Analysis of RA Symptom Severity using Transcriptional Profiling and Genetics

Singh AK [1], Lekstrom-Himes J [1], Meyer J [1], Ginsburg GS [1], Shadick NA [2], Maher NE [2], Weinblatt ME [2], Fedyk E [1],Koeing E[1], Pickard M[1], Tirrell S[1], Parker A [1]

1. Millennium Pharmaceuticals, Cambridge, MA; 2. Brigham and Women's Hospital, Boston MA

INTRODUCTION

BRASS: Brigham and Women's Rheumatoid Arthritis Sequential Study

Collaboration with Brigham and Women's Hospital

- BRASS: prospective cross-sectional study
  - Structure
    - 1000 patients followed biannually for 5 years (30% new onset RA)
  - Data/specimens collected
    - Physician data (yearly)
    - Patient reported data (q6mon)
    - DNA (once)
    - RNA (yearly)
    - Serum (yearly)
    - Whole blood (fresh; yearly)

- A well annotated sample bank at Millennium and Harvard Partners for future discovery
- Large registry structured to provide epidemiological data
- Opportunity to test biological hypotheses in disease-related pathways
- Potential to do biomarker discovery and validation activities for application towards
  - Early phase clinical trials (internal decision making around compound efficacy)
  - Late phase clinical trials (efficacy markers, pharmacogenomic markers)
  - Clinical assessment tools for practicing physicians
  - Opportunities for development of molecular diagnostics

BACKGROUND

- Predisposition to rheumatoid arthritis has been associated with several genetic variants, including the HLA-DRB1 shared epitope (SE)
- The effect of genetic variation may be mediated through RNA expression:

  GENE → mRNA → OUTCOME
  - Genotype dependent gene expression patterns have not been studied in relation to clinical severity of RA
  - Genotype-specific mRNA biomarkers that change with disease severity would
    - Facilitate design and interpretation of clinical trials
    - Help define disease mechanisms that are dependent upon genetic background

HYPOTHESIS

Do mRNA expression patterns interact with risk-associated genotypes to predict clinical severity of RA?

METHODS

- Samples
  - RNA extracted from whole blood of RA patients
- Data Generation
  - Eliminated RBC signature using globin mRNA reduction protocol
  - Labeled and hybridized to Affymetrix gene arrays (U133A array; 22,283 probesets)
- Analysis Strategy

Discovery Data Set

Simple correlation of gene expression data from RA patient samples with disease severity and risk/severity associated genetic variants

Validation Data Set

Patients were selected to represent high (DAS 28 >5.1) and low (DAS 28 < 3.2) disease activity within each of the three HLA-DRB1 shared epitope genotypic classes. The purpose of this validation study is to identify associations of mRNA expression levels with HLA genotype. Interactions between HLA SE, gene expression, and disease severity were further explored.

DERIVED VARIABLES

Genetic Variants

Table 1: Genetic Variables -Dichotomized patient cohort based on presence or absence of the 12 risk- and/or severity-associated genetic variants

GENE	Number of Variants	Associated With	
		Risk	Severity
TNF	2	2	
FCGR3A	1	1	
NFKBIL1	2	2	1
FcGR2A	1		1
IL18	1		1
TNFRSF11B	1		1
MIF	1		1
HLA_DRB1	1	1	1
IL1B	1	1	1
SLC11A1	1	1	1

DAS 28

DAS28-3(crp) = [0.56*sqrt(TJC28) + 0.28*sqrt(SJC28) + 0.36*ln(CRP+1)] * 1.10 + 1.15

TJC28: 28 Tender joint count, SJC28: 28 Swollen joint count, CRP: C-reactive protein (mg/L)

- High Disease Activity- DAS28 above 5.1 -Low Disease Activity - DAS28 below 3.2

STATISTICAL METHODS

- Univariate association of mRNA expression with DAS28

  Test Statistics
  - SNR (signal to noise ratio)= (μ1 - μ2)/(σ1+ σ2)
  - Poof- average thresholded fold change
  - Permutation testing used correct p values for multiple hypothesis testing

- Association analysis of mRNA expression pattern and HLA_DRB1 genotype

  Test Statistics
  - ANOVA used to test multiple groups for differential gene expresion simultaneously
  - FDR used to correct for multiple hypothesis testing

- Cluster analysis
  - Used to identify homogenous subgroups of patients. Dendrograms (tree diagrams) depict patients with high similarity as adjacent. Depth of nodes indicates the degree of similarity or dissimilarity among patients.

- Multivariate logistic regression
  - Used to analyze the response variable (DAS_28) with respect to HLA_DRB1 and gene expression data.
  - FDR used to correct for multiple hypothesis testing

RESULTS

Patient Characteristics by HLA Genotype

Table 2: Discovery Data Set

No. HLA Alleles	0	1	2
No. Pts (%)	22 (35%)	22 (35%)	18 (29%)
Age mean (sd)	54.8 (± 15.8)	60.9 (± 13.4)	63.7 (± 8.0)
No. Females (%)	17 (77%)	20 (91%)	13 (72%)
No. on anti-TNF tx (%)	5 (23%)	3 (14%)	4 (22%)
DAS28 mean (sd)	4.28 (± 1.69)	4.39 (± 1.31)	4.87 (± 1.28)

Table 3: Validation Data Set

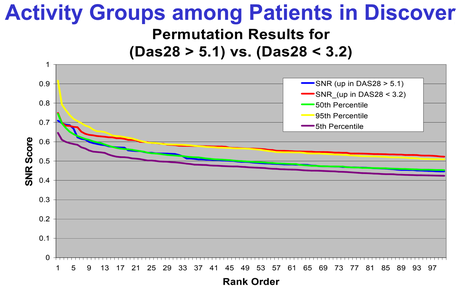
No. HLA Alleles	0	1	2
No. Pts (%)	24 (32%)	31 (42%)	19 (26%)
Age mean (sd)	57.4 (± 14.5)	61.2 (± 10.6)	55.3 (± 12.3)
No. Females (%)	20 (83%)	27 (87%)	16 (84%)
No. on anti-TNF tx (%)	8 (33%)	15 (48%)	8 (50%)
DAS28 mean (sd)	3.24 (± 1.87)	4.62 (± 2.15)	4.31 (± 2.19)
LOW Cohort No. [mean(sd)]	16 1.98 (± 0.40)	12 2.06 (± 0.47)	8 1.92 (± 0.51)
HIGH Cohort No. [mean(sd)]	8 5.75 (± 0.55)	19 6.23 (± 0.66)	11 6.05 (± 0.74)

Table 4: Discovery Data Set- Differential gene expression between individuals of different genotype

GENE	Significant Findings (p values < 0.07)*
TNF	X
FCGR3A	
NFKBIL1	X
FcGR2A	
IL18	
TNFRSF11B	
MIF	X
HLA-DRB1	X
IL1B	
SLC11A1	X

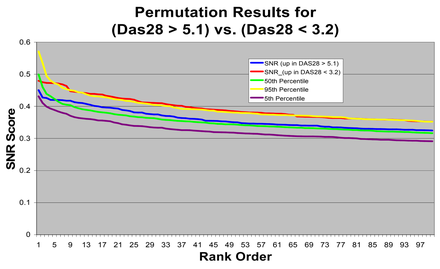
*Significant differential expression was observed when comparing presence/absence of allelic variants

Chart 1: Permutation Test Results using SNR comparing High vs. Low Disease Activity Groups among Patients in Discovery Data Set



Probe Set	LocusLink description	SNR	P-Value	Corrected P-Value
219673_at	hypothetical protein FLJ20170	0.74877	0.0001756	0.48
211894_x_at	seizure related gene 6 (mouse)-like	0.69767	0.0007346	0.51
222253_s_at	DKFZp434P211 protein	0.68338	0.0006993	0.35
201014_s_at	multifunctional polypeptide similar to SAICAR synthetase and Ai	0.68204	0.0012419	0.26
214975_s_at	myotubularin related protein 1	0.6771	0.0013398	0.23
204937_s_at	zinc finger protein 274	0.67527	0.0003894	0.16
217668_at		0.65188	0.0002758	0.22
205004_at	transcription factor NRF	0.64131	0.003683	0.24
201284_s_at	N-acylaminoacyl-peptide hydrolase	0.63555	0.0008528	0.2
213077_at	hypothetical protein FLJ21940	0.63326	0.0005695	0.17
202020_s_at	LanC (bacterial lantibiotic synthetase component C)-like 1	0.62949	0.0009535	0.15
203717_at	dipeptidylpeptidase IV (CD26, adenosine deaminase complexin	0.62814	0.0080016	0.15
220893_at	uncharacterized gastric protein ZA52P	0.62587	0.0009686	0.14
200050_at	zinc finger protein 146	0.62304	0.000412	0.11
215260_s_at	transcription factor 3 (E2A immunoglobulin enhancer binding fa	0.6226	0.0013188	0.09
213088_s_at	KIAA0974 protein	0.61844	0.001507	0.08
202960_s_at	methylmalonyl Coenzyme A mutase	0.618	0.0031556	0.07
219715_s_at	hypothetical protein FLJ11090	0.61545	0.0023377	0.07
215794_x_at	glutamate dehydrogenase 2	0.60978	0.0012358	0.08
216713_at	cerebral cavernous malformations 1	0.60906	0.0009225	0.08
217687_at	SEC14 (S. cerevisiae)-like 1	0.60806	0.0026985	0.08
203733_at	MYLE protein	0.5994	0.0017798	0.09
208136_s_at	hypothetical protein MGC3771	0.59902	0.0011639	0.07
57082_at	LDL receptor adaptor protein	0.59774	0.0008558	0.06
205565_s_at	Friedreich ataxia	0.59396	0.0044104	0.07

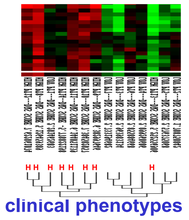
Chart 2: Permutation Test Results using SNR comparing High vs. Low Disease Severity Groups among Patients in Validation Data Set



Probe Set	LocusLink	SNR	P-Values	Corrected P-Values
210028_s_at	origin recognition complex, subunit 3 (yeast homolog)-like	0.47957	0.000186	0.64
212750_at	protein phosphatase 1, regulatory (inhibitor) subunit 16B	0.47525	0.0002248	0.35
218764_at	protein kinase C, eta	0.47283	0.00015	0.2
205097_at	solute carrier family 26 (sulfate transporter), member 2	0.47257	0.000147	0.11
209827_s_at	interleukin 16 (lymphocyte chemoattractant factor)	0.47219	0.000163	0.07
213794_s_at	DKFZP564O092 protein	0.47108	0.0001158	0.04
206633_at	cholinergic receptor, nicotinic, alpha polypeptide 1 (muscle)	0.46846	0.000463	0.03
219531_at	hypothetical protein FLJ10565	0.46361	0.00025	0.04
201189_s_at	inositol 1,4,5-triphosphate receptor, type 3	0.44716	0.0004422	0.07
204014_at	dual specificity phosphatase 4	0.44545	0.0005502	0.07
212314_at	KIAA0746 protein	0.44402	0.0003339	0.06
200050_at	zinc finger protein 146	0.44227	0.0003374	0.07
210645_s_at	tetratricopeptide repeat domain 3	0.44088	0.00037	0.06
220896_at	hypothetical protein FLJ11467	0.43885	0.0003388	0.06
209306_s_at	SWAP-70 protein	0.43868	0.0007701	0.05
219908_at	dickkopf (Xenopus laevis) homolog 2	0.43732	0.0008841	0.05
220097_s_at	hypothetical protein FLJ20255	0.43692	0.00086	0.04
216221_s_at	pumilio (Drosophila) homolog 2	0.43289	0.0004484	0.04
207097_at	solute carrier family 16 (monocarboxylic acid transporters), member 7	0.43093	0.0005527	0.04
204143_s_at	rTS beta protein	0.42853	0.0009945	0.05
213455_at	hypothetical gene supported by AB011149; NM_003587	0.42609	0.0009649	0.05
215168_at	translocase of inner mitochondrial membrane 17 (yeast) homolog A	0.42424	0.0005598	0.04
221290_s_at	hypothetical protein FLJ14868	0.42344	0.0023368	0.04
211985_s_at		0.42292	0.0005644	0.04
221969_at		0.42051	0.0006893	0.04

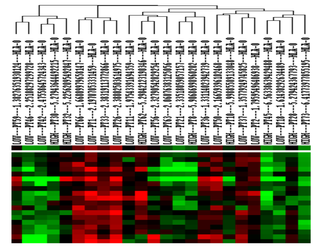
Discovery Data Set: TxP marker discovery for RA clinical phenotypes may only be possible in genetically stratified populations

- Patient cohort stratified by HLA-DRB1 shared epitope carrier status
- Carrier group has high DAS scores with little variation among patients
- Non-carrier group has more variation, significant gene expression predictors of DAS score



Validation Data Set: TxP marker discovery for RA clinical phenotypes among genetically stratified populations

- Stratified by HLA-DRB1 Shared Epitope carrier status
- Carrier group has high DAS scores with little variation
- Non-carrier group has low DAS scores with little variation



Genes Upregulated in HLA-DRB1 SE carriers

Table 5: Discovery Dataset

Probe Set	LocusLink description	Adjusted Fold Change (up in HLA-DRB1+ vs. HLA-DRB1 -)	p value	Corrected p value
209728_at	major histocompatibility complex, class II, DR beta 4	6.23297	0.004647	0.01
212671_s_at	major histocompatibility complex, class II, DQ alpha 1; majo	3.84225	0.001411	0.09
AFFX-HUMRGE/M10098_3_at	MAP/microtubule affinity-regulating kinase 3	2.33328	0.214579	0.57
AFFX-r2-Hs18S/rRNA-3_s_at	MAP/microtubule affinity-regulating kinase 3	2.32162	0.211314	0.3
204141_at	tubulin, beta polypeptide	1.9641	0.397816	0.74
217022_s_at	similar to immunoglobulin heavy chain/T-cell receptor J-alpl	1.89815	0.977572	0.74
205950_s_at	carbonic anhydrase I	1.60108	0.925711	1
203911_at	RAP1, GTPase activating protein 1	1.58768	0.653959	1
213515_x_at	hemoglobin, gamma G	1.55734	0.522402	1
203153_at	interferon-induced protein with tetratricopeptide repeats 1	1.54484	0.709036	1

Table 6: Validation Dataset

Probe Set	LocusLink description	Adjusted Fold Change (up in HLA-DRB1+ vs. HLA-DRB1 -)	p values	Corrected p values
209728_at	major histocompatibility complex, class II, DR beta 4	7.5417	0.0005209	0.01
204141_at	tubulin, beta polypeptide	4.27241	0.0422087	0.01
212671_s_at	major histocompatibility complex, class II, DQ alpha 1; major	3.47819	0.0015678	0.01
202018_s_at	lactotransferrin	2.10834	0.1933442	0.65
AFFX-HUMRGE/M10098_3_at	MAP/microtubule affinity-regulating kinase 3	1.94015	0.7815876	0.73
206371_at	folate receptor 3 (gamma)	1.92282	0.0860414	0.6
AFFX-r2-Hs18S/rRNA-3_s_at	MAP/microtubule affinity-regulating kinase 3	1.91295	0.8140173	0.46
205403_at	interleukin 1 receptor, type II	1.84263	0.2993491	0.52
207269_at	defensin, alpha 4, corticostatin	1.78157	0.1964196	0.59
211372_s_at	interleukin 1 receptor, type II	1.73724	0.3549031	0.61

Table 7: Validation Dataset- Gene vs. HLA ANOVA Results

Probeset	Locus Link Description	ANOVA Raw p value	ANOVA FDR Corrected p value
203290_at	no annotation found - check ncbi	0.0000	0.0011
209728_at	D-major histocompatibility complex, class II, DR beta 4	0.0000	0.0016
212671_s_at	major histocompatibility complex, class II, DQ alpha 1	0.0000	0.0032
221949_at	D-hypothetical gene supported by AK056566	0.0000	0.0043
221491_x_at	D-major histocompatibility complex, class II, DR beta 1	0.0000	0.0386

Table 8: Multivariate Logistic Regression Model: Assessment of Whether Gene Expression and HLA_DRB1 Genotype Interact to Predict Disease Activity (High vs. Low DAS Score)

Top 10 Gene

Probset	Locus Link	Gene*HLA Interaction Raw p	Interaction FDR p	Gene Raw p	Gene FDR p
207284_s_at	aspartate beta-hydroxylase	0.4046	0.9250	0.0014	0.8602
212750_at	protein phosphatase 1, regulatory (inhibitor) subunit 16B	0.3481	0.9105	0.0020	0.8602
219908_at	dickkopf (Xenopus laevis) homolog 2	0.2553	0.8839	0.0021	0.8602
215135_at	aspartyl aminopeptidase	0.3159	0.9005	0.0023	0.8602
205557_at	bactericidal/permeability-increasing protein	0.3368	0.9080	0.0023	0.8602
210127_at	RABGB, member RAS oncogene family	0.8303	0.9851	0.0026	0.8602
211868_x_at	similar to Ig heavy chain precursor V region (VDH26) - h	0.0374	0.8711	0.0028	0.8602
204014_at	dual specificity phosphatase 4	0.8593	0.9868	0.0029	0.8602
221662_s_at	solute carrier family 22 (organic anion transporter), mem	0.9162	0.9932	0.0029	0.8602
206871_at	elastase 2, neutrophil	0.1266	0.8711	0.0030	0.8602

Top 10 Gene*HLA Interaction Genes

Probeset	Locus Link	Gene*HLA Interaction Raw p	Interaction FDR p	Gene Raw p	Gene FDR p
219258_at	hypothetical protein FLJ20516	0.0014	0.8711	0.1443	0.8602
216893_s_at	collagen, type IV, alpha 3 (Goodpasture antigen)	0.0016	0.8711	0.8248	0.9873
220426_at	hypothetical protein MGC5366	0.0021	0.8711	0.3571	0.9264
206159_at	growth differentiation factor 10	0.0027	0.8711	0.6860	0.9760
210028_s_at	origin recognition complex, subunit 3 (yeast homolog)-like	0.0031	0.8711	0.0032	0.8602
203670_at	DKFZP434B103 protein	0.0035	0.8711	0.1002	0.8602
205413_at	chromosome 11 open reading frame 8	0.0036	0.8711	0.8680	0.9917
206331_at	calcitonin receptor-like	0.0039	0.8711	0.8570	0.9910
202987_at	chromosome 6 open reading frame 4	0.0047	0.8711	0.1032	0.8602
213354_s_at	nuclear receptor subfamily 2, group F, member 6	0.0048	0.8711	0.1012	0.8602

CONCLUSIONS

- Presence of a disease-associated HLA-DRB1 genotype is associated with reproducible differential expression of HLA-DRB4, HLA-DQA1, MARK3, and beta-tubulin in peripheral blood cell populations
- Differential gene regulation is also observed between high and low disease activity patients, but a common gene set has not yet emerged
- Since HLA-DRB1 is itself associated with disease activity, a larger patient cohort is needed to fully characterize the interaction between disease activity, HLA-DRB1 genotype and mRNA expression