This is an author produced version of Genetic data: The new challenge of personalized medicine, insights for rheumatoid arthritis patients.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/98153/

Article:
Goulielmos, GN, Zervou, MI, Myrthianou, E et al. (3 more authors) (2016) Genetic data: The new challenge of personalized medicine, insights for rheumatoid arthritis patients. Gene, 583 (2). pp. 90-101. ISSN 0378-1119

https://doi.org/10.1016/j.gene.2016.02.004

(c) 2016, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International http://creativecommons.org/licenses/by-nc-nd/4.0/
Genetic data: the new challenge of personalized medicine, insights for Rheumatoid Arthritis patients

George N. Goulielmos¹*, Maria I. Zervou¹, Effie Myrthianou¹, Agata Burska², Timothy B. Niewold³ and Frederique Ponchel²

¹Laboratory of Molecular Medicine and Human Genetics, Department of Internal Medicine, Medical School of Crete, Heraklion, Greece

²Leeds Institute of Rheumatic and Musculoskeletal Medicine, The University of Leeds, Leeds, UK

³Division of Rheumatology, Department of Immunology, Mayo Clinic, Rochester, MN, USA

Corresponding author:
George N. Goulielmos, (e-mail: goulielmos@med.uoc.gr);
Molecular Medicine and Human Genetics Section,
Department of Medicine,
Voutes, 715 00 Heraklion,
Crete, Greece.

Key words:

Abbreviated title:
Abstract

Rapid advances in genotyping technology, analytical methods, and the establishment of large cohorts for population genetic studies have resulted in a large new body of information about the genetic basis of human rheumatoid arthritis (RA). Improved understanding of the root pathogenesis of the disease holds the promise of improved diagnostic and prognostic tools based upon this information. In this review, we summarize the nature of new genetic findings in human RA, including susceptibility loci and gene-gene and gene-environment interactions, as well as genetic loci associated with sub-groups of patients and those associated with response to therapy. Possible uses of these data are discussed, such as prediction of disease risk as well as personalized therapy and prediction of therapeutic response and risk of adverse events. While these applications are largely not refined to the point of clinical utility in RA, it seems likely that multi-parameter datasets including genetic, clinical, and biomarker data will be employed in the future care of RA patients.
Contents

1. Introduction
2. The rapidly changing technologies enable new approaches
   2.1. Genome Wide Association Study (GWAS)
   2.2. Allelic discrimination by Taqman - Real Time PCR (qPCR)
   2.3. Sequencing - Next generation sequencing
   2.4. Exome sequencing
   2.5. eQTLs
   2.6. Epigenetics and EWAS (Epigenome Wide Association Studies)
3. Genetics of rheumatoid arthritis
   3.1. Genetic risk factors
   3.2. The role of genetic factors in the context of ACPA+ and ACPA-
   3.3. Gene-to-gene interactions
   3.4. Gene-to-environment interactions
4. Pharmacogenetics vs. Pharmacogenomics: The promise of personalized medicine
   4.1. Pharmacogenomics in RA
   4.2. The personalized medicine achievements
5. Looking towards the future
Conflict of interest statement
Acknowledgments
References
1. Introduction

The previous era of 'blockbuster' drugs is now giving way to an era of stratified medicine in most diseases, with the ultimate goal of delivering the right drug to the right patient at the right time, a task that represents a key objective of the modern translational medicine. To date, only few studies have demonstrated the impact of stratified medicine interventions at population level, probably because of the high costs of studies using proper stratification and tailored interventions directly compared to universal interventions.

Taking Rheumatoid Arthritis (RA) as an archetype disease, it can be treated with a broad variety of immunotherapeutic agents, and even though outcomes have improved dramatically recently with targeted therapies, few patients are cured and we are still from having sufficient knowledge to allow preventive measures in “at risk” individuals. In RA, further complexity is brought with duration of disease and history of responses to current therapies. Heterogeneity between patients is in fact one of the major features of RA, also reflected in its wide range of responses / non-response to therapeutic agents.

The past decade has seen astonishing progress in our ability to decipher genetic and molecular reasons behind diseases as complex as cancers, obesity, neurodegenerative disorders and autoimmune diseases. Equally exciting, the advances in stem cell manipulation, cellular reprogramming, tissue engineering, and genome editing have offered possible therapeutic solutions to conditions previously considered untreatable. Identifying genetic variants of clinical significance remains however very complex and far from perfect technically but such progress has been made over the past few years that we are now approaching clinical utility.

There are different approaches using detailed knowledge of human genetic variation to tailor treatments to patients. One main approach is to tailor drug treatments based upon genetic variations that may affect the metabolism of the drug itself (pharmacogenomics). Another aims to choose a drug according to the best chances of response to that particular agent (Personalized Medicine). However, it should be pointed that this “personalized medicine” approach does not yet allow treatment to be tailored to the needs of each individual but rather allows patients to be stratified into groups with a better chance to respond to a particular drug before treatment is started.

In this review, we will use the example of RA to discuss many factors shaping the future use of genetic information in personalized medicine, ranging from the discovery of RA-associated variants to the resulting insights into disease biology and the potential for clinical applications of such findings.

2. The rapidly changing technologies enable new approaches

The history of modern human genetics research is largely the history of the rapidly changing technologies. Although the vast size of the human genome appeared at first impossible to work with, the development of new analytical approaches has allowed us to reach the point where individual genome sequencing is a feasible task for many laboratories (Kere 2010). The sequencing of the human genome represents a critical
milestone in the scientific landscape and a springboard for genetic studies (International Human Genome Sequencing Consortium, 2001).

The most crucial argument for seeking genetic data is the high heritability of many important medical conditions. The extraordinary technical advances in the field of human molecular genetics over the past few years have led to an explosion of new information about the genetics of complex, multigenic human diseases, notably including autoimmune disorders. The design of genetic studies has relied on three main components: availability of population (with well-defined ethnic/racial groups for original and replication studies), technology (with comprehensive coverage of the entire genome) and data (as comprehensive as possible) to be associated with the outcome of interest. Advances in DNA sequencing and genotyping technology have put us in a unique position to consider issues raised by the use of genetics data in personalized medicine.

2.1. Genome-wide association studies (GWAS)

During the last decade, many breakthroughs have contributed to the unraveling of the genetic etiology and pathophysiology of complex autoimmune diseases and other genetic disorders. GWAS have uncovered thousands of variants involved in the pathogenesis of many complex human disorders and have proven a powerful hypothesis-free method to identify common disease-associated variants. This method identifies SNPs that are present in the general population, but it cannot predict functional consequences. Such genetic associations identify a region or genetic locus that is associated with disease, but the specific causative mutation within that locus is not immediately identified.

Despite the success of GWAS, a substantial heritability gap remains. GWAS have identified a high number of loci common to several autoimmune diseases, such variants despite being common have a modest-effect size and thus a substantial fraction of heritability remains unexplained and/or hidden (Manolio et al., 2009). Some of this missing heritability should be accounted for by low-frequency and rare variants, which would be expected to have large biological consequences (Zeggini 2011). Rare variants can be discovered by re-sequencing a small sample size cohort and then genotype the discovered variants in a larger sample set (Rivas et al., 2011; Momozawa et al., 2011). The study of these rare variants can be strengthened by focusing on isolated or well-defined populations, where an appropriate combination of data from whole genome sequencing and GWAS as well as imputation of variants into a reasonable study may lead to the detection of susceptibility loci for complex diseases (1000 Genomes Project Consortium et al., 2010; Holm et al., 2011). GWAS have not yet reached their limits and although the majority of GWAS that have discovered common variants for human diseases were performed using a case-control design, an interest has been expressed in using family-based designs for GWAS. The reason for this new tendency caused by the expansion of new generation sequencing (NGS) methodology, which outlined the importance of the rare variants in disease susceptibility (Ionita-Laza and Ottman, 2011). Also, the use of GWAS to study patient subgroups, comparing patients with different subtypes of a disease has been fruitful recently (Kariuki 2015).

2.2. Allelic discrimination by Taqman Real Time PCR
Real-time PCR is a very popular procedure for the quantification of gene expression (Ponchel et al., 2003). This technique monitors the progress of a PCR reaction as it develops to quantify a relatively small amount of initial sequence (DNA, cDNA or RNA). Quantification is based on the detection of the fluorescence produced by a reporter molecule, which increases as the reaction proceeds. These fluorescent reporter molecules include dyes that bind to the double-stranded DNA (i.e. SYBR® Green) or sequence specific probes (i.e. Molecular Beacons or TaqMan® Probes) (Bustin 2000). An important application of the TaqMan PCR is the detection of known gene mutations or polymorphisms. This is based on the design of two TaqMan probes, each specific for one allele (A or B) representing the 2 alternative sequences. Both probes are labeled with two different fluorescent tags. The TaqMan probe is then designed to bind the gene sequence flanking the mutation (Livak et al., 1995). TaqMan probe-based assays are widely used in research and medical laboratories for the purposes of SNP genotyping and Pharmacogenomics.

2.3. Sequencing - Next Generation Sequencing (NGS)

DNA sequencing and genotyping technologies have advanced quickly over the past decade, with the development of novel methodologies and machinery and the utilization of new types of molecular tools. The first DNA sequencing technologies were the product of academic research in the 1970s, with the near simultaneous development of two methods based on the electrophoretic separation of a nested set of labeled DNA products. The first one was developed by Allan Maxam and Walter Gilbert while the second one, developed by Frederick Sanger, is the basis of today’s high-throughput DNA sequencing. The second major advance in DNA sequencing in the 1980s was the development of automated DNA sequencing instruments, introducing base-specific fluorescent dyes and capillary electrophoresis into the process. More recently “next-generation DNA sequencing (NGS) technologies” are summarized in Table 1.

The use of NGS technologies has assisted in the identification of common and rare variants determining the response to drugs or serious adverse drug reactions (Wadelius et al., 2009; Carr et al., 2014). Personal genome sequencing is becoming a reality, and targeting whole exome sequencing is being explored to facilitate diagnosis and guide treatment in some conditions, mainly regarding monogenic diseases such as primary congenital glaucoma, Wolfram syndrome, and many others (McLaughlin et al 2014). Recently, guidelines were developed representing initial steps to ensure that results from tests based on NGS are reliable and useful for clinical decision making [http://www.cdc.gov/osels/lspppo/Genetic Testing Quality Practices/Nex-StoCT.html].

2.4. Exome Sequencing

Sequencing the coding regions, the exome of the human genome, is one of the major current strategies to identify rare, low frequency variants as well as de novo mutations that are associated with human diseases. Sequencing the entire protein-coding sequence is one of the major applications of NGS. The GENCODE consortium (Coffey et al., 2011) notably designed an extended set of primers to capture the complete human genome, covering an additional fragment of 10.3 Mb harboring 5594 more genes compared with the widely used commercial kit from the consensus coding sequence consortium (CCDS) database-base (Pruitt et al., 2009). These additional
regions include potential disease genes, which were previously inaccessible, thus resulting in the identification of over 5000 new SNP variants. Exome strategies are believed to be more potent towards the identification of Mendelian disease genes, and this method is a commonly used tool to identify causal variants in these diseases (Gilissen et al., 2013).

2.5. Expression quantitative trait loci (eQTLs)

Gene expression is differentially regulated across tissues, and many human transcripts are expressed in a limited set of cell types or during particular stages of a disease. The abundance of a gene transcript can be directly modified by polymorphisms in regulatory elements. Consequently, transcript abundance might be considered as a quantitative trait that can be mapped with considerable power. Expression quantitative trait loci (eQTLs) are genomic loci that regulate expression levels of mRNAs or proteins (Consoli et al., 2002). Importantly, expression traits differ from most other classical complex profile in one important respect—the measured mRNA or protein is the product of a single gene with a specific chromosomal location. Studies have shown that SNPs reproducibly associated with complex disorders (Nicolae et al., 2010) as well as certain pharmacologic phenotypes (Gamazon et al., 2010) are significantly enriched for eQTLs, and high-throughput measurement of genome-wide transcription using microarray technology has greatly facilitated the systematic mapping of eQTLs. Considering that genes mediate their biological roles in groups rather than in isolation, an interesting challenge will be the extension of the concept of eQTLs to whole genome pathway analysis.

2.6. Epigenetics and Epigenome-Wide Association Studies (EWAS)

DNA sequences carry genetic information within a dynamic complex of chromatin and regulatory sequences, such as transcription factors binding sites, which control the readout of the DNA sequence (Kilpinen and Dermitzakis 2012). A higher hierarchy also exists at the chromatin levels, with an overall backbone defined as the epigenome, consisting of nucleosomes that are made up of wrapped DNA around histone proteins. The functional importance of epigenetic changes lies in their ability to regulate gene expression. Recent technological advances have allowed the detection of a plethora of histone modifications and related chromatin structures (methylation, acetylation, and ubiquitination) (Zhou et al., 2011). DNA methylation markers may be useful as diagnostic and prognostic markers in many human diseases (Rakyan et al., 2011). Thus, the genome gives information about sequence and structure, while the epigenome provides information regarding genome function (Verma 2012).

Epigenetic mechanisms contribute to many fundamental biological processes and to the pathogenesis of complex human diseases. For example, epigenetics may help to explain individual variation in drug responses, thus representing important potential in personalised medicine (Feinberg 2010). Epigenome-wide association studies (EWAS) provide an opportunity to identify genome-wide epigenetic variants that are associated with disease. EWAS offer a potential to increase our understanding of complex disease, but also pose some challenges that are not encountered in GWAS, such as the fact that epigenetic changes can vary over time in relation to disease course and medication use. Interestingly, Rakyan et al (2011) suggested a way that an integration
of EWAS with GWAS could help to dissect complex GWAS haplotypes for further functional analysis.

3. Genetics of Rheumatoid Arthritis

3.1. Genetic risk factors

Rheumatoid Arthritis is a chronic, progressive systemic inflammatory disease, affecting approximately 0.5–1% of the population (Alamanos and Drosos 2005), that if uncontrolled may lead to significant joint damage, pain, dysfunction, work disability that results in large economic losses. Although genetic influences on susceptibility to and clinical course of RA have been known for a long time, only recently have they been approached systematically through the advances in genotyping technology and GWAS. Findings from elegant twin studies (Silman et al., 1993; MacGregor et al., 2000) have estimated the heritability of RA at about 50-60%, thus suggesting the remarkable influence of genetic factors on disease susceptibility, while the remaining part can be attributed to environmental factors or chance.

Before the GWAS era, RA susceptibility loci were detected using the traditional candidate-gene approach and linkage analysis, discovering the two most significant RA risk loci, HLA-DRB1 that exhibits the largest genetic contribution to RA susceptibility, and the group of alleles referred to as the shared epitope (SE) (Welcome Trust Case–Control Consortium 2007). The PTPN22 (Bottini et al., 2004), PAD14 (Suzuki et al., 2003) and CTLA-4 genes (Plenge et al., 2005) were also identified as risk factors in the pre-GWAS era prior to 2007. In the next few years, many more RA-susceptibility genes were discovered using GWAS study designs, which thus far have identified over 110 susceptibility loci for RA in European and Asian populations (Eyre et al., 2012; Orozco et al., 2014). Somewhat surprisingly, these susceptibility loci account for less than 50% of the total genetic heritability (Okada et al., 2014), suggesting that there is a significant portion of the genetic architecture of RA that remains to be discovered. Many of these loci associated with risk for RA are lying in one main biological pathway related to T-cell activation and signaling (HLA-DRB1, PTPN22, PTPN2, STAT4, CD40, CTLA4, IL2, IL21, IL-6R, GATA3, CCR6, IL-2R, IL-7R, CD28) (Plenge et al., 2007; Martinez et al., 2008; Raychaudhuri et al., 2008; Barton et al., 2009; Orozco et al., 2009; Cobb et al., 2013).

Following the GWAS era, an important development was the use of a novel genotyping array called Immunochip. The Immunochip was a custom Illumina Infinium genotyping array designed by the Immunochip Consortium to include risk loci confirmed in 12 major autoimmune diseases (Cortes and Brown 2011), mapping in total 196,524 SNPs of which ~3,000 were specific for each disease and including dense coverage of the extended HLA region and 186 non-HLA risk loci. Considering that a remarkable number of loci associated with RA susceptibility were also associated with multiple other autoimmune diseases (Hinks et al., 2005; Behrens et al., 2008; Zervou et al., 2008; Zhernakova et al., 2009; Kurreeman et al., 2010), it was thought that the mapping of loci discovered in other diseases might establish the same loci as RA-risk loci. Additionally, the fine-mapping nature of the chip would be important to understand how individual variants contribute to risk of disease. Use of the platform in RA allowed the detection of novel RA-associated loci such as BACH2 and RAD51B.
The small odds ratios for the majority of these risk factors make these findings of limited importance with regard to their potential use in the prediction of the risk for the disease for individual patient. The current challenge is to translate this genetic information into concrete mechanism of how these variants interact with each other or with environmental factors in order to lead to RA (Gregersen and Olsson 2009).

Joint destruction is an accurate measure for RA severity and reflective of the cumulative disease burden. Therefore, risk factors for this outcome measure may be used to arrive at individualized treatment strategies (van der Helm-van Mil et al., 2010). It has been shown that genetic variants of IL-15 (Knevel et al 2012), IL4R (Krabben et al., 2013), IL2RA (Knevel et al., 2013a), GZMB (Knevel et al., 2013b), Dkk-1 (de Rooy et al., 2013), MMP-9 (de Rooy et al., 2014), MMP-3 (Knevel et al., 2014a), OPG (Knevel et al., 2014b), PTGER4 (Rodriguez-Rodriguez et al., 2015), TRAF1/C5 (van Steenbergen et al., 2015) as well as the rs2833522 SNP (De Rooy et al., 2013) are associated with RA severity as reflected by radiological damage and joint destruction appeared in RA patients. Moreover, it has been reported that RA patients with the Blimp-1 risk allele show more synovial inflammation than those without this allele (Hereniush et al., 2011).

### 3.2. Role of genetic factors in the context of ACPA+ and ACPA -

During recent years, it has become clear that RA is composed of several phenotypes with defined and different genetic and environmental risk factors. One of such major phenotyping criteria is the presence/absence of anti–citrullinated peptide antibodies (ACPAs) despite continuing debate as to whether ACPA positive and negative RA represent two different subsets of the same disease or are two distinct diseases altogether (van der Helm-van Mil and Huizinga 2008; Daha and Toes 2011). Indeed, there are clearly significant differences in risk allele frequency that are mainly associated with loci in the HLA region supporting the distinct genetic etiologies hypothesis for ACPA+ and ACPA- RA (Padyukov et al., 2011). The presence of these antibodies is associated with HLA–DRB1 shared epitope (SE) alleles (Klareskog et al., 2006). ACPA positivity has been associated with more progressive, more destructive RA and extraarticular manifestations as well as cardiovascular disease (Visser et al., 2010). Importantly, it was shown that a non-SE HLA–DRB1 allele (HLA–DRB1*15) is associated with production of high levels of ACPAs in RA patients, suggesting a role for this genetic background in adaptive immunity (Laki et al., 2012).

Numerous genetic studies have now been conducted in RA patients stratified as either ACPA positive or negative. Genetic associations appear stronger in general in the ACPA+ subgroup (Huizinga et al., 2005; de Vries et al., 2011). Some loci, such as CD2/CD58, AFF3, CD28, CTLA-4, PXK, RBPI, TAGAP, CCR6, IRF5, TRAF1/C5, CD40, HLA–DRB1*15, HLA–DRB1*03, BACH2, RAD51B, AFF3 locus 1, CCR6, CCL21 locus 2, IL2RA locus 2, CD28, CD40, PXK, REL, RBPI, TNFRSF14, TNFAIP3 locus 3, HLA-DRB1*01, HLA-DRB1*04, HLA-DRB1*15, associated with ACPA+ RA were not associated with ACPA- disease (Verpoort et al., 2005; Thabet et al., 2009; Padyukov et al., 2010; Kurreeman et al., 2011; Daha et al., 2012; Laki et al., 2012; Kurko et al., 2013; McAllister et al., 2013; de Rooy et al., 2015) while others, such as TNFAIP3, STAT4, BLK, PTPN22, ANKRD55/IL6ST and GIN1/C5orf30, were found to be associated with susceptibility for RA irrespective of the serological status of the patients (Pierer et al., 2006; Viatte et al., 2012). In addition, IRF5, HLA-
DRB1*03 and HLA-DRB1*13 were found to be associated with susceptibility for ACPA-RA only (Padyukov et al., 2010, Kurko et al., 2013). Together, the number and identity of the risk loci associated with the two disease serotypes are different, thus strengthening the concept that they might be two genetically different diseases (Viatte et al 2012).

ACPAs are considered to be strong predictors of disease activity and radiological damage (Bukhari et al., 2007). Positivity for ACPAs is considered to be an important biomarker for the classification of RA patients in two subgroups (Klæreskog et al 2009), considering that ACPA positivity is associated with worse response to anti–tumor necrosis factor therapy (Potter et al., 2009). However, despite genetic differences and/or similarities established between ACPA positive and negative RA and serotypes showing significant differences in disease severity and course, ACPAs are not currently used to guide treatment decision in clinical practice (Viatte et al 2012).

### 3.3. Gene-Gene interactions

Gene x Gene interaction (GxG) refers to the situation in which the effect of one genotype is conditional on genotypes at one or more other unlinked loci. Interactions occur when the effect of one factor upon phenotype (i.e. disease), is modulated by other factors. The current debate over the source of genetic variation in disease may therefore move from ‘is it caused by (rare or common) variants?’ or even from ‘how much does each variant contribute?’ to rather ‘how do they work together?’ (Schork et al., 2009). The models above which calculate the amount of heritability accounted for by the known genetic variants in RA generally do not account for GxG interactions.

Multiple gene interactions have been reported in RA so far but only the HLA-DRB1 (SE) × PTPN22 interaction has been replicated convincingly, which leads to the development of ACPA+ RA. HLA-DRB1 and PTPN22 genetic polymorphisms were previously shown to be associated with an increased risk for ACPA+ but not ACPA-RA. Thus, a study conducted by Kallberg et al. (2007) showed that in RA, the R620W polymorphism of PTPN22, which alone provides a moderate effect, can exert much more significant effects in combination with the HLA-DRB1 SE alleles that account for approximately one-third of the genetically determined susceptibility to the disease (Wordsworth et al., 1992). The fact that the PTPN22 risk (T) allele appears to contribute a risk mainly when HLA-DRB1 SE alleles are present, provides strong evidence that MHC class II–dependent T-cell activation is of central pathogenetic importance for the subset of RA characterized by presence of ACPA antibodies (Kallberg et al., 2007).

Another interaction was shown between SNPs in the HLA region and a SNP on chromosome 1 (rs2476601) (González-Recio 2009), while pairwise epistatic interactions were detected between PTPN22 and seven SNPs corresponding to CEP72, MYO3A and CDH13 genes (Briggs et al., 2010). In another study, a deletion polymorphism in glutathione S-transferase Mu-1 (GSTM1-null) was found to exert significant additive interaction with SE on the risk of ACPA positivity in RA (Mikuls et al., 2010). A genome-wide epistasis analysis of SNP x SNP combinations identified interacting SNP pairs between the SULF1 and 2 SNPs from the AKAP1 gene (Julià et al., 2007). In another interaction was shown a high-degree of synergy between IL6
and IL4I-1 promoter polymorphisms, leading to the aggressive phenotype of synovial fibroblast in the development of RA (Julia et al., 2007). Finally, it was recently shown that a CD86 and a CD40LG genotype were interactively related with a significant risk of developing RA (Lee et al., 2014).

### 3.4. Gene-to-Environment interactions

Many factors such as diet, exposure to toxins and microbes, and socioeconomic status all profoundly influence health. This information is especially important in understanding susceptibility to disease. Gene-to-Environment (G×E) interactions refer to the situation in which the effect of the genotype is conditional depending on the environment. It has been long suspected that genetic and environmental factors work together to cause RA, and evidence in support of these interactions suggests that there may be many pathways to RA.

Many environmental factors have been associated with RA but only smoking has been proved to interact with certain genes in RA. A large body of evidence shows the role for cigarette smoking in causing RA, although this is now clearly related to the presence of a specific genetic background. The first studies conducted in GxE interaction in RA showed that the disease risk of Rheumatoid Factor (RF)-seropositive RA associated with SE was strongly influenced by cigarette smoking (Symmons et al., 1997; Karlson et al., 1999). The effect of smoking on increasing the risk of RA with respect to intensity and duration or cessation has been studied and results have indicated that the risk of developing RA increases in a dose-response manner with the intensity (Heliovaara et al. 1993) and the duration of smoking (Stolt et al., 2003; Costenbader et al., 2006). A recent prospective study highlighted that even light cigarette smoking is associated with increased risk of RA in women and that smoking cessation may reduce, though not remove, this risk (Di Giuseppe et al., 2013). Furthermore, excessive caffeine consumption, oral contraceptive use and having a first-degree relative with schizophrenia were found to confer a higher risk to develop ACPA+ RA, while moderate alcohol consumption has been suggested that might protect against RA (Pedersen et al., 2006). Alcohol consumption was shown to reduce the risk of developing ACPA+ RA in smokers carrying HLA-DRB1 SE alleles (Källberg et al. 2009). A meta-analysis indicated that both smoking and the PTPN22 risk allele ("T") are associated with the risk of ACPA positivity (Taylor et al., 2013). Finally, in a study specifically undertaken to examine the impact of smoking and genetic on RA susceptibility, a significant GxE interaction was detected between the GSTT1 gene and heavy smoking (Keenan et al., 2010). Notably, more recent findings showed that smoking and the interaction between smoking and SE is associated with the concurrent presence of all three RA-associated autoantibodies, that is rheumatoid factor (RF), ACPA and anti-carbamylated protein antibodies (CarP) (Gan et al., 2015).

Several microorganisms have been implicated in the development of RA. One hypothesis is that these microorganisms trigger the development of RA in individuals who carry genetic susceptibility factors to the disease (Tobon et al., 2010); however, the role for microorganisms as initiating factors of RA remains controversial. There are findings supporting the hypothesis that oral infections play a role in RA pathogenesis, while of special importance are the impact of periodontal pathogens, such as Porphyromonas gingivalis on citrullination, and the association of
Periodontitis (PD) in RA patients with seropositivity toward RF and ACPA (Detert et al., 2010).

An association has been suggested in human between RA and exposure to mineral oil (Kleinau et al., 1991; Cannon et al., 1993), although it has not yet been possible to determine to what extent this risk is restricted to individuals with a certain genetic constitution. However, a family study conducted in the UK showed an increased risk of RA in individuals carrying a certain allele of a gene on human chromosome 17, which is syntenic to the genetic region that determines susceptibility to oil-induced arthritis in rats (Klareskog et al. 2002). Another study showed that exposure to mineral oil is associated with an increased risk to develop RF+ and ACPA+ RA, respectively (Sverdrup et al., 2005). Mineral oils used in the aforementioned study varied between motor oil, hydraulic oil, form oil, or asphalt.

4. Pharmacogenetics and Pharmacogenomics: The promise of personalized medicine

Humans vary in their responses to drugs, and genetic factors are hypothesized to account for 20-95% of this variation (Harper and Topol 2012). Pharmacogenetics refers to inherited differences in the ways patients process drugs, while pharmacogenomics encompasses the overall genetic information that can help predict drug response. In RA, pharmacogenomics has been used to identify SNPs or multiple genetic signatures that may be associated with responses to traditional disease-modifying drugs (DMARDs) and biologics. These genetic markers could be useful in daily practice because they do not vary with time, and analysis can be carried out using samples derived from patient’s blood.

In the past decade, advances in pharmacogenetics and pharmacogenomics have gradually unveiled the genetic basis of inter-individual differences in drug responses (Evans and Relling 1999; 2004). Considering that genes code for protein targeted by drugs or enzymes that regulate drug metabolism, the use of pharmacogenomics predictions in combination with the development and improvement of genetic/diagnostic tests would allow the stratification of patients into genetic subgroups, allowing the prescription of drugs only to subjects that are predicted to have high drug efficacy without any undesired toxicity or adverse events. Relatively few GWAS investigating drug response have been published. Nowadays, the application of pharmacogenomic tests is still limited to a relatively small number of diseases. The use of a recommended genetic test prior to prescribing a drug can vary widely across disciplines but the most immediate applications of pharmacogenomics are in the field of cancer where drugs are already targeted to small groups of patients and the level of clinical specialization is high. Thus, it is well known the role of CYP2D6 in the prediction of response to tamoxifen (Goetz et al 2008) or the association of polymorphisms in the metabolic enzymes P450 (CYP) with drug response, toxicity and cancer risk (Bozina et al 2009).

4.1. Pharmacogenomics in RA

In addition to susceptibility to RA, some genes may have relevance for disease outcome and prognosis (Davila and Ranganathan 2011; Plant et al., 2011). To date, biologic agents for RA include TNF blockers (infliximab, etanercept, adalimumab,
certolizumab, and golimumab), the B cell-depleting anti-CD20 antibody (rituximab),
the cytotoxic T-lymphocyte antigen-4 fusion protein (abatacept) that inhibits co-
stimulation, the IL-1 receptor antagonist (anakinra), the IL-6 receptor inhibitor
(tocilizumab) (Atzeni et al., 2012) and the Janus-associated kinase (Jak) inhibitor
( tofacitinib), which blocks the cytokine pathway that leads to the activation of
lymphocytes (Lee et al., 2014). Although existing recommendations do not
distinguish between ACPA-positive and ACPA-negative RA with regards to the
choice of biological therapy, there are convincing data to suggest superior efficacy of
agents such as anti-CD20 mAb and CTLA4.Ig in ACPA-positive patients. SNPs in
several genes have been associated with altered efficacy or toxicity of methotrexate
(MTX), such as those in genes such as MTHFR, SLC19A1, SHMT1, ABCB1, IL1RNα
as well as ATIC and TYMS (Berkun et al., 2004; Dervieux et al., 2004; Tolusso et al.,
2006; Kurko et al., 2013, Salazar et al 2014). Moreover, an association was detected
between efficacy of MTX and a single SNP within the GGH gene while 5 SNPs
mapped to DHFR and 2 mapped to FPGS were significantly associated with adverse
events (Owen et al., 2012). An increased toxicity of azathioprine has been associated
with SNPs of TMPT and ITPA genes (Stolk et al., 1998). Additionally, it was found
that mutations in the NAT-2 gene resulted in adverse events in patients with RA taking
sulfasalazine (Tanaka et al., 2002). SNPs in the DHODH, ESR1 or CYP1A2 genes
have been associated with increased efficacy of leflunomide, an isoxazole derivative
used as a DMARD for the treatment of RA (Powlik et al., 2004; Dziedziejko et al.,
2010).

Treatment strategies blocking tumor necrosis factor (anti-TNF) have proven very
successful in patients with RA, showing beneficial effects in approximately 50-60%
of the patients only, for reasons that are still unknown. Many studies have been
conducted recently based either on single SNPs or on more extensive
pharmacogenomic investigations, aiming to analyze the efficacy of the biological
agents used (mostly etanercept, infliximab, adalimumab). Such pharmacogenetic
studies have unraveled genes involved in various signaling pathways that regulate key
immune and inflammatory processes. Studies that were conducted following a
candidate-gene approach confirmed a role for several know susceptibility genes
including HLA-DRB1, PTPRC, PTPN2, AFF3 and CD226 (Criswell et al., 2004;
Potter et al., 2009; Tan et al., 2009; Cui et al., 2010; Plant et al., 2012). Using GWAS,
more gene loci have been identified (including PDE3A–SLCO1C1, NUBPL, CNTN5,
VAV1, SPRED2, PDZD2, EYA4, CENTD1, MAFB2, QKI, LASS6, IFNK, CST5,
GBP6, PON1, LMO4 and MOBKL2B) (Liu et al., 2008; Plant et al., 2011; Acosta-
Colman et al., 2013; Mirkov et al., 2013). Data of many of these studies are however,
still highly inconclusive given that different publications have observed different
associations. Some of these differences may be attributed to: differences in ethnic
populations, demographics and disease stages, previous drug history, definition of
response, and size of cohorts (Adarichev et al 2005; Davila and Ranganathan 2011;
Zervou et al., 2013;Burska et al., 2014).

A pharmacogenomic study in a small number of RA patients treated with tocilizumab
showed differential expression levels for 4 genes, namely DHFR, CCDC32, EPHA
and TRAV8 between baseline and after 4 weeks of treatment (Kurko et al., 2013) but
the information about the role of these genes in RA is still poorly understood.

Some TGFβ1 SNPs are associated with a good response to rituximab therapy, and as
such could be useful genetic biomarkers in predicting therapy outcome (Daien et al.,
Furthermore, it has been suggested that BAFF gene 871C>T promoter polymorphism may be useful in prediction of treatment outcome in RA patients treated with rituximab after failure or intolerance to TNF blockers (Ruyssen-Witrand et al., 2012a), while genotyping for 158V/F polymorphism of FcGRIIIa may be helpful in prediction of rituximab treatment outcome as well (Kastbom et al., 2012; Ruyssen-Witrand et al., 2012b) (for details regarding all pharmacogenetics information see Table 2). Thus far, none of the above markers has been validated sufficiently to allow for use in the clinical prediction of response to rituximab.

4.2. Personalized medicine achievements

The success of personalized medicine is critically dependent on performance accuracy, given that the reliability and specificity of any laboratory test limits the degree to which it can be used to personalize a treatment. Furthermore, accurate patient phenotyping will always remain the most crucial part of studies attempting at gaining insights on novel genetic markers for drug safety and efficacy (Gurwitz and McLeod 2013). The success of personalized medicine at the level of a single individual is dependent on “tailoring” therapy based upon information obtained at a group level (population/cohort) and the specific biological, genetic, and clinical characteristics of an individual. Also, other factors need to be considered such as ethnicity, differences in life-style, and environmental factors (as well as co-existing susceptibility factors for other diseases); as these will bring additional complexity not reflected by the genetic make-up of the individual (Gurwitz and Pirmohamed 2010).

In RA, although there is not yet any guidance about which treatment will be the most suitable for any particular patient, it seems likely that genetic variation will be useful in predicting treatment response. Most likely a multi-parameter approach will be employed that utilizes a combination of predictors based on profiling DNA, RNA and proteins, with clinical and epidemiological information to establish an appropriate clinical decision-making algorithm (Acosta-Colman et al., 2013). To maximize the use of genetic information recently gained, it is necessary to obtain matched knowledge of the molecular pathogenesis of the different subsets of RA.

Clinical risk assessment for complex disease is currently based on family history, lifestyle and shared environmental factors, with a low predictive value (Khoury et al., 2010; McClellan et al., 2013). The development of assessment tools that will take advantage of individual genomic information may have a higher predictive value, allowing for translation of recent genetic research findings into clinical practice (Green and Guyer 2011).

5. Looking towards the future

The past decade has seen astonishing progress in our ability to decipher genetic and molecular factors underlying complex diseases such as RA. Meanwhile, advances in the etiology of disease and the ability to manipulate the genome should make it possible to personalize the diagnoses and treatment of RA. Developments in genomics, including NGS technologies, are expected to enable a more personalized approach to clinical care, with improved risk stratification and treatment selection based upon detailed knowledge of an individual’s genetic background.
Following the discoveries made using GWAS, the promise of NGS technologies are the analysis of the whole genome to identify rare variants, which could have larger effect sizes (Heinzen et al., 2012). Genome-wide re-sequencing efforts are now under way and will provide a more comprehensive description of the relations between genome variation and major clinical phenotypes. Despite the large numbers of GWASs performed to date in RA, a significant portion of the heritability of the disease remains to be explained (Heap and Heel 2008). Many complementary strategies have been developed in an attempt to explain this missing heritability. Also, larger cohorts of patients (and controls) obtained via international collaborations will increase the power of genetic studies and facilitate discovery. The majority of these new methods have potential to greatly improve our knowledge of the genetic architecture of RA. Exome sequencing has been very useful in rare diseases with Mendelian inheritance, and some recent studies using this method in complex autoimmune disease have been successful (i.e. in scleroderma, Gao et al., 2016).

Whilst most of the eQTL studies in RA have been done after a locus has been discovered as a risk locus (Okada et al., 2014), there is potential for this method to be used earlier on in the discovery process as a screen or filter of likely candidate loci, as autoimmune disease risk loci are enriched for eQTLs (Nicolae et al., 2010). In gene discovery efforts it seems likely that using extremes of phenotype or stratification by immune system phenotypes will also be valuable, and thus far there is precedent that completely new loci can be discovered when subphenotypes are examined (Kariuki et al., 2015).

By incorporating gene expression as a biological output related to the genetic variation, it is possible that eQTL methods could enhance pharmacogenomic efforts in RA. Similarly, the dense genotype information provided by NGS could provide additional resolution at the genetic level, although with a burden of a larger data set and greater difficulty in excluding false positive findings. It seems likely that prediction of treatment response and pharmacogenomics in RA will need to incorporate many variables to provide clinically useful prediction – including genetic, gene expression, protein expression and immune cell parameters, and clinical and demographic information.

In the last 5 years, the field of epigenetics has progressed impressively, and the role of diverse epigenetic modifications in the development of many diseases including RA has been extensively studied. In RA, these modifications notably affect synovial fibroblasts, resulting in the development of an aggressive, intrinsically activated phenotype (Klein et al., 2012). These studies suggest the possibility of epigenetic targeting agents as a potential new treatment for RA (Gray 2013). Linking genetic and epigenetic data with gene expression and proteomics/metabolics is certainly the biggest challenge we are facing, and novel tools needs to be developed to this end.

It is clear from the above that genetic testing offers great potential for understanding the biology of RA as well as for patient management and personalisation of treatment. As for any biomarker/therapy/device/test, it takes several years for any product to reach clinics, starting from basic research and initial discovery, followed by validation, clinical utility, manufacturing of a robust test, and its final approval by regulatory authorities. The use of genetic information will not escape this long winding road, but work done in the cancer field to link particular genetic variants with clinical decisions and outcomes has paved the way for many other diseases to take this journey.
Conflict of interest statement

All authors have read the journal's policy on disclosure of potential conflict of interest.

Acknowledgements

The authors wish to acknowledge the support of the European Union Seventh Framework Programme IMI-funded project BeTheCure, contract no. 115142-2. Although it was tried to refer as much as possible to the original research papers, it is impossible to cite all of the more than 1000 papers on RA genetics, RA Pharmacogenomics and new technologies used in these topics, listed in Pubmed today, considering space limitation. The authors therefore apologize to all researchers whose original articles were not directly cited.

References


TRAF1–C5 locus on chromosome 9 with juvenile idiopathic arthritis. Arthritis Rheum. 58, 2206–2207.


Plenge, R.M., Seielstad, M., Padyukov, L., Lee, A.T., Remmers, E.F., Ding, B., Liew, A., Khalili, H., Chandrasekaran, A., Davies, L.R., Li, W., Tan, A.K.,


<table>
<thead>
<tr>
<th>Name</th>
<th>Principle</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanger sequencing</td>
<td>enzymatic dideoxy DNA sequencing technique based on the chain-terminating dideoxynucleotide analogues</td>
<td>Sanger et al., 1977</td>
</tr>
<tr>
<td>Allan Maxam and Walter Gilbert sequencing</td>
<td>chemical degradation DNA sequencing technique in which terminally labeled DNA fragments were chemically cleaved at specific bases and separated by gel electrophoresis</td>
<td>Maxam and Gilbert 1977</td>
</tr>
<tr>
<td>Automated sequencing</td>
<td>separated fluorescent bands of DNA are detected near the bottom of a tube, and the sequence information is acquired directly by computer.</td>
<td>Smith et al., 1986</td>
</tr>
<tr>
<td>Pyrosequencing</td>
<td>depends on the detection of pyrophosphate release on nucleotide incorporation</td>
<td>Ronaghi et al., 1996</td>
</tr>
<tr>
<td>SOLiD</td>
<td>repeated rounds of ligation and cleavage of oligonucleotide probes to amplify DNA fragments in an emulsion PCR with small magnetic beads</td>
<td>Mardis 2008</td>
</tr>
<tr>
<td>combinatorial probe-anchor ligation (cPAL)</td>
<td>repeated rounds of ligation and cleavage of oligonucleotide probes</td>
<td>Drmanac et al., 2009</td>
</tr>
<tr>
<td>GWAS</td>
<td>detects genetic variation that is relatively common, e.g., genetic variants that have frequencies of 5% or more in the general population.</td>
<td>Hardy and Singleton 2009</td>
</tr>
<tr>
<td>single-molecule, real-time SMRT</td>
<td>Ion semiconductor sequencing that is based on the detection of hydrogen ions released during DNA synthesis or to Pyrosequencing, a form of sequencing by synthesis that relies on the detection of pyrophosphate released upon nucleotide incorporation into the nascent DNA strand</td>
<td>Nejentsev et al., 2009</td>
</tr>
<tr>
<td>Taqman PCR</td>
<td>real-time PCR based high throughput genotyping technique</td>
<td>Popa et al., 2009</td>
</tr>
<tr>
<td>Genome Analyzer (GA)</td>
<td>utilizes a form of ‘sequencing by-synthesis’ technology</td>
<td>Gravina et al., 2013</td>
</tr>
</tbody>
</table>

**Table 1:** DNA genotyping technologies from the origin to next generation
<table>
<thead>
<tr>
<th>Drug</th>
<th>Gene</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methotrexate</td>
<td>ATIC, SLC19A1, GGH,</td>
<td>de novo purine synthesis, transports folate and MTX into cell, catalyzes</td>
<td>Owen et al., 2012, Salazar et al., 2014,</td>
</tr>
<tr>
<td></td>
<td>DHFR, FPGS, MTHFR,</td>
<td>the hydrolysis of folypoly-gamma-glutamates, reductase converting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SHMT1, ABCB1, IL1RN</td>
<td>dihydrofolate into tetrahydrofolate, maintains cytosolic and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TMS</td>
<td>mitochondrial folypolyglutamate, generation of 5-methyl tetrahydrofolate</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>efflux pump, blocks induction of inflammation by IL-1, de novo</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pyrimidine synthesis, B cell inhibitor</td>
<td></td>
</tr>
<tr>
<td>Leflunomide</td>
<td>DHODH, C19A, ESRI,</td>
<td>de novo pyrimidine synthesis, unknown, estrogen receptor</td>
<td>Powlik et al., 2004, Mirkov et al., 2013,</td>
</tr>
<tr>
<td></td>
<td>CYP1A2</td>
<td>cytochrome P450</td>
<td>Dziedziejko et al., 2010, Bolhance et al., 2008</td>
</tr>
<tr>
<td>Tocilizumab</td>
<td>DHFR, CCDC32, EPHA,</td>
<td>reductase converting dihydrofolate into tetrahydrofolate, unknown,</td>
<td>Mirkov et al., 2013, Kurko et al., 2013, Kurko</td>
</tr>
<tr>
<td></td>
<td>IRAV8</td>
<td>cytochrome P450</td>
<td>et al., 2013</td>
</tr>
<tr>
<td>Anti-TNFs</td>
<td>HLA-DRB1, PTPRC,</td>
<td>shared epitope, HLA gene, regulates BCR and TCR signaling</td>
<td>Potter et al., 2009, Qui et al, 2010</td>
</tr>
<tr>
<td></td>
<td>PTPN22, AFF3, CD226,</td>
<td>Negative regulator of T cell activation, transcription factor for</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PDE3A-SLC01C1, PL2G4A</td>
<td>lymphpoid development, type 1 membrane protein, phosphodiesterase /</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NUBPL, CNTN5, VAV1,</td>
<td>cell membrane protein, assembly of the NADH dehydrogenase, member of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPRED2, PDZD2, EYA4,</td>
<td>immunoglobulin superfamily, T cells protection from apoptosis,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CENTD1, MAFB2, LASS6,</td>
<td>involved in intracellular signaling, formation of axon connections,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IFNK, CSTS, GBP6,</td>
<td>influences insulin secretion, co-transcription factor, cysteine protease</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PON1, LMO4, MOBKL2B,</td>
<td>hydrolyzes GTP to both GDP and GMP, hydrolyzes organophosphate pesticides</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C9orf72, C9orf72,</td>
<td>and nerve gasses, transcriptional regulator, spindle pole body duplication</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DRC1, FOXP1, MAPK1,</td>
<td>membrane trafficking regulation, unknown, regulates tissue- and cell</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MAPK14, TANK,</td>
<td>type-specific gene transcription, serine/threonine kinase of signal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TNFRSF1B, TGFBI,</td>
<td>transduction pathway, TNF-receptor, NF-kappa-B activator, cell signaling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FCGR3A, BAFP</td>
<td>protein involved in systemic inflammation, regulates the NF-κB transcription</td>
<td></td>
</tr>
<tr>
<td>Rituximab</td>
<td></td>
<td>factor, Toll/IL-R immune signal transduction, key enzyme in</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>prostaglandin biosynthesis, membrane protein involved in immune system,</td>
<td></td>
</tr>
<tr>
<td>Sulfasalazine</td>
<td></td>
<td>pathogen recognition and activation of innate immunity, involved in</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>stress and inflammatory responses, phophorylation and activation of p38</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAP kinase, B cells and immunoglobulin secretion inhibitor, receptor for</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IgG, activation factor of TNF, acetylation of sulfapyridine</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2:** Overview of the drugs used for the treatment of Rheumatoid Arthritis, the genes involved in the Pharmacogenomics of these drugs, their function and pertinent references.