A national genome wide association study to predict the response to treatment in rheumatoid arthritis patients

Proponents - Helena Canhão and João Eurico Fonseca

Participants - All the rheumatology centers are invited to participate.

Co-authors (from participating centers) - 2 authors from HEM, HUC, HGO, HSJ, IPR, HPonte Lima and 2 additional authors from HSM.
1 author from each participating small center.
From laboratory and bioinformatics work: Robert Plenge, Daniel Solomon, Eli Stahl, Diana Fernandes and Fernando Martins.

Duration - 4 years

INTRODUCTION

Rheumatoid arthritis (RA) is the most prevalent inflammatory chronic joint disease, affecting ~1% of the population worldwide. It is characterized by sustained inflammation that leads to joint destruction, disability and increased morbidity and mortality.

In the last 10 years more effective therapies have emerged and although not curative, they showed the ability to suppress inflammation and arrest structural damage.

Drugs that block tumor necrosis factor (TNF) are currently the first choice for treatment in patient refractory to methotrexate and other conventional disease modifying anti-rheumatic drugs (DMARDs). Despite anti-TNF have proven their efficacy and safety in a large proportion of patients, ~30% of RA patients fail to respond or show toxicity, delaying inflammation control with the subsequent development of erosions and irreversible joint damage. In addition, in clinical practice, only ~20% of RA patients achieve remission with this drug's treatment which eventually leads to switch to other class of biological drugs, targeted to different cytokines or cells.

Long-term outcome in patients with RA is highly dependent upon aggressive pharmacological control of inflammation early in the disease course. Despite the importance of selecting the optimal medication soon after disease onset, there is no validated biomarker predictor of severity neither of drug treatment response. As a consequence, RA patients often suffer irreversible joint destruction while a physician searches for an effective drug combination.

A genetic biomarker would be particularly useful for drugs that block the inflammatory cytokine TNF, as these drugs are first-line biological DMARDs yet, as said before, induce remission in only ~20% of patients.

The absence of genetic biomarkers for RA susceptibility, prognosis severity and for predicting response to anti-TNF therapy is largely due to limitations in conducting and interpreting genetic studies.

Very few genetic studies of response to anti-TNF therapy have been conducted, and there is no single nucleotide polymorphism (SNP) that reproducibly predicts response.

A handful of candidate gene studies, based on RA susceptibility alleles and known biology of the TNF pathway, have been conducted. These studies suggest the MHC region may be important in predicting response to anti-TNF therapy, although this effect is not observed in all populations.

Furthermore, no single allele within the MHC was associated across all studies.

We have performed one of these studies evaluating 22 infliximab treated patients at 24 weeks and we found a better response in patients with the -308 GG alleles (JE Fonseca et al. Ann Rheum Dis 2005).

Important limitations of published genetic studies of TNF treatment outcome are the small sample size and the heterogeneous definition of treatment and of treatment response. Some studies used a quantitative change in the DAS28 (delta-DAS before and after treatment), others have used the EULAR response criteria, and yet others have used the ACR response criteria. Some studies have focused on patients who have received anti-TNF therapy for the first time, while other studies allowed patients treated with multiple anti-TNF agents and other biological DMARDs. These variable definitions make it difficult to combine results across studies.

However in the last years new techniques have emerged and a huge advance in genetic and genomic research has occurred.

It is increasingly practical to test systematically common genetic variants for their role in disease – genome-wide association studies (GWAS) – but this has not yet been done in a large collection of RA patients treated with anti-TNF therapy. Common SNP variants (i.e., those present at a population frequency >1%) explain much of the genetic diversity in our species, a consequence of the historically small size and shared ancestry of the human population. There are approximately 10 million common SNPs in the human genome. GWAS of common variants are now feasible due to an improved understanding of linkage desequilibrium (LD) structure across the human genome, technical capacity to genotype hundreds of thousands of SNPs, and methods to analyze large datasets. GWAS have a distinct advantage over previous genetic studies in that they are able to test, in an unbiased manner, the majority of common variants across the genome in a single experiment. Implementation of GWAS has greatly expanded the number of true positive loci (i.e., those that have been replicated consistently in more than one study at a high level of statistical confidence) that are associated with complex traits. Only a small GWAS of 89 patients has been conducted in RA patients to address the response to therapy.

It is now clear from several studies of the genetic basis of RA that the vast majority of individual common DNA variants associated with complex traits are of modest effect size. Most individual SNPs increase risk of disease by only ~10% (corresponding odds ratio (OR) of ~1.10). As a consequence, thousands of patient samples are required to achieve convincing statistical evidence in favor of a true positive association at a stringent level of significance (e.g., $P<5x10-8$) for a single SNP. Even then, validated SNP associations explain only a small fraction of the genetic contribution to variance of most complex traits. For some complex traits, where sample size is limited – for example, response to anti-TNF therapy – there are no validated SNP associations.

Thus a critical question remains: what fraction of the variance is explained by a polygenic model, where many common alleles of modest effect size contribute to the phenotype?

The answer to this question will (a) provide insight into the genetic architecture of complex traits such as response to anti-TNF therapy (which will in turn guide future genetic studies), and (b) facilitate predictive modeling in a clinical setting. In this application we consider a novel approach: to combine data across

multiple SNPs to improve power – effectively creating an aggregate score for all SNPs. This approach will provide estimates of the amount of variation in response to anti-TNF therapy that can be explained by common alleles of modest effect size.

HYPOTHESIS

Common alleles of modest effect predict the response to anti-TNF therapy.

AIMS

General aim: Conduct a replication GWAS on ~500 Portuguese RA patients treated with anti-TNF therapy selected from the national database from the Portuguese Society of Rheumatology (BioRePortAR)

Specific aims:

a) Generate and analyze GWAS data on ~500 Portuguese RA patients to search for common SNPs that predict response to anti-TNF therapy.

b) Develop a clinical prediction model of response to treatment using aggregate SNP data.

METHODOLOGY

We will include 500 patients from BioRePortAR, the national database for biological therapies developed by the Portuguese Society of Rheumatology.

This database is a structured, codified electronic medical record linked to a database. At least every 3 month patients are observed in a specialized and certified center - the rheumatology day care unit - and information related to disease activity, functional status, adverse events and current treatment are collected and introduced in BioRePortAR.

Blood samples to extract DNA will be collected from eligible patients, processed and stored at -80º in the Biobank of SPR located at IMM.

According to the Portuguese recommendations to start biological therapy in RA patients, the patient should have active disease prior to start an anti-TNF drug (DAS28 >3.2) and should be refractory to methotrexate therapy. In addition, we

will establish strict inclusion criteria to minimize heterogeneity across sample collections. These criteria are:

(1) EULAR “good responder” or “non responder”; the first given by a delta DAS28>1,2 and a current DAS28<3,2; and non-responders showing a delta DAS28<0,6 or <1,2 if the baseline DAS28 was higher than 5,1.

(2) First ever use of an anti-TNF drug; this will remove the bias that may be present among patients who initially responded to treatment, but then become secondary non-responders, and from those that never responded, representing primary failures.

(3) Data on post treatment DAS28 evaluated at 3 months and, if the patient is a responder, the evaluation should be extended up to 12 month; this will assure that accurate response data are available within a consistent time frame to assess primary response and to detect secondary failures.

In particular, we have chosen to focus on those patients that are either ‘good responders’ or ‘non-responders’ according to EULAR criteria, meaning that we will exclude those patients classified as EULAR ‘moderate responders’. We hypothesize that selecting more extreme phenotypes will add statistical power to detect a true result, as (a) genetic factors that influence response will be enriched in the more extreme categories, and (b) accuracy of disease activity measurements is much higher at more extreme phenotypes.

Moreover, good response is important clinically, as a different drug would likely be chosen in patients with a lower probability of having a response.

The samples will be genotyped with the Illumina platform, which contains 660,000 SNPs. The Illumina 660W array captures >85% of known common genetic variation potentially relevant for immune mediate diseases, as assessed using HapMap.

We will test for potential clinical confounders using univariate and multivariate analyses. Any clinical predictor that is significant at $p<0.05$ will be included as a covariate in our regression models.

We will use multivariate logistic regression to assess the significance of a SNP predicting response category, including age, sex, and concurrent MTX as covariates. Other factors such as the type of anti-TNF (etanercept, infliximab or adalimumab), disease duration and pre-treatment DAS will be also evaluated.

We will correct for multiple tests by permutation, and adjust for potential population stratification using EIGENSTRAT.

The GWAS will be conducted under the supervision of Dr Robert Plenge (Harvard Medical School, Boston, USA) who has a large experience in conducting this kind of genetic studies in RA.

Aggregate genetic risk scores (GRS)

The classic theory of polygenic inheritance, described by Fisher in 1918 and supported empirically by recent GWAS results, implicates a large number of loci of very small individual effects that collectively account for a substantial proportion of phenotypic variation. In the early 1900s, geneticists noted that most naturally occurring trait variation, while showing strong correlation among relatives, involves the action of multiple genes and non-genetic factors. Recently, GWAS has provided empirical support for these early observations: very few diseases have common alleles of large effect size (e.g., OR >3), and those alleles reproducibly associated with phenotypic variation have modest effect size (OR 1.1 – 1.3). However, the exact number of loci that contribute to trait variation is not clear, and could represent thousands of independent alleles of even smaller effect (OR~1.05). In order to achieve genome-wide significance after correcting for the number of common alleles in the human genome, thousands of patients are required. Even then, validated SNP associations explain only a small fraction of the genetic contribution to variance of most complex traits. For some complex traits, where sample size is limiting, there are no validated SNP associations.

Thus, how is it possible to test the polygenic model, especially in the absence of very large sample sizes or in the absence of a large number of variants that are genome-wide significant?

A study led by Drs. Karlson and Plenge from Harvard Medical School developed a simple aggregate genetic risk scores to quantify risk of RA (ARD 2010).

We describe this study briefly to provide a context for the conventional approach to aggregate genetic risk scores: to model only those SNPs that are genome-wide significant.

In this study, the authors examined known RA susceptibility alleles, all of which were independent predictors of risk. At the time they initiated the study, there were 14 validated non-MHC SNPs, in addition to 8 MHC alleles. They created a weighted genetic risk score for 22 alleles (wGRS22), where the weight for each risk allele is the log of the published odds ratio. In an ordinal model of the highest vs. lowest risk categories, they observed an OR of RA between 6.24 and 12.31.

Other authors have used aggregate SNP data from GWAS to demonstrate that a polygenic model of inheritance underlies a substantial proportion of risk in patients with schizophrenia and bipolar disease (Nature 2009).

While GWAS data are generally used to test the role of individual variants, it is also possible to use GWAS data to test whether a large number of SNPs of small effect contribute to trait variation, as proposed by Fisher in 1918.

The estimate that common polygenic variation accounts for one-third of the total variation in schizophrenia risk should be thought of as a lower bound for the true value, which could be much higher. In summary, these statistical methods use GWAS data to predict disease risk, even for those SNPs that do not reach a conservative genome-wide level of significance. This approach allows one to estimate the variance explained by common SNPs.

Genotyping will occur in three steps:

(1) DNA preparation. We will determine the exact concentration of the DNA samples using a Picogreen assay. Picogreen measures double stranded DNA and is a much better predictor of genotyping success than other quantitation methods. We will normalize all samples to 50 ng / ul.

(2) DNA fingerprinting. We will assign a unique genetic ‘fingerprint’ to each sample to ensure DNA plate integrity and sample identity throughout: we will use the Sequenom iPLEX platform to genotype a panel of 24 highly polymorphic SNPs that overlaps SNPs on the Illumina array. We will obtain a new aliquot of DNA from any sample that does not pass these QC steps.

(3) GWAS genotyping. Approximately 750 ng of genomic DNA will be used to genotype each sample. Briefly, each sample is whole-genome amplified, fragmented, precipitated and resuspended in appropriate hybridization buffer.

Denatured samples are hybridized on prepared Illumina 660W-quad beadchips for a minimum of 16 hours at 48°C. Following hybridization, the beadchips are processed for the single base extension reaction, signal amplification and imaged on an Illumina Bead Array Reader. Normalized bead intensity data obtained for each sample are loaded into the Illumina Beadstudio 2.0 software, which converts fluorescent intensities into SNP genotypes.

The following steps will be performed by Dr. Eli Stahl (BWH) and Diana Fernandes (IMM), under the guidance of Dr. Plenge: (1) Quality control filters. We have established strict quality control criteria based on empirical data to minimize bias due to genotyping artifact. Within the cohorts, all individuals with >5% missing genotype data and SNPs with >5% missing genotype and/or deviation from Hardy-Weinberg Equilibrium ($p<10-5$) will be excluded from analysis in the GWAS. In addition, we will apply identity-by-descent (IBD) probabilities to capture expected and unrecognized relationships among samples. Examination of these values can identify sample contamination, unexpectedly close relationships between individuals (for example, cousins or siblings), and unanticipated duplicate samples. We will use PLINK as our primary tool to manage our GWAS data.

(2) Population stratification. Having applied strict QC filters, we will next analyze the extent of population stratification by determining the amount of statistical inflation, and correct for population stratification using a principal components method. Despite the majority of samples being self-identified European ancestry in the samples collected to date, some degree of population stratification may still be undetected (including patients with missing information about race). To correct for population stratification, we will run Eigenstrat, remove genetic outliers, and calculate the top 10 eigenvectors in the dataset. The eigenvectors will be used as covariates in our multivariate logistic regression model.

(3) Imputation of all CEU HapMap SNPs. We will use IMPUTE 100 to determine genotype calls and confidence scores for all polymorphic CEU HapMap SNPs. This will facilitate the analysis of common SNPs across the genome not genotyped directly. To impute all CEU HapMap SNPs, we will conduct separate runs for each chromosome using default IMPUTE parameters. All computations will be performed on the Broad Institute Load Sharing Facility (Boston, USA).

(4) GWAS of genotyped and imputed SNPs. Our primary analysis will test whether a SNP is associated with a dichotomous outcome, classifying patients as either “Good Responders” or “Non-Responders”, accordingly to EULAR criteria described above. We will use multivariate logistic regression, where eigenvectors and clinical factors are included as covariates. We will use dummy variables to account for unknown confounders.

For any SNP directly genotyped, the actual genotype count will be used in our analysis. If a SNP is imputed, then we will use probabilistic allele dosages. This approach accounts for some uncertainty in imputation, and avoids potential bias of estimating an exact genotype call per individual.

(5) Interpret statistical significance. A major goal of this analysis is to find a ‘big hit’ – an allele of large effect (OR>1.5) that is associated with severity and/or response to anti-TNF therapy with high statistical certainty (at $p<5x10-8$).

The only GWAS of response to anti-TNF therapy published to date was in fewer than 100 patients. Thus, it is possible that such an allele does indeed exist, as has been shown for the MHC and risk of RA. For any single SNP, we will consider a $p<5x10-8$ as genome-wide significant, given the number of independent tests in the human genome. We have >80% power to detect common variants with odds ratios (OR) >1.5.

We will control for clinical variables that may confound our analysis by including covariates in our analysis. Although clinical trials suggest that treatment response to anti-TNF therapy does not vary according to disease duration, CCP or RF status, or prior DMARD failure, we will be sure to assess these variables.

At the end of this analysis, we will have conducted a GWAS of treatment response for individual SNPs. Based on our sample size, we have power to detect common variants with OR>1.5.

Aggregate genetic risk score analysis of GWAS data

Even if an allele(s) is identified, it is exceedingly unlikely that this will account for most of the variance in treatment response. Therefore, a critical question remains: what fraction of the variance is explained by a polygenic model, where many common alleles of modest effect size contribute to treatment response? To address this critical question with empirical data, we will apply methods recently developed and described above.

We will subset our GWAS data into target and discovery samples, and test the predictive value of SNPs at varying levels of statistical significance. This will occur in 4 steps:

(1) Divide dataset in half (discovery & target): we will randomly select approximately half of the responders and non-responders as the discovery group, and the other half as the target group. This approach is feasible because we have focused on categorical response criteria, rather than using DAS as a continuous trait.

(2) Filter SNPs: We will prune our GWAS data to have a set of high-quality autosomal SNPs that are in linkage equilibrium (LE). We will select autosomal SNPs with a total sample MAF of 2% or greater and a genotyping rate threshold of 99% or greater; we will not use imputed SNPs in this analysis. We will next prune the SNP panel to remove SNPs in strong linkage disequilibrium with other SNPs (based on a pairwise r2 threshold of 0.25, within a 200-SNP sliding window). Focusing these analyses on a subset of SNPs in approximate linkage equilibrium has several advantages and makes interpretation more straightforward, for example, the calibration of results with simulated data and the comparison of results across the frequency spectrum, but most importantly to ensure the score represents the aggregate effect of a large number of independent SNPs. Focusing on autosomal SNPs avoids the issue of how to score haploid and diploid genotypes in males and females without creating artificial mean differences between the sexes. We will evaluate coverage of the pruned dataset based on mean-max r2 in CEU HapMap 181.

(3) Create score in discovery samples: We will conduct a GWAS in the discovery samples for SNPs that predict response to anti-TNF therapy, as described above. Each SNP will be assigned a weight based on the log of the odds ratio. We will develop an aggregate score using 11 overlapping sets of SNPs: all SNPs with PT <0.01, PT < 0.05, PT < 0.1, PT < 0.2, PT < 0.3, PT < 0.4, and PT < 0.5; for some analyses we will also consider sets excluding the most highly associated SNPs, 0.01 < PT < 0.2, 0.05 < PT < 0.2, 0.05 < PT < 0.5 and 0.2 < PT < 0.5. The score is expressed as the mean score per SNP in the set; the number of non-missing genotypes used to calculate each score will also be record per individual, for use as a covariate in subsequent target sample analysis.

(4) Test score in target samples. We will test for association between the score and treatment response in the target samples. The primary target sample test uses a logistic regression of disease state on score. Critical non-clinical covariates include the number of non-missing genotypes of all SNPs used to calculate the score, to control for potential differences in genotyping rate between cases and controls. Study sample is taken into account by inclusion of dummy-coded covariates to represent the strata. We will also consider clinical covariates. Study sample is also accounted for in all discovery sample analyses, by use of a Cochran-Mantel-Haenszel stratified analysis to calculate the common odds ratios. In the target sample logistic regression analysis, we estimate the variance explained in treatment response by the score as the difference in the Nagelkerke pseudo R-squared from a model including the score and covariates versus a model including only the covariates. All tests reported are two-sided.

We will report significant associations relative to the direction in the discovery set, such that a higher score is associated with an increased treatment response rate.

(5) Interpret statistical significance. We will consider a polygenic model with $p<0.004$ as significant, given the 11 overlapping sets of SNPs we will test (conservative Bonferroni correction). If we do not observe this level of significance, then we will perform secondary analyses to consider non-genetic sources of variation: phenotype definition of treatment response (we are analyzing our clinical data as a dichotomous categorical trait); drug-specific effects (three anti-TNF drugs, two general drug classes) and influence of known confounding variables (e.g., concurrent medication). If we cannot identify any obvious nongenetic explanation, then we will conclude that common SNPs do not contribute substantially to response to anti-TNF therapy in patients with RA.

(6) Replication. We will replicate our findings, as these studies will also be done in USA patients applying the same methodology.

Based on data from schizophrenia and simulations in 600 responders and 600 nonresponders we calculate that we have >95% power to detect a polygenic association at $p<0.004$ if common SNPs contribute at least 50% to the variance in treatment response. We also note that we have ~10% power at this same p-

value threshold if common SNPs explain 10% of variance and 50% power to explain 30% of the variance.

Upon completion with the American patients we will have >80% power at $p<0.004$ to detect a signal if common SNPs explain 20% of variance. Thus, if we fail to detect a polygenic signal, then this suggests either that common SNPs do not contribute substantially to variation in treatment response, or that our method of defining treatment response (e.g., responders vs. non-responders) is too noisy. In this context, a negative result would provide great insight into genetics of anti-TNF therapy.

To gain insight into the percent variance explained, we will perform simulations based on our observed data. The steps include:

(1) simulate discovery and target datasets that are comparable to those used in this study, under a variety of genetic models;

(2) repeat the score analyses across a range of PT thresholds for each pair of simulated discovery/target datasets, in order to identify models that produce profiles of results similar to the real data in terms of variance explained by the score, R2; and

(3) calculate the implied variance explained by the subset of true risk alleles from the selected model.

Considering the number of variants and their average effect size, there will be many different combinations, all other things being equal, which lead to the same variance explained. For example, 5% of the additive genetic variance could be accounted for by: 5 loci that each explain, on average, 1% of the variance; 50 loci that each explain, on average, 0.1% of the variance; 500 loci that each explain, on average, 0.01% of the variance.

To assess the robustness of our score under different clinical conditions, we will define and test different discovery and target groups. As a secondary analysis, we will divide our dataset according to class of anti-TNF therapy (infliximab/adalimumab) vs. the other (etanercept), and ask how well a score defined in one class predicts response in the other, and vice versa. Also, we will divide our collections in other ways: new-onset and long-standing disease; patients collected in Portugal vs. those collected in the United States and male / female gender.

ETHICAL ISSUES AND FUNDING

The project will be developed at a National level, involving all the Portuguese rheumatology centers which agree to participate.

BioRePortAR will be the clinical tool used.

Part of the laboratory and bioinformatics work will be developed at Harvard Medical School, Boston, USA.

The patients will be included after giving a written informed consent for all the study procedures.

The study will be submitted to local ethics committee.

The BioRePortAR was previously approved by Comissão Nacional de Protecção de Dados and by local ethics committees.

The fund for the project will be obtained from the senior clinical award from Harvard Medical School - Portugal Program 2010-2014.

TIMETABLE AND GENERAL TASK DESCRIPTION

The total duration will be 4 years.

Some tasks will be performed at local level and others at National level. Specific experimental work will be undertaken at Harvard Medical School, Boston, USA.

Every center will recruit at least 50 RA patients. The patients will be registered in BioRePortAR in the beginning of the first anti-TNF treatment, in accordance to the routine clinical practice.

A blood sample (whole blood and serum) will be drawn and sent to Instituto de Medicina Molecular for processing and assessment. The DNA extraction will be done at IMM and will be stored at -80ºC allocated to the SPR's national Biobank.

The follow-up will be done in accordance to the national guidelines. None of the study procedures would modify the management usually performed to RA patients in daily clinical practice. No additional clinical procedures will be made as all the evaluations are routinely performed by the Rheumatology Centers as they registry the information in BioRePortAR.

At 3 months a patient will be classified as non-responder, moderate responder or good-responder. If a moderate response occurs, the 6 months evaluation in BioRePortAR will be reviewed to capture late good-responders. For good-

responders all the information on disease activity captured in BioRePortAR will be assessed during the following 12 months to ensure a sustained response.
The patients' recruitment and clinical assessment will take ~~2,5 years-3 years.
The laboratory work will be done during ~6 month.
Statistical analysis and manuscript preparation will be performed in the last 6 month of the project.