

**A negatív szelekciót meghatározó *AIRE* gén  
allélpolimorfizmusainak vizsgálata reumatoid  
arthritis kockázatára klinikai minták  
felhasználásával**

Doktori (Ph.D.) értekezés  
Bérczi Bálint



Klinikai Orvostudományok Doktori Iskola  
Daganatok Molekuláris Epidemiológiája  
Doktori Iskola vezetője: Prof. Dr. Bogár Lajos,  
Programvezető: Prof. Dr. Kiss István  
Témavezető: Dr. Gyöngyi Zoltán  
Pécsi Tudományegyetem Általános Orvostudományi Kar  
Orvosi Népegészségtani Intézet

Pécs, 2024.

# Tartalomjegyzék

Rövidítések jegyzéke.....	5
I. Bevezetés .....	7
I.1. Az érésben lévő T-sejtek autoimmun regulációja .....	7
I.1.1. Centrális tolerancia.....	7
I.1.1.1. A timociták bejutása és érése a neonatális tímusz kéregállományában .....	7
I.1.1.1.1. $\beta$ -szelekció .....	7
I.1.1.1.2. Pozitív szelekció.....	8
I.1.1.1.3. Negatív szelekció .....	8
I.1.1.2. A timociták bejutása és érése a tímusz velőállományában.....	8
I.1.1.2.1. Bejutás és felhalmozódás .....	9
I.1.1.2.2. Migráció a medullában és negatív szelekció.....	9
I.1.1.2.3. A mTEC sejtek promiskuis génexpressziója .....	10
I.1.1.2.4. AIRE fehérje jellemzői .....	10
I.1.1.2.4.1. AIRE fehérje strukturális doménjei.....	11
I.1.1.2.4.2. AIRE expressziója és az mTEC fejlődési stádiumai .....	12
I.1.1.2.4.3. Transzkripciós regulációja .....	13
I.1.1.2.4.4. Poszttranszkripciós regulációja .....	14
I.1.1.2.4.5. Transzportja a citoplazmából a sejtmagba .....	15
I.1.1.2.4.6. Az AIRE-mediált TRA kifejeződés fehérjepartnerei .....	16
I.1.1.2.5. A TRA útja a sejtmagtól a medulláris sejtek MHC receptoráig.....	18
I.1.1.2.6. Az autoreaktív timocita lebontása .....	19
I.1.2. Az érett timociták kijutása a perifériára .....	21
I.2. Az AIRE szerepe az autoimmun kórképek kialakulásában.....	22
I.2.1. <i>In vivo</i> manifesztációk.....	22
I.2.2. Humán manifesztációk.....	24
I.2.3. AIRE allélpolimorfizmusok autoimmun kórképekben .....	26
I.2.3.1 AIRE allélpolimorfizmusok és a rheumatoid arthritis .....	27
I.3. Reumatoid artritisz.....	28
I.3.1. Kialakulásának stádiumai.....	29
I.3.1.1. Iniciális szakasz.....	29
I.3.1.1.1. Környezeti kockázati tényezők .....	29
I.3.1.1.2. Genetikai kockázati tényezők.....	31
I.3.1.1.3. Immunológiai folyamatok.....	31

I.3.1.2. Centrális szakasz.....	32
I.3.1.3. Destruktív szakasz.....	32
I.3.2. Általános epidemiológiája.....	33
II. Célkitűzések.....	34
II.1. Az <i>AIRE</i> allélpolimorfizmusok meta-analízise autoimmun kórképekben.....	34
II.2. <i>In silico</i> vizsgálat <i>AIRE</i> allélpolimorfizmusok meghatározására .....	34
II.3. Pilot vizsgálat <i>AIRE</i> rs878081 és rs1003854 allélpolimorfizmusok vizsgálatára reumatoid artritisz betegek mintáiban .....	35
II.4. Az <i>AIRE</i> exon 5 és 6 valamint intron 5 és 7-es régióinak vizsgálata reumatoid artritisz kockázatára.....	35
II.5. Az <i>AIRE</i> exon 5 és 6 valamint intron 5 és 7-es régiók összefüggéseinek vizsgálata a betegséget meghatározó klinikai paraméterekkel.....	35
II.6. Allélpolimorfizmusok regulatórikus kötőmotívumokra gyakorolt hatásainak <i>in silico</i> elemzése .....	36
III. Anyagok és módszerek.....	36
III.1. <i>AIRE</i> allélpolimorfizmusok meta-analízise.....	36
III.1.1. Publikációk keresési stratégiája.....	36
III.1.2. Publikációk beválasztási és kizárási kritériumai .....	36
III.1.3. Statisztikai analízis .....	36
III.2. <i>In silico</i> vizsgálat <i>AIRE</i> allélpolimorfizmusok meghatározására .....	37
III.2.1. <i>AIRE</i> promóter régiójában található variánsok beválogatási és kizárási kritériumai .....	37
III.3. Pilot vizsgálat <i>AIRE</i> rs878081 és rs1003854 allélpolimorfizmusok vizsgálatára reumatoid artritisz betegek mintáiban .....	38
III.3.1. Etikai engedély .....	38
III.3.2. Mintagyűjtés .....	38
III.3.3. Klinikai paraméterek vizsgálata .....	39
III.3.4. Genomiális DNS extrakciója teljes vérből .....	40
III.3.5. Az <i>AIRE</i> rs878081 és rs1003854 SNP-k genotipizálása .....	40
III.3.6. Statisztikai analízis .....	41
III.4. Az <i>AIRE</i> exon 5 és 6 valamint intron 5 és 7-es régiókban található allélpolimorfizmusok vizsgálata reumatoid artritisz kockázatára.....	41
III.4.1. Etikai engedély .....	42
III.4.2. Mintagyűjtés .....	42
III.4.3. Klinikai paraméterek vizsgálata .....	42
III.4.4. Genomiális DNS extrakciója teljes vérből .....	43
III.4.5. Az <i>AIRE</i> allélpolimorfizmusok genotipizálása.....	43

III.4.6. Statisztikai analízis .....	44
III.5. <i>AIRE</i> allélpolimorfizmusok regulatorikus kötőmotívumokra gyakorolt hatásainak <i>in silico</i> elemzése .....	45
IV. Eredmények.....	46
IV.1. Az <i>AIRE</i> allélpolimorfizmusok meta-analízise autoimmun kórképekben .....	46
IV.1.1. Beválasztott közlemények jellemzői.....	46
IV.1.2. Az <i>AIRE</i> rs2075876 (G/A) SNP meta-analízise.....	46
IV.1.3. Az <i>AIRE</i> rs760426 (A/G) SNP meta-analízise.....	51
IV.1.4. Szenzitivitás-analízis.....	54
IV.1.5. Publikációs torzítás analízise .....	54
IV.1.6. Trial sequential analízis.....	55
IV.2. <i>In silico</i> vizsgálat <i>AIRE</i> allélpolimorfizmusok meghatározására .....	56
IV.3. Pilot vizsgálat <i>AIRE</i> rs878081 és rs1003854 allélpolimorfizmusok vizsgálatára reumatoid artritisz betegek mintáiban .....	57
IV.3.1. A pilot vizsgálatban résztvevő alanyok demográfiai és klinikai jellemzői .....	57
IV.3.2. Genotípus-frekvenciák és a Hardy–Weinberg equilibrium.....	58
IV.3.3. Egy-lókuszos genetikai modellek betegségkockázat meghatározására.....	59
IV.3.4. Az <i>AIRE</i> rs878081 (C/T) allélpolimorfizmusok és az RA kockázata közötti kapcsolat ...	59
IV.3.5. Az <i>AIRE</i> rs1003854 (T/C) allélpolimorfizmusok és az RA kockázata közötti kapcsolat .	60
IV.3.6. A klinikai paraméterek összehasonlítása az <i>AIRE</i> rs878081 és rs1003854 genetikai modelljei alapján RA betegeknél és kontroll személyeknél .....	62
IV.4. Az <i>AIRE</i> exon 5 és 6 valamint intron 5 és 7-es régiókban található allélpolimorfizmusok vizsgálata reumatoid artritisz kockázatára.....	64
IV.4.1. A vizsgálatban résztvevő alanyok demográfiai és klinikai jellemzői .....	65
IV.4.2. Genotípus-frekvenciák és a Hardy–Weinberg equilibrium.....	67
IV.4.3. Exon 5 és intron 5 és 7-es kódoló régiókban található allélpolimorfizmusok asszociációt mutatnak az RA kockázatával .....	67
IV.4.4. Genetikai modellek genotípusos-alcsoportjainak korrelációja és asszociációja klinikai paraméterekkel RA és kontrollcsoportban .....	70
IV.4.5. Regulatorikus kötőmotívumok allélspecifikus affinitása transzkripciós faktorokhoz .....	73
V. Megbeszélés .....	74
VI. Új eredmények .....	77
VII. Irodalomjegyzék .....	79
VIII. Függelék .....	98

## Rövidítések jegyzéke

AA	alopecia areata
aCCP	anti-cyclic citrullinated protein antibody, anti-ciklikus citrullinált peptid antitest
ACR	American College of Rheumatology
AIRE	autoimmun regulator transzkripció faktor fehérje
<i>AIRE</i>	AIRE transzkripció faktor kódoló gén
AP-2	Krüppel-like zinc finger protein
APC	antigen presenting cell, antigén prezentáló sejtek
APECED	autoimmune polyendocrinopathy-candidiasis ectodermal dystrophy, autoimmun poliendokrinopátia-kandidiázis
ATF7IP	activating transcription factor 7-interacting protein
BM	bone marrow, csontvelő
BMP	bone morphogenetic protein, csont morfogenetikus protein
CARD	caspase-activation and recruitment domain, kaszpáz-aktivációs és recruitment domain
CBP	CREB-binding protein, CREB-kötő fehérje
CCR7	C-C chemokine receptor type 7, 7-es típusú C-C kemokin receptor
CD	cluster of differentiation
CI	confidence interval, konfidencia intervallum
CIA	collagen induced arthritis
CMA	chaperone mediated autophagy
CMJ	cortico-medullar junction, kortiko-medulláris junkció
CNS1	conserved noncoding sequence 1, 1-es konzervált nemkódoló szekvencia
CNV	copy number variation, nagyméretű kópiaszám változás
CpG	CpG dinukleotid melyben a citozin foszfodiészter kötéssel kötődik guaninhoz
CRP	C-reaktív protein
cTEC	cortical thymoeptithelial cells, kortikális timoeptiteliális sejtek
DAS28	disease activity score in 28 joints, betegség aktivitási pontszám 28 ízületben
DC	dendritic cell, dendritikus sejt
dHPLC	denaturing High-Performance Liquid Chromatography
DMARD	disease modifying antirheumatic drugs, betegségmódosító antireumatikus terápia
DN	double negative, egyszerűen negatív
DNA PK	DNA-activated protein kinase, DNS-aktivált protein kináz
DNMT3L	DNS metil-transzferáz
DP	double positive, kettős pozitív
EBV	Eppstein-Barr vírus
ER	endoplazmatikus retikulum
ESET	Erg-associated protein with SET
ESR	erythrocyte sedimentation rate, eritrocita szedimentációs ráta
ETT TUKEB	Egészségügyi Tudományos Tanács és Kutatásaitikai Bizottsága
FIM	functional independence measure, funkcionális függetlenség mértéke
FLS	fibroblast like synoviocyte, fibroblasztszerű szinoviocita
FOXO3A	forkhead box O3A
gamma-GT	gamma-glutamiltranszferáz
gDNS	genomiális DNS
GFP	green fluorescent protein, zölden fluoreszkáló fehérje
GOT	glutamát-oxalacetát-transzamináz
GWAS	genome wide association studies, genomszintű asszociációs vizsgálatok
H3K4m0	metilálatlan H3 hiszton K4-es lizin molekulája
HDAC	histone deacetylases, hiszton deacetiláz
HLA	humán leukocita antigén
HNRNPL	heterogeneous nuclear ribonucleoprotein L
HSR	homogeneously staining region, homogéneen festődő régió
HWE	Hardy-Weinberg equilibrium, Hardy-Weinberg egyensúly
ICAM-1	intercellular Adhesion Molecule 1
IFN	interferon
IgG	immunglobulin G
IGRP	islet-specific glucose-6-phosphatase-related protein
IL	interleukin
JMJD6	jumonji domain containing 6
JNK	c-Jun N-terminális kináz
KO	knock-out
LAMP2	lysosomal associated membrane protein 2
LD	linkage disequilibrium
LOD	logarithm of odds

LTi	lymphoid tissue inducer, limfoid szöveti induktor
LxxLL	Leu-Xaa-Xaa-Leu-Leu motívum
MAF	minor allele frequency, ritka allél frekvenciája
MBD1	methyl-CpG-binding domain protein 1
MGAT1	alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase
MHC	major histocompatibility complex, fő hisztokompatibilitási komplex
MHC IIIhi CD80hi	mTEC MHC II-t és CD80-at magasan expresszáló mTEC
MHC IIIlo CD80lo	mTEC MHC II-t és CD80-at alacsonyan expresszáló mTEC
miRNS	mikro RNS
MMP	mátrix metalloproteináz enzim
mRNS	messenger RNS
mTEC	medullar thymoeptelial cells, medulláris timoepteliális sejtek
MTX	metotrexát
MZF1	myeloid zinc finger 1, myeloid cinkujj 1-es fehérje
NB	nuclear bodies, nukleáris testek
NF- $\kappa$ B	nukleáris kappa-B
NLS	nuclear localisation signal, nukleáris lokalizációs jel domain
NOD	nem-obéz diabéteszes egérpopuláció
N-TEF	negatív transzkripció faktor
OR	odds ratio, esélyhányados
PAD	peptidil-arginin-deamináz
PADI14	peptidylarginine deiminase type 14
PBM	protein-binding microarray, fehérje-kötődési microarray vizsgálat
PCR	polymerase chain reaction, polimeráz láncreakció
pGE	promiscuous gene expression, promiszkuis génexpresszió
PHD	plant homeodomain
PKC	protein kináz C
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PS	phosphatidil serin, foszfatidil-szerin
P-TEFb	positive transcription elongation factor b
PTX	pertussis toxin
PWM	position weight matrix, helysúlyozott mátrix
QC	quality check
RA	reumatoid arthritisz
RANK	nukleáris kappa-B faktor receptor aktivátora
RANKL	nukleáris kappa-B faktor receptor aktivátorának ligandja
Rf	reumatoid faktor
RNA-pol II	RNS-polimeráz II
SAND	Sp100 AIRE NucP41/75 és DEAF domain
SD	standard deviáció
SE	közös (shared) epitóp
Shn2	Schnurr-fehérje
SIRT1	Sirtuin 1
SLE	szisztémás lupusz eritematózus
SNP	single nucleotide polymorphism, egy pontos nukleotid polimorfizmus
SP	single positive, egyszeresen pozitív
SSc	Szisztémás szklerózis
STAT	signal transducer and activator of transcription
TAP	transporter associated with antigen processing
Tbx21	T-box transcription factor 21, T-box transzkripció faktor 21
Tc	citotoxikus T-sejtek
TCR	T-cell receptor, T-sejt receptor
Teff	effektor T-sejtek
TF	transzkripció faktor
Th	T-helper sejtek
T <sub>M</sub>	memória T-sejtek
TNF	tumor nekrosis faktor
TRA	tissue restricted antigen, saját test-antigének
Treg	regulatórikus T-sejtek
TSA	trial sequential analízis
TSS	transcription start site, transzkripció kezdő helye
UTR	untranslated region, nemtranszlálódó régió
VCAM-1	vascular Cell Adhesion Molecule 1
YY1	Yin-Yang 1

## **I. Bevezetés**

### **I.1. Az érésben lévő T-sejtek autoimmun regulációja**

Az immunrendszer finoman egyensúlyoz a kórokozó által vezérelt immunitás és a saját antigén által vezérelt tolerancia között. Egészséges egyénben ez a két állapot jelenti a válaszkészség spektrumát. Ennek a skálának szélsőséges végeit jelenti a válasz hiánya, immundeficiencia vagy a nem megfelelő túlzott válasz, autoimmunitás vagy allergia, mely szélsőségek felé történő elmozdulás patofiziológiai állapotokat, betegségeket eredményezhetnek. Az immunológiai válaszkészség komplexitása szépen tükröződik a saját antigénnel szembeni tolerancia kialakulásában – immunológiai érésben –, melynek végeredménye a káros autoreaktivitás elkerülése. Sajáttal szembeni immunológiai tolerancia már a neonatális (embereknél a születéstől az első hónapig) periódusban elkezdődik és kialakításában a magas fajlagosságú, a veleszületett immunrendszerrel együttműködő adaptív immunitás központi szereplői, a T-sejtek játszanak központi szerepet. A folyamat szabályozása két szinten történik. Elsődlegesen a centrális tolerancia során, az érésben lévő T-sejtek (timociták) bejutása és érése az újszülött csecsemőmirigyben (tímusz), ezt követően a már érett repertoár további másodlagos szelekciója a limfoid és nem-limfoid szervekben (perifériás tolerancia) (1,2).

#### **I.1.1. Centrális tolerancia**

##### **I.1.1.1. A timociták bejutása és érése a neonatális tímusz kéregállományában**

A timociták érésének helyszíne a neonatális tímusz sztrómája, melynek mérete és aktivitása az életkor előrehaladtával csökken, pubertáskorra már alig kimutatható, a tímusz zsíros elcsökevényesedésével (involúció) gyakorlatilag megszűnik. A posztnatális élet során a hematopoetikus őssejtek egy része timocitává alakul, melyek a tímusz kortiko-medulláris régióján (cortico-medullar junction, CMJ) keresztül jutnak be a perivaszkuláris térbe (1,2). A timociták felületén fejlődésük során változik a sejtfelületen megjelenő molekulák összetétele. A fő fejlődési stádiumokat a cluster of differentiation (CD) markerek, a CD4, CD8 molekulák megjelenése illetve hiánya alapján különböztetjük meg (1-3). Az érés során kezdetben megkülönböztetünk kettős negatív (double negative, DN) sejteket ahol sem a CD4 sem a CD8 nem expresszálódik a T-sejt felületén (1-3). A későbbiekben a kettős pozitív (double positive, DP) sejtek mind a CD4, mind a CD8 molekulákat expresszálják a timocita felületén (1-3). Az érés során végül megkülönböztetünk egyszeresen pozitív (single positive, SP) sejteket, ahol vagy csak a CD4, vagy csak a CD8 jelenik meg a timocitán (így CD4<sup>+</sup> vagy CD8<sup>+</sup>) (1-3).

##### **I.1.1.1.1. $\beta$ -szelekció**

A prezentált antigént felismerő receptor, a T-sejt receptor (T-cell receptor, TCR) kialakulása TCR gének rekombinációjának eredménye, amelynek során a variábilis (variable, V), diverzitás (diversity, D) és a csatlakozó (joining, J) DNS szegmensek váltakozva csatlakoznak az érés során (2,4). Ennek eredménye a TCR  $\beta$ - és  $\alpha$ -lánc, mely minden egyes sejtben egyedi átrendeződés eredménye (2,4).

A neonatális timocita érés első ellenőrzőpontja a  $\beta$ -szelekció (2,4). Ennek során az éretlen DN T-sejteknek kialakul a TCR receptora, melyet tehát a receptor fehérjét meghatározó TCR- $\beta$  és

$\alpha$ -génjeinek rekombinációja és kifejeződése határoz meg (2,4). A folyamat során először csak azok a DN T-sejtek élnek túl a szelekciót, amelyeknek a TCR  $\beta$ -lókusza sikeresen átrendeződött és kifejeződött (2,4). Ekkor lépnek a DP stádiumba. Ezt követően megtörténik a TCR receptor  $\alpha$ -lókuszának az átrendeződése és kifejeződése (4). Ezen a ponton, ha a timocita nem ismeri fel a kortikális timoepiteliális sejtek (cortical thymopitheiális cells, cTECs) és dendritikus sejtek (dendritic cell, DC) felszínén prezentált fő hisztokompatibilitási komplex (major histocompatibility complex, MHC) saját peptidét, apoptózissal elpusztul. Ez akkor történik, ha a TCR nem megfelelően fejeződik ki (4).

#### **I.1.1.1.2. Pozitív szelekció**

Amennyiben a  $\beta$ -szelekciót túlélő  $\alpha\beta$ TCR, DP timociták kis affinitással kötődnek a cTEC és a DC sejteken megjelenő saját peptid-MHC komplexhez, az érésben lévő T-sejt továbbléphet a medullába (2-6). Ennek következménye, hogy a DP timocita SP T-sejtté alakul, vagyis onnantól kezdve csak egy felszíni markert fog termelni (2-6).

#### **I.1.1.1.3. Negatív szelekció**

Előfordulhat az is, hogy a  $\beta$ -szelekciót túlélő  $\alpha\beta$ TCR, DP timociták nagy affinitással kötődnek a saját peptid-MHC komplexhez, ekkor az érésben lévő T-sejt apoptózissal elpusztul. A kéregállományban a negatív szelekciót a DC sejtek végzik (2-6). Az a DN vagy DP sejt, amely túlélte a DC sejtek negatív szelekcióját átalakulnak SP CD4<sup>+</sup> vagy SP CD8<sup>+</sup> sejtekké (2-6).

A túlélő sejtek érési migrációjuk során elérkeznek a CMJ területére, melyen keresztül már SP timocitákként belépnek a tímusz velőállományába, ahol a medulláris timoepiteliális sejtek (medullar thymopitheiális cells, mTECs) kizárólag negatív szelekciót hajtanak végre rajtuk (5,6).

#### **I.1.1.2. A timociták bejutása és érése a tímusz velőállományában**

A tímusz belső medullájában foglalnak helyet az mTEC sejtek. Érési folyamatuk kezdeti stádiumában éretlen velősejtként alacsony affinitású II-es típusú MHC-peptideket tartalmaznak a felületükön (mTEC CD80lo) (2,7). Ekkor kromatinjuk zárt heterokromatin. Ez az állapot önmagában epigenetikus, mely jelzések a DNS-nek olyan kémiai-fizikai – de nem szerkezeti – konformációváltozása, ami nem jelent mutációt, de nagymértékben meghatározza a génexpressziót (2). Ezt követően még nem bizonyított úton – epigenetikus, expressziót megengedő jelzéseket – követően átalakulnak érett mTEC sejtekké, amelyek már magas affinitású MHC-peptideket tartalmaznak a felszínen (mTEC CD80hi) (2,7). Az érett mTEC sejt genomja eukromatin formában kész a transzkripcióra, a sejtek tartózkodási helye migrációjukat követően a külső medulla, ahol kész fogadni az érésben lévő SP timocitákat (5). Az mTEC sejt genomjában olyan génklaszterek foglalnak helyet, melyek a szövet-specifikus antigéneket kódolnak. Az irodalomban jelenleg hiányosak az eredmények arra vonatkozóan, mik is ezek a szövet-specifikus antigének (tissue restricted antigen, TRA) és mennyi van belőlük, átfogó adatbázis még nem készült róluk, megelőző tanulmányok még *in vivo* stádiumban vannak (1,6). Ilyen például a gyomorban egy H<sup>+</sup>/K<sup>+</sup> ATPáz ionpumpa, ami H<sup>+</sup>-iont szállít K<sup>+</sup> ellenében a parietális sejtek apikális membránján keresztül (8). Ennek 1-es  $\alpha$ -láncát (potassium-transporting ATPase alpha chain 1) kódolja az ATP4A gén, kifejezve az ionpumpa  $\alpha$ -láncának megfelelő antigént, melyet az mTEC sejt a MHC-peptidként a



felszínére helyez (8). Ilyen saját, test-antigén a bone morphogenetic protein (BMP) 2 és 4-nek a részletei, melyek felelősek a limfocita homeosztázisért, citokin produkcióért és biztosítja a T-sejt differenciációt (8). Az mTEC sejt továbbá kifejez olyan fehérjét is, mely felelős az embriogenezisért, fertilitásért és a T-sejt érésért (alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase, MGAT1) (8). Az itt bemutatott saját antigének példaként szolgálnak, valódi kvantitatív és kvalitatív adatok nem állnak róluk rendelkezésre, számuk több ezerre tehető, jelenleg nem készült róluk átfogó tanulmány (1,5). Saját antigéneket kódoló gének expressziója tehát az mTEC sejt genomjában a legnagyobb, míg a cTEC vagy DC sejtekben kisebb mértékű, más sejtekben kimutathatatlan, feltehetőleg epigenetikus jelzésekkel represszált (2,5).

#### **I.1.1.2.1. Bejutás és felhalmozódás**

*Ehrlich és mtsai.* valós idejű, 2-foton pásztázó lézermikroszkópos fénykép- és videó felvételekkel sikeresen vizualizálták, hogyan is történik a timociták bejutása és interakciója az mTEC sejtekkel (9). Amíg a kettős negatív (DN) CD4<sup>-</sup>8<sup>-</sup> és kettős pozitív (DP) CD4<sup>+</sup>8<sup>+</sup> sejtek a CMJ területén maradtak és felhalmozódtak, nem jutottak be a medullába, éles határt képeztek (9). Ezzel szemben a SP (single positive) timociták közvetlen a medullába migrálnak a CMJ területéről (9). Ennek magyarázata az, hogy a pozitív szelekción túljutó SP timocitákon kemotaktikus anyagok hatására aktiválódik egy receptor (6,9). Az érésben lévő T-sejtek ezen receptora a 7-es típusú C-C kemokin receptor (C-C chemokine receptor type 7, CCR7) (6). Ehhez a receptorhoz medulláris kemokinek kötődnek (CCL19, CCL21), melyek a későbbiekben kemotaxissal irányítják a T-sejtet a medullába (10,11). CCL19 kemokint az mTEC termeli (9). Jelenleg tisztázatlan a CCR7 hatásmechanizmusa, feltehetőleg további kemokineket vagy integrineket indukál. Átfogó vizsgálatukban *Ueno és mtsai.* megállapították, hogy a CCR7 végsősoron az, ami a timocita migrációt idézi elő a medullába és ennek következményeként fognak a sejtek felhalmozódni (5,12).

Ezt bizonyította *Ehrlich és mtsai.* azon kísérlete, melynek során pertussis toxinnal (PTX) kezeltek SP timocita sejteket (9). A következőket tapasztalták. (1) SP éretlen T-sejtek medullába jutása torzulást szenvedett; (2) mozgásuk sebessége alacsonyabb a medullában, a kezeletlenekhez képest; (3) éretlen SP T-sejtek medulláris denzitása a kortikálishoz képest a PTX kezeletlenek esetében  $6,5 \pm 0,7$  és  $4,9 \pm 2,1$ , míg a kezeletlenek esetében viszont csak  $0,4 \pm 0,2$ , vagyis a PTX-kel kezeltek esetében nincs különbség a timociták denzitásban a kortikális és medulláris régió között, ami azt jelenti, hogy nem történt felhalmozódás a medullában (9).

#### **I.1.1.2.2. Migráció a medullában és negatív szelekció**

A medullában számos, timocita szelekciót végző sejt foglal helyet: migratórikus DC, rezidens vagy szesszilis DC, plazmacitoid DC (plasmacytoid, pDC), B-sejtek és mTEC sejtek (13). Ezek közül szövetspecifikus TRA-t prezentáló antigén prezentáló sejtek (antigen presenting cells, APCs) a migráló, rezidens DC és az mTEC sejtek (5,6). Egy bejutott SP timocita mozgása egy random mintázatot mutató „séta” az mTEC és DC sejtek között – hozzávetőlegesen 10µm/perc sebességgel – ami megfelelő az elegendő mennyiségű APC sejtekkel való találkozáshoz (5,9). Ez átlagosan 5 APC kontaktust jelent óránként, vagyis egy SP timocita az 5 napos medullában tartózkodása alatt több száz APC sejtrel találkozik (1-8). Az mTEC sejtek genomja – szemben más sejtek genomjával – a legnagyobb mértékben expresszál TRA-t, melyet a sejt felszínén prezentál MHC-peptid formában az érkező SP timocitának (2,5,7). A folyamat, melynek során a medulla építelsejt megtermeli a TRA-t, a

promiszkuis génextpresszió (promiscuous gene expression, pGE). Ezekből részesülnek a DC sejtek is az antigen-handover segítségével, melyek így indirekt módon vesznek részt a szelekcióban (2,5,7,13,14).

Amennyiben a mTEC sejthez érkező timocita TCR receptorával felismeri, és nagy affinitással köti az MHC receptoron prezentált MHC-peptidet – vagyis a saját TRA-t – a timocita apoptózissal elpusztul (1,2,5). A negatív szelekció további érdekes aspektusa, hogy maga az mTEC sejt is elpusztul, mert kontaktusba lépett egy fejlődésben lévő, de autoimmun timocitával (1,2,5). Ezzel a végső szelekciós lépéssel a centrális tolerancia két irányból biztosítja az egészséges immunrendszer kialakulását. Elsősorban biztosítja, hogy a keringésbe kizárólag olyan érett T sejt kerülhet, amely nem köti nagy affinitással a TRA-t vagyis nem autoreaktív (1,2,5). Másodsorban biztosítja azt, hogy mindig friss mTEC populáció álljon rendelkezésre a timocita populáció negatív szelekciójához (1,2,5).

### **I.1.1.2.3. A mTEC sejtek promiszkuis génextpressziója**

A pGE esetleges génextpressziót jelent, mely jól kifejezi a centrális tolerancia ezen szakaszának az érdekességét (1,2,5,7). Az mTEC sejtek genomjában expresszálódnak legnagyobb mértékben azok a gének, amelyek a TRA termelésért felelősek. Hozzávetőlegesen eddig 400 TRA-gént írtak le, melyek klaszterekbe rendeződnek a genomon (7,15). Ezek közül egyes gének aktiváltak, míg mások represszáltak, expressziójuk nem teljesen valószínűségi vagy véletlenszerű. Mindemellett vannak olyan génklaszterek, amelyek egyszerre expresszálódnak és adott kromoszómát preferálnak. Jelenlegi álláspont szerint az mTEC sejtek összessége termeli meg a teljes TRA repertoárt, de ha egyet kiemelünk – pl. a fent említett gasztrikus  $H^+/K^+$  ATPáz ionpumpa alfa láncát, mint TRA-t – azt csak az mTEC sejtek 1-3%-ka termeli (1,2,6,7,15). Vagyis egy mTEC sejt nem termeli meg az összes TRA-t, hanem az mTEC populáció összessége mozaikosan fogja expresszálni mindazon TRA-k készletét, amely az SP timociták negatív szelekciójához szükséges. Ehhez az expresszióhoz nélkülözhetetlenek transzkripciós faktorok, hiszton és non-hiszton acetilázok, transzkripciós elongátorok, szerin/treonin kinázok és egyéb fehérjék (7,15).

### **I.1.1.2.4. AIRE fehérje jellemzői**

Az AIRE (autoimmun regulator) transzkripciós faktor a centrális tolerancia, medulláris negatív szelekciójának központi szereplője, a TRA-t produkáló pGE legfőbb, központi irányítója (1-15).

A transzkripciós faktort kódoló génről (*AIRE*) először 1997-ben tettek említést, amikor is két egymástól független kutatócsoportnak először sikerült a gént izolálni és klónozni emberi kromoszómáról (16-18). Az új gént akkor fedezték fel, amikor egy ritka monogénes autoimmun deficienciát okozó mutációkat vizsgálták. A betegség az autoimmun poliendokrinopátiás-kandidiázis ektodermális disztrófia (autoimmune polyendocrinopathy-candidiasis ectodermal dystrophy, APECED) vagy más néven 1-es típusú autoimmun poliendokrinopátia szindróma (autoimmun polyendocrinopathy syndrome type 1, APS-1) (16-18). További vizsgálatok során meghatározták, hogy a kódolt fehérje kettő plant homeodomain (PHD) típusú cinkujj doménnel is rendelkezik, mely szerint egy DNS-hez kötődő, nukleáris transzkripciós faktorról lehet szó (15-18).

Az *AIRE* gén a 21. kromoszóma hosszú karján, a 21q22.3 sávnál található, körülbelül 12,5 kilobázis hosszú, 14 exoniális és 13 introniális régióval egy 545 aminosavból álló fehérjét kódol, mely 58 kDa méretű (7,15). Az AIRE fehérje transzkripciós faktorként működik,

kulcsszerepet játszva a timociták negatív szelekciójának irányításában, mely magában foglalja az mTEC sejtek pGE folyamatát, ahol a DNS-kötő fehérjékkel együtt lehetővé teszi a szervezet teljes szöveti proteomját képviselő TRA-k termelését. Több ezer, klaszterekben rendezett gén expresszióját irányítja (7,15,19).

#### **I.1.1.2.4.1. AIRE fehérje struktúrális doménjei**

Az AIRE multidomén szerkezete alapján azon fehérjék csoportjába tartozik, amelyek képesek a kromatinhoz kötődni és szabályozni a génátírási folyamatát (5,15). Az N-terminustól az C-terminusig az AIRE öt funkcionálisan fontos szerkezeti doménből áll: homogénen festődő régió (homogeneously staining region, HSR) vagy más néven kaszpáz-aktivációs és recruitment domén (caspase-activation and recruitment domain, CARD), nukleáris lokalizációs jel domén (nuclear localisation signal, NLS), Sp100 AIRE NucP41/75 és DEAF domén (SAND), valamint két PHD domén (PHD1, PHD2) (5,15). Az N- és C-terminus között négy Leu-Xaa-Xaa-Leu-Leu (LxxLL) motívum is helyet foglal (5,15).

##### **HSR/CARD-domén**

A CARD domén egy 100 aminosavat magába foglaló szakasz, melyet kódoló AIRE szekvencián zömével itt fordulnak elő az APECED-et okozó mutációk (7,15,20). A domént az AIRE kódoló régiójában az 1,2 és 3. exon régiója kódolja. Funkciója, hogy az apoptózis során a kaszpáz aktivitást biztosítsa, továbbá a kötőpartnerekkel való kapcsolatot szabályozza (15,20). Ez biztosítja, hogy a sejtmagba jutott AIRE fehérje denz, nukleáris testekké alakuljon (nuclear bodies, NB). Ebben a formában olyan erős affinitással kötődik a nukleáris mátrixhoz, hogy DNáz I endonukleáz kezelést követően is intakt marad (15,19). A CARD domén aminosav-terminális régiója a HSR, mely dimerizációt, fehérje-fehérje kapcsolódást irányító szakasz (7,15,19).

##### **NLS domén**

Az AIRE NLS doménje egy hármas NLS motívum, melyet elsősorban az importin  $\alpha$ -cargókomplex (karyopherin), majd közvetlen utána az importin- $\beta$  fehérjék ismernek fel (15). Az előbbi felelős a nukleusz-citoplazma és citoplazma-nukleusz transzportért, míg az utóbbi segíti a transzkripciós faktor maghártyához kapcsolását, majd ezt követően porushoz kíséri a szállított fehérjét (15,19). Az NLS domain szintén lizinben gazdag (15,19).

##### **SAND domén**

Az AIRE fehérje egyik legfontosabb szakasza, mely nélkül nem lenne képes a medullában funkcióját kifejteni. 1998-ban a PHD domént követően, ez volt az AIRE második struktúrális szekvenciája, melyet leírtak (21). Nevét onnan kapta, hogy ezt az evolúciósan konzervált 80 aminosavnyi szekvenciát az Sp100, AIRE-1, NucP41/75 és a *Drosophila melanogaster*-ben is megtalálható deformált epidermális autoregulációs faktor-1 (deformed epidermal autoregulatory factor-1, DEAF-1) illetve annak gerincesekben megtalálható analógját a szupresszin fehérjecsaládok is tartalmazzák (21). Az AIRE-t kódoló aminosavlánc közepén elhelyezkedő SAND domén szerkezete egy globuláris  $\alpha/\beta$  elemet tartalmaz, melynek  $\alpha$ -hélix részében egy konzervált KDWK motívum található (7,15). Ez az elem biztosítja, hogy az AIRE fehérje a DNS foszfátcsoportjaihoz kötődjön, mely a SAND domén legfőbb kromatinfüggő transzkripciós szabályozó funkciója. Ennek segítségével kötődik az AIRE a

medulláris mTEC sejtek DNS-éhez (7,15,22). A kötődést követően elkezdődik a klaszterbe rendezett TRA-gének expressziója, mellyel a medulláris negatív szelekciót biztosítják az ide érkező SP CD4<sup>+</sup> vagy CD8<sup>+</sup> timociták számára (23). Ezt a domént az exon 5,6 és 7-es régió kódolja az *AIRE* kódoló régiójában. Az itt található mutációk és exoniális egy pontos nukleotid polimorfizmusok (single nucleotide polymorphism, SNP) nagymértékben befolyásolhatják az *AIRE* mTEC DNS-hez kötődését és a következményes TRA expresszióját, így a potenciális autoimmun timociták negatív szelekciójának hatékonyságát (24-26).

## PHD domének

Közel a C-terminális véghez, az *AIRE*-t kódoló aminosavlánc két PHD1- és 2-típusú cink-ujj doménnel, egyenként 41 aminosavból álló szakasszal ér véget, melyeket egy prolinban gazdag régió (prolyne rich region, PRR) választ el egymástól (7,15,27). PHD1-et az *AIRE* exon 8 és 9-es régiója, míg a PHD2-öt az exon 11 és 12-es régiója kódolja. Általában a PHD domének leolvassák a kromatin epigenetikus jelzéseit, legfőképp a metiláció mértékét a H3 hiszton fehérjén (7,15,27). Itt a PHD1-es doménon keresztül az *AIRE* fehérje közvetlenül kötődik a metilálatlan H3 hiszton K4-es lizinjéhez (H3K4m0), mellyel génexpressziót aktivál. A PHD2 a PHD1-gyel való szerkezeti hasonlósága ellenére pozitív töltést mutat, ami alkalmatlanná teszi a hisztonokkal való kölcsönhatásra. *Yang és mtsai.* vizsgálatuk során bemutatták, a PHD2 feladata a kromatin szerkezetét meghatározó és a transzkripcióban részt vevő fehérjepartnerekkel való interakció (27). Azonban nem vesz részt az *AIRE* nukleáris transzlokációjában és a pre-mRNS splicing mechanizmusában (27). Eredményeik azt mutatták, hogy a PHD2 leginkább a mTEC sejtek transzkriptomjának kifejeződéséért, vagyis az immunológiai toleranciáért felelősek (27).

## LxxLL motívumok

Az *AIRE* aminosavlánca négy közbeiktatott LXXLL szekvenciát tartalmaz. Ez a motívum azokra a fehérjékre jellemző, melyek nukleáris receptorokhoz kötődnek (7,15,28). Számos transzkripciós koaktivátorban megtalálható, mint például a CREB-kötő fehérje (CREB-binding protein, CBP) és a signal transducer and activator of transcription (STAT) fehérjében (15,28). A transzkripciós események során koaktivátorként és korepresszorként is működik valamint fehér kapcsolatokért felelős (28).

### I.1.1.2.4.2. *AIRE* expressziója és az mTEC fejlődési stádiumai

Az *AIRE* expressziója és az mTEC fejlődési stádiumai közötti összefüggést két modell, a terminális és a progresszív restriktív modell írja le (2,29,30). A terminális modell szerint az embrionális endodermából egy Bi-potent progenitor sejt származik, melyből a cTEC és az mTEC sejtek alakul ki (29). Ezt követően a medulláris epitelsejteknek három populációja alakul ki: mTEC MHC II<sup>lo</sup>, CD80<sup>lo</sup> (lo: low vagyis alacsony mennyiségű MHC II receptort és CD80 markert termelő sejtek), mTEC MHC II<sup>int</sup>, CD80<sup>int</sup> (int: intermedier vagyis köztes mennyiségű MHC II receptort és CD80 markert termelő sejtek) és mTEC MHC II<sup>hi</sup>, CD80<sup>hi</sup> (hi: high vagyis magas vagy kellő affinitású MHC II receptort és CD80 markert termelő sejtek) (29). Ezekkel az mTEC fejlődési stádiumokkal párhuzamosan emelkedik az *AIRE* és a pGE szintje is. Az MHC II<sup>hi</sup>, CD80<sup>hi</sup> mTEC sejteket további sejtpopulációkra oszthatjuk: MHC II<sup>hi</sup> CD80<sup>hi</sup> *AIRE*<sup>negative and/or low</sup> és MHC II<sup>hi</sup> CD80<sup>hi</sup> *AIRE*<sup>high</sup> sejtekre. Előbbiben nem vagy csak alig kimutatható, míg az utóbbiban magasan expresszálódik az *AIRE*. Ez utóbbi mTEC hajlamos apoptózisra, valamint termelheti az "involucrin" nevű anyagot, amely

markere a terminálisan differenciálódott keratinocitáknak. A terminális modell szerint ez a típusú sejt a végleges formája az mTEC-nek (29).

A progresszív restriktációs modell szerint is az mTEC előalakja a Bi-potent progenitor sejt (30). A különbséget az adja, hogy az éretlen mTECben rengeteg féle transzkripciós aktivitás van, beleértve pGE-t és az azt kialakító AIRE expressziót (30). Ahogy fejlődik a sejt a pGE-n kívüli transzkripciós aktivitások folyamatosan elcsendesednek és kiemelik a pGE-t (30). A terminális modellel ellentétben itt azt állítják, hogy a pGE növekedése nem következménye a korábbi érési stádiumoknak, hanem az mindig is magas szintű volt, ahogy az érésben folyamatosan halad előre az mTEC úgy csillapodik és csendesedik el a sejt pGE-n kívüli transzkripciós környezete és hangsúlyozódik ki a pGE (2,29,30).

A két elmélet közül az előbbi látszik valószínűnek, számos bizonyíték alátámasztotta, hogy az AIRE expressziója a terminálisan differenciálódott MHC II<sup>hi</sup> CD80<sup>hi</sup> sejtekre jellemző, ami azt jelenti, hogy az mTEC sejtek körülbelül 30-50%-a lesz AIRE<sup>+</sup>. Gardner és mtsai. ezt azzal bizonyították, hogy olyan *Aire* transzgen egérs populációt hoztak létre, ahol az AIRE expressziója egyet jelent két fehérje, a zölden fluoreszkáló fehérje (green fluorescent protein, GFP) és az islet-specific glucose-6-phosphatase-related protein (IGRP) fehérje expressziójával (31). Flow citométeres eredményeik azt mutatták, hogy a transzgen egerek mTEC sejtjeinek 30-50%-a bizonyult GFP<sup>+</sup>-nak így AIRE<sup>+</sup>-nak (31).

#### **I.1.1.2.4.3. Transzkripciós regulációja**

Az *AIRE* már az embriogenezis korai szakaszában, a három csíralemez megjelenése előtt expresszálódik. Az embrionális tímuszban a CD3<sup>-</sup>CD4<sup>+</sup>IL-7R $\alpha$ <sup>+</sup>-ként azonosított limfoid szöveti induktor (lymphoid tissue inducer, LTi) sejtek szabályozzák az *AIRE* expressziót a korai perinatális mTEC sejtekben, melyek a gestáció 16. napjára alakulnak ki (32,33). A korai mTEC sejtekben az *AIRE* expressziót a tumor nekrozis faktor (TNF) szupercsalád egyik tagja, a nukleáris kappa-B (NF- $\kappa$ B) faktor receptor aktivátora (RANK) és ennek megfelelő ligandja a RANKL indukálja, amelyet az LTi sejtek expresszálnak (32). Posztnatális tímuszban a RANKL ligandot a CD4<sup>+</sup> timociták szolgáltatják (32,34). A RANK-RANKL tengely tehát az egyik legfontosabb, a kapcsolat hiányában nincs *AIRE* expresszió és TRA termelés (32,34).

Transzkripciót továbbá enhancer elemek határozzák meg. Az *AIRE* és tőle upstream *DNMT3L* kódoló régiók között foglal helyet egy 90 nukleotidnyi intergénikus szekvencia, melyről feltételezik, hogy irányíthatja az *AIRE* expresszióját. Ezt a szekvenciát conserved noncoding sequence 1 (CNS1) nevezték el (35). Haljasorg és mtsai. *in vivo* vizsgálatukban bemutatták, hogy a szekvencia eltávolításával az *AIRE* expressziója drasztikusan alacsony a vad típusú egerekhez képest, vagyis a szekvencia meghatározza a transzkripciót (35). További eredményeik szerint a CNS1 szekvencia egy olyan enhancer szekvencia, amin NF- $\kappa$ B kötőhelyek vannak (c-Rel, RelA, p50), melyen az NF- $\kappa$ B elindítja a RANK jelzést, ami pedig aktiválja az *AIRE* expressziót és részt vesz az mTEC sejtek érésében (35). Az *AIRE* promotor régiójában Sp1, AP-1, NF-Y és ETS transzkripciós faktor családok kötőhelyeit találták (36). Posztnatális tímuszban az *AIRE* expresszió leginkább mTEC sejtekre korlátozódik, majd az involúciót követően helyben kimutathatatlan (32,33). Szintén kimutathatatlan az autoimmun timocitával kontaktáló, később funkcióvesztett mTEC sejtekben. Mindezek alapján elmondható, az *AIRE* expressziója térben és időben szigorúan szabályozott (32,33).

Ezt támasztja alá a legutóbbi olyan átfogó analízis, mely az *AIRE* expresszió helybeni (cis) és távoli (transz) regulátoraira irányult. Herzig és mtsai. először azt vizsgálták, melyek azok az epigenetikus, cisz-regulatórikus jelzések, melyek meghatározzák az *AIRE* expressziót (37). Ilyen a DNS-metiláció, mely a DNS-en a foszfátkötéssel guaninhoz kapcsolódó citozinyűrű

(CpG dinukleotid) 5. pozíciójában lévő szénatomját CH<sub>3</sub> molekulával metilálja. Ez a jelenség transzkripció repressziót vagy blokkot jelent, melynek során így sem a transzkripció faktorok sem pedig az RNS-polimeráz II (RNA-pol II) nem képes a promóter régióhoz hozzáférni, vagy ha mégis, C-terminális doménjének foszforilációja elmarad így a mediátor komplexhez láncolva nem képes elindulni és a pre-mRNS képződés gátlódik. A vizsgálat ilyen metilálható CpG dinukleotidokat talált az *AIRE* transzkripció start helyétől (transcription start site, TSS) 300 bázispárnnyi (base pair, bp) távolságra 5' irányban valamint az *AIRE* lókuszon exon 2-es régiójában (37). Eredményeik alapján, MHC II<sup>hi</sup> CD80<sup>hi</sup> (mTEC<sup>hi</sup>) sejtekben ezek a régiók demetiláltak, vagyis nincs metilcsoport a CpG dinukleotidokon, a transzkripció ezekben a sejtekben epigenetikusan engedélyezett (37). MHC II<sup>lo</sup> CD80<sup>lo</sup> (mTEC<sup>lo</sup>) sejtekben ennek a fordítottja volt jellemző, mely eredmény szintén a terminális modellt erősítette és kiemelve az exon 2-es régió fontosságát az *AIRE* expresszióban (37). A továbbiakban MHC II<sup>lo</sup> CD80<sup>lo</sup> sejtekben az *AIRE* lókuszon vizsgálatok során egy speciális protein, a CCCTC-kötő faktor (CCCTC-binding factor, CTCF) jelenlétét sikerült kimutatni, mely egy inzulátor fehérje. Az ilyen típusú szekvensspecifikus fehérjék kötődésükkel megváltoztatják a kromatin architektúráját, genomialis hurkok képzésével megszakítják az enhancerek és promoterek közötti kommunikációt, így a transzkripciót blokkolják (38). Ennek pozíciója a MHC II<sup>lo</sup> CD80<sup>lo</sup> sejtek DNS-én az *AIRE* lókuszon TSS szekvenciája és a 3' vég (37). MHC II<sup>hi</sup> CD80<sup>hi</sup> sejtekben a CTCF jelenléte nem kimutatható a TSS és 3' végen, mely a demetilált CpG dinukleotidokkal együtt a transzkripciót serkenti, képesség teszi a MHC II<sup>hi</sup> CD80<sup>hi</sup> sejteket *AIRE* expresszióra. *Herzig és mtsai.* azt vizsgálták, melyek azok a transzkripció faktorok, melyek mTEC<sup>hi</sup> és mTEC<sup>lo</sup> sejtekhez kötődnek (37). Ehhez egysejtes ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) szekvenálási technikát használtak, mely módszer egyes sejteket elemel, és azonosítja azokat kromatin régiókat, amelyekhez a bejuttatott transzpozáz enzim hozzáfér (39). Ezek a régiók nem kompakt nukleosómák, hanem elsősorban promóter, enhancer, silencer vagy inzulátor régiókban transzkripció faktorok kötődnek hozzájuk (39). Ez alapján az egyes sejt típusok különbséget mutatnak, mely transzkripció faktorokat preferálják (39). *Herzig és mtsai.* felfedezték, hogy a mTEC<sup>lo</sup> sejtek *AIRE* lókusza zárolt vagyis enhancerekhez és promóter régiókhoz kötődő transzkripció faktorok nem mutathatók ki. Ellenben az mTEC<sup>hi</sup> sejtekben az *AIRE* lókuszon két pontján, a TSS és a CNS1 régióhoz transzkripció aktivátorok sokasága egy nagy multimolekuláris komplexként kapcsolódik úgy, mint T-box transzkripció faktor 21 (T-box transcription factor 21, Tbx21), transzkripció faktor 7 (transcription factor 7, Tcf7), interferon regulatorikus faktor 4 és 8 (interferon regulatory factor 4,8 Irf4,8) (37). Továbbá megerősítették, hogy a CNS1 régió egy enhancer, melyhez az mTEC<sup>hi</sup> sejtekben NF-κB kötődik 3 kilobázisra (kilobase, kb) 5' irányban a TSS régiótól (37).

#### **I.1.1.2.4.4. Poszttranszkripció regulációja**

Transzkripció mechanizmusok mellett poszttranszkripció folyamatok is befolyásolják fehérje szinten az *AIRE* expresszióját. Ebben többek között központi szerepet játszik a jumoni domain containing 6 (*JMJD6*) enzim (40). *Yanagihara és mtsai.* vizsgálati elrendezésükben vad típusú (*Jmjd6*<sup>+/+</sup>) és KO (*Jmjd6*<sup>-/-</sup>) egerek tímuszának sztrómáját *in vitro* 2-deoxyguanozinnal (2-DG) kezelték RANKL jelenlétében és a nélkül (40). *Jmjd6*<sup>-/-</sup> KO egerek *AIRE* expressziója statisztikailag szignifikánsan elmaradt a vad típusúaktól még RANKL jelenlétében is. A következő lépésben RNS-szekvenálással (RNA-seq) megvizsgálták a 2-DG+RANKL kezelt *JMJD6*<sup>-/-</sup> és vad típusú populáció sztrómális transzkriptómáját (40). Dacára annak, hogy a *JMJD6* expresszió változatlan 1850 gén indukálódott melyből 1020 gén expressziója szignifikánsan magasabbnak bizonyult a vad

típusúaknál (40). A vizsgálat utolsó lépéseként pre-mRNS splicing analízist végeztek (40). Ennek során a detektált 1850 génnek 84708 intronját vizsgálták, melyből 1058 magasan expresszáldót választottak ki. Ezen belül 57 olyan intronikus szekvenciát találtak, melyek RANKL stimuláció hatására inkább *Jmjd6*<sup>-/-</sup> KO egereknél expresszáldódtak (40). Ezek között szerepelt az AIRE intron 2-es szekvenciája, mely a RANKL stimulált *JMJD6*<sup>-/-</sup> KO egereknél szignifikánsan 1,9-szer magasabb a vad típusúakhoz képest (40). *Yanagihara és mtsai.* a jelenséget az intron retenciónak tulajdonítja, melynél a splicing során az intron régió nem vágódik ki a transzkriptumból (36). *Jmjd6*<sup>-/-</sup> KO egereknél megnőtt a 2-es és 9-es intron régió retenciója és következményesen csökkent az AIRE expresszió illetve 23 AIRE-függő TRA expressziója (40). Ebből következik, hogy a normál vad típusú egérben a JMJD6 gátolja az intron retenciót és fenntartja a normál AIRE expressziót. Működését tekintve U2 small nuclear RNA auxiliary factor 2, U2AF65 splicing fehérje lizin oldalláncának hidroxilálásán keresztül befolyásolja a splicing folyamatát és határozza meg az expressziót (41). *Yanagihara és mtsai.* 2 évvel később megállapították, hogy a *Jmjd6*<sup>-/-</sup> KO egereknél autoimmun diabétesz alakult ki. AIRE expresszió az mTEC sejtekből nem volt kimutatható, az expresszió hiányát jelen esetben is az intron 2-es régiójának retenciója okozta (42).

A fent említetteken kívül az utóbbi években az AIRE posztttranszkripciók regulatorának bizonyultak a miRNS-ek, melyek 21-25 nukleotid hosszú, kettős szálú, fehérjét nem kódoló struktúrák. RNS interferenciájuk során egy endonukleáz aktivitású RNS-indukált némító fehérjekomplexbe töltődnek be, mely miután a miRNS hozzákötődött a célgen mRNS-ének 3' nemtranszlálódó régiójához (untranslated region, UTR), elhasítja azt így csökkentve a mRNS mennyiséget (43). *Tanaka és mtsai.* *in silico* miRNS-mRNS interakciós vizsgálatok során termodinamikailag stabil hibridizációt fedeztek fel az AIRE mRNS 3'UTR régiója és a miR-155 között, így feltételezték, hogy az hatással lehet az expresszióra (44). Vizsgálatuk során szintetizált, endogén miR-155-öt mimikáló szekvenciával transzfektáltak egér mTEC 3.10 sejteket, ahol a negatív kontrollhoz képest szignifikánsan alacsonyabb AIRE mRNS és fehérjeszintézist mutattak ki (44). Továbbá RNS-szekvenálással a teljes transzkriptomból 311 gén mRNS-e downregulált, melyek közül 58 TRA és 5 a timocita migrációért felelős mRNS volt (44). Annak érdekében, hogy az utóbbit vizsgálják C57BL/6 egerek tímuszából 10 µm-es pórusméretű membránt használva timocitákat szeparáltak, majd azokat Transwell kamrákban kétféle 0,5 µm pórusméretű polikarbonát membránt tartalmazó betétekre helyezték (44). Az egyik betét olyan táptalajba került, mely kezeletlen mTEC 3.10 sejteket tartalmazott, a másik miR-155-5p-t tartalmazó mTEC sejteket (44). A vizsgálat végén, a betéteken átjutott timocitákat képkötő módszerrel számolták le. Transwell kamrákban a miR-155-5p-vel kezelt táptalaj esetében szignifikánsan alacsonyabb számú timocita jutott át a betéten (44). Az eredmények alapján miR-155 nem csak az AIRE mRNS és fehérjexpresszióját hanem a timociták migrációjának sebességét is meghatározza (44).

#### **I.1.1.2.4.5. Transzportja a citoplazmából a sejtmagba**

Az AIRE elsősorban az mTEC sejtmagjában, denz magtestek formájában helyezkedik el és irányítja a TRA termelődését. Mikrotubulusokhoz kötődve másodlagosan megtalálható a citoskeletonban (15). Amennyiben a sejtmagban szükség van utánpótlásra, a citoplazmából transzportfehérjék segítségével jut be. Az AIRE fehérje N-terminális végén helyet foglal egy háromoldalú (tripartite) NLS motívum (15). Ezek közül *in vitro* vizsgálatokban a 131KRK133 NLS szakaszon tapasztalták az importin- $\alpha$  fehérjék kötődését (45). Importin  $\alpha$ -nak van egy flexibilis N-terminális importin- $\beta$ -binding doménje és egy nagyméretű NLS-binding doménje, amely 10 tandem Armadillo (AMR) repeat szekvenciát tartalmaz (20). Ez kötődik az AIRE 131KRK133 aminosav szekvenciájához. Legerősebben az importin  $\alpha$ 3 és  $\alpha$ 5, leggyengébben

az  $\alpha 7$  kötődik (20). Az importin- $\alpha$  fehérjék kötődését követően importin- $\beta$  segíti az AIRE fehérjét hozzákapcsolni a maghoz a RAN-GTPáz útvonalon (20). Ezen a helyen a komplex egy 9nm-es csatornához viszi a fehérjét, ahol csak 30-60 kDa-os fehérjék képesek áthaladni (20). Az AIRE 57,7 kDa-os méretével így a magba jut, ahol NB-átalakuláson megy keresztül. A nukleáris göccá alakulását és az erős kötődést az AIRE fehérje CARD doménje biztosítja (46). Itt a nukleáris mátrixhoz olyan erős affinitással kötődik, hogy DNase I endonukleáz kezelést követően is érintetlen marad (46).

#### **I.1.1.2.4.6. Az AIRE-mediált TRA kifejeződés fehérjepartnerei**

Az AIRE pontszerű, nukleáris magtestek formájában számos partner fehérjével együttesen fejt ki a hatását a medulláris epitélsejtek DNS-én. Ezen partnerfehérjék is részt vesznek az AIRE poszttranszkripció regulációjában, de a fent említett regulátorokhoz képest azzal a különbséggel, hogy egyben a TRA expressziót is meghatározzák (47). Ezekhez a regulátorokhoz az AIRE különböző régiói kötődnek és komplexet képeznek. Számos tanulmány készült a partnerfehérjék és azok funkcionális jelentőségének tisztázására. Ezen a területen nagy előrelépést jelentett *Abramson és mtsai.*-nak a tanulmánya, ahol immunprecipitációval, tömegspektrometriával és RNS-interferencia alapú mRNS degradációval meghatározták az AIRE partnereit (47,48).

#### **CBP**

Ubikviter hiszton- és nonhiszton acetiláz enzim, hatásával transzkripciót aktivál, az AIRE fehérjével szinergiában az expresszió szintjét növeli. RANK-RANKL interakció során a CBP a citoplazmából a sejtmagba vándorol és az AIRE magtestébe lokalizálódva, annak CARD régiójához kötődve indukálja a mTEC sejtek TRA expresszióját (15,47,49). Transzkripció aktivátorként együttműködik a transzkripció irányító általános szereplőkkel úgy, mint az RNS-polimeráz II (47).

#### **DNA-PK**

Számos közlemény a DNS-aktivált protein kinázt (DNA-activated protein kinase, DNA PK) tartja az AIRE legfőbb regulátorának (47). A fehérje kináz aktivitásával foszforilálja az AIRE Thr68 és Ser156 molekuláját (47). Az enzim egy fehérjekomplex tagja (DNA-PK/PARP-1/TOPII $\alpha$ /FACT/RNAPII), mely a DNS-töréseket és nem-homológ végek javítását végzi (50). Proteomikai vizsgálatok alapján elmondható, hogy a komplex tagjai nagyfokú kapcsoltságot mutatnak az AIRE és a DNA-PK fehérjékkel. Ilyen molekula a DNS topoizomeráz II $\alpha$  (DNA-TOPII $\alpha$ ), amely enzim működése során kettős töréseket hoz létre a nukleoszomális DNS-en miközben a helyszínre vonzza a DNA-PK-t (7). Ez elősegíti a kanonikus hisztonvariáns átmeneti eltávolítását, annak érdekében, hogy a transzkripció elkezdődjön. DNA-PK lazítja a környező kromatin régiókat valamint nem-kanonikus hisztonvariáns-cserét is elősegít a kanonikus hisztonvariáns átmeneti eltávolításával azért, hogy a DNS száltörést megjavítsa (15). Továbbá biztosítja az AIRE utánpótlást TRA expresszióhoz, ami azt mutatja, hogy elengedhetetlen partner a megfelelő mennyiségű saját antigén termeléshez (47). Az elmúlt években egyes esetleírások arról számoltak be, hogy egyes APECED betegek hátterében a DNA-PK katalitikus alegységét kódoló DNS szekvencia mutációi a felelősek (47).

#### **P-TEFb**



Az AIRE önmagában nem képes elindítani a TRA expressziót, úgy idézi elő, hogy felszabadítja az RNS-polimeráz II-t (15,20,47). Ehhez érkezik segítségül a positive transcription elongation factor b (P-TEFb) fehérje, melyhez az AIRE karboxil-csoportot tartalmazó C-terminálisával kapcsolódik (20). A transzkripció az RNS pol II utánpótlással kezdődik a TSS régiótól, a preiniciációs komplex (PIC) részeként. Ezen a ponton bár a transzkripció elkezdődik, az elongációt blokkolja a negatív transzkripciós faktor (N-TEF). Ezt az N-TEF-et foszforilálja a P-TEFb (pozitív transzkripciós elongációs faktor b) így ez a változás hozzáférhetővé teszi a DNS-t az elongációt végző RNS pol II-nak. Amennyiben ezek az események elmaradnak *Yoshida és mtsai.* kimutatták, hogy egy bromodomain-containing protein 4 (Brd4) fehérjehidat képez az AIRE és P-TEFb között, mely stabilizálja a kettejük kapcsolatát, egyensúlyban tartja az AIRE CARD régiójának foszforilációját és acetilációját.

## **HNRNPL**

A heterogeneous nuclear ribonucleoprotein L-nek (HNRNPL) szerepe az alternatív splicingban van, a 3'-UTR felismeréséért és RNS elongációért felelős, kapcsolódik az AIRE-rel és P-TEFb-vel (47,52). Általa biztosított a magas mRNS-diverzitás a TRA expresszió során az alternatív splicing során. Az AIRE-rel való partnerséget *Giraud és mtsai.* fedezték fel (52). Vizsgálatuk során egy AIRE-függő génexpressziós rendszert építettek ki, melynek során genom szintű lentivirális RNS-interferencia könyvtárat hoztak létre, amihez kettősszálú kis hajtű (short hairpin, sh) RNS-eket használtak (52). Ennek célja feltérképezni azokat az elemeket, amelyek a kromatin architektúrájához/funkciójához, a transzkripcióhoz és az mRNS-feldolgozáshoz szükségesek (52).

## **ATF7IP-MBD1-ESET komplex**

A fehérjekomplex magába foglalja a the activating transcription factor 7-interacting protein (ATF7IP), methyl-CpG-binding domain protein 1 (MBD1) és az Erg-associated protein with SET (ESET) fehérjéket (47,53). Az AIRE a PHD aminosavszekvenciájával specifikusan kötődik a H3K4me<sub>0</sub>-jelhez, vagyis a H3 hiszton 4. Lys aminosav metilálatlan oldalláncához. Ez a hisztonjel egy génexpressziós jel, melyhez a kapcsolódás önmagában nem elég (53). Az AIRE indirekt módon kötődik egy másik epigenetikus jelhez, SAND doménjén keresztül kooperál az ATF7IP-MBD1-ESET fehérjével (53). Ez a komplex, amikor nincs kapcsolatban az AIRE-rel egy represszív komplexként van jelen, ahol az MBD1 domén folyamatosan represszál a CpG szigetek metilációjával, az ESET tag pedig hiszton metiltranszferáz aktivitása révén, ezeken a helyeken fenntartja a transzkripciós tiltást, mely helyek a TRA génszakaszok, így alapesetben némítottak (53). AIRE úgy használja a komplexet, hogy felismeri SAND doménjével az ATF7IP tagot, mely az MBD1 segítségével felismeri a specifikus metilált CpG szigeteket (53). Amennyiben ezt megteszi, akkor már meg is célozta az AIRE ezen keresztül azt a TRA génszakaszt, aminek a transzkripcióját indirekt módon el fogja indítani a megrekedt RNA-pol II felszabadításával, valamint a DNA-PK transzkripciót aktiváló funkciójával és egyéb partnerekkel (53).

## **SIRT1**

Az AIRE számos fehérjével lép kölcsönhatásba, melyek poszttranszlációs módosításokon keresztül biztosítják az AIRE működését a TRA expresszió során (47). Például vannak olyan fehérjék melyek acetil- és deacetyltranszferáz aktivitással rendelkeznek. Ilyen a Sirtuin 1 (SIRT1), melynek expressziója az mTEC sejtekben a legmagasabb (47,54). Hatásmechanizmusa által deacetylálja az AIRE lizin oldalláncait, mely esemény aktiválja az

AIRE-mediált TRA transzkripciót (54). *Chuprin és mtsai.* kísérletesen *SIRT1* némítással elérték, hogy az mTEC sejtek TRA transzkripciója úgy álljon le, hogy ezalatt az AIRE expressziója változatlan marad, ezáltal bizonyítva, hogy a SIRT1 szerepének végpontja a TRA termelődés az AIRE aktivációján keresztül (54).

#### **I.1.1.2.5. A TRA útja a sejtmagtól a medulláris sejtek MHC receptoráig**

A pGE TRA termelésének egyediségét az adja, hogy ilyen sajátos jellegű génexpresszió jelenlegi tudásunk szerint bizonyítottan a centrális tolerancia során fordul elő, a tímusz involúcióját követően pedig kimutathatatlan (55). Egyes közlemények feltételezik, előfordulhat B-sejtekben illetve perifériás, szekunder limfoid szervekben úgy, mint nyirokcsomókban és a lépben bár ezek esetében pusztán *AIRE* expressziós csúcspontokat sikerült kimutatni, TRA-kat nem (55).

A pGE pontos mechanizmusára számos elmélet született. A magyarázó, jelenlegi modelleknek egyaránt vannak előnyei és hátrányai. Egyik legfőbb akadály, hogy semelyik sem képes jelenleg minden igényt kielégítően megmagyarázni, hogy az AIRE és partner komplexei mi alapján választják ki, adott klaszteren belül, mely TRA-gén expresszálódik és melyik marad néma. Kezdeti feltételezések szerint az expresszió véletlenszerű repressziós oldások eredménye, a szervezet saját antigénjeinek termelődése a tímuszban egy "zaj", melyet az AIRE kihangsúlyoz, biztosítva a negatív szelekció hatékonyságát (7). Később rájöttek, hogy ennél többről van szó. *Meredith és mtsai.* egysejtes RNS-szekvenáláson és genomális DNS-metilációs vizsgálatokon keresztül azt vizsgálta, mutatnak-e különbséget a pGE prezentált TRA-k szerkezetükben, mennyiségükben az AIRE-t termelő és nem termelő mTEC sejtek között (56). Eredményeik azt mutatták, hogy az AIRE által regulált TRA gén-pool teljes mértékben csak kisszámú mTEC sejtben indukálódott (56). A jellemző inkább az volt, hogy egyes mTEC sejtekben a TRA-géneknek igen kis hányada expresszálódott, mely kvalitatíve sejtenként és egyedenként is változó volt. Ez adja a pGE sztochasztikus jellegét (56). *Pinto és mtsai.* ugyanakkor felfedezték, hogy egyes sejteken belül jellemző volt az, hogy a TRA-gének nem külön-külön, hanem egymással határos mikroklaszter-lókuszkba rendezetten, együtt és egyszerre expresszálódtak, vagyis nem egy TRA-gén, hanem maga a klaszter mutatott expressziós csúcspontot, mely jelenség ellentétben a sztochasztikus jelleggel a determinisztikusságot erősítette, vagyis azt nem lehet kijelenteni, hogy a folyamat teljes mértékben valószínűségi (57). Ugyanakkor, ezekben az egymás melletti génklaszterekben semmilyen közös tulajdonság nincs (57). Ezt a közös, együttes expressziót a 19. kromoszómán fedezték fel, melyet *Pinto és mtsai.* „kromoszomális predilekciónak” nevezték el (57). Mind ezek alapján jelenlegi tudásunk szerint az AIRE indukált pGE és a következményes TRA termelődés egyszerre sztochasztikus és rendezett (47,57).

A pGE során létrejön a TRA transzkriptuma, amiből majd a fehérjét adó aminosavsorrendnek kell kialakulnia a transláció során. Ezt követően a fehérjeszintézis klasszikus útvonalán halad a folyamat, melynek végeredményeképpen a TSA kikerül az mTEC sejt felszíni MHC II receptoraira. Annak meghatározására, hogy hogyan is történik a saját antigének betöltődése az MHC receptorokba két elmélet született. Az egyik elmélet szerint az endoplazmatikus retikulumból (ER) a proteaszóma és transzporter komplex (transporter associated with antigen processing, TAP) tölti be az MHC II fogadó egységeibe (58). A másik elmélet szerint a termelt TRA autofágiával töltődik be az MHC receptor fogadóegységeibe. A két elméletben közös tulajdonság annak állítása, hogy a TSA először az ER-ből lizoszómába kerül majd onnan az MHC II betöltő egységeibe (58). A két elmélet kapcsán *Klein és mtsai.* összefoglaló közleményükben megállapították, hogy a TRA betöltődésnek négy típusa különíthető el (58). A konvencionális MHC I „loading” folyamata során a citoplazmatikus fehérjeantigént a

proteaszóma proteolitikusan kisebb peptidekre hasítja (58). Ezt követően a transporter associated with antigen processing (TAP) transzporterek ezeket a peptideket az ER-be transzlokálják, ahol hozzáférnek az újonnan szintetizált MHC I molekulákhoz (58). A peptiddel töltött MHC I molekulák ezután a Golgi-apparátuson keresztül a sejtfelszínre kerülnek, ahol az odaérkező SP CD8<sup>+</sup> és CD4<sup>+</sup> timociták felismerik őket (58). A konvencionális MHC II „loading” során az extracelluláris fehérjeantigéneket fagocitózis vagy pinocitózis útján kerülnek felvételre. Ezek az endocitotikus kompartmentek ezután összeolvadnak a lizoszómákkal, amelyek enzimeket szállítanak az endoszómális tartalom proteolitikus fragmentálásához (58). Annak érdekében, hogy a rendszer elkerülje a MHC II molekulákba történő idő előtti TRA betöltődést, invariáns láncok kötik a saját antigéneket az ER-hez (58). Amint az MHC II tartalmú vezikulák fuzionálnak endoszómális vagy lizoszómális kompartmentekkel az invariáns lánc degradálódik és a TRA peptidek bekötődnek az MHC II molekulák antigénkötő barázdáihoz (58). Az MHC II tartalmú vezikulák endoszómális/lizoszómális kompartmentekkel való fúziója az invariáns láncok degradációjához vezet és lehetővé teszi a peptidek hozzáférését az MHC II molekulák antigénkötő barázdájához (58). A nemkonvencionális, exogén MHC I „loading” (kereszt-prezentáció) során az antigén endocitózissal a citoszólba kerül majd a proteaszóma emésztését követően a TAP transzporterek konvencionális MHC I útvonalon folytatják az antigén prezentációt (58). Végül a nemkonvencionális endogén MHC II „loading” 3 részfolyamatból áll. Az első a makroautofágia, melynek során a saját-antigének kettős membránba vannak csomagolva, melyek majd fuzionálnak az endoszóma/lizoszóma rendszerrel (58). A folyamat következő lépésében a lysosomal associated membrane protein 2a (LAMP-2a) transzporter a chaperone mediated autophagy (CMA) szubsztrátjait a lizoszómába tölti (58). Az utolsó lépésben egy még kevésbé tisztázott folyamat, a mikroautofágia juttatja a lizoszómába az intracelluláris komponenseket (58). Mind a 3 folyamatban a közös, hogy citoplazmatikus anyagot szállít lizoszómális degradációra majd azt követően facilitálja az intracelluláris TRA-k felhasználását és betöltését az MHC II receptorokba (58).

#### **I.1.1.2.6. Az autoreaktív timocita lebontása**

Amennyiben a medullába érkező SP CD4<sup>+</sup> vagy CD8<sup>+</sup> timocita a TCR receptorával felismeri és nagy affinitással köti az mTEC sejt MHC receptorán prezentált saját TRA-t, az érésben lévő T-sejt autoreaktivitása bebizonyosodik (2,15,47). Az egyik legszemléletesebb vizsgálatot *Dzhagalov és mtsai.* végezték el, melynek során 2-foton pásztázó lézermikroszkóppal megvizsgálták, mely események követik egymást időben a negatív szelekció során (59). Bone marrow (BM) kiméra egerek tímusz szeleteit inkubálták TCR aktivált timocitákkal 30 percre, majd TRA-t adtak a sejtekhez növekvő koncentrációkban, hogy lássák van-e összefüggés a TRA koncentráció és a timocita apoptózisának időtartama között (59). A saját-antigénnel való kontaktust követően nagymértékű Ca<sup>2+</sup> beáramlás volt mérhető (59). A timocita ekkor nem törlődik azonnal, hanem migrációja során további APC főképp DC sejtekkel találkozik (59). Eközben a vándorlás sebessége és kiterjedése csökken, végül egy 30 µm-nyi, ún. „confinement zone” területre korlátozódik (60). Ekkor folyamatosan újra találkozik azokkal a TRA-kal, amelyek alapján autoreaktívnak bizonyult, a negatív szelekció így biztosítja, hogy az autoimmun timocita a legmagasabb megbízhatósággal bizonyuljon autoreaktívnak (59). A medullában egy timocita hozzávetőlegesen 500 DC sejtrel találkozik 4 napos tartózkodása során (59,60). Vizsgálatok alapján a negatív szelekcióhoz elegendő 5 DC sejtrel való találkozás (60). Két óra elteltével aktiválódik a kaszpáz-2 enzim, majd újabb 3 órával később már mutatkoznak a sejthalál első jelei (59). Növekszik a timocita membránpermeabilitása, a kromatin struktúrája változik, a mitokondriumok szétesnek (59). Ezt követően megindul a

nukleáris kondenzáció, végül foszfatidil-serin (phosphatidil serin, PS) molekulák a membrán belső oldaláról a külső sejtfelületre jutnak, mely egy jelzés a makrofágoknak, elkezdhetik a sejtfragmentek eltakarítását (59).

A vizsgálat tisztázta azt a kérdést, hogy alacsony vagy magas TRA koncentráció szükséges a timocita apoptózisához. Alapesetben az AIRE-függő saját TRA antigének kis koncentrációban vannak jelen a tímuszban, így valószínűtlen, hogy mindenképp magas koncentrációjú TRA peptidbőség kellene ahhoz, hogy a timocita negatívan szelektálódjon (59). A vizsgálat során alacsony TRA peptid koncentráció is elegendő volt arra, hogy leállítsa a timocita migrációt az ún. „mindent-vagy-semmit” elv alapján (59). A timocita abban a pillanatban, ahogy találkozik a negatív szelekciót előidéző liganddal (peptiddel, saját test-antigénnel) kalcium jelátviteli útvonalak nyílnak meg majd a migráció fokozatosan leáll, de ennek mértéke függ attól, hogy a kapott jel erőssége mekkora (59). Nagy koncentrációjú liganddal való találkozás során az APC stimulusa nagymértékű és néhány órán belül a migráció leáll, a sejt elpusztul (59). Kis koncentrációjú liganddal való találkozás során az APC stimulusa kismértékű és a timocita motilitása megjavulhat, visszanyerheti eredeti mozgékonyágát és folytathatja a csecsemőmirigy környezetében a mintavételt egészen addig, amíg az apoptózis és fagocitózis folyamatai el nem érik (59). Más szóval, az alacsony koncentrációjú liganddal való találkozás alacsony jelintenzitást jelent de végsősoron megpecsételi a timocita sorsát (59). Nagymértékben nem lassítja le, így a timocita tovább mintavételezi a medulláris környezetet egészen addig, amíg a fagociták meg nem jelennek, melyek bekebelezve megsemmisítik a haldokló timocita sejtsztruktúráit (59).

A vizsgálat arra a kérdésre is választ adott, hogy melyik folyamat kezdődik előbb, a fagocitózis vagy az apoptózis? Megfigyeléseik alapján arra jutottak, hogy maga a timocita éri el a helyben lévő, szeszilis fagocitákat nem pedig fordítva (59). Ezek a sejtek felismerik a timociták kaszpáz-3 aktivációját és sejtfelületi PS szignáljait, mely jelzés felhívás a bekebelezés elkezdésére (59). *In vitro* eredményeik azt mutatták, hogy a haldokló sejtek belső antigénjeiket sokkal jobban őrzik, mint azt eddig gondolták, vagyis a fagociták bekebelezik a timocitát az apoptózis elején, még a timocita membránjának átszakadása, a végleges nukleáris kondenzáció előtt (59). Ezt kiegészítve feltehetőleg egy fagocita is prezentálhat TRA antigént (59).

*Kirsten és mtsai.* vizsgálatában a timocita apoptózisának további szereplőit fedezték fel (61). Meghatározták, hogy a pozitív szelekció  $Ca^{2+}$ -függő foszfatáz calcineurint igényel, míg a negatív szelekció proapoptotikus Bcl-2 családdhoz tartozó Bcl-2-vel reagáló sejthalál mediátort (Bcl-2-interacting mediator of cell death, Bim) (61). Eredményeik azt mutatták, hogy a Bim expresszió és  $Ca^{2+}$ -szignalizáció szükséges a negatív szelekcióhoz.  $Ca^{2+}$  szükséges a Bim indukciójához, ebből következik, hogy a *Bim* transzkripciót a protein kináz C út indítja (PKC) de nem igényli c-Jun N-terminális kináz (JNK), mitogén-aktivált protein kináz (MAPK), p38 és a MEK/ERK szignáltranszdukciós útvonalakat (61). A vizsgálat során azt találták, hogy egy transzkripció faktor, a forkhead box O3A (FOXO3A) irányíthatja a *Bim* transzkripciót (61). Nem-obéz diabéteszes (Non-obese diabetic, NOD) egerek microarray eredménye igazolta, hogy mutáció vagy SNP következtében a *Bim* mRNS képzése valamint a következményes timocita apoptózis zavart szenved a negatív szelekció során, vagyis az autoreaktívnek bizonyult éretlen T-sejt nem törlődik (61).

A törlésre ítélt timocita apoptózisának további központi szereplője a Schnurr-fehérje (*Shn2*) vagy másik nevén *Hivep2* egy fontos "sejthalál-csillapító" (death-dampener) (62). *Shn2*<sup>-/-</sup> egerek DP timocitái pozitív szelekciós indító jelek hatására nem pozitív, hanem negatív szelekción mennek keresztül (62). Más szóval, normál esetben pozitív szelekció alatt a timocita TCR receptorával felismeri, és kis affinitással köti az MHC receptoron prezentált TRA-t, ezt követően SP timocitává válik és beléphet a medullába, hogy a rendszer elvégezze rajta a negatív szelekciós lépéseket (62). *Shn2*<sup>-/-</sup> egerek esetében a kis affinitású kötődés során

az érésben lévő timocitát a rendszer nem viszi tovább az érésben, hanem negatív szelekcióval lebontja (62). Egészséges, Shn2-t expresszáló egyedekben tehát a protein szerepét úgy kell értelmezni, mint egy féket, mely biztosítja, hogy a pozitív szelekció alatt az egészséges timocita le ne bontódjon negatív szelekcióval (62). Sejthalál-csillapító hatását az is bizonyítja, hogy az Shn2 szintek az MHC-TRA és TCR kis affinitású kapcsolódásokor emelkednek, pajzsként biztosítva az egészséges, nem autoreaktív timocita védetségét a pozitív szelekció alatt (62). Mindezekből következik, hogy a Schnurr-fehérje az egészséges timocita lehetséges őrzője, egy őrszem, mely védetséget jelent az egészségesnek bizonyult az érésben lévő T-sejt számára, mely protektív hatás egészen a medulláris folyamatok megkezdéséig kíséri a timocitát (62).

Az autoreaktív, éretlen T-sejt klonális törlésére tehát a következő megállapítások tehetőek: TRA ligandokkal való találkozást követően megemelkedik az intracelluláris  $Ca^{2+}$  koncentráció; a timocita motilitása térben és időben csökken majd leáll; alacsony koncentrációjú liganddal való találkozás, alacsony jelintenzitást jelent de végsősoron halálraítéli a timocitát, mely a sejthalál megkezdéséig a medulláris környezetet mintavételezi; a folyamat utolsó lépéseként először fagociták kebelezik be a haldokló timocitát majd csak ezt követően kezdődnek meg a klasszikus apoptotikus lépések, úgy mint a membrán átszakadása és a végleges nukleáris kondenzáció (59-62).

### I.1.2. Az érett timociták kijutása a perifériára

Miután azon T-sejtek, melyek túlélték a negatív szelekciót, vagyis változatlanul kis affinitással vagy egyáltalán nem kötődtek az MHC receptoron prezentált TRA peptidekkel, felsorakoznak a CMJ perivaszkuláris tér kapillárisainak mentén egyvonalban. Innen megkezdik emigrációjukat, melynek során diapedezissel áttörnek és belépnek a perifériás keringés kapillárisaiba, ahol készek az adaptív immunválaszra (63,64). Ehhez a kijutáshoz alapvető fontosságú a sphingosine-1 phosphate (S1P) ligandfehérje, melyet  $CD4^+$  és  $CD8^+$  érett, naív T-sejtek termelnek a felületükön (63).  $S1p1^{-/-}$  egerek embrionális állapotban elpusztultak vaszkuláris defektusban (63). Továbbá  $S1p1^{-/-}$  egereknél a T-sejtek hiányoztak minden perifériás nyirokszervből dacára annak, hogy a tímusz jól működött (63). Az S1P1-nek a transzkripcióját a Krupper like factor2 (KLF2) transzkripciófaktor irányítja, melyet pedig az ERK5 szignalizáció kapcsol be már magzatkorban (64). Ez eredményezi azt, hogy a pozitív szelekciót követően csak késve, hozzávetőlegesen 4 nappal aktiválódik a KLF2 és általa az S1P1 (64). Azt is megállapították, hogy az IL-7 stimuláció az ERK5 foszforilációját okozta, ami egy olyan modellre utal, amelyben az IL-7 jelátvitel aktiválja az ERK5-öt és indukálja a KLF2 expresszióját (64). Az IL-7 jelek DP timocitákban aktívan elnyomódnak és a pozitív szelekciót követően újra érzékenyek lesznek, ami magyarázatot adhat a KLF2 expresszió időzítésére (64,65). Az érett, naív T-sejtek emigrációja a tímuszból tehát egy szigorúan rendezett, lineáris folyamat melyet egy futószalaghoz lehetne hasonlítani, szemben a korábbi elgondolásokkal, melyek szerint a kijutás pusztán eseti és sztochasztikus (65). A kijutást követően az érett és naív T-sejtek másodlagos nyirokszervekbe jutnak, ahol a B-sejtektől elkülönülve várják az adaptív immunválasz során a feladatukat. Az antigénnel történő kölcsönhatás során aktiválódnak, proliferálódnak és differenciálódnak  $CD4^+$  T helper (Th),  $CD8^+$  citotoxikus T-sejtekké (Tc),  $CD4^+$  regulatórikus T-sejtekké (Treg) és  $CD4^+$  vagy  $CD8^+$  memória T-sejtekké ( $T_M$ ) (65,66). Treg sejtek közül megkülönböztetünk a tímuszból származó  $CD4^+$  és  $CD25^+$  (natural) nTreg és más sejtek által indukált iTreg sejteket (66). Az autoimmun betegségek indukcióját a közelmúlt eredményei alapján a tímuszból származó nTreg sejtek képesek féken tartani (66).

## I.2. Az AIRE szerepe az autoimmun kórképek kialakulásában

### I.2.1. *In vivo* manifesztációk

Az autoimmunitást koordináló központi szerv a tímusz, így annak vagy egyes elemeinek kiiktatása autoimmun betegségeket okoz. A születést követő 3. napon, a neonatálisan timektomizált experimentális egerek esetében leggyakoribb a szervspecifikus autoimmunitás, azon belül is az autoimmun gasztritisz, melyet a  $H^+/K^+$  ATPáz ionpumpa mint autoantigén ellen az autoreaktív  $CD4^+$  T-sejt által termelt autoantitestek okoznak (67,68). Feltételezések szerint az ok az lehet, hogy a timektomizált egér képtelen az autoreaktív  $CD4^+$  T-sejteket eltávolítani. Az autoreaktív T-sejtek kis populációja így kölcsönhatásba lép az APC sejtek MHC receptorain megjelenő saját antigénnel, melynek megfelelően aktiválódik és proliferálódik. Ezt követően nem-limfoid szervekbe vándorolnak, ahol citotoxikus autoimmun válaszok útján elpusztítják az útjukba kerülő célsejteket (69). Mivel tímusz hiányában nincs megfelelő mennyiségű tímusz eredetű, fékentartó nTreg sejt, kialakul az autoimmunitás. Ezt támasztja alá az a megfigyelés, mely szerint közvetlenül a születést követően, de timektómia előtt javarészt  $CD4^+$  és  $CD25^+$  nTreg kering a periférián, míg a 3. napon a timektómiát követően kizárólag  $CD4^+$  de  $CD25^-$  nem prezentáló sejtek detektálhatóak, vagyis a  $CD25^+$  felületi markert prezentálni képes nTreg sejt nem detektálható. Ennek megfelelően timektomizált egérben kevesebb Treg sejt van, mint nem-timektomizáltban (70). Neonatális timektómiát a 3. napon követheti egyéb autoimmun manifesztáció úgy, mint tireoiditisz, oophoritis, epididimorchitis (71-73).

Az első *Aire*<sup>-/-</sup> KO egértörzset 2002-ben hozta létre két egymástól független kutatócsoport azzal a céllal, hogy megvizsgálják, ennek a génnek a hiánya mennyiben érinti a centrális toleranciát. *Anderson és mtsai.* *Aire*<sup>-/-</sup> egerek exon 2-es régióját deletálták, *Aire*<sup>-/-</sup> egerek erre a mutáns allélra nézve homozigóták (74). A szekvenciaelemzés igazolta, hogy gén transzkriptjei rövidebbek a deletálódott exon régióknak megfelelően és ezekből a transzkriptumokból nem keletkezhetett funkcionális fehérje, mivel a deletálódást követően az exon 1 és 3 abnormális, egymás melletti elhelyezkedése frame shiftet váltott ki (74). A hiányzó AIRE fehérje számos szervben nagymértékű limfocita infiltrációt eredményezett úgy, mint a máj portális tereiben, a gyomor parietális sejtjeiben és a retinában (74). Ezen struktúrák autoantigénként detektálódtak, citokineket és autoantitesteket termeltek ellenük a T-sejtek. Microarray vizsgálatok során azt találták, hogy az AIRE hiányában nem termelődött megfelelő mennyiségű és összetételű TRA készlet (74). Amikor csontvelősejteket ültettek vad típusú *Aire*<sup>+/+</sup> és *Aire*<sup>-/-</sup> egerekből letális dózissal besugárzott *Aire*<sup>+/+</sup> és *Aire*<sup>-/-</sup> egerek csoportjaiba szervi infiltrációt és humorális autoimmunitást kizárólag az *Aire*<sup>-/-</sup> recipiensben találtak függetlenül annak állapotától (74). A másik kutatócsoport, *Ramsey és mtsai.* 2002-ben hozta létre az AIRE KO egérpopulációt. Ennek során az APECED betegséget okozó leggyakoribb finn R257X-es mutációt utánozták neomicin-kazetta exon 6-os kezdő régiójába inzertálásával (75). Az egerek betegség fenotípusa viszonylag enyhe volt, klinikai endokrin betegségre utaló jelek nélkül, ugyanakkor számos szervben limfociták infiltrációját, a tímusz és a mellékvesék atrófiáját, sterilitást és szervspecifikus keringő autoantitesteket figyeltek meg (75). A harmadik *Aire* KO egér modellt 2005-ben hozták létre, itt exon 5-12-es régiót célozták meg. Fenotípust tekintve az egerek  $CD4^+$  sejtjei autoreaktívnak bizonyultak az alfa-fodrin ellen, mely a későbbiekben Sjögren-szindróma kiváltó autoantigénjének bizonyult (76). Továbbiakban gasztrikus mukóza limfocita infiltrációját és a fertilitás csökkenését figyelték meg (76). 2009-ben az APECED-et okozó második leggyakoribb mutációt mimikálták az exon 8-os régió 13 bp-nyi deletálásával, mely a PHD domén kódolóhelye, aminek segítségével az AIRE partnerkomplex leolvassa a kromatin H3K4m0 motívumát, mely a

TRA-expresszió megengedő epigenetikus kulcsjelzése (15,27,77). Szervspecifikus limfocita infiltráció, számos szerv elleni autoantitest termelés és hím sterilitás volt megfigyelhető (77). Az APECED egyik környezeti kiváltó oka a *Candida albicans* fertőzés, így megvizsgálták exon 8 *Aire*<sup>-/-</sup> egerek orális és szisztémás fertőzése előidézi-e a betegséget. Fertőzött és nem fertőzött egyedek között nem mutatkozott különbség a betegség fogékonyságát tekintve (77). Dacára annak, hogy az *Aire*<sup>-/-</sup> KO-egerek jól mimikálták a humán APECED-et okozó mutációkat, vizsgálatok különbséget mutattak ki immunpatogén mechanizmusokban és tünetekben az APECED-ben szenvedő emberekhez képest. Az egyik jelentős különbség az, hogy egyik *Aire*<sup>-/-</sup> egérmodell sem produkálta a klasszikus humán APECED tüneti mintázatot, kevésbé súlyos fenotípust mutattak. Továbbá *Aire*-deficiens egerek nem termeltek ugyanolyan autoantitesteket, mint az APECED betegek valamint nem támadják meg a betegek általában érintett szerveit. Például, *Aire*<sup>-/-</sup> egereknél gyakran a májat támadják meg, APECED betegeknél ez ritkán fordul elő (74-77).

Az *in vivo* tüneti inkonzisztencia orvoslására *Ossart és mtsai.* hozták létre 2018-ban az első patkány *Aire*<sup>-/-</sup> törzset, melynél irányított *Aire* gén specifikus deletálást embriókban cink-ujj nukleáz technológiával érték el (78,79). Ennek eredményeként a patkányok produkálták a jellegzetes APECED tüneteket, mint például ektodermális disztrófia (alopecia, vitiligo, körömsorvadás). Szervek tekintetében átfedést mutattak a humán APECED-érintett szervekkel (pankreasz) és a szervspecifikus autantitestekkel úgy, mint I-es típusú interferonok (IFN) és interleukin-17 (IL-17), így a törzs egy APECED állatmodell, melyen lehetséges terápiás megközelítések vizsgálhatók (79).

Sjögren-szindróma (SS) modellezésére – mely egy krónikus, autoimmun exokrinopátia súlyos száj- és szemszárazsággal – *Chen és mtsai.* szintén létrehoztak a humán tüneteket produkáló állatmodellt *Anderson és mtsai.* exon 2-es deletálási technológiáját felhasználva (80). *Aire*<sup>+/+</sup> törzshöz képest a deletáltak eltérő innervációt, lakrimális produkciót és MIST1+ sejtek nagymértékű csökkenését produkálták, mely tünetek konzisztensek humán Sjögren-szindrómával (80).

*AIRE* KO/BALB/cAnNCr1Cr1j (*AIRE* KO/BALB/cAnN) egerek kiválóan mimikálták a humán autoimmun gastro-pankreatitist, melynél limfocita infiltrációt szervspecifikus autoantitest termelést detektáltak, 2-es típusú izomeráz-asszociált diszulfid protein (protein disulfide isomerase-associated 2, Pdia2) autoantigén ellen, melyet tipikusan a gyomor és a hasnyálmirigy termel immundeficiens betegeknél (81).

Alopecia areata (AA) – mely egy hajhullást okozó, gyógyíthatatlan autoimmun betegség – modellezésére exon 2, intron 1 és 3-as régió deletálásával 2023-ban létrehozott C57BL/6J *Aire*<sup>-/-</sup> egérmodellt alkalmaztak, mely egerek közel 62%-a abban a pozícióban produkálta a tüneteket ahol az AA betegek (82).

Reumatoid arthritisz (RA) tanulmányozására *Campbell és mtsai.* 2009-ben hoztak létre *Aire*<sup>-/-</sup> C57BL/6 egértörzset, melynél az exon 8-as régiót deletálták. Kollagén indukálta arthritiszt (collagen induced arthritis, CIA) vad típusú (*Aire*<sup>+/+</sup>) és deficiens (*Aire*<sup>-/-</sup>) populációban csirke típusú II-es kollagén (type II collagen, CII) injektálásával érték el. *Aire*<sup>-/-</sup> egerek súlyos arthritiszt produkáltak, abnormálisan magas anti-CII immunglobulin (IgG) szinttel, hozzávetőlegesen tízszerese a vad típusú szérum anti-CII IgG értéknek (83). Továbbá a két kísérleti csoportból, az *AIRE* deficiens egyedek izolált mTEC sejtei szignifikánsan alacsonyabb (p<0,0001) mértékben expresszáltak CII TRA-t, ennek megfelelően vad típusban a CII-vel szembeni tolerancia nagyobb mértékűnek bizonyult az ízületi porcban szemben az *Aire*<sup>-/-</sup> egerekkel, melyet az egyes kísérleti csoportokban a CIA mértéke jól tükrözött (83). A vizsgálat meghatározta, hogy az arthritiszt előidéző, annak CD4<sup>+</sup> T-sejt és B-sejtes aktivációja *Aire*-dependens (83).

## I.2.2. Humán manifesztációk

Az elmúlt harminc évben az újszülöttkori nyitott szívűtét a kritikus veleszületett szívhibák kezelésének standard módszerévé vált, amelyet gyakran kísérnek egyre népszerűbb perkután beavatkozások. A veleszületett szívhibák korai műtéti beavatkozásai általában timektómiát foglalnak magukban, ami a tímusz részleges vagy teljes eltávolítását jelenti (84). Érintett újszülötteknél ezt a beavatkozást technikai okokból rendszeresen elvégzik, mivel a tímusz mérete és elhelyezkedése akadályozhatja a szívhez való hozzáférést a műtét során. A tímusz eltávolításával a korai posztoperatív időszakban kiküszöbölhető az intrakapszuláris infiltráció által érintett duzzadt mirigy által a szívre és a főbb erekre gyakorolt kompresszió kockázata is (84). Mára a beavatkozást követően az újszülöttek 90%-a megéri a felnőttkort, azonban a sebészeti eljárás immunológiai következményeiről viszonylag kevés követéses vizsgálat készült. Az egyik legutóbbi 2017-ben zárult le, ahol újszülöttkori timektómián átesett gyermekeket követtek 30 évig illesztett kontroll kíséretében, ahol az esetpopuláció alacsonyabb T-sejtszámot ( $0,71 \times 10^9/L$ ;  $2,75 \times 10^9/L$ ) és megemelkedett  $T_M$  arányt (57,43%; 19,72%) mutatott (84). Autoantigén microarray vizsgálatok során képződött autoantitestek az esetpopuláció 62%-ában, míg a controlok 33%-ában volt kimutatható szérumból (84). A témában *Cavalcanti és mtsai.* elkészítették a 2021-ig publikált közlemények szisztémás analízisét, mely alapján elmondható, hogy az 5 év alatti, neonatális timektómián átesett gyermekek szérumban alacsonyabb számú  $CD4^+$ ,  $CD8^+$  és Treg sejt volt jelen, alacsonyabb diverzitású TCR repertoárral rendelkeztek (85). A neonatálisan timektomizált populáció még egy paraméterben mutatott eltérést, ez pedig a T-sejt-receptor-kivágási kör (T cell receptor excision circle, TREC) szintje. A T-sejtek differenciálódását a tímuszban a TCR gének átrendeződésének hierarchikus sorrendje jellemzi, amely a V, D és J génszakaszok egyesüléséhez vezet. Az egyes átrendeződési lépések során az átrendeződő V, D és J génszakaszok közötti DNS-töredékek körkörös kivágási termékeként, úgynevezett TREC-ként, az érési folyamat melléktermékeként vannak jelen, de az érésben már nem vesznek részt (86). A TREC-ek nagy időbeli stabilitással rendelkeznek, a T-sejtek proliferációja során felhígulnak, vérből történő mennyiségi kimutatásuk az újonnan képződött T-sejtek illetve a tímusz aktivitásának markereként használható. Alacsony szintjükből következtetni lehet a nem túl széles TCR repertoárra, mely immun és autoimmun betegségekre jelenthet fogékonyságot. A technológiát szérumban határértékekkel laboratóriumi irányelvként először 2001-ben *Hazenberg és mtsai.* dolgozták ki (87). Azóta rutinszerűen használják újszülöttek súlyos kombinált immundeficiencia (severe combined immunodeficiency, SCID) szűrésére. SCID-ben szenvedő csecsemők TREC szintje nagyon alacsony, alig kimutatható, mely a funkcionális T-sejtek hiányát jelzi. Ugyanakkor korai felismerése lehetővé teszi az azonnali kezelést, például a csontvelő-átültetést, amely életmentő lehet (86). Visszatérve *Cavalcanti és mtsai.*-nak 2021-es szisztémás analíziséhez, az 5 év alatti, neonatális timektómián átesett gyermekek szérumban szignifikánsan alacsonyabb TREC szintet mértek a kontroll csoporthoz képest. Összefoglalva, jelenleg nem készült elegendő követéses vizsgálat annak tisztázására, hogy konkrét autoimmun betegségekre jelent-e kockázatot az újszülöttek timektómiája, mindenesetre a T-sejt repertoár különbözősége és az alacsonyabb TCR repertoárt jelentő alacsonyabb TREC szint a centrális és perifériás tolerancia lehetséges érintettségére enged következtetni (85). Felnőttkori timektómián átesettekről a legátfogóbb 5-éves kohorsz vizsgálatot *Kooshesh és mtsai.* készítették el 2023-ban (88). Ennek során olyan felnőtt betegeket vizsgáltak, melyek kardiológiai sebészeti eljárás után, exponált csoportot itt a timektomizált felnőttek jelentették (1420 fő), míg az illesztett nem exponált csoport (6021 fő) jelentette azon egyének csoportját, melyek esetében a sebészeti eljárás következtében nem volt szükség timektómiára (88). Sebészeti beavatkozást követő 5-éves periódusban az összes halálra kimenetre a relatív rizikó (RR) értéke 2,9 (95%-os konfidencia



intervallum, 95% confidence interval, 95%CI 1,7-4,8) (88). Timektomizált felnőttek a nem exponáltakhoz képest szignifikánsan alacsonyabb CD4<sup>+</sup> és CD8<sup>+</sup> sejtszámot és az azokhoz tartozó TREC szintet produkáltak. Exponált csoport CD4<sup>+</sup> sejthez kötött TREC szintje 1451/μg DNS szemben a nem exponáltak 526/μg DNS (p = 0.009) szintjével (88). Timektomizált felnőttekre jellemző volt a proinflammatorikus citokinek magasabb szintje. A teljes kohorszban az autoimmun betegségek kockázata nem mutatott különbséget az egyes csoportokban (RR=1,1, 95%CI 0,8-1,4), de ha a preoperatív autoimmun betegségtől, daganattól és infekciótól mentes alcsoportot képeztek, a posztoperatív autoimmun betegségek kockázata szignifikánsan magasabb volt a nem exponáltakhoz képest (RR=1,5, 95%CI 1,02-2-2) (88).

### **APECED (autoimmun poliendokrinopátiás- kandidiázis)**

Az AIRE fehérjét kódoló szekvenciában előforduló funkcióvesztéses mutációk az esetek többségében APECED betegséget manifesztálnak. A rendellenesség egy ritka monogénes, autoszomális recesszív, autoimmun poliendokrinopátiás-kandidiázis ektodermális disztrófia (autoimmune polyendocrinopathy-candidiasis ectodermal dystrophy, APECED) vagy más néven 1-es típusú autoimmun poliendokrinopátia szindróma (autoimmun polyendocrinopathy syndrome type 1, APS-1). A betegség kiváltó oka a mutáció miatti AIRE fehérje defektus, melynek eredményeként az érésben lévő timociták közül az autoimmun példányok a negatív szelekció hiányában nem válogatódnak és törölődnek ki, hanem autoimmun naív T-sejtekként kijutnak a periféria (90). Eddig több 100 betegséget okozó nonszensz és misszensz mutációt írtak le. Ezek közül 5'-3' irányban a következő domén-kódoló régiók mutációit érdemes kiemelni. CARD/HSR régió mutációi, p.L28P és a p.L29P-t, melyek döntő fontosságúak a szerkezeti integritás szempontjából, gátolják a nukleáris lokalizációt és az AIRE által közvetített transzkripciót, míg a p.K83E enyhébb hatást gyakorol a transzkripcióra és a lokalizációra, valószínűleg a fehérje-fehérje kölcsönhatások kötőhelyeinek megváltoztatásával, ami magyarázatot adhat a perzsa zsidóknál megfigyelt enyhébb betegség fenotípusra (90). A SAND régió fehérje-fehérje kapcsolatok kialakításáért és a DNS kötődésért felelős, itt számos populáció specifikus mutáció megtalálható úgy, mint a szardíniai vagy a finn típus. Populációsan a leggyakoribb az R257X – vagy új nevén c.769C>T –, melynek hatására az exon 6-os régió citozinjának helyére timin kerül. Ennek következménye, hogy a 257. aminosav mely arginin (R257), stop (terminációs) kodonná (X) alakul, funkcióképtelen fehérjét eredményezve (90). A p.K237A, p.K253A, p.K253E, p.R257A pontmutációk hatására a fehérje a citoplazmában halmozódik fel ahelyett, hogy a sejtmagba transzportálódna (90). 2023. novemberében fedezték fel a legújabb, c.260 T>C névre hallgató AIRE misszensz mutációt, melynek során az Exon 2-es CARD régióban a timin citozinra szubsztituálódik, ennek poszttranszkripciós következményeként a 87. aminosav oldalláncon a Leucint Prolin helyettesíti (91). Az eredmény egy csonkolt, funkcióképtelen AIRE fehérje, mely a 6 éves páciens esetében hypergammaglobulinémiát, malabszorpciós szindrómát, ektodermális diszpláziát, mukokután kandidiázist, vitiligót és hipotiroidizmust okozott, továbbá szérumában anti IFN-ω autoantitesteket találtak mely tovább erősítette az APECED diagnózisát (91).

Az AIRE génben előforduló funkcióvesztő mutációk eredményeként defektust szenved a centrális tolerancia negatív szelekciója, melynek során a perifériára olyan felszíni TCR-rel rendelkező naív T-sejtek jutnak ki melyek autoreaktivitást mutatnak. Később ennek megfelelően autoantitest termeléssel reagálnak a saját-antigénekkel történő találkozást követően (89). Ezek közül a TRA-k közül számos meghatározásra került, melyek közül a leggyakoribb KH domain containing 3 like, subcortical maternal complex member (KHDC3L), regulatory factor X6 (RFX6) és kolipáz (90). Autoreaktivitás ennél a betegségnél

nem kizárólag a saját TRA szöveti antigénekre terjed ki, hanem citokinekre is úgy, mint IFN-család és Interleukin-17 (IL-17), mely az APECED betegek 95%-ában kimutatható (90). IFN- $\omega$  és - $\alpha$  13 szubtypusa ellen termelt autoantitestek az APECED betegek 95%-ában kimutatható. Az IFN- $\omega$  ellen termelt autoantitest további érdekessége, hogy magas szenzitivitással és specificitással kimutatható újszülöttekben még a klinikai tünetek kialakulása előtt (90).

Az első eseteírást 1929-ben publikálták, ahol a közlemény egy olyan négy és fél éves kislánnyól számolt be, aki krónikus tetániás görcsöktől, hipoparatiroidizmustól, orális gombafertőzéstől és szaruhártya fekélytől szenvedett (92). Magát az APECED kifejezést egy 1980-ban publikált esetet követően kezdték használni (93). Diagnózisa az ún. Whitaker-triáddal történik, melynél a krónikus nyálkahártya kandidiázis, hipoparatiroidizmus és mellékvesekéreg-elégtelenség együttállása jelenti az APECED meglétét (92). Prevalenciája az európai és észak-amerikai populációban 1:100000-200000. Genetikailag izolált populációkban magasabb, mint például az iráni zsidók (1:9000), a finnek (1:14 000) és a szardíniaiak (1:25 000) (94). Az első tünet általában a krónikus mukokután kandidiázis, mely átlagosan 5 éves korban fordul elő, a diagnosztikus triád másik két tagja – hipoparatiroidizmus és mellékvesekéreg-elégtelenség – 3-5 és 11-15 éves korban fordul elő. Születéskor várható átlagos élettartama 40 év, nemek eloszlásában nincs különbség (94). Leggyakoribb komorbiditása az I-es típusú diabétesz mellitusz, melyet csökkenő sorrendű előfordulási gyakorisággal: autoimmun hepatitisz, gasztritisz, alopecia areata, psoriasis és vitiligo követ, RA komorbiditása 1-2% (94). Hipoparatiroidizmust orális calcitriol szupplementációval, mellékvese elégtelenséget glükokortikoid és mineralokortikoid, kandidiázist antifungális szerekkel pl. fluconazzal kezelik. A betegség sajátos megnyilvánulása és súlyossága minden egyes betegnél más és más, így terápiája nagymértékben egyénre szabott (94).

### **I.2.3. AIRE allépolimorfizmusok autoimmun kórképekben**

Az első vizsgálat mely az allépolimorfizmusokat összefüggésbe hozta a centrális toleranciával Lovewell és mtsai. publikálták (95). Ennek során az AIRE promotert és a TSS régiójától 5' irányban egy 591 bp-nyi szakaszt vizsgáltak illetve lehetséges allépolimorfizmusok AIRE transzkripcióra mért hatását. Humán limfocita DNS vizsgálatára a denaturáló nagy hatékonyságú folyadékkromaotográfia (denaturing High-Performance Liquid Chromatography, dHPLC) módszerét alkalmazták (95). Ezekben a régióban két allépolimorfizmust találtak AIRE-655R (rs117557896) és az AIRE-230Y-t (rs751032), mely régiókat pCR2.1 TA vektorba ligálták és TOP10 sejtekbe transzfektálták (95). Minden egyes promoterváltozatot a luciferáz riportergén elé, a pGL3-basic vektorba klónozták, hogy létrehozzanak egy AIRE riporterkonstrukciót (95). A transzkripciós vizsgálat eredménye az lett, hogy az AIRE-655G AIRE-230T allépolimorfizmusok alkotta genotípus esetében a promóter aktivitás nagyon alacsony mértékű, szinte alig kimutatható, míg AIRE-655G AIRE-230C genotípus a legmagasabb értéket produkálta. Az érdekessége a vizsgálatnak, hogy ez által kimondható, adott lókuszon allépolimorfizmusai alkotta genotípus képes a transzkripciót oly mértékben befolyásolni, hogy az expresszió alig vagy egyáltalán nem kimutatható (95). Vagyis amennyiben ez lehetséges, akkor az is kimondható, hogy genotípusonként a centrális toleranciát meghatározó negatív szelekció újszülöttkorban egyénekenként eltérhet és adott allépolimorfizmusok alapjaiban határozhatják meg annak hatékonyságát, alacsony AIRE expressziót produkáló genotípus emelheti az autoimmun megbetegedések iniciációját már a posztnatális időszakból.

Az első *AIRE* allélpolimorfizmussal kapcsolatos molekuláris epidemiológiai vizsgálatot 2002-ben *Tazi-Ahnini és mtsai.* végezték el alopecia areata autoimmun betegségek vizsgálva (96). A kórlefolyás során T-sejt populáció szűrődik a folliculusokba, melyek autoantitest és citokintermeléssel kopaszodást idéz elő. Ennek egy súlyos fajtájánál - alopecia universalis - azt találták, hogy a G961C allél frekvenciája a 202 betegnél szignifikánsan magasabb a 175 kaukázusi kontrollhoz képest (96). Szisztémás szklerózis (SSc) vagy más néven szkleroderma egy komplex autoimmun betegség, amelyet fibrózis, vaszkuláris rendellenességek és immunrendszeri aktiváció jellemez T-sejt infiltrációval és B-sejt aktivációval, melynek következménye autoantitest és citokin termelés (98). 2007-ben 41 SSc beteg és 100 kontroll vizsgálatánál *Ferrara és mtsai.* azt találták, hogy *G11107A* allélikus intron-polimorfizmus esélye szignifikánsan magasabb a betegpopulációban. Az összefüggés még erősebbnek mutatkozott az SSc populáción belüli autoimmun tiroditisz-alcsoportban (97). Autoimmun bőrbetegség, a vitiligoval való összefüggést 2008-ban fedezték fel *Tazi-Ahnini és mtsai* (98). 86 beteg és 363 kontroll vizsgálatok az *AIRE 7215C* allélja szignifikánsan magasabb eséllyel volt jelen a vitiligo betegpopulációjában (odds ratio, OR=3,12; 95%CI 1,87-5,46;  $p=1,36 \times 10^{-5}$ ) (98). *AIRE* allélpolimorfizmus miatti nem megfelelő hatékonyságú centrális tolerancia, hatással lehet a szervezet azon képességére, hogy a tumorsejteket és azok antigénjeit felismerje és elpusztítsa. *Conteduca és mtsai.* arra voltak kíváncsiak, hogy kapcsolatban állhatnak-e bizonyos allélpolimorfizmusok melanomával (99). 2010-es vizsgálatukban inverz asszociáció részeként a rs1055311 (C8385T), rs1800520 (C8723G) és a rs1800522 (T16366C) allélja szignifikánsan gyakoribbak voltak az egészségesekben, mint a melanomás betegeknél, függetlenül a nemtől, az életkortól és a melanoma stádiumától, sugallva az allél esetleges protektív hatását (99). Autoimmun alapú súlyos izomgyengeséget, myasthenia gravist 2017-ben hozták összefüggésbe *Zhang és mtsai.* egy kínai eset-kontroll kapcsán, melynél az rs3761389 G-allélja illetve az erre az allélra homozigóta genotípus szignifikánsan magasabb eséllyel volt jelen a betegeknél a kontrollhoz képest (OR=1,68; 95%CI 1,14–2,48;  $p=0,027$ ) (100). A mellékvesekéreg autoimmun pusztulásával, az Addison-kórral 2018-ban sikerült inverz asszociációt találni (101). Az rs2075875 C-, rs2075876 A-, rs62220374 G-, rs9983695 C-allélja szignifikánsan magasabb eséllyel volt jelen a kontrollpopulációban, ezek az allélok protektívnek bizonyultak a betegséggel kapcsolatban (101). Szisztémás lupusz eritematózus (SLE) esetében az immunrendszer tévesen megtámadja több szerv szöveteit az egész testben, különösen a bőrt, a veséket, az agyat és az ízületeket. Ezt a ritka autoimmun betegséget 2021-ben hozták összefüggésbe a timociták negatív szelekcióját irányító, *AIRE* fehérje expresszióját meghatározó intron 5-ös régióban található rs2075876 A alléljával, mely protektívnek bizonyult az esetcsoportban (OR=0,16, 95%CI=0,09–0,28) (102). Végül az *AIRE* allélpolimorfizmusok terén a legutóbbi molekuláris epidemiológiai eredményt 2023-ban publikálták és találtak összefüggést immun trombocitopéniás purpura betegséggel, mely autoimmun folyamatok által okozott, autoreaktív trombocita pusztulás következtében kialakuló vérzékenység (103). Ebben a rendellenességben szintén az rs2075876 lókus intron 5-ös régiójában található A allélja és az arra homozigóta genotípus jelenti a kockázatot (OR=4,299; 95%CI 1,650–11,202;  $p=0,003$ ) (103). Ahogy látjuk, ezen autoimmun betegségek esetében tömegével nem készítettek asszociációs vizsgálatokat. Szakirodalom szerint viszont egy autoimmun betegség összefüggését kiterjedten vizsgálták az *AIRE* kódoló régió allélpolimorfizmusaiival, ez pedig a reumatoid artritisz (RA). A következő fejezetben áttekintjük az ezzel kapcsolatos szakirodalmat.

### **1.2.3.1 *AIRE* allélpolimorfizmusok és a reumatoid artritisz**

Az RA egy gyakori autoimmun betegség, amely krónikus ízületi gyulladással jár. Az ebből eredő szimmetrikus, poliartikuláris artritisz extraartikuláris szövődeményekkel kombinálva funkcionális károsodáshoz vezet. Az *AIRE* promóter régiójában lévő polimorfizmusokat 2011-ben genomszintű asszociációs vizsgálat keretében először *Terao és mtsai* vizsgálták (104). Ennek során 5415 RA és 6489 ázsiai kontroll körében 277 420 SNP vizsgálatokor azt találták, hogy az intronrégióban lévő rs2075876 A- és rs760426 G-alléja és ezen allélok alkotta homozigóta genotípusok szignifikánsan magasabb eséllyel van jelen a betegpopulációban (OR=1,18; 95%CI 1,11–1,24;p=3,6x10<sup>-9</sup>; OR=1,16; 95%CI 1,10–1,22; p=4,4 × 10<sup>-8</sup>) (104). További eredményeikben publikálták, hogy az rs2075876 splicing régióban foglal helyet, az AA genotípus jelenlétében az expresszió szignifikánsan alacsonyabb (p=6,8×10<sup>-5</sup>) szemben a GG genotípussal, mely eredmény felveti annak a lehetőségét, hogy a betegségre való kitétség már újszülöttkorban elkezdődik, a nem megfelelően működő negatív szelekció által (104). Hasonló eredményt értek el spanyol, japán, kínai, egyiptomi vizsgálatok során melyek döntő többsége eset-kontroll vizsgálat. Összességében elmondható, hogy maguk az allélpolimorfizmusok alkotta genotípusok nagymértékben hozzájárulhatnak az *AIRE* expresszió alakulásához, mint azt *Lovewell és mtsai.*-nak vizsgálatában láttuk, képesek alacsony expressziót eredményezni, mely az immunológiai tolerancia esetében az újszülöttkori negatív szelekciót alapjaiban határozhatja meg (95). Alacsony *AIRE* expresszió esetén, magasabb valószínűséggel kerülhetnek ki autoreaktív CD4<sup>+</sup> effektor sejtek a perifériára, melyek autoantigén jelenlétében elindíthatják a betegség iniciációs szakaszát már újszülöttkorban (106).

### I.3. Reumatoid artritisz

Az Európából és Észak-Afrikából származó ősi csontvázmaradványok vizsgálata egyértelműen kimutatta az ízületi gyulladás különböző típusait, beleértve az osteoartritist, köszvényt és a Bechterew-kórt. Érdekes módon a kisízületek csont-porc határfelületén a marginális erózióval járó, határozott reumás elváltozások feltűnően hiányoztak. Ezzel szemben az évezredekre visszanyúló ősi mintákon végzett vizsgálatok kimutatták a reumás ízületi gyulladás egyértelmű jeleit az észak-amerikai indián törzseknél. Jelenleg az RA prevalenciája továbbra is feltűnően magas ezekben a régiókban, egyes csoportokban az egyének több mint 5%-át érinti. RA-t Európában először a 15-17. században figyelték meg a művészi alkotásokon, különösen a holland mesterek festményein. Ennek egy legszembetűnőbb példája Jacob Jordaens flamand mester 1630 körüli alkotása, a *La Familia de Jordaens en un Jardín* címmel. Középen elhelyezkedő feleségén - Catharina van Noort - megfigyelhetőek a kézközépcsonti és ízületi duzzanatok (107). Az első esetleírást Sydenham tette közzé 1676-ban. Bár ezt követően időszakos esetsorozatokot dokumentáltak, a betegséget teljes mértékben elismerték, amikor Garrod 1859-ben definiálta azt. Ő nevezte el „reumatoid” artritisznek, hogy megkülönböztesse az ízületi gyulladás akkoriban jól ismert formáitól, a reumás láztól és a köszvénytől. A 20. század elejére az RA-t az osteoartritistől („arthritis deformans”) megkülönböztethetőnek ismerték el. 1957-ben Charles Short átfogó és pontos leírást adott az RA-ról, amelyben a szeronegatív spondyloarthropathiáktól, a kristályok okozta betegségtől, az osteoartritistől, a szisztémás lupus erythematosustól és sok más betegségtől különálló klinikai állapotként határozta meg (107).

A reumatoid artritisz (RA) szimmetrikus poliartikuláris ízületi gyulladás, amely elsősorban a kéz és a láb kis diarthrodialis ízületeit érinti. A tüneti stádiumban infiltrálódó autoreaktív CD4<sup>+</sup> effektor T-sejtek (Th1, Th2, Th17), plazmasejtek és autoantigén ellen termelt autoantitestek - többek között reumatoid faktor (Rf), anti-ciklikus citrullinált peptid antitest (anti-cyclic citrullinated protein antibody, aCCP) - jellemzik (108). Beszűrődve az ízületi

tokba a CD4<sup>+</sup> sejtpopuláció citokineket, IFN- $\gamma$  és IL-17-et termel és a makrofágokkal karöltve, együtt a felgyülemlett autoantitestek alkotta immunkomplexekkel szinovitiszt és a szinoviális sejtek növekedését valamint heges kötőszövet (pannus) kialakulását eredményezi (108).

### **I.3.1. Kialakulásának stádiumai**

Az RA kialakulása egy több szakaszon keresztül történő progresszióként értelmezhető, amelyek mindegyikét különböző klinikai és patológiai jellemzők jellemzik. A fejezet során egészen az újszülöttkortól indulva próbáljuk meg felfejteni a kiváltó okokat egészen a késői tünetek megjelenéséig. Ennek megfelelően az RA kialakulását immunbiológiai szempontból 3 szakaszra osztjuk: iniciális, centrális és késői destruktív szakasz.

#### **I.3.1.1. Iniciális szakasz**

Az RA ezen fázisát nevezik kezdetinek, gyakran „preklinikai” vagy „korai” fázisnak (109). Ezt a szakaszt genetikai és környezeti kockázati tényezők jellemzik illetve következményesen olyan immunológiai folyamatok elindulása, melyek egy élethosszon át tartó kaszkádfolyamatot tartanak fenn. A szakasz megágyaz a későbbi fázisoknak, kitétséget teremt, melynek talaján a későbbi fázisokban elindulnak olyan immun és autoimmun folyamatok melyek már a tüneti fázist jelentik.

##### **I.3.1.1.1. Környezeti kockázati tényezők**

Az RA környezeti kockázati tényezőit az életmód és a fizikális (élettelen) és a biológiai (élő) környezet adja (109). Dohányzás a legerősebb életmódi kockázati tényezője a betegségnek. Ezen életmódi faktor hatása kifejezetten az aCCP pozitív betegek alpopulációjára korlátozódik, vagyis megelőző vizsgálatokban ezen alcsoport betegei szignifikánsan magasabb eséllyel dohányoztak, mint az illesztett kontrollok. A betegség pathomechanizmusát tekintve a kockázat mértéke olyan alacsony, hogy klinikai tünetek kialakulásának szempontjából jelenleg nem releváns, vagyis kijelenteni nem lehet, hogy szükséges és elégséges oka a betegségnek (109). Egyéb lehetséges környezeti kockázati tényezők közé tartozik az alkoholfogyasztás, kávéfogyasztás, orális fogamzásgátló használata bár ezen életmódi tényezők evidencia-háttere jelenleg nem elégséges és ellentmondásosak (109). Például egyes *in vivo* vizsgálatok szerint az alkohol csökkenti az RA NF- $\kappa$ B gyulladásozó út vonalainak aktivitását, míg mások az ellenkezőjét állítják, vagy nem találtak összefüggést (109). Egyes közlemények az omega-3 zsírsav lehetséges protektivitásáról számoltak be (109). Más vizsgálatok arra jutottak, hogy a magas nátriumtartalmú élelmiszerek fogyasztása fokozza az RA kockázatát, különösen akkor, ha az étrend, dohányzással párosul (110).

Az RA kockázatát befolyásoló fizikai, élettelen környezeti kockázati tényezők a következők lehetnek a jelenlegi állás szerint. Kristályos vagy amorf szilícium-dioxidnak (SiO<sub>2</sub>) való kitétség olyan iparágakban, mint a bányászat, az építőipar és a feldolgozóipar dohányzással párosulva emelheti az RA kockázatát, kifejezetten aCCP<sup>+</sup> betegnél találtak összefüggést (111). Itt 11 285 eset és a 15 249 kontrollnál azt állapították meg, hogy aCCP<sup>+</sup> RA betegek körében szignifikánsan magasabb eséllyel fordult elő SiO<sub>2</sub> por (OR=1,4, 95%CI 1,2-1,6) és azbeszt rostnak (OR=1,2; 95%CI 1,0-1,4) való kitétség (111). Szénbányászok pneumoconiosisakor megfigyelési vizsgálatok magasabb RA előfordulásról számoltak be

(112). Közlekedésből származó légszennyezés esetében sem találtak tiszta összefüggést, ugyanis az RA prevalenciája és éves incidenciája is változatlan maradt olyan városokban, ahol az átlaghoz képest tisztább a levegő (Stockholm) (113). Tajvani vizsgálatban a légszennyezés NO<sub>2</sub> frakciója mutatott szignifikáns összefüggést RA kockázattal (114). Mindezekon kívül 2006-ban egy dán eset-kontroll vizsgálat számolt az alacsony szocio-ökonómiai státuszról, mint kockázati tényezőről (115).

Az RA további endogén biológiai, kockázati tényezője a menopauzális vagy más ok miatti drasztikus ösztrogénszint csökkenés, mely alátámasztja azt a tényt, hogy a betegség kialakulása nőknél gyakoribb (116). Utóbbi évek vizsgálatai alapján elmondható, hogy az ösztrogén protektív. Elősegíti az anti-inflammatorikus citokinek (pl. IL-10) és csökkenti a proinflammatorikus citokinek (pl. TNF- $\alpha$ , IL-1 és -6) termelődését. T-sejtes regulációja során Treg sejteket aktivál, melyek így az autoimmun Th sejteket klonálisan elnyomják (116). Ösztrogén fenntartja a csontmátrixot, késlelteti az eróziót (116). Receptorok útján közvetve részt vesz gyulladáscsökkentő folyamatok promotálásában (116). Mindezek alapján az alacsony ösztrogénszint kockázati tényező a betegségaktivitását tekintve (116). A biológiai, élő környezet exogén kockázati tényezőinek kiterjedtebb szakirodalma van, szorosabb összefüggést mutatnak, mint az fizikális és életmódi tényezők. Ezen biológiai kockázati tényezők a különböző bakteriális és virális fertőzések (117). Természetesen ezek esetében is nehéz egy adott patogént úgy értékelni, hogy szükséges és elégséges oka az RA-nak, de esetükben az immunológiai összefüggés jelentősen erősebb és magyarázatot ad számos folyamat elindulására. Esetükben fontos leszögezni, hogy bizonyos esetekben előfordulhat, hogy egy kiváltó kórokozó helyett több egyidejűleg jelenlévő bakteriális közösségek azok, amelyek hozzájárulnak a betegséghez. Ezeken a fertőzéseken való átesés gyakorisága is számottevő tényező, kiváltképp a preklinikai szakaszban (117). Továbbá, előfordulhat, hogy a fertőző ágensek nem minden esetben váltják ki az RA-t, hanem bizonyos esetekben életmódi vagy élettelen környezet kockázati kombinációjának eredményeként járulnak hozzá a korai stádiumok beindításához. Végül pedig a genetikai kockázati tényezők nagymértékben befolyásolhatják a környezeti fertőzések hatását (117). Szakirodalom alapján a következő kórokozó és gazda közti kapcsolatok válhatnak ki autoimmun pathomechanizmust: molekuláris mimikri, epitóp kiterjedés, poliklonális limfocita aktiváció és bystander mechanizmus. Molekuláris mimikri akkor fordul elő, amikor az idegen antigének kellő szerkezeti hasonlóságot mutatnak a saját antigénnel, melyet a közös (shared) epitópnak (SE). Ennek eredményeként a kórokozókra adott immunválasz keresztreaktivitáshoz vezethet a saját antigénnel. Eppstein-Barr vírus (EBV) ellen termelt antitestek keresztreakálhatnak saját, TRA-antigénnel az ízületi tokban. Erre a folyamatra jó példa a *Proteus mirabilis* baktérium. A mikroba bakteriális ureáz enzimének aminosavszekvenciája, a 337. és 341. aminosav között tartalmaz egy ún. IRRET motívumot, mely szekvenciálisan és alakilag identikus a humán alfa-2(XI) kollagén 421-425 aminosavszekvenciájának LRREI motívumával (118). A XI. típusú kollagén a kis ízületekben található hialinporc alkotóeleme melynél, ha az érintett húgyúti *Proteus* fertőzésen esik át az immunrendszer által termelt antiureáz antitest megtámadhatja a hialinporc kollagénjét (118). Epitóp kiterjedés ("epitóp spreading") akkor alakul ki, ha amikor az immun T- vagy B-sejtes válasz az eredeti epitópon túlra is kiterjed. Ez különösen azért hasznos, mert így hozzájárul azon képességhez, hogy az immunrendszer egyszerre több kórokozót is megtámadjon és elpusztítson. Epitóp kiterjedés akkor is kialakulhat, ha a fehérje aminosav összetétele megváltozik argininről citrullinra, mely ellen a rendszer autoantitesteket termel. Ez megvalósulhat egy baktérium, a *Porphyromonas gingivalis* peptidil-arginin-deamináz (PAD) enzimeje által. Ezt a citrullinált fehérjét az APC sejtek felveszik és a T- majd B-sejtes útvonalon autoantitest termelődik ellene (117).

### I.3.1.1.2. Genetikai kockázati tényezők

Az RA lehetséges genetikai hátterét először 1953-ban feltételezték, rokonsági kapcsolatok és ikervizsgálatok eredményei alapján. Későbbiekben meghatározásra került, hogy RA betegek elsőfajú rokonai négyszeres kockázatnak vannak kitéve az átlagpopulációhoz képest (119). Bár a genomszintű asszociációs vizsgálatok (genome wide association studies, GWAS) több mint 100 lókuszt detektáltak RA kockázatára, ezek közül még mindig az egyetlen, amely erős összefüggést mutat az RA kialakulásával az a humán leukocita antigén (HLA) régió (118). A HLA-molekulákat a 6-os kromoszóma rövid karján, a 6p21.1-21.3 pozícióban található I-es és II-es osztályú géncsoportok kódolják az MHC I és II-es receptormolekulákat. A HLA-DR expresszió az APC sejtek (DC sejtek, makrofágok és B-sejtek) felületén történik, nélkülözhetetlenek a CD4<sup>+</sup> T-sejtek antigén prezentációjához. Az előbb ismertetett "shared epitope" jelensége során, a patogén vagy az endogén, exogén módon módosított saját fehérje peptidrészlete hasonló szekvenciát mutat saját peptidrészlettel. Ez a peptidrészlet azért képes kötődni a HLA-DR gének által meghatározott receptorok peptidkötőzseibe, mert a gének ritka allélpolimorfizmust hordoznak és ritka esetben, de képesek olyan peptidkötőzsebeket létrehozni polimorfizmusuk okán, hogy abba a saját, önmagában nem káros, természetesen jelen levő fehérje is képes belekötődni (118). Ilyen fehérje többek között a citrullinált fehérje. A HLA gének hozzájárulását először *in vitro* vizsgálatokkal 1975-ben igazolták, a SE jelenséget 1986-ban *J. Silver és S. M. Goyer* említették először. Későbbiekben meghatározták, hogy kifejezetten a HLA-DR4-es haplotípus megléte jelenti a kockázatot (118). Később *Huizinga és mtsai.* meghatározták, hogy ez kifejezetten csak az aCCP<sup>+</sup> betegekre vonatkozik, mely az esetek 60%-a (119). A fennmaradó 40%-ban a HLA-DR4-es haplotípus nem kimutatható. Ennek megfelelően számos GWAS vizsgálat indult további genetikai kockázati tényezők felfedezésére. Ezek közül érdemes kiemelni a peptidylarginine deiminase type 14 (PADI14), mely a citrullinálciót végző enzim, illetve a protein tyrosine phosphatase, non-receptor type 22 (PTPN22), mely a T-sejt érésért valamint a citokintermelésért felelős fehérje. Azóta számos nagy áteresztőképességű GWAS vizsgálat mutatta ki a nem-HLA gének jelentőségét az RA patogenezisében (120,121). A mai napig három különálló, nemrégiben végzett GWAS 124 kockázati lókuszt hozott összefüggésbe az RA-val, 69 825 beteg és 330 798 illesztett kontroll bevonásával (122-124). Míg ez a 124 lókuszt többsége általános sejtes és gyulladásos folyamatokért felelős, addig a gének közül kizárólag egy volt felelős olyan embrionális és posztnatális immunológiai folyamatért, mely a saját antigénnel szembeni centrális toleranciát biztosítja. Az ezt kódoló szekvencia az *AIRE* (122-124).

### I.3.1.1.3. Immunológiai folyamatok

Az RA patomechanizmusa során az egyes szakaszokat nehéz pontosan demarkálni, átmenetet képeznek. Az iníciális szakasz a preklínikai stádium első része, az események legfőképpen a másodlagos limfoid szervekben zajlanak, immunológiai folyamatai genetikai kockázati tényezők által vezérelt kitettség (125). Ide soroljuk a lehetséges *AIRE* allélpolimorfizmusok miatti, a centrális tolerancia posztnatális negatív szelekciójának nem megfelelő működéséből fakadó autoreaktív CD4<sup>+</sup> T-sejtek nyirokcsomóba szüremelését (7-15). Az iníciális szakasz része a szintén allélpolimorfizmusok miatt kialakuló speciális HLA-DR szubtypusok, melyek speciális peptidkötőzsebeket kódolnak az APC sejtek (makrofágok és DC sejtek) MHC receptorain. Ezek a kötőhelyek várnak az autoantigén érkezésére és bekötődésére, mely autoantigén kialakulhat exogén – pl. fertőzés okozta molekuláris mimimkri vagy bakteriális citrullináció – vagy endogén ok miatt, mikor pl. a szervezet citrullinálja a humán enoláz enzimet (117,118). Tehát a genetikai kitettség során kialakul egy immunológiai státusz, mely

később képes lesz fogadni és reagálni az autoantigénnel (126). Az iníciaális szakasz során kialakuló autoimmunitás mellett az exogén, környezeti okok miatt folyamatosan termelődnek az autoantigének, pl. dohányzás emeli a PAD enzim expresszióját, mely a fehérjéket poszttranszlációban citrullinálja (127). Az autoimmunitás kialakulása mellett folyamatosan termelődnek az autoantigének, melyek ellen autoantitestek termelődnek. Iníciaális szakasz korai stádiumában HLA-DRB1 jelenlétében pl. már termelődnek Rf és aCCP autoantitestek, melyek szérumból kimutathatók (125). A képet tovább bonyolítja, hogy nem feltétlenül szükséges HLA-DR szubtípus által meghatározott MHC receptor az RA kialakulásához, később ezek a betegek lesznek a szeronegatív RA esetek. Összességében az iníciáció során, genetikai talajon kialakuló autoimmunitás az idő előrehaladtával egyre kevésbé tolerálja a saját struktúrákat, autoantigéneket melynek eredménye az autoantitestek képződése és végül a gyulladás (126).

### **I.3.1.2. Centrális szakasz**

A poszttranszlációs módosulásokon (citrullináció, acetiláció, karbamiláció) átesett fehérjét (pl. II-es típusú kollagén, vimentin) ebben a stádiumban olyan APC sejtek veszik fel, melyeknek az MHC receptorába az antigén képes bekötődni (128). A makrofág vagy DC sejtek az autoantigént a legközelebbi nyirokcsomóba viszik, ahol számos CD4<sup>+</sup> T-sejtpopuláció várja (Th1, Th2, Th17), melyek a B-sejt prezentáción keresztül autoantitestet termelnek a plazmasejttel az autoantigén ellen. A Th1-sejtcsoport proinflammatorikus citokineket (IFN- $\gamma$ , IL-2 és TNF- $\alpha$ ) termel (128). Th2-es sejtek gyulladásgátló citokineket, IL-4 és IL-5-öt választanak el, melyeknek fontos szerepük van a B-sejt aktivációban. Mindemellett, a Th17 sejtek IL-17-et termelnek, melyek proinflammatorikus citokintermelést, mátrix metalloproteináz (MMP) enzim elválasztást idéznek elő (128). Ezzel párhuzamosan Treg sejtek IL-10 termeléssel segítenek az autoreaktív T-sejtek klonális elnyomásában. Ebben a stádiumban alakul ki egy gyulladással körülvett környezet, melynek hatására többek között a szinóvium fibroblasztszerű szinoviocitái (fibroblast like synoviocytes, FLS) kemokineket kezdenek el termelni (CCL2, CCL5, CXCL10). Ezen a helyen endoteliális sejtek 1-es típusú intercelluláris adhéziós és vaszkuláris molekulákat (Intercellular Adhesion Molecule 1, ICAM-1, Vascular Cell Adhesion Molecule 1, VCAM-1) termelnek, melyek a kemokinekkal karöltve segítik a cirkuláló aktivált T-sejtek extravazációját a szinóviumba (129). A betegség ezen pontján az aktivált T-sejtek IFN- $\gamma$  és IL-17 termeléssel, B-sejtek, plazmasejtek autoantitest és citokin termeléssel gyulladást idéznek elő. Az esemény további makrofágokat vonz a szinóviumba, melyek TNF- $\alpha$ -t, IL-1 és IL-6 citokineket termelnek (129). A keringésben már meglévő autoantitestek is felgyűlnek a szinóviumban, ahol immunkomplexeket képezve tovább súlyosbítják a gyulladást. Mindezek hatására FLS sejtek elkezdenek növekedni és kialakul a heges kötőszövetet jelentő pannusz (129). A centrális szakasszal a preklinikai szakasz véget ért, az események már a tüneteket adó klinikai stádiumban folytatódnak (129). Ekkorra már mindkét kéz második és harmadik kézközépcsont-ízületének és több proximális interphalangeális ízület, alig észlelhető duzzanata alakul ki (129).

### **I.3.1.3. Destrukatív szakasz**

Az FLS sejtek agresszív növekedése az ízületi tok struktúráját rombolja, mely később csontterozióban folytatódik, amit az oszteoklaszt sejtek aktivációja és érése idéz elő. Proteázokat, pl. cathepsin K-t termelnek, melyek együtt a neutrofil granulociták által termelt reaktív oxigén szabadgyökökkel egyrészt fenntartják, a gyulladással körülvett másrészt



lebontják a csontmátrixot, az ízületi porc folyamatosan degenerálódik. Ezen a ponton az akut szinovitisz krónikussá kezd alakulni. Az itt termelt aCCP kötődik az oszteoklaszt sejtek prekursoraihoz, csontreszorpciót okozva, melyek radiológiai felvételeken foltok formájában mutatkoznak. A betegség fenotípusa ekkorra már deformításokat foglal magába, beleértve a kézközépcsont-ízületek részleges vagy hiányos elmozdulását (szubluxációját) (127,128). Több ujjon jellegzetes "hattyúnyak"-deformitás jön létre, mely tünet leginkább a kisujjon látható valamint kialakul a hüvelykujj Z-deformitása. Idővel a különböző kis kéztőcsontok közötti ízületi tér szűkül, és szinte egyetlen csontot alkotva összeolvadnak (128). Az RA késői destruktív fázisát jellemzi a boka- és lábízület érintettsége és deformitása. Ebben a stádiumban a betegség már régóta fennáll és kezeletlen. Ekkorra már a test többi része is érintett lehet, mint pl. a szem sclerájának degenerációjával járó scleromalacia perforans oculi sinistri vagy bilaterális reumatoid csomók és fibrózis kialakulása a tüdőben (128).

### I.3.2. Általános epidemiológiája

Az RA prevalenciája közepes és magas jövedelmű országokban 0,5-1%, éves incidenciája 5-50 eset/100 000 fő, nemi eloszlása 2:1-hez a nők javára (130). 1990 és 2017 között a globális prevalencia becslések szerint 7,4%-kal, az incidencia 8,2%-kal emelkedett (130). Egyesült Államok kohorsz vizsgálata során a szeropozitív esetek csökkenését és a szeronegatívak emelkedését figyelték meg 1985 és 2014 között ugyanazon 1987-es diagnosztikus kritériumokat használva (130). Nyugati országokban a dohányzás és a SiO<sub>2</sub> expozíció – mint exogén kockázati tényező – folyamatosan és enyhén csökken, míg a világ más országaiban az aritrogén kockázati tényezők okozta expozíció (dohányzás, légszennyezés) emelkedik (130). Összességében globális trendekről elmondható, hogy az incidencia és prevalencia illetve a betegség kezdetének átlagéletkora emelkedik. A mortalitás (beleértve a kardiovaszkuláris komorbiditást), sebészeti beavatkozások száma, a diagnózis és az első terápia közötti idő folyamatosan csökken, mely az egyre hatékonyabb felismerésnek és terápiának köszönhető (130). A halál második gyakori oka a diffúz interstitiális tüdőfibrózis, mely szintén csökkenő tendenciát mutat. Renális amiloidózis és RA vasculitis incidenciája is folyamatos csökkenést mutat. A betegség aktivitása az első prezentációkor szintén csökkenő tendenciát mutat a legutóbbi szisztémás irodalmi áttekintés alapján, mely azt mutatja, hogy a betegséget egyre korábban diagnosztizálják (132).

Európai epidemiológiája alapján az észak-és nyugat-európai országokban prevalenciája 0,4% és éves incidenciája 20-30 eset/100 000 fő, míg a déli és keleti országok mérőszámai ennél alacsonyabbak, 0,2%-os prevalenciával és 7-15 eset/100 000 fő éves incidenciával (133). A nemi eloszlás tekintetében 2:1-hez a nők javára, de egyes országokban, mint pl. Franciaországban ez az érték 5,7:1 szintén a nők javára (133).

Hazánkban az országos prevalencia mediánja 2018-ban 0,33%, mely a franciaországi és olaszországi értékekhez hasonlít (134). *Kiss és mtsai.*-nak 10 000 fős összehasonlító vizsgálata alapján elmondható, hogy Magyarországon a dél-dunántúli régió prevalenciájának mediánja 2005-ben az országosnál valamivel magasabb, 0,37% a 14 és 65 év közöttiek körében, nemi eloszlása 3,3:1-hez a nők javára (135).

A betegség primer prevencijáról nem igazán beszélhetünk tekintve, hogy a betegség kialakulásának legkorábbi eseménye egy genetikai kitettség során fiatal korban kialakuló immunológiai státusz, mely később képes lesz reagálni saját TRA antigénekkkel. Továbbá a vizsgált AIRE fehérjének az újszülöttkori negatív szelekcióban van központi szerepe, mely felveti az allélpolimorfizmusok szerepét annak tekintetében, mekkora hatékonysággal törölődnek az autoimmun timociták. *AIRE* allélpolimorfizmusok nagymértékben csökkenthetik az *AIRE* expresszióját, melynek következményeként előfordulhat, hogy nem termelődik

elegendő mennyiségű TRA az mTEC sejtekben, így az ellenőrző rendszeren átcúsúszhatnak olyan timociták melyek autoreaktívak (7-15). Ennek tükrében a betegség immunológiai státusza, talaja, az autoimmunitás genetikai alapon már újszülöttkorban kialakulhat. Ezt célzó újszülöttkori kezelés nem létezik. Az exogén, környezeti kockázati tényezők eliminálása önmagában nem jelent védelmet a betegség biológia kezdetének megakadályozására, mert a betegség multifaktoriális. Szekunder prevenciója során a cél a preklinikai státusz feltérképezése még a tünetek megjelenése előtt. A korai stádiumú betegség kezelésének célja a remisszió vagy az alacsony aktivitáson tartás (108). Szerológiai biomarkerek alkalmazása úgy, mint aCCP alkalmas a preklinikai fázis detektálására. RA-val diagnosztizált betegek 60%-a 6-10 évvel a diagnózis előtt aCCP<sup>+</sup> illetve prezentálják a HLA-DRB1-es genetikai szubtypust (108). A fennálló 40% viszont szero- és HLA-DRB1 negatív így esetükben ezek a markerek nem használhatóak. Kivizsgálásra indikációt jelent a fent ismertetett környezeti és genetikai kockázati tényezők valamelyikének megléte illetve a reggeli ízületi merevség (108). A betegség diagnózisa a 1987-es klasszifikáció (American College of Rheumatology, ACR) alapján történik (136). Gradálására a betegség aktivitási pontszámot 28 ízületben (disease activity score in 28 joints, DAS28) és a funkcionális függetlenség mértékén (functional independence measure, FIM) score-rendszert használják (137). Gyógyszeres kezelését összefoglalóan betegségmódosító antireumatikus terápiának hívják (disease modifying antirheumatic drugs, DMARD). Ennek konvencionális részét képezi a metotrexát (MTX) kezelés, mely többek között gátolja a proinflammatorikus citokinek felszabadulását, MMP hatását, gyulladáscsökkentő (138). Konvencionális terápia részét képezi továbbá a leflunomide, sulfasalazine és chloroquine terápia, melyek pirimidin szintézis inhibitorok, gyulladáscsökkentők és immunmodulátorok (128). A konvencionális terápián kívül a biológiai DMARD kezelés az egyes immunológiai útvonalak során egyes citokinek termelődését befolyásolja. Infliximab, etanercept, golimumab terápiák TNF-alfa gátlók (128). Léteznek B-sejtválasz gátlására terápiák úgy, mint rituximab, T-sejtes gátlást biztosító abatacept, interleukineket gátló tocilizumab és növekedési faktor gátló mavrilimumab (128). Jövőbeni terápiás kutatások az RA szignáltranszdukciós útvonalait, Toll-like receptor útvonalait illetve az össejtterápiát célozzák meg (128).

## **II. Célkitűzések**

### **II.1. Az AIRE allélpolimorfizmusok meta-analízise autoimmun kórképekben**

Az első AIRE allélpolimorfizmussal kapcsolatos molekuláris epidemiológiai megfigyelést 2002-ben végezték, azóta számos eset-kontroll analízis született különböző autoimmun betegségeket vizsgálva. Mielőtt elkezdtük kutatásunkat, szisztematikus szakirodalmi áttekintéssel vizsgálódtunk, mely meta-analízisnek a célja az volt, hogy tisztázzuk, mely autoimmun betegség körében született a legtöbb molekuláris epidemiológiai eset-kontroll vizsgálat, az adott betegségen belül pedig, az egyes eset-kontroll vizsgálatok összefoglaló elemzésével új eredményeket és összefüggéseket találjunk kandidáns allélpolimorfizmusok és autoimmun betegségek között.

### **II.2. In silico vizsgálat AIRE allélpolimorfizmusok meghatározására**

Annak érdekében, hogy rendelkezésünkre álljon egy adatbázis, mely lókuszek allélpolimorfizmusaival kívánunk a későbbiekben foglalkozni, célul tűztük ki egy olyan

átfogó adatbázis létrehozását, melyben 5'-3' irányban szerepel az összes olyan allélpolimorfizmus, melynek allélfrekvenciája legalább 1% vagy afölötti.

### **II.3. Pilot vizsgálat *AIRE* rs878081 és rs1003854 allélpolimorfizmusok vizsgálatára reumatoid artritisz betegek mintáiban**

Mivel számunkra kezdetben ez egy ismeretlen terület volt, meg kellett figyelniük, vajon a kiválasztott *AIRE* domén-kódoló régió lókusza adta allélfrekvenciák és vizsgálati elrendezésünk képesek-e értékelhető eredményt adni. Annak tisztázására, hogy az allélpolimorfizmusok diszkriminációjára választott allélspecifikus próbák és a genotipizálás során kialakított protokoll megfelelőek-e, célul tűztünk ki egy olyan pilot vizsgálat megvalósítását, ahol ezekre a kérdésekre választ kaphatunk. Ennek érdekében kiválasztottuk a DNS-kötő domént kódoló SAND régió két lókuszt, rs878081-et, rs1003854-et és célul tűztük ki ezen lókuszon adta allélpolimorfizmusok és genotípusok összefüggésének vizsgálatát reumatoid artritisz kockázatával, eset-kontroll vizsgálat keretében, melyhez 6 különböző genetikai modellt használtunk. Ehhez szintén célul tűztük ki egy olyan átfogó adatbázis létrehozását, mely magába foglalja a pilot study és az esetleges későbbi vizsgálat RA és kontrollpopulációinak teljes leletanyagát, rögzítve minden olyan paramétert, mely a leletanyagból kinyerhető, beleértve demográfiai adatokat, életmódi adatokat, laborvizsgálatok eredményeit és a múltban rögzített fertőzéseket is. További célul tűztük ki, amennyiben lehetséges az egyes genotípus-alcsoportok klinikai paramétereinek átlagértékeit statisztikailag összehasonlítsuk és amennyiben az szignifikánsnak bizonyul, korrelációs vagy asszociációs összefüggéseket vizsgálunk a klinikai paraméterek és a genotípusok között.

### **II.4. Az *AIRE* exon 5 és 6 valamint intron 5 és 7-es régióinak vizsgálata reumatoid artritisz kockázatára**

Annak függvényében, hogy a pilot vizsgálatunk milyen eredménnyel jár, a vizsgálat folytatásaként célul tűztük ki nagyobb elemszámú eset és kontroll populáció és több allélpolimorfizmus vizsgálatát annak tisztázására, vajon a DNS-kötő régió allélpolimorfizmusai kapcsolatban állnak-e a betegséggel. Ennek érdekében az *in silico* adatbázist és a pilot study eredményeit felhasználva, egy olyan panelt terveztünk, mely magába foglalja mindazon potenciális lókusztokat és azok adta allélpolimorfizmusokat, melyek SAND domént kódoló régiót (exon 5 és 6), valamint a köztük lévő splicing intronális régiót (intron 5 és 7) határozzák meg. Célunk a panelben szereplő allélpolimorfizmusok illetve az allélok adta genotípusok vizsgálata reumatoid artritisz kockázatára, 6 különböző genetikai modell felhasználásával.

### **II.5. Az *AIRE* exon 5 és 6 valamint intron 5 és 7-es régiók összefüggéseinek vizsgálata a betegséget meghatározó klinikai paraméterekkel**

A továbbiakban kíváncsiak voltunk, hogy az egyes allélpolimorfizmusok adta genotípusok összefüggésben állnak-e az RA-t meghatározó klinikai paraméterekkel és a betegség aktivitásával. Célul tűztük ki a panelben szereplő lókuszon alléljai által meghatározott genotípusok vizsgálatát, 5 különböző genetikai modellt felhasználva a következő klinikai paraméterek vizsgálatára: vérsüllyedést meghatározó eritrocita szedimentációs ráta

(erythrocyte sedimentation rate, ESR), C-reaktív protein (CRP), Rf (reumatoid faktor), aCCP, DAS28 érték.

## **II.6. Allélpolimorfizmusok regulatórikus kötőmotívumokra gyakorolt hatásainak *in silico* elemzése**

A vizsgálat zárásaként arra voltunk kíváncsiak, az egyes lókuszek alléljai milyen hatással lehetnek a DNS azon kötőhelyeire, melyekhez regulatórikus egységek pl. transzkripció faktorok kötődhetnek, így célul tűztük ki a panelon szereplő azon allélpolimorfizmusok bevonását, melyekkel kapcsolatban statisztikailag szignifikáns eredményt kaptunk az alléldiszkriminációs vizsgálat során RA-val illetve ha ott nem, de adott allélok adta genotípus statisztikailag szignifikáns összefüggésben volt klinikai paraméterekkel. A kötőmotívumok és az azokhoz kapcsolódó regulatorikus egységre mért hatást célul tűztük ki a HaploReg v4.2 *in silico* adatbázisban elemezni.

## **III. Anyagok és módszerek**

### **III.1. AIRE allélpolimorfizmusok meta-analízise**

#### **III.1.1. Publikációk keresési stratégiája**

A szakirodalmat a PubMed, Embase, Cochrane Library és Web of Science adatbázisokban kerestük a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) ajánlásainak megfelelően (105,139). Két egymástól független kutató végezte a kulcsszavas keresést, melyben a kulcsszavak a következők voltak: “autoimmune regulator”, “AIRE”, “polymorphism”, “rheumatoid arthritis”. A beválogatott közlemények 2011. április és 2016. június között jelentek meg. A keresését 2017. május 16-án fejeztük be (105).

#### **III.1.2. Publikációk beválasztási és kizárási kritériumai**

A közleményt beválasztottuk amennyiben (1) összefüggést talált az AIRE gént kódoló DNS-szekvenciában található SNP-k vagy azoknak haplotípusai és a RA fogékonysága között; (2) eset-kontroll vizsgálat volt; (3) az összes beteg megfelelt az American College of Rheumatology diagnosztikus és besorolási kritériumainak (105,136); (4) rendelkezésre álltak az egyes résztvevők alléljainak valamint genotípusainak OR, 95%CI és p-érték adatai. A talált publikáció kizárássra került, amennyiben (1) az duplikátum volt, valamint (2) a közleményben vizsgált allélpolimorfizmus legalább 4 további publikációban nem volt megtalálható, (3) a publikáció szisztematikus szakirodalmi áttekintés volt. A beválasztási és kizárási kritériumok alapján történő keresést két kutató, egymástól függetlenül végezte (105).

#### **III.1.3. Statisztikai analízis**

Hardy-Weinberg egyensúlyt (Hardy-Weinberg equilibrium, HWE)  $\chi^2$  próba alapján számoltunk minden egyes kontroll csoporthoz (105). Annak érdekében, hogy az rs2075876, rs760426 lókuszek allélpolimorfizmusai és RA közti összefüggés erősségét meghatározzuk,

súlyozott OR-t (pooled OR) és 95%-os konfidencia intervallumot számoltunk (105). A közleményekben különböző etnikumokat vizsgáltak, ennek megfelelően DerSimonian és Laird random effekt modelljét használtuk (105,140). Heterogenitás vizsgálatára elsőként, a Cochran-féle Q-tesztet alkalmaztuk, ahol szignifikáns heterogenitásnak az számított, ha a Q értéke meghaladja a  $\chi^2$  próba k-1 szabadsági fokú felső kritikus értékét (105). Ebben az esetben szignifikánsnak tekintettük a tesztet, ha p-értéke kisebb volt, mint 0,1. Heterogenitás vizsgálatára másodikként, az inkonzisztencia ( $I^2$ ) tesztet alkalmaztuk, ahol a kapott  $I^2$  érték a vizsgálatok közötti variabilitás százalékos aránya a teljes variabilitáshoz képest. 25, 50 és 75%-os  $I^2$  értékek megfelelnek az alacsony, mérsékelt és magas heterogenitásnak a Cochran-féle kézikönyv alapján (105,141). Annak vizsgálatára, hogy az egyes tanulmányok eredményei hogyan befolyásolják az összesített súlyozott OR-t és annak 95% CI tartományát, szenzitivitás-analízis vizsgálatot végeztünk. Publikációs torzítást (publication bias) a funnel plot vizuális inspekciójával vizsgáltuk (105).

A standard hiba az esélyhányados logaritmusának függvényében került ábrázolásra a funnel plotban. Meta-analízisünk során alkalmazott számításokat és az ábrák (forest és funnel plot) generálását a Comprehensive Meta-Analysis Version 3 (Biostat, Inc., Englewood, NJ, USA) szoftvercsomaggal végeztük (105).

A továbbiakban, statisztikai analízisünk során annak érdekében, hogy elkerüljük az elsőfajú hibák okozta torzítást – mely a nem megfelelő mennyiségű adat esetén a random hiba megemelkedett kockázata – valamint növeljük a levont következtetés erősségét trial sequential analízist (TSA) végeztünk (105,142-144). TSA egy olyan becslés, mely meghatározza szükséges elemszámot a korrigált statisztikai döntési szint küszöbértékének tekintetében. A meta-analízis meggyőző erejét a vizsgálat során a kumulált z-görbe (cumulative z-curve) és a TSA monitoráló (ellenőrző) határvonalainak (the trial sequential monitoring boundary) viszonya mutatja meg (105). Amennyiben a TSA-határvonalakat a kumulatív z-görbe metszi, és a meta-analízis összesített elemszáma eléri a szükséges értéket, a következtetésünk kellően megalapozottnak és meggyőzőnek mondható. Abban az esetben, ha a kumulatív z-görbe metszi a határvonalakat, de az elemszám nem éri el a szükséges küszöbértéket, akkor is tekinthetjük úgy, hogy a következtetés levonásához elégséges bizonyíték áll rendelkezésre, további tanulmányok bevonására nincs szükség. Amennyiben a kumulatív z-görbe nem metszi a határvonalak egyikét sem és a meta-analízis elemszáma nem éri el a szükséges küszöbértéket, akkor a meta-analízisben nincs kellő mennyiségű információ erős következtetés levonására (105,145). A heterogenitáson alapuló megközelítést alkalmaztuk a szükséges elemszám kiszámításához, a relatív rizikó csökkenését 10 %-ra, az elsőfajú hiba kockázatát 5%-ra, a másodfajú hiba kockázatát pedig szintén 10%-ra választottuk. Az rs2075876 és rs760426 allépolimorfizmusok esetében felhasznált, illesztett konfidencia intervallumok 1,13-1,31 és 1,11-1,26 voltak. A számolásokhoz a Trial Sequential Analysis programot használtuk (Copenhagen Trial Unit, Center for Clinical Intervention Research, Denmark, version 0.9 beta, [www.ctu.dk/tsa](http://www.ctu.dk/tsa)) (105).

### **III.2. *In silico* vizsgálat AIRE allépolimorfizmusok meghatározására**

#### **III.2.1. AIRE promóter régiójában található variánsok beválogatási és kizárási kritériumai**

Az eset-kontroll vizsgálatunkhoz előzetesen szükség volt egy olyan adatbázis létrehozására, mely katalogizálja és rögzíti mindazon allépolimorfizmusokat, melyek az AIRE kódoló régiójában előfordulnak. Ehhez elengedhetetlen volt egy olyan többlépcsős, előzetes

szelekciós rendszer létrehozása, mely elkülöníti az egypontos nukleotid polimorfizmusokat (SNP-eket) a nagyméretű kópiaszám változásoktól (copy number variation, CNV) pontmutációktól, inzercióktól és delécióktól (indel). Ehhez két genom-szintű annotációs rendszert, UCSC Genome Browser (GRCh38 assembly) valamint az Ensembl automatikus humán annotációs rendszert használtuk (146,147).

A variáns akkor került beválogatásra, amennyiben (1) a ritka allél frekvenciája (minor allele frequency, MAF) 0,01 (1%) és 0,5 (50%) érték közé esett; (2) a variáns a következő adatbázisok legalább egyikében szerepel: NCBI dbSNP, 1000 Genomes Project Phase III (including the HapMap project), NHLBI Trans-Omics for Precision Medicine (human only), NHLBI GO Exome Sequencing Project (human only), Genome Aggregation Data-base (human only), Exome Aggregation Consortium (human only) (148-151). A variánsok közül előfordultak olyanok, melyeknek MAF értéke 0,01 és 0,5 között esett, de ha európai (EUR) alpopulációkat vizsgálatunk – mint pl. olyan utahi lakosok genomját melyeknek észak- és nyugat-európai felmenőik vannak (Utah residents with Northern and Western European ancestry, CEU) – a MAF érték 0,01-nél alacsonyabb volt, ami felvetette ezen variánsok detektálhatóságát. Így ezeket a variánsokat az alacsony detektálhatósági valószínűség miatt kizártuk (152).

### **III.3. Pilot vizsgálat *AIRE* rs878081 és rs1003854 allélpolimorfizmusok vizsgálatára reumatoid artritisz betegek mintáiban**

#### **III.3.1. Etikai engedély**

Vizsgálatunk beavatkozással járó kutatás, így ennek engedélyeztetésére etikai engedélyre irányuló kérelmet nyújtottunk be az Országos Tisztifőorvosi Feladatokért Felelős Helyettes Államtitkársága, Egészségügyi Igazgatási Főosztályára (továbbiakban: országos tisztifőorvos) 2018-ban, négy éves intervallumra, a Harkányi Termál Rehabilitációs Centrum Közhasznú Nonprofit Kft. bevonásával. Ehhez az országos tisztifőorvos felkérte az Egészségügyi Tudományos Tanács és Kutatásetikai Bizottságát (ETT TUKEB) szakhatósági állásfoglalás kialakítására. Ennek során az ETT TUKEB a kutatás engedélyezése iránti kérelmet szakmai-etikai szempontból megfelelőnek találta, a kutatás országos tisztifőorvos általi engedélyezéséhez szakhatósági hozzájárulását adta. Ennek alapján az országos tisztifőorvos 2018. április 24-én kérelmünket 4 évre hagyta jóvá, 11871-7/2018/EÜIG iktatószámmal (152).

Pilot vizsgálatunk mintagyűjtése COVID-19-es világvárvány miatt elhúzódott, így etikai engedélyünket szükséges volt időben meghosszabbítani további négy évre, valamint a mintagyűjtés helyszínét a későbbiekre tekintettel, nagyobb elemszámú vizsgálat érdekében a mintagyűjtésünk helyszínét a meglévő Harkányi Termál Rehabilitációs Centrum Közhasznú Nonprofit Kft. mellett a Hévízgyógyfürdő és Szent András Reumakórház bevonásával kiterjeszteni. Benyújtott módosítási kérelmünket a szakhatósági véleményezésre felkért ETT TUKEB szakmai és etikai szempontból megfelelőnek találta és a kutatási terv kérelemben foglalt módosításának engedélyezéséhez hozzájárult, melyek alapján a Nemzeti Népegészségügyi Központ a módosítási kérelmünket az 57142-5/2022/EÜIG iktatószámú határozatban 2022. november 8-án engedélyezte (152).

#### **III.3.2. Mintagyűjtés**

A vizsgálatban szereplő valamennyi személy a Harkányi Termál Rehabilitációs Központ ambuláns és bennfekvéses betegei közül kerültek ki. A vizsgálatra jelentkezés önkéntes alapon történt, melyhez betegtájékoztatót és beleegyező nyilatkozatot csatoltunk. A vérmintákat 4mL-es citrátcsövekbe vettük személyenként. A teljes vért  $-75^{\circ}\text{C}$ -on tároltuk az analízisig. Az önkéntesen vállalkozó személy RA csoportba került, amennyiben (1) 18 évnél idősebb volt, (2) a beleegyező nyilatkozatot aláírta, valamint (3) az American College of Rheumatology diagnosztikus és besorolási kritériumai alapján az RA diagnózisával rendelkezik (136). Azokat az RA betegeket, akiknek a kórtörténetében egyidejűleg más autoimmun kórképek, pl. Sjögren-szindróma, pszoriázis, SSc, SLE szerepeltek, kizártuk a további elemzésből, annak érdekében, hogy csökkentjük az RA-tól eltérő autoimmun betegségek RA kockázatára mért zavaró hatását (152).

Mivel már a vizsgálat elején tudtuk, hogy gondosan meg kell terveznünk a megfelelő kontrollpopulációt, azon gondolkodtunk, hogyan korlátozhatnánk az esetleges fals negatív esetek előfordulási gyakoriságát az RA kimenetelére vonatkozóan. A betegség lefolyásával kapcsolatban azt tapasztaltuk, hogy az első tünetek és a kórházi kezelés átlagéletkora 65 év körüli. A szakirodalomból és a betegek orvosi jelentésének áttekintése során azonban azt láttuk, hogy az esetek kisebb részében az első kórházi kezelés 70 éves korban is jelentkezhet. Ezért úgy döntöttünk, hogy a kontrollok életkori felvételi kritériumát 75 év vagy annál idősebbre emeljük, mert ennek hiányában esetleg olyan alanyok kerültek volna a kontrollcsoportunkba, ahol az RA a preklinikai stádiumban van, és ahol a betegség patomechanizmusa az iníciációs vagy centrális stádiumban vannak, tünetek jelenléte nélkül. 75 éves korukra a betegek nagyobb valószínűséggel jelentkeznek tünetekkel, és egyszer már az ellátórendszer részévé válnak. Így csökkentettük a zavaró hatásokat a vizsgálatunkban. Kontroll személyekre vonatkozó felvételi kritériumaink tehát a következők voltak: (1) 75 éves vagy annál idősebbnek kellett lennie, (2) aláírta a beleegyező nyilatkozatot, (3) az American College of Rheumatology diagnosztikus és besorolási kritériumai alapján nem szenvedett RA-ban, (4) leletanyaga alapján nem szenvedett egyéb autoimmun betegségben (pl. SLE, SSc, psoriasis, Sjögren-szindróma), (5) CRP-értéke  $\leq 20$  mg/dl és Rf szintje  $\leq 25$  IU/ml (153). Végül pilot vizsgálatunk során 100 RA és 100 kontrollszemély vérmintáit sikerült összegyűjtenünk (152).

### III.3.3. Klinikai paraméterek vizsgálata

Az RA és kontroll vérmintákból elsődlegesen teljes vérképet, koleszterinszintet, vércukorszintet, ESR-értéket, CRP-értéket, nátrium-, kálium-, karbamid-, kreatinin-, glutamát-oxalacetát-transzamináz-szintet (GOT), glutamát-piruvát-transzamináz-szintet (GPT), a gamma-glutamiltranszferáz-szintet (gamma-GT), alkalikus foszfatáz, laktát-dehidrogenáz, kreatin-kináz, fehérjealbumin, magnézium, foszfor és a szérum bilirubin-értéket határoztunk meg. Az RA szeropozitivitásának és -negativitásának meghatározására Rf (IgG) és aCCP autoantitest szinteket határoztunk meg mind az RA mind pedig a kontrollcsoportban, az utóbbiban legfőképp azért, hogy az autoimmun álnegatívak ezúton is kizárásra kerüljenek (152,154,155). Az ESR meghatározása a Harkányi Termál Rehabilitációs Központ laboratóriumában, Westergren módszerrel történt (156). A CRP és Rf (IgG) meghatározása szintén a rehabilitációs központban történt, Konelab Arena 20XT klinikai kémiai analizátorral (Thermo Fisher Scientific, Waltham, MA, Egyesült Államok) a gyártó útmutatásainak megfelelően. Az aCCP szintjét Immunoscan CCPlus enzimhez kapcsolt immunsorbent assay kit (SvarLife Sciences, Malmö, Svédország) segítségével vizsgálták a Pécsi Tudományegyetem, Klinikai Központ, Immunológiai és Biotechnológiai Intézetében a

gyártó protokolljának megfelelően. Végül DAS28 pontrendszer alapján kiszámítottuk az RA-s betegek betegségaktivitását (137,152).

### III.3.4. Genomiális DNS extrakciója teljes vérből

Az RA betegek és a kontrollok genomiális DNS-ét (gDNS) 0,5 ml teljes vérből izoláltuk, amelyet 1 ml DNAzol BD reagenssel (Thermo Fisher Scientific, Waltham, USA) lizáltunk a gyártó utasításai szerint. A gDNS kicsapását szobahőmérsékleten 0,4 mL izopropanollal 5 percig végeztük majd a mintát centrifugáltuk (152). A felülúszó eltávolítása után a pelletet először 0,5 mL DNAzol BD reagenssel, majd ezt követően másodszor 1 mL 75%-os etanollal mostuk. A szárított gDNS-t 200 µL Tris-EDTA pufferben oldottuk fel, és NanoDrop One (Thermo Fisher Scientific, Waltham, USA) segítségével kvantitatívan megmértük. Reakciónként 20 ng gDNS-t használtunk a kvantitatív valós idejű PCR vizsgálathoz (qPCR) (152).

### III.3.5. Az *AIRE* rs878081 és rs1003854 SNP-k genotipizálása

A gDNS extrakcióját és tisztítását követően multiplex qPCR vizsgálatot végeztünk az *AIRE* rs878081 és rs1003854 lókuszaiban található allélpolimorfizmusok genotipizálására, melyet mind az RA betegek és kontroll személyek esetében TaqMan technológiával végeztünk (152). A pontos allélikus diszkrimináció érdekében Minor Groove Binder (MGB) TaqMan® Assay-eket (Thermo Fisher Scientific, Waltham, USA) használtunk, ahol az MGB az oligonukleotid próba 3' végében megnöveli az olvadási hőmérsékletet, így stabilizálva a próba-target hibrideket (152). Az MGB TaqMan próbák lényegesen rövidebbek, mint a hagyományos próbák, így jobban és pontosabban képesek megkülönböztetni a nagymértékben homológ allélszekvenciákat (152). Egy lókusza tervezett TaqMan® Assay tartalmaz egy primerpárt, egy VIC festékkel jelölt oligonukleotid próbát (530 nm-es sárga csatorna), amely az 1-es allél szekvenciát detektálja, és egy FAM festékkel jelölt próbát (470 nm-es zöld csatorna), amely a 2-es allél szekvenciáját detektálja. Azok a minták, amelyek kizárólag a VIC vagy FAM csatornákon mutatnak jelet, homozigóta genotípusnak minősülnek az adott szekvencia tekintetében, míg azok a minták, amelyek mindkét csatornán jelet mutatnak, heterozigóta genotípusnak minősülnek az adott lókusznak megfelelően (152). Minőségileg ellenőrzött (quality check, QC) és funkcionálisan tesztelt Taqman Assay-eket használtunk az rs878081 (assay ID: C\_\_2978265\_10) és rs1003854 lókuszon (assay ID: C\_\_9480542\_10), katalógus számuk: 4351379 (152).

A protokollunk szerint 9 µl gDNS-hez (20ng), 1 µl 20x TaqMan® Assay-t és 10 µl 4371357 katalógusszámú TaqMan Genotyping 2x Master Mix-et (Thermo Fisher Scientific, Waltham, USA) adtunk, így kialakítva a reakcióoldatot 20 µl-es össztérfogatát. A fluoreszcencia intenzitását a Rotor-Gene 3000 QPCR gépet (Corbett Life Science, Qiagen, Crawley, Egyesült Királyság) használva a FAM és VIC csatornán, a Rotor-Gene 6 v6.1 szoftverrel detektáltuk (152). A PCR protokoll a következő volt: 10 perc 95°C-on az AmpliTaq Gold® DNS-polimeráz, UP enzimaktiválás és 43 ciklus denaturáció és szintézis (15 másodperc 95°C-on a DNS denaturáció, 1 perc 60°C-on az annealing és az extenzió). Az allél diszkrimináció értékelését a Rotor-Gene 6000 v1.7 szoftver (Corbett Life Science, Qiagen, Crawley, Egyesült Királyság) segítségével végeztük (152). Genotípus-hozzárendelés megerősítése érdekében minden egyes minta spektrumát vizuális ellenőrzéssel felülvizsgáltuk. A minták 100%-át sikeresen genotipizáltuk (152).



A következő lókuszkra tervezett specifikus Taqman Assay-eket használtuk:

**rs878081:**

[http://www.ensembl.org/Homo\\_sapiens/Variation/Population?db=core;r=21:44287894-44288894;v=rs878081;vdb=variation;vf=1078950520](http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=21:44287894-44288894;v=rs878081;vdb=variation;vf=1078950520)

**assay ID: C\_\_2978265\_10; katalógus szám: 4351379**

[https://www.thermofisher.com/order/genome-database/details/genotyping/C\\_\\_2978265\\_10](https://www.thermofisher.com/order/genome-database/details/genotyping/C__2978265_10)  
TCCAGAGAGCTGTGGCCATGTCCTC[C/T]GGGGACGTCCCGGGAGCCCGAGGGG

**rs1003854**

[http://www.ensembl.org/Homo\\_sapiens/Variation/Population?db=core;r=21:44289724-44290724;v=rs1003854;vdb=variation;vf=1078952272](http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=21:44289724-44290724;v=rs1003854;vdb=variation;vf=1078952272)

**assay ID: C\_\_9480542\_10; katalógus szám: 4351379**

[https://www.thermofisher.com/order/genome-database/details/genotyping/C\\_\\_9480542\\_10](https://www.thermofisher.com/order/genome-database/details/genotyping/C__9480542_10)  
CTCTTTTCTTTAATAGACAGTATT[C/T]TTTTCTGATAATACGCAATGGTAA

### III.3.6. Statisztikai analízis

Statisztikai elemzéseinket az SPSS szoftver 28.0 verziójával végeztük (IBM Corp. Released 2021, IBM SPSS Statistics for Windows, Version 28.0; IBM Corp, Armonk, NY, USA). A várható genotípus-gyakoriságokat a Hardy-Weinberg-egyensúly (HWE) alapján számoltuk ki, és a megfigyelt gyakoriságokkal való különbséget statisztikailag összehasonlítottuk nem-paraméteres  $\chi^2$  próbával, hogy lássuk, mennyire illeszkednek a megfigyelt genotípus-gyakoriságok a HWE szerint várható gyakoriságokhoz (152). A klinikai paramétereket jelentő folytonos változókat átlagérték  $\pm$  standard deviáció (SD) formában ábrázoltuk. A normalitás meghatározására Shapiro-Wilk-próbát, a Levene-tesztet pedig a varianciák egyezésének elemzésére használtuk (152). A csoportok összehasonlítására a folytonos változók esetében független mintás t-próbát és varianciaanalízist alkalmaztunk (152). A nem normáloszlású változók mediánjainak összehasonlítására Mann-Whitney U és Kruskal-Wallis H próbákat alkalmaztunk kétszélű szignifikanciával. A korrelációt kétváltozós korrelációval, Pearson korrelációs együtthatóval és kétszélű szignifikanciával vizsgáltuk. Bináris logisztikus regressziót használtunk OR-értékek kiszámítására 95% CI intervallumot alkalmazva. Statisztikai szignifikanciát akkor határoztuk meg, ha a p-értékek kisebbek voltak 0,05-nél. Több okból is a hagyományos 0,05-ös p-értéket választottuk. Először is, ez egy szabványosított küszöbérték a tudományos kutatásban (152). Ezen kívül, tanulmányunk biológiai alapú kutatás, amelynek célja a feltételezett expozíció és a kimenetel közötti új összefüggések kimutatása, ezért el akartuk kerülni ennek a standardizált határértéknek a korrigálását vagy módosítását. Ebben az esetben a 0,05 választása ésszerű, mert elfogadható egyensúlyt teremt az első- és másodfajú hibák között. Ez azt jelenti, hogy az érték ésszerű egyensúlyt hoz létre a túl megengedő – ami a fals pozitív eredmények kockázatának emelkedését okozza – és a túlságosan szigorú értékelés között, ahol a fals negatív eredmények kockázata emelkedett volna. Bár a statisztikai szignifikanciát 0,05-re állítottuk be, a táblázatokban magasabb szignifikanciák ( $< 0,01$ ,  $< 0,001$ ) is megtalálhatók.

### III.4. Az AIRE exon 5 és 6 valamint intron 5 és 7-es régiókban található allépolimorfizmusok vizsgálata reumatoid artritisz kockázatára

### III.4.1. Etikai engedély

Annak érdekében, hogy vizsgálatunkat folytatni tudjuk szükségünk volt további minták gyűjtésére és időre, melyhez elengedhetetlen volt a pilot study kapott etikai engedély módosítására. Ennek során mintagyűjtésünk helyszínét a meglévő Harkányi Termál Rehabilitációs Centrum Közhasznú Nonprofit Kft. mellett a Hévízgyógyfürdő és Szent András Reumakórházzal együttműködésben kibővítettük és kérelmeztünk az etikai engedély meghosszabbítását további négy évre. Benyújtott módosítási kérelmünket a szakhatósági véleményezésre felkért ETT TUKÉB szakmai és etikai szempontból megfelelőnek találta és a kutatási terv kérelemben foglalt módosításának engedélyezéséhez hozzájárult, melyek alapján a Nemzeti Népegészségügyi Központ a módosítási kérelmünket az 57142-5/2022/EÜIG iktatószámú határozatban, további négy évre 2022. november 8-án engedélyezte.

### III.4.2. Mintagyűjtés

A mintagyűjtés inentől kezdve párhuzamosan két intézményben zajlott. A Harkányi Termál Rehabilitációs Centrumban történt RA és kontroll személyek vérmintáinak gyűjtése, míg a Hévízgyógyfürdő és Szent András Reumakórházzal együttműködve kizárólag RA betegek vérmintáinak gyűjtése zajlott. A vizsgálatra jelentkezés jelen esetben is önkéntes alapon történt, melyhez beleegyező nyilatkozatot és betegtájékoztatót csatoltunk. A teljes vérmintákat személyenként 4mL-es citrátcsövekbe vettük, melyet intézetünkben  $-75^{\circ}\text{C}$ -on tároltuk az analízisig. A vizsgálatra önkéntesen vállalkozó személy RA csoportba került, amennyiben (1) 18 évnél idősebb volt, (2) aláírta a beleegyező nyilatkozatot, valamint (3) az American College of Rheumatology diagnosztikus és besorolási kritériumai alapján az RA-t diagnosztizáltak nála (136). Azon RA betegeket, akiknek a leletanyagában fellelhető volt egyidejűleg más autoimmun kórkép is, mint pl. Sjögren-szindróma, pszoriázis, SSc, SLE, kizártuk a további elemzésből, annak érdekében, hogy kizárjuk az RA-tól eltérő autoimmun betegségek a kiválasztott allépolimorfizmusokkal való lehetséges összefüggését, valamint hogy csökkentsük az RA-tól eltérő autoimmun betegségek RA kockázatára mért zavaró hatását. Jelen vizsgálatunkban a pilot studyban alkalmazott kontroll csoport életkori kritériumát használtuk a fals RA negatív esetek elkerülésére, így a kontrollok életkori felvételi kritériumát 75 év vagy annál idősebbre emeltük. Ennek hiányában esetleg olyan alanyok kerültek volna a kontrollcsoportunkba, akik az RA preklinikai, tünetmentes stádiumában vannak. Az önkéntesen jelentkező kontrollokra vonatkozó felvételi kritériumaink ennek megfelelően a következők voltak: (1) 75 éves vagy annál idősebbnek kellett lennie, (2) aláírta a beleegyező nyilatkozatot, (3) az American College of Rheumatology diagnosztikus és besorolási kritériumai alapján nem szenvedett RA-ban, (4) leletanyaga alapján nem szenvedett egyéb autoimmun betegségben (pl. SLE, SSc, psoriasis, Sjögren-szindróma), (5) CRP-értéke  $\leq 20$  mg/dl és Rf szintje  $\leq 25$  IU/ml (153). A teljes mintagyűjtés végére, 2023. decemberére összesen 592 jelentkező vérmintáit sikerült összegyűjteni, melyből 270 RA beteg és 322 kontroll személy volt.

### III.4.3. Klinikai paraméterek vizsgálata

Az 592 személy vérmintáját elsődlegesen a következő klinikai paraméterek meghatározására használtuk: teljes vérkép, koleszterin, vércukor, ESR, CRP, nátrium, kálium, karbamid, kreatinin, GOT, GPT, gamma-GT, alkalikus foszfatáz, laktát-dehidrogenáz, kreatin-kináz, fehérjealbumin, magnézium, foszfor és a szérumbilirubin. Szérumbilirubin (IgG) és aCCP

meghatározás annak érdekében történt, hogy erősítsük RA-csoportunkat, valamint a kontrollból az álnegatívakat kizárjuk (154,155). ESR meghatározása a Harkányi Termál Rehabilitációs Központ valamint a Hévízgyógyfürdő és Szent András Reumakórház laboratóriumában, Westergren módszerrel történt (156). A CRP és Rf (IgG) értékek meghatározása a Harkányi Termál Rehabilitációs Központban, Konelab Arena 20XT klinikai kémiai analizátorral (Thermo Fisher Scientific, Waltham, MA, Egyesült Államok) a gyártó útmutatásainak megfelelően történt. A szérumban aCCP szintet Immunoscan CCPlus enzimhez kapcsolt immunszorbent assay kit (SvarLife Sciences, Malmö, Svédország) segítségével a Pécsi Tudományegyetem, Klinikai Központ, Immunológiai és Biotechnológiai Intézetében határozták meg a gyártó protokolljának megfelelően. Mindezen adatok ismeretében kiszámítottuk az RA-s személyek betegségének aktivitását a DAS28 pontrendszer alapján (137).

#### **III.4.4. Genomiális DNS extrakciója teljes vérből**

Az 592 személy gDNS-ének extrakciójához jelen esetben is 0,5 ml teljes vérhez 1 ml DNAzol BD reagenst (Thermo Fisher Scientific, Waltham, USA) adva lizáltunk a gyártó utasításai szerint. Szobahőmérsékleten a gDNS kicsapását 0,4 mL izopropanollal 5 percig végeztük majd ezt követően a mintát centrifugáltuk. A pelletet először 0,5 mL DNAzol BD reagenssel, majd ezt követően másodszor 1 mL 75%-os etanollal mostuk. A szárított gDNS-t 200 µL Tris-EDTA pufferben oldottuk fel, és NanoDrop One (Thermo Fisher Scientific, Waltham, USA) segítségével kvantitatívan megmértük.

#### **III.4.5. Az *AIRE* allélpolimorfizmusok genotipizálása**

Reakciónként 20 ng gDNS-t használtunk multiplex qPCR vizsgálatunkhoz. Ebben a kiterjedtebb második vizsgálatban az allélpolimorfizmusok genotipizálására szintén TaqMan technológiával alkalmaztunk. A Minor Groove Binder (MGB) TaqMan® Assay-eket (Thermo Fisher Scientific, Waltham, USA) alkalmazva stabilabb próba-target hibrideket sikerült létrehozni az allélikus diszkrimináció során. A pilot vizsgálatunk során sikeresen alkalmaztuk a genotipizáláshoz szükséges termékeket, így a jelen esetben is azt a vizsgálati elrendezést választottuk. TaqMan® Assay-ek tartalmaznak egy primerpárt, egy VIC festékkel jelölt oligonukleotid próbát (530 nm-es sárga csatorna), amely az 1-es allél szekvenciát detektálja, és egy FAM festékkel jelölt próbát (470 nm-es zöld csatorna), amely a 2-es allél szekvenciáját detektálja. Homozigóta a genotípus, ha kizárólag a VIC vagy a FAM csatornán mutatkozik fluoreszcens jel. Heterozigóta a genotípus, amennyiben mindkét csatornán mutatkozik jel. QC és funkcionálisan tesztelt Taqman Assay-eket használtunk az rs878081 (assay ID: C\_\_2978265\_10), rs1003854 (assay ID: C\_\_9480542\_10), rs1003853 (assay ID: C\_\_2978264\_1\_), rs2075876 (assay ID: C\_\_15863141\_20) és rs1055311 (assay ID: C\_\_9480540\_20) lókuszon, katalógus számuk: 4351379. Jelen esetben a pilot vizsgálat során sikeresen alkalmazott protokollt alkalmaztuk, mely szerint 9 µl gDNS-hez (20ng), 1 µl 20x TaqMan® Assay-t és 10 µl 4371357 katalógusszámú TaqMan Genotyping 2x Master Mix-et (Thermo Fisher Scientific, Waltham, USA) adtunk, így kialakítva 20 µl-es ösztérfogatot. A fluoreszcencia intenzitását ez esetben is a Rotor-Gene 3000 QPCR gépet (Corbett Life Science, Qiagen, Crawley, Egyesült Királyság) használva a FAM és VIC csatornán, a Rotor-Gene 6 v6.1 szoftverrel detektáltuk. A PCR protokoll során 10 percig 95°C-on az AmpliTaq Gold® DNS-polimeráz, UP enzimének aktiválása melyet 43 ciklus denaturáció és szintézis követett (15 másodperc 95°C-on a DNS denaturáció, 1 perc 60°C-on

az annealing és az extenzió). Az allél diszkrimináció értékelését a Rotor-Gene 6000 v1.7 szoftver (Corbett Life Science, Qiagen, Crawley, Egyesült Királyság) segítségével végeztük. Minden egyes minta spektrumát vizuális ellenőrzéssel felülvizsgáltuk, annak érdekében, hogy a genotípusok hozzárendelését megerősítsük. Ebben a vizsgálatban is sikerült a minták 100%-át sikeresen genotipizálni.

A következő lókuszkra tervezett specifikus Taqman Assay-eket használtuk:

**rs878081:**

[http://www.ensembl.org/Homo\\_sapiens/Variation/Population?db=core;r=21:44287894-44288894;v=rs878081;vdb=variation;vf=1078950520](http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=21:44287894-44288894;v=rs878081;vdb=variation;vf=1078950520)

**assay ID: C\_\_2978265\_10; katalógus szám: 4351379**

[https://www.thermofisher.com/order/genome-database/details/genotyping/C\\_\\_2978265\\_10](https://www.thermofisher.com/order/genome-database/details/genotyping/C__2978265_10)  
TCCAGAGAGCTGTGGCCATGTCCTC[C/T]GGGGACGTCCCGGGAGCCCGAGGGG

**rs1003854**

[http://www.ensembl.org/Homo\\_sapiens/Variation/Population?db=core;r=21:44289724-44290724;v=rs1003854;vdb=variation;vf=1078952272](http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=21:44289724-44290724;v=rs1003854;vdb=variation;vf=1078952272)

**assay ID: C\_\_9480542\_10; katalógus szám: 4351379**

[https://www.thermofisher.com/order/genome-database/details/genotyping/C\\_\\_9480542\\_10](https://www.thermofisher.com/order/genome-database/details/genotyping/C__9480542_10)  
CTCTTTTCTTTAATAGACAGTATT[C/T]TTTTCTGATAATACGCAATGGTAA

**rs1003853**

[http://www.ensembl.org/Homo\\_sapiens/Variation/Population?db=core;r=21:44288465-44289465;v=rs1003853;vdb=variation;vf=722412392](http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=21:44288465-44289465;v=rs1003853;vdb=variation;vf=722412392)

**assay ID: C\_\_2978264\_1\_1; katalógus szám: 4351379**

[https://www.thermofisher.com/order/genome-database/details/genotyping/C\\_\\_2978264\\_1\\_1](https://www.thermofisher.com/order/genome-database/details/genotyping/C__2978264_1_1)  
GCAGGGCCGCTGGTGGCAGACCCAC[C/T]GTGCCATCGGGGCATTCCATCTCAA

**rs2075876**

[https://www.ensembl.org/Homo\\_sapiens/Variation/Population?db=core;r=21:44288770-44289770;v=rs2075876;vdb=variation;vf=1078962793](https://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=21:44288770-44289770;v=rs2075876;vdb=variation;vf=1078962793)

**assay ID: C\_\_15863141\_20; katalógus szám: 4351379**

[https://www.thermofisher.com/order/genome-database/details/genotyping/C\\_\\_15863141\\_20?pluginName=&CID=&ICID=](https://www.thermofisher.com/order/genome-database/details/genotyping/C__15863141_20?pluginName=&CID=&ICID=)  
TGTTCTTGTCTGCTGCTGTGAGG[A/G]TAGTAGGTCTACTGTGCACAGACCC

**rs1055311**

[https://www.ensembl.org/Homo\\_sapiens/Variation/Population?db=core;r=21:44289185-44290185;v=rs1055311;vdb=variation;vf=1078953309](https://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=21:44289185-44290185;v=rs1055311;vdb=variation;vf=1078953309)

**assay ID: C\_\_9480540\_20; katalógus szám: 4351379**

[https://www.thermofisher.com/order/genome-database/details/genotyping/C\\_\\_9480540\\_20?CID=&ICID=&subtype=](https://www.thermofisher.com/order/genome-database/details/genotyping/C__9480540_20?CID=&ICID=&subtype=)  
GCTCCAAGAAGTGCATCCAGGTTGG[C/T]GGGGAGTTCTACACTCCCAGCAAGT

### III.4.6. Statisztikai analízis

Statisztikai elemzéseinket a vizsgálat során SPSS szoftver 28.0 verziójával végeztük (IBM Corp. Released 2021, IBM SPSS Statistics for Windows, Version 28.0; IBM Corp, Armonk,

NY, USA). A HWE alapján várható és megfigyelt genotípus frekvenciákat nem-paraméteres  $\chi^2$  próbával vizsgáltuk, hogy lássuk van-e szignifikáns eltérés a kettő között. A kapcsoltságot a Lewontin-féle „D” módszerrel értékeltük, ahol az LD-struktúrát magába foglaló haplotípus blokkot a Haploview szoftver (4.2. verzió) segítségével generáltuk (157). A klinikai folytonos változókat átlagérték  $\pm$  SD formában ábrázoltuk. A normalitás meghatározására Shapiro-Wilk-próbát, a Levene-tesztet pedig a varianciák egyezésének elemzésére használtuk. A csoportok összehasonlítására a folytonos változók esetében független mintás t-próbát és varianciaanalízist alkalmaztunk. Mann-Whitney U és Kruskal-Wallis H próbákat alkalmaztunk kétszélű szignifikanciával a nem normáloszlású változók mediánjainak összehasonlítására. A korrelációt kétváltozós korrelációval, Pearson korrelációs együtthatóval és kétszélű szignifikanciával vizsgáltuk. OR-értékek kiszámítására bináris logisztikus regressziót alkalmaztunk, melynek során 95% CI és p-értéket generáltunk. Statisztikai szignifikanciát a 0,05 értéknél kisebb p-érték jelentette.

### **III.5. AIRE allélpolimorfizmusok regulatórikus kötőmotívumokra gyakorolt hatásainak *in silico* elemzése**

Annak érdekében, hogy meghatározzuk, az egyes allélpolimorfizmusoknak a jelenléte milyen hatással lehet a transzkripciós faktorok kötődési affinitására, azokat az allélokat, melyek statisztikailag szignifikáns asszociációt mutattak RA kockázatával illetve korrelációt klinikai paraméterekkel *in silico* elemeztük a HaploReg v4.2 adatbázis segítségével (158,159).

Az adatbázisban a referencia SNP (rs)ID számot begépelve az adott lókusza a következő alapadatokat kapjuk meg egy összefoglaló táblázatban: melyik kromoszómán (chr) és hol (pos) helyezkedik el; linkage disequilibrium (LD ( $r^2$ ,  $D'$ )) értékek; mely allélpolimorfizmusok találhatóak rajta (Ref, Alt); mi a MAF az egyes alpopulációkban (afrikai: AFR, amerikai: AMR, ázsiai: ASN, európai: EUR); transzkripciós faktor (TF) DNS kötőmotívumaihoz való kötődési affinitás az egyes allélpolimorfizmusok jelenlétében mutatnak-e különbséget, és ha igen, mely TF esetében (Motifs changed); a lókuszon jelen lévő allélpolimorfizmusok szerepelnek expressziós adatbázisban (Selected eQTL hits); melyik génkódoló régiójában foglal helyet (GENCODE genes); mi az allélpolimorfizmus annotált funkcionális következménye (dbSNP func annot) (159).

Vizsgálatunk során arra voltunk kíváncsiak, hogy az egyes allélpolimorfizmusok helyet foglalnak-e bizonyos TF-k DNS kötőmotívumának motívumaiban, és ha igen akkor hatással vannak-e az egyes allélok a TF kötődési affinitására (Regulatory motifs altered). Az erre vonatkozó eredményeket a HaploReg adatbázis helysúlyozott mátrix (position weight matrix, PWM) predikciós modell segítségével adja meg, melyhez forrásként TF-kat, genomialis kötőhelyeket valamint DNS-TF kötődési affinitás értékeket tartalmazó adatbázisokat (TRANSFAC, JASPAR) illetve fehérje-kötődési microarray vizsgálatok (protein-binding microarray, PBM) eredményeit használják fel (160-163). A Ref- és Alt-allélokat mindkét oldalon 29 bp-nyi genomialis kontextussal importálták UCSC Genome Browser adatbázisából (146).

A PWM egy scoring rendszer, mely egyrészt az adott nukleotid előfordulási gyakoriságát adja meg a kötőhely minden pozíciójában, másrészt az adott nukleotidhoz tartozó TF kötődési affinitását is mutatja. PWM-pontszám kiszámításához összegezni kell az egyes nukleotidok logaritmusos valószínűségeit az egyes pozíciókban (159). A magasabb PWM-pontszámmal rendelkező szekvenciák általában nagyobb kötődési affinitásnak jelentenek, mivel jobban megfelelnek a TF preferált kötőhelyének. A PWM pontszám egy valószínűségi érték, melyet esélyhányadosá alakítása után, annak 10-es alapú logaritmusát véve kapjuk meg a logarithm of odds (LOD)-értéket, mely kiszámítható ugyanabban a pozícióban lévő Ref- és Alt-allélok

adta kötőmotívum szekvenciákra (159). Relatív magasabb LOD-értékből kivonva az alacsonyabbat a kapott eredmény megmutatja, melyik allél biztosítja a kötőmotívumnak magasabb prediktív kötődési affinitást a TF-hoz (159).

A kapott eredményeket táblázat összesíti, melynek Position Weight Matrix ID oszlopa azon TF-at jelöli melyek kötőmotívum szekvenciájában megtalálható az adott SNP-t adó lókus. 'Strand' oszlop azt mutatja, hogy a TF melyik DNS szálát, a pozitív (sense) vagy negatív (antisense) szálát preferálja (159). A Ref és Alt oszlop értékei az adott Ref és Alt allélpolimorfizmushoz tartozó LOD-értékek, míg az utolsó oszlop a kötőmotívum lehetséges szekvenciáját mutatja (159).

## IV. Eredmények

### IV.1. Az *AIRE* allélpolimorfizmusok meta-analízise autoimmun kórképekben

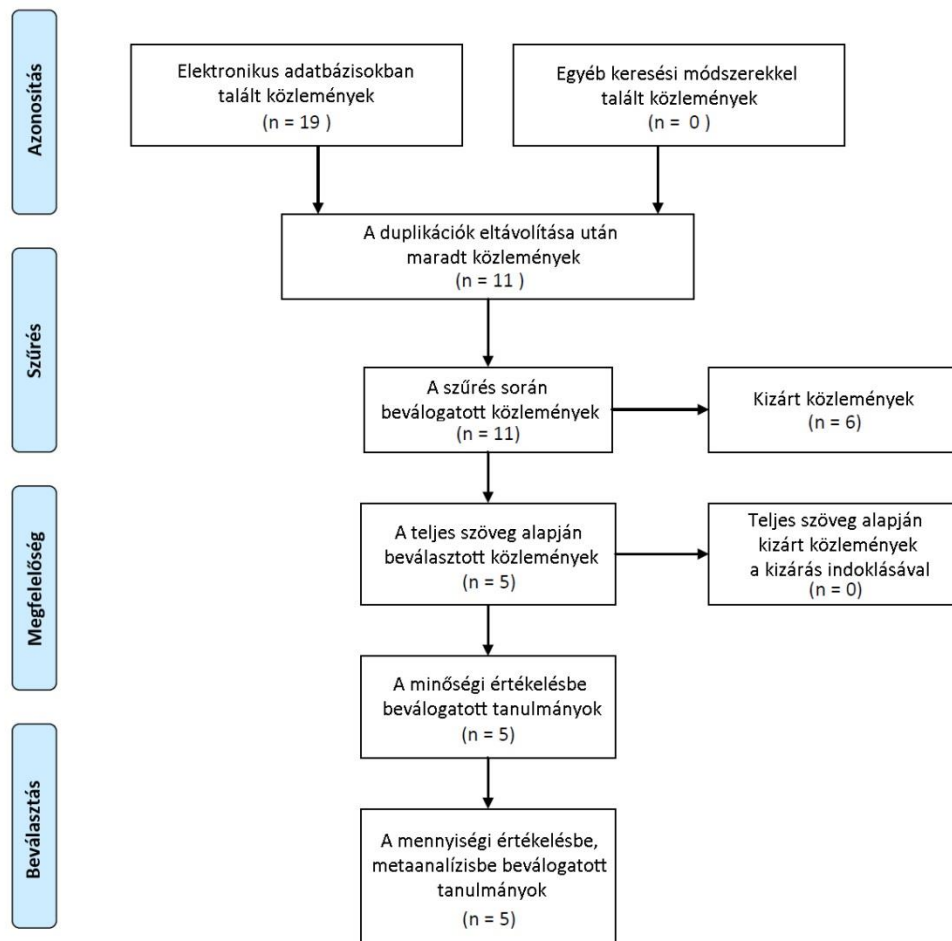
#### IV.1.1. Beválasztott közlemények jellemzői

Keresés során a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) irányelveket követtük, melynek eredményeképpen Embase, PubMed, Cochrane Library, és a Web of Science adatbázisokban 19 releváns közleményt azonosítottunk. A duplikátumok eltávolítását követően a maradék 11 közleményt vizsgálva azt tapasztaltuk, hogy egyedül a RA esetében öt, nagy elemszámú eset-kontroll vizsgálat volt jelen, míg a többi autoimmun betegség esetében pusztán 1 vizsgálatot publikáltak, így meta-analízisünket az RA irányába állítottuk. A PRISMA irányelvek szerinti keresésünket az **1. ábra** mutatja. A beválasztott eset-kontroll vizsgálatok ázsiai és kaukázusi populációkat vizsgáltak. A RA diagnózisa és besorolása kivétel nélkül minden betegnél az ACR 1987-es kritériumai alapján történt (136). A teljes RA betegpopulációnak átlagéletkora  $54,1 \pm 2,4$  év, ebből a női esetek megoszlása 73,34% volt. A genotipizálás során microarray, SNaPshot és Taqman SNP-Assay vizsgálatok alkalmaztak. Az öt beválasztott eset-kontroll analízis vizsgálata során az rsID-szám alapján 11 SNP-t azonosítottunk az *AIRE* kódoló génszakaszon (rs2075876, rs760426, rs1800250, rs2776377, rs878081, rs1055311, rs933150, rs1003854, rs2256817, rs374696, rs1078480), melyek közül egyedül az rs2075876 és rs760426 allélpolimorfizmust vizsgálta négynél több eset-kontroll analízis. Ennek megfelelően elvégeztük az rs2075876 (7145 eset, 8579 kontroll) valamint az rs760426 (6696 eset, 8164 kontroll) SNP meta-analízisét. A kontrollcsoportok genotípus-frekvencia értékei Hardy-Weinberg egyensúlyban (HWE) voltak. A beválasztott vizsgálatok jellemzőit az **I. táblázat** tartalmazza.

#### IV.1.2. Az *AIRE* rs2075876 (G/A) SNP meta-analízise

A beválogatott közlemények közül öt vizsgálat elemezte az rs2075876 (G/A) SNP és az RA fogékonysága közötti összefüggést (105). A beválasztott közlemények allélokkal számoltak, így a résztvevők számát megduplázták. Annak érdekében, hogy a számolás egységes legyen, mi is ezt követtük. A publikációk közül *Terao és mtsai.* által készített GWAS három egymástól független eset-kontroll vizsgálatot foglalt magában (A, B, C-vel jelölve) (105). *García-Lozano és mtsai.* közleményének kivételével az összes vizsgálat részletesen leírta az egyes GG, AG, AA genotípusok megoszlását, ebből adódóan esélyhányadosokat tudtunk

számolni olyan esetekben, ha azok nem álltak rendelkezésre (**II. táblázat**). Az egyes genetikai modellekre kapott eredményeket a **2. ábra** mutatja.



**1. ábra: A meta-analízisbe beválogatott és a kizárt közlemények PRISMA irányelvek szerinti folyamatábrája**

Az allélikus modellre (A vs. G, **2. ábra**) számolt esélyhányadosok értékei: 1,21 (95% CI 1,09-1,36,  $p < 0,001$ ), 1,18 (1,07-1,30,  $p = 0,001$ ), 1,15 (1,06-1,24,  $p < 0,001$ ), 1,02 (0,42-2,42,  $p = 0,964$ ), 1,32 (1,04-1,69,  $p = 0,021$ ), 1,30 (1,12-1,50,  $p < 0,00$ ), és 1,41 (1,16-1,70,  $p < 0,001$ ) (105). A domináns modellre (AG+AA vs. GG, **2. ábra**) számolt esélyhányadosok értékei: 1,18, (95% CI 1,06-1,32,  $p = 0,002$ ), 1,31 (1,19-1,45,  $p < 0,001$ ), 1,18 (1,09-1,27,  $p < 0,001$ ), 1,41 (1,08-1,84,  $p = 0,010$ ), 1,55 (1,32-1,82,  $p < 0,001$ ), és 1,48 (1,22-1,78,  $p < 0,001$ ) (105). Recesszív modellre (AA vs. AG+GG, **2. ábra**) számolt OR értékek: 1,53 (95% CI 1,31-1,79,  $p < 0,001$ ), 1,09 (0,95-1,26,  $p = 0,204$ ), 1,25 (1,12-1,39,  $p < 0,001$ ), 1,52 (1,13-2,05,  $p = 0,006$ ), 1,25 (1,05-1,49,  $p = 0,010$ ), és 1,78 (1,36-2,35,  $p < 0,001$ ) (105). Kodomináns heterozigóta modell (AG vs. GG) esetében számolt esélyhányadosok értékei: 1,08 (0,96-1,21,  $p = 0,168$ ), 1,32 (1,20-1,47,  $p < 0,001$ ), 1,13 (1,05-1,23,  $p = 0,002$ ), 1,28 (0,97-1,70,  $p = 0,077$ ), 1,51 (1,28-1,79,  $p < 0,001$ ), és 1,34 (1,10-1,64,  $p = 0,003$ ) (105).

		Publikálás éve	Ország	Etnicitás	Diagnosztikus kritérium	Genotipizálás	Átlagéletkor		Nők %		Kontroll forrása	
							eset	kontroll	eset	kontroll		
<b>SNP rs2075876 (G/A)</b>	<b>Terao C és mtsai.</b>	A	2011	Japán	ázsiai	Revised criteria (1987) of the American College of Rheumatology for RA	microarray	63,0±12,5	52,0±15,8	82,1	60,6	K
		B	2011				microarray	60,8±11,5	38,1±11,9	84,1	39,6	K
		C	2011				microarray	61,4±11,5	52,5±15,2	81,4	44,4	K
	<b>García-Lozano JR és mtsai.</b>		2013	Spanyolország	kaukázusi	Revised criteria (1987) of the American College of Rheumatology for RA	Taqman SNP genotyping assay	49,23 ±14,8	NA	74,3	NA	K
	<b>Shao S és mtsai.</b>		2014	Kína	ázsiai	Revised criteria (1987) of the American College of Rheumatology for RA	SNaPshot assay	48,69±14,18	47,02±16,34	80,6	37,0	K
	<b>Feng ZJ és mtsai.</b>		2015	Kína (Han)	ázsiai	Revised criteria (1987) of the American College of Rheumatology for RA	Taqman SNP genotyping assay	54,1 ± 11,2	52,4 ± 11,8	53,5	58,5	K
<b>Li X és mtsai.</b>		2016	Kína (Shaanxi)	ázsiai	Revised criteria (1987) of the American College of Rheumatology for RA	Snapshot Assay	43,5 ± 19,2	44,3 ± 17,8	64,3	59,7	K	
<b>SNP rs760426 (A/G)</b>	<b>Terao C és mtsai.</b>	A	2011	Japán	ázsiai	Revised criteria (1987) of the American College of Rheumatology for RA	microarray	63,0±12,5	52,0±15,8	82,1	60,6	K
		B	2011				microarray	60,8±11,5	38,1±11,9	84,1	39,6	K
		C	2011				microarray	61,4±11,5	52,5±15,2	81,4	44,4	K
	<b>Shao S és mtsai.</b>		2014	Kína	ázsiai	Revised criteria (1987) of the American College of Rheumatology for RA	SNaPshot assay	48,69±14,18	47,02±16,34	80,6	37,0	K
	<b>Feng ZJ és mtsai.</b>		2015	Kína (Han)	ázsiai	Revised criteria (1987) of the American College of Rheumatology for RA	Taqman SNP genotyping assay	54,1 ± 11,2	52,4 ± 11,8	53,5	58,5	K
	<b>Li X és mtsai.</b>		2016	Kína (Shaanxi)	ázsiai	Revised criteria (1987) of the American College of Rheumatology for RA	Snapshot Assay	43,5 ± 19,2	44,3 ± 17,8	64,3	59,7	K

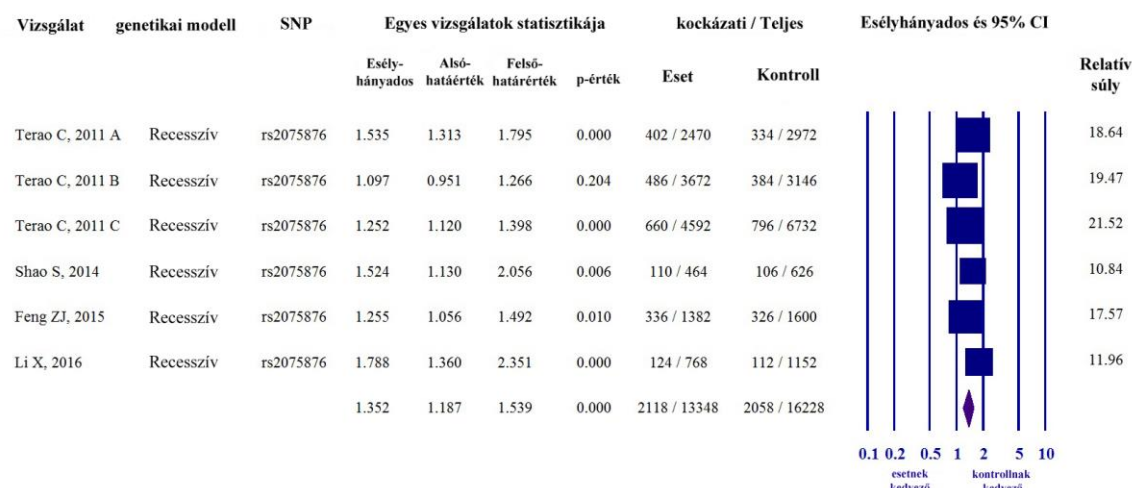
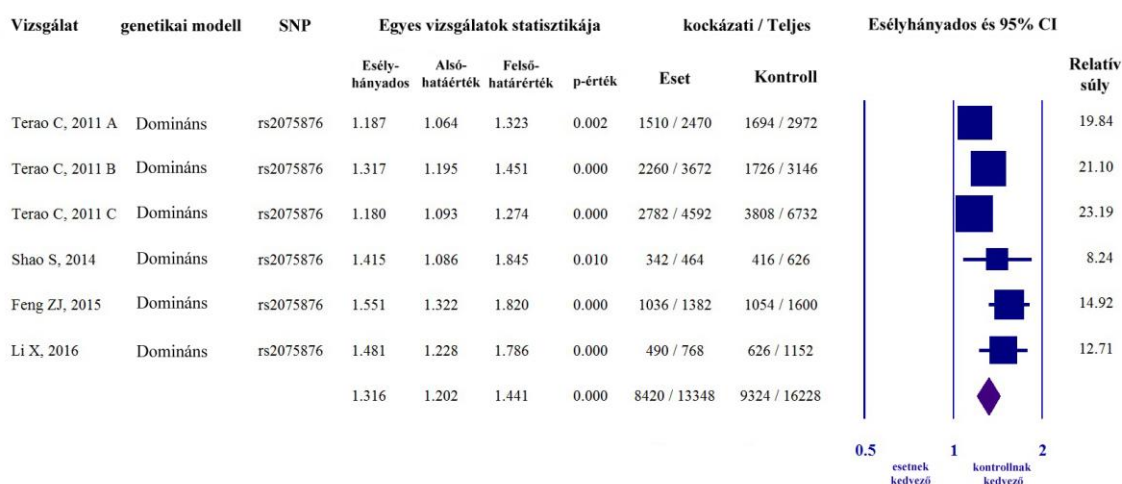
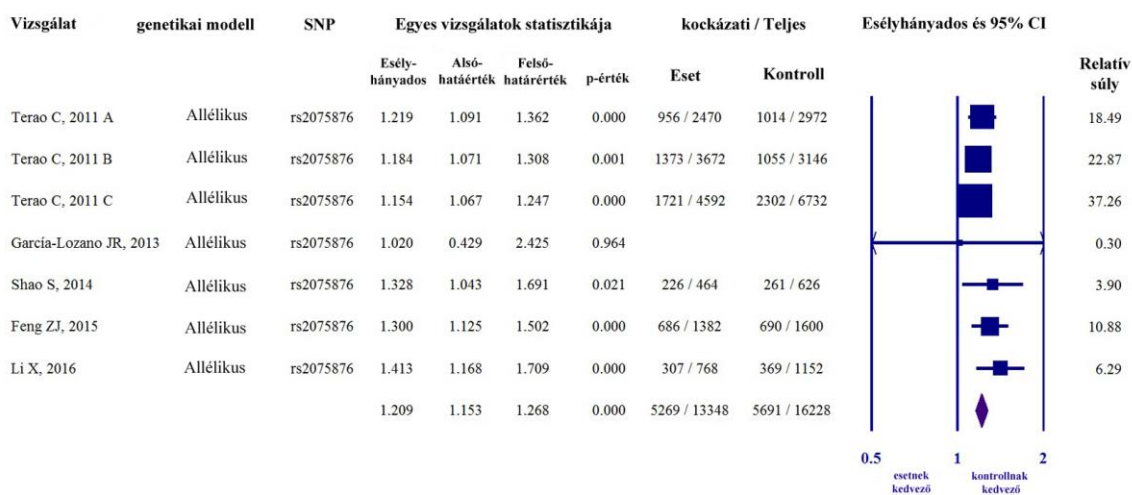
I. táblázat: rs2075876 (G/A) és rs760426 (A/G) SNP meta-analíziséhez beválasztott vizsgálatok jellemzői (NA: nincs adat, K: kórház)



Polimorfizmus	Publikáció	Genetikai modell		OR	95%CI	p	
SNP rs2075876 (G/A)	Terao C, 2011	A	Allelikus*	(A vs G)	1,21	1,09-1,36	< 0,001
			Domináns	(AG+AA vs GG)	1,18	1,06-1,32	0,002
			Recesszív	(AA vs AG+GG)	1,53	1,31-1,79	< 0,001
			Kodomináns heterozigóta	(AG vs GG)	1,08	0,96-1,21	0,168
			Kodomináns homozigóta	(AA vs GG)	1,60	1,35-1,89	< 0,001
		B	Allelikus*	(A vs G)	1,18	1,07-1,30	< 0,001
			Domináns	(AG+AA vs GG)	1,31	1,19-1,45	< 0,001
			Recesszív	(AA vs AG+GG)	1,09	0,95-1,26	0,204
			Kodomináns heterozigóta	(AG vs GG)	1,32	1,20-1,47	< 0,001
			Kodomináns homozigóta	(AA vs GG)	1,27	1,09-1,48	0,002
		C	Allelikus*	(A vs G)	1,15	1,06-1,24	< 0,001
			Domináns	(AG+AA vs GG)	1,18	1,09-1,27	< 0,001
			Recesszív	(AA vs AG+GG)	1,25	1,12-1,39	< 0,001
			Kodomináns heterozigóta	(AG vs GG)	1,13	1,05-1,23	0,002
			Kodomináns homozigóta	(AA vs GG)	1,34	1,19-1,51	< 0,001
	García-Lozano JR, 2013	Allelikus	(A vs G)	1,02	0,42-2,42	0,964	
		Domináns	(AG+AA vs GG)	NA			
		Recesszív	(AA vs AG+GG)				
		Kodomináns heterozigóta	(AG vs GG)				
		Kodomináns homozigóta	(AA vs GG)				
	Shao S, 2014	Allelikus*	(A vs G)	1,32	1,04-1,69	0,021	
		Domináns	(AG+AA vs GG)	1,41	1,08-1,84	0,010	
		Recesszív	(AA vs AG+GG)	1,52	1,13-2,05	0,006	
		Kodomináns heterozigóta	(AG vs GG)	1,28	0,97-1,70	0,077	
		Kodomináns homozigóta	(AA vs GG)	1,78	1,26-2,52	0,001	
	Feng ZJ, 2015	Allelikus	(A vs G)	1,30	1,12-1,50	< 0,001	
		Domináns	(AG+AA vs GG)	1,55	1,32-1,82	< 0,001	
		Recesszív	(AA vs AG+GG)	1,25	1,05-1,49	0,010	
		Kodomináns heterozigóta	(AG vs GG)	1,51	1,28-1,79	< 0,001	
Kodomináns homozigóta		(AA vs GG)	1,62	1,32-1,99	< 0,001		
Li X, 2016	Allelikus*	(A vs G)	1,41	1,16-1,70	< 0,001		
	Domináns	(AG+AA vs GG)	1,48	1,22-1,78	< 0,001		
	Recesszív	(AA vs AG+GG)	1,78	1,36-2,35	< 0,001		
	Kodomináns heterozigóta	(AG vs GG)	1,34	1,10-1,64	0,003		
	Kodomináns homozigóta	(AA vs GG)	2,09	1,56-2,81	< 0,001		

II. táblázat: rs2075876 (G/A) SNP vizsgálata során, az egyes genetikai modellekhez számolt OR, 95%CI és p-érték (NA = nincs adat; OR = odds ratio; CI = konfidencia intervallum)

A kodomináns homozigóta modellre (AA vs. GG) számolt OR értékek: 1,60 (95% CI 1,35-1,89, p<0,001), 1,27 (1,09-1,48, p=0,002), 1,34 (1,19-1,51, p<0,001), 1,78 (1,26-2,52, p=0,001), 1,62 (1,32-1,99, p<0,001), és 2,09, (1,56-2,81, p<0,001) (105).



**2. ábra: Allélikus (A vs. G), domináns (AG+AA vs. GG) és recesszív (AA vs. AG+GG) számított forest plot analízisek, mely az rs2075876 (G/A) SNP és a RA kockázata közötti összefüggést mutatja (CI: konfidencia intervallum)**

A heterogenitási tesztek eredménye az allélikus modell esetében  $P_h=0,439$ ,  $I^2=0\%$ ; domináns modellnél  $P_h=0,011$ ,  $I^2=66,2\%$ ; recesszív modellnél  $P_h=0,005$ ,  $I^2=69,9\%$ ; kodomináns heterozigóta modell esetében  $P_h=0,004$ ,  $I^2=70,5\%$ ; kodomináns homozigóta modellnél  $P_h=0,012$ ,  $I^2=65,4\%$  (105). Mérsékelt heterogenitást találtunk a domináns, recesszív, kodomináns heterozigóta és kodomináns homozigóta modelleknél. A 31 esélyhányados közül

egyedül 4 bizonyult statisztikailag inszignifikánsnak, 27 szignifikáns OR érték egyezményesen azt mutatta, hogy az rs2075876 (G/A) SNP előfordulásának esélye nagyobb a RA esetcsoportban szemben a kontroll csoporttal, így a kapott eredmények kapcsolatot mutattak az AIRE transzkripciós faktort kódoló DNS-szekvenciában található rs2075876 (G/A) SNP és RA fogékonyság között (105).

#### IV.1.3. Az AIRE rs760426 (A/G) SNP meta-analízise

A beválasztott közlemények közül négy vizsgálat elemezte az rs760426 (A/G) SNP és a RA fogékonysága közötti összefüggést. A beválasztott közlemények allélokkal számoltak, így a résztvevők számát megduplázták. Annak érdekében, hogy számolásunk egységes legyen, mi is ezt követtük. Ez esetben is, *Terao és mtsai.* által készített GWAS analízis három egymástól független eset-kontroll vizsgálattal szolgált (A, B, C-vel jelölve). *Feng és mtsai.* közleményének kivételével az összes vizsgálat részletesen leírta az egyes AA, GA, GG genotípusok megoszlását, ebből adódóan esélyhányadosokat tudunk számolni olyan esetekben, ha azok az eredeti közleményekben nem álltak rendelkezésre (**III. táblázat**). *Feng és mtsai.* aszimmetrikus esélyhányadost publikált, így a továbbiakban kizártuk a kapcsolódó OR, 95% CI és p-értékeket a meta-analízisünkből. Az egyes genetikai modellekre kapott eredményeket a **3. ábra** mutatja (105).

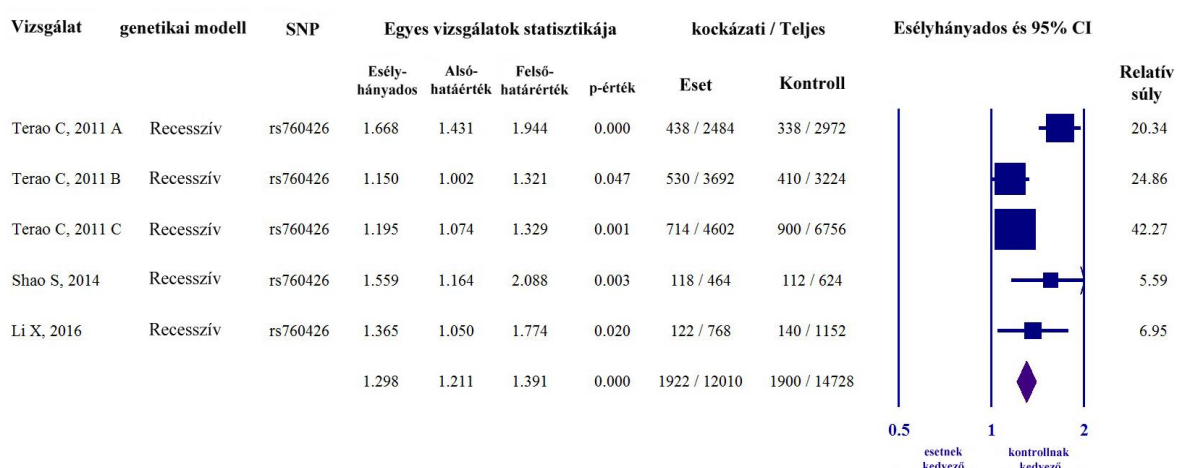
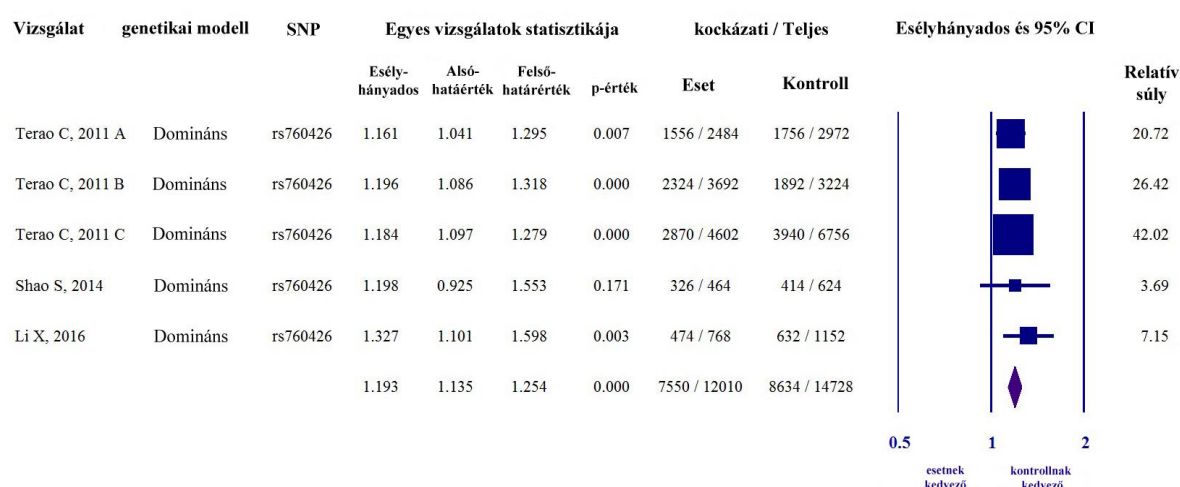
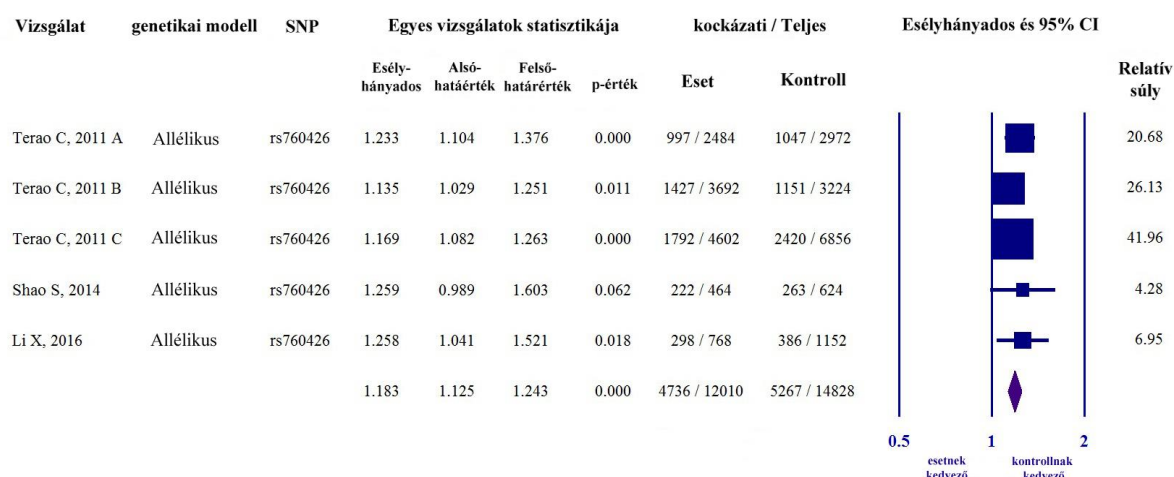
Az allélikus modellre (G vs. A, **3. ábra**) számolt esélyhányadosok értékei: 1,23 (95% CI 1,10-1,37,  $p < 0,001$ ), 1,13 (1,02-1,25,  $p = 0,011$ ), 1,16 (1,08-1,26,  $p < 0,001$ ), 1,25 (0,98-1,60,  $p = 0,062$ ), 1,25 (1,04-1,52,  $p = 0,018$ ). A domináns modellre számolt esélyhányadosok értékei (GG+GA vs. AA, **3. ábra**): 1,16 (1,04-1,29,  $p = 0,007$ ), 1,19 (1,08-1,31,  $p < 0,001$ ), 1,18 (1,09-1,27,  $p < 0,001$ ), 1,19 (0,92-1,55,  $p = 0,171$ ), és 1,32 (1,10-1,59,  $p = 0,003$ ). Recesszív modellre (GG vs. GA+AA, **3. ábra**) számolt esélyhányadosok értékei: 1,66 (95%CI 1,43-1,94,  $p < 0,001$ ), 1,15 (1,00-1,32,  $p = 0,047$ ), 1,19 (1,07-1,32,  $p = 0,001$ ), 1,55 (1,16-2,08,  $p = 0,003$ ), és 1,36 (1,05-1,77,  $p = 0,020$ ). Kodomináns heterozigóta modellre (GA vs. AA) számolt esélyhányadosok értékei: 1,03, (95% CI 0,92-1,16,  $p = 0,582$ ), 1,17 (1,06-1,30,  $p = 0,002$ ), 1,15 (1,06-1,25,  $p = 0,001$ ), 1,04 (0,79-1,38,  $p = 0,741$ ), and 1,26 (1,03-1,54,  $p = 0,020$ ). Kodomináns homozigóta modellre (GG vs. AA) számolt esélyhányadosok értékei: 1,69 (1,43-2,00,  $p < 0,001$ ), 1,25 (1,08-1,46,  $p = 0,003$ ), 1,29 (1,15-1,44,  $p < 0,001$ ), 1,60 (1,15-2,24,  $p = 0,006$ )

Polimorfizmus	Publikáció	Genetikai modell		OR	95%CI	p	
SNP rs760426 (A/G)	Terao C és mtsai.	A	Allelikus*	(G vs A)	1,23	1,10-1,37	< 0,001
			Domináns	(GG+GA vs AA)	1,16	1,04-1,29	0,007
			Recesszív	(GG vs GA+AA)	1,66	1,43-1,94	< 0,001
			Kodomináns heterozigóta	(GA vs AA)	1,03	0,92-1,16	0,582
			Kodomináns homozigóta	(GG vs AA)	1,69	1,43-2,00	< 0,001
		B	Allelikus*	(G vs A)	1,13	1,02-1,25	0,011
			Domináns	(GG+GA vs AA)	1,19	1,08-1,31	< 0,001
			Recesszív	(GG vs GA+AA)	1,15	1,00-1,32	0,047
			Kodomináns heterozigóta	(GA vs AA)	1,17	1,06-1,30	0,002
			Kodomináns homozigóta	(GG vs AA)	1,25	1,08-1,46	0,003

		<b>C</b>	Allelikus*	(G vs A)	1,16	1,08-1,26	< 0,001	
			Domináns	(GG+GA vs AA)	1,18	1,09-1,27	< 0,001	
			Recesszív	(GG vs GA+AA)	1,19	1,07-1,32	0,001	
			Kodomináns heterozigóta	(GA vs AA)	1,15	1,06-1,25	0,001	
			Kodomináns homozigóta	(GG vs AA)	1,29	1,15-1,44	< 0,001	
	<b>Shao S és mtsai.</b>			Allelikus*	(G vs A)	1,25	0,98-1,60	0,062
				Domináns	(GG+GA vs AA)	1,19	0,92-1,55	0,171
				Recesszív	(GG vs GA+AA)	1,55	1,16-2,08	0,003
				Kodomináns heterozigóta	(GA vs AA)	1,04	0,79-1,38	0,741
				Kodomináns homozigóta	(GG vs AA)	1,60	1,15-2,24	0,006
	<b>Feng ZJ és mtsai.</b>			Allelikus*	(G vs A)	1,87	1,09-2,45	0,074
				Domináns	(GG+GA vs AA)	<b>NA</b>		
				Recesszív	(GG vs GA+AA)			
				Kodomináns heterozigóta	(GA vs AA)			
				Kodomináns homozigóta	(GG vs AA)			
	<b>Li X és mtsai.</b>			Allelikus*	(G vs A)		1,25	1,04-1,52
				Domináns	(GG+GA vs AA)	1,32	1,10-1,59	0,003
				Recesszív	(GG vs GA+AA)	1,36	1,05-1,77	0,020
				Kodomináns heterozigóta	(GA vs AA)	1,26	1,03-1,54	0,020
				Kodomináns homozigóta	(GG vs AA)	1,54	1,16-2,04	0,003

**III. táblázat: rs760426 (G/A) SNP vizsgálata során, az egyes genetikai modellekhez számolt OR, 95%CI és p-érték (NA: nincs adat, OR: odds ratio, CI:konfidencia intervallum)**

és 1,54 (1,16-2,04, p=0,003) (105). A heterogenitási tesztek eredménye az allélikus modell esetében  $P_h=0,737$ ,  $I^2=0\%$ ; domináns modellnél  $P_h=0,822$ ,  $I^2=0\%$ ; recesszív modellnél  $P_h=0,001$ ,  $I^2=76,7\%$ ; kodomináns heterozigóta modell esetében  $P_h=0,323$ ,  $I^2=14,2\%$ ; kodomináns homozigóta modellnél  $P_h=0,038$ ,  $I^2=60,5\%$ . Mérsékelt heterogenitást találtunk a recesszív és kodomináns homozigóta modelleknél (105).

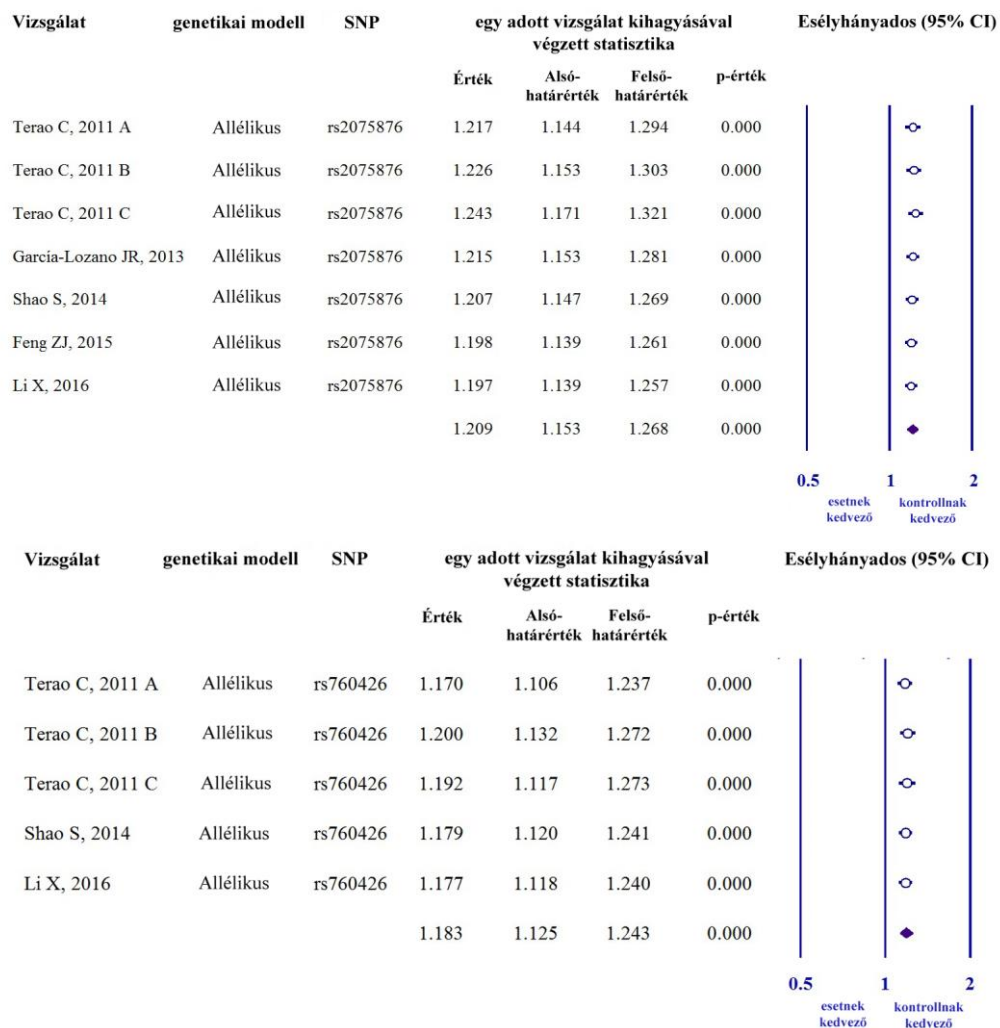


**3. ábra: Allélikus (G vs. A), domináns (GG+GA vs. AA) és recesszív modellre (GG vs. GA+AA) számított forest plot analízisek, mely az rs760426 (A/G) SNP és a RA kockázata közötti összefüggést mutatja (CI: konfidencia intervallum)**

A 26 esélyhányados közül egyedül 4 bizonyult statisztikailag inszignifikánsnak, 22 szignifikáns OR érték egyezményesen azt mutatta, hogy az rs760426 (A>G) SNP előfordulásának esélye nagyobb a RA esetscsoportban szemben a kontroll csoporttal, így a kapott eredmények kapcsolatot mutattak az AIRE transzkripciós faktort kódoló DNS-szekvenciában található rs760426 (A/G) SNP és RA fogékonyság között (105).

#### IV.1.4. Szenzitivitás-analízis

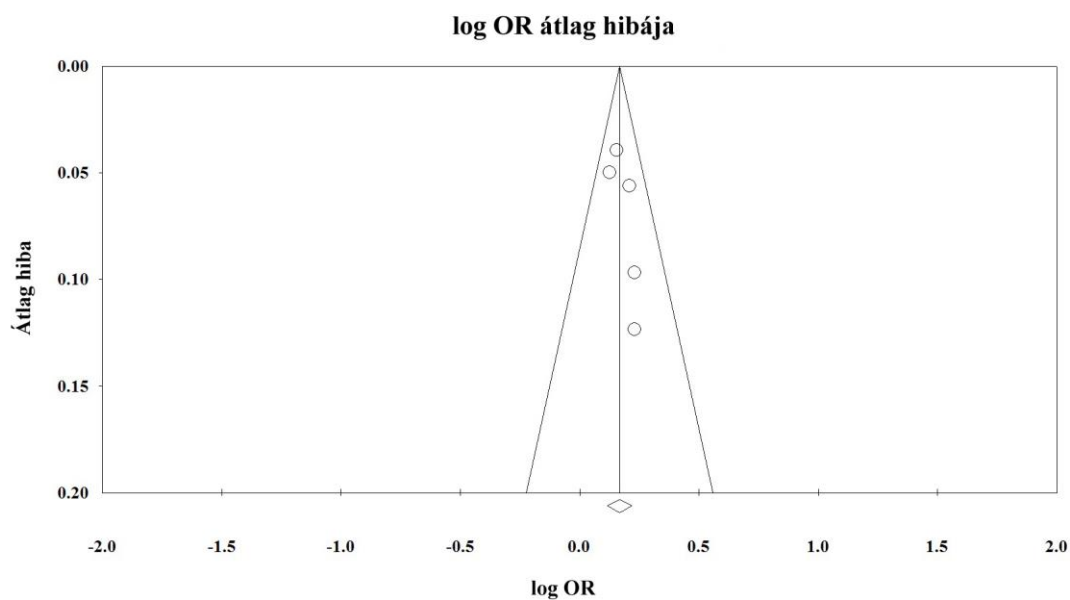
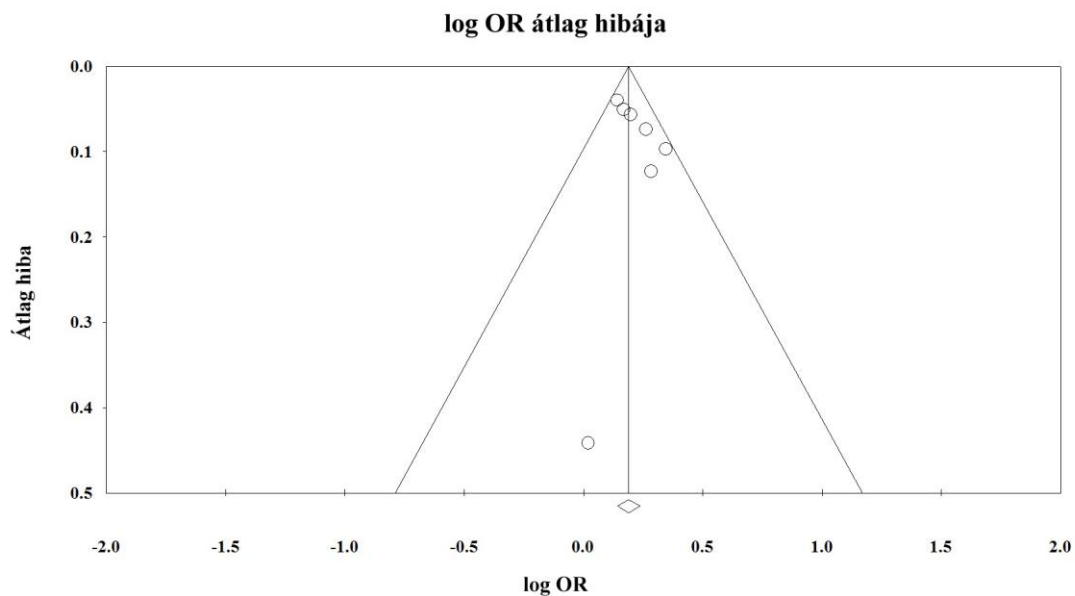
Annak érdekében, hogy meghatározzuk az egyes eset-kontroll vizsgálatok hatását a teljes meta-analízisre, szenzitivitás-analízist végeztünk egy adott vizsgálat kihagyásával, melynek eredményeképpen, az allélikus (**4. ábra**), domináns, recesszív, kodomináns hetero- és homozigóta genetikai modelleknél nem találtunk heterogenitást (105).



**4. ábra Allélikus modellek szenzitivitás analízisének eredménye rs2075876 (G/A) és rs760426 (A/G) SNP esetében**

#### IV.1.5. Publikációs torzítás analízise

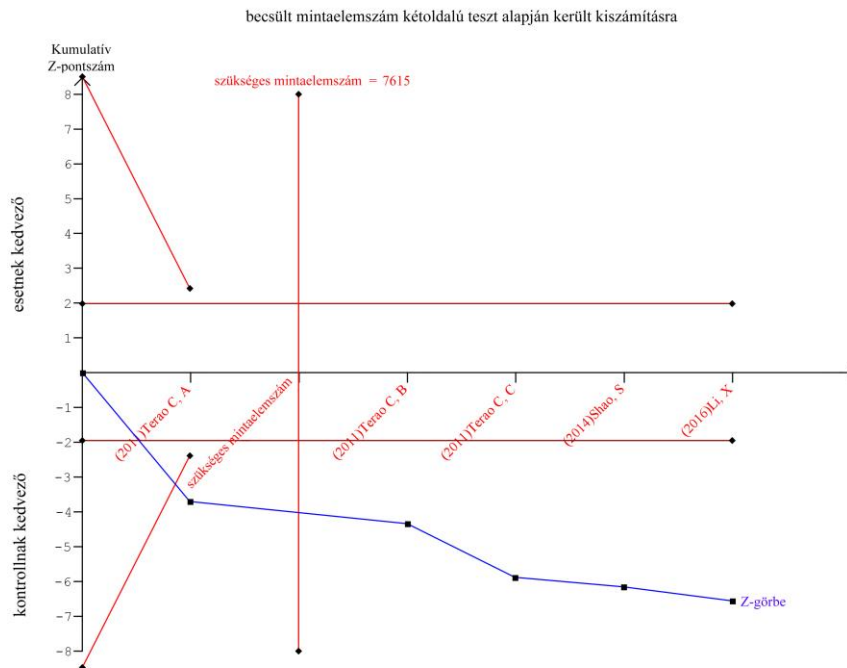
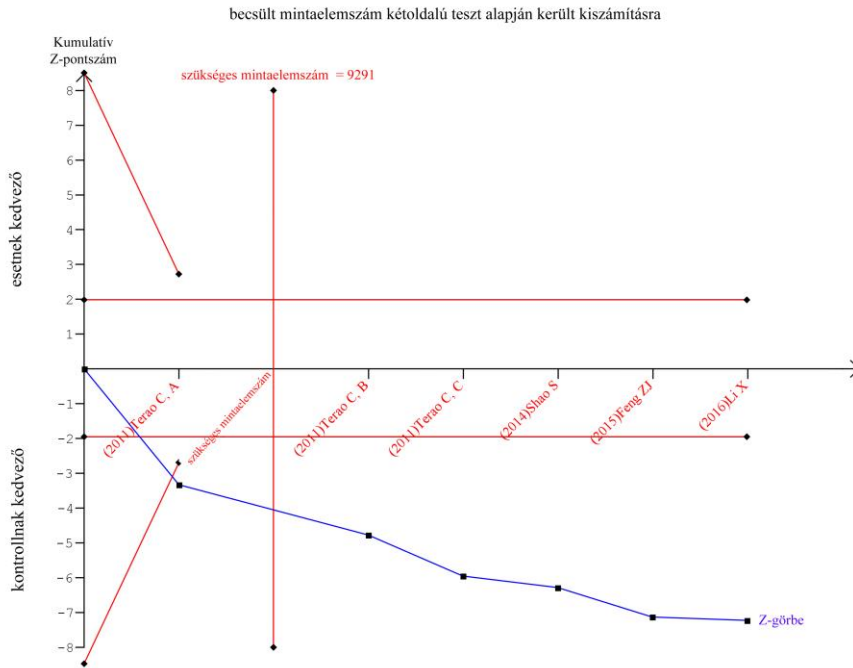
Az elemzés során funnel plotokat hoztunk létre az egyes polimorfizmusok esetében az allélikus (**5. ábra**), domináns, recesszív, kodomináns hetero- és homozigóta genetikai modellekre. Elemzésünk során szemrevételezéssel azt tapasztaltuk, hogy az összes funnel plot tökéletesen szimmetrikus, így publikációs torzítás nem volt jelen rs2075876 SNP és rs760426 SNP vizsgálatakor.



**5. ábra: Allélikus modell funnel plot analízisének eredménye rs2075876 (G/A) és rs760426 (A/G) SNP esetében (OR: odds ratio)**

#### IV.1.6. Trial sequential analízis

Az rs2075876 (G/A) és rs760426 (A/G) SNP-k allélikus modelljeire elvégzett TSA eredményeként a kumulált z-görbék (kék vonal) metszették a TSA monitoráló (ellenőrző) határvonalait, így a meta-analízis összesített elemszáma eléri a szükséges értéket (**6. ábra**).



**6. ábra: Allélikus modellek TSA analizisének eredménye rs2075876 (G>A) és rs760426 (A>G) SNP esetében**

Ebből adódóan, a levont következtetés nagy erősséggel alátámasztja az rs2075876 (G>A), rs760426 (A>G) SNP és a RA fogékonysága közötti kapcsolatot. Ezek az eredmények egyúttal azt is mutatják, hogy az összefüggés szempontjából meta-analízisünkbe további vizsgálatot bevonni szükségtelen.

#### **IV.2. *In silico* vizsgálat AIRE allélpolimorfizmusok meghatározására**

UCSC Genome Browser és az Ensembl automatikus annotációs rendszerében a keresésünket az AIRE kódoló szekvenciájára (HGNC Symbol; Acc: HGNC:360) állítottuk be, mely a 21.



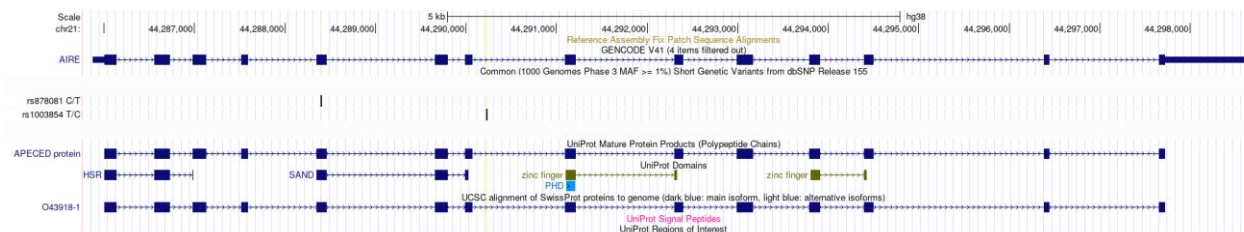
kromoszóma hosszú karján, a 21q22.3 sávnál található a 44,285,838-44,298,648 közötti régióban, 5'-3' irányban (147,148). Korlátozások nélkül 23 023 variációt azonosítottunk, melyek közé CNV, pontmutáció, indel és SNP is tartozik. A globális MAF 0,01-0,5-re történő beállítása után 68 olyan SNP-t detektáltunk, melyek az NCBI dbSNP, 1000 Genomes Project Phase III (beleértve a HapMap projektet), NHLBI Trans-Omics for Precision Medicine (human only), NHLBI GO Exome Sequencing Project (human only), Genome Aggregation Database (human only) és Exome Aggregation Consortium (human only) adatbázisok legalább egyikében szerepelnek és MAF értékük 0,01 és 0,5 között helyezkedik el (148-151). Kizárási kritériumainkat követve, annak érdekében, hogy az alacsony detektálhatósági valószínűséget elkerüljük 24 olyan SNP-t zártunk ki melyeknél a globális MAF érték 0,01-0,5 közötti de az EUR és a CEU alpopulációkban ez az érték 0,01 alatti. *In silico* vizsgálatunk ezen pontján 44 SNP-t választottunk ki a 44,285,838-44,298,648 régióból, melyek 0,01-0,5 MAF-értékkel rendelkeznek EUR és CEU alpopulációban és nagy valószínűséggel alacsonyabb elemszámú vizsgálatnál is már detektálhatóak.

### **IV.3. Pilot vizsgálat AIRE rs878081 és rs1003854 allépolimorfizmusok vizsgálatára reumatoid artritisz betegek mintáiban**

Első lépésként az AIRE kódoló régiójában található 14 exon régió közül el kellett döntenünk, mely régiót tartjuk fontosnak a pilot vizsgálat kivitelezésére. A strukturális domének közül a SAND régió sajátos, mert fehérje-fehérje kapcsolatok és az AIRE expressziójának fenntartása mellett egyik legfontosabb funkciója az mTEC sejtek DNS-éhez való kötődés, mely feltétele a klaszterba rendezett TRA-gének CBP-függő promiszkuis génexpressziójának, ami pedig szükséges a centrális tolerancia során a medullába érkező SP CD4<sup>+</sup> vagy CD8<sup>+</sup> timociták negatív szelekciójához, az autoreaktív T-sejtek klonális törléséhez (23). Mutációk az AIRE SAND régióját kódoló exon 5 és 6 valamint intron 5-ös régiójában APECED betegséghez vezetnek, mely diszfunkció a negatív szelekció defektusa (21,164,165). Tekintettel a SAND régió jelentőségére a negatív szelekcióban és az autoimmunitásban, a vizsgálódásunkat SAND régióban található allépolimorfizmusok felé fordítottuk. Pilot vizsgálatunk célja, ezen régió genotipizálásának elkezdése, valamint hogy megtudjuk, az allépolimorfizmusok diszkriminációjára választott allélspecifikus próbák és a genotipizálás során kialakított protokoll megfelelőek-e? A célkitűzés teljesítésére a következő két lókuszt választottunk ki az előzetesen *in silico* módszerrel kialakított 44 SNP-t magába foglaló adatbázisunkból. Az exon 5-ös régióban található rs878081-t és a vele magas genetikai kapcsoltságban lévő, az intron 7-es régióban lévő rs1003854-t (**7. ábra**) (152).

#### **IV.3.1. A pilot vizsgálatban résztvevő alanyok demográfiai és klinikai jellemzői**

RA-ban szenvedő eset és kontroll személyek demográfiai és klinikai adatait a **IV. táblázat** tartalmazza. RA betegek átlagéletkora  $65,38 \pm 9,84$  év, nemi eloszlásuk alapján 85%-uk nő (152,165). Annak érdekében, hogy minimalizáljuk a látens esetek valószínűségét a kontrollok között, 75 éves vagy annál idősebb résztvevőket vontunk be. A betegek és a kontrollszemélyek mind Magyarországon születtek. Az RA csoportban szignifikánsan magasabb ESR, CRP, Rf (IgG) és aCCP autoantitest szinteket mutattak ki ( $p < 0,001$ ) és 60%-os szeropozitivitást mutattak 100 kontroll személyhez képest, akiknél az említett klinikai paraméterek a normális tartományban voltak, szeronegativitása 100%-os volt (152, 166-169). A DAS28-at RA-s betegek esetében is kiszámítottuk, ahol a DAS28 átlagértéke mérsékelt betegségaktivitásra utalt (170).



**7. ábra: Az AIRE exon és intron kódoló szekvenciái (GENCODE V44), az AIRE fehérje doménjei (Uni-Prot domének), az rs878081 és rs1003854 pozíciói (UCSC Genome Browser humán (GRCh38/hg38) adatbázis lekérdezése által generált, az UCSC Genome Browser weboldaláról szabadon letölthető kép) <https://genome.ucsc.edu/index.html> (146)**

### IV.3.2. Genotípus-frekvenciák és a Hardy–Weinberg equilibrium

Kontrollcsoportokban az rs878081 és rs1003854 lókuszek alléljai által meghatározott genotípusok előfordulási gyakoriságai megfeleltek a Hardy–Weinberg egyensúlynak, ugyanis szignifikánsan nem tért el a várt és a megfigyelt genotípusfrekvencia egymástól. Az rs878081 esetében  $p=0,316$  valamint az rs1003854 esetében  $p=0,214$ . Ennek megfelelően mintagyűjtésünkkel elkerültük a szelekciós torzítást, a non-random mintavételt valamint a rokonok halmozódását.

Jellemzők	RA betegek (N=100)	Kontrollok (N=100)
Életkor (évek), átlag $\pm$ SD	65,38 $\pm$ 9,84*	81,17 $\pm$ 4,56
Nem		
Nő, no. (%)	85 (85%)	68 (68%)
Férfi, no. (%)	15 (15%)	32 (32%)
ESR (We) (mm/h), átlag $\pm$ SD (normális tartomány: 3–13 mm/h)	40,24 $\pm$ 23,90*	15,43 $\pm$ 9,97
CRP (mg/dL), mean $\pm$ SD (normális tartomány: 0–10 mg/dL)	28,80 $\pm$ 29,21*	6,19 $\pm$ 2,53
Rf (IgG) (IU/mL), átlag $\pm$ SD (normális tartomány: 0–15 IU/mL)	166,15 $\pm$ 156,39*	11,79 $\pm$ 7,61
aCCP (U/mL), mean $\pm$ SD (normális tartomány: < 20,0 U/mL; gyenge pozitivitás: $\geq$ 20,0–39,0 U/mL; mérsékelt pozitivitás: $\geq$ 39,0–59,0 U/mL; erős pozitivitás: $\geq$ 59,0 U/mL)	561,10 $\pm$ 583,07*	4,566 $\pm$ 2,65
Szeropozitivitás, no. (%)	66 (66%)*	0 (0%)
Szeronegativitás, no. (%)	14 (14%)*	100 (100%)

DAS28, átlag ± SD

(betegség remissziója: < 2,6;

alacsony: ≥ 2,6–3,1;

mérsékelt: ≥ 3,1–5,1;

magas: ≥ 5,1)

4,93 ± 1,46

-

RA: reumatoid arthritisz; SD: standard deviáció; ESR (We): eritrocita szedimentációs ráta; CRP: C-reaktív protein; Rf: reumatoid faktor; IgG: immunglobulin G; aCCP: anti-ciklikus citrullinált peptid antitest; DAS28: betegség aktivitási pontszám 28 ízületben, disease activity score with 28-joint counts; \*p<0,001, kontroll alanyok összehasonlításában.

#### IV. táblázat: A pilot vizsgálatban résztvevő alanyok demográfiai és klinikai jellemzői

### IV.3.3. Egy-lókuszos genetikai modellek betegségkockázat meghatározására

Allélikus modellünk keretében a kockázatot jelentő allélt keressük az rs878081 és rs1003854 lókuszt esetében, vagyis először megvizsgáltuk, mely allélok frekvenciája magasabb a betegpopulációban, majd bináris logisztikus regressziós vizsgálattal OR értéket számoltunk, hogy megtudjuk, a betegségkockázat mely allél irányába mutat. Ezt követően a betegségkockázatot jelentő allélra alakítottuk ki domináns, recesszív, kodomináns heterozigóta és kodomináns homozigóta és over-domináns populációgenetikai modelljeinket mind az rs878081 és rs1003854 lókuszt esetében (152).

### IV.3.4. Az AIRE rs878081 (C/T) allélpolimorfizmusok és az RA kockázata közötti kapcsolat

AIRE rs878081	RA betegek N=100 (%)	Kontrollok N=100 (%)	OR (95% CI)	p <sup>a</sup>
Allél				
T	46 (23,0)	42 (21,0)	1,12 (0,70–1,80)	0,629
C	154 (77,0)	158 (79,0)		
Genotípus				
TT	10 (10,0)	2 (2,0)	4,68 (0,98–22,27)	0,052
CT	26 (26,0)	38 (38,0)	0,64 (0,34–1,18)	0,154
CC	64 (64,0)	60 (60,0)		
Domináns genetikai modell				
CT+TT	36 (36,0)	40 (40,0)	0,84 (0,47–1,49)	0,629
CC	64 (64,0)	60 (60,0)		
Recesszív genetikai modell				
TT	10 (10,0)	2 (2,0)	5,44 (1,16–25,52)	0,032*
CC+CT	90 (90,0)	98 (98,0)		
Over-domináns genetikai modell				

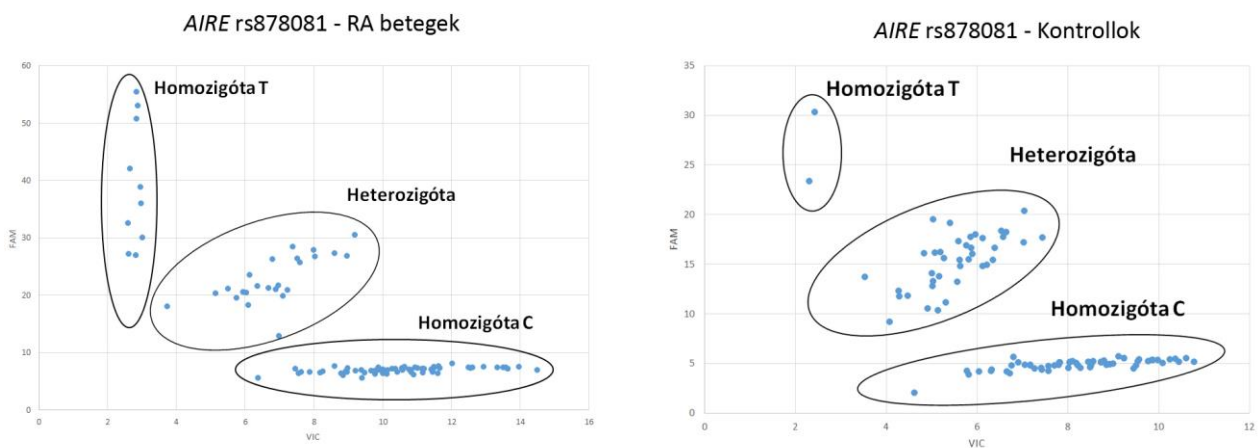
CT	26 (26,0%)	38 (38,0%)	0,57 (0,31–1,04)	0,070
CC+TT	74 (74,0%)	62 (62,0%)		

SNP: single-nucleotide polymorphism, egy pontos nukleotid polimorfizmus; OR: odds ratio, esélyhányados; CI: confidence interval, konfidencia intervallum.

p<sup>a</sup>: bináris logisztikus regresszió szignifikanciája; \* p<0,05

**V. táblázat: AIRE rs878081 lókusznak allélfrekvenciái és genotípus gyakoriságai, asszociációjuk RA-val**

A recesszív genetikai modell alapján statisztikailag szignifikáns összefüggést találtunk az RA-val, ahol a TT homozigóták gyakorisága szignifikánsan magasabb volt az RA-s betegekénél, mint a kontroll személyeknél, ami alátámasztja az RA kockázatát (OR=5,44, 95% CI=1,16-25,52, p=0,032) (152). A TT homozigóták genotípus-gyakorisága szintén magasabb volt az RA-s betegek körében a kodomináns homozigóta genetikai modell alapján, mint a kontroll személyek körében; az összefüggés azonban nem volt szignifikáns (OR=4,68, 95% CI=0,98-22,27, p=0,052), de ugyanazon irányba mutatott, mint a recesszív modell (152). Az allélikus,



domináns, kodomináns heterozigóta vagy over-domináns genetikai modellekben nem találtunk további szignifikáns összefüggést. Az rs878081 allél- és genotípus-frekvenciáit RA-s betegekénél és kontroll személyeknél, valamint a kockázati számításokat az **V. táblázat** tartalmazza. Az rs878081 alléldiszkriminációs teszteredményeit RA-s betegekénél és kontroll személyeknél a **8. ábra** szemlélteti.

**8. ábra: AIRE rs878081 lókusznak alléldiszkriminációs teszteredményei RA és kontrollcsoportban**

**IV.3.5. Az AIRE rs1003854 (T/C) allélpolimorfizmusok és az RA kockázata közötti kapcsolat**

Az rs1003854 allélpolimorfizmusait allélikus, domináns, recesszív, kodomináns heterozigóta, kodomináns homozigóta és over-domináns genetikai modellek alapján vizsgáltuk (152). A genotípus gyakorisága az RA-s betegekénél statisztikailag szignifikánsan magasabb volt a TT homozigótáknál, mint a kontroll személyeknél, ami alátámasztja az RA kockázatát (OR=4,84, 95% CI=1,02-23,02, p=0,047) (152). Az allélikus, domináns, kodomináns heterozigóta vagy over-domináns genetikai modellekben nem találtunk további szignifikáns összefüggést. Az rs1003854 allél- és genotípus-frekvenciáit RA-s betegekénél és kontroll személyeknél, valamint a kockázati számításokat a **VI. táblázat** tartalmazza. Az rs1003854 allél-

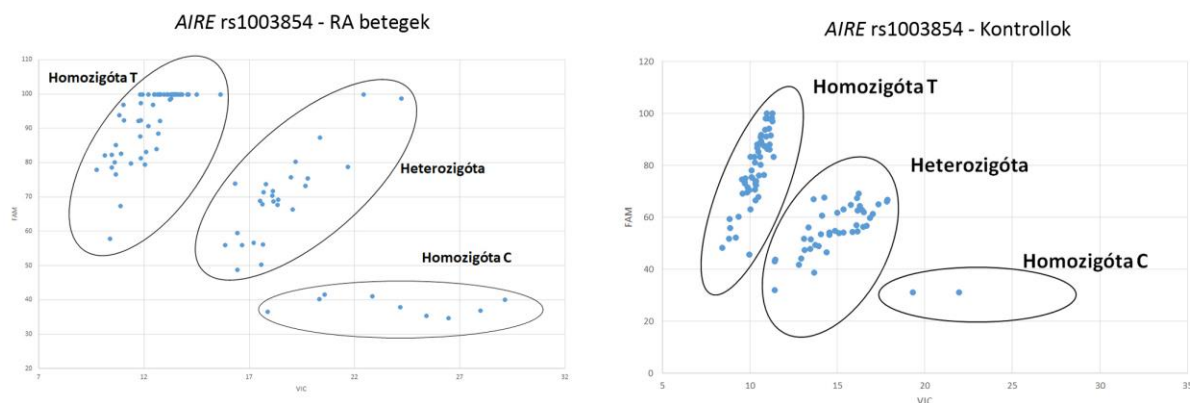
diszkriminációs tesztek eredményeit RA-s betegeknél és kontroll személyeknél a **9. ábra** szemlélteti.

<i>AIRE</i> rs1003854	RA betegek N=100 (%)	Kontrollok N=100 (%)	OR (95% CI)	p <sup>a</sup>
<b>Allél</b>				
C	47 (23,5)	45 (22,5)	1,05 (0,66–1,68)	0,812
T	153 (76,5)	155 (77,5)		
<b>Genotípus</b>				
CC	9 (9,0)	2 (2,0)	4,13 (0,85–19,96)	0,077
TC	29 (29,0)	41 (41,0)	0,65 (0,35–1,18)	0,157
TT	62 (62,0)	57 (57,0)		
<b>Domináns genetikai modell</b>				
TC+CC	38 (38,0)	43 (43,0)	0,81 (0,46–1,43)	0,471
TT	62 (62,0)	57 (57,0)		
<b>Recesszív genetikai modell</b>				
CC	9 (9,0)	2 (2,0)	4,84 (1,02–23,02)	0,047*
TT+TC	91 (91,0)	98 (98,0)		
<b>Over-domináns genetikai modell</b>				
TC	29 (29,0)	41 (41,0)	0,58 (0,32–1,05)	0,076
TT+CC	71 (71,0)	59 (59,0)		

SNP: single-nucleotide polymorphism, egy pontos nukleotid polimorfizmus; OR: odds ratio, esélyhányados; CI: confidence interval, konfidencia intervallum.

p<sup>a</sup>: bináris logisztikus regresszió szignifikanciája; \* p<0,05

**VI. táblázat: *AIRE* rs1003854 lókuszának allélfrekvenciái és genotípus gyakoriságai, asszociációjuk RA-val**



**9. ábra: *AIRE* rs1003854 lókusz alléldiszkriminációs teszteredményei RA és kontrollcsoportban**

#### **IV.3.6. A klinikai paraméterek összehasonlítása az *AIRE* rs878081 és rs1003854 genetikai modelljei alapján RA betegeknél és kontroll személyeknél**

Vizsgálatunk célja az ESR, CRP, Rf, aCCP és DAS28 szintek összehasonlítása volt az *AIRE* rs878081 és rs1003854 genetikai modellek genotípus-alcsoportjai között RA-s betegek és kontrollszemélyek esetében. Első körben a cél az volt, hogy a genotípusok alcsoportátlagainak összehasonlításával szignifikáns különbségeket találjunk, a következő lépésben pedig a klinikai paraméterek és a genotípusok közötti összefüggéseket szándékoztuk feltárni (152). RA betegeknél a CC genotípus-alcsoportban szignifikánsan magasabb volt az ESR szint a TT homozigótákhoz képest az rs1003854 kodomináns homozigóta modellje alapján ( $p=0,048$ ). A CC genotípus-alcsoport szignifikánsan magasabb CRP-szintet mutatott a TT homozigótákhoz képest a kodomináns homozigóta modellben valamint a TT+TC alcsoporthoz képest a

recesszív modellben ( $p=0,043$  és  $p=0,029$ ) (152). Továbbá, a CC homozigóta RA-s betegek esetében szignifikáns korreláció mutatkozott a CRP-vel a kodomináns homozigóta és a recesszív modellben ( $p=0,006$  és  $p=0,002$ , illetve  $r=0,350$  és  $r=0,338$ ) (152). Az rs878081 és az rs1003854 egyéb klinikai változói és genetikai modelljei között nem találtunk további szignifikáns különbségeket. Ezeket az eredményeket a **VII. táblázat** szemlélteti.

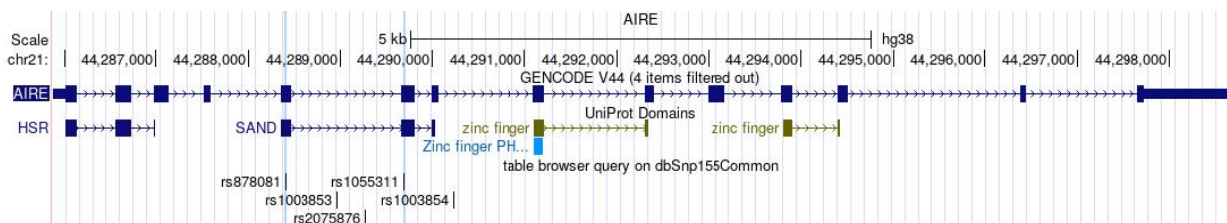
<i>AIRE</i> rs1003854	RA betegek				Kontrollok			
	CRP	$p^a$	Pearson korreláció	$p^b$	CRP	$p^c$	Pearson korreláció	$p^d$
Genotípus								
CC	58,41 ± 47,30	0,043*	0,350	0,006**	6,00 ± 0,00	0,486	0,002	0,990
TC	22,04 ± 23,03	0,094	-0,086	0,469	6,49 ± 3,14	0,486	0,099	0,337
TT	27,21 ± 25,75				5,98 ± 2,06			
Domináns genetikai modell								
TC+CC	32,47 ± 35,36	0,687	0,092	0,416	6,47 ± 3,06	0,430	0,096	0,347
TT	26,87 ± 25,55				5,98 ± 2,06			
Recesszív genetikai modell								
CC	58,41 ± 47,30	0,029*	0,338	0,002**	6,00 ± 0,00	0,571	-0,011	0,915
TT+TC	25,56 ± 24,95				6,19 ± 2,56			
Over-domináns genetikai modell								
TC	22,09 ± 23,57	0,050	-0,132	0,239	6,49 ± 3,14	0,527	0,100	0,329
TT+CC	31,01 ± 30,69				5,98 ± 2,02			

SNP: single-nucleotide polymorphism, egy pontos nukleotid polimorfizmus; RA: rheumatoid arthritis; CRP: C-reactive protein; SD: standard deviáció;  $p^a$ : az egyes genetikai modellekben az RA betegek genotípus-alcsoport szerinti CRP szintjeinek összehasonlításából eredő szignifikanciája;  $p^b$ : az egyes genetikai modellekben az RA betegek genotípus-alcsoportjainak CRP-vel való Pearson korrelációjának a szignifikanciája;  $p^c$ : az egyes genetikai modellekben az kontroll személyek genotípus-alcsoport szerinti CRP szintjeinek összehasonlításából eredő szignifikanciája;  $p^d$ : az egyes genetikai modellekben a kontroll személyek genotípus-alcsoportjainak CRP-vel való Pearson korrelációjának a szignifikanciája; \* $p<0,05$ , \*\* $p<0,01$ ; A klinikai paraméterek átlagérték ± SD-ben vannak megadva

## VII. táblázat: CRP szintek az *AIRE* rs1003854 genetikai modell genotípus-alcsoportjai között RA betegek és kontrollszemélyek csoportjában

### IV.4. Az *AIRE* exon 5 és 6 valamint intron 5 és 7-es régiókban található allélpolimorfizmusok vizsgálata reumatoid arthritisz kockázatára

A pilot vizsgálat sikerességére alapozva és felhasználva az in silico vizsgálatunk során detektált 44 SNP-t tartalmazó adatbázisunkat – a SAND-domén kódoló *AIRE* régióban maradván – olyan panelt terveztünk, mely magába foglalja mindazon potenciális lókuszokat és azok adta allélpolimorfizmusokat, melyek SAND domént kódoló régiót (exon 5 és 6) valamint a közöttük lévő splicing intronális régiót (intron 5 és 7) határozzák meg. A négy *AIRE* régióból tehát a pilotban vizsgált két lókusz mellé, még választottunk három olyan lókuszt, melyek nagymértékben hozzájárulnak a SAND domén expressziójához. A vizsgálatához kiválasztott lókuszok 5'-3' irányban így az rs878081 az exon 5-ös, rs1003853 és rs2075876 az intron 5-ös, rs1055311 az exon 6-os és rs1003854 az intron 7-es régió. A két exon és három intronális splice variánsnak a pozícióját a **10. ábra** mutatja.



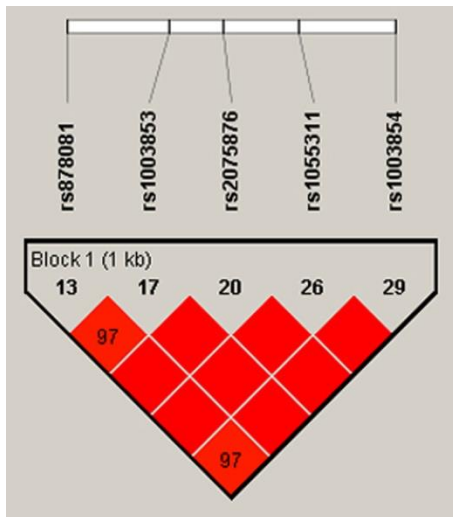
**10. ábra:** Az *AIRE* exon és intron kódoló szekvenciái (GENCODE V44), az *AIRE* fehérje doménjei (Uni-Prot domének), valamint az rs878081, rs1003853, rs2075876, rs1055311 és rs1003854 pozíciói (UCSC Genome Browser humán (GRCh38/hg38) adatbázis lekérdezése által generált, az UCSC Genome Browser weboldaláról szabadon letölthető kép) <https://genome.ucsc.edu/index.html> (146)

Annak érdekében, hogy az öt lókusz kapcsoltságát megvizsgáljuk, HaploView 4.2-es szoftver segítségével haplotípus analízist végeztünk melyből kiderült, hogy az öt lókusz egy haplotípus blokkon helyezkedik el és közöttük magas kapcsoltság van, melynek LD D'-értéke 0,977 és 1,000 közötti (157,172,173). A lókuszok elhelyezkedését a haplotípus blokkokban a **11. és 12. ábra** illusztrálja.





11. ábra: Az öt lókuszt az LD-térképen és Lewontin „D” értékei. A piros szín magas kapcsoltságot jelez, a számok nélküli piros blokkok a maximális kapcsoltságot jelzik ( $D'=1$ ), kék blokkok nagyon alacsony vagy a kapcsoltság hiányát jelzik ( $D'=0-0,3$ ) (Species: Human; Assembly: GRCh38.p14; Region Lookup: 21:44285838-44298648). LD, linkage disequilibrium (171,172).



12. ábra: Az öt lókuszt és Lewontin „D” értékei összesített LD-térképen. A piros szín magas kapcsoltságot jelez, a számok nélküli piros blokkok a maximális kapcsoltságot jelzik ( $D'=1$ ) (Species: Human; Assembly: GRCh38.p14; Region Lookup: 21:44285838-44298648). LD, linkage disequilibrium (171,172).

#### IV.4.1. A vizsgálatban résztvevő alanyok demográfiai és klinikai jellemzői

A teljes populációnk 592 személyt tartalmaz, 270 RA-s beteget és 322 egészséges kontrollszemélyt. Közülük 100-100 résztvevőt pilot study keretében vizsgáltunk, melyhez rs878081 és rs1003854 lókuszt választottuk és elemeztük allélpolimorfizmusukat, annak

érdekében, hogy lássuk a MAF tendenciáit, allélok detektálhatóságát és teszteljük a genotipizálási munkafolyamatunk eredményességét (152).

Ebben a vizsgálatunkban két csoportot alakítottunk ki, melynek megfelelően az 1. csoport a teljes populációt jelenti 270 RA és 322 kontrollal, melyekhez választottuk a három további AIRE rs1003853, rs2075876 és rs1055311 lókuszt. A 2. csoport a teljes populációnak egy alcsoportja 170 RA és 222 kontrollal, melyekhez az AIRE rs878081 és rs1003854-es lókuszt választottuk. Értelem szerint esetükben azért nem számolhattunk erre a két lókuszt a teljes populációval, mert abból már 100-100-at a rs878081 és rs1003854 lókusztokra már vizsgáltunk és publikáltunk (152). Ennek megfelelően a rs878081 és rs1003854 lókusztokra a teljes populációval nem számolhattunk (152).

Az egyes vizsgálati csoportok jellemzői az **VIII. táblázatban** láthatók. Az RA-betegek átlagéletkora  $65,47 \pm 9,85$  év és  $65,53 \pm 9,89$  év volt, az 1. és 2. csoportban 84,0% és 85,5% nő volt, megoszlásuk szignifikánsan magasabb volt, mint a férfiaké az RA és a kontroll csoportokban (157,165). A teljes populációban részt vevők mind Magyarországon születtek. A szérum ESR, CRP, Rfs és aCCP szintek  $p < 0,001$  értékkel szignifikánsan magasabbak voltak mindkét RA csoportban a kontrollhoz képest (157).

Az 1. csoport RA betegeinek 60,7%-ában sikerült szeropozitivást, 8,5%-ában szeronegativitást kimutatni, míg 30,8%-ukban nem készült szerológia (157). Az 1. csoport kontroll alanyainak 90%-ában sikerült szeronegativitást, míg 10%-uknál nem készült szerológia. A 2. csoport RA betegeinek 71,1%-ában sikerült szeropozitivást, 19,4%-ában szeronegativitást kimutatni, míg 9,5%-ukban nem készült szerológia. Az 2. csoport kontroll alanyainak 84,2%-ában sikerült szeronegativitást, míg 15,8%-uknál nem készült szerológia. Az aCCP tekintetében az 1. és 2. csoport RA betegek 590,69  $\pm$  947,96 és 596,27  $\pm$  1003,71 U/mL értékkel erősen pozitívnak bizonyultak, valamint 5,1 feletti, magas betegségaktivitást jelentő DAS28 pontszámot demonstráltak az 1. és 2. csoportban 5,35  $\pm$  1,20 és 5,44  $\pm$  1,12 értékekkel (157). A demográfiai és klinikai adatokat a **VIII. táblázat** tartalmazza.

Jellemzők	1. csoport		2. csoport	
	RA betegek (n=270)	Kontrollok (n=322)	RA betegek (n=170)	Kontrollok (n=222)
Életkor (évek), átlag $\pm$ SD	65.47 $\pm$ 9.85 <sup>a</sup>	78.66 $\pm$ 5.86	65.53 $\pm$ 9.89 <sup>d</sup>	77.53 $\pm$ 6.03
Sex				
Female, no. (%)	227 (84.0) <sup>b</sup>	252 (78.2) <sup>c</sup>	142 (83.5) <sup>e</sup>	184 (82.8) <sup>f</sup>
Male, no. (%)	43 (15.9)	70 (21.7)	28 (16.4)	38 (17.1)
ESR (We) (mm/h), átlag $\pm$ SD (normális tartomány: 3–13 mm/h)	39.52 $\pm$ 25.21 <sup>a</sup>	17.87 $\pm$ 12.81	39.09 $\pm$ 26.04 <sup>d</sup>	19.02 $\pm$ 13.82
CRP (mg/dL), mean $\pm$ SD (normális tartomány: 0–10 mg/dL)	29.79 $\pm$ 29.88 <sup>a</sup>	7.10 $\pm$ 6.80	30.36 $\pm$ 30.34 <sup>d</sup>	7.53 $\pm$ 8.03
Rf (IgG) (IU/mL), mean $\pm$ SD (normális tartomány: 0–15 IU/mL)	118.80 $\pm$ 126.63 <sup>a</sup>	13.49 $\pm$ 5.55	104.71 $\pm$ 113.31 <sup>d</sup>	14.41 $\pm$ 13.75
aCCP (U/mL), mean $\pm$ SD (normális tartomány: < 20.0 U/mL; gyenge pozitívitas: $\geq$ 20.0–39.0 U/mL; mérsékelt pozitívitas: $\geq$ 39.0–59.0 U/mL; erős pozitívitas: $\geq$ 59.0 U/mL)	590.69 $\pm$ 947.96 <sup>a</sup>	4.56 $\pm$ 2.29	596.27 $\pm$ 1003.71 <sup>d</sup>	4.55 $\pm$ 1.66
Szeropozitivitas, no. (%)	164 (60.7)		121 (71.1)	
Szeronegativitas, no. (%)	23 (8.5) <sup>a</sup>	290 (90.0)	33 (19.4) <sup>d</sup>	187 (84.2)
Szerológia nélkül (%)	83 (30.8)	32 (10.0)	16 (9.5)	35 (15.8)

DAS28, átlag ± SD (betegség remissziója: < 2.6; alacsony: ≥ 2.6–3.1; mérsékelt: ≥ 3.1–5.1; magas: ≥ 5.1)	5.35 ± 1.20	-	5.44 ± 1.12	-
--	-------------	---	-------------	---

1. csoport: teljes populáció; 2. csoport: a teljes populáció alsó csoportja; SD: standard deviáció; ESR (We): eritrocita szedimentációs ráta; CRP: C-reaktív protein; Rf: reumatoid faktor; IgG: immunglobulin G; aCCP: anti-ciklikus citrullinált peptid antitest; DAS28: betegség aktivitási pontszám 28 ízületben, disease activity score with 28-joint counts; <sup>a</sup>: p<0.001 az 1. csoport kontroll személyeivel való összehasonlítás szignifikanciája; <sup>b</sup>: p<0.001 az 1. csoport RA férfijaival való összehasonlítás szignifikanciája; <sup>c</sup>: p<0.001 az 1. csoport kontroll férfijaival való összehasonlítás szignifikanciája; <sup>d</sup>: p<0.001 a 2. csoport kontroll személyeivel való összehasonlítás szignifikanciája; <sup>e</sup>: p<0.001 a 2. csoport RA férfijaival való összehasonlítás szignifikanciája; <sup>f</sup>: p<0.01 a 2. csoport kontroll férfijaival való összehasonlítás szignifikanciája.

### VIII. táblázat: Az 1. és 2. csoport demográfiai és klinikai adatai

#### IV.4.2. Genotípus-frekvenciák és a Hardy–Weinberg equilibrium

Az 1. csoport RA betegpopuláció várt és megfigyelt genotípus-frekvenciái közötti differencia szignifikanciája az rs2075876, rs1055311 és rs1003853 esetében p=0,916, p=0,985 és p=0,715 volt. Az 1. csoport kontrolljai ezek esetében p=0,627, p=0,899 és p=0,503 értéket adtak. A 2. csoport RA betegeknek a várt és a megfigyelt genotípus-frekvenciái között a különbség statisztikai szignifikanciája az rs878081 és az rs1003854 SNP-k esetében p=0,991, és p=0,995 (157). A 2. csoport kontrolljai esetében p=0,870 és p=0,658 értéket mutattak. Nem-paraméteres  $\chi^2$  próbák mind az 1. és 2. csoportban azt mutatták, hogy a megfigyelt frekvenciák jól illeszkednek a HWE egyensúly alapján várható genotípus-frekvenciákhoz. Ezért kimondható, vizsgálati populációnk mentes a szelekciós torzítástól, a non-random mintavételtől és a rokonok felhalmozódásától (157).

#### IV.4.3. Exon 5 és intron 5 és 7-es kódoló régiókban található allélpolimorfizmusok asszociációt mutatnak az RA kockázatával

Jelen esetben is először az allélikus genetikai modellt használtuk annak érdekében, hogy megtaláljuk mely allélok előfordulási gyakorisága magasabb a betegpopulációban, majd bináris logisztikus regressziós vizsgálattal OR értéket számoltunk, hogy megtudjuk, a betegségkockázat melyik irányába mutat. Ezt követően a betegségkockázatot jelentő allélra alakítottuk ki domináns, recesszív, kodomináns heterozigóta és kodomináns homozigóta és over-domináns populációgenetikai modelljeinket. Ezt a metodikát elvégeztük mind az öt lókuszt, az rs878081, rs1003853, rs2075876, rs1055311 és rs1003854 esetében, az 1. és 2. csoport alkotta teljes, 592 fős vizsgálati populáció esetében, melyhez 1184 allél diszkriminációs teszteredményeit használtuk fel (157).

Az öt lókuszt közül először a 2. csoportban az *AIRE* SAND exon 5-ös SAND domént kódoló rs878081 C-alléja mutatott szignifikáns összefüggést az RA kockázatával, az allélikus modell szerint (OR=1,48, 95% CI 1,05-2,09, p=0,023). Itt a recesszív genetikai modell alapján a CC homozigóta RA-betegek genotípus-frekvenciája szignifikánsan magasabb volt a CT heterozigóták és a TT homozigóták genotípus-frekvenciáinak összegéhez képest. Recesszív genetikai modell alapján kijelenthető, hogy a CC genotípus és RA között az összefüggés statisztikailag szignifikáns (OR=1,64, 95% CI 1,09-2,48, p=0,01). Az over-domináns genetikai modellben a CT genotípus gyakorisága statisztikailag szignifikánsan alacsonyabb, mint a TT és a CC genotípusok frekvenciájának az összege. OR-érték és szignifikancia alapján elmondható, hogy a CT genotípus inverz asszociációt mutat a betegséggel (OR=0,64,

95% CI 0,42-0,97,  $p=0,037$ ) (157). Az 2. csoport rs878081 lókusza esetében a VIC és FAM csatornák fluoreszcenz jelintenzitását, a jelek pozícióját valamint az allél-diszkriminációs teszteredményeket RA és kontrollcsoportban a **13. ábra** mutatja.

Vizsgálatunk során a továbbiakban az 1. csoportot jelentő teljes, 592 fős populáció 1184 allélját vizsgálva határoztunk meg szignifikáns összefüggést. Az *AIRE* intron 5-ös régiójában található rs1003853 splice variáns C-alléljának frekvenciája szignifikánsan magasabb volt RA betegek körében, mint a kontroll személyeknél, bináris logisztikus regresszió eredménye alapján elmondható, hogy az *AIRE* rs1003853 C-alléja és az RA betegség között statisztikailag szignifikáns összefüggés van ( $OR=1,33$ , 95% CI 1,01-1,74,  $p=0,037$ ) (157). Az 592 fős populációban a rs1003853 CC homozigóta genotípusának frekvenciája szintén statisztikailag szignifikánsan magasabb az RA betegek körében, mint kontrollokban, melynek eredményeképpen a recesszív genetikai modellben statisztikailag szignifikáns asszociációt tapasztaltunk a CC homozigóta genotípus és az RA között ( $OR=1,618$ , 95% CI 1,16-2,25,  $p=0,004$ ) (157). Az over-domináns modell alapján a CT heterozigóta genotípus statisztikailag szignifikánsan, de kisebb mértékben volt jelen a beteg populációban, ennek megfelelően bináris logisztikus regressziós analízis alapján az összefüggés szignifikáns és inverz ( $OR=0,59$ , 95% CI 0,42-0,82,  $p=0,002$ ) (157). Az 1. csoport rs1003853 lókuszára vonatkozó VIC és FAM csatornák jelintenzitása valamint az alléldiszkriminációs tesztek eredményét a **14. ábra** foglalja össze.

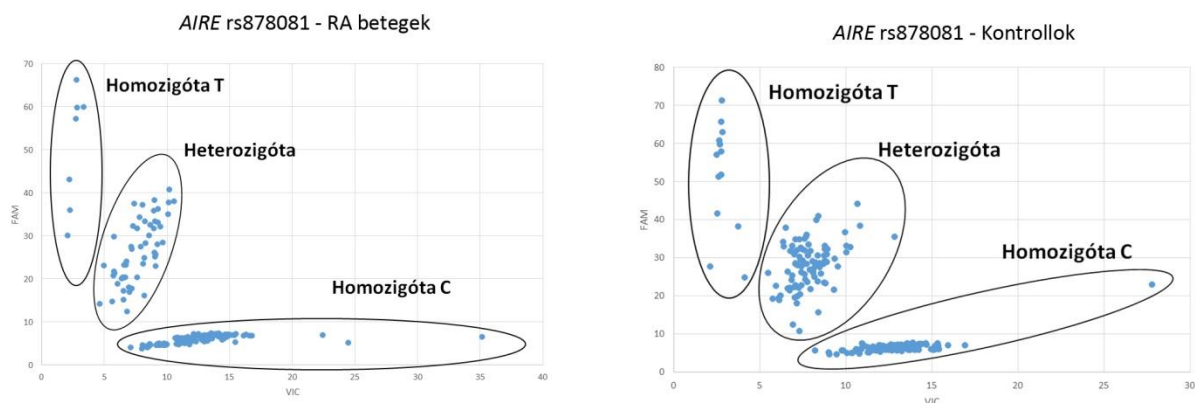
A 2. csoportban, a továbbiakban az *AIRE* intron 7-es rs1003854 lókusza volt a következő, ahol a T-allél előfordulásának az esélye szignifikánsan magasabb volt az RA allélok között a C-allélhoz képest ( $OR=1,52$ , 95% CI 1,08-2,12,  $p=0,014$ ) (157). A recesszív genetikai modell alapján a TT homozigóta genotípus esélye szignifikánsan magasabb volt RA betegek körében, a kontrollokhoz képest ( $OR=1,72$ , 95% CI 1,15-1,44,  $p=0,008$ ). Az over-domináns genetikai modellben a TC heterozigóta genotípus esélye RA betegek körében szignifikánsan alacsonyabb, a kontrollokhoz képest ( $OR=0,63$ , 95% CI 0,42-0,96,  $p=0,034$ ), mely genotípus a betegséggel inverz asszociációt mutat (157). Az erre vonatkozó alléldiszkriminációs teszteredményeket a **15. ábra** mutatja. A fent ismertett három lókusza adta genetikai eredményeket a **IX. táblázat** foglalja össze.

<i>AIRE</i> rs878081 (Exon 5)	RA betegek N=170 (%)	Kontrollok N=222 (%)	OR (95% CI)	$p^a$
Allél				
T	66 (19,4)	117 (26,4)		
C	274 (80,6)	327 (73,6)	1,48 (1,05-2,09)	0,023*
Genotípus				
TT	7 (4,1)	13 (5,9)		
CT	53 (31,2)	92 (41,4)	1,07 (0,40-2,84)	0,892
CC	110 (64,7)	117 (52,7)	1,74 (0,67-4,53)	0,253
Domináns genetikai modell				
TT	7 (4,1)	13 (5,9)		
CC+CT	163 (95,9)	209 (94,1)	1,44 (0,56-3,71)	0,441
Recesszív genetikai modell				
CT+TT	60 (35,3)	105 (47,3)		
CC	110 (64,7)	117 (52,7)	1,64 (1,09-2,48)	0,017*
Over-domináns genetikai modell				
CC+TT	117 (68,8)	130 (58,6)		
CT	53 (31,2)	92 (41,4)	0,64 (0,42-0,97)	0,037*
<i>AIRE</i> rs1003853 (Intron 5)	RA betegek N=270 (%)	Kontrollok N=322 (%)	OR (95% CI)	$p^a$
Allél				
T	116 (21,5)	172 (26,7)		
C	424 (78,5)	472 (73,3)	1,33 (1,01-1,74)	0,037*

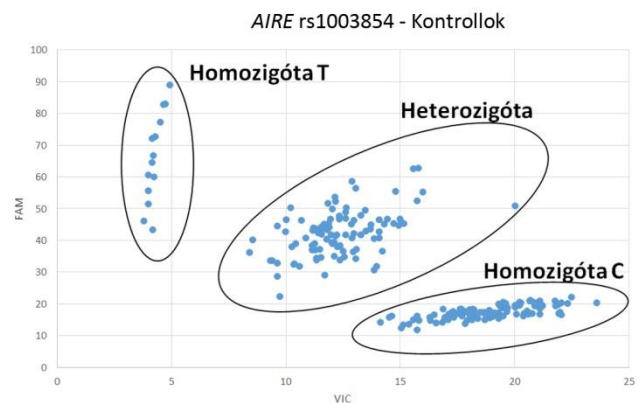
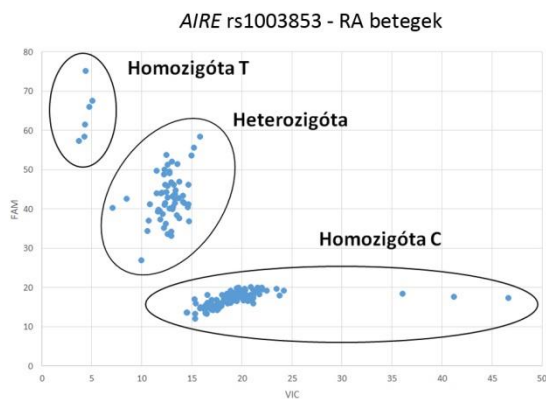
Genotípus				
TT	15 (5,6)	16 (5,0)		
CT	85 (31,5)	141 (43,8)	0,64 (0,30-1,36)	0,251
CC	170 (63,0)	165 (51,2)	1,09 (0,52-2,29)	0,802
Domináns genetikai modell				
TT	15 (5,6)	16 (5,0)		
CC+CT	255 (94,4)	306 (95,0)	0,88 (0,43-1,83)	0,750
Recesszív genetikai modell				
CT+TT	100 (37,0)	157 (48,8)		
CC	170 (63,0)	165 (51,2)	1,618 (1,16-2,25)	0,004*
Over-domináns genetikai modell				
CC+TT	185 (68,5)	181 (56,2)		
CT	85 (31,5)	141 (43,8)	0,59 (0,42-0,82)	0,002*
<i>AIRE</i> rs1003854 (Intron 7)	RA betegek N=170 (%)	Kontrollok N=222 (%)	OR (95% CI)	p <sup>a</sup>
Allél				
C	69 (20,3)	124 (27,9)		
T	271 (79,7)	320 (72,1)	1,52 (1,08-2,12)	0,014*
Genotípus				
CC	6 (3,5)	13 (5,9)		
TC	57 (33,5)	99 (44,6)	1,34 (0,48-3,69)	0,567
TT	106 (62,4)	109 (41,1)	2,10 (0,77-5,74)	0,146
Domináns genetikai modell				
CC	6 (3,5)	14 (6,3)		
TC+TT	164 (96,5)	208 (93,7)	1,84 (0,69-4,89)	0,222
Recesszív genetikai modell				
TC+CC	63 (37,1)	112 (50,5)		
TT	107 (62,9)	110 (49,5)	1,72 (1,15-1,44)	0,008*
Over-domináns genetikai modell				
CC+TT	113 (66,5)	124 (55,9)		
TC	57 (33,5)	98 (44,1)	0,63 (0,42-0,96)	0,034*

CI: confidence interval, konfidencia intervallum; OR: odds ratio, esélyhányados; RA: reumatoid artritisz. p<sup>a</sup>: bináris logisztikus regresszió szignifikanciája; \*p<0,05.

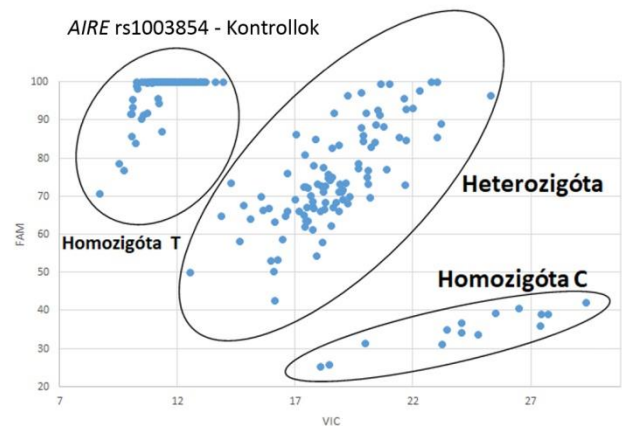
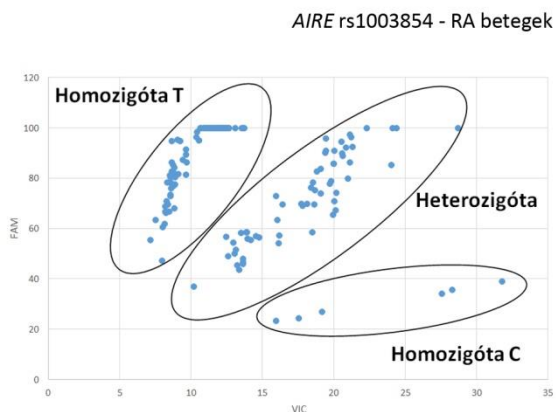
**IX. táblázat: *AIRE* rs878081, rs1003853 és rs1003854 lókuszaik allélfrekvenciái és genotípus gyakoriságai, asszociációjuk RA-val**



**13. ábra: 2. csoport *AIRE* rs878081 lókusza alléldiszkriminációs teszt eredményei RA és kontrollcsoportban**



**14. ábra: 1. csoport AIRE rs1003853 lókuszs alléldiszkriminációs teszteredményei RA és kontrollcsoportban**



**15. ábra: 2. csoport AIRE rs1003854 lókuszs alléldiszkriminációs teszteredményei RA és kontrollcsoportban**

A 2. csoportban további AIRE rs2075876 és rs1055311 lókuszsok vizsgálata során egyik csoportban sem találtunk szignifikáns összefüggést RA betegségkockázattal a 6 genetikai modell alapján, mely eredményeket az I. és II. melléklet táblázat, alléldiszkriminációs eredményeiket pedig a 1. és 2. melléklet ábra tartalmazza.

#### IV.4.4. Genetikai modellek genotípusos-alsocsoportjainak korrelációja és asszociációja klinikai paraméterekkel RA és kontrollcsoportban

Vizsgálatunk egyik célkitűzése az volt, hogy statisztikailag elemezzük az ESR, CRP, Rf, aCCP és DAS28 klinikai paramétereket RA és kontrollcsoportokhoz rendelt kodomináns homozigóta, kodomináns heterozigóta, domináns, recesszív és over-domináns genetikai modellekben, annak érdekében, hogy lássuk van-e olyan genotípus vagy genetikai modell adta genotípuscsoport, mely statisztikailag szignifikánsan összefügg egy klinikai paraméter adta folytonos változóval (157).

A klinikai jellemzők közül az ESR jól jellemezte a RA betegségaktivitását a DAS28-zal való szignifikáns pozitív korrelációjával ( $p=0,013$ ,  $r=0,258$ ) a teljes RA-populációban (157). Hasonló összefüggést találtunk CRP érték és DAS28 között ( $p<0,001$ ,  $r=0,372$ ). Ennek hatására úgy döntöttünk, hogy az RA-populációt kodomináns homozigóta és heterozigóta,

domináns, recesszív és overdomináns genetikai modellek alapján stratifikáljuk, hogy lássuk, vannak-e további korrelációk vagy asszociációk a genetikai modellek genotípus-alcsoportjai és a klinikai paraméterek között (157). Az RA betegek körében statisztikailag szignifikáns különbségeket találtunk a 2. csoport rs878081 recesszív modelljében, ahol a CC homozigóták szignifikánsan magasabb átlag ESR szinteket mutattak ( $p=0,027$ ) a TT+CT genotípusú alpopulációhoz képest. Pearson-féle kétváltozós korrelációval statisztikailag szignifikáns összefüggést találtunk az ESR és a CC homozigóta genotípus között ( $p=0,023$ ,  $r=0,190$ ), bináris logisztikus regresszióval szintén statisztikailag szignifikáns összefüggést találtunk ( $OR=1,01$ , 95% CI 1,002-1,032,  $p=0,026$ ) a két változó között (157). Az 1. csoport 592 alanyával 270 RA-betegnél szignifikánsan magasabb aCCP-szintet mutatott ki az rs1055311 lókuszt kodomináns heterozigóta modell CT és domináns modell CC+CT genotípus-csoportja ( $p=0,028$ ,  $r=0,250$  illetve  $p=0,044$ ,  $r=0,150$ ) (157). Továbbá a CT és CC genotípusok korrelációs elemzése szignifikánsan pozitív korrelációt mutatott ki a heterozigóták szignifikánsan magasabb aCCP-szintje között ( $p=0,036$ ,  $r=0,255$ ), ami a genotípus által meghatározott betegségaktivitásra utal (157). Az alcsoport-elemzések eredményeit a **X. és XI. táblázat** mutatja be.

AIRE rs1055311 (Exon 6)	RA betegek				Kontrollok			
	aCCP	$p^a$	Pearson correlation	$p^b$	aCCP	$p^c$	Pearson correlation	$p^d$
Genotípus								
TT	213,23 ± 258,53				4,44 ± 2,43			
CT	725,33 ± 998,49	0,028*	0,25	0,036*	4,43 ± 2,32	0,950	-0,001	0,997
CC	591,51 ± 999,63	0,093	0,16	0,116	4,83 ± 2,26	0,779	0,06	0,679
Domináns genetikai modell								
TT	213,23 ± 258,53				4,44 ± 2,43			
CC+CT	644,19 ± 997,37	0,044*	0,15	0,071	4,62 ± 2,29	0,918	0,02	0,843
Recesszív genetikai modell								
TT+CT	589,77 ± 893,25				4,43 ± 2,31			
CC	591,51 ± 999,63	0,982	0,001	0,991	4,83 ± 2,26	0,187	0,085	0,403
Over-domináns genetikai modell								
TT+CC	519,84 ± 917,72				4,77 ± 2,26			
CT	725,33 ± 998,49	0,157	0,103	0,216	4,43 ± 2,32	0,209	-0,07	0,467

aCCP: anti-ciklikus citrullinált peptid antitest; RA: reumatoid arthritisz; SD: standard deviáció;  $p^a$ : aCCP szintek közti különbség szignifikanciája RA csoportban;  $p^b$ : aCCP és genotípus-alcsoportok Pearson korrelációjának szignifikanciája RA csoportban;  $p^c$ : aCCP szintek közti különbség szignifikanciája kontroll csoportban;  $p^d$ : aCCP és genotípus-alcsoportok Pearson korrelációjának szignifikanciája kontroll csoportban. A klinikai paraméterek átlagérték ± SD-ben vannak megadva; \*  $p<0,05$

**X. táblázat: 1. csoport aCCP-szintjeinek korrelációja genotípus-alcsoportokkal az AIRE rs1055311 öt különböző genetikai modelljében**

AIRE rs878081 (Exon 5)	RA betegek						Kontrollok					
	ESR	p <sup>a</sup>	Pearson korreláció	p <sup>b</sup>	OR (95% CI)	p <sup>c</sup>	ESR	p <sup>d</sup>	Pearson korreláció	p <sup>e</sup>	OR (95% CI)	p <sup>f</sup>
Genotípus												
TT	24,25 ± 13,50						16,69 ± 13,11					
CT	33,16 ± 23,37	0,599	0,10	0,458	1,02 (0,96-1,08)	0,454	21,73 ± 14,96	0,147	0,11	0,255	1,02 (0,97-1,08)	0,259
CC	42,77 ± 27,11	0,196	0,13	0,179	1,03 (0,98-1,09)	0,196	17,33 ± 12,79	0,708	0,01	0,86	1,004 (0,95-1,05)	0,863
Domináns genetikai modell												
TT	24,25 ± 13,50						16,69 ± 13,11					
CC+CT	39,52 ± 26,22	0,295	0,09	0,24	1,03 (0,97-1,09)	0,262	19,18 ± 13,88	0,391	0,04	0,530	1,01 (0,96-1,06)	0,530
Recesszív genetikai modell												
TT+CT	32,46 ± 22,79						21,04 ± 14,76					
CC	42,77 ± 27,11	0,027*	0,19	0,023*	1,01 (1,002-1,03)	0,026*	17,33 ± 12,79	0,063	-0,13	0,054	0,98 (0,96-1,001)	0,980
Over-domináns genetikai modell												
TT+CC	42,00 ± 26,90						17,26 ± 12,77					
CT	33,16 ± 23,37	0,059	-0,16	0,056	0,98 (0,97-1,001)	0,059	21,73 ± 14,96	0,020*	0,15	0,023*	1,02 (1,003-1,04)	0,025*

ESR: eritrocita szedimentációs ráta; OR: odds ratio, esélyhányados; RA: reumatoid arthritisz; SD: standard deviáció; p<sup>a</sup>: ESR szintek közti különbség szignifikanciája RA csoportban; p<sup>b</sup>: ESR és genotípus-alcsoportok Pearson korrelációjának szignifikanciája RA csoportban; p<sup>c</sup>: ESR és genotípus-alcsoportok bináris logisztikus regressziójának szignifikanciája RA csoportban; p<sup>d</sup>: ESR szintek közti különbség szignifikanciája kontroll csoportban; p<sup>e</sup>: ESR és genotípus-alcsoportok Pearson korrelációjának



szignifikanciája kontroll csoportban;  $p^f$ : ESR és genotípus-alcsoportok bináris logisztikus regressziójának szignifikanciája kontroll csoportban;  $^*$ :  $p < 0.05$ ; A klinikai paraméterek átlagérték  $\pm$  SD-ben vannak megadva;  $^*$   $p < 0,05$

## **XI. táblázat: 2. csoport ESR-szintjeinek korrelációja genotípus-alcsoportokkal az AIRE rs878081 öt különböző genetikai modelljében**

### **IV.4.5. Regulatórikus kötőmotívumok allélspecifikus affinitása transzkripció faktorokhoz**

A SAND doménben az 5-ös exonon rs878081 és a 6-os exonon található rs1055311 allélpolimorfizmusok szinonim polimorfizmusok, melyek a fehérjében strukturális alapú funkcióvesztést nem, de allélspecifikus funkcionális különbségeket képesek okozni. A lehetséges transzkripció következmények tisztázása érdekében allélspecifikusan, a HaploReg v4.2 adatbázis segítségével *in silico* analízist végeztünk annak érdekében, hogy meglássuk, az egyes allélpolimorfizmusok helyet foglalnak-e bizonyos TF-k DNS kötőregiójának motívumaiban, és ha igen akkor hatással vannak-e az egyes allélok a TF kötődési affinitására (Regulatory motifs altered) (157).

Azt a négy lókuszt vizsgáltuk, melyeknek allélpolimorfizmusai összefüggésbe hozhatóak RA kockázatával illetve RA csoporton belül klinikai paraméterekkel úgy, mint ESR vagy aCCP (157). A helysúlyozott mátrix predikciós modellje az rs878081 esetében kimutatta, hogy maga a lókuszt pozíciója NF- $\kappa$ B és a Yin-Yang 1 (YY1) kötőmotívumainak szekvenciájában helyezkedik el. Az NF- $\kappa$ B NF-kappaB\_disc2 DNS kötőmotívumához számított LOD-érték +11,1 értékkel magasabb a T-allél esetében, a NF-kappaB\_known4 esetében pedig ez az érték +2, mely eredmények azt mutatják, hogy a referencia C-allél jelenlétében az NF- $\kappa$ B gyengébben kötődik a DNS-hez (157). Ugyanezen a lókuszon a YY1\_disc1 kötőmotívumhoz az alternatív T-allél esetében az affinitást jelentő LOD-érték + 0,9 értékkel magasabb, ami szintén azt jelenti, hogy a referencia C-allél jelenlétében ez a TF gyengébben kötődik (157).

A továbbiakban a helysúlyozott mátrix predikciós modellje az rs1003853 kimutatta, hogy maga a lókuszt pozíciója a Krüppel-like zinc finger protein AP-2rep (AP-2rep) és a Rad21 transzkripció faktorok kötőmotívumainak szekvenciájában helyezkedik el. Az AP-2rep DNS kötőmotívumához számított LOD-érték +7,1 értékkel, a Rad21\_disc5 kötőmotívum esetében +0,5 értékkel magasabb a T-allél esetében, vagyis C referencia allél jelenlétében ez a két TF gyengébben kötődik.

A helysúlyozott mátrix predikciós modellje szerint az rs1003853 esetében a lókuszt pozíciója a myeloid cinkujj 1-nek (myeloid zinc finger 1, MZF1) a MZF1::1-4\_1 motívumában, NF- $\kappa$ B transzkripció faktor NF-kappaB\_disc2 és p300 transzkripció koaktivátor p300\_known1 kötőmotívumának szekvenciájában foglal helyet. Prediktív LOD-érték következetesen az MZF1::1-4\_1 és NF-kappaB\_disc2 esetében +1,6 és +11,1, mely azt mutatja, hogy az alternatív T-allél jelenlétében erősebb és a referencia C-allél jelenlétében gyengébb affinitással kötődik a TF. A p300\_known1 esetében a LOD-érték megfordul a referencia C-allél javára + 1,3 értékkel, mely fordított helyzetnél az előzőektől eltérően az alternatív T-allél köti gyengébben a p300 transzkripció koaktivátort.

A predikciós modell alapján az rs1003854 lókuszt a GATA a cink-ujj fehérje család – melyek képesek a (T/A)GATA(A/G) konszenzusos DNS-szekvenciához kötődni – GATA\_disc3 kötőmotívumának szekvenciájában foglal helyet (157). A prediktív LOD-érték +1 értékkel magasabb az alternatív C-allél esetében, vagyis a referencia T-allél ezt a transzkripció faktort gyengébben köti (157). A helysúlyozott mátrix predikciós modelljének *in silico* eredményeit a **XII. táblázat** mutatja be.

Lókusz	Ref	Alt	Position Weight Matrix (helysúlyozott mátrix) ID	LOD value Ref	LOD value Alt
rs878081	C	T	NF-kappaB_disc2	1,5	12,6
			NF-kappaB_known4	12,5	14,5
			YY1_disc1	8,6	9,5
rs1003853	C	T	AP-2rep	3,0	10,1
			Rad21_disc5	10,2	10,7
rs1055311	C	T	MZF1::1-4_1	10,3	11,9
			NF-kappaB_disc2	1,7	12,8
			p300_known1	11,1	9,8
rs1003854	T	C	GATA_disc3	12,2	13,2

Ref: referencia allél; Alt: alternatív allél; LOD: log odds ratio

**XII. táblázat: HaploReg4.2 adatbázis predikciós modellje alapján az AIRE allélpolimorfizmusok pozíciója transzkripciós faktorok DNS kötőrégiójának motívumaiban, és allélspecifikus hatásuk a TF kötődési affinitására, <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php> (172)**

## V. Megbeszélés

Kutatásunk egyik legfőbb célja az volt, hogy behatóan megismerjük az AIRE transzkripciós faktor összefüggését lehetséges autoimmun betegségekkel. Mivel az AIRE funkciója a CD4<sup>+</sup> autoreaktív timociták kiszűrése és törlése, a tímusz involúcióját – mely pubertáskorban kezdődik – követően fiatal felnőttkorban az expresszió helyileg kimutathatatlan, igazolja azt a feltevést mely szerint az AIRE transzkripciós faktor működése a csecsemőkorral van összefüggésben. Továbbá a tímusz életkorhoz köthető természetes funkcióvesztése eredményezi azt, hogy az immunrendszer felnőttkorra egy meglévő, a továbbiakban nagymértékben nem gyarapodó T-sejt repertoárral dolgozik, a naív T-sejtek termelése szignifikánsan alábbhagy. Amennyiben tehát a genetikai alapú betegségekre megléte bebizonyosodik azon túl, hogy ez összefüggést jelent az AIRE transzkripciós faktorról, egyben összefüggést jelent a csecsemőkorral, ami a betegség patomechanizmusát új megvilágításba helyezné. Az AIRE a negatív szelekció alapjául szolgáló TRA termelésért felelős. Amennyiben nem vagy nem kellő mértékben termelődik, a centrális tolerancia ezen lépése hiányos lesz vagy elmaradhat, megemelve az autoreaktív, naív T-sejtek klonális túlélését. A fentiekben ismertetem Lovewell és mtsai.-nak munkáját, melyben kettő, az rs117557896, rs751032 lókuszon található AIRE-655G AIRE-230T allélpolimorfizmusok alkotta genotípus esetében a promóter aktivitása nagyon alacsony, a nullához közelített, továbbá a negatív szelekció elmaradt (95). Amennyiben két allélpolimorfizmus megléte ilyen drasztikus expressziós különbséget képes eredményezni, elképzelhető, hogy az AIRE promóter régióján allélpolimorfizmusok sora és az azok általi haplotípusok egy személyre szabott AIRE expressziós profilt hoznak létre. A populáció egészét tekintve az esetek többségében ez elegendő expressziót biztosít, viszont előfordulhat olyan AIRE transzkripciós érintő haplotípus kombináció, mely nem biztosít kellő mennyiségű AIRE fehérjét a TRA expresszióhoz, a negatív szelekció elmaradhat, mely az autoreaktív, naív T-sejtek klonális expanziójához vezethet. Számos autoimmun betegség kiváltó oka az autoreaktív T-sejt populáció, így amennyiben az AIRE egy kulcsszereplő az immunológiai toleranciában elképzelhető, hogy allélpolimorfizmus által számos autoimmun betegség közös nevezője, oki tényezője lehet, ennél fogva az AIRE számos autoimmun betegség kiindulópontja lehet.

Ennek kutatásához csecsemőkori *AIRE* expressziós profil kialakítására, haplotípus analízisre lenne szükség tímusz szövetmintákon, valamint ezen csecsemők élethossz-kohorsz utánkötése, egy nagy elemszámú vizsgálatban, kontrollcsoport bevonásával. Továbbá mindemelett a centrális tolerancia teljes szereplőgárdájának vizsgálata adhatna egy részletesebb képet. Ilyen fokú komplex vizsgálatok kivitelezésére nem törekedtünk. Vizsgálatsorozatunk egyes lépései egymásból alakultak ki. Meta-analízisünkkel 2018-ban a vizsgálat legelején arra törekedtünk, hogy meghatározzuk a tudomány jelenlegi állása szerint melyik az az autoimmun betegség, mely leginkább összefüggésbe hozható az *AIRE* allélpolimorfizmusokkal. Publikációnk során miután meghatároztuk, hogy ez a reumatoid artritisz azt tisztáztuk, mely allélpolimorfizmusok melyik rasszt tekintve adnak szignifikáns összefüggést (105). A legnagyobb populáció melyet bevonhattunk ázsiai volt, melyben meghatároztunk RA-ra nézve az rs2075876 lókuszt A-allélja és a rs760426 G-allélja, illetve ezeket az allélokot hordozó genotípusok – legyen homo- vagy heterozigóta – szignifikáns összefüggésben állnak a betegséggel, majd hogyan az összes genetikai modellt tekintve. Meta-analízisünk tehát megállapította, mely allélpolimorfizmusok jelentenek kockázatot ázsiai populációban RA-ra a közlemények 2018-as állása alapján (105). Miután bebizonyosodott, hogy az allélpolimorfizmusok összefüggésben állnak a betegséggel, ugyanakkor szignifikáns összefüggést nem kaptunk európai populációra így kíváncsiak voltunk, létezik-e kockázatot jelentő allélpolimorfizmus kaukázusi személyek között. Ennek érdekében megterveztük a további lépéseket, melynek során *in silico* vizsgálattal 44 SNP-t detektáltunk az *AIRE* régióban melyek potenciálisan hozzájárulhatnak autoimmun körkép kialakulásához. Ebből a 44-ből választottunk ki kettőt a DNS-kötő régiót kódoló SAND domén vizsgálatára: rs878081 és rs1003854 lókuszt. Pilot vizsgálatunk elsődleges szerepe az összefüggések keresése és a vizsgálati módszerek optimalizálása volt (152). Eredményeink alapján rs878081 és rs1003854-es lókusza szignifikáns összefüggést mutatott RA-val, a betegségkockázatot jelentő allél hordozása kockázatot jelentett ESR és CRP emelkedésre. Közleményünk elsőként hozta összefüggésbe rs878081 és rs1003854 lókusztok adta genotípusokat RA-val, továbbá szintén elsőként sikerült összefüggést találni genotípusok és klinikai paraméterek (ESR és CRP) között a betegpopulációban (152). Azon RA betegek, melyek kettő kockázatot jelentő allélt hordoztak az rs1003854-es lókuszon, szignifikánsan magasabb CRP-értékkel rendelkeztek mind a recesszív ( $p=0,029$ ), mind pedig a kodomináns ( $p=0,043$ ) genetikai modellben, mely genetikai konstitúció egyben pozitív korrelációt is mutatott CRP-vel mind a két modellben ( $p=0,002$  és  $p=0,006$ ) (152). Ennek magyarázatára *Bergström és mtsai.*-nak vizsgálata adhat magyarázatot (173). A betegség centrális szakaszában alakul ki az ízületi tokban a gyulladási környezet, melynek során az FLS sejtek kemokineket (CXCL10, CCL2, CCL5, CCL8) termelnek, a jelenlévő VCAM-1 sejtekkel együtt segítik az itt keringő autoreaktív T-sejtek extravazációját. A kísérlet során azt tapasztalták, hogy ha az *AIRE* expressziót siRNA-sel némítják FLS populációban, a sejtek 26x magasabb kemokin produkciót mutatnak szemben a nem stimuláltakhoz képest, ahol kemokintermelés kimutathatatlan volt, vagyis az *AIRE*-némított FLS sejtpopuláció gyulladási aktivitása magasabb volt (173). Allélpolimorfizmusok mint láttuk nagymértékben hozzájárulhatnak az *AIRE* expressziójához, vagyis elképzelhető hogy az *AIRE* rs1003854 lókusztán a kockázati allélra homozigóta genotípus nagymértékben csökkentette az expressziót, mely évtizedekkel később az FLS sejteken keresztül kialakított egy kemokinekben gazdag környezetet autoreaktív T-sejtek számára. Ezt az is alátámasztja, hogy a kemokinek (CXCL10, CCL8 és CCL5) erősen asszociálnak CRP szintekkel (174-176). Pilot studyt követően vizsgálatunkat kiterjesztettük további 3 lókusztúra és 592 före (157). A genetikai expozíció és a betegség közti összefüggés érdekében öt lókuszt allélpolimorfizmusait vizsgáltuk, melyek közül a SAND domén 5-ös exonjában található rs878081, az 5-ös intron splice helyen található rs1003853 és a 7-es intron splice helyen található rs1003854

szignifikáns pozitív asszociációt mutatott a betegségkockázattal abban az esetben, ha a kockázatot jelentő allélra a genotípus homozigóta ( $p=0.017$ ,  $p=0.004$ ,  $p=0.008$ ) (157). Két lókus esetében (rs878081, rs1055311) pedig összefüggés mutatkozott RA betegek genetikai alcsoportjai és ESR valamint aCCP klinikai paraméterek között (157). Eredményeink alapján az rs1055311 lókus allél- és genotípus-frekvenciája nem mutat szignifikáns összefüggést az RA-val, de korrelál az aCCP-szintekkel. Ez többek között azért fordulhat elő, mert az allélpolimorfizmus episztatikus úton, köztes változóra hat, mely az autoantitest termelés, vagyis egy köztes fenotípust határoz meg, mely allélpolimorfizmus így közvetlenül nincs összefüggésben a betegségkockázattal. Befolyásolhatja az autoreaktív T-sejtek túlélését, amelyek belsőleg elősegítik az aCCP antitestek termelődését, jellemezve a betegség köztes fenotípusát.

Mindezen betegségkockázat eredmények alapján tehát úgy gondoljuk, hogy mind az exonok szinonim, mind pedig splice régiók introniális allélpolimorfizmusai illetve az azok által kialakított genotípus jelenlétében az AIRE expresszió kellően alacsony ahhoz, hogy a TRA expresszió megvalósuljon, így a kontroll alól autoreaktív timociták szabadulhatnak ki, melyek naív T-sejtté alakulva a másodlagos limfoid szövetekben klonálisan túlélnek és egészen a centrális szakasz extravazációjáig rejtve maradnak. Innentől több irányban triggerelik a patomechanizmust, egyrészt citokineket termelnek (IFN- $\gamma$ , IL-17), melyek később makrofágokat idevonzva circulus vitiosus fenntartva, másrészt autoreaktív CD4<sup>+</sup> T-sejtek az FLS sejtpopuláció expanszióját okozza, melynek eredménye a heges kötőszövet képződése. Kockázatot jelentő allélra homozigóta egyének esetében elképzelhető, hogy genetikai kockázatból adódó alacsony AIRE expresszió egy az átlagnál aktívabb gyulladáshoz vezető környezetet teremt, azáltal, hogy kemokinek széles spektruma termelődik AIRE hiányában az FLS sejtekben, melynek hatására további citokinek, autoantitestek, felgyülemlett immunkomplexek képződnek (173-176).

Vizsgálatunk utolsó lépéseként ok-okozati összefüggések mélyebb feltárása érdekében *in silico* elemeztük a kockázati alléljaink molekuláris következményeit. Ennek során azt találtuk, hogy az 5 lókuszból 4, melyek vagy a betegségkockázattal, vagy pedig a klinikai paraméterekkel mutattak statisztikailag szignifikáns összefüggést, transzkripciós faktorok DNS kötőmotívumainak szekvenciájában foglalnak helyet és *in silico* adatok alapján hatással lehetnek azok kötődési affinitására. A kapott eredményeket az alapján választottuk ketté, hogy a referencia (Ref) allél jelenlétében erősebb-e vagy gyengébb a transzkripciós faktor kötődés a DNS kötőmotívum szekvenciájához, melyben az allélpolimorfizmus helyet foglal. A transzkripciós fátumot nem minden esetben lehetséges meghatározni, egy-két kivételtől eltekintve erősen kontextusfüggőek. Azt találtuk, hogy az exon 5-ös rs878081 és exon 6-os rs1055311 Ref alléljának (C) jelenlétében gyengébben kötődik az NF- $\kappa$ B. *Haljasorg és mtsai*-nak vizsgálata alapján ennek a transzkripciós faktornak az erős affinitású kötődése az mTEC DNS-hez szükséges a negatív szelekciós AIRE expresszióhoz a tímuszban (35). Mindkét lókus esetében a kockázatot jelentő allél a Ref C-allél volt, melynek homozigóta genotípusa asszociált a betegségkockázattal rs878081 esetében és aCCP-vel rs1055311 esetében (157). Eset-kontroll és *in silico* eredményeink alapján elképzelhetőnek tartjuk azt a feltevést, hogy ezen allél jelenlétében nem volt képes megfelelő affinitással kötődni az NF- $\kappa$ B az mTEC DNS-éhez, melynek következménye az alacsony AIRE expresszió illetve az abból fakadó TRA expresszió, mely eredményezhette a negatív szelekció részleges vagy teljes elmaradását. Továbbiakban az rs878081 esetében Ref allélhoz gyengébben kötődött YY1, mely egy cink-ujj multifunkciós transzkripciós faktor, mely kontextus függően aktivátor vagy represszor. Az utóbbira példa, amennyiben hiszton deacetiláz (histone deacetylase, HDAC) jelenlétében kötődik a target promóter régiójához, transzkripciós repressziót, kromatinkondenzációt okoz (177). Nagyságrendekkel gyengébben kötődik az intron 5-ös rs1003853 Ref (C) alléljához az AP-2 transzkripciós faktor, mely DNS-kötő doménjével promóter és enhancer régiókhöz

kötődik, erős affinitása transzkripciót aktivál. Ezen lókuszt Ref allélja jelenlétében, szintén gyengébben kötődik a Rad21, melynek erős affinitású kapcsolódása szükséges a DNS-repairhez, kohezin-komplex része, mely segíti a kromoszómaszegregációt meiózis és mitózis alatt. Gyenge affinitású kötődése befolyással lehet ezeknek a funkcióknak a betöltésére (178). Eredményeink alapján elmondható, hogy az rs1055311 lókuszt Ref (C) alléljához MZF1 transzkripciós faktor gyengébben kötődik, melynek szerepe többek között a mieloid DC sejtek fejlődésének a biztosítása (179). Az intron 7-es rs1003854 Ref (T) allélja esetében a GATA transzkripciós faktor család gyengébben kötődik a DNS TF kötő motívumához. A GATA alapvető fontosságú a CD4<sup>+</sup> timociták érésében, hatással van a  $\beta$ -szelekcióra és meghatározzák a CD4<sup>+</sup> sejt vonalak elköteleződését a centrális toleranciában (180). Mivel ez az allélpolimorfizmus intronikus, elképzelhető, hogy alternatív splicing folyamatán lesz képes hatni az AIRE expresszióra (181). A TF-ok közül 1 kötődött erősebben a Ref allélhoz, ez pedig a p300 az rs1055311 C-alléljához. Hiszton acetil transzferáz (HAT) enzimként acetilcsoportot transzportál a hiszton molekulák lizin oldalláncára mellyel a kromatint elektromos töltések következményeként relaxálja és elérhetővé teszi a DNS-t RNS polimeráz II (RNA pol II) számára, vagyis transzkripciót aktivál. *In silico* vizsgálatunk eredményeként elmondható, hogy az 5 lókuszból 4 esetében a Ref allélt tartalmazó DNS kötőmotívumhoz számos TF gyengébb affinitással kapcsolódik, melyek közül hangsúlyosak voltak azon transzkripciós faktorok, melyek alapvetően a génexpressziót határozzák meg, előrevetítve ez úton is az AIRE allélpolimorfizmus-függő expresszióját ennek megfelelően TRA-vezérelt negatív szelekcióját.

Eredményeink mellett vizsgálatainknak számos limitációja van. Legfőbb limitáció, mely az összefüggés erősségét negatívan befolyásolhatta a korlátozott elemszám. Ebben elsődlegesen a COVID-19 világjárvány miatti betegforgalom visszaesés játszott szerepet. Másodlagosan pedig a két kórház nettó betegforgalma, 2019 és 2023 között ennyi RA beteg és kontroll vizsgálatba kapcsolására volt képes. Egy világjárvánnyal terhelt 4 éves mintagyűjtési időszakban ez volt az a maximum, melyet a két kórház biztosítani tudott, mely teljesítményért nem győzünk hálásak lenni a Harkányi Termál Rehabilitációs Centrum és a Hévízgyógyfürdő és Szent András Reumakórház főorvosainak, orvosainak, nővéreinek akik nélkül ez a vizsgálat sorozat nem valósulhatott volna meg. A vizsgálatunkban résztvevők számának növelése jelentősen növelte volna mind az összefüggések számát, mind pedig a statisztikai szignifikanciát. További korlátot jelent, hogy a vizsgálatba egy domén kódoló régiót vontunk be, melynek eredményei nem képesek átfogó betekintést nyújtani az AIRE funkciójába és egyes allélpolimorfizmusok AIRE-re mért expressziós következményeibe. Elképzelésünk szerint az összes domén kódoló régióban található összes allélpolimorfizmus és AIRE-partnerek genomszintű elemzése és eredményei nem adnának értékelhető eredményeket anélkül, hogy azokat izoláltan, egymástól elkülönítve ne vizsgálnánk meg, melynek folytatása jövőbeli terveink közé tartozik. További limitáció, hogy az AIRE mRNS-expresszióját nem állt módunkban vizsgálni, mely nagymértékben hozzájárult volna az allél- illetve a genotípusfüggő transzkripciós következmények jobb megértéséhez. Jövőbeli terveink között szerepel az RA és a centrális tolerancia közötti kapcsolat, valamint az AIRE partnerek szerepének átfogóbb megértése további genetikai asszociációs vizsgálatokon a fennmaradó 11 intronikus és 12 exonikus kódoló szekvenciára vonatkozóan.

## VI. Új eredmények

Vizsgálatunk összességében 6 részből állt. Elsőként ázsiai populációkban az rs2075876 és rs760426 lókuszon lévő allélpolimorfizmusok meta-analízisét végeztük el. Itt 2018-ban elsőként határoztunk meg, hogy az rs2075876 kockázatot jelentő A-allélja valamint rs760426

kockázatot jelentő G-allélja illetve az általuk meghatározott genotípusok és genotípus csoportok statisztikailag szignifikáns összefüggést mutattak RA-val, kizárólag ázsiai populációban. Ezt az összefüggést kutatócsoportunk publikálta először (105).

Következő lépcsőként MAF allélfrekvenciák alapján meghatároztunk 44 olyan SNP-t, melyek potenciálisan vizsgálhatóak európai, kaukázusi populációkban melyekből kettőt, rs878081 és rs1003854-t kiválasztva elvégeztük pilot vizsgálatunkat, melynek végére elsőként határoztuk meg, hogy az rs878081 és rs1003854 kockázatot jelentő alléljai és genotípusai statisztikailag összefüggnek betegségkockázattal illetve ESR és CRP klinikai paraméterekkel (152). Európai populációban elsőként kutatócsoportunk pilot vizsgálata elemezte az allélpolimorfizmusok és RA, valamint klinikai paraméterek közötti összefüggést hat különböző (allélikus, domináns, recesszív, kodomináns hetero- és homozigóta) genetikai modell alapján (152). Az addigi eredmények alapján következő vizsgálatunk elemszámát 529 főre és lókuszt-állományát 3 további pozícióra kiterjesztve, azt az eredményt kaptuk, hogy az öt kiválasztott lókuszból az rs878081 és rs1003853 kockázati C- valamint rs1003854 T-allélja illetve azok genotípusai statisztikailag szignifikáns összefüggésben állnak RA betegségkockázattal. Klinikai paraméterek közül ESR statisztikailag szignifikáns pozitív asszociációt és az aCCP pedig pozitív korrelációt mutatott a kockázatot jelentő C-allélra homozigóta genotípussal az rs878081 és rs1055311 esetében (157). Vizsgálatsorozatunk utolsó lépéseként *in silico* elemeztük az egyes allélpolimorfizmusok adta transzkripciós következményeket, melynek során elsőként határoztuk meg 7 különböző transzkripciós faktor viselkedését kockázatot jelentő allélpolimorfizmus jelenlétében. Ezek közül egyezményesen kijelenthető, hogy a kockázatot jelentő SNP-k jelenlétében 7-ből 6 TF gyengébben kötődik a DNS kötőmotívumhoz, melyek közül 4-ből 2 SNP az NF- $\kappa$ B gyengébb kötődését mutatták, mely a neonatális negatív szelekció során felelős a TRA expressziót biztosító *AIRE* transzkripcióért. Eredményeink alapján az egyik lehetséges következmény a kockázatot jelentő allélok jelenlétében az NF- $\kappa$ B gyengébb kötődése, így a neonatális negatív szelekció során nem termelődhet megfelelő mennyiségű *AIRE* illetve TRA, mely előíranyozza az autoreaktív timociták lehetséges klonális túlélését, megteremtve ez által autoimmun betegségek, köztük RA kialakulásának kezdeti immunológiai talaját.

## VII. Köszönetnyilvánítás

Szeretnék köszönetet mondani az Orvosi Népegészségtani Prof. Dr. Kiss István intézetvezetőnek, hogy lehetővé tette, hogy munkámat az általa vezetett intézetben végezhessem.

Hálásan köszönöm Dr. Gyöngyi Zoltán egyetemi docensnek, témavezetőmnek, hogy szakmailag és emberileg is támogatott munkám során.

Köszönettel tartozom Dr. Péter Ivánnak, a Harkányi Termál Rehabilitációs Központ kórházigazgatójának, hogy lehetővé tette, hogy munkámat az általa vezetett kórházban végezhessem.

Köszönöm Dr. Németh Balázs, egyetemi adjunktusnak, hogy a kutatás kezdeti lépéseiben segítséget nyújtott.

Köszönettel tartozom Dr. Nusser Nórának, a Harkányi Termál Rehabilitációs Központ orvosigazgatójának, azért a mérhetetlen szakmai segítségért, szervezésért, továbbá jóindulatért és bizalomért, emberségért, melyet munkám során tőle kaptam, melyek nélkül ez a vizsgálat, ebben a formájában elképzelhetetlen lett volna.

Köszönettel tartozom Feledi Zsuzsannának, a Harkányi Termál Rehabilitációs Központ laboratórium vezetőjének, hogy a mintákon dolgozó csapatot összefogta, továbbá Tóthné Kamanczi Zita laboránsnak, hogy szisztematikusan és fáradhatatlanul gyűjtötte és

katalogizálta mintáinkat. Köszönettel tartozom továbbá Pappné Netik Natáliának, Óbert Anitának és Szűcs Adriennek a harkányi mintagyűjtésben nyújtott segítségükért.

Köszönettel tartozom Dr. Holléné Dr. Mándó Zsuzsannának, a Hévízgyógyfürdő és Szent András Reumakórház kórházigazgatójának, hogy lehetővé tette, hogy munkámat az általa vezetett kórházzal együttműködésben végezhessem.

Köszönettel tartozom Dr. Kulisch Agotának, a Hévízgyógyfürdő és Szent András Reumakórház osztályvezető főorvosának, azért a rengeteg segítségért mellyel lehetővé tette, hogy az együttműködés és mintagyűjtés létrejöhsen. Nélküle szintén elképzelhetetlen lett volna, hogy vizsgálatunk megvalósuljon.

Köszönettel tartozom Dr. Fülöp Jánosnének, a Hévízgyógyfürdő és Szent András Reumakórház főnővérének, a hévízi mintagyűjtés során nyújtott mérhetetlen segítségért.

Továbbá köszönet illeti az Orvosi Népegészségtani Intézet összes munkatársát munkám során nyújtott segítségükért.

Végül, de nem utolsó sorban köszönöm feleségemnek azt a támogatást és biztatást mellyel nagymértékben hozzájárult a dolgozat elkészüléséhez. Továbbá köszönöm egész családomnak szintén azt a támogatást és biztatást, mellyel nagyban hozzájárultak a dolgozat elkészüléséhez. A kutatások támogatásáért köszönet illeti a PTE ÁOK, Kari Kutatási Alap (ÁOK-KA) KA-304516 számú pályázatát.

### **VIII. Irodalomjegyzék**

1. Yang L, Jin R, Lu D, Ge Q. T cell Tolerance in Early Life. *Front Immunol.* 2020 Nov 20;11:576261. doi: 10.3389/fimmu.2020.576261. PMID: 33329542; PMCID: PMC7714828.
2. Kyewski B, Klein L. A central role for central tolerance. *Annu Rev Immunol.* 2006;24:571-606. doi: 10.1146/annurev.immunol.23.021704.115601. PMID: 16551260.
3. Dzhagalov IL, Chen KG, Herzmark P, Robey EA. Elimination of self-reactive T cells in the thymus: a timeline for negative selection. *PLoS Biol.* 2013;11(5):e1001566. doi: 10.1371/journal.pbio.1001566. Epub 2013 May 21. PMID: 23700386; PMCID: PMC3660248.
4. Chann AS, Charnley M, Newton LM, Newbold A, Wiede F, Tiganis T, Humbert PO, Johnstone RW, Russell SM. Stepwise progression of  $\beta$ -selection during T cell development involves histone deacetylation. *Life Sci Alliance.* 2022 Oct 25;6(1):e202201645. doi: 10.26508/lsa.202201645. PMID: 36283704; PMCID: PMC9595210.
5. Tong Q, Yao L, Su M, Yang YG, Sun L. Thymocyte migration and emigration. *Immunol Lett.* 2024 Apr 30;267:106861. doi: 10.1016/j.imlet.2024.106861. Epub ahead of print. PMID: 38697225.
6. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol.* 2014 Jun;14(6):377-91. doi: 10.1038/nri3667. Epub 2014 May 16. PMID: 24830344; PMCID: PMC4757912.
7. Mathis D, Benoist C. Aire. *Annu Rev Immunol.* 2009;27:287-312. doi: 10.1146/annurev.immunol.25.022106.141532. PMID: 19302042.

8. Ruan QG, Tung K, Eisenman D, Setiady Y, Eckenrode S, Yi B, Purohit S, Zheng WP, Zhang Y, Peltonen L, She JX. The autoimmune regulator directly controls the expression of genes critical for thymic epithelial function. *J Immunol*. 2007 Jun 1;178(11):7173-80. doi: 10.4049/jimmunol.178.11.7173. PMID: 17513766.
9. Ehrlich LI, Oh DY, Weissman IL, Lewis RS. Differential contribution of chemotaxis and substrate restriction to segregation of immature and mature thymocytes. *Immunity*. 2009 Dec 18;31(6):986-98. doi: 10.1016/j.immuni.2009.09.020. Epub 2009 Dec 3. PMID: 19962328; PMCID: PMC4106268.
10. Mori K, Itoi M, Tsukamoto N, Kubo H, Amagai T. The perivascular space as a path of hematopoietic progenitor cells and mature T cells between the blood circulation and the thymic parenchyma. *Int Immunol*. 2007 Jun;19(6):745-53. doi: 10.1093/intimm/dxm041. Epub 2007 May 9. PMID: 17493961.
11. Kwan J, Killeen N. CCR7 directs the migration of thymocytes into the thymic medulla. *J Immunol*. 2004 Apr 1;172(7):3999-4007. doi: 10.4049/jimmunol.172.7.3999. PMID: 15034011.
12. Ueno T, Saito F, Gray DH, Kuse S, Hieshima K, Nakano H, Kakiuchi T, Lipp M, Boyd RL, Takahama Y. CCR7 signals are essential for cortex-medulla migration of developing thymocytes. *J Exp Med*. 2004 Aug 16;200(4):493-505. doi: 10.1084/jem.20040643. Epub 2004 Aug 9. Erratum in: *J Exp Med*. 2004 Oct 4;200(7):following 946. PMID: 15302902; PMCID: PMC2211934.
13. Kohli K, Janssen A, Förster R. Plasmacytoid dendritic cells induce tolerance predominantly by cargoing antigen to lymph nodes. *Eur J Immunol*. 2016 Nov;46(11):2659-2668. doi: 10.1002/eji.201646359. Epub 2016 Oct 18. PMID: 27592607; PMCID: PMC5129535.
14. Klein L, Hinterberger M, von Rohrscheidt J, Aichinger M. Autonomous versus dendritic cell-dependent contributions of medullary thymic epithelial cells to central tolerance. *Trends Immunol*. 2011 May;32(5):188-93. doi: 10.1016/j.it.2011.03.002. Epub 2011 Apr 12. PMID: 21493141.
15. Perniola R. Twenty Years of AIRE. *Front Immunol*. 2018 Feb 12;9:98. doi: 10.3389/fimmu.2018.00098. PMID: 29483906; PMCID: PMC5816566.
16. Finnish-German APECED Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet*. 1997 Dec;17(4):399-403. doi: 10.1038/ng1297-399. PMID: 9398840.
17. Heino M, Peterson P, Sillanpää N, Guérin S, Wu L, Anderson G, Scott HS, Antonarakis SE, Kudoh J, Shimizu N, Jenkinson EJ, Naquet P, Krohn KJ. RNA and protein expression of the murine autoimmune regulator gene (Aire) in normal, RelB-deficient and in NOD mouse. *Eur J Immunol*. 2000 Jul;30(7):1884-93. doi: 10.1002/1521-4141(200007)30:7<1884::AID-IMMU1884>3.0.CO;2-P. PMID: 10940877.



18. Derbinski J, Schulte A, Kyewski B, Klein L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat Immunol.* 2001 Nov;2(11):1032-9. doi: 10.1038/ni723. PMID: 11600886.
19. Michelson DA, Hase K, Kaisho T, Benoist C, Mathis D. Thymic epithelial cells co-opt lineage-defining transcription factors to eliminate autoreactive T cells. *Cell.* 2022 Jul 7;185(14):2542-2558.e18. doi: 10.1016/j.cell.2022.05.018. Epub 2022 Jun 16. PMID: 35714609; PMCID: PMC9469465.
20. Abramson J, Goldfarb Y. AIRE: From promiscuous molecular partnerships to promiscuous gene expression. *Eur J Immunol.* 2016 Jan;46(1):22-33. doi: 10.1002/eji.201545792. Epub 2015 Nov 2. PMID: 26450177.
21. Gibson TJ, Ramu C, Gemünd C, Aasland R. The APECED polyglandular autoimmune syndrome protein, AIRE-1, contains the SAND domain and is probably a transcription factor. *Trends Biochem Sci.* 1998 Jul;23(7):242-4. doi: 10.1016/s0968-0004(98)01231-6. PMID: 9697411.
22. Carles CC, Fletcher JC. Missing links between histones and RNA Pol II arising from SAND? *Epigenetics.* 2010 Jul 1;5(5):381-5. doi: 10.4161/epi.5.5.11956. Epub 2010 Jul 1. PMID: 20458168.
23. Qi Y, Zhang R, Lu Y, Zou X, Yang W. Aire and Fezf2, two regulators in medullary thymic epithelial cells, control autoimmune diseases by regulating TSAs: Partner or complements? *Front Immunol.* 2022 Aug 30;13:948259. doi: 10.3389/fimmu.2022.948259. PMID: 36110862; PMCID: PMC9468217.
24. Abbott JK, Huoh YS, Reynolds PR, Yu L, Rewers M, Reddy M, Anderson MS, Hur S, Gelfand EW. Dominant-negative loss of function arises from a second, more frequent variant within the SAND domain of autoimmune regulator (AIRE). *J Autoimmun.* 2018 Mar;88:114-120. doi: 10.1016/j.jaut.2017.10.010. Epub 2017 Nov 10. PMID: 29129473; PMCID: PMC5846191.
25. Santos JC, Dametto M, Masson AP, Faça VM, Bonacin R, Donadi EA, Passos GA. The AIRE G228W mutation disturbs the interaction of AIRE with its partner molecule SIRT1. *Front Immunol.* 2022 Sep 6;13:948419. doi: 10.3389/fimmu.2022.948419. PMID: 36148232; PMCID: PMC9485725.
26. Oftedal BE, Assing K, Baris S, Safgren SL, Johansen IS, Jakobsen MA, Babovic-Vuksanovic D, Agre K, Klee EW, Majcic E, Ferré EMN, Schmitt MM, DiMaggio T, Rosen LB, Rahman MO, Chrysis D, Giannakopoulos A, Garcia MT, González-Granado LI, Stanley K, Galant-Swofford J, Suwannarat P, Meyts I, Lionakis MS, Husebye ES. Dominant-negative heterozygous mutations in AIRE confer diverse autoimmune phenotypes. *iScience.* 2023 May 5;26(6):106818. doi: 10.1016/j.isci.2023.106818. PMID: 37235056; PMCID: PMC10206195.
27. Yang S, Bansal K, Lopes J, Benoist C, Mathis D. Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. *Proc Natl Acad Sci U S A.* 2013 Jan 29;110(5):1833-8. doi: 10.1073/pnas.1222023110. Epub 2013 Jan 14. PMID: 23319629; PMCID: PMC3562810.

28. Meloni A, Incani F, Corda D, Cao A, Rosatelli MC. Role of PHD fingers and COOH-terminal 30 amino acids in AIRE transactivation activity. *Mol Immunol*. 2008 Feb;45(3):805-9. doi: 10.1016/j.molimm.2007.06.156. Epub 2007 Aug 1. PMID: 17675238.
29. Derbinski J, Gäbler J, Brors B, Tierling S, Jonnakuty S, Hergenahn M, Peltonen L, Walter J, Kyewski B. Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *J Exp Med*. 2005 Jul 4;202(1):33-45. doi: 10.1084/jem.20050471. Epub 2005 Jun 27. PMID: 15983066; PMCID: PMC2212909.
30. Gillard GO, Farr AG. Contrasting models of promiscuous gene expression by thymic epithelium. *J Exp Med*. 2005 Jul 4;202(1):15-9. doi: 10.1084/jem.20050976. Epub 2005 Jun 27. PMID: 15983067; PMCID: PMC2212887.
31. Gardner JM, Devoss JJ, Friedman RS, Wong DJ, Tan YX, Zhou X, Johannes KP, Su MA, Chang HY, Krummel MF, Anderson MS. Deletional tolerance mediated by extrathymic Aire-expressing cells. *Science*. 2008 Aug 8;321(5890):843-7. doi: 10.1126/science.1159407. PMID: 18687966; PMCID: PMC2532844.
32. Lopes N, Sergé A, Ferrier P, Irla M. Thymic Crosstalk Coordinates Medulla Organization and T-Cell Tolerance Induction. *Front Immunol*. 2015 Jul 20;6:365. doi: 10.3389/fimmu.2015.00365. PMID: 26257733; PMCID: PMC4507079.
33. van de Pavert SA, Mebius RE. New insights into the development of lymphoid tissues. *Nat Rev Immunol*. 2010 Sep;10(9):664-74. doi: 10.1038/nri2832. Epub 2010 Aug 13. PMID: 20706277.
34. Rossi SW, Kim MY, Leibbrandt A, Parnell SM, Jenkinson WE, Glanville SH, McConnell FM, Scott HS, Penninger JM, Jenkinson EJ, Lane PJ, Anderson G. RANK signals from CD4(+)3(-) inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. *J Exp Med*. 2007 Jun 11;204(6):1267-72. doi: 10.1084/jem.20062497. Epub 2007 May 14. PMID: 17502664; PMCID: PMC2118623.
35. Haljasorg U, Bichele R, Saare M, Guha M, Maslovskaja J, Kõnd K, Remm A, Pihlap M, Tomson L, Kisand K, Laan M, Peterson P. A highly conserved NF- $\kappa$ B-responsive enhancer is critical for thymic expression of Aire in mice. *Eur J Immunol*. 2015 Dec;45(12):3246-56. doi: 10.1002/eji.201545928. Epub 2015 Oct 26. PMID: 26364592.
36. Murumägi A, Silvennoinen O, Peterson P. Ets transcription factors regulate AIRE gene promoter. *Biochem Biophys Res Commun*. 2006 Sep 22;348(2):768-74. doi: 10.1016/j.bbrc.2006.07.135. Epub 2006 Jul 31. PMID: 16890195.
37. Herzig Y, Nevo S, Bornstein C, Brezis MR, Ben-Hur S, Shkedy A, Eisenberg-Bord M, Levi B, Delacher M, Goldfarb Y, David E, Weinberger L, Viukov S, Ben-Dor S, Giraud M, Hanna JH, Breiling A, Lyko F, Amit I, Feuerer M, Abramson J. Transcriptional programs that control expression of the autoimmune regulator gene Aire. *Nat Immunol*. 2017 Feb;18(2):161-172. doi: 10.1038/ni.3638. Epub 2016 Dec 12. PMID: 27941786.
38. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med*. 2015 Jun 5;47(6):e166. doi: 10.1038/emm.2015.33. PMID: 26045254; PMCID: PMC4491725.

39. Grandi FC, Modi H, Kampman L, Corces MR. Chromatin accessibility profiling by ATAC-seq. *Nat Protoc.* 2022 Jun;17(6):1518-1552. doi: 10.1038/s41596-022-00692-9. Epub 2022 Apr 27. PMID: 35478247; PMCID: PMC9189070.
40. Yanagihara T, Sanematsu F, Sato T, Uruno T, Duan X, Tomino T, Harada Y, Watanabe M, Wang Y, Tanaka Y, Nakanishi Y, Suyama M, Yoshinori F. Intronic regulation of Aire expression by *Jmjd6* for self-tolerance induction in the thymus. *Nat Commun.* 2015 Nov 4;6:8820. doi: 10.1038/ncomms9820. PMID: 26531897; PMCID: PMC4667615.
41. Unoki M, Masuda A, Dohmae N, Arita K, Yoshimatsu M, Iwai Y, Fukui Y, Ueda K, Hamamoto R, Shirakawa M, Sasaki H, Nakamura Y. Lysyl 5-hydroxylation, a novel histone modification, by Jumonji domain containing 6 (JMJD6). *J Biol Chem.* 2013 Mar 1;288(9):6053-62. doi: 10.1074/jbc.M112.433284. Epub 2013 Jan 9. PMID: 23303181; PMCID: PMC3585044.
42. Yanagihara T, Tomino T, Uruno T, Fukui Y. Thymic epithelial cell-specific deletion of *Jmjd6* reduces Aire protein expression and exacerbates disease development in a mouse model of autoimmune diabetes. *Biochem Biophys Res Commun.* 2017 Jul 15;489(1):8-13. doi: 10.1016/j.bbrc.2017.05.113. Epub 2017 May 22. Erratum in: *Biochem Biophys Res Commun.* 2017 Sep 30;491(4):1126. PMID: 28546003.
43. Svoboda P. Key Mechanistic Principles and Considerations Concerning RNA Interference. *Front Plant Sci.* 2020 Aug 13;11:1237. doi: 10.3389/fpls.2020.01237. PMID: 32903622; PMCID: PMC7438612.
44. Tanaka PP, Oliveira EH, Vieira-Machado MC, Duarte MJ, Assis AF, Bombonato-Prado KF, Passos GA. miR-155 exerts posttranscriptional control of autoimmune regulator (Aire) and tissue-restricted antigen genes in medullary thymic epithelial cells. *BMC Genomics.* 2022 May 28;23(1):404. doi: 10.1186/s12864-022-08631-4. PMID: 35643451; PMCID: PMC9145475.
45. Ilmarinen T, Melén K, Kangas H, Julkunen I, Ulmanen I, Eskelin P. The monopartite nuclear localization signal of autoimmune regulator mediates its nuclear import and interaction with multiple importin alpha molecules. *FEBS J.* 2006 Jan;273(2):315-24. doi: 10.1111/j.1742-4658.2005.05065.x. PMID: 16403019.
46. Akiyoshi H, Hatakeyama S, Pitkänen J, Mouri Y, Doucas V, Kudoh J, Tsurugaya K, Uchida D, Matsushima A, Oshikawa K, Nakayama KI, Shimizu N, Peterson P, Matsumoto M. Subcellular expression of autoimmune regulator is organized in a spatiotemporal manner. *J Biol Chem.* 2004 Aug 6;279(32):33984-91. doi: 10.1074/jbc.M400702200. Epub 2004 May 17. PMID: 15150263.
47. Anderson MS, Su MA. AIRE expands: new roles in immune tolerance and beyond. *Nat Rev Immunol.* 2016 Apr;16(4):247-58. doi: 10.1038/nri.2016.9. Epub 2016 Mar 14. PMID: 26972725; PMCID: PMC4831132.
48. Abramson J, Giraud M, Benoist C, Mathis D. Aire's partners in the molecular control of immunological tolerance. *Cell.* 2010 Jan 8;140(1):123-35. doi: 10.1016/j.cell.2009.12.030. PMID: 20085707.

49. Pitkänen J, Rebane A, Rowell J, Murumägi A, Ströbel P, Möll K, Saare M, Heikkilä J, Doucas V, Marx A, Peterson P. Cooperative activation of transcription by autoimmune regulator AIRE and CBP. *Biochem Biophys Res Commun*. 2005 Aug 5;333(3):944-53. doi: 10.1016/j.bbrc.2005.05.187. PMID: 15964547.
50. Liiv I, Rebane A, Org T, Saare M, Maslovskaja J, Kisand K, Juronen E, Valmu L, Bottomley MJ, Kalkkinen N, Peterson P. DNA-PK contributes to the phosphorylation of AIRE: importance in transcriptional activity. *Biochim Biophys Acta*. 2008 Jan;1783(1):74-83. doi: 10.1016/j.bbamcr.2007.09.003. Epub 2007 Oct 4. PMID: 17997173; PMCID: PMC2225445.
51. Oven I, Brdicková N, Kohoutek J, Vaupotic T, Narat M, Peterlin BM. AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. *Mol Cell Biol*. 2007 Dec;27(24):8815-23. doi: 10.1128/MCB.01085-07. Epub 2007 Oct 15. PMID: 17938200; PMCID: PMC2169392.
52. Giraud M, Jmari N, Du L, Carallis F, Nieland TJ, Perez-Campo FM, Bensaude O, Root DE, Hacohen N, Mathis D, Benoist C. An RNAi screen for Aire cofactors reveals a role for Hnrnp1 in polymerase release and Aire-activated ectopic transcription. *Proc Natl Acad Sci U S A*. 2014 Jan 28;111(4):1491-6. doi: 10.1073/pnas.1323535111. Epub 2014 Jan 16. PMID: 24434558; PMCID: PMC3910647.
53. Waterfield M, Khan IS, Cortez JT, Fan U, Metzger T, Greer A, Fasano K, Martinez-Llordella M, Pollack JL, Erle DJ, Su M, Anderson MS. The transcriptional regulator Aire coopts the repressive ATF7ip-MBD1 complex for the induction of immunotolerance. *Nat Immunol*. 2014 Mar;15(3):258-65. doi: 10.1038/ni.2820. Epub 2014 Jan 26. Erratum in: *Nat Immunol*. 2014 Sep;15(9):894. PMID: 24464130; PMCID: PMC4172453.
54. Chuprin A, Avin A, Goldfarb Y, Herzig Y, Levi B, Jacob A, Sela A, Katz S, Grossman M, Guyon C, Rathaus M, Cohen HY, Sagi I, Giraud M, McBurney MW, Husebye ES, Abramson J. The deacetylase Sirt1 is an essential regulator of Aire-mediated induction of central immunological tolerance. *Nat Immunol*. 2015 Jul;16(7):737-45. doi: 10.1038/ni.3194. Epub 2015 May 25. PMID: 26006015.
55. Rackaityte E, Halkias J. Mechanisms of Fetal T Cell Tolerance and Immune Regulation. *Front Immunol*. 2020 Apr 9;11:588. doi: 10.3389/fimmu.2020.00588. PMID: 32328065; PMCID: PMC7160249.
56. Meredith M, Zemmour D, Mathis D, Benoist C. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat Immunol*. 2015 Sep;16(9):942-9. doi: 10.1038/ni.3247. Epub 2015 Aug 3. PMID: 26237550; PMCID: PMC4632529.
57. Pinto S, Michel C, Schmidt-Glenewinkel H, Harder N, Rohr K, Wild S, Brors B, Kyewski B. Overlapping gene coexpression patterns in human medullary thymic epithelial cells generate self-antigen diversity. *Proc Natl Acad Sci U S A*. 2013 Sep 10;110(37):E3497-505. doi: 10.1073/pnas.1308311110. Epub 2013 Aug 26. PMID: 23980163; PMCID: PMC3773787.

58. Klein L, Hinterberger M, von Rohrscheidt J, Aichinger M. Autonomous versus dendritic cell-dependent contributions of medullary thymic epithelial cells to central tolerance. *Trends Immunol.* 2011 May;32(5):188-93. doi: 10.1016/j.it.2011.03.002. Epub 2011 Apr 12. PMID: 21493141.
59. Dzhagalov IL, Chen KG, Herzmark P, Robey EA. Elimination of self-reactive T cells in the thymus: a timeline for negative selection. *PLoS Biol.* 2013;11(5):e1001566. doi: 10.1371/journal.pbio.1001566. Epub 2013 May 21. PMID: 23700386; PMCID: PMC3660248.
60. Le Borgne M, Ladi E, Dzhagalov I, Herzmark P, Liao YF, Chakraborty AK, Robey EA. The impact of negative selection on thymocyte migration in the medulla. *Nat Immunol.* 2009 Aug;10(8):823-30. doi: 10.1038/ni.1761. Epub 2009 Jun 21. PMID: 19543275; PMCID: PMC2793676.
61. Canté-Barrett K, Gallo EM, Winslow MM, Crabtree GR. Thymocyte negative selection is mediated by protein kinase C- and Ca<sup>2+</sup>-dependent transcriptional induction of bim [corrected]. *J Immunol.* 2006 Feb 15;176(4):2299-306. doi: 10.4049/jimmunol.176.4.2299. Erratum in: *J Immunol.* 2006 Mar 15;176(6):3843. PMID: 16455986.
62. Staton TL, Lazarevic V, Jones DC, Lanser AJ, Takagi T, Ishii S, Glimcher LH. Dampening of death pathways by schnurri-2 is essential for T-cell development. *Nature.* 2011 Apr 7;472(7341):105-9. doi: 10.1038/nature09848. PMID: 21475200; PMCID: PMC3077958.
63. Weinreich MA, Hogquist KA. Thymic emigration: when and how T cells leave home. *J Immunol.* 2008 Aug 15;181(4):2265-70. doi: 10.4049/jimmunol.181.4.2265. PMID: 18684914; PMCID: PMC2861282.
64. Hunter MC, Teixeira A, Halin C. T Cell Trafficking through Lymphatic Vessels. *Front Immunol.* 2016 Dec 21;7:613. doi: 10.3389/fimmu.2016.00613. PMID: 28066423; PMCID: PMC5174098.
65. James KD, Jenkinson WE, Anderson G. T-cell egress from the thymus: Should I stay or should I go? *J Leukoc Biol.* 2018 Aug;104(2):275-284. doi: 10.1002/JLB.1MR1217-496R. Epub 2018 Feb 27. PMID: 29485734; PMCID: PMC6174998.
66. Chatzileontiadou DSM, Sloane H, Nguyen AT, Gras S, Grant EJ. The Many Faces of CD4<sup>+</sup> T Cells: Immunological and Structural Characteristics. *Int J Mol Sci.* 2020 Dec 23;22(1):73. doi: 10.3390/ijms22010073. PMID: 33374787; PMCID: PMC7796221.
67. Kojima A, Prehn RT. Genetic susceptibility to post-thymectomy autoimmune diseases in mice. *Immunogenetics.* 1981;14(1-2):15-27. doi: 10.1007/BF00344296. PMID: 7035348.
68. Rajendeeran A, Tenbrock K. Regulatory T cell function in autoimmune disease. *J Transl Autoimmun.* 2021 Oct 30;4:100130. doi: 10.1016/j.jtauto.2021.100130. PMID: 35005594; PMCID: PMC8716637.

69. Bonomo A, Kehn PJ, Shevach EM. Post-thymectomy autoimmunity: abnormal T-cell homeostasis. *Immunol Today*. 1995 Feb;16(2):61-7. doi: 10.1016/0167-5699(95)80089-1. PMID: 7888068.
70. Asano M, Toda M, Sakaguchi N, Sakaguchi S. Autoimmune disease as a consequence of developmental abnormality of a T cell subpopulation. *J Exp Med*. 1996 Aug 1;184(2):387-96. doi: 10.1084/jem.184.2.387. PMID: 8760792; PMCID: PMC2192701.
71. Kojima A, Tanaka-Kojima Y, Sakakura T, Nishizuka Y. Spontaneous development of autoimmune thyroiditis in neonatally thymectomized mice. *Lab Invest*. 1976 Jun;34(6):550-7. PMID: 947081.
72. Taguchi O, Nishizuka Y, Sakakura T, Kojima A. Autoimmune oophoritis in thymectomized mice: detection of circulating antibodies against oocytes. *Clin Exp Immunol*. 1980 Jun;40(3):540-53. PMID: 6998618; PMCID: PMC1538942.
73. Tung KS, Smith S, Teuscher C, Cook C, Anderson RE. Murine autoimmune oophoritis, epididymoorchitis, and gastritis induced by day 3 thymectomy. *Immunopathology. Am J Pathol*. 1987 Feb;126(2):293-302. PMID: 3548402; PMCID: PMC1899560.
74. Anderson MS, Venanzi ES, Klein L, Chen Z, Berzins SP, Turley SJ, von Boehmer H, Bronson R, Dierich A, Benoist C, Mathis D. Projection of an immunological self shadow within the thymus by the aire protein. *Science*. 2002 Nov 15;298(5597):1395-401. doi: 10.1126/science.1075958. Epub 2002 Oct 10. PMID: 12376594.
75. Ramsey C, Winqvist O, Puhakka L, Halonen M, Moro A, Kämpe O, Eskelin P, Pelto-Huikko M, Peltonen L. Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. *Hum Mol Genet*. 2002 Feb 15;11(4):397-409. doi: 10.1093/hmg/11.4.397. PMID: 11854172.
76. Kuroda N, Mitani T, Takeda N, Ishimaru N, Arakaki R, Hayashi Y, Bando Y, Izumi K, Takahashi T, Nomura T, Sakaguchi S, Ueno T, Takahama Y, Uchida D, Sun S, Kajiura F, Mouri Y, Han H, Matsushima A, Yamada G, Matsumoto M. Development of autoimmunity against transcriptionally unrepressed target antigen in the thymus of Aire-deficient mice. *J Immunol*. 2005 Feb 15;174(4):1862-70. doi: 10.4049/jimmunol.174.4.1862. PMID: 15699112.
77. Hubert FX, Kinkel SA, Crewther PE, Cannon PZ, Webster KE, Link M, Uibo R, O'Bryan MK, Meager A, Forehan SP, Smyth GK, Mittaz L, Antonarakis SE, Peterson P, Heath WR, Scott HS. Aire-deficient C57BL/6 mice mimicking the common human 13-base pair deletion mutation present with only a mild autoimmune phenotype. *J Immunol*. 2009 Mar 15;182(6):3902-18. doi: 10.4049/jimmunol.0802124. PMID: 19265170.
78. Ossart J, Moreau A, Autrusseau E, Ménoret S, Martin JC, Besnard M, Ouisse LH, Tesson L, Flippe L, Kisand K, Peterson P, Hubert FX, Anegon I, Josien R, Guillonnet C. Breakdown of Immune Tolerance in AIRE-Deficient Rats Induces a Severe Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy-like Autoimmune Disease. *J Immunol*. 2018 Aug 1;201(3):874-887. doi: 10.4049/jimmunol.1701318. Epub 2018 Jun 29. PMID: 29959280.

79. Geurts AM, Cost GJ, Rémy S, Cui X, Tesson L, Usal C, Ménoret S, Jacob HJ, Anegon I, Buelow R. Generation of gene-specific mutated rats using zinc-finger nucleases. *Methods Mol Biol.* 2010;597:211-25. doi: 10.1007/978-1-60327-389-3\_15. PMID: 20013236.
80. Chen FY, Lee A, Ge S, Nathan S, Knox SM, McNamara NA. Aire-deficient mice provide a model of corneal and lacrimal gland neuropathy in Sjögren's syndrome. *PLoS One.* 2017 Sep 19;12(9):e0184916. doi: 10.1371/journal.pone.0184916. PMID: 28926640; PMCID: PMC5605119.
81. Kurisaki H, Nagao Y, Nagafuchi S, Mitsuyama M. Autoimmune gastro-pancreatitis with anti-protein disulfide isomerase-associated 2 autoantibody in Aire-deficient BALB/cAnN mice. *PLoS One.* 2013 Aug 26;8(8):e73862. doi: 10.1371/journal.pone.0073862. PMID: 23991207; PMCID: PMC3753263.
82. Maglakelidze N, Gao T, Feehan RP, Hobbs RP. AIRE Deficiency Leads to the Development of Alopecia Areata–Like Lesions in Mice. *J Invest Dermatol.* 2023 Apr;143(4):578-587.e3. doi: 10.1016/j.jid.2022.09.656. Epub 2022 Oct 19. PMID: 36270546.
83. Campbell IK, Kinkel SA, Drake SF, van Nieuwenhuijze A, Hubert FX, Tarlinton DM, Heath WR, Scott HS, Wicks IP. Autoimmune regulator controls T cell help for pathogenetic autoantibody production in collagen-induced arthritis. *Arthritis Rheum.* 2009 Jun;60(6):1683-93. doi: 10.1002/art.24501. PMID: 19479827.
84. van den Broek T, Madi A, Delemarre EM, Schadenberg AWL, Tesselaar K, Borghans JAM, Nierkens S, Redegeld FA, Otten HG, Rossetti M, Albani S, Sorek R, Cohen IR, Jansen NJG, van Wijk F. Human neonatal thymectomy induces altered B-cell responses and autoreactivity. *Eur J Immunol.* 2017 Nov;47(11):1970-1981. doi: 10.1002/eji.201746971. Epub 2017 Sep 6. PMID: 28691750; PMCID: PMC5697610.
85. Cavalcanti NV, Palmeira P, Jatene MB, de Barros Dorna M, Carneiro-Sampaio M. Early Thymectomy Is Associated With Long-Term Impairment of the Immune System: A Systematic Review. *Front Immunol.* 2021 Nov 25;12:774780. doi: 10.3389/fimmu.2021.774780. PMID: 34899730; PMCID: PMC8656688.
86. Puck JM. Laboratory technology for population-based screening for severe combined immunodeficiency in neonates: the winner is T-cell receptor excision circles. *J Allergy Clin Immunol.* 2012 Mar;129(3):607-16. doi: 10.1016/j.jaci.2012.01.032. Epub 2012 Jan 29. PMID: 22285280; PMCID: PMC3294074.
87. Hazenberg MD, Verschuren MC, Hamann D, Miedema F, van Dongen JJ. T cell receptor excision circles as markers for recent thymic emigrants: basic aspects, technical approach, and guidelines for interpretation. *J Mol Med (Berl).* 2001 Nov;79(11):631-40. doi: 10.1007/s001090100271. PMID: 11715066.
88. Kooshesh KA, Foy BH, Sykes DB, Gustafsson K, Scadden DT. Health Consequences of Thymus Removal in Adults. *N Engl J Med.* 2023 Aug 3;389(5):406-417. doi: 10.1056/NEJMoa2302892. PMID: 37530823; PMCID: PMC10557034.

89. Bruserud Ø, Oftedal BE, Wolff AB, Husebye ES. AIRE-mutations and autoimmune disease. *Curr Opin Immunol*. 2016 Dec;43:8-15. doi: 10.1016/j.coi.2016.07.003. Epub 2016 Aug 6. PMID: 27504588.
90. Ferré EMN, Schmitt MM, Lionakis MS. Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy. *Front Pediatr*. 2021 Nov 1;9:723532. doi: 10.3389/fped.2021.723532. PMID: 34790633; PMCID: PMC8591095.
91. Vitozzi S, Correa SG, Lozano A, Fernández EJ, Quiroga R. A novel missense mutation in the AIRE gene underlying autoimmune polyglandular syndrome type 1. *Immunogenetics*. 2024 Feb;76(1):69-74. doi: 10.1007/s00251-023-01324-6. Epub 2023 Nov 30. PMID: 38030802.
92. Thorpe, E. S. (1929). Chronic tetany and chronic mycelial stomatitis in a child aged four and one-half years. *Arch. Pediatr. Adolesc. Med.* 38, 328. doi:10.1001/archpedi.1929.01930080104011
93. Perheentupa, J. (1980). Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) (ed. A. W. Eriksson, H. R. Forsius, H. R. Nevanlinna, P. L. Workm and R. K. Norio), pp. 583-588. *Popul. Struct. Genet. Disord. Lond. Acad. Press*.
94. Gatta E, Anelli V, Cimino E, Di Lodovico E, Piovani E, Zammarchi I, Gozzoli G, Maltese V, Cavadini M, Agosti B, Delbarba A, Pirola I, Girelli A, Buoso C, Bambini F, Alfieri D, Bremi W, Facondo P, Lupo R, Bezzi F, Fredi M, Mazzola AM, Gandossi E, Saullo M, Marini F, Licini M, Pezzaioli LC, Pini L, Franceschini F, Ricci C, Cappelli C. Autoimmune polyglandular syndrome type 4: experience from a single reference center. *Front Endocrinol (Lausanne)*. 2023 Oct 23;14:1236878. doi: 10.3389/fendo.2023.1236878. PMID: 37937054; PMCID: PMC10627240.
95. Lovewell TR, McDonagh AJ, Messenger AG, Azzouz M, Tazi-Ahnini R. The AIRE - 230Y Polymorphism Affects AIRE Transcriptional Activity: Potential Influence on AIRE Function in the Thymus. *PLoS One*. 2015 May 15;10(5):e0127476. doi: 10.1371/journal.pone.0127476. PMID: 25978041; PMCID: PMC4433237.
96. Tazi-Ahnini R, Cork MJ, Gawkrödger DJ, Birch MP, Wengraf D, McDonagh AJ, Messenger AG. Role of the autoimmune regulator (AIRE) gene in alopecia areata: strong association of a potentially functional AIRE polymorphism with alopecia universalis. *Tissue Antigens*. 2002 Dec;60(6):489-95. doi: 10.1034/j.1399-0039.2002.600604.x. PMID: 12542742.
97. Ferrera F, Rizzi M, Spreccacenero B, Balestra P, Sessarego M, Di Carlo A, Filaci G, Gabrielli A, Ravazzolo R, Indiveri F. AIRE gene polymorphisms in systemic sclerosis associated with autoimmune thyroiditis. *Clin Immunol*. 2007 Jan;122(1):13-7. doi: 10.1016/j.clim.2006.09.013. Epub 2006 Nov 13. PMID: 17101293.
98. Tazi-Ahnini R, McDonagh AJ, Wengraf DA, Lovewell TR, Vasilopoulos Y, Messenger AG, Cork MJ, Gawkrödger DJ. The autoimmune regulator gene (AIRE) is strongly associated with vitiligo. *Br J Dermatol*. 2008 Sep;159(3):591-6. doi: 10.1111/j.1365-2133.2008.08718.x. Epub 2008 Jul 4. PMID: 18616774.



99. Conteduca G, Ferrera F, Pastorino L, Fenoglio D, Negrini S, Sormani MP, Indiveri F, Scarrà GB, Filaci G. The role of AIRE polymorphisms in melanoma. *Clin Immunol.* 2010 Jul;136(1):96-104. doi: 10.1016/j.clim.2010.03.002. Epub 2010 Apr 2. PMID: 20363194.
100. Zhang X, Ding XJ, Wang Q, Yue YX, Xie Y, Hao HJ, Liang B, Zhang XJ, Song M, Gao Z, Jiang P, Qin Y, Li HF. Rs3761389 polymorphism in autoimmune regulator (AIRE) gene is associated with susceptibility of myasthenia gravis in Chinese patients. *J Clin Neurosci.* 2017 Jun;40:180-184. doi: 10.1016/j.jocn.2017.02.049. Epub 2017 Mar 3. PMID: 28262400.
101. Eriksson D, Bianchi M, Landegren N, Dalin F, Skov J, Hultin-Rosenberg L, Mathioudaki A, Nordin J, Hallgren Å, Andersson G, Tandre K, Rantapää Dahlqvist S, Söderkvist P, Rönnblom L, Hulting AL, Wahlberg J, Dahlqvist P, Ekwall O, Meadows JRS, Lindblad-Toh K, Bensing S, Rosengren Pielberg G, Kämpe O. Common genetic variation in the autoimmune regulator (AIRE) locus is associated with autoimmune Addison's disease in Sweden. *Sci Rep.* 2018 May 30;8(1):8395. doi: 10.1038/s41598-018-26842-2. PMID: 29849176; PMCID: PMC5976627.
102. Alghamdi SA, Kattan SW, Toraih EA, Alrowaili MG, Fawzy MS, Elshazli RM. Association of AIRE (rs2075876), but not CTLA4 (rs231775) polymorphisms with systemic lupus erythematosus. *Gene.* 2021 Feb 5;768:145270. doi: 10.1016/j.gene.2020.145270. Epub 2020 Oct 26. PMID: 33122082.
103. Abdel Ghafar MT, Elshora OA, Allam AA, Mashaal RG, Hamous SAA, Abd El-Khalik SR, Abd-Ellatif RN, Mariah RA, Eissa R, Mwafy M, Shalaby RE, Nasif E, Elkholy RA. Autoimmune Regulator Gene Polymorphisms and the Risk of Primary Immune Thrombocytopenic Purpura: A Case-Control Study. *Int J Mol Sci.* 2023 Mar 5;24(5):5007. doi: 10.3390/ijms24055007. PMID: 36902438; PMCID: PMC10003684.
104. Terao C, Yamada R, Ohmura K, Takahashi M, Kawaguchi T, Kochi Y; Human Disease Genomics Working Group; RA Clinical and Genetic Study Consortium; Okada Y, Nakamura Y, Yamamoto K, Melchers I, Lathrop M, Mimori T, Matsuda F. The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Hum Mol Genet.* 2011 Jul 1;20(13):2680-5. doi: 10.1093/hmg/ddr161. Epub 2011 Apr 19. PMID: 21505073.
105. Bérczi B, Gerencsér G, Farkas N, Hegyi P, Veres G, Bajor J, Czopf L, Alizadeh H, Rakonczay Z, Vigh É, Eröss B, Szemes K, Gyöngyi Z. Association between AIRE gene polymorphism and rheumatoid arthritis: a systematic review and meta-analysis of case-control studies. *Sci Rep.* 2017 Oct 26;7(1):14096. doi: 10.1038/s41598-017-14375-z. PMID: 29074995; PMCID: PMC5658331.
106. Mellado M, Martínez-Muñoz L, Cascio G, Lucas P, Pablos JL, Rodríguez-Frade JM. T Cell Migration in Rheumatoid Arthritis. *Front Immunol.* 2015 Jul 27;6:384. doi: 10.3389/fimmu.2015.00384. PMID: 26284069; PMCID: PMC4515597.
107. Pouya Entezami, BS, David A. Fox, Philip J. Clapham and Kevin C. Chung. Historical Perspective on the Etiology of Rheumatoid Arthritis. *Hand Clin.* 2011 Feb; 27(1): 1–10. doi: 10.1016/j.hcl.2010.09.006

108. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. *Lancet*. 2016 Oct 22;388(10055):2023-2038. doi: 10.1016/S0140-6736(16)30173-8. Epub 2016 May 3. Erratum in: *Lancet*. 2016 Oct 22;388(10055):1984. PMID: 27156434.
109. Klareskog L, Rönnelid J, Saevarsdottir S, Padyukov L, Alfredsson L. The importance of differences; On environment and its interactions with genes and immunity in the causation of rheumatoid arthritis. *J Intern Med*. 2020 May;287(5):514-533. doi: 10.1111/joim.13058. PMID: 32176395.
110. Sundström B, Johansson I, Rantapää-Dahlqvist S. Interaction between dietary sodium and smoking increases the risk for rheumatoid arthritis: results from a nested case-control study. *Rheumatology (Oxford)*. 2015 Mar;54(3):487-93. doi: 10.1093/rheumatology/keu330. Epub 2014 Sep 10. PMID: 25209067.
111. Ilar A, Klareskog L, Saevarsdottir S, Wiebert P, Askling J, Gustavsson P, Alfredsson L. Occupational exposure to asbestos and silica and risk of developing rheumatoid arthritis: findings from a Swedish population-based case-control study. *RMD Open*. 2019 Jul 11;5(2):e000978. doi: 10.1136/rmdopen-2019-000978. PMID: 31413868; PMCID: PMC6667973.
112. Gough J, Rivers D, Seal RM. Pathological studies of modified pneumoconiosis in coal-miners with rheumatoid arthritis; Caplan's syndrome. *Thorax*. 1955 Mar;10(1):9-18. doi: 10.1136/thx.10.1.9. PMID: 14373655; PMCID: PMC1019442.
113. Hart JE, Källberg H, Laden F, Bellander T, Costenbader KH, Holmqvist M, Klareskog L, Alfredsson L, Karlson EW. Ambient air pollution exposures and risk of rheumatoid arthritis: results from the Swedish EIRA case-control study. *Ann Rheum Dis*. 2013 Jun;72(6):888-94. doi: 10.1136/annrheumdis-2012-201587. Epub 2012 Jul 24. PMID: 22833374; PMCID: PMC3654032.
114. Chang KH, Hsu CC, Muo CH, Hsu CY, Liu HC, Kao CH, Chen CY, Chang MY, Hsu YC. Air pollution exposure increases the risk of rheumatoid arthritis: A longitudinal and nationwide study. *Environ Int*. 2016 Sep;94:495-499. doi: 10.1016/j.envint.2016.06.008. Epub 2016 Jun 11. PMID: 27302847.
115. Pedersen M, Jacobsen S, Klarlund M, Frisch M. Socioeconomic status and risk of rheumatoid arthritis: a Danish case-control study. *J Rheumatol*. 2006 Jun;33(6):1069-74. Epub 2006 Apr 15. PMID: 16622905.
116. Cutolo M, Gotelli E. Complex role of oestrogens in the risk and severity of rheumatoid arthritis in menopause. *RMD Open*. 2023 Jun;9(2):e003176. doi: 10.1136/rmdopen-2023-003176. PMID: 37321667; PMCID: PMC10277547.
117. Arleevskaya MI, Kravtsova OA, Lemerle J, Renaudineau Y, Tsibulkin AP. How Rheumatoid Arthritis Can Result from Provocation of the Immune System by Microorganisms and Viruses. *Front Microbiol*. 2016 Aug 17;7:1296. doi: 10.3389/fmicb.2016.01296. PMID: 27582741; PMCID: PMC4987382.

118. Ebringer A, Rashid T. Rheumatoid arthritis is caused by Proteus: the molecular mimicry theory and Karl Popper. *Front Biosci (Elite Ed)*. 2009 Jun 1;1(2):577-86. doi: 10.2741/e56. PMID: 19482674.
119. Díaz-Peña R, Quiñones LA, Castro-Santos P, Durán J, Lucia A. Latin American Genes: The Great Forgotten in Rheumatoid Arthritis. *J Pers Med*. 2020 Oct 26;10(4):196. doi: 10.3390/jpm10040196. PMID: 33114702; PMCID: PMC7711650.
120. Huizinga TW, Amos CI, van der Helm-van Mil AH, Chen W, van Gaalen FA, Jawaheer D, Schreuder GM, Wener M, Breedveld FC, Ahmad N, Lum RF, de Vries RR, Gregersen PK, Toes RE, Criswell LA. Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. *Arthritis Rheum*. 2005 Nov;52(11):3433-8. doi: 10.1002/art.21385. PMID: 16255021.
121. MacKay K, Eyre S, Myerscough A, Milicic A, Barton A, Laval S, Barrett J, Lee D, White S, John S, Brown MA, Bell J, Silman A, Ollier W, Wordsworth P, Worthington J. Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum*. 2002 Mar;46(3):632-9. doi: 10.1002/art.10147. Erratum in: *Arthritis Rheum* 2002 May;46(5):1406. PMID: 11920398.
122. Kurkó J, Besenyei T, Laki J, Glant TT, Mikecz K, Szekanecz Z. Genetics of rheumatoid arthritis - a comprehensive review. *Clin Rev Allergy Immunol*. 2013 Oct;45(2):170-9. doi: 10.1007/s12016-012-8346-7. PMID: 23288628; PMCID: PMC3655138.
123. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Kawaguchi T, Stahl EA, Kurreeman FA, Nishida N, Ohmiya H, Myouzen K, Takahashi M, Sawada T, Nishioka Y, Yukioka M, Matsubara T, Wakitani S, Teshima R, Tohma S, Takasugi K, Shimada K, Murasawa A, Honjo S, Matsuo K, Tanaka H, Tajima K, Suzuki T, Iwamoto T, Kawamura Y, Tanii H, Okazaki Y, Sasaki T, Gregersen PK, Padyukov L, Worthington J, Siminovitch KA, Lathrop M, Taniguchi A, Takahashi A, Tokunaga K, Kubo M, Nakamura Y, Kamatani N, Mimori T, Plenge RM, Yamanaka H, Momohara S, Yamada R, Matsuda F, Yamamoto K. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet*. 2012 Mar 25;44(5):511-6. doi: 10.1038/ng.2231. PMID: 22446963.
124. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, Graham RR, Manoharan A, Ortmann W, Bhangale T, Denny JC, Carroll RJ, Eyler AE, Greenberg JD, Kremer JM, Pappas DA, Jiang L, Yin J, Ye L, Su DF, Yang J, Xie G, Keystone E, Westra HJ, Esko T, Metspalu A, Zhou X, Gupta N, Mirel D, Stahl EA, Diogo D, Cui J, Liao K, Guo MH, Myouzen K, Kawaguchi T, Coenen MJ, van Riel PL, van de Laar MA, Guchelaar HJ, Huizinga TW, Dieudé P, Mariette X, Bridges SL Jr, Zhernakova A, Toes RE, Tak PP, Miceli-Richard C, Bang SY, Lee HS, Martin J, Gonzalez-Gay MA, Rodriguez-Rodriguez L, Rantapää-Dahlqvist S, Arlestig L, Choi HK, Kamatani Y, Galan P, Lathrop M; RACI consortium; GARNET consortium; Eyre S, Bowes J, Barton A, de Vries N, Moreland LW, Criswell LA, Karlson EW, Taniguchi A, Yamada R, Kubo M, Liu JS, Bae SC, Worthington J, Padyukov L, Klareskog L, Gregersen PK, Raychaudhuri S, Stranger BE, De Jager PL, Franke L, Visscher PM, Brown MA, Yamanaka H, Mimori T, Takahashi A, Xu H, Behrens TW, Siminovitch KA, Momohara S, Matsuda F, Yamamoto K, Plenge RM. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014 Feb 20;506(7488):376-81. doi: 10.1038/nature12873. Epub 2013 Dec 25. PMID: 24390342; PMCID: PMC3944098.

125. Ishigaki K, Sakaue S, Terao C, Luo Y, Sonehara K, Yamaguchi K, Amariuta T, Too CL, Laufer VA, Scott IC, Viatte S, Takahashi M, Ohmura K, Murasawa A, Hashimoto M, Ito H, Hammoudeh M, Emadi SA, Masri BK, Halabi H, Badsha H, Uthman IW, Wu X, Lin L, Li T, Plant D, Barton A, Orozco G, Verstappen SMM, Bowes J, MacGregor AJ, Honda S, Koido M, Tomizuka K, Kamatani Y, Tanaka H, Tanaka E, Suzuki A, Maeda Y, Yamamoto K, Miyawaki S, Xie G, Zhang J, Amos CI, Keystone E, Wolbink G, van der Horst-Bruinsma I, Cui J, Liao KP, Carroll RJ, Lee HS, Bang SY, Siminovitch KA, de Vries N, Alfredsson L, Rantapää-Dahlqvist S, Karlson EW, Bae SC, Kimberly RP, Edberg JC, Mariette X, Huizinga T, Dieudé P, Schneider M, Kerick M, Denny JC; BioBank Japan Project; Matsuda K, Matsuo K, Mimori T, Matsuda F, Fujio K, Tanaka Y, Kumanogoh A, Traylor M, Lewis CM, Eyre S, Xu H, Saxena R, Arayssi T, Kochi Y, Ikari K, Harigai M, Gregersen PK, Yamamoto K, Louis Bridges S Jr, Padyukov L, Martin J, Klareskog L, Okada Y, Raychaudhuri S. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat Genet.* 2022 Nov;54(11):1640-1651. doi: 10.1038/s41588-022-01213-w. Epub 2022 Nov 4. PMID: 36333501; PMCID: PMC10165422.
126. Weyand CM, Goronzy JJ. The immunology of rheumatoid arthritis. *Nat Immunol.* 2021 Jan;22(1):10-18. doi: 10.1038/s41590-020-00816-x. Epub 2020 Nov 30. PMID: 33257900; PMCID: PMC8557973.
127. Makrygiannakis D, Hermansson M, Ulfgren AK, Nicholas AP, Zendman AJ, Eklund A, Grunewald J, Skold CM, Klareskog L, Catrina AI. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Ann Rheum Dis.* 2008 Oct;67(10):1488-92. doi: 10.1136/ard.2007.075192. Epub 2008 Apr 15. PMID: 18413445.
128. Guo Q, Wang Y, Xu D, Nossent J, Pavlos NJ, Xu J. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. *Bone Res.* 2018 Apr 27;6:15. doi: 10.1038/s41413-018-0016-9. PMID: 29736302; PMCID: PMC5920070.
129. Yap HY, Tee SZ, Wong MM, Chow SK, Peh SC, Teow SY. Pathogenic Role of Immune Cells in Rheumatoid Arthritis: Implications in Clinical Treatment and Biomarker Development. *Cells.* 2018 Oct 9;7(10):161. doi: 10.3390/cells7100161. PMID: 30304822; PMCID: PMC6211121.
130. Finckh A, Gilbert B, Hodkinson B, Bae SC, Thomas R, Deane KD, Alpizar-Rodriguez D, Lauper K. Global epidemiology of rheumatoid arthritis. *Nat Rev Rheumatol.* 2022 Oct;18(10):591-602. doi: 10.1038/s41584-022-00827-y. Epub 2022 Sep 6. PMID: 36068354.
131. Myasoedova E, Davis J, Matteson EL, Crowson CS. Is the epidemiology of rheumatoid arthritis changing? Results from a population-based incidence study, 1985-2014. *Ann Rheum Dis.* 2020 Apr;79(4):440-444. doi: 10.1136/annrheumdis-2019-216694. Epub 2020 Feb 17. PMID: 32066556; PMCID: PMC7085464.
132. Minichiello E, Semerano L, Boissier MC. Time trends in the incidence, prevalence, and severity of rheumatoid arthritis: A systematic literature review. *Joint Bone Spine.* 2016 Dec;83(6):625-630. doi: 10.1016/j.jbspin.2016.07.007. Epub 2016 Sep 5. PMID: 27616690.

133. Abhishek A, Doherty M, Kuo CF, Mallen CD, Zhang W, Grainge MJ. Rheumatoid arthritis is getting less frequent-results of a nationwide population-based cohort study. *Rheumatology (Oxford)*. 2017 May 1;56(5):736-744. doi: 10.1093/rheumatology/kew468. PMID: 28064207; PMCID: PMC5850292.
134. Elmer D, Endrei D, Sebestyén A, Csákvári T, Németh N, Horváth L, Pónusz R, Kumánovics G, Boncz I, Ágoston I. A rheumatoid arthritis okozta országos epidemiológiai és egészségbiztosítási betegségteher Magyarországon [Nationwide epidemiological and health insurance disease burden of rheumatoid arthritis in Hungary]. *Orv Hetil*. 2021 Mar 28;162(162 Suppl 1):30-37. Hungarian. doi: 10.1556/650.2021.32157. PMID: 33774606.
135. Kiss CG, Lövei C, Sütö G, Varjú C, Nagy Z, Füzesi Z, Illés T, Czirják L. Prevalence of rheumatoid arthritis in the South-Transdanubian region of Hungary based on a representative survey of 10,000 inhabitants. *J Rheumatol*. 2005 Sep;32(9):1688-90. PMID: 16142861.
136. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum*. 1988 Mar;31(3):315-24. doi: 10.1002/art.1780310302. PMID: 3358796.
137. Graham JE, Granger CV, Karmarkar AM, Deutsch A, Niewczyk P, Divita MA, Ottenbacher KJ. The Uniform Data System for Medical Rehabilitation: report of follow-up information on patients discharged from inpatient rehabilitation programs in 2002-2010. *Am J Phys Med Rehabil*. 2014 Mar;93(3):231-44. doi: 10.1097/PHM.0b013e3182a92c58. PMID: 24088780; PMCID: PMC3944381.
138. Brown PM, Pratt AG, Isaacs JD. Mechanism of action of methotrexate in rheumatoid arthritis, and the search for biomarkers. *Nat Rev Rheumatol*. 2016 Dec;12(12):731-742. doi: 10.1038/nrrheum.2016.175. Epub 2016 Oct 27. PMID: 27784891.
139. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015 Jan 1;4(1):1. doi: 10.1186/2046-4053-4-1. PMID: 25554246; PMCID: PMC4320440.
140. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. *Contemp Clin Trials*. 2015 Nov;45(Pt A):139-45. doi: 10.1016/j.cct.2015.09.002. Epub 2015 Sep 4. PMID: 26343745; PMCID: PMC4639420.
141. The Cochrane Collaboration *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (eds. Higgins, J. P. T. & Green, S.; John Wiley & Sons, 2011)
142. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PLoS One*. 2013;8(3):e59202. doi: 10.1371/journal.pone.0059202. Epub 2013 Mar 27. PMID: 23544056; PMCID: PMC3609745.
143. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol*. 2008 Jan;61(1):64-75. doi: 10.1016/j.jclinepi.2007.03.013. Epub 2007 Aug 23. PMID: 18083463.

144. Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive--Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *Int J Epidemiol*. 2009 Feb;38(1):287-98. doi: 10.1093/ije/dyn188. Epub 2008 Sep 29. PMID: 18824466.
145. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. *BMC Med Res Methodol*. 2009 Dec 30;9:86. doi: 10.1186/1471-2288-9-86. PMID: 20042080; PMCID: PMC2809074.
146. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res*. 2002 Jun;12(6):996-1006. doi: 10.1101/gr.229102. PMID: 12045153; PMCID: PMC186604.
147. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, Berry A, Bhai J, Bignell A, Billis K, Boddu S, Brooks L, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, Donaldson S, El Houdaigui B, El Naboulsi T, Fatima R, Giron CG, Genev T, Martinez JG, Gujjarro-Clarke C, Gymer A, Hardy M, Hollis Z, Hourlier T, Hunt T, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Marugán JC, Mohanan S, Mushtaq A, Naven M, Ogeh DN, Parker A, Parton A, Perry M, Piližota I, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Pérez-Silva JG, Stark W, Steed E, Sutinen K, Sukumaran R, Sumathipala D, Suner MM, Szpak M, Thormann A, Tricomi FF, Urbina-Gómez D, Veidenberg A, Walsh TA, Walts B, Willhoft N, Winterbottom A, Wass E, Chakiachvili M, Flint B, Frankish A, Giorgetti S, Haggerty L, Hunt SE, Iisley GR, Loveland JE, Martin FJ, Moore B, Mudge JM, Muffato M, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Dyer S, Harrison PW, Howe KL, Yates AD, Zerbino DR, Flicek P. Ensembl 2022. *Nucleic Acids Res*. 2022 Jan 7;50(D1):D988-D995. doi: 10.1093/nar/gkab1049. PMID: 34791404; PMCID: PMC8728283.
148. Cavalli-Sforza, L.L.; Bodmer, W.F. *The genetics of human populations*, 1st edition; Dover Publications, San Francisco, USA, 1971, 41-118.
149. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res*. 1999 Aug;9(8):677-9. PMID: 10447503.
150. Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res*. 2020 Jan 8;48(D1):D941-D947. doi: 10.1093/nar/gkz836. PMID: 31584097; PMCID: PMC6943028.
151. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts

K, Tolonen C, Wade G, Talkowski ME; Genome Aggregation Database Consortium; Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020 May;581(7809):434-443. doi: 10.1038/s41586-020-2308-7. Epub 2020 May 27. Erratum in: *Nature*. 2021 Feb;590(7846):E53. Erratum in: *Nature*. 2021 Sep;597(7874):E3-E4. PMID: 32461654; PMCID: PMC7334197.

152. Berczi B, Nusser N, Peter I, Nemeth B, Gyongyi Z. Association Between AIRE Polymorphisms rs870881(C>T), rs1003854(T>C) and Rheumatoid Arthritis Risk: A Hungarian Case-control Study. *In Vivo*. 2024 Mar-Apr;38(2):774-784. doi: 10.21873/invivo.13501. PMID: 38418155; PMCID: PMC10905445.

153. Ingegnoli F, Castelli R, Gualtierotti R. Rheumatoid factors: clinical applications. *Dis Markers*. 2013;35(6):727-34. doi: 10.1155/2013/726598. Epub 2013 Nov 13. PMID: 24324289; PMCID: PMC3845430.

154. Steiner G, Smolen J. Autoantibodies in rheumatoid arthritis and their clinical significance. *Arthritis Res*. 2002;4 Suppl 2(Suppl 2):S1-5. doi: 10.1186/ar551. Epub 2002 Apr 26. PMID: 12110150; PMCID: PMC3238219.

155. Niewold TB, Harrison MJ, Paget SA. Anti-CCP antibody testing as a diagnostic and prognostic tool in rheumatoid arthritis. *QJM*. 2007 Apr;100(4):193-201. doi: 10.1093/qjmed/hcm015. PMID: 17434910.

156. Jou JM, Lewis SM, Briggs C, Lee SH, De La Salle B, McFadden S; International Council for Standardization in Haematology. ICSH review of the measurement of the erythrocyte sedimentation rate. *Int J Lab Hematol*. 2011 Apr;33(2):125-32. doi: 10.1111/j.1751-553X.2011.01302.x. Epub 2011 Feb 25. PMID: 21352508.

157. Bérczi B, Nusser N, Péter I, Németh B, Kulisch Á, Kiss Z, Gyöngyi Z. Genetic Polymorphisms in Exon 5 and Intron 5 and 7 of AIRE Are Associated with Rheumatoid Arthritis Risk in a Hungarian Population. *Biology*. 2024; 13(6):439. <https://doi.org/10.3390/biology13060439>

158. HaploReg v4.2. Available online:  
<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>

159. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012 Jan;40(Database issue):D930-4. doi: 10.1093/nar/gkr917. Epub 2011 Nov 7. PMID: 22064851; PMCID: PMC3245002.

160. Kheradpour P, Kellis M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res*. 2014 Mar;42(5):2976-87. doi: 10.1093/nar/gkt1249. Epub 2013 Dec 13. PMID: 24335146; PMCID: PMC3950668.

161. Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*. 2003 Jan 1;31(1):374-8. doi: 10.1093/nar/gkg108. PMID: 12520026; PMCID: PMC165555.

162. Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A, Wasserman WW. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D110-5. doi: 10.1093/nar/gkv1176. Epub 2015 Nov 3. PMID: 26531826; PMCID: PMC4702842.
163. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Coburn D, Newburger DE, Morris Q, Hughes TR, Bulyk ML. Diversity and complexity in DNA recognition by transcription factors. *Science.* 2009 Jun 26;324(5935):1720-3. doi: 10.1126/science.1162327. Epub 2009 May 14. PMID: 19443739; PMCID: PMC2905877.
164. Horita N, Kaneko T. Genetic model selection for a case-control study and a meta-analysis. *Meta Gene.* 2015 May 22;5:1-8. doi: 10.1016/j.mgene.2015.04.003. PMID: 26042205; PMCID: PMC4443430.
165. Di Matteo A, Bathon JM, Emery P. Rheumatoid arthritis. *Lancet.* 2023 Nov 25;402(10416):2019-2033. doi: 10.1016/S0140-6736(23)01525-8. Epub 2023 Oct 27. PMID: 38240831.
166. Sharland DE. Erythrocyte sedimentation rate: the normal range in the elderly. *J Am Geriatr Soc.* 1980 Aug;28(8):346-8. doi: 10.1111/j.1532-5415.1980.tb01096.x. PMID: 7400503.
167. Miller A, Green M, Robinson D. Simple rule for calculating normal erythrocyte sedimentation rate. *Br Med J (Clin Res Ed).* 1983 Jan 22;286(6361):266. doi: 10.1136/bmj.286.6361.266. PMID: 6402065; PMCID: PMC1546487.
168. Wener MH, Daum PR, McQuillan GM. The influence of age, sex, and race on the upper reference limit of serum C-reactive protein concentration. *J Rheumatol.* 2000 Oct;27(10):2351-9. PMID: 11036829.
169. Coenen D, Verschueren P, Westhovens R, Bossuyt X. Technical and diagnostic performance of 6 assays for the measurement of citrullinated protein/peptide antibodies in the diagnosis of rheumatoid arthritis. *Clin Chem.* 2007 Mar;53(3):498-504. doi: 10.1373/clinchem.2006.078063. Epub 2007 Jan 26. PMID: 17259232.
170. Fuchs HA, Brooks RH, Callahan LF, Pincus T. A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis. *Arthritis Rheum.* 1989 May;32(5):531-7. doi: 10.1002/anr.1780320504. PMID: 2719728.
171. Lewontin, R. C., & Kojima, K. (1960). The Evolutionary Dynamics of Complex Polymorphisms. *Evolution*, 14(4), 458–472. <https://doi.org/10.2307/2405995>
172. HaploReg v4.2. Available online: <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php> (accessed on 25 March 2024).
173. Bergström B, Lundqvist C, Vasileiadis GK, Carlsten H, Ekwall O, Ekwall AH. The Rheumatoid Arthritis Risk Gene AIRE Is Induced by Cytokines in Fibroblast-Like



- Synoviocytes and Augments the Pro-inflammatory Response. *Front Immunol.* 2019 Jun 18;10:1384. doi: 10.3389/fimmu.2019.01384. PMID: 31275320; PMCID: PMC6591464.
174. Pandya JM, Lundell AC, Andersson K, Nordström I, Theander E, Rudin A. Blood chemokine profile in untreated early rheumatoid arthritis: CXCL10 as a disease activity marker. *Arthritis Res Ther.* 2017 Feb 2;19(1):20. doi: 10.1186/s13075-017-1224-1. PMID: 28148302; PMCID: PMC5289001.
175. Han JH, Suh CH, Jung JY, Nam JY, Kwon JE, Yim H, Kim HA. Association of CXCL10 and CXCL13 levels with disease activity and cutaneous manifestation in active adult-onset Still's disease. *Arthritis Res Ther.* 2015 Sep 19;17(1):260. doi: 10.1186/s13075-015-0773-4. PMID: 26385705; PMCID: PMC4575437.
176. Ruth JH, Rottman JB, Katschke KJ Jr, Qin S, Wu L, LaRosa G, Ponath P, Pope RM, Koch AE. Selective lymphocyte chemokine receptor expression in the rheumatoid joint. *Arthritis Rheum.* 2001 Dec;44(12):2750-60. doi: 10.1002/1529-0131(200112)44:12<2750::aid-art462>3.0.co;2-c. PMID: 11762935.
177. Verheul TCJ, van Hijfte L, Perenthaler E, Barakat TS. The Why of YY1: Mechanisms of Transcriptional Regulation by Yin Yang 1. *Front Cell Dev Biol.* 2020 Sep 30;8:592164. doi: 10.3389/fcell.2020.592164. PMID: 33102493; PMCID: PMC7554316.
178. Cheng H, Zhang N, Pati D. Cohesin subunit RAD21: From biology to disease. *Gene.* 2020 Oct 20;758:144966. doi: 10.1016/j.gene.2020.144966. Epub 2020 Jul 17. PMID: 32687945; PMCID: PMC7949736.
179. Nagamura-Inoue T, Tamura T, Ozato K. Transcription factors that regulate growth and differentiation of myeloid cells. *Int Rev Immunol.* 2001 Feb;20(1):83-105. doi: 10.3109/08830180109056724. PMID: 11342299.
180. Wan YY. GATA3: a master of many trades in immune regulation. *Trends Immunol.* 2014 Jun;35(6):233-42. doi: 10.1016/j.it.2014.04.002. Epub 2014 Apr 28. PMID: 24786134; PMCID: PMC4045638.
181. Rose AB. Intron-mediated regulation of gene expression. *Curr Top Microbiol Immunol.* 2008;326:277-90. doi: 10.1007/978-3-540-76776-3\_15. PMID: 18630758.

## IX. Függelék

<i>AIRE</i> rs2075876 (Intron 7)	RA betegek N=270 (%)	Kontrollok N=322 (%)	OR (95% CI)	p <sup>a</sup>
Allél				
A	59 (10,9)	72 (11,2)		
G	481 (89,1)	572 (88,8)	1,02 (0,71-1,47)	0,890
Genotípus				
AA	3 (1,1)	8 (2,5)		
GA	49 (18,1)	54 (16,8)	2,42 (0,607-9,63)	0,210
GG	218 (80,7)	260 (80,7)	2,23 (0,58-8,53)	0,239
Domináns genetikai modell				
AA	3 (1,1)	8 (2,5)		
GG+GA	267 (98,9)	314 (97,5)	2,26 (0,59-8,63)	0,230
Recesszív genetikai modell				
GA+AA	52 (19,3)	62 (19,3)		
GG	218 (80,7)	260 (80,7)	1,00 (0,66-1,507)	0,999
Over-domináns genetikai modell				
AA+GG	221 (81,9)	268 (83,2)		
GA	49 (18,1)	54 (16,8)	1,100 (0,71-1,86)	0,660

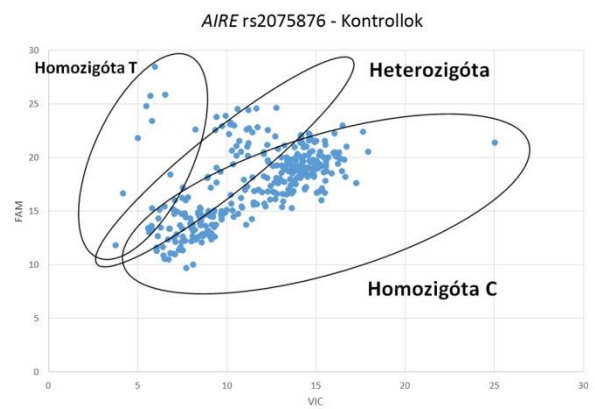
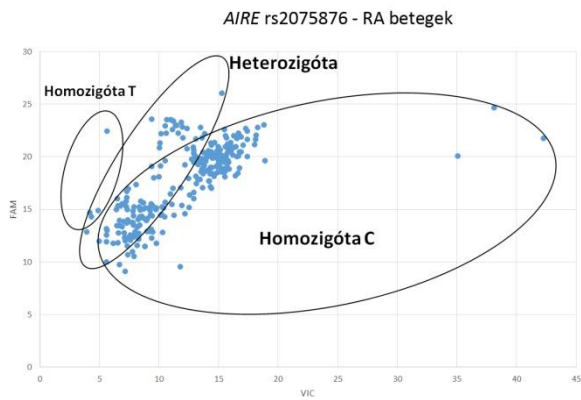
RA: reumatoid arthritisz; OR: odds ratio, esélyhányados; CI: confidence interval, konfidencia intervallum; p<sup>a</sup>: bináris logisztikus regresszió szignifikanciája

### I. melléklet táblázat: Az 1. csoport *AIRE* rs2075876 lókusának allélfrekvenciái és genotípus gyakoriságai, asszociációjuk RA-val

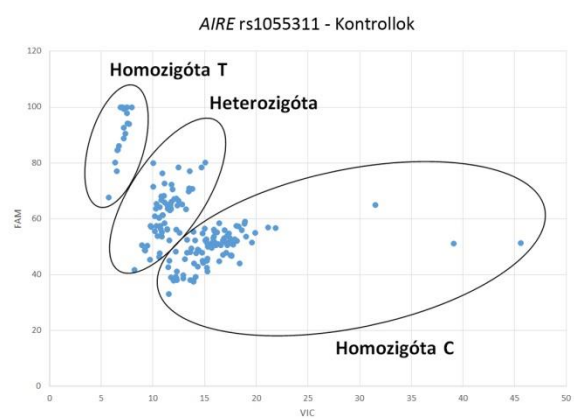
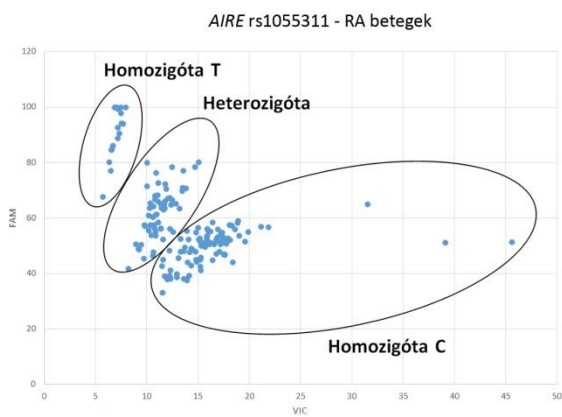
<i>AIRE</i> rs1055311 (Exon 6)	RA betegek N=270 (%)	Kontrollok N=322 (%)	OR (95% CI)	p <sup>a</sup>
Allél				
T	161 (29,8)	184 (28,6)		
C	397 (70,2)	460 (71,4)	0,94 (0,73-1,21)	0,639
Genotípus				
TT	25 (9,3)	24 (7,5)		
CT	111 (41,1)	135 (41,9)	0,798 (0,42-1,458)	0,450
CC	134 (49,6)	162 (50,3)	0,794 (0,43-1,454)	0,455
Domináns genetikai modell				
TT	25 (9,3)	24 (7,5)		
CC+CT	245 (90,7)	298 (92,5)	0,78 (0,44-1,41)	0,428
Recesszív genetikai modell				
CT+TT	136 (50,4)	160 (49,7)		
CC	134 (49,6)	162 (50,3)	0,97 (0,704-1,34)	0,86
Over-domináns modell				
CC+TT	159 (58,9)	186 (57,8)		
CT	111 (41,1)	136 (42,2)	0,95 (0,68-1,32)	0,782

RA: reumatoid arthritisz; OR: odds ratio, esélyhányados; CI: confidence interval, konfidencia intervallum; p<sup>a</sup>: bináris logisztikus regresszió szignifikanciája

### II. melléklet táblázat: Az 1. csoport *AIRE* rs1055311 lókusának allélfrekvenciái és genotípus gyakoriságai, asszociációjuk RA-val



**1. melléklet ábra: AIRE rs2075876 lókuszt alléldiszkriminációs teszteredményei RA és kontrollcsoportban**



**2. melléklet ábra: AIRE rs1055311 lókuszt alléldiszkriminációs teszteredményei RA és kontrollcsoportban**

A dolgozat alapjául szolgáló közlemények:

105. **Bérczi B**, Gerencsér G, Farkas N, Hegyi P, Veres G, Bajor J, Czopf L, Alizadeh H, Rakonczay Z, Vigh É, Erőss B, Szemes K, Gyöngyi Z. Association between AIRE gene polymorphism and rheumatoid arthritis: a systematic review and meta-analysis of case-control studies. *Sci Rep.* 2017 Oct 26;7(1):14096. doi: 10.1038/s41598-017-14375-z. PMID: 29074995; PMCID: PMC5658331.

Folyóirat: Scientific Reports

Kiadó: Nature

Besorolása: D1

IF: 4,6

152. **Bérczi B**, Nusser N, Peter I, Nemeth B, Gyongyi Z. Association Between AIRE Polymorphisms rs870881(C>T), rs1003854(T>C) and Rheumatoid Arthritis Risk: A Hungarian Case-control Study. *In Vivo.* 2024 Mar-Apr;38(2):774-784. doi: 10.21873/invivo.13501. PMID: 38418155; PMCID: PMC10905445.

Folyóirat: In vivo

Kiadó: IJAR Journals

Besorolása: Q2

IF: 2,4

157. **Bérczi B**, Nusser N, Péter I, Németh B, Kulisch Á, Kiss Z, Gyöngyi Z. Genetic Polymorphisms in Exon 5 and Intron 5 and 7 of AIRE Are Associated with Rheumatoid Arthritis Risk in a Hungarian Population. *Biology.* 2024; 13(6):439. <https://doi.org/10.3390/biology13060439>

Folyóirat: Biology


Kiadó: MDPI

Besorolása: Q1

IF: 4,2

Kumulált IF: 11,2

# SCIENTIFIC REPORTS



OPEN

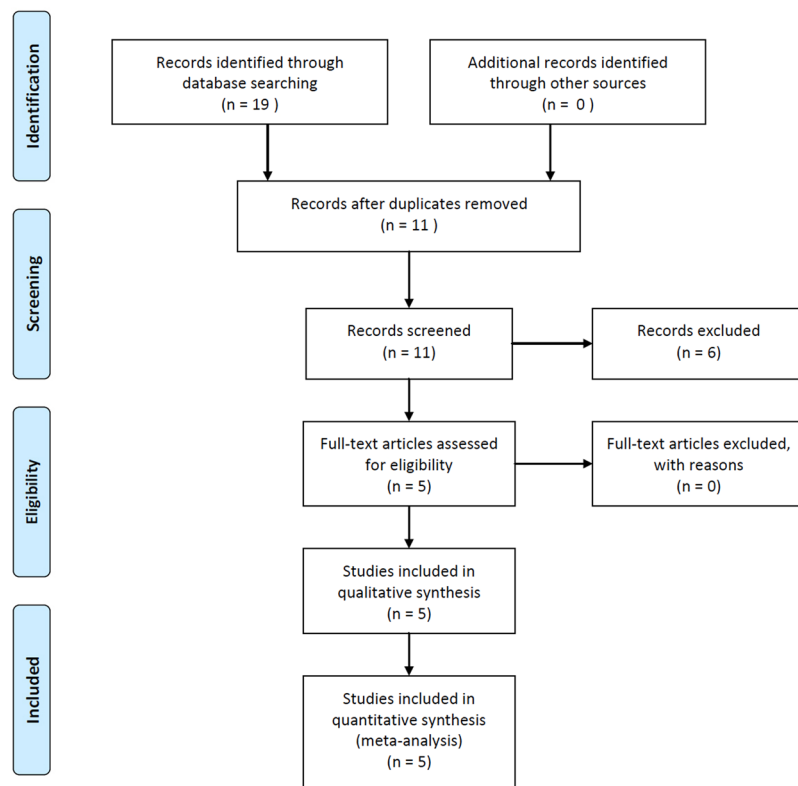
## Association between *AIRE* gene polymorphism and rheumatoid arthritis: a systematic review and meta-analysis of case-control studies

Bálint Bérczi<sup>1</sup>, Gellért Gerencsér<sup>1</sup>, Nelli Farkas<sup>2</sup>, Péter Hegyi<sup>3,4,5</sup>, Gábor Veres<sup>6</sup>, Judit Bajor<sup>7</sup>, László Czopf<sup>8</sup>, Hussain Alizadeh<sup>9</sup>, Zoltán Rakonczay<sup>10</sup>, Éva Vigh<sup>11</sup>, Bálint Erőss<sup>4</sup>, Kata Szemes<sup>7</sup> & Zoltán Gyöngyi<sup>1</sup>

Autoimmune regulator (*AIRE*) is a transcription factor that functions as a novel player in immunological investigations. In the thymus, it has a pivotal role in the negative selection of naive T-cells during central tolerance. Experimental studies have shown that single nucleotide polymorphism (SNP) alters transcription of the *AIRE* gene. SNPs thereby provide a less efficient negative selection, propagate higher survival of autoimmune T-cells, and elevate susceptibility to autoimmune diseases. To date, only rheumatoid arthritis (RA) has been analysed by epidemiological investigations in relation to SNPs in *AIRE*. In our meta-analysis, we sought to encompass case-control studies and confirm that the association between SNP occurrence and RA. After robust searches of Embase, PubMed, Cochrane Library, and Web of Science databases, we found 19 articles that included five independent studies. Out of 11 polymorphisms, two (rs2075876, rs760426) were common in the five case-control studies. Thus, we performed a meta-analysis for rs2075876 (7145 cases and 8579 controls) and rs760426 (6696 cases and 8164 controls). Our results prove that rs2075876 and rs760426 are significantly associated with an increased risk of RA in allelic, dominant, recessive, codominant heterozygous, and codominant homozygous genetic models. These findings are primarily based on data from Asian populations.

Rheumatoid arthritis (RA) is a common autoimmune disease associated with chronic synovial inflammation. The resultant symmetrical polyarticular arthritis, combined with extra-articular complications, leads to functional impairment. In developed countries, disease prevalence is 0.5–1% of the adult population, and the annual incidence has been reported to be 5–50 per 100,000<sup>1</sup>. Although the aetiology of RA has not been completely elucidated, numerous publications agree that autoimmune T-cells may escape from the adaptive immune system and, by migrating to the synovium, initiate disease development<sup>2–5</sup>. RA susceptibility is determined by multiple environmental and genetic factors, including several risk alleles. The latest trans-ethnic genome wide association study (GWAS), which involved 29,980 RA cases and 73,578 controls, completely screened novel polymorphisms in genes contributing to the disease<sup>6</sup>. One of the associated genes that seems to play a pivotal role in controlling autoimmunity is autoimmune regulator (*AIRE*). The gene is located in the 21q22.3 region, is ~12.5 kb long, and

<sup>1</sup>Department of Public Health Medicine, Medical School, University of Pécs, Pécs, Hungary. <sup>2</sup>Institute of Bioanalysis, Medical School, University of Pécs, Pécs, Hungary. <sup>3</sup>MTA-SZTE Translational Gastroenterology Research Group, Szeged, Hungary. <sup>4</sup>Institute for Translational Medicine, University of Pécs, Pécs, Hungary. <sup>5</sup>Department of Translational Medicine, First Department of Medicine, University of Pécs, Pécs, Hungary. <sup>6</sup>1st Department of Pediatrics, Semmelweis University, Budapest, Hungary. <sup>7</sup>Department of Gastroenterology, First Department of Medicine, University of Pécs, Pécs, Hungary. <sup>8</sup>Division of Cardiology and Angiology, First Department of Medicine, University of Pécs, Pécs, Hungary. <sup>9</sup>Department of Haematology, First Department of Medicine, University of Pécs, Pécs, Hungary. <sup>10</sup>Department of Pathophysiology, University of Szeged, Szeged, Hungary. <sup>11</sup>Department of Radiology, Medical School, University of Pécs, Pécs, Hungary. Correspondence and requests for materials should be addressed to Z.G. (email: [zoltan.gyongyi@aok.pte.hu](mailto:zoltan.gyongyi@aok.pte.hu))



**Figure 1.** PRISMA flow diagram for inclusion and exclusion of studies in the meta-analysis.

encodes a 545 amino acid protein of 58 kDa by 14 exonial sequences<sup>7–9</sup>. The AIRE protein is a transcription factor that is indispensable with regards to the negative selection of immature T-cells (thymocytes). Cooperating with DNA-binding proteins, AIRE controls the promiscuous expression of peripheral tissue antigens (PTA). Mutations in the protein coding gene sequence of *AIRE* results in the development of autoimmune polyendocrinopathy candidiasis-ectodermal dystrophy, an autoimmune deterioration of numerous organs<sup>10–12</sup>.

To date, increasing numbers of publications have suggested that SNPs in the gene sequence affect *AIRE* transcription. The SNPs thereby alter the functional activity of AIRE and potentially elevate disease susceptibility<sup>7</sup>. A recent experimental study described two distinct SNPs of *AIRE*. *AIRE*–230Y, and *AIRE*–655G. *AIRE*–230T haplotype transcriptionally modifies *AIRE* expression and influence negative selection, elevating the risk of autoimmunity<sup>13</sup>. Various SNPs in the *AIRE* genetic sequence have garnered attention; however, to date, only a minority of case-control studies have observed an association between gene polymorphism and susceptibility to diseases, including vitiligo<sup>7,14</sup>, alopecia areata<sup>7,15</sup>, melanoma<sup>7,16</sup>, systemic sclerosis<sup>7,17</sup> and RA<sup>7,18–22</sup>. Among the latter diseases, only RA has been analysed by multiple case-control studies and, therefore, seems to be optimal to analyse positive or negative associations<sup>7</sup>. Xu *et al.* have published that *AIRE* polymorphism was associated with the increased risk of RA<sup>23</sup>. Here, we present a systematic review and first meta-analysis that includes case-control studies to verify the association of SNPs rs2075876 and rs760426 in the *AIRE* gene with RA.

## Results

**Characteristics of included studies.** We identified 19 publications after a thorough search of Embase, PubMed, Cochrane Library, and Web of Science databases. After removing duplicates, we reviewed the remaining 11 studies for eligibility and selected five publications for inclusion in our meta-analysis. Our PRISMA flow chart of the searching process is shown in Fig. 1. Asian and Caucasian ethnicities were involved. Diagnosis of RA was determined according to the American College of Rheumatology classification criteria in 1987<sup>24</sup>. The overall mean age of RA patients was  $54.1 \pm 2.4$  years, and the percentage of female cases was 73.34%. Genotyping was conducted by microarrays, single base extension methods (SNaPshot), and Taqman SNP Genotyping Assays. By further reviewing the five eligible publications, we identified 11 SNPs of the *AIRE* gene (rs2075876, rs760426, rs1800250, rs2776377, rs878081, rs1055311, rs933150, rs1003854, rs2256817, rs374696, rs1078480). Only rs2075876 and rs760426 were involved in four or more studies; therefore, we performed meta-analysis for rs2075876 (7145 cases and 8579 controls) and rs760426 (6696 cases and 8164 controls). All genotype frequencies of the controls were in Hardy-Weinberg Equilibrium (HWE). Characteristics of the included studies on rs2075876 and rs760426 are summarized in Table 1.

**Meta-analysis of SNP rs2075876 ( $G > A$ ).** Five studies were identified that investigated the association between SNP rs2075876 and RA susceptibility<sup>18–22</sup>. Most of the publications doubled the individual number to account for alleles; thus, to normalize the data, we also calculated with duplicated values (see Supplementary

		Year	Country	Ethnicity	Diagnostic criteria	Genotyping	Mean age		Female %		Control source		
							case	control	case	control			
SNP rs2075876 (G > A)	Terao C	A	2011	Japan	American College of Rheumatology for RA (1987)	microarrays	63.0 ± 12.5	52.0 ± 15.8	82.1	60.6	HB		
		B	2011			Asian	microarrays	60.8 ± 11.5	38.1 ± 11.9	84.1	39.6	HB	
		C	2011				microarrays	61.4 ± 11.5	52.5 ± 15.2	81.4	44.4	HB	
	García-Lozano JR		2013	Spain	Caucasian	American College of Rheumatology for RA (1987)	Taqman SNP genotyping assay	49.2 ± 14.8	NA	74.3	NA	HB	
	Shao S		2014	China	Asian	American College of Rheumatology for RA (1987)	SNaPshot assay	48.7 ± 14.2	47.0 ± 16.3	80.6	37.0	HB	
	Feng ZJ		2015	China (Han)	Asian	American College of Rheumatology for RA (1987)	Taqman SNP genotyping assay	54.1 ± 11.2	52.4 ± 11.8	53.5	58.5	HB	
	Li X		2016	China (Shaanxi)	Asian	American College of Rheumatology for RA (1987)	Snapshot Assay	43.5 ± 19.2	44.3 ± 17.8	64.3	59.7	HB	
SNP rs760426 (A > G)	Terao C	A	2011	Japan	American College of Rheumatology for RA (1987)	microarrays	63.0 ± 12.5	52.0 ± 15.8	82.1	60.6	HB		
		B	2011			microarrays	60.8 ± 11.5	38.1 ± 11.9	84.1	39.6	HB		
		C	2011			microarrays	61.4 ± 11.5	52.5 ± 15.2	81.4	44.4	HB		
		Shao S		2014	China	Asian	American College of Rheumatology for RA (1987)	SNaPshot assay	48.7 ± 14.2	47.0 ± 16.3	80.6	37.0	HB
		Feng ZJ		2015	China (Han)	Asian	American College of Rheumatology for RA (1987)	Taqman SNP genotyping assay	54.1 ± 11.2	52.4 ± 11.8	53.5	58.5	HB
		Li X		2016	China (Shaanxi)	Asian	American College of Rheumatology for RA (1987)	Snapshot Assay	43.5 ± 19.2	44.3 ± 17.8	64.3	59.7	HB

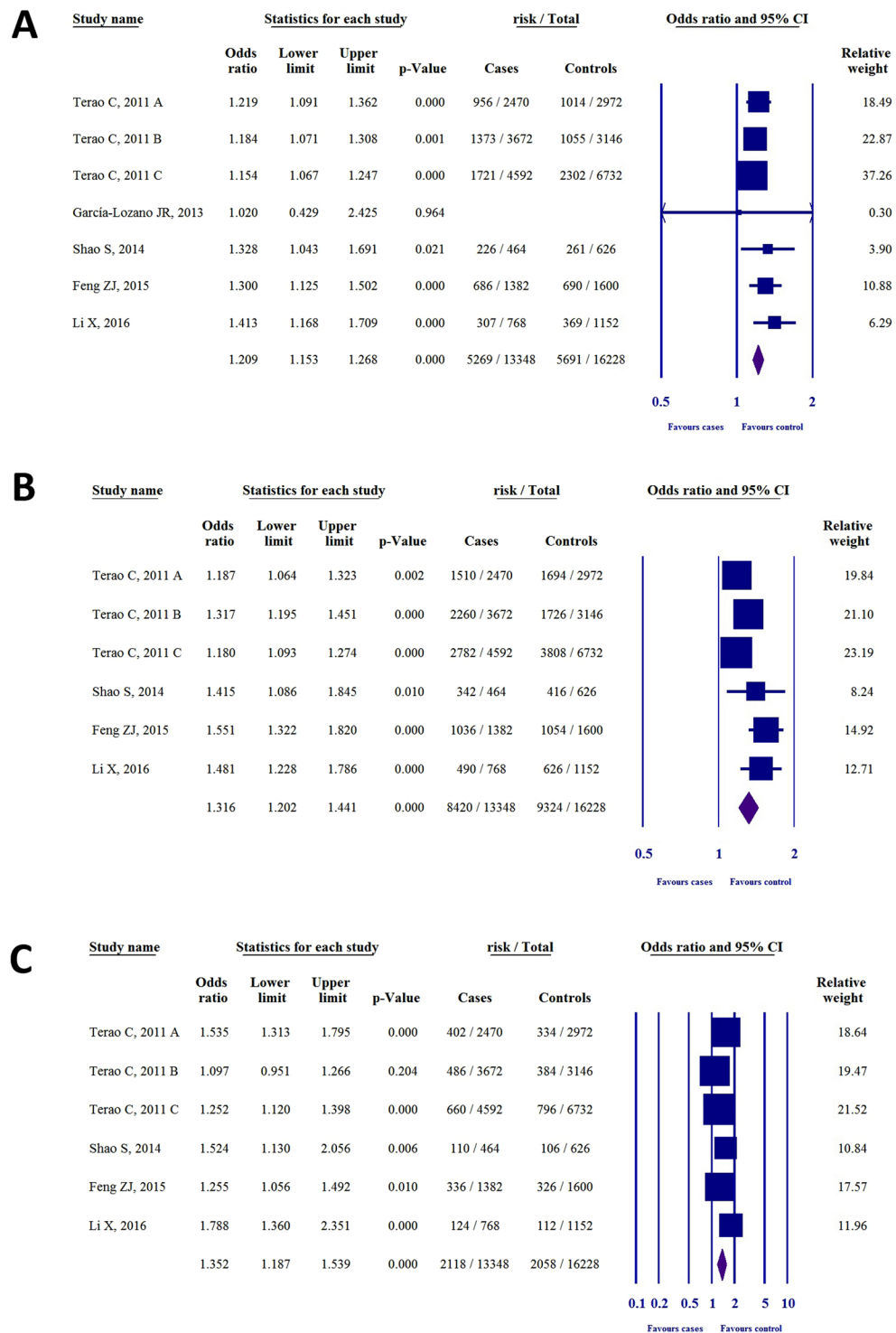
**Table 1.** Characteristics of the included studies on SNP rs2075876 (G > A) and rs760426 (A > G) (SNP = single nucleotide polymorphism; NA = not available; HB = hospital based).

Table S1). GWAS by Terao *et al.*<sup>18</sup> served as three independent case-control studies (denoted with A, B, C). With the exception of García-Lozano *et al.*<sup>19</sup>, all of the studies described the genotype distribution for GG, AG, AA. Therefore, we calculated odds ratios (ORs) for genetic models where there was no available or feasible data in the given study (Table 2). Results for each genetic model are shown in Fig. 2. For the allelic model (A vs. G, Fig. 2A) the ORs were 1.21 (95% CI 1.09–1.36,  $P < 0.001$ ), 1.18 (1.07–1.30,  $P = 0.001$ ), 1.15 (1.06–1.24,  $P < 0.001$ ), 1.02 (0.42–2.42,  $P = 0.964$ ), 1.32 (1.04–1.69,  $P = 0.021$ ), 1.30 (1.12–1.50,  $P < 0.001$ ), and 1.41 (1.16–1.70,  $P < 0.001$ ). For the dominant model (AG + AA vs. GG, Fig. 2B) the ORs were 1.18, (95% CI 1.06–1.32,  $P = 0.002$ ), 1.31 (1.19–1.45,  $P < 0.001$ ), 1.18 (1.09–1.27,  $P < 0.001$ ), 1.41 (1.08–1.84,  $P = 0.010$ ), 1.55 (1.32–1.82,  $P < 0.001$ ), and 1.48 (1.22–1.78,  $P < 0.001$ ). For the recessive model (AA vs. AG + GG, Fig. 2C) the ORs were 1.53 (95% CI 1.31–1.79,  $P < 0.001$ ), 1.09 (0.95–1.26,  $P = 0.204$ ), 1.25 (1.12–1.39,  $P < 0.001$ ), 1.52 (1.13–2.05,  $P = 0.006$ ), 1.25 (1.05–1.49,  $P = 0.010$ ), and 1.78 (1.36–2.35,  $P < 0.001$ ). For the codominant heterozygous model (AG vs. GG, see Supplementary Fig. S1), the ORs were 1.08 (0.96–1.21,  $P = 0.168$ ), 1.32 (1.20–1.47,  $P < 0.001$ ), 1.13 (1.05–1.23,  $P = 0.002$ ), 1.28 (0.97–1.70,  $P = 0.077$ ), 1.51 (1.28–1.79,  $P < 0.001$ ), and 1.34 (1.10–1.64,  $P = 0.003$ ). For the codominant homozygous model (AA vs. GG, see Supplementary Fig. S1) the ORs were 1.60 (95% CI 1.35–1.89,  $P < 0.001$ ), 1.27 (1.09–1.48,  $P = 0.002$ ), 1.34 (1.19–1.51,  $P < 0.001$ ), 1.78 (1.26–2.52,  $P = 0.001$ ), 1.62 (1.32–1.99,  $P < 0.001$ ), and 2.09, (1.56–2.81,  $P < 0.001$ ).

Results of heterogeneity analysis for each genetic model are shown in Supplementary Table S2. For the allelic model  $P_h = 0.439$  and  $I^2 = 0\%$ , for the dominant model  $P_h = 0.011$  and  $I^2 = 66.2\%$ , for the recessive model  $P_h = 0.005$  and  $I^2 = 69.9\%$ , for the codominant heterozygous model  $P_h = 0.004$  and  $I^2 = 70.5\%$ , and for the codominant homozygous model  $P_h = 0.012$  and  $I^2 = 65.4\%$ . Moderate heterogeneity was found in dominant, recessive, codominant heterozygous, and codominant homozygous models.

Only four out of 31 ORs were statistically insignificant, and the ORs revealed that SNP rs2075876 (G > A) is associated with an elevated risk of RA. These results therefore suggest a link between AIRE SNP rs2075876 (G > A) and RA susceptibility.

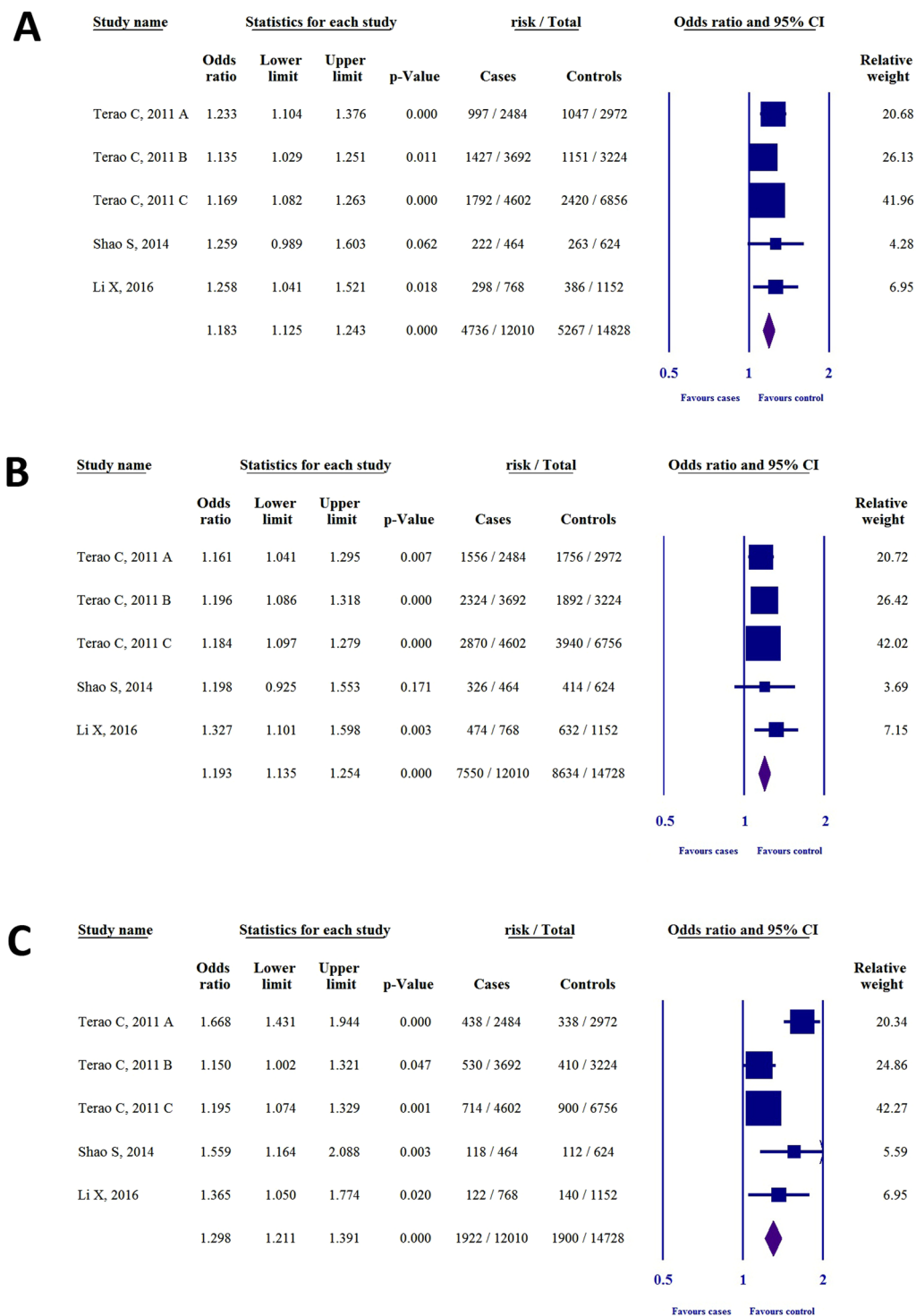
**Meta-analysis of SNP rs760426 (A > G).** Four studies investigated the association between SNP rs760426 and RA susceptibility<sup>18,20–22</sup>. Most of the publications doubled the individual number; thus, to normalize the data, we also calculated with duplicated values, as was conducted with rs2075876 SNP (see Supplementary Table S1). Again, GWAS by Terao *et al.*<sup>18</sup> served as three independent case-control studies (denoted with A, B, C).



**Figure 2.** The association of SNP rs2075876 (G > A) with RA risk in different genetic models. (A) Allelic model (A vs. G). (B) Dominant model (AG + AA vs. GG). (C) Recessive model (AA vs. AG + GG).

With the exception of *Feng et al.*<sup>21</sup>, all studies described the genotype distribution for AA, GA, GG. We therefore calculated ORs for all the genetic models that were not published in the original articles (Table 2). Furthermore, we excluded the OR, 95% CI and p-value of *Feng et al.*<sup>21</sup> from the statistical analysis due to the asymmetry of the OR. Results for each genetic model are shown in Fig. 3. For the allelic model (G vs. A), Fig. 3A, the ORs were 1.23 (95% CI 1.10–1.37,  $P < 0.001$ ), 1.13 (1.02–1.25,  $P = 0.011$ ), 1.16 (1.08–1.26,  $P < 0.001$ ), 1.25 (0.98–1.60,  $P = 0.062$ ), 1.25 (1.04–1.52,  $P = 0.018$ ). For the dominant model (GG + GA vs. AA, Fig. 3B), the ORs were 1.16 (1.04–1.29,  $P = 0.007$ ), 1.19 (1.08–1.31,  $P < 0.001$ ), 1.18 (1.09–1.27,  $P < 0.001$ ), 1.19 (0.92–1.55,  $P = 0.171$ ), and OR 1.32 (1.10–1.59,  $P = 0.003$ ). For the recessive model (GG vs. GA + AA, Fig. 3C) the ORs were 1.66 (95%CI

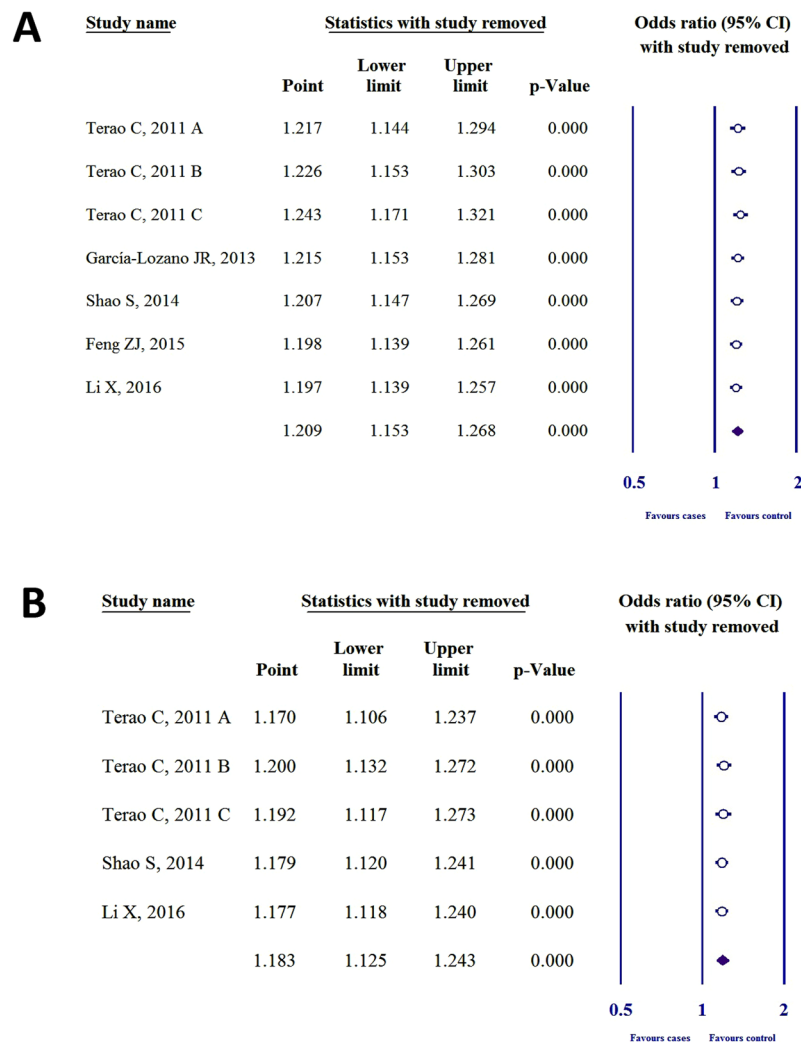




**Figure 3.** The association of SNP rs760426 (A > G) with RA risk in different genetic models. (A) Allelic model (A vs. G). (B) Dominant model (AG + AA vs. GG). (C) Recessive model (AA vs. AG + GG).

1.43–1.94,  $P < 0.001$ ), 1.15 (1.00–1.32,  $P = 0.047$ ), 1.19 (1.07–1.32,  $P = 0.001$ ), 1.55 (1.16–2.08,  $P = 0.003$ ), and 1.36 (1.05–1.77,  $P = 0.020$ ). For the codominant heterozygous model (GA vs. AA, see Supplementary Fig. S2) the ORs were 1.03, (95% CI 0.92–1.16,  $P = 0.582$ ), 1.17 (1.06–1.30,  $P = 0.002$ ), 1.15 (1.06–1.25,  $P = 0.001$ ), 1.04 (0.79–1.38,  $P = 0.741$ ), and 1.26 (1.03–1.54,  $P = 0.020$ ). For the codominant homozygous model (GG vs. AA, see Supplementary Fig. S2), the ORs were 1.69 (1.43–2.00,  $P < 0.001$ ), 1.25 (1.08–1.46,  $P = 0.003$ ), 1.29 (1.15–1.44,  $P < 0.001$ ), 1.60 (1.15–2.24,  $P = 0.006$ ), and 1.54 (1.16–2.04,  $P = 0.003$ ).

Results of heterogeneity analysis for each genetic model are shown in Supplementary Table S2. For the allelic model  $P_h = 0.737$ ,  $I^2 = 0\%$ , for the dominant model  $P_h = 0.822$ ,  $I^2 = 0\%$ , for the recessive model:  $P_h = 0.001$ ,



**Figure 4.** Sensitivity analysis for the allelic models of (A) SNP rs2075876 (G > A) and (B) rs760426 (A > G).

$I^2 = 76.7\%$ , for the codominant heterozygous model:  $P_h = 0.323$ ,  $I^2 = 14.2\%$ , and for the codominant homozygous model  $P_h = 0.038$ ,  $I^2 = 60.5\%$ . Moderate heterogeneity was found in recessive and codominant homozygous models.

Only four out of 26 ORs were statistically insignificant, and the ORs showed that rs760426 (A > G) SNP is associated an elevated risk. These results therefore suggest a link between *AIRE* SNP rs760426 (A > G) and RA susceptibility.

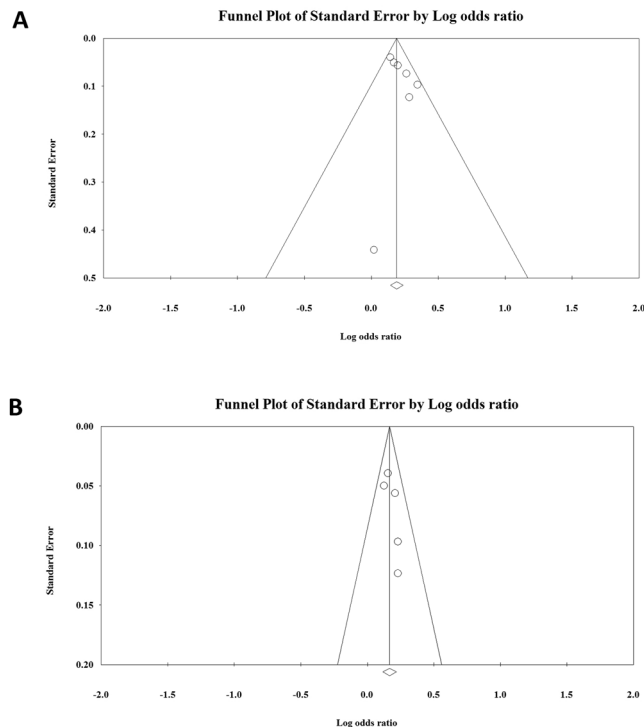
**Sensitivity analysis.** To detect the influence of each case-control study on the whole meta-analysis, we performed sensitivity analysis by omitting one individual study. Heterogeneity was not found in SNP rs2075876 or rs760426 by investigating allelic (Fig. 4), dominant, recessive, codominant heterozygous, and codominant homozygous genetic models (see Supplementary Figs S3 and S4).

**Publication bias.** Bias analysis was performed by generating funnel plots for each polymorphism of the allelic (Fig. 5), dominant, recessive, codominant heterozygous, and codominant homozygous genetic models (see Supplementary Figs S5 and S6). After analysis, all funnel plots were perfectly symmetric, and no publication bias was detected for SNP rs2075876 or rs760426.

**Trial sequential analysis.** We performed a TSA for the allelic models (Fig. 6) of SNPs rs2075876 (G > A) and rs760426 (A > G). Results of allelic models for both polymorphisms showed that the blue line of cumulative z-curve crossed the TSA monitoring boundary and the cumulative sample size was reached. Therefore, we observed robust evidence in the association between SNPs rs2075876 (G > A) and rs760426 (A > G) and RA risk. These results suggest that no further studies are necessary to confirm the association.

polymorphism	study	Genetic model	OR	95%CI	P		
SNP rs2075876 (G > A)	Terao C, 2011	A	Allelic*	(A vs. G)	1.21	1.09–1.36	<0.001
			Dominant	(AG + AA vs. GG)	1.18	1.06–1.32	0.002
			Recessive	(AA vs. AG + GG)	1.53	1.31–1.79	<0.001
			Codominant heterozygous	(AG vs. GG)	1.08	0.96–1.21	0.168
			Codominant homozygous	(AA vs. GG)	1.60	1.35–1.89	<0.001
			Allelic*	(A vs. G)	1.18	1.07–1.30	<0.001
		B	Dominant	(AG + AA vs. GG)	1.31	1.19–1.45	<0.001
			Recessive	(AA vs. AG + GG)	1.09	0.95–1.26	0.204
			Codominant heterozygous	(AG vs. GG)	1.32	1.20–1.47	<0.001
			Codominant homozygous	(AA vs. GG)	1.27	1.09–1.48	0.002
			Allelic*	(A vs. G)	1.15	1.06–1.24	<0.001
			Dominant	(AG + AA vs. GG)	1.18	1.09–1.27	<0.001
	C	Recessive	(AA vs. AG + GG)	1.25	1.12–1.39	<0.001	
		Codominant heterozygous	(AG vs. GG)	1.13	1.05–1.23	0.002	
		Codominant homozygous	(AA vs. GG)	1.34	1.19–1.51	<0.001	
		Allelic	(A vs. G)	1.02	0.42–2.42	0.964	
		García-Lozano JR, 2013	Dominant	(AG + AA vs. GG)	NA		
			Recessive	(AA vs. AG + GG)			
	Codominant heterozygous		(AG vs. GG)				
	Codominant homozygous		(AA vs. GG)				
	Allelic*		(A vs. G)	1.32			
	Shao S, 2014	Dominant	(AG + AA vs. GG)	1.41	1.08–1.84	0.010	
		Recessive	(AA vs. AG + GG)	1.52	1.13–2.05	0.006	
		Codominant heterozygous	(AG vs. GG)	1.28	0.97–1.70	0.077	
		Codominant homozygous	(AA vs. GG)	1.78	1.26–2.52	0.001	
		Allelic	(A vs. G)	1.30	1.12–1.50	<0.001	
	Feng ZJ, 2015	Dominant	(AG + AA vs. GG)	1.55	1.32–1.82	<0.001	
		Recessive	(AA vs. AG + GG)	1.25	1.05–1.49	0.010	
		Codominant heterozygous	(AG vs. GG)	1.51	1.28–1.79	<0.001	
		Codominant homozygous	(AA vs. GG)	1.62	1.32–1.99	<0.001	
		Allelic*	(A vs. G)	1.41	1.16–1.70	<0.001	
	Li X, 2016	Dominant	(AG + AA vs. GG)	1.48	1.22–1.78	<0.001	
Recessive		(AA vs. AG + GG)	1.78	1.36–2.35	<0.001		
Codominant heterozygous		(AG vs. GG)	1.34	1.10–1.64	0.003		
Codominant homozygous		(AA vs. GG)	2.09	1.56–2.81	<0.001		
Allelic*		(G vs. A)	1.23	1.10–1.37	<0.001		
SNP rs760426 (A > G)	Terao C, 2011	A	Dominant	(GG + GA vs. AA)	1.16	1.04–1.29	0.007
			Recessive*	(GG vs. GA + AA)	1.66	1.43–1.94	<0.001
			Codominant heterozygous*	(GA vs. AA)	1.03	0.92–1.16	0.582
			Codominant homozygous*	(GG vs. AA)	1.69	1.43–2.00	<0.001
			Allelic*	(G vs. A)	1.13	1.02–1.25	0.011
			Dominant	(GG + GA vs. AA)	1.19	1.08–1.31	<0.001
		B	Recessive	(GG vs. GA + AA)	1.15	1.00–1.32	0.047
			Codominant heterozygous	(GA vs. AA)	1.17	1.06–1.30	0.002
			Codominant homozygous	(GG vs. AA)	1.25	1.08–1.46	0.003
			Allelic*	(G vs. A)	1.16	1.08–1.26	<0.001
			Dominant	(GG + GA vs. AA)	1.18	1.09–1.27	<0.001
			Recessive	(GG vs. GA + AA)	1.19	1.07–1.32	0.001
	C	Codominant heterozygous	(GA vs. AA)	1.15	1.06–1.25	0.001	
		Codominant homozygous	(GG vs. AA)	1.29	1.15–1.44	<0.001	
		Allelic*	(G vs. A)	1.25	0.98–1.60	0.062	
		Dominant	(GG + GA vs. AA)	1.19	0.92–1.55	0.171	
		Recessive	(GG vs. GA + AA)	1.55	1.16–2.08	0.003	
		Codominant heterozygous	(GA vs. AA)	1.04	0.79–1.38	0.741	
	Shao S, 2014	Codominant homozygous	(GG vs. AA)	1.60	1.15–2.24	0.006	
		Allelic*	(G vs. A)	1.87	1.09–2.45	0.074	
		Dominant	(GG + GA vs. AA)	NA			
		Recessive	(GG vs. GA + AA)				
		Codominant heterozygous	(GA vs. AA)				
	Codominant homozygous	(GG vs. AA)					
	Allelic*	(G vs. A)	1.25				1.04–1.52
	Feng ZJ, 2015	Dominant	(GG + GA vs. AA)	1.32	1.10–1.59	0.003	
		Recessive	(GG vs. GA + AA)	1.36	1.05–1.77	0.020	
		Codominant heterozygous	(GA vs. AA)	1.26	1.03–1.54	0.020	
		Codominant homozygous	(GG vs. AA)	1.54	1.16–2.04	0.003	
		Allelic*	(G vs. A)	1.25	1.04–1.52	0.018	
	Li X, 2016	Dominant	(GG + GA vs. AA)	1.32	1.10–1.59	0.003	
		Recessive	(GG vs. GA + AA)	1.36	1.05–1.77	0.020	
Codominant heterozygous		(GA vs. AA)	1.26	1.03–1.54	0.020		
Codominant homozygous		(GG vs. AA)	1.54	1.16–2.04	0.003		
Allelic*		(G vs. A)	1.25	1.04–1.52	0.018		

**Table 2.** ORs, 95% CIs, and P-values for each genetic model in the association of SNPs rs2075876 (G > A) and rs760426 (A > G) with RA risk (NA = not available; OR = odds ratio; CI = confidence interval; \*literature data.



**Figure 5.** Funnel plots of allelic genetic models of (A) SNP rs2075876 (G > A) and (B) rs760426 (A > G).

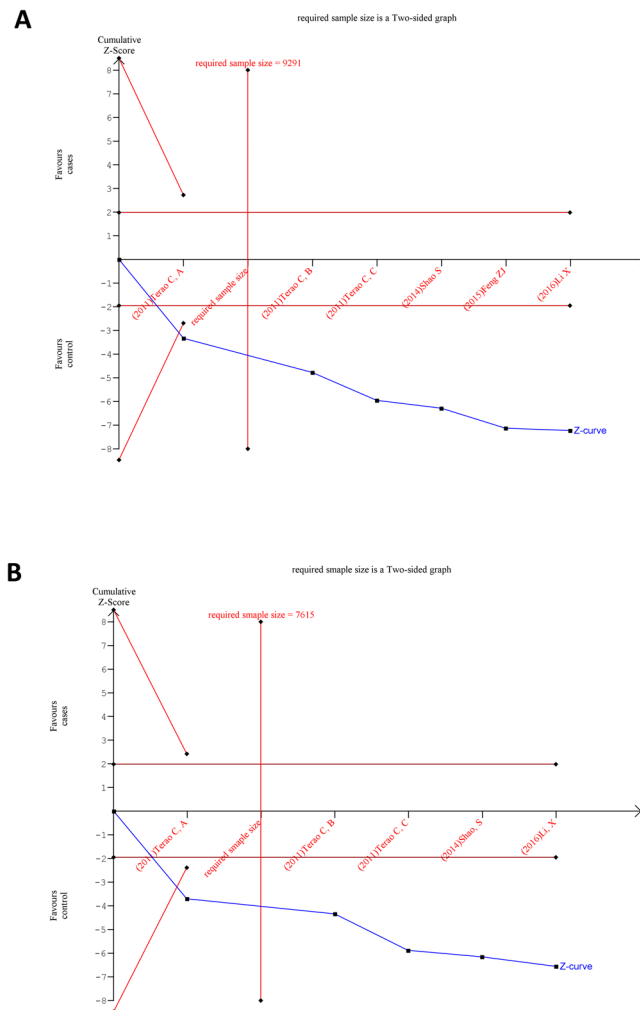
## Discussion

RA is a multifactorial disorder where genetic and environmental events equally contribute to disease commencement<sup>1,25</sup>. The latest GWAS meta-analysis discovered and screened 42 novel RA risk SNPs at a genome level from 98 candidate biological RA risk genes<sup>6</sup>. The detected risk genes, including *AIRE*, are mainly in the category of primary immunodeficiency (PID), HIV, and immune dysregulation. With the exception of *AIRE*, none of the other associated proteins have been directly related to central tolerance<sup>6</sup>.

Self-tolerance involving negative selection, the machinery of which is directed by *AIRE*, is a central immuno-physiological process required to create a normal adaptive immune system. We believe that polymorphisms in this indispensable gene lower the protein expression of *AIRE*, decrease the presentation of self-antigens, reduce negative selection, and contribute to the escape and survival of autoimmune T-cells. Reaching the periphery, matured, autoimmune T-cells are a source of autoantibodies and serve as a medium for numerous immune disorders, including RA. In support of this belief, Lovewell *et al.*<sup>13</sup> have concluded, through a gene reporter assay, that specific haplotypes (*AIRE*–655G *AIRE*-230T) can dramatically reduce *AIRE* transcription. However, with *in vitro* and *in vivo* experiments, Kont *et al.*<sup>26</sup> have demonstrated that the presentation of PTAs from mTECs is quantitatively affected by these reductions in *AIRE* expression.

An *in silico* investigation by Terao *et al.*<sup>18</sup>, which analysed the expression profile of 210 lymphoblastoid cells in the Gene Expression Omnibus (GEO) database, has demonstrated a statistically significant ( $p < 0.001$ ) correlation between the rs2075876 risk allele (A) and decreased *AIRE* transcription<sup>27</sup>. No association was found in GEO between rs760426 (G) and *AIRE* expression. Additionally, García-Lozano *et al.*<sup>19</sup> found statistically significant decreases in the expression levels of rs878081 C allele by analysing GEO database. This SNP is located in the Exon 5 region of *AIRE*; however, rs2075876 (G > A) is located in Intron 5 and rs760426 (A > G) in Intron 12<sup>7</sup>. The latter SNPs may affect the transcription of *AIRE* by modifying alternative splicing or intron-mediated enhancement<sup>28</sup>. The reduction in transcription, in turn, provides lower amounts of PTAs ectopically on the major histocompatibility complex/human leukocyte antigen of mTECs, which thereby contributes to the failure of negative selection in the thymus and increases the survival of autoimmune T-cells. In individuals who carry these SNPs, this sequence increases RA susceptibility. By analysing allelic, dominant, recessive, codominant heterozygous, and codominant homozygous models, we demonstrated that the SNPs rs2075876 (G > A) and rs760426 (A > G) occur more frequently in RA patients than in controls.

There are some limitations in our meta-analysis. We cannot extrapolate the findings of rs2075876 (G > A) and rs760426 (A > G) to Caucasians due to the limited study number. Based on García-Lozano *et al.*<sup>19</sup>, the results are not statistically significant; however, rs878081 C allele seemed to occur more frequently in RA patients. Furthermore, considering GWAS of Terao *et al.*<sup>18</sup>, the association of *AIRE* with RA among Caucasians was not supported. The number of the included studies also limited our meta-analysis; however, Terao *et al.*<sup>18</sup> provided three case-control studies in one publication, which elevated the number of the included epidemiological studies. In the future, further European case-control, GWAS, and stratified subgroup analyses (age, smoking) are needed in order to better elucidate the association between RA and *AIRE* polymorphism.



**Figure 6.** Trial sequential analysis for allelic genetic models of (A) SNP rs2075876 ( $G > A$ ) and (B) rs760426 ( $A > G$ ).

To our knowledge, this is the first time that the association between SNP rs2075876 ( $G > A$ ), rs760426 ( $A > G$ ), and RA susceptibility was statistically estimated in one meta-analysis. We used multiple haplotype investigations for each polymorphism, sensitivity analyses, and TSA to confirm the robustness of association. In conclusion, our meta-analysis clearly confirmed with each genetic model that the presence of SNPs rs2075876 ( $G > A$ ) and rs760426 ( $A > G$ ) is significantly associated with an increased risk for RA.

## Methods

**Search strategy.** We searched for related literature in the PubMed, Embase, Cochrane Library, and Web of Science databases in accordance with the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement<sup>29</sup>. On 16th May 2017 we completed the search. Keywords (“autoimmune regulator”; “AIRE”; “polymorphism”; “rheumatoid arthritis”) were thoroughly used by two independent investigators. All studies were published from April 2011 to June 2016.

**Inclusion and Exclusion criteria.** In order for studies to be included, publications had to demonstrate that (1) the study focused on the association between SNPs or haplotypes within the *AIRE* gene and RA susceptibility, (2) the study was case-control-designed, (3) all RA patients met the American College of Rheumatology classification and diagnostic criteria, and (4) detailed genotype data and feasible ORs, 95% CIs, and p-values were available. Publications were excluded if (1) a previous study was duplicated or (2) the given polymorphism was not found in at least four studies. Review articles were also excluded. Inclusion and exclusion criteria were independently screened by two investigators.

**Statistical analysis.** HWE was calculated by the chi-squared test for each study in the control groups. Pooled ORs and 95% CIs were calculated to examine the strength of the association between rs2075876 and rs760426 polymorphisms and RA. We used the random effect model by DerSimonian and Laird<sup>30</sup> because of the different ethnicities of those included. Heterogeneity between trials was tested with two methods. First, we employed the Cochrane’s Q homogeneity test, which exceeds the upper-tail critical value of chi-square on  $k-1$

degrees of freedom, with a p-value of less than 0.10 considered suggestive of significant heterogeneity. Second, we used the inconsistency ( $I^2$ ) index.  $I^2$  is the proportion of total variation contributed by between-study variability.  $I^2$  values of 25, 50 and 75% correspond to low, moderate, and high degrees of heterogeneity, respectively, based on Cochran's handbook<sup>31</sup>. Sensitivity analyses were performed to identify the influence of each study on the pooled ORs and 95% CIs. Publication bias was examined by visual inspection of funnel plots where the standard error was plotted against the log odds ratio. Meta-analytic calculations were performed with Comprehensive MetaAnalysis software Version 3 (Biostat, Inc., Englewood, NJ, USA).

**Trial sequential analysis (TSA).** Meta-analyses may be biased in type I errors owing to an increased risk of random error when sparse data are analysed, combined with reduplicative testing on accumulating data. To avoid this problem and to increase the robustness of conclusions, we used trial sequential analysis (TSA)<sup>32–34</sup>. TSA combines an estimation of the required sample size with an adjusted threshold for statistical significance. The relationship between the cumulative z-curve and the trial sequential monitoring boundary shows the expressiveness of the meta-analysis. If the cumulative z-curve crosses the trial sequential monitoring boundary, and the cumulative sample size of the meta-analysis reaches the required sample size, firm evidence can be observed. When the cumulative z-curve crosses the boundaries, but the sample size does not reach the required information size, a sufficient level of evidence for the anticipated intervention effect may have been reached and no further trials are needed. If the z-curve does not cross any of the boundaries and the required sample size has not been reached, evidence to reach a conclusion is insufficient<sup>35</sup>. For calculation of the information size, we used a heterogeneity adjusted assumption with 10% of relative risk reduction, 5% of overall Type-I-Error, and 10% of Type-II-Error for the case of both gene alleles. The adjusted CIs for rs2075876 and rs760426 are 1.13–1.31 and 1.11–1.26, respectively. For calculations we used the Trial Sequential Analysis software tool from Copenhagen Trial Unit, Center for Clinical Intervention Research, Denmark (version 0.9 beta, [www.ctu.dk/tsa](http://www.ctu.dk/tsa)).

All data generated or analysed during this study are included in this published article and its Supplementary information file.

## References

1. Scott, D. L., Wolfe, F. & Huizinga, T. W. Rheumatoid arthritis. *Lancet* **376**, 1094–1108 (2010).
2. Cosway, E., Anderson, G., Garside, P. & Prendergast, C. The thymus and rheumatology: should we care? *Curr. Opin. Rheumatol.* **28**, 189–195 (2016).
3. Viatte, S., Plant, D. & Raychaudhuri, S. Genetics and epigenetics of rheumatoid arthritis. *Nat. Rev. Rheumatol.* **9**, 141–153 (2013).
4. Starr, T. K., Jameson, S. C. & Hogquist, K. A. Positive and negative selection of T cells. *Annu. Rev. Immunol.* **21**, 139–176 (2003).
5. Mellado, M. *et al.* T Cell Migration in Rheumatoid Arthritis. *Front. Immunol.* **6**, 1–12 (2015).
6. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
7. Colobran, R., Giménez-Barcons, M., Marín-Sánchez, A., Porta-Pardo, E. & Pujol-Borrell, R. AIRE genetic variants and predisposition to polygenic autoimmune disease: The case of Graves' disease and a systematic literature review. *Hum. Immunol.* **77**, 643–651 (2016).
8. Nagamine, K. *et al.* Positional cloning of the APECED gene. *Nat. Genet.* **17**, 393–398 (1997).
9. Finnish-German, A. P. E. C. E. D. Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat. Genet.* **17**, 399–403 (1997).
10. Mathis, D. & Benoist, C. Aire. *Annu. Rev. Immunol.* **27**, 287–312 (2009).
11. Anderson, M. S. & Su, M. A. Aire and T cell development. *Curr. Opin. Immunol.* **23**, 198–206 (2011).
12. Heino, M. *et al.* APECED mutations in the autoimmune regulator (AIRE) gene. *Hum. Mutat.* **18**, 205–211 (2001).
13. Lovewell, T. R., McDonagh, A. J., Messenger, A. G., Azzouz, M. & Tazi-Ahnini, R. The AIRE -230Y Polymorphism Affects AIRE Transcriptional Activity: Potential Influence on AIRE Function in the Thymus. *PLoS One*. **10**, e0127476 (2015).
14. Tazi-Ahnini, R. *et al.* The autoimmune regulator gene (AIRE) is strongly associated with vitiligo. *Br. J. Dermatol.* **159**, 591–596 (2008).
15. Wengraf, D. A. *et al.* Genetic analysis of autoimmune regulator haplotypes in alopecia areata. *Tissue Antigens*. **71**, 206–212 (2008).
16. Conteduca, G. *et al.* The role of AIRE polymorphisms in melanoma. *Clin. Immunol.* **136**, 96–104 (2010).
17. Ferrera, F. *et al.* AIRE gene polymorphisms in systemic sclerosis associated with autoimmune thyroiditis. *Clin. Immunol.* **122**, 13–17 (2007).
18. Terao, C. *et al.* The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Hum. Mol. Genet.* **20**, 2680–2685 (2011).
19. García-Lozano, J. R. *et al.* Association of the AIRE gene with susceptibility to rheumatoid arthritis in a European population: a case control study. *Arthritis Res. Ther.* **15**, R11 (2013).
20. Shao, S., Li, X. R., Cen, H. & Yin, Z. S. Association of AIRE polymorphisms with genetic susceptibility to rheumatoid arthritis in a Chinese population. *Inflammation*. **37**, 495–499 (2014).
21. Feng, Z. J., Zhang, S. L., Wen, H. F. & Liang, Y. Association of rs2075876 polymorphism of AIRE gene with rheumatoid arthritis risk. *Hum. Immunol.* **76**, 281–285 (2015).
22. Li, X., Li, T., Chen, M. & Chai, Y. Association of AIRE gene polymorphisms with susceptibility to rheumatoid arthritis among ethnic Han Chinese from Shaanxi. *Chin. J. Med. Genet.* **33**, 373–377 (2016).
23. Xu, Y. S., Jiang, X. J. and Chen, J. M. A single nucleotide polymorphism of AIRE gene located in the 21q22. 3 increases the risk of rheumatoid arthritis. *Oncotarget* (2017).
24. Arnett, F. C. *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988).
25. van der Woude, D. *et al.* Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum.* **60**, 916–923 (2009).
26. Kont, V. *et al.* Modulation of Aire regulates the expression of tissue-restricted antigens. *Mol. Immunol.* **45**, 25–33 (2008).
27. Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
28. Rose, A. B. Intron-mediated regulation of gene expression. *Curr. Top. Microbiol. Immunol.* **326**, 277–290 (2008).
29. Moher, D. *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **4**, 1 (2015).
30. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials revisited. *Contemp. Clin. Trials* **45**, 139–145 (2015).
31. The Cochrane Collaboration Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (eds. Higgins, J. P. T. & Green, S.; John Wiley & Sons, 2011)



32. Turner, R. M., Bird, S. M. & Higgins, J. P. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PLoS One* **8**, e59202 (2013).
33. Wetterslev, J., Thorlund, K., Brok, J. & Gluud, C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J. Clin. Epidemiol.* **61**, 64–75 (2008).
34. Brok, J., Thorlund, K., Wetterslev, J. & Gluud, C. Apparently conclusive meta-analyses may be inconclusive—Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *Int. J. Epidemiol.* **38**, 287–298 (2009).
35. Wetterslev, J., Thorlund, K., Brok, J. & Gluud, C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. *BMC Med. Res. Methodol.* **9**, 86 (2009).

## Acknowledgements

This study was supported by grants from the Medical School, University of Pécs, Hungary (No. PTE-AOK-KA-2017). The present scientific contribution is dedicated to the 650th anniversary of the foundation of the University of Pécs, Hungary.

## Author Contributions

B.B., P.H., and Z.G. designed this study; B.B. and G.G. searched databases and collected full-text papers; B.B., G.V. and J.B. extracted data; K.F., B.E., and L.C. analysed data; B.B., H.A., and É.V. wrote the manuscript; Z.R., K.S., B.K., and Z.G. reviewed the manuscript. All of the authors have read and approved the final version of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-14375-z>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

# Association Between *AIRE* Polymorphisms rs870881(C>T), rs1003854(T>C) and Rheumatoid Arthritis Risk: A Hungarian Case-control Study

BALINT BERCZI<sup>1</sup>, NÓRA NUSSER<sup>2</sup>, IVAN PETER<sup>2</sup>, BALAZS NEMETH<sup>1,2</sup> and ZOLTAN GYONGYI<sup>1</sup>

<sup>1</sup>Department of Public Health Medicine, Medical School, University of Pécs, Pécs, Hungary;

<sup>2</sup>Harkány Thermal Rehabilitation Centre, Harkány, Hungary

**Abstract.** *Background/Aim:* Autoimmune regulator (*AIRE*) is a transcription factor that plays pivotal role in controlling autoimmunity. In the thymus, it supports the presentation of peripheral tissue antigens to developing T cells, where recognition of these self-antigens negatively selects the autoimmune naïve T-cells by central tolerance. Studies demonstrated that single-nucleotide polymorphisms (SNPs) in *AIRE* alter transcription and propagate clonal survival of autoimmune T cells, therefore increase susceptibility to autoimmune diseases. This study intended to identify SNPs in exon and intron sequences that determine *AIRE* transcription, where their genotypes are associated with rheumatoid arthritis (RA) risk and clinical parameters. *Patients and Methods:* After a thorough *in silico* research, we enrolled 100 patients with RA and 100 healthy controls to analyze the association of SNP rs870881(C>T) and rs1003854(T>C) in *AIRE* coding sequence with RA risk by using five different genetic models and selected clinical parameters. Multiplex quantitative polymerase chain reaction was used to determine allelic discrimination of SNPs. RA risk was assessed by odds ratios (ORs) and confidence intervals (CIs). *Results:* In a recessive model of rs870881, minor allele TT homozygotes were associated with RA ( $p=0.032$ ,  $OR=5.44$ ,  $95\%CI=1.16-25.52$ ); in a recessive model of rs1003854, minor allele CC

homozygotes were associated with RA ( $p=0.047$ ,  $OR=4.84$ ,  $95\%CI=1.02-23.02$ ). Higher C-reactive protein (CRP) levels in patients with RA were significantly associated with minor allele homozygotes in recessive and codominant genetic models ( $p=0.029$  and  $p=0.043$ , respectively) of rs1003854. *Conclusion:* Genotypes for minor alleles of rs870881 and rs1003854 might be involved in RA pathogenesis and risk prediction.

Rheumatoid arthritis (RA) is a common disease of immune dysfunction characterized by chronic synovitis; the resultant polyarticular arthritis with extra-articular symptoms causes functional impairment. The disease prevalence is 0.5-1% among adults, and the annual incidence in developed countries is 5-50 per 100,000 adults (1).

The immunopathology of RA has not been completely clarified; however, many studies have reported that some autoimmune T cells clonally expand and survive negative selection of the adaptive immune system. These autoimmune T cells then infiltrate the synovium and cause RA (2, 3).

Despite the significant association between major histocompatibility complex (MHC) *HLA-DRB1* and RA by numerous genome-wide association studies (4-7), the two latest transethnic, high-throughput genome-wide association studies have associated 51 novel risk loci with RA using a combined total of 33,954 cases and 90,649 controls (8, 9). One of these risk loci is the coding sequence for *AIRE*, which is indispensable for the control of autoimmunity.

*AIRE* is located in the 21q22.3 chromosomal region and encodes the 545 amino acid transcription factor autoimmune regulator (*AIRE*), which is located in the nucleus of medullary thymic epithelial cells (mTECs) (10). During postnatal development of the immune system, *AIRE* binds to the DNA of mTECs to ensure the promiscuous, ectopic expression of peripheral tissue self-antigens in the thymic environment, which is the “educational” site of the incoming naïve T cell precursors, thymocytes (2). Self-antigens are presented on MHC molecules of mTECs to the T cell

*Correspondence to:* Balázs Németh, MD, Department of Public Health Medicine, Medical School, University of Pécs, Szigeti str. 12, H-7624 Pécs, Hungary. E-mail: balazs.nemeth@aok.pte.hu

**Key Words:** Autoimmune regulator, single-nucleotide polymorphism, allelic discrimination, rheumatoid arthritis, C-reactive protein.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) 4.0 international license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



receptors on thymocytes (2). If self-antigens are recognized by the T-cell receptor (TCR) of thymocytes, these naïve T-cells become autoreactive and negative selection clonally deletes them by apoptosis to ensure a normal T cell population without autoreactivity (2, 10). This negative selection is the last step in the process of immunological central tolerance (2). AIRE consists of several structural domains: the caspase-recruitment domain controls interactions with binding partners; the SAND domain has DNA-binding and transcription properties and controls peripheral tissue antigen production; the PHD fingers read the histone code; and the LXXLL sequences interact with the ubiquitous transcription activators CREB-binding protein (CBP) and signal transducers and activators of transcription (STAT) protein (10-13). Certain functional mutations in the *AIRE* protein-coding sequence result in the loss of negative selection, causing development of the systemic autoimmune disease autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), also known as autoimmune polyendocrine syndrome type 1 (APS-1), which causes multiorgan devastation (10, 14).

Quantitatively inappropriate presentation of self-antigens may decrease the effectiveness of central and peripheral tolerance, causing the survival and maturation of autoreactive thymocytes which leads to various autoimmune diseases, including RA. This effect has been confirmed for the glucose-6-phosphate isomerase self-antigen in the K/BxN mouse model of RA (15). Numerous studies describe the fundamental contribution to RA development of defective selection of autoimmune T cells and the resultant synovial infiltration of CD4<sup>+</sup> type 1 T helper (Th) 1, Th2, Th17, and regulatory T cells (3, 16-18). We endorse the rising scientific attention on thymic T cell selection and the necessity to analyze the role of gene variants of its key player, AIRE. The role of allelic polymorphisms in *AIRE* transcription was first described in an *in vivo* study that detected two distinct *AIRE* single-nucleotide polymorphisms (SNPs), *AIRE*-230Y and *AIRE*-655G. The haplotype *AIRE*-230T transcriptionally modified *AIRE* gene expression, which elevated the risk of autoimmunity through the alteration of negative selection (19). Although *AIRE* SNPs attract scientific attention, only few case-control studies have analyzed the association between allelic polymorphisms of *AIRE* and autoimmune diseases, such as vitiligo (20), alopecia areata (21), Addison's disease (22), and systemic sclerosis (23). The majority of these molecular epidemiological studies analyzed the association of *AIRE* SNPs with RA. Similarly, we previously conducted a meta-analysis to verify the association of *AIRE* allelic polymorphisms with RA in Asian populations (24). So far, the association in European populations is poorly understood. To date, one study has analyzed *AIRE* and RA in the Spanish population (25). According to our knowledge, our report is the first European

study to examine the association of genotypes of *AIRE* allelic polymorphisms and RA risk using five different genetic models (allelic, dominant, recessive, codominant heterozygous, codominant homozygous, and overdominant) to determine risk genotypes.

## Patients and Methods

**Preliminary *in silico* research.** We performed preliminary *in silico* research to identify SNPs that could be relevant to RA risk. We used the Ensembl automatic annotation of the human genome sequence (GRCh38.p13) and the UCSC Genome Browser (GRCh38 assembly) as a human genome database (26, 27). Our inclusion criteria for SNPs to be tested were global minor allele frequency (MAF) of 0.01-0.5 and presence of the variant in at least one of the following databases: 1000 Genomes Project Phase III (including the HapMap project), NCBI dbSNP, NHLBI GO Exome Sequencing Project (human only), Exome Aggregation Consortium (human only), Genome Aggregation Database (human only), and NHLBI Trans-Omics for Precision Medicine (human only) (28-32).

Hits were excluded as they had a low probability of detection in our study population. Therefore, we established a criterion to exclude a genetic variant if it existed in the 1000 Genomes Project database with a global MAF of 0.01-0.5 but the MAF was less than 0.01 in the EUR or CEU subpopulations.

**Study subjects.** Patients with RA and healthy control subjects were consecutively recruited by voluntary participation from individuals receiving outpatient or inpatient care at the Harkány Thermal Rehabilitation Centre, Hungary. We collected blood samples from the recruited subjects from May 2018 to September 2022; the subjects gave informed consent and received a study information sheet.

Blood samples were obtained and collected in citrate tubes (one 4 ml tube per subject). Whole blood was stored at -75°C until analysis. Our inclusion criteria for the patients with RA were that the subjects aged more than 18 years, had signed a consent statement, and had been clinically diagnosed with RA, based on the American College of Rheumatology classification (33). Our inclusion criteria for the healthy control subjects were that they aged more than 18 years, had signed a consent statement, and did not have RA, based on the American College of Rheumatology classification and had blood values of C-reactive protein (CRP) ≤20 mg/dl and rheumatoid factor (Rf) ≤25 IU/ml (34). Ultimately, we analyzed data from 100 patients with RA and 100 healthy control subjects.

**Laboratory assessment.** The blood samples were analyzed for complete blood count, cholesterol profile, blood sugar profile, erythrocyte sedimentation rate (ESR), CRP, sodium, potassium, urea, creatinine, glutamate-oxaloacetate transaminase, glutamate-pyruvate transaminase, gamma-glutamyl transferase, alkaline phosphatase, lactate dehydrogenase, creatine kinase, protein albumin, magnesium, phosphorus, and serum bilirubin. We measured rheumatoid factor (Rf), anti-cyclic citrullinated peptide (aCCP), and antinuclear antibodies to confirm seropositivity in patients with RA and seronegativity in control subjects (35, 36). Finally, we calculated the functional status of patients with RA by the Functional Independence Measure and disease activity score with 28-joint counts (DAS28) (37). ESR was quantified using the Westergren method (38). CRP and Rf were quantified at the

Table I. Demographic and clinical characteristics of study populations.

Characteristics	Patients with RA (N=100)	Control subjects (N=100)
Age (years), mean±SD	65.38±9.84*	81.17±4.56
Sex		
Female, no. (%)	85 (85%)	68 (68%)
Male, no. (%)	15 (15%)	32 (32%)
ESR (mm/h), mean±SD (normal range=3-13 mm/h)	40.24±23.90*	15.43±9.97
CRP (mg/dl), mean±SD (normal range=0-10 mg/dl)	28.80±29.21*	6.19±2.53
Rf (IgG) (IU/ml), mean±SD (normal range=0-15 IU/ml)	166.15±156.39*	11.79±7.61
aCCP (U/ml), mean±SD (normal range: <20.0 U/ml; weak positivity: ≥20.0-39.0 U/ml; moderate positivity: ≥39.0-59.0 U/ml; strong positivity: ≥59.0 U/ml)	561.10±583.07*	4.566±2.65
Seropositive, no. (%)	66 (66%)*	0 (0%)
Seronegative, no. (%)	14 (14%)*	100 (100%)
DAS28, mean±SD (disease remission: <2.6; low: ≥2.6-3.1; moderate: ≥3.1-5.1; high: ≥5.1)	4.93±1.46	–

RA: Rheumatoid arthritis; SD: standard deviation; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein; Rf: rheumatoid factor; IgG: immunoglobulin G; aCCP: anti-cyclic citrullinated peptide; DAS28: disease activity score with 28-joint counts; \* $p < 0.001$ , compared with control subjects.

Harkány Thermal Rehabilitation Centre using Konelab Arena 20XT Clinical Chemistry Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) in accordance with the manufacturer's recommendations. Levels of aCCP were analyzed using an Immunoscan CCPlus enzyme-linked immunosorbent assay kit (Svar Life Sciences, Malmö, Sweden) at the Department of Immunology and Biotechnology, Medical School, University of Pécs, Hungary, in accordance with the manufacturer's protocols.

**Genomic DNA extraction from blood.** Genomic DNA was isolated using 1 ml DNAzol™ BD reagent (Thermo Fisher Scientific) mixed with 0.5 ml whole blood in accordance with the manufacturer's instructions. Genomic DNA was precipitated with 0.4 ml isopropanol for 5 min at room temperature and then centrifuged. After the removal of the supernatant, the pellet was washed with 0.5 ml DNAzol™ BD reagent and then washed with 1 ml 75% ethanol. The genomic DNA was dried, dissolved in 200 µL Tris EDTA buffer (10mM Tris, pH 8.0, and 1mM EDTA), and quantified with a NanoDrop One (Thermo Fisher Scientific). For genotyping by quantitative polymerase chain reaction, 20 ng genomic DNA was used per reaction.

**Genotyping AIRE SNPs rs878081 (C>T) and rs1003854 (T>C).** We used the predesigned Minor Groove Binder (MGB) TaqMan probe assays (Thermo Fisher Scientific) to detect SNPs in AIRE. The presence of the MGB at the 3' ends of the probe increases the melting temperature, stabilizes the probe–target hybrids, and enables more accurate discrimination between highly homologous allele sequences. After genomic DNA purification from blood samples, multiplex quantitative polymerase chain reaction was performed. FAM and VIC fluorescence intensity signals were detected using the Rotor-Gene 3000 thermal cycler (Corbett Life Science, Qiagen, Crawley, UK). Samples with signals in either the VIC or FAM channels were considered homozygous for the SNP site of interest, whereas samples with signals in both VIC and FAM channels were considered heterozygous for the SNP site of interest.

**Statistical analysis.** Our statistical analyses were performed using SPSS software, version 28.0 (IBM Corp. Released 2021, IBM SPSS Statistics for Windows, Version 28.0; IBM Corp, Armonk, NY, USA). Chi-square tests were used to determine the contingency of genotype distributions with Hardy–Weinberg equilibrium. We also used chi-square tests to check significant differences in the genotype frequencies of control subjects. Clinical continuous variables are represented as mean±standard deviation (SD).

The Shapiro-Wilk test was used to determine normality, and Levene's test was used to analyze the equality of variances. To compare groups for continuous variables, we used independent samples *t*-tests and analysis of variance. To compare medians of non-normally distributed variables, we used Mann-Whitney *U* and Kruskal-Wallis *H* tests with two-tailed significance. Correlation was tested using bivariate correlation with Pearson correlation coefficient and two-tailed significance. Binary logistic regression was used to calculate odds ratios (ORs) with a 95% confidence interval (CI). Statistical significance was determined when *p*-values were less than 0.05. The power was calculated and indicated if it exceeded 0.8 (39).

## Results

**Demographic and clinical characteristics of the study subjects.** The demographic and clinical data of the study subjects are presented in Table I. The mean±SD age of RA patients was 65.38±9.84 years; 85% were female. According to the literature, the mean age of the disease onset is 65 years (1). To minimize the probability of latent cases among controls, we included participants of age 75 or older. Patients and control individuals were all born in Hungary. The RA group demonstrated significantly higher ESR, CRP, Rf immunoglobulin G (IgG) antibody, aCCP levels (all  $p < 0.001$ ) and had 60% seropositivity compared with 100

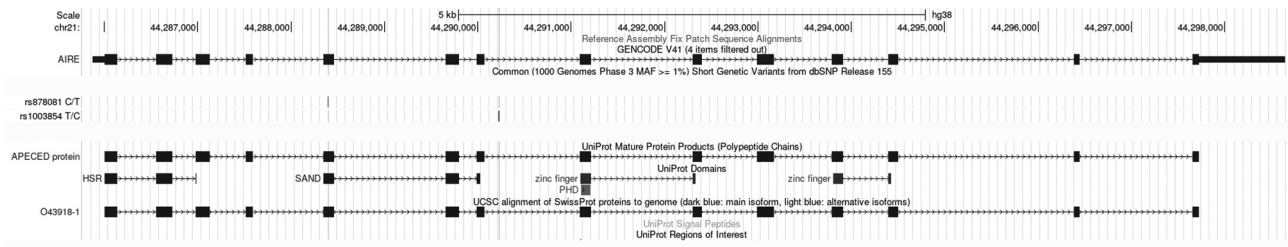


Figure 1. Domain-coding sequences of *AIRE*, and the positions of rs878081 and rs1003854 (image generated in UCSC Genome Browser).

control subjects in whom the aforementioned clinical parameters were in the normal range and whose seronegativity was 100% (34, 40-43). We also calculated DAS28 for patients with RA, where the mean±SD of DAS28 suggested moderate disease activity (44).

**Results of preliminary in silico research.** Using the Ensembl automatic annotation of the human genome sequence and the UCSC Genome Browser as a human genome database, we set up our parameters in the *AIRE* (HGNC Symbol:Acc:HGNC:360) gene sequence (also known as *APECED*, *APS1*, or *PGA1*), which is located in the 21q22.3 chromosomal region between 44,285,838-44,298,648 on the forward strand (26, 27). Without using any restrictions, we identified 23,023 variations (point mutations, indels, and SNPs). To find detectable SNPs, we adjusted the MAF to 0.01-0.5 with at least one representation in each database. We had 68 hits; however, some SNPs had MAFs less than 0.01 in the European (EUR) subpopulation of Utah residents with Northern and Western European ancestry (CEU). We excluded these SNPs by adjusting the MAF to 0.01-0.5 in EUR, because of the low probability of SNP detection in this population. This allowed us to exclude 24 SNPs, and we subsequently shortlisted 44 SNPs in our *in silico* setting. Each of the shortlisted SNPs is found in the 1000 Genomes Project Phase III data set with MAFs 0.01-0.5 and is present in EUR and CEU subpopulations.

Among the 14 exon regions of *AIRE*, exon 5 codes for the SAND domain, which is responsible for transcriptional regulation and DNA binding in mTECs during negative selection. Therefore, we analyzed synonymous polymorphisms in the SAND domain-coding sequence that had MAFs of 0.01-0.5 that might alter messenger RNA (mRNA) splicing, mRNA stability, mRNA structure, protein translation, or cotranslational protein folding (10, 11, 26, 45). In the coding sequence of exon 5, we found one synonymous polymorphism, rs878081, and analyzed its strongly linked intron variant rs1003854, for which the pairwise  $r^2$  value for linkage disequilibrium was 0.912. Domain-coding structural sequences of *AIRE* and the positions of rs878081 and rs1003854 are illustrated in Figure 1.

**Genotype frequencies and Hardy–Weinberg equilibrium.** Genotype frequencies in control groups for rs878081 and rs1003854 were in accordance with Hardy–Weinberg equilibrium. No statistically significant difference was found between the expected and observed genotype frequencies in rs878081 ( $p=0.316$ ) and rs1003854 ( $p=0.214$ ); therefore, our sample collection avoided selection bias, nonrandom sampling, and accumulation of relatives.

**Association between rs878081 (C>T) allelic polymorphism and the risk of rheumatoid arthritis.** Allelic polymorphism in rs878081 was analyzed under allelic, dominant, recessive, codominant heterozygous, codominant homozygous, and overdominant genetic models (46). Under the recessive genetic model, we found statistically significant association with RA, where the genotype frequency of TT homozygotes was significantly higher in patients with RA than in healthy control subjects, supporting RA risk (OR=5.44, 95%CI=1.16-25.52,  $p=0.032$ , power >0.8). The genotype frequency of TT homozygotes was also higher among patients with RA in a codominant homozygous setting than among control subjects; however, the association was barely insignificant (OR=4.68, 95% CI=0.98-22.27,  $p=0.052$ , power >0.8) but suggested the same direction as the recessive model. No other significant associations were found in the allelic, dominant, codominant heterozygous, or overdominant genetic models. Allele and genotype frequencies of rs878081 in patients with RA and control subjects and risk calculations are presented in Table II. The results of allelic discrimination tests for rs878081 in patients with RA and control subjects are illustrated in Figure 2.

**Association between rs1003854 (T>C) allelic polymorphism and the risk of rheumatoid arthritis.** Allelic polymorphism of rs1003854 was analyzed under allelic, dominant, recessive, codominant heterozygous, codominant homozygous, and overdominant genetic models. The genotype frequency in patients with RA was statistically significantly higher in TT homozygotes than in control subjects, supporting RA risk (OR=4.84, 95%CI=1.02-23.02,  $p=0.047$ , power >0.8). No other statistically significant association was found under the

Table II. Distribution of allele and genotype frequencies of AIRE rs878081 in patients with rheumatoid arthritis and control subjects.

AIRE SNP rs878081 (C>T)	Patients with RA N=100 (%)	Control subjects N=100 (%)	OR (95%CI)	p-Value <sup>a</sup>
Allele				
T	46 (23.0)	42 (21.0)	1.12 (0.70-1.80)	0.629
C	154 (77.0)	158 (79.0)		
Genotype				
TT	10 (10.0)	2 (2.0)	4.68 (0.98-22.27)	0.052 <sup>†</sup>
CT	26 (26.0)	38 (38.0)	0.64 (0.34-1.18)	0.154
CC	64 (64.0)	60 (60.0)		
Dominant				
CT+TT	36 (36.0)	40 (40.0)	0.84 (0.47-1.49)	0.629
CC	64 (64.0)	60 (60.0)		
Recessive				
TT	10 (10.0)	2 (2.0)	5.44 (1.16-25.52)	0.032* <sup>†</sup>
CC+CT	90 (90.0)	98 (98.0)		
Overdominant				
CT	26 (26.0%)	38 (38.0%)	0.57 (0.31-1.04)	0.070
CC+TT	74 (74.0%)	62 (62.0%)		

SNP: Single-nucleotide polymorphism; OR: odds ratio; CI: confidence interval. abased on logistic regression; \* $p < 0.05$ , <sup>†</sup>power  $> 0.8$ .

allelic, dominant, codominant heterozygous, codominant homozygous, or overdominant genetic models. Allele and genotype frequencies of rs1003854 in patients with RA and control subjects and risk calculations are presented in Table III. The results of allelic discrimination tests for rs1003854 in patients with RA and control subjects are illustrated in Figure 3.

*Comparison of clinical parameters under the genetic models of rs878081 and rs1003854 in patients with rheumatoid arthritis and control subjects.* Our study aimed to compare ESR, CRP, Rf, aCCP, and DAS28 levels between the genotype subgroups of genetic models of rs878081 and rs1003854 in patients with RA and control subjects. The first stage was designed to find significant differences by comparing group means, and the second stage was designed to reveal correlations between clinical parameters and genotypes.

In patients with RA, ESRs were significantly higher in the CC subgroup compared to TT homozygotes under the codominant homozygous model of rs1003854 ( $p=0.048$ ). The CC subgroup demonstrated significantly higher CRP levels compared to TT homozygotes in the codominant homozygous model and the TT+TC subgroup in the recessive model ( $p=0.043$  and  $p=0.029$ , respectively). Furthermore, the CC homozygote patients with RA were significantly correlated with CRP under the codominant homozygous and recessive models ( $p=0.006$  and  $p=0.002$ , respectively;  $r=0.350$  and  $r=0.338$ , respectively). No other significant differences were found in other clinical variables

and genetic models of rs878081 and rs1003854. These results are illustrated in Table IV.

## Discussion

RA is a complex, multifactorial disorder in which genetic and environmental causes contribute to disease initiation (1-3). *In vitro* and *in vivo* studies have demonstrated that presentation of peripheral tissue self-antigens is quantitatively determined by AIRE transcription, and AIRE SNPs may cause clonal survival (19, 47). Autoreactive CD4<sup>+</sup> T cells are recruited to the synovium where they initiate an inflammatory environment by the production of interferon-gamma, tumor necrosis factor-alpha, interleukins (ILs), proinflammatory cytokines, and chemokines, causing tissue destruction (48-50). The inflammatory environment activates fibroblast-like synoviocytes (FLSs) to produce chemokines in order to attract more autoreactive CD4<sup>+</sup> T cells, creating a vicious circle (50, 51). This was demonstrated by an *in vitro* study by Bergström *et al.* in which IL-1 $\beta$  and tumor necrosis factor-stimulated FLSs from patients with RA and in which there was 26-fold statistically significantly higher CXCL10 and CCL8 production compared with cells with AIRE expression silenced by sense siRNA (51). Nonstimulated FLSs did not produce chemokines (51). Based on our current understanding, AIRE SNPs are important in two ways: they allow the escape of autoreactive CD4<sup>+</sup> T cells during central tolerance, and they cause FLSs to produce chemokines to attract increasingly more self-reactive CD4<sup>+</sup> cells, thereby maintaining the hyperplastic inflammatory environment (19,



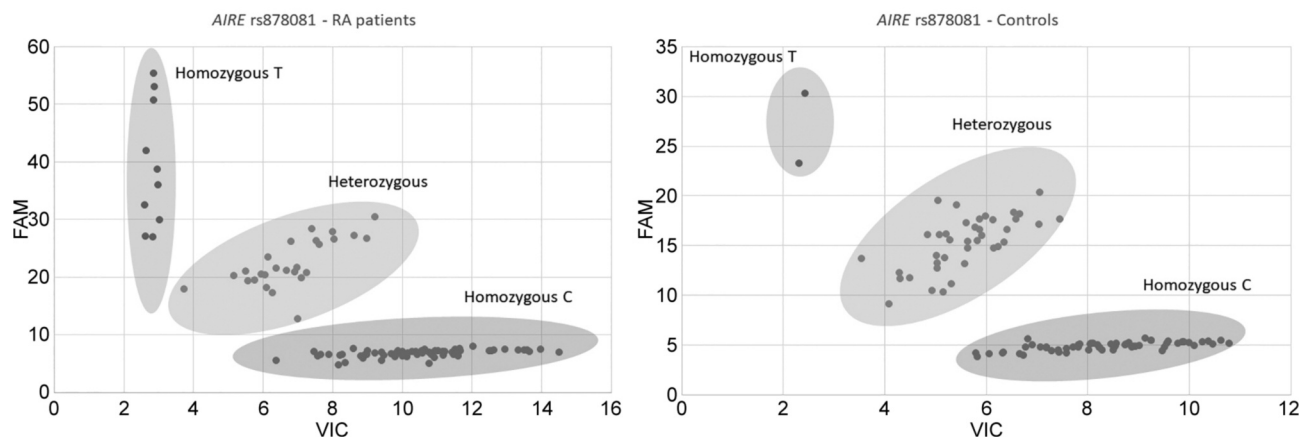


Figure 2. Result of allelic discrimination tests for rs878081 in patients with rheumatoid arthritis and control subjects.

Table III. Distribution of allele and genotype frequencies of AIRE rs1003854 in patients with rheumatoid arthritis and control subjects.

AIRE SNP rs1003854 (T>C)	Patients with RA N=100 (%)	Control subjects N=100 (%)	OR (95%CI)	p-Value <sup>a</sup>
Allele				
C	47 (23.5)	45 (22.5)	1.05 (0.66-1.68)	0.812
T	153 (76.5)	155 (77.5)		
Genotype				
CC	9 (9.0)	2 (2.0)	4.13 (0.85-19.96)	0.077 <sup>†</sup>
TC	29 (29.0)	41 (41.0)	0.65 (0.35-1.18)	0.157
TT	62 (62.0)	57 (57.0)		
Dominant				
TC+CC	38 (38.0)	43 (43.0)	0.81 (0.46-1.43)	0.471
TT	62 (62.0)	57 (57.0)		
Recessive				
CC	9 (9.0)	2 (2.0)	4.84 (1.02-23.02)	0.047 <sup>*†</sup>
TT+TC	91 (91.0)	98 (98.0)		
Overdominant				
TC	29 (29.0)	41 (41.0)	0.58 (0.32-1.05)	0.076
TT+CC	71 (71.0)	59 (59.0)		

SNP: Single-nucleotide polymorphism; OR: odds ratio; CI: confidence interval; <sup>a</sup>based on logistic regression; \* $p < 0.05$ , <sup>†</sup>power  $> 0.8$ .

51). Hence, synonymous SNPs and their strongly related intron variants can be associated with disease activity. In Asian populations, a strong association was found between rs2075776, rs760426, and RA in different genetic models (24, 52-55). Furthermore, Terao *et al.* demonstrated genotype-specific AIRE expression in lymphoblastoid cells using the Gene Expression Omnibus database and raised the idea that different genotypes may alter disease activity (56). Allelic polymorphisms of rs878081 have been analyzed in Chinese, Egyptian, and Spanish populations (25, 56, 57). Yang *et al.* measured AIRE expression in blood lymphocytes and found significant differences between rs878081 TT and CC homozygotes in Asian populations (56).

Our investigation is the first attempt to analyze different genotypes and RA risk among Hungarian-born, European individuals in allelic, dominant, recessive, codominant heterozygous, codominant homozygous, and overdominant genetic models combined with clinical parameters, such as ESR and CRP. After our preliminary *in silico* selection of SNPs, our results showed in a recessive model (TT vs. CC+CT) that rs878081 TT is significantly associated with RA risk ( $p=0.032$ ). The codominant homozygous model (TT vs. CC) provided the same result; however, only borderline significance was detected ( $p=0.052$ ). These results made us consider that two copies of the minor alleles may be necessary for a quantitative association with RA risk, as the

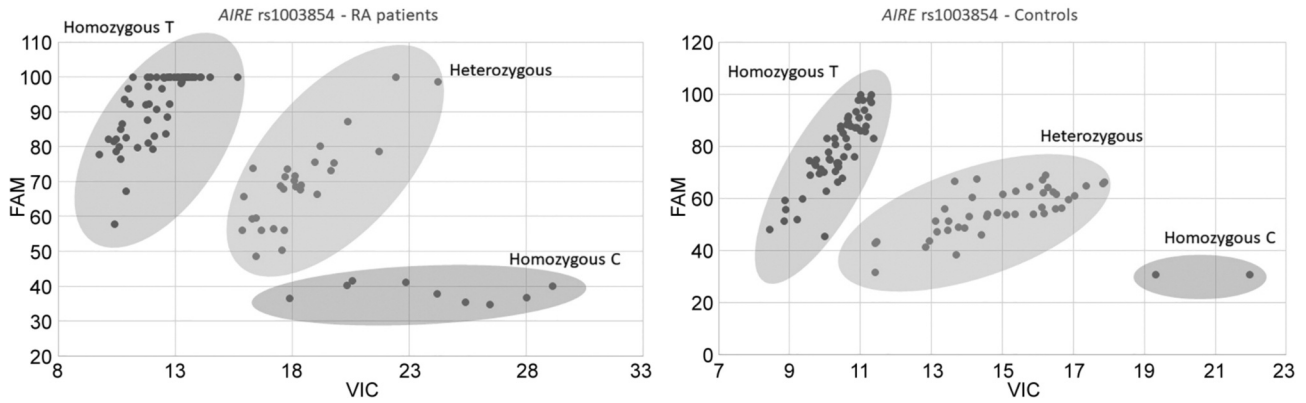


Figure 3. Result of allelic discrimination tests for rs1003854 in patients with rheumatoid arthritis and control subjects.

Table IV. C-reactive protein levels among the genotypic subgroups of genetic models of rs1003854 within patients with rheumatoid arthritis and control subjects.

AIRE SNP rs1003854 (T>C)	Patients with RA				Control subjects			
	CRP	p-Value <sup>a</sup>	Pearson correlation	p-Value <sup>b</sup>	CRP	p-Value <sup>c</sup>	Pearson correlation	p-Value <sup>d</sup>
Genotype								
CC	58.41±47.30	0.043*	0.350	0.006**	6.00±0.00	0.486	0.002	0.990
TC	22.04±23.03	0.094	-0.086	0.469	6.49±3.14	0.486	0.099	0.337
TT	27.21±25.75				5.98±2.06			
Dominant								
TC+CC	32.47±35.36	0.687	0.092	0.416	6.47±3.06	0.430	0.096	0.347
TT	26.87±25.55				5.98±2.06			
Recessive								
CC	58.41±47.30	0.029*	0.338	0.002**	6.00±0.00	0.571	-0.011	0.915
TT+TC	25.56±24.95				6.19±2.56			
Overdominant								
TC	22.09±23.57	0.050	-0.132	0.239	6.49±3.14	0.527	0.100	0.329
TT+CC	31.01±30.69				5.98±2.02			

SNP: Single-nucleotide polymorphism; RA: rheumatoid arthritis; CRP: C-reactive protein; <sup>a</sup>genotype subgroup comparison of CRP levels in patients with RA in each genetic model; <sup>b</sup>Pearson correlation of CRP with genotype subgroups of patients with RA in each genetic model; <sup>c</sup>genotype subgroup comparison of CRP levels in control subjects in each genetic model; <sup>d</sup>Pearson correlation of CRP with genotype subgroups of control subjects in each genetic model; \*p<0.05, \*\*p<0.01; Clinical parameters are reported as mean±SD.

SNP is synonymous and resides in exon 5. Exon 5 is in the SAND coding sequence of AIRE that is responsible for DNA-binding and cooperation with transcriptional factors, so allelic polymorphisms in this region may functionally determine AIRE production and affect central tolerance (10-13).

Our study analyzed intron variant rs1003854, where significant correlations were found between clinical parameters and genotypes in recessive (CC vs. TT+TC) and homozygous codominant (CC vs. TT) genetic models. The allelic polymorphism rs1003854 resides in intron 7 and may modify the transcription of AIRE by altering alternative-splicing mechanisms with the process of intron-mediated

enhancement (58). Interestingly, after we determined RA risk for SNP rs1003854, we concluded that only the recessive and codominant genetic models demonstrated significantly higher levels of CRP (p=0.029 and p=0.043, respectively) and positive, significant correlation with CRP (p=0.002 and p=0.006, respectively) when patients with RA were homozygous for the minor C allele. This finding supports the notion that CRP levels vary in genetic models of AIRE rs1003854 among patients with RA, reflecting a possible association between AIRE genotypes and RA disease activity, where chemokines (such as CXCL10, CCL8, and CCL5) are essential and are strongly correlated with CRP levels (59-61).

To conclude, rs878081(C>T) and rs1003854(T>C) SNPs together with ESR and CRP levels would be useful as potential biomarkers to determine RA susceptibility, follow-up, and prognosis. Before future clinical application, further studies are necessary. To continue contributing to this research area, we intend to carry out studies on larger cohorts of patients with RA and control subjects and to consider multiple loci as well as possible molecular partners of AIRE.

## Conclusion

In conclusion, our data demonstrate the effects of gene polymorphisms in rs878081 and rs1003854 on the risk of RA in the European population, where two copies of the minor alleles and the related homozygous genotypes provided susceptibility. Analyses of the clinical parameters ESR and CRP supported our assumption that the homozygous subgroup for the minor allele of patients with RA may have higher ESR and CRP levels, which might reflect disease activity. Therefore, these clinical parameters may be suitable to detect at-risk populations in the future. Based on medical history, a previous family history of RA may be an indication for a risk genotype test, where the detection would prospectively allow a follow-up of the healthy individual and possible preventive physiotherapy and drug therapy.

## Conflicts of Interest

The Authors declare no conflicts of interest in relation to this study.

## Authors' Contributions

Bálint Bérczi: conceptualization, methodology, formal analysis, investigation, writing – original draft preparation, writing – review & editing, funding acquisition. Nóra Nusser: methodology, resources, writing—original draft preparation, writing—review & editing. Iván Péter: resources, writing—original draft preparation, writing—review & editing. Balázs Németh: resources, writing—original draft preparation, writing—review & editing. Zoltán Gyöngyi: conceptualization, investigation, methodology, data curation, writing—original draft preparation, writing—review & editing, supervision, funding acquisition. All Authors have read and agreed to the published version of the manuscript.

## Acknowledgements

The Authors thank all the patients and control subjects for participation in the study. The Authors would like to also acknowledge the conscientious laboratory assistance by the team of the Harkány Thermal Rehabilitation Centre, Hungary. This research was funded by the ÁOK János Szolcsányi Research Fund supported by the Medical School, University of Pécs, Hungary. The study was conducted in accordance with the guidelines of the Declaration of Helsinki, and the protocol was approved by the Secretary of the Medical Research Council, Scientific and Research Ethics Committee of Hungary (11871-7/2018/EÜIG of 24 April 2018, and

57142-5/2022/EÜIG of 8 November 2022). Informed consent was obtained from all subjects involved in the study.

## References

- 1 Scott DL, Wolfe F, Huizinga TW: Rheumatoid arthritis. *Lancet* 376(9746): 1094-1108, 2010. DOI: 10.1016/S0140-6736(10)60826-4
- 2 Starr TK, Jameson SC, Hogquist KA: Positive and negative selection of T cells. *Annu Rev Immunol* 21(1): 139-176, 2003. DOI: 10.1146/annurev.immunol.21.120601.141107
- 3 Mellado M, Martínez-Muñoz L, Cascio G, Lucas P, Pablos JL, Rodríguez-Frade JM: T cell migration in rheumatoid arthritis. *Front Immunol* 6: 384, 2015. DOI: 10.3389/fimmu.2015.00384
- 4 McInnes IB, Schett G: The pathogenesis of rheumatoid arthritis. *N Engl J Med* 365(23): 2205-2219, 2011. DOI: 10.1056/NEJMra1004965
- 5 Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, Alfredsson L, Padyukov L, Klareskog L, Worthington J, Siminovitch KA, Bae SC, Plenge RM, Gregersen PK, de Bakker PI: Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 44(3): 291-296, 2012. DOI: 10.1038/ng.1076
- 6 MacKay K, Eyre S, Myerscough A, Milicic A, Barton A, Laval S, Barrett J, Lee D, White S, John S, Brown MA, Bell J, Silman A, Ollier W, Wordsworth P, Worthington J: Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum* 46(3): 632-639, 2002. DOI: 10.1002/art.10147
- 7 Kurkó J, Besenyei T, Laki J, Glant TT, Mikecz K, Szekanecz Z: Genetics of rheumatoid arthritis - a comprehensive review. *Clin Rev Allergy Immunol* 45(2): 170-179, 2013. DOI: 10.1007/s12016-012-8346-7
- 8 Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Kawaguchi T, Stahl EA, Kurreeman FA, Nishida N, Ohmiya H, Myouzen K, Takahashi M, Sawada T, Nishioka Y, Yukioka M, Matsubara T, Wakitani S, Teshima R, Tohma S, Takasugi K, Shimada K, Murasawa A, Honjo S, Matsuo K, Tanaka H, Tajima K, Suzuki T, Iwamoto T, Kawamura Y, Tani H, Okazaki Y, Sasaki T, Gregersen PK, Padyukov L, Worthington J, Siminovitch KA, Lathrop M, Taniguchi A, Takahashi A, Tokunaga K, Kubo M, Nakamura Y, Kamatani N, Mimori T, Plenge RM, Yamanaka H, Momohara S, Yamada R, Matsuda F, Yamamoto K: Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet* 44(5): 511-516, 2012. DOI: 10.1038/ng.2231
- 9 Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, Graham RR, Manoharan A, Ortmann W, Bhangale T, Denny JC, Carroll RJ, Eyler AE, Greenberg JD, Kremer JM, Pappas DA, Jiang L, Yin J, Ye L, Su DF, Yang J, Xie G, Keystone E, Westra HJ, Esko T, Metspalu A, Zhou X, Gupta N, Mirel D, Stahl EA, Diogo D, Cui J, Liao K, Guo MH, Myouzen K, Kawaguchi T, Coenen MJ, van Riel PL, van de Laar MA, Guchelaar HJ, Huizinga TW, Dieudé P, Mariette X, Bridges SL Jr, Zernakova A, Toes RE, Tak PP, Miceli-Richard C, Bang SY, Lee HS, Martin J, Gonzalez-Gay MA, Rodriguez-Rodriguez L, Rantapää-Dahlqvist S, Arlestig L, Choi HK, Kamatani Y, Galan P, Lathrop M, RACI consortium, GARNET consortium, Eyre S, Bowes J, Barton A, de Vries N, Moreland LW, Criswell LA, Karlson EW, Taniguchi A, Yamada

- R, Kubo M, Liu JS, Bae SC, Worthington J, Padyukov L, Klareskog L, Gregersen PK, Raychaudhuri S, Stranger BE, De Jager PL, Franke L, Visscher PM, Brown MA, Yamanaka H, Mimori T, Takahashi A, Xu H, Behrens TW, Siminovitch KA, Momohara S, Matsuda F, Yamamoto K, Plenge RM: Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506(7488): 376-381, 2014. DOI: 10.1038/nature12873
- 10 Mathis D, Benoist C: Aire. *Annu Rev Immunol* 27(1): 287-312, 2009. DOI: 10.1146/annurev.immunol.25.022106.141532
- 11 Bottomley MJ, Collard MW, Huggenvik JI, Liu Z, Gibson TJ, Sattler M: The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. *Nat Struct Biol* 8(7): 626-633, 2001. DOI: 10.1038/89675
- 12 Ragvin A, Valvatne H, Erdal S, Årskog V, Tufteland KR, Breen K, Øyan AM, Eberharter A, Gibson TJ, Becker PB, Aasland R: Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. *J Mol Biol* 337(4): 773-788, 2004. DOI: 10.1016/j.jmb.2004.01.051
- 13 Plevin MJ, Mills MM, Ikura M: The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem Sci* 30(2): 66-69, 2005. DOI: 10.1016/j.tibs.2004.12.001
- 14 Akirav EM, Ruddle NH, Herold KC: The role of AIRE in human autoimmune disease. *Endocrinol* 7(1): 25-33, 2011. DOI: 10.1038/nrendo.2010.200
- 15 Perera J, Liu X, Zhou Y, Joseph NE, Meng L, Turner JR, Huang H: Insufficient autoantigen presentation and failure of tolerance in a mouse model of rheumatoid arthritis. *Arthritis Rheum* 65(11): 2847-2856, 2013. DOI: 10.1002/art.38085
- 16 Walser-Kuntz DR, Weyand CM, Weaver AJ, O'Fallon WM, Goronzy JJ: Mechanisms underlying the formation of the T cell receptor repertoire in rheumatoid arthritis. *Immunity* 2(6): 597-605, 1995. DOI: 10.1016/1074-7613(95)90004-7
- 17 Kouskoff V, Korganow AS, Duchatelle V, Degott C, Benoist C, Mathis D: Organ-specific disease provoked by systemic autoimmunity. *Cell* 87(5): 811-822, 1996. DOI: 10.1016/s0092-8674(00)81989-3
- 18 Goronzy JJ, Weyand CM: Thymic function and peripheral T-cell homeostasis in rheumatoid arthritis. *Trends Immunol* 22(5): 251-255, 2001. DOI: 10.1016/s1471-4906(00)01841-x
- 19 Lovewell TR, McDonagh AJ, Messenger AG, Azzouz M, Tazi-Ahnini R: The AIRE -230Y polymorphism affects AIRE transcriptional activity: Potential influence on AIRE function in the thymus. *PLoS One* 10(5): e0127476, 2015. DOI: 10.1371/journal.pone.0127476
- 20 Tazi-Ahnini R, McDonagh AJ, Wengraf DA, Lovewell TR, Vasilopoulos Y, Messenger AG, Cork MJ, Gawkrödger DJ: The autoimmune regulator gene (*AIRE*) is strongly associated with vitiligo. *Br J Dermatol* 159(3): 591-6, 2008. DOI: 10.1111/j.1365-2133.2008.08718.x
- 21 Wengraf DA, McDonagh AJ, Lovewell TR, Vasilopoulos Y, Macdonald-Hull SP, Cork MJ, Messenger AG, Tazi-Ahnini R: Genetic analysis of autoimmune regulator haplotypes in alopecia areata. *Tissue Antigens* 71(3): 206-212, 2008. DOI: 10.1111/j.1399-0039.2007.00992.x
- 22 Eriksson D, Bianchi M, Landegren N, Dalin F, Skov J, Hultin-Rosenberg L, Mathioudaki A, Nordin J, Hallgren Å, Andersson G, Tandre K, Rantapää Dahlqvist S, Söderkvist P, Rönnblom L, Hulting AL, Wahlberg J, Dahlqvist P, Ekwall O, Meadows JRS, Lindblad-Toh K, Bensing S, Rosengren Pielberg G, Kämpe O: Common genetic variation in the autoimmune regulator (*AIRE*) locus is associated with autoimmune Addison's disease in Sweden. *Sci Rep* 8(1): 8395, 2018. DOI: 10.1038/s41598-018-26842-2
- 23 Ferrera F, Rizzi M, Spreccacenero B, Balestra P, Sessarego M, Di Carlo A, Filaci G, Gabrielli A, Ravazzolo R, Indiveri F: *AIRE* gene polymorphisms in systemic sclerosis associated with autoimmune thyroiditis. *Clin Immunol* 122(1): 13-17, 2007. DOI: 10.1016/j.clim.2006.09.013
- 24 Bérczi B, Gerencsér G, Farkas N, Hegyi P, Veres G, Bajor J, Czopf L, Alizadeh H, Rakonczay Z, Vigh É, Erőss B, Szemes K, Gyöngyi Z: Association between *AIRE* gene polymorphism and rheumatoid arthritis: a systematic review and meta-analysis of case-control studies. *Sci Rep* 7(1): 14096, 2017. DOI: 10.1038/s41598-017-14375-z
- 25 García-Lozano JR, Torres-Agrela B, Montes-Cano MA, Ortiz-Fernández L, Conde-Jaldón M, Teruel M, García A, Núñez-Roldán A, Martín J, González-Escribano MF: Association of the *AIRE* gene with susceptibility to rheumatoid arthritis in a European population: a case control study. *Arthritis Res Ther* 15(1): R11, 2013. DOI: 10.1186/ar4141
- 26 Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, Berry A, Bhai J, Bignell A, Billis K, Boddu S, Brooks L, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, Donaldson S, El Houdaigui B, El Naboulsi T, Fatima R, Giron CG, Genev T, Martinez JG, Guizarro-Clarke C, Gymer A, Hardy M, Hollis Z, Hourlier T, Hunt T, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Marugán JC, Mohanan S, Mushtaq A, Naven M, Ogeh DN, Parker A, Parton A, Perry M, Piližota I, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Pérez-Silva JG, Stark W, Steed E, Sutinen K, Sukumaran R, Sumathipala D, Suner MM, Szpak M, Thormann A, Tricomi FF, Urbina-Gómez D, Veidenberg A, Walsh TA, Walts B, Willhoft N, Winterbottom A, Wass E, Chakiachvili M, Flint B, Frankish A, Giorgetti S, Haggerty L, Hunt SE, Ilesley GR, Loveland JE, Martin FJ, Moore B, Mudge JM, Muffato M, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Dyer S, Harrison PW, Howe KL, Yates AD, Zerbino DR, Flicek P: Ensembl 2022. *Nucleic Acids Res* 50(D1): D988-D995, 2022. DOI: 10.1093/nar/gkab1049
- 27 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: The human genome browser at UCSC. *Genome Res* 12(6): 996-1006, 2002. DOI: 10.1101/gr.229102
- 28 Cavalli-Sforza LL, Bodmer WF: The genetics of human populations. 1<sup>st</sup> ed. San Francisco, CA, USA, Dover Publications, pp. 41-118, 1971.
- 29 Fairley S, Lowy-Gallego E, Perry E, Flicek P: The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res* 48(D1): D941-D947, 2020. DOI: 10.1093/nar/gkz836
- 30 Sherry ST, Ward M, Sirotkin K: dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 9: 677-679, 1999. DOI: 10.1101/gr.9.8.677
- 31 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek



- M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database Consortium, Neale BM, Daly MJ, MacArthur DG: The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581(7809): 434-443, 2020. DOI: 10.1038/s41586-020-2308-7
- 32 Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, Jain D, Argos M, Arnett DK, Avery C, Barnes KC, Becker LC, Bien SA, Bis JC, Blangero J, Boerwinkle E, Bowden DW, Buyske S, Cai J, Cho MH, Choi SH, Choquet H, Cupples LA, Cushman M, Daya M, de Vries PS, Ellinor PT, Faraday N, Fornage M, Gabriel S, Ganesh SK, Graff M, Gupta N, He J, Heckbert SR, Hidalgo B, Hodonsky CJ, Irvin MR, Johnson AD, Jorgenson E, Kaplan R, Kardia SLR, Kelly TN, Kooperberg C, Lasky-Su JA, Loos RJJ, Lubitz SA, Mathias RA, McHugh CP, Montgomery C, Moon JY, Morrison AC, Palmer ND, Pankratz N, Papanicolaou GJ, Peralta JM, Peyser PA, Rich SS, Rotter JI, Silverman EK, Smith JA, Smith NL, Taylor KD, Thornton TA, Tiwari HK, Tracy RP, Wang T, Weiss ST, Weng LC, Wiggins KL, Wilson JG, Yanek LR, Zöllner S, North KE, Auer PL, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Hematology & Hemostasis Working Group, Raffield LM, Reiner AP, Li Y: Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet* 15(12): e1008500, 2019. DOI: 10.1371/journal.pgen.1008500
- 33 Arnett FC, Edworthy SM, Bloch DA, Mcshane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger TA, Mitchell DM, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG: The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 31(3): 315-324, 1988. DOI: 10.1002/art.1780310302
- 34 Ingegnoli F, Castelli R, Gualtierotti R: Rheumatoid factors: clinical applications. *Dis Markers* 35(6): 727-734, 2013. DOI: 10.1155/2013/726598
- 35 Niewold TB, Harrison MJ, Paget SA: Anti-CCP antibody testing as a diagnostic and prognostic tool in rheumatoid arthritis. *QJM* 100(4): 193-201, 2007. DOI: 10.1093/qjmed/hcm015
- 36 Steiner G, Smolen J: Autoantibodies in rheumatoid arthritis and their clinical significance. *Arthritis Res* 4 Suppl 2(Suppl 2): S1-S5, 2002. DOI: 10.1186/ar551
- 37 Graham JE, Granger CV, Karmarkar AM, Deutsch A, Niewczyk P, Divita MA, Ottenbacher KJ: The Uniform Data System for Medical Rehabilitation: report of follow-up information on patients discharged from inpatient rehabilitation programs in 2002-2010. *Am J Phys Med Rehabil* 93(3): 231-244, 2014. DOI: 10.1097/PHM.0b013e3182a92c58
- 38 Jou JM, Lewis SM, Briggs C, Lee SH, De La Salle B, McFadden S, International Council for Standardization in Haematology: ICSH review of the measurement of the erythrocyte sedimentation rate. *Int J Lab Hematol* 33(2): 125-132, 2011. DOI: 10.1111/j.1751-553X.2011.01302.x
- 39 Skol AD, Scott LJ, Abecasis GR, Boehnke M: Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38(2): 209-213, 2006. DOI: 10.1038/ng1706
- 40 Sharland DE: Erythrocyte sedimentation rate: The normal range in the elderly. *J Am Geriatr Soc* 28(8): 346-348, 1980. DOI: 10.1111/j.1532-5415.1980.tb01096.x
- 41 Miller A, Green M, Robinson D: Simple rule for calculating normal erythrocyte sedimentation rate. *Br Med J (Clin Res Ed)* 286(6361): 266, 1983. DOI: 10.1136/bmj.286.6361.266
- 42 Wener MH, Daum PR, McQuillan GM: The influence of age, sex, and race on the upper reference limit of serum C-reactive protein concentration. *J Rheumatol* 27(10): 2351-2359, 2000.
- 43 Coenen D, Verschuere P, Westhovens R, Bossuyt X: Technical and diagnostic performance of 6 assays for the measurement of citrullinated protein/peptide antibodies in the diagnosis of rheumatoid arthritis. *Clin Chem* 53(3): 498-504, 2007. DOI: 10.1373/clinchem.2006.078063
- 44 Fuchs HA, Brooks RH, Callahan LF, Pincus T: A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis. *Arthritis Rheum* 32(5): 531-537, 1989. DOI: 10.1002/anr.1780320504
- 45 Zeng Z, Bromberg Y: Predicting functional effects of synonymous variants: a systematic review and perspectives. *Front Genet* 10: 914, 2019. DOI: 10.3389/fgene.2019.00914
- 46 Horita N, Kaneko T: Genetic model selection for a case-control study and a meta-analysis. *Meta Gene* 5: 1-8, 2015. DOI: 10.1016/j.mgene.2015.04.003
- 47 Chan AY, Anderson MS: Central tolerance to self revealed by the autoimmune regulator. *Ann N Y Acad Sci* 1356(1): 80-89, 2015. DOI: 10.1111/nyas.12960
- 48 Tuncel J, Holmberg J, Haag S, Hopkins MH, Wester-Rosenlöf L, Carlsen S, Olofsson P, Holmdahl R: Self-reactive T cells induce and perpetuate chronic relapsing arthritis. *Arthritis Res Ther* 22(1): 95, 2020. DOI: 10.1186/s13075-020-2104-7
- 49 Zhu L, Ji F, Wang Y, Zhang Y, Liu Q, Zhang JZ, Matsushima K, Cao Q, Zhang Y: Synovial autoreactive T cells in rheumatoid arthritis resist IDO-mediated inhibition. *J Immunol* 177(11): 8226-8233, 2006. DOI: 10.4049/jimmunol.177.11.8226
- 50 Yap HY, Tee SZ, Wong MM, Chow SK, Peh SC, Teow SY: Pathogenic role of immune cells in rheumatoid arthritis: Implications in clinical treatment and biomarker development. *Cells* 7(10): 161, 2018. DOI: 10.3390/cells7100161
- 51 Bergström B, Lundqvist C, Vasileiadis GK, Carlsten H, Ekwall O, Ekwall AH: The rheumatoid arthritis risk gene AIRE is induced by cytokines in fibroblast-like synoviocytes and augments the pro-inflammatory response. *Front Immunol* 10: 1384, 2019. DOI: 10.3389/fimmu.2019.01384
- 52 Terao C, Yamada R, Ohmura K, Takahashi M, Kawaguchi T, Kochi Y, Human Disease Genomics Working Group, RA Clinical and Genetic Study Consortium, Okada Y, Nakamura Y, Yamamoto K, Melchers I, Lathrop M, Mimori T, Matsuda F: The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Hum Mol Genet* 20(13): 2680-2685, 2011. DOI: 10.1093/hmg/ddr161
- 53 Shao S, Li XR, Cen H, Yin ZS: Association of AIRE polymorphisms with genetic susceptibility to rheumatoid arthritis in a Chinese population. *Inflammation* 37(2): 495-499, 2014. DOI: 10.1007/s10753-013-9763-3
- 54 Feng ZJ, Zhang SL, Wen HF, Liang Y: Association of rs2075876 polymorphism of AIRE gene with rheumatoid arthritis risk. *Hum*

- Immunol 76(4): 281-285, 2015. DOI: 10.1016/j.humimm.2015.01.026
- 55 Li X, Li T, Chen M, Chai Y: [Association of AIRE gene polymorphisms with susceptibility to rheumatoid arthritis among ethnic Han Chinese from Shaanxi]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 33(3): 373-377, 2016. DOI: 10.3760/cma.j.issn.1003-9406.2016.03.022
- 56 Yang H, Li J, Jiang L, Jiang X, Zhou X, Xu N: The rs878081 polymorphism of AIRE gene increases the risk of rheumatoid arthritis in a Chinese Han population: a case-control study. *Braz J Med Biol Res* 51(12): e7944, 2018. DOI: 10.1590/1414-431X20187944
- 57 A Moneim NH, Mansour MF, Omar HH, Fouad MM, Metwally L, El-Abaseri TB, Abdelnaby MM: Association of autoimmune regulator gene polymorphism with susceptibility to rheumatoid arthritis in Egyptian population. *Immunol Res* 68(2): 90-96, 2020. DOI: 10.1007/s12026-020-09127-7
- 58 Rose AB: Intron-Mediated Regulation of Gene Expression. *Curr Top Microbiol Immunol* 326: 277-290, 2008. DOI: 10.1007/978-3-540-76776-3\_15
- 59 Pandya JM, Lundell AC, Andersson K, Nordström I, Theander E, Rudin A: Blood chemokine profile in untreated early rheumatoid arthritis: CXCL10 as a disease activity marker. *Arthritis Res Ther* 19(1): 20, 2017. DOI: 10.1186/s13075-017-1224-1
- 60 Han JH, Suh CH, Jung JY, Nam JY, Kwon JE, Yim H, Kim HA: Association of CXCL10 and CXCL13 levels with disease activity and cutaneous manifestation in active adult-onset Still's disease. *Arthritis Res Ther* 17(1): 260, 2015. DOI: 10.1186/s13075-015-0773-4
- 61 Ruth JH, Rottman JB, Katschke KJ Jr, Qin S, Wu L, LaRosa G, Ponath P, Pope RM, Koch AE: Selective lymphocyte chemokine receptor expression in the rheumatoid joint. *Arthritis Rheum* 44(12): 2750-2760, 2001. DOI: 10.1002/1529-0131(200112)44:12<2750::aid-art462>3.0.co;2-c

*Received October 25, 2023*  
*Revised November 21, 2023*  
*Accepted November 22, 2023*

## Article

# Genetic Polymorphisms in Exon 5 and Intron 5 and 7 of *AIRE* Are Associated with Rheumatoid Arthritis Risk in a Hungarian Population

Bálint Bérczi <sup>1</sup>, Nóra Nusser <sup>2</sup>, Iván Péter <sup>2</sup>, Balázs Németh <sup>1,2</sup>, Ágota Kulisch <sup>3</sup>, Zsuzsanna Kiss <sup>3</sup> and Zoltán Gyöngyi <sup>1,\*</sup>

<sup>1</sup> Department of Public Health Medicine, Medical School, University of Pécs, Szigeti út 12, 7624 Pécs, Hungary; berczi.balint84@gmail.com (B.B.); balazs.nemeth@aok.pte.hu (B.N.)

<sup>2</sup> Harkány Thermal Rehabilitation Centre, Zsigmondy Sétány 1, 7815 Harkány, Hungary; orvosigazgato@harkanykorhaz.hu (N.N.); foigazgato@harkanykorhaz.hu (I.P.)

<sup>3</sup> St. Andrew Hospital for Rheumatology and Medicinal Spa of Hévíz, Dr. Schulhof Vilmos Sétány. 1, 8380 Hévíz, Hungary; agotakulich@gmail.com (Á.K.); zsuzsanna92kiss@gmail.com (Z.K.)

\* Correspondence: zoltan.gyongyi@aok.pte.hu; Tel.: +36-72-501-500 (ext. 38107)

**Simple Summary:** It is a fundamental question: what is the earliest commencement of autoimmune diseases? In recent years, one process deemed to be essential is central self-tolerance during the neonatal period, when the immune system selects and destroys autoimmune T-cells. This mechanism is genetically governed by the master gene *AIRE*. Genetic variations in this gene may cause a less efficient clearance of autoimmune T-cells, causing severe autoimmune diseases decades later. According to numerous research studies and our preliminary results, genetic polymorphisms in this master gene may be at the root of several autoimmune diseases. Although the interaction of genes and disease has been well studied in Asian populations, results are scarce in European populations. Our study is intended to analyse the association between genetic variations in this master gene and rheumatoid arthritis in subjects of Hungarian origin. We successfully identified an association between polymorphisms and the disease, reinforcing the notion that immunologic events during the neonatal period may contribute to autoimmune diseases. A further strength of our research is that in the future, because of its neonatal involvement, *AIRE* can be used as a risk biomarker in cases in which other regular markers cannot predict the disease risk.

**Abstract:** Background: Rheumatoid arthritis (RA) is chronically persistent synovitis and systemic inflammation. Although multiple contributors are detected, only one is pivotal in the neonatal period: the negative selection of autoimmune naïve T-cells by the autoimmune regulator (*AIRE*) transcriptional factor. Methods: Single-nucleotide polymorphisms (SNPs) in the DNA-binding site of *AIRE* may determine its function and expression. We intended to analyse site-specific allelic polymorphisms in two exon (rs878081 and rs1055311) and three intron (rs1003853, rs2075876, and rs1003854) loci with an RA risk. Our analytical case-control study analysed 270 RA patients and 322 control subjects in five different genetic models using quantitative real-time PCR (qPCR) with TaqMan<sup>®</sup> assays. Results: Statistically significant differences were found between the odds of allelic polymorphisms in the loci of rs878081, rs1003854, and rs1003853 among the controls and RA patients, and the disease activity seemed to be significantly associated with the genotypic subgroups of rs878081 and rs1055311. Our in silico analysis supported this, suggesting that allele-specific alterations in the binding affinity of transcriptional factor families might determine RA activity. Conclusion: Our findings highlight the involvement of neonatal self-tolerance in RA pathogenesis, providing novel insights into disease development and paving the way for an analysis of further site-specific genetic polymorphisms in *AIRE* to expand the intervention time for RA.

**Keywords:** single-nucleotide polymorphism; central tolerance; rheumatoid arthritis; autoimmune disease



**Citation:** Bérczi, B.; Nusser, N.; Péter, I.; Németh, B.; Kulisch, Á.; Kiss, Z.; Gyöngyi, Z. Genetic Polymorphisms in Exon 5 and Intron 5 and 7 of *AIRE* Are Associated with Rheumatoid Arthritis Risk in a Hungarian Population. *Biology* **2024**, *13*, 439. <https://doi.org/10.3390/biology13060439>

Academic Editor: Paul Potter

Received: 12 May 2024

Revised: 6 June 2024

Accepted: 12 June 2024

Published: 15 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Rheumatoid arthritis (RA) is a chronic autoimmune dysfunction associated with synovial inflammation in which consequent tissue destruction causes functional impairment and disability [1]. In medium- and high-income countries, the reported prevalence of RA is 0.5–1%, with an annual incidence of 5–50 cases per 100,000 people [1]. While the exact cause is not fully understood, it is widely accepted that in the central stage, the exo- and endogenous propagation of citrullination and autoreactive T-cells plays an essential role in the production of autoantibodies in lymph nodes, such as anti-cyclic citrullinated peptide (aCCP) antibodies, in the early stages of the disease [2,3]. The percolation of autoreactive CD4+ T-cells, plasma cells, and autoantibodies into the synovium initiates the later, destructive phase of RA, characterised by cytokine-driven synovitis and the formation of pannus by the proliferation of fibroblast-like synoviocytes (FLSs), leading to irreversible clinical manifestations [2–6].

RA was firstly found to be significantly associated with HLA-DRB1 alleles carrying a common amino acid motif, which is the human leukocyte antigen (HLA)-DRB shared epitope (HLA-DRB-SE); however, the dosage of the shared epitope allele is only associated with patients who present aCCP antibodies [7]. Subsequent high-throughput genome-wide association studies (GWASs) have highlighted the importance of non-HLA genes in the pathogenesis of RA [7–9]. As of the time of writing this article, three recent GWASs have identified 124 risk loci with RA, involving 69,825 patients and 330,798 adjusted controls [10–12].

The majority of these 124 loci are associated with general cellular and inflammatory processes. However, among these genes, only one, the autoimmune regulator (*AIRE*) is specifically linked to a neonatal immunologic event. Its crucial role is to suppress the autoimmune potential in neonates [13]. The *AIRE* gene's coding sequence is located in the 21q22.3 chromosomal region with 14 exonic and 13 intronic regions. It encodes a 540-amino-acid transcription factor known as the autoimmune regulator (AIRE). This regulator is primarily produced in the thymic medulla and is localised in the nucleus of medullar thymoepithelial cells (mTECs). Its primary and crucial function is to eliminate autoimmune thymocytes from the repertoire by producing tissue-restricted antigens (TRAs) during promiscuous gene expression (pGE) [13–17]. The central role and expression of *AIRE* gradually disappear by the post-adolescent involution of the thymus, indicating its crucial role only in the neonatal period.

In the existing literature, the precise biological origin of RA has remained elusive. Thus, we aim to shed light on the early onset of the disease, suggesting neonatal self-tolerance as a potential premature contributor to RA. This suggests that the disease may start in the neonatal period.

*AIRE* comprises five functionally important structural domains from the N- to C-termini. These include the homogeneously staining region (HSR) or caspase-activation and recruitment domain (CARD), which are responsible for caspase activity during apoptosis and for controlling the connection with binding partners. The nuclear localisation signal (NLS) is recognised by specific importins. The SAND (Sp100 AIRE NucP41/75 and DEAF) domain contains a DNA-binding sequence for mTECs and has transcriptional characteristics that determine the production of TRAs. Additionally, two plant homeodomain (PHD1 and PHD2, respectively) proteins are isolated by one proline-rich region (PRR). These domains epigenetically interact with unmethylated histone H3 lysine 4 (H3K4) and DNA-dependent protein kinase (DNA-PK) and activate gene transcription, respectively [13–15,18,19]. Moreover, four LxxLL motifs are encased between the N- and C-termini, where their role is to interact with CREB-binding protein (CBP) and signal transducers and activators of transcription (STAT) partners [19].

Over 100 nonsense, frameshift, and missense mutations have been identified in AIRE protein coding sequences. These mutations lead to autoimmune polyendocrine syndrome type 1 (APS-1), also known as autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED) (OMIM #240300). APS-1 is characterised by multi-organ, generalised deterioration attributed to the failure to eliminate self-reactive T-cells during early immunologic development [19,20].

Research studies using K/BxN mouse models of RA have provided evidence supporting the role of central tolerance in the development of RA. This indicated that a low presentation of self-antigens may lead to the synovial percolation of autoreactive CD4+ type 1 T helper (Th) 1, Th2, Th17, and regulatory T cells (Treg) in the late, destructive stage of RA [5,21,22]. The association between negative selection and single-nucleotide polymorphisms (SNPs) in AIRE was first described by an *in vivo* study in which AIRE-655R (rs117557896) and AIRE-230Y (rs751032) were detected [23]. The study demonstrated that the haplotypes AIRE-655G and AIRE-230T dramatically reduced AIRE transcription, influenced negative selection, and, therefore, elevated the risk of autoimmunity [23]. Albeit the growing scientific interest in AIRE, only a small number of studies have associated autoimmune disease susceptibility to allelic polymorphisms in AIRE, including vitiligo [24], alopecia areata [25], Addison's disease [26], systemic sclerosis [27], autoimmune thyroiditis [27], and melanoma [28]. Among most molecular epidemiological studies, RA has been associated with AIRE SNPs [29]. In order to comprehensively address the allelic polymorphisms in AIRE, associated with RA in Asian populations, we performed a meta-analysis with rs2075876 and rs760426 [30]. To date, only one study has investigated AIRE SNPs in a Spanish population, analysing allelic frequencies [31]. The SAND domain caught our attention due to its perceived importance in binding AIRE and its partners to mTEC DNA, as well as in protein–protein interactions that play a role in determining the negative selection of central tolerance through the transcriptional expression of TRAs [32]. Our study's primary aim was to assess the relationship between AIRE and RA using five different genetic models (allelic, dominant, recessive, codominant heterozygous, codominant homozygous, and overdominant) in order to establish risk associations and associated clinical phenotypes.

## 2. Materials and Methods

### 2.1. Inclusion and Exclusion Criteria of Variations

We performed a multistep preliminary selection method to separate the variations of interest from copy number variations (CNVs), point mutations, insertions, and deletions (indels). UCSC Genome Browser (GRCh38 assembly) and Ensembl automatic annotation of the human genome sequence (GRCh38.p14) were used as human genome databases [33,34].

The inclusion criteria were set with the utmost thoroughness, requiring the global minor allele frequency (MAF) to be between the values of 0.01 and 0.5 and the variation to be present in at least one of the following databases: NCBI dbSNP, 1000 Genomes Project Phase III (including the HapMap project), NHLBI Trans-Omics for Precision Medicine (human only), NHLBI GO Exome Sequencing Project (human only), Genome Aggregation Database (human only), and Exome Aggregation Consortium (human only) [35–39]. Some variations presented global MAF between 0.01 and 0.5. Still, when we scrutinised the allele frequencies in subpopulations of European (EUR) and Utah residents with Northern and Western European ancestry (CEU), MAF was lower than 0.01, which raised the probability of detection failure in our study population. Therefore, we excluded those variations due to low detection probability.

### 2.2. Ethical Approval

Our analytical case-control study was conducted following the guidelines of the Declaration of Helsinki. Our protocol was also approved by the Secretary of the Medical Research Council, Scientific and Research Ethics Committee of Hungary (protocol numbers



11871-7/2018/EÜIG on 24 April 2018 and 57142-5/2022/EÜIG on 8 November 2022). Informed consent was obtained from all subjects.

### 2.3. Study Subjects

Individuals receiving outpatient or inpatient care at the Harkány Thermal Rehabilitation Centre and St. Andrew Hospital for Rheumatology of Hévíz formed the study population of RA patients and controls, and we successfully included a total of 592 individuals. Voluntary participants received our study description sheet and gave their informed consent. Whole blood samples were consecutively collected in 4 mL citrate tubes after patients' admission and stored at  $-75^{\circ}\text{C}$ . The patients in the RA study group were required to be older than 18, sign a consent statement, and have clinically diagnosed RA, based on the American College of Rheumatology classification, to be included [40]. Patients with concomitant autoimmune pathology in their past medical history, such as Sjögren's syndrome, psoriasis, and systemic sclerosis, were excluded from the further analysis to lower the probability of confounding effects of autoimmune diseases other than RA. We finally included 270 patients in the RA group. For the control study group, we required subjects to be older than 18, to sign a consent statement, to not have symptomatic RA, based on the American College of Rheumatology classification, and to not have any autoimmune manifestation recorded in their medical history. The mean age of RA symptomatological onset is 65 years; therefore, to minimise the probability of latent cases among controls, we included participants aged 75 or older [1]. Furthermore, levels of C-reactive protein (CRP) lower than 20 mg/dL and rheumatoid factor (Rf) lower than 25 IU/mL IgG were necessary for inclusion in the control group, and we successfully included 322 subjects [41].

### 2.4. Assessment of Clinical Parameters

Complete blood count, cholesterol profile, blood sugar profile, erythrocyte sedimentation rate (ESR), CRP, sodium, potassium, urea, creatinine, glutamate-oxaloacetate transaminase, glutamate-pyruvate transaminase, gamma-glutamyl transferase, alkaline phosphatase, lactate dehydrogenase, creatine kinase, protein albumin, magnesium, phosphorus, and serum bilirubin levels were thoroughly assessed. To ensure the seropositivity of patients with RA and the seronegativity of healthy controls, we determined levels of Rf, aCCP, and antinuclear antibodies [42,43]. Based on the aforementioned clinical parameters and diagnostic criteria, functional status was calculated for each subject with RA by disease activity score with 28-joint counts (DAS28) and functional independence measure (FIM) [44]. To determine ESR, the Westergren method was applied to measure C-reactive protein (CRP) and Rf (IgG). Konelab Arena 20XT Clinical Chemistry Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) was used, following the manufacturer's recommendations [45]. All of the abovementioned basic laboratory tests, including ESR and CRP in addition to Rf, were performed at the Harkány Thermal Rehabilitation Centre. Immunoscan CCPlus enzyme-linked immunosorbent assay kit (Svar Life Sciences, Malmö, Sweden) was used to determine aCCP levels at the Department of Immunology and Biotechnology, Medical School, University of Pécs, Hungary, according to the manufacturer's protocols.

### 2.5. DNA Extraction

Genomic DNA of patients with RA and controls was isolated using 0.5 mL whole blood lysed with 1 mL DNAzol BD reagent (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. The precipitation of genomic DNA (gDNA) was performed at room temperature with 0.4 mL isopropanol for 5 min, and then the sample was centrifuged. After the supernatant was removed, the pellet was first washed with 0.5 mL DNAzol BD reagent, subsequently followed by a second wash with 1 mL 75% ethanol. Dried genomic DNA was dissolved in 200  $\mu\text{L}$  Tris-EDTA buffer and quantitatively measured using NanoDrop One (Thermo Fisher Scientific, Waltham, MA, USA). Per reaction, 20 ng gDNA was used for quantitative real-time PCR (qPCR).

## 2.6. Genotyping AIRE Variants

Purification of gDNA was subsequently followed by multiplex qPCR testing. Samples of patients with RA and controls were genotyped using TaqMan technology. To perform accurate discrimination, we used Minor Groove Binder (MGB) TaqMan<sup>®</sup> Assays (Thermo Fisher Scientific, Waltham, MA, USA) in which MGB at the 3' ends of the probe increased the melting temperature, therefore stabilising probe–target hybrids. MGB TaqMan probes are significantly shorter than traditional probes, enabling these probes to better and more accurately discriminate between highly homologous allele sequences. Assays use one primer pair, a VIC dye-labelled probe (530 nm yellow channel) that detects the Allele 1 sequence and an FAM dye-labelled probe (470 nm green channel) that detects the Allele 2 sequence. Samples with signals only in VIC or FAM channels are homozygous for the given sequence, while samples with signals in both channels are considered heterozygous for the analysed SNP site. Quality-checked and functionally tested Taqman assays were utilised for rs878081 (assay ID: C\_\_2978265\_10), rs1003854 (assay ID: C\_\_9480542\_10), rs2075876 (assay ID: C\_\_15863141\_20), rs1055311 (assay ID: C\_\_9480540\_20), and rs1003853 (assay ID: C\_\_2978264\_1\_) (catalogue 4351379). As per our protocol, 9  $\mu$ L template gDNA (20 ng) was mixed with 1  $\mu$ L TaqMan<sup>®</sup> Assay and 10  $\mu$ L TaqMan Genotyping Master Mix (catalogue 4,371,357) (Thermo Fisher Scientific, Waltham, MA, USA) to a total volume of 20  $\mu$ L. Fluorescence intensity was detected in both FAM and VIC channels in the Rotor-Gene 3000 thermal cycler (Corbett Life Science, Qiagen, Crawley, UK) with its software Rotor-Gene 6 v6.1 for which the protocol was as follows: 10 min at 95 °C for AmpliTaq Gold<sup>®</sup> DNA Polymerase, UP enzyme activation, and 43 cycles of denaturation and synthesis (15 s at 95 °C for DNA denaturation, 1 min at 60 °C for annealing and extension). Assessment of allelic discrimination was conducted using Rotor-Gene 6000 v1.7 software (Corbett Life Science, Qiagen, Crawley, UK). Spectra of each sample were reviewed by visual inspection to confirm genotype assignment. The final call rate for genotyping was 100%.

## 2.7. Statistical Analyses

Clinical characteristics as continuous variables are represented as mean  $\pm$  standard deviation (SD). Levene's test was applied to analyse the equality of variances. Furthermore, the Shapiro–Wilk test was used to determine normality. Independent-sample t-tests and analysis of variance were used to compare groups for continuous variables. Non-normally distributed variables were compared by their medians, and Mann–Whitney U and Kruskal–Wallis H tests with two-tailed significance were used. To compare the significance of sex distributions among males and females, the one-sample Kolmogorov–Smirnov test was applied. Linkage disequilibrium (LD) was evaluated by Lewontin's 'D', in which LD structure was constructed using Haploview software (version 4.2) [46,47]. Expected genotype frequencies were calculated based on Hardy–Weinberg equilibrium (HWE), and the difference with observed frequencies was statistically compared with non-parametric chi-square tests to see how well our observed frequencies fit the expected frequencies according to the HWE. To test the association between genotype and allele frequencies with RA, odds ratios (OR) with 95% confidence interval (CI) were calculated by binary logistics in which *p*-value less than 0.05 determined the statistical significance in allelic, codominant homozygous, codominant heterozygous, dominant, recessive, and overdominant genetic models. Bivariate correlation was applied with the Pearson correlation coefficient and two-tailed significance to determine the correlation between genotype frequencies and clinical characteristics. IBM SPSS software (version 28.0, IBM Company, Armonk, NY, USA) was used for statistical analyses.

## 2.8. Analysing the Effects of Variants on Regulatory Binding Motifs

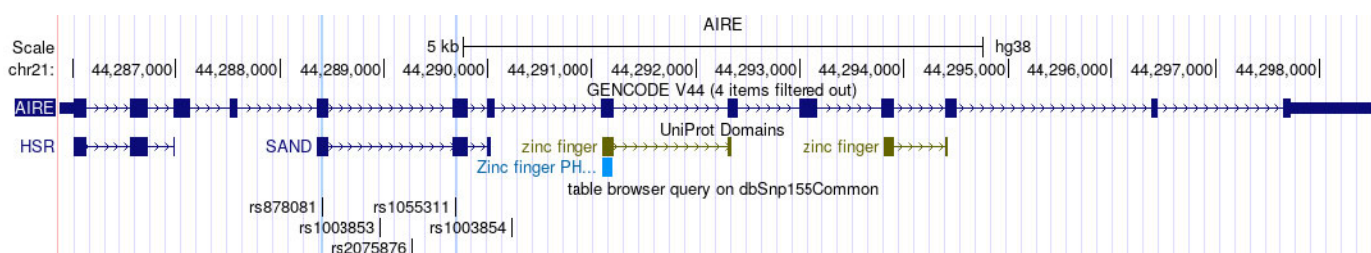
To estimate the outcomes of allelic variants on *AIRE* expression in silico, we used HaploReg v4.2 databases [47]. Allele-specific affinity to transcription factor binding motifs was determined by using the logarithm of the odds (LOD) values of reference (ref) and alternative (alt) sequences [48]. When a positive result is provided by the subtraction of LOD (ref) from the LOD (alt) value for the alternative sequence, the predicted relative affinity is higher, suggesting that the alternate sequence binds the regulatory transcriptional factor with higher affinity than that of the reference sequence [48].

## 3. Results

### 3.1. Preliminarily Included Variations

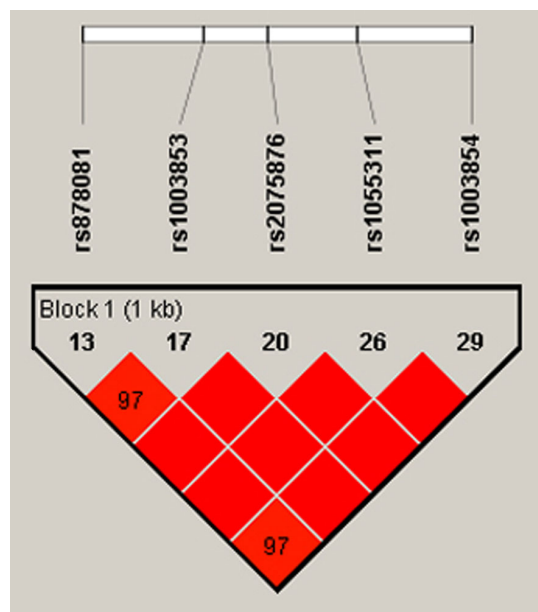
In the UCSC Genome Browser and Ensembl automatic annotation system, our scope was restricted to the *AIRE* gene sequence (HGNC Symbol; Acc: HGNC:360), which resides in the 21q22.3 chromosomal region between 44,285,838 and 44,298,648 on the forward strand [33,34]. Without restrictions, we identified 23,023 variations. After adjusting the global MAF to 0.01–0.5, we detected 68 SNPs with at least one representation in NCBI dbSNP, 1000 Genomes Project Phase III (including the HapMap project), NHLBI Trans-Omics for Precision Medicine (human only), NHLBI GO Exome Sequencing Project (human only), Genome Aggregation Database (human only), and Exome Aggregation Consortium (human only) [35–39]. Following our exclusion criteria, we eliminated 24 SNPs with an MAF lower than 0.01 from the EUR and CEU subpopulations to avoid low chances of allelic discrimination; therefore, 44 variations were shortlisted in the 44,285,838–44,298,648 region.

Out of the 14 exons in *AIRE*, Exon 5 and Exon 6 encode the SAND structural domain in particular, which is responsible for binding AIRE and its partners to mTECs' DNA and further governing the pGE of TRAs by protein–protein interactions with CBP [32]. Mutations in SAND-encoding Exon 5 and Exon 6 and the splice site between them of Intron 5 cause autoimmune APECED [49–52]. In light of the importance of the SAND region in negative selection and autoimmunity, our scope of interest included the analysis of SNPs in Exons 5–6. Consequently, we chose the loci of rs878081 in Exon 5 and rs1055311 in Exon 6. Furthermore, we examined their strongly linked intron variants, including rs1003853, rs2075876 in Intron 5, and rs1003854 in Intron 7, respectively, to clarify the possible association of alleles and genotypes with RA and the clinical characteristics of patients. The positions of the two exon and three intron variations in *AIRE* and the resultant domains of AIRE are demonstrated in Figure 1. The five selected loci are on one strongly linked haplotype block with 'D' values ranging between 0.977 and 1.000. The linkage map is illustrated in Figure 2.



**Figure 1.** Exonic and intronic coding sequences in *AIRE* (GENCODE V44), domains of *AIRE* protein (UniProt domains), and positions of the five loci (table browser query on dbSnp155Common generated by UCSC Genome Browser on Human (GRCh38/hg38)). See <https://genome.ucsc.edu/index.html> (accessed on 25 March 2024) [33].





**Figure 2.** Positions of the five loci on the LD map and Lewontin's 'D' values. The dark red colour indicates high LD, and blocks without numbers indicate the maximum LD value with 'D' = 1 (Species: Human; Assembly: GRCh38.p14; Region Lookup: 21:44285838-44298648). LD, linkage disequilibrium [46,47].

### 3.2. Characteristics of Study Populations

Our total population includes 592 individuals, 270 patients with RA, and 322 healthy controls. Among these, 100-100 participants were preliminarily investigated in our pilot study for allelic polymorphisms rs878081 and rs1003854 to see their tendencies of MAFs and to test our genotyping workflow and chances of detectability [50]. From then on, we formed two groups as follows: Group 1 was the total population with 270 RA patients and 322 controls for the rs1003853, rs2075876, and rs1055311 loci; and Group 2 had 170 RA patients and 222 controls, which was the subpopulation of Group 1 for rs878081 and rs1003854 because the results for 100 RA patients and 100 controls of rs878081 and rs1003854 from the total population in Group 1 had already been published. The characteristics of each study group can be seen in Table 1. The mean age of the RA patients was  $65.47 \pm 9.85$  years and  $65.53 \pm 9.89$  years, with 84.0% and 85.5% females in Groups 1 and 2, respectively. The distribution of females was significantly higher than that of males in the RA and control groups [1]. The participants in our total population were all born in Hungary. Their ESRs, CRPs, Rfs, and aCCPs were significantly higher at  $p < 0.001$  in all of the RA groups with seropositivity values of 60.7% and 71.1%, respectively, and seronegativity values for control Groups 1 and 2 of 90.0% and 84.2%, respectively. Our RA groups seemed to be strongly positive with aCCP levels of  $590.69 \pm 947.96$  and  $596.27 \pm 1003.71$  U/mL and demonstrated a highly active disease ( $\geq 5.1$ ) with DAS28 values calculated as  $5.35 \pm 1.20$  and  $5.44 \pm 1.12$  for Groups 1 and 2, respectively.

**Table 1.** Clinical characteristics of study populations.

Characteristics	Group 1		Group 2	
	Patients with RA (n = 270)	Control Subjects (n = 322)	Patients with RA (n = 170)	Control Subjects (n = 222)
Age (years), mean $\pm$ SD	$65.47 \pm 9.85^a$	$78.66 \pm 5.86$	$65.53 \pm 9.89^d$	$77.53 \pm 6.03$
Sex				
Female, no. (%)	227 (84.0) <sup>b</sup>	252 (78.2) <sup>c</sup>	142 (83.5) <sup>e</sup>	184 (82.8) <sup>f</sup>
Male, no. (%)	43 (15.9)	70 (21.7)	28 (16.4)	38 (17.1)
ESR (mm/h), mean $\pm$ SD (normal range: 3–13 mm/h)	$39.52 \pm 25.21^a$	$17.87 \pm 12.81$	$39.09 \pm 26.04^d$	$19.02 \pm 13.82$

Table 1. Cont.

Characteristics	Group 1		Group 2	
	Patients with RA (n = 270)	Control Subjects (n = 322)	Patients with RA (n = 170)	Control Subjects (n = 222)
CRP (mg/dL), mean ± SD (normal range: 0–10 mg/dL)	29.79 ± 29.88 <sup>a</sup>	7.10 ± 6.80	30.36 ± 30.34 <sup>d</sup>	7.53 ± 8.03
Rf (IgG) (IU/mL), mean ± SD (normal range: 0–15 IU/mL)	118.80 ± 126.63 <sup>a</sup>	13.49 ± 5.55	104.71 ± 113.31 <sup>d</sup>	14.41 ± 13.75
aCCP (U/mL), mean ± SD (normal range: <20.0 U/mL; weak positivity: ≥20.0–39.0 U/mL; moderate positivity: ≥39.0–59.0 U/mL; strong positivity: ≥59.0 U/mL)	590.69 ± 947.96 <sup>a</sup>	4.56 ± 2.29	596.27 ± 1003.71 <sup>d</sup>	4.55 ± 1.66
Seropositive, no. (%)	164 (60.7)		121 (71.1)	
Seronegative, no. (%)	23 (8.5) <sup>a</sup>	290 (90.0)	33 (19.4) <sup>d</sup>	187 (84.2)
Without serology (%)	83 (30.8)	32 (10.0)	16 (9.5)	35 (15.8)
DAS28, mean ± SD (disease remission: <2.6; low: ≥2.6–3.1; moderate: ≥3.1–5.1; high: ≥5.1)	5.35 ± 1.20	-	5.44 ± 1.12	-

Group 1: total population; Group 2: subpopulation of the total; aCCP: anti-cyclic citrullinated peptide; CRP: C-reactive protein; DAS28: disease activity score with 28-joint counts; ESR: erythrocyte sedimentation rate; RA: rheumatoid arthritis; Rf: rheumatoid factor; SD: standard deviation; <sup>a</sup>:  $p < 0.001$  compared with control subjects in Group 1; <sup>b</sup>:  $p < 0.001$  compared with RA males in Group 1; <sup>c</sup>:  $p < 0.001$  compared with control males in Group 1; <sup>d</sup>:  $p < 0.001$  compared with control subjects in Group 2; <sup>e</sup>:  $p < 0.001$  compared with RA males in Group 2; <sup>f</sup>:  $p < 0.01$  compared with control males in Group 2.

### 3.3. Results of HWE Analysis

The statistical significance levels of the difference between the expected and observed genotype frequencies in Group 1 for rs2075876, rs1055311, and rs1003853 among the RA patients were  $p = 0.916$ ,  $p = 0.985$ , and  $p = 0.715$ , respectively. Those of the controls were  $p = 0.627$ ,  $p = 0.899$ , and  $p = 0.503$ , respectively. In Group 2 for rs878081 and rs1003854, the statistical significance levels of the difference between the expected and observed genotype frequencies for each SNP among the RA patients were  $p = 0.991$  and  $p = 0.995$ , respectively. Those of the controls were  $p = 0.870$  and  $p = 0.658$ , respectively. Non-parametric chi-square tests demonstrated that the expected genotype frequencies fit well with the observed frequencies, following the HWE. Therefore, our study population is free from selection bias, non-random sampling, and the accumulation of relatives.

### 3.4. Allelic Polymorphisms in Exon 5 and Intron 5 and 7 Are Associated with RA Risk

To clarify significant associations between allelic and genotype frequencies and RA, allelic, codominant homozygous, codominant heterozygous, dominant, recessive, and overdominant genetic models were utilised for the 592 individuals and their 1184 alleles [51]. Among the five loci, the C-allele of SAND coding rs878081 on Exon 5 demonstrated firstly a significant association with RA risk under the allelic model (OR = 1.48, 95% CI 1.05–2.09,  $p = 0.023$ ). The genotype frequency of CC homozygote RA patients was significantly higher and associated with RA risk under the recessive model in contrast with CT heterozygotes and TT homozygotes (OR = 1.64, 95% CI 1.09–2.48,  $p = 0.01$ ). In the overdominant model, the CT genotype was negatively associated with RA risk, contrasting with the subpopulation of TT and CC homozygotes (OR = 0.64, 95% CI 0.42–0.97,  $p = 0.037$ ). Supplementary Figures S1 and S2 show the allelic discrimination results of signals in VIC<sup>TM</sup> or FAM<sup>TM</sup> channels of rs878081 for the RA patients and controls.

Our study revealed a significant association between the intronic variation in rs1003853 in Intron 5 of AIRE and RA risk. Analysing the frequencies of 1184 alleles, we found that the C-allele was significantly more represented among the RA alleles than the alleles of the controls (OR = 1.33, 95% CI 1.01–1.74,  $p = 0.037$ ). Interestingly, the CC homozygosity of rs1003853 was positively associated with RA risk by genotype frequency, with significantly higher odds for the RA patients than those for the controls in the recessive genetic model (OR = 1.618, 95% CI 1.16–2.25,  $p = 0.004$ ). The overdominant model, however, suggested a negative association with RA risk (OR = 0.59, 95% CI 0.42–0.82,  $p = 0.002$ ).

In the sequence of Intron 7 of *AIRE*, rs1003854 was the next locus where the odds of the T-allele were significantly higher for the RA alleles compared to those for the C-allele (OR = 1.52, 95% CI 1.08–2.12,  $p = 0.014$ ). In the recessive model, TT homozygosity was significantly more frequent in the patients with RA than those with the TC and CC genotypes compared with the healthy controls (OR = 1.72, 95% CI 1.15–1.44,  $p = 0.008$ ). In the overdominant model, heterozygotes were significantly overrepresented in the controls compared to in the RA cases (OR = 0.63, 95% CI 0.42–0.96,  $p = 0.034$ ). The allele and genotype frequencies, as well as the results of our bivariate logistic regression associating rs878081, rs1003854, and rs1003853 with RA, are presented in Table 2. The association results of rs2075876 and rs1055311 with RA are published in Supplementary Tables S1 and S2.

**Table 2.** Allele and genotype frequencies, as well as association of RA with allelic polymorphisms rs878081, rs1003853, and rs1003854 in *AIRE*.

<i>AIRE</i> rs878081 (Exon 5)	RA n = 170 (%)	Control Subjects n = 222 (%)	OR (95% CI)	$p^a$
Alleles				
T	66 (19.4)	117 (26.4)		
C	274 (80.6)	327 (73.6)	1.48 (1.05–2.09)	0.023 *
Genotypes				
TT	7 (4.1)	13 (5.9)		
CT	53 (31.2)	92 (41.4)	1.07 (0.40–2.84)	0.892
CC	110 (64.7)	117 (52.7)	1.74 (0.67–4.53)	0.253
Dominant model				
TT	7 (4.1)	13 (5.9)		
CC + CT	163 (95.9)	209 (94.1)	1.44 (0.56–3.71)	0.441
Recessive model				
CT + TT	60 (35.3)	105 (47.3)		
CC	110 (64.7)	117 (52.7)	1.64 (1.09–2.48)	0.017 *
Overdominant model				
CC + TT	117 (68.8)	130 (58.6)		
CT	53 (31.2)	92 (41.4)	0.64 (0.42–0.97)	0.037 *
<i>AIRE</i> rs1003853 (Intron 5)	RA n = 270 (%)	Control subjects n = 322 (%)	OR (95% CI)	$p^a$
Alleles				
T	116 (21.5)	172 (26.7)		
C	424 (78.5)	472 (73.3)	1.33 (1.01–1.74)	0.037 *
Genotypes				
TT	15 (5.6)	16 (5.0)		
CT	85 (31.5)	141 (43.8)	0.64 (0.30–1.36)	0.251
CC	170 (63.0)	165 (51.2)	1.09 (0.52–2.29)	0.802
Dominant model				
TT	15 (5.6)	16 (5.0)		
CC + CT	255 (94.4)	306 (95.0)	0.88 (0.43–1.83)	0.750
Recessive model				
CT + TT	100 (37.0)	157 (48.8)		
CC	170 (63.0)	165 (51.2)	1.618 (1.16–2.25)	0.004 *
Overdominant model				
CC + TT	185 (68.5)	181 (56.2)		
CT	85 (31.5)	141 (43.8)	0.59 (0.42–0.82)	0.002 *
<i>AIRE</i> rs1003854 (Intron 7)	RA n = 170 (%)	Control subjects n = 222 (%)	OR (95% CI)	$p^a$
Alleles				
C	69 (20.3)	124 (27.9)		
T	271 (79.7)	320 (72.1)	1.52 (1.08–2.12)	0.014 *
Genotypes				
CC	6 (3.5)	13 (5.9)		
TC	57 (33.5)	99 (44.6)	1.34 (0.48–3.69)	0.567
TT	106 (62.4)	109 (41.1)	2.10 (0.77–5.74)	0.146
Dominant model				
CC	6 (3.5)	14 (6.3)		
TC + TT	164 (96.5)	208 (93.7)	1.84 (0.69–4.89)	0.222
Recessive model				
TC + CC	63 (37.1)	112 (50.5)		
TT	107 (62.9)	110 (49.5)	1.72 (1.15–1.44)	0.008 *
Overdominant model				
CC + TT	113 (66.5)	124 (55.9)		
TC	57 (33.5)	98 (44.1)	0.63 (0.42–0.96)	0.034 *

CI: confidence interval; OR: odds ratio; RA: rheumatoid arthritis.  $p^a$ : significance of binary logistic regression; \*  $p < 0.05$ .

### 3.5. Bivariate Correlation and Association of Clinical Parameters with Genotypic Subgroups of Different Genetic Models in Patients with RA and Control Subjects

Our study was intended to statistically analyse clinical parameters (ESR, CRP, Rf, aCCP, and DAS28) in the codominant homozygous, codominant heterozygous, dominant, recessive, and overdominant genetic models in subpopulations of RA patients and controls to provide possible evidence of differences between them in terms of clinical parameters and genotypes.

Among their clinical features, the clinical parameter ESR was able to characterise RA activity well by its significant positive correlation with DAS28 ( $p = 0.013$ ) in the total RA population. Therefore, we decided to stratify the RA population based on codominant homozygous and heterozygous, dominant, recessive, and overdominant genetic models to see if there were further correlations and associations between genotypic subgroups and clinical parameters. Among the RA patients, statistically significant differences were found in the recessive model of rs878081, in which CC homozygotes presented significantly higher mean ESR levels ( $p = 0.027$ ) compared with the TT + CT subpopulation. For the Pearson bivariate correlation, the ESR was shown to be significantly correlated with CC homozygotes ( $p = 0.023$ ,  $r = 0.190$ ); furthermore, with a binary logistic regression, a significant association (OR = 1.01, 95% CI 1.002–1.032,  $p = 0.026$ ) was found. With 592 subjects, Group 1 demonstrated significantly higher aCCP levels among the 270 RA patients in the codominant heterozygous and dominant model of the rs1055311 locus ( $p = 0.028$  and  $0.044$ , respectively). Furthermore, our correlation analysis of the CT and CC genotypes revealed a significantly positive correlation of heterozygotes with significantly higher aCCP levels ( $r = 0.255$ ,  $p = 0.036$ ), suggesting a genotype-determined disease activity. The results of our subgroup analyses are demonstrated in Tables 3 and 4.

**Table 3.** Correlation and association of ESR levels with genotype subgroups in five different genetic models of rs878081 in AIRE.

AIRE rs878081 (Exon 5)	Patients with RA						Control Subjects					
	ESR	$p^a$	Pearson Correlation	$p^b$	OR (95% CI)	$p^c$	ESR	$p^d$	Pearson Correlation	$p^e$	OR (95% CI)	$p^f$
Genotype												
TT	24.25 ± 13.50				1.02		16.69 ± 13.11				1.02	
CT	33.16 ± 23.37	0.599	0.10	0.458	(0.96–1.08)	0.454	21.73 ± 14.96	0.147	0.11	0.255	(0.97–1.08)	0.259
CC	42.77 ± 27.11	0.196	0.13	0.179	1.03 (0.98–1.09)	0.196	17.33 ± 12.79	0.708	0.01	0.86	1.004 (0.95–1.05)	0.863
Dominant												
TT	24.25 ± 13.50				1.03		16.69 ± 13.11				1.01	
CC + CT	39.52 ± 26.22	0.295	0.09	0.24	(0.97–1.09)	0.262	19.18 ± 13.88	0.391	0.04	0.530	(0.96–1.06)	0.530
Recessive												
TT + CT	32.46 ± 22.79				1.01		21.04 ± 14.76				0.98	
CC	42.77 ± 27.11	0.027*	0.19	0.023*	(1.002–1.03)	0.026*	17.33 ± 12.79	0.063	−0.13	0.054	(0.96–1.001)	0.980
Overdominant												
TT + CC	42.00 ± 26.90				0.98		17.26 ± 12.77				1.02	
CT	33.16 ± 23.37	0.059	−0.16	0.056	(0.97–1.001)	0.059	21.73 ± 14.96	0.020*	0.15	0.023*	(1.003–1.04)	0.025*

ESR: erythrocyte sedimentation rate; OR: odds ratio; RA: rheumatoid arthritis; SD: standard deviation; ESR levels were expressed as mean ± SD;  $p^a$ : significance of differences in ESR levels in patients with RA;  $p^b$ : significance of Pearson correlation in genetic subgroups and ESR levels in patients with RA;  $p^c$ : significance of bivariate logistic regression of genetic subgroups and ESR levels in patients with RA;  $p^d$ : significance of differences in ESR levels in healthy controls;  $p^e$ : significance of Pearson correlation in genetic subgroups and ESR levels in healthy controls;  $p^f$ : significance of bivariate logistic regression of genetic subgroups and ESR levels in healthy controls; \*:  $p < 0.05$ .

**Table 4.** Correlation of aCCP levels with genotype subgroups in five different genetic models of rs1055311 in AIRE.

AIRE rs1055311 (Exon 6)	Patients with RA				Control Subjects			
	aCCP	$p^a$	Pearson Correlation	$p^b$	aCCP	$p^c$	Pearson Correlation	$p^d$
Genotype								
TT	213.23 ± 258.53				4.44 ± 2.43			
CT	725.33 ± 998.49	0.028*	0.25	0.036*	4.43 ± 2.32	0.950	−0.001	0.997
CC	591.51 ± 999.63	0.093	0.16	0.116	4.83 ± 2.26	0.779	0.06	0.679

Table 4. Cont.

AIRE rs1055311 (Exon 6)	Patients with RA				Control Subjects			
	aCCP	<i>p</i> <sup>a</sup>	Pearson Correlation	<i>p</i> <sup>b</sup>	aCCP	<i>p</i> <sup>c</sup>	Pearson Correlation	<i>p</i> <sup>d</sup>
Dominant TT	213.23 ± 258.53				4.44 ± 2.43			
CC + CT	644.19 ± 997.37	0.044 *	0.15	0.071	4.62 ± 2.29	0.918	0.02	0.843
Recessive TT + CT	589.77 ± 893.25				4.43 ± 2.31			
CC	591.51 ± 999.63	0.982	0.001	0.991	4.83 ± 2.26	0.187	0.085	0.403
Overdominant TT + CC	519.84 ± 917.72				4.77 ± 2.26			
CT	725.33 ± 998.49	0.157	0.103	0.216	4.43 ± 2.32	0.209	−0.07	0.467

aCCP: anti-cyclic citrullinated peptide; RA: rheumatoid arthritis; SD: standard deviation; aCCP levels were expressed as mean ± SD; *p*<sup>a</sup>: significance of differences in aCCP levels in patients with RA; *p*<sup>b</sup>: significance of Pearson correlation in genetic subgroups and aCCP levels in patients with RA; *p*<sup>c</sup>: significance of differences in aCCP levels in healthy controls; *p*<sup>d</sup>: significance of Pearson correlation in genetic subgroups and aCCP levels in healthy controls. \* *p* < 0.05.

### 3.6. Allele-Specific Affinity to Transcriptional Factor Binding Motifs

In the SAND domain, rs878081 on Exon 5 and rs1055311 on Exon 6 are synonymous polymorphisms that are not structural, but alterations in transcription may be associated with different alleles. To clarify the possible transcriptional consequences on alleles specifically, we performed an in silico analysis of the affinity of transcription factor binding motifs using HaploReg v4.2, in which we tested the four loci that seemed to be associated with RA risk and clinical parameters such as ESR and aCCP levels [48]. The database predicted two important motif-binding sites for nuclear factor-kappa B (NF-κB) and GATA transcription factors for the rs878081 and rs1003854 loci, respectively. GATA is a family of zinc-finger proteins with the ability to bind the consensus DNA sequence (T/A)GATA(A/G). The calculated LOD scores for affinity to regulatory motifs were +2 for the rs878081 T-allele and +1 for the rs1003854 C-allele, demonstrating a higher affinity of binding with the aforementioned transcriptional factors by alternate sequences of the rs878081 T-allele and rs1003854 C-allele, respectively. Furthermore, a higher affinity to bind to NF-κB was also predicted for the alternate rs1055311 T-allelic polymorphism with a +11.1 LOD score. Our results are demonstrated in Table 5.

Table 5. Regulatory motifs altered at rs878081, rs1003854, and rs1055311.

Locus	Ref	Alt	Position Weight Matrix ID	Strand	LOD Value of Ref	LOD Value of Alt
rs878081	C	T	NF-kappaB_known4	+	12.5	14.5
rs1003854	T	C	GATA_disc3	+	12.2	13.2
rs1055311	C	T	NF-kappaB_disc2	+	1.7	12.8

Ref: Reference allele; Alt: Alternative allele; LOD: logarithm of the odds; Available online: <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php> (accessed on 25 March 2024) [47,48].

## 4. Discussion

RA is a multifactorial autoimmune disease in which endo- and exogenous events contribute to the disease's initiation [52]. Three of the latest GWASs have associated RA with specific loci posing a genetic risk, such as *AIRE*, which play a role in regulating immunological self-tolerance to ensure a healthy adaptive immune system. Postnatally, in a thymic environment, the thymocyte repertoire first undergoes an education against self-reactivity by central tolerance. Following this, peripheral organs further refine and eliminate potentially autoimmune T-cells through a process known as clonal deletion [53]. Central tolerance involves three checkpoints. The first is the β-selection, in which abnormal thymocytes with incorrectly assembled TCRs are removed [53]. The subsequent second checkpoint is the positive selection still in the cortex, where thymocytes are neglected if



they are unable to recognise major histocompatibility complex (MHC) epitopes on cortical myoepithelial cells (ECs) [53]. Upon reaching the medulla, surviving thymocytes undergo negative selection at the final checkpoint [53]. At this stage, mTECs epigenetically express a diverse range of unique gene clusters, generating TRAs to mirror the immunologic “self” with the highest presentation from all parenchymal organs (such as the heart, stomach, bone, CNS, liver, and intestine) [53]. This phenomenon has been denoted as pGE, whereby AIRE binds to the DNA of mTECs by its SAND domain and regulates the expression of many downstream genes in clusters by binding their promoter or enhancer regions, ubiquitously and promiscuously synthesising TRAs as self-epitopes which are presented on MHC/HLA molecules of mTECs [53,54]. Incoming T-cells are immediately selected and clonally deleted if their TCRs recognise and interact with self-epitopes of mTECs because they seem to be autoimmune; moreover, medullar cells are also stopped due to apoptosis [53,54]. Binding AIRE to mTEC’s DNA is an essential step to establish a self-tolerant, mature T-cell repertoire. The DNA-binding SAND-domain coding region of *AIRE* plays an equally essential role in this process. Intronic and Exon 5 and 6 mutations in the SAND-domain coding sequence contribute to disrupted central tolerance, resulting in different forms of APECED [55–58].

The association of allelic polymorphisms and RA risk was initially identified in Asian populations, particularly in Japanese GWASs. In these study, it was found that rs2075876 and rs760426 were significantly associated with RA risk, with  $p$ -values of  $5.1 \times 10^{-4}$  and  $2.0 \times 10^{-4}$ , respectively, among 471 SNPs [30,59–63]. Our study population accurately reflects the European population, and it is important to note that the allele frequencies and functions of the rs2075876 locus may differ in RA development compared to those in Asian populations. Possibly, that is why we could not find significant associations of rs2075876 with RA.

We believe the allelic polymorphisms in the DNA-binding SAND domain of this indispensable gene may lead to decreased AIRE expression, resulting in a diminished representation of TRAs during negative selection and allowing self-reactive thymocytes to evade elimination. The persistence of mature autoreactive CD4<sup>+</sup> T-cells promotes an environment conducive to RA, in which recruited autoimmune CD4<sup>+</sup> T-cells in the synovium contribute to the development of inflammation by producing tumour necrosis factor-alpha, interleukins, and other proinflammatory cytokines, leading to tissue damage [64]. To associate the genetic exposure and disease endpoint part of our hypothesis, we intended to analyse five loci, for which the allelic polymorphism of rs878081 in Exon 5 of the SAND domain, rs1003853 in the Intron 5 splice site, and rs1003854 in the Intron 7 splice site were associated with RA risk.

To delve further into causality, we analysed the molecular consequences of our alleles of risk. The HaploReg database predicted that the two synonymous variations that reference the C-alleles of Exon 5 rs878081 and Exon 6 rs1055311 have lower affinities of binding to the DNA-binding motif of the NF- $\kappa$ B family, although rs1055311 was not associated with an RA risk but with a high level of aCCP and disease activity. NF- $\kappa$ B (RANK) signalling is essential in determining central and peripheral tolerance [65]. Numerous in vivo studies have demonstrated that RANK signaling significantly influences the expression of AIRE in both the thymus and extrathymic tissues. This is mediated through an evolutionarily conserved upstream NF- $\kappa$ B-responsive enhancer region. Studies on RANK-deficient mice have shown a decrease in thymic AIRE expression in mTECs, leading to the development of severe organ-specific autoimmunity. As a result, the NF- $\kappa$ B pathway plays a crucial role in regulating the expression of TRAs and the CCL19, CCL21, and CCR7 chemokines, which are responsible for facilitating the migration of thymocytes from the cortex to the medulla. [66,67]. NF- $\kappa$ B binding to the DNA of mTECs seemed to be crucial to *AIRE* expression. Thus, an allele-specific affinity for the transcriptional factor binding motif may provide different expression levels of *AIRE*. It is suggested that having two copies of low-affinity rs878081 C-alleles in homozygotes may genetically result in the lowest levels of AIRE and TRA expression, potentially leading to a higher number of surviving

mature autoreactive CD4<sup>+</sup> T-cells during development. This genetic variation may account for the significant association of the CC genotype with an increased risk of RA in the recessive genetic model. Within the RA population, it was observed that individuals with CC homozygosity exhibited significantly higher levels and a positive association with ESR, indicating disease activity. This finding supports the idea that genotype-dependent *AIRE* expression may not only determine one's risk but also influence disease activity through clinical characteristics.

Furthermore, the T-allelic splice variant rs1003854 located in Intron 7 demonstrated a lower affinity to bind DNA-binding motifs of the GATA family, which is essential in the early and late stages of thymic and CD4<sup>+</sup> thymocyte development from lymphoid progenitors [68]. In this transcriptional factor family, GATA3 governs the expression of TCR subunits, impacting  $\beta$ -selection and determining the commitment of CD4<sup>+</sup> cell lineages in central tolerance [68]. As this allelic polymorphism is intronic, it is possible that it could influence alternative splicing through intron-mediated enhancement, thereby affecting *AIRE* expression [69]. Our research emphasizes the significance of allelic polymorphisms in negative selection, impacting the immunologic educator mTECs, the objects of selection, the naïve thymocytes, and their TCR development and  $\beta$ -selection.

The allelic and genotype frequencies of rs1055311 do not show a significant association with RA but are correlated with aCCP levels. The allelic polymorphism may be indirectly related to the disease through an intermediate variable, serum aCCP levels, without impacting the overall RA risk. Additionally, it may influence the survival of autoreactive T-cells that internally promote aCCP antibody production in an epistatic manner, characterising an intermediate phenotype of the disease. It is our view that the chosen allelic polymorphisms may impact both immunologic educators and educated sides.

Our research yields important tripartite clinical implications. Firstly, we have successfully associated the *AIRE* gene, crucial for neonatal self-tolerance with the onset of the disease. The presence of rs878081, rs1003854, and rs1003853 risk variants suggests a potential neonatal origin of RA, offering new insights into its development. Secondly, we found that rs878081 and rs1055311 are genotypically associated with ESR and aCCP levels, respectively, in an RA population, providing further evidence that the lack of neonatal negative selection of self-reactive CD4<sup>+</sup> T-cells against citrullinated peptides may lead to significantly higher disease activity and may explain the age-related aggregation of citrullination and autoantibodies [70]. Thirdly, *AIRE* variants are suitable for assessing the risk of RA, even in cases in which the patient is HLA-negative. Since the biological timing of the onset of RA starts around the neonatal period, there is longer time for intervention and monitoring. Our initial approach involves the study of *AIRE*, the principal regulator in central tolerance, and its DNA-binding coding region, given its pivotal role. It is crucial to note that TRA production relies on DNA binding. Once we achieve significant results, our next step will involve genetically mapping additional domain coding regions and analysing the genetics of the *AIRE* partners. We aim to broaden our focus beyond the medullary centre to include the corticomedullary and cortical regions, and ultimately the peripheral regions. Considering the intricate nature of neonatal self-tolerance, a stepwise analysis of the involved proteins and their coding regions is imperative.

Our limited ability to significantly associate allelic polymorphisms and clinical phenotypes with RA risk was mainly due to constraints in our sample size. Increasing the number of participants in our study would have significantly boosted our statistical power and provided greater certainty about the effect size. A further limitation is the scope of one domain coding region, SAND, whose results are not able to provide insight on *AIRE*. However, a genome-wide analysis of all allelic polymorphisms in the coding regions of all potential partners and their outcomes would not yield meaningful results without isolating and examining each player individually. Our findings could have been greatly corroborated with the analysis of the mRNA expression of *AIRE*, specifically associated with the five loci. To fully understand the link between RA and central tolerance, as well as the role of

AIRE partners, it is necessary to obtain further genetic association results for the remaining 11 intronic and 12 exonic coding sequences.

## 5. Conclusions

In conclusion, our study's findings are based on the analysis of the AIRE coding sequence. We have established significant associations between allelic polymorphisms and genotypes of the SAND domain coding Exon 5 and the Intron 5 and 7 splicing sites in relation to RA. We have also observed that the genetic subgroups of rs878081 and rs1055311 among the cases were significantly associated and correlated with the clinical parameters of the ESR and aCCP levels, respectively. These findings suggest a possible role of structural domain coding regions in AIRE in the development of RA, with implications for the potential origin of RA during neonatal self-tolerance. Moving forward, we aim to broaden our investigation to other domains with regulatory and epigenetic elements in order to enhance our understanding of gene function.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biology13060439/s1>, Figure S1: Result plot of allelic discrimination test of rs878081 among RA patients. Figure S2: Result plot of allelic discrimination test of rs878081 among control subjects. Table S1: Allele and genotype frequencies, association of RA with allelic polymorphism rs2075876 in AIRE. Table S2: Allele and genotype frequencies, the association of RA with allelic polymorphism rs1055311 in AIRE.

**Author Contributions:** Conceptualization, B.B. and Z.G.; data curation, B.B., N.N. and Á.K.; formal analysis, B.B. and N.N.; methodology, B.B., N.N., B.N. and Z.G.; resources, B.B., Z.G., N.N. and Á.K.; investigation, B.B., N.N., I.P., Z.K., B.N., Á.K. and Z.G.; validation, B.B., N.N. and Z.G.; writing—original draft, B.B., N.N. and Z.G.; writing review and editing, N.N., I.P., B.N., Z.K., Á.K. and Z.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the ÁOK János Szolcsányi Research Fund (Z.G.; Grant Number: KA-304516) supported by the Medical School, University of Pécs, Hungary.

**Institutional Review Board Statement:** Our study was conducted following the guidelines of the Declaration of Helsinki and approved by the Secretary of the Medical Research Council, Scientific and Research Ethics Committee of Hungary (protocol numbers 11871-7/2018/EÜIG as of 24 April 2018 and 57142-5/2022/EÜIG as of 8 November 2022).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in this study.

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding authors.

**Acknowledgments:** The authors express their gratitude towards the patients and control subjects who participated in this study. They also acknowledge the diligent laboratory assistance provided by the team at the Harkány Thermal Rehabilitation Centre and St. Andrew Hospital for Rheumatology and Medicinal Spa of Hévíz in Hungary.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Finckh, A.; Gilbert, B.; Hodkinson, B.; Bae, S.C.; Thomas, R.; Deane, K.D.; Alpizar-Rodriguez, D.; Lauper, K. Global epidemiology of rheumatoid arthritis. *Nat. Rev. Rheumatol.* **2022**, *18*, 591–602. [[CrossRef](#)] [[PubMed](#)]
2. McInnes, I.B.; Schett, G. The pathogenesis of rheumatoid arthritis. *N. Engl. J. Med.* **2011**, *365*, 2205–2219. [[CrossRef](#)] [[PubMed](#)]
3. Sakkas, L.I.; Bogdanos, D.P.; Katsiari, C.; Platsoucas, C.D. Anti-citrullinated peptides as autoantigens in rheumatoid arthritis—relevance to treatment. *Autoimmun. Rev.* **2014**, *13*, 1114–1120. [[CrossRef](#)] [[PubMed](#)]
4. Liu, J.; Gao, J.; Wu, Z.; Mi, L.; Li, N.; Wang, Y.; Peng, X.; Xu, K.; Wu, F.; Zhang, L. Anti-citrullinated Protein Antibody Generation, Pathogenesis, Clinical Application, and Prospects. *Front. Med.* **2022**, *8*, 802934. [[CrossRef](#)] [[PubMed](#)]
5. Mellado, M.; Martínez-Muñoz, L.; Cascio, G.; Lucas, P.; Pablos, J.L.; Rodríguez-Frade, J.M. T Cell Migration in Rheumatoid Arthritis. *Front. Immunol.* **2015**, *6*, 384. [[CrossRef](#)] [[PubMed](#)]
6. Sokolove, J.; Bromberg, R.; Deane, K.D.; Lahey, L.J.; Derber, L.A.; Chandra, P.E.; Edison, J.D.; Gilliland, W.R.; Tibshirani, R.J.; Norris, J.M.; et al. Autoantibody epitope spreading in the pre-clinical phase predicts progression to rheumatoid arthritis. *PLoS ONE* **2012**, *7*, e35296. [[CrossRef](#)] [[PubMed](#)]



7. Amariuta, T.; Luo, Y.; Knevel, R.; Okada, Y.; Raychaudhuri, S. Advances in genetics toward identifying pathogenic cell states of rheumatoid arthritis. *Immunol. Rev.* **2020**, *294*, 188–204. [[CrossRef](#)] [[PubMed](#)]
8. MacKay, K.; Eyre, S.; Myerscough, A.; Milicic, A.; Barton, A.; Laval, S.; Barrett, J.; Lee, D.; White, S.; John, S.; et al. Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum.* **2002**, *46*, 632–639. [[CrossRef](#)] [[PubMed](#)]
9. Kurkó, J.; Besenyei, T.; Laki, J.; Glant, T.T.; Mikecz, K.; Szekanecz, Z. Genetics of rheumatoid arthritis—A comprehensive review. *Clin. Rev. Allergy Immunol.* **2013**, *45*, 170–179. [[CrossRef](#)] [[PubMed](#)]
10. Okada, Y.; Terao, C.; Ikari, K.; Kochi, Y.; Ohmura, K.; Suzuki, A.; Kawaguchi, T.; Stahl, E.A.; Kur-reeman, F.A.; Nishida, N.; et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat. Genet.* **2012**, *44*, 511–516. [[CrossRef](#)] [[PubMed](#)]
11. Okada, Y.; Wu, D.; Trynka, G.; Raj, T.; Terao, C.; Ikari, K.; Kochi, Y.; Ohmura, K.; Suzuki, A.; Yoshida, S.; et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **2014**, *506*, 376–381. [[CrossRef](#)] [[PubMed](#)]
12. Ishigaki, K.; Sakaue, S.; Terao, C.; Luo, Y.; Sonehara, K.; Yamaguchi, K.; Amariuta, T.; Too, C.L.; Laufer, V.A.; Scott, I.C.; et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat. Genet.* **2022**, *54*, 1640–1651. [[CrossRef](#)] [[PubMed](#)]
13. Mathis, D.; Benoist, C. Aire. *Annu. Rev. Immunol.* **2009**, *27*, 287–312. [[CrossRef](#)] [[PubMed](#)]
14. Perniola, R. Twenty Years of AIRE. *Front. Immunol.* **2018**, *9*, 98. [[CrossRef](#)] [[PubMed](#)]
15. Kyewski, B.; Klein, L. A central role for central tolerance. *Annu. Rev. Immunol.* **2006**, *24*, 571–606. [[CrossRef](#)] [[PubMed](#)]
16. Klein, L.; Kyewski, B.; Allen, P.M.; Hogquist, K.A. Positive and negative selection of the T cell repertoire: What thymocytes see (and don't see). *Nat. Rev. Immunol.* **2014**, *14*, 377–391. [[CrossRef](#)]
17. Derbinski, J.; Schulte, A.; Kyewski, B.; Klein, L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat. Immunol.* **2001**, *2*, 1032–1039. [[CrossRef](#)] [[PubMed](#)]
18. Engelmann, R.; Biemelt, A.; Cordshagen, A.; Johl, A.; Kuthning, D.; Müller-Hilke, B. The Prerequisites for Central Tolerance Induction against Citrullinated Proteins in the Mouse. *PLoS ONE* **2016**, *11*, e0158773. [[CrossRef](#)] [[PubMed](#)]
19. Constantine, G.M.; Lionakis, M.S. Lessons from primary immunodeficiencies: Autoimmune regulator and autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *Immunol. Rev.* **2019**, *287*, 103–120. [[CrossRef](#)] [[PubMed](#)]
20. Miller, C.N.; Waterfield, M.R.; Gardner, J.M.; Anderson, M.S. Aire in Autoimmunity. *Annu. Rev. Immunol.* **2024**. *epub ahead of print.* [[CrossRef](#)]
21. Walser-Kuntz, D.R.; Weyand, C.M.; Weaver, A.J.; O'Fallon, W.M.; Goronzy, J.J. Mechanisms under-lying the formation of the T cell receptor repertoire in rheumatoid arthritis. *Immunity* **1995**, *2*, 597–605. [[CrossRef](#)] [[PubMed](#)]
22. Goronzy, J.J.; Weyand, C.M. Thymic function and peripheral T-cell homeostasis in rheumatoid arthritis. *Trends Immunol.* **2001**, *22*, 251–255. [[CrossRef](#)] [[PubMed](#)]
23. Lovewell, T.R.; McDonagh, A.J.; Messenger, A.G.; Azzouz, M.; Tazi-Ahnini, R. The AIRE -230Y Polymorphism Affects AIRE Transcriptional Activity: Potential Influence on AIRE Function in the Thymus. *PLoS ONE* **2015**, *10*, e0127476. [[CrossRef](#)] [[PubMed](#)]
24. Tazi-Ahnini, R.; McDonagh, A.J.; Wengraf, D.A.; Lovewell, T.R.; Vasilopoulos, Y.; Messenger, A.G.; Cork, M.J.; Gawkrödger, D.J. The autoimmune regulator gene (AIRE) is strongly associated with vitiligo. *Br. J. Dermatol.* **2008**, *159*, 591–596. [[CrossRef](#)] [[PubMed](#)]
25. Wengraf, D.A.; McDonagh, A.J.; Lovewell, T.R.; Vasilopoulos, Y.; Macdonald-Hull, S.P.; Cork, M.J.; Messenger, A.G.; Tazi-Ahnini, R. Genetic analysis of autoimmune regulator haplotypes in alopecia areata. *Tissue Antigens* **2008**, *71*, 206–212. [[CrossRef](#)] [[PubMed](#)]
26. Eriksson, D.; Bianchi, M.; Landegren, N.; Dalin, F.; Skov, J.; Hultin-Rosenberg, L.; Mathioudaki, A.; Nordin, J.; Hallgren, Å.; Andersson, G.; et al. Common genetic variation in the autoimmune regulator (AIRE) locus is associated with autoimmune Addison's disease in Sweden. *Sci. Rep.* **2018**, *8*, 8395. [[CrossRef](#)] [[PubMed](#)]
27. Ferrera, F.; Rizzi, M.; Spreccacenero, B.; Balestra, P.; Sessarego, M.; Di Carlo, A.; Filaci, G.; Gabrielli, A.; Ravazzolo, R.; Indiveri, F. AIRE gene polymorphisms in systemic sclerosis associated with autoimmune thyroiditis. *Clin. Immunol.* **2007**, *122*, 13–17. [[CrossRef](#)] [[PubMed](#)]
28. Conteduca, G.; Ferrera, F.; Pastorino, L.; Fenoglio, D.; Negrini, S.; Sormani, M.P.; Indiveri, F.; Scarrà, G.B.; Filaci, G. The role of AIRE polymorphisms in melanoma. *Clin. Immunol.* **2010**, *136*, 96–104. [[CrossRef](#)] [[PubMed](#)]
29. Colobran, R.; Giménez-Barcons, M.; Marín-Sánchez, A.; Porta-Pardo, E.; Pujol-Borrell, R. AIRE genetic variants and predisposition to polygenic autoimmune disease: The case of Graves' disease and a systematic literature review. *Hum. Immunol.* **2016**, *77*, 643–651. [[CrossRef](#)] [[PubMed](#)]
30. Bérczi, B.; Gerencsér, G.; Farkas, N.; Hegyi, P.; Veres, G.; Bajor, J.; Czopf, L.; Alizadeh, H.; Ra-konczay, Z.; Vigh, É.; et al. Association between AIRE gene polymorphism and rheumatoid arthritis: A systematic review and meta-analysis of case-control studies. *Sci. Rep.* **2017**, *7*, 14096. [[CrossRef](#)] [[PubMed](#)]
31. García-Lozano, J.R.; Torres-Agrela, B.; Montes-Cano, M.A.; Ortiz-Fernández, L.; Conde-Jaldón, M.; Teruel, M.; García, A.; Núñez-Roldán, A.; Martín, J.; González-Escribano, M.F. Association of the AIRE gene with susceptibility to rheumatoid arthritis in a European population: A case control study. *Arthritis Res. Ther.* **2013**, *15*, R11. [[CrossRef](#)] [[PubMed](#)]
32. Qi, Y.; Zhang, R.; Lu, Y.; Zou, X.; Yang, W. Aire and Fezf2, two regulators in medullary thymic epithelial cells, control autoimmune diseases by regulating TSAs: Partner or complements? *Front. Immunol.* **2022**, *13*, 948259. [[CrossRef](#)] [[PubMed](#)]

33. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at UCSC. *Genome Res.* **2002**, *12*, 996–1006. [[CrossRef](#)] [[PubMed](#)]
34. Cunningham, F.; Allen, J.E.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Austine-Orimoloye, O.; Azov, A.G.; Barnes, I.; Bennett, R.; et al. Ensembl 2022. *Nucleic Acids Res.* **2022**, *50*, D988–D995. [[CrossRef](#)] [[PubMed](#)]
35. Cavalli-Sforza, L.L.; Bodmer, W.F. *The Genetics of Human Populations*, 1st ed.; Dover Publications: San Francisco, CA, USA, 1971; pp. 41–118.
36. Sherry, S.T.; Ward, M.; Sirotkin, K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* **1999**, *9*, 677–679. [[CrossRef](#)] [[PubMed](#)]
37. Fairley, S.; Lowy-Gallego, E.; Perry, E.; Flicek, P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* **2020**, *48*, D941–D947. [[CrossRef](#)] [[PubMed](#)]
38. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alfoldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **2020**, *581*, 434–443. [[CrossRef](#)] [[PubMed](#)]
39. Kowalski, M.H.; Qian, H.; Hou, Z.; Rosen, J.D.; Tapia, A.L.; Shan, Y.; Jain, D.; Argos, M.; Arnett, D.K.; Avery, C.; et al. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet.* **2019**, *15*, e1008500. [[CrossRef](#)] [[PubMed](#)]
40. Arnett, F.C.; Edworthy, S.M.; Bloch, D.A.; McShane, D.J.; Fries, J.F.; Cooper, N.S.; Healey, L.A.; Kaplan, S.R.; Liang, M.H.; Luthra, H.S.; et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **1988**, *31*, 315–324. [[CrossRef](#)] [[PubMed](#)]
41. Ingegnoli, F.; Castelli, R.; Gualtierotti, R. Rheumatoid factors: Clinical applications. *Dis. Markers* **2013**, *35*, 727–734. [[CrossRef](#)] [[PubMed](#)]
42. Niewold, T.B.; Harrison, M.J.; Paget, S.A. Anti-CCP antibody testing as a diagnostic and prognostic tool in rheumatoid arthritis. *QJM* **2007**, *100*, 193–201. [[CrossRef](#)] [[PubMed](#)]
43. Steiner, G.; Smolen, J. Autoantibodies in rheumatoid arthritis and their clinical significance. *Arthritis Res.* **2002**, *4* (Suppl. 2), S1–S5. [[CrossRef](#)]
44. Graham, J.E.; Granger, C.V.; Karmarkar, A.M.; Deutsch, A.; Niewczyk, P.; Divita, M.A.; Ottenbacher, K.J. The Uniform Data System for Medical Rehabilitation: Report of follow-up information on patients discharged from inpatient rehabilitation programs in 2002–2010. *Am. J. Phys. Med. Rehabil.* **2014**, *93*, 231–244. [[CrossRef](#)] [[PubMed](#)]
45. Jou, J.M.; Lewis, S.M.; Briggs, C.; Lee, S.H.; De La Salle, B.; McFadden, S.; International Council for Standardization in Haematology. ICSH review of the measurement of the erythrocyte sedimentation rate. *Int. J. Lab. Hematol.* **2011**, *33*, 125–132. [[CrossRef](#)] [[PubMed](#)]
46. Lewontin, R.C.; Kojima, K. The Evolutionary Dynamics of Complex Polymorphisms. *Evolution* **1960**, *14*, 458–472.
47. HaploReg v4.2. Available online: <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php> (accessed on 25 March 2024).
48. Ward, L.D.; Kellis, M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **2012**, *40*, D930–D934. [[CrossRef](#)] [[PubMed](#)]
49. Gibson, T.J.; Ramu, C.; Gemünd, C.; Aasland, R. The APECED polyglandular autoimmune syndrome protein, AIRE-1, contains the SAND domain and is probably a transcription factor. *Trends Biochem. Sci.* **1998**, *23*, 242–244. [[CrossRef](#)] [[PubMed](#)]
50. Berczi, B.; Nusser, N.; Peter, I.; Nemeth, B.; Gyongyi, Z. Association Between AIRE Polymorphisms rs870881(C>T), rs1003854(T>C) and Rheumatoid Arthritis Risk: A Hungarian Case-control Study. *In Vivo* **2024**, *38*, 774–784. [[CrossRef](#)] [[PubMed](#)]
51. Horita, N.; Kaneko, T. Genetic model selection for a case-control study and a meta-analysis. *Meta Gene* **2015**, *22*, 1–8. [[CrossRef](#)] [[PubMed](#)]
52. Di Matteo, A.; Bathon, J.M.; Emery, P. Rheumatoid arthritis. *Lancet* **2023**, *402*, 2019–2033. [[CrossRef](#)] [[PubMed](#)]
53. Brown, C.C.; Rudensky, A.Y. Conceiving the Inconceivable: The Function of Aire in Immune Tolerance to Peripheral Tissue-Restricted Antigens in the Thymus. *J. Immunol.* **2021**, *206*, 245–247. [[CrossRef](#)] [[PubMed](#)]
54. Kumar, P.G.; Laloraya, M.; Wang, C.Y.; Ruan, Q.G.; Davoodi-Semiro, A.; Kao, K.J.; She, J.X. The autoimmune regulator (AIRE) is a DNA-binding protein. *J. Biol. Chem.* **2001**, *276*, 41357–41364. [[CrossRef](#)] [[PubMed](#)]
55. Bottomley, M.J.; Collard, M.W.; Huggenvik, J.L.; Liu, Z.; Gibson, T.J.; Sattler, M. The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. *Nat. Struct. Biol.* **2001**, *8*, 626–633. [[CrossRef](#)] [[PubMed](#)]
56. Mora, M.; Hanzu, F.A.; Pradas-Juni, M.; Aranda, G.B.; Halperin, I.; Puig-Domingo, M.; Aguiló, S.; Fernández-Rebollo, E. New splice site acceptor mutation in AIRE gene in autoimmune polyendocrine syndrome type 1. *PLoS ONE* **2014**, *9*, e101616. [[CrossRef](#)] [[PubMed](#)]
57. Giordano, C.; Modica, R.; Allotta, M.L.; Guarnotta, V.; Cervato, S.; Masiero, S.; Giordano, R.; Garelli, S.; Betterle, C. Autoimmune polyendocrinopathy-candidiasis-ectodermal-dystrophy (APECED) in Sicily: Confirmation that R203X is the peculiar AIRE gene mutation. *J. Endocrinol. Investig.* **2012**, *35*, 384–388. [[CrossRef](#)] [[PubMed](#)]
58. Aytac, G.; Guven, B.; Aydin, I.; Topyildiz, E.; Aykut, A.; Durmaz, A.; Edeer Karaca, N.; Aksu, G.; Kutukculer, N. An Extraordinary Case of Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy (APECED) Syndrome Misdiagnosed as Juvenile Idiopathic Arthritis on Admission. *Case Rep. Immunol.* **2023**, *2023*, 2363760. [[CrossRef](#)] [[PubMed](#)]

59. Terao, C.; Yamada, R.; Ohmura, K.; Takahashi, M.; Kawaguchi, T.; Kochi, Y.; Human Disease Genomics Working Group; RA Clinical and Genetic Study Consortium; Okada, Y.; Nakamura, Y.; et al. The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Hum. Mol. Genet.* **2011**, *20*, 2680–2685. [[CrossRef](#)] [[PubMed](#)]
60. Shao, S.; Li, X.R.; Cen, H.; Yin, Z.S. Association of AIRE polymorphisms with genetic susceptibility to rheumatoid arthritis in a Chinese population. *Inflammation* **2014**, *37*, 495–499. [[CrossRef](#)] [[PubMed](#)]
61. Feng, Z.J.; Zhang, S.L.; Wen, H.F.; Liang, Y. Association of rs2075876 polymorphism of AIRE gene with rheumatoid arthritis risk. *Hum. Immunol.* **2015**, *76*, 281–285. [[CrossRef](#)] [[PubMed](#)]
62. Li, X.; Li, T.; Chen, M.; Chai, Y. Association of AIRE gene polymorphisms with susceptibility to rheumatoid arthritis among ethnic Han Chinese from Shaanxi. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **2016**, *33*, 373–377. (In Chinese) [[CrossRef](#)]
63. Yang, H.; Li, J.; Jiang, L.; Jiang, X.; Zhou, X.; Xu, N. The rs878081 polymorphism of AIRE gene increases the risk of rheumatoid arthritis in a Chinese Han population: A case-control study. *Braz. J. Med. Biol. Res.* **2018**, *51*, e7944. [[CrossRef](#)] [[PubMed](#)]
64. Yap, H.Y.; Tee, S.Z.; Wong, M.M.; Chow, S.K.; Peh, S.C.; Teow, S.Y. Pathogenic Role of Immune Cells in Rheumatoid Arthritis: Implications in Clinical Treatment and Biomarker Development. *Cells* **2018**, *7*, 161. [[CrossRef](#)] [[PubMed](#)]
65. van Delft, M.A.; Huitema, L.F.; Tas, S.W. The contribution of NF- $\kappa$ B signalling to immune regulation and tolerance. *Eur. J. Clin. Invest.* **2015**, *45*, 529–539. [[CrossRef](#)] [[PubMed](#)]
66. Haljasorg, U.; Bichele, R.; Saare, M.; Guha, M.; Maslovskaia, J.; Kõnd, K.; Remm, A.; Pihlap, M.; Tomson, L.; Kisand, K.; et al. A highly conserved NF- $\kappa$ B-responsive enhancer is critical for thymic expression of Aire in mice. *Eur. J. Immunol.* **2015**, *45*, 3246–3256. [[CrossRef](#)] [[PubMed](#)]
67. Barnabei, L.; Laplantine, E.; Mbongo, W.; Rieux-Laucat, F.; Weil, R. NF- $\kappa$ B: At the Borders of Autoimmunity and Inflammation. *Front. Immunol.* **2021**, *12*, 716469. [[CrossRef](#)] [[PubMed](#)]
68. Wan, Y.Y. GATA3: A master of many trades in immune regulation. *Trends Immunol.* **2014**, *35*, 233–242. [[CrossRef](#)] [[PubMed](#)]
69. Rose, A.B. Intron-mediated regulation of gene expression. *Curr. Top. Microbiol. Immunol.* **2008**, *326*, 277–290. [[CrossRef](#)] [[PubMed](#)]
70. Osaki, D.; Hiramatsu, H. Citrullination and deamidation affect aggregation properties of amyloid  $\beta$ -proteins. *Amyloid* **2016**, *23*, 234–241. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.