LD information from the 1000 Genomes Phase 1[271]. Evidences from ENCODE[272] about regulatory protein binding, chromatin structure, chromatin state of the region, and putative transcription factor binding motifs that are altered by the variant are also included.

For all input variants, the VEP contains detailed annotation for effects on transcripts, proteins and regulatory regions[268]. The Ensembl Regulatory Build[273], which uses data from ENCODE[272], BLUEPRINT[274], and the NIH Epigenomics Roadmap[275], is the primary regulatory annotation[268].

RegulomeDB (http://www.regulomedb.org/) expounds SNPs with known and predicted regulatory elements in the intergenic regions using data included in public datasets from GEO[276], ENCODE project[272], and published literature as well as computational predictions and manual annotations to identify putative regulatory potential and identify functional variants. Known and predicted regulatory DNA elements include regions of DNAase hypersensitivity, binding sites of transcription factors, and promoter regions that have been biochemically characterized to regulation transcription. Each variant includes a score (1-6) based on functional evidence, lower scores indicate increasing evidence for a variant to be located in a functional region[269].

RESULTS

Several bioinformatic tools such as, RegulomeDB[269], HaploReg[270] and VEP[268], are useful to determine the effect of the variant and identify candidate regulatory SNPs for prioritizing SNPs for functional analyses.

HaploReg showed that four SNPs (rs6873385, rs10065637, rs71624119 and rs10213692) are in strong LD (r^2>0.9) with rs6859219. However, rs11377254 and rs7731626 did not have any proxy in strong LD (r^2>0.8) (Table 5). Haploreg v4.1 also showed that three SNPs (rs6859219, rs11377254 and rs7731626) overlap with a ChromHMM-predicted promoter region in diverse primary T helper cells, monocytes-$CD14^+$ RO01746 primary cells and GM12878 lymphoblastoid cell line. Moreover, analysis of chromatin state showed that five SNPs (rs6873385, rs6859219, rs71624119, rs11377254 and rs7731626) mapped to enhancer regions in several tissues, including numerous cell lines and primary monocytes, neutrophils, and T cell subtypes, among others. ChIP-Seq experiments also detected that rs6859219, rs10065637, rs11377254 and rs7731626 bound to different proteins and six out seven SNPs affected diverse protein-binding motifs (Table 5).

According to the VEP, rs6859219 and rs10065637, overlap with a transcription enhancer region in lymphoblastoid cell lines (GM12878). Moreover, rs71624119, rs11377254 and rs7731626 are located in regulatory regions in neutrophil, $CD14^+CD16^-$ monocyte, M0 macrophage, GM12878 and monocytes-$CD14^+$ (Table 5).

In RegulomeDB analysis, two (rs11377254 and rs7731626) out of seven DNA variants achieved the highest regulatory score (i.e., 2a) likely to affect binding. Rs6859219 scored 2b likely to affect binding. rs10065637 and rs71624119 showed lower regulatory score, 3a and 5, respectively. However, two SNPs in LD with rs6859219 (rs10213692 and rs6873385) did not have any available Regulome DB data (Table 5).

Table 5. *In silico* analysis for regulatory potential evaluation of rs6859219, rs11377254 and rs7731626 and their proxies. Proxies for three DNA variants were found using HaploReg, according to European ancestral super-population from 1000 Genomes Project Phase 1 release. SNP with r^2>0.8 were included. The predicted regulatory effects from each SNP were studied using HaploReg, VEP and RegulomeDB and summarized in the table. Abbreviations: TF: transcription factor

	chr	Pos (hg38)	LD (r^2)	LD (D')	Variant	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	Regulome DB	VEP
rs6859219	5	56141024	0.99	1	rs6873385		4 tissues				No data	Trancript 001: intron variant; trancript 002: intron variant; trancript 007: intron variant, non-coding transcript variant
	5	56142753	1	1	rs6859219	BLD	13 tissues	8 tissues	EBF1, PAX5N19, TCF12, AP2ALPHA, AP2GAMMA, CMYC	ERalpha-a	2b (TF binding + any motif + DNase Footprint + DNase peak)	Regulatory region variant (GM12878, TF binding site); trancript 001: intron variant; trancript 002: intron variant; trancript 006: downstream gene variant; trancript 007: intron variant, non-coding transcript variant
	5	56143024	0.99	1	rs10065637			4 tissues	EBF1, NFKB, PU1	Maf, Pou5f1	3a (TF binding + any motif + DNase peak)	Regulatory region variant (GM12878, TF binding site); trancript 001: intron variant; trancript 002: intron variant; trancript 006: downstream gene variant; trancript 007: intron variant, non-coding transcript variant
	5	56144903	0.9	0.99	rs71624119		8 tissues			Bbx, DMRT1, STAT	5 (TF binding or DNase peak)	Regulatory region variant (neutrophil (VB), $CD14^+CD16^-$ monocyte (VB), M0 macrophage (VB), GM12878, Monocytes-$CD14^+$, promoter flanking region); trancript 001: intron variant; trancript 002: intron variant; trancript 006: intron variant; trancript 007: intron variant, non-coding transcript variant
	5	56146422	0.9	0.99	rs10213692					Lmo2-complex_1, Myf_4, ZEB1_known3, Znf143_disc3	No data	Trancript 001: intron variant; trancript 002: intron variant; trancript 006: intron variant; trancript 007: intron variant, non-coding transcript variant

rs11377254	5	56148654	1	1	rs11377254	BLD	6 tissues	6 tissues	BATF, BCL11A, EBF1, IRF4, JUND, MEF2A, NFKB, PAX5C20, PAX5N19, RAD21, WHIP	Evi-1_4, GR_known10, HDAC2_disc6, Zfp105	2a (TF binding + matched TF motif + matched DNase Footprint + DNase peak)	Regulatory region variant (neutrophil (VB), $CD14^{+}CD16^{-}$ monocyte (VB), M0 macrophage (VB), GM12878, Monocytes-$CD14^{+}$, promoter flanking region); trancript 001: intron variant; trancript 002: intron variant; trancript 006: intron variant; trancript 007: intron variant, non-coding transcript variant
rs7731626	5	56148856	1	1	rs7731626	BLD	6 tissues	8 tissues	TCF12	CAC-binding-protein, Egr-1_known2, MAZ, MZF1::1-4_2, MZF1::1-4_3, PPAR_2, PU.1_known2, Pax-4_5, Pou2f2_disc2, RREB-1_2, RXRA_known1, RXRA_known6, SP1_disc3, SP1_known4, SRF_disc2, Sp4, TFII-I, UF1H3BETA, WT1, Zfp281	2a (TF binding + matched TF motif + matched DNase Footprint + DNase peak)	Regulatory region variant (neutrophil (VB), $CD14^{+}CD16^{-}$ monocyte (VB), M0 macrophage (VB), GM12878, Monocytes-$CD14^{+}$, promoter flanking region); trancript 001: intron variant; trancript 002: intron variant; trancript 006: intron variant; trancript 007: intron variant, non-coding transcript variant.

The risk (C) allele of rs6859219 is associated with higher expression of ANKRD55 001, 005, and 007 in CD4⁺ T cells

Considering that *ANKRD55* is a genetic risk factor for diverse autoimmune diseases, it is likely to have a role in the immune system. In a previous work from our research group, Lopez de Lapuente *et al*.[265] characterized *ANKRD55* expression in immune and non-immune cells by conventional RT-PCR and concluded that 001, 005 and 007 transcript variants were expressed in PBMC. Moreover, we analyzed the association of rs6859219 with *ANKRD55* expression by qPCR in MS patients and healthy controls. When individuals were grouped by genotype, homozygotes for the risk allele (CC) showed significantly higher expression of the three *ANKRD55* transcripts than carriers of the protective allele (AA). No significant differences were found between MS patients and controls.

PBMC are composed of lymphocytes (T cells, B cells and NK cells), monocytes and dendritic cells. In humans, the frequencies of these populations vary across individuals, but typically, lymphocytes are in the range of 70–90 %, monocytes range from 10 to 20 %, while dendritic cells are rare, accounting for 1–2 %. The frequencies of cell types within the lymphocyte population include 70–85 % $CD3^+$ T cells, 5–10 % B cells and 5–20 % NK cells. The $CD3^+$ lymphocytes are composed of $CD4^+$ and $CD8^+$ T cells, roughly in a 2:1 ratio[277].

We set out to identify the main cellular source of *ANKRD55* expression in this heterogeneous mixture of cell types. For this purpose, $CD4^+$, $CD8^+$, $CD14^+$, $CD19^+$ and $CD56^+$ immune subsets were isolated from PBMC of 23 healthy controls, subsets purity was determined using flow cytometric analysis showing a mean of 92.26%, 91.36%, 89.53%, 51.53% and 61.88% for $CD4^+$, $CD8^+$, $CD14^+$, $CD19^+$ and $CD56^+$, respectively, and *ANKRD55* transcripts were analyzed by qPCR using specific primers. The results showed that the three *ANKRD55* transcripts were uniquely and highly expressed in $CD4^+$ T cells but were undetectable or minimally expressed in the other subsets (Figure 8, A-C). Individual transcript levels were highly correlated (Figure 8, G-I) and the homozygotes for the risk allele (CC) expressed significantly higher levels of 001 and 005 transcripts in $CD4^+$

T cells (Figure 8, D-F). Consequently, we found that rs6859219 regulated the expression of ANKRD55 in PBMC and CD4⁺ T cells, qualifying rs6859219 as a novel eQTL for ANKRD55.

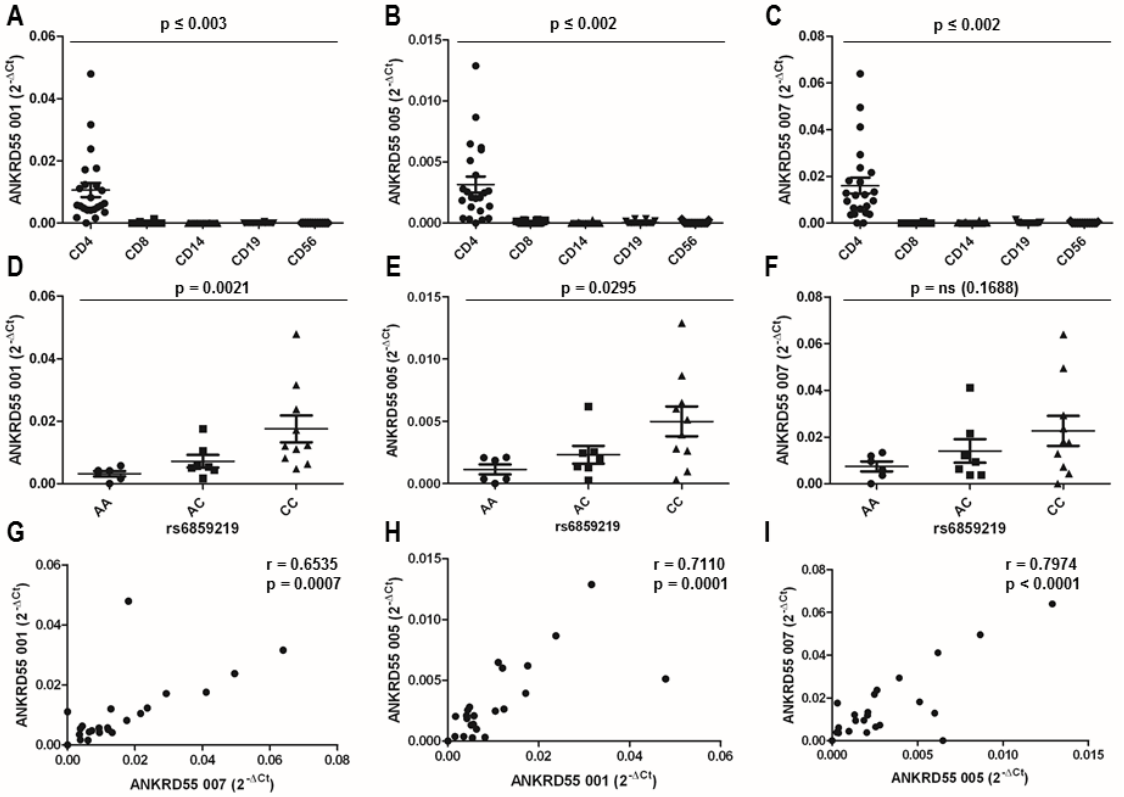


Figure 8. qPCR analysis of *ANKRD55* transcripts 001, 005, and 007 in CD4⁺, CD8⁺, CD14⁺, CD19⁺ and CD56⁺ immune subsets of PBMC from healthy controls. (A-C) The expression of three *ANKRD55* transcripts was analyzed by qPCR in five purified immune subsets of PBMC from 23 healthy controls. Expression differences between CD4⁺ T cells and the other subsets was analyzed by paired t-test. The highest p-values from the comparisons are shown. (D–F) CD4⁺ *ANKRD55* transcripts were grouped according to genotype and compared using the Kruskal–Wallis test. (G–I) Correlation analysis among three *ANKRD55* transcripts in CD4⁺ T cells was determined using the Pearson correlation coefficient. Expression of each transcript was normalized to internal controls (*GAPDH* and *ACTB*) and for each individual $2^{-\Delta Ct}$ value is represented with a dot.

ANKRD55 transcript 001 is regulated by rs71624119 in HEK293-T-Cas9^{p300} cells

In this study dCas9-p300 activator expression vector was used because of the low expression levels of *ANKRD55* transcripts in HEK293-T cells, and this was done with the aim to assess any potential increase in *ANKRD55, IL6ST* and *IL31RA* expression due to the presence of predicted enhancer variants. Preliminary data from two independent experiments showed that expression of *ANKRD55* transcript 001 was increased in engineered HEK293-T-Cas9^{p300} cells which stably expressed 4B sgRNA construct that targets the region 2 including rs71624119 (Figure 9). *ANKRD55* transcript 005 expression levels were highly increased in HEK293-T-Cas9^{p300}-C12A cells (Figure 9). This sgRNA was designed in the promoter region as the positive control for transcript 005. *IL6ST* expression in experiment 1 was highly increased in HEK293-T-Cas9^{p300} transduced with 9A (region 3 including rs11377154 and rs7731626) as well as 6FA (positive control, *IL6ST* promoter region) compared to the control (CT) (Figure 9). Moreover, in both experiments HEK293-T-Cas9^{p300}-7B cell line targeting also region 3 showed more modest expression increased compared to the *IL6ST* positive control. For *IL31RA* expression, HEK293-T-Cas9^{p300}-4B showed upregulation compared to the control (CT) in experiment 1 (Figure 9).

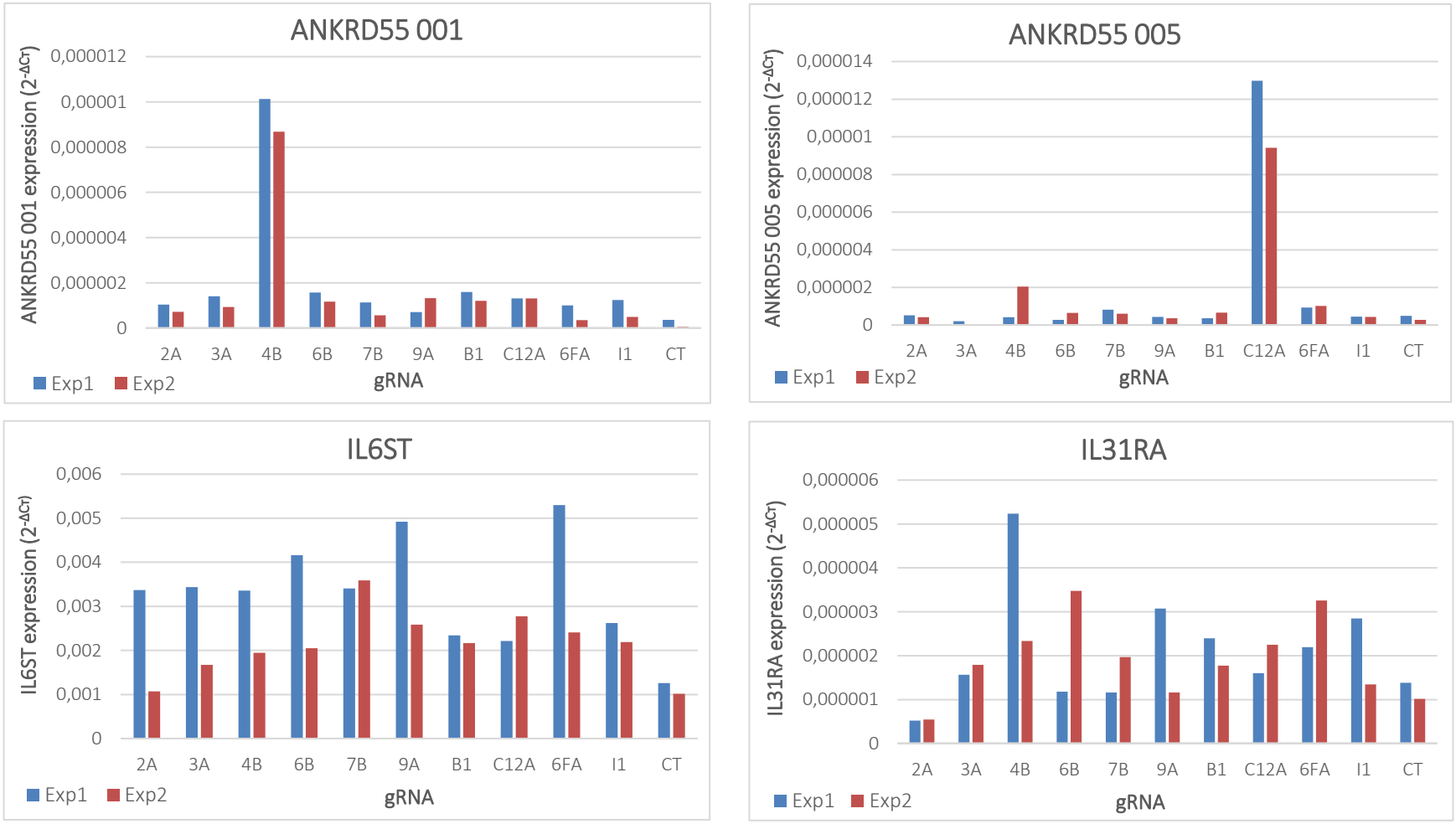


Figure 9. Expression analysis of *ANKRD55* transcripts 001 and 005, *IL6ST* and *IL31RA* in HEK293T-Cas9^{p300}-specific sgRNA cells. The expression of *ANKRD55, IL6ST* and *IL31RA* was analyzed in two independent experiments (Exp 1 and 2) by qPCR in HEK293T stably expressing Cas9^{p300} and individual sgRNA, as negative control empty sgRNA delivery vector stably expressing cells were used (CT). Expression of each transcript was normalized to internal controls (*GAPDH* and *ACTB*) and for each $2^{-\Delta Ct}$ value is represented.

DISCUSSION

Much progress has been made to increase understanding of the role of the immune system in MS[278]. All genetic studies point to the immune system as being the primary cause of MS pathology, with a complex interaction occurring between genes and environment[56]. Genetic susceptibility and environmental factors prime the immune response in MS that targets the CNS. Genetic predisposition accounts for about only 25% of the life-time risk of patients developing MS. To define this risk at the molecular level, the associated gene for every SNP needs to be identified and its biological effect needs to be characterized[278].

The first identified MS-associated SNP in *ANKRD55*, rs6859219[190,191], is located in a genomic region with a high recombination rate[265]; it has only four proxies with $r^2 > 0.8$ (Table 5). It is interesting to note that four SNPs at this locus that are reported to be associated with MS and other autoimmune diseases, such as RA[193,195], CD[196], JIA[199] and T1D[198], correspond to rs6859219 itself or its closest proxies. rs6859219 in *ANKRD55* emerged originally from a meta-analysis of GWAS on RA[193]. However, other DNA variants in moderate (rs7731626[194]; $r^2<0.5$) or non-LD (rs11377254[260]) with rs6859219 in *ANKRD55* locus have also been associated with RA. The most strongly associated SNP at the *ANKRD55* locus of the latest comprehensive association studies of MS[179,192] appears to be a perfect proxy of rs6859219.

The qPCR results from this study indicate that 001 is the most highly expressed protein-coding transcript in human $CD4^+$ T cells, whereas the noncoding transcript variant 007 is by far the most abundant transcript overall. *ANKRD55* 007 is 853 bp long and contains seven exons, some of which partially or completely overlap with the exons of the protein-coding transcripts, whereas others are made up of intronic sequences. According to these features, it can be categorized as a long noncoding RNA (lncRNA)[279]. LncRNAs are RNAs > 200 nt that do not contain an open reading frame; although their biological meaning and function still need to be elucidated, ample evidence, including specific subcellular localization, tissue specificity, and association with disease, indicates that they may be essential regulators of cell function[279–281]. The importance of lncRNAs in MS is highlighted

by some studies showing that various SNPs associated with MS in GWAS are located in genomic regions that code for lncRNAs[282] or regulate lncRNA expression in cis in human monocytes[283]. Abundant expression of *ANKRD55* 007 in $CD4^+$ T cells may be an indication of the unique cell-specific functional role for this noncoding RNA[265].

Most autoimmune-disease-risk effects identified by GWAS localize to open chromatin with gene-regulatory activity. GWAS loci are also enriched in eQTLs, thus suggesting that most risk variants alter gene expression[183,284]. Importantly, we found that the risk (C) allele of rs6859219 is associated with higher expression of *ANKRD55* mRNA in PBMC and $CD4^+$ T cells, qualifying this SNP as a novel eQTL for *ANKRD55*. The genotype seems to increase the overall expression of the gene but not the splicing or differential expression of transcripts, because the effect was observed for all three transcripts assessed by qPCR (005, 007, and 001)[265]. Given that higher levels of *ANKRD55* were observed in healthy carriers of the risk allele, and not only in patients, it is conceivable that the high level of *ANKRD55* increases susceptibility to pathologic inflammation, rather than being a consequence of it[265].

Recently, Chun *et al.* (2017) identified in *ANKRD55* an eQTL effect driven by rs71624119, a perfect proxy of rs6859219, in $CD4^+$ T cells that exhibited shared associations with MS, CD and RA[228]. Moreover, James *et al.* (2018) also showed that rs71624119 was significantly associated with *ANKRD55* expression and was found to be located in an active regulatory region (H3K4me1) with potential binding for STAT1 transcription factors, known to be involved in immunity[285]. Trynka *et al.* (2013) showed, based on chromatin mark datasets from the ENCODE and NIH Roadmap Epigenomics Mapping Consortium, that rs10065637, the perfect proxy of rs6859219 (separated by 271 bp), was found to occur very close to a H3K4me3 chromatin mark (distance of only 5 bp) that is specific to primary $CD4^+$ TReg cells and $CD4^+$ memory T cells[226]. In a study on RA, rs7731626, which is located in another *ANKRD55* intron 6.1 kb from rs6859219, was found to regulate the expression of *ANKRD55* in $CD4^+$ T cells but not in $CD14^+CD16^-$ cells[194] (no ANKRD55 transcript-specific information). Westra *et al.* (2018) in a fine-mapping and functional study including *ANKRD55* T1DM and RA associated DNA variants observed that rs10213692 does not overlap with any DNAse I hypersensitive sites, while

rs11377254 overlaps with such sites in 101 experiments performed on distinct cell lines, and rs7731626 overlaps with 55 such, many of which are involved in the immune system. Similarly, rs11377254 and rs7731626 overlap with different transcription factor binding site peaks in lymphoblastoid cell lines and H3K4me3 sites in T-helper cells and monocytes, while rs10213692 does not. rs11377254 did overlap with a peak in their time course of Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq), a method for mapping chromatin accessibility genome-wide. Only rs10213692 showed differential binding in Jurkat T cells in the gel electrophoresis mobility shift assay (EMSA) used to detect protein complexes with nucleic acids[260].The prediction of MS and RA-associated *ANKRD55* DNA variants to be localized to putative enhancer elements encouraged us to perform a study aiming to validate this data through their possible capacity to regulate the expression of *ANKRD55* and its potential interacting genes, *IL6ST* and *IL31RA*[259], using CRISPR/dCas9-based epigenome editing. The preliminary analysis of five *ANKRD55* predicted enhancer variants showed that expression of *ANKRD55* transcript 001 was increased in engineered HEK293-T-Cas9^{p300} cells targeting the region including rs71624119. For *IL31RA* expression, only one experiment from HEK293-T-Cas9^{p300} modifying the region containing rs71624119 showed upregulation compared to the control. *IL6ST* expression in a unique experiment was highly increased in HEK293-T-Cas9^{p300} with modified region 3 including rs11377154 and rs7731626. Moreover, another HEK293-T-Cas9^{p300} stable cell line targeting also region 3 showed more modest expression increased compared to the *IL6ST* positive control.

In summary, this study contributes toward understanding ANKRD55 in the context of MS and the immune system. Correlation of rs6859219 with expression of *ANKRD55* in CD4^{+} T cells implies a functional link between the protein and ncRNA expression products of the *ANKRD55* gene with MS, and higher expression of the gene in carriers of the risk allele points to a potential proinflammatory role for ANKRD555. While confirmation in additional experiments is needed, our data suggest that rs71624119 is located in an enhancer region and regulates ANKRD55 transcript 001 expression in HEK293-T-Cas9p300 cells. Moreover, further functional analysis, using CRISPR/dCas9-based epigenome editing approach, could help us to identify other *ANKRD55* variants in

enhancer elements and to study their possible capacity to regulate the expression of *ANKRD55*, and its potential interacting genes *IL6ST* and *IL31RA*.

CHAPTER 3: PROTEIN-BASED ANALYSIS

INTRODUCTION

Spatial partitioning of biological functions is a phenomenon fundamental to life. In humans, this spatial partitioning constitutes a hierarchy of specialized systems ranging across scales, from organs to specialized cells, to subcellular structures, down to macro-molecular complexes[286]. At the cellular level, proteins function at specific times and locations. These subcellular locations such as organelles provide a specific chemical environment and set of interaction partners that are necessary to fulfill the protein's function[286]. Using high-throughput protein labeling and imaging techniques, several proteome-wide studies have found that protein localization and function are strongly correlated to each other[287–291]. In 2017, the first analysis of how proteins are organized in a cell using two different high throughput approaches, high content imaging and spatial proteomic, revealed that approximately half of the proteins can be found in more than one compartment[286]. Mislocalization of proteins can be associated with cellular dysfunction and disease[292,293]. Thus, knowledge of the spatial distribution of proteins at a subcellular level is essential for understanding protein function, interactions, and cellular mechanisms[286].

Moreover, cellular functions are mediated by complex 'interactome' networks of physical and functional interactions between biological macromolecules, including DNA, RNA, proteins, and lipids, as well as smaller molecules such as metabolites[294,295]. Many proteins exert their functions within cells in the context of protein complexes. Therefore, identifying protein–protein interactions (PPIs), defined as a direct physical contact between two proteins, is pivotal to gaining insight into the biological function of proteins[296]. If an uncharacterized protein binds to a protein whose role in the cell is understood, its function is likely to be related[297].

These interactions may occur on their own in a binary manner or may require additional interaction partners. Proteins can form multimeric complexes in which there are many direct interactions as well as indirect protein–protein associations between the different proteins of the complex[294]. Binary mapping approaches interrogate pairs of proteins for the existence of direct interactions between them. Protein-complex mapping

approaches, by contrast, aim to identify the set of proteins that belong to a multimeric protein complex, where in pairs of proteins form either direct contacts or are linked by indirect associations[294]. Pioneering interactomic techniques, such as yeast two-hybrid (Y2H), were important initial tools for dissecting PPIs in cells[298]. In this system, two recombinant fusion proteins are generated: the bait is fused with a DNA-binding domain and the prey with a transcription activation domain. These two elements are expressed in yeast along with a reporter gene whose regulation is under a promoter recognised by the recombinant transcription activation domain. If these two proteins interact with each other, the transcription activation domain becomes functional leading to expression of the reporter gene[299].

During the past decades, mass spectrometry-based proteomics has become the method of choice to comprehensively identify PPIs[296], for a single or a few proteins of interest and also in large-scale studies[300–302]. The first step in MS-based interaction proteomics is the purification nowadays mostly done via immunoprecipitation of the protein of interest with the aim to copurify and identify its interactors. Commonly used workflows are based on affinity-purification mass spectrometry of the protein of interest using specific antibodies or antibodies against epitope tags[303,304], such as FLAG[303,305], TAP[304,306,307], and GFP tags[308,309]. This approach has been the most commonly used purification strategy for large-scale interactomes in, for example, human cell lines[300,301] and yeast cultures[306,307]. After immunoprecipitation of the protein of interest, specific interactions must be identified. Target immunoprecipitations are compared to matched control samples, and specific interactors are identified by their significant enrichment in the target immunoprecipitations, whereas background binding proteins are equally present in target immunoprecipitations and controls.

As mentioned previously, ANKRD55 belongs to a family containing AR, which exclusively functions to mediate PPIs. ANKRD55 isoform 001 and 005 includes nine and one ANK repeats, respectively, consequently should take part in diverse PPI as has been previously confirmed for other members of the ANK family.

Therefore, the present study was conducted to provide clues about the biological functions of ANKRD55 and increase our understanding of its role in autoimmune disease.

For this purpose, in first instance we analyzed the subcellular localization of endogenous and recombinant ANKRD55 isoforms in several immune and non-immune cells. Then, we identified and validated potential protein partners of ANKRD55 in diverse subcellular compartments.

MATERIAL AND METHODS

1. Cell culture

Human embryonic kidney 293 (HEK293) and cervical cancer (HeLa) cell lines were maintained in DMEM supplemented with 10% FBSi and 2mM L-glutamine (all from Sigma-Aldrich). Human T-lymphocyte (Jurkat) and monocyte (U937) cell lines were kindly provided by Prof. Carlos Matute (University of the Basque Country, Leioa, Spain) and maintained in RPMI 1640 medium supplemented with 10% FBSi and 2mML-glutamine (all from Sigma-Aldrich). Neuroblastoma (SH-SY5Y) cell line was kindly provided by Prof. Pablo Villoslada (Institute of Biomedical Research August Pi Sunyer, Barcelona, Spain) and maintained in DMEM/F12 GlutaMAX supplement medium (Thermo Fisher Scientific) complemented with 10% FBSi (Sigma-Aldrich). All cultures were maintained at 37°C and 5% CO_2 in a humidified atmosphere.

2. DNA constructs and cloning procedures

pCMV6 containing TrueORF Gold Expression validated cDNA clone including C-terminal FLAG/myc tagged ORF of human ANKRD55 transcript 001 (NM_024669; Cat. No. RC221211) was purchased from OriGene.

Human ANKRD55 transcript 005 coding sequence was amplified by PCR from pCR-BluntII-TOPO clone IRCMp5012E115D (Source Bioscience) with specific primers including restriction sites (Supplementary Table 1) from IDT using Pfu DNA Polymerase (Agilent Technologies) following manufacturer's protocol. The cycling conditions were set to initial denaturation at 94°C 45 sec, followed by 30 cycles of denaturation at 94°C 45 sec, annealing at 50°C 45 sec and extension at 72°C 1 min 30 sec and a final extension at 72°C for 10 min. The PCR product size was verified by agarose gel electrophoresis, and the products were purified using GeneJET Gel Extraction and DNA Cleanup Micro Kit (Thermo Fisher Scientific) following manufacturer's protocol. Amplified product was digested using SgfI and MluI restriction enzymes (New England Biolabs) and verified by agarose gel electrophoresis. Purified insert was ligated with pCMV6 vector digested with the same

enzymes using T4 ligase (New England Biolabs). Individual colonies were screened by colony PCR and double digestion to identify those that contain the insert and confirmed by Sanger sequencing.

pCMV6 vectors including ANKRD55 transcripts were modified by annealed oligo cloning to exchange myc for his-tag. Specific overlapping primers containing his-tag sequence (Supplementary Table 1) were designed, annealed and cloned directly into pCMV6 vector. Briefly, oligos were resuspended in annealing buffer (10mM Tris, pH7.5, 50mM NaCl, 1mM EDTA), mixing in equimolar concentrations and incubate in the hot block at 94°C for 5 min, which was removed from the heat source allowing for slow cooling to RT. Annealed oligos were ligated with purified and digested pCMV6 vector with EcoRV and XhoI using T4 ligase (all from New England Biolabs). The insert was confirmed by Sanger sequencing.

3. ANKRD55 expression analysis and immunoprecipitation

3.1. *ANKRD55 isoform overexpression*

HEK293 and HeLa cells were cultured in T75 flasks (Sigma-Aldrich) and transfected with FLAG/myc or his-tagged ANKRD55 001 or 005 using MACsfectin Reagent (Miltenyi Biotec) according to manufacturer's protocol and incubated for 48h.

Jurkat cells were transfected with FLAG/myc-tagged ANKRD55 001 and pmaxGFP™ (positive control) using Amaxa 4D-Nucleofector device (Lonza). Jurkat cells were suspended in fresh complete RPMI 1640 medium and seeded into a six-well plate one day before transfection. The following day, $1x10^6$ cells/reaction were transfected with 6µg of specified plasmids using the SE Cell Line 4D-Nucleofector Kit (Cat. No. V4XC-1012, Lonza), according to the manufacturer's protocol and incubated for 48h.

3.2. *Cross-linking*

Cells were cross-linked using the membrane-permeable thiol-cleavable homobifunctional cross-linker dithiobis[succinimidyl propionate] (DSP;Thermo Fisher) dissolved in dimethyl sulfoxide. The cells were incubated in PBS supplemented with 100 µg/ml of DSP for 30

min, with gentle vortexing every 5 min. Reactions were quenched by adding 1 M Tris-HCl pH 7.5 to final concentration of 50 mM 15 min at RT.

3.3. *Obtaining cellular extracts*

For analysis of endogenous and recombinant ANKRD55 isoform expression, Jurkat, SH-SY5Y, HeLa and HEK293 were fractionated into different subcellular compartments following the protocol from Horton and Holden[310]. Nuclei from human $CD4^+$ T cells and Jurkat, SH-SY5Y, U937 and HEK293 cell lines were isolated by sucrose gradient centrifugation[311].

For silver staining analysis, FLAG-ANKRD55 immunoprecipitation (IP) was carried out from nuclear extracts obtained by sucrose gradient centrifugation; membranous organelles from subcellular fractionation and total protein extracts using lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 1% triton X-100 and 1% cOmplete EDTA-free protease inhibitor cocktail pH adjusted to 8.0) or RIPA buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate and 0.1% SDS, 1U/ml Benzonase and 1% cOmplete EDTA-free protease inhibitor cocktail). His-tagged ANKRD55 isoform-overexpressing cells were lysed with lysis buffer for IP under non-denaturing conditions and lysis buffer containing urea (50mM NaH_2PO_4, 300mM NaCl, 15mM Imidazole, 8M Urea, 1% cOmplete EDTA-free protease inhibitor cocktail at pH 8) for IP under denaturing conditions. Part of the lysates were kept as input controls.

For interactome analysis, ANKRD55 IP was done from nuclear extracts lysed with RIPA buffer and total protein extracts using lysis buffer. Both buffers included phosphatase inhibitors (1mM sodium pervanadate, 5 mM beta-glycerophosphate and 5 mM NaF). Part of the lysates were kept as input controls.

Protein concentration was estimated using BCA protein assay kit (Pierce).

3.4. *Subcellular fractionation*

Cells were fractionated into different subcellular compartments, sequentially generating fractions enriched for cytosolic, membrane bound organellar and nuclear proteins. Cells were washed three times with PBS from Gibco. Cell pellet was resuspended by gentle pipetting in ice cold buffer 1 (150mM NaCl, 50mM HEPES pH 7.4, 25 µl/ml digitonin (Sigma-Aldrich) and 1% cOmplete EDTA-free protease inhibitor cocktail (Roche)), incubated on ice for 10 min and centrifuged at 4°C, 2000 g for 3 min. This supernatant was the cytosol enriched fraction. The pellet was washed with ice cold PBS and centrifuge at 4°C, 100 g for 3 min to remove any remaining digitonin and resuspended by vortexing in ice cold buffer 2 (150mM NaCl, 50mM HEPES pH 7.4, 1% NP40 and 1% cOmplete EDTA-free protease inhibitor cocktail), incubated on ice for 30 min and centrifuged at 4°C, 7000 g for 5 min. This supernatant included membrane bound organelles such as the endoplasmic reticulum, golgi, mitochondria and some nuclear lumenal proteins. The pellet was washed with ice cold PBS and centrifuge at 4°C, 100 g for 3 min to remove any remaining NP40 extract and resuspended by vortexing in ice cold buffer 3 (150mM NaCl, 50mM HEPES pH 7.4, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 1U/ml Benzonase (Sigma-Aldrich) and 1% cOmplete EDTA-free protease inhibitor cocktail), incubated on ice for 1h and centrifuged at 4°C, 7000 g for 10 min to pellet insoluble proteins. This supernatant included nuclear membranes and nuclear proteins.

3.5. *Nuclear enrichment by sucrose gradient centrifugation*

Cells were washed three times with PBS and resuspended on ice-cold sucrose buffer I (0.32 M sucrose, 3 mM $CaCl_2$, 2 mM magnesium acetate, 0.1 mM EDTA, 10 mM Tris-HCl pH 8.0, 1 mM DTT, 0.5% NP-40) in a 50ml tube. A small aliquot of cell was examined with a phase-contrast microscope to ensure that they were uniformly lysed. Sucrose buffer II (2 M sucrose, 5 mM magnesium acetate, 0.1 mM EDTA, 10 mM Tris-HCl pH 8.0, 1 mM DTT) was added and mixed by gentle pipetting and inversion. Sucrose buffer II was added to 50 ml polypropylene centrifuge tubes for JA-20 rotor (sucrose cushion). Nuclei mixture were carefully layered onto the sucrose cushion, the gradient was completed with sucrose buffer I to have the same volume and centrifuged for 45 min at 30000 g and 4°C

in Beckman Coulter J2-MC High Speed Centrifuge. The supernatant was carefully removed, and nuclei were lysed with RIPA buffer.

3.6. *Immunoprecipitation*

For the IP of ANKRD55 isoforms together with their binding partners, ANKRD55 and control samples were incubated with Anti-FLAG G1 mouse monoclonal antibody coupled to the resin (cat.no. L00432; Genscript) at 4°C overnight (o/n), on rotating wheel. After incubation, resin was washed six times with washing buffer containing 50 mM Tris-HCl and 150 mM NaCl, pH adjusted to 7.4 (phosphatase inhibitors were included for interactome analysis). Elution of ANKRD55 001 complexes was subjected to an optimization process in which different elution methods were tested: (A) 0.1M glycine-HCl pH 2, neutralized with 1M Tris-HCl pH 10; (B) 0.1M glycine HCl pH 3.5 neutralized with 0.5M Tris, 1.5M NaCl pH 7.4; (C) 0.1M glycine, 0.15M NaCl and 0.5% SDS pH 2, neutralized with 1M Tris-HCl pH 10 (D) 0.1M glycine-NaOH pH 10 and neutralized with 1M glycine-HCl, pH 2 and (E) 100µg/ml FLAG peptide (Cat. No. F3290; Sigma-Aldrich).

For his-tagged ANKRD55 isoform IP, ANKRD55 and control samples were incubated with Ni^{2+}-NTA agarose (Qiagen) at 4°C o/n, on rotating wheel. After incubation, resin was washed three times with washing buffer 1 (50mM NaH_2PO_4, 300mM NaCl, 20mM Imidazole with or w/o 8M urea at pH 6.3) and another three times with washing buffer 2 (50mM NaH_2PO_4, 300mM NaCl, 20mM Imidazole at pH 5.5). ANKRD55 001 complexes were eluted using elution buffer (50mM NaH_2PO_4, 300mM NaCl, 10mM Imidazole, 100mM EDTA, with or w/o 8M urea at pH 8).

4. Western blot analysis

Protein of each subcellular fraction, input controls and elution samples were separated by SDS-PAGE on 10% gels, transferred onto polyvinylidene difluoride membrane (Merck Millipore) and blocked with 2% casein in Tris-buffered saline for 1h at RT. The blots from endogenous ANKRD55 intracellular localization study were incubated with anti-ANKRD55 Ab (Cat. No. ARP68892_P50; 1:1000) from Aviva Systems Biology at 4°C o/n, in the presence or absence of a 200-fold molar excess of ANKRD55 blocking peptide (Cat. No.

AAP68892; 2 mg/ml), whose sequence is part of both Ensembl isoforms 001 and 005 and is the original immunogen used to generate the polyclonal Ab, and then with horseradish peroxidase (HRP)-conjugated anti-rabbit (Cat.No.711-035-152; 1:3000) from Jackson Immunoresearch for 1 h at RT. Anti-GAPDH (Cat. No. MAB374; 1:1000) from Merck Millipore, anti-histone H3 (Cat. No. 9715; 1:1000) from Cell Signaling Technology, and anti-GRP94 (Cat. No. ADI-SPA-850-D; 1:1000) from Enzo Life Sciences Abs were used to validate fraction purity. The blots from recombinant ANKRD55 isoform overexpression analysis and interactome validation were incubated with anti-ANKRD55 (Cat. No. HPA051049; 1:500) from Sigma-Aldrich or anti-ANKRD55 (1:1000) from Aviva Systems Biology, anti-FLAG (Cat. No. 20543-1-AP; 1:1000), anti-RPS3 (Cat. No. 66046-1-Ig; 1:2000), anti-CLTC (Cat. No. 66487-1-Ig; 1:1000), anti-VIM (Cat. No. 60330-1-Ig; 1:1000), anti-TUBB (Cat. No. 66240-1-Ig; 1:5000) and anti-SMC1A (Cat. No. 21695-1-AP; 1:500) from Proteintech; anti-SMC3 (Cat. No. 5696; 1:2000), anti-14-3-3 pan (Cat. No. 8312; 1:1000) from Cell Signaling Technology; anti-PRKDC (Cat. No. NBP2-22128SS; 1:1000) and anti-HIF1AN (Cat. No. NB100-428; 1:1000) from Novus Biologicals and then with HRP-conjugated anti-rabbit, anti-mouse (Cat. No. 715-035-150; 1:3000) from Jackson Immunoresearch and anti-rat (Cat. No. 7077; 1:3000) from Cell Signaling Technology secondary antibodies for 1h at RT. The membranes were incubated with clarity Western ECL Substrate (Bio-Rad) for chemiluminescent signal detection using ChemiDoc camera (Bio-Rad).

5. Interactome analysis

5.1. In-gel tryptic digestion

Immunoprecipitated samples were separated by SDS-PAGE on 6% or 10% gels and stained with Pierce Silver Staining kit or SYPRO Ruby Protein Gel Stain (both from Thermo Fisher) according to manufacturer's protocol. Either selected protein bands from silver stained gels were excised, or entire lanes from SYPRO Ruby-stained gels were cut into ten contiguous pieces (done for both the ANKRD55 001 and control lanes). These gel slabs were then cut into small pieces and washed in Milli-Q water. Reduction and alkylation was achieved by incubation with dithiothreitol (DTT, 10 mM in 50 mM ammonium bicarbonate) at 56°C for 20 min, followed by an incubation in iodoacetamide (IA, 50 mM

in 50 mM ammonium bicarbonate) for another 20 min, in the dark. Gel pieces were dried and incubated with trypsin (12.5 µg/mL, in 50 mM ammonium bicarbonate) for 20 min on ice. After rehydration, the trypsin supernatant was discarded; gel bands were covered with 50 mM ammonium bicarbonate and incubated o/n at 37 ºC. After digestion, acidic peptides were further extracted with TFA 0.1% and dried in a RVC2 25 SpeedVac concentrator (Christ). The peptides were resuspended in 0.1% FA and sonicated for 5 min prior to their analysis.

5.2. *NanoLC-MS/MS and Data Analysis*

Identification of ANKRD55 interacting partners by nano-flow liquid chromatography coupled to tandem mass spectrometry (nLC-MS/MS) analysis was carried out by the Proteomics Platform from CICbioGUNE under the supervision of Dr. Felix Elortza. The functional analysis of the obtained results was also done in collaboration with Proteomics Platform.

Peptide mixtures obtained from trypsin digestion were separated by online nanoLC and analyzed by electrospray tandem mass spectrometry. Peptide separation was performed on a nanoAcquity UPLC system (Waters) connected to an LTQ Orbitrap XL mass spectrometer (Thermo Electron). Approximately 40% of each sample was loaded onto a Symmetry 300 C18 UPLC Trap column, 180 µm x 20 mm, 5 µm (Waters). The precolumn was connected to a BEH130 C18 column, 75 µm x 200 mm, 1.7 µm (Waters) equilibrated in 3% acetonitrile and 0.1% FA, and peptides were eluted at 300 nl/min using a 60 min linear gradient of 3–50% acetonitrile directly onto the nanoelectrospray ion source (Proxeon Biosystems). The mass spectrometer automatically switched between MS and MS/MS acquisition in DDA mode. Survey full scan MS spectra (m/z 400–2000) were acquired in the orbitrap with a resolution of 30,000 at m/z 400. The 6 most intense ions were sequentially subjected to collision-induced dissociation (CID) fragmentation in the linear ion trap. Precursors with charge states of 2 and 3 were specifically selected for CID. Collision-energy applied to each peptide was automatically normalized as a function of the m/z and charge state. Analyzed peptides were excluded for further analysis during 30 s using dynamic exclusion lists. Searches were performed using Mascot Search engine (Matrix Science) on Proteome Discoverer 1.2. software (Thermo Electron).

Carbamidomethylation of cysteines as fixed modification, oxidation of methionines as variable modification, 10 ppm of peptide mass tolerance, and 0.5 Da fragment mass tolerance were adopted as search parameters. 2 missed cleavages were allowed. Spectra were searched against a UNIPROT/Swissprot database restricted to *Homo sapiens*. A false discovery rate estimation procedure was applied for peptide identification (FDR <1%). Proteins identified with at least two peptides passing that cutoff were considered for further discussion.

For interactome analysis, three replicates were run for each sample type, and a spectral counting approach was carried out in order to select specifically enriched proteins. This approach relies on the counting of the MS/MS spectra matching to a certain protein[312]. The comparison of this value for a certain protein in two different samples provides an idea of its putative differential abundance. In our case, proteins with an ANKRD55 IP/Negative Ctrl IP spectral count ratio >5 in 3/3 replicates were selected for further analysis and discussion. In order to determine the relative abundance of each protein in each of the samples, a Normalized Spectral Abundance Factor (NSAF) approach was carried out[313]. The spectral counts for a protein were normalized to its length and further expressed as a % of the total of normalized spectral counts present in that particular sample.

5.3. Phosphopeptide enrichment

To enrich for phosphorylated peptides, titanium dioxide chromatography was performed essentially as described by Rigbolt and colleagues (2011)[314] with some modifications. Briefly, an equal volume of Titansphere material (GL Sciences) was mixed with a buffer containing 80% ACN / 1% TFA / 0.6M glycolic acid. The slurry was incubated for 20 min at RT with end-over-end rotation. Sample buffer was adjusted to a final concentration of 60% ACN / 1% TFA before incubation with the beads. Titansphere beads were added to the sample in a 0.6 mg TiO_2 / 100 peptide ug proportion and mixed 20 min at RT with end-over-end rotation. Finally, Titansphere material with bound phosphopeptides was washed three times with 150 µl 60% ACN / 1% TFA and transferred on top of a C8 disc (Empore) placed in 200 µL pipette-tip. Phopshopeptide samples were sequentially eluted using a buffer containing 5% NH_4OH and a buffer containing 25%

ACN/10% NH_4OH, dried in a RVC2 25 SpeedVac concentrator (Christ), and resuspended in 2% ACN/0.3% TFA before mass spectrometry analysis.

5.4. *Functional and pathway analysis*

The DAVID tool from the NIH (https://david.ncifcrf.gov/)[315] was used for Gene Ontology (GO) enrichment analysis. Results from the functional annotation clustering were used to select the interesting GO terms, and only results with $p<0.05$ for the modified Fisher's test were considered.

Ingenuity Pathway Analysis was used for more detailed characterization of the molecular events lying behind the differential protein patterns under analysis. The calculated p-values determine the probability that the association between proteins in the dataset and a given canonical pathway is explained by chance alone, based on a Fisher's exact test (p-value<0.05 being considered significant).

5.5. Validation of ANKRD55 partners

5.5.1. Validation by immunocytochemistry

For endogenous ANKRD55 subcelullar localization study, Jurkat, U937, and SH-SY5Y cells were seeded on coverslips coated with poly-D-lysine (Sigma-Aldrich). U937 cells were differentiated with 25 ng/ml PMA for 40 h and stimulated with 5 mg/ml LPS (both from Sigma- Aldrich) for 1, 4, and 24 h. Subsequently, cells were fixed in 4% paraformaldehyde-PBS for 20 min, incubated in 0.2% Triton X-100 in PBS for 30 min, blocked in 3% BSA (all from Sigma-Aldrich) in 0.2% Triton X-100 in PBS for 30 min, and stained with anti-ANKRD55 (1:500; Aviva Systems Biology) for 1 h at RT. Staining with Alexa Fluor 488–conjugated anti-rabbit (1:500; Invitrogen) was performed for 1 h at RT, followed by staining with DAPI (1:1000). To confirm specificity of ANKRD55 staining, a blocking peptide for the ANKRD55 Ab (Aviva Systems Biology) was used (see Western blot analysis section). Images were obtained using a Leica TCS CW SP8 STED Super-Resolution microscope with a 63x immersion objective and excitation wavelengths of 408 and 488 nm. For SH-SY5Y cells, we used a zoom factor ranging from 0.75 to 1.53.

For colocalization study of recombinant ANKRD55 and interacting partners, HEK293 cells were plated on coverslips coated with poly-D-lysine (Sigma-Aldrich) and transfected with ANKRD55 001 using MACsfectin Reagent (Miltenyi Biotec) according to manufacturer's instructions. 24 and 48h later and cells were fixed with 4% paraformaldehyde in PBS for 20 min at RT. Those coverslips destined to nuclear colocalization were subjected to antigen retrieval treatment incubating the cells with R-Universal buffer (Cat. No. AP0501-500; Aptum). Then cells were permeabilizated with 0.2% Triton X-100 in PBS for 10 min and blocked with 3% BSA (all from Sigma-Aldrich) in PBS for 30 min. Staining with anti-FLAG rabbit polyclonal (1:200) and mouse monoclonal (Cat. No. 66008-2-Ig; 1:100), anti-RPS3 (1:500), anti-TUBB (1:200), anti-VIM (1:500) from Proteintech, and anti-14-3-3 (pan) (1:200) from Cell Signaling for 1h at RT, followed by staining with Alexa Fluor 647 conjugated anti-rabbit (ab150075; 1:500) and DyLight 550 conjugated anti-mouse (ab98713; 1:500) from Abcam for 1h at RT and counter-stained with DAPI (1:500; Sigma-Aldrich). Images were obtained using a Leica TCS CW SP8 STED Super-Resolution microscope with a 63x immersion objective and excitation wavelengths of 405, 550 and 650nm.

5.5.2. Quantification of colocalization

For the cytoplasmatic colocalization, quantification was carried out using LAS AF (Leica Application Suite Advanced Fluorescence) software and Pearson's correlation coefficient. First, the background signal was extracted in all the studied images. Then, the Pearson's correlation coefficient values from 30 regions of interest (ROIs) were annotated, corresponding each ROI to a unique cell (n=30 cellular ROIs/condition). In transfected cells the ROIs were selected from transfected cells.

5.5.3. Statistical analysis

Analysis of the colocalization quantification was performed with GraphPad v.6. The data is represented as mean ± SEM and the Mann–Whitney test was applied. The level of significance was set at $p \leq 0.05$, with $p < 0.0001$ (****) as extremely significant, $p < 0.001$ (***) highly significant, $p < 0.01$ (**) very significant, $p < 0.05$ (*) significant and $p > 0.05$ (ns) not significant.

5.5.4. Validation by western blot

ANKRD55 immunoprecipitates were analysed by western blot using antibodies against target proteins, as described above (see section Western blot analysis).

6. Bioinformatics analysis

ANKRD55 phosphosites were identified *in silico* using several tools: (i) PhosphoSitePlus[316] and dbPAF[317] databases were consulted to look for any ANKRD55 phosphosites described in scientific literature or public databases, and (ii) 14-3-3-Pred web server (http://www.compbio.dundee.ac.uk/1433pred)[318] was used to predict putative 14-3-3 binding phosphosites on ANKRD55 by combining predictions from three different classifiers: ANN – Artificial Neural Network (cut-off = 0.55), PSSM – Position-Specific Scoring Matrix (cut-off = 0.80), and SVM – Support Vector Machine (cut-off = 0.25).

The ability of ANKRD55 for binding to nucleotides, DNA and RNA was analyzed using the following web-servers: (i) DRNApred (http://biomine.cs.vcu.edu/servers/DRNApred)[319] for the prediction of DNA and RNA-binding proteins and (ii) NsitePred (http://biomine.ece.ualberta.ca/nSITEpred/)[320] predicted binding residues from the protein sequence for ATP, ADP, AMP, GTP and GDP.

RESULTS

1. CHARACTERIZATION OF ANKRD55 ISOFORMS EXPRESSION AND SUBCELLULAR DISTRIBUTION

Previous mRNA-based functional studies in *ANKRD55* showed different expression patterns between the immune and nervous systems[265]. As mentioned before, *ANKRD55* transcripts 001, 005 and 007 were expressed in human PBMCs. $CD4^+$ T cells were identified as the main cellular source of the three transcripts in PBMCs, pointing to a role in cell-mediated immune responses. Moreover, transcript 001 was detected in SH-SY5Y and U937. Transcript 005 was detected in Jurkat and U937. No amplification of transcripts 002, 006 and 007 were found in any of the studied cell lines[265]. Considering this data, we extended the analysis to assess the protein expression and subcellular distribution of endogenous ANKRD55 in immune and non-immune cells, which will provide valuable information to elucidate protein functions and underlying mechanism of ANKRD55.

1.1. Endogenous ANKRD55 proteins locate primarily to the nucleus in primary $CD4^+$ T cells and Jurkat, U937, and SH-SY5Y human cell lines

We studied ANKRD55 endogenous expression and intracellular localization in primary $CD4^+$ T cells and Jurkat, U937, and SH-SY5Y human cell lines by confocal microscopy, nuclear enrichment through sucrose gradient ultracentrifugation, and biochemical fractionation to obtain cytosolic, membranous organelles, and nuclear fractions.

Considering the results about the expression of ANKRD55 in the primary mouse microglia cells[265], where a significant upregulation of the expression of ANKRD55 was observed after 24h LPS stimulation as assessed by ELISA and WB and the immunocytochemical data showed that it occurred mainly in the nucleus we study this possible effect in the human cell line U937 differentiated into macrophages upon PMA-treatment and LPS stimulation. The results did not show any differences in expression and localization of ANKRD55 between PMA/LPS-treated and PMA-treated U937 cells.

In PMA/LPS-treated U937 cells (Supplementary Figure 1) and PMA-treated U937 cells, as well as in SH-SY5Y cells, the cytosol along with nucleus to a greater extent, showed a strong signal for ANKRD55 reactivity (Figure 10 A, B). The nuclear signal, but not the cytoplasmic one, largely disappeared, as demonstrate by confocal microscopy, in the presence of a specific ANKRD55-blocking peptide coinciding with the immunogen used to raise the Ab. Subcellular ANKRD55 distribution was more difficult to interpret in undifferentiated U937 cells (data not shown) and Jurkat cells (Figure 10C) because of the relatively large nucleus. However, the presence of a nuclear ANKRD55 signal was confirmed in all cell types studied.

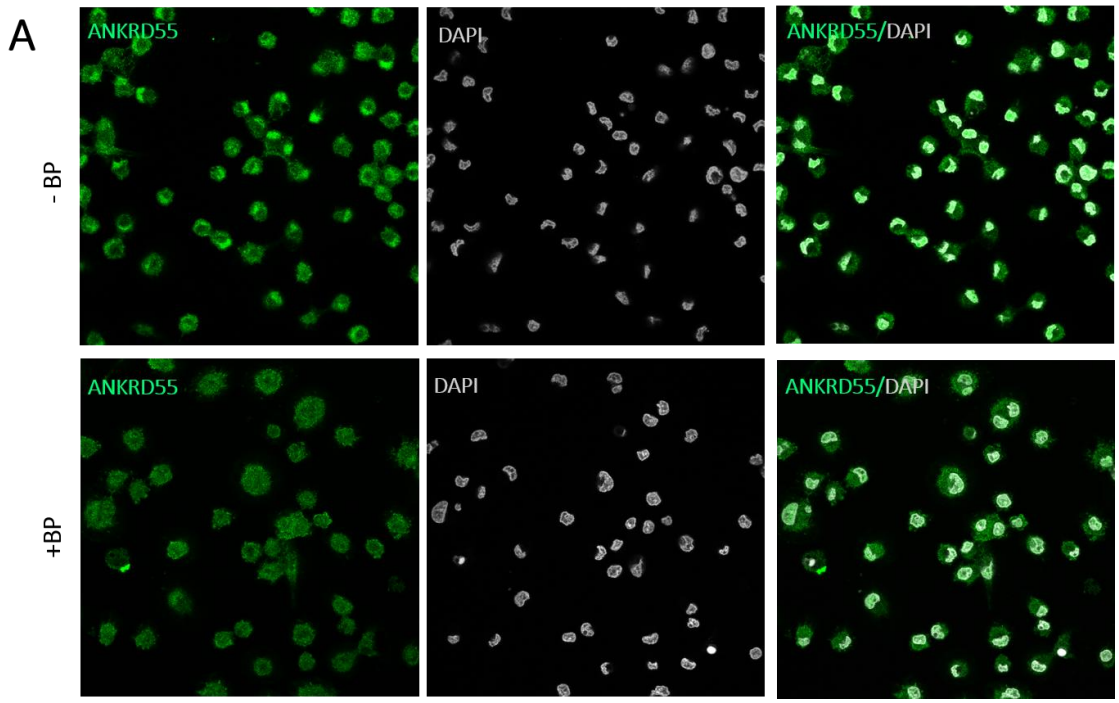


Figure 10. Intracellular localization of ANKRD55 in Jurkat, SH-SY5Y and PMA-treated U937 cells.

Inmunostaining for ANKRD55 (Aviva Systems Biology Ab; green) and nuclei (DAPI; grey) in (A) PMA-treated U937, (B) SHSY5Y, and (C) Jurkat cells analyzed by confocal microscopy. To confirm the specificity, immunostaining was performed either in the absence (-BP) or presence (+BP) of ANKRD55 blocking peptide.

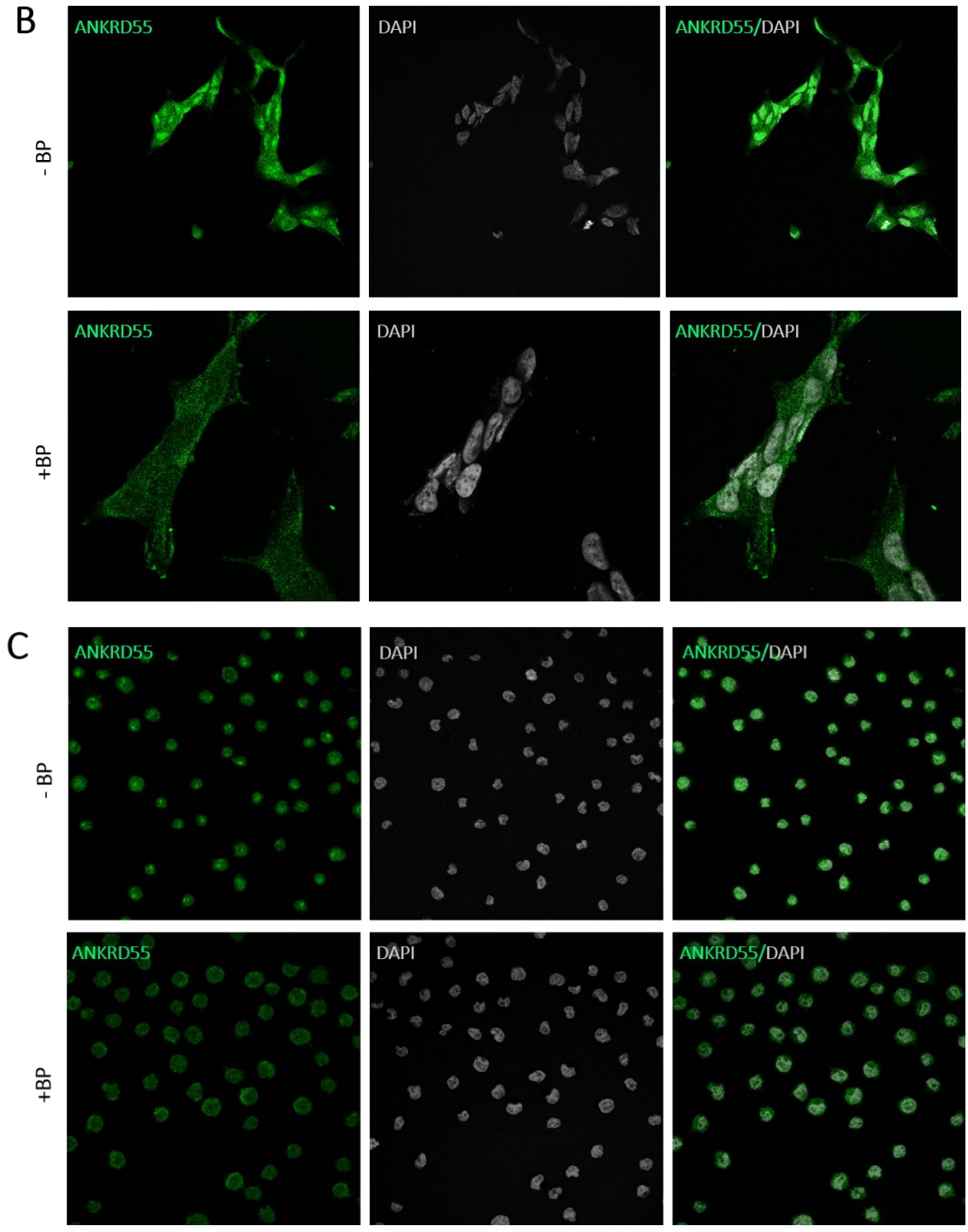


Figure 10. (Continued)

To validate the expression patterns of endogenous ANKRD55 observed using confocal microscopy, the subcellular localization of endogenous isoforms 001 and 005 of ANKRD55 in primary $CD4^+$ T cells and Jurkat, SH-SY5Y, and U937 cells was assessed in highly enriched nuclear fractions obtained by sucrose gradient centrifugation (Figure 11A). The data confirmed the location of isoforms 001 (68kDa) and 005 (37kDa) in the nucleus of U937 and Jurkat cells, as well as of primary $CD4^+$ T cells, and revealed a high abundance

of isoform 001 in the nucleus of SHSY5Y cells. Biochemical fractionation showed that isoform 001 of ANKRD55 was located mainly in the nucleus and membranous organelles of SH-SY5Y and Jurkat cell lines (Figure 11B). However, isoform 005 distributed between cytoplasm and membranous organelles in Jurkat cells, whereas in SH-SY5Y cells it was weakly detectable in cytosol and membrane organelle fractions but was virtually absent from the nuclei as shown through sucrose ultracentrifugation or biochemical fractionation (Figure 11 A, B). Instead, a smaller protein (30 kDa), not seen in any of the other cell types analyzed, was detectable in the nucleus but not in the cytosol and membrane organelles. Thus, this analysis revealed the predominantly nuclear location of ANKRD55 001 and 005.

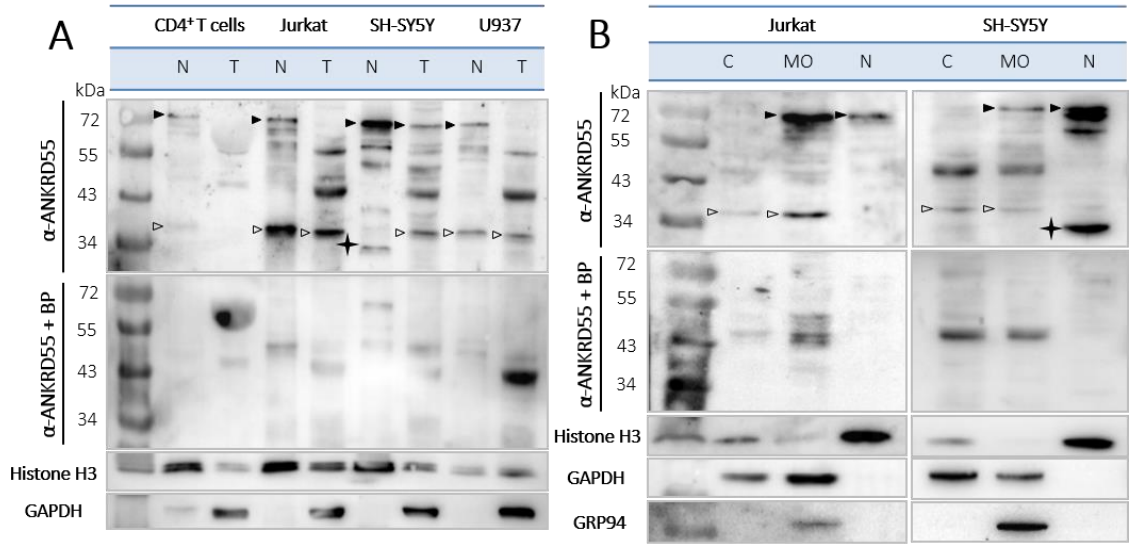


Figure 11. Subcellular localization of endogenous ANKRD55 isoforms in CD4⁺ T cells and Jurkat, SH-SY5Y, and U937 cells. (A) Nuclei from human CD4⁺ T cells (healthy donor) and Jurkat, SH-SY5Y, and U937 cell lines were isolated by sucrose gradient centrifugation. ANKRD55 was detected by WB in the absence or presence (+BP) of ANKRD55 blocking peptide in nuclear (N) and total (T) cell extracts. Enrichment for nuclear proteins was validated using histone H3 and GAPDH. (B) Jurkat and SH-SY5Y cells were biochemically fractionated to obtain cytosolic (C), membranous organelles (M), and nuclear (N) fractions. ANKRD55 was detected with specific Ab (Aviva Systems Biology) by WB in the absence or presence (+BP) of ANKRD55 blocking peptide (Aviva Systems Biology). Enrichment for subcellular fractions was validated using histone H3 (nuclear), GRP94 (membranous organelles), and GAPDH (cytosol). Specific bands corresponding to isoforms 001 and 005 are indicated by black and white arrowheads, respectively. Black cross in SH-SY5Y marks a lower-Mr nuclear immunoreactive band.

1.2. Overexpressed ANKRD55 isoforms are located in cytosol, membranous organelles and nucleus in HEK293 and HeLa human cell lines

To further study the subcellular localization of ANKRD55 isoforms, we transiently overexpressed human ANKRD55 001 and 005 isoforms in HEK293 and HeLa cells and fractionated the cells into three subcellular compartments. In transfected HEK293 cells,

similar levels of isoform 005 were detected in the three fractions. Isoform 001 (72 kDa) showed a more abundant signal than 005 (43 kDa) and, was as well present in the three compartments but slightly less abundant in the cytosol than in the other two fractions (Figure 12A). In HeLa cells, the isoform 001 from the cytosolic enriched fraction showed a stronger signal than that of the other two fractions. Isoform 005 detection was slightly weaker in nuclei than in the rest of the fractions (Figure 12B). Moreover, truncated immunoreactive products were present, probably generated by specific proteolytic events from the intact isoform proteins. These included (i) a 55-60-kDa band in HEK293 and HeLa cells in isoform 001-transfected cells present in all 3 compartments though at levels differing from the full-length proteins, (ii) 20-kDa band in membranous organelles of HEK293 cells, and (iii) a series of bands with Mr<34 kDa especially enriched in the cytosol of isoform-005 transfected HeLa cells and membranous organelles and cytosol of HEK293 cells. Additionally, the results showed the endogenous isoform 001 in both cell lines demonstrating the capacity of HEK293 and Hela cells to produce ANKRD55 and therefore to engage its biological pathway(s). Thus, western blot analysis of biochemically fractionated cells revealed overexpression of the isoforms 001 and 005 of ANKRD55 in all three analyzed subcellular compartments of both cell lines.

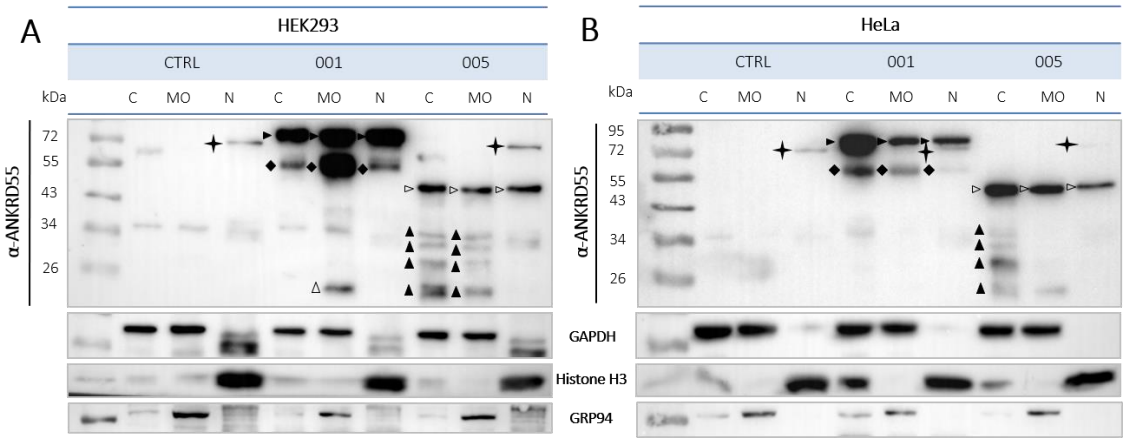


Figure 12. Subcellular localization of recombinant ANKRD55 isoforms in HEK293 and HeLa cells. (A) HEK293 and (B) HeLa cells were transfected with isoforms (001) and (005) of ANKRD55 or not (negative control) and these cells were cultivated under the same conditions, but without transfection agent. Cells were fractionated at 48h after transfection. ANKRD55 was detected with specific Ab (Sigma-Aldrich) by WB. Enrichment for subcellular fractions was verified using histone H3 as nuclear marker (N), GRP94 as ER marker (MO) and GAPDH as cytosolic marker (C). Specific bands corresponding to overexpressed isoforms 001 and 005 are indicated by black and white arrowheads, respectively. Black cross marks endogenous ANKRD55 isoform 001. Black diamonds, and black and white triangles show three different truncated immunoreactive products.

2. FUNCTIONAL STUDY OF ANKRD55 THROUGH ITS INTERACTOME ANALYSIS

To identify ANKRD55 interactomes, we performed an IP of tagged recombinant ANKRD55 and its interacting partners followed by nLC-MS/MS analysis (Figure 13).

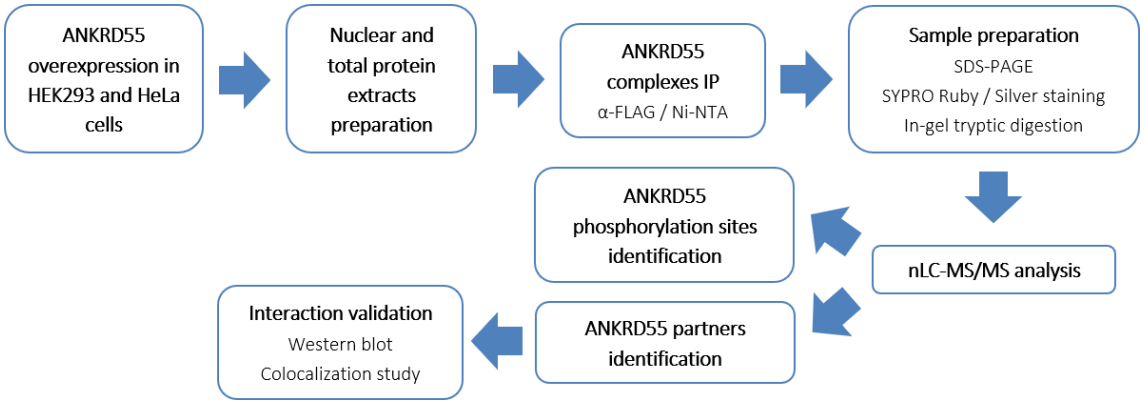


Figure 13. Overall experimental design and analysis scheme for the identification of ANKRD55-interacting partners.

2.1. ANKRD55 isoform 001 was overexpressed and immunoprecipitated from HEK293 and HeLa cells

Initially, we planned to overexpress C-terminally FLAG-myc-tagged human ANKRD55 001 and 005 by electroporation in Jurkat cells. However, we could not obtain a sufficient amount of recombinant proteins to perform IP and mass spectrometry analysis, due to either a low electroporation efficiency or to low levels of recombinant protein produced (Figure 14A). Therefore, we overexpressed and IPed FLAG/myc- and his-tagged ANKRD55 isoforms from HEK293 and HeLa cells, which as we showed in Figure 12, produce much higher amounts of recombinant ANKRD55 (Figure 14 B-H). Briefly, to express the recombinant proteins, pCMV6 containing FLAG/myc tagged ANKRD55 001 cDNA was purchased from OriGene, ANKRD55 005 cDNA was subcloned into pCMV6 vector including FLAG/myc tags and pCMV6 including ANKRD55 transcripts were modified by annealed oligo cloning to exchange myc for his-tag.

FLAG/myc-tagged ANKRD55 constructs were overexpressed in HEK293 cells and ANKRD55 complexes were IPed from highly enriched nuclear fractions, membranous organelles fraction and total protein extracts using anti-FLAG affinity resin at 48h post-

transfection. This was done in the presence or absence of DSP crosslinker treatment so as to verify whether any chemical modification of amines in the side chain of lysine (K) residues by the crosslinker in the FLAG-tag (which contains two such, DYKDDDK) would compromise purification efficiency based on anti-FLAG IP Recombinant ANKRD55 isoforms and interacting proteins were eluted with acid buffer (method A described in immunoprecipitation section from materials and methods). The elution step was subjected to an optimization process using FLAG peptide and different alkaline and acid elution buffers. The results showed that acid elutions at pH 2 in absence or presence of SDS were the best options for our IP (Supplementary Figure 2). ANKRD55 isoform 001 was efficiently IPed from total cell lysates as well as from the three cell fractions as shown by western blot (Figure 14 B, C, E). In contrast, recombinant isoform 005 was less efficiently IPed from HEK293 cells (Figure 14 B, C). As expected, endogenous ANKRD55 was not detected in the control IP. DSP treatment, which stabilizes transient protein interactions, did not block the interaction of ANKRD55 and FLAG resin even if the signal was somewhat decreased with the treatment (Figure 14 B, C). The overexpression and IP of FLAG-tagged ANKRD55 isoform 001 was validated in total protein extracts and membranous organelles fraction of HeLa cells by WB with anti-FLAG (Figure 14 D, E).

His-tagged ANKRD55 constructs were also overexpressed in HEK293 cells and ANKRD55 complexes were IP from total protein extracts using Ni^{2+}–NTA agarose 48h post-transfection under non-denaturing conditions. IP of ANKRD55 isoform 001 was confirmed by WB using anti-His, whereas we could hardly detect recombinant isoform 005. Thus, IP of isoform 005 under non-denaturing conditions was unsuccessful (Figure 14F). The Ni^{2+}-NTA-based procedure was modified to include DSP crosslinker treatment and carried out under denaturing conditions by adding high molar concentrations of urea, a chaotropic reagent. This results in disruption of tertiary structure, full unfolding of the protein and enhances exposure of the his-tag to improve the efficiency of the IP procedure. The results demonstrated an improvement of the IP efficiency under denaturing conditions, but the signal was decreased with the DSP treatment (Figure 14G). The modified his-tagged ANKRD55 isoform 001 IP was also validated in total protein extracts of HeLa cells by WB with anti-ANKRD55 (Figure 14H). These data demonstrated that tagged ANKRD55

isoform 001 more so than isoform 005 can be efficiently and specifically immunoprecipitated from different cell extracts.

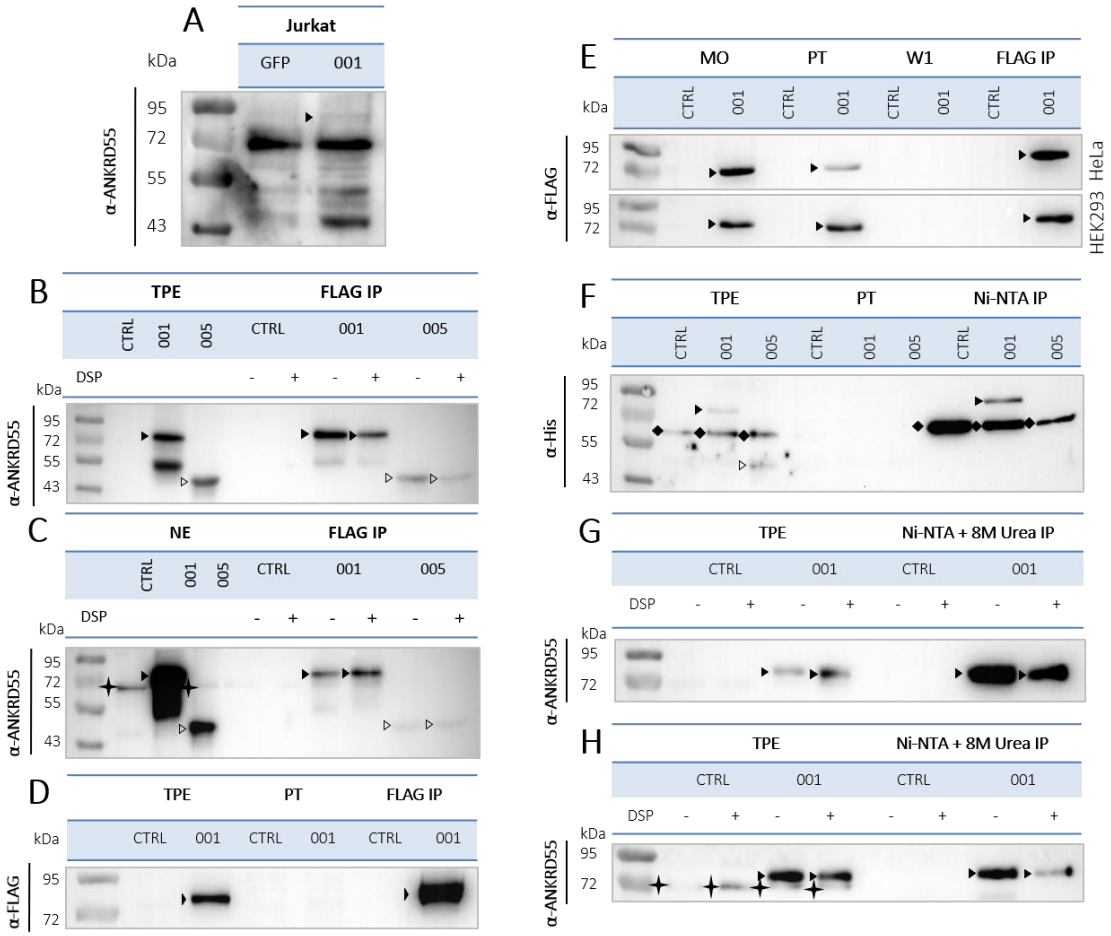


Figure 14. Overexpression and immunoprecipitation of recombinant ANKRD55 isoforms. (A) Jurkat cells were electroporated with isoform 001 and GFP was used as negative control. Isoform 001 was detected with specific Ab (Sigma-Aldrich) by WB. (B, C, E) HEK293 and (D, E) HeLa cells were transfected with FLAG-tagged isoforms (001) and (005) of ANKRD55. Cells were treated +/- DSP, when indicated, and IP was performed on enriched nuclear fractions (NE), total protein extracts (TPE) or membranous organelles fraction (MO) after 48h of transfection. Recombinant ANKRD55 isoforms were detected in NE, TPE, MO and eluted fractions (FLAG IP) with specific Ab (Sigma-Aldrich) and FLAG Ab by WB. (F, G) HEK293 and (H) HeLa cells were transfected with his-tagged isoforms 001 or 005 of ANKRD55. Cells were treated +/- DSP, when indicated, and ANKRD55 was IPed from TPE after 48h of transfection under non-denaturing (Ni-NTA IP, F) or denaturing (8M urea Ni-NTA IP, G & H) conditions. Recombinant ANKRD55 isoforms were detected in TPE and eluted fractions with specific Ab (Aviva Systems Biology) or His Ab by WB. As negative control cells were cultivated under the same conditions, but without transfection agent (CTRL). Specific bands corresponding to overexpressed isoforms 001 and 005 are indicated by black and white arrowheads, respectively. Black cross marks endogenous ANKRD55 isoform 001. Diamond marks a specific protein interacting with both the Ni^{2+}-NTA agarose and the anti-histidine Ab in (F). PT, pass through fraction; W1, first washing step.

2.2. Identification of ANKRD55 001-interacting proteins

To determine the interacting partners of ANKRD55, two methods were used for mass spectrometry-based proteomics studies. First, bands specifically detected in the ANKRD55 IPed samples were cut and trypsin-digested in order to identify proteins responsible for those differences. Although this approach is highly valuable for the identification of the most abundant interactors, many putative interacting proteins may get lost since we only cut a limited number of bands. Therefore, a second approach was carried out where the whole gel lanes were cut, digested and analyzed by nLC-MS/MS. The substraction of specific ANKRD55-interacting proteins was carried out at mass spectrometry level. FLAG- and his-tagged ANKRD55 IPs from nuclear fractions, membranous organelles fraction and total protein extracts were loaded onto a SDS-PAGE and silver stained. The results showed several specific bands in FLAG-ANKRD55 001 and 005 IP that were not present in control samples. For his-ANKRD55 001, only a single specific band was found that corresponded to the overexpressed ANKRD55 (Figure 15). Bands corresponding to the Mr of recombinant FLAG-ANKRD55 isoforms and interacting partners not present in the control samples were cut, trypsin digested and identified by nLC-MS/MS analysis (Table 6).

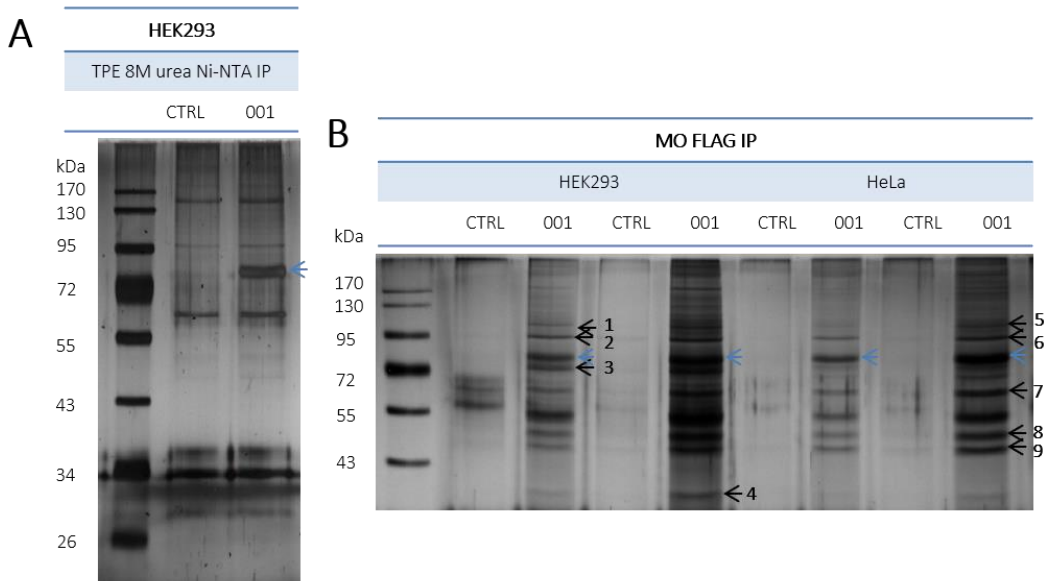


Figure 15. Silver-staining analysis of ANKRD55 complexes. Silver-stained SDS-PAGE showing ANKRD55 IP with (A) Ni^{2+}-NTA agarose under denaturing conditions from total protein extracts (TPE) and FLAG resin from (B) membranous organelles fraction (MO), (C) total protein extracts (TPE), and (D) nuclear fractions (NE) of HEK293 and HeLa cells expressing FLAG/his-tagged ANKRD55 isoforms (001) and (005). As negative control cells were cultivated under the same conditions, but without transfection agent (CTRL). The proteins identified using nLC-MS/MS are listed in **Table 6** and marked by black arrowheads. The recombinant ANKRD55 001 and 005 are indicated by blue and red arrowheads, respectively.

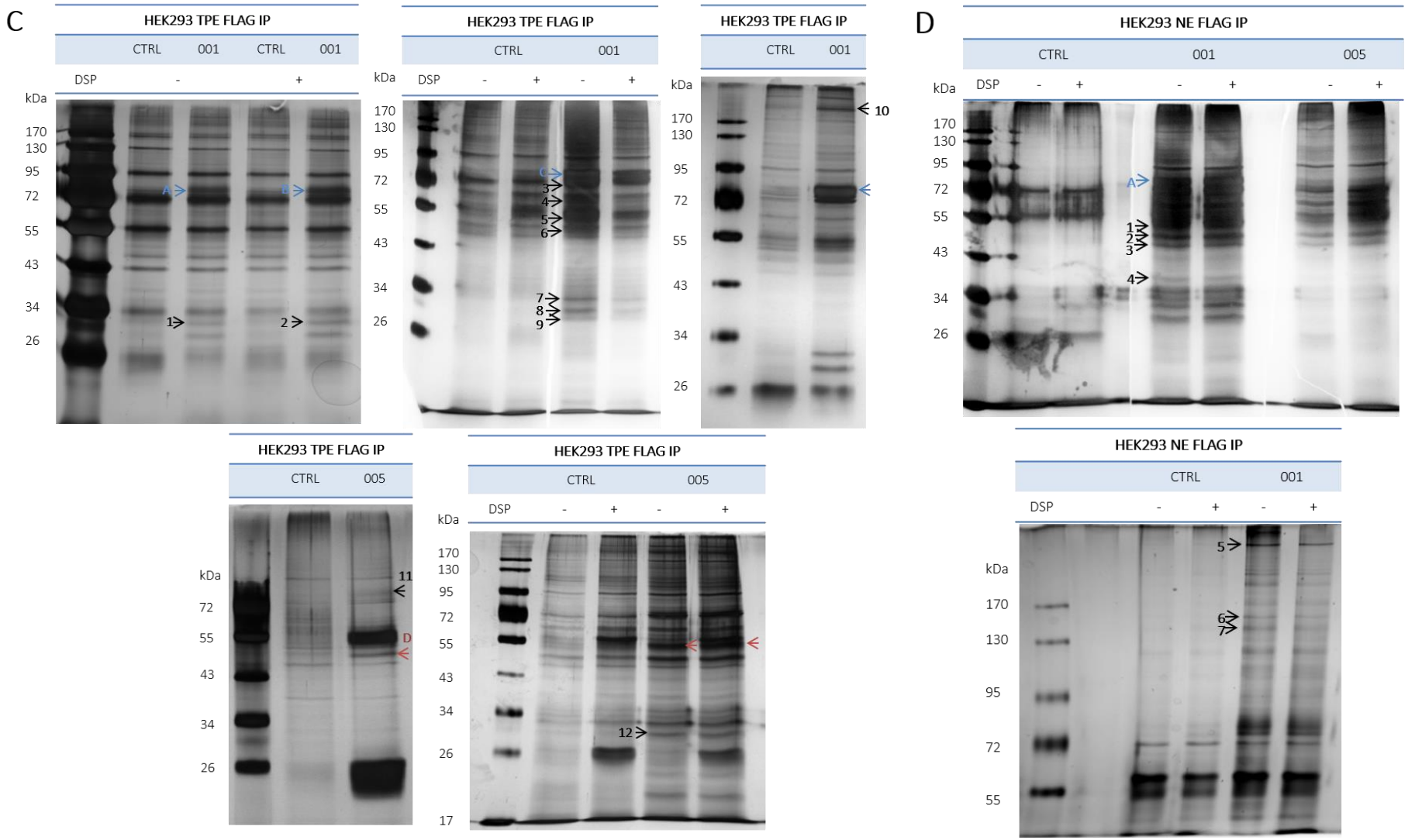


Figure 15. (Continued)

Table 6. Identified ANKRD55-interacting proteins from three different cell extracts by silver staining and nLC-MS/MS analysis

	IP	Band	Gene symbol	Accession	MW (kDa)	Protein name	Peptides
TOTAL PROTEIN EXTRACTS	HEK293 001	A	ANKRD55	Q3KP44	68.41	Ankyrin repeat domain-containing protein 55	12
			HSPA9	P38646	73.68	Stress-70 protein, mitochondrial	4
		B	ANKRD55	Q3KP44	68.41	Ankyrin repeat domain-containing protein 55	9
			HSPA9	P38646	73.68	Stress-70 protein, mitochondrial	4
		1	YWHAE	P62258	29.17	14-3-3 protein epsilon	4
		2	YWHAE	P62258	29.17	14-3-3 protein epsilon	3
		C	ANKRD55	Q3KP44	68.41	Ankyrin repeat domain-containing protein 55	19
			HSPA5	P11021	72.28	78 kDa glucose-regulated protein	8
		3	LMNB1	P20700	66.408	Lamin-B1	21
			HSPA1A	P0DMV8	70	Heat shock 70 kDa protein 1A/1B	16
			LMNB2	Q03252	69.948	Lamin B2	15
			RPN1	P04843	68.569	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	4
		4	VIM	P08670	53.65	Vimentin	13
			CCT4	P50991	57.924	T-complex protein 1 subunit delta	7
			PKM	P14618	57.937	Pyruvate kinase PKM	3
			CCT7	Q99832	59.367	T-complex protein 1 subunit eta	2
			TCP1	P17987	60.344	T-complex protein 1 subunit alpha	2
		5	TUBB	P07437	49.6	Tubulin beta chain	10
			TUBA1B	P68363	50.1	Tubulin alpha-1B chain	9
			ATP5B	P06576	56.56	ATP synthase subunit beta, mitochondrial	7
		6	EEF1A1	P68104	50.14	Elongation factor 1-alpha 1	4
			TUFM	P49411	49.542	Elongation factor Tu, mitochondrial	4
			EIF4A1	P60842	46.15	Eukaryotic initiation factor 4A-I	2
		7	YWHAE	P62258	29.174	14-3-3 protein epsilon	13
			GNB2L1	P63244	35.07	Guanine nucleotide-binding protein subunit beta-2-like 1 Receptor of activated protein C kinase 1	4
			RPS3	P23396	26.688	40S ribosomal protein S3	4
		8	YWHAZ	P63104	27.745	14-3-3 protein zeta/delta	25
			PHB	P35232	29.8	Prohibitin	3
			SLC25A5	P05141	32.852	ADP/ATP translocase 2	6
			YWHAE	P62258	29.174	14-3-3 protein epsilon	3
			SLC25A6	P12236	32.866	ADP/ATP translocase 3	5
			YWHAH	Q04917	28.219	14-3-3 protein eta	3
			YWHAB	P31946	28.082	14-3-3 protein beta/alpha	3
			YWHAQ	P27348	27.76	14-3-3 protein theta	2
			YWHAG	P61981	28.3	14-3-3 protein gamma	3
		9	YWHAZ	P63104	27.745	14-3-3 protein zeta/delta	6
			YWHAG	P61981	28.3	14-3-3 protein gamma	5
			PHB	P35232	29.8	Prohibitin	3
			YWHAH	Q04917	28.219	14-3-3 protein eta	6
			YWHAQ	P27348	27.76	14-3-3 protein theta	5
			SLC25A5	P05141	32.852	ADP/ATP translocase 2	6

		SLC25A6	P12236	32.866	ADP/ATP translocase 3	6
		YWHAB	P31946	28.082	14-3-3 protein beta/alpha	6
	10	CAD	P27708	242.984	CAD protein	18
		MYH10	P35580	228.99	Myosin-10	4
		USP9X	Q93008	292.28	Probable ubiquitin carboxyl-terminal hydrolase FAF-X	3
		GCN1L1	Q92616	292.75	Translational Activator GCN1	3
		PRKDC	P78527	469.089	DNA-dependent protein kinase catalytic subunit	3
		FASN	P49327	273.427	Fatty acid synthase	2
	11	HSPA5	P11021	72.28	78 kDa glucose-regulated protein	2
HEK293 005	D	ANKRD55	Q3KP44	36.9	Isoform 2 Ankyrin repeat domain-containing protein 55	13
		EEF1A1	P68104	50.1	Elongation factor 1-alpha 1	2
	12	YWHAZ	P63104	27.745	14-3-3 protein zeta/delta	5
	A	ANKRD55	Q3KP44	68.41	Ankyrin repeat domain-containing protein 55	13
		HSPA5	P11021	72.28	78 kDa glucose-regulated protein	2
	1	TUBB	P07437	49.60	Tubulin beta chain	12
		TUBB4B	P68371	49.80	Tubulin beta-4B chain	12
		TUBB4A	P04350	49.60	Tubulin beta-4A chain	10
		TUBA1B	P68363	50.10	Tubulin alpha-1B chain	9
		ATP5B	P06576	56.56	ATP synthase subunit beta, mitochondrial	5
	2	EIF4A1	P60842	46.15	Eukaryotic initiation factor 4A-I	5
		EEF1A1	P68104	50.14	Elongation factor 1-alpha 1	4
		DDOST	P39656	50.801	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	5
		EIF4A3	P38919	46.871	Eukaryotic initiation factor 4A-III	5
		RPL3	P39023	46.109	60S ribosomal protein L3	3
		TUBB	P07437	49.60	Tubulin beta chain	3
		EEF1G	P26641	50.119	Elongation factor 1-gamma	3
	3	ACTB	P60709	41.737	Actin, cytoplasmic 1	10
		CKB	P12277	42.644	Creatine kinase B-type	5
		OAT	P04181	48.53	Ornithine aminotransferase, mitochondrial	3
		CS	O75390	51.712	Citrate synthase, mitochondrial	3
		PAICS	P22234	47.079	Multifunctional protein ADE2	3
	4	GAPDH	P04406	36.05	Glyceraldehyde-3-phosphate dehydrogenase	3
		RPLP0	P05388	36.05	60S acidic ribosomal protein P0	3
	5	PRKDC	P78527	469.089	DNA-dependent protein kinase catalytic subunit	23
	6	SMC1A	Q14683	143.23	Structural maintenance of chromosomes protein 1A	6
	7	SMC3	Q9UQE7	141.54	Structural maintenance of chromosomes protein 3	13
		DHX9	Q08211	140.95	ATP-dependent RNA helicase A	6

NUCLEAR EXTRACTS

HEK293 001

MEMBRANOUS ORGANELLES	HEK293 001	1	HSP90B1	P14625	92.469	Endoplasmin	6
			GANAB	Q14697	106.874	Neutral alpha-glucosidase AB	4
		2	HSP90AB1	P08238	83.264	Heat shock protein HSP 90-beta	26
			HSP90AA1	P07900	84.66	Heat shock protein HSP 90-alpha	23
			MCM5	P33992	82.286	DNA replication licensing factor MCM5	3
		3	HSPA8	P11142	70.89	Heat shock cognate 71 kDa protein	21
			HSPA1A	P0DMV8	70.052	Heat shock 70 kDa protein 1A	17
			HSPA9	P38646	73.68	Stress-70 protein, mitochondrial	8
			SLC25A13	Q9UJS0	74.176	Calcium-binding mitochondrial carrier protein Aralar2	5
			HSPA5	P11021	72.28	78 kDa glucose-regulated protein	3
			XRCC6	P12956	69.843	X-ray repair cross-complementing protein 6	3
			IFT74	Q96LB3	69.239	Intraflagellar transport protein 74 homolog	3
		4	GAPDH	P04406	36.05	Glyceraldehyde-3-phosphate dehydrogenase	13
	HeLa 001	5	HSP90B1	P14625	92.469	Endoplasmin	9
			EEF2	P13639	95.338	Elongation factor 2	4
			GANAB	Q14697	106.874	Neutral alpha-glucosidase AB	4
		6	HSP90AB1	P08238	83.264	Heat shock protein HSP 90-beta	16
			HSP90AA1	P07900	84.66	Heat shock protein HSP 90-alpha	10
		7	PKM	P14618	57.937	Pyruvate kinase PKM	14
			HSPD1	P10809	61.055	60 kDa heat shock protein, mitochondrial	7
			TCP1	P17987	60.344	T-complex protein 1 subunit alpha	4
		8	EIF4A1	P60842	46.154	Eukaryotic initiation factor 4A-I	5
			EEF1A1	P68104	50.14	Elongation factor 1-alpha 1	3
		9	TUFM	P49411	49.542	Elongation factor Tu, mitochondrial	10
			EIF4A1	P60842	46.154	Eukaryotic initiation factor 4A-I	6

Some of the identified interacting proteins belong to 14-3-3 family (YWHAE, YWHAZ, YWHAH, YWHAB, YWHAQ, YWHAG), heat shock proteins (HSPA9, HSPA5, HSPA1A, HSP90B1, HSP90AB1, HSP90AA1, HSPA8, HSPD1), tubulin superfamily (TUBB, TUBA1B, TUBB4B, TUBB4A), CCT complex (CCT4, CCT7, TCP1) and initiation and elongation phases of eukaryotic translation (EEF1A1, EIF4A1, TUFM, EIF4A3, EEF1G, EEF2), among others. YWHAE was identified in presence or absence of cross-linker as well as in two independent experiments from HEK293 total protein extracts. HSPA5 was detected in the three different HEK293 cell extracts. PRKDC was found in three independent experiments from HEK293 nuclear extracts. Several tubulin isoforms were found in nuclear and total HEK293 cell extracts. HSP90AB1, HSP90AA1, and HSP90B1 were identified in HEK293 and HeLa membranous organelles extracts. Thus, what clearly emerges from this analysis is that ANKRD55 interacts with an unusual combination of functionally very distinct proteins

on the basis of which no clear conclusion can be made on its precise biological role. However, of these, 14-3-3 proteins appear to present the best-resolved interactors in silver-stained gels (various examples in Figure 15).

2.3. The ANKRD55 interactome

In our first approach, several interacting candidates were identified corresponding to discrete bands IPed with ANKRD55 from three different cell extracts. Therefore, we decided to compare and analyze in more detail ANKRD55 and control IPs to search for and identify further interacting partners and to confirm the ones previously identified. In essence, rather than focusing on single visualized gel bands as we did before, our goal here was to analyze the full lanes corresponding to ANKRD55 001 or mock (not transfected cells) eluates. Indeed, low abundant ANKRD55-interacting proteins could be present that are not resolved as single bands and may be missed by the first method. Together such bands tend to produce a darker stainable protein smear in ANKRD55 compared to mock lanes (several such examples can be seen in Figure 15). By subtracting the 'mock proteome' from the ANKRD55 proteome, a specific and comprehensive ANKRD55 interactome can be deduced. For this second method, we focused our proteomic analysis on FLAG-tagged ANKRD55 isoform 001, which was more efficiently IPed than isoform 005, and showed highest number of potential interacting partners. FLAG-ANKRD55 isoform 001 was overexpressed in HEK293 cells and IPed from total protein extracts and nuclear extracts in triplicates (three independent experiments each that were performed on distinct days from distinct batches of cells). WB with FLAG antibody was performed to verify the quality of ANKRD55 overexpression and IP (Figure 16A). ANKRD55 complexes were then separated by SDS-PAGE and stained using SYPRO Ruby. The full gel lanes starting from the boundary of the stacking gel with the resolving gel up to the gel front were cut into ten pieces (one contained ANKRD55 001 or its control), trypsin digested, and identified by nLC-MS/MS (Figure 16B).

To build the ANKRD55 interactome, all potential ANKRD55 interacting partners were filtered according to their spectral count enrichment. Spectral counting accounts for the total number of MS/MS spectra matching to peptides that belong to a certain protein and provides a valuable parameter for relative protein abundance comparison (see materials

and methods for more details). We reproducibly identified enrichment of 148 proteins in total protein extracts and 22 proteins in nuclear extracts over the three replicates, based on stringent selection criteria including both at least two unique peptides and a 5-fold increase in ANKRD55 spectral counts in the ANKRD55 IP compared to control IP in all three replicates. These proteins were selected for further functional analysis and validation (Figure 16C, Table 7). The average NSAF of these proteins was calculated in order to estimate their relative abundance within the IPed samples (see materials and methods for more details). Although not considered for further analyses, an additional, less robust, dataset of putative ANKRD55 interactors was substracted in addition to the set of reliably enriched proteins. For this purpose, proteins enriched with a more relaxed criteria of ANKRD55 IP/negative control IP spectral count ratio of at least 5 in only 2 out of 3 replicates were selected, providing additional 144 putative-ANKRD55 partners in total protein extracts and 34 in nuclear extracts (Supplementary Table 2).

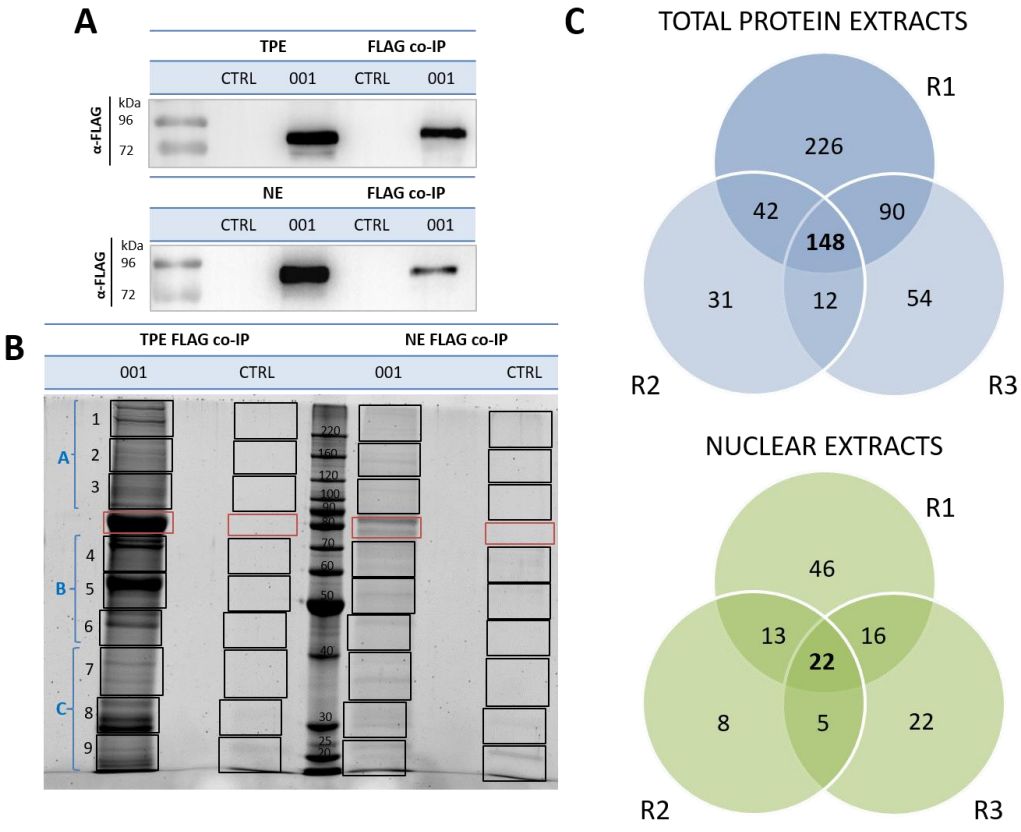


Figure 16. Isolation and identification of ANKRD55 interactome components by nLC-MS/MS. (A) Recombinant ANKRD55 isoform (001) was detected in nuclear extracts (NE), total protein extracts (TPE) and eluted fractions (FLAG IP) with FLAG Ab by WB. As negative control cells were cultivated under the same conditions, but without transfection agent (CTRL). (B) A SYPRO Ruby-stained SDS-PAGE showing analysis strategy of ANKRD55 complexes IP with FLAG resin from NE and TPE of HEK293 cells expressing isoform 001. WB and SDS-PAGE are representative of three independent biological replicates. (C) Venn diagram represents total number and common significant interactors in each replicate of ANKRD55 complexes IP from TPE and NE.

Table 7. Identification of ANKRD55-interacting partners from nuclear and total protein extracts of HEK293 cells by nLC-MS/MS

	Gene symbol	Accession	MW (kDa)	Protein	NSAF	
					ANKRD55	CTRL
TOTAL PROTEIN EXTRACTS	YWHAE	P62258	29.17	14-3-3 protein epsilon	2.64	0.21
	YWHAB	P31946	28.082	14-3-3 protein beta/alpha	1.14	0.00
	SLC25A5	P05141	32.852	ADP/ATP translocase 2	1.08	0.06
	YWHAH	Q04917	28.219	14-3-3 protein eta	0.97	0.00
	RPS3	P23396	26.688	40S ribosomal protein S3	0.81	0.26
	YWHAG	P61981	28.3	14-3-3 protein gamma	0.81	0.00
	TUBB6	Q9BUF5	49.857	Tubulin beta-6 chain	0.76	0.00
	SLC25A11	Q02978	34.062	Mitochondrial 2-oxoglutarate/malate carrier protein	0.66	0.08
	NTPCR	Q9BSD7	20.713	Cancer-related nucleoside-triphosphatase	0.57	0.00
	RPS15A	P62244	14.84	40S ribosomal protein S15a	0.53	0.00
	SLC25A13	Q9UJS0	74.176	Calcium-binding mitochondrial carrier protein Aralar2	0.47	0.00
	RPL9	P32969	21.863	60S ribosomal protein L9	0.46	0.00
	ARF4	P18085	20.511	ADP-ribosylation factor 4	0.46	0.00
	CCT4	P50991	57.924	T-complex protein 1 subunit delta	0.43	0.12
	PGAM5	Q96HS1	32.004	Serine/threonine-protein phosphatase PGAM5, mitochondrial	0.40	0.06
	MCM7	P33993	81.308	DNA replication licensing factor MCM7	0.39	0.00
	PCBP2	Q15366	38.58	Poly(rC)-binding protein 2	0.39	0.00
	PCBP1	Q15365	37.498	Poly(rC)-binding protein 1	0.37	0.00
	TIMM50	Q3ZCQ8	39.646	Mitochondrial import inner membrane translocase subunit TIM50	0.37	0.00
	RPL23	P62829	14.865	60S ribosomal protein L23	0.34	0.00
	PPP3CA	Q08209	58.688	Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	0.34	0.00
	PCNA	P12004	28.769	Proliferating cell nuclear antigen	0.33	0.00
	CFL1	P23528	18.502	Cofilin-1	0.31	0.00
	CCT8	P50990	59.621	T-complex protein 1 subunit theta	0.31	0.05
	ATP5O	P48047	23.277	ATP synthase subunit O, mitochondrial	0.31	0.00
	PPP2R1A	P30153	65.309	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	0.31	0.03
	RPS9	P46781	22.591	40S ribosomal protein S9	0.31	0.00
	DNAJA1	P31689	44.868	DnaJ homolog subfamily A member 1	0.30	0.00
	SLC25A12	O75746	74.762	Calcium-binding mitochondrial carrier protein Aralar1	0.30	0.00
	HSPA9	P38646	73.68	Stress-70 protein, mitochondrial	0.29	0.07
	SRPRB	Q9Y5M8	29.702	Signal recognition particle receptor subunit beta	0.28	0.00
	CAD	P27708	242.984	CAD protein	0.27	0.00
	RPS3A	P61247	29.945	40S ribosomal protein S3a	0.27	0.00
	GNB2L1	P63244	35.077	Guanine nucleotide-binding protein subunit beta-2-like 1	0.27	0.00
	SFXN1	Q9H9B4	35.619	Sideroflexin-1	0.27	0.06
	DPM1	O60762	29.634	Dolichol-phosphate mannosyltransferase subunit 1	0.27	0.00
	ARL1	P40616	20.418	ADP-ribosylation factor-like protein 1	0.27	0.00
	SLC25A1	P53007	34.013	Tricarboxylate transport protein, mitochondrial	0.27	0.00
	RPL10	P27635	24.604	60S ribosomal protein L10	0.27	0.00
	PRKDC	P78527	469.089	DNA-dependent protein kinase catalytic subunit	0.26	0.01

	RPLP0	P05388	36.05	60S acidic ribosomal protein P0	0.25	0.00
	RPN1	P04843	68.569	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	0.24	0.04
	PPP2R2A	P63151	51.692	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	0.24	0.00
	RPS2	P15880	31.324	40S ribosomal protein S2	0.24	0.00
	SLC25A3	Q00325	40.095	Phosphate carrier protein, mitochondrial	0.24	0.05
	VDAC2	P45880	31.567	Voltage-dependent anion-selective channel protein 2	0.24	0.00
	TTC26	A0AVF1	64.178	Intraflagellar transport protein 56	0.24	0.00
	EIF4A3	P38919	46.871	Eukaryotic initiation factor 4A-III	0.23	0.00
	DDX3X	O00571	73.243	ATP-dependent RNA helicase DDX3X	0.23	0.00
	C1QBP	Q07021	31.362	Complement component 1 Q subcomponent-binding protein, mitochondrial	0.22	0.00
	RCN2	Q14257	36.876	Reticulocalbin-2	0.22	0.00
	RBM39	Q14498	59.38	RNA-binding protein 39	0.22	0.00
	CDK5	Q00535	33.304	Cyclin-dependent-like kinase 5	0.22	0.00
	OAT	P04181	48.53	Ornithine aminotransferase, mitochondrial	0.21	0.00
	AIFM1	O95831	66.901	Apoptosis-inducing factor 1, mitochondrial	0.21	0.00
	RAB5C	P51148	23.483	Ras-related protein Rab-5C	0.20	0.00
	RPL7A	P62424	29.996	60S ribosomal protein L7a	0.20	0.00
	ATP1A1	P05023	112.896	Sodium/potassium-transporting ATPase subunit alpha-1	0.20	0.00
	SERPINH1	P50454	46.44	Serpin H1	0.20	0.00
	CSE1L	P55060	110.417	Exportin-2	0.20	0.00
	IRS4	O14654	133.768	Insulin receptor substrate 4	0.19	0.00
	RARS	P54136	75.379	Arginine--tRNA ligase, cytoplasmic	0.19	0.00
	IFT74	Q96LB3	69.239	Intraflagellar transport protein 74 homolog	0.19	0.00
	LRPPRC	P42704	157.905	Leucine-rich PPR motif-containing protein, mitochondrial	0.19	0.03
	NCL	P19338	76.614	Nucleolin	0.18	0.06
	RTCB	Q9Y3I0		tRNA-splicing ligase RtcB homolog	0.18	0.04
	SSR1	P43307	32.235	Translocon-associated protein subunit alpha	0.18	0.00
	CLTC	Q00610	191.615	Clathrin heavy chain 1	0.18	0.02
	PDHB	P11177	39.233	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	0.17	0.00
	ATAD3B	Q5T9A4	72.573	ATPase family AAA domain-containing protein 3B	0.17	0.00
	ETFA	P13804	35.08	Electron transfer flavoprotein subunit alpha, mitochondrial	0.17	0.00
	PYCRL	Q53H96	28.663	Pyrroline-5-carboxylate reductase 3	0.17	0.00
	PTBP1	P26599	57.221	Polypyrimidine tract-binding protein 1	0.17	0.00
	XPO1	O14980	123.386	Exportin-1	0.17	0.00
	HSD17B12	Q53GQ0	34.324	Very-long-chain 3-oxoacyl-CoA reductase	0.16	0.00
	FARSA	Q9Y285	57.564	Phenylalanine--tRNA ligase alpha subunit	0.16	0.04
	CBSL	P0DN79	60.587	Cystathionine beta-synthase-like protein	0.16	0.00
	ALDH18A1	P54886	87.302	Delta-1-pyrroline-5-carboxylate synthase	0.15	0.00
	IARS	P41252	144.498	Isoleucine--tRNA ligase, cytoplasmic	0.15	0.00
	MCM6	Q14566	92.889	DNA replication licensing factor MCM6	0.15	0.00
	XPOT	O43592	109.964	Exportin-T	0.15	0.00
	ATP5C1	P36542	32.996	ATP synthase subunit gamma, mitochondrial	0.14	0.00
	SMC1A	Q14683	143.23	Structural maintenance of chromosomes protein 1A	0.14	0.00

	DNAJA2	O60884	45.746	DnaJ homolog subfamily A member 2	0.14	0.00
	MCM3	P25205	90.981	DNA replication licensing factor MCM3	0.14	0.00
	AARS	P49588	106.81	Alanine--tRNA ligase, cytoplasmic	0.14	0.00
	BSG	P35613	42.2	Basigin	0.14	0.00
	DNM2	P50570	98.064	Dynamin-2	0.14	0.00
	CTPS1	P17812	66.69	CTP synthase 1	0.13	0.00
	CAND1	Q86VP6	136.376	Cullin-associated NEDD8-dissociated protein 1	0.13	0.00
	DDX21	Q9NR30	87.344	Nucleolar RNA helicase 2	0.13	0.00
	ATP2A2	P16615	114.757	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	0.13	0.00
	SLC1A5	Q15758	56.598	Neutral amino acid transporter B(0)	0.12	0.00
	TIMM44	O43615	51.35	Mitochondrial import inner membrane translocase subunit TIM44	0.12	0.00
	RPL4	P36578	47.697	60S ribosomal protein L4	0.12	0.00
	FARSB	Q9NSD9	66.116	Phenylalanine--tRNA ligase beta subunit	0.12	0.00
	FASN	P49327	273.427	Fatty acid synthase	0.12	0.01
	PSMC3	P17980	49.204	Proteasome 26S Subunit, ATPase 3	0.12	0.00
	IMPDH2	P12268	55.805	Inosine-5'-monophosphate dehydrogenase 2	0.12	0.00
	EFTUD2	Q15029	109.436	116 kDa U5 small nuclear ribonucleoprotein component	0.11	0.00
	MMS19	Q96T76	113.29	MMS19 nucleotide excision repair protein homolog	0.11	0.00
	IPO4	Q8TEX9	118.715	Importin-4	0.11	0.00
	MARS	P56192	101.116	Methionine--tRNA ligase, cytoplasmic	0.11	0.00
	KPNA2	P52292	57.862	Importin subunit alpha-1	0.11	0.00
	ATXN10	Q9UBB4	53.489	Ataxin-10	0.11	0.00
	GCN1L1	Q92616	292.75	Translational activator GCN1	0.10	0.00
	SNRNP200	O75643	244.508	U5 small nuclear ribonucleoprotein 200 kDa helicase	0.10	0.01
	NOP56	O00567	66.05	Nucleolar protein 56	0.10	0.00
	ABCD3	P28288	75.476	ATP-binding cassette sub-family D member 3	0.10	0.00
	ARAF	P10398	67.585	Serine/threonine-protein kinase A-Raf	0.09	0.00
	COPG1	Q9Y678	97.718	Coatomer subunit gamma-1	0.09	0.00
	TRIP13	Q15645	48.551	Pachytene checkpoint protein 2 homolog	0.09	0.00
	SLC27A4	Q6P1M0	72.064	Long-chain fatty acid transport protein 4	0.09	0.00
	RARS2	Q5T160	65.505	Probable arginine--tRNA ligase, mitochondrial	0.08	0.00
	LARS	Q9P2J5	134.466	Leucine--tRNA ligase, cytoplasmic	0.08	0.00
	MIB1	Q86YT6	110.136	E3 ubiquitin-protein ligase MIB1	0.08	0.00
	ADSL	P30566	54.889	Adenylosuccinate lyase	0.08	0.00
	PRMT5	O14744	72.684	Protein arginine N-methyltransferase 5	0.07	0.00
	MYBBP1A	Q9BQG0	148.855	Myb-binding protein 1A	0.07	0.00
	USP9X	Q93008	292.28	Probable ubiquitin carboxyl-terminal hydrolase FAF-X	0.07	0.00
	ADCK3	Q8NI60	71.95	Atypical kinase ADCK3, mitochondrial	0.07	0.00
	UPF1	Q92900	124.345	Regulator of nonsense transcripts 1	0.06	0.00
	DYNC1H1	Q14204	532.408	Cytoplasmic dynein 1 heavy chain 1	0.06	0.00
	WDR6	Q9NNW5	121.725	WD repeat-containing protein 6	0.06	0.00
	USP7	Q93009	128.302	Ubiquitin carboxyl-terminal hydrolase 7	0.06	0.00
	CNOT1	A5YKK6	266.939	CCR4-NOT transcription complex subunit 1	0.06	0.00
	FANCI	Q9NVI1	149.324	Fanconi anemia group I protein	0.06	0.00

	ACLY	P53396	120.839	ATP-citrate synthase	0.06	0.00
	MTHFD1L	Q6UB35	105.79	Monofunctional C1-tetrahydrofolate synthase, mitochondrial	0.05	0.00
	PRPF8	Q6P2Q9	273.6	Pre-mRNA-processing-splicing factor 8	0.05	0.00
	IPO5	O00410	123.63	Importin-5	0.05	0.00
	VARS	P26640	140.476	Valine--tRNA ligase	0.05	0.00
	SF3B3	Q15393	135.577	Splicing factor 3B subunit 3	0.05	0.00
	GANAB	Q14697	106.874	Neutral alpha-glucosidase AB	0.05	0.00
	PNPLA6	Q8IY17	149.995	Neuropathy target esterase	0.05	0.00
	UBE3C	Q15386	123.923	Ubiquitin-protein ligase E3C	0.04	0.00
	DDB1	Q16531	126.968	DNA damage-binding protein 1	0.04	0.00
	EIF3A	Q14152	166.569	Eukaryotic translation initiation factor 3 subunit A	0.04	0.00
	NUP160	Q12769	162.121	Nuclear pore complex protein Nup160	0.04	0.00
	TLN1	Q9Y490	269.767	Talin-1	0.04	0.00
	UBR5	O95071	309.352	E3 ubiquitin-protein ligase UBR5	0.04	0.00
	PI4KA	P42356	236.83	Phosphatidylinositol 4-kinase alpha	0.03	0.00
	NCAPD3	P42695	168.891	Condensin-2 complex subunit D3	0.03	0.00
	FLNA	P21333	280.739	Filamin-A	0.03	0.00
	MDN1	Q9NU22	632.82	Midasin	0.03	0.00
	NUP205	Q92621	227.922	Nuclear pore complex protein Nup205	0.03	0.00
	LRBA	P50851	319.108	Lipopolysaccharide-responsive and beige-like anchor protein	0.02	0.00
	ACACA	Q13085	265.554	Acetyl-CoA carboxylase 1	0.02	0.00
NUCLEAR EXTRACTS	VIM	P08670	53.65	Vimentin	1.11	0.19
	PHB2	Q99623	33.296	Prohibitin-2	1.03	0.08
	DDX3X	O00571	73.243	ATP-dependent RNA helicase DDX3X	0.81	0.11
	PGAM5	Q96HS1	32.004	Serine/threonine-protein phosphatase PGAM5, mitochondrial	0.81	0.00
	TUBA1C	Q9BQE3	49.895	Tubulin alpha-1C chain	0.73	0.00
	RPS3	P23396	26.688	40S ribosomal protein S3	0.64	0.00
	SMC1A	Q14683	143.23	Structural maintenance of chromosomes protein 1A	0.60	0.02
	DDX21	Q9NR30	87.344	Nucleolar RNA helicase 2	0.58	0.00
	ATAD3A	Q9NVI7	71.369	ATPase family AAA domain-containing protein 3A	0.57	0.04
	SMC3	Q9UQE7	141.54	Structural maintenance of chromosomes protein 3	0.49	0.02
	PRKDC	P78527	469.089	DNA-dependent protein kinase catalytic subunit	0.46	0.04
	NXF1	Q9UBU9	70.182	Nuclear RNA export factor 1	0.40	0.04
	RUVBL1	Q9Y265	50.22	RuvB-like 1	0.36	0.00
	RBM14	Q96PK6	69.492	RNA-binding protein 14	0.32	0.03
	AIFM1	O95831	66.901	Apoptosis-inducing factor 1, mitochondrial	0.30	0.00
	ELAVL1	Q15717	36.092	ELAV-like protein 1	0.29	0.00
	IRS4	O14654	133.768	Insulin receptor substrate 4	0.24	0.00
	WDR6	Q9NNW5	121.725	WD repeat-containing protein 6	0.22	0.00
	NUP205	Q92621	227.922	Nuclear pore complex protein Nup205	0.19	0.01
	NUP160	Q12769	162.121	Nuclear pore complex protein Nup160	0.16	0.00
	PDCD11	Q14690	208.701	Protein RRP5 homolog	0.09	0.00
	RANBP2	P49792	358.199	E3 SUMO-protein ligase RanBP2	0.09	0.00

2.4. Functional analysis of ANKRD55-protein partners

To interpret the potential functional role of ANKRD55-interacting proteins, ANKRD55 interactome results were subjected to enrichment analysis using DAVID bioinformatics resource tool. We used functional annotation clustering to select the interesting Gene Ontology (GO) and other terms. The results from GO biological process, GO cellular compartment, GO molecular function and Uniprot (UP) keywords with a FDR<5% are shown in Figures 17 and 18. A more detailed summary of the functional annotation clustering with $p<0.05$ is provided in Supplementary Table 3.

The functional analysis suggested that the ANKRD55-protein partners from total protein extracts were tightly related to nucleotide and ATP binding ($p<1.18x10^{-13}$). GO categories were enriched in nuclear transport terms including Ran GTPase binding, intracellular protein transport, protein import into nucleus and nuclear pore ($p<2.00x10^{-03}$). Total protein extracts dataset was also linked to processes that are associated with cell cycle and RNA, lipid and amino acid metabolism. Additionally, bioinformatics identified protein biosynthesis-related terms such as translation, structural constituent of ribosome, SRP-dependent cotranslational protein targeting to membrane, tRNA aminoacylation for protein translation and regulation of translational fidelity. This enrichment could be due to ANKRD55 overexpression. GO cellular component annotated focal adhesion, ribosome, nuclear pore, and nuclear chromosome, among others, as the localization of the involved proteins in the total protein interactome.

In the enrichment analysis of the ANKRD55-protein partners from nuclear extracts, ~20 and 30% of proteins are related to sumoylation and RNA binding, respectively. GO categories were involved in processes that are associated with cell cycle including cell division, mitotic nuclear envelope disassembly, mitosis, DNA repair and damage. Nuclear extracts interactome was also linked to processes including RNA transport, nucleotide and ATP binding, among others. GO cellular component analysis indicated nuclear pore as the localization of the involved proteins in the nuclear interactome.

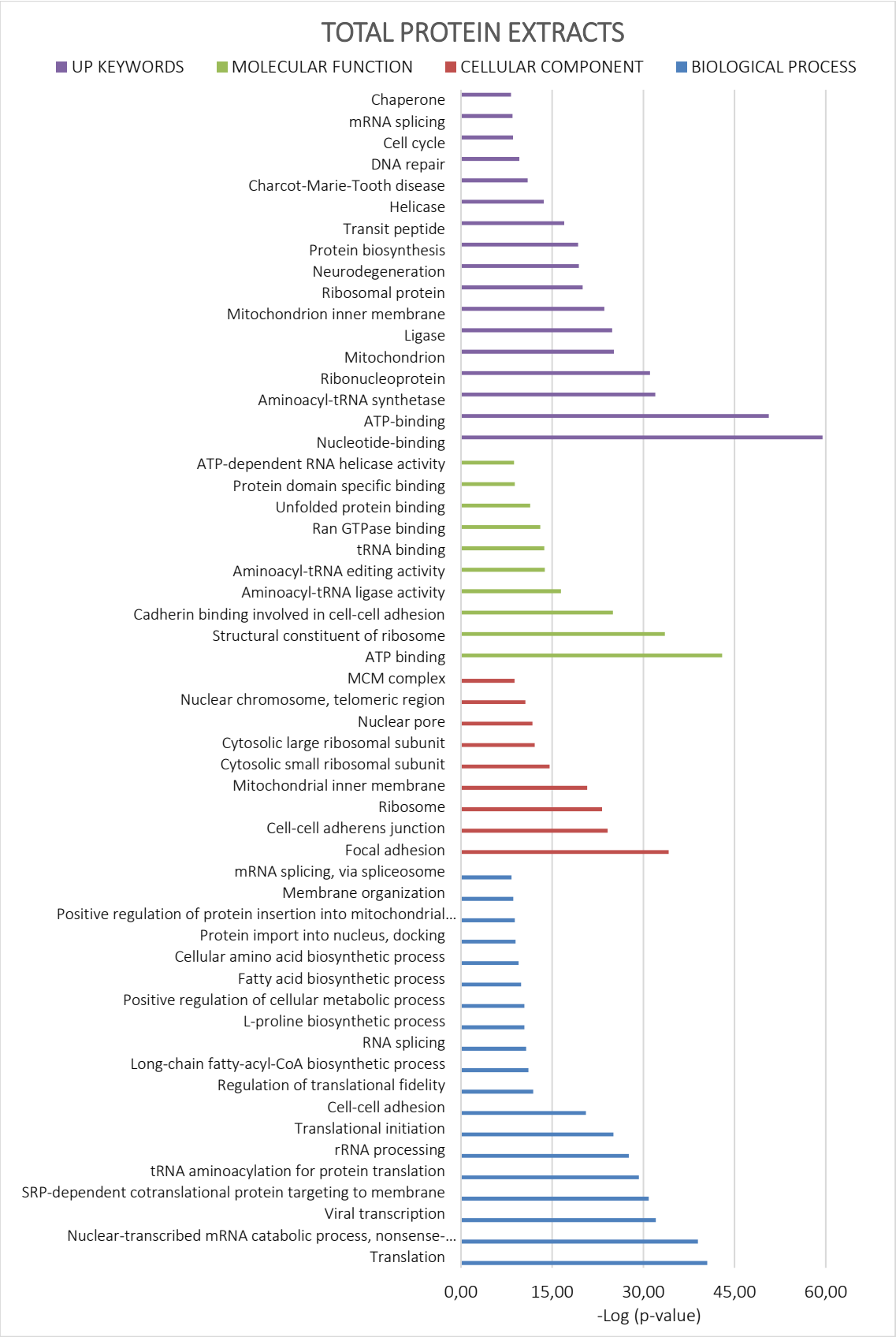


Figure 17. Functional enrichment analysis of ANKRD55 interactome from total protein extracts. The algorithm DAVID was used to analyze the ANKRD55 interactome using the GO terms Cellular Component (maroon), Biological Process (blue), Molecular Function (green) and Up Keywords (purple). The GO and other terms with FDR<5% are shown. Plotted as –log (p-value) significance.

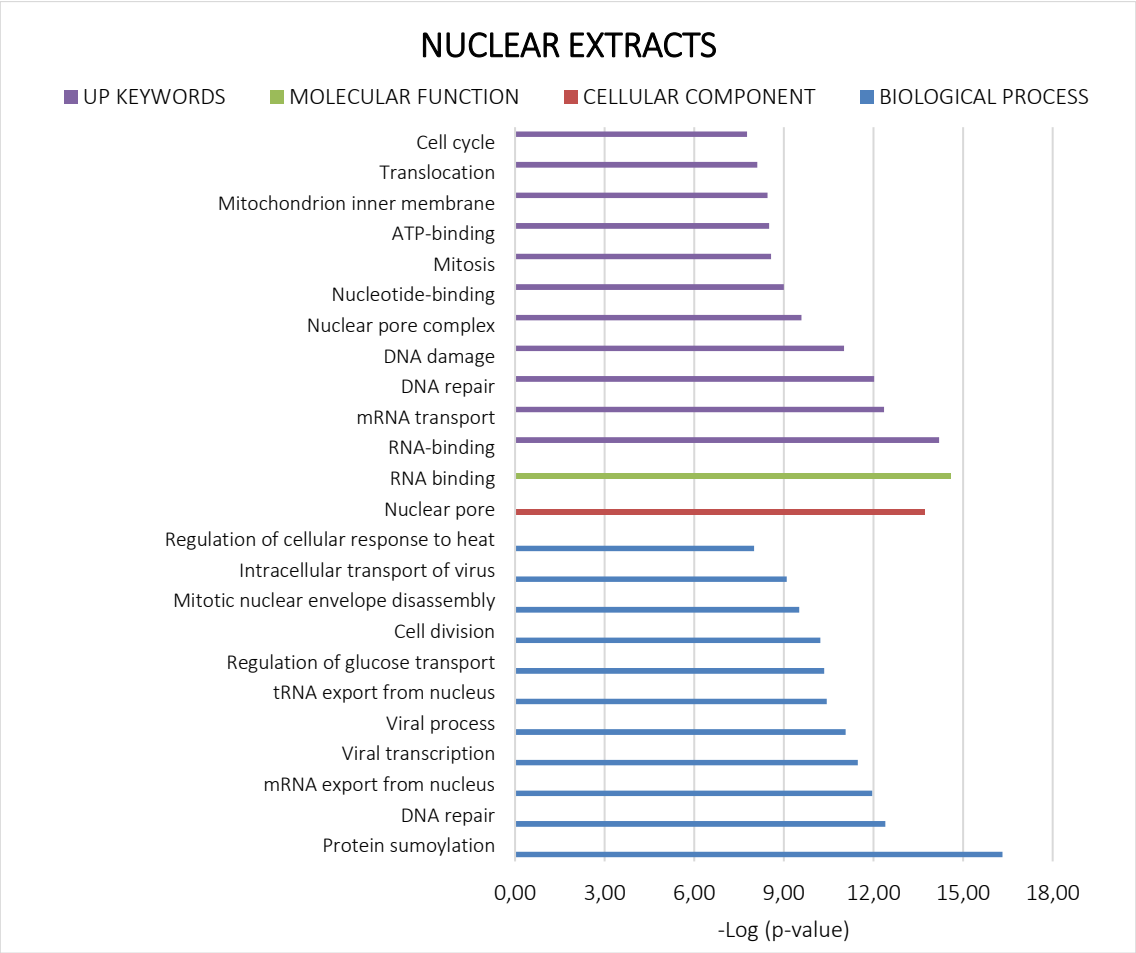


Figure 18. Functional enrichment analysis of ANKRD55 interactome from nuclear extracts. The algorithm DAVID was used to analyze the ANKRD55 interactome using the GO terms Cellular Component (maroon), Biological Process (blue), Molecular Function (green) and Up Keywords (purple). The GO and other terms with FDR<5% are shown. Plotted as –log (p-value) significance.

To further examine the ANKRD55 interactome, we performed an Ingenuity Pathway Analysis (IPA, QIAGEN) to identify major canonical pathways associated with ANKRD55-interacting proteins. Sixty out of one hundred ninety-six pathways were significantly enriched in total protein extracts ($p<0.05$; -log (p-value) >1.30). The top ten pathways, ranked by significance included tRNA Charging, EIF2 Signaling, Regulation of eIF4 and p70S6K Signaling, RAN Signaling, mTOR Signaling, Cell cycle: G2/M DNA Damage Checkpoint Regulation, HIPPO Signaling, Cell Cycle Control of Chromosomal Replication, PI3K/AKT Signaling and p70S6K Signaling (Figure 19). In nuclear extract interactome analysis, SMC1A and SMC3 proteins were related to Mitotic Roles of Polo-Like Kinase and ATM Signaling; PRKDC and PHB2 to Estrogen Receptor Signaling, VIM and TUBA1C to 14-3-3-mediated Signaling and PRKDC and TUBA1C to Sirtuin Signaling Pathway. All

information on the IPA canonical pathways analysis of both interactomes is included in Supplementary Table 4.

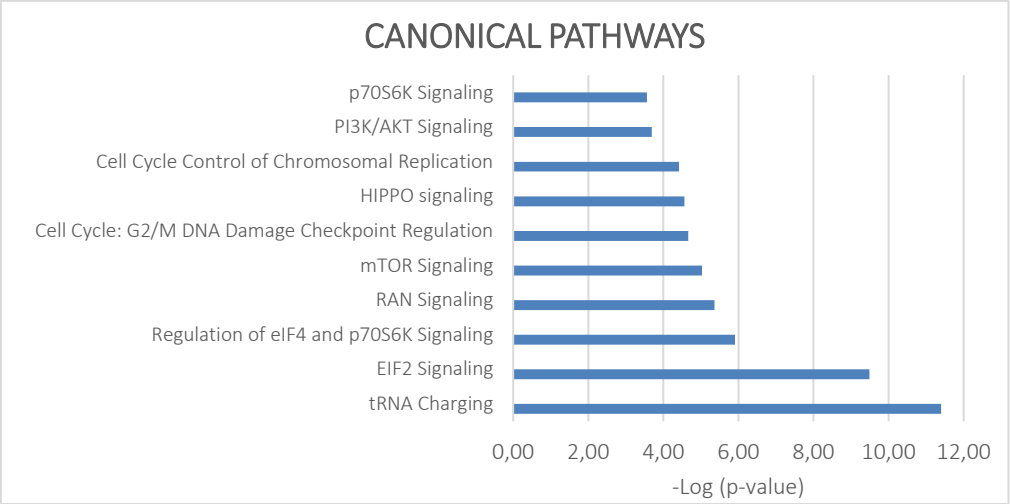


Figure 19. Top canonical signaling pathways enrichment in the ANKRD55 interactome from total protein extract. IPA software was used to identify major canonical pathways. Plotted as –log (p-value) significance.

2.5. Validation of selected ANKRD55 interacting partners

We confirmed specific interactions of several identified candidate proteins with recombinant ANKRD55 isoform 001 by IP followed by WB using antibodies (Supplementary Table 5 for details) against the selected partners. From the cytosolic interactome, we selected (i) six of the first ten proteins ranked according to NSAF value (14-3-3 isoforms, RPS3, and TUBB6); (ii) CLTC with a lower NSAF value and mainly localized to the vesicles, in addition to cytosol and mitotic spindle, and (iii) proteins that were found to be present as well in the NE interactome (SMC1A and PRKDC). From the nuclear interactome, three proteins (VIM, SMC1A and SMC3) from the top ten in the list were analyzed. Several of these proteins had at first been identified in individual bands cut from silver-stained gels including PRKDC, SMC1A and SMC3 (see Table 6). Regarding TUBB6, we used an anti-β-tubulin antibody which can recognize different β-tubulin isoforms including TUBB3, TUBB2B, TUBB, TUBB4B, TUBB2A, and TUBB4A (three of these were identified in two out of three interactome replicates, and several of these had been identified as well as in individual bands analyzed in the first approach, see Table 6). These β-tubulin isoforms showed more than 92% of homology with the immunogen used for the antibody production. We transiently expressed FLAG-ANKRD55 in HEK293 cells and performed IP from total protein extracts and nuclear extracts. Anti-FLAG IP experiments

and WB demonstrated that endogenous SMC1A, TUBB6, RPS3, 14-3-3 isoforms, PRKDC, and CLTC in total protein extracts (Figure 20 A, B) and SMC1A, SMC3, and VIM in nuclear extracts interacted specifically with FLAG-ANKRD55 (Figure 20C). In contrast, neither ANKRD55 nor interacting partners were detected in the control IP. The IP of FLAG-ANKRD55 with 14-3-3 isoforms and PRKDC was also validated in total protein extracts of HeLa cells by WB (Figure 20D).

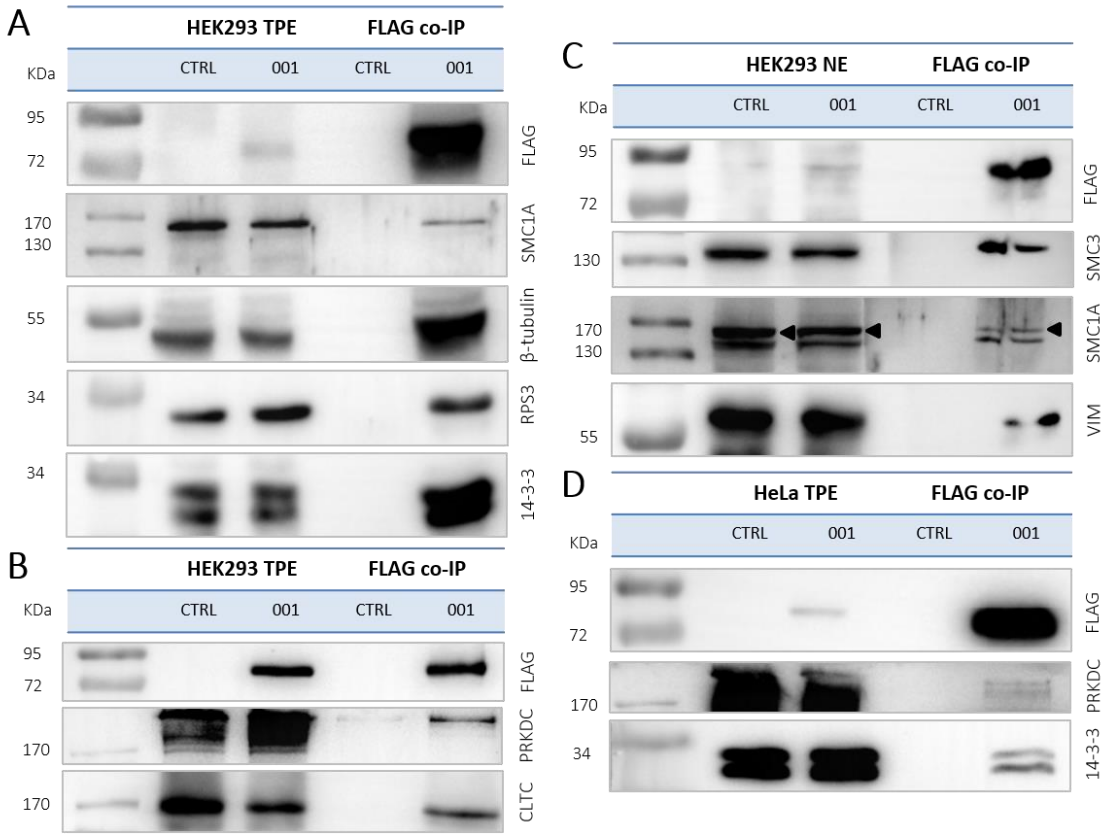


Figure 20. Validation of ANKRD55 isoform 001 interacting proteins. Overexpressed FLAG-ANKRD55 isoform 001 (001) and endogenous candidate proteins were IPed from enriched nuclear fractions (NE) and total protein extracts (TPE) after 48h of transfection in HEK293 and HeLa cells. As negative control cells were cultivated under the same conditions, but without transfection agent (CTRL). Recombinant ANKRD55 and endogenous proteins were detected by WB in NE, TPE and eluted fractions (FLAG IP) using FLAG Ab and specific interacting partners Ab.

2.6. Immunofluorescence colocalization study on ANKRD55 interactome

Proteins that interact tend to reside within the same or adjacent subcellular compartments[321]. Fluorescence colocalization microscopy is frequently used to assess potential links between distinct molecules[322]. For this reason, we study the subcellular localization of recombinant ANKRD55 and a selection of interacting partners at different time points to verify if they share at least one common subcellular localization.

The results showed a highly significant colocalization ($p < 0.0001$) between recombinant ANKRD55 isoform 001 and RPS3 in the cytoplasm (Figure 21), and β-tubulin (Figure 22) and 14-3-3 isoforms (Figure 23) in the cytoplasm and cell membrane upon comparison of transfected and non-transfected cells. Moreover, the colocalization with TUBB increased with time, decreased with RPS3 and there were no differences between two-time points for 14-3-3 isoforms (Figure 21B, 22B, 23B).

It is necessary to mention that overexpressed ANKRD55 itself showed different staining patterns in the colocalization experiments with RPS3 and β-tubulin, where it was localized mainly in the cytoplasm, versus 14-3-3 where it occurred more pronouncedly in the vicinity of the cell membrane. These staining differences are due to the type of antibodies used to mark ANKRD55. In the former analysis the anti-FLAG antibody used was monoclonal and, therefore, arguably more specific than the one used in the colocalization with 14-3-3 which was polyclonal and not antigen-purified (Figure 23).

As we can see in Figures 21A, 22A and 23A, it was not possible to visualize recombinant ANKRD55 in the nuclei. Given that in previous analysis the presence of ANKRD55, in its endogenous and overexpressed form, was validated in the nucleus of several cell lines including HEK293 (Figures 10, 11 and 12) it was thought that the signal in immunofluorescence may have been too weak to facilitate visualization with the previous protocol. Thus, the procedure was modified to include an antigen retrieval treatment to try to increase the fluorescent signal of the nuclei.

The modification of the immunofluorescence protocol was successful, so it was subsequently applied to study the nuclear colocalization of overexpressed ANKRD55 together with β-tubulin (Figure 24) and VIM (Figure 25) as representatives of the validated proteins of ANKRD55 interactomes (Table 7).

After antigen retrieval treatment, the results showed recombinant ANKRD55 in the nuclei of cells in division. Therefore, the nuclear colocalization with β-tubulin (Figure 24) and VIM (Figure 25) could only be studied in this stage, and quantification was not performed due to the small sample size. It should be mentioned that cells in mitosis transfected with ANKRD55 showed a lower intensity of fluorescent signal, in comparison with the rest of

the fixed cells. In the images, we can see a positive colocalization with diffuse β-tubulin in the nucleus and specific in the spindle (Figure 24), and with diffuse VIM in the nucleus without an associated specific structure (Figure 25).

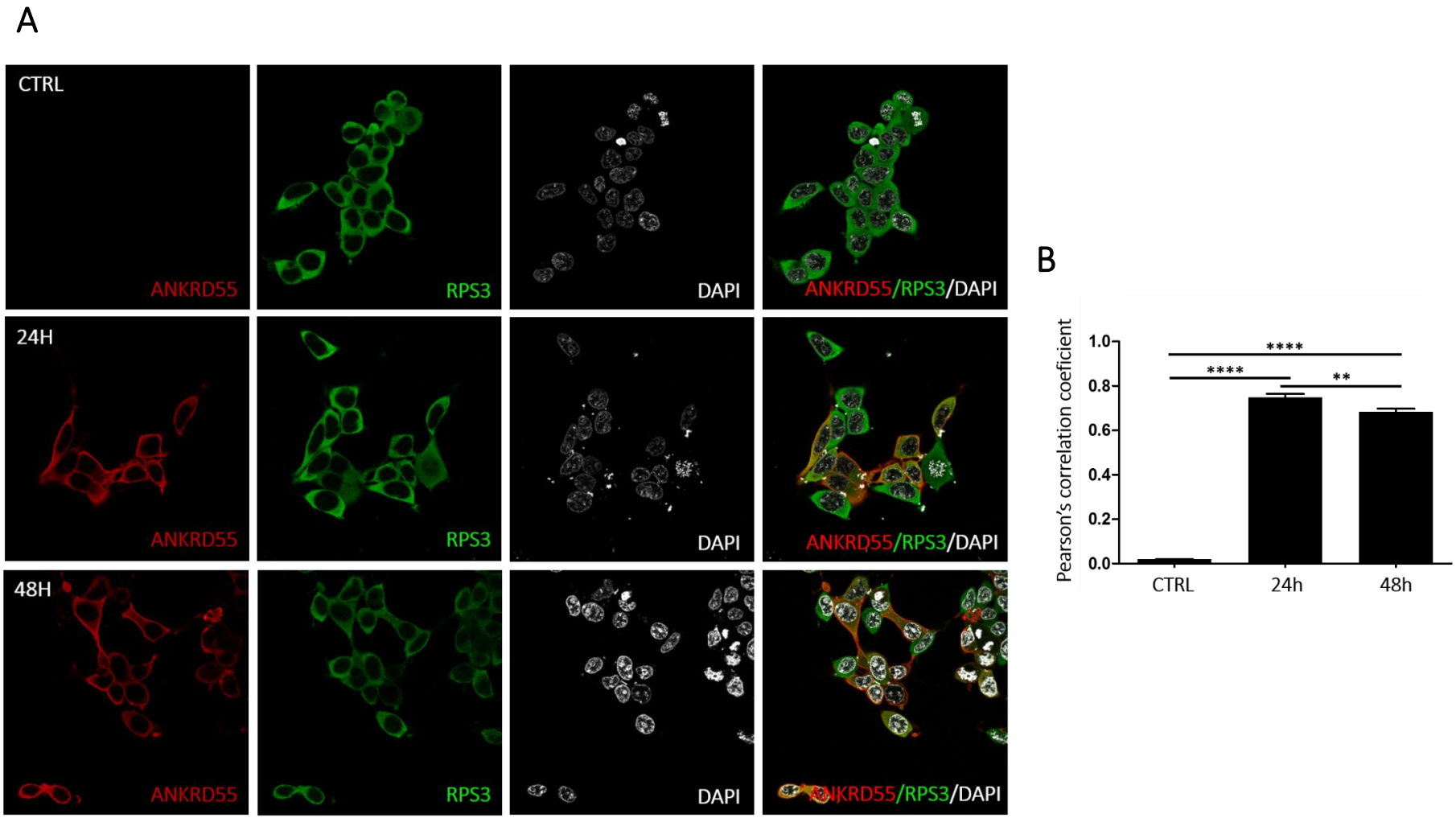


Figure 21. Recombinant ANKRD55 colocalized with endogenous RPS3 in HEK293 cells. A) Representative microphotographs of immunostaining for ANKRD55 (FLAG Ab; red), 40S ribosomal protein S3 (RPS3 Ab; green) and nuclei (DAPI; white) in non-transfected (CTRL) and transfected 24h (24h) and 48h (48h) HEK293 cells with ANKRD55 isoform 001. B) Colocalization quantification using Pearson's correlation coefficient. Data are mean ± SEM (n=30 cellular ROIs/condition), $p<0.0001$ (****) and $p<0.01$ (**) comparing three conditions, Mann-Whitney test.

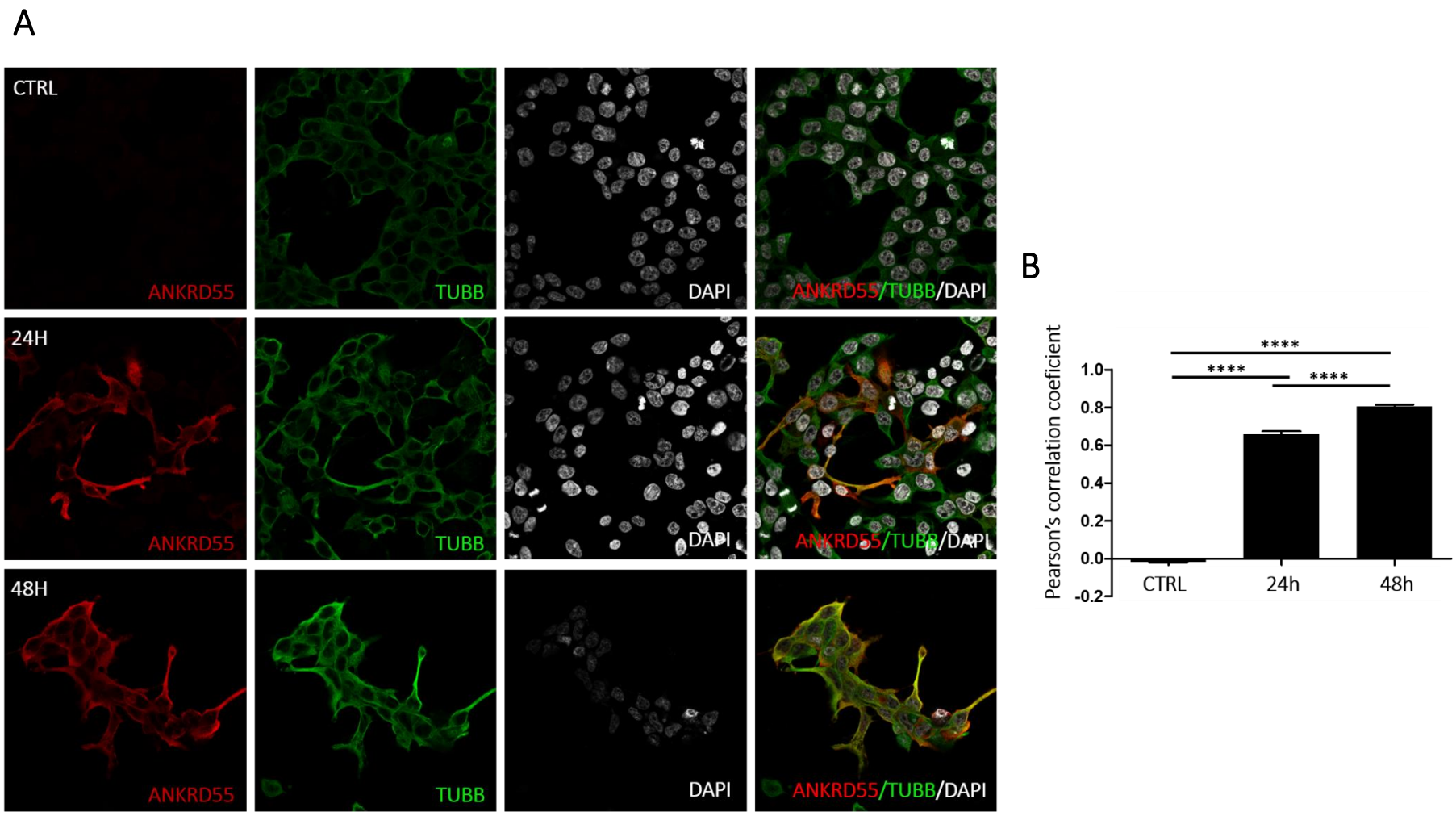


Figure 22. Recombinant ANKRD55 colocalized with endogenous β-tubulin in HEK293 cells. A) Representative microphotographs of immunostaining for ANKRD55 (FLAG Ab; red), tubulin beta chain (TUBB Ab; green) and nuclei (DAPI; white) in non-transfected (CTRL) and transfected 24h (24h) and 48h (48h) HEK293 cells with ANKRD55 isoform 001. B) Colocalization quantification using Pearson's correlation coefficient. Data are mean ± SEM (n=30 cellular ROIs/condition), $p<0.0001$ (****) comparing three conditions, Mann-Whitney test.

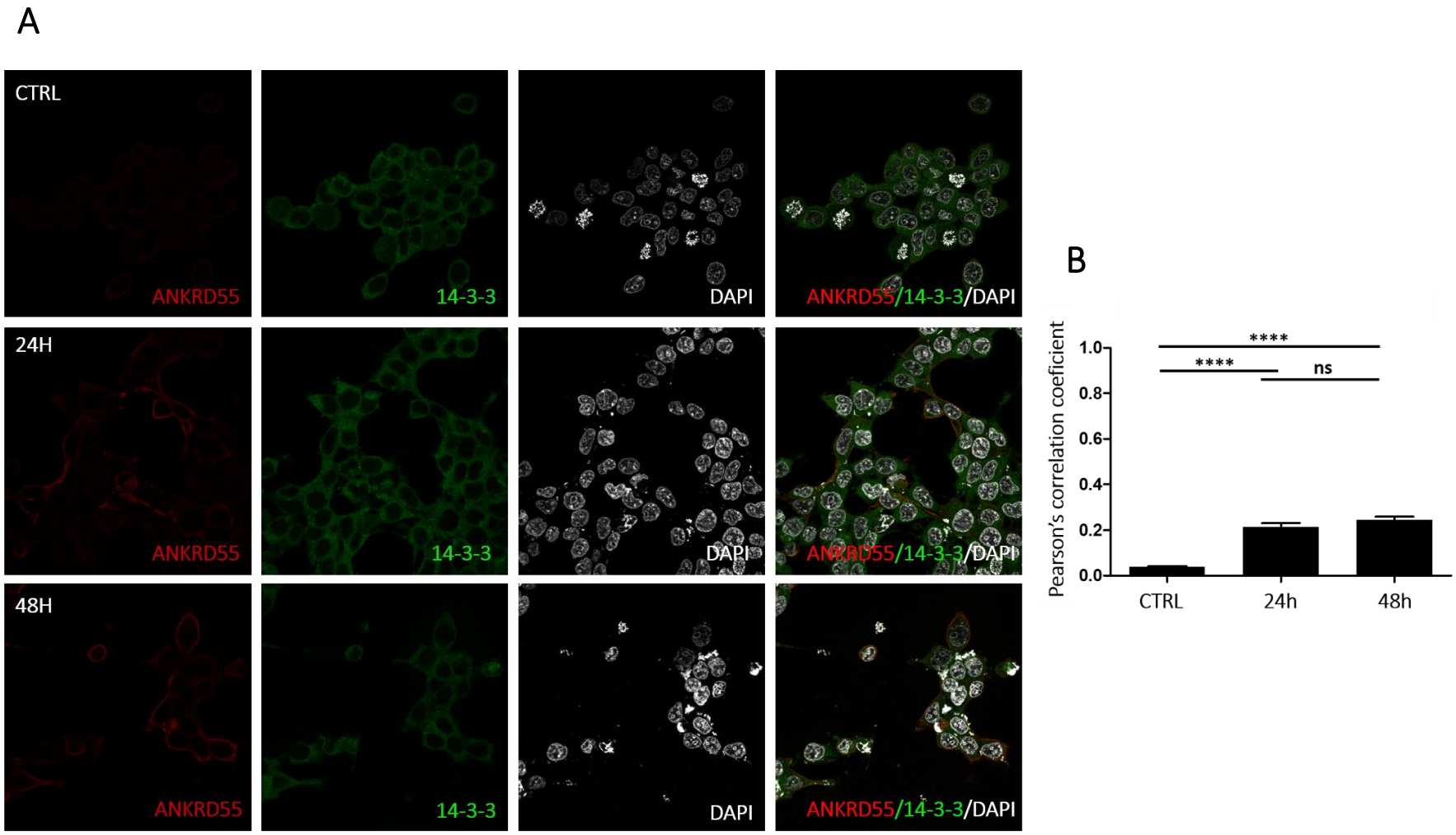


Figure 23. Recombinant ANKRD55 colocalized with endogenous 14-3-3 protein isoforms in HEK293 cells. A) Representative microphotographs of immunostaining for ANKRD55 (FLAG Ab; red), 14-3-3- protein isoforms (14-3-3 (pan) Ab; green) and nuclei (DAPI; white) in non-transfected (CTRL) and transfected 24h (24h) and 48h (48h) HEK293 cells with ANKRD55 isoform 001. B) Colocalization quantification using Pearson's correlation coefficient. Data are mean ± SEM (n=30 cellular ROIs/condition), $p<0.0001$ (****) and no significant (ns) comparing three conditions, Mann-Whitney test.

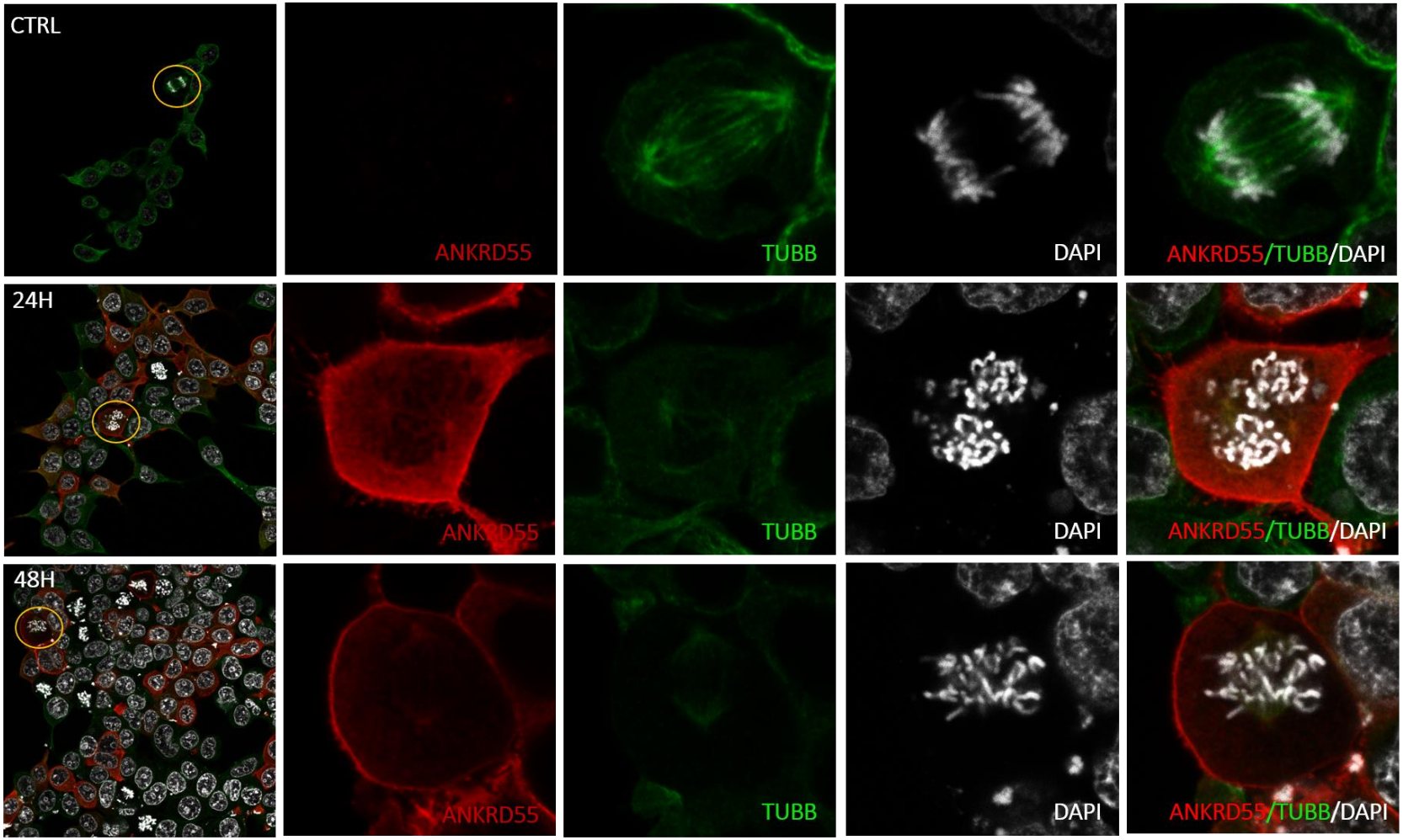


Figure 24. Recombinant ANKRD55 colocalized with endogenous β-tubulin in HEK293 cells. Representative microphotographs of immunostaining for ANKRD55 (FLAG Ab; red), tubulin beta chain (TUBB Ab; green) and nuclei (DAPI; white) in non-transfected (CTRL) and transfected 24h (24h) and 48h (48h) HEK293 cells with ANKRD55 isoform 001 subjected to antigen retrieval treatment. Original image included zoon-in region (yellow), the signal intensity in zoom-in images was increased.

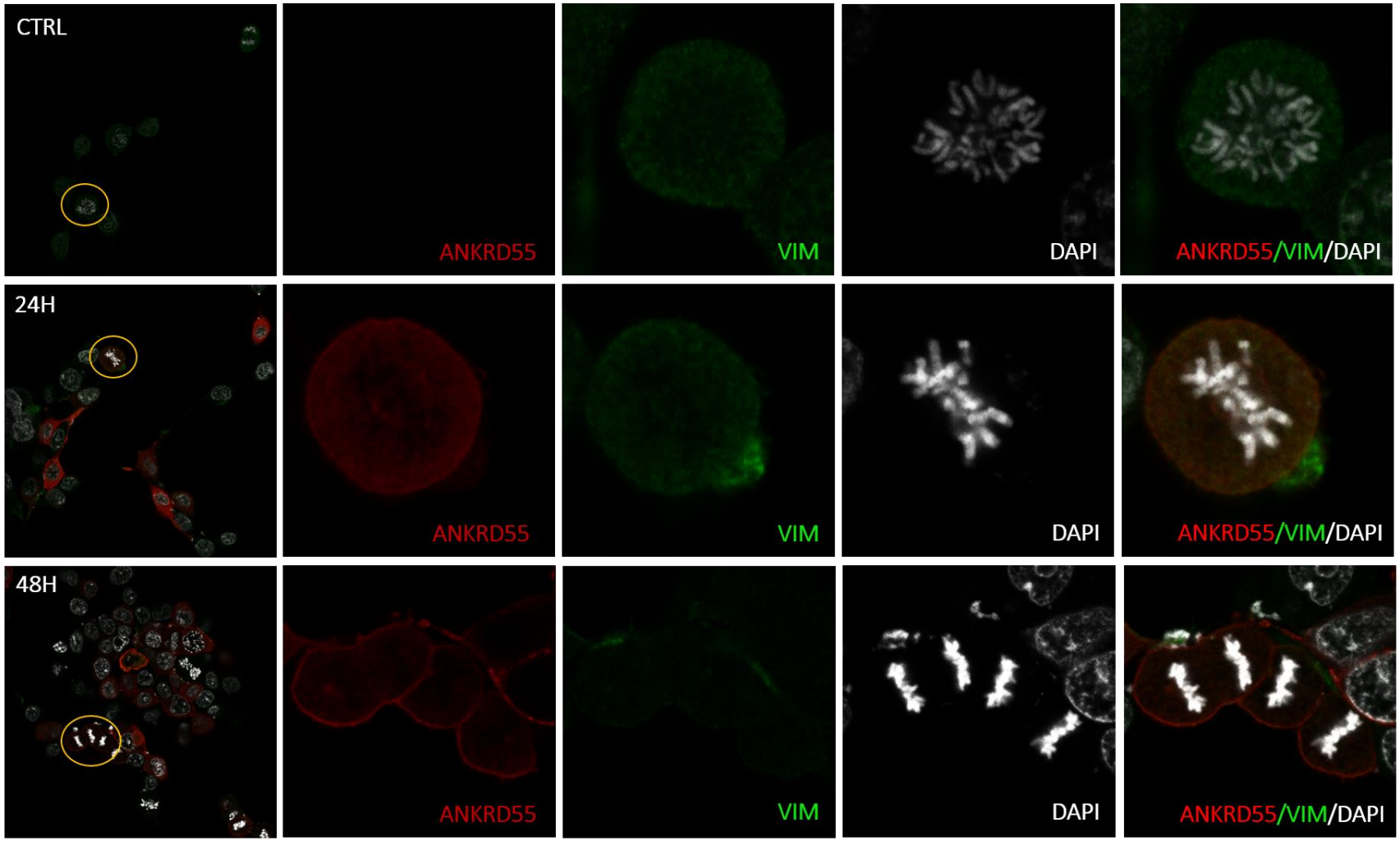


Figure 25. **Recombinant ANKRD55 colocalized with endogenous VIM in HEK293 cells** Representative microphotographs of immunostaining for ANKRD55 (FLAG Ab; red), vimentin (VIM Ab; green) and nuclei (DAPI; white) in non-transfected (CTRL) and transfected 24h (24h) and 48h (48h) HEK293 cells with ANKRD55 isoform 001 subjected to antigen retrieval treatment. Original image included zoon-in region (yellow), the signal intensity in zoom-in images was increased.

2.7. ANKRD55 is a phosphoprotein

An important subset of proteins identified in this interactome analysis belong to the 14-3-3 protein family. 14-3-3 proteins are ubiquitously expressed phosphoserine/phosphothreonine (pS/pT)-binding proteins that are members of a large family of isoforms[323]. In general, they bind to their ligands through RSXpS/TXP or RXXXpS/TXP sequences[324], but they can also bind their ligands in a phosphorylation-independent manner[325,326]. Thus, we determined whether ANKRD55 was phosphorylated.

Our primary sources for this step were phosphoproteome repositories, such as PhosphoSitePlus[316] and dbPAF[317] databases. The identified phosphorylation sites are summarized in Table 8. Moreover, 14-3-3-Pred web server[318] was used to predict putative 14-3-3 binding phosphosites on ANKRD55. Two out of ninety predicted phosphorylation sites (S436 and S597) fulfill the cut-off (Supplementary Table 6). Finally, we identified specific pS/pT sites in overexpressed and IPed ANKRD55 isoform 001 from HEK293 from nuclear and total proteins extracts by TiO_2-based phosphopeptide enrichment prior to nLC–MS/MS analysis. The phosphorylation sites at T6 and S436 in ANKRD55 were confirmed and the S475 site was found in nuclear and total protein extracts previously identified in mouse ANKRD55 (Table 8). In conclusion, the data revealed five phosphorylation sites in human ANKRD55 at T6, T11, T189, S436 and S475. 14-3-3-Pred web server predicted with the highest score S436 as 14-3-3 binding phosphosites. ANKRD55 identified phosphorylation sites are summarized in Table 8.

Table 8. Summary of the identified Ser/Thr Phosphorylation Sites in ANKRD55. The first five columns show the phosphorylation sites identified in the phosphoproteome repositories including position and sequence of the peptide, specie in which it was found, reference for the experiment data and database. The last column includes our experimental data about the ANKRD55 phosphosites: N, nuclear extracts and T, total protein extracts.

Position	Peptides	Species	References	Sources	Identified phoshorylation sites
T6	MMRQAtMDFSTPS	Human	[327]	PhosphoSitePlus	N/T
T11	QATMDFStPSVFDQQ	Human	[327]	PhosphoSitePlus	
T189	GADPtLVDK	Human	[328]	dbPAF	
S436	TQsLPPITLGNNFLTASHR	Human	[329]	dbPAF	T
S473	MAQRSQKsRSEQDLL	Mouse	[330,331]	PhosphoSitePlus	
S475	QRSQKSRsEQDLLNN	Mouse	[330–335]	PhosphoSitePlus	N/T

DISCUSSION

ANKRD55 mainly has been identified as a risk gene for several autoimmune diseases in multiple GWAS[191–193,196,198–200,202]. However, the functions of ANKRD55 have not been elucidated. Knowledge of protein localization is invaluable for understanding their function as well as the interaction of different proteins[336]. It has been revealed that over 80% of proteins do not operate alone but in complexes[337]. The functionality of unidentified proteins can be predicted on the evidence of their interaction with a protein, whose function is already revealed[338]. Moreover, ANKRD55 belongs to a family containing AR, which exclusively functions to mediate PPIs, consequently should take part in diverse PPI as has been reported for other members of the ANK family.

First, we identified the endogenous and recombinant ANKRD55 subcellular localization using an Ab that recognizes 001 and 005 isoforms. Endogenous ANKRD55 immunoreactivity was detected in the nucleus of SH-SY5Y, Jurkat, and U937 cell lines by immunofluorescence microscopy. ANKRD55 reactivity was also present in the cytosol of all of the studied cell types, although, unlike the nuclear signal, the cytosolic one did not disappear upon addition of an ANKRD55-specific blocking peptide (Figure 10). Immunoblotting of sucrose gradient–purified nuclei confirmed the presence of 005 and 001 in the nucleus of primary $CD4^+$ T cells and in those of the Jurkat and U937 cell lines, as well as of 001 in the nucleus of SH-SY5Y cells (Figure 11A). Biochemical fractionation also validated that isoform 001 of ANKRD55 was located mainly in the nucleus as well as in membranous organelles of SH-SY5Y and Jurkat cell lines. However, isoform 005 distributed between cytoplasm and membranous organelles in Jurkat cells, whereas in SH-SY5Y cells it was weakly detectable in cytosol and membrane organelle fractions but was virtually absent from the nuclei as shown through sucrose ultracentrifugation or biochemical fractionation. Instead, a smaller immunoreactive protein (30 kDa), which is detectable in the nucleus but not in the cytosol and is not seen in any of the other cell types analyzed, may correspond to a lineage-specific immunoreactive truncation product of a larger nuclear isoform because an ANKRD55 splice variant coding for a protein of this size has not been reported (Figure 11B). Both recombinant ANKRD55 isoforms 001 and

005 were detected in cytosol, membranous organelles and nucleus of HeLa and HEK293 cells by biochemical fractionation. Additionally, truncated immunoreactive products were present, probably generated by specific proteolytic events from the intact isoform proteins (Figure 12).

Then, we characterized ANKRD55 interactome from total protein extracts and nuclear extracts which revealed unusual combination of functionally very distinct proteins. A larger group of ANKRD55-interacting partners was present in total protein extract than in nuclear interactome. Functional analysis suggested that the ANKRD55-protein partners from nuclear (eight and seven proteins) and total protein extracts (fifty-two and forty-three proteins) are related to nucleotide and ATP binding, respectively (Figures 17 and 18). 30% of proteins from nuclear interactome are related to RNA binding. Bioinformatics analysis predicted the binding ability of ANKRD55 to DNA, ATP, ADP and GTP nucleotides, but not to RNA (Supplementary Table 8).

GO categories from total protein extracts were enriched in nuclear transport terms providing a possible explanation for the nucleocytoplasmic transport of ANKRD55, which seems to be mediated by importins/exportins transport system enriched in the interactome (CSE1L, XPO1, IPO4, XPOT, IPO5, NUP160, NUP205, KPNA2, COPG1, CDK5, YWHAH, CLTC) (Figure 17). However, a number of AR, such as IκBα, ASPP2, and GABPβ, have been reported to enter the nucleus via an unknown mechanism independent of a canonical nuclear localization signal[339]. As demonstrated by Lu *et al.* (2014), this mechanism is characterized by the presence of components of the RanGDP/AR pathway which represents a general importin-independent nuclear import pathway frequently used by AR-containing proteins[340]. Ran GTPase binding, which obtained high significance in the GO, is known to regulate protein trafficking through the nuclear envelope and is present in both nuclear transport systems.

Both interactomes were also linked to processes that are associated with the cell cycle (Figures 17 and 18). ANKRD55-interacting partners validation by confocal microscopy reported that ANKRD55 is localized in the nucleus only during mitosis (Figures 24 and 25). In the different phases of mitosis, ANKRD55 showed a more intense signal located in the region of the mitotic spindle and less diffuse intensity in the remainder of the

nucleoplasm. The mitotic spindle has a crucial role in ensuring the accurate segregation of chromosomes into the two daughter cells during cell division. It is a self-organized and dynamic macromolecular structure that is constructed from microtubules, microtubule-associated proteins and motor proteins[341]. Its assembly is dependent on the tightly regulated nucleation of microtubules, which form the major structural component of the mitotic spindle. Microtubules are assembled from dimers of α- and β-tubulin, a process that is initiated from γ-tubulin ring complexes[341,342]. These data agree with the results obtained in the ANKRD55 interactomes due to the identification of diverse α- and β-tubulin isoforms and other proteins involved in mitotic spindle (Table 7 and Supplementary Table 2). Moreover, in the images of nuclear immunofluorescence, β-tubulin presented a specific marking in the mitotic spindle that colocalized with ANKRD55. This indicates that the mitotic spindle, although it has been a well-defined structure for years, ANKRD55 could be an associated protein. Thus, for a better study of ANKRD55 presence in the cell nucleus during the different phases of mitosis and a better understanding about its nuclear function, in the future the cells should be subjected to appropriate treatments to arrest them in the different phases of mitosis. SMC1A and SMC3 were also validated as part of the interactome of ANKRD55 (Figure 20 A and C). In fact, both proteins are members of cohesin complex crucial for chromatid cohesion. By holding the two sister chromatids together from their formation during replication until their separation in anaphase, cohesin creates a counterforce to the pulling of the mitotic spindle, which allows correct chromosome alignment and segregation[343]. Therefore, although these colocalizations have not been performed by immunofluorescence, one would expect the colocalization in the mitotic pole of SMC1A, SMC3 and ANKRD55. This association between the ANKRD55-interacting proteins and the mitotic spindle is also supported by the presence of RPS3 protein involved in spindle dynamics[344]. Similarly, the Ran GTPase binding, enriched in the GO (Figure 17), has been related to the spindle formation[345].

Intriguingly, numerous members of diverse protein families, such as 14-3-3, mitochondrial carriers, minichromosome maintenance proteins (MCM), serine/threonine phosphatases and S (RPS) and L (RPL) risobosomal proteins were present.

Within the 14-3-3 protein family, the isoforms present in the analysis of the total protein extract were epsilon, beta/alpha, eta and gamma. 14-3-3 proteins, being pS/pT-binding proteins, led us to identify and confirm three phosphorylation sites in ANKRD55. 14-3-3 proteins modulate the action of proteins that are involved in cell cycle and transcriptional control, signal transduction, intracellular trafficking and regulation of ion channels[346] through a variety of mechanisms: (i) direct conformational change of the target protein; (ii) physical occlusion of sequence specific or structural features; and (iii) scaffolding that anchors proteins within close proximity of one another[347]. 14-3-3 proteins have been detected in the CSF of patients with MS[346] and could be a useful marker to identify a subgroup of demyelinating diseases patients[348]. Additionally, reactive astrocytes in chronic demyelinating lesions of MS patients intensely expressed 14-3-3 beta, epsilon, zeta, eta and sigma isoforms, among which the epsilon isoform is a highly specific marker for reactive astrocytes[348] as well as 14-3-3 gamma plays a pivotal role in signaling pathways for oligodendrocytes protection in neuroinflammation[349].

Although ribosomal proteins (RP) are known for playing an essential role in ribosome assembly and protein translation, their ribosome-independent functions are also increasingly being appreciated[350]. Eleven RP were identified in ANKRD55 interactomes. Mutations in RPL10 are associated with T-cell acute lymphoblastic leukemia[351] and autism, probably by affecting brain development[352]. RPL23 has been found to regulate the MDM2/MDMX–p53 cascade, consequently suppressing tumor cell proliferation[353,354]. RPS9 induces CDK1[355], RPS3A activates NF-κB[356], RPLP0 is involved in gynecologic tumors[357], RPL4[358] inhibits and RPL9[359] promotes virus production. RPS3, a validated ANKRD55-interacting partner in both cytosolic and nuclear interactomes, was shown to induce apoptosis by collaborating with E2F1[360]. Phospho-RPS3 was shown to translocate into the nucleus and upregulate prosurvival gene expression via association with NF-κB in non-small cell lung cancer cells[361,362]. RPS3 is also involved in immune signaling by selectively modulating NF-κB target gene expression. NF-kB is a family of transcription factors that were originally identified to regulate genes crucial for immune response, but later on were found to also regulate genes implicated in cell survival or proliferation[350]. Given the fact that RPS3 also activates the p53 tumor suppressive

pathway[363], this RP is regarded as one of the most fascinating RP with pivotal multifunctions[364].

The HUGO Gene Nomenclature Committee (HGNC) provides a list of transporter families of the solute carrier (SLC) gene series (http://www.genenames.org)[365] which include over 300 members organized into 60 families. Solute carriers are eukaryotic membrane proteins that control the uptake and efflux of various solutes, including amino acids, sugars, and drugs[366]. Six members of mitochondrial carrier family (MCF) or SLC25 protein family were identified in the ANKRD55 interactome. Additionally, three further members of the same family were detected in two out of three replicates and in the first approach that was based on excision of silver stained gel bands. All SLC25 members have common sequence features: a tripartite structure, a 3-fold repeated signature motif, and six transmembrane α-helices (two in each of the three repeats). These structural features are different from those of any other transporter family and facilitate to unequivocally recognize them. The characterized SLC25 family members transport an extensive range of solutes[367]. Those identified in the interactome transport: SLC25A1, citrate, isocitrate, malate and phosphoenolpyruvate; SLC25A3, phosphate; SLC25A5, ADP and ATP; SLC25A11, 2-oxoglutarate and malate; SLC25A12 and SLC25A13; aspartate and glutamate. Mutations in *SLC25* genes mainly cause rare metabolic disorders. Their symptoms follow from the specific metabolism affected[368]. However, a disease caused by a mutation in the *SLC25A12* gene (present in the ANKRD55 interactome) is related with global hypomyelination and impaired neuronal aspartate transport, which prevents myelin formation by failing to provide N-acetylaspartate (NAA) to oligodendrocytes[369]. *Slc25a12* knockout mice showed impaired myelination, a marked decrease of the myelin lipid precursor NAA and lower aspartate concentration in the brain[369,370]. These results provide genetic evidence in support of an important role for SLC25A12 in myelin formation[371]. Other members from different SLC families were also identified. SLC1A5, a glutamine and essential amino acids transporter, is involved in several human cancers, amyotrophic lateral sclerosis[372], and schizophrenia[373]; SLC27A4, a fatty acid transporter, is associated with autism and potentially has important roles in the developing brain[374].

MCM2 to 7 is a heterohexamer complex that contains six MCM homologues, and three of these were identified in the ANKRD55 interactome (MCM3, 6 and 7). MCMs have been determined to be DNA helicases involved in the initiation of DNA replication. Some studies have determined that there are more MCMs bound onto chromatin than required by replication origins, and the phenomenon is coined as the 'MCM paradox'[375,376]. MCM7, a key component of the MCM complex, forms a double trimeric complex with MCM4 and MCM6. The double trimeric complex initiates DNA replication through unwinding DNA double strands. During the S phase, the complex participates in elongating DNA strands. Recently, MCM7 was reported to co-localize with tubulin in mitotic cells with MCM7 depletion resulting in aberrant mitosis. This is in line with observations that MCM7 exerts certain functions on spindle formation to prevent cytokinesis during early mitosis by regulating CDK1 activity[377]. Therefore, although these colocalizations have not been performed by immunofluorescence, one would expect the colocalization in the mitotic pole of MCM7 and ANKRD55.

Regarding the two approaches we used for identification of ANKRD55-interacting partners, almost 60% of proteins identified with the first approach that was based on excision of silver stained gel bands, were confirmed in at least two out of the three replicates from full-gel interactome analysis. Seven out of fifty-nine were identified in a different cell extract and sixteen were not validated. Moreover, we checked for freely available protein-protein databases (BioGRID[378], STRING[379] and HuRI) to confirm some of the identified interactions. The BioPlex 2.0 dataset[300,380], which has been deposited into the BioGRID database, described thirteen ANKRD55-interacting partners; seven were included in at least two out of three replicates from our interactome analysis (YWHAE, YWHAQ, IFT52, IFT74, MIB1, YWHAZ, and YWHAG). Intriguingly, HIF1AN, an ANKRD55-interacting partner described only in the BioPlex 2.0 dataset, catalyzes hydroxylation of highly conserved asparaginyl residues within AR of NFKB1, NFKBIA, NOTCH1, ASB4, PPP1R12A and several other AR-containing proteins[381]. Therefore, this could suggest the hydroxylation of asparaginyl residues from ANKRD55. However, in our experimental set-up, we did not clearly detect HIF1AN in ANKRD55 interactomes by mass spectrometry, and neither were we able to demonstrate clearly its interaction with ANKRD55 by WB

analysis using specific antibodies (Supplementary Figure 3), or that of another BioGrid partner, ZSCAN1 (data not shown).

STRING database also shared ANKRD55-partners, specifically RPS3 and ACLY in *Homo sapiens* and Rps3 and Hsp90ab1 in *Mus Musculus*. Human Reference Protein Interactome Project (HuRI) included unpublished thirty ANKRD55-interacting partners; none of these were classified in our protein list (Supplementary Table 7). However, some of the HuRI ANKRD55 partners are involved in ciliogenesis and its regulation (CFAP206, LRGUK, and DEUP1), just as some proteins from our ANKRD55 interactome (MIB1, IFT74, ARF4, CCT4, CCT8, TTC26, FLNA, and IFT52), and ANKRD55 itself was described as a novel intraflagellar transport complex protein[382]. Moreover, the BioGRID database also identified other members included in the same process (IFT46, TTC30A, and TTC30B). HuRi also described the interaction between ANKRD55 and relevant proteins in T-cell development, differentiation and activation (BACH2[383,384], BANP[385] and LEF1[386,387]) as well as CARD9[388], PHLPP1[389] and CTPS1[390] from other interactome studies.

We observed that ANKRD55 interacted with other autoimmune risk genes, such as TIMMDC1, KPNB1, ANAPC1, COPB1, BACH2 (HuRI) and LEF1 (HuRI) MS risk genes and C1QBP RA risk genes. Additionally, we also noticed that some ANKRD55 partners (PRMT5, DDX17, CDK5, NDUFA13, KPNA2, IPO7, MYBBP1A, RQCD1, PGAM5, GNB2L1, CNOT1, RUVBL1, RUVBL2, DHX30, IRS4, RPS3, ILF2, WDR36, ELAVL1, YBX1, SMC1A) interacted with MS and RA risk genes (Supplementary Table 9).

The large number of identified interactors suggests that ANKRD55 can act as a scaffold, probably exerting its function(s) in multiple protein complexes. No clear conclusion can be made on its precise biological role. However, based on interactome and subcellular localization analysis, ANKRD55 potentially is transported into the nucleus by the classical nuclear import pathway and it is involved somehow in: (i) mitosis, probably associated to mitotic spindle dynamics, as described for ANKRD53[391], another AR-containing protein; (ii) the control of the uptake and efflux of various solutes, (iii) ciliogenesis and its regulation. ANKRD55 could also be important in the T-cell development, activation and differentiation, a key immune cell in the pathogenesis of MS and RA.

CONCLUSIONS

The present work resulted in the following main conclusions:

1. The three *ANKRD55* transcripts were uniquely and highly expressed in $CD4^+$ T cells. Individual transcript levels were highly correlated and the homozygotes for the risk allele (CC) expressed significantly higher levels of 001 and 005 transcripts in $CD4^+$ T cells. rs6859219 regulated the expression of *ANKRD55* in $CD4^+$ T cells, qualifying rs6859219 as a novel eQTL for *ANKRD55*.
2. Preliminary data showed that rs71624119 is located in an enhancer region and regulated *ANKRD55* transcript 001 in HEK293-T-$Cas9^{p300}$ cells.
3. Endogenous ANKRD55 proteins is located primarily to the nucleus in primary $CD4^+$ T cells and Jurkat, U937, and SH-SY5Y human cell lines.
4. Overexpressed ANKRD55 isoforms are located in cytosol, membranous organelles and nucleus in HEK293 and HeLa human cell lines.
5. Retrieval of 005 and 001 ANKRD55 interactomes from transfected HEK293 and HeLa cell lines was optimized by experimentally testing and adjusting immunoprecipitation conditions.
6. For ANKRD55 interactome analysis a two-stage approach was adopted: (i) mass spectrometry of individual protein bands IPed with ANKRD55 in silver-stained gels; and (ii) mass spectrometry of gel slabs corresponding to the complete running lane of Sypro-stained gels. In this last approach, following subtraction of proteins of mock from those of 001 ANKRD55-transfected cells under stringent parameters, 148 specifically interacting proteins were found in total protein extracts and 22 in nuclear extracts.
7. The functional analysis suggested that the ANKRD55-protein partners from total protein extracts were related to nucleotide and ATP binding, enriched in nuclear transport terms and associated with cell cycle and RNA, lipid and amino acid metabolism. The enrichment analysis of the ANKRD55-protein partners from nuclear extracts are related to sumoylation, RNA binding, processes associated with cell cycle, RNA transport, nucleotide and ATP binding, among others.
8. The interaction between overexpressed ANKRD55 isoform 001 and endogenous RPS3, SMC1A, SMC3, CLTC, PRKDC, VIM, β-tubulin isoforms, and 14-3-3 isoforms were validated by WB or confocal microscopy.

9. Human ANKRD55 isoform 001 includes five Ser/Thr phosphorylation sites.

RESUMEN

La esclerosis múltiple (EM) es una enfermedad inflamatoria desmielinizante del sistema nervioso central (SNC), que produce una discapacidad crónica progresiva en la mayoría de las personas con este trastorno. Esta condición tiene una presentación heterogénea que puede involucrar los sistemas motor, sensorial, visual y autónomo. La variación en las manifestaciones clínicas se correlaciona con la diseminación espaciotemporal de las zonas dañadas de la patología dentro del SNC, afectando tanto a los tractos de la sustancia blanca como a la sustancia gris cortical y profunda. Estas lesiones son una marca distintiva de la EM y son causadas por la infiltración de células inmunes a través de la barrera hematoencefálica, que promueven la inflamación, desmielinización, gliosis y degeneración neuroaxonal, lo que conduce a la interrupción de la señalización neuronal.

La clasificación de las formas clínicas de la EM no es sencilla debido a la heterogeneidad del curso de la enfermedad y la sintomatología. La clasificación distingue tres patrones clínicos que son: 1) EM recurrente-remitente, 2) EM progresiva secundaria y 3) EM progresiva primaria. Por lo general, se diagnostica entre los 20 y 40 años. La enfermedad tiene una prevalencia heterogénea en todo el mundo, es más alta en América del Norte y Europa, y más baja en Asia Oriental y África subsahariana. La EM se presenta con más frecuencia en mujeres que en hombres (3:1).

Aunque algunos sugieren que la EM es principalmente un proceso neurodegenerativo con la intervención secundaria del sistema inmune, diversas evidencias apuntan a mecanismos de enfermedades mediadas por el sistema inmune debido a los siguientes hallazgos: 1) presencia de linfocitos T y B en las lesiones desmielinizantes; 2) las respuestas inmunes específicas de antígeno del SNC se detectan en la sangre periférica de pacientes con EM; 3) la enfermedad puede ser suprimida por terapias inmunomoduladoras; y 4) los estudios genéticos apuntan hacia el sistema inmune adaptativo, particularmente su componente de linfocitos T $CD4^+$.

La patogenia de la EM se explica mediante un modelo multifactorial que incorpora interacciones entre influencias genéticas, epigenéticas, infecciosas, nutricionales, climáticas u otras influencias ambientales. Todos estos factores pueden intervenir en la inmunidad adaptativa y/o innata, que se cree que es la vía principal modulada por las variantes de riesgo genético de la EM. El estilo de vida y los factores ambientales que

aumentan el riesgo de EM incluyen la exposición al humo del tabaco, la infección por el virus de Epstein-Barr, la obesidad de los adolescentes, la falta de exposición al sol o los bajos niveles de vitamina D. Varios estudios de asociación de genoma completo (GWAS) y estudios a gran escala dirigidos en los últimos 10 años, mostraron evidencia estadística inequívoca de la asociación de 200 variantes de susceptibilidad autosómicas fuera del complejo mayor de histocompatibilidad (MHC), una variante del cromosoma X y 32 asociaciones independientes dentro del MHC. Estos datos apoyan un modelo de herencia poligénica para la EM, en el que el riesgo está determinado por un único alelo de efecto moderado y muchos alelos de efecto mucho menor.

Una variante intrónica en *ANKRD55*, rs6859219, se identificó como factor de riesgo genético para la EM. Otros estudios han relacionado a *ANKRD55* con varias enfermedades autoinmunes, como la artritis reumatoide (AR), enfermedad de Crohn, diabetes tipo 1, enfermedad celiaca, miopatías inflamatorias, y otro tipo de enfermedades, como el trastorno de estrés postraumático, enfermedad de Alzheimer (disminución cognitiva) y diabetes tipo 2. Recientemente, *ANKRD55* se asoció con niveles de IgA. Además, otra variante cercana a este gen también se vinculó con cambios en la N-glicosilación de IgG, que se altera en la AR. Curiosamente, la defectuosa N-glicosilación también está implicada en la EM.

El locus *ANKRD55*, ubicado en el cromosoma 5q11.2, codifica para la proteína con dominio de repetición de anquirina 55, cuya función actualmente se desconoce. La versión *75* de la base de datos *Ensembl,* incluye seis variantes de transcripción: cuatro son codificantes para proteínas (001, 002, 005 y 006), y las dos restantes transcritos no codificantes (007 y 008). Sin embargo, solo los transcritos codificantes de *Ensembl* 001 (correspondientes a la forma completa) y 005, están incluidos en las bases de datos UniProt. ANKRD55 pertenece a una familia que contiene dominios de repetición de anquirina (RA). Este motivo está formado por 30-34 residuos aminoacídicos, es uno de los más comunes de la naturaleza, modular y de interacción proteína-proteína (PPI). La RA se ha encontrado en proteínas con diversas funciones, como la transcripción, regulación del ciclo celular, integridad del citoesqueleto, respuesta inflamatoria, desarrollo, señalización célula-célula y diversos fenómenos de transporte.

El propósito de este estudio es conocer más, en torno a las funciones biológicas de ANKRD55 asociadas con la EM y AR y aumentar nuestro conocimiento sobre su papel en las enfermedades autoinmunes. Para abordar el objetivo del proyecto, perseguimos los siguientes objetivos específicos: 1) Identificación de la principal fuente celular de *ANKRD55* en células mononucleares de sangre periférica (PBMC); 2) determinación de la variante intrónica de *ANKRD55* asociada a la EM y AR como *expression quantitative trait loci* (eQTL) para tres transcritos de *ANKRD55* en cinco subpoblaciones de PBMC; 3) estudio de cinco variantes intrónicas de *ANKRD55* asociadas a la EM y AR, localizadas en regiones *enhancer* bioinformaticamente predichos, por su capacidad para regular la expresión de *ANKRD55* y los genes con los que posiblemente interacciona, *IL6ST* e *IL31RA*; 4) análisis de la localización subcelular de la forma endógena y recombinante de ANKRD55 en células inmunes y no inmunes; 5) identificación y validación de posibles proteínas con las que interacciona ANKRD55 en varios compartimentos subcelulares.

Teniendo en cuenta que *ANKRD55* es un factor de riesgo genético para diversas enfermedades autoinmunes, es probable que tenga un papel en el sistema inmunológico. Previamente en el grupo de investigación de Neurogenomiks analizamos la asociación de rs6859219 con la expresión de *ANKRD55* por qPCR en pacientes con EM y controles sanos. Cuando los individuos se agruparon por genotipo, los homocigotos para el alelo de riesgo (CC), mostraron una expresión significativamente mayor de los tres transcritos *ANKRD55* que los portadores del alelo protector (AA). No se encontraron diferencias significativas entre los pacientes con EM y los controles. A causa de la mezcla heterogénea de tipos celulares que componen las PBMC, nos propusimos identificar la principal fuente celular de expresión de *ANKRD55*. Para ello, se aislaron subpoblaciones de células $CD4^+$, $CD8^+$, $CD14^+$, $CD19^+$ y $CD56^+$ de PBMC de 23 controles sanos. Los resultados mostraron que los tres transcritos de *ANKRD55* se expresaron de forma única y elevada en las células T $CD4^+$. Los niveles de transcritos individuales estaban altamente correlacionados y los homocigotos para el alelo de riesgo expresaron niveles significativamente más altos de transcritos 001 y 005 en células T $CD4^+$. Por ello, encontramos que rs6859219 regulaba la expresión de *ANKRD55* en PBMC y linfocitos T $CD4^+$, calificando al rs6859219 como un nuevo eQTL para *ANKRD55*.

Posteriormente, analizamos cinco variantes *enhancer* bioinformáticamente predichas de *ANKRD55* para determinar su capacidad moduladora de la expresión de *ANKRD55* y los genes con los que posiblemente interacciona, *IL6ST* e *IL31RA*. En este caso, se usó el sistema CRISPR/*d*Cas9 acoplado a la histona acetiltransferasa p300 que consigue la acetilación del residuo 27 de lisina de la histona H3, aumentando la transcripción. Los datos preliminares de dos experimentos independientes mostraron que la expresión del transcrito *ANKRD5*5 001, se incrementó en las células HEK293-T-Cas9^{p300}, que expresan de forma estable el constructo que modifica la región que incluye el rs71624119.

En 2017, el primer análisis sobre la organización de las proteínas en una célula utilizando dos enfoques diferentes, imágenes de alto contenido y proteómica espacial, reveló que aproximadamente la mitad de las proteínas se pueden encontrar en más de un compartimento subcelular. La mala localización de proteínas puede estar asociada con disfunción celular y enfermedad. Por lo tanto, el conocimiento de la distribución espacial de las proteínas a nivel subcelular es esencial para comprender su función, las interacciones y los mecanismos celulares. Además, las funciones celulares están mediadas por complejas redes de interacciones físicas y funcionales entre macromoléculas, como ADN, ARN, proteínas y lípidos, así como moléculas más pequeñas, como los metabolitos. Muchas proteínas ejercen sus funciones dentro de las células en el contexto de complejos de proteínas. Por lo tanto, identificar las PPI, definidas como un contacto físico directo entre dos proteínas, es fundamental para obtener información sobre su función biológica, ya que una proteína no caracterizada que se une a una proteína cuyo papel en la célula se entiende, es probable que posea una función relacionada.

Por este motivo, estudiamos la expresión endógena de ANKRD55 y la localización intracelular en células T CD4$^+$ primarias y en las líneas celulares humanas Jurkat, U937 y SH-SY5Y mediante microscopía confocal, enriquecimiento nuclear mediante ultracentrifugación en gradiente de sacarosa y fraccionamiento bioquímico para obtener fracción citosólica, orgánulos membranosos y fracción nuclear. Los resultados mostraron que las isoformas endógenas de ANKRD55 se ubican principalmente en el núcleo de las células T CD4$^+$ primarias y las líneas celulares humanas Jurkat, U937 y SH-SY5Y.

Para estudiar más a fondo la localización subcelular de las isoformas de ANKRD55, sobreexpresamos transitoriamente las isoformas humanas de ANKRD55 001 y 005 en células HEK293 y HeLa y fraccionamos las células en tres compartimentos subcelulares. Las dos isoformas sobreexpresadas de ANKRD55 se detectaron en citosol, orgánulos membranosos y núcleos en ambas líneas celulares. Además, había presentes productos inmunorreactivos truncados, probablemente generados por eventos proteolíticos específicos de las proteínas de isoformas intactas. Los resultados también mostraron la isoforma endógena 001 en ambas líneas celulares, lo que demuestra la capacidad de las células HEK293 y Hela para producir ANKRD55 y, por lo tanto, comprometerse con su(s) vía(s) biológica(s).

Para identificar los interactomas de ANKRD55, realizamos una inmunoprecipitación (IP) de ANKRD55 recombinante incluyendo FLAG *tag* y las proteínas con las que interacciona, seguido de un análisis mediante cromatografía líquida de nano flujo acoplada a análisis de espectrometría de masas en tándem (nLC-MS/MS). Este análisis se realizó en dos etapas: 1) espectrometría de masas de bandas de proteínas individuales inmunoprecipitadas con ANKRD55 en geles teñidos con plata de distintos extractos proteicos de las líneas celulares HEK293 y HeLa; y 2) espectrometría de masas de cortes de gel correspondientes a la IP cargada en geles teñidos con Sypro de extracto proteico total y extracto nuclear de HEK293. En este último enfoque, después de eliminar, bajo parámetros estrictos, las proteínas presentes en el control negativo a las de células transfectadas con ANKRD55 001, se encontraron 148 proteínas que interactúan específicamente en extractos de proteínas totales y 22 en extractos nucleares. La interacción entre la isoforma 001 de ANKRD55 sobreexpresada y la forma endógena de RPS3, SMC1A, SMC3, CLTC, PRKDC, VIM, isoformas de β-tubulina e isoformas de 14-3-3 fueron validadas mediante WB y microscopía confocal.

El análisis funcional a través de la herramienta DAVID, sugirió que las proteínas que interaccionan con ANKRD55 de los extractos de proteínas totales, se relacionan con la unión de nucleótidos y ATP, están enriquecidas en términos de transporte nuclear y se asociaron con el ciclo celular y metabolismo de ARN, lípidos y aminoácidos. El análisis de enriquecimiento de las proteínas que interaccionan con ANKRD55 de extractos nucleares,

se relaciona con la sumoilación, unión de ARN, procesos asociados con el ciclo celular, el transporte de ARN, unión de nucleótidos y ATP, entre otros.

Un importante subconjunto de proteínas identificadas en este análisis pertenece a la familia de proteínas 14-3-3. Las proteínas 14-3-3 son proteínas de unión a fosfoserina / fosfotreonina expresadas de forma ubicua que son miembros de una gran familia de isoformas. Por ello, analizamos si ANKRD55 se encontraba fosforilada. Los resultados mostraron que la isoforma 001 de ANKRD55 incluye cinco sitios de fosforilación de serina/treonina.

La identificación de un gran número de proteínas que interaccionan con ANKRD55, sugiere que ANKRD55 puede actuar como un *scaffold*, probablemente ejerciendo su(s) función(es) en complejos compuestos por múltiples proteínas. Por ello, no se puede concluir de forma clara su papel biológico preciso. Sin embargo, según el análisis de localización subcelular y el interactoma, ANKRD55 se transporta potencialmente al núcleo por la vía de importación nuclear clásica y está involucrada de alguna forma en: 1) mitosis, probablemente asociado a la dinámica del huso mitótico; 2) el control de captación y flujo de salida de varios solutos y 3) ciliogénesis y su regulación. ANKRD55 puede ser importante en el desarrollo, activación y diferenciación de los linfocitos T, una célula inmunitaria clave en la patogénesis de EM y AR.

REFERENCES

1. Sawcer, S., Franklin, R. J. M. & Ban, M. Multiple sclerosis genetics. *Lancet Neurol.* **13,** 700–709 (2014).
2. Compston, A. & Coles, A. Multiple sclerosis. *Lancet* **372,** 1502–1517 (2008).
3. Kearney, H. *et al.* Cervical cord lesion load is associated with disability independently from atrophy in MS. *Neurology* **84,** 367–373 (2015).
4. Harrison, D. In the Clinic. Multiple Sclerosis. *Ann. Intern. Med.* **160,** ITC4-2-ITC4-18 (2014).
5. Frischer, J. M. *et al.* The relation between inflammation and neurodegeneration in multiple sclerosis brains. *Brain* **132,** 1175–1189 (2009).
6. Poser, C. M. The dissemination of multiple sclerosis: A Viking saga? A historical essay. *Ann. Neurol.* **36,** S231–S243 (1994).
7. Medaer, R. Does the history of multiple sclerosis go back as far as the 14th century? *Acta Neurol. Scand.* **60,** 189–192 (1979).
8. Hauser, S. L. & Oksenberg, J. R. The Neurobiology of Multiple Sclerosis: Genes, Inflammation, and Neurodegeneration. *Neuron* **52,** 61–76 (2006).
9. Reich, D. S., Lucchinetti, C. F. & Calabresi, P. A. Multiple Sclerosis. *N. Engl. J. Med.* **378,** 169–180 (2018).
10. Murray, T. J. Multiple Sclerosis: The History of a Disease. *J. R. Soc. Med.* **98,** 289 (2005).
11. Charcot, J. M. Histologie de la sclérose en plaques. *Gaz. des Hop.* **41,** 554–555 (1868).
12. Raphael, I., Webb, J., Stuve, O., Haskins, W. & Forsthuber, T. Body fluid biomarkers in multiple sclerosis: ow far we have come and how they could affect the clinic now and in the future. *Expert Rev Clin Immunol* **11,** 69–91 (2015).
13. Dawson, J. W. The Histology of Disseminated Sclerosis. *Edinb Med J* **17,** 229–241 (1916).
14. Rolak, L. A. MS: The Basic Facts. *Clin Med Res* **1,** 61–62 (2003).
15. Rivers, T. M. Observations on Attempts to Produce Acute Disseminated Encephalomyelitis in Monkeys. *J. Exp. Med.* **58,** 39–53 (1933).
16. Young, I. R. *et al.* Nuclear magnetic resonance imaging of the brain in multiple sclerosis. *Lancet* **2,** 1063–1066 (1981).
17. Brass, S. D., Weiner, H. L. & Hafler, D. A. Multiple Sclerosis. in *The Autoimmune Diseases* 615–632 (Elsevier Inc., 2006).
18. Arnason, B. G. W. Interferon beta in multiple sclerosis. *Neurology* **43,** 641 (1993).
19. Didonna, A. & Oksenberg, J. The Genetics of Multiple Sclerosis. in *Multiple Sclerosis: Perspectives in Treatment and Pathogenesis* 3–16 (Codon Publications, 2017).
20. Hofker, M. H., Fu, J. & Wijmenga, C. The genome revolution and its role in understanding complex diseases. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842,** 1889–1895 (2014).
21. Huang, Q. Genetic Study of Complex Diseases in the Post-GWAS Era. *J. Genet. Genomics* **42,** 87–98 (2015).
22. Schaeffer, J., Cossetti, C., Mallucci, G. & Pluchino, S. Multiple Sclerosis. in *Neurobiology of Brain Disorders: Biological Basis of Neurological and Psychiatric Disorders* 497–520 (Elsevier, 2014).
23. Compston, A. & Coles, A. Multiple sclerosis. *Lancet* **359,** 1221–1231 (2002).
24. Steinman, L. Multiple sclerosis : a two-stage disease. *Nat. Immunol.* **2,** 762–764 (2001).
25. Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B. & Ciccarelli, O. Multiple sclerosis.

Lancet **6736,** 1–15 (2018).

26. Katz, S. Classification, diagnosis, and differential diagnosis of multiple sclerosis. *Curr. Opin. Neurol.* **28,** 193–205 (2015).
27. Fisniku, L. K. *et al.* Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain* **131,** 808–817 (2008).
28. Dendrou, C. A., Fugger, L. & Friese, M. A. Immunopathology of multiple sclerosis. *Nat. Rev. Immunol.* **15,** 545–558 (2015).
29. Okuda, D. T. *et al.* Radiologically isolated syndrome: 5-year risk for an initial clinical event. *PLoS One* **9,** e90509 (2014).
30. Kantarci, O. H. *et al.* Primary Progressive Multiple Sclerosis Evolving from Radiologically Isolated Syndrome. *Ann. Neurol.* **79,** 288–294 (2016).
31. Thompson, A. J. *et al.* Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* **17,** 162–173 (2018).
32. Browne, P. *et al.* Atlas of Multiple Sclerosis 2013: A growing global problem with widespread inequity. *Neurology* **83,** 1022–1024 (2014).
33. Antel, J., Antel, S., Caramanos, Z., Arnold, D. L. & Kuhlmann, T. Primary progressive multiple sclerosis: Part of the MS disease spectrum or separate disease entity? *Acta Neuropathol.* **123,** 627–638 (2012).
34. Leray, E., Moreau, T., Fromont, A. & Edan, G. Epidemiology of multiple sclerosis. *Rev. Neurol. (Paris).* **172,** 3–13 (2016).
35. Baranzini, S. E. & Oksenberg, J. R. The Genetics of Multiple Sclerosis: From 0 to 200 in 50 Years. *Trends Genet.* **33,** 960–970 (2017).
36. Koch-Henriksen, N. & Sørensen, P. S. The changing demographic pattern of multiple sclerosis epidemiology. *Lancet Neurol.* **9,** 520–532 (2010).
37. Orton, S. M. *et al.* Sex ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol.* **5,** 932–936 (2006).
38. Eskandarieh, S., Heydarpour, P., Minagar, A., Pourmand, S. & Sahraian, M. A. Multiple Sclerosis Epidemiology in East Asia, South East Asia and South Asia: A Systematic Review. *Neuroepidemiology* **46,** 209–221 (2016).
39. Benito-León, J. Are the prevalence and incidence of multiple sclerosis changing? *Neuroepidemiology* **36,** 148–149 (2011).
40. Bergamaschi, R. Prognostic Factors in Multiple Sclerosis. *Int. Rev. Neurobiol.* **79,** 423–447 (2007).
41. Stys, P. K., Zamponi, G. W., Minnen, J. Van & Geurts, J. J. G. Will the real mulitple sclerosis please stand up? *Nat Rev Neurosci* **13,** 507–514 (2012).
42. Lassmann, H. Pathology and disease mechanisms in different stages of multiple sclerosis. *J. Neurol. Sci.* **333,** 1–4 (2013).
43. Weissert, R. The immune pathogenesis of multiple sclerosis. *J. Neuroimmune Pharmacol.* **8,** 857–866 (2013).
44. Hartung, H. P. & Kieseier, B. C. The new therapeutic landscape in multiple sclerosis: Exciting times and new perspectives. *Curr. Opin. Neurol.* **27,** 243–245 (2014).
45. Yadav, S. K., Mindur, J. E., Ito, K. & Dhib-Jalbut, S. Advances in the immunopathogenesis of multiple sclerosis. *Curr. Opin. Neurol.* **28,** 206–219 (2015).
46. Gandhi, R., Laroni, A. & Weiner, H. L. Role of the innate immune system in the

pathogenesis of multiple sclerosis. *J. Neuroimmunol.* **221,** 7–14 (2010).

47. Grimholt, U. MHC and Evolution in Teleosts. *Biology (Basel).* **5,** 6 (2016).

48. Hanna, S. & Etzioni, A. MHC class i and II deficiencies. *J. Allergy Clin. Immunol.* **134,** 269–275 (2014).

49. Legroux, L. & Arbour, N. Multiple Sclerosis and T Lymphocytes: An Entangled Story. *J. Neuroimmune Pharmacol.* **10,** 528–546 (2015).

50. Harkiolaki, M. *et al.* T Cell-Mediated Autoimmune Disease Due to Low-Affinity Crossreactivity to Common Microbial Peptides. *Immunity* **30,** 348–357 (2009).

51. Münz, C., Lünemann, J. D., Getts, M. T. & Miller, S. D. Antiviral immune responses: Triggers of or triggered by autoimmunity? *Nat. Rev. Immunol.* **9,** 246–258 (2009).

52. Olson, J. K., Ercolini, a M. & Miller, S. D. A virus-induced molecular mimicry model of multiple sclerosis. *Curr. Top. Microbiol. Immunol.* **296,** 39–53 (2005).

53. Ji, Q., Perchellet, A. & Goverman, J. M. Viral infection triggers central nervous system autoimmunity via activation of CD8 + T cells expressing dual TCRs. *Nat. Immunol.* **11,** 628–634 (2010).

54. Traugott, U., Reinherz, E. L. & Raine, C. S. Multiple sclerosis. Distribution of T cells, T cell subsets and Ia-positive macrophages in lesions of different ages. *J. Neuroimmunol.* **4,** 201–221 (1983).

55. Hauser, S. L. *et al.* Immunohistochemical analysis of the cellular infiltrate in multiple sclerosis lesions. *Ann. Neurol.* **19,** 578–587 (1986).

56. Baecher-Allan, C., Kaskow, B. J. & Weiner, H. L. Multiple Sclerosis: Mechanisms and Immunotherapy. *Neuron* **97,** 742–768 (2018).

57. Ransohoff, R. M. & Engelhardt, B. The anatomical and cellular basis of immune surveillance in the central nervous system. *Nat. Rev. Immunol.* **12,** 623–635 (2012).

58. Louveau, A. *et al.* Structural and functional features of central nervous system lymphatic vessels. *Nature* **523,** 337–341 (2015).

59. Heneka, M. T., Kummer, M. P. & Latz, E. Innate immune activation in neurodegenerative disease. *Nat. Rev. Immunol.* **14,** 463–477 (2014).

60. Popescu, B. F. G. & Lucchinetti, C. F. Pathology of Demyelinating Diseases. *Annu. Rev. Pathol. Mech. Dis.* **7,** 185–217 (2012).

61. Peterson, P., Org, T. & Rebane, A. Transcriptional regulation by AIRE: Molecular mechanisms of central tolerance. *Nat. Rev. Immunol.* **8,** 948–957 (2008).

62. Tada, T., Takemori, T., Okumura, K., Nonaka, M. & Tokuhisa, T. Two distinct types of helper T cells involved in the secondary antibody response: independent and synergistic effects of Ia- and Ia+ helper T cells. *J. Exp. Med.* **147,** 446–58 (1978).

63. Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature* **441,** 235–238 (2006).

64. Panitch, H. S., Haley, A. S., Hirsch, R. L. & Johnson, K. P. Exacerbations of Multiple Sclerosis in Patients Treated With Gamma Interferon. *Lancet* **329,** 893–895 (1987).

65. Havrdová, E. *et al.* Activity of secukinumab, an anti-IL-17A antibody, on brain lesions in RRMS: results from a randomized, proof-of-concept study. *J. Neurol.* **263,** 1287–1295 (2016).

66. Luger, D. *et al.* Either a Th17 or a Th1 effector response can drive autoimmunity: conditions of disease induction affect dominant effector category. *J. Exp. Med.* **205,** 799–

810 (2008).

67. Lee, S. Y. & Goverman, J. M. The Influence of T Cell Ig Mucin-3 Signaling on Central Nervous System Autoimmune Disease Is Determined by the Effector Function of the Pathogenic T Cells. *J. Immunol.* **190,** 4991–4999 (2013).

68. Kozovska, M. E. *et al.* Interferon beta induces T-helper 2 immune deviation in MS. *Neurology* **53,** 1692–1692 (1999).

69. Miller, A. *et al.* Treatment of multiple sclerosis with Copolymer-1 (Copaxone®): Implicating mechanisms of Th1 to Th2/Th3 immune-deviation. *J. Neuroimmunol.* **92,** 113–121 (1998).

70. Zoghi, S. *et al.* Cytokine secretion pattern in treatment of lymphocytes of multiple sclerosis patients with fumaric acid esters. *Immunol. Invest.* **40,** 581–596 (2011).

71. Barry, M. & Bleackley, R. C. Cytotoxic T lymphocytes: all roads lead to death. *Nat. Rev. Immunol.* **2,** 401–409 (2002).

72. Zang, Y. C. Q. *et al.* Increased CD8+ cytotoxic T cell responses to myelin basic protein in multiple sclerosis. *J. Immunol.* **172,** 5120–5127 (2004).

73. Melzer, N., Meuth, S. G. & Wiendl, H. CD8+ T cells and neuronal damage: direct and collateral mechanisms of cytotoxicity and impaired electrical excitability. *FASEB J.* **23,** 3659–3673 (2009).

74. Huber, M., Heink, S. & Pagenstecher, A. IL-17A secretion by CD8+ T cells supports Th17-mediated autoimmune encephalomyelitis. *J Clin Invest.* **123,** 247–260 (2013).

75. Larochelle, C. *et al.* Melanoma cell adhesion molecule identifies encephalitogenic T lymphocytes and promotes their recruitment to the central nervous system. *Brain* **135,** 2906–2924 (2012).

76. Duan, H. *et al.* Targeting endothelial CD146 attenuates neuroinflammation by limiting lymphocyte extravasation to the CNS. *Sci. Rep.* **3,** 1–11 (2013).

77. Greenfield, A. L. & Hauser, S. L. B Cell Therapy for Multiple Sclerosis: Entering an Era. *Ann. Neurol.* **83,** 13–26 (2017).

78. Lovato, L. *et al.* Related B cell clones populate the meninges and parenchyma of patients with multiple sclerosis. *Brain* **134,** 534–541 (2011).

79. Henderson, A. P. D., Barnett, M. H., Parratt, J. D. E. & Prineas, J. W. Multiple sclerosis: Distribution of inflammatory cells in newly forming lesions. *Ann. Neurol.* **66,** 739–753 (2009).

80. Cepok, S. *et al.* Patterns of cerebrospinal fluid pathology correlate with disease progression in multiple sclerosis. *Brain* **124,** 2169–2176 (2001).

81. Hauser, S. L. *et al.* B-cell Depletion with Rituximab in Relapsing-Remitting Multiple Sclerosis. *N Engl J Med* **358,** 676–688 (2008).

82. Skov, A. G., Skov, T. & Frederiksen, J. L. Oligoclonal bands predict multiple sclerosis after optic neuritis: A literature survey. *Mult. Scler. J.* **17,** 404–410 (2011).

83. Brettschneider, J. *et al.* IgG antibodies against measles, rubella, and varicella zoster virus predict conversion to multiple sclerosis in clinically isolated syndrome. *PLoS One* **4,** 1–5 (2009).

84. Sakaguchi, S. *et al.* Immunologic tolerance maintained by CD25+ CD4+ regulatory T cells: Their common role in controlling autoimmunity, tumor immunity, and transplantation tolerance. *Immunol. Rev.* **182,** 18–32 (2001).

85. Venken, K. *et al.* Natural Naive CD4 $^{+}$ CD25 $^{+}$ CD127 ^{low} Regulatory T Cell (Treg)

Development and Function Are Disturbed in Multiple Sclerosis Patients: Recovery of Memory Treg Homeostasis during Disease Progression. *J. Immunol.* **180,** 6411–6420 (2008).

86. Lu, L. F. & Rudensky, A. Molecular orchestration of differentiation and function of regulatory T cells. *Genes Dev.* **23,** 1270–1282 (2009).

87. Martinez-Forero, I. *et al.* IL-10 suppressor activity and ex vivo Tr1 cell function are impaired in multiple sclerosis. *Eur. J. Immunol.* **38,** 576–586 (2008).

88. Gregori, S., Goudy, K. S. & Roncarolo, M. G. The cellular and molecular mechanisms of immuno-suppression by human type 1 regulatory T cells. *Front. Immunol.* **3,** 1–12 (2012).

89. Baecher-Allan, C., Brown, J. A., Freeman, G. J. & Hafler, D. A. CD4+CD25high Regulatory Cells in Human Peripheral Blood. *J. Immunol.* **167,** 1245–1253 (2001).

90. Schmidt, A., Oberle, N. & Krammer, P. H. Molecular mechanisms of Treg-mediated T cell suppression. *Front. Immunol.* **3,** 1–20 (2012).

91. Pacholczyk, R. & Kern, J. The T-cell receptor repertoire of regulatory T cells. *Immunology* **125,** 450–458 (2008).

92. Lee, H. M., Bautista, J. L., Scott-Browne, J., Mohan, J. F. & Hsieh, C. S. A Broad Range of Self-Reactivity Drives Thymic Regulatory T Cell Selection to Limit Responses to Self. *Immunity* **37,** 475–486 (2012).

93. Schadenberg, A. W. L. *et al.* FOXP3+CD4+ Tregs lose suppressive potential but remain anergic during transient inflammation in human. *Eur. J. Immunol.* **41,** 1132–1142 (2011).

94. O'Connor, R. A., Malpass, K. H. & Anderton, S. M. The Inflamed Central Nervous System Drives the Activation and Rapid Proliferation of Foxp3+ Regulatory T Cells. *J. Immunol.* **179,** 958–966 (2007).

95. Viglietta, V., Baecher-Allan, C., Weiner, H. L. & Hafler, D. A. Loss of Functional Suppression by $CD4^+CD25^+$ Regulatory T Cells in Patients with Multiple Sclerosis. *J. Exp. Med.* **199,** 971–979 (2004).

96. Haas, J. *et al.* Reduced suppressive effect of CD4+CD25high regulatory T cells on the T cell immune response against myelin oligodendrocyte glycoprotein in patients with multiple sclerosis. *Eur. J. Immunol.* **35,** 3343–3352 (2005).

97. Costantino, C. M., Baecher-Allan, C. M. & Hafler, D. A. Human regulatory T cells and autoimmunity. *Eur. J. Immunol.* **38,** 921–924 (2008).

98. Schneider, A. *et al.* In Active Relapsing-Remitting Multiple Sclerosis, Effector T Cell Resistance to Adaptive Tregs Involves IL-6-Mediated Signaling. *Sci. Transl. Med.* **5,** 170ra15 (2013).

99. Bhela, S. *et al.* Nonapoptotic and Extracellular Activity of Granzyme B Mediates Resistance to Regulatory T Cell (Treg) Suppression by HLA-DR^- $CD25^{hi}$ $CD127^{lo}$ Tregs in Multiple Sclerosis and in Response to IL-6. *J. Immunol.* **194,** 2180–2189 (2015).

100. Kimura, A. & Kishimoto, T. IL-6: Regulator of Treg/Th17 balance. *Eur. J. Immunol.* **40,** 1830–1835 (2010).

101. Becher, B. & Segal, B. M. T H17 cytokines in autoimmune neuro-inflammation. *Curr. Opin. Immunol.* **23,** 707–712 (2011).

102. Groux, H. *et al.* A CD4+ T-cell subset inhibits antigen-specific T-cell responses and prevents colitis. *Nature* **389,** 737–742 (1997).

103. Astier, A. L., Meiffren, G., Freeman, S. & Hafler, D. A. Alterations in CD46-mediated Tr1 regulatory T cells in patients with multiple sclerosis. *J Clin Invest* **116,** 3252–3257 (2006).

104. Mayo, L. *et al.* IL-10-dependent Tr1 cells attenuate astrocyte activation and ameliorate chronic central nervous system inflammation. *Brain* **139,** 1939–1957 (2016).

105. Chiarini, M. *et al.* Modulation of the central memory and Tr1-like regulatory T cells in multiple sclerosis patients responsive to interferon-beta therapy. *Mult. Scler. J.* **18,** 788–798 (2012).

106. Mayo, L., Quintana, F. J. & Weiner, H. L. The innate immune system in demyelinating disease. *Immunol. Rev.* **248,** 170–187 (2012).

107. Comabella, M., Montalban, X., Münz, C. & Lünemann, J. D. Targeting dendritic cells to treat multiple sclerosis. *Nat. Rev. Neurol.* **6,** 499–507 (2010).

108. Mishra, M. K. & Yong, V. W. Myeloid cells — targets of medication in multiple sclerosis. *Nat. Rev. Neurol.* **12,** 539–551 (2016).

109. Wu, H. Y. *et al.* In vivo induction of Tr1 cells via mucosal dendritic cells and AHR signaling. *PLoS One* **6,** e23618 (2011).

110. Quintana, F. J. *et al.* Control of Treg and TH17 cell differentiation by the aryl hydrocarbon receptor. *Nature* **453,** 65–71 (2008).

111. Severson, C. & Hafler, D. A. T-Cells in Multiple Sclerosis. *Results Probl. Cell Differ.* **51,** 75–98 (2010).

112. Barreira Da Silva, R. & Münz, C. Natural killer cell activation by dendritic cells: Balancing inhibitory and activating signals. *Cell. Mol. Life Sci.* **68,** 3505–3518 (2011).

113. Miller, S. D., McMahon, E. J., Schreiner, B. & Bailey, S. L. Antigen presentation in the CNS by myeloid dendritic cells drives progression of relapsing experimental autoimmune encephalomyelitis. *Ann. N. Y. Acad. Sci.* **1103,** 179–191 (2007).

114. McMahon, E. J., Bailey, S. L., Castenada, C. V., Waldner, H. & Miller, S. D. Epitope spreading initiates in the CNS in two mouse models of multiple sclerosis. *Nat. Med.* **11,** 335–339 (2005).

115. Greter, M. *et al.* Dendritic cells permit immune invasion of the CNS in an animal model of multiple sclerosis. *Nat. Med.* **11,** 328–334 (2005).

116. Stasiolek, M. *et al.* Impaired maturation and altered regulatory function of plasmacytoid dendritic cells in multiple sclerosis. *Brain* **129,** 1293–1305 (2006).

117. Weiner, H. L. A shift from adaptive to innate immunity: A potential mechanism of disease progression in multiple sclerosis. *J. Neurol.* **255,** 3–11 (2008).

118. Karni, A. *et al.* Innate Immunity in Multiple Sclerosis: Myeloid Dendritic Cells in Secondary Progressive Multiple Sclerosis Are Activated and Drive a Proinflammatory Immune Response. *J. Immunol.* **177,** 4196–4202 (2006).

119. Karni, A., Koldzic, D. N., Bharanidharan, P., Khoury, S. J. & Weiner, H. L. IL-18 is linked to raised IFN-γ in multiple sclerosis and is induced by activated CD4+T cells via CD40-CD40 ligand interactions. *J. Neuroimmunol.* **125,** 134–140 (2002).

120. Balashov, K. E., Smith, D. R., Khoury, S. J., Hafler, D. A. & Weiner, H. L. Increased interleukin 12 production in progressive multiple sclerosis: Induction by activated CD4+ T cells via CD40 ligand. *Proc. Natl. Acad. Sci.* **94,** 599–603 (1997).

121. Hernández-pedro, N. Y., Espinosa-ramirez, G., Pérez De Cruz, V., Pineda, B. & Sotelo, J. Initial Immunopathogenesis of Multiple Sclerosis : Innate Immune Response. *Clin Dev Immunol* **2013,** 1–15 (2013).

122. Nimmerjahn, A., Kirchhoff, F. & Helmchen, F. Resting microglial cells are highly dynamic surveillants of brain parenchyma in vivo. *Science.* **308,** 1314–1318 (2005).

123. Ponomarev, E. D., Shriver, L. P., Maresz, K. & Dittel, B. N. Microglial cell activation and proliferation precedes the onset of CNS autoimmunity. *J. Neurosci. Res.* **81,** 374–389 (2005).

124. Rasmussen, S. *et al.* Persistent activation of microglia is associated with neuronal dysfunction of callosal projecting pathways and multiple sclerosis-like lesions in relapsing-remitting experimental autoimmune encephalomyelitis. *Brain* **130,** 2816–2829 (2007).

125. Takeuchi, H. *et al.* Tumor necrosis factor-α induces neurotoxicity via glutamate release from hemichannels of activated microglia in an autocrine manner. *J. Biol. Chem.* **281,** 21362–21368 (2006).

126. Czeh, M., Gressens, P. & Kaindl, A. M. The yin and yang of microglia. *Dev. Neurosci.* **33,** 199–209 (2011).

127. Saijo, K. & Glass, C. K. Microglial cell origin and phenotypes in health and disease. *Nat. Rev. Immunol.* **11,** 775–787 (2011).

128. Rawji, K. S. & Yong, V. W. The Benefits and Detriments of Macrophages / Microglia in Models of Multiple Sclerosis. *Clin. Dev. Immunol.* **2013,** 1–13 (2013).

129. Casano, A. M. & Peri, F. Microglia: Multitasking specialists of the brain. *Dev. Cell* **32,** 469–477 (2015).

130. Hume, D. A. The mononuclear phagocyte system. *Curr. Opin. Immunol.* **18,** 49–53 (2006).

131. Fogel, L. A., Yokoyama, W. M. & French, A. R. Natural killer cells in human autoimmune disorders. *Arthritis Res. Ther.* **15,** 216 (2013).

132. Schleinitz, N., Vély, F., Harlé, J. R. & Vivier, E. Natural killer cells in human autoimmune diseases. *Immunology* **131,** 451–458 (2010).

133. Caruana, P., Lemmert, K., Ribbons, K., Lea, R. & Lechner-Scott, J. Natural killer cell subpopulations are associated with MRI activity in a relapsing-remitting multiple sclerosis patient cohort from Australia. *Mult. Scler.* **23,** 1479–1487 (2017).

134. Laroni, A. *et al.* Dysregulation of regulatory CD56bright NK cells/T cells interactions in multiple sclerosis. *J. Autoimmun.* **72,** 8–18 (2016).

135. Traugott, U. Characterization and distribution of lymphocyte subpopulations in multiple sclerosis plaques versus autoimmune demyelinating lesions. *Springer Semin. Immunopathol.* **8,** 71–95 (1985).

136. Bielekova, B. *et al.* Regulatory CD56(bright) natural killer cells mediate immunomodulatory effects of IL-2Ralpha-targeted therapy (daclizumab) in multiple sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 5941–5946 (2006).

137. Morandi, B. *et al.* Role of natural killer cells in the pathogenesis and progression of multiple sclerosis. *Pharmacol. Res.* **57,** 1–5 (2008).

138. Infante-Duarte, C. Frequency of blood CX3CR1-positive natural killer cells correlates with disease activity in multiple sclerosis patients. *FASEB J.* **19,** 1902–1904 (2005).

139. Saraste, M., Irjala, H. & Airas, L. Expansion of CD56Bright natural killer cells in the peripheral blood of multiple sclerosis patients treated with interferon-beta. *Neurol. Sci.* **28,** 121–126 (2007).

140. Olsson, T., Barcellos, L. F. & Alfredsson, L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat. Rev. Neurol.* **13,** 26–36 (2016).

141. Berg-Hansen, P. *et al.* Prevalence of multiple sclerosis among immigrants in Norway. *Mult. Scler.* **21,** 695–702 (2015).

142. Gale, C. R. & Martyn, C. N. Migrant studies in multiple sclerosis. *Prog. Neurobiol.* **47,** 425–448 (1995).

143. Lucas, R., Byrne, S. N., Correale, J., Ilschner, S. & Hart, P. Ultraviolet radiation, vitamin D and multiple sclerosis. *Neurodegener. Dis. Manag.* **5,** 413–24 (2015).

144. Munger, K. L., Levin, L. I., Hollis, B. W., Howard, N. S. & Ascherio, A. Serum 25-Hydroxyvitamin D Levels and Risk of Multiple Sclerosis. *JAMA* **296,** 2832–2838 (2015).

145. Bjørnevik, K. *et al.* Sun exposure and multiple sclerosis risk in Norway and Italy: The EnvIMS study. *Mult. Scler. J.* **20,** 1042–1049 (2014).

146. Cortese, M. *et al.* Timing of use of cod liver oil, a vitamin D source, and multiple sclerosis risk: The EnvIMS study. *Mult. Scler.* **21,** 1856–1864 (2015).

147. Bäärnhielm, M., Olsson, T. & Alfredsson, L. Fatty fish intake is associated with decreased occurrence of multiple sclerosis. *Mult. Scler. J.* **20,** 726–732 (2014).

148. Sandberg, L. *et al.* Vitamin D and axonal injury in multiple sclerosis. *Mult. Scler.* **22,** 1027–1031 (2016).

149. Fitzgerald, K. C. *et al.* Association of Vitamin D Levels With Multiple Sclerosis Activity and Progression in Patients Receiving Interferon Beta-1b. *JAMA Neurol.* **72,** 1458–1465 (2015).

150. Aranow, C. Vitamin D and the immune system. *J. Investig. Med.* **59,** 881–886 (2011).

151. Ascherio, A. & Munger, K. L. Environmental risk factors for multiple sclerosis. Part I: The role of infection. *Ann. Neurol.* **61,** 288–299 (2007).

152. Wingerchuk, D. M. Smoking: Effects on multiple sclerosis susceptibility and disease progression. *Ther. Adv. Neurol. Disord.* **5,** 13–22 (2012).

153. Sundström, P. & Nyström, L. Smoking worsens the prognosis in multiple sclerosis. *Mult. Scler.* **14,** 1031–1035 (2008).

154. Manouchehrinia, A. *et al.* Tobacco smoking and disability progression in multiple sclerosis: United Kingdom cohort study. *Brain* **136,** 2298–2304 (2013).

155. Hedström, A. K. *et al.* Smokers run increased risk of developing anti-natalizumab antibodies. *Mult. Scler. J.* **20,** 1081–1085 (2014).

156. Hedström, A. K. *et al.* Smoking and risk of treatment-induced neutralizing antibodies to interferon β-1a. *Mult. Scler. J.* **20,** 445–450 (2014).

157. Vessey, M. P., Villard-Mackintosh, L. & Yeates, D. Oral contraceptives, cigarette smoking and other factors in relation to arthritis. *Contraception* **35,** 457–464 (1987).

158. Nagata, C. *et al.* Systemic lupus erythematosus: a case-control epidemiologic study in Japan. *Int. J. Dermatol.* **34,** 333–337 (1995).

159. Shan, M. *et al.* Lung myeloid dendritic cells coordinately induce TH1 and TH17 responses in human emphysema. *Sci Transl Med* **1,** 4ra10 (2009).

160. Odoardi, F. *et al.* T cells become licensed in the lung to enter the central nervous system. *Nature* **488,** 675–679 (2012).

161. Hedström, A. K. *et al.* Smoking and two human leukocyte antigen genes interact to increase the risk for multiple sclerosis. *Brain* **134,** 653–664 (2011).

162. Hedström, A. K., Olsson, T. & Alfredsson, L. Body mass index during adolescence, rather than childhood, is critical in determining MS risk. *Mult. Scler.* **22,** 878–883 (2016).

163. Wesnes, K. *et al.* Body size and the risk of multiple sclerosis in Norway and Italy: The EnvIMS study. *Mult. Scler. J.* **21,** 388–395 (2015).

164. Matarese, G. *et al.* Leptin increase in multiple sclerosis associates with reduced number of CD4+CD25+ regulatory T cells. *Proc. Natl. Acad. Sci.* **102,** 5150–5155 (2005).

165. Wortsman, J., Matsuoka Y, L., Chen C, T., Zhiren, L. & Holick F, M. Decreased bioavailability of vitamin D in obesity. *Am. J. Clin. Nutr.* **72,** 690–693 (2000).

166. Hedström, A. K. *et al.* Interaction between adolescent obesity and HLA risk genes in the etiology of multiple sclerosis. *Neurology* **82,** 865–872 (2014).

167. Oksenberg, J. R., Baranzini, S. E., Sawcer, S. & Hauser, S. L. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nat. Rev. Genet.* **9,** 516–526 (2008).

168. Robertson, N. P. *et al.* Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain* **119,** 449–455 (1996).

169. Westerlind, H. *et al.* Modest familial risks for multiple sclerosis: A registry-based study of the population of Sweden. *Brain* **137,** 770–778 (2014).

170. Kuusisto, H. *et al.* Concordance and heritability of multiple sclerosis in Finland: Study on a nationwide series of twins. *Eur. J. Neurol.* **15,** 1106–1110 (2008).

171. Willer, C. J., Dyment, D. A., Risch, N. J., Sadovnick, A. D. & Ebers, G. C. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc. Natl. Acad. Sci.* **100,** 12877–12882 (2003).

172. Oksenberg, J. R. & Baranzini, S. E. Multiple sclerosis genetics--is the glass half full, or half empty? *Nat. Rev. Neurol.* **6,** 429–437 (2010).

173. Horton, R. *et al.* Gene map of the extended human MHC. *Nat Rev Genet* **5,** 889–899 (2004).

174. Horton, R. *et al.* Variation analysis and gene annotation of eight MHC haplotypes: The MHC Haplotype Project. *Immunogenetics* **60,** 1–18 (2008).

175. Patsopoulos, N. A. *et al.* Fine-Mapping the Genetic Association of the Major Histocompatibility Complex in Multiple Sclerosis: HLA and Non-HLA Effects. *PLoS Genet.* **9,** (2013).

176. Cotsapas, C. & Mitrovic, M. Genome-wide association studies of multiple sclerosis. **7,** 1–9 (2018).

177. Manolio, T. A. Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363,** 166–176 (2010).

178. Lander, E. The new genomics:global views of biology. *Science.* **274,** 536–539 (1996).

179. Patsopoulos, N. *et al.* The Multiple Sclerosis Genomic Map: Role of peripheral immune cells and resident microglia in susceptibility. *BioRxiv* (2017).

180. Jakkula, E. *et al.* Genome-wide Association Study in a High-Risk Isolate for Multiple Sclerosis Reveals Associated Variants in STAT3 Gene. *Am. J. Hum. Genet.* **86,** 285–291 (2010).

181. Barrett, J. C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40,** 955–962 (2008).

182. Maier, L. M. *et al.* IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet.* **5,** e1000322 (2009).

183. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science.* **337,** 1190–1195 (2012).

184. Farh, K. K. H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease

variants. *Nature* **518,** 337–343 (2015).

185. Gregory, S. G. *et al.* Interleukin 7 receptor α chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat. Genet.* **39,** 1083–1091 (2007).

186. Galarza-Muñoz, G. *et al.* Human Epistatic Interaction Controls IL7R Splicing and Increases Multiple Sclerosis Risk. *Cell* **169,** 72–84 (2017).

187. Gregory, A. P. *et al.* TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* **488,** 508–511 (2012).

188. Couturier, N. *et al.* Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility. *Brain* **134,** 693–703 (2011).

189. Didonna, A. *et al.* A non-synonymous single-nucleotide polymorphism associated with multiple sclerosis risk affects the EVI5 interactome. *Hum. Mol. Genet.* **24,** 7151–7158 (2015).

190. Alloza, I. *et al.* ANKRD55 and DHCR7 are novel multiple sclerosis risk loci. *Genes Immun.* **13,** 253–257 (2012).

191. Lill, C. M. *et al.* Genome-wide significant association of ANKRD55 rs6859219 and multiple sclerosis risk. *J Med Genet* **50,** 140–143 (2013).

192. Beecham, A. H. *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45,** 1353–1362 (2013).

193. Stahl, E. A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42,** 508–514 (2010).

194. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506,** 376–381 (2014).

195. Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* **44,** 1336–1340 (2012).

196. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491,** 119–124 (2012).

197. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47,** 979–986 (2015).

198. Fortune, M. D. *et al.* Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. *Nature Genetics* **47,** 839–846 (2015).

199. Hinks, A. *et al.* Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat. Genet.* **45,** 664–669 (2013).

200. Zhernakova, A. *et al.* Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.* **7,** e1002004 (2011).

201. Li, L. *et al.* Positive association between ankrd55 polymorphism 7731626 and dermatomyositis/polymyositis with interstitial lung disease in Chinese han population. *Biomed Res. Int.* **2017,** 2905987 (2017).

202. Houtman, M. *et al.* T-cell transcriptomics from peripheral blood highlights differences between polymyositis and dermatomyositis patients. 1–15 (2018).

203. Stein, M. B. *et al.* Genome-wide Association Studies of Posttraumatic Stress Disorder in 2 Cohorts of US Army Soldiers. *JAMA Psychiatry* **73,** 695–704 (2016).

204. Sherva, R. *et al.* Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimer's Dement.* **10,** 45–52 (2014).

205. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44,** 981–990 (2012).

206. Imamura, M. *et al.* Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nat. Commun.* **7,** 10531 (2016).

207. Jonsson, S. *et al.* Identification of sequence variants influencing immunoglobulin levels. *Nat. Genet.* **49,** 1182–1191 (2017).

208. Lauc, G. *et al.* Loci Associated with N-Glycosylation of Human Immunoglobulin G Show Pleiotropy with Autoimmune Diseases and Haematological Cancers. *PLoS Genet.* **9,** e1003225 (2013).

209. Mkhikian, H. *et al.* Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. *Nat. Commun.* **2,** 313–334 (2011).

210. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42,** 749–755 (2014).

211. Wu, C., Jin, X., Tsueng, G., Afrasiabi, C. & Su, A. I. BioGPS: Building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res.* **44,** D313–D316 (2016).

212. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–585 (2013).

213. Papatheodorou, I. *et al.* Expression Atlas: Gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* **46,** D246–D251 (2018).

214. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science.* **347,** 1260419–1260419 (2015).

215. Mosavi, L. K., Minor, D. L. & Peng, Z. Y. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16029–16034 (2002).

216. Mosavi, L. & Cammett, T. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13,** 1435–1448 (2004).

217. Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally? *Proteins Struct. Funct. Bioinforma.* **17,** 363–374 (1993).

218. Byeon, I. J. L. *et al.* Tumor suppressor p16INK4A: Determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4. *Mol. Cell* **1,** 421–431 (1998).

219. Tevelev, A. *et al.* Tumor suppressor p16INK4A: structural characterization of wild-type and mutant proteins by NMR and circular dichroism. *Biochemistry* **35,** 9475–9487 (1996).

220. Li, J., Mahajan, A. & Tsai, M. D. Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry* **45,** 15168–15178 (2006).

221. Main, E. R. G., Jackson, S. E. & Regan, L. The folding and design of repeat proteins: Reaching a consensus. *Curr. Opin. Struct. Biol.* **13,** 482–489 (2003).

222. Sedgwick, S. G. & Smerdon, S. J. The ankyrin repeat: A diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24,** 311–316 (1999).

223. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42,** 1001–1006 (2014).

224. Trynka, G. *et al.* Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci. *Am. J. Hum. Genet.*

97, 139–152 (2015).

225. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genet.* **6,** e1000888 (2010).

226. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45,** 124–130 (2013).

227. Corradin, O. & Scacheri, P. C. Enhancer variants: Evaluating functions in common disease. *Genome Med.* **6,** 1–14 (2014).

228. Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* **49,** 600–605 (2017).

229. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311–318 (2007).

230. Heintzman, N. D. *et al.* Histone modification at human enhancers reflect global cell-type specific gene expression. *Nature* **459,** 108–112 (2009).

231. Taylor, G., Eskeland, R. & Hekimoglu-balkan, B. H4K16 acetylation marks active genes and enhancers of embryonic stem cells , but does not alter chromatin compaction. *Genome Res.* **23,** 2053–2065 (2013).

232. Zentner, G. E. *et al.* Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* **21,** 1273–1283 (2011).

233. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–285 (2011).

234. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107,** 21931–21936 (2010).

235. Ong, C. T. & Corces, V. G. Enhancer function: New insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12,** 283–293 (2011).

236. Canver, M. C., Bauer, D. E. & Orkin, S. H. Functional interrogation of non-coding DNA through CRISPR genome editing. *Methods* **121,** 118–129 (2017).

237. Boyaval, P., Moineau, S., Romero, D. a & Horvath, P. Against Viruses in Prokaryotes. *Science.* **315,** 1709–1712 (2007).

238. Mali, P. *et al.* RNA-Guide Human Genome Engineering via Cas9. *Science.* **339,** 823–826 (2013).

239. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science.* **339,** 819–823 (2013).

240. Jiang, F. & Doudna, J. A. CRISPR – Cas9 Structures and Mechanisms. *Annu.Rev.Biophys* **46,** 505–529 (2017).

241. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8,** 2180–2196 (2013).

242. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183 (2013).

243. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: Repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17,** 5–15 (2016).

244. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167,** 233–247 (2016).

245. Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation.

Nucleic Acids Res. **44,** 5615–5628 (2016).

246. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–588 (2015).

247. Thakore, P. I., Black, J. B., Hilton, I. B. & Gersbach, C. A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods* **13,** 127–37 (2016).

248. Thakore, P. I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12,** 1143–1149 (2015).

249. Klann, T. S. *et al.* CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* **35,** 561–568 (2017).

250. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33,** 510–517 (2015).

251. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489,** 109–113 (2012).

252. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503,** 290–294 (2013).

253. Schoenfelder, S., Clay, I. & Fraser, P. The transcriptional interactome: Gene expression in 3D. *Curr. Opin. Genet. Dev.* **20,** 127–133 (2010).

254. Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nat Commun* **6,** 10069 (2015).

255. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25,** 582–597 (2015).

256. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **47,** 598–606 (2015).

257. Jäger, R. *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.* **6,** 1–9 (2015).

258. Dryden, N. H. *et al.* Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res.* **24,** 1854–1868 (2014).

259. Mcgovern, A. J. Functional Characterisation of Rheumatoid Arthritis Risk Loci. (2016).

260. Westra, H. J. *et al.* Fine-mapping and functional studies highlight potential causal variants for rheumatoid arthritis and type 1 diabetes. *Nat. Genet.* **50,** 1366–1374 (2018).

261. McAllister, K., Orozco, G., Worthington, J. & Eyre, S. Characterization of rheumatoid arthritis susceptibility locus, 5q11 (ANKRD55). in *Rheumatology* i48 (2015).

262. Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **44,** W272–W276 (2016).

263. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **32,** 941–946 (2014).

264. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72,** 8463–8471 (1998).

265. Lopez de Lapuente, A. *et al.* Novel Insights into the Multiple Sclerosis Risk Gene ANKRD55. *J. Immunol.* **196,** 4553–4565 (2016).

266. Boniface, K. *et al.* Oncostatin M secreted by skin infiltrating T lymphocytes is a potent

keratinocyte activator involved in skin inflammation. *J. Immunol.* **178,** 4615–4622 (2007).

267. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3,** 1101–1108 (2008).

268. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17,** 1–14 (2016).

269. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22,** 1790–1797 (2012).

270. Ward, L. D. & Kellis, M. HaploReg v4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44,** D877–D881 (2016).

271. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

272. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

273. Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T. & Flicek, P. R. The Ensembl Regulatory Build. *Genome Biol.* **16,** 1–8 (2015).

274. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol.* **30,** 224–226 (2012).

275. Romanoski, C. E., Glass, C. K., Stunnenberg, H. G., Wilson, L. & Almouzni, G. Epigenomics: Roadmap for regulation. *Nature* **518,** 314–316 (2015).

276. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Res.* **41,** 991–995 (2013).

277. Kleiveland, C. R. Peripheral Blood Mononuclear Cells. in *The Impact of Food Bioactives on Health: In Vitro and Ex Vivo Models* 161–167 (Springer International Publishing, 2015).

278. Hemmer, B., Kerschensteiner, M. & Korn, T. Role of the innate and adaptive immune responses in the course of multiple sclerosis. *The Lancet Neurology* **14,** 406–419 (2015).

279. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs. *Genome Res.* **22,** 1775–1789 (2012).

280. Clark, B. S. & Blackshaw, S. Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease. *Front. Genet.* **5,** 1–19 (2014).

281. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15,** 17–29 (2006).

282. Ning, S. *et al.* A global map for dissecting phenotypic variants in human lincRNAs. *Eur. J. Hum. Genet.* **21,** 1128–1133 (2013).

283. Almlöf, J. C. *et al.* Single nucleotide polymorphisms with Cis-regulatory effects on long non-coding transcripts in human primary monocytes. *PLoS One* **9,** 1–11 (2014).

284. Bernstein, B. E. *et al.* The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* **28,** 1045–1048 (2010).

285. James, T. *et al.* Impact of genetic risk loci for multiple sclerosis on expression of proximal genes in patients. *Hum. Mol. Genet.* **27,** 912–928 (2018).

286. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science.* **356,** pii: eaal3321 (2017).

287. Loo, L. H., Laksameethanasan, D. & Tung, Y. L. Quantitative Protein Localization Signatures Reveal an Association between Spatial and Functional Divergences of Proteins. *PLoS Comput. Biol.* **10,** e1003504 (2014).

288. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425,** 686–691 (2003).

289. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28,** 1248–1250 (2010).

290. Kumar, A. *et al.* Subcellular localization of the yeast proteome Subcellular localization of the yeast proteome. *Genes Dev.* **16,** 707–719 (2002).

291. Matsuyama, A. *et al.* ORFeome cloning and global analysis of protein localization in the fission yeast Schizosaccharomyces pombe. *Nat. Biotechnol.* **24,** 841–847 (2006).

292. Laurila, K. & Vihinen, M. Prediction of disease-related mutations affecting protein localization. *BMC Genomics* **10,** 122 (2009).

293. Park, S. *et al.* Protein localization as a principal feature of the etiology and comorbidity of genetic diseases. *Mol. Syst. Biol.* **7,** 1–11 (2011).

294. Luck, K., Sheynkman, G. M., Zhang, I. & Vidal, M. Proteome-Scale Human Interactomics. *Trends Biochem. Sci.* **42,** 342–354 (2017).

295. Vidal, M., Cusick, M. E. & Barabási, A. L. Interactome networks and human disease. *Cell* **144,** 986–998 (2011).

296. Smits, A. H. & Vermeulen, M. Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. *Trends Biotechnol.* **34,** 825–834 (2016).

297. Alberts, B. *et al.* Analyzing protein structure and function. in *Molecular Biology of the Cell* (Garland Science, 2002).

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

299. Carneiro, D. G., Clarke, T., Davies, C. C. & Bailey, D. Identifying novel protein interactions: Proteomic methods, optimisation approaches and data analysis pipelines. *Methods* **95,** 46–54 (2016).

300. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162,** 425–440 (2015).

301. Hein, M. Y. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163,** 712–723 (2015).

302. Malovannaya, A. *et al.* Analysis of the human endogenous coregulator complexome. *Cell* **145,** 787–799 (2011).

303. Ho, Y. *et al.* Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415,** 180–183 (2002).

304. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147 (2002).

305. Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell* **138,** 389–403 (2009).

306. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440,** 637–643 (2006).

307. Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–636 (2006).

308. Hubner, N. C. *et al.* Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J. Cell Biol.* **189,** 739–754 (2010).

309. Cristea, I. M., Williams, R., Chait, B. T. & Rout, M. P. Fluorescent Proteins as Proteomic

Probes. *Mol. Cell. Proteomics* **4,** 1933–1941 (2005).

310. Holden, P. & Horton, W. Crude subcellular fractionation of cultured mammalian cell lines. *BMC Res. Notes* **2,** 243 (2009).

311. Greenberg, M. E. & Bender, T. P. Identification of newly transcribed RNA. in *Current Protocols in Molecular Biology* **78,** 4.10.5–4.10.7 (2007).

312. Blein-Nicolas, M. & Zivy, M. Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochim. Biophys. Acta - Proteins Proteomics* **1864,** 883–895 (2016).

313. Zybailov, B. *et al.* Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J. Proteome Res.* **5,** 2339–2347 (2006).

314. Rigbolt, K. T. G. *et al.* System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci. Signal.* **4,** rs3 (2011).

315. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).

316. Hornbeck, P. V *et al.* PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43,** D512–D520 (2015).

317. Ullah, S. *et al.* DbPAF: An integrative database of protein phosphorylation in animals and fungi. *Sci. Rep.* **6,** 23534 (2016).

318. Madeira, F. *et al.* 14-3-3-Pred: Improved methods to predict 14-3-3-binding phosphopeptides. *Bioinformatics* **31,** 2276–2283 (2015).

319. Yan, J. & Kurgan, L. DRNApred, fast sequence-based method that accurately predicts and discriminates DNA-and RNA-binding residues. *Nucleic Acids Res.* **45,** 1–16 (2017).

320. Chen, K., Mizianty, M. J. & Kurgan, L. Prediction and analysis of nucleotide-binding residues using sequence and sequence-derived structural descriptors. *Bioinformatics* **28,** 331–341 (2012).

321. Jiang, J. Q. & Wu, M. Predicting multiplex subcellular localization of proteins using protein-protein interaction network: a comparative study. *BMC Bioinformatics* **13,** S20 (2012).

322. Lutz, M. I., Schwaiger, C., Hochreiter, B., Kovacs, G. G. & Schmid, J. A. Novel approach for accurate tissue-based protein colocalization and proximity microscopy. *Sci. Rep.* **7,** 2668 (2017).

323. Wilker, E. & Yaffe, M. B. 14-3-3 Proteins - A focus on cancer and human disease. *J. Mol. Cell. Cardiol.* **37,** 633–642 (2004).

324. Yaffe, M. B. Master of all things phosphorylated. *Biochem. J.* **379,** e1–e2 (2004).

325. Masters, S. C., Pederson, K. J., Zhang, L., Barbieri, J. T. & Fu, H. Interaction of 14-3-3 with a nonphosphorylated protein ligand, exoenzyme S of Pseudomonas aeruginosa. *Biochemistry* **38,** 5216–5221 (1999).

326. Wang, B. *et al.* Isolation of high-affinity peptide antagonists of 14-3-3 proteins by phage display. *Biochemistry* **38,** 12499–12504 (1999).

327. Hornbeck, P. V. *et al.* PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **40,** 261–270 (2012).

328. Song, C. *et al.* Systematic Analysis of Protein Phosphorylation Networks From Phosphoproteomic Data. *Mol. Cell. Proteomics* **11,** 1070–1083 (2012).

329. Zhou, H. *et al.* Toward a comprehensive characterization of a human cancer cell phosphoproteome. *J. Proteome Res.* **12,** 260–271 (2013).

330. Weintz, G. *et al.* The phosphoproteome of toll-like receptor-activated macrophages. *Mol. Syst. Biol.* **6,** 371 (2010).

331. Zanivan, S. *et al.* Solid Tumor Proteome and Phosphoproteome Analysis by High Resolution Mass Spectrometry. *J. Proteome Res.* **7,** 5314–5326 (2008).

332. Minard, A. Y. *et al.* mTORC1 Is a Major Regulatory Node in the FGF21 Signaling Network in Adipocytes. *Cell Rep.* **17,** 29–36 (2016).

333. Pinto, S. M. *et al.* Quantitative phosphoproteomic analysis of IL-33 mediated signaling. *Proteomics* **14,** 532–444 (2015).

334. Humphrey, S. J. *et al.* Dynamic adipocyte phosphoproteome reveals that akt directly regulates mTORC2. *Cell Metab.* **17,** 1009–1020 (2013).

335. Trost, M. *et al.* The Phagosomal Proteome in Interferon-γ-Activated Macrophages. *Immunity* **30,** 143–154 (2009).

336. Nevo-Dinur, K., Govindarajan, S. & Amster-Choder, O. Subcellular localization of RNA and proteins in prokaryotes. *Trends Genet.* **28,** 314–322 (2012).

337. Berggård, T., Linse, S. & James, P. Methods for the detection and analysis of protein-protein interactions. *Proteomics* **7,** 2833–2842 (2007).

338. Rao, V. S., Srinivas, K., Sujini, G. N. & Kumar, G. N. Protein-Protein Interaction Detection: Methods and Analysis. *Int J Proteomics* **2014,** 147648 (2014).

339. Sachdev, S., Hoffmann, A. & Hannink, M. Nuclear localization of IkappaB alpha is mediated by the second ankyrin repeat: the IkappaB alpha ankyrin repeats define a novel class of cis-acting nuclear import sequences. *Mol. Cell. Biol.* **18,** 2524–2534 (1998).

340. Lu, M. *et al.* A code for RanGDP binding in ankyrin repeats defines a nuclear import pathway. *Cell* **157,** 1130–1145 (2014).

341. Prosser, S. L. & Pelletier, L. Mitotic spindle assembly in animal cells: A fine balancing act. *Nat. Rev. Mol. Cell Biol.* **18,** 187–201 (2017).

342. Kollman, J. M., Merdes, A., Mourey, L. & Agard, D. A. Microtubule nucleation by γ-tubulin complexes. *Nat. Rev. Mol. Cell Biol.* **12,** 709–721 (2011).

343. Tanaka, T., Fuchs, J., Loidl, J. & Nasmyth, K. Cohesin ensures bipolar attachment of microtubules to sister centromeres and resists their precocious separation. *Nat. Cell Biol.* **2,** 492–499 (2000).

344. Jang, C. Y., Kim, H. D., Zhang, X., Chang, J. S. & Kim, J. Ribosomal protein S3 localizes on the mitotic spindle and functions as a microtubule associated protein in mitosis. *Biochem. Biophys. Res. Commun.* **429,** 57–62 (2012).

345. Hetzer, M., Gruss, O. J. & Mattaj, I. W. The Ran GTPase as a marker of chromosome position in spindle formation and nuclear envelope assembly. *Nat. Cell Biol.* **4,** E177-184 (2002).

346. Berg, D., Holzmann, C. & Riess, O. 14-3-3 proteins in the nervous system. *Nat. Rev. Neurosci.* **4,** 752–762 (2003).

347. Obsil, T. & Obsilova, V. Structural basis of 14-3-3 protein functions. *Semin. Cell Dev. Biol.* **22,** 663–672 (2011).

348. Colucci, M. *et al.* The 14-3-3 protein in multiple sclerosis: a marker of disease severity. *Mult Scler* **10,** 477–481 (2004).

349. Lee, D. H. *et al.* Role of glial 14-3-3 gamma protein in autoimmune demyelination. *J. Neuroinflammation* **12,** 187 (2015).

350. Zhou, X., Liao, W. J., Liao, J. M., Liao, P. & Lu, H. Ribosomal proteins: Functions beyond the ribosome. *J. Mol. Cell Biol.* **7,** 92–104 (2015).

351. De Keersmaecker, K. *et al.* Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **45,** 186–190 (2013).

352. Klauck, S. M. *et al.* Mutations in the ribosomal protein gene RPL10 suggest a novel modulating disease mechanism for autism. *Mol. Psychiatry* **11,** 1073–1084 (2006).

353. Jin, A., Itahana, K., Keefe, K. O. & Zhang, Y. Inhibition of HDM2 and Activation of p53 by Ribosomal Protein L23 Inhibition of HDM2 and Activation of p53 by Ribosomal Protein L23. *Society* **24,** 7669–7680 (2004).

354. Dai, M. S. *et al.* Ribosomal Protein L23 Activates p53 by Inhibiting MDM2 Function in Response to Ribosomal Perturbation but Not to Translation Inhibition. *Mol. Cell. Biol.* **24,** 7654–7668 (2004).

355. Iizumi, Y. *et al.* The Flavonoid Apigenin Downregulates CDK1 by Directly Targeting Ribosomal Protein S9. *PLoS One* **8,** 1–10 (2013).

356. Lim, K. H. *et al.* RPS3a over-expressed in HBV-associated hepatocellular carcinoma enhances the HBx-induced NF-κB signaling via its novel chaperoning function. *PLoS One* **6,** 1–17 (2011).

357. Artero-Castro, A. *et al.* Expression of the ribosomal proteins Rplp0, Rplp1, and Rplp2 in gynecologic tumors. *Hum. Pathol.* **42,** 194–203 (2011).

358. Green, L., Houck-Loomis, B., Yueh, A. & Goff, S. P. Large Ribosomal Protein 4 Increases Efficiency of Viral Recoding Sequences. *J. Virol.* **86,** 8949–8958 (2012).

359. Beyer, A. R. *et al.* Nucleolar Trafficking of the Mouse Mammary Tumor Virus Gag Protein Induced by Interaction with Ribosomal Protein L9. *J. Virol.* **87,** 1069–1082 (2013).

360. Lee, S. B. *et al.* Ribosomal protein S3, a new substrate of Akt, serves as a signal mediator between neuronal apoptosis and DNA repair. *J. Biol. Chem.* **285,** 29457–29468 (2010).

361. Wan, F. *et al.* Ribosomal Protein S3: A KH Domain Subunit in NF-κB Complexes that Mediates Selective Gene Regulation. *Cell* **131,** 927–939 (2007).

362. Yang, H. J. *et al.* Phosphorylation of ribosomal protein S3 and antiapoptotic TRAF2 protein mediates radioresistance in non-small cell lung cancer cells. *J. Biol. Chem.* **288,** 2965–2975 (2013).

363. Yadavilli, S. *et al.* Ribosomal protein S3: A multi-functional protein that interacts with both p53 and MDM2 through its KH domain. *DNA Repair (Amst).* **8,** 1215–1224 (2009).

364. Gao, X. & Hardwidge, P. R. Ribosomal protein S3: A multifunctional target of attaching/effacing bacterial pathogens. *Front. Microbiol.* **2,** 1–6 (2011).

365. Yates, B. *et al.* Genenames.org: The HGNC and VGNC resources in 2017. *Nucleic Acids Res.* **45,** D619–D625 (2017).

366. Hediger, M. A. *et al.* The ABCs of solute carriers: Physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflugers Arch. Eur. J. Physiol.* **447,** 465–468 (2004).

367. Palmieri, F. The mitochondrial transporter family SLC25: Identification, properties and physiopathology. *Mol. Aspects Med.* **34,** 465–484 (2013).

368. Palmieri, F. Mitochondrial transporters of the SLC25 family and associated diseases: A

review. *J. Inherit. Metab. Dis.* **37,** 565–575 (2014).

369. Wibom, R. *et al.* AGC1 Deficiency Associated with Global Cerebral Hypomyelination. *N. Engl. J. Med.* **361,** 489–495 (2009).

370. Falk, M. *et al.* AGC1 Deficiency Causes Infantile Epilepsy, Abnormal Myelination, and Reduced N-Acetylaspartate. *JIMD Rep.* **4,** 77–85 (2012).

371. Amoedo, N. D. *et al.* AGC1/2, the mitochondrial aspartate-glutamate carriers. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863,** 2394–2412 (2016).

372. Lee, N. Y., Kim, Y., Ryu, H. & Kang, Y. S. The alteration of serine transporter activity in a cell line model of amyotrophic lateral sclerosis (ALS). *Biochem. Biophys. Res. Commun.* **483,** 135–141 (2017).

373. Maucler, C., Pernot, P., Vasylieva, N., Pollegioni, L. & Marinesco, S. In vivo d-serine hetero-exchange through alanine-serine-cysteine (ASC) transporters detected by microelectrode biosensors. *ACS Chem. Neurosci.* **4,** 772–781 (2013).

374. Maekawa, M. *et al.* Investigation of the fatty acid transporter-encoding genes SLC27A3 and SLC27A4 in autism. *Sci. Rep.* **5,** 1–15 (2015).

375. Aparicio, T., Megías, D. & Méndez, J. Visualization of the MCM DNA helicase at replication factories before the onset of DNA synthesis. *Chromosoma* **121,** 499–507 (2012).

376. Das, M., Singh, S., Pradhan, S. & Narayan, G. MCM Paradox: Abundance of Eukaryotic Replicative Helicases and Genomic Integrity. *Mol. Biol. Int.* **2014,** 1–11 (2014).

377. Zheng, D. *et al.* Pre-RC protein MCM7 depletion promotes mitotic exit by inhibiting CDK1 activity. *Sci. Rep.* **7,** 1–10 (2017).

378. Chatr-Aryamontri, A. *et al.* The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* **43,** D470–D478 (2015).

379. Szklarczyk, D. *et al.* STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43,** D447–D452 (2015).

380. Huttlin, E. L. *et al.* Architecture of the human interactome defines protein communities and disease networks. *Nature* **545,** 505–509 (2017).

381. Cockman, M. E. *et al.* Posttranslational hydroxylation of ankyrin repeats in IkappaB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc Natl Acad Sci U S A* **103,** 14767–14772 (2006).

382. Fuhrer, T. Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol Syst Biol* **13,** 932 (2017).

383. Huang, Q. & Belz, G. T. Bach2: An Instrument of Heterogeneity for Long-Term Protection. *Immunity* **48,** 618–620 (2018).

384. Roychoudhuri, R. *et al.* BACH2 represses effector programs to stabilize T reg-mediated immune homeostasis. *Nature* **498,** 506–510 (2013).

385. Kaul-Ghanekar, R. *et al.* Abnormal V(D)J recombination of T cell receptor β locus in SMAR1 transgenic mice. *J. Biol. Chem.* **280,** 9450–9459 (2005).

386. Travis, A., Amsterdam, A., Belanger, C. & Grosschedl, R. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function. *Genes Dev* **5,** 880–894 (1991).

387. Yu, S. *et al.* The TCF-1 and LEF-1 Transcription Factors Have Cooperative and Opposing Roles in T Cell Development and Malignancy. *Immunity* **37,** 813–826 (2012).

388. Wang, H. *et al.* C-Type Lectin Receptors Differentially Induce Th17 Cells and Vaccine Immunity to the Endemic Mycosis of North America. *J. Immunol.* **192,** 1107–1119 (2014).

389. Patterson, S. J. *et al.* Cutting Edge: PHLPP Regulates the Development, Function, and Molecular Signaling Pathways of Regulatory T Cells. *J. Immunol.* **186,** 5533–5537 (2011).

390. Martin, E. *et al.* CTP synthase 1 deficiency in humans reveals its central role in lymphocyte proliferation. *Nature* **510,** 288–292 (2014).

391. Kim, S. & Jang, C. Y. ANKRD53 interacts with DDA3 and regulates chromosome integrity during mitosis. *Biochem. Biophys. Res. Commun.* **470,** 484–491 (2016).

SUPPLEMENTARY FILES

Supplementary Table 1. Specific primers for ANKRD55 expression vector cloning

Vector name	Primer	Sequence 5'-3'
pCMV6 005 FLAG/myc tag	005 SgfI Fw	ATTCGCGATCGCCATGGACAGCAACCTG
	005 MluI Rv	ACCACGCGTATTTTCATCACTGGTGGGGTTGGCAGA
pCMV6 001/005 FLAG/his tag	His pCMV6 Fw	TCGAGCATCATCACCACCACCATGCAGCAAATGAT
	His pCMV6 Rv	ATCATTTGCTGCATGGTGGTGGTGATGATGC

Supplementary Table 2. Identification of ANKRD55-interacting partners in a minimum of two replicates of nuclear and total protein extracts by nLC-MS/MS

	Gene symbol	Accession	MW (kDa)	Protein	NSAF	
					ANKRD55	CTRL
TOTAL PROTEIN EXTRACTS	TUBB	P07437	49.60	Tubulin beta chain	2.23	1.48
	TUBB4B	P68371	49.80	Tubulin beta-4B chain	1.92	1.35
	TUBA1B	P68363	50.152	Tubulin alpha-1B chain	1.60	0.00
	YWHAZ	P63104	27.745	14-3-3 protein zeta/delta	1.60	0.67
	YWHAQ	P27348	27.76	14-3-3 protein theta	1.32	0.67
	HSPA1A	P0DMV8	63.89	Isoform 2 of Heat shock 70 kDa protein 1A/1B	1.31	0.77
	TUBB2A	Q13885	49.907	Tubulin beta-2A chain	1.14	0.00
	EIF4A1	P60842	46.15	Eukaryotic initiation factor 4A-I	1.10	0.81
	HSPD1	P10809	61.055	60 kDa heat shock protein, mitochondrial	0.75	0.19
	RPS18	P62269	17.719	40S ribosomal protein S18	0.66	1.44
	SLC25A6	P12236	32.866	ADP/ATP translocase 3	0.61	0.00
	RUVBL1	Q9Y265	50.22	RuvB-like 1	0.55	0.24
	CCT2	P78371	57.488	T-complex protein 1 subunit beta	0.50	0.10
	ARF3	P61204	20.601	ADP-ribosylation factor 3	0.47	0.00
	PKM	P14618	57.937	Pyruvate kinase PKM	0.47	0.21
	TUFM	P49411	49.542	Elongation factor Tu, mitochondrial	0.47	0.24
	PPIA	P62937	18.012	Peptidyl-prolyl cis-trans isomerase A	0.46	0.33
	CCT6A	P40227	58.024	T-complex protein 1 subunit zeta	0.37	0.00
	PRDX1	Q06830	22.11	Peroxiredoxin-1	0.37	1.10
	RPS16	P62249	16.445	40S ribosomal protein S16	0.37	0.38
	UQCRC2	P22695	48.443	Cytochrome b-c1 complex subunit 2, mitochondrial	0.36	0.00
	RPS14	P62263	16.273	40S ribosomal protein S14	0.35	0.36
	RPS4X	P62701	29.598	40S ribosomal protein S4, X isoform	0.31	0.21
	ATAD3A	Q9NVI7	71.369	ATPase family AAA domain-containing protein 3A	0.29	0.17
	CCT5	P48643	59.671	T-complex protein 1 subunit epsilon	0.29	0.20
	DDX39A	O00148	49.13	ATP-dependent RNA helicase DDX39A	0.28	0.00
	ALDH1B1	P30837	57.206	Aldehyde dehydrogenase X, mitochondrial	0.27	0.00
	DARS	P14868	57.136	Aspartate--tRNA ligase, cytoplasmic	0.26	0.00
	HSP90AB1	P08238	83.264	Heat shock protein HSP 90-beta	0.26	0.53
	CCT7	Q99832	59.367	T-complex protein 1 subunit eta	0.26	0.10
	DHX9	Q08211	140.95	ATP-dependent RNA helicase A	0.26	0.09
	ILF2	Q12905	43.06	Interleukin enhancer-binding factor 2	0.24	0.00
	RPS20	P60866	13.373	40S ribosomal protein S20	0.24	0.00

	RPS11	P62280	18.431	40S ribosomal protein S11	0.22	0.00
	EEF2	P13639	95.338	Elongation factor 2	0.22	0.06
	RPN2	P04844	69.284	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	0.21	0.00
	DDX17	Q92841	80.272	Probable ATP-dependent RNA helicase DDX17	0.21	0.00
	CCT3	P49368	60.534	T-complex protein 1 subunit gamma	0.20	0.00
	NDUFA13	Q9P0J0	16.698	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	0.19	0.00
	PRPS1	P60891	34.834	Ribose-phosphate pyrophosphokinase 1	0.19	0.00
	SLC25A10	Q9UBX3	31.282	Mitochondrial dicarboxylate carrier	0.18	0.00
	MTHFD1	P11586	101.559	C-1-tetrahydrofolate synthase, cytoplasmic	0.18	0.23
	DDOST	P39656	50.801	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	0.17	0.00
	RPL11	P62913	20.12	60S ribosomal protein L11	0.17	0.00
	HNRNPH1	P31943	49.229	Heterogeneous nuclear ribonucleoprotein H	0.16	0.00
	CDIPT	O14735	23.539	CDP-diacylglycerol--inositol 3-phosphatidyltransferase	0.15	0.00
	PABPC1	P11940	70.67	Polyadenylate-binding protein 1	0.15	0.00
	OTUB1	Q96FW1	31.284	Ubiquitin thioesterase OTUB1	0.14	0.00
	MCM5	P33992	82.286	DNA replication licensing factor MCM5	0.14	0.00
	SRSF3	P84103	19.33	Serine/arginine-rich splicing factor 3	0.13	0.00
	EPRS	P07814	170.591	Bifunctional glutamate/proline--tRNA ligase	0.13	0.14
	TARDBP	Q13148	44.74	TAR DNA-binding protein 43	0.13	0.00
	ANXA2P2	A6NMY6	38.659	Putative annexin A2-like protein	0.13	0.00
	TIMM23	O14925	21.943	Mitochondrial import inner membrane translocase subunit Tim23	0.13	0.00
	FAF2	Q96CS3	52.623	FAS-associated factor 2	0.13	0.00
	PSMD2	Q13200		26S proteasome non-ATPase regulatory subunit 2	0.12	0.06
	DIMT1	Q9UNQ2	35.236	Probable dimethyladenosine transferase	0.12	0.00
	PCMT1	P22061	24.636	Protein-L-isoaspartate(D-aspartate) O-methyltransferase	0.12	0.00
	RQCD1	Q92600	33.631	Cell differentiation protein RCD1 homolog	0.12	0.00
	KPNB1	Q14974	97.17	Importin subunit beta-1	0.12	0.06
	YBX1	P67809	35.924	Nuclease-sensitive element-binding protein 1	0.12	0.00
	SURF4	O15260	30.394	Surfeit locus protein 4	0.12	0.00
	PSMC4	P43686	47.366	26S protease regulatory subunit 6B	0.12	0.13
	TUBG1	P23258	51.17	Tubulin gamma-1 chain	0.12	0.00
	SRM	P19623	102.335	Spermidine synthase	0.11	0.00
	STUB1	Q9UNE7	34.856	E3 ubiquitin-protein ligase CHIP	0.11	0.00
	ACOT8	O14734	35.914	Acyl-coenzyme A thioesterase 8	0.11	0.00
	PSMC6	P62333	44.173	26S protease regulatory subunit 10B	0.10	0.00
	HNRNPU	Q00839	90.584	Heterogeneous nuclear ribonucleoprotein U	0.10	0.00
	PSMD3	O43242	60.978	26S proteasome non-ATPase regulatory subunit 3	0.10	0.00
	HAX1	O00165	31.621	HCLS1-associated protein X-1	0.09	0.00
	PSMD12	O00232	52.904	26S proteasome non-ATPase regulatory subunit 12	0.09	0.00
	SEC61A1	P61619	52.265	Protein transport protein Sec61 subunit alpha isoform 1	0.09	0.00
	VDAC3	Q9Y277	30.659	Voltage-dependent anion-selective channel protein 3	0.09	0.00

	MSH2	P43246	104.743	DNA mismatch repair protein Msh2	0.09	0.00
	SLC25A18	Q9H1K4	33.849	Mitochondrial glutamate carrier 2	0.09	0.00
	HNRNPA0	Q13151	30.841	Heterogeneous nuclear ribonucleoprotein A0	0.09	0.00
	SMC3	Q9UQE7	141.54	Structural maintenance of chromosomes protein 3	0.09	0.00
	COPB1	P53618	107.142	Coatomer subunit beta	0.09	0.00
	DHRS7B	Q6IAN0	35.119	Dehydrogenase/reductase SDR family member 7B	0.09	0.00
	RPL8	P62917	28.025	60S ribosomal protein L8	0.08	0.00
	AASDHPPT	Q9NRN7	35.776	L-aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase	0.08	0.00
	SCO2	O43819	29.81	Protein SCO2 homolog, mitochondrial	0.08	0.00
	TRIM28	Q13263	88.55	Transcription intermediary factor 1-beta	0.08	0.07
	GALK1	P51570	42.272	Galactokinase	0.08	0.00
	MATR3	P43243	94.623	Matrin-3	0.08	0.00
	IFT52	Q9Y366	49.706	Intraflagellar transport protein 52 homolog	0.08	0.00
	TIMMDC1	Q9NPL8	32.178	Complex I assembly factor TIMMDC1, mitochondrial	0.08	0.00
	RFC3	P40938	40.556	Replication factor C subunit 3	0.07	0.00
	EIF3L	Q9Y262	66.727	Eukaryotic translation initiation factor 3 subunit L	0.07	0.00
	CDC45	O75419	65.569	Cell division control protein 45 homolog	0.07	0.00
	WDR36	Q8NI36	105.322	WD repeat-containing protein 36	0.07	0.00
	LBR	Q14739	70.703	Lamin-B receptor	0.07	0.00
	COPA	P53621	138.346	Coatomer subunit alpha	0.07	0.00
	STT3A	P46977	80.53	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	0.07	0.00
	AGK	Q53H12	47.137	Acylglycerol kinase, mitochondrial	0.07	0.00
	DHX15	O43143	90.933	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	0.07	0.00
	KHSRP	Q92945	73.115	Far upstream element-binding protein 2	0.06	0.00
	EI24	O14681	38.965	Etoposide-induced protein 2.4 homolog	0.06	0.00
	UMPS	P11172	52.222	Uridine 5'-monophosphate synthase	0.06	0.00
	GARS	P41250	83.166	Glycine--tRNA ligase	0.06	0.00
	POLD1	P28340	123.631	DNA polymerase delta catalytic subunit	0.06	0.00
	FAR1	Q8WVX9	59.357	Fatty acyl-CoA reductase 1	0.06	0.00
	MCM4	P33991	96.558	DNA replication licensing factor MCM4	0.06	0.00
	DDX1	Q92499	82.432	ATP-dependent RNA helicase DDX1	0.06	0.00
	UNC45A	Q9H3U1	103.077	Protein unc-45 homolog A	0.06	0.00
	COPB2	P35606	102.487	Coatomer subunit beta'	0.06	0.00
	HSPH1	Q92598	96.865	Heat shock protein 105 kDa	0.05	0.00
	SARS	P49591	58.777	Serine--tRNA ligase, cytoplasmic	0.05	0.00
	TELO2	Q9Y4R8	91.747	Telomere length regulation protein TEL2 homolog	0.05	0.00
	TRAP1	Q12931	80.11	Heat shock protein 75 kDa, mitochondrial	0.05	0.00
	NSUN2	Q08J23	86.471	tRNA (cytosine(34)-C(5))-methyltransferase	0.05	0.00
	PTPN1	P18031	49.967	Tyrosine-protein phosphatase non-receptor type 1	0.05	0.00
	NCLN	Q969V3	62.974	Nicalin	0.05	0.00
	ASNS	P08243	64.37	Asparagine synthetase [glutamine-hydrolyzing]	0.05	0.00
	MSH6	P52701	152.786	DNA mismatch repair protein Msh6	0.04	0.00
	HSP90B1	P14625	92.469	Endoplasmin	0.04	0.00

	CCDC8	Q9H0W5	59.374	Coiled-coil domain-containing protein 8	0.04	0.00
	TNPO1	Q92973	102.355	Transportin-1	0.04	0.00
	POLR2B	P30876	133.897	DNA-directed RNA polymerase II subunit RPB2	0.04	0.00
	IQGAP1	P46940	189.252	Ras GTPase-activating-like protein IQGAP1	0.04	0.00
	DIS3	Q9Y2L1	109.003	Exosome complex exonuclease RRP44	0.04	0.00
	AARS2	Q5JTZ9	107.34	Alanine--tRNA ligase, mitochondrial	0.04	0.00
	IPO9	Q96P70	115.963	Importin-9	0.03	0.00
	TFRC	P02786	84.871	Transferrin receptor protein 1	0.03	0.00
	IARS2	Q9NSE4	113.792	Isoleucine--tRNA ligase, mitochondrial	0.03	0.00
	DHX37	Q8IY37	129.545	Probable ATP-dependent RNA helicase DHX37	0.03	0.00
	RPTOR	Q8N122	149.038	Regulatory-associated protein of mTOR	0.03	0.00
	ESYT2	A0FGR8	102.357	Extended synaptotagmin-2	0.03	0.00
	SMC4	Q9NTJ3	147.182	Structural maintenance of chromosomes protein 4	0.03	0.00
	IPO7	O95373	119.517	Importin-7	0.03	0.00
	PPP6R3	Q5H9R7	97.669	Serine/threonine-protein phosphatase 6 regulatory subunit 3	0.02	0.00
	WDR11	Q9BZH6	136.685	WD repeat-containing protein 11	0.02	0.00
	USO1	O60763	107.895	General vesicular transport factor p115	0.02	0.00
	MYO1B	O43795	131.985	Unconventional myosin-Ib	0.02	0.00
	RFC1	P35251	128.255	Replication factor C subunit 1	0.02	0.00
	DHX30	Q7L2E3	133.938	Putative ATP-dependent RNA helicase DHX30	0.02	0.00
	TARBP1	Q13395	181.675	Probable methyltransferase TARBP1	0.02	0.00
	HEATR1	Q9H583	242.37	HEAT repeat-containing protein 1	0.02	0.00
	TBC1D4	O60343	146.563	TBC1 domain family member 4	0.02	0.00
	BTAF1	O14981	206.887	TATA-binding protein-associated factor 172	0.02	0.00
	HUWE1	Q7Z6Z7	481.891	E3 ubiquitin-protein ligase HUWE1	0.01	0.00
	ANAPC1	Q9H1A4	216.5	Anaphase-promoting complex subunit 1	0.01	0.00
	NUP188	Q5SRE5	196.043	Nucleoporin NUP188 homolog	0.01	0.00
NUCLEAR EXTRACTS	HSPA1A	P0DMV8	63.89	Isoform 2 of Heat shock 70 kDa protein 1A/1B	1.59	0.29
	EIF4A1	P60842	46.15	Eukaryotic initiation factor 4A-I	0.89	0.16
	DHX9	Q08211	140.95	ATP-dependent RNA helicase A	0.82	0.24
	NONO	Q15233	54.232	Non-POU domain-containing octamer-binding protein	0.73	0.26
	HSPA8	P11142	70.89	Heat shock cognate 71 kDa protein	0.70	0.14
	TUBB	P07437	49.60	Tubulin beta chain	0.62	0.00
	RPL22	P35268	14.787	60S ribosomal protein L22	0.60	0.40
	RUVBL2	Q9Y230	51.157	RuvB-like 2	0.59	0.05
	DDX47	Q9H0S4	50.647	Probable ATP-dependent RNA helicase DDX47	0.36	0.00
	MATR3	P43243	94.623	Matrin-3	0.36	0.17
	HNRNPH3	P31942	36.926	Heterogeneous nuclear ribonucleoprotein H3	0.33	0.15
	EIF4A3	P38919	46.871	Eukaryotic initiation factor 4A-III	0.32	0.00
	C1QBP	Q07021	31.362	Complement component 1 Q subcomponent-binding protein, mitochondrial	0.32	0.00
	MIB1	Q86YT6	110.136	E3 ubiquitin-protein ligase MIB1	0.30	0.00
	ATAD3B	Q5T9A4	72.573	ATPase family AAA domain-containing protein 3B	0.27	0.00
	LBR	Q14739	70.703	Lamin-B receptor	0.21	0.00
	SF3B3	Q15393	135.577	Splicing factor 3B subunit 3	0.19	0.02
	MTHFD1L	P11586	105.79	Monofunctional C1-tetrahydrofolate synthase, mitochondrial	0.18	0.00

	RPL3	P39023	46.109	60S ribosomal protein L3	0.18	0.00
	TCP1	P17987	60.344	T-complex protein 1 subunit alpha	0.17	0.00
	RCC1	P18754	44.969	Regulator of chromosome condensation	0.14	0.00
	PRPF31	Q8WWY3	55.456	U4/U6 small nuclear ribonucleoprotein Prp31	0.14	0.00
	LAS1L	Q9Y4W2	83.065	Ribosomal biogenesis protein LAS1L	0.12	0.00
	SNRNP200	O75643	244.508	U5 small nuclear ribonucleoprotein 200 kDa helicase	0.12	0.00
	NAT10	Q9H0A0	115.73	RNA cytidine acetyltransferase	0.10	0.00
	AMOT	Q4VCS5	118.09	Angiomotin	0.09	0.00
	MYO1C	O00159	121.682	Unconventional myosin-Ic	0.09	0.00
	EFTUD2	Q15029	109.436	116 kDa U5 small nuclear ribonucleoprotein component	0.08	0.00
	DSG2	Q14126	122.294	Desmoglein-2	0.07	0.00
	SMARCA5	O60264	121.905	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	0.06	0.00
	PRPF8	Q6P2Q9	273.6	Pre-mRNA-processing-splicing factor 8	0.05	0.00
	DHX30	Q7L2E3	133.938	Putative ATP-dependent RNA helicase DHX30	0.05	0.00
	NUMA1	Q14980	238.26	Nuclear mitotic apparatus protein 1	0.05	0.00
	MDN1	Q9NU22	632.82	Midasin	0.04	0.00

Supplementary Table 3. Functional enrichment analysis of ANKRD55 interactome from nuclear and total protein extracts. The algorithm DAVID was used to analyze the ANKRD55 interactome. The GO and other terms with p value<0.05 are shown and grouped by enrichment score.

TOTAL PROTEIN EXTRACTS

Annotation Cluster 1 - Enrichment Score: 12.73

Category	Term	Count	%	PValue	Genes	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KEYWORDS	Nucleotide-binding	52	34,90	1,26E-18	P17812, Q9Y5M8, O43615, O00571, Q5T9A4, Q15029, P10398, P18085, Q9Y3I0, P53396, Q00535, P05023, Q9BUF5, Q14566, Q9Y285, Q6P1M0, Q5T160, P49588, P41252, P50570, P28288, P38646, Q8NI60, P42356, Q9NU22, Q9P2J5, Q9NSD9, Q14683, P56192, Q9BSD7, P45880, Q6UB35, P54136, P16615, P78527, P27708, P40616, P17980, P50990, P50991, P38919, Q92900, Q9NR30, O75643, P54886, P51148, P25205, P33993, P26640, Q13085, Q14204, Q15645	4,02	2,93E-16	9,75E-17	1,61E-15
UP_KEYWORDS	ATP-binding	43	28,86	5,74E-16	P17812, P56192, O43615, Q9BSD7, Q5T9A4, O00571, P10398, Q6UB35, P54136, P16615, P78527, P27708, Q9Y3I0, P17980, P53396, Q00535, P50990, P50991, P38919, P05023, Q92900, Q14566, Q9Y285, Q9NR30, Q5T160, P49588, O75643, P41252, P54886, P28288, P38646, P25205, P33993, P26640, Q13085, Q8NI60, P42356, Q14204, Q14683, Q9NSD9, 9P2J5, Q9NU22, Q15645	4,27	1,29E-13	3,23E-14	7,11E-13

GOTERM_MF_DIRECT	GO:0005524~ATP binding	45	30,20	1,18E-13	P17812, P31689, P56192, O43615, Q9BSD7, Q5T9A4, O00571, P10398, Q6UB35, P54136, P16615, P78527, P27708, Q9Y3I0, P17980, P53396, Q00535, P50990, P50991, P38919, P05023, Q92900, Q14566, Q9Y285, Q9NR30, Q5T160, P49588, O75643, P41252, O60884, P54886, P28288, P38646, P25205, P33993, P26640, Q13085, Q8NI60, P42356, Q14204, Q14683, Q9NSD9, Q9P2J5, Q9NU22, Q15645	3,48	4,39E-11	2,19E-11	1,62E-10
UP_SEQ_FEATURE	nucleotide phosphate-binding region:ATP	22	14,77	1,33E-05	Q9NR30, O43615, O75643, Q9BSD7, O00571, Q5T9A4, P28288, P10398, Q6UB35, P25205, P33993, Q13085, Q14204, Q14683, Q9NU22, P17980, Q00535, Q15645, P53396, P38919, Q92900, Q14566	2,98	7,91E-03	1,13E-03	1,96E-02
Annotation Cluster 2 - Enrichment Score: 8.13										
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**	
GOTERM_BP_DIRECT	GO:0006412~translation	20	13,42	6,31E-13	P15880, P36578, P53007, P05388, Q92616, P27635, Q15029, P32969, P05141, Q00325, P62424, O75746, Q9NSD9, P61247, P62829, Q02978, P62244, P23396, Q9UJS0, P46781	9,09	6,09E-10	6,09E-10	9,93E-10	
GOTERM_BP_DIRECT	GO:0000184~nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	15	10,07	1,85E-12	P15880, P36578, P63151, P05388, P30153, P27635, P32969, P62424, P61247, P62829, P38919, P62244, P23396, Q92900, P46781	14,50	1,78E-09	8,90E-10	2,91E-09	

GOTERM_CC_DIRECT	GO:0005925~focal adhesion	21	14,09	5,15E-11	P15880, P36578, Q9Y490, P35613, P05388, P50570, P21333, P31946, P23528, P61981, P38646, P32969, P42356, P62258, P62424, P61247, P62829, Q00610, P23396, Q15366, P46781	6,61	1,31E-08	2,61E-09	6,70E-08
GOTERM_MF_DIRECT	GO:0003735~structural constituent of ribosome	17	11,41	8,05E-11	P15880, P36578, P53007, P05388, P27635, P32969, P05141, Q00325, P62424, O75746, P61247, P62829, Q02978, P62244, P23396, Q9UJS0, P46781	8,85	3,00E-08	7,51E-09	1,11E-07
GOTERM_BP_DIRECT	GO:0019083~viral transcription	13	8,72	2,29E-10	Q12769, P15880, P36578, Q92621, P62424, P05388, P61247, P62829, P27635, P62244, P23396, P46781, P32969	13,35	2,21E-07	7,36E-08	3,60E-07
UP_KEYWORDS	Ribonucleoprotein	17	11,41	4,44E-10	P15880, P36578, O75643, P05388, P27635, Q15029, O00567, P32969, P62424, P61247, P62829, Q6P2Q9, P62244, P23396, Q15366, Q15365, P46781	7,93	1,04E-07	1,73E-08	5,71E-07
GOTERM_BP_DIRECT	GO:0006614~SRP-dependent cotranslational protein targeting to membrane	12	8,05	5,08E-10	Q9Y5M8, P15880, P36578, P62424, P05388, P61247, P62829, P27635, P62244, P23396, P46781, P32969	14,68	4,90E-07	1,22E-07	7,99E-07
GOTERM_BP_DIRECT	GO:0006364~rRNA processing	15	10,07	4,88E-09	P15880, Q9NR30, P36578, P05388, P27635, O00567, P32969, P62424, Q9NU22, P61247, P62829, P38919, P62244, P23396, P46781	8,06	4,71E-06	7,85E-07	7,68E-06
GOTERM_BP_DIRECT	GO:0006413~translational initiation	12	8,05	2,84E-08	P15880, P36578, P62424, P05388, P61247, Q14152, P62829, P27635, P62244, P23396, P46781, P32969	10,07	2,74E-05	3,91E-06	4,47E-05

GOTERM_CC_DIRECT	GO:0005840~ribosome	12	8,05	1,05E-07	P15880, P36578, P62424, P05388, Q92616, P61247, P62829, P27635, P62244, P23396, P46781, P32969	8,90	2,66E-05	2,05E-06	1,36E-04
UP_KEYWORDS	Ribosomal protein	11	7,38	9,41E-07	P15880, P36578, P62424, P05388, P61247, P62829, P27635, P62244, P23396, P46781, P32969	8,21	2,19E-04	1,57E-05	1,21E-03
KEGG_PATHWAY	hsa03010:Ribosome	11	7,38	4,00E-05	P15880, P36578, P62424, P05388, P61247, P62829, P27635, P62244, P23396, P46781, P32969	5,22	6,06E-03	3,04E-03	4,79E-02
GOTERM_CC_DIRECT	GO:0022627~cytosolic small ribosomal subunit	6	4,03	4,13E-05	P15880, P61247, O00571, P62244, P23396, P46781	15,39	1,04E-02	6,56E-04	5,38E-02
GOTERM_CC_DIRECT	GO:0022625~cytosolic large ribosomal subunit	6	4,03	2,21E-04	P36578, P62424, P05388, P62829, P27635, P32969	10,86	5,46E-02	3,11E-03	2,87E-01
Annotation Cluster 3 - Enrichment Score: 6.987									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_MF_DIRECT	GO:0098641~cadherin binding involved in cell-cell adhesion	16	10,74	3,00E-08	P15880, Q9Y490, P35613, Q92616, O00571, P49327, P21333, P31946, O00567, P54136, P42356, P62258, P62424, P63244, P50990, Q15365	6,38	1,12E-05	1,87E-06	4,14E-05
GOTERM_CC_DIRECT	GO:0005913~cell-cell adherens junction	16	10,74	5,55E-08	P15880, Q9Y490, P35613, Q92616, O00571, P49327, P21333, P31946, O00567, P54136, P42356, P62258, P62424, P63244, P50990, Q15365	6,10	1,41E-05	1,28E-06	7,23E-05
GOTERM_BP_DIRECT	GO:0098609~cell-cell adhesion	14	9,40	6,56E-07	P15880, P35613, Q92616, O00571, P49327, P31946, O00567, P54136, P42356, P62258, P62424, P63244, P50990, Q15365	5,94	6,33E-04	7,91E-05	1,03E-03

Annotation Cluster 4 - Enrichment Score: 6.97									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	Mitochondrion	28	18,79	2,69E-08	P31689, P53007, O43615, Q3ZCQ8, P45880, Q07021, O00571, Q5T9A4, Q6UB35, O95831, P05141, O75746, P13804, Q02978, P36542, P23396, Q9UJS0, Q96HS1, Q5T160, P54886, P04181, P38646, P11177, Q8NI60, P42704, Q9H9B4, P48047, Q00325	3,46	6,26E-06	8,95E-07	3,45E-05
UP_KEYWORDS	Mitochondrion inner membrane	14	9,40	8,07E-08	P53007, O43615, Q3ZCQ8, Q5T9A4, P54886, O95831, P05141, P48047, Q00325, O75746, Q02978, P36542, P23396, Q9UJS0	7,16	1,88E-05	1,88E-06	1,04E-04
GOTERM_CC_DIRECT	GO:0005743~mitochondrial inner membrane	17	11,41	5,65E-07	P53007, O43615, Q3ZCQ8, P45880, Q5T9A4, P28288, P54886, O95831, P05141, Q9H9B4, P48047, Q00325, O75746, Q02978, P36542, P23396, Q9UJS0	4,75	1,43E-04	1,02E-05	7,36E-04
Annotation Cluster 5 - Enrichment Score: 5.30									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
INTERPRO	IPR016024:Armadillo-type fold	19	12,75	3,28E-10	Q96T76, P55060, Q92616, Q8TEX9, P30153, Q9UBB4, Q93008, Q86VP6, P78527, P42695, Q9Y678, P42356, O14980, P52292, Q9BQG0, P50851, Q00610, O43592, O00410	6,83	1,57E-07	1,57E-07	4,70E-07
UP_SEQ_FEATURE	repeat:HEAT 2	10	6,71	3,96E-10	Q96T76, P42695, Q9Y678, O14980, Q92616, Q8TEX9, P30153, Q86VP6, P78527, O00410	23,22	2,37E-07	2,37E-07	5,85E-07

UP_SEQ_FEATURE	repeat:HEAT 1	10	6,71	3,96E-10	Q96T76, P42695, Q9Y678, O14980, Q92616, Q8TEX9, P30153, Q86VP6, P78527, O00410	23,22	2,37E-07	2,37E-07	5,85E-07
UP_SEQ_FEATURE	repeat:HEAT 4	9	6,04	8,45E-10	Q96T76, P42695, Q9Y678, O14980, Q92616, Q8TEX9, P30153, Q86VP6, O00410	28,18	5,06E-07	2,53E-07	1,25E-06
UP_SEQ_FEATURE	repeat:HEAT 3	9	6,04	2,99E-09	Q96T76, P42695, Q9Y678, O14980, Q92616, Q8TEX9, P30153, Q86VP6, O00410	24,24	1,79E-06	5,98E-07	4,42E-06
INTERPRO	IPR011989:Armadillo-like helical	14	9,40	2,65E-08	Q96T76, P55060, Q92616, Q8TEX9, P30153, Q9UBB4, Q86VP6, P78527, P42695, Q9Y678, O14980, P52292, O43592, O00410	7,86	1,27E-05	6,33E-06	3,79E-05
UP_SEQ_FEATURE	repeat:HEAT 6	7	4,70	7,47E-08	Q96T76, O14980, Q92616, Q8TEX9, P30153, Q86VP6, O00410	31,42	4,47E-05	1,12E-05	1,10E-04
UP_SEQ_FEATURE	repeat:HEAT 5	7	4,70	1,98E-07	Q96T76, O14980, Q92616, Q8TEX9, P30153, Q86VP6, O00410	26,93	1,19E-04	1,98E-05	2,92E-04
UP_SEQ_FEATURE	repeat:HEAT 7	5	3,36	2,69E-05	Q96T76, O14980, Q92616, P30153, Q86VP6	28,05	1,60E-02	1,61E-03	3,98E-02
UP_SEQ_FEATURE	repeat:HEAT 9	4	2,68	1,68E-04	O14980, Q92616, P30153, Q86VP6	35,91	9,56E-02	7,15E-03	2,47E-01
UP_SEQ_FEATURE	repeat:HEAT 10	4	2,68	1,68E-04	O14980, Q92616, P30153, Q86VP6	35,91	9,56E-02	7,15E-03	2,47E-01
UP_SEQ_FEATURE	repeat:HEAT 8	4	2,68	3,50E-04	O14980, Q92616, P30153, Q86VP6	28,35	1,89E-01	1,30E-02	5,15E-01
UP_SEQ_FEATURE	repeat:HEAT 15	3	2,01	1,11E-03	Q92616, P30153, Q86VP6	57,71	4,85E-01	3,83E-02	1,62E+00
UP_SEQ_FEATURE	repeat:HEAT 14	3	2,01	1,88E-03	Q92616, P30153, Q86VP6	44,88	6,76E-01	5,76E-02	2,74E+00
UP_SEQ_FEATURE	repeat:HEAT 12	3	2,01	2,34E-03	Q92616, P30153, Q86VP6	40,40	7,54E-01	6,78E-02	3,40E+00
UP_SEQ_FEATURE	repeat:HEAT 13	3	2,01	2,34E-03	Q92616, P30153, Q86VP6	40,40	7,54E-01	6,78E-02	3,40E+00
UP_SEQ_FEATURE	repeat:HEAT 11	3	2,01	2,85E-03	Q92616, P30153, Q86VP6	36,72	8,19E-01	7,81E-02	4,12E+00
INTERPRO	IPR000357:HEAT	3	2,01	4,53E-03	Q8TEX9, P30153, O00410	29,14	8,86E-01	1,44E-01	6,29E+00
INTERPRO	IPR021133:HEAT, type 2	3	2,01	9,63E-03	Q92616, Q8TEX9, P30153	19,93	9,90E-01	2,65E-01	1,29E+01

Annotation Cluster 6 - Enrichment Score: 5.25									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
INTERPRO	IPR027417:P-loop containing nucleoside triphosphate hydrolase	24	16,11	3,77E-07	P17812, Q9Y5M8, Q9NR30, O75643, Q9BSD7, P50570, O00571, Q5T9A4, Q15029, P28288, Q6UB35, P51148, P25205, P33993, P18085, Q14204, Q14683, Q9NU22, P40616, P17980, Q15645, P38919, Q92900, Q14566	3,45	1,80E-04	3,60E-05	5,39E-04
INTERPRO	IPR003593:AAA+ ATPase domain	9	6,04	1,22E-05	P33993, Q14204, Q9BSD7, Q9NU22, P17980, Q15645, Q5T9A4, P28288, P25205	8,35	5,81E-03	7,28E-04	1,74E-02
UP_SEQ_FEATURE	nucleotide phosphate-binding region:ATP	22	14,77	1,33E-05	Q9NR30, O43615, O75643, Q9BSD7, O00571, Q5T9A4, P28288, P10398, Q6UB35, P25205, P33993, Q13085, Q14204, Q14683, Q9NU22, P17980, Q00535, Q15645, P53396, P38919, Q92900, Q14566	2,98	7,91E-03	1,13E-03	1,96E-02
SMART	SM00382:AAA	9	6,04	1,53E-05	P33993, Q14204, Q9BSD7, Q9NU22, P17980, Q15645, Q5T9A4, P28288, P25205	7,92	1,28E-03	1,28E-03	1,64E-02
Annotation Cluster 7 - Enrichment Score: 5.25									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	Aminoacyl-tRNA synthetase	9	6,04	2,41E-10	P26640, Q9Y285, P56192, Q5T160, P49588, Q9P2J5, Q9NSD9, P41252, P54136	32,71	5,61E-08	1,12E-08	3,09E-07
GOTERM_BP_DIRECT	GO:0006418~tRNA aminoacylation for protein translation	9	6,04	1,55E-09	P26640, Q9Y285, P56192, Q5T160, P49588, Q9P2J5, Q9NSD9, P41252, P54136	25,88	1,49E-06	2,99E-07	2,44E-06

UP_KEYWORDS	Ligase	16	10,74	3,24E-08	P17812, Q9Y285, Q86YT6, P56192, Q6P1M0, Q5T160, P49588, P41252, P54136, Q6UB35, P26640, Q13085, P27708, Q9Y3I0, Q9P2J5, Q9NSD9	6,37	7,56E-06	8,39E-07	4,17E-05
INTERPRO	IPR009080:Aminoacyl-tRNA synthetase, class 1a, anticodon-binding	6	4,03	3,44E-08	P26640, P56192, Q5T160, Q9P2J5, P41252, P54136	58,27	1,64E-05	5,48E-06	4,92E-05
INTERPRO	IPR001412:Aminoacyl-tRNA synthetase, class I, conserved site	6	4,03	7,93E-08	P26640, P56192, Q5T160, Q9P2J5, P41252, P54136	50,50	3,79E-05	9,47E-06	1,13E-04
UP_SEQ_FEATURE	short sequence motif:"HIGH" region	6	4,03	1,62E-07	P26640, P56192, Q5T160, Q9P2J5, P41252, P54136	44,88	9,70E-05	1,94E-05	2,39E-04
INTERPRO	IPR014729:Rossmann-like alpha/beta/alpha sandwich fold	7	4,70	8,88E-07	P26640, P56192, Q5T160, Q9P2J5, P41252, P13804, P54136	21,04	4,24E-04	7,08E-05	1,27E-03
UP_KEYWORDS	Protein biosynthesis	10	6,71	1,58E-06	P26640, Q9Y285, P56192, Q5T160, P49588, Q9P2J5, Q9NSD9, P41252, Q14152, P54136	9,09	3,67E-04	2,30E-05	2,03E-03
KEGG_PATHWAY	hsa00970:Aminoacyl-tRNA biosynthesis	9	6,04	6,46E-06	P26640, Q9Y285, P56192, Q5T160, P49588, Q9P2J5, Q9NSD9, P41252, P54136	8,81	9,82E-04	9,82E-04	7,74E-03
GOTERM_MF_DIRECT	GO:0004812~aminoacyl-tRNA ligase activity	5	3,36	1,14E-05	P26640, Q9Y285, P56192, Q9P2J5, P41252	34,01	4,24E-03	5,31E-04	1,57E-02
GOTERM_MF_DIRECT	GO:0002161~aminoacyl-tRNA editing activity	4	2,68	7,13E-05	P26640, P49588, Q9P2J5, P41252	46,25	2,62E-02	2,95E-03	9,84E-02
GOTERM_MF_DIRECT	GO:0000049~tRNA binding	6	4,03	7,46E-05	Q9Y285, P56192, P49588, P41252, O43592, P54136	13,60	2,74E-02	2,78E-03	1,03E-01
UP_SEQ_FEATURE	short sequence motif:"KMSKS" region	4	2,68	2,05E-04	P26640, P56192, Q9P2J5, P41252	33,66	1,16E-01	8,17E-03	3,03E-01
GOTERM_BP_DIRECT	GO:0006450~regulation of translational fidelity	4	2,68	2,66E-04	P26640, P49588, Q9P2J5, P41252	30,67	2,26E-01	2,53E-02	4,17E-01

INTERPRO	IPR013155:Valyl/Leucyl/Isoleucyl-tRNA synthetase, class I, anticodon-binding	3	2,01	9,03E-04	P26640, Q9P2J5, P41252	63,13	3,51E-01	3,85E-02	1,28E+00
INTERPRO	IPR009008:Valyl/Leucyl/Isoleucyl-tRNA synthetase, class Ia, editing domain	3	2,01	9,03E-04	P26640, Q9P2J5, P41252	63,13	3,51E-01	3,85E-02	1,28E+00
INTERPRO	IPR002300:Aminoacyl-tRNA synthetase, class Ia	3	2,01	9,03E-04	P26640, Q9P2J5, P41252	63,13	3,51E-01	3,85E-02	1,28E+00
Annotation Cluster 8 - Enrichment Score: 4.95									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	Mitochondrion	28	18,79	2,69E-08	P31689, P53007, O43615, Q3ZCQ8, P45880, Q07021, O00571, Q5T9A4, Q6UB35, O95831, P05141, O75746, P13804, Q02978, P36542, P23396, Q9UJS0, Q96HS1, Q5T160, P54886, P04181, P38646, P11177, Q8NI60, P42704, Q9H9B4, P48047, Q00325	3,46	6,26E-06	8,95E-07	3,45E-05
UP_KEYWORDS	Transit peptide	16	10,74	7,73E-06	Q5T160, P53007, O43615, Q3ZCQ8, Q07021, Q6UB35, P04181, P38646, O95831, P11177, Q8NI60, P42704, P48047, Q00325, P13804, P36542	4,12	1,80E-03	1,00E-04	9,93E-03
UP_SEQ_FEATURE	transit peptide:Mitochondrion	15	10,07	1,38E-05	Q5T160, P53007, O43615, Q3ZCQ8, Q07021, Q6UB35, P38646, O95831, P11177, Q8NI60, P42704, P48047, Q00325, P13804, P36542	4,19	8,23E-03	1,03E-03	2,04E-02
GOTERM_CC_DIRECT	GO:0005759~mitochondrial matrix	9	6,04	5,10E-03	P11177, Q5T160, O43615, Q07021, P13804, P36542, Q6UB35, P23396, P04181	3,39	7,27E-01	5,27E-02	6,45E+00

Annotation Cluster 9 - Enrichment Score: 3.84

Category	Term	Count	%	PValue	Genes	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KEYWORDS	Neurodegeneration	13	8,72	1,44E-06	P56192, Q5T160, P49588, P50570, P54886, O00567, Q9UBB4, O95831, Q8NI60, Q14204, Q8IY17, Q00535, P12004	6,13	3,36E-04	2,24E-05	1,85E-03
UP_KEYWORDS	Charcot-Marie-Tooth disease	5	3,36	4,93E-04	O95831, P56192, P49588, Q14204, P50570	13,54	1,09E-01	4,78E-03	6,32E-01
UP_KEYWORDS	Neuropathy	5	3,36	4,07E-03	O95831, P56192, P49588, Q14204, P50570	7,67	6,13E-01	2,75E-02	5,10E+00

Annotation Cluster 10 - Enrichment Score: 3.44

Category	Term	Count	%	PValue	Genes	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KEYWORDS	Mitochondrion inner membrane	14	9,40	8,07E-08	P53007, O43615, Q3ZCQ8, Q5T9A4, P54886, O95831, P05141, P48047, Q00325, O75746, Q02978, P36542, P23396, Q9UJS0	7,16	1,88E-05	1,88E-06	1,04E-04
UP_SEQ_FEATURE	repeat:Solcar 3	6	4,03	3,65E-05	P05141, P53007, O75746, Q00325, Q02978, Q9UJS0	15,84	2,16E-02	1,98E-03	5,39E-02
UP_SEQ_FEATURE	repeat:Solcar 1	6	4,03	4,83E-05	P05141, P53007, O75746, Q00325, Q02978, Q9UJS0	14,96	2,85E-02	2,41E-03	7,12E-02
UP_SEQ_FEATURE	repeat:Solcar 2	6	4,03	4,83E-05	P05141, P53007, O75746, Q00325, Q02978, Q9UJS0	14,96	2,85E-02	2,41E-03	7,12E-02
INTERPRO	IPR018108:Mitochondrial substrate/solute carrier	6	4,03	5,96E-05	P05141, P53007, O75746, Q00325, Q02978, Q9UJS0	14,29	2,81E-02	2,84E-03	8,52E-02
INTERPRO	IPR023395:Mitochondrial carrier domain	6	4,03	5,96E-05	P05141, P53007, O75746, Q00325, Q02978, Q9UJS0	14,29	2,81E-02	2,84E-03	8,52E-02
INTERPRO	IPR002067:Mitochondrial carrier protein	4	2,68	1,21E-03	P05141, P53007, O75746, Q9UJS0	18,70	4,41E-01	4,73E-02	1,72E+00

GOTERM_BP_DIRECT	GO:0006094~gluconeogenesis	4	2,68	6,45E-03	P53007, O75746, Q02978, Q9UJS0	10,46	9,98E-01	2,00E-01	9,67E+00
Annotation Cluster 11 - Enrichment Score: 3.05									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
INTERPRO	IPR001494:Importin-beta, N-terminal	5	3,36	8,08E-06	P55060, O14980, Q8TEX9, O43592, O00410	37,13	3,85E-03	5,51E-04	1,16E-02
GOTERM_MF_DIRECT	GO:0008536~Ran GTPase binding	5	3,36	1,20E-04	P55060, O14980, Q8TEX9, O43592, O00410	19,27	4,39E-02	4,07E-03	1,66E-01
UP_SEQ_FEATURE	domain:Importin N-terminal	4	2,68	1,35E-04	P55060, O14980, Q8TEX9, O00410	38,47	7,76E-02	6,20E-03	1,99E-01
SMART	SM00913:SM00913	4	2,68	2,81E-04	P55060, O14980, Q8TEX9, O43592	29,93	2,33E-02	7,84E-03	3,01E-01
GOTERM_CC_DIRECT	GO:0005643~nuclear pore	6	4,03	2,89E-04	Q12769, Q92621, P52292, Q8TEX9, O43592, O00410	10,26	7,08E-02	3,86E-03	3,76E-01
GOTERM_BP_DIRECT	GO:0000059~protein import into nucleus, docking	3	2,01	2,00E-03	Q92621, Q8TEX9, O00410	43,13	8,55E-01	9,68E-02	3,11E+00
GOTERM_BP_DIRECT	GO:0006886~intracellular protein transport	8	5,37	4,44E-03	Q9Y678, O14980, Q8TEX9, Q00535, Q04917, Q00610, O43592, O00410	3,90	9,86E-01	1,52E-01	6,76E+00
GOTERM_CC_DIRECT	GO:0034399~nuclear periphery	3	2,01	1,23E-02	Q92621, Q8TEX9, O00410	17,59	9,57E-01	1,10E-01	1,49E+01
Annotation Cluster 12 - Enrichment Score: 2.79									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
INTERPRO	IPR023410:14-3-3 domain	4	2,68	1,63E-05	P62258, Q04917, P31946, P61981	72,14	7,77E-03	8,66E-04	2,33E-02
INTERPRO	IPR023409:14-3-3 protein, conserved site	4	2,68	1,63E-05	P62258, Q04917, P31946, P61981	72,14	7,77E-03	8,66E-04	2,33E-02
INTERPRO	IPR000308:14-3-3 protein	4	2,68	1,63E-05	P62258, Q04917, P31946, P61981	72,14	7,77E-03	8,66E-04	2,33E-02
SMART	SM00101:14_3_3	4	2,68	1,85E-05	P62258, Q04917, P31946, P61981	68,41	1,55E-03	7,78E-04	1,99E-02
UP_SEQ_FEATURE	site:Interaction with phosphoserine on interacting protein	4	2,68	2,14E-05	P62258, Q04917, P31946, P61981	67,33	1,28E-02	1,43E-03	3,17E-02
PIR_SUPERFAMILY	PIRSF000868:14-3-3 protein	4	2,68	4,09E-05	P62258, Q04917, P31946, P61981	48,34	6,95E-04	6,95E-04	2,96E-02

KEGG_PATHWAY	hsa04110:Cell cycle	10	6,71	1,10E-04	P33993, P62258, Q14683, Q04917, P31946, P12004, P61981, P78527, Q14566, P25205	5,21	1,65E-02	5,54E-03	1,31E-01
GOTERM_MF_DIRECT	GO:0019904~protein domain specific binding	8	5,37	2,15E-03	A5YKK6, P62258, P40616, O14980, Q04917, P31946, P05023, P61981	4,45	5,51E-01	5,98E-02	2,92E+00
GOTERM_BP_DIRECT	GO:1900740~positive regulation of protein insertion into mitochondrial membrane involved in apoptotic signaling pathway	4	2,68	2,16E-03	P62258, Q04917, P31946, P61981	15,34	8,75E-01	9,44E-02	3,34E+00
GOTERM_BP_DIRECT	GO:0061024~membrane organization	4	2,68	2,60E-03	P62258, Q04917, P31946, P61981	14,38	9,19E-01	1,04E-01	4,02E+00
KEGG_PATHWAY	hsa04114:Oocyte meiosis	7	4,70	6,39E-03	P62258, Q14683, P30153, Q08209, Q04917, P31946, P61981	4,15	6,23E-01	2,16E-01	7,39E+00
GOTERM_BP_DIRECT	GO:0000086~G2/M transition of mitotic cell cycle	6	4,03	6,74E-03	P63151, P62258, Q14204, P50570, P30153, P61981	5,04	9,99E-01	2,01E-01	1,01E+01
GOTERM_BP_DIRECT	GO:0006605~protein targeting	3	2,01	4,46E-02	P62258, P31946, P61981	8,85	1,00E+00	6,32E-01	5,12E+01
Annotation Cluster 13 - Enrichment Score: 2.40									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_MF_DIRECT	GO:0051082~unfolded protein binding	7	4,70	3,73E-04	P31689, Q9NU22, P50990, P50991, O60884, P50454, P38646	7,36	1,30E-01	1,15E-02	5,13E-01
UP_KEYWORDS	Chaperone	7	4,70	3,37E-03	P31689, Q9NU22, P50990, P50991, O60884, P50454, P38646	4,81	5,44E-01	2,43E-02	4,24E+00
GOTERM_BP_DIRECT	GO:0006457~protein folding	7	4,70	4,75E-03	P31689, Q14697, P49588, P50990, P50991, O60884, P38646	4,47	9,90E-01	1,56E-01	7,22E+00
GOTERM_BP_DIRECT	GO:1901998~toxin transport	3	2,01	3,85E-02	P31689, P50990, P50991	9,58	1,00E+00	5,86E-01	4,61E+01

Annotation Cluster 14 - Enrichment Score: 2.35

Category	Term	Count	%	PValue	Genes	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KEYWORDS	DNA repair	9	6,04	1,30E-03	Q96T76, Q16531, Q9NVI1, Q14683, O95071, P23396, P12004, Q93009, P78527	4,24	2,61E-01	1,15E-02	1,65E+00
UP_KEYWORDS	DNA damage	9	6,04	4,02E-03	Q96T76, Q16531, Q9NVI1, Q14683, O95071, P23396, P12004, Q93009, P78527	3,53	6,08E-01	2,80E-02	5,04E+00
GOTERM_BP_DIRECT	GO:0006281~DNA repair	7	4,70	1,64E-02	Q96T76, Q16531, Q9NVI1, Q14683, O95071, P23396, Q92900	3,43	1,00E+00	3,83E-01	2,29E+01

Annotation Cluster 15 - Enrichment Score: 2.24

Category	Term	Count	%	PValue	Genes	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_DIRECT	GO:0055129~L-proline biosynthetic process	3	2,01	7,28E-04	Q53H96, P54886, P04181	69,01	5,04E-01	4,89E-02	1,14E+00
GOTERM_BP_DIRECT	GO:0008652~cellular amino acid biosynthetic process	4	2,68	1,42E-03	P0DN79, Q53H96, P54886, P04181	17,69	7,45E-01	8,19E-02	2,21E+00

Annotation Cluster 16 - Enrichment Score: 1.97

Category	Term	Count	%	PValue	Genes	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KEYWORDS	Helicase	8	5,37	7,97E-05	P33993, Q9NR30, O75643, O00571, P38919, Q92900, Q14566, P25205	7,73	1,84E-02	8,84E-04	1,02E-01
GOTERM_CC_DIRECT	GO:0000784~nuclear chromosome, telomeric region	7	4,70	6,53E-04	P33993, Q16531, P12004, Q92900, P78527, Q14566, P25205	6,63	1,53E-01	8,26E-03	8,47E-01
INTERPRO	IPR018525:Mini-chromosome maintenance, conserved site	3	2,01	9,03E-04	P33993, Q14566, P25205	63,13	3,51E-01	3,85E-02	1,28E+00
UP_SEQ_FEATURE	domain:MCM	3	2,01	1,47E-03	P33993, Q14566, P25205	50,49	5,86E-01	4,78E-02	2,15E+00

INTERPRO	IPR001208:Mini-chromosome maintenance, DNA-dependent ATPase	3	2,01	2,13E-03	P33993, Q14566, P25205	42,08	6,40E-01	7,55E-02	3,01E+00
GOTERM_CC_DIRECT	GO:0042555~MCM complex	3	2,01	2,24E-03	P33993, Q14566, P25205	41,05	4,34E-01	2,56E-02	2,88E+00
SMART	SM00350:MCM	3	2,01	2,33E-03	P33993, Q14566, P25205	39,91	1,78E-01	4,79E-02	2,48E+00
GOTERM_BP_DIRECT	GO:0000082~G1/S transition of mitotic cell cycle	5	3,36	1,18E-02	P33993, Q08209, P12004, Q14566, P25205	5,64	1,00E+00	3,09E-01	1,70E+01
KEGG_PATHWAY	hsa03030:DNA replication	4	2,68	1,74E-02	P33993, P12004, Q14566, P25205	7,18	9,30E-01	3,58E-01	1,89E+01
GOTERM_MF_DIRECT	GO:0003678~DNA helicase activity	3	2,01	1,79E-02	P33993, Q14566, P25205	14,45	9,99E-01	2,74E-01	2,20E+01
BIOCARTA	h_mcmPathway:CDK Regulation of DNA Replication	3	2,01	2,77E-02	P33993, Q14566, P25205	10,83	8,30E-01	5,87E-01	2,48E+01
UP_KEYWORDS	DNA replication	4	2,68	3,04E-02	P33993, P12004, Q14566, P25205	5,88	9,99E-01	1,51E-01	3,27E+01
GOTERM_BP_DIRECT	GO:0006270~DNA replication initiation	3	2,01	3,10E-02	P33993, Q14566, P25205	10,78	1,00E+00	5,15E-01	3,91E+01
GOTERM_BP_DIRECT	GO:0006260~DNA replication	5	3,36	4,55E-02	P33993, P12004, Q92900, Q14566, P25205	3,71	1,00E+00	6,31E-01	5,19E+01
Annotation Cluster 17 - Enrichment Score: 1.90									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_CC_DIRECT	GO:0005643~nuclear pore	6	4,03	2,89E-04	Q12769, Q92621, P52292, Q8TEX9, O43592, O00410	10,26	7,08E-02	3,86E-03	3,76E-01
GOTERM_BP_DIRECT	GO:0006607~NLS-bearing protein import into nucleus	3	2,01	1,66E-02	P52292, Q8TEX9, O00410	15,00	1,00E+00	3,79E-01	2,32E+01
COG_ONTOLOGY	Intracellular trafficking and secretion	4	2,68	2,08E-02	Q9Y678, P52292, Q8TEX9, O00410	6,43	2,71E-01	2,71E-01	1,36E+01
GOTERM_MF_DIRECT	GO:0008139~nuclear localization sequence binding	3	2,01	2,39E-02	P52292, Q8TEX9, O00410	12,39	1,00E+00	3,37E-01	2,84E+01

Annotation Cluster 18 - Enrichment Score: 1.88									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_BP_DIRECT	GO:0035338~long-chain fatty-acyl-CoA biosynthetic process	5	3,36	4,63E-04	Q13085, P53007, P53396, P49327, Q53GQ0	13,69	3,60E-01	3,65E-02	7,25E-01
GOTERM_BP_DIRECT	GO:0031325~positive regulation of cellular metabolic process	3	2,01	7,28E-04	Q13085, P53396, P49327	69,01	5,04E-01	4,89E-02	1,14E+00
GOTERM_BP_DIRECT	GO:0006633~fatty acid biosynthetic process	5	3,36	1,05E-03	Q13085, P53396, P49327, P28288, Q53GQ0	11,06	6,36E-01	6,51E-02	1,63E+00
Annotation Cluster 19 - Enrichment Score: 1.86									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_BP_DIRECT	GO:0008380~RNA splicing	8	5,37	5,99E-04	Q14498, P26599, Q07021, P30153, Q15029, Q15393, Q6P2Q9, P38919	5,54	4,39E-01	4,35E-02	9,38E-01
UP_KEYWORDS	mRNA splicing	8	5,37	2,80E-03	Q14498, P26599, O75643, Q07021, Q15029, Q15393, Q6P2Q9, P38919	4,25	4,79E-01	2,08E-02	3,54E+00
GOTERM_BP_DIRECT	GO:0000398~mRNA splicing, via spliceosome	8	5,37	3,18E-03	P26599, O75643, Q15029, Q15393, Q6P2Q9, P38919, Q15366, Q15365	4,14	9,54E-01	1,20E-01	4,89E+00
GOTERM_CC_DIRECT	GO:0071013~catalytic step 2 spliceosome	5	3,36	6,56E-03	O75643, Q15029, Q15393, Q6P2Q9, P38919	6,69	8,12E-01	6,23E-02	8,21E+00
UP_KEYWORDS	mRNA processing	8	5,37	1,03E-02	Q14498, P26599, O75643, Q07021, Q15029, Q15393, Q6P2Q9, P38919	3,33	9,11E-01	6,32E-02	1,25E+01
UP_KEYWORDS	Spliceosome	5	3,36	1,34E-02	O75643, Q15029, Q15393, Q6P2Q9, P38919	5,44	9,57E-01	7,75E-02	1,59E+01
GOTERM_BP_DIRECT	GO:0006397~mRNA processing	6	4,03	1,96E-02	Q14498, P26599, Q07021, Q15029, Q15393, Q6P2Q9	3,86	1,00E+00	3,95E-01	2,68E+01
GOTERM_CC_DIRECT	GO:0016607~nuclear speck	6	4,03	2,36E-02	Q14498, Q3ZCQ8, O00571, Q15029, Q6P2Q9, P38919	3,68	9,98E-01	1,78E-01	2,67E+01

GOTERM_CC_DIRECT	GO:0005681~spliceosomal complex	4	2,68	4,05E-02	O75643, Q15029, Q15393, Q6P2Q9	5,24	1,00E+00	2,66E-01	4,16E+01
Annotation Cluster 20 - Enrichment Score: 1.85									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	mRNA transport	5	3,36	7,24E-03	Q12769, Q92621, P42704, O14980, P38919	6,52	8,16E-01	4,72E-02	8,91E+00
GOTERM_BP_DIRECT	GO:0075733~intracellular transport of virus	4	2,68	9,70E-03	Q12769, Q92621, O14980, P52292	9,02	1,00E+00	2,69E-01	1,42E+01
GOTERM_CC_DIRECT	GO:0005635~nuclear envelope	5	3,36	3,98E-02	Q12769, Q92621, P55060, Q9Y3I0, O14980	3,87	1,00E+00	2,75E-01	4,11E+01
Annotation Cluster 21 - Enrichment Score: 1.72									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	mRNA transport	5	3,36	7,24E-03	Q12769, Q92621, P42704, O14980, P38919	6,52	8,16E-01	4,72E-02	8,91E+00
KEGG_PATHWAY	hsa03013:RNA transport	8	5,37	1,62E-02	Q12769, O14744, Q92621, O14980, Q14152, O43592, P38919, Q92900	3,00	9,17E-01	3,92E-01	1,78E+01
Annotation Cluster 22 - Enrichment Score: 1.57									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	Helicase	8	5,37	7,97E-05	P33993, Q9NR30, O75643, O00571, P38919, Q92900, Q14566, P25205	7,73	1,84E-02	8,84E-04	1,02E-01
GOTERM_MF_DIRECT	GO:0004004~ATP-dependent RNA helicase activity	5	3,36	2,36E-03	Q9NR30, O75643, O00571, P38919, Q92900	8,89	5,85E-01	6,09E-02	3,20E+00
INTERPRO	IPR011545:DNA/RNA helicase, DEAD/DEAH box type, N-terminal	4	2,68	1,71E-02	Q9NR30, O75643, O00571, P38919	7,32	1,00E+00	3,85E-01	2,19E+01
INTERPRO	IPR014014:RNA helicase, DEAD-box type, Q motif	3	2,01	3,42E-02	Q9NR30, O00571, P38919	10,24	1,00E+00	5,83E-01	3,92E+01

UP_SEQ_FEATURE	short sequence motif:Q motif	3	2,01	3,51E-02	Q9NR30, O00571, P38919	10,10	1,00E+00	5,90E-01	4,10E+01
GOTERM_MF_DIRECT	GO:0004386~helicase activity	4	2,68	3,68E-02	O75643, O00571, P38919, Q92900	5,44	1,00E+00	3,93E-01	4,04E+01
Annotation Cluster 23 - Enrichment Score: 1.44									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_BP_DIRECT	GO:0008652~cellular amino acid biosynthetic process	4	2,68	1,42E-03	P0DN79, Q53H96, P54886, P04181	17,69	7,45E-01	8,19E-02	2,21E+00
UP_KEYWORDS	Amino-acid biosynthesis	3	2,01	1,49E-02	P0DN79, Q53H96, P54886	15,94	9,70E-01	8,38E-02	1,76E+01
Annotation Cluster 24 - Enrichment Score: 1.24									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_MF_DIRECT	GO:0004722~protein serine/threonine phosphatase activity	4	2,68	1,36E-02	P63151, Q3ZCQ8, P30153, Q08209	7,97	9,94E-01	2,35E-01	1,72E+01
GOTERM_BP_DIRECT	GO:0006470~protein dephosphorylation	5	3,36	2,38E-02	Q96HS1, P63151, Q3ZCQ8, P30153, Q08209	4,56	1,00E+00	4,48E-01	3,15E+01
Annotation Cluster 25 - Enrichment Score: 1.19									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	Cell cycle	13	8,72	2,62E-03	Q9NVI1, Q9NNW5, Q5T9A4, Q93008, P25205, P33993, P42695, Q14683, Q00535, P63244, Q00610, P23396, Q14566	2,76	4,57E-01	2,02E-02	3,31E+00
Annotation Cluster 26 - Enrichment Score: 1.08									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_MF_DIRECT	GO:0003924~GTPase activity	7	4,70	1,57E-02	P18085, P40616, P50570, O00571, Q15029, Q9BUF5, P51148	3,46	9,97E-01	2,56E-01	1,96E+01

	INTERPRO	IPR024156:Small GTPase superfamily, ARF type	3	2,01	2,61E-02	Q9Y5M8, P18085, P40616	11,84	1,00E+00	5,05E-01	3,15E+01
	UP_SEQ_FEATURE	nucleotide phosphate-binding region:GTP	7	4,70	2,76E-02	Q9Y5M8, P18085, P40616, P50570, Q15029, Q9BUF5, P51148	3,04	1,00E+00	5,18E-01	3,39E+01
	UP_KEYWORDS	GTP-binding	7	4,70	3,79E-02	Q9Y5M8, P18085, P40616, P50570, Q15029, Q9BUF5, P51148	2,82	1,00E+00	1,71E-01	3,91E+01
NUCLEAR EXTRACTS	**Annotation Cluster 1 - Enrichment Score: 2.79**									
	Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
	GOTERM_BP_DIRECT	GO:0016925~protein sumoylation	5	21,74	1,22E-05	Q12769, Q92621, P49792, Q14683, Q9UQE7	32,62	2,69E-03	1,34E-03	1,56E-02
	GOTERM_CC_DIRECT	GO:0005643~nuclear pore	4	17,39	7,47E-05	Q12769, Q9UBU9, Q92621, P49792	46,02	6,33E-03	6,33E-03	8,04E-02
	UP_KEYWORDS	mRNA transport	4	17,39	1,90E-04	Q12769, Q9UBU9, Q92621, P49792	33,77	1,76E-02	2,53E-03	2,08E-01
	GOTERM_BP_DIRECT	GO:0006406~mRNA export from nucleus	4	17,39	2,52E-04	Q12769, Q9UBU9, Q92621, P49792	30,53	5,40E-02	1,38E-02	3,20E-01
	GOTERM_BP_DIRECT	GO:0019083~viral transcription	4	17,39	3,52E-04	Q12769, Q92621, P49792, P23396	27,26	7,45E-02	1,54E-02	4,47E-01
	GOTERM_BP_DIRECT	GO:0016032~viral process	5	21,74	4,64E-04	Q12769, Q9UBU9, Q92621, P49792, P08670	12,76	9,71E-02	1,69E-02	5,89E-01
	GOTERM_BP_DIRECT	GO:0006409~tRNA export from nucleus	3	13,04	7,22E-04	Q12769, Q92621, P49792	71,56	1,47E-01	2,25E-02	9,16E-01
	GOTERM_BP_DIRECT	GO:0010827~regulation of glucose transport	3	13,04	7,68E-04	Q12769, Q92621, P49792	69,39	1,56E-01	2,09E-02	9,74E-01
	UP_KEYWORDS	Nuclear pore complex	3	13,04	1,30E-03	Q12769, Q92621, P49792	53,69	1,14E-01	1,20E-02	1,41E+00
	GOTERM_BP_DIRECT	GO:0007077~mitotic nuclear envelope disassembly	3	13,04	1,37E-03	Q12769, Q92621, P49792	52,04	2,60E-01	2,96E-02	1,72E+00
	GOTERM_BP_DIRECT	GO:0075733~intracellular transport of virus	3	13,04	1,83E-03	Q12769, Q92621, P49792	44,90	3,32E-01	3,60E-02	2,31E+00
	UP_KEYWORDS	Translocation	3	13,04	3,61E-03	Q12769, Q92621, P49792	31,96	2,85E-01	2,08E-02	3,88E+00

GOTERM_BP_DIRECT	GO:1900034~regulation of cellular response to heat	3	13,04	3,91E-03	Q12769, Q92621, P49792	30,53	5,78E-01	6,93E-02	4,87E+00
KEGG_PATHWAY	hsa03013:RNA transport	4	17,39	5,53E-03	Q12769, Q9UBU9, Q92621, P49792	10,04	1,39E-01	1,39E-01	4,49E+00
GOTERM_BP_DIRECT	GO:0031047~gene silencing by RNA	3	13,04	8,38E-03	Q12769, Q92621, P49792	20,63	8,43E-01	1,33E-01	1,02E+01
GOTERM_CC_DIRECT	GO:0005635~nuclear envelope	3	13,04	1,42E-02	Q12769, Q92621, P49792	15,63	7,05E-01	1,15E-01	1,43E+01
Annotation Cluster 2 - Enrichment Score: 8.13									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_BP_DIRECT	GO:0006281~DNA repair	5	21,74	1,86E-04	Q96PK6, Q14683, Q9UQE7, Q9Y265, P23396	16,24	4,00E-02	1,35E-02	2,36E-01
UP_KEYWORDS	DNA repair	5	21,74	2,40E-04	Q14683, Q9UQE7, Q9Y265, P23396, P78527	15,27	2,21E-02	2,79E-03	2,63E-01
UP_KEYWORDS	DNA damage	5	21,74	4,82E-04	Q14683, Q9UQE7, Q9Y265, P23396, P78527	12,71	4,39E-02	4,97E-03	5,27E-01
GOTERM_BP_DIRECT	GO:0051301~cell division	5	21,74	8,39E-04	Q14683, Q9BQE3, Q9UQE7, Q9Y265, P23396	10,90	1,69E-01	2,03E-02	1,06E+00
UP_KEYWORDS	Mitosis	4	17,39	2,63E-03	Q14683, Q9UQE7, Q9Y265, P23396	13,66	2,17E-01	1,86E-02	2,84E+00
UP_KEYWORDS	Cell cycle	5	21,74	4,58E-03	Q14683, Q9NNW5, Q9UQE7, Q9Y265, P23396	6,88	3,48E-01	2,35E-02	4,90E+00
UP_KEYWORDS	Cell division	4	17,39	7,85E-03	Q14683, Q9UQE7, Q9Y265, P23396	9,23	5,19E-01	3,60E-02	8,26E+00
GOTERM_BP_DIRECT	GO:0007067~mitotic nuclear division	3	13,04	3,79E-02	Q9UQE7, Q9Y265, P23396	9,23	1,00E+00	4,12E-01	3,89E+01
Annotation Cluster 3 - Enrichment Score: 2.66									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
GOTERM_MF_DIRECT	GO:0003723~RNA binding	7	30,43	4,03E-05	Q96PK6, Q15717, Q9UBU9, P49792, O00571, P23396, Q14690	9,82	4,07E-03	2,03E-03	4,48E-02

UP_KEYWORDS	RNA-binding	7	30,43	5,33E-05	Q96PK6, Q15717, Q9NR30, Q9UBU9, P49792, O00571, P23396	9,42	4,95E-03	8,26E-04	5,83E-02
INTERPRO	IPR012677:Nucleotide-binding, alpha-beta plait	4	17,39	3,13E-03	Q96PK6, Q15717, Q9NR30, Q9UBU9	12,78	2,22E-01	1,18E-01	3,28E+00
GOTERM_MF_DIRECT	GO:0003729~mRNA binding	3	13,04	1,04E-02	Q15717, Q9UBU9, P23396	18,42	6,53E-01	1,40E-01	1,10E+01
UP_KEYWORDS	Methylation	5	21,74	2,02E-02	Q15717, Q9UBU9, P49792, O00571, P23396	4,47	8,50E-01	7,92E-02	2,00E+01
Annotation Cluster 4 - Enrichment Score: 2.53									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	DNA repair	5	21,74	2,40E-04	Q14683, Q9UQE7, Q9Y265, P23396, P78527	15,27	2,21E-02	2,79E-03	2,63E-01
UP_KEYWORDS	DNA damage	5	21,74	4,82E-04	Q14683, Q9UQE7, Q9Y265, P23396, P78527	12,71	4,39E-02	4,97E-03	5,27E-01
UP_KEYWORDS	Nucleotide-binding	8	34,78	1,96E-03	Q9NR30, Q9NVI7, Q14683, Q9BQE3, O00571, Q9UQE7, Q9Y265, P78527	4,00	1,66E-01	1,51E-02	2,12E+00
INTERPRO	IPR027417:P-loop containing nucleoside triphosphate hydrolase	6	26,09	2,54E-03	Q9NR30, Q9NVI7, Q14683, O00571, Q9UQE7, Q9Y265	5,76	1,84E-01	1,84E-01	2,67E+00
UP_KEYWORDS	ATP-binding	7	30,43	2,75E-03	Q9NR30, Q9NVI7, Q14683, O00571, Q9UQE7, Q9Y265, P78527	4,50	2,26E-01	1,81E-02	2,97E+00
UP_SEQ_FEATURE	nucleotide phosphate-binding region:ATP	6	26,09	3,84E-03	Q9NR30, Q9NVI7, Q14683, O00571, Q9UQE7, Q9Y265	5,27	4,66E-01	2,69E-01	4,56E+00
GOTERM_MF_DIRECT	GO:0005524~ATP binding	7	30,43	8,10E-03	Q9NR30, Q9NVI7, Q14683, O00571, Q9UQE7, Q9Y265, P78527	3,59	5,60E-01	1,28E-01	8,64E+00
UP_KEYWORDS	Helicase	3	13,04	1,01E-02	Q9NR30, O00571, Q9Y265	18,77	6,11E-01	4,40E-02	1,05E+01
KEGG_PATHWAY	hsa04110:Cell cycle	3	13,04	2,88E-02	Q14683, Q9UQE7, P78527	10,45	5,46E-01	3,26E-01	2,15E+01

Annotation Cluster 5 - Enrichment Score: 1.03									
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	Mitochondrion inner membrane	4	17,39	2,86E-03	O95831, Q9NVI7, P23396, Q99623	13,26	2,34E-01	1,76E-02	3,09E+00
UP_KEYWORDS	Mitochondrion	6	26,09	5,71E-03	O95831, Q96HS1, Q9NVI7, O00571, P23396, Q99623	4,80	4,13E-01	2,76E-02	6,08E+00
GOTERM_CC_DIRECT	GO:0005743~mitochondrial inner membrane	4	17,39	1,35E-02	O95831, Q9NVI7, P23396, Q99623	7,51	6,86E-01	1,21E-01	1,36E+01
Annotation Cluster 6	**Enrichment Score: 0.9551892733349611**								
Category	**Term**	**Count**	**%**	**PValue**	**Genes**	**Fold Enrichment**	**Bonferroni**	**Benjamini**	**FDR**
UP_KEYWORDS	Mitochondrion	6	26,09	5,71E-03	O95831, Q96HS1, Q9NVI7, O00571, P23396, Q99623	4,80	4,13E-01	2,76E-02	6,08E+00

Supplementary Table 4. IPA canonical signaling pathways enrichment in the ANKRD55 interactome from nuclear and total protein extract. Significantly enriched pathways (-Log (p-value)>1.30) are shown.

	Ingenuity Canonical Pathways	-Log(p-value)	Molecules
TOTAL PROTEIN EXTRACTS	tRNA Charging	11,40	LARS,RARS,RARS2,AARS,VARS,MARS,IARS,FARSA,FARSB
	EIF2 Signaling	9,49	PTBP1,RPL4,RPS3A,EIF4A3,RPS9,EIF3A,RPL10,RPS15A,RPL23,RPS2,RPS3,RPL7A,RPL9,RPLP0
	Regulation of eIF4 and p70S6K Signaling	5,91	PPP2R1A,RPS3A,PPP2R2A,EIF4A3,RPS9,EIF3A,RPS15A,RPS2,RPS3
	RAN Signaling	5,36	CSE1L,XPO1,KPNA2,IPO5
	mTOR Signaling	5,03	PPP2R1A,RPS3A,PPP2R2A,EIF4A3,RPS9,EIF3A,RPS15A,RPS2,RPS3
	Cell Cycle: G2/M DNA Damage Checkpoint Regulation	4,66	PRKDC,YWHAG,YWHAH,YWHAE,YWHAB
	HIPPO signaling	4,56	PPP2R1A,YWHAG,YWHAH,YWHAE,YWHAB,PPP2R2A
	Cell Cycle Control of Chromosomal Replication	4,42	MCM3,MCM6,PCNA,CDK5,MCM7
	PI3K/AKT Signaling	3,69	PPP2R1A,YWHAG,YWHAH,YWHAE,YWHAB,PPP2R2A
	Proline Biosynthesis I	3,57	PYCR3,ALDH18A1
	p70S6K Signaling	3,56	PPP2R1A,YWHAG,YWHAH,YWHAE,YWHAB,PPP2R2A
	ERK/MAPK Signaling	3,37	PPP2R1A,YWHAG,YWHAH,ARAF,YWHAB,PPP2R2A,TLN1
	Proline Biosynthesis II (from Arginine)	3,18	PYCR3,OAT
	Arginine Degradation VI (Arginase 2 Pathway)	3,18	PYCR3,OAT
	ERK5 Signaling	2,99	YWHAG,YWHAH,YWHAE,YWHAB
	Myc Mediated Apoptosis Signaling	2,89	YWHAG,YWHAH,YWHAE,YWHAB
	Caveolar-mediated Endocytosis Signaling	2,87	RAB5C,FLNA,DNM2,COPG1
	Citrulline Biosynthesis	2,80	OAT,ALDH18A1
	14-3-3-mediated Signaling	2,71	YWHAG,YWHAH,YWHAE,TUBB6,YWHAB
	Purine Nucleotides De Novo Biosynthesis II	2,62	ADSL,IMPDH2
	Sirtuin Signaling Pathway	2,43	PRKDC,ATP5F1C,TIMM44,TIMM50,ACLY,SLC25A5,VDAC2
	Pyrimidine Ribonucleotides De Novo Biosynthesis	2,41	CAD,CTPS1,DDX3X
	ATM Signaling	2,36	USP7,PPP2R1A,PPP2R2A,SMC1A
	Superpathway of Citrulline Metabolism	2,35	OAT,ALDH18A1

	Dopamine-DARPP32 Feedback in cAMP Signaling	2,29	PPP2R1A,CDK5,PPP2R2A,ATP2A2,PPP3CA
	IGF-1 Signaling	2,24	YWHAG,YWHAH,YWHAE,YWHAB
	Protein Kinase A Signaling	2,23	YWHAG,YWHAH,YWHAE,YWHAB,FLNA,RACK1,TIMM50,PPP3CA
	Role of CHK Proteins in Cell Cycle Checkpoint Control	2,17	PCNA,PPP2R1A,PPP2R2A
	Acetyl-CoA Biosynthesis III (from Citrate)	2,17	ACLY
	Huntington's Disease Signaling	2,15	ATP5F1C,CDK5,CLTC,RACK1,HSPA9,DNM2
	Induction of Apoptosis by HIV1	2,09	SLC25A13,SLC25A3,SLC25A5
	Mitotic Roles of Polo-Like Kinase	2,00	PPP2R1A,PPP2R2A,SMC1A
	ILK Signaling	1,96	PPP2R1A,CFL1,PPP2R2A,FLNA,IRS4
	Cysteine Biosynthesis III (mammalia)	1,95	CBS/CBSL,PRMT5
	Remodeling of Epithelial Adherens Junctions	1,94	RAB5C,TUBB6,DNM2
	Clathrin-mediated Endocytosis Signaling	1,88	RAB5C,CLTC,USP9X,PPP3CA,DNM2
	Palmitate Biosynthesis I (Animals)	1,87	FASN
	Uridine-5'-phosphate Biosynthesis	1,87	CAD
	Fatty Acid Biosynthesis Initiation II	1,87	FASN
	Cysteine Biosynthesis/Homocysteine Degradation	1,87	CBS/CBSL
	Cardiac β-adrenergic Signaling	1,81	PPP2R1A,PPP2R2A,RACK1,ATP2A2
	Inosine-5'-phosphate Biosynthesis II	1,70	ADSL
	Biotin-carboxyl Carrier Protein Assembly	1,70	ACACA
	Cell Cycle Regulation by BTG Family Proteins	1,59	PPP2R1A,PPP2R2A
	Superpathway of Methionine Degradation	1,59	CBS/CBSL,PRMT5
	Arginine Degradation I (Arginase Pathway)	1,57	OAT
	Salvage Pathways of Pyrimidine Ribonucleotides	1,55	PYCR3,CDK5,ARAF
	Mitochondrial Dysfunction	1,54	ATP5F1C,ATP5PO,AIFM1,VDAC2
	CTLA4 Signaling in Cytotoxic T Lymphocytes	1,53	PPP2R1A,PPP2R2A,CLTC
	CDK5 Signaling	1,53	PPP2R1A,CDK5,PPP2R2A
	Tetrahydrofolate Salvage from 5,10-methenyltetrahydrofolate	1,48	MTHFD1L

	Folate Polyglutamylation	1,48	MTHFD1L
	Protein Ubiquitination Pathway	1,47	USP7,USP9X,HSPA9,DNAJA1,PSMC3
	Stearate Biosynthesis I (Animals)	1,45	FASN,SLC27A4
	Pyrimidine Ribonucleotides Interconversion	1,43	CTPS1,DDX3X
	Virus Entry via Endocytic Pathways	1,42	FLNA,CLTC,DNM2
	p53 Signaling	1,41	PRKDC,PCNA,COQ8A
	Arginine Biosynthesis IV	1,40	OAT
	Acetyl-CoA Biosynthesis I (Pyruvate Dehydrogenase Complex)	1,33	PDHB
	3-phosphoinositide Biosynthesis	1,32	ATP1A1,PGAM5,PI4KA,PPP3CA
NUCLEAR EXTRACTS	Mitotic Roles of Polo-Like Kinase	2,66	SMC3,SMC1A
	ATM Signaling	2,33	SMC3,SMC1A
	Estrogen Receptor Signaling	2,10	PRKDC,PHB2
	14-3-3-mediated Signaling	2,08	VIM,TUBA1C
	DNA Double-Strand Break Repair by Non-Homologous End Joining	1,84	PRKDC
	Granzyme B Signaling	1,78	PRKDC
	RAN Signaling	1,75	RANBP2
	ILK Signaling	1,75	IRS4,VIM
	Tumoricidal Function of Hepatic Natural Killer Cells	1,60	AIFM1
	Role of JAK2 in Hormone-like Cytokine Signaling	1,46	IRS4
	Sirtuin Signaling Pathway	1,43	PRKDC,TUBA1C
	Pyrimidine Ribonucleotides Interconversion	1,34	DDX3X
	Pyrimidine Ribonucleotides De Novo Biosynthesis	1,32	DDX3X

Supplementary Table 5. List of antibodies for the validation of selected ANKRD55 interacting partners. Abbreviations: WB, western blot and IF, immunofluoresce.

Antibody	Catalog Number	Host	Applications
TUBB	66240-1-Ig Proteintech	Mouse monoclonal CloneNo.: 1D4A4	WB and IF in combination with FLAG rabbit for recombinant ANKRD55
VIM	60330-1-Ig Proteintech	Mouse monoclonal CloneNo.: 3H9D1	WB and IF in combination with FLAG rabbit for recombinant ANKRD55
14-3-3 (pan)	8312 Cell Signaling	Rabbit polyclonal	WB and IF in combination with FLAG rabbit and mouse for recombinant ANKRD55, respectively.
RPS3	66046-1-Ig Proteintech	Mouse monoclonal CloneNo.: 2G7H4	WB and IF in combination with FLAG rabbit for recombinant ANKRD55
CLTC	66487-1-Ig Proteintech	Mouse monoclonal CloneNo.: 1B3D7	WB in combination with FLAG rabbit for recombinant ANKRD55
SMC1A	21695-1-AP Proteintech	Rabbit polyclonal	WB in combination with FLAG rabbit for recombinant ANKRD55
SMC3	5696 Cell Signaling	Rabbit monoclonal	WB in combination with FLAG rabbit for recombinant ANKRD55
PRKDC	NBP2-22128SS Novus Biologicals	Mouse monoclonal	WB in combination with FLAG rabbit for recombinant ANKRD55
HIF1AN	NB100-428 Novus Biologicals	Rabbit Polyclonal	WB in combination with FLAG rabbit for recombinant ANKRD55
FLAG	20543-1-AP Proteintech	Rabbit polyclonal	
FLAG	66008-2-Ig Proteintech	Mouse monoclonal CloneNo.: 1E7B4	

Supplementary Table 6. List of potential 14-3-3 binding sites on ANKRD55. Abbreviations: ANN, Artificial Neural Network (cut-off = 0.55); PSSM Position-Specific Scoring Matrix (cut-off = 0.80); SVM, Support Vector Machine (cut-off = 0.25).

14-3-3 binding site	Predicted phosphopeptides	ANN	PSSM	SVM	Consensus – average of the three scores
10	QATMDFsTPSV	0.314	0.299	-0.494	0.04
11	ATMDFStPSVF	0.03	-0.49	-2.021	-0.827
13	MDFSTPsVFDQ	0.145	-0.078	-0.918	-0.284
22	DQQRGDsSEEV	0.444	0.691	-0.435	0.233
23	QQRGDSsEEVD	0.206	0.186	-0.474	-0.028
29	SEEVDLtMVYQ	0.036	-0.231	-1.947	-0.714
36	MVYQAAsNGDV	0.577	0.239	-0.278	0.18
44	GDVNALtAVIR	0.141	-0.023	-0.952	-0.278
52	VIREDPsILEC	0.144	-0.0	-1.138	-0.331
59	ILECCDsEGCT	0.607	0.107	-0.203	0.17
63	CDSEGCtPLMH	0.018	-0.704	-2.349	-1.011
70	PLMHAVsGRQA	0.543	0.029	-0.006	0.189
76	SGRQADtVKLL	0.389	0.187	-0.343	0.077
96	QDAYGRtSLCL	0.034	-0.18	-1.503	-0.55
97	DAYGRTsLCLA	0.357	0.045	-0.403	0.0

102	TSLCLAtYLGW	0.088	-0.259	-1.337	-0.503
112	WLEGCVsLLRN	0.609	0.064	-0.058	0.205
135	LPLHAAtAEPD	0.279	-0.0	-0.637	-0.12
144	PDMRLLtVLLQ	0.199	0.674	-0.486	0.129
150	TVLLQQsNISE	0.04	-0.197	-1.524	-0.56
153	LQQSNIsEINH	0.131	-0.072	-0.871	-0.271
164	QDNEGMtPLHW	0.016	-0.492	-2.028	-0.835
178	HNQPQHtQMLL	0.141	0.065	-0.985	-0.259
189	KKGADPtLVDK	0.392	-0.039	-0.383	-0.01
197	VDKDFKtALHW	0.108	-0.114	-1.083	-0.363
205	LHWAVQsGNRI	0.184	-0.008	-0.851	-0.225
212	GNRILCsIILS	0.208	0.098	-0.82	-0.172
216	LCSIILsHHQG	0.204	-0.045	-0.716	-0.186
222	SHHQGPsIINY	0.048	-0.243	-1.723	-0.639
230	INYDDEsGKTC	0.065	-0.39	-1.526	-0.617
233	DDESGKtCVHI	0.138	-0.145	-0.843	-0.283
244	AAAAGFsDIIH	0.098	-0.27	-1.256	-0.476
267	LDVDDRtPLHW	0.019	-0.557	-2.135	-0.891
283	KAECVQsLLEL	0.342	-0.002	-0.609	-0.089
291	LELGMDsNLRD	0.099	-0.336	-1.465	-0.567
299	LRDINEsTPLA	0.503	0.584	-0.063	0.341
300	RDINEStPLAY	0.035	-0.397	-1.677	-0.68
311	ALYCGHtACVK	0.233	0.001	-1.003	-0.256
318	ACVKLLsQESR	0.091	-0.166	-1.56	-0.545
321	KLLSQEsRTEP	0.118	-0.242	-1.424	-0.516
323	LSQESRtEPTR	0.609	0.691	0.236	0.512
326	ESRTEPtRPPP	0.383	0.301	-0.481	0.068
331	PTRPPPsQSSR	0.126	-0.144	-1.152	-0.39
333	RPPPSQsSRPQ	0.139	-0.122	-0.943	-0.309
334	PPPSQSsRPQK	0.174	0.082	-0.937	-0.227
366	AHQKDPsRDRY	0.306	-0.156	-1.196	-0.349
375	RYREEDtSEVN	0.258	0.152	-1.059	-0.216
376	YREEDTsEVND	0.217	0.263	-0.629	-0.05
383	EVNDIItTFDS	0.167	-0.028	-0.809	-0.223
384	VNDIITtFDSI	0.168	-0.05	-0.474	-0.118
387	IITTFDsIVGT	0.15	-0.22	-0.992	-0.354
391	FDSIVGtNCQE	0.056	-0.355	-1.697	-0.665
410	VEFKKKtSDNS	0.248	0.003	-0.822	-0.19
411	EFKKKTsDNSK	0.258	-0.011	-0.426	-0.06
414	KKTSDNsKYLL	0.311	0.03	-0.529	-0.062
434	GLPPIRtQSLP	0.071	-0.193	-1.485	-0.536
436	PPIRTQsLPPI	0.917	1.494	1.525	1.312
441	QSLPPItLGNN	0.313	-0.052	-0.429	-0.056
448	LGNNFLtASHR	0.049	-0.23	-1.341	-0.508
450	NNFLTAsHRAT	0.201	-0.065	-0.708	-0.191
454	TASHRAtSHAG	0.367	0.076	-0.509	-0.022
455	ASHRATsHAGL	0.569	0.733	0.228	0.51

460	TSHAGLsSAPH	0.018	-0.594	-2.171	-0.916
461	SHAGLSsAPHH	0.1	0.127	-1.307	-0.36
470	HHMAQRsQKSR	0.04	-0.127	-1.801	-0.629
473	AQRSQKsRSEQ	0.174	-0.029	-0.931	-0.262
475	RSQKSRsEQDL	0.715	0.46	0.213	0.463
484	DLLNNRtGCQM	0.174	-0.163	-0.794	-0.261
496	LDNPWKsDSNQ	0.115	-0.138	-1.207	-0.41
498	NPWKSDsNQVF	0.314	0.214	-0.59	-0.021
503	DSNQVFsYKVW	0.154	-0.111	-1.025	-0.327
508	FSYKVWtVSSS	0.13	-0.129	-1.107	-0.369
510	YKVWTVsSSDK	0.336	-0.027	-0.44	-0.043
511	KVWTVSsSDKL	0.076	-0.036	-1.299	-0.42
512	VWTVSSsDKLL	0.193	0.168	-0.595	-0.078
521	LLDRLLsVRPG	0.45	0.713	0.067	0.41
530	PGHQEVsVPPH	0.485	0.241	-0.251	0.158
542	RHLHNPsSGQN	0.227	-0.09	-0.867	-0.244
543	HLHNPSsGQNF	0.064	-0.102	-0.985	-0.341
551	QNFQHLsPNRH	0.052	-0.568	-1.567	-0.695
563	IRDLPFtRNNL	0.229	0.136	-0.648	-0.094
577	PDQKFLsGEPL	0.109	-0.137	-1.059	-0.363
583	SGEPLRtNRVL	0.055	-0.281	-1.598	-0.608
592	VLPAIPsQRRH	0.102	-0.44	-1.49	-0.609
597	PSQRRHsTAAE	0.846	1.106	0.778	0.91
598	SQRRHStAAEE	0.635	0.964	0.134	0.577
603	STAAEEsEHSA	0.107	-0.217	-1.398	-0.503
606	AEESEHsANPT	0.177	-0.123	-1.123	-0.356
610	EHSANPtSDEN	0.155	-0.15	-1.096	-0.364
611	HSANPTsDEN-	0.122	-0.034	-0.822	-0.245

Supplementary Table 7. List of identified ANKRD55-interacting proteins in protein-protein databases

Gene Symbol	Accession	Protein name	Resources	Specie
FHL2	Q14192	Four and a half LIM domains protein 2	HuRI (HI-III)	*Homo sapiens*
HIF1AN	Q9NWT6	Hypoxia-inducible factor 1-alpha inhibitor	BioGrid	
IFT52	Q9Y366	Intraflagellar transport protein 52 homolog	BioGrid	
ANKS1A	Q49AR9	Ankyrin repeat and SAM domain-containing protein 1A	HuRI (HI-III)	
PTK6	Q13882	Protein-tyrosine kinase 6	HuRI (HI-III)	
HSF2BP	O75031	Heat shock factor 2-binding protein	HuRI (HI-III)	
BANP	Q8N9N5	Protein BANP	HuRI (HI-III)	
CABP5	Q9NP86	Calcium-binding protein 5	HuRI (HI-III)	
CLIC3	O95833	Chloride intracellular channel protein 3	HuRI (HI-III)	
ABI2	Q9NYB9	Abl interactor 2	HuRI (HI-III)	
BOLA2	A0A087WZT3	BolA-like protein 2	HuRI (HI-III)	

HGS	O14964	Hepatocyte growth factor-regulated tyrosine kinase substrate	HuRI (HI-III)
FRS3	O43559	Fibroblast growth factor receptor substrate 3	HuRI (HI-III)
NCK2	O43639	Cytoplasmic protein NCK2	HuRI (HI-III)
TRIM23	P36406	E3 ubiquitin-protein ligase TRIM23	HuRI (HI-III)
TRIP6	Q15654	Thyroid receptor-interacting protein 6	BioGrid, HuRI (HI-III, unpublished GS test space)
NOXA1	Q86UR1	NADPH oxidase activator 1	HuRI (HI-III)
TTC30A	Q86WT1	Tetratricopeptide repeat protein 30A	BioGrid
CFAP206	Q8IYR0	Cilia- and flagella-associated protein 206	HuRI (HI-III)
TTC30B	Q8N4P2	Tetratricopeptide repeat protein 30B	BioGrid
TEKT4	Q8WW24	Tektin-4	HuRI (HI-III)
ZMYND19	Q96E35	Zinc finger MYND domain-containing protein 19	HuRI (HI-III)
IFT74	Q96LB3	Intraflagellar transport protein 74 homolog	BioGrid
PFDN5	Q99471	Prefoldin subunit 5	HuRI (HI-III)
BACH2	Q9BYV9	Transcription regulator protein BACH2	HuRI (HI-III)
IFT46	Q9NQC8	Intraflagellar transport protein 46 homolog	BioGrid
ZSCAN1	Q8NBB4	Zinc finger and SCAN domain-containing protein 1	BioGrid
YWHAZ	P63104	14-3-3 protein zeta/delta	BioGrid,
MIB1	Q86YT6	E3 ubiquitin-protein ligase MIB1	BioGrid
YWHAG	P61981	14-3-3 protein gamma	BioGrid
RABGEF1	Q9UJ41	Rab5 GDP/GTP exchange factor	HuRI (HI-III)
LEF1	Q9UJU2	Lymphoid enhancer-binding factor 1	HuRI (HI-III, unpublished GS test space)
SUFU	Q9UMX1	Suppressor of fused homolog	HuRI (HI-III, unpublished GS test space)
CARD9	Q9H257	Caspase recruitment domain-containing protein 9	HuRI (HI-III, unpublished GS test space)
DEUP1	Q05D60	Deuterosome assembly protein 1	HuRI (HI-III, unpublished GS test space)
LRP2BP	P98164	Low-density lipoprotein receptor-related protein 2	HuRI (HI-III, unpublished GS test space)
SF3A3	Q12874	Splicing factor 3A subunit 3	HuRI (HI-III, unpublished GS test space)
TCAP	O15273	Telethonin	HuRI (HI-III, unpublished GS test space)
FANCL	Q9NW38	E3 ubiquitin-protein ligase FANCL	HuRI (HI-III, unpublished GS test space)
CIAO1	O76071	Probable cytosolic iron-sulfur protein assembly protein CIAO1	HuRI (HI-III, unpublished GS test space)
YWHAE	P62258	14-3-3 protein epsilon	BioGrid

YWHAQ	P27348	14-3-3 protein theta	BioGrid
RPS3	P23396	40S ribosomal protein S3	STRING
CNOT6L	Q96LI5	CCR4-NOT transcription complex subunit 6-like	STRING
LRKK2	Q5S007	Leucine-rich repeat serine/threonine-protein kinase 2	STRING
NNT	Q13423	NAD(P) transhydrogenase, mitochondrial	STRING
ASH1L	Q9NR48	Histone-lysine N-methyltransferase ASH1L	STRING
PHLPP1	O60346	PH domain leucine-rich repeat-containing protein phosphatase 1	STRING
PHLPP2	Q6ZVD8	PH domain leucine-rich repeat-containing protein phosphatase 2	STRING
LRGUK	Q96M69	Leucine-rich repeat and guanylate kinase domain-containing protein	STRING
GART	P22102	Trifunctional purine biosynthetic protein adenosine-3	STRING
ACLY	P53396	ATP-citrate synthase	STRING
Nnt	Q61941	NAD(P) transhydrogenase, mitochondrial	STRING
Rps3	P62908	40S ribosomal protein S3	STRING
Gart	Q64737	Trifunctional purine biosynthetic protein adenosine-3	STRING
Phlpp2	Q8BXA7	PH domain leucine-rich repeat-containing protein phosphatase 2	STRING
Nos	Q9Z0J4	Nitric oxide synthase, brain	STRING
Hsp90aa1	P07901	Heat shock protein HSP 90-alpha	STRING
Hsp90ab1	P11499	Heat shock protein HSP 90-beta	STRING
Phlpp1	Q8CHE4	PH domain leucine-rich repeat-containing protein phosphatase 1	STRING
Lrrk1	Q3UHC2	Leucine-rich repeat serine/threonine-protein kinase 1	STRING
Lrrk2	Q5S006	Leucine-rich repeat serine/threonine-protein kinase 2	STRING

Mus musculus

Supplementary Table 8. ***In silico*** **analysis for DNA, RNA and nucleotides binding prediction.** The first column displays the ANKRD55 isoform 001 protein sequence, which is followed by the results of predictions for 7 ligand types: DNA, RNA, ATP, ADP, AMP, GTP, and GDP. Predictions for each ligand type include two columns that provide annotation of predicted binding (each ligand is annotated as either binding (B or 1) or non-binding (N or 0)) and prediction scores which estimate probability of binding to a given ligand. Residues with the probability > 0.4727 for the prediction of DNA-binding and > 0.1493 for the prediction of RNA-binding are annotated as putative DNA- and RNA-binding residues, respectively.

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
M	0	0.1894	0	0.0470	N	0.024	N	0.016	N	0.022	N	0.017	N	0.016
M	0	0.1969	0	0.0485	N	0.028	N	0.019	N	0.020	N	0.020	N	0.019
R	0	0.3463	0	0.0409	N	0.031	N	0.031	N	0.048	N	0.016	N	0.029
Q	0	0.2369	0	0.0418	N	0.028	N	0.025	N	0.027	N	0.014	N	0.016
A	0	0.1557	0	0.0399	N	0.043	N	0.033	N	0.033	N	0.060	N	0.016
T	0	0.2356	0	0.0417	N	0.034	N	0.033	N	0.030	N	0.040	N	0.016
M	0	0.1887	0	0.0406	N	0.032	N	0.024	N	0.031	N	0.022	N	0.016
D	0	0.2516	0	0.0420	N	0.039	N	0.025	N	0.042	N	0.038	N	0.018
F	0	0.3632	0	0.0414	N	0.026	N	0.015	N	0.027	N	0.015	N	0.015
S	1	0.5018	0	0.0420	N	0.023	N	0.027	N	0.033	N	0.011	N	0.019
T	0	0.4313	0	0.0416	N	0.031	N	0.034	N	0.030	N	0.030	N	0.017
P	0	0.2912	0	0.0410	N	0.037	N	0.027	N	0.027	N	0.015	N	0.018
S	0	0.3016	0	0.0405	N	0.025	N	0.023	N	0.045	N	0.029	N	0.013
V	0	0.1961	0	0.0393	N	0.020	N	0.026	N	0.018	N	0.017	N	0.017
F	0	0.2466	0	0.0395	N	0.032	N	0.020	N	0.021	N	0.012	N	0.015
D	0	0.3975	0	0.0406	N	0.025	N	0.016	N	0.020	N	0.015	N	0.014
Q	0	0.3233	0	0.0402	N	0.026	N	0.016	N	0.019	N	0.014	N	0.019
Q	0	0.2170	0	0.0435	N	0.022	N	0.014	N	0.022	N	0.026	N	0.013
R	0	0.4190	0	0.0450	N	0.023	N	0.026	N	0.030	N	0.025	N	0.017
G	0	0.2052	0	0.0444	N	0.030	N	0.035	N	0.026	N	0.027	N	0.016
D	0	0.2056	0	0.0441	N	0.026	N	0.034	N	0.033	N	0.033	N	0.019

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
S	0	0.1915	0	0.0436	N	0.029	N	0.028	N	0.031	N	0.021	N	0.021
S	0	0.1671	0	0.0431	N	0.026	N	0.024	N	0.034	N	0.019	N	0.019
E	0	0.1322	0	0.0425	N	0.031	N	0.027	N	0.031	N	0.024	N	0.017
E	0	0.1136	0	0.0424	N	0.036	N	0.032	N	0.028	N	0.023	N	0.020
V	0	0.1281	0	0.0424	N	0.027	N	0.035	N	0.026	N	0.019	N	0.016
D	0	0.1236	0	0.0579	N	0.029	N	0.020	N	0.031	N	0.049	N	0.021
L	0	0.1023	0	0.0567	N	0.063	N	0.032	N	0.021	N	0.014	N	0.014
T	0	0.1676	0	0.0583	N	0.035	N	0.024	N	0.034	B	0.077	N	0.021
M	0	0.1131	0	0.0481	N	0.034	N	0.059	N	0.047	N	0.057	N	0.023
V	0	0.0916	0	0.0473	N	0.036	N	0.039	N	0.038	N	0.033	N	0.026
Y	0	0.2154	0	0.0501	N	0.056	N	0.046	N	0.044	N	0.060	N	0.029
Q	0	0.1534	0	0.0493	N	0.050	N	0.037	N	0.030	N	0.021	N	0.022
A	0	0.1287	0	0.0480	N	0.025	N	0.028	N	0.037	N	0.036	N	0.022
A	0	0.1304	0	0.0487	N	0.027	N	0.022	N	0.035	N	0.031	N	0.020
S	0	0.2208	0	0.0504	B	0.135	N	0.027	N	0.027	N	0.026	N	0.021
N	0	0.3360	0	0.0507	B	0.129	N	0.027	N	0.023	N	0.028	N	0.018
G	0	0.3106	0	0.0521	N	0.029	N	0.025	N	0.033	N	0.029	N	0.021
D	0	0.2663	0	0.0515	N	0.027	N	0.026	N	0.028	N	0.028	N	0.017
V	0	0.2056	0	0.0489	N	0.029	N	0.021	N	0.027	N	0.023	N	0.019
N	0	0.2386	0	0.0500	N	0.025	N	0.020	N	0.028	N	0.023	N	0.016
A	0	0.1345	0	0.0487	N	0.028	N	0.021	N	0.028	N	0.021	N	0.016
L	0	0.0971	0	0.0481	N	0.029	N	0.029	N	0.034	N	0.022	N	0.016
T	0	0.1451	0	0.0507	N	0.021	N	0.028	N	0.026	N	0.032	N	0.020
A	0	0.1265	0	0.0519	N	0.023	N	0.022	N	0.027	N	0.017	N	0.016
V	0	0.0894	0	0.0534	N	0.019	N	0.012	N	0.022	N	0.018	N	0.011

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
I	0	0.0901	0	0.0551	N	0.025	N	0.019	N	0.014	N	0.020	N	0.010
R	0	0.1825	0	0.0576	N	0.019	N	0.020	N	0.021	N	0.015	N	0.017
E	0	0.1203	0	0.0636	N	0.029	N	0.021	N	0.026	N	0.026	N	0.015
D	0	0.1098	0	0.0634	N	0.028	N	0.022	N	0.021	N	0.019	N	0.017
P	0	0.1034	0	0.0626	N	0.029	N	0.021	N	0.024	N	0.027	N	0.018
S	0	0.1117	0	0.0630	N	0.031	N	0.026	N	0.027	N	0.026	N	0.019
I	0	0.0899	0	0.0615	N	0.033	N	0.027	N	0.029	N	0.020	N	0.020
L	0	0.0876	0	0.0606	N	0.029	N	0.025	N	0.025	N	0.019	N	0.019
E	0	0.1117	0	0.0630	N	0.031	N	0.032	N	0.030	N	0.026	N	0.020
C	0	0.0987	0	0.0511	N	0.029	N	0.035	N	0.027	N	0.021	N	0.019
C	0	0.1075	0	0.0513	N	0.041	N	0.023	N	0.033	N	0.018	N	0.018
D	0	0.1339	0	0.0533	N	0.030	N	0.027	N	0.031	N	0.021	N	0.023
S	0	0.1548	0	0.0527	N	0.035	N	0.027	N	0.025	N	0.022	N	0.016
E	0	0.1739	0	0.0544	N	0.029	N	0.032	N	0.030	N	0.021	N	0.017
G	0	0.1991	0	0.0538	N	0.027	N	0.034	N	0.033	N	0.068	N	0.026
C	0	0.1813	0	0.0520	N	0.061	N	0.036	N	0.026	N	0.020	N	0.018
T	0	0.2997	0	0.0544	N	0.033	N	0.038	N	0.036	N	0.050	N	0.030
P	0	0.1264	0	0.0510	N	0.061	B	0.153	N	0.036	N	0.053	N	0.028
L	0	0.0956	0	0.0370	N	0.031	N	0.062	N	0.036	N	0.034	N	0.031
M	0	0.2082	0	0.0408	N	0.074	N	0.054	N	0.038	N	0.066	N	0.030
H	0	0.2133	0	0.0387	N	0.046	N	0.038	N	0.032	N	0.028	N	0.023
A	0	0.1353	0	0.0377	N	0.033	N	0.032	N	0.037	N	0.036	N	0.022
V	0	0.1391	0	0.0380	N	0.028	N	0.020	N	0.035	N	0.028	N	0.020
S	0	0.2713	0	0.0396	N	0.094	N	0.029	N	0.029	N	0.031	N	0.021
G	0	0.3358	0	0.0404	B	0.110	N	0.025	N	0.024	N	0.031	N	0.018

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
R	1	0.5380	0	0.0407	N	0.029	N	0.024	N	0.035	N	0.027	N	0.021
Q	1	0.4985	0	0.0408	N	0.029	N	0.029	N	0.029	N	0.026	N	0.018
A	0	0.3152	0	0.0389	N	0.030	N	0.024	N	0.027	N	0.023	N	0.019
D	0	0.2738	0	0.0392	N	0.025	N	0.020	N	0.030	N	0.024	N	0.017
T	0	0.1578	0	0.0384	N	0.029	N	0.035	N	0.031	N	0.025	N	0.018
V	0	0.1156	0	0.0373	N	0.029	N	0.031	N	0.033	N	0.023	N	0.018
K	0	0.2670	0	0.0394	N	0.023	N	0.029	N	0.031	N	0.032	N	0.023
L	0	0.1247	0	0.0381	N	0.026	N	0.023	N	0.028	N	0.018	N	0.017
L	0	0.0906	0	0.0389	N	0.029	N	0.026	N	0.027	N	0.027	N	0.014
L	0	0.1002	0	0.0430	N	0.025	N	0.018	N	0.021	N	0.033	N	0.012
K	0	0.2469	0	0.0457	N	0.025	N	0.029	N	0.028	N	0.019	N	0.018
M	0	0.1848	0	0.0449	N	0.031	N	0.030	N	0.025	N	0.031	N	0.020
G	0	0.2655	0	0.0461	N	0.030	N	0.023	N	0.022	N	0.037	N	0.018
A	0	0.1703	0	0.0439	N	0.030	N	0.029	N	0.024	N	0.028	N	0.018
N	0	0.3425	0	0.0396	N	0.024	N	0.024	N	0.031	N	0.038	N	0.024
I	0	0.1781	0	0.0370	N	0.030	N	0.028	N	0.032	N	0.027	N	0.021
N	0	0.4023	0	0.0394	N	0.032	N	0.034	N	0.036	N	0.031	N	0.019
M	0	0.2137	0	0.0374	N	0.030	N	0.029	N	0.030	N	0.032	N	0.023
Q	0	0.3082	0	0.0392	N	0.034	N	0.026	N	0.033	N	0.024	N	0.019
D	0	0.3498	0	0.0398	N	0.031	N	0.020	N	0.030	N	0.025	N	0.019
A	0	0.4037	0	0.0388	N	0.033	N	0.030	N	0.026	N	0.022	N	0.021
Y	1	0.5458	0	0.0392	N	0.030	N	0.039	N	0.027	N	0.024	N	0.016
G	0	0.3100	0	0.0390	N	0.030	N	0.059	N	0.033	N	0.060	N	0.023
R	0	0.4123	0	0.0399	N	0.049	N	0.040	N	0.026	N	0.019	N	0.019
T	0	0.4272	0	0.0414	N	0.033	N	0.038	N	0.037	N	0.056	N	0.032

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
S	0	0.1567	0	0.0383	N	0.070	B	0.162	N	0.035	N	0.052	N	0.028
L	0	0.1061	0	0.0364	N	0.031	N	0.063	N	0.035	N	0.034	N	0.029
C	0	0.2287	0	0.0398	N	0.088	N	0.048	N	0.038	B	0.075	N	0.033
L	0	0.1319	0	0.0372	N	0.046	N	0.033	N	0.031	N	0.028	N	0.023
A	0	0.1190	0	0.0427	N	0.032	N	0.034	N	0.038	N	0.036	N	0.022
T	0	0.1240	0	0.0444	N	0.030	N	0.025	N	0.034	N	0.029	N	0.019
Y	0	0.2464	0	0.0449	B	0.100	N	0.028	N	0.027	N	0.027	N	0.020
L	0	0.1622	0	0.0566	N	0.078	N	0.026	N	0.025	N	0.031	N	0.019
G	0	0.2602	0	0.0583	N	0.027	N	0.027	N	0.045	N	0.030	N	0.020
W	0	0.2341	0	0.0582	N	0.023	N	0.024	N	0.028	N	0.026	N	0.017
L	0	0.1342	0	0.0553	N	0.027	N	0.020	N	0.025	N	0.025	N	0.019
E	0	0.1658	0	0.0485	N	0.024	N	0.018	N	0.030	N	0.021	N	0.016
G	0	0.1505	0	0.0475	N	0.027	N	0.031	N	0.029	N	0.023	N	0.016
C	0	0.1127	0	0.0467	N	0.026	N	0.028	N	0.030	N	0.023	N	0.017
V	0	0.1611	0	0.0475	N	0.023	N	0.028	N	0.030	N	0.029	N	0.022
S	0	0.1750	0	0.0368	N	0.025	N	0.021	N	0.026	N	0.019	N	0.016
L	0	0.1192	0	0.0350	N	0.026	N	0.020	N	0.025	N	0.025	N	0.014
L	0	0.1231	0	0.0351	N	0.026	N	0.017	N	0.015	N	0.036	N	0.012
R	0	0.3640	0	0.0372	N	0.024	N	0.024	N	0.029	N	0.019	N	0.017
N	0	0.2973	0	0.0384	N	0.031	N	0.030	N	0.027	N	0.033	N	0.019
G	0	0.2929	0	0.0389	N	0.032	N	0.024	N	0.019	N	0.033	N	0.017
A	0	0.2073	0	0.0369	N	0.031	N	0.027	N	0.029	N	0.030	N	0.020
K	1	0.5568	0	0.0428	N	0.026	N	0.023	N	0.030	N	0.032	N	0.021
H	0	0.2382	0	0.0386	N	0.031	N	0.029	N	0.033	N	0.030	N	0.021
N	0	0.3232	0	0.0408	N	0.031	N	0.032	N	0.031	N	0.032	N	0.019

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
I	0	0.1944	0	0.0382	N	0.030	N	0.031	N	0.029	N	0.032	N	0.026
P	0	0.2703	0	0.0403	N	0.037	N	0.029	N	0.031	N	0.021	N	0.018
D	0	0.3733	0	0.0408	N	0.028	N	0.023	N	0.031	N	0.027	N	0.017
K	0	0.4029	0	0.0403	N	0.031	N	0.028	N	0.027	N	0.021	N	0.020
N	0	0.2414	0	0.0391	N	0.028	N	0.040	N	0.027	N	0.022	N	0.015
G	0	0.2265	0	0.0486	N	0.029	N	0.038	N	0.033	N	0.059	N	0.024
R	0	0.3035	0	0.0482	N	0.045	N	0.040	N	0.025	N	0.022	N	0.020
L	0	0.2770	0	0.0478	N	0.032	N	0.038	N	0.036	N	0.058	N	0.029
P	0	0.1219	0	0.0470	N	0.054	B	0.239	N	0.034	N	0.063	N	0.028
L	0	0.0956	0	0.0444	N	0.038	N	0.061	N	0.032	N	0.037	N	0.028
H	0	0.1990	0	0.0511	N	0.055	N	0.051	N	0.042	N	0.068	N	0.032
A	0	0.1101	0	0.0537	N	0.044	N	0.040	N	0.035	N	0.035	N	0.023
A	0	0.0985	0	0.0518	N	0.031	N	0.038	N	0.039	N	0.036	N	0.022
T	0	0.0990	0	0.0523	N	0.031	N	0.019	N	0.034	N	0.027	N	0.020
A	0	0.1375	0	0.0553	B	0.185	N	0.029	N	0.027	N	0.029	N	0.021
E	0	0.1561	0	0.0624	B	0.189	N	0.027	N	0.023	N	0.031	N	0.019
P	0	0.1352	0	0.0638	N	0.028	N	0.024	N	0.032	N	0.032	N	0.021
D	0	0.1376	0	0.0628	N	0.025	N	0.025	N	0.027	N	0.028	N	0.016
M	0	0.1072	0	0.0664	N	0.025	N	0.021	N	0.021	N	0.023	N	0.018
R	0	0.1442	0	0.0678	N	0.026	N	0.020	N	0.028	N	0.022	N	0.015
L	0	0.0960	0	0.0660	N	0.028	N	0.028	N	0.033	N	0.024	N	0.018
L	0	0.0891	0	0.0653	N	0.030	N	0.032	N	0.032	N	0.023	N	0.017
T	0	0.1114	0	0.0672	N	0.022	N	0.029	N	0.030	N	0.029	N	0.021
V	0	0.0990	0	0.0673	N	0.022	N	0.017	N	0.023	N	0.018	N	0.014
L	0	0.0843	0	0.0659	N	0.032	N	0.017	N	0.025	N	0.023	N	0.015

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
L	0	0.0898	0	0.0648	N	0.025	N	0.019	N	0.017	N	0.018	N	0.011
Q	0	0.1229	0	0.0680	N	0.026	N	0.027	N	0.020	N	0.020	N	0.018
Q	0	0.1376	0	0.0641	N	0.035	N	0.028	N	0.027	N	0.021	N	0.019
S	0	0.1297	0	0.0643	N	0.033	N	0.033	N	0.026	N	0.033	N	0.022
N	0	0.1313	0	0.0634	N	0.029	N	0.029	N	0.023	N	0.021	N	0.018
I	0	0.1386	0	0.0592	N	0.032	N	0.023	N	0.030	N	0.026	N	0.023
S	0	0.1446	0	0.0561	N	0.035	N	0.016	N	0.037	N	0.042	N	0.023
E	0	0.1565	0	0.0564	N	0.030	N	0.032	N	0.033	N	0.052	N	0.022
I	0	0.1058	0	0.0546	N	0.034	N	0.047	N	0.029	N	0.027	N	0.021
N	0	0.1839	0	0.0572	N	0.035	N	0.032	N	0.031	N	0.047	N	0.017
H	0	0.1660	0	0.0564	N	0.029	N	0.030	N	0.027	N	0.021	N	0.019
Q	0	0.1730	0	0.0576	N	0.041	N	0.035	N	0.033	N	0.026	N	0.019
D	0	0.2142	0	0.0600	N	0.029	N	0.023	N	0.029	N	0.028	N	0.016
N	0	0.2499	0	0.0597	N	0.029	N	0.029	N	0.028	N	0.021	N	0.021
E	0	0.2075	0	0.0587	N	0.028	N	0.037	N	0.025	N	0.022	N	0.015
G	0	0.1777	0	0.0574	N	0.029	N	0.040	N	0.033	N	0.071	N	0.020
M	0	0.1794	0	0.0570	N	0.050	N	0.045	N	0.027	N	0.019	N	0.020
T	0	0.2285	0	0.0577	N	0.031	N	0.038	N	0.036	N	0.051	N	0.030
P	0	0.1058	0	0.0488	N	0.059	B	0.275	N	0.034	N	0.073	N	0.029
L	0	0.0881	0	0.0473	N	0.041	N	0.069	N	0.033	N	0.038	N	0.030
H	0	0.1554	0	0.0485	N	0.055	N	0.061	N	0.045	N	0.062	N	0.045
W	0	0.1102	0	0.0469	N	0.037	N	0.035	N	0.037	N	0.040	N	0.023
A	0	0.0988	0	0.0467	N	0.033	N	0.038	N	0.037	N	0.036	N	0.022
A	0	0.1126	0	0.0475	N	0.030	N	0.022	N	0.036	N	0.029	N	0.020
F	0	0.1937	0	0.0494	B	0.153	N	0.030	N	0.029	N	0.028	N	0.021

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
H	0	0.3408	0	0.0518	B	0.127	N	0.027	N	0.025	N	0.030	N	0.018
N	0	0.3168	0	0.0522	N	0.029	N	0.025	N	0.039	N	0.031	N	0.021
Q	0	0.3642	0	0.0516	N	0.028	N	0.028	N	0.028	N	0.026	N	0.018
P	0	0.2327	0	0.0492	N	0.028	N	0.021	N	0.025	N	0.024	N	0.019
Q	0	0.2652	0	0.0486	N	0.025	N	0.019	N	0.030	N	0.022	N	0.017
H	0	0.1839	0	0.0471	N	0.029	N	0.040	N	0.030	N	0.023	N	0.017
T	0	0.1235	0	0.0467	N	0.031	N	0.030	N	0.031	N	0.023	N	0.017
Q	0	0.1937	0	0.0485	N	0.023	N	0.028	N	0.030	N	0.032	N	0.022
M	0	0.1194	0	0.0471	N	0.026	N	0.022	N	0.027	N	0.020	N	0.017
L	0	0.0923	0	0.0458	N	0.027	N	0.023	N	0.025	N	0.027	N	0.015
L	0	0.0962	0	0.0529	N	0.028	N	0.018	N	0.018	N	0.031	N	0.012
K	0	0.1815	0	0.0557	N	0.022	N	0.025	N	0.025	N	0.021	N	0.017
K	0	0.1902	0	0.0624	N	0.030	N	0.029	N	0.027	N	0.035	N	0.018
G	0	0.1720	0	0.0628	N	0.032	N	0.023	N	0.020	N	0.032	N	0.018
A	0	0.1163	0	0.0595	N	0.031	N	0.024	N	0.024	N	0.028	N	0.018
D	0	0.2019	0	0.0621	N	0.027	N	0.025	N	0.031	N	0.033	N	0.022
P	0	0.1282	0	0.0598	N	0.030	N	0.029	N	0.033	N	0.029	N	0.021
T	0	0.1864	0	0.0609	N	0.030	N	0.032	N	0.035	N	0.031	N	0.018
L	0	0.1221	0	0.0587	N	0.025	N	0.029	N	0.027	N	0.031	N	0.023
V	0	0.1271	0	0.0596	N	0.039	N	0.032	N	0.031	N	0.019	N	0.018
D	0	0.1463	0	0.0610	N	0.029	N	0.022	N	0.029	N	0.026	N	0.020
K	0	0.2065	0	0.0566	N	0.031	N	0.031	N	0.028	N	0.018	N	0.025
D	0	0.1394	0	0.0563	N	0.026	N	0.039	N	0.026	N	0.021	N	0.015
F	0	0.1496	0	0.0558	N	0.029	N	0.034	N	0.032	N	0.058	N	0.027
K	0	0.2199	0	0.0573	N	0.052	N	0.035	N	0.024	N	0.022	N	0.018

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
T	0	0.2165	0	0.0553	N	0.033	N	0.033	N	0.036	N	0.055	N	0.032
A	0	0.1158	0	0.0523	N	0.058	B	0.251	N	0.036	N	0.060	N	0.030
L	0	0.0959	0	0.0506	N	0.037	N	0.060	N	0.034	N	0.040	N	0.030
H	0	0.2455	0	0.0555	N	0.052	N	0.053	N	0.046	B	0.075	N	0.031
W	0	0.1478	0	0.0518	N	0.038	N	0.041	N	0.033	N	0.033	N	0.023
A	0	0.1241	0	0.0508	N	0.031	N	0.037	N	0.039	N	0.036	N	0.023
V	0	0.1508	0	0.0515	N	0.033	N	0.025	N	0.035	N	0.029	N	0.020
Q	0	0.2954	0	0.0557	B	0.162	N	0.029	N	0.027	N	0.028	N	0.021
S	0	0.3059	0	0.0559	B	0.142	N	0.029	N	0.027	N	0.027	N	0.019
G	0	0.2800	0	0.0557	N	0.028	N	0.026	N	0.038	N	0.029	N	0.020
N	0	0.2753	0	0.0577	N	0.027	N	0.027	N	0.032	N	0.033	N	0.017
R	0	0.3091	0	0.0548	N	0.028	N	0.023	N	0.026	N	0.028	N	0.020
I	0	0.1900	0	0.0539	N	0.023	N	0.018	N	0.031	N	0.022	N	0.017
L	0	0.1231	0	0.0523	N	0.028	N	0.031	N	0.031	N	0.023	N	0.016
C	0	0.0994	0	0.0523	N	0.028	N	0.027	N	0.033	N	0.024	N	0.017
S	0	0.1425	0	0.0540	N	0.023	N	0.028	N	0.028	N	0.030	N	0.020
I	0	0.1061	0	0.0526	N	0.028	N	0.019	N	0.024	N	0.021	N	0.014
I	0	0.0895	0	0.0514	N	0.020	N	0.016	N	0.026	N	0.016	N	0.012
L	0	0.1014	0	0.0516	N	0.025	N	0.019	N	0.021	N	0.024	N	0.013
S	0	0.2073	0	0.0613	N	0.021	N	0.019	N	0.026	N	0.015	N	0.019
H	0	0.2351	0	0.0625	N	0.030	N	0.024	N	0.029	N	0.017	N	0.014
H	0	0.2410	0	0.0613	N	0.023	N	0.021	N	0.025	N	0.009	N	0.015
Q	0	0.1563	0	0.0591	N	0.026	N	0.026	N	0.016	N	0.020	N	0.017
G	0	0.1902	0	0.0599	N	0.026	N	0.014	N	0.021	N	0.024	N	0.014
P	0	0.1586	0	0.0593	N	0.020	N	0.020	N	0.020	N	0.021	N	0.010

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
S	0	0.1621	0	0.0586	N	0.020	N	0.036	N	0.023	N	0.028	N	0.013
I	0	0.1332	0	0.0564	N	0.016	N	0.020	N	0.022	N	0.015	N	0.014
I	0	0.1214	0	0.0562	N	0.020	N	0.034	N	0.030	N	0.022	N	0.019
N	0	0.1542	0	0.0504	N	0.023	N	0.031	N	0.036	N	0.049	N	0.015
Y	0	0.1531	0	0.0487	N	0.029	N	0.020	N	0.028	N	0.024	N	0.017
D	0	0.1525	0	0.0500	N	0.030	N	0.018	N	0.034	N	0.015	N	0.021
D	0	0.1839	0	0.0507	N	0.028	N	0.021	N	0.026	N	0.033	N	0.015
E	0	0.2100	0	0.0499	N	0.031	N	0.029	N	0.028	N	0.023	N	0.020
S	0	0.2005	0	0.0498	N	0.027	N	0.032	N	0.026	N	0.023	N	0.015
G	0	0.2168	0	0.0497	N	0.030	N	0.045	N	0.033	N	0.066	N	0.022
K	0	0.3493	0	0.0503	N	0.072	N	0.040	N	0.027	N	0.021	N	0.022
T	0	0.3346	0	0.0507	N	0.032	N	0.036	N	0.037	N	0.046	N	0.028
C	0	0.1443	0	0.0486	N	0.047	B	0.266	N	0.036	N	0.072	N	0.029
V	0	0.1014	0	0.0470	N	0.032	N	0.063	N	0.033	N	0.036	N	0.029
H	0	0.2583	0	0.0519	N	0.048	N	0.055	N	0.039	N	0.059	N	0.041
I	0	0.1254	0	0.0483	N	0.037	N	0.039	N	0.035	N	0.039	N	0.023
A	0	0.1182	0	0.0479	N	0.032	N	0.035	N	0.038	N	0.035	N	0.022
A	0	0.1240	0	0.0480	N	0.034	N	0.023	N	0.036	N	0.029	N	0.021
A	0	0.2600	0	0.0507	B	0.202	N	0.031	N	0.028	N	0.027	N	0.021
A	0	0.2557	0	0.0556	B	0.169	N	0.028	N	0.024	N	0.029	N	0.019
G	0	0.2469	0	0.0564	N	0.026	N	0.027	N	0.038	N	0.032	N	0.020
F	0	0.2060	0	0.0555	N	0.026	N	0.026	N	0.031	N	0.027	N	0.017
S	0	0.1426	0	0.0560	N	0.026	N	0.019	N	0.023	N	0.022	N	0.018
D	0	0.1317	0	0.0559	N	0.027	N	0.016	N	0.030	N	0.021	N	0.016
I	0	0.1027	0	0.0538	N	0.027	N	0.029	N	0.030	N	0.025	N	0.015

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
I	0	0.0978	0	0.0535	N	0.030	N	0.027	N	0.032	N	0.019	N	0.016
H	0	0.1550	0	0.0572	N	0.020	N	0.023	N	0.022	N	0.026	N	0.016
E	0	0.0969	0	0.0561	N	0.029	N	0.032	N	0.024	N	0.019	N	0.017
L	0	0.0833	0	0.0588	N	0.023	N	0.018	N	0.034	N	0.022	N	0.014
A	0	0.0906	0	0.0645	N	0.023	N	0.029	N	0.018	N	0.024	N	0.018
R	0	0.1488	0	0.0682	N	0.034	N	0.022	N	0.021	N	0.018	N	0.019
V	0	0.1137	0	0.0659	N	0.031	N	0.022	N	0.038	N	0.019	N	0.018
P	0	0.1161	0	0.0672	N	0.028	N	0.024	N	0.028	N	0.018	N	0.019
E	0	0.1193	0	0.0668	N	0.026	N	0.021	N	0.030	N	0.039	N	0.018
C	0	0.1020	0	0.0663	N	0.030	N	0.031	N	0.028	N	0.037	N	0.019
N	0	0.1178	0	0.0679	N	0.030	N	0.025	N	0.029	N	0.028	N	0.023
L	0	0.0906	0	0.0654	N	0.030	N	0.033	N	0.034	N	0.020	N	0.021
Q	0	0.1230	0	0.0585	N	0.031	N	0.032	N	0.034	N	0.043	N	0.018
A	0	0.1057	0	0.0548	N	0.025	N	0.036	N	0.028	N	0.044	N	0.020
L	0	0.1092	0	0.0546	N	0.037	N	0.033	N	0.033	N	0.021	N	0.018
D	0	0.1319	0	0.0562	N	0.030	N	0.023	N	0.030	N	0.029	N	0.017
V	0	0.1405	0	0.0553	N	0.028	N	0.028	N	0.026	N	0.021	N	0.021
D	0	0.1398	0	0.0560	N	0.026	N	0.034	N	0.030	N	0.029	N	0.016
D	0	0.1311	0	0.0564	N	0.029	N	0.034	N	0.033	N	0.057	N	0.028
R	0	0.2238	0	0.0567	N	0.040	N	0.037	N	0.024	N	0.025	N	0.018
T	0	0.2476	0	0.0575	N	0.035	N	0.037	N	0.036	N	0.062	N	0.030
P	0	0.1162	0	0.0558	N	0.054	B	0.210	N	0.035	N	0.059	N	0.028
L	0	0.0923	0	0.0539	N	0.051	N	0.062	N	0.038	N	0.042	N	0.034
H	0	0.2316	0	0.0559	N	0.060	N	0.056	N	0.042	N	0.067	N	0.032
W	0	0.1185	0	0.0521	N	0.040	N	0.035	N	0.040	N	0.039	N	0.023

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
A	0	0.1055	0	0.0513	N	0.034	N	0.044	N	0.039	N	0.036	N	0.022
A	0	0.1115	0	0.0524	N	0.031	N	0.022	N	0.036	N	0.027	N	0.020
A	0	0.1587	0	0.0538	B	0.145	N	0.029	N	0.026	N	0.027	N	0.021
A	0	0.1808	0	0.0541	B	0.147	N	0.025	N	0.025	N	0.032	N	0.018
G	0	0.2303	0	0.0545	N	0.028	N	0.024	N	0.035	N	0.033	N	0.021
K	0	0.2876	0	0.0552	N	0.028	N	0.028	N	0.029	N	0.028	N	0.017
A	0	0.1524	0	0.0526	N	0.030	N	0.023	N	0.025	N	0.023	N	0.019
E	0	0.1937	0	0.0532	N	0.026	N	0.020	N	0.031	N	0.023	N	0.016
C	0	0.1254	0	0.0516	N	0.030	N	0.034	N	0.031	N	0.026	N	0.017
V	0	0.1113	0	0.0513	N	0.031	N	0.032	N	0.031	N	0.026	N	0.018
Q	0	0.1492	0	0.0524	N	0.024	N	0.030	N	0.031	N	0.031	N	0.022
S	0	0.1033	0	0.0515	N	0.024	N	0.020	N	0.027	N	0.019	N	0.016
L	0	0.0860	0	0.0503	N	0.031	N	0.021	N	0.026	N	0.027	N	0.014
L	0	0.0934	0	0.0557	N	0.028	N	0.020	N	0.018	N	0.032	N	0.013
E	0	0.1376	0	0.0577	N	0.026	N	0.026	N	0.026	N	0.022	N	0.019
L	0	0.1432	0	0.0582	N	0.032	N	0.028	N	0.027	N	0.031	N	0.019
G	0	0.1418	0	0.0581	N	0.035	N	0.028	N	0.024	N	0.038	N	0.019
M	0	0.1081	0	0.0563	N	0.031	N	0.024	N	0.025	N	0.028	N	0.019
D	0	0.1613	0	0.0592	N	0.026	N	0.025	N	0.030	N	0.034	N	0.023
S	0	0.1207	0	0.0578	N	0.029	N	0.029	N	0.033	N	0.029	N	0.021
N	0	0.2208	0	0.0598	N	0.031	N	0.030	N	0.033	N	0.030	N	0.019
L	0	0.1228	0	0.0527	N	0.025	N	0.028	N	0.027	N	0.036	N	0.023
R	0	0.2000	0	0.0540	N	0.037	N	0.026	N	0.033	N	0.023	N	0.019
D	0	0.2152	0	0.0547	N	0.030	N	0.028	N	0.030	N	0.029	N	0.019
I	0	0.1912	0	0.0535	N	0.033	N	0.033	N	0.029	N	0.021	N	0.023

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
N	0	0.1906	0	0.0549	N	0.027	N	0.035	N	0.027	N	0.028	N	0.015
E	0	0.1834	0	0.0543	N	0.034	N	0.037	N	0.035	N	0.060	N	0.024
S	0	0.1688	0	0.0538	N	0.042	N	0.036	N	0.025	N	0.023	N	0.019
T	0	0.2692	0	0.0480	N	0.036	N	0.032	N	0.037	N	0.054	N	0.028
P	0	0.1340	0	0.0465	N	0.057	B	0.204	N	0.034	N	0.052	N	0.028
L	0	0.0941	0	0.0451	N	0.040	N	0.068	N	0.036	N	0.036	N	0.028
A	0	0.1777	0	0.0477	N	0.080	N	0.052	N	0.039	B	0.076	N	0.029
Y	0	0.1390	0	0.0462	N	0.044	N	0.033	N	0.032	N	0.033	N	0.023
A	0	0.1100	0	0.0456	N	0.040	N	0.034	N	0.039	N	0.035	N	0.022
L	0	0.1281	0	0.0464	N	0.031	N	0.027	N	0.035	N	0.028	N	0.019
Y	0	0.3624	0	0.0479	B	0.125	N	0.030	N	0.028	N	0.031	N	0.021
C	0	0.2689	0	0.0485	B	0.123	N	0.027	N	0.027	N	0.030	N	0.019
G	0	0.2845	0	0.0492	N	0.031	N	0.026	N	0.048	N	0.034	N	0.021
H	0	0.3635	0	0.0495	N	0.027	N	0.026	N	0.031	N	0.037	N	0.017
T	0	0.2049	0	0.0485	N	0.024	N	0.023	N	0.023	N	0.026	N	0.019
A	0	0.1871	0	0.0482	N	0.025	N	0.020	N	0.030	N	0.022	N	0.016
C	0	0.1248	0	0.0474	N	0.029	N	0.035	N	0.032	N	0.025	N	0.018
V	0	0.1313	0	0.0467	N	0.028	N	0.030	N	0.033	N	0.024	N	0.018
K	0	0.3383	0	0.0491	N	0.023	N	0.030	N	0.031	N	0.032	N	0.021
L	0	0.1102	0	0.0467	N	0.027	N	0.021	N	0.023	N	0.019	N	0.016
L	0	0.0901	0	0.0456	N	0.027	N	0.013	N	0.026	N	0.021	N	0.015
S	0	0.1215	0	0.0474	N	0.020	N	0.013	N	0.014	N	0.014	N	0.013
Q	0	0.1780	0	0.0514	N	0.022	N	0.017	N	0.030	N	0.010	N	0.019
E	0	0.1434	0	0.0501	N	0.028	N	0.020	N	0.034	N	0.022	N	0.016
S	0	0.1596	0	0.0506	N	0.028	N	0.024	N	0.014	N	0.031	N	0.017

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
R	0	0.2187	0	0.0506	N	0.032	N	0.024	N	0.023	N	0.029	N	0.018
T	0	0.1836	0	0.0600	N	0.020	N	0.017	N	0.027	N	0.021	N	0.017
E	0	0.1469	0	0.0566	N	0.028	N	0.025	N	0.030	N	0.021	N	0.019
P	0	0.1608	0	0.0610	N	0.022	N	0.025	N	0.026	N	0.022	N	0.019
T	0	0.1483	0	0.0539	N	0.024	N	0.026	N	0.030	N	0.016	N	0.018
R	0	0.2312	0	0.0558	N	0.029	N	0.024	N	0.029	N	0.014	N	0.020
P	0	0.1726	0	0.0529	N	0.021	N	0.028	N	0.026	N	0.013	N	0.029
P	0	0.1736	0	0.0513	N	0.037	N	0.030	N	0.029	N	0.016	N	0.020
P	0	0.2472	0	0.0546	N	0.020	N	0.031	N	0.028	N	0.012	N	0.026
S	0	0.2884	0	0.0577	N	0.040	N	0.027	N	0.032	N	0.030	N	0.020
Q	0	0.2388	0	0.0560	N	0.024	N	0.026	N	0.034	N	0.030	N	0.014
S	0	0.3872	0	0.0522	N	0.029	N	0.047	N	0.035	N	0.037	N	0.046
S	0	0.2420	0	0.0495	N	0.042	N	0.033	N	0.035	N	0.037	N	0.071
R	0	0.4087	0	0.0488	N	0.068	N	0.034	N	0.045	N	0.026	N	0.021
P	0	0.1716	0	0.0474	N	0.030	N	0.037	N	0.031	N	0.021	N	0.020
Q	0	0.1595	0	0.0471	N	0.036	N	0.032	N	0.033	N	0.054	N	0.020
K	0	0.2281	0	0.0479	N	0.030	N	0.034	N	0.029	N	0.034	N	0.013
K	0	0.2593	0	0.0497	N	0.029	N	0.024	N	0.029	N	0.024	N	0.011
E	0	0.1661	0	0.0472	N	0.031	N	0.020	N	0.027	N	0.015	N	0.016
R	0	0.3077	0	0.0451	N	0.032	N	0.025	N	0.028	N	0.017	N	0.015
R	0	0.2532	0	0.0450	N	0.020	N	0.030	N	0.056	N	0.017	N	0.018
F	0	0.1545	0	0.0508	N	0.021	N	0.014	N	0.020	N	0.014	N	0.014
N	0	0.2557	0	0.0530	N	0.024	N	0.018	N	0.030	N	0.023	N	0.014
V	0	0.1741	0	0.0511	N	0.027	N	0.024	N	0.036	N	0.016	N	0.015
L	0	0.1332	0	0.0496	N	0.019	N	0.023	N	0.047	N	0.019	N	0.016

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
N	0	0.1961	0	0.0514	N	0.019	N	0.023	N	0.019	N	0.028	N	0.017
Q	0	0.2291	0	0.0520	N	0.023	N	0.017	N	0.021	N	0.016	N	0.014
I	0	0.1312	0	0.0500	N	0.024	N	0.018	N	0.016	N	0.011	N	0.009
F	0	0.1361	0	0.0506	N	0.030	N	0.022	N	0.018	N	0.012	N	0.013
C	0	0.2191	0	0.0530	N	0.021	N	0.018	N	0.019	N	0.009	N	0.016
K	0	0.3025	0	0.0533	N	0.023	N	0.021	N	0.019	N	0.018	N	0.015
N	0	0.1774	0	0.0521	N	0.032	N	0.022	N	0.022	N	0.013	N	0.018
K	0	0.2521	0	0.0522	N	0.028	N	0.021	N	0.023	N	0.013	N	0.010
K	0	0.2899	0	0.0534	N	0.023	N	0.019	N	0.028	N	0.016	N	0.015
E	0	0.1494	0	0.0492	N	0.036	N	0.041	N	0.029	N	0.022	N	0.015
E	0	0.1313	0	0.0491	N	0.024	N	0.034	N	0.029	N	0.021	N	0.016
Q	0	0.2065	0	0.0511	N	0.022	N	0.024	N	0.036	N	0.013	N	0.011
R	0	0.2431	0	0.0509	N	0.026	N	0.026	N	0.027	N	0.008	N	0.008
A	0	0.1296	0	0.0492	N	0.027	N	0.023	N	0.035	N	0.020	N	0.017
H	0	0.2437	0	0.0532	N	0.032	N	0.033	N	0.027	N	0.014	N	0.012
Q	0	0.1642	0	0.0549	N	0.029	N	0.017	N	0.019	N	0.011	N	0.013
K	0	0.3189	0	0.0605	N	0.035	N	0.024	N	0.027	N	0.010	N	0.019
D	0	0.1691	0	0.0554	N	0.023	N	0.020	N	0.023	N	0.023	N	0.015
P	0	0.1718	0	0.0551	N	0.028	N	0.030	N	0.024	N	0.012	N	0.019
S	0	0.1746	0	0.0544	N	0.024	N	0.013	N	0.019	N	0.014	N	0.037
R	0	0.2080	0	0.0555	N	0.020	N	0.046	N	0.030	N	0.028	N	0.031
D	0	0.1356	0	0.0558	N	0.026	N	0.039	N	0.026	N	0.032	N	0.050
R	0	0.1598	0	0.0547	N	0.027	N	0.033	N	0.034	N	0.033	N	0.026
Y	0	0.1434	0	0.0509	N	0.032	N	0.046	N	0.022	N	0.033	N	0.016
R	0	0.1278	0	0.0622	N	0.024	N	0.022	N	0.030	N	0.057	N	0.014

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
E	0	0.1095	0	0.0605	N	0.025	N	0.026	N	0.033	N	0.044	N	0.012
E	0	0.0998	0	0.0601	N	0.043	N	0.026	N	0.022	N	0.031	N	0.009
D	0	0.1019	0	0.0602	N	0.026	N	0.021	N	0.021	N	0.025	N	0.012
T	0	0.1125	0	0.0672	N	0.027	N	0.018	N	0.026	N	0.011	N	0.013
S	0	0.1077	0	0.0663	N	0.027	N	0.021	N	0.023	N	0.015	N	0.014
E	0	0.1033	0	0.0687	N	0.018	N	0.020	N	0.020	N	0.015	N	0.014
V	0	0.0935	0	0.0698	N	0.031	N	0.019	N	0.022	N	0.018	N	0.013
N	0	0.1151	0	0.0643	N	0.029	N	0.011	N	0.018	N	0.015	N	0.013
D	0	0.1292	0	0.0573	N	0.027	N	0.031	N	0.026	N	0.021	N	0.016
I	0	0.0975	0	0.0555	N	0.017	N	0.030	N	0.032	N	0.024	N	0.018
I	0	0.0945	0	0.0552	N	0.027	N	0.031	N	0.024	N	0.025	N	0.017
T	0	0.1136	0	0.0562	N	0.025	N	0.019	N	0.026	N	0.017	N	0.012
T	0	0.1017	0	0.0565	N	0.034	N	0.024	N	0.023	N	0.016	N	0.011
F	0	0.1211	0	0.0529	N	0.022	N	0.013	N	0.025	N	0.018	N	0.012
D	0	0.1783	0	0.0545	N	0.021	N	0.018	N	0.019	N	0.018	N	0.016
S	0	0.2399	0	0.0535	N	0.028	N	0.021	N	0.028	N	0.018	N	0.016
I	0	0.1987	0	0.0521	N	0.030	N	0.022	N	0.028	N	0.016	N	0.017
V	0	0.1536	0	0.0523	N	0.031	N	0.029	N	0.030	N	0.022	N	0.019
G	0	0.1979	0	0.0536	N	0.021	N	0.019	N	0.030	N	0.020	N	0.020
T	0	0.1374	0	0.0527	N	0.031	N	0.027	N	0.032	N	0.026	N	0.021
N	0	0.2619	0	0.0546	N	0.023	N	0.022	N	0.030	N	0.028	N	0.020
C	0	0.1387	0	0.0475	N	0.027	N	0.025	N	0.029	N	0.017	N	0.019
Q	0	0.2761	0	0.0490	N	0.035	N	0.023	N	0.030	N	0.024	N	0.015
E	0	0.1731	0	0.0479	N	0.024	N	0.027	N	0.027	N	0.020	N	0.019
Q	0	0.1478	0	0.0467	N	0.036	N	0.030	N	0.028	N	0.021	N	0.015

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
P	0	0.1745	0	0.0481	N	0.025	N	0.035	N	0.025	N	0.016	N	0.016
G	0	0.2091	0	0.0487	N	0.032	N	0.041	N	0.034	N	0.050	N	0.015
D	0	0.1907	0	0.0487	N	0.054	N	0.035	N	0.020	N	0.013	N	0.017
Q	0	0.3463	0	0.0506	N	0.032	N	0.035	N	0.034	N	0.039	N	0.015
V	0	0.1391	0	0.0458	N	0.028	N	0.060	N	0.030	N	0.036	N	0.021
A	0	0.1060	0	0.0375	N	0.041	N	0.054	N	0.045	N	0.032	N	0.022
M	0	0.1314	0	0.0394	N	0.071	N	0.062	N	0.036	N	0.072	N	0.023
V	0	0.0940	0	0.0376	N	0.060	N	0.029	N	0.033	N	0.030	N	0.022
E	0	0.1154	0	0.0390	N	0.031	N	0.027	N	0.036	N	0.032	N	0.021
F	0	0.1175	0	0.0391	N	0.028	N	0.026	N	0.032	N	0.022	N	0.015
K	0	0.2818	0	0.0405	N	0.034	N	0.029	N	0.025	N	0.024	N	0.016
K	0	0.3919	0	0.0408	N	0.061	N	0.021	N	0.022	N	0.023	N	0.017
K	1	0.5325	0	0.0423	N	0.032	N	0.023	N	0.049	N	0.025	N	0.020
T	0	0.4565	0	0.0422	N	0.032	N	0.027	N	0.029	N	0.024	N	0.017
S	0	0.3848	0	0.0431	N	0.027	N	0.021	N	0.027	N	0.021	N	0.018
D	0	0.2867	0	0.0439	N	0.025	N	0.017	N	0.029	N	0.022	N	0.016
N	0	0.2181	0	0.0445	N	0.030	N	0.034	N	0.033	N	0.024	N	0.017
S	0	0.1632	0	0.0424	N	0.033	N	0.033	N	0.032	N	0.023	N	0.018
K	0	0.2515	0	0.0467	N	0.025	N	0.029	N	0.032	N	0.031	N	0.022
Y	0	0.1772	0	0.0467	N	0.029	N	0.023	N	0.025	N	0.017	N	0.017
L	0	0.0849	0	0.0436	N	0.023	N	0.019	N	0.026	N	0.025	N	0.014
L	0	0.0926	0	0.0491	N	0.026	N	0.020	N	0.020	N	0.024	N	0.014
P	0	0.1498	0	0.0675	N	0.024	N	0.023	N	0.031	N	0.017	N	0.018
E	0	0.1405	0	0.0706	N	0.029	N	0.026	N	0.027	N	0.029	N	0.020
K	0	0.1627	0	0.0672	N	0.032	N	0.025	N	0.018	N	0.034	N	0.017

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
K	0	0.1138	0	0.0544	N	0.032	N	0.019	N	0.024	N	0.030	N	0.017
P	0	0.1240	0	0.0583	N	0.026	N	0.022	N	0.029	N	0.027	N	0.022
L	0	0.1560	0	0.0612	N	0.032	N	0.032	N	0.031	N	0.020	N	0.019
A	0	0.1606	0	0.0607	N	0.031	N	0.030	N	0.029	N	0.033	N	0.019
R	0	0.2663	0	0.0506	N	0.028	N	0.024	N	0.025	N	0.014	N	0.020
K	0	0.2565	0	0.0502	N	0.026	N	0.019	N	0.026	N	0.017	N	0.018
G	0	0.2338	0	0.0506	N	0.015	N	0.029	N	0.020	N	0.012	N	0.032
L	0	0.1304	0	0.0434	N	0.021	N	0.021	N	0.025	N	0.014	N	0.015
P	0	0.2450	0	0.0499	N	0.023	N	0.018	N	0.016	N	0.009	N	0.010
P	0	0.2042	0	0.0486	N	0.013	N	0.014	N	0.022	N	0.011	N	0.006
I	0	0.1262	0	0.0414	N	0.013	N	0.016	N	0.017	N	0.015	N	0.007
R	1	0.4811	0	0.0539	N	0.024	N	0.021	N	0.022	N	0.028	N	0.011
T	0	0.3826	0	0.0494	N	0.049	N	0.037	N	0.030	N	0.021	N	0.020
Q	0	0.2083	0	0.0437	N	0.030	N	0.027	N	0.026	N	0.069	N	0.014
S	0	0.2132	0	0.0428	N	0.037	N	0.021	N	0.029	N	0.034	N	0.015
L	0	0.0957	0	0.0358	N	0.020	N	0.024	N	0.028	N	0.032	N	0.009
P	0	0.1670	0	0.0401	N	0.028	N	0.017	N	0.028	N	0.017	N	0.010
P	0	0.1465	0	0.0396	N	0.028	N	0.024	N	0.024	N	0.027	N	0.010
I	0	0.1101	0	0.0392	N	0.023	N	0.027	N	0.021	N	0.027	N	0.010
T	0	0.4036	0	0.0392	N	0.030	N	0.018	N	0.028	N	0.019	N	0.009
L	0	0.3820	0	0.0381	N	0.029	N	0.016	N	0.031	N	0.017	N	0.015
G	0	0.3619	0	0.0334	N	0.025	N	0.023	N	0.035	N	0.023	N	0.020
N	1	0.5579	0	0.0353	N	0.038	N	0.025	N	0.028	N	0.018	N	0.024
N	0	0.3566	0	0.0314	N	0.026	N	0.041	N	0.029	N	0.015	N	0.011
F	0	0.2406	0	0.0304	N	0.031	N	0.075	N	0.028	N	0.050	N	0.017

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
L	0	0.1522	0	0.0290	N	0.056	N	0.051	N	0.038	N	0.013	N	0.016
T	1	0.5703	0	0.0321	N	0.034	N	0.025	N	0.038	N	0.053	N	0.026
A	0	0.3052	0	0.0289	N	0.055	B	0.161	N	0.040	N	0.042	N	0.026
S	1	0.7072	0	0.0317	N	0.059	N	0.056	N	0.041	N	0.032	N	0.031
H	1	0.7067	0	0.0325	B	0.112	N	0.088	N	0.037	N	0.066	N	0.038
R	1	0.7105	0	0.0320	B	0.120	N	0.024	N	0.025	N	0.014	N	0.023
A	0	0.2623	0	0.0291	N	0.027	N	0.028	N	0.034	N	0.033	N	0.020
T	1	0.6075	0	0.0318	N	0.038	N	0.025	N	0.030	N	0.021	N	0.015
S	1	0.5098	0	0.0329	N	0.065	N	0.029	N	0.026	N	0.023	N	0.017
H	1	0.7260	0	0.0335	B	0.104	N	0.023	N	0.021	N	0.028	N	0.017
A	0	0.3480	0	0.0298	N	0.035	N	0.022	N	0.035	N	0.026	N	0.016
G	0	0.2524	0	0.0300	N	0.020	N	0.022	N	0.028	N	0.021	N	0.015
L	0	0.2343	0	0.0295	N	0.026	N	0.013	N	0.018	N	0.017	N	0.015
S	1	0.5036	0	0.0331	N	0.018	N	0.016	N	0.026	N	0.018	N	0.013
S	1	0.4933	0	0.0334	N	0.026	N	0.024	N	0.035	N	0.018	N	0.014
A	0	0.3340	0	0.0304	N	0.024	N	0.024	N	0.037	N	0.022	N	0.017
P	0	0.2280	0	0.0308	N	0.023	N	0.025	N	0.027	N	0.032	N	0.019
H	0	0.2769	0	0.0351	N	0.019	N	0.026	N	0.026	N	0.016	N	0.016
H	1	0.6196	0	0.0420	N	0.021	N	0.020	N	0.024	N	0.022	N	0.012
M	0	0.3242	0	0.0367	N	0.021	N	0.013	N	0.016	N	0.024	N	0.010
A	0	0.1598	0	0.0313	N	0.027	N	0.020	N	0.024	N	0.016	N	0.016
Q	0	0.3476	0	0.0358	N	0.025	N	0.020	N	0.025	N	0.019	N	0.015
R	1	0.6162	0	0.0341	N	0.035	N	0.027	N	0.019	N	0.023	N	0.016
S	1	0.5977	0	0.0345	N	0.031	N	0.017	N	0.025	N	0.025	N	0.015
Q	1	0.6657	0	0.0357	N	0.023	N	0.020	N	0.026	N	0.024	N	0.019

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
K	1	0.6683	0	0.0360	N	0.025	N	0.028	N	0.030	N	0.021	N	0.019
S	1	0.6649	0	0.0355	N	0.022	N	0.023	N	0.028	N	0.021	N	0.019
R	1	0.6931	0	0.0349	N	0.025	N	0.028	N	0.031	N	0.014	N	0.020
S	0	0.4642	0	0.0350	N	0.031	N	0.025	N	0.035	N	0.021	N	0.015
E	0	0.1156	0	0.0309	N	0.032	N	0.026	N	0.029	N	0.032	N	0.025
Q	0	0.2881	0	0.0362	N	0.046	N	0.039	N	0.036	N	0.045	N	0.020
D	0	0.1759	0	0.0345	N	0.022	N	0.023	N	0.024	N	0.024	N	0.025
L	0	0.0946	0	0.0305	N	0.036	N	0.026	N	0.033	N	0.012	N	0.018
L	0	0.1260	0	0.0312	N	0.028	N	0.021	N	0.034	N	0.015	N	0.016
N	0	0.3410	0	0.0336	N	0.023	N	0.031	N	0.025	N	0.013	N	0.015
N	0	0.3818	0	0.0392	N	0.030	N	0.023	N	0.027	N	0.015	N	0.027
R	1	0.5902	0	0.0399	N	0.026	N	0.015	N	0.045	N	0.012	N	0.011
T	0	0.4725	0	0.0380	N	0.032	N	0.032	N	0.033	N	0.017	N	0.019
G	0	0.4431	0	0.0381	N	0.026	N	0.021	N	0.034	N	0.016	N	0.024
C	0	0.1291	0	0.0367	N	0.034	N	0.029	N	0.030	N	0.018	N	0.026
Q	0	0.2728	0	0.0373	N	0.024	N	0.040	N	0.031	N	0.027	N	0.021
M	0	0.1872	0	0.0330	N	0.025	N	0.024	N	0.020	N	0.021	N	0.015
L	0	0.1004	0	0.0321	N	0.033	N	0.020	N	0.039	N	0.026	N	0.014
L	0	0.0948	0	0.0320	N	0.022	N	0.012	N	0.018	N	0.016	N	0.010
D	0	0.1645	0	0.0347	N	0.021	N	0.025	N	0.019	N	0.015	N	0.014
N	0	0.2939	0	0.0353	N	0.024	N	0.023	N	0.021	N	0.013	N	0.019
P	0	0.1565	0	0.0342	N	0.026	N	0.028	N	0.024	N	0.014	N	0.015
W	0	0.4272	0	0.0335	N	0.026	N	0.025	N	0.020	N	0.017	N	0.016
K	1	0.6468	0	0.0357	N	0.026	N	0.043	N	0.020	N	0.017	N	0.013
S	0	0.4100	0	0.0336	N	0.021	N	0.024	N	0.023	N	0.017	N	0.011

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
D	0	0.2152	0	0.0330	N	0.024	N	0.024	N	0.025	N	0.016	N	0.011
S	0	0.4308	0	0.0332	N	0.043	N	0.019	N	0.037	N	0.024	N	0.013
N	1	0.5911	0	0.0341	N	0.015	N	0.017	N	0.017	N	0.023	N	0.041
Q	1	0.5134	0	0.0341	N	0.017	N	0.014	N	0.021	N	0.020	N	0.025
V	0	0.1469	0	0.0317	N	0.037	N	0.019	N	0.030	N	0.011	N	0.014
F	0	0.1616	0	0.0325	N	0.026	N	0.034	N	0.042	N	0.037	N	0.018
S	0	0.4358	0	0.0373	N	0.026	N	0.024	N	0.036	N	0.017	N	0.014
Y	0	0.2063	0	0.0372	N	0.021	N	0.029	N	0.026	N	0.024	N	0.024
K	1	0.5171	0	0.0370	N	0.026	N	0.030	N	0.031	N	0.023	N	0.025
V	0	0.1478	0	0.0345	N	0.016	N	0.029	N	0.042	N	0.029	N	0.028
W	1	0.5383	0	0.0373	N	0.013	N	0.027	N	0.029	N	0.025	N	0.016
T	1	0.5095	0	0.0384	N	0.025	N	0.022	N	0.034	N	0.017	N	0.016
V	0	0.2221	0	0.0366	N	0.019	N	0.024	N	0.032	N	0.014	N	0.016
S	0	0.4567	0	0.0390	N	0.020	N	0.024	N	0.025	N	0.017	N	0.020
S	0	0.3740	0	0.0393	N	0.014	N	0.024	N	0.033	N	0.018	N	0.017
S	0	0.3084	0	0.0386	N	0.019	N	0.022	N	0.027	N	0.009	N	0.013
D	0	0.2038	0	0.0387	N	0.023	N	0.028	N	0.014	N	0.028	N	0.015
K	1	0.4995	0	0.0413	N	0.025	N	0.027	N	0.031	N	0.016	N	0.012
L	0	0.1054	0	0.0374	N	0.031	N	0.012	N	0.029	N	0.021	N	0.010
L	0	0.0925	0	0.0377	N	0.016	N	0.047	N	0.022	N	0.026	N	0.009
D	0	0.1322	0	0.0411	N	0.033	N	0.019	N	0.023	N	0.014	N	0.030
R	0	0.2523	0	0.0409	N	0.017	N	0.022	N	0.018	N	0.017	N	0.040
L	0	0.0898	0	0.0372	N	0.020	N	0.019	N	0.013	N	0.008	N	0.013
L	0	0.1026	0	0.0384	N	0.012	N	0.013	N	0.016	N	0.016	N	0.007
S	0	0.2088	0	0.0398	N	0.015	N	0.014	N	0.032	N	0.024	N	0.018

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
V	0	0.1917	0	0.0409	N	0.027	N	0.031	N	0.024	N	0.019	N	0.013
R	1	0.4825	0	0.0429	N	0.027	N	0.020	N	0.035	N	0.018	N	0.012
P	0	0.2391	0	0.0426	N	0.032	N	0.026	N	0.028	N	0.017	N	0.010
G	0	0.2914	0	0.0432	N	0.046	N	0.025	N	0.046	N	0.046	N	0.010
H	0	0.3042	0	0.0451	N	0.041	N	0.046	N	0.029	N	0.034	N	0.014
Q	0	0.1966	0	0.0421	N	0.027	N	0.031	N	0.024	N	0.014	N	0.009
E	0	0.1216	0	0.0403	N	0.042	N	0.040	N	0.047	N	0.026	N	0.025
V	0	0.1397	0	0.0412	N	0.027	N	0.026	N	0.061	N	0.038	N	0.011
S	0	0.1235	0	0.0405	N	0.034	N	0.026	N	0.025	N	0.015	N	0.012
V	0	0.1042	0	0.0486	N	0.029	N	0.031	N	0.035	N	0.023	N	0.014
P	0	0.1331	0	0.0512	N	0.031	N	0.026	N	0.041	N	0.023	N	0.019
P	0	0.1547	0	0.0421	N	0.026	N	0.029	N	0.025	N	0.017	N	0.018
H	0	0.3235	0	0.0456	N	0.024	N	0.057	N	0.052	N	0.028	N	0.016
L	0	0.1021	0	0.0366	N	0.037	N	0.037	N	0.049	N	0.008	N	0.025
R	0	0.2018	0	0.0423	N	0.034	N	0.020	N	0.037	N	0.020	N	0.017
H	0	0.3487	0	0.0455	N	0.038	N	0.042	N	0.024	N	0.010	N	0.007
L	0	0.1078	0	0.0364	N	0.019	N	0.015	N	0.020	N	0.014	N	0.012
H	0	0.3000	0	0.0412	N	0.016	N	0.035	N	0.034	N	0.010	N	0.011
N	1	0.5610	0	0.0436	N	0.020	N	0.031	N	0.018	N	0.010	N	0.029
P	0	0.3539	0	0.0399	N	0.018	N	0.023	N	0.020	N	0.024	N	0.010
S	0	0.3991	0	0.0409	N	0.022	N	0.031	N	0.026	N	0.027	N	0.012
S	0	0.4561	0	0.0391	N	0.011	N	0.014	N	0.021	N	0.027	N	0.011
G	0	0.3985	0	0.0383	N	0.020	N	0.014	N	0.024	N	0.020	N	0.024
Q	1	0.5160	0	0.0397	N	0.022	N	0.026	N	0.017	N	0.020	N	0.007
N	1	0.5599	0	0.0394	N	0.015	N	0.019	N	0.016	N	0.005	N	0.005

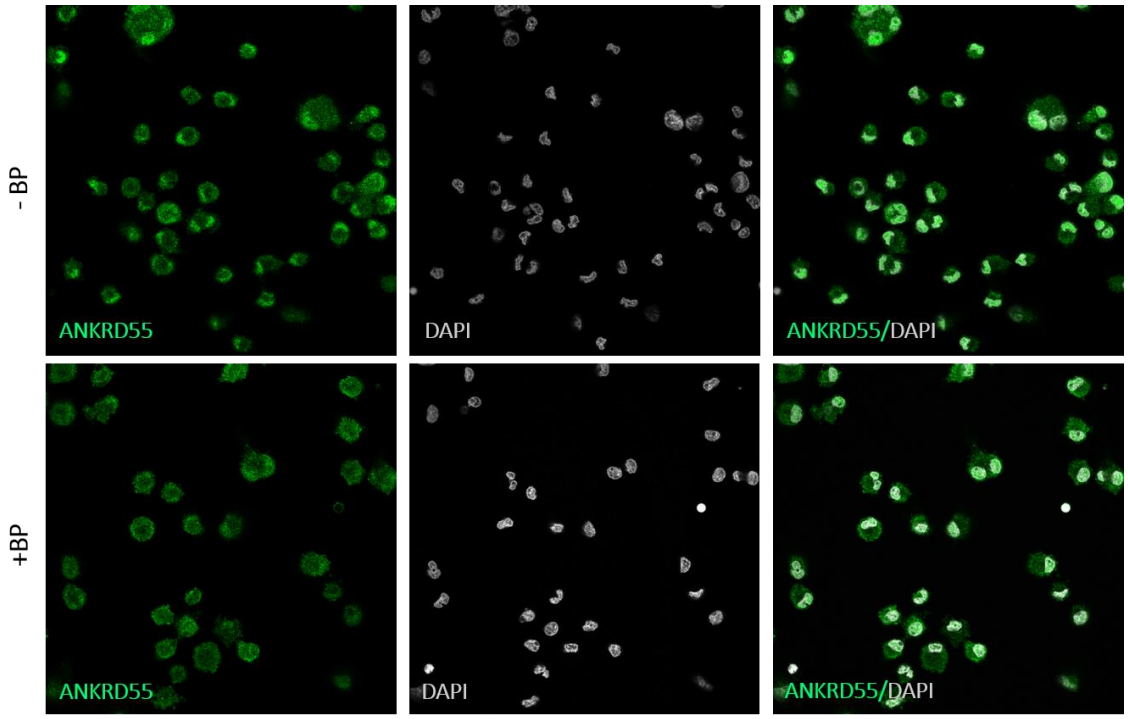
Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
F	0	0.3264	0	0.0377	N	0.042	N	0.037	N	0.020	N	0.016	N	0.015
Q	1	0.5873	0	0.0429	N	0.011	N	0.012	N	0.014	N	0.014	N	0.020
H	0	0.4597	0	0.0395	N	0.033	N	0.017	N	0.020	N	0.017	N	0.024
L	0	0.1411	0	0.0355	N	0.038	N	0.009	N	0.016	N	0.003	N	0.011
S	0	0.3823	0	0.0417	N	0.023	N	0.036	N	0.023	N	0.019	N	0.011
P	0	0.1553	0	0.0361	N	0.030	N	0.021	N	0.023	N	0.021	N	0.012
N	0	0.2520	0	0.0379	N	0.009	N	0.015	N	0.015	N	0.009	N	0.009
R	1	0.4948	0	0.0387	N	0.023	N	0.028	N	0.054	N	0.016	N	0.036
H	0	0.2401	0	0.0387	N	0.027	N	0.023	N	0.023	N	0.009	N	0.013
K	1	0.4759	0	0.0486	N	0.032	N	0.026	N	0.021	N	0.033	N	0.011
I	0	0.1323	0	0.0372	N	0.070	N	0.014	N	0.019	N	0.010	N	0.005
R	0	0.4028	0	0.0442	N	0.023	N	0.016	N	0.022	N	0.028	N	0.011
D	0	0.1334	0	0.0393	N	0.012	N	0.012	N	0.017	N	0.015	N	0.021
L	0	0.1179	0	0.0368	N	0.021	N	0.015	N	0.024	N	0.009	N	0.012
P	0	0.1416	0	0.0399	N	0.034	N	0.011	N	0.033	N	0.028	N	0.008
F	0	0.1399	0	0.0384	N	0.026	N	0.018	N	0.021	N	0.011	N	0.013
T	0	0.2907	0	0.0394	N	0.019	N	0.009	N	0.022	N	0.012	N	0.005
R	1	0.5806	0	0.0415	N	0.029	N	0.023	N	0.033	N	0.012	N	0.010
N	1	0.5489	0	0.0413	N	0.036	N	0.054	N	0.020	N	0.020	N	0.019
N	1	0.5078	0	0.0415	N	0.030	N	0.017	N	0.025	N	0.036	N	0.011
L	0	0.0997	0	0.0351	N	0.019	N	0.020	N	0.026	N	0.011	N	0.008
A	0	0.1319	0	0.0368	N	0.026	N	0.011	N	0.022	N	0.021	N	0.005
P	0	0.1592	0	0.0410	N	0.010	N	0.007	N	0.055	B	0.098	N	0.017
L	0	0.0981	0	0.0356	N	0.036	N	0.011	N	0.028	N	0.036	N	0.011
P	0	0.1393	0	0.0413	N	0.014	N	0.012	N	0.024	N	0.028	N	0.012

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
D	0	0.1566	0	0.0444	N	0.030	N	0.019	N	0.026	N	0.026	N	0.015
Q	0	0.2330	0	0.0421	N	0.031	N	0.040	N	0.025	N	0.038	N	0.017
K	0	0.4475	0	0.0453	N	0.023	N	0.029	N	0.021	N	0.009	N	0.011
F	0	0.1354	0	0.0355	N	0.024	N	0.031	N	0.022	N	0.023	N	0.023
L	0	0.0958	0	0.0332	N	0.025	N	0.032	N	0.021	N	0.017	N	0.017
S	0	0.3820	0	0.0394	N	0.034	N	0.026	N	0.028	N	0.015	N	0.006
G	0	0.1922	0	0.0371	N	0.028	N	0.028	N	0.027	N	0.013	N	0.013
E	0	0.1178	0	0.0341	N	0.037	N	0.016	N	0.025	N	0.017	N	0.011
P	0	0.1617	0	0.0367	N	0.012	N	0.013	N	0.033	N	0.011	N	0.008
L	0	0.1073	0	0.0331	N	0.024	N	0.014	N	0.027	N	0.019	N	0.012
R	1	0.5988	0	0.0377	N	0.053	N	0.028	N	0.028	N	0.012	N	0.015
T	0	0.4553	0	0.0371	N	0.034	N	0.016	N	0.034	N	0.007	N	0.013
N	0	0.4112	0	0.0372	N	0.028	N	0.028	N	0.021	N	0.013	N	0.010
R	1	0.5750	0	0.0385	N	0.029	N	0.018	N	0.017	N	0.011	N	0.010
V	0	0.1369	0	0.0348	N	0.021	N	0.009	N	0.019	N	0.007	N	0.017
L	0	0.0883	0	0.0287	N	0.018	N	0.017	N	0.026	N	0.008	N	0.008
P	0	0.1355	0	0.0332	N	0.026	N	0.023	N	0.033	N	0.019	N	0.015
A	0	0.1042	0	0.0295	N	0.021	N	0.025	N	0.032	N	0.010	N	0.016
I	0	0.1033	0	0.0292	N	0.025	N	0.018	N	0.021	N	0.009	N	0.016
P	0	0.1575	0	0.0330	N	0.017	N	0.017	N	0.027	N	0.018	N	0.022
S	1	0.4750	0	0.0342	N	0.030	N	0.033	N	0.023	N	0.017	N	0.017
Q	1	0.5741	0	0.0340	N	0.036	N	0.048	N	0.027	N	0.020	N	0.021
R	1	0.7255	0	0.0331	N	0.056	N	0.016	N	0.030	N	0.025	N	0.027
R	1	0.6229	0	0.0331	N	0.021	N	0.025	N	0.017	N	0.021	N	0.015
H	1	0.6440	0	0.0342	N	0.024	N	0.020	N	0.025	N	0.019	N	0.008

Amino Acid	Binary DNA	DNA prob	Binary RNA	RNA prob	ATP binding	ATP prob	ADP binding	ADP prob	AMP binding	AMP prob	GTP binding	GTP prob	GDP binding	GDP prob
S	0	0.3627	0	0.0324	N	0.031	N	0.014	N	0.026	N	0.022	N	0.009
T	0	0.1686	0	0.0305	N	0.035	N	0.012	N	0.028	N	0.009	N	0.005
A	0	0.2083	0	0.0313	N	0.032	N	0.020	N	0.027	N	0.020	N	0.006
A	0	0.1250	0	0.0297	N	0.009	N	0.009	N	0.015	N	0.010	N	0.007
E	0	0.1100	0	0.0296	N	0.008	N	0.015	N	0.015	N	0.012	N	0.009
E	0	0.1347	0	0.0332	N	0.013	N	0.023	N	0.018	N	0.018	N	0.005
S	0	0.1445	0	0.0336	N	0.022	N	0.017	N	0.028	N	0.012	N	0.010
E	0	0.1020	0	0.0469	N	0.019	N	0.022	N	0.024	N	0.012	N	0.009
H	0	0.2295	0	0.0567	N	0.013	N	0.020	N	0.021	N	0.011	N	0.004
S	0	0.2174	0	0.0568	N	0.024	N	0.019	N	0.029	N	0.018	N	0.010
A	0	0.1881	0	0.0568	N	0.017	N	0.017	N	0.022	N	0.017	N	0.010
N	0	0.1734	0	0.0568	N	0.027	N	0.024	N	0.021	N	0.016	N	0.008
P	0	0.1209	0	0.0544	N	0.024	N	0.017	N	0.017	N	0.011	N	0.011
T	0	0.2552	0	0.0579	N	0.027	N	0.023	N	0.026	N	0.014	N	0.008
S	0	0.1806	0	0.0582	N	0.032	N	0.020	N	0.022	N	0.014	N	0.015
D	0	0.2091	0	0.0584	N	0.028	N	0.036	N	0.026	N	0.017	N	0.022
E	0	0.1374	0	0.0551	N	0.026	N	0.034	N	0.028	N	0.026	N	0.019
N	0	0.2206	0	0.0555	N	0.021	N	0.027	N	0.021	N	0.019	N	0.016

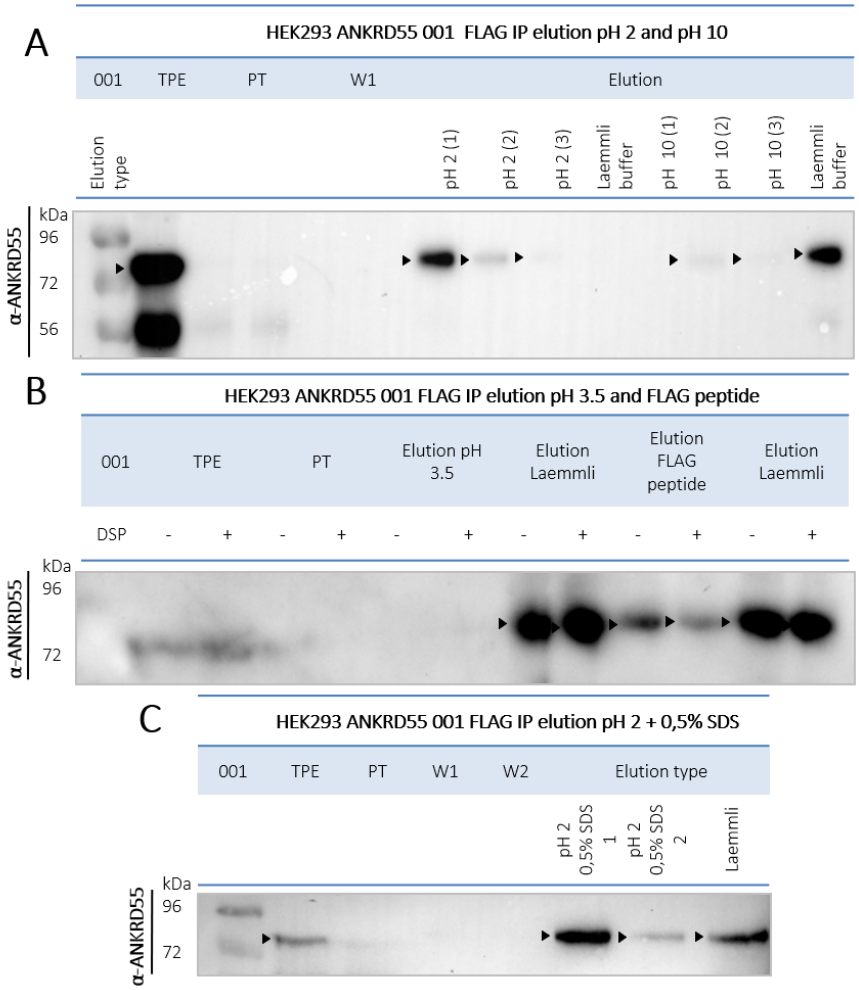
Supplementary Table 9. Interaction between MS/RA risk genes and ANKRD55 and/or ANKRD55-interacting partners. In bold the proximal risk gene involved in the interaction. Abbreviations: TPE, total protein extracts and NE, nuclear extracts.

	SNP	Proximal Gene(s)	Interaction with ANKRD55	Interaction with ANKRD55-interacting partners
Multiple sclerosis	rs9843355	TIMMDC1	X (TPE 2/3)	
	rs72928038	BACH2	X (HuRI)	
	rs37749689	STAT3		CDK5, NDUFA13 (TPE 2/3)
	rs11079784	NPEPPS-**KPNB1**	X (TPE 2/3)	KPNA2, IPO7 (TPE 2/3)
	rs9321490	HBS1L-**MYB**		MYBBP1A, RQCD1
	rs2283792	MAPK1		PRMT5, CDK5, DDX17
	rs10438843	PDE4A-**KEAP1**		PGAM5
	rs10914539	LCK		GNB2L1
	rs12086448	VANGL2		GNB2L1
	rs12373588	MIR4435-1HG-**ANAPC1**	X (TPE 2/3)	
	rs7690934	LEF1	X (HuRI)	RUVBL1, RUVBL2
	rs32954092	TRIM71-**CCR4**		CNOT1, RQCD1 (TPE 2/3)
	rs10271373	ZC3HAV1		DDX17, DHX30 (TPE-NE 2/3)
	rs11023242	RRAS2-**COPB1**	X (TPE 2/3)	
	rs9913257	GRB2		IRS4
Rheumatoid arthritis	rs6715284	CFLAR-**CASP8**		RPS3
	rs45475795	**IL2-IL21**		ILF2, WDR36
	rs657075	**IL3-CSF2**		ELAVL1
	rs9268839	**HLA-DRB1**		YBX1
	chr11:107967350	**ATM**		SMC1A, CDK5
	rs773125	**CDK2**		DDX17
	rs72634030	**C1QBP**	X (TPE 3/3 NE 2/3)	
	rs1877030	**MED1**		DDX17
	chr19:10771941	**ILF3**		ILF2

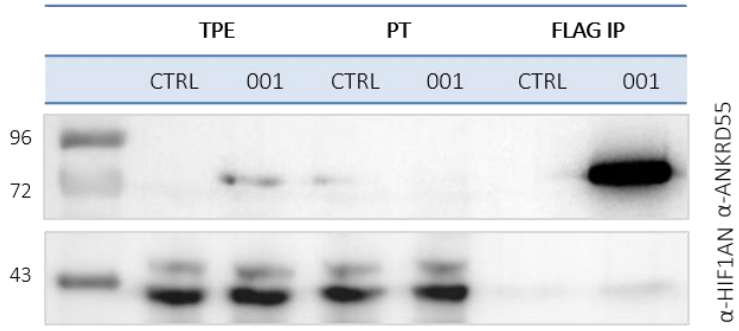


Supplementary Figure 1. Intracellular localization of ANKRD55 in PMA/LPS-treated U937 cells.

Inmunostaining for ANKRD55 (Aviva Systems Biology Ab; green) and nuclei (DAPI; grey) in PMA/LPS-treated U937 analyzed by confocal microscopy. To confirm the specificity, immunostaining was performed either in the absence (-BP) or presence (+BP) of ANKRD55 blocking peptide.



Supplementary Figure 2. Optimization process in the immunoprecipitation elution step from ANKRD55 isoform 001. HEK293 cells were transfected with FLAG-tagged isoform 001 of ANKRD55. Cells were treated +/- DSP, as indicated, and IP was performed in total protein extracts (TPE) after 48h of transfection. Elution step was carried out with (A) pH 2 or pH 10, each three consecutive times (B) pH 3.5 or FLAG peptide and (C) pH 2 + 0,5% SDS. The five distinct elution procedures were followed by a final elution step by means of Laemmli buffer. This removes all remaining protein of the gel matrix and facilitates to semi-quantitatively estimate the amount of protein not eluted by the specific condition. Recombinant ANKRD55 isoforms were detected in TPE, pass-through (PT), washing (W1, W2) and elution steps with specific Ab (Sigma-Aldrich) by WB.



Supplementary Figure 3. Validation of ANKRD55-interacting partners from BioGrid database. HEK293 cells were transfected with FLAG-tagged isoform 001 of ANKRD55 and IPed from total protein extracts (TPE) after 48h of transfection. Endogenous HIF1AN and overexpressed ANKRD55 were detected in TPE, pass-through (PT), and elution with specific Ab by WB.

APPENDIX

LIST OF PUBLICATIONS

1. Lopez de Lapuente A*, Feliú A*, Ugidos N*, Mecha M, Mena J, Astobiza I, Riera J, Carrillo-Salinas FJ, Comabella M, Montalban X, Alloza I, Guaza C, Vandenbroeck K. A Novel Insights into the Multiple Sclerosis Risk Gene ANKRD55. *J. Immunol.* 196, 4553–4565 (2016). (*shared first autor)